

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

UNIVERSITE MOULOUD MAMMERI DE TIZI-OUZOU



FACULTE DU GENIE ELECTRIQUE ET DE L'INFORMATIQUE
DEPARTEMENT D'AUTOMATIQUE

Mémoire de Fin d'Etudes de MASTER ACADEMIQUE

Domaine : Sciences et Technologies

Filière : Automatique

Spécialité : Automatique et Système

Présenté par

Youcef AMROUN

Amar BOUABBA

Thème

Application de l'Apprentissage Profond pour la segmentation sémantique en Vidéosurveillance

Mémoire soutenu publiquement le 03/10/ 2024 devant le jury composé de :

M Arezki BENFADILA

Professeur, UMMTO, Président

Mme Sadia ALKAMA

MCA, UMMTO, Encadrant

Mme Ouiza NAIT BELAID

MAA, UMMTO, Examineur

Mme Zohra HALICHE

MAA, UMMTO, Examineur

Remerciements

Tout d'abord, merci à Dieu le Tout Puissant de nous avoir donné la force et la volonté d'acheminer ce travail.

Tout travail de recherche n'est jamais totalement l'œuvre d'une seule personne.

À cet effet, nous tenons à exprimer nos sincères reconnaissances et nos vifs remerciements à tous ceux qui ont contribué de près ou de loin à l'élaboration de ce travail.

Toutes nos gratitudes à notre promotrice Madame ALKAMA Sadia, pour sa gentillesse et son suivi durant la réalisation de ce mémoire.

Enfin, nous adressons nos plus sincères remerciements à nos parents, et à tous nos proches et amis qui nous ont toujours encouragées au cours de la réalisation de ce mémoire.

On vous remercie infiniment

Dédicaces

Je dédie ce travail premièrement à mes chers parents pour leurs encouragements, leur soutien. Vous êtes ma source d'inspiration et ma force.

Je le dédie également à mon frère et ma sœurs, ainsi qu'à tous ceux qui, de près ou de loin, ont cru en moi et m'ont soutenu durant ce parcours.

Enfin, je dédie ce mémoire à tous mes amis qui ont partagé cette aventure avec moi, pour leur précieuse amitié et leur aide.

Avec toute ma gratitude.

Youcef.

Dédicaces

Je dédie ce travail en premier lieu à toute ma famille pour leurs encouragements, leur soutien et leurs contribution durant mon cursus d'étude. Et je tien aussi à souligner mes amies, qui sont pour moi une source d'encouragement à tout moment.

Je le dédie également à tous ceux qui ont cru en moi, de près ou de loin, et m'ont soutenu durant ce parcours.

Enfin, je dédie ce mémoire à tous ceux qui ont partagé cette aventure avec moi, pour leur aide et encouragement.

Avec toute ma gratitude.

Amar.

Table des matières

Introduction générale	1
I Vidéosurveillance et Traitement d'images	2
I.1 Introduction	2
I.2 Vidéosurveillance	2
I.2.1 Définition	2
I.2.2 Historique	3
I.2.3 Composants d'un système de vidéosurveillance	4
I.3 Traitement d'images pour vidéosurveillance automatique	5
I.4 Segmentation d'images	6
I.4.1 Définition	6
I.4.2 Différentes approches de la Segmentation d'images	7
I.4.3 Application de la Segmentation d'images à la Détection d'objets	13
I.4.4 Rôle de la Segmentation d'images dans la Vidéosurveillance	14
I.5 Conclusion	15
II Intelligence artificielle et Deep Learning	16
II.1 Introduction	16
II.2 Intelligence artificielle	16
II.2.1 Définition	16
II.2.2 Historique de l'IA	17
II.2.3 Domaine d'utilisation de l'IA	18
II.2.4 Différentes branches de l'IA	19
II.2.4.1 IA symbolique	19
II.2.4.2 IA basée sur les connaissance	19
II.2.4.3 Apprentissage automatique (Machine learning)	19
II.3 Machine learning	20
II.3.1 Définition	20
II.3.2 Types d'apprentissage automatique	20

II.3.2.1	Apprentissage supervisé	21
II.3.2.2	Apprentissage non supervisé	21
II.3.2.3	Apprentissage par renforcement	22
II.3.3	Diagramme des relations entre les concepts	22
II.4	Réseaux de neurones	23
II.4.1	Historique des réseaux de neurones	23
II.4.2	Neurones biologiques	24
II.4.3	Neurone formel	24
II.4.4	Définition et exemple de réseaux de neurones	26
II.4.4.1	Définition	26
II.4.4.2	Perceptron multicouches	26
II.4.5	Processus d'apprentissage MLP	27
II.5	Deep Learning	29
II.5.1	Définition	29
II.5.2	Types de modèles de Deep Learning	29
II.5.3	Les réseaux de neurones convolutifs (CNN)	30
II.5.4	Applications du Deep Learning	31
II.5.5	Segmentation sémantique des images	32
II.6	Conclusion	33
III	Mise en Œuvre Tests et Résultats	34
III.1	Introduction	34
III.2	Schéma de principe	34
III.3	Segmentation sémantique	36
III.3.1	Chargement des données	36
III.3.2	Prétraitement des données	37
III.3.3	Augmentation des données	38
III.3.4	Entraînement du modèle	38
III.4	Choix du réseau	41
III.5	Résultats	44
III.6	Difficultés Rencontrées lors de la Mise en Œuvre du Programme	48
III.7	Conclusion	48
	Conclusion générale	50
	Bibliographie	51

Table des figures

I.1	Vidéosurveillance	3
I.2	Processus de traitement d'images pour la vidéosurveillance . .	5
I.3	Exemple de segmentation	6
I.4	Méthodes de Seuillage	8
I.5	Segmentation basée sur les régions	9
I.6	Segmentation basée sur les contours	11
II.1	Schéma de l'apprentissage supervisé	21
II.2	Schéma de l'apprentissage non supervisé	22
II.3	Diagramme des relations entre les concepts	23
II.4	Le neurone biologique	24
II.5	Fonction réalisée par le neurone formel	25
II.6	Architecture d'un perceptron multicouches	27
II.7	Différentes couches de CNN	30
II.8	Structure d'un CNN	32
III.1	Schéma de principe du système de vidéo surveillance autonome	35
III.2	Images et Images des étiquettes de la base Cam-Vid	37
III.3	Architecture de U-Net	39
III.4	Architecture de Segnet	39
III.5	La structure du réseau combinant Deeplabv3+ et ResNet 50 .	40
III.6	Segmentation sémantique en utilisant plusieurs réseaux	41
III.7	Métrique d'évolution durant l'entraînement	42
III.8	Résultats des tests sur la vidéo 1	45
III.9	Résultats des tests sur la vidéo 2	46
III.10	Exemple d'une occlusion	47

Abréviations

Abréviations	Définitions
IA	Intelligence Artificielle
ML	Machine Learning (Apprentissage Automatique)
DL	Deep Learning (Apprentissage Profond)
IoT	Internet of Things (Internet des Objets)
GPU	Graphics Processing Unit (Unité de Traitement Graphique)
CPU	Central Processing Unit (Unité Centrale de Traitement)
MLP	Multi-Layer Perceptron (Perceptron Multicouche)
IP	Internet Protocol (Protocole Internet)
ID	Identifiant

Introduction générale

La vidéosurveillance est devenue un outil incontournable dans la gestion moderne de la sécurité, jouant un rôle crucial dans la protection des espaces publics et privés. Avec l'essor des technologies numériques et de l'intelligence artificielle, l'analyse automatique des flux vidéo a pris une importance primordiale, permettant de détecter des comportements suspects, de surveiller des zones sensibles et d'assurer la sécurité des individus et des biens. La segmentation d'images, en tant que technique clé du traitement d'images, permet d'isoler efficacement les objets d'intérêt dans des environnements complexes, facilitant ainsi leur détection et leur suivi.

Ce mémoire explore les principes fondamentaux de la vidéosurveillance et du traitement d'images en mettant particulièrement l'accent sur la segmentation sémantique.

Nous abordons les différentes approches et technologies utilisées pour améliorer la précision et l'efficacité des systèmes de surveillance.

À travers une étude approfondie des modèles de Deep Learning, nous analyserons les résultats obtenus sur des images capturées en temps réel dans un flux vidéo, tout en identifiant les défis associés à leur mise en œuvre. Ce travail contribue à la compréhension des enjeux contemporains liés à la vidéosurveillance intelligente.

Nous avons organisé ce mémoire en trois chapitres.

Le premier chapitre est consacré à la vidéosurveillance, et explore les concepts de la segmentation d'images pour la détection d'objets.

Le second chapitre porte sur le domaine de l'intelligence artificielle et Deep Learning.

Le dernier chapitre est consacré à la mise en œuvre, tests et résultats.

Nous terminons par une conclusion générale.

Chapitre I

Vidéosurveillance et Traitement d'images

I.1 Introduction

La vidéosurveillance est un outil essentiel pour la gestion de la sécurité dans de nombreux contextes, permettant une observation en temps réel des espaces publics et privés. L'analyse automatique des flux vidéo est devenue indispensable pour détecter des comportements suspects, surveiller des zones sensibles et garantir la sécurité des individus et des biens.

La capacité de détecter et de suivre des objets en temps réel repose sur des techniques avancées de traitement d'images, notamment la segmentation d'images, qui permet d'isoler des objets d'intérêt dans des environnements complexes. Cette interaction entre la détection d'objets et la vidéo surveillance ouvre la voie à des applications variées, allant de la surveillance urbaine à la sécurité des entreprises, en passant par la gestion des foules lors d'événements publics.

Ce chapitre vise à explorer les concepts fondamentaux de la segmentation d'images, à examiner son rôle dans la détection d'objets et à analyser les défis et enjeux associés à son application dans le domaine de la vidéo surveillance.

I.2 Vidéosurveillance

I.2.1 Définition

La surveillance est l'action de surveiller, de contrôler quelque chose ou quelqu'un. Cela implique une observation attentive, un suivi rigoureux, et parfois une intervention pour garantir la sécurité, le bon fonctionnement, ou la discipline. La surveillance peut être exercée dans divers contextes tels que

les établissements scolaires, la sécurité publique, la santé, ou même la protection des lieux et des personnes. Elle peut également inclure des activités de contrôle, d'inspection, et de vigilance pour prévenir des actions délictueuses ou criminelles [1].

La vidéosurveillance est une technologie qui utilise des caméras pour surveiller des espaces spécifique.

Ce système permet de suivre des événements dans les lieux publics ou privés et d'intervenir en cas de comportements suspects, d'infraction ou d'accidents.(voir figure I.1)

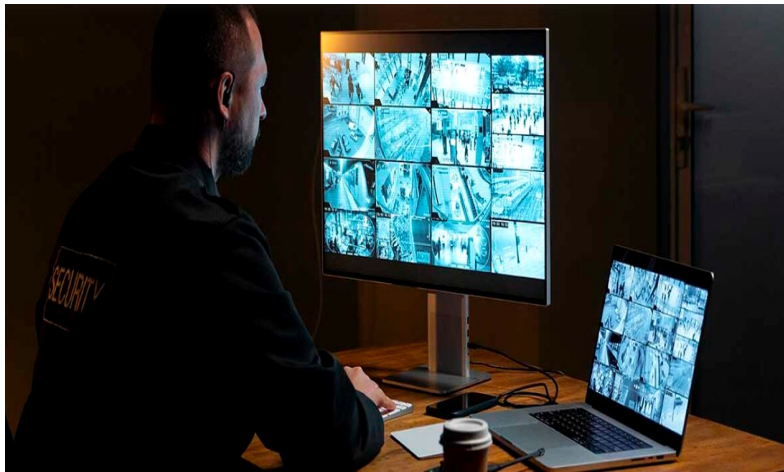


FIGURE I.1 – Vidéosurveillance .

I.2.2 Historique

Initialement, les système de vidéosurveillance se limitaient à des caméra analogiques reliées à des moniteurs, offrant une vision limitée et de qualité médiocre. Avec l'évolution technologique, la vidéosurveillance a intégré des cameras numériques, la compression vidéo et surtout des systèmes intelligents capables de détecter et d'analyser des situations anormales. L'avènement des caméras IP permet la transmission des données vidéo sur des réseaux informatiques, offrant ainsi la possibilité de Télé-surveiller.

La miniaturisation des caméras a également ouvert de nouvelles perspectives. Ces technologies compactes et discrètes ont permis l'intégration de la surveillance dans des dispositifs de plus en plus petits, offrant ainsi des solutions de surveillance invisibles pour des applications variées, telles que la sécurité personnelle et la surveillance clandestine.

L'intelligence artificielle (IA) et l'apprentissage automatique ont révolutionné la manière dont les systèmes de surveillance analysent les données. Les algo-

algorithmes avancés peuvent désormais détecter automatiquement les comportements suspects, identifier les objets et les personnes, et même anticiper les incidents potentiels, ce qui améliore considérablement la précision et l'efficacité de la surveillance.

Parmi les avancées significatives de la vidéosurveillance on trouve :

- La résolution et la qualité des images
- La vision nocturne avancée
- L'utilisation d'algorithmes d'intelligence artificielles
- L'intégration avec d'autres systèmes
- La surveillance à distance
- La protection de la vie privée

I.2.3 Composants d'un système de vidéosurveillance

Un système de vidéosurveillance moderne est composé de plusieurs éléments qui permettent de capturer, stocker, transmettre et analyser le flux vidéo pour fournir une surveillance en temps réel ou différée. Les éléments des systèmes de vidéosurveillance sont :

- Les caméras qui sont utilisées pour capturer les images les vidéos. Elles peuvent être de différents types. Elles sont fixes pour surveiller des angles fixes et capable de pivoter de s'incliner et de zoomer pour une couverture large et dynamique d'une zone. Pour la surveillance des endroits à faible luminosité des caméras infrarouges sont utilisées. Afin de transmettre des vidéos à travers un réseau internet pour une surveillance à distance, des caméras IP sont utilisées.
- Les systèmes de stockage qui sont essentiels pour enregistrer les flux vidéo capturés par les caméras. Le stockage peut être réalisé sur des disques durs ou dans des centres de stockage des données ou sur le cloud.
- Les réseaux de communication qui sont nécessaires pour transmettre les vidéos des caméras vers les systèmes de stockage ou aux postes de contrôle. Ils peuvent être câblés ou sans fil (wi-fi) ou internet.
- Les moniteurs ou interfaces utilisateur qui permettent aux opérateurs de visualiser les images en temps réel ou de consulter les enregistrements.
- Les logiciels de gestion des vidéos qui sont intégrés dans la surveillance moderne. Ils permettent d'analyser les événements et d'alerter les opérateurs en cas de situation anormale.

I.3 Traitement d'images pour vidéosurveillance automatique

Le traitement d'images est un composant essentiel dans la mise en œuvre d'un système de vidéo de surveillance automatique. Il est utilisé pour analyser en temps réel le flux vidéo et extraire des informations pour identifier des événements et des comportements anormaux.

Le processus de traitement d'images pour la vidéosurveillance est assez semblable au processus de traitement d'images, il se décompose en plusieurs étapes comme le montre la figure I.2.

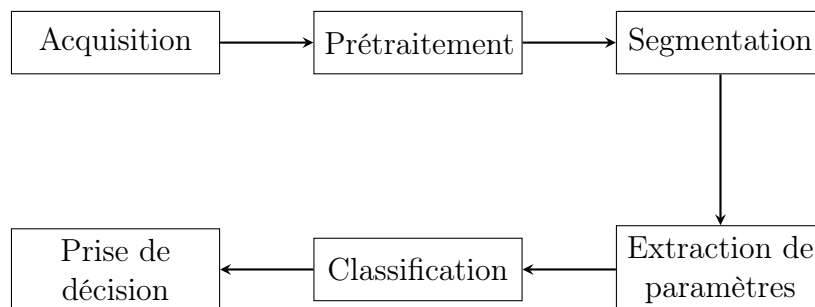


FIGURE I.2 – Processus de traitement d'images pour la vidéosurveillance

Ce schéma montre le processus de traitement d'image de la télésurveillance automatique. La première étape de processus consiste à l'acquisition, qu'est pour recevoir les données (images) à partir du système de vidéosurveillance. En suite le pré-traitement, c'est la deuxième étape de processus, qui consiste à améliorer la qualité des données collectées afin de les rendre utilisables et également réduire le bruit. La troisième étape est la segmentation, cette étape consiste à diviser les données (image) en différentes catégories. En suite l'étape de l'extraction des paramètres, qui consiste à extraire les caractéristiques qui vont permettre de décrire chaque objet de l'image. En suite la classification, cette étape utilise les caractéristiques extraites de l'étape précédente pour classer chaque objet. En fin, l'étape de prise de décision, cette étape a pour déclencher une action appropriée.

Les différentes étapes de ce processus ont leur importance mais l'étape de la segmentation est cruciale car elle conditionne la qualité des étapes suivantes. Nous la détaillerons dans le paragraphe suivant.

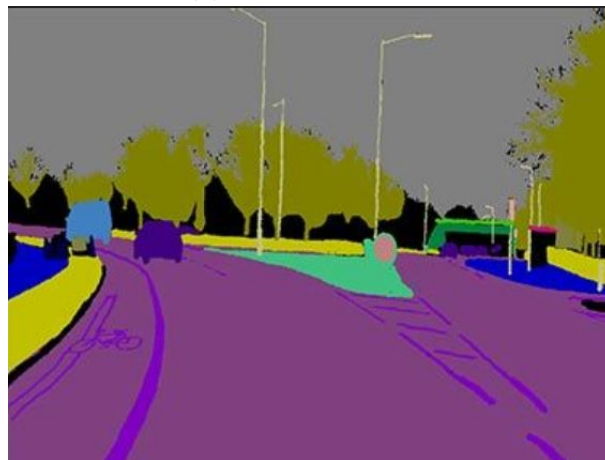
I.4 Segmentation d'images

I.4.1 Définition

La segmentation d'images est une technique fondamentale qui permet l'extraction des éléments constitutifs d'une image, jouant un rôle crucial dans le processus d'interprétation d'une scène. Elle consiste à partitionner l'image en un certain nombre de régions disjointes, facilitant ainsi l'analyse et la compréhension des objets présents comme le montre la figure suivante.



(a) Image originale.



(b) Image segmentée.

FIGURE I.3 – Exemple de segmentation

La segmentation est une décomposition de l'image I en n régions R_i telle que :

$$\begin{aligned} \cup_i R_i \\ R_i \cap R_j &= \emptyset \\ P(R_i) &= \text{vrai} \\ P(R_i \cap R_j) &= \text{faux} \end{aligned} \tag{I.1}$$

En conclusion la segmentation consiste à :

- Regrouper les pixels de l'image qui partagent une même propriété pour former des régions homogènes.
- Répartir l'ensemble de pixels de l'image en différents groupes.
- découper l'image en région. Une région est caractérisée par contours et par homogénéité (par exemple, même couleur).
- Partitionner une image en un ensemble de régions connexes et disjointes.
- La recherche de zones de l'image possédant des attributs communs, comme la luminosité, la couleur ou plus rarement la texture.

I.4.2 Différentes approches de la Segmentation d'images

Les techniques les plus répandues de segmentation d'images traditionnelles comprennent :

- **Méthodes de Seuillage :** Le seuillage consiste à choisir un seuil de luminosité qui détermine la classification des pixels. Les pixels dont l'intensité est inférieure à ce seuil sont généralement considérés comme appartenant à l'arrière-plan (souvent représentés par la couleur noire), tandis que ceux dont l'intensité est supérieure sont classés comme appartenant à l'objet d'intérêt (souvent représentés par la couleur blanche). Ce processus transforme l'image en une image binaire, facilitant ainsi l'analyse ultérieure.

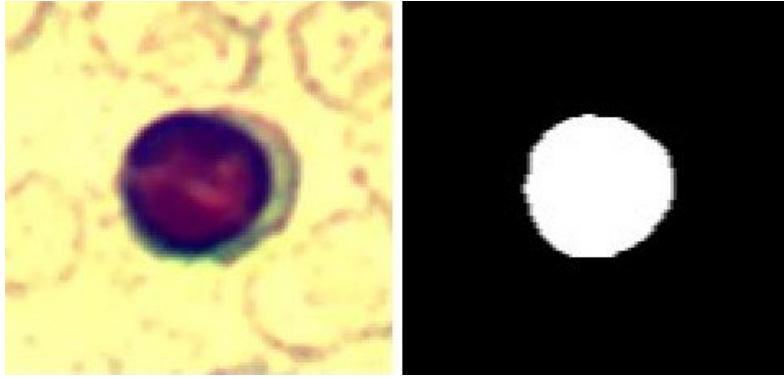


FIGURE I.4 – Méthodes de Seuillage

La méthode de seuillage consiste à appliquer une fonction de seuil à chaque pixel d'une image pour décider s'il doit être classé comme appartenant à l'objet d'intérêt ou à l'arrière-plan. Mathématiquement, cela peut être exprimé comme suit :

Soit $I(x, y)$ l'intensité du pixel à la position (x, y) dans l'image. Soit T le seuil choisi.

Fonction de Seuillage

La fonction de seuillage S peut être définie comme :

$$S(x, y) = \begin{cases} 255, & \text{si } I(x, y) \geq T \\ 0, & \text{sinon} \end{cases} \quad (\text{I.2})$$

Cela signifie que si l'intensité d'un pixel est supérieure ou égale au seuil, il devient blanc (255), sinon il devient noir (0).

Algorithmes Connus de Seuillage

1. Seuillage Binaire Simple

- **Description** : Le seuil est fixé manuellement. L'utilisateur choisit un seuil T et l'applique à tous les pixels de l'image.
- **Complexité** : $O(n)$, où n est le nombre de pixels dans l'image.

2. Algorithme d'Otsu

- **Description** : Cet algorithme détermine automatiquement le seuil optimal T en minimisant la variance intra-classe des niveaux de gris. Il analyse l'histogramme de l'image et choisit le seuil qui maximise la séparation entre les classes (objet et fond).

— **Étapes de l'algorithme :**

- (a) Calculer l'histogramme de l'image.
- (b) Normaliser l'histogramme pour obtenir la probabilité de chaque niveau de gris.
- (c) Calculer la moyenne globale de l'image.
- (d) Pour chaque seuil possible, calculer la variance intra-classe et sélectionner le seuil qui minimise cette variance.

— **Complexité :** $O(n + k)$, où n est le nombre de pixels et k est le nombre de niveaux de gris.

3. Seuillage Adaptatif

— **Description :** Contrairement au seuillage global, qui utilise un seul seuil pour toute l'image, le seuillage adaptatif calcule des seuils pour des sous-régions de l'image. Cela est particulièrement utile dans les cas où l'éclairage n'est pas uniforme.

— **Étapes de l'algorithme :**

- (a) Diviser l'image en blocs.
- (b) Pour chaque bloc, calculer un seuil (par exemple, en utilisant la moyenne ou la médiane des pixels dans le bloc).
- (c) Appliquer le seuillage local à chaque bloc en fonction du seuil calculé.

— **Complexité :** $O(n)$, où n est le nombre de pixels, mais peut varier selon la taille des blocs.

— **Segmentation basée sur les régions :** L'approche de segmentation basée sur les régions consiste à diviser l'image en régions distinctes. Les pixels adjacents sont divisés en régions distinctes sur la base d'un critère d'uniformité ou de cohérence, compte tenu de la proximité. Ce critère peut être par exemple le niveau de gris, la couleur, la texture, etc. Une procédure de regroupement est répétée jusqu'à ce que tous les pixels de l'image soient intégrés aux régions. Cette méthode tente donc d'isoler l'image caractéristique selon les territoires implicites.

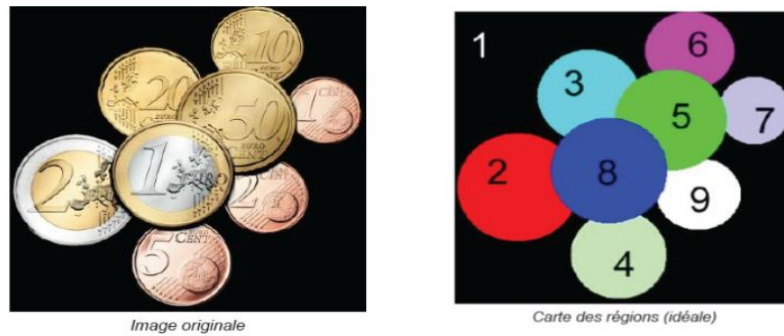


FIGURE I.5 – Segmentation basée sur les régions

Algorithmes Connus de Segmentation Basée sur les Régions

1. Croissance de Régions

- **Description** : Cet algorithme commence par une ou plusieurs graines (pixels initiaux) et agrandit ces régions en ajoutant des pixels voisins qui répondent à des critères de similarité.
- **Étapes de l'algorithme** :
 - (a) Choisir une ou plusieurs graines initiales.
 - (b) Évaluer les pixels voisins de chaque pixel de la région.
 - (c) Ajouter un pixel à la région si sa similarité avec la région existante est suffisante.
 - (d) Répéter jusqu'à ce qu'aucun pixel supplémentaire ne puisse être ajouté.
- **Complexité** : $O(n)$, où n est le nombre de pixels dans l'image.

2. Division et Fusion

- **Description** : Cette méthode divise l'image en régions et fusionne ensuite les régions adjacentes qui répondent à des critères de similarité.
- **Étapes de l'algorithme** :
 - (a) Diviser l'image en blocs de taille fixe.
 - (b) Évaluer chaque bloc selon des critères de similarité.
 - (c) Fusionner les blocs adjacents qui sont similaires.
 - (d) Répéter jusqu'à ce que toutes les régions soient stables.
- **Complexité** : $O(n \log n)$, en fonction de la méthode de fusion utilisée.

3. Segmentation par Graphes

- **Description** : La segmentation par graphes modélise l'image comme un graphe, où les pixels sont des nœuds, et les similarités entre pixels sont représentées par des arêtes. Les algorithmes de coupe de graphes sont utilisés pour segmenter l'image.
- **Étapes de l'algorithme** :
 - (a) Construire un graphe à partir de l'image.
 - (b) Déterminer des régions en coupant les arêtes du graphe qui représentent des différences importantes entre les pixels.
 - (c) Obtenir les segments en analysant les composantes connexes du graphe.
- **Complexité** : $O(n \log n)$, dépend de l'algorithme de coupe de graphes utilisé.
- **Segmentation basée sur les contours** : La segmentation basée sur les contours est une technique de traitement d'images qui se concentre sur la détection et l'extraction des contours des objets présents dans une image. Cette méthode est particulièrement efficace pour isoler des formes complexes et des objets dont les bords sont bien définis.

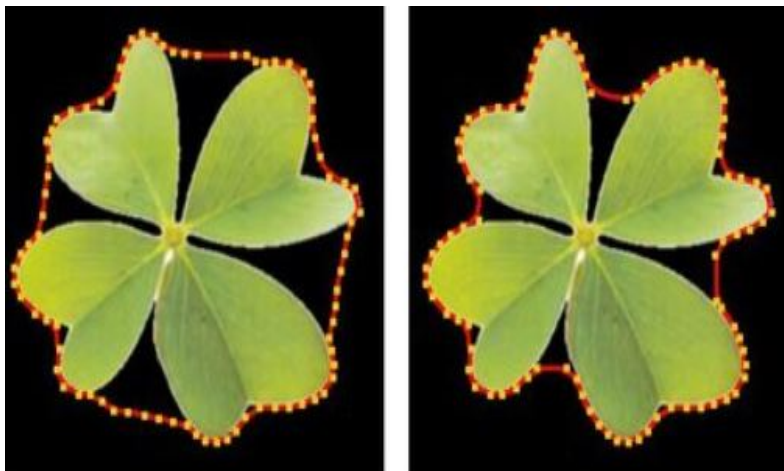


FIGURE I.6 – Segmentation basée sur les contours

Algorithmes Connus de Segmentation Basée sur les Contours

1. Détection de Contours avec le Filtre de Sobel

- **Description** : Le filtre de Sobel est utilisé pour détecter les contours en calculant le gradient de l'intensité de l'image. Les gradients sont calculés dans les directions horizontale et verticale.
- **Équations** : Les dérivées approximatives sont données par les noyaux suivants :

$$G_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}, \quad G_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \quad (\text{I.3})$$

où G_x et G_y sont les noyaux pour les dérivées en x et y , respectivement. Le gradient total G est ensuite calculé comme suit :

$$G = \sqrt{G_x^2 + G_y^2} \quad (\text{I.4})$$

- **Complexité** : $O(n)$, où n est le nombre de pixels dans l'image.

2. Filtre de Canny

- **Description** : Le filtre de Canny est une méthode de détection de contours qui utilise plusieurs étapes, y compris la suppression du bruit, le calcul des gradients, et l'application de la non-maximisation et du seuillage.
- **Étapes de l'algorithme** :
 - (a) Appliquer un filtre gaussien pour réduire le bruit.
 - (b) Calculer le gradient de l'image en utilisant des noyaux de Sobel.
 - (c) Appliquer la non-maximisation pour supprimer les pixels qui ne sont pas des maxima locaux.
 - (d) Utiliser un seuil double pour identifier les contours.
- **Complexité** : $O(n \log n)$, en raison de l'application de la non-maximisation et de la recherche des maxima locaux.

3. Segmentation par Évolution de Contours

- **Description** : Cette méthode utilise des modèles de contours actifs qui évoluent à travers l'image pour s'ajuster aux bords des objets. Un exemple courant est le modèle de Snakes.
- **Équation** : L'énergie totale E du contour actif est donnée par :

$$E = \int (E_{\text{internal}} + E_{\text{external}}) ds \quad (\text{I.5})$$

où E_{internal} contrôle la régularité du contour et E_{external} attire le contour vers les bords détectés.

- **Complexité** : $O(n)$, dépendant du nombre d'itérations pour atteindre la convergence.

I.4.3 Application de la Segmentation d'images à la Détection d'objets

La segmentation d'images joue un rôle crucial dans la détection d'objets, qui est une tâche fondamentale en vision par ordinateur et en intelligence artificielle. Voici comment la segmentation d'images s'applique à la détection d'objets :

- **Segmentation Sémantique pour la Détection d'objets** : La segmentation sémantique est une technique qui permet de classifier chaque pixel d'une image dans des catégories prédéfinies, comme "voiture", "personne", "bâtiment", etc. En appliquant un algorithme de segmentation sémantique, on obtient une carte de l'image où chaque région correspond à un objet détecté.

Cette segmentation sémantique fournit une base solide pour la détection d'objets, car elle identifie précisément les zones de l'image qui contiennent des objets d'intérêt. Les algorithmes de détection peuvent ensuite se concentrer sur ces régions pour localiser et classifier les objets présents[2].

- **Segmentation par Instance pour la Détection d'objets** : La segmentation par instance va encore plus loin en distinguant les différentes instances d'un même type d'objet dans l'image. Par exemple, si plusieurs voitures sont présentes, la segmentation par instance délimitera chaque voiture individuellement.

Cette approche est particulièrement utile pour la détection d'objets, car elle permet non seulement d'identifier les objets, mais aussi de les compter et de les localiser précisément dans l'image. Les boîtes englobantes autour de chaque instance d'objet fournissent les coordonnées nécessaires pour la détection.

- **Segmentation panoptique** : Ce type de segmentation consiste à la classification sémantique de tous les pixels et différencie chaque instance d'objet au sein d'une même image, combinant les avantages de la segmentation sémantique et d'instance.

Dans une tâche de segmentation panoptique, chaque pixel doit se voir attribuer à la fois une étiquette sémantique et un « ID d'instance ». Les

pixels partageant le même libellé et le même ID appartiennent au même objet ; pour les pixels déterminés comme ensembles, l'ID d'instance est ignoré. La segmentation panoptique fournit ainsi aux systèmes de vision par ordinateur une compréhension globale et exhaustive d'une image donnée. Bien qu'il soit tentant de l'utiliser, réaliser une segmentation panoptique de manière cohérente et efficace sur le plan informatique constitue un défi de taille.

Cette approche présente deux inconvénients majeurs : elle nécessite beaucoup de temps de calcul et a du mal à résoudre les écarts entre les points de données produits par le réseau de segmentation sémantique et les points de données générés par le réseau de segmentation d'instance.

En combinant la puissance de la segmentation d'images et les techniques avancées de détection d'objets, il est possible de créer des systèmes de vision par ordinateur très performants, capables de comprendre et d'analyser des scènes visuelles de manière robuste et précise.

I.4.4 Rôle de la Segmentation d'images dans la Vidéosurveillance

La segmentation d'images joue un rôle crucial dans le domaine de la Vidéosurveillance, en permettant une analyse approfondie et efficace des scènes capturées. Cette technique consiste à diviser une image ou une séquence vidéo en segments distincts, facilitant ainsi l'identification et le suivi des objets d'intérêt.

Parmi les rôles de la Segmentation d'images dans la Vidéosurveillance on trouve :

- **Détection et suivi d'objets :** La segmentation d'images permet de détecter et de suivre des objets tels que des personnes, des véhicules ou des bagages laissés sans surveillance. En isolant ces éléments, les systèmes de surveillance peuvent mieux analyser les comportements, identifier les menaces potentielles et réduire les fausses alertes. Par exemple, la capacité à différencier les comportements normaux des comportements suspects est essentielle pour assurer la sécurité dans des environnements sensibles comme les aéroports ou les centres commerciaux.
- **Analyse comportementale :** En segmentant les images, les systèmes de surveillance peuvent également analyser les interactions entre les personnes. Cela permet de comprendre les schémas de comportement, d'évaluer les réactions émotionnelles et d'améliorer les réponses des agents de sécurité face à des situations potentiellement dangereuses. L'analyse

comportementale est rendue plus précise grâce à la segmentation, qui fournit des données détaillées sur les mouvements et les actions des individus dans une scène.

- **Amélioration de la précision :** L'utilisation de techniques avancées de segmentation, souvent basées sur l'intelligence artificielle, améliore la précision des systèmes de détection et de classification. Ces algorithmes peuvent traiter des images en temps réel, permettant ainsi une surveillance dynamique et réactive. Par exemple, les systèmes peuvent identifier des objets laissés sans surveillance ou des comportements anormaux avec une grande fiabilité, ce qui est essentiel pour la sécurité publique.

I.5 Conclusion

Dans ce chapitre on a exploré l'importance de la détection d'objets par segmentation d'images dans le domaine de la vidéosurveillance. En mettant en lumière les techniques avancées de traitement d'images. La segmentation d'images, en particulier, se révèle être un outil indispensable pour isoler et analyser les objets d'intérêt dans des environnements complexes, permettant ainsi une surveillance plus précise et efficace.

En somme, ce chapitre a posé les bases nécessaires pour comprendre les mécanismes fondamentaux de la détection d'objets et son application dans des systèmes de surveillance sophistiqués, ouvrant la voie à des discussions plus approfondies sur l'intelligence artificielle et le Deep Learning dans le chapitre suivant.

Chapitre II

Intelligence artificielle et Deep Learning

II.1 Introduction

Ces derniers temps, on entend toujours le terme "Intelligence artificielle" partout, que ce soit à la maison, au travail ou dans les médias. L'intelligence Artificielle (IA) est présente dans de nombreux domaines et sur beaucoup de machines.

Elle intervient dans plusieurs secteurs tels que la santé, l'industrie, les transports, le militaire, l'aérospatiale, les finances, l'agriculture, la sécurité, etc. En réalisant diverses tâches tel que la reconnaissance vocale, la vision par ordinateur, la conduite autonome, l'optimisation des itinéraires, la prise de décision, etc.

L'IA est omniprésente et continue à transformer le monde. Elle se divise en plusieurs branches et le Deep Learning (ou apprentissage profond) est l'une des branches les plus marquantes. Le Deep Learning est une sous-catégorie du Machine Learning (apprentissage automatique). Il repose sur les réseaux de neurones artificiels. Il est souvent utilisé dans le traitement des données massives.

Dans ce chapitre, nous définirons l'IA et ces principales branches avant de nous concentrer sur le Deep Learning.

II.2 Intelligence artificielle

II.2.1 Définition

Selon le Larousse, l'intelligence artificielle est définie comme un ensemble de théories et de techniques mises en œuvre en vue de réaliser des machines capables de simuler l'intelligence humaine.

Plusieurs auteurs ont tentés de définir l'intelligence humaine et beaucoup d'entre eux [3, 4, 5] s'accordent sur le fait qu'elle inclut deux capacités, la

capacité de raisonnement (c'est à dire penser, comprendre, et analyser) et la capacité de communication (qui signifie interagir avec l'environnement extérieur).

Transposée au domaine artificiel, ces deux capacités correspondent à la possibilité de réaliser des calculs complexes, d'apprendre et d'interagir avec l'environnement notamment à travers les données fournies au système.

Autrement, l'intelligence artificielle (IA) est un processus d'imitation de l'intelligence humaine qui repose sur la création et l'application d'algorithmes exécutés dans un environnement informatique dynamique. Son but est de permettre à la machine d'agir comme un être humain. Pour y parvenir, trois composants sont nécessaires :

- des systèmes informatiques.
- des données avec des systèmes de gestion.
- des algorithmes de l'IA (codes).

Pour se rapprocher le plus possible du comportement humain, l'intelligence artificielle a besoin d'une quantité de données et d'une capacité de traitement élevées.

II.2.2 Historique de l'IA

L'intelligence artificielle a fait son apparition en 1943 dans une publication de l'article ("**A Logical Calculus of Ideas Immanent in Nervous Activity**") par Warren McCulloch et Walter Pitts. Dans ce document, les chercheurs dévoilent pour la première fois le modèle mathématique d'un réseau de neurones.

Sept ans plus tard des jeunes étudiants de Havard, nommés Marvin Minsky et Dean Edmonds, ont pu créer le premier ordinateur à réseau de neurones, et la même année Alan Turing publie le Turing Test, qui sert encore à évaluer les IA, dans le but de répliquer ou simuler l'intelligence humaine dans les machines. En 1956 le terme de l'IA se fait dévoiler au public durant la conférence ("**Dartmouth Summer Research Project on Artificial Intelligence**"). Beaucoup considère cet événement comme la naissance de l'intelligence artificielle [6].

Plusieurs années s'écoulent et les travaux sur l'intelligence artificielle continuent. En 1959, Arthur Samuel invente le terme de Machine Learning en travaillant chez IBM. Rapidement, les limites de l'IA apparaissent face à des problèmes complexes.

Vers les années 1980, elle connaît une nouvelle vague d'enthousiasme. En 1989, le français Yann Lecun met au point le premier réseau de neurones

capable de reconnaître des chiffres écrits à la main, cette invention sera à l'origine du développement du Deep Learning.

Cette vague s'est aussi essoufflée. Ce n'est que vers la fin des années 1990 et début des années 2000 qu'elle connaît une véritable renaissance grâce à l'augmentation de la puissance de calcul, et la disponibilité de grandes quantités de données et aux avancées en machine Learning.

En 1997, un événement majeur marque l'histoire de l'IA, le système Deep Blue d'IBM triomphe du champion du monde d'échecs Gary Kasparov. Pour la première fois, la machine a vaincu l'Homme [6].

II.2.3 Domaine d'utilisation de l'IA

L'intelligence artificielle (IA), trouve de nombreuses applications dans la société actuelle. Elle a été utilisée pour améliorer et faire progresser divers domaines, tels que :

- **La santé** : De nombreux systèmes utilisant l'intelligence artificielle sont capables de diagnostiquer des maladies, notamment des cancers, de manière aussi précise, que les spécialistes. L'intelligence artificielle est aussi utilisée pour accélérer le processus de développement de nouveaux médicaments, réduisant ainsi le délai entre la découverte d'une nouvelle molécule et sa commercialisation.
- **L'industrie** : La robotique moderne repose largement sur l'intelligence artificielle. Cela se manifeste notamment dans des robots capables d'apprendre des gestes à partir de manipulations effectuées par des opérateurs. De plus, elle facilite l'automatisation, garantissant ainsi une production rapide et une planification efficace.
- **L'agriculture** : L'intelligence artificielle (IA) offre aux agriculteurs une palette d'outils novateurs pour optimiser la surveillance des cultures, accroître les rendements et piloter leur activité de manière plus durable et efficace.
- **Transport** : L'IA permet aux véhicules de circuler sans intervention humaine, en utilisant des algorithmes de prise de décision. Les systèmes de navigation utilisent L'IA pour optimiser les trajets en fonction de trafic et des incidents routiers.
- **sécurité** : Les systèmes de caméras équipés d'IA surveillent les espaces privés et publics et identifient des comportements suspects et des intrusions.

II.2.4 Différentes branches de l'IA

Il existe plusieurs branches de l'intelligence artificielle, chacune adaptée à des problématiques spécifiques. Les trois branches principales sont :

II.2.4.1 IA symbolique

Manipule des symboles et cette approche utilise des règles logiques pour résoudre des problèmes complexes. Ces principes de base sont les suivants :

- **Représentation symbolique** : Utilise des représentations symboliques de problème, de concept et de relations.
- **Raisonnement logique** : Les systèmes d'IA symbolique appliquent des règles logiques à ces représentations, pour tirer des conclusions, faire des inférences et résoudre des problèmes.
- **Connaissances explicites** : Repose sur des connaissances préalablement définies des règles codées à la main.

II.2.4.2 IA basée sur les connaissances

L'IA basée sur les connaissances repose sur l'utilisation d'une base de connaissances pour simuler le raisonnement humain et prendre des décisions. Ces caractéristiques principales sont :

- **Représentation des connaissances** : Utilise des structures de données pour représenter les informations, cela peut inclure des ontologies, des graphes de connaissances ou des systèmes de règles.
- **Raisonnement** : Cela implique des techniques comme le raisonnement déductif ou inductif.
- **système expert** : Conçu pour résoudre des problèmes spécifiques dans un domaine particulier en utilisant des règles dérivées de l'expertise humaine.
- **interactivité** : Affiner les réponses de l'utilisateur ou pour apprendre de nouvelles informations, ce qui les rend adaptables et personnalisables.

II.2.4.3 Apprentissage automatique (Machine learning)

L'apprentissage automatique (Machine Learning) est une branche de l'intelligence artificielle qui se concentre sur le développement d'algorithmes permettant aux machines d'apprendre et de faire des prédictions ou des décisions basées sur des données.

II.3 Machine learning

II.3.1 Définition

Un chercheur informaticien qui porte le nom de Arthur Samuel, a été le premier à définir l'apprentissage automatique en 1959. Il l'a défini comme un domaine d'étude qui permet aux ordinateurs d'apprendre sans programmation explicite [7]. En suite un professeur nommé Tom Mitchell développe cette définition en 1998, il a déclaré qu'un programme informatique apprend de l'expérience E pour une tâche T et une mesure de performance P si ces performances sur T , mesurées par P à enseigné à une machine à accumuler des connaissances à partir d'expériences afin de résoudre un problème donné de manière optimale [8].

Le machine learning, ou apprentissage automatique, est un sous-ensemble de l'intelligence artificielle (IA) qui se concentre sur le développement d'algorithmes et de modèles capables d'apprendre à partir de données. Contrairement aux systèmes traditionnels qui nécessitent des instructions explicites pour effectuer des tâches, le machine learning permet aux ordinateurs d'identifier des motifs, de faire des prédictions et d'améliorer leurs performances au fil du temps sans intervention humaine directe.

Cette discipline utilise des techniques variées, telles que les réseaux de neurones, les arbres de décision et les machines à vecteurs de support, pour traiter des ensembles de données massifs et complexes. Le machine learning peut être appliqué à une multitude de domaines, allant de la reconnaissance d'images et du traitement du langage naturel à la prévision des ventes et à la détection de fraudes.

L'un des aspects les plus fascinants du machine learning est sa capacité à s'adapter et à évoluer en fonction des nouvelles données, ce qui le rend particulièrement puissant dans un monde où les informations changent rapidement. En intégrant des concepts de statistiques, d'algorithmique et de théorie de l'information, le machine learning représente une avancée majeure dans la manière dont les machines peuvent interagir avec le monde qui les entoure, ouvrant la voie à des applications innovantes et à des solutions intelligentes pour des problèmes complexes [8].

II.3.2 Types d'apprentissage automatique

Il existe plusieurs types de machine learning employant chacun des techniques algorithmiques différentes. Le choix du type est dicté par la nature des données et le résultat souhaité. Les trois principaux types d'apprentissage

sont :

II.3.2.1 Apprentissage supervisé

Dans ce type d'apprentissage, la machine apprend par l'exemple. C'est une technique qui utilise donc des données étiquetées au préalable. les ensembles de données comportent a la fois des variables d'entrées X, et des variables de sorties Y qui ont déjà été associées a la bonne réponse. Un algorithme pour apprendre la fonction de correspondance entre l'entrée et la sortie est utilisé [9](voir figure II.1) .

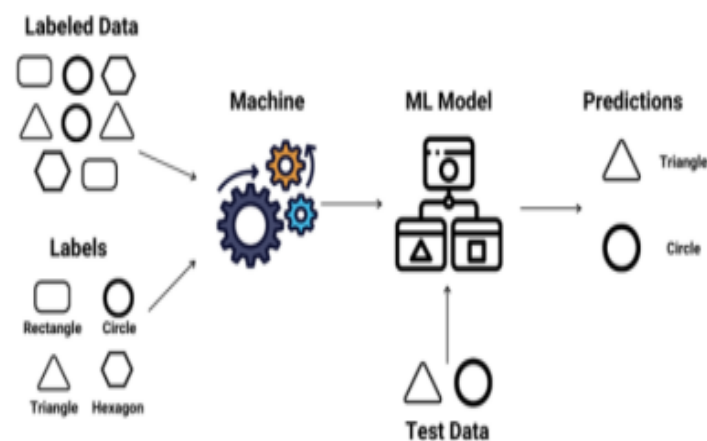


FIGURE II.1 – Schéma de l'apprentissage supervisé

II.3.2.2 Apprentissage non supervisé

Contrairement à l'apprentissage supervisé, l'apprentissage non supervisé se fait sans la présence des données étiquetées. Aucune donnée d'apprentissage n'est fournie au réseau, mais simplement des données d'entrée X sans sortie associée. Dans ce cas la c'est le réseau qui doit trouver seul des distributions intéressantes et des corrélations dans les données (voir figure II.2)

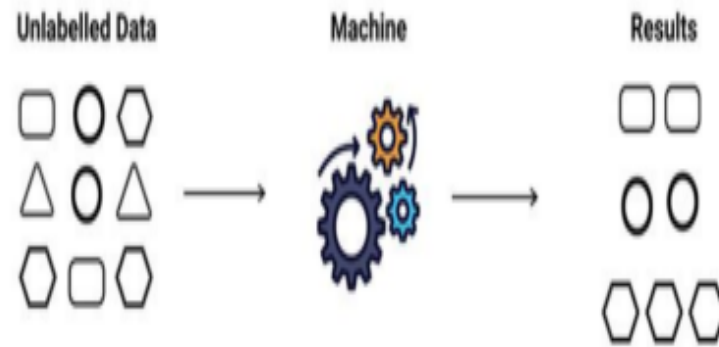


FIGURE II.2 – Schéma de l'apprentissage non supervisé

II.3.2.3 Apprentissage par renforcement

L'apprentissage par renforcement est considéré comme une sorte d'apprentissage supervisé. Dans cette approche le réseau doit apprendre la corrélation entrée/sortie via une estimation de son erreur, c'est à dire du rapport échecs/succès.

Le réseau va donc tendre à maximiser un index de performance qui lui est fourni, appelé signal de renforcement. Le système étant capable de savoir si la réponse qu'il fournit est correcte ou non, mais il ne connaît pas la bonne réponse.

II.3.3 Diagramme des relations entre les concepts

Les termes " intelligence artificielle " et " machine learning " et " Deep Learning " employés dans ce chapitre, pour mieux comprendre comment ces concepts sont structurés, et les relations entre eux, nous représentons un schéma explicatif en II.3 qui regroupe tous ces éléments.

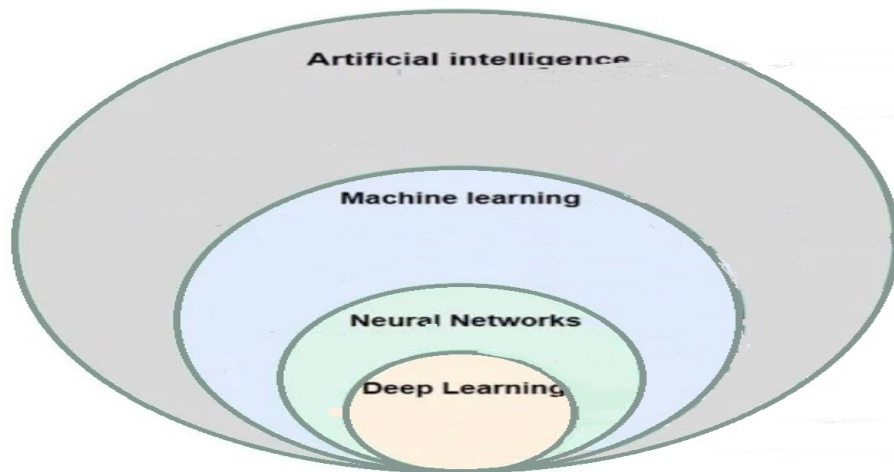


FIGURE II.3 – Diagramme des relations entre les concepts

Ce schéma montre clairement que le Machine Learning est un sous-ensemble de l'intelligence artificielle, tandis que le Deep Learning est lui-même un sous-ensemble du Machine Learning, basé sur l'utilisation des réseaux de neurones.

Avant de développer la notion de Deep Learning nous allons d'abord développer celle de réseaux de neurones dans le paragraphe suivant.

II.4 Réseaux de neurones

Les réseaux de neurones, inspirés du cerveau humain, sont constitués de couches inter-connectées de neurones qui apprennent à accomplir des tâches en ajustant les poids des connexions. Leur utilisation s'est largement popularisée grâce à l'augmentation de la puissance de calcul et à l'accès à de grandes quantités de données.

II.4.1 Historique des réseaux de neurones

Le terme de réseaux de neurones, signifie la possibilité de modéliser le cerveau humain en ensemble de fonction mathématique pour reproduire certaines de ces fonctions. Les origines de cette discipline sont très diversifiées.

- En 1943, Mc Culloch et Pitt étudièrent un ensemble de neurones formels interconnecter et montrèrent leur capacité à calculer certaines fonction logiques.
- En 1958, Rosenblatt décrivit le premier modèle opérationnel de réseaux de neurones : le Perceptron inspiré du système visuel.

- En 1969, Minsky et Papert, deux mathématiciens démontrent les limites théoriques du Perceptron ce qui ralentit la recherche dans ce domaine.
- Dans les années 1980, l'invention de l'algorithme de rétro-propagation du gradient ravive de l'intérêt pour les réseaux de neurones. Depuis les années 2000 les réseaux de neurones jouent un rôle centrale dans l'intelligence artificielle et particulièrement le Deep Learning.

II.4.2 Neurones biologiques

Le neurone, unité fondamentale du tissu nerveux, se compose de dendrites, qui servent d'entrées, d'un axone, qui constitue sa sortie, et de synapses, assurant la transmission du signal entre les neurones (voir figure II.4).

Le neurone est une entité polarisée, transmettant l'information dans un seul sens : des dendrites vers l'axone.

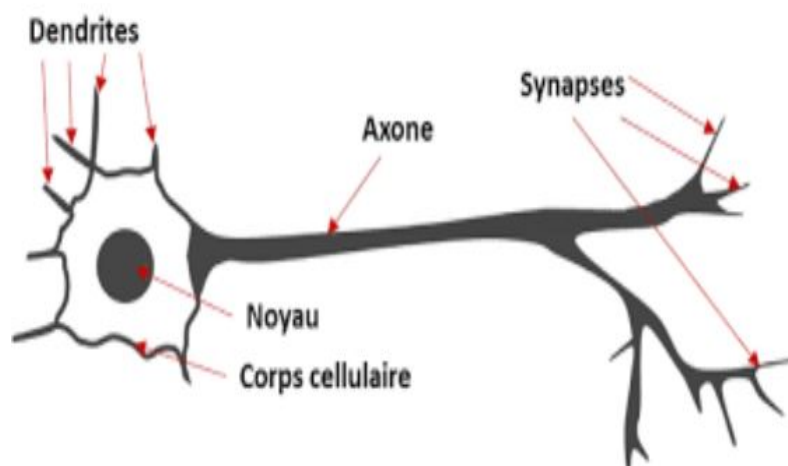


FIGURE II.4 – Le neurone biologique

II.4.3 Neurone formel

Un neurone formel est une fonction algébrique non linéaire est bornée. Les variables de cette fonction sont appelées entrées de neurone et la valeur de la fonction est appelée sortie. La valeur de la sortie dépend des paramètres appelés coefficients ou poids (voir figure II.5).

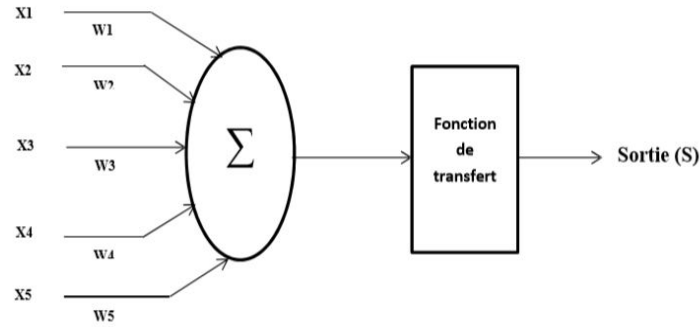


FIGURE II.5 – Fonction réalisée par le neurone formel

Soit S : la sortie calculé par le neurone

P : la somme des entrées pondérées ;

F : une fonction d'activation ;

La formule pour P est donnée par :

$$P = \sum_{i=1}^n x_i w_i + b \quad (\text{II.1})$$

x_i : est une entrées du neurone.

$w_i = w1.....wn$: Poids synaptique.

b : Est le biais représentant le seuil d'activation du neurone.

La sortie du neurone est donnée par :

$$S = F(P) \quad (\text{II.2})$$

La fonction d'activation F définit la valeur de sortie du neurone. Cette fonction peut prendre différentes formes possibles, comme le montre le tableau ci-dessous :

Nom de la fonction	Relation d'entrée/sortie	Icône	Nom Matlab
seuil	$a = 0$ si $n < 0$ $a = 1$ si $n \geq 0$		hardlim
seuil symétrique	$a = -1$ si $n < 0$ $a = 1$ si $n \geq 0$		hardlims
linéaire	$a = n$		purelin
linéaire saturée	$a = 0$ si $n < 0$ $a = n$ si $0 \leq n \leq 1$ $a = 1$ si $n > 1$		satlin
linéaire saturée symétrique	$a = -1$ si $n < -1$ $a = n$ si $-1 \leq n \leq 1$ $a = 1$ si $n > 1$		satlins
linéaire positive	$a = 0$ si $n < 0$ $a = n$ si $n \geq 0$		poslin
sigmoïde	$a = \frac{1}{1+\exp^{-n}}$		logsig
tangente hyperbolique	$a = \frac{e^n - e^{-n}}{e^n + e^{-n}}$		tansig
compétitive	$a = 1$ si n maximum $a = 0$ autrement		compet

II.4.4 Définition et exemple de réseaux de neurones

II.4.4.1 Définition

Un réseau de neurones artificiels (RNA) est un ensemble de neurones formels interconnectés.

En général, ces neurones sont associés en couches qui fonctionnent en parallèle.

Chaque neurone d'une couche est connecté à tous les neurones de la couche précédente et à tous les neurones de la couche suivante.

Les RNA sont classés en deux grandes catégories : les réseaux non bouclés (Feed Forward) et les réseaux bouclés (Feed Back).

Les réseaux non bouclés est un ensemble de neurones connectés où l'information circule des entrées vers les sorties sans retour en arrière alors que dans les réseaux bouclés, il y'a retour en arrière de l'information.

Le perceptron est l'exemple type d'un réseau non bouclé

II.4.4.2 Perceptron multicouches

Un Perceptron multicouches (MLP) est constitué des neurones d'entrées, des neurones cachés et des neurones de sorties, permet de réaliser des fonc-

tions non linéaires comme le montre la figure suivante. Le neurone le plus actif de la couche de sortie représente la classe d'appartenance du vecteur d'entrée[10].

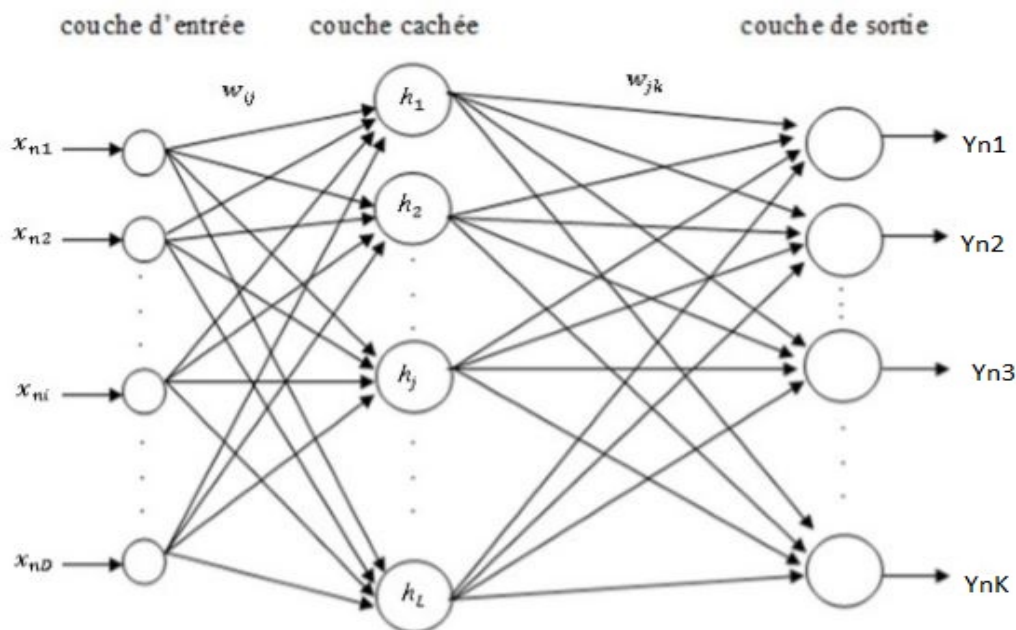


FIGURE II.6 – Architecture d'un perceptron multicouches

L'apprentissage du réseau de neurone est le processus de détermination de ces poids.

II.4.5 Processus d'apprentissage MLP

Lors du processus d'apprentissage, un réseau de neurones reçoit un ensemble d'exemples, chacun étant constitué d'une paire (entrée, sortie désirée). Le réseau effectue alors des calculs pour déterminer les sorties correspondantes en se propageant de la couche d'entrée à la couche de sortie [11]. Soit l'ensemble de données $x_n = [x_{n1}, x_{n2}, \dots, x_{ni}, \dots, x_{nD}]$ est présenté en entrée, la réponse du j -ème neurone de la couche cachée h est calculée par :

$$h_j = F \left(\sum_{i=1}^D w_{ij} x_{ni} + x_0 \right) \quad (\text{II.3})$$

Avec D correspondant à la dimension des données d'entrée.

Au niveau de la couche de sortie, la réponse du k -ème neurone est exprimée par :

$$y_k = F \left(\sum_{j=1}^L w_{jk} h_j + h_0 \right) \quad (\text{II.4})$$

L est le nombre de neurone de couche h .

Les valeurs x_0 et h_0 sont des biais, fixés généralement à 1, F est la fonction d'activation. L'apprentissage du réseau MLP est effectué en utilisant la règle de rétro-propagation de gradient dont le principe consiste à propager vers les couches internes l'erreur quadratique, en commise entre la sortie du réseau $F(x_n) = y_n$ et celle désirée d_n [11].

$$E_n = \frac{1}{2} \left(\sum_{k=1}^K (d_{nk} - y_{nk})^2 \right) \quad (\text{II.5})$$

K est le nombre de neurone de la couche de sortie.

Cette erreur est rétropropagée à travers les couches du réseau afin de modifier les poids synaptiques, de manière à minimiser l'erreur entre les sorties du réseau et les sorties désirées lors de la prochaine itération. Ce processus est répété jusqu'à convergence, c'est-à-dire jusqu'à ce que l'erreur quadratique moyenne E observée sur l'ensemble des données d'apprentissage soit inférieure à un seuil prédéfini ou que le nombre d'itérations spécifié soit atteint.

$$E = \frac{1}{N} \left(\sum_{n=1}^N E_n \right) \quad (\text{II.6})$$

L'adaptation des poids synaptiques par la méthode du gradient est basée sur les formules itératives suivantes :

$$w_{jk}(t+1) = w_{jk}(t) - \eta \Delta w_{jk}(t) \quad (\text{II.7})$$

$$w_{ij}(t+1) = w_{ij}(t) - \eta \Delta w_{ij}(t) \quad (\text{II.8})$$

où η est le pas d'apprentissage.

et :

$$\Delta w_{jk} = \frac{\partial E_n}{\partial w_{jk}} = h_j E_k = h_j y_k (1 - y_k) (d_k - y_k) \quad (\text{II.9})$$

$$\Delta w_{ij} = \frac{\partial E_n}{\partial w_{ij}} = x_i E_j = x_i h_j (1 - h_j) \sum_{k=1}^K w_{jk} E_k \quad (\text{II.10})$$

II.5 Deep Learning

II.5.1 Définition

Le Deep Learning ou apprentissage profond est une branche du Machine Learning qui utilise des réseaux de neurones à pour apprendre aux machines à faire ce qui semble naturel aux humains. Dans le Deep Learning, un modèle apprend à effectuer des tâches de classification ou de régression directement à partir de données telles que des images, du texte ou du son. Les modèles de Deep Learning sont fondés sur des architectures de réseaux de neurones. Le terme « deep » fait généralement référence au nombre de couches cachées dans le réseau de neurones [12][13].

II.5.2 Types de modèles de Deep Learning

Il existe trois types de modèles d'apprentissage profond : les réseaux de neurones convolutifs (CNN), les réseaux de neurones récurrents (RNN) et les modèles de transformateur .

- **Réseaux de neurones à convolution (CNN)** : Les CNN sont spécialement conçus pour traiter des données structurées sous forme de grille .Il effectuent une convolution des caractéristiques apprises avec les données d'entrée et utilisent des couches convolutives 2D, ce qui rend cette architecture bien adaptée au traitement des données 2D, telles que les images[12]. Le CNN fonctionne en extrayant des caractéristiques directement à partir des images. Les caractéristiques pertinentes sont apprises pendant que le réseau s'entraîne sur une collection d'images.
- **Réseaux de neurones récurrents (RNN)** : Les RNN sont utilisés pour traiter des séquences de données, telles que des séquences temporelles ou des séquences de mots dans un texte. Ils sont capables de conserver une mémoire interne, ce qui les rend adaptés à des tâches telles que la traduction automatique ou la génération de texte[13].
- **Transformateurs** : Les transformateurs sont des architectures de réseau de neurones qui ont révolutionné le traitement du langage naturel. Ils sont utilisés pour des tâches telles que la traduction automatique, la génération de texte et la compréhension du langage naturel[13].

II.5.3 Les réseaux de neurones convolutifs (CNN)

Les CNN sont largement utilisées pour l'analyse d'images. Ils sont conçus pour exploiter la structure spatiale des données en appliquant des filtres qui balayent l'image pour extraire des caractéristique locales.

Un réseau CNN est généralement composé de plusieurs types de couches (voir figure II.7) :

- la couche de convolution (conv).
- la couche de pooling (POOL).
- la couche de correction (ReLU).
- la couche entièrement connectée (FC).
- la couche de perte (LOSS).

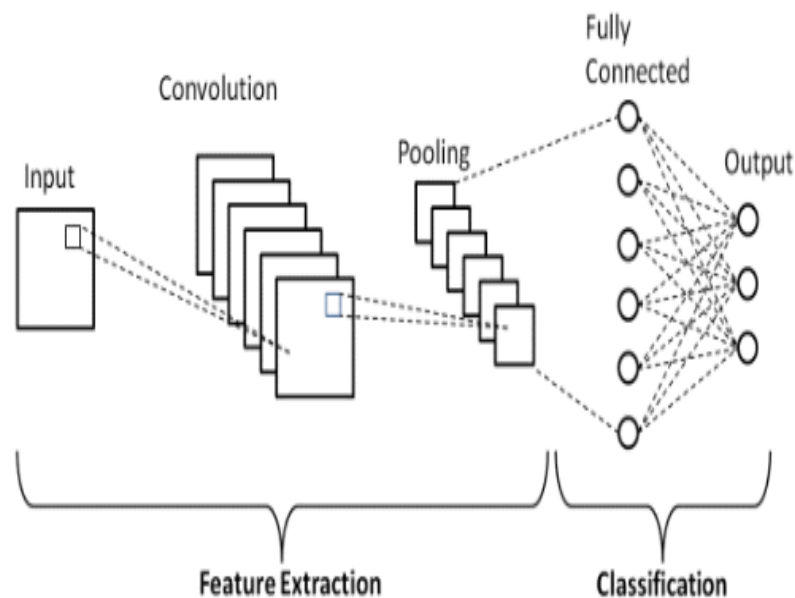


FIGURE II.7 – Différentes couches de CNN

Explications de ces différentes couches :

- **Couche de Convolution** : C'est la couche clé qui applique des filtres (ou noyaux de convolution) sur l'image d'entrée pour extraire des caractéristiques. Chaque filtre est conçu pour détecter des motifs spécifiques, comme des bords ou des textures. Les résultats de cette opération sont appelés feature maps.
- **Couche de Pooling** : Cette couche réduit la dimensionnalité des feature maps, permettant de diminuer le nombre de paramètres et de calculs dans le réseau. Le pooling aide également à rendre les caractéristiques détectées plus invariantes aux transformations comme la translation.
- **Couche d'activation (ReLU)** : Après chaque opération de convolution, une fonction d'activation comme ReLU (Rectified Linear Unit) est souvent appliquée pour introduire de la non-linéarité dans le modèle.
- **La couche entièrement connectée** : À la fin du réseau, une ou plusieurs couches entièrement connectées (MLP) sont utilisées pour la classification finale. Ces couches prennent les caractéristiques extraites par les couches précédentes et les transforment en prédictions de classes.
- **La couche de perte (LOSS)** : La couche de loss, ou fonction de perte, est un élément essentiel dans l'entraînement des réseaux de neurones, car elle évalue la performance du modèle en quantifiant l'écart entre les prédictions du modèle et les valeurs cibles réelles. Cette mesure est cruciale pour guider le processus d'optimisation qui permet d'ajuster les poids et les biais du modèle afin d'améliorer ses performances.

II.5.4 Applications du Deep Learning

Les applications du Deep Learning sont nombreuses et on les retrouve dans plusieurs domaines :

- **La reconnaissance faciale** : Les yeux, le nez, la bouche, tout autant de caractéristiques qu'un algorithme de DL va apprendre à détecter sur une photo. Il va s'agir en premier lieu de donner un certain nombre d'images à l'algorithme, puis à force d'entraînement, l'algorithme va être en mesure de détecter et reconnaître un visage sur une image.
- **Le traitement automatique de langage naturel** : Le traitement automatique de langage naturel est une autre application du DL. Son but étant d'extraire le sens des mots, voire des phrases pour faire de l'analyse de sentiments. L'algorithme va par exemple comprendre ce qui est dit dans un avis Google, ou va communiquer avec des personnes via

des chatbots. La lecture et l'analyse automatique de textes est aussi un des champs d'application du DL.

- **Voitures autonomes** : Les entreprises qui construisent de tels types de services d'aide à la conduite, ainsi que des voitures autonomes doivent apprendre à un ordinateur à maîtriser certaines parties essentielles de la conduite à l'aide de systèmes de capteurs numériques. Pour ce faire, ces entreprises commencent généralement par entraîner des algorithmes en utilisant une grande quantité de données. A fin de fournir des modèles commerciaux inattendus aux entreprises.
- **Reconnaissance d'image** : Un autre domaine populaire en matière de DL est la reconnaissance d'images. Son objectif est de reconnaître et d'identifier les personnes et les objets dans les images, ainsi que de comprendre le contenu et le contexte. La reconnaissance d'image est déjà utilisée dans plusieurs secteurs tels que les jeux, les médias sociaux, la vente au détail, le tourisme, etc. Cette tâche nécessite la classification des objets d'une photo parmi un ensemble d'objets connus auparavant. Une variante plus complexe de cette tâche, appelée détection d'objet, consiste à identifier spécifiquement un ou plusieurs objets dans la scène de la photo et à dessiner un cadre autour d'eux.

II.5.5 Segmentation sémantique des images

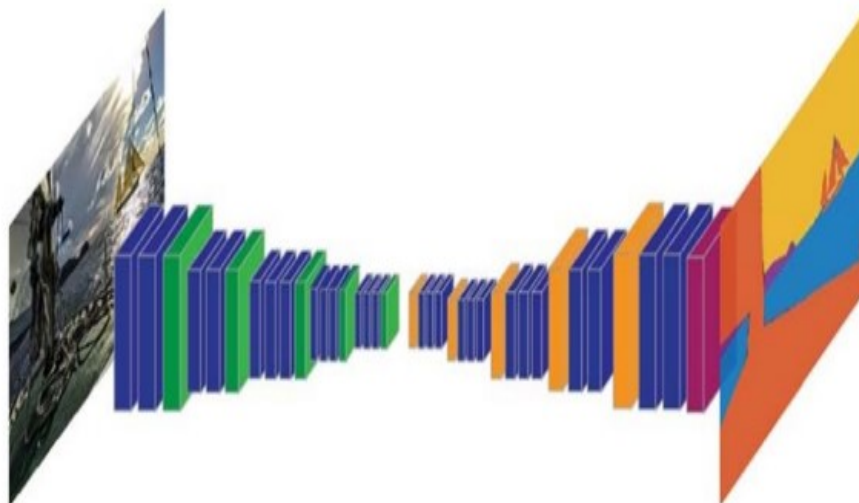


FIGURE II.8 – Structure d'un CNN

Ce réseau est une architecture de réseau neuronal profond qui est utilisée souvent pour la segmentation d'images et la détection d'objet et la classification d'images. Il se distingue par sa capacité d'extraire des caractéristiques pertinentes à partir d'images grâce à ces trois couches de convolution qui sont entièrement connectées effectuent le raisonnement de haut niveau dans le réseau neuronal en connectant chaque nœud de la couche de sortie à un nœud de la couche précédente. Elle exploite généralement une fonction d'activation softmax pour classer les entrées du manière appropriée, produisant une probabilité de 0 à 1.

les réseaux de neurones convolutifs combinent ces trois types de couches pour traiter les images de manière hiérarchique. Les couches de convolution extraient des caractéristiques locales, les couches de pooling réduisent la dimensionnalité et les couches entièrement connectées classifient les caractéristiques extraites. Cette architecture permet aux CNN de réaliser des performances exceptionnelles dans de nombreuses tâches de vision par ordinateur et d'autres domaines de l'intelligence artificielle.

II.6 Conclusion

En conclusion, ce chapitre a permis de faire un tour d'horizon des techniques d'apprentissage modernes, en mettant en lumière deux méthodes particulièrement influentes : l'intelligence Artificielle (IA) et le Deep Learning. L'IA et plus particulièrement le Deep Learning représentent non seulement des outils puissants pour résoudre des problèmes complexes, mais aussi des catalyseurs de transformation pour notre société moderne. À mesure que ces technologies continuent de se développer, elles promettent de créer de nouvelles opportunités tout en nécessitant une vigilance constante pour en gérer les impacts. Ce chapitre a donc posé les bases pour comprendre l'importance de ces techniques et leur potentiel à façonner l'avenir.

À la fin de ce chapitre, nous avons développé une application du Deep Learning qui est la segmentation sémantique.

En effet, cette segmentation sémantique sera utilisée pour la mise en œuvre de notre système de surveillance.

Chapitre III

Mise en Œuvre Tests et Résultats

III.1 Introduction

Dans ce chapitre, nous détaillerons la mise en œuvre du système de surveillance vidéo.

Celui-ci se divise en deux parties : la partie matérielle et la partie logicielle . La partie matérielle est relativement simple, c'est pour cela qu'elle ne sera pas détaillée dans ce chapitre. Elle consiste essentiellement à positionner la caméra dans le lieu à surveiller et à la connecter au système de traitement à travers des liaisons adaptées, qu'il s'agisse d'une connexion filaire, sans fil (WI-fi), ou autre.

Cette connexion garantit la transmission des vidéos au PC central où se trouve la partie logicielle du système.

C'est justement cette dernière partie qui sera détaillée dans ce chapitre.

Nous présenterons un schéma de principe donnant les différentes étapes du fonctionnement du système, chaque étape sera par la suite détaillée.

III.2 Schéma de principe

Le schéma de principe de la mise en œuvre du système de vidéo surveillance est donné dans la figure III.1.

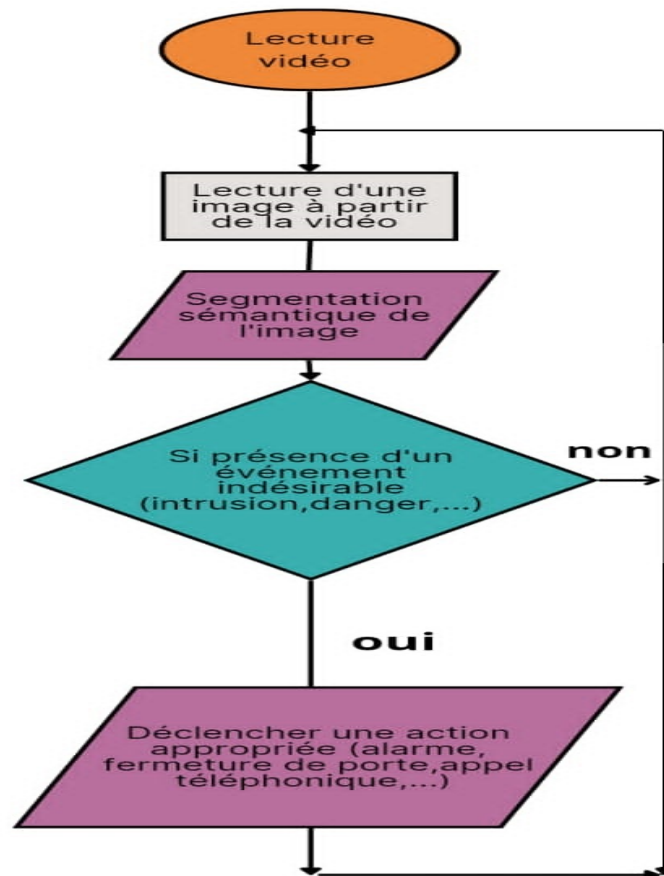


FIGURE III.1 – Schéma de principe du système de vidéo surveillance autonome

Comme le montre la figure III.1, le système de vidéo de surveillance mis en œuvre se compose de plusieurs parties. Bien sûr la vidéo doit être lue par le système et pour réaliser la surveillance automatique, il faut traiter image par image. Donc l'image au temps t est extraite de la vidéo et subira un traitement approprié avant de lire une autre image de la vidéo et ainsi de suite. La partie principale de ce système est la segmentation sémantique de l'image. Elle fera l'objet d'une attention particulière et sa mise en œuvre sera détaillée dans le paragraphe suivant.

Cette partie est essentielle car chaque objet de l'image sera étiqueté et reconnu, et ces résultats conditionneront la suite du système.

Si elle est bien réalisée, l'étape suivante qui consiste à détecter s'il y a présence d'un événement indésirable sera aisée à mettre en œuvre et par conséquent l'étape qui consiste à déclencher une action appropriée le sera aussi.

III.3 Segmentation sémantique

La mise en œuvre du programme de segmentation sémantique est cruciale pour tirer parti des techniques de Deep Learning et obtenir des résultats performants. Cette section détaille la structure du programme et les étapes clés de son implémentation.

Pour réaliser la segmentation sémantique de l'image lu à partir de la vidéo, deux options s'offrent à nous : Soit concevoir et entraîner notre propre réseau de neurones convolutif, soit utiliser un réseau de neurones convolutif pré-entraîné et continuer son entraînement avec des images semblables à celles de notre application.

La première option serait complexe à mettre en œuvre et nécessiterait un travail considérable en termes et temps et de ressources.

Pour toutes ces raisons, nous avons choisi d'utiliser un réseau pré-entraîné. Nous avons testé dans ce travail trois réseaux pré-entraînés les résultats de chacun seront donnés dans les sections suivantes.

Pour continuer l'entraînement de ce réseaux nous avons implémenter un programme en quatre sections clés :

- chargement des données.
- Pré-traitement des données.
- Augmentation des données.
- Entraînement du modèle.

III.3.1 Chargement des données

Pour continuer l'entraînement du modèle choisit, il faut lui fournir une base de données images avec sa base de données des images étiquettes.

Dans ce travail, nous avons opté pour l'utilisation de la base Cam-vid (Cambridge-driving labeled vidéo dataset) qui fournit des images prises durant la conduite et leurs images des étiquettes qui associent chaque pixel à l'une des 32 classes sémantique(Ciel, Bâtiment, Poteau, Marquage routier, Route, Trottoir, Arbre, Symbole de panneau, Clôture, Voiture, Piéton, Cycliste, Motocycliste, Passage pour piétons, Enfant, Bus, Moto, Animal, Feu de circulation, Trottoir, Poteau de colonne, Pont, Piste cyclable, Camion, Train, Bloc de stationnement, Cône de signalisation, Garage, Terrain, Mur, Végétation, Autre objet en mouvement.).La base que nous avons récupérer contient 641 images de taille 480*360 et 641 images des étiquettes de taille 960*720.

Sur la figure ci-dessous nous afficherons trois exemples d'images et d'images étiquettes correspondantes tirées de la base Cam-vid.

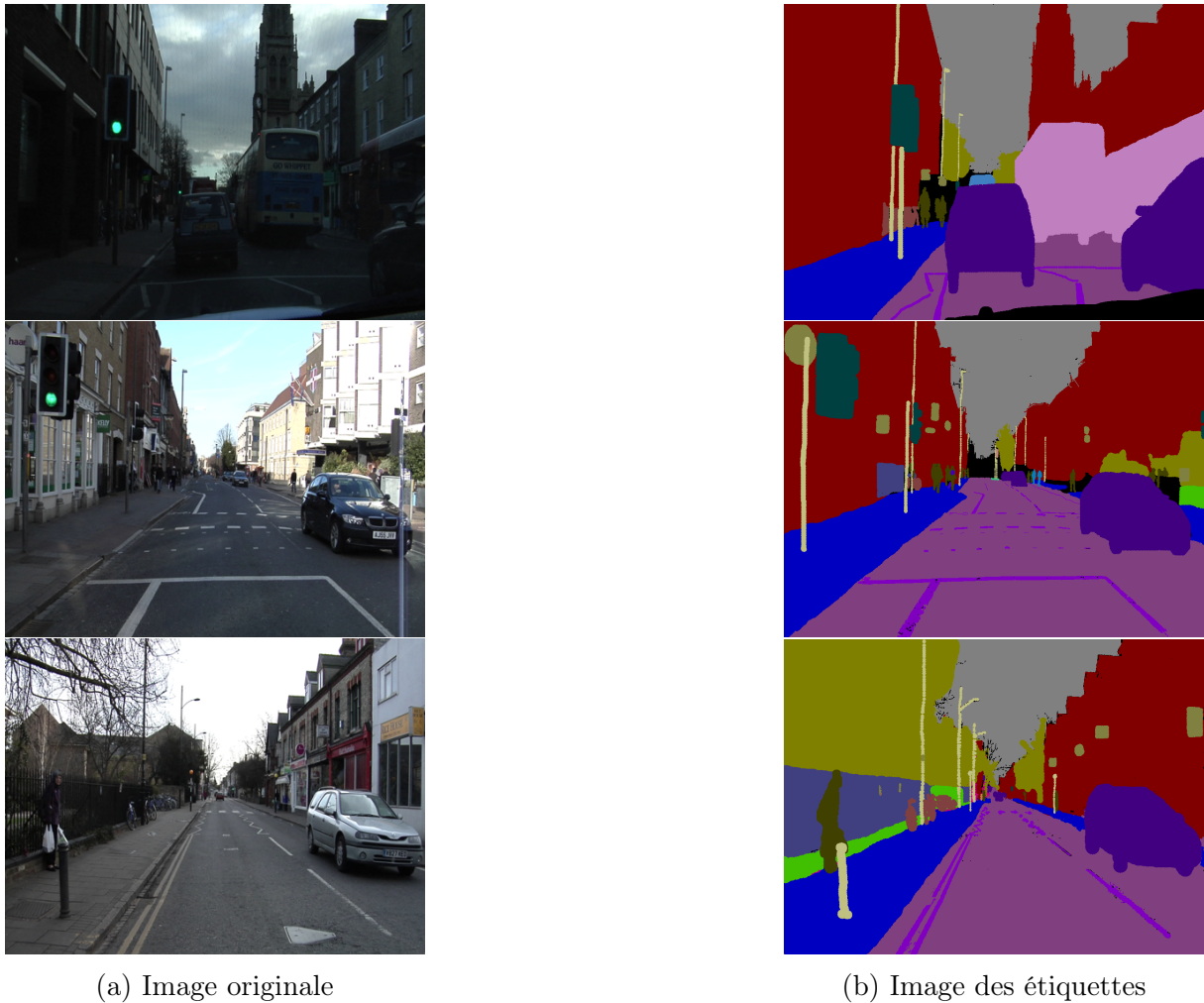


FIGURE III.2 – Images et Images des étiquettes de la base Cam-Vid

III.3.2 Prétraitement des données

Dans cette section, les images et les images des étiquettes sont redimensionnées à la même dimension à savoir 960*720 pixels.

En effet cette étape est essentielle car pour chaque pixel de l'image originale, doit être associé à un pixel de l'image des étiquettes.

Nous avons utiliser 11 classes sémantiques sur les 32 classes. Ces 11 classes sont (Ciel, Bâtiment, Poteau, Route, Trottoir, Arbre, Panneau, Clôture, Voiture, Piéton, Bicycliste.)

III.3.3 Augmentation des données

L'augmentation des données est une technique essentielle dans le processus d'entraînement des modèles de segmentation sémantique. Elle consiste à créer des variations des données d'entraînement existantes afin d'enrichir le jeu de données sans avoir besoin de collecter de nouvelles images.

Elle permet d'améliorer la généralisation en augmentant la diversité des données d'entraînement. Le modèle devient plus robuste et capable de généraliser ses performances sur des données non vues. L'augmentation aide aussi à prévenir le sur-apprentissage en fournissant des exemples variés, ce qui permet au modèle d'apprendre des caractéristiques pertinentes plutôt que de mémoriser les données d'entraînement.

Les techniques d'augmentation appliquées dans le programme incluent :

- **La rotation** : Les images sont tournées aléatoirement dans une plage spécifiée pour simuler différentes orientations d'objets.
- **La translation** : Des translations horizontales et verticales sont appliquées pour décaler les objets dans l'image, ce qui aide à simuler des variations de position.
- **La réflexion** : Un retournement horizontal des images est effectué pour introduire des variations supplémentaires, ce qui est particulièrement utile pour les objets symétriques.

Ces techniques d'augmentation contribuent à enrichir l'ensemble de données d'entraînement, améliorant ainsi la performance du modèle de segmentation sémantique.

III.3.4 Entraînement du modèle

Cette section définit l'architecture du modèle de segmentation, configure les options d'entraînement, et lance le processus d'entraînement en utilisant les données augmentées.

Dans notre étude, nous avons exploré trois architectures de réseaux de neurones convolutifs pour la segmentation sémantique : U-Net, SegNet, DeepLabv3+ associé à ResNet50.

U-Net est une architecture largement utilisée dans le domaine de la segmentation d'images, en particulier pour les applications biomédicales. Sa structure en forme de U permet une combinaison efficace des informations d'encodage et de décodage, ce qui est essentiel pour capturer des détails fins (voir la Figure III.3).

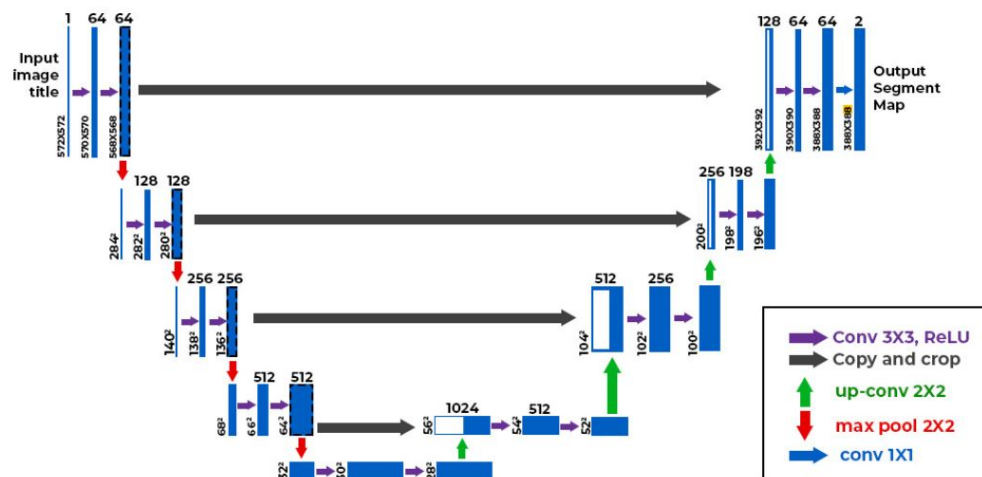


FIGURE III.3 – Architecture de U-Net

SegNet est une architecture qui se concentre également sur la segmentation sémantique. Il utilise une approche d'encodage-décodage, mais se distingue par son utilisation de couches de décodage pour reconstruire les images segmentées à partir des caractéristiques extraites (voir la Figure III.4).

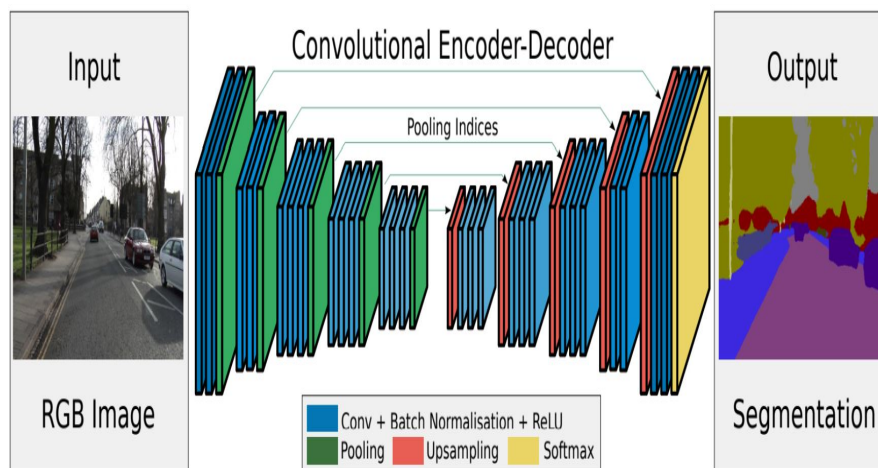


FIGURE III.4 – Architecture de Segnet

DeepLabv3+ associé à **ResNet50** est une architecture avancée qui utilise des convolutions dilatées et un module ASPP (Atrous Spatial Pyramid Pooling) pour capturer des informations contextuelles à différentes échelles. Il a montré des résultats compétitifs dans la segmentation sémantique. En utilisant la fonction `deeplabv3plusLayers(imageSize, numClasses, "resnet50")`, nous créons un réseau de segmentation sémantique qui combine

les avantages de l'architecture DeepLabv3+ avec la robustesse et l'efficacité de ResNet-50 comme réseau de base. Cela permet d'obtenir des résultats de segmentation de meilleure qualité.

La structure de ce réseau est donnée par la Figure III.5.

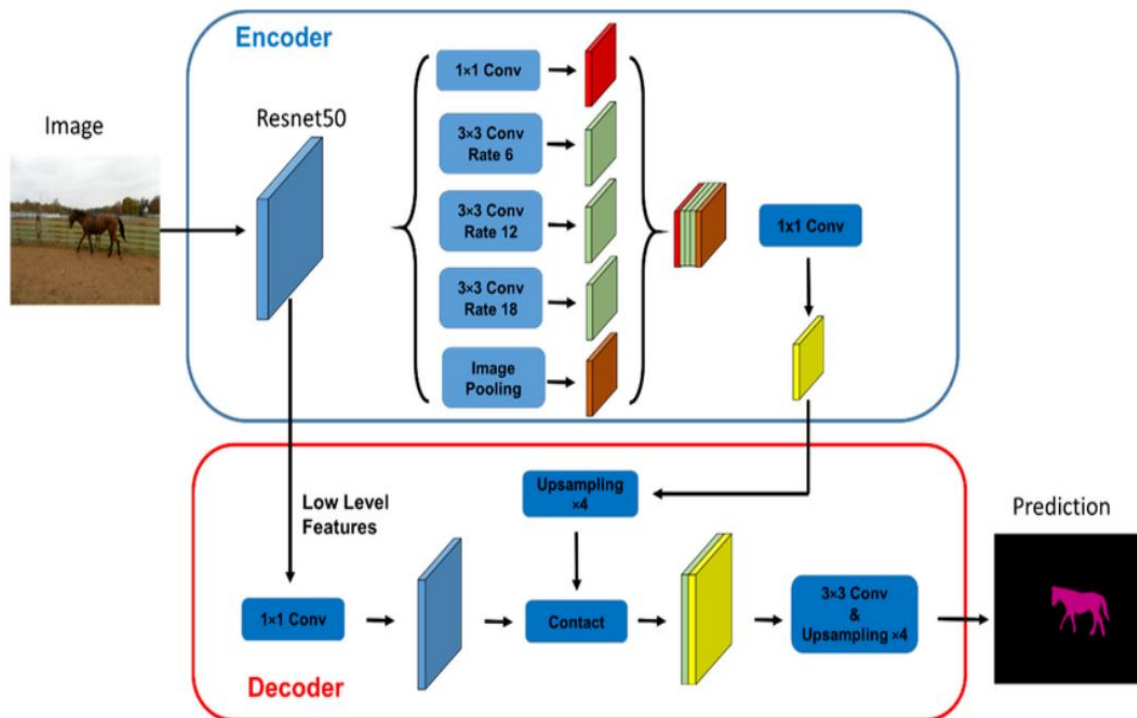


FIGURE III.5 – La structure du réseau combinant Deeplabv3+ et ResNet 50

Durant l'entraînement du modèle, pour chaque réseau, la base de données est partitionnée en trois parties 60% de cette base est utilisé pour l'entraînement, 20% pour la validation et 20% pour le Test. Après affichage des résultats, sur la figure III.6, nous avons constaté que le réseau DeepLabV3+ associée a ResNet50 fournit de meilleurs résultats.

Le temps d'entraînement de chaque réseau est donnée sur le tableau ci-dessous.

Réseau	U-Net	Segnet	DeepLabv3+ associé a ResNet50
Temps d'entraînement en min	1414min	186 min	893 min
Temps d'entraînement en h	23h34min	3h6min	14h53 min

TABLE III.1 – Le temps d'entraînement de chaque réseau

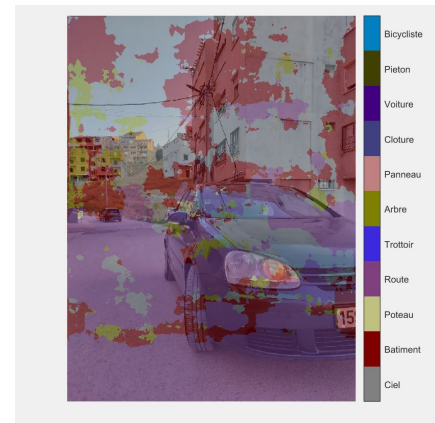
III.4 Choix du réseau

Dans cette section, nous commençons par choisir un réseau avant de présenter les résultats de segmentation sur des exemples d'images capturées à partir d'un flux vidéo.

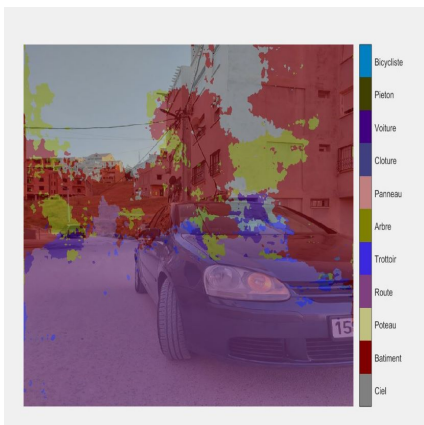
Pour choisir le réseau qui donne les meilleurs résultats, nous affichons ces résultats dans la figure III.6.



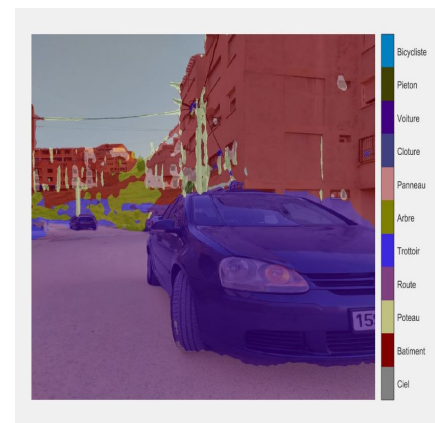
(a) Image originale



(b) Image segmentée avec U-Net



(c) Image segmentée avec SegNet



(d) Image segmentée avec DeepLabv3+ associée à ResNet50

FIGURE III.6 – Segmentation sémantique en utilisant plusieurs réseaux

En observant la figure III.6, nous constatons que les meilleurs résultats sont obtenus pour le réseau DeepLabV3+ associé à ResNet50. Les métriques d'évaluation précision et perte d'entraînement sont affichées sur la figure III.7.



FIGURE III.7 – Métrique d'évolution durant l'entraînement

Terminologie en Apprentissage Automatique

1. Accuracy (Précision)

L'accuracy (précision) est une métrique utilisée pour évaluer la performance d'un modèle de classification. Elle mesure la proportion de prédictions correctes faites par le modèle par rapport au nombre total de prédictions. L'accuracy est calculée comme suit :

$$\text{Accuracy} = \frac{\text{Nombre de prédictions correctes}}{\text{Nombre total de prédictions}}$$

Exemple : Si un modèle a fait 80 prédictions correctes sur 100, son accuracy serait de 0,8 ou 80 %. Cependant, l'accuracy peut être trompeuse dans des contextes où les classes sont déséquilibrées (par exemple, si une classe représente 95 % des données), car un modèle peut avoir une haute précision tout en négligeant complètement la classe minoritaire.

2. Loss (Fonction de Perte)

La loss (fonction de perte) mesure la différence entre les prédictions du modèle et les véritables étiquettes des données. C'est une mesure quantitative qui indique à quel point le modèle fait des erreurs. L'objectif de l'entraînement d'un modèle est de minimiser cette perte. Différentes fonctions de perte peuvent être utilisées en fonction de la tâche :

— **Pour la classification binaire** : On utilise souvent la *binary cross-entropy* :

$$\text{Loss} = -\frac{1}{N} \sum_{i=1}^N [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)] \quad (\text{III.1})$$

où y_i est la vraie étiquette et p_i est la prédiction du modèle.

— **Pour la régression** : La *mean squared error (MSE)* est souvent utilisée :

$$\text{Loss} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (\text{III.2})$$

où $\hat{y}_i$ est la valeur prédite par le modèle et y_i est la valeur réelle. Cette fonction mesure la moyenne des carrés des erreurs entre les prédictions du modèle et les vraies étiquettes. Un objectif commun lors de l'entraînement d'un modèle est de minimiser cette perte afin d'améliorer la précision des prédictions.

— Epoch

Une epoch est une itération complète à travers l'ensemble de données d'entraînement. Pendant une epoch, le modèle reçoit chaque exemple d'entraînement au moins une fois, ajuste ses poids en fonction de la fonction de perte, et essaie d'apprendre à partir des données.

L'entraînement d'un modèle implique généralement plusieurs epochs, car une seule itération est souvent insuffisante pour que le modèle converge vers une solution optimale.

Exemple : Si vous avez 1000 échantillons de données et que vous exécutez 10 epochs, cela signifie que chaque échantillon sera utilisé 10 fois dans le processus d'entraînement.

Nous constatons que le réseau U-Net, L'accuracy passe directement à 100% et le Loss à 0% ceci est le cas typique d'un surapprentissage d'un réseau. Le modèle devient trop spécialisé sur les données d'entraînement de sa capacité à généraliser à de nouvelles données, c'est pour cela que les résultats affichés sur la figure III.6.b, est mauvais.

Pour les deux autres réseaux, à savoir SegNet et DeepLabv3+ associé à Resnet50, leurs métriques d'évolution durant l'entraînement sont assez semblables. Nous avons choisi d'utiliser celui qui fournit les meilleurs résultats visuels. C'est pour toutes ces raisons que nous avons opté pour le réseau DeepLabv3+ associé à Resnet50 dans la suite de notre travail.

III.5 Résultats

Dans cette section, nous fournirons les résultats de la segmentation sur deux flux vidéo.

Nous afficherons pour chacun des flux le résultat sur trois images et sur chaque image nous donnons sa position temporelle dans la vidéo. Les résultats de la vidéo 1 sont fournis sur la figure III.8 et ceux de la vidéo 2 sur la figure III.9.



FIGURE III.8 – Résultats des tests sur la vidéo 1

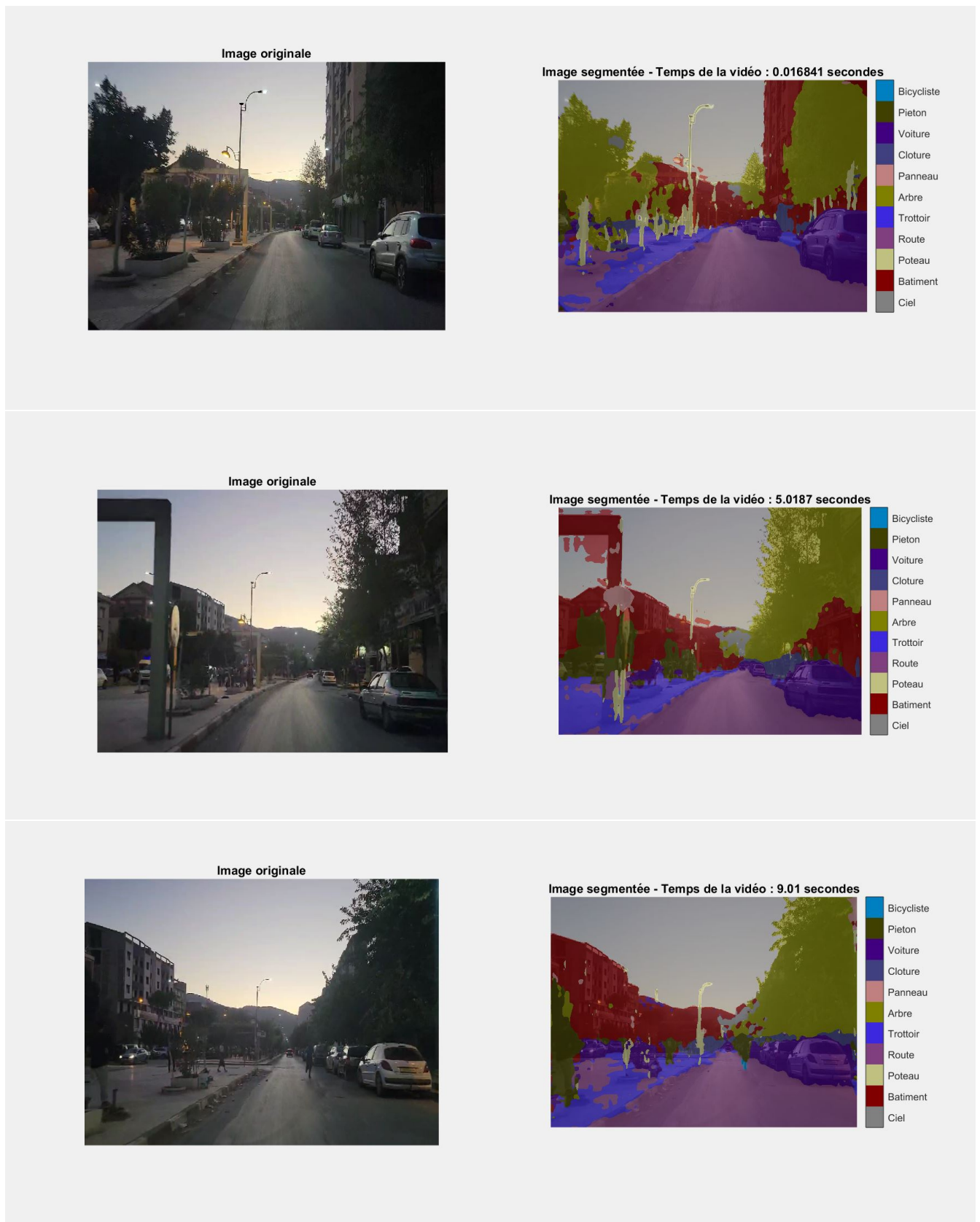


FIGURE III.9 – Résultats des tests sur la vidéo 2

L'efficacité du modèle de segmentation sémantique est mise en lumière par les résultats obtenus sur les images capturées à partir de flux vidéo. Les

performances globales du modèle DeepLab v3+ avec ResNet-50 comme backbone s'avèrent satisfaisantes.

En effet les objets d'intérêt sont bien identifiés tel que le trottoir, les bâtiments, les arbres, le ciel, les piétons et les panneaux, dans la majorité des cas, les contours des objets sont clairement délimités, ce qui est essentiel pour une analyse approfondie et fiable. Cette précision permet de garantir que les informations extraites des images sont non seulement pertinentes, mais également exploitables pour l'étape suivante de surveillance ou l'analyse comportementale. Sachant que ces images sont prises d'une vidéo et que la segmentation sémantique dure environ 4 secondes, il n'est pas possible de traiter toutes les images des vidéos qui présente 30 images par seconde. Le traitement sera réalisé sur 1 image parmi environ 120 images.

En outre la gestion des occlusion est abordé avec une certaine efficacité par le système.

Nous pouvons le constater sur la figure III.10 où l'objet piéton, bien que masqué partiellement par la voiture, est bien identifié.

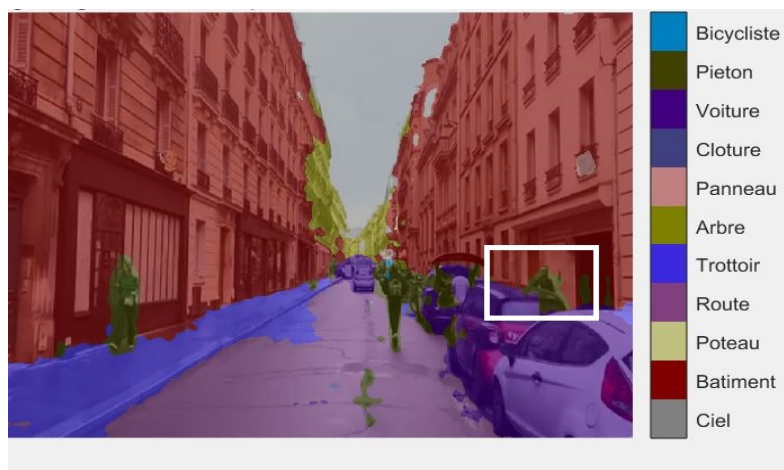


FIGURE III.10 – Exemple d'une occlusion

Ses performances tendent à se dégrader dans des contextes plus complexes comportant de multiples occlusions. Cela met en évidence la nécessité d'améliorer la robustesse du modèle face à des environnements dynamiques et encombrés, ce qui est crucial pour des applications de surveillance en temps réel.

III.6 Difficultés Rencontrées lors de la Mise en Œuvre du Programme

Malgré les avancées réalisées dans la mise en œuvre de la segmentation sémantique, plusieurs défis et limitations ont été rencontrés au cours du développement du programme. Ces problèmes ont nécessité des ajustements pour améliorer les performances et la précision des résultats.

- **Formation du Réseau :** La création d'un réseau de neurones performant pour la segmentation sémantique est une tâche complexe qui nécessite une grande quantité de données étiquetées de haute qualité. La constitution d'un ensemble de données suffisamment varié et représentatif s'est avérée être un défi majeur. Pour surmonter ce problème, l'ensemble de données CamVid disponible sous MATLAB a été utilisé. Cependant, même avec cet ensemble de données, la diversité des images reste limitée et peut entraîner des biais dans le modèle.
- **Temps de Simulation :** L'entraînement d'un réseau de neurones profond pour la segmentation sémantique est extrêmement gourmand en ressources de calcul. L'utilisation d'une unité de traitement graphique (GPU) est indispensable pour accélérer le processus d'apprentissage. Lors des tests initiaux sur CPU, le temps de simulation était très long [893 minutes], rendant l'entraînement impraticable. L'utilisation d'une GPU puissante peu réduire considérablement le temps de calcul, en particulier lorsque le réseau est profond et complexe.
- **Qualité des Résultats :** Bien que les algorithmes de segmentation sémantique aient démontré de bonnes performances sur l'ensemble de test CamVid, il subsiste encore des erreurs et des imprécisions dans les résultats, notamment dans des conditions difficiles (faible luminosité, occultations, scènes encombrées). L'amélioration continue de la précision et de la robustesse des algorithmes reste un défi majeur pour permettre une utilisation fiable dans des applications réelles de vidéo surveillance.

III.7 Conclusion

Ce chapitre a exploré en profondeur la mise en œuvre et les résultats de la segmentation sémantique appliquée à la vidéo surveillance, avec une exploration de trois architectures de réseaux de neurones convolutifs pour la segmentation sémantique : U-Net, SegNet, DeepLabv3+ associé à ResNet50.

Après avoir évalué les performances des trois architectures, DeepLabv3+ associée à ResNet50 a été retenue pour sa capacité à fournir des résultats précis et robustes en segmentation sémantique.

Les résultats de segmentation ont été présentés à travers des exemples d'images capturées à partir d'un flux vidéo, illustrant la capacité du modèle à identifier et à segmenter avec précision des objets d'intérêt dans des environnements complexes. La visualisation des résultats a permis d'évaluer la performance du modèle, mettant en lumière sa capacité à gérer des objets de différentes tailles et à maintenir une résolution élevée des contours.

La formation du réseau a été un défi majeur, notamment en raison de la nécessité d'un ensemble d'images étiquetées de grande qualité. Pour pallier ce problème, nous avons utilisé l'ensemble de données CamVid, disponible sur MATLAB, qui a permis de fournir des exemples variés pour l'entraînement. Néanmoins, la diversité des images reste une contrainte qui peut influencer la généralisation du modèle.

De plus, le temps de simulation s'est révélé être un obstacle significatif. Lors de l'utilisation de la CPU, le processus d'entraînement était extrêmement long, ce qui a rendu l'optimisation et l'expérimentation difficile. L'adoption d'une unité de traitement graphique (GPU) est essentielle pour réduire le temps de calcul, ce qui nécessite une gestion efficace des ressources.

En conclusion, ce chapitre a non seulement mis en évidence les capacités techniques du modèle de segmentation sémantique, mais a également démontré son potentiel dans des applications pratiques de vidéo surveillance, malgré les plusieurs problèmes rencontrés lors de la simulation et de l'entraînement du modèle.

Conclusion générale

Dans ce mémoire, nous avons abordé les différentes tâches liées à la détection d'objets par segmentation d'images dans le cadre de la vidéosurveillance. Nous avons d'abord exploré les principes fondamentaux de la segmentation d'images, soulignant son rôle crucial dans l'isolement des objets d'intérêt au sein d'environnements complexes. L'intégration de l'intelligence artificielle et du Deep Learning a également été mise en avant, démontrant comment ces technologies améliorent la précision des algorithmes de segmentation, permettant ainsi une analyse plus fine des flux vidéo.

Cependant, plusieurs difficultés ont été rencontrées tout au long du processus. La formation du réseau a été un défi majeur, notamment en raison de la nécessité d'un ensemble d'images étiquetées de haute qualité. Pour pallier ce problème, nous avons utilisé l'ensemble de données CamVid, disponible sur MATLAB, qui a fourni un bon point de départ, mais qui a également révélé des limitations en termes de diversité des images, et l'entraînement du modèle sur CPU était extrêmement long. L'adoption d'une unité de traitement graphique (GPU) aurait permis de réduire considérablement ce temps, mais le processus d'entraînement reste exigeant en ressources, nécessitant une gestion efficace des hyperparamètres et des techniques d'augmentation de données.

Ce mémoire a mis l'importance de la vidéosurveillance et du traitement d'images dans le contexte actuel de sécurité. L'utilisation de techniques avancées telles que la segmentation sémantique a démontré son efficacité dans l'identification et le suivi des objets d'intérêt dans un flux vidéo en temps réel. Les résultats obtenus avec le modèle DeepLabv3+ associé à ResNet-50 montrent une performance prometteuse, bien que certaines limitations aient été identifiées, notamment dans la gestion des occlusions complexes.

Bibliographie

- [1] Manuel Guzman et Bert Verstappen. Qu'est-ce que la surveillance ? Série sur la surveillance et la documentation sur les droits de l'homme, vol. 2, p. 3.
- [2] Innovatiana, *Object Detection : Our Guide*, Disponible sur : <https://www.innovatiana.com/post/object-detection-our-guide>
- [3] Howard Gardner. *Frames of Mind : The Theory of Multiple Intelligences*. Basic Books, 1983.
- [4] David Wechsler. *The Measurement of Adult Intelligence*. Williams et Wilkins, 1944.
- [5] Jean Piaget. *The Origins of Intelligence in Children*. International Universities Press, 1952.
- [6] datascientest, *intelligence artificielle definition*, Disponible sur : <https://datascientest.com/intelligence-artificielle-definition5>
- [7] Samuel, A. L. *Some studies in machine learning using the game of checkers*. IBM Journal of Research and Development, 3(3), 210-229, 1959.
- [8] Mitchell, T. M. *Machine Learning*. McGraw-Hill Education, 1st Edition, 01 March 1997.
- [9] B. MOHAMMED et B. BRAHIM. *L'apprentissage profond (Deep Learning) pour la classification et la recherche d'images par le contenu* K. MERBAH Faculté des Nouvelles Technologies de l'Information et de la Communication, OUARGLA, ALGERIE 2017.
- [10] Mellah, R. *Cours de commande intelligente*, Chapitre Réseaux de neurones, Université Mouloud MAMMERI, TIZI OUZOU, ALGERIE, [2023/2024].
- [11] Malika Ait Aider. *Contribution à la reconnaissance automatique de caractères numériques*. Thèse de doctorat, Université Mouloud MAMMERI, TIZI OUZOU, ALGERIE, 2019, pages 74-76.
- [12] MathWorks, *Deep Learning with MATLAB and Simulink*, 2024, Disponible sur : <https://ch.mathworks.com/fr/discovery/deep-learning.html>.

-
- [13] GeeksforGeeks, *Introduction to Deep Learning*, Disponible sur : <https://www.geeksforgeeks.org/introduction-deep-learning/>.
- [14] Ian Goodfellow, Yoshua Bengio, Aaron Courville. *Deep Learning*. MIT Press, 2016.