

الجمهورية الجزائرية الديمقراطية الشعبية
وزارة التعليم العالي والبحث العلمي

UNIVERSITE BADJI MOKHTAR - ANNABA
BADJI MOKHTAR – ANNABA UNIVERSITY



جامعة باجي مختار – عنابة

Faculté : TECHNOLOGIE
Département : ELECTRONIQUE
Domaine : SCIENCES ET TECHNIQUES
Filière : télécommunication
Spécialité : systèmes de télécommunication

Mémoire

Présenté en vue de l'obtention du Diplôme de Master
Thème :

**Segmentation sémantique des images de drones par le
modèle U-Net**

Présenté par : Khettache Rania / Berrahmoune Intissar

Encadrant : BOUKARI Karima

Grade: MCA

UBM Annaba

Jury de Soutenance :

ZERMI Narima	M.C.A	UBM Annaba	President
Boukari Karima	M.C.A	UBM Annaba	Encadrant
			Co-encadrant
BOUCHAALA Ali	M.A.A	UBM Annaba	Examination

Année Universitaire : 2021/2022

Remerciement

Je voudrais tout d'abord adresser toute ma reconnaissance à la directrice de ce mémoire, Madame Karima Boukari pour sa patience, sa disponibilité et surtout ses judicieux conseils, qui ont contribué à alimenter ma réflexion. Pour m'avoir fait confiance

Pour le temps conséquent qu'il m'a accordé, ses qualités pédagogiques et scientifiques, sa franchise et sa sympathie

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

Je désire aussi remercier les professeurs de département d'électronique, qui m'ont fourni les outils nécessaires à la réussite de mes études universitaires.

Je tiens également à exprimer mes vifs remerciements à toutes les personnes qui ont contribué de près ou de loin à la réalisation de notre travail.

Dédicaces

Je dédie Ce modeste travail

A mes chers parents, pour tous leurs sacrifices, leur amour, leur tendresse, leur soutien et leurs prières tout au long de mes études,

A celle qui m'a donné la vie, la source de la tendresse, ma chère mère Mouhelkhir Fatma Zohra qui m'a apporté son appui durant toute mes années d'étude, pour son sacrifice et soutien qui m'ont donné l'amour, le courage et la sécurité ;

A mon cher père Mohamed qui m'a entouré de tous ses encouragements et son aide durant toute la période de mes études.

Si je vais bouillir le dictionnaire, je ne trouve pas les mots pour remercier, merci Pour votre soutien moral, matériel et financier durant toutes ces années d'études, Pour votre patience et pour avoir toujours cru en moi, Ma réussite est aussi la vôtre, Que dieu vous protège et vous garde pour moi.

A mes chères sœurs : Hamida, Sara et Yasmine pour leurs encouragements permanents, et leur soutien moral,

A mes chers frères, Yacine, Saad Eddine et Karim pour leur appui et leur encouragement,

A mes chers grands- parents que dieu vous protège et vous garde pour moi.

A toute ma famille, mes oncles, mes cousines pour leur soutien tout au long de mon parcours universitaire,

Que ce travail soit l'accomplissement de vos vœux tant allégués, et le fruit de votre soutien infaillible,

Merci d'être toujours là pour moi.

K.Rania

Dédicaces

Tout d'abord, je tiens à remercier DIEU, De m'avoir donné la force et le courage de mener.

Je dédie ce travail en signe de reconnaissance.

A celui qui a lutté et sacrifié pour m'offrir les conditions propices à ma réussite : Mon très cher père

A celle qui m'a étreint de tendresse et d'affection et qui a constitué la première école de mon existence .Ma très précieuse, chaleureuse et aimable mère

A mes sœurs, merci pour votre soutien et votre encouragement.

A mon âme sœur et mon amie Aya et mon amie Fulla et Rania Merci de rester avec moi et pour tout ce travail acharné et cette motivation

Et à tous ceux qui ont contribué de près ou de loin pour que ce mémoire soit possible, je vous dis merci.

Berrahmoune Intissar



Résumé :

L'utilisation de drones dans les activités liées aux ressources naturelles a considérablement augmenté ces dernières années en raison de leur disponibilité supérieure et de leur capacité à déployer rapidement des drones.

La détection et le suivi des routes dans les enregistrements de drones jouent un rôle important dans la vérification de l'état du réseau routier, permettant la navigation à distance sur le même terrain, la surveillance des routes ou la sécurité des données. La segmentation est une méthode de classification largement utilisée qui permet l'utilisation d'étiquettes sémantiques. Ce mémoire explore la segmentation sémantique des images aériennes pour identifier plus précisément le trafic routier à l'aide d'un modèle U-Net d'apprentissage en profondeur. Pour éviter le traitement fastidieux et chronophage de grandes images aériennes, la formation est effectuée sur de petites parcelles tirées au hasard des images de formation, suivie d'une segmentation. La méthode a été testée et validée sur des images de drone (Image UAV).

Mots-clés:

Image UAV, Deep Learning, patch Segmentation, U-Net

Abstract :

The use of drones in natural resource activities has increased significantly in recent years due to their superior availability and ability to deploy drones quickly.

Road detection and tracking in drone recordings play an important role in verifying the condition of the road network, enabling remote navigation on the same terrain, road monitoring or data security. Segmentation is a widely used classification method that enables the use of semantic labels. This dissertation explores semantic segmentation of aerial images to more accurately identify road traffic using a deep learning U-Net model. To avoid tedious and time-consuming processing of large aerial images, training is performed on small patches randomly drawn from the training images, followed by segmentation. The method has been tested and validated on UAV images.

Key Words : Image UAV, Deep Learning, patch Segmentation, U-Net

ملخص:

زاد استخدام الطائرات بدون طيار في أنشطة الموارد الطبيعية بشكل كبير في السنوات الأخيرة بسبب توافرها الفائق وقدرتها على نشر الطائرات بدون طيار بسرعة.

يلعب اكتشاف الطريق وتتبعه في تسجيلات الطائرات بدون طيار دورًا مهمًا في التحقق من حالة شبكة الطرق، مما يتيح التنقل عن بُعد على نفس التضاريس أو مراقبة الطريق أو أمان البيانات. التجزئة هي طريقة تصنيف مستخدمة على نطاق واسع تتيح استخدام العلامات الدلالية. تستكشف هذه الرسالة التقسيم الدلالي للصور الجوية لتحديد حركة المرور على الطرق بشكل أكثر دقة باستخدام نموذج U-Net للتعلم العميق. لتجنب المعالجة الشاقة والمستهلكة للوقت للصور الجوية الكبيرة، يتم إجراء التدريب على رقع صغيرة مستمدة عشوائيًا من صور التدريب، متبوعة بالتقسيم. تم اختبار الطريقة والتحقق من صحتها على صور الطائرات بدون طيار.

الكلمات الدالة :

صورة الطائرات بدون طيار، التجزئة، التعلم العميق U-Net .

Liste des tableaux :

Tableau (3.1) Matrice de confusion des prédictions d'un classifieur à C classes.....	(32)
Tableau (3.2) Matrice de confusion des prédictions d'un classifieur à deux classes....	(34)
Tableau (3.3) Le rapport d'entrainement pour PC1.....	(39)
Tableau (3.4) Les résultats des paramètres d'apprentissage pour image 10.....	(43)
Tableau (3.5) comparaison des résultats entre le modèle unet-1 et unet-2.....	(51)
Tableau (3.6) Trois classes du réseau U-NET en fonction du temps d'apprentissage.	(56)
Tableau (3.7) Le rapport d'entrainement	(56)

Liste des figures :

Chapitre 1 «Segmentation des images aériennes»

Figure (1.1) Principe de base de la télédétection	(5)
Figure (1.2) multi rotors.....	(6)
Figure (1.3) micro drone.....	(7)
Figure (1.4) exemple d’image segmentée.....	(11)

Chapitre 2 «Deep Learning»

Figure (2.1) Relations entre les ensembles "intelligence artificielle", "apprentissage automatique" et "apprentissage profond.....	(14)
Figure (2.2) Réseaux neuronaux, Les réseaux peuvent avoir des dizaines ou des centaines de couches cachées.....	(15)
Figure (2.3) Flux de travail de l'apprentissage profond. Les images sont Transmises au CNN.	(19)
Figure (2.4) Réseau de neurones artificiels.....	(20)
Figure (2.5) Détection des objets ET segmentation de instances.....	(21)
Figure (2.6) ConvNet architecture.....	(22)
Figure (2.7) architecture de couche de convolution.....	(22)
Figure (2.8) types de La couche de pooling.....	(23)
Figure (2.9) Allure de la fonction ReLU.....	(23)

Figure (2.10) Démonstration des couches entièrement connectées dans le CNN.....	(24)
Figure (2.11) Architecture U-Net.	(25)
Figure (2.12) Architecture U-Net couches de convolution	(26)
Figure (2.13) Architecture U-Net couches de convolution.....	(27)
Figure (2.14) Architecture U-Net couches de convolution	(27)
Figure (2.15) Architecture U-Net chemin expansif.....	(28)
Figure (2.16) Architecture U-Net couche de convolution avec un filtre de taille 1x1.....	(28)

Chapitre 3 «Travail personnel»

Figure (3.1) Le premier modèle unet-1.....	(38)
Figure (3.2) Architecture du Modèle1 encodeur –décodeur U-net1	(39)
Figure (3.3) image 10 avec parameter <i>Size patch [32*32]</i> MAX-EPOCH 20 MINI-PATCH 8 (A, B, C, D)	(44)
Figure (3.4) image 10 avec parameter <i>Size patch [64*64]</i> (A, B, C).....	(45)
Figure(3.5)image10avecparameterMAX-EPOCH15 MINI-PATCH8(A,B,C, D)....	(46)
Figure (3.6) image 10 avec parameter MAX-EPOCH 30 MINI-PATCH 8 (A, B,C,D).....	(47)
Figure(3.7)image 10 avec parameter MINI-PATCH 16 MAX-EPOCH 20(A,B,C,D).....	(48)
Figure (3.8) image 10 avec parameter MAX-EPOCH 30 MINI-PATCH 16 (A, B, C, D).....	(49)
Figure (3.9) deuxième modèle U-net2.....	(50)
Figure (3.10) Architecture du Modèle2 encodeur –décodeur U-net2.....	(51)
Figure (3.11) image 10 (A, B, C, D)	(52)

Figure (3.12) image 330 (A, B, C, D) (53)

Figure (3.13) notre modèle U-NET 2 sur une autre base de données à plusieurs classes (building, tree, road) (55)

Figure (3.14) : Résultat accuracy modèle U-NET 2 avec trois classes (building, tree, road).....(57)

Liste des symboles :

UAV : Unmanned Aerial Vehicle.

MP : M-P neurone, la fonctionnalité d'un neurone biologique..

SVM : Support Vector Machine

BP : La propagation des convictions

NTIC : nouvelles technologies de l'information et de la communication

CNN : Convolution Neural Networks.

GPU : Graphical Processing Unit.

TP : True Positive

TN : True Négative

FN : False Négative

FP : False Positive

Ppos : échantillons classés positivement

Pneg : négativement échantillons classés

P : nombre d'échantillons total

VP : Vrais Positifs

FN : Faux Négatifs

VN : Vrais Négatifs

FP : Faux Positifs

Liste de formule :

Numéro de la formule	La formule	page
(3.1)	$\frac{VP}{VP+FN} = \frac{VP}{P}$	34
(3.2)	$\frac{VN}{VN+FP} = \frac{VN}{N}$	35
(3.3)	$\frac{VP}{VP+FP} = \frac{VP}{P_{pos}}$	35
(3.4)	$\frac{VP + VN}{VP + VN + FP + FN} = \frac{P}{\Sigma}$	35
(3.5)	$F\beta = \frac{(1 + \beta^2)(Precision:Rappel)}{\beta^2:Precision:Rappel}$	35

Liste des matières :

⇒ <i>Remerciement</i>	(I)
⇒ <i>Dédicace</i>	(II)
⇒ <i>Dédicace</i>	(III)
⇒ <i>Résumés</i>	(IV)
⇒ <i>Liste Des Tableaux</i>	(VI)
⇒ <i>Liste Des Figures</i>	(VII)
⇒ <i>Liste Des Symboles</i>	(X)
⇒ <i>Liste Des Formules</i>	(XI)
⇒ <i>Table Des Matières</i>	(XII)
⇒ <i>Introduction Générale</i>	(I)

Chapitre 1 : Segmentation des images aériennes

1. Introduction	(4)
2. Les images utilisées dans la télédétection.....	(4)
3. Les images de drones.....	(6)
3.1. Les applications de la photographie aérienne.....	(7)
4. Domaines d'application.....	(8)
5. Segmentation des images UAV.....	(9)
5.1. L'analyse supervisée.....	(9)
5.2. L'analyse non supervisée.....	(10)
6. Segmentation sémantique d'images	(10)
6.1. La segmentation d'instance.....	(10)
7. Conclusion.....	(11)

Chapitre 2 : Deep Learning

1. Introduction.....	(13)
----------------------	------

2. Définition	(14)
3. Historique de Deep Learning.....	(16)
4. Domaine d'application	(18)
5. Types de Deep Learning.....	(19)
5.1.convolution neural network (CNN, or ConvNet)	(19)
5.2.Encoder-Décoder (U-Net)	(24)
6. Conclusion.....	(29)

Chapitre 03 : Travail personnel

1. Introduction.....	(31)
2. Base de donnes.....	(31)
3. Évaluation des méthodes.....	(32)
4. Principe de fonctionnement	(36)
5. résultats et interprétation.....	(37)
5.1.Partie 01 : génération du premier modelé.....	(37)
5.2.Partie 02 : training du modèle U-Net 1.....	(39)
5.3.Partie 03 : test sur le modèle U-net 2.....	(50)
5.4.Partie 04 : modèle U-NET 2 sur une autre base de données à plusieurs classes.....	(54)
6. Conclusion.....	(58)
⇒ Conclusion générale.....	(59)
⇒ Bibliographie.....	(60)

Introduction générale :

Le traitement mathématique des visuels utilisant autre chose qu'un traitement du signal dans lequel l'entrée est une image, une collection d'images ou une vidéo telle qu'une prise de vue ou une position cible, et la sortie est soit une image, soit des paramètres liés à l'image est connu sous le nom de prétraitement des données. La surveillance du trafic, le transport en commun, Les UAV pourraient être utilisés dans la stratégie et les opérations de transport pour examiner l'occupation régulière des itinéraires et surveiller l'utilisation des espaces de stationnement. Au cours de l'apprentissage automatique et de l'apprentissage en profondeur, l'utilisation de drones (UAV) pour la surveillance est un sujet majeur. Dans le but d'identifier, de reconnaître et de suivre des objets sur une série de photographies tout en essayant d'interpréter. Les routes, les rivières, les champs pétrolifères et les données de trafic sont également tous des usages fréquents des UAV. Les méthodes de détection routière sont nécessaires bien que difficiles. Dans de nombreuses Applications d'apprentissage en profondeur et applications de surveillance ou de navigation automobile.

Les webcams des drones sont considérées comme une plate-forme à faible coût capable de fournir des données aux systèmes de transport intelligents. À mesure que le nombre de voitures sur la route augmente, la demande de contrôle du trafic augmente également, et ce type de plate-forme se généralise. Pour la navigation, la détection et la collecte de données, les utilisateurs peuvent utiliser leur connaissance des zones routières pour trouver des lieux d'intérêt. La segmentation d'image génère une série de segments qui, ensemble, couvrent l'intégralité de l'image. La chrominance, l'intensité et la texture font partie des propriétés de pixel utilisées dans la segmentation.

Plusieurs algorithmes ont été développés pour la segmentation de plusieurs objets dans des images aériennes ou satellites, telles que des routes, des routes, des bâtiments, des toits et des avions. Dans ces systèmes, le traitement des données et des réponses, y compris une telle représentation caractéristique de ces qualités, est utilisé pour générer des règles.

L'apprentissage en profondeur peut également créer une représentation multicouche mais également des représentations de données hiérarchiques. Des méthodes telles qu'U-Net pour la segmentation d'images ont été développées. Ces méthodes reposent sur la simulation du processus d'enseignement pour porter des jugements sans avoir besoin

d'algorithmes complexes. Les algorithmes basés sur l'apprentissage en profondeur ont récemment dominé. Les exigences de précision les plus élevées pour une gamme de tâches de segmentation numérique.

Les réseaux d'apprentissage en profondeur (DL) ont fait émerger une nouvelle génération de modèles de segmentation avec des augmentations de performances incroyables, Avec les marges d'erreur les plus faibles sur les liens significatifs, cela a culminé avec la façon dont la plupart considèrent un changement de paradigme dans le domaine. La méthode de segmentation utilisant l'apprentissage en profondeur a été explorée dans ce mémoire. Pour la segmentation des routes, U-Net prend en charge deux architectures de réseau de neurones artificiels distinctes.

En tant que identification de pixels, notre travail aborde le sujet de l'interprétation d'images aériennes. L'objectif principal est de faire correspondre chaque pixel à la classe appropriée.

Le premier chapitre est dédié au cadre de ce mémoire. Nous y présentons l'importance de la télédétection dans le réseau routier. L'utilisation des images de drones dans ce type d'applications. Nous allons citer différentes techniques de segmentation d'image. Puis dans le deuxième chapitre, nous allons ensuite faire un survol des réseaux de neurones convolution en particulier le réseau UNET. Le chapitre trois détaillera les étapes de l'algorithme de segmentation. Les résultats seront alors présentés et commentés.

Chapitre I :

Segmentation des images aériennes

1. Introduction :

Dans ce premier chapitre, nous introduirons les notions fondamentales liées au contexte de notre travail. Dans un premier temps, nous aborderons quelques notions sur La segmentation d'image qui est une étape essentielle dans l'objet de recherche en traitement d'image, elle consiste à regrouper les pixels des images qui partagent une même propriété pour former des régions connexes en définissant les termes et les concepts essentiels qui permettent de mieux comprendre ce que nous observons avec l'image segmentée.

Nous allons parler d'images de drones, leurs utilisations et le domaine de notre application.

Puis, la segmentation des images UAV et la segmentation sémantique d'images.

2. Les images utilisées dans la télédétection :

L'expression "télédétection", est composée des mots "télé" et "détection", qui signifient "à distance" et "décélérer" (ou "découvrir", respectivement). La mesure ou l'acquisition d'informations sur un objet, une région ou un phénomène en évaluant les données acquises par un appareil qui n'est pas en contact avec un autre équipement est appelée télédétection. Le processus complet de capture et d'enregistrement de l'énergie du rayonnement électromagnétique émis ou réfléchi, de traitement et d'évaluation des données, puis d'utilisation de ces données est appelé télédétection. Elle peut se faire depuis une plateforme aérienne (ballon ou avion), depuis un satellite, ou depuis tout autre type d'équipement pouvant collecter des données environnementales.

Bien que le traitement numérique soit couramment utilisé dans l'analyse des données de télédétection, des méthodes non numériques peuvent également être utilisées. La télédétection a été théoriquement créée en combinant deux vieilles inventions : la montgolfière et la photographie. En fait, la télédétection moderne a fait ses débuts aux États-Unis dans les années 1960, lorsque de nouveaux capteurs ont complété la photographie aérienne traditionnelle, qui avait gagné en popularité au cours du XXe

siècle. Le lancement du satellite Spoutnik en 1957 a marqué un tournant dans la télédétection moderne. De nombreux pays, dont les États-Unis, le Canada, la France, l'ancienne Union soviétique, puis la Russie, la Chine, le Japon et l'Inde, ont construit leurs propres systèmes de télédétection. Les photos de télédétection sont désormais une source essentielle de données pour une variété d'applications.

Le premier domaine important d'utilisation de la télédétection a été l'étude de l'atmosphère (climatologie et météorologie) (notamment pour suivre l'évolution spatio-temporelle de la couverture nuageuse, mesurer la température et les précipitations, etc.). Le champ d'application a ensuite été élargi pour inclure à la fois des applications militaires et civiles (planification de mission) (comme par exemple dans l'urbanisme, les sciences de l'environnement ou la géologie). L'imagerie par télédétection fournit des informations utiles qui peuvent être utilisées dans une variété de domaines, y compris la couverture terrestre et la cartographie de l'utilisation des terres, pour détecter les zones de végétation, les surfaces imperméables et une étude plus approfondie des changements.

L'aménagement du territoire, l'urbanisme, la gestion de l'environnement, l'information, les transports et les télécommunications font partie des utilisations. La forme de télédétection qui nous intéresse dans ce mémoire est celle qui génère des photographies aériennes.

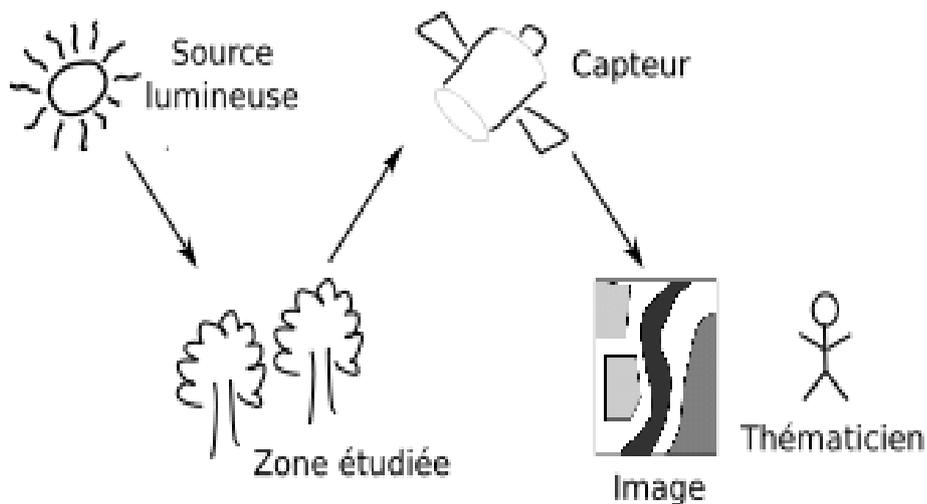


Figure (1.1) : Principe de base de la télédétection

3. les images de drones :

Les drones sont des aéronefs capables de voler et d'effectuer des missions sans la présence humaine. Cette première caractéristique essentielle justifie leur désignation de véhicules aériens sans pilote (Unmanned Aerial Vehicle.) (Ou UAV). Le nom UAV est très limité car il ne couvre que les véhicules aériens. Un drone n'est en réalité qu'un des éléments d'un système conçu et déployé pour effectuer une ou plusieurs missions. C'est pourquoi les experts parlent de "systèmes aériens sans pilote" Drone. Ainsi, tous les drones, autonomes ou non, nécessitent au moins un opérateur au sol pour recueillir en temps réel les bénéfices de la mission : ce dernier reçoit, analyse et enregistre les informations transmises par le drone. Après le décollage, Le drone navigue de manière autonome vers la zone d'observation sélectionnée, patrouille et envoie des données importantes, puis revient à sa position de départ.

Un drone est utilisé à faible hauteur pour capturer des photos très précises qui seraient impossibles à obtenir en hélicoptère, ainsi que pour réaliser des prises de vue aériennes en haute définition à un coût minime.

Nous avons différents types de drones à titre d'exemples :

1) Multi rotors:



Figure (1.2): multi rotors

2) Micro drone :**Figure (1.3) : micro drone****3.1 Les applications de la photographie aérienne :**

Avec l'avènement de la Première Guerre mondiale, les autorités militaires ont rapidement compris l'intérêt de prendre des photos au-dessus des troupes ennemies. À l'époque, les photos étaient prises à l'aide de ballons et d'avions. Par la suite, le renseignement militaire a été servi par des équipements sophistiqués, y compris un mirage pendant la guerre froide de la série avec l'Union soviétique. Au XXI^e siècle, dans le domaine militaire, les images prises depuis l'espace ont remplacé les photos prises depuis les avions. . Les missions militaires assignées aux drones peuvent être divisées en trois catégories :

- Renseignement et surveillance.
- Assister au combat
- Le combat proprement dit Les drones de nouvelle génération.

Des photos aériennes sont également commandées par des municipalités ayant des projets de rénovation soucieuses du respect de l'authenticité architecturale de leurs bâtiments.

Qui n'a pas entendu parler du site archéologique. Les traces vues des photos aériennes ont révélé que les fondations qui étaient bien enfouies dans le sol ont changé la couleur de la surface. La distance de l'objectif peut donner un sens à ces éléments énormes qui ne peuvent pas être vus dans leur ensemble sur le sol comme les géoglyphes de Nazca.

4. domaines d'application :

La photographie aérienne est utilisée en raison de son originalité de perspective. Parmi les informations recherchées dans les images de télédétection, les réseaux routiers présentent un intérêt particulier. L'extraction des réseaux routiers a fait l'objet de nombreux travaux en traitement d'images et en télédétection depuis deux décennies. En effet, il est nécessaire de développer des procédures de mise à jour automatique des fichiers cartographiques pour des fonctions allant de la gestion du trafic à la navigation automatique des véhicules. Il s'agit en effet d'un élément de la carte topographique qui doit être constamment révisé au fur et à mesure que de nouvelles routes sont construites ou que le tracé change chaque année. De plus, les réseaux routiers fournissent les informations contextuelles les plus importantes, permettant de localiser d'autres types d'éléments, tels que des structures résidentielles ou commerciales, sur des images de télédétection. Aujourd'hui encore, la cartographie des réseaux routiers se fait en grande partie avec l'apport de l'interprétation d'images et de mesures de terrain. Ce processus est un processus long et coûteux et doit être exécuté à chaque fois que le réseau est mis à jour. Il existe un intérêt indéniable pour les méthodes automatiques et robustes d'extraction de routes à partir d'images de télédétection. Diverses méthodes d'extraction de réseaux routiers à partir d'images aériennes et satellitaires ont émergé ces dernières années, et deux grandes catégories peuvent être identifiées dans la littérature : les méthodes semi-automatiques et les méthodes automatiques. Habituellement, La méthode semi-automatique permet une extraction précise et efficace des itinéraires spécifiés par l'utilisateur. Cependant, ces méthodes offrent des gains de productivité moindres par rapport aux travaux d'extraction effectués par des experts. La tendance vers des approches purement automatisées a ensuite fait l'objet de plusieurs efforts de recherche. Des exemples de méthodes existantes incluent la morphologie mathématique, les réseaux de neurones, la programmation dynamique, les contours actifs, les modèles de Markov ou l'analyse multi-résolution. Ces méthodes exploitent les principales caractéristiques des routes dans les images, à savoir la géométrie, la radiométrie et les propriétés sur le réseau

telles que la connectivité. Certains travaux considèrent le contexte spatial afin de se concentrer sur les zones les plus susceptibles de contenir des routes.

5. segmentation des images UAV :

La première étape de l'analyse d'images de télédétection consiste à extraire les informations recherchées (divers objets géographiques présents dans la scène analyse, leur nombre, etc.). Cela peut être fait manuellement (traduction de photos). L'extraction manuelle d'informations remonte aux premiers stades de la technologie de télédétection et est interprétée à travers des photographies aériennes. Avec l'émergence et la propagation des nouvelles technologies (émergence des enregistrements numériques de données le développement de la télédétection et de l'informatique), la quantité de données tend à augmenter, de sorte que ce processus d'interprétation visuelle prend de plus en plus de temps. L'analyse automatique des données (extraction automatique d'informations) devient alors de plus en plus nécessaire.

Il ressort de la littérature que deux aspects indissociables coexistent dans le problème de l'extraction automatique d'informations à partir d'images de télédétection. Le premier aspect concerne les techniques de traitement d'image (dans la plupart des cas de segmentation), tandis que le deuxième résultat traite des données complexes (telles que la classification) dans l'analyse et l'exploration de données. Ces deux éléments bénéficient d'années de recherche.

Ils diffèrent principalement par leur finalité, mais aussi par les algorithmes mis en œuvre.

Le Domaine de la segmentation d'images et de la classification des données.

Ces deux grands domaines peuvent alors être déterminées par analyse supervisée ou non supervisée :

5.1. L'analyse supervisée :

Nécessite des processus basés sur des techniques de segmentation et de classification supervisées. La segmentation supervisée intègre des connaissances préalables (forme, informations spectrales, etc.) des objets à extraire lors de la segmentation.

La classification supervisée consiste à générer, à partir d'un ensemble d'objets de classe connus appelé ensemble d'apprentissage, un modèle de classe qui caractérise les données, et Il est important de noter qu'il peut être généralisé à des ensembles de données plus

importants. Appliqué à l'image, la segmentation de l'image revient à classer ou associer à chaque pixel p de l'image, un index k de la classe C_k à laquelle il appartient.

5.2 L'analyse non supervisée :

Doit être similaire à celles La méthode précédente, mais cette fois de manière non supervisée. La segmentation non supervisée ne nécessite aucune connaissance préalable des objets extraits. La classification non supervisée est également connue sous le nom de partitionnement. D'autre part, le clustering des données vise à regrouper les données en groupes au sein d'un ensemble homogène (classe) grâce à des critères de similarité. Les données, les entités de la classe sont les plus similaires possibles (homogénéité au sein d'une classe) et les entités de classes différentes sont les plus distinctes possible (séparabilité entre les classes).

6. segmentation sémantique d'images :

Segmentation d'images et détection d'objets en optique les images de télédétection souffrent généralement de plusieurs défis croissants tels que les variations de vue, l'ombre, et l'occlusion. Lors des premières études, la faible résolution spatiale des images satellites antérieures (telles que Landsat), ne pouvaient pas permettre la détection de substances artificielles ou naturelles objets. Ces études ont généralement porté sur l'extraction des propriétés de la région à partir de ces images.

Avec le progrès de la technologie de télédétection, le très haut satellite de résolution (VHR) et des images aériennes qui ont été générées généralement par des véhicules aériens sans pilote (UAV) fournissent des informations spatiales et texturales plus détaillées. Vu que

La résolution spatiale est accrue, une plus grande gamme d'objets reconnus par l'homme peut être identifiée séparément.

6.1 La segmentation d'instance :

La segmentation d'image est une technique de vision par ordinateur qui consiste à diviser automatiquement une image en régions de pixels appartenant à la même classe d'objets. La segmentation d'images a de nombreuses applications, notamment en imagerie aérienne. La segmentation d'image est divisée en segmentation sémantique et segmentation d'instance. Dans le cadre de la segmentation sémantique, on cherche à

classer les pixels d'une image comme appartenant ou non à une certaine classe. D'autre part, avec la segmentation d'instance, plus d'informations sur une image peuvent être obtenues en divisant différentes instances du même objet.

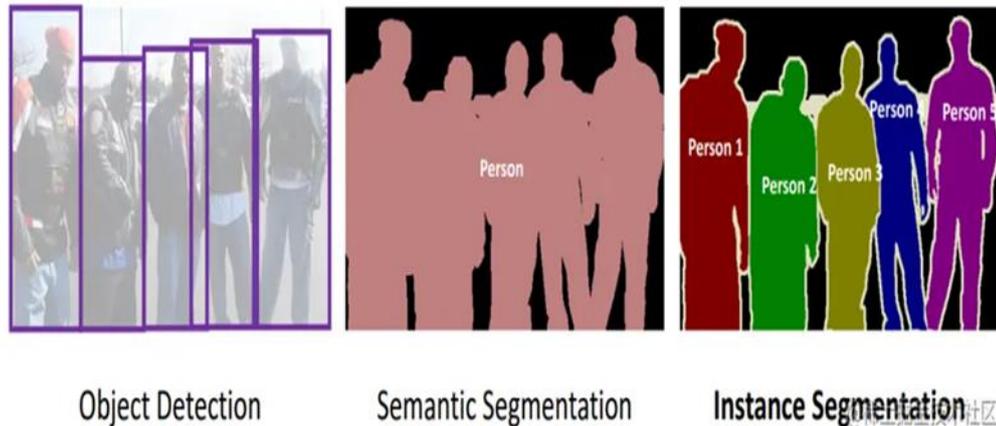


Figure (1.4) : exemple d'image segmentée

7. Conclusion :

La segmentation prend une place importante dans le processus global. Nous avons Aussi constaté à travers cet état de l'art que peu de travaux traitent de la détection et la Segmentation sémantique à partir d'images aériennes. Les travaux de ce mémoire portent sur la segmentation sémantique d'images de drones par le modèle U-Net. Avant de détailler ces travaux dans les chapitres 3. Nous présentons dans le chapitre suivant un historique sur les méthodes de segmentation d'images et bien définie le modèle U-Net

Chapitre II :

Deep Learning

1. Introduction :

L'apprentissage profond est l'un des techniques les plus débattues en ce moment en raison de son énergie et de son pouvoir d'imiter l'esprit humain, qui est à la fois merveilleux et dangereux.

Dernièrement, nous entendons parler dans les nouvelles des nombreuses réalisations remarquables de l'intelligence artificielle, et bien que beaucoup d'entre elles contiennent de fausses affirmations, certaines d'entre elles sont réelles et ont un impact sur le travail et la vie quotidienne, La plupart de ces succès au cours des 10 dernières années sont dus à des applications dans l'un des sous-domaines de l'intelligence artificielle, à savoir l'apprentissage automatique.

L'un des groupes d'algorithmes d'apprentissage automatique est constitué de réseaux de neurones, qui ont été initialement inspirés par le travail du cerveau humain, et ont ensuite évolué vers diverses formes.

Il est devenu possible de former des modèles de réseaux de neurones plus compliqués mais aussi multicouches que jamais grâce aux développements de cette discipline, ainsi que le domaine concerné par ces réseaux de neurones profonds et complexes est connu sous le nom d'apprentissage profond (Deep Learning), C'est aussi une simulation du cerveau humain à certains égards seulement.

Le concept d'apprentissage en profondeur n'est pas nouveau ; il a été découvert dans les années 1990, mais ses applications se sont considérablement développées en 2015 du fait de la quantité massive de données traitées. La majorité des modèles d'apprentissage en profondeur contemporains sont construits sur des réseaux de neurones. Un réseau de neurones convolutifs synthétiques (CNN), basé notamment sur U-Net.

Et ce dernier est un réseau de neurones convolutifs développé pour la segmentation d'images biomédicales au Département d'informatique de l'Université de Fribourg. [1] Le réseau est basé sur l'ensemble du réseau convolutifs [2].

- C'est notre sujet ici dans ce chapitre, donc C'est quoi le Deep Learning ?

2. Définition :

L'apprentissage profond (Deep Learning) est une forme d'intelligence artificielle dérivée de l'apprentissage automatique (machine Learning).

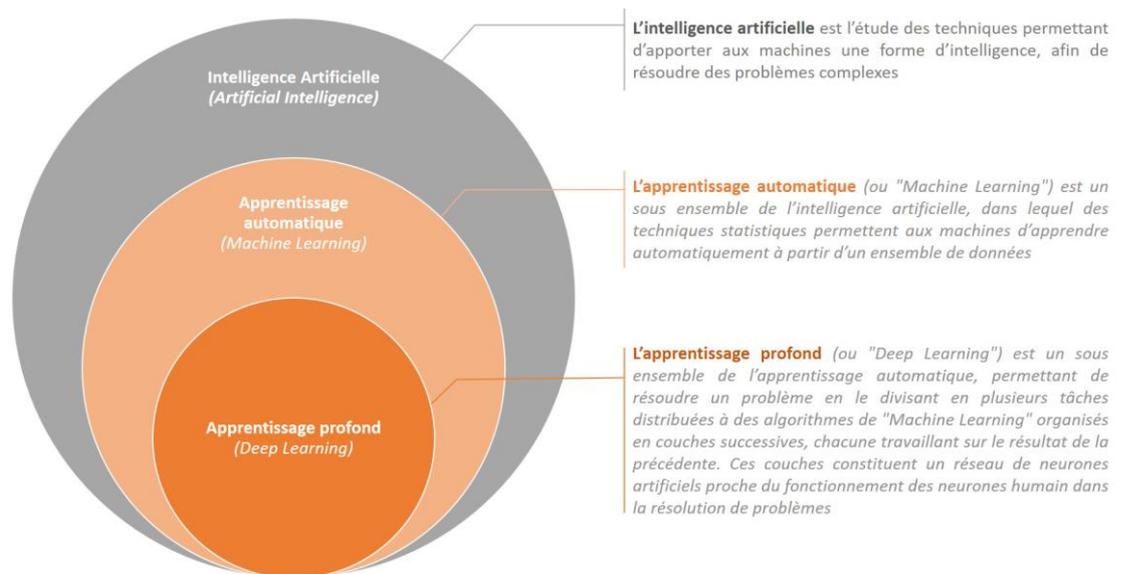


Figure (3.1) : Relations entre les ensembles "intelligence artificielle", "apprentissage automatique" et "apprentissage profond"

L'apprentissage en profondeur vise à imiter toutes les capacités de l'esprit humain, y compris la vision, la compréhension de la parole, la composition, l'ouïe et d'autres capacités puissantes. Les chercheurs étudient l'esprit humain et comment il développe des algorithmes capables de le simuler. Pour cette raison, ces algorithmes sont inspirés de la recherche médicale et neurologique humaine.

Par exemple, imaginez un enfant dont le premier mot est chien. Le jeune enfant apprend ce qu'est un chien – et ce qu'il n'est pas – en pointant des objets et en prononçant le mot chien. Le parent dit : « Oui, c'est un chien » ou « Non, ce n'est pas un chien ». En continuant à montrer des objets du doigt, le jeune enfant devient plus conscient des caractéristiques que possèdent tous les chiens. Ce que le jeune enfant fait, sans le savoir, c'est clarifier une abstraction complexe – le concept de chien – en construisant une hiérarchie dans laquelle

chaque niveau d'abstraction est créé avec les connaissances acquises au niveau précédent de la hiérarchie. [2].

La plupart des méthodes d'apprentissage en profondeur utilisent des architectures de réseaux de neurones, c'est pourquoi les modèles d'apprentissage en profondeur sont souvent appelés réseaux de neurones profonds. Le terme "profondeur" fait généralement référence au nombre de couches cachées dans un réseau de neurones [3].

Les réseaux de neurones Traditionnels ne contiennent que 2 ou 3 couches cachées, tandis que les réseaux profonds peuvent en avoir jusqu'à 150.

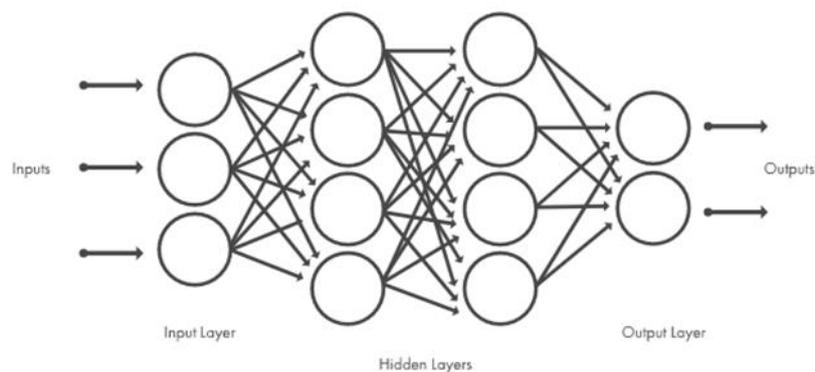


Figure (2.2) : Réseaux neuronaux, Les réseaux peuvent avoir des dizaines ou des centaines de couches cachées.

Les modèles d'apprentissage en profondeur sont formés à l'aide de grands ensembles de données étiquetés et d'architectures de réseaux neuronaux qui apprennent les fonctionnalités directement à partir des données, éliminant ainsi le besoin d'extraire manuellement les fonctionnalités. D'autre part, les techniques d'apprentissage en profondeur sont très complexes et il existe de nombreux types de réseaux de neurones qui peuvent résoudre un problème ou un ensemble de données spécifique :

- Les réseaux de neurones convolutifs (CNN), souvent utilisés dans les applications de vision par ordinateur et de classification d'images, peuvent identifier des caractéristiques et des modèles dans les images, permettant des tâches telles que la détection et la reconnaissance d'objets.

- L'Encodeur-Décodeur est un réseau de neurones et une structure largement utilisée dans l'apprentissage en profondeur et utilisé dans de nombreux projets. C'est un pilier fondamental dans les logiciels de traduction. [4]

3. Historique de Deep Learning :

L'un des domaines les plus importants de l'apprentissage automatique, l'apprentissage en profondeur a gagné en popularité au fil des ans et a attiré l'attention nationale et internationale. D'autre part, le succès de l'apprentissage en profondeur n'est pas un hasard. Il a une longue histoire de développement.

Examinons le processus de croissance de l'apprentissage en profondeur. Il a traversé trois étapes historiques de développement :

Partie 1 :

En 1943, le psychologue McCulloch et le mathématicien logicien Pitts publient l'article «Logical Calculus of Inner Thought in Neural Activity» et proposent le modèle MP (M-P Neuron, la fonctionnalité d'un neurone biologique.). En tant qu'origine du réseau neuronal artificiel, le modèle MP a créé une nouvelle ère de réseau neuronal artificiel et jeté les bases du modèle de réseau neuronal.

En 1949, Donald Hebb, un célèbre psychologue canadien, proposa une règle basée sur l'apprentissage non supervisé, la règle de Hebb, dans « L'organisation du comportement». À la fin des années 1950, sur la base du modèle MP et de l'étude des règles d'apprentissage de Hebb, le scientifique américain Rosenblatt a découvert un algorithme d'apprentissage similaire au processus d'apprentissage humain - l'apprentissage par le perceptron. Et en 1958, un réseau de neurones composé de deux couches de neurones a été formellement proposé, appelé "perceptron".

Mais avec l'approfondissement de la recherche, en 1969, le «père de l'IA» Marvin Minsky et le fondateur du langage LOGO Simon Piper ont co-écrit un livre «Perceptron», dans lequel ils ont prouvé que les perceptrons à une seule couche ne peuvent pas résoudre des problèmes linéairement inséparables. [5]

Partie 2 :

En 1982, le célèbre physicien John Hopfield a inventé le réseau de neurones Hopfield. Le réseau de neurones de Hopfield est un réseau de neurones récurrent qui combine un système de mémoire et un système binaire. Le réseau de Hopfield peut également simuler la mémoire humaine.

Jusqu'en 1986, Jeffrey Hinton, le père de l'apprentissage profond, a proposé un algorithme de rétro propagation adapté aux perceptrons multicouches - l'algorithme BP (La propagation des convictions). Sur la base de la propagation vers l'avant du réseau neuronal traditionnel, l'algorithme BP ajoute le processus de propagation vers l'arrière de l'erreur.

De plus, au milieu des années 1990, d'autres algorithmes d'apprentissage automatique peu profonds représentés par SVM (Support Vector Machine) ont été proposés et ont obtenu de bons résultats dans les problèmes de classification et de régression, et leurs principes étaient évidemment différents des modèles de réseaux de neurones. [6]

Partie 3 :

En 2006, Jeffrey Hinton et son étudiant Ruslan Salakhudinov ont formellement proposé le concept d'apprentissage en profondeur. Dans un article publié dans la plus grande revue universitaire au monde «Science», ils ont donné une solution détaillée au problème de la «disparition du gradient» - en entraînant l'algorithme couche par couche grâce à une méthode d'apprentissage non supervisée, puis en utilisant un algorithme de rétropropagation supervisé pour régler.

En 2012, lors du célèbre concours de reconnaissance d'images Image Net, l'équipe dirigée par Jeffrey Hinton a adopté le modèle d'apprentissage en profondeur Alex Net pour remporter le championnat. Alex Net utilise la fonction d'activation ReLU pour résoudre fondamentalement le problème de la disparition des gradients et utilise le GPU pour améliorer considérablement la vitesse de fonctionnement du modèle.

Avec l'avancement continu de la technologie d'apprentissage en profondeur et l'amélioration continue des capacités de traitement des données, en 2014, le projet DeepFace de Facebook basé sur la technologie d'apprentissage en profondeur a atteint un taux de précision de plus de 97 % dans la reconnaissance faciale, ce qui est presque le même que le taux de précision. De la reconnaissance humaine Pas de différence De tels résultats prouvent

une fois de plus que l'algorithme d'apprentissage en profondeur est inégalé dans la reconnaissance d'images. [7]

4. Domaine d'application :

Deep Learning serait utilisé dans une gamme de domaines NTIC (nouvelles technologies de l'information et de la communication), notamment :

- La robotique.
- La sécurité.
- L'art.
- L'intelligence artificielle en général.
- Détection de la fraude financière : l'apprentissage en profondeur est appliqué avec succès à la détection de la fraude financière, à la détection de l'évasion fiscale. [8]
- Militaire : le département de la Défense des États-Unis a appliqué l'apprentissage en profondeur pour former des robots à de nouvelles tâches grâce à l'observation. [9] est utilisé pour détecter des objets à partir de satellites qui identifient des zones d'intérêt, ainsi que des zones sûres ou dangereuses pour les troupes. [10]
- Reconstruction d'image : la reconstruction d'image est la reconstruction des images sous-jacentes à partir des mesures liées à l'image. Plusieurs travaux ont montré les performances meilleures et supérieures des méthodes d'apprentissage en profondeur par rapport aux méthodes analytiques pour diverses applications, par exemple l'imagerie spectrale [11] et l'imagerie par ultrasons. [12]
- Restauration d'image : l'apprentissage en profondeur a été appliqué avec succès à des problèmes inverses tels que le dé bruitage, la super-résolution, l'inpainting et la colorisation de film. Image Prior, qui s'entraîne sur l'image qui doit être restaurée. [13]
- La reconnaissance d'image, La reconnaissance vocale. [14]

5. Types de Deep Learning :

5.1. convolution neural network (CNN, or ConvNet)

La plupart des modèles modernes d'apprentissage profond sont basés sur des réseaux neuronaux artificiels, en particulier des réseaux neuronaux convolutifs (CNN),

Un réseau neuronal convolutifs ou CNN est un type de réseau neuronal artificiel, largement utilisé pour la reconnaissance et la classification des images/objets. Le Deep Learning reconnaît donc les objets dans une image en utilisant un CNN.

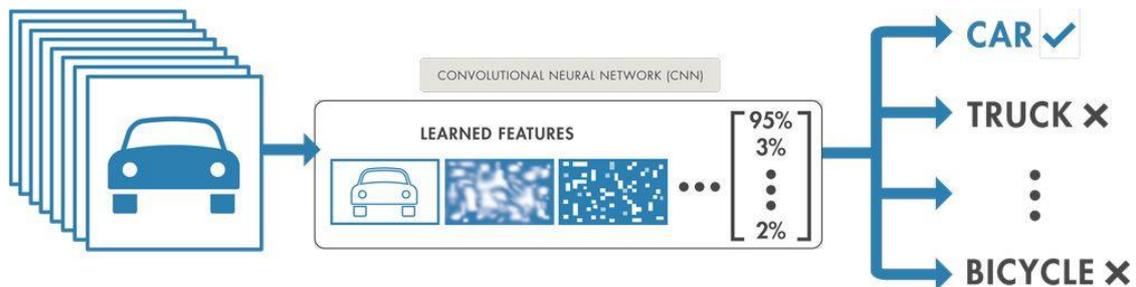


Figure (2.3): Flux de travail de l'apprentissage profond. Les images sont transmises au CNN.

CNN jouent un rôle majeur dans diverses tâches/fonctions telles que les problèmes de traitement d'images, les tâches de vision par ordinateur comme la localisation et la segmentation, l'analyse vidéo, la reconnaissance des obstacles dans les voitures à conduite autonome, ainsi que la reconnaissance vocale dans le traitement du langage naturel. Les CNN jouent un rôle important dans ces domaines émergents et à croissance rapide. Un réseau neuronal typique comporte une couche d'entrée, des couches cachées et une couche de sortie. [15]

Tout comme un neurone dans le cerveau traite et transmet des informations dans tout le corps, les neurones artificiels ou les nœuds dans les réseaux neuronaux prennent des entrées, les traitent et envoient le résultat comme sortie. L'image est fournie en entrée. La couche

d'entrée accepte les pixels de l'image sous la forme de tableaux. Dans les CNN, il peut y avoir plusieurs couches cachées, qui effectuent l'extraction de caractéristiques de l'image en effectuant des calculs. Il peut s'agir de couches de convolution, de mise en commun, d'unités linéaires rectifiées et de couches entièrement connectées. La convolution est la première couche qui effectue l'extraction des caractéristiques d'une image d'entrée.

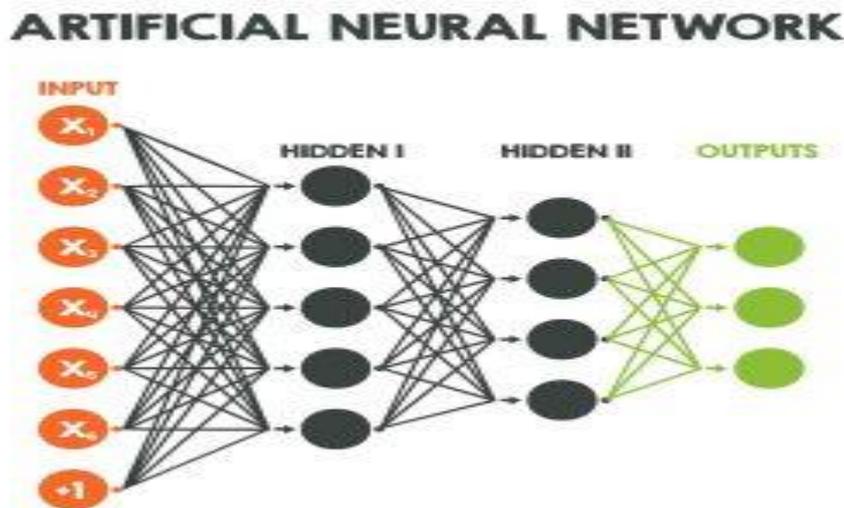


Figure (2.4): Réseau de neurones artificiels

La couche entièrement connectée classe l'objet et l'identifie dans la couche de sortie. "Les réseaux CNN sont des réseaux de type feedforward, c'est-à-dire que le flux d'informations se fait dans une seule direction, de leurs entrées vers leurs sorties.

Les CNN ont fondamentalement changé notre approche de la reconnaissance d'images, car ils peuvent détecter des modèles et leur donner un sens. Ils sont considérés comme l'architecture la plus efficace pour les tâches de classification, d'extraction et de détection d'images, car la précision de leurs résultats est très élevée. [16]

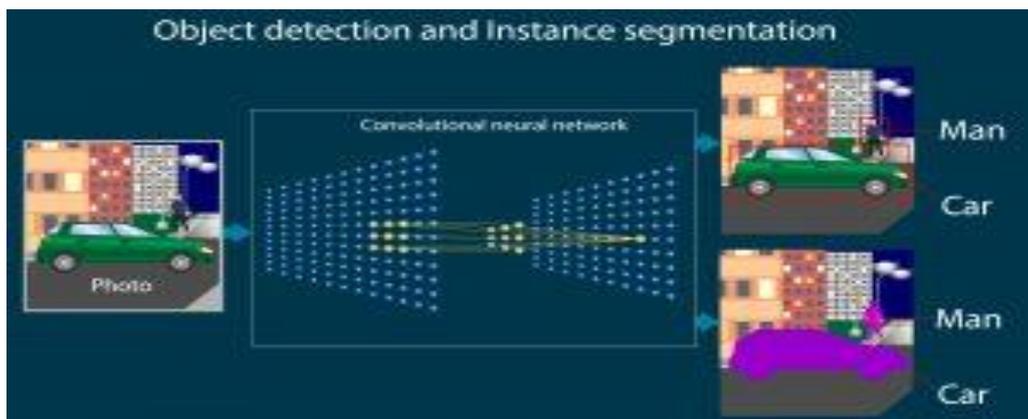


Figure (2.5): Détection des objets ET segmentation des instances

Ils ont de nombreuses applications dans les tests du monde réel, où ils produisent des résultats de haute qualité et peuvent faire un bon travail de localisation et d'identification de l'emplacement d'une personne, d'une voiture, d'un oiseau, etc. dans une image. Cet aspect a fait d'eux la méthode de prédilection pour les prédictions impliquant toute image en entrée.

Une caractéristique essentielle des CNN est leur capacité à atteindre une "invariance spatiale", ce qui implique qu'ils peuvent apprendre à reconnaître et à extraire des caractéristiques d'image n'importe où dans l'image. Il n'est pas nécessaire de procéder à une extraction manuelle, car les CNN apprennent les caractéristiques par eux-mêmes à partir de l'image/des données et effectuent l'extraction directement à partir des images. Cela fait des CNN un outil puissant au sein du Deep Learning pour obtenir des résultats précis.

L'objectif de CNN est de réduire l'image afin qu'elle puisse être traitée plus facilement sans perdre les fonctionnalités qui aident à des prédictions précises. L'architecture ConvNet comporte quatre couches sont :

1. La couche de convolution (CONV)
2. La couche de pooling (POOL)
3. La couche de correction (ReLU)
4. La couche "entièrement connectée" (FC)

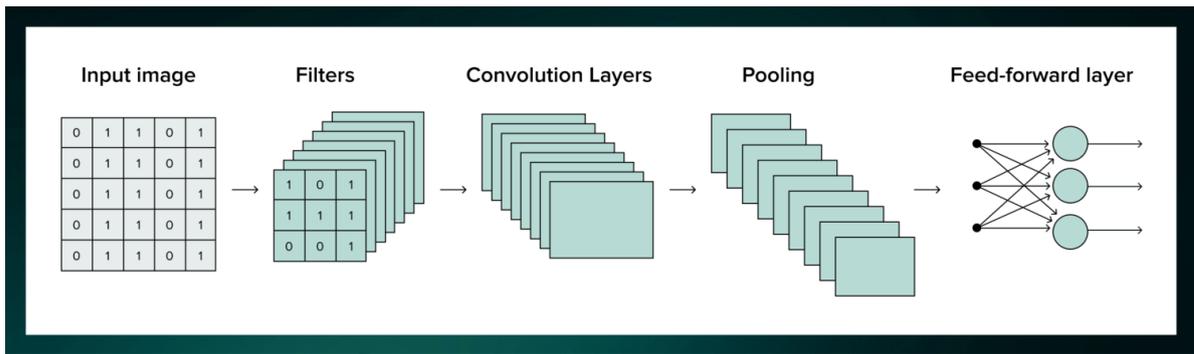


Figure (2.6): ConvNet architecture

1. **La couche de convolution (CONV) :** Couche de convolution est le composant central du CNN, où la plupart des calculs sont effectués. Cela nécessite certains composants tels que des données d'entrée, des filtres et des cartes d'entités. Un détecteur de caractéristiques est un tableau bidimensionnel de poids qui représentent une partie d'une image. La taille varie, mais la taille du filtre est généralement une matrice 3x3.

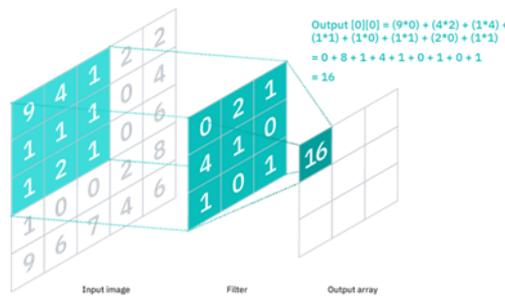


Figure (2.7): architecture de couche de convolution

Il détermine également la taille du champ récepteur, il n'est pas nécessaire de connecter toutes les valeurs de sortie de la carte des caractéristiques à toutes les valeurs de pixel de l'image d'entrée [17]. Vous devez vous connecter au champ récepteur auquel le filtre est appliqué.

2. **La couche de pooling (POOL) :** La couche de pooling, également appelées sous-échantillonnage, effectuent une réduction de la dimensionnalité, en réduisant le nombre de paramètres en entrée.[18] Comme pour la couche convolution, l'opération de mise en

commun balaie un filtre sur l'ensemble de l'entrée, mais la différence est que ce filtre n'a pas de poids.

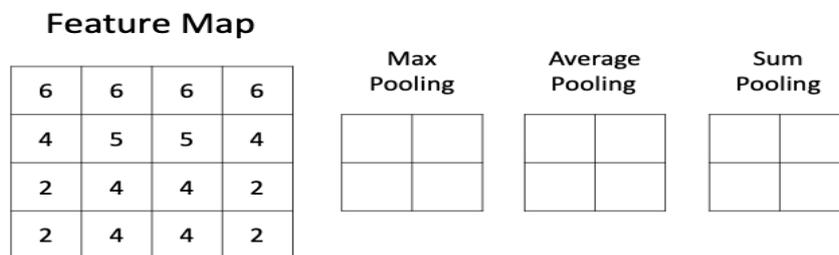


Figure (2.8): types de La couche de pooling

Il existe deux principaux types de La couche de pooling :

Max pooling : Lorsque le filtre se déplace sur l'entrée, il sélectionne le pixel ayant la valeur maximale à envoyer au tableau de sortie.

Average pooling : À mesure que le filtre se déplace sur l'entrée, il calcule la valeur moyenne dans le champ réceptif pour l'envoyer au tableau de sortie. Bien qu'une grande quantité d'informations soit perdue dans la couche de mise en commun, elle présente également un certain nombre d'avantages pour le CNN.

3. **La couche de correction (ReLU) :** ReLU (Rectified Linear Units) est une fonction d'activation. Permet une formation plus rapide et plus efficace en faisant correspondre les valeurs négatives à zéro et en conservant les valeurs positives. On parle parfois d'activation, car seules les caractéristiques activées sont reportées dans la couche suivante.

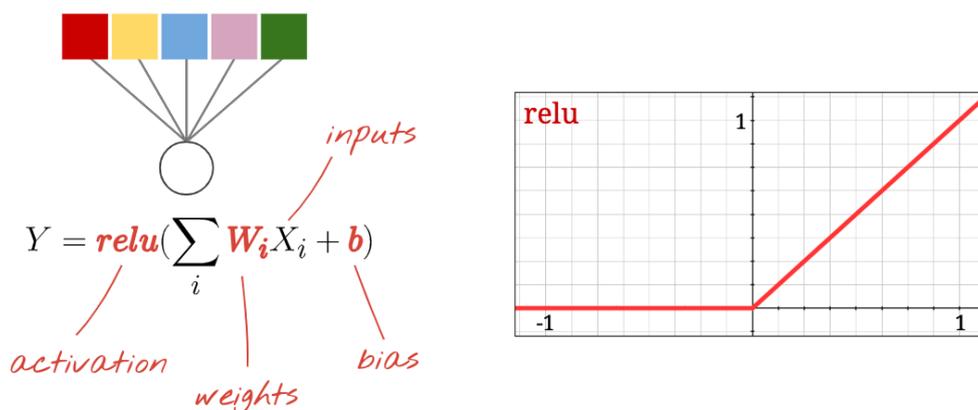


Figure (2.9) : Allure de la fonction ReLU

4. **La couche "entièrement connectée" (FC) :** Le nom de la couche entièrement connectée se décrit bien. Dans la couche entièrement connectée, chaque nœud de la couche de sortie est directement connecté à un nœud de la couche précédente.

Alors que les couches de convolution et de mise en commun ont tendance à utiliser des fonctions ReLu, les couches FC s'appuient généralement sur une fonction d'activation softmax pour classer les entrées de manière appropriée, en produisant une probabilité de 0 à 1

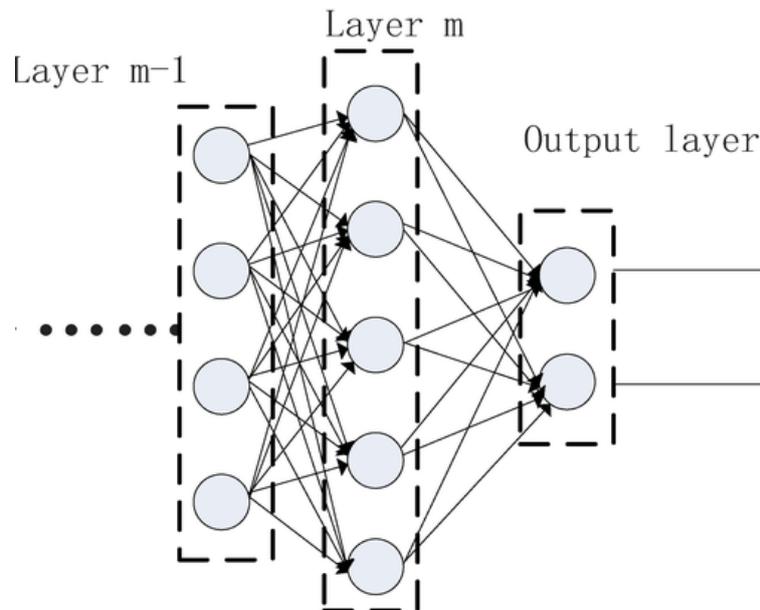


FIGURE (2.10) : Démonstration des couches entièrement connectées dans le CNN.

5.2. Encoder-Decoder (U-Net) :

L'architecture U-Net, issu du réseau de neurones CNN traditionnel, publiée pour la première fois au cours de l'année 2015, a constitué une révolution dans le domaine de l'apprentissage profond. L'architecture a remporté avec une grande marge le défi de suivi des cellules de l'International Symposium on Biomédical Imaging (ISBI) de 2015 dans de nombreuses catégories. Certains de leurs travaux incluent la segmentation de structures neuronales dans des piles de microscopie électronique et des images de microscopie à lumière transmise.

Avec cette architecture U-Net, la segmentation d'images de taille 512X512 peut être calculée avec un GPU moderne en peu de temps. De nombreuses variantes et modifications de cette architecture ont vu le jour en raison de son succès phénoménal.

L'U-Net est une architecture élégante qui résout la plupart des problèmes rencontrés. Il utilise le concept de réseaux entièrement convolution pour cette approche. L'intention de l'U-Net est de capturer à la fois les caractéristiques du contexte et de la localisation. Ce processus est complété avec succès par le type d'architecture construit. L'idée principale de l'implémentation est d'utiliser des couches contractantes successives, qui sont immédiatement suivies par des opérateurs de sur échantillonnage pour obtenir des sorties de plus haute résolution sur les images d'entrée.

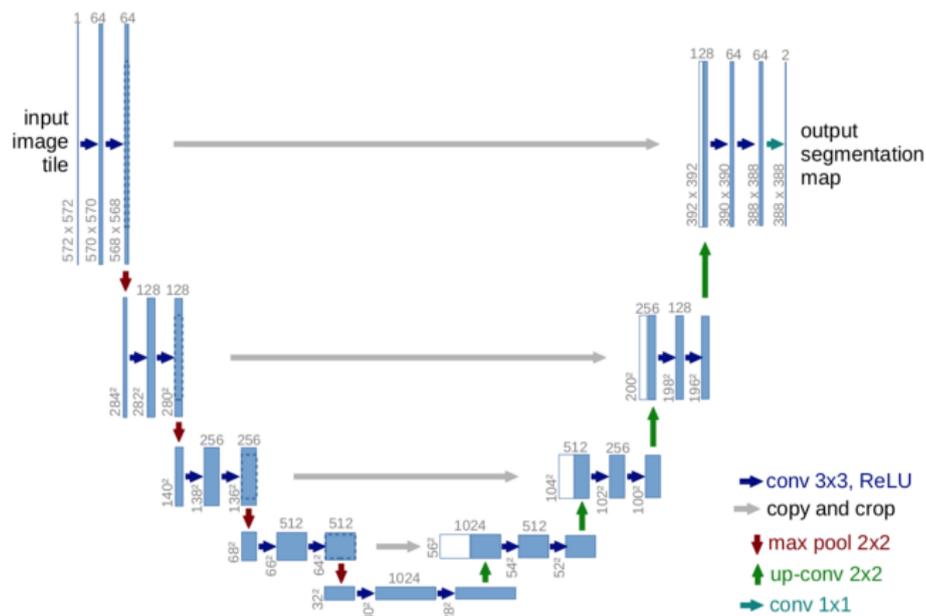


Figure (2.11): Architecture U-Net.

- Chaque rectangle bleu correspond à une carte de caractéristiques multicanaux. Le nombre de canaux est indiqué en haut du rectangle
- Il a une forme en "U". L'architecture est symétrique et se compose de deux parties principales
- la partie gauche est appelée chemin de contraction, qui est constitué par le processus convolution général ; la partie droite est le chemin expansif, qui est constitué par des couches convolutions 2d transposées.

- 1) Le premier bloc est aussi appelé l'encodeur et sert à récupérer le contexte de l'image. Ce bloc se compose de couches convolution et de couches de regroupement maximal pour capturer les caractéristiques de l'image et réduire sa taille pour réduire le nombre de paramètres réseau. Cela implique l'application répétée de deux couches convolution 3x3. Chaque couche est suivie d'une fonction d'activation ReLU et d'une normalisation par lots. [19] Ensuite, une opération de regroupement 2x2 max est appliquée pour réduire la dimension spatiale.
- 2) Un pont ou un goulot d'étranglement relie les réseaux d'encodeurs et de décodeurs pour compléter le flux d'informations. Il se compose de deux couches de convolutions 3x3, chacune suivie d'une fonction d'activation ReLU. Le deuxième bloc est le bloc du décodeur. Grâce à la convolution transposée, il permet une localisation précise et retrouve également la taille initiale de l'image. Le bloc décodeur sur échantillonne d'abord la carte des caractéristiques, suivi d'une couche convolutive 2x2 transposée. Ensuite, utilisez deux couches de convolutions 3x3, où chaque convolution est suivie d'une fonction d'activation ReLU. La sortie du dernier décodeur passe par une couche convolutive 1x1 avec une fonction d'activation sigmoïde

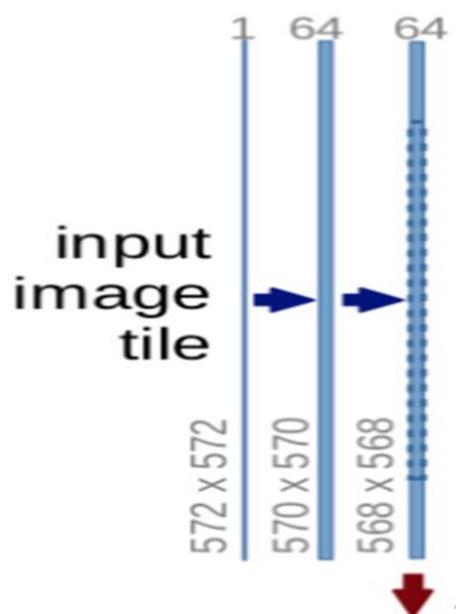


Figure (2.12) : Architecture U-Net couches de convolution

- Notez que chaque processus constitue deux couches de convolution, et le nombre de canaux change de $1 \rightarrow 64$, car le processus de convolution augmentera la profondeur de l'image. La flèche rouge pointant vers le bas est le processus de max pooling qui réduit de moitié la taille de l'image (la taille réduite de $572 \times 572 \rightarrow 568 \times 568$ est due à des problèmes de padding, mais l'implémentation ici utilise padding="same").

Ce processus est répété 3 fois de plus

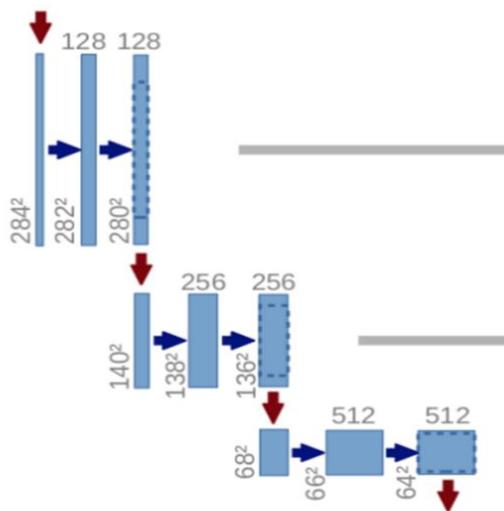


FIGURE (2.13): Architecture U-Net couches de convolution

Et nous arrivons au plus bas :



Figure (2.14): Architecture U-Net couches de convolution

- L'image à ce moment a été redimensionnée à $28 \times 28 \times 1024$. Maintenant, passons au chemin expansif.
- Dans le chemin expansif, l'image va être redimensionnée à sa taille originale.



Figure (2.15) : Architecture U-Net chemin expansif.

- La convolution transposée est une technique de suréchantillonnage qui agrandit la taille des images. Il y a une démo visualisée ici et une explication ici. Fondamentalement, elle effectue un remplissage de l'image originale suivi d'une opération de convolution. Après cette convolution transposée, l'image est agrandie de $28 \times 28 \times 1024 \rightarrow 56 \times 56 \times 512$, et ensuite, cette image est concaténée avec l'image correspondante du chemin de contraction et ensemble, elles forment une image de taille $56 \times 56 \times 1024$. La raison ici est de combiner les informations des couches précédentes afin d'obtenir une prédiction plus précise. Comme précédemment, ce processus est répété 3 fois de plus
- La dernière étape consiste à remodeler l'image pour répondre à nos besoins de prédiction.

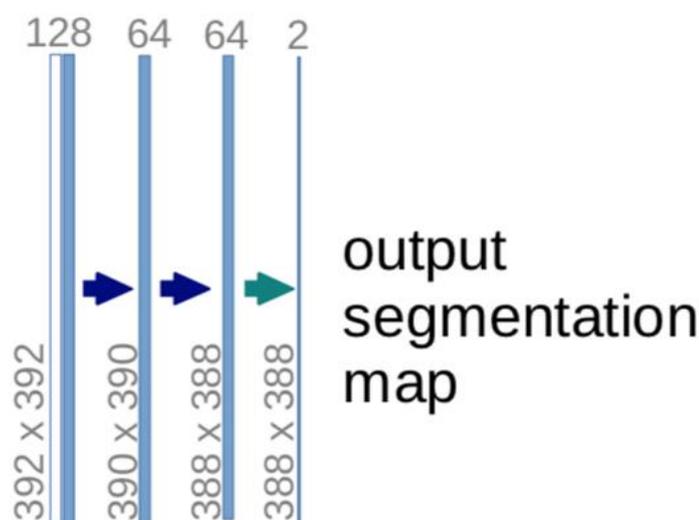


Figure (2.16) : Architecture U-Net couche de convolution avec un filtre de taille 1x1

- La dernière est une couche de convolution avec un filtre de taille 1x1 (notez qu'il n'y a pas de couche dense dans tout le réseau). Et le reste est le même pour la formation du réseau neuronal.

6. Conclusion :

Le réseau U-Net est capable de localiser des images en prédisant l'image pixel par pixel .le réseau est suffisamment puissant pour réaliser des prédictions de qualité à partir de quelques ensembles de données en utilisant des techniques d'augmentation excessive des données.

L'augmentation des données excessives. Il présente aussi une précision supérieure à celle des modèles de convolution.

En plus de sa grande robustesse, l'U-Net a une architecture simple fondée sur des auto-encodeurs et sur des couches de convolution facilement applicable.

Il y a plusieurs variantes de ce réseau adaptées à différentes situations, certaines sont très précises, d'autres sont très rapides.

Chapitre III :

Travail personnel

1. Introduction :

L'utilisation des images UAV pour la détection du réseau routier dans des plusieurs applications notamment civiles ou militaires a bénéficié d'un grand intérêt pour les chercheurs. Les routes sont un élément important aux infrastructures urbaines. Par conséquent, leur détection et leur modélisation est une étape essentielle dans la segmentation sémantique des images UAV.

Dans ce travail on va utiliser une méthode de segmentation à l'aide de réseaux de neurones convolutifs sur les images acquises par drone à haute résolution spatiale.

Le travail se résume comme suit :

- Concevoir notre propre réseau basique d'encodeur –décodeur du modèle U-Net. Nous proposons deux architectures U net (avec 5 couches de convolutions) et un deuxième moins profond ((avec deux convolutions).
- Choisir les paramètres optimaux de la phase d'apprentissage. Car la phase d'apprentissage conditionnera les résultats de la segmentation.
- Vu la taille des images on procèdera à l'apprentissage sur des patches de taille 32*32 sélectionnés aléatoirement sur l'image.
- Segmentation sur l'image test en partitionnant l'image test sur des fenêtres de la taille choisie. La segmentation se fera individuellement sur toutes les fenêtres de l'image.
- Le résultat de la segmentation sera affiné grâce au filtrage médian pour homogénéiser le résultat. (enlever le bruit impulsionnel)
- test et comparaison des résultats des deux modèles à l'aide des paramètres de scores calculés et définis plus bas.

2. Base de données :

Lors de l'apprentissage du test du réseau de segmentation sémantique, nous avons besoin d'un ensemble d'images et de l'ensemble correspondant d'images étiquetées (groundtruth), où chaque valeur de pixel dans cette dernière, représente l'étiquette de ce pixel dans l'image réelle correspondante. Nous évaluons les deux réseaux sur une base de données d'images de drones composée d'images semi-urbaines et non urbaines collectées sur Internet.

Ils se composent de plusieurs routes pavées et non pavées avec différentes résolutions, largeurs et formes. La couleur, l'éclairage et le contraste de l'image de l'itinéraire actuel changent, dans les 101 images UAV utilisées.

On peut télécharger cette base de données : <https://sites.google.com/site/hailingzhouwei/>.

3. Évaluation des méthodes :

Les techniques d'analyse des données utilisent habituellement Comme point de départ une représentation sous forme de tableau, que l'on appelle matrice de confusion ou tableau de contingence, mettant en relation le nombre de prédictions " $X_i ; j$ " (Classe ; prédiction) des échantillons de la classe "i" attribués à une classe "j" (parmi C classes).

Le nombre d'échantillons formant la classe "i" est noté " K_i ", et le nombre total de prédictions

Attribuées à cette classe est nommé " M_i ". La somme des " K_i " ainsi que la somme des " M_i " correspondant au nombre total d'échantillons. [20]

		prédictions			Total / classes
		classes 1	Classe i	Classe C	
Classe réelle	Classe 1	$X_1 ; 1$	$X_1 ; i$	$X_1 ; C$	K_1
	Classe i	$X_i ; 1$	$X_i ; i$	$X_i ; C$	K_i
	Classe C	$X_C ; 1$	$X_C ; i$	$X_C ; C$	K_C
Total		M_1	M_i	M_C	Σ

Tableau (3.1) : Matrice de confusion des prédictions d'un classifieur à C classes

En considérant le problème précédent, pour chacune des classes "i", comme étant binaire (Classe "i" : positif ; Toutes les autres classes "j, i" : négatif), ou directement dans le

cas d'un problème à deux classes, les prédictions sont résumées sous la forme de quatre informations principales :

- Vrais Positifs - VP ("True Positive" - TP) : Échantillons de la classe positive ("i") correctement classés (X_i, i)
- Faux Négatifs - FN ("False Négative" - FN) : Échantillons de la classe positive ("i") incorrectement classés ($X_i, j, \forall j, i$)
- Vrais Négatifs - VN ("True Négative" - TN) : Échantillons des classes négatives ("j") correctement classés ($X_{j,t}, \forall (j, t) \in [1, C], i$)
- Faux Positifs - FP ("False Positive" - FP) : Échantillons des classes négatives ("j") incorrectement classés ($X_{j,i}, i, i, \forall j, i$)

Pour un problème à N classes, considéré comme binaire pour le test de chacune des classes, il y a autant de matrices de confusion qu'il y a de classes, chacune représentant les performances de la classification vis à vis de la classe "i".

La matrice de confusion pour un problème à deux classes établit ainsi le lien entre le nombre d'échantillons total (P) de la classe positive, le nombre d'échantillons total (N) de la classe négative, les quatre informations précédentes et par conséquent le nombre total d'échantillons classés positivement (Ppos) et négativement (Pneg). [21]

Cette matrice de confusion, tableau (3.2), donne une indication sur les résultats de classification.

Cependant, du fait que la quantité d'échantillons par classe peut varier entre les différentes classes et des différentes problématiques, son exploitation n'est pas toujours facile.

		prédictions		Total / classes
		classes 1	Classe i	
Classe réelle	Classe positive	VP	FN	P
	Classe négative	FP	VN	N
Total prédictions		Ppos	Pneg	Σ

Tableau (3.2) : Matrice de confusion des prédictions d'un classifieur à deux classes

Des critères de performances dérivent de cette matrice de confusion et permettent d'évaluer rapidement les performances de classification.

Ces nouvelles métriques représentent des indications normalisées de ces performances sous forme d'un pourcentage. [22]

Il est ainsi possible d'analyser et de comparer les performances de divers classifieurs. Ces mesures sont les suivantes :

Sensibilité ou Rappel :

Proportion de solutions pertinentes correctement déterminées (taux positif réel). Mesure l'aptitude du système à fournir toutes les solutions appropriées.

$$\text{Sensibilité} = \frac{VP}{VP+FN} = \frac{VP}{P} \tag{3.1}$$

Spécificité (Spécificité) :

Part des solutions négatives correctement identifiées (taux négatif réel). Mesure la capacité du système de prendre en compte toutes les situations négatives. C'est un indicateur complémentaire de sensibilité.

$$\text{Spécificité} = \frac{VN}{VN+FP} = \frac{VN}{N} \quad (3.2)$$

Précision (Précision) :

Proportion des solutions exactes parmi les prévisions positives. Mesure la capacité du système de refuser des solutions sans importance.

$$\text{Précision} = \frac{VP}{VP+FP} = \frac{VP}{P_{\text{pos}}} \quad (3.3)$$

Robustesse, exactitude ou justesse (Accuracy) :

Pourcentage de solutions correctes (positives et négatives) dans toutes les prédictions. Mesure la capacité de l'ensemble du système pour donner toutes les bonnes solutions.

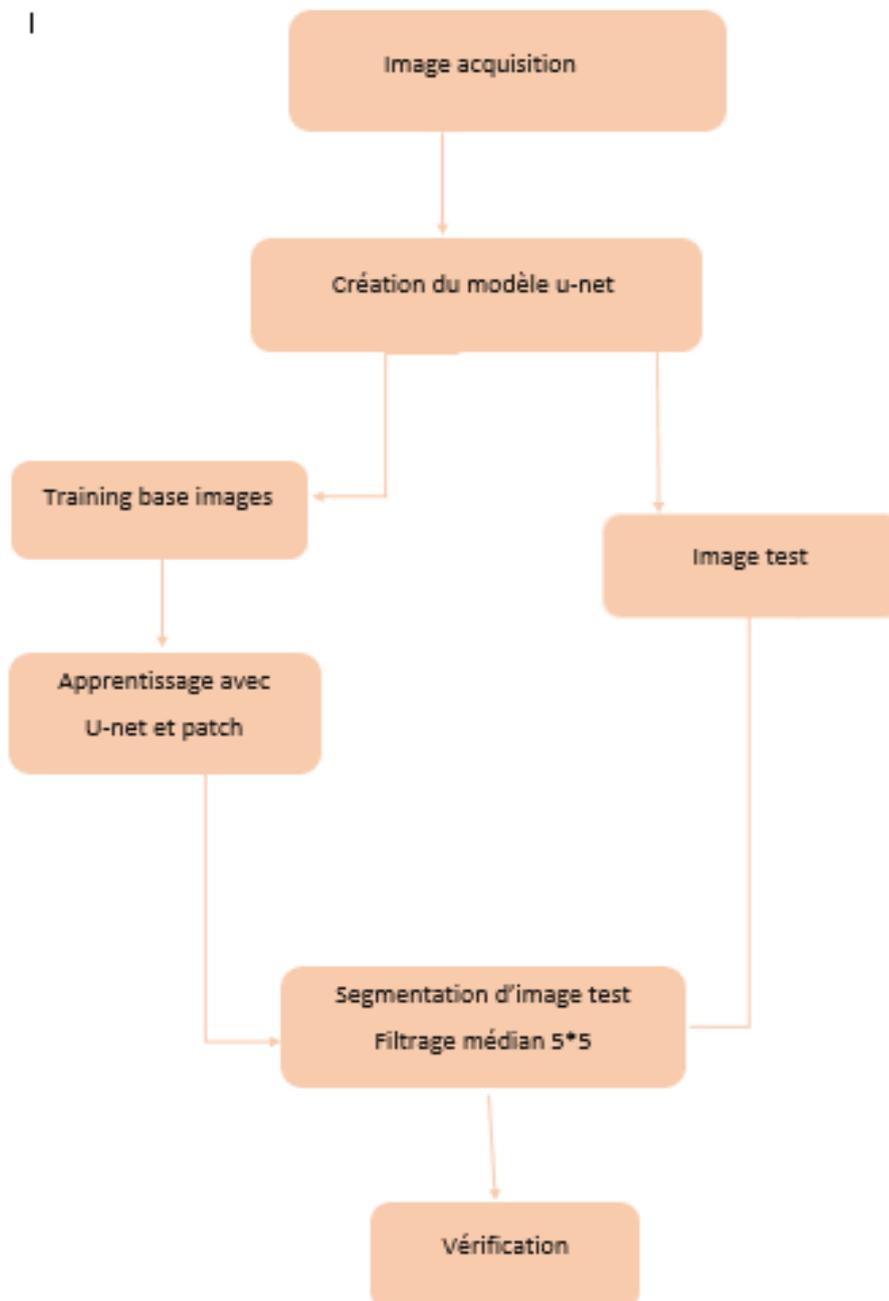
$$\text{Justesse} = \frac{VP + VN}{VP + VN + FP + FN} = \frac{P}{\Sigma} \Rightarrow \frac{\text{Échantillons correctement attribués}}{\text{Totalité des échantillons}} \quad (3.4)$$

F -mesure (F -measure) or F-Score:

Allie l'exactitude et la mémoire grâce à leur moyenne harmonique. Offre une mesure qui favorise soit le rappel, soit l'exactitude dans le cas de la classification fondée sur les données non équilibrées

$$F\beta = \frac{(1 + \beta^2)(\text{Precision:Rappel})}{\beta^2 \cdot \text{Precision:Rappel}} \quad (3.5)$$

4. Principe de fonctionnement :



5. Résultats et interprétation :

5.1. Partie 01 : génération du premier modelé Unet1

La segmentation sémantique de l'image test a été réalisée sur deux modèles Unet .le premier modèle unet-1 est montré sur la figure ci dessous.il comprend

Les couches suivantes :

'ImageInputLayer'	Image Input	32x32x3 images with 'zerocenter' normalization
'Encoder-Section-1-Conv-1'	Convolution	64 3x3x3 convolutions with stride [1 1] and padding [1 1 1 1]
'Encoder-Section-1-ReLU-1'	ReLU	ReLU
'Encoder-Section-1-Conv-2'	Convolution	64 3x3x64 convolutions with stride [1 1] and padding [1 1 1 1]
'Encoder-Section-1-ReLU-2'	ReLU	ReLU
'Encoder-Section-1-MaxPool'	Max Pooling	2x2 max pooling with stride [2 2] and padding [0 0 0 0]
'Encoder-Section-2-Conv-1'	Convolution	128 3x3x64 convolutions with stride [1 1] and padding [1 1 1 1]
'Encoder-Section-2-ReLU-1'	ReLU	ReLU
'Encoder-Section-2-Conv-2'	Convolution	128 3x3x128 convolutions with stride [1 1] and padding [1 1 1 1]
'Encoder-Section-2-ReLU-2'	ReLU	ReLU
'Encoder-Section-2-MaxPool'	Max Pooling	2x2 max pooling with stride [2 2] and padding [0 0 0 0]
'Encoder-Section-3-Conv-1'	Convolution	256 3x3x128 convolutions with stride [1 1] and padding [1 1 1 1]
'Encoder-Section-3-ReLU-1'	ReLU	ReLU
'Encoder-Section-3-Conv-2'	Convolution	256 3x3x256 convolutions with stride [1 1] and padding [1 1 1 1]
'Encoder-Section-3-ReLU-2'	ReLU	ReLU
'Encoder-Section-3-MaxPool'	Max Pooling	2x2 max pooling with stride [2 2] and padding [0 0 0 0]
'Encoder-Section-4-Conv-1'	Convolution	512 3x3x256 convolutions with stride [1 1] and padding [1 1 1 1]
'Encoder-Section-4-ReLU-1'	ReLU	ReLU
'Encoder-Section-4-Conv-2'	Convolution	512 3x3x512 convolutions with stride [1 1] and padding [1 1 1 1]
'Encoder-Section-4-ReLU-2'	ReLU	ReLU
'Encoder-Section-4-DropOut'	Dropout	50% dropout
'Encoder-Section-4-MaxPool'	Max Pooling	2x2 max pooling with stride [2 2] and padding [0 0 0 0]
'Mid-Conv-1'	Convolution	1024 3x3x512 convolutions with stride [1 1] and padding [1 1 1 1]
'Mid-ReLU-1'	ReLU	ReLU
'Mid-Conv-2'	Convolution	1024 3x3x1024 convolutions with stride [1 1] and padding [1 1 1 1]
'Mid-ReLU-2'	ReLU	ReLU
'Mid-DropOut'	Dropout	50% dropout
'Decoder-Section-1-UpConv'	Transposed Convolution	512 2x2x1024 transposed convolutions with stride [2 2] and cropping [0 0 0 0]
'Decoder-Section-1-UpReLU'	ReLU	ReLU
'Decoder-Section-1-DepthConcatenation'	Depth concatenation	Depth concatenation of 2 inputs
'Decoder-Section-1-Conv-1'	Convolution	512 3x3x1024 convolutions with stride [1 1] and padding [1 1 1 1]
'Decoder-Section-1-ReLU-1'	ReLU	ReLU

'Decoder-Section-1-ReLU-1'	ReLU	ReLU
'Decoder-Section-1-Conv-2'	Convolution	512 3x3x512 convolutions with stride [1 1] and padding [1 1 1 1]
'Decoder-Section-1-ReLU-2'	ReLU	ReLU
'Decoder-Section-2-UpConv'	Transposed Convolution	256 2x2x512 transposed convolutions with stride [2 2] and cropping [0 0 0 0]
'Decoder-Section-2-UpReLU'	ReLU	ReLU
'Decoder-Section-2-DepthConcatenation'	Depth concatenation	Depth concatenation of 2 inputs
'Decoder-Section-2-Conv-1'	Convolution	256 3x3x512 convolutions with stride [1 1] and padding [1 1 1 1]
'Decoder-Section-2-ReLU-1'	ReLU	ReLU
'Decoder-Section-2-Conv-2'	Convolution	256 3x3x256 convolutions with stride [1 1] and padding [1 1 1 1]
'Decoder-Section-2-ReLU-2'	ReLU	ReLU
'Decoder-Section-3-UpConv'	Transposed Convolution	128 2x2x256 transposed convolutions with stride [2 2] and cropping [0 0 0 0]
'Decoder-Section-3-UpReLU'	ReLU	ReLU
'Decoder-Section-3-DepthConcatenation'	Depth concatenation	Depth concatenation of 2 inputs
'Decoder-Section-3-Conv-1'	Convolution	128 3x3x256 convolutions with stride [1 1] and padding [1 1 1 1]
'Decoder-Section-3-ReLU-1'	ReLU	ReLU
'Decoder-Section-3-Conv-2'	Convolution	128 3x3x128 convolutions with stride [1 1] and padding [1 1 1 1]
'Decoder-Section-3-ReLU-2'	ReLU	ReLU
'Decoder-Section-4-UpConv'	Transposed Convolution	64 2x2x128 transposed convolutions with stride [2 2] and cropping [0 0 0 0]
'Decoder-Section-4-UpReLU'	ReLU	ReLU
'Decoder-Section-4-DepthConcatenation'	Depth concatenation	Depth concatenation of 2 inputs
'Decoder-Section-4-Conv-1'	Convolution	64 3x3x128 convolutions with stride [1 1] and padding [1 1 1 1]
'Decoder-Section-4-ReLU-1'	ReLU	ReLU
'Decoder-Section-4-Conv-2'	Convolution	64 3x3x64 convolutions with stride [1 1] and padding [1 1 1 1]
'Decoder-Section-4-ReLU-2'	ReLU	ReLU
'Final-ConvolutionLayer'	Convolution	2 1x1x64 convolutions with stride [1 1] and padding [0 0 0 0]
'Softmax-Layer'	Softmax	softmax
'Segmentation-Layer'	Pixel Classification Layer	Cross-entropy loss

Figure (3.1) : Le premier modèle unet-1

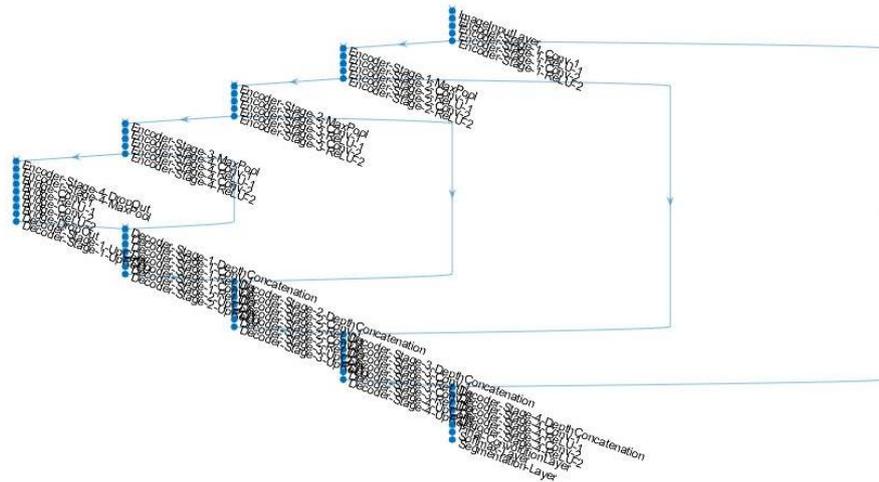


Figure (3.2) : Architecture du Modèle encodeur –décodeur U-net1

5.2. Partie 02 : training du modèle Unet1

TEST 1 :

- Nous opérons sur une base de données de 101 images, comme le montre ce tableau, qui explique également les résultats d'exploitation du réseau U-NET en fonction du temps d'apprentissage.

Epoch	Iteration	Time Elapsed (hh:mm:ss)	Mini-batch Accuracy	Mini-batch Loss	Base Learning Rate
1	1	00:00:25	82.58%	1.9738	0.0010
1	50	00:03:43	91.22%	0.2372	0.0010
1	100	00:06:41	92.38%	0.2109	0.0010
1	150	00:09:42	89.89%	0.1955	0.0010
1	200	00:12:53	86.49%	0.3573	0.0010
2	250	00:16:14	84.86%	0.5364	0.0010
2	300	00:19:42	95.76%	0.1413	0.0010

	2		350		00:23:06		92.55%		0.1664		0.0010	
	2		400		00:26:24		95.72%		0.1846		0.0010	
	3		450		00:29:35		93.12%		0.1481		0.0010	
	3		500		00:32:41		94.82%		0.1308		0.0010	
	3		550		00:35:46		91.69%		0.1800		0.0010	
	3		600		00:38:50		97.29%		0.1249		0.0010	
	4		650		00:41:56		95.54%		0.1469		0.0010	
	4		700		00:45:06		94.57%		0.1272		0.0010	
	4		750		03:27:35		96.47%		0.1157		0.0010	
	4		800		03:30:35		97.60%		0.0763		0.0010	
	5		850		03:33:30		95.52%		0.1044		0.0010	
	5		900		03:36:25		92.40%		0.1894		0.0010	
	5		950		03:39:24		91.75%		0.3367		0.0010	
	5		1000		03:42:45		98.88%		0.0472		0.0010	
	6		1050		03:46:48		87.95%		0.3271		0.0010	
	6		1100		03:50:46		91.48%		0.4692		0.0010	
	6		1150		03:54:43		96.19%		0.0909		0.0010	
	6		1200		03:58:50		95.83%		0.1008		0.0010	
	7		1250		04:02:50		96.95%		0.1119		0.0010	
	7		1300		04:07:00		98.25%		0.0700		0.0010	
	7		1350		04:10:51		96.14%		0.0933		0.0010	
	7		1400		04:14:52		99.30%		0.0439		0.0010	
	8		1450		04:18:38		98.79%		0.0738		0.0010	
	8		1500		04:21:55		98.60%		0.0444		0.0010	
	8		1550		04:25:12		94.32%		0.1666		0.0010	
	8		1600		04:28:30		73.24%		0.9357		0.0010	
	9		1650		04:31:46		81.42%		0.5762		0.0010	
	9		1700		04:35:03		98.25%		0.0481		0.0010	
	9		1750		04:38:19		99.06%		0.0273		0.0010	
	9		1800		04:42:01		99.05%		0.0445		0.0010	
	10		1850		04:45:50		89.31%		0.4170		0.0010	
	10		1900		04:49:44		77.49%		1.1949		0.0010	

	10		1950		04:53:23		97.88%		0.0512		0.0010	
	10		2000		04:57:08		99.04%		0.0387		0.0010	
	11		2050		05:00:40		91.81%		0.2824		0.0010	
	11		2100		05:04:15		99.34%		0.0248		0.0010	
	11		2150		05:07:54		98.90%		0.0291		0.0010	
	11		2200		05:11:28		95.21%		0.2471		0.0010	
	12		2250		05:15:33		93.27%		0.3901		0.0010	
	12		2300		05:19:32		98.78%		0.0411		0.0010	
	12		2350		05:23:27		96.94%		0.0923		0.0010	
	12		2400		05:27:18		99.10%		0.0458		0.0010	
	13		2450		05:31:15		98.79%		0.0373		0.0010	
	13		2500		05:35:00		97.20%		0.0891		0.0010	
	13		2550		05:38:55		99.32%		0.0266		0.0010	
	13		2600		05:42:27		97.81%		0.0610		0.0010	
	14		2650		05:46:10		98.84%		0.0313		0.0010	
	14		2700		05:49:40		99.08%		0.0306		0.0010	
	14		2750		05:53:28		98.67%		0.0471		0.0010	
	14		2800		05:57:15		94.52%		0.1647		0.0010	
	15		2850		06:00:58		98.54%		0.0465		0.0010	
	15		2900		06:04:26		99.21%		0.0369		0.0010	
	15		2950		06:07:53		99.50%		0.0371		0.0010	
	15		3000		06:11:36		99.48%		0.0287		0.0010	
	16		3050		06:15:42		99.43%		0.0226		0.0010	
	16		3100		06:19:19		97.79%		0.0671		0.0010	
	16		3150		06:22:29		99.00%		0.0363		0.0010	
	16		3200		06:25:44		99.17%		0.0503		0.0010	
	17		3250		06:29:01		96.52%		0.0977		0.0010	
	17		3300		06:32:26		94.26%		0.1508		0.0010	
	17		3350		06:36:00		99.10%		0.0252		0.0010	
	17		3400		06:39:23		99.40%		0.0264		0.0010	
	18		3450		06:43:00		96.77%		0.0850		0.0010	
	18		3500		06:46:31		98.05%		0.0623		0.0010	

	18		3550		06:50:03		99.48%		0.0198		0.0010	
	18		3600		06:54:02		98.68%		0.0398		0.0010	
	19		3650		06:59:44		97.72%		0.0620		0.0010	
	19		3700		07:02:58		96.45%		0.0977		0.0010	
	19		3750		07:06:19		97.69%		0.0649		0.0010	
	19		3800		07:09:34		96.00%		0.1309		0.0010	
	20		3850		07:12:47		99.73%		0.0221		0.0010	
	20		3900		07:15:59		99.10%		0.0320		0.0010	
	20		3950		07:19:12		99.13%		0.0375		0.0010	
	20		4000		07:22:25		89.38%		0.2150		0.0010	
	20		4040		07:25:01		82.85%		0.5607		0.0010	

Tableau (3.3): Le rapport d'entrainement pour PC1

Lors du premier test, les résultats ont pris 7 heures et 25 minutes (accuracy) et se termineront à 82,85 %. Sachant que nous avons utilisé un ordinateur avec un processeur Intel(R) Core(TM) i3-2310M à 2,10 GHz et 4,00 Ram.

TESTE 2 :

Nous avons remplacé l'ordinateur par un processeur Intel(R) Core(TM) i5-7400M fonctionnant à 3,00 GHz et 8,00 Go de RAM.

L'apprentissage (précision) s’est terminé à 99,39% avec une durée d'exécution de 01 heures et 17 minutes.

En conséquence, le PC joue une influence critique dans la fiabilité des résultats.

Le temps de traitement a été réduit grâce à la capacité de traitement de la machine.

⇒ Les résultats dépendent en fonction des parametres de training choisis :

Le tableau résume les résultats de la segmentation avec UNET 1 défini plus bas.

Pour optimiser les résultats, nous l'avons exécuté plusieurs fois en changeant les paramètres d'apprentissage qui sont :

✚ Max Epoch : Nombre maximal d'époques à utiliser pour l'entraînement, spécifié sous la forme d'un entier positif. Une itération est une étape franchie dans l'algorithme de descente

de gradient vers la minimisation de la fonction de perte à l'aide d'un mini-batch. Une époque est le passage complet de l'algorithme d'apprentissage sur l'ensemble d'apprentissage.

✚ Mini batch : Taille du mini-batch à utiliser pour chaque itération d'entraînement, spécifiée sous la forme d'un entier positif. Un mini-lot est un sous-ensemble de l'ensemble d'apprentissage utilisé pour évaluer le gradient de la fonction de perte et mettre à jour les poids. Si la taille du mini-batch ne divise pas uniformément le nombre d'échantillons d'apprentissage, alors train Network rejette les données d'apprentissage qui ne rentrent pas dans le mini-lot complet final de chaque époque

✚ Size patch : Taille de patch, spécifiée comme l'une des valeurs suivantes.

Un vecteur à 2 éléments d'entiers positifs pour les patches 2D. Patch Size a la forme [r c] où r spécifie le nombre de lignes et c spécifie le nombre de colonnes dans le patch.

Un vecteur à 3 éléments d'entiers positifs pour les patches 3D. Patch Size a la forme [r c p] où r spécifie le nombre de lignes, c spécifie le nombre de colonnes et p spécifie le nombre de plans dans le patch.

Image	Max-Epoch	Mini-Epoch	Size patch	Accuracy	loss
10	15	8	[32*32]	98%	0.0541
10	20	8	[32*32]	82.8 %	0.5607
10	30	8	[32*32]	99.1%	0.0502
10	20	16	[32*32]	97.66%	0.0802
10	20	8	[64*64]	98.85 %	0.0499
10	30	16	[32*32]	79.33 %	0.5967

Tableau (3.4) : Les résultats des paramètres d'apprentissage pour image 10

⇒ *Size patch [32*32] MAX-EPOCH 20 MINI-PATCH 8 :*

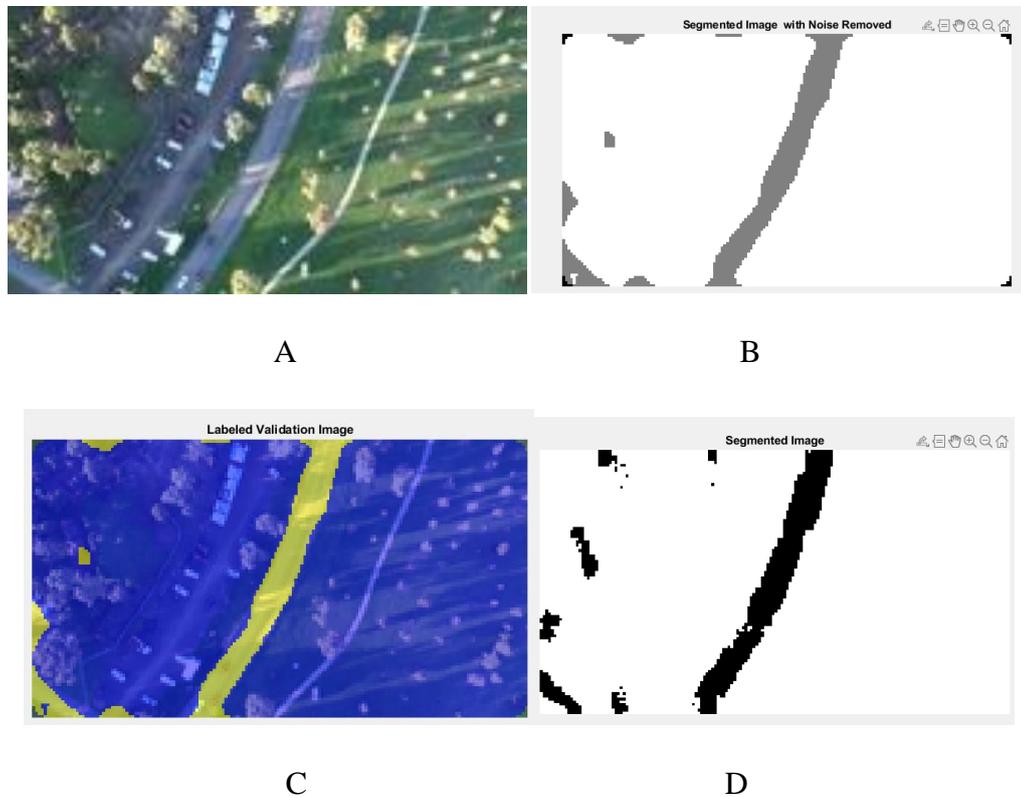


Figure (3.3): image 10 avec parameter *Size patch [32*32] MAX-EPOCH 20 MINI-PATCH 8 (A, B, C, D)*

A : image originale

B : image segmentée après filtrage

C : image superposée sur l'image originale

D : image segmentée avant filtrage

⇒ Size patch [64*64] MAX-EPOCH 20 MINI-PATCH 8 :

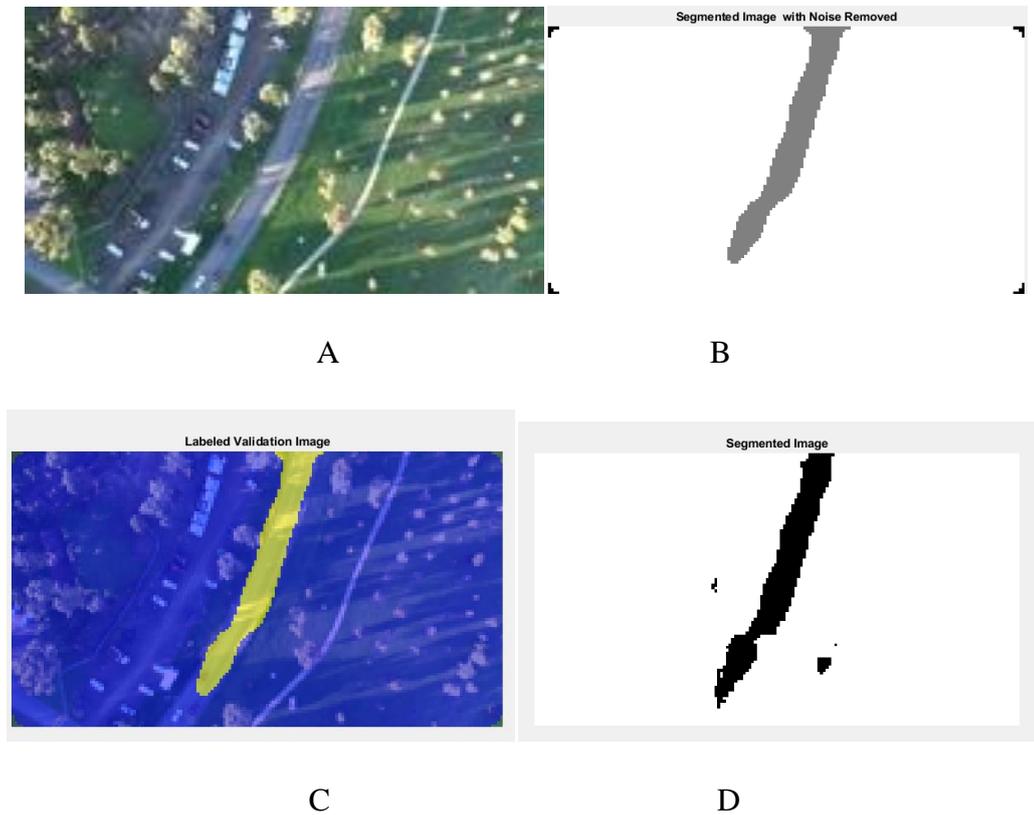


Figure (3.4): image 10 avec parameter Size patch [64*64] (A, B, C, D)

A : image originale

B : image segmentée après filtrage

C : image superposée sur l'image originale

D : image segmentée avant filtrage

⇒ Size patch [32*32] MAX-EPOCH 15 MINI-PATCH 8:

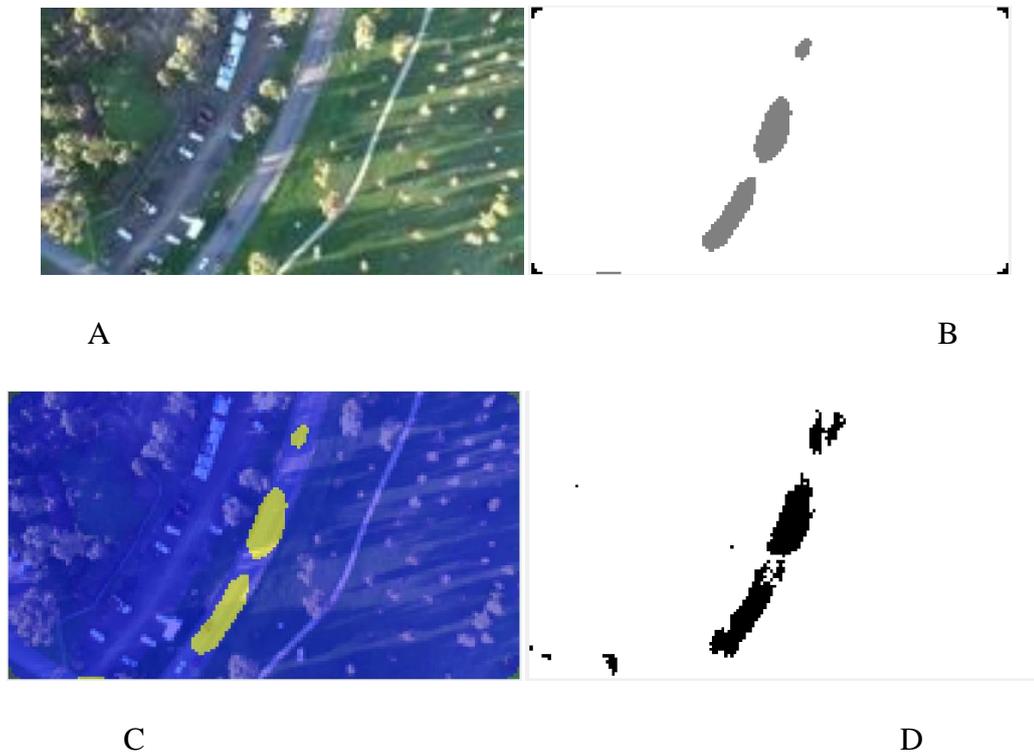


Figure (3.5): image 10 avec parameter MAX-EPOCH 15 MINI-PATCH 8 (A, B, C, D)

A : image originale

B : image segmentée après filtrage

C : image superposée sur l'image originale

D : image segmentée avant filtrage

⇒ Size patch [32*32] MAX-EPOCH 30 MINI-PATCH 8 :

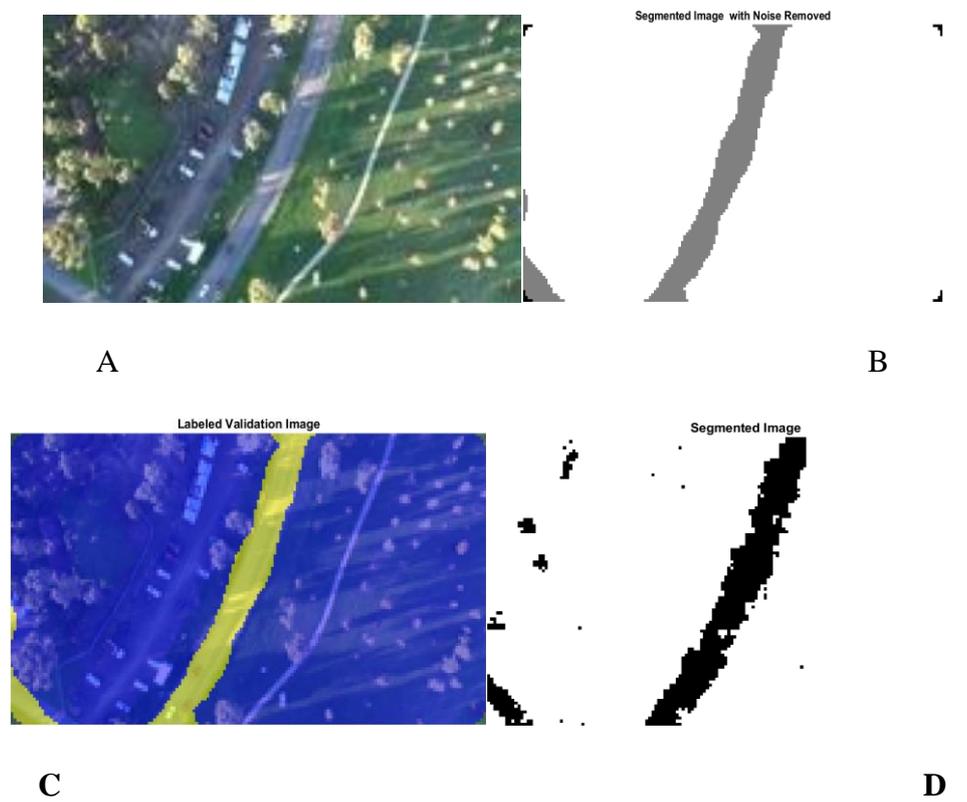


Figure (3.6): image 10 avec parameter MAX-EPOCH 30 MINI-PATCH 8 (A, B, C, D)

A : image originale

B : image segmentée après filtrage

C : image superposée sur l'image originale

D : image segmentée avant filtrage

⇒ Size patch [32*32] MINI-PATCH 16 MAX-EPOCH 20 :

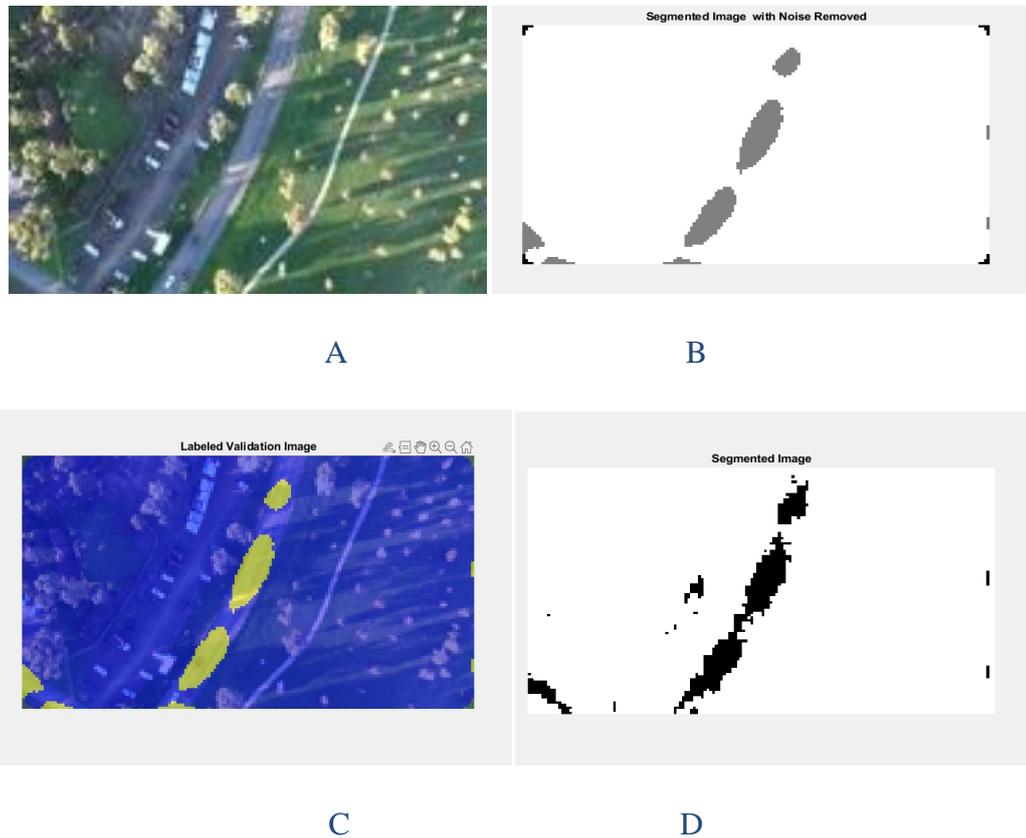


Figure (3.7): image 10 avec parameter MINI-PATCH 16 MAX-EPOCH 20 (A,B,C,D)

A : image originale

B : image segmentée après filtrage

C : image superposée sur l'image originale

D : image segmentée avant filtrage

⇒ Size patch [32*32] MAX-EPOCH 30 MINI-PATCH 16 :

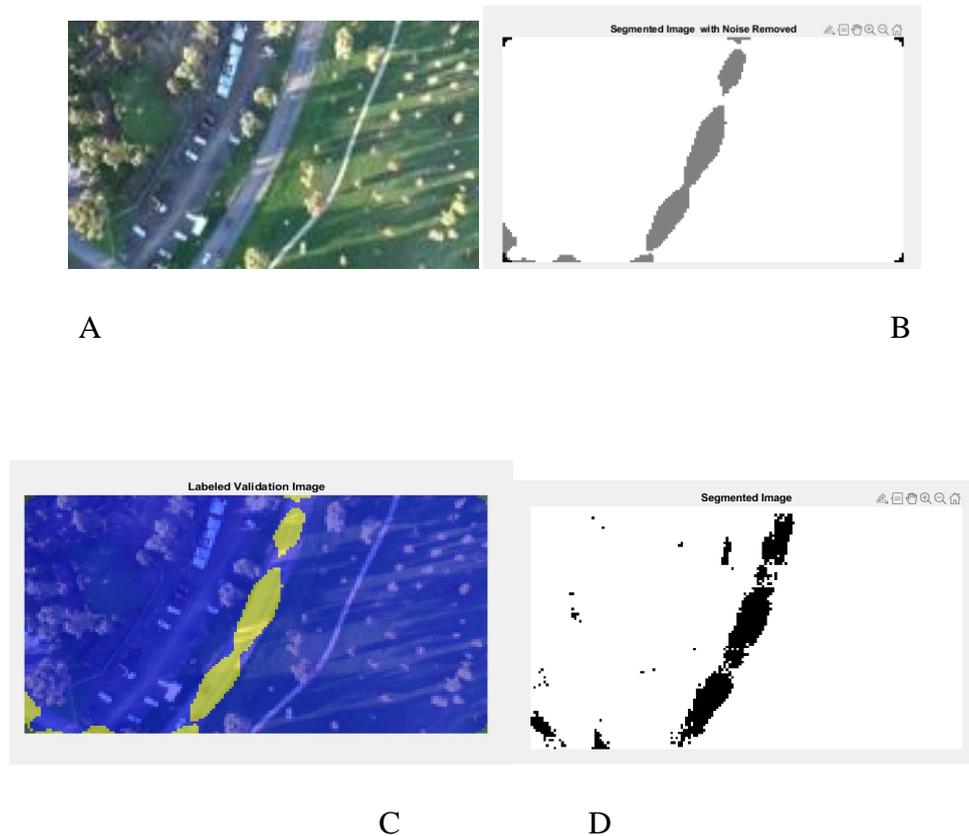


Figure (3.8): image 10 avec parameter MAX-EPOCH 30 MINI-PATCH 16 (A, B,C,D)

A : image originale

B : image segmentée après filtrage

C : image superposée sur l'image originale

D : image segmentée avant filtrage

- **Remarque :**

✚ les résultats varient largement en fonction de paramètres. Nous voyons que le réglage du paramètre (Epoch max) à 30 nous donne un excellent résultat B C D. la classe donnée à une voie correspond à la classe réelle avec une précision de 99,10%.

✚ Le choix des paramètres influe sur le temps de l'étape d'apprentissage. Un compromis est nécessaire entre le temps de training et le maxepoch, le mini batch le nombre d'images d'apprentissage et aussi la taille des images traitées.

5.2 Partie 03 : test sur le modèle Unet 2

Nous avons donc créé un deuxième modèle U-net2 présenté sur la figure (3.9) (plus petit CAD diminué les couches de convolution déconvolution pooling) :

Image Input	648x1152x3 images with 'zerocenter' normalization
Convolution	64 3x3 convolutions with stride [1 1] and padding [1 1 1 1]
ReLU	ReLU
Max Pooling	2x2 max pooling with stride [2 2] and padding [0 0 0 0]
Convolution	64 3x3 convolutions with stride [1 1] and padding [1 1 1 1]
ReLU	ReLU
Transposed Convolution	64 4x4 transposed convolutions with stride [2 2] and cropping [1 1 1 1]
Convolution	3 1x1 convolutions with stride [1 1] and padding [0 0 0 0]
Softmax	softmax
Pixel Classification Layer	Cross-entropy loss

Figure (3.9): deuxième modèle U-net2

- Et voici le résultat obtenu en utilisant la même base de données contenant 101 images et une sélection aléatoire des images 10 et 330 pour le test :
- Nous avons utilisé les meilleurs paramètres d'apprentissage cités plus haut ; puis les scores de la segmentation sont répertoriés sur le tableau suivant :

Image	Accurac y	Sensitivit y	F measure	Precisio n	MCC	Dice	Jaccard	Specitivity
10 Unet -2	0.1230	0.0114	0.0017	9.0314e- 04	-0.5411	0.0017	8.3770e- 04	0.1306
10 Unet-1	0.9357	0	0	-	-	0	0	1

Tableau (3.5) : comparaison des résultats entre le modèle unet-1 et unet-2



Figure (3.10) : Architecture du Modèle2 encodeur –décodeur U-net2

⇒ **Résultat de la segmentation :**

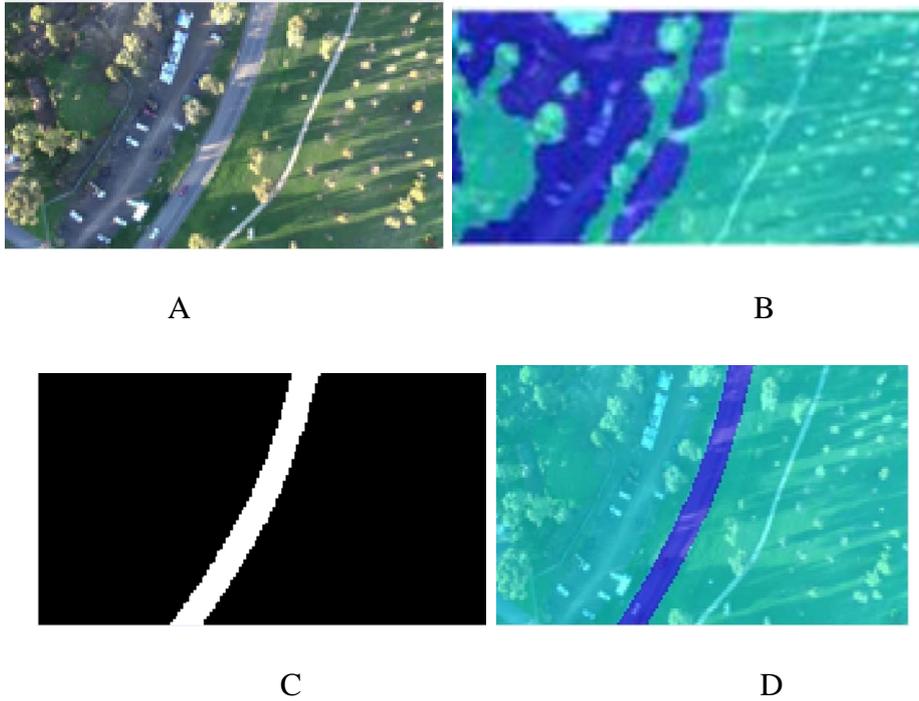


Figure (3.11) : image 10 (A, B, C, D)

A. : image couleur originale

B. : image segmentée

C. : image groundtruth

D. : image originale juxtaposée avec le groundtruth en cyan

⇒ **Résultat de la segmentation avec U-Net 2 image 330 :**

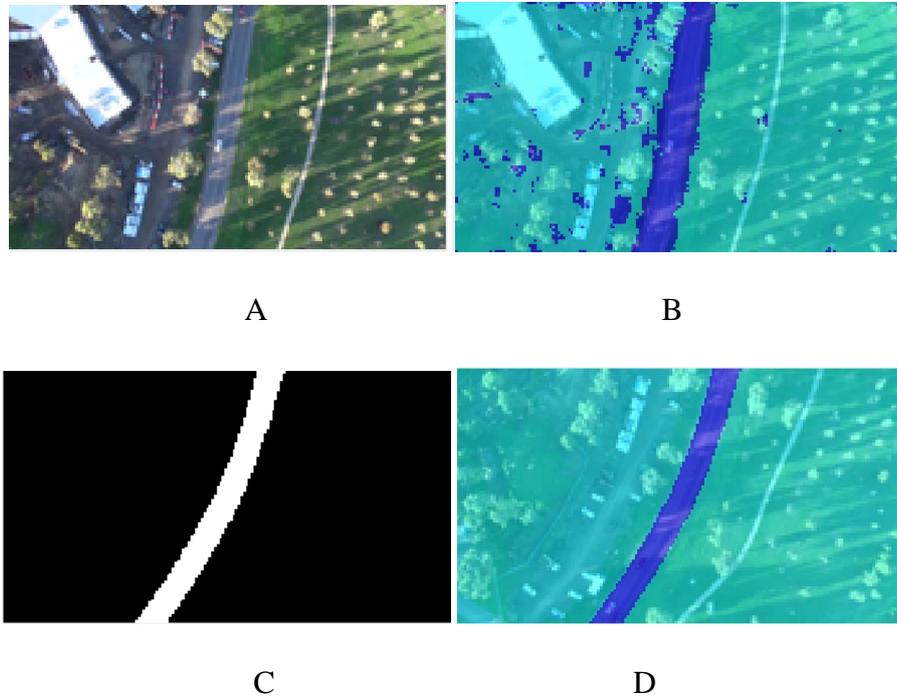


Figure (3.12) : image 30 (A, B, C, D)

A. : image couleur originale

B. : image segmenté avant filtrage

C. : image groundtruth

D. : image originale juxtaposée avec le groundtruth en cyan

- **Remarque :**

⇒ Nous avons constaté que dans ce second réseau, nous avons réussi à réduire le temps d'apprentissage de façon plus efficace avec un pourcentage allant à 91,22 % pour l'apprentissage. Une comparaison entre les deux réseaux, montre un gain énorme de temps pour le deuxième modèle U-Net au détriment de la qualité du résultat de la segmentation.

5.3. Partie 04 : modèle U-NET 2 sur une autre base de données à plusieurs classes

⇒ Base de données : Lors de l'apprentissage du test du réseau de segmentation sémantique, nous avons besoin d'un ensemble d'images et de l'ensemble correspondant d'images étiquetées (groundtruth), où chaque valeur de pixel dans cette dernière, représente l'étiquette de ce pixel dans l'image réelle correspondante. Nous évaluons les deux réseaux sur une base de données d'images de drones composée d'images semi-urbaines et non urbaines collectées sur Internet. Ils se composent de plusieurs routes pavées et non pavées avec différentes résolutions, largeurs et formes. La couleur, l'éclairage et le contraste de l'image de l'itinéraire actuel changent, dans les 10 images UAV utilisées. On peut télécharger cette base de données : <https://sites.google.com/site/hailingzhouwei/>.

⇒ Nous avons testé notre modèle U-NET 2 sur la base de données à plusieurs classes (building, tree, road)

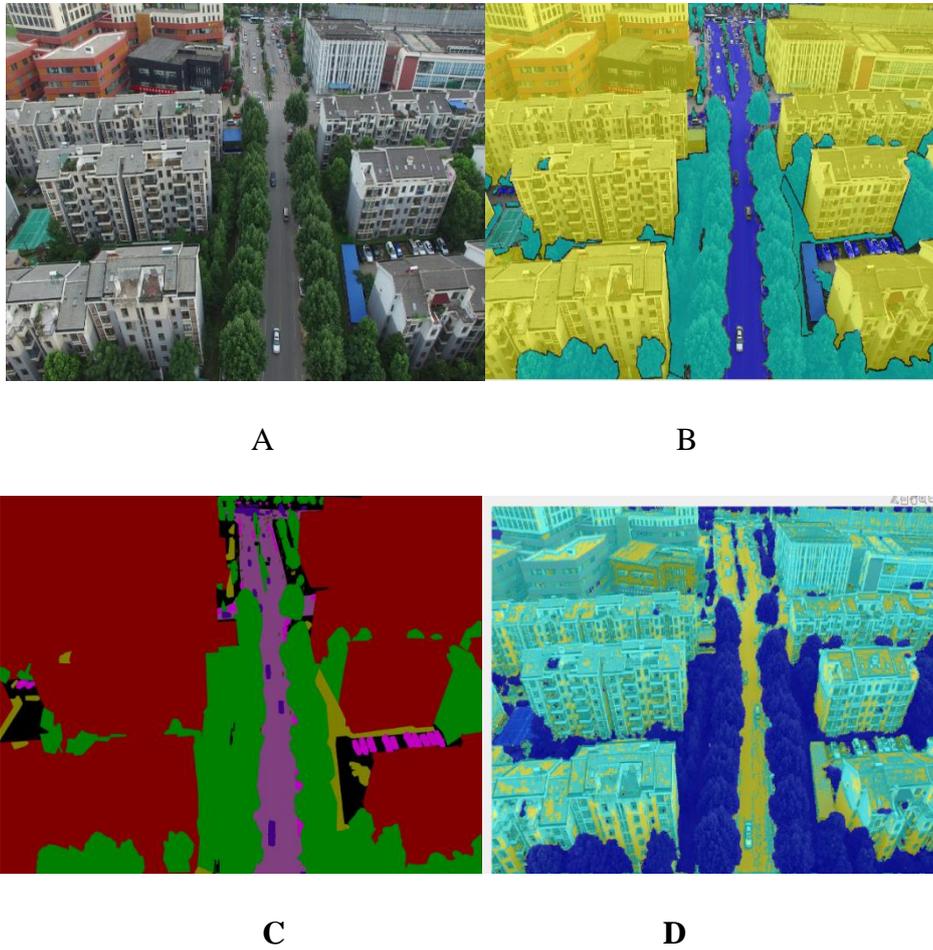


Figure (3.13) : notre modèle U-NET 2 avec trois classes (building, tree, road)

A. : image couleur originale

B. : image originale juxtaposée avec le groundtruth en cyan

C. : image groundtruth

D. : image segmenté

- Ce tableau explique également les résultats d'exploitation du réseau U-NET en fonction du temps d'apprentissage :

Name	Pixel Count	ImagePixelCount
{'végétation'}	6.6568e+05	6.7185e+06
{'road' }	1.2917e+06	6.7185e+06
{'building'}	3.851e+06	6.7185e+06

Tableau (3.6) : Trois classes du réseau U-NET en fonction du temps d'apprentissage

```

=====
| Epoch | Iteration | Time Elapsed | Mini-batch | Mini-batch | Base Learning |
|       |          | (hh :mm:ss) | Accuracy   | Loss       | Rate         |
=====
| 1 | 1 | 00 :00 :26 | 30.14% | 3.5422 | 0.0010 |
| 50 | 50 | 00 :12:14 | 69.32% | 0.6246 | 0.0010 |
| 100 | 100 | 00 :24:07 | 74.54% | 0.5812 | 0.0010 |
| 150 | 150 | 19:11:00 | 74.68% | 0.5630 | 0.0010 |
| 200 | 200 | 19:23:08 | 75.47% | 0.5438 | 0.0010 |
=====

```

Tableau (3.7) : Le rapport d'entrainement



```
Command Window
2 2 2 2
2 2 2 2
2 2 2 2
2 2 2 2
2 2 2 2
3 1 1 1

>> accuracy=mean(mean(C1==gt))

accuracy =

    1
```

Figure (3.14) : Résultat accuracy modèle U-NET 2 avec trois classes (building, tree, road)

⇒ Nous avons présenté un nouvel ensemble de données UAV pour faire progresser le développement de la segmentation sémantique dans les scènes de rue urbaines Images de drones. La reconnaissance d'objets en mouvement dans les scènes de rue, et la cohérence temporelle sur plusieurs trames. Trois classes ont été définies (building road végétation)

⇒ Nous aimerions élargir davantage l'ensemble de données dans la taille et le nombre de catégories pour le rendre plus bénéfiques et utile pour faire avancer la recherche de segmentation sémantique pour le drone Imagerie

6. Conclusion :

Rappelons que La segmentation des images UAV a été faite avec deux modèles UNET l'apprentissage a été réalisé sur des patches choisis aléatoirement sur les images.

La segmentation est réalisée sur des fenêtres de l'image avec des tailles égales à au moins celle du patch. Les résultats sont satisfaisants mais dépendent des paramètres de training, du nombre de patches, de la taille des images et la base d'apprentissage surtout.

Une comparaison entre les deux réseaux, montre un gain énorme de temps le deuxième modèle U-Net au détriment de la qualité du résultat de la segmentation.

Conclusion générale :

La reconnaissance du réseau routier acquiert une importance grandissante grâce à la démocratisation des drones qui sont un outil de détection à faible cout. Notre but dans ce mémoire est dans le même contexte.la segmentation des images a été réalisée par les réseaux profonds plus spécifiquement les modelés Unet. Nous avons exploré dans notre travail les concepts de base de la segmentation d'images, ainsi que le Deep Learning basé sur les réseaux de neurones convolutifs, Nous avons expliqué les différentes couches du modèle U-Net (convolution /déconvolution, pooling,

L'apprentissage a été réalisé sur des patches choisis aléatoirement sur les images.la segmentation est réalisée sur des fenêtres de l'image avec des tailles égales à au moins celle du patch .deux structures du modèles UNET avec 5couches puis avec deux couches ont été utilisés pour l'apprentissage et les résultats montrent une préférence pour la première structure mais le facteur temps d'apprentissage reste le problème prépondérant à cause des machines peu compétentes.

Le Deep Learning impose une machine très performante en termes de vitesse de traitement et d'espace mémoire. Pour pallier à ce problème en perspective des solutions sont proposées comme :

- un PC doté (processeur plus exigeant et carte CPU et GPU)
- Le choix des paramètres influe sur le temps de l'étape d'apprentissage. Un compromis est nécessaire entre le temps de training et le maxepoch, le mini batch le nombre d'images d'apprentissage et aussi la taille des images traitées)

Bibliographie :

- [1] Image Segmentation" Ronneberger O, Fischer P, Brox T (2015). "U-Net : Convolution Networks for Biomédicale Image Segmentation". <https://arxiv.org/abs/1505.04597>
- [2] Shelhamer E, Long J, Darrell T (April 2017). "Fully Convolutional Networks for Semantic Segmentation". *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 39 (4): 640–65 <https://arxiv.org/abs/1411.4038>
- [3] Kleanthous, Christos; Chatzis, Sotirios (2020). "Gated Mixture Variational Autoencoders for Value Added Tax audit case selection". *Knowledge-Based Systems*. 188 : 105048. doi:10.1016/j.knosys.2019.105048.
- [4] "Army researchers develop new algorithms to train robots". *EurekAlert!*. Archived from the original on 28 August 2018. Retrieved 29 August 2018.
- [5] TOM KELDENICH/ 17 OCTOBRE 2021 /<https://www.inside-machinelearning.com/encodeur-decodeur/>
- [6] <http://blog.itpub.net/29829936/viewspace-2217861/>
- [7] <https://arabicprogrammer.com/article/4283136996/>
- [8] <http://blog.itpub.net/29829936/viewspace-2217861/>
- [9]March 2021 <https://www.techtarget.com/searchenterpriseai/definition/deep-learning-deep-neural-network>
- [10]Oktem, Figen S.; Kar, Oğuzhan Fatih; Bezek, Can Deniz; Kamalabadi, Farzad (2021). "High-Resolution Multi-Spectral Imaging With Diffractive Lenses and Learned Reconstruction". *IEEE Transactions on Computational Imaging*. 7: 489–504. arXiv:2008.11625. doi:10.1109/TCI.2021.3075349.
- [11]Bernhardt, Melanie; Vishnevskiy, Valery; Rau, Richard; Goksel, Orcun (December 2020). "Training Variational Networks With Multidomain Simulations: Speed-of-Sound Image Reconstruction". *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*. 67 (12): 2584–2594. arXiv:2006.14395.
- [12] https://en.wikipedia.org/wiki/Deep_learning#Applications

- [13] <https://www.journaldunet.fr/web-tech/guide-de-l-intelligence-artificielle/1501333-deep-learning-definition-et-principes-de-l-apprentissage-profond/>
Antoine Crochet-Damais, Mis à jour le 23/02/22 18:10, Deep Learning : définition et principes de l'apprentissage profond
- [14] <https://actualiteinformatique.fr/intelligence-artificielle/definition-deep-learning>
- [15] <https://www.futura-sciences.com/tech/definitions/informatique-reseau-neuronal-601/>
- [16] <https://fr.scribd.com/document/504948072/77504-BONAZZA-2019-archivage>
- [17] <https://fr.scribd.com/document/504948072/77504-BONAZZA-2019-archivage>
- [18] <https://www.invivoo.com/reduction-dimensionnalite-machine-learning/> 3 juin 2021
- [19] <https://blent.ai/unet-computer-vision/> Nada Belaidi Publié le 8 mars 2022
- [20] [21] [22] <https://www.scribd.com/document/504948072/77504-BONAZZA-2019-archivage> BONAZZA 2019

- Wikipédia