



Rapport du Stage

Candidat de soutenance

Nom et prénom : Qianben CHEN
Année d'étude : Deuxième année d'ingénieur
Formation : Ingénierie Généraliste Ei2 - STING

Supervisé par

Madame Diana MATEUS
Enseignant Chercheur
Centrale Nantes/LS2N

Monsieur Saïd MOUSSAOUI
Enseignant Chercheur
Centrale Nantes/LS2N

Monsieur Oudom Somphone
Scientifique Senior à Philips
Philips Research Suresnes

Monsieur Sègbédji Junior GOUBALAN
Ingénieur de Recherche
Philips Research Suresnes

Ce document présente le travail achevé par Qianben CHEN durant la période du 4 mai au 30 août en tant que stagiaire chez Philips Research Suresnes, ce qui est considéré le stage d'ingénieur (STING) demandé par Ecole Centrale de Nantes.

Entreprise d'accueil

Philips Research Suresnes

Durée du stage

Du 4 mai au 30 août (17 semaines), 2020

Titre du stage

Stage entreprise pour le débruitage de l'imagerie échographique médicale en Deep Learning

Mission du stage

- 1) Établissement d'une base de données avec le bruit synthétique
- 2) Réalisation des approches de bout en bout avec Deep Learning
- 3) Compréhension et interprétation des comportements des réseaux neuronaux
- 4) Exploration des méthodes hybrides pour rendre le modèle explicable et interprétable
- 5) Visualisation, explication et perfectionnement du comportement des réseaux neuronaux

Profile

BAC+4/+5 ;

Ecole d'ingénieur/Master ou équivalent ;

Connaissances solides des statistiques, de Deep Learning et du traitement d'image

Connaissances ou une expérience en débruitage et filtrage conventionnel

Anglais/français à l'oral et à l'écrite ;

Expérience en Python et Matlab ;

Bonnes aptitudes à la communication et capacité à travailler en équipe

Table des matières

I RÉSUMÉ.....	5
II INTRODUCTION.....	6
III PRÉSENTATION DE L'ENTREPRISE.....	7
INTRODUCTION GLOBALE.....	7
PHILIPS RESEARCH SURESNES	9
IV PRÉSENTATIONS DU CONTEXTE OU DE LA PROBLÉMATIQUE DU STAGE.....	10
V ORGANISATION DU STAGE.....	11
VI PRÉSENTATIONS DE LA MÉTHODOLOGIE DE DEEP LEARNING	12
ÉTABLISSEMENT D'UNE BASE DE DONNEES.....	12
<i>Données d'ImageNet.....</i>	<i>12</i>
<i>Données DIV2K.....</i>	<i>13</i>
<i>Données d'Ultrason (par Philips).....</i>	<i>13</i>
<i>Prétraitement des Données.....</i>	<i>14</i>
CONSTRUCTION DES MODELES DE DEEP LEARNING.....	15
<i>DnCNN.....</i>	<i>15</i>
<i>FFDNet.....</i>	<i>15</i>
<i>U-net.....</i>	<i>16</i>
<i>CBDNet.....</i>	<i>17</i>
<i>Réseau des Filtres Dynamiques.....</i>	<i>18</i>
Approche Globale	19
Approche Locale	19
<i>Conclusion.....</i>	<i>20</i>
EXPLICATION, INTERPRETATION ET AMELIORATION DES MODELES	21
<i>Comportement de la Structure</i>	<i>21</i>
Pour Tous les Modèles.....	21
Pour le Réseau des Filtres Dynamiques.....	22
<i>Études de la Capacité de la Généralisation.....</i>	<i>22</i>
Entraînement avec le bruit de variances variées	23
Carte d'Estimation du bruit	24
Modèle sans biais.....	24
VII PRÉSENTATIONS DES RÉSULTATS DE DEEP LEARNING	26
CHOIX POUR L'ENTRAÎNEMENT DES MODÈLES	26
<i>Fonction du coût.....</i>	<i>26</i>
<i>Initialisation.....</i>	<i>28</i>
<i>Préapprentissage.....</i>	<i>28</i>
COMPARAISON DES MODÈLES.....	29
<i>U-net, Réseaux des Filtres Dynamiques Globaux et Locaux.....</i>	<i>30</i>
<i>Les Autres.....</i>	<i>32</i>

AMÉLIORATION DU RÉSEAU DES FILTRES DYNAMIQUES LOCAUX.....	32
REPORTAGE DES ÉTUDES DE LA CAPACITÉ DE LA GÉNÉRALISATION.....	33
VIII POUR ALLER PLUS LOIN: MODÈLE HYBRIDE.....	35
ANALYSE DES FILTRES	36
FONCTION CORRELEE ENTRE PATCHES ET FILTRES	37
IX CONCLUSION.....	41
X REMERCIEMENT.....	42
XI RÉFÉRENCE	42

I Résumé

J'ai effectué un stage de quatre mois dans le département du Philips Research Suresnes. Ce stage, qui vise à faire du débruitage des images médicales avec Deep Learning, correspond bien à ce que j'ai appris à Centrale Nantes et ce que je vais faire aux études futures. Pendant mon stage, je suis resté vigoureux du début à la fin. J'ai maintenu un esprit dynamique et bien motivant au travail. J'ai accompli les tâches que j'ai été assignées sous la supervision de mes tuteurs. Grâce à eux, j'ai pu apprécier la gestion des projets, la méthodologie des outils informatiques et des idées mathématiques. Même si on vivait dans une situation difficile à cause de l'épidémie et on organisait les activités professionnelles sur Internet pour la plupart du temps, je me suis bien entendu avec eux et on partageait une bonne ambiance. Ce qui est noté, c'est que j'ai maintenu des communications deux fois par mois avec mes tuteurs, des échanges d'expérience hebdomadaires avec les autres stagiaires et des présentations de stage mensuelles envers toute l'équipe avec lesquelles j'ai bien compris l'art de communication et j'ai amélioré l'efficacité de la transmission des messages. A part du travail journalier, j'ai aussi beaucoup appris des connaissances professionnelles grâce aux opportunités de communiquer avec les scientifiques et les ingénieurs supérieurs, ce qui me permet de me perfectionner dans mon domaine spécialisé voire plus dans ma carrière professionnelle.

Généralement, c'est une opportunité précieuse pour moi qui m'a enseigné non seulement des compétences professionnelles mais aussi des visions globales dans l'industrie. Ma capacité concernant la gestion des projets, des outils informatiques et des idées mathématiques est bien entraînée. De plus, grâce à cette chance inoubliable, ma conception du monde industriel s'agrandit de sorte que j'obtiens une telle conclusion : Ce que l'on a appris à l'école est peu suffisant. Ce qui est vraiment primordial, c'est l'idée d'étudier et absorber de nouvelles connaissances continuellement et pratiquer et maîtriser la compétence exigée du poste au plus vite. La pensée d'apprendre à long terme et de renouveler le système de connaissances sans cessation est essentielle et cela me permet d'avoir une place et une compétitivité dans le marché des talents.

II Introduction

Grâce à mon stage de quatre mois chez Philips, j'ai eu l'opportunité d'apercevoir le développement avancé de l'application informatique au domaine du traitement des images médicales, spécialement la contribution énorme de l'apprentissage automatique aux besoins plus rigoureux et divers des clients. Sur cette plate-forme, j'ai pu réaliser la gestion des projets et j'ai vu comment élargir le marché local en satisfaisant les clients différents et j'ai fait une contribution à la réalisation des idées de Deep Learning les plus avancées et au développement des outils informatiques pour le débruitage des images médicales.

Ce que j'ai fait peut se diviser en 5 niveaux.

Au premier niveau, j'ai établi une base de données obtenue sur ImageNet avec le bruit synthétique qui est soit gaussien soit 'speckled'. Ce premier essai me permet de me familiariser avec les données et les caractéristiques des bruits.

Au deuxième niveau, j'ai construit plusieurs modèles de Deep Learning inspirés par les thèses les plus avancées sur le débruitage et j'ai considéré un des modèles comme le baseline en comparant leur performance et efficacité. J'ai enregistré des informations importantes afin de rédiger des diapositives de réunion pour la référence des membres dans toute l'équipe.

Au troisième niveau, lors de la réception des conseils de mes tuteurs, j'ai examiné les détails réalisé la visualisation et les performances statistiques de chaque composante des réseaux neuronaux afin d'expliquer et d'interpréter leur comportement. Durant ce processus, j'ai mieux compris la structure des réseaux neuronaux et j'ai trouvé les manières de construire un modèle robuste et explicable comme ce que je voulais.

Au quatrième niveau, je me suis contribué énormément aux méthodes hybrides qui non seulement utilisent des outils de Deep Learning et profite des avantages de l'apprentissage statistique afin de non seulement rendre le modèle robuste et précis mais aussi intègrent les idées des méthodes traditionnelles en mettant des contraintes à l'intérieur des modèles pour qu'ils soient interprétable et contrôlable.

Au cinquième niveau, j'utilise des outils de Machine Learning pour étudier et visualiser le comportement des réseaux neuronaux, ce qui permet d'avoir assuré la sécurité et la stabilité des modèle de Deep Learning. De plus, je peux ainsi aller plus loin pour chercher une manière d'économiser du temps sans perdre la performance des modèles.

Afin de passer en revue l'ensemble de mon stage et de faire un résumé systématique de mon travail au cours de cette période, ce rapport du stage est donc rédigé avec mes enregistrements et réflexions personnelles.

III Présentation de l'Entreprise

Introduction Globale

La société Philips est fondée en 18912 par les frères Gerard et Anton Philips à Eindhoven, aux Pays-Bas. L'entreprise a commencé par produire des lampes à filament de carbone devenant au début du XX^e siècle l'un des plus grands fabricants d'Europe. La première manufacture sert aujourd'hui de musée.

Lancé tout au début dans les premiers essais des soins corporels, des outils d'éclairage et des disques électriques, Philips s'est énormément agrandi. Cependant, Philips a aussi poursuivi ses efforts de recherche dans le domaine de haute-technologie, spécialement dans le domaine médical. Philips a intégré l'ensemble des produits à destination du grand public au sein d'une grande et même activité appelée Consumer Lifestyle (« Mode de vie du Consommateur »). Engagé dans l'ensemble du processus de soins de santé, de la prévention des modes de vie sains et des maladies au diagnostic, au traitement et aux soins à domicile, Philips a pour objectif d'améliorer la santé personnelle et en même temps perfectionner les résultats médicaux par innovation. Elle est l'un des leaders sur les marchés des soins du cœur, des soins de courte durée et des soins à domicile, des systèmes d'éclairage économes en énergie et des nouvelles applications en éclairage, ainsi que dans les domaines des produits de rasage, des tondeuses pour hommes et de l'hygiène bucco-dentaire. Basée aux Pays-Bas, Philips a réalisé un chiffre d'affaires de 17,8 milliards d'euros en 2017, avec environ 74 000 employés dans 100 pays à travers le monde.

En tant qu'entreprise en quête de leadership dans les technologies de la santé, Philips pense que l'innovation peut améliorer la santé et les conditions sanitaires, ainsi que de rendre les soins plus accessibles et abordables. Chez Philips, l'objectif est d'améliorer la vie de 3 milliards de personnes par an d'ici 2030, dont 400 millions dans les communautés de soins de santé mal desservies. Guidé par la vision de rendre le monde plus sain et plus durable grâce à l'innovation, la stratégie de Philips est de diriger avec des solutions innovantes les marchés clés le long du continuum de la santé - aider nos clients à atteindre l'objectif quadruple (des meilleures conditions pour la santé, une meilleure expérience pour les patients et le personnel, un moindre coût des soins) et aider les gens à mieux prendre soin de leur santé à chaque stade de la vie. Philips cherche à agir de manière responsable et durable, en tirant parti de nos ressources pour maximiser la création de valeur pour toutes les parties prenantes. Reflétant son engagement envers les objectifs de développement durable des Nations Unies 3 (assurer une vie saine et promouvoir le bien-être pour tous à tout âge), 12 (Assurer des modes de consommation et de production durables) et 13 (Prendre des mesures urgentes pour lutter contre le changement climatique et ses impacts), Philips continue à intégrer la durabilité plus profondément dans nos activités, en mettant accès aux soins, économie circulaire et action pour le climat.

En ce qui concerne la structure du business en 2019, les segments à déclarer étaient Business

de diagnostic et de traitement, Business de soins connectés et Business de santé personnelle, chacune ayant été responsable de la gestion de ses affaires dans le monde entier.

Pour le Business de diagnostic et de traitement, les activités de diagnostic et de traitement de Philips sont fondamentales à sa stratégie en matière de technologies de la santé, comme le diagnostic de précision et les thérapies guidées par imagerie.

Pour le Business de soins connectés, cela couvre le continuum de la santé et se charge d'améliorer les résultats patients, augmenter l'efficacité et améliorer le niveau de la satisfaction des soignants. Les solutions s'appuient sur les forces de Philips dans les secteurs verticaux (surveillance et analyse, sommeil et soins respiratoires et thérapeutiques) et horizontaux (gestion de la santé de la population et soins connectés informatiques) pour améliorer les résultats cliniques et économiques dans tous les milieux de soins, à l'intérieur et à l'extérieur de l'hôpital.

Pour le Business de santé personnelle, cela joue un rôle important sur le continuum de la santé - dans une vie saine, étapes de la prévention et des soins à domicile – prestation des solutions intégrées, connectées et personnalisées qui soutiennent des modes de vie plus sains et ceux qui vivent avec des maladies chroniques.

Philips			
Diagnosis & Treatment businesses	Connected Care businesses	Personal Health businesses	Other
Diagnostic Imaging Image-Guided Therapy Ultrasound Enterprise Diagnostic Informatics	Monitoring & Analytics Sleep & Respiratory Care Therapeutic Care Population Health Management Connected Care Informatics	Oral Healthcare Mother & Child Care Personal Care Domestic Appliances*	Innovation IP Royalties Central costs Other
Focus of external reporting		*On January 28, 2020, Philips announced that it will review options for future ownership of the Domestic Appliances business, and start the process of creating a separate legal structure for this business.	

Figure 1 Structure du business 2019

En ce qui concerne les résultats opérationnels en 2019, le chiffre d'affaires du Groupe s'est élevé à 19 482 millions d'euros en 2019, 8% de plus en valeur nominale. Ajusté à 3,0% effet de change et impact de consolidation positifs, les ventes comparables étaient de 4,5% plus que celles en 2018.

Philips Group
Sales in millions of EUR unless otherwise stated
 2017 - 2019

	2017	2018	2019
Diagnosis & Treatment businesses	7,365	7,726	8,485
Nominal sales growth (%)	2.9	4.9	9.8
Comparable sales growth (%) ¹⁾	3.4	6.6	5.5
Connected Care businesses	4,331	4,341	4,674
Nominal sales growth (%)	2.0	0.2	7.7
Comparable sales growth (%) ¹⁾	4.5	2.7	3.1
Personal Health businesses	5,685	5,524	5,854
Nominal sales growth (%)	2.4	(2.8)	6.0
Comparable sales growth (%) ¹⁾	5.4	2.3	5.0
Other	400	530	469
Philips Group	17,780	18,121	19,482
Nominal sales growth (%)	2.1	1.9	7.5
Comparable sales growth (%) ¹⁾	3.9	4.7	4.5

¹⁾ Non-IFRS financial measure. For the definition and reconciliation of the most directly comparable IFRS measure, refer to [Reconciliation of non-IFRS information, starting on page 176](#).

Figure 2 Composition de Revenue de Philips

Source : <https://www.results.philips.com/publications/ar19>

Philips Research Suresnes

En ce qui concerne Philips Research, c'est est une organisation mondiale fondée en 1914 qui aide Philips à introduire des innovations significatives qui améliorent la vie personnelle. C'est maintenant l'une des organisations privées de recherche les plus grandes du monde. Philips Research propose des options technologiques pour des innovations dans le domaine de la santé et du bien-être, ciblant à la fois les marchés développés et émergents. Positionné à l'avant-garde du processus d'innovation, Philips Research travaille sur la détection des tendances et des idées à la preuve de concept et au développement de produits unique en son genre.

En ce qui concerne le département en France où je réalise mon stage, c'est nommé Philips Research Suresnes. C'est un département qui rassemble le siège de tous les secteurs de Philips en France. Il a été l'un des premiers à recevoir le label HQE (Haute Qualité Environnementale) en région parisienne pour son utilisation efficace de l'énergie et le meilleur recyclage de ses déchets et il est situé juste à côté du Haut-Seine.



Figure 3 Philips Research Suresnes

IV Présentations du Contexte ou de la Problématique du Stage

Le problème principal concernant mon stage réside dans le débruitage des images médicales. Dans la réalité, malheureusement, toutes les images médicales contiennent du bruit visuel. La présence de bruit donne à l'image un aspect marbré, granuleux, texturé ou flocon de neige qui sont inoubliables. Bien que le bruit donne aux images une apparence indésirable, le facteur le plus important est que le bruit peut couvrir et réduire la visibilité de certaines fonctionnalités des images, ce qui n'est pas ce que souhaitent les médecins. En conséquence, le débruitage reste un sujet populaire parmi le traitement des images médicales.

Afin d'atteindre ces objectifs, en avant, on utilisait des outils traditionnels qui se basent fortement sur les connaissances mathématiques, comme l'ondelette. L'expérience montre que ces outils ont une influence positive mais insuffisante. Autrement dit, les résultats attendus ne sont pas assez robustes et performants. Cependant, grâce au développement avancé du Deep Learning, on crée plusieurs réseaux neuronaux profonds qui ont les performances excellentes dans le domaine du débruitage au cours des dernières années, d'où l'idée principale provoque de réseaux de neurones convolutifs[1].

Par conséquent, ce que je fais, se concentre principalement dans la recherche des méthodes de Deep Learning qui aident au débruitage des images médicales. Pourtant, bien que les méthodes de Deep Learning atteignent le résultat à la pointe de la technologie, elles manquent du support de la base mathématique. Ce que j'étudie plus profondément, c'est de

combiner les méthodes traditionnelles et les méthodes de Deep Learning (on les appelle donc les méthodes hybrides), ce qui non seulement joue l'avantage de l'apprentissage statistique, mais aussi rend les modèles plus explicables et interprétables à l'aide des méthodes traditionnelles. Un exemple classique est la recherche du lien entre l'ondelette et U-net, sachant qu'ils ont tous une structure hiérarchique.

Puisque les modèles appliqués à l'industrie nécessitent la sécurité et la stabilité, il faut donc les expliquer et interpréter selon la base théorique. Afin d'atteindre cet objectif, je crée des outils pour analyser et visualiser le comportement des réseaux neuronaux à l'aide du Machine Learning. Eventuellement, j'établis un modèle hybride qui non seulement remporte les avantages statistiques mais aussi se rend explicable avec les fondements théoriques.

Ce qu'il faut souligner, c'est que je participe à la réquisition d'un brevet d'entreprise au sein de mon équipe. Ce brevet, appelant « Deep Learning-based filter analysis for speckle reduction and image sharpness improvement », vise à apprendre les filtres locaux par Deep Learning qui vont être appliqués aux pixels pour le débruitage et vont s'adapter selon nos besoins. Un autre sujet est le renforcement de l'acuité sur le bord, ce qui est la clé pour garder de l'information mais est difficile à résoudre.

En résumé, nous avons prolongé les outils de Deep Learning dans le domaine du débruitage d'imagerie médicale alors que les supports mathématiques s'appliquent à l'aide des modèles hybrides. Finalement, nos contributions aux modèles hybrides permettent de non seulement nous approcher d'une performance de l'art mais aussi les contrôler et ajuster pour satisfaire les demandes industrielles.

V Organisation du Stage

Mon stage se déroule avec les autres enseignants seniors qui visent au même objectif. En conséquence, je participe à continuer leur travail en cherchant et déployant les nouvelles méthodes ainsi que le comportement des composantes de chaque structure.

Malheureusement, à cause de la situation sanitaire, mon stage réalise principalement à distance. Un ordinateur de Philips m'est livré pour perfectionner le stage en ligne. En raison de la nécessité de certains travaux ou le besoin des contacts humains, comme le point du stage avec mes tuteurs, je travaille dans le bureau une fois par semaine.

Afin de nous échanger des idées dans l'équipe, j'organise plusieurs présentations, soient envers mes encadrants chaque semaine, soient envers tous les employeurs dans le laboratoire chaque moi. Les échanges, qui non seulement renforcent la communication technique mais aussi permettent d'avoir une meilleure manière du contact humain dans la période COVID-19, compte significativement. De plus, parce que Philips est une entreprise internationale, je rejoins aussi régulièrement la réunion organisée par la succursale américaine pour un partage des techniques les plus récentes, ce qui est une chance précieuse pour moi.

En tant qu'un des six stagiaires, je tiens la chance de communiquer avec les autres qui sont autant jeunes et débutants que moi pour espérer une amélioration dans la vie professionnelle.

Au total, je suis reconnaissant pour les nombreuses chances de partager avec les autres enseignants de la même profession dans la plate-forme de Philips. Je grave dans la mémoire l'organisation complète et ordonnée du déroulement du stage.

VI Présentations de la Méthodologie de Deep Learning

Établissement d'une Base de Données

Nous établissons trois types de données :

1. Les données d'ImageNet sont obtenus en ligne. Ces données sont nombreuses et simples.
2. Les données DIV2K sont obtenus en ligne et réputées comme les images de haute qualité pour plusieurs missions dans le domaine du traitement d'image
3. Les données ultrason, qui sont fournies par Philips et considérées comme les objectifs finaux pour le débruitage.

Nous travaillons tout d'abord sur les données d'ImageNet pour accéder au premier essai de la faisabilité des modèles. Ensuite, nous nous concentrons sur les données DIV2K qui sont plus grandes et compliqués. Dans les deux premiers cas, nous sommes capables de mettre manuellement les bruits avec le type et le niveau espérés selon nos besoins.

Finalement, nous transférons nos études en données objectives : les données d'ultrason par Philips. Dans ce dernier cas, nous fixons le type de bruit en chatoiement parce que en réalité, les images d'ultrason sont prises avec le bruit en chatoiement.

Données d'ImageNet

Nous travaillons tout d'abord sur les données d'ImageNet en raison de sa accessibilité facile et sa taille petite, ce qui permet d'avoir un premier essai rapide.

Ce que nous mettons en entrée des modèles sont les données d'animal, de plante et d'artefact. Les images grises, qui comptent environ 2500 en total, sont tous de taille fixe en 128 carrée. Pour la normalisation, nous mettons tous les valeurs des pixels entre 0 et 1.

En ce qui concerne le bruit, nous ajoutons manuellement le bruit gaussien blanc avec son écart-type égal à 0.05 tout au début.

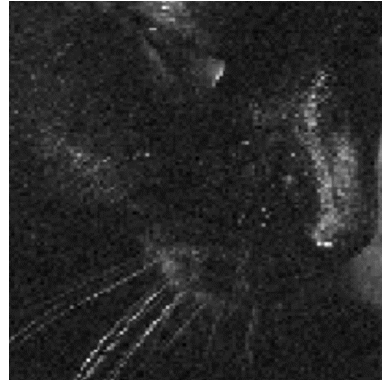
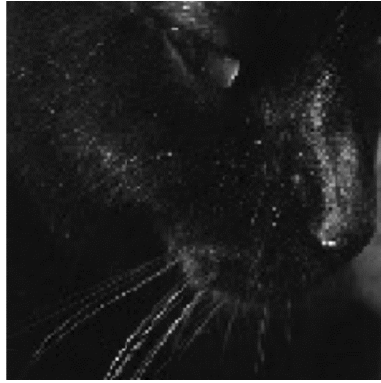


Figure 4 Un exemplaire sur ImageNet de vérité terrain (à gauche) et avec bruit gaussien manuel (à droite)

Données DIV2K

Nous nous intéressons ensuite aux données DIV2K parce qu'elles sont de qualité plus grande ainsi que leur taille, ce qui se rapproche plus du fait des données objectives.

Ce que nous mettons en entrée des modèles sont 800 images pour l'entraînement et 100 images pour le test. Considérons que les images sont de taille grande, nous mettons les patches qui sont capturés aléatoirement des images entières pour l'entraînement alors que nous mettons les images entières pour le test. Pour la normalisation, nous mettons tous les valeurs des pixels entre 0 et 1.

En ce qui concerne le bruit, nous simulons un bruit d'ultrason du pseudo B mode qui peut être considéré comme le bruit de chatolement. Premièrement, nous mettons son écart-type fixé en 0.01. Les codes réalisés sont en Matlab.



Figure 5 Un exemplaire de DIV2K de vérité terrain (à gauche) et avec bruit de chatolement (à droite)

Données d'Ultrason (par Philips)

Après l'accumulation des expériences des deux premiers types de données, nous transférons éventuellement nos études en données d'ultrason pour vérifier si les travaux satisfont les

besoins des clients.

Les images bruitées sont prises par les caméras alors que les images de vérité terrain sont calculées et générées avec les sources d'angles différents en fonction d'un algorithme.

Ce que nous mettons en entrée des modèles sont 1101 images pour l'entraînement et 150 images pour le test. Nous les traitons de la même manière que les images DIV2K en raison de leur taille grande. Pour la normalisation, nous mettons tous les valeurs des pixels entre 0 et 1.

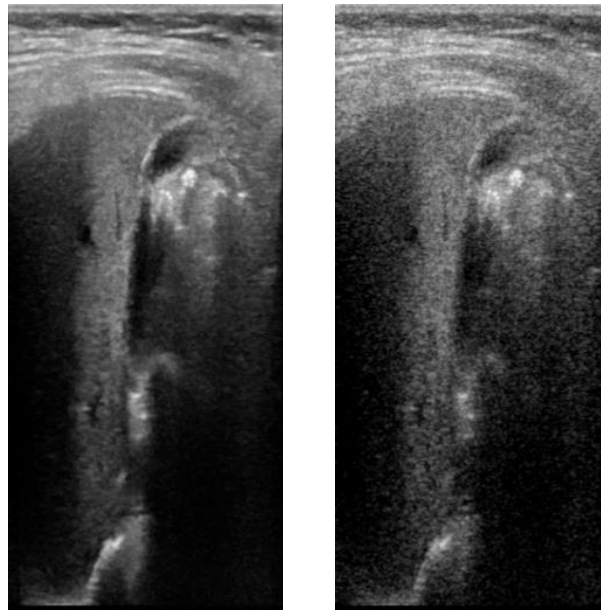


Figure 6 Un exemplaire d'Ultrason de vérité terrain (à gauche) et avec bruit de chatolement (à droite)

Prétraitement des Données

En ce qui concerne la taille des images, toutes les images d'entraînement sont coupées en patches de taille 128 carrée à cause de leur taille originale infixe alors que les images de test sont prises en compte de taille pleine.

En ce qui concerne la normalisation, nous utilisons la Normalisation Min-Max pour rédiger toutes les valeurs des images bruitées et non bruitées entre 0 et 1. La formule de cette normalisation est donnée ci-dessous :

$$\chi^* = \frac{\chi - \min}{\max - \min}$$

En ce qui concerne le type de données, nous mettons *float32* qui est assez précis dans notre cas.

Construction des Modèles de Deep Learning

DnCNN[2]

Le modèle DnCNN est le premier en Deep Learning appliqué dans le débruitage des images. Auparavant, les méthodes sont principalement basées sur les antécédents et calculées selon leur signification physique. Cependant, le modèle DnCNN, qui emprunte la structure de réseaux de neurones convolutifs qui est efficace pour augmenter la capacité et flexibilité dans l'exploitation des caractéristiques des images, atteint une performance plus élevée et stable que les méthodes traditionnelles précédentes. Son architecture est dessinée ci-dessous :

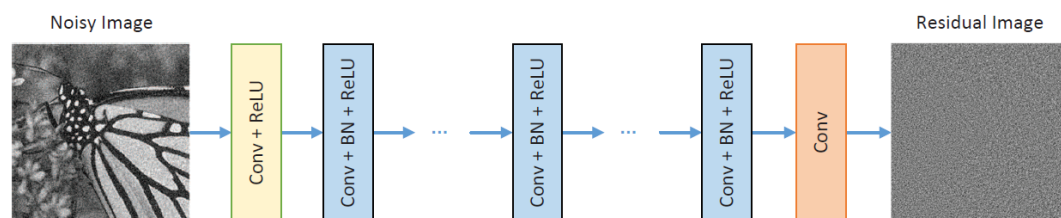


Figure 7 Architecture proposée du réseau DnCNN

L'architecture principale sert à apprendre le résidu (autrement dit, le bruit) de l'image en entrée d'après une combinaison linéaire de Conv + BN + ReLU¹. Ensuite, le résidu est appliqué à l'image d'entrée de manière soustractive.

Ce modèle contribue à nos études dans les aspects suivants :

- 1) CNN a une capacité grande pour apprendre les caractéristiques des images.
- 2) L'apprentissage d'un filtre se rend plus efficace au lieu de l'apprentissage direct des images débruitées.

FFDNet[3]

Le modèle FFDNet s'établit sur le fait que les modèles de CNN existants ne généralisent pas bien les bruits réels qui sont variés et inconnus. Autrement dit, le CNN tend toujours à surapprendre le bruit et fonctionne mal si les données de test possèdent un bruit différent que celui entraîné. Donc, FFDNet concourt à la flexibilité et l'efficacité des réseaux de neurones convolutifs par l'ajout d'une carte d'estimation du niveau du bruit, ce qui permet au modèle de prévoir la quantité du bruit et s'y adapter pour avoir plus de généralité en face de bruits différents. En outre, ce modèle prend en entrée les images sous-échantillonnées afin d'augmenter son efficacité et précision. Son architecture est dessinée ci-dessous :

¹ Conv, BN, ReLU sont respectivement appelés couche convolutionnelle, normalisation de lots et fonction d'activation ReLU

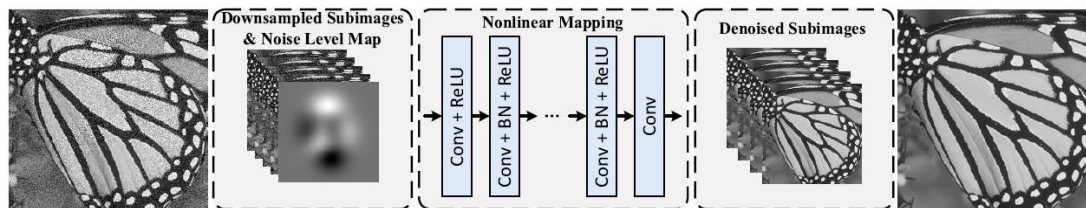


Figure 8 Architecture proposée du réseau FFDNet

Ce modèle contribue à nos études dans les aspects suivants :

- 1) La concaténation d'une carte d'estimation du niveau du bruit aide à augmenter la capacité de généralisation des modèles de Deep Learning.
- 2) La concaténation des images sous-échantillonnées rend le modèle plus efficace et précis.

U-net[4]

Le modèle U-net exerce une influence significative et profonde sur l'abstrait et la reformulation de l'Information des images. U-net est donc réputé dans les vastes problèmes du traitement d'image et considéré comme une méthode beaucoup mieux que les CNN purs. Les sous-échantillonnages et suréchantillonnages au sein du modèle servent à fuser les caractéristiques du niveau élevé et bas pour collecter les informations multi-échelles. De plus, le saut de connexion aide à passer l'information du niveau bas avec succès. Son architecture est dessinée ci-dessous :

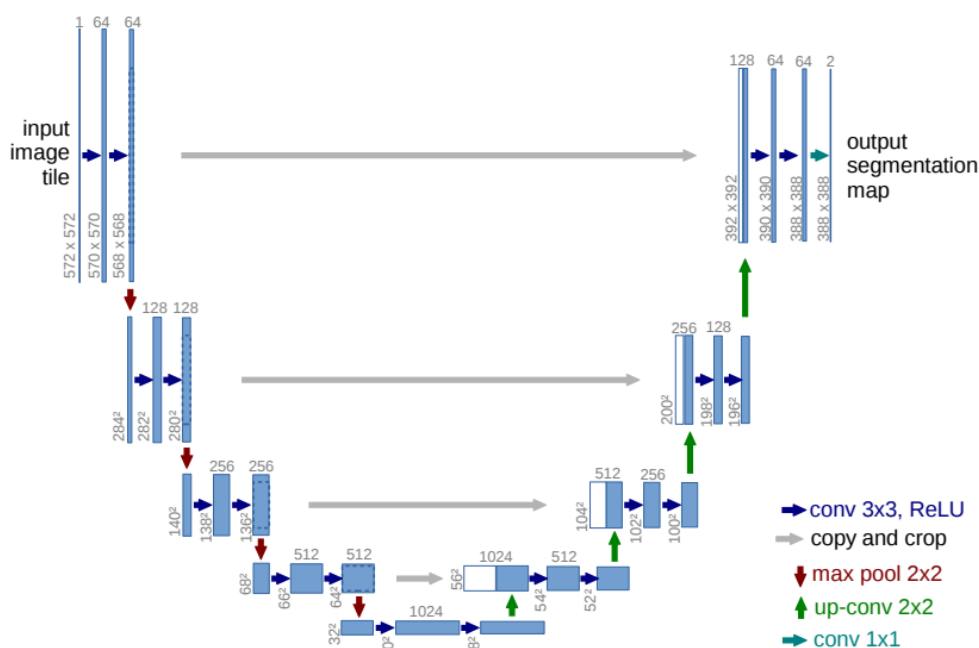


Figure 9 Architecture proposée du réseau U-net

Ce modèle contribue à nos études dans les aspects suivants :

- 1) Le remplacement d'une combinaison de CNN purs par les couches de sous-échantillonnages et suréchantillonnages envisage d'améliorer la capacité de la reformulation de l'information des images par le modèle.
- 2) Le saut de connexion permet de passer mieux l'information du niveau bas qui est utile pour garder corps dominant (voire plus le bord) des images originales dans le domaine du débruitage des images.

CBDNet[5]

Le modèle CBDNet continue ses études suite au modèle FFDNet et U-net. Ce modèle est composé de deux parties : la partie d'estimation du bruit et la partie du débruitage non aveugle. La partie précédente emprunte l'idée du FFDNet en ajoutant une carte d'estimation du niveau du bruit alors qu'il faut l'apprendre par un réseau de neurones parce qu'en réalité, le niveau du bruit est inconnu. La partie suivante emprunte la structure du Unet pour réaliser le débruitage. D'ailleurs, pendant l'entraînement, un terme du coût asymétrique est ajouté pour pénaliser plus si le niveau estimé du bruit est plus petit que le niveau réel considérant chaque pixel, ce qui permet d'avoir une capacité plus grande pour généraliser le bruit. Son architecture est dessinée ci-dessous :

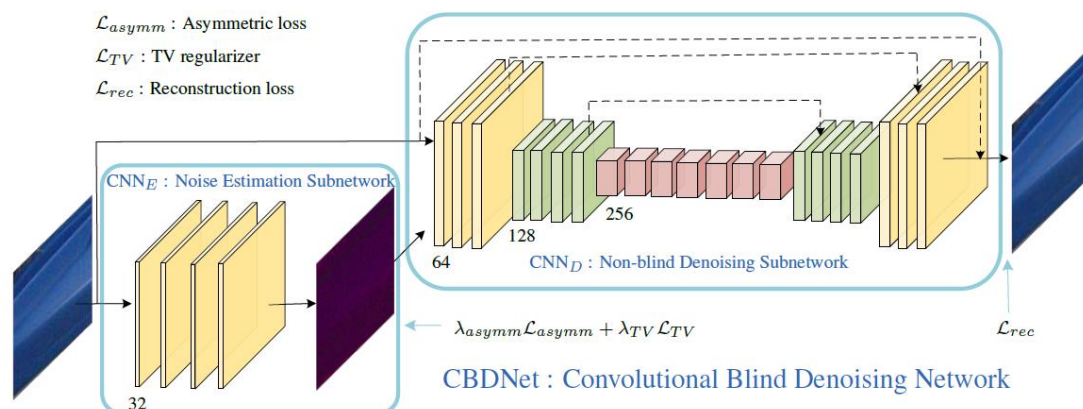


Figure 10 Architecture proposée du réseau CBDNet

Ce modèle contribue à nos études dans les aspects suivants :

- 1) Proposition de l'apprentissage d'une carte d'estimation de bruit permet de traiter le cas réel.
- 2) Le remplacement d'une combinaison de CNN purs par les couches de sous-échantillonnages et suréchantillonnages envisage d'améliorer la capacité de la reformulation de l'information des images par le modèle.
- 3) L'ajout des contraintes comme un terme de coût permet de contrôler le processus d'entraînement, ce qui est utile pour étudier le mécanisme interne des réseaux de neurones.

Réseau des Filtres Dynamiques[6]

Les modèles précédents envisagent d'apprendre de manière directe les caractéristiques objectives des images. L'expérience montre que les caractéristiques objectives non seulement comptent beaucoup plus que les bruits adhérents mais aussi sont variés par rapport aux contenus des images. Par rapport au bruit qui est moins grand en valeur et plus régulier, il vaut mieux qu'on apprenne la caractéristique du bruit comme les filtres et ensuite les appliquions sur les images bruitées, d'où la méthode est appelée Réseau des Filtres Dynamiques (Dynamic Filter Networks en anglais). La raison pourquoi nous l'appelons **dynamique** est que les filtres appris sont appris et s'adaptent aux images d'entrée.

Le modèle est composé de deux parties : un réseau pour la génération de filtres et une couche de filtrage dynamique (voir figure 11). Le réseau pour la génération de filtres génère dynamiquement des filtres spécifiques dont les paramètres sont conditionnés à l'entrée du réseau. Notez que ceux-ci ne sont pas fixes après l'entraînement, différents aux paramètres de modèle. La couche de filtrage dynamique applique ensuite ces filtres spécifiques aux images d'entrée.

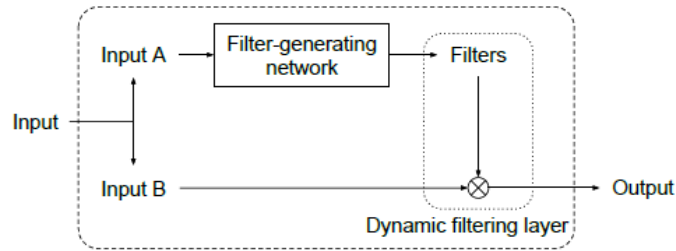


Figure 11 Architecture proposée du Réseau des Filtres Dynamiques

Ce modèle contribue à nos études dans les aspects suivants :

- 1) Au lieu d'apprendre directement les caractéristiques objectives, il vaut mieux apprendre les filtres, ce qui est significatif parce que le bruit est moins grand en valeur et plus régulier.
- 2) Les filtres appris sont spécifiques aux échantillons, ce qui rend le modèle plus robuste.
- 3) Le mécanisme de l'application des filtres permet de lier mathématiquement les méthodes traditionnelles et les méthodes de Deep Learning pour le débruitage d'imagerie, ce qui donne la possibilité pour ajouter des contraintes afin de créer des filtres qui sont de formes explicables et interprétables.

Le réseau pour la génération de filtres prend en entrée $I_A \in \mathbb{R}^{h \times w \times c_A}$, où h , w et c_A sont respectivement la hauteur, la largeur et le nombre de canaux de l'entrée A. Il prend en sortie des filtres F_θ paramétrés par $\theta \in \mathbb{R}^{s \times s \times c_B \times n \times d}$ où s est la taille du filtre, c_B est le nombre de canaux de l'entrée B et n est le nombre de filtres. Ce qu'il faut noter, c'est que d est égal à 1 pour l'approche globale et $h \times w$ pour le l'approche locale, que nous allons discuter ci-

dessous. Les filtres sont appliqués à l'entrée $I_B \in \mathbb{R}^{h \times w \times c_B}$ pour générer une sortie $G = F_\theta(I_B)$, avec $G \in \mathbb{R}^{h \times w \times n}$. La taille du filtre s détermine le champ réceptif et est choisie en fonction de l'application.

D'après la manière de l'application des filtres, nous pouvons séparer cette idée en deux types de modèles : Réseau des Filtres Dynamiques Globaux (approche global) et Réseau des Filtres Dynamiques Locaux (approche locale). Dans la suite, nous choisissons U-net comme le réseau pour la génération de filtres.

Approche Globale

L'approche globale sert à apprendre les filtres globaux qui sont ensuite appliqués aux images d'entrée de manière convolutionnelle.

En ce qui concerne le réseau pour la génération des filtres, la partie du sous-échantillonnage (la partie à gauche, voir la figure 10) a la même structure que celle du modèle U-net, sauf que le saut de connexion s'oriente vers la même couche intermédiaire pour intégrer l'information de plusieurs couches en réduisant la taille spatiale. La partie à droite sert à reformuler l'information et prend ensuite en sortie les filtres dynamiques. Éventuellement, les filtres sont appliqués aux images d'entrée de manière convolutionnelle. Son architecture est dessinée ci-dessous :

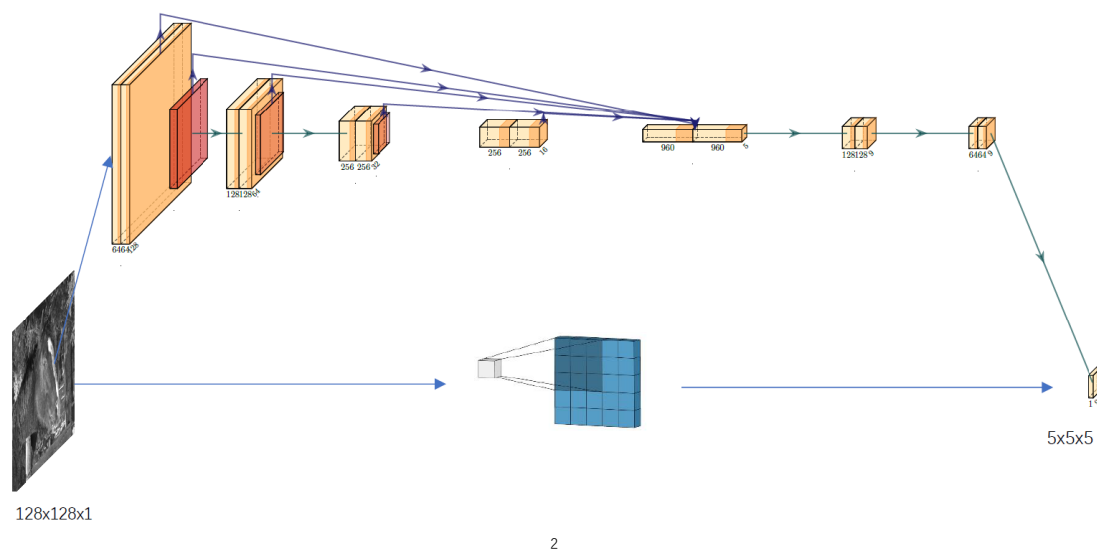


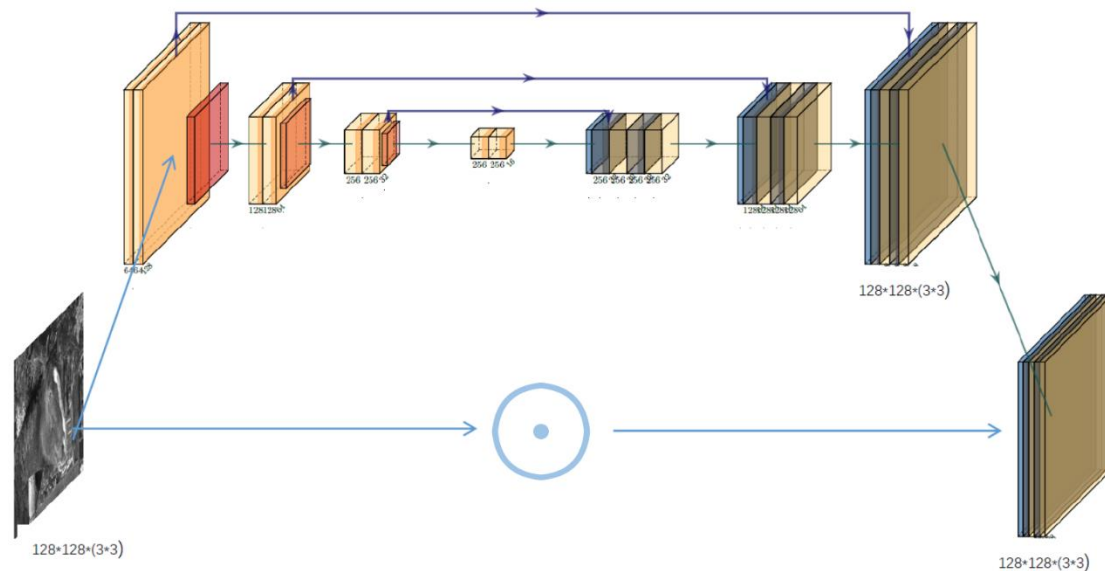
Figure 12 Architecture proposée du Réseau des Filtres Dynamiques Globaux

Approche Locale

² Dans nos expériences, nous mettons la taille spatiale des filtres globaux en 5 carrée, avec le nombre total des filtres globaux égal à 5. Donc, le réseau pour la génération de filtres prend en sortie les filtres de dimension (5x5)x5.

L'approche locale sert à apprendre les filtres locaux qui sont spécifiques à chaque position (autrement dit, chaque pixel) des images en entrée. Nous les appliquons ensuite aux images d'entrée par produit scalaire.

En ce qui concerne le réseau pour la génération des filtres, c'est exactement la structure du modèle U-net. Après la génération de filtres, ils sont appliqués aux images d'entrée par produit scalaire. Son architecture est dessinée ci-dessous :



3

Figure 13 Architecture proposée du Réseau des Filtres Dynamiques Locaux

Conclusion

Nous faisons des expériences sur tous les modèles avec les données de niveau et de type différents. Nous constatons que le Réseau des Filtres Dynamiques Locaux non seulement arrivent au meilleur résultat qui satisfait les besoins des clients dans le domaine du traitement des images médicales mais aussi se rend possible pour la fusion des méthodes traditionnelles par le mécanisme de l'application des filtres. Par conséquent, nous choisissons ce modèle pour la base du modèle hybride ainsi que le déploiement industriel.

De l'autre côté, les autres modèles sont intéressants à étudier par certaines des idées créatives qui peuvent être implémentées dans le modèle hybride :

- 1) L'efficacité de la couche convolutionnelle (DnCNN).
- 2) L'efficacité de la transformation des images en sous-échantillonnage. (FFDNet)
- 3) La capacité de la généralisation par l'ajout d'une carte d'estimation du bruit. (FFDNet)

³ Dans nos expériences, nous mettons la taille spatiale des filtres locaux en 3 carrée, c'est-à-dire, nous nous concentrons à l'entourage de chaque pixel de taille 3x3 et nous générons un filtre de taille 3x3 qui est appliqué à l'entourage de ce pixel par produit scalaire.

- 4) La capacité de la reformulation de l'information des images. (U-net)
- 5) La possibilité de l'apprentissage automatique d'une carte d'estimation du bruit. (CBDNet)
- 6) La possibilité de l'ajout d'un terme de contrainte pour contrôler le comportement des réseaux de neurones. (CBDNet)

Explication, Interprétation et Amélioration des Modèles

Comportement de la Structure

Pour Tous les Modèles

En ce qui concerne la profondeur du modèle, autrement dit, le nombre de couches convolutionnelles, plus il y en a, plus de capacité dont le modèle dispose. L'expérience montre qu'une combinaison successive de plusieurs couches convolutionnelles au lieu d'une seule aide fortement à augmenter la capacité d'apprentissage sans rédiger la structure fondamentale du modèle ni ajouter beaucoup de paramètres. Ainsi, avant chaque sous-échantillonnage ou suréchantillonnage, nous mettons une combinaison de trois couches convolutionnelles. Le schéma illustratif est affiché ci-dessous :

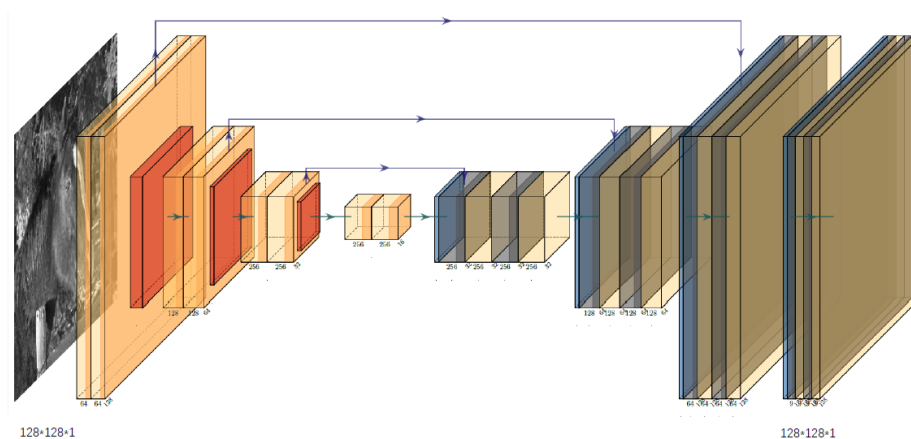


Figure 14 Schéma: dans chaque profondeur, il y a une combinaison de trois cocouches convolutionnelles

Cependant, il ne faut pas que la profondeur soit trop grande pour que le modèle risque de produire le problème du surapprentissage. Donc, éventuellement nous mettons environ 20 couches à tous les modèles dont les nombres de paramètres sont presque 7 millions.

En ce qui concerne la normalisation des lots, cette composante joue le rôle de normaliser la couche d'entrée en ajustant et en mettant à l'échelle les activations afin d'améliorer la vitesse, la performance et la stabilité des réseaux de neurones. Bien que l'effet de la normalisation des lots soit évident, les raisons de son efficacité restent débattues. Dans un cas, ceci accélère le roulement du gradient et évite la disparition et l'exposition des gradients, ce qui ne nous pose pas d'inquiétude. Dans d'autre cas, en considérant que chaque image est spécifique et

grandement variée par rapport à la distribution des valeurs, nous constatons que théoriquement et pratiquement⁴, la normalisation des lots n'est pas utile pour le débruitage d'imagerie.

Pour le Réseau des Filtres Dynamiques

Ce modèle nous intéresse le plus en raison de sa meilleure performance et sa possibilité de la fusion des méthodes traditionnelles. Donc, nous pensons à l'améliorer en ajoutant plusieurs composantes qui sont probablement utiles.

1) La carte d'estimation de biais.

Cette carte est apprise par le réseau afin de déduire la différence de distribution entre les images d'entrée et les images de sortie. Nous faisons l'hypothèse statistique que les images bruitées et les images débruitées sont Indépendantes et distribuées à l'identique, ce qui est approché par l'ajout d'une carte d'estimation de biais. Ce qu'il faut noter, c'est que la carte est de la même taille spatiale que les images d'entraînement et nous l'y ajoutons après que les autres filtres y sont appliqués.

2) Attention spatiale

Nous mettons le mécanisme d'attention sur l'entourage de chaque pixel, c'est-à-dire pour le pixel voisin qui est de valeur plus corrélée avec le pixel central, nous mettons plus de poids sur la position correspondante dans le filtre qui est lié avec ce pixel central. L'idée riche est le principe essentiel de la création des filtres d'après les méthodes traditionnelles dans le domaine du débruitage d'imagerie.

3) Superposition simultanée de plusieurs filtres locaux de tailles différentes

Nous apprenons plusieurs filtres qui fonctionnent en même temps pour un seul pixel. Par exemple, pour chaque pixel, nous pouvons mettre trois filtres qui sont de tailles respectivement 3x3, 5x5 et 7x7 et qui vont y être appliqués successivement. L'idée se propose pour augmenter la complexité du modèle, souhaitant renforcer sa robustesse.

Études de la Capacité de la Généralisation

La capacité de la généralisation des modèles de Deep Learning reste toujours un sujet critique parce que dans la réalité, nous souhaitons que les modèles soient adaptables aux situations variées et réelles. Spécifiquement, dans le domaine du débruitage d'imagerie, il existe deux

⁴ Nos résultats montrent que le coût du test diminue avec l'ajout de la normalisation des lots. Pour les critères, PSNR diminue de 4% à 8% et SSIM diminue de 3% à 5% pour tous les modèles sur les données d'ImageNet.

problèmes inoubliables :

- 1) Si nous traitons des images avec la variance inconnue (c'est-à-dire nous ne pouvons pas savoir l'information sur le bruit) ?
- 2) Si nous traitons des images avec la variance hors de la plage de celles dans l'entrée des modèles (c'est-à-dire la variance qui est différente que celles données dans l'entrée des modèles) ?

Les deux problèmes sont réels et pratiques mais aussi difficiles à résoudre (voir la figure 22). Dans la suite, nous proposons trois méthodes qui peuvent renforcer la capacité de la généralisation des modèles.

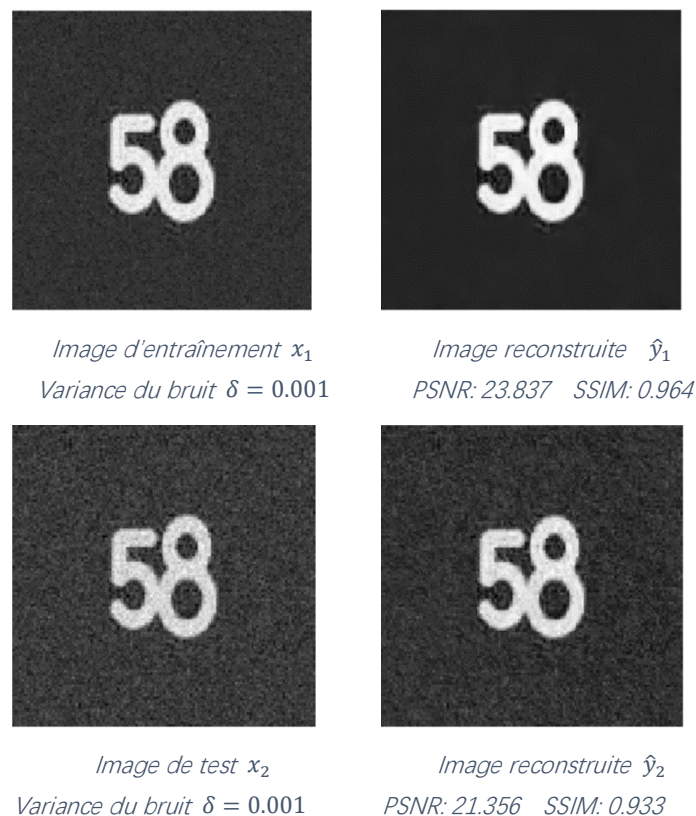


Figure 15 Démonstration sur l'insuffisance de la capacité de la généralisation des modèles. Prenons par exemple le Réseau des Filtres Dynamiques Locaux qui est entraîné par l'image x_1 et prend en sortie l'image reconstruite \hat{y}_1 , si nous testons ce modèle avec l'image x_2 , nous obtenons en sortie l'image \hat{y}_2 qui n'est pas bien débruitée. Donc, nous constatons que le modèle est surappris par le bruit spécifique d'entraînement.

Entraînement avec le bruit de variances variées

Au lieu de prendre en entrée les images bruitées avec la variance fixe, nous rendons la variance changeable dans une plage de valeurs. C'est considéré comme l'un des avantages concernant les images synthétiques d'où le bruit est mis selon la volonté subjective de l'humain. Ainsi, le modèle se rend plus robuste face aux images avec le bruit de variances différentes, ce qui est extrêmement utile dans la réalité où nous ne connaissons pas le bruit.

Carte d'Estimation du bruit

L'idée d'apprentissage de l'estimation du bruit est utilisée pour la première fois dans le papier qui introduit le modèle CBDNet[5]. Le papier montre qu'il est facile à estimer la variance du bruit quelques soient les images d'entrée en utilisant un sous-réseau de neurones. Le sous-réseau, qui prend en entrée les images bruitées et en sortie la carte d'estimation de la variance du bruit, est simplement une combinaison linéaire des couches convolutionnelles avec le résidu (voir la figure 12). La carte apprise est ensuite concaténée avec les images bruitées pour l'entrée du U-net.

L'ajout de la carte d'estimation du bruit sert à estimer l'information sur le bruit avant de débruiter les images, ce qui rend le modèle plus robuste et stable face aux images avec le bruit de variances différentes.

Modèle sans biais[7]

Les biais des modèles existent dans les deux parties : les couches de la normalisation des lots et les couches convolutionnelles. En effet, tous les nouveaux modèles de Deep Learning pour le débruitage d'imagerie risquent de se sur-adaptent systématiquement aux niveaux de bruit des images d'entraînement, ce qui est dû à la présence des biais dans les réseaux neurones.

Dans la plupart des modèles, pour une image bruitée d'entrée $y \in R^N$ avec N pixels, la fonction $f: R^N \rightarrow R^N$ calculée par un réseau de neurones peut être écrite comme :

$$f(y) = W_L R(W_{L-1} \dots R(W_1 y + b_1) + \dots b_{L-1}) + b_L = A_y y + b_y$$

D'où :

- $A_y \in R^{N \times N}$ est le Jacobien de $f(\cdot)$ évalué à l'entrée y
- $b_y \in R^N$ représente le biais du réseau de neurones

Nous notons que les modèles avec les biais ne sont pas invariants de mise à l'échelle à cause de l'existence des biais. Cependant, l'invariance de mise à l'échelle est intuitivement souhaitable pour une méthode de débruitage opérée sur les images naturelles.

Nous proposons un modèle sans biais pour le débruitage d'imagerie avec la fonction calculée qui peut être écrite comme :

$$f_{BF}(y) = W_L R(W_{L-1} \dots R(W_1 y)) = A_y y$$

Nous notons que ce modèle est strictement linéaire et donc invariant de mise à l'échelle. L'influence du biais est éliminée. Nous constatons que l'élimination des biais disposent d'une influence importante : les réseaux de neurones acquièrent la capacité de se généraliser à des niveaux de bruit non rencontrés lors de l'entraînement.

D'après le papier[7] qui sort cette idée, un schéma des normes en fonction du niveau de bruit est affiché pour démontrer l'effet négatif des biais :

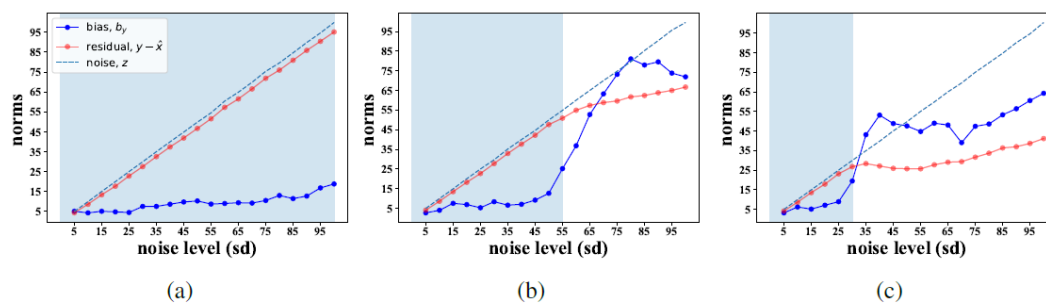


Figure 16 Analyse du résidu d'un réseau de neurones convolutif de débruitage en fonction du niveau de bruit. Les graphiques montrent les normes du résidu et du biais net calculées en moyenne sur 100 20x20 patchs d'images naturelles pour l'entraînement du réseau formé sur les différentes plages de formation. **(a)** Lorsque le réseau est entraîné sur toute la plage de niveau de bruit ($\delta \in [0, 100]$) le biais est faible, augmentant légèrement à mesure que le bruit augmente. **(b-c)** Lorsque le réseau est entraîné sur une plage plus petite ($\delta \in [0, 55]$) et ($\delta \in [0, 30]$), le biais augmente de manière explosive pour le bruit de niveau au-delà de la plage de formation. Le réseau de neurones utilisé pour cet exemple est DnCNN.

Un schéma des images peut aussi expliquer l'effet relatif entre résidu et le biais :

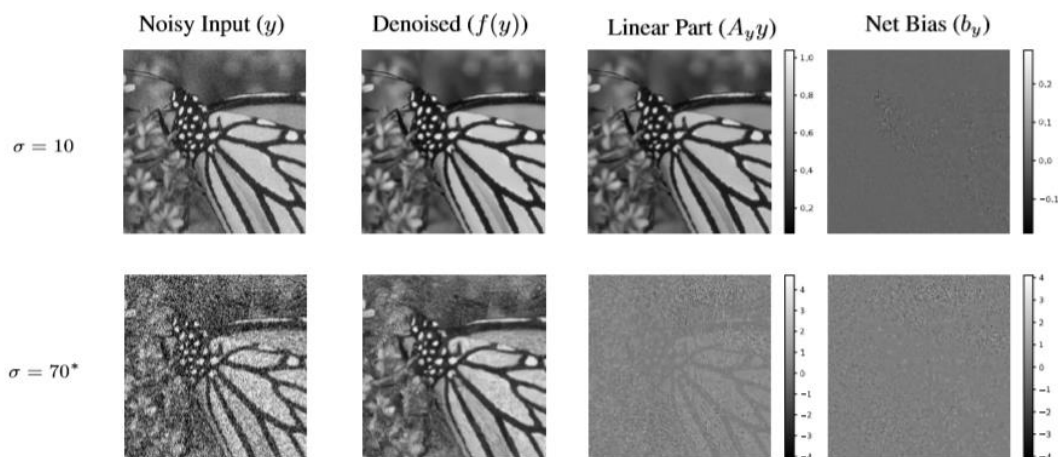


Figure 17 Visualisation de la décomposition de la sortie du réseau de neurones U-net qui. L'image bruitée en haut est dans la plage de niveau de bruit de l'entraînement alors que l'image bruitée en bas est en dehors de la plage. Nous constatons que sur la plage d'entraînement, le biais net est faible et la partie linéaire est responsable de la majeure partie de l'effort de débruitage. Cependant, quand le réseau est évalué hors de plage d'entraînement, la contribution du biais augmente considérablement, ce qui coïncide avec une baisse significative de la performance du débruitage.

Nous allons montrer nos résultats dans le chapitre suivant : La Présentation des Résultats de Deep Learning.

VII Présentations des Résultats de Deep Learning

Dans cette partie, nous montrons les résultats expérimentaux.

Choix pour l'Entraînement des Modèles

Nous expliquons les choix optionnels qui optimisent le cours d'entraînement des modèles. Les choix optionnels sont la fonction du coût et la manière d'initialisation des poids des modèles.

Fonction du coût

Nous testons trois types de fonctions du coût : l'erreur moyenne quadratique (L2), l'erreur moyenne absolue (L1) et l'erreur moyenne combinée. Notons que l'erreur moyenne combinée est la combinaison de L2 et L1 qui sont respectivement appliquées aux pixels et aux gradients des pixels.

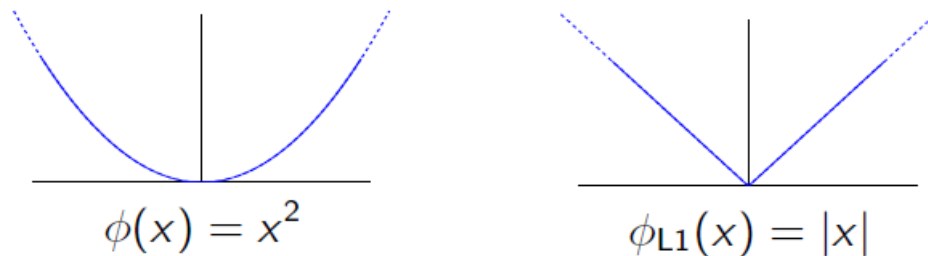


Figure 18 Schéma de L2 (à gauche) et L1 (à droite)

En effet, L2 est une fonction lisse qui pénalise plus les grandes différences mais moins les petites différences, ce qui accélère la convergence de l'entraînement. Néanmoins, L1 est une fonction dure qui pénalise de même niveau toutes les différences. Même si L1 converge moins vite, ceci a plus de capacité de garder les bords, qui est l'information la plus importante pour la mission du débruitage. Un schéma ci-dessous du l'effet de ces deux termes peut expliquer la différence :

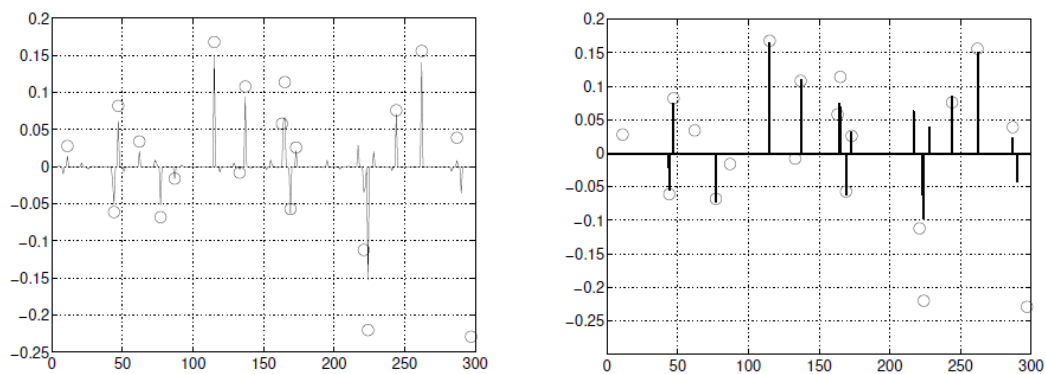


Figure 19 Schéma du l'effet de L2 (à gauche) et L1 (à droite)

Les cercles représentent les pixels et les pics représentent les valeurs après la pénalisation

Source : cours IMINV, Chap 3 déconvolution parcimonieuse, DATASIM, 2020

Donc, théoriquement, nous préférons L1 que L2 afin de garder les bords des objets dans les images. De plus, pour ne pas perdre trop de détails, nous pensons joindre L2. Donc, l'erreur moyenne combinée[8] est créée. Nous testons ces trois fonctions du coût sur les données d'ImageNet et nous visualisons la différence :

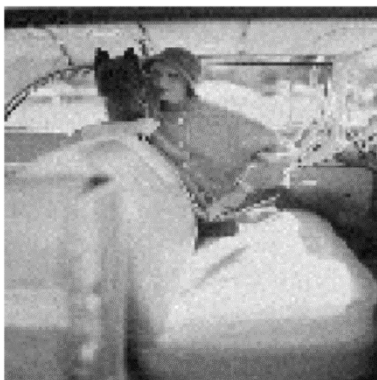


Image bruitée



L2

PSNR : 29.382

SSIM : 0.934



L1

PSNR : 30.013

SSIM : 0.991



Coût combiné

PSNR : 30.312

SSIM : 0.992

Conclusion : nous constatons que théoriquement et statistiquement, L2 fonctionne pire que L1 ou le coût combiné parce que l'image reconstruite devient plus floue en perdant un peu de détails, ce qui n'est pas ce que nous souhaitons. Donc, en gardant l'idée du débruitage d'imagerie que le bruit est éliminé alors que les détails sont gardés, nous choisissons éventuellement le coût combiné pour toutes les expériences à la suite.

Initialisation

L'initialisation est énormément importante et influence beaucoup les résultats. Certaines stratégies d'initialisation non seulement ralentit la vitesse de convergence mais aussi influence le résultat final de manière négative.

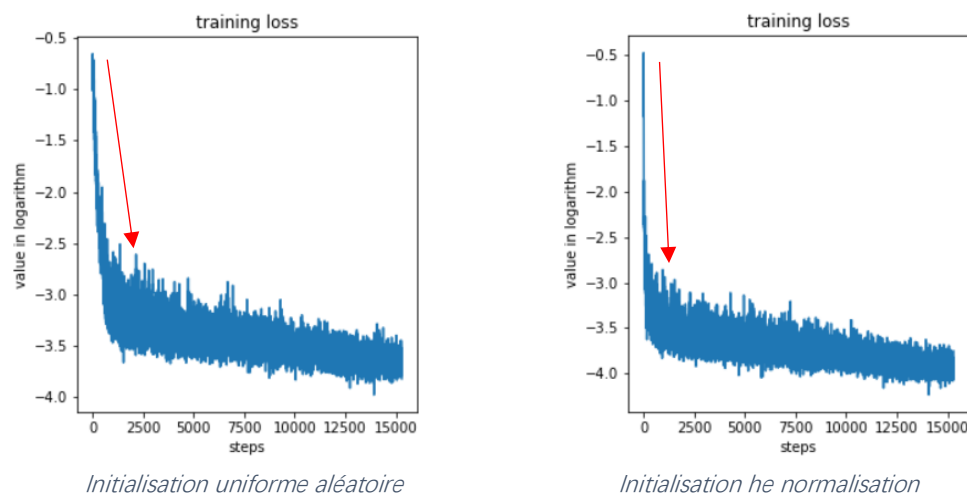


Figure 20 Schéma du terme du coût en fonction d'étapes des initialisations différentes

La figure 20 montre que l'initialisation pose une influence inoubliable sur la vitesse de convergence. Dans notre cas, nous préférons utiliser he normalisation comme initialisation en raison de l'existence de plusieurs fonctions d'activation ReLU dans le réseau de neurones et de la vitesse de convergence plus vite.

D'ailleurs, nous pouvons aussi entraîner plusieurs fois et moyenner les poids afin de réduire la variance et obtenir un modèle plus juste.

Préapprentissage

Vue que nous avons des milliers de données de DIV2K alors que les données d'ultrason sont spécifiques et rares en fonction de la confidentialité des clients et de la particularité des opérations hospitalisées. Donc, dans la vie normale, nous n'avons pas assez de données d'ultrason pour l'entraînement des modèles. En revanche, nous choisissons le préapprentissage des modèles par les données DIV2K et ensuite réaliser l'apprentissage

transféré des mêmes modèles par les données d'ultrason en considérant que dans n'importe quel domaine d'imagerie, les modèles ont tendance d'apprendre les filtres similaires pour procéder le débruitage.

Nous décidons donc deux étapes pour réaliser le débruitage des données d'ultrason en utilisant le Réseau des Filtres Dynamiques Locaux :

- 1) Le préapprentissage du modèle par les 1000 données de DIV2K
- 2) Le réglage fin (en anglais : fine-tuning) du modèle par certaines données objectives d'ultrason.

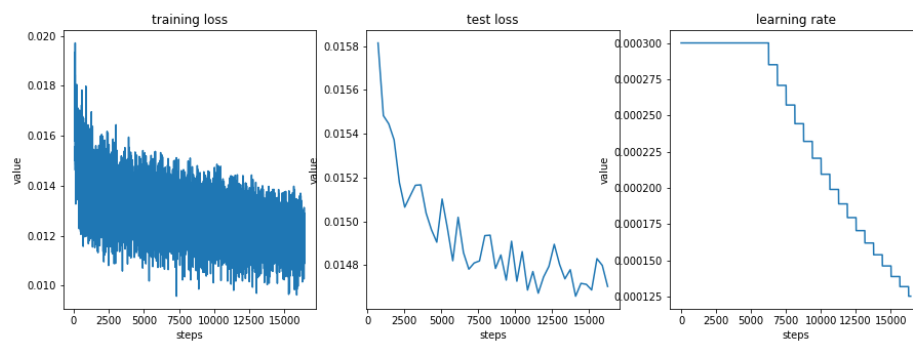


Figure 21 Etape 1) préapprentissage du modèle. Nous exigeons le taux d'apprentissage haut au début et le diminue en fonction d'étapes. Le coût d'apprentissage et de test baissent totalement.

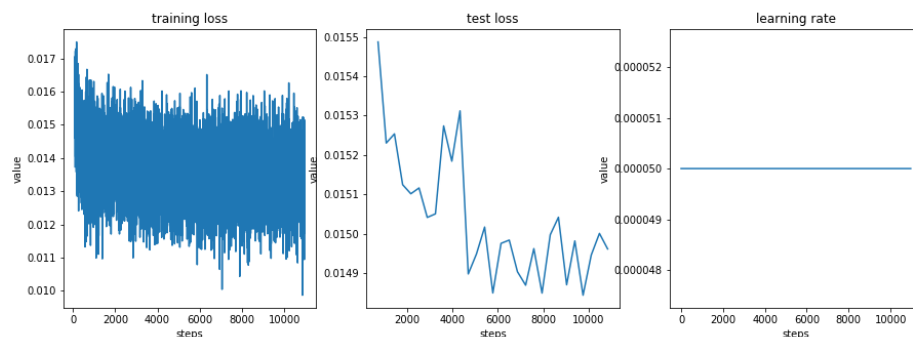


Figure 22 Etape 2) réglage fin. Nous maintenons le taux d'apprentissage bas pour régler un petit peu les paramètres du modèle. Le coût d'apprentissage et de test baissent aussi totalement mais lentement, ce qui signifie que le modèle réussit à s'adapter aux nouvelles données.

Éventuellement, si nous comparons la performance du modèle qui s'entraîne par les deux étapes et celui qui ne s'entraîne que par les certaines données d'ultrason, nous constatons sûrement que le premier modèle fonctionne beaucoup mieux.

Comparaison des Modèles

Dans cette partie, nous envisageons de comparer la performance des modèles en affichant les résultats visuels et statistiques.

U-net, Réseaux des Filtres Dynamiques Globaux et Locaux

Dans cette partie, nous comparons les trois modèles : U-net, le Réseau des Filtres Dynamiques Globaux et le Réseau des Filtres Dynamiques Locaux. Pour que c'est soit juste, la structure de U-net est strictement la même que celle dans le Réseau des Filtres Dynamiques Locaux, ou est partagée d'une semi partie dans le Réseau des Filtres Dynamiques Globaux, d'où le nombre des paramètres est presque le même dans les trois modèles.

Les données utilisées sont les données d'ImageNet, comptant 2000 pour l'entraînement et 500 pour le test. Le bruit synthétisé est le bruit blanc gaussien avec la variance égale à 0.05.

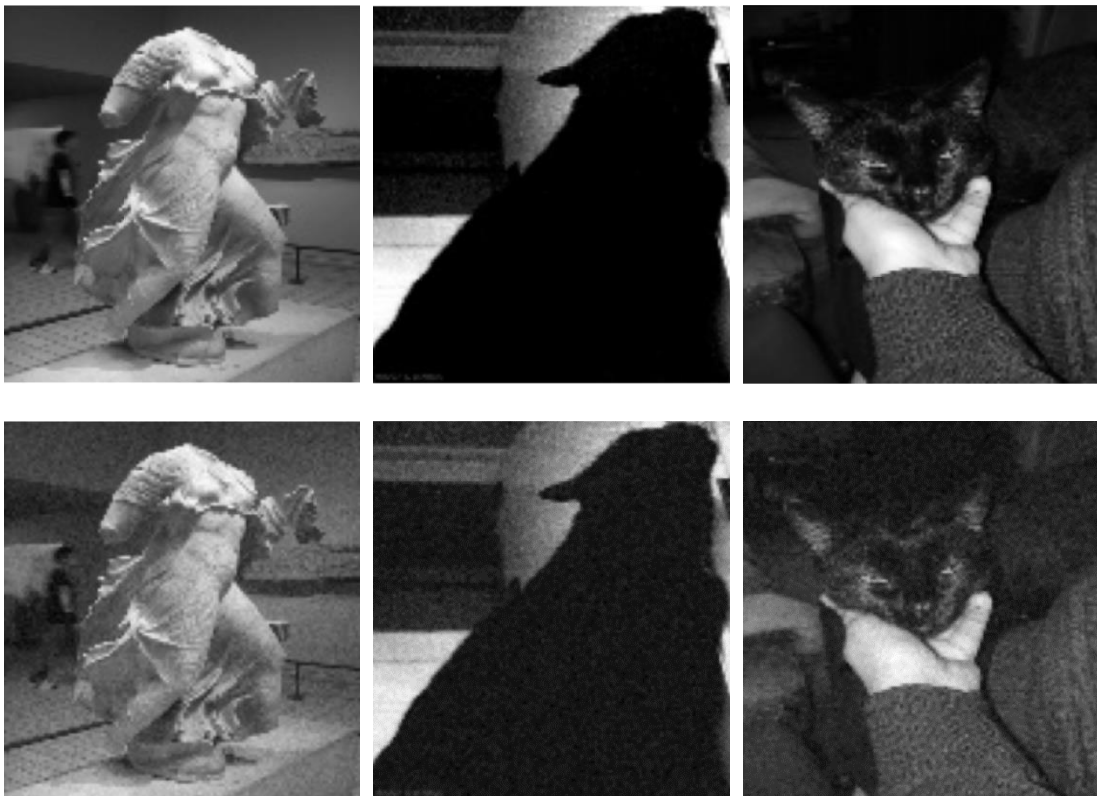


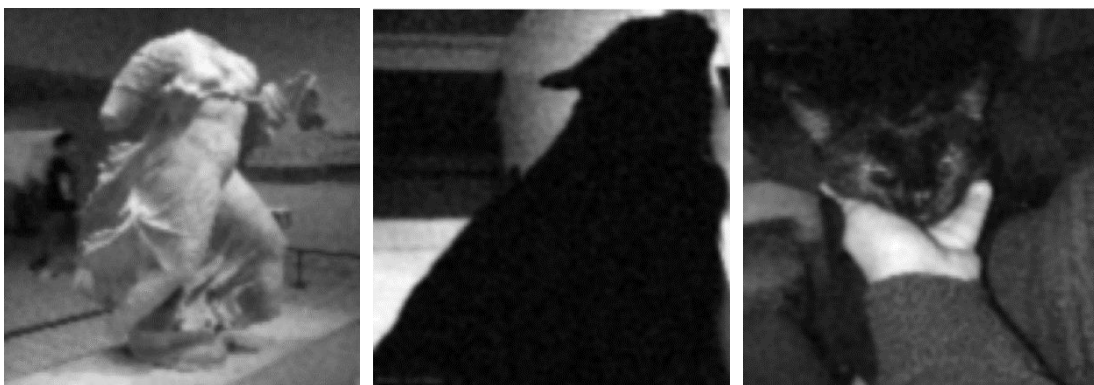
Figure 23 Exemples d'ImageNet des vérités terrains (en haut) et des images bruitées (en bas)

Nous mettons les données dans les modèles et nous obtenons les résultats ci-dessous :



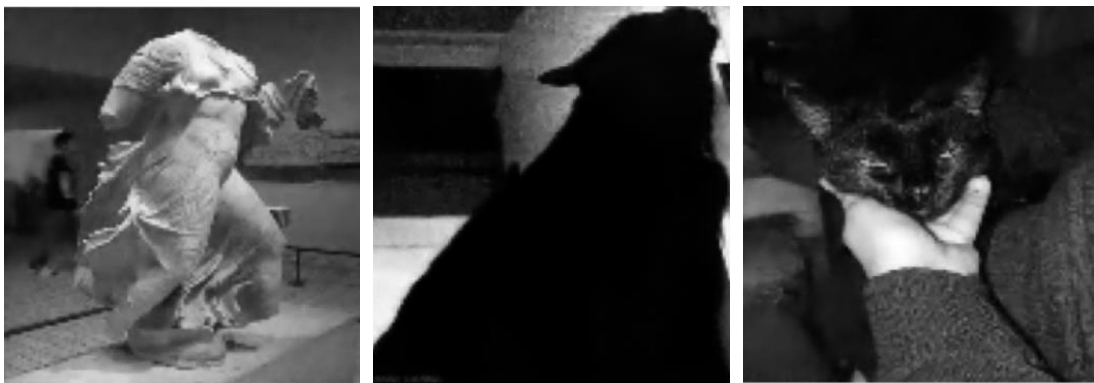
U-net

PSNR 24.475 SSIM 0.988



Réseau des Filtres Dynamiques Globaux

PSNR 23.872 SSIM 0.961



Réseau des Filtres Dynamiques Locaux

PSNR 30.409 SSIM 0.992

Figure 24 Les images reconstruites par les trois modèles

Nous constatons que :

- 1) U-net fonctionne déjà bien à l'aide de sa capacité forte de la reformulation de l'information mais cela n'est pas suffisant parce qu'il existe des lignes d'interférence, ce qui signifie que le modèle arrive facilement au point d'optimisation locale.
- 2) Le Réseau des Filtres Dynamiques Globaux ne fonctionne pas bien. C'est probablement parce que nous n'arrivons pas à apprendre un filtre global

représentatif qui peut marcher partout dans une image pleine. Il faut encore étudier la structure de ce réseau de neurones.



Figure 25 Visualisation des 5 filtres appris qui sont appliqués successivement aux images entières de manière convolutionnelle. Constatons que c'est difficile d'apprendre les filtres globaux qui sont réguliers et représentatifs pour tous les pixels des images d'entrée

- 3) Le Réseau des Filtres Dynamiques Locaux arrive au meilleur résultat à cause de son mécanisme local, d'où chaque filtre est spécifique par rapport à un seul pixel. Le réseau a donc une grande capacité d'adaptativité.

Les Autres

Nous n'étudions que la faisabilité du réseau DnCNN, FFDNet et CBDNet parce que les idées créatives de ces modèles sont tous empruntées par les modèles présentés dans la partie précédente. Nous obtenons des conclusions ci-dessous :

- 1) La normalisation des lots n'est pas utile pour la mission du débruitage d'imagerie. (DnCNN)
- 2) La concaténation des images sous-échantillonnées rend le modèle plus efficace et précis. (FFDNet)
- 3) La concaténation d'une carte d'estimation du niveau du bruit aide à augmenter la capacité de généralisation des modèles de Deep Learning. (FFDNet)
- 4) C'est facile pour qu'un réseau de neurones apprenne une carte d'estimation de bruit. (CBDNet)

Amélioration du Réseau des Filtres Dynamiques Locaux

Ce modèle nous intéresse le plus en raison de sa meilleure performance et sa possibilité de la fusion des méthodes traditionnelles. Donc, nous pensons à l'améliorer en testant plusieurs composantes introduites auparavant qui sont probablement utiles :

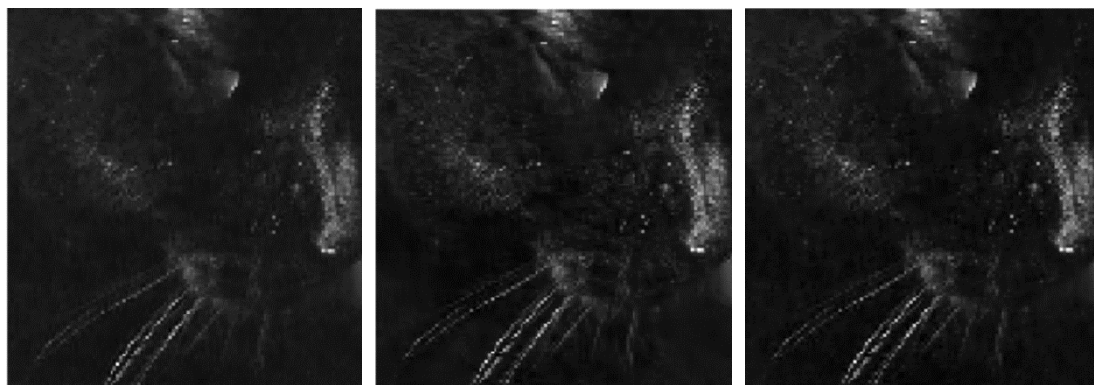
- 1) La carte d'estimation de biais.
- 2) Attention spatiale
- 3) Superposition simultanée de plusieurs filtres locaux de tailles différentes

L'idée 1) sera testée dans le chapitre suivant. Afin de tester ces deux idées dernières, nous ajoutons les composantes au modèle basé (le Réseau des Filtres Dynamiques Locaux) et comparer leur performance sur les données d'ImageNet d'où le bruit est gaussien blanc avec

la variance égale à 0.05:



Figure 26 Exemple d'ImageNet de vérité terrain (à gauche) et de l'image bruitée (à droite)



Modèle basé

Filtre de taille 3x3

PSNR 29.897

SSIM 0.991

Modèle avec

Filtres de tailles 3x3 et 5x5

PSNR 30.013

SSMI 0.991

Modèle avec

Filtres de tailles 3x3 et 5x5

Attention spatiale

PSNR 30.663

SSMI 0.993

Figure 27 images reconstruites

Conclusion :

- 1) Les progrès des deux idées s'affirment en pouvant contribuer visuellement et statistiquement à l'effet du débruitage.
- 2) L'idée de la superposition simultanée de plusieurs filtres locaux de tailles différentes donne au modèle un champs réceptif plus grand pour chaque pixel et rend le modèle plus robuste et puissant.
- 3) L'idée de l'attention spatiale aide les filtres à s'adapter plus intellectuellement à leur pixel correspondant en accordant plus d'attention aux pixels voisins similaires. De ce point de vue, nous arrivons à garder plus le bord des objets dans les images d'entrée.

Reportage des Études de la Capacité de la Généralisation

La capacité de la généralisation des modèles de Deep Learning reste toujours un sujet critique parce que dans la réalité, nous souhaitons que les modèles soient adaptables aux situations variées et réelles. Dans cette partie, nous pensons à tester les trois techniques introduites auparavant qui visent à améliorer la capacité de la généralisation des réseaux de neurones :

- 1) Entraînement avec le bruit de variances variées
- 2) Carte d'Estimation du bruit
- 3) Modèle sans biais

Les données utilisées proviennent d'ImageNet avec le bruit de variance différente afin de tester si le modèle s'y adapte. La variance du bruit, qui varie entre 0.001 et 0.01, se répartie uniformément dans les données d'entraînement alors que celle des données de test est fixée à trois valeurs : 0.001, 0.05 et 0.01.

Nous proposons les quatre essais pour la comparaison :

- 1) Les données d'entraînement sont fixées du bruit de la variance de 0.025. (C'est une inférence)
- 2) Les données d'entraînement sont attribuées du bruit de la variance entre 0.005 et 0.05.
- 3) Les données d'entraînement sont attribuées du bruit de la variance entre 0.005 et 0.05. De plus, une carte d'estimation de la variance du bruit est apprise par un réseau de neurones et concaténée aux images d'entrée.
- 4) Les données d'entraînement sont attribuées du bruit de la variance entre 0.005 et 0.05. De plus, le modèle est éliminé de tous les biais.





Figure 28 Résultats des données de test sur les quatre essais 1) 2) 3) 4) avec les quatre lignes d'affichage respectives. Le modèle étudié est le Réseau des Filtres Dynamiques Locaux.

Nous constatons donc :

- 1) Un modèle qui prend en entrée le bruit de variance différente atténue le surapprentissage d'un certain bruit et fonctionne bien face au bruit d'une grande plage de variance. (2^{ème} ligne)
- 2) La concaténation d'une carte d'estimation du bruit donne plus d'information au modèle, ce qui est énormément efficace. (3^{ème} ligne)
- 3) Les réseaux de neurones convolutifs sans biais profitent bien de la suppression du biais et se traduisent par une capacité de la généralisation efficace en niveaux de bruit hors de la plage d'entraînement.

VIII Pour Aller Plus Loin: Modèle hybride

Dans cette partie, nous choisissons le Réseau des Filtres Dynamiques Locaux avec la taille des filtres en 3x3 comme le modèle basé. Nous pensons aussi à ajouter des contraintes pour que le Réseau des Filtres Dynamiques Locaux avec les tailles des filtres respectivement en 3x3, 5x5 et 7x7 (il y a donc une superposition simultanée de trois filtres) marchent le contrôle manuel.

Analyse des Filtres

Afin d'étudier le modèle hybride, nous avons besoins d'exploiter le mécanisme de la création des filtres par Deep Learning. Nous adoptons donc les outils mathématiques et de visualisation pour pouvoir expliquer le comportement des filtres.

Premièrement, nous calculons la valeur moyenne de tous les filtres pour avoir une impression globale. Nous constatons que les filtres ont une forme gaussienne, ce qui est raisonnable et explicable par le filtre gaussien traditionnel qui joue le rôle fondamental de débruitage.

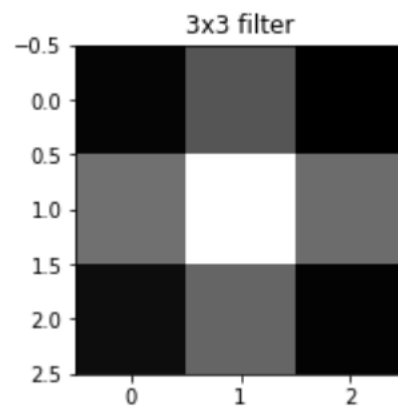


Figure 29 Visualisation globale des filtres

Cependant, même si les filtres sont de forme similaire, ils sont distribués des valeurs différentes avec un écart-type inoubliable :

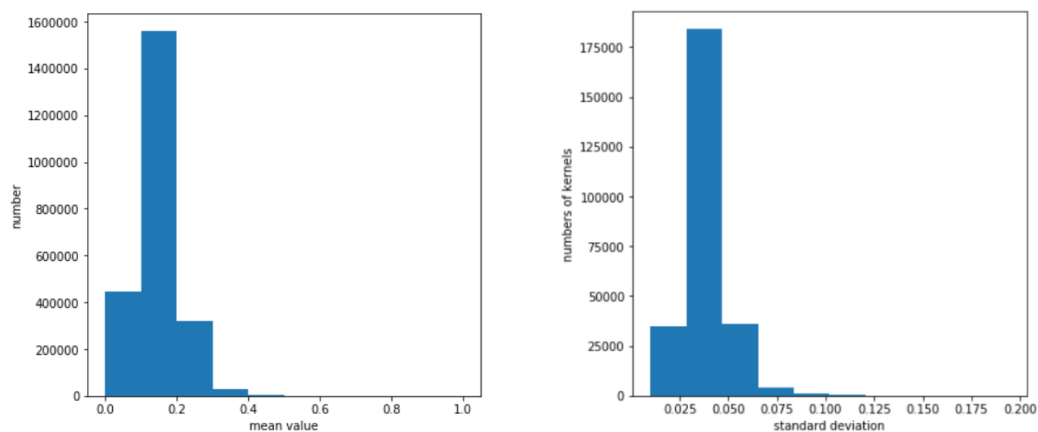


Figure 30 Distribution de la valeur moyenne et l'écart-type des filtres

Deuxièmement, pour étudier plus profondément la différence et la similarité parmi les filtres, nous utilisons l'Analyse des Composantes Indépendantes. Nous affichons les deux premières composantes indépendantes de tous les filtres pour visualiser leurs caractéristiques.

Nous constatons que l'Analyse des Composantes Indépendantes semble distinguer un petit peu la différence des filtres (les points jaunes, verts et les autres ont une divergence). D'ailleurs,

la plus part des filtres ont des caractéristiques similaires (les points noirs, bleus, oranges, violets etc...).

Troisièmement, nous faisons le clustering des filtres afin de trouver leur différence et similarité. Nous utilisons le modèle de mélange gaussien (en anglais : gaussian mixture model) pour faire du regroupement des filtres. Le nombre des clusters est mis en 9.

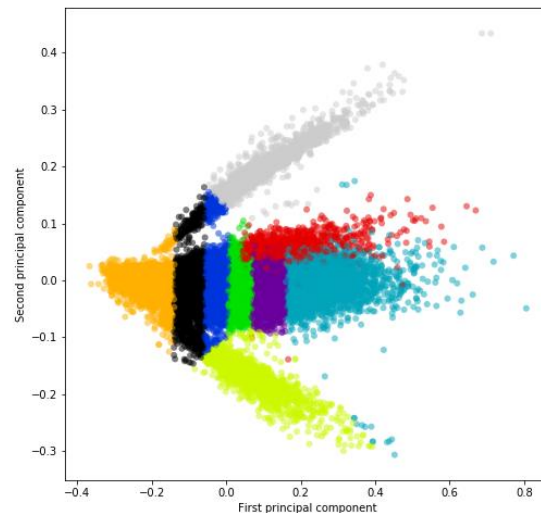


Figure 31 Visualisation des caractéristiques des filtres. Les couleurs représentent le résultat du regroupement par le modèle de mélange gaussien.

Quatrièmement, afin de visualiser la forme des filtres en regroupement, nous choisissons d'utiliser la décomposition de valeur singulière (en anglais : singular value decomposition). Ce qu'il faut noter, c'est qu'un filtre de taille 3x3 possède 9 caractéristiques, d'où nous pouvons donc tirer 9 vecteurs propres qui représente les formes de tous les filtres :

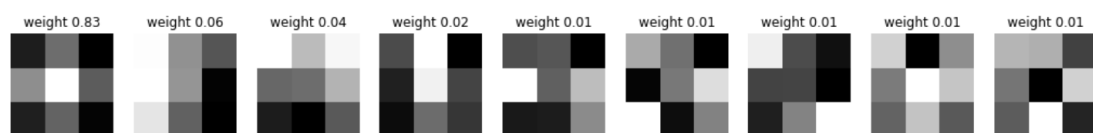


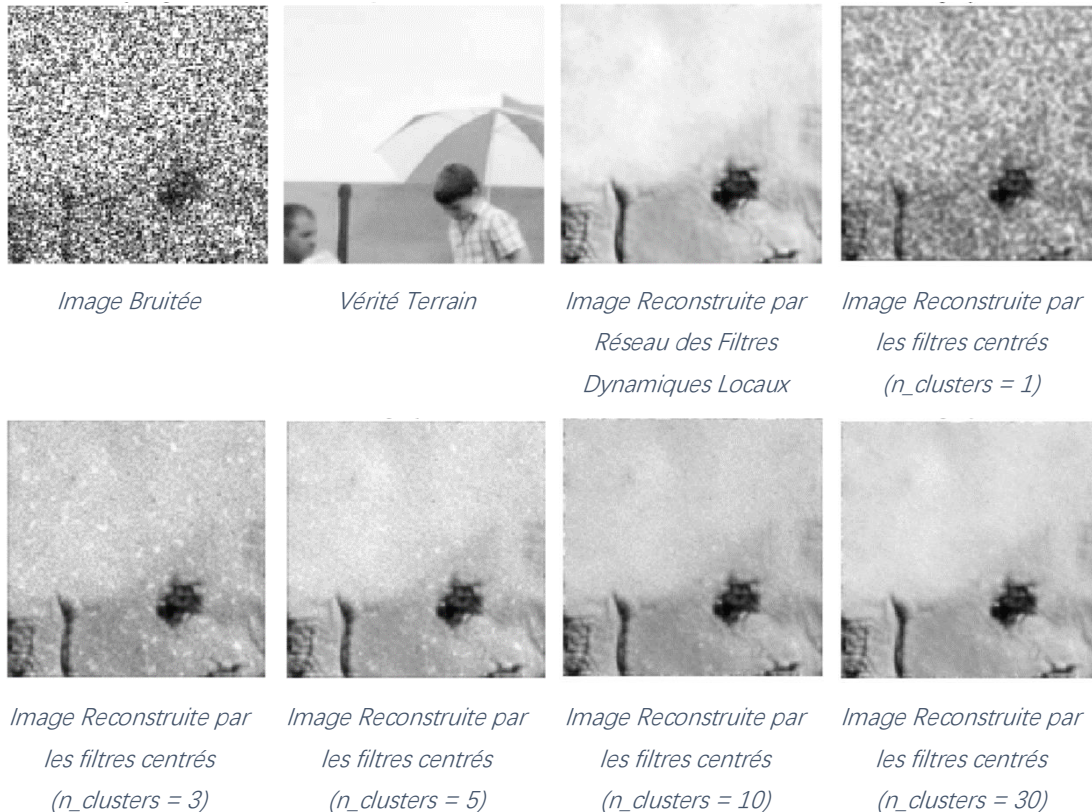
Figure 32 Visualisation des vecteurs propres tirés par la DVS

Nous constatons que les filtres similaires de la forme gaussienne (le 1er vecteur propre) comptent le plus, suivis par ceux de formes particulières (les autres vecteurs propres) qui sont liés avec des pixels spéciaux. Nous concluons donc que la plupart des filtres jouent le rôle d'une mode gaussienne mais aussi se distinguent dans les petites différences, ce qui montre l'adaptativité et la robustesse du Réseau des Filtres Dynamiques Locaux.

Fonction Corrélée entre Patches et Filtres

Nous réalisons qu'il y a des millions de filtres appris pour les images d'entraînement. Notons que nous avons par exemple 100 images de taille 128 carrée, nous obtenons donc

$100 \times 128 \times 128 \approx 1,600,000$ filtres au total, ce qui est, d'après nous, redondant. Afin de vérifier notre hypothèse, nous faisons le regroupement des filtres et remplacer tous les filtres par leur centre des filtres regroupé correspondant. Nous utilisons le modèle de mélange gaussien (en anglais : gaussian mixture model) pour faire du regroupement des filtres. Nous visualisons les résultats des images débruitées par l'application des filtres centrés :



Ce qu'il faut souligner, c'est que $n_clusters$ signifie le nombre de regroupement par le modèle de mélange gaussien et est équivalent donc au nombre des filtres centrés totaux qui sont appliqués sélectivement à tous les patches. Dans ce cas, un **patch** signifie un entourage d'un pixel de la même taille qu'un filtre. Nous constatons que plus grand est le nombre de regroupement, plus proches sont les images reconstruites par les filtres centrés à celles par le Réseau des Filtres Dynamiques Locaux. C'est évident parce que les dernières sont débruitées par des millions des filtres par rapport aux premières débruitées que par certains filtres (le chiffre des filtres est égal à $n_clusters$). Les résultats sont impressionnants parce que nous réussissons à débruiter les images d'entrée par seulement un certain nombre de filtres. Si nous choisissons 30 filtres et les appliquer aux pixels de manière correcte, nous arrivons à presque une même performance élevée que le Réseau des Filtres Dynamiques Locaux.

Ensuite, il faut réfléchir sur comment tirer certains filtres représentatifs et comment mettre le lien entre ces filtres et tous les patches. Si nous arrivons à trouver une fonction corrélée entre eux, nous pourrions ensuite choisir certains filtres représentatifs et nous serons capables de réaliser le débruitage en n'utilisant que ces filtres!

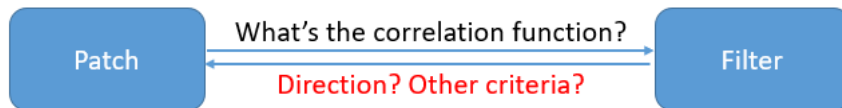


Figure 33 Pensée au lien entre les patches et les filtres

Nous construisons donc les étapes suivantes pour réaliser un modèle hybride :

- 1) Choisir des millions de filtres et leurs filtres correspondants appris par le Réseau des Filtres Dynamiques Locaux.
- 2) Envisager de trouver une fonction de corrélation entre les patches et les filtres.
- 3) Tirer certains filtres représentatifs par un algorithme de regroupement lié avec la fonction de corrélation.
- 4) Étant donné un nouvel patch, choisir un filtre parmi les filtres tirés par l'étape 3).

Nous pensons donc à trouver une fonction corrélée. En considérant que les patches peuvent avoir une direction ainsi que les filtres, nous réfléchissons à la possibilité de les regrouper par la direction (nous décidons qu'il y a au total 8 directions). Nous visualisons le résultat de regroupement :

Nous constatons que les filtres de forme gaussienne partagent un poids assez élevé, ce qui nous empêche de trouver les filtres de forme représentative sauf la forme gaussienne. Ainsi, il faut trouver une fonction de corrélation plus intelligente.

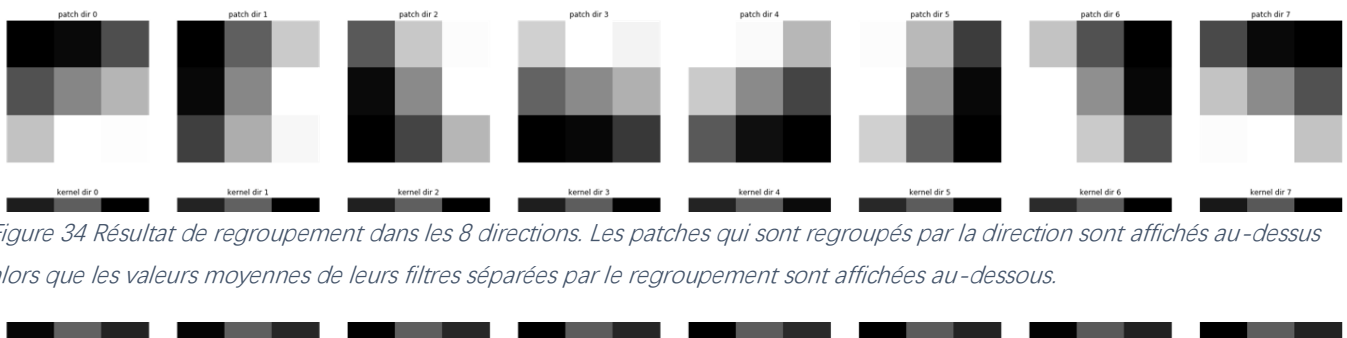


Figure 34 Résultat de regroupement dans les 8 directions. Les patches qui sont regroupés par la direction sont affichés au-dessus alors que les valeurs moyennes de leurs filtres séparées par le regroupement sont affichées au-dessous.

Une méthode traditionnelle, d'après nous, a une tendance de toujours tirer les filtres gaussiens. Nous pensons donc à une méthode automatique à l'aide des outils d'apprentissage non supervisé. Nous choisissons donc le modèle de mélange gaussien pour effectuer le regroupement automatique des patches et trouver ensuite les filtres représentatifs en calculant les valeurs moyennes des filtres regroupés d'après leurs patches.

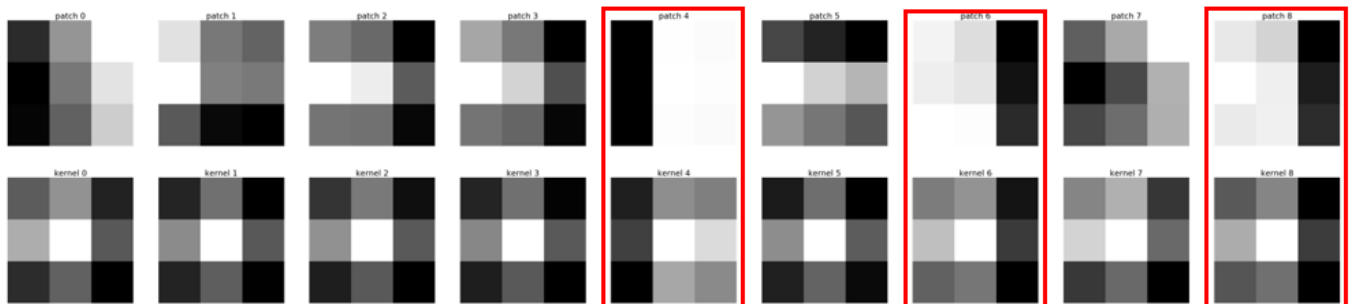


Figure 35 Résultat de regroupement des patches par le modèle de mélange gaussien (dans ce cas, le nombre de clusters est 9). Les patches qui sont regroupés sont affichés au-dessus alors que les valeurs moyennes de leurs filtres séparées par le regroupement sont affichées au-dessous.

Nous constatons qu'à l'exception du tirage des filtres gaussiens, il y a encore des filtres tirés qui sont capables d'évoquer la direction des patches d'un point de vue (les filtres sont cerclés des cadres rouges). Nous arrivons donc à trouver des filtres représentatifs. Nous concluons que :

- 1) Plus grand est le nombre des clusters, plus de filtres représentatifs sont trouvés.
- 2) Plus grand est le nombre des clusters, plus grand est PSNR dans pour les données de test.

Cependant, nous admettons que le modèle hybride établie dans la base du modèle de mélange gaussienne rencontre des difficultés à atteindre les bons résultats :

- 1) Les filtres gaussiens tirés comptent encore le plus et pose une influence négative à trouver les filtres d'autres formes.
- 2) La réalisation du débruitage avec moins de cent filtres n'a pas d'effet favorable. Cependant, plus grand est le nombre de clusters, plus de temps est consommé pour le regroupement, plus difficile est le modèle de mélange gaussienne de terminer la boucle de l'algorithme, ce qui est coûteux.

Pour conclure, nous remarquons qu'une fonction corrélée moins coûteuse et plus efficace reste à rechercher. Du point de vue de la durée serrée du stage, le modèle hybride reste encore à étudier dans l'avenir.

IX Conclusion

A la fin de ce rapport, je fais une conclusion pour mon stage de quatre mois chez Philips Research Suresnes à partir de ces 4 aspects : capacité professionnelle, soft skills, fonctionnement d'entreprise et perspectif de carrière.

Franchement dit, bien que la mission du débruitage médical cible les clients dans les hôpitaux, je n'ai pas traité assez d'affaires qui nécessitent l'interaction avec les clients. Puisque le département à Suresnes est par essence un département de recherche, il ne faut qu'un esprit pur pour l'innovation des idées. En bref, c'est une meilleure opportunité de pratiquer les connaissances au lieu de perfectionner la capacité professionnelle.

Dans mon équipe, j'ai travaillé avec deux collègues qui sont aussi mes tuteurs. A part des communications quotidiennes, j'ai des occasions de contacter d'autres gens venant d'autres équipes, comme le département de business aux États-Unis. Ainsi, les présentations inter-équipes sont demandées. Grâce aux chances précieuses, je me suis entraîné par rapport de capacité de coordination et communication. De plus, j'ai appris le moyen de diriger un projet.

L'une de plus précieuses obtentions que j'ai acquise de ce stage est une vision globale. Du fait que Philips est une entreprise très connue à l'échelle du monde, au fur et à mesure de son développement, au sein de l'entreprise, un système de gestion et d'organisation des employés est établi afin de rendre le fonctionnement de l'entreprise plus efficace. Les employés de première ligne disposent d'une grande autonomie et un droit de décision par rapport au gens dans le HQ. Car ils affrontent le business directement et donc ils connaissent bien les vraies situations. Dans ce cas-là, moins de temps sera gaspillé puisqu'il n'y a pas trop de négociations entre le HQ et chaque bureau de représentation. Donc Philips dispose une plus haute efficacité en ce mode.

De plus, à l'aide de ce poste, j'ai connu l'importance de l'innovation dans le monde entier car Philips est un des leaders dans les techniques de la santé. Donc, c'est vraiment une excellente occasion pour moi d'accéder à l'essence des équipes et communiquer avec les autres gens pour connaître le développement le plus récent dans ce domaine, ce qui m'aide au perspectif de carrière. D'ailleurs, les besoins de traiter une masse de données existent aujourd'hui dans presque toutes les industries. En conséquent, la capacité de savoir comment traiter et analyser les données est primordiale. C'est aussi la raison pourquoi j'ai choisi mon option DataSIM. J'espère que je pourrai intégrer cette compétence dans mon perspectif de carrière.

X Remerciement

En ce qui me concerne, en général, je fais du travail de l'innovation des idées pour le débruitage médical, implémente les codes, teste leur faisabilité et m'échange avec les autres au sein de mon équipe. Je voudrais sincèrement exprimer mon remerciement à tout ce qui m'aide pendant mon stage. Je ne peux pas bien me réaliser sauf leur aide.

D'abord, c'est pour moi un grand honneur de travailler pour Philips Research Suresnes pendant 17 semaines, du début de mai à la fin d'août. Étant donné que stagiaire dans l'équipe concernant le débruitage d'imagerie, j'ai trois collègues, Oudom, Segbedji et Bo. Ils sont tous scientifiques seniors dans ce domaine et sont chargés d'innover et créer des méthodes non seulement efficaces mais aussi interprétables. Ils ont tous un bon sens vers le développement le plus récent et connaissent bien l'opération des projets. J'ai beaucoup appris l'évaluation des projets grâce à eux. Ils m'ont aussi enseigné beaucoup sur l'organisation de présentation et la rédaction de rapport.

D'ailleurs, j'ai de la chance de travailler avec d'autre stagiaires. Ils sont tous vraiment gentils et nous nous entraînons dans la carrière professionnelle. Comme il s'agit de ma première expérience professionnelle dans une grande entreprise, j'ai rencontré aussi des difficultés concernant les procédures administratives. Par conséquent, je dois apprendre de nouvelles choses du niveau le plus élémentaire et, progressivement, je peux gérer les affaires réelles.

XI Référence

- [1] LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* 521, 436–444 (2015)
- [2] Zhang, Kai, et al. "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising." *IEEE Transactions on Image Processing* 26.7 (2017): 3142-3155.
- [3] Zhang, Kai, Wangmeng Zuo, and Lei Zhang. "FFDNet: Toward a fast and flexible solution for CNN-based image denoising." *IEEE Transactions on Image Processing* 27.9 (2018): 4608-4622.
- [4] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." *International Conference on Medical image computing and computer-assisted intervention*. Springer, Cham, 2015.
- [5] Guo, Shi, et al. "Toward convolutional blind denoising of real photographs." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2019.
- [6] Jia, Xu, et al. "Dynamic filter networks." *Advances in neural information processing systems*. 2016.
- [7] Mohan, Sreyas, et al. "Robust and interpretable blind image denoising via bias-free convolutional neural networks." *arXiv preprint arXiv:1906.05478* (2019).
- [8] Mildenhall, Ben, et al. "Burst denoising with kernel prediction networks." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018.