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

Université Saad Dahlab Blida (USDB)
Faculté des sciences
Département d'informatique

MEMOIRE DE FIN D'ETUDE
En vue de l'obtention du diplôme de Master
Filière : Informatique
Option : Ingénierie des Logiciels (IL)

Par : HAMIDI Kamel

Réalisation d'un système de détection d'intrus exploitant la reconnaissance faciale

Soutenue publiquement devant le jury :

Mme.	FERDI Imene	MCB	USDB 1	Président
Mr.	HAMOUDA Mohamed	MAA	USDB 1	Examinateur
Mlle.	DJEDDAR Afrah	MCB	USDB 1	Encadreur

PROMOTION 2022/2023

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

Université Saad Dahlab Blida (USDB)
Faculté des sciences
Département d'informatique

MEMOIRE DE FIN D'ETUDE
En vue de l'obtention du diplôme de Master
Filière : Informatique
Option : Ingénierie des Logiciels (IL)

Par : HAMIDI Kamel

Réalisation d'un système de détection d'intrus exploitant la reconnaissance faciale

Soutenue publiquement devant le jury :

Mme.	FERDI Imene	MCB	USDB 1	Président
Mr.	HAMOUDA Mohamed	MAA	USDB 1	Examinateur
Mlle.	DJEDDAR Afrah	MCB	USDB 1	Encadreur

PROMOTION 2022/2023

Dédicaces

Mes chers parents,

Je tiens à vous exprimer ma profonde gratitude pour tout le soutien inconditionnel que vous m'avez apporté tout au long de ma vie. Votre amour, votre guidance et votre présence ont été essentiels pour mon parcours et mes réussites. Je vous suis infiniment reconnaissant pour les sacrifices que vous avez consentis pour moi et pour les valeurs que vous m'avez inculquées.

À ma chère femme,

Ma complice de vie, je tiens à te dédier ces mots empreints de tendresse et de reconnaissance. Tu as été mon roc, mon inspiration et ma source de bonheur. Ta présence à mes côtés a donné un sens profond à chaque étape de notre vie ensemble. Je te suis infiniment reconnaissant pour ton amour inconditionnel, ton soutien indéfectible et ta confiance en moi.

À mes chers Assil et Amir,

Mes chers enfants, vous êtes la lumière de ma vie, ma plus grande fierté. Chaque jour passé à vos côtés est un cadeau précieux. Votre innocence, votre joie et votre amour inconditionnel illuminent mon existence. Je souhaite que vous grandissiez en étant conscients de mon amour infini pour vous et en sachant que je serai toujours là pour vous soutenir dans tous vos rêves et projets.

À vous tous, ma famille bien-aimée, vous êtes mes piliers et ma plus grande source de bonheur. Vous m'avez encouragé, inspiré et soutenu dans chaque étape de ma vie, et je ne pourrais pas demander de meilleures personnes à mes côtés. Votre amour inconditionnel et votre présence sont des trésors inestimables.

Avec tout mon amour,

Kamel

Remerciements

Tous d’abord, je souhaite exprimer mon profonde gratitude à ma promotrice Madame Djeddar Afrah pour avoir dirigé ce mémoire. J'ai eu le plaisir de travailler sous votre direction. Je vous remercie pour votre gentillesse et spontanéité avec lesquelles vous avez dirigé ce travail, ainsi que pour votre disponibilité et vos conseils que grâce à eux j'ai pu améliorer mon travail. J'espère que votre confiance que vous vous m'accordez et que ce mémoire est à la hauteur de vos espérances.

Un grand merci à ma mère et mon père, pour leur amour, leurs conseils ainsi que leur soutien inconditionnel, à la fois moral et économique, qui m’a permis de réaliser les études que je voulais et par conséquent ce mémoire.

Je tiens à exprimer ma profonde reconnaissance envers ma femme pour son soutien inconditionnel tout au long de l'élaboration de ce mémoire. Sa présence aimante, son encouragement constant et sa patience infinie ont été d'une valeur inestimable pour moi. Sans elle, ce travail n'aurait pas été possible. Je suis reconnaissant de l'amour, de la compréhension et du soutien qu'elle m'a apportés à chaque étape de ce parcours académique.

Enfin, je remercie mon cousin Amrouche DjamelEddine qui a toujours été là pour m'aider à continuer mes études. Son soutien inconditionnel et ses encouragements ont été d’une grande aide.

À tous ces intervenants, je présente mes remerciements, mon respect et ma gratitude.

Résumé

Le présent mémoire propose une étude approfondie sur la création d'un système de détection d'intrusion en exploitant les technologies de vidéosurveillance. Avec l'augmentation croissante des infractions et des menaces à la sécurité, il devient primordial de développer des outils efficaces pour renforcer la surveillance et réduire les risques liés à l'intrusion.

Dans ce contexte, notre recherche se concentre sur la conception et la mise en œuvre d'un système innovant capable d'analyser en temps réel les flux vidéo provenant des caméras de vidéosurveillance et de détecter toute intrusion. Le système repose sur des algorithmes avancés de vision par ordinateur et d'apprentissage automatique, qui pourraient indiquer la présence d'un intrus.

L'étude comprend une revue exhaustive de la littérature scientifique sur les techniques de détection d'intrusion existantes et les avancées récentes dans le domaine de la vidéosurveillance. Nous explorons également les différentes approches de traitement d'image et d'apprentissage automatique utilisées pour analyser et interpréter les flux vidéo en temps réel.

Ensuite, nous décrivons en détail l'architecture du système de détection d'intrusion proposé, en mettant l'accent sur les composants clés tels que la capture et l'acquisition des flux vidéo, le prétraitement des images, l'extraction des caractéristiques, la classification et la décision des notifications.

Pour évaluer les performances de notre système, nous avons réalisé une expérience sur un jeu de données réels, comme on a suivi un scénario pour tester le fonctionnement de notre système. Les résultats obtenus démontrent l'efficacité et la précision de notre système de détection d'intrusion, avec un taux de détection élevé et un faible taux de fausses alertes.

En conclusion, ce mémoire contribue à l'avancement des connaissances dans le domaine de la sécurité et de la vidéosurveillance en proposant un système novateur de détection d'intrusion. Les résultats obtenus offrent des perspectives prometteuses pour l'application pratique de ce système dans des environnements réels, tels que les installations industrielles, les bâtiments publics et les espaces urbains, afin d'améliorer significativement la sécurité et de prévenir les intrusions.

Mots-clés : système de vidéosurveillance, détection de personnes, détection de visages, reconnaissance faciale, détection d’intrus.

Abstract

This thesis proposes an in-depth study on the creation of an intrusion detection system by exploiting video surveillance technologies. With the growing increase in breaches and security threats, it becomes essential to develop effective tools to strengthen surveillance and reduce the risks associated with intrusion.

In this context, our research focuses on the design and implementation of an innovative system capable of analyzing real-time video streams from CCTV cameras and detecting any intrusion. The system relies on advanced computer vision and machine learning algorithms, which could indicate the presence of an intruder.

The study includes an exhaustive review of the scientific literature on existing intrusion detection techniques and recent advances in the field of video surveillance. We also explore the different

image processing and machine learning approaches used to analyze and interpret real-time video streams.

Next, we describe in detail the architecture of the proposed intrusion detection system, focusing on key components such as video stream capture and acquisition, image preprocessing, feature extraction, classification and decision of notifications.

To evaluate the performance of our system, we performed an experiment on a real data set, as we followed a scenario to test the operation of our system. The results obtained demonstrate the efficiency and accuracy of our intrusion detection system, with a high detection rate and a low rate of false alarms.

In conclusion, this thesis contributes to the advancement of knowledge in the field of security and video surveillance by proposing an innovative intrusion detection system. The results obtained offer promising prospects for the practical application of this system in real environments, such as industrial installations, public buildings and urban spaces, in order to significantly improve security and prevent intrusions.

Keywords: video surveillance system, person detection, face detection, facial recognition, intruder detection.

ملخص

تقترح هذه المذكرة دراسة معمقة حول إنشاء نظام كشف التسلل من خلال استغلال تقنيات المراقبة بالفيديو. مع الزيادة السريعة في الانتهاكات والتهديدات الأمنية، يصبح من الضروري تطوير أدوات فعالة لتعزيز المراقبة وتقليل المخاطر المرتبطة بالتطفل.

في هذا السياق، يركز بحثنا على تصميم وتنفيذ نظام مبتكر قادر على تحليل تدفقات الفيديو في الوقت الفعلي من كاميرات الدوائر التلفزيونية المغلقة واكتشاف أي اختراق. يعتمد النظام على رؤية الكمبيوتر المتقدمة وخوارزميات التعلم الآلي، والتي يمكن أن تشير إلى وجود دخيل.

تتضمن الدراسة مراجعة شاملة للأدبيات العلمية حول تقنيات كشف التسلل الحالية والتطورات الحديثة في مجال المراقبة بالفيديو. نستكشف أيضًا مناهج معالجة الصور والتعلم الآلي المختلفة المستخدمة لتحليل وتفسير تدفقات الفيديو في الوقت الفعلي.

بعد ذلك، نصف بالتفصيل بنية نظام الكشف عن التسلل المقترح، مع التركيز على المكونات الرئيسية مثل التقاط دفق الفيديو والحصول عليه، والمعالجة المسبقة للصور، واستخراج الميزات، والتصنيف، وقرار الإخطارات.

لتقييم أداء نظامنا، أجرينا تجربة على مجموعة بيانات حقيقية، كما اتبعنا سيناريو شامل لاختبار تشغيل نظامنا. توضح النتائج التي تم الحصول عليها كفاءة ودقة نظام الكشف عن التسلل، مع معدل اكتشاف مرتفع ومعدل منخفض من الإنذارات الكاذبة.

في الختام، تساهم هذه المذكرة في تقدم المعرفة في مجال الأمن والمراقبة بالفيديو من خلال اقتراح نظام مبتكر للكشف عن التسلل. تقدم النتائج التي تم الحصول عليها آفاقًا واعدة للتطبيق العملي لهذا النظام في بيئات حقيقية، مثل المنشآت الصناعية والمباني العامة والأماكن الحضرية، من أجل تحسين الأمن بشكل كبير ومنع الاقتحام.

الكلمات المفتاحية: نظام المراقبة بالفيديو، كشف الأشخاص، كشف الوجه، التعرف على الوجه، كشف الدخلاء.

Table des matières

Liste des tableaux

Table des figures

Introduction générale

De nos jours, la télésurveillance a pris une place de plus en plus importante dans la société, ceci est principalement dû à l'intérêt qu'elle procure dans la prévention et dans la sécurité [1]. Inventé dans les années 1950, ce système prévu au départ pour un usage militaire a été adopté par les états et les particuliers [2]. La Grande-Bretagne est le premier pays au monde à généraliser ce système où l'on dénombre des centaines de milliers de caméras disposées dans les parcs, les banques, les écoles, etc. [3, 4]. Tout comme ce dernier, la majorité des pays occidentaux a suivi cette tendance [5]. En Algérie l'usage des caméras de surveillance, soit le système de vidéosurveillance, sera généralisé à toutes les wilayas du pays [6].

Les systèmes de vidéosurveillance classiques ont besoin du côté humain comme élément de base, car les séquences vidéo enregistrés par les caméras de surveillance doivent être analysés par les agents affectés à ces systèmes, et en raison de la nature de l'être humain, sujet à la fatigue, au stress et à l'oubli, ils ont des limites en termes de disponibilité et de capacité pour analyser un grand nombre de séquences vidéo qui s'accumulent plus rapidement qu'ils ne peuvent les visionner, de nombreuses erreurs peuvent se produire, et c'est ce qui constitue des vulnérabilités de sécurité dans le système.

Par conséquent, le travail que font ces analystes pour découvrir des intrus ou dans des enquêtes sur des personnes suspectes après un événement important, par exemple, peut être vain ou peut donner des résultats inattendus, surtout si le nombre de clips est énorme, car ils sont obligés de les regarder encore et encore. Nous devons donc implémenter une fonction de détection d'intrus en temps réel dans les systèmes de vidéosurveillance, en particulier avec le développement observé dans le domaine du traitement d'image, de l'intelligence artificielle et de la reconnaissance faciale.

Dans ce mémoire, nous avons concentré notre attention sur ce sujet important, qui est très nécessaire dans le domaine de la sécurité, pour cela nous avons réalisé un projet d'application intégrée au système de vidéosurveillance, son rôle est de détecter les intrus et de lancer une alarme, en utilisant la technologie de la reconnaissance faciale, Où elle fonctionne sur l'approche suivante : Lors de la détection du mouvement d'une personne, l'application enregistre des séquences vidéo résumées (Enregistre que des clips qui contenant des personnes), En parallèle, elle tente d'identifier ces personnes grâce à la technologie de reconnaissance faciale, elle vérifie également les autorisations de lieux et d'heures pour ces personnes, afin de déterminer si la personne qui apparaît dans la vidéo est un intrus ou non, si c'est le cas, l'application envoie une notification.

Notre étude porte, essentiellement, sur trois parties : dans la première, nous effectuons une revue de littérature sur les systèmes de vidéosurveillance, les méthodes de détection de mouvement et enfin les spécificités de la reconnaissance faciale appliquée aux systèmes de vidéosurveillance.

Dans la seconde partie nous exposons le matériel et les méthodes utilisés pour l'étude de la faisabilité et conception d'un système informatique de détection d'intrus en temps réel. Enfin, dans la dernière partie de cette étude, nous présentons les résultats obtenus après les tests de notre solution ainsi qu'une analyse des résultats.

Chapitre 1 : La vidéosurveillance et la détection de personnes

1.1 Introduction

Dans le cadre de ce mémoire, nous abordons le thème de la création d'un système de détection de l'intrus en exploitant les possibilités offertes par la vidéosurveillance. La vidéosurveillance est un domaine en constante évolution, et elle joue un rôle crucial dans la sécurisation des espaces publics et privés. Grâce à l'avancée technologique, les systèmes de vidéosurveillance sont devenus plus sophistiqués, offrant des fonctionnalités avancées pour la détection précoce et la prévention des intrusions.

Ce chapitre vise à explorer les différents aspects des systèmes de vidéosurveillance, en mettant l'accent sur leur utilisation dans la détection des intrus. Nous commencerons par une présentation générale de la vidéosurveillance, en expliquant son importance et ses objectifs en matière de sécurité. Ensuite, nous examinerons les différents composants d'un système de vidéosurveillance, tels que les caméras, les enregistreurs et les moniteurs, ainsi que les technologies associées telles que la vision par ordinateur et l'intelligence artificielle.

Nous aborderons ensuite les techniques de détection d'intrusion les plus couramment utilisées dans les systèmes de vidéosurveillance. Cela inclut la détection de personnes, qui est basé sur la technique de détection des objets, ainsi que les approches de détection de personnes.

En conclusion, ce chapitre fournira une base solide pour comprendre les systèmes de vidéosurveillance et leur utilisation dans la détection d'intrusion. En explorant les différents composants et techniques, nous serons en mesure de formuler une approche solide pour la création d'un système efficace de détection de l'intrus en exploitant les capacités de la vidéosurveillance.

1.2 La vidéosurveillance

1.2.1 Définition

La vidéo surveillance est un système de caméras permettant de surveiller un espace privé ou public. Des images sont enregistrées avec ce système et sont par la suite visionnées et sauvegardées. Les systèmes de la vidéo surveillance sont composés de différents types de matériel en fonction des besoins de son utilisateur (les caméras de surveillance, l'écran de la vidéo surveillance, l'alimentation des caméras de la vidéo surveillance, les enregistreurs de la vidéo surveillance, les câbles de la vidéo surveillance ou les liaisons sans fil …etc.). [7]

1.2.2 Les systèmes de vidéosurveillance

1.2.2.1 Système de vidéosurveillance analogique

A leur début les systèmes de vidéosurveillance étaient entièrement analogiques c.à.d. que la transmission se faisait comme celle des signaux de téléphoniques.

1.2.2.2 La vidéosurveillance sur IP

La vidéo sur IP, souvent appelée IP-Surveillance, est un système permettant à ses utilisateurs de visualiser et d'enregistrer des images vidéo via un réseau IP (LAN/WAN/Internet).

À la différence des systèmes analogiques, la vidéo sur IP utilise le réseau informatique plutôt qu'un système de câblage point-à-point pour transmettre les informations. Le terme vidéo sur IP englobe à la fois les sources vidéo et audio véhiculées par le système. Dans une application de vidéo sur IP, les flux d'images vidéo numériques peuvent être transférés n'importe où dans le monde via un réseau IP sécurisé, câblé ou sans fil, permettant une visualisation et un enregistrement vidéo en tout point du réseau.

La vidéo sur IP permet aux utilisateurs d'obtenir à tout instant et en tout lieu des informations sur une opération en cours, et de la suivre en temps réel. Cette caractéristique en fait une technologie idéale pour assurer le contrôle des installations, des personnes et des locaux, sur place ou à distance comme le contrôle de la circulation, le contrôle des lignes de production ou le contrôle des points de vente.

Une caméra réseau peut être définie comme l'association d'une caméra et d'un ordinateur. Elle capte et transmet des images en direct sur un réseau IP, ce qui permet aux utilisateurs autorisés de suivre en local ou à distance, d'enregistrer et de gérer la vidéo à l'aide d'une infrastructure réseau IP standard.

Outre ses fonctions vidéo, la caméra réseau possède bien d'autres fonctions permettant notamment la transmission d'autres types d'informations via la même connexion réseau : entrées et sorties numériques, audio, ports série pour des données série ou mécanismes de contrôle des mouvements en panoramique/inclinaison/zoom.

Ces dernières années, les caméras réseau ont rattrapé la technologie analogique et répondent aujourd'hui aux mêmes exigences et spécifications. Les caméras réseau ont même dépassé les caméras analogiques en termes de performances, grâce à l'intégration d'un ensemble de fonctions avancées.

1.2.2.3 Les systèmes analogique/IP (hybride)

Les systèmes mentionnés ici sont des systèmes réunissant des systèmes analogiques et des réseaux IP. Ceci permet par exemple d'étendre un système analogique afin de le rendre plus efficace, de l'ouvrir sur l'extérieur. Ils sont généralement caractérisés par la présence d'un serveur vidéo. Un serveur vidéo permet de migrer vers un système de vidéo sur IP en conservant les installations analogiques existantes et en leur octroyant de nouvelles fonctionnalités. Il permet par ailleurs d'éliminer certains équipements spécifiques (câbles coaxiaux, moniteurs ou enregistreurs numériques), ceux-ci devenant en effet superflus puisque les enregistrements vidéo peuvent se faire à l'aide de serveurs informatiques classiques. Un serveur vidéo possède en général de un à quatre ports analogiques pour la connexion de caméras analogiques, et un port Ethernet pour la connexion au réseau. Tout comme les caméras réseau, un serveur vidéo possède un serveur web intégré, une puce de compression et un système d'exploitation permettant la conversion des flux entrants en images vidéo numériques, ainsi que leur transmission et leur enregistrement sur le réseau informatique où elles pourront être visualisées et consultées plus facilement. [8]

1.3 Détection de personnes

La détection d'intrus par le biais de systèmes de télésurveillance nécessite de la technologie de détection de personnes, qui est l'un des domaines les plus importants de la vision par ordinateur impliquant la détection d'une ou plusieurs personnes dans une image numérique. Ceci est un cas particulier de détection d'objet.

Des études ont commencé à partir de la fin des années 1990 [9, 10], prouvant qu'il s'agit d'un sujet très difficile, compte tenu de la grande variété d'apparences des personnes, des conditions difficiles et en environnement non contraint. S'appuyant sur les avancées systématiques qui ont été faites dans la détection des visages, la détection des personnes a inspiré des méthodes spécifiques, telles que les histogrammes de gradient orienté, qui sont particulièrement efficaces. Le moyen le plus efficace de construire des modèles statistiques est l'apprentissage supervisé, à partir des caractéristiques de forme ou d'apparence, calculées sur de nombreux exemples d'images de personnes.

Depuis les années 2000, le domaine a bénéficié des avancées de la détection de visage et notamment de la méthode de Viola et Jones, qui a été étendue à la détection de personnes utilisant le mouvement. Là où le domaine a connu un véritable essor [11]. En 2005, des chercheurs de l'Institut national de recherche en informatique et en automatique (INRIA) ont proposé les histogrammes de gradient orienté (HOG) [12] dont les excellents résultats en ont fait une méthode standard. Puis plusieurs technologies ont vu le jour au fil des années pour permettre la détection de personnes humaine grâce à des séquences vidéo.

1.3.1 Les approches de détection de personnes

En raison de son importance et de ses nombreux défis, la détection de personnes est un domaine de recherche très actif. Dans cette section, différentes approches de détection de personnes sont présentées. Ces méthodes utilisent principalement des approches basées sur l'apprentissage d'un classificateur pour la création d'un modèle de classe prédéfinie, utilisé par la suite dans la phase de détection. Les méthodes de détection de personnes utilisent dans la majorité des cas des caméras perspectives. Dans les méthodes basées sur la classification, la détection est effectuée en deux étapes distinctes : L'apprentissage et la détection comme on peut le voir dans la (figure 1). L'étape d'apprentissage se concentre sur l'extraction de caractéristiques discriminantes de l'image. Plusieurs méthodes d'apprentissage peuvent être utilisées pour créer un modèle capable de discriminer la classe "personne". Le modèle obtenu est ensuite utilisé pour la prise de décision lors de la phase de détection. [13]

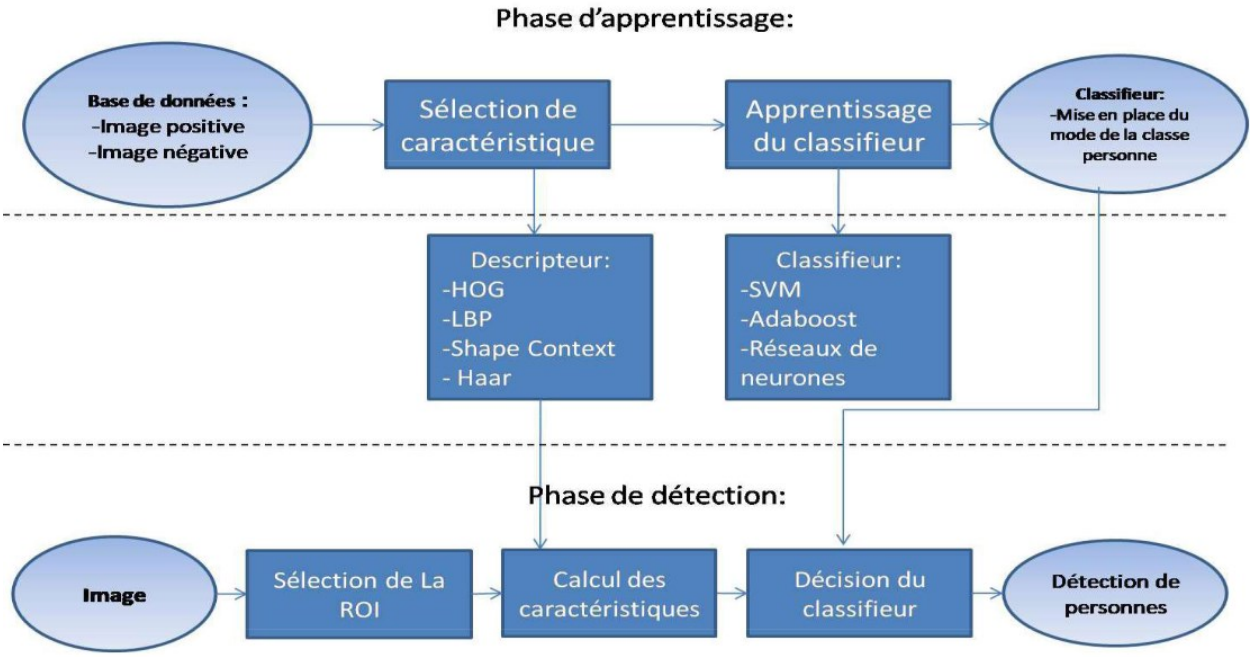


Figure 1 - Fonctionnement des méthodes basées descripteurs-classifieurs [13]

1.3.1.1 Les descripteurs

Beaucoup de travaux ont porté sur la détection de personnes. Ils ont comme phase préalable, l'extraction de caractéristiques pertinentes afin de faire la distinction entre différentes classes. Or une image pour un ordinateur est représentée comme une matrice de pixels. L'utilisation de chaque pixel indépendamment, ne permet pas l'extraction d'informations pertinentes sur le contenu de l'image. De nombreuses approches d'extraction de caractéristiques visuelles utilisent des groupes de pixels. Ces caractéristiques peuvent être calculées à partir d'informations de bas niveau tels que le contour, la texture ou le mouvement quand il y a un mouvement dans la scène. Dans cette section, plusieurs caractéristiques sont revues en fonction des aspects de la forme humaine qu'elles décrivent : forme, texture et mouvement. Ces caractéristiques permettent d'extraire des informations significatives utilisées pour la détection de personnes.

- **Edge Orientation Histograms (EOH)**

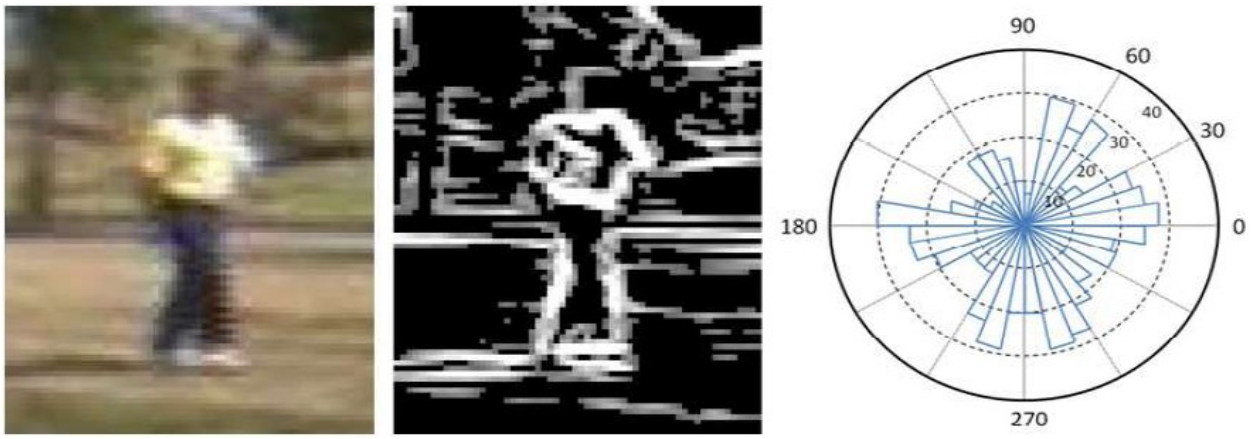


Figure 2 - Exemple des caractéristiques EOH [14]

Les informations de silhouette et de contour sont des caractéristiques importantes pour discriminer une personne dans les images. Pour coder ces informations, les histogrammes orientés de bord (EOH), ont été proposés initialement pour la détection de visages par Levi et Weiss [15]. Ces fonctionnalités permettent de conserver une invariance face aux changements globaux (sur toute l'image) de luminosité, mais aussi de décrire des propriétés géométriques difficiles à capturer avec d'autres descripteurs. Plus tard, les EOH ont été utilisés pour la détection des personnes. Une combinaison de caractéristiques Haar-like et d'histogrammes orientés de bord est utilisée comme caractéristique discriminante lors de la classification [14].

- **Les ondelettes de Haar**

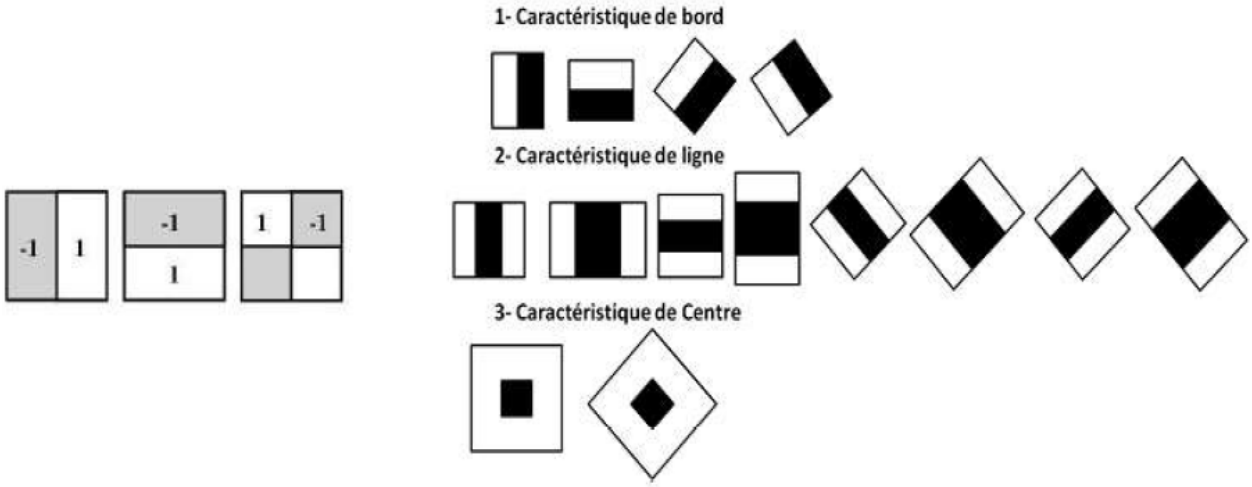


Figure 3 - Les ondelettes de Haar [16]

Les ondelettes de Haar ont été introduites pour la première fois dans le contexte de la détection d'objets à la fin des années 90 par Papageorgiou [17]. Viola et Jones [16] ont adapté l'idée d'utiliser des ondelettes Haar et ont développé les caractéristiques dites Haar. Ils ont introduit la notion d'image intégrale afin de calculer ces caractéristiques de manière rapide. Les caractéristiques de type Haar codent les relations entre les intensités moyennes des régions voisines selon des orientations différentes capturant des bords ou des changements de texture. Cela les rend capables de détecter les similitudes structurelles entre différentes instances d'une classe. Ces caractéristiques captent le changement d'intensité locale selon les directions horizontales, verticales et diagonales. Lorsque ce détecteur est appliqué aux images, la valeur d'une caractéristique de deux rectangles est la différence entre la somme des pixels situés dans la zone non ombrée avec la somme des pixels situés dans la zone ombrée. Une caractéristique de quatre rectangles calcule la différence entre les paires diagonales de rectangles. Leinhart et al ont introduit un ensemble de fonctionnalités Haar étendues en ajoutant des fonctions rectangulaires orientées vers le haut, permettant aux prototypes d'être mis à l'échelle indépendamment dans les axes vertical et horizontal [18].

- **Les motifs binaires locaux**

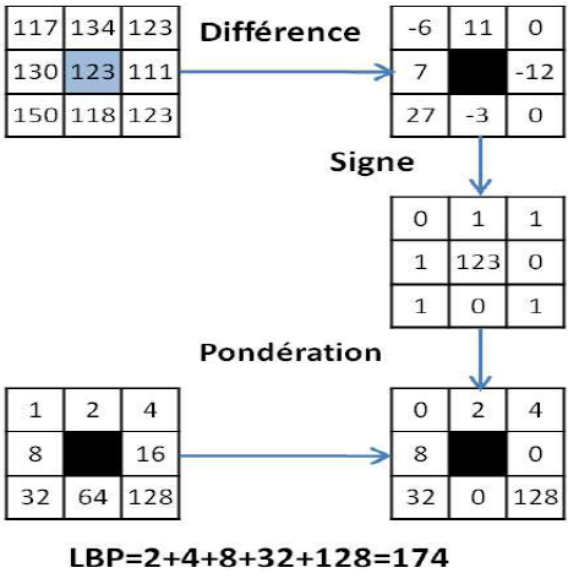


Figure 4 - Fonctionnement du descripteur LBP [19]

Les motifs binaires locaux (*Local Binary Pattern* LBP) sont une caractéristique de codage de texture. C'est un cas particulier du modèle de texture [19, 20]. La version originale de la caractéristique des motifs binaires locaux pour chaque pixel est basée sur un bloc de 3x3 pixels d'une image. Les pixels de ce bloc sont définis par la valeur du pixel central, multipliée par des puissances de deux, puis additionnées pour obtenir une étiquette pour le pixel central. Comme le quartier se compose de 8 pixels, l'étiquette d'une valeur est comprise entre 0 et 256 en fonction des valeurs de gris du centre et des pixels dans le voisinage après la pondération. Une fonctionnalité LBP plus générique est proposée dans [21]. Elle permet une meilleure extraction d'informations à partir du voisinage circulaire autour du pixel central, selon deux paramètres qui sont le rayon du cercle de voisinage "R" et le nombre de points de quartier considérés "P".

- **Contexte de forme**

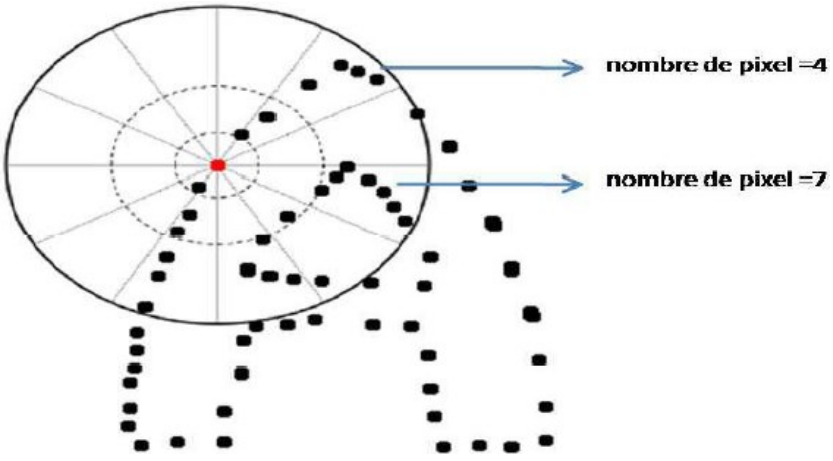


Figure 5 - Représentations de la distribution des pixels du contour dans chaque bloc [22]

Les contextes de forme ont été introduits par [22] pour des tâches de détection d'objets ou de reconnaissance de caractères. L'approche consiste à choisir n points sur les contours de forme

de l'objet à classer. Le contour est extrait à l'aide d'un détecteur de contour, ensuite, chaque point du bord est décrit selon un diagramme de block "log-polaire", l'algorithme a été testé sur la base de données MNIST, et donne de bons résultats. Ces descripteurs sont très bien adaptés à des fins de correspondance et ont également été utilisés pour la détection de personnes dans [23] ou pour la reconnaissance de caractères [24, 25].

▪ Histogramme Orienté du Gradient

Un autre descripteur extrêmement répondu basé sur la silhouette, est l'Histogramme Orienté du Gradient (HOG), proposé par Dalal et Triggs dans [26] pour la détection des personnes. L'extraction des caractéristiques est plus complexe que dans les Histogrammes d'orientation du bord, améliorant les performances discriminatoires du descripteur tout en assurant un certain degré d'invariance. Le calcul du descripteur HOG se fait en cinq étapes :

1. Une normalisation globale de l'image, en utilisant une compression gamma, est effectuée pour réduire l'influence que peut avoir une variation d'éclairage

2. Calcul du gradient de l'image grâce à différents opérateurs (Roberts, Prewitt ou Sobel).

3. La fenêtre d'image est dans un premier temps divisée en petites régions spatiales, appelées « cellules » qui regroupent plusieurs pixels. On génère un histogramme d'orientation du contour en accumulant toutes les orientations de chaque pixel formant la cellule. Chaque pixel du contour contribue avec une valeur proportionnelle à la valeur de son orientation.

4. Une étape de normalisation s'effectue en accumulant une mesure de l'intensité de l'histogramme local sur des groupes de cellules appelées « blocs ». Chaque cellule est normalisée par rapport au bloc auquel elle appartient. Les blocs se chevauchent et donc une même cellule peut contribuer à la formation de plusieurs blocs.

5. Le descripteur HOG de la fenêtre de détection est obtenu en concaténant tous les descripteurs HOG de tous les blocs. Ce vecteur permet de caractériser la forme de l'objet.

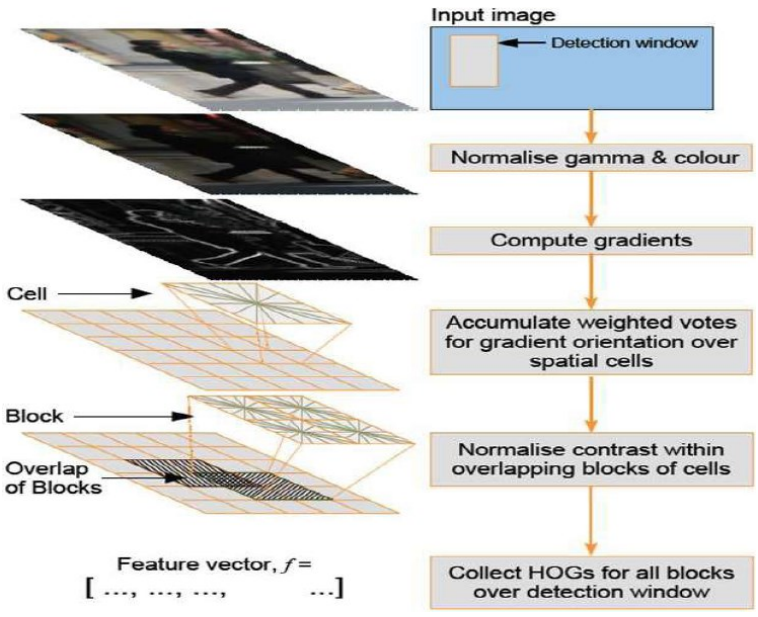


Figure 6 - Histogramme orienté du gradient pour la détection de personnes [26]

1.3.1.2 Les classifieurs

La classification est la dernière étape de la détection de personnes. Au cours de la classification, une région candidate est évaluée et une décision est prise si cette région contient une personne ou non. Dans l'état de l'art la majorité des algorithmes de détection de personnes utilisent des approches basées sur des classificateurs. La plupart de ces approches utilisent des variantes de l'apprentissage par Boosting et SVMs. Mais d'autres techniques d'appariement de silhouettes peuvent être utilisées comme celles basées sur la distance de Chanfrein. Ces deux méthodes de classification sont présentées dans les paragraphes suivants.

- **Machine à vecteurs de support SVM**

La fonction du classifieur est de donner une décision sur l'appartenance du candidat à la classe recherchée. Elle repose sur une base de données d'apprentissage. En prenant en entrée les caractéristiques des exemples contenant un individu de la classe (ici une personne) et des exemples ne contenant pas d'individus de la classe, le classifieur doit déterminer de qui les caractéristiques de l'image candidate sont les plus proches. Dans la plupart des cas, cette étape est la dernière du processus puisqu'une fois reconnues par le classifieur, il suffit d'afficher les fenêtres de détection. Cette méthode élaborée par Vapnik et al [27], vise à déterminer un hyperplan séparateur entre les espaces des deux classes. L'idée est de maximiser la marge, c'est-à-dire la distance entre les frontières de séparation des échantillons les plus proches. Pour cela, l'algorithme transforme l'espace de représentation des données d'entrée en un espace de plus grande dimension, dans lequel il est probable qu'il existe une droite séparatrice linéaire. Du fait de sa très grande efficacité, cette méthode est très couramment utilisée notamment dans le domaine de la détection de personnes [28, 29].

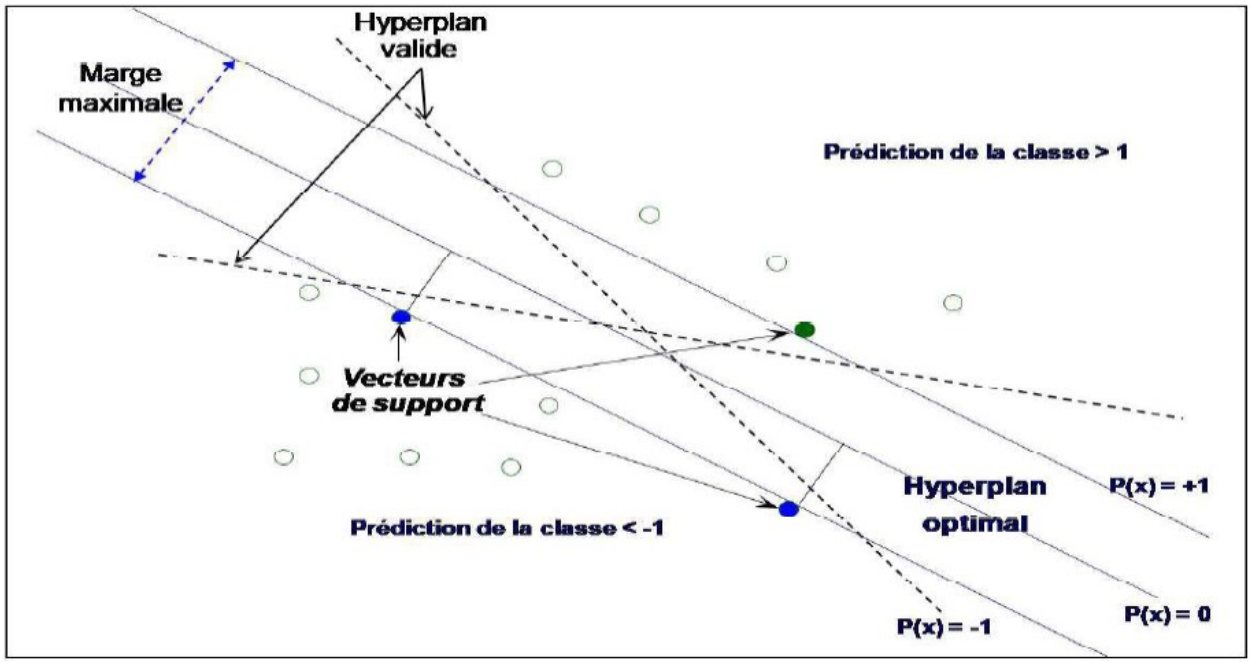


Figure 7 - Machine à vecteurs de support SVM [27]

- **AdaBoost**

AdaBoost (ou *adaptive boosting*) est un algorithme d'apprentissage, Il peut être associé avec un autre algorithme d'apprentissage pour en améliorer les performances [30, 31]. Le principe consiste à combiner des classifieurs, par itérations successives. La connaissance d'un classifieur faible est ajoutée au classifieur final. Le classifieur ajouté est pondéré par la qualité de sa classification : plus il permet de bien classer, plus il sera important. Les exemples mal classés sont boostés pour qu'ils aient davantage d'importance vis à vis de l'apprenant faible au prochain tour, afin de pallier à ce manque.

- **Réseaux de neurones**

Un réseau de neurones est un algorithme d'apprentissage basé sur des concepts inspirés du fonctionnement d'un cerveau humain. Un réseau de neurones peut être considéré comme un graphe dirigé reliant un certain nombre de neurones par des liens pondérés, le fonctionnement des neurones est simulé par des fonctions d'activation. Les réseaux de neurones peuvent être classés en trois catégories basées sur l'architecture adoptée [Jain et al., 2000] : les réseaux feed forward (FFs) ou Perceptrons multicouches (MLP), les réseaux de fonctions à base radiale (RBFs) et les cartes auto-organisatrices (SOMs). Dans les réseaux MLP, une combinaison linéaire des entrées est calculée en utilisant une fonction qui renvoie le produit scalaire entre les entrées et les poids synaptiques correspondants. Par contre les RBFs emploient des fonctions de combinaison qui renvoient les distances euclidiennes entre les entrées et les centres de la couche cachée. Les cartes auto-organisatrices (SOMs) possèdent une structure différente où des nœuds (neurones) sont régulièrement placés dans une grille hexagonale ou rectangulaire. Les SOMs sont basés sur le principe de transformation des données d'entrée de grandes dimensions à des données de dimensions inférieures [32]. L'apprentissage d'un réseau de neurones implique l'ajustement de poids basé sur un processus itératif qui optimise une fonction d'erreur particulière. Les méthodes de la descente de gradient par exemple sont des méthodes d'apprentissage supervisées qui ont été beaucoup utilisées grâce aux résultats raisonnables qu'ils peuvent générer en un temps minimal. Cependant, elles peuvent converger vers des solutions locales dans la plupart des cas et plus particulièrement lorsque la tâche est difficile, telle que la reconnaissance de caractères manuscrits. Les réseaux de neurones, avec leur capacité de généralisation, peuvent reconnaître et détecter des formes imprévues qui sont difficiles à détecter par d'autres méthodes de classification. Un réseau de neurones entraîné peut être considéré comme un "expert" capable de fournir des réponses dans de nouvelles circonstances (une situation inédite).

1.4 Conclusion

Dans ce chapitre, nous avons passé en revue un aperçu des concepts de base de la vidéosurveillance, qui sont essentiels pour comprendre les idées traitées dans ce travail. Nous avons fourni quelques concepts importants des systèmes de la vidéosurveillance en général et les techniques de la détection de personnes, qui est la première étape pour la détection de l'intrus.

Dans le chapitre suivant, nous présenterons la notion fondamentale de la reconnaissance faciale par les techniques d'apprentissage profond et les réseaux de neurones à convolution.

Chapitre 2 : La reconnaissance faciale par les techniques d'apprentissage profond et les réseaux de neurones à convolution

2.1 Introduction

Ce chapitre consacré à la reconnaissance faciale par l'apprentissage profond permettra d'explorer les avancées technologiques récentes et les meilleures pratiques dans ce domaine. Il fournira une base solide pour comprendre les principes fondamentaux, les techniques et les applications de la reconnaissance faciale par l'apprentissage profond, tout en mettant en évidence les défis et les opportunités futures.

2.2 L'apprentissage profond

Le *deep learning* ou apprentissage profond est un sous-domaine du *machine learning* ou apprentissage machine, sous-domaine de l'intelligence artificielle, (Voir Figure 8 - L'intelligence artificielle et ses sous-domaines).

Par intelligence artificielle on comprend l'acte de faire reproduire par des machines des tâches qui sont jugées comme complexes par les humains, typiquement : faire raisonner une machine, apprendre à une machine la planification de tâches, ou encore apprendre à une machine à représenter des connaissances d'une manière structurée.

Le *deep learning* ou apprentissage profond, regroupe actuellement les méthodes les plus efficaces et les plus performantes appliquées dans la communauté de l'apprentissage automatique. [33]

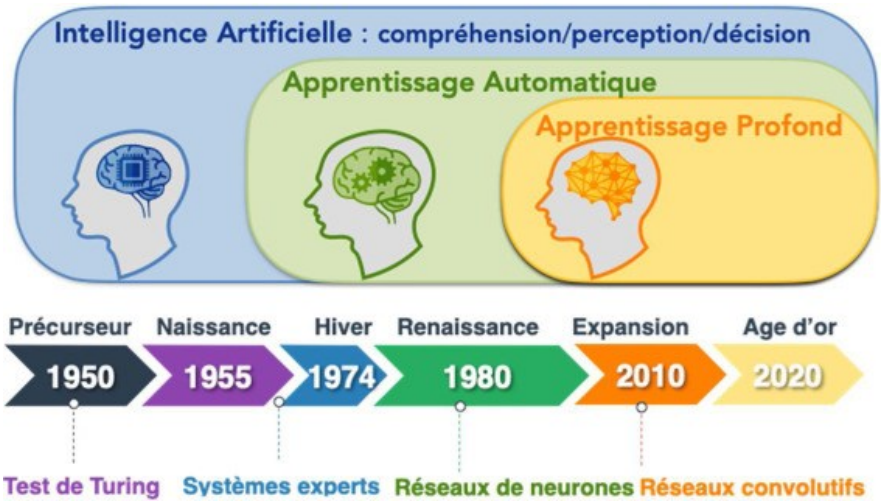


Figure 8 - L'intelligence artificielle et ses sous-domaines [33]

2.2.1 Définitions

2.2.1.1 Le réseau neuronal

Se situe au croisement de l'informatique et de la biologie. Il est calqué sur le paradigme du cerveau humain dont il démultiplie la puissance, sans lui ressemble tout à fait puisqu'il est dépourvu d'émotion.

$$y = f\left(\sum_{i=0}^{n} w_i x_i + b\right)$$

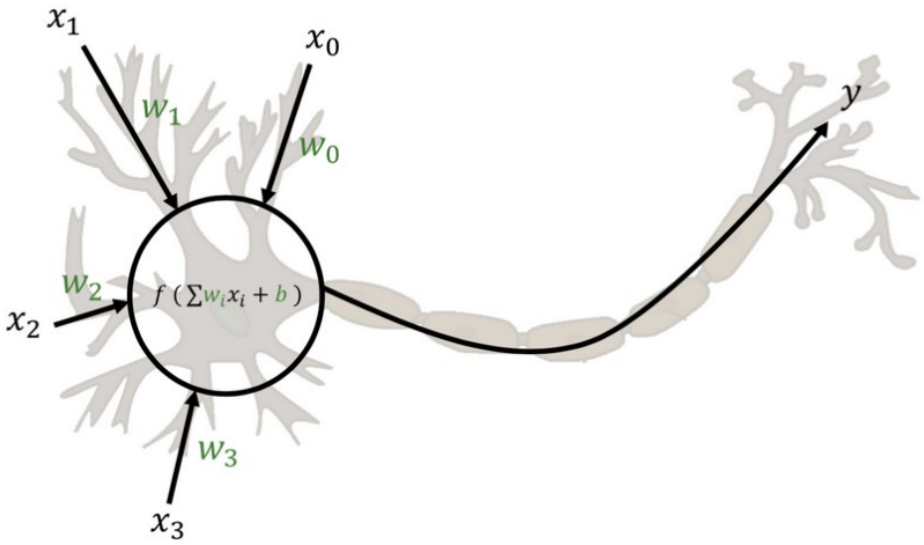


Figure 9 - Schéma d'un neurone informatique superposé à un schéma de neurone biologique [33]

2.2.1.2 L'intelligence artificielle

Est un champ de recherche qui regroupe l'ensemble des techniques et méthodes qui tendent à comprendre et reproduire le fonctionnement d'un cerveau humain.

2.2.1.3 La machine learning

Est un ensemble de techniques donnant la capacité aux machines d'apprendre automatiquement un ensemble de règles à partir de données. Contrairement à la programmation qui consiste en l'exécution de règles prédéterminées.

- **Techniques d'apprentissage automatique**

a. Régression

Elle est basée sur les principes de base de la physique qui aident à prédire l'avenir à partir des données actuelles. Elle vous aide également à trouver la corrélation entre deux variables pour définir la relation de cause à effet. Vous pouvez tracer un graphique basé sur ces variables et faire des prévisions en continu, en vous basant sur la variable prédictive.

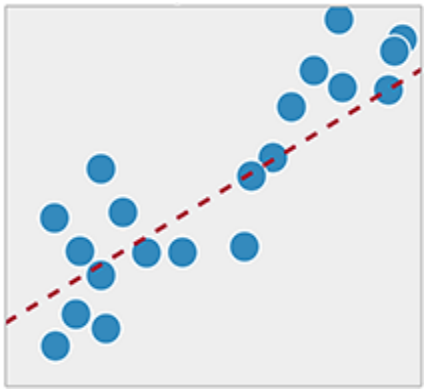


Figure 10 - Illustration de la régression linéaire [33]

Cependant, il existe différentes formes de régression, allant de la régression linéaire à la régression complexe, en passant par le calcul et la représentation de données polynomiales.

Les exemples courants de régression linéaire sont les suivants :

- Prévisions météorologiques
- Prévoir les tendances du marché
- Identifier les risques potentiels

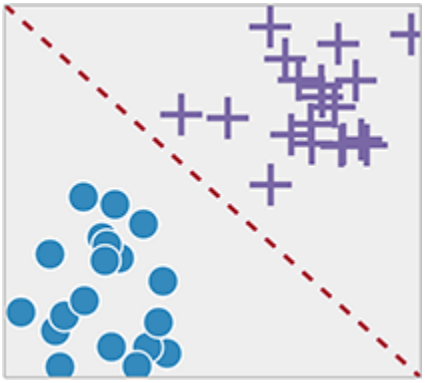


Figure 11 - Illustration de la classification linéaire [33]

b. Classification

La classification est un processus de catégorisation d'un ensemble de données en classes. Elle peut être effectuée sur des données structurées ou non structurées. Le processus commence par la prévision de la classe des points de données. Les classes sont souvent appelées cible, étiquette ou catégories (en anglais : Label).

c. Regroupement

Il s'agit d'une technique d'apprentissage automatique non supervisée, dans laquelle les traits similaires sont utilisés pour faire une prédiction, au lieu des données passées. L'algorithme utilise des repères visuels pour concevoir une solution. K-Means est la méthode la plus populaire de regroupement des données d'entrée, qui permet de fixer la valeur de K et de classer les données en fonction de cette valeur.

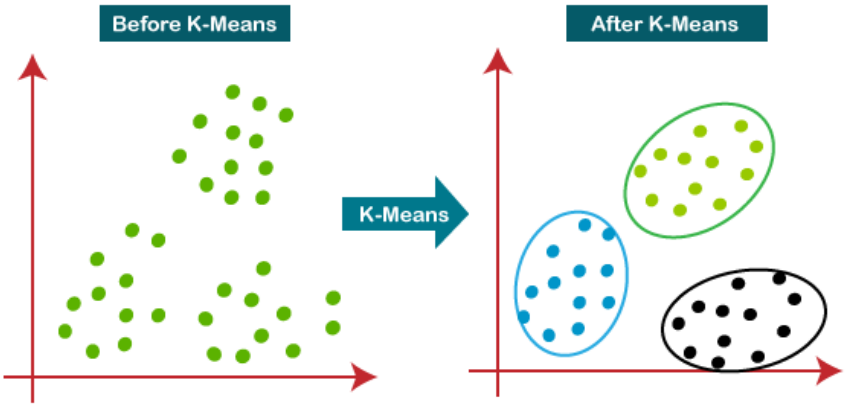


Figure 12 - Algorithme de clustering K-Means [33]

d. Réduction de la dimensionnalité

Il s'agit du processus de réduction des variables aléatoires tout en catégorisant les données. Plus le nombre de variables est élevé, plus les résultats seront complexes, ce qui rendra difficile leur consolidation.

La sélection et l'extraction des caractéristiques sont au cœur de la réduction de la dimensionnalité dans l'apprentissage machine. Elles permettent d'éliminer les variables non pertinentes.

L'exemple le plus courant de réduction dimensionnelle est le processus de classification des courriers électroniques utilisé pour trier les courriers indésirables. En général, il utilise un grand nombre de variables telles que les titres, le contenu et le modèle du courriel, entre autres. Mais il est possible que l'algorithme recoupe certains facteurs qui peuvent affecter le résultat.

Ainsi, pour faire des suppositions précises, le logiciel intègre la réduction de dimensionnalité afin d'atténuer les chances de répétition et de vous fournir des résultats précis.

e. Méthode d'ensemble

Il s'agit d'une technique permettant d'empiler des données en utilisant des variables de prédiction provenant de divers modèles. Elle combine donc divers modèles prédictifs pour former une sortie prédictive très précise et optimisée. La méthode est utilisée pour prendre des décisions en tenant compte de divers facteurs.

L'algorithme *Random* Forest est un exemple typique de méthodes d'ensemble qui combinent divers arbres de décision basés sur des ensembles de données multiples. Grâce à cela, la sortie prédictive est de bien meilleure qualité que les estimations d'un seul arbre de décision.

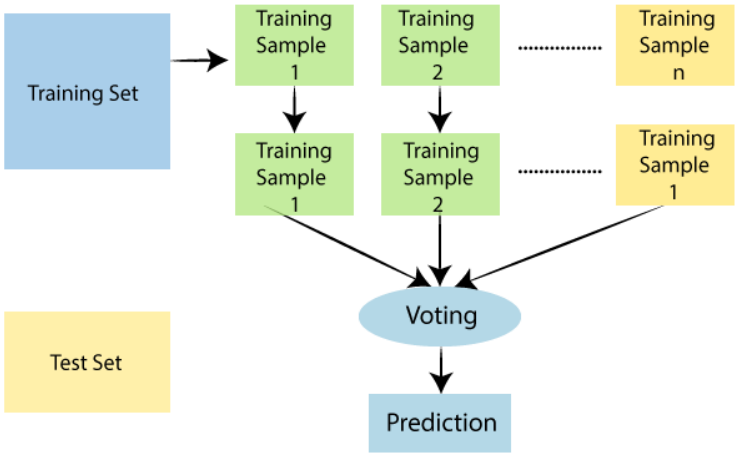


Figure 13 - L'algorithme de forêt aléatoire [34]

f. Réseaux neuronaux et apprentissage approfondi

Contrairement aux modèles linéaires, le réseau de neurones est basé sur un modèle de données complexe et divisionnaire. Il comprend plusieurs couches d'un paramètre pour vous fournir une sortie unique et précise. Cependant, le modèle est toujours basé sur la régression linéaire mais utilise de multiples couches cachées ; c'est pourquoi on l'appelle un réseau neuronal.

Le terme d'apprentissage approfondi indique les connaissances complexes requises pour résumer ces multiples paramètres. La technique est encore en phase de développement, ce qui rend difficile de se tenir au courant des dernières avancées.

Les scientifiques spécialisés dans l'apprentissage approfondi ont besoin d'unités de traitement graphique de haut niveau pour traiter de grandes quantités de données. C'est pourquoi ces techniques connaissent un grand succès dans les genres liés aux images, au son et à la vidéo.

2.2.1.4 Le deep learning ou apprentissage profond

Est une technique de machine learning reposant sur le modèle des réseaux de neurones, des dizaines voire des centaines de couches de neurones sont empilées pour apporter une plus grande complexité à l'établissement des règles.

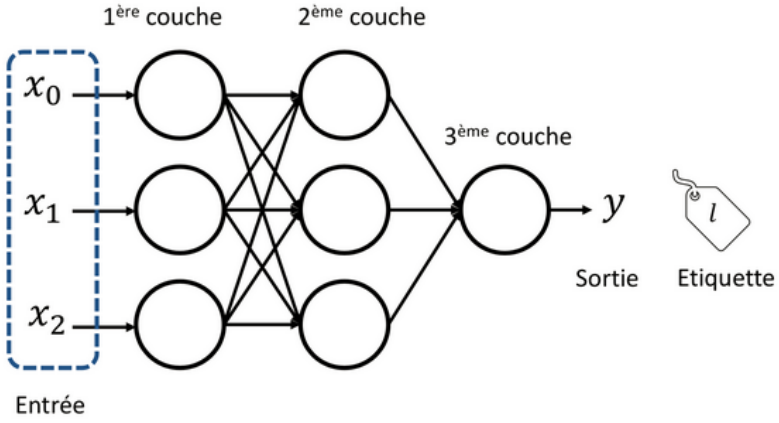


Figure 14 - Un perceptron multicouches ou MLP composé de trois couches [34]

2.2.2 Types d'approches d'apprentissage profond

Comme l'apprentissage automatique, les approches d'apprentissage en profondeur peuvent être classées comme suit : supervisées, semi-supervisées et non supervisées. En outre, il existe une autre catégorie d'apprentissage appelée apprentissage par renforcement (RL) ou Deep RL (DRL) qui sont souvent abordées dans le cadre d'approches d'apprentissage semi-supervisé ou parfois non supervisé [34].

1) **Apprentissage supervisé :** C'est une technique d'apprentissage qui utilise des données étiquetées. Dans le cas des approches DL supervisées, l'environnement dispose d'un ensemble d'entrées et de sorties correspondantes. Il modifiera alors itérativement les paramètres du réseau pour une meilleure approximation des sorties souhaitées. Il existe différentes approches d'apprentissage supervisé pour le deep learning, notamment les réseaux de neurones profonds (*Deep Neural Networks*- DNN), les réseaux de neurones convolutifs (*Convolutional Neural Networks*- CNN), les réseaux de neurones récurrents (*Recurrent Neural Networks*- RNN), y compris la mémoire à long terme (Long Short Terme Memory-LSTM) et les unités récurrentes fermées (GRU) [34].

2) **Apprentissage semi-supervisé :** C'est un apprentissage basé sur des ensembles de données partiellement étiquetés (également appelés apprentissage par renforcement). Danscertains cas, les DRL et les Réseaux Adversaires Génératifs (GAN) sont utilisés comme techniques d'apprentissage semi-supervisé. De plus, RNN, y compris LSTM et GRU, sont également utilisées pour l'apprentissage semi-supervisé [34].

3) **Apprentissage non supervisé :** C'est un apprentissage qui se fait sans la présence d'étiquettes de données. Dans ce cas, le réseau apprend la représentation interne ou des fonctionnalités importantes pour découvrir des relations ou une structure inconnue dans les données d'entrée. Le regroupement, la réduction de la dimensionnalité et les techniques génératives sont souvent considérés comme des approches d'apprentissage non supervisé. Il y a plusieurs membres de la famille du deep learning qui sont bons pour le *clustering* et la réduction de dimensionnalité non linéaire, y compris : encodeurs automatiques (AE), Machines Boltzmann restreintes (RBM) et le GAN récemment développé. En outre, les RNN, tels que LSTM (*Long-Short Term Memory*

neural network) et RL, sont également utilisés pour l'apprentissage non supervisé dans de nombreux domaines d'application [35].

4) **Apprentissage par transfert :** L'idée de l'apprentissage par transfert a contribué à réduire les demandes de données. L'apprentissage par transfert est le processus consistant à prendre un modèle pré-entrainé (les poids et les paramètres d'un réseau qui a été formé sur une grande base de données) et à affiner le modèle avec notre propre base de données. L'idée est que ce modèle pré-entraîné agira comme un extracteur de fonctionnalités. Nous supprimerons la dernière couche du réseau et la remplacerons par notre propre classificateur. Nous figeons ensuite les poids de toutes les autres couches et formons le réseau normalement (geler les couches signifie ne pas changer les poids lors de la descente /optimisation du gradient) [36].

Le deep Learning se concentre sur cinq réseaux neuronaux de base, y compris :

- Percepteur multicposition.
- Réseau de base radiale.
- Réseaux neuronaux récurrents (RNN).
- Réseaux accusatoires génératifs.
- Réseaux neuronaux convolutionnels (CNN).

2.3 Réseau de neurones à convolution (CNN)

Le nom '*Réseau de neurones à convolution*' indique que le réseau emploi une opération mathématique appelée la convolution. Les réseaux de convolution sont un type spécialisé de réseaux neuronaux qui utilisent la convolution à la place de la multiplication matricielle générale dans au moins une de leurs couches. Les CNN sont l'un des meilleurs algorithmes d'apprentissage pour faire l'opération de convolution qui aide à l'extraction de fonctionnalitésutiles à partir de points de données corrélés localement. La sortie des noyaux convolutifs est ensuite affectée à l'unité de traitement non linéaire (fonction d'activation), qui non seulement aide à apprendre les abstractions, mais intègre également la non-linéarité dans l'espace des fonctionnalités. Cette non-linéarité génère différents modèles d'activations pour différentes réponses et facilite ainsi l'apprentissage des différences sémantiques dans les images [37].

La topologie de CNN est divisée en plusieurs étapes d'apprentissage composées d'une combinaison des couches convolutives, des unités de traitement non linéaires et des couches de sous-échantillonnage [38, 37]. (voir Figure 15 - Réseau de neurones avec de nombreuses couches convolutives)

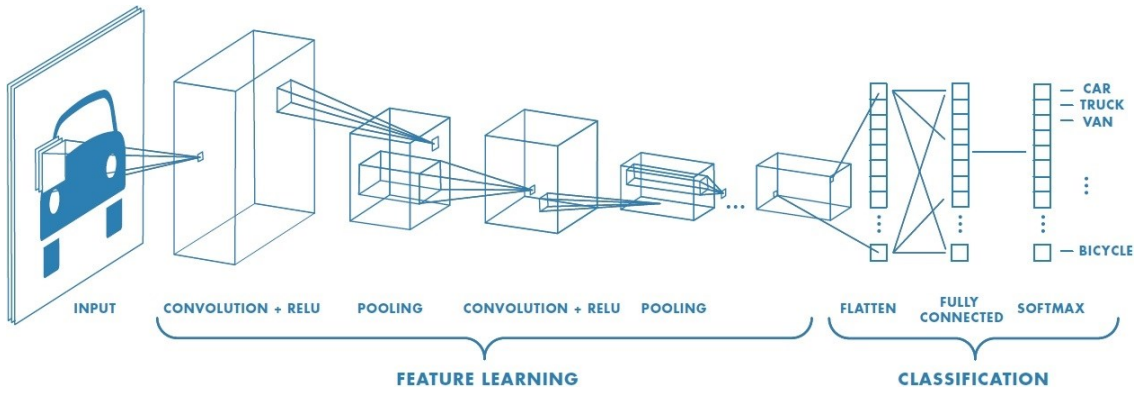


Figure 15 - Réseau de neurones avec de nombreuses couches convolutives [37, 38]

2.3.1 Les principaux blocs de construction utilisés dans les réseaux de neurones convolutifs

2.3.1.1 Couche de convolution

La convolution est la première couche à extraire des entités d'une image d'entrée. La convolution préserve la relation entre les pixels en apprenant les caractéristiques de l'image à l'aide de petits carrés de données d'entrée. Il s'agit d'uneopération mathématique qui prend deux entrées telles qu'une matrice d'image et un filtreou un noyau [39].(voir la Figure 16 - Opération de convolution appliquée à la reconnaissance d'images).

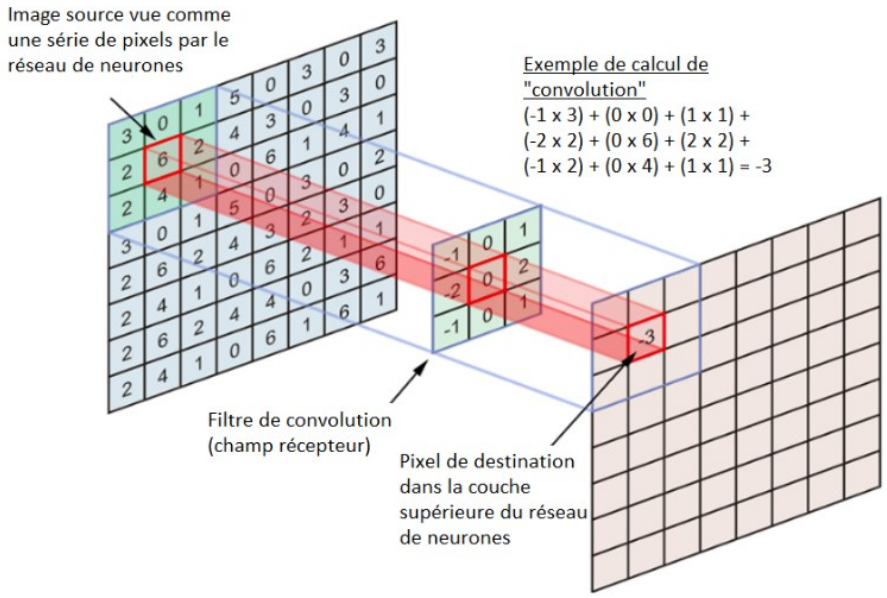


Figure 16 - Opération de convolution appliquée à la reconnaissance d'images [39]

2.3.1.2 Couche de pooling

La couche *pooling* généralement placé entre deux couches de convolution, le *pooling* est un processus de discrétisation à base d'échantillon. L'objectif est à en-bas-de-promotion une représentation de contribution (l'image, la matrice de production de cacher-couche, etc.), en réduisant sa dimensionnalité et en tenant compte des hypothèses à être rendues des caractéristiques contenues dans les sous-régions regroupées. Il existe différents types de pooling : *(1)* **Pooling moyen** (*Average Pooling*) qui prend la moyenne de tous les pixels de la sélection, *(2)* **Pooling maximal** (*Max Pooling en*) qui prend le pixel qui a la valeur maximale entre tous les pixels de la sélection. (voir la Figure 17 - Illustration des techniques de Max pooling & Average pooling)

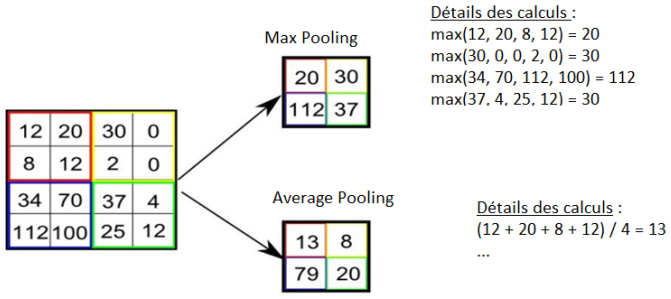


Figure 17 - Illustration des techniques de Max pooling & Average pooling [39]

2.3.1.3 Couche de « entièrement connectée » (*fully-connected*)

La couche entièrement connectée est un traditionnel perceptron multicouche (Multi Layer Perceptron). Le terme « entièrement connecté » implique que chaque neurone dans la couche précédente est connecté à chaque neurone sur la couche suivante. La représente un exemple d'une couche entièrement connectée [40].

La sortie des couches de convolution et de *Pooling* représente les fonctions de haut niveau de l'image d'entrée. Le but de la couche entièrement connectée est de pouvoir utiliser ces fonctions pour classer l'image d'entrée dans différentes classes en fonction de l'ensemble de données d'apprentissage [40].

Comme pour les couches de convolution, les couches « fully connected » sont associées à une fonction d'activation. Il s'agit de la fonction softmax qui permet d'associer à chaque sortie du réseau de neurones convolutifs un score, ensuite transformé en probabilité d'appartenir à un certain label d'image.

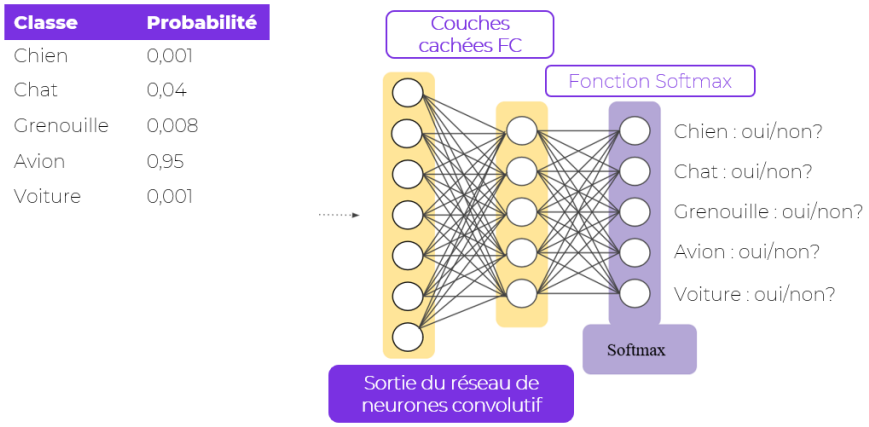


Figure 18 - Couches Fully Connected (FC) & Fonction Softmax [40]

2.3.1.4 Fonction d'activation

La fonction d'activation est une technique pour résoudre un problème majeur lié au processus de rétropropagation de l'erreur dans le réseau de neurones.

En effet, le réseau de neurones convolutif apprend par propagations-avants (apprentissage de l'entrée du réseau de neurones vers la couche de sortie) et propagations-arrières (rétropropagation de l'erreur dans le réseau, c'est-à-dire couche de sortie vers la couche d'entrée).

Au cours de ces passes, une mise à jour des poids de connexions est réalisée.

- **Les fonctions d'activation les plus couramment utilisées**

✓ **Sigmoïd** : fonction populaire et ancienne mais hélas, elle présente des problèmes de saturation. La convergence du modèle de deep learning n'est alors plus garantie. Pour en dire quelques mots, cette fonction prend n'importe quelle valeur et retourne des valeurs comprises entre 0 et 1.

✓ **ReLU** : fonction linéaire rectifiée (ReLU) est l'une des fonctions d'activation les plus populaires et efficaces. Elle est simple d'utilisation et permet un entraînement plus rapide en nous évitant les problèmes de la fonction Sigmoid. Elle présente, néanmoins, un

problème majeur connu sous le nom de “Dying ReLu” où certains neurones peuvent devenir inactifs/mourir.

✓ **Leaky ReLU** : fonction inspirée de la ReLU mais qui permet de surmonter les problèmes de “Dying ReLu”.

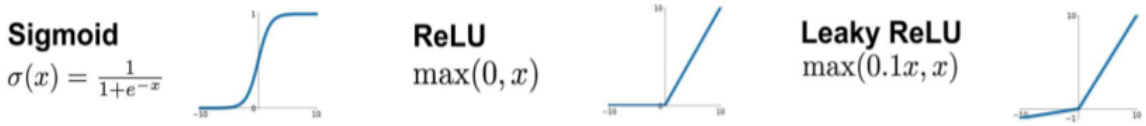


Figure 19 - Formules mathématiques des principales fonctions d’activation [40]

2.4 La reconnaissance faciale

La reconnaissance faciale est une technologie qui vise à identifier et à vérifier l'identité d'une personne à partir de ses caractéristiques faciales. Avec l'avènement des techniques d'apprentissage profond et des réseaux de neurones à convolution (CNN), la reconnaissance faciale a connu des avancées significatives en termes de précision et de performances.

Dans cette section nous aborderons les différentes étapes impliquées dans la reconnaissance faciale utilisant des techniques d'apprentissage profond. Nous discuterons également des défis et des considérations associées à la reconnaissance faciale utilisant des techniques d'apprentissage profond. Cela peut inclure la gestion des variations d'éclairage, des expressions faciales, des occlusions et des données d'entraînement insuffisantes. Nous explorerons les techniques de prétraitement, de régularisation et d'augmentation des données qui peuvent être utilisées pour améliorer la robustesse du système de reconnaissance faciale.

2.4.1 Les étapes de la reconnaissance de visage

La reconnaissance faciale est un système permettant d'identifier et de confirmer les personnes en contrôlant si celles-ci appartiennent à la base de données du système. L'image suit un processus de reconnaissance faciale spécifique contenant plusieurs étapes qui peuvent être illustrées dans le diagramme de la figure 2 ci-dessous :

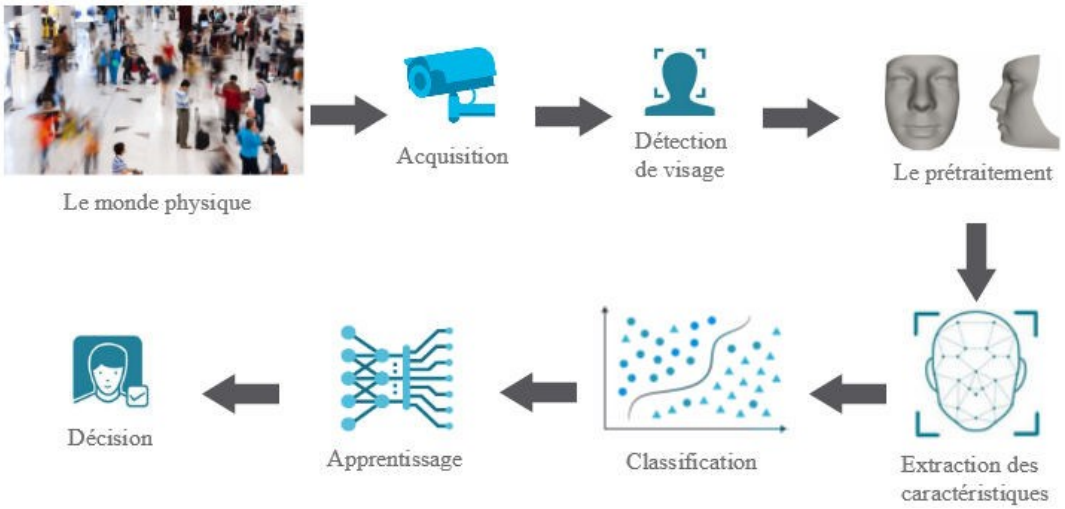


Figure 20 - Processus d’un système de reconnaissance de visage [41]

2.4.1.1 **Le monde physique**

C'est le monde réel en dehors du système avant l'acquisition de l'image. Dans cette étape, on tient compte généralement de trois paramètres essentiels : L'éclairage, la variation de posture et l'échelle. La variation de l'un de ces trois paramètres peut conduire à une distance entre deux images du même individu, supérieure à celle séparant deux images de deux individus différents, et par conséquence une fausse identification [41].

2.4.1.2 **Acquisition**

Le système d'acquisition est généralement équipé d'un capteur qui permet aux utilisateurs d'obtenir une fonction spécifique (par exemple, un microphone pour enregistrer le son et une caméra pour capter une photo, etc.). Un appareil photo nous permet d'avoir une image 2D du visage à partir d'une scène 3D.

2.4.1.3 **Détection de visage**

La détection de visage est une étape très intéressante dans le domaine de reconnaissance de visage. Plusieurs travaux de recherches ont été effectués dans ce domaine. Ils ont donné lieu au développement d'une multitude de techniques allant de la simple détection du visage, à la localisation précise des régions caractéristiques du visage, tels que les yeux, le nez, les sourcils, la bouche, les lèvres, les oreilles, etc. [42] .Un visage est considéré correctement détecté si la taille d'image extraite ne dépasse pas 20% de la taille réelle de la région faciale. Cette étape peut faire la détection de la couleur, de peau, la forme de la tête, il existe plusieurs méthodes détectant les différentes caractéristiques du visage [43].

Les solutions proposées jusqu'à présent ne sont pas suffisamment satisfaites, car elles fonctionnent sous certaines conditions et ne fonctionnent pas dans des acquisitions normales, notamment en présence ou en absence d'aspects structurels du visage, tels que barbe, moustache, lunettes, etc.

Figure 21 - Détection de visage. [42]

- **Les difficultés liées à la détection de visage**

Les difficultés liées à la détection de visage peuvent être attribuées aux facteurs suivants :

- ✓ **Posture :** L'image d'un visage change en raison de la position relative caméra- visage (de face, de profil ou dans une position intermédiaire), et certains attributs faciaux tels que les yeux ou le nez peuvent être partiellement oucomplètement occultés.
- ✓ **Présence ou absence de composantes structurelles :** Les attributs faciaux tels que la barbe, la moustache et ou des lunettes peuvent être présents ou pas et cela avec une forte variabilité. De plus, ces attributs peuvent eux-mêmes revêtir des formes très différentes d'un individu à l'autre : géométrie, couleur, taille, *etc*.
- ✓ **Expression faciale :** L'expression faciale d'une personne affecte directement l'aspect de son visage.

- ✓ **Occultation :** Des visages peuvent être partiellement cachés par d'autres objets. Dans une image avec un groupe de personnes, certains visages peuvent partiellement ou entièrement en cacher d'autres.
- ✓ **Orientation de l'image :** Les images de visage changent directement pour différentes rotations autour de l'axe optique de la caméra.
- ✓ **Conditions de prise de vue :** Des facteurs tels que l'éclairage (distribution, orientation et intensité de la source) et les caractéristiques de la caméra (capteur, optique) affectent l'aspect d'un visage dans l'image. [44]

2.4.1.4 Le prétraitement

Les données délivrées par les capteurs primaires ne sont qu'une représentation initiale de celles-ci d'où la nécessité d'un traitement antérieur. L'image brute peut être affectée par divers facteurs provoquant sa dégradation, pouvant être bruyante, c'est-à-dire contenir de fausses informations dues à des dispositifs optiques ou électroniques. Le rôle de cette étape est d'éliminer les parasites accompagnants l'image, provoqués par la qualité de ces dispositifs. Ceci est nécessaire car l'image ne peut jamais être sans bruit, car le fond et la lumière sont généralement inconnus. Il existe plusieurs types de traitement et d'optimisation de la qualité d'image, tels que la normalisation, les graphiques, le filtrage, la correction gamma ou des méthodes plus complexes telles que le lissage anisotrope [45].

2.4.1.5 Extraction des caractéristiques

L'extraction des caractéristiques est le cœur du système de reconnaissance qui extrait les informations d'image qui seront stockées dans la mémoire pour une utilisation ultérieure dans l'étape de décision. Le choix de cette information utile réside dans la création d'un modèle de visage, qui doit être discriminatoire. Cette analyse est appelée propriétés d'indexation, de représentation, de modélisation ou d'extraction. L'efficacité de cette étape a un impact direct sur la performance du système de reconnaissance faciale [41].

2.4.1.6 Classification

Lorsque les formulaires sont stockés dans la base de données, le système se compose d'échantillons similaires de nombreuses personnes sélectionnées ainsi que d'une liste limitée de candidats. Cette étape consiste à modéliser les paramètres extraits de la ou les faces de chaque individu en fonction de leurs caractéristiques communes. Un modèle est une collection d'informations utiles, uniques et non récurrentes qui identifie une ou plusieurs personnes ayantdes similitudes.

2.4.1.7 Apprentissage

L'apprentissage consiste à retenir les modèles calculés pendant la phase d'analyse des personnes connues. Ce modèle est une représentation intégrée d'images pour faciliter l'identification, mais aussi la quantité de données stockées sous une forme ou une autre. Cette étape correspond aux références interactives réelles qui seront enregistrées dans la base de données.

2.4.1.8 Décision

La décision fait partie du système dans lequel nous décidons si l'individu appartient à tous les visages ou non. Dans cette phase, le système d'identification consiste à trouver le modèle

correspondant au visage pris à partir de ceux stockés dans la base de données, dans ce cas quelle est son identité. Par conséquent, la résolution est l'aboutissement de ce processus, il peut être évalué au taux de reconnaissance (fiabilité), déterminé par le taux de résolution de la décision.

2.4.2 Méthodes de reconnaissance de visages

De nombreuses méthodes de reconnaissance de visages ont été proposées au cours des trente dernières années. La reconnaissance faciale automatique est un challenge tel qu'il a suscité de nombreuses recherches dans des disciplines différentes : psychologie, neurologie, mathématiques, physique et informatique (reconnaissance des formes, réseaux de neurones, vision par ordinateur). C'est la raison pour laquelle la littérature sur la reconnaissance de visages est vaste et diversifiée. Les neurologistes étudient depuis longtemps le mécanisme par lequel le cerveau reconnaît les visages, et beaucoup croient que le cerveau perçoit les visages d'une manière spécifique et très différente des autres objets visuels. Par exemple, les études ont constaté qu'une rotation d'image faciale de 180° dégrade la reconnaissance beaucoup plus qu'une même rotation pour des objets quelconques (voir Figure 22 - Quand les visages ne sont pas vus dans leur état naturel, la capacité du système visuel humain à les distinguer

Figure 22 - Quand les visages ne sont pas vus dans leur état naturel, la capacité du système visuel humain à les distinguer est dégradée. [46]

est dégradée.) Dans un travail innovant [46], Moscovitch et al. ont démontré que le cerveau des hommes traite les visages et les objets dans des aires séparées à savoir que les visages sont traités dans une aire spéciale. Cette approche a fait office de référence pendant la dernière décennie jusqu'au travail ré- cent de Jiang *et al.* [47] qui montre que le traitement des visages et des objets ne repose finalement pas sur des mécanismes différents.

Les systèmes de reconnaissance de visages sont très souvent classés à partir des conclusions d'études psychologiques sur la façon dont les hommes utilisent les caractéristiques faciales pour reconnaitre les autres. De ce point de vue, on distingue les trois catégories : les méthodes globales, les méthodes locales et les méthodes hybrides(Voir Figure 23).

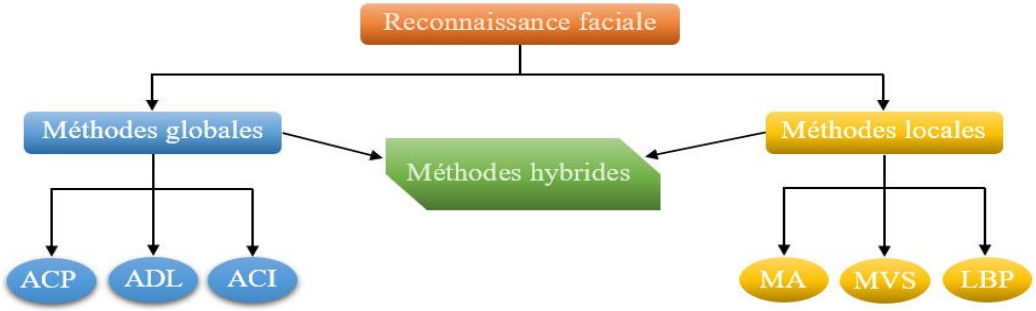


Figure 23 - Schéma illustrant les différentes méthodes de la reconnaissance faciale [48]

2.4.2.1 **Méthodes globales**

Le principe de ces méthodes est de représenter une image faciale par un seul vecteur de grande dimension en concaténant les niveaux de gris de tous les pixels du visage. Cette représentation, appelée description basée sur *l'apparence globale*, a deux avantages. Premièrement, elle conserve implicitement toutes les informations de texture et de forme utiles pour différentier des visages. Deuxièmement, elle peut tenir compte des aspects d'organisation structurelle globaux du visage. Toutefois, son inconvénient majeur réside dans la dimension très grande de l'espace image qu'elle nécessite ce qui rend très difficile la classification.

Pour traiter le problème des données de grande dimension, des techniques de réduction de la dimensionnalité peuvent être utilisées. L'une des techniques les plus courantes pour la reconnaissance de visages est la description par visages propres [48], qui est basée sur l'analyse en composantes principales (ACP).

- **Analyse en composantes principales (ACP)**

Une méthode très populaire, basée sur la technique ACP, est la méthode *Eigenface* [48]. Son principe est le suivant : étant donné un ensemble d'images de visages exemples, il s'agit tout d'abord de trouver les composantes principales de ces visages. Ceci revient à déterminer les vecteurs propres de la matrice de covariance formée par l'ensemble des images exemples. Chaque visage exemple peut alors être décrit par une combinaison linéaire de ces vecteurs propres. Pour construire la matrice de covariance, chaque image de visage est transformée en vecteur. Chaque élément du vecteur correspond à l'intensité lumineuse d'un pixel. Cette méthode sera présentée avec davantage de détails dans le chapitre suivant.

L'ACP est une technique rapide, simple et populaire dans l'identification de modèle, c'est l'une des meilleures techniques. Les projections de l'ACP sont optimales pour la reconstruction d'une base de dimension réduite. Cependant, l'ACP n'est pas optimisé pour la séparabilité (discrimination) de classe. Une alternative qui est l'analyse discriminante linéaire LDA tient compte de ceci.

- **Analyse Discriminante Linéaire (ADL)**

Une autre méthode très connue est celle basée sur l'ADL (Analyse discriminante linéaire). L'objectif de la plupart des algorithmes basés sur l'ADL [49, 50] est de trouver les directions de projection les plus discriminantes dans l'espace propre, en maximisant le ratio entre les variations interpersonnelles et les variations intra-personnelles. Comme les variations intra-personnelles peuvent être petites (notablement quand il n'y a pas beau- coup d'images par individu), ce ratio est difficile à maximiser puisqu'il est déjà grand.

Ce problème est encore appelé *Small Sampler Size*. Pour l'éviter, on peut utiliser tout d'abord l'ACP et ensuite l'ADL, et cette méthode est appelée *Fisherfaces*. Voilà pourquoi

Les méthodes basées sur l'ADL ne fonctionnent bien que lorsque beaucoup d'images par personne sont disponibles dans la base d'apprentissage. En revanche, quand il n'y a pas beaucoup d'images par personne, les méthodes basées sur l'ADL marchent moins bien que celles basées sur l'ACP [49].

- **L'Analyse en Composantes Indépendantes (ACI)**

L'ACI est un algorithme global, basé sur le concept intuitif de contraste. En effet, pour extraire une information pertinente d'un ensemble riche de données complexes et non structurées, il faut optimiser le contraste, c'est-à-dire disposer de différents points de vue, à partir de directions les plus éloignées les unes des autres. L'algorithme ACI appliqué à la reconnaissance de visage a été proposé initialement par Bartlett. Ainsi, en reconnaissance faciale, l'ACI minimise les dépendances statistiques d'ordre élevé des données d'entrée et non pas seulement du second ordre, comme c'est le cas dans l'ACP (matrice de covariance), en tentant de trouver les bases sur lesquelles les données projetées sont statistiquement indépendantes entre-elles.

L'avantage principal des méthodes globales est qu'elles sont rapides à mettre en œuvre, les calculs reposent sur des opérations matricielles relativement simples. En revanche, elles sont très sensibles aux variations d'éclairage, de pose et d'expression faciale, puisque la moindre variation des conditions de l'environnement ambiant entraîne des changements inéluctables dans les valeurs des pixels qui sont traités directement [51].

Bien que les méthodes globales aient eu beaucoup de succès, leur inconvénient majeur réside dans le fait qu'elles utilisent uniquement des photos 2D d'apparence faciale. Or, on sait qu'une telle représentation est sensible aux changements d'expression, d'illuminationet de poses. Une manière d'éviter ce problème consiste à utiliser des représentations faciales locales. En effet, les caractéristiques locales ne sont généralement pas aussi sensibles aux changements d'apparence que les caractéristiques globales.

2.4.2.2 **Méthodes locales**

Les méthodes locales peuvent être classées en deux catégories, les méthodes basées sur les points d'intérêt et celles basées sur l'apparence du visage.

- **Méthodes locales basées sur les points d'intérêts**

On détecte tout d'abord les points d'intérêts et ensuite on extrait des caractéristiques localisées sur ces points d'intérêt. Les méthodes les plus anciennes en reconnaissancede visages appartiennent à cette catégorie [52, 53] Elles s'appuient toutes [53, 54] sur l'extraction de caractéristiques géométriques spécifiques telles que la largeur de la tête,les distances entre les yeux,... Ces données sont ensuite utilisées par des classificateurs afinde reconnaître des individus. Ces méthodes présentent les deux inconvénients suivants :

1. Les caractéristiques géométriques sont difficiles à extraire dans certains cas puisque la tâche de la détection précise de points caractéristiques n'est pas facile, en particulier dans les cas où des occultations ou des variations (pose, expression) de visages sont présentes.

2. Les caractéristiques géométriques seules ne sont pas suffisantes pour représenter réellement un visage, alors que d'autres informations utiles telles que les valeurs des niveaux de gris de l'image sont complètement écartées.

Ces deux limites ont engendré deux directions de recherche. La première se concentre surles performances des détecteurs de points caractéristiques du visage. Brunelli et Poggio [54] ont proposé d'utiliser un ensemble d'apprentissage pour détecter la position de l'œil dans une image. Ils ont tout d'abord calculé pour chaque point des coefficients de corrélation entre l'image de test et les images de l'ensemble d'apprentissage et ensuite ils ont cherché les

valeurs maximales. Rowley *et al.* [55] ont utilisé plusieurs détecteurs de traits spécifiques correspondant à chaque partie du visage, tels que les yeux, le nez, la bouche,...Malgré toutes ces recherches, il n'existe pas encore de détecteur de points caractéristiques qui soit suffisamment précis.

Dans la deuxième direction, les méthodes se concentrent sur des représentations plus élaborées des informations portées par les points caractéristiques du visage, plutôt que simplement sur des caractéristiques géométriques. Manjunath *et al*. [56] ont proposé des algorithmes pour détecter et représenter des caractéristiques faciales à partir d'ondelettes de Gabor. Pour chaque point détecté, deux types d'information sont stockées : sa positionet ses caractéristiques.

Pour modéliser la relation entre les points caractéristiques, un graphe topologique est construit pour chaque visage. Plus tard, Wiskott *et al.* [57] ont proposé une méthodetrès connue appelée *Elastic Bunch Graph Matching* (EBGM), où les nœuds des graphes sont situés sur un certain nombre de points sélectionnés du visage (voir Figure 24) De manière similaire à la méthode décrite dans l'étude de Manjunath *et al.* [56], Wiskott *et al.* ont utilisé les ondelettes de Gabor pour extraire les caractéristiques des points détectés car les filtres de Gabor sont robustes aux changements d'illumination, aux distorsions et aux variations d'échelle.

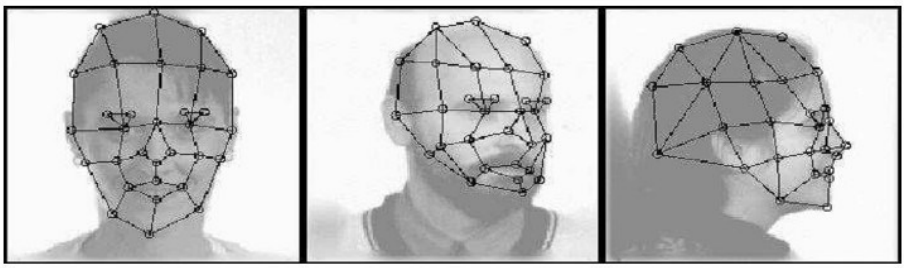


Figure 24 - EBGM [56]

- **Les méthodes locales basées sur l'apparence du visage**

Dans les méthodes locales basées sur l'apparence du visage, on divise le visage en petites régions (ou patches) sur lesquelles les caractéristiques locales sont extraites directement. Martinez a présenté une approche probabiliste locale pour la reconnaissance de visages qui sont occultés partiellement et avec des variations d'expression dans le cas où une seule image de référence est disponible [58]. De nombreux échantillons virtuels sont d'abord générés à l'aide d'une méthode de perturbation de l'image, puis, chaque visage estdivisé en six régions locales en forme d'ellipse. Ensuite, tous les patches locaux à la mêmeposition de chaque visage sont regroupés séparément dans un sous-espace de visages (donc six sous-espaces au total). Dans la phase d'identification, les images de test sont égalementdivisées en six zones locales et sont projetées sur les espaces propres calculés ci-dessus, respectivement. Une approche probabiliste est utilisée pour mesurer la similarité d'une paire d'images. Des expériences sur un ensemble de 2600 images montrent que l'approche probabiliste locale ne réduit pas le taux de reconnaissance même lorsque $\frac{1}{6}$ du visage est occulté. Cependant, les complexités de calcul et de stockage ainsi que la procédure de génération des échantillons virtuels sont très compliquées (6615 échantillons par individu), en particulier si on a une base de référence avec de nombreux visages. La variation de poseest l'une des questions les plus

importantes et difficiles en reconnaissance automatique devisages, en particulier dans le cas où une seule image de référence est disponible. Pour traiter ce problème, Kanade et Yamada [59] ont proposé une méthode probabiliste qui est similaire à celle de Moghaddam et Pentland [60].

Tableau 1 - Comparaison des propriétés des caractéristiques locales et des caractéristiques globales

Variations	Caractéristiques locales	Caractéristiques globales
Petites variations	Pas sensible	Sensible
Grandes variations	Sensible	Très sensible
Illuminations	Pas sensible	Sensible
Expressions	Pas sensible	Sensible
Pose	Sensible	Très sensible
Bruit	Très sensible	Sensible
Occultations	Pas sensible	Très sensible

Les méthodes mentionnées ci-dessus ne considèrent pas explicitement les relations existantes entre les caractéristiques locales. Il est concevable que l'utilisation de cette information soit bénéfique pour le système de reconnaissance. Une solution possible est de construire un modèle flexible de géométrie sur les caractéristiques locales comme cela se fait dans la méthode EBGM.

- **Méthodes hybrides**

Les méthodes hybrides (ou méthodes de fusion) sont des approches utilisant à la fois des caractéristiques globales et des caractéristiques locales. Les facteurs clés qui influent les performances des méthodes de fusion comprennent le choix des caractéristiques pour la combinaison et la manière de les combiner de telle sorte que leurs avantages soient préservés et que leurs inconvénients soient évités.

Les caractéristiques locales et les caractéristiques globales ont des propriétés très différentes et peuvent offrir des informations complémentaires utiles à la tâche de classification. Notons aussi que d'un certain point de vue, les méthodes locales peuvent être considérées comme des méthodes hybrides car des informations globales sont générale- ment prises en compte. Dans la méthode probabiliste locale [58] de nouveaux échantillons d'apprentissage sont d'abord produits pour chaque personne par méthode globale, puis une méthode locale est utilisée pour la reconnaissance.

2.4.3 **Principales difficultés de la reconnaissance de visage**

2.4.3.1 **Influence des changements d'éclairage**

L'intensité et la direction d'éclairage lors de la prise de vue influent énormément sur l'apparence du visage dans l'image. En effet, dans la plupart des applications courantes, des changements dans les conditions d'éclairage sontinévitables, notamment lorsque les vues sont collectées à des heures différentes, en intérieur ou en extérieur. Etant donnée la forme spécifique d'un visage humain, ces variations d'éclairage peuvent y faire apparaître des ombres accentuant ou masquant certaines caractéristiques faciales [44].

Figure 25- Exemple de variation d'éclairage [44]

2.4.3.2 **Influence des variations de la pose**

Les changements d'orientation et les changements de l'angle d'inclinaison du visage engendrent de nombreuses modifications d'apparence dans les images collectées. Une phase préliminaire de normalisation de l'image du visage permet de corriger d'éventuelles rotations dans le plan de celle-ci. Les rotations en profondeur engendrent l'occultation de certaines parties du visage comme pour les vues de trois- quarts. D'autre part, elles amènent des différences de profondeur qui, projetées sur le plan 2D de l'image, provoquent des déformations qui font varier la forme globale du visage. Ces déformations qui correspondent à l'étirement de certaines parties du visageet la compression d'autres régions font varier aussi les distances entre lescaractéristiques faciales. [44]

Figure 26 - Exemples de variation de poses [44]

2.4.3.3 **Influence des expressions faciales**

Les visages sont des éléments non rigides. Les expressions faciales véhiculant des émotions, combinées avec les déformations induites par la parole, peuvent produire des changements d'apparence importants, et le nombre de configurations possibles est trop important pour que celles-ci soient décrites *in extenso* de façon réaliste. L'influence de l'expression faciale sur la reconnaissance est donc difficile à évaluer avec précision. Cependant, du fait que ce facteur

affecte la forme géométrique et les positions des caractéristiques faciales, les techniques globales ou hybrides y sont généralement plus robustes que la plupart des techniques géométriques [44]

Figure 27 - Exemples de variation d'expressions [44]

2.4.3.4 **Influence des occultations partielles**

Le visage peut être partiellement masqué par des objets ou par le port d'accessoires tels que des lunettes, un chapeau, une écharpe, *etc*. Les occultations peuvent être intentionnelles ou non. Dans le contexte de la vidéosurveillance, il peut s'agir d'une volonté délibérée d'empêcher la reconnaissance. Il est clair que la reconnaissance sera d'autant plus difficile que peu d'éléments discriminants seront simultanément visibles. [44]

Figure 28 - Examples des occultations partielles [44]

2.4.4 **Les applications de la reconnaissance faciale**

La reconnaissance faciale utilisant des techniques d'apprentissage profond a ouvert la voie à de nombreuses applications dans divers domaines. Voici quelques exemples d'applications de la reconnaissance faciale basée sur l'apprentissage profond :

Tableau 2 - Applications typiques de la reconnaissance de visage [61]

Domaines	**Applications spécifiques**
Divertissement	Jeux vidéo, réalité virtuelle, programmes de formation

	Interaction homme-robot, Interaction homme-ordinateur
Cartes intelligentes	Permis de conduite,
	Immigration, ID nationaux, passeports, enregistrement d'électeurs
	Fraude d'assistance sociale
Sécurité de l'information	Contrôle parental de TV, ouverture personnelle de dispositifs ou de bureaux
	Sécurité des applications, Sécurité des bases de données, Encryptage de fichiers
	Sécurité Intranet, Accès Internet, Enregistrements Médicaux
	Sécurité des terminaux publics
Renforcement de la loi etsurveillance	Vidéo surveillance avancée
	Commande de portique, Analyse après événements
	Recherche et poursuite de suspects

2.5 Travaux antérieures connexes

2.5.1 La reconnaissance faciale

Plusieurs recherches antérieures ont abordé le sujet de la reconnaissance faciale et ses applications dans différents domaines. Cette section présente un aperçu des travaux connexes, en mettant l'accent sur les études les plus pertinentes et les avancées récentes.

Reconnaissance faciale dans la sécurité : Jones et al. (2018) ont réalisé une étude approfondie sur l'utilisation de la reconnaissance faciale dans les systèmes de sécurité. Leur recherche a démontré une précision élevée dans la détection d'individus suspects dans les lieux publics, mais a également souligné la nécessité d'aborder les problèmes de biais et de protection de la vie privée. [62]

Smith et coll. (2019) ont examiné l'utilisation de la reconnaissance faciale comme méthode d'authentification biométrique. Leur étude a révélé une précision significative dans la vérification des identités, mais a également mis en évidence les défis liés à la robustesse de la technologie face aux attaques de contrefaçon et aux variations de l'apparence due à des facteurs externes. [63]

Reconnaissance faciale dans la réalité virtuelle : Chen et al. (2020) ont exploré les applications de la reconnaissance faciale dans la réalité virtuelle. Leur recherche a montré comment la reconnaissance des expressions faciales peut améliorer l'immersion et l'interaction dans les environnements virtuels, ouvrant ainsi de nouvelles possibilités pour les jeux, la formation et la simulation. [64]

Biais et équité dans la reconnaissance faciale : Zhang et al. (2021) ont étudié les problèmes de biais et d'équité dans les systèmes de reconnaissance faciale. Leur recherche a mis en évidence les disparités dans les performances des algorithmes en fonction de l'origine ethnique et du genre

des individus, soulignant la nécessité d'une collecte de données plus diversifiée et de techniques de détection de biais. [65]

Vie privée et protection des données dans la reconnaissance faciale : Brown et al. (2022) ont examiné les questions de vie privée et de protection des données liées à l'utilisation de la reconnaissance faciale. Leur recherche a proposé des méthodes de protection des données sensibles, telles que la cryptographie homomorphique et l'apprentissage fédéré, pour préserver la confidentialité des informations biométriques. [66]

2.5.2 Réseaux Convolutifs Profonds

Cette section examine les travaux de recherche antérieurs liés aux Deep Convolutional Networks (Réseaux Convolutifs Profonds). Les travaux connexes se divisent en trois grandes catégories : (1) les avancées dans l'architecture des réseaux convolutifs, (2) les améliorations de l'apprentissage et de l'entraînement des réseaux convolutifs, et (3) les applications spécifiques des réseaux convolutifs dans différents domaines.

2.5.2.1 Avancées dans l'architecture des réseaux convolutifs

Depuis l'introduction des réseaux convolutifs profonds par LeCun et al. (1998), de nombreuses avancées ont été réalisées dans l'architecture des réseaux convolutifs. L'un des développements majeurs est l'introduction des architectures en profondeur, telles que le réseau VGG (Simonyan et al., 2014) et le réseau ResNet (He et al., 2015). Ces architectures profondes ont démontré une amélioration significative des performances de classification en exploitant des structures plus complexes et des couches plus profondes.

2.5.2.2 Améliorations de l'apprentissage et de l'entraînement des réseaux convolutifs

Plusieurs recherches se sont concentrées sur l'amélioration de l'apprentissage et de l'entraînement des réseaux convolutifs. Parmi ces travaux, on peut citer l'utilisation de techniques d'optimisation avancées, telles que l'optimisation stochastique par descente de gradient (SGD) avec moment (Sutskever et al., 2013) et l'optimisation adaptative du taux d'apprentissage (Adam) (Kingma et Ba, 2015). De plus, des techniques de régularisation, comme l'ajout de couches de normalisation par lots (Batch Normalization) (Ioffe et al., 2015), ont été proposées pour améliorer la généralisation des réseaux convolutifs.

2.5.2.3 Applications spécifiques des réseaux convolutifs

Les réseaux convolutifs ont été largement utilisés dans divers domaines et ont donné des résultats impressionnants. Dans le domaine de la vision par ordinateur, les réseaux convolutifs ont été appliqués avec succès à des tâches telles que la détection d'objets (Girshick et al., 2014), la segmentation sémantique (Long et al., 2015) et la reconnaissance faciale (Sun et al., 2014). De plus, les réseaux convolutifs ont également été utilisés dans des domaines tels que la traduction automatique (Wu et al., 2016) et la génération d'images (Goodfellow et al., 2014).

2.5.3 La différence entre notre travail et les travaux antérieurs

Dans notre travail de recherche, nous avons réalisé une analyse approfondie des travaux antérieurs dans le domaine de la détection d'intrusion et de la vidéosurveillance. Cette revue de la littérature nous a permis de comprendre les différentes techniques, méthodes et algorithmes existants utilisés dans ce domaine.

Cependant, nous avons identifié certaines différences significatives entre notre travail et les travaux antérieurs, ce qui a constitué une contribution originale à la recherche. Voici les principales différences :

Méthodologie : Nous avons développé une approche spécifique pour la détection d'intrusion en exploitant la vidéosurveillance, en utilisant des algorithmes et des techniques différents de ceux utilisés dans les travaux antérieurs. Nous avons pris en compte les spécificités de notre contexte de recherche pour adapter la méthodologie à notre objectifs et contraintes.

Techniques de détection : Nous avons utilisé des techniques avancées d'apprentissage approfondi pour améliorer la détection des objets et la reconnaissance faciale. En appliquant des réseaux neuronaux profonds, nous avons pu obtenir des résultats précis et fiables. Ces modèles d'apprentissage approfondi ont été entraînés sur de vastes ensembles de données afin de mieux comprendre les caractéristiques des objets et des visages, permettant ainsi une détection et une reconnaissance plus précises dans des scénarios variés.

Évaluation et résultats : Nous avons évalué les performances de notre système de détection d'intrusion en utilisant des jeux de données spécifiques et en comparant les résultats obtenus avec ceux des travaux antérieurs. Nous avons démontré comment notre approche a de bon résultat en termes de précision, de sensibilité, de spécificité ou d'autres mesures de performance pertinentes.

En utilisant les travaux antérieurs comme base, Nous avons intégré leurs concepts, méthodes et résultats dans notre propre travail. Nous avons également cité et référencé ces travaux pour montrer comment notre recherche s'inscrit dans le contexte plus large de la détection d'intrusion en vidéosurveillance.

En résumé, notre recherche se distingue des travaux antérieurs par sa méthodologie spécifique, les caractéristiques extraites, l'algorithme de détection développé, ainsi que les évaluations et les résultats obtenus. Nous avons utilisé les travaux antérieurs comme point de départ pour construire et améliorer notre propre travail, tout en apportant des contributions originales au domaine de la détection d'intrusion en vidéosurveillance.

2.6 Conclusion

Dans ce chapitre, nous avons défini le deep learning, ainsi que ses différents architecteurs. Nous avons focalisé notre attention sur les réseaux de neurones convolutifs CNN et leurs structures, et ses différentes couches. Ce chapitre est un état de l’art sur les concepts reliés à la problématique traitée dans ce mémoire et sa conception. Dans ce chapitre aussi, nous avons défini la reconnaissance faciale et leurs systèmes, ainsi ces techniques locale, global et hybride, enfin nous avons mis en évidence les différentes difficultés inhérentes à la reconnaissance de visages, notamment l’invariance à l’illumination, les changements de pose et les expressions faciales. Dans le chapitre suivant nous allons expliquer les étapes de conception et d’implémentation de notre système de détection de l’intrus exploitant la reconnaissance faciale.

Chapitre 3 : Conception expérimentale

3.1 Introduction

Ce chapitre joue un rôle essentiel dans notre mémoire. Il nous permet de présenter notre méthodologie expérimentale, de discuter des résultats obtenus et d'évaluer les performances réelles de notre système de détection de l'intrus. En fournissant une analyse approfondie et des discussions pertinentes, nous contribuerons à une meilleure compréhension du fonctionnement de notre système et à l'identification des opportunités d'amélioration pour des applications futures.

3.2 Méthodologie de conception

Cette section détaillée sur la méthodologie de conception offrira une vision claire et transparente du processus suivi pour développer notre système de détection d'intrusion. Il permettra de mieux comprendre les décisions prises et les raisons qui les sous-tendent, tout en fournissant un cadre méthodologique solide pour évaluer et analyser les performances du système.

3.2.1 Vue globale

Le système que nous avons développé se divise en deux parties distinctes, chacune jouant un rôle essentiel dans la détection de l'intrusion en temps réel et la gestion du système. La première partie concerne le système de détection de l'intrusion en temps réel, tandis que la deuxième partie concerne la plateforme web dédiée à la gestion et au contrôle du système.

La première partie, le système de détection de l'intrusion en temps réel, est responsable de la surveillance continue des flux vidéo provenant des caméras de vidéosurveillance. Il utilise des algorithmes avancés de traitement d'images et d'apprentissage automatique pour analyser en temps réel les images et détecter toute intrusion. Ces algorithmes peuvent inclure la détection intelligente de mouvement, la reconnaissance faciale, etc. Lorsqu'une intrusion est détectée, le système déclenche des alertes ou des notifications pour informer les utilisateurs de l'incident en cours.

La deuxième partie est une plateforme web conviviale conçue pour la gestion et le contrôle du système de détection d'intrusion. Cette plateforme permet aux utilisateurs autorisés de surveiller les flux vidéo en direct, d'accéder aux enregistrements vidéo archivés et de visualiser les alertes générées par le système. Les utilisateurs peuvent également configurer les paramètres du système, tels que les zones de détection, les horaires de surveillance, etc. La plateforme offre une interface intuitive pour faciliter la navigation et l'utilisation, permettant aux utilisateurs de gérer efficacement le système de détection d'intrusion.

Les deux parties du système, à savoir le système de détection en temps réel et la plateforme web de gestion, communiquent entre elles via un réseau informatique. Cela permet la transmission des données de détection d'intrusion en temps réel vers la plateforme de gestion, où elles peuvent être analysées, enregistrées et traitées. La communication entre les deux parties peut être établie à l'aide de protocoles réseau standard tels que TCP/IP, XML-RPC, permettant un échange de données sécurisé et fiable.

Le système de détection de l'intrusion en temps réel surveille toutes les caméras enregistrées via la plate-forme Web et celles qui ont été défini comme caméra active. Lorsque le

système détecte une ou plusieurs personnes avec la technologie de la détection intelligente de mouvement dans une de ces caméras, le système commence à enregistrer un résumé de vidéo, qui ne montre que des séquences contenant des personnes, et en même temps commence le processus de reconnaissance faciale pour identifier les personnes montrées par cette caméra.

Si le système ne reconnaît pas un visage, une notification d'intrus est envoyée à la plateforme web (notification de personne anonyme).

Dans le cas où l'identité de la personne est connue, le système vérifie que cette personne est habilitée à être présente à l'endroit où la caméra l'a captée. Dans le cas où elle n'en a pas l'habilitation, une notification de présence d'un intrus est envoyé à la plateforme web (une notification de non-conformité à l'autorité).

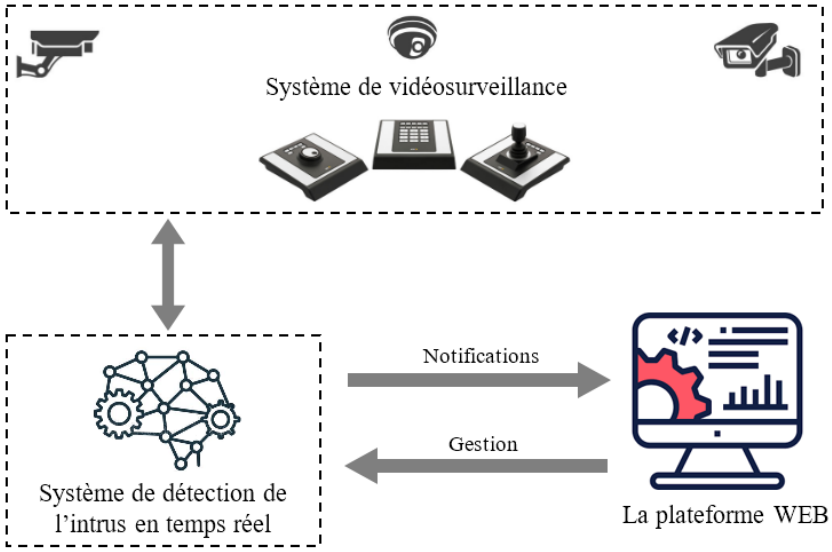


Figure 29 - Architecture globale du système proposé

Le système proposé fonctionne en deux phases, la première est la phase d'enrôlement au cours de laquelle il est question d'ajouter au système les personnes et les privilèges dont elles disposent, qui sont liés aux différents lieux surveillés par des caméras, l'ajout de ces personnes se fera par la plateforme web et consistera à renseigner des informations telle que les informations personnelles des personnes et le groupe auquel appartiennent ils qui détermine leurs privilèges, en plus de la chose la plus importante, qui est les photos de ces personnes. Ces photos serviront à entrainer le système pour leur identification ultérieure.

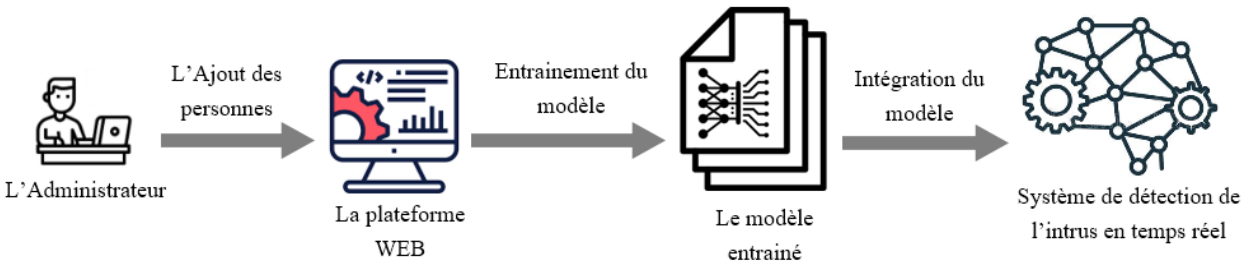


Figure 30 - La phase d'enrôlement

La seconde phase qui est la phase opérationnelle fera l'objet de la sous-section suivante.

3.2.2 Fonctionnement du système

Le fonctionnement du système repose sur une combinaison de technologies de détection, de traitement d'image, de communication réseau et de gestion centralisée. En utilisant ces composants interconnectés, le système offre une solution globale pour la détection d'intrusion, la collecte de preuves et la gestion des incidents, contribuant ainsi à renforcer la sécurité des sites surveillés.

3.2.2.1 Système de détection d'intrusion en temps réel

Le système de détection d'intrus est un ensemble de modules. Il s'agit principalement du module chargé de la détection intelligente de mouvement et de l'enregistrement des séquences vidéos détectées par et du module chargé de l'identification des personnes la reconnaissance faciale.

- **Module de détection intelligente de mouvement**

Il s'agit de module qui travaille constamment tant que le système de détection d'intrusion est actif, qui vérifie chaque image de vidéo de chaque caméra pour détecter toute apparition humaine devant n'importe quelle caméra de surveillance. Ce module joue le rôle d'un déclencheur, qui lance les opérations d'enregistrement et de détection faciale s'il détecte une personne. Comme il termine l'opération d'enregistrement, s'il ne capture aucune personne (Voir la Figure 31 - Fonctionnement du module de détection de personnes).

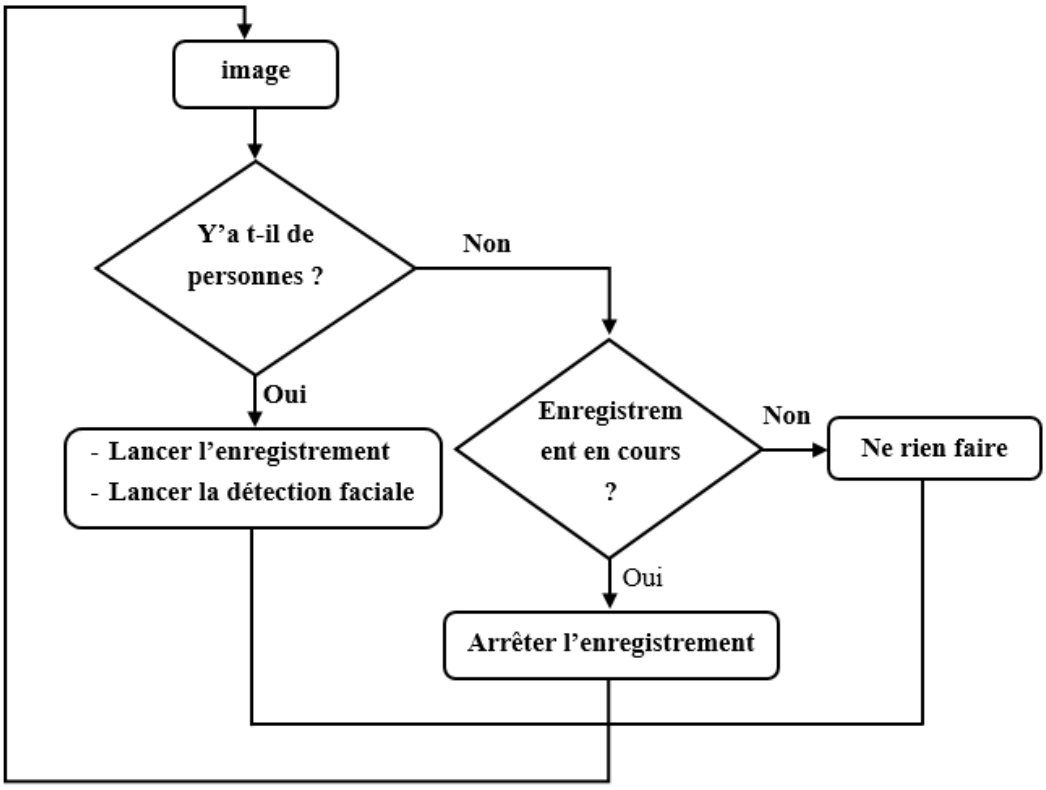


Figure 31 - Fonctionnement du module de détection de personnes

- **Module de la reconnaissance faciale**

Ce module se lance la reconnaissance faciale dès qu'une personne a été détectée en parallèle avec l'enregistrement de la vidéo, il commence à rechercher des visages dans chaque image de

vidéo de la caméra, il identifie chaque visage détecté en comparant les caractéristiques du visage détecté avec celles dans la base de données. Si le visage n'est pas identifié, une notification d'existence d'intrus envoyer, sinon on va lancer le module de vérification des privilèges(voir la Figure 32 - Fonctionnement du module de la reconnaissance faciale).

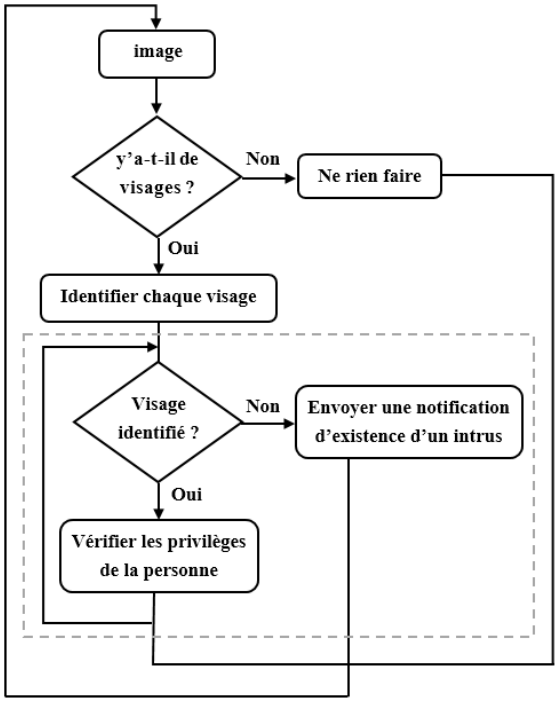


Figure 32 - Fonctionnement du module de la reconnaissance faciale

- **Module de vérification des privilèges**

Après une identification d'une personne, le module de vérification de privilèges se lance pour vérifier les autorisations en fonction du site, qui concernes la personne détectée, si cette personne n'a pas des autorisations de site, une notification de l'intrus envoyer à la plateforme web. Sinon, en lance la vérification en fonction du temps, si elle n'a pas des autorisations de temps, une notification de l'intrus envoyer (voir la Figure 33 - Fonctionnement du module de vérification des privilèges).

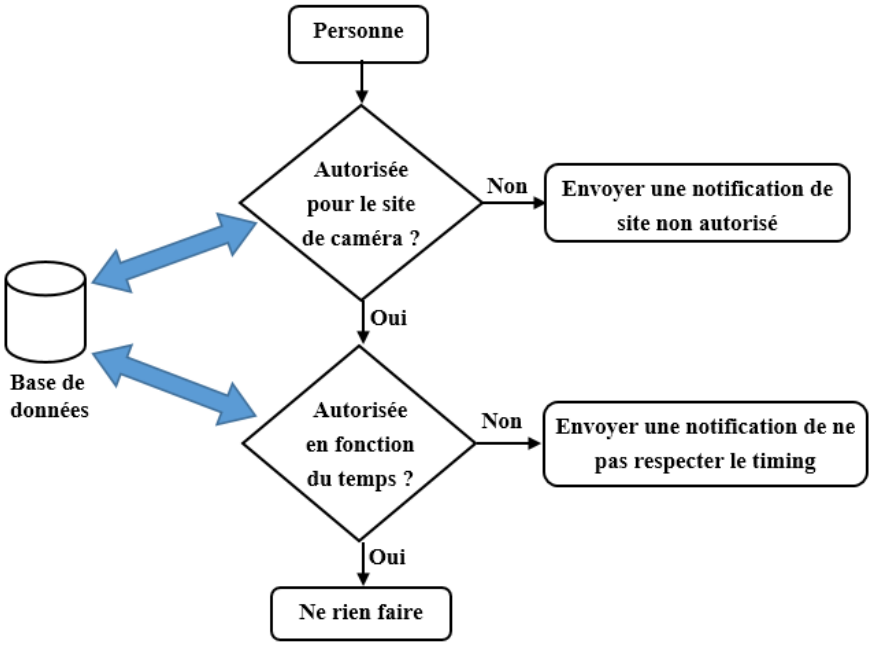


Figure 33 - Fonctionnement du module de vérification des privilèges

3.2.2.2 La plateforme Web

La plateforme web est un complémentaire pour notre système de détection de l'intrus en temps réel, elle garantit l'interaction entre le système et ses administrateurs.

- **Le rôle de la plateforme Web**

La plateforme web dans ce système a deux rôles principaux. Tout d'abord, elle assure la gestion et la configuration du système de détection de l'intrus en temps réel. Elle offre une interface conviviale qui permet aux utilisateurs autorisés de gérer les paramètres du système, tels que les caméras à surveiller, les zones sensibles, les seuils de détection, etc. Les utilisateurs peuvent également configurer les préférences de notification. La plateforme permet ainsi une administration centralisée du système, facilitant la surveillance et la maintenance efficaces du système de détection d'intrusion.

a. La gestion du système

Cette plateforme nous permet de gérer le système dans ses différents axes, comme suit :

1. **Enrôlement des personnes :** est l'action d'inscrire toutes les personnes qui circules dans l'organisme contrôlé par notre système, cette opération se déroule en deux étapes :
    a. **Enregistrement des personnes :** enregistrement des personnes avec leurs coordonnées et leurs photos, Il faut que les photos d'une personne soient variées en termes de poses, d'angles d'éclairage et d'expressions faciales.
    b. **Entrainement du modèle :** La plateforme web nous permet d'entraîner le modèle de deep learning en utilisant les photos des personnes enregistrés dans la base de données, afin de pouvoir les identifier par le système ultérieurement.
2. **Gestion des sites :** La gestion des emplacements des caméras, nous permet de contrôler la gestion des autorisations des personnes.

3. **Gérer les caméras :** A l'aide de la plateforme web, on peut ajouter, supprimer, activer ou désactiver des caméras contrôlées par le système de détection de l'intrus.
4. **Contrôler l'archive des résumés de vidéos :** La plate-forme nous permet de visualiser des vidéos, de les supprimer ou de consulter des notifications à leur sujet.
5. **La gestion des autorisations pour les personnes :** nous permet de gérer les autorisations liées aux personnes, ce qui concerne les lieux et le temps.
6. **Contrôler les types des objets à observer :** Notre système offre la possibilité de sélectionner les types des objets à observer par le système, En plus des humains, le système peut détecter et identifier d'autres types d'objets et les signaler

b. Recevoir les notifications d'intrusion

L'un des deux principaux rôles de la plateforme, à côté de la gestion de système, est de recevoir et la gestion des différents types de notifications, qui vient du système de détection de l'intrus en temps réel. On distingue trois types de notification de détection de l'intrus :

1. Personne inconnue.
2. Site non autorisé.
3. Ne pas respecter le timing.
4. Présence d'animaux.

▪ La structure conceptuelle de la plateforme Web

La plateforme Web est une application Web qui se base sur une base de données, cette base de données nous a permet de stocker les données nécessaires pour la gestion de notre système.

a. Le diagramme de classes

Le diagramme de classes décrive clairement la structure de notre système en modélisant ses classes, ses attributs, ses opérations et les relations entre ses objets (voir la Figure 34 - Le diagramme de classes qui décrive la structure de la base de données de la plateforme Web).

➢ Les classes

Voici une description des différentes classes du diagramme de classes :

- **Caméra :** pour stocker les différentes caméras qui fonctionnes dans le système, et de les gérer, configurer et consulter les vidéos enregistrées par ces caméras.
- **Enregistrement :** Permet de gérer les enregistrement vidéos qui sont capturées à la détection des personnes par le système.
- **Notification :** C'est la classe qui détermine la façon de stockage des différents types de notifications envoyées par le système.
- **Personne :** Cette classe représente les personnes que le système connaît, où leurs images sont stockées, ces images sont utilisées dans le processus de l'apprentissage du système et de la création du modèle utilisé pour les identifier.

- **Groupe :** C'est le groupe de personnes qui ont les mêmes autorisations concernant l'espace et le temps.
- **Lieu :** Représente l'ensemble des lieux surveillés par le système de détection d'intrusion en temps réel.
- **Temps** : Une classe qui définit les temps utilisés pour définir les autorisations.

Ces classes représentent les principaux composants du système de détection d'intrusion et de sa gestion.

➢ Les relations

Voici une description des différentes relations entre les classes du diagramme de classes :

- **Notification←Enregistrement :** Lier des enregistrements vidéo et des notifications, pour afficher les enregistrements liés à une notification spécifique.
- **Caméra→Enregistrement :** Pour voir quelle caméra a enregistré le clip vidéo.
- **Personne→Notification :** Lier l'identité de l'intrus à la notification, si elle a été déterminée par le système.
- **Groupe→Personne :** Pour pouvoir attribuer la personne à un groupe, où chaque groupe a ses propres pouvoirs et autorisations.
- **Lieu→Caméra :** Pour déterminer l'emplacement de la caméra.
- **Lieu→Lieu/Temps :** Spécifie les autorisations de lieu pour un groupe particulier.
- **Temps→Lieu/Temps :** Spécifie les autorisations d'heure pour un groupe particulier.

Ces relations entre les classes permettent de modéliser les interactions et les dépendances entre les différentes entités du système. Elles aident à représenter la structure et le comportement des classes.

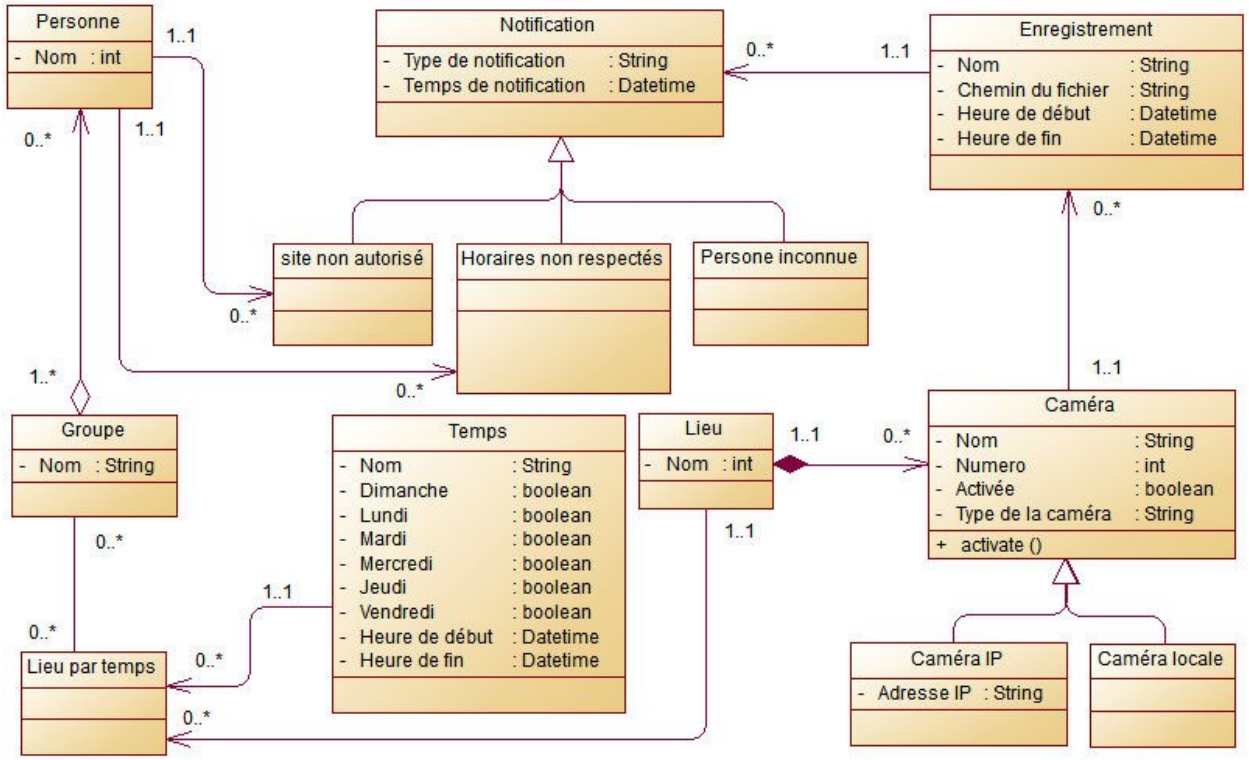


Figure 34 - Le diagramme de classes qui décrive la structure de la base de données de la plateforme Web

Dans cette section, nous avons décrit en détail l'approche que nous avons suivie pour concevoir notre système de détection d'intrusion basé sur la vidéosurveillance. Nous avons mis en évidence les différentes étapes clés de notre méthodologie, en mettant l'accent sur la préparation du prochaine section "Algorithmes et métriques".

3.3 Algorithmes et métriques

Cette section constitue une étape cruciale de notre mémoire, où nous nous concentrons sur les algorithmes spécifiques utilisés dans notre système de détection d'intrusion basé sur la vidéosurveillance, ainsi que les métriques utilisées pour évaluer les performances de notre système. Ce chapitre nous permettra de plonger plus en profondeur dans les détails techniques de notre recherche et de présenter les résultats de manière précise et mesurable.

3.3.1 Algorithme du module de la détection intelligente de mouvement

Notre algorithme de détection intelligente de mouvement est basé sur la technique de détection des objets, Nous avons opté pour cela à l'utilisation des modèles de Deep Learning pré-entrainés de reconnaissance d'objets. Ces modèles sont conçus pour détecter et identifier des objets tels que des personnes, des téléphones, des tables, des chaises... etc. Nous avons utilisé cette technologie dans notre système pour filtrer uniquement les personnes. À cet effet, nous avons combiné MobileNet avec *Single Shot Detector* (SSD) afin de répondre aux limites pratiques de l'exploitation de réseaux de neurones à haute puissance, il est désormais appelé MobileNet-SSD.

3.3.1.1 C’est quoi MobileNet-SSD ?

MobileNet est un réseau de neurones convolutifs, développé par Google, permettant de classifier 21 types d’objets et présentant des avantages tels que légèreté, rapidité et précision. Ce

réseau de neurones est associé au réseau de neurones Single *Shot Multibox Detector* (SSD) permettant d'identifier toutes les zones dans l'image présentant un élément à classifier. La combinaison de MobileNet et de SSD permet d'obtenir une méthode rapide et efficace de détection d'objets basée sur le Deep Learning. Le modèle que nous utilisons est une implémentation sous Caffe entrainée.

La question se pose de savoir pourquoi nous utilisons MobileNet, pourquoi nous ne pouvons pas utiliser ResNet, VGG ou AlexNet ?

La réponse est simple. ResNet ou VGG ou AlexNet a une grande taille de réseau et augmente le taux de calcul, alors que dans MobileNet il existe une architecture simple consistant en une convolution en profondeur 3 × 3 suivie d'une convolution point par point 1 × 1.

3.3.1.2 **Fonctionnement de l'algorithme**

Avant tout on a téléchargé des fichiers nécessaires pour le modèle déjà entrainé de MobileNet SSD :

- Le fichier du modèle entraîné aux 21 types d'objets : « **MobileNetSSD_deploy.caffemodel** »
- Le fichier de configuration : « **MobileNetSSD_deploy.prototxt** »

1. Dans le code python, on doit importer des bibliothèques comme suite :

```
from datetime import datetime
from imutils.video import FPS
import numpy as np
import time
import cv2
import pickle
```

2. Puis, initialiser la liste des étiquettes de classes MobileNet SSD, qui a été formé pour les détecter :

```
CLASSES = ["background", "aeroplane", "bicycle", "bird", "boat", "bottle", "bus",
"car", "cat", "chair", "cow", "diningtable", "dog", "horse", "motorbike", "person",
"pottedplant", "sheep", "sofa", "train", "tvmonitor"]
```

3. Charger le modèle sérialisé à partir du disque, Pour ce faire, on utilise la fonction d'OpenCV **readNetFromCaffe()** :

```
net = cv2.dnn.readNetFromCaffe("MobileNetSSD_deploy.prototxt.txt",
      "MobileNetSSD_deploy.caffemodel")
```

4. Pour contrôler une caméra pour vérifier la présence de personnes, on va boucler sur les images du flux vidéo de cette caméra, comme suite :

```
while True:
```

Pour chaque image du flux vidéo, on va faire le traitement suivant :

a) Obtenir l'heure actuelle, pour l'utiliser ultérieurement pour nommer les enregistrements vidéo, et les notifications :

```
now = datetime.now()
current_time = now.strftime("%d/%m/%Y %H:%M:%S")
```

b) Récupérez le cadre (frame) du flux vidéo fileté et redimensionnez-le :

```
_, frame = self.cap.read()
# redimensionnez l'image
dim = (self.width, self.height)
resized = cv2.resize(frame, dim)
```

c) Saisir les dimensions du cadre et le convertir en blob :

```
(h, w) = resized.shape[:2]
blob = cv2.dnn.blobFromImage(cv2.resize(resized, (300, 300)), 0.007843, (300, 300),
127.5)
```

d) Passer le blob à travers le réseau et obtenir les détections et les prédictions :

```
net.setInput(blob)
detections = net.forward()
```

e) Initialiser le détecteur de personnes

```
person_detected = False
```

f) Boucle sur les détections (la variable `detections`), Pour vérifier l'existence des personnes :

```
for i in np.arange(0, detections.shape[2]):
```

Pour chaque détection, faire :

1- Extraire la confiance (c'est-à-dire la probabilité) associée à la prédiction :

```
confidence = detections[0, 0, i, 2]
```

2- Filtrer les détections faibles en s'assurant que la "confiance" est supérieure à la confiance minimale :

```
if confidence > self.confidence:
```

3- Extraire l'index de l'étiquette de classe de la variable `*detections*` :

```
idx = int(detections[0, 0, i, 1])
```

4- Si l'objet détecté est une personne :

```
if CLASSES[idx] == "person":
```

Alors :

➢ Lancer la détection faciale, pour identifier la personne :

```
person_detected = True
# Lancer la reconnaissance faciale
self.face_recognition(frame, face_cascade, model, labels)
```

La méthode 'face_recognition' est résponsable à la reconnaissance faciale, pour identifier la personne détectée, on va la détaillée dans la partie suivante.

➢ Lancer l'enregistrement de vidéo :

```
if CLASSES[idx] == "person" and not self.recording:

   # Turn on the recording mode
   self.recording = True
```

```
print("[INFO] Person detected, at ", current_time)
print("[INFO] Recording ...")

# Create a video writer
fourcc = cv2.VideoWriter_fourcc('m', 'p', '4', 'v')
# File name
starting_time = now.strftime("%d-%m-%Y_%H-%M-%S")
file_name = "recorded_videos/rec_" + starting_time + ".mp4"
writer = cv2.VideoWriter(file_name, fourcc, 30.0, (self.width, self.height))
```

3.3.2 Algorithme du module de la reconnaissance faciale

La première étape dans la reconnaissance des visages est la détection des visages. Avec la bibliothèque OpenCV, il est assez facile de détecter un visage de face dans une image en utilisant son détecteur de visage Haar Cascade (connu comme la méthode de Viola-Jones).

Etant donné un fichier ou une vidéo en direct, le détecteur de visage examine chaque emplacement de l'image et classifie comme visage ou non visage. Le classifieur utilise des données stockées dans un fichier XML pour décider comment classifier chaque localisation image. OpenCV est livré avec plusieurs classifieur différents pour la détection de visage dans des poses frontales, ainsi que la détection de certains visages de profil, la détection des yeux, détection des corps...etc.

Pour la détection des données XML, nous pouvons choisir l'un de ces classifieurs Haar Cascade d'OpenCV (dans le répertoire "opencv/data/haarcascade") :

- haarcascade_frontalface_default.xml
- haarcascade_frontalface_alt.xml
- haarcascade_frontalface_alt2.xml
- haarcascade_frontalface_alt_tree.xml

Nous avons choisi le fichier `'haarcascade_frontalface_default.xml'` pour l'application car il correspond parfaitement à notre objectif.

Pour charger des données XML, nous utilisons la fonction `CascadeClassifier()` avec le chemin vers le fichier XML contenant cascade de Haar comme son premier paramètre d'entrée :

```
face_cascade = cv2.CascadeClassifier('haarcascade_frontalface_default.xml')
```

Il ne reste plus qu'à effectuer une détection de visage sur l'image capturée, en utilisant haar Cascade et retourner une liste des visages détectés dans l'image donnée :

```
rgb = cv2.cvtColor(frame, cv2.COLOR_BGR2RGB)
faces = face_cascade.detectMultiScale(rgb, scaleFactor=1.3, minNeighbors=5)
```

Maintenant que nous avons détecté un visage, nous pouvons utiliser cette image de visage pour la reconnaissance faciale. Notre système de reconnaissance est basé sur le réseau de neurones convolutionnels VGG16.

3.3.2.1 VGG : en quoi consiste ce modèle ?

VGG est un réseau de neurones convolutionnels proposés par K. Simonyan et A. Zisserman de l'université d'Oxford et qui a acquis une notoriété en gagnant la compétition **ILSVRC** (*ImageNet Large Scale Visual Recognition Challenge*) en 2014. Le modèle a atteint une précision de 92.7% sur *Imagenet* ce qui est un des meilleurs scores obtenus. Il a marqué une progression par rapport aux modèles précédents en proposant, dans les couches de convolution, **des noyaux de convolution de plus petites dimensions (3×3)** que ce qui avait été fait jusque-là. Le modèle a été entraîné sur des semaines en utilisant des cartes graphiques de pointe. [67]

- **ImageNet**

ImageNet est une gigantesque base de données de plus de **14 millions d'images labellisées** réparties dans plus de 1000 classes, en 2014. En 2007, une chercheuse du nom de Fei-Fei Li a commencé à travailler sur l'idée de créer un tel jeu de données. Certes la modélisation est un aspect très important pour obtenir des bonnes performances, mais disposer de données de grande qualité l'est tout autant pour avoir un apprentissage de qualité. Les données ont été collectées et étiquetées depuis le web par des humains.

Method	top-1 val. error (%)	top-5 val. error (%)	top-5 test error (%)
VGG (2 nets, multi-crop & dense eval.)	**23.7**	**6.8**	**6.8**
VGG (1 net, multi-crop & dense eval.)	24.4	7.1	7.0
VGG (ILSVRC submission, 7 nets, dense eval.)	24.7	7.5	7.3
GoogLeNet (Szegedy et al., 2014) (1 net)	-	7.9	
GoogLeNet (Szegedy et al., 2014) (7 nets)	-	**6.7**	
MSRA (He et al., 2014) (11 nets)	-	-	8.1
MSRA (He et al., 2014) (1 net)	27.9	9.1	9.1
Clarifai (Russakovsky et al., 2014) (multiple nets)	-	-	11.7
Clarifai (Russakovsky et al., 2014) (1 net)	-	-	12.5
Zeiler & Fergus (Zeiler & Fergus, 2013) (6 nets)	36.0	14.7	14.8
Zeiler & Fergus (Zeiler & Fergus, 2013) (1 net)	37.5	16.0	16.1
OverFeat (Sermanet et al., 2014) (7 nets)	34.0	13.2	13.6
OverFeat (Sermanet et al., 2014) (1 net)	35.7	14.2	-
Krizhevsky et al. (Krizhevsky et al., 2012) (5 nets)	38.1	16.4	16.4
Krizhevsky et al. (Krizhevsky et al., 2012) (1 net)	40.7	18.2	-

Figure 35- La figure ci-dessus compare les résultats de différents modèles de 2014 ou bien des années précédentes. Nous pouvons voir que VGG donne les meilleurs résultats aussi bien sur le jeu de validation que sur le jeu de test. Remarquons également que le modè [67]

- **L'architecture**

Il existe deux algorithmes disponibles : **VGG16 et VGG19**. Dans ce memoire, nous allons nous concentrer sur l'architecture VGG16. Si les deux architectures sont très proches et respectent la même logique, VGG19 présente un plus grand nombre de couches de convolution.

Figure 36 - Architecture Algoritme VGG16 [67]

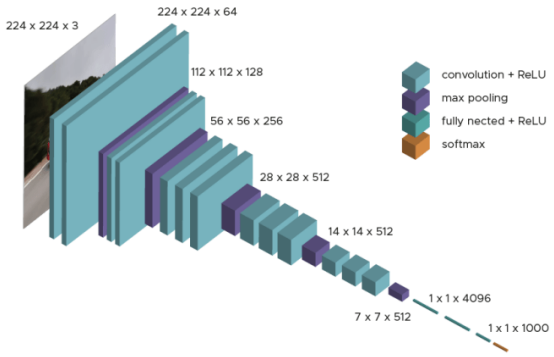


Figure 37 - Structure Algoritme VGG16 [67]

Le modèle ne demande qu'un prétraitement spécifique qui consiste à soustraire la valeur RGB moyenne, calculée sur l'ensemble d'apprentissage, de chaque pixel.

Durant l'apprentissage du modèle, l'input de la première couche de convolution est une image RGB de taille 224 x 224. Pour toutes les couches de convolution, le noyau de convolution est de taille 3×3: la plus petite dimension pour capturer les notions de haut, bas, gauche/droite et centre. C'était une spécificité du modèle au moment de sa publication. Ces couches ont pour but de filtrer l'image en ne gardant que des informations discriminantes comme des formes géométriques atypiques.

Ces couches de convolution s'accompagnent de couche de Max-Pooling, chacune de taille 2×2, pour réduire la taille des filtres au cours de l'apprentissage.

En sortie des couches de convolution et pooling, nous avons 3 couches de neurones Fully-Connected. Les deux premières sont composées de 4096 neurones et la dernière de 1000 neurones avec une fonction d'activation softmax pour déterminer la classe de l'image.

Comme on a pu le constater l'architecture est claire et simple à comprendre ce qui est aussi une force de ce modèle. C'est pour cela on a choisi le VGG16.

3.3.2.2 L'algorithme de l'apprentissage automatique

1) Importer les bibliothèques nécessaires :

```
import os
import pandas as pd
import numpy as np
import tensorflow.keras as keras
import matplotlib.pyplot as plt
from tensorflow.keras.layers import Dense, GlobalAveragePooling2D
from tensorflow.keras.preprocessing import image
from tensorflow.keras.applications.mobilenet import preprocess_input
from tensorflow.keras.preprocessing.image import ImageDataGenerator
from tensorflow.keras.models import Model
from tensorflow.keras.optimizers import Adam
from keras_vggface.vggface import VGGFace
import pickle
```

2) Augmenter les images d'entraînement

```
train_datagen = ImageDataGenerator(preprocessing_function=preprocess_input)
```

```
train_generator = train_datagen.flow_from_directory(
    './Headshots',
    target_size=(224, 224),
    color_mode='rgb',
    batch_size=32,
    class_mode='categorical',
    shuffle=True)
# Obtenir le nombre de classes d'images
train_generator.class_indices.values()
# dict_values([0, 1, 2])
NO_CLASSES = len(train_generator.class_indices.values())
```

3) Construire le modèle :

```
base_model = VGGFace(include_top=False, model='vgg16', input_shape=(224, 224, 3))
# ajouter les couches personnalisés
x = base_model.output
x = GlobalAveragePooling2D()(x)

x = Dense(1024, activation='relu')(x)
x = Dense(1024, activation='relu')(x)
x = Dense(512, activation='relu')(x)

# couche finale avec activation softmax
preds = Dense(NO_CLASSES, activation='softmax')(x)
```

4) Créer un nouveau modèle avec l'entrée d'origine du modèle de base et la sortie du nouveau modèle :

```
model = Model(inputs=base_model.input, outputs=preds)
```

5) Étant donné que les 19 premières couches ont déjà été entrainées par le modèle VGGFace16, nous n'avons qu'à entrainer les nouvelles couches que nous avons ajoutées au modèle. Essentiellement, les nouvelles couches que nous avons ajoutées seront entrainées pour reconnaître nos images :

```
# ne pas entraîner les 19 premières couches - 0..18
for layer in model.layers[:19]:
    layer.trainable = False

# entraîner le reste des couches - à partir de 19
for layer in model.layers[19:]:
    layer.trainable = True
```

6) Compilation et entrainement du modèle :

```
# compiler le modèle à l'aide de l'optimiseur Adam et de la perte d'entropie
catégorielle
opt = Adam(lr=0.001)
model.compile(optimizer=opt, loss=keras.losses.categorical_crossentropy,
metrics=['accuracy'])
# Entrainer le modèle
hist = model.fit(train_generator, validation_data=test_generator,
validation_steps=10, batch_size=1, verbose=1, epochs=20)
```

```
    # visualiser la précision de l'entrainement et de la validation
    plt.plot(hist.history["accuracy"])
    plt.plot(hist.history['val_accuracy'])
    plt.plot(hist.history['loss'])
    plt.plot(hist.history['val_loss'])
    plt.title("model accuracy")
    plt.ylabel("Accuracy")
    plt.xlabel("Epoch")
    plt.legend(["Accuracy","Validation Accuracy","loss","Validation Loss"])
    plt.show()


    # Enregistrement du modèle
    # ========================
    # crée un fichier HDF5
    model.save('transfer_learning_trained' + '_face_cnn_model.h5')
    # Saving the Training Labels
    class_dictionary = train_generator.class_indices
    class_dictionary = {value: key for key, value in class_dictionary.items()}
    print(class_dictionary)
    # save the class dictionary to pickle
    face_label_filename = 'face-labels.pickle'
    with open(face_label_filename, 'wb') as f: pickle.dump(class_dictionary, f)
```

3.3.2.3 **L'algorithme de la reconnaissance faciale**

```
# pour tout les visages détectés
for (x, y, w, h) in faces:
```

Appliquer le traitement suivant pour tous les images des visages détectés :

```
# Convertir l'images en RGB
roi_rgb = rgb[y:y + h, x:x + w]

# Redimensionner l'image
size = (224, 224)
resized_image = cv2.resize(roi_rgb, size)
image_array = np.array(resized_image, "uint8")
img = image_array.reshape(1, image_width, image_height, 3)
img = img.astype('float32')
img /= 255
```

Ensuite, on va prédire avec la fonction `model.predict(img)`, avec l'image RGB comme paramètre, qui renvoie une liste des probabilités, où chaque classe avec sa probabilité prédite :

```
# prévoir l'image
predicted_prob = model.predict(img)
```

On récupère la classe qui a le taux prédictif le plus élevé :

```
max_predicted_prob = np.max(predicted_prob)
```

Selon le taux prédictif, on décide que le visage détecté appartient a une personne connue ou non :

```
if max_predicted_prob > 0.5:
   name = labels[predicted_prob[0].argmax()]
   print(name, str(max_predicted_prob * 100), ' %')

   # Cette vérification a pour ne pas vérifier la même personne plus d'une fois
   if name not in self.detection_list:
      self.detection_list.append(name)
      # Vérifiez les autorisations de la personne connu
      # Cette fonction sert à vérifier les autorisations de site et du temps
      # et d'envoyer des notifications au cas où la personne ne dispose pas de l'une
de ces autorisations.
      xmlrpc_connection.check_permissions(name, self.id, datetime.now())
else:
   if 'unknown' not in self.detection_list:
      self.detection_list.append('unknown')
      print('[Avertissement] Une personne inconnue a été détectée !')
      # Envoyer une notification de personne "inconnue".
      xmlrpc_connection.set_notification('unknown', self.id, 0)
```

3.3.3 Le module de l'interconnexion entre le système de détection de l'intrus et la plateforme web

Ce module assure la connexion entre le système de détection d'intrusion en temps réel et la plate-forme Web, à travers la technologie XML-RPC de odoo, car elle est considérée comme un médiateur entre eux, pour assurer la cohérence nécessaire au fonctionnement du système en tant qu'unité unifiée, et pour effectuer ce qu'on attend. Cette unité fournit les services suivants :

1. **Obtenir la liste des caméras**

   Lorsque de l'activation du système de détection d'intrusion en temps réel, il lit la liste des caméras déjà enregistrées au niveau de la plate-forme Web, où toutes les caméras du système sont gérées.

2. **Vérifiez les autorisations pour l'heure et le lieu**

   Lorsque l'identité d'une personne est reconnue par le système, celui-ci passe à l'étape de vérification des autorisations dont dispose cette personne, il doit donc contacter la plateforme web, à travers laquelle ces autorisations sont gérés.

3. **Envoi de notifications d'intrus**

   Au cas un intrus est détecté, le système envoie une notification à la plateforme web, où elle est visualisée par les opérateurs en charge de la surveillance du système.

4. **Envoi des vidéos enregistrés à la plateforme web**

   Une fois la vidéo enregistrée, le système l'envoie à la plate-forme Web, pour que les opérateurs puissent la visionner ultérieurement.

3.3.3.1 XML-RPC : qu'est-ce que c'est ?

XML-RPC est un protocole RPC (*remote procedure call*), une spécification simple et un ensemble de codes qui permettent à des processus s'exécutant dans des environnements différents de faire des appels de méthodes à travers un réseau. XML-RPC permet d'appeler une fonction sur un serveur distant à partir de n'importe quel système (Windows, Mac OS X, GNU/Linux) et avec n'importe quel langage de programmation. Le serveur est lui-même sur n'importe quel système et est programmé dans n'importe quel langage. Cela permet de fournir un **service web** utilisable par tout le monde sans restriction de système ou de langage. Les processus d'invocation à distance utilisent le protocole **HTTP** pour le transfert des données et la norme XML pour la structuration des données. XML-RPC est conçu pour permettre à des structures de données complexes d'être transmises, exécutées et renvoyées très facilement.

3.3.4 Métriques

Les métriques permettent d'évaluer le niveau de performance d'un système donné. Lors de l'utilisation de Deep Learning, nous avons différentes métriques qui nous montrent comment fonctionne notre modèle. Cependant, ces mesures peuvent confondre ce qu'elles signifient, comment elles peuvent être interprétées ou ce qu'elles sont exactement. Sachant cela, nous pourrions en déduire plus d'informations sur nos modèles.

Dans ce mémoire, nous nous concentrerons sur la perte (**loss**) et la précision (**Accuracy**). Ces deux valeurs sont essentielles à prendre en compte lors de la formation de nos modèles.

3.3.4.1 Loss

La perte (*loss*) est une valeur qui représente la somme des erreurs dans notre modèle. Il mesure à quel point (ou mal) notre modèle se porte bien. Si les erreurs sont élevées, la perte sera élevée, ce qui signifie que le modèle ne fait pas du bon travail. Sinon, plus il est bas, mieux notre modèle fonctionne. Pour calculer la perte, une fonction de perte ou de coût est utilisée. Il existe plusieurs fonctions de coût différentes à utiliser. Chacun pénalise les erreurs de différentes manières, et le problème détermine lequel est le meilleur à utiliser. L'entropie croisée et l'erreur quadratique moyenne sont les plus couramment utilisées pour les problèmes de classification et de régression, respectivement. Mais comment savoir si la perte est élevée ou faible ? Eh bien, cela dépend du problème et de la fonction de coût utilisée.

Cependant, que la perte soit élevée ou faible n'est pas l'indication la plus importante que nous puissions en tirer. Si nous traçons les résultats de perte au fil du temps, nous pouvons voir si notre modèle apprend et à quelle vitesse. En effet, dans Deep Learning, la fonction de perte est utilisée par le modèle pour apprendre. Le but du modèle est de minimiser la valeur de la perte. Cela se fait en utilisant des techniques telles que la descente de gradient, qui modifie les paramètres du modèle en utilisant les informations du résultat de la perte. Voir la perte au fil du temps peut donner des résultats intéressants pour nos modèles. Si la valeur de perte ne diminue pas, mais qu'elle oscille simplement, le modèle peut ne pas apprendre du tout. Cependant, s'il diminue dans l'ensemble d'apprentissage mais pas dans l'ensemble de validation (ou s'il diminue mais qu'il y a une différence notable), le modèle peut être surajusté. En d'autres termes, il pourrait s'agir d'un surapprentissage à partir des exemples de formation, devenant inutile dans de nouveaux exemples. Si tel est le cas, il serait intéressant d'utiliser la régularisation, d'utiliser des modèles plus simples ou, en Deep Learning, de simplement réduire le taux d'apprentissage.

3.3.4.2 Accuracy

La précision (*accuracy*) est plus simple. Il mesure la qualité de prédiction de notre modèle en comparant les prédictions du modèle avec les vraies valeurs en termes de pourcentage.

Par exemple, supposons que nous ayons un modèle de classification d'images qui détecte s'il y a ou non un chat dans l'image. Nous avons 5 images de test. Si le modèle est capable de prédire correctement s'il y a ou non un chat dans 3 des images, il en résulte une précision de 60%

3.3.4.3 La relation entre ces deux métriques

Il n'y a pas de relation entre ces deux métriques, mais si nous analysons ces deux mesures ensemble, nous pouvons en déduire plus d'informations sur le fonctionnement de notre modèle :

Avoir une précision faible mais une perte élevée signifierait que le modèle fait de grosses erreurs dans la plupart des données. Mais si la perte et la précision sont faibles, cela signifie que le modèle fait de petites erreurs dans la plupart des données. Cependant, s'ils sont tous les deux élevés, cela fait de grosses erreurs dans certaines des données. Enfin, si la précision est élevée et la perte faible, le modèle fait de petites erreurs sur certaines données seulement, ce qui serait le cas idéal.

Tableau 3 - Comment déduire des informations en analysant la précision et la perte.

	Faible perte	**Perte élevée**
Faible précision	beaucoup de petites erreurs	beaucoup de grandes erreurs
Haute précision	quelques petites erreurs	quelques grandes erreurs

3.4 Matériel et environnement de travail

3.4.1 Matériel

La présente section aborde l'aspect matériel utilisé pour la mise en œuvre et l'exécution des expérimentations de notre système de détection de l'intrus. L'utilisation du bon matériel joue un rôle crucial dans les performances, l'efficacité et la fiabilité des systèmes, puisque on a utilisé des réseaux convolutifs, qui sont connus pour être des modèles computationnellement intensifs.

Pour réaliser ce projet, on a utilisé un ensemble des équipements avec les caractéristiques suivantes :

3.4.1.1 Laptop

On a utilisé un laptop lenovo E560 pour la programmation et pour faire des tests, dont les principales caractéristiques sont les suivantes :

- Processeur : Intel(R) Core(TM) i5-6200U CPU @ 2.30GHz 2.40 GHz.
- RAM : 12 Go.
- Système d'exploitation : Windows 10 Pro.
- Carte graphique : Intel® HD Graphics 520 / 1Go de RAM.

- GPU : Video Adapter Intel Skylake-U GT2 - Integrated Graphics Controller.

3.4.1.2 **Serveur**

Pour faire des tests de l'apprentissage automatique de notre modèle, on a utilisé un Serveur DELL PowerEdge T330, avec les caractéristiques suivantes :

- Processeur : Intel® Xeon® CPU E3-1220 v5 @ 3.00GHz 3.00 GHz.
- RAM : 16 Go.
- Système d'exploitation : Windows Server 2016 Standard 64bits.

3.4.1.3 **Caméra**

On a utilisé une caméra IP de la marque Avaya le modèle CU360 pour tester notre système, avec les caractéristiques suivantes :

- Capteur : 4kp30
- Résolution vidéo : 1080p30
- Champ de vision horizontal : 102°
- Champ de vision vertical : 68°
- Zoom numérique : 3x

3.4.2 **Environnement de travail**

L'environnement de développement est un autre élément essentiel pour l'implémentation et l'exécution des systèmes. Dans ce mémoire, nous avons utilisé un environnement de développement intégré (IDE) PyCharm qui est largement utilisé par les développeurs qui utilise python, des modules de python tel que OpenCV, TensorFlow ou Keras. Ces IDE fournissent des outils puissants pour la construction, l'entraînement et l'évaluation des modèles de réseaux convolutifs. Nous détaillerons l'environnement de développement utilisé, y compris les versions spécifiques des bibliothèques logicielles et les packages supplémentaires requis.

3.4.2.1 **Python**

Le langage Python est un langage de programmation open source multi-plateformes et orienté objet. Grâce à des bibliothèques spécialisées, Python s'utilise pour de nombreuses situations comme le développement logiciel, l'analyse de données, ou la gestion d'infrastructures.

Langage de programmation interprété, Python permet l'exécution du code sur n'importe quel ordinateur. Utilisable aussi bien par des programmeurs débutants qu'experts, Python permet de créer des programmes de manière simple et rapide.

Langage principalement utilisé pour le **machine learning** et la data science, Python a fortement contribué à l'essor du big data. Grâce à ses nombreuses bibliothèques telles Panda, Bokeh, Numpy, Scipy, Scrapy, Matpotlib, Scikit-Learn ou encore TensorFlow, Python offre une grande flexibilité dans les tâches à effectuer et une grande compatibilité quelle que soit la plateforme utilisée. C'est pour ça qu'on a choisi python, la **version 3.7.9** afin de faire réaliser ce travail.

3.4.2.2 **Pycharm**

PyCharm est un environnement de développement intégré utilisé pour programmer en **Python**. Il permet l'analyse de code et contient un débogueur graphique. Il permet également la gestion des tests unitaires, l'intégration de logiciel de gestion de versions, et supporte le développement web avec Django. Développé par l'entreprise tchèque JetBrains, c'est un logiciel multiplateforme qui fonctionne sous Windows, Mac OS X et GNU/Linux.

On a utilisé la **version 2022.3.2 (*Community Edition*)** de pycharm.

3.4.2.3 **OpenCV**

Initialement développée par Intel, OpenCV (Open Computer Vision) est une bibliothèque graphique. Elle est spécialisée dans le **traitement d'images**, que ce soit pour de la photo ou de la vidéo. Sa première version est sortie en juin 2000. Elle est disponible sur la plupart des systèmes d'exploitation et existe pour les langages Python, C++ et Java.

3.4.2.4 **TensorFlow**

TensorFlow est une bibliothèque de Machine Learning, il s'agit d'une boîte à outils permettant de résoudre des problèmes mathématiques extrêmement complexes avec aisance. Elle permet aux chercheurs de développer des architectures d'apprentissage expérimentales et de les transformer en logiciels. On a travaillé avec la **version 2.11.0** de TensorFlow comme une bibliothèque de python.

3.4.2.5 **Keras**

Ecrite en Python, Keras est bibliothèque open source de prototypage rapide de modèles de deep learning. A la portée des débutants en IA, elle s'articule autour d'une API de haut niveau supportant différentes librairies de réseaux de neurones artificiels récurrents ou convolutifs, comme Tensorflow, Microsoft Cognitive Toolkit, PlaidML ou Theano. L'objectif de Keras est d'offrir un cadre pour développer au plus vite des réseaux de neurones artificiels. Initiée en 2015, cette technologie repose sur le travail de François Chollet, un développeur de Google. Elle s'inscrit dans le cadre du projet Oneiros (pour Open-ended Neuro-Electronic Intelligent Robot Operating System).

3.4.2.6 **Odoo**

Odoo, anciennement OpenERP et Tiny ERP, est initialement un progiciel de gestion intégré open-source comprenant de très nombreux modules permettant de répondre à de nombreux besoins de gestion des entreprises ou de gestion de la relation client (CRM). Le logiciel est utilisé par plus de cinq millions d'utilisateurs pour gérer leurs entreprises à travers le monde. Odoo est le système ERP open-source le plus populaire. Il existe une version community gratuite sous licence LGPLv3, et une version entreprise sous licence propriétaire Odoo Enterprise Edition License v1.05. À l'origine progiciel de gestion intégré (ERP), le logiciel s'est vu étendre ses fonctionnalités à des applications de front office (CMS, e-commerce, blogs, forums, news, événements, live chat, job offers…). On a travaillé sur la **version 12.0 *community edition*** de odoo, pour réaliser la partie de la gestion du système de la détection de l'intrus.

3.4.2.7 **La base de données des images**

Les bases de données servent à valider les travaux effectués puisqu'elles fournissent un grand ensemble de données avec lequel on peut tester les systèmes développés. Pour la validation de nos travaux, on a eu recours à la base de données **CelebFaces Attributes (CelebA) Dataset** téléchargée depuis le site Kaggle qui a servie à tester le système de reconnaissance faciale, qui est la base du système détection d'intrus en temps réel. Cette base de donnée constitue un grand ensemble de données utiles pour les projets de reconnaissance faciale. Elle contient **12191 images** de **98 personnes**. Ces images ont capturé sous différentes conditions de luminosité et d'images de fond.

3.5 Conclusion

Ce chapitre de conception expérimentale constitue le pilier central de notre travail, où nous avons détaillé les méthodologies, les algorithmes, le matériel et l'environnement de travail utilisés. Ces informations seront essentielles pour évaluer et discuter des résultats dans le chapitre suivant.

Chapitre 4 : Résultats et discussion

4.1 Introduction

Ce chapitre constitue une étape essentielle de notre travail, où nous présentons et interprétons les résultats obtenus à partir de notre recherche ou de notre développement. Il fournit une évaluation détaillée de ces résultats et ouvre la voie à une réflexion approfondie sur leur signification et leurs implications.

4.2 Résultats

Dans cette section, nous présentons les résultats majeurs de nos travaux. Premièrement, nous présentons les résultats obtenus après un test de l'entrainement de notre modèle, ensuit on va présenter les résultats d'un test complet de notre système de détection de l'intrus en temps réel.

4.2.1 Résultat de l'apprentissage

Après avoir mené une expérience d'apprentissage de notre modèle sur la base de données d'images précédemment citée préparée à cet effet, sur un serveur de type «DELL PowerEdge T330», nous avons obtenu les résultats suivants (Voir le Tableau 4 - Tableau montre les valeurs de l'accuracy et loss dans chaque Epoch de l'apprentissage.).

Tableau 4 - Tableau montre les valeurs de l'accuracy et loss dans chaque Epoch de l'apprentissage.

Epoch	loss	accuracy	val_loss	val_accuracy
1	1.4741	0.6485	0.3444	0.9
2	0.2648	0.9313	0.1253	0.9688
3	0.1813	0.9505	0.1836	0.9469
4	0.144	0.9571	0.1138	0.9812
5	0.1146	0.9676	0.0944	0.9656
6	0.1036	0.9667	0.0565	0.9875
7	0.0894	0.9728	0.0778	0.9719
8	0.0683	0.9797	0.0207	1
9	0.0555	0.9821	0.0581	0.9906
10	0.0803	0.9735	0.0446	0.9812
11	0.055	0.9822	0.0539	0.9844
12	0.0635	0.9798	0.0624	0.9844
13	0.0357	0.9865	0.0255	0.9937
14	0.0274	0.9906	0.1667	0.975
15	0.0562	0.9812	0.0422	0.9875
16	0.0269	0.991	0.0477	0.975
17	0.0184	0.9944	0.0045	1
18	0.0579	0.9837	0.069	0.9781
19	0.0315	0.9898	0.0092	1
20	0.0151	0.995	0.0072	0.9969

Nous pouvons constater que le système s'est entraîné avec 20 épochs, et nous avons atteint une précision de 0.995% et un faible taux de perte et la même chose concernant la validation(Voir Figure 38 - Un graphe montre le résultat du plot après l'apprentissage de notre modèle.).

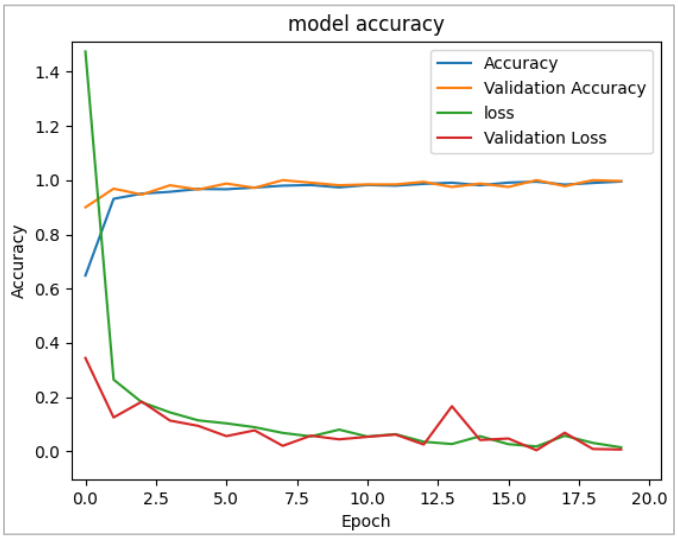


Figure 38 - Un graphe montre le résultat du plot après l'apprentissage de notre modèle.

4.2.2 Résultats de test de fonctionnement du système

Afin de tester les performances de notre système, il est nécessaire de réaliser plusieurs expériences différentes les unes des autres, dans le but de couvrir tous les cas possibles, et comme nous ne pouvons pas décrire tous ces cas dans ce mémoire, nous avons choisi un scénario spécifique qui nous suivons pour montrer les résultats résultant de chaque unité de notre système.

Pour ce faire, nous allons montrer les résultats générés par chaque module séparément.

4.2.2.1 Le démarrage du système

Lorsque le système démarre, il vérifie les paramètres de la plateforme Web afin d'obtenir la liste des caméras à surveiller, puis pour chaque caméra le système va créer un thread, pour pouvoir surveiller plusieurs caméras à la fois. Le système charge également les deux modèles, le premier pour le module de détection intelligente de mouvement et l'autre pour le module de reconnaissance faciale (voir la Figure 39 - Le démarrage du système.).

```
2023-05-26 22:01:19 [INFO] Démarrage de système de détection de l'intrus en temps réel ...
2023-05-26 22:01:19 [INFO] Obtention de la liste des cameras, veillez patienter s'il vous plais ...
2023-05-26 22:01:23 [OK] La liste des cameras est prête.
2023-05-26 22:01:23 [INFO] Création de la liste des threads ...
2023-05-26 22:01:23 [OK] Tout les threads sont démarrés.
2023-05-26 22:01:25 [INFO] Chargement du modèle pour la détection de personne ...
2023-05-26 22:01:25 [INFO] Chargement de modèle pour la reconnaissance faciale ...
{0: 'Amrouche DjamelEddine', 1: 'Amrouche Nadjib', 2: 'Amrouche Oussama', 3: 'Boufatis Mourad', 4:
 5: 'Hamidi Kamel', 6: 'Harmel Abdelkader', 7: 'Kikout SalahEddine', 8: 'Ksab Kaddour', 9: 'Moussaou
mahi Elhachmi', 11: 'Youcef Achira Abderrazak'}
2023-05-26 22:01:26 [INFO] Démarrage de la capture vidéo...
```

Figure 39 - Le démarrage du système.

4.2.2.2 Le module de la détection intelligente de mouvement

Lorsqu'une personne se présente devant l'une des caméras surveillées par le système de détection d'intrusion en temps réel, le module de détection intelligente de mouvement analyse la prise de vue de la caméra pour savoir le type d'objet qui est devant la caméra, après l'analyse le module renvoie le type d'objet détecté avec le pourcentage de la prédiction et lance enregistrement du clip vidéo Immédiatement (voir Figure 40 - Le résultat de détection de personne par le module de détection intelligente de mouvement.).

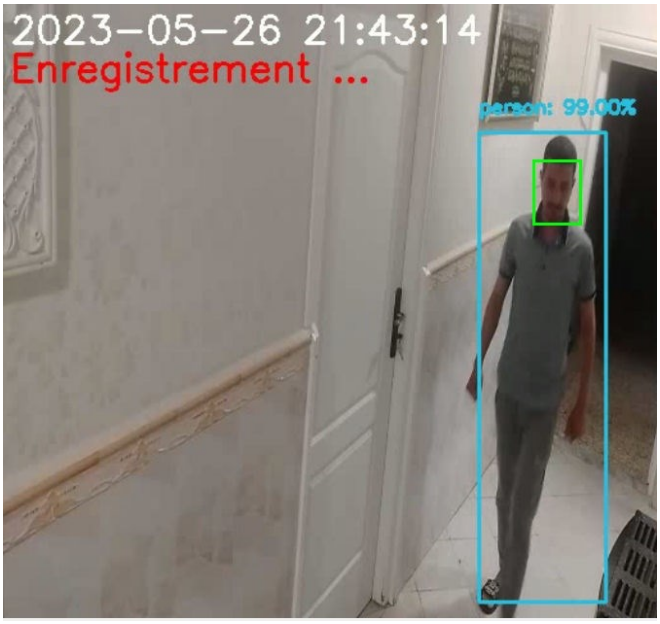


Figure 40 - Le résultat de détection de personne par le module de détection intelligente de mouvement.

Le système commence à enregistrer immédiatement après l'apparition d'un objet de tout type, (voir la Figure 41 - L'affichage dans la console au moment de détection de personne.)

```
2023-05-26 22:01:30[INFO] Personne détectée.
2023-05-26 22:01:30 [INFO] Enregistrement de vidéo ...
```

Figure 41 - L'affichage dans la console au moment de détection de personne.

4.2.2.3 Le module de la reconnaissance faciale

Si le type de l'objet détecté par le module de détection intelligente de mouvement est une "**personne**", le module de reconnaissance faciale identifie d'abord la zone du visage (voir la Figure 40 - Le résultat de détection de personne par le module de détection intelligente de mouvement.), puis la prend et l'analyse pour tenter d'identifier l'identité de cette personne.

En cas de la personne détectée n'est pas identifiée, le système envoie une notification à la plateforme web et affiche dans la console un avertissement de détection d'une personne inconnue,(voir la Figure 42 - Avertissement d'une personne inconnue.)

```
2023-05-26 22:04:09 [Avertissement] Une personne inconnue a été détectée !
```

Figure 42 - Avertissement d'une personne inconnue.

Dans le cas où le système reconnaît l'identité de la personne, il affichera son nom sur la console ainsi que le pourcentage de prédiction qui y est attaché (voir la Figure 43 - Affichage du nom de la personne détectée et le pourcentage de prédiction.).

```
2023-05-26 22:19:21[INFO] Personne détectée.
2023-05-26 22:19:21 [INFO] Enregistrement de vidéo ...
1/1 [==============================] - 0s 308ms/step
Hamidi Kamel 80.3812325000763  %
```

Figure 43 - Affichage du nom de la personne détectée et le pourcentage de prédiction.

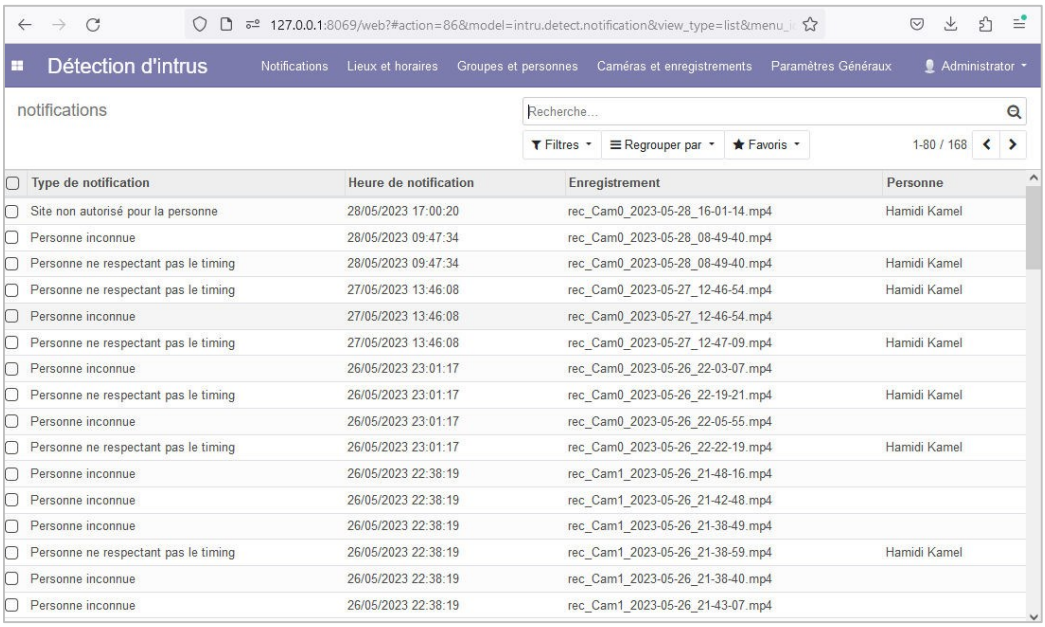
Après avoir reconnu l'identité de la personne, le système vérifie immédiatement que cette personne a le droit d'être à l'endroit où la caméra l'a capturée. Dans le cas où elle n'a pas le droit, le système envoie une notification au plate-forme Web, et un message apparaît sur la console pour Informer de l'intrusion. Mais s'il a le droit d'être à l'endroit spécifié, mais que l'heure à laquelle il a été découvert n'est pas autorisée, le système envoie également un avis indiquant que cette personne ne respecte pas l'heure autorisée.

Lorsque des personnes disparaissent de la caméra, le processus d'enregistrement vidéo s'arrête (voir la Figure 44 - Enregistrement terminé.).

Figure 44 - Enregistrement terminé.

4.2.2.4 **La plateforme WEB**

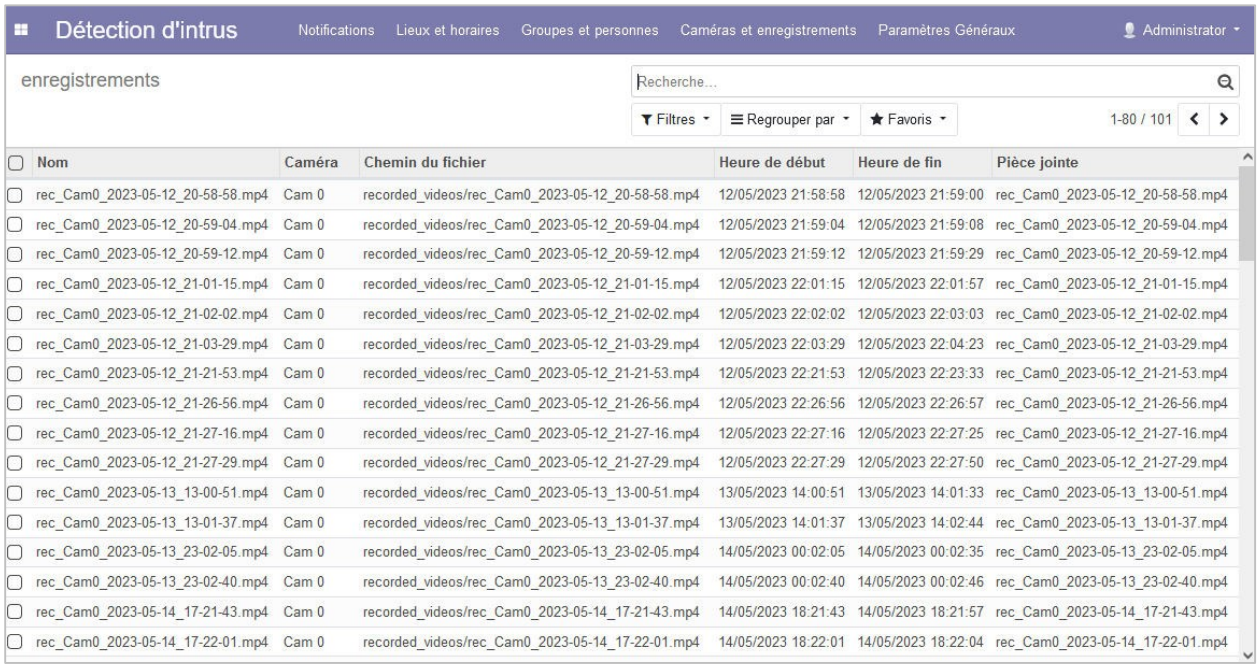
Au niveau de la plate-forme Web, nous trouvons toutes les notifications que le système a envoyées via la technologie XMLRPC, de toutes sortes, où nous trouvons à chaque notification, le temps d'envoi, l'enregistrement vidéo qui lui est lié et le nom de la personne découverte si elle est identifiée,(voir la Figure 45 - La liste des notifications dans la plateforme WEB)



Type de notification	Heure de notification	Enregistrement	Personne
Site non autorisé pour la personne	28/05/2023 17:00:20	rec_Cam0_2023-05-28_16-01-14.mp4	Hamidi Kamel
Personne inconnue	28/05/2023 09:47:34	rec_Cam0_2023-05-28_08-49-40.mp4	
Personne ne respectant pas le timing	28/05/2023 09:47:34	rec_Cam0_2023-05-28_08-49-40.mp4	Hamidi Kamel
Personne ne respectant pas le timing	27/05/2023 13:46:08	rec_Cam0_2023-05-27_12-46-54.mp4	Hamidi Kamel
Personne inconnue	27/05/2023 13:46:08	rec_Cam0_2023-05-27_12-46-54.mp4	
Personne ne respectant pas le timing	27/05/2023 13:46:08	rec_Cam0_2023-05-27_12-47-09.mp4	Hamidi Kamel
Personne inconnue	26/05/2023 23:01:17	rec_Cam0_2023-05-26_22-03-07.mp4	
Personne ne respectant pas le timing	26/05/2023 23:01:17	rec_Cam0_2023-05-26_22-19-21.mp4	Hamidi Kamel
Personne inconnue	26/05/2023 23:01:17	rec_Cam0_2023-05-26_22-05-55.mp4	
Personne ne respectant pas le timing	26/05/2023 23:01:17	rec_Cam0_2023-05-26_22-22-19.mp4	Hamidi Kamel
Personne inconnue	26/05/2023 22:38:19	rec_Cam1_2023-05-26_21-48-16.mp4	
Personne inconnue	26/05/2023 22:38:19	rec_Cam1_2023-05-26_21-42-48.mp4	
Personne inconnue	26/05/2023 22:38:19	rec_Cam1_2023-05-26_21-38-49.mp4	
Personne ne respectant pas le timing	26/05/2023 22:38:19	rec_Cam1_2023-05-26_21-38-59.mp4	Hamidi Kamel
Personne inconnue	26/05/2023 22:38:19	rec_Cam1_2023-05-26_21-38-40.mp4	
Personne inconnue	26/05/2023 22:38:19	rec_Cam1_2023-05-26_21-43-07.mp4	

Figure 45 - La liste des notifications dans la plateforme WEB

En plus des notifications, la plateforme Web contient une archive des vidéos qui ont été enregistrées lorsque des intrus ont été détectés (voir la Figure 46 - La liste des enregistrements vidéo.).



Nom	Caméra	Chemin du fichier	Heure de début	Heure de fin	Pièce jointe
rec_Cam0_2023-05-12_20-58-58.mp4	Cam 0	recorded_videos/rec_Cam0_2023-05-12_20-58-58.mp4	12/05/2023 21:58:58	12/05/2023 21:59:00	rec_Cam0_2023-05-12_20-58-58.mp4
rec_Cam0_2023-05-12_20-59-04.mp4	Cam 0	recorded_videos/rec_Cam0_2023-05-12_20-59-04.mp4	12/05/2023 21:59:04	12/05/2023 21:59:08	rec_Cam0_2023-05-12_20-59-04.mp4
rec_Cam0_2023-05-12_20-59-12.mp4	Cam 0	recorded_videos/rec_Cam0_2023-05-12_20-59-12.mp4	12/05/2023 21:59:12	12/05/2023 21:59:29	rec_Cam0_2023-05-12_20-59-12.mp4
rec_Cam0_2023-05-12_21-01-15.mp4	Cam 0	recorded_videos/rec_Cam0_2023-05-12_21-01-15.mp4	12/05/2023 22:01:15	12/05/2023 22:01:57	rec_Cam0_2023-05-12_21-01-15.mp4
rec_Cam0_2023-05-12_21-02-02.mp4	Cam 0	recorded_videos/rec_Cam0_2023-05-12_21-02-02.mp4	12/05/2023 22:02:02	12/05/2023 22:03:03	rec_Cam0_2023-05-12_21-02-02.mp4
rec_Cam0_2023-05-12_21-03-29.mp4	Cam 0	recorded_videos/rec_Cam0_2023-05-12_21-03-29.mp4	12/05/2023 22:03:29	12/05/2023 22:04:23	rec_Cam0_2023-05-12_21-03-29.mp4
rec_Cam0_2023-05-12_21-21-53.mp4	Cam 0	recorded_videos/rec_Cam0_2023-05-12_21-21-53.mp4	12/05/2023 22:21:53	12/05/2023 22:23:33	rec_Cam0_2023-05-12_21-21-53.mp4
rec_Cam0_2023-05-12_21-26-56.mp4	Cam 0	recorded_videos/rec_Cam0_2023-05-12_21-26-56.mp4	12/05/2023 22:26:56	12/05/2023 22:26:57	rec_Cam0_2023-05-12_21-26-56.mp4
rec_Cam0_2023-05-12_21-27-16.mp4	Cam 0	recorded_videos/rec_Cam0_2023-05-12_21-27-16.mp4	12/05/2023 22:27:16	12/05/2023 22:27:25	rec_Cam0_2023-05-12_21-27-16.mp4
rec_Cam0_2023-05-12_21-27-29.mp4	Cam 0	recorded_videos/rec_Cam0_2023-05-12_21-27-29.mp4	12/05/2023 22:27:29	12/05/2023 22:27:50	rec_Cam0_2023-05-12_21-27-29.mp4
rec_Cam0_2023-05-13_13-00-51.mp4	Cam 0	recorded_videos/rec_Cam0_2023-05-13_13-00-51.mp4	13/05/2023 14:00:51	13/05/2023 14:01:33	rec_Cam0_2023-05-13_13-00-51.mp4
rec_Cam0_2023-05-13_13-01-37.mp4	Cam 0	recorded_videos/rec_Cam0_2023-05-13_13-01-37.mp4	13/05/2023 14:01:37	13/05/2023 14:02:44	rec_Cam0_2023-05-13_13-01-37.mp4
rec_Cam0_2023-05-13_23-02-05.mp4	Cam 0	recorded_videos/rec_Cam0_2023-05-13_23-02-05.mp4	14/05/2023 00:02:05	14/05/2023 00:02:35	rec_Cam0_2023-05-13_23-02-05.mp4
rec_Cam0_2023-05-13_23-02-40.mp4	Cam 0	recorded_videos/rec_Cam0_2023-05-13_23-02-40.mp4	14/05/2023 00:02:40	14/05/2023 00:02:46	rec_Cam0_2023-05-13_23-02-40.mp4
rec_Cam0_2023-05-14_17-21-43.mp4	Cam 0	recorded_videos/rec_Cam0_2023-05-14_17-21-43.mp4	14/05/2023 18:21:43	14/05/2023 18:21:57	rec_Cam0_2023-05-14_17-21-43.mp4
rec_Cam0_2023-05-14_17-22-01.mp4	Cam 0	recorded_videos/rec_Cam0_2023-05-14_17-22-01.mp4	14/05/2023 18:22:01	14/05/2023 18:22:04	rec_Cam0_2023-05-14_17-22-01.mp4

Figure 46 - La liste des enregistrements vidéo.

L'application nous permet de télécharger et de visionner le clip vidéo directement via l'interface de notification, en cliquant uniquement sur le lien vidéo (voir la Figure 47 - Les détails d'une notification.).

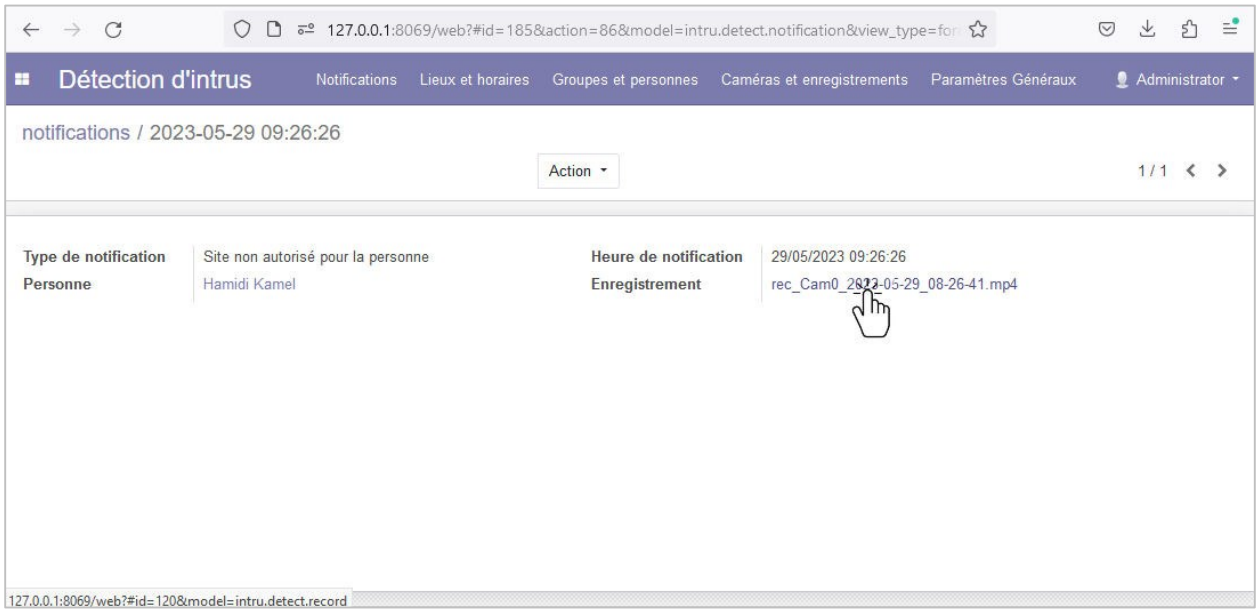


Figure 47 - Les détails d'une notification.

Dans cette section, nous avons présenté les résultats de l'expérience d'apprentissage du modèle sur une base de données d'images, qui comprend 12191 images de 98 personnes différentes, et nous avons également présenté les résultats résultant de chaque module de notre système, chacune séparément, après avoir suivi un scénario de détection un intrus par le système, avec une indication Pour les résultats au niveau de la plateforme web. Dans la section suivante, nous discuterons ces résultats et montrerons les limites du système.

4.3 Discussion

Dans cette section, nous interprétons les résultats exposés dans la section précédente. Cette interprétation nous a permis de mieux évaluer le niveau de réalisation et d'aboutissement des objectifs que nous nous sommes initialement fixés.

4.3.1 Interprétation des résultats obtenus

4.3.1.1 L'apprentissage

La forme et la dynamique d'une courbe d'apprentissage peuvent être utilisées pour diagnostiquer le comportement d'un modèle d'apprentissage automatique, et à leur tour, peut-être suggérer le type de modifications de configuration qui peuvent être apportées pour améliorer l'apprentissage et/ou les performances. Il existe trois dynamiques communes que nous somme susceptible d'observer dans les courbes d'apprentissage ; ils sont :

1. **Insuffisant (*Underfit*)**

   Le sous-apprentissage fait référence à un modèle qui ne peut pas apprendre l'ensemble de données d'apprentissage. Il peut afficher une ligne plate ou des valeurs bruyantes de perte relativement élevée, indiquant que le modèle n'a pas du tout appris l'ensemble de données d'apprentissage (voir la Figure 49 - Exemple de courbe d'apprentissage de formation montrant un modèle sous-ajusté qui n'a pas une capacité suffisante).

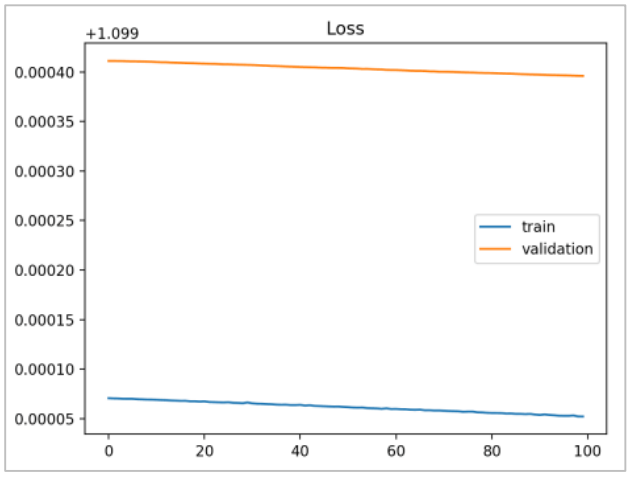


Figure 49 - Exemple de courbe d'apprentissage de formation montrant un modèle sous-ajusté qui n'a pas une capacité suffisante

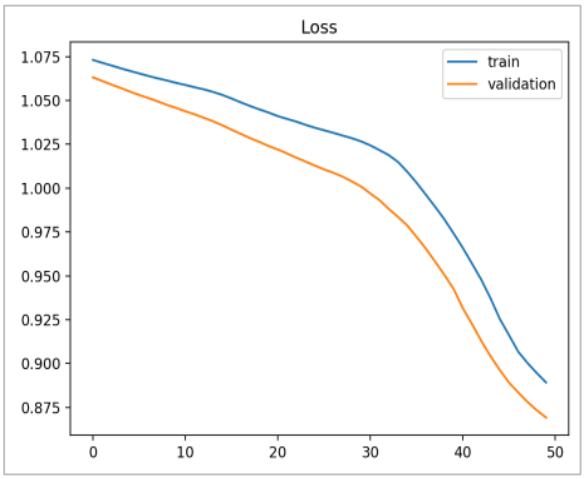


Figure 48 - Exemple de courbe d'apprentissage de la formation montrant un modèle sous-ajusté nécessitant une formation supplémentaire

Un modèle sous-ajusté peut également être identifié par une perte d'entraînement qui diminue et continue de diminuer à la fin du graphique. Cela indique que le modèle est capable d'apprendre davantage et d'éventuelles améliorations supplémentaires et que le processus de formation a été interrompu prématurément (voir la Figure 48 - Exemple de courbe d'apprentissage de la formation montrant un modèle sous-ajusté nécessitant une formation supplémentaire).

Un tracé des courbes d'apprentissage montre un sous-ajustement si :

- ✓ La perte de formation reste stable quelle que soit la formation.
- ✓ La perte d'entraînement continue de diminuer jusqu'à la fin de l'entraînement.

2. Sur-apprentissage (*Overfit*)

Le Sur-apprentissage fait référence à un modèle qui a trop bien appris l'ensemble de données d'apprentissage, y compris le bruit statistique ou les fluctuations aléatoires dans l'ensemble de données d'apprentissage.

Le problème avec le sur-apprentissage est que plus le modèle devient spécialisé dans les données d'apprentissage, moins il est capable de généraliser à de nouvelles données, ce qui entraîne une augmentation de l'erreur de généralisation. Cette augmentation de l'erreur de généralisation peut être mesurée par les performances du modèle sur le jeu de données de validation.

Cela se produit souvent si le modèle a plus de capacité que nécessaire pour le problème et, par conséquent, trop de flexibilité. Cela peut également se produire si le modèle est entraîné trop longtemps.

Un tracé des courbes d'apprentissage montre un sur-apprentissage si :

- ✓ L'intrigue de la perte d'entraînement continue de diminuer avec l'expérience.
- ✓ Le tracé de la perte de validation diminue jusqu'à un certain point et recommence à augmenter.

Le point d'inflexion de la perte de validation peut être le point auquel la formation pourrait être interrompue, car l'expérience après ce point montre la dynamique du sur-apprentissage.

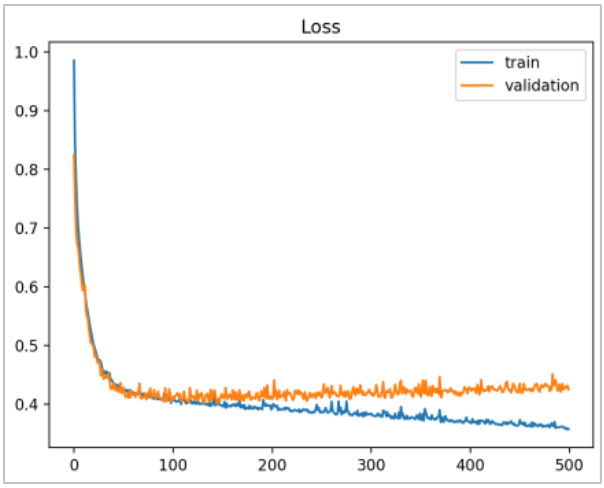


Figure 50 - Exemple de courbes d'apprentissage d'entraînement et de validation montrant un modèle de sur-apprentissage.

3. Bon ajustement (*Good Fit*)

Un bon ajustement est l'objectif de l'algorithme d'apprentissage et existe entre un modèle de sur ajustement et de sous-ajustement. Un bon ajustement est identifié par une perte d'entraînement et de validation qui diminue jusqu'à un point de stabilité avec un écart minimal entre les deux valeurs de perte finales. La perte du modèle sera presque toujours plus faible sur le jeu de données

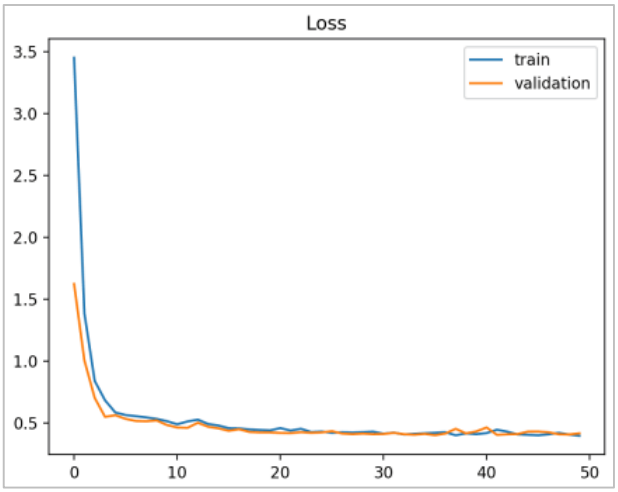


Figure 51 - Exemple de courbes d'apprentissage d'entraînement et de validation montrant un bon ajustement.

d'apprentissage que sur le jeu de données de validation. Cela signifie que nous devrions nous attendre à un certain écart entre les courbes d'apprentissage de perte de train et de validation. Cet écart est appelé « écart de généralisation ».

Tracé des courbes d'apprentissage montre un bon ajustement si :

- ✓ Le tracé de la perte d'entraînement diminue jusqu'à un point de stabilité.
- ✓ Le tracé de la perte de validation diminue jusqu'à un point de stabilité et présente un petit écart avec la perte d'apprentissage.

L'entraînement continu d'un bon ajustement entraînera probablement un surentraînement.

Après l'observation des résultats de test de l'apprentissage de notre modèle en les comparant par les trois cas que nous avons mentionné précédemment, on a constaté que le graphe de résultat de notre teste d'apprentissage, illustre un cas de bon ajustement (voir la Figure 38 - Un graphe montre le résultat du plot après l'apprentissage de notre modèle.). Où en observant le courbe de *loss* et le courbe de *Validation loss*, on constate que le courbe de *loss* 'entraînement diminue jusqu'à un point de stabilité, et le courbe de Validation *loss* diminue jusqu'à un point de stabilité et présente un petit écart avec la perte d'apprentissage.

Nous avons également obtenu un pourcentage élevé de précision, qui a atteint **0,995 %**, ce qui indique un taux élevé d'obtention de prédictions correctes.

4.3.1.2 **Le fonctionnement du système**

Pour mieux apprécier les résultats obtenus, rappelons quels sont les objectifs fixés pour atteindre notre but. Les objectifs fixés pour le système de détection d'intrus sont :

1. arriver à détecter un mouvement dans le flux d'image envoyé par la caméra, et reconnaitre le type de l'objet détecté et lancer l'enregistrement d'une vidéo ;

2. procéder à la reconnaissance faciale des visages détectés et renvoyer l'identité de la personne détectée si c'est possible ;

3. Envoyer des notifications à la plateforme WEB.

4. envoyer la vidéo au serveur de la plateforme WEB.

L'évaluation de chaque objectif revient à interpréter les résultats de chaque sous-système présenté au chapitre précédent. L'objectif **(1)** a été pleinement atteint au vu des résultats. Le système a bien interagi avec la présence de personnes devant les caméras, car il détecte rapidement la présence de personnes, dans différentes positions et formes. Le système détecte également efficacement la présence d'autres objets, comme la présence de chiens, chats, chevaux...etc. Il a également enregistré des clips vidéo dans tous les cas où l'un des objets concernés par la surveillance a été effectivement détecté par le système, et le système a immédiatement arrêté l'enregistrement lorsque les objets ont disparu de l'avant de la caméra.

L'objectif **(2)** a été pleinement atteint, l'identité des personnes découvertes a été identifiée avec succès, avec un faible pourcentage d'erreurs causées par la nature de l'environnement, telles que la qualité de l'éclairage ou l'angle sous lequel l'image a été prise, et ces erreurs peuvent parfois être expliqué par le manque de données sur lesquelles le modèle a été formé.

Les objectifs **(3 et 4)** ont été également pleinement atteints puisque les résultats montrent que les notifications et les vidéos enregistrées ont été envoyés avec succès à la plateforme WEB.

4.3.2 Les Limites

Le système développé est utilisable dans les environnements de type semi contrôlé c'est-à-dire dans le scénario où les caméras sont placées dans les corridors, les couloirs et les individus sont obligés de prendre par des points de contrôle spécifiques afin qu'une bonne image de leurs visages soit obtenue.

De ce fait, il ne peut être déployé dans les environnements non contrôlés où le visage des individus peut ne pas être capturé ou dans les environnements où les visages sont vus de profil ou assez éloignés des caméras. Les images de la figure suivante présentent ces cas où le système se révélera inefficace.

Aussi, les erreurs d'identification : Le système peut commettre des erreurs d'identification, tant en termes de fausses positives (identification incorrecte) que de fausses négatives (non-identification d'une personne connue).

4.4 Conclusion

Ce chapitre a été essentiel pour évaluer les réalisations de notre projet et en tirer des conclusions significatives. Les résultats ont été présentés de manière claire et analysés en profondeur, en tenant compte du contexte existant. Cette analyse a permis de comprendre les implications de nos résultats et d'identifier des pistes pour de futurs travaux. En somme, ce chapitre constitue une étape cruciale dans la progression de notre recherche ou de notre développement.

Conclusion générale

La création d'un système de détection de l'intrus en exploitant la vidéosurveillance constitue une avancée majeure dans le domaine de la sécurité. Ce mémoire a permis d'explorer les différentes techniques et méthodes disponibles pour détecter de manière efficace les intrusions à partir des flux vidéo.

L'étude approfondie des algorithmes de vision par ordinateur et d'apprentissage automatique a démontré leur potentiel dans l'amélioration des capacités de détection et de reconnaissance des objets. En utilisant des approches telles que la détection d'objets et la reconnaissance faciale, il est possible de détecter rapidement les intrusions et de prendre des mesures préventives.

Cependant, il est important de noter que la mise en œuvre d'un tel système ne se limite pas seulement à l'aspect technologique. Des considérations telles que la protection de la vie privée, la conformité aux réglementations en matière de données personnelles et l'optimisation des ressources sont également cruciales pour assurer le bon fonctionnement et l'acceptation de ce type de système.

Pour atténuer les erreurs d'identification dans les systèmes de reconnaissance faciale, voici quelques recommandations :

Améliorer la diversité des ensembles de données : Les erreurs d'identification peuvent être causées par des ensembles de données d'entraînement peu diversifiés. Il est essentiel de veiller à ce que les ensembles de données utilisés pour former les systèmes de reconnaissance faciale comprennent des visages représentatifs de différentes ethnies, sexes, âges et autres caractéristiques démographiques. Cela contribuera à réduire les biais et à améliorer la précision pour tous les groupes.

Contrôle de la qualité des données : Il est important de s'assurer que les ensembles de données utilisés pour l'entraînement des systèmes de reconnaissance faciale sont de haute qualité et représentatifs de la population cible. Cela peut nécessiter des efforts supplémentaires pour collecter et vérifier les données, ainsi que pour éliminer les éventuels biais ou erreurs.

Utilisation de seuils ajustables : Les systèmes de reconnaissance faciale peuvent inclure des seuils ajustables pour équilibrer les taux de fausses positives et de fausses négatives. Cela permet de régler la sensibilité du système en fonction des besoins spécifiques de l'application.

Validation humaine : Lorsque des décisions critiques sont prises sur la base des résultats de la reconnaissance faciale, il est recommandé d'inclure une validation humaine supplémentaire pour confirmer l'identification.

En conclusion, la création d'un système de détection de l'intrus en exploitant la vidéosurveillance représente une avancée prometteuse pour renforcer la sécurité des espaces surveillés. Avec une conception minutieuse, une implémentation adéquate et une gestion efficace des données, ce système peut jouer un rôle essentiel dans la prévention et la dissuasion des actes criminels, contribuant ainsi à un environnement plus sûr et plus sécurisé.

Bibliographie

[1] M. Dahmane, «Système de Vidéosurveillance et de Monitoring,» Université de Montréal, Montréal, 2004.

[2] W. Dornberger, V-2. Hurst/Scientific Book, 1954.

[3] B. Yesil, "Watching ourselves," *Cultural Studies,* vol. 20, no. 4-5, pp. 400-416, 2006.

[4] J.-P. Langellier, «La grande-bretagne se transforme en une "société sous surveillance"[archive],» *Le Monde,* november 2006.

[5] J. Durand, "Un jeu de dupes entre État et élus," *Libération,* december 2009.

[6] S. BENALIA, «Lutte contre la criminalité,» *L'EXPRESSION,* 2023.

[7] Y.Asma et L.Abderrahim, «Mise au point d'une application de télésurveillance,» Université Abou Berk Belkaid, 18 juin 2017.

[8] R. Armand et N. Ngaamou, «Etude et mise en place d'un système de vidéosurveillance,» Institut supérieur des technologies et du design industriel - Cameroun, 2011. [En ligne]. Available: https://www.memoireonline.com/01/13/6765/m_Etude-et-mise-en-place-d-un-systeme-de-videosurveillance-Cas-de-l-immeuble-Folepe--Bali.html.

[9] Heisele and Wöhler, "Motion-based recognition of pedestrians," in *Proceedings of the 14th International Conference on Pattern Recognition*, Brisbane, Australie, 16-20 août 1998.

[10] Franke and Kutzbach, "Fast stereo based object detection for Stop & Go traffic," in *Proceedings of the IEEE Intelligent Vehicles Symposium*, Tokyo, 1996.

[11] Viola, Jones and Snow, "Detecting Pedestrians using Patterns of Motion and Appearance," *IJCV,* vol. 63, no. 2, pp. 153-161, 2005.

[12] D. Navneet and T. Bill, "Histograms of oriented gradients for human," in *Conference on Computer Vision and Pattern Recognition*, 2005 [archive du 23 novembre 2008] [PDF], sur acemedia.org.

[13] B. Marouane, «Détection et suivi de personnes par vision omnidirectionnelle: approches 2D et 3D,» Université d'Evry Val d'Essonne, 2018.

[14] D. Geronimo, A. Lopez, D. Ponsa and A. Sappa, "Haar wavelets and edge orientation histograms for on–board pedestrian detection," *Pattern Recognition and Image Analysis,* p. 418–425, 2007.

[15] K. Levi and Y. Weiss, "Learning object detection from a small number of examples : the importance of good feature," in *Computer Vision and Pattern Recognition*, 2004.

[16] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE Computer Society Conference on, volume 1, pages I–I. IEEE*, 2001.

[17] C. P. Papageorgiou, M. Oren et T. Poggio, «A general framework for object detection,» chez *Computer vision, sixth international conference on, pages 555–562*, 1998.

[18] R. Lienhart and J. Maydt, "An extended set of haar like features for rapid object detection," in *Image Processing. 2002. Proceedings. 2002 International Conference on, volume 1, pages I–I. IEEE*, 2002.

[19] L. Wang and D.-C. He, "Texture classification using texture spectrum," *Pattern Recognition,* vol. 23, no. 8, p. 05–910., 1990.

[20] D.-C. He et L. Wang, «Texture unit, texture spectrum, and texture analysis,» *IEEE transactions on Geoscience and Remote Sensing,* vol. 28, n° %14, p. 509–512, 1990.

[21] T. Ojala, M. Pietikainen et a. M. T. , «Multiresolution gray-scale and rotation invariant texture classification with local binary patterns,» *IEEE Transactions on pattern analysis and machine intelligence,* vol. 24, n° %17, p. 971–987, 2002.

[22] S. Belongie et J. Malik, «Matching with shape contexts,» *Proceedings Workshop on Content-based Access of Image and Video Libraries,* p. 20–26, 2000.

[23] B. Leibe, E. Seemann and B. Schiele, "Pedestrian detection in crowded scenes," *Computer Vision and Pattern Recognition,* vol. 1, p. 878–885, 2005.

[24] S. Belongie, J. Malik and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE transactions on pattern analysis and,* vol. 24, no. 4, p. 509–522, 2002.

[25] G. Mori, S. Belongie et J. and Malik, «Efficient shape matching using shape contexts,» *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 27, n° %111, p. 1832–1837, 2005.

[26] N. Dalal et B. Triggs, «Histograms of oriented gradients for human detection,» *Computer Vision and Pattern Recognition,* vol. 1, p. 886–893, 2005.

[27] C. Cortes et V. Vapnik, «Support-vector networks,» *Machine learning,* vol. 20, n° %13, p. 273–297, 1995.

[28] C. Papageorgiou et T. Poggio, «A trainable system for object detection,» *International Journal of Computer Vision,* vol. 38, n° %11, p. 15–33, 2000.

[29] S. Kang, H. Byun et S.-W. Lee, «Real-time pedestrian detection using support vector machines,» *Pattern Recognition with Support Vector Machines,* p. 273–281, 2002.

[30] A. Shashua, Y. Gdalyahu and G. and Hayun, "Pedestrian detection for driving assistance systems : Single-frame classification and system level performance," *In Intelligent Vehicles Symposium,* pp. 1-6, 2004.

[31] J. Zhu, H. Zou, S. Rosset, T. Hastie et al, «Multi-class adaboost,» *Statistics and its Interface,* vol. 2, n° %13, p. 349–360, 2009.

[32] J. Vesanto et E. Alhoniemi, «Clustering of the self-organizing map,» *IEEE Transactions on neural networks,* vol. 11, n° %13, p. 586–600, 2000.

[33] C. D. David Rousseau, «Introduction à l'apprentissage profond (deep learning) de l'intelligence artificielle,» CultureSciences Physique, octobre 2021. [En ligne]. Available: https://culturesciencesphysique.ens-lyon.fr/ressource/IA-apprentissage-Rousseau.xml.

[34] Alom and autres, "The history began from alexnet : A comprehensive survey on deep learning approaches," in *arXiv preprint arXiv :1803.01164*, 2018.

[35] K. Arulkumaran, M. P. Deisenroth, M. Brundage and A. A. Bharath, A brief survey of deep reinforcement learning, arXiv preprint arXiv :1708.05866, 2017.

[36] M. Hon and N. M. Khan, "Towards alzheimer's disease classification through transfer learning," in *International conference on bioinformatics and biomedicine (BIBM)*, 2017.

[37] A. Khan, A. Sohail, U. Zahoora and A. S. Qureshi, A survey of the recent architectures of deep convolutional neural networks, Artificial Intelligence Review, 2020, p. 1–62.

[38] K. Jarrett, K. Kavukcuoglu, M. Ranzato and Y. LeCun, "What is the best multi-stage architecture for object recognition ?," in *12th international conference on computer vision*, 2009.

[39] S. Indolia, A. K. Goswami, S. Mishra et P. Asopa, «Conceptual understanding of convolutional neural network-a deep learning approach,» *Procedia computer science,* vol. 132, p. 679–688, 2018.

[40] I. AKCHA et A. AMMARI, «Développement d'un système de reconnaissance faciale,» Mémoire de Master, Blida, 2020.

[41] A. Ghali, «Amélioration de la reconnaissance par le visage,» Thèse de Magister en informatique,Université Mohamed Boudiaf, Oran, 2015.

[42] S. Djedi, «Etude comparative de PCA et KPCA associées au SVM en biométrie,» Thèse de Doctorat en informatique, Université Mohamed Khider, Biskra, 2012.

[43] S. Boudjellal, «Détection et identification de personne par méthode biométrique,» Thèse de Magister, Université Mouloud Mammeri, 2017.

[44] L. HAMOUDI, «Application de techniques d'apprentissage,» Lille, 2011.

[45] K. Y. Ghoulia. B, «Etude comparative d'ensemble des descripteurs de texture pour la reconnaissance de visages,» Thèse de master en génie électrique, Université Kasdi Merbah, Ouargla, 2017.

[46] Behrman, Moscovitch et Winocur, «Facing the issue : New research shows that the brain process faces and objects in separate brain systems,» *Journal of Cognitive Neuroscience,* 1997.

[47] Jiang, Rosen and others, "Evaluation of a shape-based model of human face discrimination using fmri and bechavioral techniqueq," *Neuron,* vol. 50, no. 1, pp. 72-159, 2006.

[48] Pentland, Turk and A. P., "Face Recognition using Eigenfaces," *Proc. IEEE,* pp. 586-591, 1991.

[49] Belhumeur, Hespanha et Kriegman, «Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection,» IEEE Trans. PAMI, 1997.

[50] Kak, Martinez et A. C., «Pca versus lda,» *IEEE Trans. PAMI,* vol. 23, n° %12, pp. 228-233, 2001.

[51] [En ligne]. Available: http://www.univ-usto.dz/theses_en_ligne/doc_num.php. [Accès le 18 03 2022].

[52] M. D. Kelly, "Visual identification of people by computer," PhD thesis, Stanford, CA, USA, 1971.

[53] T. KANADE, Computer recognition of human faces, Birkhauser, Basel, Switzerland, and Stuttgart, Germany, 1973.

[54] Brunelli et Poggio, «Face Recognition : Features versus Template,» *IEEE Trans. Pattern Anal. Machine Intell,* vol. 15, pp. 1042-1052, 1993.

[55] Rowley, Baluja et Kanade, «Neural Network-based face detection,» *IEEE Transaction on Pattern Analysis and Machine Intelligence,* vol. 20, n° %11, pp. 23-38, 1998.

[56] Manjunath, Chellapa et Andmalsburg, «A feature based approach to face recognition,» *In Proceedings, IEEE Conference on Computer Vision and Pattern Recognition,* pp. 373-378, 1992.

[57] Wiskott, Fellous, Kruger et v. d. Malsburg, «IEEE Trans. Face recog- nition by elastic bunch graph matching,» vol. 19, n° %17, pp. 775-779, 1997.

[58] A. Martinez, Recognizing imprecisely localized, partially occluded and expression variant faces from a single sample per class, vol. 24, IEEE Trans. Patt. Anal., 2002, pp. 748-763.

[59] A. Y. Takeo Kanade, "Multi-Subregion Based Probabilistic Approach To- ward Pose-Invariant Face Recognition," in *IEEE International Symposium on Computational Intelligence in Robotics and Automation*, 2003.

[60] A. P. Moghaddam, «Probabilistic Visual Learning for Object Representation,» *IEEE Trans. on Pattern Analysis and Machine Intelligence,* vol. 19, n° %17, pp. 696-710, 1997.

[61] BARKI et Hicham, «DETECTION ET RECONNAISSANCE DE VISAGE,» UNIVERSITE FERHAT ABBAS, SETIF.

[62] Jones et al, «Reconnaissance faciale dans la sécurité,» 2018.

[63] Smith et coll, «Reconnaissance faciale pour l'authentification,» 2019.

[64] Chen et al, «Reconnaissance faciale dans la réalité virtuelle,» 2020.

[65] Zhang et al, «Biais et équité dans la reconnaissance faciale,» 2021.

[66] Brown et al, «Vie privée et protection des données dans la reconnaissance faciale,» 2022.

[67] Daniel, «VGG : en quoi consiste ce modèle ? Daniel vous dit tout !,» OMNES EDUCATION, 27 Avril 2023. [En ligne]. Available: https://datascientest.com/quest-ce-que-le-modele-vgg.

[68] Wiesel and Hubel, "Functional architecture of macaque monkey," *Proc. Royal Soc. B (London),* vol. 198, pp. 1-59, 1978.