



Apprendre de données positives et non étiquetées: application à la segmentation et la détection d'évènements calciques

Mémoire

Gabriel Leclerc

Maîtrise en génie électrique - avec mémoire
Maître ès sciences (M. Sc.)

Québec, Canada

**Apprendre de données positives et non étiquetées:
application à la segmentation et la détection
d'événements calciques**

Mémoire

Gabriel Leclerc

Sous la direction de:

Christian Gagné, directeur de recherche
Flavie Lavoie-Cardinal, codirectrice de recherche

Résumé

Deux types de neurotransmission se produisent dans les neurones du cerveau : la transmission évoquée et la transmission spontanée. Contrairement à la transmission évoquée, le rôle de la transmission spontanée sur la plasticité synaptique - un mécanisme utilisé pour doter le cerveau de capacités d'apprentissage et de mémorisation - reste incertain. Les neurotransmissions spontanées sont localisées et se produisent aléatoirement dans les synapses des neurones. Lorsqu'un tel événement spontané se produit, ce que l'on appelle un influx synaptique miniature d'ions calcium (*miniature Synaptic Ca²⁺ Transient*, mSCT), des ions calcium messagers secondaires pénètrent dans la synapse, activant les voies de signalisation en aval de la plasticité synaptique. L'utilisation de l'imagerie calcique du neurone *in vitro* permet la visualisation spatiotemporelle de l'entrée des ions calcium. Les vidéos calciques qui en résultent permettent une étude quantitative de l'impact du mSCT sur la plasticité synaptique. Cependant, la localisation des mSCTs dans l'imagerie du calcium est difficile en raison de leur petite taille, de leur faible intensité par rapport au bruit de l'imagerie et de leur caractère aléatoire inhérent. Dans ce mémoire, nous présentons une méthode d'analyse quantitative à grande échelle des vidéos d'imagerie calcique limitant la variabilité induite par les interventions humaines pour obtenir des données probantes, dans le but de caractériser l'impact des mSCTs sur la plasticité synaptique. En nous basant sur un outil semi-automatique de détection à seuil d'intensité (*Intensity Thresholded Detection*, ITD), nous sommes capables de générer des données pour entraîner un réseau pleinement convolutionnel (*Fully Convolutional Network*, FCN) afin de détecter rapidement et automatiquement les mSCTs à partir de vidéos calciques. En utilisant les segmentations bruitées de l'ITD comme données d'entraînement, combinées à un schéma d'entraînement positif (P) et non étiqueté (*Unlabeled*, U), les performances du FCN surpassent ITD. Le FCN détecte des mSCTs de faible intensité non détectés auparavant par ITD et offre une segmentation supérieure à ITD. Nous avons ensuite caractérisé l'impact des paramètres PU tels que le nombre de P et le ratio P:U. Le FCN entraîné est intégré dans une routine tout-en-un pour permettre une analyse à grande échelle des mSCTs. La routine offre la détection, la segmentation, la caractérisation et la visualisation des mSCTs ainsi qu'une solution logicielle pour gérer plusieurs vidéos avec différentes métadonnées.

Abstract

Two types of neurotransmission occur in brain's neurons: evoked transmission and spontaneous transmission. Unlike the former, the role of spontaneous transmission on synaptic plasticity – a mechanism used to endow the brain learning and memory abilities – remain unclear. Spontaneous neurotransmissions are localized and randomly happening in neuron's synapses. When such spontaneous events happen, so-called miniature synaptic Ca^{2+} transients (*mSCT*), second messenger calcium ions entered the spine, activating downstream signaling pathways of synaptic plasticity. Using calcium imaging of *in vitro* neuron enable spatiotemporal visualization of the entry of calcium ions. Resulting calcium videos enable quantitative study of *mSCT*'s impact on synaptic plasticity. However, *mSCT* localization in calcium imaging can be challenging due to their small size, their low intensity compared with the imaging noise and their inherent randomness. In this master's thesis, we present a method for quantitative high-throughput analysis of calcium imaging videos that limits the variability induced by human interventions to obtain evidence for characterizing the impact of *mSCT*s on synaptic plasticity. Based on a semi-automatic intensity thresholded detection (ITD) tool, we are able to generate data to train a fully convolutional neural network (FCN) to rapidly and automatically detect *mSCT* from calcium videos. Using ITD noisy segmentations as training data combine with a positive and unlabeled (PU) training schema, we leveraged FCN performances and could even detect previously undetected low intensity *mSCT*s missed by ITD. The FCN also provide better segmentation than ITD. We then characterized the impact of PU parameters such as the number of P and the ratio P:U. The trained FCN is bundled in a all-in-one pipeline to permit a high-throughput analysis of *mSCT*. The pipeline offers detection, segmentation, characterization and visualization of *mSCT*s as well as a software solution to manage multiple videos with different metadatas.

Table des matières

Résumé	ii
Abstract	iii
Table des matières	iv
Liste des tableaux	vi
Liste des figures	vii
Remerciements	xi
Introduction	1
1 Mise en contexte	6
1.1 Le neurone	7
1.2 La synapse chimique	8
1.3 Influx synaptique miniature d'ions calcium (mSCT)	9
1.4 L'imagerie calcique	10
1.5 Méthode d'analyse d'imagerie calcique classique	12
1.6 Apprentissage machine	13
1.7 Réduction de dimensionnalité	14
1.8 Apprentissage profond appliqué à la vision	15
1.9 L'entraînement des réseaux profonds	15
1.10 Le réseau de neurones à convolution	16
1.11 Le réseau pleinement convolutionnel	18
1.12 Architectures de type U-Net	19
1.13 Apprentissage profond appliqué à l'imagerie calcique	21
2 Méthodologie	23
2.1 Données d'imagerie calcique	23
2.2 Méthodes de référence	29
2.3 Apprentissage PU	32
2.4 Extraction des échantillons positifs (P)	35
2.5 Extraction des échantillons non étiquetée (U)	36
2.6 Algorithmes de segmentation	37
2.7 Prétraitement des vidéos	40
2.8 Détails de l'entraînement des algorithmes	42

2.9	Post-traitement des prédictions	44
2.10	Évaluation des performances	44
3	Analyses et comparaisons	49
3.1	Analyse U-Net 2D et U-Net 3D en segmentation	49
3.2	Analyse U-Net 2D et U-Net 3D en détection	55
3.3	Analyse de l'apprentissage PU	59
3.4	Discussion	62
4	Outils développés	64
4.1	Gestion des données	64
4.2	Visualisation des données	68
4.3	Routine automatisée d'analyse à grande échelle	70
4.4	Résultats d'analyses biologiques préliminaires	71
	Conclusion	76
A	Annexe	80
A.1	Types d'événements calciques pour la segmentation précise	80
A.2	Liste des métadonnées	80
A.3	Fichier de sauvegarde des résultats	81
A.4	Interface de visualisation	82
	Bibliography	86

Liste des tableaux

2.1	Composition du jeu de données d'imagerie calcique Ca^{2+} :	35
3.1	Performances globales des différentes approches proposées pour la détection et la segmentation des mSCTs :	63

Liste des figures

1.1	Premières images de neurones :	6
1.2	Le neurone et ses composantes :	7
1.3	La synapse et ses composantes :	9
1.4	Fluorescence émise lors de l'entrée d'ions calciums :	11
1.5	Exemple d'apprentissage non supervisé pour la réduction de dimensionnalité :	14
1.6	Schémas des réseaux encodeur-décodeur et U-Net :	20
2.1	Vidéo calcique d'un dendrite imagé par microscopie TIRF :	24
2.2	Transients Ca^{2+} synaptiques miniatures (mSCTs) dans un vidéo :	25
2.3	Zoom d'un mSCT sur une épine dendritique :	26
2.4	Normalisation d'une trace calcique de fluorescence :	27
2.5	Variation spatiale de fluorescence en fonction du temps :	28
2.6	Procédure de seuillage <i>MiniFinder</i> :	29
2.7	Exemples de distributions de données PU :	33
2.8	Extraction des échantillons positifs (P) :	36
2.9	Exemples des différentes méthodes de seuillage :	37
2.10	Extraction des étiquettes non étiquetées (U) :	38
2.11	Réseaux U-Net 2D et U-Net 3D :	40
2.12	Prétraitement des vidéos d'imagerie calcique :	41
2.13	Variation du coefficient de segmentation <i>Dice</i> pour de petites instances :	46
3.1	Comparaison en segmentation des approches U-Net 2D et U-Net 3D avec <i>MiniFinder</i> :	50
3.2	Analyse intra- et inter-expert :	52
3.3	Comparaison en segmentation des approches avec les annotations d'experts :	53
3.4	Exemples de segmentations comparées avec l'annotation précise d'un expert :	54
3.5	Comparaison en détection des approches U-Net 2D et U-Net 3D avec <i>MiniFinder</i> :	56
3.6	Réannotation des faux positifs (FP) du jeu de données de test :	57
3.7	Comparaison en segmentation des approches U-Net 2D, U-Net 3D, <i>MiniFinder</i> et <i>SICT</i> avec les experts :	58
3.8	Analyse de l'impact des régions <i>Positives</i> (P) et <i>Unlabeled</i> (U) :	60
4.1	Application développée sous Kivy pour la visualisation des mSCTs dans un vidéo :	69
4.2	Schéma de la routine automatisée d'analyse à grande échelle pour la segmentation, la détection, la caractérisation et la comparaison d'événements calciques :	70
4.3	Deux exemples de la comparaison entre vidéos :	73
4.4	Exemple d'analyse de l'aire des mSCTs :	74

4.5	Exemple d'application d'analyse non-supervisée pour la réduction de dimensionnalité avec l'algorithme <i>UMAP</i> :	75
A.1	Exemple de feuille Google pour sauvegarder les métadonnées :	81
A.2	<i>HDFView</i> : Logiciel d'inspection des fichiers HDF :	82
A.3	Téléchargement de la base de données et choix du vidéo :	83
A.4	Explorateur visuel des vidéos :	84
A.5	Panneaux pour contrôler le visuel et afficher de l'informations :	84
A.6	Visualisation des traces de fluorescence :	85

*À mes parents
Alain Leclerc et Chantale Savard*

A new philosophy, a new way
of life, is not given for nothing.
It has to be paid dearly for and
only acquired with much patience
and great effort.

- *Fyodor Dostoevsky*

Remerciements

Je remercie mes directeurs de recherche, Flavie Lavoie-Cardinal et Christian Gagné, qui m'ont guidé tout au long de ce parcours ardu, mais ô combien enrichissant ! Vos judicieux et perspicaces commentaires m'ont indéniablement permis de mener à terme ce mémoire. Grâce à votre confiance, j'ai pu découvrir deux domaines captivants : l'apprentissage machine et la neuroscience. Merci.

Je remercie mes parents, Chantale et Alain, de m'avoir inculqué la ténacité et le dévouement nécessaires à l'accomplissement de ce travail. L'importance de l'éducation à vos yeux est le phare qui guide mon cheminement académique. Vos mots, vos valeurs et vous-mêmes êtes et serez toujours un exemple. Merci.

Je remercie mes nombreuses soeurs, Marie-France, Olivia et Félicia. Votre écoute active et votre présence immuable constituent les fondations de mon épanouissement personnel. Merci.

Je remercie ma défunte grand-mère, grand-maman-Pierrette, qui a toujours été un exemple d'humilité. Elle était consciente du privilège de l'accès à l'éducation, ce qui a toujours été très inspirant. « Tu es très chanceux de pouvoir aller à l'école ! », me disait-elle souvent. Merci.

À tous ceux qui ont contribué directement ou indirectement à la réalisation de ce mémoire, je vous remercie.

Introduction

La neuroscience est le domaine qui s'intéresse à l'étude des mécanismes biologiques, physiques, chimiques et électriques oeuvrant dans le cerveau avec l'aspiration, entre autres, de comprendre son fonctionnement. Cette science évolue de pair avec les avancées dans le domaine de l'imagerie cérébrale qui nous permettent d'observer les composantes principales du cerveau. Ainsi, la microscopie nous révèle que le neurone constitue l'unité fondamentale permettant de faire circuler de l'information à même le cerveau, tout comme au reste de nos organes. Des structures neuronales complexes lui confèrent la possibilité de modifier la circulation de l'information dans le connectome. Les modifications ont lieu à l'échelle microscopique au sein du neurone et se répercutent jusqu'à l'échelle macroscopique, c'est-à-dire nos fonctions cognitives. À titre d'exemple, nos capacités d'apprentissage et de mémorisation illustrent la plasticité du cerveau (Kandel *et al.*, 2000).

Différentes hypothèses ont été émises pour tenter d'expliquer les changements qui ont lieu dans le cerveau. La fausse croyance populaire veut que l'apprentissage soit corrélé avec une augmentation du nombre de neurones. Par conséquent, une personne savante aurait nécessairement plus de neurones. Une hypothèse plus réaliste propose que l'apprentissage soit lié à l'augmentation d'interconnexions neuronales (Greenough *et al.*, 2002). À cette proposition s'ajoute la découverte des synapses, situées à la jonction entre deux neurones. Ces synapses régulent la facilité avec laquelle l'information circule entre les neurones. Ainsi, les évidences scientifiques sont maintenant orientées vers les changements biologiques opérant au sein des synapses (Humeau et Choquet, 2019). Une prémisse naïve serait de supposer que la taille de la synapse permet de réguler la transmission neuronale (p. ex. un influx nerveux circulerait plus difficilement via de petites synapses). Toutefois, la relation est plus complexe. Il semble y avoir un échafaudage éclectique de protéines à l'intérieur même de la synapse. De récentes avancées proposent que cet échafaudage permette un réarrangement de protéines d'intérêts, lequel causerait une modification de la capacité de transmission, c'est-à-dire la force synaptique (Abbott et Nelson, 2000; Martin *et al.*, 2000; Citri et Malenka, 2008).

La force synaptique décrit la propriété d'une synapse à transférer de l'information entre deux neurones. Une synapse comportant une force synaptique élevée signifie qu'un signal nerveux y circule aisément. Ainsi, en modifiant la force synaptique de leurs synapses, les neurones

peuvent faciliter ou entraver la propagation de leur influx nerveux. Actuellement, deux méthodes connues permettent au neurone de modifier leur force synaptique : la transmission évoquée – véhiculée par l’influx nerveux – et la transmission spontanée (TS). La transmission évoquée et ses effets sur la force synaptique ont été largement étudiés (Zucker et Regehr, 2002). Par exemple, dans une certaine mesure, l’augmentation du nombre d’influx nerveux augmente la force synaptique. Cela signifie que plus un chemin neuronal est utilisé, plus facilement l’influx nerveux y circule. D’un autre côté, la TS et ses répercussions sur la force synaptique sont encore peu étudiées. On sait néanmoins que la TS se produit indépendamment de la transmission évoquée et qu’elle est localisée sur une synapse (Sutton *et al.*, 2006; Kavalali, 2015).

Lors de la TS, une synapse transmet aléatoirement un signal au neurone voisin, générant des courants miniatures, de l’ordre du picoampère, sur la membrane neuronale. Simultanément, des ions calciques entrent dans la synapse. Les ions Ca^{2+} sont les messagers à l’origine des mécanismes de modulation de la force synaptique (Augustine *et al.*, 2003; Sutton et Schuman, 2009; Choi *et al.*, 2014; Reese et Kavalali, 2015). L’entrée d’ions calciques lors d’une TS porte le nom d’événement miniature calcique spontané (mSCT). L’imagerie calcique optique permet une détection temporelle de la synapse active par l’entremise de l’entrée d’ions Ca^{2+} lors d’un mSCT (Grienberger et Konnerth, 2012). Cette technique génère un vidéo d’un dendrite dans lequel les mSCTs sont visibles par l’augmentation du signal de fluorescence. Ces vidéos sont des données essentielles pour évaluer l’impact des mSCTs sur la plasticité synaptique.

Cependant, l’étude des vidéos générés par l’imagerie calcique optique nécessite la détection des mSCTs. Leur détection n’est pas triviale pour plusieurs raisons, dont : 1) les mSCTs ont lieu aléatoirement dans le temps et sur le dendrite ; 2) les mSCTs sont de petites tailles par rapport au champ de vue ; 3) la faible fluorescence rend certains mSCTs difficiles à visualiser par rapport au bruit de la caméra du microscope ; et 4) la variabilité inhérente des systèmes biologiques (Hoffman *et al.*, 2021) nécessite une analyse quantitative pour obtenir des données probantes. Un vidéo peut contenir jusqu’à plusieurs centaines de mSCTs et leur détection manuelle est longue, fastidieuse et sujette aux erreurs (Vohs *et al.*, 2014). Ainsi, des méthodes informatiques semi-automatisées par seuillage sont proposées pour accélérer les analyses (Rueckl *et al.*, 2017; Mancini *et al.*, 2018; Sun et Südhof, 2021). Celles-ci extraient de probables mSCTs, lesquels sont ensuite manuellement inspectés par un expert. L’efficacité des méthodes par seuillage dépend notamment de la sélection d’un seuil quantifiant la variation de fluorescence minimale pour détecter un mSCT. Un seuil élevé permet de capter moins de bruits, mais omet la détection des mSCTs ayant une intensité inférieure au seuil. Inversement, avec un seuil bas, la méthode capte plus de mSCTs et de bruits, ce qui augmente le temps d’inspection de l’expert. Le choix du seuil ajoute une variabilité supplémentaire dans l’analyse quantitative de l’effet des mSCTs sur la force synaptique. De plus, la tâche d’inspection par un expert est laborieuse et répétitive causant des erreurs d’annotation – un biais humain inévitable (Vohs

et al., 2014).

En fait, le biais humain lié à l’annotation peut être diminué en limitant les interventions humaines dans la boucle d’analyse. Pour ce faire, des méthodes d’apprentissage automatique peuvent effectuer une tâche donnée, sans intervention humaine. Les méthodes conventionnelles utilisent une série de caractéristiques soigneusement choisies pour représenter les données brutes – qu’on appelle des représentations. Avec ces représentations, aussi appelée des représentations, des algorithmes apprennent, à partir d’exemples, à effectuer une tâche, telle que de la classification, de la détection ou de la segmentation. Par contre, l’humain reste indirectement dans la boucle d’analyse lorsqu’il définit les caractéristiques pour représenter les données. Mais encore, les méthodes sont limitées par la qualité des caractéristiques sélectionnées. En effet, la réussite de la phase d’apprentissage de l’algorithme est liée à la pertinence des représentations pour effectuer la tâche.

Le prochain jalon logique pour l’apprentissage automatique est de permettre à la méthode d’apprendre automatiquement la représentation qui lui est utile pour la tâche. C’est ce que permettent les modèles du domaine de l’apprentissage profond (LeCun *et al.*, 2015). Par le biais d’une grande quantité de données de pair avec une augmentation de la puissance de calcul des ordinateurs, des modèles complexes peuvent apprendre automatiquement des représentations pertinentes de la donnée brute. L’architecture de ces modèles, basée sur des couches de filtres, permet une composition hiérarchique de représentations de bas niveau (p. ex. des lignes de différentes orientations) vers des représentations de haut niveau (p. ex. un oeil ou un visage). Avec des images, entre autres, cet apprentissage de représentation automatique dépasse la capacité de l’humain à créer des caractéristiques visuelles efficaces pour la classification (Krizhevsky *et al.*, 2012; He *et al.*, 2016).

Les modèles d’apprentissage profond ont rapidement été adaptés au domaine d’imagerie biomédicale (Long *et al.*, 2015; Ronneberger *et al.*, 2015; Milletari *et al.*, 2016). Dans ce domaine, la segmentation sémantique – où chaque pixel d’une image est catégorisé – est une tâche récurrente. En effet, les différentes techniques d’imagerie biomédicale sont développées pour visualiser des structures biologiques qui ne sont pas visibles à l’oeil nu, d’où l’intérêt pour des méthodes automatiques de segmentation de ces structures (Litjens *et al.*, 2017). Pour ce faire, les modèles apprennent simultanément deux tâches : 1) transformer la donnée brute vers une représentation abstraite, dans laquelle les structures biologiques sont facilement distinguables ; et 2) retourner de la représentation abstraite vers l’image, où chaque pixel est associé à une catégorie. Par exemple, des modèles ont été entraînés avec des données brutes d’imagerie calcique de culture de neurones pour y segmenter le soma des neurones actifs afin de caractériser la transmission évoquée (Apthorpe *et al.*, 2016; Klibisz *et al.*, 2017; Soltanian-Zadeh *et al.*, 2019).

Appliquer un modèle d’apprentissage profond supervisé pour extraire automatiquement les

mSCTs des vidéos nécessite l’annotation de vidéos au pixel près par un expert. Ce processus d’étiquetage standard comporte quatre problématiques :

1. *Étiquetage fastidieux* – L’annotation manuelle d’un vidéo au pixel près est extrêmement longue. Par exemple, imager des mSCTs pendant une minute génère environ 100 M de pixels (pour un vidéo de 512×512 pixels à 6 Hz).
2. *Bordures bruitées* – Contrairement aux images naturelles où le périmètre des instances à segmenter est bien défini (Cordts *et al.*, 2016), les mSCTs ont une bordure inexorablement floue. La microscopie optique de fluorescence est naturellement bruitée (Haider *et al.*, 2016) et les mSCTs sont observables lors d’une variation de fluorescence parfois du même ordre de grandeur que le bruit. La bordure spatiale des mSCTs est une transition vers le bruit ambiant, ce qui la rend floue et difficile à délimiter pour un expert humain.
3. *Difficilement discernables* – La faible intensité, selon le rapport signal sur bruit, de fluorescence et la petite taille des mSCTs dans le champ de vue les rendent difficiles à détecter à l’oeil nu. D’autant plus que leur détection nécessite l’inspection du champ de vue dans le temps (p. ex. les vidéos sont imagés à 6 Hz pendant plus d’une minute).
4. *Plusieurs étiquetages requis* – La variabilité inhérente des systèmes biologiques nécessite l’annotation de plusieurs neurones (vidéos) pour avoir une bonne représentation de la distribution des mSCTs pour entraîner des modèles profonds (Goodfellow *et al.*, 2016).

Par conséquent, une annotation manuelle invariable et impartiale exige que l’expert humain soit concentré et rigoureux pendant plusieurs heures. Il doit définir des frontières claires du mSCT, malgré le bruit, et détecter les petites intensités locales qui se détachent du bruit de fond. Ce n’est pas une tâche évidente pour un humain (Vohs *et al.*, 2014).

Problématique de recherche : *Ce mémoire vise à développer une méthode d’analyse quantitative à grande échelle des vidéos d’imagerie calcique limitant la variabilité induite par les interventions humaines pour obtenir des données probantes, dans le but, lors de futurs travaux, de caractériser l’impact des mSCTs sur la plasticité synaptique.*

Les analyses quantitatives sont très répandues dans les domaines possédant une importante variabilité (Gygi *et al.*, 1999; Karaman *et al.*, 2008; Zhao, 2013; Van Valen *et al.*, 2016; Chen *et al.*, 2019). L’analyse est basée sur des métriques numériques (p. ex. la taille, l’intensité et la fréquence des mSCTs), échantillonnées sur une grande population pour décrire, comprendre ou prédire des phénomènes (Somekh et Lewin, 2005). Des tests statistiques permettent de confirmer des hypothèses de corrélation à partir de données (p. ex. vérifier si la taille des mSCTs est corrélée avec leur intensité), à condition de réduire au maximum l’introduction de biais humain dans la prise des métriques numériques. Pour y parvenir, deux objectifs sont précisés :

- **1^{er} objectif :** *Adapter la procédure d’entraînement des modèles d’apprentissage profond aux contraintes spécifiques de l’imagerie calcique des mSCTs.*

— **2^{ème} objectif :** *Développer une routine d'analyse automatique tout-en-un pour de l'analyse à grande échelle d'imagerie calcique de mSCTs.*

La suite du mémoire est divisée en quatre chapitres. Le chapitre 1 met en contexte les concepts biologiques et informatiques dans le but de comprendre la nature des mSCTs et les outils d'analyse disponibles. Le chapitre 2 présente les particularités des vidéos d'imagerie calcique de mSCTs justifiant la méthodologie d'entraînement basée sur l'apprentissage à partir de données positives et non étiquetées. Le chapitre 3 analyse et compare la méthodologie proposée sous plusieurs angles afin d'exposer les performances et les limitations de celle-ci. Finalement, le chapitre 4 présente la routine automatique d'analyse développée, ses fonctionnalités et des résultats biologiques préliminaires.

Chapitre 1

Mise en contexte

La structure du neurone et sa physiologie sont d'abord présentées. Présent dans le neurone, la neurotransmission spontanée et la plasticité synaptique sont détaillées. S'ensuit une présentation des outils actuellement disponibles pour détecter et analyser la neurotransmission spontanée par le biais de la microscopie optique. Par la suite sont présentées des notions relatives aux récentes avancées dans le domaine de l'apprentissage machine – l'apprentissage profond. Notamment, les architectures de segmentation et leurs composantes spécifiques sont décrites. Finalement, des pistes récentes d'application de méthodes d'apprentissage profond sur des vidéos d'imagerie de microscopie à fluorescence sont présentées.

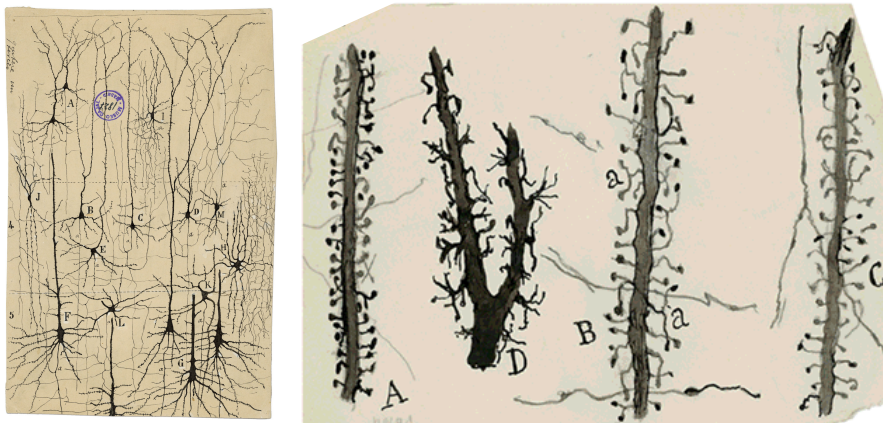


FIGURE 1.1 – Premières images de neurones. De la collection Garcia-Lopez *et al.* (2010), première représentation de neurones dessinée à la main par Santiago Ramon y Cajal (y Cajal, 1911). Sur la gauche, quelques neurones isolés, pour montrer que le système nerveux n'est pas un réseau continu tel qu'on le croyait à l'époque. À droite, on y voit des épines dendritiques de cellules pyramidales provenant du cortex frontal d'un humain.

1.1 Le neurone

Les premiers neurones imagés ont été tout d'abord dessinés à la main vers la fin du XIX^e siècle (figure 1.1) par le neuroscientifique espagnol Ramon y Cajal (y Cajal, 1911). Il fut le premier à proposer correctement, à partir de ses observations, que le cerveau soit composé de milliards de neurones individuels connectés entre eux (Abbott, 2008; Garcia-Lopez *et al.*, 2010). On sait maintenant que le neurone typique comporte cinq sections principales : le corps cellulaire (ou soma), les dendrites, l'axone, les épines et les boutons synaptiques (figure 1.2). Le soma comporte des prolongements dendritiques et un seul prolongement axonal. Les neurones sont des cellules spécifiquement conçues pour faire circuler de l'information dans le corps humain par le biais de l'influx nerveux, appelé également potentiel d'action. Les dendrites reçoivent les potentiels d'action, tandis que l'axone transmet ce potentiel. L'axone est de longueur variable et agit principalement comme médium pour transmettre l'influx nerveux entre deux neurones. La transmission de l'influx nerveux au sein du neurone a lieu via sa membrane cellulaire grâce à des échanges d'ions. Le neurone au repos possède une polarisation d'environ -70 mV aux bornes de sa membrane. Lorsque la polarisation membranaire atteint le seuil critique d'environ -55 mV , un potentiel d'action émerge du soma du neurone (Kandel *et al.*, 2000). Ce potentiel se propage dans l'axone jusqu'aux boutons synaptiques sous forme de courant bioélectrique. Par contre, ce courant électrique ne se propage pas d'un neurone à l'autre tel quel. Il est généralement propagé chimiquement par le biais de protubérances situées sur les dendrites (épines) et sur les axones (boutons). Ces zones d'échanges chimiques sont appelées synapses chimiques.

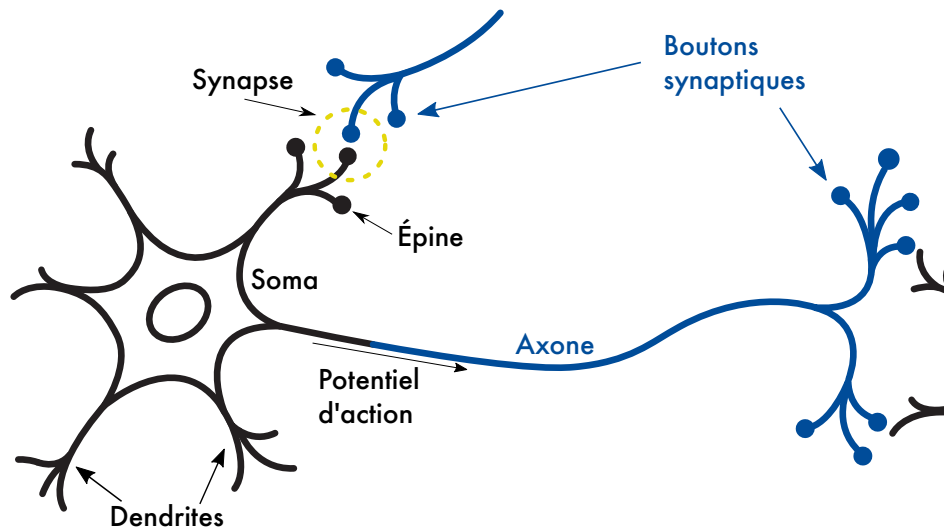


FIGURE 1.2 – Le neurone et ses composantes. Le soma du neurone comporte des prolongements dendritiques et un axone. Le potentiel d'action se propage du soma vers les boutons synaptiques par le biais de l'axone. La synapse chimique est la jonction entre une épine et un bouton synaptique. Elle connecte chimiquement deux neurones.

1.2 La synapse chimique

La synapse est la région qui connecte deux neurones via l'axone et le dendrite (figure 1.3). Une synapse est constituée de deux composantes : un bouton présynaptique et une épine postsynaptique, situés respectivement sur l'axone et le dendrite. La distance séparant ces composantes, l'espace intersynaptique, est d'environ 25 nm (O'Rourke *et al.*, 2012). L'espace intersynaptique prévient la circulation électrique de l'influx nerveux et agit comme médium permettant une circulation chimique du signal. En effet, lorsqu'un potentiel d'action atteint le bouton présynaptique, des neurotransmetteurs sont relâchés dans la zone intersynaptique par les vésicules synaptiques. Ils diffusent et s'attachent à des récepteurs situés sur l'épine postsynaptique et permettent l'entrée d'ions pour dépolariser la membrane et ainsi transférer l'influx entre deux neurones. Cette méthode de communication porte le nom de neurotransmission évoquée (Kandel *et al.*, 2000). Parallèlement, des neurotransmetteurs secondaires activent d'autres récepteurs postsynaptiques, les récepteurs NMDA, pour permettre l'entrée d'ions calciques. Les ions calcium agissent à titre de messagers secondaires pour activer les voies de signalisation modulant des fonctions de la synapse telle que la force synaptique (Tsien, 1983).

La force synaptique est l'efficacité de transmission de l'influx nerveux d'un neurone à l'autre. Cette force est dynamique et complexe (Hebb, 2005). Par exemple, une augmentation du nombre d'influx nerveux augmente la force synaptique (l'influx passe plus facilement), mais si la fréquence est élevée, la force synaptique va diminuer. En outre, des mécanismes peuvent également modifier à court ou à long terme la force synaptique (Citri et Malenka, 2008; Buonomano et Merzenich, 1998; Turrigiano *et al.*, 1998). La plasticité synaptique est le nom donné aux changements pré et postsynaptiques qui affectent la force synaptique (Collingridge *et al.*, 2004). La plasticité synaptique serait la source pour encoder, ou mémoriser, de la nouvelle information dans le cerveau (Martin *et al.*, 2000; Abbott et Nelson, 2000). Pour tenter de comprendre ses fonctions, des potentiels d'action sont induits dans le soma – ce qui simule une neurotransmission évoquée – permettant de générer des relâchements vésiculaires synchronisés dans les synapses du neurone. Ainsi, la plasticité synaptique est étudiée et caractérisée selon divers paradigmes d'expérimentations (Walter *et al.*, 2014). Toutefois, des relâchements aléatoires de neurotransmetteurs localisés sans lien apparent avec la neurotransmission évoquée sont observés. Historiquement, ces relâchements étaient considérés comme des résidus sans importance, mais de récentes études prouvent le contraire. Ils ont leurs importances dans les mécanismes d'apprentissage et de mémoire (Ramirez et Kavalali, 2011; Zucker, 2005; Choi *et al.*, 2014). Dans la littérature, ces relâchements sont groupés dans une deuxième catégorie de neurotransmission : la neurotransmission spontanée.

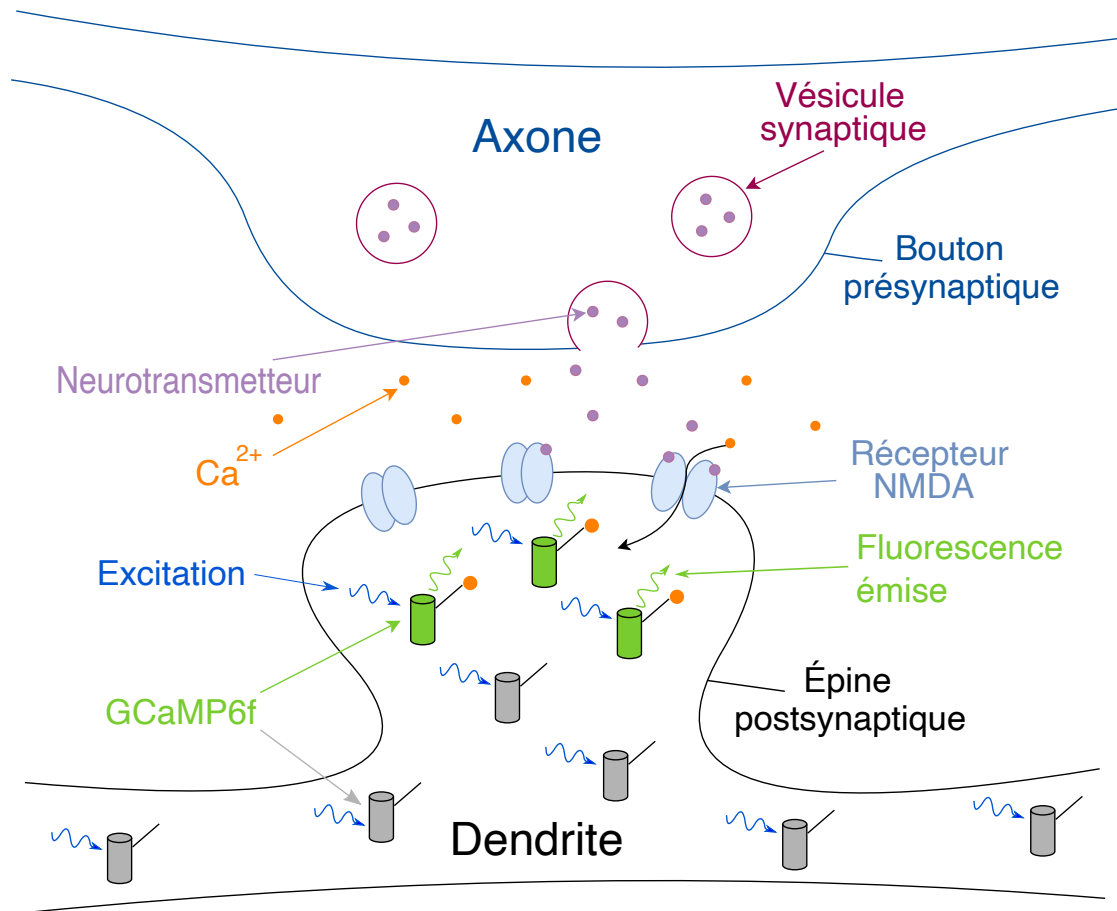


FIGURE 1.3 – La synapse et ses composantes. Schéma simplifié d'une synapse lors d'une transmission chimique de l'influx nerveux. Le bouton présynaptique (*bleu*), situé sur l'axone, libère des neurotransmetteurs via les vésicules synaptiques (*rouge*). Ces neurotransmetteurs migrent vers des récepteurs glutamatergiques, p. ex. récepteurs NMDA (*bleu pâle*) situés sur l'épine postsynaptique (*noir*). Activé par les neurotransmetteurs, le récepteur agit comme une pompe à ions et fait entrer des ions calciums (*orange*) dans l'épine dendritique (*noir*). Pour visualiser ce flux d'ions, des senseurs calciques GCaMP6f (*vert*) sont préalablement transfectés dans le neurone. Lorsqu'un atome de calcium s'attache au senseur, un changement conformationnel de la protéine induit l'émission de fluorescence de photons dans le vert (*flèches ondulées vertes*) lorsque le senseur est excité avec une lumière bleue (*flèche ondulée bleue*). Finalement, la fluorescence émise est captée par microscopie-vidéo.

1.3 Influx synaptique miniature d'ions calcium (mSCT)

La neurotransmission spontanée, découverte en 1952, est une propriété générale des synapses (Fatt et Katz, 1952; Choi *et al.*, 2014). Il s'agit d'un événement localisé, où, spontanément, une vésicule synaptique relâche des neurotransmetteurs et active un petit nombre de récepteurs NMDA sur l'épine postsynaptique (figure 1.3). Les récepteurs font entrer des ions calciques, lesquels sont à l'origine des mécanismes de modulation de la force synaptique (Augustine *et al.*, 2003). Cet événement est connu sous les noms : influx synaptique miniature de Ca^{2+} (*miniature*

Synaptic Ca²⁺ Transient, mSCT), événement calcique ou neurotransmission spontanée. Ces événements calciques ont lieu aléatoirement dans les synapses d’un neurone et jouent un rôle important dans la plasticité synaptique (Kavalali, 2015; Choi *et al.*, 2014). Il a été montré, entre autres, que les mSCTs régularisent la force synaptique (Sutton et Schuman, 2009), la maturation synaptique (Choi *et al.*, 2014) et la stabilité synaptique (Reese et Kavalali, 2015).

La détection des neurotransmissions spontanée est une tâche difficile. En effet, sans modifier le neurone, une détection spatiotemporelle est actuellement impossible. Par exemple, l’électrophysiologie consiste à insérer une électrode dans une cellule pour y mesurer les courants présents. Les courants mesurés dans le soma indiquent précisément le moment où des ions calciums entrent dans une des épines postsynaptiques (Murphy *et al.*, 1995). Cependant, cette méthode ne donne aucune information relative à la localisation des ions entrant dans le neurone, c.-à-d. la localisation de la synapse dans laquelle les ions sont entrés. La microscopie à fluorescence pallie ce manque d’information spatiale en imageant, dans un large champ de vue, les influx calciques à l’aide de senseurs fluorescents. Cette méthode porte le nom d’imagerie calcique.

1.4 L’imagerie calcique

En 1985, Grynkiewicz *et al.* (1985) propose une famille d’indicateur de Ca²⁺, une protéine qui émet une fluorescence maximale à la longueur d’onde du vert (avec une lumière d’excitation bleue) lorsqu’un ion de calcium se lie à son complexe. Ainsi, la présence d’ions Ca²⁺ peut être localisé en détectant la fluorescence émise. Il est alors possible de corrélérer l’entrée des ions calciums dans un neurone et un événement calcique. Or, plusieurs mSCTs peuvent survenir dans un petit laps de temps. Pour obtenir une bonne résolution temporelle, l’indicateur doit retourner rapidement dans un état sans fluorescence lorsque le complexe libère l’ion Ca²⁺. Dans le cas contraire, plusieurs événements calciques indépendants émettent la fluorescence équivalente à un seul événement calcique intense (p. ex. l’étoile à la figure 1.4). La classe d’indicateur de fluorescence *GCaMP6*, basé sur la famille d’indicateur de Ca²⁺, exhibe une bonne sensibilité et la détection des potentiels d’action avec une grande précision (Chen *et al.*, 2013). Parmi la classe *GCaMP6*, l’indicateur *GCaMP6f* (*f* signifie *fast*) comporte la meilleure précision et permet la résolution de mSCTs jusqu’à 50-75ms d’interval (Chen *et al.*, 2013). La fluorescence émise par les senseurs $F(t)$ peut être approximée par la convolution du signal d’entrée d’ions calciums $s(t)$ avec le noyau d’une exponentielle décroissante $K(t)$ (Pachitariu *et al.*, 2018) :

$$F(t) = s(t) * K(t) + n(t) \tag{1.1}$$

où $n(t)$ est un bruit gaussien indépendant. À la figure 1.4, dans le signal $s(t)$, chaque regroupement représente un mSCT. L’intensité du pic de fluorescence d’un mSCT dépend du nombre d’ions relâchés. Lorsque deux mSCTs ont lieu près temporellement, le signal de fluorescence permet difficilement de les dissocier (p. ex. l’étoile dans la figure 1.4).

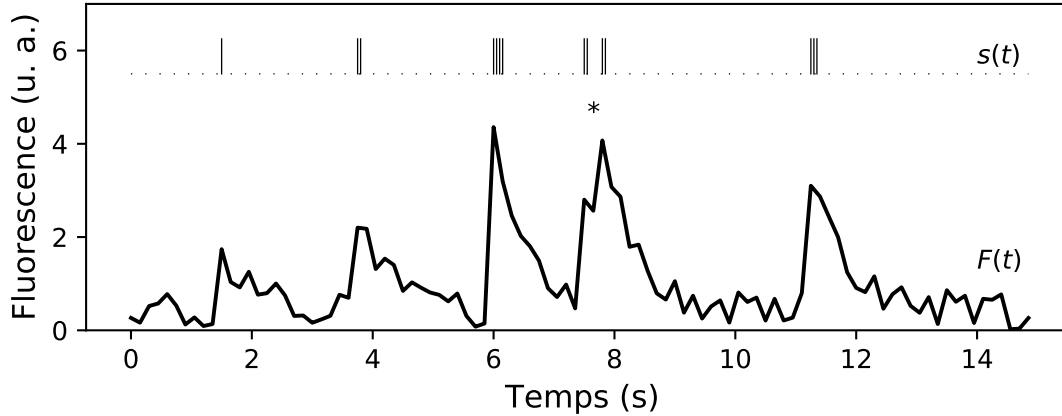


FIGURE 1.4 – Fluorescence émise lors de l’entrée d’ions calciums. Le signal binaire $s(t)$ représente le moment où des ions calciums entrent par les récepteurs NMDA et se lient à un senseur Ca^{2+} (un mSCT). La fluorescence émise par les senseurs est représentée par le signal $F(t)$. Lorsque des mSCTs surviennent dans un petit interval de temps, leurs distinctions sont difficiles uniquement à partir du signal de fluorescence (exemple *).

Le senseur calcique doit être dans le neurone au moment où un ion calcique y entre. Pour y arriver, le code génétique du senseur est transfecté dans le neurone pour qu’il synthétise lui-même l’indicateur calcique (Hanawa *et al.*, 2004). Après quelques jours de culture, les senseurs, synthétisés par ce neurone, baignent dans son corps cellulaire, ses dendrites et son axone. Ainsi, les Ca^{2+} qui entrent dans le neurone se lient aux senseurs *GCaMP6f*, lesquels émettent de la fluorescence verte après l’excitation (fig. 1.3). Il ne reste qu’à capter les photons émis avec un microscope adéquat.

Les sous-modules composant la synapse nécessitent une résolution nanoscopique pouvant être acquise par microscopie super-résolution à fluorescence (Hell et Wichmann, 1994; Huang *et al.*, 2009). Par contre, son application sur une longue période de temps avec une bonne résolution temporelle est limitée. En effet, dû au photoblanchiment (la perte des fonctions d’une protéine causée par la lumière), les méthodes d’acquisition rapide du champ de vue altèrent l’imagerie et limitent le temps d’acquisition (Hell et Wichmann, 1994; Schermelleh *et al.*, 2019). La microscopie optique conventionnelle est donc utilisée pour faire de l’imagerie calcique (Wilt *et al.*, 2009; Grienberger et Konnerth, 2012). La microscopie à fluorescence conventionnelle permet de capter l’émission de photons avec une résolution maximale dictée par la limite de diffraction, soit environ 250 nm dans le spectre visible (Abbe, 1873; Lichtman et Conchello, 2005). Cela permet de visualiser les composantes formant le neurone, dont les épines avec une taille de l’ordre du μm . La microscopie TIRF (*total internal reflection microscopy*) permet d’imager un plan focal complet avec une résolution suffisante pour discerner les synapses et une fréquence d’acquisition convenable pour détecter la dynamique de fluorescence des senseurs calciques (Axelrod, 2001). C’est le type de microscopie utilisé pour acquérir les données

présentées dans ce mémoire. L'imagerie calcique permet ainsi d'étudier la plasticité synaptique par le biais de vidéos du signal de fluorescence (Grienberger et Konnerth, 2012). L'investigation quantitative de la plasticité synaptique en présence de neurotransmission spontanée requiert l'analyse de quantités énormes de données où l'annotation manuelle par des experts est fastidieuse et peut être sujette à des biais (Rueckl *et al.*, 2017). Ainsi, des méthodes d'analyse peuvent être utilisées pour extraire l'information importante contenue dans ces vidéos.

1.5 Méthode d'analyse d'imagerie calcique classique

Les méthodes disponibles d'analyse d'imagerie calcique diffèrent selon l'échelle utilisée, définie par la taille du champ de vue : la micro-échelle, qui étudie les unités subcellulaires du neurone, comme les dendrites et les épines (Svoboda *et al.*, 1997) ; la méso-échelle, qui étudie des populations de neurones, sans résolution épineuse (Stosiek *et al.*, 2003; Ohki *et al.*, 2005; Pachitariu *et al.*, 2017) ; la macro-échelle, qui étudie les interactions entre les différentes régions du cerveau, sans résolution neuronale (Wang *et al.*, 2003). Des phénomènes différents sont étudiés à chaque échelle, nécessitant des analyses spécifiques. Par exemple, les analyses macro-échelle génèrent des quantités énormes de données dans lesquelles les régions du cerveau sont préalablement identifiées afin d'observer les interactions entre celles-ci (Fleishman *et al.*, 2019). À l'échelle cellulaire (méso-échelle), de grandes quantités de données sont également générées, dans lesquelles les neurones sont inventoriés pour caractériser les interactions entre eux (Giovannucci *et al.*, 2019). Du point de vue micro-échelle, la quantité de données générée est également imposante. Le défi à cette échelle consiste à identifier les régions d'intérêts à travers le bruit. En effet, dans un événement calcique où peu d'ions calciums entrent dans l'épine, la fluorescence émise est faible et se mélange au bruit ambiant (Mancini *et al.*, 2018; Tian *et al.*, 2019).

Peu importe l'échelle, des changements de fluorescence localisée sont observés dans un vidéo d'imagerie calcique. Outre l'annotation manuelle, différentes stratégies permettent de détecter ces changements de fluorescences : 1) l'application d'un seuillage (Rueckl *et al.*, 2017; Mancini *et al.*, 2018; Tian *et al.*, 2019) ; 2) la décomposition en ondelettes (*wavelet decomposition*) (Illaste *et al.*, 2019) ; 3) l'analyse en composantes indépendantes (ICA) (Mukamel *et al.*, 2009; Zoccoler et de Oliveira, 2020) ; 4) la factorisation par matrice non négatives (NMF) (Maruyama *et al.*, 2014; Pnevmatikakis *et al.*, 2016). Les méthodes ICA et NMF utilisent la réduction de dimensionnalité pour conserver celles expliquant les variations (p. ex. les augmentations de fluorescence) dans le vidéo. Ces méthodes fonctionnent bien pour la détection des transmissions évoquées (méso-échelle), mais ne sont pas adaptées pour la détection des transmissions spontanées en raison du faible rapport signal sur bruit. Illaste *et al.* (2019) utilise la décomposition en ondelettes pour détecter des régions d'intérêts qui sont filtrées automatiquement selon un seuil (Du *et al.*, 2006). La méthode est fiable, mais ne caractérise pas la forme de la propagation des ions calcium dans le neurone. Finalement, la majorité des

méthodes développées pour faire la détection et la segmentation des mSCTs sont basées sur le seuillage du signal de fluorescence. Les vidéos sont d’abord alignés pour corriger les mouvements possibles dans le champ de vue lors de l’imagerie (Pnevmatikakis et Giovannucci, 2017). Chaque vidéo est filtré avec un noyau gaussien et l’intensité est normalisée par rapport au signal basal (Macleod, 2012). Un seuillage est appliqué pour segmenter les mSCTs. Le nombre de détections et la précision de celles-ci sont fortement reliés au choix du seuil. Un seuil plus petit permet de repérer des mSCTs de faibles intensités en contrepartie d’une baisse de la précision due au bruit détecté. C’est pourquoi l’inspection manuelle de ce type d’approche est nécessaire pour conserver uniquement les mSCTs (Rueckl *et al.*, 2017; Mancini *et al.*, 2018). Avec un seuil adaptatif au niveau de bruit, Tian *et al.* (2019) propose une segmentation sans inspection visuelle. Par contre, leur approche est développée pour la détection d’événements calciques provenant des myocytes cardiaques.

Essentiellement, les outils présentés pour détecter les mSCTs comportent des paramètres qui doivent être ajustés selon le bruit de l’image, le champ de vue, la fréquence d’échantillonnage et la microscopie utilisée. Le temps d’ajustement de ces paramètres est considérable. La performance des outils est tributaire de l’expert qui effectue l’ajustement selon son expérience et sa fatigue (Vohs *et al.*, 2014). De plus, la détection des outils présentés est basée sur une caractéristique unique : la variation d’intensité induite par les mSCTs. Au lieu d’imposer la caractéristique nécessaire à la tâche et de devoir modifier des paramètres pour différentes conditions, des approches d’apprentissage machine peuvent s’adapter à plusieurs conditions et apprendre eux-mêmes les caractéristiques utiles à la tâche.

1.6 Apprentissage machine

L’apprentissage automatique, ou apprentissage machine (*Machine Learning*, ML), fait partie d’un domaine plus général appelé l’intelligence artificielle. Les données, les algorithmes et les processus d’apprentissage sont les constituants élémentaires du ML. Le but du ML consiste à résoudre une tâche spécifique en utilisant un algorithme, lequel s’adapte à partir d’exemples de la tâche au moyen d’un processus d’apprentissage. L’apprentissage automatique est possible en supposant l’existence d’un modèle expliquant, pas nécessairement parfaitement, les données observées ou la relation entre deux variables. La tâche d’un algorithme de ML consiste à approximer ce modèle. La possession de ce modèle permet de synthétiser un jeu de données ou encore de l’utiliser pour en faire des prédictions (Alpaydin, 2020). Par exemple, soit un modèle $G(x|\theta)$ qui approxime un jeu de données $x \sim \mathbf{X}$ et une nouvelle donnée \hat{x} , la vraisemblance que la donnée soit générée par ce modèle $\mathbb{E}(\hat{x}|G)$ permet de conclure sur son appartenance au jeu de données $\hat{x} \stackrel{?}{\sim} \mathbf{X}$. On peut ainsi prédire si \hat{x} possède les mêmes caractéristiques que les données issues de \mathbf{X} . Par contre, une difficulté avec cet exemple survient avec des données à haute dimensionnalité. Dans ce cas, le nombre d’exemples nécessaire pour faire l’approximation du modèle augmente exponentiellement avec l’augmentation de dimensions (Vapnik, 2013). Cet

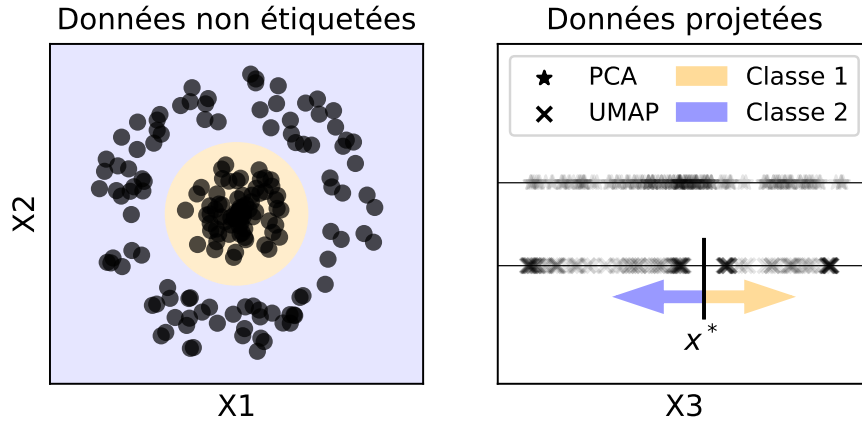


FIGURE 1.5 – Exemple d’apprentissage non supervisé pour la réduction de dimensionnalité. Les données non étiquetées (rond noir) possèdent deux caractéristiques : $X1$ et $X2$. Les zones bleu et orange montrent qu’il semble exister une distribution simple, mais non linéaire, qui explique ces données. Les données non étiquetées sont projetées dans un espace comportant une caractéristique, $X3$, avec deux méthodes non supervisées : PCA et UMAP. La composante $X3$ calculée avec la PCA ne permet pas de séparer les données. Avec l’algorithme *UMAP*, la nouvelle caractéristique permet une séparation des deux classes triviale (p. ex. Classe 1 si $X3 > x^*$)

obstacle porte le nom de malédiction de la dimension (Bellman, 1966).

1.7 Réduction de dimensionnalité

Pour pallier cette malédiction, des méthodes mathématiques ou statistiques permettent de réduire la dimensionnalité des données (Wold *et al.*, 1987; Maaten et Hinton, 2008; McInnes *et al.*, 2020). L’idée est de diminuer le nombre de caractéristiques d’une donnée avec une perte minimale d’information. L’analyse en composante principale est une approche populaire et consiste à déterminer des caractéristiques indépendantes expliquant au mieux la variance des données (Wold *et al.*, 1987). Cependant, elle est limitée aux données avec des distributions simples où une combinaison linéaire des caractéristiques est suffisante. Une autre approche possible consiste à approximer un *manifold* non linéaire comportant moins de dimensions tout en minimisant la perte d’informations : cette méthode se nomme le UMAP (McInnes *et al.*, 2020). Le *manifold* est complexe, mais la projection des données sur celui-ci permet de simplifier la représentation de la donnée. Par exemple, à la figure 1.5, des données non linéaires sont distribuées dans le plan cartésien selon leurs caractéristiques $X1$ et $X2$. Visuellement, deux classes différentes semblent présentes, les zones bleu et orange. Naïvement et sans *a priori*, la séparation dans l’espace initial nécessite plusieurs données pour apprendre une frontière circulaire entre les deux groupes. Les données à deux dimensions sont projetées dans un espace à une dimension $X3$ avec les méthodes PCA et UMAP (fig. 1.5). Dans le nouvel espace de

la PCA, la séparation n'est pas possible. Or, avec la réduction de dimensionnalité UMAP, la séparation entre les deux groupes est triviale et nécessite un seul paramètre, le seuil x^* . C'est un exemple simplifié, mais cette méthode peut être appliquée avec données complexes pour des analyses non supervisées (Wiesner *et al.*, 2020).

L'application d'approches classiques d'apprentissage automatique sont basées, entre autres, sur deux éléments : 1) le choix des caractéristiques et 2) le choix du modèle à optimiser. En effet, en fonction de la tâche et des données, le praticien doit fixer les caractéristiques nécessaires et choisir le modèle adéquat. Comme les caractéristiques à utiliser sont fortement corrélées avec la tâche à automatiser, il est possible d'augmenter la complexité du modèle afin de lui permettre d'apprendre automatiquement les caractéristiques nécessaires à son fonctionnement. Ce type d'apprentissage machine porte le nom d'apprentissage profond.

1.8 Apprentissage profond appliqué à la vision

L'apprentissage profond (*Deep Learning*, DL) est un sous-domaine de l'apprentissage machine où les modèles comporte beaucoup de paramètres et ont besoin, par conséquent, de nettement plus de données (LeCun *et al.*, 2015). Contrairement aux approches classiques de ML, les caractéristiques, c'est-à-dire la représentation des données, sont apprises par les modèles de DL. Dans le traitement d'image, cela implique qu'un *même* modèle peut être entraîné dans différents domaines, sans l'aide d'une connaissance spécialisée. C'est avantageux dans le domaine médical, car les méthodes d'imagerie sont très variées et chacune d'elle comporte des caractéristiques spécifiques qui peuvent être capturées dans l'apprentissage d'une représentation adaptée au problème.

1.9 L'entraînement des réseaux profonds

Comme les modèles de réseaux profonds supervisés comportent beaucoup de paramètres (plusieurs milliers à quelques millions), leur ajustement nécessite énormément de données annotées. La méthodologie reconnue pour faire l'ajustement de ces paramètres est basée sur des optimisations stochastiques selon l'impact de chaque paramètre sur la performance du réseau – la descente de gradient stochastique (*Stochastic Gradient Descent*, SGD) (LeCun *et al.*, 1998). La performance est calculée avec une fonction de perte dérivable où celle-ci définit l'exactitude de la prédiction selon l'annotation. Minimiser la distance euclidienne entre la prédiction et la donnée souhaitée est un exemple de fonction de perte simple :

$$L = \|G_\theta(x) - y\|^2 \tag{1.2}$$

où L est la fonction de perte, $G_\theta(x)$ la prédiction du réseau paramétré par θ pour l'entrée x et y l'annotation. Lorsque la fonction de perte est égale à zéro, la prédiction correspond parfaitement à l'annotation. Le but, pendant le processus d'ajustement des paramètres, est

donc de minimiser cette fonction de perte. Comme toutes les opérations mathématiques reliant les paramètres à la fonction de perte sont dérivables, l'impact de chaque paramètre sur la fonction de perte peut être approximé avec la dérivée en chaîne. L'application de la dérivée en chaîne approxime un gradient spécifique à chaque paramètre. Le gradient calculé indique si une petite variation de ce paramètre a un impact, positif ou négatif, sur la fonction de perte. Ainsi, avec un algorithme d'optimisation comme la descente de gradient stochastique (LeCun *et al.*, 1998), les paramètres sont ajustés jusqu'à ce que la performance du réseau à effectuer la tâche est convenable (que la fonction de perte est suffisamment faible).

$$\theta \leftarrow \theta - \eta g \tag{1.3}$$

$$\text{où } g = \frac{\delta L}{\delta \theta} \tag{1.4}$$

où θ sont les paramètres d'un réseau, η définit le taux d'apprentissage pour éviter que l'apprentissage diverge, g représente le gradient issu de la dérivée en chaîne de la fonction de perte L par rapport aux paramètres θ . La différentiation automatique permet d'accélérer le calcul du gradient en utilisant une représentation des opérations sous forme de *graph* (Paszke *et al.*, 2017). Le gradient de la fonction de perte est calculé (le premier noeud du *graph*) et est rétropropagé jusqu'aux paramètres (les feuilles du *graph*). L'entraînement de réseaux avec la méthode SGD est un processus stochastique puisque chaque étape d'optimisation utilise un petit lot (une *Batch*, \mathfrak{B}) d'exemples pour estimer le gradient moyen sur l'ensemble des exemples. L'estimation du gradient est ensuite utilisée pour mettre à jour les paramètres θ du réseau (équation 1.3). Ainsi, dans un paradigme d'entraînement de réseaux par le biais du gradient, toutes les opérations qui relient les paramètres à la fonction de perte doivent être dérivables pour rétropropager le gradient jusqu'aux paramètres à optimiser. Lorsque le gradient est faible, voire nul, les paramètres du réseau ne sont pas bien mis à jour, ce qui entraîne une faible amélioration du réseau à effectuer la tâche. La convergence et la rapidité du processus d'optimisation découlent directement de la qualité du gradient à circuler dans le réseau (He *et al.*, 2016; Huang *et al.*, 2017).

1.10 Le réseau de neurones à convolution

Pour l'analyse d'image, le modèle de DL le plus utilisé est le réseau de neurones à convolution (*Convolutional Neural Network*, CNN). Le CNN comporte des couches successives de convolution où des caractéristiques progressivement complexes sont apprises (Goodfellow *et al.*, 2016). Les caractéristiques sont apprises en optimisant les poids des noyaux de convolution, les paramètres du réseau, présents dans chaque couche. Généralement, les premières couches de convolution apprennent des caractéristiques de bas niveau (p. ex. des lignes verticales, horizontales et diagonales) qui sont par la suite utilisées par la couche subséquente (Zeiler et Fergus, 2014). La composition hiérarchique des caractéristiques est optimisée pour obtenir des filtres de représentations abstraites capables de détecter la présence d'instances spécifiques

dans l'image originale. Les instances peuvent être des notions abstraites qui sont difficiles à définir au niveau des pixels d'une image (p. ex. l'ensemble des chats ou des chiens) (LeCun *et al.*, 2015).

Chaque couche de convolution comporte, entre autres, deux opérations distinctes. Premièrement, une opération de corrélation croisée, communément appelée convolution, extraie des caractéristiques du tenseur \mathcal{T} en entrée. Ces caractéristiques C , également appelées activations, canaux, dimensions ou représentations, sont apprises lors de la phase d'entraînement. Pour extraire différentes caractéristiques, plusieurs filtres \mathcal{F} sont appris. La taille des noyaux de convolution est généralement de 3×3 . Un noyau de convolution est appris pour chaque canal du tenseur en entrée, p. ex. avec un tenseur \mathcal{T} de taille $C \times H \times W$ (hauteur (*Height*, H) et largeur (*Width*, W)), les noyaux de convolution peuvent être exprimés sous la forme d'un tenseur de poids \mathcal{W} de taille $C \times 3 \times 3$. La convolution est calculée indépendamment pour chaque dimension, puis additionnée pour n'obtenir qu'une seule caractéristique par filtre par pixel :

$$\mathcal{F}_j = b_j + \sum_{k=0}^{C-1} \mathcal{W}_k \star \mathcal{T}_k \quad (1.5)$$

où k représente le $k^{\text{ième}}$ canal ou équivalent, la caractéristique du $k^{\text{ième}}$ filtre précédent. Au $j^{\text{ième}}$ filtre \mathcal{F}_j , correspond un tenseur de poids \mathcal{W} avec k filtre et un biais b_j . Ainsi, chaque filtre d'une couche de convolution utilise toutes les réponses des filtres de la couche précédente. Cela implique que le nombre de couches d'un réseau à convolution fait augmenter exponentielle la combinaison de ses filtres. La capacité des réseaux à acquérir des représentations abstraites provient, entre autres, de ces multiples combinaisons. Les opérations convolutives constituent la majorité des paramètres à optimiser. Par exemple, la section d'extraction de caractéristiques du célèbre modèle VGG (16 couches) (Simonyan et Zisserman, 2015) compte au total 14 723 136 paramètres, dont 99,94% sont des poids de convolution et le reste des paramètres pour la normalisation (Ioffe et Szegedy, 2015).

Deuxièmement, une opération non linéaire, dite fonction d'activation, est appliquée sur les caractéristiques résultantes de la convolution. Le rôle de la fonction d'activation est d'assurer la présence d'une non-linéarité entre les couches. L'imposition de la non-linéarité entre les couches de convolution permet au réseau d'approximer des fonctions hautement non linéaires. Sans elle, comme les fonctions de convolutions sont linéaires, peu importe le nombre de couches de convolution, le réseau peut être réécrit comme une fonction linéaire, une fonction qui ne peut que découper l'espace d'entrée en régions très simples (Hart *et al.*, 2000). Or, une tâche telle que la segmentation nécessite des transformations complexes de l'espace en entrée. C'est pourquoi les fonctions d'activation sont importantes et permettent d'obtenir un réseau pouvant approximer des fonctions remarquablement complexes. Les fonctions d'activations couramment utilisées sont : la sigmoïde qui comprime son entrée dans la plage 0 à 1, la tangente hyperbolique qui comprime dans la plage -1 à 1 et l'unité de rectification linéaire

(Rectifier Linear Unit, ReLU), exprimée par $ReLU(x) = \max(0, x)$. L'avantage des deux premières est qu'elles sont dérivables sur toute leur plage d'action $x \in \mathbb{R}$. Tandis que la ReLU est dérivable dans la plage d'action positive uniquement $x \in \mathbb{R}^+$. Cela signifie qu'avec une optimisation basée sur le gradient, les unités convolutionnelles qui sont désactivées par la ReLU n'obtiennent pas de gradient et les poids de convolution en amont ne sont pas modifiés. Pour permettre au gradient de circuler, une pente légèrement négative est ajoutée à la portion $x \in \mathbb{R}^-$ de la fonction ReLU, appelé *Leaky ReLU*, défini par :

$$LReLU(x) = \begin{cases} x, & \text{si } x \geq 0 \\ -a * x, & \text{si } x < 0 \end{cases}$$

où a représente la pente négative de la LReLU et est normalement défini à 0,01 (Maas *et al.*, 2013). Ainsi, dans la plage $x \in \mathbb{R}^-$, le gradient est retropropagé avec un facteur de a . Empiriquement, ce sont les opérations non linéaires rectifiantes (de type ReLU) qui offrent de meilleurs résultats pour l'entraînement de réseaux profonds (Xu *et al.*, 2015; He *et al.*, 2015).

Ces deux opérations forment la neurone à convolution, la composante de base de l'architecture des réseaux de neurones à convolution (CNN).

1.11 Le réseau pleinement convolutionnel

Les premiers CNN développés terminent avec un classificateur afin de classer l'image à son entrée (LeCun *et al.*, 1998). Malgré l'utilité de la classification des images, des applications nécessitent des tâches plus complexes telles que la détection et la segmentation. Cela est particulièrement vrai dans le domaine biomédical, où plus de 75% des tâches sont réparties entre la détection et la segmentation (Litjens *et al.*, 2017). Afin d'étendre les capacités du CNN à cette tâche et de permettre la segmentation d'une image, la dernière couche de classification est retirée et remplacée par des couches de convolutions (Long *et al.*, 2015). Le réseau pleinement convolutionnel (*Fully Convolutional Network*, FCN) obtenu permet de réaliser une segmentation, dite dense, d'une image pour une ou plusieurs classes.

Comme le CNN de classification, la première section du FCN, l'encodeur, permet principalement de transformer la représentation de l'image – dans l'espace du pixel – vers une représentation plus abstraite. Dans ce nouvel espace, la valeur des caractéristiques, des \mathbb{R} , renseigne sur des notions abstraites dans l'image originale, p. ex. la présence ou non d'un chat. L'apprentissage de cette représentation est possible en imposant une structure spécifique au réseau de neurones. L'idée est de transférer l'information spatiale de l'image vers une représentation abstraite de l'image. Ainsi, en règle générale, à chaque fois que le tenseur \mathcal{T} , de taille $C \times H \times W$, traverse une couche convolutionnelle, la taille spatiale, H et W , est réduite de moitié et le nombre de caractéristiques C , c.-à.-d. le nombre de filtres \mathcal{F} , est doublé (LeCun *et al.*, 1998). La réduction spatiale impose aux filtres d'extraire des caractéristiques pour

compresser l'image d'entrée. L'augmentation du nombre de caractéristiques, tel que mentionné précédemment, permet un agencement exponentiellement plus complexe de celles-ci.

La deuxième section du FCN, le décodeur, permet au réseau de retourner vers un nouvel espace des pixels, donc de décoder la représentation abstraite. Le nouvel espace obtenu à la fin du décodeur est dépendant de la tâche à effectuer. Pour une tâche de segmentation à M classes, la sortie du décodeur \mathcal{S} est de taille $M \times H \times W$. Elle comporte, pour chaque pixel en sortie, un vecteur de longueur M représentant la pseudo probabilité de ce pixel d'appartenir à chacune des M classes. Ainsi, la segmentation est effectuée en assignant la classe ayant la probabilité la plus haute pour chaque pixel. Pour retourner vers l'espace des pixels, à l'inverse de l'encodeur, la taille spatiale de la représentation apprise est doublée tandis que le nombre de caractéristiques est diminué par deux à chaque couche. Pour augmenter la taille spatiale, deux méthodes sont généralement utilisées : la convolution transposée (déconvolution) (Noh *et al.*, 2015) et le sur échantillonnage suivi d'une convolution. La convolution transposée offre de bons résultats en segmentation, mais comporte des artéfacts en damier, particulièrement visible dans des tâches de génération d'images (Odena *et al.*, 2016; ?). Le sur échantillonnage avec une méthode d'interpolation linéaire ou bicubique permet d'augmenter la taille spatiale du tenseur en entrée sans nécessiter de paramètres supplémentaires et élimine l'artéfact en damier. L'interpolation est suivie d'une ou deux opérations de convolution (Badrinarayanan *et al.*, 2017; Ronneberger *et al.*, 2015). La tâche de reconstruction du décodeur à partir de la représentation abstraite uniquement est ardue, particulièrement pour obtenir des frontières de segmentation précises. En effet, les frontières en contiennent beaucoup d'information puisqu'elles sont un signal haute fréquence dans l'image (Shannon, 1948).

1.12 Architectures de type U-Net

Pour améliorer les performances en segmentation des FCN de type encodeur-décodeur, des liens sont ajoutés pour transférer les caractéristiques apprises dans la section encodeur vers la section décodeur du réseau, et ce, pour chaque étage (Ronneberger *et al.*, 2015). La classe de FCN comportant des liens est communément appelée U-Net dans la littérature. Ces liens facilitent la tâche du décodeur puisqu'il a maintenant accès aux caractéristiques apprises par l'encodeur. Par exemple, les caractéristiques de bas niveau, telles que les frontières, sont extraites par l'encodeur et utilisées par le décodeur pour raffiner les frontières de segmentation à la sortie du FCN (figure 1.6). Du même coup, avec ces liens, la tâche de l'encodeur est également diminuée en enlevant la nécessité d'encoder l'information nécessaire à la reconstruction des frontières. Les frontières ne sont pas un concept abstrait dans une image, au contraire, un simple filtre de Sobel (Kanopoulos *et al.*, 1988) permet leur détection. Ainsi, encoder l'information associée aux frontières dans la représentation abstraite est contre-productif pour l'encodeur.

L'état de l'art de la segmentation médicale utilise des réseaux FCN avec des variantes de

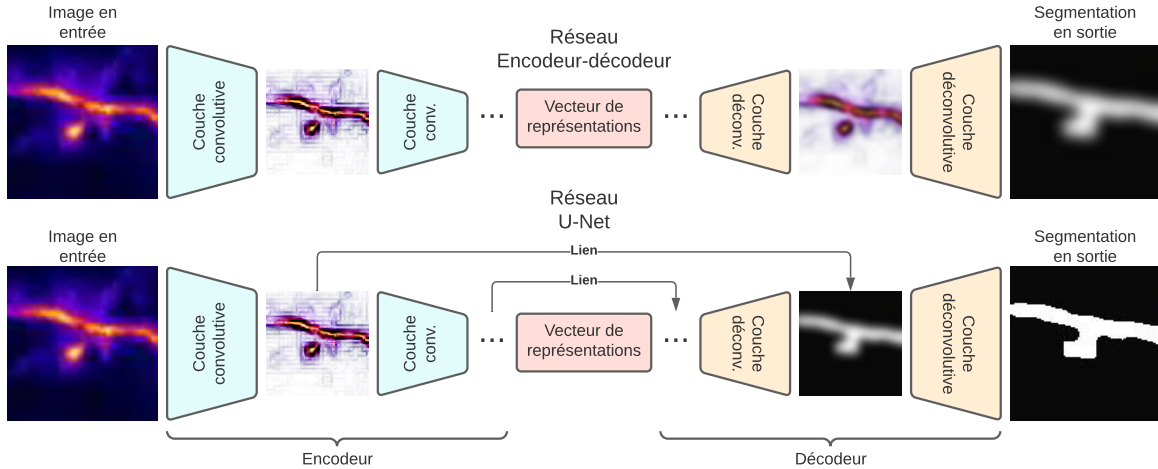


FIGURE 1.6 – Schémas des réseaux encodeur-décodeur et U-Net. Une tâche simple où les réseaux sont utilisés pour faire la détection de l’avant-plan de l’image d’entrée vers une segmentation en sortie. Les deux réseaux comportent une hiérarchie de couches convolutives dans la section encodeur. Symétriquement, la section décodeur comporte une hiérarchie de couches déconvolutives. L’encodeur-décodeur apprend à reconstruire la segmentation uniquement à partir de la représentation compressée au centre du réseau (le vecteur de représentations). Pour réduire la tâche du décodeur, celui-ci a également accès aux caractéristiques générées par l’encodeur. Avec cette information supplémentaire, le décodeur produit une segmentation plus précise en sortie.

type U-Net dans lesquelles les découvertes architecturales sont introduites dans les FCN. Tout d’abord, les réseaux résiduels (*ResNets*) améliorent l’entraînement de réseau profond par le biais de blocs résiduels (He *et al.*, 2016; Szegedy *et al.*, 2017). Le bloc résiduel somme les caractéristiques à son entrée et la transformation non linéaire de celles-ci. L’ajout de connexions identités, c’est-à-dire que l’entrée est additionnée à la sortie sans modification, permet de faire circuler efficacement le gradient dans des architectures profondes. Les *ResNet* ont été adaptés aux FCN pour la segmentation d’images médicales (Drozdzal *et al.*, 2016) en remplaçant les blocs convolutionnels du U-Net standard par des blocs résiduels.

Inspiré par les *ResNets*, le réseau *DenseNet* est ensuite proposé (Huang *et al.*, 2017). Ce réseau comporte des blocs, dits *dense*, dans lesquels les caractéristiques des couches sont additionnées itérativement aux caractéristiques des couches précédentes. Par exemple, dans un bloc *dense* de 3 couches, les caractéristiques de la couche 1 sont additionnées à celles de la couche 2 et les caractéristiques des couches 1 et 2 sont additionnées à celles de la couche 3. L’ajout de bloc *dense* permet une réutilisation intensive des caractéristiques apprises pour les couches subséquentes grâce au lien direct. La réutilisation des caractéristiques est avantageuse pour la tâche de segmentation et c’est cette motivation qui poussa au développement d’un réseau U-Net basé sur le *DenseNet* Jégou *et al.* (2017).

Les FCN de type U-Net ont également été élargis pour des données volumétriques (Milletari *et al.*, 2016; Çiçek *et al.*, 2016). Dans ce paradigme tridimensionnel, les filtres de convolutions ont trois dimensions et la corrélation croisée est effectuée sur les trois axes. L’impact immédiat avec les réseaux appliqués aux données volumétriques est l’augmentation d’un ordre de grandeur du nombre de paramètres. Les noyaux de convolution passent d’une taille de $C \times H \times W$ (en deux dimensions) à une taille de $C \times D \times H \times W$ où D est la profondeur (*depth*, D). Cette augmentation radicale du nombre de paramètres implique une augmentation du nombre de données pour les ajuster (Vapnik, 2013). De plus, le calcul de la corrélation croisée sur une dimension supplémentaire engendre un nombre d’opérations d’un ordre de grandeur supplémentaire par rapport aux modèles 2D. Le FCN tridimensionnel (FCN 3D) extrait une hiérarchie de caractéristiques tridimensionnelles. Il apprend à comprimer le volume en entrée vers un tenseur de représentation abstraite. Contrairement au FCN 2D, les représentations peuvent décrire des instances tridimensionnelles complexes, p. ex. des tumeurs (Zhao *et al.*, 2020). La représentation abstraite est décodée vers l’espace des voxels – un voxel correspond à un *pixel* volumétrique. Les FCN 3D peuvent, entre autres, segmenter un volume de voxels selon différentes classes.

Dû à leur capacité de segmentation précise, les architectures de type U-Net (2D et 3D) sont massivement utilisées dans le domaine biomédical (Drozdal *et al.*, 2016; Litjens *et al.*, 2017; Deniz *et al.*, 2018; Haque et Neubert, 2020; Taghanaki *et al.*, 2021; Zhou *et al.*, 2020; Ibtehaz et Rahman, 2020).

1.13 Apprentissage profond appliqué à l’imagerie calcique

Les approches de DL ont été maintes fois appliquées avec succès sur des données d’imagerie de fluorescence (Van Valen *et al.*, 2016; Caicedo *et al.*, 2019; Falk *et al.*, 2019; Raza *et al.*, 2017; Sadanandan *et al.*, 2017). Plus précisément, l’application du DL à l’imagerie calcique de fluorescence a connu un essor grâce à la disponibilité de données publiques via la compétition NeuroFinder¹ et les données publiques du Allen Brain Observatory². L’imagerie calcique de culture de neurones génère des données spatiotemporelles, soit une vidéo temporelle d’un plan focal. Trois approches sont utilisées pour l’entraînement de réseaux profonds sur des données calciques spatiotemporelles : 1) 2D : analyse spatiale sur la projection du vidéo 2) (2+1)D : analyse spatiale et temporelle 3) 3D : analyse entièrement tridimensionnelle. La première méthode utilise une projection moyenne du vidéo pour ensuite entraîner un FCN à détecter les neurones (Klibisz *et al.*, 2017). La projection consiste à calculer une moyenne temporelle pour chaque pixel du vidéo. Comme deux neurones peuvent se chevaucher dans ce type d’imagerie, la projection fusionne ces neurones et complexifie la segmentation. Pour pallier à cette difficulté, une architecture proposée consiste à appliquer un FCN sur chaque image du vidéo et de

1. <http://neurofinder.codeneuro.org/>

2. <http://observatory.brain-map.org/visualcoding>

fusionner cette information via une mise en commun maximale temporelle (*max-pooling*) (Apthorpe *et al.*, 2016). Par contre, les performances n'étaient pas significativement meilleures que l'architecture 2D. La dernière approche propose de traiter les données d'imagerie spatiotemporelle comme un volume 3D de dimensions $T \times Y \times X$ (Tran *et al.*, 2015). Le premier avantage avec ce type de réseau est qu'il va apprendre lui-même les caractéristiques spatiotemporelles utiles à la tâche. Le deuxième avantage est la rapidité d'inférence grâce aux accélérations GPU disponibles avec les bibliothèques standards de DL. Le réseau FCN 3D offre les meilleurs résultats pour la segmentation de neurones dans la compétition NeuroFinder (Soltanian-Zadeh *et al.*, 2019).

Cependant, l'application de méthodes de DL sur la détection et la segmentation de mSCTs ne connaît pas autant d'essor pour deux raisons : 1) le manque de jeu de données publiques disponibles pour cette tâche 2) l'annotation précise et dense des mSCTs par un expert est laborieuse. Les mSCTs de faibles intensités sont difficilement discernables du bruit de fond de l'imagerie calcique. Leurs frontières sont floues. De plus, leur petite dimension par rapport au champ de vue facilite l'omission de mSCTs par mégarde pendant l'exercice d'annotation. Pour ces raisons les méthodes de références actuelles sont basées sur des approches classiques, telles que mentionnées à la section 1.5 (Rueckl *et al.*, 2017; Mancini *et al.*, 2018; Tian *et al.*, 2019).

Chapitre 2

Méthodologie

La méthodologie pour adapter la procédure d'entraînement des modèles d'apprentissage profond aux contraintes spécifiques de l'imagerie calcique des mSCTs est décrite. Les données et leurs caractéristiques sont d'abord présentées. S'ensuit une brève description des méthodes classiques de référence pour la segmentation des mSCTs. Les segmentations de l'une d'elles sont utilisées pour extraire les échantillons positifs (P) du jeu d'entraînement. Étant donné les caractéristiques spécifiques des données calciques, il est possible d'extraire des échantillons non étiquetés (U) et d'appliquer une méthode d'entraînement inspirée de l'apprentissage de données positives et non étiquetées (apprentissage PU). Les détails associés à l'extraction des échantillons P et U, l'entraînement des algorithmes d'apprentissage profond et l'analyse des performances y sont finalement décrits.

2.1 Données d'imagerie calcique

La microscopie TIRF permet d'imager tout le champ de vue à chaque temps d'acquisition (Axelrod, 2001). Par exemple, la figure 2.3(a) montre plusieurs plans d'un neurone à différent temps. L'acquisition du champ de vue à 150 ms permet l'échantillonnage de la décroissance de fluorescence du senseur GCaMP6f qui possède un temps de demi-vie d'environ 135 ms (Helassa *et al.*, 2015). La taille du champ de vue est maximisée pour détecter le plus grand nombre de mSCTs. L'image du plan focal comporte 512×512 pixels où chaque pixel a une taille de 160×160 nm. Une surface d'environ $82 \times 82 \mu\text{m}$ est ainsi imagée. Avec cette surface, les mSCTs occupent au minimum 3×3 pixels. La surface imagée ($82 \times 82 \mu\text{m}$) couvre seulement une section du dendrite (un exemple de dendrite est présenté à la figure 2.1(b)).

La sensibilité de la caméra image les variations de fluorescence avec une plage dynamique large. Par exemple, à la figure 2.1(b), l'intensité varie d'environ 1k dans l'arrière-plan à 4k sur le dendrite. Cela permet de caractériser les augmentations drastiques de fluorescence (jusqu'à des dizaines de milliers d'unités arbitraires) lors d'un mSCT. L'unité arbitraire (u. a.) est utilisée pour quantifier l'intensité d'un pixel pour deux raisons : 1) comme une caméra

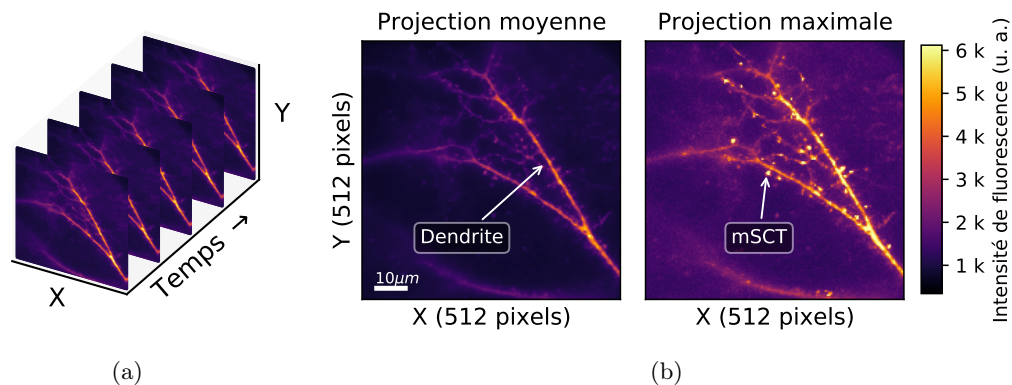


FIGURE 2.1 – Vidéo calcique d’un dendrite imagé par microscopie TIRF. (a) Le champ de vue complet est capturé à chaque période d’acquisition du TIRF. (b) La résolution des vidéos est de 512×512 . La projection moyenne et maximale représente la valeur maximale et moyenne de chaque pixel pour le temps d’imagerie. La plage dynamique est grande, comme le montre l’intensité moyenne de l’arrière-plan et du dendrite. Lors d’un événement calcique (mSCT dans la projection maximale), l’intensité de fluorescence peut aller jusqu’à 7 fois l’intensité moyenne.

conventionnelle, la valeur mesurée est une conversion non linéaire des photons reçus ; et 2) l’intensité de fluorescence détectée peut changer d’une expérience à l’autre, puisqu’elle est ajustée pour ne pas saturer les capteurs de la caméra. Par exemple, la figure 2.12(b) montre comment la fluorescence moyenne de l’avant-plan (la zone sur le dendrite) varie d’un vidéo à l’autre (entre 2k et 8k).

Dans un vidéo, des événements calciques ont lieu aléatoirement dans le temps et sur le dendrite (fig. 2.2(a)). La dimension moyenne d’un mSCT est d’environ 14×14 pixels et dure en moyenne 1,5 s (10 acquisitions à 150 ms). La figure 2.2(b) montre que les mSCTs sont de petites dimensions par rapport à la taille de l’image. Les mSCTs ont lieu sur une épine dendritique ou sur un dendrite. Ceux sur le dendritique sont normalement allongés, car le calcium entrant diffuse dans celui-ci. Tandis que les mSCTs provenant d’une épine restent localisés sur celle-ci. La figure 2.2(c) présente la durée des mSCTs sur une plage de 7,5 secondes. Un vidéo complet dure au minimum 90 secondes afin de détecter une quantité suffisante de mSCTs récurrents. La figure 2.2(c) montre un épine active où deux mSCTs récurrents y ont lieu. L’impact de la récurrence des mSCTs sur la plasticité synaptique est très intéressant à quantifier. Bien que l’exemple de la figure 2.2 comporte beaucoup de mSCTs pour le bien de la visualisation, les vidéos ont en moyenne un ratio de pixels représentant un mSCT d’environ 0,03 %. Les mSCTs sont donc très épars dans un vidéo.

Les dimensions d’un mSCT varient de quelques μm sur une synapse à une dizaine de μm sur l’arbre dendritique. À la figure 2.3, un mSCT est isolé et présenté sous différentes coupes spatiotemporelles (le mSCT est encadré en jaune à la figure 2.2(b) et 2.2(c)). Le mSCT a lieu

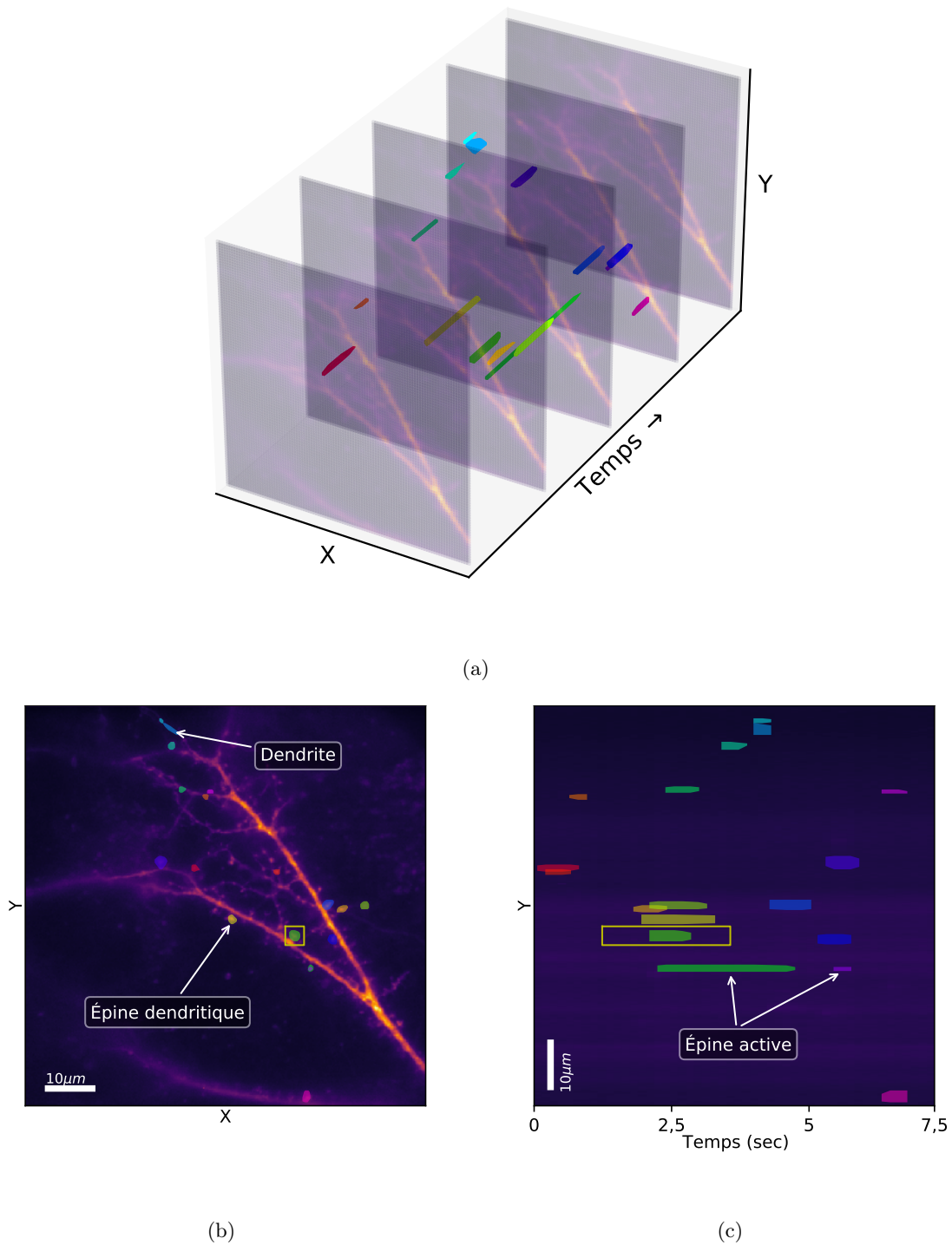


FIGURE 2.2 – Transients Ca^{2+} synaptiques miniatures (mSCTs) dans un vidéo. (a) Une section de 7,5 secondes avec les mSCTs segmentés. (b) Les mSCTs sont ajoutés sur la projection moyenne du vidéo. Les événements calciques sont localisés et ont lieu au niveau d'une épine dendritique ou directement sur le dendrite (c) Vue temporelle de la segmentation des mSCTs. Plus d'un événement calcique (mSCT) peut avoir lieu dans une même épine, par exemple l'épine pointée comporte deux mSCTs. La région jaune est présentée en détail à la figure 2.3.

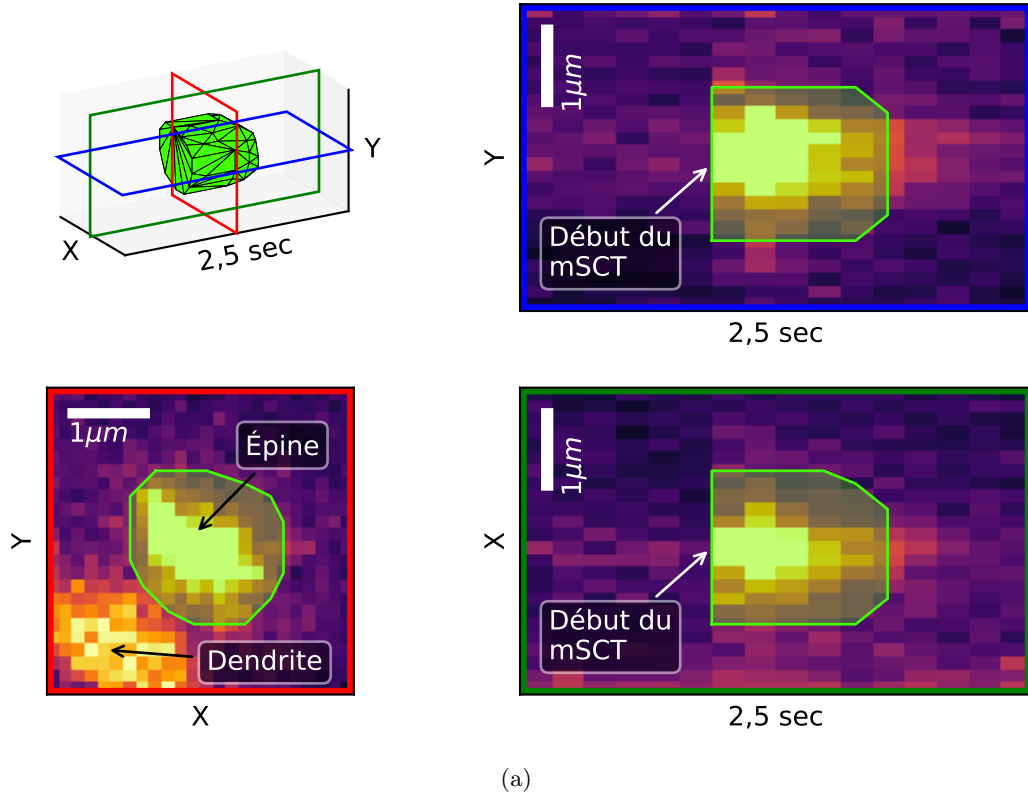


FIGURE 2.3 – Zoom d'un mSCT sur une épine dendritique. Une coupe spatiale (bas-gauche) et deux coupes spatiotemporelles centrées sur le mSCT encadré à la figure 2.2. Le mSCT, de dimension de $1,5\mu\text{m}\times 1,5\mu\text{m}$ dure approximativement 1 seconde. L'augmentation rapide d'intensité de fluorescence est visible dans les deux coupes spatiotemporelles de droites. Les zones vertes représentent la région de segmentation de l'événement calcique.

sur une épine et est situé à proximité du dendrite. Sans la composante temporelle, la distinction entre un événement calcique actif et un dendrite peut être difficile. Ainsi, avec l'information des deux coupes spatiotemporelles de la figure 2.2, le début d'un mSCT est facilement visible par l'augmentation drastique d'intensité.

Une méthode couramment utilisée pour comparer et analyser les changements d'intensité dans un signal calcique est la normalisation $\Delta F/F$ (Macleod, 2012). La normalisation est essentielle puisque la concentration de senseurs calciques n'est pas constante entre les neurones de même qu'à travers le dendrite. La concentration de senseurs calciques influence directement le niveau de fluorescence basale. Par exemple, à la figure 2.4, deux traces calciques situées à environ $10\mu\text{m}$ l'une de l'autre sont extraites. La trace bleue comporte un signal basal (F_0) de 1,4k et un pic d'intensité de 5,0k, tandis que la trace rouge comporte un F_0 de 1,3k et un pic de 2,2k. La trace basale (F_0) est le niveau de fluorescence moyen du pixel et peut être approximée avec la méthode airPLS (Baek *et al.*, 2015). La méthode est supérieure à une moyenne mobile puisqu'elle évite de tirer l'intensité basale vers le haut en présence des pics d'intensité (la

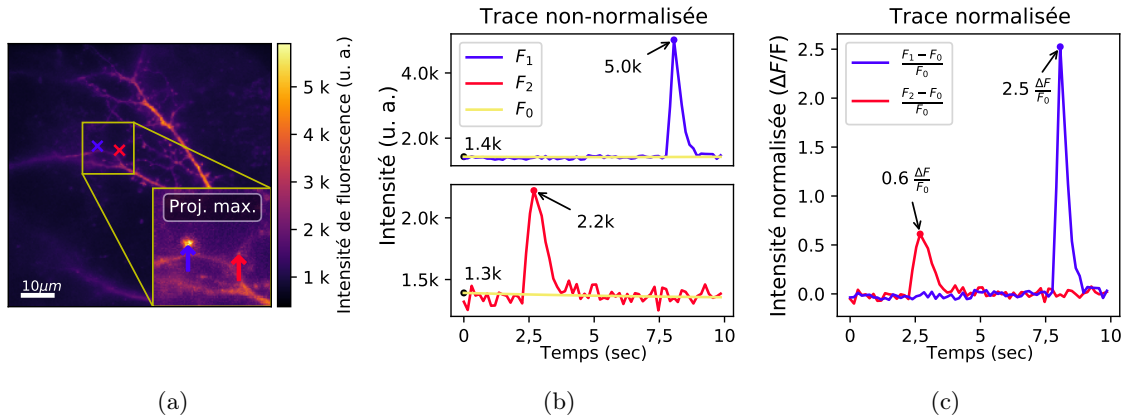
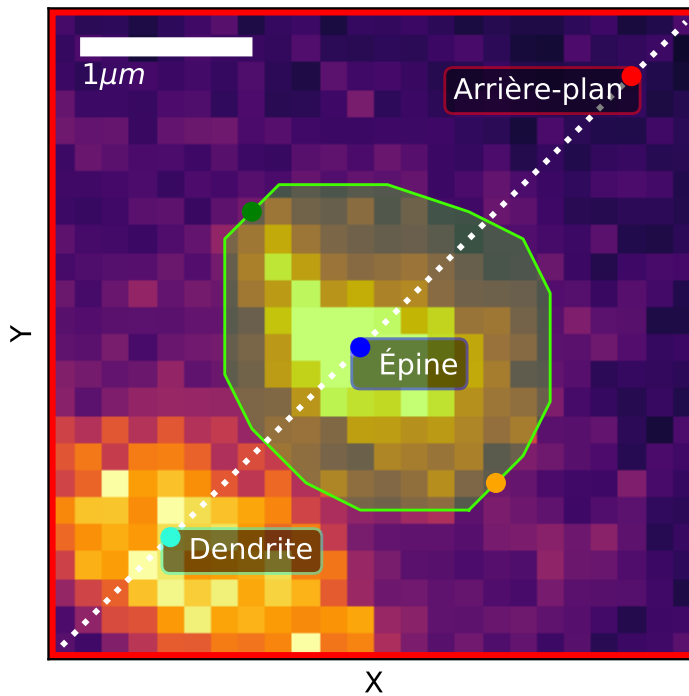


FIGURE 2.4 – Normalisation d’une trace calcique de fluorescence. (a) Deux événements calciques (mSCTs) de différente intensité de fluorescence (bleu et rouge) sont représentés sur la projection maximale de la section zoomée. (b) L’intensité de fluorescence (en unité arbitraire) des deux événements détectés ainsi que la trace basale (F_0) utilisée pour normaliser. L’intensité basale est l’intensité maximale des événements sont différentes. (c) Les traces normalisées avec la méthode $\Delta F/F_0 = F - F_0/F_0$. Avec cette normalisation, l’intensité des mSCTs est dans un même référentiel. Les deux pics annotés sont les valeurs d’intensités utilisées (la valeur maximale) pour quantifier l’amplitude normalisée d’un mSCT.

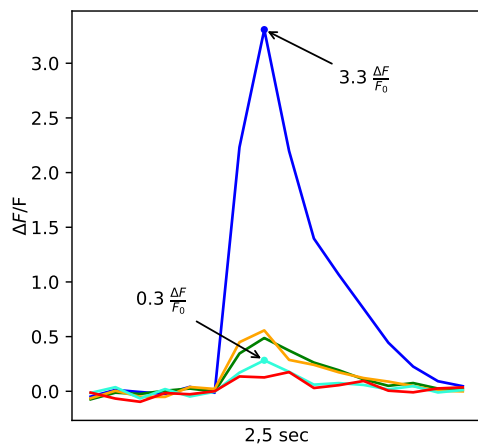
trace jaune reste stable en présence des pics à la figure 2.4(b)). Avec la normalisation $\Delta F/F$, les traces bleu et rouge peuvent effectivement être comparées. L’intensité normalisée d’un événement représente l’intensité maximale, centrée sur l’événement, atteinte pendant toute la durée du mSCT (p. ex. 0,6 et 2,5 à la figure 2.4(c)).

La détection et la segmentation d’un mSCT sont grandement corrélées avec l’intensité normalisée. Les méthodes actuelles de segmentation d’événements calciques sont d’ailleurs basées sur un seuillage de l’intensité normalisée (Mancini *et al.*, 2018; Tian *et al.*, 2019; Sun et Südhof, 2021). La difficulté majeure avec ces méthodes est le choix du seuil. Un seuil élevé permet généralement une segmentation précise des événements de fortes intensités, mais ignore ceux de faible intensité. Cependant, quand le seuil de détection défini est trop bas, la région segmentée n’est plus contrainte sur l’événement calcique, mais sur les structures aux alentours également. Par exemple, dans la figure 2.5, un seuil inférieur à $0,3 \Delta F/F_0$, inclura la zone dendritique (bleu pâle) qui ne fait pas partie de l’épine dans laquelle a lieu le mSCT. L’information temporelle c’est-à-dire la variation de l’intensité normalisée, utilisée par ces méthodes pour segmenter les mSCTs est limitative. Il serait intéressant d’incorporer des caractéristiques spatiales dans un algorithme de segmentation d’événements calciques.

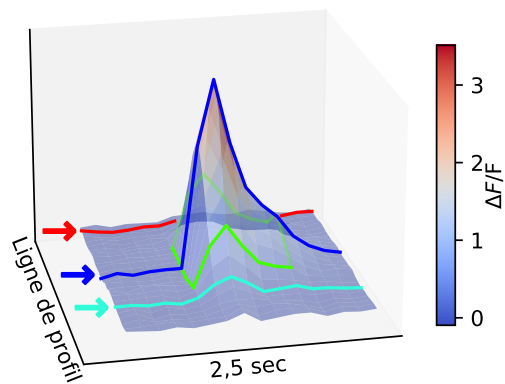
Pour conclure sur les données calciques, les caractéristiques importantes, qui impactent les choix subséquents de la méthodologie, sont : 1) les instances d’intérêt à segmenter, les mSCTs, sont très petites par rapport au champ de vu ; 2) Les mSCTs sont également très épars et



(a)



(b)



(c)

FIGURE 2.5 – Variation spatiale de fluorescence en fonction du temps. (a) Vu de face du mSCT (même mSCT qu'à la figure 2.3). La trace calcique normalisée est calculée à divers endroits autour de l'événement calcique (points de couleurs). Entre autres, elle est calculée sur l'arbre dendritique (magenta), au centre de l'épine dendritique (bleu) et dans l'arrière-plan (rouge). (b) Les traces calciques normalisées ($\Delta F/F$) de fluorescence. L'intensité normalisée maximale atteinte pendant le mSCT est de 3,3 sur l'épine et de 0,3 sur le dendrite. (c) L'intensité calcique normalisée de la trace de profil (ligne pointillée blanche en (a)) est présentée. Les traces d'intérêt y sont présentées. La ligne verte englobe la zone de segmentation spatiale et temporelle de l'événement calcique.

émergent aléatoirement dans un vidéo; et 3) la segmentation manuelle des événements calciques dans un vidéo est une tâche laborieuse.

2.2 Méthodes de référence

Les méthodes de référence pour la détection et la segmentation de petits événements calciques spontanés sont actuellement basées sur des analyses par pixel de seuillage du signal de fluorescence.

2.2.1 Procédure de seuillage *MiniFinder*

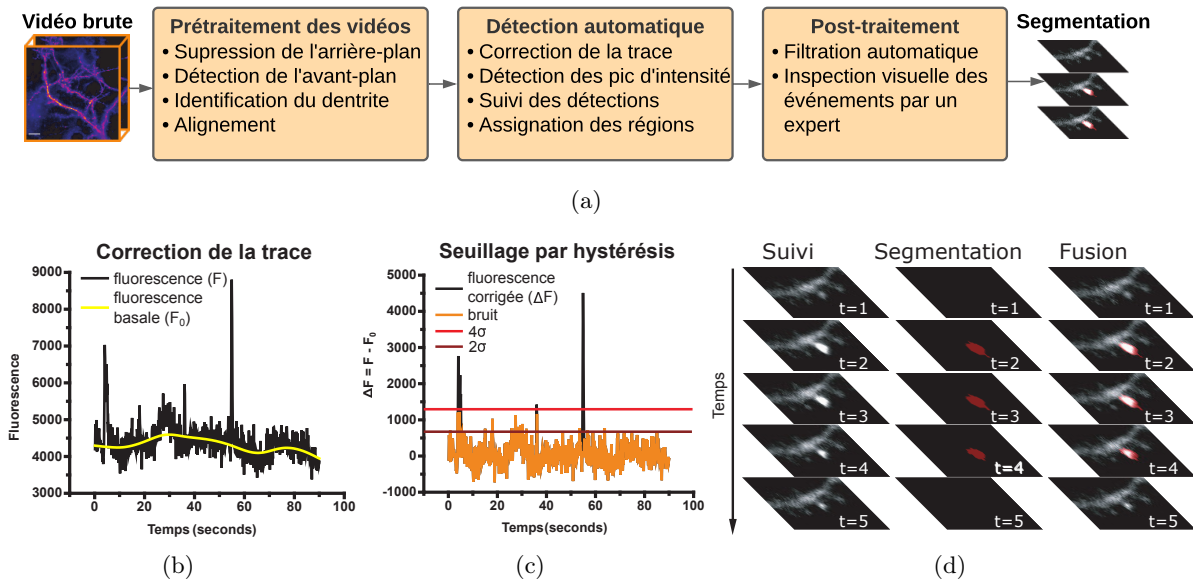


FIGURE 2.6 – Procédure de seuillage *MiniFinder* pour la détection et la segmentation des événements calciques spontanés. (a) Séquence d’opérations appliquées sur le vidéo brut jusqu’à la segmentation. (b) Trace calcique brute, moyennée sur un noyau spatial 5×5 ainsi que la fluorescence basale (*jaune*). (c) Après correction de la trace, il y a détection si le signal est supérieur au seuil 4σ , où σ représente l’écart-type du bruit calculé à partir du signal calcique corrigé (*orange*). Cette détection est suivie dans le temps et se termine lorsque le signal est inférieur au seuil 2σ . (d) Exemple de détection d’un événement calcique. La procédure est semi-automatique, certaines étapes du pré- et post-traitement sont effectuées manuellement. Adaptée de Wiesner (2021).

La procédure *MiniFinder* est une méthode semi-automatique, développée à l’interne, qui permet la détection et la segmentation des événements calciques spontanés. La routine *MiniFinder* est utilisée pour générer les masques de segmentation avec lesquelles les réseaux de neurones sont entraînés. Son fonctionnement général est présenté ici, pour de plus amples détails, se référer à la thèse de Theresa Wiesner (Wiesner, 2021).

Prétraitement des vidéos. Dû à des facteurs externes, des décalages minimes de l'échantillon sur le plan d'imagerie sont possibles (de l'ordre du μm). C'est pourquoi une étape d'alignement est nécessaire pour éliminer ce décalage (Guizar-Sicairos *et al.*, 2008). Par la suite, les pixels de l'arrière-plan sont retranchés des pixels à analyser. En effet, 98,4% des mSCTs sont situés dans la zone dendritique formant l'avant-plan (Mancini *et al.*, 2018). Comme environ 89,3% des pixels font partie de l'arrière-plan (table 2.1), un gain computationnel équivalent est possible au détriment des 1,6% mSCTs écartés. Pour ce faire, un masque grossier est manuellement tracé autour du neurone à analyser. Dans ce masque, un seuil est appliqué pour raffiner le masque davantage autour du neurone. Par la suite, le squelette du neurone est manuellement tracé afin d'estimer sa longueur pour ainsi normaliser le nombre d'événements détectés selon la longueur dendritique présente dans le champ de vue.

Détection automatique. Ensuite, pour chaque pixel à analyser, la trace calcique est extraite avec un noyau moyennant de 5×5 dans le plan spatial (*signal noir*, fig. 2.6(b)), et la fluorescence basale F_0 est calculée avec la méthode airPLS (Baek *et al.*, 2015) (*signal jaune*, fig. 2.6(b)) puis soustraite à la trace calcique originale F pour obtenir la trace corrigée $\Delta F = F - F_0$ (*signal noir*, fig. 2.6(c)). L'écart-type σ du bruit de la trace corrigée est calculé en enlevant les pics de fluorescence (*signal orange*, fig. 2.6(c)). L'écart-type σ est ensuite utilisé pour définir deux seuils, 2σ et 4σ . Une méthode par hystérésis est utilisée pour identifier les pixels d'intérêt. Un pixel est sélectionné lorsque son intensité dépasse le seuil 4σ (*ligne rouge*, fig. 2.6(c)). Par la suite, le pixel est suivi dans le temps jusqu'à ce que son intensité descende en dessous du deuxième seuil 2σ (*ligne bourgogne*, fig. 2.6(c)). Finalement, les pixels voisins sont agglomérés. Chaque agglomérat constitue la segmentation d'un événement calcique.

Post-traitement. Les événements calciques sont filtrés selon leurs caractéristiques spatiales et temporelles. Les détections avec une dimension plus petite que 3 pixels en x ou y sont éliminés. De même, celles d'une durée de 150 ms (une seule image) sont éliminées. En effet, un événement avec ces dimensions est fort probablement une mauvaise segmentation causée par la variation d'intensité due, entre autres, à du bruit de grenaille ou encore à un corps étranger. Finalement, un expert effectue une inspection visuelle pour assurer la qualité des détections et éliminer les faux positifs. Une bonne détection peut être vérifiée lorsque la trace calcique correspond à une augmentation rapide de l'intensité du marqueur suivi d'une décroissance exponentielle de l'intensité. En moyenne, la fluorescence maximale est atteinte après 300 ms et redescend au niveau basale après 3 s.

L'avantage de cette méthode est qu'elle est manuellement paramétrable et facilement interprétable, contrairement à des méthodes d'apprentissage profond par exemple. Effectivement, grâce aux seuils de détection définis manuellement, l'interprétation se résume à la comparaison de l'intensité par rapport aux seuils. Par contre, un inconvénient prépondérant est l'impact de la valeur des seuils choisis. Ceux-ci influencent à la fois la segmentation d'un événement, la détection d'événements la quantité de faux positifs. Par exemple, si le seuil est trop bas, le

bruit présent dans le signal dépasse le premier seuil et crée de faux positifs. Cela permet de détecter des événements de petites amplitudes, au détriment du temps d’inspection visuelle par un expert pour éliminer les faux positifs. Les seuils ont été choisis empiriquement pour avoir un bon compromis entre la quantité d’événements détectée et le temps d’inspection visuelle nécessaire. Pour ce faire, les traces calciques ainsi que les détections de plusieurs vidéos choisis aléatoirement sont inspectées par un expert qui, par essai-erreur, détermine les seuils selon ce compromis. La méthode est implémentée sous MATLAB (MATLAB, 2016). *MiniFinder* est utilisé pour générer les masques de segmentations utilisés pour extraire les échantillons positifs (P) de la méthode d’apprentissage PU.

2.2.2 Procédure de seuillage *SICT*

L’approche qui s’apparente le plus à *MiniFinder* est la méthode proposée par Mancini *et al.* (2018), appelée *SICT*. Leur méthode est développée pour détecter et segmenter des événements Ca^{2+} spontanés sur des dendrites. La procédure proposée, tout comme *MiniFinder*, est semi-automatique et basée sur une analyse par seuillage. L’intensité de fluorescence à chaque pixel est seuillée avec un seuil calculé à partir des statistiques de l’image (moyenne et variance). L’effet du choix de ce seuil sur la sensibilité de l’algorithme est également présent. En effet, pour détecter un maximum d’événements calciques, le seuil doit être diminué et cela engendre beaucoup de fausses détections. Afin de pallier à cette lacune, une interface intuitive est disponible pour faciliter et accélérer l’inspection visuelle. La méthode proposée comporte cinq étapes ; 1) le vidéo est filtré avec un filtre gaussien 3D pour éliminer les hautes fréquences spatiales et le bruit temporels ; 2) le signal de fluorescence est normalisé pour chaque pixel avec une transformation $\Delta F/F_0$ où F_0 est la fluorescence basale calculée avec une moyenne glissante ; 3) à partir des statistiques du vidéo, un seuil est calculé et appliqué sur le signal normalisé ; 4) les agglomérats de pixels d’intérêts interconnectés sont assignés à des régions ; et 5) les régions sont inspectées visuellement.

Le calcul automatique du seuil, dans la méthode de Mancini *et al.* (2018), semble être bien adapté pour leur type de données exclusivement. En l’utilisant tel quel avec les données d’imagerie calcique TIRF, les masques des événements calciques sont trop larges. Afin de faire une comparaison équitable, le calcul du seuil automatique a été adapté aux données d’imagerie calcique TIRF. En augmentant le seuil, les masques sont moins larges, mais le nombre d’événements calciques détectés diminue. Par essai-erreur, le seuil a été ajusté au mieux pour avoir des performances similaires à *MiniFinder*. Le code source est ouvert à tous et implémenté sous la version 2016b de MATLAB (MATLAB, 2016).

2.3 Apprentissage PU

Les algorithmes classiques d'apprentissage supervisés sont entraînés en utilisant des données et leurs étiquettes associées. Les étiquettes peuvent être regressées ou classifiées (Alpaydin, 2020). Considérant le cas d'une classification binaire, où l'étiquette $\{0, 1\}$ est associée à la classe négative et positive respectivement. Soit N paires d'échantillons $\{(x_1, y_1), \dots, (x_N, y_N)\}$ où $x_i \in \mathbb{R}^d$ est un vecteur de caractéristiques de d dimensions et $y_i \in \{0, 1\}$ est une étiquette binaire associée au vecteur x_i . Le but de l'apprentissage machine, dans ce cas binaire, est donc d'apprendre une fonction discriminative $g(x|y) \rightarrow y$ qui permet d'inférer l'étiquette associée à une nouvelle donnée (Friedman *et al.*, 2009).

Dans certains cas particuliers, cette dernière affirmation ne peut pas toujours être respectée. Par exemple, supposons que l'on souhaite prédire les intérêts d'un usager à partir de sa navigation sur une page internet (l'étiquette $y = 1$ est associée à un sujet qui intéresse l'utilisateur) (Chang *et al.*, 2016). Lorsque l'utilisateur clique sur un lien, il est possible de dire avec une certaine confiance que le lien a un intérêt pour l'utilisateur ($y = 1$), mais cela n'implique pas nécessairement que tous les autres liens de la page web étaient inintéressants ($y \stackrel{?}{=} 0$). De cette façon, à moins de demander explicitement à l'utilisateur d'identifier les liens qui ne l'intéressent pas, la classe négative ne sera jamais étiquetée.

Un autre exemple dans le domaine du diagnostic médical expose comment les étiquettes négatives ne peuvent pas être étiquetées, où du moins, sans un processus très coûteux (Claesen *et al.*, 2015). Pour un patient, il est relativement facile de valider si celui-ci possède une maladie quelconque (l'étiquette $y = 1$ est associée à une maladie contractée par le patient), en effectuant le test de dépistage associé et c'est cette information qui est enregistrée dans le dossier du patient. Par contre, pour vérifier si le patient n'a pas toutes autres maladies ($y \stackrel{?}{=} 0$), il aurait fallu effectuer un dépistage pour chacune d'entre elles, ce qui n'est pas réaliste. Le dossier d'un patient contient donc uniquement les données étiquetées positives et les autres ne sont pas connues (non étiquetées).

Ces domaines d'applications, entre autres, motivent la recherche dans le domaine de l'apprentissage à partir de données positives et non étiquetées (*Positive and Unlabeled*, PU). L'apprentissage PU a été formalisé mathématiquement au début des années 2000 (Denis, 1998; Liu *et al.*, 2003; Denis *et al.*, 2005). Ce domaine se positionne dans un contexte de classification binaire à partir d'étiquettes positives (*Positive*, P) et d'étiquettes non étiquetées (*Unlabeled*, U). Cette méthode d'apprentissage est également efficace lorsque les échantillons U sont disponibles à moindre coût et en grandes quantités.

Il est possible à partir du formalisme de classification binaire exprimé plus haut, soit N paires d'échantillon $\{(x_1, y_1), \dots, (x_N, y_N)\}$ avec $x_i \in \mathbb{R}^d$ et $y_i \in \{0, 1\}$, de reformuler le problème.

En effet, les étiquettes négatives ne sont pas accessibles, selon (Bekker et Davis, 2020), soit :

$$\begin{aligned} \mathbf{x} &\sim f(x) \\ &\sim \alpha f_+(x) + (1 - \alpha) f_-(x) \end{aligned} \quad (2.1)$$

où $f(x)$ représente la distribution théorique à laquelle appartiennent les données x . Les échantillons x pigés de la distribution $f(x)$ ont une étiquette binaire. Il est ainsi possible de définir distribution $f(x)$ comme la somme pondérée de deux distributions : f_+ et f_- . Elles sont associées à la distribution de données x étiquetées positives ou négatives respectivement. Le facteur $\alpha = P(y = 1)$ représente la probabilité à priori qu’une donnée x appartienne à la classe positive. La connaissance du facteur α est important puisqu’il permet d’adapter la stratégie d’entraînement. Par contre, en pratique, il est difficile de le calculer en raison des étiquettes manquantes.

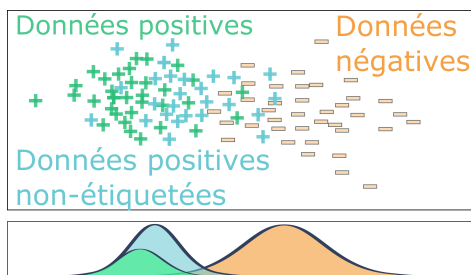


FIGURE 2.7 – Exemples de distributions de données PU où les vraies distributions des données positives (*vert*), positives non étiquetées (*bleu*) et négatives (*orange*) sont connues. Il y a un biais présent dans l’étiquetage des données positives. Plus une donnée positive est *semblable* à une donnée négative, moins sa probabilité d’être étiqueté est grande, d’où l’augmentation de la courbe en bleu entre les données positives et négatives. Inspirée de Bekker et Davis (2020).

Dans un contexte d’apprentissage PU, le but reste le même, soit d’effectuer la classification binaire des instances x . Par conséquent, les algorithmes d’apprentissage PU suggère différentes solutions pour traiter les étiquettes U (Bekker et Davis, 2020). Les solutions à employer sont fondées sur des hypothèses *a priori* sur la distribution des données U. Voyons tout d’abord les deux possibilités d’une donnée U : 1) la donnée U est réellement négative et elle n’est pas étiquetée ; ou 2) la donnée U est positive, mais elle n’a pas été sélectionné pour être étiquetée (on considère que le mécanisme d’étiquetage ne fait pas de faux positif). Cela signifie que si le mécanisme d’étiquetage a manqué un échantillon P, cet échantillon va se retrouver dans la distribution d’échantillon U – le mécanisme d’étiquetage influence la distribution des U. Or, le mécanisme d’étiquetage est possiblement influencé par les données elles-mêmes. À la figure 2.7, parmi la vraie distribution des données P, seule une sous-distribution est effectivement sélectionnée et étiquetée par le mécanisme d’étiquetage. La sélection du mécanisme semble moins efficace lorsque la donnée se *rapproche* de la distribution des données négative. Par exemple, les mSCTs sont plus difficiles à annoter lorsque leur intensité est faible et près du rapport signal sur bruit. Pour ces raisons, des hypothèses par rapport aux données elles-mêmes, c.-à-d.

le ratio α et des hypothèses par rapport au mécanisme d'étiquetage permettent de définir comment les étiquettes U sont gérées dans le contexte de l'imagerie calcique des mSCTs.

Les données d'imagerie Ca^{2+} de mSCTs sont introduites à la section 2.1 et le mécanisme d'étiquetage utilisé est la procédure *MiniFinder* détaillée à la section 2.2.1. Trois affirmations reliées aux données et au mécanisme d'étiquetage peuvent être formalisées :

Affirmation 1 *Les mSCTs et leur étiquetage.*

- (1) *Chaque pixel d'un vidéo est classifié $y \in \{0, 1\}$ par MiniFinder. Un pixel de la classe 1 appartient à un mSCT, tandis qu'un pixel de la classe 0 correspond à un pixel non étiqueté par la procédure ou rejeté par l'expert lors de l'annotation manuelle.*
- (2) *Les mSCTs sont petits et apparaissent aléatoirement sur le neurone.*
- (3) *Les mSCTs sont épars dans le temps et dans le champ de vue.*

À partir de ces affirmations, il est possible d'émettre de fortes hypothèses sur les données et permettre d'établir une méthodologie efficace pour entraîner un algorithme selon l'aspect *PU*.

Pour le mécanisme d'étiquetage, la procédure de seuillage génère une étiquette binaire pour chaque pixel. Les étiquettes positives sont inspectées par un ou des experts. L'inspection visuelle permet de minimiser le taux de faux positif, ce qui permet d'accorder une grande confiance dans ces étiquettes positives. De plus, la procédure de seuillage *MiniFinder* permet d'annoter une majorité des événements calciques de forte intensité, mais ne permet pas d'étiqueter ceux de faible intensité. Ainsi, les pixels classifiés $y = 0$ possèdent une probabilité non nulle d'appartenir à la classe $y = 1$. C'est pourquoi un pixel étiqueté $y = 0$ par *MiniFinder* correspond à un pixel non étiqueté – le pixel n'est pas inspecté par un expert et a une probabilité non nulle d'appartenir à un mSCT. L'inspection visuelle des étiquettes fait ressortir deux caractéristiques : 1) la probabilité qu'un pixel appartienne à la classe $y = 0$ est beaucoup plus grande que la probabilité d'appartenir à la classe $y = 1$, $P(y = 0|x) \gg P(y = 1|x)$; et 2) la probabilité qu'un pixel x annoté dans la classe $y = 0$ par l'algorithme de seuillage provienne de la distribution f_+ est très faible, $P(y = 0|f_+(x)) \ll 1$. En d'autres termes, le ratio du nombre de pixels positifs non étiquetés relativement au nombre total des pixels non étiquetés est faible. Cela permet d'approximer avec confiance le facteur α comme le ratio du nombre de pixels annotés positivement et négativement par *MiniFinder*. La colonne **P** de la table 2.1 présente les résultats empiriques du ratio α pour chaque pli de validation ainsi que le jeu de donnée en test. Puisque ce ratio est très faible, les étiquettes qui ne sont pas positives peuvent être raisonnablement approximées comme des étiquettes négatives.

L'imagerie de neurones biologiques admet des caractéristiques bien spécifiques qui se distinguent des images naturelles et il est possible d'en tirer avantage. Effectivement, ce n'est pas tout le champ de vue imagé par le microscope qui contient des pixels informatifs. L'arrière-plan

contient peu ou pas d’informations pour la tâche à imager, dans ce cas présent des événements calciques (voir la zone exclue à la figure 2.10(a)). Ainsi, ce n’est pas bénéfique d’utiliser ces régions pour entraîner un algorithme d’apprentissage automatique. Par contre, dans l’avant-plan du vidéo, sont présents la majorité des mSCTs (98,4%, Mancini *et al.* (2018)). Le ratio α calculé sur l’avant-plan est encore faible (colonne P/FG dans la table 2.1), ce qui implique que les étiquettes qui ne sont pas positives dans l’avant-plan peuvent être raisonnablement approximées comme des étiquettes négatives. L’utilisation de l’avant-plan pour extraire des échantillons U plus informatifs permet en quelques sortes de choisir des exemples négatifs qui sont plus près de la distribution positive (fig. 2.7). Ces exemples raffinent la frontière de décision apprise.

En résumé, il est possible de stipuler les hypothèses suivantes pour l’apprentissage PU appliqué à l’imagerie d’événements calciques.

Hypothèse 1 *Apprentissage PU .*

- (1) *Les exemples positifs $y = 1$, obtenus par l’algorithme de seuillage, sont traités avec certitude comme positifs.*
- (2) *Les exemples non étiquetés de l’algorithme de seuillage sont définis comme des exemples négatifs $y = 0$.*
- (3) *Le gain d’information variable dans les exemples négatifs permet de raffiner le choix des exemples négatifs.*

La méthodologie décrite dans les prochaines sections est basée sur ces hypothèses.

TABLE 2.1 – Composition du jeu de données d’imagerie calcique Ca^{2+} . Le pourcentage de pixels d’arrière-plan (BG), pixels d’avant-plan (FG), pixels positifs (P) et le ratio P sur FG pixels, avec le nombre de mSCTs et neurones analysés sont décrits.

Splits	BG	FG	P	P/FG	mSCTs	Neurons
Fold A	89.93 %	10.07 %	0.03 %	0.26 %	4190	25
Fold B	90.65 %	9.35 %	0.03 %	0.32 %	4108	24
Fold C	87.53 %	12.47 %	0.02 %	0.15 %	3873	24
Test set	96.23 %	3.77 %	0.01 %	0.33 %	951	14

2.4 Extraction des échantillons positifs (P)

Selon l’hypothèse 1-1, les étiquettes positives provenant de *MiniFinder* sont considérées comme de vrais positifs. Considérant aussi que spatialement, les étiquettes positives sont toujours en agglomération (bas de la figure 2.8(a)). La stratégie employée consiste à regrouper les pixels positifs voisins pour obtenir un élément, identifié comme un événement et de les extraire. Par contre, pour donner plus de contexte à l’algorithme, une plus grande région est extraite.

Un pixel est considéré comme voisin s'il touche un côté ou un coin. La région d'intérêt est centrée sur l'événement en calculant la moyenne pondérée par l'intensité de fluorescence des pixels du masque (2.8(a)). Puis, à partir de ce centre, une région de 64×64 pixels est extraite (figure 2.8(b)). Le même processus est appliqué à la segmentation effectuée par MiniFinder pour extraire le masque d'étiquettes. La figure 2.8(c) présente des exemples finaux des paires d'échantillons (x, y) positives utilisées pour l'entraînement de l'algorithme de prédiction.

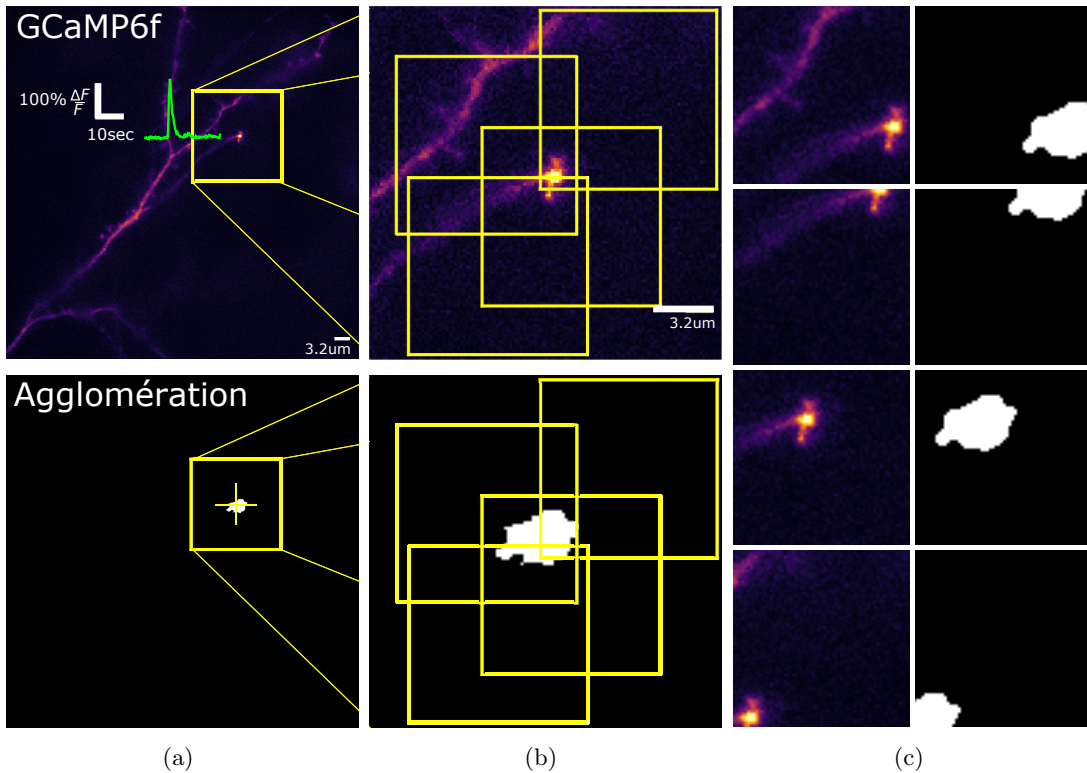


FIGURE 2.8 – Extraction des échantillons positifs (P). (a) La variation d'intensité de fluorescence (en vert) déclenche la procédure de seuillage pour générer l'agglomération d'étiquettes positives (les pixels blancs dans l'image du bas). Une région d'intérêt est découpée centrée sur l'agglomérat (b) À partir de la région d'intérêt extraite précédemment, des sous-régions aléatoires sont découpées dans le vidéo de fluorescence et dans le masque de segmentation (c) Exemples des entrées et sorties utilisées pour entraîner l'algorithme.

2.5 Extraction des échantillons non étiquetée (U)

Selon les hypothèses 1-2 et 1-3, les échantillons non étiquetés U sont extraits en prenant des régions au hasard sur l'avant-plan, c'est-à-dire sur le dendrite 2.10(a). Un seuil est calculé pour extraire l'avant-plan. La valeur du seuil peut être calculée automatiquement selon différentes caractéristiques de l'image basée sur : la moyenne, l'entropie (Li et Tam, 1998), les attributs des objets dans l'image (Huang et Wang, 1995), des méthodes de regroupement (Otsu, 1979), la forme de l'histogramme (Zack *et al.*, 1977). La figure 2.9 présente un exemple d'applica-

tion du seuillage calculé pour chaque méthode de seuillage. Les méthodes *Otsu* et *Triangle* se démarquent des autres puisqu'elles permettent d'extraire l'avant-plan sans empiéter sur l'arrière-plan. Comme le calcul du seuillage *Triangle* (3,76 ms) est légèrement plus rapide que celui d'*Otsu* (3,95 ms) sur une image, le seuillage *Triangle* est utilisé. La méthode du seuillage *Triangle* consiste à tracer l'histogramme d'intensité de l'image, tracer une droite entre les pics les plus éloignés et de déterminer la valeur d'intensité, perpendiculaire à cette droite qui maximise la distance de la droite à l'histogramme (voir la figure 2.9).

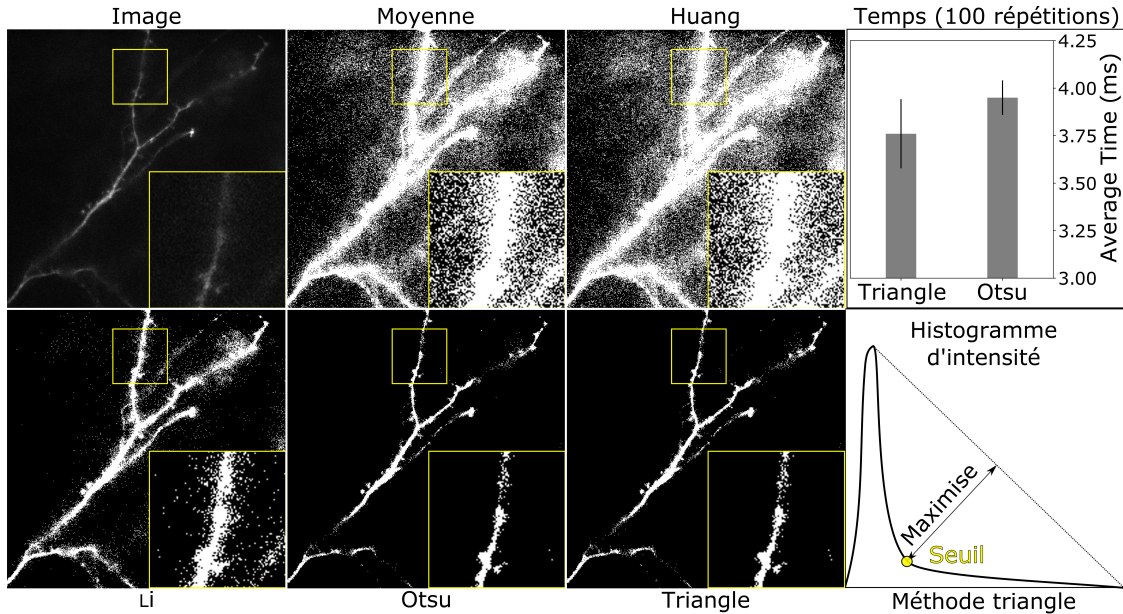


FIGURE 2.9 – Exemples des différentes méthodes de seuillage. Basée sur la moyenne de l'image, sur les attributs de l'image (*Huang*), sur l'entropie (*Li*), par méthode de regroupement (*Otsu*) et selon la forme de l'histogramme (*Triangle*). (*haut-droit*) Comparaison du temps de calcul de *Triangle* et *Otsu* pour 100 répétitions. (*bas-droit*) Méthode graphique du calcul du seuillage *Triangle*.

2.6 Algorithmes de segmentation

Deux approches de réseaux profonds à convolution ont été implémentées pour résoudre la tâche de segmentation des événements calciques. Le premier algorithme, U-Net 2D, est inspiré de Ronneberger *et al.* (2015) qui est reconnu dans le domaine de la segmentation biomédicale. La figure 2.11(a) présente le schéma de ce réseau. Le deuxième algorithme, U-Net 3D (Milletari *et al.*, 2016; Çiçek *et al.*, 2016) est l'extension du U-Net 2D sur des données tridimensionnelles. Pour les deux architectures testées, chaque étage de l'encodeur et du décodeur comporte deux couches de convolution, une normalisation par lot et une fonction de non-linéarité. Dans la partie encodeur, chaque étage diminue la résolution spatiale par un facteur deux et augmente le nombre de caractéristiques du même facteur. Le processus inverse est appliqué dans la partie décodeur de l'architecture. Les filtres de convolution utilisés comportent un noyau de

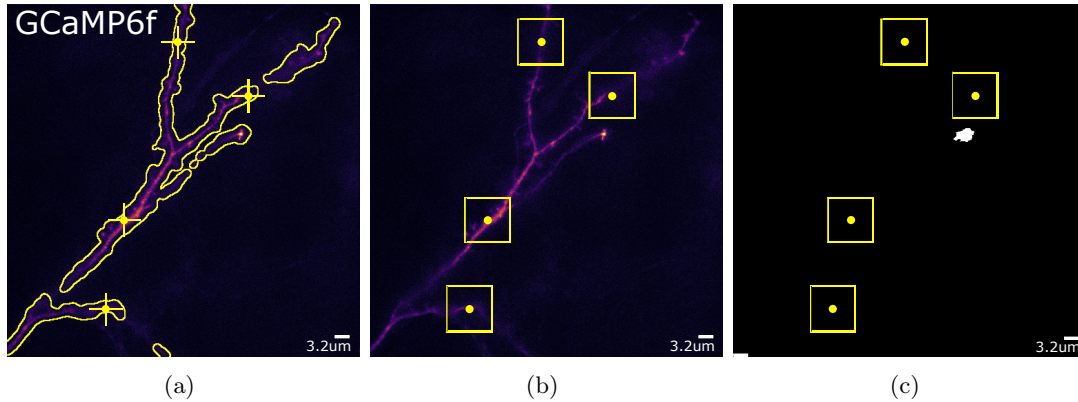


FIGURE 2.10 – Extraction des étiquettes non étiquetées (U). (a) Le seuillage triangle (Zack *et al.*, 1977) permet d’extraire la région dendritique (contour jaune). Dans ce masque, des pixels sont pigés aléatoirement (4 croix jaunes). (b) Centrées sur les pixels pigés, des régions d’intérêts sont découpées. (c) De la même façon, des régions d’intérêts sont découpées dans le masque de segmentation pour extraire les régions U.

dimension 3 dans chaque dimension de convolution. Contrairement à la publication originale du U-Net (Ronneberger *et al.*, 2015), comme suggéré par (Maas *et al.*, 2013; Xu *et al.*, 2015; Isola *et al.*, 2017) des unités de rectification linéaire *leaky* (*Leaky ReLU*) sont utilisées dans la section encodeur. La section décodeur du réseau utilise des unités de rectification linéaire (*ReLU*). Le *padding* des convolutions est ajusté afin que les dimensions de l’image soient conservées suite à la convolution. Tous les réseaux ont été implémentés avec la librairie PyTorch (Paszke *et al.*, 2019).

2.6.1 U-Net 2D

Le U-Net (Ronneberger *et al.*, 2015) originale a été proposé pour effectuer la segmentation dense et précise de structures biologiques. Son architecture extrait automatiquement des caractéristiques spatiales spécifiques à la structure biologique à segmenter. L’application du U-Net original n’est pas adaptée pour la segmentation des mSCTs. En effet, comme montré à la section 2.1, les caractéristiques spatiales ne permettent pas de différencier une augmentation d’intensité causée par un mSCT d’une zone où l’intensité est constamment élevée (ces zones sont présentes, entre autres, sur le dendrite, voire la fig. 2.3). De plus, la variation de fluorescence est nécessaire pour définir les temps où commence et termine un mSCT. Par exemple, la détection d’une augmentation drastique de la fluorescence, liée à un mSCT (exemple à la fig. 2.5(b)), n’est visible qu’en ayant accès au signal basal avant cette augmentation. C’est pourquoi l’intégration d’information temporelle dans le réseau est nécessaire pour une segmentation précise des mSCTs.

Pour ce faire, le U-Net est modifié afin d’accepter des canaux additionnels à son entrée. L’ar-

chitecture originale utilise les différents canaux en entrée pour fusionner différentes modalités spatiales de l’image. Par exemple, une image couleur comporte trois modalités : les signaux chromatiques rouge, vert et bleu (*Red, Green, Blue*, RGB). Chaque modalité chromatique représente un canal spécifique à l’entrée du U-Net, donc pour une image 32×32 , les canaux sont concaténés pour obtenir un tenseur \mathcal{T} de dimensions $3 \times 32 \times 32$. Chaque pixel de l’image comporte maintenant les trois caractéristiques chromatiques, lesquelles sont présentées au U-Net. Pour adapter l’entrée du U-Net, la profondeur des noyaux de convolution de la première couche est ajustée selon le nombre de canaux en entrées. Ainsi, les noyaux à convolution \mathcal{W} ont une dimension de $C \times 3 \times 3$, où C est le nombre de canaux de l’image ($C = 3$ avec une image RGB). Cependant, bien que le noyau \mathcal{W} comporte trois dimensions, l’opération convolutionnelle est effectuée uniquement sur les axes spatiaux de l’image (hauteur et largeur) :

$$\mathcal{F} = b + \sum_{k=0}^{C-1} \mathcal{W}_k \star \mathcal{T}_k \quad (2.2)$$

où k représente le $k^{\text{ième}}$ canal C , b le biais du neurone à convolution et \mathcal{F} les caractéristiques en sortie. Ainsi, l’information temporelle est injectée dans le U-Net en ajoutant des canaux à son entrée. Chaque canal comprend une image provenant de différents temps, tel que présenté à gauche dans la figure 2.11(a). Par exemple, avec cinq canaux en entrée, la stratégie consiste à utiliser des images suivant le patron $[t_{-2}, t_{-1}, t_0, t_1, t_2]$, dans lequel on s’intéresse à segmenter l’image située au temps t_0 . Deux acquisitions sont utilisées, de part et d’autre de l’image au temps t_0 pour créer un tenseur \mathcal{T} de dimensions $5 \times 32 \times 32$. La sortie prédite comporte un seul canal, associé au temps t_0 .

Le désavantage avec cette méthodologie est que le U-Net compresse l’information temporelle dès la première couche de convolution par le biais d’une somme pondérée, c.-à-d. la somme sur les canaux C dans l’équation 2.2. La puissance d’extraction de ce réseau est principalement limitée à des caractéristiques spatiales dues à la composition hiérarchique des filtres de convolutions spatiales (Goodfellow *et al.*, 2016). Ainsi, l’extraction de caractéristiques temporelles complexes n’est pas garant de la profondeur du U-Net, car les canaux temporels sont fusionnés à son entrée. Pour ces raisons, l’architecture adoptée comporte 128 filtres en entrée (contrairement à 64 suggéré par Ronneberger *et al.* (2015)) et 2 couches de profondeur (contrairement à 4 dans Ronneberger *et al.* (2015)). La grande quantité de filtres en entrée permet d’extraire beaucoup de caractéristiques sur les canaux temporels à l’entrée du U-Net.

2.6.2 U-Net 3D

Pour pallier le manque de puissance d’extraction de caractéristiques temporelles, la généralisation en trois dimensions du U-Net 2D est aussi analysée, référée par U-Net 3D (Çiçek *et al.*, 2016). La figure 2.11(b) présente un schéma de ce type de réseau. Les noyaux de convolutions comportent trois dimensions. Cette approche est particulièrement utilisée avec des données

volumétriques (Zhao *et al.*, 2020; Falk *et al.*, 2019). Dans le cas des données d'imagerie calcique, il y a les deux dimensions spatiales et une dimension temporelle ce qui forme un volume, voir en exemple le cube à gauche de la figure 2.11(b). Dans ce cas, la puissance d'extraction combine l'information spatiale et temporelle. Un désavantage du U-Net 3D est qu'il comporte irrémédiablement plus de paramètres à optimiser que le U-Net 2D, environ un ordre de grandeur plus de paramètres, ce qui augmente le temps d'entraînement à environ une demi-journée. L'architecture utilisée est la même que Çiçek *et al.* (2016) avec l'ajout de fonction d'activation *LeakyReLU* dans la section décodeur tel que suggéré par Isola *et al.* (2017).

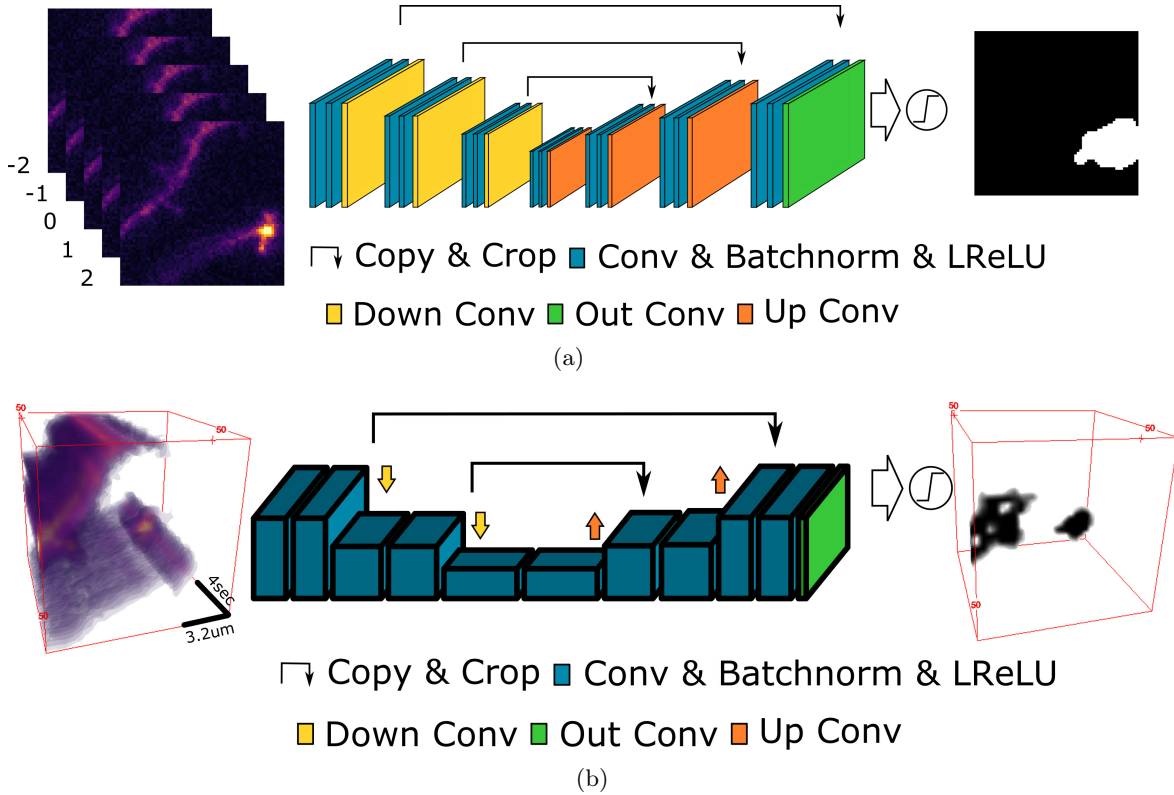


FIGURE 2.11 – Réseaux U-Net 2D et U-Net 3D. (a) Exemple de réseau UNet-2D avec en entrée une région $5 \times 32 \times 32$ comportant cinq modalités temporelles consécutives $[t_{-2}, t_{-1}, t_0, t_1, t_2]$ et en sortie un seul canal 32×32 pour le temps t_0 . (b) Exemple de réseau UNet-3D avec, en entrée et en sortie, une région $32 \times 32 \times 32$.

2.7 Prétraitement des vidéos

Dans le domaine de l'apprentissage machine, les données doivent être formatées pour retirer les biais présents qui complique la tâche de l'algorithme. Par exemple, comme les réseaux ont des fonctions d'activation qui opère autour de zéro (*ReLU, tanh, sigmoid*), il est considéré comme une bonne pratique de normaliser les données en entrée (Goodfellow *et al.*, 2016).

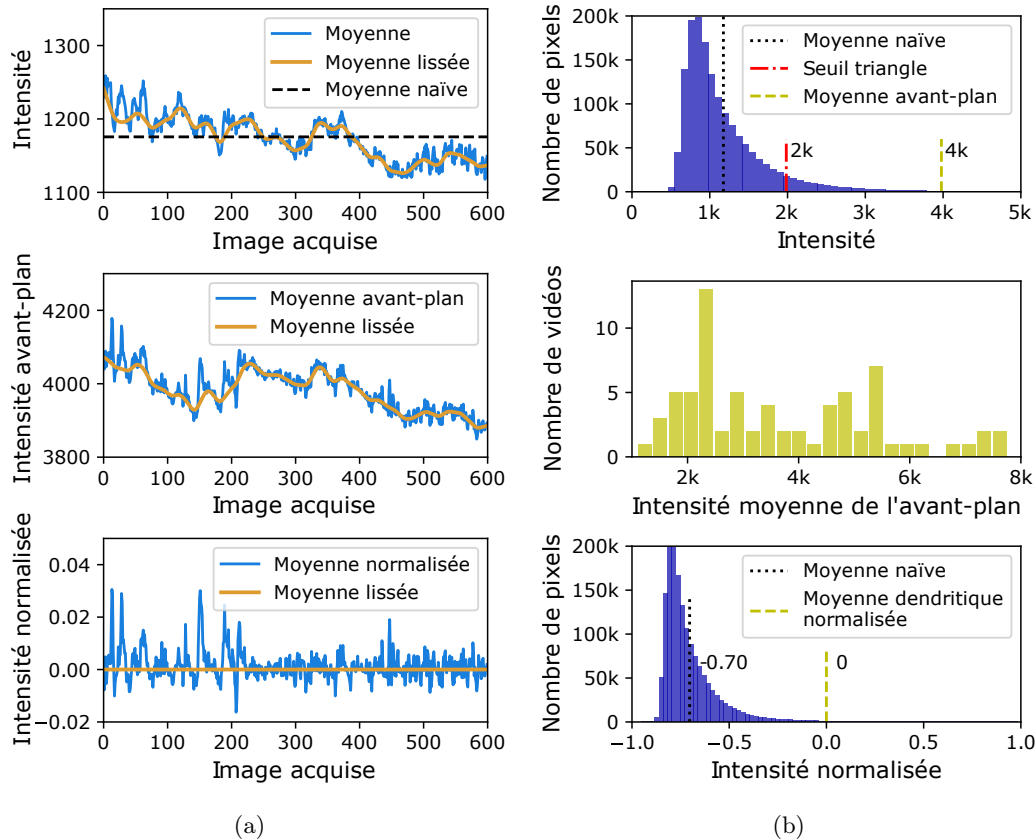


FIGURE 2.12 – Prétraitement des vidéos d’imagerie calcique. (a) Variation d’intensité pour un exemple de vidéo de 1min 40sec où la moyenne lissée (*orange*) est le signal lissé avec la méthode airPLS (Baek *et al.*, 2015) du signal bleu. De haut en bas, la fluorescence moyenne de tout le champ de vue, la fluorescence moyenne de l’avant-plan (p. ex. dans la zone jaune de la fig. 2.10(a)) et, finalement, l’intensité moyenne suite au prétraitement du vidéo. (b) de haut en bas; distribution d’intensité d’un vidéo original, la moyenne naïve (*noire*) est nettement plus basse que la moyenne de l’avant-plan (*jaune*) extraite après seuillage du vidéo avec le seuil triangle (*rouge*); la distribution de l’intensité moyenne de l’avant-plan de tous les neurones du jeu d’entraînement; finalement, la distribution de l’intensité suite au prétraitement du vidéo. L’intensité est comprise entre -1 et 1, avec une moyenne naïve à -0,70 (*noire*), par contre la moyenne de l’avant-plan, où la majorité des mSCTs ont lieu (Mancini *et al.*, 2018), est à 0.

Ceci étant dit, les vidéos d’imagerie calcique comportent de nombreuses caractéristiques qui doivent être prises en compte pour entraîner un réseau à convolution profond. Tout d’abord, des petits mouvements, de l’ordre de dizaines de pixels, sont présents lors de l’imagerie des dendrites et sont retirés en alignant le dendrite tout au long du vidéo (Guizar-Sicairos *et al.*, 2008).

Ensuite, l’intensité de fluorescence moyenne du champ de vue diminue après chaque acquisition du microscope. En effet, l’activation des indicateurs de fluorescence *GCaMP6*, présent

dans le neurone, induit une réaction photochimique qui dégrade leur capacité à émettre de la fluorescence – ce phénomène porte le nom de photoblanchiment (Greenbaum *et al.*, 2000). Le graphique du haut de la figure 2.12(a) montre la tendance descendante de la moyenne de l’intensité de fluorescence pour l’image entière (*ligne bleue*) dû au photoblanchiment. La majorité du signal provient de l’arrière-plan (la distribution de l’intensité est présentée à la figure 2.12(b)). Ainsi, normaliser les vidéos avec la moyenne du plan entier n’est pas adaptée, puisque la majorité des mSCTs sont situés sur l’avant-plan (Mancini *et al.*, 2018). La normalisation doit donc être basée sur la moyenne du signal de l’avant-plan, extraite avec la méthode de seuillage triangle (voir figure 2.9). À la figure 2.12(b), la moyenne de l’avant-plan est autour de 4000, contrairement à la moyenne naïve qui est d’environ 1200. Ainsi, la moyenne de l’avant-plan est calculée pour chaque acquisition (*ligne bleue* de la fig. 2.12(a)-*milieu*). Pour éviter d’éliminer des fluctuations rapides dues à des mSCTs, le signal est finalement lissé avec une méthode itérative (*ligne orange*) (Baek *et al.*, 2015). La moyenne de l’avant-plan lissée est utilisée pour normaliser le vidéo (*ligne bleue* à la fig. 2.12(a)-*bas*). La distribution de l’intensité normalisée (fig :method :pp :hist-*bas*) montre que l’intensité du vidéo, dans la zone d’avant-plan, est bien normalisée à zéro. Cette normalisation est effectuée pour chaque vidéo puisque la moyenne de l’avant-plan est très variable (de 1000 à 8000, fig. 2.12(b)-*milieu*) d’un vidéo à l’autre. D’autant plus que le photoblanchiment n’est pas constant entre les vidéos.

2.8 Détails de l’entraînement des algorithmes

Le jeu de donnée de microscopie d’événements miniatures comprend 87 neurones biologiques. Les prochaines sous-sections rapportent les détails relatifs au processus d’entraînement, de validation et d’évaluation.

2.8.1 Séparation des données

Parmi les 87 neurones, 14 neurones jamais vus en entraînement composent le jeu de données de test utilisé pour caractériser précisément les performances des algorithmes.

Étant donnée la variance biologique présente dans les données et afin de ne pas sur-apprendre sur les données sélectionnées pour le jeu de validation, la validation croisée est utilisée. Le test de validation croisée estime l’efficacité générale d’un algorithme, assurant un équilibre entre le biais et la variance (Kohavi, 1995). Compte tenu du nombre de neurones et pour limiter le nombre d’entraînements nécessaire, un nombre de trois plis est utilisé. Les 73 neurones restants sont utilisés pour créer les trois plis. Pour la création des plis, les vidéos ont été séparés afin d’avoir, dans la mesure du possible, le même nombre de mSCTs et le même nombre de neurones. Ainsi, pour les trois entraînements, 2 plis sont sélectionnés et utilisés pour entraîner le réseau, et le troisième pli est utilisé comme validation. Les résultats rapportés sont une moyenne des algorithmes entraînés indépendamment sur les trois plis.

La table 2.1 présente les statistiques du jeu de données pour chaque pli et le jeu de données de test.

2.8.2 Paramètres d’entraînement

Pour le U-Net 2D, l’entraînement comprend au total 80k itérations, où chaque itération représente une inférence dans le réseau (*forward pass*) pour un lot de \mathfrak{B} images (*batch size*). Le nombre d’itérations est augmenté à 160k itérations pour le modèle U-Net 3D puisqu’il comporte plus de paramètres à optimiser. La taille du lot d’image est de 128 et 16 pour le U-Net 2D et 3D respectivement. Tous les entraînements utilisent l’optimiser Adam (Kingma et Ba, 2017) avec un taux d’apprentissage initial de $\alpha = 2 \times 10^{-4}$ et les facteurs de décroissance exponentielle $\beta_1 = 0.9$ et $\beta_2 = 0.999$. Pendant l’entraînement, des transformations sont appliquées à la volée : des retournements verticaux et horizontaux et des découpes aléatoires. Les échantillons P et U extraits (64×64) sont précisément le double de la dimension de l’image à découper aléatoirement (32×32). Cela permet d’augmenter artificiellement le jeu de données pour éviter le surapprentissage (Goodfellow *et al.*, 2016). L’objectif de l’entraînement consiste à minimiser l’erreur quadratique moyenne (*MSE : Mean Square Error*) entre les pixels en sortie du réseau et ceux provenant de la procédure de seuillage. Une étape de validation est effectuée suite à chaque 100 itérations pour déterminer les performances du réseau sur le pli de validation. Pour avoir le réseau avec la meilleure généralisation, celui avec la plus petite erreur en validation pendant la procédure d’entraînement est conservé et est évalué avec le jeu de données en test.

2.8.3 Validation

La procédure de validation utilise les données du pli de validation. Tous les échantillons P du pli sont extraits, comme présenté à la section 2.4. Ces échantillons sont inférés dans le réseau, sans faire l’optimisation de ses poids, et la fonction de perte MSE entre la prédiction du réseau et la procédure de seuillage est calculée. Pour ne pas pénaliser les bonnes prédictions du réseau qui ne sont pas détectées par la procédure de seuillage, la validation n’est pas effectuée sur tout le vidéo. La validation est plutôt calculée uniquement sur les échantillons P. Cela permet d’avoir une approximation de la performance du modèle basé uniquement sur les mSCTs bien étiquetés par la procédure de seuillage. Il n’y a aucune stochasticité dans la procédure de validation, tous les échantillons P sont utilisés à chaque étape de validation.

2.8.4 Choix du seuil en sortie

Suite à la procédure d’entraînement, le réseau produit une pseudo-probabilité pour chaque pixel inféré. Pour obtenir une segmentation, il est nécessaire de déterminer le meilleur seuil à appliquer en sortie pour classifier le pixel $y \in \{0, 1\}$. Le choix du seuil à utiliser n’est pas trivial, ceci est d’autant plus vrai dans un contexte d’apprentissage *PU*. Il est difficile d’optimiser le

seuil en sortie tout en se comparant avec le masque obtenu par la procédure de seuillage. Du fait que des régions peuvent avoir été détectées par le réseau entraîné, mais n’ont pas été détectées dans la *vérité terrain* (et c’est ce qui se produit, voir la section 3.6). C’est pourquoi le choix du seuil est déterminé avec les échantillons P uniquement. Comme il y a un débalancement entre la classe positive et négative, Saito et Rehmsmeier (2015) recommande d’utiliser la courbe de précision-rappel pour obtenir une meilleure approximation des performances de l’algorithme. La courbe est calculée en utilisant tous les échantillons P et le seuil ayant le meilleur compromis rappel et précision est sélectionné. Cela correspond au seuil le plus près d’une précision et d’un rappel parfait dans le graphique de la courbe précision-rappel. Avec le seuil obtenu dans la section précédente, les pixels en sortie du réseau sont classifiés selon la règle suivante :

$$y \in \begin{cases} 1 & \text{si } g(x) > \text{seuil} \\ 0 & \text{sinon.} \end{cases} \quad (2.3)$$

2.9 Post-traitement des prédictions

Une procédure d’identification et de filtration est appliquée pour extraire les événements calciques d’une segmentation générée par le réseau. L’algorithme 1-*Extraction des événements* explique en détail les étapes appliquées. Tout d’abord, les pixels voisins sont agglomérés en instances pour représenter les événements calciques. Ensuite, les événements extraits sont filtrés selon leurs dimensions temporelles, spatiales et leurs intensités normalisées. Un événement est éliminé si l’une de ces caractéristiques n’est pas respectée : 1) une hauteur plus petite que trois pixels 2) une largeur plus petite que trois pixels 3) une durée de moins de trois images 4) une fluorescence normalisée inférieure à 0,12 (obtenu en maximisant le compromis de vrais positifs et de faux positifs). Dû à leurs petites dimensions, ces événements sont retirés puisqu’ils ne sont pas intéressants à considérer pour l’analyse biologique subséquente. Finalement, l’ensemble contenant tous les événements calciques identifiés est retourné.

Par la suite, des métriques relatives à la morphologie des mSCTs, à la récurrence des mSCTs dans un vidéo et à leur intensité de fluorescence sont extraites et sauvegardées pour permettre de faire l’analyse quantitative de ces données. Les outils développés à cette fin sont expliqués en détail dans le chapitre 4.

2.10 Évaluation des performances

Cette section présente les différentes méthodologies utilisées pour analyser les performances des réseaux entraînés. Dû à l’incertitude reliée à la *vérité terrain* obtenue par la procédure de seuillage (sec. 2.2.1), plusieurs étapes d’évaluations sont nécessaires, autant qualitatives que quantitatives. Plus précisément, cette section propose de bien comprendre les métriques qui sont utilisées et, plus particulièrement, leurs limitations.

Algorithme 1 Extraction des événements

- 1: **Soit** : T, H, W le temps, la hauteur et la largeur du vidéo
 - 2: **Soit** : $p_{t,h,w}$ un pixel dénoté par la coordonnée de l'image, la hauteur et la largeur dans le vidéo
 - 3: **Soit** : $(T_{min}, H_{min}, W_{min})$ les dimensions minimales d'un événement
 - 4: **Soit** : $\frac{\Delta F}{F}_{min}$ la fluorescence normalisée minimale d'un événement
 - 5: **Argument** : Un vidéo binaire $\Omega \in \{0, 1\}^T \times \{0, 1\}^H \times \{0, 1\}^W$
 - 6: **Initialisation** : L'ensemble des événements calcique $\mathcal{C} = \{\}$
 - 7: **while** $\exists p_{t,h,w} \mid \Omega[t, h, w] = 1$ **do**
 - 8: Choisir aléatoirement un pixel avec une valeur de 1 : $p' \sim \mathcal{U}\{p_{t,h,w} \mid \Omega[t, h, w] = 1\}$
 - 9: Identifier l'événement : $\mathcal{E} = \text{trouverVoisins}(p')$
 - 10: **if** $\text{dimensions}(\mathcal{E}) < (T_{min}, H_{min}, W_{min})$ **or**
 $\text{calculDeltaF}(\mathcal{E}) < \frac{\Delta F}{F}_{min}$ **then**
 - 11: Effacer l'événement : $\Omega[\mathcal{E}] = 0$
 - 12: **else**
 - 13: Ajouter l'événement : $\mathcal{C}.\text{append}(\mathcal{E})$
 - 14: **end if**
 - 15: **end while**
 - 16: **return** \mathcal{C} l'ensemble d'événements calciques
-

2.10.1 Description des métriques

Les éléments de bases pour les familles de métriques de recoupement sont les cardinalités de la matrice de confusion : les vrais positifs (VP), les vrais négatifs (VN), les faux positifs (FP) et les faux négatifs (FN) (Taha et Hanbury, 2015). De ces éléments, découlent directement la précision et le rappel :

$$\text{Précision} = \frac{VP}{VP + FP} = \frac{|\hat{y} \cap y|}{|\hat{y}|} \quad (2.4)$$

$$\text{Rappel} = \frac{VP}{VP + FN} = \frac{|\hat{y} \cap y|}{|y|} \quad (2.5)$$

La prédiction et la vérité terrain sont représentées par \hat{y} et y respectivement. La précision peut être interprétée comme une mesure de la qualité des prédictions, alors que le rappel mesure la quantité des prédictions parmi les éléments positifs. Afin d'avoir un portrait global des performances, la précision et le rappel sont combinés en une seule métrique : le coefficient de Dice (Dice, 1945). Celui-ci est obtenu en calculant la moyenne harmonique de la précision et du rappel. Le coefficient de Dice, aussi appelé *F-score* ou *F1-score*, s'exprime donc selon ces trois formulations mathématiques équivalentes.

$$\text{Dice} = 2 \cdot \frac{\text{Précision} \cdot \text{Rappel}}{\text{Précision} + \text{Rappel}} = \frac{2 \times |\hat{y} \cap y|}{|\hat{y}| + |y|} = \frac{2 \times VP}{2 \times VP + FP + FN} \quad (2.6)$$

Une autre métrique communément utilisée est le coefficient de Jaccard, aussi connu sous le nom *Intersection over Union (IoU)*, pour quantifier la similarité de recoupement entre

deux ensembles (Jaccard, 1901). Ce coefficient représente le ratio de l'intersection sur l'union des deux ensembles. On le retrouve, entre autres, sous ces trois formulations mathématiques équivalentes.

$$\mathbf{Jaccard} = \frac{\text{Dice}}{2 - \text{Dice}} = \frac{|\hat{y} \cap y|}{|\hat{y} \cup y|} = \frac{VP}{VP + FP + FN} \quad (2.7)$$

Il est évident que cette métrique est très semblable au coefficient de Dice. En effet, en comparant les membres de droite des équations 2.6 et 2.7, la seule différence est le coefficient 2 devant le nombre de vrais positifs dans le cas du Dice. Cela implique que le coefficient de Jaccard accorde plus d'importance aux faux négatifs et faux positifs par rapport au nombre de vrais positifs. Ainsi, sa valeur est souvent plus faible comparativement au coefficient de Dice pour une situation donnée. Cela étant dit, il n'y a pas un véritable gain en information de rapporter le coefficient de Jaccard et de Dice. Les résultats de ce mémoire sont décrits selon leur **précision**, **rappel** et coefficient de **Dice**.

Interprétation des métriques de segmentation

De faibles valeurs des métriques de segmentations sont observées dans la de petites instances (Yeghiazaryan et Voiculescu, 2018), par exemple : des véhicules dans un grand champ de vue (Kampffmeyer *et al.*, 2016), des organes dans une numérisation CT (Linguraru *et al.*, 2012), des anomalies rétinal dans de l'imagerie OCT (Seeböck *et al.*, 2020). Dans ce cas, une petite erreur sur le périmètre de la segmentation ajoute une grande quantité de faux positifs

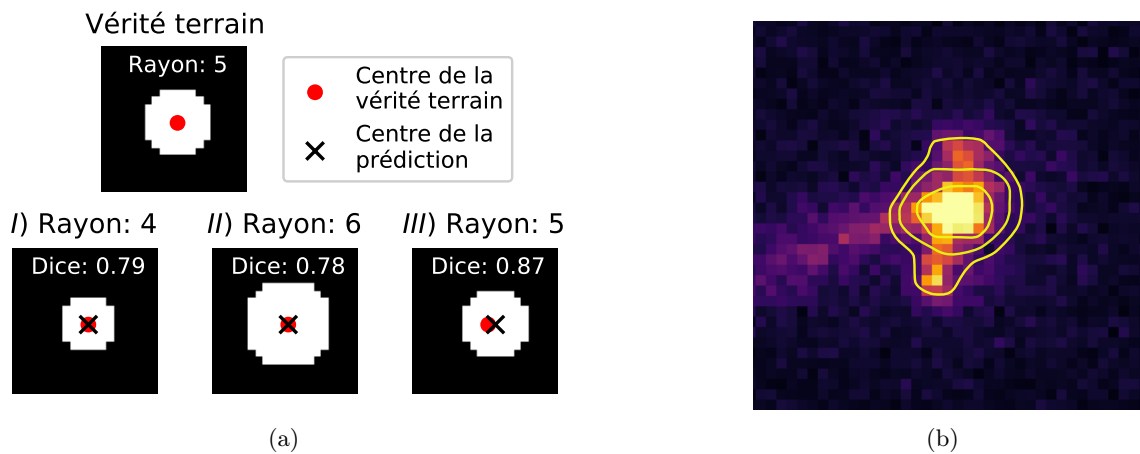


FIGURE 2.13 – Variation du coefficient de segmentation *Dice* pour de petites instances. (a) Illustration de trois cas : I) un pixel de moins sur le périmètre de segmentation II) un pixel supplémentaire sur le périmètre III) le masque est exactement le même, mais décalé d'un pixel vers la droite. (b) Trois bordures de segmentation valides pour illustrer la variabilité inhérente associée à la définition subjective de la bordure de segmentation d'un événement calcique.

(ou faux négatifs). Comme le nombre de vrais positifs est petit, la précision est grandement pénalisée, pénalisant ainsi le coefficient de Dice. Dans le cas d'une tâche de segmentation de petite taille, donc avec un nombre de pixels vrais positifs, une petite erreur sur la bordure de la prédiction affecte énormément la performance rapportée par le coefficient de *Dice* et de *Jaccard*. La figure 2.13(a) présente trois exemples typiques illustrant cette limitation. Un autre problème qui doit être pris en compte dans le contexte présent est que la vérité terrain n'est pas bien définie, ce qui influence grandement les mesures prises avec ces différentes métriques (fig. 2.13(b)). C'est pourquoi les valeurs de coefficient de *Dice* élevés sont difficilement atteignables. Par exemple, la figure 3.2(c) présente la valeur du Dice obtenue en comparant deux segmentations d'un expert pour un même événement calcique.

2.10.2 Analyse intra- et inter-experts

L'analyse de la variabilité intra- et inter-experts permet de quantifier empiriquement les valeurs de coefficient de *Dice* maximale en segmentation pouvant être obtenue. La variabilité intra-experts est la variabilité présente lorsqu'un expert humain annote deux fois la même image. La variabilité inter-experts, quant à elle, permet de quantifier la variabilité présente entre deux experts humains. En suivant la méthodologie proposée par Udupa *et al.* (2006), un ensemble d'images à segmenter sont extraites du jeu de données de test. Les images à segmenter ont été choisies pour avoir une représentation adéquate de tous les types d'événements calciques, la liste complète se retrouve dans l'annexe A.1. Afin de faciliter la tâche de segmentation, seulement l'image au maximum d'intensité de fluorescence est annotée, cela est suffisant pour présenter la variabilité de segmentation présente dans la tâche. Un ensemble de 78 images est extrait. Chacune de ces images est annotée deux fois par chacun des deux experts, pour un total de quatre segmentations par image. Des rotations aléatoires de 0, 90, 180 ou 270 degrés sont appliquées avant chaque segmentation. Le processus d'annotation a été effectué en deux étapes à une semaine d'intervalle. À chaque étape, toutes les images sont annotées par chacun des experts. Les segmentations du même expert (à une semaine d'intervalle) permettent de quantifier la variabilité intra-experts. La variation inter-experts, est quant à elle calculée entre les annotations des deux experts.

2.10.3 Évaluation en détection

L'évaluation en détection permet de quantifier les performances de l'algorithme sans tenir compte de la qualité de segmentation. Une détection est donc toujours définie par le centroïde de la segmentation de cet événement.

Les performances en détection sont d'abord évaluées par rapport aux détections provenant de la procédure de seuillage. Un nombre important de détections additionnelles, inférées par les réseaux, ne concordaient pas avec la procédure de seuillage. Pour confirmer ou infirmer la validité des faux positifs prédits par le réseau, les événements sont réannotés par un expert.

Avec la validation croisée et les différents hyperparamètres testés de l’approche PU, 57 réseaux ont été évalués sur les 28 vidéos du jeu de données de test. Afin de simplifier la procédure de réannotation, une application d’annotation a été développée (décrite à la section 4.2), ce qui permet à l’expert d’inspecter tous les événements et de les ajouter dans l’ensemble des événements positifs. Si une détection est déjà dans l’ensemble, l’expert n’a pas à la réannoter. Cela permet d’itérativement créer un ensemble de nouveaux événements non détectés par la procédure de seuillage et de diminuer la charge d’annotation de l’expert.

Les résultats en détection par rapport à cet ensemble sont ensuite évalués pour tous les réseaux entraînés. Les métriques utilisées pour évaluer la détection sont les mêmes que celles en segmentation : la précision (éq. 2.4), le rappel (éq. 2.5) et le *Dice* (éq. 2.6). Elles sont calculées avec la définition de ces métriques selon les VP, FP et FN. Un VP a lieu lorsque deux événements sont associés entre ceux prédits et ceux de validation. Une association entre événements a lieu lorsque la distance de leur centroïde est inférieure à 6 pixels. L’association utilise l’algorithme d’assignation Hungarian (Kuhn, 1955) qui impose des associations uniques entre événements. Si un mSCT prédit n’est pas associé avec les mSCTs de validation, il s’agit d’un FP. De même, si un mSCT de validation n’est pas associé à un mSCT prédit, il s’agit d’un FN.

Cette étape de réannotation est cruciale dans un contexte d’apprentissage PU. Cela vient du fait que la population d’étiquettes positives sélectionnée par le mécanisme d’étiquetage contient rarement toutes les étiquettes positives (p. ex. la figure 2.7). Ainsi, si l’algorithme généralise bien sur la distribution de données positives, des prédictions positives, non sélectionnées par le mécanisme d’étiquetage sont possibles. Celles-ci doivent être évaluées par un expert externe.

Chapitre 3

Analyses et comparaisons

La méthodologie décrite dans le chapitre précédent est maintenant caractérisée et validée. Tout d’abord, la capacité de segmentation des réseaux de neurones est analysée selon *MiniFinder*, *SICT* et un jeu de données annotées précisément par deux experts. Les performances en détection de chaque approche sont ensuite analysées. Finalement, l’effet des différents paramètres de l’entraînement PU sont analysé empiriquement sur la tâche de détection des événements calciques.

3.1 Analyse U-Net 2D et U-Net 3D en segmentation

Dans cette section, les performances en segmentation sont mises de l’avant pour les deux approches d’apprentissage profond utilisées pour segmenter les événements calciques : le U-Net 2D et le U-Net 3D. Ces deux approches sont tout d’abord comparées à la procédure par seuillage *MiniFinder* qui est utilisée pour générer les étiquettes d’entraînement. Une analyse intra- et inter-expert avec un jeu de données segmenté précisément par deux experts permet de quantifier le domaine concevable du coefficient de *Dice* en segmentation pour cette tâche. Finalement, *MiniFinder*, le U-Net 2D et 3D et *SICT* sont tous comparés aux segmentations précises des experts.

3.1.1 Comparaison avec *MiniFinder*

La figure 3.1(a) montre la distribution du coefficient *Dice* en segmentation pour les événements associés avec *MiniFinder* (les vrais positifs). La distribution a une médiane faible et comporte une queue allongée et large vers des petites valeurs. Il y a beaucoup de petites valeurs du coefficient *Dice* (pour le U-Net 2D, 25 % sont en dessous de 0,5). À première vue, il semble que la segmentation soit mauvaise. Or, après inspection visuelle, les masques des U-Net 2D et U-Net 3D semblent qualitativement meilleurs que *MiniFinder* (fig. 3.1(b)). Par exemple, la première rangée de la figure 3.1(b) montre un faible coefficient de *Dice*. C’est difficile à

voir, mais la segmentation des U-Nets délimite mieux le mSCT que celle de *MiniFinder*. Une analyse avec des annotations manuelles d’experts est faite ultérieurement à la section 3.1.3.

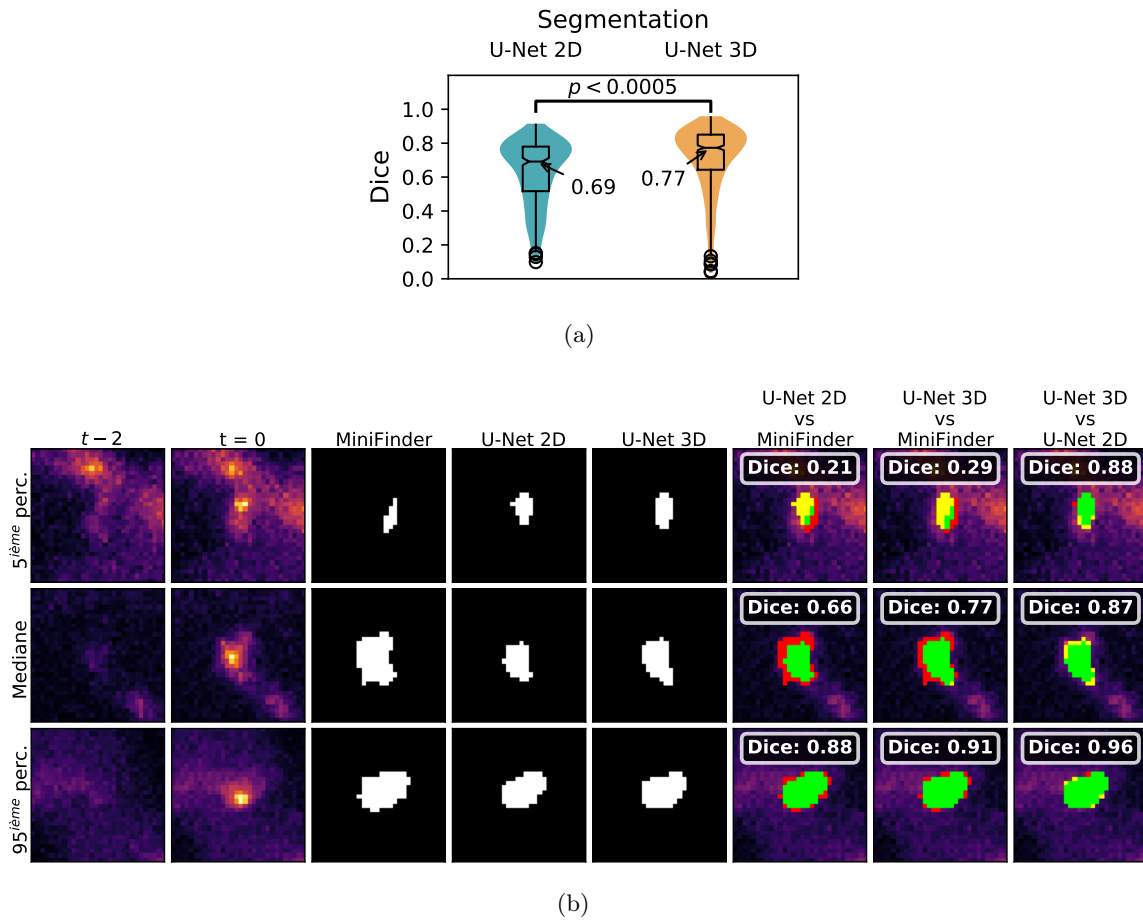


FIGURE 3.1 – Comparaison en segmentation des approches U-Net 2D et U-Net 3D avec *MiniFinder* pour tous les mSCTs du jeu de test. (a) Distribution du coefficient de *Dice* en segmentation avec la médiane. Le U-Net 3D génère des segmentations significativement plus fiables que le U-Net 2D par rapport à *MiniFinder*. (b) Exemples de segmentation dans le 5^e percentile, médiane et 95^e percentile. La colonne segmentée correspond au temps $t = 0$. La colonne $t - 2$ correspond à la même région, mais 2 pas d’acquisition auparavant (300ms avant). Le code de couleur vert, jaune, rouge correspond aux VP, FP, FN respectivement.

De plus, comme mentionné à la section 2.10.1, la petite taille des masques des mSCTs explique en partie les faibles performances. La rangée de milieu de la figure 3.1(b) est un bon exemple d’un coefficient *Dice* de 0,66 qui résulte d’une différence de quelques pixels en bordure seulement.

La distribution du coefficient *Dice* est significativement meilleure pour le U-Net 3D lorsqu’on compare ses segmentations avec *MiniFinder*. On remarque aussi à la dernière colonne de la figure 3.1(b) que le U-Net 2D et 3D ont des segmentations similaires malgré leur différence

avec *MiniFinder*. Par exemple, dans la première rangée de la figure 3.1(b), les deux U-Nets ont une segmentation différente de *MiniFinder* (*Dice* de 0,21 et 0,29), mais la comparaison entre la segmentation des U-Nets montre qu'ils sont similaires avec un *Dice* de 0,88. Une explication possible est que le U-Net 2D et 3D ont tous les deux appris des caractéristiques générales semblables pour segmenter les mSCTs.

3.1.2 Analyse intra- et inter-experts

Les performances en segmentation sont difficilement interprétables : en présence des petites instances à segmenter, une erreur de quelques pixels est très pénalisante, l'instance à segmenter est difficile à définir à une précision d'un pixel (sec. 2.10.1). La figure 3.1(b) expose quelques exemples concrets avec des événements calciques. Pour avoir une meilleure compréhension de la distribution du coefficient de *Dice* en segmentation et tenter de quantifier la qualité des masques de segmentation pour différents mSCTs, l'analyse suivante propose d'évaluer la variabilité intra-expert ainsi que la variabilité inter-expert (pour plus de détails sur la méthodologie, voir la section 2.10.2).

Varibilité intra-expert

La variabilité intra-expert permet de quantifier la variabilité présente dans la tâche de segmentation pour un même expert. La quantification de cette variabilité permet d'évaluer le consensus entre les experts sur la forme exacte d'un événement calcique. C'est excellent si un algorithme performe aussi bien ou mieux que le consensus d'experts, car l'erreur provient de la difficulté de la définition et de l'interprétation de la tâche à automatiser. Les mêmes événements calciques ont été présentés à chaque expert avec une semaine d'interval. Avec un expert *parfait* et une segmentation sans ambiguïté (aucune variabilité), le coefficient de *Dice* attendu est un score parfait, soit 1. Par contre, le coefficient de *Dice* médian est de 0,78 et 0,77 pour les deux experts (figure 3.2(a)). Des exemples de segmentation sont présentés dans la portion gauche de la figure 3.2(c). Les valeurs de coefficient de *Dice* obtenues représentent approximativement une borne supérieure difficilement atteignable par l'algorithme de segmentation, particulièrement lorsque celui n'est pas entraîné avec des annotations d'expert. Par exemple, les masques de *MiniFinder*, utilisées pour l'entraînement, ont un coefficient de *Dice* moyen de 0,52 et 0,56 par rapport aux deux experts respectivement, mais le U-Net 3D a un *Dice* moyen de 0,57 et 0,64 (fig. 3.3(a) et 3.3(b)). La segmentation du U-Net 3D est donc meilleure que *MiniFinder* selon des experts, mais inférieure à la variabilité présente dans ces experts.

Variabilité inter-expert

La variabilité inter-expert permet de quantifier la variabilité présente entre deux experts pour une même tâche de segmentation. La variabilité inter-expert est plus réaliste pour déterminer une borne supérieure à atteindre. Dans Giovannucci *et al.* (2019), la variabilité inter-expert

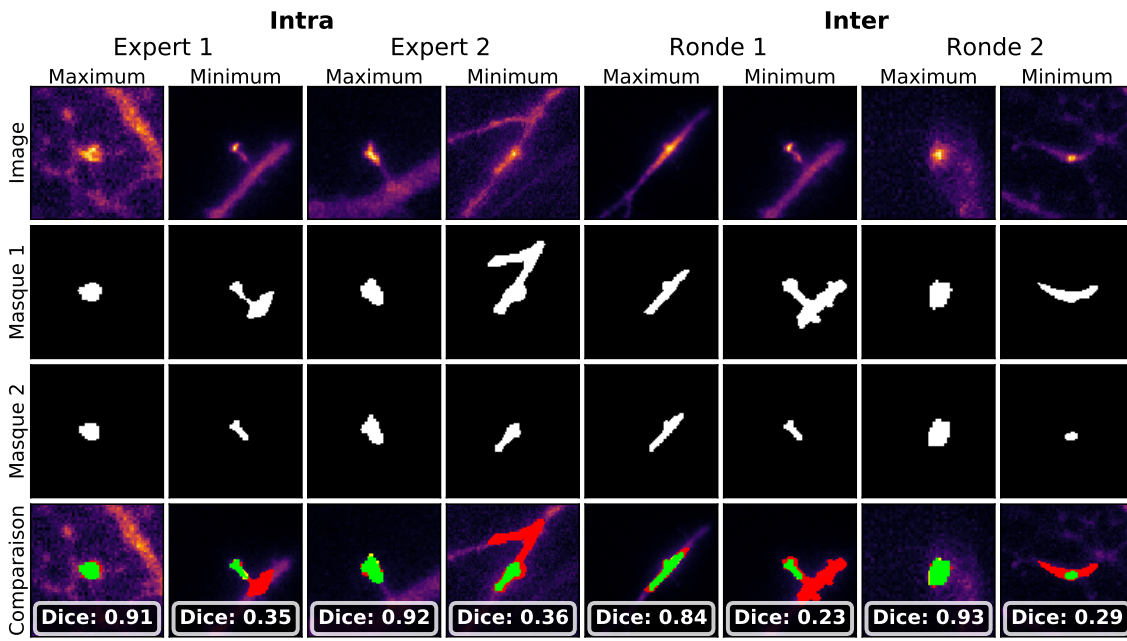
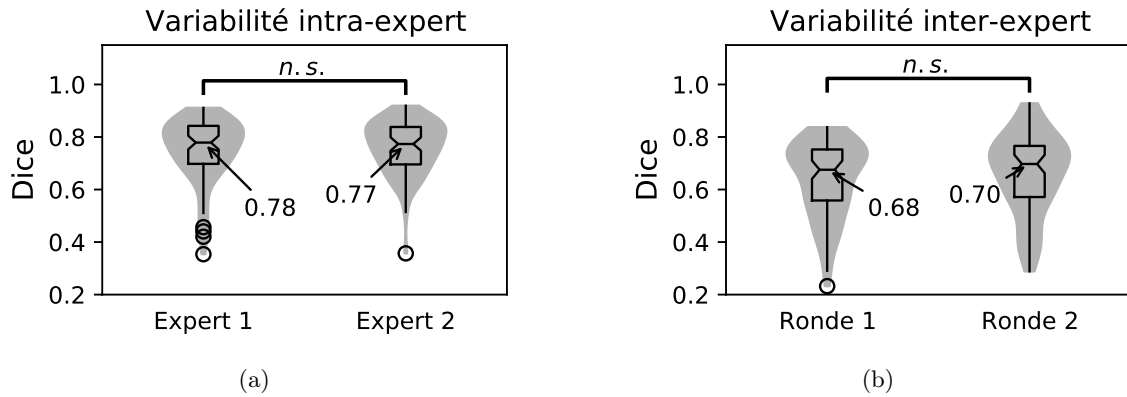


FIGURE 3.2 – Analyse intra- et inter-expert. Les experts ont accès à 10 images avant et après pour faire la segmentation. (a) Distribution du *Dice* entre deux segmentations d’un même expert, soit la variabilité intra-expert. Les variabilités intra-expert ne sont pas significativement différentes entre les deux experts testés. (b) Distribution du *Dice* entre les masques de segmentation de deux experts, soit la variabilité inter-expert (non significative également). Par contre, les valeurs de *Dice* intra-expert et inter-expert sont significativement différentes entre elles, avec un $p < 0.0005$. (c) Exemples du coefficient de *Dice* maximum et minimum pour chaque graphique en (a) et (b). Le code de couleur vert, jaune, rouge correspond aux VP, FP, FN respectivement.

est utilisée pour quantifier le niveau de désaccord entre les experts afin de mieux quantifier leurs performances. La variabilité inter-expert est évidemment plus grande que la variabilité intra-expert, donc une valeur de *Dice* plus faible. En effet, le *Dice* médian, entre les deux

experts, est de 0,68 et 0.70 pour les deux rondes d’annotations respectives (figure 3.2(b)). La moitié droite de la figure 3.2(c) présente les segmentations avec le coefficient de *Dice* maximum et minimum pour les deux rondes d’annotation pour chaque expert.

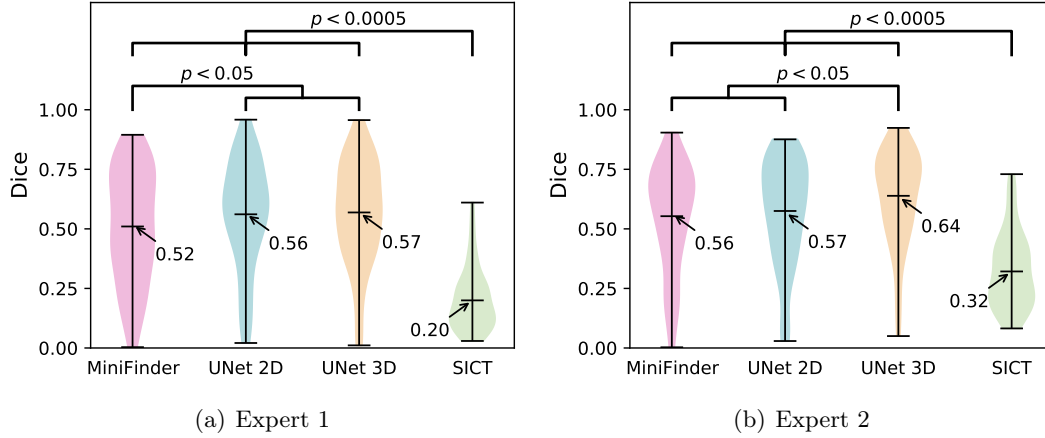


FIGURE 3.3 – Comparaison en segmentation des approches avec les annotations d’experts. Les annotations précises de deux experts sont présentées à la figure 3.2(c). (a) et (b) représente la distribution du coefficient de *Dice* pour toutes les segmentations pour *MiniFinder*, U-Net 2D, U-Net 3D et *SICT*. Chaque graphique représente les résultats pour un seul expert. La segmentation de *SICT* est significativement inférieure aux trois autres approches, selon les deux experts. Selon l’expert 1, les U-Net 2D et 3D sont significativement meilleurs que *MiniFinder*, mais le U-Net 3D n’est pas supérieur. Selon l’expert 2, le U-Net 3D est significativement meilleur que *MiniFinder* et le U-Net 2D.

3.1.3 Comparaison avec annotation précise

La segmentation précise est celle utilisée pour quantifier la variabilité des experts. Le jeu de données, annoté à une précision au niveau du pixel, comporte 78 masques d’événement calcique ayant été segmenté au niveau maximal d’intensité (l’image où l’intensité normalisée $\Delta F/F_0$ est maximale sur la trace de ce mSCT). L’utilisation du jeu de données précis permet d’avoir une référence pour comparer toutes les approches équitablement. En effet, contrairement à *SICT*, les réseaux U-Net 2D et 3D obtiennent de meilleurs résultats en segmentation lorsqu’ils sont comparés à *MiniFinder* puisqu’ils ont été entraînés avec les masques de *MiniFinder*. Les figures 3.3(a) et 3.3(b) présente la distribution du coefficient *Dice* pour les différentes approches proposées pour segmenter automatiquement ou semi automatiquement les mSCTs. Tout d’abord, pour les deux experts, *MiniFinder*, U-Net 2D et U-Net 3D sont toujours statistiquement supérieurs à *SICT*. La procédure développée à l’interne performe mieux que la référence. De plus, pour les deux experts également, le U-Net 3D est aussi statistiquement supérieur que *MiniFinder*. En utilisant les masques de *MiniFinder* comme jeu d’entraînement, le réseau apprend à faire des segmentations morphologiquement meilleures que *MiniFinder*. Ce résultat renforce l’hypothèse que le réseau apprend des caractéristiques plus adaptées à la

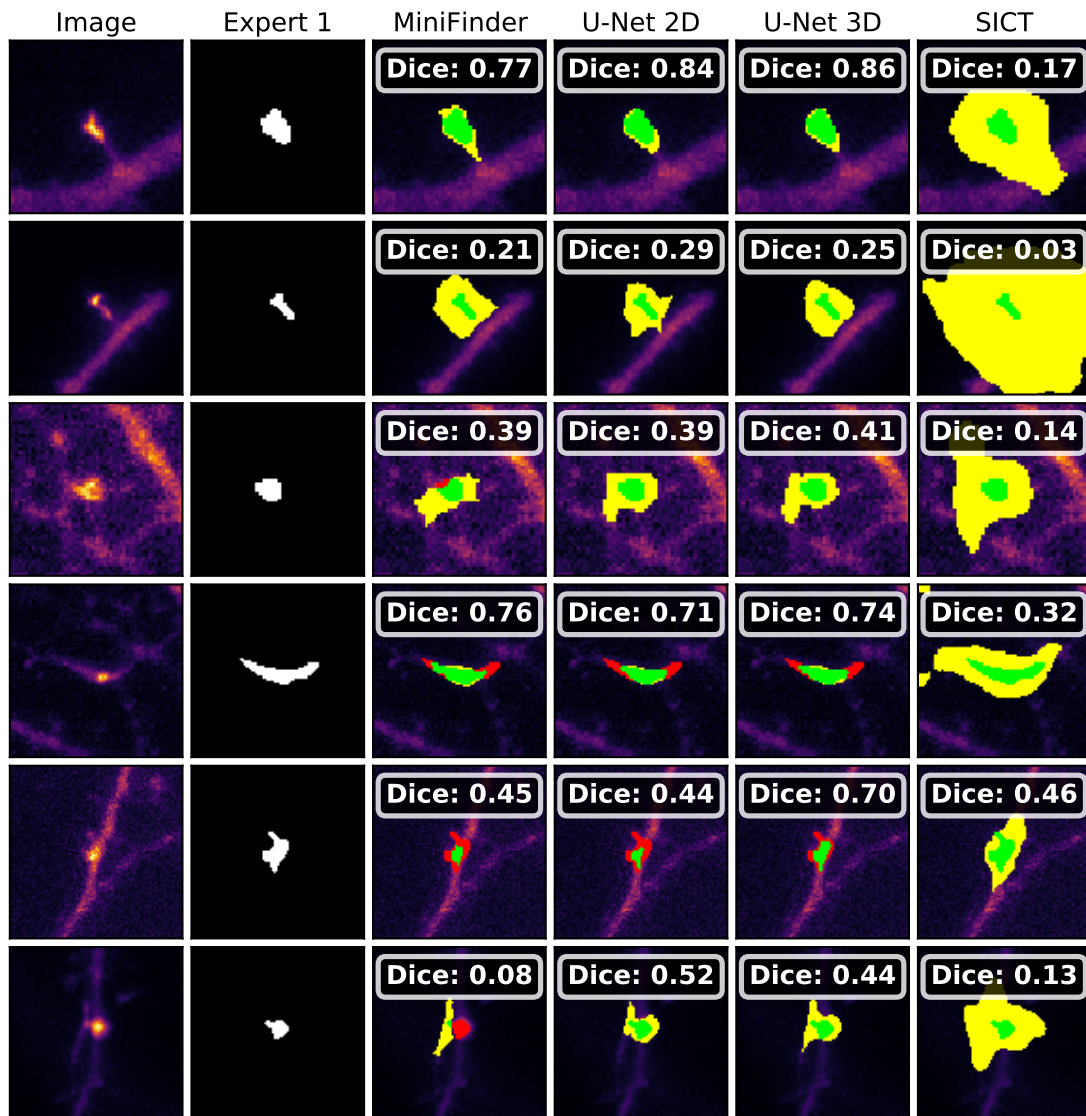


FIGURE 3.4 – Exemples de segmentations comparées avec l’annotation précise d’un expert. Le code de couleur vert, jaune, rouge correspond aux VP, FP, FN respectivement.

segmentation des mSCTs que la variation de fluorescence utilisée *MiniFinder*. Par contre, le U-Net 2D est significativement meilleur à *MiniFinder* uniquement en comparant avec l’expert 1. Notamment, le U-Net 3D comparé avec les annotations de l’expert 2 (figure 3.3(b)) a un *Dice* (0,64) égal au *Dice* quantifié par la variation inter-expert à la figure 3.2(b). Cela signifie que la segmentation de ce réseau est comparable à la variation présente entre deux experts du domaine.

L’inconvénient de l’approche *SICT*, visible dans les exemples de segmentation dans la colonne de droite à la figure 3.4, est que la segmentation est beaucoup plus large que la zone d’intérêt réel du mSCT. Cela se répercute fortement dans les faibles valeurs de *Dice* de cette méthode. Plusieurs exemples de segmentation, pour chaque méthode, sont affichés à la figure 3.4.

3.2 Analyse U-Net 2D et U-Net 3D en détection

La prochaine section évalue les performances de détection des différentes approches. Pour ce faire, le centre pondéré de chaque événement calcique segmenté est calculé et défini comme une détection. Les détections prédites et celles de références sont affectées entre elles avec la méthode d’assignation Hungarian (la méthode est détaillée à la section 2.10.3). Ces résultats sont calculés sur tout le jeu de données de test.

3.2.1 Comparaison avec *MiniFinder*

Les performances sont comparées avec la détection d’un événement calcique de *MiniFinder* utilisées pour entraîner les réseaux. La figure 3.5(a) présente le nombre d’événements détecté, associé, ou non, avec *MiniFinder*. Les zones vertes, jaunes et rouges indiquent le nombre de vrais positifs, faux positifs et faux négatifs respectivement.

Au total, il y a 909 et 1060 événements détectés pour le U-Net 2D et U-Net 3D respectivement. Le rappel est élevé pour les deux approches, soit 86,0% et 92,2%. En nombre de détections, le U-Net 3D comporte 34 faux négatifs de moins que le U-Net 2D. À première vue, dû au grand nombre de faux positifs, la précision est basse, soit $474/(474+358) = 57,0\%$ et $508/(508+509) = 50,0\%$. En effet, comme le U-Net 3D produit beaucoup plus de faux positifs, sa précision est plus basse que le U-Net 2D. Pour mieux visualiser les événements calciques détectés dans la figure 3.5(a), les figures 3.5(b) et 3.5(c) montrent la distribution des VP, FP et FN selon l’intensité calcique normalisée. Les FP, en jaune, sont majoritairement présents dans la zone de faible intensité, par exemple, dans le cas du U-Net 3D, la zone d’intensité $\Delta F/F_0 < 0,5$ comporte 45,8% de tous les FP. Les événements calciques de référence, déterminés à partir des détections de *MiniFinder*, sont basés sur la définition d’un seuil minimum absolu devant être respecté avant d’effectuer le reste de la détection. Il est alors nécessaire de se demander si ces faux positifs sont effectivement des faux positifs ou bien si ce sont des événements manqués par *MiniFinder*.

3.2.2 Réannotation des faux positifs

Pour vérifier que des événements calciques détectés par le U-Net 2D ou le U-Net 3D n'ont pas été manqués par *MiniFinder*, tous les FP sont inspectés visuellement et réannotés. La procédure de réannotation est représentée à la figure 3.6(a). Un événement calcique FP ainsi que sa trace calcique sont présentés à l'expert, lequel doit prendre une décision de la validité de cet événement. Un événement calcique détecté est valide si sa segmentation est jugée bonne par l'expert et si la trace calcique au temps de détection comporte le pic de fluorescence caractéristique d'une entrée calcique dans un neurone (tel que le schéma de trace calcique présenté dans la figure 3.6(a)). Le cas échéant, celui-ci est considéré comme un vrai positif (VP) ou, dans le cas contraire, conserve l'étiquette faux positif (FP). La figure 3.6(b) dépeint le processus de réannotation avec le nombre de détections avant et après la réannotation. Le rappel ne change pas pour les deux approches après la réannotation (86,0% et 92,2%). La précision par événement, quant à elle, augmente. Pour le U-Net 2D, elle passe de 57,0% à $785/(785+47) = 94,3\%$ et pour le U-Net 3D, elle passe de 50,0% à $955/(955+62) = 94,0\%$.

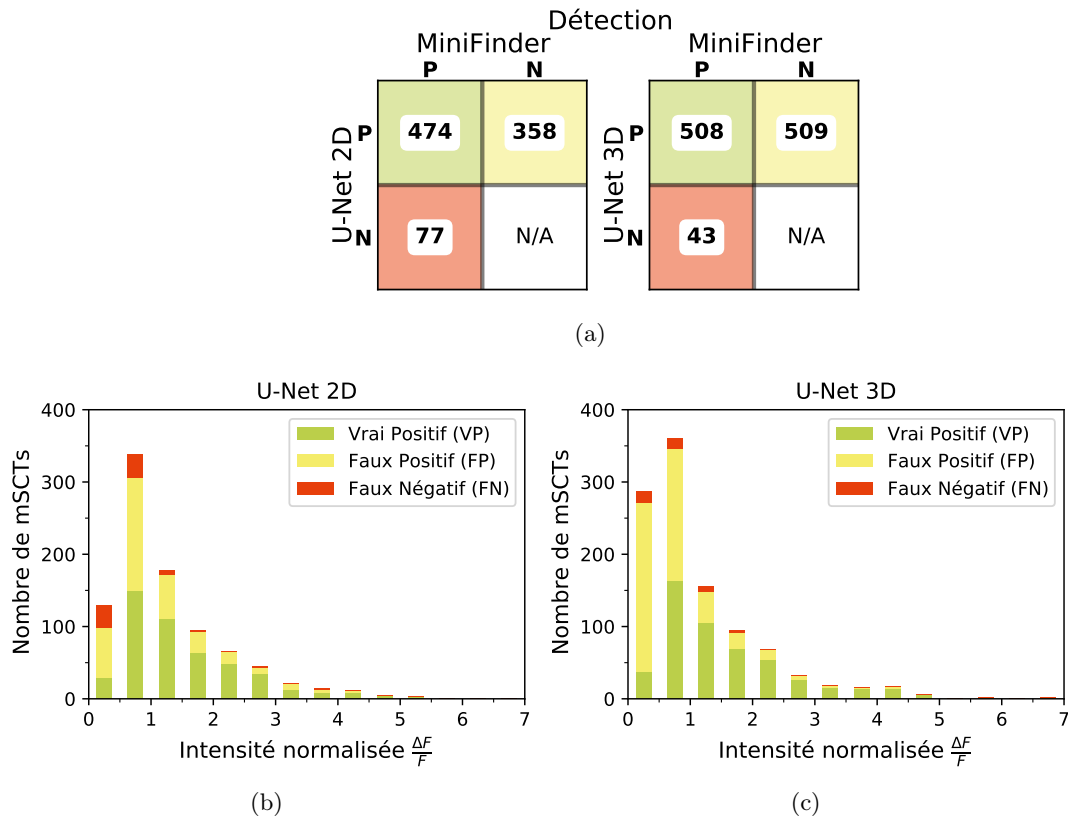


FIGURE 3.5 – Comparaison en détection des approches U-Net 2D et U-Net 3D avec *MiniFinder* pour tous les mSCTs du jeu de test. (a) Nombre de détections : vrais positifs (*vert*), faux positifs (*jaune*) et faux négatifs (*rouge*). (b) et (c) Distribution des détections en (a) selon l'intensité normalisée des événements pour les deux approches respectives.

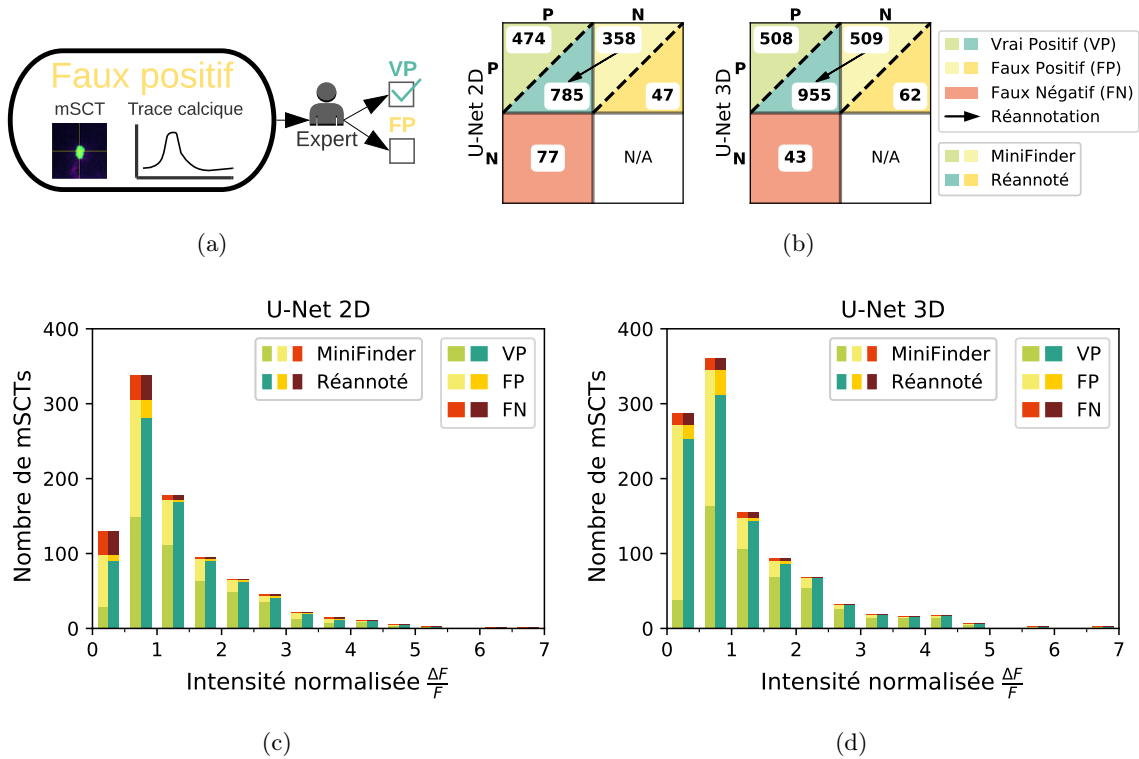
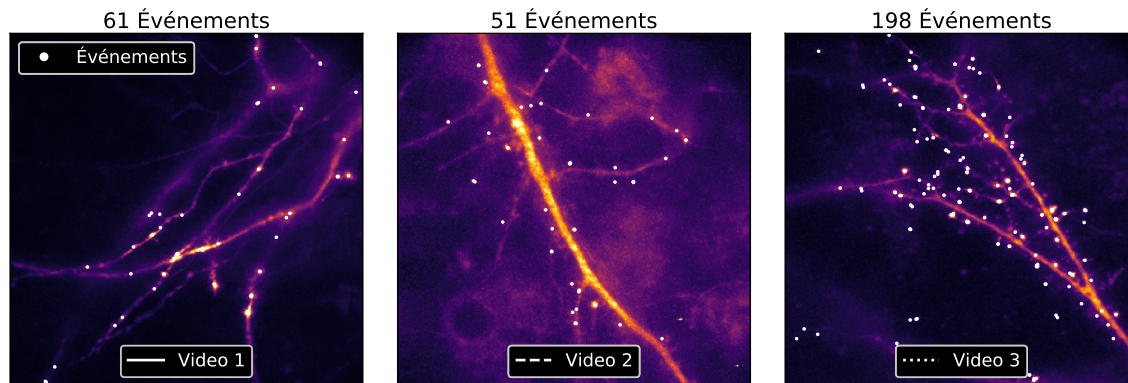


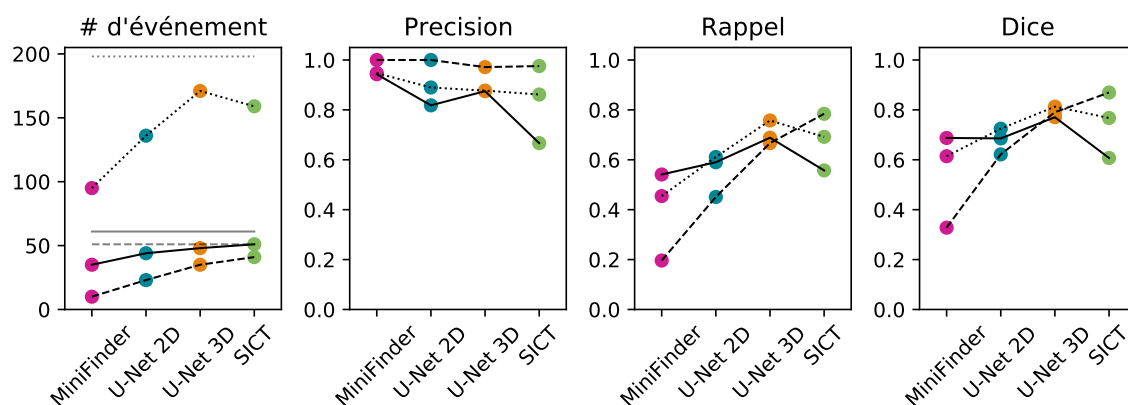
FIGURE 3.6 – Réannotation des faux positifs (FP) du jeu de données de test. (a) Schéma du processus de réannotation par un expert. Les FP sont identifiés et sont présentés à un expert. L'expert peut inspecter visuellement tout l'événement ainsi que sa trace calcique (l'application de visualisation est présentée à la section 4.2). L'expert prend une décision si le FP en est réellement un ou si l'il s'agit d'un VP. (b) Nombre de détections avant et après la réannotation de tous les FP (séparé par la diagonale pointillée). La flèche noire représente les FP réannotés en VP. (c) et (d) La distribution des mSCTs selon leur intensité normalisée avant et après la réannotation pour le U-Net 2D et le U-Net 3D respectivement.

Le *dice* par événement est calculé pour avoir une performance globale des deux approches : $Dice_{2D} = 2 \times 785 / (2 \times 785 + 47 + 77) = 0,927$ et $Dice_{3D} = 2 \times 955 / (2 \times 955 + 62 + 43) = 0,948$. Le U-Net 3D offre une performance globalement meilleure. En effet, bien que sa précision soit légèrement plus basse que le U-Net 2D, le U-Net 3D détecte $955 - 785 = 170$ vrais positifs supplémentaires et affiche un *dice* par événement supérieur.

Les figures 3.6(c) et 3.6(d) présente les distributions d'événements selon leur intensité normalisée pour le cas avant et après la réannotation (comparaison avec *MiniFinder* et réannoté). Dans le cas du U-Net 3D, après réannotation, 83,9% des FP et 72,1% des FN ont une intensité normalisée $\Delta F / F_0 < 1$. Le réseau a donc plus tendance à faire des erreurs pour des événements de faible intensité. C'est un résultat entendu puisque ces événements de faible intensité ne font pas partie du jeu d'entraînement.



(a)



(b)

FIGURE 3.7 – Comparaison en segmentation des approches U-Net 2D, U-Net 3D, *MiniFinder* et *SICT* avec les experts. (a) Trois vidéos du jeu de test ont été inspectés minutieusement et uniquement le centre de chaque mSCT a été annoté (aucune segmentation). Toutes les détections sont représentées par un point blanc sur la projection maximale du vidéo. En (b), de gauche à droite, le nombre d'événements détecté (les lignes pâles horizontales montrent le nombre d'événements détectés par l'expert pour chaque vidéo), la précision, le rappel et le Dice par événement pour chaque approche. Les scores reliés par une ligne proviennent du même vidéo inspecté.

3.2.3 Comparaison avec annotation précise

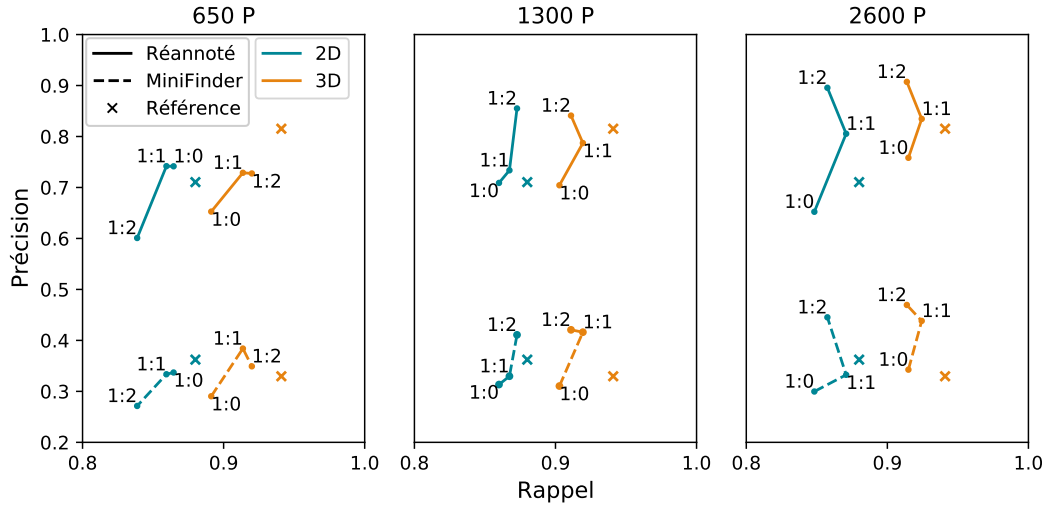
La section précédente permet d'analyser la validité des événements détectés selon *MiniFinder* et, ensuite, en réannotant les FP des réseaux. Cela a comme inconvénient de biaiser l'analyse vers ces événements détectés et de négliger de possibles événements non détectés par *MiniFinder*, U-Net 2D ou U-Net 3D. Il est donc essentiel d'analyser les performances en détection avec un petit jeu de données annoté manuellement par un expert. L'expert a inspecté minutieusement trois vidéos du jeu de test et annoté le centre des mSCTs détectés (aucune segmentation). Les détections (rond blanc) sont affichées sur la projection maximale des vidéos

à la figure 3.7(a). Ensuite, les performances des différentes approches sont comparées avec ce jeu de données (fig. 3.7(b)). La ligne noire reliant les performances entre les approches fait le lien entre un même vidéo. Aucune approche ne détecte tous les événements calciques (à gauche dans fig. 3.7(b)). Bien que la procédure de seuillage *MiniFinder* comporte un taux de détection d'événements le plus faible (rappel bas), c'est la méthode la plus précise. Pour toutes les approches, la variation du rappel, entre les trois vidéos, montre que ces méthodes sont peu robustes. Le U-Net 3D présente un rappel ayant le moins de variation, autour de 0,7, suggérant ainsi une robustesse quant au pourcentage d'événements détectés selon différents vidéos. D'un point de vue de la détection, le U-Net 3D est meilleur que le U-Net 2D. Le U-Net 3D fait toujours plus de détections (meilleur rappel) que le U-Net 2D avec une précision comparable. *SICT* affiche des performances semblables au U-Net 3D. Finalement, le U-Net 3D a taux de détection semblable à *SICT*, mais offre une segmentation nettement meilleure (sec. 3.1.3).

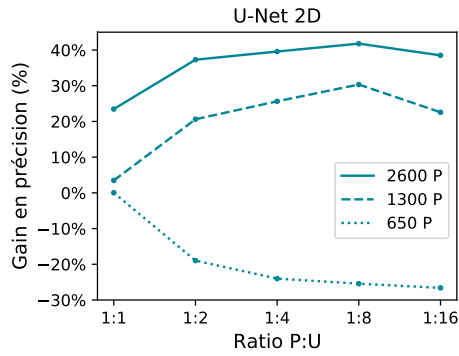
3.3 Analyse de l'apprentissage PU

Suite aux analyses détaillées des performances de détection et de segmentation des U-Net 2D et 3D, l'analyse qui suit présente l'impact du nombre de *Positive* (P), du nombre de *Unlabeled* (U) et du ratio PU dans un entraînement d'apprentissage PU.

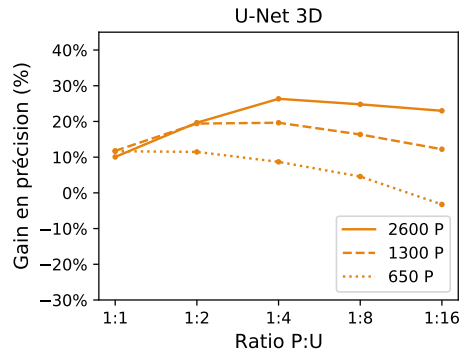
Pour valider l'approche d'apprentissage PU proposée, elle est d'abord comparée avec une approche d'entraînement *classique* ne comportant aucune région U. Le nombre de mSCTs maximum dans un pli d'entraînement est de 7800 (7800 P). Un entraînement, sans la méthodologie d'apprentissage PU, est effectué en utilisant les 7800 P. Cet entraînement *classique*, nommé la référence (dénotée par un \times dans la figure 3.8(a)), utilise toutes les données disponibles et permet donc de définir la performance maximale atteignable avec un entraînement *classique* pour un nombre de données défini. Les comparaisons avec l'entraînement de référence utilisent au maximum 7800 régions (P et/ou U) en faisant varier le nombre de P et le ratio P :U. Spécifiquement, trois ratios P :U {1:0 ; 1:1 ; 1:2} sont analysés pour trois collections de P {650 ; 1300 ; 2600}, et ce, pour le U-Net 2D et 3D. De cette façon, le nombre maximal de régions utilisées pour l'entraînement, c'est-à-dire 2600 P et un ratio 1:2 (2600 P + 5200 U = 7800 régions), ne dépasse jamais la référence de 7800 régions P. Chaque combinaison P et PU est comparée avec *MiniFinder* et la réannotation de l'expert. Le rappel varie peu au cours des différentes combinaisons P et U testées, cela implique que les performances pour prédire les étiquettes générées par *MiniFinder* pour entraîner les réseaux ne sont pas affectées par l'approche PU tel que présentée. L'impact du PU est observable par la variation de la précision après réannotation. Chaque donnée représente la moyenne des trois entraînements (un par pli).



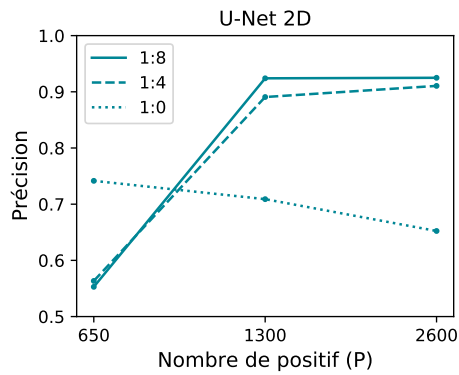
(a)



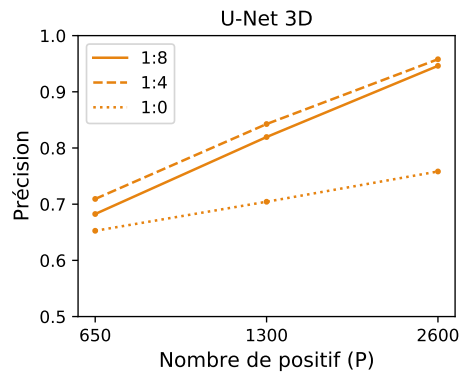
(b)



(c)



(d)



(e)

FIGURE 3.8 – Analyse de l’impact des régions *Positives* (P) et *Unlabeled* (U). (a) Précision et rappel pour 650, 1300, 2600 P et différents ratios PU de 1:0, 1:1, 1:2 pour les deux approches de réseaux. La référence \times représente 7800 P et 0 U, soit l’équivalent P du nombre de régions maximales utilisées en entraînement (ratio 1:2 avec 2600 = $1 * 2600 + 2 * 2600 = 7800$). (b) et (c) Pourcentage de gain en précision selon le ratio 1:0 en fonction de différents ratios PU pour différents P pour le U-Net 2D et le U-Net 3D respectivement. (d) et (e) Précision selon le nombre de P pour différent ratio PU.

3.3.1 Impact du ratio PU

Les graphiques 3.8(b) et 3.8(c) présentent le gain en précision par rapport au cas où il n’y a pas de U, c’est-à-dire un ratio PU de 1:0. Dans le cas du U-Net 2D, un ratio PU de 1:8 semble offrir le meilleur gain précision. Pour 1300 et 2600 P, le gain en précision est maximal avec un ratio PU de 1:8 (fig. 3.8(b)). De plus, il semble qu’avec un trop grand ratio PU (un ratio de 1:16), la précision commence à diminuer. Avec un nombre trop faible de positifs, l’ajout de régions U fait diminuer la précision par rapport au cas sans U selon la courbe pointillée 650 P de la figure 3.8(b). Il est donc important d’avoir une quantité suffisante de régions P afin d’ajouter des régions U à l’entraînement dans le cas du U-Net 2D.

La dynamique de gain en précision selon le ratio PU est différente pour le U-Net 3D. Tout d’abord, le gain maximal en précision, soit 26,3%, a lieu avec un ratio de 1:4 pour 2600 P. Ce gain est nettement plus faible (26,3% vs 41,8%) que celui du U-Net 2D. Cela est causé par la faible précision du ratio 1:0 avec 2600 P pour le U-Net 2D ce qui a comme effet d’augmenter les pourcentages dans le calcul du gain (dans le graphique de droite de la figure 3.8(a), les précisions après réannotation pour un ratio 1:0). D’ailleurs, malgré le gain en précision plus faible du U-Net 3D, la valeur maximale absolue de la précision est tout même plus grande que le U-Net 2D, voir figure 3.8(d) et 3.8(e) (3D : précision de 0,958 vs 2D : précision de 0,925). Comme dans le cas du U-Net 2D, un ratio trop élevé détériore le gain en précision. Il y a donc un compromis à faire pour le choix du ratio PU.

La performance par rapport à la référence de 7800 P (\times dans la figure 3.8(a)) est semblable pour les deux approches. Avec 650 P, quel que soit le ratio PU, la référence est toujours supérieure. En effet, le nombre de régions P semble trop faible pour permettre aux réseaux à convolution d’apprendre des filtres adéquats. Par contre, avec nombre de P inférieur à 7800, il est possible de dépasser cette référence. Par exemple, pour le U-Net 3D avec seulement 2600 P et un ratio PU de 1:1 (fig. 3.8(a), à droite), la précision est meilleure que la référence. Cela permet d’entraîner un réseau en utilisant beaucoup moins d’échantillons P, donc moins d’événements calciques, qui sont coûteux à annoter dans un contexte d’imagerie de neurones biologiques.

3.3.2 Impact des P

Dans le cas du U-Net 2D, le nombre de P utilisé a un impact considérable sur les résultats en précision après entraînement. À la figure 3.8(d), la précision diminue abruptement avec l’ajout de U pour 650 P. Quand le nombre de P augmente, la précision plafonne rapidement autour de 0,9 avec 1300 P. De même, la précision plafonne aussi avec un ratio de 1:4, car la différence entre les courbes de ratio 1:4 et 1:8 est près. Pour le U-Net 3D, la précision augmente linéairement avec le nombre de P, du moins jusqu’à 2600P. La précision est maximale pour le ratio 1:4 et diminue légèrement pour tous les P avec le ratio 1:8.

Selon cette analyse empirique, il semble que l’architecture 3D est capable d’utiliser les P additionnels afin d’améliorer la précision, ce qui n’est pas le cas avec l’architecture 2D où la précision plafonne rapidement. De plus, dans le cas de l’architecture 2D, l’ajout de U avec une trop faible quantité dans P détériore les performances lors de l’entraînement.

3.4 Discussion

La segmentation du U-Net 3D est significativement supérieure au U-Net 2D lorsqu’elle est comparée à la segmentation de *MiniFinder*. Le *Dice*, utilisée comme métrique de segmentation, a ici des valeurs relativement basses (une médiane de 0,69 et 0,77 pour le U-Net 2D et 3D) dues à la difficulté d’une segmentation précise au niveau des pixels et aux petites instances à segmenter. Pour quantifier ces faibles valeurs du coefficient *Dice*, l’analyse intra- et inter-expert est faite. La variabilité inter-expert, c’est-à-dire lorsque deux experts annotent les mêmes masques, comporte un coefficient de *Dice* médian de 0,68 et 0,70 pour deux rounds d’annotation. Lorsque la segmentation du U-Net 3D est comparée aux annotations d’experts, sa performance est significativement meilleure que les autres approches et semblable à la variabilité présente entre deux experts. Notamment, même si les segmentations de *MiniFinder* sont utilisées pour entraîner le U-Net 3D, le réseau apprend à faire des segmentations significativement meilleures que *MiniFinder*. Cela signifie que les filtres appris par le réseau généralisent mieux que la procédure utilisée pour faire la segmentation.

Le U-Net 2D et 3D vont tous les deux apprendre à des caractéristiques permettant de mieux généraliser la détection des événements calciques de faible intensité, c’est-à-dire avec une intensité normalisée $\Delta F/F_0 < 1$, difficile à détecter pour *MiniFinder*. Pour arriver à cette conclusion, il a fallu réannoter tous les événements faux positifs des réseaux. Après réannotation, le U-Net 3D détecte 170 événements supplémentaires que le U-Net 2D. En contrepartie, il fait 15 mauvaises détections d’événements supplémentaires. Globalement, le U-Net 3D performe mieux que le U-Net 2D avec un *Dice* par événement de 0,948 vs 0,927. La majorité (83,9%) des fausses détections du U-Net 3D ont lieu également pour une intensité normalisée $\Delta F/F_0 < 1$.

L’ajout des régions U dans l’entraînement semble permettre aux réseaux de régulariser davantage ses filtres, ce qui augmente la précision et, par le fait même, diminue le nombre de faux positifs. Par contre, l’ajout de régions U peut détériorer les performances si le nombre de P n’est pas suffisant, comme c’est le cas avec 650 P pour le U-Net 2D. De plus, un trop grand nombre de U peut aussi impacter négativement les performances du réseau, 2D ou 3D. Il y a un compromis à faire dans le choix du ratio PU. En effet, avec un nombre suffisant de données P (par exemple 2600P pour le U-Net 3D) et un ratio PU de 1:4, les meilleurs résultats sont obtenus. Avec moins de P ou un ratio PU supérieur, les performances diminuent. L’avantage supplémentaire (outre le gain en précision) de l’utilisation de régions U est

TABLE 3.1 – Performances globales des différentes approches proposées pour la détection et la segmentation des mSCTs. [†]Les métriques de performances sont comparées avec le jeu de données annoté précisément. [‡]Les temps sont calculés pour un vidéo de 600 images avec une résolution 512×512.

Méthode	Detection [†]			Segmentation [†]	mSCTs	Temps inférence [‡]	Temps annotation [‡]
	Precision	Rappel	<i>Dice</i>	<i>Dice</i>			
MiniFinder	0.96	0.40	0.54	0.53	140	10 min	15 min
U-Net 2D	0.90	0.55	0.68	0.57	203	6 min	N/A
U-Net 3D	0.91	0.70	0.79	0.61	254	3 min	N/A
SICT	0.83	0.68	0.75	0.30	251	10 min	10 min

qu'il est possible d'avoir des résultats supérieurs avec une quantité moindre de régions P par rapport à n'utiliser aucune région U. Selon l'application, si les P sont coûteux à obtenir (coût humain, monétaire, éthique) contrairement au U, la diminution des P et l'augmentation des U est une méthode intéressante pour limiter le nombre d'annotations coûteuses. De manière générale, ajouter une dimension de convolution supplémentaire au réseau permet d'obtenir de meilleures performances en segmentation, mais plus particulièrement en détection. C'est conclusions sont très spécifiques à l'application analysée ici, elles peuvent servir de point de départ pour d'autres applications, mais n'assurent en rien leur validité sur d'autres types de données.

La table 3.1 dresse un portrait global des avantages et inconvénients de chacune des méthodes analysées dans ce chapitre. Le U-Net 3D ressort par rapport aux autres. En détection, il comporte une précision acceptable, le meilleur rappel et le meilleur *Dice*. En comparant avec une segmentation d'expert, il se démarque significativement par rapport aux autres. L'avantage certain, tout comme le U-Net 2D, est sa vitesse d'inférence et le fait que cette approche soit entièrement automatique. Enlever l'humain de la boucle est un avantage indéniable par rapport aux autres méthodes de référence, *MiniFinder* et *SICT*, qui elles contiennent un temps non négligeable de visualisation par un expert du domaine pour éliminer les détections erronées. Une note importante, l'étape de réannotation (à la section 3.2.2) pour les approches U-Net 2D et U-Net 3D est effectuée seulement à titre d'analyse des performances, elle n'est pas nécessaire pour inférer sur de nouveaux vidéos.

Chapitre 4

Outils développés

Ce chapitre décrit les outils développés pour permettre une analyse quantitative des événements calciques. Les besoins essentiels pour mener cette analyse sont : 1) la gestion des données ; 2) la sauvegarde des résultats ; 3) la visualisation efficace des événements calciques ; et 4) la possibilité de faire des routines d’analyses automatisées à grande échelle. Les choix technologiques proposés pour répondre à ces besoins pourraient servir à d’autres projets, malgré qu’ils aient été développés spécifiquement pour cette analyse.

4.1 Gestion des données

Les besoins associés à la gestion des données dérivent de la nature des expérimentations biologiques effectuées, en voici une brève description.

4.1.1 Expérimentations biologiques

Le but est de caractériser l’activité neuronale spontanée en quantifiant son effet. Différentes conditions (ou traitements) sont appliqués aux neurones pour amoindrir ou accentuer l’activité calcique. Avant un traitement, il est nécessaire d’imager le niveau d’activité initial de chaque neurone. Ce niveau d’activité basal est mesuré par le biais d’un vidéo contrôle, acquis quelques minutes avant d’appliquer le traitement sur le neurone. Plusieurs caractéristiques spécifiques peuvent, ou non, avoir un impact sur le niveau d’activité calcique : l’âge de la culture de neurones, la journée d’imagerie, l’heure d’imagerie, etc. Il est essentiel de conserver toutes ces métadonnées pour les analyses *post-hoc* de biais dans les données. L’obtention de résultats significatifs nécessite plusieurs neurones imagés pour un même traitement. Finalement, les neurones semblables sont combinés et comparés pour caractériser la plasticité synaptique selon l’activité calcique pour différents traitements. Pour de plus amples détails sur les expérimentations et les conclusions biologiques, se référer à la thèse de doctorat de Theresa Wiesner (Wiesner, 2021).

Un vidéo d'imagerie de l'activité calcique dure au minimum 90 secondes, avec une taille de 512×512 pixels et une résolution temporelle de 150 ms, pour un total de 600 captures. Étant donnée la grande plage dynamique de ce type d'imagerie (exemple à la sec. 2.1), l'intensité est enregistrée sur 16 bits. Un vidéo occupe au minimum 300 MB d'espace sur le disque. L'espace total occupé par les vidéos est environ 1 TB (soit 2000 vidéos). Le réseau de neurones (le U-Net 3D développé) est utilisé pour inférer une segmentation complète des événements calciques présents dans les vidéos. À partir de caractéristiques extraites des mSCTs, des analyses comparatives sont effectuées afin de déceler des différences induites par les traitements appliqués.

Il y a deux tâches de gestion de données : la gestion des métadonnées et la gestion des résultats provenant de la caractérisation.

4.1.2 Sauvegarde des métadonnées

Chaque vidéo comporte des métadonnées qui lui sont propres (elles sont cataloguées dans l'annexe A.2). Les métadonnées sont des ressources fréquemment visitées qui permettent d'avoir une vision globale des données. Ainsi, les métadonnées doivent être facilement accessibles, modifiables et partageables par différents usagers sur différents postes de travail. Par exemple, sur l'ordinateur du microscope pendant l'acquisition des données et sur le poste d'analyse. La solution adoptée pour organiser les métadonnées est la feuille Google (*Google Spread Sheet*)¹. Chaque colonne de la feuille correspond à une métadonnée, tandis que chaque ligne représente un vidéo unique. La procédure de sauvegarde consiste à ajouter les métadonnées associées à un vidéo en l'ajoutant à la feuille.

Les avantages de la feuille Google sont :

1. Accès concurrent : la visualisation et la modification, en parallèle, des métadonnées.
2. Centralisation : la sauvegarde des métadonnées à un seul endroit élimine les duplications entre différents usagers et assure leurs concordances.
3. Ressources minimales : nécessite uniquement un appareil possédant un navigateur internet.
4. Partage : l'ajout d'usagers, et ce, avec un contrôle des droits de modification.
5. Accès informatique : la feuille est accessible par des ressources externes, en lecture et en écriture, grâce à l'API².
6. Simplicité : aucune programmation ni installation n'est requise.

Or, la version actuelle d'utilisation de la feuille Google comporte les désavantages suivants :

1. Rigidité : la contrainte de devoir travailler avec des métadonnées sous forme de tables. Par exemple, ajouter une caractéristique pour un seul vidéo n'est pas possible puisqu'il

1. https://workspace.google.com/intl/en_ca/products/sheets/

2. <https://developers.google.com/sheets/api>

faut créer une nouvelle colonne dans la table et assigner une valeur nulle aux autres vidéos.

2. Visualisation : l'interface d'analyse et d'inspection des métadonnées est limitative, particulièrement quand la feuille contient beaucoup de vidéos.
3. Accès internet : nécessite un accès internet pour accéder aux métadonnées.

La liste des avantages liés à l'utilisation de la feuille Google démontre que cet outil est adéquat pour cette expérimentation. La rigidité associée à l'utilisation d'une table pour représenter les métadonnées des vidéos n'est pas restrictive puisque toutes les métadonnées sont définies au préalable et ne changent pas au cours du temps. Par contre, lorsqu'une feuille contient beaucoup de vidéos (beaucoup de lignes) l'inspection des métadonnées devient quelque peu incommode, attribuable au manque d'outils pour faciliter la navigation. Néanmoins, puisque Google fournit un API permettant d'interfacer la feuille, ce problème de visualisation peut être réglé avec d'autres outils, lesquels sont discutés dans les prochaines sections.

4.1.3 Sauvegarde des résultats

Les résultats provenant de la caractérisation des vidéos doivent être sauvegardés et facilement accessibles a posteriori. De plus, ils doivent être liés aux métadonnées associées aux vidéos afin d'effectuer des analyses comparatives des résultats selon différentes métadonnées. Heureusement, la gestion de données est une situation rencontrée dans beaucoup d'autres domaines et de multiples solutions sont offertes. Le choix de la technologie utilisée est basé sur les caractéristiques et les contraintes associées au projet.

La liste suivante dénombre les spécifications souhaitées :

1. La possibilité de sauvegarder une grande quantité de données.
2. La possibilité de sauvegarder des résultats sous différentes formes, c'est-à-dire, du texte, des images, des vidéos, des tableaux de données.
3. La flexibilité de pouvoir ajouter ou supprimer des données pour un vidéo, indépendamment des autres.
4. La possibilité de sauvegarder des métadonnées hétérogènes.
5. L'utilisation et la visualisation simple de la base de données.

Avec ces spécifications, il est possible de générer plusieurs types de résultats par vidéo et de les sauvegarder. La sauvegarde de métadonnées hétérogènes permet ainsi d'associer les métadonnées définies dans la feuille Google (sec. 4.1.2) au vidéo. Ainsi, cette association permet de filtrer les résultats selon les métadonnées (par exemple, déterminer le nombre de mSCTs détecté pour une expérimentation spécifique). La simplicité d'utilisation et de visualisation de l'outil de gestion facilite grandement le développement logiciel.

Le fichier HDF5

La gestion des données provenant des résultats d'analyse et de caractérisation est basée sur l'utilisation des fichiers *Hierarchical Data Format*, version 5 (HDF5) (Koranne, 2011). Le fichier HDF5 répond aux spécifications énumérées précédemment. Le format HDF est utilisé dans le domaine académique et dans le domaine industriel (Anderson *et al.*, 2010; Folk *et al.*, 2011; Giovannucci *et al.*, 2019). C'est le format choisi pour la gestion de données en neurosciences dans l'effort de normalisation des standards de sauvegarde proposée par *Neurodata Without Border*³.

Le groupe HDF⁴ offre le logiciel et le format de fichier HDF pour faciliter son utilisation. Le code source est ouvert à tous. Les systèmes d'exploitation les plus utilisés (Windows, Mac, Linux) sont supportés. Le logiciel *HDFView*⁵ est l'interface disponible pour visualiser l'état courant d'un fichier HDF. De plus, un interface de programmation Python est aussi offert sous la librairie *h5py*⁶. Un fichier HDF n'a pas de limite de taille et celui-ci permet des accès concourants en lecture.

Un fichier HDF5 agit, grossièrement, comme un conteneur de fichiers. Sa structure s'apparente à l'arborescence de dossiers que l'on retrouve dans le système de fichiers d'un ordinateur courant. Le dossier *root* est le point de départ de l'arborescence, duquel il est possible de créer des dossiers et des sous-dossiers. Ce type de fichier implémente différents objets et abstractions. Parmi ceux-ci, deux sont utiles à définir : les *group* et les *dataset*.

Group. L'abstraction *group* est l'équivalent d'un dossier dans un système de fichiers standard. Il est possible d'insérer dans un *group* tous les autres types d'objets, incluant un *group*. Ceux-ci permettent de structurer l'arborescence du fichier. L'objet *group* est un concept intuitif qui permet de créer facilement une structure flexible et adaptée. Le *group* supporte l'ajout de métadonnées par le biais de ses *attribute*.

Dataset. Le *dataset* est l'objet qui permet de créer une abstraction vers un tableau NumPy. Il permet de sauvegarder un tableau de données homogènes directement sur le disque (la taille du tableau peut donc être aussi grande que le disque le permet). Contrairement au tableau NumPy, des fonctionnalités transparentes de stockage sont ajoutées telles que la compression, la détection d'erreurs et les écritures/lectures par morceaux. Dans notre cas, l'objet *dataset* est particulièrement pratique pour ces quatre raisons :

1. Il supporte la majorité des opérations d'un tableau NumPy⁷.
2. Il permet un chargement paresseux par défaut (*lazy loading*). Cela permet d'accéder à une section d'un tableau sans le charger entièrement en mémoire.

3. <https://www.nwb.org/>

4. <https://www.hdfgroup.org/solutions/hdf5/>

5. <https://www.hdfgroup.org/downloads/hdfview/>

6. <https://www.h5py.org/>

7. <https://numpy.org/>

3. Il supporte la compression *gzip* selon différents niveaux de compression.
4. Il supporte l'ajout de métadonnées par le biais des *attribute*.

Ainsi, la gestion des données est basée sur la feuille Google et le format de fichier HDF5. Les résultats de chaque vidéo sont compartimentés dans un *group*. À cet objet *group* sont associées les métadonnées définies dans la feuille Google. Les métadonnées permettent de filtrer les vidéos. Tous les résultats découlant de l'inférence du U-Net 3D et de la caractérisation morphologiques des mSCTs sont enregistrés dans le *group* du vidéo respectif. Par exemple, la segmentation dense des vidéos est enregistrée sous forme de tableaux NumPy dans un objet *dataset* et compressée avec l'algorithme de compression *gzip* intégrée. Puisque les mSCTs sont très épars dans le vidéo (beaucoup de pixels à zéro), la compression est efficace avec un ratio *900 :1* de compression. Les caractéristiques des mSCTs (p. ex. la taille, l'intensité et la fréquence des mSCTs) sont enregistrées sous forme de tableau NumPy dans un objet *dataset* sans compression. Finalement, le chargement paresseux d'un *dataset* permet d'accéder à une section du vidéo ou de la segmentation, sans les charger en mémoire vive. Les détails relatifs aux types de résultats sauvegardés ainsi qu'un exemple de la base de données avec le logiciel *HDFView* sont présentés à l'annexe A.3.

4.2 Visualisation des données

La visualisation, en général, est très importante pour valider le fonctionnement des algorithmes proposés. Dans le cas de l'entraînement de réseaux de neurones avec une approche d'apprentissage PU à partir des étiquettes de *MiniFinder*, la validation est d'autant plus importante puisque des événements additionnels (comparé à *MiniFinder*) sont détectés, nécessitant une étape de réannotation par un expert (sec. 3.2.2) pour valider adéquatement les performances. À titre indicatif, 57 réseaux ont été entraînés et chacun d'eux a été évalué sur les 28 vidéos du jeu de données de test. La charge de réannotation est donc importante, d'où le besoin d'avoir une interface adaptée. De plus, les mSCTs (un exemple de mSCT est présenté à la section 2.1) sont de petites instances dans le champ de vue et dans le temps. Dues à la caractéristique éparses des mSCTs, leurs inspections sont laborieuses avec un logiciel standard de visualisation, tel que Fiji (Schindelin *et al.*, 2012), un logiciel libre développé pour visualiser plusieurs types de données biologiques. En effet, le temps nécessaire pour trouver la région d'intérêt, dans un vidéo, est ardu puisqu'il faut naviguer spatialement et temporellement dans celui-ci pour trouver un mSCT. Pour ces raisons, une interface de visualisation adaptée a été développée.

L'interface de visualisation est développée avec la librairie Kivy⁸. Kivy est une librairie récente, distribuée librement, offrant un environnement très riche pour le développement d'interfaces utilisateurs en Python. D'emblée, une application développée avec Kivy est supportée sur plusieurs plateformes (Linux, Windows, OS X, Android, iOS et Raspberry Pi) et des accélé-

8. <https://kivy.org/>

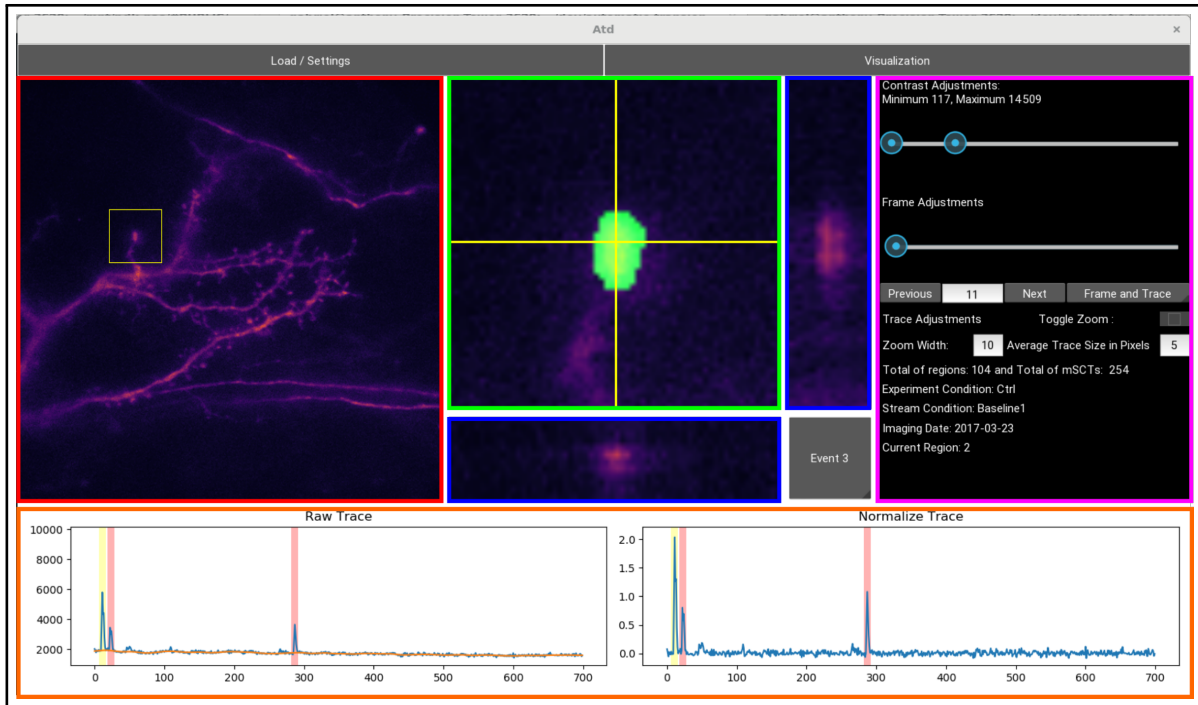


FIGURE 4.1 – Application développée sous Kivy pour la visualisation des mSCTs dans un vidéo. (*Rouge*) Le champ de vue complet est visible dans cet encadré, avec en jaune, la zone d'intérêt. (*Vert*) Le zoom sur le mSCT courant. (*Bleu*) Les deux coupes temporelles, verticale et horizontale, sélectionnées à l'endroit où se trouvent les deux lignes jaunes dans l'encadré vert. (*Rose*) La section pour permettre à l'utilisateur de modifier le contraste de l'image, l'image courante, le type de visionnement et les paramètres pour extraire la trace calcique de fluorescence. Des métadonnées sur le vidéo y sont aussi affichées à titre indicatif. (*Orange*) Les deux traces calciques, originale et normalisée, avec en jaune pâle (sur la trace) la localisation du mSCT affiché. Les zones rouge pâle sont les autres mSCTs faisant partie de la même région que le mSCT affiché. Les traces calciques sont calculées à l'intersection des deux droites jaunes dans la zone verte. La trace peut être zoomée en cliquant sur le bouton *toggle Zoom* dans l'encadré rose.

rations graphiques sont automatiquement intégrées. Un autre aspect supportant le choix de cette librairie est sa documentation claire et adoptée par la communauté.

L'application développée sous Kivy permet d'interfacer la base de données décrite à la section 4.1.3 afin de faciliter la visualisation et la réannotation des mSCTs dans un vidéo d'imagerie calcique. La figure 4.1 présente uniquement la fenêtre consacrée à la visualisation des données. Le reste de l'application est présenté dans l'annexe A.4. Un vidéo de la base de données est sélectionné par l'utilisateur afin d'être visualisé. L'application permet de voir simultanément un mSCT d'intérêt et son positionnement dans le champ de vue. Afin de donner un contexte temporel autour de l'événement calcique, des coupes centrées sur le mSCT sont affichées. Cela permet à l'utilisateur de visualiser l'évolution de la fluorescence dans le temps lo-

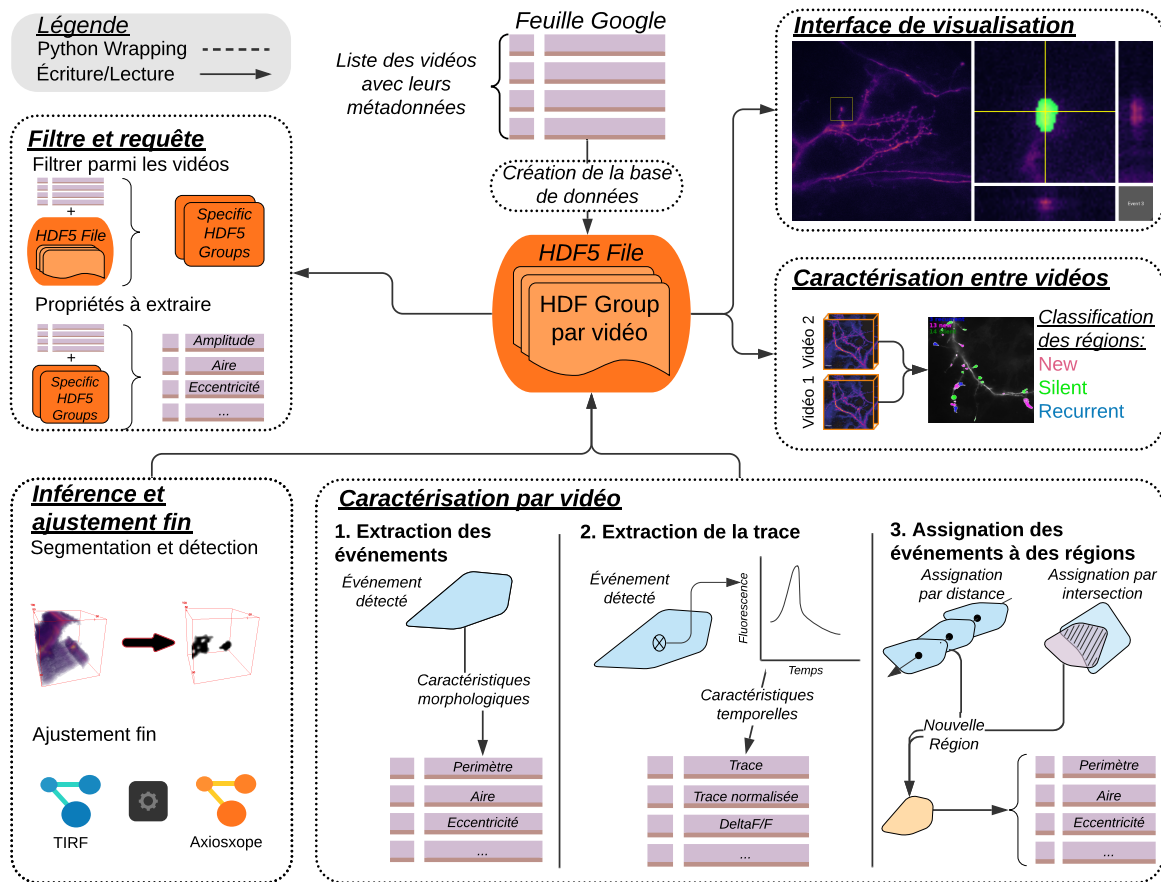


FIGURE 4.2 – Schéma de la routine automatisée d’analyse à grande échelle pour la segmentation, la détection, la caractérisation et la comparaison d’événements calciques.

calement autour de l’événement afin de vérifier sa validité. Les traces de fluorescence calcique, normalisées ou non, sont de bons indicateurs de la validité des mSCTs pour un expert. Les traces sont accessibles, pour la durée du vidéo ou localement autour de l’événement. Divers raccourcis clavier sont aussi disponibles pour naviguer à travers les mSCTs d’un vidéo et spécifier la validité (ou l’invalidité) d’un mSCT.

4.3 Routine automatisée d’analyse à grande échelle

Les outils présentés et le réseau U-Net 3D sont utilisés pour le développement de la routine automatisée à grande échelle pour la segmentation, la détection, la caractérisation et la comparaison d’événements calciques. Cette routine est implémentée en Python, la figure 4.2 présente toutes ses composantes.

Voici un sommaire des différentes étapes de cette routine .

Création de la base de données. La base de données est initialisée à partir d’une feuille Google (sec. 4.1.2) contenant toutes les métadonnées associées aux vidéos. Un *group* HDF est créé pour chaque vidéo ajouté. C’est dans ce *group* que toutes les métadonnées et tous les résultats d’analyse subséquents sont sauvegardés. Les résultats tels que la segmentation dense sont compressés, alors que les données morphologiques et les traces de fluorescence ne le sont pas.

Filtre et requête. Il est possible d’extraire des vidéos en filtrant selon les métadonnées. Les propriétés souhaitées des vidéos filtrés peuvent ensuite être extraites.

Inférence et ajustement fin. **Inférence**. Lorsqu’un réseau de neurones profonds a préalablement été entraîné pour la segmentation des mSCTs, cet algorithme peut être utilisé pour analyser les vidéos dans la base de données. Les résultats de détection et de segmentation des mSCTs sont enregistrés dans le *group* du vidéo respectif sous forme de matrice. **Ajustement fin**. À partir d’un réseau préalablement entraîné avec un jeu de données spécifiques (par exemple avec des données acquises par microscopie TIRF), il est possible d’effectuer une étape d’ajustement fin avec un nouveau jeu de données (par exemple avec des données acquises par microscopie Axioscope) pour spécialiser le réseau de neurones avec ces nouvelles données. Il suffit de spécifier les données à utiliser pour faire l’entraînement et la validation, ainsi que les paramètres d’entraînement (p. ex. le nombre d’itérations).

Caractérisation par vidéo. La caractérisation des mSCTs détectés lors de l’inférence est effectuée en trois étapes distinctes : 1) l’extraction des propriétés pour chaque mSCT ; 2) l’extraction de la trace et le calcul de la dynamique de fluorescence pour chaque mSCT ; et 3) l’identification des correspondances entre les mSCTs appartenant à une même région et l’extraction des propriétés de ces régions. Les résultats sont enregistrés dans le *group* du vidéo sur lequel la caractérisation est effectuée.

Comparaison entre vidéos. Cette méthode permet d’effectuer une analyse comparative entre des vidéos précédemment caractérisés. Elle permet de déterminer, entre deux vidéos, l’apparition, la disparition et la récurrence de régions après l’application d’un traitement chimique ou non.

4.4 Résultats d’analyses biologiques préliminaires

L’analyse automatique de milliers de vidéos d’imagerie de fluorescence calcique permet ainsi de générer une quantité massive de résultats. Ces résultats se prêtent à des analyses quantitatives des phénomènes biologiques, afin de pallier à l’immense variabilité inhérente au système biologique (Hoffman *et al.*, 2021). Trois exemples d’analyses sont présentés utilisant la routine proposée.

Analyse de la distribution des mSCTs. La segmentation des mSCTs dans les vidéos permet de visualiser la distribution des événements calciques ayant lieu sur le neurone. Or,

avec le repérage de ces régions d'intérêt, une analyse comparative entre deux vidéos permet de visualiser (et de quantifier) comment cette distribution d'événements varie. Par exemple, pour un même neurone, la comparaison du vidéo contrôle avec celui acquis après l'application d'un traitement. La figure 4.3 présente les régions récurrentes, nouvelles et silencieuses entre le neurone contrôle et traité.

Analyse d'une caractéristique spécifique. L'analyse de caractéristiques spécifiques est également possible pour tenter de déterminer la variation de celles-ci due au traitement. Par exemple, les distributions de l'aire des mSCTs présents dans les neurones contrôles et cLTP sont obtenues en recueillant cette propriété pour quelques dizaines de neurones (fig. 4.4(a)). Également, la distribution de l'aire des mSCTs avant et après un traitement est présentée pour ces mêmes neurones contrôle ou traité (fig. 4.4(b)). Il semble y avoir une tendance vers l'augmentation de l'aire des mSCTs avant et après, qu'il y ait eu un traitement ou non.

Analyse non supervisée. Finalement, au lieu de faire une analyse par caractéristique des mSCTs, il est également possible d'utiliser des approches non supervisées pour les caractériser. Par exemple, à partir des caractéristiques morphologiques, il est possible de les représenter dans un espace multidimensionnel et, par la suite, d'utiliser des méthodes de réduction de dimensionnalité (Wold *et al.*, 1987; Maaten et Hinton, 2008; McInnes *et al.*, 2020). Dans ce nouvel espace, des groupes peuvent être visibles, ce qui permet de classifier selon les caractéristiques du groupe (Wiesner *et al.*, 2020). La figure 4.5 utilise l'algorithme UMAP (McInnes *et al.*, 2020) pour réduire les caractéristiques morphologiques des mSCTs (aire, aire projetée, amplitude, longueur de l'axe majeur et longueur de l'axe mineur) vers deux dimensions. Les mSCTs nouveaux, silencieux, récurrents avant et après sont présentés en différenciant ceux provenant de neurones contrôles (noir) et traités (coloré). Selon ces résultats préliminaires, l'utilisation du UMAP pour réduire la dimensionnalité des mSCTs ne semble pas être suffisante pour obtenir des groupes distincts.

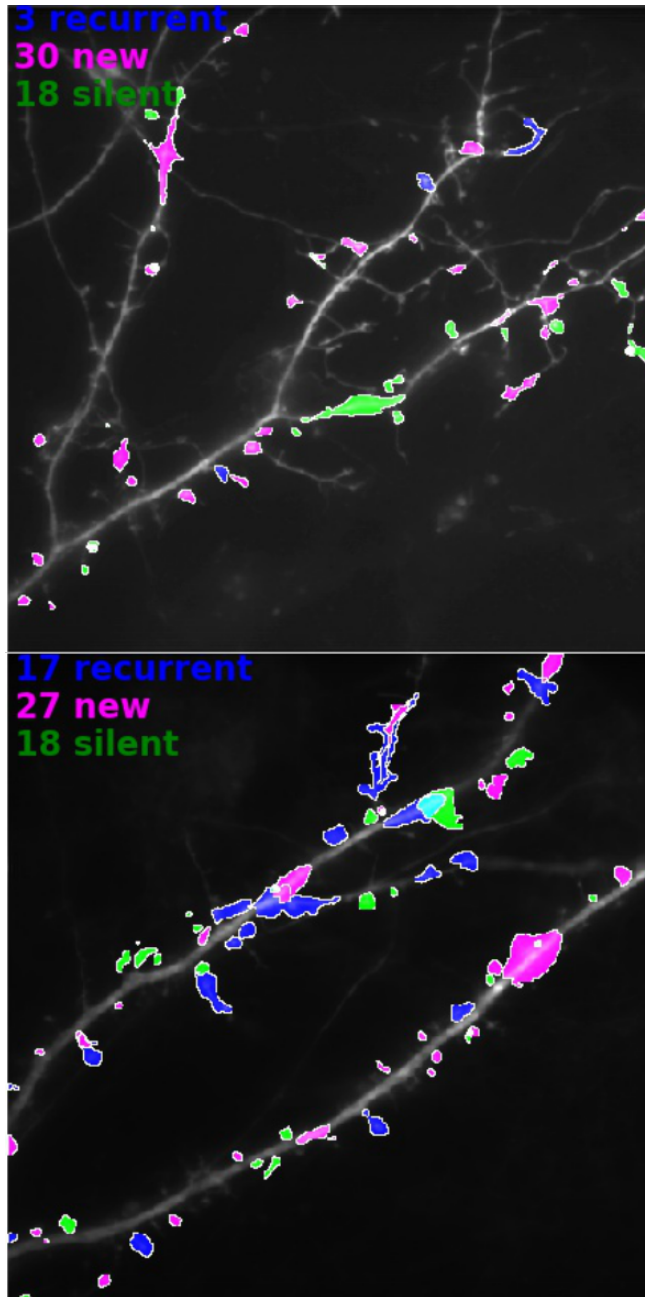


FIGURE 4.3 – Deux exemples de la comparaison entre vidéos. Pour chaque exemple, un vidéo contrôle est utilisé pour effectuer la comparaison. Le neurone est ensuite traité avec une solution pour induire chimiquement la potentialisation à long terme (cLTP) (Lu *et al.*, 2001) et imagé 10 minutes après le traitement. Les régions silencieuses (*vertes*) sont des sites d'activité calcique ayant lieu uniquement dans le vidéo contrôle. Les régions nouvelles (*roses*) ont lieu uniquement dans le vidéo du neurone traité. Les régions récurrentes (*bleues*) sont des régions actives dans les vidéos contrôle et traité.

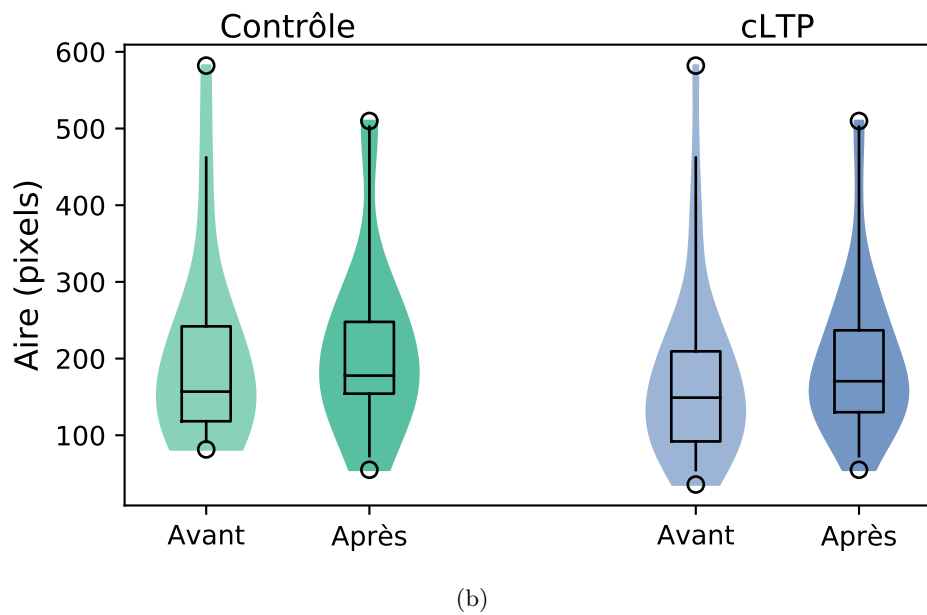
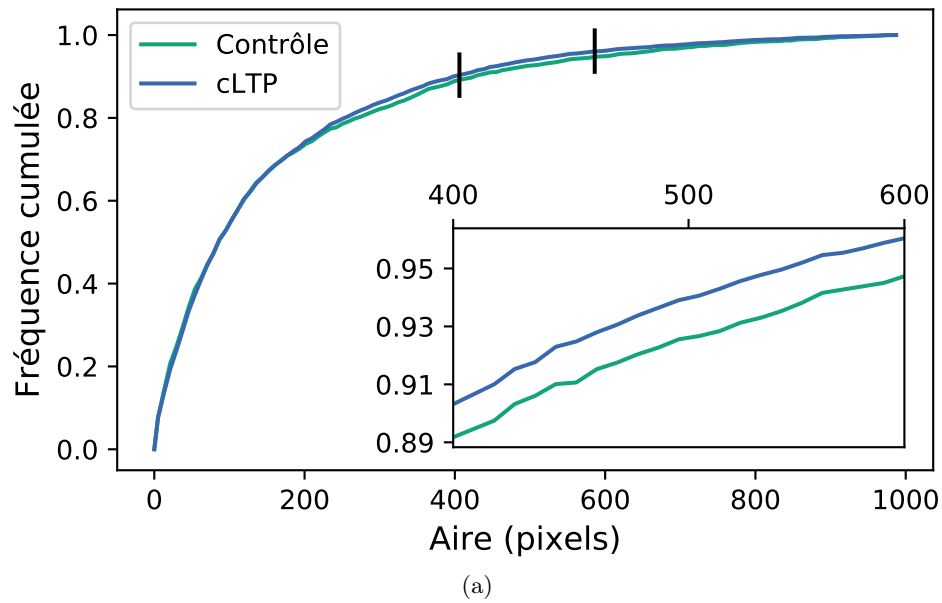


FIGURE 4.4 – Exemple d’analyse de l’aire des mSCTs. (a) Distribution normalisée de l’aire des mSCTs de neurones contrôles et de neurones traités chimiquement pour induire la potentialisation à long terme (cLTP). L’accent est mis à l’endroit où la différence est maximale entre les deux conditions. (b) Distribution de l’aire des mSCTs, pour des neurones contrôle et cLTP, avant et après le traitement.

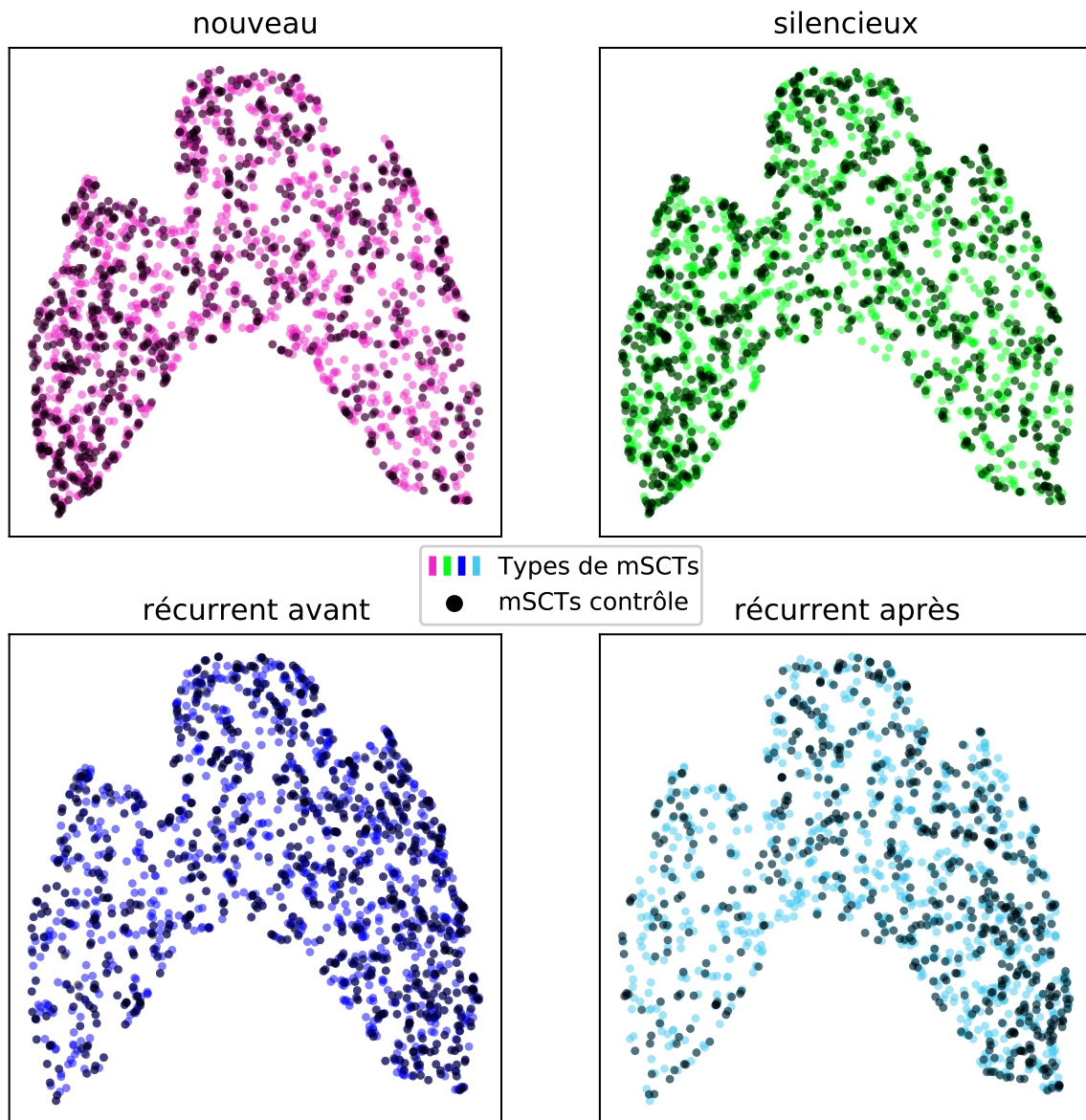


FIGURE 4.5 – Exemple d’application d’analyse non-supervisé pour la réduction de dimensionnalité avec l’algorithme *UMAP*. Cinq caractéristiques morphologiques des mSCTs sont réduites vers un espace à deux dimensions. Les mSCTs provenant des neurones contrôles sont représentés en noir. Les caractéristiques morphologiques des mSCTs sont : l’aire, l’aire projetée, l’amplitude, la longueur de l’axe majeur et la longueur de l’axe mineur.

Conclusion

Ce mémoire propose une méthode d'analyse quantitative à grande échelle des vidéos d'imagerie calcique limitant la variabilité induite par les interventions humaines pour obtenir des données probantes, dans le but, lors de futurs travaux, de caractériser l'impact des mSCTs sur la plasticité synaptique. La méthode est basée sur l'utilisation des modèles de réseaux de neurones artificiels à convolution pour effectuer le traitement des vidéos. Le modèle est intégré dans une routine tout-en-un permettant de réaliser une analyse pour quantifier l'effet des mSCTs sur la plasticité synaptique.

Le premier objectif du projet présenté dans l'introduction consiste à : **1) Adapter la procédure d'entraînement des modèles d'apprentissage profond aux contraintes spécifiques de l'imagerie calcique des mSCTs.** Les chapitres 2 et 3 présentent le travail effectué pour répondre à cet objectif. Ainsi, la contribution 1 définit une méthodologie afin d'*appliquer une méthode inspirée de l'apprentissage PU pour entraîner des réseaux de neurones pleinement convolutionnels (FCN)*. Cette méthodologie d'entraînement permet de tirer avantage des contraintes associées aux mSCTs, soit : i) la petite taille des mSCTs dans le champ de vue ; et ii) la présence clairsemée des mSCTs dans un vidéo. Ces deux contraintes peuvent être combinées et identifiées à un problème commun de l'apprentissage machine, à savoir le déséquilibre des classes (Japkowicz et Stephen, 2002). Ceci implique un déséquilibre élevé entre la classe positive (un échantillon P contenant un mSCT) et la classe négative (tout le reste). Dans ce contexte, on peut extraire des échantillons non étiquetés (U) avec une faible probabilité de contenir un mSCT, pour améliorer les performances des FCN lors de l'entraînement. Non seulement l'ajout d'échantillons U, traités comme des négatifs, améliore les performances, mais cela permet également de dépasser les performances lorsqu'une plus grande quantité d'échantillons P est utilisée. À titre d'exemple, un entraînement avec 7800 échantillons P mène à des performances moindres qu'effectuer un entraînement avec seulement 2600 échantillons P auxquels 5200 échantillons U y ont été ajoutés. Additionnellement, cette méthodologie réduit le nombre d'échantillons à étiqueter (les échantillons P) sans aucune perte de performance. Suivant l'exemple précédent, le nombre d'échantillons P est réduit de 7800 à 2600, soit une réduction de 66,7% du nombre d'échantillons à étiqueter. C'est un avantage crucial puisqu'avec des données biomédicales, les échantillons P sont coûteux à obtenir (coût humain, monétaire, éthique) contrairement aux échantillons U, qui sont des sous-produits de

l’acquisition des échantillons P. Par contre, tirer avantage des échantillons U tel que présenté dans ce travail est dépendant de la nature déséquilibrée des données. Pour traiter les échantillons U comme des négatifs, ils doivent avoir une faible probabilité de contenir un échantillon P – l’instance à segmenter (Elkan et Noto, 2008). De plus, à la lumière des expériences empiriques effectuées, il semble qu’ajouter des échantillons U lorsqu’il n’y a pas suffisamment d’échantillons P détériore les performances du FCN. C’est un résultat tout de même attendu, puisqu’avec peu de P, l’ajout des U dilue le signal d’apprentissage au réseau à détecter les mSCTs. À notre connaissance, c’est la première application d’une méthode d’entraînement de réseaux profonds utilisant l’apprentissage PU sur des données d’imagerie calcique.

Complémentairement à la première contribution du premier objectif, la contribution 2 de ce travail montre comment les réseaux FCN combinés à l’apprentissage PU permettent de *valoriser les segmentations bruitées d’une méthode de seuillage semi-automatique pour obtenir des segmentations de meilleure qualité*. Les segmentations sont bruitées pour deux raisons. Premièrement, le processus de sélection des mSCTs à segmenter dépend de la variation de fluorescence minimale à respecter. Deuxièmement, lorsqu’un mSCT est sélectionné, la segmentation est délimitée par la variation d’intensité locale dans le vidéo sans notion *a priori* sur la structure dendritique sous-jacente. Cela engendre des segmentations grossières, particulièrement au niveau de leurs bordures. Ainsi, compte tenu du pouvoir d’extraction de caractéristiques haut-niveau des FCN (LeCun *et al.*, 2015), entraîner un FCN avec ces étiquettes bruitées lui permet d’apprendre des caractéristiques plus abstraites que la variation d’intensité utilisée pour générer les étiquettes bruitées. Le FCN produit maintenant des segmentations, qualitativement et quantitativement, plus semblables à celle d’un expert humain. En fait, les segmentations du FCN se situent dans la marge de la variabilité des segmentations présentes entre des experts, c’est-à-dire l’inter-variabilité des experts.

Ensuite, l’ajout d’échantillons U joue un rôle régularisateur pour les caractéristiques apprises par le FCN. En effet, avec seulement des échantillons P contenant une variation de fluorescence, cela biaise le réseau à apprendre des caractéristiques basées uniquement sur la variation d’intensité. Cela conduit à une détection d’autres sources de bruits causant des variations d’intensité (p. ex. des variations locales d’intensité sont parfois présentes sur le dendrite). Ajouter des échantillons U situés aléatoirement sur l’avant-plan ajoute des exemples du bruit présent sur le dendrite, ce qui permet au réseau d’apprendre des caractéristiques moins basées sur l’intensité. De plus, le FCN ainsi entraîné peut détecter une forte population de mSCTs de faible intensité, autrement difficilement détectable par les méthodes classiques semi-automatiques (Mancini *et al.*, 2018). Leur détection est possible grâce aux caractéristiques plus générales apprises par le FCN. Cette contribution s’intègre bien dans le premier objectif, puisque la valorisation des segmentations de la méthode de seuillage pour générer les étiquettes d’entraînement du FCN diminue considérablement la tâche d’annotation, limitant ainsi le biais humain introduit dans la génération des étiquettes. L’expert humain doit seulement valider les mSCTs extraits par

la méthode de seuillage, la segmentation au pixel près n'est pas requise.

La **contribution 3** connexe au premier objectif est *la création de deux modèles FCN pour segmenter les mSCTs présents dans l'imagerie calcique de dendrite*. Au-delà de la méthode d'entraînement proposée, les FCN obtenus et caractérisés sont des outils forts utiles pour la communauté neuroscientifique intéressée à la détection des mSCTs. Deux types de FCN sont proposés : le U-Net 2D et le U-Net 3D. Le U-Net 3D a des performances en segmentation et en détection significativement supérieures au U-Net 2D et ses segmentations sont comparables à l'inter-variabilité des experts. Les deux modèles FCN offrent des performances en détection supérieures aux méthodes de seuillage de référence *MiniFinder* et *SICT* (Mancini *et al.*, 2018). Les deux FCN offrent des possibilités uniques d'analyse des vidéos d'imagerie calcique des mSCTs. Afin d'inférer un masque de segmentation, le U-Net 3D utilise au minimum 32 acquisitions – le tenseur en entrée doit être de dimensions $32 \times 32 \times 32$, où la première dimension est temporelle. Ainsi, son utilisation pour des applications temps réel est limitée. Par exemple, avec une période d'acquisition de 150 ms, le U-Net 3D introduit un décalage de 4,8 s ($32 * 150$ ms) nettement plus long que la durée d'un mSCT qui est d'environ 2 à 3 s. C'est là que le U-Net 2D est avantageux à utiliser, puisqu'il utilise cinq acquisitions $[t_{-2}, t_{-1}, t_0, t_1, t_2]$ pour effectuer une inférence au temps t_0 . Donc, dans un paradigme temps réel, un décalage de deux temps d'acquisition est nécessaire pour imager les temps t_1 et t_2 . Avec une période d'acquisition de 150 ms, le U-Net 2D introduit ainsi un décalage total de 300 ms, ce qui lui confère des performances suffisantes pour une application temps réel d'imagerie calcique de mSCTs. Par rapport au U-Net 3D, le U-Net 2D détecte 18 % moins de mSCTs (la majorité de ces mSCTs manqués sont de faible intensité), c'est le compromis à faire pour une détection temps réel. Pour des analyses hors-ligne, le U-Net 3D offre les meilleures performances.

Le deuxième objectif consiste à : **2) développer une routine d'analyse automatique tout-en-un pour l'analyse à grande échelle d'imagerie calcique de mSCTs**. Le chapitre 4 répond à cet objectif et présente la routine développée et ses fonctionnalités. La **contribution 4** de ce mémoire constitue *le développement d'une application tout-en-un qui permet la détection, la segmentation, la visualisation et la caractérisation à grande échelle et automatique des mSCTs*. Des outils semblables sont disponibles pour la caractérisation de l'activité neuronale évoquée (Pachitariu *et al.*, 2017; Giovannucci *et al.*, 2019). Par contre, des écarts comme la petite taille des mSCTs et leur faible rapport signal sur bruit de leur intensité de fluorescence font que ces outils ne sont pas adaptés à l'analyse des vidéos d'imagerie calcique des mSCTs. Au coeur de la routine développée, le FCN présenté dans ce mémoire est utilisé pour détecter et segmenter les mSCTs dans les vidéos. Cette étape ne nécessite *aucune* intervention humaine et utilise des accélérations graphiques pour accélérer le traitement. Une interface intuitive permet de naviguer rapidement dans les vidéos pour inspecter les mSCTs détectés, leur segmentation et leur signal de fluorescence. À partir des mSCTs détectés, des métriques sont automatiquement extraites de la morphologie du masque de seg-

mentation des mSCTs, de leur récurrence dans un vidéo et de leur intensité de fluorescence. Toutes ces métriques, les vidéos et les segmentations sont sauvegardées dans une base de données pour faciliter leur accès. Avec l'outil développé, des milliers de vidéos ont pu être analysés en quelques jours seulement, contrairement aux mois d'analyse nécessaires avec les méthodes semi-automatiques de l'état de l'art, *MiniFinder* et *SICT* (Mancini *et al.*, 2018). Ainsi, la routine développée est un outil indispensable à la communauté neuroscientifique qui cherche à réaliser des analyses quantitatives de la plasticité synaptique causée par l'activité neuronale.

Les contributions de ce mémoire offrent les outils nécessaires au développement d'une analyse quantitative à grande échelle limitant la variabilité humaine pour obtenir des données probantes, dans le but de caractériser l'impact des mSCTs sur la plasticité synaptique. Cela étant dit, un autre aspect intéressant de ce travail est les prochaines avancées qu'offrent les contributions développées. Notamment, la procédure d'apprentissage PU pourrait être adaptée à d'autres domaines ayant des caractéristiques semblables aux vidéos d'imagerie calcique de mSCTs. Par exemple, l'imagerie par microscopie électronique de tissus biologiques permet une résolution jusqu'aux dizaines de picomètres, ce qui génère des images de très grande taille dans lesquelles des organelles d'intérêts peuvent représenter un faible pourcentage des pixels de l'image (Goodhew et Humphreys, 2000). Une autre avenue intéressante serait l'utilisation de la méthode de détection automatique et rapide des mSCTs pour permettre une microscopie multimodale intelligente. Par exemple, un microscope possédant deux modalités d'imagerie : une acquisition grand champ et une acquisition super-résolution par balayage (p. ex. STED). Même si la microscopie super-résolution STED est destructive pour l'échantillon biologique, son imagerie par balayage permet l'acquisition de zones précises dans le plan focal. Avec le flux vidéo grand champ, le U-Net 2D est utilisé en temps réel pour repérer des synapses actives. Lorsqu'une synapse active est repérée, une image super-résolution de cette synapse est prise. Ainsi, il serait possible d'obtenir une séquence temporelle d'images super-résolues de synapses actives *in vitro*. Ce sont des données indispensables, et encore inexistantes, pour comprendre le réarrangement des protéines au coeur des synapses.

Ce travail démontre indubitablement que c'est avec des travaux multidisciplinaires, combinant le savoir de sciences telles que l'informatique, la biochimie et la physique optique, que des progrès seront possibles pour découvrir les interactions fondamentales à la base de l'intelligence humaine.

Annexe A

Annexe

A.1 Types d'événements calciques pour la segmentation précise

Pour avoir une représentation non biaisée de la segmentation d'événements calciques faite par les experts, les mSCTs ont été séparés en 8 catégories. Les événements de chaque catégorie se distinguent, soit par leur apparence ou par leur intensité. Il y a au total 78 régions segmentées, voici la liste :

- Présent sur une épine dendritique (x10)
- Présent sur l'arbre dendritique (x14)
- Petit mSCT \implies autour de 15 par 15 pixels (x13)
- Gros mSCT \implies autour de 100 par 100 pixels (x7)
- Hors du plan focal du microscope \implies généralement de faible intensité (x10)
- mSCT long \implies fluorescence qui se propage sur l'arbre dendritique (x8)
- mSCT de forte intensité de fluorescence par rapport au signal basal (x9)
- mSCT de faible intensité de fluorescence par rapport au signal basal (x7)

A.2 Liste des métadonnées

Voici la liste des métadonnées ainsi qu'une brève description :

- *Neuron ID* : Identifiant unique du neurone.
- *Coverslip ID* : Identifiant unique de la lamelle de culture de neurones.
- *Date* : Date d'imagerie.
- *Stream name* : Nom du fichier du vidéo.
- *Folder path* : Chemin vers le dossier contenant le vidéo.

Neuron ID	Coverslip ID	Date	Stream name	Folder Path	Imaging Period	Experimentation	Stream condition
1	1	2017-03-23	Stream1.tif	NAS:/path/to/the/folder	150	Control	Baseline 1
			Stream2.tif				Baseline 2
			Stream3.tif				5 min
			Stream4.tif				10 min
			Stream5.tif				15 min
2	1	2017-03-23	Stream1.tif	NAS:/path/to/the/folder	150	cLTP	Baseline 1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

FIGURE A.1 – Exemple de feuille Google pour sauvegarder les métadonnées. Chaque ligne est un vidéo et chaque colonne correspond à un type de métadonnée.

- *Imaging period* : Période d'échantillonnage d'un plan d'imagerie (environ 150ms).
- *Experimentation* : Une expérimentation comporte plusieurs vidéos et permet de tester un traitement.
 - *Control* : Expérimentation sans traitement.
 - *cLTP* : Traitement chimique pour induire la potentialisation à long terme (Lu *et al.*, 2001).
 - *APV* : Traitement chimique pour bloquer les récepteurs NMDA (Kleinschmidt *et al.*, 1987).
 - *High Magnesium* : Traitement pour bloquer les récepteurs NMDA (Mayer *et al.*, 1984).
 - *Synaptic Scaling* : Traitement pour induire la plasticité homéostatique (Turrigiano *et al.*, 1998).
 - *Ifenprodyl* : Traitement pour bloquer des sous-unités des récepteurs NMDA (Williams, 2001).
 - *Different Glycine Concentrations* : Pour moduler la concentration des récepteurs NMDA (Papouin *et al.*, 2012).
- *Stream condition* : Pendant une expérimentation, différentes conditions sont imagées.
 - *Baseline 1* : Premier vidéo de référence (aucun traitement).
 - *Baseline 2* : Deuxième vidéo de référence (aucun traitement).
 - *X min* : Vidéo X minutes après l'application du traitement.

A.3 Fichier de sauvegarde des résultats

Voici la liste des quelques caractéristiques importantes extraites pour caractériser les mSCTs ainsi qu'une brève description. Un vidéo a une dimension de $T \times H \times W$ où T est le nombre d'images, H et W le nombre de pixels en hauteur et en largeur. N est le nombre de mSCTs et M le nombre de régions. Une région est un endroit sur le neurone où surviennent un ou plusieurs mSCTs.

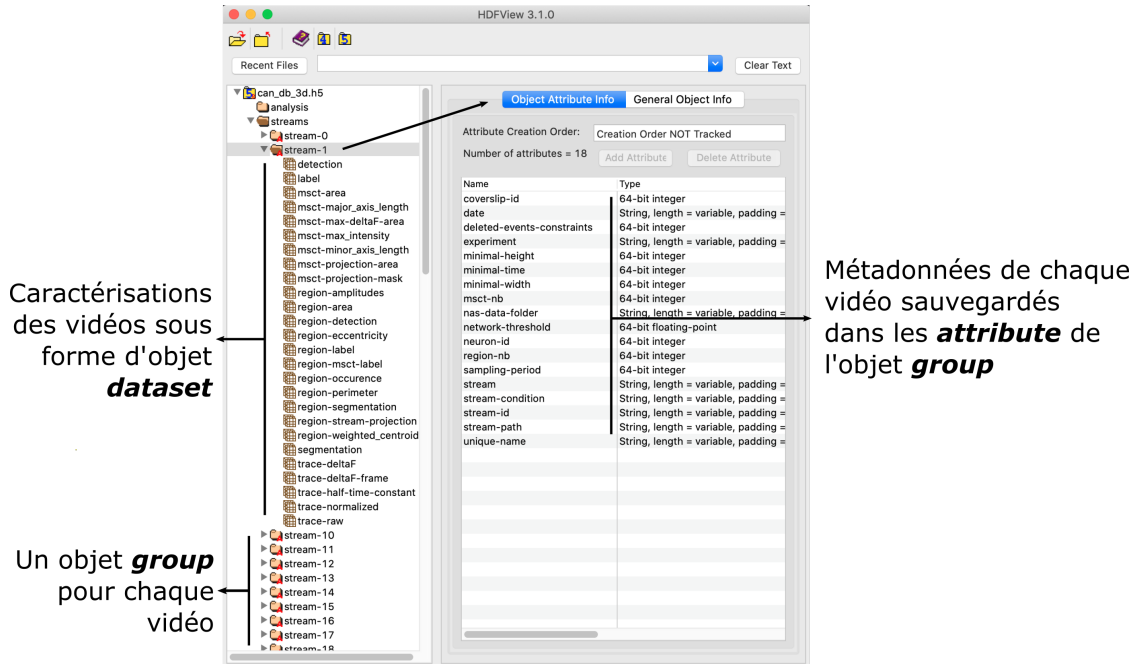


FIGURE A.2 – *HDFView* : Logiciel d’inspection des fichiers HDF. Les objets *group* pour chaque vidéo permettent d’entreposer tous les résultats de caractérisations des mSCTs. Les caractérisations des vidéos sont sauvegardées sous forme d’un objet *dataset* imitant un tableau NumPy. Le panel de droite comporte les *attribute* contenus pour le *group* nommé *stream-1*. Les métadonnées sont sauvegardées dans les *attribute* de l’objet.

- *segmentation* : Segmentation dense du vidéo. Chaque pixel a une valeur de 0.. N , où 0 est un pixel appartenant à l’arrière-plan, et N , un pixel appartenant au N^{e} mSCT. Dimension : $[T \times H \times W]$.
- *detection* : Centroides des mSCTs détectés dans le vidéo. Dimension : $[N \times 3]$.
- *msct-X* : Caractéristiques morphologiques des mSCTs. Par exemple, *msct-area* correspond à l’aire des mSCTs. Dimension : $[N \times 1]$
- *region-X* : Caractéristiques morphologiques des régions. Dimension : $[M \times 1]$
- *trace-X* : Toutes les caractéristiques relatives à la trace calcique de fluorescence extraite pour chaque mSCT. Dimension : $[N \times T]$.

A.4 Interface de visualisation

L’interface de visualisation comporte deux fenêtres principales : une pour télécharger le fichier HDF5 et sélectionner le vidéo à inspecter et une fenêtre consacrée à la visualisation et à l’annotation.

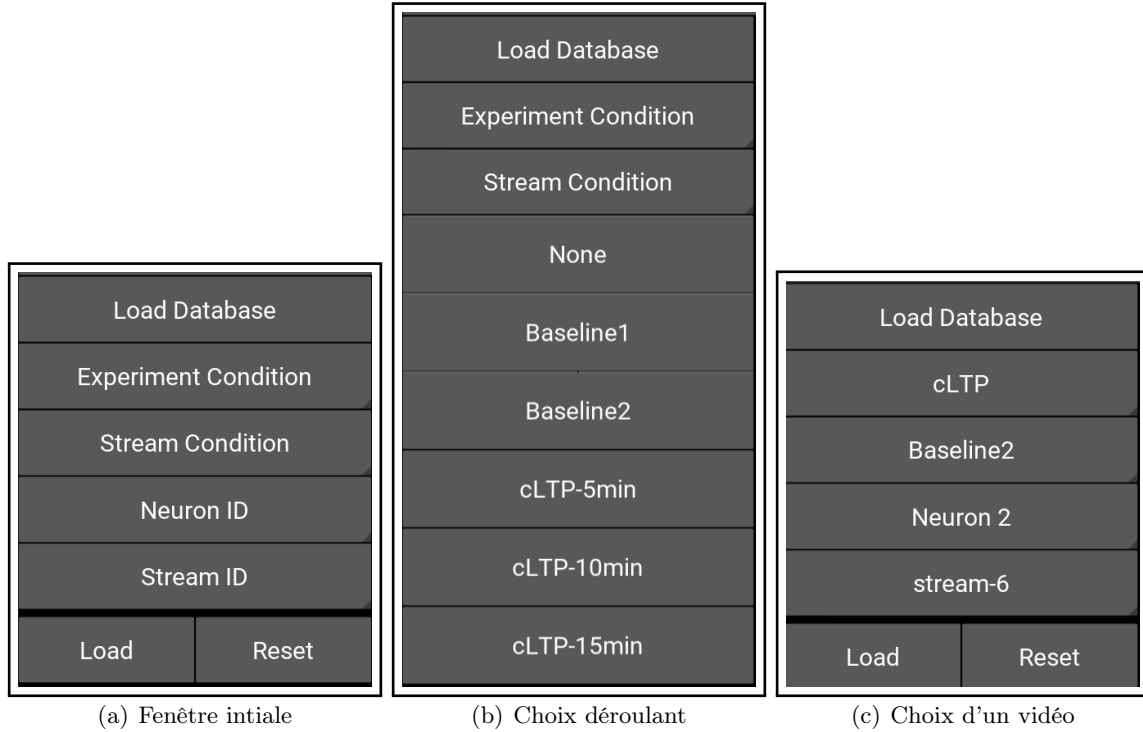


FIGURE A.3 – Téléchargement de la base de données et choix du vidéo.

A.4.1 Sélection du vidéo

L'interface de sélection de vidéo (fig. A.5(a)) permet de sélectionner la base de données (*Load Database*) et de choisir un vidéo à inspecter. Il est possible de filtrer les vidéos selon ses métadonnées : *Experiment Condition*, *Stream Condition*, *Neuron ID*, *Stream ID*. Lors d'un clic sur l'une des métadonnées (fig. A.5(b)), un menu déroulant apparaît avec tous les choix possibles dans la base de données. Quand un vidéo est choisi (fig. A.5(a)), les métadonnées pour le vidéo choisi sont affichées. En appuyant sur le bouton *Load* le téléchargement du vidéo est effectué la fenêtre de visualisation est initialisée.

A.4.2 Visualisation

La panneau complet de visualisation est présenté à la figure 4.1. Chaque sous-section de ce panneau est présentée ici.

La visualisation du champ de vue du vidéo est synchronisée avec le mSCT inspecté (fig. A.4). La zone centrée sur l'événement calcique est visible dans le champ de vue par le biais de l'encadré jaune. Des coupes temporelles, verticale et horizontale, extraites à l'endroit où se trouvent les lignes jaunes centrées sur l'événement.

Le contraste de ces vues est modifiable avec les deux boutons coulissants du haut dans la figure

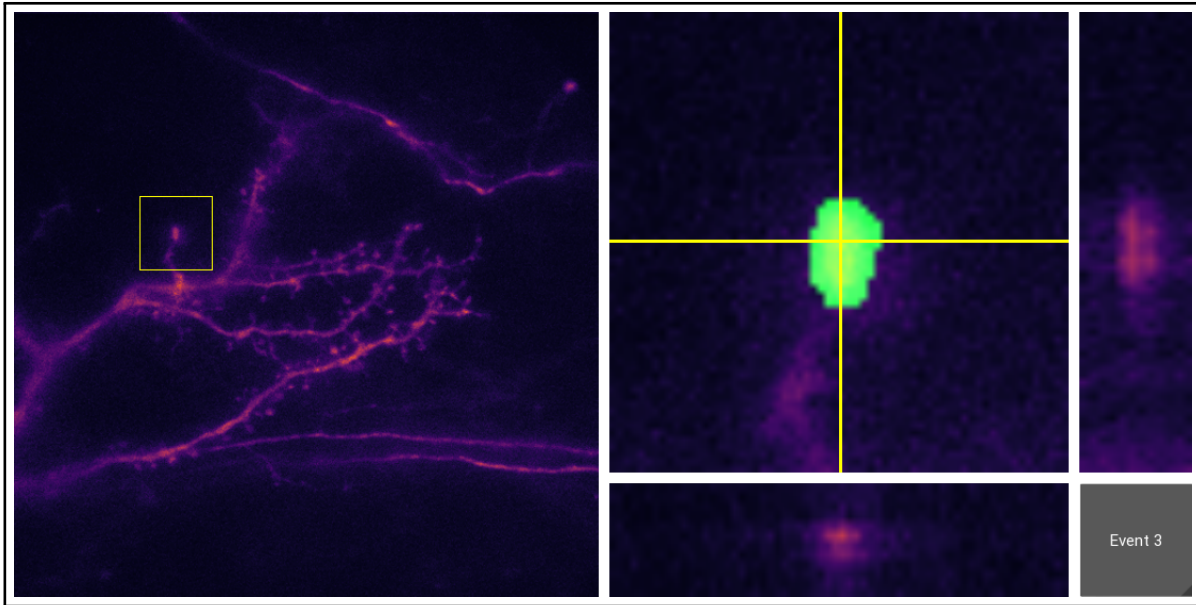
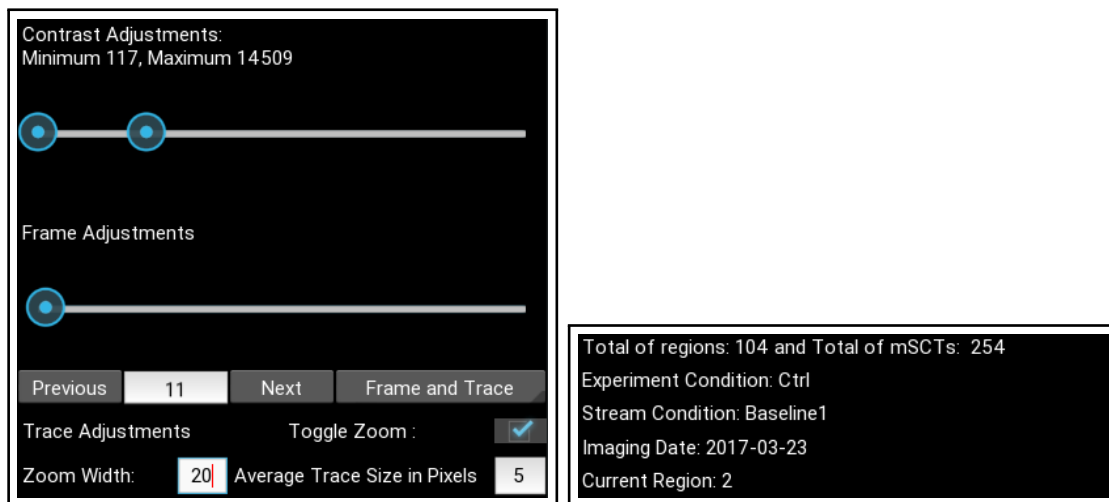


FIGURE A.4 – Explorateur visuel des vidéos. (*Gauche*) : le champ de vision complet du vidéo. (*Droite*) : cadre de la section jaune dans l’image de gauche. Les droites horizontale et verticale représentent une tranche dans le temps. La segmentation du mSCT sélectionné (*Event 3*) est recouverte en vert sur la section cadrée.

A.5(a). Le deuxième bouton coulissant permet de changer l’image courante dans le vidéo, donc de se déplacer dans le temps. La section *Trace Adjustments* permet d’ajuster les paramètres de calcul et de visualisation de la trace calculée de fluorescence : *Toggle Zoom* permet d’afficher



(a) Contrôle du visuel du vidéo

(b) Affichage statique d’informations sur le vidéo et le mSCT couramment affiché

FIGURE A.5 – Panneaux pour contrôler le visuel et afficher de l’information.

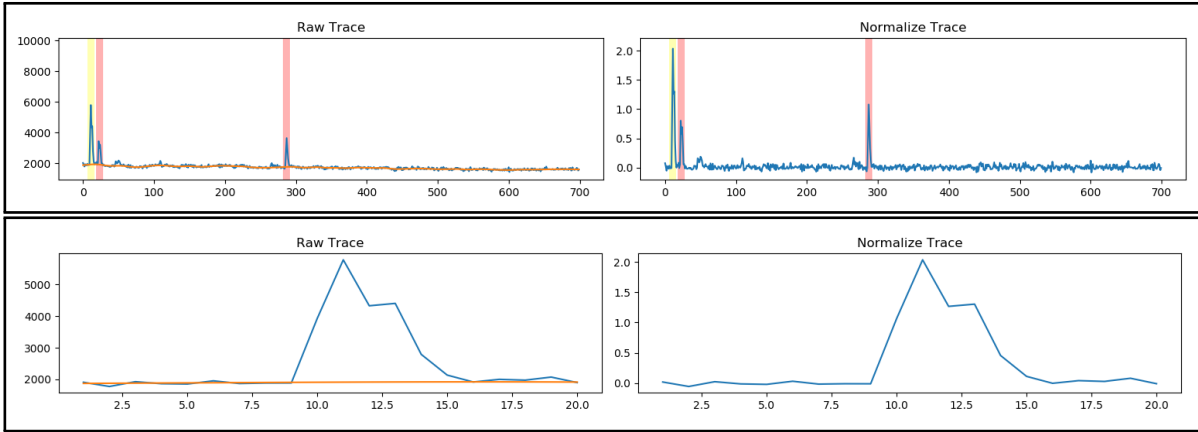


FIGURE A.6 – Visualisation des traces de fluorescence. Les traces calciques de fluorescence sont calculées à l’intersection de la croix jaune dans la figure A.4. (*Gauche*) Trace de la fluorescence calcique sans modification. (*Droite*) Trace normalisée. La section jaune représente la section du mSCT affichée et les sections rouges représentent les mSCTs présents sur cette même région. La trace du bas est celle obtenue lorsque l’option *Toggle Zoom* est activée.

un zoom de la trace, *Zoom Width* définit la largeur du zoom et *Average Trace Size in Pixels* définit la grosseur du noyau utilisé pour moyenner le calcul de la trace. Les métadonnées associées au vidéo et au mSCT courant sont présentées à l’usager (fig. A.5(b)).

Finalement, la figure A.6 permet de visualiser la trace de fluorescence calcique sans modification et celle normalisée. La section jaune représente le mSCT courant et les rouges présentent les mSCTs dans cette même région. La trace calcique peut être zoomée en appuyant sur le bouton *Toggle Zoom*

Quelques raccourcis clavier sont aussi disponibles pour faciliter la navigation :

1. *Touche a* a : Valider le mSCT et l’ajouter.
2. *Touche d* d : Invalider le mSCT et le supprimer.
3. *Flèche Gauche* ← : région précédente.
4. *Flèche Droite* → : région suivante.
5. *Flèche Bas* ↓ : mSCT précédent.
6. *Flèche Haut* ↑ : mSCT suivant.
7. *Touche d’échappement* ESC : quitter.

Bibliography

- ABBE, E. (1873). Beiträge zur theorie des mikroskops und der mikroskopischen wahrnehmung. *Archiv für mikroskopische Anatomie*, 9(1):413–468.
- ABBOTT, A. (2008). Hidden treasures: The Cajal collection in Madrid. *Nature*, 452(7190):940–940.
- ABBOTT, L. F. et NELSON, S. B. (2000). Synaptic plasticity: Taming the beast. *Nature Neuroscience*, 3(11):1178–1183.
- ALPAYDIN, E. (2020). *Introduction to machine learning*. MIT press.
- ANDERSON, K., ALEXOV, A., BAEHREN, L., GRIESSMEIER, J.-M., WISE, M. et RENTING, A. (2010). LOFAR and HDF5: Toward a new radio data standard. *ArXiv e-print 1012.2266*, <https://arxiv.org/abs/1012.2266>.
- APTHORPE, N. J., RIORDAN, A. J., AGUILAR, R. E., HOMANN, J., GU, Y., TANK, D. W. et SEUNG, H. S. (2016). Automatic neuron detection in calcium imaging data using convolutional networks. *ArXiv e-print 1606.07372*, <https://arxiv.org/abs/1606.07372>.
- AUGUSTINE, G. J., SANTAMARIA, F. et TANAKA, K. (2003). Local calcium signaling in neurons. *Neuron*, 40(2):331–346.
- AXELROD, D. (2001). Total internal reflection fluorescence microscopy in cell biology. *Traffic*, 2(11):764–774.
- BADRINARAYANAN, V., KENDALL, A. et CIPOLLA, R. (2017). SegNet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12):2481–2495.
- BAEK, S.-J., PARK, A., AHN, Y.-J. et CHOO, J. (2015). Baseline correction using asymmetrically reweighted penalized least squares smoothing. *The Analyst*, 140(1):250–257.
- BEKKER, J. et DAVIS, J. (2020). Learning from positive and unlabeled data: A survey. *Machine Learning*, 109(4):719–760.
- BELLMAN, R. (1966). Dynamic programming. *Science*, 153(3731):34–37.

- BUONOMANO, D. V. et MERZENICH, M. M. (1998). Cortical plasticity: From synapses to maps. *Annual review of neuroscience*, 21(1):149–186.
- CAICEDO, J. C., ROTH, J., GOODMAN, A., BECKER, T., KARHOHS, K. W., BROISIN, M., MOLNAR, C., MCQUIN, C., SINGH, S., THEIS, F. J. *et al.* (2019). Evaluation of deep learning strategies for nucleus segmentation in fluorescence images. *Cytometry Part A*, 95(9):952–965.
- CHANG, S., ZHANG, Y., TANG, J., YIN, D., CHANG, Y., HASEGAWA-JOHNSON, M. A. et HUANG, T. S. (2016). Positive-Unlabeled Learning in streaming networks. *In Proc. of the Conference on Knowledge Discovery Data Mining (KDD)*, pages 755–764.
- CHEN, J.-H., BLANPIED, T. A. et TANG, A.-H. (2019). Quantitative analysis of trans-synaptic protein alignment. *BioRxiv e-print 10.1101/573063*, <https://www.biorxiv.org/content/early/2019/03/11/573063>.
- CHEN, T.-W., WARDILL, T. J., SUN, Y., PULVER, S. R., RENNINGER, S. L., BAOHAN, A., SCHREITER, E. R., KERR, R. A., ORGER, M. B., JAYARAMAN, V. *et al.* (2013). Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature*, 499(7458):295–300.
- CHOI, B. J., IMLACH, W. L., JIAO, W., WOLFRAM, V., WU, Y., GRBIC, M., CELA, C., BAINES, R. A., NITABACH, M. N. et MCCABE, B. D. (2014). Miniature neurotransmission regulates drosophila synaptic structural maturation. *Neuron*, 82(3):618–634.
- ÇIÇEK, Ö., ABDULKADIR, A., LIENKAMP, S. S., BROX, T. et RONNEBERGER, O. (2016). 3D U-Net: Learning dense volumetric segmentation from sparse annotation. *In Proc. of the Conference on Medical Computing and Computer Assisted Intervention (MICCAI)*.
- CITRI, A. et MALENKA, R. C. (2008). Synaptic plasticity: Multiple forms, functions, and mechanisms. *Neuropsychopharmacology*, 33(1):18–41.
- CLAESSEN, M., DE SMET, F., GILLARD, P., MATHIEU, C. et DE MOOR, B. (2015). Building classifiers to predict the start of glucose-lowering pharmacotherapy using belgian health expenditure data. *ArXiv e-print 1504.07389*, <https://arxiv.org/abs/1504.07389>.
- COLLINGRIDGE, G. L., ISAAC, J. T. et WANG, Y. T. (2004). Receptor trafficking and synaptic plasticity. *Nature Reviews Neuroscience*, 5(12):952–962.
- CORDTS, M., OMRAN, M., RAMOS, S., REHFELD, T., ENZWEILER, M., BENENSON, R., FRANKE, U., ROTH, S. et SCHIELE, B. (2016). The cityscapes dataset for semantic urban scene understanding. *In Proc. of the International Conference on Computer Vision and Pattern Recognition (CVPR)*.
- DENIS, F. (1998). PAC Learning from positive statistical queries. *In Proc. of the Conference on Algorithmic Learning Theory (ALT)*.

- DENIS, F., GILLERON, R. et LETOUZEY, F. (2005). Learning from positive and unlabeled examples. *Theoretical Computer Science*, 348(1):70–83.
- DENIZ, C. M., XIANG, S., HALLYBURTON, R. S., WELBECK, A., BABB, J. S., HONIG, S., CHO, K. et CHANG, G. (2018). Segmentation of the proximal femur from MR images using deep convolutional neural networks. *Scientific reports*, 8(1):1–14.
- DICE, L. R. (1945). Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302.
- DROZDZAL, M., VORONTSOV, E., CHARTRAND, G., KADOURY, S. et PAL, C. (2016). The importance of skip connections in biomedical image segmentation. In *Proc. of the Conference on Deep Learning in Medical Image Analysis (DLMIA)*.
- DU, P., KIBBE, W. A. et LIN, S. M. (2006). Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching. *Bioinformatics*, 22(17):2059–2065.
- ELKAN, C. et NOTO, K. (2008). Learning classifiers from only positive and unlabeled data. In *Proc. of the Conference on Knowledge Discovery Data Mining (KDD)*.
- FALK, T., MAI, D., BENSCH, R., ÇIÇEK, Ö., ABDULKADIR, A., MARRAKCHI, Y., BÖHM, A., DEUBNER, J., JÄCKEL, Z., SEIWALD, K., DOVZHENKO, A., TIETZ, O., DAL BOSCO, C., WALSH, S., SALTUKOGLU, D., TAY, T. L., PRINZ, M., PALME, K., SIMONS, M., DIESTER, I., BROX, T. et RONNEBERGER, O. (2019). U-Net: Deep learning for cell counting, detection, and morphometry. *Nature Methods*, 16(1):67–70.
- FATT, P. et KATZ, B. (1952). Spontaneous subthreshold activity at motor nerve endings. *The Journal of Physiology*, 117(1):109–128.
- FLEISHMAN, G. M., ZHANG, M., TUSTISON, N. J., ESPINOSA-MEDINA, I., MU, Y., KHAIRY, K. et AHRENS, M. (2019). Deformable registration of whole brain zebrafish microscopy using an implementation of the flash algorithm within ants. In *Proc. of the International Symposium on Biomedical Imaging (ISBI)*.
- FOLK, M., HEBER, G., KOZIOL, Q., POURMAL, E. et ROBINSON, D. (2011). An overview of the HDF5 technology suite and its applications. In *Proc. of the Conference on Extending Database Technology (EDBT/ICDT) Workshop on Array Databases*.
- FRIEDMAN, J., HASTIE, T. et TIBSHIRANI, R. (2009). *The elements of statistical learning: data mining, inference and prediction*. Springer, 2e édition.
- GARCIA-LOPEZ, P., GARCIA-MARIN, V. et FREIRE, M. (2010). The histological slides and drawings of Cajal. *Frontiers in Neuroanatomy*.

- GIOVANNUCCI, A., FRIEDRICH, J., GUNN, P., KALFON, J., BROWN, B. L., KOAY, S. A., TAXIDIS, J., NAJAFI, F., GAUTHIER, J. L., ZHOU, P., KHAKH, B. S., TANK, D. W., CHKLOVSKII, D. B. et PNEVMATIKAKIS, E. A. (2019). Caiman: An open source tool for scalable calcium imaging data analysis. *eLife e-print 8:e38173*, <https://elifesciences.org/articles/38173>.
- GOODFELLOW, I., BENGIO, Y., COURVILLE, A. et BENGIO, Y. (2016). *Deep learning*. MIT press Cambridge.
- GOODHEW, P. J. et HUMPHREYS, J. (2000). *Electron microscopy and analysis*. CRC Press.
- GREENBAUM, L., ROTHMANN, C., LAVIE, R. et MALIK, Z. (2000). Green fluorescent protein photobleaching: A model for protein damage by endogenous and exogenous singlet oxygen. *Biological Chemistry*, 381(12):1251–1258.
- GREENOUGH, W. T., BLACK, J. E. et WALLACE, C. S. (2002). *Experience and brain development*. Blackwell Publishing.
- GRIENBERGER, C. et KONNERTH, A. (2012). Imaging calcium in neurons. *Neuron*, 73(5):862–885.
- GRYNKIEWICZ, G., POENIE, M. et TSIEN, R. Y. (1985). A new generation of Ca²⁺ indicators with greatly improved fluorescence properties. *Journal of Biological Chemistry*, 260(6):3440–3450.
- GUIZAR-SICAIROS, M., THURMAN, S. T. et FIENUP, J. R. (2008). Efficient subpixel image registration algorithms. *Optics Letters*, 33(2):156–158.
- GYGI, S. P., RIST, B., GERBER, S. A., TURECEK, F., GELB, M. H. et AEBERSOLD, R. (1999). Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature Biotechnology*, 17(10):994–999.
- HAIDER, S. A., CAMERON, A., SIVA, P., LUI, D., SHAFIEE, M. J., BORROOMAND, A., HAIDER, N. et WONG, A. (2016). Fluorescence microscopy image noise reduction using a stochastically-connected random field model. *Scientific Reports*, 6(1):20640.
- HANAWA, H., HEMATTI, P., KEYVANFAR, K., METZGER, M. E., KROUSE, A., DONAHUE, R. E., KEPES, S., GRAY, J., DUNBAR, C. E., PERSONS, D. A. et NIENHUIS, A. W. (2004). Efficient gene transfer into rhesus repopulating hematopoietic stem cells using a simian immunodeficiency virus-based lentiviral vector system. *Blood*, 103(11):4062–4069.
- HAQUE, I. R. I. et NEUBERT, J. (2020). Deep learning approaches to biomedical image segmentation. *Informatics in Medicine Unlocked*, 18(1):100297.
- HART, P. E., STORK, D. G. et DUDA, R. O. (2000). *Pattern classification*. Wiley New York.

- HE, K., ZHANG, X., REN, S. et SUN, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification. *In Proc. of the International Conference on Computer Vision (ICCV)*.
- HE, K., ZHANG, X., REN, S. et SUN, J. (2016). Deep residual learning for image recognition. *In Proc. of the International Conference on Computer Vision and Pattern Recognition (CVPR)*.
- HEBB, D. O. (2005). *The Organization of Behavior: A Neuropsychological Theory*. Psychology Press.
- HELASSA, N., ZHANG, X.-h., CONTE, I., SCARINGI, J., ESPOSITO, E., BRADLEY, J., CARTER, T., OGDEN, D., MORAD, M. et TÖRÖK, K. (2015). Fast-Response Calmodulin-based fluorescent indicators reveal rapid intracellular calcium dynamics. *Scientific Reports*, 5(1):1–15.
- HELL, S. W. et WICHMANN, J. (1994). Breaking the diffraction resolution limit by stimulated emission: Stimulated-emission-depletion fluorescence microscopy. *Optics Letters*, 19(11):780–782.
- HOFFMAN, D. P., SLAVITT, I. et FITZPATRICK, C. A. (2021). The promise and peril of deep learning in microscopy. *Nature Methods*, 18(2):131–132.
- HUANG, B., BATES, M. et ZHUANG, X. (2009). Super-resolution fluorescence microscopy. *Annual Review of Biochemistry*, 78:993–1016.
- HUANG, G., LIU, Z., VAN DER MAATEN, L. et WEINBERGER, K. Q. (2017). Densely connected convolutional networks. *In Proc. of the International Conference on Computer Vision and Pattern Recognition (CVPR)*.
- HUANG, L.-K. et WANG, M.-J. J. (1995). Image thresholding by minimizing the measures of fuzziness. *Pattern recognition*, 28(1):41–51.
- HUMEAU, Y. et CHOQUET, D. (2019). The next generation of approaches to investigate the link between synaptic plasticity and learning. *Nature Neuroscience*, 22(10):1536–1543.
- IBTEHAZ, N. et RAHMAN, M. S. (2020). MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation. *Neural Networks*, 121(1):74–87.
- ILLASTE, A., WULLSCHLEGER, M., FERNANDEZ-TENORIO, M., NIGGLI, E. et EGGER, M. (2019). Automatic detection and classification of Ca²⁺ release events in line- and frame-scan images. *Biophysical Journal*, 116(3):383–394.
- IOFFE, S. et SZEGEDY, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *In Proc. of the International Conference on Machine Learning (ICML)*.

- ISOLA, P., ZHU, J.-Y., ZHOU, T. et EFROS, A. A. (2017). Image-to-image translation with conditional adversarial networks. *In Proc. of the International Conference on Computer Vision and Pattern Recognition (CVPR)*.
- JACCARD, P. (1901). Distribution de la flore alpine dans le bassin des Dranses et dans quelques régions voisines. *Bull Soc Vaudoise Sci Nat*, 37(1):241–272.
- JAPKOWICZ, N. et STEPHEN, S. (2002). The class imbalance problem: A systematic study. *Intelligent Data Analysis*, 6(5):429–449.
- JÉGOU, S., DROZDZAL, M., VAZQUEZ, D., ROMERO, A. et BENGIO, Y. (2017). The one hundred layers tiramisu: Fully convolutional DenseNets for semantic segmentation. *In Proc. of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*.
- KAMPPFMEYER, M., SALBERG, A.-B. et JENSSEN, R. (2016). Semantic segmentation of small objects and modeling of uncertainty in urban remote sensing images using deep convolutional neural networks. *In Proc. of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*.
- KANDEL, E. R., SCHWARTZ, J. H., JESSELL, T. M., SIEGELBAUM, S., HUDSPETH, A. J. et MACK, S. (2000). *Principles of neural science*, volume 4. McGraw-hill New York.
- KANOPOULOS, N., VASANTHAVADA, N. et BAKER, R. L. (1988). Design of an image edge detection filter using the Sobel operator. *IEEE Journal of Solid-state Circuits*, 23(2):358–367.
- KARAMAN, M. W., HERRGARD, S., TREIBER, D. K., GALLANT, P., ATTERIDGE, C. E., CAMPBELL, B. T., CHAN, K. W., CICERI, P., DAVIS, M. I., EDEEN, P. T., FARAONI, R., FLOYD, M., HUNT, J. P., LOCKHART, D. J., MILANOV, Z. V., MORRISON, M. J., PALLARES, G., PATEL, H. K., PRITCHARD, S., WODICKA, L. M. et ZARRINKAR, P. P. (2008). A quantitative analysis of kinase inhibitor selectivity. *Nature Biotechnology*, 26(1):127–132.
- KAVALALI, E. T. (2015). The mechanisms and functions of spontaneous neurotransmitter release. *Nature Reviews Neuroscience*, 16(1):5–16.
- KINGMA, D. P. et BA, J. (2017). Adam: A method for stochastic optimization. *ArXiv e-print 1412.6980*, <https://arxiv.org/abs/1412.6980>.
- KLEINSCHMIDT, A., BEAR, M. F. et SINGER, W. (1987). Blockade of NMDA receptors disrupts experience-dependent plasticity of kitten striate cortex. *Science*, 238(4825):355–358.
- KLIBISZ, A., ROSE, D., EICHOLTZ, M., BLUNDON, J. et ZAKHARENKO, S. (2017). Fast, simple calcium imaging segmentation with fully convolutional networks. *In Proc. of the Conference on Deep Learning in Medical Image Analysis (DLMIA)*.

- KOHAVI, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *In Proc. of the International Joint Conference on Artificial Intelligence (IJCAI)*.
- KORANNE, S. (2011). Hierarchical Data Format 5: HDF5. *In Handbook of Open Source Tools*, pages 191–200. Springer US.
- KRIZHEVSKY, A., SUTSKEVER, I. et HINTON, G. E. (2012). ImageNet classification with deep convolutional neural networks. *In Proc. of the Conference on Advances in Neural Information Processing Systems (NIPS)*, volume 25.
- KUHN, H. W. (1955). The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2(1-2):83–97.
- LECUN, Y., BENGIO, Y. et HINTON, G. (2015). Deep learning. *Nature*, 521(7553):436–444.
- LECUN, Y., BOTTOU, L., BENGIO, Y. et HAFFNER, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324.
- LI, C. H. et TAM, P. K. S. (1998). An iterative algorithm for minimum cross entropy thresholding. *Pattern Recognition Letters*, 19(8):771–776.
- LICHTMAN, J. W. et CONCHELLO, J.-A. (2005). Fluorescence microscopy. *Nature Methods*, 2(12):910–919.
- LINGURARU, M. G., PURA, J. A., PAMULAPATI, V. et SUMMERS, R. M. (2012). Statistical 4D graphs for multi-organ abdominal segmentation from multiphase CT. *Medical Image Analysis*, 16(4):904–914.
- LITJENS, G., KOOI, T., BEJNORDI, B. E., SETIO, A. A. A., CIOMPI, F., GHAFOORIAN, M., VAN DER LAAK, J. A. W. M., VAN GINNEKEN, B. et SÁNCHEZ, C. I. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42(1):60–88.
- LIU, B., DAI, Y., LI, X., LEE, W. S. et YU, P. S. (2003). Building text classifiers using positive and unlabeled examples. *In Proc. of the International Conference on Data Mining (ICDM)*, pages 179–186.
- LONG, J., SHELHAMER, E. et DARRELL, T. (2015). Fully convolutional networks for semantic segmentation. *In Proc. of the International Conference on Computer Vision and Pattern Recognition (CVPR)*.
- LU, W.-Y., MAN, H.-Y., JU, W., TRIMBLE, W. S., MACDONALD, J. F. et WANG, Y. T. (2001). Activation of synaptic nmda receptors induces membrane insertion of new ampa receptors and LTP in cultured hippocampal neurons. *Neuron*, 29(1):243–254.

- MAAS, A. L., HANNUN, A. Y. et NG, A. Y. (2013). Rectifier nonlinearities improve neural network acoustic models. *In Proc. of the International Conference on Machine Learning (ICML)*.
- MAATEN, L. v. d. et HINTON, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9(Nov):2579–2605.
- MACLEOD, G. T. (2012). Imaging and analysis of nonratiometric calcium indicators at the drosophila larval neuromuscular junction. *Cold Spring Harbor Protocols*, 2012(7):pdb.prot070110.
- MANCINI, R., van der BIJL, T., BOURGEOIS-JAARSMA, Q., LASABUDA, R. et GROFFEN, A. J. (2018). SICT: Automated detection and supervised inspection of fast Ca²⁺ transients. *Scientific Reports*, 8(1):1–13.
- MARTIN, S. J., GRIMWOOD, P. D. et MORRIS, R. G. (2000). Synaptic plasticity and memory: An evaluation of the hypothesis. *Annual Review of Neuroscience*, 23(1):649–711.
- MARUYAMA, R., MAEDA, K., MORODA, H., KATO, I., INOUE, M., MIYAKAWA, H. et AONISHI, T. (2014). Detecting cells using non-negative matrix factorization on calcium imaging data. *Neural Networks*, 55(1):11–19.
- MATLAB (2016). *MATLAB Release 2016b*. The MathWorks Inc. <https://www.mathworks.com/help/releases/R2016b/index.html>, Natick, Massachusetts.
- MAYER, M. L., WESTBROOK, G. L. et GUTHRIE, P. B. (1984). Voltage-dependent block by Mg²⁺ of NMDA responses in spinal cord neurones. *Nature*, 309(5965):261–263.
- MCINNES, L., HEALY, J. et MELVILLE, J. (2020). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *ArXiv e-print 1802.03426*, <https://arxiv.org/abs/1802.03426>.
- MILLETARI, F., NAVAB, N. et AHMADI, S.-A. (2016). V-net: Fully convolutional neural networks for volumetric medical image segmentation. *In Proc. of the International Conference on 3D Vision (3DV)*.
- MUKAMEL, E. A., NIMMERJAHN, A. et SCHNITZER, M. J. (2009). Automated analysis of cellular signals from large-scale calcium imaging data. *Neuron*, 63(6):747–760.
- MURPHY, T. H., BARABAN, J. M. et WIER, W. G. (1995). Mapping miniature synaptic currents to single synapses using calcium imaging reveals heterogeneity in postsynaptic output. *Neuron*, 15(1):159–168.
- NOH, H., HONG, S. et HAN, B. (2015). Learning deconvolution network for semantic segmentation. *In Proc. of the International Conference on Computer Vision (ICCV)*.

- ODENA, A., DUMOULIN, V. et OLAH, C. (2016). Deconvolution and checkerboard artifacts. *Distill*, <http://doi.org/10.23915/distill.00003>.
- OHKI, K., CHUNG, S., CH'NG, Y. H., KARA, P. et REID, R. C. (2005). Functional imaging with cellular resolution reveals precise micro-architecture in visual cortex. *Nature*, 433(7026):597–603.
- O'ROURKE, N. A., WEILER, N. C., MICHEVA, K. D. et SMITH, S. J. (2012). Deep molecular diversity of mammalian synapses: Why it matters and how to measure it. *Nature Reviews Neuroscience*, 13(6):365–379.
- OTSU, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1):62–66.
- PACHITARIU, M., STRINGER, C., DIPOPPA, M., SCHRÖDER, S., ROSSI, L. F., DALGLEISH, H., CARANDINI, M. et HARRIS, K. D. (2017). Suite2p: beyond 10,000 neurons with standard two-photon microscopy. *BioRxiv e-print 10.1101/061507* <https://www.biorxiv.org/content/early/2017/07/20/061507>.
- PACHITARIU, M., STRINGER, C. et HARRIS, K. D. (2018). Robustness of spike deconvolution for neuronal calcium imaging. *Journal of Neuroscience*, 38(37):7976–7985.
- PAPOUIN, T., LADÉPÊCHE, L., RUEL, J., SACCHI, S., LABASQUE, M., HANINI, M., GROG, L., POLLEGIONI, L., MOTHET, J.-P. et OLIET, S. H. R. (2012). Synaptic and extrasynaptic NMDA receptors are gated by different endogenous coagonists. *Cell*, 150(3):633–646.
- PASZKE, A., GROSS, S., CHINTALA, S., CHANAN, G., YANG, E., DEVITO, Z., LIN, Z., DESMAISON, A., ANTIGA, L. et LERER, A. (2017). Automatic differentiation in pytorch.
- PASZKE, A., GROSS, S., MASSA, F., LERER, A., BRADBURY, J., CHANAN, G., KILLEEN, T., LIN, Z., GIMELSHEIN, N., ANTIGA, L., DESMAISON, A., KOPF, A., YANG, E., DEVITO, Z., RAISON, M., TEJANI, A., CHILAMKURTHY, S., STEINER, B., FANG, L., BAI, J. et CHINTALA, S. (2019). PyTorch: An imperative style, high-performance deep learning library. *In Proc. of the Conference on Advances in Neural Information Processing Systems (NeurIPS)*.
- PNEVMATIKAKIS, E. A. et GIOVANNUCCI, A. (2017). NoRMCorre: An online algorithm for piecewise rigid motion correction of calcium imaging data. *Journal of Neuroscience Methods*, 291:83–94.
- PNEVMATIKAKIS, E. A., SOUDRY, D., GAO, Y., MACHADO, T. A., MEREL, J., PFAU, D., REARDON, T., MU, Y., LACEFIELD, C., YANG, W. *et al.* (2016). Simultaneous denoising, deconvolution, and demixing of calcium imaging data. *Neuron*, 89(2):285–299.

- RAMIREZ, D. M. et KAVALALI, E. T. (2011). Differential regulation of spontaneous and evoked neurotransmitter release at central synapses. *Current Opinion in Neurobiology*, 21(2):275–282.
- RAZA, S. E. A., CHEUNG, L., EPSTEIN, D., PELENGARIS, S., KHAN, M. et RAJPOOT, N. M. (2017). Mimo-net: A multi-input multi-output convolutional neural network for cell segmentation in fluorescence microscopy images. *In Proc. of the International Symposium on Biomedical Imaging (ISBI)*.
- REESE, A. L. et KAVALALI, E. T. (2015). Spontaneous neurotransmission signals through store-driven Ca²⁺ transients to maintain synaptic homeostasis. *eLife e-print 4:e09262*, <https://elifesciences.org/articles/09262>.
- RONNEBERGER, O., FISCHER, P. et BROX, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. *In Proc. of the Conference on Medical Computing and Computer Assisted Intervention (MICCAI)*.
- RUECKL, M., LENZI, S. C., MORENO-VELASQUEZ, L., PARTHIER, D., SCHMITZ, D., RUEDIGER, S. et JOHENNING, F. W. (2017). SamuROI, a Python-based software tool for visualization and analysis of dynamic time series imaging at multiple spatial scales. *Frontiers in Neuroinformatics e-print 11(1):44*, <https://www.frontiersin.org/articles/10.3389/fninf.2017.00044/full>, 11(1).
- SADANANDAN, S. K., RANEFALL, P., LE GUYADER, S. et WÄHLBY, C. (2017). Automated training of deep convolutional neural networks for cell segmentation. *Sci Rep e-print 7(1):7860*, <https://www.nature.com/articles/s41598-017-07599-6>.
- SAITO, T. et REHMSMEIER, M. (2015). The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets. *PLoS computational biology e-print 10(3):e0118432*, <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0118432>.
- SCHERMELLEH, L., FERRAND, A., HUSER, T., EGGELING, C., SAUER, M., BIEHLMAIER, O. et DRUMMEN, G. P. (2019). Super-resolution microscopy demystified. *Nature Cell Biology*, 21(1):72–84.
- SCHINDELIN, J., ARGANDA-CARRERAS, I., FRISE, E., KAYNIG, V., LONGAIR, M., PIETZSCH, T., PREIBISCH, S., RUEDEN, C., SAALFELD, S., SCHMID, B., TINEVEZ, J.-Y., WHITE, D. J., HARTENSTEIN, V., ELICEIRI, K., TOMANCAK, P. et CARDONA, A. (2012). Fiji: An open-source platform for biological-image analysis. *Nature Methods*, 9(7):676–682.
- SEEBÖCK, P., ORLANDO, J. I., SCHLEGL, T., WALDSTEIN, S. M., BOGUNOVIĆ, H., KLIMSCHA, S., LANGS, G. et SCHMIDT-ERFURTH, U. (2020). Exploiting epistemic uncertainty

- of anatomy segmentation for anomaly detection in retinal OCT. *IEEE Transactions on Medical Imaging*, 39(1):87–98.
- SHANNON, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423.
- SIMONYAN, K. et ZISSERMAN, A. (2015). Very deep convolutional networks for large-scale image recognition. *ArXiv e-print 1409.1559*, <https://arxiv.org/abs/1409.1556>.
- SOLTANIAN-ZADEH, S., SAHINGUR, K., BLAU, S., GONG, Y. et FARSIU, S. (2019). Fast and robust active neuron segmentation in two-photon calcium imaging using spatiotemporal deep learning. *Proc. of the National Academy of Sciences (PNAS)*, 116(17):8554–8563.
- SOMEKH, B. et LEWIN, C. (2005). *Research methods in the social sciences*. Sage.
- STOSIEK, C., GARASCHUK, O., HOLTHOFF, K. et KONNERTH, A. (2003). In vivo two-photon calcium imaging of neuronal networks. *Proc. of the National Academy of Sciences (PNAS)*, 100(12):7319–7324.
- SUN, Z. et SÜDHOF, T. C. (2021). A simple Ca²⁺-imaging approach to neural network analyses in cultured neurons. *Journal of Neuroscience Methods e-print 349(1):109041*, <https://pubmed.ncbi.nlm.nih.gov/33340555/>, 349:109041.
- SUTTON, M. A., ITO, H. T., CRESSY, P., KEMPF, C., WOO, J. C. et SCHUMAN, E. M. (2006). Miniature neurotransmission stabilizes synaptic function via tonic suppression of local dendritic protein synthesis. *Cell*, 125(4):785–799.
- SUTTON, M. A. et SCHUMAN, E. M. (2009). Partitioning the synaptic landscape: Distinct microdomains for spontaneous and spike-triggered neurotransmission. *Science Signaling e-print 2(65):pe19*, <https://stke.sciencemag.org/content/2/65/pe19>.
- SVOBODA, K., DENK, W., KLEINFELD, D. et TANK, D. W. (1997). In vivo dendritic calcium dynamics in neocortical pyramidal neurons. *Nature*, 385(6612):161–165.
- SZEGEDY, C., IOFFE, S., VANHOUCHE, V. et ALEMI, A. (2017). Inception-v4, Inception-ResNet and the impact of residual connections on learning. *In Proc. of the Conference on Association for the Advancement of Artificial Intelligence (AAAI)*.
- TAGHANAKI, S. A., ABHISHEK, K., COHEN, J. P., COHEN-ADAD, J. et HAMARNEH, G. (2021). Deep semantic segmentation of natural and medical images: a review. *Artificial Intelligence Review*, 54(1):137–178.
- TAHA, A. A. et HANBURY, A. (2015). Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool. *BMC Medical Imaging e-print 15(1):29*, <https://bmcmimedimaging.biomedcentral.com/articles/10.1186/s12880-015-0068-x>.

- TIAN, Q., SCHRÖDER, L., SCHWARZ, Y., FLOCKERZI, A., KAESTNER, L., ZEUG, A., BRUNS, D. et LIPP, P. (2019). Large scale, unbiased analysis of elementary calcium signaling events in cardiac myocytes. *Journal of Molecular and Cellular Cardiology*, 135(1):79–89.
- TRAN, D., BOURDEV, L., FERGUS, R., TORRESANI, L. et PALURI, M. (2015). Learning spatiotemporal features with 3D convolutional networks. *In Proc. of the International Conference on Computer Vision (ICCV)*.
- TSIEN, R. W. (1983). Calcium channels in excitable cell membranes. *Annual Review of Physiology*, 45(1):341–358.
- TURRIGIANO, G. G., LESLIE, K. R., DESAI, N. S., RUTHERFORD, L. C. et NELSON, S. B. (1998). Activity-dependent scaling of quantal amplitude in neocortical neurons. *Nature*, 391(6670):892–896.
- UDUPA, J. K., LEBLANC, V. R., ZHUGE, Y., IMIELINSKA, C., SCHMIDT, H., CURRIE, L. M., HIRSCH, B. E. et WOODBURN, J. (2006). A framework for evaluating image segmentation algorithms. *Computerized Medical Imaging and Graphics*, 30(2):75–87.
- VAN VALEN, D. A., KUDO, T., LANE, K. M., MACKLIN, D. N., QUACH, N. T., DEFELICE, M. M., MAAYAN, I., TANOUCI, Y., ASHLEY, E. A. et COVERT, M. W. (2016). Deep learning automates the quantitative analysis of individual cells in live-cell imaging experiments. *PLoS computational biology e-print 12(11):e1005177*, <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005177>.
- VAPNIK, V. (2013). *The nature of statistical learning theory*. Springer science & business media.
- VOHS, K. D., BAUMEISTER, R. F., SCHMEICHEL, B. J., TWENGE, J. M., NELSON, N. M. et TICE, D. M. (2014). Making choices impairs subsequent self-control: a limited-resource account of decision making, self-regulation, and active initiative. *Educational Publishing Foundation e-print 94(5)*, <https://pubmed.ncbi.nlm.nih.gov/18444745/>.
- WALTER, A. M., HAUCKE, V. et SIGRIST, S. J. (2014). Neurotransmission: Spontaneous and evoked release filing for divorce. *Current Biology*, 24(5):R192–R194.
- WANG, J. W., WONG, A. M., FLORES, J., VOSSHALL, L. B. et AXEL, R. (2003). Two-photon calcium imaging reveals an odor-evoked map of activity in the fly brain. *Cell*, 112(2):271–282.
- WIESNER, T. (2021). *Quantitative Assessment of Synaptic Plasticity at the Molecular Scale with Multimodal Microscopy and Computational Tools*. Thèse de doctorat, Université Laval.

- WIESNER, T., BILODEAU, A., BERNATCHEZ, R., DESCHÊNES, A., RAULIER, B., DE KONINCK, P. et LAVOIE-CARDINAL, F. (2020). Activity-dependent remodeling of synaptic protein organization revealed by high throughput analysis of sted nanoscopy images. *Frontiers in Neural Circuits e-print 14:57*, <https://www.frontiersin.org/articles/10.3389/fncir.2020.00057/full>.
- WILLIAMS, K. (2001). Ifenprodil, a novel NMDA receptor antagonist: Site and mechanism of action. *Current Drug Targets*, 2(3):285–298.
- WILT, B. A., BURNS, L. D., WEI HO, E. T., GHOSH, K. K., MUKAMEL, E. A. et SCHNITZER, M. J. (2009). Advances in light microscopy for neuroscience. *Annual Review of Neuroscience*, 32:435–506.
- WOLD, S., ESBENSEN, K. et GELADI, P. (1987). Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1-3):37–52.
- XU, B., WANG, N., CHEN, T. et LI, M. (2015). Empirical evaluation of rectified activations in convolutional network. *ArXiv e-print 1505.00853*, <https://arxiv.org/abs/1505.00853>.
- y CAJAL, S. R. (1911). *Histologie du système nerveux de l'homme & des vertébrés: Cervelet, cerveau moyen, rétine, couche optique, corps strié, écorce cérébrale générale & régionale, grand sympathique*. A. Maloine.
- YEGHIAZARYAN, V. et VOICULESCU, I. D. (2018). Family of boundary overlap metrics for the evaluation of medical image segmentation. *Journal of Medical Imaging*, 5(1):015006.
- ZACK, G. W., ROGERS, W. E. et LATT, S. A. (1977). Automatic measurement of sister chromatid exchange frequency. *Journal of Histochemistry & Cytochemistry*, 25(7):741–753.
- ZEILER, M. D. et FERGUS, R. (2014). Visualizing and understanding convolutional networks. *In Proc. of the European Conference on Computer Vision (ECCV)*.
- ZHAO, L. (2013). The gut microbiota and obesity: From correlation to causality. *Nature Reviews Microbiology*, 11(9):639–647.
- ZHAO, W., JIANG, D., QUERALTA, J. P. et WESTERLUND, T. (2020). Multi-scale supervised 3D U-Net for kidneys and kidney tumor segmentation. <https://arxiv.org/abs/2004.08108>.
- ZHOU, Z., SIDDIQUEE, M. M. R., TAJBAKHS, N. et LIANG, J. (2020). UNet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE Transactions on Medical Imaging*, 39(6):1856–1867.
- ZOCCOLER, M. et DE OLIVEIRA, P. X. (2020). METROID: An automated method for robust quantification of subcellular fluorescence events at low SNR. *BMC Bioinformatics*, 21(1):332.

ZUCKER, R. S. (2005). Minis: Whence and wherefore? *Neuron*, 45(4):482–484.

ZUCKER, R. S. et REGEHR, W. G. (2002). Short-term synaptic plasticity. *Annual Review of Physiology*, 64(1):355–405.