Ministère de l'enseignement Supérieur et de la recherche Scientifique

وزارة التعليم العالي والبحث العلمي

Badji Mokhtar Annaba University
Université Badji Mokhtar – Annaba
Faculté de Technologie

Département Electronique

جامعة باجي مختار – عنابـــة

كـليـة التكـنـولـوجيـا

قســم الإلكترونيك

Thèse

Présentée pour obtenir le diplôme de

Doctorat Troisième Cycle

Filière : Télécommunication

Spécialité : Multimédia et Communication Numérique

Par :

KEBIR Abdeldjalil

Thème :

Segmentation Fond/Forme par Apprentissage machine pour la Détection d'Objets en Vidéosurveillance et l'Extraction de Routes par vues Aériennes.

Thèse soutenue le 20 juillet 2023 devant le jury composé de :

N°	Nom et prénom	Grade	Etablissement	Qualité
01	BOUGHAZI Mohamed	Prof	Université Badji Mokhtar -Annaba	Président
02	TAIBI Mahmoud	Prof.	Université Badji Mokhtar -Annaba	Rapporteur
03	TALEB-AHMED Abdelmalik	Prof.	U - UPHF-Valenciennes – FRANCE	Co-rapporteur
04	ZITOUNI Athmane	Prof	Université - Mohamed Khider – BISKRA	Examinateur
05	BOUTERAA Nadia	MCA	Université Badji Mokhtar -Annaba	Examinateur
06	OUAFI Abdelkrim	Prof	Université - Mohamed Khider – BISKRA	Invité

تجزئة المقدمة / الخلفية باستخدام التعلم الآلي لاكتشاف الأجسام في المر اقبة " " بالفيديو واستخراج شبكة الطرقات من الرؤية الجوية.

الملخص:

في هذه الأطروحة، ندرس إشكالية تجزئة المنطقة الأمامية والتي تمثل الكائنات المتحركة في المراقبة بالفيديو وتجزئة المنطقة الأمامية في سياق الصورة الجوية لاستخراج شبكة الطرق بواسطة نموذج التعلم العميق الذي تم تطويره للتجزئة التلقائية على أساس متعدد العمق (مضغوط) من خلال نظام التشفير التلقائي، يتم تدريب هذه الأنظمة من الصفر ومن خلال التعلم بالنقل في الوضع الخاضع للإشراف عند استخدام بيانات تدريب محدودة. يعتمد الهدف الرئيسي للأنظمة المقترحة على طرق زيادة البيانات الجديدة، مبنية على التحولات الهندسية من خلال تحليل المكونات الرئيسية والتحولات القائمة على المرشحات (اكتشاف الحواف، وتحسين التفاصيل، وتحسين الحواف)، لإنشاء تمثيلات جديدة لـ المتغيرات ولإبراز المعلومات المميزة. تم تصميم تقنيات التعزيز الجديدة هذه لأداء تجزئة دقيقة وتقليل أخطاء التجزئة بين القناع المتوقع والحقيقة الأساسية مع تقليل وقت التدريب وتجنب الإفراط في التعلم. أظهرت النتائج التجريبية أن طرق التعزيز المقترحة قد حسنت أداء النماذج وتفوقت على النماذج الشائعة الأخرى باستخدام نفس المبدأ عند استخدام بيانات تدريب محدودة، من حيث الدقة وتعميم النموذج وسرعة التعلم. بالإضافة إلى ذلك، يتم تقديم توصيات وشروحات مفصلة للتقنيات في مجال التعلم العميق في هذه الأطروحة. نعتقد أن استراتيجيتنا المطورة ستكون ذات أهمية كبيرة في مختلف التطبيقات مثل المجالات الطبية والبيولوجية وخاصة في المراحل الأولى من التجارب حيث يوجد عدد قليل من العينات.

كلمات مفتاحية: تعلم الآلة ، التعلم العميق ، تجزئة الصور / الفيديو ، التشفير التلقائي ، زيادة البيانات ، البيانات المحدودة ، شبكة الطرق ، الكائنات المتحركة، الإفراط في التعلم.

« Segmentation Fond/Forme par Apprentissage machine pour la Détection d'Objets en Vidéosurveillance et l'Extraction de Routes par vues Aériennes. »

Résumé :

Dans cette thèse, nous étudions le problème de la segmentation des régions de premier plan qui représentent des objets en mouvement dans la vidéosurveillance et la segmentation des régions de premier plan dans le contexte de l'image aérienne pour l'extraction des réseaux routier. Nous proposons un modèle d'apprentissage profond développé pour la segmentation automatique basée sur des architectures d'auto-encoder a multi-profondeur (compressé). Ces systèmes sont entrainées à zéro et par transfert en mode supervisé lorsqu'en utilisant des données d'apprentissage limitées. L'objectif principal des systèmes proposés sont basées sur des nouvelles méthodes d'augmentation de données. Nous utilisons la transformation géométrique par l'analyse en composantes principales ACP et des transformations basées sur des filtres (détection des contours, amélioration des détails et amélioration EDGE) pour créer des nouvelles représentations des variables et d'accentuer les informations caractéristiques. Ces nouvelles techniques d'augmentation sont conçues pour effectuer une segmentation précise et de minimiser les erreurs de segmentation entre le masque prédit et la vérité terrain, tout en réduisant le temps de calcul en évitant le sur-apprentissage. Les résultats expérimentaux ont montré que les méthodes d'augmentation proposées ont amélioré les performances des modèles et surpasse d'autres modèles populaires qui utilisent le même principe lorsqu'en utilisant des données d'entraînement limitées, en terme de précision, généralisation du modèle et rapidité d'apprentissage. De plus, des recommandations et des explications détaillées des techniques dans le domaine de l'apprentissage profond sont fournies dans cette thèse. Nous croyons que notre stratégie développée sera d'un intérêt considérable dans diverses applications telles que les domaines médicaux et biologiques et surtout dans les premières étapes d'expérimentation où il y a peu d'échantillons.

Mots clés : Apprentissage machine, Apprentissage profond, Segmentation d'Images/Vidéo, Auto-encoder, ACP, Augmentation des données, Données limitées, Réseau routier, Objets en mouvement, Sur-ajustement.

« Title Foreground/Background Segmentation using Machine Learning for Object Detection in Video Surveillance and Road Extraction from Aerial Views. »

Abstract :

In this thesis, we study the problem of foreground region segmentation which represents moving objects in video surveillance and foreground region segmentation in the context of aerial image for road network extraction by a deep learning model developed for automatic segmentation based on multi-depth (compressed) autoencoder architectures, these systems are trained from scratch and by transfer in supervised mode when using limited training data. The main objective of the proposed systems are based on new data augmentation methods, with geometric transformation through Principal Component Analysis (PCA) and filter-based transformations (edge detection, detail enhancement, and edge enhancement), to create new representations of variables and to accentuate the characteristic information. These new augmentation techniques are designed to perform precise segmentation and to minimize segmentation errors between the predicted mask and the ground truth with reduced computation time while avoiding overfitting. Experimental results have shown that the proposed augmentation methods have improved the performance of the models and outperformed other popular models using the same principle when using limited training data in terms of accuracy, model generalization, and learning speed. In addition, detailed recommendations and explanations of techniques in the deep learning field are provided in this thesis. We believe our developed strategy will be of significant interest in various applications such as medical and biological domains and especially in the early stages of experiments where there are few samples.

Key words : Machine Learning, Deep Learning, Image/Video Segmentation, Autoencoder, PCA, Data Augmentation, Limited Data, Road Network, Moving Objects, Overfitting.

Remerciement

Je tiens tout d'abord à remercier Allah pour Son soutien constant et sa grâce tout au long de ce voyage d'écriture de thèse. Sa guidance divine a été la source de ma force et de mon inspiration, et je suis profondément reconnaissant pour les bénédictions qu'Il a accordées sur moi et ma famille.

Je tiens à exprimer ma plus profonde gratitude à tous ceux qui ont contribué à la réalisation de cette thèse. Tout d'abord, je tiens à remercier mon directeur de thèse **Pr Taibi Mahmoud** pour son orientation, son soutien et sa confiance en moi tout au long de ce projet. Ses conseils judicieux, sa disponibilité et son encouragement ont été précieux pour moi.

Je tiens à remercier mon Co-encadrant **Taleb-Ahmed Abdelmalik** pour les précieuses corrections et les conseils judicieux qu'il m'a prodigués tout au long de la rédaction de ma thèse.

Je tiens également à exprimer ma reconnaissance envers **Pr Francissco Searadilla**, qui m'a accueilli durant ma période de stage. Sa sagesse, son expertise et sa générosité ont été précieuses pour moi et ont grandement contribué à la réalisation de ce projet. Je suis reconnaissant pour ses conseils avisés, et je suis fier d'avoir eu la chance de travailler sous sa direction.

Je tiens à exprimer ma profonde gratitude envers **les membres du jury**, le Professeur **BOUGHAZI Mohamed** de l'Université Badji Mokhtar - Annaba, le Professeur **ZITOUNI Athmane** de l'Université Mohamed Khider - Biskra, la Docteure **BOUTERAA Nadia** de l'Université Badji Mokhtar - Annaba, et le Professeur **OUAFI Abdelkrim** de l'Université Mohamed Khider - Biskra. Je suis reconnaissant pour leur temps, leurs commentaires constructifs et leur soutien inestimable. Leur expertise a grandement enrichi ma recherche et a contribué à l'amélioration de ma thèse. Je les remercie sincèrement pour leur participation en tant que membres du jury et pour leur engagement envers l'excellence académique.

Je suis reconnaissant envers **mes collègues et amis du Labo LASA** pour leur soutien moral, leur encouragement et leur camaraderie tout au long de ce parcours.

Je tiens à remercier chaleureusement mon épouse **KHALFA Sara** pour son amour et son soutien constant durant tout le processus de recherche pour ma thèse. Son encouragement et sa compréhension ont été essentiels à ma réussite. Merci d'être toujours là pour moi.

Je tiens à remercier chaleureusement **ma famille** pour leur soutien constant, leur amour inconditionnel et leur encouragement. Leur présence dans ma vie est source de force et d'inspiration.

Je tiens à exprimer ma reconnaissance envers **mes collègues de travail au centre de recherche CRTI** pour leur soutien, leur collaboration et leur aide pour la réalisation de ce projet.

Enfin, je tiens à remercier toutes les personnes qui ont directement ou indirectement contribué à la réalisation de cette thèse. Je suis particulièrement reconnaissant envers les membres de la communauté scientifique pour leur aide, leurs conseils et leur soutien.

Dédicaces

A mes chers parents,

Je dédie ce travail à vous en reconnaissance de votre amour inconditionnel, de votre soutien constant et de votre dévouement envers moi. Vous m'avez donné la force de persévérer dans mes rêves et de poursuivre mes objectifs avec détermination. Je suis reconnaissant pour toutes les sacrifices que vous avez faits pour moi, pour les encouragements que vous m'avez prodigués, et pour la confiance que vous avez placée en moi. Vous êtes les personnes les plus importantes de ma vie et je vous aime plus que tout. Ce travail est dédié à vous avec tout mon amour et ma gratitude éternelle.

Table de matières

Remerciement …… II

Dédicace …… III

Table de matières …….. IV

Liste des figures …… XII

Liste des tableaux …… IX

Liste des abréviations …… X

Résume …… XI

Abstract …… XII

ملخص …… XIII

Introduction générale …… 1

Introduction générale …… 2

Chapitre 1 État de l'art …… 10

1. Segmentation Fond/Forme par apprentissage machine …… 11

1.1. Introduction …… 11

1.2. Segmentation …… 11

1.2.1. Segmentation des objets en mouvement (vidéosurveillance) …… 12

1.2.2. Segmentation des routes à partir d'images aériennes (cartographie routière) …… 16

1.3. Conclusion …… 19

Chapitre 2 : Théorie …….. 20

2. Contexte …… 21

2.1. Introduction …… 21

2.2. Intelligence artificielle …… 21

2.3. Apprentissage machine (automatique) …… 22

2.3.1. Types d'apprentissage machine …… 22

2.4. Apprentissage profond …… 27

2.4.1. Réseau de neurone …… 28

2.5. Réseaux de neurons convolutionnels CNNs …….. 31

2.5.1. Convolution …… 31

2.5.2. Le pas (Stride) …… 33

2.5.3. Fonction d'activation …… 33

2.5.4. Pooling …… 33

2.5.5. Couche entièrement connecté …… 34

2.6. La propagation vers l'avant (Forword propagation) ……..…… 35
2.7. Réseaux entièrement convolutifs ……...…… 35
2.8. SegNet ……...…… 36
2.9. U-Nets ……..…… 36
2.10. Sur-ajustement (overfitting) ……...…… 37
2.10.1. L'arrêt précoce (early stopping) ……..…… 38
2.10.2. La régularisation ……..…… 38
2.10.3. L'augmentation des données ……...…… 38
2.10.4. L'apprentissage par transfert (transfer learning) ……... ……. 38
2.10.5. Dropout ……..…… 39
2.11. Généralité de l'apprentissage profond pour la segmentation ……..…… 39
2.11.1. Optimiseurs ……..…… 39
2.11.2. Fonction de perte ……...…… 40
2.11.3. Couche de sortie (Sigmoid/Softmax) ……..…… 40
2.12. Défis de l'apprentissage profond pour la segmentation ……..…… 41
2.13. Conclusion ……...…… 41
Chapitre 3 : Méthodologie ……..…… 42
3. Méthodologie ……...…… 43
3.1. Introduction ……...…… 43
3.2. Auto-Encoder ……..…… 43
3.3. Formation des données ……...…… 45
3.3.1. **Partie 1** : Approche de la segmentation des objets en mouvement (vidéosurveillance) ……..…… 45
3.3.1.1. Dataset Cdnet 2014 ……..…… 45
3.3.1.2. Augmentation des données avec l'analyse en composantes principales (ACP) ……...……..…… 47
3.3.1.3. Protocole d'expérience ……...…… 50
3.3.1.4. Processus des données d'apprentissage ……...…… 51
3.3.2. **Partie 2** : Approche de la segmentation des routes à partir d'images aériennes (cartographie routière) ……...…… 53
3.3.2.1. Dataset Teselas ……...…… 53
3.3.2.2. Augmentation des données à base des transformations par filtres ……..…… 53
3.3.2.3. Protocole d'expérience ……...…… 54
3.3.2.4. Processus des données d'apprentissage ……...…… 56
3.4. Test ……..…… 57
3.4.1. Métriques des testes ……...…… 57
3.5. Implémentation (logiciels et librairies) ……...…… 58

3.5.1 Python ……..…… 58
3.5.2. Librairie ……..…… 59
3.6. Conclusion ……..…… 60
Chapitre 4 : Analyse et discussion des résultats ……..…… 61
4. Résultats et analyse ……..…… 62
4.1. Introduction ……...…… 62
4.2. Apprentissage et plateforme de traitement ……..…… 62
4.3. Analyse et discussion des résultats ……..…… 63
4.3.1. **Partie 1** : Approche de la segmentation des objets en mouvement (Vidéosurveillance) ……...…… 63
4.3.1.1. Test et évaluation des approches proposées ……...…… 63
4.3.1.2. Test de l'approche basé sur l'apprentissage par transfert ……...…… 63
4.3.1.3. Test de l'approche basé sur l'apprentissage multi-profondeur ……...…… 65
4.3.1.4. Comparaison avec les algorithmes de référence ……...…… 67
4.3.1.5. Discussion ……..…… 69
4.3.2. **Partie 2** : Approche de la segmentation des routes à partir d'images aériennes (cartographie routière) ……...…… 70
4.3.2.1. Test et évaluation des approches proposées ……...…… 70
4.3.2.2. Temps d'apprentissage ……...…… 70
4.3.2.3. Comparaison des trois modèles compressées ……..…… 73
4.3.2.4. Discussion ……...…… 74
4.4. Conclusion ……..…… 74
Conclusion et perspectives ……..…… 76
Conclusion et perspectives ……..…… 77
Conclusion ……..…… 77
Perspectives ……...…… 78
Références ……...…… 81
Liste des publications ……..…… 89

Liste des figures

Figure 1 :	structure du chapitre 2. …….....……	7
Figure 2 :	structure du chapitre 3. ……...……	7
Figure 3 :	structure du chapitre 4……..……	8
Figure 4 :	structure du chapitre 5……..……	9
Figure 5 :	Segmentation des objets en mouvement (voitures) dans la vidéo et la segmentation des routes à partir d'images aériennes….....…….……	12
Figure 6 :	Segmentation au niveau des objets (a), et au niveau des instances (b)…… ……	13
Figure 7 :	La chronologie des approches traditionnelles et les approches d'apprentissage profond utilisées pour segmenter les objets en mouvement……...……	14
Figure 8 :	Diagramme des approches traditionnelles et les approches profond utilisées pour la segmentation des routes……...…….…….…	17
Figure 9 :	La relation entre l'intelligence artificielle, l'apprentissage automatique et l'apprentissage profond …….....…….…….……	21
Figure 10 :	Schéma explicatif du type d'apprentissage supervisé……...……	23
Figure 11 :	Schéma explicatif du type d'apprentissage non supervisé……...……	24
Figure 12 :	Schéma explicatif du type d'apprentissage semi-supervisé…….……	26
Figure 13 :	Schéma explicatif du type d'apprentissage par renforcement……..……	26
Figure 14 :	La différence entre l'apprentissage machine et l'apprentissage profond…….……	27
Figure 15 :	La relation entre le réseau neurone artificiel et le cerveau humain. ….	28
Figure 16 :	La structure d'un réseau neuronal de base……...……	29
Figure 17 :	Boucle d'apprentissage d'un réseau neurone……...……	30
Figure 18 :	L'architecture du réseau neurone convolutionnel classique……...……	31
Figure 19 :	Opération de convolution sur une image d'entrée avec un seul noyau, en utilisant le pas (1,1)……. …… ……	32
Figure 20 :	Fonction d'activation……...……	33
Figure 21 :	Couche de Pooling……...……	34
Figure 22 :	Réseaux entièrement convolutif……...……	35
Figure 23 :	Architectures du SegNet……..……	36
Figure 24 :	Architecture du U-Net…….……	37
Figure 25 :	L'architecture de l'approche générale de l'Auto-Encoder…….……	44
Figure 26 :	L'expression de base d'un Auto-Encoder général…….……	45
Figure 27 :	Les images des catégories et des noms de vidéos dans le jeu de données CDnet2014…….…….…….…	46

Figure 28 : La structure des couches de l'approche d'auto-encodeur basé sur l'apprentissage par transfert…… 51
Figure 29 : La structure des couches de l'approche d'auto-encodeur basé sur l'apprentissage a multi-profondeur…… 51
Figure 30 : Exemples des différentes transformations de chaque stratégie…..…… 52
Figure 31 : Exemples d'images qui ne contiennent pas des routes (A) et d'images contenant des routes avec la cartographie associée (B)… ……. ……. … 53
Figure 32 : Les approches générales des trois architectures Auto-Encoder compressées…….…… 55
Figure 33 : Transformation des différentes stratégies d'augmentation des données utilisées……..……..……..… 57
Figure 34 : Graphiques des pertes et des précisions d'apprentissage et de validation des stratégies utilisées……..……..……..…… 64
Figure 35 : Graphiques de la fonction de précision entre les données d'apprentissage et de validation des stratégies utilisées……...…… 65
Figure 36 : Graphiques de la fonction de perte entre les données d'apprentissage et de validation des stratégies utilisées……..……..……..……. 66
Figure 37 : Graphiques des pertes et des précisions d'apprentissage et de validation du modèle F……..……..……..…. 67
Figure 38 : Comparaison visuelle entre les masques de segmentation générés par le modèle proposé et les modèles de référence……..……..…… 68
Figure 39 : Les courbes de fonction de précision entre les données d'apprentissage et de validation des trois architectures……...……..……..…… 71
Figure 40 : Les courbes de fonction de perte entre les données d'apprentissage et de validation des trois architectures……...……..……..……. 72
Figure 41 : une comparaison visuelle entre les trois modèles compressées…..…… 74

Liste des tableaux

Tableau 1 : Liste des catégories et description de chaque vidéos dans le jeu de données CDnet2014**……...……** 46

Tableau 2 : Description et nombre d'échantillons d'entraînement utilisés pour chaque stratégie**……..……** 52

Tableau 3 : : Description et nombre d'échantillons d'entraînement utilisés pour chaque stratégie**……..……** 56

Tableau 4 : Les résultats des tests obtenus par les stratégies PCA, DA et PCA+DA. 64

Tableau 5 : Une comparaison des résultats de nos approches avec des algorithmes de référence**……..……..……..** 68

Tableau 6 : Temps de d'apprentissage des stratégies pour chaque approche**…. ….** 70

Tableau 7 : Une comparaison entre les trois modèles compressées **……..……** 73

Liste des abréviations

ACP	Analyse en composantes principales
AE	Auto-encoder
BCE	Binary Cross-Entropy
CNNs	Réseaux de neurones convolutionnels
DA	Data augmentation
FCN	Réseaux entièrement convolutifs
GAN	Réseaux adversariens génératifs
GMM	Modèle Gaussien Mixture
GPS	Global Positioning System
GPU	Graphics Processing Unit
IA	Intelligence artificielle
IoU	Intersection over Union
KDE	Estimateur de Densité de Kernel
LSTM	Long Short-Term Memory
MD AE	Multi depth Auto-encoder
MOS	Segmentation des objets en mouvement
NLP	Traitement du langage naturel
ReLU	Rectified linear unit
RNA	Réseau neurone artificiel
RNN	Réseau de neurones récurrent
SVM	Support Vector Machine
TL	Transfer learning
TL AE	Trasfer learning Auto-encoder
VGG	Visual Geometry Group

Introduction générale

1. Introduction générale

La segmentation des objets dans le monde réel est plus facile pour le cerveau humain, nous pouvons facilement distinguer un objet et le classer dans des catégories générales telles qu'une voiture en mouvement ou un arbre statique, même lorsque l'environnement de vision est changeant. Par contre lorsque la même segmentation est réalisée à l'aide d'un traitement informatique ou bien de l'intelligence artificielle. Il est difficile pour ces méthodes d'extraire avec précision les objets de premier plan.

L'objectif de la segmentation est de séparer les objets de premier plan des objets du fond dans les images et les vidéos[1, 2]. Donc la segmentation d'objet divise les pixels de la séquence en deux sous-ensembles : la cible de premier plan et la région d'arrière-plan, et génère le masque de segmentation d'objet d'intérêt, qui est le cœur du problème. La segmentation est également considérée comme un problème de classification, où l'objectif est de classer chaque pixel de l'image entre l'arrière-plan et le premier plan.

La segmentation des régions de premier plan est une tâche classique de traitement d'image et de vidéo, largement utilisée dans la vidéosurveillance et d'autres domaines, elle est l'étape de base de nombreuses tâches de vision par ordinateur. Le schéma de segmentation d'objet retire un ensemble de caractéristiques dans une séquence d'images et les utilise pour un apprentissage ultérieur dans la phase d'apprentissage.

Ces dernières années, les architectures d'apprentissage profond ont fourni des résultats de qualités dans diverses applications liées à la vision par ordinateur, notamment la classification d'images, la détection d'objets et le traitement du langage naturel (NLP) [1, 3, 4]. Le concept d'apprentissage profond est un sous-domaine de l'intelligence artificielle qui se distingue des techniques d'apprentissage automatique par la manière dont il apprend les représentations à partir des données. Contrairement aux techniques traditionnelles d'apprentissage automatique, les modèles d'apprentissage profond extraient de manière autonome les caractéristiques cachées des données en utilisant un réseau hiérarchique à travers de nombreuses couches. Au cours des dernières années, un large choix d'architectures d'apprentissage profond a été développé, examiné et discuté [1, 5].

Dans cette thèse, on a analysé les techniques de segmentation d'objets dans le contexte des images et des vidéos par les méthodes d'apprentissage machine profond.

- **La segmentation des régions de premier plan qui représentent des objets en mouvement dans les vidéos** est le concept central de la plupart des systèmes de vision par ordinateur comme (l'analyse de mouvement, le suivi d'objets, l'amélioration de la qualité d'image et la compression de données). La segmentation des objets en mouvement est considérée comme une étape cruciale, alors qu'elle représente une opération complexe pour de nombreuses applications de vidéosurveillance comme le comptage de personnes, la reconnaissance d'actions et la surveillance du trafic [6-8]. En outre, la construction d'un modèle précis capable de segmenter des objets en mouvement dans des vidéos de faible qualité est encore plus délicate. En outre, d'autres problèmes tels que la présence d'ombre, les changements d'éclairage, l'arrière-plan dynamique et les mauvaises conditions météorologiques peuvent rendre la tâche de modélisation plus complexe.
- **La segmentation des régions de premier plan dans le contexte de l'image aérienne**, est une source d'information utile pour de nombreux domaines comme les systèmes développés pour l'extraction de routes à partir d'images de télédétection, ce sujet est important dans un certain nombre d'applications, par exemple la gestion du trafic, la planification urbaine, la surveillance des routes, la navigation GPS et la mise à jour de la cartographie [9, 10]. Cependant, nous nous concentrons sur la segmentation de routes dans les images aériennes, notamment, la nécessité de mettre régulièrement à jour les cartes routières. Actuellement, au niveau national, l'opération de cartographie des routes est un processus manuel dur et long, Il existe aussi de nombreux facteurs qui rendent l'extraction de réseaux routiers à partir d'images de télédétection difficile. Cependant, en raison de l'absence de bordures bien marquées, la présence d'occlusions (par exemple, une végétation dense), la différence de largeur et des grands changements de courbure, leur surveillance et leur fournir une cartographie à jour représentent un travail considérable pour l'administration publique. Étant donné que les routes sont fréquemment réparées et construites, la mise à jour de la cartographie routière est une mission importante pour les organismes publics et le processus de mise à jour est souvent coûteux et long. Jusqu'à présent, aucune méthode d'extraction de réseau routier entièrement automatisée n'est appliquée dans la pratique. La raison réside dans la difficulté du problème de classification.

Récemment, les modèles d'apprentissage profond sont devenus l'un des plus importants concepts pour résoudre plusieurs activités liées à la vision par ordinateur, notamment pour les applications basées sur la segmentation d'images et la modélisation dynamique de l'arrière-plan [11-13], car ils offrent de meilleures performances que les méthodes traditionnelles d'apprentissage automatique. Ces modèles d'apprentissage profond présentent de nombreux problèmes qui doivent être résolus. Le problème le plus courant est le manque d'échantillons d'entraînement dans les ensembles de données. Cependant, la collecte d'une grande quantité d'images pour l'entraînement est très contrariante et coûteuse. C'est pourquoi de diverses méthodes d'augmentation des données ont été proposées. Plusieurs études dans la littérature ont ciblé le développement de modèles de segmentation d'objets basés sur l'apprentissage profond, comme les travaux mentionnés dans les références [11, 14]. La plupart de ces travaux ont étudié les performances des méthodes de segmentation par apprentissage profond dans le cas de l'utilisation de grands ensembles de données pour entraîner les modèles basés sur l'apprentissage profond. Et seulement quelques études ont exploré comment former et améliorer les performances d'un modèle d'apprentissage profond dans le cas de petits ensembles de données.

L'entraînement de modèles d'apprentissage profond sur un petit ensemble de données a attiré une attention particulière dans les recherches récentes dans plusieurs domaines. Cependant, seuls quelques travaux ont abordé un tel problème. Par exemple, pour surmonter le problème de la petite taille des ensembles de données, Salehinejad et al. [15] Ont utilisé une technique de transformation cylindrique dans un système de coordonnées cylindriques. En appliquant de telles transformations, ils ont pu réaliser une segmentation d'objets à partir d'une tomographie abdominale 3D en obtenant des performances supérieures à celles des FCN [16]. Afin de pallier le manque de données d'apprentissage, les auteurs de [17] ont proposé une architecture SSF-CNNN qui modifie la structure et le poids des filtres obtenus par CNN afin de réduire le nombre de paramètres d’apprentissage. La technique proposée a prouvé son efficacité pour les problèmes de reconnaissance de visages des nourrissants dans le monde réel et la classification multi-objets. Salehinejad et al. [18] Ont utilisé une transformation radiale au niveau du pixel dans un système de coordonnées polaires pour chaque image afin d'augmenter le nombre d'échantillons de la base de données. L'approche proposée a augmenté les performances de généralisation des modèles pour divers ensembles de données.

Les algorithmes d'apprentissage profond nécessitent souvent des milliers de données d'entraînement pour bien se généraliser. Cependant, il existe de nombreux domaines et applications où la quantité de données d'apprentissage disponibles est limitée en raison des coûts élevés engendrés par la collecte ou l'annotation de données appropriées. Dans tels scénarios, le pré-entraînement (l'apprentissage par transfert) sur des tâches similaires avec de grandes quantités de données telles que le jeu de données ImageNet [19] est devenu illégal pour les applications commerciales à cause de réglementations sur les droits d'auteur imposées par de nombreux pays. Néanmoins, la majorité des recherches qui appliquent l'apprentissage profond à de petits ensembles de données utilisent les méthodes d'augmentation des données et se concentrent uniquement sur la transformation physique, par exemple la rotation, le décalage et la mise à l'échelle. Ce qui rend le système moins authentique.

Un modèle de segmentation précis et bien généralisé appliquée à de petits ensembles de données reste un défi crucial en vision par ordinateur, ce qui est aussi souvent le cas dans de nombreuses applications du monde réel. Dans le cadre de la pandémie mondiale de COVID-19 [20, 21], l'utilisation de l'intelligence artificielle pour analyser l'image de la radiographie pulmonaire (CXR) pour le diagnostic du COVID-19 et le triage des patients devient important. Malheureusement, en raison de la nature récente de la pandémie de COVID-19, il est difficile de collecter les données CXR pour l'entraînement des réseaux neuronaux profonds.

L'étude actuelle fait partie d'un modèle d'apprentissage profond développé pour la segmentation automatique d'objets en mouvement et l’extraction des réseaux routiers. L'apprentissage profond est appris à partir de données provenant des caractéristiques de haut niveau générées par les différentes couches du réseau en utilisant des méthodes d'apprentissage simples. En outre, la présence de grandes bases de données est nécessaire pour reconstruire efficacement le masque de segmentation résultant et obtenir de meilleurs résultats à partir de ces caractéristiques précises. Cependant, dans le scénario du monde réel, les grandes bases de données ne sont pas toujours disponibles.

Suite à ce constat, dans ce travail nous introduisons un système pour améliorer la précision de la segmentation d'objets en mouvement dans les vidéos surveillance et d’extraction des réseaux routiers dans des images aériennes à haute résolution avec des données d'entraînement limitées en utilisant un auto-encodeur de bout en bout a des architectures a multi-profondeurs (compressé) par formation à partir de zéro et aussi par l’apprentissage par transfert. Afin d’améliorer les performances du modèle la plupart des

recherches tendent à fixer les données d'entraînement et modifie l'architecture du modèle. Tant dis que d'autres chercheurs n'utilisent que des techniques traditionnelles et générales d'augmentation des données pour élargir la base de données dans de différents domaines, telles que (rotation, translation, etc.). Dans le présent travail, certaines modifications des données sont effectuées pour augmenter le nombre d'échantillons. Pour cette raison, on propose et on compare les techniques orientées vers la segmentation du premier plan pour augmenter et améliorer la qualité de l'ensemble de données d'entraînement qui aide le modèle à extraire les caractéristiques pertinentes et à couvrir le manque d'échantillons nécessaires. Ces techniques reposent sur les transformations des filtres de base (détection des contours, amélioration des détails, amélioration EDGE et recherche des bords), et autres basées sur la transformation géométrique avec l'analyse en composantes principales (ACP). Ces nouvelles techniques proposées ont pour but d'éviter l'effet de sur-ajustement et d'améliorer le processus de distinction du premier plan lorsque les données disponibles sont limitées, et d'atteindre un temps d'apprentissage de segmentation rapide.

L'objectif de ce travail est de savoir comment segmenter rapidement et précisément les objets d'intérêt dans l'image et la vidéo. Une extraction pertinente et une réduction des informations caractéristiques de haute dimension avec différentes stratégies de filtrage cités ci-avant afin de créer un ensemble de données d'apprentissage adapté pour la segmentation du premier plan avec un temps de calcul réduit et de créer un modèle compressé pour des données limitées.

Les résultats obtenus à partir des expériences comparatives entre les approches proposées donnent de bons résultats de segmentation du premier plan et surpasse les autres modèles populaires en terme de rapidité et précision, qui utilisent le même principe d'apprentissage avec des données limitées. Notre méthodologie constitue des stratégies utiles pour augmenter les échantillons dans le domaine de l'apprentissage profond et appliquée pour la segmentation et aident à la généralisation du modèle. Nous pensons que la stratégie que nous avons développée sera d'un intérêt considérable dans diverses applications telles que la biologie et le domaine médical, notamment dans les premières étapes des expériences où les échantillons sont limités. Compte tenu du fait que l'apprentissage profond est un sujet hautement recommandé dans plusieurs domaines, nous avons mené notre travail pour être l'une des premières contributions qui traite le problème de la phase d'apprentissage avec des données limitées.

Cette thèse est organisée comme suit :

- **Le chapitre 1** présente un aperçu sur la segmentation Fond/Forme dans le domaine de l'image et la vidéo, ainsi qu'une revue des travaux connexes sur les techniques d'apprentissage machine traditionnelles et profondes utilisée pour la segmentation des régions de premier plan qui représentent des objets en mouvement dans les vidéos et la segmentation des régions de premier plan dans le contexte de l'image aérienne. Un bref aperçu du chapitre est structuré dans la **figure1**.

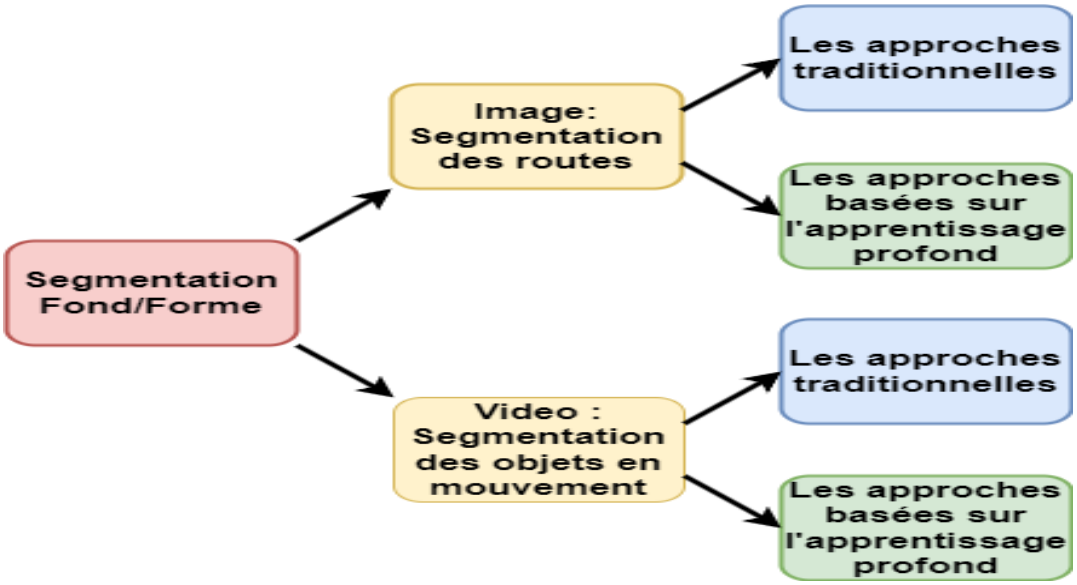

Figure 1 : structure du chapitre 1.

- **Le chapitre 2** présente un aperçu technique et les connaissances de base de l'apprentissage automatique. Les réseaux de neurones convolutionelles (CNNs) et leurs méthodes de fonctionnement dans le défi de la classification d'images seront abordés, ainsi que les architecturent variantes de l'apprentissage profond lié à la tache de segmentation au cours de ce processus. En outre, les types d'apprentissage et les techniques de la régularisation seront étudiés. Un bref aperçu du chapitre est structuré dans la **figure2**.

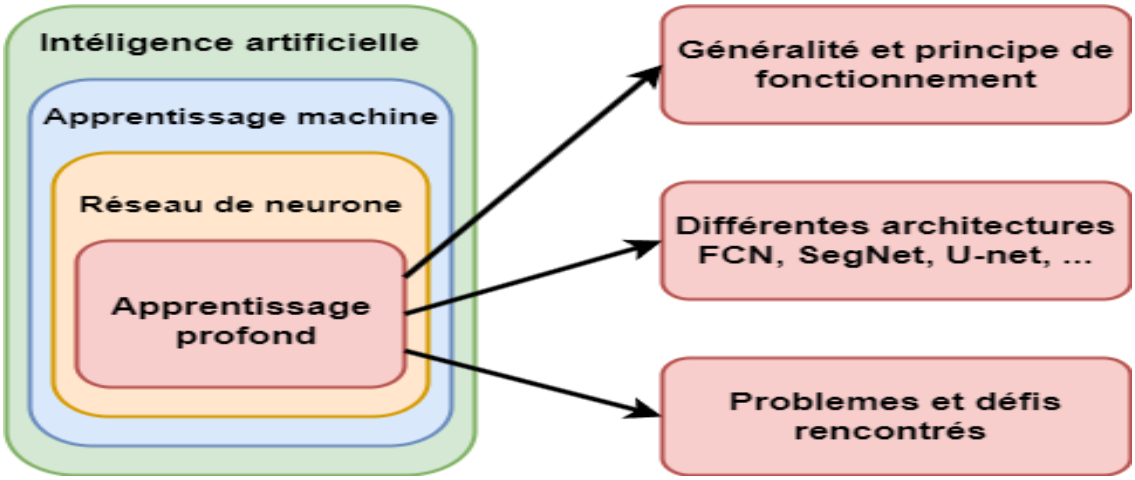

Figure 2 : structure du chapitre 2.

- **Le chapitre 3** présente les méthodes et le matériel utilisés pour construire un système de segmentation d'objets en mouvement robustes à partir de la vidéosurveillance et un système de segmentation des routes rapide et précis à partir des images aériennes. Nous fournissons une description détaillée de chacun des principaux éléments constitutifs introduits dans notre méthodologie. Ainsi, nous présentons l'implémentation des approches proposées. Par contre, les détails des approches proposées et les différentes stratégies d'augmentation des données seront présentés. Un bref aperçu du chapitre est structuré dans la **figure3**.

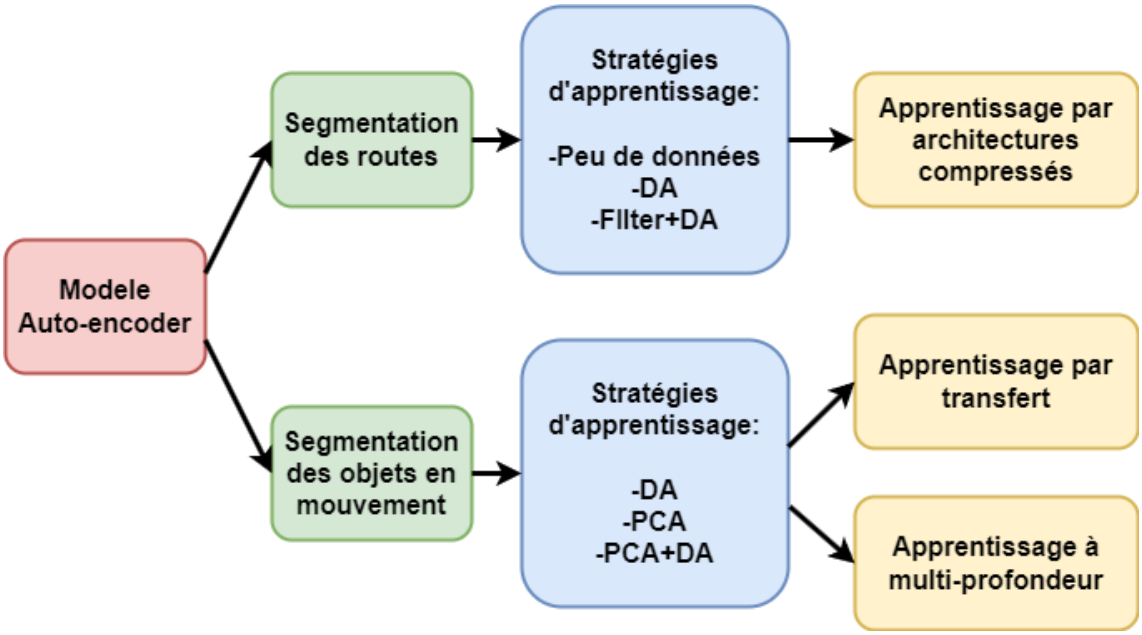

Figure 3 : structure du chapitre 3.

- **Le chapitre 4** présente les résultats expérimentaux de la phase d'apprentissage et du test de tous modèles développés et l'analyse de chaque approche et la manière d'effectuer la segmentation pour les deux domaines d'applications. Ainsi, une discussion des performances des nouvelles stratégies d'augmentation des données sur chaque approche. De même, les résultats qualitatifs et quantitatifs de chaque modèle sont présentés à des fins de comparaison entre approches proposées et avec les méthodes populaires. Un bref aperçu du chapitre est structuré dans la **figure4**.

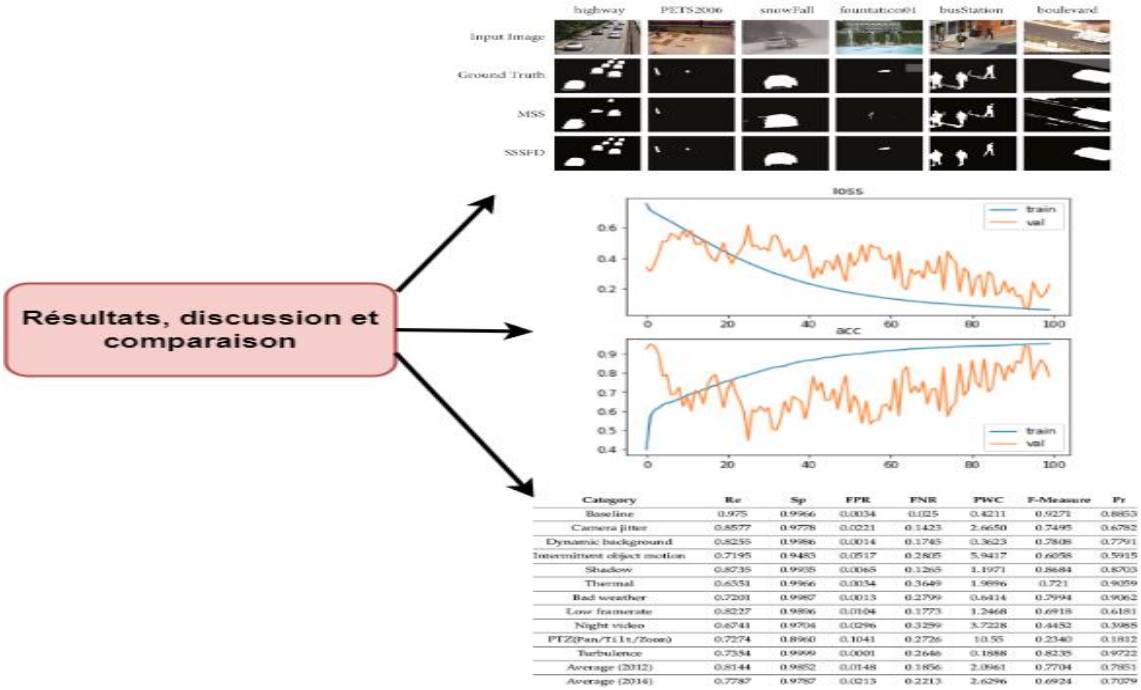

Figure 4 : structure du chapitre 4.

- Nous concluons la thèse et discutons des travaux futurs dans une dernière section.

Chapitre 1
État de l’art

1. Segmentation Fond/Forme par apprentissage machine

1.1. Introduction

Dans le monde réel, la segmentation des objets est une tâche facile pour le cerveau humain, qui peut rapidement distinguer et classifier les objets dans des catégories générales, même dans des environnements changeants. Cependant, lorsque cette segmentation est réalisée par des méthodes informatiques ou d'intelligence artificielle, extraire avec précision les objets de premier plan devient un défi.

Ce chapitre fournit un aperçu de la segmentation Fond/Forme, ainsi que des informations générales sur les techniques d'apprentissage machine utilisées dans le domaine de la segmentation d'images et de vidéos.

1.2. Segmentation

La segmentation d'images est la capacité d'une machine à associer les pixels d'une image à leurs étiquettes de classe d'objet correspondantes ou de la région qui l'entoure. L'objectif de la segmentation est de séparer les objets de premier plan des objets du fond dans les images et les vidéos. Donc, la segmentation divise les pixels de la séquence en deux sous-ensembles : la cible de premier plan et la région d'arrière-plan, et génère le masque de segmentation d'objet d'intérêt, qui est le cœur du problème[2, 22].

La segmentation des régions de premier plan est une tâche classique et une étape de base de nombreuses applications dans la vidéosurveillance et d'autres domaines [1, 2]. De nos jours, la segmentation est appliquée aux images fixes et aux vidéos, elle est l'une des étapes essentielles dans le domaine de la vision par ordinateur. Le processus de segmentation d'objet consiste à extraire un ensemble de caractéristiques d'une séquence d'images, qui seront ensuite utilisées dans la phase d'apprentissage.

Dans cette thèse, on va analyser les techniques de segmentation d'objets dans le contexte des images (Segmentation des routes à partir d'images aériennes) et des vidéos (Segmentation des objets en mouvement) par les méthodes d'apprentissage profond. Comme le montre la **figure5**

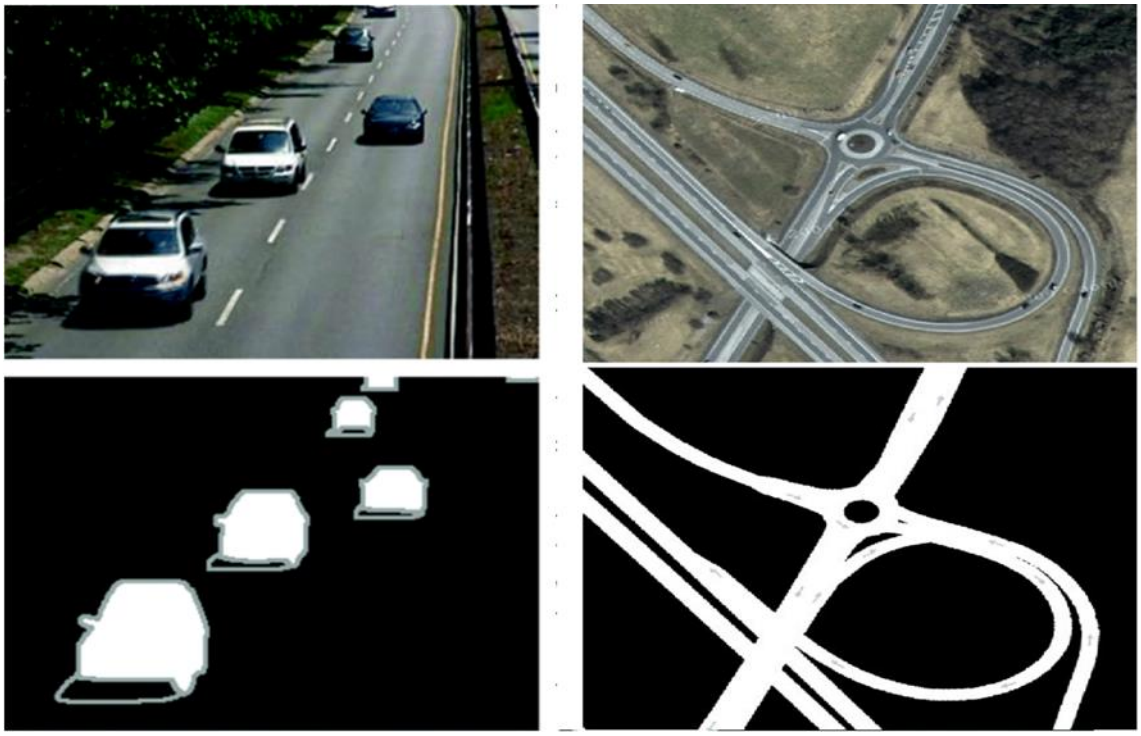

Figure 5 : Segmentation des objets en mouvement (voitures) dans la vidéo et la segmentation des routes à partir d'images aériennes.

1.2.1. Segmentation des objets en mouvement (vidéosurveillance)

La segmentation des régions de premier plan qui représentent des objets en mouvement dans les vidéos (MOS) [23] est une étape fondamentale en vision par ordinateur. Il s'agit du processus d'extraction du contenu dynamique de premier plan dans la vidéo, tels que les véhicules ou les piétons en mouvement, tout en écartant l'arrière-plan. Elle représente un rôle essentiel dans de nombreuses applications du monde réel, comme la conduite autonome [24], le suivi des objets [25] et les systèmes de surveillance [22].

L'objectif de la MOS est de diviser les pixels de la séquence vidéo en deux sous-ensembles : la cible de premier plan (les objets mobiles) et la région d'arrière-plan (les objets statiques), et de générer le masque de segmentation d'objet. La MOS est également considéré comme un problème de classification binaire, où l'objectif est de classer chaque pixel des vidéos entre l'arrière-plan et le premier plan, c'est pourquoi elle est également connue sous le nom de soustraction d'arrière-plan [7].

La segmentation des objets en mouvement est classée en deux catégories : la segmentation au niveau des objets et la segmentation au niveau des instances [1, 26]. Comme l'illustre la **figure6(a)**, l'objectif de **la segmentation au niveau de l'objet** est de détecter tous les objets en mouvement au premier plan et de générer un masque de

segmentation binaire au niveau du pixel sans distinguer les différentes instances. Par contre, comme l'illustre la **figure6(b)**, **la segmentation au niveau de l'instance** permet non seulement de détecter les objets mobiles de premier plan, mais aussi d'attribuer à chaque instance une étiquette différente, qui peut être utilisée pour l'identification et le suivi des objets dans le temps.

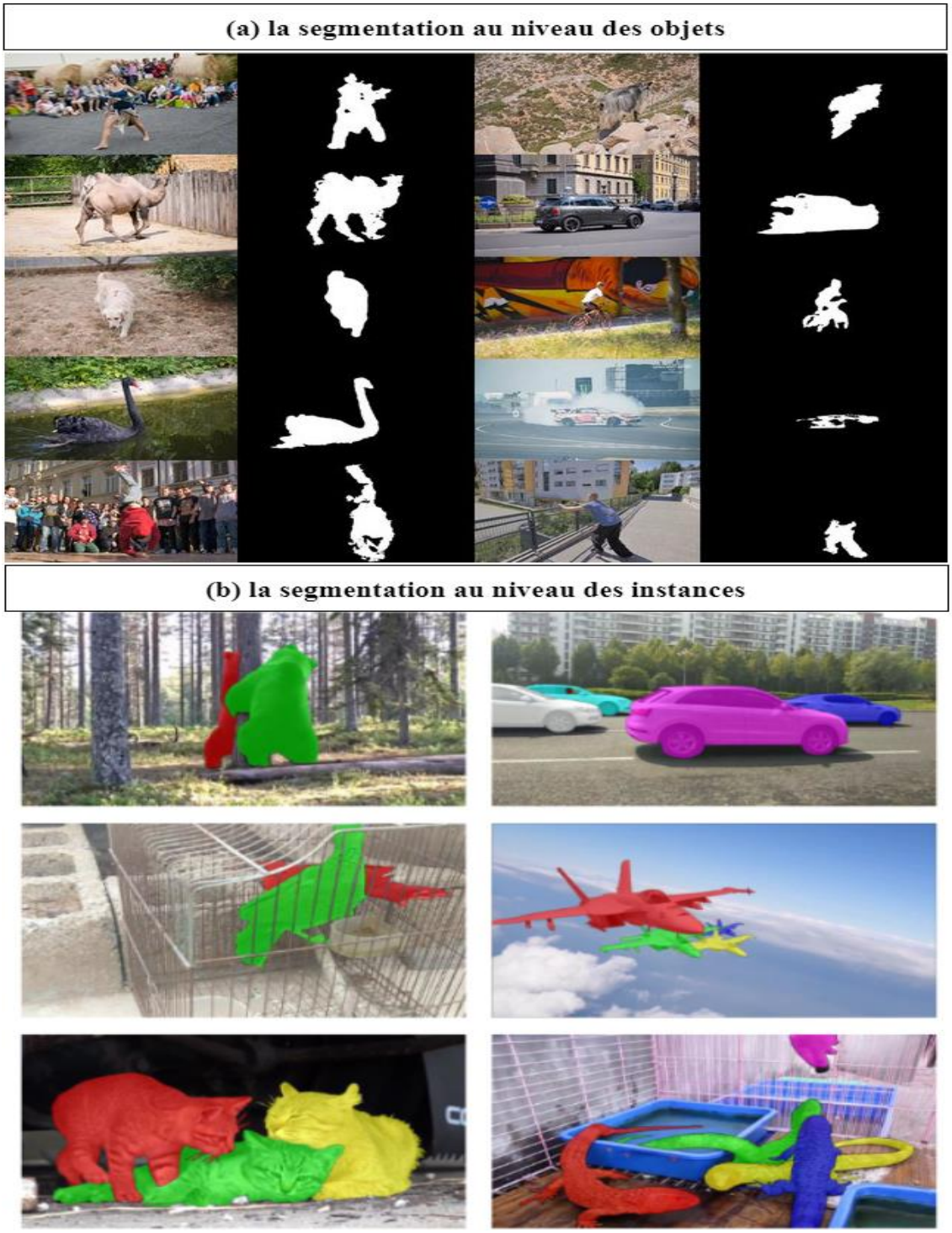

Figure 6 : Segmentation au niveau des objets (a), et au niveau des instances (b).

Les approches utilisées pour segmenter les objets en mouvement peuvent être classées en deux catégories : **les approches traditionnelles** et **les approches basées sur l'apprentissage profond**. L'évolution chronologique des algorithmes est illustrée à la **figure7**. Les approches basées sur l'apprentissage profond sont de plus en plus développées à partir de 2016 et les approches traditionnelles ont commencé au début des années 2000.

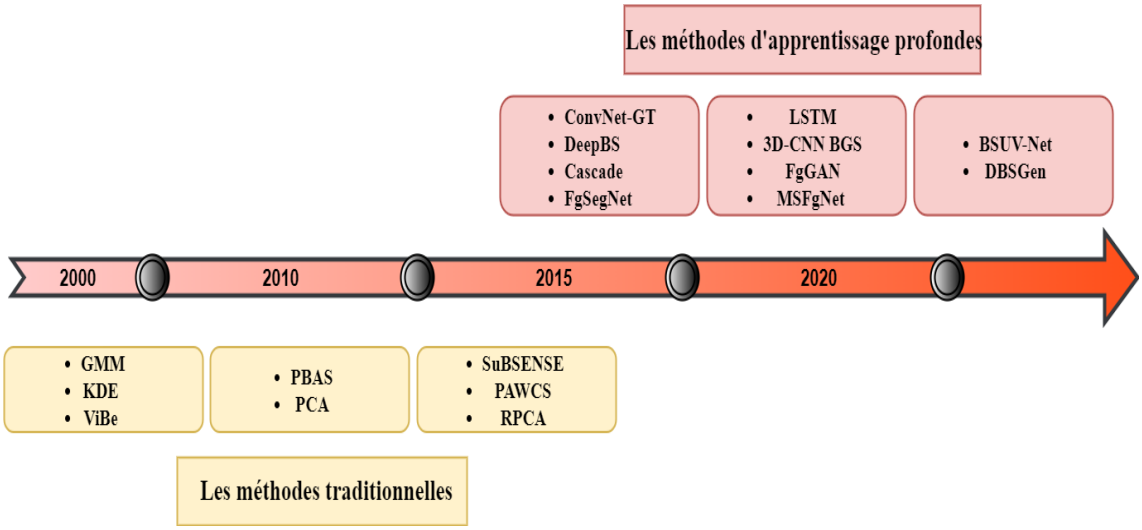

Figure 7 : La chronologie des approches traditionnelles et les approches d'apprentissage profond utilisées pour segmenter les objets en mouvement.

Les approches traditionnelles se basent sur la modélisation du fond de scène d'arrière-plan est initialisée et mise à jour au fil du temps, puis chaque pixel est classé comme avant-plan ou arrière-plan en fonction d'un seuil. Ces méthodes traditionnelles ne nécessitent pas de masques de segmentation étiquetés mais rencontrent des difficultés lorsqu'elles sont appliquées à des scénarios complexes, tels que des vidéos avec des arrière-plans dynamiques, des ombres, des changements d'illumination et des scènes de nuit.

Les mécanismes de modélisation de l'arrière-plan peuvent être paramétriques ou non paramétriques. Les approches paramétriques représentent l'arrière-plan (statistique) utilisent une fonction de densité de probabilité comme une simple gaussienne ou un mélange de gaussiennes (GMM) [27], et nécessitent d'apprendre les paramètres de fonction de densité de probabilité.

Les approches non paramétriques n'ont pas besoin d'apprendre les paramètres des distributions de probabilité, et utilisent la méthode d'estimation de la densité du noyau (KDE) [28]. Il existe également plusieurs sous-catégories d'approches non paramétriques : les méthodes basées sur des filtres, le consensus d'échantillon et l'analyse en

composantes principales (ACP) [29]. Parmi les approches basées sur des filtres, on peut citer le filtrage de Kalman [30] et le filtrage de la médiane temporelle [31]. Les méthodes de consensus d'échantillons à savoir le ViBe [32], SuBSENSE [33] et PAWCS [34]. L'analyse robuste des composantes principales (RPCA) [35] est également développée pour résoudre le problème du mouvement de la caméra dans la phase de la modélisation de l'arrière-plan basée sur l'ACP.

Avec le développement d'une infrastructure informatique puissante, **les approches basées sur l'apprentissage profond** ont montré un succès remarquable dans la segmentation des objets en mouvement qui sont basés sur les réseaux de neurone profond. Ils ont démontré que les réseaux de neurone profond peuvent extraire automatiquement des caractéristiques spatiales de bas, moyen et haut niveau qui s'avèrent très utiles dans les problèmes de segmentation des objets en mouvement.

Le premier travail basé sur l'apprentissage profond est ConvNet-GT [36], qui effectue la détection des changements en remplaçant l'étape de classification des pixels par un réseau neurone. L'arrière-plan est estimé à l'aide d'un filtre médian, puis il est empilé avec les images vidéo originales pour former l'entrée du CNN qui produit les masques binaires. Dans un modèle plus avancé, le MSFgNet [37], un réseau d'atténuation du mouvement (MSNet) est utilisé pour l'étape d'estimation de l'arrière-plan, qui est ensuite soustrait des images originales, suivi d'un réseau d'extraction d'avant-plan (FgNet) qui détecte les objets en mouvement. Un autre type de CNN est le CNN basé sur l'apprentissage de caractéristiques multi-échelles, qui extrait des caractéristiques multi-échelles pour obtenir une meilleure précision, comme le MSCNN+Cascade [11], le Guided Multi-Scale CNN [38], le MCSCNN [39], et les réseaux basés sur le VGG-16 [40] FgSegNet_M [41] et FgSegNet_v2 [42]. Dans MSCNN+Cascade [11], les résultats de la segmentation sont générés pixel par pixel. Bien qu'il atteigne une bonne précision, le traitement pixel par pixel prend beaucoup de temps. Dans FgSegNet_M [41], un CNN 2D prend chaque image vidéo à trois échelles de résolution différentes en parallèle comme entrée du réseau de codage.

L'avantage de la convolution 3D pour les problèmes MOS est que la convolution 3D peut utiliser les informations spatio-temporelles des données visuelles, ce qui permet d'améliorer la précision du modèle. Par exemple, dans le CNN-LSTM 3D [43], les mouvements temporels à court terme d'une séquence vidéo sont capturés par des convolutions 3D, tandis que les mouvements temporels à long et court terme sont capturés par des modules LSTM 2D. Un autre exemple est 3D-CNN-BGS [44] utilise la convolution 3D pour suivre les changements temporels dans les séquences vidéo. Il

effectue une convolution 3D sur 10 frames consécutifs de la vidéo, et sur-échantillonne les couches de caractéristiques de bas, moyen et haut niveau du réseau dans une approche multi-échelle pour améliorer la précision de la segmentation. Dans [45], le réseau CNN 3D combiné avec une couche entièrement connectée est utilisé dans une méthode en patch par lot.

Les réseaux adversariens génératifs (GAN) sont également appliqués aux problèmes de MOS. Le BScGAN [46], est basé sur le GAN conditionnel (cGAN), dans lequel le discriminateur ne prend pas seulement le masque de premier plan comme entrée, mais aussi l'image couleur empilée et son arrière-plan comme entrée conditionnelle pour améliorer la précision de la segmentation. FgGAN [47], a adopté l'apprentissage non couplé basé sur le GAN pour résoudre les problèmes de MOS tels que l'arrière-plan dynamique, le mauvais temps, l'effet d'ombre et le mouvement irrégulier des objets. Comme l'ensemble d'apprentissage comporte des images d'entrée et de sortie non associées, ce réseau est formé en ajoutant une perte compatible avec le cycle à la perte traditionnelle du GAN. Tout d'abord, l'arrière-plan d'une vidéo est estimé à l'aide d'un réseau d'apprentissage non apparié basé sur le GAN (réseau I). Ensuite, pour extraire les informations de mouvement liées au premier plan, la pertinence du mouvement est estimée en utilisant l'arrière-plan estimé et l'image vidéo actuelle. DBSGen [48] a également été conçu pour faire face au problème de la soustraction dynamique de l'arrière-plan. Il repose sur deux réseaux neuronaux génératifs, l'un pour la suppression dynamique du mouvement et l'autre pour la génération de l'arrière-plan. Les réseaux ont été optimisés de bout en bout. Enfin, les objets mobiles d'avant-plan ont été obtenus par un seuil de distance par pixel basé sur une carte d'entropie dynamique.

1.2.2. Segmentation des routes à partir d'images aériennes (cartographie routière)

Les routes sont les informations fondamentales qui sont traitées dans le système d'information géographique. Les informations routières peuvent être utilisées dans de nombreux secteurs de la vie publique, tels que la navigation des véhicules, la gestion du trafic, la mise à jour des cartes et les secours en cas de catastrophe géologique [14].

La segmentation des routes à partir d'images aérienne est un domaine de recherche important, au cours des dernières années, utilisé par exemple dans la gestion du trafic, la planification urbaine, la surveillance des routes, la navigation GPS et la mise à jour de la cartographie [9, 10]. Un grand nombre de méthodes ont été développées pour traiter ce problème. Le processus de base de ces méthodes consiste à obtenir des informations pour

la classification et la reconnaissance des objets en fonction des caractéristiques spectrales, ainsi que des caractéristiques de forme et des liens spatiaux.

Les méthodes utilisées pour la segmentation des routes à partir d'images aérienne sont aussi classées en deux catégories : **les approches traditionnelles** et **les approches basées sur l'apprentissage profond**. Les méthodes traditionnelles sont les plus populaires dans les premières années, et les méthodes basées sur l'apprentissage profond ont été récemment développées rapidement. La **figure8** représente le diagramme des approches traditionnelles et les approches profondes utilisées pour la segmentation des routes.

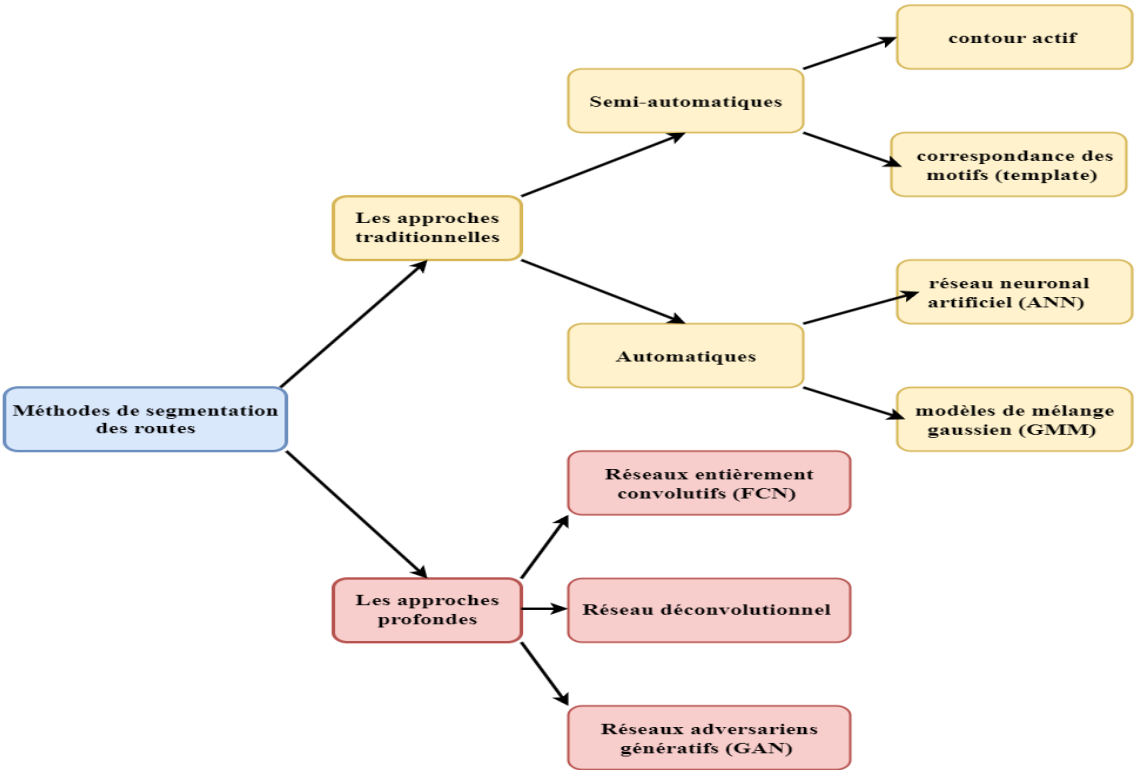

Figure 8 : Diagramme des approches traditionnelles et les approches profond utilisées pour la segmentation des routes.

En ce qui concerne **Les approches traditionnelles,** elles sont divisées en méthodes semi-automatiques et en méthodes automatiques selon le degré d'intervention humaine. Les chercheurs souhaitent vivement que la machine extraie automatiquement le réseau routier et met à jour la cartographie en temps réel. Cependant, les méthodes d'extraction automatique ne peuvent pas atteindre des résultats convaincants compte tenu de la complexité des images aériennes. Par conséquent, les scientifiques ont adopté une solution de compromis, par l'extraction semi-automatique avec l'intervention humaine.

Dans **les algorithmes semi-automatiques** d'extraction de routes, les données (généralement les contours et les directions des tronçons de route) sont fournies manuellement. Ensuite, d'autres tronçons de route sont recherchés en fonction de certaines

caractéristiques de la route et reliés pour former des segments de route. En mode semi-automatique, les algorithmes peuvent demander une intervention aux utilisateurs de juger si les résultats de l'extraction sont précis ou annuler les résultats antérieurs lorsque les algorithmes extraient des routes erronées. Certaines approches traditionnelles des méthodes semi-automatiques comme ; Le modèle de contour actif [49], également appelé Snake, dans ce modèle, une fonction d'énergie est définie sur le contour dessiné par l'utilisateur ; ensuite, la valeur minimale de la fonction d'énergie est calculée de manière itérative en ajustant le contour jusqu'à ce qu'il converge. Les méthodes basées sur l'algorithme de correspondance des motifs (template) [50] sont applicables lorsque les caractéristiques et la surface de la route sont stables et présente une grande similarité dans une certaine zone. Dans ces méthodes de comparaison de modèles, une fenêtre de comparaison spécifique (rectangulaire ou circulaire) est construite au point de départ. Ensuite, la fenêtre est déplacée vers l'avant de manière itérative pour trouver une série de points de correspondance par des critères de similarité par rapport à certaines caractéristiques (couleur, intensité, et texture). Enfin, les points de correspondance sont connectés pour former des segments de route.

Dans **les algorithmes d'extraction automatique**, les paramètres doivent généralement être prédéfinis en fonction de diverses images. Ensuite, l'ensemble du processus peut être exécuté de manière indépendante pour améliorer l'automatisation. Des travaux ont été proposés pour l'extraction automatique des routes ; Méthodes basées sur la segmentation : Il existe de nombreux travaux appliqués à la segmentation d'images, qui peuvent également être utilisés dans l'extraction automatique des routes pour obtenir des régions routières [51]. Les algorithmes suivants sont couramment utilisés : réseau neuronal artificiel (ANN) [52], machine à vecteur de support (SVM) [53], modèles de mélange gaussien (GMM) [54], etc.

Les approches basées sur l'apprentissage profond sont classées en plusieurs types selon l'architecture du réseau neurone utilisé. Parmi les premières tentatives d'extraction de routes à partir des images aérienne à l'aide d'un réseau neurone profond il y a celles qui ont été effectuées par Minh et Hinton [54]. Ils ont construit un réseau de neurone convolutif basé sur des patchs avec des millions de paramètres apprenables et un grand champ réceptif pour prédire la classe des pixels centrés sur le patch et ont amélioré de manière significative de la précision et du recall. Les expériences ont prouvé que le réseau neurone profond offre une capacité remarquable de représentation des caractéristiques. Parmi les approches d'extraction de routes basées sur le réseau neurone profond au cours des dernières années ; on note les méthodes basées sur les réseaux

entièrement convolutifs (FCN) [55], méthodes basées sur le réseau déconvolutionnel [56], méthodes basées sur les réseaux adversariens génératifs (GAN) [57].

1.3. Conclusion

Les approches traditionnelles d'apprentissage automatique pour la segmentation d'objets en mouvement comprennent généralement l'utilisation de méthodes basées sur la modélisation de fond et sur caractéristiques, tels que la colorimétrie et la forme des objets. Ils peuvent être efficaces pour certaines tâches spécifiques, mais peuvent être limités par des défauts dans la compréhension de la scène. De même, pour les approches traditionnelles de segmentation de routes à partir d'images aériennes, on note généralement l'utilisation de techniques basées sur les règles telles que la détection de contour et les algorithmes de filtrage, qui sont souvent limitées par leur manque de souplesse pour traiter les variations de l'environnement et les conditions de prise de vue difficiles.

Les approches récentes d'apprentissage automatique, telles que les réseaux de neurones convolutionnels (CNNs), ont fourni des améliorations considérables en termes de performance pour la segmentation des objets en mouvements et pour l’extraction des routes. Les CNNs peuvent apprendre à détecter les caractéristiques visuelles importantes des objets en mouvement et des routes et les séparent des autres éléments, ce qui les rendent plus robustes aux variations de l'environnement et aux conditions de prise de vue difficiles.

En résumé, les approches traditionnelles sont souvent limitées par leur dépendance aux caractéristiques définies manuellement, alors que les approches récentes peuvent bénéficier de grandes quantités de données pour améliorer les résultats, ce qui les rendent plus flexibles et plus robustes que les approches traditionnelles.

Chapitre 2 : Théorie

2. Contexte

2.1. Introduction

L'aperçu technique et les connaissances de base de l'apprentissage machine sont présentés dans ce chapitre. Les réseaux de neurones convolutionnels (CNNs) et leur méthode de fonctionnement dans le défi de la classification et la segmentation d'images seront abordés, ainsi que les variantes architecturent de l'apprentissage profond lié à la tache de segmentation au cours de ce processus. En outre, les types d'apprentissage et les techniques de la régularisation seront étudiés.

2.2. Intelligence artificielle

Pour accomplir la tâche finale de segmentation dans des conditions environnementales variables, une compréhension de base de l'intelligence artificielle (IA), de l'apprentissage automatique et de l'apprentissage profond est nécessaire.

L'intelligence artificielle [58, 59] est l'idée générale selon laquelle l'effort d'automatisation les machines à des tâches intelligentes normalement effectuées par les humains aussi précisément que possible. C'est-à-dire, Les machines sont développées et codées pour imiter l'intelligence et le comportement humains. Bien que cette définition comprenne à la fois l'apprentissage automatique et l'apprentissage profond, l'IA englobe l'apprentissage automatique qui lui-même englobe l'apprentissage profond.

La relation entre l'IA, l'apprentissage automatique et l'apprentissage profond est illustrée dans la **figure9**.

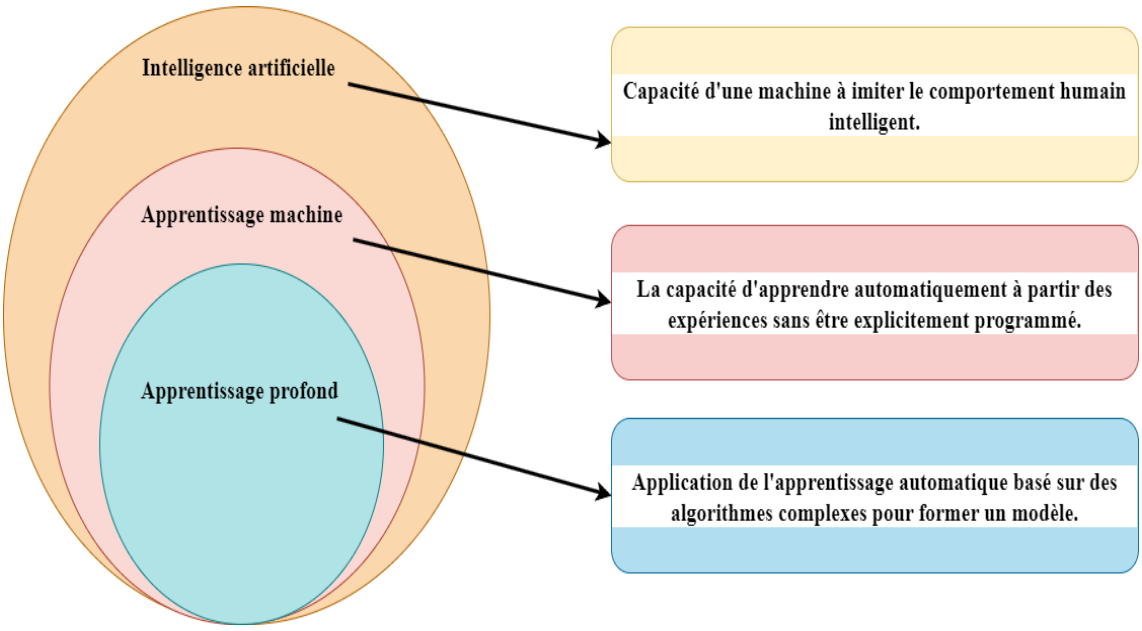

Figure 9 : La relation entre l'intelligence artificielle, l'apprentissage automatique et l'apprentissage profond.

2.3. Apprentissage machine (automatique)

L'apprentissage machine [58-60] (y compris sa branche d'apprentissage profond) est l'étude des "algorithmes d'apprentissage". L'apprentissage automatique exige que le système soit entraîné plutôt que programmé à l'aide des expériences adaptés à une tâche donnée. Le système d'apprentissage automatique détermine alors une structure statistique qui lui permet d'apprendre des représentations significatives des données d'entrée. Ces représentations apprises peuvent ensuite permettre au système de développer ses propres règles pour automatiser la tâche.

Il existe quatre types d'apprentissage automatique : supervisé, non supervisé, auto-supervisé et par renforcement.

Dans cette thèse on utilise l'apprentissage supervisé, qui nécessite trois composantes :

- Des éléments de données d'entrée
- Des exemples de résultats attendus
- Des paramètres de performance

Les paramètres de performance utilisés dans un système d'apprentissage automatique relient l'entrée et la sortie en fournissant un signal de retour pour corriger les sorties de l'algorithme afin de mieux correspondre à la sortie attendue. Cela permet au système d'apprendre à partir de données fournies et de transformer l'entrée en représentations les plus performantes en fonction d'une mesure de performance donnée.

2.3.1 Types d'apprentissage machine

L'apprentissage machine se divise généralement en quatre types : l'apprentissage supervisé, semi supervisé, non supervisé et l'apprentissage par renforcement, en fonction du fait que l'ensemble de données d'apprentissage contient ou non des étiquettes annotées. [59, 60]

- **L'apprentissage supervisé**

Dans l'apprentissage supervisé, la machine est entraînée en lui fournissant des modèles d'entrée et de sortie correspondants. Les principales caractéristiques de l'apprentissage supervisé sont la disponibilité de données d'entraînement "étiquetées" qui peuvent être un ensemble de données à classer. Une variable d'entrée (x) est affectée à une variable de sortie (y) à l'aide d'une fonction d'affectation qui est apprise par un modèle d'apprentissage machine. L'algorithme utilise cet ensemble de variables pour créer une fonction qui associe les entrées aux sorties souhaitées. On peut dire que le type

d'apprentissage supervisé fournit un outil puissant pour classer et traiter les données. Les détails sont illustrés dans la **figure10**.

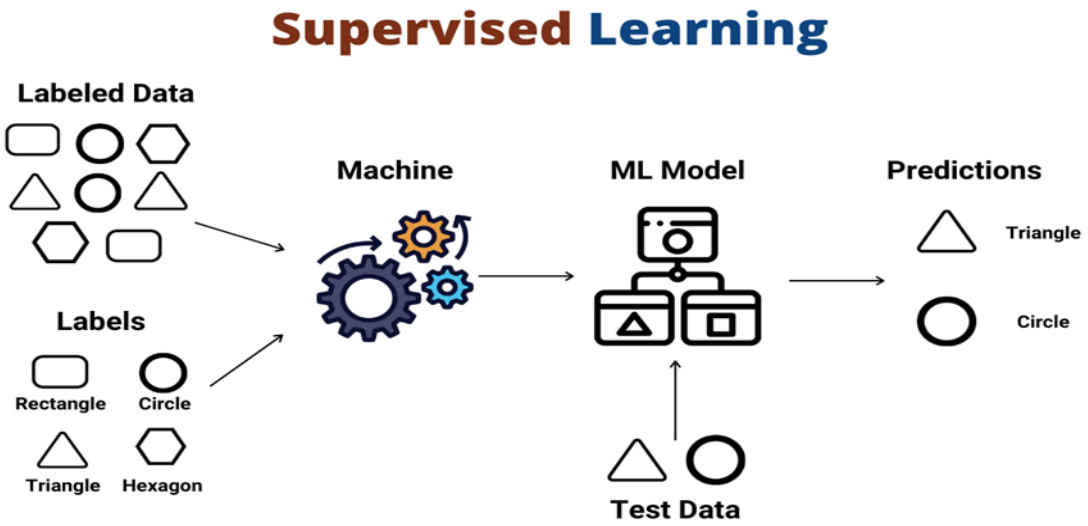

Figure 10 : Schéma explicatif du type d'apprentissage supervisé.

Les techniques d'apprentissage automatique supervisé sont de deux types :

- **Régression**

  La régression est un processus statistique qui permet de trouver une relation significative entre les variables dépendantes et indépendantes. Un algorithme de régression, il prédit un nombre continu. Par exemple, on peut utiliser un algorithme de régression pour déterminer la note de test d'un étudiant en fonction du nombre d'heures qu'il a étudiées dans une semaine. Dans cette situation, les heures d'étude deviennent la variable indépendante et la note finale de l'étudiant est la variable dépendante. La méthode de régression prédit une valeur unique en tant que résultat sur les bases d'apprentissage données. Elle est couramment utilisée pour faire des projections, par exemple ; pour prédire le prix d'une maison à partir des données d'apprentissage en situation réelle ou pour prédire le chiffre d'affaires d'une entreprise.

  Les algorithmes de régression courants comprennent ; La régression linéaire, la régression logistique et la régression polynomiale.

- **Classification**

  La classification peut être définie comme le regroupement d'éléments similaires dans une classe. Si l'algorithme a classé les étiquettes d'entrée dans deux classes distinctes, il s'agit d'une classification binaire. En outre, si les étiquettes sont classées dans plus de deux classes, on parle de classification multi-classe. Avec la

classification, une valeur d'entrée est prise et une classe ou une catégorie lui est attribuée, en fonction des données d'entraînement fournies. Par exemple, juger si un e-mail est un spam ou non est un exemple de classification. Les algorithmes de classification courants sont les suivants : les machines à vecteurs de support, les forêts aléatoires, les arbres de décision et les k-voisins les plus proches.

- **L'apprentissage non supervisé**

L'apprentissage non supervisé est un type d'apprentissage automatique dans lequel les algorithmes reçoivent des données d'apprentissage qui ne contiennent pas d'étiquettes ou d'instructions explicites sur ce qu'il faut faire et sont capable à agir sur les données sans supervision. Les algorithmes d'apprentissage automatique non supervisés sont utilisés pour regrouper les données non structurées en fonction de leurs similarités et des caractéristiques distinctes de l'ensemble de données.

L'objectif est de permettre à l'algorithme d'apprentissage de trouver par lui-même la structure des données d'entrée et particulièrement utile pour trouver des modèles inconnus dans un ensemble de données. Il aide à trouver les caractéristiques nécessaires à la catégorisation et il est spécialement utilisé dans le traitement du contenu multimédia pour le regroupement ou le partitionnement des connaissances.

Pour simplifier, l'apprentissage non supervisé est une sorte d'auto-apprentissage où l'algorithme peut trouver des caractéristiques cachées dans les ensembles de données non étiquetées et fournir les résultats requis sans aucune interférence. L'identification de ces caractéristiques cachées facilite le regroupement, l'association et la détection d'anomalies et d'erreurs dans les données. La **figure11** donne une illustration explicative du type d'apprentissage non supervisé.

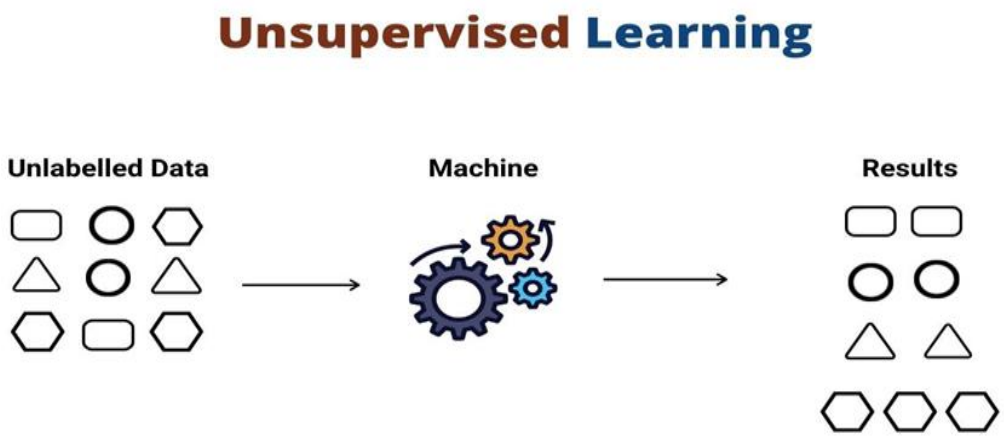

Figure 11 : Schéma explicatif du type d'apprentissage non supervisé.

- **Regroupement (clustering)**

  Le clustering consiste à trouver un modèle dans une collection de données non catégorisées. Les algorithmes de clustering traitent les données et trouvent les clusters naturels existant dans les données. Il s'agit de trouver des caractéristiques cachées dans les données en fonction de leurs similarités ou de leurs différences. Ces modèles peuvent être liés à la forme, à la taille ou à la couleur …etc.

- **Association**

  La technique d'association consiste à trouver les relations qui existent entre les variables dans les grandes bases de données. Nous pouvons ensuite utiliser ces dépendances pour établir des associations entre les objets de données et les mettre en correspondance de manière avantageuse. Par exemple, les personnes qui achètent une nouvelle maison sont plus susceptibles d'acheter de nouveaux meubles. Ces techniques sont souvent utilisées dans l'analyse du comportement des clients sur les sites Web de commerce électronique.

- **Réduction de dimensionnalités**

  L'algorithme vise à réduire les dimensions de caractéristiques d'entrée dans un ensemble de données. Il est utilisé pour l'extraction de caractéristiques.

  L'extraction des caractéristiques importantes d'un ensemble de données est un aspect essentiel des algorithmes d'apprentissage automatique. Cela permet de réduire le nombre de variables aléatoires dans l'ensemble de données en filtrant les caractéristiques non pertinentes.

- **L'apprentissage semi supervisé**

Dans les deux types d'apprentissage automatique précédents, il existe des données étiquetées ou non étiquetées pour assister l'apprentissage. L'apprentissage automatique semi-supervisé tente de résoudre les problèmes qui se situent entre ces deux techniques en découvrant et en apprenant simultanément la structure des variables d'entrée.

L'étiquetage des données est un processus coûteux et long qui nécessite des ressources humaines hautement qualifiées. À cet effet, il existe des cas où les étiquettes ne sont pas disponibles dans la plupart des observations, mais présentes dans une partie seulement, et c'est là qu'intervient l'apprentissage automatique semi-supervisé. Plus nous utilisons de données (pertinentes) pour la formation, plus notre modèle devient robuste. La **figure12** donne une illustration explicative du type d'apprentissage semi supervisé.

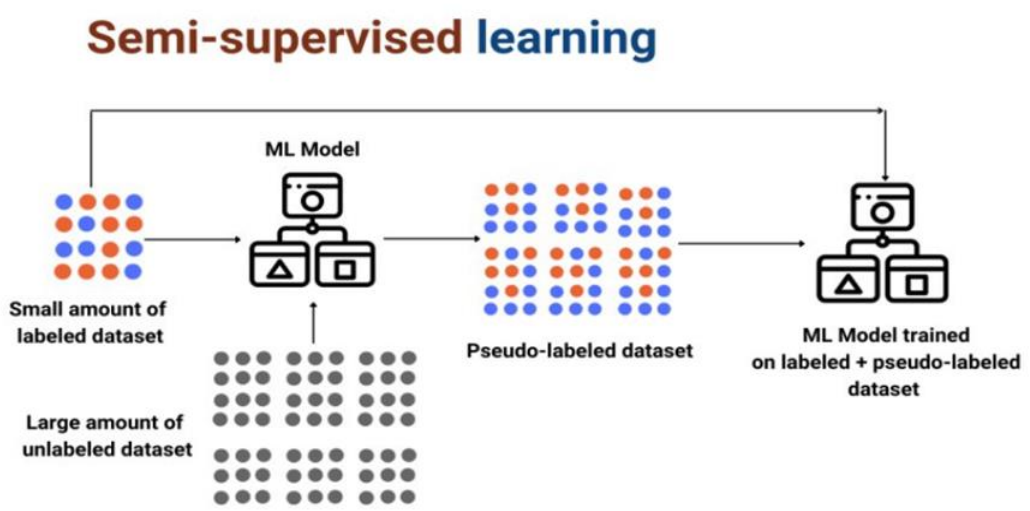

Figure 12 : Schéma explicatif du type d'apprentissage semi-supervisé.

- **L'apprentissage par renforcement**

Contrairement à l'apprentissage supervisé et non supervisé, l'apprentissage par renforcement est un algorithme de type rétroactif. En d'autres termes, pour chaque observation obtenue, l'algorithme donne un feedback au modèle en cours d'apprentissage pour agir de manière à maximiser la récompense ou à minimiser le risque. L'algorithme (également appelé agent) étudie continuellement son environnement jusqu'à ce qu'il explore toutes les possibilités.

L'objectif de l'apprentissage par renforcement est d'entraîner un agent à accomplir une tâche dans un environnement inconnu et dynamique. L'agent reçoit des observations et une récompense de l'environnement et envoie des actions à l'environnement. La récompense est une mesure de la réussite d'une action par rapport à l'objectif sans intervention humaine et sans être explicitement programmé pour réaliser la tâche. La **figure13** présente une illustration explicative du type d'apprentissage par renforcement.



Figure 13 : Schéma explicatif du type d'apprentissage par renforcement.

2.4. Apprentissage profond

L'augmentation de la puissance de calcul grâce au GPU (Graphics Processing Unit) a permis de favoriser l'utilisation d'algorithmes d'apprentissage profond, de sorte que des ensembles de données de plus en plus importants peuvent être utilisés pour la formation. L'apprentissage profond est un sous-domaine de l'apprentissage machine qui approfondit l'approche en apprenant de nombreuses couches successives de représentations des données. Ces couches constituent ce que l'on appelle la profondeur du modèle. L'apprentissage profond doit être exploité à l'aide de réseaux neurones pour apprendre les nombreuses caractéristiques et représentations différentes des données [61, 62].

La catégorie de méthodes la plus couramment appliquée pour analyser l'imagerie visuelle dans le cadre de l'apprentissage profond est le réseau neuronal convolutionels (CNNs) [61]. Contrairement aux méthodes traditionnelles d'apprentissage automatique où les caractéristiques sont choisies manuellement et extraites par des algorithmes dirigés, l'apprentissage profond est composé de méthodes qui alimentent d'abord le système en valeurs de pixels ou en données brutes et lui permettent de découvrir automatiquement des caractéristiques ou des représentations qui peuvent être utilisées pour détecter, distinguer et classer des modèles. L'apprentissage profond est très utile pour apprendre et découvrir des structures fines et complexes dans des données hautement dimensionnelles.

La **figure14** représente la différence entre l'apprentissage machine et l'apprentissage profond.

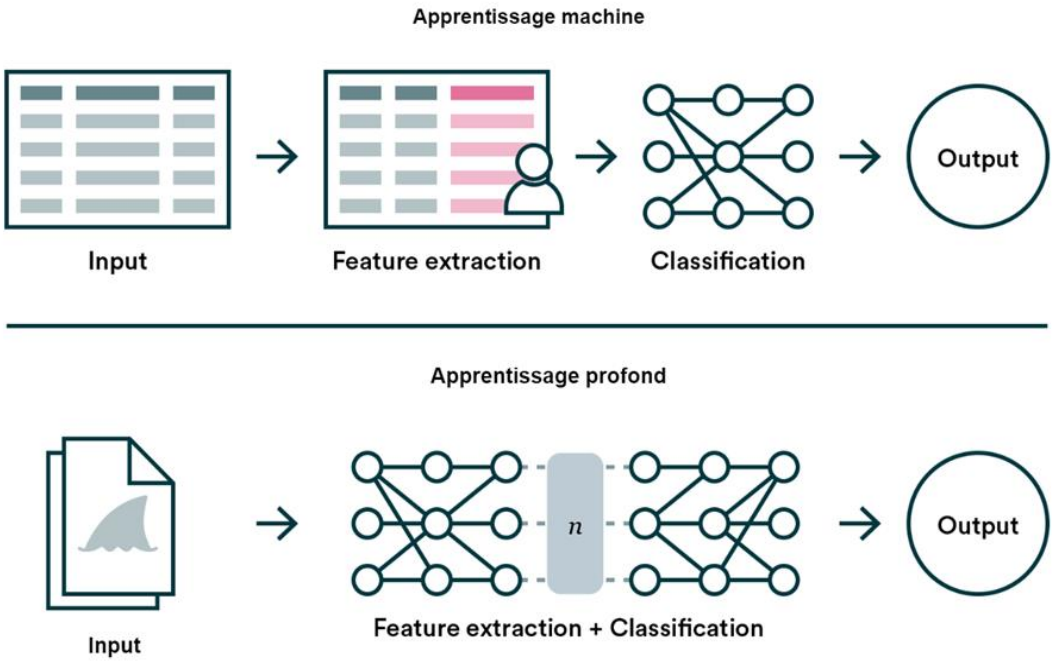

Figure 14 : La différence entre l'apprentissage machine et l'apprentissage profond.

2.4.1 Réseau de neurone

Un réseau neurone artificiel (RNA) est un algorithme d'apprentissage automatique inspiré du fonctionnement des réseaux neurones biologiques. Chaque RNA contient des nœuds (similaire aux corps cellulaires) qui communiquent avec d'autres nœuds via des connexions (similaire aux axones et aux dendrites). De la même manière que les synapses entre les neurones sont renforcées lorsque leurs sorties sont corrélées dans un réseau neuronal biologique, les connexions entre les nœuds d'un RNA sont pondérées en fonction de leur capacité à fournir un résultat souhaité [58].

La **figure15** compare la structure d'un neurone dans un cerveau humain (A) et d'un neurone dans un réseau neuronal artificielle (B).

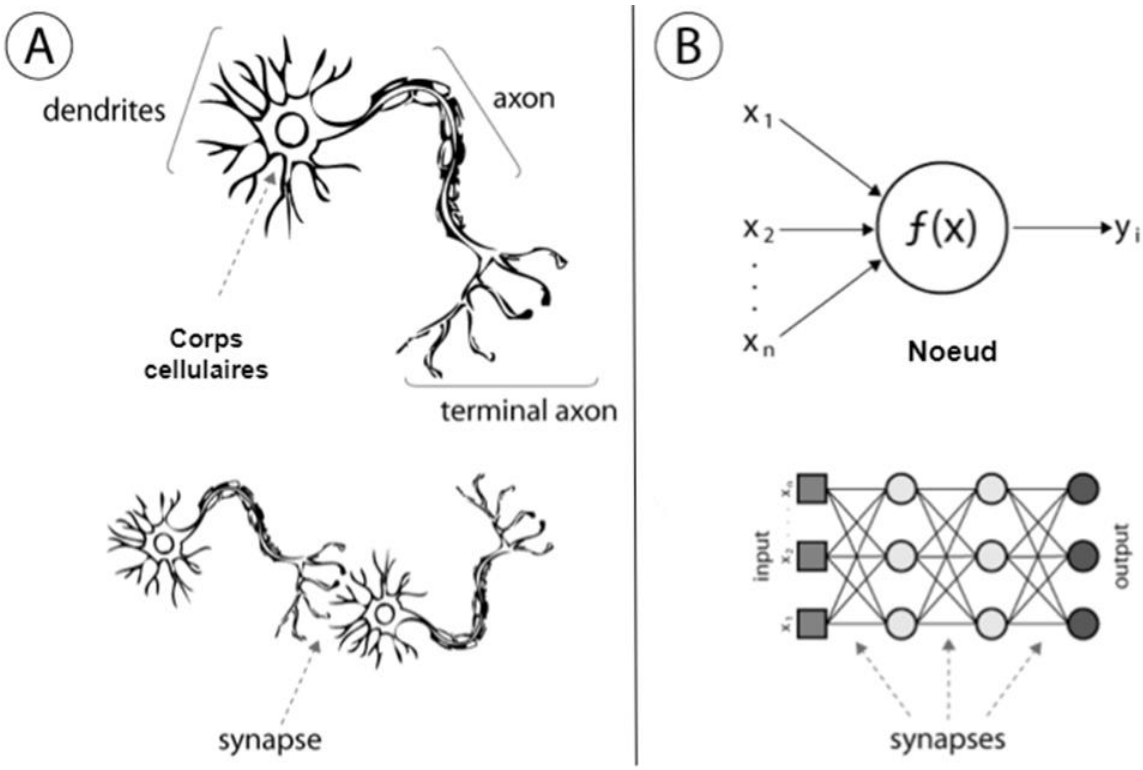

Figure 15 : La relation entre le réseau neurone artificiel et le cerveau humain.

Les réseaux de neurones sont des modèles structurés par des couches empilées les unes sur les autres, chaque étape filtrant successivement les données d'entrée pour apprendre une représentation finale utile pour une tâche donnée [58, 59].

La structure d'un réseau neurone de base est illustrée à la **figure16**.

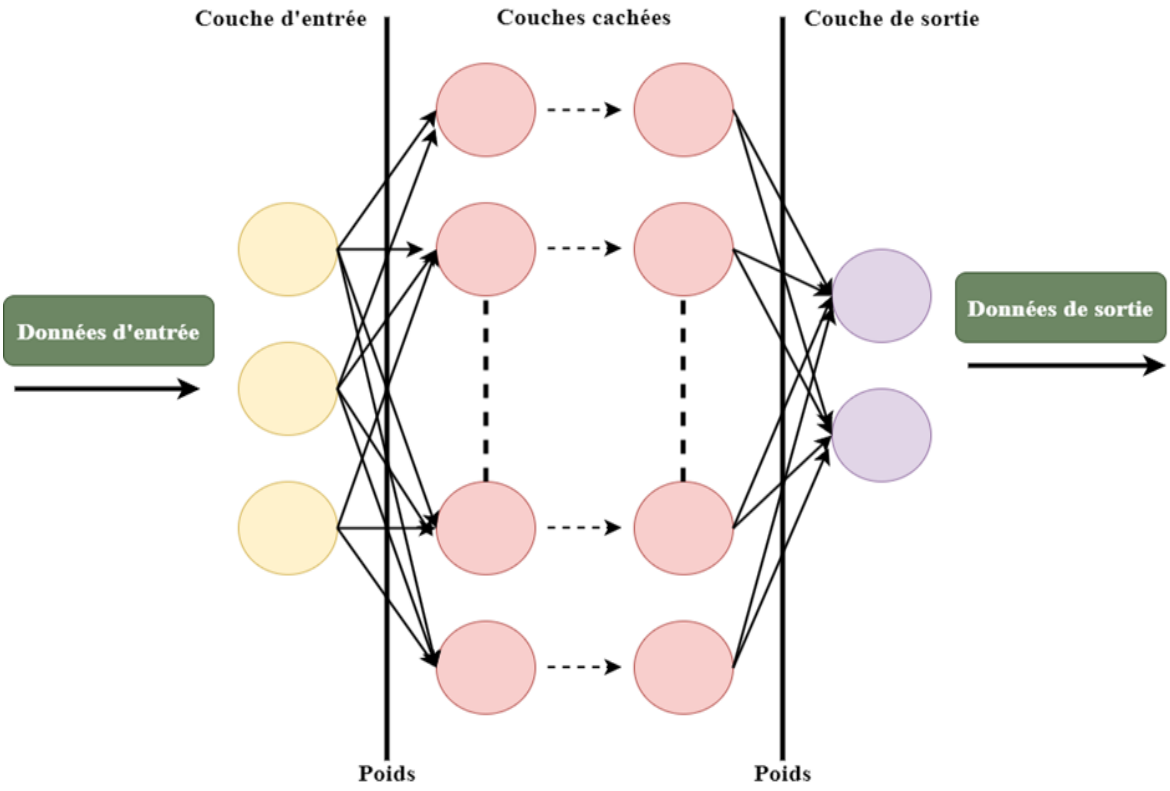

Figure 16 : La structure d'un réseau neuronal de base.

La formation d'un réseau neurone implique quatre composants :

- Données d'entrée

- Couches

- Fonction de perte

- Optimiseur

Cette méthode d'apprentissage en plusieurs étapes de diverses représentations de données est le principe fondamental de l'apprentissage profond.

Les couches peuvent être définies comme la structure de données fondamentale d'un réseau neurone. Chaque couche traite les entrées et les sorties dans des conteneurs de données appelés tenseurs. Les couches ont aussi fréquemment des valeurs appelées poids. Le but final de la formation d'un réseau neurone est de trouver les valeurs correctes de ces poids, qui spécifient les transformations que la couche effectue sur les données entrantes [58, 59]. Ce paramétrage des couches du réseau est illustré à la **figure17**.

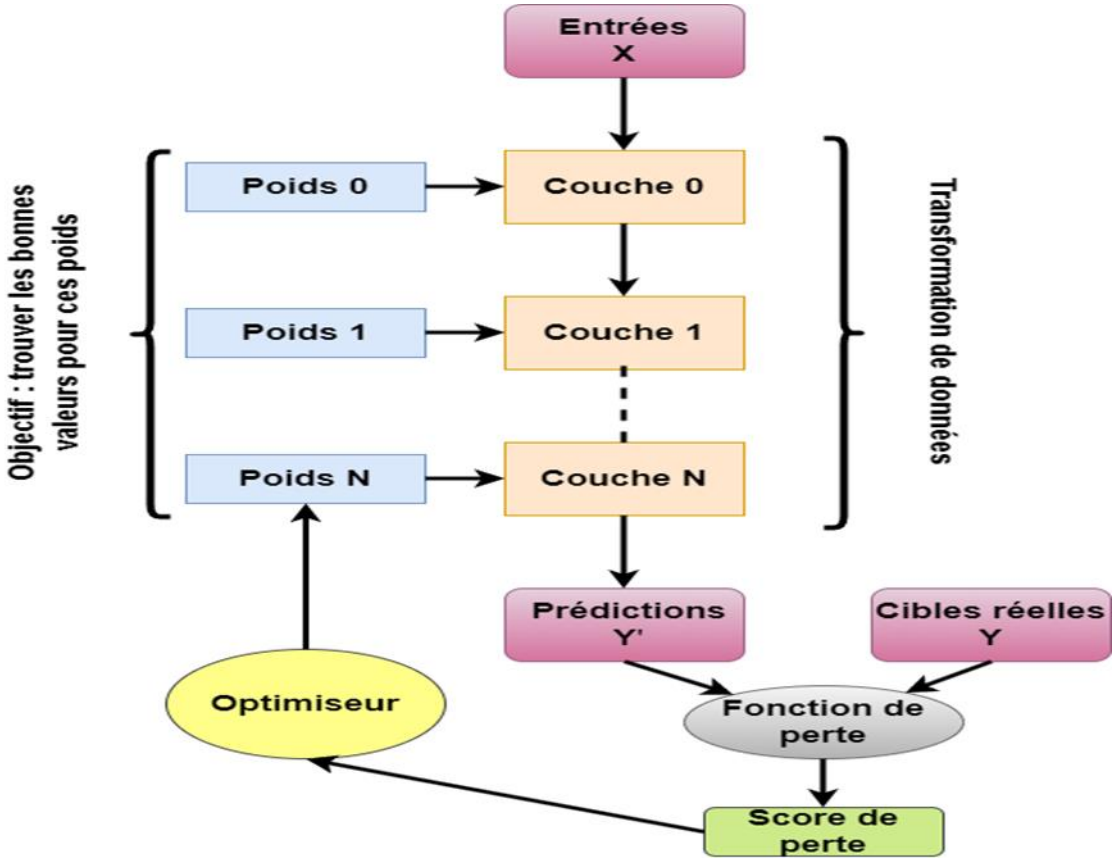

Figure 17 : Boucle d'apprentissage d'un réseau neurone.

La difficulté d'optimiser ces réseaux pour diverses tâches réside dans la configuration de ces poids, car il peut exister des dizaines de millions de paramètres qui dépendent les uns des autres. Pour déterminer une configuration de poids, une fonction de perte est utilisée comme mesure de la performance des représentations des couches de la sortie du réseau par rapport à la sortie attendue. Ce score est ensuite utilisé comme un signal de retour pour l'optimiseur afin d'ajuster les poids des couches du réseau à l'aide d'un algorithme de rétro-propagation [58, 59].

L'algorithme de rétro-propagation commence par la valeur de perte finale et travaille du haut vers le bas des couches pour calculer la contribution de chaque paramètre à cette valeur [58, 59]. **La descente de gradient stochastique** est le processus qui consiste à prendre la dérivée de la fonction de perte à un point local de la formation pour déterminer les ajustements nécessaires pour minimiser la perte et optimiser le modèle par rapport à une métrique d'évaluation définie [58, 59].

La représentation finale d'une boucle de formation de réseau neuronal est illustrée à la **figure17**. Cette boucle de formation commence par une initialisation aléatoire des poids, ce qui entraîne un score de perte élevé puisque ces poids initiaux sont probablement incorrects pour représenter les données de manière utile. Cependant, avec chaque lot de

formation, le réseau ajuste ses poids et commence lentement à s'améliorer jusqu'à ce que le réseau ait un score de perte minimal pouvant être généralisé à l'ensemble de test [58, 59].

2.5. Réseaux de neurons convolutionnels CNNs

Un CNNs [19, 63] contient plusieurs couches de Convolution, alternées entre plusieurs couches de Pooling, et de couches entièrement connectées. Réseaux de neurones convolutionnels ont de bonnes performances dans le traitement des images.
La figure14 montre un exemple classique de CNNs.

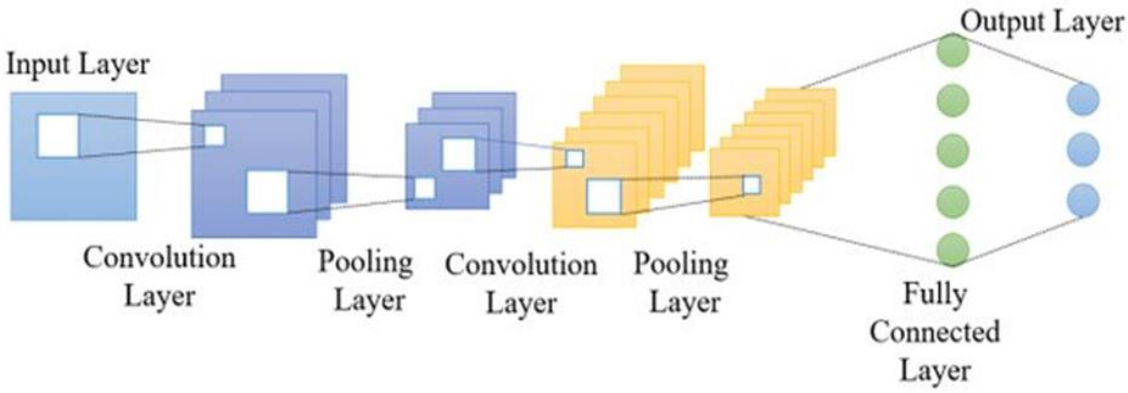

Figure 18 : L'architecture du réseau neurone convolutionnel classique.

2.5.1. Convolution

Une couche de convolution se compose d'un ensemble de noyaux **K** (ou filtres) qui sont utilisés pour convoluer l'image d'entrée **X**, en extrayant une variété de caractéristiques de l'image d'entrée, générant une carte de caractéristiques **Y** pour chaque noyau. Une fois les cartes de caractéristiques créées, une fonction d'activation est utilisée pour introduire une non-linéarité dans les cartes de caractéristiques pré-activées **Y**, ce qui donne une sortie **A**. Lors de la création d'une couche de convolution, on peut modifier le nombre et la taille des noyaux, la taille du pas, la taille de la fenêtre et la présence ou l'absence de remplissage.

La convolution d'une image d'entrée **X** avec un noyau **K** de taille r×s donnera une sortie **Y**. Chaque position dans la sortie **Y** est définie comme suit :

$$Y_{ij} = (X * K)_{ij} = \sum_{pq} X_{(i+p),(j+q)}\ K_{(p),(q)} \qquad (1)$$

Où p va de 0 à r et q de 0 à s.

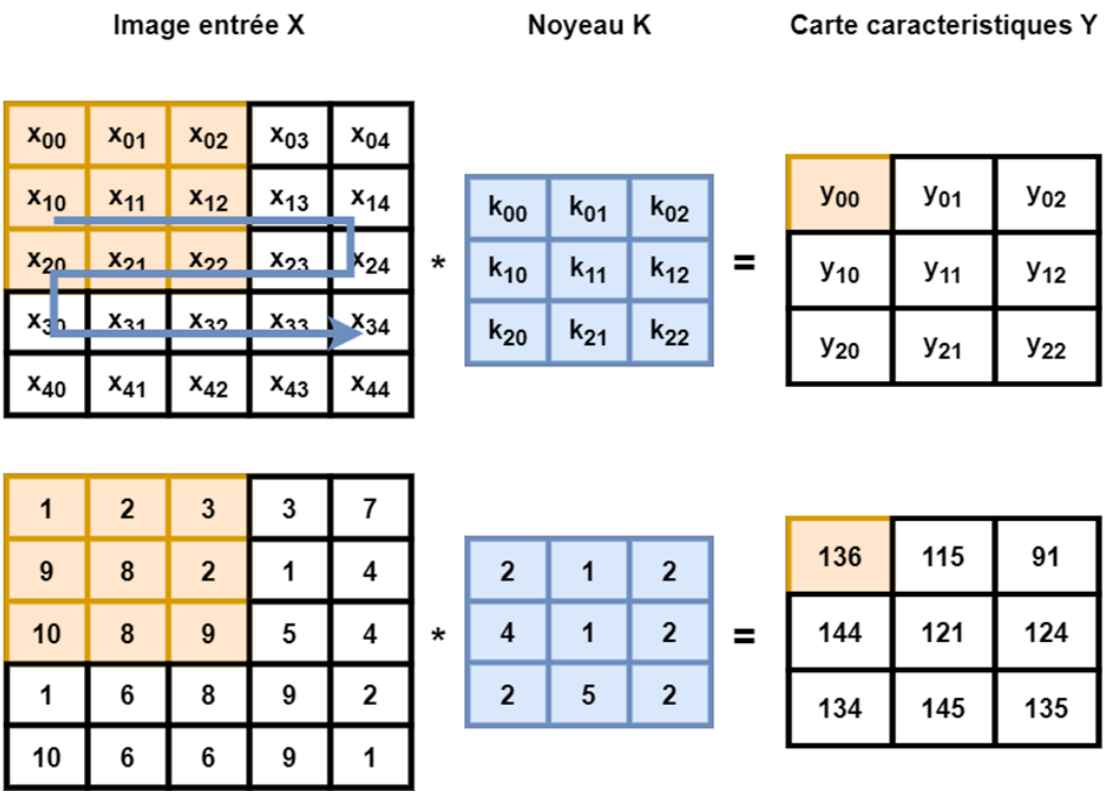

Figure 19 : Opération de convolution sur une image d'entrée avec un seul noyau, en utilisant le pas (1,1).

La couche de convolution est utilisée comme extracteur de caractéristiques en appliquant certains poids apprenables (filtres) sur l'image d'entrée. L'image d'entrée est convoluée avec les filtres afin de calculer les nouvelles cartes de caractéristiques. Pour comprendre l'opération de convolution, nous pouvons observer la **figure19**, et d'après la formule, nous constatons que des convolutions à l'entrée **X** ont été appliquées en utilisant un ensemble de filtres. Chaque filtre **K** est appliqué à un endroit de l'entrée, où il effectue le produit entre chaque élément du noyau et la partie de l'entrée avec laquelle il se superpose. Le résultat de la convolution est additionné pour obtenir la sortie à chaque emplacement **Y**. La nouvelle sortie composée est connue sous le nom de carte de caractéristiques.

Un exemple de la sortie de la première ligne et de la première colonne (0,0) dans la **figure19** est présenté ci-dessous.

$$Y_{00} = (X * K)_{00}$$

$$= X_{0,0}K_{0,0} + X_{0,1}K_{0,1} + X_{0,2}K_{0,2} + X_{1,0}K_{1,0} + X_{1,1}K_{1,1} + X_{1,2}K_{1,2} + X_{2,0}K_{2,0} + X_{2,1}K_{2,1} + X_{2,2}K_{2,2}$$

$$= 1 \text{ x } 2 + 2 \text{ x } 1 + 3 \text{ x } 2 + 9 \text{ x } 4 + 8 \text{ x } 1 + 2 \text{ x } 2 + 10 \text{ x } 2 + 8 \text{ x } 5 + 9 \text{ x } 2$$

$$Y_{00} = 136$$

2.5.2. Le pas (Stride)

Le pas, également appelé stride, consiste à un paramètre de l'opération de convolution qui définit la distance entre les patches qui sont extraits de la carte de caractéristiques d'entrée. Avec un pas de 2, la largeur et la hauteur de la carte de caractéristiques de sortie sont sous-échantillonnées par un facteur de 2. C'est un facteur qui peut contribuer à une taille de sortie différente dans les CNN.

2.5.3. Fonction d'activation

La couche de convolution se termine par une fonction d'activation. Traditionnellement, une fonction Tanh ou sigmoïde était utilisée. Cependant, des travaux récents ont montré que les fonctions ReLU (rectified linear unit) nécessitent moins de puissance de traitement et entraînent le réseau plus rapidement.

On augmente la non-linéarité en appliquant la fonction d'activation f(x) = max(0, x) à chaque valeur de la sortie de pré-activation Y. La couche de correction ReLU remplace donc toutes les valeurs négatives reçues en entrées par des zéros. La **figure20** illustre ce fonctionnement.

$$f(x) = \begin{cases} x, & x > 0 \\ 0, & x < 0 \end{cases}$$

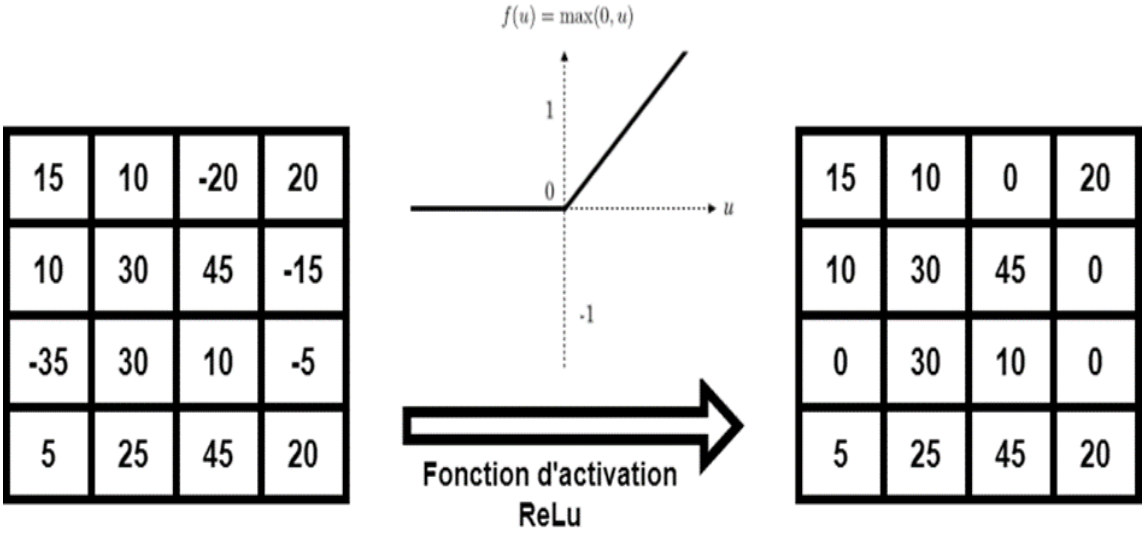

Figure 20 : Fonction d'activation.

2.5.4. Pooling

La couche Pooling est couramment utilisée dans les réseaux entre les couches de convolution pour réduire la taille spatiale de l'entrée, la quantité de paramètres et le temps de calcul du réseau. Elle applique une opération similaire à la convolution, mais remplace la combinaison linéaire par une autre fonction, telle que le maximum Max-Pooling ou la

moyenne Mean-pooling de l'entrée des voisins. Dans le cas du mean-pooling, la valeur mise en commun est la moyenne de toutes les valeurs comprises dans la fenêtre de pooling. En ce qui concerne le max-pooling, la valeur mise en commun est la valeur maximale de toutes les valeurs de la fenêtre de pooling. Il a été démontré que le max-pooling a un effet plus favorable sur les performances du réseau. De la même manière que la couche convolution, la couche pooling a les paramètres de taille de noyau et de pas.

La **figure21** illustre le fonctionnement des exemples de max-pooling et de mean-pooling.

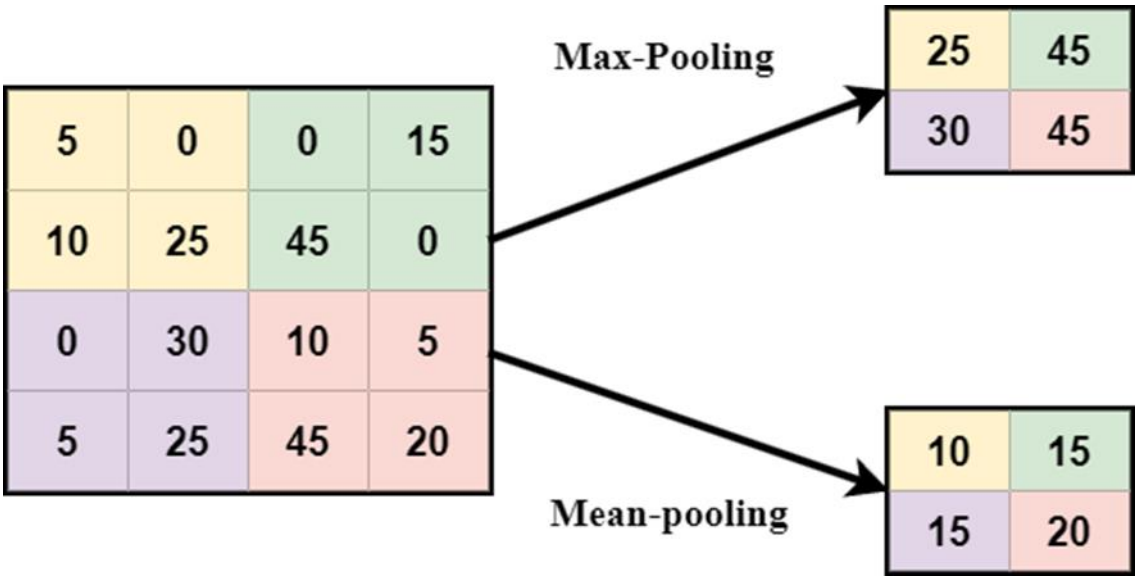

Figure 21 : Couche de Pooling.

2.5.5. Couche entièrement connecté

La couche entièrement connectée suit généralement le groupe de couches de convolution et de Pooling pour extraire les représentations des caractéristiques abstraites de l'image. Un CNN peut avoir une ou plusieurs couches entièrement connectée. Elles expliquent la présentation des caractéristiques et exécutent une fonction de raisonnement de haut niveau. En pratique, la dernière couche entièrement connecté produit les sorties du réseau et contient autant de neurones que le nombre de classes à prédire. Selon le problème, une fonction d'activation peut être ajoutée pour favoriser les valeurs de sortie, par exemple, il utilise les valeurs linéaires pour un problème de régression, mais il est nécessaire d'utiliser la fonction d'activation Sigmoïde [61] dans le cas de la classification binaire.

2.6. La propagation vers l’avant (Forword propagation)

La propagation vers l'avant d'un réseau est le processus par lequel les données d'entrée se déplacent dans une seule direction à travers un réseau. Elle est contrairement à la rétro-propagation, qui se produit lorsqu'une erreur est traitée par le réseau dans la direction opposée, entraînant et mettant à jour chaque couche par la suite [58-60].

La propagation vers l'avant est utilisée dans la passe de formation avec des données de test connues, ainsi qu'après la formation du réseau, pour classer de nouvelles données de test inconnues.

2.7. Réseaux entièrement convolutifs

Le FCN [16] diffère du CNN en changeant les couches entièrement connectées à la fin du CNN par des couches de convolution et ne comprend que des couches convolutives. Ce faisant, on crée un réseau qui calcule un filtre non linéaire pour les vecteurs de sortie de chaque couche. Ainsi, le réseau est équipé pour traiter des entrées de taille variable et produire des images de segmentation avec les mêmes dimensions correspondantes à l’entrée. Cela permet au réseau de fournir une carte segmentation spatiale des classes d'objets désirée au lieu de scores de classification. En ajoutant une perte spatiale au réseau pour obtenir un modèle efficace pour l’apprentissage de bout en bout. Le FCN est aussi capable de préserver l'information spatiale de la source, ce qui est essentiel pour l'application de la segmentation sémantique puisque cette opération demande à la fois la localisation et la classification. La **figure22** présente un FCN.

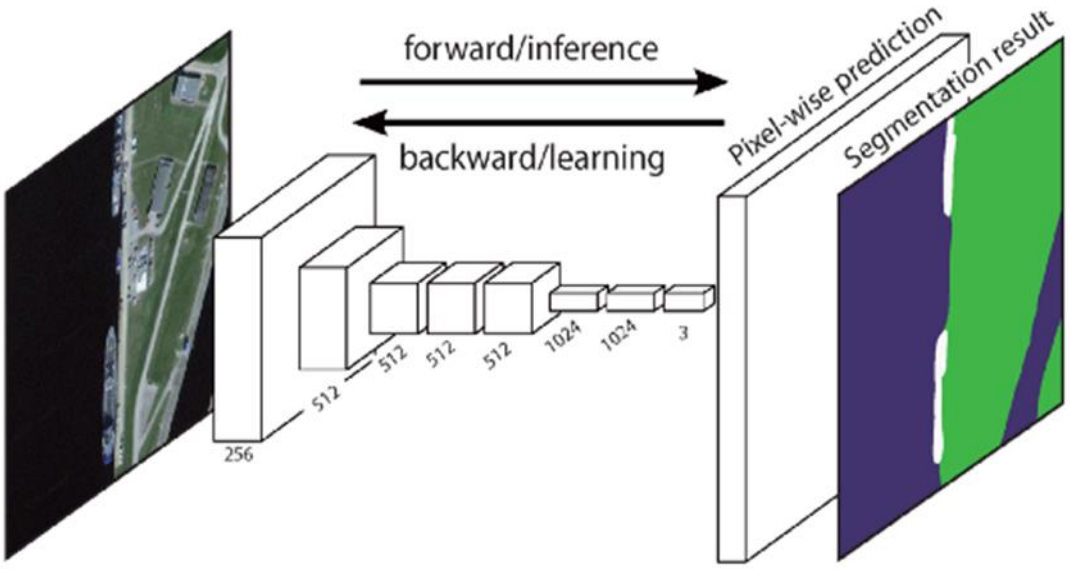

Figure 22 : Réseaux entièrement convolutif.

2.8. SegNet

Le modèle SegNet est un travail connu pour la segmentation, Badrinarayanan et al. [64] ont proposé une architecture qui se compose d'une partie d'encodeur et une partie décodeur convolutif dédies à la segmentation d'images comme le montre la **figure23**. Le réseau de segmentation SegNet, Chaque couche de la partie encodeur a une couche correspondante dans la partie décodeur, les deux parties sont identiques et comprennent 13 couches convolutives du réseau VGG16, suivi par une couche de classification par pixel pour générer des possibilités de classification indépendantes est introduite dans le dernier réseau décodeur. La nouveauté de SegNet réside dans la manière dont le décodeur suréchantillonne (upsamples) les cartes de caractéristiques d'entrée à faible résolution. Plus précisément, il utilise les coefficients de la couche Pooling calculés lors de l'étape du codeur pour effectuer un suréchantillonnage (upsamples) non linéaire. Cela élimine le besoin d'apprendre à sur-échantillonner pour le décodeur. Les cartes sur-échantillonnées sont ensuite convoluées avec des filtres entraînables pour produire des cartes de caractéristiques denses.

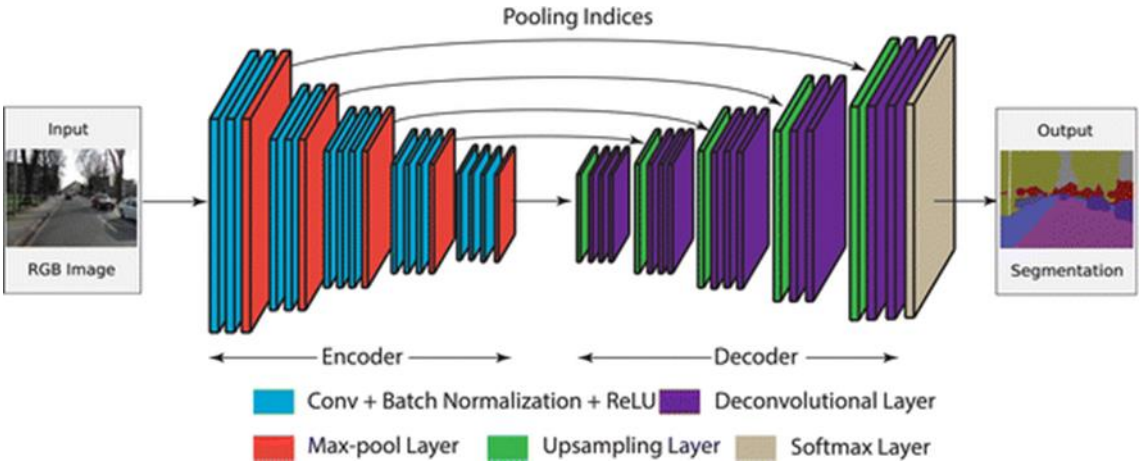

Figure 23 : Architectures du SegNet.

2.9. U-Nets

Le U-Net a été proposé pour la segmentation des images biomédicales [65] . Leur concept d'apprentissage se base sur l'amélioration des données pour apprendre efficacement à partir des quelques images annotées. L'architecture de U-Net comprend deux parties comme le montre la **figure24**, un chemin de compression (côté droit) pour extraire les caractéristiques contextuelle à haute résolution, et un chemin de décompression symétrique (côté gauche) qui permet une localisation précise. La partie de sous-échantillonnage ou de compression possède une architecture de type FCN qui extrait les caractéristiques avec des convolutions 3 × 3. La partie de sur-échantillonnage ou de

décompression utilise la convolution ascendante (ou déconvolution) réduise le nombre de cartes de caractéristiques et augmente leurs dimensions. Les cartes de caractéristiques de la partie de sous-échantillonnage du réseau sont copiées dans la partie de sur-échantillonnage afin d'éviter la perte d'informations caractéristiques. Enfin, une convolution 1×1 exploite les cartes de caractéristiques pour générer une carte de segmentation qui classifie chaque pixel de l'image d'entrée.

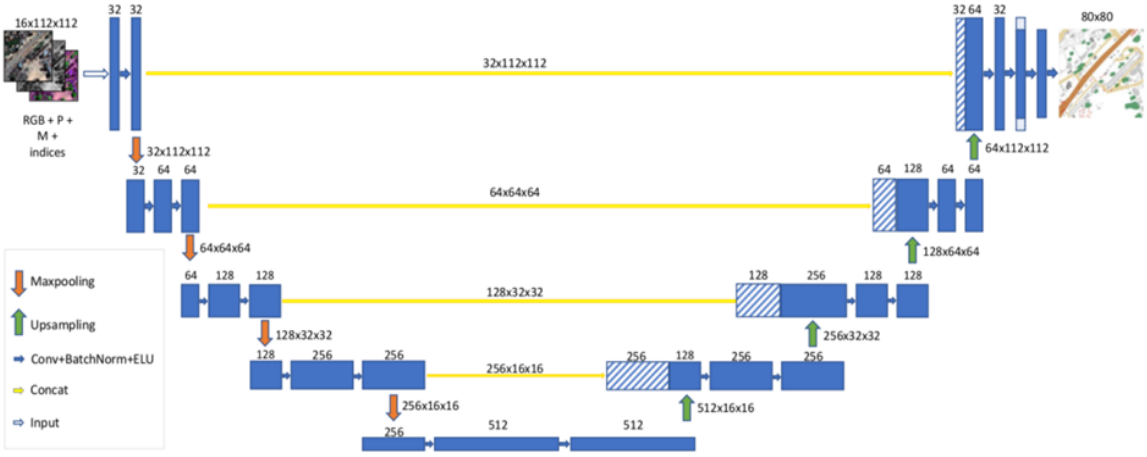

Figure 24 : Architecture du U-Net.

2.10. Sur-ajustement (overfitting)

L'overfitting [58-60] est un phénomène courant dans l'apprentissage automatique où un modèle est trop complexe et se conforme trop étroitement aux données d'entraînement. Cela peut entraîner un fort score sur les données d'entraînement, mais une performance médiocre sur les nouvelles données. Il existe plusieurs raisons pour lesquelles l'overfitting peut se produire dans un modèle de deep learning, notamment :

- Trop de capacité : lorsque le modèle est trop complexe par rapport à la taille des données d'entraînement, il peut sur-ajuster les données d'entraînement en apprenant les fluctuations et les bruit plutôt que les informations caractéristiques générales.
- Peu de données d'entraînement : lorsqu'il y a peu de données d'entraînement, le modèle peut sur-ajuster en apprenant les détails spécifiques des données d'entraînement plutôt que les informations brutes générales.
- Informations caractéristique redondants : lorsque le modèle est formé avec des features redondants, il peut sur-ajuster en apprenant les relations entre les features au lieu de la fonction d'association cible.

Pour éviter l'overfitting, il est souvent nécessaire d'utiliser des techniques de régularisation telles que le Dropout.

Les approches courantes pour éviter l'overfitting sont **l'arrêt précoce**, **la régularisation**, **l'augmentation des données, l'apprentissage par transfert et le Dropout** [58-60]**.**

2.10.1. L'arrêt précoce (early stopping)

Pour mesurer si le modèle est sur-ajusté, nous pouvons diviser l'ensemble de données en ensembles de formation et de test. L'ensemble de test est utilisé pour déterminer la performance sur de nouvelles données non vues et ne sont pas directement utilisé pour mettre à jour les poids. Seul l'ensemble d'entraînement est utilisé pour modifier les paramètres. Pendant la formation, nous comparons l'erreur de classification sur l'ensemble de formation et l'ensemble de test et lorsque la précision sur l'ensemble de test commence à diminuer, la formation est arrêtée pour éviter le sur-ajustement. Cette procédure est connue sous le nom d'arrêt précoce.

2.10.2. La régularisation

En outre, pour éviter l'ajustement excessif, un terme de régularisation [58] est généralement ajouté à la fonction de perte. En général, les valeurs extrêmes des paramètres sont observées lorsque le sur-ajustement se produit. Ainsi, la norme des valeurs de poids peut être incluse dans la fonction de perte, de sorte que les poids sont favorisés pour être dans une gamme raisonnable. L'ajustement excessif est minimisé et a été donné par l'entropie croisée et la normalité des paramètres.

2.10.3. L'augmentation des données

L'augmentation des données [66, 67] est souvent utilisée pour élargir virtuellement le nombre d'échantillons pour la formation. L'augmentation des échantillons de données se fait par l'application aléatoire d'un ensemble prédéfini de transformations sur les données d'apprentissage. Par exemple, dans la classification d'images, les échantillons de données peuvent être inversés ou tournés de manière aléatoire pour augmenter la taille de l'ensemble d'apprentissage.

2.10.4. L'apprentissage par transfert (transfer learning)

L'apprentissage par transfert est une approche intéressante pour former des modèles d'apprentissage profond efficaces lorsque seuls de petits ensembles de données sont disponibles. Par rapport aux modèles d'apprentissage profond formés à partir de zéro, la technique d'apprentissage par transfert vise à améliorer la précision du modèle en utilisant une puissance de calcul moindre. Le concept d'apprentissage par transfert peut

être considéré comme une technique en deux étapes. Tout d'abord, il vise à apprendre la représentation des données en entraînant un modèle sur des ensembles de données disponibles contenant un grand nombre de données annotées. Ensuite, il utilise cette représentation pour construire un nouveau modèle basé sur le modèle pré-entraîné en utilisant un ensemble de données plus petit, en n'entraînant que certaines couches sélectionnées ou la couche de décision finale [67, 68].

L'apprentissage par transfert [69, 70] est une méthode d'apprentissage automatique, où un modèle développé pour une tâche donnée est réutilisé comme référence pour un autre modèle sur une seconde tâche. Le concept est d'utiliser les connaissances apprises du premier modèle lors de la résolution d'un nouveau problème. En d'autres termes, on peut dire qu'il s'agit d'un transfert de connaissances. Cependant, l'avantage d'utiliser l'apprentissage par transfert est qu'il n'est pas nécessaire de former de grands ensembles de données pour éviter le sur-ajustement et que peu de ressources de calculs sont nécessaires.

2.10.5. Dropout

Le dropout [19, 58] est une technique de régularisation utilisée dans les réseaux de neurones en profondeur. Il consiste à aléatoirement "ignorer" certains neurones lors de chaque itération de l'entraînement, ce qui permet d'éviter l'overfitting et de rendre le modèle plus robuste. Cela se fait en mettant à zéro une fraction de la sortie de chaque neurone pendant l'entraînement, ce qui force le modèle à ne pas se reposer sur un petit nombre de neurones fortement corrélés.

2.11. Généralité de l’apprentissage profond pour la segmentation

2.11.1. Optimiseurs

Les optimiseurs [58] de deep learning sont des algorithmes utilisés pour entraîner les réseaux de neurones profonds. Ils permettent de mettre à jour les poids des réseaux de neurones de manière à minimiser une fonction de coût donnée. Il existe de nombreux optimiseurs différents qui ont été développés pour la tâche de segmentation d'images, les optimiseurs les plus couramment utilisés incluent :

- Descente de gradient stochastique (SGD) : un optimiseur simple qui met à jour les poids du réseau de neurones en utilisant la dérivée de la fonction de coût par rapport à chaque poids.
- Adam : un optimiseur qui utilise une combinaison de la méthode de descente de gradient stochastique et de la méthode de descente de gradient à mémoire à long

terme. Il est souvent utilisé pour les tâches de classification d'images et de segmentation car il est généralement rapide et performant.

- RMSProp : un optimiseur qui utilise une moyenne pondérée des dérivées passées pour mettre à jour les poids du réseau de neurones. Il est souvent utilisé pour les tâches de segmentation car il est efficace pour éviter les oscillations dans les gradients.

2.11.2. Fonction de perte

La fonction de perte [58] (aussi appelée fonction de coût) est utilisée pour mesurer la qualité des prédictions du modèle par rapport aux valeurs attendues (annotations). Plus précisément, la fonction de coût calcule une valeur numérique qui représente l'erreur du modèle. En segmentation d'image, il existe plusieurs fonctions de coût couramment utilisées, notamment :

- Binary Cross-Entropy (BCE) : une fonction de coût binaire qui mesure la différence entre les prédictions binaires et les annotations binaires.
- Dice Loss : une fonction de coût qui mesure la similitude entre les prédictions et les annotations en utilisant le coefficient Dice.
- Intersection over Union (IoU) : une fonction de coût qui mesure la similitude entre les prédictions et les annotations en utilisant le rapport d'intersection sur l'union.
- Focal Loss : une fonction de coût qui se concentre davantage sur les erreurs dans les prédictions qui sont les plus difficiles à corriger.

2.11.3. Couche de sortie (Sigmoid/Softmax)

En segmentation d'image, il est courant d'utiliser soit la fonction sigmoïde, soit la fonction softmax comme une fonction d’activation pour la couche de sortie du modèle.

- La fonction sigmoïde : Elle est utilisée lorsque la segmentation est une tâche de classification binaire, c'est-à-dire que chaque pixel de l'image doit être classé soit comme appartenant à l'objet d'intérêt (classe 1), soit en dehors de celui-ci (classe 0). La fonction sigmoïde produit une sortie entre 0 et 1, qui peut être interprétée comme la probabilité d'appartenance à la classe positive.
- La fonction softmax : Elle est utilisée lorsque la segmentation est une tâche de classification multi-classes, c'est-à-dire que chaque pixel de l'image doit être classé dans l'une des plusieurs classes possibles. La fonction softmax produit une sortie qui représente les probabilités d'appartenance à chaque classe.

2.12. Défis de l’apprentissage profond pour la segmentation

La segmentation d'images et vidéo peut présenter de nombreux défis, notamment en ce qui concerne l'overfitting, la nécessité de quantités importantes de données, le temps d'apprentissage long et le choix des hyper-paramètres.

- **Sur-apprentissage (Overfitting) :** Les réseaux de neurones profonds peuvent facilement surapprendre les données d'entraînement et ne pas se généraliser sur les données de test, ce qui peut entraîner une mauvaise performance.
- **Nécessité de quantités importantes de données :** Les réseaux de neurones profonds nécessitent souvent de grandes quantités de données d'entraînement pour fonctionner efficacement, ce qui peut poser des problèmes lorsque les données sont rares ou coûteuses à obtenir.
- **Temps de formation long :** L'entraînement des réseaux de neurones profonds peut être très lent et nécessite beaucoup de temps de calcul, ce qui peut être un défi pour les applications en temps réel.
- **Choix des hyper-paramètres :** Le choix des hyper-paramètres, tels que la taille du réseau, le nombre d'époques (itérations) d'entraînement et le choix des optimiseurs adéquat, peut être difficile et requiert souvent une analyse expérimentale pour trouver les meilleurs réglages.

3.13. Conclusion

Ce chapitre traite principalement la définition et l'évolution des réseaux de neurones, des règles d’apprentissage, et des connaissances liées aux réseaux de neurone profond. Des modèles populaires relatives à la segmentation à base d’auto-encodeur sont présenté tels que CNN, FCN, SegNet et U-nets. En outre, ce chapitre présente également les problèmes d’apprentissage automatique couramment rencontrés dans l'ingénierie de l'intelligence artificielle et plus précisément les défis rencontrés lors des taches de segmentation d’image et vidéo comme le sur-ajustement et leurs solutions.

Chapitre 3 : Méthodologie

3. Méthodologie

3.1. Introduction

Dans cette section, nous allons présenter les méthodes et le matériel utilisés pour construire un système de segmentation d'objets en mouvement robustes à partir de la vidéosurveillance et un système de segmentation des routes rapide et précis à partir des images aériennes. Nous fournissons une description détaillée de chacun des principaux éléments constitutifs introduits dans notre méthodologie. Ainsi, nous présentons l'implémentation des approches proposées. Par contre, les détails des approches proposées et les différentes stratégies d'augmentation des données sont présentés.

3.2. Auto-Encoder

L'auto-encodeur est un type de réseau neurone profond qui connaît un grand succès et est considéré comme l'un des algorithmes non supervisés. Il vise à reproduire les données d'entrée à la sortie [71, 72], où les couches d'entrée et de sortie ont le même nombre de neurones.

L'auto-encodeur se compose de deux parties principales, à savoir l'encodeur et le décodeur. Le rôle principal de l'encodeur est de compresser les données d'entrée en une représentation de dimension inférieure par l'utilisation d'une transformation non linéaire tout en préservant les caractéristiques essentielles des données en supprimant les éléments inutiles. Ensuite, les sorties du codeur sont transmises au décodeur qui les décompresse et reconstruit les données d'origine à partir des données de dimension inférieure générées. Selon la littérature, il existe quatre architectures principales d'auto-codeurs, à savoir l'auto-codeur convolutif, l'auto-codeur variationnel, l'auto-codeur de débruitage et l'auto-codeur clairsemé. Les auto-encodeurs peuvent être adoptés dans plusieurs applications comme le débruitage des données et la réduction de la dimensionnalité [73]. La **figure25** montre l'architecture globale du réseau.

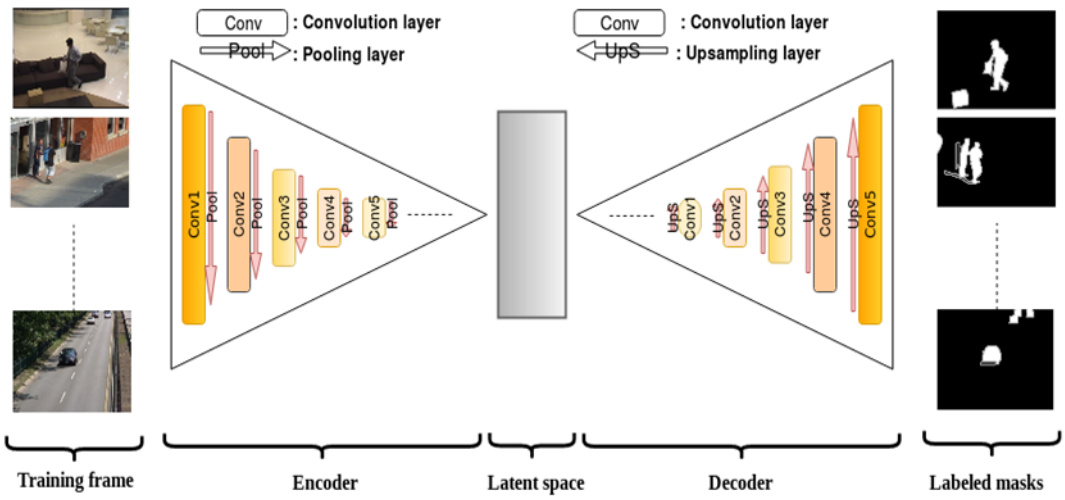

Figure 25 : L'architecture de l'approche générale de l'Auto-Encoder.

L'auto-encodeur comporte trois blocs essentiels :

- **La partie encodeur** : L'encodeur vise à encoder toutes les informations pertinentes sur l'entrée dans l'espace latent.
- **Espace latent** : Il représente l'espace représenté par une forme comprimée de l'entrée.
- **Partie décodeur** : Le décodeur vise à reproduire les données d'entrée au niveau de la sortie en se concentrant uniquement sur les données de l'espace latent.

Le codage des données d'entrée X via la fonction de codage non linéaire E donnant Z=E(X), puis le décodage de Z donnant Y=D(Z) via la fonction de décodage non linéaire D qui se rapproche des données d'origine X. Comme le montre la **figure26**.

Nous pouvons décrire cet algorithme dans sa forme la plus simple comme suit :

$$Y = D\big(E(X)\big) \qquad (2)$$

Le processus d'apprentissage minimise la fonction de perte entre l'entrée X et la sortie Y comme :

$$Loss(\mathrm{X},\mathrm{Y}) = \big|\big|\mathrm{X}-\mathrm{Y}\big|\big|^2 = \big|\big|\mathrm{X}-\mathrm{D}(\mathrm{Z})\big|\big|^2 = \big|\big|\mathrm{X}-\mathrm{D}(\mathrm{E}(\mathrm{X}))\big|\big|^2 \qquad (3)$$

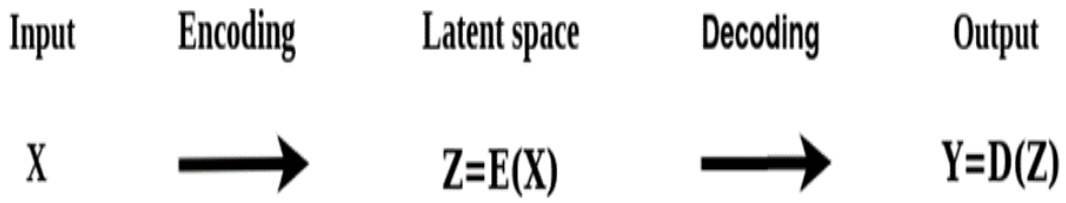

Figure 26 : L'expression de base d'un Auto-Encoder général.

3.3. Formation des données

Dans cette thèse, nous proposons des nouvelles techniques d'augmentation des données basée sur l'amélioration des images d'apprentissage par les méthodes de bases de l'apprentissage machine. Lors de l'entraînement des modèles d'apprentissage profond, il est nécessaire de préparer suffisamment de données d'entraînement prises dans différentes conditions. Cependant, les méthodes traditionnelles d'augmentation des données sont limitées à la transformation physique et aux conditions de prise de vue des images, de sorte que les conditions de prise de vue sont insuffisantes pour l'augmentation des données. Par conséquent, nous utilisons des méthodes d'amélioration d'image pour générer plus d'images avec différentes orientations pour construire une base d'apprentissage de haute qualité.

3.3.1. Partie 1 : Approche de la segmentation des objets en mouvement (vidéosurveillance)

3.3.1.1. Dataset Cdnet 2014

Le jeu de données Cdnet2014 (détection des changements) [74] est adopté pour entraîner et tester notre modèle. Il est composé de vidéos réelles capturées dans des scénarios difficiles, comme le montre le **tableau 1** et la **figure27**. Pour une plus grande généralisation du processus d'entraînement, nous sélectionnons toutes les séquences vidéo (53 scènes), qui contiennent 11 catégories de vidéos du jeu de données CDnet2014 ; chaque vidéo a une moyenne de 2000 images. Chaque catégorie comporte plusieurs vidéos, ce qui permet de tester les algorithmes de segmentation de manière complète et détaillée dans un large éventail de scénarios de surveillance vidéo.

Le **tableau1** et la **figure27** montrent la liste et les images des catégories et une description de chaque catégorie vidéo dans le jeu de données CDnet 2014.

Tableau 1 : Liste des catégories et description de chaque vidéos dans le jeu de données CDnet2014.

Categories	Challenge et description
Baseline	vidéos de scènes statiques sans mouvements importants.
Camera Jitter	vidéos montrant des tremblements de caméra.
Bad Weather	vidéos montrant des conditions météorologiques défavorables, telles que la pluie, la neige, le brouillard, etc.
Dynamic Background	vidéos montrant des mouvements d'arrière-plan, tels que des voitures en mouvement, des feuilles soufflées par le vent, etc.
Intermittent Object Motion	vidéos montrant des objets en mouvement intermittents, tels que des personnes qui entrent et sortent d'un champ de vue.
Low Frame rate	vidéos à faible fréquence d'images.
Night Videos	vidéos nocturnes montrant des scènes dans l'obscurité.
PTZ	vidéos montrant des mouvements de caméra télécommandée, tels que des panoramiques, des inclinaisons et des zooms.
Shadow	vidéos montrant des ombres causées par des objets en mouvement.
Thermal	vidéos thermiques montrant des scènes à la lumière infrarouge.
Turbulence	Vidéos des scènes affectées par des perturbations visuelles telles que les vibrations, les ondulations ou les secousses.

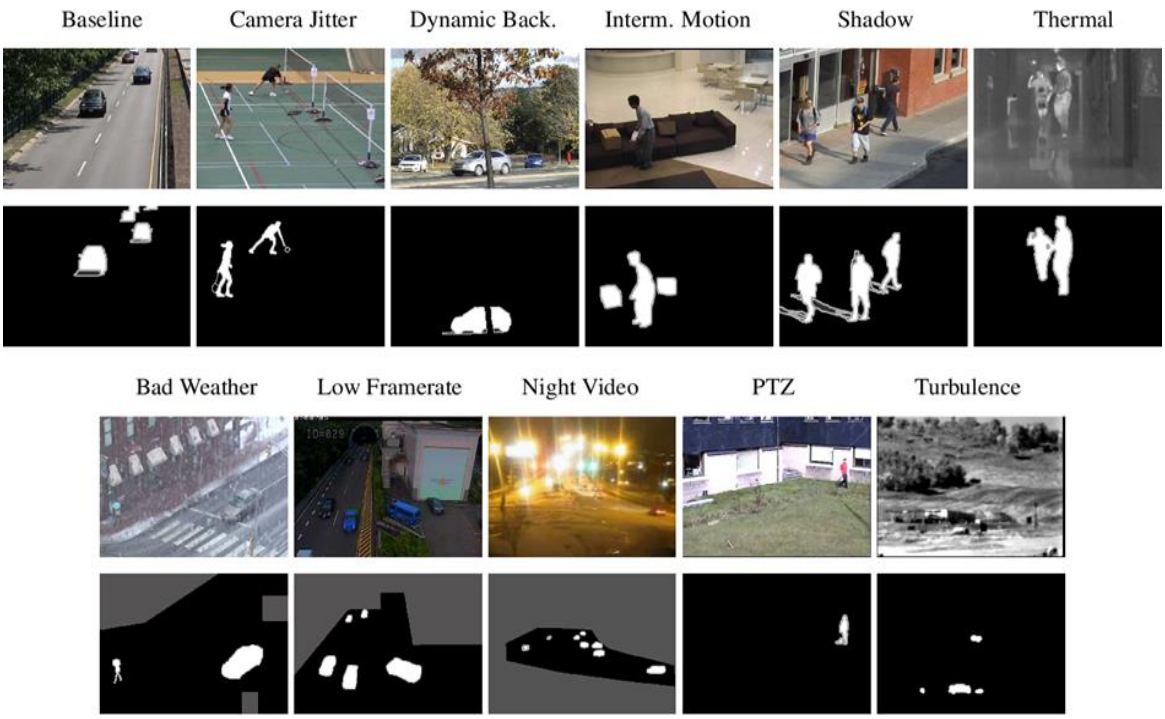

Figure 27 : Les images des catégories et des noms de vidéos dans le jeu de données CDnet2014.

3.3.1.2. Augmentation des données avec l'analyse en composantes principales (ACP)

- **l'analyse en composantes principales (ACP)**

L'analyse en composantes principales (ACP) [75] est une technique statistique non supervisée basée sur une simple transformation linéaire, c'est une technique de réduction de la dimensionnalité [76]. Cependant, l'objectif principal d'une ACP est de compresser les données en conservant le plus d'informations possible. Cela peut être particulièrement utile pour les images, qui ont souvent un grand nombre de pixels et de canaux de couleur. Elle est utilisée dans de nombreuses applications du traitement d'image comme la compression d'image[77] et la reconnaissance de visage[78].

Lorsque l'on utilise ACP pour une image en entrée, le processus consiste à transformer l'espace des caractéristiques (les pixels) en un espace de composantes principales. Cela permet de réduire le nombre de dimensions des données sans perdre trop d'informations. Les composantes principales sont des vecteurs qui sont orthogonaux les uns aux autres et qui décrivent les directions dans lesquelles les données varient le plus. En effet, dans notre cas, elle sera utilisée comme nouvelle technique d'extraction de caractéristiques, de compression et de reconstruction pour préserver et extraire de nouvelles caractéristiques essentielles de manière linéaire à différents niveaux de la distribution des données, dans le but de l'utiliser comme une nouvelle technique d'augmentation des données.

La mise en œuvre utilisée pour la compression et la reconstruction des images couleur pour notre proposition à l'aide de l'analyse en composantes principales (ACP) peut être divisée en 3 étapes principales :

- Découpage des images en 3 canaux R, G et B : convertir l'image en une matrice de pixels.
- Réalisation de l'ACP et sélection des N valeurs propres les plus dominantes sur chaque matrice de valeurs de couleur : calculer la matrice de covariance des pixels. Ensuite, calculer les vecteurs propres de la matrice de covariance, qui sont les composantes principales de l'image.
- Recréer les images originales en fusionnant les composantes R, G et B (on peut utiliser ces composantes principales pour réduire le nombre de dimensions de l'image).

- **Augmentation des données**

La performance des modèles d'apprentissage profond dépend de la taille de l'ensemble de données d'entraînement. Cependant, le manque de jeux de données dans plusieurs domaines, est l'un des problèmes les plus critiques auxquels sont confrontés les

chercheurs. Pour résoudre un tel problème, plusieurs solutions ont été proposées au fil des ans, notamment l'apprentissage par transfert et l'augmentation des données.

L'augmentation des données est une procédure qui vise à élargir la taille de l'ensemble de données en appliquant certaines transformations, où les images originales et créées sont utilisées pour entraîner le modèle [66]. Par conséquent, notre principal objectif est d'utiliser l'ensemble de données existantes pour générer de nouvelles données afin d'éviter le problème de sur-ajustement tout en améliorant les performances du modèle. L'une des principales techniques d'augmentation des données consiste à effectuer certains ajustements et transformations géométriques, notamment le recadrage, la translation, la mise à l'échelle, le miroir, la rotation et la modification des conditions d'éclairage. Ces méthodes sont largement utilisées dans la littérature pour résoudre des problèmes liés au traitement des images et des vidéos, notamment la détection, la reconnaissance et la segmentation.

Les augmentations géométriques citées ci-dessous.

- Réflexions (Flip)

Les réflexions d'images est une technique d'augmentation de données qui consiste à renverser horizontalement ou verticalement les images. Le Flip peut être effectué horizontalement, verticalement ou les deux. Cette augmentation est l'une des plus faciles à mettre en œuvre et s'est avérée utile sur des jeux de données impliquant la reconnaissance en permettant au modèle d'apprendre à reconnaître les objets sous différents angles et orientations.

- Recadrage (Cropping)

Le recadrage ou cropping des images peut être utilisé comme une étape de traitement pratique pour les données d'image ayant des dimensions différente en hauteur et en largeur, en recadrant une partie centrale de chaque image. En outre, le recadrage peut également aider à améliorer la généralisation du modèle, en permettant au modèle d'apprendre à reconnaître les objets sous différents angles et dans différentes positions.

- Rotation

La rotation est une technique courante d'augmentation de données qui consiste à tourner les images autour d'un axe central compris entre 1° et 360°. Lors de la rotation, l'image est transformée en utilisant une transformation géométrique, telle qu'une matrice de rotation. Les augmentations par rotation sont déterminées par le paramètre du degré de rotation. Lorsque plusieurs angles sont utilisés, cela peut entraîner une augmentation significative de la taille du jeu de données.

- Translation (Shift)

La translation est une technique d'augmentation de données qui consiste à déplacer les images de différentes manières vers la gauche, la droite, le haut ou le bas, ou à un emplacement aléatoire. Ces transformations sont utiles pour éviter les problèmes de position dans les données d'apprentissage. Lorsque l'image originale est translatée dans une direction, l'espace restant peut être rempli soit d'une valeur constante telle que 0 ou 255, soit d'un bruit aléatoire ou gaussien. Ce remplissage préserve les dimensions spatiales de l'image après l'augmentation.

- Zoom

Le zoom est une technique courante d'augmentation de données qui consiste à agrandir ou à réduire les images pour créer de nouvelles données. Lors du zoom, l'image est transformée en utilisant une transformation géométrique telle qu'une matrice de zoom. Le zoom peut également aider à améliorer la généralisation du modèle, en permettant au modèle d'apprendre à reconnaître les objets à différentes échelles.

- Décalage de couleur (Color shifting)

Le color shifting est une technique qui consiste à changer les couleurs des images. Le décalage de couleur peut inclure des opérations telles que le changement de luminosité, de contraste, de saturation et de balance des couleurs. Ces opérations sont généralement effectuées en modifiant les valeurs des pixels dans l'espace de couleur. Cette augmentation peut également aider à améliorer la généralisation du modèle, en permettant au modèle d'apprendre à reconnaître les objets sous différentes conditions de lumière et de couleur.

- L'ajout de bruit

Les images peuvent être modifiées en ajoutant du bruit pour créer de nouvelles données. Le bruit peut inclure des anomalies telles que le bruit blanc gaussien, le bruit poissonnier et le bruit salt & pepper. Ce bruit peut être ajouté aux images en modifiant aléatoirement les valeurs des pixels. L'ajout de bruit peut également aider à améliorer la généralisation du modèle, en permettant au modèle d'apprendre à reconnaître les objets en présence de bruit ou de perturbations.

- Combiner de multiples techniques

Ce procédé consiste à combinées les différentes techniques d'augmentation de données pour créer de nouvelles données d'apprentissage.

3.3.1.3. Protocole d’expérience

Les approches proposées dans cette étude sont basées sur l'apprentissage par transfert et d’architecture à multi-profondeur. Pour effectuer la segmentation de l'objet en mouvement, nous utilisons l’architecture d'auto-encodeur comme un apprentissage supervisé pour les deux approches.

Pour la première approche de l'apprentissage par transfert, nous construisons le réseau d’auto-encodeur en ajustant le réseau VGG-16 [40], [79] qui a été pré-entraîné sur le populaire jeu de données ImageNet [40], [19] comme partie encodeur. Ensuite, nous avons remplacé les couches entièrement connectées par un espace latent. D'autre part, l'architecture transposée du VGG-16 a été utilisée dans la partie décodeur pour reconstruire le masque résultant des images d'entrée. Le processus de reconstruction vise à augmenter la taille de la sortie d’encodeur pour reconstruire les données d'entrée originales par des opérations de sur-échantillonnage et de convolution, qui sont appelées architecture VGG-16 transposée. Enfin, nous entraînons uniquement l'espace latent et la partie décodeur avec le jeu de données CDnet2014 [74], alors qu'il n'y a pas de processus d'entraînement sur la partie pré-entraînée d’encodeur.

Pour la première approche, nous gelons les 14 premières couches du VGG16 (partie encodeur), puis nous exécutons l'apprentissage pour les couches restantes de l'espace latent et toute la partie décodeur (VGG16 transposé). La couche de Dropout [19] est appliquée après chaque couche de convolution de la partie décodeur est fixée à un taux d'apprentissage de 0,2.

Dans la deuxième approche d’architecture à multi-profondeur, des couches de convolution et de sou-échantillonnage (pooling) sont empilées pour construire la partie encodeur, tandis que des couches de convolution et de sur-échantillonnage (up-sampling) sont utilisées dans la partie décodeur pour sur-échantillonner les images dans l'espace latent. Les couches cachées sont en multi-profondeur. Pour cette approche, nous avons entraîné le modèle entier avec le jeu de données CDnet2014.

Nous avons développé cinq modèles dans la deuxième approche, en commençant par quatre couches cachées et en passant à dix-huit couches cachées. La couche de Dropout [19] est appliquée après chaque couche de convolution est fixée à un taux d'apprentissage de 0,2 pour généraliser le modèle.

Les **figures28 et 29** montrent les architectures d'auto-encodeur proposées dans cette étude, qui sont basées sur les architectures d'apprentissage par transfert et multi-profondeur.

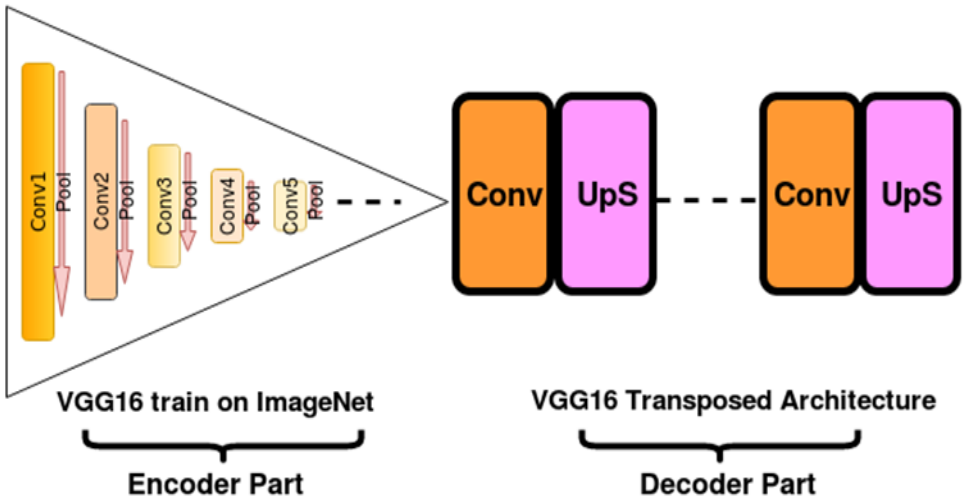

Figure 28 : La structure des couches de l'approche d'auto-encodeur basé sur l'apprentissage par transfert.

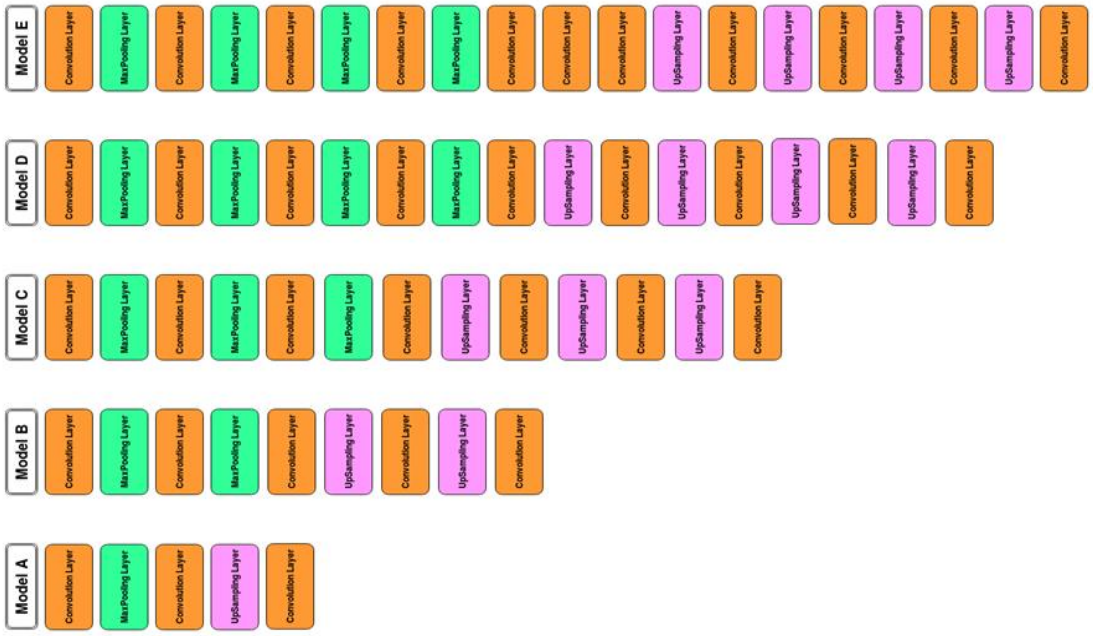

Figure 29 : La structure des couches de l'approche d'auto-encodeur basé sur l'apprentissage a multi-profondeur.

3.3.1.4. Processus des données d'apprentissage

Dans cette partie, nous présentons le processus de sélection de données de nos stratégies pour le choix de la base d'apprentissage, nous avons choisi manuellement vingt-cinq images pour chaque vidéo de CDnet qui contiennent des objets de premier plan importants pour aider notre modèle à apprendre et à segmenter le premier plan avec précision. Cependant, le modèle a été formé à travers une variété de stratégies de formation de données, dont :

- **La stratégie PCA** :

Cette stratégie consiste à générer quatre différentes projections des composantes principales (transformation géométrique) en utilisant la technique ACP à un taux de 80 %, 60 %, 40 % et 20 %, pour éliminer les variables non informatives de chaque image d'entraînement sélectionnée.

- **La stratégie Data augmentation traditionnelle DA** :

La stratégie DA vise à élargir les données en générant quatre images de transformation morphologique pour chaque image de formation sélectionnée, y compris la translation, le retournement, le zoom et la rotation.

- **La stratégie PCA+DA :**

Cette stratégie consiste à fusionner les stratégies précédentes (ACP et DA) pour construire plus d'images de données pour chaque vidéo.

La **figure30** montre les résultats des différentes techniques et stratégies de transformation. Le **tableau2** fournit la description et la quantité d'échantillons utilisés pour le processus d'apprentissage de toutes les stratégies.

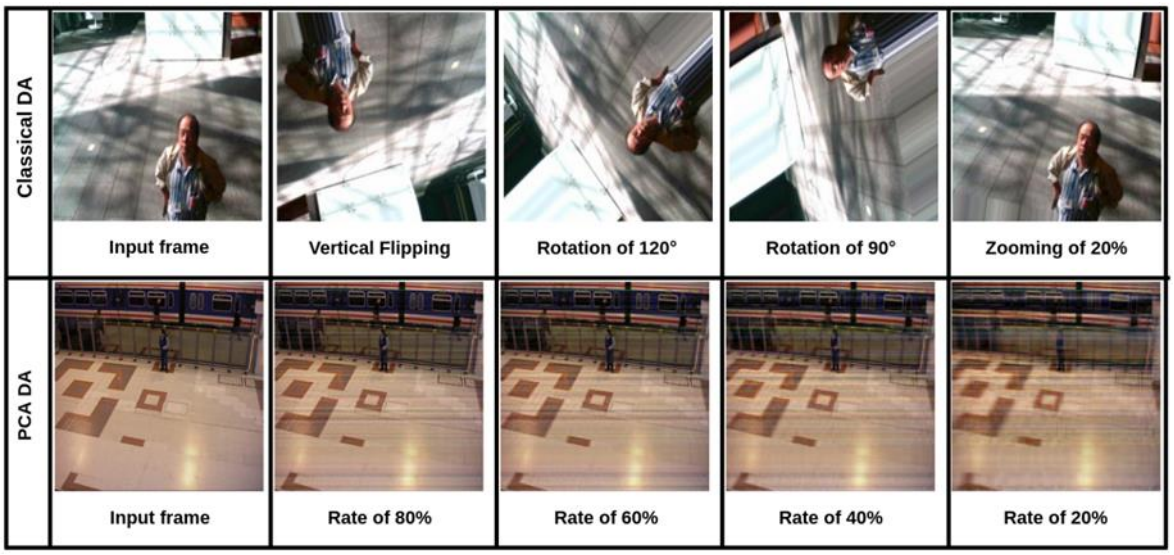

Figure 30 : Exemples des différentes transformations de chaque stratégie.

Tableau 2 : Description et nombre d'échantillons d'entraînement utilisés pour chaque stratégie.

Strategies	Description	Nombre d'échantillons utilisés pour entraîner le modèle
DA	25(frames) x 53(all video) x 5	6625
PCA	25(frames) x 53(all video) x 5	6625
PCA+DA	25(frames) x 53(all video) x 9	11925

3.3.2. Partie 2 : Approche de la segmentation des routes à partir d'images aériennes (cartographie routière)

3.3.2.1. Dataset Teselas

Nous avons entraîné et testé notre modèle sur le jeu de données Teselas [80] qui contient des images aériennes réelles de routes de transport et des routes secondaires, collectées dans des régions espagnoles et capturées dans des scénarios difficiles. Le jeu de données contient 2 catégories (catégorie 1-pas de route, catégorie 2-route existante) comme le montre la **figure31**, environ 9000 images étiquetées (format png) de dimension 256x256.

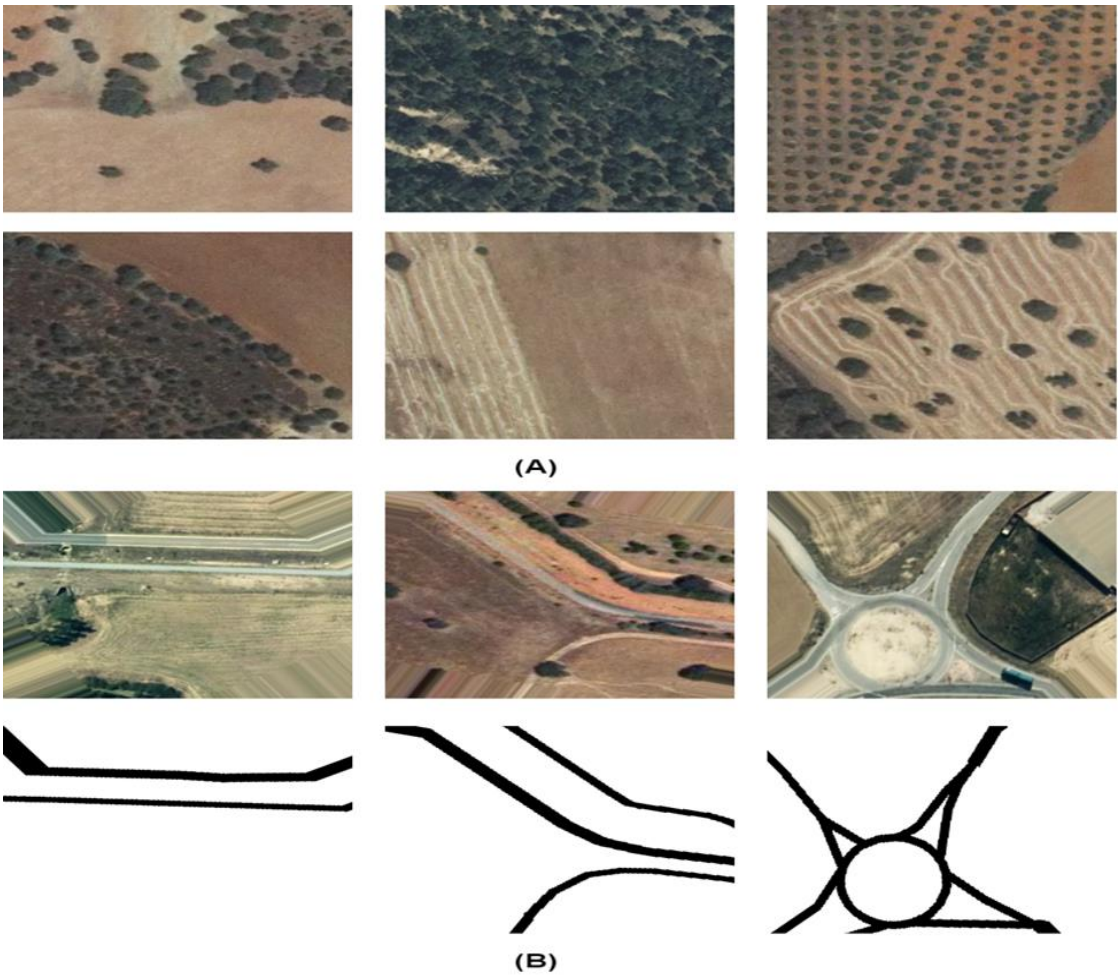

Figure 31 : Exemples d'images qui ne contiennent pas des routes (A) et d'images contenant des routes avec la cartographie associée (B).

3.3.2.2. Augmentation des données à base des transformations par filtres

De nombreuses techniques courantes de traitement d'image telles que l'égalisation d'histogramme, les augmentations et le renforcement de luminosité, le brouillage et les filtres sont des techniques très répandues.

La transformation par filtre est une technique utilisée pour améliorer ou modifier les caractéristiques d'une image. Ce sont des techniques courantes utilisées dans le traitement de l'image et la vision par ordinateur, elles peuvent être appliquées à une image en utilisant différents types de filtres tels que les filtres de Sobel, Canny et Laplacien en glissant une matrice n×m sur l'ensemble de l'image.

- La détection des contours (Contour detection) :

Une technique utilisée pour identifier les limites des objets ou des régions dans une image. Elle peut être utilisée pour mettre en évidence les bords et les limites d'une image, ce qui facilite l'identification et l'isolement d'objets ou de régions spécifiques.

- L'amélioration des détails (Detail enhance) :

Une technique utilisée pour augmenter le niveau de détail d'une image. Elle peut être utilisée pour accentuer ou clarifier une image, ce qui permet de mieux voir les petits détails ou les textures.

- L'amélioration des contours (Edge enhance) :

Une technique utilisée pour augmenter la visibilité des contours dans une image. Elle peut être utilisée pour rendre une image plus nette ou plus claire, afin de mieux voir les limites entre différents objets ou régions.

- Trouver les bords (Find edges) :

Une technique utilisée pour identifier les bords dans une image, elle peut être utilisée pour extraire les bords d'un objet ou d'une région, elle est couramment utilisée dans les tâches de détection et de reconnaissance d'objets.

3.3.2.3. Protocole d’expérience

Dans cette expérience, nous proposons une approche basée sur l'auto-encodeur compressé à trois profondeurs en mode supervisé pour effectuer la segmentation des routes pour toutes les architectures.

Dans les trois architectures de l’approche d’auto-encodeur basé sur VGG16 compressé, nous avons empilé des couches de convolution et de pooling dans les parties d’encodeur pour sous-échantillonner les images d'entrée, et dans la partie du décodeur, nous avons placé des couches de convolution et de upsampling pour sur-échantillonner les images dans l'espace latent. En fait, les couches cachées sont en multi-profondeur. De cette manière, le nombre de couches sera réduit, c'est-à-dire que le modèle d'auto-encodeur VGG16 sera compressé en VGG12 et VGG10. Enfin, nous entraînons l'ensemble des modèles à partir de zéro avec le jeu de données Teselas [80]. Les approches

générales des trois architectures Auto-Encoder compressé sont visualisées dans la **figure32**.

La première architecture du VGG16 Auto-Encoder. Dans les deux autres approches nous avons construit des modèles compressés, nous débutons avec 12 couches cachées puis diminuons progressivement à 10 couches cachées.

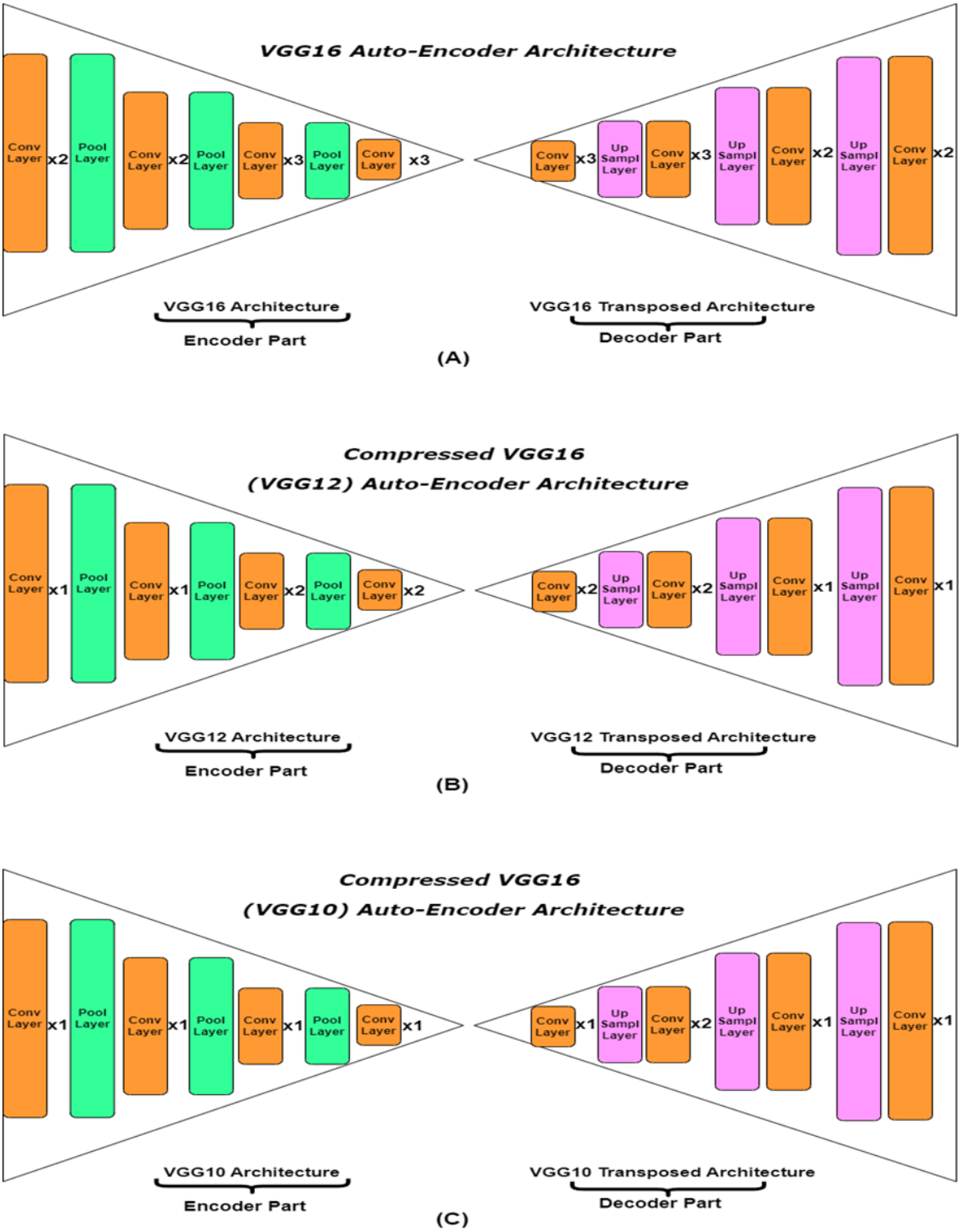

Figure 32 : Les approches générales des trois architectures Auto-Encoder compressées.

3.3.2.4. Processus des données d'apprentissage

Dans cette partie, nous décrivons le processus de sélection de base de données de nos stratégies pour la phase d'apprentissage, nous avons choisi aléatoirement 520 images pour chaque catégories de la base Teselas [80] qui contiennent. Cependant, le modèle a été formé à travers différentes stratégies d'apprentissage de données, dont :

- **La stratégie DA :**

Consiste à augmenter les données en générant trois images de transformation classiques (rotation pour couvrir les angles essentiels) pour chaque image de formation sélectionnée.

- **La stratégie Filtre :**

Consiste à générer trois opérations de transformations par filtre (détection des contours, amélioration des détails, et amélioration EDGE) pour chaque image d'entraînement sélectionnée afin de faciliter la visualisation des données et d'accentuer les variables informatives.

La **figure33** et le **tableau3** montre les résultats des différentes techniques et stratégies de transformation pour l'apprentissage. Ainsi, la description et la quantité d'échantillons utilisés.

Tableau 3 : Description et nombre d'échantillons d'entraînement utilisés pour chaque stratégie.

Stratégies	**Description**	**Nombre d'échantillons utilisés pour entraîner le modèle**
Peu de données	520 images de chaque (catégorie)x2	1040
DA	Peu de données+(3 transformation DA)	4160
Filter + DA	Peu de données+(3 transformation Filter)	4160

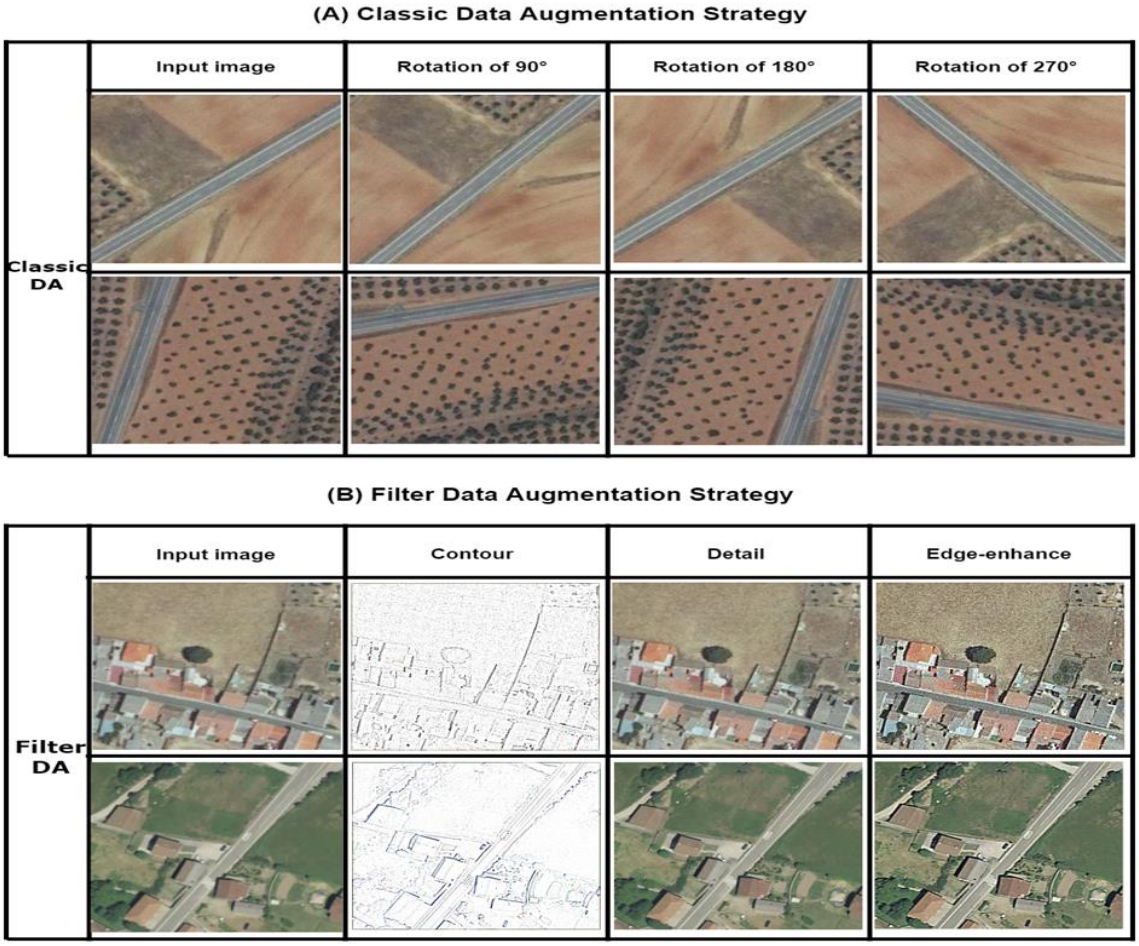

Figure 33 : Transformation des différentes stratégies d'augmentation des données utilisées.

3.4. Test

3.4.1. Métriques des testes

Plusieurs métriques sont adoptées pour évaluer les modèles basés sur l'apprentissage profond [74], notamment la spécificité, la précision, la F-Mesure, le taux de faux positifs, le taux de faux négatifs et le pourcentage de classifications erronées, en utilisant les quatre paramètres de la matrice de confusion. Ces métriques peuvent être mesurées selon les équations suivantes :

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{4}$$

$$\text{False Positive Rate} = \frac{FP}{TN + FP} \tag{5}$$

$$\text{False Negative Rate} = \frac{FN}{TP + FN} \tag{6}$$

$$\text{Percentage of the Wrong Classifications} = 100 \times \frac{FN + FP}{TP + FN + FP + TN} \tag{7}$$

$$\text{F} - \text{Measure} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{8}$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (9)$$

Où TP, FP, FN et TN désignent respectivement un vrai positif, un faux positif, un faux négatif et un vrai négatif.

- **Vrai positif** : c'est le nombre de pixels segmentés d'objets mobiles correspondant aux pixels segmentés dans la vérité terrain.
- **Faux positif** : c'est le nombre de pixels d'objets mobiles segmentés correspondant à des pixels non segmentés dans la vérité terrain.
- **Vrai négatif** : c'est le nombre de pixels non segmentés d'objets mobiles correspondant aux pixels non segmentés dans la vérité terrain.
- **Faux négatif** : il s'agit du nombre de pixels non segmentés d'objets mobiles correspondant à des pixels segmentés dans la vérité terrain.

3.5. Implémentation (logiciels et librairies)

3.5.1. Python

Python [81] est un langage de programmation interprété de haut niveau, connu pour sa lisibilité, sa simplicité et sa polyvalence. Il est largement utilisé dans une variété d'applications telles que le développement web, le calcul scientifique, l'analyse de données et l'intelligence artificielle. La syntaxe de Python est facile à apprendre et à comprendre, ce qui en fait un excellent langage pour les débutants. Il dispose également d'une communauté importante et active, ce qui signifie qu'il existe de nombreuses bibliothèques, modules et frameworks disponibles pour aider les développeurs dans diverses tâches.

Python est souvent utilisé dans la science des données et l'apprentissage automatique, plusieurs bibliothèques telles que NumPy, pandas et Scikit-learn sont largement utilisées pour la manipulation et l'analyse des données [81]. En outre, il existe des bibliothèques telles que TensorFlow, Keras et Pytorch qui sont utilisées pour construire et former des modèles d'apprentissage automatique. En termes de traitement d'images et de vision par ordinateur, Python fournit plusieurs bibliothèques telles que ; OpenCV, PIL et matplotlib qui sont largement utilisées pour les tâches de traitement d'images et de vision par ordinateur.

Python peut être exécuté sur de nombreuses plateformes, notamment Windows, macOS, Linux, Unix et même sur des appareils mobiles tels que Android. Il peut également être exécuté sur des serveurs et dans le cloud, ce qui en fait un choix populaire pour les déploiements d'apprentissage automatique en production. De plus, il existe de

nombreuses distributions Python pré-compilées pour les différentes platesformes, ainsi que des outils en ligne de commande pour l'installation et la gestion des paquets. Cela facilite grandement la mise en place de l'environnement de développement Python sur n'importe quelle plateforme.

3.5.2. Librairie

Plusieurs bibliothèques ont été développées afin de simplifier le processus de mise en œuvre des algorithmes d'apprentissage automatique. Elles fournissent des outils qui exécutent automatiquement un grand nombre des tâches nécessaires à la mise en œuvre, entre autres, d'un réseau neuronal fonctionnel. Bien que les bibliothèques soient basées sur les mêmes modèles théoriques d'apprentissage automatique, elles diffèrent dans leur approche de leur mise en œuvre. Vous trouverez ci-dessous de brefs résumés de certaines des bibliothèques les plus couramment utilisées.

- **Tensorflow**

Développé à l'origine par Google dans le cadre du projet Google Brain. Tensorflow [82] est une bibliothèque Python et se distingue en étant la seule des grandes bibliothèques développée de A à Z par une grande entreprise, alors que les autres ont leur origine dans la communauté des chercheurs. Tensorflow dispose de quelques outils de qualité de vie intégrés, tels que TensorBoard, qui permet à l'utilisateur de produire facilement des graphiques visualisant des éléments tels que le taux d'apprentissage, les poids des modèles, les fonctions de perte, etc. Tensorflow est également la seule bibliothèque capable de distribuer la charge de travail non seulement sur les GPU d'un même appareil, mais aussi sur plusieurs appareils connectés, ce qui peut constituer un avantage majeur en termes de calcul.

- **Keras**

Keras [83] est une bibliothèque Python de haut niveau construite sur la base de TensorFlow et destinée à simplifier le processus de développement et d'apprentissage automatique. En outre, elle est conçue pour minimiser le nombre d'actions requises pour les cas d'utilisation courants et elle fournit des messages d'erreur clairs et exploitables. Il dispose également d'une documentation complète et de guides du développeur, pour permettre un prototypage rapide et facile des algorithmes d'apprentissage automatique. Nous pouvons aussi exporter les modèles Keras en JavaScript pour les exécuter directement dans le navigateur, en TF Lite pour les exécuter sur iOS, Android et les appareils embarqués. Il est également facile de servir les modèles Keras via une API Web.

Keras est le framework d'apprentissage profond le plus utilisé par les cinq premières équipes gagnantes du concours Kaggle. Keras est la solution d'apprentissage profond de choix pour de nombreux cours universitaires. Elle est largement recommandée comme l'un des meilleurs moyens d'apprendre l'apprentissage profond.

3.6. Conclusion

Dans ce chapitre, nous avons présenté le matériel utilisé pour construire deux approches de segmentation à base des auto-encodeurs. Le premier est un système de segmentation d'objets en mouvement robustes à partir de la vidéosurveillance. Le second est un système de segmentation des routes compressé à partir des images aériennes.

Nous fournissons également une description détaillée de chacun des principaux éléments constitutifs introduits dans notre méthodologie. Ainsi, nous présentons le protocole d'expérimentation des approches proposées. Par contre, les détails des approches proposées et les différentes stratégies d'augmentation des données seront présentés.

Chapitre 4 : Analyse et discussion des résultats

4. Résultats et analyse

4.1. Introduction

Ce chapitre présente les résultats et l'analyse de chaque approche et la manière d'effectuer la segmentation pour les deux domaines d'applications. Les sections de ce chapitre présentent les résultats de la phase d'apprentissage et du test de tous modèles développés. Ainsi, discutant les performances des nouvelles stratégies d'augmentation des données sur chacune des approches. De même, les résultats qualitatifs et quantitatifs de chaque modèle sont présentés, des exemples de segmentation seront également présentés à des fins de comparaison entre approches proposées et avec les populaires méthodes.

4.2. Apprentissage et plateforme de traitement

Tous les modèles d'EA ont été implémentés (formé et testé) en utilisant la bibliothèque Keras [83] programmée en Python. Les modèles **de la segmentation des objets en mouvement** et **de la segmentation des routes à partir d'images aériennes** ont été entraînés respectivement pendant 100 et 150 itérations. Après avoir mené plusieurs essais et effectué une comparaison approfondie entre différents optimiseurs et les fonctions du cout, nous avons opté pour l'optimiseur RMSprop dans la phase d'apprentissage en raison de ses performances supérieures dans notre cas. Cet optimiseur a démontré une convergence plus rapide et une meilleure stabilité. Nous avons également choisi d'utiliser la fonction de coût BCE (Binary Cross-Entropy) pour mesurer la performance de notre modèle. Cette fonction est particulièrement adaptée aux problèmes de classification binaire, tels que la segmentation, qui peut être mesurée à l'aide de la formule suivante :

$$BCE\left(Y, \widetilde{\mathrm{X}}\right) = -\left(Y * \log\left(\tilde{X}\right) + (1 - Y) * \log\left(1 - \widetilde{\mathrm{X}}\right)\right) \qquad (10)$$

Où $\tilde{\mathrm{X}}$ et Y désignent respectivement l'étiquette de vérité-terrain et l'étiquette prédite par les modèles.

L'ensemble du réseau a été entraîné avec 80% des images provenant des données d'apprentissage, et 20% des images pour valider l'entraînement des réseaux. Nous faisons également l'évaluation des modèles avec 50% de l'ensemble de la base de données.

Dans nos expériences, le processus d'apprentissage et de validation est effectué en ligne sur la plateforme Google Colaboratory à travers un GPU Tesla K80 [84, 85], pour **la segmentation des objets en mouvement**, et l'utilisation d'une station de travail

équipée d'un GPU Nvidia GTX 1080TI pour **la segmentation des routes à partir d'images aériennes.**

- **Google Colaboratory Vs GPU Nvidia GTX 1080TI**

La GeForce GTX 1080 Ti est une carte graphique haut de gamme qui peut être installée dans un ordinateur de bureau ou un ordinateur portable conçue pour les applications de calcul intensives, telles que la simulation, l'analyse de données et notamment pour les tâches liées à l'apprentissage automatique et à l'intelligence artificielle. Elle dispose de 3584 cœurs CUDA et de 11 Go de mémoire GDDR5X, ce qui la rend idéale pour les tâches nécessitant beaucoup de puissance de calcul.

Google Colaboratory [84, 85], est un service de calcul en nuage gratuit qui vous permet d'exécuter des notebooks Jupyter en utilisant des GPU de haute performance. Il est basé sur le cloud computing et vous permet de partager et de collaborer facilement sur des projets de d'apprentissage machine et de data science. Il est souvent utilisé pour l'apprentissage automatique et la recherche dans les domaines de la science des données et de l'intelligence artificielle. Il propose des ressources de calcul limitées (comme un processeur standard, un GPU Tesla K80 ou T4, et une certaine quantité de mémoire vive), mais permet aux utilisateurs d'accéder à ces ressources à partir de n'importe où avec une connexion internet.

4.3. Analyse et discussion des résultats

4.3.1. Partie 1 : Approche de la segmentation des objets en mouvement (Vidéosurveillance)

4.3.1.1. Test et évaluation des approches proposées

Dans cette sous-section, nous analysons les stratégies de formation à l'aide des résultats obtenus et présentons leurs influences sur le jeu de données adopté.

4.3.1.2. Test de l'approche basé sur l'apprentissage par transfert

Les résultats obtenus en utilisant l'approche d'apprentissage par transfert sont présentés dans le **tableau4** et la **figure34**. Les courbes de la fonction de perte et de précision entre les données d'apprentissage et de validation ont été tracées durant la phase d'apprentissage.

Tableau 4 : Les résultats des tests obtenus par les stratégies PCA, DA et PCA+DA.

Stratégies	Sp	FPR	FNR	PWC	FM	Pr
PCA	0.993	0.0058	0.3458	1.383	0.692	0.748
DA	0.997	0.0029	**0.3338**	1.073	0.742	0.853
PCA+DA	**0.997**	**0.0020**	0.3571	**1.027**	**0.811**	**0.873**

Les résultats montrent clairement que la stratégie PCA+DA a surpassé les autres stratégies en fournissant de meilleures performances. La **figure34** montre les graphiques des précisions et des pertes de formation et de validation.

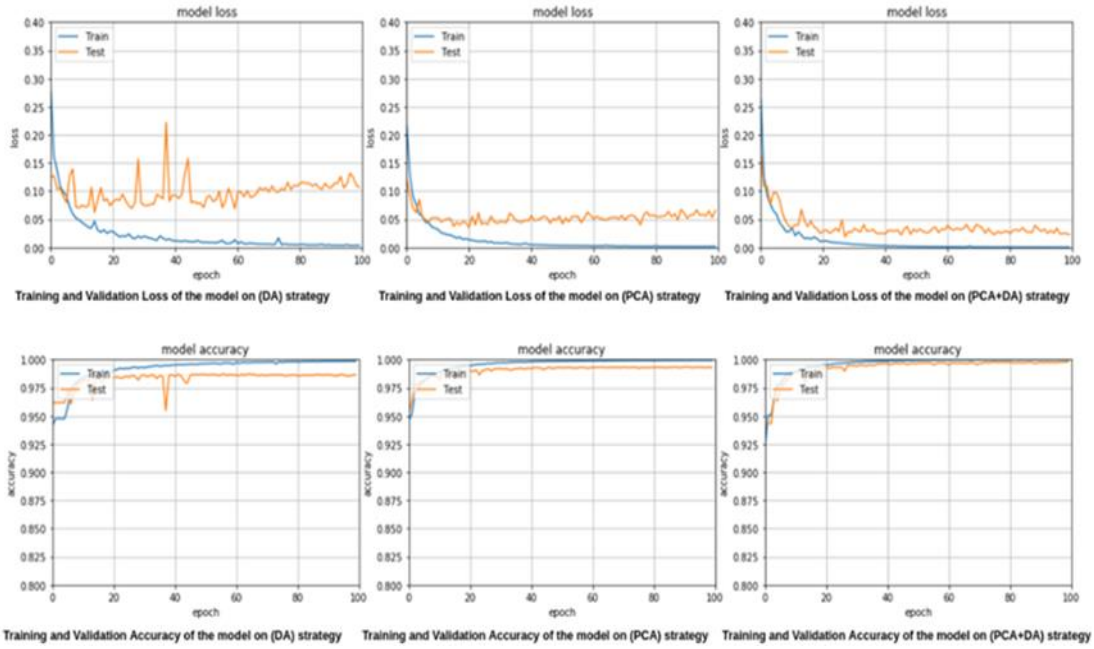

Figure 34 : Graphiques des pertes et des précisions d'apprentissage et de validation des stratégies utilisées.

D'après la courbe de perte de la stratégie classique DA, nous pouvons clairement voir l'écart entre la perte de validation et la perte de formation, ce qui indique que le modèle est sur-ajusté. Cette sur-adaptation peut être due au manque d'échantillons d'entraînement.

De plus, comme le montre la **figure34**, l'écart entre la perte de formation et de test est réduit dans le cas de l'utilisation de la stratégie PCA, et il y a une légère différence entre les valeurs de perte de formation et de validation. Par conséquent, nous pouvons voir que l'effet de sur-ajustement a été réduit en raison de la présence de la transformation PCA avec des caractéristiques plus précises qui aident et facilitent le système pour la tâche d'apprentissage.

La stratégie PCA+DA proposée fournit les meilleures et les plus faibles pertes d'apprentissage et de test par rapport aux deux autres stratégies. Ainsi, le modèle évite le sur-ajustement grâce au nombre d'échantillons ajoutés à l'ensemble de données d'apprentissage.

Pour les courbes de précision de la **figure34**, on observe qu'avec chaque augmentation des échantillons par PCA dans les données d'apprentissage, la précision augmente et les deux courbes convergent encore plus.

4.3.1.3. Test de l'approche basé sur l'apprentissage multi-profondeur

L'étude actuelle vise à évaluer le modèle d'auto-encodeur dans différentes profondeurs, nous avons analysé l'influence et les résultats du modèle d'apprentissage multi-profondeurs. Pour cela, les courbes de la fonction de perte et de précision entre les données d'apprentissage et de validation ont été tracées.

Nous avons formé les cinq modèles uniquement avec la stratégie DA. Les résultats sont présentés dans la **figure35** et la **figure36**.

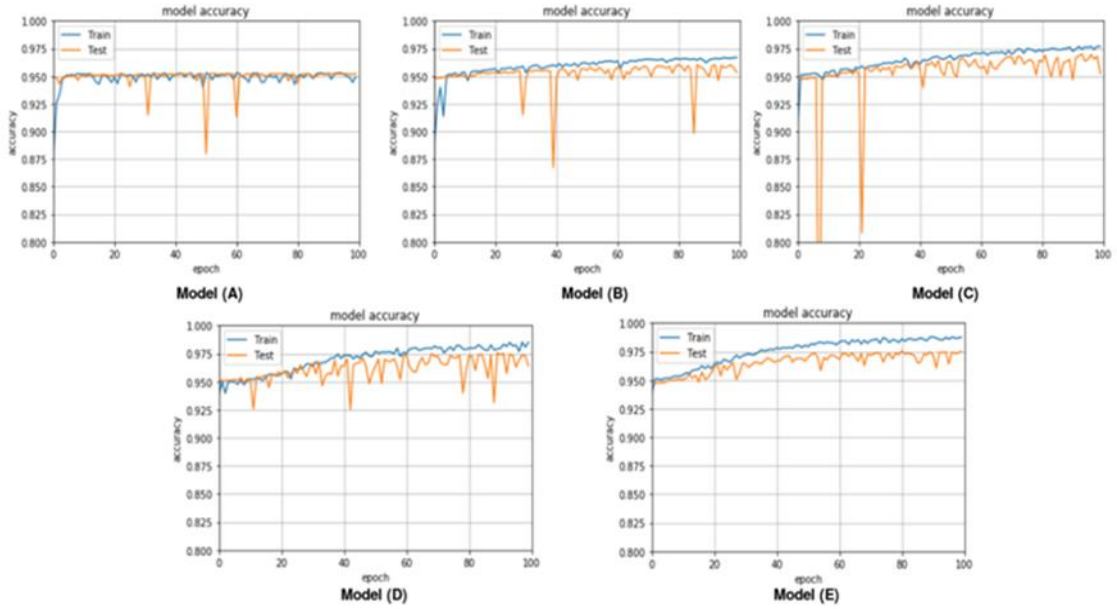

Figure 35 : Graphiques de la fonction de précision entre les données d'apprentissage et de validation des stratégies utilisées.

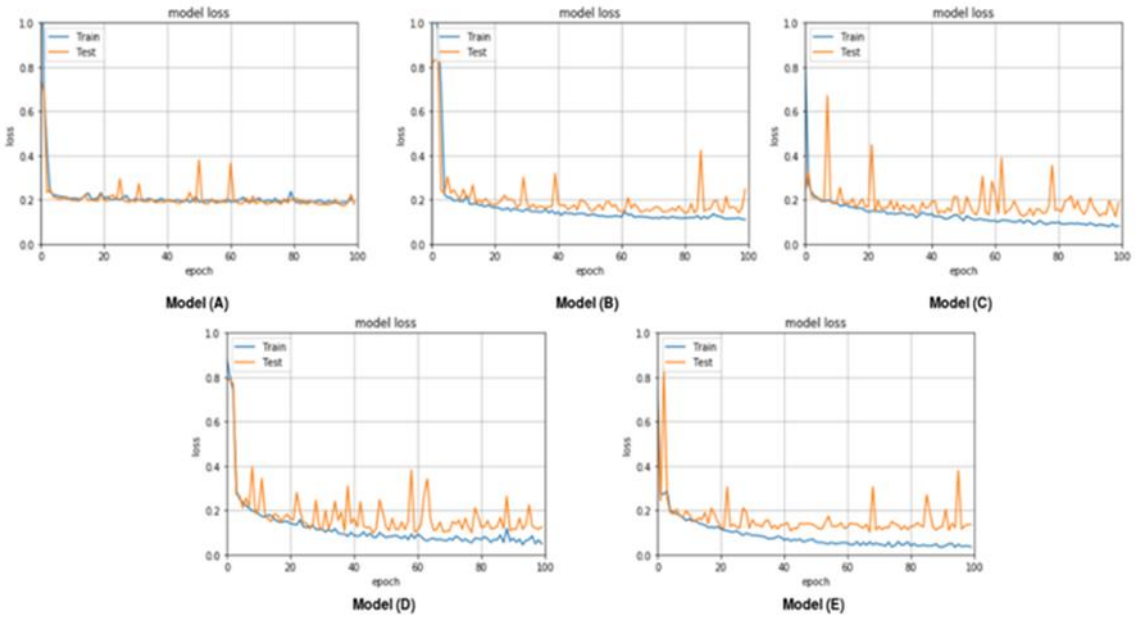

Figure 36 : Graphiques de la fonction de perte entre les données d'apprentissage et de validation des stratégies utilisées.

Nous pouvons voir que pour les modèles (A), (B), et (C), les pertes de validation et d'apprentissage sont très proches mais ont une erreur plus élevée. La précision de validation et de formation a une faible valeur de précision.

D'après la **figure36**, nous pouvons voir que les modèles (D) et (E) commencent à se sur-ajuster, et que l'erreur commence à diminuer. La précision de validation et la précision de formation commencent à augmenter.

En général, lorsque nous augmentons le nombre de couches, il peut en résulter une meilleure précision et une perte minimale, l'augmentation de la profondeur signifie l'augmentation de la capacité du modèle (peut apprendre des représentations complexes), mais avec peu de données d'entraînement, cela peut engendrer un risque élevé de sur-ajustement.

Nous constatons que le modèle (E) fournit de meilleurs résultats que les autres méthodes en termes de précision et d'erreur. Mais le modèle (E) commence à se sur-ajuster à partir de l'itération 50. À cette fin, nous avons choisi le modèle (E) comme modèle de base pour l'améliorer en ajoutant plus de données d'entraînement.

Les résultats de la version améliorée sont présentés à la **figure37**.

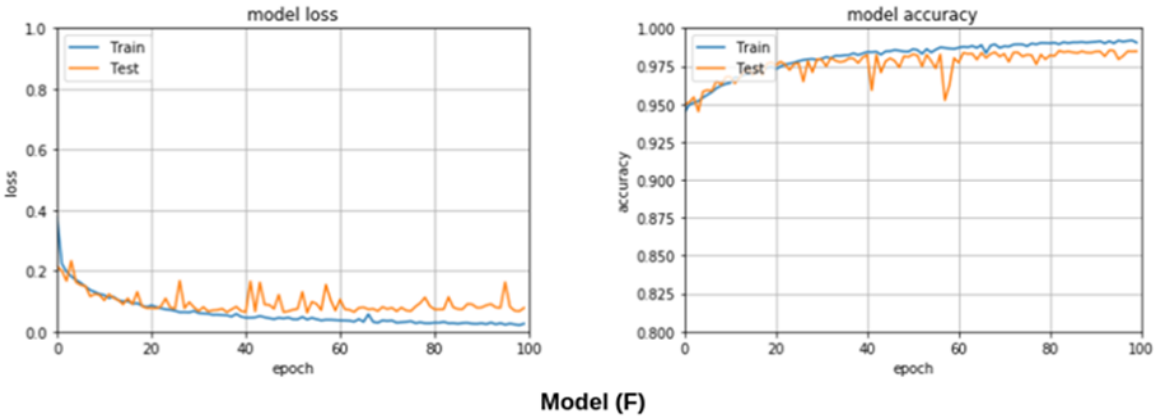

Figure 37 : Graphiques des pertes et des précisions d'apprentissage et de validation du modèle F.

Le modèle (F) est le même modèle (E) mais entraîné avec plus de données. En plus des données générées à l'aide des techniques classiques d'AD, les données utilisées pour entraîner le modèle (F) sont générées à l'aide de la stratégie PCA. Par conséquent, le modèle est entraîné à l'aide des deux stratégies DA et PCA. Comme le montre la **figure37**, nous pouvons remarquer que la perte de validation et la précision du modèle (F) sont améliorées, ce qui le rend plus robuste à l'over-fitting.

Notre objectif principal est d'atteindre des performances élevées en termes de précision et de perte en utilisant des architectures CNN profondes avec de petits ensembles de données tout en évitant le sur-apprentissage.

4.3.1.4. Comparaison avec les algorithmes de référence

La quantité de données d'apprentissage utilisée pour former les modèles est différente d'une approche à une autre. Nous avons donc comparé nos modèles avec des approches qui utilisent le même principe d'apprentissage avec peu de données d'entraînement.

Nous avons comparé notre modèle avec populaire modèle développé par Babaee et al[14]. Ce modèle est entraîné par 5% (100 images) des images de chaque séquence vidéo. De plus, BSUV-Net [86] et Fast BSUV-Net 2.0 [87] ont proposé des algorithmes de soustraction d'arrière-plan pour des vidéos basés sur un réseau de neurones entièrement convolutif. Ils ont introduit une technique d'augmentation des données spatio-temporelles pour surmonter le problème du manque d'échantillons d'entraînement. Notre approche est également comparée à d'autres algorithmes traditionnels, notamment SuBSENSE[33], IUTIS-5 [88] et PAWCS [89], où les modèles sont entraînés à partir de petites images de chaque séquence vidéo de l'ensemble de données.

On fait la comparaison avec nos meilleurs modèles pour chaque approche dAuto-Encoder Multi-Profondeur (MD AE) et d'apprentissage par transfert (TL AE) avec la stratégie PCA+DA.

Les résultats de comparaison obtenus sont montrés dans le **tableau5**.

Tableau 5 : Une comparaison des résultats de nos approches avec des algorithmes de référence.

Method	Sp	FPR	FNR	PWC	FM	Pr
TL AE	**0.997**	**0.0020**	0.3571	1.027	**0.811**	**0.873**
Fast BSUV-Net 2.0 [87]	0.995	0.0044	0.1819	**0.905**	0.803	0.842
BSUV-Net [86]	0.995	0.0054	**0.1797**	1.140	0.787	0.811
CNN [14]	0.990	0.0095	0.2455	1.992	0.755	0.833
MD AE	0.990	0.0071	0.4467	1.836	0.747	0.809
IUTIS-5 [88]	0.995	0.0052	0.2151	1.198	0.772	0.808
PAWCS [89]	0.995	0.0051	0.2280	1.199	0.740	0.786
SuBSENSE [33]	0.990	0.0096	0.1876	1.678	0.741	0.751

Pour une évaluation plus approfondie, nous comparons nos approches avec les modèles de références. La **figure38** présente les résultats de la comparaison qualitative. Trois images de différentes séquences vidéo provenant du jeu de données CDnet 2014 sont sélectionnées comme exemples démonstratifs.

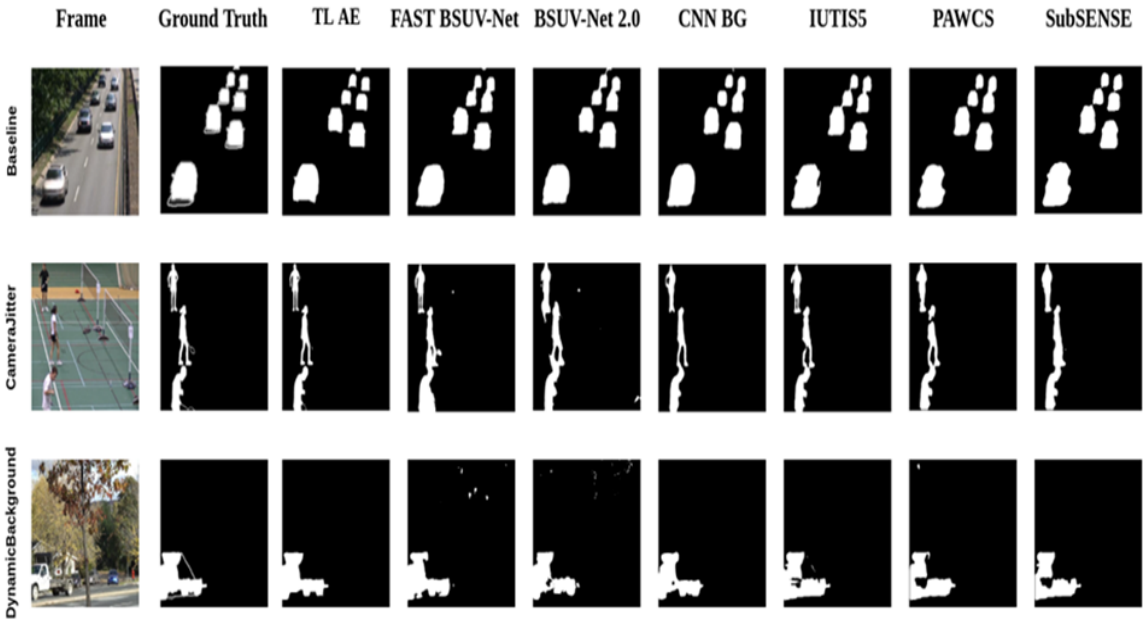

Figure 38 : Comparaison visuelle entre les masques de segmentation générés par le modèle proposé et les modèles de référence.

La première et la deuxième colonne de la **figure38** montrent respectivement les images d'entrée et la vérité terrain. La troisième colonne présente notre modèle d'auto-encodeur profond basé sur l'apprentissage par transfert formé par la stratégie (PCA+DA). Les autres colonnes de la **figure38** représentent les résultats des modèles de référence.

Les résultats du **tableau5** et de la **figure38** montrent que notre approche basée sur la stratégie de formation PCA+DA fournit de meilleurs résultats et une segmentation sélective par rapport aux autres modèles.

4.3.1.5. Discussion

Le modèle a été entraîné à l'aide d'une technique d'apprentissage par transfert dans la première expérience, tandis que dans la deuxième expérience, nous avons exploré plus en profondeur l'architecture de l'auto-encodeur pour entraîner un modèle en multi-profondeur.

Comme il est montré dans cette section, la stratégie PCA+DA apporte des améliorations significatives sur toutes les métriques objectives par rapport aux autres stratégies adoptées pour l'apprentissage par transfert et les approches multi-profondeurs. Par conséquent, nous avons utilisé les modèles basés sur ces stratégies pour les comparer avec le populaire modèle CNN [14] et les méthodes les plus connues. D'après les résultats obtenus dans le **tableau5** et la **figure38**, nous pouvons constater que les résultats de notre modèle basé sur (TL AE) obtiennent de meilleures performances que celles obtenues par les algorithmes de référence.

En comparant nos deux approches développées, les performances obtenues par l'approche d'apprentissage par transfert sont dues aux connaissances transférées aux premières couches de la partie encodeur entraînée avec le modèle VGG16 (jeu de données ImageNet). En outre, les informations pertinentes fournies par les principales composantes de la transformation ACP sont ordonnées en fonction de leur capacité de représentation, ce qui rend les caractéristiques du jeu de données à être statistiquement indépendantes et donnent une meilleure représentation.

L'ajout d'images transformées par ACP à plusieurs taux a pour objectif d'augmenter la taille du jeu de données d'apprentissage tout en préservant les informations essentielles. En outre, fournir des données adaptées et significatives pour compenser le manque d’information dans les couches de la partie décodeur du réseau. Par conséquent, nous prouvons l'efficacité de la nouvelle stratégie d'augmentation des données proposée. Cette stratégie est basée sur la préservation des informations nécessaires, ce qui prouve

sa capacité à éviter l'impact de l'over-fitting grâce à une meilleure représentation des données d'apprentissage.

4.3.2. Partie 2 : Approche de la segmentation des routes à partir d'images aériennes (cartographie routière)

4.3.2.1. Test et évaluation des approches proposées

Dans cette sous-section, nous analysons l'influence et les résultats des stratégies d'apprentissage pour chaque approche sur le jeu de données adopté.

4.3.2.2. Temps d'apprentissage

Le temps d'apprentissage des stratégies de formation pour chaque approche est présenté dans le **tableau 6**.

Tableau 6 : Temps de d'apprentissage des stratégies pour chaque approche.

Architectures	Nombre de Paramètres	Le temps de d'apprentissage des stratégies en seconde (S)		
		Peu de données	DA traditionnelle	DA filtre
VGG16 AE	18087715	328045	1315534	1312389
VGG12 AE	9590145	218174	860137	883612
VGG10 AE	**5460097**	163077	658538	**635932**

Les résultats montrent clairement que le modèle VGG10 AE formé avec la stratégie d'augmentation des données a base des filtres a un temps de calcul plus court et un petit nombre de paramètres que les autres stratégies.

Pour mieux analyser la performance de nos modèles, nous avons tracé la fonction de précision et de perte entre les données d'apprentissage et de validation pour toutes les architectures formées sur les différentes stratégies dans les **figures39** et **40**.

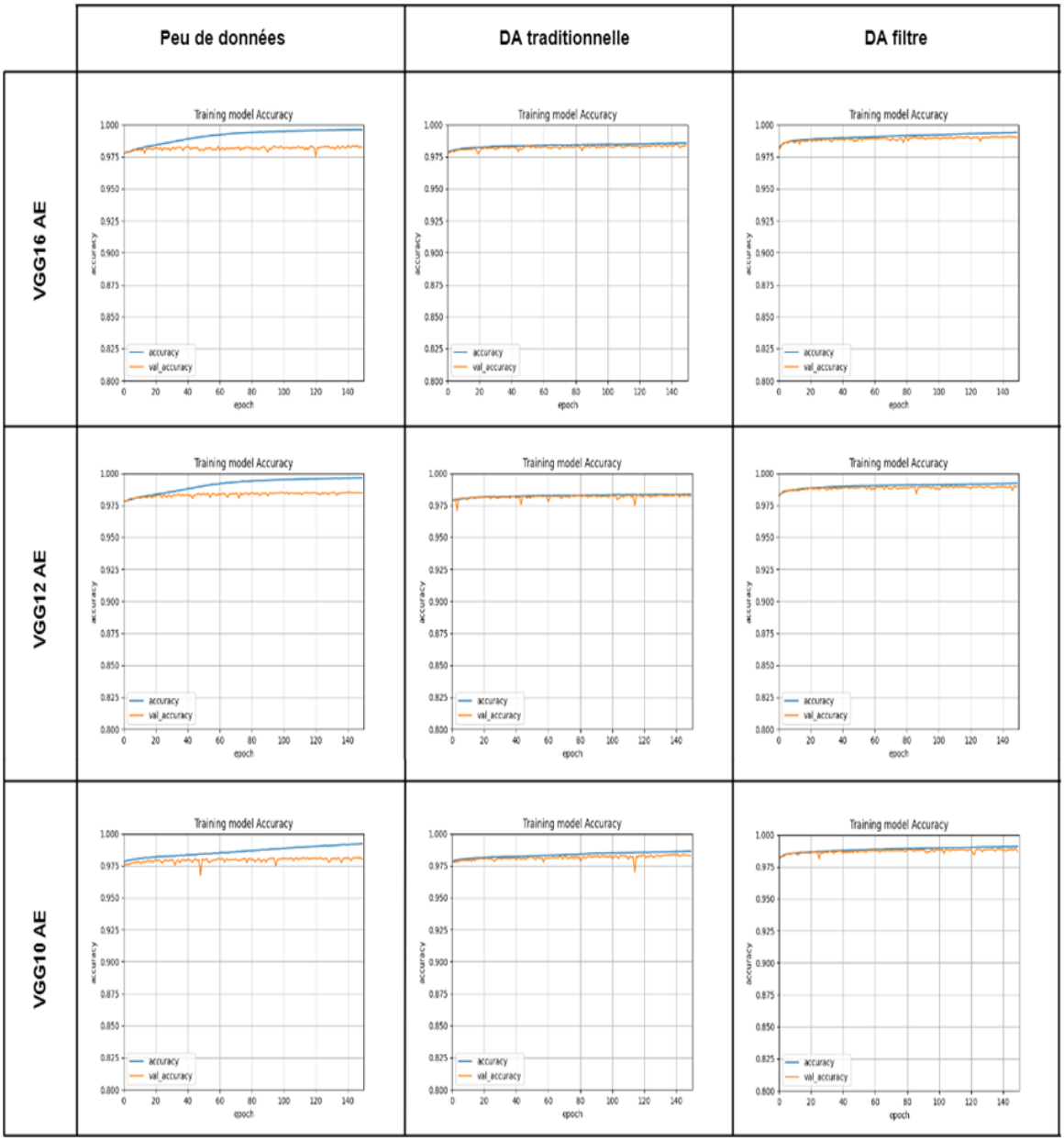

Figure 39 : Les courbes de fonction de précision entre les données d'apprentissage et de validation des trois architectures.

Pour les courbes de la fonction de précision, on observe qu'avec chaque diminution du nombre de couches cachées des modèles compressés, la précision augmente et les deux courbes d'apprentissage et de validation convergent encore plus. Cette amélioration de précision est remarquable, lorsqu'on a ajouté plus de données d'entraînement par la stratégie transformation de filtre pour les trois modèles.

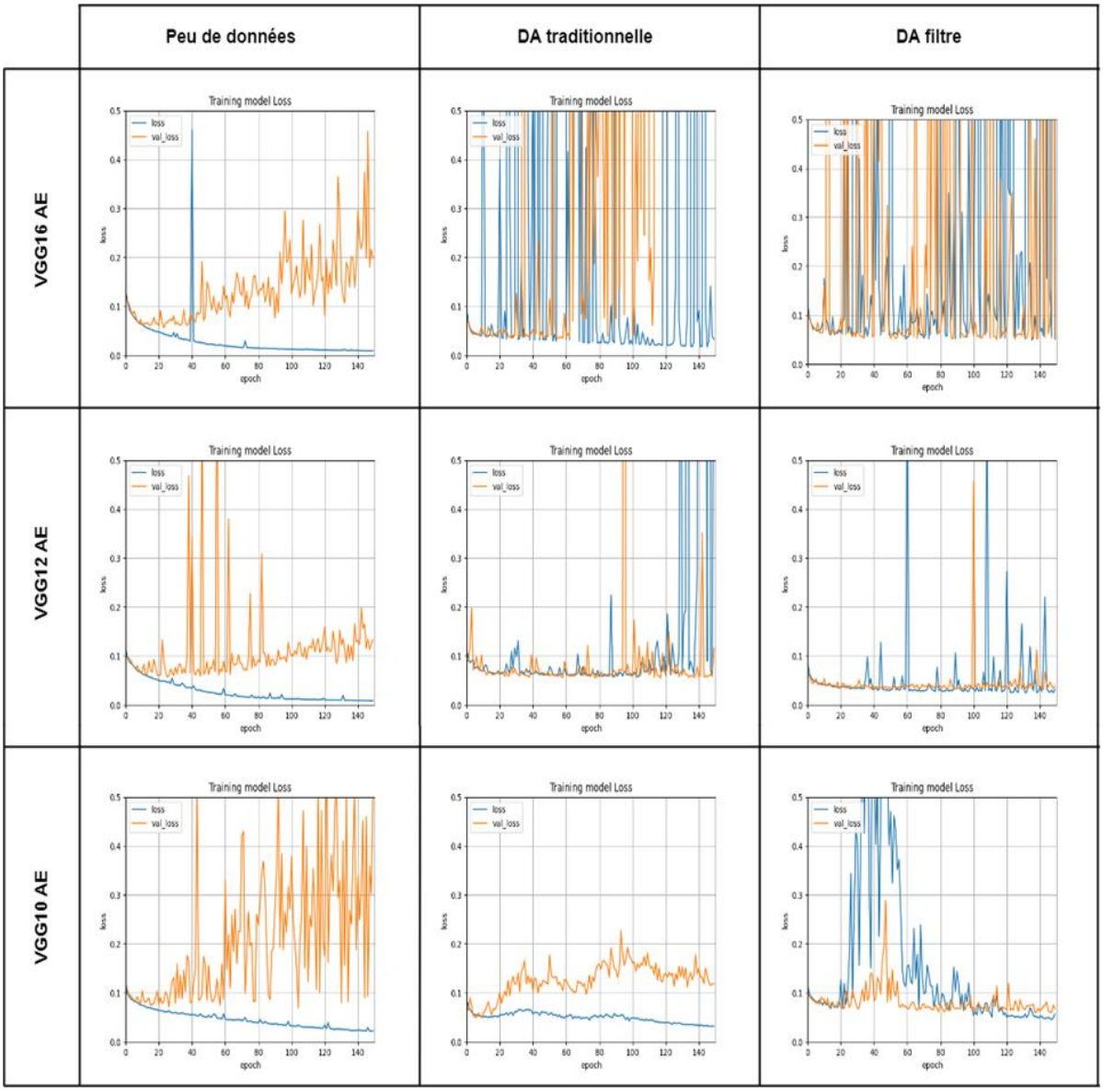

Figure 40 : Les courbes de fonction de perte entre les données d'apprentissage et de validation des trois architectures.

Nous pouvons voir que pour la courbe de fonction de perte sur les données d'apprentissage sans faire aucune augmentation (peu de donnée) pour toutes les trois approches, l'écart de la perte entre les données de validation et de d'apprentissage est important, il signifie que les modèles sont sur-ajustés en raison du manque de données d'apprentissage.

A partir des courbes de fonction de perte sur les données d'apprentissage ajouteés par l'augmentation classique des données de toutes les approches, on peut clairement voir que le modèle commence à être sur-ajusté après 70 itérations.

D'après la courbe de fonction de perte du modèle VGG10 AE formé sur les données d'augmentation classique, il y a rarement des différences entre les courbes

d’apprentissage et de validation, par conséquent, nous pouvons voir que l'effet de sur-ajustement a diminué en raison de la présence de la transformation d'augmentation des données. D'après le tracé de la fonction de perte du VGG10 AE formé à la stratégie d'augmentation des données par filtre, les courbes de validation et d’apprentissage convergent toutes les deux, de sorte que le modèle évite le sur-ajustement grâce aux nouveaux échantillons de transformations par filtres ajoutés à l'ensemble de données d’apprentissage avec des caractéristiques plus précises qui aident et facilitent la tâche d'apprentissage du système, et au nombre réduit de couches.

Nous constatons que le modèle (VGG10) formé sur l'augmentation des données FILTER donne de meilleurs résultats par rapport aux autres en termes de précision et d'erreur. Lorsque nous avons ajouté plus de données d'entraînement par transformation de filtre dans le modèle, nous pouvons remarquer que la perte de validation a diminué alors que la validation de la précision a augmenté, par conséquent, nous avons évité le sur-ajustement. Dans ce cas, le modèle paraît plus robuste.

Notre objectif est de faire un compromis entre les quatre facteurs (diminution du nombre de couches cachées, diminution de l'erreur sans présence de sur-ajustement et augmentation de la précision avec moins de temps d'apprentissage possible).

4.3.2.3. Comparaison des trois modèles compressées

Le t**ableau7** montre une comparaison entre nos résultats du test des approches proposées.

Tableau 7 : une comparaison entre les trois modèles compressées

Méthodes	FM	Pr	Sp
VGG10 AE	**0.841**	**0.913**	**0.976**
VGG12 AE	0.797	0.899	0.971
VGG16 AE	0.785	0.863	0.966

Pour faire une meilleure comparaison visuelle entre nos approches, nous présentons dans la **figure41**, les résultats de la comparaison qualitative, nous avons choisi comme exemple démonstratif deux images du jeu de données Teselas.

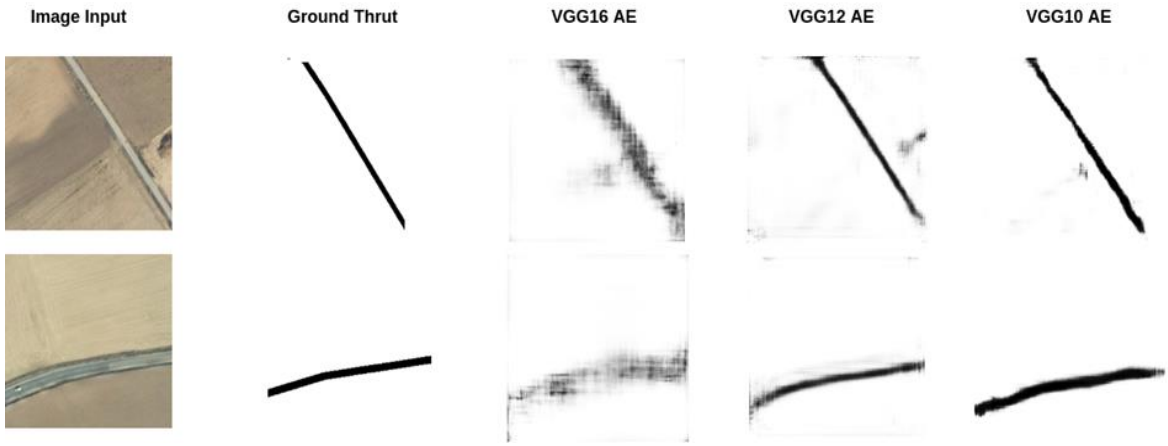

Figure 41 : une comparaison visuelle entre les trois modèles compressées.

La première et la deuxième colonne de la **figure41** montrent respectivement les images d'entrée et les images étiquetées. Les autres colonnes représentent les résultats de nos modèles d'auto-encodeur compressés à 16, 12 et 10 couches basées sur l'apprentissage par la stratégie des transformations des FILTER.

Les résultats du **tableau7** et de la **figure41** montrent que l'approche la plus compressé (VGG10 AE) basée sur la stratégie de formation par filtre a une meilleure performance de segmentation avec un temps d'apprentissage moindre que les autres modèles.

4.3.2.4. Discussion

Les résultats indiquent que le modèle d'apprentissage comprimé VGG10 AE entraîné sur l'augmentation des données FILTER peut apprendre avec succès la généralisation de la segmentation des routes dans le temps le plus court possible avec peu de données d'apprentissage. En outre, la préservation des informations essentielles fournies par les transformations basées sur le filtre (amélioration des contours et des bords) a pour but d'augmenter la taille des données sans perdre les informations primaires. De même, il s'agit de créer des informations adaptées et nécessaires. Par conséquent, une nouvelle stratégie d'augmentation des données a été démontrée sur la base de la préservation des informations essentielles et nécessaires pour adapter la technique de l'encodeur automatique compressé et une nouvelle technique pour éviter l'effet de sur-ajustement.

5.4. Conclusion

Dans ce chapitre nous avons présenté nos contributions à la segmentation à base d'auto-encoder compressé et à l'apprentissage par transfert, nous avons proposé des

nouvelles méthodes d'augmentation des données basée sur l'amélioration des caractéristiques de la base de données d'apprentissage (transformation APC et FILTER).

Les résultats expérimentaux ont montré que les méthodes d'augmentation proposées ont amélioré les performances du modèle. Les résultats ont également montré que la combinaison des méthodes proposées avec les méthodes d'augmentation des données couramment utilisées a donné de meilleures performances dans la segmentation des images lorsqu'on utilise des données d'entraînement limitées.

Les approches basées sur les nouvelles méthodes d'augmentation de données sont généralement construites dans le but d'effectuer une segmentation précise et de minimiser les erreurs de segmentation entre l'image prédite et la vérité terrain avec un temps de calcul moindre en évitant le sur-apprentissage.

Conclusion et perspectives

Conclusion et perspectives

Conclusion

Motivés par le développement récent des méthodes de segmentation d'objets basées sur l'apprentissage profond, nous avons présenté un système de détection des objets en mouvement dans la vidéosurveillance qui a un aspect important pour (l'analyse de mouvement, le suivi d'objets, l'amélioration de la qualité d'image et la compression de données). On présente également un deuxième système pour l'extraction des routes à partir des images aériennes, ce sujet est important dans nombreuse applications, comme (la gestion du trafic, la planification urbaine, la surveillance des routes, la navigation GPS et la mise à jour de la cartographie). Ce système a été proposé pour résoudre le problème d'existence d'éléments géo-spatiaux (réseau routier). Ces deux systèmes ont été étudiés avec l'utilisation des données d'entraînement limitées. La catégorie d'apprentissage profonde utilisée dans nos expériences est basée sur l'auto-encodeur en mode supervisé, qui est principalement utilisé pour des tâches de segmentation d'images non supervisé.

Dans le premier système, nous avons présenté des expériences comparant deux approches d'auto-encoder (apprentissage par transfert et a multi-profondeur) entraînées en utilisant trois stratégies différentes (DA, PCA et PCA+DA). Dans le deuxième système, nous avons proposé trois modèles compressées d'auto-encoder formé par deux stratégies (DA et transformation Filtre + DA), ces nouvelle stratégies de formation et d'entrainement ayant pour but d'augmenter la taille et d'améliorer les informations caractéristiques des données d'apprentissage.

Les résultats expérimentaux ont montré que les méthodes d'augmentation proposées ont amélioré les performances des modèles, et ont également montré que la combinaison des méthodes proposées avec les méthodes d'augmentation des données couramment utilisées a donné de meilleures performances dans la segmentation des images l'lorsqu'on utilise des données d'entraînement limitées. Ces nouvelles augmentations utilisent des concepts de transformation géométrique qui permet au modèle d'apprentissage automatique d'étudier les propriétés et de faire des relations structurales significatives des objets dans l'image.

L'objectif principal des systèmes proposés basés sur les nouvelles méthodes d'augmentation de données qui sont construits pour effectuer une segmentation précise et de minimiser les erreurs de segmentation entre le masque prédit et la vérité terrain avec un temps de calcul réduit en évitant le sur-apprentissage.

Cependant, on peut conclure que la combinaison de la transformation morphologique et géométrique pour l'apprentissage du système, aide le modèle à

améliorer ses capacités de généralisation et à générer un modèle précis avec un minimum de données d'entraînement. En outre, par rapport aux techniques traditionnelles d'augmentation des données (rotation, zoom et décalage) qui reposent sur le changement de l'emplacement des coordonnées dans le même plan mathématique, ce qui produit des variables corrélées et peut offrir une amélioration minimale.

Notre travail montre l'intérêt d'enrichir les données d'apprentissage comme avec l'ACP et les transformations basées sur des filtres (amélioration des bords et détection des contours) pour créer une nouvelle représentation des variables et d'accentuer les informations caractéristiques dans un nouveau plan avec une variance importante. De même, avec ces transformations, l'extraction des variables nécessaires à la tâche de segmentation et la suppression des variables inutiles qui peuvent fausser les résultats de la prédiction et retardé l'apprentissage. Donc, ces transformations rendent les données plus simples et faciles à traiter par un système moins complexe en réduisant la taille du modèle et la complexité de calcul, tout en obtenant une précision de segmentation meilleure.

En conclusion, les défis associés à l'apprentissage automatique, tels que l'overfitting, la nécessité de grandes quantités de données, le temps d'apprentissage long et le choix des hyper-paramètres, ont été abordés dans ce travail. Grâce à l'utilisation de nos stratégies d'augmentation de données, nous avons pu minimiser l'impact de ces défis sur la qualité de nos modèles liée à la tache de segmentation. Le choix judicieux des hyper-paramètres a également été un facteur clé dans l'obtention de bons résultats. Enfin, la mise en œuvre de méthodes efficaces pour traiter les défis mentionnés a permis d'améliorer considérablement les performances de nos modèles d'apprentissage profond.

Nous pensons que les stratégies que nous avons développées seront d'un intérêt considérable dans diverses applications telles que la biologie et le domaine médical, notamment dans les premières étapes des expériences où les échantillons sont limités. Compte tenu du fait que l'apprentissage profond est un sujet hautement recommandé dans plusieurs domaines, nous avons mené notre travail pour être l'une des premières contributions qui traite le problème de la phase d'apprentissage avec des données limitées et il a montré d'excellents résultats en termes de précision et de rapidité.

Perspectives

À la fin de notre travail, il est important d'identifier des axes de recherche prometteurs pour les travaux futurs. En général, il y a deux façons d'améliorer les travaux que nous avons présentés. La première approche au niveau data-set (ensembles de données), consiste à trouver une amélioration pour les méthodes d'augmentation des

données dans lequel les algorithmes existants sont basés sur des techniques d'apprentissage machine classique (transformations par ACP et Filtres) et à combler les lacunes des méthodes existantes dans ce cadre par des techniques d'apprentissage profond. La seconde consiste à améliorer l'architecture du réseau et à apporter les modifications appropriées à la modélisation du problème "Segmentation et classification par pixel". Ces deux points peuvent aider à améliorer les performances du modèle en cas de données d'entraînement limitées.

Dans ce qui suit, nous proposons quelques travaux perspectifs pour les deux domaines d'applications étudiés dans cette thèse :

- **L'extraction et la différenciation entre les routes primaires et secondaires à partir de la segmentation instance :** c'est un sujet important dans le domaine de l'analyse d'images aériennes. Les routes primaires et secondaires ont souvent des caractéristiques distinctes en termes de largeur, de trafic et d'importance pour les réseaux de transport. Il est important de faire l'extraction et différenciation précises entre les routes primaires et secondaires pour avoir des implications importantes pour la planification et la gestion du transport, ainsi que pour l'analyse de la sécurité routière et de l'impact environnemental.

Les perspectives pour aborder ce problème incluent :

1. Utilisation de caractéristiques géométriques, telles que la largeur de la route et la distance aux bâtiments adjacents, pour différencier les routes primaires et secondaires et exploitation des données de trafic pour améliorer la classification des routes.
2. L'utilisation des modèles génératifs tels que GANs ou VAEs pour augmenter les données d'apprentissage, ces modèles génératifs peuvent être utilisés pour générer de nouvelles images de réseaux routiers en différents éléments, tels que les voies et les intersections.
3. Développement de réseaux de neurones plus avancés pour prendre en compte des caractéristiques plus complexes des routes, telles que la qualité de la surface et les patterns de circulation et les segmenter en éléments pertinents..
4. L'intégration de ces modèles dans des systèmes de navigation et de surveillance de la circulation routière embarquée.

- **l'amélioration de la segmentation d'objets en mouvement par l'utilisation de modèles en profondeur en temps réel :** est un domaine de recherche actif en

intelligence artificielle. La segmentation d'objets en mouvement est un défi pour les algorithmes d'analyse vidéo en raison de la nature changeante des objets en mouvement. Les modèles en profondeur récents peuvent être utilisés pour améliorer la segmentation d'objets en mouvement en temps réel et qui vise à améliorer la précision et la rapidité de la segmentation.

Les perspectives prometteuses dans ce domaine :

1. L’utilisation des réseaux de neurones plus complexes pour extraire des caractéristiques clés des images tels que les modèles génératifs GANs.
2. L’intégration des nouvelles séquences dans des environnements plus complexes pour l’apprentissage tels que les scènes de circulation routière et les environnements industriels.
3. Combiner des algorithmes de segmentation avec l’algorithme de suivi d'objets en temps réel : Les algorithmes de suivi d'objets en temps réel pourraient fournir des informations de localisation plus précises des objets en mouvement.

Références

[1] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, P. Martinez-Gonzalez, and J. Garcia-Rodriguez, "A survey on deep learning techniques for image and video semantic segmentation," *Applied Soft Computing*, vol. 70, pp. 41-65, 2018.

[2] S. Ghosh, N. Das, I. Das, and U. Maulik, "Understanding deep learning techniques for image segmentation," *ACM Computing Surveys (CSUR)*, vol. 52, pp. 1-35, 2019.

[3] N. Rachburee and W. Punlumjeak, "An assistive model of obstacle detection based on deep learning: YOLOv3 for visually impaired people," *International Journal of Electrical & Computer Engineering (2088-8708)*, vol. 11, 2021.

[4] J. Li, A. Sun, J. Han, and C. Li, "A survey on deep learning for named entity recognition," *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, pp. 50-70, 2020.

[5] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, "Deep learning for visual understanding: A review," *Neurocomputing*, vol. 187, pp. 27-48, 2016.

[6] W. Ge, Z. Guo, Y. Dong, and Y. Chen, "Dynamic background estimation and complementary learning for pixel-wise foreground/background segmentation," *pattern recognition*, vol. 59, pp. 112-125, 2016.

[7] T. Bouwmans, "Traditional and recent approaches in background modeling for foreground detection: An overview," *Computer science review*, vol. 11, pp. 31-66, 2014.

[8] H. Liu, X. Han, X. Li, Y. Yao, P. Huang, and Z. Tang, "Deep representation learning for road detection using Siamese network," *Multimedia Tools and Applications*, vol. 78, pp. 24269-24283, 2019.

[9] A. Abdollahi, B. Pradhan, N. Shukla, S. Chakraborty, and A. Alamri, "Deep learning approaches applied to remote sensing datasets for road extraction: A state-of-the-art review," *Remote Sensing*, vol. 12, p. 1444, 2020.

[10] W. Wang, N. Yang, Y. Zhang, F. Wang, T. Cao, and P. Eklund, "A review of road extraction from remote sensing images," *Journal of traffic and transportation engineering (english edition)*, vol. 3, pp. 271-282, 2016.

[11] Y. Wang, Z. Luo, and P.-M. Jodoin, "Interactive deep learning method for segmenting moving objects," *Pattern Recognition Letters*, vol. 96, pp. 66-75, 2017.

[12] J. Gracewell and M. John, "Dynamic background modeling using deep learning autoencoder network," *Multimedia Tools and Applications,* vol. 79, pp. 4639-4659, 2020.

[13] A. Bouguettaya, H. Zarzour, A. Kechida, and A. M. Taberkit, "Vehicle detection from UAV imagery with deep learning: A review," *IEEE Transactions on Neural Networks and Learning Systems,* 2021.

[14] M. Babaee, D. T. Dinh, and G. Rigoll, "A deep convolutional neural network for video sequence background subtraction," *pattern recognition,* vol. 76, pp. 635-649, 2018.

[15] H. Salehinejad, S. Naqvi, E. Colak, J. Barfett, and S. Valaee, "Cylindrical transform: 3d semantic segmentation of kidneys with limited annotated images," in *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 2018, pp. 539-543.

[16] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431-3440.

[17] R. Keshari, M. Vatsa, R. Singh, and A. Noore, "Learning structure and strength of CNN filters for small sample size training," in *proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 9349-9358.

[18] H. Salehinejad, S. Valaee, T. Dowdell, and J. Barfett, "Image augmentation using radial transform for training deep neural networks," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 3016-3020.

[19] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM,* vol. 60, pp. 84-90, 2017.

[20] F. Shi, J. Wang, J. Shi, Z. Wu, Q. Wang, Z. Tang, K. He, Y. Shi, and D. Shen, "Review of artificial intelligence techniques in imaging data acquisition, segmentation, and diagnosis for COVID-19," *IEEE reviews in biomedical engineering*, vol. 14, pp. 4-15, 2020.

[21] Y. Oh, S. Park, and J. C. Ye, "Deep learning COVID-19 features on CXR using limited training data sets," *IEEE transactions on medical imaging,* vol. 39, pp. 2688-2700, 2020.

[22] S. Wan, X. Xu, T. Wang, and Z. Gu, "An intelligent video analysis method for abnormal event detection in intelligent transportation systems," *IEEE transactions on intelligent transportation systems,* vol. 22, pp. 4487-4495, 2020.

[23] B. Garcia-Garcia, T. Bouwmans, and A. J. R. Silva, "Background subtraction in real applications: Challenges, current models and future directions," *Computer science review,* vol. 35, p. 100204, 2020.

[24] A. Rasouli and J. K. Tsotsos, "Autonomous vehicles that interact with pedestrians: A survey of theory and practice," *IEEE transactions on intelligent transportation systems,* vol. 21, pp. 900-918, 2019.

[25] W. Brendel and S. Todorovic, "Video object segmentation by tracking regions," in *2009 IEEE 12th international conference on computer vision*, 2009, pp. 833-840.

[26] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2961-2969.

[27] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *Proceedings. 1999 IEEE computer society conference on computer vision and pattern recognition (Cat. No PR00149)*, 1999, pp. 246-252.

[28] A. Elgammal, D. Harwood, and L. Davis, "Non-parametric model for background subtraction," in *Computer Vision—ECCV 2000: 6th European Conference on Computer Vision Dublin, Ireland, June 26–July 1, 2000 Proceedings, Part II 6*, 2000, pp. 751-767.

[29] Y. Liu, Z. Bellay, P. Bradsky, G. Chandler, and B. Craig, "Edge-to-fog computing for color-assisted moving object detection," in *Big Data: Learning, Analytics, and Applications*, 2019, pp. 9-17.

[30] S. C. Sen-Ching and C. Kamath, "Robust techniques for background subtraction in urban traffic video," in *Visual Communications and Image Processing 2004*, 2004, pp. 881-892.

[31] Q. Zhou and J. Aggarwal, "Object tracking in an outdoor environment using fusion of features and cameras," *Image and Vision Computing,* vol. 24, pp. 1244-1255, 2006.

[32] O. Barnich and M. Van Droogenbroeck, "ViBe: A universal background subtraction algorithm for video sequences," *IEEE Transactions on Image Processing,* vol. 20, pp. 1709-1724, 2010.

[33] P.-L. St-Charles, G.-A. Bilodeau, and R. Bergevin, "SuBSENSE: A universal change detection method with local adaptive sensitivity," *IEEE Transactions on Image Processing*, vol. 24, pp. 359-373, 2014.

[34] P.-L. St-Charles, G.-A. Bilodeau, and R. Bergevin, "A self-adjusting approach to change detection based on background word consensus," in *2015 IEEE winter conference on applications of computer vision*, 2015, pp. 990-997.

[35] S. Javed, A. Mahmood, S. Al-Maadeed, T. Bouwmans, and S. K. Jung, "Moving object detection in complex scene using spatiotemporal structured-sparse RPCA," *IEEE Transactions on Image Processing*, vol. 28, pp. 1007-1022, 2018.

[36] M. Braham and M. Van Droogenbroeck, "Deep background subtraction with scene-specific convolutional neural networks," in *2016 international conference on systems, signals and image processing (IWSSIP)*, 2016, pp. 1-4.

[37] P. W. Patil and S. Murala, "MSFgNet: A novel compact end-to-end deep network for moving object detection," *IEEE transactions on intelligent transportation systems*, vol. 20, pp. 4066-4077, 2018.

[38] X. Liang, S. Liao, X. Wang, W. Liu, Y. Chen, and S. Z. Li, "Deep background subtraction with guided learning," in *2018 IEEE International Conference on Multimedia and Expo (ICME)*, 2018, pp. 1-6.

[39] J. Liao, G. Guo, Y. Yan, and H. Wang, "Multiscale cascaded scene-specific convolutional neural networks for background subtraction," in *Advances in Multimedia Information Processing–PCM 2018: 19th Pacific-Rim Conference on Multimedia, Hefei, China, September 21-22, 2018, Proceedings, Part I 19*, 2018, pp. 524-533.

[40] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[41] L. A. Lim and H. Y. Keles, "Foreground segmentation using convolutional neural networks for multiscale feature encoding," *Pattern Recognition Letters*, vol. 112, pp. 256-262, 2018.

[42] L. A. Lim and H. Y. Keles, "Learning multi-scale features for foreground segmentation," *Pattern Analysis and Applications*, vol. 23, pp. 1369-1380, 2020.

[43] T. Akilan, Q. J. Wu, A. Safaei, J. Huo, and Y. Yang, "A 3D CNN-LSTM-based image-to-image foreground segmentation," *IEEE transactions on intelligent transportation systems*, vol. 21, pp. 959-971, 2019.

[44] D. Sakkos, H. Liu, J. Han, and L. Shao, "End-to-end video background subtraction with 3d convolutional neural networks," *Multimedia Tools and Applications,* vol. 77, pp. 23023-23041, 2018.

[45] Y. Gao, H. Cai, X. Zhang, L. Lan, and Z. Luo, "Background subtraction via 3D convolutional neural networks," in *2018 24th International Conference on Pattern Recognition (ICPR)*, 2018, pp. 1271-1276.

[46] M. C. Bakkay, H. A. Rashwan, H. Salmane, L. Khoudour, D. Puig, and Y. Ruichek, "BSCGAN: Deep background subtraction with conditional generative adversarial networks," in *2018 25th IEEE International Conference on Image Processing (ICIP)*, 2018, pp. 4018-4022.

[47] P. Patil and S. Murala, "Fggan: A cascaded unpaired learning for background estimation and foreground segmentation," in *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2019, pp. 1770-1778.

[48] F. Bahri and N. Ray, "Dynamic Background Subtraction by Generative Neural Networks," in *2022 18th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 2022, pp. 1-8.

[49] O. Germain and P. Réfrégier, "Edge location in SAR images: Performance of the likelihood ratio filter and accuracy improvement with an active contour approach," *IEEE Transactions on Image Processing,* vol. 10, pp. 72-78, 2001.

[50] S.-R. Park, "Semi-automatic road extraction algorithm from IKONOS images using template matching," in *Proc. 22nd Asian Conference on Remote Sensing, 2001,* 2001.

[51] J. B. Mena, "State of the art on automatic road extraction for GIS update: a novel classification," *Pattern Recognition Letters,* vol. 24, pp. 3037-3058, 2003.

[52] A. Kirthika and A. Mookambiga, "Automated road network extraction using artificial neural network," in *2011 International Conference on Recent Trends in Information Technology (ICRTIT)*, 2011, pp. 1061-1065.

[53] M. Song and D. Civco, "Road extraction using SVM and image segmentation," *Photogrammetric Engineering & Remote Sensing,* vol. 70, pp. 1365-1371, 2004.

[54] Z. Miao, B. Wang, W. Shi, H. Wu, and Y. Wan, "Use of GMM and SCMS for accurate road centerline extraction from the classified image," *Journal of Sensors,* vol. 2015, 2015.

[55] Z. Zhong, J. Li, W. Cui, and H. Jiang, "Fully convolutional networks for building and road extraction: Preliminary results," in *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2016, pp. 1591-1594.

[56] Y. Wei, Z. Wang, and M. Xu, "Road structure refined CNN for road extraction in aerial image," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, pp. 709-713, 2017.

[57] Q. Shi, X. Liu, and X. Li, "Road detection from remote sensing images by generative adversarial networks," *IEEE Access*, vol. 6, pp. 25486-25494, 2017.

[58] F. Chollet, *Deep learning with Python*: Simon and Schuster, 2021.

[59] L. Huawei Technologies Co., "Huawei MindSpore AI Development Framework," in *Artificial Intelligence Technology*, ed: Springer, 2022, pp. 137-162.

[60] J. Verbraeken, M. Wolting, J. Katzy, J. Kloppenburg, T. Verbelen, and J. S. Rellermeyer, "A survey on distributed machine learning," *ACM Computing Surveys (CSUR)*, vol. 53, pp. 1-33, 2020.

[61] S. Dong, P. Wang, and K. Abbas, "A survey on deep learning and its applications," *Computer science review*, vol. 40, p. 100379, 2021.

[62] C. Janiesch, P. Zschech, and K. Heinrich, "Machine learning and deep learning," *Electronic Markets*, vol. 31, pp. 685-695, 2021.

[63] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, pp. 2278-2324, 1998.

[64] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, pp. 2481-2495, 2017.

[65] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, 2015, pp. 234-241.

[66] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of big data*, vol. 6, pp. 1-48, 2019.

[67] K. Seddiki, P. Saudemont, F. Precioso, N. Ogrinc, M. Wisztorski, M. Salzet, I. Fournier, and A. Droit, "Towards cnn representations for small mass spectrometry data classification: From transfer learning to cumulative learning," *bioRxiv*, p. 2020.03. 24.005975, 2020.

[68] I. Idrissi, M. Azizi, and O. Moussaoui, "Accelerating the update of a DL-based IDS for IoT using deep transfer learning," *Indones. J. Electr. Eng. Comput. Sci*, vol. 23, pp. 1059-1067, 2021.

[69] M. Al-Smadi, M. Hammad, Q. B. Baker, and A. Sa'ad, "A transfer learning with deep neural network approach for diabetic retinopathy classification,"

International Journal of Electrical and Computer Engineering, vol. 11, p. 3492, 2021.

[70] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He, "A comprehensive survey on transfer learning," *Proceedings of the IEEE,* vol. 109, pp. 43-76, 2020.

[71] D. Bank, N. Koenigstein, and R. Giryes, "Autoencoders," *arXiv preprint arXiv:2003.05991,* 2020.

[72] S. Leonov, A. Vasilyev, A. Makovetskii, V. Kuznetsov, and J. Diaz-Escobar, "An algorithm for selecting face features using deep learning techniques based on autoencoders," in *Applications of Digital Image Processing XLI*, 2018, pp. 792-798.

[73] S. A. Ebiaredoh-Mienye, E. Esenogho, and T. G. Swart, "Artificial neural network technique for improving prediction of credit card default: A stacked sparse autoencoder approach," *International Journal of Electrical and Computer Engineering,* vol. 11, p. 4392, 2021.

[74] Y. Wang, P.-M. Jodoin, F. Porikli, J. Konrad, Y. Benezeth, and P. Ishwar, "CDnet 2014: An expanded change detection benchmark dataset," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2014, pp. 387-394.

[75] S. M. Shaharudin, N. Ahmad, and S. M. C. M. Nor, "A modified correlation in principal component analysis for torrential rainfall patterns identification," *IAES International Journal of Artificial Intelligence,* vol. 9, p. 655, 2020.

[76] C. Kamlaskar and A. Abhyankar, "Multilinear principal component analysis for iris biometric system," *Indonesian Journal of Electrical Engineering and Computer Science,* vol. 23, pp. 1458-1469, 2021.

[77] C. Clausen and H. Wechsler, "Color image compression using PCA and backpropagation learning," *pattern recognition,* vol. 33, pp. 1555-1560, 2000.

[78] G.-C. Feng, P. C. Yuen, and D.-Q. Dai, "Human face recognition using PCA on wavelet subband," *Journal of electronic imaging,* vol. 9, pp. 226-233, 2000.

[79] A. W. Reza, M. M. Hasan, N. Nowrin, and M. A. Shibly, "Pre-trained deep learning models in automatic COVID-19 diagnosis," *Indonesian Journal of Electrical Engineering and Computer Science,* vol. 22, pp. 1540-1547, 2021.

[80] C.-I. Cira, R. Alcarria, M.-Á. Manso-Callejo, and F. Serradilla, "A deep convolutional neural network to detect the existence of geospatial elements in

high-resolution aerial imagery," *Multidisciplinary Digital Publishing Institute Proceedings,* vol. 19, p. 17, 2019.

[81] S. Raschka, *Python machine learning*: Packt publishing ltd, 2015.

[82] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, and M. Isard, "Tensorflow: a system for large-scale machine learning," in *Osdi*, 2016, pp. 265-283.

[83] F. Chollet, "Keras: The python deep learning library," *Astrophysics source code library*, p. ascl: 1806.022, 2018.

[84] E. Bisong, *Building machine learning and deep learning models on Google cloud platform*: Springer, 2019.

[85] T. S. Gunawan, A. Ashraf, B. S. Riza, E. V. Haryanto, R. Rosnelly, M. Kartiwi, and Z. Janin, "Development of video-based emotion recognition using deep learning with Google Colab," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 18, pp. 2463-2471, 2020.

[86] O. Tezcan, P. Ishwar, and J. Konrad, "BSUV-Net: A fully-convolutional neural network for background subtraction of unseen videos," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2020, pp. 2774-2783.

[87] M. O. Tezcan, P. Ishwar, and J. Konrad, "BSUV-Net 2.0: Spatio-temporal data augmentations for video-agnostic supervised background subtraction," *IEEE Access*, vol. 9, pp. 53849-53860, 2021.

[88] S. Bianco, G. Ciocca, and R. Schettini, "Combination of video change detection algorithms by genetic programming," *IEEE Transactions on Evolutionary Computation*, vol. 21, pp. 914-928, 2017.

[89] P.-L. St-Charles, G.-A. Bilodeau, and R. Bergevin, "Universal background subtraction using word consensus models," *IEEE Transactions on Image Processing*, vol. 25, pp. 4768-4781, 2016.

Liste des publications

Les publications et communications relatives à cette thèse sont les suivantes :

Publication Internationale :

- Kebir, Abdeldjalil, and Mahmoud Taibi. "**End-to-end deep auto-encoder for segmenting a moving object with limited training data.**" International Journal of Electrical & Computer Engineering (2088-8708) 12.6 (2022).

Communication Internationale :

- Kebir, Abdeldjalil, Mahmoud Taibi, and Francisco Serradilla. "**Compressed VGG16 auto-encoder for road segmentation from aerial images with few data training.**" Conference Proceedings ICCSA. 2021.
- Kebir, Abdeldjalil, Mahmoud Taibi, and Ahmed Kechida. " **Auto-encoder Method for Segmenting Moving Object with Few Data Training.**" Conference Proceedings ICESTI.2019.
- Kebir, Abdeldjalil, Mahmoud Taibi, and Houda Bouaziz. " **Background subtraction using Deep learning methods: Review**." Conference Proceedings ICEEB.2018.

Communication nationale :

- Kebir, Abdeldjalil, and Mahmoud Taibi. " **A Deep Convolution Auto-encoder method for Segmenting Moving object with few Data Training.**" Conference Proceedings DeepLearning IndabaXAlgeria.2019.