الجمهورية الجزائرية الديمقراطية الشعبية

وزارة التعليم العالي والبحث العلمي

UNIVERSITE BADJI MOKHTAR - ANNABA
BADJI MOKHTAR – ANNABA UNIVERSITY

جامعة باجي مختار – عنابـــة

Faculté : Sciences de l'Ingéniorat

Département : Electronique

Domaine : Sciences et technologie

Filière : télécommunication

Spécialité : systèmes de télécommunication

Mémoire

Présenté en vue de l'obtention du Diplôme de Master

Thème:

Segmentation des images de drones par apprentissage profond ou Deep Learning

- **Présenté par** : *Boutalba Med Amine*

Encadrant : *BOUKARI Karima* *Grade :MCA* *UBM Annaba*

Jury de Soutenance :

HAFS Toufik	MCB	UBM Annaba	Président
BOUKARI Karima	MCA	UBM Annaba	Encadrant
			Co-encadrant
AMARA Fethi	MCB	UBM Annaba	Examinateur
			Deuxième examinateur / Membre invité

Année Universitaire : 2019/2020

ملخـص:

قد استفادت الصور الجوية ذات الارتفاعات المنخفضة والمتوسطة سلسلة كاملة من أنواع تطبيقات الطائرات بدون طيار مثل مراقبة الكوارث ومراقبة حركة المرور وخطط الإخلاء الطارئة لهذا الغرض. شبكات الطرق أمر ضروري.

أظهرت التطورات الحديثة في التعلم الآلي الأداء الرائع للشبكات العصبية التلافيفية. يتعامل هذا العمل مع استخدام الشبكات العصبية التلافيفية العميقة لتصنيف كثيف لصور الطائرات بدون طيار. على وجه الخصوص، نقوم بتدريب شبكتين، الأولى هي مشفر decoder_غير مدرب مسبقًا، والثانية متغير من بنية SegNet مع التعلم المسبق على Vgg16 على الصور الجوية.

يتم الحصول على النتائج خلال مرحلة الاختبار، مما يجعل من الممكن عرض الصور المجزأة بفضل خطوة التفكيك والامتصاص. النتائج ليست دقيقة للغاية، ولكن يمكن صقلها في العمل المستقبلي.

الكلمات الدالة :

صورة الطائرات بدون طيار، التجزئة، الشبكات العصبية التلافيفية، التعلم العميق.

ABSTRACT :

Aerial imagery at low and medium altitudes has benefited a whole series of types of drone applications such as disaster monitoring, traffic monitoring and evacuation emergency plans for this purpose network extraction road is essential. Recentadvances in machine learning have shown the great performance of convolutional neural networks. This workconcerns the use of deep convolutional neural networks for the dense classification of uav images. In particular, we train two networks the first a non-pre-trained _decoder encoder, the second a variant of the SegNet architecture with a pre-learning on Vgg16 on aerial images. Semi urban and non urban.

The results are obtainedduring the test phase, which makesit possible to display segmented images thanks to the deconvolution and oversampling stage. The results are not very precise, but they canberefined in future work.

Keywords :

UAV image, segmentation, Convolutional neural networks, Deep Learning.

RESUME :

L’imagerie aérienne à basse et moyenne altitude, a bénéficié à toute une série de types d'applications des drones telles que le suivi des catastrophes, la surveillance du trafic et les plans d'urgence pour l'évacuation pour cela l’extraction des réseaux routiers est primordial.

Les récents progrès en apprentissage automatique ont montré les très grandes performances des réseaux de neurones convolutifs. Ce travail porte sur l’utilisation des réseaux de neurones convolutifs profonds pour la classification dense des images uav. En particulier, nous entraînons deux réseaux le premier un encodeur _décodeur non pré-entrainé, le deuxième une variante de l’architecture SegNet avec un pre-apprentissage sur Vgg16 sur des images aériennes .semi urbaines et non urbaines.

Les résultats sont obtenus à la phase de test, ce qui permet d’afficher des images segmentées grâce à l’étape de déconvolution et sur-échantillonnage. Les résultats ne sont pas d’une grande précision, mais ils peuvent être affinés lors de travaux futurs.

Mots-clés :

Image UAV, segmentation, Réseaux de neurones convolutionnels, Deep Learning.

REMERCIMENTS

Je tiens tout d'abord à remercier l'encadreur pour m'avoir fait confiance , guidé, encouragé et conseillé, j'espère avoir été à la hauteur.
Je lui suis également reconnaissant pour le temps conséquent qu'il m'a accordé, ses qualités pédagogiques et scientifiques, sa franchise et sa sympathie. J'ai beaucoup appris à ses côtés et je lui adresse ma gratitude pour tout cela.

J'adresse de sincères remerciements à nos profs et aux membres du jury pour m'avoir fait l'honneur de participer au jury de soutenance.

Enfin, je voudrais remercier toutes les personnes qui ont participé de près ou de loin à mes recherches et à l'élaboration de ce mémoire.

DEDICACES

Que ce travail témoigne de mes respects :

A mes parents.

A Mes frères, mes sœurs et à toute ma famille

Pour leur encouragement et leurs grands sacrifices. Aucune dédicace ne pourrait exprimer mon respect, ma considération et mes profonds sentiments envers eux. Je prie Dieu de les bénir, de veiller sur eux, en espérant qu'ils seront toujours fiers de moi.

A tous mes professeurs

Leur générosité et leur soutien m'oblige de leurs témoigner mon profond respect et ma loyale considération.

A mes amis et collègues

Et à tous ceux qui me sont chers
Ils vont trouver ici l'expression d'une fidélité et d'une amitié infinie, de mes sentiments de reconnaissance pour le soutien qu'ils n'ont cessé de me porter.
Trouvez dans ce modeste travail mes sincères gratitudes et
Reconnaissance.
Ce travail est votre.

Liste des Tableaux

Liste des Figures

Liste des Figures

Liste Des Symboles

CNN : Convolutional Neural Nettworks.

UAV : Unmanned Aerial Vehicle.

GPU : Graphical Processing Unit.

TP : True Positive.

TN : True Negative.

FN : False Negative.

FP : False Positive.

Wo : Le nombre de neurones du volume de sortie.

Wi : La taille de volume d'entrée (nombre de champs récepteur).

K : La surface de traitement.

P : La taille de la marge.

S : Pas de convolution.

Liste Des Formules

Liste Des Matières :

Liste Des Matières :

Introduction Générale

.....

Les véhicules aériens sans pilote (UAV) ont été largement utilisés dans de nombreux domaines, notamment dans les transports. Les principales applications sont la surveillance de la sécurité, le contrôle du trafic, l'inspection de la construction des routes et la surveillance de la circulation, fleuve, littoral, pipeline, etc.

Les drones équipés de caméras sont considérés comme une sorte de plate-forme à faible coût qui peut fournir des mécanismes d'acquisition de données pour les systèmes de transport intelligents.

Avec l'utilisation croissante des véhicules et leurs exigences en matière de gestion du trafic, ce type de plate-forme devient de plus en plus populaire. La collecte conventionnelle de données sur le trafic s'appuie sur des infra structures limitées à une région locale et, par conséquent, elle est coûteuse et exigeante en main-d'œuvre pour surveiller les activités de circulation à travers de grands domaines. En comparaison, le drone présente des avantages, notamment

(1) le coût de la surveillance sur de longues distances est faible ;

(2) il est flexible pour voler sur de larges échelles spatiales et temporelles

(3) il est capable de transporter divers types de capteurs pour collecter des données abondantes. Pour collecter des informations pour le transport il est important de savoir où se trouvent les routes dans les UAV

vidéos. La connaissance des zones routières peut fournir aux utilisateurs les régions d'intérêt pour la poursuite de la navigation, la détection et la collecte de données en profitant de leur efficacité et de leur précision.

La segmentation d'une image aérienne, consiste à attribuer à chaque pixel de cette image une classe. L'attribution de cette classe s'appuie sur l'analyse visuelle propre au pixel mais peut aussi s'appuyer sur la description visuelle de son voisinage. Les travaux existants utilisent des outils de classification supervisée, reposant sur des algorithmes d'apprentissage automatique. Le terme supervisé provient de l'étape d'entraînement de l'algorithme qui consiste à modéliser les classes en jeu à partir d'un jeu de données de référence : on peut ensuite inférer les classes de données non étiquetées en leur appliquant le modèle ainsi entraîné. Le jeu de données de référence est un ensemble de pixels, décrits par des attributs (observations), et dont on connaît à l'avance les classes : chaque classe est ainsi modélisée par rapport à ces attributs. Il est important de noter que la qualité d'une classification dépend du choix des attributs pour discriminer au mieux les classes entre elles.

Les techniques d'apprentissage profond, et plus particulièrement les réseaux de neurones convolutifs (CNNs) répondent parfaitement aux critères énoncés. A l'inverse des méthodes nécessitant l'extraction d'attribut pertinent en amont du classifieur, les CNNs apprennent de bout en bout.

Ainsi, on s'affranchit de l'étape de d'extraction d'attributs, requérant des connaissances expertes, différentes selon les tâches à effectuer. Les attributs appris par le CNN dépendant du jeu d'apprentissage, ils prennent implicitement en compte les relations entre les classes d'objets. La conséquence directe de l'apprentissage des attributs en même temps que du classifieur est le besoin massif de données d'apprentissage, puisqu'il s'agit de déterminer l'ensemble des filtres constituant le CNN.

Introduction Générale

.....

Les algorithmes de classification constituent un outil essentiel pour la segmentation des images. Les récents progrès en apprentissage automatique ont montré les très grandes performances des réseaux de neurone convolutifs pour de nombreuses applications, y compris la classification d'images aériennes .Le présent document est axé sur la détection des routes dans les zones semi urbaines à faible et à moyen altitudes.

Ce travail établit une stratégie quant à l'utilisation d'un réseau de neurone convolutif pour la classification d'images uav à haute résolution spatiale Des recherches ont démontré que les performances d'un CNN augmentent avec sa profondeur ; cependant multiplier les couches de convolution d'un CNN multiplie également le nombre de paramètres à optimiser, pouvant mener à un cas de sur-apprentissage (le modèle s'adapte exactement au jeu d'entraînement, menant à de faibles performances sur de nouvelles données non étiquetées).

Le premier chapitre est dédié au cadre de ce mémoire. Nous y présentons quelques notions sur la segmentation de l'image. Nous allons citer différentes techniques de segmentation d'image.

Puis dans le deuxième chapitre, ensuite détailler les méthodes neuronales avec le cas particulier les réseaux de neurones convolutionnels.

Le chapitre trois donne plus de détails sur l'algorithme de segmentation et les étapes qu'il faut respecter pour parvenir à un résultat. Les résultats seront alors exposés et commentés.

Chapitre 01

Segmentation des images aériennes

Chapitre1 Segmentation des images aériennes

1.1 Introduction :

Notre travail s'articule autour d'un grand outil de traitement d'image : la segmentation. La segmentation d'images est un processus visant à décomposer une image en un ensemble de régions ou classes ou sous-ensembles homogènes au sens d'un ou plusieurs critères. En imagerie aérienne, la segmentation est très importante, que ce soit pour l'extraction de paramètres ou de mesures sur les images, ainsi que pour la représentation et la visualisation. Dans notre application la segmentation sera exploitée pour détecter (visualiser) le les zones routières.

La segmentation fait appel à plusieurs branches des mathématiques et de l'informatique. Comprendre les enjeux réels et déterminer les performances, les avantages et les inconvénients de chaque approche est donc une tâche difficile. Le choix d'une méthode sera toutefois guidé par des critères comme le type d'image, le bruit, les artéfacts d'acquisition, l'information a priori disponible, le temps de calcul, le niveau d'interaction acceptable ou encore la possibilité de corriger efficacement le résultat. D'autre part, l'expérience, le savoir-faire, sont indéniablement la clef du succès.

Plusieurs méthodes de segmentation existent, que ce soit avec l'approche région qu'avec l'approche contour (filtre spatial, les approches multi-résolutions (ondelettes, quadtree…), l'approche morphologie mathématique…). Toutes ces méthodes seules ou combinées nous ont permis de résoudre le problème posé mais à chaque fois on fait intervenir un paramètre manuel. Bien qu'il ait été prouvé que la segmentation est un problème mal posé (solutions multiples pour un même problème), nous sommes toujours à la recherche de « la méthode » de segmentation idéale, en tout cas pour une application donnée.

Dans ce chapitre Nous allons d'abord expliquer le contexte de notre application, exposer les méthodes de segmentations. Nous mettrons l'accent sur la segmentation sémantique objet de notre intérêt.

1.2 Contexte :

L'imagerie aérienne par drone pour la détection de routes :

L'imagerie aérienne à basse et moyenne altitude, a bénéficié à toute une série de types d'applications des drones telles que la surveillance du trafic et la surveillance, la planification des parcours et l'inspection.

Les images Satellitaires sont utilisées pour extraire les réseaux routiers, notamment pour l'acquisition et la mise à jour des données GPS. Cependant, certaines applications, telles que le suivi des catastrophes, la surveillance du trafic et les plans d'urgence pour l'évacuation, ont besoin d'un traitement en temps réel des données détectées, où les images satellites ne sont pas aussi appropriées que les données saisies par les drones de terrain.

Les drones sont largement utilisés comme une plate-forme à faible coût qui peut efficacement acquérir des données de terrain avec le développement de matériel. Il s'agit de la principale force motrice à laquelle le traitement et l'analyse d'images des drones font de plus en plus attention récemment. Pour de nombreuses applications basées sur des drones, il serait souhaitable pour que les drones sachent où se trouvent les régions routières.

Dans la littérature, les images satellites des routes a fait l'objet de travaux basés sur le raisonnement heuristique et sur la ligne droite en comparaison, la détection des routes pour extraire les routes dans les images des drones (en particulier pour les drones de basse et moyenne altitude), n'en est encore qu'à ses débuts pour les UAV à base d'images.

Sur les images satellites, les routes sont généralement très étroites et les routes peuvent donc être modélisées sous forme de lignes ou de courbes.

Les méthodes existantes utilisent le plus souvent un seuil simple basé sur la couleur ou une ligne droite pour la détection pour identifier les zones routières.

Dans les images de drones, les intensités / les couleurs dans les aires de la route peuvent être inhomogènes, la largeur des routes intra- et/ou inter-images varie beaucoup, et les routes ne sont probablement pas bien pavées. Par conséquent, les méthodes de détection des routes utilisées dans les images satellitaires ne sont généralement pas adaptées aux images de drones.

Les observations ci-dessus nous motivent à concevoir notre propre modèle de route pour les images de drones. Comme nous le savons, essentiellement, une route se compose de deux propriétés : la géométrie et la radiométrie. On peut toujours représenter une route comme des zones allongées "presque" homogènes avec une largeur "plus ou moins" constante.1 Cela signifie que la variation de couleurs et de largeurs pour une route individuelle est délimitée.

La répartition de la largeur d'une route se distingue de celle des autres contextes (tels que maisons, buissons et arbres), permettant d'indiquer la route initiale de la région. En complément de la forme, la couleur de la route se distingue également de celle des contextes dans de nombreuses scènes de route.

1.3Différentes approches de segmentation :

Beaucoup de méthodes de segmentation existent dans la littérature et peuvent être séparées en deux grandes familles.

a. Les méthodes de segmentation par contours.

b. Les méthodes de segmentation en région homogènes.

Dans la première approche, On s'intéresse aux frontières des régions et dans la deuxième approche on s'intéresse au contenu de région [1] (figure 1.1).

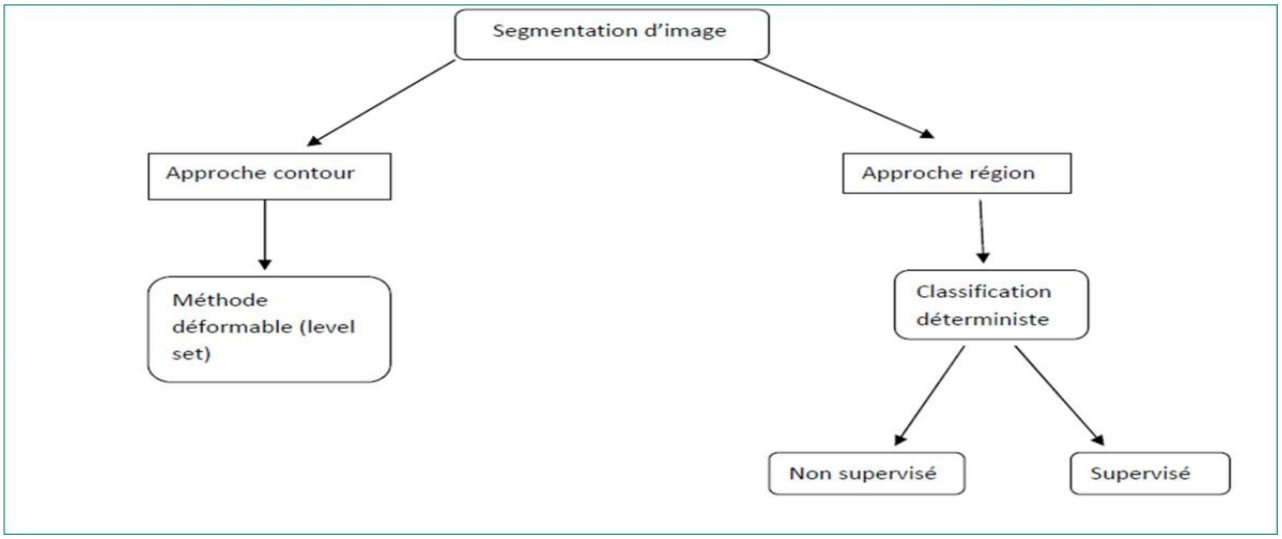

Figure (1.1) : Les approches de segmentation d'image.

- **Approche région :**

Les approches de segmentation régions visent à créer une partition de l'image en un ensemble de régions homogènes au sens d'un ou plusieurs critères pour avoir des ensembles de pixels partageant des propriétés communes. Les régions sont différenciées entre elles par des propriétés élémentaires fondées sur des critères locaux tels que le niveau de gris de chaque pixel, ou bien sur un attribut estimé dans le voisinage du pixel tel que la valeur moyenne, la variance, ou des paramètres de texture.

L'ensemble des regroupements de pixels constitue une segmentation d'image [2].

- **Approches frontières (contour) :**

Contrairement aux approches régions, qui cherchent à former des zones homogènes, les approches contours se basent sur les discontinuités des images pour déterminer les contours des régions et trouver les di similarités avec la recherche à exploiter le fait qu'il existe une transition détectable entre deux régions connexes. La détection de contours consiste à repérer les points d'une image numérique qui correspondent à un changement brutal du niveau de gris [3].

Les méthodes de détection de contours donnent de bons résultats quand les contours de l'image sont bien définis. Cependant, dans le cas des images bruitées ou faiblement contrastées les contours obtenus ne sont pas connexes et fermés, alors c'est une méthode qui nécessite une étape supplémentaire afin de fermer les bords des régions [4].

- **Approches classification :**

La Classification est un processus qui permet de rassembler les pixels d'une image dans des sous-ensembles qui présentent une similitude et une uniformité selon un critère prédéfini, on parle de partitionnement ou de clustering (classe).

Approche s'appuie sur les concepts de la logique floue [5].

Les méthodes de classification sont issues des méthodes statistiques multidimensionnelles. Il n'existe pas une méthode de classification qui peut s'appliquer à tout type d'image et qui peut fournir un partitionnement optimal. Ce qui explique la grande diversité de méthodes de classification qui existe dans la littérature.

Le choix d'une méthode est déterminé par différents facteurs tels que le nombre de classes attendues, la forme des classes extraites ou encore le chevauchement ou non des classes [6] (figure 1.2).

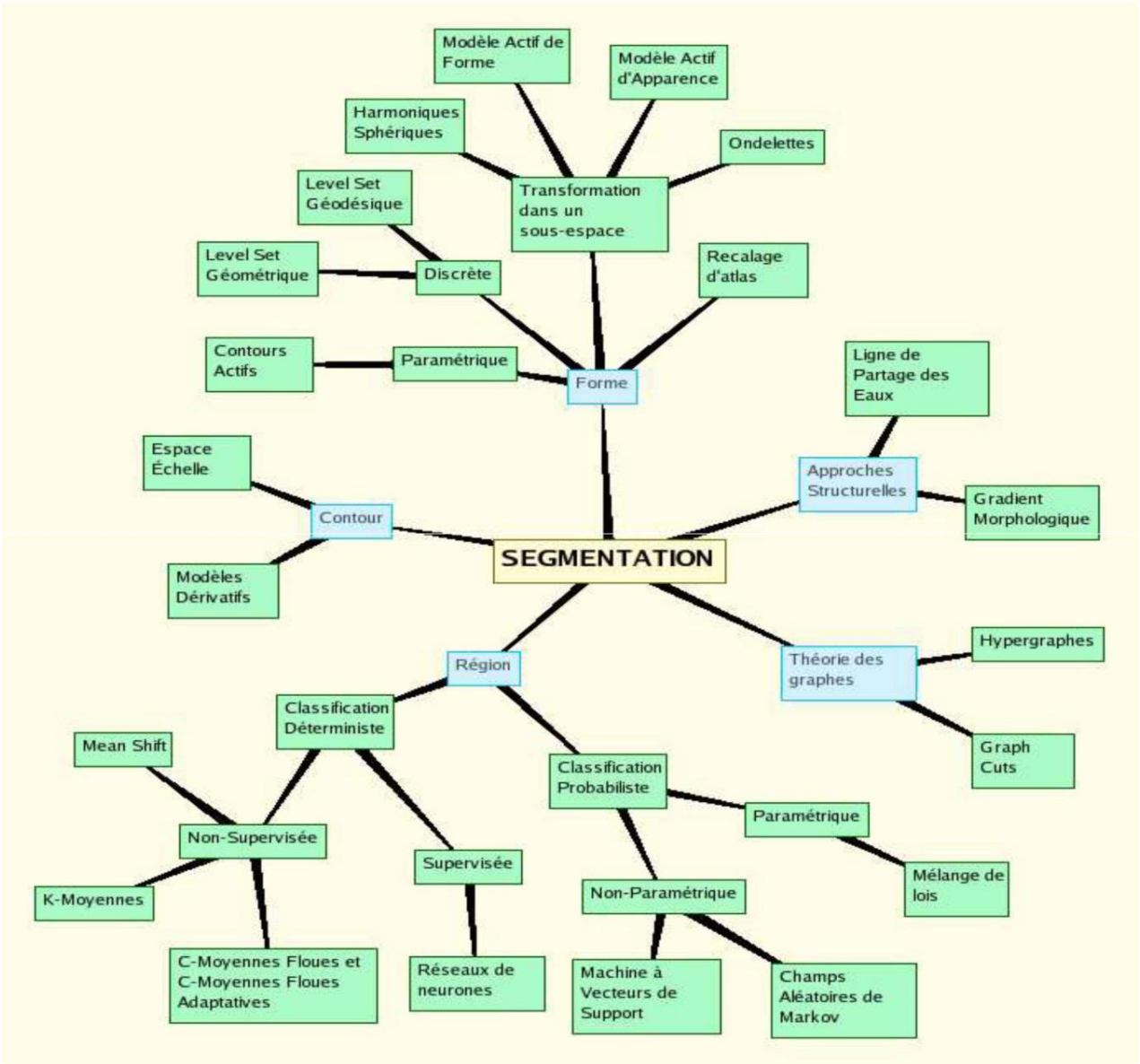

Figure (1.2) : Classification des différentes méthodes de segmentation

1.4Segmentation des images aériennes par des techniques de classification :

Les méthodes de classification permettent de regrouper des objets en groupes ou classes d'objets plus homogènes. Les objets regroupés ont des caractéristiques communes, ils sont similaires mais se distinguent clairement des objets des autres classes. Les résultats des méthodes de classification sont plus ou moins différents.

Méthode de classification :

La mise en œuvre d'une procédure de classification, ayant pour objectif de classifier automatiquement des objets, comporte généralement deux phases fondamentales :

- **Une phase d'apprentissage** : dont le but est de déterminer un espace de représentation des signaux et de rechercher les paramètres discriminants capables de caractériser chaque classe d'objets.
- **Une phase de reconnaissance** : au cours de laquelle on attribue à une classe chacun des objets inconnus dans l'espace de représentation déterminé durant l'apprentissage. La classification peut être supervisée ou non supervisée [7].

1.4.1 : Méthodes non supervisées :

Les méthodes non supervisées, elles ne nécessitent aucune base d'apprentissage et aucune tâche préalable d'étiquetage manuel. La seule intervention de l'expert se situe à la fin du processus pour identifier les classes trouvées. Parmi ces méthodes, on peut citer l'algorithme des K-moyennes (K-Means), l'algorithme des C-moyennes floues et les approches probabilistes [8]

1.4.2 : Méthodes supervisées :

Les approches supervisées nécessitent une étape d'apprentissage sur un échantillon avant de pouvoir être l'appliquer sur de nouvelles données. On répertorie, entre autres, dans ce type d'approche : les réseaux de neurones, les Support Vector Machine (SVM), et les K-plus proche voisins. Les approches supervisées nécessitent généralement une interaction avec l'utilisateur pour le choix de l'échantillon d'apprentissage, source de variabilité et de non-reproductibilité des résultats. Ce type de méthodes est cependant intégré dans des approches combinées.

Dans la cadre de ce projet nous nous intéresserons essentiellement à la segmentation par approche classification. Nous allons présenter quelques algorithmes qui répondent à la classification supervisée tel que l'algorithme des réseaux de neurones [8] (figure 1.3).

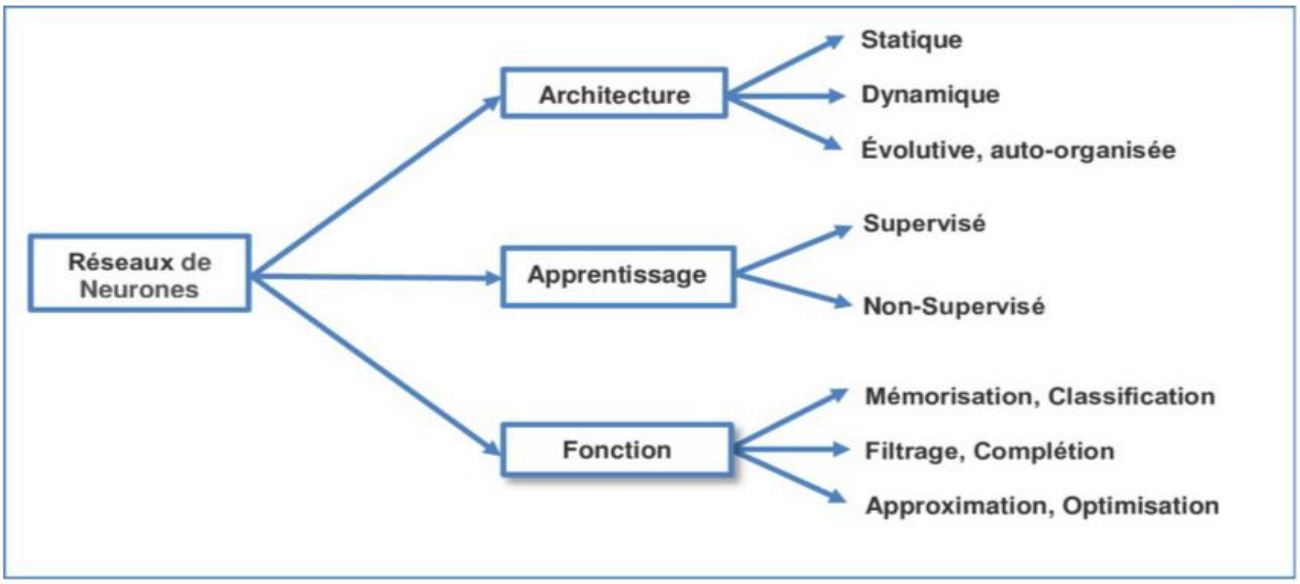

Figure (1.3) : Différentes possibilités de classification des réseaux de neurones

1.5 Segmentation sémantique d'images

La segmentation sémantique associe une étiquette ou une catégorie à chaque pixel d'une image. Elle permet de reconnaître un ensemble de pixels qui forment des catégories distinctes.

La séparation d'images en deux classes est un exemple simple de segmentation sémantique. Par exemple, à la figure (1.4), une image présentant une personne à la plage est associée à une version montrant les pixels de l'image segmentés en deux classes distinctes : la personne et l'arrière-plan

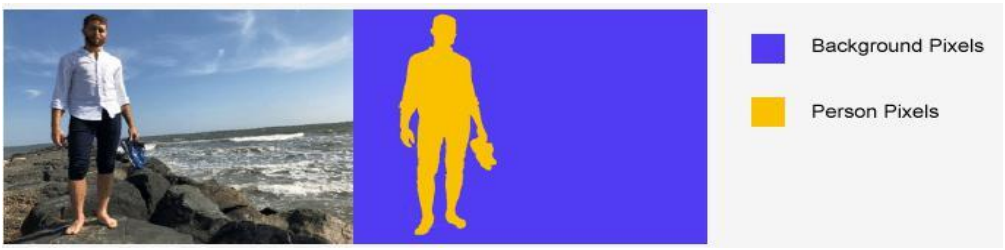

Figure (1.4) : Image et étiquette Des pixels

La segmentation sémantique étiquette les pixels d'une image c'est ce qui la rend utile dans des applications de divers domaines :

Conduite autonome : pour identifier un parcours conduisible pour les véhicules en distinguant la route des obstacles tels que les piétons, trottoirs, poteaux et autres véhicules.

Contrôles industriels : pour détecter les défauts dans des matériaux, comme le contrôle des composants électroniques.

Imagerie satellite et aérienne par drone : pour identifier les montagnes, les rivières, les déserts, les routes et autres terrains.

Imagerie médicale : pour analyser et détecter les anomalies cancéreuses dans les cellules.

Vision robotique : pour identifier les objets et le terrain et s'y déplacer.

Chapitre 02 :

Deep learning

2.1. Historique et domaine d'application :

Depuis 2012, les algorithmes à base de Deep Learning semblent prêts à résoudre bien des problèmes : reconnaitre des visages comme le propose DeepFace, vaincre des joueurs de go (le jeu de go est un jeu de stratégie, très répandu en extrême-orient, qui se joue à deux et qui consiste à former des territoires en posant des pions, ou pierres, sur un plateau. Le but est de marquer plus de points que l'adversaire en créant plus de territoires que lui et/ou en capturant ses pierres), ou de poker ou bientôt permettre la conduite de voitures autonomes ou encore la recherche de cellules cancéreuses.

Les fondements de ces méthodes ne sont pas si récents :

Le Deep Learning a été formalisé en 2007 à partir des nouvelles architectures de réseaux de neurones dont les précurseurs sont McCulloch et Pitts en 1943 [9]**.** Suivront de nombreux développements comme les réseaux de neurones convolutifsde Yann Le Cun et Yoshua Bengio en 1998 [10], et les réseaux de neurones profonds qui en découlent en 2012 et ouvrent la voie à de nombreux champs d'application comme la vision, le traitement du langage ou la reconnaissance de la parole, ce développement c'est fait récemment ces nouvelles techniques de Machine Learning qui utilise des données massives (big data) ainsi que aux grande capacité de capacités de calcul notamment grâce aux processeurs graphiques.

Dans ce chapitre nous développerons le principe de réseaux neuronaux convolutionnels outil essentiel du Deep Learning et mettrons l'accent sur le principe de l'auto encodeur et décodeur utilisé dans la segmentation sémantique.

2.2. Définition :

Le Deep Learning est un ensemble de méthodes d'apprentissage automatique (Machine Learning) tentant de modéliser avec un haut niveau d'abstraction des données grâce à des architectures articulées de différentes transformations non linéaires .Ces techniques ont permis des progrès importants et rapides dans les domaines de l'analyse du signal sonore ou visuel et notamment de la reconnaissance faciale, de la reconnaissance vocale, de la vision par ordinateur , du traitement automatisé du language.

Le Deep Learning est une méthode de machine Learning qui consiste à enseigner à des ordinateurs ce dont les humains sont naturellement capables

Le Deep Learning apprend à un modèle informatique comment réaliser des tâches de classification directement à partir d'images, de textes ou d'audio. Les modèles de Deep Learning peuvent atteindre un niveau de précision exceptionnel,. L'entraînement des modèles s'effectue via un vaste ensemble de données labellisées et d'architectures de réseaux de neurones qui contiennent de nombreuses couches.

Le Deep Learning est une branche particulière du Machine Learning. Un processus de Machine Learning commence par l'extraction manuelle de caractéristiques pertinentes à partir d'images. En s'appuyant sur ces caractéristiques, un modèle qui catégorise les objets de l'image est ensuite créé. Dans un processus de Deep Learning, l'extraction de caractéristiques pertinentes à partir d'images est automatique. En outre, le Deep Learning effectue un apprentissage « de bout en bout » : à partir de données brutes, un réseau se voit assigner des tâches à accomplir (une classification, par exemple) et apprend comment les automatiser.

Un des avantages majeurs des réseaux de Deep Learning réside dans leur capacité à continuer à s'améliorer en même temps que le volume de vos données augmente.

Figure (2.1) : Comparaison de méthodes de classification de véhicules de Machine Learning (gauche) et de Deep Learning (droite).

Pour classer des images avec le Machine Learning, les choix de caractéristiques et de classificateur doivent être effectués manuellement. Avec le Deep Learning, l'extraction de caractéristiques et le processus de modélisation sont automatiques.

2.3. Types de Deep Learning :

- ***CNN :*** **Les réseaux de neurones convolutionnels**

Les réseaux de neurones convolutionnels sont à ce jour les modèles les plus performants pour classer des images. Désignés par l'acronyme CNN, de l'anglais Convolutional Neural Network, ils comportent deux parties bien distinctes. En entrée, une image est fournie sous la forme d'une matrice de pixels. Elle a 2 dimensions pour une image en niveaux de gris. La couleur est représentée par une troisième dimension, de profondeur 3 pour représenter les couleurs fondamentales [Rouge, Vert, Bleu]. La première partie d'un CNN est la partie convolutive à proprement parler. Elle fonctionne comme un extracteur de caractéristiques des images. Une image est passée à travers une succession de filtres, ou noyaux de convolution, créant de nouvelles images appelées cartes de convolutions. Certains filtres intermédiaires réduisent la résolution de l'image par une opération de maximum local. Au final, les cartes de convolutions sont mises à plat et concaténées en un vecteur de caractéristiques, appelé code CNN [11].

Ce code CNN en sortie de la partie convolutive est ensuite branché en entrée d'une deuxième partie, constituée de couches entièrement connectées (perceptron multicouche). Le rôle de cette partie est de combiner les caractéristiques du code CNN pour classer l'image. La sortie est une dernière couche comportant un neurone par catégorie. Les valeurs numériques obtenues sont généralement normalisées entre 0 et 1, de somme 1, pour produire une distribution de probabilité sur les catégories.

Une architecture CNN est formée par un empilement de couches de traitement indépendantes (figure 2.2) :

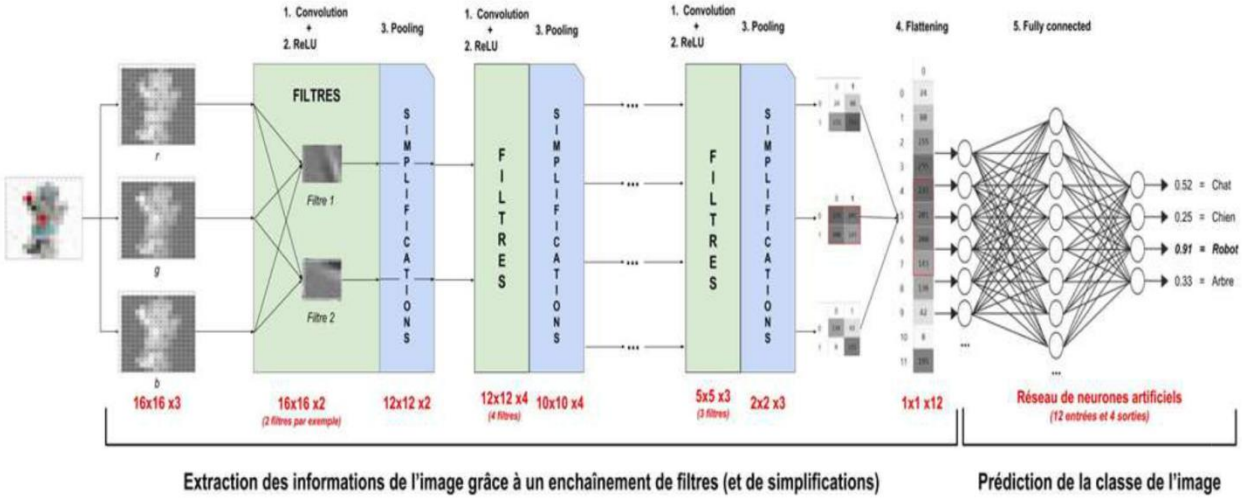

Figure (2 .2) : couches de traitement indépendant

La couche de convolution (CONV) qui traite les données d'un champ récepteur.

La couche de pooling (POOL), qui permet de compresser l'information en réduisant la taille de l'image intermédiaire (souvent par sous-échantillonnage).

La couche de correction (ReLU), souvent appelée par abus 'ReLU' en référence à la fonction d'activation (Unité de rectification linéaire).

La couche "entièrement connectée" (FC), qui est une couche de type perceptron. La couche de perte (LOSS).

A/Couche de convolution(CONV) :

La couche de convolution est le bloc de construction de base d'un CNN. Trois paramètres permettent de dimensionner le volume de la couche de convolution la profondeur, le pas et la marge.

1. Profondeur de la couche : nombre de noyaux de convolution (ou nombre de neurones associés à un même champ récepteur).

2. Le pas : contrôle le chevauchement des champs récepteurs. Plus le pas est petit, plus les champs récepteurs se chevauchent et plus le volume de sortie sera grand.

3. La marge (à 0) ou zéro padding : parfois, il est commode de mettre des zéros à la frontière du volume d'entrée. La taille de ce 'zéro-padding' est le troisième hyper paramètre. Cette marge permet de contrôler la dimension spatiale du volume de sortie. En particulier, il est parfois souhaitable de conserver la même surface que celle du volume d'entrée. Si le pas et la marge appliquée à l'image d'entrée permettent de contrôler le nombre de champs récepteurs à gérer (surface de traitement), la profondeur permet d'avoir une notion de volume de sortie, et de la même manière qu'une image peut avoir un volume, si on prend une profondeur de 3 pour les trois canaux RGB d'une image couleur, la couche de convolution va également

présenter en sortie une profondeur. C'est pour cela que l'on parle plutôt de "volume de sortie" et de "volume d'entrée", car l'entrée d'une couche de convolution peut être soit une image soit la sortie d'une autre couche de convolution. La taille spatiale du volume de sortie peut être calculée en fonction de la taille du volume d'entrée Wi la surface de traitement K (nombre de champs récepteurs), le pas S avec lequel ils sont appliqués, et la taille de la marge P.

Le nombre de neurones du volume de sortie est calculé selon la formule :

$$Wo = \frac{\mathrm{wi}-\mathrm{k}+2\mathrm{P}}{\mathrm{s}} + 1 \quad \text{..... (2.1)}$$

Si W0 n'est pas entier, les neurones périphériques n'auront pas autant d'entrée que les autres. Il faudra donc augmenter la taille de la marge (pour recréer des entrées virtuelles) [11]

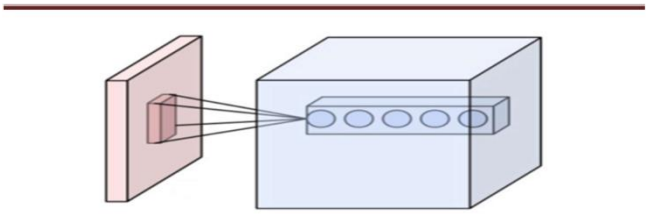

Figure (2.3) : Ensemble de neurones(Cercles) créant la profondeur d'une Couche de convolution (bleu).

Ils sont liés à un même champ récepteur (rouge).

B/**Couche de pooling (POOL) :**

Un autre concept important des CNNs est le pooling, ce qui est une forme de sous-échantillonnage de l'image. L'image d'entrée est découpée en une série de rectangles de n pixels de côté ne se chevauchant pas (pooling). Chaque rectangle peut être vu comme une tuile. Le signal en sortie de tuile est défini en fonction des valeurs prises par les différents pixels de la tuile. Le pooling réduit la taille spatiale d'une image intermédiaire, réduisant ainsi la quantité de paramètres et de calcul dans le réseau. Il est donc fréquent d'insérer périodiquement une couche de pooling entre deux couches convolutives successives d'une architecture CNN pour contrôler l'overfitting (sur-apprentissage).

L'opération de pooling crée aussi une forme d'invariance par translation. La couche de pooling fonctionne indépendamment sur chaque tranche de profondeur de l'entrée et la redimensionne uniquement au niveau de la surface. La forme la plus courante est une couche de mise en commun avec des tuiles de taille 2x2 (largeur/hauteur) et comme valeur de sortie la valeur maximale en entrée. On parle dans ce cas de « Max-Pool 2x2 ».

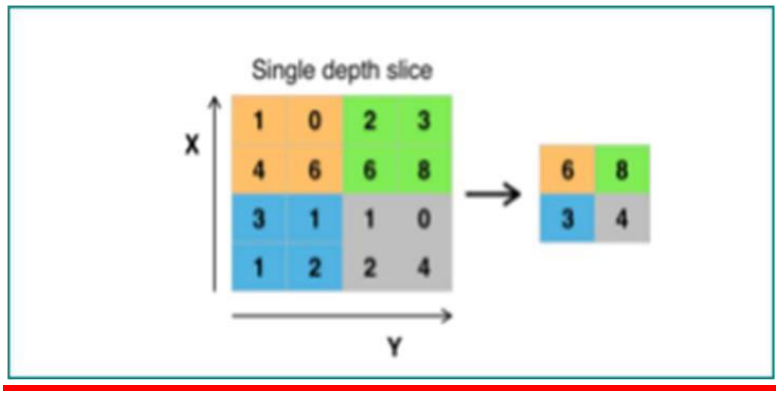

Figure (2.4) : Pooling avec un filtre 2x2 et un pas de 2

Il existe plusieurs types de pooling : (figure 2.4)

- Le « max pooling« qui revient à prendre la valeur maximale de la sélection. C'est le type le plus utilisé car il est rapide à calculer (immédiat), et permet de simplifier efficacement l'image
- Le « meanpooling » (ou averagepooling), soit la moyenne des pixels de la sélection : on calcule la somme de toutes les valeurs et on divise par le nombre de valeurs. On obtient ainsi une valeur intermédiaire pour représenter ce lot de pixels
- Le « sumpooling « c'est la moyenne sans avoir divisé par le nombre de valeurs (on ne calcule que leur somme) [12].

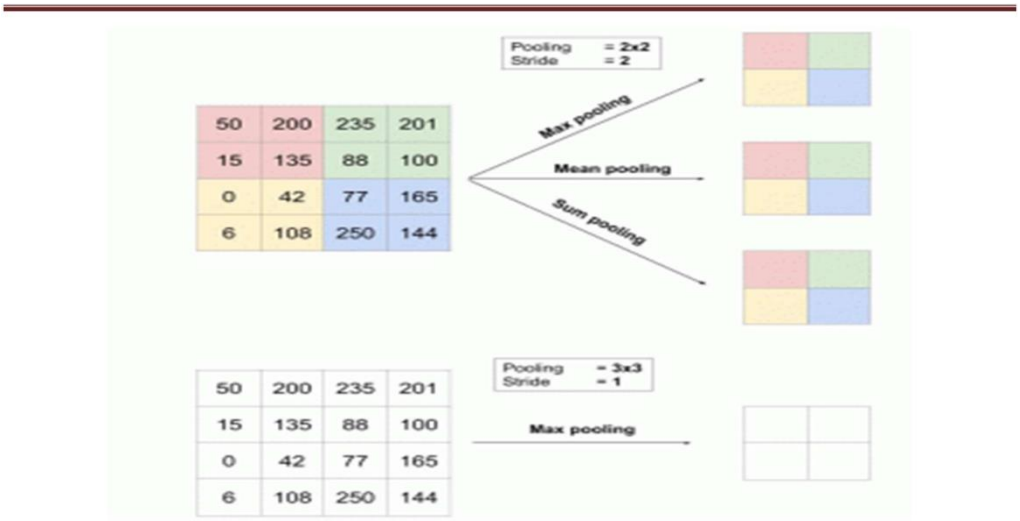

Figure (2.5) : Calcul du pooling sur une image 4×4. Un pooling de 2×2 signifie que l'on sélectionne les pixels en carrés de 2×2.

Le stride indique de combien de cases décaler le carré à chaque fois.

Initialement l'average pooling était souvent utilisé, mais il s'est avéré que le max-pooling est plus efficace car celui-ci augmente plus significativement l'importance des activations fortes. En d'autres circonstances, on pourra utiliser un pooling stochastique. Le pooling permet de gros gains en puissance de calcul. Cependant, en raison de la réduction agressive de la taille de la représentation (et donc de la perte d'information associée), la tendance actuelle est d'utiliser de petits filtres (type 2x2). Il est aussi possible d'éviter la couche de pooling mais cela implique un risque sur-apprentissage plus important.

C/Couches de correction (ReLU) :

Il est possible d'améliorer l'efficacité du traitement en intercalant entre les couches de traitement une couche qui va opérer une fonction mathématique (fonction d'activation) surles signaux de sortie.

La fonction ReLU (abréviation de Unités Rectifié linéaires) est définie comme suit :

F(x)=max (0, x). Cette fonction force les neurones à retourner des valeurs positives [13].

D/Couche entièrement connectée (FC) :

Après plusieurs couches de convolution et de max-pooling, le raisonnement de haut niveau dans le réseau neuronal se fait via des couches entièrement connectées. Les neurones dans une couche entièrement connectée ont des connexions vers toutes les sorties de la couche précédente. Leurs fonctions d'activations peuvent donc être calculées avec une multiplication matricielle suivie d'un décalage de polarisation [12].

E/ Couche de perte (LOSS) :

La couche de perte spécifie comment l'entrainement du réseau pénalise l'écart entre le signal prévu et réel. Elle est normalement la dernière couche dans le réseau. Diverses fonctions de perte adaptées à différentes tâches peuvent y être utilisées. La fonction « Softmax » permet de calculer la distribution de probabilités sur les classes de sortie [12].

- ***AUTO ENCODEUR :***

Au meilleur de notre connaissance, l'utilisation des couches de sur-échantillonage a été d'abord présentée par Huang *et al.* (2007) dans une architecture d'apprentissage non supervisée. Un réseau de déconvolution a également été introduit par Zeiler*et al.* (2011) pour reconstruire les images d'entrée à partir de sa représentation fonctionnelle. Zeiler&Fergus (2014) ont utilisé des couches de déconvolution dans leur travail pour comprendre le comportement des réseaux de neurones convolutionnels en visualisant des cartes de caractéristiques de différents niveaux.

Afin de mieux récupérer les pertes d'informations causées par les couches de sous-échantillonnage, les chercheurs ont poussé l'idée proposée dans le travail de Long *et al.* (2015) un peu plus loin en élargissant la partie de sur-échantillonnage des cartes de caractéristiques et en ajoutant davantage de liaisons de sauts entre les couches de différents niveaux. Ce type d'architecture est connu sous le nom de réseau encodeur-décodeur.

La première partie de l'architecture du réseau est appelée encodeur. Elle est similaire à la partie de convolution d'un réseau de classification. Des blocs de couches de convolution suivis chacun d'une couche de sous-échantillonnage sont utilisés pour convertir l'image d'entrée en cartes de caractéristiques.

La deuxième partie est le décodeur qui est une version miroir de l'encodeur. Elle comprend plusieurs ensembles de couches de déconvolution et de sur-échantillonnage pour récupérer les informations spatiales à partir de la sortie de l'encodeur. La dernière couche du décodeur est une couche de classification softmax qui produit une segmentation finale de la même taille que l'image d'entrée.

Il existe généralement des connexions d'encodeur vers décodeur pour aider le décodeur à réduire la perte d'informations et à mieux récupérer les détails de l'objet d'intérêt. De nombreuses architectures ont été développées sur la base du concept encodeur-décodeur (Badri-narayanan*et al.* (2015) ; Nohet *al.* (2015) ; Ronnebergeret *al.* (2015) ; Badrinarayananet *al.* (2017)). Ce qui différencie une architecture des autres est la façon dont elle relie les différents niveaux de l'encodeur avec leurs parties inversées dans le décodeur.

Badrinarayanan*et al.* (2017) ont développé une architecture de forme encodeur-décodeur, connue sous le nom SegNet, pour la segmentation multi-classes des pixels. La partie encodeur de SegNet est similaire aux 13 couches convolutivesdu réseau VGG-16 (figure 2.6). Elle est composée de cinq blocs. Chaque bloc est constitué de couches de convolution 2D, de normalisation par lots et d'une activation d'unité rectifiée linéaire (ReLU). La dernière couche du bloc est une couche de pool maximum qui vise à réduire la dimension spatiale des cartes de caractéristiques. Pour récupérer la haute résolution des cartes de caractéristiques, SegNet a utilisé un processus inverse de la partie encodeur en tant que décodeur avec le même nombre de blocs. Chaque bloc est constitué d'une couche de sur-échantillonnage suivie par des couches convolutives entraînées, normalisation par lots et d'une activation (ReLU). Un élément clé de l'architecture SegNet est l'utilisation des emplacements restaurés de pool maximum pour effectuer un sur-échantillonnage dans la partie décodeur. Ce processus permet de conserver des détails haute fréquence dans les images segmentées avec un faible coût de consommation de mémoire et un nombre réduit de paramètres d'apprentissage dans la phase de décodeur.

Une autre architecture de type encodeur-décodeur, connue sous le nom UNet, a été proposé par Ronnebergeret *al.* (2015). UNet a montré d'excellentes performances en segmentant différentes cibles dans différentes modalités d'images notamment médicales. L'architecture est composée de 23 couches de convolution au total. Elle consiste en un chemin de contraction (encodeur) et un chemin d'expansion (décodeur). Comme SegNet, la partie encodeur est composée de blocs de convolution répétés. Chaque bloc est constitué de deux couches de convolutions avec des filtres de taille (3×3), chacune suivie d'une activation (ReLU) et d'une opération de pool maximum de (2×2). Le nombre de cartes de caractéristiques est doublé après chaque sous-échantillonnage. La partie d'expansion d'UNet est une version inversée de la partie contraction. Le nombre de cartes de caractéristiques est réduit de moitié après chaque bloc. Pour connecter la partie codeur à la partie décodeur correspondante, une copie des cartes de caractéristiques est concaténée avec les cartes correspondantes de la partie décodeur. La dernière couche utilise une convolution avec des filtres de tailles (1×1) pour fournir des cartes de classification de même nombre que les classes

souhaitées. La taille de sortie est inférieure à la taille d'entrée (voir figure 2.7) en raison de l'utilisation de la convolution sans l'ajout de pixels autour de l'image. Pour obtenir la même taille que l'entrée, l'image entière est prédite partie par partie à l'aide d'une stratégie de mosaïque de chevauchement.

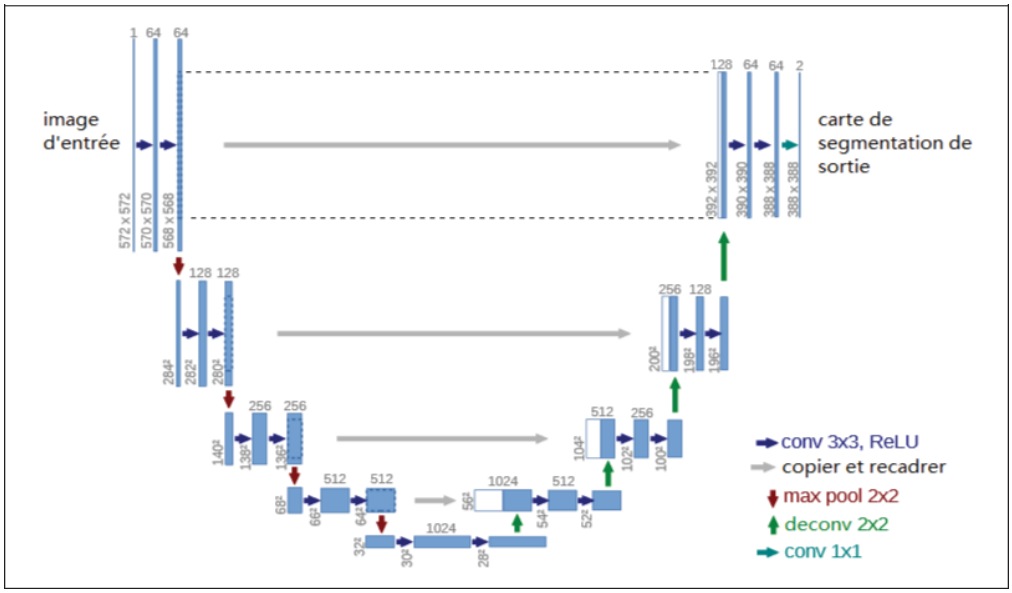

Figure (2.6) : Architecture U-net. Chaque rectangle bleu correspond à une carte de caractéristiques multicanaux. Le nombre de canaux est indiqué en haut du rectangle.

La taille spatiale est indiquée sur le bord inférieur gauche du rectangle. Les rectangles blancs représentent les cartes de caractéristiques copiées. Les flèches indiquent les différentes opérations. Figure reproduite et adaptée avec l'autorisation de (Ronnebergeret al. (2015)).

Le nombre d'échantillons annotés est généralement limité car cela devrait être fait par des experts. D'autre part, les réseaux d'apprentissage dans un tel domaine nécessitent un grand nombre d'échantillons. Pour surmonter cette limite, Ronneberger*etal.* (2015) ont augmenté la taille de l'ensemble d'apprentissage en utilisant des transformations aléatoires sur les images d'entrée telles que les inversions, les distorsions, les rotations et en particulier les déformations élastique.

Chapitre 03 :

Travail personnel

3.1. Introduction :

Pour un grand nombre d'applications tant civiles que militaires, les routes sont une partie essentielle de l'infrastructure urbaine. Par conséquent, leur détection et la modélisation représentent une étape importante sur le plan sémantique

Ce travail établit une stratégie quant à l'utilisation d'un réseau de neurone convolutif pour la classification d'images uav à haute résolution spatiale,

Dans une première partie nous avons conçu notre propre réseau basique d'encodeur –décodeur avec crash c'est dire sans utilisation d'un modèle pré-entrainé puis dans une deuxième étape nous avons utilisé le modèle encodeur –décodeur Segnet avec un préapprentissage sur VGG16. Nous évaluerons et comparerons les résultats des deux réseaux.

3.2. Base de données :

Pour évaluer un réseau de segmentation sémantique, nous avons besoin d'une collection d'images et de la collection correspondante d'images étiquetées au pixel près (groundtruth). Une image étiquetée par pixel est une image où chaque valeur de pixel représente l'étiquette ou label de ce pixel.

Nous avons testé nos deux réseaux sur une base de données collectée sur internet. Les images collectées par drone, se composent d'images semi urbaines et non urbaines. Elles sont composées de plusieurs routes bien pavées et non pavées, de différentes résolutions, largeurs et formes. Les images de la route présentent également des variations de couleur, d'éclairage et de contraste. L'ensemble de données peut être téléchargé de https://sites.google.com/site/hailingzhouwei/.

100 images aériennes sont utilisées pour l'apprentissage et 12 pour le test. (soit environ10%).

3.3. Indices de scores de la segmentation :

Des évaluations quantitatives sont calculées en comparant les résultats avec la vérité de terrain au pixel près. Nous nous référons à la vérité de terrain comme les zones routières qui sont annotées manuellement dans chaque image test. **[14]**

Des mesures quantitatives sont introduites .Les indices sont exprimés en termes de vrais positifs (TP), faux positifs (FP) et faux négatifs (FN) et (TN) vrais négatifs définis comme suit :

	Valeur observée		
valeur prédite		"positif"	"négatif"
	"positif"	TP	FP
	"négatif"	FN	TN

La valeur observée représente le résultat binaire de notre segmentation (road, nonroad) et la valeur prédite représente l'image vérité terrain ou groundtruth.

On peut alors définir toute une batterie d'indicateurs permettant de juger de la qualité de notre prédicteur (ou plutôt de notre score).

- Accuracy ou taux d'erreur = (TP+TN)/ (FN+FP+TP+TN) ; ……(3,1)
- Precision = TP/ (TP+FP) ; ……..…… (3,2)

le taux de positifs prédits (*positive predictive value*)

- Fmeasure = 2*TP/ (2*TP+FP+FN) ; ……..... (3,3)

Le score de correspondance des contours de la frontière F1 (BF) indique dans quelle mesure la frontière prédite de chaque classe s'aligne sur la frontière réelle.

- Specitivity (spécificité**)** = TN/ (TN+FP) ;……..... (3,4)

Taux de vrais négatifs (*TrueNegative Rate*)

- Sensitivity (la sensibilité)= TP/ (TP+FN) ; …… (3,5)

Correspondant au taux de vrais positifs (*true positive rate*)

- MCC (Matthews corrélation coefficient)=

(TP*TN-FP*FN)/sqrt ((TP+FP)*(TP+FN)*(TN+FP)*(TN+FN)) ; ……...(3,6)

- Jaccard Similarity Coefficient =TP / (TP + FP + FN) ; ……. (3,7)

Jaccard Similarity Coefficient : L'intersection sur l'union (IoU), également connue sous le nom de coefficient de similarité Jaccard, est la métrique la plus utilisée. Pour chaque classe, l'IoU est le rapport entre les pixels correctement classés et le nombre total de pixels réels et prévus dans cette classe.

- Dice = 2*TP/ (2*TP+FP+FN) ; …….……(3,8)

3.4. Création d'un réseau encodeur décodeur de segmentation sémantique :

Un modèle commun et basique dans la segmentation sémantique.

Nécessite le sous-échantillonnage d'une image entre la convolution et ReLU

Puis sur échantillonner la sortie pour qu'elle corresponde à la taille de l'entrée.

La (Figure (3.1)) illustre l'architecture du réseau utilisé avec ses différentes couches de convolution. **[15]**

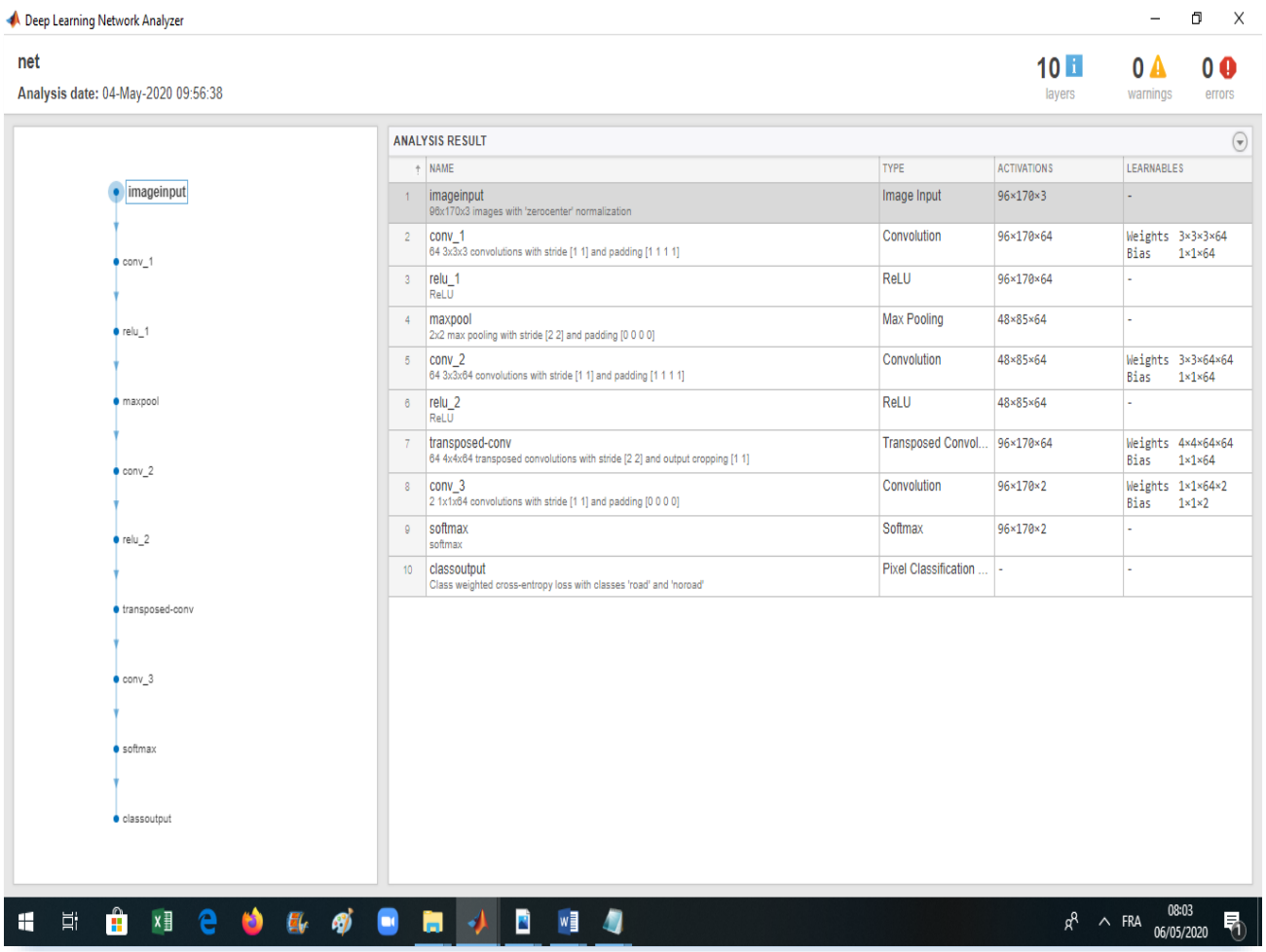

Figure (3.1) : Architecture du Modèle1encodeur –décodeur

Les couches détaillées du modèle sont définies comme suit :

1. **Convolution et ReLU** :

Commençons par les couches de convolution et de ReLU, nous avons appliqué deux convolutions avec des pas de 1. Ensuite, Toutes les convolutions sont suivies d'un ReLU pour introduire la non-linéarité.

Le remplissage de la couche de convolution est sélectionné de telle sorte que la taille de sortie de la couche de convolution soit la même que la taille d'entrée. Cela facilite la construction d'un réseau [14].

2. Réseau de sous échantillonnage (downsampling/ pooling)

Le sous-échantillonnage est effectué à l'aide d'une couche de regroupement maximale.

Nous avons créé une couche de regroupement maximale pour échantillonner l'entrée par un facteur de 2 en définissant le paramètre « Stride » égale à 2.

Nous avons empilé les couches de convolution, ReLU et Max pour créer un réseau qui minimise son entrée par un facteur de 2.

3. Réseau de sur échantillonnage (upsampling)

L'échantillonnage est effectué à l'aide de la couche de convolution transposée appelée couche «deconv».

Nous avons créé une couche de convolution transposée avec un pas de sur échantillonnage de 2.

Le paramètre recadrage est défini avec une valeur de 1 pour rendre la taille de sortie égale à la taille d'entrée.

Nous avons empilé les couches de convolution et de Relu transposés. Une entrée de cet ensemble de calques est sur échantillonnée avec un facteur de 2.

L'ensemble final des calques est responsable de la classification des pixels. Ces couches finales traitent une entrée qui a les mêmes dimensions spatiales que l'image d'entrée.

Nous avons ensuite créé une couche de convolution pour combiner la troisième dimension de l'entité en entrée qui correspond au nombre de classes.

Après cette couche de convolution 1-par-1 nous trouvons les couches de classification de SoftMax et de pixel. Ces deux calques se combinent pour prédire l'étiquette catégorielle pour chaque pixel d'image.

A présent, ce réseau est prêt à être formé à l'aide de train Network du Deep Learning Toolbox de MATLAB 2018.

Résultat et interprétation :

1. L'entrainement du réseau :

Les expériences se sont déroulées sur un ordinateur doté d'un processeur Intel® Core™_i3 4600 CPU et d'un processeur de (2,70 GHz) et une RAM de (4Go).

Le tableau nous montre le rapport complet d'entrainement du réseau de neurones et nous donnent les informations relatives à son fonctionnement en fonction du temps en précisant le pourcentage de perte et de précision et les nombres d'itérations effectuées.

Training on single CPU.

Initializing image normalization.

Epoch	Iteration	Time Elapsed (hh:mm:ss)	Mini-batch Accuracy	Mini-batch Loss	Base Learning Rate
1	1	00:00:07	61.35%	0.6931	0.0010
17	50	00:05:28	94.26%	0.6922	0.0010
34	100	00:11:31	94.19%	0.6912	0.0010
50	150	00:17:15	94.37%	0.6879	0.0010
67	200	00:22:33	94.22%	0.6822	0.0010
84	250	00:27:46	94.18%	0.6679	0.0010
100	300	00:33:00	92.75%	0.6328	0.0010
117	350	00:38:12	86.06%	0.6171	0.0010
134	400	00:43:28	80.43%	0.5920	0.0010
150	450	00:48:41	73.91%	0.5503	0.0010
167	500	00:53:49	74.20%	0.5438	0.0010
184	550	00:58:41	75.99%	0.5483	0.0010
200	600	01:03:40	74.50%	0.5018	0.0010

Tableau (3.1) : Le rapport d'entrainement

Le résultat de l'apprentissage sur notre base d'images dotée de 100 images du réseau a pris environ 1h heures et 03 minutes parce que notre PC n'est pas doté
de GPUs qui peuvent réduire considérablement le temps de traitement nécessaire pour entraînerun modèle .le pourcentage de précision augmente jusqu'à 74.5%.

Le test :

Après l'entraînement du réseau de neurone une phase de test est recommandée pour confirmer sice dernier fonctionne correctement ou qu'il n'a pas bien appris les informations et les données liées à la route. Rappelons que 12 images ont été réservées pour le test choisies aléatoirement.

Un tableau des scores résume les scores obtenus :

paramètres / images	Accuracy	Sensitivity	F mesure	Precision	MMC	Dice	Jaccard	Specitivity
6	0.1091	0.0631	0.0094	0.0051	-0.5517	0.0094	0.0047	0.1123
9	0.1126	0.0445	0.0062	0.0033	-0.5347	0.0062	0.0031	0.1169
18	0.1046	0.0151	0.0022	0.0012	-0.5769	0.0022	0.0011	0.1108
25	0.1094	0.0202	0.0027	0.0015	-0.5528	0.0027	0.0014	0.1151
30	0.1138	0.0259	0.0033	0.0018	-0.5295	0.0033	0.0017	0.1191
46	0.1187	0.0808	0.0106	0.0057	-0.5010	0.0106	0.0053	0.1210
65	0.1223	0.1081	0.0140	0.0075	-0.4804	0.0140	0.0071	0.1232
78	0.1512	0.3452	0.4015	0.0221	-0.3093	0.0415	0.0212	0.1402
83	0.1567	0.3986	0.0484	0.0238	-0.2769	0.0484	0.0248	0.1430
97	0.1926	0.3764	0.0571	0.0309	-0.2691	0.0571	0.0294	0.1798
106	0.1697	0.2242	0.0394	0.0216	-0.3961	0.0394	0.0201	0.1652
112	0.1514	0.1625	0.0302	0.0167	-0.4643	0.0302	0.0154	0.1504

Tableau (3.2) : les scores pour les 12 images tests

Afin d'illustrer les résultats de la segmentation nous avons illustré à titre d'exemple pour 4 images tests toutes les étapes de la classification.

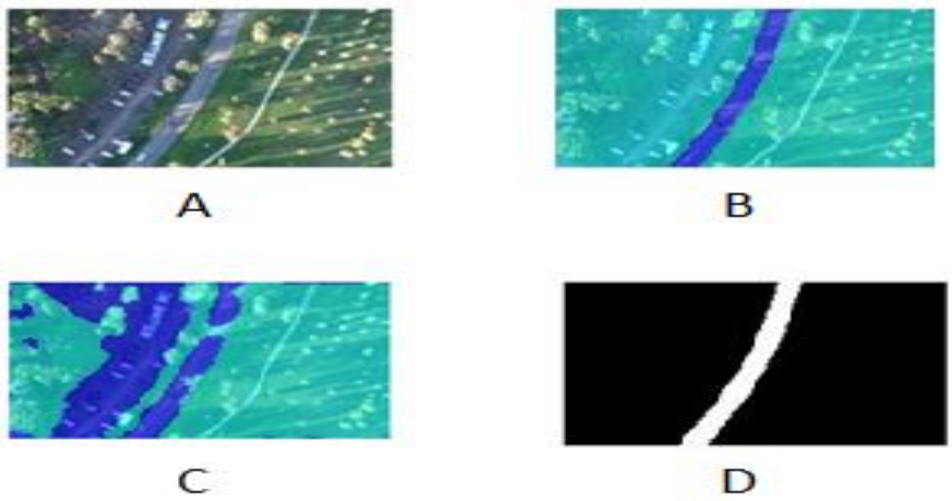

Figure (3.2) : image 6 (A, B, C, D)

Figure Image 6 :

A. : image couleur originale
B. :image originale juxtaposée avec le groundtruth en cyan
C. : image resultat apres segmentation
D. : image vérité ou grountruth

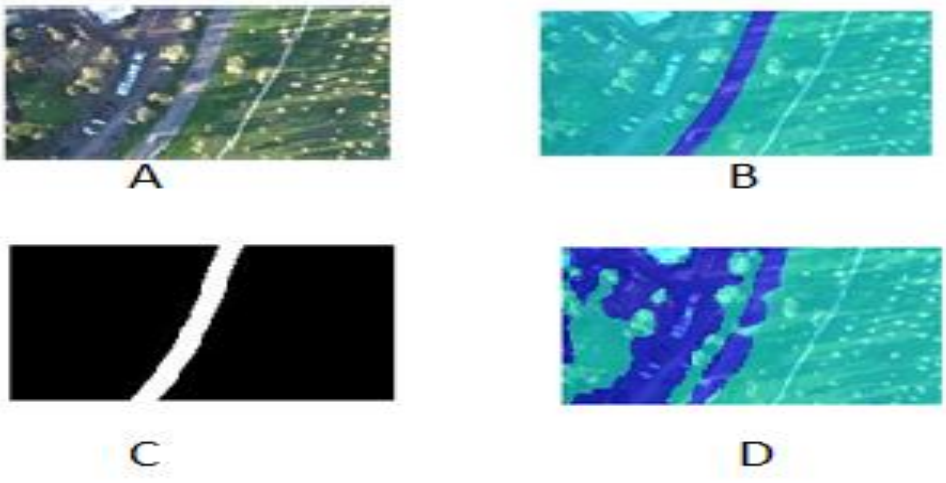

Figure (3.3) : image 18 (A, B, C, D)

Figure Image 18 :

A. : image couleur originale
B. :image originale juxtaposée avec le groundtruth en cyan
C. : image resultat apres segmentation
D. : image vérité ou grountruth

Figure (3.4) : image 65 (A, B, C,D)

Figure Image 65 :

A. : image couleur originale
B. :image originale juxtaposée avec le groundtruth en cyan
C. : image resultat apres segmentation
D. : image vérité ou grountruth

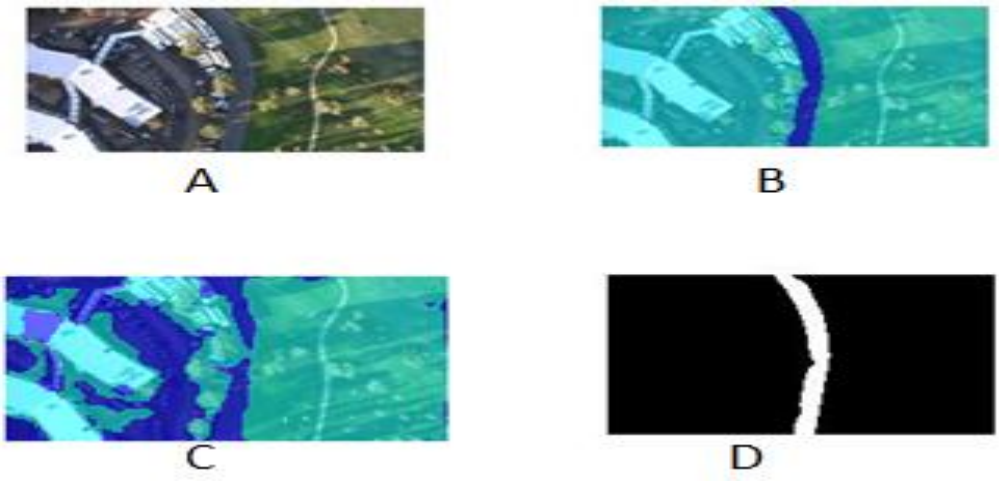

Figure (3.5) : image 78 (A, B, C, D)

Figure Image 78 :

A. : image couleur originale
B. :image originale juxtaposée avec le groundtruth en cyan
C. : image resultat apres segmentation
D. : image vérité ou grountruth

3.5.Segmentation sémantique par un réseau pré entrainé pour l'apprentissage par transfert :

De nombreux travaux récents ont montré l'efficacité des réseaux de neurones entièrement convolutifs .dans le cadre de la segmentation sémantique. SegNet présente une architecture encodeur-décodeur conçue sur la base des couches de convolution du modèleVGG-16 Le réseau de neurones convolutionnels que nous avons choisis est le SegNet combiné avec VGG-16.

3.5.1. Définition de VGG-16

VGGNet se compose de 16 couches convolutives et est très attrayant en raison de son architecture très uniforme. Similaire à AlexNet, seulement 3x3 convolutions, mais beaucoup de filtres. Formé sur 4 GPU pendant 2 à 3 semaines. Il s'agit actuellement du choix le plus populaire dans la communauté pour l'extraction de fonctionnalités à partir d'images. La configuration de poids du VGGN et est disponible publiquement et a été utilisée dans de nombreuses autres applications et défis en tant qu'extracteur de fonctionnalités de base (Figure (3.3))

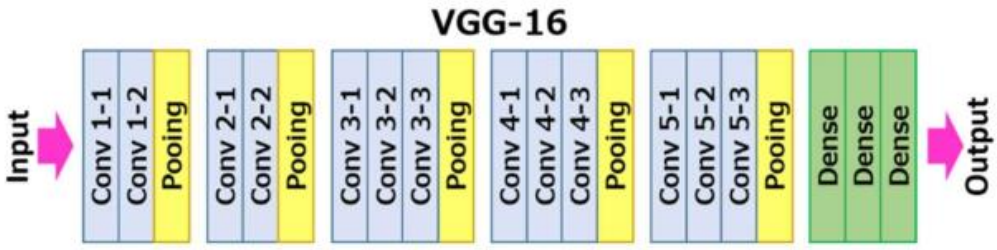

Figure 3.6 : Architecture de VGG-16

3.5.2. Définition de SegNet :

Il possède une architecture de réseau neuronal totalement convolutionelle profonde et novatrice pour la segmentation sémantique par pixels. Ce moteur de segmentation comprend un réseau décodeur qui est typologiquement identique aux 13 couches convolutives du réseau VGG16, un réseau de décodeurs

correspondant, suivi d'une couche de classification en pixels. Le réseau de décodage a pour rôle de mapper les mappages de caractéristiques de codeurs basse résolution avec les mappages de caractéristiques de résolution d'entrée totale pour une classification par pixel.

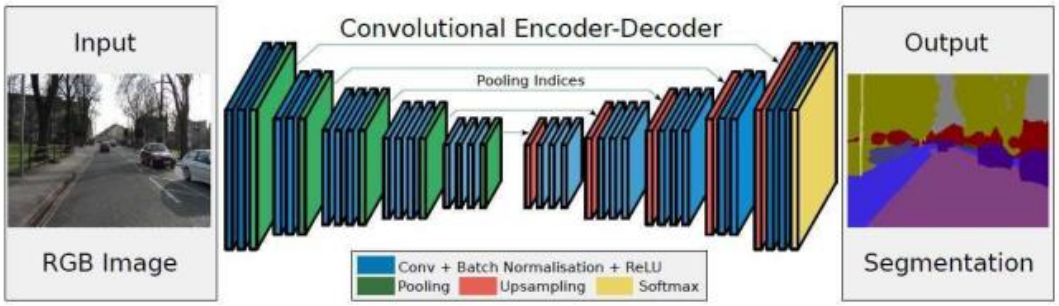

Figure (3.7) :L'architecture de segmentation par SegNet

- **Résultats :**

Les mêmes étapes seront réalisées pour le deuxième modèle donc création du modèle, apprentissage et test. La figure illustre l'architecture du réseau utilisé avec ses différentes couches de convolution.

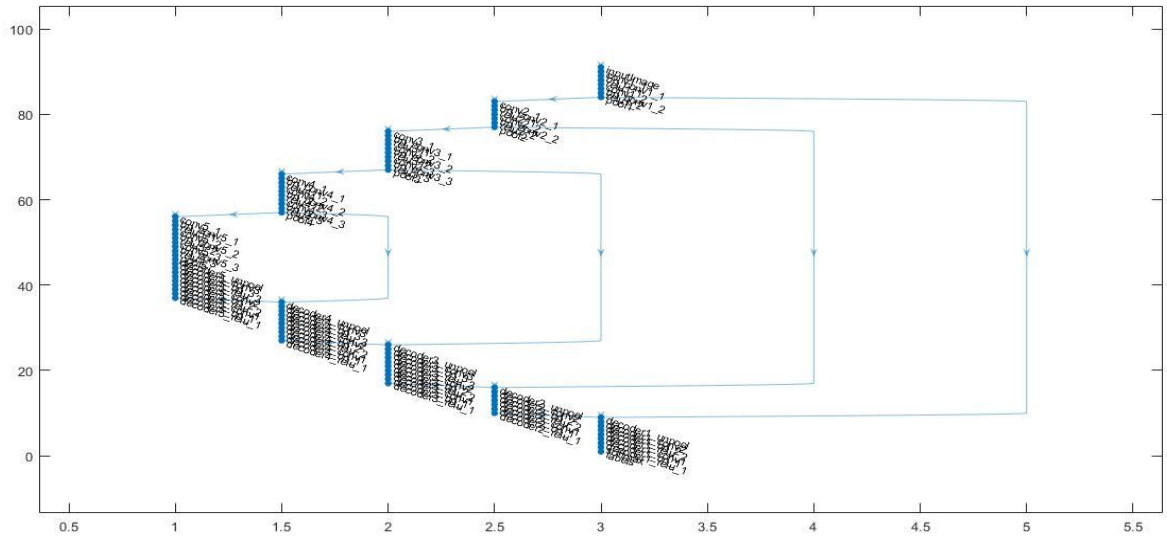

Figure (3.8) : Architecture du réseau utilisé.

- **Explication** :

SegNet présente une architecture encodeur-décodeur conçue sur la base des couches de convolution du modèle VGG-16 [6, 30]. L'encodeur est une succession de couches convolutives suivies par une

normalisation par batch et des fonctions de transfert non linéaires. Chaque bloc de 2 ou 3 convolutions est suivi par une couche de sous échantillonnage de pas égal à 2.

Le décodeur est une symétrie de l'encodeur et possède le même nombre de convolutions et le même nombre de blocs. Les réductions de dimensions sont remplacées par des sur-échantillonnages. Ceux-ci replacent les valeurs des activations intermédiaires aux indices ("argmax") calculés lors du sous-échantillonnage.

L'encodeur étant calqué sur VGG-16, ses poids sont initialisés à partir de ce même CNN pré-entraîné sur le jeu de network données ImageNet.

Le résultat de l'apprentissage sur notre base d'images du réseau a pris environ 8h heures et 30 minutes parce que notre PC n'est pas doté de GPUset le modèle segnet __VGG16 est beaucoup plus volumineux que le premier modèle. Le pourcentage de précision augmente jusqu'à 84.5%. .

Après l'entraînement du réseau de neurone une phase de test est recommandée pour confirmer si ce dernier fonctionne correctement ou qu'il n'a pas bien appris les informations et les données liées à la route. Rappelons que les mêmes 12 images ont été réservées pour le test.

Un tableau des scores résume les scores obtenus :

paramètres / images	Accuracy	Sensitivity	F mesure	Precision	MMC	Dice	Jaccard	Specitivity
6	0.3328	0.133	0.0228	0.0125	-0.3481	0.0228	0.0115	0.2359
9	0.2432	0.1073	0.0169	0.0092	-0.3365	0.0169	0.0085	0.2520
18	0.3746	0.1602	0.0279	0.0153	-0.2948	0.0279	0.0142	0.2826
25	0.2939	0.1171	0.0197	0.0108	-0.2911	0.0197	0.0100	0.3054
30	0.2993	0.0703	0.0118	0.0064	-0.2981	0.0118	0.0059	0.3128
46	0.3070	0.0734	0.0117	0.0046	-0.2979	0.0117	0.0059	0.3217
65	0.2939	0.0805	0.0130	0.0071	-0.3013	0.0130	0.0069	0.3070
78	0.2822	0.2336	0.0318	0.0171	-0.2647	0.0318	0.0161	0.2426
83	0.3233	0.1948	0.0263	0.0141	-0.2999	0.0263	0.0133	0.2250
97	0.4856	0.8922	0.2731	0.1612	0.0674	0.2731	0.1582	0.1782
106	0.3820	0.8569	0.2767	0.1650	0.0287	0.2767	0.1606	0.1723
112	0.2890	0.8335	0.2990	0.1822	0.0015	0.2990	0.1758	0.1679

Tableau (3.3) : Les résultats pour les 12 images avec VGG16

Afin d'illustrer les résultats de la segmentation nous avons illustré pourles mêmes 4 images tests toutes les étapes de la classification .ceci à titre de comparaison avec le premier modèle. Nous avons ajouté le résultat de la segmentation du reseau1.

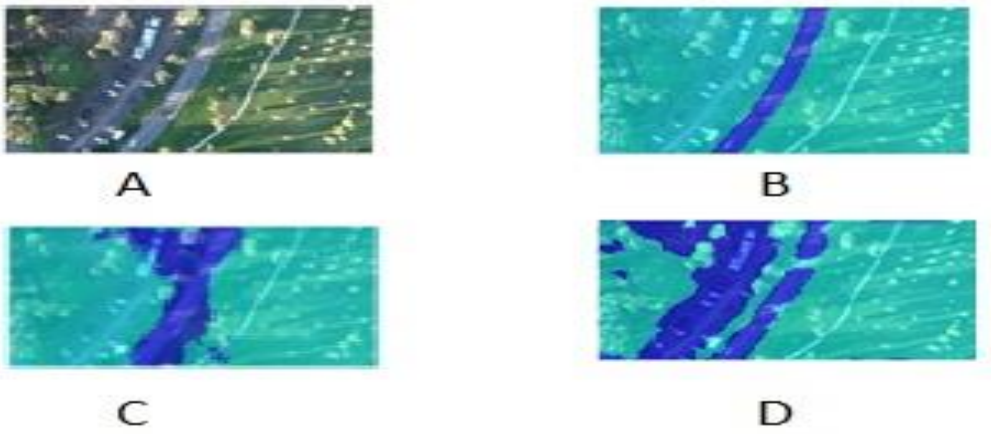

Figure (3.9) : image 6 (A, B, C, D**)**

Figure Image 6 :

A. : image couleur originale
B. : image originale juxtaposée avec le groundtruth en cyan
C. : image résultat après segmentation avec le réseau 2 segnet
D. : image résultat après segmentation avec le réseau 1

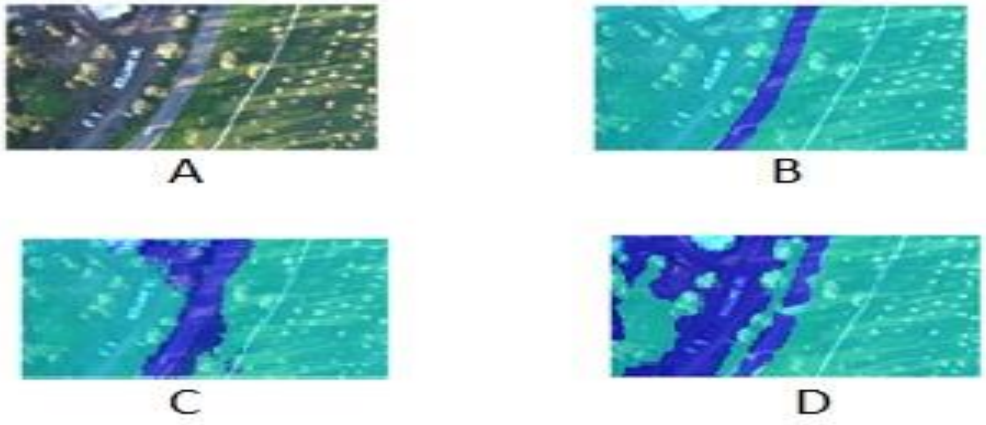

Figure (3.10) : image 18 (A, B, C, D)

Figure Image 18 :

A. image couleur originale
B. image originale juxtaposée avec le groundtruth en cyan
C. image résultat après segmentation avec le réseau 2 segnet
D. image résultat après segmentation avec le réseau 1

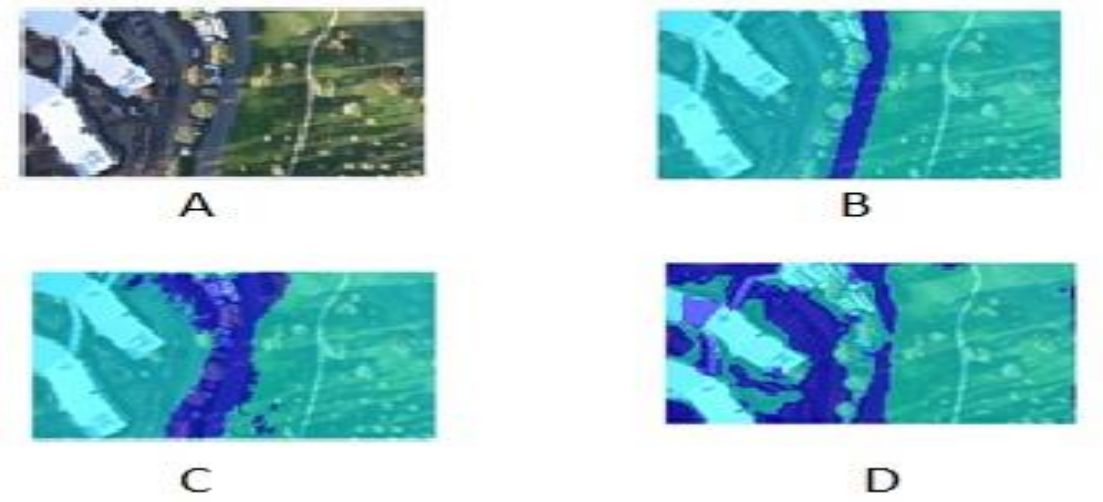

Figure (3.11) : image 65 (A,B ,C,D)

Figure Image 65 :

A. image couleur originale
B. image originale juxtaposée avec le groundtruth en cyan
C. image résultat après segmentation avec le réseau 2 segnet
D. image résultat après segmentation avec le réseau 1

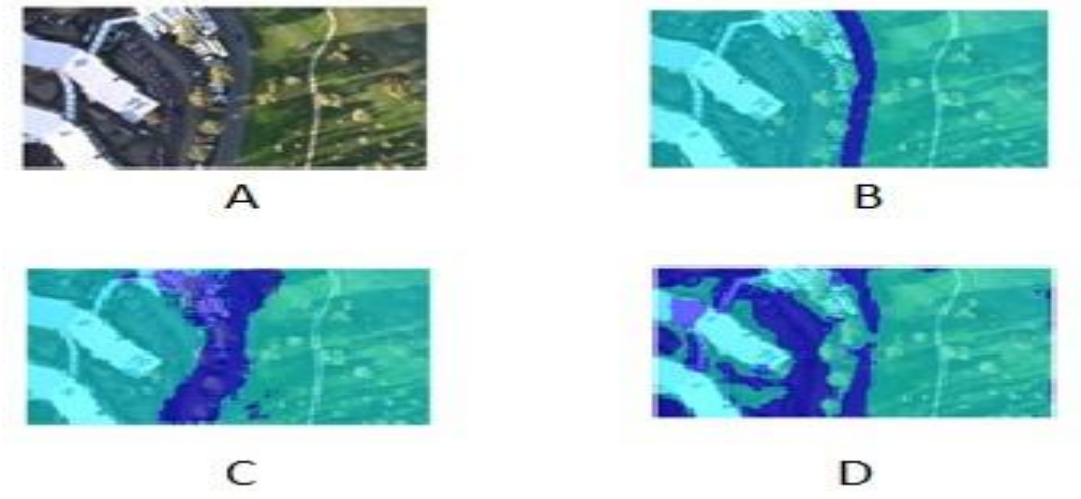

Figure (3.12) : image 78 (A,B ,C,D)

Figure Image 78 :

A. image couleur originale
B. image originale juxtaposée avec le groundtruth en cyan
C. image résultat après segmentation avec le réseau 2 segnet
D. image résultat après segmentation avec le réseau 1

En comparant les résultats des de réseaux, nous remarquons que La classe attribuée à la route par le réseau couvre la classe réelle avec une meilleure précision pour le modèle Segnet.

Pour affiner ce résultat il faut intervenir à différents stades :

_ augmenter la base d'apprentissage qui est très petite.

_Régler les paramètres un à un (nombre de convolution/déconvolution, pooling…), choix de la fonction d'activation, réglage du dropout

_ refaire Les expériences sur un ordinateur doté d'un processeur plus puissant et dotée GPUs afin de réduire considérablement le temps de traitement nécessaire pour entraîner le modèle.

Conclusion générale :

Dans ce projet nous avons discuté des notions fondamentales de segmentation d'image et du Deep Learning basé sur les réseaux de neurones convolutionnels en particulier. Nous avons introduit ces réseaux de neurones convolutionnels en présentant les différents types découches utilisées dans la classification.

Nous avons exploité et fait fonctionner deux modelés de segmentation sémantique sur Matlab 2018. Dans la phase d'implémentation, l'utilisation d'un CPU ne fait que le temps d'exécution soit trop couteux. Les résultats sont obtenus à la phase de test, ce qui permet d'afficher des images segmentées grâce à l'étape de déconvolution et sur-échantillonnage. Les résultats ne sont pas d'une grande précision, mais ils peuvent être affinés lors de travaux futurs.

En perspective Pour affiner ce résultat il faut intervenir à différents stades :

_ augmenter la base d'apprentissage qui est très petite Pour réaliser sa propre segmentation sémantique, il est nécessaire d'utiliser sa propre base étiquetée, Ceci est un travail fastidieux mais qui est nécessaire dans le cas d'un domaine d'utilisation bien défini au préalable..

_Régler les paramètres un à un (nombre de convolution/déconvolution, pooling…), choix de la fonction d'activation, réglage du dropout

_ refaire Les expériences sur un ordinateur doté d'un processeur plus puissant et dotée de GPUs afin de réduire considérablement le temps de traitement nécessaire pour entraîner le modèle.

Bibliographie :

[1] :S. Ameur et Z. Ameur, " Revue des approches de segmentation d'images textures:exemple des images météorologiques", 3rd International Conférence : Science of Electronic,Technologies of Information and Telecommunication, Tunisia, 2005.

[2] : J.-P.COCQUEREZ et S.PHILIPP, "Analyse d'images : filtrage et segmentation", Paris, 2003.

[3] : J. J. ROUSSELLE, "Les contours actifs une méthode de segmentation : Application àl'imagerie médicale", Thèse doctorat, juillet 2003.

[4] : A. Mekhmoukh, "Segmentation d'images IRM par améliorations de l'algorithmeFCM", thèse doctorat, Université de A.mira-Béjaia le 30/01/2016.

[5] : A.N.Benaichouche, "Conception de métaheuristiques d'optimisation pour la segmentation d'images. Application aux images IRM du cerveau et aux images de Tomographie par Émission de positons ", thèse de doctorat université paris 12, 2012.

[6] : Lotfi A Zadeh. "Fuzzy sets". Information and control, 8(3) : 338–353, 1965.

[7] : Mlle Hadjer LAGUEL, « Déploiement sur une plateforme de visualisation, d'unalgorithme coopératif pour la segmentation d'images IRM basé sur les systèmes multiagents», Projet de Fin d'Étude pour l'obtention du diplôme d'ingénieur d'état de l'universitédes Sciences et de la Technologie Houari Boumediene, 12 octobre 2010.

[8] : MoualhiMouloud, SehakiMenad, « Segmentation 3D des structures internes du cerveauet des tumeurs cérébrales dans des IRM-3D de têtes d'individus»,Mémoire de fin d'étudesPour l'obtention du diplôme d'Ingénieur d'Etat en Informatique, 22 Juin 2014 .

[9] :Warren McCulloch& Walter Pitts, *A LogicalCalculus of Ideas Immanent in NervousActivity*, 1943, Bulletin of MathematicalBiophysics 5:115-133.

[10] :LeCun, Yann, L. Bottou, Yoshua Bengio et P. Haffner (nov. 1998). « Gradient basedlearningapplied to document recognition ». Anglais. In : Proceedings of the IEEE.

[11] : Mokri Mohammed Zakaria, « Classification des images avec les réseaux de neurones,Convolutionnels», Mémoire de fin d'études Pour l'obtention du : diplôme de Master eninformatique, Université Abou BakrBelkaid Tlemcen, 2017.

[12] : Benjamin Graham, Fractional Max-Pooling, 12 May 2015.

[13] : A. Krizhevsky, I. Sutskever et G. E. Hinton, « Image Net Classification withDeepConvolutional Neural Networks », Advances in neural ProcessingSystemsdetraitement, vol. 1, 2012, p. 1097–1105.

[14] : https://towardsdatascience.com/metrics-to-evaluate-your-semantic-segmentation-model-6bcb99639aa2

[15] : Deep Learning Toolbox™

User's Guide

Mark Hudson Beale

Martin T. Hagan

Howard B. Demuth