



École doctorale : **Mathématiques, Sciences de l'Information et de l'Ingénieur ED 269**

Unité de recherche : **ICube - CNRS UMR 7357**

THÈSE

Rédigée par : **Thibaut Eloy**

Pour obtenir le grade de : **Docteur de l'Université de Strasbourg**

Discipline/Spécialité : **Informatique**

Soutenue le 20/09/2024

Reconstruction à partir de particule isolée en microscopie par fluorescence

Thèse dirigée par :

BAUDRIER Étienne

Maître de conférence, Université de Strasbourg

Encadrant :

FORTUN Denis

Chargé de recherche CNRS, Laboratoire ICUBE

Rapporteurs :

**KERVANN Charles
DENIS Loïc**

Directeur de recherche INRIA, Institut Curie
Professeur, Université de Saint-Etienne

Examineurs :

**SOUBIÈS Emmanuel
JONIC Slavica
LIEBLING Michaël**

Chargé de recherche CNRS, Laboratoire IRIT
Directrice de recherche CNRS, Laboratoire IMPMC
Chercheur Senior, Idiap Research Institute

Glossaire

Fluofire Méthode de reconstruction hétérogène développée dans le chapitre 4.

ShiReGmm Méthode de reconstruction homogène développée dans le chapitre 3, qui utilise une représentation sous forme de mélange de gaussiennes du volume.

ShiReGmm2 Méthode de reconstruction homogène développée dans le chapitre 3, qui utilise une représentation sous forme de mélange de gaussiennes du volume et des vues.

SHiReVol Méthode de reconstruction homogène développée dans le chapitre 3, qui utilise une représentation voxelique du volume.

ACP analyse en composante principale.

cFSC conical fourier shell correlation.

CLSM confocal laser scanning microscopy.

cryo-ME microscopie électronique cryogénique.

CTF contrast transfer function.

DL degré de labellisation.

ECS escalade de la colline stochastique.

EM espérance-maximisation.

FSC fourier shell correlation.

iU-ExM iterative ultrastructure expansion microscopy.

MLP multilayer perceptron.

PSF point spread function.

RMN résonance magnétique nucléaire.

RPI reconstruction à partir de particules isolées.

RSB rapport signal à bruit.

SIM structured illumination light sheet microscopy.

SIREN sinusoidal representation network.

SMLM single-molecule localization microscopy.

SSIM structural similarity.

STED stimulated-emission-depletion.

U-ExM ultrastructure expansion microscopy.

Résumé

La microscopie par fluorescence, technique de microscopie optique permettant l'acquisition d'images tridimensionnelles, a connu des avancées majeures ces dernières décennies grâce au développement des méthodes de super-résolution optique et de la microscopie par expansion. Cette technique de microscopie commence même à émerger comme un outil prometteur pour la biologie structurale, discipline qui étudie la structure tridimensionnelle des assemblages macromoléculaires afin de comprendre leurs rôles dans les mécanismes cellulaires.

Néanmoins, ces progrès sont freinés par deux limitations principales. D'une part, l'anisotropie de résolution : la résolution selon l'axe du microscope est de 3 à 5 fois inférieure à la résolution dans le plan latéral. D'autre part, le faible degré de labellisation (DL) : les fluorophores utilisés pour labelliser les protéines d'intérêts d'un assemblage macromoléculaire, ne couvrent pas uniformément la structure observée. L'anisotropie de résolution et le faible DL sont des limitations physiques intrinsèques au processus d'acquisition. L'objectif de la thèse est de les contourner à l'aide de l'outil numérique en développant une méthode de reconstruction à partir de particules isolées (RPI). Le principe consiste à combiner plusieurs acquisitions de particules biologiques identiques observées selon différentes orientations inconnues afin de reconstruire un modèle de particule. Grâce à la fusion des informations complémentaires dans chaque vue, nous visons à reconstruire un volume avec une résolution isotrope et une densité de labellisation uniforme. Dans le cadre de ma thèse, nous nous sommes concentrés sur les modalités d'acquisition convolutionnelles : les vues sont modélisées comme la convolution du modèle de particule par la réponse impulsionnelle du microscope.

Aujourd'hui, la microscopie par fluorescence est encore peu utilisée en biologie structurale, car les progrès qui ont permis d'améliorer la résolution des images sont relativement récents. C'est pourquoi, il existe très peu de méthodes de RPI en microscopie par fluorescence. En revanche, la cryomicroscopie électronique (cryo-ME) est un outil bien plus établi en biologie structurale, et de nombreuses méthodes de RPI dédiées à la cryo-ME ont déjà été développées. Nous nous sommes donc en partie inspiré de ces méthodes, en les adaptant au cas particulier de la microscopie par fluorescence.

Les méthodes de RPI (en microscopie par fluorescence ou en cryo-ME) peuvent être appliquées à des données homogènes ou hétérogènes. La reconstruction homogène considère que toutes les vues sont issues du même modèle de particule. En revanche, la reconstruction hétérogène ne fait pas cette hypothèse, et chaque vue peut être issue d'un état différent de la particule considérée. C'est ce qu'on appelle l'hétérogénéité des conformations. Dans ce cas, la reconstruction consiste non seulement à estimer les poses associées à chacune des vues, mais également l'ensemble (pouvant être discret ou continu) des conformations dont elles sont issues.

Je propose dans ma thèse deux contributions. La première est le développement et l'analyse d'une méthode de RPI homogène en microscopie par fluorescence. Cette dernière a fait l'objet, en 2023, d'une publication dans le journal IEEE Transactions on Computational Imaging. Elle se

base sur une reformulation du problème d’optimisation conjoint original, par rapport au volume, aux rotations et aux translations en un schéma d’optimisation hiérarchique qui découple la minimisation par rapport à chaque groupe de variables, en 3 niveaux imbriqués. Le volume est estimé au niveau le plus élevé grâce à de l’optimisation stochastique, et le succès de notre approche réside dans le développement d’optimiseurs approximatifs rapides pour la minimisation par rapport aux rotations et translations aux niveaux inférieurs. Grâce à cette approche, le coût de calcul d’une reconstruction est réduit à moins de 30 minutes, et ne requiert pas d’initialisation. De plus, notre approche est peu sensible aux hyper-paramètres, ce qui simplifie son utilisation en pratique. Nous démontrons sur plusieurs ensembles de données synthétiques que nous surpassons l’état de l’art en termes de résolution et d’évaluation visuelle.

La seconde contribution est le développement et l’analyse d’une méthode de RPI hétérogène en microscopie par fluorescence. Cette dernière généralise à la 3D la méthode CryoFIRE publiée en 2022 par Levy et al. Comme eux, nous nous basons sur une architecture prenant la forme d’un système encodeur-décodeur entraîné de bout en bout et qui utilise une représentation dans un espace à faible dimension de l’hétérogénéité des conformations. Ce système repose sur une estimation amortie des poses et du paramètre d’hétérogénéité (l’encodeur), et une représentation implicite du volume à l’aide d’un réseau de neurones densément connecté (le décodeur). La différence avec CryoFIRE est que nous adaptons l’architecture au modèle d’acquisition de la microscopie par fluorescence. Outre cette adaptation, nos contributions sont l’ajout d’une contrainte pose à pose qui permet d’améliorer la robustesse de l’estimation des poses, la prise en compte de l’information a priori de symétrie C9 du centriole pour affiner la reconstruction et la reconstruction d’un modèle paramétrique dans le cas où peu de données réelles sont disponibles. Ce travail forme le contenu d’un article en cours de rédaction.

Table des matières

1	Contexte et problématique	8
1.1	Biologie structurale	8
1.2	Cristallographie aux rayons X, spectroscopie à résonance magnétique nucléaire (RMN) et microscopie électronique cryogénique (cryo-ME)	9
1.3	Microscopie par fluorescence	10
1.4	Problématique	15
1.4.1	Reconstruction homogène et hétérogène	17
1.4.2	Cas d'application : le centriole	19
2	État de l'art	22
2.1	Reconstruction en microscopie par fluorescence	22
2.2	Reconstruction en cryo-ME	23
2.2.1	Estimation conjointe des poses et du volume	24
2.2.2	Marginalisation par rapport aux poses	26
2.2.3	Apprentissage profond	31
2.2.4	Hétérogénéité discrète des conformations	34
2.2.5	Hétérogénéité continue des conformations	35
2.3	Comparaison des modalités cryo-ME et microscopie par fluorescence	42
3	Reconstruction homogène	44
3.1	Positionnement et contributions	44
3.2	Formulation du problème	45
3.2.1	Prérequis sur la représentation des rotations	45
3.2.2	Modèle d'observation	46
3.2.3	Problème d'optimisation joint	46
3.3	Méthode	47
3.3.1	Reformulation multi-niveaux du problème d'optimisation	47
3.3.2	Optimisation par rapport au volume	48
3.3.3	Optimisation par rapport aux orientations	48
3.3.4	Optimisation par rapport aux translations	50
3.3.5	Critère d'arrêt	53
3.4	Variante de la méthode : représentation du volume sous forme d'un mélange de gaussiennes	54
3.5	Deuxième variante : représentation des vues sous forme d'un mélange de gaussiennes	57
3.5.1	Principe	57
3.5.2	Méthode pour représenter les vues sous une forme de mélange de gaussiennes	59
3.6	Résultats	61

3.6.1	Résultats sur données simulées	61
3.6.2	Résultats sur des données réelles	84
3.7	Résumé du chapitre et perspectives	85
3.7.1	Résumé	85
3.7.2	Perspectives méthodologiques	87
4	Reconstruction hétérogène	89
4.1	Positionnement et contributions	89
4.1.1	Cas d'application : hétérogénéité compositionnelle	89
4.1.2	Adaptation de CryoFIRE	90
4.1.3	Apprentissage pose à pose	90
4.1.4	Information <i>a-priori</i> sur la symétrie de la reconstruction	90
4.1.5	Reconstruction d'un modèle paramétrique	91
4.2	Méthode	91
4.2.1	Formulation du problème	91
4.2.2	Modèle	92
4.2.3	Apprentissage pose à pose	95
4.2.4	Imposer la symétrie C9 à la reconstruction	97
4.2.5	Imposer un modèle paramétrique à la reconstruction	98
4.3	Résultats sur des données générées synthétiquement	98
4.3.1	Reconstruction homogène	98
4.3.2	Reconstruction hétérogène	99
4.4	Résultats obtenus sur des données réelles	110
4.4.1	Description des données	110
4.4.2	Reconstruction monocanal de la Tubuline	114
4.4.3	Reconstruction bicanal Tubuline/Sas6	116
4.5	Imposer un modèle paramétrique à la reconstruction	116
4.5.1	Génération du tore à partir de ces paramètres	117
4.5.2	Validation sur des données synthétiques	119
4.5.3	Résultats obtenus sur des données réelles	121
4.6	Résumé du chapitre et perspectives méthodologiques	123
4.6.1	Résumé	123
4.6.2	Perspectives méthodologiques	124
5	Discussions et perspectives	127
5.1	Bilan	127
5.2	Perspectives générales	128
5.2.1	Acquisition de données	128
5.2.2	Logiciel Scipion	129
5.2.3	Estimation de la réponse impulsionnelle	129
5.2.4	Prise en compte de la déformation liée à l'expansion	129
6	Appendices	131
6.1	Calcul des gradients dans le cas d'une représentation voxélique	131
6.2	Calcul des gradients dans le cas d'une représentation sous forme de mélange de gaussiennes	131
6.3	Distance entre mélanges de gaussiennes	133

6.4	Projet annexe sur le pore nucléaire	135
-----	---	-----

Chapitre 1

Contexte et problématique

1.1 Biologie structurale

Ma thèse s’inscrit dans le contexte général de la biologie structurale, discipline qui étudie la structure tridimensionnelle des protéines et assemblages macromoléculaires. Un assemblage macromoléculaire est une structure 3D stable énergétiquement constituée de diverses protéines, liées entre elles par des liaisons non-covalentes, c’est-à-dire des liaisons moléculaires ne mettant pas en jeu le partage d’électrons entre atomes voisins mais plutôt des forces électromagnétiques. Les protéines sont elles-mêmes constituées d’acides aminés, liés entre eux à la fois par des liaisons covalentes (ce qui forme la chaîne polypeptidique), mais aussi via des liaisons non-covalentes. La cellule, unité de base du vivant, contient dans son cytoplasme et son noyau de nombreuses protéines et assemblages macro-moléculaires, qui par le biais de leurs interactions, donnent lieu à des mécanismes complexes essentiels à la vie et la reproduction de la cellule. Ces mécanismes sont par exemple le transport de molécules de l’extérieur vers l’intérieur de la cellule (endocytose), la libération de molécules de l’intérieur vers l’extérieur (exocytose), les échanges de molécules entre le noyau et le cytoplasme (translocation), la division cellulaire (mitose), la production de protéines à partir de séquences de nucléotide (transcription), les mécanismes épigénétiques de contrôle de l’expression des gènes, la différenciation cellulaire, l’autodestruction de la cellule en réponse à un signal (apoptose), l’ingestion de particules étrangères (phagocytose) et bien d’autres encore. Chaque assemblage macromoléculaire ou protéine isolée admet une ou plusieurs fonctions nécessaires au bon fonctionnement d’un ou plusieurs mécanismes cellulaires. Par exemple, le ribosome transcrit le code de l’ARN messager en protéines, le centriole joue un rôle majeur dans la mitose notamment en permettant la séparation des brins d’ADN, le pore nucléaire permet de réguler les échanges entre le noyau et le cytoplasme, les vésicules transportent des produits et déchets cellulaires le long des fibres du cytosquelette (réseau de fibres qui confère sa structure à la cellule), le récepteur AMPA¹ laisse passer certains ions à travers la membrane cellulaire lorsqu’il reçoit un signal d’un messenger chimique. Déchiffrer les structures de ces assemblages macromoléculaires est indispensable pour comprendre leurs rôles dans les mécanismes cellulaires, permettant ainsi d’élucider certains mystères de la vie.

L’observation de la première cellule vivante est attribuée à Robert Hooke en 1665, identifiée à l’aide d’un microscope optique qu’il avait lui-même fabriqué, et qui permettait d’obtenir un grossissement d’un rapport 30 [1]. Par la suite, Antoni Van Leeuwenhoek améliora considérablement la

1. AMPA pour α -amino-3-hydroxy-5-méthylisoxazol-4-propionate

résolution des microscopes de l'époque, et on lui attribue la découverte des protistes², des bactéries et des spermatozoïdes [2]. Ces découvertes pionnières ont initiées des avancées majeures dans notre compréhension du vivant. Elles ont par exemple fourni des preuves tangibles à la théorie microbienne des maladies [3], affirmant que des micro-organismes peuvent être à l'origine de nombreuses maladies. Cette théorie s'opposait à la théorie des miasmes de l'époque (aujourd'hui infirmée), qui supposait que des maladies telles que le choléra, la malaria ou la peste noire venaient d'une forme nocive de mauvais air. Elles ont également donné lieu, au courant du XIX^{ème} siècle, aux axiomes de la théorie cellulaires énoncés par Matthias Jakob Schleiden, Theodor Schwann et Rudolf Virchow [4]. Les 3 principaux axiomes sont les suivants : 1) tous les organismes sont constitués de petites unités : les cellules, 2) Toute cellule provient d'une autre cellule, 3) La cellule est non seulement l'unité de base du vivant, mais est aussi une unité vivante autonome capable de réaliser elle-même un certain nombre de fonctions nécessaires et suffisantes à sa vie. Cette autonomie relative³ est permise par la membrane plasmique qui filtre les échanges avec l'environnement extérieur, et garantit l'individualité de la cellule.

Néanmoins, les avancées en microbiologie ont longtemps été restreintes par la limite de diffraction des microscopes optiques traditionnels (dont on reparlera plus loin), qui permettaient uniquement l'observation, avec très peu de détails, des assemblages macromoléculaires les plus gros de la cellule (comme le centriole) ainsi que de ses organites⁴. Les premières réelles avancées de la biologie structurale, qui ont contribué à l'observation des détails de structures complexes, datent du XX^{ème} siècle avec le développement de la cristallographie aux rayons X, la spectroscopie à RMN (résonance magnétique nucléaire), la cryo-ME (microscopie électronique cryogénique) et, vers la fin du siècle, des techniques de super-résolution en microscopie optique.

1.2 Cristallographie aux rayons X, spectroscopie à RMN et cryo-ME

Historiquement, les principaux apports en biologie structurale viennent de la cristallographie aux rayons X [6], née au début du XX^{ème} siècle, grâce à la découverte en 1912 par Max Von Laue et ses collègues de la possibilité de diffracter les rayons X avec un cristal. Ces travaux, ainsi que ceux de William Henry Bragg et William Lawrence Bragg sur la mise en équation du phénomène, ont donné les bases à la cristallographie aux rayons X qui a finalement permis en 1957 d'imager la première structure d'une macromolécule, à une résolution de 6 Angström (Å) : la myoglobine⁵ [7]. Elle a par la suite permis de révéler la structure atomique de l'architecture ADN [8], et a aussi joué un rôle majeur dans la conception de nouveaux médicaments et de biomatériaux [9, 10]. Elle reste néanmoins limitée par l'impossibilité d'observer des assemblages macromoléculaires trop complexes [11].

Parallèlement, la spectroscopie RMN ainsi que la cryo-ME se sont développées et sont venues

2. Les protistes sont des micro-organismes eucaryotes à organisation cellulaire simple (unicellulaire ou multicellulaire avec peu de cellules)

3. Relative parce que la cellule est un système ouvert qui autorise les transferts de matière et d'énergie avec l'environnement extérieur, et c'est justement cette ouverture qui permet son organisation interne, contrairement à par exemple une chaise qui est un système fermé. Ce phénomène est autrement appelé "auto-éco-organisation" par le philosophe de la complexité Edgar Morin [5]

4. Un organite est une structure spécialisée, contenue dans le cytoplasme, et délimitée du reste de la cellule par une membrane (par exemple le noyau, les mitochondries, l'appareil de Golgi)

5. la myoglobine est une protéine assurant une fonction de transfert et de stockage de l'oxygène

compléter les apports de la cristallographie aux rayons X à la biologie structurale. La spectroscopie RMN repose sur l'interaction entre les noyaux atomiques et un champ magnétique, qui provoque la résonance des noyaux à une fréquence spécifique. En mesurant les fréquences de résonance, on peut déterminer la composition chimique de l'échantillon.

Le principe de la cryo-ME repose sur la congélation de l'échantillon à observer. Celui-ci est d'abord purifié⁶ puis plongé dans un bain d'éthane liquide à une température extrêmement basse (de l'ordre de -180°C). La congélation rapide permet de le préserver dans un état vitreux, évitant ainsi la formation de cristaux de glace qui pourraient endommager la structure. L'échantillon congelé est ensuite transféré dans un microscope électronique, et entouré de vide. Pour acquérir les images, un faisceau d'électrons est dirigé vers l'échantillon et leur interaction avec ce dernier génère des signaux qui sont détectés, après passage à travers une lentille magnétique, générant ainsi une projection 2D de l'échantillon [11, 12]. Deux grandes catégories de méthodes numériques permettent ensuite de reconstruire l'objet 3D original à partir d'un ensemble de projections issues de cet objet [6]. La première est la reconstruction à partir de particules isolées (RPI). Les données d'entrées sont de larges images contenant un grand nombre de projections de particules (*i.e.* protéines ou assemblages macromoléculaires) identiques orientées aléatoirement. Ces projections sont d'abord détectées⁷ puis le volume et les poses (orientations et translations associées à chaque vues) sont conjointement estimées à partir de cet ensemble de vues. La seconde est la cryo-tomographie, où contrairement à l'analyse de particules isolées, une seule particule est observée. L'échantillon est tourné en différentes orientations, et une image est acquise pour chacune d'entre elles. Dans ce cas, l'estimation des poses n'est pas nécessaire. Dans les deux cas, le modèle d'observation se formalise de la manière suivante :

$$y_l = CTF \otimes \mathcal{T}_{t_l}(\mathcal{P}(\mathcal{R}_{\theta_l}(f))) + \epsilon, \quad (1.1)$$

où y_l est la $l^{\text{ème}}$ vue, $\mathcal{R}_{\theta_l}$ la rotation d'angle $\theta_l \in \mathbb{R}^3$, $\mathcal{T}_{t_l}$ la translation de vecteur $t_l \in \mathbb{R}^2$, $\mathcal{P}$ la projection selon l'axe du microscope, CTF la fonction de transfert de contraste du microscope, modélisant le passage à travers la lentille magnétique, et ϵ un bruit blanc.

La cryo-ME présente l'avantage de ne pas nécessiter la cristallisation de l'échantillon, ce qui permet d'observer des structures plus complexes sans les dégrader [13]. Longtemps limitée à quelques dizaines d'Angströms, la résolution en cryo-ME a connu d'importants progrès ces dernières années, lui permettant, en 2013, d'égaler la cristallographie aux rayons-X en termes de résolution [11]. Même si la cryo-ME permet d'imager des structures plus complexes que la cristallographie aux rayons-X, la préparation de l'échantillon peut dégrader la structure si cette dernière dépasse plusieurs centaines de mégadaltons. En outre, cette méthode ne permet pas, contrairement à la microscopie par fluorescence, de cibler des protéines spécifiques de l'assemblage macromoléculaire observé.

1.3 Microscopie par fluorescence

Contrairement à la cryo-ME, ou la cristallographie aux rayons-X, la microscopie par fluorescence, technique de microscopie optique 3D, [14] permet l'observation de structures de grande taille sans les dégrader [8]. Elle se base sur l'utilisation de marqueurs fluorescents, chacun composé d'un anticorps et d'un fluorophore. Le choix de l'anticorps est effectué en fonction du type

6. La purification permet de conserver uniquement les particules d'intérêt

7. La densité de spécimen présents dans l'échantillon permet que leurs images ne soient pas superposées dans l'acquisition, ce qui justifie le nom de "particules isolées"

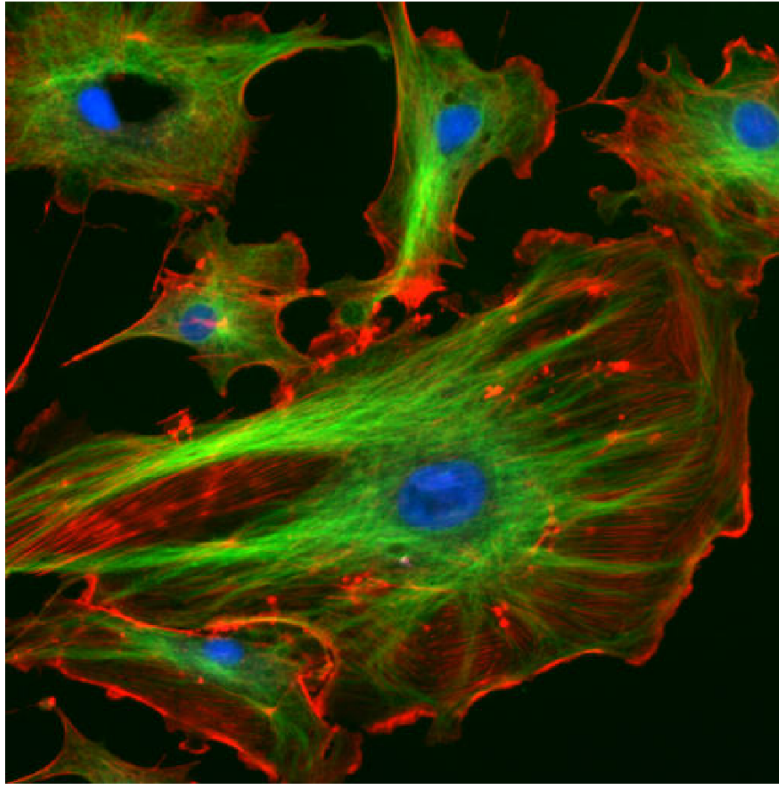


FIGURE 1.1 – Exemple de microscopie à fluorescence : cellules endothéliales d’une artère pulmonaire bovine (Cell Image Library)

de protéine que l’on souhaite observer. L’utilisation de différents types de marqueurs donne alors lieu à l’obtention d’images multi-canaux, chaque canal étant associé à une protéine particulière, comme illustré sur la figure 1.1 représentant une image à trois canaux.

Plus précisément, les marqueurs sont introduits dans l’échantillon et les anticorps s’attachent aux protéines ciblées, elles sont alors dites ‘labellisées’. L’échantillon est ensuite éclairé à une longueur d’onde choisie de sorte à faire passer les fluorophores de leur état fondamental à un état excité. Ils dissipent rapidement une partie de l’énergie absorbée, par collision avec les molécules voisines, passant alors dans un état excité d’énergie inférieure. Finalement, ils retournent dans leur état fondamental, générant l’émission spontanée d’un photon d’énergie inférieure, donc de longueur d’onde supérieure à celle utilisée pour l’excitation. La lumière ainsi émise est détectée par un capteur, ce qui permet de localiser les protéines ciblées. Ce phénomène d’excitation, puis de retour à l’état fondamental est illustré sur la figure 1.2. La figure 1.3 donne le schéma général d’un microscope à fluorescence. L’élément central est le miroir dichroïque qui réfléchit la longueur d’onde d’excitation mais laisse passer la lumière émise par les marqueurs fluorescents. Par ailleurs, un filtre d’entrée permet de sélectionner la longueur d’onde d’excitation exacte et un filtre de sortie de s’assurer que seule la lumière émise par les marqueurs arrive jusqu’au détecteur.

Il existe différents types de microscopes à fluorescence, le plus simple d’entre eux est le microscope à champ large. L’échantillon entier est éclairé en une seule fois par des rayons de lumière parallèles. Cela permet de visualiser simultanément l’ensemble des protéines marquées situées dans le plan focal de l’objectif. L’échantillon est déplacé mécaniquement à l’aide d’une platine motorisée dans la direction axiale (la direction de l’axe du microscope) permettant d’acquérir un ensemble d’images 2D qui sont concaténées pour former une image 3D. Si cette technique est simple à

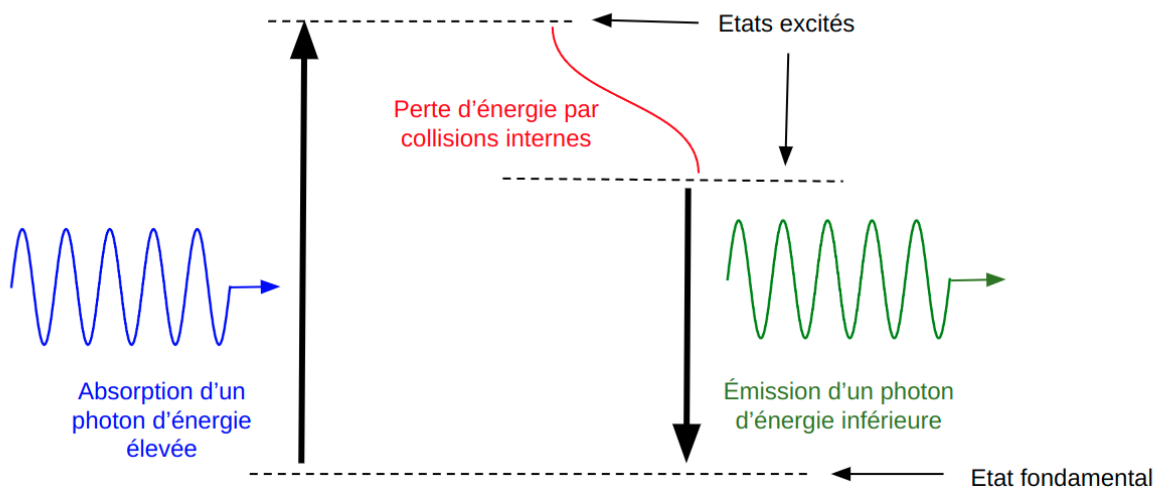


FIGURE 1.2 – Excitation et retour à l'état fondamental d'un marqueur fluorescent

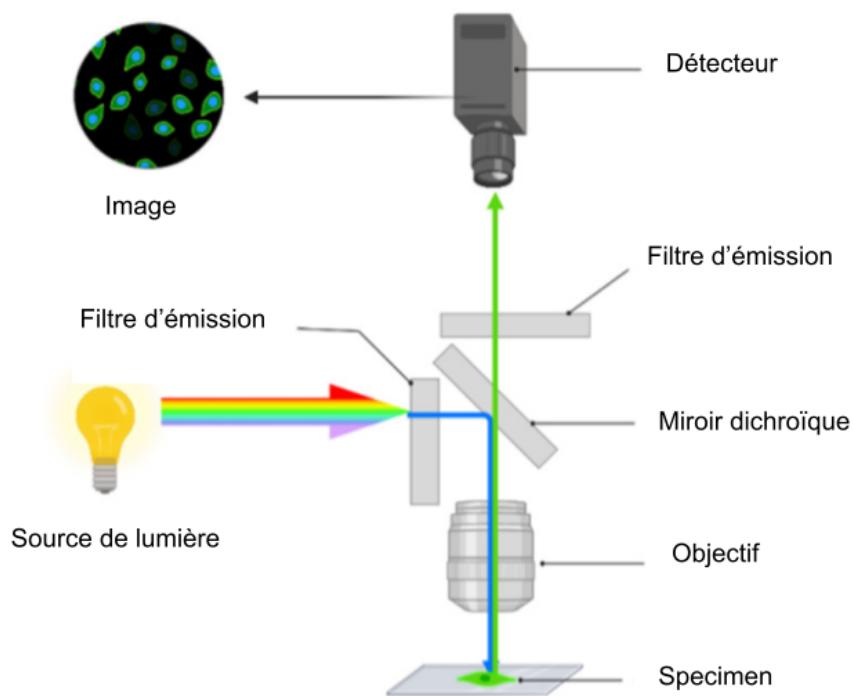


FIGURE 1.3 – Schéma d'un microscope par fluorescence. Image extraite de [15]

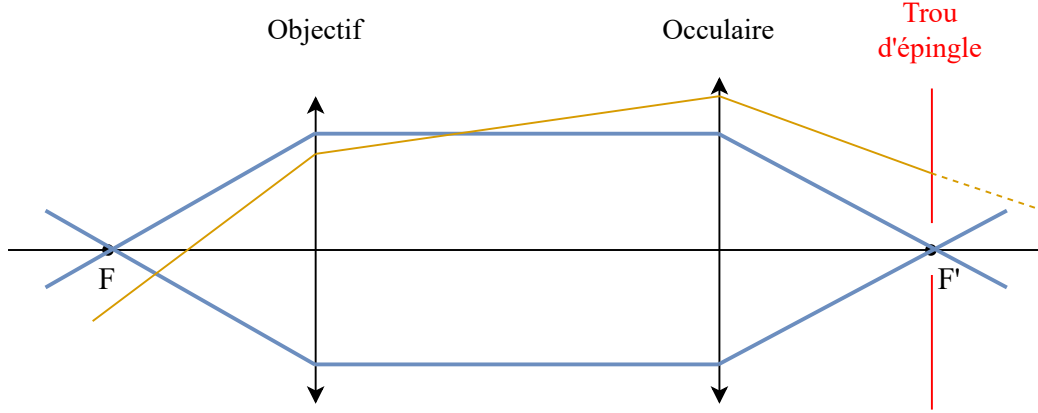


FIGURE 1.4 – Tracé de rayons d'un microscope confocal. La lumière ne provenant pas du point focal objet de l'objectif F (ou d'un voisinage proche) ne passe pas à travers le trou d'épingle, placé dans le plan focal image de l'oculaire (qui passe par F')

mettre en place et rapide à exécuter, elle reste très limitée en termes de résolution spatiale, particulièrement dans la direction axiale, car une partie non négligeable de la lumière ne provenant pas du plan focal atteint le détecteur [15], et dégrade le signal.

C'est pour pallier ces limitations que Marvin Minsky inventa, en 1955, le microscope confocal [15], [16]. La différence majeure réside dans l'utilisation d'un trou d'épingle, qui est placé dans le plan conjugué au plan focal objet de l'objectif (c'est-à-dire dans le plan focal image de l'oculaire, voir figure 1.4). Par conséquent, il bloque la lumière parasite en laissant passer uniquement les rayons provenant du point focal de l'objectif. L'échantillon est balayé point par point à l'aide d'un laser dont la cible est réglée par deux miroirs motorisés, et l'ensemble des différents points sont assemblés pour former l'image 3D finale. La taille du trou d'épingle doit être judicieusement choisie, il ne doit être ni trop petit pour limiter la diffraction de la lumière qui le traverse, ni trop grand pour éviter de laisser passer de la lumière ne provenant pas du point focal.

Malgré les avancées apportées par le microscope confocal, la résolution des microscopes à fluorescence est longtemps restée contrainte par la limite de diffraction de la lumière. En effet, de part le phénomène de diffraction, l'image d'un point matériel par le microscope n'est pas un point mais une tache d'Airy, un ensemble de ronds concentriques dont l'intensité décroît avec la distance au centre. Par conséquent, les images de deux points trop proches l'un de l'autre dans l'échantillon observé ne sont pas distinguables. La résolution latérale d'un microscope confocal traditionnel est limitée par le critère de Rayleigh [17] :

$$R = \frac{\lambda}{2n \sin(\alpha)} = \frac{\lambda}{2NA} \quad (1.2)$$

où λ représente la longueur d'onde de la lumière utilisée, $NA = n \sin(\alpha)$ l'ouverture numérique, n l'indice de réfraction du milieu et α l'angle de demi-ouverture de l'objectif. En pratique, les biologistes utilisent une huile d'immersion ayant un indice de réfraction supérieur à celui de l'air ambiant, de l'ordre de $n = 1,5$. Pour $n = 1,5$, NA ne peut dépasser cette même valeur, par conséquent pour des longueurs d'onde appartenant au spectre de la lumière visible, entre 400 et 700 nm, la résolution est au mieux entre 130 et 230 nm.

Néanmoins, depuis le début des années 90, des techniques de super-résolution optique, telle que la déplétion par émission stimulée (stimulated-emission-depletion (STED)) [17], ont permis

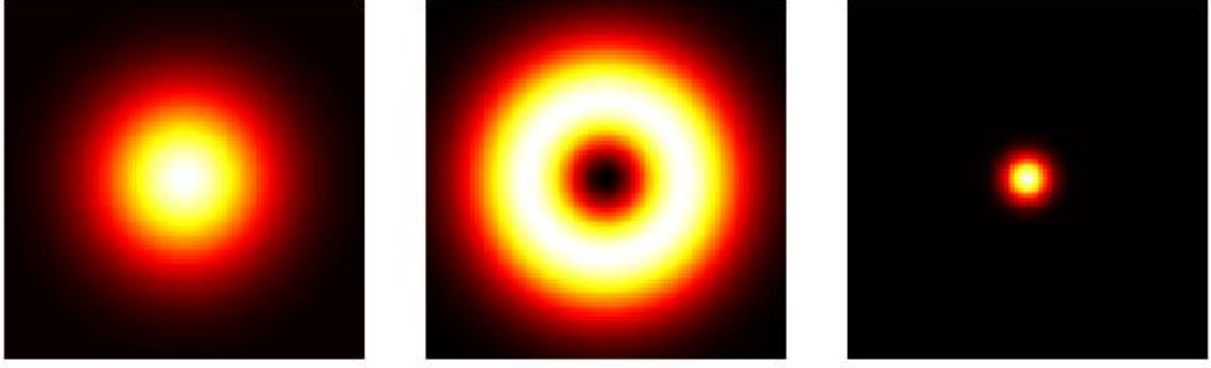


FIGURE 1.5 – Illustration du principe de la microscopie STED. A Gauche : le faisceau d'excitation, au centre le faisceau de désexcitation et à droite la fluorescence résultante. Crédit : Marcel Lauterbach

de dépasser cette limite de résolution et d'atteindre pour les meilleurs microscopes une résolution de l'ordre de 10 nm [18]. Le principe de la microscopie STED repose sur l'utilisation de deux faisceaux laser. Le premier faisceau, similaire à celui utilisé par le microscope confocal, excite les molécules fluorescentes de l'échantillon, les amenant à un état énergétique élevé. Le second faisceau, le faisceau STED, a une forme de tore et est conçu pour annuler la fluorescence émise par toutes les molécules, sauf celles situées dans une région très petite et centrale. Les faisceaux d'excitation et de désexcitation ainsi que la fluorescence résultante sont illustrés sur la figure 1.5. La figure 1.6 illustre le même échantillon observé par microscopie confocale et par STED. Plus récemment, en 2015, des chercheurs ont développé une méthode de microscopie par expansion [19], basée sur l'utilisation d'un gel faisant grossir l'échantillon, ce qui permet d'observer des détails invisibles avec un microscope confocal classique. Elle a depuis connu différentes variations et améliorations [20]. Aujourd'hui, la microscopie par expansion permet d'obtenir une dilatation d'un facteur 4 (donc d'améliorer la résolution du même facteur), mais des travaux sont en cours pour faire fonctionner la méthode avec un facteur de dilatation de 16. La figure 1.7 illustre une coupe de centriole obtenue par microscopie confocale, avant et après expansion.

Les microscopies à champ large, confocales et STED sont des modalités d'acquisition dites convolutionnelles, l'image acquise est modélisée comme la convolution de l'objet original par la réponse impulsionnelle du microscope (point spread function (PSF)). En effet, le système optique est considéré comme linéaire et invariant par translation, l'image de l'objet est donc la somme des images des points le constituant, ce qui revient à effectuer une convolution avec la PSF. Plus formellement, on a :

$$y = h \otimes f + \epsilon, \quad (1.3)$$

avec h la PSF du microscope, f l'objet original, y l'image acquise et ϵ le bruit. L'intensité détectée par le microscope dépend du nombre de photons qui atteignent le détecteur. Ce nombre est une variable aléatoire suivant une loi de Poisson dont la moyenne est proportionnelle à l'intensité du signal. Par ailleurs, le bruit provenant de l'appareil de mesure est supposé gaussien. Le bruit ϵ est donc en général modélisé comme combinaison linéaire d'un bruit gaussien et d'un bruit de Poisson [21].

En dehors de STED, il existe d'autres catégories de modalités d'acquisition, telles que la mi-

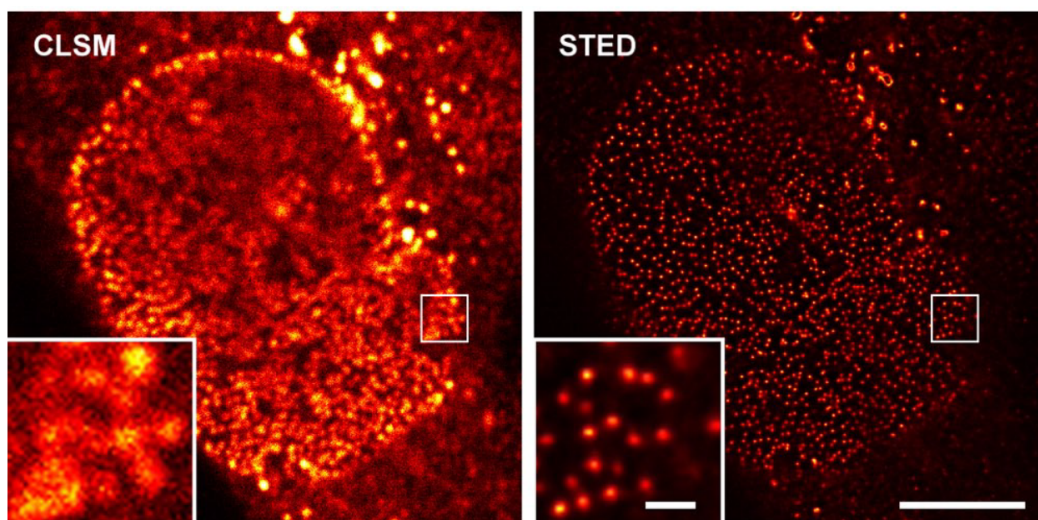


FIGURE 1.6 – Comparaison d'un échantillon observé par microscopie confocale à balayage laser (confocal laser scanning microscopy (CLSM)) et STED. Image extraite de l'article [17]

microscopie de localisation à molécule unique (single-molecule localization microscopy (SMLM)) [22] ou la microscopie à illumination structurée (structured illumination light sheet microscopy (SIM) [23]) qui permettent de dépasser la limite de diffraction de la lumière, et qui ne sont pas des modalités convolutionnelles. SMLM n'acquiert pas une image mais un nuage de points, dont les points sont les coordonnées 3D des fluorophores. Cette technique utilise des fluorophores particuliers qui alternent aléatoirement entre un état allumé et un état éteint. La probabilité de passage d'un état à l'autre peut-être modulée par une irradiation laser, ce qui permet d'imager successivement chacun des fluorophores, ainsi les images de fluorophores voisins ne se chevauchent pas. La méthode SIM module le signal que l'on souhaite imager à l'aide de différents filtres sinusoïdaux de phases différentes, appliqués avant le passage à travers la PSF du microscope. Cela donne lieu à un ensemble d'images dont les informations sont combinées afin de reconstruire un objet, dont la résolution surpasse la limite de diffraction.

Le développement de ces méthodes de super-résolution et de microscopie par expansion a rendu possible l'observation d'assemblages macromoléculaires en microscopie par fluorescence, et a motivé son utilisation croissante en biologie structurale, avec des résultats de plus en plus prometteurs [18, 24, 25].

1.4 Problématique

Les améliorations de résolution des microscopes optiques effectuées ces dernières décennies sont freinées par deux limitations majeures. Premièrement, la résolution est fortement anisotrope. Dans toutes les modalités convolutionnelles, la résolution selon l'axe du microscope est de 3 à 5 fois inférieure à la résolution dans le plan latéral. Deuxièmement, les variations d'intensité parasites le long des structures protéiques peuvent gêner l'interprétation biologique. Ce phénomène s'explique par un degré de labellisation (DL) faible et non uniforme. Ce dernier caractérise la densité des fluorophores qui se fixent sur la protéine d'intérêt et émettent la lumière fluorescente. Ces deux phénomènes sont illustrés sur la figure 1.8. Il s'agit d'une image de centriole acquise par microscopie confocale que l'on représente à l'aide de trois coupes orthogonales, dans les plans XY , XZ et YZ ,

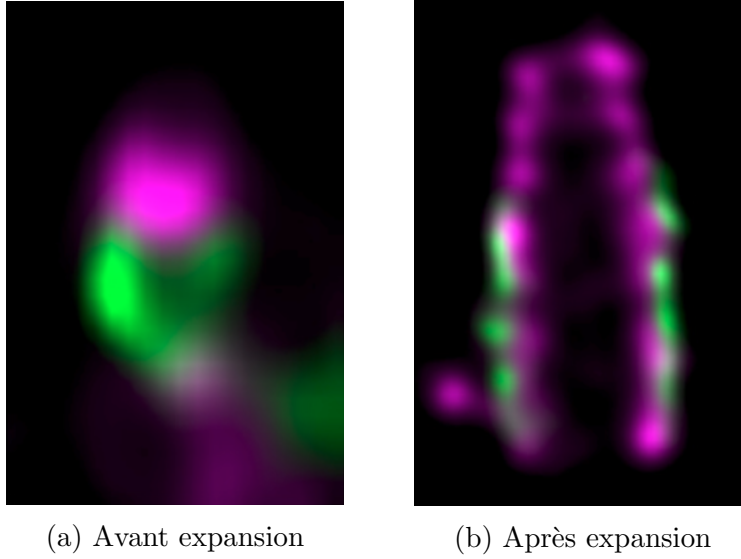


FIGURE 1.7 – Microscopie par expansion, illustrations extraites de [20]. Facteur de dilatation : 4

Z étant l'axe du microscope.

L'anisotropie de résolution et le faible DL sont des limitations physiques intrinsèques au processus d'acquisition. L'objectif de la thèse est de les contourner à l'aide de l'outil numérique en développant une méthode de reconstruction à partir de particules isolées (RPI). Le principe, similaire à la RPI en cryo-ME, consiste à combiner plusieurs acquisitions de particules biologiques identiques observées selon différentes orientations inconnues afin de reconstruire un modèle de particule. Grâce à la fusion des informations complémentaires dans chaque vue, nous visons à reconstruire un volume avec une résolution isotrope et une densité de labellisation uniforme. Nous nous concentrons dans cette thèse sur les modalités d'acquisition convolutionnelles (à champ large, confocal, et STED), nous n'aborderons pas les modalités SMLM et SIM.

Le processus de reconstruction se divise en 3 étapes distinctes. La première est la détection, à partir d'images contenant un grand nombre de particules identiques orientées aléatoirement, nous identifions les coordonnées tridimensionnelles des patchs contenant les vues. La détection n'est pas l'objet de cette thèse car une méthode a déjà été développée au sein de l'équipe IMAGeS [26]. Elle se base sur un apprentissage supervisé par fenêtre glissante : une taille de fenêtre est fixée et pour chaque position de l'image, un réseau de neurones convolutif tente de prédire si une particule est présente à cet emplacement. La contrainte majeure du problème de détection est de ne disposer que d'un très faible nombre d'observations, quelques dizaines qui proviennent d'une détection manuelle par l'utilisateur. Néanmoins, le nombre de particules par image est grossièrement connu, ce qui permet de développer une méthode de *Positive-Unlabelled learning*. L'idée générale est d'étiqueter une faible proportion de l'image comme positive, et de considérer le reste non pas comme des exemples négatifs mais comme des exemples non-labellisés pour lesquels on dispose d'*a priori* statistiques.

La seconde étape, qui est l'objet de cette thèse, est la reconstruction *ab-initio* : une estimation conjointe du volume à reconstruire et des poses (*i.e.* orientations et translations) permettant d'obtenir une estimation grossière de la forme de la particule. Elle ne suppose aucune connaissance *a priori* sur les poses et sur le volume à reconstruire : on part d'une initialisation aléatoire. Il s'agit donc d'un problème inverse aveugle : on cherche à reconstruire la structure qui a donné lieu à l'observation des vues, sans connaître *a priori* une partie des paramètres du modèle d'acquisition

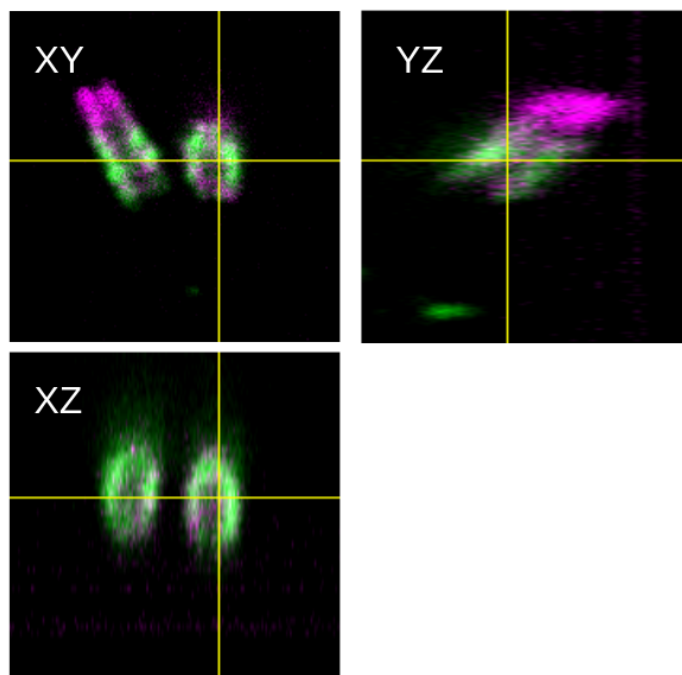


FIGURE 1.8 – Vue 3D bi-canal d’un centriole, représentée à l’aide de 3 vues coupes orthogonales. La résolution dans le plan XY est 4 fois meilleures que celles dans les plans YZ et XZ .

(en l’occurrence les poses).

Enfin une méthode de raffinement, initialisée à partir de la reconstruction *ab-initio*, permet d’obtenir une estimation plus fine du volume estimé et des poses. Une méthode a été développée par Denis Fortun, un de mes encadrants de thèse [27]. Elle se base sur une alternance entre la minimisation du volume à poses fixe, et la minimisation des poses à volume fixe. Les étapes du processus de reconstruction - détection, reconstruction *ab-initio* et raffinement - sont représentées sur la figure 1.9.

Ma thèse s’inscrit dans une thématique autour du traitement d’images pour la microscopie par fluorescence, financée en partie par le projet SPfluor. L’objectif est d’intégrer le processus de reconstruction complet au sein du logiciel Scipion, habituellement utilisé pour la RPI en cryo-ME. En même temps que ma thèse, le stage de M2 effectué par Luc Vedrenne [26] a été dédié à la méthode de détection des vues. Un poste d’ingénieur de recherche en CDD, encore en cours, et tenu par Jean Plumail est dédié à l’intégration des méthodes de détection, reconstruction *ab-initio* et raffinement au sein du logiciel Scipion. Enfin la reconstruction dédiée aux modalités basées sur l’acquisition de nuages de points (tel que SMLM) plutôt que d’image 3D fait l’objet d’une autre thèse, encore en cours, et effectuée par Luc Vedrenne.

1.4.1 Reconstruction homogène et hétérogène

La reconstruction à partir de vues à résolution anisotrope peut être qualifiée d’homogène ou d’hétérogène. La reconstruction homogène suppose que toutes les vues sont issues du même modèle 3D que l’on cherche à estimer. Cependant, cette hypothèse n’est pas toujours valide, car l’assemblage macromoléculaire étudié peut exister sous différents états stables. C’est ce que l’on appelle l’hétérogénéité des conformations. On ne cherche alors plus à reconstruire un volume unique, mais

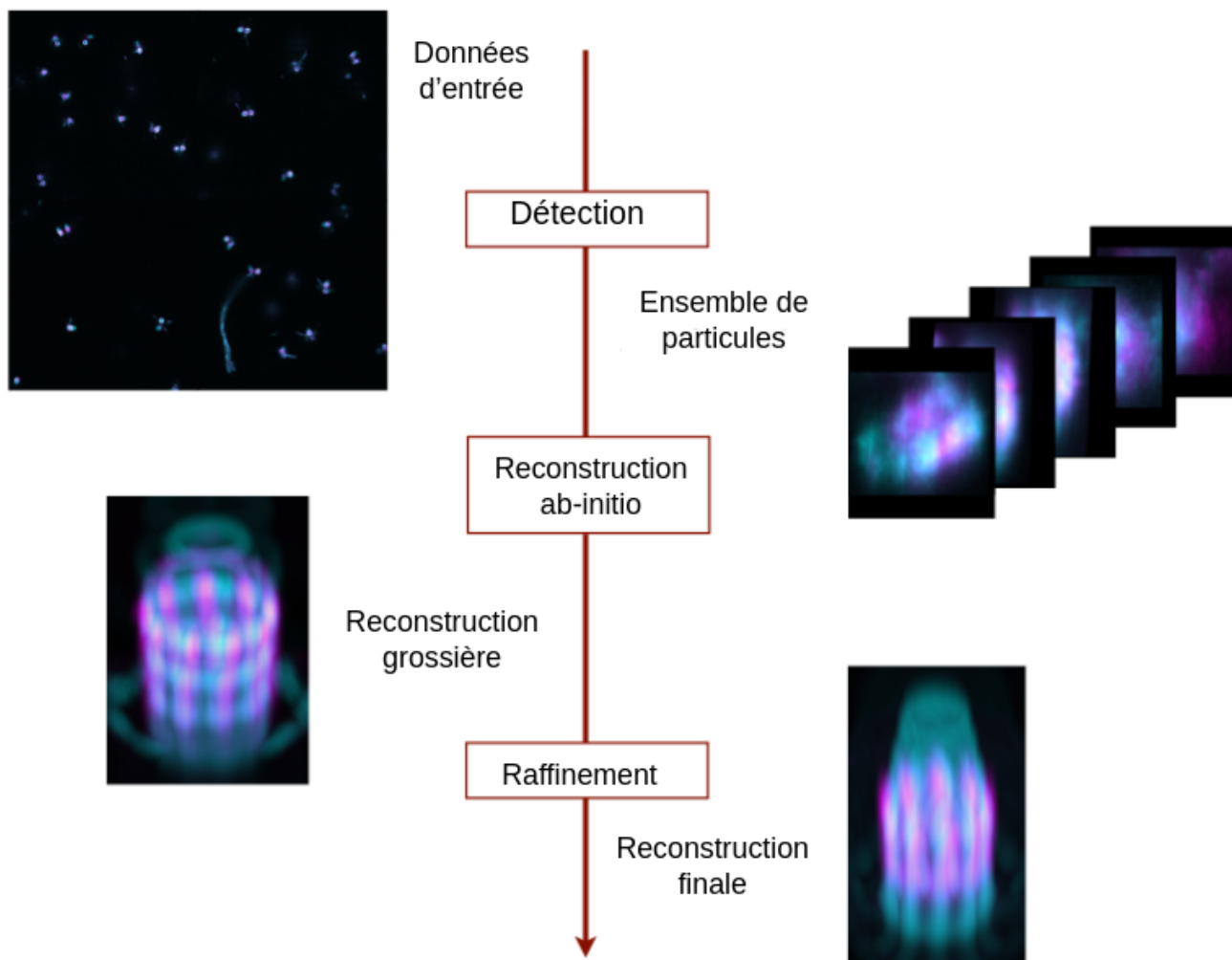


FIGURE 1.9 – Processus de reconstruction (exemple du centriole)

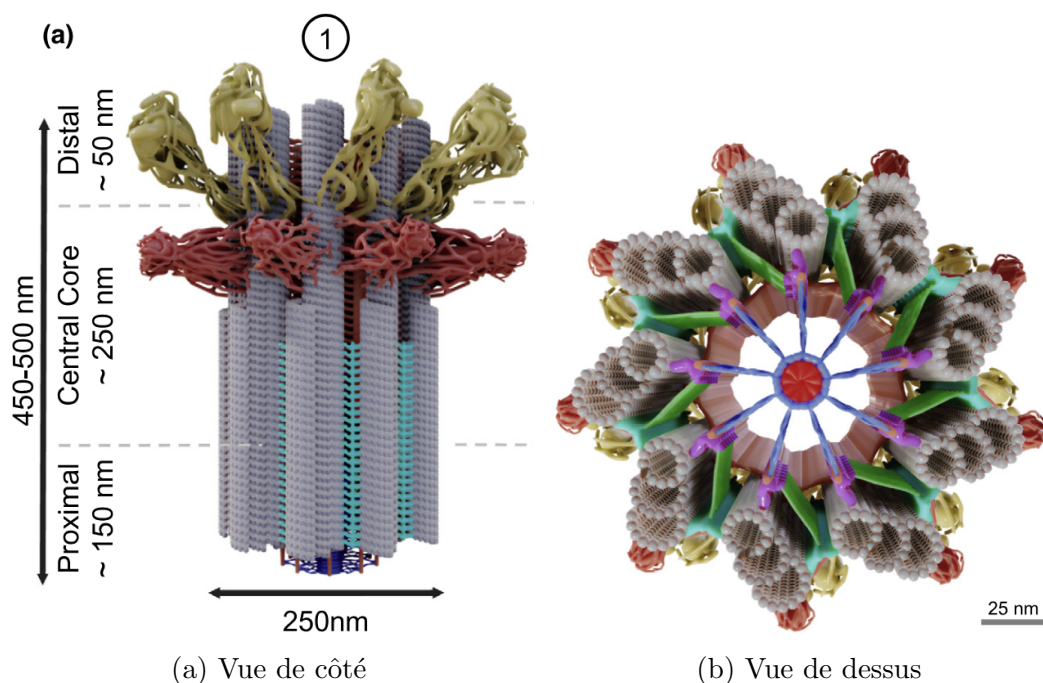


FIGURE 1.10 – Modèle de centriole [28]

un ensemble de volumes, ainsi que d'associer à chacune des vues le volume dont elle est issue. Durant ma thèse, j'ai dans un premier temps développé une méthode de reconstruction homogène (qui sera l'objet du chapitre 3), et dans un deuxième temps une méthode de reconstruction hétérogène (objet du chapitre 4). Cette dernière peut néanmoins également s'appliquer à la reconstruction homogène.

1.4.2 Cas d'application : le centriole

Nous avons appliqué nos méthodes de reconstruction, homogène et hétérogène, au cas particulier du centriole. Les données nous sont fournies par le Centriole Lab⁸, dirigé par Paul Guichard et Virginie Hamel. Cet assemblage macromoléculaire prend la forme d'un cylindre creux admettant une symétrie C9 (voir figure 1.10). La taille relativement grande de cet assemblage macromoléculaire (400 nm de longueur et 250 nm diamètre environ) se prête assez mal à son observation en cryo-ME. Il est en effet difficile de congeler de trop gros objets sans les dégrader. C'est pourquoi, les biologistes avec qui nous travaillons préfèrent utiliser la microscopie par fluorescence pour son observation.

Le centriole joue un rôle central dans le cycle cellulaire des cellules eucaryotes (*i.e.* les cellules qui admettent un noyau). Le cycle cellulaire constitue l'ensemble des étapes qu'une cellule traverse depuis sa création après la division d'une cellule mère jusqu'à ce qu'elle achève sa propre division pour donner naissance à deux cellules filles présentant des caractéristiques similaires à celles de la cellule mère. Le cycle cellulaire se divise en deux parties principales : l'interphase et la division cellulaire (ou mitose). Lors de l'interphase, la cellule synthétise les protéines nécessaires à son fonctionnement et prépare sa division notamment en répliquant son ADN. Vient ensuite la mitose, phase de division cellulaire durant laquelle le centriole est impliqué. La mitose est elle-même divisée

8. <https://mocel.unige.ch/research-groups/guichard-hamel/overview>

en 4 sous étapes : la prophase, la métaphase, l'anaphase et la télophase. Ces dernières sont illustrées sur la figure 1.11, extraites du site web de nagwa⁹. Nous pouvons voir qu'avant la division, chaque cellule contient 2 centrosomes. Un centrosome est une paire de centrioles orientés orthogonalement. Ces figures illustrent le rôle clé des 2 centrosomes de la cellule : ils permettent la séparation de chaque chromosome en 2 chromatides, qui sont réparties dans les 2 cellules filles. De même, les deux paires de centrioles sont répartis dans les deux cellules filles. Ces dernières ne contiennent donc que 2 centrioles, alors que la cellule mère en contient 4. Lors de l'interphase, les centrioles vont donc se dupliquer de sorte à ce que chacune des cellules filles puissent se diviser à leur tour. Dans un premier temps les deux centrioles d'une cellule fille, initialement regroupés dans un centrosome se séparent pour former 2 centrioles isolés. Un procentriole apparaît ensuite à proximité de chacun d'entre eux. Le procentriole a, tout comme le centriole, une structure de tore avec symétrie C9 mais sa longueur et son diamètre sont inférieurs. Le procentriole croît petit à petit lors de l'interphase pour arriver, à la fin de cette dernière, à état de maturité. La cellule fille admet alors 4 centrioles identiques (répartis en 2 centrosomes) et elle peut à son tour se diviser (elle devient donc cellule mère). Le procentriole est donc un assemblage pouvant exister sous différentes conformations, chacune d'entre elles étant associées à un état d'avancement différent de l'interphase : il y a hétérogénéité des conformations. La figure 1.12 illustre de manière schématique différents états de croissance du procentriole, situé à proximité d'un centriole mature.

Les modèles déjà existants de centriole, tels que la structure présentée dans l'article [28] (illustrée sur la figure 1.10), ont été déterminés à partir de deux sources principales d'observation. La première est la cryotomographie. Comme évoqué en section 1.2, cette technique se base sur l'observation d'un échantillon unique selon différentes orientations (ce n'est donc pas de la RPI, la RPI en cryo-ME n'est pas adaptée au centriole car il a une trop grande taille). L'inconvénient de la cryotomographie est que le nombre de vues que l'on peut utiliser pour la reconstruction est relativement limité. D'autre part, les données de cryotomographie ne permettent pas l'obtention d'une reconstruction hétérogène du centriole à différents états de croissance (car un seul objet est observé). La seconde source est la microscopie par fluorescence. Étant donné qu'avant ma thèse, il n'existait quasiment aucune méthode de RPI en microscopie par fluorescence (exceptée celle développée par Denis Fortun [27]), ces données étaient difficilement exploitables. Des modèles de centriole ont tout de même été générés à partir des observations de microscopie par fluorescence en mesurant manuellement les dimensions des vues [28, 29]. Cette méthodologie est d'une part très laborieuse mais est également fortement biaisée par les mesures manuelles. Les méthodes de RPI que j'ai développé pendant ma thèse permette quant à elles d'obtenir des reconstructions quasiment sans intervention manuelle.

Nous traiterons la reconstruction hétérogène du procentriole au chapitre 4. La reconstruction homogène du centriole mature sera quant à elle détaillée au chapitre 3. Mais avant de détailler les deux méthodes que j'ai implémenté, je propose une revue des méthodes de RPI existantes, en microscopie par fluorescence et en cryo-ME : c'est l'objet du chapitre suivant.

9. <https://www.nagwa.com/fr/explainers/650191397968/>

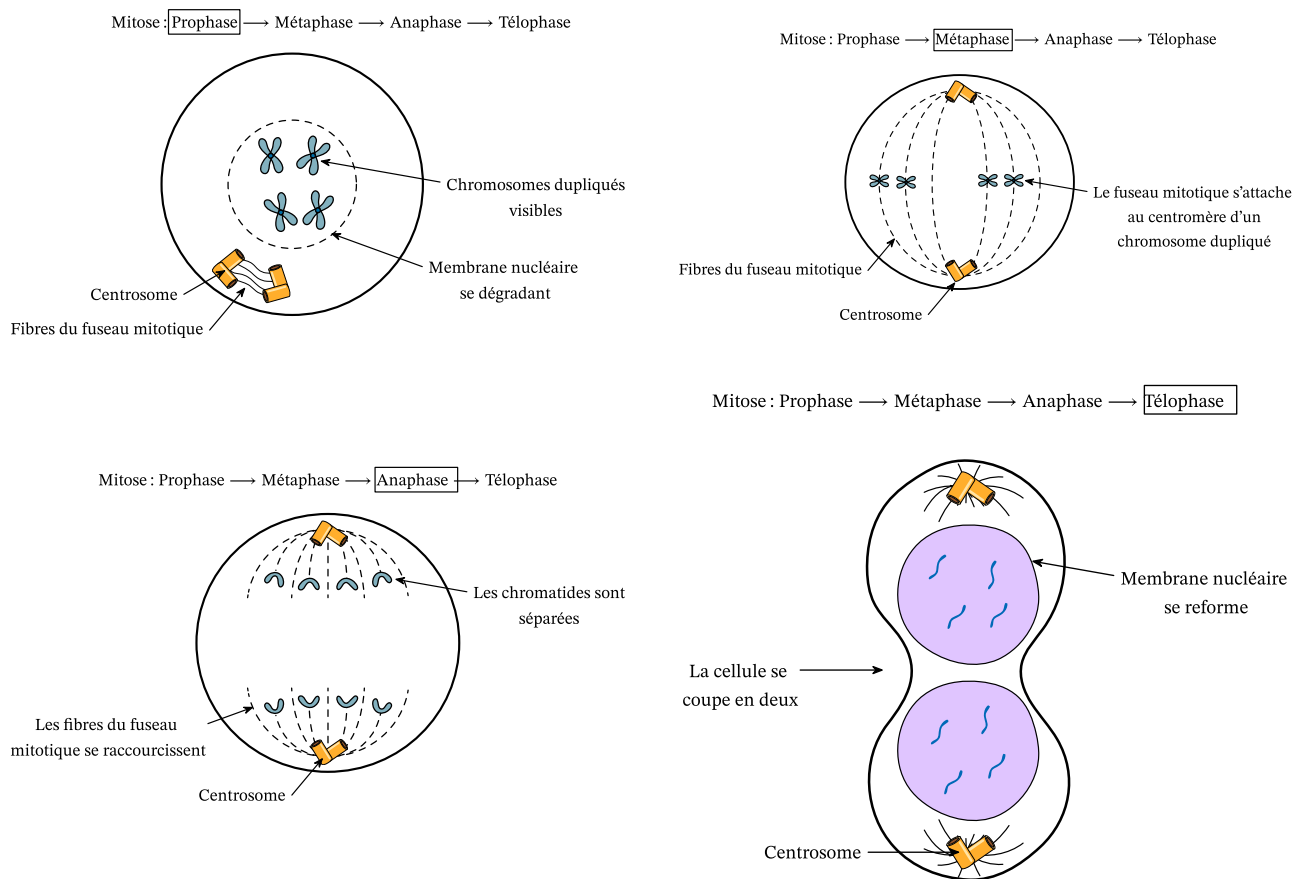


FIGURE 1.11 – Illustration des 4 étapes de la mitose, et du rôle du centrosome dans cette dernière. Un centrosome est constitué de 2 centrioles orientés orthogonalement. Ces illustrations sont issues du site de nagwa (citation dans le texte).

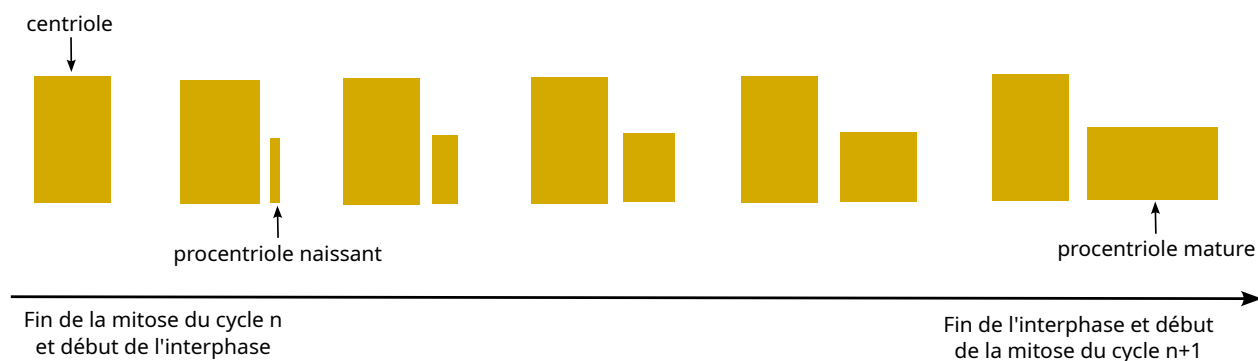


FIGURE 1.12 – Illustration schématique de différents états de croissance du procentriole lors de l'interphase. Le centriole, de forme cylindrique, est illustrée en vue de côté par un rectangle

Chapitre 2

État de l’art

La reconstruction multi-vues est une pratique courante dans plusieurs techniques de microscopie pour lesquels des procédures de calibration physique fournissent une estimation précise de la position et de l’orientation de chaque vue (par exemple, la microscopie à feuille de lumière [30, 31], la microscopie 3D à fluorescence par réflexion totale interne [32]). En revanche, la spécificité du problème de reconstruction multi-vues en RPI réside dans l’absence de connaissances préalables sur les poses (orientations et translations associées à chacune des vues). Cela donne lieu à un problème d’optimisation non-convexe particulièrement difficile qui nécessite le développement de méthodes dédiées.

Cette section a pour but d’introduire les travaux existants en RPI pour la microscopie par fluorescence. Le problème est similaire à la RPI en cryo-ME dans une certaine mesure, car dans les deux cas il s’agit de reconstruire un volume 3D à partir d’un ensemble de vues dont les poses sont inconnues. Ce qui distingue les deux modalités est que les vues acquises en microscopie par fluorescence sont modélisées comme la convolution de l’objet original avec la réponse impulsionnelle du microscope, alors que le modèle d’acquisition en cryo-ME est basé sur une projection 2D de l’objet 3D d’origine. Les deux problèmes étant très similaires, nous présenterons donc également les méthodes de RPI dédiées à la cryo-ME.

2.1 Reconstruction en microscopie par fluorescence

La microscopie par fluorescence a commencé à être utilisée comme outil pour la biologie structurale bien plus récemment que la microscopie électronique (et plus spécifiquement la cryo-ME). En effet, comme nous l’avons vu, la résolution des microscopes optiques traditionnels était restreinte par la limite de diffraction de la lumière, et les premières techniques de super-résolution en microscopie optique datent des années 90. Les microscopes électroniques sont en revanche utilisés en biologie structurale depuis les années 1930 [33]. Par conséquent, la littérature sur la RPI en microscopie par fluorescence est beaucoup moins vaste que celle de la RPI en cryo-ME.

Cela explique également le fait que, malgré la différence entre les modèles d’acquisition, les premières reconstructions en microscopie par fluorescence étaient obtenues en appliquant les méthodes dédiées à la cryo-ME à des images 2D, obtenues en sélectionnant un seul plan dans les acquisitions 3D, ou en projetant artificiellement les volumes 3D [34, 35]. Cela donne lieu à des reconstructions grossières très largement sous-optimales, car ces méthodes ignorent le modèle d’acquisition réel des microscopes optiques et, par ailleurs, une partie considérable de l’information contenue dans les images 3D est perdue lors de la projection. La plupart des méthodes existantes

développées spécifiquement pour la microscopie par fluorescence sont soit restreintes à des images d’entrées 2D [36, 37, 38], ou nécessitent une initialisation ou un modèle paramétrique [39, 40, 41], ce qui introduit inévitablement un biais dans la reconstruction.

La méthode proposée par Fortun *et al.* dans [27] est la seule, à notre connaissance, à prendre en compte un modèle d’imagerie 3D approprié. Elle est basée sur une procédure de recuit simulé qui explore l’espace des poses en minimisant une fonction de coût, définie comme l’écart au modèle d’observation auquel s’ajoute un terme de régularisation. Malgré de bonnes performances sur des données réelles, elle présente trois limitations pratiques majeures. La première est son coût de calcul élevé, qui peut s’avérer limitant, en particulier pour des particules non symétriques, pour lesquelles l’espace de recherche des orientations ne peut pas être réduit. En effet, le calcul de l’écart au modèle d’observation implique une étape de reconstruction du volume à poses constantes, qui doit être répétée à chaque itération du recuit simulé. La deuxième limitation est la nécessité d’une étape de classification pour réduire le nombre de particules. La classification automatique manque souvent de robustesse sur des données réelles, et la sélection manuelle des particules est fastidieuse et introduit un biais arbitraire. Finalement, la reconstruction repose sur une procédure d’optimisation séquentielle qui considère les vues une par une dans un ordre prédéfini, ce qui entraîne des erreurs de reconstruction pour des degrés de labellisation faibles.

Quelques études se sont intéressées à la RPI pour les données SMLM, en se basant sur des méthodes de recalage de nuages de points [40, 42]. Notre problème de reconstruction dédié aux modalités convolutionnelles est radicalement différent, car nos données sont des volumes ne pouvant pas être traités avec les mêmes méthodes que les nuages de points. De plus, notre objectif est la reconstruction et non seulement le recalage.

2.2 Reconstruction en cryo-ME

Si peu de méthodes de RPI en microscopie par fluorescence ont été développées, il existe en revanche une littérature conséquente sur la reconstruction en cryo-ME. Historiquement, les premières méthodes de RPI en cryo-ME se basaient sur l’identification de lignes communes entre paires de vues [43, 44, 45, 46]. En effet, du théorème de la coupe centrale de Fourier découle que deux projections de l’objet 3D sont, dans le domaine de Fourier, des plans orthogonaux à ces projections qui, si les directions sont différentes, s’intersectent en une ligne. Son identification donne l’axe d’inclinaison commun à ces deux vues mais n’est pas suffisant pour déterminer leurs orientations relatives, on a pour cela besoin d’une troisième vue, dont on détermine les axes d’inclinaison communs avec les deux autres vues. Ces trois axes sont suffisants pour déterminer les orientations relatives des 3 vues. L’identification d’une ligne commune entre deux images s’effectue en maximisant la corrélation croisée normalisée sur l’ensemble des paires de lignes des deux images. Les méthodes de lignes communes restent néanmoins peu robustes au bruit. Elles ne sont d’autre part pas applicables à la reconstruction en fluorescence, car le modèle d’observation en fluorescence se base sur une convolution et non une projection, ce qui empêche l’utilisation du théorème de la coupe centrale de Fourier¹.

Nous proposons dans cette section une revue des méthodes de RPI en cryo-ME reposant sur des principes pouvant être transposés au cas de la reconstruction en microscopie par fluorescence. Nous séparons l’état de l’art en quatre parties distinctes. Dans un premier temps, nous présentons les méthodes se basant sur une estimation conjointe des poses et du volume par des algorithmes

1. Le théorème de la coupe centrale de Fourier énonce que la projection d’une image selon une direction donnée est équivalente dans le domaine de Fourier à une coupe 2D orthogonale à cette même direction

itératifs stochastiques. Dans un second temps, nous exposons les méthodes qui ne cherchent pas à estimer les poses mais les considèrent comme variables latentes par rapport auxquelles la fonction à minimiser est marginalisée. Puis, nous présentons les algorithmes les plus récents utilisant des réseaux de neurones pour la représentation du volume, pour l'estimation des poses, ou pour les deux à la fois. Enfin, nous détaillons les méthodes dédiées au cas de la reconstruction hétérogène, qui supposent non plus l'existence d'un volume unique que l'on cherche à reconstruire, mais d'un ensemble de volumes, pouvant être discret ou continu.

La plupart des méthodes présentées dans cette section sont dédiées à la reconstruction *ab-initio*. Nous décrivons néanmoins certaines méthodes incontournables dédiées au raffinement (*e.g.* RELION [47]), et le précisons dans le cas échéant.

Dans certains cas, la reconstruction *ab-initio* n'est non pas obtenue directement à partir des vues du jeu de données mais à partir de classes moyennes. Ces dernières sont calculées en regroupant les vues similaires entre elles, puis en moyennant les vues d'un même groupe [48]. Cela présente l'avantage de réduire l'influence du bruit sur la qualité de la reconstruction, le rapport signal à bruit (RSB) des images de cryo-ME étant particulièrement élevé. Cependant, cela introduit une source d'erreur supplémentaire au processus de reconstruction. C'est pourquoi tous les papiers que nous présentons dans cette section n'utilisent pas des classes moyennes (appelées de manière équivalente « moyennes » pour plus de légèreté) comme entrée de leurs algorithmes. Nous le précisons lorsqu'une méthode se base sur l'utilisation de classes moyennes.

2.2.1 Estimation conjointe des poses et du volume

Une première catégorie de méthodes cherche à estimer conjointement les orientations $(\theta_l)_{l \in \llbracket 1, N \rrbracket}$, les translations $(t_l)_{l \in \llbracket 1, N \rrbracket}$ et le volume f de sorte à maximiser la vraisemblance des données $(y_l)_{l \in \llbracket 1, N \rrbracket}$. Plus formellement, il s'agit de résoudre le problème de maximisation suivant :

$$f^*, \Theta^*, T^* = \arg \max_{f, \Theta, T} p(Y|f, \Theta, T), \quad (2.1)$$

avec $Y = (y_l)_{l \in \llbracket 1, N \rrbracket}$, $\Theta = (\theta_l)_{l \in \llbracket 1, N \rrbracket}$ et $T = (t_l)_{l \in \llbracket 1, N \rrbracket}$ ². On suppose que les vues Y sont indépendamment distribuées. Le problème (2.1) est donc équivalent à :

$$f^*, \Theta^*, T^* = \arg \max_{f, \Theta, T} \prod_{l=1}^N p(y_l|f, \theta_l, t_l). \quad (2.2)$$

De plus, il est courant de faire l'hypothèse que le bruit ϵ dans (2.3) est un bruit blanc gaussien. Sous cette hypothèse on a :

$$p(y_l|f, \theta_l, t_l) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{\|y_l - CTF \otimes \mathcal{T}_{t_l}(\mathcal{P}(\mathcal{R}_{\theta_l}(f)))\|^2}{2\sigma^2}\right), \quad (2.3)$$

donc

$$-\log(p(y_l|f, \theta_l, t_l)) = \frac{1}{2\sigma^2} \|y_l - CTF \otimes \mathcal{T}_{t_l}(\mathcal{P}(\mathcal{R}_{\theta_l}(f)))\|^2 + \log(\sigma\sqrt{2\pi}). \quad (2.4)$$

Par application de la fonction $-\log$ à (2.2), le problème de maximum de vraisemblance est donc équivalent au problème de minimisation suivant :

2. Nous utiliserons des notations similaires pour la RPI en fluorescence, que nous expliciterons davantage dans la partie 3.2

$$f^*, \Theta^*, T^* = \arg \min_{f, \Theta, T} E(f, \Theta, T), \quad (2.5)$$

avec

$$E(f, \Theta, T) = \sum_{l=1}^N \|CTF * \mathcal{T}_{t_l}(\mathcal{P}(\mathcal{R}_{\theta_l}(f))) - y_l\|^2. \quad (2.6)$$

En pratique l'équation (2.5) est résolue dans le domaine de Fourier, car cela permet d'effectuer les projections très rapidement en tirant partie du théorème de la coupe centrale de Fourier³. L'équation (2.5) donne un problème non-convexe particulièrement difficile pour lequel il n'existe pas de solution analytique. Par conséquent, sa résolution nécessite de développer des méthodes stochastiques permettant d'explorer au mieux l'espace de recherche, sans que cela garantisse nécessairement l'obtention d'une solution optimale. C'est ce que proposent les auteurs des articles [49, 50, 51, 52, 53].

Dans [49], Ogura et Sato reconstruisent à partir de classes moyennes, et se basent sur l'utilisation d'un schéma itératif de recuit simulé. Des poses aléatoires sont assignées initialement à chacune des moyennes et le volume est reconstruit en utilisant une méthode de rétroprojection [54]. À chaque itération, les poses associées à une moyenne tirée au sort sont mises à jour et le volume est recalculé. La mise à jour est acceptée si elle réduit l'énergie, sinon, elle peut encore être acceptée avec une probabilité dépendant de la différence d'énergie et d'un paramètre de température décroissant. Ce processus permet d'éviter les minima locaux en début d'apprentissage et de se concentrer sur la solution courante à mesure que l'algorithme converge. Leur algorithme donne des résultats corrects sur 2 jeux de données simulées, et un jeu de données réelles, et ils mettent en évidence la nécessité de définir un schéma de décroissance du paramètre de température suffisamment lent, pour permettre à l'algorithme d'éviter les minima locaux.

Les auteurs de plusieurs études ([50, 51, 52] et [53]) proposent des variantes de l'algorithme de « correspondance de projections ». Cet algorithme d'optimisation alterne entre l'estimation des poses à volume fixe et l'estimation du volume à poses fixes. Les poses sont déterminées par recherche exhaustive, en projetant le volume selon de nombreuses poses sur une grille discrète et en comparant ces projections aux vues du jeu de données via corrélation ou différence quadratique moyenne. Lors de la reconstruction, chaque vue est associée à la pose avec le meilleur score de reprojection, et le volume est reconstruit en utilisant ces poses. Cette méthode est habituellement utilisée pour le raffinement d'une reconstruction grossière, mais les auteurs de [50, 51, 52] et [53] lui apportent un certain nombre d'améliorations et/ou l'appliquent à des cas particuliers afin de la faire fonctionner pour la reconstruction *ab-initio*.

Dans [50], Yan *et al.* proposent une modification de l'algorithme de correspondance de projection en utilisant uniquement les vues avec un score de reprojection supérieur à un seuil. Leur méthode fonctionne pour des objets avec symétrie icosaédrique, commune chez les virus, et cette symétrie est imposée comme connaissance a priori. Cependant, l'algorithme doit être relancé plusieurs fois pour obtenir des résultats corrects, sélectionnés ensuite par une analyse visuelle et le score de fourier shell correlation (FSC) (que l'on explicitera en section 3.6.1.2.2).

La méthode VIPER [52], proposée par Penczek *et al.*, est une variante de l'algorithme de correspondance de projections pour la reconstruction *ab-initio* à partir de classes moyennes. Elle inclut deux modifications pour améliorer la robustesse : d'abord, l'orientation retenue pour une classe

3. Dans le cadre de la RPI en fluorescence, ce théorème ne peut pas être utilisé, mais le domaine de Fourier est tout de même utilisé afin d'effectuer la convolution plus rapidement (par une multiplication terme à terme)

moyenne est la première à améliorer le score de reprojection par rapport à l’itération précédente, réduisant ainsi le temps de calcul et évitant certains minima locaux (il s’agit d’un algorithme d’escalade de la colline stochastique (ECS)). Ensuite, un algorithme génétique est implémenté, avec une population de six individus initialisés aléatoirement. A chaque individu est appliqué une reconstruction via l’ECS, et les six individus sont appariés pour générer des descendants qui sont utilisés pour les générations suivantes. Bien que l’algorithme VIPER produise de meilleurs résultats que ceux de Yan *et al.* et donne des reconstructions cohérentes sur des jeux de données réelles, il ne garantit pas l’obtention de l’optimum global. Pour augmenter les chances de converger vers ce dernier, les auteurs lancent l’algorithme VIPER un grand nombre de fois et éliminent les vues dont les poses estimées varient de manière trop importante d’une expérience à l’autre, ou qui obtiennent en moyenne de mauvais scores de reprojections. Si cette procédure améliore la qualité des résultats, elle est néanmoins très coûteuse en temps de calcul.

L’algorithme Cryo-RANSAC [53], proposé par Vargas *et. al.*, reconstruit à partir de classes moyennes et est basé sur une approche de réduction de dimensionnalité non linéaire. Chaque moyenne est projetée sur un espace bi-dimensionnel par la méthode LTSA (Local Tangent Space Alignment). Cet espace est séparé en un nombre prédéfini M de carreaux (typiquement 9), et une moyenne est sélectionnée aléatoirement dans chaque carreau, des orientations aléatoires leurs sont assignées, et le volume est reconstruit à partir de ces M moyennes. Le processus est répété indépendamment L fois générant ainsi L reconstructions. Pour chacune d’entre elles, le nombre de *inliers* est déterminé. Un *inlier* est une classe moyenne dont le score de reprojection est supérieur à un certain seuil. Les k reconstructions admettant le plus grand nombre de *inliers* sont conservées et raffinées par la méthode de correspondance de projections. Après raffinement, ces k volumes sont classés en sommant les coefficients de corrélation de leurs *inliers* et le volume qui obtient le score le plus élevé est conservé. L’algorithme Cryo-RANSAC parvient à donner des résultats satisfaisants sur un jeu de données simulé et 4 jeux de données réelles. Leur méthode présente deux avantages par rapport à l’état de l’art au moment de la publication (2014). D’une part, les résultats dépendent très peu des paramètres initiaux de l’algorithme. D’autre part, la reconstruction est très rapide : moins d’une heure sur la plupart des exemples qu’ils présentent (sans prendre en compte le calcul des classes moyennes). Néanmoins, tout comme les méthodes déjà présentées dans cette section, Cryo-RANSAC ne parvient pas dans tous les cas de figure à produire des résultats satisfaisants, en raison de la présence de nombreux minima locaux.

Les méthodes présentées dans cette section proposent des algorithmes stochastiques itératifs, et heuristiques, permettant d’estimer conjointement le volume et les poses associées à chacune des vues ou classes moyennes. Nous avons exposé un algorithme basé sur un recuit simulé [49], et 3 variantes de la méthode de correspondance de projections [50, 52, 53] pouvant être appliquées à la reconstruction *ab-initio*. Néanmoins, ces méthodes relativement anciennes (Cryo-RANSAC [53] et VIPER [52] datent de 2014 et Yan *et al.* de 2007) sont loin d’être à la pointe de l’art des méthodes de RPI en cryo-ME (RELION, CryoSPARC, CryoAI, CryoDRGN, que nous aborderons dans les sections 2.2.2 et 2.2.3). En particulier, l’algorithme de correspondance de projections, basé sur une optimisation alternée, est très peu robuste, et même si les auteurs mentionnés dans cette section ont tenté d’y remédier à l’aide de stratégies heuristiques, cela reste très loin d’être suffisant.

2.2.2 Marginalisation par rapport aux poses

En raison du faible rapport signal à bruit des images de cryo-ME, il est risqué de tenter de trouver une seule pose optimale pour chacune des vues. En effet, une erreur d’estimation des poses pour une vue peut avoir une influence considérable sur le résultat final. C’est pourquoi, un

certain nombre d'articles de la littérature ne cherchent plus à estimer directement les poses mais les considèrent comme variables latentes par rapport auxquelles une marginalisation est effectuée. Intuitivement, cela revient à effectuer un moyennage sur les poses pondérées par des poids qui sont d'autant plus élevés que la probabilité que la pose soit correcte est élevée. On notera dans la suite $v = (\theta, t) \in \Lambda$ le paramètre de poses, avec $\Lambda = SO(3) \times \mathbb{R}^2$. On notera $\Upsilon \in \Lambda^N$ le paramètre de poses associé à l'ensemble des vues.

2.2.2.1 Optimisation par Espérance Maximisation

2.2.2.1.1 Principe de la méthode L'algorithme d'espérance-maximisation (EM) est une approche itérative visant à estimer les paramètres du maximum de vraisemblance d'un modèle probabiliste lorsque ce dernier dépend de variables latentes non observables. La méthode EM a été appliquée pour la première fois à la RPI en cryo-ME par Sjors Scheres [55, 56, 47], ce qui a donné lieu en 2012 à la méthode RELION [47]. Dans ce paragraphe nous présentons le principe général de l'algorithme EM, puis explicitons formellement comment il se décline au cas particulier de la RPI en cryo-ME.

L'algorithme EM alterne entre deux phases, une phase d'espérance et une phase de maximisation. Lors de la phase d'espérance (E), on calcule la probabilité *a-posteriori* $p(\Upsilon|Y, f^{(n)})$ des variables latentes connaissant le modèle et les données. Elle s'obtient à partir de la vraisemblance des données, par utilisation de la formule de bayes :

$$p(\Upsilon|Y, f^{(n)}) = \frac{p(Y|\Upsilon, f^{(n)})p(\Upsilon|f^{(n)})}{\int_{\Upsilon} p(Y|\Upsilon, f^{(n)})p(\Upsilon|f^{(n)})d\Upsilon}. \quad (2.7)$$

Puis, lors de la phase de maximisation (M), on calcule la nouvelle estimation des paramètres du modèle en maximisant l'espérance mathématique du logarithme de la vraisemblance par rapport aux variables latentes Υ , connaissant les données Y et l'estimation courante du modèle $f^{(n)}$. Formellement, la mise à jour est :

$$f^{(n+1)} = \arg \max_f Q(f, f^{(n)}), \quad (2.8)$$

avec :

$$\begin{aligned} Q(f, f^{(n)}) &= E_{\Upsilon|Y, f^{(n)}}(\log p(Y, \Upsilon|f)) \\ &= \int_{\Upsilon} p(\Upsilon|Y, f^{(n)}) \log p(Y, \Upsilon|f) d\Upsilon \\ &= \int_{\Upsilon} p(\Upsilon|Y, f^{(n)}) \log(p(Y|\Upsilon, f)p(\Upsilon|f)) d\Upsilon. \end{aligned} \quad (2.9)$$

Les nouveaux paramètres du modèle sont ensuite utilisés lors de la phase d'espérance suivante, et le processus est répété jusqu'à convergence.

Si l'on décline l'équation (2.9) au cas particulier du modèle d'observation en cryo-ME, on obtient :

$$Q(f, f^{(n)}) = \sum_{l=1}^N \int_v p(v|y_l, f^{(n)}) \log(p(y_l|v, f)p(v|f)) dv. \quad (2.10)$$

En remplaçant la vraisemblance $p(y_l|v, f)$ par sa valeur dans (2.3), on obtient :

$$Q(f, f^{(n)}) = - \sum_{l=1}^N \int_v p(v|y_l, f^{(n)}) \left(\frac{1}{2\sigma^2} \|y_l - \mathcal{H}_v(f)\|^2 + C \right) dv, \quad (2.11)$$

avec $\mathcal{H}_v(f) = CTF \otimes \mathcal{T}_t(\mathcal{P}(\mathcal{R}_\theta((f))))$ et $C = \log(\sigma\sqrt{2\pi}) - \log(p(v|f))$. C ne dépend pas de f , car on ne suppose aucun *a-priori* sur les poses, $p(v|f)$ est donc une distribution uniforme. La nouvelle estimation du volume, $f^{(n+1)}$, s'obtient en résolvant un problème des moindres carrés pondéré par les poids $p(v|y_l, f^{(n)})$.

L'algorithme d'espérance maximisation appliquée à la RPI en cryo-ME présente des similitudes avec l'algorithme de correspondance de projections. Dans les deux cas, on alterne entre une phase durant laquelle le volume courant f est projeté selon un grand nombre de directions et comparé avec l'ensemble des vues du jeu de données (*i.e.* le calcul de la vraisemblance $p(y_l|v, f^{(n)})$ pour tout v et pour tout $l \in \llbracket 1, N \rrbracket$), et une phase de reconstruction. Mais l'algorithme de correspondance de projections se contente de reconstruire en associant à chaque vue la pose maximisant cette vraisemblance, alors que l'algorithme EM reconstruit en utilisant toutes les poses et en leur associant des poids d'autant plus élevés que leur probabilité *a-posteriori* est élevée. Cette méthode s'avère par conséquent plus robuste que l'algorithme de correspondance de projections, car si on associe une seule pose à chaque vue, on prend le risque que cette dernière soit non seulement mauvaise mais surtout éloignée de la pose réelle, ce qui peut être le cas en particulier en début d'apprentissage. Malgré cette plus forte robustesse, l'algorithme EM est très sensible à l'initialisation $f^{(0)}$ et ne peut pas être utilisé pour la reconstruction *ab-initio*, mais uniquement pour le raffinement de structures grossières.

2.2.2.1.2 Maximum *a-posteriori* L'algorithme EM présenté dans le paragraphe précédent tente d'estimer le modèle maximisant la vraisemblance des données. Il est possible de transformer ce problème de maximum de vraisemblance en un problème de maximum *a posteriori*, permettant de modéliser une connaissance *a priori* $p(f)$ sur la structure du modèle à reconstruire f [47, 57]. On utilise pour cela la formule de Bayes, ce qui donne :

$$p(f|\Upsilon, Y) = \frac{p(Y|\Upsilon, f)p(f)}{p(Y)} \propto p(Y|\Upsilon, f)p(f). \quad (2.12)$$

La relation de proportionnalité est vraie puisque $p(Y)$ ne dépend pas de f , et c'est par rapport à cette variable qu'on optimise.

2.2.2.1.3 Réduction du temps de calcul La méthode EM (appliquée à le RPI en cryo-ME) peut s'avérer très coûteuse en temps de calcul. En effet, à chaque itération une intégrale sur un grand nombre de poses doit être calculée. C'est pourquoi plusieurs stratégies ont été intégrée à la méthode RELION pour réduire le nombre de poses sur lesquelles l'intégrale est effectuée.

La première consiste à utiliser deux grilles différentes de discrétisation des poses. Les poids $p(v|y_l, f^{(n)})$ sont d'abord évalués en chacun des points d'une grille grossière de valeurs de v . Ils sont ensuite triés et le plus petit sous-ensemble de cette grille dont la somme des poids associés est égale à 99,9 % est sélectionné. Ensuite, les poids sont évalués sur une grille plus fine dont les valeurs sont centrées autour des points de ce sous-ensemble. Leur seconde stratégie est d'intégrer seulement sur un voisinage des orientations optimales trouvées à l'itération précédente. Cela permet de réduire considérablement le temps de calcul, mais peut également avoir un effet important sur la qualité de la reconstruction, si les poses estimées à l'itération précédente sont éloignées dans l'espace de recherche des poses optimales. C'est pourquoi il est préférable d'utiliser cette stratégie lorsque le processus de reconstruction est déjà bien avancé.

2.2.2.2 Variantes de l’algorithme EM

Les auteurs de [58] et [59] proposent des variantes de l’algorithme EM, pouvant être appliquées à la reconstruction *ab-initio* (contrairement à l’algorithme EM classique, utilisé uniquement pour le raffinement d’une structure grossière).

Le méthode [58], proposée par Sorzano *et. al*, est très similaire dans le concept à l’algorithme EM : elle alterne entre une phase de projections du volume courant $f^{(n)}$ selon un grand nombre de poses et une phase de reconstruction résolvant un problème des moindres carrés pondéré par des poids calculés lors de la phase de projections. Néanmoins, le formalisme probabiliste est différent de l’algorithme EM. Par conséquent, contrairement à l’algorithme EM, les poids ne sont pas définis comme la probabilité *a-posteriori* des poses connaissant le volume courant et les données (c.a.d. $p(v|y_l, f^{(n)})$). Afin de définir leur schéma de poids, ils définissent les concepts de ”direction étant significative pour une image” et une ”image étant significative pour une direction”. Intuitivement, une direction θ est significative pour une image I si la distance entre la projection du volume courant selon θ et l’image I est suffisamment inférieure à la moyenne des distances entre l’image I et les projections selon toutes les autres directions. De même, une image I est significative pour une direction θ si la distance entre l’image I et la projection du volume courant selon θ est suffisamment inférieure à la moyenne des distances entre la projection du volume courant selon θ et toutes les autres images. Ensuite, le schéma de poids est défini de sorte à ce que le poids associé à la vue d’indice l et à la direction d’indice k soit nul si (et seulement si) la direction k n’est pas significative pour l’image l ou l’image l n’est pas significative pour la direction k . Un des avantages est de pouvoir éliminer automatiquement certaines images *outliers*, qui ne seront significatives par rapport à aucune direction. Ils montrent expérimentalement sur deux jeux de données réelles qu’à partir d’une initialisation aléatoire leur méthode converge dans tous les cas (10 fois sur 10) en un temps raisonnable (6h et 16h sur les deux jeux de données), alors que Cryo-RANSAC et RELION ne convergent qu’occasionnellement sur ces mêmes données (RELION est un algorithme dédié au raffinement de structure grossière, il est donc prévisible qu’il ne donne pas de bons résultats pour la reconstruction *ab-initio*).

La méthode PRIME [59], développée par Dominika et Hans Elmlund, est également une variante de l’algorithme EM. Une différence importante est qu’à chaque itération une seule vue sélectionnée au hasard est considérée, alors que l’algorithme EM considère l’ensemble des vues à chaque itération. Deuxièmement, tout comme [58] leur définition des poids est différente de celle de l’algorithme EM usuel. Ces deux modifications sacrifient la garantie d’améliorer la fonction objectif à chaque itération, mais permettent d’augmenter la diversification de l’exploration de l’espace de recherche. Ainsi, certains minima locaux peuvent être évités ce qui permet d’obtenir une reconstruction correcte sans initialisation, contrairement à l’algorithme EM classique. Leur algorithme donne des résultats satisfaisants sur trois jeux de données réelles et un jeu de données simulées, à partir d’une initialisation aléatoire.

2.2.2.3 Montée de gradient stochastique sur la vraisemblance des données

L’algorithme EM présenté en section 2.2.2.1 permet de résoudre le problème de RPI sans avoir à estimer les poses : il tente pour cela de maximiser une espérance par rapport à ces dernières. Un autre formalisme possible permettant également d’éviter leur estimation est de tenter de maximiser, non pas une espérance, mais la vraisemblance des données connaissant le volume. La vraisem-

blance d'une vue est calculée à l'aide de la formule des probabilités totales :

$$p(y_l|f) = \int_v p(y_l|f, v)p(v)dv. \quad (2.13)$$

La vraisemblance de l'ensemble du jeu de données s'obtient par produit des vraisemblances associées à chacune des vues :

$$p(Y|f) = \prod_{l=1}^N p(y_l|f). \quad (2.14)$$

En appliquant la fonction $\log$ à (2.14), on obtient la fonction de coût à maximiser :

$$f^* = \arg \max_f E(f) = \arg \max_f \sum_{l=1}^N E_l(f), \quad (2.15)$$

avec

$$E_l(f) = \log \int_v \exp(-\frac{\|y_l - \mathcal{H}_v(f)\|^2}{2\sigma^2})p(v)dv. \quad (2.16)$$

Le problème de maximum de vraisemblance peut être transformé en un problème de maximum *a-posteriori* permettant d'intégrer une connaissance *a-priori* $p(f)$ sur la structure du volume. Le problème d'optimisation n'est donc plus (2.15) mais :

$$f^* = \arg \max_f \sum_{l=1}^N E_l(f) + \log p(f). \quad (2.17)$$

Les problème (2.15), dans le cas du maximum de vraisemblance, ou (2.17) dans le cas du maximum *a-posteriori* peuvent ensuite être résolus à l'aide d'un algorithme de montée de gradient stochastique. Le volume est initialisé aléatoirement, et à chaque itération t , un batch $\mathcal{I}^{(t)}$ contenant un petit sous ensemble de vues est tiré au sort. Le gradient de la fonction de coût par rapport au volume est alors estimé en sommant uniquement sur ce sous-ensemble, puis le volume est modifié par un petit déplacement dans la direction du gradient estimé. Formellement, l'équation de mise à jour est :

$$f^{(t+1)} = f^{(t)} + \lambda \sum_{l \in \mathcal{I}^{(t)}} \nabla_f E_l(f^{(t)}). \quad (2.18)$$

L'avantage de la méthode de montée de gradient stochastique est qu'elle assure que le volume en cours de reconstruction ne soit que très légèrement modifié entre deux itérations consécutives, ce qui permet d'éviter les mauvais minima locaux au début de l'apprentissage

Les auteurs de [60] utilisent ce formalisme, avec une régularisation prenant la forme de la somme d'une régularisation de type variation totale (qui pénalise les variations d'intensité entre pixels voisins) et d'une régularisation de type $\mathcal{L}_1$. Néanmoins, l'utilisation de l'algorithme de montée de gradient stochastique sur la vraisemblance des données pour la reconstruction *ab-initio* est restée relativement marginale, jusqu'à ce qu'en 2016 Punjani *et al.* publient CryoSPARC [57]. Ce qui a freiné son développement est que le calcul du gradient nécessite l'estimation d'une intégrale sur l'ensemble des poses, ce qui est très coûteux en temps de calcul. L'algorithme EM et ses variantes nécessitent également l'estimation d'une intégrale sur l'ensemble des poses, mais ces derniers n'ont pas besoin d'autant d'itérations pour converger que la montée de gradient stochastique, ce qui

les rend moins coûteux en temps. Mais CryoSPARC [57] parvient à considérablement réduire le temps dédié à l'estimation de l'intégrale, à l'aide d'une méthode d'échantillonnage préférentiel. Au lieu de l'estimer sur une grille fine et régulière, elle est estimée sur un petit sous-ensemble tiré aléatoirement selon des distributions d'importance. Ces distributions sont la somme de deux composantes : une composante uniforme qui permet d'assurer la diversité de l'exploration et une composante centrée autour des poses ayant donné de bon résultats à l'itération précédente. Grâce à cette amélioration basée sur échantillonnage préférentiel, ils parviennent à obtenir une reconstruction en environ 24h, une durée très courte par rapport aux temps de calcul des autres méthodes de l'époque (qui étaient de l'ordre de une semaine).

2.2.2.4 Comparaisons des méthodes

Nous avons vu, dans les sections 2.2.1 et 2.2.2, qu'il existe un nombre considérable de méthodes dédiées à la reconstruction *ab-initio*, se basant sur une estimation conjointe des poses et du volume (section 2.2.1), ou sur une marginalisation par rapport aux poses (section 2.2.2). Il est donc naturel de se demander quel algorithme est le plus performant, afin de pouvoir identifier lequel est le plus adapté pour être transféré au cas de la microscopie par fluorescence.

Malheureusement, les méthodes précédemment exposées ne se comparent que très peu entre elles. Néanmoins, quelques indices peuvent nous éclairer. Sorzano *et. al* [58] se comparent à Cryo-RANSAC [53] et montrent que leur méthode est plus robuste sur des jeux de données compliqués, bien que plus coûteuse en temps de calcul. Cela suggère, comme on pouvait s'y attendre, que les méthodes qui reposent sur une marginalisation plutôt qu'une estimation d'une seule pose par vue sont plus robustes. Ceci est confirmé par les statistiques d'utilisation des différents logiciels de RPI en cryo-ME présentées sur le site web Electron Microscopy Data Bank⁴, qui sont clairement dominés par les logiciels CryoSPARC (de l'article de Punjani *et. al* [57]) et RELION (de l'article de Scheres [56]), tous les deux basés sur une marginalisation par rapport aux poses. De plus, la méthode RELION est utilisée uniquement pour le raffinement, pour la reconstruction *ab-initio* le logiciel RELION utilise une méthode de montée de gradient stochastique inspirée de celle de CryoSPARC (ceci est précisé dans la documentation de RELION⁵). La méthode CryoSPARC, ou plus généralement une méthode à base de gradient stochastique, semble donc être le choix le plus adapté pour la reconstruction *ab-initio* parmi toutes les méthodes présentées jusque-là. En revanche, pour le raffinement, l'algorithme d'espérance maximisation proposée par RELION est plus adapté (le logiciel CryoSPARC utilise également une méthode d'espérance maximisation pour le raffinement [61]). Par ailleurs, Cryo-RANSAC peut éventuellement s'avérer pertinent dans certains cas, lorsqu'on a à faire à des jeux de données simples, car elle est très rapide : elle permet d'obtenir une reconstruction en moins d'une heure.

2.2.3 Apprentissage profond

Depuis 2012, après la publication de AlexNet par Krizhevsky *et al.* [62], l'apprentissage profond a révolutionné la vision par ordinateur en montrant des résultats sans précédents pour des tâches variées telles que la classification, la détection d'objets, la segmentation, la génération d'images ou encore la reconstruction. C'est donc naturellement que les experts du domaine de la RPI en cryo-ME ont souhaité expérimenter des méthodes d'apprentissage profond, afin de savoir si elles

4. https://www.ebi.ac.uk/emdb/statistics/emdb_software_year

5. https://relion.readthedocs.io/en/release-3.1/SPA_tutorial/InitialModel.html

étaient en mesure de surpasser les résultats déjà très satisfaisants de CryoSPARC et RELION. C’est l’objet de la section suivante.

L’apprentissage profond a fait son entrée dans le domaine de la RPI en cryo-ME en 2021 avec les publications des méthodes CryoPoseNet [63], CryoGAN [64], CryoDRGN [65] et CryoAI [66] qui sont venues, pour certaines, concurrencer les résultats de CryoSPARC, sans pour autant les dépasser significativement. Ces algorithmes, dédiés à la reconstruction *ab-initio*, diffèrent les uns des autres dans la manière dont les poses sont traitées (estimation amortie [63, 66] , estimation non-amortie [65] ou pas d’estimation [64]) et dans la représentation du volume (représentation explicite [63, 64] ou représentation implicite [65, 66]).

2.2.3.1 Estimation des poses amortie, non-amortie ou sans estimation

L’inférence non-amortie considère indépendamment les uns des autres les N problèmes d’estimation des poses associées aux N vues. Tous les algorithmes présentés en section 2.2.1 se basent sur de l’inférence non-amortie. L’inconvénient majeur est que le temps de calcul augmente alors avec le nombre de vues N , car il y aura d’autant plus de poses à estimer.

En revanche, l’inférence amortie ne cherche plus à résoudre indépendamment les N problèmes mais les fusionne en un seul. Elle tente pour cela d’estimer les paramètres η d’une fonction g_η , qui prend en entrée une vue (ou classe moyenne) et renvoie la pose qui lui est associée. Cette fonction prend la forme d’un réseau de neurones convolutionnel, suivi d’un réseau densément connecté, dont les poids sont progressivement ajustés au cours de l’entraînement. Il est important de préciser que l’apprentissage de g_η ne nécessite pas de supervision sur les poses, grâce à un apprentissage de bout en bout. A chaque itération, une vue y_l (ou un mini-batch) est sélectionnée au hasard et passe à travers le réseau g_η , qui renvoie les poses qui lui sont associées. Ces poses sont utilisées pour tourner et traduire le volume en cours de reconstruction, qui est dans le même temps projeté puis convolué avec la CTF du microscope, pour donner la vue estimée $\hat{y}_l$ qui est comparée à la vue d’entrée à l’aide d’une fonction de distance. Le gradient de cette fonction de coût est rétro-propagé pour ajuster les paramètres du volume et de l’estimateur de poses g_η . La figure 2.1 illustre le schéma général d’un entraînement de bout en bout utilisant une estimation amortie des poses. L’avantage majeur des algorithmes utilisant l’inférence amortie des poses est que leur temps de reconstruction est indépendant du nombre de vues, ils sont donc plus rapide en général que les algorithmes basés sur une inférence non-amortie. Néanmoins, l’inférence amortie peut s’avérer moins précise, surtout s’il n’y a pas suffisamment de vues pour permettre l’apprentissage de g_η .

La méthode CryoGAN [64], développée par Gupta et McCann, ne requiert pas l’estimation des poses associées à chacune des vues, et n’est pas non plus basée sur une marginalisation par rapport aux poses. Les auteurs proposent un algorithme qui se fonde sur le principe des réseaux de neurones antagonistes. A chaque itération, un ensemble de vues est généré à partir de l’estimation courante du volume en simulant le modèle d’observation, avec des poses tirées aléatoirement. Un discriminateur tente ensuite de distinguer les vues ainsi simulées des vues réelles du jeu de données. Au cours de l’apprentissage, l’estimation du volume s’ajuste petit à petit de sorte à ce que le discriminateur ne puisse distinguer les vues simulées des vues du jeu de données. De son côté, le discriminateur ne cesse de progresser dans sa capacité à distinguer le vrai du faux. Tout comme les algorithmes basés sur l’inférence amortie des poses, le temps de calcul de CryoGAN n’augmente pas avec le nombre de vues. Néanmoins, leur algorithme ne donne pas pour l’instant de résultats compétitifs avec l’état de l’art, contrairement à la méthode CryoAI (basée sur une estimation amortie des poses).

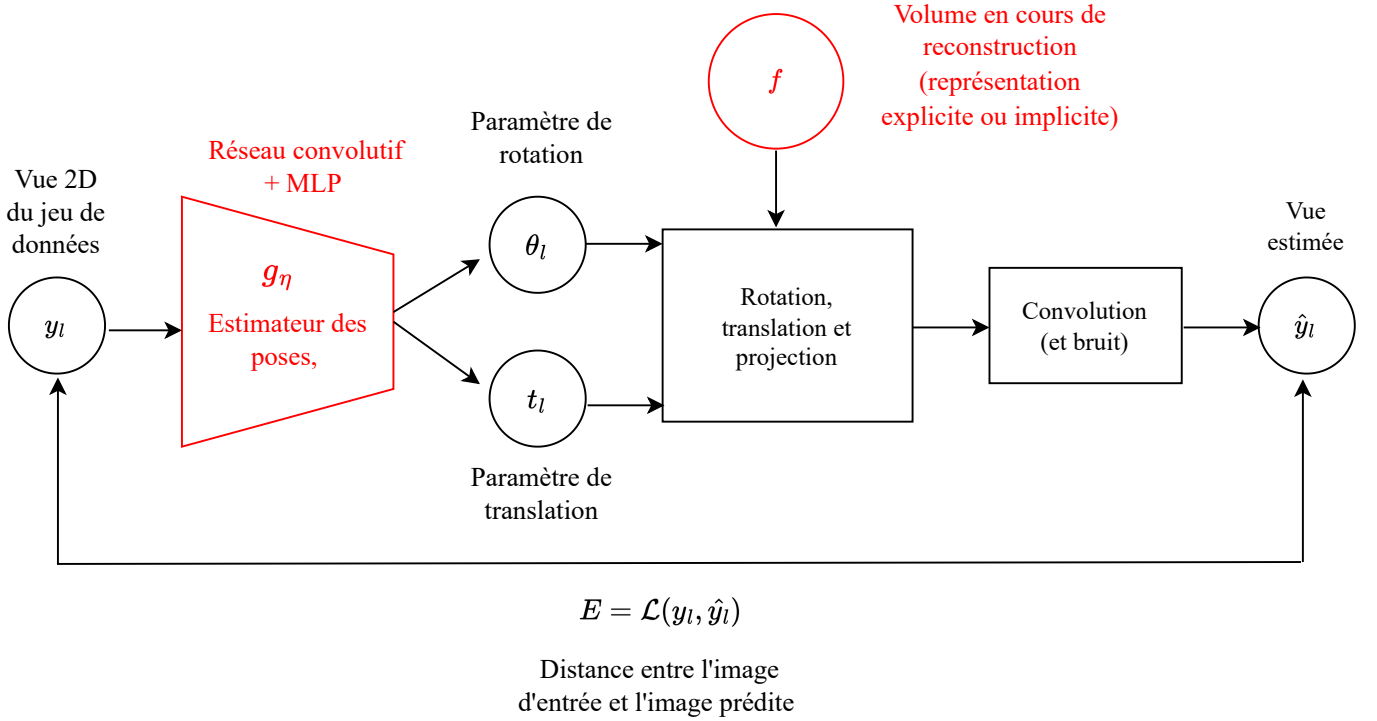


FIGURE 2.1 – Schéma-bloc général des méthodes utilisant une estimation amortie des poses. Les blocs dont les paramètres sont appris sont représentés en rouge (le volume f et l'estimateur des poses g_η). Leurs paramètres sont ajustés à l'aide du gradient de la fonction de coût E .

2.2.3.2 Représentation du volume : explicite ou implicite

Les algorithmes exposés précédemment, dans les sections 2.2.1 et 2.2.2, utilisent une représentation explicite du volume, il est représenté par les valeurs de ses voxels (ou frequels dans le domaine fréquentiel) sur un domaine 3D Ω discret et fini.

La représentation implicite suppose en revanche un domaine de définition de l'image continu, et le volume n'est pas représenté par les valeurs de chacun de ses voxels, mais par une fonction $f_\gamma : x \in \Omega \mapsto f_\gamma(x)$ qui prend en entrée les coordonnées x d'un voxel et renvoie l'intensité correspondante. L'apprentissage cherche alors à estimer les paramètres γ de cette fonction.

La méthode CryoDRGN [65], développée par Zhong *et al.*, utilise un perceptron multi-couche⁶ comme représentation implicite du volume f_γ . Ce dernier admet 3 neurones d'entrée (les coordonnées des voxels), un neurone de sortie (la valeur prédite associée aux coordonnées d'entrée) et 2 couches intermédiaires. Levy *et al.* utilisent dans CryoAI [66] une représentation implicite sous la forme de la somme de deux réseaux de neurones à représentation sinusoïdal (SIRENs pour "sinusoïdal representation network"). Le SIREN [67] est, tout comme le perceptron multi-couches, un réseau densément connecté mais utilise des fonctions d'activation sinusoïdales au lieu des ReLU.

Ce type de représentation peut potentiellement atteindre de meilleures résolutions que si une représentation explicite est adoptée, car une grille 3D fixe se prête assez mal aux opérations de

6. C'est-à-dire un réseau densément connecté admettant des fonctions d'activation ReLU (pour Rectified Linear Unit, c'est une fonction nulle dans les entiers négatifs et égale à l'identité dans les entiers positifs)

rotations. En effet, le volume f tourné d'un angle θ évalué au point x s'écrit $f(R_\theta^t x)$. Or, lorsque l'on considère une représentation explicite, f n'est définie que sur une grille discrète de point, si $R_\theta^t x$ ne tombe pas sur un des points de la grille, une interpolation doit être effectuée pour calculer $f(R_\theta^t x)$; et cette interpolation peut induire des erreurs. En revanche, dans le cas de la représentation implicite, n'importe quelle coordonnée peut être mise en entrée de la fonction f_γ , sans qu'il n'y ait nécessité d'effectuer une interpolation.

2.2.3.3 Apport de l'apprentissage profond

CryoPoseNet [63], la première méthode publiée utilisant une estimation amortie des poses, ne permet pas d'obtenir des résultats satisfaisants sur des données réelles. D'autre part, ils ne comparent pas leurs résultats avec ceux de l'état de l'art.

CryoAI [66], basé sur une estimation amortie des poses et une représentation implicite du volume, expose des résultats comparatifs de leur méthode avec CryoPoseNet et CryoSPARC. Ils montrent sur plusieurs jeux de données, qu'après raffinement⁷, leur méthode donne des résultats similaires à ceux de CryoSPARC, et bien meilleurs que ceux de CryoPoseNet.

CryoDRGN [65], basé sur une estimation non amortie des poses et une représentation implicite du volume, se compare également à CryoSPARC. Sur des données synthétiques, ils obtiennent des résultats similaires à CryoSPARC. Sur des données réelles, leurs résultats sont moins bons sur deux jeux de données considérés comme simples et meilleurs sur un jeu de données considéré comme difficile.

Les auteurs de CryoGAN [64] ne se comparent pas explicitement à d'autres méthodes, mais ils précisent dans le texte que leur méthode n'est pas encore compétitive avec l'état de l'art.

Pour conclure, les algorithmes exposés dans cette section présentent pour certains des résultats analogues à l'état l'art des méthodes n'utilisant pas l'apprentissage profond (CryoDRGN et CryoAI), et pour d'autres de moins bons résultats (CryoPoseNet et CryoGAN). Néanmoins, ces résultats sont encore à considérer avec précaution, puisqu'il s'agit uniquement de comparaisons effectuées par les auteurs concernés. Par conséquent, elles sont sujettes au biais de pression à la publication. Des études comparatives effectuées par des chercheurs tiers sont donc encore attendues.

2.2.4 Hétérogénéité discrète des conformations

Nous avons pour l'instant abordé le cas de la reconstruction homogène. Cela suppose que toutes les vues du jeu de données sont issues du même modèle f , que l'on cherche à reconstruire. Cependant, cette hypothèse n'est pas toujours valide, car il existe des assemblages pouvant exister sous différents états stables. C'est ce que l'on appelle l'hétérogénéité des conformations. L'apprentissage consiste alors, non pas à reconstruire un volume unique, mais un ensemble de volumes, ainsi que d'associer à chacune des vues le volume dont elle est issue.

Les premières méthodes de reconstruction hétérogène classifiaient l'ensemble des vues en K ensembles disjoints [68], chacun étant supposé regrouper les vues issues d'une même conformation. A l'issue de cette classification, K reconstructions sont effectuées séparément. Cette méthode n'est aujourd'hui que très peu utilisée en pratique. En effet, le rapport signal à bruit des images de cryo-ME est trop important pour permettre l'obtention de résultats satisfaisants. Une approche plus robuste consiste à effectuer les reconstructions des K volumes conjointement à la classification [69]. Cela s'effectue en résolvant le problème de maximum de vraisemblance suivant :

7. Les méthodes CryoAI, CryoSPARC et CryoPoseNet sont utilisés pour la rconstruction *ab-initio* et les résultats sont raffinés par un algorithme EM

$$\{\hat{f}_k\}_{k \in \llbracket 1, K \rrbracket} = \arg \min_{\{f_k\}_{k \in \llbracket 1, K \rrbracket}} \mathcal{L}(\{f_k\}_{k \in \llbracket 1, K \rrbracket}), \quad (2.19)$$

avec $\{f_k\}_{k \in \llbracket 1, K \rrbracket}$ les K conformations à estimer et $\mathcal{L}$ la fonction de coût définie par :

$$\mathcal{L}(\{f_k\}_{k \in \llbracket 1, K \rrbracket}) = \sum_{l=1}^N \log \left(\sum_{k=1}^K \int_v p(y_l | f_k, v) dv \right). \quad (2.20)$$

Ce problème d'optimisation peut être résolu par un algorithme d'espérance-maximisation, similaire à celui utilisé pour la reconstruction homogène présenté en 2.2.2.1.1. La formule de mise à jour est dans ce cas :

$$f^{(n+1)} = \arg \max_f \mathcal{Q}^K(f, f^{(n)}), \quad (2.21)$$

avec

$$\mathcal{Q}^K(f, f^{(n)}) = \sum_{l=1}^N \sum_{k=1}^K \int_v p(v | y_l, f_k^{(n)}) \log(p(y_l | v, f) p(f)) dv. \quad (2.22)$$

Une limite importante de cette méthodologie est qu'elle considère un ensemble discret et fini de K conformations. Cela suppose d'une part de connaître à l'avance le nombre de conformations, mais cela omet surtout de prendre en compte que les assemblages macromoléculaires ne sont pas des entités parfaitement rigides mais ont un certain degré de flexibilité, les particules assignées à une même classe auront rarement des structures parfaitement identiques. Si l'on veut pouvoir observer cette flexibilité dans l'objet reconstruit, il y a donc nécessité de modéliser une hétérogénéité continue des conformations. C'est l'objet de la section suivante.

2.2.5 Hétérogénéité continue des conformations

Nous séparons l'état de l'art de la reconstruction hétérogène continue en 2 catégories, d'une part les méthodes opérant directement sur les densité de la structure (dans l'espace réel ou de Fourier), d'autre part les méthodes se basant sur une représentation dans un espace latent à faible dimension. Il existe une grande variété de types d'hétérogénéité continue, certaines méthodes sont donc plus adaptées que d'autres à traiter une hétérogénéité particulière.

2.2.5.1 Opérations sur les densités

Raffinement à corps multiples Les méthodes de raffinement à corps multiples [70], ou de raffinement focalisé [71, 72], représentent le volume comme une ensemble de sous-parties mobiles. Ces sous-parties, appelées corps, sont identiques pour toutes les conformations, ce qui varie d'une conformation à l'autre est leurs orientations relatives. Le raffinement à corps multiples cherche alors à estimer ces orientations, via un algorithme d'espérance-maximisation. Afin d'estimer l'orientation relative associée à un corps rigide particulier, un masque est appliqué de sorte à soustraire la contribution des autres corps à la vraisemblance conditionnelle.

Il existe une légère différence entre le raffinement focalisé et le raffinement à corps multiple. Pour le premier, les masques sont appliqués une seule fois avant de commencer le raffinement, alors que pour le second ils sont mis à jour à chaque itération en prenant en compte les modifications du volume effectuées aux itérations précédentes.

L'avantage de ces méthodes est leur simplicité de mise en œuvre. Bien qu'utiles dans certains cas, elles sont fortement limitées par le type d'hétérogénéité qu'elles peuvent modéliser et ne peuvent donc pas s'appliquer à tous les cas de figure.

Combinaison linéaire de volumes principaux Une autre approche effectuant des opérations sur les densités et offrant une plus grande flexibilité dans la modélisation de l’hétérogénéité, consiste à représenter le volume comme la somme d’un volume de référence f_0 et d’une combinaison linéaire de K composantes orthogonales ($f_1, \dots, f_k, \dots, f_K$) appelées ”volumes principaux” qui modélisent la variabilité autour de la référence. Formellement, une conformation quelconque se met sous la forme :

$$f = f_0 + \sum_{k=1}^K z_k f_k. \quad (2.23)$$

Le volume de référence f_0 et les poses associées à chacune des vues sont supposés connus, et sont obtenus grâce à une reconstruction homogène grossière. Selon cette formulation, le raffinement hétérogène consiste alors à estimer les volumes principaux (f_k) ainsi que les coefficients (z_k).

Si les volumes dans différentes conformations étaient connus, un choix naturel pour les volumes principaux serait d’appliquer la méthode d’analyse en composante principale (ACP) à ces conformations, afin d’obtenir les K directions de plus grande variabilité. Ces dernières s’obtiendraient en sélectionnant les K premiers vecteurs propres (i.e dont les valeurs propres associées sont les plus élevées) de la matrice de covariance Σ associée à ces conformations connues.

Étant donné qu’avant d’avoir effectué la reconstruction hétérogène, nous avons uniquement les vues à disposition et non le volume sous différentes conformations, cette méthode n’est pas directement applicable. Néanmoins, les auteurs de [73] (Tagare *et al.*) ont montré qu’il était possible, sous certaines conditions, d’estimer la matrice de covariance Σ directement à partir des vues, à l’aide d’une méthode d’ACP probabiliste. Les auteurs de 3DVA [74] (Punjani *et al.*) ont ensuite amélioré la méthode proposée par [73], lui permettant d’inclure un plus grand nombre de composantes, et d’atteindre de meilleures résolutions en un temps de calcul inférieur.

Les auteurs de [73] et [74] montrent que cette méthode témoigne de résultats satisfaisants pour certains cas d’application. Néanmoins, elle se base sur une représentation linéaire de l’espace des conformations. Dans des situations où cet espace présente une géométrie complexe et non-linéaire, il est nécessaire d’utiliser des modèles non-linéaires fondés sur une représentation dans un espace latent. C’est l’objet du paragraphe suivant.

2.2.5.2 Représentation des conformations dans un espace latent

Une deuxième catégorie de méthodes permettant d’estimer l’hétérogénéité continue se base sur une représentation des conformations dans un espace latent à faible dimension (CryoFIRE [75], CryoDRGN [65], ACE-HetEM [76], Cryofold [77], AtomVAE [78], E2gmm [79], Multi-CryoGAN [80], 3DFlex [81], DynaMight [82]). L’hétérogénéité est alors modélisée par une variable latente (ou variable de conformation) z_l . À chaque valeur de variable de conformation correspond une conformation possible de la structure dans le domaine réel, par l’intermédiaire d’une fonction non-linéaire pouvant prendre diverses formes, appelée décodeur.

Plus précisément, on considère itérativement chaque vue (ou chaque mini-batch) du jeu de données et, à chaque itération, la variable latente (ou variable de conformation) z_l correspondant à la vue courante passe à travers le décodeur. Cela permet de construire le volume associé f_{z_l} à partir de cette dernière. Puis, le volume ainsi construit est projeté selon les poses associées à la vue courante (qui peuvent être connues ou elles-même estimées), et la convolution avec la CTF est appliquée. L’algorithme cherche à faire en sorte que la projection ainsi obtenue soit proche de la vue courante du jeu de données. La première est donc comparée à la seconde à l’aide d’une

certaines fonction de coût (usuellement une différence quadratique moyenne), et le gradient de cette dernière est rétro-propagé afin d’ajuster les paramètres du modèle.

Bien que similaires dans le principe général, ces méthodes varient selon plusieurs points :

- Le décodeur utilisé : cela peut être une représentation implicite sous forme de réseau de neurones densément connectés [75, 65, 76], l’estimation d’un flux par rapport à un volume 3D canonique (et l’application de ce flux à la structure canonique) [81, 82] ou l’estimation d’une représentation sous la forme d’un mélange de fonctions gaussiennes [77, 78, 79, 82]
- La manière dont les variables de conformations z_i sont estimées : dans la plupart des cas, elles sont estimées de manière amortie [65, 75, 77, 78, 79, 82, 76] à l’aide d’un encodeur (un réseau de neurones convolutif suivi d’un réseau densément connecté) qui prend en entrée la vue courante y_i . Mais dans le cas de 3DFlex [81], elles sont estimées de manière non-amortie.
- Les auteurs de [75, 65, 76, 77, 78, 79, 82] proposent un système auto-encodeur (un encodeur suivi d’un décodeur), qui peut-être variationnel [65, 77, 78, 82] ou non [75, 79, 76]
- Une régularisation peut être imposée à la variable latente [65, 81], et/ou au volume [77, 78, 81, 82] mais pas toujours. La régularisation sur le volume permet d’imposer des contraintes qui modélisent une réalité physique.
- La reconstruction hétérogène peut être utilisée dans un second temps, après une reconstruction *ab-initio* grossière ne prenant pas en compte l’hétérogénéité [77, 79, 81, 82]. Cela permet d’estimer un volume canonique correspondant le mieux possible aux vues issues de diverses conformations. Lors du raffinement, les poses sont fixées à leurs valeurs estimées lors de la reconstruction *ab-initio*, simplifiant ainsi l’apprentissage des conformations. Néanmoins, dans des situations où les conformations sont trop éloignées les une des autres, l’estimation d’un volume canonique ne s’avère pas adapté, les poses doivent donc être estimées conjointement aux conformations [75, 65, 76, 78].

La figure 2.2 illustre le schéma bloc général de ces méthodes. Nous explicitons dans ce qui vient leurs spécificités, en séparant les méthodes en fonction du type de décodeur utilisé.

2.2.5.2.1 Représentation implicite du volume à l’aide d’un réseau de neurones Dans les méthodes CryoFIRE [75] et CryoDRGN [65], le décodeur est une représentation implicite du volume dans le domaine de Fourier sous la forme d’un réseau de neurones densément connecté (un SIREN ou un perceptron multicouche (MLP pour ”Multilayer perceptron” en anglais)). Nous avons déjà évoqué la représentation implicite dans le cadre de la reconstruction homogène (partie 2.2.3.2). Dans le cadre de la reconstruction hétérogène, la fonction f_γ , représentation implicite du volume, prend non seulement en entrée les coordonnées x d’un voxel mais également la variable latente z . La valeur $f_\gamma(x, z)$ correspond donc à l’intensité, dans le domaine de Fourier, de la conformation associée à la variable latente z au voxel de coordonnées x . La méthode CryoFIRE est une simple extension de CryoAI [66] au cas hétérogène. L’encodeur de CryoFIRE est constitué d’un bloc commun, qui est ensuite séparé en 2 parties, l’une estimant les poses et l’autre la variable de conformation z . CryoFIRE et CryoDRGN ne supposent pas de connaissances *a-priori* sur les poses (*i.e.* il n’y a pas de reconstruction canonique effectuée en amont), et la distinction majeure entre ces deux méthodes est que la première estime les poses de manière amortie, et la seconde de manière non-amortie.

Ces deux méthodes effectuent les étapes du passage à travers le décodeur et de la projection en une seule fois. Il suffit en effet de considérer une grille 2D, qui est tournée et translatée à l’aide des poses ξ_i et de mettre les coordonnées de cette grille en entrée du décodeur (*i.e.* la représentation implicite du volume dans le domaine de Fourier). Cela revient à prendre une section du volume

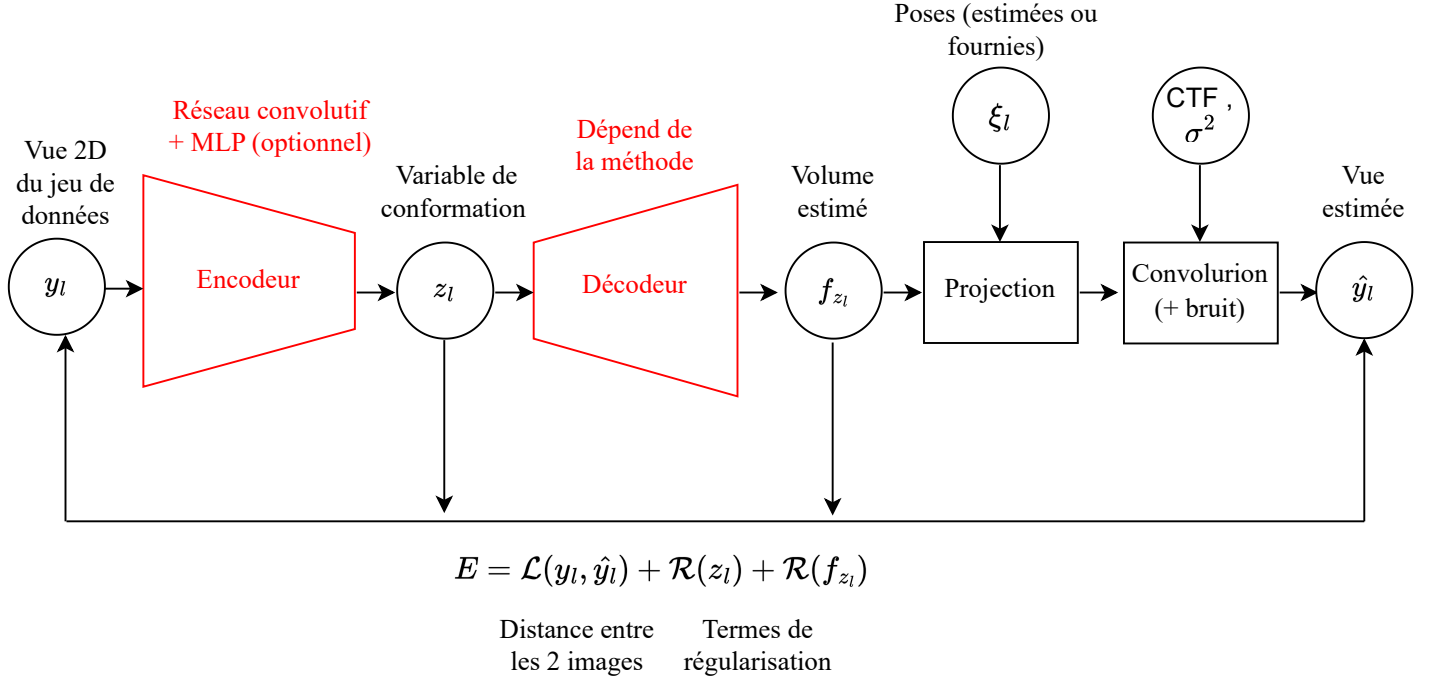


FIGURE 2.2 – Schéma-bloc général des méthodes de reconstruction d'hétérogénéité continue basée sur l'estimation d'une variable latente. Les blocs dont les paramètres sont appris sont représentés en rouge (le décodeur et l'encodeur). Leurs paramètres sont ajustés à l'aide du gradient de la fonction de coût E . L'encodeur est optionnel, lorsque celui-ci n'est pas présent, c'est directement la variable z_l qui est apprise (inférence non-amortie).

dans le domaine de Fourier selon une certaine orientation, ce qui correspond à une projection dans le domaine réel selon cette même orientation (d'après le théorème de la coupe centrale de Fourier).

L'avantage de ces représentations est qu'elles peuvent modéliser une très grande variabilité d'hétérogénéité, ce qui peut être également considéré comme un inconvénient, car lorsque l'on a un *a-priori* sur le type d'hétérogénéité en présence, il peut être plus adapté d'utiliser une méthode lui étant spécifiquement dédiée.

La méthode CryoFIRE, de Levy *et al.*, est bien moins coûteuse en temps de calcul que CryoDRGN, car elle ne nécessite pas d'effectuer une recherche exhaustive sur les poses à chaque itération. Néanmoins, cette méthode basée sur une estimation amortie des poses est plus susceptible de rester bloquée dans des minima locaux que CryoDRGN, ce qui a tendance à générer en particulier des reconstructions ayant une fausse symétrie planaire. C'est pourquoi les auteurs de CryoFIRE ont implémenté ce qu'ils appellent une « fonction de coût symétrique », qui a permis d'améliorer significativement les résultats. Cependant, en pratique cette fonction de coût symétrique ne permet pas d'éviter systématiquement les minima locaux. C'est pourquoi Chen *et al.* ont développé une amélioration de CryoFIRE qu'ils appellent ACE-HetEM [76]. Cette méthode adopte une architecture similaire à celle de CryoFIRE et l'apprentissage alterne entre une phase d'entraînement image-à-image, similaire à l'entraînement de CryoFIRE (mais sans la fonction de

coût symétrique) et une phase d’entraînement pose-à-pose. L’entraînement pose-à-pose permet de stabiliser l’apprentissage et ainsi d’éviter les minima locaux. Nous reviendrons sur cette approche dans la section 4.2.3.

2.2.5.2.2 Représentation pseudo-atomique Les méthodes CryoFold [77], AtomVAE [78] et E2gmm [79] (publiées la même année (2021)), ainsi que DynaMight [82] (publiée en 2023), utilisent un décodeur qui fait correspondre la variable latente z à une représentation sous forme de mélange (i.e combinaison linéaire) de fonctions gaussiennes (ou de fonctions radiales). Ces gaussiennes sont autrement appelées *pseudo-atomes*. Un pseudo-atome peut représenter en fonction des cas une protéine, un acide aminé ou même un atome. L’avantage de ce type de représentation est de pouvoir intégrer plus aisément une information *a-priori* à la structure, en ajoutant un terme de régularisation qui modélise une contrainte stéréochimique entre les pseudo-atomes. Un autre avantage est que la reconstruction est directement interprétable car cela permet de visualiser directement la position des éléments de la structure. Il est en effet courant, dans le cas des méthodes reconstruisant une densité (i.e une représentation voxélique ou fréquentielle), d’appliquer une étape de post traitement faisant correspondre cette densité à un modèle pseudo-atomique, ce qui permet ensuite d’étudier plus aisément la topologie du complexe ou ses propriétés hydrodynamiques [83]. Reconstruire directement sous la forme d’un modèle pseudo-atomique permet d’éviter cette étape de post traitement, qui peut être coûteuse en temps et difficile à mettre en œuvre [77]. Néanmoins, cette représentation n’est pas adaptée à tous les types d’hétérogénéité. En particulier, elle ne modélise pas l’hétérogénéité compositionnelle puisqu’elle suppose que la composition atomique de la structure est connue et fixe. Nous exposons dans la suite de cette section les spécificités des 4 articles précédemment cités.

La méthode AtomVAE [78] utilise une structure de référence appelée squelette constituée d’un nombre prédéfini d’atomes. Son décodeur permet de faire correspondre ce squelette à la structure estimée. Il prédit pour cela, à partir de la vue d’entrée, la rotation et la translation de chaque atome par rapport à sa position dans le squelette. Une régularisation sur la représentation pseudo-atomique est imposée en ajoutant à la fonction de coût la différence quadratique moyenne entre les distances séparant deux atomes consécutifs du squelette et la longueur supposée d’une liaison. AtomVAE ne se base pas sur l’utilisation d’un volume canonique, et estime donc conjointement les poses et les conformations. Leur système démêle (“disentangle”) néanmoins l’estimation des poses de celle des conformations. Un premier réseau de neurones fait correspondre la vue d’entrée à un espace latent intermédiaire. Ce dernier est ensuite séparé à l’aide de deux réseaux de neurones, l’un estimant les poses et l’autre la variable de conformation. D’autre part, les auteurs précisent que pour assurer la convergence il est nécessaire d’effectuer une première phase où seulement les poses sont estimées. Malheureusement, leur algorithme doit encore faire ses preuves puisqu’il ne permet l’obtention de résultats corrects que sur des données synthétiques.

La méthode CryoFold [77] utilise une représentation des acides aminés de la structure. Chaque acide aminé est représenté par deux fonctions radiales (FR), l’une faisant partie du squelette de la structure à un écart type fixe (identique pour toutes les FR), et l’autre FR n’en faisant pas partie a un écart type variable⁸. Le décodeur prédit directement les positions et écart-types variables des FR (contrairement à AtomVAE qui prédit une rotation et translation par rapport à une structure de référence). Ils imposent deux contraintes de régularisation. Le premier terme, identique à celui

8. Pour cela il y a une incertitude. Leur schéma montre des écart-types variables mais ce n’est pas précisé dans le texte. D’autre part, ils exposent le nombre de variables que le modèle estime et celui-ci laisse penser que les écart-types des FR qui ne sont pas dans le squelette sont effectivement variables

de AtomVAE, assure que la distance entre deux atomes consécutifs soit proche de celle supposée (3,8 Å dans leur cas). Le second terme permet de limiter la distance entre les deux FR d'un acide aminé, il est similaire au premier mais ne pénalise la fonction de coût que lorsque le lien est supérieur à la valeur supposée de la liaison. Contrairement à AtomVAE, CryoFold suppose les poses connues, obtenues à l'aide d'une reconstruction *ab-initio* grossière, d'autre part il ne parvient pas non plus à obtenir des résultats sur un jeu de données réelles.

La méthode E2Gmm [79] est plus simple que AtomVAE et CryoFold. Leur décodeur prédit les positions et écart-types des gaussiennes représentant le volume mais n'imposent pas de contrainte de régularisation à la structure, et les poses sont supposées connues. Les auteurs parviennent à montrer des résultats sur des données réelles issues du Ribosome et de la protéine spike du SARS-VoV-2.

La méthode DynaMight [82], développée par Johannes Schwab en 2023 (de l'équipe de Scheres) et implémentée dans le logiciel RELION, parvient également à obtenir des résultats de reconstruction hétérogène satisfaisants sur des données réelles, en partant d'une reconstruction canonique *ab-initio*. Cette méthode utilise un décodeur qui, pour chaque gaussienne de la reconstruction canonique, estime un vecteur de translation modélisant le déplacement de cette dernière (DynaMight est donc également dans la catégorie des méthodes estimant un champ de déformation, dont nous reparlerons dans la section suivante). Les auteurs de DynaMight tentent, comme AtomVAE et CryoFold, d'imposer une contrainte de régularisation, modélisant une connaissance *a-priori* sur la structure du volume. Ils testent deux types de régularisation. La première est similaire à celles proposées par AtomVAE et CryoFold, les indices des pseudo-atomes liés entre eux sont supposés connus, ainsi que la longueur des liaisons, et un terme de régularisation tente de minimiser la variation autour de ces longueurs supposées. Leur seconde régularisation utilise moins de connaissance *a-priori*, et ne requiert pas de modèle atomique, elle impose simplement une contrainte de proximité entre gaussiennes voisines. Les auteurs mettent en avant le risque d'utiliser le premier type de régularisation, cela peut dans certains cas améliorer les résultats, mais peut également introduire un biais dans la reconstruction si le modèle atomique s'avère légèrement erroné.

Pour conclure ce paragraphe, un des intérêts majeurs de la représentation pseudo-atomique est de pouvoir imposer des contraintes stéréochimiques à la reconstruction, mais comme précisé par les auteurs de DynaMight, ce n'est pas sans risque car cela peut biaiser la reconstruction, et nous conduire à surestimer la résolution des résultats obtenus. En effet, la procédure courante de validation des résultats, consiste à reconstruire indépendamment sur deux sous-ensembles disjoints des données, et de comparer les deux reconstructions à l'aide de la corrélation des anneaux de Fourier (FSC) (nous en reparlerons dans le paragraphe 3.6.1.2.3). Or, si le même biais est introduit dans les deux reconstructions séparées, la FSC sera élevée alors même que la reconstruction n'est pas fidèle à la réalité. Le champ de recherche reste donc encore ouvert, à la fois pour développer de nouvelles procédures de validation et pour explorer des méthodes de régularisation plus sophistiquées, introduisant moins de biais au processus de reconstruction, tout en étant plus complexe qu'une simple modélisation de proximité entre pseudo-atomes voisins.

2.2.5.2.3 Estimation d'un champ de déformation (3DFlex) La méthode 3DFlex [81], proposée par Punjani *et al.* et implémentée dans le logiciel CryoSPARC, se base sur l'hypothèse centrale que les différentes conformations f_z de l'assemblage macro-moléculaire sont reliées les unes aux autres par la déformation d'une structure unique V , appelée structure canonique. 3DFlex cherche alors à estimer V ainsi que les champs de déformation reliant chaque conformation à V . Formellement, la conformation f_{z_l} associée à la vue y_l s'écrit :

$$f_{z_l} = D(u(z_l), V), \quad (2.24)$$

avec $u(z_l)$ le champ de déformation associé à la vue d'indice l , et D un opérateur de convection, qui déforme la structure canonique V en utilisant le champ de déformation $u(z_l)$. La fonction u est modélisée par un réseau de neurones densément connecté dont les poids sont ajustés au cours de l'apprentissage. Par ailleurs, l'algorithme impose une contrainte de continuité et de rigidité locale du champ de déformation. Cela permet en particulier d'éviter de trop grandes déformations qui sont physiquement improbables. Bien que non précisé dans l'article, cette contrainte de régularisation pose le même problème que celui évoqué par les auteurs de DynaMight dans le paragraphe précédent (c.a.d le risque de biaiser la reconstruction et de conduire à une surestimation de la qualité des résultats).

La méthode DynaMight [82] se base également sur l'estimation d'un champ de déformation, en représentant la structure canonique, non pas dans le domaine voxelique comme 3DFlex, mais comme une combinaison de fonctions gaussiennes. En raison de l'absence de résultats comparatifs entre les deux méthodes, mais surtout de l'absence de procédure de validation fiable, il n'est pas possible de savoir laquelle des deux représentations est la plus adaptée à l'estimation du champ de déformation.

L'avantage de l'estimation d'un champ de déformation, par rapport à des méthodes comme CryoFIRE ou CryoDRGN, est de pouvoir modéliser le déplacement des protéines et la préservation de la géométrie locale tout en permettant des déformations non-rigides, ce qui n'est pas possible avec les méthodes de raffinement à corps multiples). D'autre part, contrairement aux méthodes pseudo-atomiques précédemment présentées, 3DFlex estime non seulement le champ de déformation mais également la structure canonique par rapport à laquelle toutes les conformations peuvent s'aligner. De plus, 3DFlex a fait ses preuves sur 5 jeux de données réelles, même si la validation quantitative des résultats de reconstruction hétérogène est, comme nous l'avons vu, encore un problème ouvert [84, 82].

Néanmoins, 3DFlex et DynaMight présentent deux faiblesses par rapport à CryoFIRE ou CryoDRGN. D'une part, ils supposent que les poses sont connues, donc obtenues à partir d'une reconstruction homogène *ab-initio*, il faut donc que l'hétérogénéité ne soit pas trop importante pour permettre une reconstruction homogène initiale. D'autre part, tout comme les méthodes pseudo-atomiques, 3DFlex n'est pas adapté pour traiter l'hétérogénéité compositionnelle.

2.2.5.3 Quelle méthode de reconstruction hétérogène choisir ?

Étant donné le grand nombre de méthodes de reconstruction hétérogène à disposition, la question du choix de la méthode se pose naturellement. Il n'y a en réalité pas de réponse absolue, ce choix dépend en bonne partie de connaissances *a-priori* sur le type d'hétérogénéité que l'on souhaite modéliser. Si l'on sait qu'il existe un nombre discrets de conformations et que les légères variations (tel que des petits mouvements d'atomes ou de filaments) autour de ces conformations sont négligeables (et qu'on ne souhaite pas les modéliser), une méthode de reconstruction hétérogène discrète sera plus adaptée.

Dans le cas de l'hétérogénéité continue, plusieurs cas de figure se posent également. Si l'on sait que la structure est composée d'un nombre discret de corps rigides et que l'on connaît leurs positions et formes, la méthode de raffinement à corps multiple est le choix idéal. Néanmoins, si l'on veut également modéliser des déformations non-rigides, les méthodes 3DFlex ou DynaMight, estimant un champ de déformation sont plus adaptées. Si l'on veut imposer une contrainte

stéréochimique à la structure, imposant par exemple la distance entre 2 atomes consécutifs du squelette de la structure, la représentation pseudo-atomique est plus adaptée (mais la régularisation présente le risque de biaiser les résultats). Enfin, si l'on a aucune connaissance *a-priori* sur le type d'hétérogénéité en présence CryoFIRE ou CryoDRGN seront plus adaptées, ces méthodes permettent en particulier de modéliser une hétérogénéité compositionnelle. L'avantage de CryoFIRE par rapport à CryoDRGN est qu'elle est moins coûteuse en temps de calcul, mais elle est plus susceptible de rester bloquer dans un minimum local que CryoDRGN. Néanmoins l'entraînement pose-à-pose ajouté à CryoFIRE par Chen *et al.* [76] rend la méthode plus robuste, sans augmenter significativement le temps de calcul. Avec cette amélioration, CryoFIRE sera très probablement plus adapté que CryoDRGN dans la plupart des cas.

Il est également important de savoir si les variations entre différentes conformations sont suffisamment faibles pour permettre d'obtenir une première reconstruction canonique, à l'aide d'un algorithme de reconstruction homogène. Si ce n'est pas le cas, il faudra choisir une approche qui permet d'estimer conjointement l'hétérogénéité des conformations et les poses (comme CryoFIRE ou CryoDRGN par exemple).

2.3 Comparaison des modalités cryo-ME et microscopie par fluorescence

Nous avons vu en partie 2.1 qu'il existait très peu de méthodes de RPI en microscopie par fluorescence, et une seule prenant en compte le modèle d'imagerie 3D approprié, basées sur une convolution avec la réponse impulsionnelle du microscope [27]. La partie 2.2 montre clairement qu'il existe un nombre bien plus important de méthodes dédiées à la RPI en cryo-ME, à la fois pour la reconstruction homogène et pour la reconstruction hétérogène. Le domaine est en effet bien plus ancien que la RPI en fluorescence.

Il est néanmoins important de rappeler que la RPI en fluorescence diffère sur plusieurs points de la RPI en cryo-ME, ce qui empêche la simple application de méthodes de reconstruction en cryo-ME sur des données de fluorescence.

D'abord, les données d'entrées sont différentes. Dans le cas de la fluorescence, les images sont des données 3D : le modèle d'acquisition est une convolution 3D avec la réponse impulsionnelle. Ensuite, les images ont un RSB (rapport signal à bruit) de l'ordre de 10 (10 db) [85]. Dans le cas de la cryo-ME il s'agit de données 2D : le modèle d'observation se caractérise par une projection selon l'axe du microscope suivie d'une convolution 2D. De plus, les images sont fortement bruitées : le RSB est de l'ordre de 0,1 (-10 db) [86]. Le faible niveau du RSB en cryo-ME s'explique par le fait que les rayons d'électrons utilisés en cryo-ME dégradent l'échantillon observé, ce qui limite la dose totale d'électrons qui peut être utilisée [11].

Le modèle d'observation sous forme de projection ainsi que le RSB très faible des images de cryo-ME ont pour conséquence qu'une vue acquise par cryo-ME contient très peu d'information sur l'objet original. Pour obtenir une reconstruction à une résolution satisfaisante en cryo-ME, il est donc nécessaire d'utiliser un très grand nombre de vues (de l'ordre de quelques dizaines de milliers). Ce grand nombre de données à traiter n'est pas si problématique en cryo-ME car l'application du modèle d'observation se fait très rapidement en tirant partie du théorème de la coupe centrale de Fourier.

En microscopie par fluorescence, le RSB plus faible ainsi que la reconstruction à partir de vues 3D anisotropes plutôt que de projections 2D permet l'obtention d'une reconstruction à résolution

satisfaisante à partir d'un nombre bien moins important de vues qu'en cryo-ME (quelques dizaines). Une vue 3D anisotrope contient en effet bien plus d'information sur l'objet original qu'une projection 2D. En revanche, l'application du modèle d'observation, qui doit être effectuée à de nombreuses reprises au cours de l'apprentissage, est bien plus longue qu'en cryo-ME. Ce dernier est en effet constitué d'une rotation 3D et d'une convolution 3D, ne pouvant pas être accélérées par le théorème de la coupe centrale de Fourier.

Je propose donc dans ma thèse deux nouvelles méthodes de reconstruction *ab-initio* pour la microscopie par fluorescence, qui s'inspirent des algorithmes de RPI en cryo-ME. La première, présentée au chapitre 3, est dédiée à la reconstruction homogène et a fait l'objet, en 2023, d'une publication dans le journal IEEE Transactions on Computational Imaging [87]. La seconde, exposée au chapitre 4 est une méthode basée sur l'apprentissage profond que nous avons appliqué à la reconstruction hétérogène, mais qui peut également être utilisée dans le cadre de la reconstruction homogène.

Chapitre 3

Reconstruction homogène

3.1 Positionnement et contributions

Nous exposons dans ce chapitre la méthode de RPI en microscopie par fluorescence que nous avons développée. Elle se base sur une reformulation du problème d’optimisation conjoint original, par rapport au volume, aux rotations et aux translations en un schéma d’optimisation hiérarchique qui découple la minimisation par rapport à chaque groupe de variables, en 3 niveaux imbriqués. Le volume est estimé au niveau le plus élevé grâce à de l’optimisation stochastique, et le succès de notre approche réside dans le développement de solveurs approximatifs rapides pour la minimisation par rapport aux rotations et translations aux niveaux inférieurs. Grâce à cette approche, le coût de calcul d’une reconstruction est réduit à moins de 30 minutes, et ne requiert pas d’initialisation. De plus, notre approche est peu sensible aux hyper-paramètres, ce qui simplifie son utilisation en pratique. Nous démontrons sur plusieurs ensembles de données synthétiques que nous surpassons les résultats de [27] en termes de résolution et d’évaluation visuelle. Enfin, nous montrons le potentiel de notre approche sur des données réelles de centrioles acquises grâce à une technique de microscopie par expansion.

Notre travail s’appuie sur la méthode de reconstruction dédiée à la cryo-ME CryoSPARC [57]. Nous nous en éloignons sur les principaux points suivants :

- CryoSPARC ne peut pas être appliqué à nos données, car il est conçu pour un modèle direct de projection en 2D. Nous développons une méthode dédiée aux modèles de convolution en 3D
- Nous estimons les poses avec un schéma d’optimisation multi-niveau, tandis que la stratégie de [57] consiste à reconstruire le volume sans estimation de poses en maximisant une vraisemblance marginalisée.
- Nous estimons les translations au dernier niveau de notre schéma hiérarchique à l’aide d’un algorithme rapide de corrélation de phase. La séparation de l’estimation des translations et des rotations est nécessaire pour obtenir un temps de calcul raisonnable dans notre cas 3D

Nous avons par ailleurs développé deux variantes de la méthode. La première utilise une représentation sous forme de mélange de gaussiennes (gmm) du volume à reconstruire, et la seconde utilise une représentation gmm à la fois du volume et des vues. Néanmoins, ces variantes (surtout la seconde) présentent des performances inférieures à la méthode originale.

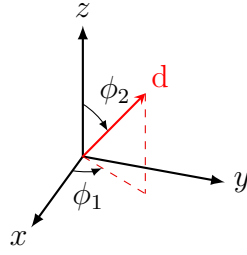


FIGURE 3.1 – Représentation d’une direction d (point sur la sphère unité) avec son azimuth ϕ_1 et son inclinaison ϕ_2 .

3.2 Formulation du problème

Cette section expose la formulation mathématique du problème à résoudre. Nous commençons par introduire le système de coordonnées que nous utilisons pour représenter les rotations, ainsi qu’une discrétisation de l’espace des matrices de rotation ($\text{SO}(3)$). Nous exposons ensuite le modèle d’acquisition des images (ou modèle d’observation), qui modélise le passage à travers le microscope, c’est-à-dire l’obtention des vues à résolution anisotrope à partir de l’objet original (que l’on souhaite reconstruire). Enfin, nous donnons l’expression de la fonction de coût que l’on cherche à minimiser.

3.2.1 Prérequis sur la représentation des rotations

Nous utilisons la représentation axe-angle (d, ψ) pour représenter une rotation vectorielle 3D, où $d \in \mathbb{R}^3$ est l’axe de rotation, et $\psi \in [0, 2\pi[$ est l’angle de rotation autour de cet axe. L’axe d est un vecteur unitaire défini comme suit :

$$d = \begin{pmatrix} \cos(\phi_1) \sin(\phi_2) \\ \sin(\phi_1) \sin(\phi_2) \\ \cos(\phi_2) \end{pmatrix}, \quad (3.1)$$

où $\phi_1 \in [0, 2\pi[$ est l’azimut et $\phi_2 \in [0, \pi[$ est l’inclinaison (voir figure 3.1). Notre méthode nécessite une discrétisation de $\text{SO}(3)$. Nous l’écrivons comme suit : $\{\theta_{i,j} = (\phi_{1,j}, \phi_{2,j}, \psi_i) \mid (j, i) \in \llbracket 0, M_d - 1 \rrbracket \times \llbracket 0, M_\psi - 1 \rrbracket\}$, où M_d et M_ψ sont les tailles des échantillons discrétisés de (ϕ_1, ϕ_2) et d , respectivement. Pour ψ , nous choisissons une discrétisation uniforme de $[0, 2\pi[: \forall i \in \llbracket 0, M_\psi - 1 \rrbracket, \psi_i = 2\pi i / M_\psi$. Pour (ϕ_1, ϕ_2) , nous utilisons la discrétisation de Fibonacci, qui donne une discrétisation quasiment uniforme de la sphère unité. Elle se formalise de la manière suivante :

$$\forall j \in \llbracket 0, M_d - 1 \rrbracket, \quad \phi_{1,j} \equiv \frac{2\pi j}{G} [2\pi] \quad (3.2)$$

$$\phi_{2,j} = \arccos\left(1 - \frac{2j + 1}{M_d}\right), \quad (3.3)$$

avec $G = \frac{1+\sqrt{5}}{2}$ le nombre d’or.

3.2.2 Modèle d'observation

Soit $\{y_l\}_{l \in \llbracket 1, N \rrbracket}$ un ensemble de N volumes (nous appelons volume une fonction $\Omega \rightarrow \mathbb{R}$ où Ω est un domaine borné inclus dans $\mathbb{R}^3$), où y_l est une vue contenant une seule particule, avec une orientation $\theta_l \in \mathbb{R}^3$ et une translation $t_l \in \mathbb{R}^3$. À partir de cet ensemble de vues, l'objectif est de reconstruire un volume f qui représente la particule observée. Le modèle d'observation appartient à la catégorie des modalités convolutionnelles et est défini par :

$$y_l(x) = (h * \mathcal{T}_{t_l}(\mathcal{R}_{\theta_l}(f)))(x) + \varepsilon(x), \quad (3.4)$$

où h est la PSF du microscope, ε est un bruit additif gaussien, $\mathcal{T}_t$ est un opérateur de translation de vecteur t défini par

$$\mathcal{T}_t(f)(x) = f(x - t), \quad (3.5)$$

et $\mathcal{R}_\theta$ est un opérateur de rotation d'angle θ défini par :

$$\mathcal{R}_\theta(f)(x) = f(R_\theta^T x), \quad (3.6)$$

où $R_\theta \in SO_3(\mathbb{R})$ est la matrice de rotation d'angle θ . Des exemples de volumes f et de vues y_l sont représentés sur la figure 3.4. Nous ne faisons pas d'hypothèse particulière sur la PSF, même si en pratique elle a une forme anisotrope, plus allongée selon l'axe Z [88], ce qui crée l'anisotropie de résolution.

3.2.3 Problème d'optimisation joint

Nous formulons le problème de reconstruction comme une estimation du maximum de vraisemblance du volume, des paramètres de rotation et de translation, ce qui revient à résoudre le problème des moindres carrés suivant :

$$f^*, \Theta^*, T^* = \arg \min_{f, \Theta, T} E(f, \Theta, T), \quad (3.7)$$

où

$$E(f, \Theta, T) = \sum_{l=1}^N \|y_l - h * \mathcal{T}_{t_l}(\mathcal{R}_{\theta_l}(f))\|_2^2, \quad (3.8)$$

avec

$$\Theta = \{\theta_l\}_{l \in \llbracket 1, N \rrbracket} \text{ et } T = \{t_l\}_{l \in \llbracket 1, N \rrbracket}.$$

Nous avons constaté expérimentalement que l'utilisation d'une régularisation dans l'équation (3.8) n'améliorait pas les résultats (nous avons envisagé une pénalisation l_2 des coefficients de Fourier). Alors que la régularisation joue un rôle important dans la déconvolution à partir d'une seule vue pour éviter l'apprentissage du bruit, la somme sur N vues dans l'équation (3.8) induit un effet de moyennage qui agit comme une contrainte de régularité. Par conséquent, l'ajout d'un terme de régularisation s'est avéré inutile dans notre cas (pour plus de détails, voir 3.6.1.4.5).

Pour réduire le temps de calcul, nous formulons l'énergie (3.8) dans le domaine de Fourier. Cela nous permet d'effectuer les convolutions avec de simples multiplications terme à terme. De plus, la rotation et la transformée de Fourier commutent, et la translation dans le domaine de Fourier est un simple décalage de phase. Formellement, nous avons donc :

$$\mathcal{F}(h * \mathcal{T}_{t_l}(\mathcal{R}_{\theta_l}(f))) = \hat{h}\rho_{t_l}\mathcal{R}_{\theta_l}(\hat{f}), \quad (3.9)$$

où $\mathcal{F}$ est la transformée de Fourier, ρ_{t_l} est un facteur de phase défini par $\rho_{t_l} : \omega \mapsto e^{it_l\omega}$, et nous utilisons la notation $\mathcal{F}(f) = \hat{f}$. L'énergie dans le domaine de Fourier est obtenue par le théorème de Parseval :

$$E(f, \Theta, T) = \sum_{l=1}^N \|\hat{y}_l - \hat{h}\rho_{t_l}\mathcal{R}_{\theta_l}(\hat{f})\|_2^2. \quad (3.10)$$

L'objet de la section suivante est de proposer une méthode d'optimisation robuste pour résoudre le problème (3.7).

3.3 Méthode

Le problème d'optimisation (3.7) est particulièrement non-convexe avec de nombreux minima locaux, ce qui rend impossible une optimisation simultanée par rapport à toutes les variables. Il est similaire dans une certaine mesure aux problèmes d'auto-calibration en microscopie, où le volume et certains paramètres du modèle d'observation sont estimés conjointement. Le schéma d'optimisation alternée couramment utilisé pour l'auto-calibration repose sur des modèles d'observation théoriques (tels que les modèles PSF dans la déconvolution aveugle) qui fournissent une bonne initialisation des paramètres, permettant ainsi d'éviter plus aisément les minima locaux. Dans notre cas, un schéma alternatif est voué à l'échec car nous ne disposons d'aucune estimation initiale du volume ou des poses.

3.3.1 Reformulation multi-niveaux du problème d'optimisation

Pour pallier ces difficultés, nous reformulons le problème (3.7) en une formulation multi-niveaux équivalente :

$$\hat{f}^* = \arg \min_{\hat{f}} E(f, \Theta^*, T^*) \quad (3.11a)$$

$$\text{t.q. } \Theta^* = \arg \min_{\Theta} E(f, \Theta, T^*) \quad (3.11b)$$

$$\text{t.q. } T^* = \arg \min_T E(f, \Theta, T). \quad (3.11c)$$

Dans ce schéma, un sous-problème à un niveau donné dans (3.11) est imbriqué en tant que contrainte dans le sous-problème du niveau supérieur. L'intérêt principal de cette reformulation est de décomposer le problème d'origine en trois sous-problèmes plus simples, comme dans le cas de l'optimisation alternée. Cependant, à la différence d'un schéma d'optimisation alternée, les points localement optimaux du problème joint (3.7) sont également des points optimaux de la formulation multi-niveaux. En contrepartie, si une méthode itérative est utilisée pour résoudre un problème à un niveau donné, le problème au niveau inférieur doit être résolu à chaque itération. Ainsi, le coût de calcul de cette optimisation imbriquée peut devenir prohibitif si les solveurs des sous-problèmes ne sont pas suffisamment efficaces (pour plus de détails sur l'optimisation multi-niveaux, voir [89]). Dans les trois sous-sections suivantes, nous détaillons nos stratégies d'optimisation rapides et complémentaires pour les sous-problèmes de (3.11)

3.3.2 Optimisation par rapport au volume

Lorsque les paramètres des poses Θ et T sont connus, le problème de minimisation (3.11a) peut être résolu facilement de manière analytique [27]. Cependant, comme l'estimation des poses est formulée sous forme de contraintes (3.11b) et (3.11c), nous devons adopter une approche itérative afin de pouvoir mettre à jour les poses à chaque itération. De plus, le problème contraint n'est pas convexe, et la somme sur un nombre potentiellement élevé de N particules peut augmenter considérablement le coût de calcul. Pour surmonter ces problèmes, nous utilisons une stratégie simple de descente de gradient stochastique par batch. Le volume est mis à jour à l'aide de l'équation suivante :

$$\hat{f}^{(n+1)} = \hat{f}^{(n)} - \lambda \sum_{l \in b_n} \nabla_{\hat{f}} E_l(f^{(n)}, \theta_l^*, t_l^*), \quad (3.12)$$

où $E_l = \|\hat{y}_l - \hat{h} \rho_{t_l} \mathcal{R}_{\theta_l}(\hat{f}^{(n)})\|_2^2$ est l'énergie associée à l'indice l , λ est le pas de descente du gradient, et b_n est le batch considéré à l'itération n . Notre procédure d'optimisation est divisée en époques de sorte que nous imposons de parcourir l'ensemble des données à chaque époque. Le calcul du gradient est détaillé en Annexe 6.1.

La robustesse du gradient stochastique pour des problèmes non-convexes a été établie pour des fonctions objectif similaire dans le contexte de l'apprentissage de fonctions ou pour la cryo-ME [57], bien que les propriétés de convergence ne puissent être garanties que dans le cas convexe.

3.3.3 Optimisation par rapport aux orientations

Le problème d'optimisation secondaire (3.11b), dédié à l'estimation des orientations, doit être résolu à chaque itération du problème principal (3.11a) pour l'évaluation du gradient $\nabla_{\hat{f}} E_{l_n}$. Les N vues sont considérées indépendamment, de sorte que (3.11b) peut être réécrit comme suit :

$$\forall l \in \llbracket 1, N \rrbracket \quad \theta_l^* = \arg \min_{\theta_l} E_l(f, \theta_l, t_l^*). \quad (3.13)$$

Une approche basée sur le gradient pour résoudre (3.13) est très susceptible d'atteindre un minimum local médiocre en raison de la non-convexité du problème, de sorte qu'elle ne peut être utilisée que pour le raffinement. D'autre part, une recherche exhaustive sur une discrétisation de $SO(3)$ serait trop coûteuse en temps de calcul. En effet, comme nous le verrons plus loin, la résolution de (3.13) nécessite de calculer numériquement la rotation de l'objet en cours de reconstruction $f^{(n)}$ pour un grand nombre d'orientations différentes. Si l'on considère un pas de discrétisation de 1° , il y aurait à chaque époque $360^3 N = 4,6 \times 10^7 N$ rotations à effectuer, et comme la rotation d'une image de taille $50 \times 50 \times 50$ prend $7 \times 10^{-4} s$ sur une carte graphique Titan X, si l'on considère $N = 20$, et que l'algorithme met en général une trentaine d'époques à converger, le temps total de reconstruction serait supérieur à 223 jours.

Pour surmonter cette difficulté, nous nous basons sur une recherche exhaustive sur un ensemble restreint d'orientations $\{\theta_{i,j} | (j, i) \in \mathcal{I}^{\psi,l} \times \mathcal{I}^{d,l}\}$, où $\mathcal{I}^{\psi,l} \subset \llbracket 0, M_\psi - 1 \rrbracket$ et $\mathcal{I}^{d,l} \subset \llbracket 0, M_d - 1 \rrbracket$ sont des sous-ensemble d'indices de la discrétisation de $SO(3)$ définie dans la section 3.2.1. Les tailles N_d et N_ψ de $\mathcal{I}^{d,l}$ et $\mathcal{I}^{\psi,l}$, respectivement, sont choisis de sorte à ce que $N_d \ll M_d$ et $N_\psi \ll M_\psi$. Soit $E_{i,j}^l = E_l(f, \theta_{i,j}, t_l^*)$ l'énergie associée à l'orientation $\theta_{i,j}$. La solution de (3.13) est approximée par θ_{i^*,j^*} avec

$$i^*, j^* = \arg \min_{i \in \mathcal{I}^{\psi,l}, j \in \mathcal{I}^{d,l}} E_{i,j}^l.$$

La question cruciale est de concevoir une méthode rapide pour construire des sous-ensembles $\mathcal{I}^{\psi,l}$ et $\mathcal{I}^{d,l}$ qui sont adaptés au paysage de l'énergie $E_{i,j}^l$. Définissons $\mathcal{Q}^{l,d}$ et $\mathcal{Q}^{\psi,d}$ les distributions d'échantillonnage à partir desquelles nous créons $\mathcal{I}^{\psi,l}$ et $\mathcal{I}^{d,l}$. Puisque nous recherchons le maximum de vraisemblance dans (3.13), nous pouvons suivre l'approche de [57] et définir $\mathcal{Q}^{l,d}$ et $\mathcal{Q}^{\psi,d}$ à partir des vraisemblances marginales par rapport à d et ψ , respectivement. Ces vraisemblances marginales sont définies en chaque point de la discrétisation par

$$\forall i \in \llbracket 0, M_d - 1 \rrbracket, \pi_i^{l,\psi} = \sum_{j=0}^{M_d-1} p_{i,j}^l, \quad (3.14)$$

$$\forall j \in \llbracket 0, M_\psi - 1 \rrbracket, \pi_j^{l,d} = \sum_{i=0}^{M_\psi-1} p_{i,j}^l, \quad (3.15)$$

où $p_{i,j}^l \propto \exp(-E_{i,j}^l)$ est la vraisemblance associée à l'orientation $\theta_{i,j}$. Cependant, les vraisemblances sont inconnues et nous devons les approximer. L'observation clé est que la variation du volume $\hat{f}$ entre deux itérations consécutives de (3.12) est faible. Par conséquent, la solution de (3.13) à une itération donnée est proche de la solution de l'itération précédente. Nous pouvons donc raisonnablement approximer les vraisemblances dans (3.14) et (3.15) par leurs valeurs aux itérations précédentes (dans ce qui suit, $p_{i,j}^l$ fait référence à sa valeur à l'itération précédente, sans changement de notation pour des raisons de lisibilité).

La difficulté réside dans le fait que la vraisemblance n'est pas évaluée pour tous les éléments des sommes dans (3.14) et (3.15), mais seulement aux indices $\mathcal{I}^{l,d}$ et $\mathcal{I}^{l,\psi}$. Ce problème est résolu en utilisant l'échantillonnage préférentiel [90] pour approximer les vraisemblances marginales (3.14) et (3.15). Étant donné que les indices sont tirés des distributions $\mathcal{Q}^{l,d}$ et $\mathcal{Q}^{\psi,d}$, elles peuvent être utilisées comme distributions d'importance pour créer les approximations suivantes de $\pi^{l,\psi}$ et $\pi^{l,d}$:

$$\forall i \in \mathcal{I}^{l,\psi} \quad \tilde{\pi}_i^{l,\psi} = \sum_{j \in \mathcal{I}^{l,d}} \frac{p_{i,j}^l}{\mathcal{Q}_j^{l,d}}, \quad (3.16)$$

$$\forall j \in \mathcal{I}^{l,d} \quad \tilde{\pi}_j^{l,d} = \sum_{i \in \mathcal{I}^{l,\psi}} \frac{p_{i,j}^l}{\mathcal{Q}_i^{l,\psi}}. \quad (3.17)$$

Pour calculer les vraisemblances approximatives $\tilde{\pi}^{l,\psi}$ et $\tilde{\pi}^{l,d}$ associées à tous les éléments de discrétisation de $\text{SO}(3)$, une méthode d'estimation par noyau est utilisée :

$$\forall i \in \llbracket 0, M_\psi - 1 \rrbracket, \tilde{\pi}_i^{l,\psi} = Z_\psi^{-1} \sum_{k \in \mathcal{I}^{l,\psi}} K_{i,k}^\psi \tilde{\pi}_k^{l,\psi}, \quad (3.18)$$

$$\forall j \in \llbracket 0, M_d - 1 \rrbracket, \tilde{\pi}_j^{l,d} = Z_d^{-1} \sum_{k \in \mathcal{I}^{l,d}} K_{j,k}^d \tilde{\pi}_k^{l,d}, \quad (3.19)$$

où Z_ψ et Z_d sont des constantes de normalisation données par

$$Z_\psi = \sum_{i=1}^{M_\psi} \sum_{k \in \mathcal{I}^{l,\psi}} K_{i,k}^\psi \tilde{\pi}_k^{l,\psi} \text{ et } Z_d = \sum_{j=1}^{M_d} \sum_{k \in \mathcal{I}^{l,d}} K_{j,k}^d \tilde{\pi}_k^{l,d},$$

et K^ψ et K^d sont des noyaux définis par :

$$\forall i, k \in \llbracket 0, M_\psi - 1 \rrbracket, K_{i,k}^\psi = \exp(\beta_\psi \cos(\psi_i - \psi_k)), \quad (3.20)$$

$$\forall j, k \in \llbracket 0, M_d - 1 \rrbracket, \quad K_{j,k}^d = \exp(\beta_d d_j^T d_k), \quad (3.21)$$

où β_ψ et β_d sont des hyper-paramètres.

Durant les premières itérations de (3.12), l'hypothèse de faible variation n'est pas vérifiée, jusqu'à ce que l'on parvienne à une reconstruction grossière stable. Pour en tenir compte, les distributions d'échantillonnage finales sont définies comme des combinaisons linéaires des vraisemblances approximatives avec une distribution uniforme :

$$\mathcal{Q}^{l,d} = \alpha \mathcal{U}^d + (1 - \alpha) \tilde{\pi}^{l,d}, \quad (3.22)$$

$$\mathcal{Q}^{l,\psi} = \alpha \mathcal{U}^\psi + (1 - \alpha) \tilde{\pi}^{l,\psi}, \quad (3.23)$$

où $\mathcal{U}^\psi = \frac{1}{M_\psi}$ et $\mathcal{U}^d = \frac{1}{M_d}$ sont les composantes uniformes, et α est le coefficient ajustant la proportion de la composante uniforme. La composante uniforme permet d'explorer l'ensemble de l'espace de rotation au début du processus de reconstruction, et l'évolution de α restreint progressivement l'espace de recherche autour de solutions plausibles (c'est-à-dire les solutions qui ont donné des vraisemblances élevées à l'itération précédente). Elle est fixée à 1 au début de la descente de gradient et est divisée par α_r entre deux époques de la descente de gradient (3.12).

La figure 3.2 illustre la méthode d'estimation des angles par échantillonnage préférentiel décrite dans cette section.

3.3.4 Optimisation par rapport aux translations

Le problème (3.11c), dédié à l'estimation des translations, doit être résolu à chaque évaluation de l'énergie dans la minimisation du problème dédié à l'estimation des orientations (3.11b), décrite dans la section précédente. Il peut être formulé comme suit :

$$\forall l \in \llbracket 1, N \rrbracket, \quad t_l = \arg \min_{t_l} \|y_l - \mathcal{T}_{t_l}(b_l)\|^2, \quad (3.24)$$

avec $b_l = h * \mathcal{R}_{\theta_l}(f)$.

Or on a :

$$\|y_l - \mathcal{T}_{t_l}(b_l)\|^2 = \int_{x \in \Omega} (y_l(x) - b_l(x - t_l))^2 dx \quad (3.25)$$

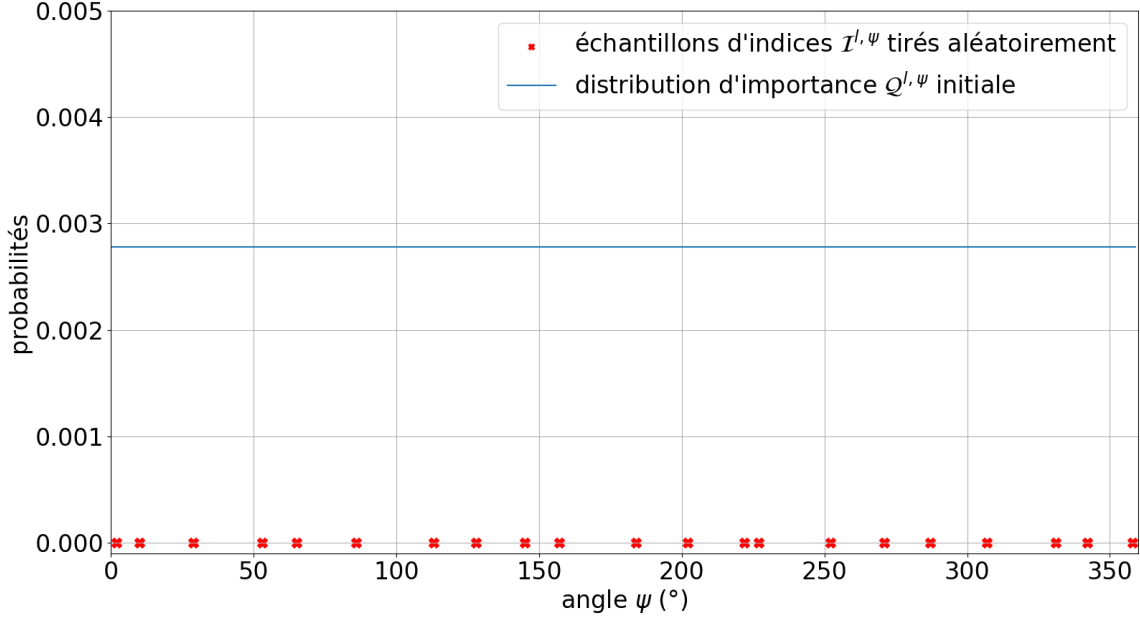
$$= \int_{x \in \Omega} (y_l(x)^2 - 2b_l(x - t_l)y_l(x) + b_l(x - t_l)^2) dx \quad (3.26)$$

$$= \int_{x \in \Omega} y_l(x)^2 dx - 2 \int_{x \in \Omega} b_l(x - t_l)y_l(x) dx + \int_{x \in \Omega} b_l(x - t_l)^2 dx \quad (3.27)$$

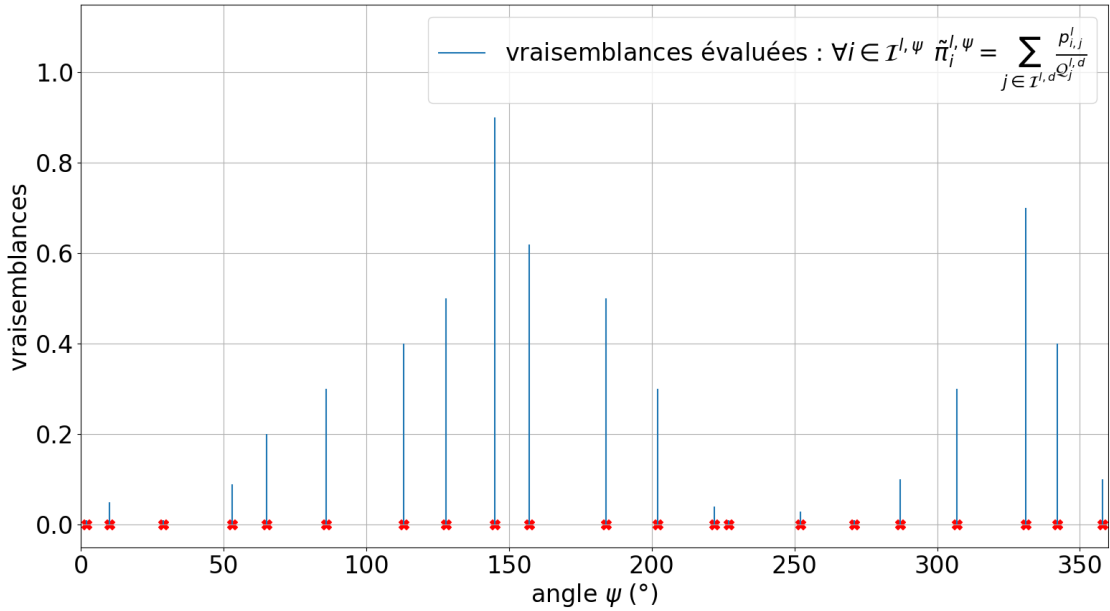
$$= \int_{x \in \Omega} y_l(x)^2 dx - 2 \int_{x \in \Omega} b_l(x - t_l)y_l(x) dx + \int_{x \in \Omega} b_l(x)^2 dx. \quad (3.28)$$

La dernière égalité est vérifiée, si l'on fait l'hypothèse que le support de l'image Ω est non borné, et qu'on pose le changement de variable $x = x - t$ dans la troisième intégrale. On remarque donc que le seul terme qui dépend de t est le second. Par conséquent :

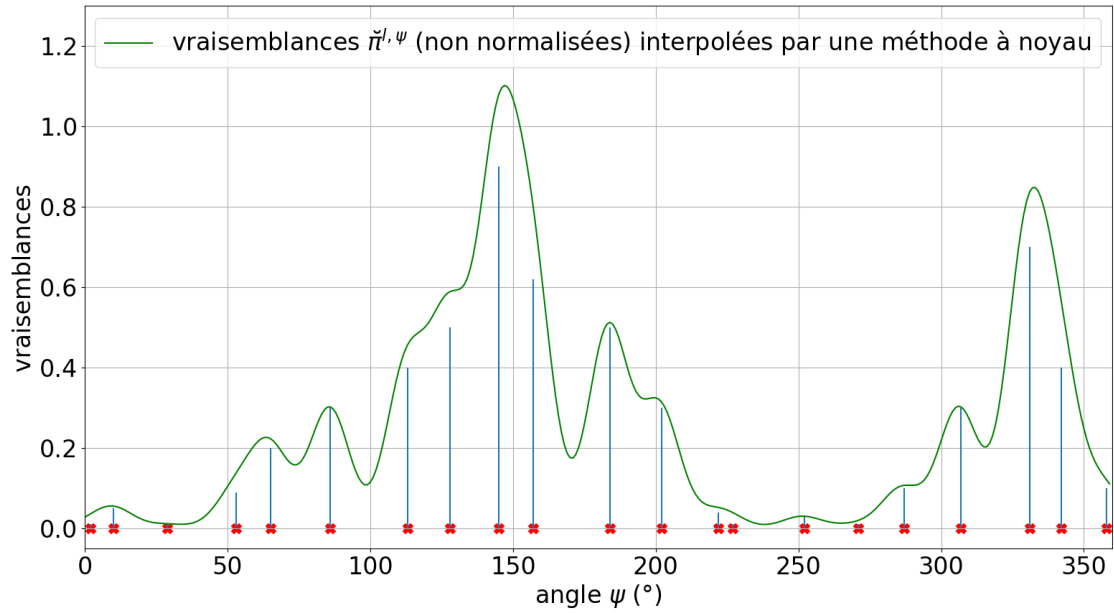
$$\arg \min_{t_l} \|y_l - \mathcal{T}_{t_l}(b_l)\|^2 = \arg \max_{t_l} \int_{x \in \Omega} b_l(x - t_l)y_l(x) dx = \arg \max_{t_l} \text{cor}_{b_l, y_l}(t_l), \quad (3.29)$$



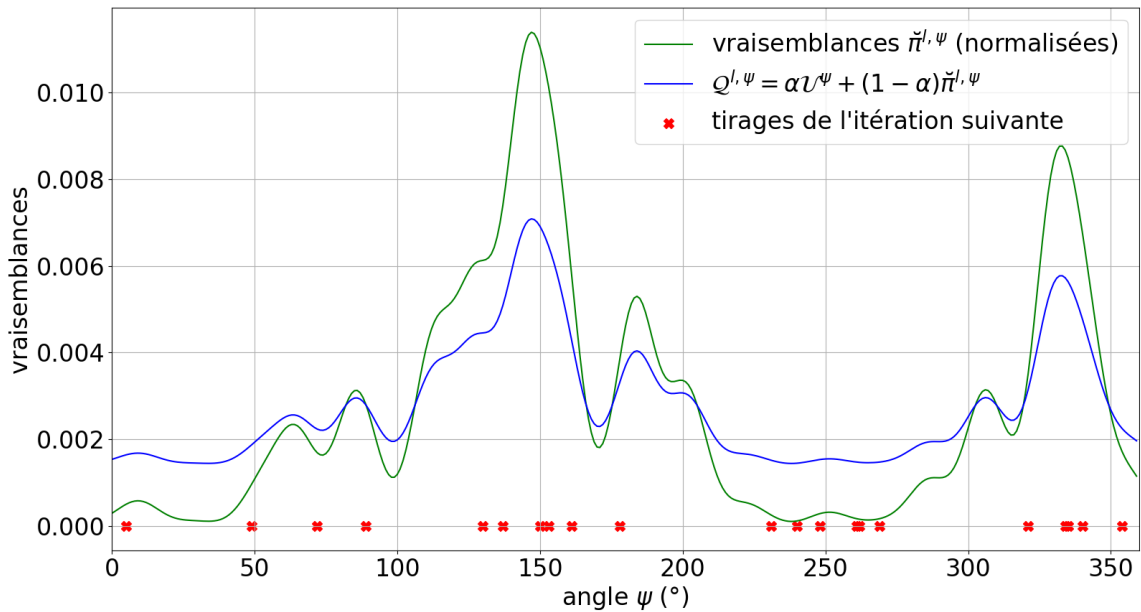
(a) N_ψ angles ψ sont tirées au sort suivant la distribution d'importance initiale $Q^{l,\psi}$, initialisée comme une distribution uniforme.



(b) Les vraisemblances associées à chacun de ces angles sont calculées, il faut pour cela intégrer par rapport à l'autre variable utilisée pour représenter les rotations (la variable d qui représente les axes de rotations). Il n'est pas possible d'intégrer par rapport à l'ensemble des éléments de la discrétisation de d , car seulement N_d axes sont tirés au sort. L'intégrale est donc estimée par la méthode d'échantillonnage préférentiel. L'orientation ψ donnant la vraisemblance la plus élevée est assignée à la vue d'indice l .



(c) La vraisemblance est interpolée en l'ensemble des éléments de la discrétisation de ψ , par une méthode d'estimation à noyau



(d) La vraisemblance est combinée linéairement à une distribution uniforme, pour donner la distribution d'importance de l'itération suivante. Cette nouvelle distribution $Q^{l,\psi}$ est utilisée pour échantillonner les angles de l'itération suivante

FIGURE 3.2 – Illustration de la méthode d'estimation des angles par échantillonnage préférentiel. Le processus illustré par ces quatre figures est effectué pour toutes les vues d'indice $l \in \llbracket 1, N \rrbracket$. Seule l'estimation de ψ est illustrée, mais d est estimée de manière similaire. Les meilleures angles ψ et axes d sont utilisées pour effectuer une itération de la descente de gradient, et le processus est répété jusqu'à convergence

avec cor_{b_l, y_l} la corrélation croisée entre b_l et y_l . L'intérêt de transformer le problème de minimisation de la MSE en une maximisation de la corrélation croisée entre les deux images est que cette dernière peut être calculée très rapidement dans le domaine de Fourier par une multiplication terme à terme. On a en effet :

$$cor_{b_l, y_l} = \mathcal{F}^{-1}(\hat{y}_l \bar{\hat{b}_l}). \quad (3.30)$$

où $\bar{z}$ fait référence au conjugué de z . Une fois la corrélation calculée, la position spatiale du pic de corrélation nous donne la solution du problème.

En pratique, nous utilisons un algorithme standard de corrélation de phase qui considère une corrélation croisée normalisée $corn$ au lieu d'une simple corrélation croisée, afin d'obtenir un pic de corrélation plus net et plus précis [91]. $corn$ se calcule de la manière suivante :

$$corn_{b_l, y_l} = \mathcal{F}^{-1} \left(\frac{\hat{y}_l(\omega) \bar{\hat{b}_l(\omega)}}{|\hat{y}_l(\omega)| |\hat{b}_l(\omega)|} \right) \quad (3.31)$$

Le coût computationnel de cette étape réside essentiellement dans une seule transformée de Fourier inverse. En effet, comme la reconstruction s'effectue dans le domaine de Fourier, la transformée de Fourier de b_l est déjà connue, et l'on peut calculer les transformées de Fourier des vues y_l en amont de la reconstruction. Cette efficacité est cruciale pour la faisabilité computationnelle de l'algorithme de reconstruction global, car cette étape est imbriquée dans les deux problèmes supérieurs (3.11a) et (3.11b).

Les coordonnées du pic de la corrélation croisée normalisée (3.31) donnent une estimation au pixel près du vecteur de translation t_l . Elle est ensuite raffinée en suréchantillonnant d'un facteur 10 la transformée de Fourier discrète dans un voisinage proche de l'estimation. Ce suréchantillonnage local est effectué par la méthode présentée dans [92].

Les étapes détaillées de la reconstruction sont regroupées dans l'Algorithme 1.

3.3.5 Critère d'arrêt

Une problématique importante est de définir le critère d'arrêt de la méthode. L'algorithme doit effectuer suffisamment d'itérations pour assurer une reconstruction correcte, mais il ne doit pas en effectuer trop pour éviter des calculs inutiles.

Nous avons choisi un critère d'arrêt basé sur deux conditions. La première condition $\mathcal{C}_1$ nécessaire à l'arrêt de l'algorithme repose sur les variations de l'énergie lissée. Plus précisément, notons n_1 , n_2 deux itérations quelconques et $\overline{E_{[n_1, n_2]}}$ la moyenne des énergies sur l'intervalle $[n_1, n_2]$, et n l'itération courante. On définit alors $\mathcal{C}_1$ par :

$$\mathcal{C}_1 : \overline{E_{[n-N, n-2N]}} - \overline{E_{[n, n-N]}} < \epsilon_1, \quad (3.32)$$

avec ϵ_1 et N des paramètres fournis par l'utilisateur. En pratique, on fixe $N = 4$ et $\epsilon_1 = 0$. Le lissage de la fonction de coût par moyennage est essentiel à la définition du critère d'arrêt car le caractère stochastique de la méthode induit de légères remontées locales bien avant que l'algorithme ait convergé.

La seconde condition nécessaire $\mathcal{C}_2$ est vérifiée lorsque la proportion de distribution uniforme α dans (3.23) devient inférieure à un certain seuil ϵ_2 . On considère en effet que si α est trop élevée, l'algorithme échantillonnage préférentiel n'a pas encore suffisamment affiné son exploration de l'espace de recherche, il serait donc prématuré de stopper l'algorithme. On fixe en pratique

$\epsilon_2 = 0, 1$. Le coefficient α_r par lequel α est divisé à chaque itération est constant et fixé au début de l'apprentissage. Par conséquent la condition $\mathcal{C}_2$ est vérifiée au bout d'un nombre d'itérations N_{min} , défini comme le plus petit entier vérifiant $\frac{1}{\alpha_r^{N_{min}}} \leq \epsilon_2$, c'est-à-dire :

$$N_{min} = \lceil \frac{-\log(\epsilon_2)}{\log(\alpha_r)} \rceil. \quad (3.33)$$

Les conditions $\mathcal{C}_1$ et $\mathcal{C}_2$ forment ensemble une condition nécessaire et suffisante à l'arrêt de l'algorithme.

3.4 Variante de la méthode : représentation du volume sous forme d'un mélange de gaussiennes

Nous avons développé une variante de la méthode en ne représentant plus le volume par les valeurs de ses voxels mais par une représentation implicite sous la forme d'un mélange (ou combinaison linéaire) de gaussiennes isotropes, ayant toutes le même écart-type σ . Notre idée originale était de pouvoir réduire le temps de calcul, car les rotations de la représentation implicite s'effectuent bien plus rapidement que les rotations du volume représenté de manière explicite. Or cette opération est effectuée à de nombreuses reprises au cours de l'apprentissage. On écrit le volume f à reconstruire de la façon suivante :

$$f = \sum_{p=1}^{N_g} \alpha_p \mathcal{N}(\mu_p, \sigma^2 I), \quad (3.34)$$

avec

$$\mathcal{N}(\mu, \Sigma) : x \mapsto \mathcal{N}(x; \mu, \Sigma) = e^{-\frac{1}{2}(x-\mu)^t \Sigma^{-1}(x-\mu)}, \quad (3.35)$$

la gaussienne de moyenne μ et de matrice de covariance Σ . $\mathcal{N}$ n'est en réalité pas exactement une fonction gaussienne, car elle n'est pas normalisée, nous avons en effet constaté que le facteur de normalisation perturbe l'apprentissage. Si on note $\mathcal{N}^*$ une vraie fonction gaussienne (avec le facteur de normalisation), on a $\mathcal{N} = (2\pi)^{\frac{3}{2}} |\Sigma|^{\frac{1}{2}} \mathcal{N}^*$. D'autre part, nous choisissons d'attribuer le même écart-type σ à toutes les gaussiennes et de le fixer avant le début de l'apprentissage. Nous avons en effet constaté que l'optimisation par rapport à l'écart-type est fortement instable et conduit presque systématiquement à l'obtention de solutions aberrantes. Pour $\Sigma = \sigma^2 I$, on a $\mathcal{N}(\mu, \sigma^2 I)(x) = e^{-\frac{1}{2\sigma^2}(\|x-\mu\|^2)}$ avec $\|\cdot\|$ la norme $\mathcal{L}_2$.

Les paramètres à apprendre sont donc les coefficients des gaussiennes $\alpha = (\alpha_p)_{p \in [1, N_g]}$ et leurs centres $\mu = (\mu_p)_{p \in [1, N_g]}$. On cherche alors à reparamétriser l'énergie en fonction de ces variables. Pour rappel, l'énergie s'écrit :

$$E(f, \Theta, T) = \sum_{l=1}^N \|h * \mathcal{T}_{t_l}(\mathcal{R}_{\theta_l}(f)) - y_l\|^2 \quad (3.36)$$

Dans un premier temps on exprime le volume tourné et translaté en fonction des paramètres du mélange de gaussiennes :

Algorithm 1: Algorithme de reconstruction SHiReVol

Entrée Ensemble de vues $\{y_l\}_{l \in [1, N]}$ à résolution anisotrope, PSF h du microscope

Sortie volume reconstruit f , poses associées à chacune des vues : $(\theta_l^*)_{l \in [1, N]}$ et $(t_l^*)_{l \in [1, N]}$

Paramètres pas de gradient λ , nombre de batches N_b , coefficient α_r par lequel α est divisé à chaque époque (voir (3.23)), β_d et β_ψ (voir (3.20) et (3.21)), nombres d'angles N_ψ et nombre d'axes N_d tirés à chaque itération

- 1: Initialiser $\hat{f}$ aléatoirement, ou comme la moyenne des vues
- 2: Initialiser les distributions d'importance, avec des distributions uniformes $\mathcal{Q}^{l, \psi} = \mathcal{U}$ and $\mathcal{Q}^{l, d} = \mathcal{U}$
- 3: Diviser le jeu de données en batches $B = \{b_n\}_{n \in [1, N_b]}$, avec N_b le nombre de batches
- 4: **while** critère d'arrêt non atteint **do**
- 5: mélanger aléatoirement les batches
- 6: **for** $n = 1$ to N_b **do**
- 7: **for** $l \in b_n$ **do**
- 8: tirer au hasard les indices $\mathcal{I}^{l, \psi}$ et $\mathcal{I}^{l, d}$ selon les distributions $\mathcal{Q}^{l, \psi}$ et $\mathcal{Q}^{l, d}$
- 9: **for** $(i, j) \in \mathcal{I}^{l, \psi} \times \mathcal{I}^{l, d}$ **do**
- 10: $\theta_{i, j} = (d_j, \psi_i)$
- 11: calculer $\hat{h}\mathcal{R}_{\theta_{i, j}}(\hat{f})$
- 12: évaluer $t_l^*(\theta_{i, j})$, en utilisant la corrélation de phase entre $\hat{h}\mathcal{R}_{\theta_{i, j}}(\hat{f})$ et $\hat{y}_l$.
- 13: calculer $E_{i, j}^l = \|\hat{h}\rho_{t_l^*(\theta_{i, j})}\mathcal{R}_{\theta_{i, j}}(\hat{f}) - \hat{y}_l\|_2^2$
- 14: **end for**
- 15: $\theta_l^* = (d_{j^*}, \psi_{i^*})$ avec $i^*, j^* = \arg \min_{i \in \mathcal{I}^{l, \psi}, j \in \mathcal{I}^{l, d}} E_{i, j}^l$
- 16: $\forall (i, j) \in \mathcal{I}^{l, \psi} \times \mathcal{I}^{l, d}$, calculer $p_{i, j}^l = \exp(-E_{i, j}^l)$.
- 17: Calculer les distributions marginales
- 18: $\forall i \in \mathcal{I}^{l, \psi} \quad \tilde{\pi}_i^{l, \psi} = \sum_{j \in \mathcal{I}^{l, d}} \frac{p_{i, j}^l}{\mathcal{Q}_j^{l, d}}$
- 19: $\forall j \in \mathcal{I}^{l, d} \quad \tilde{\pi}_j^{l, d} = \sum_{i \in \mathcal{I}^{l, \psi}} \frac{p_{i, j}^l}{\mathcal{Q}_i^{l, \psi}}$
- 20: Mettre à jour les distributions de probabilités
- 21: $\mathcal{Q}_j^{l, d} = \alpha \mathcal{U}_j^d + (1 - \alpha) Z_d^{-1} \sum_{k \in \mathcal{I}^{l, d}} K_{j, k}^d \tilde{\pi}_k^{l, d}$
- 22: $\mathcal{Q}_i^{l, \psi} = \alpha \mathcal{U}_i^\psi + (1 - \alpha) Z_\psi^{-1} \sum_{k \in \mathcal{I}^{l, \psi}} K_{i, k}^\psi \tilde{\pi}_k^{l, \psi}$
- 23: $\alpha = \alpha / \alpha_r$
- 24: **end for**
- 25: $\hat{f} = \hat{f} - \lambda \sum_{l \in b_n} \nabla_{\hat{f}} E_l(f, \theta_l^*, t_l^*(\theta_l^*))$
- 26: **end for**
- 27: **end while**
- 28: $\hat{f}^* = \hat{f}$
- 29: **return** $\mathcal{F}^{-1}(\hat{f}^*), (\theta_l^*)_{l \in [1, N]}, (t_l^*)_{l \in [1, N]}$

$$\mathcal{T}_{t_l}(\mathcal{R}_{\theta_l}(f))(x) = \mathcal{T}_{t_l}(f(R_{\theta_l}^t x)) \quad (3.37)$$

$$= f(R_{\theta_l}^t(x - t_l)) \quad (3.38)$$

$$= \sum_{p=1}^{N_g} \alpha_p e^{-\frac{1}{2\sigma^2} \|R_{\theta_l}^t(x - t_l) - \mu_p\|^2}. \quad (3.39)$$

Or on sait qu'une matrice de rotation R quelconque vérifie la propriété suivante : $\forall x \in \mathbb{R}^3, \|Rx\| = \|x\|$. Donc en multipliant par R_{θ_l} dans la norme on obtient :

$$\mathcal{T}_{t_l}(\mathcal{R}_{\theta_l}(f))(x) = \sum_{p=1}^{N_g} \alpha_p e^{-\frac{1}{2\sigma^2} \|x - (R_{\theta_l}\mu_p + t_l)\|^2} \quad (3.40)$$

$$= \sum_{p=1}^{N_g} \alpha_p \mathcal{N}(x; R_{\theta_l}\mu_p + t_l, \sigma^2 I). \quad (3.41)$$

Nous venons de démontrer qu'appliquer une rotation de matrice de rotation R_{θ_l} puis une translation de vecteur t_l à un mélange de gaussiennes revient à appliquer ces mêmes opérations à chacun des centres de ce mélange.

Ensuite, on applique la convolution par la PSF h au volume tourné et translaté pour obtenir $h * \mathcal{T}_{t_l}(\mathcal{R}_{\theta_l}(f))$. Tout d'abord, la convolution est un opérateur linéaire, la convolution de la somme des gaussiennes est donc la somme des gaussiennes convoluées. D'autre part on utilise le résultat énonçant que la convolution d'une gaussienne (normalisée) par une gaussienne est une gaussienne obtenue en sommant les moyennes et les matrices de covariance, c'est-à-dire :

$$\forall \mu_1, \mu_2 \in \mathbb{R}^3, \forall \Sigma_1, \Sigma_2 \in \mathcal{S}^3 \quad \mathcal{N}^*(\mu_1, \Sigma_1) * \mathcal{N}^*(\mu_2, \Sigma_2) = \mathcal{N}^*(\mu_1 + \mu_2, \Sigma_1 + \Sigma_2) \quad (3.42)$$

Donc, pour la gaussienne p on a :

$$\mathcal{N}(R_{\theta_l}\mu_p + t_l, \sigma^2 I) * h = (2\pi)^{\frac{3}{2}} \sigma \mathcal{N}^*(R_{\theta_l}\mu_p + t_l, \sigma^2 I) * \mathcal{N}^*(0, \Sigma_h) \quad (3.43)$$

$$= (2\pi)^{\frac{3}{2}} \sigma \mathcal{N}^*(R_{\theta_l}\mu_p + t_l, \sigma^2 I + \Sigma_h) \quad (3.44)$$

$$= \frac{\sigma}{|\Sigma_h + \sigma^2 I|^{\frac{1}{2}}} \mathcal{N}(R_{\theta_l}\mu_p + t_l, \sigma^2 I + \Sigma_h). \quad (3.45)$$

On notera $C = \frac{\sigma}{|\Sigma_h + \sigma^2 I|^{\frac{1}{2}}}$ pour ne pas alourdir les notations. Le volume tourné translaté et convolué s'écrit donc :

$$h * \mathcal{T}_{t_l}(\mathcal{R}_{\theta_l}(f)) = C \sum_{p=1}^N \alpha_p \mathcal{N}(R_{\theta_l}\mu_p + t_l, \sigma^2 I + \Sigma_h). \quad (3.46)$$

La convolution par la PSF revient donc à ajouter la matrice de covariance de la PSF aux matrices de covariances de chacune des gaussiennes représentant le volume (à un facteur multiplicatif C près).

On peut alors réécrire l'énergie E :

$$E(\gamma, \Theta, T) = \sum_{l=1}^N \|y_l - C \sum_{p=1}^{N_g} \alpha_p \mathcal{N}(R_{\theta_l} \mu_p + t_l, \sigma^2 I + \Sigma_h)\|^2, \quad (3.47)$$

où $\gamma = (\alpha, \mu)$ sont les paramètres du mélange de gaussiennes. La dépendance de E en f a été changé en une dépendance en γ pour indiquer que l'on ne cherche plus directement f mais les paramètres du mélange qui le représentent. L'algorithme 2 expose, étapes par étapes, la procédure d'optimisation qui utilise la représentation sous forme de mélange de gaussiennes. Les différences par rapport à la méthode originale sont écrites en rouge. Par ailleurs, les gradients de l'énergie par rapport à γ sont présentés en Annexe 6.2.

3.5 Deuxième variante : représentation des vues sous forme d'un mélange de gaussiennes

3.5.1 Principe

L'avantage de la représentation du volume à reconstruire sous forme d'un mélange de gaussiennes est de permettre d'effectuer les rotations, translations et convolutions rapidement, par simple application d'opérateurs linéaires aux centres et matrices de covariances des gaussiennes. Cependant, la majorité du temps de calcul est dédié à l'évaluation du mélange de gaussiennes aux points de la grille, c'est-à-dire le calcul de $h * \mathcal{T}_{t_l}(\mathcal{R}_{\theta_l}(f))(x)$ pour tout $x \in \Omega$. Ce calcul est nécessaire pour pouvoir ensuite comparer voxel à voxel le volume tourné et convolué avec chacune des vues y_l dans l'évaluation de l'énergie à chaque tirage d'angle. Nous avons constaté expérimentalement que ce temps de calcul est du même ordre que le temps de rotation d'une image. Cela réduit l'intérêt de l'utilisation de cette représentation, que nous avons initialement choisie pour limiter le coût computationnel.

Pour éviter de devoir évaluer le mélange de gaussiennes sur la grille, nous avons eu l'idée de représenter également les vues y_l par un mélange de gaussiennes et de calculer la distance entre y_l et $h * \mathcal{T}_{t_l}(\mathcal{R}_{\theta_l}(f))$ en utilisant uniquement les paramètres des mélanges qui les représentent : ce n'est plus une distance entre images qui est calculée mais une distance entre distributions. On écrira $\tilde{y}_l$ la représentation en mélange de gaussiennes de y_l . On notera ses coefficients et centres respectivement $(\beta_q^l)_{q \in \llbracket 1, M_l \rrbracket}$ et $(m_q^l)_{q \in \llbracket 1, M_l \rrbracket}$ avec M_l le nombre de gaussiennes pour représenter la vue d'indice l . Les vues ayant une résolution anisotrope, nous avons choisi de les représenter par des gaussiennes également anisotropes de matrice de covariance $\Delta = \sigma^2 I + \Sigma_h$.

$$\tilde{y}_l = \sum_{q=1}^{M_l} \beta_q^l \mathcal{N}(m_q^l, \Delta) \quad (3.48)$$

Nous donnons en annexe 6.3 le calcul de la distance ℓ_2 entre $\tilde{y}_l$ et $h * \mathcal{T}_{t_l}(\mathcal{R}_{\theta_l}(f))$ par comparaison de leurs distributions - c'est-à-dire en utilisant seulement les paramètres de leur représentation implicite - ainsi que le calcul des gradients de la fonction de coût associée.

Algorithm 2: Algorithme de reconstruction ShiReGmm, **représentation GMM du volume**

Input Ensemble de vues $\{y_l\}_{l \in \llbracket 1, N \rrbracket}$ à résolution anisotrope, PSF h du microscope

Output volume reconstruit f^* , poses associées à chacune des vues : $(\theta_l^*)_{l \in \llbracket 1, N \rrbracket}$ et $(t_l^*)_{l \in \llbracket 1, N \rrbracket}$

Paramètres Même paramètres que ceux de la méthode voxélique SHiReVol avec en plus : nombre de gaussiennes N_g , écart type des gaussiennes σ .

- 1: Initialiser les centres $\{\mu_p\}_{p \in \llbracket 1, N_g \rrbracket}$ et coefficients $\{\alpha_p\}_{p \in \llbracket 1, N_g \rrbracket}$ des gaussiennes aléatoirement
 - 2: Initialiser les distributions d'importance, avec des distributions uniformes $\mathcal{Q}^{l, \psi} = \mathcal{U}$ and $\mathcal{Q}^{l, d} = \mathcal{U}$
 - 3: Diviser le jeu de données en batchs $B = \{b_n\}_{n \in \llbracket 1, N_b \rrbracket}$, avec N_b le nombre de batchs
 - 4: **while** critère d'arrêt non atteint **do**
 - 5: mélanger aléatoirement les batchs
 - 6: **for** $n = 1$ to N_b **do**
 - 7: **for** $l \in b_n$ **do**
 - 8: tirer au hasard les indices $\mathcal{I}^{l, \psi}$ et $\mathcal{I}^{l, d}$ selon les distributions $\mathcal{Q}^{l, \psi}$ et $\mathcal{Q}^{l, d}$
 - 9: **for** $(i, j) \in \mathcal{I}^{l, \psi} \times \mathcal{I}^{l, d}$ **do**
 - 10: $\theta_{i, j} = (d_j, \psi_i)$
 - 11: tourner la représentation implicite de $\theta_{i, j}$ en calculant $R_{\theta_{i, j}} \mu_p$ pour tout $p \in \llbracket 1, N_b \rrbracket$
 - 12: calculer $h * \mathcal{R}_{\theta_{i, j}}(f)(x) = C \sum_{p=1}^{N_g} \alpha_p \mathcal{N}(x; R_{\theta_{i, j}} \mu_p, \sigma^2 I + \Sigma_h)$ pour tout $x \in \Omega$
 - 13: évaluer $t_l^*(\theta_{i, j})$, en utilisant la corrélation de phase entre $h * \mathcal{R}_{\theta_{i, j}}(f)$ et y_l
 - 14: calculer $h * \mathcal{T}_{t_l^*(\theta_{i, j})}(\mathcal{R}_{\theta_{i, j}}(f))(x) = C \sum_{p=1}^{N_g} \alpha_p \mathcal{N}(x; R_{\theta_{i, j}} \mu_p + t_l^*(\theta_{i, j}), \sigma^2 I + \Sigma_h)$ pour tout $x \in \Omega$
 - 15: calculer $E_{i, j}^l = \|y_l - h * \mathcal{T}_{t_l^*(\theta_{i, j})}(\mathcal{R}_{\theta_{i, j}}(f))\|^2$
 - 16: **end for**
 - 17: $\theta_l^* = (d_{j^*}, \psi_{i^*})$ avec $i^*, j^* = \arg \min_{i \in \mathcal{I}^{l, \psi}, j \in \mathcal{I}^{l, d}} E_{i, j}^l$
 - 18: $\forall (i, j) \in \mathcal{I}^{l, \psi} \times \mathcal{I}^{l, d}$, calculer $p_{i, j}^l = \exp(-E_{i, j}^l)$.
 - 19: Calculer les distributions marginales
 - 20: $\forall i \in \mathcal{I}^{l, \psi} \quad \tilde{\pi}_i^{l, \psi} = \sum_{j \in \mathcal{I}^{l, d}} \frac{p_{i, j}^l}{\mathcal{Q}_j^{l, d}}$
 - 21: $\forall j \in \mathcal{I}^{l, d} \quad \tilde{\pi}_j^{l, d} = \sum_{i \in \mathcal{I}^{l, \psi}} \frac{p_{i, j}^l}{\mathcal{Q}_i^{l, \psi}}$
 - 22: Mettre à jour les distributions de probabilités
 - 23: $\mathcal{Q}_j^{l, d} = \alpha \mathcal{U}_j^d + (1 - \alpha) Z_d^{-1} \sum_{k \in \mathcal{I}^{l, d}} K_{j, k}^d \tilde{\pi}_k^{l, d}$
 - 24: $\mathcal{Q}_i^{l, \psi} = \alpha \mathcal{U}_i^\psi + (1 - \alpha) Z_\psi^{-1} \sum_{k \in \mathcal{I}^{l, \psi}} K_{i, k}^\psi \tilde{\pi}_k^{l, \psi}$
 - 25: $\alpha = \alpha / \alpha_r$
 - 26: **end for**
 - 27: $\mu_p = \mu_p - \mu \sum_{l \in b_n} \nabla_{\mu_p} E_l(\gamma, \theta_l^*, t_l^*(\theta_l^*))$ pour tout $p \in \llbracket 1, N_g \rrbracket$
 - 28: $\alpha_p = \alpha_p - \lambda \sum_{l \in b_n} \nabla_{\alpha_p} E_l(\gamma, \theta_l^*, t_l^*(\theta_l^*))$ pour tout $p \in \llbracket 1, N_g \rrbracket$
 - 29: **end for**
 - 30: **end while**
 - 31: $\mu_p^* = \mu_p$ pour tout $p \in \llbracket 1, N_g \rrbracket$
 - 32: $\alpha_p^* = \alpha_p$ pour tout $p \in \llbracket 1, N_g \rrbracket$
 - 33: calculer $f^* = \sum_{p=1}^{N_g} \alpha_p^* \mathcal{N}(\mu_p^*, \sigma^2 I)$
 - 34: **return** $f^*, (\theta_l^*)_{l \in \llbracket 1, N \rrbracket}, (t_l^*)_{l \in \llbracket 1, N \rrbracket}$
-

3.5.2 Méthode pour représenter les vues sous une forme de mélange de gaussiennes

Pour pouvoir représenter chaque vue y_l sous la forme d'un mélange de gaussiennes, c'est-à-dire trouver les paramètres M_l , β_q^l et m_q^l , nous avons développé une méthode qui combine une approche heuristique et une descente de gradient. La première étape de l'heuristique consiste à estimer le nombre de gaussiennes M_l . Soit $S = \sum_{x \in \Omega} \mathcal{N}(x; m_q^l, \Delta)$ la somme des niveaux de gris d'une gaussienne, et $S_l = \sum_{x \in \Omega} y_l(x)$ la somme des niveaux de gris de la vue d'indice l . On pose alors :

$$M_l = \lceil \frac{2S_l}{S} \rceil. \quad (3.49)$$

On veut en effet que la somme des niveaux de gris de l'ensemble des gaussiennes soit égale à la somme des niveaux de gris de l'image. De plus, on suppose que la valeur moyenne des coefficients β_q^l est de $\frac{1}{2}$, d'où le facteur 2. Cette hypothèse est un peu arbitraire mais reste cohérente car les vues sont normalisées entre 0 et 1. (i.e. les valeurs des pixels vont de 0 à 1). Par ailleurs, $S = (2\pi)^{\frac{3}{2}} |\Delta|^{\frac{1}{2}} \sum_{x \in \Omega} \mathcal{N}^*(x) \approx (2\pi)^{\frac{3}{2}} |\Delta|^{\frac{1}{2}} 1$. On a donc :

$$M_l = \lceil \frac{2S_l}{(2\pi)^{\frac{3}{2}} |\Delta|^{\frac{1}{2}}} \rceil \quad (3.50)$$

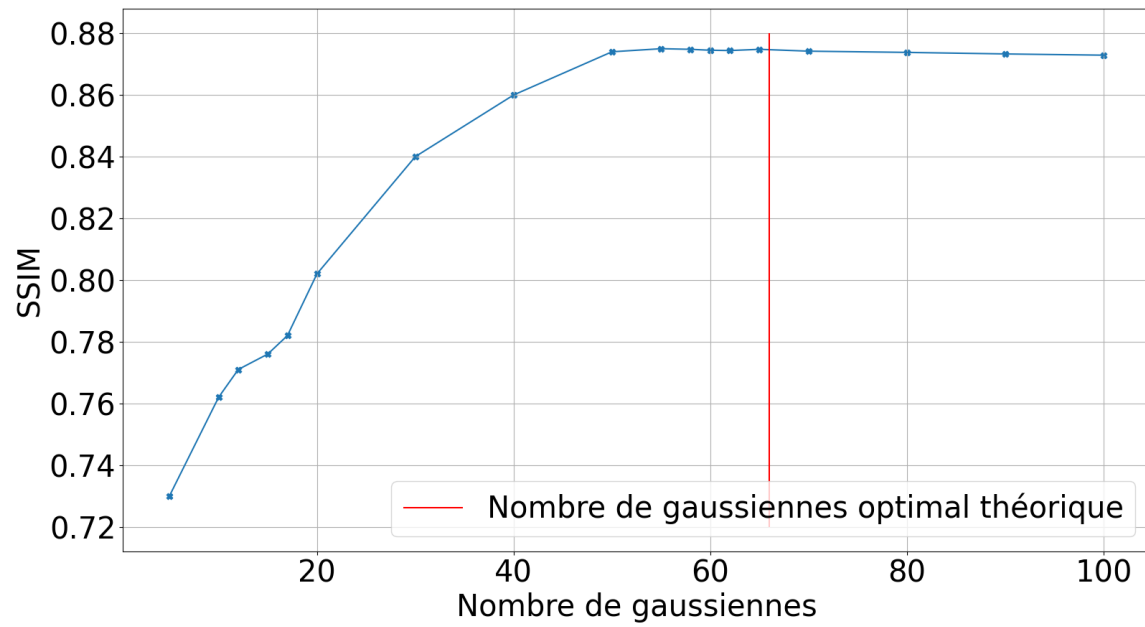
Une fois M_l estimé, les gaussiennes sont positionnées une à une aux emplacements les plus lumineux de l'image. Plus précisément, la première gaussienne est positionnée au pixel d'intensité la plus élevée $m_1^l = \arg \max_{x \in \Omega} y_l(x)$. Par la suite, la q^{ieme} gaussienne est placée au pixel d'intensité la plus élevée de l'image y_l à laquelle on a retiré la mixture composée des gaussiennes déjà placées. c'est-à-dire :

$$m_q^l = \arg \max_{x \in \Omega} (y_l(x) - \sum_{j=1}^{q-1} \beta_j^l \mathcal{N}(x; m_j^l, \Delta)) \quad (3.51)$$

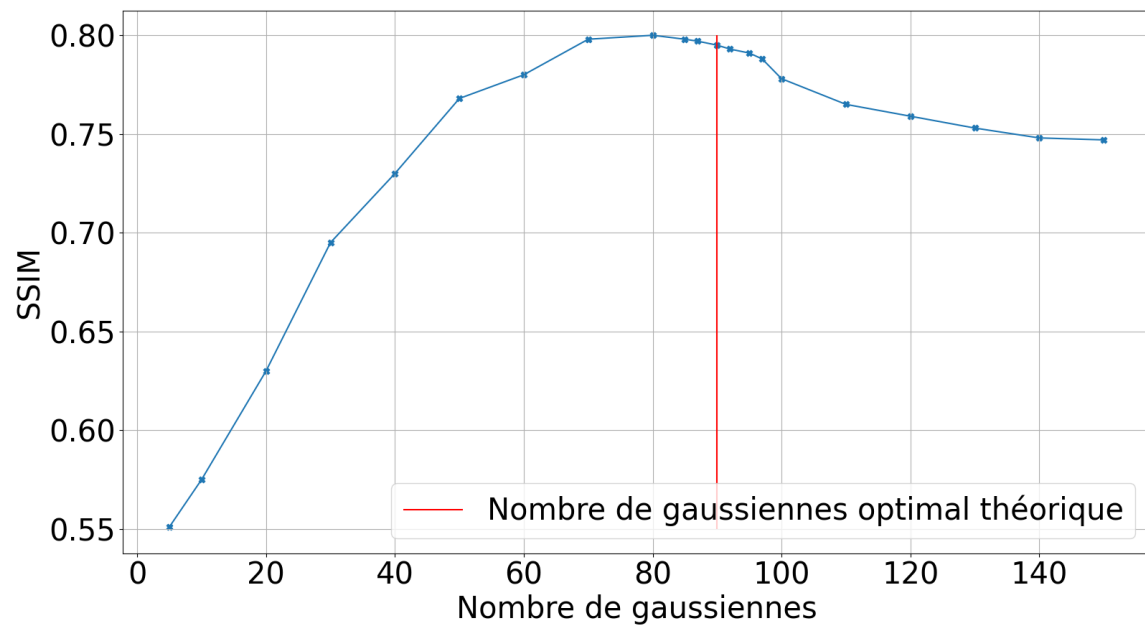
Le coefficient β_q^l est obtenu comme l'intensité de l'image au point où la gaussienne est placée : $\beta_q^l = y_l(m_q^l)$. Une fois les centres et coefficients initialisés ainsi, une descente de gradient sur ces paramètres permet de raffiner la représentation.

Pour évaluer la pertinence d'estimer le nombre de gaussiennes pour représenter une vue par $M_l = \lceil \frac{2S_l}{S} \rceil$, nous avons appliqué cette méthode heuristique à des vues issues de deux objets différentes (Récepteur AMPA et HIV-Vaccin, qu'on montrera plus loin dans la partie résultats 3.6.1.1). Pour différents nombres de gaussiennes, on applique la méthode heuristique sur 25 vues et on note l'index de similarité de structure (structural similarity (SSIM)) (voir définition du SSIM en partie 3.6.1.2) moyen entre les mixtures de gaussiennes et les vues. La figure 3.3 illustre les résultats obtenus. Le nombre optimal théorique de gaussiennes $M_l = \lceil \frac{2S_l}{S} \rceil$ est représenté par une barre verticale rouge. On constate, dans le cas de Récepteur AMPA, que le SSIM n'augmente plus à partir d'environ 55 gaussiennes ce qui est proche du nombre idéal estimé. Pour HIV-Vaccin, on remarque même une diminution du SSIM pour des nombres de gaussiennes supérieurs à M_l .

1. on utilise $\approx$ plutôt que $=$ car la somme des valeurs de $\mathcal{N}^*$ s'effectue sur un ensemble fini discret



(a) Récepteurs AMPA



(b) HIV-Vaccine

FIGURE 3.3 – SSIM entre une vue et le mélange de gaussiennes estimé par la méthode heuristique. Le nombre de gaussiennes optimal théorique $M_l = \lceil \frac{2S_l}{S} \rceil$ est représenté par une barre verticale rouge

3.6 Résultats

3.6.1 Résultats sur données simulées

Pour valider nos méthodes, nous les testons dans un premier temps sur des données simulées. Dans cette section, nous présentons d’abord comment sont générées les données à partir d’un modèle vérité terrain. Ensuite nous exposons les différentes métriques que nous utilisons pour comparer la reconstruction obtenue sur le jeu de données simulées au volume vérité terrain dont elles sont issues. Puis, nous analysons la sensibilité de nos méthodes aux différents hyper-paramètres mis en jeu. Enfin, nous comparons les méthodes entre elles ainsi qu’avec celles de l’état de l’art, et évaluons la robustesse aux paramètres utilisés pour générer les données (bruit et degré de labellisation).

3.6.1.1 Génération des données

Nous utilisons des modèles de particules obtenus par reconstruction en cryo-ME, et disponibles sur la base de données EMDDataRessource². Nous avons retenu les particules Récepteurs AMPA [93], NLR-Résistosome [94], Clathrine [95], HIV-Vaccine [96]. Nous considérons ces modèles comme vérité terrain et simulons le processus d’acquisition convolutionnel (3.4) pour obtenir un ensemble de $N = 20$ vues, avec orientations aléatoires tirées uniformément selon la discrétisation définie en partie 3.2. Les translations sont quant à elles tirées uniformément selon une loi gaussienne centrée en 0, ce qui est réaliste car dans le cas de données réelles, les vues sont centrées en amont de la reconstruction de sorte à ce que leurs barycentres coïncident avec le centre de l’image. Nous utilisons une PSF gaussienne h , ayant pour matrice de covariance Σ_h et que l’on définit par :

$$\Sigma_h = \begin{pmatrix} \sigma_{xy}^2 & 0 & 0 \\ 0 & \sigma_{xy}^2 & 0 \\ 0 & 0 & \sigma_z^2 \end{pmatrix}, \quad (3.52)$$

où σ_z et σ_{xy} sont les écart-types respectivement selon la direction du microscope et perpendiculaire à cette dernière. Pour imiter un flou anisotrope, nous choisissons $\sigma_z = 5$ et $\sigma_{xy} = 1$. Nous utilisons ce modèle particulier de PSF car il correspond à une PSF réaliste, mais notre méthode peut-être appliquée à n’importe quel type de PSF. Les volumes vérités terrain sont redimensionnés en une taille de $50 \times 50 \times 50$ avant la génération des vues. Nous appliquons un bruit additif gaussien, avec un rapport signal à bruit de 5. Les variations d’intensité parasites dues à un degré de labellisation non uniforme sont simulées en soustrayant de petits spots gaussiens (des fonctions gaussiennes 3D et isotropes) à des emplacements aléatoires dans l’image. L’écart-type des spots gaussiens est uniformément échantillonné entre 0 et 1 (les images sont normalisées entre 0 et 1) et le nombre de spots à soustraire est fixé à 120. Sur la figure 3.4, nous montrons les volumes vérités terrain et des exemples de vues simulées, alignées à la vérité terrain. Pour les représenter, nous utilisons deux types de visualisations complémentaires : trois vues orthogonales (plans XY, YZ et XZ), et une visualisation 3D de l’isosurface correspondant à un certain seuil (avec Chimera³). L’anisotropie de résolution peut être visualisée sur les cartes d’intensité XZ et YZ.

2. <https://www.emdataresource.org/>

3. <https://www.cgl.ucsf.edu/chimera/>

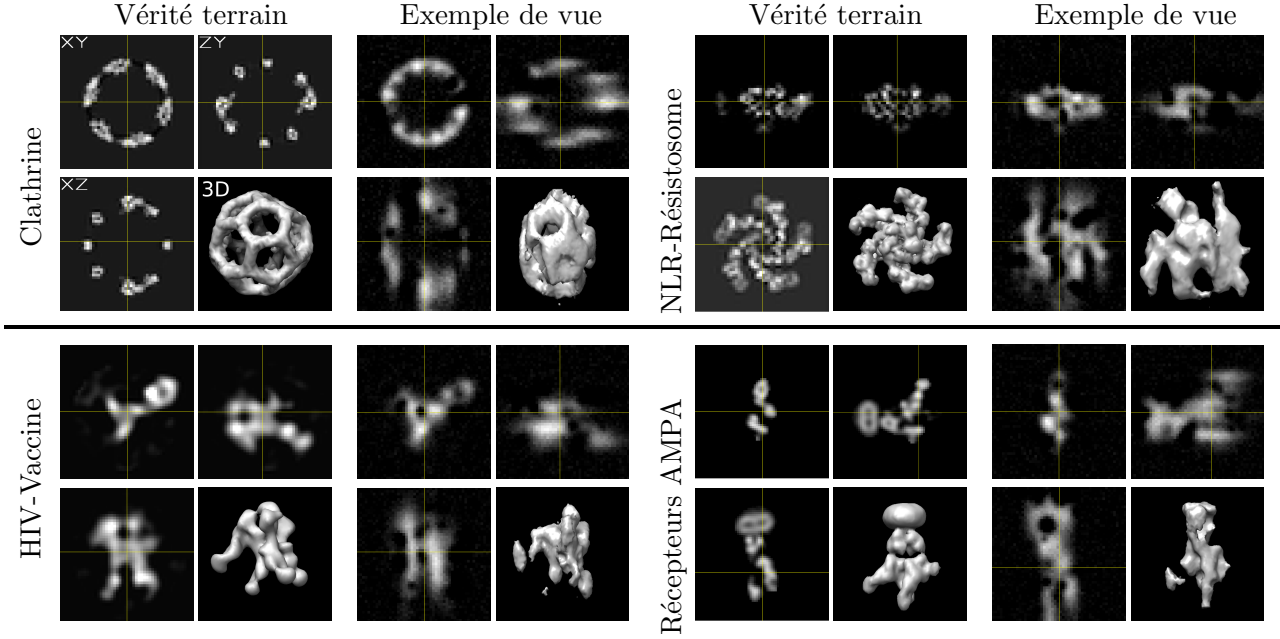


FIGURE 3.4 – Exemples de données simulées. Pour chaque volume nous montrons 3 vues orthogonales et une représentation iso-surfacique 3D

3.6.1.2 Métriques

Nous utilisons les données ainsi simulées comme entrées de nos trois méthodes pour obtenir des résultats de reconstruction que nous comparons à la vérité terrain, en utilisant trois différentes métriques que nous présentons ci-dessous.

3.6.1.2.1 SSIM Pour quantifier la qualité visuelle de la reconstruction, nous calculons l'index de similarité de structure [97](SSIM) entre la reconstruction f^* et la vérité terrain f_{gt} . Il se définit par :

$$SSIM(f_{gt}, f^*) = l(f_{gt}, f^*) \cdot c(f_{gt}, f^*) \cdot s(f_{gt}, f^*), \quad (3.53)$$

avec $l(f_{gt}, f^*)$, $c(f_{gt}, f^*)$ et $s(f_{gt}, f^*)$ les composantes respectivement de *luminance*, de *contrast* et de *structure*.

Ils sont définis par

$$l(f_{gt}, f^*) = \frac{2\mu_{f_{gt}}\mu_{f^*} + C_1}{\mu_{f_{gt}}^2 + \mu_{f^*}^2 + C_1}, \quad (3.54)$$

$$c(A, B) = \frac{2\sigma_{f_{gt}}\sigma_{f^*} + C_2}{\sigma_{f_{gt}}^2 + \sigma_{f^*}^2 + C_2} \quad (3.55)$$

et

$$s(f_{gt}, f^*) = \frac{\sigma_{f_{gt}, f^*} + C_3}{\sigma_{f_{gt}}\sigma_{f^*} + C_3}, \quad (3.56)$$

avec $\mu_{f_{gt}}$ et μ_{f^*} les moyennes des images f_{gt} et f^* , $\sigma_{f_{gt}}^2$ et $\sigma_{f^*}^2$ les variances, σ_{f_{gt}, f^*} la covariance entre les deux images, et C_1 , C_2 , C_3 des constantes proches de 0. Le SSIM peut également être défini de manière local en chaque point de l'image et en calculant les grandeurs statistiques par pondération

avec un noyau gaussien centré au point d'intérêt. Néanmoins, nous considérons dans cette thèse uniquement la définition globale. Le SSIM prend ses valeurs entre 0 et 1 et est d'autant plus proche de 1 que les moyennes et variances des deux images sont similaires, et que leur corrélation est proche de 1. Si le SSIM peut nous donner une indication sur la ressemblance de deux images, il admet cependant des limites, car il n'est pas toujours en accord avec les propriétés du système visuel humain [98].

3.6.1.2.2 Corrélation des anneaux de Fourier Pour mesurer la résolution de la reconstruction, nous suivons la pratique usuelle en microscopie électronique et calculons la Corrélation des Anneaux de Fourier [99] (FSC pour "Fourier Shell Correlation"). La FSC a initialement été développée pour évaluer la résolution d'une reconstruction lorsqu'on ne connaît pas la vérité terrain. L'idée générale est d'effectuer deux reconstructions indépendantes sur deux sous-ensembles disjoints du jeu de données et d'évaluer la fréquence maximale pour laquelle la corrélation entre les deux reconstructions est supérieure à un certain seuil. Néanmoins, dans notre cas, nous n'avons pas à disposition suffisamment de données issues d'acquisitions réelles pour que cette procédure donne des résultats fiables. Nous avons donc choisi d'utiliser la FSC uniquement sur des données synthétiques, pour lesquelles nous avons accès à une vérité terrain. La FSC entre la reconstruction f^* et la vérité terrain f_{gt} est une fonction $FSC : \mathbb{R} \mapsto \mathbb{R}$ définie par :

$$FSC(\rho) = \frac{\sum_{\nu \in \mathcal{S}_\rho} \mathcal{F}(f^*)(\nu) \overline{\mathcal{F}(f_{gt})(\nu)}}{\sqrt{\sum_{\nu \in \mathcal{S}_\rho} \mathcal{F}(f^*)^2(\nu) \sum_{\nu \in \mathcal{S}_\rho} \mathcal{F}(f_{gt})^2(\nu)}}, \quad (3.57)$$

où $\mathcal{S}_\rho$ désigne la sphère de rayon ρ . Ainsi, la FSC mesure la corrélation croisée normalisée entre des sphères de f^* et f_{gt} à différentes amplitudes de fréquences ρ dans le domaine de Fourier. En pratique, lorsque deux objets sont similaires sans être parfaitement identiques, la FSC prend des valeurs élevées (proches de 1) dans les basses fréquences, et diminue progressivement pour atteindre 0 dans les hautes fréquences. Les 2 objets comparés seront d'autant plus similaires que la FSC décroît lentement, car une corrélation élevée dans les hautes fréquences témoigne d'une similarité des détails. La résolution de f^* peut alors être déterminée à partir de la FSC en déterminant la fréquence limite à partir de laquelle la corrélation devient inférieure à un certain seuil (nous considérons le seuil standard de 0,143). La figure 3.5 illustre un exemple de FSC, et la manière dont est déterminée la résolution à partir de cette courbe. Les valeurs de résolution sont exprimées comme l'inverse de la taille du pixel. Dans ce qui suit, nous utiliserons le terme FSC pour nous référer à la résolution déterminée à partir de la courbe de la FSC.

3.6.1.2.3 Corrélation Conique des Anneaux de Fourier Dans notre cas, la résolution des données d'entrée étant anisotrope, il est important d'évaluer la capacité de notre méthode à la rendre isotrope. Pour quantifier le degré d'isotropie de résolution, nous utilisons une variante de la FSC appelée FSC conique (conical fourier shell correlation (cFSC)) [100]. Le principe de la cFSC est de calculer une courbe FSC localement pour chaque direction.

Plus formellement, un frequel $\nu \in \mathbb{R}^3$ est représenté en coordonnées sphériques par un rayon $\rho(\nu)$ et deux angles $\psi(\nu) = (\phi_1(\nu), \phi_2(\nu))$. L'espace de Fourier est discrétisé par un ensemble de rayons $P = (\rho_i)_{i \in \mathbb{N}}$ et un ensemble d'orientations, $\Psi = (\psi_j)_{j \in \mathbb{N}}$. La cFSC sur l'espace de Fourier discrétisé est alors définie par :

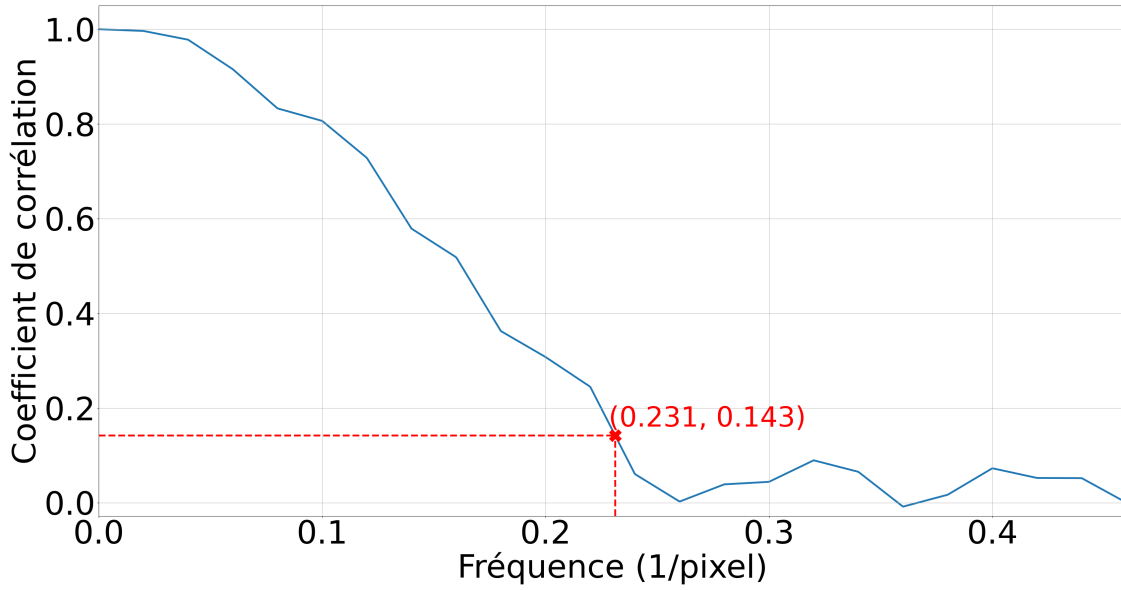


FIGURE 3.5 – Courbe de FSC et la résolution associée (0,231)

$$cFSC: P \times \Psi \rightarrow \mathbb{R} \quad (3.58)$$

$$(\rho_i, \psi_j) \mapsto \frac{\sum_{\nu \in \mathbb{R}^3} \omega_{i,j}(\nu) F_1(\nu) F_2(\nu)^*}{\sqrt{\sum_{\nu \in \mathbb{R}^3} \omega_{i,j}(\nu) F_1(\nu) F_1(\nu)^* \sum_{\nu \in \mathbb{R}^3} \omega_{i,j}(\nu) F_2(\nu) F_2(\nu)^*}} \quad (3.59)$$

avec F_1 et F_2 les transformées de Fourier des deux images considérées et $\omega_{i,j}(\nu)$ des poids définis par : $\omega_{i,j}(\nu) = a_i(\nu)b_j(\nu)$, avec

$$a_i(\nu) = \exp\left(-\frac{(\rho(\nu) - \rho_i)^2}{2\sigma_a^2}\right), \quad (3.60)$$

et

$$b_j(\nu) = \exp\left(-\frac{\mathcal{L}(\psi(\nu), \psi_j)}{2\sigma_b^2}\right), \quad (3.61)$$

avec $\mathcal{L}(\psi(\nu), \psi_j)$ la distance angulaire entre $\psi(\nu)$ et ψ_j . Ainsi, la cFSC se calcule en chaque point de la discrétisation par une corrélation croisée pondérée, donnant plus d'importance aux fréquences proches du point considéré.

À titre d'exemple, nous représentons en haut à gauche de la figure 3.10 les courbes de la cFSC entre l'objet vérité terrain, le Récepteurs AMPA, et une vue générée (alignée sur la vérité terrain). Les courbes associées à 2000 directions sont représentées, en gris clair pour la plus part, et les courbes associées aux directions X , Y et Z sont mises en valeur. Nous traçons également une carte de résolution, associant à chaque direction (ϕ_1, ϕ_2) l'antécédent de 0,143 par la courbe correspondante (en utilisant le système de coordonnées 3.1). La carte associée aux courbes précédemment mentionnées est représentée à droite de ces dernières. Ces figures mettent en évidence l'anisotropie de la résolution : la résolution selon l'axe Z ($\phi_2 = 0^\circ$ et $\phi_2 = 180^{circ}$) est beaucoup plus faible que dans les autres directions.

L'orientation de l'objet reconstruit dans l'espace est aléatoire et n'a aucune raison de correspondre à l'orientation de la vérité terrain. Par conséquent, les orientations associées à chacune des vues ne sont pas estimées de manière absolue mais relativement les unes par rapport aux autres. Il est donc nécessaire de recalibrer la reconstruction sur l'objet original avant de calculer le SSIM, la FSC ou la cFSC. Le recalage est effectué avec une première étape de recherche exhaustive maximisant l'information mutuelle entre les deux images [101]. Le pas de la discrétisation est fixé à 10 degrés. Pour raffiner le recalage, nous utilisons ensuite une descente de gradient.

3.6.1.2.4 Erreur sur les orientations Nous calculons également l'erreur sur les orientations. Comme les poses sont estimées de manière relative, nous ne nous attendons pas à ce que la matrice de rotation estimée $R_{\theta_l}^*$ soit proche de la matrice de rotation réelle R_{θ_l} , mais à ce que $M_l = R_{\theta_l}^{*-1} R_{\theta_l}$ soit proche de $M_m = R_{\theta_m}^{*-1} R_{\theta_m}$ pour tout $(l, m) \in \llbracket 1, N_v \rrbracket$. L'erreur sur les orientations η est alors définie comme la moyenne sur l'ensemble des couples (l, m) de l'écart entre $M_l M_m^{-1}$ et la matrice identité. Formellement :

$$\eta = \frac{1}{N_v^2} \sum_{l=1}^{N_v} \sum_{m=1}^{N_v} \|R_{\theta_l}^{*-1} R_{\theta_l} R_{\theta_m}^{-1} R_{\theta_m}^* - I\|_F, \quad (3.62)$$

où $\|\cdot\|_F$ désigne la norme de Frobenius.

L'objet NLR-Résistosome (voir figure 3.4) admet une symétrie C5 : il est invariant par toute matrice de rotation d'angle $\frac{2k\pi}{5}$ autour de l'axe Y , pour $k \in \llbracket 1, 5 \rrbracket$. Par conséquent, la métrique qui mesure l'erreur sur les orientations doit être adaptée. En effet il n'y a plus pour chaque vue une seule orientation correcte (à une référence près) mais 5. L'erreur sur les orientations se calcule dans ce cas de la manière suivante :

$$\eta = \min_{k \in \llbracket 1, 5 \rrbracket} \eta_k, \quad (3.63)$$

avec

$$\eta_k = \frac{1}{N_v^2} \sum_{l=1}^{N_v} \sum_{m=1}^{N_v} \|R_{\theta_l}^{*-1} R_{\theta_l} R_y(k) R_{\theta_m}^{-1} R_{\theta_m}^* - I\|_F, \quad (3.64)$$

avec $R_y(k)$ la matrice de rotation d'angle $\frac{2k\pi}{5}$ selon l'axe Y .

De même l'objet Clathrine admet une symétrie, mais nous ne sommes pas parvenu à la caractériser, nous n'indiquerons pas l'erreur sur les orientations associées à cet objet dans le tableau 3.2.

3.6.1.3 Analyse des hyperparamètres

Les trois méthodes présentées en parties 3.3, 3.4 et 3.5 admettent un certain nombre d'hyperparamètres. Nous présentons dans cette section une analyse de la qualité de la reconstruction, mesurée par le SSIM et la FSC, en fonction des valeurs prises par ces derniers. Toutes les expériences de cette sous-section sont effectuées sur l'objet Récepteurs AMPA.

3.6.1.3.1 Paramètres de l'échantillonnage préférentiel L'estimation des orientations s'effectue de la même manière quelque soit la méthode retenue, en utilisant la procédure d'échantillonnage préférentiel présentée en partie 3.3.3. Elle met en jeu 4 hyper-paramètres différents :

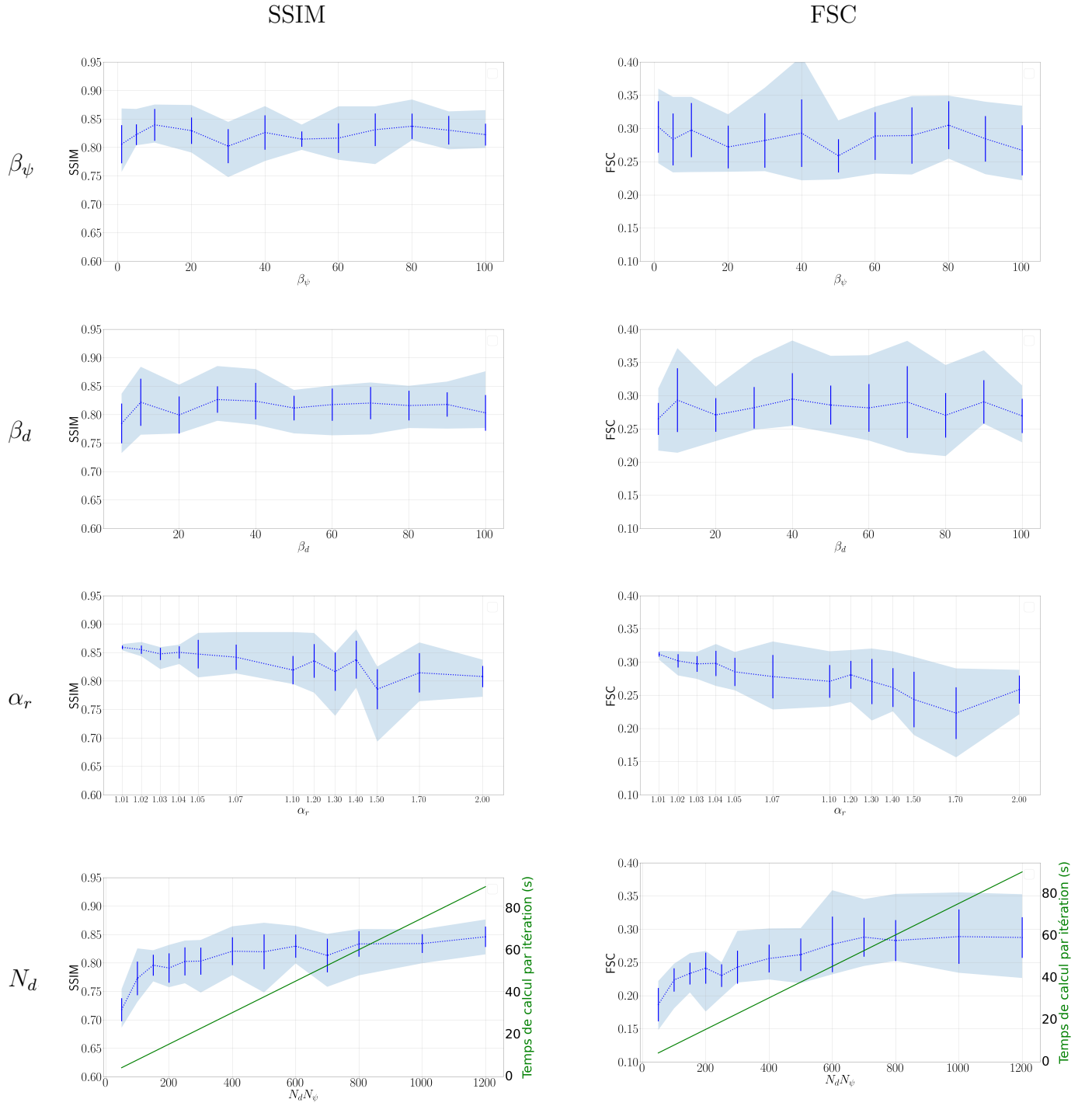


FIGURE 3.6 – Influences des hyper-paramètres β_d , β_ψ , α_r et $N_d N_\psi$ sur le SSIM et la FSC à la vérité terrain. Chaque point est une moyenne calculée sur 10 reconstructions. Les barres verticales représentent l'écart-type sur les 10 expériences, et l'aire en bleu est l'écart entre la valeur maximale et la valeur minimale. Méthode SHiReVol

- Les paramètres β_d et β_ψ qui déterminent la largeur des noyaux utilisées dans l'estimation des vraisemblances marginalisées (voir 3.20 et 3.21).
- Le ratio α_r par lequel est divisée à chaque itération la proportion de la composante uniforme α dans les distributions $\mathcal{Q}^{l,\psi}$ et $\mathcal{Q}^{l,d}$ (voir 3.23).
- Le nombre d'orientations tirées aléatoirement à chaque itération, $N_d N_\psi$, nous fixons $N_d = N_\psi^2$.

Nous testons seulement la sensibilité à ces paramètres de la méthode utilisant une représentation voxélique du volume. Nous supposons que les deux autres méthodes, présentées en 3.4 et 3.5, se comportent de manière similaire lorsque ces paramètres varient. De plus, nous testons chacun des 4 paramètres en fixant les 3 autres à leurs valeurs par défaut. Il aurait été en effet bien trop coûteux en temps et en énergie de tester chaque combinaison d'une grille de dimension 4. Nous fixons les valeurs par défaut à $N_d = 64$, $N_\psi = 8$, $\alpha_r = 1,2$, $\beta_\psi = \beta_d = 50$. Le volume $\hat{f}$ est initialisé avec des valeurs aléatoires uniformément distribuées entre 0 et 1. Le paramètre α dans 3.23 est initialisé à 1 de sorte à ce qu'au début de l'apprentissage la discrétisation de Fibonacci soit explorée uniformément. Nous utilisons 20 vues pour la reconstruction.

Sur la figure 3.6, nous avons tracé les courbes de SSIM et FSC en fonction de ces paramètres. Chaque point est une moyenne obtenue sur 10 reconstructions, les barres verticales représentent l'écart-type sur les 10 expériences, et l'aire en bleu est l'écart entre la valeur maximale et la valeur minimale. De manière générale, on constate que les courbes de FSC ont une allure similaire à celles des SSIM. L'influence des paramètres des noyaux β_d et β_ψ sur les deux métriques est relativement faible et on ne repère pas de tendance générale, ce qui laisse supposer que les légères variations sont dues au caractère stochastique des expériences.

En revanche, concernant le paramètre α_r on observe deux tendances. Premièrement, le SSIM (et la FSC) décroît légèrement avec α_r . Ceci s'explique par le fait que pour des valeurs de α_r élevées, la proportion de distribution uniforme dans 3.23 décroît plus rapidement que pour de faibles valeurs. Le module d'estimation des poses par échantillonnage préférentiel va donc rapidement centrer la recherche autour des orientations ayant donné de bon résultats aux premières itérations. Or ces dernières peuvent être éloignées des orientations réelles, car au début de l'apprentissage le volume en cours de reconstruction est très loin de la vérité terrain. Par conséquent, pour des valeurs élevées de α_r , l'algorithme sera plus susceptible de rester bloqué dans un minimum local. Deuxièmement, l'écart type du SSIM (et de la FSC) sur les 10 expériences tend à croître avec α_r . En effet, pour des valeurs faibles de α_r , le résultat de l'algorithme sera bien moins dépendant des orientations tirées aléatoirement lors des premières itérations, ce qui se traduit par une plus faible variabilité du SSIM (et de la FSC). Ces deux constats ne signifient pas que choisir une valeur de α_r la plus proche possible de 1 soit la meilleure solution. Comme on l'a vu en partie 3.3.5, on impose un nombre minimum d'itération qui est d'autant plus élevé que α_r est faible, l'algorithme sera donc d'autant plus long que α_r est faible. Choisir α_r relève donc d'un compromis entre la qualité de la reconstruction et le temps de calcul. En pratique, nous pensons que $\alpha_r = 1,05$ est un bon compromis.

Enfin, on constate que le SSIM (et la FSC) croît avec le nombre d'orientations échantillonnées à chaque itération, $N_\psi N_d$. En effet, si $N_\psi N_d$ est trop faible, il est plus probable que l'algorithme n'explore pas la région de l'espace de recherche contenant la solution optimale. D'un autre côté, le temps de calcul par itération est proportionnel à $N_d N_\psi$. Choisir N_d et N_ψ est donc un compromis entre la qualité de la reconstruction et le temps de calcul. En pratique, nous pensons que $N_\psi = 8$ et $N_d = 64$ est un bon compromis.

Nous aimerions préciser que les graphes des SSIMs en fonction de α_r et β_ψ sont différents de ceux présentés dans notre article [87]. Nous avons en effet reconduit ces expériences, à causes d'im-

σ		0,06	0,05	0,04	0,035	0,025	0,015
Récepteurs AMPA	$N_{opt}^{exp}(\sigma)$	40	50	100	175	175	175
	$N_{opt}^{est}(\sigma)$	14	23	45	66	181	838
HIV-Vaccine	$N_{opt}^{exp}(\sigma)$	30	30	75	125	150	> 250
	$N_{opt}^{est}(\sigma)$	15	26	50	74	200	936
Clathrine	$N_{opt}^{exp}(\sigma)$	40	50	100	125	100	> 250
	$N_{opt}^{est}(\sigma)$	10	18	34	50	137	630
NLR-Résistosome	$N_{opt}^{exp}(\sigma)$	30	50	100	150	150	> 250
	$N_{opt}^{est}(\sigma)$	7	12	22	33	89	410

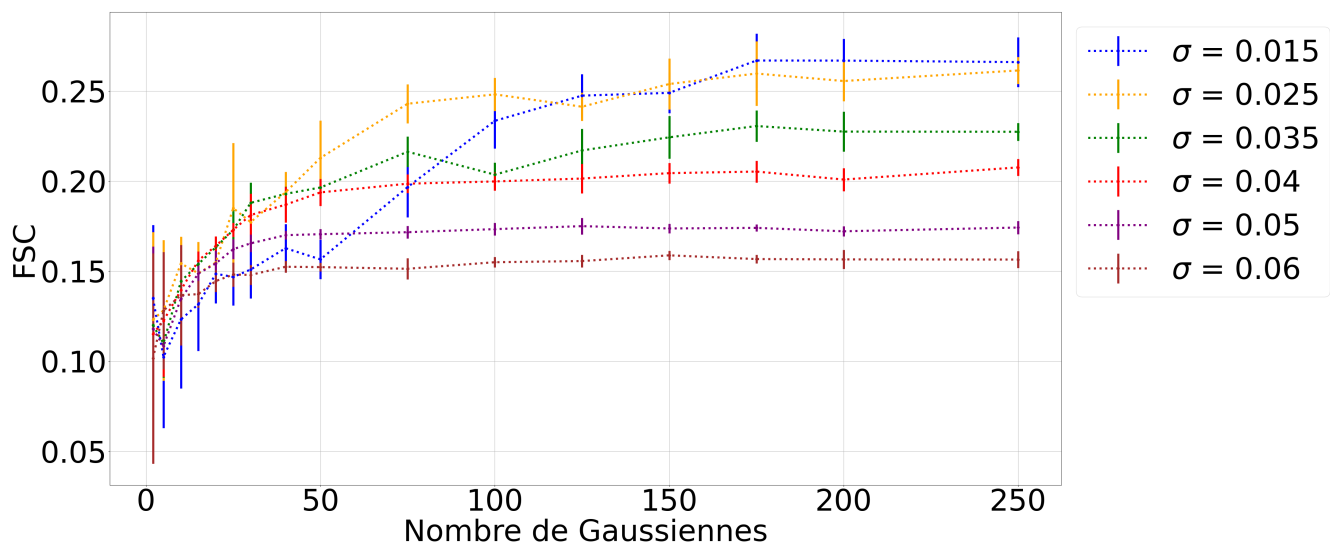
TABLE 3.1 – Nombres de gaussiennes optimaux estimés $N_{opt}^{est}(\sigma)$ et expérimentaux $N_{opt}^{exp}(\sigma)$ pour différents écart-types de gaussiennes σ . Méthode ShiReGmm

perfections dans le processus expérimental. Concernant α_r , l'algorithme s'arrêtait prématurément pour de faibles valeurs du paramètre car nous n'avions pas imposé un nombre minimal d'itérations. Concernant β_ψ , nous utilisons une valeur plus élevée de $N_d N_\psi$ que pour les autres expériences.

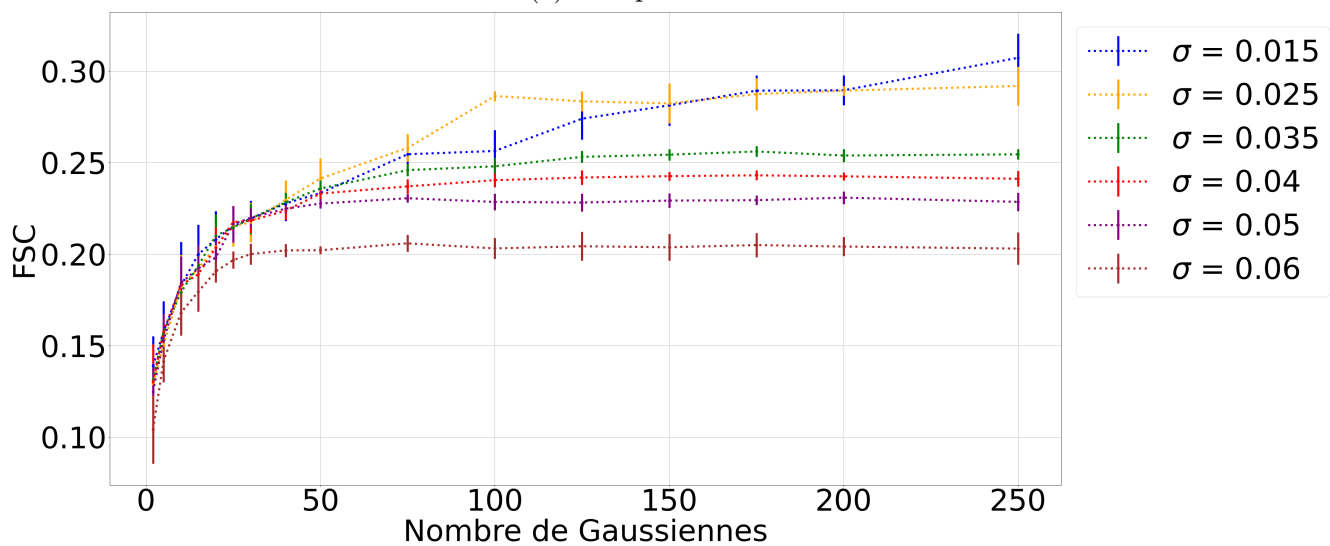
3.6.1.3.2 Paramètres de la représentation du volume sous forme de mélange de gaussiennes La méthode présentée en 3.4 met en jeu deux hyper-paramètres supplémentaires, en plus de ceux de l'échantillonnage préférentiel : l'écart-type σ des gaussiennes, et le nombre N_g de gaussiennes utilisées pour représenter le volume. Pour chaque combinaison (σ, N_g) sur une grille de valeurs discrètes, nous effectuons 10 reconstructions et notons la FSC entre le résultat et la vérité terrain. Les paramètres de l'échantillonnage préférentiel sont fixés à $\alpha_r = 1,05$, $\beta_d = \beta_\psi = 50$ et $N_d = N_\psi^2 = 64$. Nous appliquons cette procédure aux 4 objets vérité terrains. Les résultats obtenus sur les objets Récepteurs AMPA et Clathrine sont synthétisés sur la figure 3.7. On observe quatre comportements. D'abord, on constate que quelque soit l'écart-type des gaussiennes, la qualité de la reconstruction s'améliore avec le nombre de gaussiennes, ce qui est prévisible : plus de gaussiennes c'est plus de matière pour représenter le volume. Ensuite, toutes les courbes semblent se stabiliser autour d'une certaine valeur $FSC_{max}(\sigma)$ (différente pour chaque valeur de σ), ce qui laisse supposer que quelque soit la valeur de σ , il existe un nombre optimal de gaussiennes $N_{opt}(\sigma)$ à partir duquel ajouter des gaussiennes supplémentaires n'améliore pas la qualité de la reconstruction sans pour autant la dégrader. Puis, $FSC_{max}(\sigma)$ est d'autant plus élevé que σ est faible. Ce résultat est prévisible, car des gaussiennes plus petites donnent la possibilité de représenter des détails plus fins du volume. Enfin, $N_{opt}(\sigma)$ augmente également avec σ . En effet, plus les gaussiennes sont petites, plus il en faudra pour couvrir le même volume.

Pour un jeu de données quelconque associé à un objet différent de ceux sur lesquels les expérimentations ont été effectuées, nous aimerions pouvoir prédire $N_{opt}(\sigma)$ et utiliser ce nombre de gaussiennes pour reconstruire l'objet. En effet, si l'on choisit un nombre de gaussiennes N_g inférieur à $N_{opt}(\sigma)$, la reconstruction s'avérera sous-optimale et si on fixe $N_g > N_{opt}(\sigma)$, l'algorithme sera plus coûteux en temps de calcul que nécessaire. Nous constatons dans un premier temps que les valeurs de $N_{opt}(\sigma)$ ne sont pas indépendantes de l'objet considéré (par exemple, on observe sur la figure 3.7 que $N_{opt}(0,05) = 30$ pour Clathrine et $N_{opt}(0,05) = 50$ pour Récepteurs AMPA). Il ne semble donc pas pertinent d'extrapoler les valeurs de $N_{opt}(\sigma)$ observées sur les courbes de la figure 3.7 à un nouveau jeu de données.

Nous avons pensé à une autre approche. Pour rappel, les vues (y_l) sont reliées à l'objet à reconstruire par la formule suivantes



(a) Récepteurs AMPA



(b) Clathrine

FIGURE 3.7 – Influence du nombre gaussiennes et de leur écart type σ sur la FSC à la vérité terrain. Chaque point est une moyenne calculée sur 10 reconstructions. σ est exprimée en pourcentage de la taille de l'image

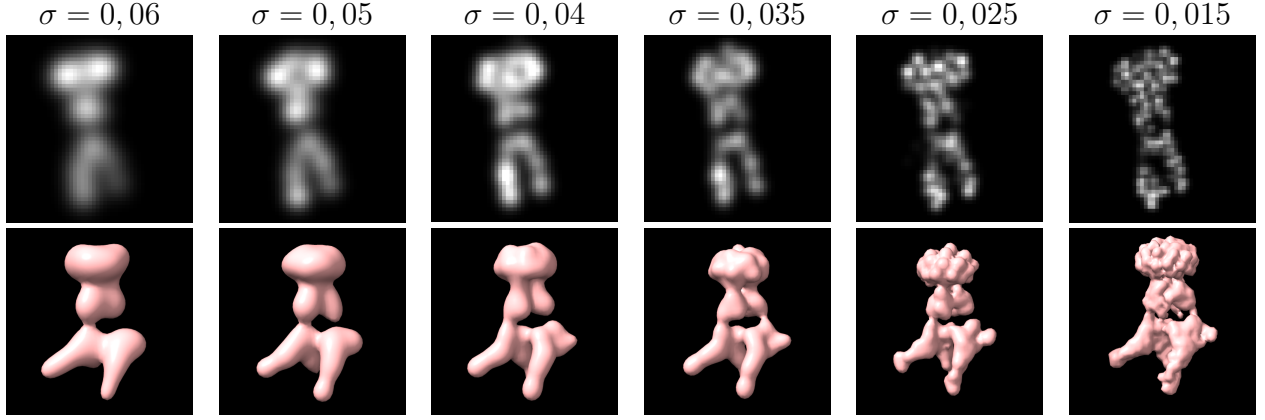


FIGURE 3.8 – Résultats visuels de la méthode utilisant une représentation sous forme de mélange de gaussiennes du volume pour différents écarts-types de gaussiennes

$$y_l = h * \mathcal{T}_{t_l}(\mathcal{R}_{\theta_l}(f)) + \epsilon. \quad (3.65)$$

Par utilisation du théorème de Fubini, et en remarquant que $\int_{x \in \Omega} \epsilon(x) dx = 0$, car le bruit est centré, on obtient :

$$\int_{x \in \Omega} y_l(x) dx = \int_{x \in \Omega} h(x) dx \int_{x \in \Omega} f(x) dx. \quad (3.66)$$

Or on a choisi h comme une fonction gaussienne, elle vérifie donc $\int_{x \in \Omega} h(x) dx = 1$. Par conséquent

$$\int_{x \in \Omega} f(x) dx = \int_{x \in \Omega} y_l(x) dx. \quad (3.67)$$

On peut donc estimer le nombre de gaussiennes optimal pour représenter une vue y_l du jeu de données, et considérer que ce nombre est égal à $M_l(\sigma)$. $M_l(\sigma)$ est estimé par la méthode présentée en 3.5.2, en posant $\Delta = \sigma^2 I$ au lieu de $\Delta = \sigma^2 I + \Sigma_h$. En pratique, on posera plutôt :

$$N_{opt}(\sigma) = \frac{1}{N_v} \sum_{l=1}^{N_v} M_l(\sigma) = \frac{1}{N_v} \sum_{l=1}^{N_v} \left\lceil \frac{2S_l}{(2\pi)^{\frac{3}{2}} \sigma^2} \right\rceil = \frac{1}{N_v} \sum_{l=1}^{N_v} \left\lceil \frac{2 \sum_{x \in \Omega} y_l(x)}{(2\pi)^{\frac{3}{2}} \sigma^2} \right\rceil, \quad (3.68)$$

en effet toutes les égalités écrites ci-dessus ne sont valides que si Ω est un ensemble infini, or il s'agit en pratique d'un ensemble discret et continu. La moyenne sur l'ensemble des vues permet donc de réduire l'influence des approximations induites par la discrétisation. Nous avons cherché à vérifier si les valeurs de $N_{opt}(\sigma)$ estimées à partir des données d'entrées par l'équation (3.68) (que l'on notera $N_{opt}^{est}(\sigma)$), correspondent aux valeurs que l'on observe à partir des courbes expérimentales 3.7 (que l'on notera $N_{opt}^{exp}(\sigma)$). Nous avons pour cela synthétisé dans le tableau 3.1 les valeurs de $N_{opt}^{est}(\sigma)$ et de $N_{opt}^{exp}(\sigma)$ pour les 4 objets considérés dans le tableau. Les valeurs $N_{opt}^{exp}(\sigma)$ sont nécessairement assez imprécises, d'abord parce que la différence en nombre de gaussiennes entre deux points du graphe est relativement importante et ensuite parce que la valeur à partir de laquelle une courbe se stabilise est estimée visuellement (nous aurions pu aussi nous baser sur les différences finies et fixer un seuil, mais le choix de ce seuil aurait également été arbitraire). Nous remarquons que les valeurs de $N_{opt}^{exp}(\sigma)$ ne correspondent pas aux valeurs de $N_{opt}^{est}(\sigma)$. D'autre

		Récepteurs AMPA	HIV-Vaccine	Clathrine	NLR-Résistosome
Poses connues	SSIM	0,91	0,93	0,92	0,93
	FSC	0,42	0,44	0,34	0,35
SHiReVol	SSIM	0,85	0,86	0,82	0,88
	FSC	0,29	0,33	0,29	0,28
	EA	4,9	4,2	—	3,1
ShiReGmm	SSIM	0,76	0,74	0,85	0,91
	FSC	0,22	0,17	0,32	0,30
	EA	6,1	5,3	—	3,4
ShiReGmm2	SSIM	0,74	0,69	0,65	0,67
	FSC	0,17	0,12	0,11	0,11
	EA	8,2	45,3	—	15,2
cryo-RANSAC	SSIM	0,54	0,53	0,55	0,53
	FSC	0,05	0,06	0,11	0,10
MP3DR	SSIM	0,72	0,72	0,67	0,81
	FSC	0,14	0,13	0,14	0,17

TABLE 3.2 – Comparaison quantitative des résultats de nos 3 méthodes, cryo-RANSAC, MP3DR et la reconstruction avec poses connues. Les métriques SSIM et FSC sont reportées pour nos 3 méthodes, et l’erreur aux angles réels (EA) , en degré (°) est reportée pour nos 3 méthodes. Nous n’indiquons par l’erreur sur les angles de l’objet Clathrine car nous ne sommes pas parvenus à caractériser sa symétrie. Chaque valeur est une moyenne effectuée sur 100 reconstructions

part, malgré les imprécisions de l’estimation des $N_{opt}^{exp}(\sigma)$, il ne semble pas s’agir d’une simple erreur d’un facteur multiplicatif. Nous en concluons que notre méthode d’estimation du nombre de gaussiennes n’est pas adaptée.

3.6.1.4 Résultats

Dans cette section, nous présentons dans un premier temps une comparaison de nos méthodes, entre elles et avec celles de l’état de l’art. Les méthodes utilisant respectivement une représentation voxélique du volume, une représentation sous forme de mélange de gaussiennes du volume, une représentation sous forme de mélange de gaussiennes des vues et du volumes sont notées respectivement SHiReVol, ShiReGmm et ShiReGmm2. SHiRe est l’acronyme de ”Stochastic Hierarchical Reconstruction”.

Ensuite, nous analysons la robustesse de la méthode SHiReVol aux différents paramètres du modèle de dégradation (niveau de bruit, d’anisotropie, degré de labellisation).

3.6.1.4.1 Résultats comparatifs Dans ce paragraphe, nous comparons les 3 méthodes présentées dans ce chapitre entre elles ainsi qu’avec deux autres méthodes de l’état de l’art. Nous utilisons les paramètres suivants pour l’échantillonnage préférentiel : $\alpha_r = 1,05$, $\beta_d = \beta_\psi = 50$ et $N_d = N_\psi^2 = 64$. Concernant la méthode ShiReGmm, nous utilisons $N_v = 250$ gaussiennes d’écart type $\sigma = 0,015$ (exprimé en proportion de la taille de l’image). Pour la méthode ShiReGmm2, nous fixons également $N_v = 250$ et $\sigma = 0,015$, et choisissons M_l en utilisant la méthode présentée en partie 3.5.2.

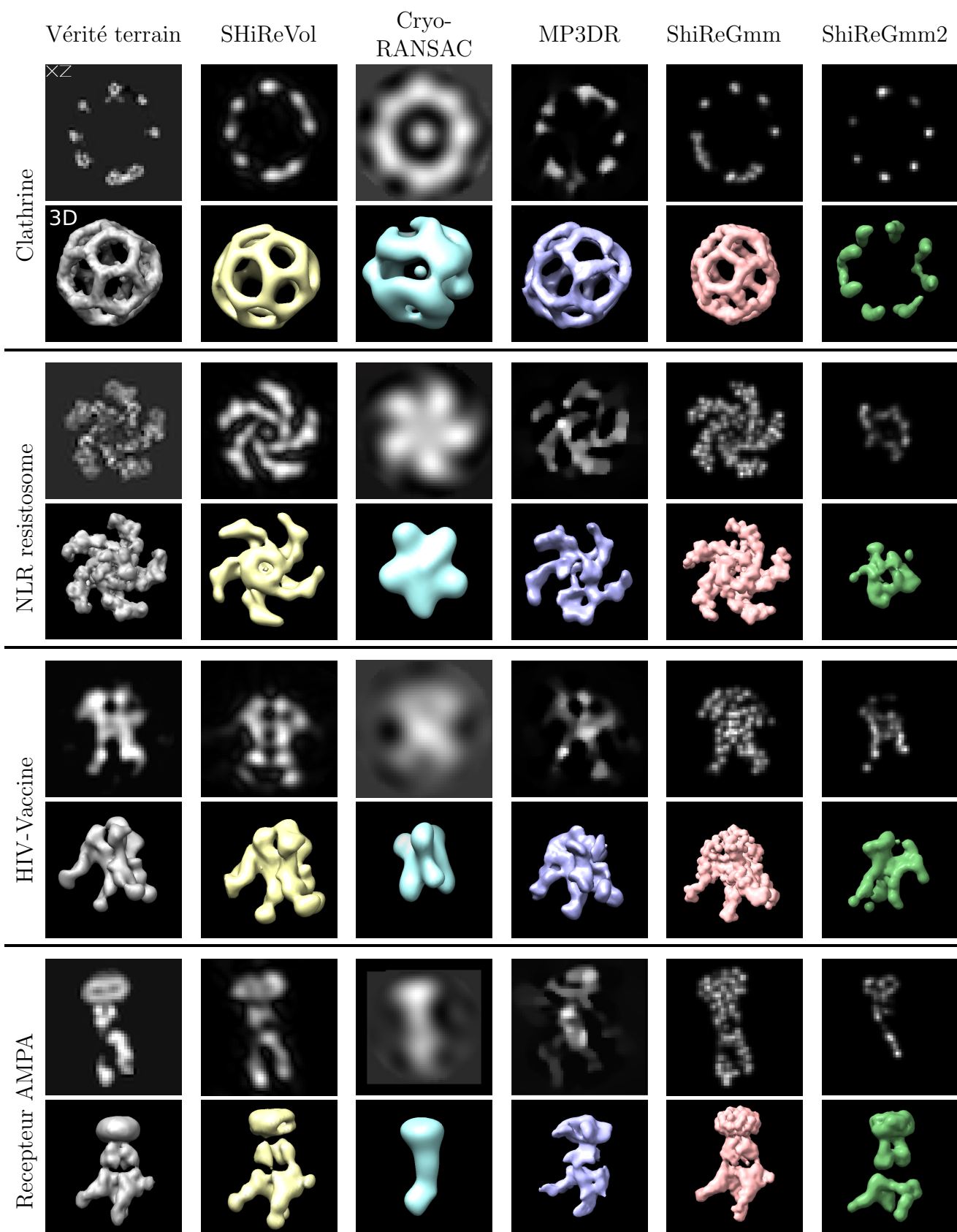


FIGURE 3.9 – Résultats de reconstructions obtenus sur des données synthétiques. Chaque volume est représenté par une coupe dans le plan XZ et une représentation iso-surfacique 3D

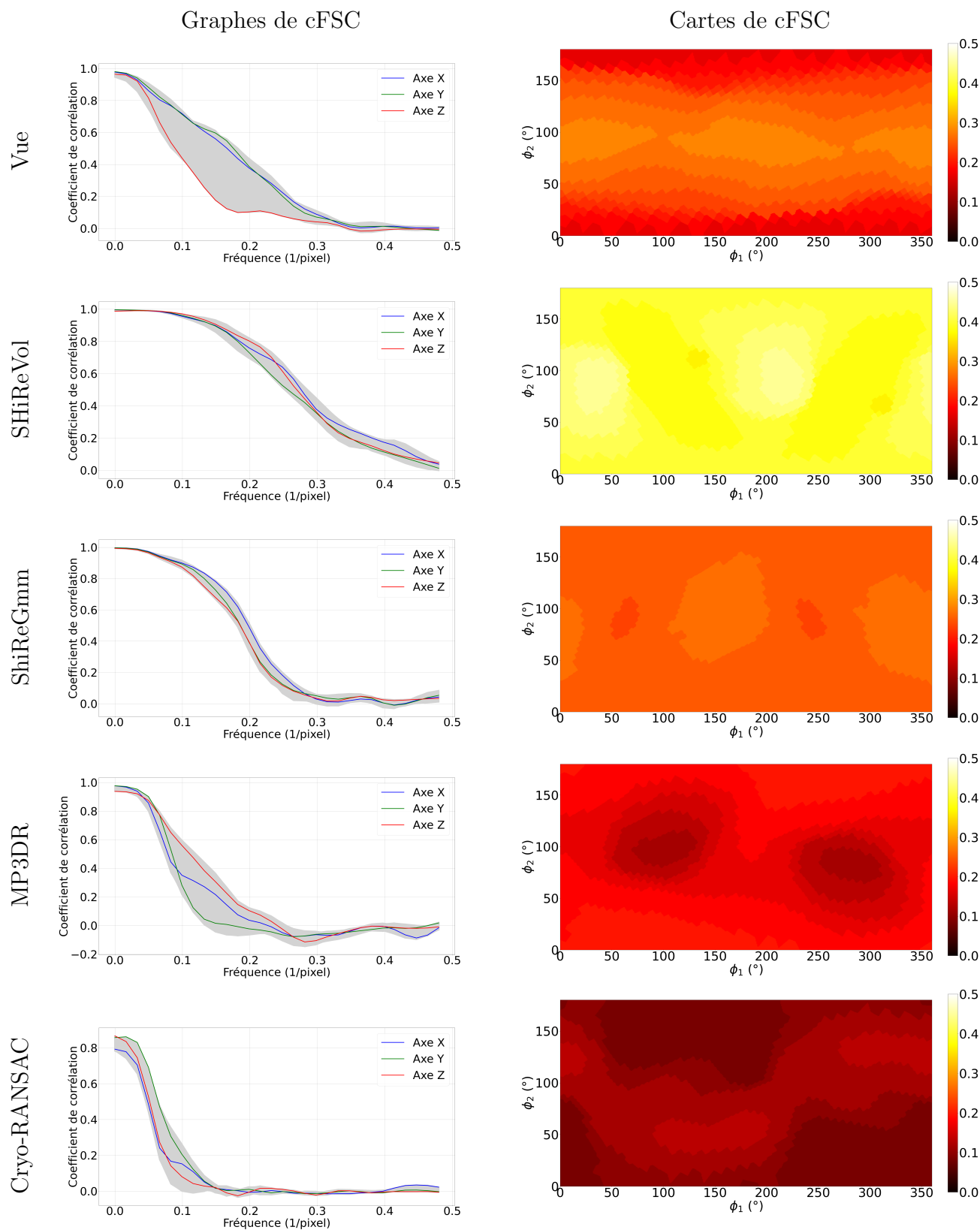


FIGURE 3.10 – Graphes et cartes de cFSC entre l’objet vérité terrain et les résultats de différentes méthodes (ligne 2 à 4) et une vue (ligne 1). Chaque point sur la carte est une direction, représentée par deux angles $\phi_1 \in [0, 360]$ et $\phi_2 \in [0, 180]$ (voir figure 3.1). Objet Récepteurs AMPA.

L’approche standard pour la RPI en microscopie par fluorescence est d’appliquer des méthodes de reconstruction habituellement dédiées à la cryo-ME sur des données 2D. Pour appliquer cette idée, nous créons artificiellement des projections 2D à partir de nos volumes d’entrées 3D et effectuons la reconstruction à partir de ces dernières en utilisant une méthode de reconstruction cryo-ME basée sur l’algorithme RANSAC [102] (intégrée dans le logiciel Scipion [103]). Nous nommons cette méthode cryo-RANSAC. Nous comparons également nos résultats avec ceux de la méthode de RPI dédiée aux modèles convolutifs 3D décrite dans [27]. Nous l’appelons MP3DR (Multiple Particles 3D Reconstruction) par commodité. Cette méthode nécessite la sélection d’un petit nombre de particules pour maintenir un coût de calcul raisonnable. Nous sélectionnons manuellement 5 particules représentatives pour chaque ensemble de données.

La figure 3.9 résume les résultats visuels obtenus sur les données simulées avec nos 3 méthodes, cryo-RANSAC et MP3DR. Nous constatons que les résultats obtenus par la méthode cryo-RANSAC sont extrêmement peu précis, bien qu’on reconnaisse la forme originale. Ce résultat illustre l’importance de prendre en compte un modèle d’imagerie 3D approprié, tandis que travailler avec des projections 2D entraîne une perte d’informations. La méthode ShiReGmm2 ne parvient pas non plus à reconstruire correctement les objets, les reconstructions présentent de nombreuses parties manquantes et des déformations par rapport à la vérité terrain. La méthode MP3DR est plus robuste que les deux précédentes, mais les reconstructions obtenues par cette dernière présentent tout de même des différences conséquentes avec les objets de référence. Les résultats de la méthode ShiReGmm sont dans l’ensemble fidèles à la forme d’origine. Néanmoins, la représentation sous forme de mélange de gaussiennes se traduit visuellement par une granularité que l’on peut considérer comme indésirable pour les objets HIV-Vaccine Récepteurs AMPA et Clathrine car l’objet vérité terrain a une forme relativement lisse. En revanche, l’objet vérité terrain NLR-Résistosome admet une certaine granularité, il se prête donc particulièrement bien à la représentation sous forme de mélange de gaussiennes. Pour finir, les reconstructions obtenues par la méthode SHiReVol nous semblent de meilleure qualité que celles obtenues par toutes les autres méthodes, sur tous les objets à l’exception du NLR-Résistosome (dont la structure se prête mieux à la représentation sous forme de mélange de gaussiennes). Voyons maintenant ce que donnent les résultats quantitatifs.

Le tableau 3.2 résume les résultats quantitatifs obtenus par les 5 méthodes, ainsi que par la reconstruction avec poses connues. Ils sont mesurés par le SSIM, la FSC et l’erreur aux angles réels. Nous constatons que, pour l’ensemble des particules, nos algorithmes SHiReVol et ShiReGmm offrent des valeurs de SSIM et de FSC (Fourier Shell Correlation) significativement meilleures que les reconstructions réalisées avec cryo-RANSAC. Nous surpassons également de manière considérable MP3DR sur les deux métriques considérées. À noter qu’en plus de cette amélioration quantitative, l’un des principaux avantages de nos méthodes par rapport à MP3DR est que nous n’avons pas besoin de sélectionner un nombre réduit de vues, ce qui est sujet à des erreurs de classification ou à des biais humains, et nous pouvons utiliser directement l’ensemble des données pour la reconstruction. Comme prévu, nos méthodes fournissent des valeurs légèrement inférieures de SSIM et FSC par rapport à la reconstruction avec poses connues. Ce résultat montre que les erreurs d’estimation d’orientation n’ont qu’un faible impact sur la reconstruction finale, alors que l’erreur moyenne sur les angles n’est pas négligeable. Par ailleurs, aucune des deux méthodes SHiReVol et ShiReGmm ne surpasse l’autre sur les 4 objets considérés : ShiReGmm donne de meilleurs résultats, à la fois en termes de SSIM et FSC, sur Clathrine et NLR-Résistosome alors que SHiReVol s’avère plus performante sur les objets Récepteurs AMPA et HIV-Vaccine. Enfin, il est à noter que ShiReGmm2 affiche des performances nettement inférieures par rapport à ShiReGmm et SHiReVol, et parvient à surpasser MP3DR uniquement sur l’objet Récepteurs AMPA.

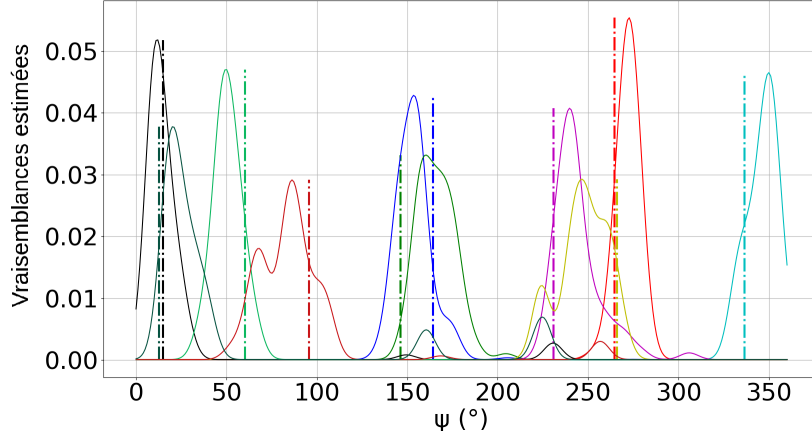
Afin d'évaluer la résolution de la reconstruction de manière directionnelle, nous illustrons sur la figure 3.10 les courbes et cartes de cFSC calculées entre l'objet vérité terrain Récepteurs AMPA et, respectivement, une vue et les résultats de reconstruction de SHiReVol, ShiReGmm, MP3DR et Cryo-RANSAC (le détail du calcul de la cFSC est explicité au paragraphe 3.6.1.2.3). Les courbes et la carte de cFSC de la reconstruction obtenue par SHiReVol montrent non seulement que la résolution de la reconstruction est bien plus uniforme que celle d'une vue, mais la résolution dans le plan latéral est elle aussi améliorée. La reconstruction obtenue par la méthode ShiReGmm présente également, d'après la carte de cFSC, une résolution relativement uniforme, bien qu'inférieure à celle de SHiReVol. Quant aux méthodes MP3DR et Cryo-RANSAC, les résultats présentent des résolutions à la fois non uniformes et inférieures à la résolution des vues d'entrée.

Après évaluation qualitative et quantitative des résultats, nous pouvons conclure que nos deux méthodes SHiReVol et ShiReGmm surpassent les résultats de reconstruction de l'état de l'art. Par ailleurs, les performances relatives de ces deux méthodes dépendent de la structure considérée : ShiReGmm sera plus adaptée pour traiter des structures présentant une certaine granularité, telle que le NLR-Résistosome. Néanmoins, dans la majorité des cas SHiReVol est plus performante, car elle présente l'avantage de pouvoir représenter tout type de structure, alors que ShiReGmm est contrainte par la représentation sous forme de mélange de gaussiennes, qui a pour conséquence néfaste d'ajouter à la structure une granularité qui n'a pas lieu d'être.

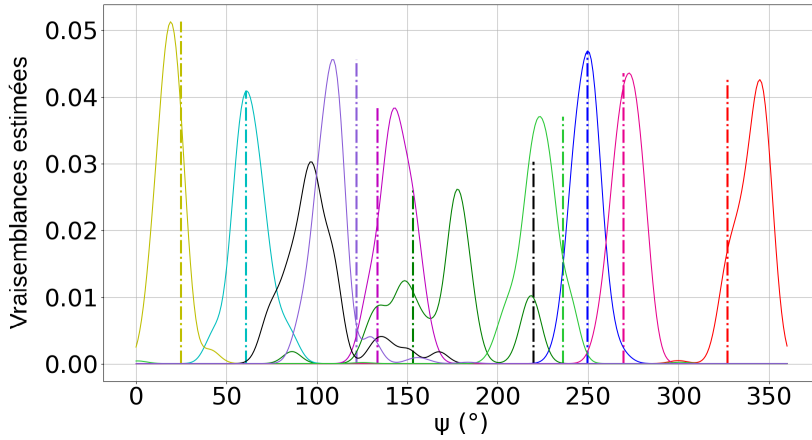
3.6.1.4.2 Estimation des distributions d'importance Afin d'évaluer plus précisément la précision de l'estimation des orientations, nous montrons les vraisemblances marginales approximatives $\check{\pi}^{l,d}$ et $\check{\pi}^{l,\psi}$ obtenues lors de la dernière itération de la reconstruction. Pour rappel, ces distributions sont utilisées pour échantillonner plusieurs axes d et angles ψ , à partir desquels la représentation optimale de l'axe et de l'angle est estimée. Les vraisemblances $\check{\pi}^{l,\psi}$ sont représentées sur les figures 3.11a et 3.11b, où chaque courbe colorée représente la vraisemblance d'une vue. Les barres verticales représentent les valeurs vérité terrain de ψ pour chaque vue. La somme sur l'ensemble des vues des vraisemblances $\check{\pi}^{l,d}$ est représentée sur la figure 3.11c sur une carte 2D où chaque axe représente une dimension de $d = (\phi_1, \phi_2)$. Les valeurs vérité terrain d sont représentées par des croix jaunes.

Pour ψ et d , nous observons que les vraisemblances sont presque toutes approximativement centrées autour des valeurs vérité terrain, ce qui montre que les orientations sont correctement estimées. Une seule estimation de vraisemblance n'est pas proche de la vérité terrain (la courbe noire de la figure 3.11b).

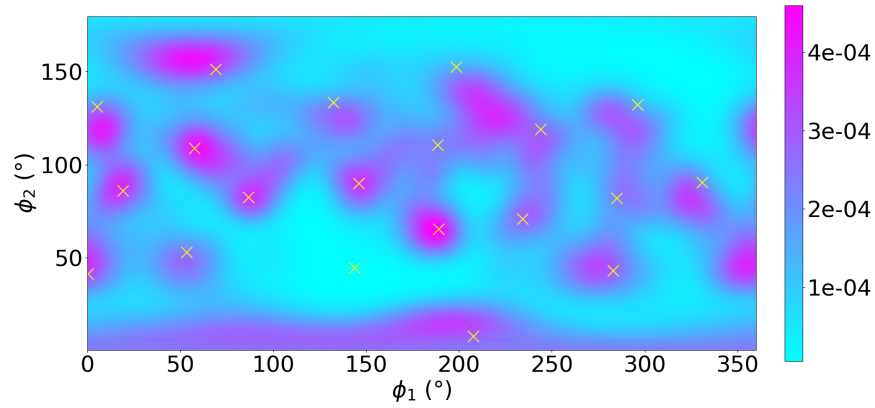
3.6.1.4.3 Robustesse à l'anisotropie Afin d'évaluer la robustesse de notre méthode SHiReVol à différents niveaux d'anisotropie, nous avons réalisé des reconstructions à partir de données générées avec différentes valeurs de σ_z . Les résultats visuels sont présentés sur la figure 3.12, où nous montrons également l'impact du nombre de vues. Lorsque le nombre de vues augmente, l'algorithme de reconstruction est en mesure de combiner davantage d'informations complémentaires, ce qui améliore la qualité de la reconstruction. À mesure que l'anisotropie augmente, une seule vue contient moins d'informations, et le nombre minimum de vues nécessaire pour obtenir une reconstruction satisfaisante augmente également. Pour $\sigma_z = 5$, le nombre minimum de vues est d'environ 20, et il est d'environ 40 vues pour $\sigma_z = 10$. Pour un niveau d'anisotropie extrêmement élevé de $\sigma_z = 15$, plus de 60 vues seraient nécessaires pour obtenir une reconstruction correcte.



(a) 10 premières vues



(b) 10 dernières vues



(c)

FIGURE 3.11 – Vraisemblances estimées $\check{\pi}^{l,\psi}$ et $\check{\pi}^{l,d}$ à la dernière itération de la reconstruction sur des données simulées à partir de Récepteurs AMPA (voir figure 3.4). (a) et (b) : Vraisemblances $\check{\pi}^{l,\psi}$, où chaque courbe représente une vue et les barres verticales représentent les valeurs réelles (les résultats sont séparés en deux graphiques pour une meilleure visualisation). (c) : Somme des vraisemblances $\check{\pi}^{l,d}$ de toutes les vues. Les croix jaunes sont les vraies valeurs de d .

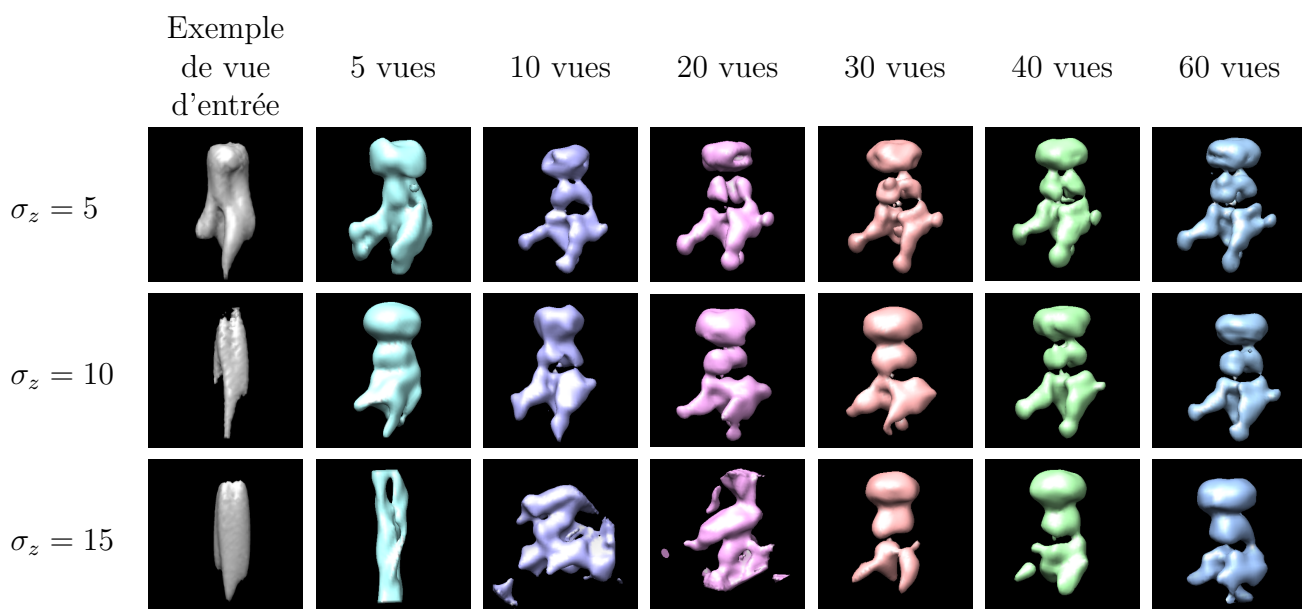


FIGURE 3.12 – Reconstructions obtenues avec différents niveaux d'anisotropie σ_z et différents nombres de vues. Méthode SHiReVol

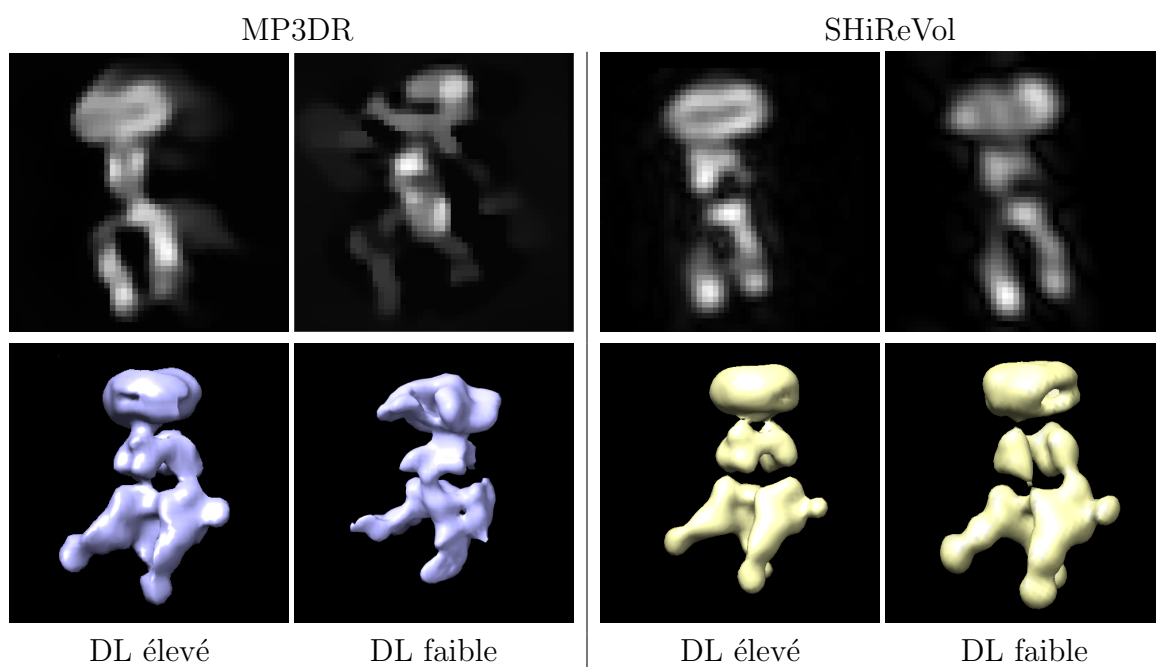


FIGURE 3.13 – Comparaison des reconstructions des méthodes SHiReVol et MP3DR dans le cas de degrés de labellisations faible et élevé

3.6.1.4.4 Robustesse à des faibles degrés de labellisation Dans cette section, nous nous concentrons sur la comparaison de la méthode SHiReVol avec MP3DR et évaluons l’impact du degré de labellisation (DL). Rappelons que nous simulons la variation de DL le long de la structure en soustrayant des tâches gaussiennes à des emplacements aléatoires. Sur la figure 3.13, nous montrons les reconstructions obtenues avec la méthode SHiReVol et MP3DR dans deux conditions expérimentales différentes :

- Faible DL : les paramètres de simulation par défaut, ceux utilisés pour simuler les données présentées sur la figure 3.4. Cela correspond à une situation réaliste.
- Haut DL : aucune tâche gaussienne n’est retirée de l’image de référence.

Notre méthode SHiReVol est capable d’obtenir des reconstructions satisfaisantes dans les deux cas, tandis que MP3DR échoue en cas de faible DL. Cela s’explique par la nature séquentielle de MP3DR : une première reconstruction avec deux vues est réalisée, puis les autres vues sont introduites une par une. Le défaut de cette procédure est que dans le cas d’un faible DL, les vues sont trop différentes pour obtenir une estimation correcte des poses avec seulement deux d’entre elles. Par conséquent, la reconstruction initiale à deux vues est inexacte et introduit un biais préjudiciable qui se propage lorsque les autres vues sont ajoutées séquentiellement. En revanche, notre approche utilise toutes les vues conjointement de manière désordonnée, ce qui nous permet d’utiliser toutes les informations et d’obtenir une bonne reconstruction.

3.6.1.4.5 Robustesse au bruit Afin d’évaluer la robustesse de SHiReVol à différents niveaux de bruit, nous avons réalisé des reconstructions à partir de données générées avec différents rapport signal à bruit (RSB). Le bruit généré est gaussien d’écart type

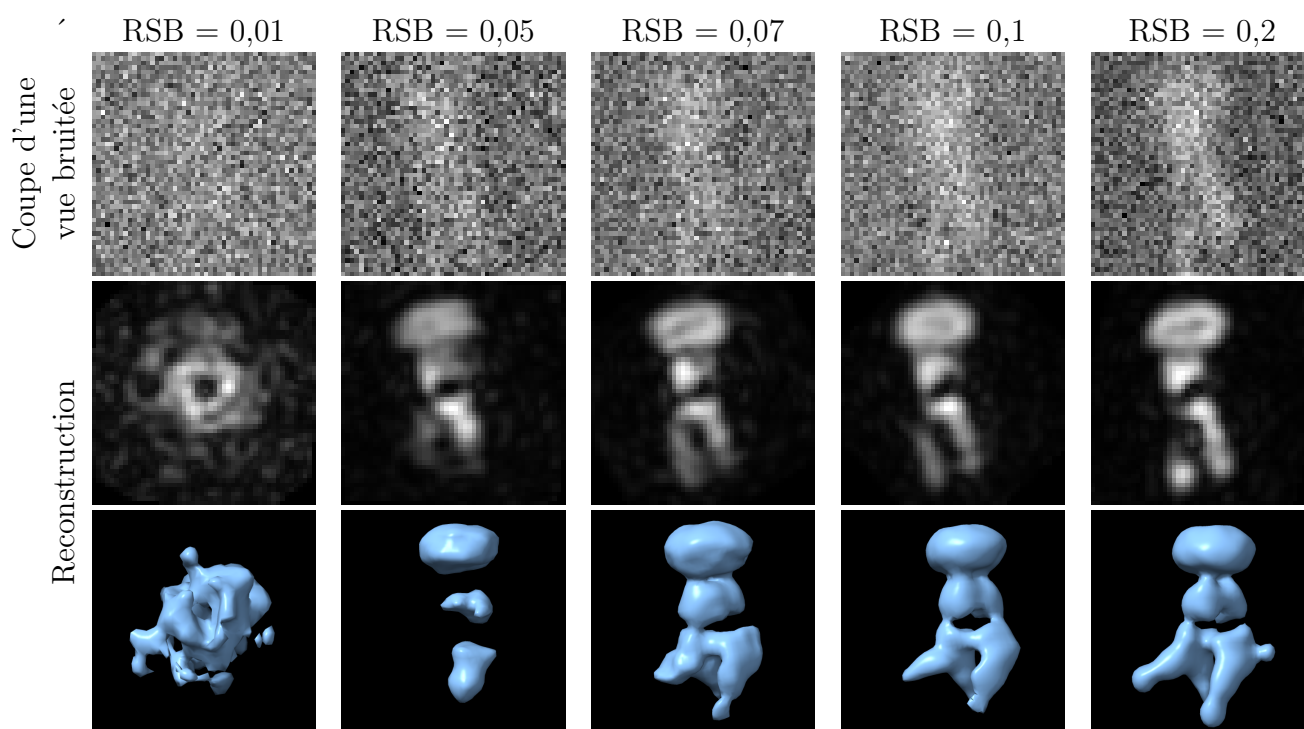
$$\xi = \sqrt{\frac{\sum_{x \in \Omega} f(x)^2}{RSB}}. \quad (3.69)$$

Les résultats visuels ainsi que la FSC à la vérité terrain sont présentés sur la figure 3.14. 20 vues sont utilisées pour la reconstruction. On remarque que la courbe de FSC se stabilise à partir de $RSB = 0,2$, ce qui est cohérent avec les résultats visuels : la reconstruction pour $RSB = 0,2$ est quasiment identique à celle obtenue avec le niveau de bruit pas défaut ($RSB = 5$). Au vu de la forte robustesse au bruit de notre méthode, nous n’avons pas envisagé l’implémentation d’un terme de régularisation. En effet, en microscopie par fluorescence, le RSB des images acquises est de l’ordre de 10 et descend rarement en dessous de 1 ce qui est largement au-dessus des valeurs de RSB à partir desquels la reconstruction se dégrade.

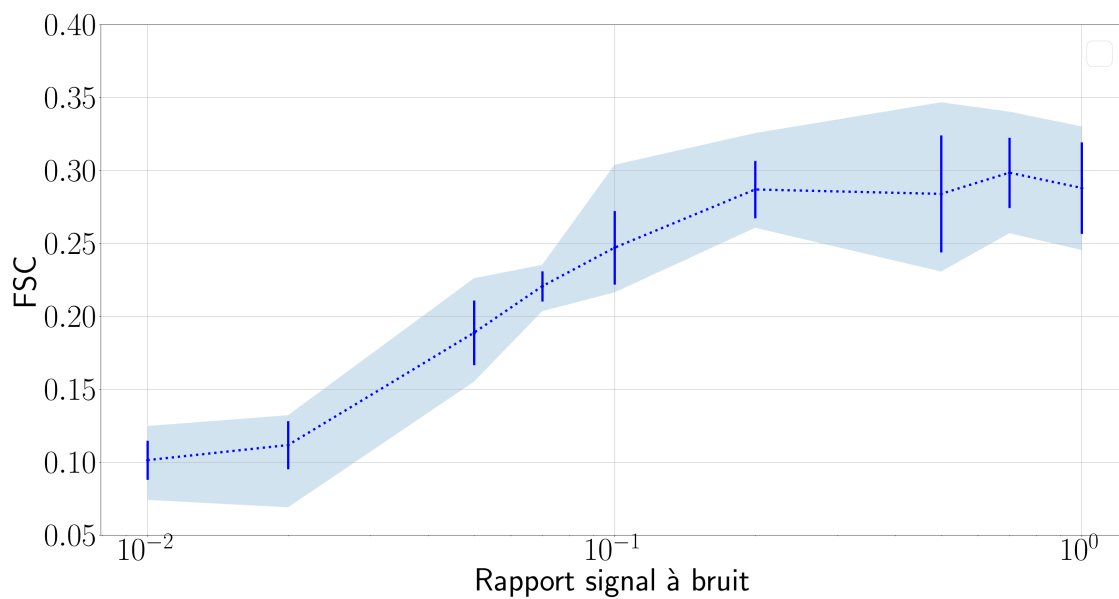
3.6.1.5 Temps de calcul

3.6.1.5.1 Méthode SHiReVol Le temps de calcul pour une époque de la méthode SHiReVol dépend de trois paramètres : le nombre de vues N , le nombre d’orientations échantillonnées $N_d N_\psi$ et le nombre de voxels $N_v = s^3$ (avec s la taille d’un côté du volume) dans le volume. Chaque époque contient autant d’itérations que le nombre de vues, donc le temps de calcul par époque est proportionnel à N . La grande majorité du temps de calcul est consacrée à l’évaluation des énergies $E_{i,j}^l$ correspondant aux rotations $\theta_{i,j}$ échantillonnées de manière aléatoire (de la ligne 11 à la ligne 13 de l’Algorithme 1). Ainsi, le temps de calcul par itération est proportionnel au nombre de rotations échantillonnées $N_d N_\psi$. Voici plus de précisions sur les lignes 11 à 13

- À la ligne 11, $\hat{f}$ est tourné d’une orientation $\theta_{i,j}$, puis une multiplication terme à terme est effectuée. Le temps de calcul dédiée à la multiplication terme à terme est négligeable devant le temps de la rotation. L’opération de rotation nécessite de calculer $\mathcal{R}(f)(x) =$



(a) Résultats visuels



(b) FSC à la vérité terrain. Chaque point du graphe est une moyenne effectuée sur 10 tests

FIGURE 3.14 – Influence du rapport signal à bruit sur la qualité de la reconstruction, méthode SHiReVol

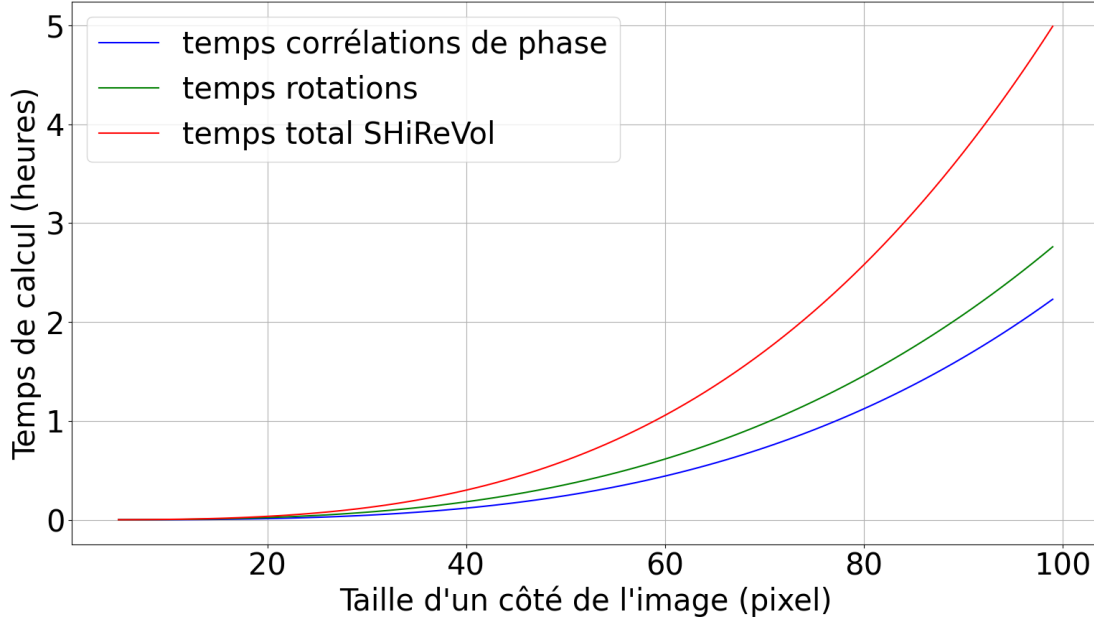


FIGURE 3.15 – Temps de calcul total de la méthode SHiReVol en fonction de la taille d'un côté de l'image ($s = N_v^{\frac{1}{3}}$). Il se décompose comme la somme du temps pour effectuer l'ensemble des rotations (courbe en vert) et le temps pour effectuer l'ensemble des corrélations de phase (courbe en bleue)

$f(R_{\theta_{i,j}}^t x)$ pour tout $x \in \Omega$, avec Ω le domaine de définition de f . Or, comme f est stockée informatiquement, Ω est en pratique un ensemble discret. Pour un x quelconque dans Ω , $R_{\theta_{i,j}}^t x$ n'appartient donc pas nécessairement à Ω , et donc le calcul de $f(R_{\theta_{i,j}}^t x)$ nécessite d'effectuer une interpolation. Cette interpolation est effectuée à l'aide de la fonction *affine transform* du module *scipy.ndimage*. Le temps de calcul de cette interpolation est de l'ordre de $O(N_v)$.

- À la ligne 12, la corrélation de phase est calculée entre $\hat{h}\mathcal{R}_{\theta_{i,j}}(\hat{f})$ et $\hat{y}_l$ pour estimer le vecteur de translation $t_l^*(\theta_{i,j})$. La corrélation de phase met en jeu trois grandes étapes. D'abord le calcul de la corrélation croisée normalisée dans le domaine de Fourier, qui met en jeu des multiplications et divisions termes à termes. Ensuite le calcul de la transformée inverse, et enfin la détection du pic de corrélation. Nous constatons expérimentalement que le temps de calcul est dominé par celui de la transformée de Fourier inverse, qui a une complexité en $O(N_v \log(N_v))$
- À la ligne 13, l'énergie $E_{i,j}^l$ est calculée étant donné le volume tourné $\mathcal{R}_{\theta_{i,j}}$ et le vecteur de translation $t_l^*(\theta_{i,j})$. Ce calcul met en jeu des multiplications et additions termes à terme, dont le temps de calcul est négligeable devant celui de la rotation et de la corrélation de phase

Globalement le temps de calcul consacré à une époque T_{ep} peut donc se mettre sous la forme :

$$T_{ep}^{vox} = NN_d N_\psi N_v (K_1 + L \log(N_v)) \quad (3.70)$$

Par ailleurs, on impose un nombre minimal d'époque $N_{min} = \lceil \frac{-\log(\epsilon_2)}{\log(\alpha_r)} \rceil$. Le temps de calcul

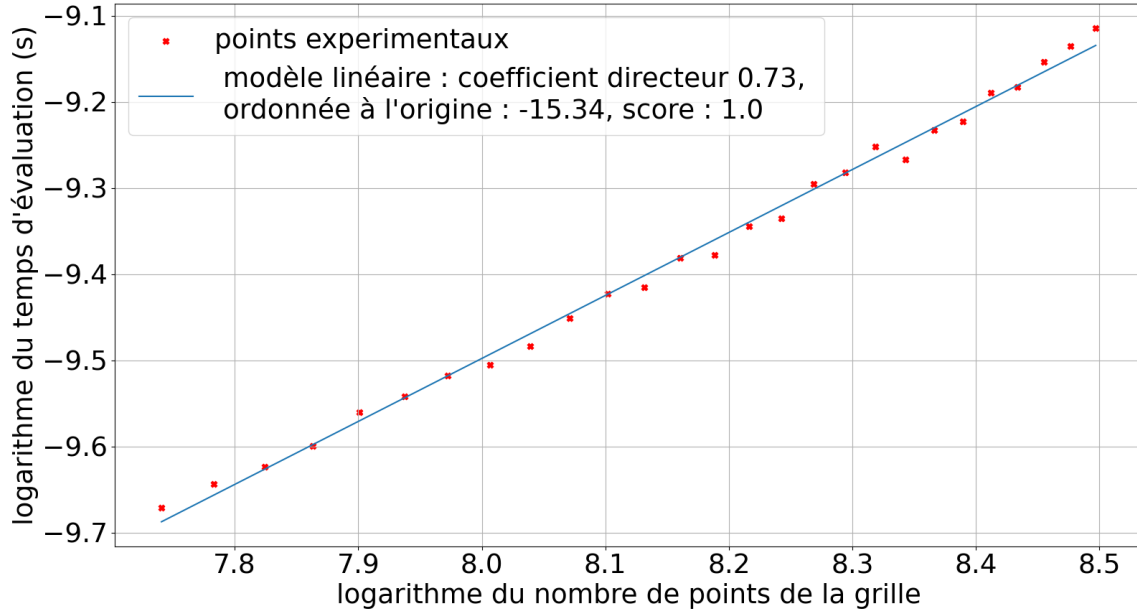


FIGURE 3.16 – Analyse du temps de calcul de l'évaluation d'une gaussienne sur une grille discrète en fonction du nombre de points de la grille. Pour chaque point expérimental, une moyenne du temps de calcul sur 50 expériences est effectuée

total minimal est donc :

$$T_{tot}^{vox} = N_{min} T_{ep}^{vox} = N_{min} N N_d N_\psi N_v (K_1 + L \log(N_v)) \quad (3.71)$$

Pour accélérer le calcul, nous utilisons une implémentation GPU de la rotation et de la corrélation de phase, disponibles dans les packages CuPy⁴ et cuCIM⁵ respectivement. Avec les paramètres par défaut $N_{min} = 52$, $N = 20$, $N_d = 64$, $N_\psi = 8$ et $N_v = 50^3$, une reconstruction complète prend environ 42 minutes sur une carte graphique Titan Xp. En comparaison, MP3DR prend 90 minutes pour une reconstruction avec $N = 5$ et le même $N_v = 50^3$.

Il est important de garder en tête que ce temps de calcul de 42 minutes est un cas particulier. Ce dernier dépend en effet de nombreux paramètres, en particulier la taille d'un côté du volume ($s = N_v^{\frac{1}{3}}$). Le temps de calcul sera d'autant plus élevé que la taille de l'image l'est. Il y a donc un compromis à trouver entre le temps de reconstruction et la qualité de la reconstruction. En effet, plus la grille contient de voxel, meilleure sera la résolution maximale que l'on peut atteindre. Sur la figure 3.15 nous présentons la dépendance du temps de calcul total de la méthode SHiReVol en fonction de s . Ce dernier se décompose comme la somme du temps pour effectuer l'ensemble des rotations et le temps pour effectuer l'ensemble des corrélations de phase (le temps de calcul des autres opérations est négligeable). Pour les autres hyper-paramètres (N , N_{min} , N_d , N_ψ) la dépendance du temps de calcul avec ces derniers est bien plus simple à envisager : il s'agit d'une simple relation de proportionnalité.

4. <https://cupy.dev/>

5. <https://pypi.org/project/cucim/>

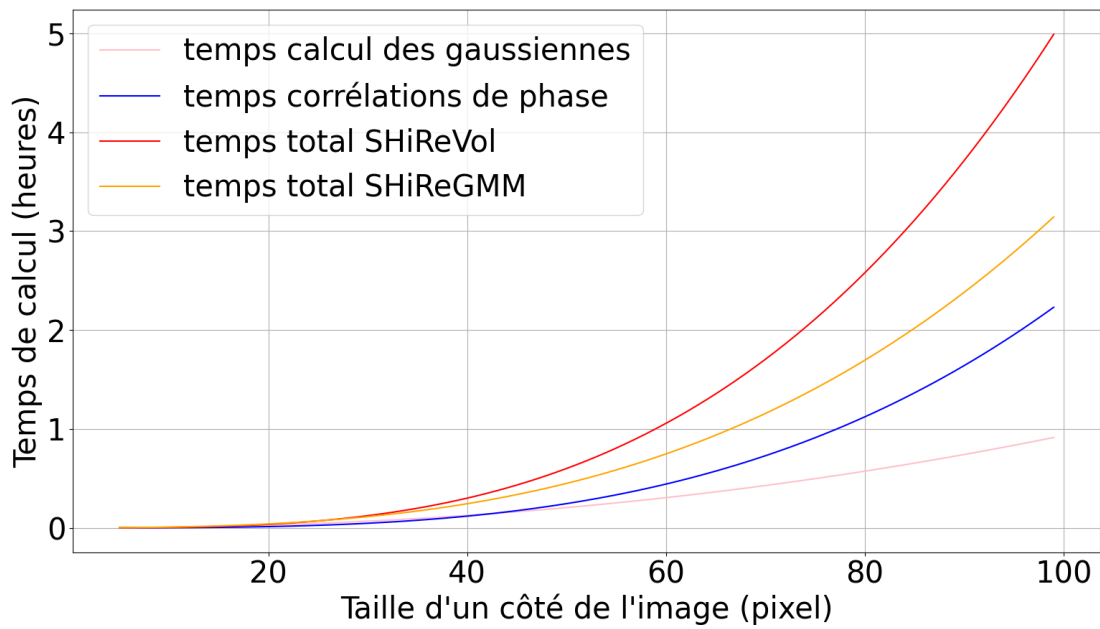


FIGURE 3.17 – Comparaison des temps de calcul totaux des méthodes SHiReVol (courbe en rouge) et ShiReGmm (courbe en orange) en fonction de la taille d'un côté de l'image ($s = N_v^{\frac{1}{3}}$). Le temps de la méthode ShiReGmm se décompose comme la somme du temps évaluer l'ensemble des mélanges de gaussiennes sur la grille 3D (courbe en rose) et le temps pour effectuer l'ensemble des corrélations de phase (courbe en bleue)

3.6.1.5.2 Méthode ShiReGmm L'analyse de la complexité de la méthode ShiReGmm diffère légèrement de celle de la méthode SHiReVol. La procédure générale reste la même, mais les deux algorithmes diffèrent dans la manière dont les énergies $E_{i,j}^l$ sont calculées (voir lignes 11 à 14 de l'algorithme 2).

- À la ligne 11, une rotation de matrice $R_{\theta_{i,j}}$ est appliquée à l'ensemble des centres $\{\mu_p\}_{p \in [1, N_g]}$. Nous avons remarqué empiriquement que le temps de calcul de ces rotations est négligeable devant le reste
- À la ligne 12, la représentation implicite du volume tourné et convolué est évaluée sur l'ensemble des points de la grille discrète Ω . Pour calculer la somme, nous effectuons une boucle sur l'ensemble des gaussiennes, le temps de calcul est donc proportionnel à N_g . Par ailleurs, pour réduire le temps de calcul, les gaussiennes sont tronquées, elles ne sont pas évaluées sur l'ensemble des points de la grille Ω mais sur un pavé droit dont les dimensions sont égales à 3 fois les écarts types de $\sigma^2 I + \Sigma_h$ dans les trois dimensions de l'espace. Par conséquent, le temps de calcul dépend de l'écart type σ des gaussiennes et des écart-types σ_z et σ_{xy} de la PSF. Sur la figure 3.16 nous avons tracé le temps d'évaluation d'une gaussienne en fonction du nombre de points n de la grille sur laquelle elle est évaluée. On remarque une dépendance en $T_{gaussian} = K_2 n^\zeta$ avec $\zeta = 0,73$. Dans notre cas, comme

$$\sigma^2 I + \Sigma_h = \begin{pmatrix} \sigma^2 + \sigma_{xy}^2 & 0 & 0 \\ 0 & \sigma^2 + \sigma_{xy}^2 & 0 \\ 0 & 0 & \sigma^2 + \sigma_z^2 \end{pmatrix}, \quad (3.72)$$

on a donc $n = 27(\sigma^2 + \sigma_{xy}^2)(\sigma^2 + \sigma_z^2)$.

- À la ligne 13 est effectué le calcul de la corrélation de phase entre $h * \mathcal{R}_{\theta_{i,j}}(f)$ et y_l .
- À la ligne 2, on doit à nouveau évaluer une combinaison linéaire de gaussiennes sur une grille discrète, le temps de calcul est donc en identique à celui de la ligne 12.

Le temps de calcul total de la méthode ShiReGmm peut donc se mettre sous la forme :

$$\begin{aligned} T_{tot}^{gmm} &= N_{min} N N_d N_\psi N (2 N_g T_{gaussian} + T_{corr}) \\ &= N_{min} N N_d N_\psi (2 K_2 N_g (27(\sigma^2 + \sigma_{xy}^2)(\sigma^2 + \sigma_z^2))^\zeta + L_2 N_v \log(N_v)). \end{aligned}$$

Avec les paramètres par défaut $N = 20$, $N_d = 64$, $N_\psi = 8$, $N_v = 50^3$, $\sigma^2 = 0,03s$, $\sigma_z^2 = 0.1s$ et $\sigma_{xy}^2 = 0.02s$ (avec s la taille d'un coté de l'image) et $N_g = 100$, une reconstruction complète prend environ 22 minutes sur une carte graphique Titan Xp. Là encore il s'agit d'un cas spécifique, le temps de calcul dépend de nombreux paramètres. Il est en particulier proportionnel au nombre de gaussiennes utilisées pour la reconstruction. Il y a donc, encore ici, un compromis à trouver entre le temps de calcul et la qualité de la reconstruction. Plus de gaussiennes signifie que ShiReGmm est en mesure de couvrir une plus grande part de la structure, mais cela augmente le temps de calcul à la même occasion. Tout comme SHiReVol, le temps de calcul de la méthode ShiReGmm est fortement dépendant de la taille d'un côté de l'image s . La figure 3.17 illustre l'évolution du temps de calcul de la méthode ShiReGmm en fonction de s . A titre de comparaison, nous traçons sur le même graphe le temps de calcul de la méthode SHiReVol. Le temps de calcul de ShiReGmm se décompose comme la somme du temps pour évaluer l'ensemble des mélanges de gaussiennes sur la grille 3D et le temps pour effectuer l'ensemble des corrélations de phase. Nous pouvons clairement remarquer que pour des tailles d'images relativement faibles, les temps de calcul des deux méthodes sont équivalents, mais que pour des valeurs de $s > 35$ le temps de calcul de SHiReVol dépasse de manière non négligeable celui de ShiReGmm.

3.6.2 Résultats sur des données réelles

Nous avons testé notre méthode SHiReVol sur des données réelles de centriole [28] acquises par le Centriole Lab⁶. Pour rappel (voir section 1.4.2), il s'agit d'un assemblage macromoléculaire indispensable à la division cellulaire et la polymérisation des microtubules de la cellule. Il est composé de neuf triplets de microtubules, organisés en cylindres. Les protéines qui composent les centrioles comprennent principalement les tubulines, qui sont les protéines constitutives des microtubules. Mais le centriole est composé de bien d'autres types de protéines, tel que la Cep135, impliquée dans la régulation de la longueur des centrioles, la Cep164 jouant un rôle de stabilisation des microtubules.

Comme les dimensions d'un centriole, autour de 250 nm de diamètre, sont de très peu supérieures à la limite de diffraction dans le visible (de 200 à 350 nm), ils ont été imagés à l'aide de la technique de microscopie par expansion U-ExM (pour ultrastructure expansion microscopy). Pour rappel (voir partie 1.3), l'U-ExM est une approche de super-résolution qui consiste à incorporer un échantillon biologique dans un polymère gonflable qui se dilatera 4 fois dans de l'eau pure, permettant d'atteindre une résolution de 50-70 nm avec un microscope à champ large conventionnel [104]. Après l'expansion, l'échantillon biologique est labellisé avec des marqueurs fluorescents dirigés contre Cep164, une protéine qui compose les appendices distaux du centriole [28], et qui apparaît avec 9 points formant un anneau admettant une symétrie cylindrique d'ordre 9 (C9). L'échantillon est imagé par un microscope à champ large Leica DM18. Les images 3D acquises ont une taille de pixel de 200 nm selon la dimension Z et de 100 nm dans le plan XY . Nous redimensionnons les images selon l'axe Z de sorte à ce que la taille des pixels selon l'axe Z correspondent à la taille des pixels dans le plan XY , à l'aide d'un algorithme d'interpolation d'ordre 3. Dans notre cas, cela revient à doubler le nombre de pixels selon Z . La PSF est mesurée expérimentalement en imageant une source ponctuelle. Nous sélectionnons manuellement 47 vues dans les volumes acquis.

Sur la figure 3.18, nous montrons 4 exemples de vues qui incluent 2 vues de dessus, une vue de côté et une vue inclinée. La résolution selon l'axe Z est clairement inférieure à la résolution dans le plan XY . De plus, la densité de labellisation non-uniforme est observable en comparant 2 vues de dessus. La symétrie C9 est également clairement visible sur la vue 3.18c.

Sur la figure 3.19, nous présentons la reconstruction obtenue avec cet ensemble de données et les mêmes hyper-paramètres que ceux présentés dans la Section 3.6.1.4.1. Nous comparons notre résultat avec celui de cryo-RANSAC. Afin d'obtenir un résultat de meilleure qualité, les reconstructions *ab initio* sont raffinées à l'aide d'un schéma d'optimisation alterné décrit dans [27]. En plus du raffinement de base, nous présentons également des reconstructions obtenues avec une contrainte de symétrie C9 imposée. Dans un souci de visualisation, les axes de symétrie des centrioles reconstruits sont alignés avec l'axe Z . L'axe de symétrie est déterminé en utilisant une ACP sur une version binarisée de la reconstruction.

La forme reconstruction obtenue par notre méthode est un anneau qui est cohérent avec les vues d'entrée. En effet, le diamètre de l'anneau observable sur le plan XY est similaire au diamètre des vues de dessus 3.18b et 3.18c. De plus, l'épaisseur de l'anneau, observable sur les coupes XZ et ZY , est similaire à ce que nous observons sur la vue de côté 3.18a. La symétrie C9 est floue dans la reconstruction *ab-initio*, mais suffisante pour permettre une symétrie clairement visible dans le volume raffiné. Si nous imposons la symétrie C9 dans le raffinement, l'anneau présente une symétrie parfaitement régulière. La reconstruction cryo-RANSAC n'est pas capable de récupérer

6. <https://mocel.unige.ch/research-groups/guichard-hamel/overview>

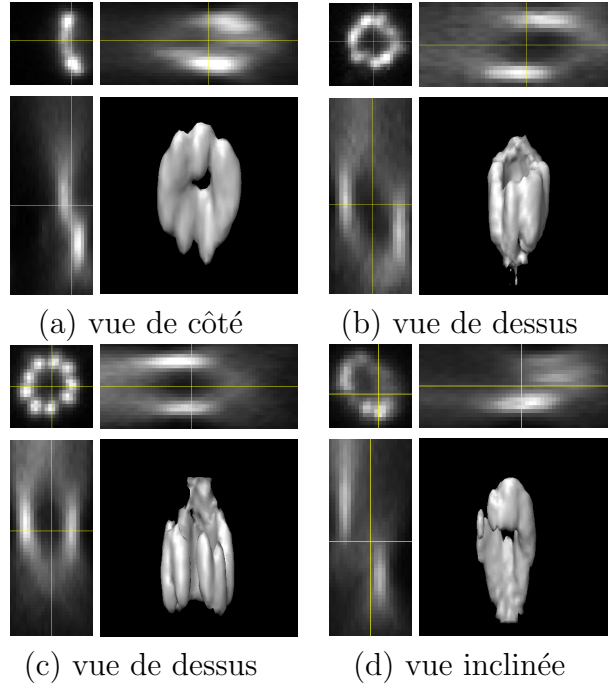


FIGURE 3.18 – Exemples de vues issues de notre jeu de données réelles, après redimensionnement selon l’axe Z de sorte à ce que les tailles de pixels en Z et en XY soient identiques. La taille d’un côté de la coupe XY (dans le coin en haut à gauche de chaque vue) est de 1150 nm, ce qui correspond à une taille avant expansion de $\frac{1150}{4} = 288$ nm.

les mêmes détails, car la forme de l’anneau n’est pas correctement reconstruite et la symétrie C_9 n’est pas visible.

3.7 Résumé du chapitre et perspectives

3.7.1 Résumé

Nous avons présenté dans ce chapitre une méthode de RPI homogène, adaptée au modèle d’acquisition de la microscopie par fluorescence. Cette dernière se base sur une reformulation hiérarchique du problème d’optimisation joint original, chaque niveau étant associé à un groupe de variables (les paramètres du volume pour le premier niveau, de rotations pour le deuxième et de translations pour le troisième). Chaque problème, à un niveau donné, est résolu à chaque itération de la résolution du problème au niveau supérieur. Nous estimons les paramètres du volume à l’aide d’un algorithme de descente de gradient stochastique. Les paramètres de rotations sont estimés par une recherche exhaustive sur un sous-ensemble de $SO(3)$ sélectionné judicieusement par une méthode échantillonnage préférentiel. Enfin, les paramètres de translations sont estimés à l’aide d’un algorithme de corrélation de phase.

Nous avons d’autre part développé 2 variantes de la méthode originale (appelée SHiReVol). La première, appelée ShiReGmm, utilise une représentation implicite du volume sous forme d’une combinaison linéaire de gaussiennes anisotropes. La seconde, appelée ShiReGmm2, utilise une représentation du volume et des vues sous forme d’un mélange de gaussiennes.

Nous avons évalué les performances des 3 méthodes SHiReVol, ShiReGmm et ShiReGmm2

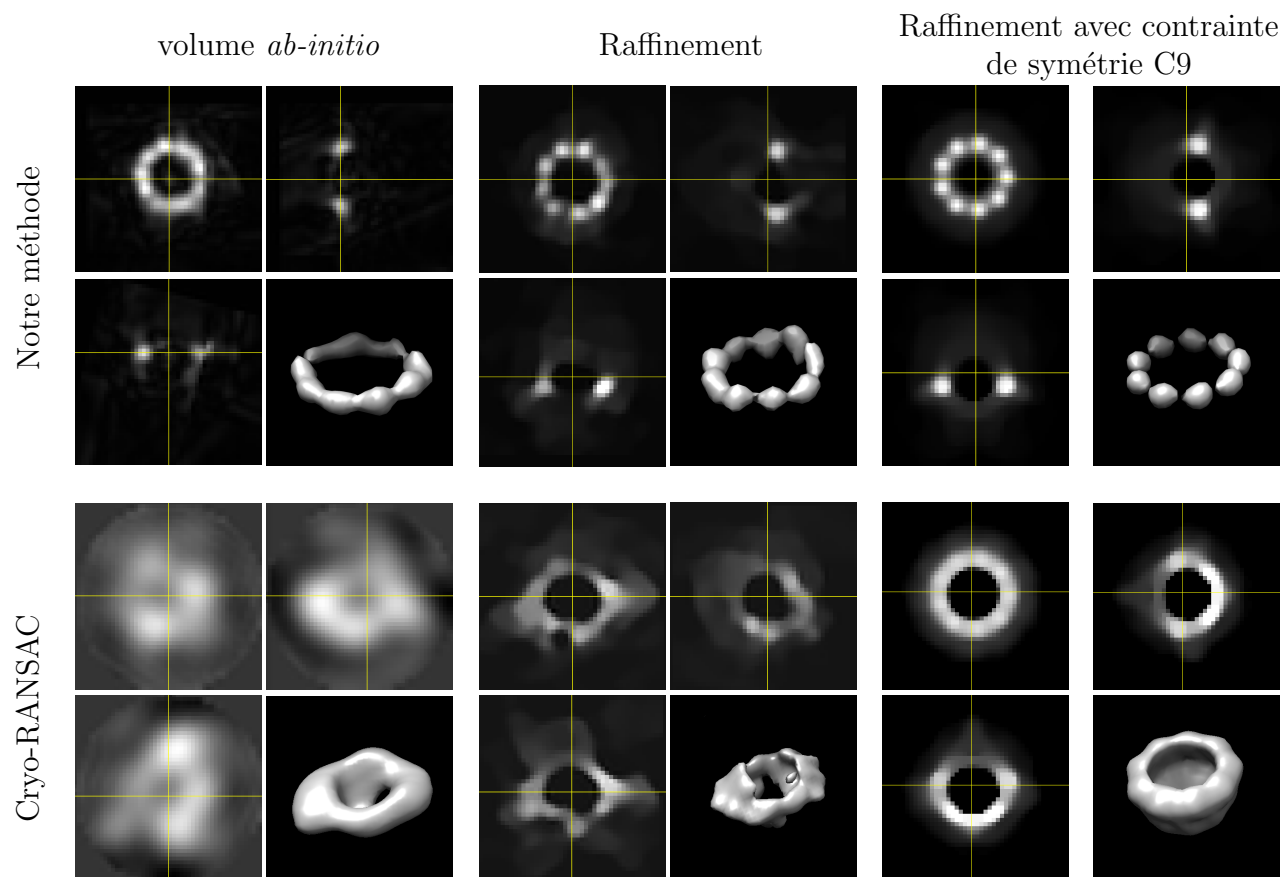


FIGURE 3.19 – Résultats de reconstruction obtenus sur un jeu de données réelles, à l'aide de notre méthode SHiReVol et de cryo-RANSAC.

sur des jeux de données générées synthétiquement à partir d’objets vérité terrain, et les avons comparés avec l’état de l’art de la RPI en microscopie par fluorescence (la méthode MP3DR [27]). Ces résultats nous ont permis d’écarter ShiReGmm2, qui témoigne de significativement moins bons résultats que les autres (en termes de SSIM, de FSC, et d’évaluation visuelle qualitative). La représentation des vues sous forme de mélange de gaussiennes induit en effet une perte d’information trop importante. Les deux méthodes SHiReVol et ShiReGmm parviennent à surpasser MP3DR. La méthode SHiReVol permet globalement d’atteindre des résultats plus précis que ShiReGmm, en raison de la granularité introduite par le modèle sous forme de mélange de gaussiennes. Nous avons également mis en évidence l’isotropie de résolution des reconstructions obtenues par SHiReVol et ShiReGmm grâce à la métrique de la Corrélacion de Fourier conique. La résolution est non seulement améliorée dans la direction axiale, mais également dans le plan latéral. Notre méthode témoigne par ailleurs d’une forte robustesse au bruit et à de faibles degrés de labellisation.

Nous avons également évalué les performances de SHiReVol sur un jeu de données réelles : des images de la protéine Cep164 du centriole. Les résultats obtenus sont visuellement cohérents avec les données d’entrée. Néanmoins, nous n’avons pas évalué quantitativement les performances sur ce jeu de données réelles. Afin d’évaluer les performances d’une méthode de RPI sur un jeu de données réelles, pour lequel nous n’avons pas d’objet vérité terrain à disposition, la pratique courante est de séparer les données en deux sous-ensembles disjoints. Deux reconstructions sont effectuées séparément sur ces deux sous-ensembles et elles sont comparées à l’aide de la métrique FSC. Néanmoins, nous n’avons pas mis en place cette procédure, et nous nous sommes contenté d’appréciations visuelles. En effet, notre jeu de données était constitué de seulement 47 vues, ce qui est un nombre trop faible pour que deux sous-ensembles sélectionnés au hasard soit statistiquement représentatifs de l’ensemble des vues. Par ailleurs, une seule protéine (la Cep164) d’un seul assemblage macromoléculaire (le centriole) a été testé. Une validation plus en profondeur de notre méthode pourrait donc faire l’objet de travaux ultérieurs, mais nous aurions pour cela besoin de disposer d’une quantité plus importante de données.

3.7.2 Perspectives méthodologiques

3.7.2.1 Utilisation d’une marginalisation par rapport aux orientations plutôt que de l’optimisation hiérarchique

Notre approche repose sur une optimisation hiérarchique à 3 niveaux. Dans ce cadre, une pose unique est estimée pour chacune des vues. Une alternative présentant le même coût calculatoire serait de marginaliser la fonction de coût par rapport aux orientations, de la même manière que ce qui est réalisé dans la méthode CryoSPARC. Néanmoins, cette marginalisation se ferait uniquement par rapport aux orientations, nous conserverions l’estimation des translations par corrélation de phase. En effet, une marginalisation par rapport aux deux groupes de variables serait trop coûteuse en temps de calcul dans le cas tridimensionnel de la microscopie par fluorescence. Elle serait par ailleurs inutile puisque la corrélation de phase fournit déjà une estimation très précise du paramètre de translation.

3.7.2.2 Contraintes stéréochimiques sur la représentation sous forme de mélange de gaussiennes

Un avantage de la représentation sous forme de mélange de gaussiennes est de pouvoir modéliser des contraintes stéréochimiques entre protéines voisines de la structure. En effet, si l’écart-type

des gaussiennes est choisis judicieusement, on peut considérer que chaque gaussienne représente une protéine particulière. Nous pourrions alors développer une variante de la méthode permettant d'imposer que la distance entre deux gaussiennes voisines soit proche d'une certaine valeur, qui correspond approximativement à la longueur supposée de la liaison non-covalent entre ces deux protéines.

Au delà de ces perspectives méthodologiques spécifiques à ce chapitre, nous présentons dans le chapitre 5 des perspectives plus générales pouvant être communes aux deux cas d'applications que nous avons abordés durant cette thèse (reconstruction homogène et reconstruction hétérogène).

Chapitre 4

Reconstruction hétérogène

La méthode présentée dans le chapitre précédent est dédiée à la reconstruction homogène. Toutes les vues du jeu de données sont supposées être issues du même volume, que l'on cherche à reconstruire. Nous proposons dans ce chapitre une méthode de reconstruction hétérogène continue pour la RPI en microscopie par fluorescence : les vues ne sont plus supposées provenir d'un volume unique mais d'un ensemble continu de volumes. La reconstruction consiste alors non seulement à estimer les poses de chacune des vues mais également les différentes conformations desquelles elles sont issues.

4.1 Positionnement et contributions

Bien qu'il n'existe pas de méthode de reconstruction hétérogène dédiée à la RPI en microscopie par fluorescence, nous avons vu dans les sections 2.2.4 et 2.2.5 qu'il existe une grande variété de méthodes de reconstruction hétérogène en cryo-ME, nous nous inspirons donc de l'état de l'art en cryo-ME. Nous avons aussi vu que certaines méthodes étaient plus adaptées à traiter un type d'hétérogénéité que d'autres. Par exemple, si l'on sait que l'assemblage étudié existe sous un nombre discret d'états stables, relativement éloignés les uns des autres, une méthode d'espérance maximisation par rapport à k volumes sera plus adaptée (hétérogénéité discrète). Si l'hétérogénéité est continue, il existe encore un large panel de méthodes pouvant être plus ou moins adaptées en fonction des cas. En particulier, certaines méthodes telles que 3DFlex [81] ou Dynamight [82] se prêtent bien à l'estimation d'une hétérogénéité non-compositionnelle. Ce type d'hétérogénéité suppose que les conformations sont obtenues à partir de la déformation d'une structure canonique unique. Dans ce cas, l'apprentissage consiste à estimer, pour chaque conformation, le champ de déformation qui la relie à la structure canonique.

4.1.1 Cas d'application : hétérogénéité compositionnelle

Les méthodes de reconstruction hétérogène discrète ou continue non-compositionnelle ne sont pas adaptées à notre cas d'application, c'est-à-dire la croissance du procentreole lors de l'interphase du cycle cellulaire (voir section 1.4.2). En effet, le procentreole existe sous un nombre continu d'états correspondant à différents états d'avancement de l'interphase (il sera d'autant plus grand que cette dernière est avancée). D'autre part, la composition du procentreole varie au cours de sa croissance : plus il est grand, plus il contient de matière. Nous devons donc choisir une méthode de reconstruction adaptée à l'hétérogénéité compositionnelle.

L'état de l'art traitant ce type d'hétérogénéité en cryo-ME sont les méthodes CryoDRGN [65] et CryoFIRE [75], qui ont été présentées au paragraphe 2.2.5.2.1. Pour rappel ces dernières utilisent une représentation implicite du volume à l'aide d'un réseau de neurones, qui prend en entrée les coordonnées d'un voxel x et un paramètre z de conformation et renvoie l'intensité du pixel correspondant à la conformation z au voxel x . Cette représentation se prête bien à notre cas, car le choix de la dimension de z permet d'imposer le nombre de degré de liberté de la reconstruction. Or, dans le cas particulier de la croissance du procentre, nous pouvant faire l'hypothèse qu'il y a 1 degré de liberté, correspondant à l'état d'avancement de la croissance (directement lié à la longueur du procentre). La différence majeure entre CryoDRGN et CryoFIRE est la manière dont les poses sont estimées. Dans CryoDRGN, les poses sont estimées de manière indépendante pour chacune des vues, alors que CryoFIRE se base sur une estimation amortie des poses.

4.1.2 Adaptation de CryoFIRE

Néanmoins les méthodes CryoDRGN et CryoFIRE ne peuvent pas directement être appliquées à la RPI en microscopie par fluorescence car elles ont été conçues pour le modèle d'acquisition de la cryo-ME, différent du modèle de microscopie par fluorescence. Nous avons donc adapté CryoFIRE au modèle de microscopie par fluorescence.

Nous avons choisi d'adapter CryoFIRE plutôt que CryoDRGN car cette dernière est moins coûteuse en temps de calcul. En effet, la méthode CryoDRGN nécessite d'effectuer une recherche exhaustive sur les poses à chaque itération de l'apprentissage, et pour chaque vue, afin d'identifier les poses optimales. En revanche, CryoFIRE utilise un réseau de neurones faisant correspondre la vue d'entrée à la pose qui lui est associée. L'apprentissage des poses n'est donc pas effectué indépendamment pour chacune des vues, ce qui se traduit par un gain computationnel considérable. Ceci est particulièrement important dans le cas de la fluorescence, car nous traitons des données d'entrée 3D.

De manière similaire à CryoFIRE, notre modèle d'apprentissage prend la forme d'un système encodeur-décodeur entraîné de bout en bout. L'encodeur permet d'estimer les poses associées à une vue d'entrée, et le décodeur est principalement constitué de la représentation implicite du volume. Notre méthode, que nous nommons Fluofire, sera présentée en détail dans la section 4.2.

En plus de cette adaptation au modèle de fluorescence, nous intégrons trois extensions du modèle, résumées dans les prochaines sous-sections.

4.1.3 Apprentissage pose à pose

L'inférence amortie des poses permet d'obtenir une reconstruction en un temps considérablement inférieur à un algorithme utilisant une inférence non-amortie. Néanmoins, ce gain s'obtient au prix d'une réduction de la précision de l'estimation des poses (et donc de la reconstruction). Afin d'améliorer la robustesse de l'estimation des poses, nous ajoutons au processus d'apprentissage des phases d'apprentissage pose à pose durant lesquelles l'ordre de l'encodeur et du décodeur est inversé. Nous présenterons cette méthodologie, inspirée de [76] en partie 4.2.3.

4.1.4 Information *a-priori* sur la symétrie de la reconstruction

Nous savons que le centriole admet une symétrie C9 : il prend la forme d'un tore invariant par rotation d'angle $\frac{2k\pi}{9}$ (avec k entier) autour de son axe. Nous avons donc ajouté à notre algorithme

la possibilité de prendre en compte cette information *a-priori*. Nous présenterons cette option en section 4.2.4.

4.1.5 Reconstruction d'un modèle paramétrique

Nous avons inclus la possibilité d'utiliser une représentation paramétrique du volume, plutôt que d'une représentation implicite sous la forme d'un réseau de neurones. L'avantage est de pouvoir inclure une contrainte *a-priori* forte à la forme de la reconstruction.

La représentation paramétrique que nous avons implémenté est celle d'un tore dont on cherche à estimer la longueur, la largeur et le rayon. Lorsqu'il y a deux canaux, nous estimons ces 3 paramètres pour les deux canaux ainsi qu'un paramètre de position du second canal par rapport au premier. Cette forme est en effet adaptée à notre cas d'application : le centriole. Cette estimation d'un modèle paramétrique sera présentée en partie 4.5.

4.2 Méthode

Notre méthode de reconstruction Fluofire est un système encodeur-décodeur entraîné de bout en bout. L'encodeur est un réseau de neurones qui estime de manière amortie les poses et le paramètre de conformation associé à une vue d'entrée. Le décodeur est un réseau de neurones densément connecté qui représente implicitement le volume à reconstruire. Avant de rentrer dans les détails de l'architecture, nous commençons par formuler le problème d'optimisation de la reconstruction hétérogène, c'est l'objet de la section suivante.

4.2.1 Formulation du problème

Nous modélisons l'hétérogénéité des conformations par une variable de conformation z contenue dans un espace à faible dimension K ($z \in \mathbb{R}^K$). Nous notons f le volume à reconstruire :

$$f : \mathbb{R}^3 \times \mathbb{R}^K \rightarrow \mathbb{R}$$

$$(x, z) \mapsto f(x, z),$$

$f(x, z)$ est donc l'intensité de la conformation de paramètre z au point x . Le modèle d'acquisition s'écrit donc, en reprenant les mêmes notations qu'en Section 3.2,

$$y_l(x) = h * \mathcal{T}_{t_l}(\mathcal{R}_{\theta_l}(f(\cdot, z_l)))(x) + \epsilon(x) \quad (4.1)$$

$$= h * f(R_{\theta_l}^t(x - t_l), z_l) + \epsilon(x). \quad (4.2)$$

La reconstruction hétérogène consiste donc à résoudre le problème d'optimisation suivant :

$$f^*, \Theta^*, T^*, Z^* = \arg \min_{f, \Theta, T, Z} E(f, \Theta, T, Z), \quad (4.3)$$

où

$$E(f, \Theta, T, Z) = \sum_{l=1}^N \sum_{x \in \mathcal{G}} [y_l(x) - h * f(R_{\theta_l}^t(x - t_l), z_l)]^2 + \alpha \|z_l\|^2, \quad (4.4)$$

avec

$$\Theta = \{\theta_l\}_{l \in [1, N]}, T = \{t_l\}_{l \in [1, N]} \text{ et } Z = \{z_l\}_{l \in [1, N]}.$$

et $\mathcal{G}$ une grille 3D de pas fixé au début de l'apprentissage.

L'objet de la section suivante est de détailler les choix (déjà évoqués en introduction de partie) que nous avons effectués pour résoudre le problème d'optimisation joint (4.3).

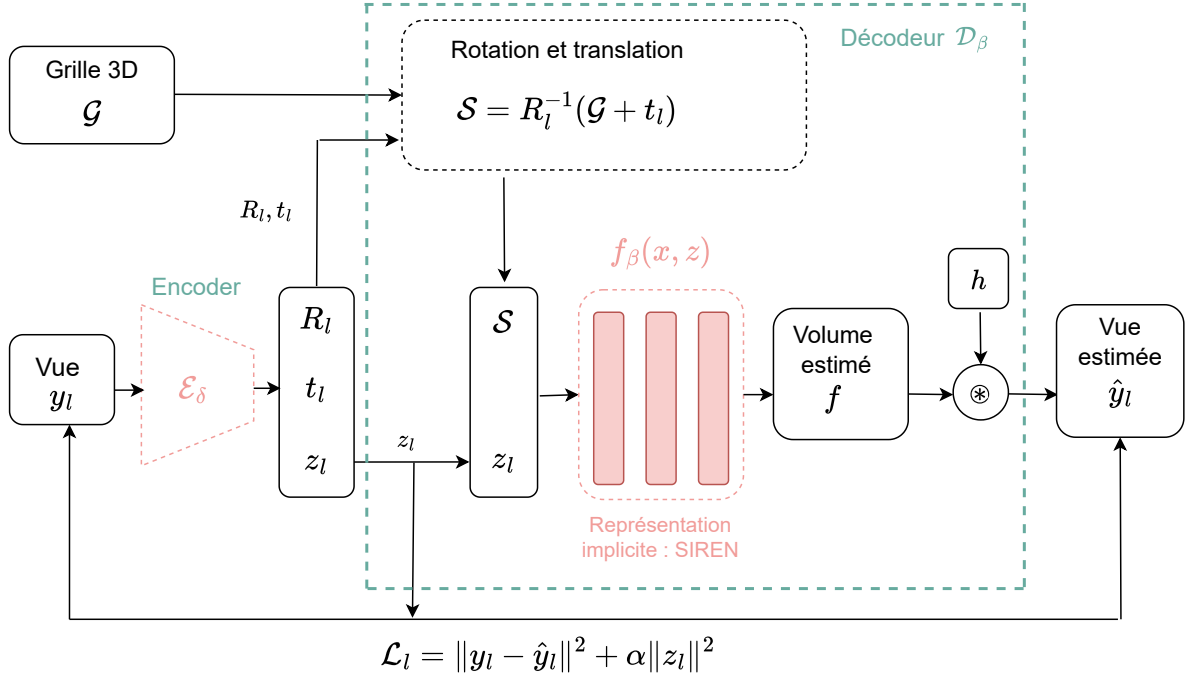


FIGURE 4.1 – Architecture de la méthode Fluofire. Les paramètres appris sont illustrés en rose.

4.2.2 Modèle

Le système d'apprentissage Fluofire est illustrée sur le schéma-bloc 4.1. Nous détaillons dans ce qui suit ces différentes composantes.

4.2.2.1 Représentation implicite du volume

Nous utilisons une représentation implicite du volume à reconstruire : nous le modélisons à l'aide d'un réseau de neurones à représentation sinusoïdale (SIREN [67]) f_β de paramètres β . Il s'agit d'un réseau de neurones densément connecté, admettant des fonctions d'activation sinusoïdales au lieu des fonctions ReLU habituellement utilisées dans les perceptrons multicouches (MLP). Le SIREN prend en entrée les coordonnées $x \in \mathbb{R}^3$ et le paramètre de conformation $z \in \mathbb{R}^K$ et renvoie $f_\beta(x, z)$ l'intensité de la conformation z au point x . Il admet donc $3 + K$ neurones d'entrée et 1 neurone de sortie. Nous utilisons par ailleurs 3 couches cachées admettant chacune 256 neurones. Au cours de l'apprentissage, les paramètres β du réseau (*i.e.* les poids et biais) sont progressivement ajustés de sorte à faire correspondre le mieux possible le volume aux données. Comme évoqué en section 2.2.3.2, cette représentation implicite peut potentiellement atteindre de meilleures résolutions qu'une représentation explicite, car n'importe quelle coordonnée x peut-être mise en entrée du réseau f_β . Nous avons constaté expérimentalement que l'utilisation d'un réseau de neurones admettant des fonctions d'activation ReLU plutôt que des sinusoides conduisait à des résultats aberrants. La figure 4.2 illustre la représentation implicite qui vient d'être décrite.

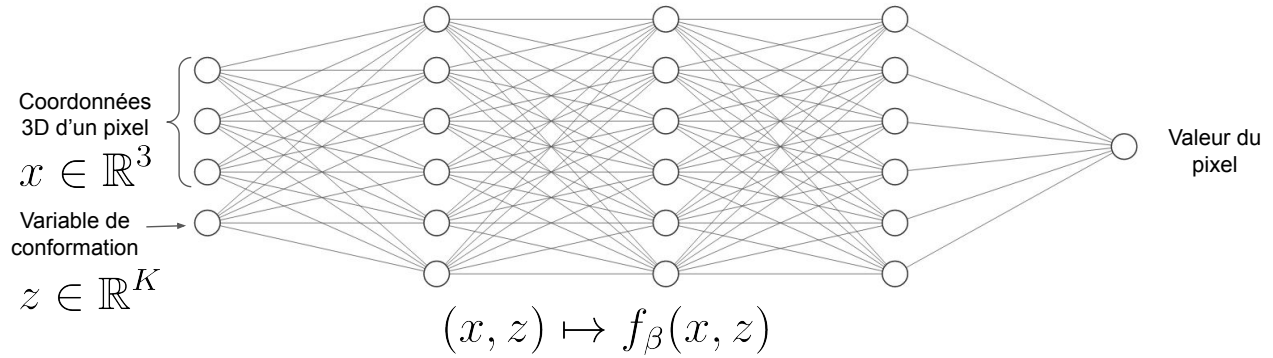


FIGURE 4.2 – Représentation implicite du volume sous forme d'un SIREN

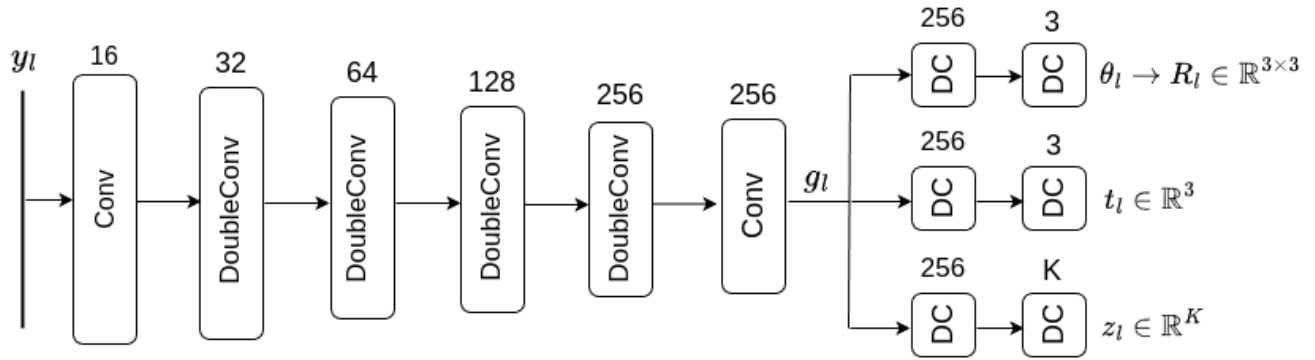


FIGURE 4.3 – Architecture de l'encodeur. Les couches notées Conv sont des couches convolutives 3D admettant un pas de 2 et dont le nombre de filtres est indiqué au dessus. Les couches notées DoubleConv sont des blocs de deux couches convolutives, la première admettant un pas de 1 et la seconde un pas de 2. Les couches notées DC sont des couches densément connectées dont le nombre de neurones est précisé au dessus. Chaque couche est séparée de la suivante par une fonction d'activation ReLU

4.2.2.2 Estimation amortie des poses et du paramètre de conformation

Nous choisissons d'autre part d'utiliser une estimation amortie des poses et du paramètre de conformation. Pour rappel, cela signifie qu'au lieu d'estimer les poses - et le paramètre de conformation - associés à chacune des vues indépendamment, on estime une fonction prenant en entrée une vue, et qui renvoie en sortie les poses - et le paramètre de conformation. Dans notre cas, la fonction estimée est un réseau de neurones. Plus précisément, notre réseau de neurones est constitué d'un bloc commun aux paramètres de rotation, translation et de conformation prenant la forme d'un réseau convolutif, et qui renvoie un vecteur de caractéristiques g_l . Ce dernier est mis en entrée de trois MLPs, prédisant respectivement les rotations, les translations et le paramètre de conformation associés à la vue d'entrée. Nous avons donc :

$$g_l = \mathcal{E}_{\delta_0}^0(y_l) \quad (4.5)$$

$$R_l = \mathcal{E}_{\delta_1}^1(g_l) \quad (4.6)$$

$$t_l = \mathcal{E}_{\delta_2}^2(g_l) \quad (4.7)$$

$$z_l = \mathcal{E}_{\delta_3}^3(g_l), \quad (4.8)$$

avec $\mathcal{E}_{\delta_0}^0$ le réseau de neurones convolutif qui extrait les caractéristiques g_l et $\mathcal{E}_{\delta_1}^1$, $\mathcal{E}_{\delta_2}^2$, $\mathcal{E}_{\delta_3}^3$ les MLPs qui prédisent respectivement les rotations, translations et paramètres de conformations. $\delta = \{\delta_0, \delta_1, \delta_2, \delta_3\}$ sont les paramètres des quatre réseaux de neurones qui sont ajustés au cours de l'apprentissage. Nous notons R_l au lieu de R_{θ_l} pour plus de légèreté. La figure 4.3 détaille l'architecture de l'encodeur.

4.2.2.3 Apprentissage de bout en bout

On peut alors reparamétriser le problème de minimisation (4.3) avec les paramètres β et δ des réseaux de neurones utilisés respectivement pour la représentation implicite du volume et pour l'estimation des poses et paramètres de conformations :

$$\beta^*, \delta^* = \arg \min_{\beta, \delta} E(\beta, \delta), \quad (4.9)$$

avec

$$E(\beta, \delta) = \sum_{l=1}^N \|y_l - \Gamma_{\beta, \delta}(y_l)\|^2 + \alpha \|\mathcal{E}_{\delta_3}^3(g_l)\|^2, \quad (4.10)$$

où $\Gamma_{\beta, \delta}$ représente l'architecture complète, c'est-à-dire le réseau de neurones qui estime les poses suivi de la représentation implicite sous forme de SIREN. Formellement :

$$\Gamma_{\beta, \delta}(y_l)(x) = h * f_{\beta}(\mathcal{E}_{\delta_1}^1(g_l)^t(x - \mathcal{E}_{\delta_2}^2(g_l)), \mathcal{E}_{\delta_3}^3(g_l)). \quad (4.11)$$

Il s'agit d'un système encodeur-décodeur, que l'on entraîne de bout en bout. L'encodeur $\mathcal{E}_{\delta}$ prédit à partir d'une vue y_l les poses, R_l et t_l , et le paramètre de conformation z_l associés. Ces dernières sont mises en entrée du décodeur $\mathcal{D}_{\beta}$ constitué de 3 blocs. Le premier consiste à utiliser les poses pour tourner et traduire une grille 3D $\mathcal{G}$. Ensuite, la grille et le paramètre de conformation z_l sont concaténés et mis en entrée de la représentation implicite f_{β} du volume. Enfin une convolution par la réponse impulsionnelle h est appliquée à la sortie de la représentation implicite pour obtenir

la vue estimée $\hat{y}_l$. Nous pouvons donc réécrire l'équation (4.10) sous la forme :

$$E(\beta, \delta) = \sum_{l=1}^N \|y_l - \hat{y}_l\|^2 + \alpha \|z_l\|^2, \quad (4.12)$$

avec

$$\hat{y}_l = (\mathcal{D}_\beta \circ \mathcal{E}_\delta)(y_l) \quad (4.13)$$

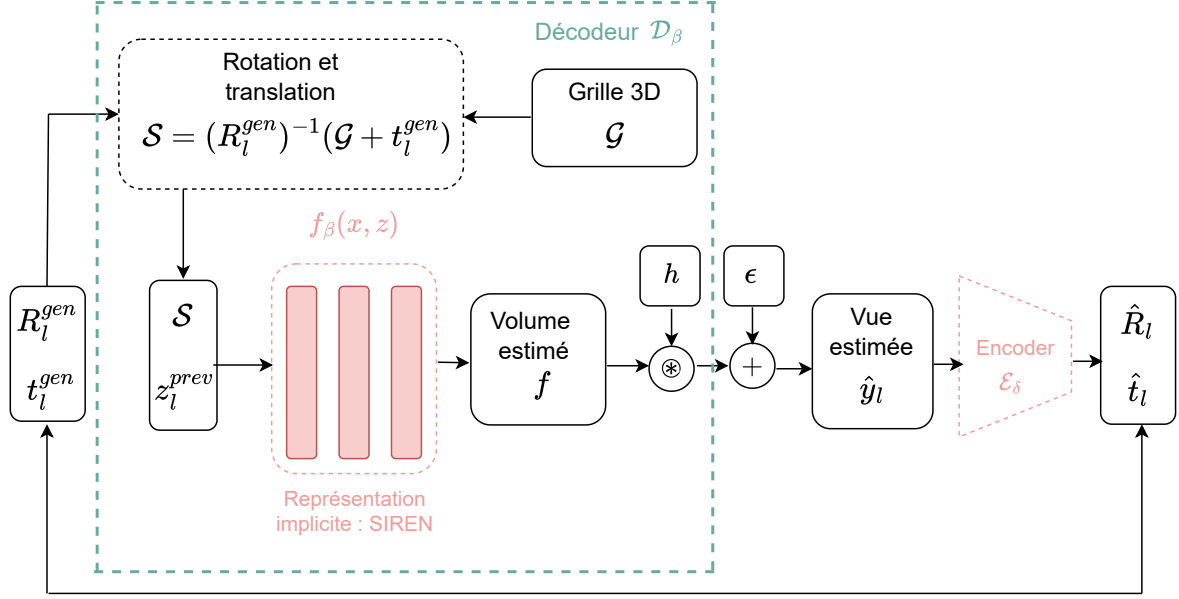
L'apprentissage de bout en bout constitue un système d'apprentissage non supervisé. Aucun préapprentissage n'est effectué : la reconstruction se fait uniquement à partir des vues à résolution anisotrope d'entrée, ce qui signifie que pour chaque nouveau jeu de données, l'encodeur et le décodeur sont ré-entraînés à partir d'initialisations aléatoires. En particulier, aucune supervision sur les poses n'est utilisée.

L'architecture présentée dans cette section précédente est, comme nous l'avons déjà mentionné en introduction de ce chapitre, inspirée de la méthode CryoFIRE. Nous l'avons néanmoins adaptée au modèle d'acquisition en fluorescence et cela pose une difficulté principale : les éventuels problèmes de mémoire de carte graphique. En effet, dans la cas de la cryo-ME le théorème de la tranche centrale de Fourier peut être mis à profit pour effectuer les étapes de rotations du volume et de projection en une seule fois, en passant à travers le décodeur f_β une grille 2D qui a été préalablement tournée et translatée. En revanche, dans le cas de la fluorescence une grille 3D doit être passée à travers le décodeur, et il y aura autant d'activations du SIREN stockées en mémoire que de points sur la grille. Cela peut donc causer des problèmes de mémoire vive de carte graphique, notamment si la taille de batch est trop importante. Nous avons constaté qu'avec une taille de batch de 4, la taille de grille limite pouvant être utilisée est de $60 \times 60 \times 60$ pour une carte graphique de 12Go.

4.2.3 Apprentissage pose à pose

Nous avons remarqué qu'en se contentant d'utiliser l'architecture de Fluofire 4.1, l'optimisation conduisait dans certains cas à l'obtention d'un minimum local largement sous optimal. Les auteurs de CryoFIRE ont également constaté ce problème dans le cas de la cryo-ME. Pour le résoudre, ils ont implémenté ce qu'ils appellent une "fonction de coût symétrique". Pour une vue d'entrée donnée, cette dernière est calculée (dans le cas de la cryo-ME) comme le minimum entre les fonctions de coût respectivement associées à la vue d'entrée et à cette même vue rotationnée de 180° . Dans un premier temps, nous avons adapté la fonction de coût symétrique au cas de la microscopie par fluorescence. Nous l'avons alors définie comme le minimum des fonctions de coût associées à 4 images d'entrée : la vue d'entrée, et les 3 vues obtenues par rotation de 180° autour des 3 axes X , Y et Z . Si cette approche nous a effectivement permis de résoudre le problème de convergence vers de mauvais minima locaux, elle présente néanmoins un inconvénient majeur : le temps de calcul et l'espace mémoire se trouvent être multipliés par 4.

C'est pourquoi, nous nous sommes, plus tard dans la thèse, inspirés d'une autre approche proposée par Chen *et al.* en Avril 2024 [76] et appelée "apprentissage pose à pose". Cette méthode permet à l'algorithme d'explorer un ensemble bien plus large de poses en couvrant la sphère unité de manière presque exhaustive. La procédure de reconstruction alterne entre deux phases : une phase d'apprentissage image à image similaire à celle exposée en section 4.2.2, et une phase d'apprentissage pose à pose. Cette dernière utilise les mêmes encodeurs et décodeurs que l'architecture image à image de la figure 4.1 mais dans un ordre inversé. Des poses R_l^{gen} et t_l^{gen} sont générées aléatoirement, selon une distribution uniforme sur la discrétisation de Fibonacci pour R_l^{gen} et selon



$$\mathcal{L}_{pose} = \frac{1}{9} \|R_l^{gen} - \hat{R}_l\|_F + \frac{1}{2} \|t_l^{gen} - \hat{t}_l\|_1$$

FIGURE 4.4 – Architecture de l’apprentissage pose à pose. L’encodeur et le décodeur sont identiques à l’architecture de base mais utilisés dans un ordre inversé.

une distribution normale centrée en 0 pour t_l^{gen} . Ces poses, ainsi que le paramètre de conformation z_l^{prev} de l’itération précédente sont mises en entrée du décodeur. Le résultat du décodeur, auquel est ajouté un bruit blanc gaussien est mis en entrée de l’encodeur pour prédire des poses $\hat{R}_l$ et $\hat{t}_l$ censées être proches des poses d’entrée. Une fonction de coût $\mathcal{L}_{pose}$ est utilisée pour évaluer cette proximité, et son gradient est rétro-propagé pour ajuster les poids de l’encodeur (les paramètres du décodeurs sont gelés). $\mathcal{L}_{pose}$ est définie par :

$$\mathcal{L}_{pose} = \frac{1}{9} \|R_l^{gen} - \hat{R}_l\|_F + \frac{1}{2} \|t_l^{gen} - \hat{t}_l\|_1, \quad (4.14)$$

avec $\|\cdot\|_F$ la norme de Frobenius (définie comme la racine carrée de la somme des coefficients au carré), et

$$\hat{R}_l, \hat{t}_l = \mathcal{E}_\delta(\mathcal{D}_\beta(R_l^{gen}, t_l^{gen}) + \epsilon) \quad (4.15)$$

Nous représentons graphiquement cette formulation sur le diagramme 4.4.

L’avantage de cet apprentissage pose à pose est double. D’une part, cela permet d’assurer une exploration relativement exhaustive de l’ensemble des poses possibles, et ainsi éviter au mieux les minima locaux. D’autre part, cela permet d’assurer le désenchevêtrement des poses du paramètre de conformation z_l . En effet, en l’absence de cet apprentissage pose à pose, le paramètre de conformation z_l peut inclure tout ou une partie de l’estimation des poses, ce qui donne lieu à des reconstructions qui tournent en même temps que le paramètre de conformation évolue.

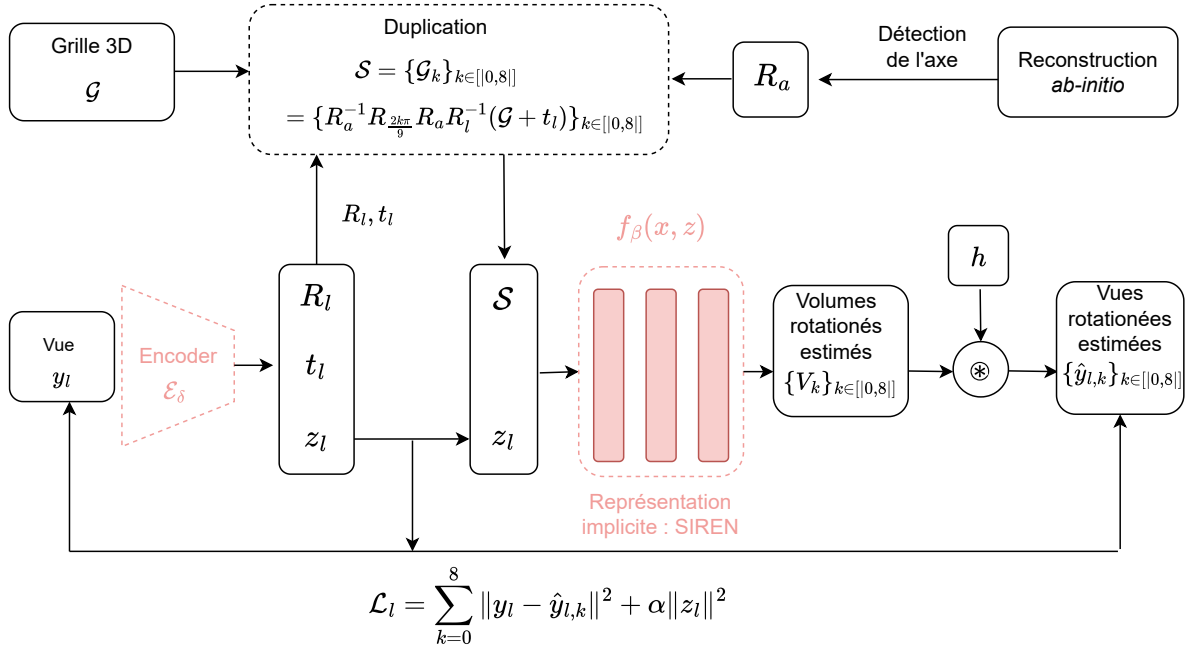


FIGURE 4.5 – Architecture permettant d’imposer la symétrie à la reconstruction

4.2.4 Imposer la symétrie C9 à la reconstruction

La représentation implicite du volume sous forme de SIREN est suffisamment générale pour permettre la reconstruction d’une structure quelconque. Néanmoins, dans certains cas nous avons accès à une information *a-priori* sur la structure de l’objet reconstruit que nous voudrions pouvoir utiliser pour faciliter la reconstruction. En particulier, dans le cas du centriole, nous savons qu’il admet une symétrie C9 autour de son axe de croissance. On dit qu’un objet admet une symétrie C9 autour d’un axe donné s’il est invariant par toute rotation de $\frac{2k\pi}{9}$ (avec k entier) autour de cet axe. Nous avons donc ajouté une option à notre méthode permettant d’imposer la symétrie C9 à la reconstruction. Ceci peut être utile lorsque la symétrie apparaît en partie dans les données mais que ces dernières admettent un nombre trop important de morceaux manquants pour qu’elle apparaisse sur la reconstruction.

Nous commençons par effectuer une reconstruction *ab-initio* à l’aide de notre méthode de reconstruction hétérogène, et l’imposition de la symétrie se fait à partir de cette reconstruction initiale. Cette reconstruction *ab-initio* est nécessaire car elle permet de pouvoir identifier l’axe de symétrie *a* sans avoir à l’estimer conjointement aux autres paramètres.

Formellement, la minimisation de la fonction de coût suivante permet d’obtenir une reconstruction admettant une symétrie C9 et qui se rapproche le plus possible des données :

$$E(f, \Theta, T, Z) = \sum_{l=1}^N \sum_{k=0}^8 \|y_l - h * (\mathcal{T}_{t_l} \circ \mathcal{R}_{\theta_l} \circ \mathcal{R}_a^{-1} \mathcal{R}_{\frac{2k\pi}{9}} \circ \mathcal{R}_a)(f(., z_l))\|^2 + \alpha \|z_l\|^2, \quad (4.16)$$

avec $\mathcal{R}_a$ la rotation permettant de passer de l’axe de rotation *a* de la reconstruction *ab-initio* à l’axe *z* et $\mathcal{R}_{\frac{2k\pi}{9}}$ la rotation de $\frac{2k\pi}{9}$ autour de l’axe *z*. Cette équation traduit que le volume estimé

f est dupliqué 9 fois, chaque version étant obtenue par rotation du volume courant de $\frac{2k\pi}{9}$ autour de l'axe de symétrie a . Ensuite, les mêmes opérations que celles de l'architecture de base sont appliquées à chacune des versions (c'est-à-dire une rotation, une translation et la convolution avec la réponse impulsionnelle du microscope). L'architecture du réseau permettant d'imposer la symétrie à la reconstruction est illustrée sur la figure 4.5. Pour la comprendre, il est à noter que l'équation 4.16 est équivalente à :

$$E(f, \Theta, T, Z) = \sum_{l=1}^N \sum_{k=0}^8 \sum_{x \in \mathcal{G}} [y_l - h * f((R_l R_a^{-1} R_{\frac{2k\pi}{9}} R_a)^t(x + t_l), z_l)]^2 + \alpha \|z_l\|^2 \quad (4.17)$$

$$= \sum_{l=1}^N \sum_{k=0}^8 \sum_{x \in \mathcal{G}} [y_l - h * f(R_a^{-1} R_{\frac{2k\pi}{9}} R_a R_l^{-1}(x + t_l), z_l)]^2 + \alpha \|z_l\|^2. \quad (4.18)$$

4.2.5 Imposer un modèle paramétrique à la reconstruction

Dans certains cas, il est possible d'approximer la structure de la protéine à reconstruire¹ par un modèle paramétrique. Dans le cas du centriole, nous savons par exemple que les protéines tubuline a une forme de tore.

Lorsque l'information contenue dans les données est insuffisante, à cause d'un nombre de vues trop faible, d'un degré de marquage partiel trop important ou d'un niveau de bruit trop élevé, imposer à la reconstruction de suivre ce modèle paramétrique peut apporter une information *a priori cruciale*.

Pour imposer un modèle paramétrique, l'architecture de base 4.1 doit être légèrement modifiée. La partie principale du décodeur, qui est un SIREN dans l'architecture de base est remplacée par le modèle paramétrique, c'est-à-dire une fonction qui génère la forme que l'on souhaite imposer à partir d'un certain nombre de paramètres (qui dépendent du choix de modèle paramétrique adopté). Ces paramètres sont directement générés à partir du décodeur : la dimension de l'espace latent doit donc être choisie de sorte à ce que le nombre de variables latentes soit égal au nombre de variables du modèle paramétrique. Le reste de l'architecture reste par ailleurs inchangé.

Nous présentons en section 4.5 un cas particulier de modèle paramétrique que nous avons implémenté : une forme de tore bicanal.

4.3 Résultats sur des données générées synthétiquement

Nous avons évalué les performances de notre méthode sur des données synthétiques. Nous présentons dans un premier temps des résultats de reconstruction homogène. Puis, nous exposons des résultats de reconstruction hétérogène. Nous avons en particulier testé l'influence du bruit ainsi que la robustesse à différents degrés d'anisotropie de résolution.

4.3.1 Reconstruction homogène

Bien que l'objectif de Fluofire soit la reconstruction hétérogène, nous commençons par valider ses performances dans le cas de la reconstruction homogène.

1. Par "structure de la protéine" il faut ici comprendre la structure formée par l'ensemble des protéines d'un même type de l'assemblage macromoléculaire considéré

4.3.1.1 Génération des données

Nous générons des données homogènes à partir des 4 objets de référence présentés en section 3.6.1.1 : Clathrine, NLR-Résistosome, HIV-Vaccine et Récepteurs AMPA. A partir de chacun de ces 4 objets sont générées 250 vues par application du modèle d’observation. Les poses sont sélectionnées aléatoirement, selon une distribution uniforme sur la discrétisation de la sphère présentée en partie 3.2.1. La réponse impulsionnelle, de forme gaussienne, admet des écarts-types de $\sigma_z = 5$ selon l’axe z et $\sigma_{xy} = 1$ dans le plan latéral. Nous ajoutons par ailleurs un bruit gaussien pour obtenir un $RSB = 1$ (0 dB).

4.3.1.2 Expérimentation

Pour chacun de ces 4 jeux de données, nous lançons 2 expérimentations. La première est une reconstruction en utilisant l’architecture de base de Fluofire 4.1. Comme il s’agit de jeux de données homogènes, l’encodeur $\mathcal{E}_\delta$ n’estime pas de paramètre de conformation z_l (il estime uniquement les poses), et le décodeur f_β prend en entrée uniquement les coordonnées d’un voxel. La seconde expérience considère une alternance entre des phases d’apprentissage image à image et pose à pose (comme présenté en section 4.2.3). Dans les deux cas, l’algorithme d’optimisation Adam du module pytorch est utilisé pour l’optimisation, et nous fixons le pas de gradient associé aux poids de l’encodeur à 10^{-4} , et le pas associé à la représentation implicite sous forme de SIREN est fixé à 5×10^{-6} . Nous utilisons par ailleurs une taille de batch de 4.

4.3.1.3 Résultats

Le tableau 4.6 présente les résultats obtenus. Sur la première colonne les 4 objets vérité terrain sont illustrés. Les deuxième et troisième colonnes illustrent respectivement les résultats obtenus par Fluofire sans ou avec utilisation de l’apprentissage pose à pose. A titre de comparaison, la dernière colonne présente les résultats de reconstruction obtenus dans le chapitre 3 par la méthode voxel sur ces mêmes objets. Nous pouvons voir que l’ajout de l’apprentissage pose à pose est indispensable pour assurer la convergence vers une solution correcte. D’autre part, nous constatons que pour les deux objets Clathrine et NLR resistosome, la reconstruction de Fluofire est plus éloignée de la vérité terrain que celle de SHiReVol. Il est à noter que ces deux objets, contrairement aux deux autres, admettent une certaine forme de symétrie. Par conséquent, une grande variation de l’orientation associée à une vue ne correspond pas nécessairement à une grande variation de la fonction de coût correspondante. Nous faisons l’hypothèse que ceci peut perturber l’apprentissage de l’encodeur. Il ne s’agit néanmoins que d’une hypothèse, que nous n’avons pas validé.

4.3.2 Reconstruction hétérogène

Nous avons ensuite évalué Fluofire sur des jeux de données hétérogènes générés synthétiquement. Nous présentons dans un premier temps un résultat de reconstruction lorsque l’espace latent admet une seule dimension. Puis, nous analysons les résultats et la structure de l’espace latent lorsqu’il admet deux dimensions. Ensuite, nous évaluons les performances de l’option permettant d’imposer la symétrie $C9$ à la reconstruction. Par la suite, nous évaluons la robustesse au bruit de Fluofire. Enfin, nous étudions l’influence du nombre de vues et de l’anisotropie.

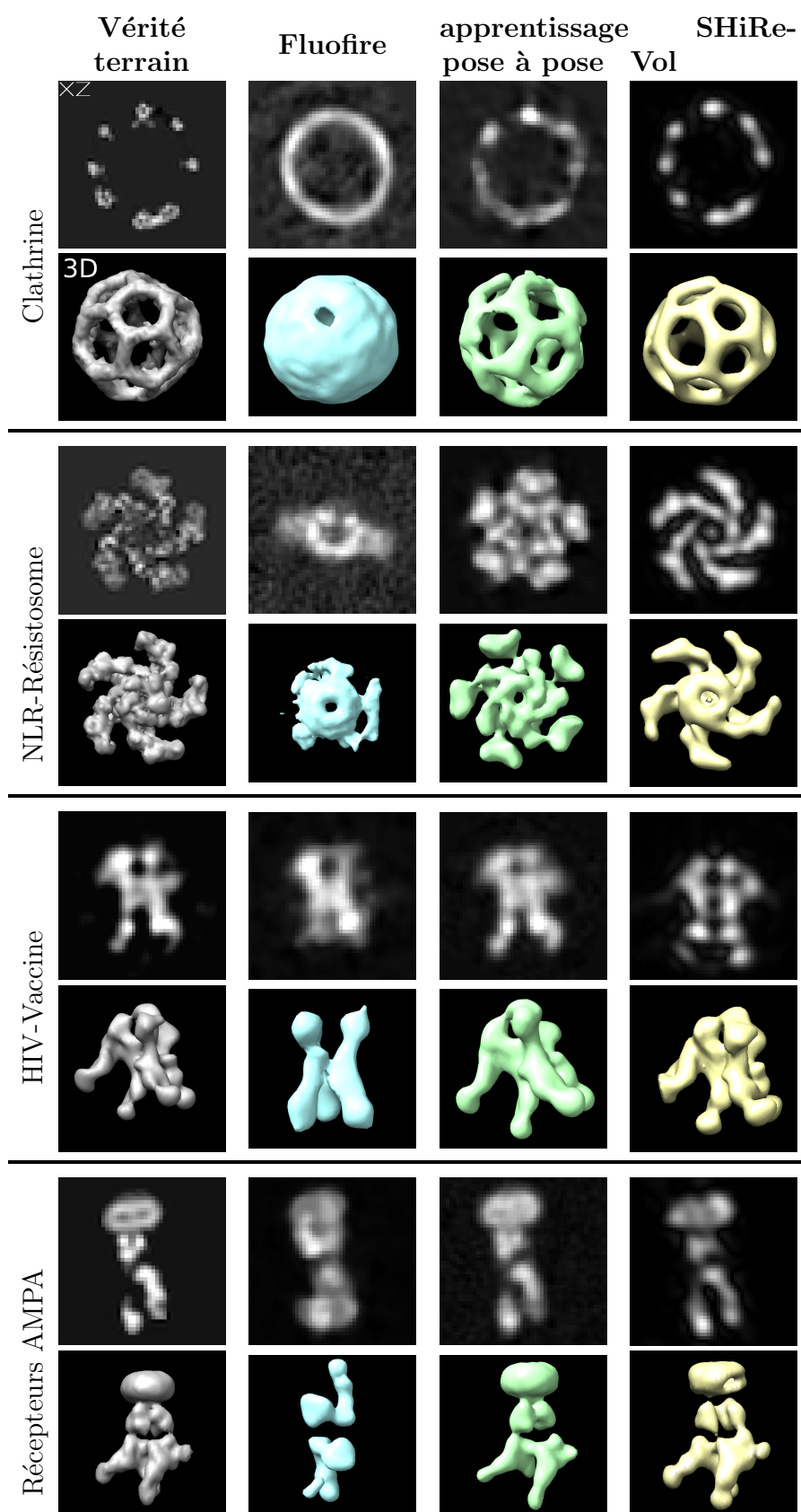


FIGURE 4.6 – Résultats de reconstruction homogène obtenus sur des données synthétiques, à l'aide des méthodes Fluofire (avec ou sans apprentissage pose à pose) et SHiReVol. Chaque volume est représenté par une coupe dans le plan XZ et une représentation iso-surfacique 3D

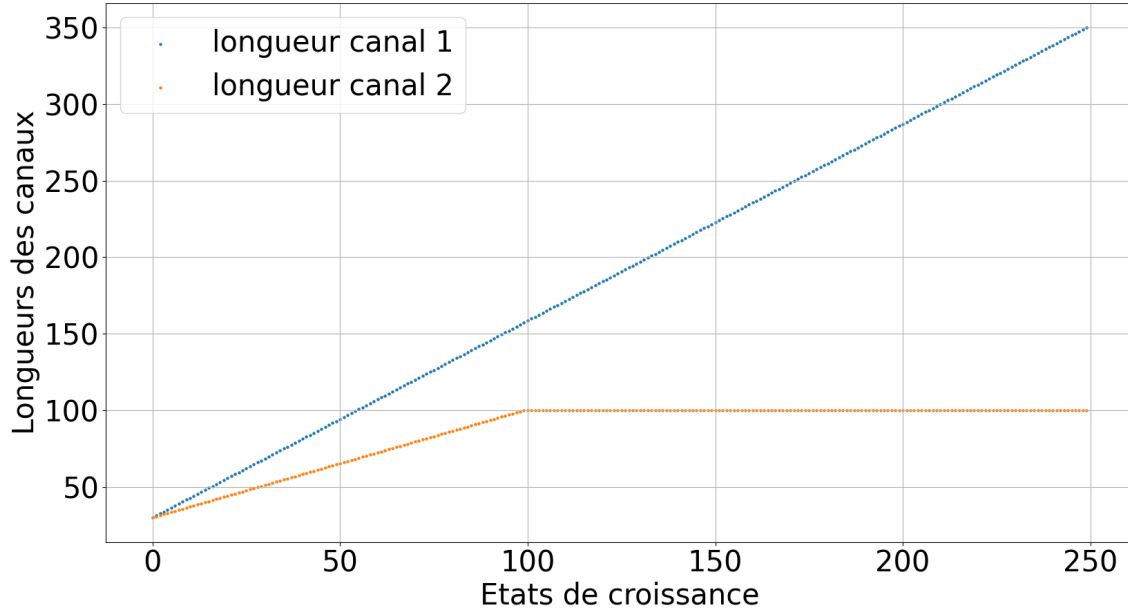


FIGURE 4.7 – Longueurs du centriole synthétique associées aux différents états de croissances des deux canaux. La taille de l’image vaut $S = 400$

4.3.2.1 Espace latent monodimensionnel

Génération des données Nous générons des données à partir d’un modèle synthétique de centriole, établi expérimentalement par Le Guennec *et al.* [105]. Ce dernier permet de générer différents états de croissance de la protéine tubuline du centriole, qui a une structure cylindrique admettant une symétrie C9. Nous générons 250 états de croissance, et pour chacun d’entre eux nous simulons une vue par application du modèle d’observation (nous utilisons encore $\sigma_z = 5$ et $\sigma_{xy} = 1$). Nous simulons un objet de référence admettant 2 canaux. Le premier canal, de rayon supérieur au second, croît linéairement d’une longueur de $\frac{S}{10}$ (avec S la taille de l’image) jusqu’à $\frac{8,75S}{10}$. Le second croît linéairement de $\frac{S}{10}$ à $\frac{S}{4}$ sur les 100 premiers états puis reste identique égal à $\frac{S}{4}$ sur les 150 états suivants. 4 états de croissance de l’objet de référence sont représentés sur la première colonne de la figure 4.8. La figure 4.7 illustre, pour les deux canaux, les longueurs du centriole synthétique associées aux différents états de croissances.

Expérimentation Nous fixons la dimension du paramètre de conformation z_l à 1 et effectuons une reconstruction en lançant l’algorithme Fluofire, avec l’option d’apprentissage pose à pose. Les pas de gradient associés à l’encodeur et au décodeur, ainsi que la taille de batch, sont fixés aux mêmes valeurs que celles utilisées pour la reconstruction hétérogène. Par ailleurs, nous fixons la variable de régularisation du paramètre d’hétérogénéité à $\alpha = 10^{-4}$. Étant donné que les données d’entrée admettent 2 canaux, nous fixons à 2 le nombre de neurones de sortie de la représentation implicite du volume sous forme de SIREN (voir figure 4.2).

Résultats Nous présentons sur la deuxième colonne de la figure 4.8 la reconstruction obtenue à 4 états de croissances différents, régulièrement espacés. Nous pouvons voir que la structure est

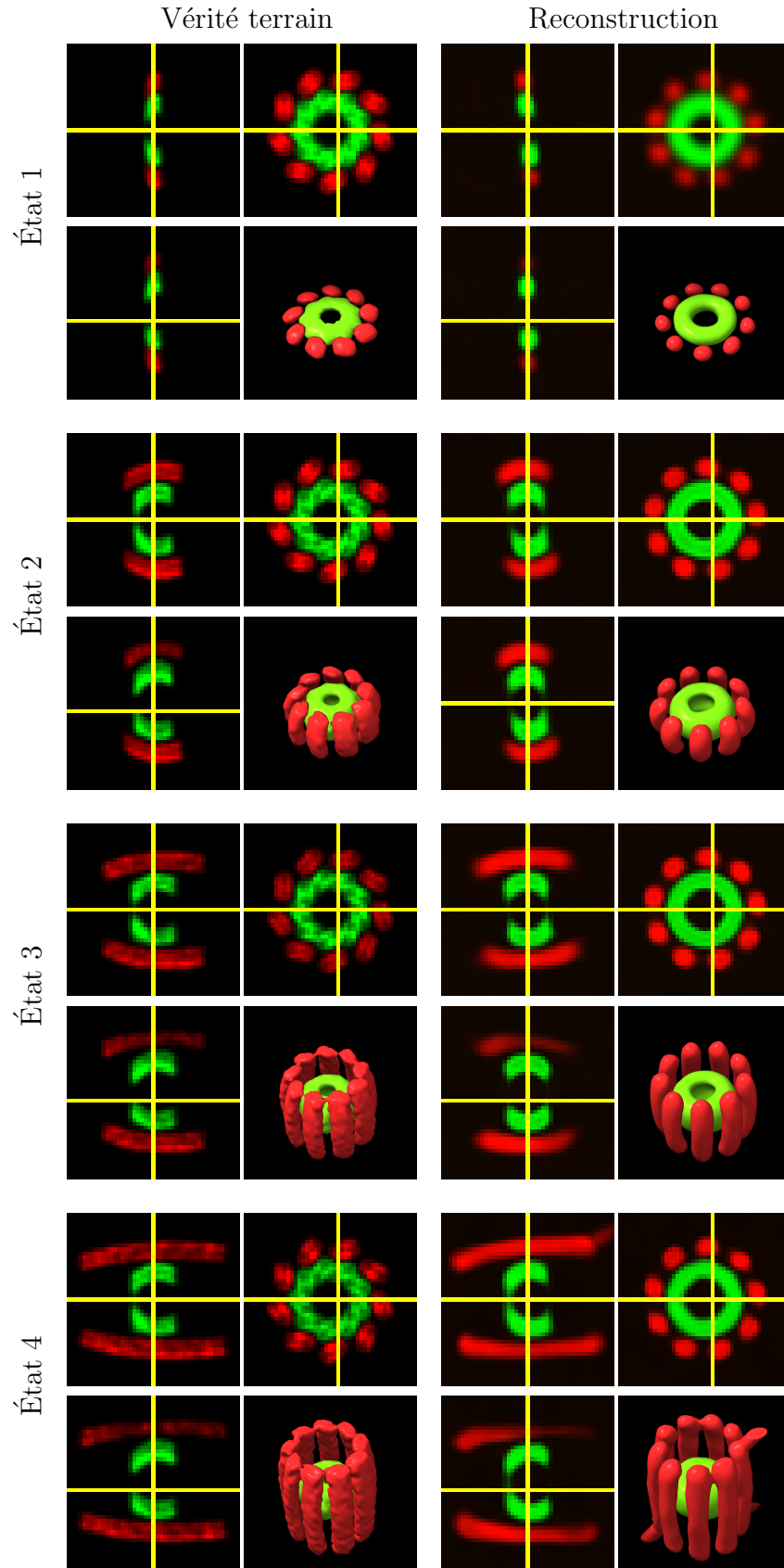


FIGURE 4.8 – Résultats de reconstruction hétérogène obtenus par méthode Fluofire sur des données de centriole synthétique. 4 états sont illustrés

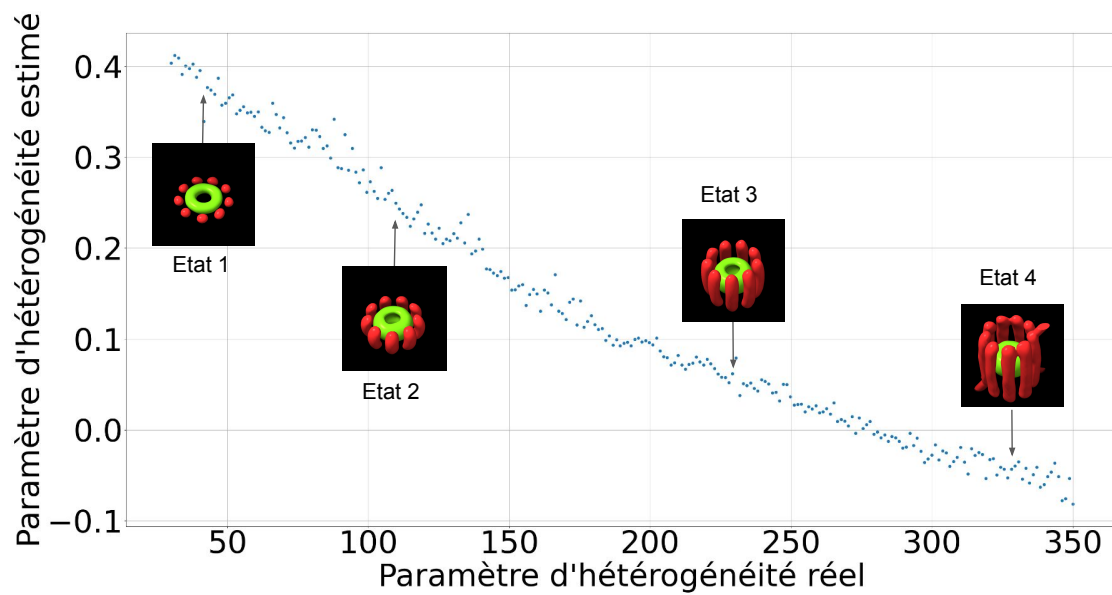


FIGURE 4.9 – Estimation du paramètre d'hétérogénéité z_l . Chaque point est associé à une vue du jeu de données. En abscisse est représenté le paramètre d'hétérogénéité réel, c'est-à-dire la longueur du canal rouge. Les valeurs (qui vont de 30 à 350) sont celles mises en entrée du modèle de génération du centriole synthétique. Selon la référence du modèle, la taille de l'image vaut $S = 400$. En ordonnée est illustré le paramètre de conformation z_l estimé par Fluofire.

globalement similaire à la vérité terrain. On observe néanmoins une légère erreur de l'algorithme de reconstruction, dû au fait que le centriole de référence présente une légère fausse symétrie. En effet, son rayon est légèrement plus grand d'un côté de la structure que de l'autre. On observe alors que, pour certains états de croissance, la structure est reconstruite dans un sens inversé (par exemple aux états 2 et 3 de la figure 4.8).

Nous présentons également sur la figure 4.9 le paramètre de conformation estimé en fonction du paramètre de conformation réel, c'est-à-dire la longueur du premier canal (le canal rouge). Les volumes associés à certaines valeurs de paramètre de conformation sont également représentés sur la figure. Nous pouvons remarquer qu'il existe une relation quasi-linéaire entre les deux, ce qui montre que ce dernier est correctement estimé.

4.3.2.2 Espace latent bidimensionnel

Génération des données Pour cette expérience deux jeux de données sont simulées. Le premier présente les mêmes caractéristiques que celui présenté au paragraphe 4.3.2.1 : il s'agit donc d'un jeu de données avec un seul degré de liberté. Le second jeu de données admet quant à lui deux degrés de liberté. Les centrioles générés ont toujours une longueur variable, mais également un rayon variable. Pour limiter le coût computationnel, nous générons cette fois-ci un seul canal dans les données. La longueur du centriole varie de $\frac{S}{10}$ à $\frac{8,75S}{10}$ avec 50 valeurs de longueurs différentes. Le rayon varie de $\frac{S}{8}$ à $\frac{5S}{16}$ avec 30 valeurs de rayon différentes.

Expérimentations L'algorithme Fluofire est ensuite lancé sur ces deux jeux de données, en utilisant les mêmes paramètres que ceux présentés dans le paragraphe 4.3.2.1 à l'exception de l'espace latent des conformations dont la dimension est fixée à 2 (au lieu de 1).

Résultats La figure 4.10 illustre la structure de l'espace latent à la fin de l'apprentissage sur le jeu de données admettant un seul degré de liberté. Nous représentons en abscisse et ordonnée respectivement les première et seconde dimensions de l'espace latent. Le paramètre de conformation réel, c'est-à-dire la longueur du centriole, est quant à lui illustré à l'aide d'un dégradé de couleur. Nous observons clairement que l'unique degré de liberté des données est correctement capté par la méthode de reconstruction. Il apparaît sur une droite, la position sur la droite représente la longueur du centriole. Nous utilisons alors une Analyse en Composante Principale (ACP) afin d'identifier la direction dans l'espace latent du degré de liberté.

La figure 4.11 illustre la structure de l'espace latent à la fin de l'apprentissage sur le jeu de données admettant deux degrés de liberté. Les deux sous-figures illustrent exactement les mêmes points, mais les labels sont différents. Sur la figure 4.11a, le gradient de couleur représente le premier paramètre de conformation réel (*i.e.* la longueur). Sur la figure 4.11b, le gradient de couleur représente le second paramètre de conformation réel (*i.e.* le rayon). Nous observons que l'espace latent est structuré autour de deux axes. Les déplacements selon les premier et second axe donnent lieu respectivement à une augmentation de la longueur et du rayon du centriole. Ceci est illustré par les représentations visuelles de 4 états, correspondant à des valeurs extrêmes de l'espace latent. Les deux axes sont ici encore détectés à l'aide d'une ACP sur l'espace latent, mais ce choix pourrait cependant être critiqué. En effet, l'ACP définit les axes du nouveau repère comme une combinaison linéaire des axes du repère d'origine. Or, dans notre cas, il semble que l'espace latent n'est pas tout à fait la forme d'un rectangle, il pourrait donc être plus pertinent de définir les dimensions du nouveau repère comme des courbes plutôt que des droites.

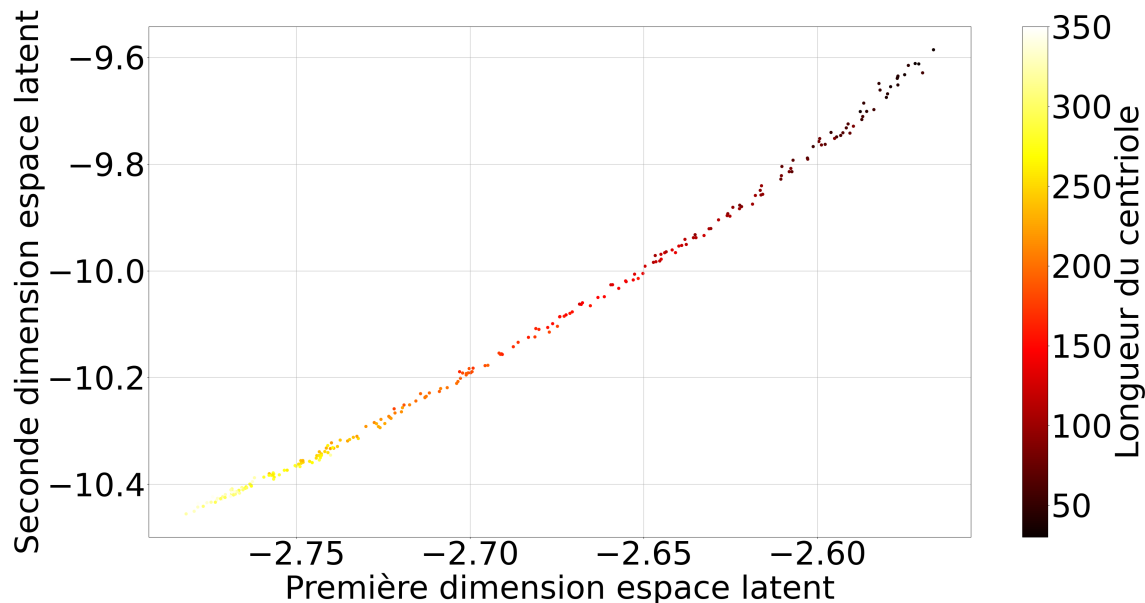


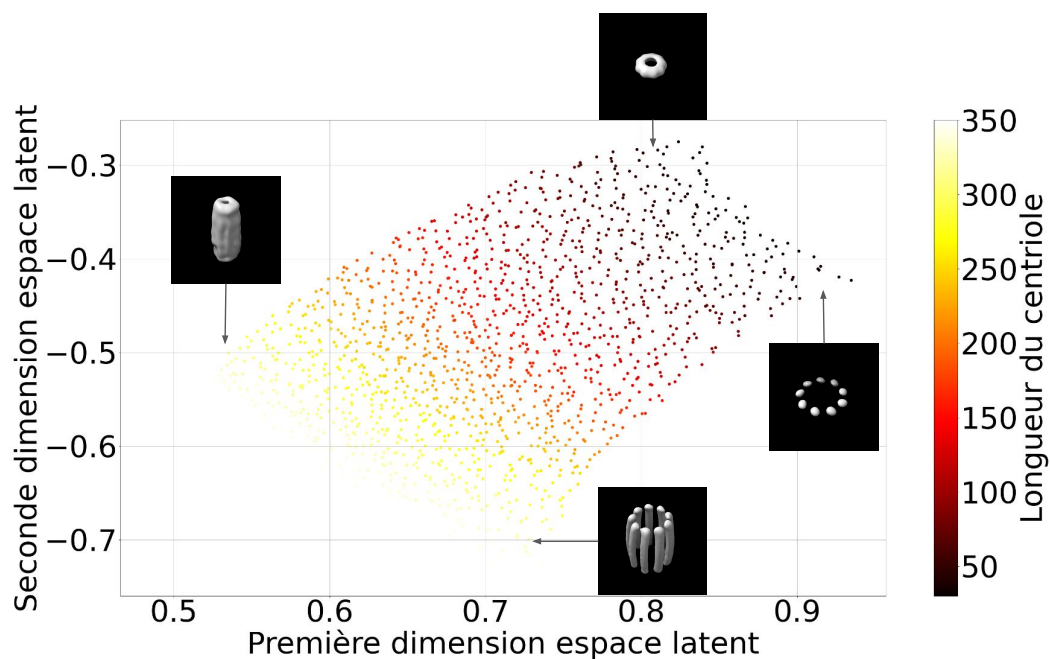
FIGURE 4.10 – Structure de l’espace latent à deux dimensions lorsque les données générées présentent un seul degré de liberté. Le paramètre de conformation réel - *i.e.* la longueur réelle du canal 1 - est représenté à l’aide d’un gradient de couleur.

En résumé, les résultats de reconstruction sur des données admettant une hétérogénéité des conformations multidimensionnelle semblent satisfaisants. Néanmoins, une analyse plus poussée de l’espace latent basée sur des transformations non-linéaires pourrait offrir une perspective de recherche.

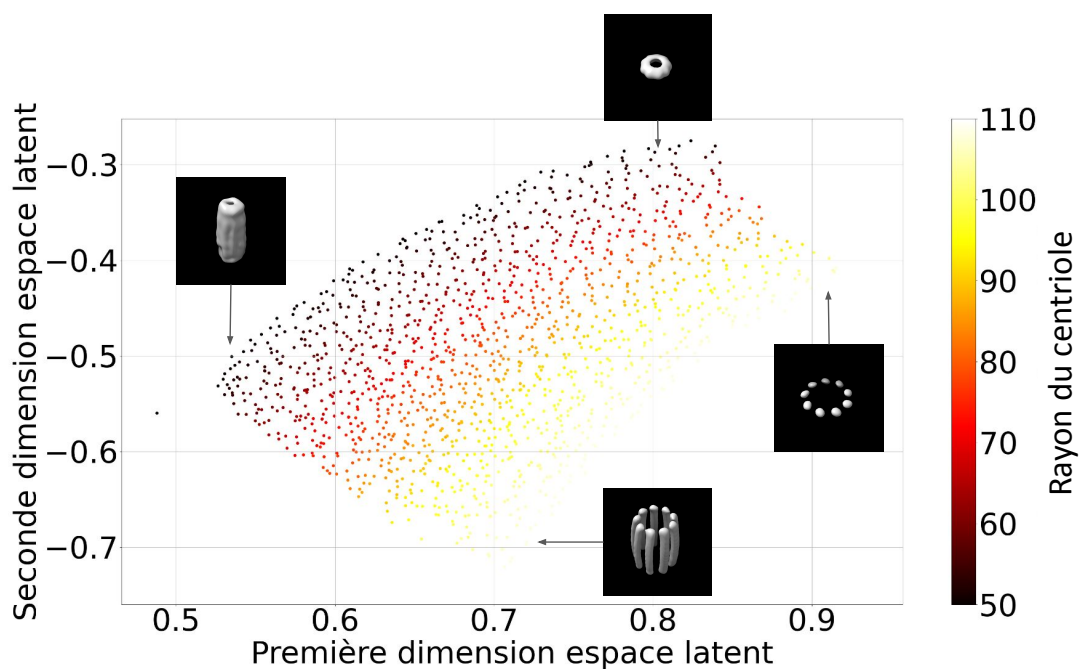
4.3.2.3 Symétrie imposée

Génération des données Afin d’évaluer l’option permettant d’imposer la symétrie C_9 à la reconstruction, nous avons généré des vues à partir du centriole synthétique utilisé dans la section 4.3.2 mais en lui retirant un certain nombre de bras. Nous aurions pu également retirer des spots gaussiens à des emplacements aléatoires de la vérité terrain, ce qui aurait constitué un jeu de données plus réaliste. Néanmoins, dans ce cas la symétrie C_9 aurait pu apparaître sans qu’on ait besoin de l’imposer, il nous semblait donc plus convainquant de traiter des jeux de données plus compliqué, où un ou plusieurs bras du centriole sont manquants. Nous avons généré en tout 6 jeux de 250 vues issues d’objets de référence ayant respectivement 1, 2, 3, 4, 5 et 8 bras manquants.

Expérimentation Sur ces 6 jeux de données, nous lançons la variante de Fluofire décrite en section 4.2.4, permettant d’imposer la symétrie C_9 . Nous utilisons les mêmes paramètres que ceux utilisés précédemment, à l’exception de la taille de batch qui est fixée à 1. En effet, l’algorithme qui impose la symétrie nécessite une duplication par 9 de la grille 3D d’entrée, et les 9 grilles doivent être passées conjointement dans le réseau SIREN. Une taille de mini-batch supérieure à 1, pour une taille d’images de $55 \times 55 \times 55$ entraîne une erreur de mémoire de carte graphique (de mémoire vive 12Go).



(a)



(b)

FIGURE 4.11 – Structure de l'espace latent à deux dimensions lorsque les données générées présentent deux degrés de liberté (données simulées). Le premier paramètre de conformation réel - la longueur du centriole - est représenté à l'aide d'un gradient de couleur sur la figure (a). Le second paramètre de conformation réel - le rayon du centriole - est représenté à l'aide d'un gradient de couleur sur la figure (b). Les états associés à quatre points de l'espace latent sont également illustrés

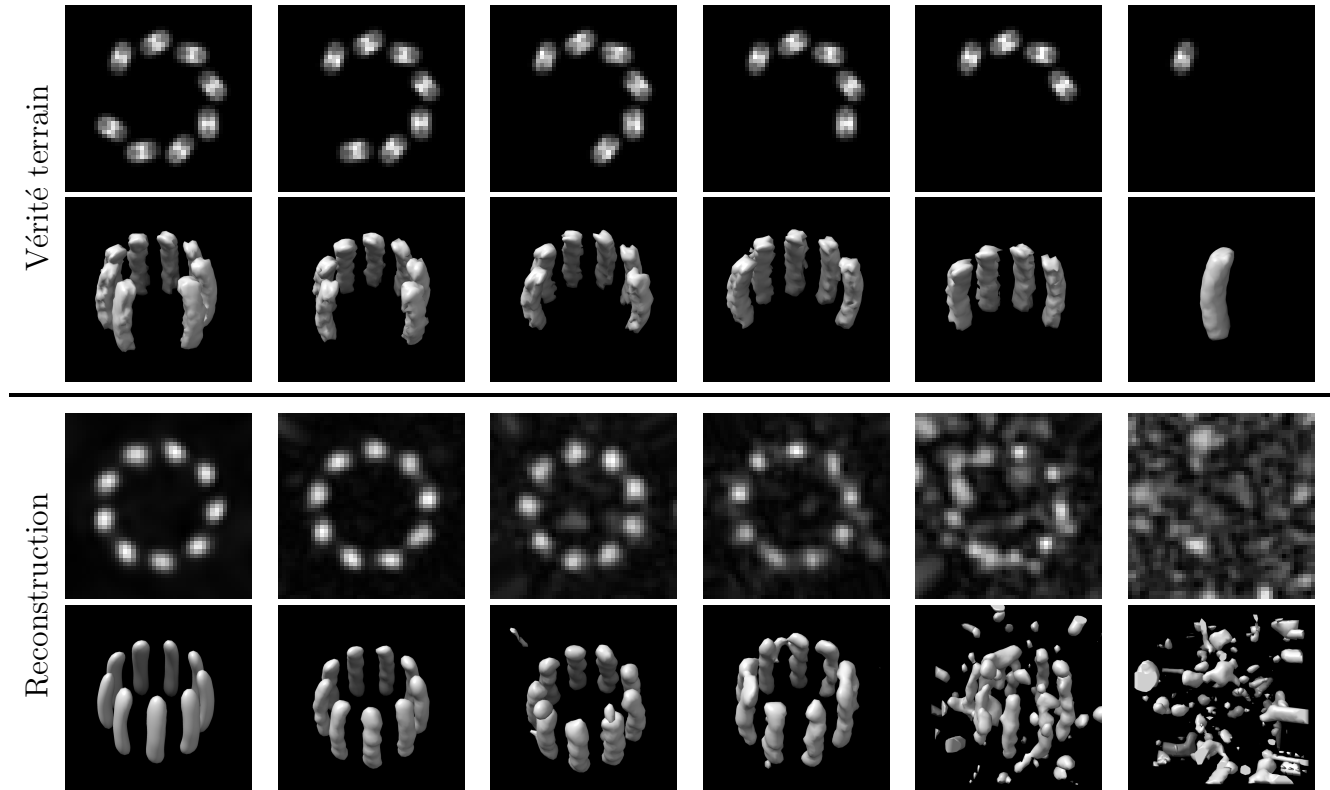


FIGURE 4.12 – Résultats de reconstruction avec l’option permettant d’imposer la symétrie. Sur les 2 premières lignes, l’objet vérité terrain est représenté à l’aide d’une coupe dans le plan XY et une représentation surfacique. Nous utilisons un centriole synthétique avec un nombre croissant de bras manquants. Sur les deux dernières, la reconstruction obtenue est illustrée. Pour une visualisation concise, nous représentons les objets vérité terrains et reconstruits à un seul état de reconstruction mais des reconstructions hétérogènes ont bien été effectuées.

Résultats Les résultats obtenus sont présentés sur la figure 4.12. Nous constatons que la méthode fonctionne correctement : on obtient un objet admettant une symétrie C_9 qui se rapproche le plus possible des données. On observe également que la reconstruction se dégrade progressivement lorsque le nombre de bras manquants augmente, ce qui est tout à fait prévisible. On distingue encore légèrement le centriole lorsque 5 bras sont manquants, mais avec 8 bras manquants plus rien n’est visible.

L’inconvénient de cette option permettant d’imposer la symétrie à la reconstruction est qu’elle multiplie par 9 le temps de calcul du passage à travers le décodeur. Étant donné que, dans l’architecture de base, la traversée de l’encodeur prend environ le même temps que la traversée du décodeur, cela a pour conséquence de multiplier par 5 le temps de calcul à travers le système encodeur-décodeur. Mais le temps de calcul par itération n’est pas si problématique car notre méthode qui impose la symétrie est appliquée à une reconstruction *ab-initio* assez proche de la structure finale, il n’y a donc besoin que de relativement peu d’itérations pour atteindre la convergence. Ce qui pose surtout problème, c’est la taille mémoire utilisée, nous sommes contraints d’utiliser une taille de mini-batch de 1 et des images de taille maximale de $55 \times 55 \times 55$ (en voxels).

4.3.2.4 Influence du bruit

Il est important d’évaluer la robustesse au bruit de notre méthode afin de s’assurer qu’elle puisse être adaptée à des cas d’application réels.

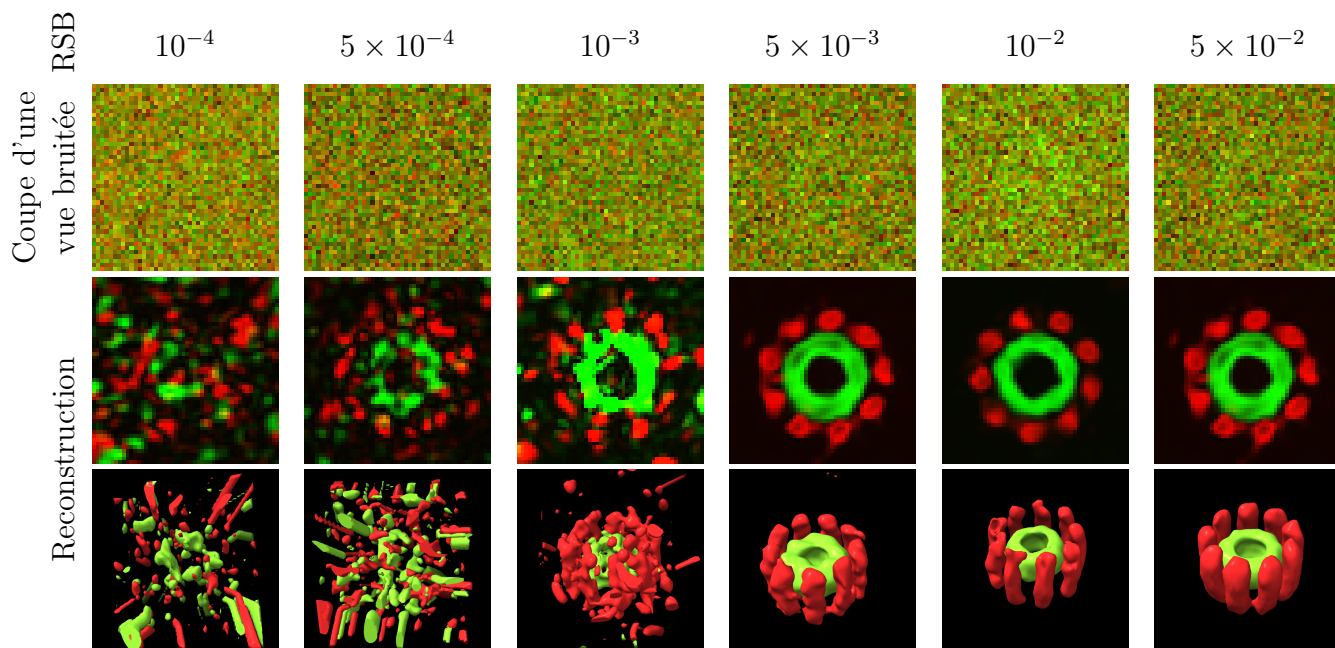
Génération des données Pour cela, nous avons généré 20 jeux de 250 vues auxquels des bruits gaussiens de différents niveaux ont été ajoutés. Le niveau de bruit varie d’un rapport signal à bruit (RSB) de 10^{-4} à un RSB de 10^2 . Les vues sont des données hétérogènes admettant un degré de liberté, elles sont générées selon le schéma décrit dans le paragraphe 4.3.2.1. La première ligne de la figure 4.13b illustre quelques exemples de vues bruitées pour différents RSB.

Expérimentation Pour chaque RSB, 10 reconstructions sont effectuées avec la méthode Fluofire. Afin d’évaluer la performance de la méthode nous évaluons pour chaque expérience la valeur absolue de la corrélation entre la paramètre d’hétérogénéité réel, c’est-à-dire la longueur du cylindre rouge, et le paramètre d’hétérogénéité estimé z_l .

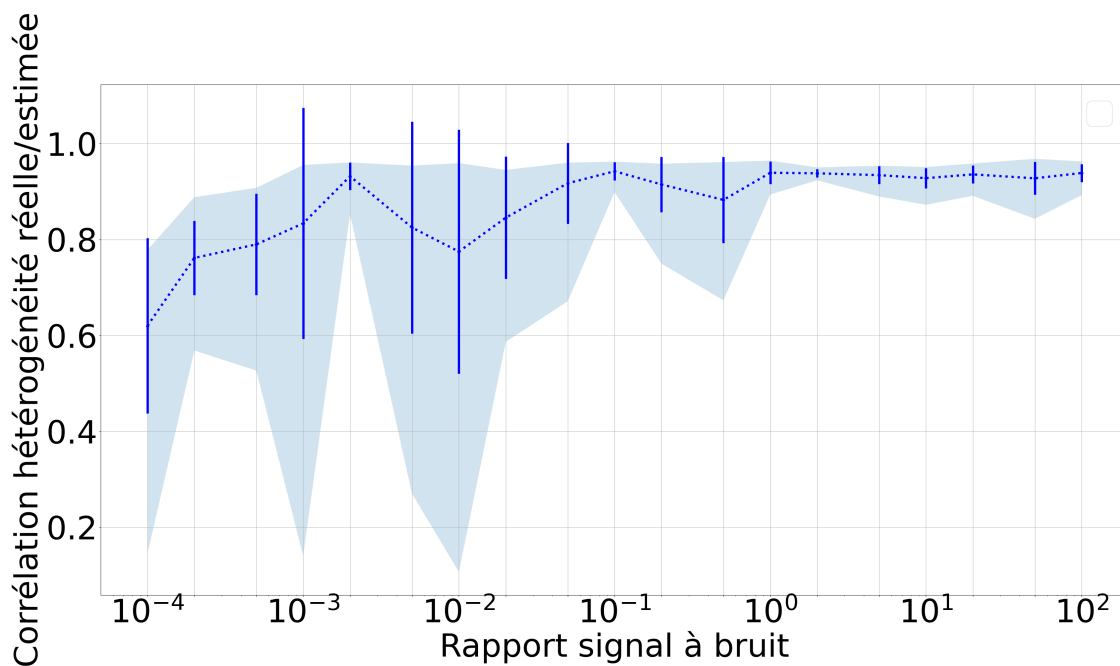
Résultats La figure 4.13a montre des reconstructions obtenues pour différents RSB. À partir de $RSB = 5 \times 10^{-2}$, la qualité visuelle de la reconstruction ne s’améliore plus, c’est pourquoi nous ne montrons pas les résultats obtenus pour des RSB supérieurs. On observe que la méthode commence à donner un résultat cohérent à partir de $RSB = 10^{-3}$, voire 5×10^{-4} . Cela illustre la forte robustesse au bruit de Fluofire. Le graphique 4.13b illustre la corrélation entre les paramètres d’hétérogénéité réel et estimé pour différents niveaux de bruit. Comme pour les précédents graphiques, les barres verticales représentent les écarts types sur les 10 tests, et les zones grisées l’étendu entre le minimum et le maximum. Nous pouvons voir que la courbe se stabilise à partir de $RSB = 5 \times 10^{-2}$, ce qui est cohérent avec ce que l’on observe visuellement.

4.3.2.5 Influence du nombre de vues et de l’anisotropie

Tout comme le RSB, le niveau d’anisotropie (*i.e.* l’écart type σ_z de la PSF selon l’axe z) des données est un paramètre important du modèle de dégradation. Nous nous attendons à ce que



(a) Résultats visuels



(b) FSC à la vérité terrain. Chaque point du graphe est une moyenne effectuée sur 10 tests

FIGURE 4.13 – Influence du rapport signal à bruit sur la qualité de la reconstruction, méthode SHiReVol

Fluofireait plus de mal à reconstruire lorsque ce dernier augmente. De même on s’attend à ce que le nombre de vues ait une influence considérable sur la qualité de la reconstruction car un nombre de vues élevé permet de tirer partie de plus d’informations.

Génération des données Nous générons des jeux de données pour diverses combinaisons (N, σ_z) . Plus précisément N (le nombre de vues) prend pour valeurs : 5, 20, 50, 100, 200, 300 et 500, et σ_z (l’écart type de la PSF selon l’axe z) : 1, 3, 5, 10 et 15. Il a donc en tout $7 \times 5 = 35$ jeux de données. Le RSB est par ailleurs fixé à 1.

Expérimentation Pour chacun des 35 jeux de données, nous lançons 10 reconstructions. Nous évaluons encore les performances de la méthode en calculant pour chaque expérience la valeur absolue de la corrélation entre la paramètre d’hétérogénéité réel et le paramètre d’hétérogénéité estimé.

Résultats La figure 4.14 montre l’évolution de la corrélation entre les hétérogénéités réelles et estimées avec le nombre de vues, pour différents niveaux d’anisotropie. Nous constatons que la corrélation augmente avec le nombre de vues, et qu’elle se stabilise à partir d’une centaine de vues. Il ne semble donc pas nécessaire d’en utiliser plus. Néanmoins, ceci n’est pas un résultat absolu : des expérimentations sur d’autres structures pourraient donner des résultats différents. Par ailleurs, le modèle de dégradation des données réelles n’est rarement parfaitement identique au modèle de dégradation théorique, il n’est donc pas du tout certain que l’on puisse généraliser ces résultats obtenus sur des données jouets dans le cadre de l’application de Fluofire sur des données réelles.

Quant à l’influence de l’anisotropie, nous ne constatons pas de différences notables entre les différentes courbes lorsque $N > 100$. Lorsque $N < 100$, la variance des résultats est trop importante pour pouvoir distinguer les courbes les unes des autres. Il aurait pu être judicieux d’effectuer des expérimentations supplémentaires, en augmentant le nombre de valeurs de N testées et le nombre de tests par point. Néanmoins, dans une approche de sobriété énergétique et compte tenu de l’apport mineur à la connaissance scientifique que ces expérimentations auraient pu apporter, nous n’avons pas souhaiter les effectuer.

4.4 Résultats obtenus sur des données réelles

Dans la section précédente, nous avons mis en évidence le bon fonctionnement de Fluofire sur des données synthétiques, et en particulier sa forte robustesse au bruit. Nous présentons dans cette section des résultats obtenus sur des données réelles de procentreiole. Nous présentons dans un premier temps les données à notre disposition, et dans un second temps les résultats obtenus sur un jeu de données monocanal, puis un jeu de données bicanal.

4.4.1 Description des données

4.4.1.1 Données brutes

Le Centriole Lab, dirigé par Paul Guichard et Virginie Hamel, biologistes du Centriole Lab a mis à notre disposition des images contenant divers centrosomes. Pour rappel, un centrosome est constitué d’un centriole mature et d’un procentreiole orientées orthogonalement (voir section 1.4.2).

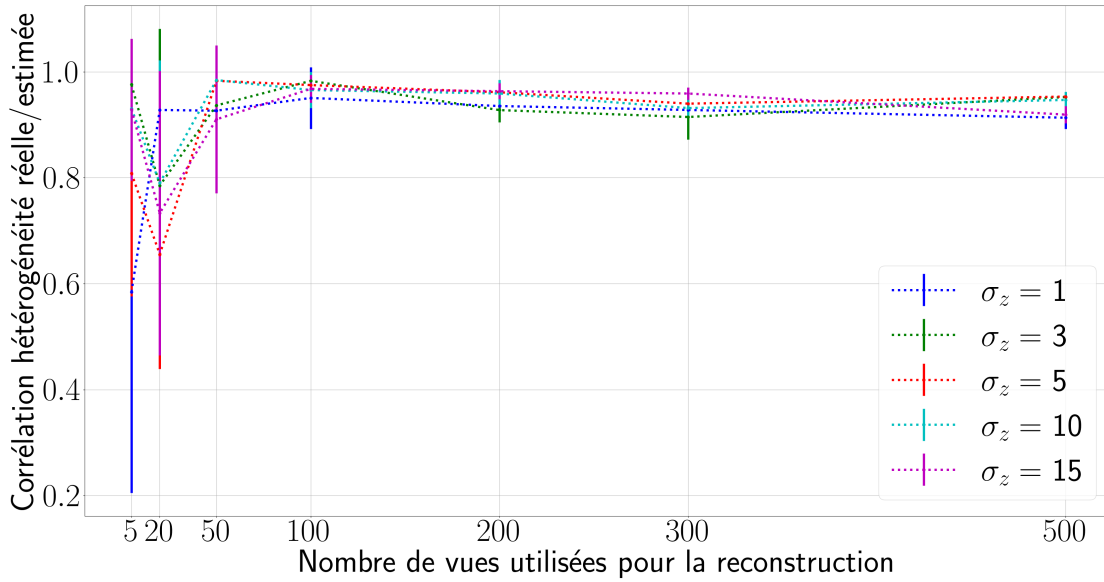


FIGURE 4.14 – Influence du nombre de vues et de l’anisotropie sur l’estimation du paramètre d’hétérogénéité

Les images ont été acquises à l’aide d’un microscope confocal Leica TCS SP8. La technique U-ExM (Ultrastructure Expansion Microscopy) est utilisée, donnant lieu à une expansion des particules d’un facteur 4. Plus de détails sur l’acquisition de la préparation de l’échantillon et l’acquisition des images sont disponibles dans le paragraphe ”Method details” de [29]

Le jeu de données est constitué de 1958 images de centrosomes, mais toutes ces images ne correspondent pas aux mêmes protéines observées. Chaque image permet la visualisation de deux protéines : l’une d’entre elles est la tubuline et est commune à toutes les images. L’autre dépend du sous-ensemble considéré et peut être la Sas6, CPAP, POC1B, POC, Centrin et bien d’autres encore (voir [28] pour plus de détails sur ces protéines). La figure 4.15 illustre 4 images de centrosomes. Les protéines visualisées sont la tubuline (en rouge), et Sas6 (en vert), présente uniquement sur le procentriole. Nous pouvons voir que le procentriole apparaît à divers états de croissance.

4.4.1.2 Prétraitement

Nous nous sommes intéressés à la reconstruction hétérogène du procentriole afin de reconstituer ses différents états de croissance. Pour cela, nous avons dans un premier temps développé un algorithme semi-automatique permettant de séparer les centrioles matures des procentrioles. Cet algorithme se compose de trois étapes. Dans un premier temps, les images sont transformées en nuages de points définis comme les coordonnées des voxels pour lesquels l’intensité de la tubuline est supérieur à un certain seuil prédéfini. Dans un second temps, l’algorithme de classification non supervisé DBSCAN [106] vise à séparer le nuage de point en deux sous-ensembles, l’un contenant les points du procentriole et l’autre les points du centriole mature. Enfin un algorithme de machine à vecteur support [107] permet de calculer le plan qui sépare au mieux les deux clusters. Ce plan est utilisé pour générer deux images, l’une contenant le centriole mature et l’autre le procentriole. Il s’agit d’un algorithme semi-automatique car une certaine intervention manuelle a été nécessaire pour obtenir une séparation correcte pour l’ensemble des images du jeu de données. En effet,

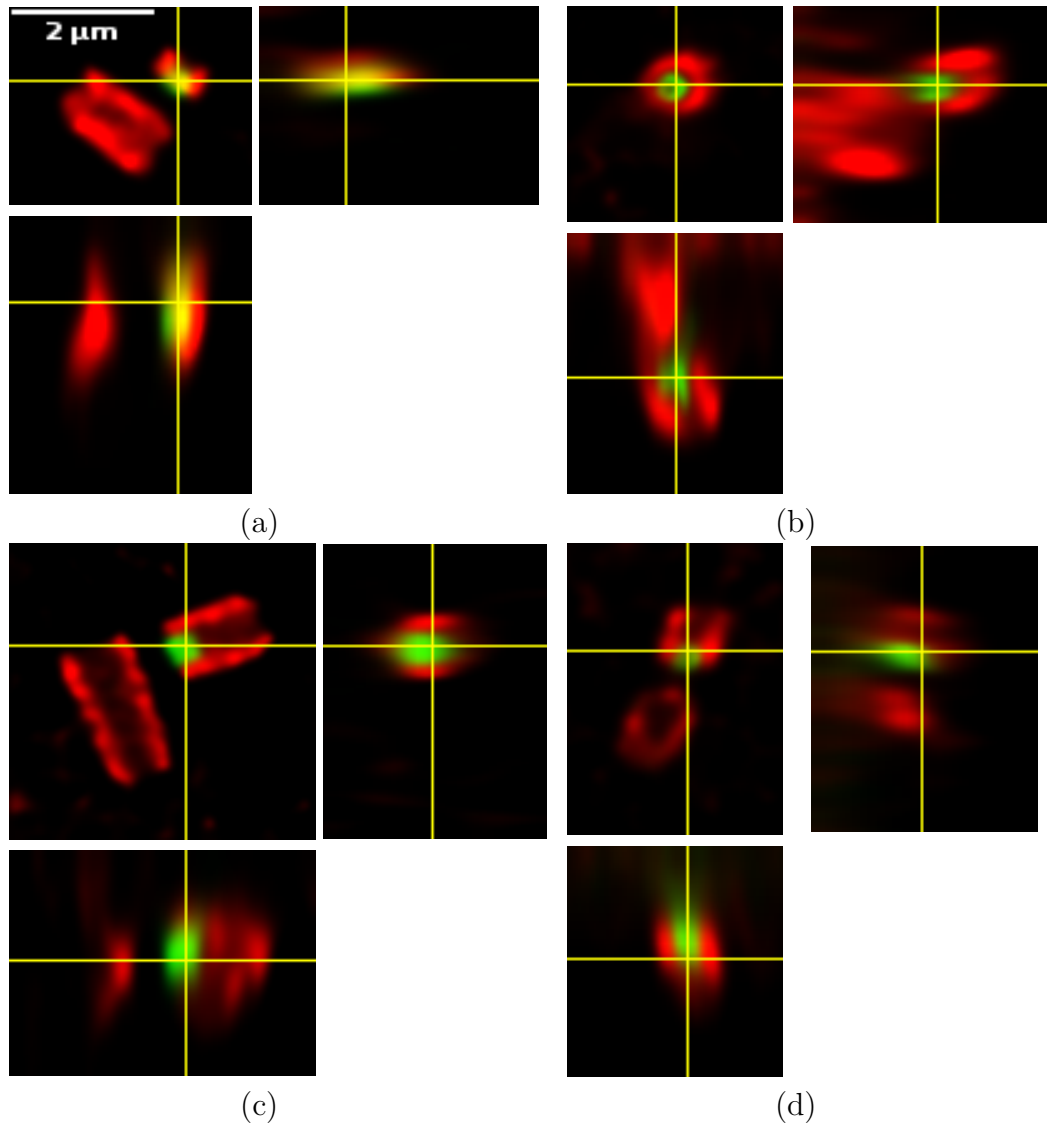


FIGURE 4.15 – Exemples de centrosomes issus de notre jeu de données. La Tubuline apparaît en rouge et Sas6 en vert. Chaque image est représentée à l'aide de 3 coupes orthogonales dans les plans XY, XZ et YZ. L'échelle, identique pour toutes les coupes et toutes les images, est indiquée sur l'image en haut à gauche. Il s'agit de l'échelle après expansion de l'échantillon. L'échelle avant expansion s'obtient en divisant par le facteur d'expansion (de valeur 4)

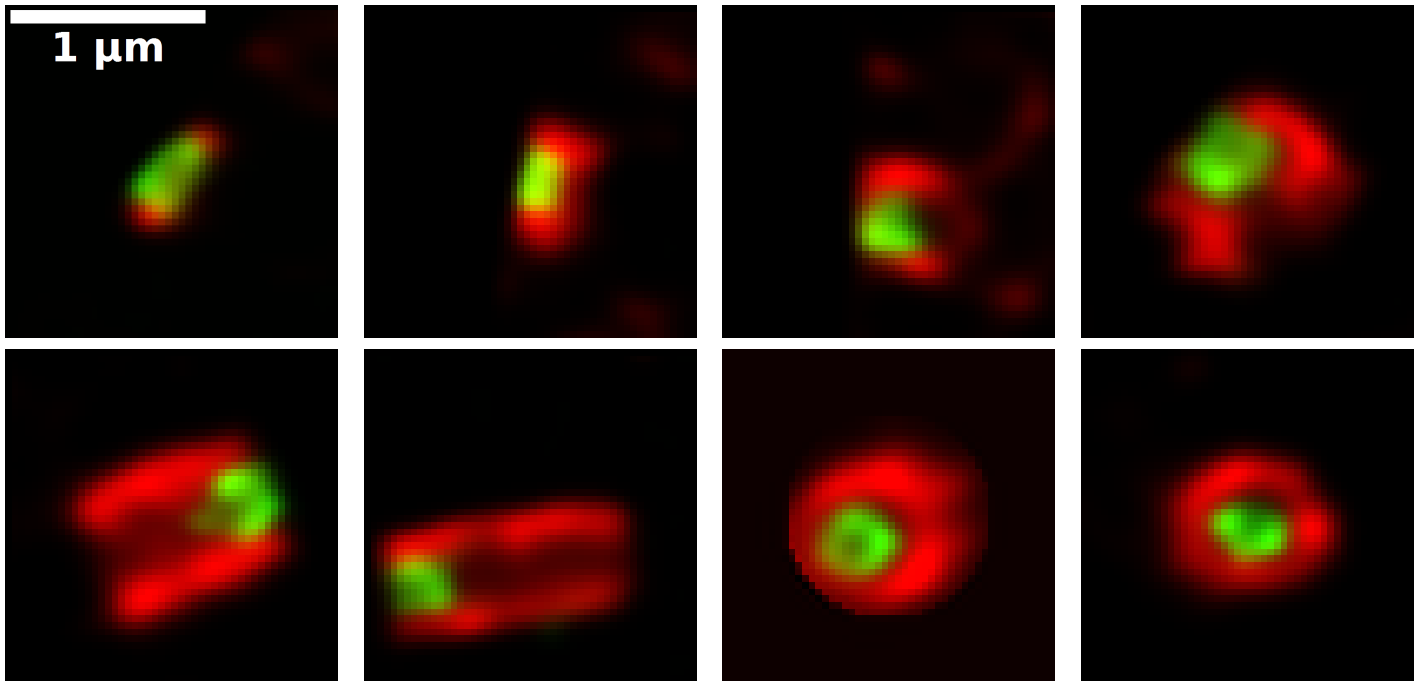


FIGURE 4.16 – Procentrioles à différents états de croissance. Ils sont visualisés par une coupe dans le plan XY. La tubuline apparaît en rouge et Sas6 en vert. L'échelle, identique pour toutes les images, est indiquée sur l'image en haut à gauche. Il s'agit de l'échelle après expansion de l'échantillon. L'échelle avant expansion s'obtient en divisant par le facteur d'expansion (de valeur 4)

l’algorithme DBSCAN ne sépare pas systématiquement le nuage de points en deux clusters comme on le souhaiterait. Parfois, il identifie des clusters correspondant à des tâches parasites de l’image. D’autre fois, le centriole mature ou le procentriole sont séparés à tort en plusieurs clusters. Nous avons donc spécifié manuellement, pour chaque image n’ayant pas été correctement séparée, quels clusters devaient être supprimés et quels clusters devaient être fusionnés.

La figure 4.16 montre des images de procentrioles, séparés du centriole mature, à différents états de croissance et selon différentes orientations.

4.4.1.3 Jeu de données monocanal protéine Tubuline

Ce prétraitement nous permet de constituer deux jeux de données. Le premier est un jeu de données monocanal de la protéine Tubuline du centriole. Il est constitué de 105 vues qui ont été sélectionnées manuellement de sorte à ce qu’elles soient d’une part de bonne qualité mais également représentatives de l’ensemble des poses et des longueurs possibles du procentriole.

4.4.1.4 Jeu de données bicanal protéines Tubuline/Sas6

Le second jeu de données est constitué de vues bicanal des protéines Tubuline et Sas6. Dans ce cas, nous n’avons pas pu effectuer une sélection manuelle des meilleures vues : nous avons en effet à disposition un nombre restreint de vues pour lesquelles ces deux protéines ont été ciblées lors de la préparation de l’échantillon. Le jeu de données, constitué de 98 vues, est donc de bien moins bonne qualité que celui utilisé pour la reconstruction monocanal. En particulier, nous avons à disposition très peu de vues de dessus de bonne qualité (3 ou 4).

4.4.2 Reconstruction monocanal de la Tubuline

4.4.2.1 Expérience

Nous avons dans un premier temps appliqué notre algorithme de reconstruction au jeu de données monocanal de la protéine Tubuline seule. Nous lançons Fluofire sur ce jeu de données, en utilisant une variable latente des conformations (z_l) de dimension 1. Nous fixons le pas de gradient associé aux poids de l’encodeur à 10^{-4} , et le pas associé à la représentation implicite sous forme de SIREN est fixé à 5×10^{-6} . La variable de régularisation du paramètre d’hétérogénéité est fixée à $\alpha = 10^{-4}$. Nous utilisons par ailleurs une taille de batch de 4. Comme pour les données synthétiques, l’algorithme d’optimisation Adam est utilisé. Nous utilisons également l’apprentissage pose à pose.

Pour les données que nous avons à disposition, il n’était pas pertinent d’utiliser l’algorithme permettant d’imposer la symétrie $C9$. En effet, cette dernière n’était pas du tout visible dans les données, elle n’avait donc aucune chance d’apparaître sur la reconstruction même en l’imposant.

4.4.2.2 Résultats

Le résultat de reconstruction, illustré à 6 états d’avancement de la croissance du procentriole, est présenté sur la figure 4.17. Les états illustrés sur la figure ont été choisis de façon à être régulièrement espacés (en considérant un ensemble de 6 valeurs de paramètre de conformation z régulièrement espacées). Nous constatons que la reconstruction est cohérente avec les données : il s’agit d’un tore dont les longueurs minimales et maximales correspondent à celles que l’on observe sur les centrioles vus de côté des données.

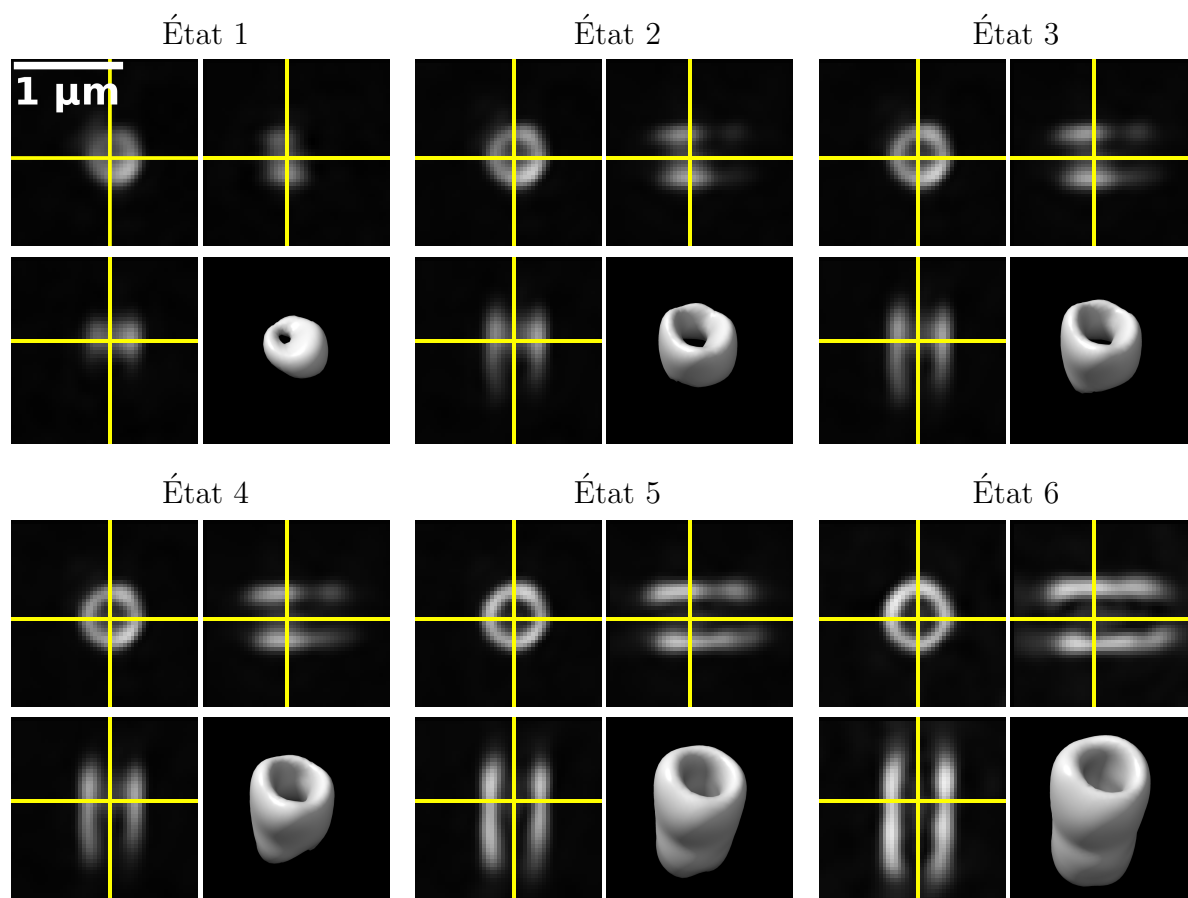


FIGURE 4.17 – Résultat de reconstruction hétérogène de la protéine Tubuline du procentre. 6 états de croissance régulièrement espacés sont illustrés. L'échelle, identique pour toutes les images, est indiquée sur l'image en haut à gauche. Il s'agit de l'échelle après expansion de l'échantillon. L'échelle avant expansion s'obtient en divisant par le facteur d'expansion (de valeur 4)

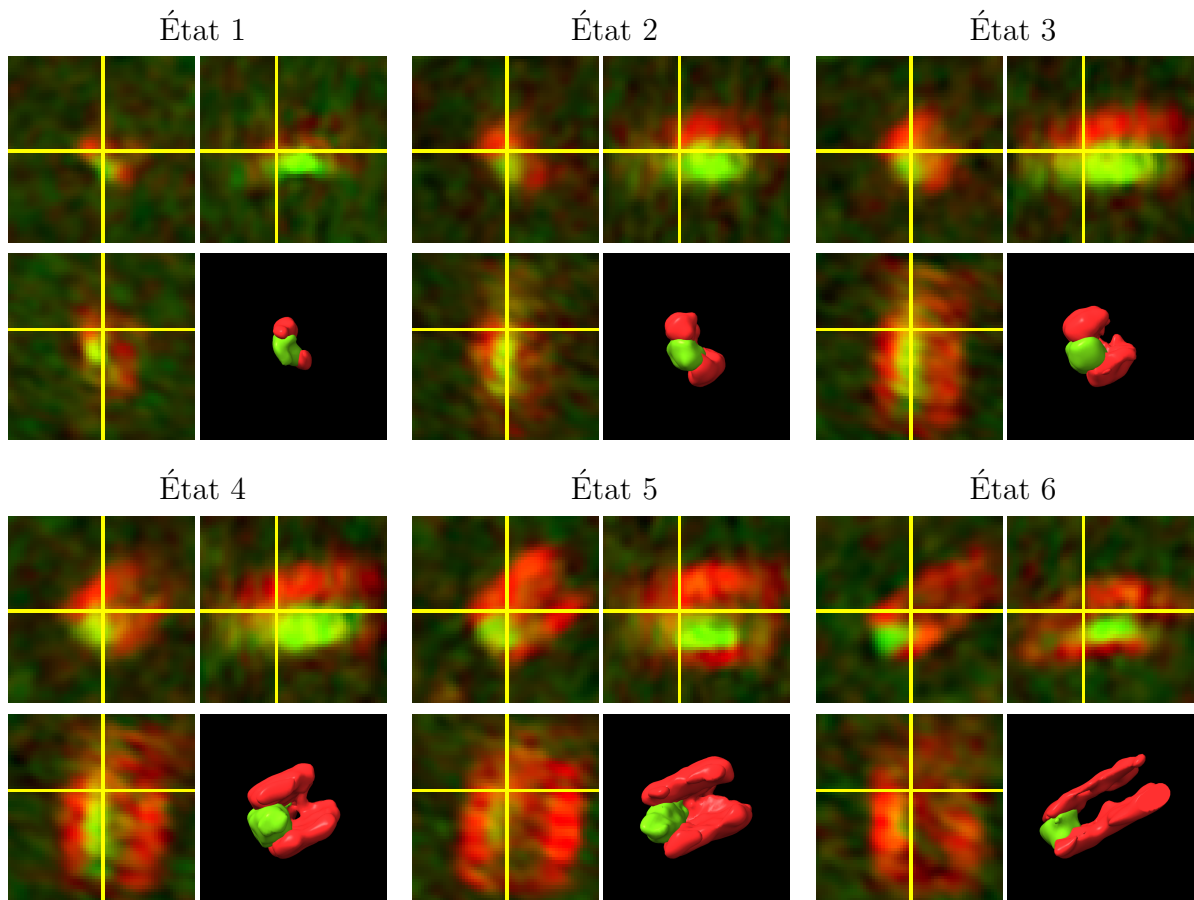


FIGURE 4.18 – Résultat de reconstruction hétérogène bicanal sur les protéines Tubuline et Sas6 du procentriole

4.4.3 Reconstruction bicanal Tubuline/Sas6

Dans un second temps nous avons tenté d'effectuer la reconstruction sur le jeu de données Tubuline/Sas6, de moins bonne qualité que les données de Tubuline seule. Les paramètres expérimentaux sont les mêmes que ceux utilisés pour la reconstruction monocanal.

Les résultats obtenus sont illustrés sur la figure 4.18. Nous constatons que la croissance du procentriole est bien reconstituée. Néanmoins, l'anisotropie de résolution n'est pas corrigée, ce qui est dû au manque de vues de dessus de bonne qualité à notre disposition.

4.5 Imposer un modèle paramétrique à la reconstruction

Pour pallier ce manque de données, nous avons décidé d'imposer une contrainte *a priori* forte à la reconstruction. Nous modélisons chacun des deux canaux comme un tore dont on estime, à chaque état de croissance, la longueur, le rayon et l'épaisseur. Nous estimons par ailleurs la position du canal Sas6 par rapport au canal Tubuline. La reconstruction à un état de croissance donnée est alors entièrement paramétrée par 7 variables, et on cherche à estimer ces dernières. Nous notons respectivement r_1 , r_2 , l_1 , l_2 , w_1 , w_2 les rayons, longueurs, et épaisseurs associées aux deux canaux. Nous notons par ailleurs p la position du canal 2 par rapport au canal 1. Ces paramètres sont

représentés sur la figure de tore bicanal 4.20. Nous supposons que les deux canaux ont le même axe de rotation, la variable p a donc une dimension 1. Le tore associé au canal 1 est placé par construction au centre de l'image. Une valeur $p = 0$ correspond à un canal 2 également placé au centre, et plus p est élevé plus le canal 2 se rapproche du bord de l'image.

Cette représentation présente l'avantage d'assurer qu'un tore soit bien reconstruit, mais elle ne permet pas la reconstruction de détails fins de la structure. Elle ne doit donc être utilisée que lorsque les données à disposition ne sont pas suffisantes pour reconstruire une structure correcte avec l'architecture de base.

4.5.1 Génération du tore à partir de ces paramètres

Nous explicitons dans cette section la manière dont le tore est construit numériquement à partir des paramètres de rayon, épaisseur et longueur.

Soit $\mathcal{G}$ une grille 3D qui est stockée à l'aide d'un tableau 4D de dimension $(S, S, S, 3)$ avec $\mathcal{G}(i, j, k) = (i, j, k)$. Nous notons respectivement $\mathcal{G}_X$, $\mathcal{G}_Y$ et $\mathcal{G}_Z$ les premières, deuxièmes et troisièmes coordonnées de $\mathcal{G}$. Formellement $\mathcal{G}_X(i, j, k) = i$, $\mathcal{G}_Y(i, j, k) = j$ et $\mathcal{G}_Z(i, j, k) = k$.

Soit $D = \sqrt{\mathcal{G}_X^2 + \mathcal{G}_Y^2}$ le tableau 3D qui stocke les distances à l'axe Z de chacun des points de la grille $\mathcal{G}$. On calcule alors pour chaque canal p ($p \in \llbracket 1, 2 \rrbracket$) :

$$\mathcal{D}_p(r_p, w_p) = \mathcal{N}(D; r_p, w_p), \quad (4.19)$$

avec $\mathcal{N}(X; \mu, \sigma)$ la gaussienne de moyenne μ et d'écart type σ évaluée en l'ensemble des points du tableau X . $\mathcal{D}_p$ est donc un tore de rayon r_p , car les points ayant une distance à l'axe proche de r_p donneront une valeur élevée dans le tableau $\mathcal{D}_p$ (et réciproquement). Par ailleurs, le tore sera d'autant plus large que w_p est élevé.

Le tore $\mathcal{D}_p(r_p, w_p)$ a une longueur de la taille de l'image alors que nous souhaitons pouvoir créer un tore de longueur quelconque l_p . Nous voudrions que le tore soit non nul pour les coordonnées (i, j, k) pour lesquelles $|G_Z(i, j, k)| \leq \frac{l_p}{2}$ et nul sinon. Nous pourrions donc appliquer une fonction porte de largeur l_p à G_Z et multiplier le résultat par $\mathcal{D}_p$ afin d'obtenir un tore de longueur l_p . Néanmoins, nous avons constaté que la non continuité de la fonction porte perturbe la reconstruction car l'algorithme de pytorch ne parvient pas à rétro-propager les gradients. Au lieu d'une fonction porte, nous avons décidé d'utiliser une fonction continue Π_H^c (avec H la largeur de la porte) d'allure similaire mais qui admet des transitions continues pour passer de 0 à 1 et de 1 à 0. Nous avons choisi d'exprimer Π_H^c (représentée sur la figure 4.19) sous la forme de la somme de deux fonctions tangentes hyperboliques ayant des sens inversés :

$$\Pi_H^c(x) = \frac{1}{2}(\tanh(\omega(x + \frac{H}{2})) + \tanh(-\omega(x - \frac{H}{2}))). \quad (4.20)$$

Le paramètre ω permet de régler la pente des transitions de 0 à 1 et de 1 à 0. En pratique, nous fixons $\omega = 50$. Le tore associé au canal 1, $\mathcal{D}_1^*(l_1, r_1, \omega_1)$, de longueur l_1 , s'obtient donc de la manière suivante :

$$\mathcal{D}_1^*(l_1, r_1, \omega_1) = \Pi_{l_1}^c(G_z) \mathcal{D}_1(r_1, \omega_1) \quad (4.21)$$

$$= \Pi_{l_1}^c(G_z) \mathcal{N}(D; r_1, \omega_1). \quad (4.22)$$

Par ailleurs, tore associé au canal 2, de longueur l_2 et situé à la position p est :

$$\mathcal{D}_2^*(l_2, r_2, \omega_2) = \Pi_{l_2}^c(G_z - p) \mathcal{D}_2(r_2, \omega_2) \quad (4.23)$$

$$= \Pi_{l_2}^c(G_z - p) \mathcal{N}(D; r_2, \omega_2). \quad (4.24)$$

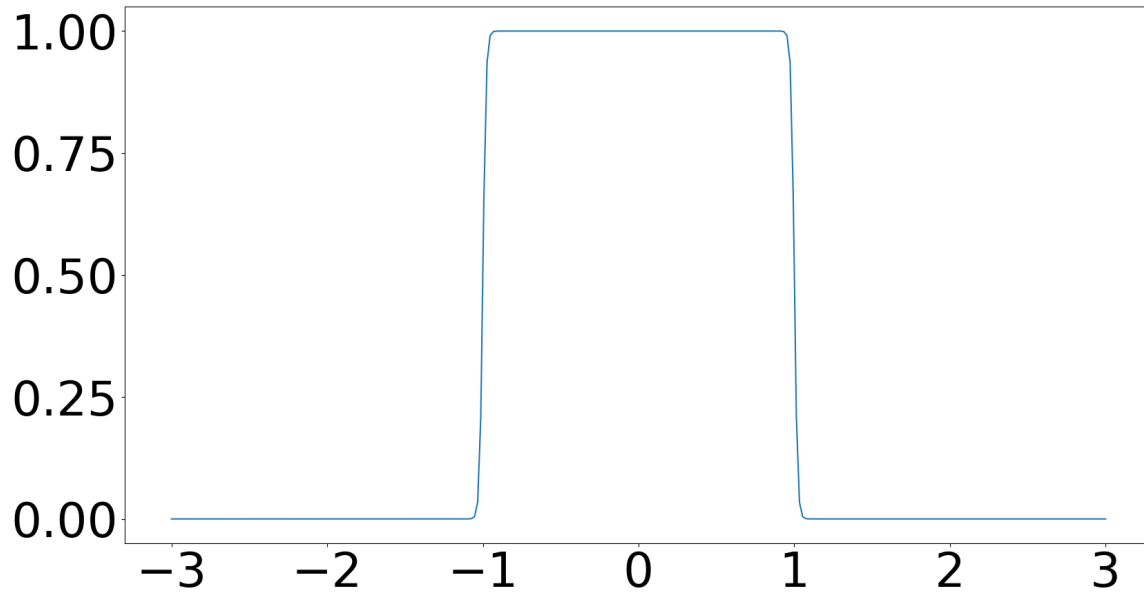


FIGURE 4.19 – Fonction rectangle continue Π_H^c , somme de deux tangentes hyperboliques.

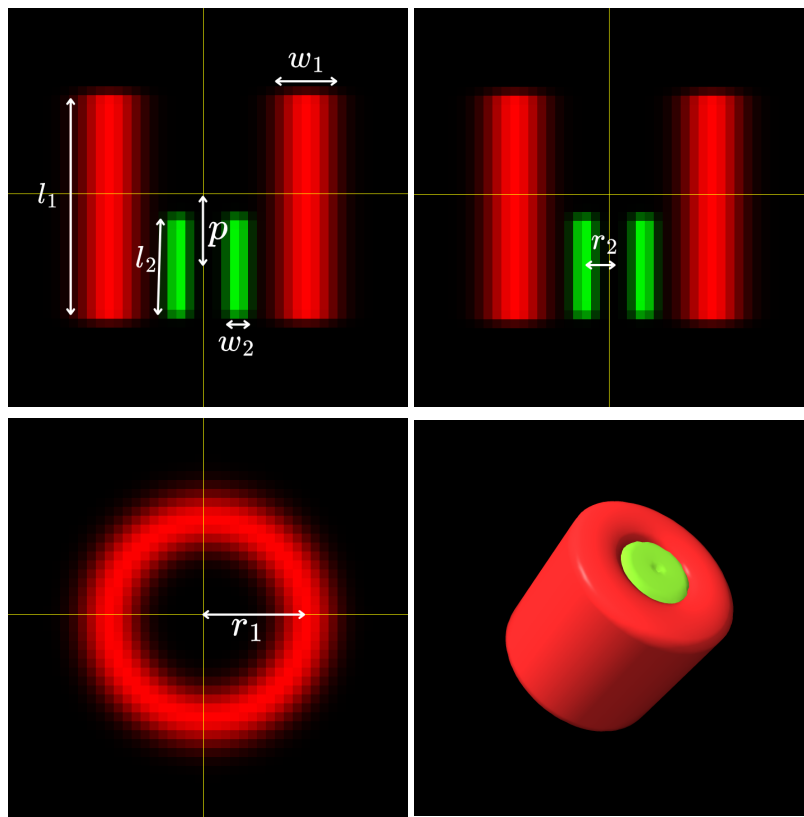


FIGURE 4.20 – Exemple de tore bicanal

La figure 4.20 illustre un exemple de tore bicanal. En pratique, nous remplaçons, dans l’architecture de base 4.1, le SIREN par les fonctions (4.22) et (4.24) qui permettent de construire un tore bicanal à partir de ses 7 paramètres, eux-même prédits à l’aide de l’encodeur. Nous contraignons les 7 paramètres du cylindre à appartenir à un certain intervalle de valeur : les longueurs, largeurs et rayons des deux tores ne doivent pas dépasser la taille de l’image et doivent être positives. Par ailleurs, la position p du tore 2 ne doit pas dépasser la moitié de la taille de l’image et doit également être positive. Cette dernière condition est indispensable, car si l’on considère deux images de tores bicaux (qui admettent les mêmes paramètres de rayons, longueurs et largeurs) l’une ayant une position du canal 2 de p et l’autre une position de $-p$ ces deux images sont équivalentes à une rotation près. Contraindre p à être positif permet donc de ne pas sur-paramétriser le problème.

Afin de contraindre les paramètres du cylindre à être dans les intervalles souhaités (notons un intervalle quelconque $[a, b]$), chacune des 7 sorties de l’encodeur sont mises en entrée d’une fonction sigmoïde et le résultat est multiplié par $b - a$ puis additionné à a .

4.5.2 Validation sur des données synthétiques

Pour valider cette variante de la méthode, nous avons commencé par la tester sur des données synthétiques.

4.5.2.1 Position p du canal 2 nulle

Génération des données Les données sont générées en utilisant le même modèle de tore que celui utilisé pour la reconstruction. Nous générons 250 vues en fixant les paramètres $r_1 = 0,5$, $r_2 = 0,15$, $w_1 = 0,04$, $w_2 = 0,1$ et $p = 0$ identiques pour toutes les vues. Par ailleurs, les longueurs des tores associés aux canaux 1 et 2 suivent le schéma de croissance illustré sur la figure 4.21.

Résultats Nous lançons sur ce jeu de données la variante de l’algorithme Fluofire qui permet d’imposer la forme de tore. A l’issue de l’optimisation, nous obtenons pour chaque vue les 7 paramètres du tore correspondant à l’état de croissance associé à cette vue. Nous représentons les résultats à l’aide de 4 graphes (voir figure 4.22). Les 3 premiers montrent respectivement les estimations des longueurs, rayons et largeurs estimés des deux canaux. Le dernier illustre l’estimation de la position du canal 2. La longueur du canal 1, qui représente l’état de croissance du centriole est en abscisse de tous les graphes. Nous constatons que les longueurs, rayons et largeurs des deux canaux sont quasiment parfaitement estimés. Pour la position p , nous constatons une erreur constante de 0,04, correspondant à la taille d’un pixel, donc relativement négligeable.

Les 7 paramètres étant correctement estimés pour tous les états de croissance, cela signifie que l’objet reconstruit l’est également, en effet la forme de tore bicanal que nous imposons à la reconstruction est entièrement définie par ces 7 paramètres.

Ces résultats sont encourageants, mais ils ont été obtenus sur une version légèrement simplifié du problème, car nous avons utilisé dans les données générées une position p du canal 2 par rapport au canal 1 nulle. Bien que cette valeur de $p = 0$ ne soit pas fournie à l’algorithme de reconstruction, cela facilite l’apprentissage. En effet, une valeur de p nulle signifie que le barycentre des deux tores (associés aux canaux 1 et 2) sont identiques. L’objet de référence admet donc une symétrie supplémentaire : il est invariant par un retournement de 180° (c’est-à-dire une rotation de 180° autour d’un axe perpendiculaire à l’axe de symétrie des tores). Dans le paragraphe suivant 4.5.2.2,

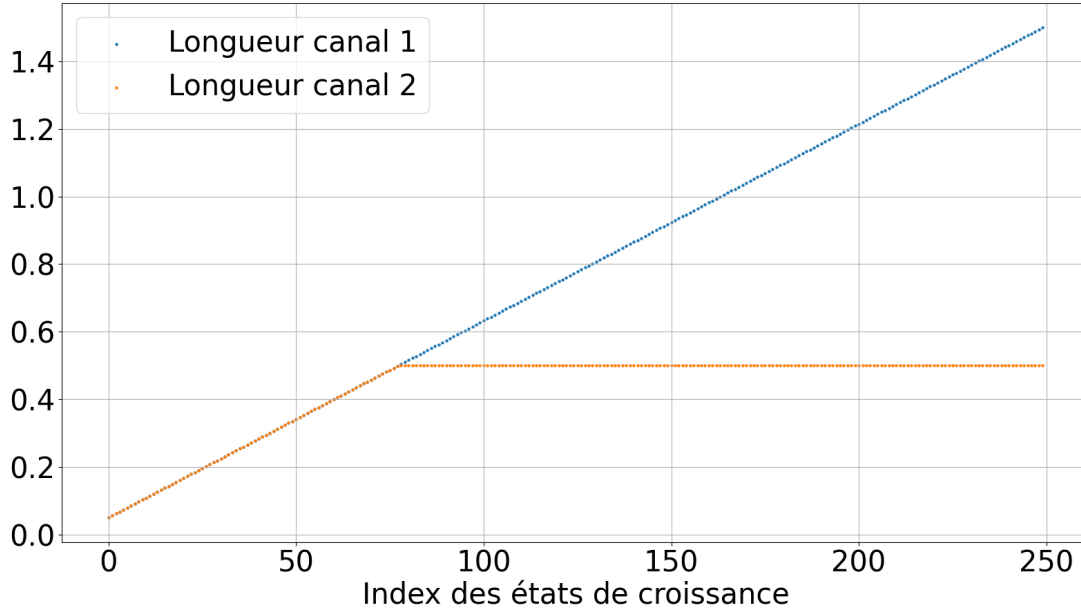


FIGURE 4.21 – Longueurs l_1 et l_2 des tores des canaux 1 et 2 pour différents états de croissance (données générées synthétiquement). Une longueur de 2 correspond à la taille de l'image.

nous émettrons une hypothèse sur la raison pour laquelle nous pensons que cette symétrie simplifie l'apprentissage.

4.5.2.2 Position p du canal 2 non nulle

Génération des données Nous générons un jeu de données très similaire à celui présenté au paragraphe 4.5.2.1, à la différence que la position n'est plus fixée à 0 mais à 0, 3.

Résultats Les résultats sont illustrés sur la figure 4.23. Un comportement problématique apparaît. Nous observons que les paramètres du canal 1 sont correctement estimés, mais pas ceux du canal 2. Ces derniers le sont seulement pour environ la moitié des vues.

Nous pensons expliquer cela par le fait que, pour une vue donnée, il existe plusieurs orientations correctes associées au canal 1 étant donné que le tore est non seulement invariant par rotation de n'importe quel angle autour de son axe de symétrie mais surtout par rotation de 180° autour d'un axe perpendiculaire à son axe de symétrie. Cependant, cette deuxième invariance n'est plus vraie lorsque le deuxième canal est ajouté. Or, le canal 1 contribue en moyenne bien plus à la fonction de coût que le canal 2 car il est plus large et plus long que ce dernier. Nous faisons donc l'hypothèse, qu'au début de l'apprentissage, l'algorithme va se baser principalement sur le canal 1 pour estimer les angles, et ces derniers vont converger vers une certaine valeur. Malheureusement, pour la moitié des vues ces angles seront uniquement adaptés au canal 1 mais ne le seront pas pour le canal 2. L'algorithme devrait alors idéalement pouvoir changer drastiquement ces angles (d'une rotation de 180°) pour qu'ils deviennent adaptés aux deux canaux. Mais il n'y parvient pas car il se retrouve bloqué dans un minimum local. Au lieu de cela, il va minimiser les longueurs des tores du canal 2 afin de réduire le plus possible leur contribution à la fonction de coût.

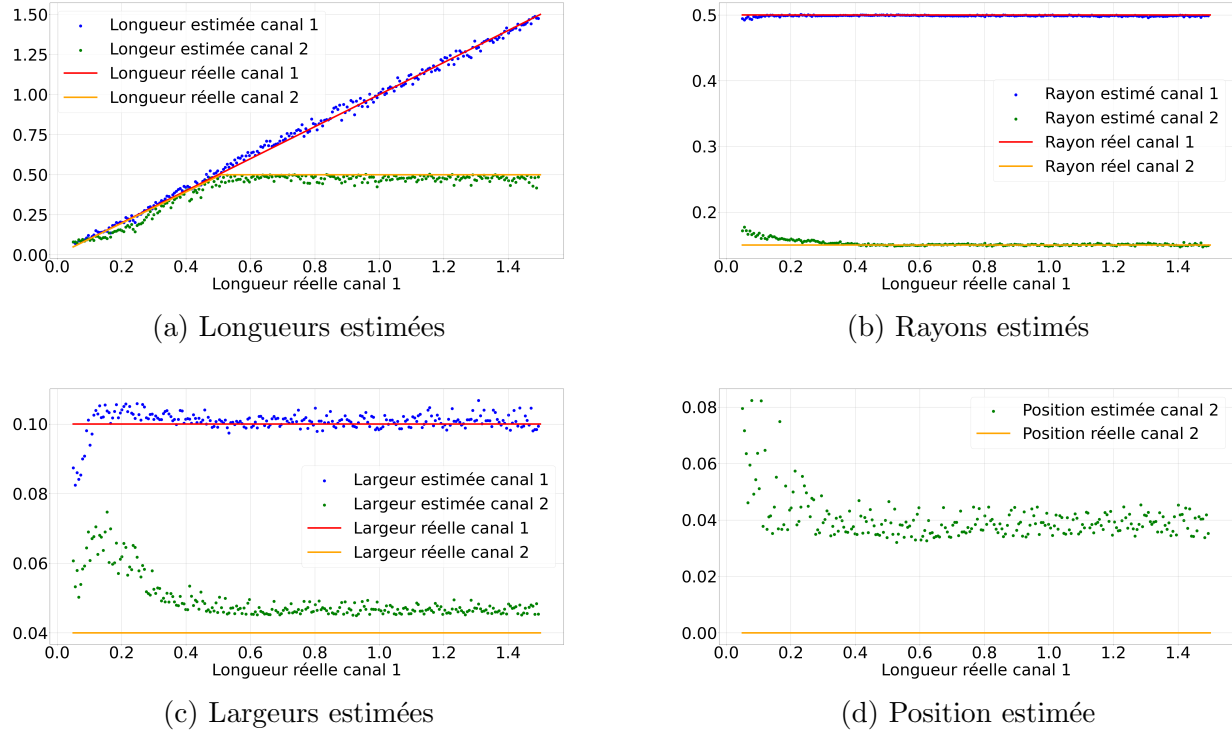


FIGURE 4.22 – Paramètres estimés des tores lorsque $p = 0$ dans les données générées. En pointillés sont représentés les paramètres estimés associés à chacune des vues du jeu de données, et les lignes représentent les paramètres réels. La longueur réelle du canal 1, qui représente l'état de croissance du centriole, est en abscisse de tous les graphes

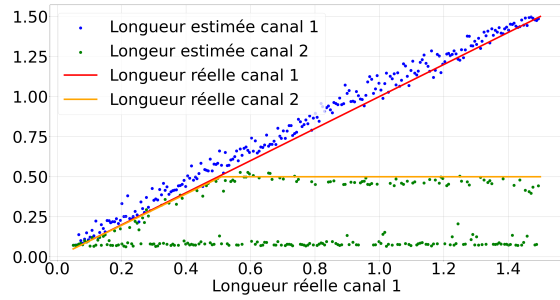
4.5.2.3 Contributions différentes des deux canaux à la fonction de coût

Pour résoudre ce problème, nous avons eu l'idée de donner un poids différent aux deux canaux dans la fonction de coût. Ce poids est calculé en faisant le rapport de la somme des pixels des vues des canaux 1 et 2. Ainsi, l'algorithme pourra se baser de manière relativement égale sur les deux canaux pour estimer les poses, réduisant les chances de tomber dans un minimum local éloigné d'une rotation de 180° de la valeur réelle.

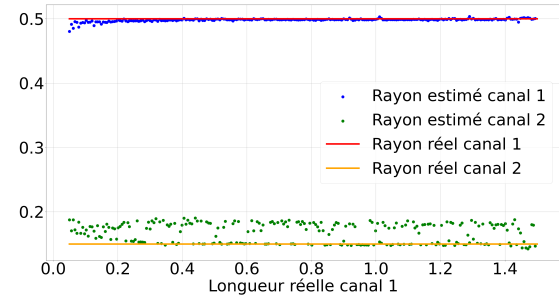
Les résultats sont illustrés sur la figure 4.24. Nous constatons que le problème précédemment rencontré n'apparaît plus. Néanmoins, nous observons un autre comportement problématique. L'algorithme ne parvient pas à estimer correctement les faibles longueurs. Nous ne sommes pas parvenus à interpréter ce dysfonctionnement.

4.5.3 Résultats obtenus sur des données réelles

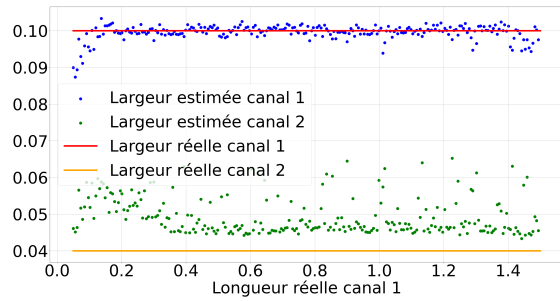
Nous avons ensuite tenté d'appliquer cette variante de la méthode de reconstruction aux données réelles présentées en section 4.6.2.1. Un résultat de reconstruction, présenté à 6 états de croissance différents, est illustré sur la figure 4.25. Le résultat obtenu n'est pas incohérent mais semble être assez loin d'être parfait. D'un côté, nous constatons que le paramètre de longueur du canal 1 est globalement bien estimé, de même pour la position p du canal 2. Néanmoins, nous constatons visuellement, que les autres paramètres ne correspondent pas très bien à ce que l'on observe dans les données.



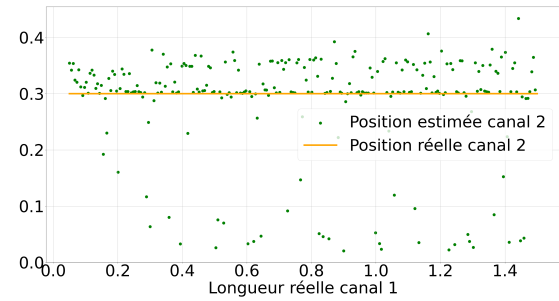
(a) Longueurs estimées



(b) Rayons estimés

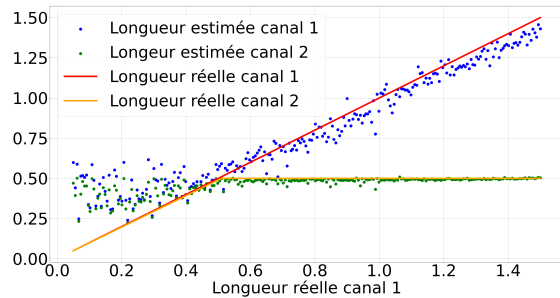


(c) Largeurs estimées

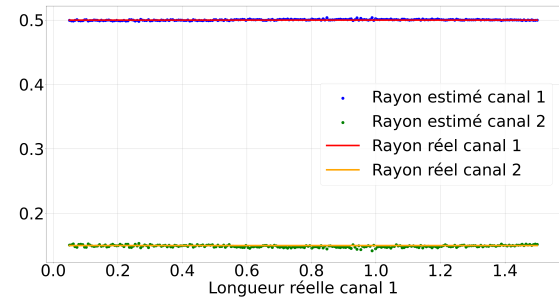


(d) Position estimée

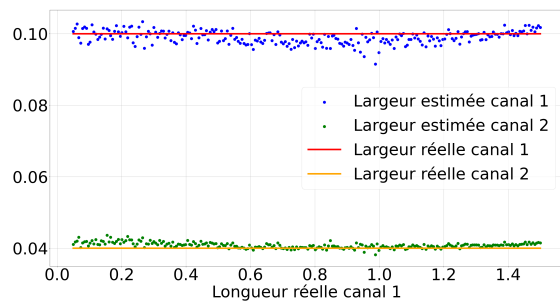
FIGURE 4.23 – Paramètres estimés des tores lorsque $p \neq 0$ dans les données générées



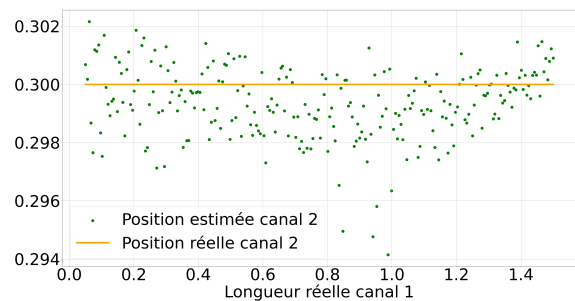
(a) Longueurs estimées



(b) Rayons estimés



(c) Largeurs estimées



(d) Position estimée

FIGURE 4.24 – Paramètres estimés des tores lorsque $p \neq 0$ dans les données générées. Un poids supérieur est associé au canal 2 dans la reconstruction afin d'éviter la convergence vers de mauvais minimum locaux.

Cette méthode visant à imposer un modèle de tore bicanal à la reconstruction n'est donc pas encore finalisée. Dans la partie 4.6.2.2, nous proposerons des perspectives de recherche susceptibles de la faire aboutir.

4.6 Résumé du chapitre et perspectives méthodologiques

4.6.1 Résumé

Dans ce chapitre, nous avons présenté notre méthode de RPI hétérogène Fluofire, dédiée à la microscopie par fluorescence. Elle se base sur une estimation amortie des poses et du paramètre de conformation et une représentation implicite du volume, constituant un système encodeur-décodeur entraîné de bout en bout. L'estimation amortie des poses permet de réduire considérablement le coût computationnel, car le problème d'estimation des poses n'a pas à être résolu de manière indépendante sur chacune des vues. Néanmoins, l'estimation peut s'avérer moins précise que lorsqu'une estimation non-amortie est utilisée, et conduire à l'obtention de minima locaux largement sous optimaux. D'autre part, une partie de l'estimation des poses peut être incluse dans le paramètre d'hétérogénéité, ce qui conduit à une structure qui tourne en même temps que ce dernier évolue. Afin d'éviter au mieux les minima locaux, et de désenchevêtrer les poses et le paramètre de conformation, tout en conservant l'avantage en termes de coût computationnel de l'estimation amortie, nous avons inclus, en plus de l'apprentissage image à image un apprentissage pose à pose. Dans ce cas, l'ordre de l'encodeur - chargé de l'estimation des poses et du paramètre de conformation - et du décodeur - chargé de l'estimation du volume - est inversé.

Nous utilisons une représentation dans un espace latent de l'hétérogénéité des conformations. Ainsi, les variations de la variable de conformation z , contenue dans cet espace latent se traduit par le passage d'une conformation à une autre dans le domaine voxélique. Ainsi, une hétérogénéité continue des conformations peut être modélisée.

Nous avons testé Fluofire sur les mêmes jeux de données homogènes que ceux présentés au chapitre 3. Sur ces données, les résultats obtenus par Fluofire sont similaires à ceux obtenus avec la méthode SHiReVol pour deux jeux de données et légèrement de moins bonne qualité pour les deux autres. Bien que Fluofire puisse s'appliquer également à la reconstruction homogène, l'utilisation de la méthode SHiReVol nous semble donc plus adaptée lorsqu'on a à disposition un jeu de données homogène.

Nous avons ensuite évalué Fluofire sur un jeu de données hétérogènes admettant un degré de liberté, simulé à partir d'un modèle de centriole existant. Les résultats de reconstruction obtenus sont très proches de l'objet vérité terrain, bien que la fausse symétrie ne soit pas parfaitement reconstruite. Le paramètre de conformation est également correctement estimé.

Nous avons également testé Fluofire sur des données simulées, générées à partir du même modèle de centriole, mais admettant cette fois-ci deux degrés de liberté : la longueur et le rayon du centriole. L'analyse de la structure de l'espace latent et des résultats visuels montre clairement que Fluofire parvient à capter correctement les deux degrés de liberté.

Nous avons aussi mis en évidence la forte robustesse au bruit de la méthode malgré l'absence de terme de régularisation. Cela confirme les résultats d'analyse de la robustesse au bruit du chapitre précédent : l'effet de moyenne dû à l'utilisation d'un grand nombre de vues pour la reconstruction permet d'éviter d'inclure un terme de régularisation.

Nous avons inclus à Fluofire une option permettant d'imposer la symétrie $C9$ à la reconstruction, pouvant être appliquée en raffinement. En particulier, elle nécessite que l'axe de symétrie de

la reconstruction soit connu, elle ne peut donc pas être utilisée en reconstruction *ab-initio*. Nous avons évalué cette option sur les mêmes données synthétiques que celles utilisées précédemment mais en retirant plus ou moins de bras au centriole.

Après avoir validé la méthode sur des jeux de données générées synthétiquement, nous l'avons testé sur des jeux de données réelles de centriole. La reconstruction monocanal de la protéine Tubuline du centriole nous semble cohérente avec les vues d'entrée. Cette évaluation résulte simplement d'appréciation visuelle qualitative : une structure de tore qui croît avec le paramètre de conformation est reconstruite, et les tailles minimales et maximales correspondent à ce que l'on observe dans les vues du jeu de données. Néanmoins, nous n'avons pas validé quantitativement les résultats (nous en reparlerons en conclusion de la thèse).

Nous ne sommes pas parvenus à obtenir une reconstruction correcte sur les données hétérogène bicanal des protéines Tubuline/Sas6 du procentriole. Fluofire parvient à reconstruire correctement la croissance de la structure, mais l'anisotropie de résolution n'est pas corrigée. Cela est dû au manque de données à notre disposition, et en particulier au manque de vue de dessus de bonne qualité.

Pour pallier ce manque de données, nous avons développé une méthode permettant d'imposer une forme de tore à la reconstruction dont on cherche à estimer les paramètres de longueurs, largeurs, rayons et positions. Néanmoins, les résultats obtenus ne sont pas satisfaisants. Sur des données simulées, la méthode ne parvient pas à estimer correctement les faibles valeurs de longueurs. Sur les données réelles, seuls le paramètre de longueur du premier canal et la position p du canal 2 sont correctement estimés. Nous ne sommes pas parvenus à identifier la cause de cette convergence vers un mauvais minima local, mais nous faisons l'hypothèse que cela est dû au trop grand nombre de paramètres d'hétérogénéité à estimer conjointement.

4.6.2 Perspectives méthodologiques

4.6.2.1 Segmentation automatique du procentriole

Dans le chapitre 4, nous nous sommes intéressés à la reconstruction de la croissance du procentriole. Nous avons donc dû effectuer une segmentation afin de séparer le procentriole de son centriole mature voisin. Notre méthode de segmentation, exposée en partie , était relativement laborieuse car elle nécessite une certaine intervention manuelle. Il pourrait être intéressant de développer une méthode de segmentation basée sur de l'apprentissage profond qui permettrait d'automatiser au maximum la tâche. Nous pourrions générer synthétiquement des couples centriole mature/procentrioles et apprendre à un réseau à les segmenter, en espérant que le réseau puisse généraliser la tâche sur des données réelles.

4.6.2.2 Imposer un tore : aboutissement de la méthode

Nous avons présenté en partie 4.5 une méthode permettant d'imposer une forme de tore bicanal à la reconstruction. Le manque de données à notre disposition nous a encouragé à suivre cette voie. Néanmoins, les résultats obtenus sont encore assez loin d'être satisfaisants, à la fois sur les données synthétiques et réelles. Nous avons envisagé plusieurs pistes de recherche pour la suite.

D'abord, nous souhaitons vérifier que les erreurs d'estimation des paramètres du tore associé au deuxième canal sont bien causées par la symétrie du modèle paramétrique. Pour cela, nous envisageons d'effectuer des expériences en considérant un modèle paramétrique de tore qui n'est pas invariant par une rotation de 180° autour d'un axe perpendiculaire à son axe de symétrie.

Nous envisageons également de tester une reconstruction en deux temps. Dans un premier temps, nous reconstruisons uniquement sur le premier canal et enregistrons les poses obtenues. Dans un second temps, une reconstruction conjointe sur les deux canaux est effectuée. Lors de cette deuxième étape, les poses estimées lors de la première expérience sont utilisées mais elles ne sont pas fixées à ces valeurs : l’algorithme cherchera à estimer pour chaque vue une variable binaire indiquant si oui ou non une rotation de 180° doit être effectué.

Ce travail sur l’amélioration de la méthode permettant d’imposer un tore fait actuellement l’objet d’un stage de M1.

4.6.2.3 Analyse des espaces latents multidimensionnels

Nous avons présenté au paragraphe 4.3.2.2 des résultats sur données synthétiques en utilisant un espace latent admettant plusieurs dimensions (2 en l’occurrence). Les données générées admettaient quant à elles, soit un soit deux degrés de liberté. En observant la structure de l’espace latent, et en labellisant les points avec les valeurs réelles des paramètres de conformation, nous observons que le degré (quand il y a un degré de liberté dans les données) ou les deux degrés (quand il y en a deux) de liberté sont correctement captés par l’algorithme de reconstruction. À la fin de l’apprentissage, nous appliquons une ACP à l’espace latent afin d’identifier les directions des degrés de liberté. Néanmoins, il semble que l’ACP ne soit pas tout à fait le choix le plus judicieux. En effet, les degrés de liberté semblent ne pas tout à fait suivre des droites, mais plutôt des lignes courbes. Une analyse non linéaire de la structure de l’espace latent pourrait donc être effectuée afin d’identifier les équations de ces courbes.

4.6.2.4 Estimation non amortie des poses : adaptation de CryoDRGN

Comme nous l’avons déjà mentionné, notre méthode de RPI hétérogène Fluofire est une adaptation au cas de la fluorescence de CryoFIRE et est basé sur une estimation amortie des poses et du paramètre de conformation. L’inférence amortie réduit la précision des résultats au profit du temps de calcul. Nous pourrions développer une méthode utilisant une inférence non amortie des poses (tout en gardant l’inférence amortie du paramètre de conformation). Cela reviendrait à adapter CryoDRGN au cas de la fluorescence. Ainsi, l’utilisateur aurait deux possibilités, en fonction de s’il préfère privilégier le temps de calcul ou la qualité de la reconstruction.

4.6.2.5 Hétérogénéité modélisée par un champ de déformation (non-compositionnelle)

La méthode Fluofire est en théorie censée pouvoir être adaptée à tout type d’hétérogénéité (discrète, continue compositionnelle, continue non-compositionnelle). Néanmoins, nous avons présenté dans ce chapitre uniquement des résultats sur des données hétérogènes compositionnelles. Bien que cela ne corresponde pas à notre cas d’application final, nous l’avons également testée sur des données simulées hétérogènes non-compositionnelles (ce qui signifie que l’ensemble des structures peuvent s’obtenir par déformation d’une structure canonique unique) de pore nucléaire. Les données ont été générées à partir du code développé par les auteurs de [108]. Les résultats obtenus sont assez loin d’être satisfaisants. On a en effet constaté que, malgré l’utilisation de la fonction de coût pose-à-pose, le paramètre de conformation capture une partie de la position et de l’orientation de l’objet reconstruit (la fonction de coût pose-à-pose a néanmoins permis de grandement réduire la proportion de l’orientation et de la position estimée par le paramètre de conformation). Pour ce cas d’application, il pourrait donc être intéressant d’adapter à la fluorescence une des méthodes

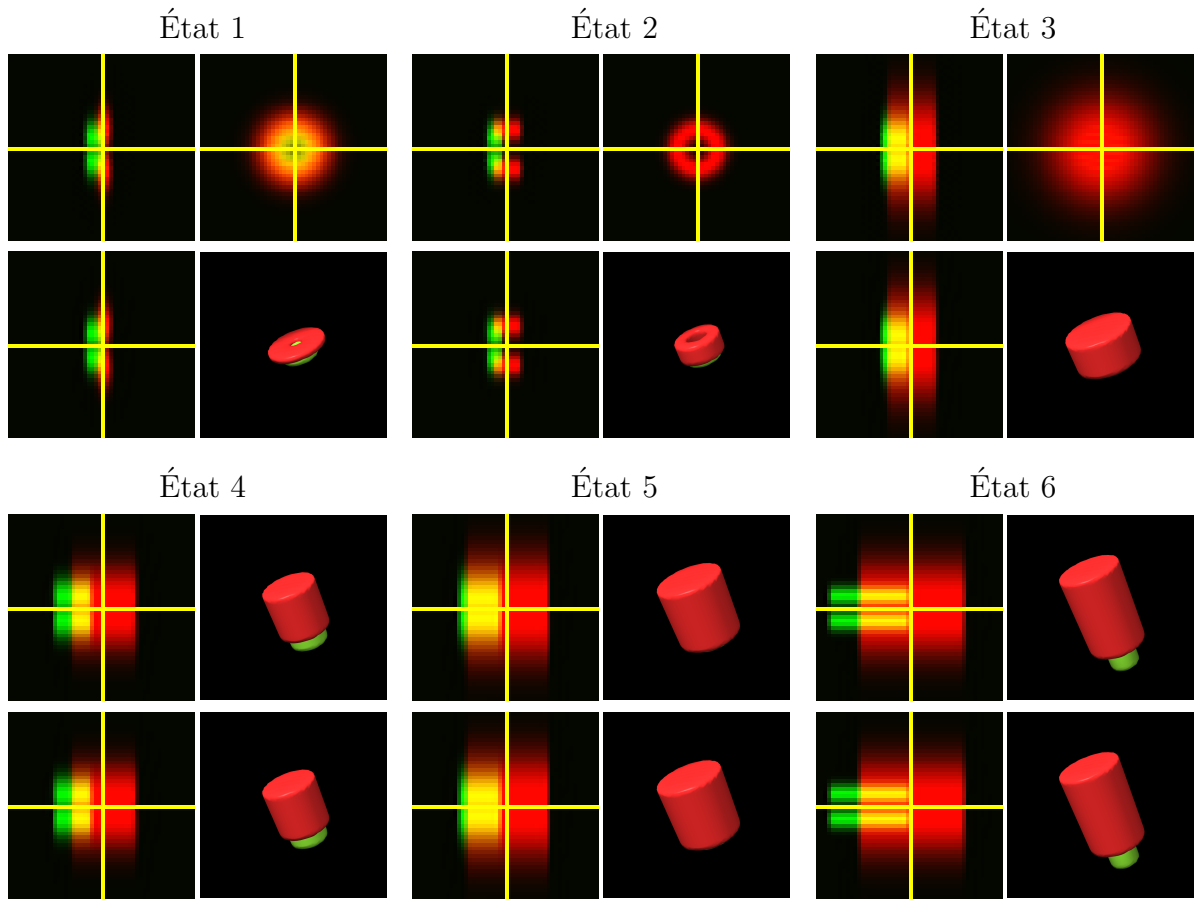


FIGURE 4.25 – Cylindre imposé données réelles

dédiées à la cryo-ME qui se base sur l'estimation d'un champ de déformation (3DFlex [81] ou DynaMight [82] par exemple).

Au delà de ces perspectives méthodologiques spécifiques à la reconstruction hétérogène, nous présentons dans le chapitre 5 des perspectives plus générales pouvant être communes aux deux cas d'applications que nous avons abordés durant cette thèse (reconstruction homogène et reconstruction hétérogène).

Chapitre 5

Discussions et perspectives

5.1 Bilan

Durant ma thèse, j’ai développé des méthodes de reconstruction à partir de particules isolées dédiées aux modalités convolutionnelles de microscopie par fluorescence (microscopies à champ large, confocal, STED).

Nous avons présenté au chapitre 3 une méthode de reconstruction homogène (*i.e.* qui suppose que toutes les vues sont issues du même objet). Cette dernière est basée sur une reformulation du problème d’optimisation conjoint original, par rapport au volume, aux rotations et aux translations en un schéma d’optimisation hiérarchique qui découple la minimisation par rapport à chaque groupe de variables, en 3 niveaux imbriqués. Nous avons développé 3 variantes de la méthode, l’une basée sur une représentation voxélique du volume (SHiReVol), la seconde utilise une représentation implicite sous forme de mélange de gaussienne du volume (ShiReGmm), et la troisième utilise une représentation sous forme de mélange de gaussiennes à la fois du volume et des vues (ShiReGmm2). Les résultats obtenus sur des données générées synthétiquement à partir d’objets de référence nous ont permis d’écarter la troisième variante, dont les performances étaient nettement inférieures aux deux premières. Nous avons également mis en évidence des résultats sur un jeu de données réelles de la protéine Cep164 du centriole. A la fois sur les données générées synthétiquement et les données réelles, nous surpassons l’état de l’art de la RPI en microscopie par fluorescence.

Au chapitre 4, nous avons présenté notre méthode de RPI hétérogène Fluofire. On ne suppose plus dans ce cas que les vues sont issues du même objet : il y a hétérogénéité des conformations. Cela rend le problème bien plus complexe, car en plus des rotations et des translations, l’ensemble des conformations doivent être estimées. Notre cas d’application est la croissance du procentriole lors de l’interphase de la mitose : il s’agit d’un cas d’hétérogénéité continu (l’objet analysé existe sous un nombre continu d’état stable) et non-compositionnelle (on ne peut pas modéliser le passage d’une conformation à l’autre par l’application d’un champ de déformation). Ce cas d’application nous a motivé à faire reposer Fluofire sur l’utilisation d’une représentation implicite des conformations sous la forme d’un SIREN. Les méthodes de RPI dédiées à la cryo-ME CryoDRGN et CryoFIRE ont en effet montré que cette représentation se prêtait bien au cas de l’hétérogénéité continue non-compositionnelle. Nous avons mis en évidence des résultats satisfaisants sur des jeux de données synthétiques, et un jeu de données réelles monocal de la protéine Tubuline du centriole. Fluofire a en revanche échoué sur le jeu de données bicanal Tubuline/Sas6.

La méthode Fluofire peut également s’appliquer à la reconstruction homogène. Néanmoins, si les données d’entrées ne présentent pas d’hétérogénéité des conformations, il nous semble plus

pertinent d'utiliser la méthode SHiReVol présentée au chapitre 3. En effet, cette méthode présente l'avantage d'être légère : elle n'a pas besoin de carte graphique de 12Go pour fonctionner. D'autre part, SHiReVol utilise une estimation non amortie des poses alors que Fluofire se base sur une estimation amortie, qui présente l'avantage de réduire le temps de calcul mais au détriment de la précision de l'estimation des poses, ce qui se traduit par une reconstruction de moins bonne qualité.

5.2 Perspectives générales

Nous avons déjà présenté en conclusion des chapitres 3 et 4 des perspectives spécifiques aux méthodes Fluofire et SHiReVol. Concernant les perspectives du chapitre 3, nous avons évoqué la possibilité d'utiliser une marginalisation par rapport aux orientations plutôt qu'une optimisation hiérarchique à 3 niveaux. Une autre perspective du chapitre est d'imposer une contrainte stéréochimique à la représentation sous forme de mélange de gaussiennes. Au chapitre 4, nous avons évoqué les perspectives de segmentation automatique du procentre, d'analyse des espaces latents multidimensionnels, d'estimation non-amortie, d'hétérogénéité modélisée par un champ de déformation ainsi que de faire aboutir la méthode permettant d'imposer un tore à la reconstruction. Pour plus de détails, nous renvoyons le lecteur aux sections concernées (sections 3.7.2 et 4.6.2).

Nous exposons dans cette section conclusive quatre autres perspectives pouvant être communes à nos deux cas d'application (reconstructions homogène et hétérogène).

5.2.1 Acquisition de données

En section 3.6.2, nous avons montré des résultats de reconstruction homogène sur la protéine Cep164 du centriole. Ils ont permis de mettre en évidence le bon fonctionnement de notre méthode SHiReVol. Des reconstructions sur des données d'entrée ayant une meilleure résolution seraient toutefois bénéfiques. Il est théoriquement possible d'acquérir de telles données. En effet, les données à notre disposition ont été obtenues par combinaison de microscopie par expansion, conduisant à une dilatation d'un facteur 4, et d'un microscope optique à champ large. Bien que l'expansion permette d'améliorer la résolution des images, le microscope à champ large ne peut atteindre les résolutions des techniques de microscopie optique modernes telles que STED. De même pour la reconstruction hétérogène, bien que nous ayons déjà obtenu des résultats prometteurs, des données de meilleure qualité et en plus grande quantité, notamment des vues de dessus, permettraient d'améliorer significativement la correction de l'anisotropie de résolution. Les biologistes du CentriolLab avec qui nous collaborons travaillent actuellement à l'acquisition de nouvelles données utilisant des techniques avancées, ce qui nous fournira davantage de données à traiter et ouvrira des opportunités de découvertes biologiques.

Le manque de données à notre disposition explique également l'absence d'évaluation quantitative des performances sur les jeux de données réelles. L'approche standard pour évaluer la qualité d'une méthode de RPI consiste à séparer le jeu de données en deux sous-ensembles, à effectuer une reconstruction sur chacun d'eux et à comparer les deux reconstructions à l'aide de la métrique FSC, présentée au paragraphe 3.6.1.2.2. Cependant, pour qu'un résultat de FSC soit correctement interprétable, il est nécessaire de disposer d'un nombre suffisant de vues afin que les distributions des orientations (et des paramètres de conformation dans le cas de la reconstruction hétérogène) des deux sous-ensembles soient très proches l'une de l'autre. L'acquisition d'un plus grand nombre

de données ouvrirait également la perspective de réaliser des analyses de FSC sur des données réelles, permettant ainsi une évaluation quantitative plus précise de la qualité des reconstructions.

5.2.2 Logiciel Scipion

Comme évoqué en introduction, cette thèse s’inscrit dans un contexte plus général : le projet SPfluor. En particulier, l’ingénieur de recherche Jean Plumail a été chargé d’intégrer le code que j’ai développé au sein du logiciel Scipion, habituellement utilisé pour la RPI en cryo-ME. En plus des méthodes de reconstruction *ab-initio* homogène et hétérogène présentées durant cette thèse, le logiciel Scipion intégrera la méthode de détection 3D des vues, développée par Luc Vedrenne [26], ainsi que la méthode de raffinement développée par Denis Fortun.

Lorsqu’il sera rendu public, ce logiciel permettra l’utilisation de nos méthodes de détection et de reconstruction par des biologistes qui ne sont pas familiers avec l’informatique. Nos méthodes pourront ainsi contribuer à de nouvelles découvertes en biologie structurale.

5.2.3 Estimation de la réponse impulsionnelle

Une hypothèse importante que nous avons faite tout au long de cette thèse est que la réponse impulsionnelle (PSF) du microscope est connue et identique pour toutes les vues. Cette dernière est obtenue par acquisition de l’image d’une source ponctuelle par le microscope. Néanmoins, son estimation peut s’avérer être imprécise, ce qui aura pour conséquence de dégrader la qualité de la reconstruction. Par ailleurs, la PSF n’est pas toujours parfaitement identique pour toutes les vues. En effet, ces dernières sont extraites de grandes images qui contiennent un grand nombre de vues, et la réponse impulsionnelle est en générale légèrement plus large sur les bords. Il pourrait donc être pertinent d’intégrer l’estimation de la PSF à nos algorithmes de reconstruction, ou mieux encore l’estimation d’une PSF pour chacune des vues.

Une première approche pourrait être d’estimer la PSF en prétraitement à l’aide d’une méthode de déconvolution aveugle appliquée sur chacune des vues. Mais son estimation pourraient également s’effectuer conjointement aux autres paramètres. Nous pourrions par exemple adapter Fluofire de sorte à ce que l’encodeur prédisent, en plus des poses et du paramètre de conformation, les écart-types de la PSF associée à une vue d’entrée selon l’axe du microscope et dans le plan latéral. Contrairement aux poses, une connaissance *a-priori* sur la PSF devra tout de même être donnée au modèle d’apprentissage : nous savons que les PSFs associées à chaque vues doivent être proches de la PSF mesurée expérimentalement. Nous ajouterions donc une contrainte de régularisation sur ces écart-types pour qu’ils ne s’écartent pas trop des écart-types de la PSF mesurée.

5.2.4 Prise en compte de la déformation liée à l’expansion

Les images que nous avons utilisées pour la reconstruction ont été acquises par microscopie par expansion : un gel est introduit dans l’échantillon afin de grossir les particules observées. Ainsi, des détails plus fins peuvent être visualisés. Néanmoins, l’inconvénient de cette procédure d’acquisition est qu’elle introduit des déformations parasites à la structure. Ces dernières s’ajoutent aux déformations dues à l’hétérogénéité des conformations. Notre méthode Fluofire n’a, à l’heure actuelle, aucun moyen de distinguer ces deux types de déformations, ce qui introduit inévitablement un biais dans la reconstruction. Une perspectives de travaux futurs pourrait être de modéliser les déformations parasites liées à l’expansion et d’inclure cette connaissance dans le modèle d’apprentissage. Ainsi, notre méthode de reconstruction hétérogène pourrait estimer de manière séparée

les déformations liées à l'hétérogénéité des conformations ; et notre méthode de reconstruction homogène pourrait utiliser cette information afin de corriger les déformations parasites et donc obtenir une reconstruction plus proche de l'objet original (avant expansion).

Chapitre 6

Appendices

6.1 Calcul des gradients dans le cas d'une représentation voxélique

Nous présentons ici le calcul du gradient de l'énergie E_l par rapport au volume $\hat{f}$ utilisé dans l'équation de mise à jour de la descente de gradient stochastique de la reconstruction homogène (3.12) (méthode SHiReVol). On a :

$$E_l(f, \theta_l, t_l) = \|\hat{y}_l - \hat{h}\rho_{t_l}\mathcal{R}_{\theta_l}(\hat{f})\|^2 \quad (6.1)$$

$$= \int [\hat{y}_l(\omega) - \hat{h}(\omega)\rho_{t_l}(\omega)\hat{f}(R_{\theta_l}^T\omega)]^2 d\omega. \quad (6.2)$$

Soit ω_0 une fréquence particulière, calculons le gradient en ω_0 : $\nabla_{\hat{f}}E_l(\omega_0) = \frac{\partial E_l}{\partial \hat{f}(\omega_0)}$. Dans l'intégrale (6.2), seul le terme qui vérifie $R_{\theta_l}^T\omega = \omega_0$ dépend de $\hat{f}(\omega_0)$, on a en effet :

$$\frac{\partial \hat{f}(R_{\theta_l}^T\omega)}{\partial \hat{f}(\omega_0)} = \begin{cases} 1 & \text{if } R_{\theta_l}^T\omega = \omega_0 \\ 0 & \text{otherwise} \end{cases}. \quad (6.3)$$

Nous avons donc :

$$\frac{\partial E_l}{\partial \hat{f}(\omega_0)} = 2 \left[\rho_{R_{\theta_l}t_l}(\omega_0)\hat{h}(R_{\theta_l}\omega_0)\hat{f}(\omega_0) - \hat{y}_l(R_{\theta_l}\omega_0) \right] \rho_{R_{\theta_l}t_l}(\omega_0)\hat{h}(R_{\theta_l}\omega_0) \quad (6.4)$$

$$= 2 \left[\rho_{R_{\theta_l}t_l}(\omega_0)\mathcal{R}_{-\theta_l}(\hat{h})(\omega_0)\hat{f}(\omega_0) - \mathcal{R}_{-\theta_l}(\hat{y}_l)(\omega_0) \right] \rho_{R_{\theta_l}t_l}(\omega_0)\mathcal{R}_{-\theta_l}(\hat{h})(\omega_0). \quad (6.5)$$

Finalement, le gradient peut s'écrire

$$\nabla_{\hat{f}}E_l = 2\rho_{R_{\theta_l}t_l}\mathcal{R}_{\theta_l}^{-1}(\hat{h}) \left[\rho_{R_{\theta_l}t_l}\mathcal{R}_{\theta_l}^{-1}(\hat{h})\hat{f} - \mathcal{R}_{\theta_l}^{-1}(\hat{y}_l) \right]. \quad (6.6)$$

6.2 Calcul des gradients dans le cas d'une représentation sous forme de mélange de gaussiennes

Nous présentons ici le calcul des gradients de $\gamma \mapsto E(\gamma, \Theta, T)$ (voir équation (3.47)) par rapport à chaque coefficient α_p et chaque centre μ_p des gaussiennes (méthode ShiReGmm). Pour rappel, on a :

$$E(\gamma, \Theta, T) = \sum_{l=1}^N \|h * \mathcal{T}_{t_l}(\mathcal{R}_{\theta_l}(f)) - y_l\|^2 = \sum_{l=1}^N \|y_l - C \sum_{p'=1}^{N_g} \alpha_{p'} \mathcal{N}(R_{\theta_l} \mu_{p'} + t_l, \sigma^2 I + \Sigma_h)\|^2, \quad (6.7)$$

— Dérivée de E_l par rapport à α_p :

$$\frac{\partial E_l}{\partial \alpha_p} = 2 \left\langle \frac{\partial(h * \mathcal{T}_{t_l}(\mathcal{R}_{\theta_l}(f)))}{\partial \alpha_p}, h * \mathcal{T}_{t_l}(\mathcal{R}_{\theta_l}(f)) - y_l \right\rangle \quad (6.8)$$

$$= 2 \sum_{x \in \Omega} \frac{\partial(h * \mathcal{T}_{t_l}(\mathcal{R}_{\theta_l}(f)))(x)}{\partial \alpha_p} (h * \mathcal{T}_{t_l}(\mathcal{R}_{\theta_l}(f))(x) - y_l(x)) \quad (6.9)$$

Comme $h * \mathcal{T}_{t_l}(\mathcal{R}_{\theta_l}(f)) = C \sum_{p'=1}^N \alpha_{p'} \mathcal{N}(R_{\theta_l} \mu_{p'} + t_l, \sigma^2 I + \Sigma_h)$, on a $\frac{\partial \mathcal{T}_{t_l}(\mathcal{R}_{\theta_l}(f))(x)}{\partial \alpha_p} = \mathcal{N}(x; R_{\theta_l} \mu_p + t_l, \sigma^2 I + \Sigma_h)$. En effet, dans cette somme, seul le p^{ieme} terme dépend de p .

Finalement :

$$\frac{\partial E_l}{\partial \alpha_p} = 2C \sum_{x \in \Omega} \mathcal{N}(x; R_{\theta_l} \mu_p + t_l, \sigma^2 I + \Sigma_h) (\mathcal{T}_{t_l}(\mathcal{R}_{\theta_l}(f))(x) - y_l(x)) \quad (6.10)$$

— Gradient de E_l par rapport aux centres μ_p :

Comme pour les coefficients, on a :

$$\nabla_{\mu_p} E_l = 2 \langle \nabla_{\mu_p} (\mathcal{T}_{t_l}(\mathcal{R}_{\theta_l}(f))), \mathcal{T}_{t_l}(\mathcal{R}_{\theta_l}(f)) - y_l \rangle \quad (6.11)$$

$$= 2 \sum_{x \in \Omega} \nabla_{\mu_p} (\mathcal{T}_{t_l}(\mathcal{R}_{\theta_l}(f)))(x) (\mathcal{T}_{t_l}(\mathcal{R}_{\theta_l}(f))(x) - y_l(x)) \quad (6.12)$$

$$= 2C \sum_{x \in \Omega} \alpha_p \nabla_{\mu_p} \mathcal{N}(x; R_{\theta_l} \mu_p + t_l, \sigma^2 I + \Sigma_h) (\mathcal{T}_{t_l}(\mathcal{R}_{\theta_l}(f))(x) - y_l(x)) \quad (6.13)$$

Le calcul du gradient $\nabla_{\mu_p} \mathcal{N}(x; R_{\theta_l} \mu_p + t_l, \sigma^2 I + \Sigma_h)$ se décompose en deux étapes. La première est le calcul du gradient d'une gaussienne par rapport à son centre. La seconde utilise la formule de dérivation des fonctions composées (règle de la chaîne) pour, de manière générique, déduire le gradient de la fonction $h : \mu \mapsto g(R\mu + t)$ connaissant le gradient de g . La démonstration est faite pour une fonction $g : \mathbb{R}^N \mapsto \mathbb{R}$ quelconque et une matrice $R \in \mathcal{M}_n(\mathbb{R})$ quelconque, dans notre cas on l'applique à $g : c \mapsto \mathcal{N}(x; c, \Sigma)$, $R = R_{\theta_l}$, $t = t_l$ en notant $\Sigma = \sigma^2 I + \Sigma_h$. La démonstration se fait en deux étapes :

Première étape :

$$g(c) = \mathcal{N}(x; c, \Sigma) = e^{-\frac{1}{2}(x-c)^t \Sigma^{-1} (x-c)} = e^{-\frac{1}{2} \phi(c)} \quad (6.14)$$

D'après [109] on a $\frac{\partial \phi}{\partial c} = -2\Sigma^{-1}(x - c)$ On obtient par la règle de la chaîne :

$$\frac{\partial g}{\partial c} = \Sigma^{-1}(x - c) \mathcal{N}(x; c, \Sigma) \quad (6.15)$$

Deuxième étape :

On note $c = R\mu + t$. c , μ et t sont des vecteurs à 3 dimensions que l'on note $c = (c_j)_{j \in \llbracket 1,3 \rrbracket}$, $\mu = (\mu_k)_{k \in \llbracket 1,3 \rrbracket}$, $t = (t_j)_{j \in \llbracket 1,3 \rrbracket}$. La relation matricielle $c = R\mu + t$ peut être écrite sous forme

explicite : $c_j = \sum_{k=1}^3 R_{jk}\mu_k + t_j$. On peut alors calculer la dérivée partielle de chaque c_j par rapport à chaque μ_k . Cela donne : $\frac{\partial c_j}{\partial \mu_k} = R_{jk}$
La règle de la chaîne donne alors :

$$\frac{\partial h}{\partial \mu_k} = \sum_{j=1}^3 \frac{\partial c_j}{\partial \mu_k} \frac{\partial g}{\partial c_j} = \sum_{j=1}^3 R_{jk} \frac{\partial g}{\partial c_j} = \sum_{j=1}^3 R_{kj}^t \frac{\partial g}{\partial c_j} \quad (6.16)$$

En écrivant cette expression sous forme matricielle, on obtient que le gradient de h par rapport à μ est le gradient de g par rapport à c multiplié par la transposée de la matrice R .

Formellement cela donne :

$$\nabla_{\mu} h = \nabla_{\mu} g(R\mu + t) = R^t \nabla_{\mu} g = R^t \Sigma^{-1} (x - R\mu - t) \mathcal{N}(x; R\mu + t, \Sigma) \quad (6.17)$$

Pour résumer, en remplaçant dans (6.9) $\Sigma = \sigma^2 I + \Sigma_h$, $\mu = \mu_p$, $R = R_{\theta_l}$ et $t = t_l$ on obtient

$$\nabla_{\mu_p} E_l = \alpha_p C R_{\theta_l}^t (\sigma^2 I + \Sigma_h)^{-1} \sum_{x \in \Omega} (x - R_{\theta_l} \mu_p - t_l) \mathcal{N}(x; R_{\theta_l} \mu_p + t_l, \sigma^2 I + \Sigma_h) (\mathcal{T}_{t_l}(\mathcal{R}_{\theta_l}(f))(x) - y_l(x)) \quad (6.18)$$

6.3 Distance entre mélanges de gaussiennes

Nous présentons ici le calcul de la distance entre deux mélanges de gaussiennes, qui est utilisé dans le cadre de la méthode ShiReGmm2 pour calculer la fonction de coût (voir section 3.5.1).

On cherche à évaluer la distance ℓ_2 entre les mélanges de gaussiennes $\tilde{y}_l$ et $h * \mathcal{T}_{t_l}(\mathcal{R}_{\theta_l}(f))$ en comparant leurs distributions.

Pour rappel on a :

$$h * \mathcal{T}_{t_l}(\mathcal{R}_{\theta_l}(f)) = \sum_{p=1}^{N_g} C \alpha_p \mathcal{N}(R_{\theta_l} \mu_p + t_l, \Delta), \quad (6.19)$$

et

$$\tilde{y}_l = \sum_{q=1}^{M_l} \beta_q^l \mathcal{N}(m_q^l, \Delta). \quad (6.20)$$

$$\begin{aligned}
\tilde{E}^l(\alpha, \mu) &= \|h * \mathcal{T}_{t_l}(\mathcal{R}_{\theta_l}(f)) - \tilde{y}_l\|^2 = \\
&\int_{x \in \Omega} \left(\sum_{p=1}^{N_g} C \alpha_p \mathcal{N}(x; R_{\theta_l} \mu_p + t_l, \Delta) - \sum_{q=1}^{M_l} \beta_q^l \mathcal{N}(x; m_q^l, \Delta) \right)^2 dx \\
&= \sum_{p=1}^{N_g} \sum_{p'=1}^{N_g} C^2 \alpha_p \alpha_{p'} \int_{x \in \Omega} \mathcal{N}(x; R_{\theta_l} \mu_p + t_l, \Delta) \mathcal{N}(x; R_{\theta_l} \mu_{p'} + t_l, \Delta) dx \\
&\quad + \sum_{q=1}^{M_l} \sum_{q'=1}^{M_l} \beta_q^l \beta_{q'}^l \int_{x \in \Omega} \mathcal{N}(x; m_q^l, \Delta) \mathcal{N}(x; m_{q'}^l, \Delta) dx \\
&\quad - 2 \sum_{p=1}^{N_g} \sum_{q=1}^{M_l} C \alpha_p \beta_q^l \int_{x \in \Omega} \mathcal{N}(x; m_q^l, \Delta) \mathcal{N}(x; R_{\theta_l} \mu_p + t_l, \Delta) dx
\end{aligned}$$

Or d'après [109],

$$\int_{x \in \Omega} \mathcal{N}(x, \mu, \Sigma) \mathcal{N}(x, \mu', \Sigma') dx = \mathcal{N}(\mu; \mu', \Sigma + \Sigma'). \quad (6.21)$$

On peut donc écrire :

$$\tilde{E}^l(\alpha, \mu) = E_1^l(\alpha, \mu) + E_2^l(\alpha, \mu) + E_3^l, \quad (6.22)$$

avec

$$E_1^l(\alpha, \mu) = C^2 \sum_{p=1}^{N_g} \sum_{p'=1}^{N_g} \alpha_p \alpha_{p'} \mathcal{N}(R_{\theta_l} \mu_{p'} + t_l; R_{\theta_l} \mu_p + t_l, 2\Delta), \quad (6.23)$$

$$E_2^l(\alpha, \mu) = -2C \sum_{p=1}^{N_g} \sum_{q=1}^{M_l} \alpha_p \beta_q^l \mathcal{N}(m_q^l; R_{\theta_l} \mu_p + t_l, 2\Delta), \quad (6.24)$$

$$E_3^l = \sum_{q=1}^{M_l} \sum_{q'=1}^{M_l} \beta_q^l \beta_{q'}^l \mathcal{N}(m_q^l; m_{q'}^l, 2\Delta). \quad (6.25)$$

On cherche maintenant à calculer les dérivées de E_1 et E_2 par rapport à chacun des coefficients et centres α_p et μ_p . Le calcul des dérivés par rapport aux coefficients s'effectue aisément. Que ce soit pour E_1 ou E_2 , seul un terme de la première somme dépend de α_p , leurs dérivées ont donc une forme similaire en remplaçant les sommes doubles par des sommes simples. C'est-à-dire :

$$\frac{\partial E_2^l}{\partial \alpha_p} = -2C \sum_{q=1}^{M_l} \beta_q^l \mathcal{N}(m_q^l; R_{\theta_l} \mu_p + t_l, 2\Delta) \quad (6.26)$$

$$\frac{\partial E_1^l}{\partial \alpha_p} = C^2 \sum_{p'=1}^{N_g} \alpha_{p'} \mathcal{N}(R_{\theta_l} \mu_{p'} + t_l; R_{\theta_l} \mu_p + t_l, 2\Delta) \quad (6.27)$$

Le calcul des gradients par rapport aux centres peut-être effectué en réutilisant le résultat (6.18) de l'annexe précédente. On remarque en effet que $R_{\theta_l}\mu_{p'} + t_l$ et m_q^l jouent le même rôle que x dans $C \sum_{p'=1}^N \alpha_{p'} \mathcal{N}(x; R_{\theta_l}\mu_{p'} + t_l, \sigma^2 I + \Sigma_h)$. On a donc :

$$\nabla_{\mu_p} E_1^l = \alpha_p C^2 R_{\theta_l}^t (2\Delta)^{-1} R_{\theta_l} \sum_{p'=1}^{N_g} \alpha_{p'} (\mu_{p'} - \mu_p) \mathcal{N}(R_{\theta_l}\mu_{p'} + t_l; R_{\theta_l}\mu_p + t_l, 2\Delta) \quad (6.28)$$

$$\nabla_{\mu_p} E_2^l = -2\alpha_p C R_{\theta_l}^t (2\Delta)^{-1} \sum_{q=1}^{M_l} \beta_q^l (m_q^l - R_{\theta_l}\mu_p - t_l) \mathcal{N}(m_q^l; R_{\theta_l}\mu_p + t_l, 2\Delta) \quad (6.29)$$

6.4 Projet annexe sur le pore nucléaire

Au-delà de mes travaux de thèse, j'ai eu l'occasion de contribuer à un article en collaboration avec l'équipe du Centriole Lab de l'Université de Genève [110]. L'article montre qu'il est possible, avec une technique de microscopie par expansion nommée iU-ExM (pour iterative ultra-structure expansion microscopy), d'obtenir des résolutions similaire à la SMLM (Single Molecule localization Microscopy) (technique de microscopie évoquée en section 1.3). Cette étude est basée sur l'acquisition d'images de pores nucléaires (NPC pour "nuclear pore complex"), assemblages macromoléculaires admettant une symétrie C8 et impliqués dans les échanges entre le noyau et le cytoplasme. Ma contribution a consisté à développer une méthode automatique permettant d'évaluer la capacité de iU-ExM à révéler la symétrie C8 de la structure. L'algorithme détecte le nombre de coins labellisés dans chaque image de pore nucléaire et est adapté de celui proposé par [111] développé uniquement pour les données SMLM.

Pour plus de détails, nous renvoyons le lecteur vers l'article mentionné [110] (en particulier la section "Automatic detection of nuclear pores corners").

Bibliographie

- [1] D. Ribatti, “An historical note on the cell theory,” *Experimental cell research*, vol. 364, no. 1, pp. 1–4, 2018.
- [2] A. Wollman, R. Nudd, E. Hedlund, and M. Leake, “From animaculum to single molecules : 300 years of the light microscope. open biol. 2015 ; 5 (4) : 150019,”
- [3] S. M. Opal, “A brief history of microbiology and immunology,” *Vaccines : A biography*, pp. 31–56, 2010.
- [4] P. Mazzarello, “A unifying concept : the history of cell theory,” *Nature cell biology*, vol. 1, no. 1, pp. E13–E15, 1999.
- [5] E. Morin, “Intoduction à la pensée complexe,” 1990.
- [6] X. Benjin and L. Ling, “Developments, applications, and prospects of cryo-electron microscopy,” *Protein Science*, vol. 29, no. 4, pp. 872–882, 2020.
- [7] Y. Shi, “A glimpse of structural biology through x-ray crystallography,” *Cell*, vol. 159, no. 5, pp. 995–1014, 2014.
- [8] S. Curry, “Structural biology : a century-long journey into an unseen world,” *Interdisciplinary Science Reviews*, vol. 40, no. 3, pp. 308–328, 2015.
- [9] S. W. Cowan-Jacob, G. Fendrich, A. Floersheimer, P. Furet, J. Liebetanz, G. Rummel, P. Rheinberger, M. Centeleghe, D. Fabbro, and P. W. Manley, “Structural biology contributions to the discovery of drugs to treat chronic myelogenous leukaemia,” *Acta Crystallographica Section D : Biological Crystallography*, vol. 63, no. 1, pp. 80–93, 2007.
- [10] T. Hu, E. R. Sprague, M. Fodor, T. Stams, K. L. Clark, and S. W. Cowan-Jacob, “The impact of structural biology in medicine illustrated with four case studies,” *Journal of Molecular Medicine*, vol. 96, pp. 9–19, 2018.
- [11] Y. Cheng, “Single-particle cryo-EM—how did it get here and where will it go,” *Science*, vol. 361, no. 6405, pp. 876–880, 2018.
- [12] Y. Cheng, “Single-particle cryo-EM at crystallographic resolution,” *Cell*, vol. 161, no. 3, pp. 450–457, 2015.
- [13] H.-W. Wang and J.-W. Wang, “How cryo-electron microscopy and x-ray crystallography complement each other,” *Protein Science*, vol. 26, no. 1, pp. 32–39, 2017.
- [14] J. W. Lichtman and J.-A. Conchello, “Fluorescence microscopy,” *Nature methods*, vol. 2, no. 12, pp. 910–919, 2005.
- [15] X. Wang and Y. Lai, “Three basic types of fluorescence microscopy and recent improvement,” vol. 290, p. 01031, 2021.
- [16] D. Semwogerere and E. R. Weeks, “Confocal microscopy,” *Encyclopedia of biomaterials and biomedical engineering*, vol. 23, pp. 1–10, 2005.

- [17] L. Schermelleh, R. Heintzmann, and H. Leonhardt, “A guide to super-resolution fluorescence microscopy,” *Journal of Cell Biology*, vol. 190, no. 2, pp. 165–175, 2010.
- [18] K. Prakash, B. Diederich, R. Heintzmann, and L. Schermelleh, “Super-resolution microscopy : a brief history and new avenues,” *Philosophical Transactions of the Royal Society A*, vol. 380, no. 2220, p. 20210110, 2022.
- [19] F. Chen, P. W. Tillberg, and E. S. Boyden, “Expansion microscopy,” *Science*, vol. 347, no. 6221, pp. 543–548, 2015.
- [20] D. Gambarotto, F. U. Zwettler, M. Le Guennec, M. Schmidt-Cernohorska, D. Fortun, S. Borgers, J. Heine, J.-G. Schloetel, M. Reuss, M. Unser, *et al.*, “Imaging cellular ultrastructures using expansion microscopy (U-ExM),” *Nature methods*, vol. 16, no. 1, pp. 71–74, 2019.
- [21] S. Haider, A. Cameron, P. Siva, D. Lui, M. Shafiee, A. Boroomand, N. Haider, and A. Wong, “Fluorescence microscopy image noise reduction using a stochastically-connected random field model,” *Scientific reports*, vol. 6, no. 1, p. 20640, 2016.
- [22] M. Lelek, M. T. Gyparaki, G. Beliu, F. Schueder, J. Griffié, S. Manley, R. Jungmann, M. Sauer, M. Lakadamyali, and C. Zimmer, “Single-molecule localization microscopy,” *Nature Reviews Methods Primers*, vol. 1, no. 1, p. 39, 2021.
- [23] M. Saxena, G. Eluru, and S. S. Gorthi, “Structured illumination microscopy,” *Advances in Optics and Photonics*, vol. 7, no. 2, pp. 241–275, 2015.
- [24] S. Pujals, N. Feiner-Gracia, P. Delcanale, I. Voets, and L. Albertazzi, “Super-resolution microscopy as a powerful tool to study complex synthetic materials,” *Nature Reviews Chemistry*, vol. 3, no. 2, pp. 68–84, 2019.
- [25] S. Pujals and L. Albertazzi, “Super-resolution microscopy for nanomedicine research,” *ACS nano*, vol. 13, no. 9, pp. 9707–9712, 2019.
- [26] L. Vedrenne, E. Baudrier, and D. Fortun, “Particle detection based on few shot learning in 3d fluorescence microscopy,” *IEEE International Symposium on Biomedical Imaging*, 2024.
- [27] D. Fortun, P. Guichard, V. Hamel, C. O. S. Sorzano, N. Banterle, P. Gönczy, and M. Unser, “Reconstruction from multiple particles for 3D isotropic resolution in fluorescence microscopy,” *IEEE transactions on medical imaging*, vol. 37, no. 5, pp. 1235–1246, 2018.
- [28] M. Le Guennec, N. Klena, G. Aeschlimann, V. Hamel, and P. Guichard, “Overview of the centriole architecture,” *Current Opinion in Structural Biology*, vol. 66, pp. 58–65, 2021.
- [29] M. H. Laporte, D. Gambarotto, É. Bertiaux, L. Bournonville, V. Louvel, J. M. Nunes, S. Borgers, V. Hamel, and P. Guichard, “Time-series reconstruction of the molecular architecture of human centriole assembly,” *Cell*, vol. 187, no. 9, pp. 2158–2174, 2024.
- [30] M. Temerinac-Ott, O. Ronneberger, P. Ochs, W. Driever, T. Brox, and H. Burkhardt, “Multi-view deblurring for 3-D images from light-sheet-based fluorescence microscopy,” *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 1863–1873, 2011.
- [31] S. Preibisch, F. Amat, E. Stamatakis, M. Sarov, R. H. Singer, E. Myers, and P. Tomancak, “Efficient bayesian-based multiview deconvolution,” *Nature methods*, vol. 11, no. 6, pp. 645–648, 2014.
- [32] J. Boulanger, C. Gueudry, D. Münch, B. Cinquin, P. Paul-Gilloteaux, S. Bardin, C. Guérin, F. Senger, L. Blanchoin, and J. Salamero, “Fast high-resolution 3D total internal reflection fluorescence microscopy by incidence angle scanning and azimuthal averaging,” *Proceedings of the National Academy of Sciences*, vol. 111, no. 48, pp. 17164–17169, 2014.

- [33] B. R. Masters, “History of the electron microscope in cell biology,” *eLS*, 2009.
- [34] D. Salas, A. Le Gall, J.-B. Fiche, A. Valeri, Y. Ke, P. Bron, G. Bellot, and M. Nollmann, “Angular reconstitution-based 3D reconstructions of nanomolecular structures from super-resolution light-microscopy images,” *Proceedings of the National Academy of Sciences*, vol. 114, no. 35, pp. 9273–9278, 2017.
- [35] C. Sieben, N. Banterle, K. M. Douglass, P. Gönczy, and S. Manley, “Multicolor single-particle reconstruction of protein complexes,” *Nature methods*, vol. 15, no. 10, pp. 777–780, 2018.
- [36] A. Löschberger, S. van de Linde, M.-C. Dabauvalle, B. Rieger, M. Heilemann, G. Krohne, and M. Sauer, “Super-resolution imaging visualizes the eightfold symmetry of gp210 proteins around the nuclear pore complex and resolves the central channel with nanometer resolution,” *Journal of cell science*, vol. 125, no. 3, pp. 570–575, 2012.
- [37] A. Szymborska, A. De Marco, N. Daigle, V. C. Cordes, J. A. Briggs, and J. Ellenberg, “Nuclear pore scaffold structure analyzed by super-resolution microscopy and particle averaging,” *Science*, vol. 341, no. 6146, pp. 655–658, 2013.
- [38] B. Blundell, C. Sieben, S. Manley, E. Rosten, Q. Ch’ng, and S. Cox, “3D structure from 2D microscopy images using deep learning,” *Frontiers in Bioinformatics*, p. 62, 2021.
- [39] V. Mennella, B. Keszthelyi, K. McDonald, B. Chhun, F. Kan, G. C. Rogers, B. Huang, and D. Agard, “Subdiffraction-resolution fluorescence microscopy reveals a domain of the centrosome critical for pericentriolar material organization,” *Nature cell biology*, vol. 14, no. 11, pp. 1159–1168, 2012.
- [40] J. Broeken, H. Johnson, D. S. Lidke, S. Liu, R. P. Nieuwenhuizen, S. Stallinga, K. A. Lidke, and B. Rieger, “Resolution improvement by 3D particle averaging in localization microscopy,” *Methods and applications in fluorescence*, vol. 3, no. 1, p. 014003, 2015.
- [41] X. Shi, G. Garcia III, Y. Wang, J. F. Reiter, and B. Huang, “Deformed alignment of super-resolution images for semi-flexible structures,” *PloS one*, vol. 14, no. 3, p. e0212735, 2019.
- [42] H. Heydarian, M. Joosten, A. Przybylski, F. Schueder, R. Jungmann, B. v. Werkhoven, J. Keller-Findeisen, J. Ries, S. Stallinga, M. Bates, *et al.*, “3D particle averaging and detection of macromolecular symmetry in localization microscopy,” *Nature communications*, vol. 12, no. 1, pp. 1–9, 2021.
- [43] P. A. Penczek, J. Zhu, and J. Frank, “A common-lines based method for determining orientations for more than 3 particle projections simultaneously,” *Ultramicroscopy*, vol. 63, no. 3-4, pp. 205–218, 1996.
- [44] A. Singer, R. R. Coifman, F. J. Sigworth, D. W. Chester, and Y. Shkolnisky, “Detecting consistent common lines in cryo-EM by voting,” *Journal of structural biology*, vol. 169, no. 3, pp. 312–322, 2010.
- [45] D. Lyumkis, S. Vinterbo, C. S. Potter, and B. Carragher, “Optimod—an automated approach for constructing and optimizing initial models for single-particle electron microscopy,” *Journal of structural biology*, vol. 184, no. 3, pp. 417–426, 2013.
- [46] I. Greenberg and Y. Shkolnisky, “Common lines modeling for reference free ab-initio reconstruction in cryo-EM,” *Journal of structural biology*, vol. 200, no. 2, pp. 106–117, 2017.
- [47] S. H. Scheres, “RELION : implementation of a Bayesian approach to cryo-EM structure determination,” *Journal of structural biology*, vol. 180, no. 3, pp. 519–530, 2012.

- [48] D. Elmlund and H. Elmlund, “Cryogenic electron microscopy and single-particle analysis,” *Annual Review of Biochemistry*, vol. 84, pp. 499–517, 2015.
- [49] T. Ogura and C. Sato, “A fully automatic 3D reconstruction method using simulated annealing enables accurate posterioric angular assignment of protein projections,” *Journal of structural biology*, vol. 156, no. 3, pp. 371–386, 2006.
- [50] X. Yan, K. A. Dryden, J. Tang, and T. S. Baker, “Ab initio random model method facilitates 3d reconstruction of icosahedral particles,” *Journal of Structural Biology*, vol. 157, no. 1, pp. 211–225, 2007. Software tools for macromolecular microscopy.
- [51] E. Sanz-García, A. B. Stewart, and D. M. Belnap, “The random-model method enables ab initio 3d reconstruction of asymmetric particles and determination of particle symmetry,” *Journal of structural biology*, vol. 171, no. 2, pp. 216–222, 2010.
- [52] P. A. Penczek and F. J. Asturias, “Ab initio cryo-EM structure determination as a validation problem,” pp. 2090–2094, 2014.
- [53] J. Vargas, A.-L. Alvarez-Cabrera, R. Marabini, J. M. Carazo, and C. O. S. Sorzano, “Efficient initial volume determination from electron microscopy images of single particles,” *Bioinformatics*, vol. 30, no. 20, pp. 2891–2898, 2014.
- [54] G. Harauz and M. van Heel, “Exact filters for general geometry three dimensional reconstruction,” *Optik*, vol. 73, no. 4, pp. 146–156, 1986.
- [55] S. H. Scheres, R. Núñez-Ramírez, Y. Gómez-Llorente, C. San Martín, P. P. Eggermont, and J. M. Carazo, “Modeling experimental image formation for likelihood-based classification of electron microscopy data,” *Structure*, vol. 15, no. 10, pp. 1167–1177, 2007.
- [56] S. H. Scheres, “A Bayesian view on cryo-EM structure determination,” *Journal of molecular biology*, vol. 415, no. 2, pp. 406–418, 2012.
- [57] A. Punjani, M. A. Brubaker, and D. J. Fleet, “Building proteins in a day : Efficient 3D molecular structure estimation with electron cryomicroscopy,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 4, pp. 706–718, 2016.
- [58] C. Sorzano, J. Vargas, J. de la Rosa-Trevín, J. Otón, A. Alvarez-Cabrera, V. Abrishami, E. Sesmero, R. Marabini, and J. Carazo, “A statistical approach to the initial volume problem in single particle analysis by electron microscopy,” *Journal of structural biology*, vol. 189, no. 3, pp. 213–219, 2015.
- [59] H. Elmlund, D. Elmlund, and S. Bengio, “PRIME : probabilistic initial 3D model generation for single-particle cryo-electron microscopy,” *Structure*, vol. 21, no. 8, pp. 1299–1306, 2013.
- [60] N. Jaitly, M. A. Brubaker, J. L. Rubinstein, and R. H. Lilien, “A Bayesian method for 3D macromolecular structure inference using class average images from single particle electron microscopy,” *Bioinformatics*, vol. 26, no. 19, pp. 2406–2415, 2010.
- [61] A. Punjani, M. A. Brubaker, and D. J. Fleet, “Supplementary Information for cryoSPARC : Algorithms for rapid unsupervised cryo-EM structure determination,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 4, pp. 706–718, 2016.
- [62] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” *Advances in neural information processing systems*, vol. 25, 2012.
- [63] Y. S. Nashed, F. Poitevin, H. Gupta, G. Woollard, M. Kagan, C. H. Yoon, and D. Ratner, “Cryoposenet : end-to-end simultaneous learning of single-particle orientation and 3d map reconstruction from cryo-electron microscopy data,” pp. 4066–4076, 2021.

- [64] H. Gupta, M. T. McCann, L. Donati, and M. Unser, “Cryogan : A new reconstruction paradigm for single-particle cryo-EM via deep adversarial learning,” *IEEE Transactions on Computational Imaging*, vol. 7, pp. 759–774, 2021.
- [65] E. D. Zhong, T. Bepler, B. Berger, and J. H. Davis, “Cryodrgn : reconstruction of heterogeneous cryo-EM structures using neural networks,” *Nature methods*, vol. 18, no. 2, pp. 176–185, 2021.
- [66] A. Levy, F. Poitevin, J. Martel, Y. Nashed, A. Peck, N. Miolane, D. Ratner, M. Dunne, and G. Wetzstein, “Cryoai : Amortized inference of poses for ab initio reconstruction of 3D molecular volumes from real cryo-EM images,” pp. 540–557, 2022.
- [67] V. Sitzmann, J. Martel, A. Bergman, D. Lindell, and G. Wetzstein, “Implicit neural representations with periodic activation functions,” *Advances in neural information processing systems*, vol. 33, pp. 7462–7473, 2020.
- [68] N. Elad, D. K. Clare, H. R. Saibil, and E. V. Orlova, “Detection and separation of heterogeneity in molecular complexes by statistical analysis of their two-dimensional projections,” *Journal of structural biology*, vol. 162, no. 1, pp. 108–120, 2008.
- [69] S. H. Scheres, H. Gao, M. Valle, G. T. Herman, P. P. Eggermont, J. Frank, and J.-M. Carazo, “Disentangling conformational states of macromolecules in 3D-EM through likelihood optimization,” *Nature methods*, vol. 4, no. 1, pp. 27–29, 2007.
- [70] T. Nakane, D. Kimanius, E. Lindahl, and S. H. Scheres, “Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION,” *elife*, vol. 7, p. e36861, 2018.
- [71] X.-c. Bai, E. Rajendra, G. Yang, Y. Shi, and S. H. Scheres, “Sampling the conformational space of the catalytic subunit of human γ -secretase,” *elife*, vol. 4, p. e11182, 2015.
- [72] S. L. Ilca, A. Kotecha, X. Sun, M. M. Poranen, D. I. Stuart, and J. T. Huiskonen, “Localized reconstruction of subunits from electron cryomicroscopy images of macromolecular complexes,” *Nature communications*, vol. 6, no. 1, p. 8843, 2015.
- [73] H. D. Tagare, A. Kucukelbir, F. J. Sigworth, H. Wang, and M. Rao, “Directly reconstructing principal components of heterogeneous particles from cryo-EM images,” *Journal of structural biology*, vol. 191, no. 2, pp. 245–262, 2015.
- [74] A. Punjani and D. J. Fleet, “3D variability analysis : Resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM,” *Journal of structural biology*, vol. 213, no. 2, p. 107702, 2021.
- [75] A. Levy, G. Wetzstein, J. N. Martel, F. Poitevin, and E. Zhong, “Amortized inference for heterogeneous reconstruction in cryo-EM,” *Advances in Neural Information Processing Systems*, vol. 35, pp. 13038–13049, 2022.
- [76] W. Chen, L. Yao, Z. Xia, and Y. Wang, “Ace-hetem for ab initio heterogenous cryo-EM 3d reconstruction,” *arXiv preprint arXiv :2308.04956*, 2023.
- [77] E. D. Zhong, A. Lerer, J. H. Davis, and B. Berger, “Exploring generative atomic models in cryo-EM reconstruction,” *arXiv preprint arXiv :2107.01331*, 2021.
- [78] D. Rosenbaum, M. Garnelo, M. Zielinski, C. Beattie, E. Clancy, A. Huber, P. Kohli, A. W. Senior, J. Jumper, C. Doersch, *et al.*, “Inferring a continuous distribution of atom coordinates from cryo-EM images using vaes,” *arXiv preprint arXiv :2106.14108*, 2021.

- [79] M. Chen and S. J. Ludtke, “Deep learning-based mixed-dimensional gaussian mixture model for characterizing variability in cryo-EM,” *Nature methods*, vol. 18, no. 8, pp. 930–936, 2021.
- [80] H. Gupta, T. H. Phan, J. Yoo, and M. Unser, “Multi-cryogan : Reconstruction of continuous conformations in cryo-EM using generative adversarial networks,” in *European Conference on Computer Vision*, pp. 429–444, Springer, 2020.
- [81] A. Punjani and D. J. Fleet, “3DFlex : determining structure and motion of flexible proteins from cryo-EM,” *Nature Methods*, pp. 1–11, 2023.
- [82] J. Schwab, D. Kimanius, A. Burt, T. Dendooven, and S. Scheres, “Dynamight : estimating molecular motions with improved reconstruction from cryo-EM images,” *bioRxiv*, pp. 2023–10, 2023.
- [83] S. Jonić and C. Ó. S. Sorzano, “Coarse-graining of volumes for modeling of structure and dynamics in electron microscopy : Algorithm to automatically control accuracy of approximation,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 1, pp. 161–173, 2015.
- [84] C. Donnat, A. Levy, F. Poitevin, E. D. Zhong, and N. Miolane, “Deep generative modeling for volume reconstruction in cryo-electron microscopy,” *Journal of Structural Biology*, p. 107920, 2022.
- [85] A. Ferrand, K. D. Schleicher, N. Ehrenfeuchter, W. Heusermann, and O. Biehlmaier, “Using the noise2noise workflow to measure signal-to-noise ratios of confocal microscopes,” *Scientific reports*, vol. 9, no. 1, p. 1165, 2019.
- [86] W. T. Baxter, R. A. Grassucci, H. Gao, and J. Frank, “Determination of signal-to-noise ratios and spectral SNRs in cryo-EM low-dose imaging of molecules,” *Journal of structural biology*, vol. 166, no. 2, pp. 126–132, 2009.
- [87] T. Eloy, É. Baudrier, M. Laporte, V. Hamel, P. Guichard, and D. Fortun, “Fast and robust single particle reconstruction in 3D fluorescence microscopy,” *IEEE Transactions on Computational Imaging*, 2023.
- [88] H. Kirshner, F. Aguet, D. Sage, and M. Unser, “3-D PSF fitting for fluorescence microscopy : implementation and localization application,” *Journal of microscopy*, vol. 249, no. 1, pp. 13–25, 2013.
- [89] R. Sato, M. Tanaka, and A. Takeda, “A gradient method for multilevel optimization,” *Advances in Neural Information Processing Systems*, vol. 34, 2021.
- [90] S. T. Tokdar and R. E. Kass, “Importance sampling : a review,” *Wiley Interdisciplinary Reviews : Computational Statistics*, vol. 2, no. 1, pp. 54–60, 2010.
- [91] X. Tong, Z. Ye, Y. Xu, S. Gao, H. Xie, Q. Du, S. Liu, X. Xu, S. Liu, K. Luan, *et al.*, “Image registration with fourier-based image correlation : A comprehensive review of developments and applications,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 10, pp. 4062–4081, 2019.
- [92] M. Guizar-Sicairos, S. T. Thurman, and J. R. Fienup, “Efficient subpixel image registration algorithms,” *Optics letters*, vol. 33, no. 2, pp. 156–158, 2008.
- [93] Y. Zhao, S. Chen, A. C. Swensen, W.-J. Qian, and E. Gouaux, “Architecture and subunit arrangement of native AMPA receptors elucidated by cryo-EM,” *Science*, vol. 364, no. 6438, pp. 355–362, 2019.

- [94] J. Wang, M. Hu, J. Wang, J. Qi, Z. Han, G. Wang, Y. Qi, H.-W. Wang, J.-M. Zhou, and J. Chai, “Reconstitution and structure of a plant NLR resistosome conferring immunity,” *Science*, vol. 364, no. 6435, p. eaav5870, 2019.
- [95] K. L. Morris, J. R. Jones, M. Halebian, S. Wu, M. Baker, J.-P. Armache, A. Avila Ibarra, R. B. Sessions, A. D. Cameron, Y. Cheng, *et al.*, “cryo-EM of multiple cage architectures reveals a universal mode of clathrin self-assembly,” *Nature structural & molecular biology*, vol. 26, no. 10, pp. 890–898, 2019.
- [96] K. M. Cirelli, D. G. Carnathan, B. Nogal, J. T. Martin, O. L. Rodriguez, A. A. Upadhyay, C. A. Enemuoh, E. H. Gebru, Y. Choe, F. Viviano, *et al.*, “Slow delivery immunization enhances HIV neutralizing antibody and germinal center responses via modulation of immunodominance,” *Cell*, vol. 177, no. 5, pp. 1153–1171, 2019.
- [97] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment : from error visibility to structural similarity,” *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.
- [98] J. Nilsson and T. Akenine-Möller, “Understanding SSIM,” *arXiv preprint arXiv :2006.13846*, 2020.
- [99] M. Van Heel and M. Schatz, “Fourier shell correlation threshold criteria,” *Journal of structural biology*, vol. 151, no. 3, pp. 250–262, 2005.
- [100] F. G. F. Christoph A. Diebolder, A. J. Koster, and R. I. Koning, “Conical fourier shell correlation applied to electron tomograms,” *Journal of Structural Biology*, vol. 190, no. 2, pp. 215–223, 2015.
- [101] C. R. Castro-Pareja and R. Shekhar, “Hardware acceleration of mutual information-based 3d image registration,” *Journal of Imaging Science and Technology*, vol. 49, no. 2, pp. 105–113, 2005.
- [102] J. Vargas, A.-L. Alvarez-Cabrera, R. Marabini, J. M. Carazo, and C. O. S. Sorzano, “Efficient initial volume determination from electron microscopy images of single particles,” *Bioinformatics*, vol. 30, no. 20, pp. 2891–2898, 2014.
- [103] J. M. D. la Rosa-Trevin *et al.*, “Scipion : A software framework toward integration, reproducibility and validation in 3D electron microscopy,” *Journal of structural biology*, vol. 195, no. 1, pp. 93–99, 2016.
- [104] V. Hamel and P. Guichard, “Improving the resolution of fluorescence nanoscopy using post-expansion labeling microscopy,” *Methods in Cell Biology*, vol. 161, pp. 297–315, 2021.
- [105] M. Le Guennec, N. Klena, D. Gambarotto, M. H. Laporte, A.-M. Tassin, H. Van den Hoek, P. S. Erdmann, M. Schaffer, L. Kovacic, S. Borgers, *et al.*, “A helical inner scaffold provides a structural basis for centriole cohesion,” *Science advances*, vol. 6, no. 7, p. eaaz4137, 2020.
- [106] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, *et al.*, “A density-based algorithm for discovering clusters in large spatial databases with noise,” vol. 96, no. 34, pp. 226–231, 1996.
- [107] W. S. Noble, “What is a support vector machine?,” *Nature biotechnology*, vol. 24, no. 12, pp. 1565–1567, 2006.
- [108] M. Theiss, J.-K. Hériché, C. Russell, D. Helekal, A. Soppitt, J. Ries, J. Ellenberg, A. Brazma, and V. Uhlmann, “Simulating structurally variable nuclear pore complexes for microscopy,” *Bioinformatics*, vol. 39, no. 10, p. btad587, 2023.
- [109] K. B. Petersen and M. S. Pedersen, “The matrix cook book,” vol. equation 85, 2012.

- [110] V. Louvel, R. Haase, O. Mercey, M. H. Laporte, T. Eloy, É. Baudrier, D. Fortun, D. Soldati-Favre, V. Hamel, and P. Guichard, “iU-ExM : nanoscopy of organelles and tissues with iterative ultrastructure expansion microscopy,” *Nature Communications*, vol. 14, no. 1, p. 7893, 2023.
- [111] J. V. Thevathasan, M. Kahnwald, K. Cieřliński, P. Hoess, S. K. Peneti, M. Reitberger, D. Heid, K. C. Kasuba, S. J. Hoerner, Y. Li, *et al.*, “Nuclear pores as versatile reference standards for quantitative superresolution microscopy,” *Nature methods*, vol. 16, no. 10, pp. 1045–1053, 2019.