الجمهورية الجزائرية الديمقراطية الشعبية
République algérienne démocratique et populaire
وزارة التعليم العالي والبحث العلمي
Ministère de l'enseignement supérieur et de la recherche scientifique
جامعة عين تموشنت بلحاج بوشعيب
Université –Ain Temouchent– Belhadj Bouchaib
Faculté des Sciences et de la Technologie
Département d'Électronique et des Télécommunications

Projet de Fin d'Etudes
Pour l'obtention du diplôme de Master en :
Domaine : Sciences et Technologie
Filière : Électronique
Spécialité : Instrumentation
Thème

Conception et Réalisation d'un Système de Vidéo-Surveillance dans une Maison Intelligente

Présenté Par :
Melle FACI Ikhlase Radjaa

Devant le jury composé de :

Dr BENZINA Amina	MCB	UAT.B.B (Ain Temouchent)	Présidente
Mme BOUTKHIL Malika	MAA	UAT.B.B (Ain Temouchent)	Examinatrice
Dr BENTAIEB Samia	MCB	UAT.B.B (Ain Temouchent)	Encadrante

Année Universitaire 2022/2023

C'est avec une profonde gratitude et un amour infini que je souhaite dédier ce mémoire à ma chère maman. Tout au long de ma vie, tu as été ma source inépuisable de soutien et d'inspiration, et ton bienveillance constante a été le moteur de ma réussite académique.
Ton encouragements sans faille et ton amour inconditionnel m'ont poussée à donner le meilleur de moi-même et à poursuivre mes rêves. Tu as toujours cru en moi, même lorsque j'ai douté de mes propres capacités, et tu as été là pour me rappeler que je suis capable de grandes choses. Tu m'a donné la force et la confiance nécessaires pour surmonter les obstacles et persévérer, même lorsque les défis semblaient insurmontables.
Merci du fond du cœur pour tout ce que tu as fait et continues de faire pour moi et pour avoir été la meilleure mère qu'il soit. Je t'aime plus que les mots ne pourront jamais l'exprimer. Cher père,Je souhaite te témoigner mon amour profond et ma gratitude infinie en t'offrant ce mémoire. Tu as toujours été là pour célébrer mes succès et me réconforter dans les moments de doute.Il est un humble hommage à l'homme exceptionnel que tu es et à l'impact positif immense que tu as eu sur ma vie .

Ikhlase Radjaa

Remerciements

Je tiens tout d'abord à exprimer ma gratitude envers le Miséricordieux, **ALLAH**, pour m'avoir accordé la santé, la volonté et la possibilité de réussir dans mes études, ainsi que le courage nécessaire pour mener à bien ce travail.

je souhaite adresser mes remerciements et ma profonde gratitude à mon encadrante, le Dr BENTAIEB Samia , pour son précieux encadrement, sa disponibilité et ses conseils avisés tout au long de ce processus. Son expertise, ses suggestions constructives et son engagement ont été essentiels pour orienter mes recherches, m'aider à progresser et me guider vers la réussite. Son expérience dans le domaine de la recherche et de l'enseignement m'a permis d'explorer le vaste univers de la recherche scientifique en matière d'intelligence artificielle et de systèmes embarqués.

Je tiens à remercier chaleureusement le Dr BENZINA A. d'avoir présidé le jury et Mme BOUTKHIL M. d'avoir accepté d'évaluer ce travail.

Mes remerciements vont également à Mme BOUZI Wissem, Mr HAMAIDA Habib, Mr BOUMEDINE Ahmed Yassine et Mme IKNI Hind pour leur aide inestimable .

Enfin, je souhaite exprimer ma sincère gratitude envers mes parents, mes frères, ma soeur et ma famille pour leur soutien inconditionnel et leur encouragement constant tout au long de cette aventure académique. Leur présence et leur soutien moral ont été d'une importance capitale. Enfin, je tiens à remercier tous ceux qui ont contribué à la réalisation de ce projet.

Table des matières

Table des figures

Liste des tableaux

Abréviations

AI: **A**rtificielle **I**ntelligence
ANN: **A**rtifical **N**eural **N**etworks
BP: **B**ack **P**ropagation
CNN: **C**onvolution **N**eural **N**etworks
CPU: **C**entral **P**rocessing **U**nit DL: **D**eep **L**earning
DPI: **D**ots **P**er **I**nch
FPS : **F**rames **P**er **S**econd
GIF : **G**raphics **I**nterchange **F**ormat
GPIO: **G**eneral **P**urpose **I**nput/**O**utput
GPU: **G**raphics **P**rocessing **U**nit
HD : **H**igh **D**efinition
HDMI: **M**ultimedia **I**nterface
JPEG : **J**oint **P**hotographic **E**xperts **G**roup
LCD: **L**iquid **C**rystal **D**isplay
ML: **M**achine **L**earning
OS: **O**perating **S**ystem
PCA: **Prinical Analysis**
ReLU: **R**ectified **L**inear **U**nits
ROI: **R**egionl **O**Of **I**nterest
RNN: **R**ecurrent **N**eural **N**etworks
SBC: **S**ingle **B**oard **C**omputer PNG : **P**ortable **N**etwork **G**raphics
SD : **S**tandard **D**efinition
SGD : **S**tochastic **G**radient **D**escent SOC: **S**ystem **O**n **C**hip
SVM: **S**upport **V**ector **M**achine

TIFF :**T**agged **I**mage **F**ile **F**ormat
USB:**U**niversal **S**erial **B**us
VGA: **V**ideo **G**raphics **A**rray
VNC: **V**irtual **N**etwork **C**omputing

Résumé

De nombreux domaines importants, tels que la sécurité, la biométrie, la surveillance et l'interface humain-ordinateur, dépendent fortement des technologies de détection faciale et de reconnaissance. Pour étendre davantage ce sujet, nous présentons une étude de recherche qui se concentre sur le développement d'une méthode basée sur Raspberry Pi pour la détection et la reconnaissance faciales automatiques. Notre approche utilise des techniques d'apprentissage profond, en particulier le modèle SSD (Single Shot MultiBox Detector), qui est réputé pour son efficacité dans la gestion des architectures profondes. Le modèle SSD sert de colonne vertébrale pour la détection faciale dans notre système.
En tirant parti d'un modèle pré-entraîné, nous sommes en mesure d'extraire des caractéristiques robustes des images faciales, permettant des représentations faciales précises et discriminatoires à des fins de reconnaissance. La méthodologie proposée vise à créer un système intégré et léger pour l'identification faciale dans les images et les vidéos, en particulier dans le contexte de la surveillance de la maison intelligente.
L'intégration de notre système basé sur Raspberry Pi permet un déploiement efficace dans divers environnements, y compris les maisons intelligentes, où la surveillance et l'identification des individus sont cruciales à des fins de sécurité. Grâce à la puissance des techniques d'apprentissage profond et à l'utilisation de modèles pré-entraînés, notre système atteint une précision élevée dans la détection et la reconnaissance des visages, fournissant une solution fiable et efficace pour l'identification faciale dans des scénarios en temps réel.
Mots clés: Détection de visage, reconnaissance faciale, Deep Learning, Transfer Learning, MobilenetV2, SSD. Raspberry Pi.

ملخص

تعتمد مجالات هامة كثيرة، مثل الأمن الاستدلال البيولوجي، والمراقبة، والربط بين البشر والحواسيب، اعتمادا كبيرا على تكنولوجيات الكشف عن الوجه والاعتراف به. ولزيادة التوسع في هذا الموضوع، نقدم دراسة بحثية تركز على تطوير طريقة تعتمد على توت البري للكشف عن الوجه والتعرف عليه تلقائياً. ويستخدم نهجنا تقنيات التعلم العميق، وعلى وجه التحديد نموذج SSD الذي يُشَهَر بفعاليته في التعامل مع العمارات العميقة. يعمل نموذج SSD بمثابة العمود الفقري لكشف الوجه في نظامنا.

ومن خلال الاستفادة من نموذج سابق للتدريب، يمكننا أن نستخرج سمات قوية من صور الوجه، مما يمكّن من عرض الوجوه بدقة وتمييز لأغراض الاعتراف. وتهدف المنهجية المقترحة إلى

إنشاء نظام مدمج وخفيف الوزن للتعرف على الوجه في الصور وأشرطة الفيديو على السواء، ولا سيما في سياق الرصد الذكي للمنازل.

ويتيح تكامل نظامنا القائم على توت التبري نشرا فعالا في بيئات مختلفة، بما في ذلك المنازل الذكية، حيث يتسم رصد الأفراد وتحديد هويتهم بأهمية حاسمة للأغراض الأمنية. وبقوة تقنيات التعلم العميق واستخدام النماذج السابقة للتدريب، يحقق نظامنا دقة عالية في الكشف عن الوجوه والتعرف عليها، ويوفر حلا موثوقا وفعالا لتحديد الوجه في سيناريوهات الوقت الحقيقي.

الكلمات الرئيسية : لكشف عن الوجه، التعرف على الوجه، التعلم العميق، التعليم عن طريق النقل، $MobilenetV2$ ، SSD ، $RaspberryPi$.

Abstract

Many important areas, such as security, biometrics, surveillance, and human-computer interface, heavily rely on facial detection and recognition technologies. To further expand on this topic, we present a research study that focuses on developing a Raspberry Pi-based method for automatic face detection and recognition. Our approach utilizes deep learning techniques, specifically the SSD (Single Shot MultiBox Detector) model, which is renowned for its effectiveness in handling deep architectures. The SSD model serves as the backbone for face detection in our system.

By leveraging a pre-trained model, we are able to extract robust features from facial images, enabling accurate and discriminative face representations for recognition purposes. The proposed methodology aims to create an embedded, lightweight system for face identification in both images and videos, particularly in the context of smart home monitoring. The integration of our Raspberry Pi-based system allows for efficient deployment in various environments, including smart homes, where monitoring and identification of individuals are crucial for security purposes. With the power of deep learning techniques and the use of pre-trained models, our system achieves high accuracy in detecting and recognizing faces, providing a reliable and effective solution for facial identification in real-time scenarios.

Keywords:Face detection, face recognition, Deep Learning, Transfer Learning,MobilenetV2, SSD Raspberry Pi.

Introduction générale

Contexte

Ces dernières années, la communauté scientifique et le secteur de la sécurité dans le monde entier ont manifesté un grand intérêt pour les systèmes de vidéo-surveillance automatisés. La création de systèmes de sécurité capables de détecter, de suivre et de signaler les risques potentiels pour la sécurité avec une intervention humaine minimale a suscité un intérêt massif, en raison de la disponibilité d'ordinateurs puissants, de caméras vidéo de haute qualité, et du besoin croissant d'analyse vidéo automatisée. En particulier pour la surveillance des lieux publics tels que les gares, les aéroports, les autoroutes ou privés tels que les maisons, ces systèmes sont devenus de plus en plus attrayants grâce aux progrès technologiques et à l'introduction de diverses techniques de vision par ordinateur et d'intelligence artificielle.

Objectifs

La majorité des systèmes de surveillance existants reposent sur une surveillance humaine constante. C'est le principal inconvénient de ces systèmes, car l'étendue de la couverture et le nombre de caméras dépendent fortement de la disponibilité des opérateurs humains. Les performances générales de ces systèmes constituent un autre point faible. Ces systèmes reposent uniquement sur la vigilance de la personne qui les surveille, car aucun processus automatisé n'a été mis en œuvre. Il suffit de 20 minutes d'observation de moniteurs de sécurité pour que l'attention de la plupart des gens tombe en dessous d'un niveau inacceptable. Pour remédier à ce problème et assurer une surveillance constante des écrans, il est devenu courant d'enregistrer les séquences de surveillance pour les utiliser

ultérieurement.

Dans ce projet, nous proposons une conception et une mise en oeuvre d'un système de vidéo-surveillance dans une maison intelligente basée sur la vision par ordinateur et de deep learning dont les objectifs sont:

1. Détecter les visages de sujets présents dans une scène capturée par une caméra
2. Identifier l'identité des sujets détectés soit étant connu ou inconnu
3. notifier le propriétaire de ladite maison de la présente de personne étrangère si le passage de cette dernière se répète

Structure

La partie restante de ce document est organisée en 2 parties: théorique et pratique.
Dans la partie I, nous présentons 3 chapitres. Dans le chapitre 1, nous présentons d'abord les concepts de base nécessaires à la compréhension des différentes méthodes de traitement d'images. Le chapitre 2 donne un aperçue détaillé sur l'intelligence artificielle, ses types, particulièrement le deep learning. Le chapitre 3 couvre la méthodologie que nous avons suivie pour la détection et la reconnaissance de visages, les outils utilisés pour mettre en œuvre les modèles et les techniques d'évaluation sont discutées.
La partie pratique II se compose de 2 chapitres. Le chapitre 4 est une présentation de l'environnement de développement basé sur le Raspberry Pi 4. Dans le 5 implémentation des modèles utilisées sur le Raspberry Pi 4 est présentée et discutée en détails. Finalement, une conclusion générale est présentée.
Finalement, une conclusion générale est présentée.

Première partie

Partie Théorique

Chapitre 1

Traitement d'Images

Sommaire

1.1 Introduction

Le système de vision est la principale source de compréhension du monde qui nous entoure nous les êtres humains. En effet, on ne se contente pas d'observer les objets afin de les identifier et de les classifier, mais également pour repérer les divergences et obtenir une vue d'ensemble rapide d'une situation en un bref regard.
Dans ce chapitre, on se propose d'explorer les différentes techniques et méthodes utilisées dans le domaine du traitement d'image. Nous nous intéresserons particulièrement à l'application de ces techniques pour la vidéo-surveillance.

1.2 Préliminaires

Le traitement d'images consiste en toutes les opérations de traitement du signal qui impliquent une image en entrée, qu'il s'agisse d'une image ou d'une image vidéo. La sortie de ce traitement peut être soit une nouvelle image, soit une série de caractéristiques ou de paramètres liés à l'image d'origine [Petrou and Petrou, 2010]. Le traitement d'images est une technique qui permet de transformer une image en une forme numérique et d'appliquer diverses opérations sur celle-ci, dans le but d'améliorer l'image ou d'en extraire des informations pertinentes.

Les ordinateurs sont utilisés pour manipuler les images numériques à l'aide de techniques de traitement numérique. Les données d'imagerie brutes issues des capteurs présentent souvent des imperfections, qui doivent être corrigées par des étapes de traitement distinctes [Castleman, 1996].

1.2.1 Image numérique

Une image I peut être décrite comme une fonction à deux dimensions, $I(x, y)$, où x et y représentent les coordonnées spatiales et la valeur de I à chaque paire de coordonnées (x, y) correspond à l'intensité de l'image en ce point [Castleman, 1996].
Une image numérique est constituée d'un ensemble de points, également appelés pixels, stockés sous forme de nombres dans une matrice. Les images sont des données spatiales qui sont indexées par deux coordonnées spatiales.

Le système de coordonnées pour une image I est illustré dans la figure 1.1. L'image est représentée par $I(x, y)$, où x correspond à la position horizontale du pixel et y à la position verticale. Pour l'image de petite taille de la figure 1.1., les valeurs d'intensité sont

$I(0,0) = 0, I(3,1) = 18 \quad et \quad (2,3) = 19.$

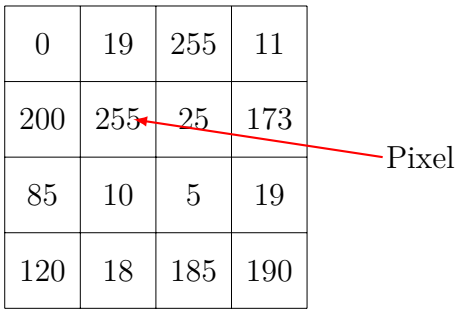


FIGURE 1.1 – Représentation d'un pixel dans une image

1.2.2 Les types d'images

L'image binaire est une image qui ne contient que deux valeurs d'intensité possibles. En général, le noir et le blanc sont utilisés dans l'image binaire, et la valeur de chaque pixel est stockée sous la forme de 0 ou de 1 respectivement.
Une image en niveaux de gris est une image qui n'a qu'une seule composante d'intensité et qui est représentée en noir et blanc. Elle contient 256 niveaux de gris, allant de 0 qui représente le noir à 255 qui représente le blanc. Le nombre 256 est lié à la quantification de l'image, car chaque niveau de gris est codé sur 8 bits, ce qui permet de représenter des valeurs entre 0 et $2^8 - 1 = 255$. Une image RGB est une image composée de trois couleurs primaires: rouge (R), vert (G) et bleu (B), représentées chacune par une composante de couleur distincte (voir la figure 1.2).

(a) Image en rouge (b) image en vert (c) image en bleu

FIGURE 1.2 – Représentation d'image en trois couleurs

Chaque composante de couleur est représentée par un nombre de 8 bits, ce qui permet de représenter 256 niveaux d'intensité pour chaque couleur.

La figure 1.3 illustre un exemple des différents types d'images mentionnés.

(a) Image binaire (b) Image en niveaux de gris (c) Image couleur

FIGURE 1.3 – Les différents types d'image

Il est possible de convertir une image couleur RGB en une image en niveaux de gris en utilisant des pondérations différentes pour chaque composante de couleur. Voici une équation couramment utilisée pour cette conversion:

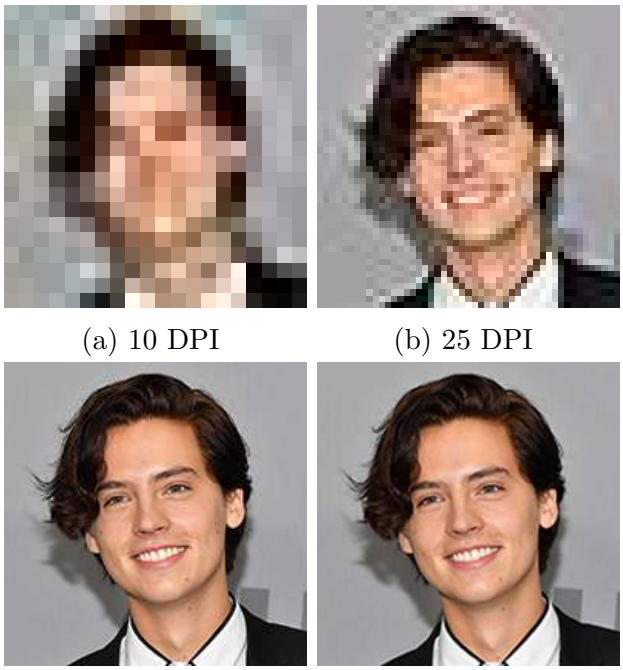
$$Gray(i,j) = 0.40 \times R(i,j) + 0.60 \times G(i,j) + 0.30 \times B(i,j) \tag{1.1}$$

Dans cette équation, $R(i,j)$, $G(i,j)$ et $B(i,j)$ représentent respectivement les valeurs de rouge, de vert et de bleu du pixel à la position (i,j) dans l'image d'origine. Les coefficients de pondération 0.30, 0.59 et 0.11 sont choisis de manière à donner plus de poids à la composante verte, qui est considérée comme plus importante pour la perception visuelle, et moins de poids aux composantes rouge et bleue.

1.2.3 La résolution d'une image

La résolution d'une image correspond au nombre de points qui se trouvent dans une longueur donnée, généralement exprimée en pouces et mesurée en Dots Per Inch (DPI). Un pouce correspond à $2,54cm$. La résolution permet de déterminer la relation entre la définition de l'image en pixels et sa taille réelle lorsqu'elle est représentée sur un support physique tel qu'un écran ou une impression papier. La figure 1.4 illustre l'impact de différentes résolutions sur la même image.

Les traits du visage du sujet dans la figure 1.4a sont indéchiffrables et son identification est impossible. La figure 1.4b, une image d'une résolution de $25DPI$, commence à montrer



(a) 10 DPI (b) 25 DPI

(c) 100 DPI (d) 300 DPI

FIGURE 1.4 – Une image avec différentes résolutions

plus de détails, mais la figure 1.4c est une image de $100DPI$ avec un niveau de détail beaucoup plus élevé.

1.2.4 Les formats d'images

1. **TIFF**

   Le format de fichier d'image appelé TIFF (Tagged Image File Format) a une capacité de stockage importante, ce qui en fait des fichiers de grande taille. Les images TIFF ne sont pas compressées, ce qui les rend très détaillées et explique leur volumétrie importante. De plus, la flexibilité de ce format est notable en matière de gestion des couleurs, car il permet de stocker des images en niveaux de gris, en CMYK pour l'impression ou en RGB pour le web.

2. **JPEG**

   JPEG est l'acronyme de Joint Photographic Experts Group, un organisme ayant créé une norme pour le format d'image portant ce nom. Les fichiers JPEG contiennent des images qui ont été compressées pour stocker de grandes quantités d'informations dans un fichier de petite taille. Le format JPEG est largement utilisé dans la photographie

numérique car il permet de stocker plus de photos sur une carte mémoire que d'autres formats. Un fichier JPEG est compressé de manière à réduire sa taille en perdant une partie des détails de l'image lors de la compression. On parle de compression "avec perte" car une partie des informations de l'image est perdue.

Les fichiers JPEG sont souvent utilisés pour les photographies sur le web, car ils permettent de créer des fichiers de petite taille tout en conservant une bonne qualité d'image.

Cependant, les fichiers JPEG ne sont pas adaptés aux dessins, logos ou graphiques contenant des lignes droites, car la compression peut créer un effet "bitmappy" ou des lignes irrégulières plutôt que des lignes nettes et précises.

3. **GIF**

   En effet, contrairement au JPEG, le format GIF utilise une compression sans perte, ce qui signifie que l'image est compressée sans perte de détails, mais cela rend difficile la compression d'images complexes en taille. Les fichiers GIF sont limités en termes de gamme de couleurs, ce qui les rend plus adaptés aux images simples et graphiques qu'aux photographies. Les GIF sont souvent utilisés pour les images avec des formes géométriques simples, des logos et des graphiques, ainsi que pour les animations simples sur le web.

4. **PNG**

   Le PNG est l'abréviation de Portable Network Graphics, est un format d'image ouvert conçu pour remplacer le format GIF, qui était soumis à des restrictions de brevet. Contrairement au GIF, le PNG utilise une compression sans perte, ce qui signifie qu'il n'y a aucune perte de détails dans l'image, mais il ne peut pas atteindre la même petite taille de fichier que le JPEG. Le PNG offre également une gamme complète de couleurs, le rendant adapté aux images web de haute qualité, mais il n'est pas aussi efficace que le JPEG pour les photographies. Le PNG est particulièrement adapté pour les images contenant du texte ou des dessins au trait, car il préserve la netteté des contours et minimise l'apparence de "bitmappy".

1.3 Le prétraitement des images

Le prétraitement des images a pour objectif de nettoyer les images en éliminant les données visuellement inutiles et indésirables telles que le bruit qui peut résulter du processus d'acquisition de l'image. Il peut également comprendre la recherche de régions d'intérêt

pour un traitement ultérieur.

1.3.1 Élimination du bruit

Le bruit est une perturbation non désirée dans une image, qui peut avoir différentes sources. Dans le cas d'une image numérique, le signal optique est converti en un signal électrique continu qui est ensuite échantillonné, et c'est à ce moment-là que le bruit peut apparaître. En plus de cela, d'autres facteurs comme des problèmes avec le capteur, des erreurs dans la transmission de données, ou des interférences électroniques peuvent également altérer la qualité de l'image de manière significative.[Castleman, 1996].
Le bruit dans une image numérique peut être modélisé par une distribution aléatoire, telle que la distribution gaussienne, uniforme ou du sel et du poivre. Ces types de bruit peuvent ajouter une valeur aléatoire à la luminosité exacte d'un pixel, créant ainsi une image avec des détails indésirables.
Les filtres spatiaux sont couramment utilisés pour supprimer différents types de bruit et améliorer la qualité de l'image. Les trois types de filtres couramment utilisés sont les filtres moyens, les filtres médians et les filtres de rehaussement. La figure 1.5 donne un exemple des types d'images avec des bruits différents. Les filtres moyens et médians sont couram-

(a) Image originale (b) Bruit de distribution Gaussienne (c) Bruit de distribution du sel et du poivre

FIGURE 1.5 – Une image avec différents types de bruit

ment utilisés pour réduire ou éliminer le bruit dans une image, mais ils peuvent également avoir d'autres applications. Par exemple, un filtre moyen peut être utilisé pour donner à une image un aspect plus doux ou flou, tandis qu'un filtre de rehaussement peut être utilisé pour faire ressortir les contours et les détails de l'image.
Les filtres spatiaux sont appliqués à l'aide de masques de convolution, qui sont de petites matrices de nombres utilisées pour calculer une nouvelle valeur pour chaque pixel de

l'image en fonction de ses valeurs et de celles de ses voisins. Comme cette opération utilise une somme pondérée des pixels voisins pour chaque pixel de l'image, elle est considérée comme un filtre linéaire[Andrews et al., 2013].
La figure 1.6 donne un exemple des images bruitées filtrées avec un filtre médian.

(a) Image originale (b) Image bruitée avec un de distribution du seldu sel et du poivre et du poivre (c) Bruit de distribution

FIGURE 1.6 – Une image filtrée par un filtre median

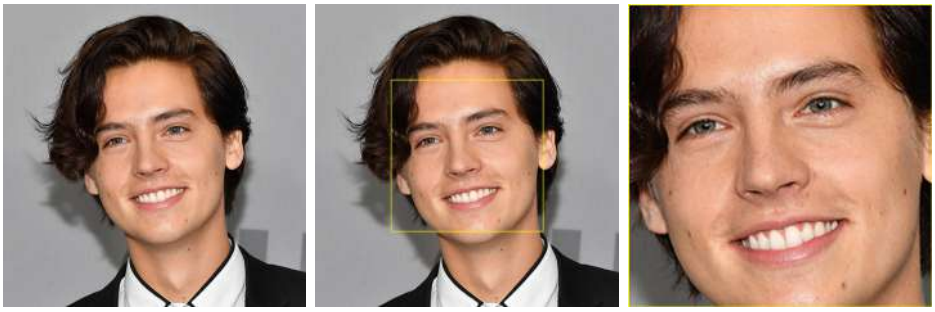
1.3.2 Recadrage d'une image

Pour effectuer une analyse d'images, il est souvent nécessaire de zoomer sur une zone d'intérêt spécifique appelée région d'intérêt (ROI). Pour cela, nous utilisons une opération appelée recadrage, qui modifie les coordonnées spatiales de l'image en sélectionnant une sous-image et en la coupant du reste de l'image. Par exemple, la figure 1.7 montre un exemple d'une image recadrée aux points $P_1 1(15, 15)$, $P_2(150, 100)$.

1.3.3 Redimenstionnement d'une image

Le redimensionnement d'images consiste à modifier l'échelle ou la grille de pixels d'une image d'origine pour obtenir une nouvelle image avec une taille différente. Cette opération est couramment utilisée dans de nombreuses applications de traitement d'images et d'apprentissage automatique. Le redimensionnement peut être utilisé pour réduire le nombre de pixels d'une image ou pour zoomer sur une zone d'intérêt.

L'opération de redimensionnement est réalisée à l'aide d'une interpolation bidimensionnelle de l'image originale, où plusieurs techniques peuvent être utilisées. La méthode



(a) Image originale (b) ROI (c) Image recadrée

FIGURE 1.7 – Exemple d'une image recadrée

du plus proche voisin, qui attribue simplement à chaque pixel de la nouvelle image la valeur de l'élément le plus proche de l'image d'origine, produit une image très pixelisée. L'interpolation bilinéaire approxime la fonction avec une interpolation linéaire, ce qui permet d'obtenir une image de meilleure qualité. L'interpolation bicubique utilise une approximation cubique pour obtenir une image encore meilleure.

La figure 1.8 illustre un exemple d'application de cette technique avec une image avant et après le redimensionnement [Thévenaz et al., 2000].

1.4 La vidéo

L'avènement de la vidéo numérique a été une avancée technologique majeure de notre ère, devenant un élément essentiel de la vie quotidienne. Une séquence vidéo contient plus d'informations visuelles qu'une seule image, notamment en raison de la capture du mouvement à travers une séquence d'images plutôt qu'une seule image statique.

1.4.1 Définition

La vidéo est une technologie qui permet d'enregistrer des images en mouvement de manière électronique. Ces images en mouvement sont en réalité une séquence d'images fixes qui changent suffisamment vite pour donner l'impression de mouvement. La fluidité d'une vidéo est mesurée en nombre d'images par seconde (FPS).
Chaque image dans une séquence vidéo est représentée par une matrice de valeurs où chaque valeur dépend de l'information entourant le point correspondant dans l'image. Un élément unique dans une matrice d'image contient des informations pour toutes les

(a) Image originale (256×256)

(b) (Image réduite (64×64)

(c) Interpolation du plus proche voisin

(d) Interpolation bilinéaire

(e) Interpolation cubique

FIGURE 1.8 – Une image redimensionnée avec différents types d'interpolationt

composantes de couleur. Pour stocker ou transmettre une vidéo, une grande capacité de stockage et un taux de transfert rapide sont nécessaires. La compression vidéo est grandement influencée par le nombre d'images par seconde et la résolution de l'image.

1.4.2 Nombre d'images par seconde

Le FPS représente la fréquence à laquelle les images successives, appelées trames, apparaissent sur un écran pour former des images animées. Bien que nous percevions les vidéos que nous visionnons comme étant en mouvement, elles sont en réalité constituées d'une séquence d'images statiques qui se succèdent rapidement. Par exemple, si une vidéo est enregistrée à 24 images par seconde, cela signifie qu'en une seconde, 24 images distinctes sont affichées (comme illustré dans la figure 1.9). Toutefois, le nombre de FPS peut varier en fonction des différents supports de visualisation, ainsi que d'autres facteurs. [FPS,]

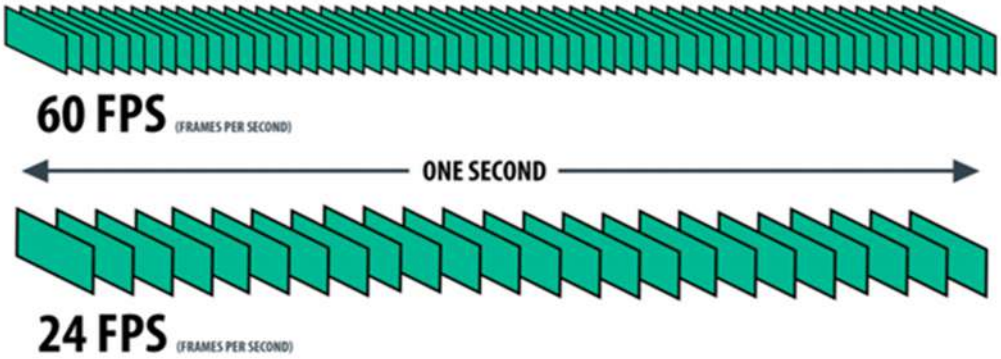


FIGURE 1.9 – Illustration de la notion de FPS

1.4.3 La résolution

Le nombre de pixels affichables sur un dispositif est appelé résolution. Cette valeur est souvent indiquée sous forme de deux nombres, représentant respectivement la largeur et la hauteur de l'écran en pixels. Par exemple, une résolution de 640 par 480 signifie que l'écran a une largeur de 640 pixels et une hauteur de 480 pixels, soit un total de 307200 pixels. Il existe plusieurs formats de résolution tels que le SD, le HD, le Full HD, etc. La figure 1.10 compare différentes résolutions.

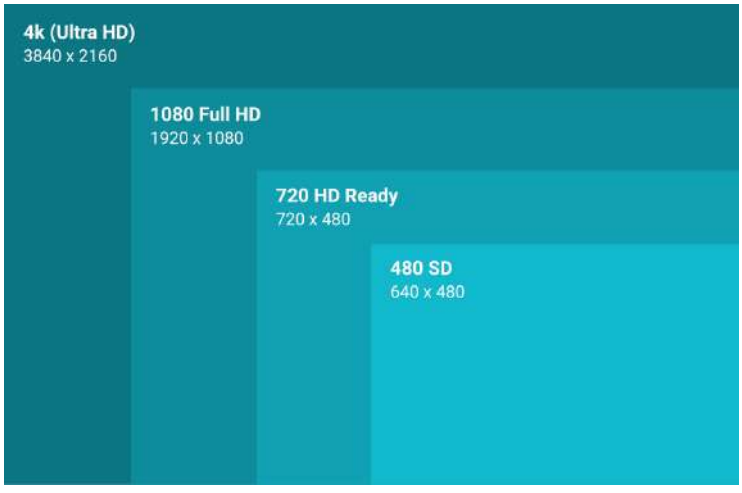


FIGURE 1.10 – Comparatif des différentes résolutions

1.4.4 Domaine d'utilisation de la vidéo

Au fil des vingt dernières années, l'utilisation de la vidéo a connu une croissance significative dans divers domaines tels que les communications, l'éducation, la médecine et les divertissements. Les avantages des applications vidéo incluent une meilleure communication, une compréhension plus rapide des concepts complexes et une accessibilité accrue à l'information. En outre, la demande pour différentes tâches de surveillance vidéo telles que la détection de mouvements, la reconnaissance faciale, la détection d'intrus, le contrôle de scènes intérieures et extérieures comme les aéroports, les gares, les parkings, les autoroutes et les magasins est en constante augmentation.

1.5 Conclusion

Le présent chapitre a présenté les notions fondamentales essentielles à la compréhension des diverses techniques de traitement d'images, ainsi que les particularités des vidéos. Les prochains chapitres seront consacrés aux méthodes de surveillance dans une maison intelligente qui s'appuient sur le traitement d'images.

Chapitre 2

Intelligence Artificielle

Sommaire

2.1 Introduction

Dans ce chapitre, nous mettons l'accent sur l'intelligence artificielle (IA) en tant que pierre angulaire de notre travail. Nous allons donc lui accorder une attention particulière et explorer ses différentes facettes afin de mieux comprendre son fonctionnement, ses applications et ses implications dans notre domaine d'intérêt.

2.2 L'intelligence artificielle

L'Intelligence Artificielle (IA), parfois appelée intelligence des machines, est une intelligence démontrée par des machines ou des ordinateurs. L'IA a été utilisée pour émuler des fonctions complexes associées à l'esprit humain, telles que la détection, l'apprentissage et la prédiction. Depuis 1956, la recherche sur l'IA est reconnue comme une discipline universitaire [Su and Yang, 2022].
L'IA peut être classée en IA analytique, inspirée par l'homme et humanisée, en fonction des types d'intelligence qu'elle présente (intelligence cognitive, émotionnelle et sociale) ou en intelligence artificielle étroite, générale et super intelligente, en fonction de son stade d'évolution [Haenlein and Kaplan, 2019].
L'intelligence artificielle [1] a de nombreuses applications pratiques dans divers secteurs et domaines, notamment:

1. **Les soins de santé:** l'IA est utilisée pour le diagnostic médical, la découverte de médicaments et l'analyse prédictive des maladies
2. **Les finances:** l'IA aide à l'évaluation du crédit, à la détection des fraudes et aux prévisions financières
3. **Le commerce de détail:** l'IA est utilisée pour les recommandations de produits, l'optimisation des prix et la gestion de la chaîne d'approvisionnement
4. **La production:** l'IA aide au contrôle de la qualité, à la maintenance prédictive et à l'optimisation de la production
5. **Les transports:** l'IA est utilisée pour les véhicules autonomes, la prédiction du trafic et l'optimisation des itinéraires
6. **Les services à la clientèle:** Les chatbots alimentés par l'IA sont utilisés pour le support client, pour répondre aux questions fréquemment posées et pour traiter les demandes simples

1. Artificial Intelligence | An Introduction - GeeksforGeeks

7. **La sécurité:** l'IA est utilisée pour la reconnaissance faciale, la détection des intrusions et l'analyse des menaces de cybersécurité
8. **Le marketing:** l'IA est utilisée pour la publicité ciblée, la segmentation des clients et l'analyse des sentiments
9. **L'éducation:** l'IA est utilisée pour l'apprentissage personnalisé, les tests adaptatifs et les systèmes de tutorat intelligents

2.2.1 L'apprentissage automatique

Le Machine Learning (ML) désigne, de manière générale, le processus consistant à adapter des modèles prédictifs à des données ou à identifier des regroupements informatifs au sein des données. En d'autres termes, les systèmes du ML sont capables de prédire les actions futures sur la base des expériences passées ([Bishop and Nasrabadi, 2006];[Murphy, 2012]) .
Le domaine de l'apprentissage automatique tente essentiellement d'approcher ou d'imiter la capacité des humains à reconnaître des modèles, mais de manière objective, en utilisant le calcul.

Le ML est particulièrement utile lorsque l'ensemble de données que l'on souhaite analyser est trop grand (beaucoup de points de données individuels) ou trop complexe (contient un grand nombre de caractéristiques) pour une analyse humaine et/ou lorsqu'on souhaite automatiser le processus d'analyse des données pour établir un pipeline reproductible et efficace en temps.

Les algorithmes de ML peuvent être classés en fonction de la nature des données utilisées dans le processus d'apprentissage et du résultat souhaité. La figure 2.1 représente le schéma général des algorithmes d'apprentissage automatique.

Nous examinerons dans ce qui suit les types les plus importants du ML.

2.2.1.1 Apprentissage supervisé

Dans ce type d'apprentissage, les données d'entrée sont généralement composées d'une paire d'éléments: le vecteur d'entrée (X) et son étiquette (y). Dans l'apprentissage supervisé, un classifieur est entraîné à l'aide des de données étiquetées. Une fois qu'il a été entraîné, il peut être utilisé pour faire des prédictions sur les étiquettes de nouvelles données non étiquetées qui sont très similaires aux données entraînées. L'apprentissage

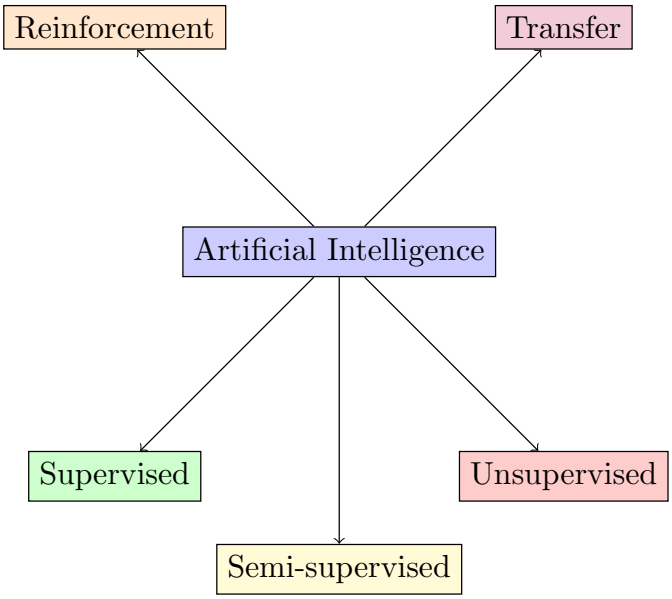


FIGURE 2.1 – Les types de Machine Learning

supervisé le plus courant utilisé est la classification, où la sortie est simplement un code entier identifiant une catégorie spécifique (une image peut être reconnue comme provenant de la catégorie 0 contenant des visages, ou de la catégorie 1 contenant des non-visages, etc.)[Goodfellow et al., 2020].
Un système est considéré comme un modèle de régression si les données de sortie sont des variables continues; c'est le cas dans des applications telles que les prévisions météorologiques et boursières.

2.2.1.2 Apprentissage non-supervisé

L'apprentissage non-supervisé est une branche dans laquelle les données d'apprentissage se composent uniquement de vecteurs d'entrée (X) sans étiquettes de sortie associées. L'objectif est de trouver des points communs ou d'identifier de nouveaux modèles dans les données d'entrée. Le regroupement et la réduction de la dimensionnalité sont des exemples courants d'apprentissage non-supervisé [Goodfellow et al., 2020].

2.2.1.3 Apprentissage semi-supervisé

L'apprentissage semi-supervisé combine des aspects de chacune de approches précédentes. Il utilise à la fois des données étiquetées et non étiquetées et est couramment déployé dans des scénarios où les données étiquetées sont insuffisantes. La classification semi-supervisée est une technique qui permet d'améliorer les résultats de la classification supervisée en entraînant un classificateur avec des données étiquetées et non étiquetées [Zur et al., 2009]

2.2.1.4 Apprentissage par renforcement

L'apprentissage par renforcement est une branche de l'apprentissage automatique, où les machines apprennent progressivement les comportements de contrôle via l'auto-exploration de l'environnement.
L'apprentissage par renforcement emploie un acteur qui interagit de manière itérative avec l'environnement et modifie ses actions de contrôle pour maximiser les récompenses reçues de l'environnement. Le principal avantage de l'algorithme d'apprentissage par renforcement est qu'il apprend à optimiser les politiques de contrôle par l'exploration de l'environnement indépendamment de la linéarité ou de la multivariabilité du système.[Ma et al., 2019].

2.2.1.5 Apprentissage par transfert

La plupart des algorithmes d'apprentissage automatique sont conçus pour résoudre des types de problèmes spécifiques; un modèle construit pour identifier les visages, par exemple, n'essaierait pas de faire la même chose pour les animaux. Cependant, cette méthode ne donne pas une image fidèle de l'intelligence humaine; par exemple, nous savons que les gens peuvent appliquer leur compréhension des visages humains sur des animaux. Les êtres humains ont une capacité innée à tirer parti d'un large éventail d'expériences et à adapter ce qu'ils ont appris à de nouvelles situations.

En s'inspirant de la manière dont les humains apprennent, les chercheurs ont étudié comment adapter les méthodes traditionnelles d'apprentissage automatique pour transférer des informations d'une tâche (la "tâche source") à une nouvelle tâche (la "tâche cible") qui est conceptuellement liée à celles qui ont été apprises en premier lieu.

2.2.2 Deep Learning

L'apprentissage profond est un sous-ensemble de l'apprentissage automatique (voir la figure 2.2) qui implique l'utilisation de réseaux neuronaux à couches multiples pour apprendre des représentations hiérarchiques des données [LeCun et al., 2015, Goodfellow et al., 2016]. L'apprentissage profond a atteint des performances de pointe sur un large éventail d'applications, notamment la reconnaissance d'images, le traitement du langage naturel et la reconnaissance vocale.

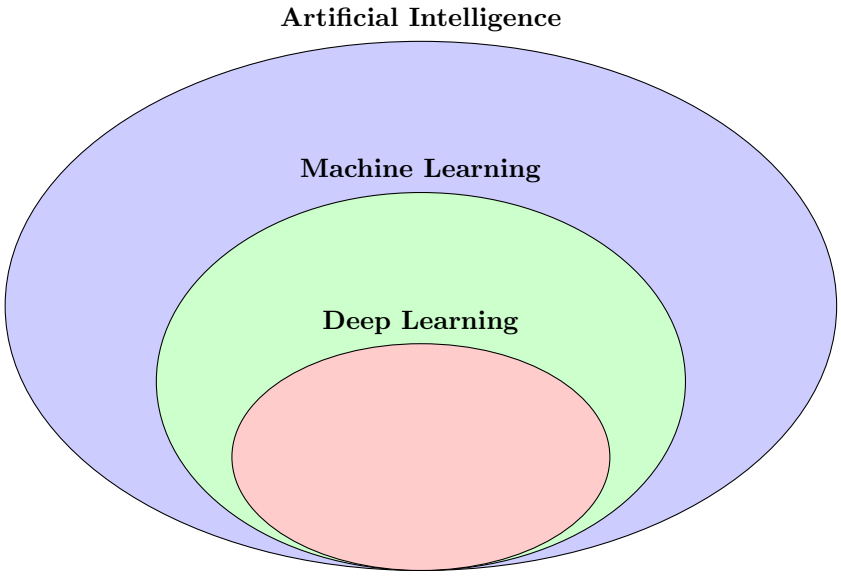


FIGURE 2.2 – Diagramme illustrant la différence entre l'intelligence artificielle, l'apprentissage automatique et l'apprentissage profond.

2.2.2.1 Le cerveau humain et les réseaux neuronaux artificiels (Artificial Neural Networks)

Le cerveau humain contenant 86 milliards de neurones, se forme de manière auto-organisée: les neurones se connectent naturellement les uns aux autres par des connexions structurelles appelées « synapses ».Un neurone se compose de 3 parties: le soma, les dendrites et l'axone, qui partagent la même membrane cellulaire. L'axone d'un neurone (neurone présynaptique) peut former des synapses avec les dendrites ou soma d'un autre neurone (neurone postsynaptique). Les synapses sont une structure qui propage des signaux de manière unidirectionnelle d'une manière«

électrique-chimique-électrique » [Chen et al., 2023]. La figure 2.3 illustre l'architecture d'un neurone biologique.

Un réseau neuronal artificiel est un modèle qui imite la structure et le comporte-

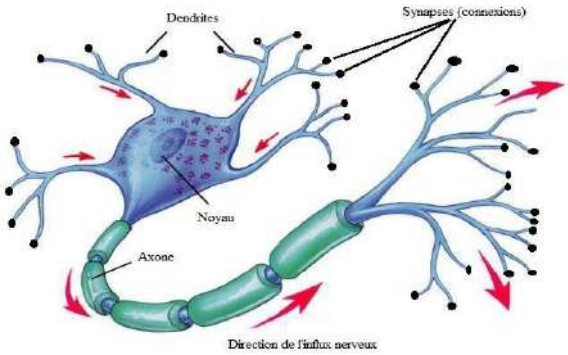


FIGURE 2.3 – Architecture d'un neurone biologique

ment des réseaux neuronaux réels, notamment le cerveau humain. Il s'agit d'un type d'algorithme d'apprentissage automatique auquel on peut apprendre à analyser des données afin d'en tirer des conclusions ou de faire des prédictions. Un neurone artificiel (ou nœud) est l'élément fondamental d'un réseau neuronal artificiel; il reçoit des données, les traite et produit un résultat. Chaque couche de neurones communique avec les couches supérieures et inférieures. La couche d'entrée, la ou les couches cachées et la couche de sortie sont les trois principaux types de couches d'un tel réseau.

2.2.2.2 L'architecture des réseaux neuronaux

Comme le montre la figure 2.4, les réseaux neuronaux comportent principalement 3 couches:

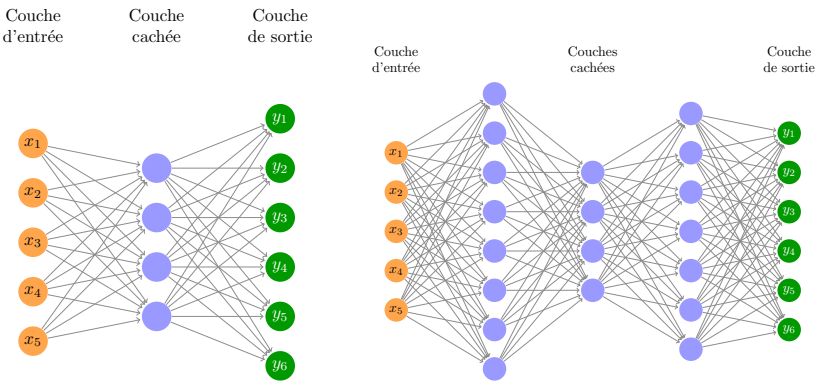
— Input Layer (couche d'entrée)

— Hidden Layers (couches cachées)

— Output Layer (couche de sortie)

Input Layer: La couche initiale du réseau de neurones comprend les neurones qui reçoivent les caractéristiques en entrée. En plus de ces caractéristiques, un biais est également inclus dans la couche d'entrée.

Hidden Layers: Les couches cachées servent d'intermédiaires entre les couches d'entrée et de sortie et il est possible d'avoir un nombre variable de couches cachées. Les réseaux



(a) Réseau de neurones simple (b) Réseau de neurones profond

FIGURE 2.4 – Illustration de réseau de neurones

qui ont plus d'une couche cachée sont considérés comme des réseaux neuronaux profonds. Les neurones de la couche cachée sont alimentés par les entrées de la couche d'entrée et produisent des sorties à destination de la couche de sortie.

Output Layer: La couche de sortie est conçue avec un nombre de neurones qui correspond au nombre de classes de sortie. Par exemple, en cas de classification multi-classes, elle aura autant de neurones que de classes différentes. En revanche, pour une classification binaire, elle ne comportera que deux neurones.

Les neurones formels suivent ce même principe. Ils sont conçus comme des fonctions mathématiques à plusieurs variables réelles, avec plusieurs entrées et une sortie qui correspondent respectivement aux dendrites et à l'axone. Le perceptron, qui est le modèle le plus simple de neurone artificiel, a été créé en 1957 par F. Rosenblatt (voir figur 2.5).

Un neurone j est défini par:

- des entrées: $x_1, x_2, ..., x_n$
- des poids: $w_1, w_2..., w_n$ qui pondèrent les entrées.
- un biais: b_j qui régule l'activation du neurone.
- une fonction d'activation $\Phi()$: détaillée ci-dessous.
- une sortie : y_j

Dans le schéma présenté dans la figure 2.5, les entrées subissent une multiplication par des poids spécifiques avant d'être transmises à la couche cachée suivante. En plus de ces entrées pondérées, un biais est également ajouté. La somme pondérée résultante est ensuite soumise à une fonction non linéaire connue sous le nom de fonction d'activation.

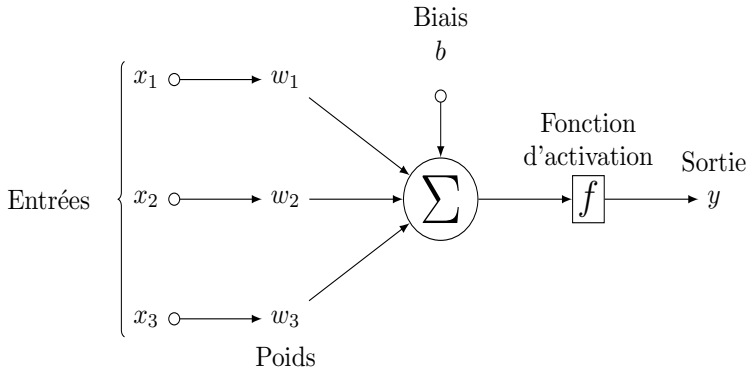


FIGURE 2.5 – Architecture d'un neurone artificiel (perceptron)

2.2.2.3 Les fonctions d'activation

Les fonctions d'activation sont un élément crucial du DL car elles déterminent si un neurone est activé ou non. Afin d'apprendre des fonctions complexes, il est nécessaire que la fonction d'activation soit non linéaire[Xiao, 2019].

Les différentes fonctions d'activation seront décrites dans la suite et illustrées dans la figure 2.6.

- **Fonction Sigmoid (sigmoïde):** La fonction a pour but principal de ramener la valeur d'entrée dans une plage entre 0 et 1, et est définie selon les termes suivants :

$$F(x) = \frac{1}{1 + e^{-x}} \tag{2.1}$$

- **Fonction Tanh:** La fonction souvent appelée "tangente hyperbolique" génère une sortie qui varie entre -1 et 1. Sa formule mathématique est la suivante:

$$F(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \tag{2.2}$$

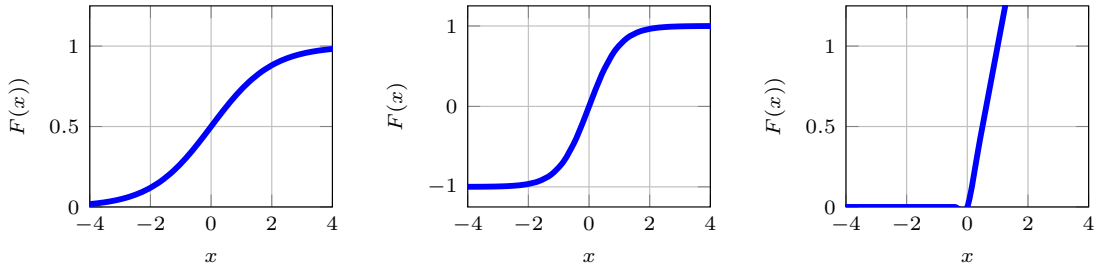
- **Unité linéaire rectifiée (ReLU):** La fonction ReLU est largement utilisée dans l'apprentissage profond pour améliorer les propriétés de la fonction de décision et de l'ensemble du réseau, sans affecter les champs réceptifs de la couche convolutive. Sa définition est la suivante:

$$F(x) = max(0, x) \tag{2.3}$$

  La fonction d'activation ReLU a été utilité dans [Krizhevsky et al., 2017] conduisant ç une accélération considérable la convergence du processus d'optimisation.

- **Fonction Softmax:** La fonction SoftMax est couramment utilisée comme dernière couche dans les modèles d'apprentissage profond pour calculer les scores de la classe. Elle prend en entrée un vecteur de caractéristiques qui a été obtenu grâce au processus d'apprentissage et renvoie une probabilité indiquant la probabilité qu'une image appartienne à une classe donnée. La formule de la fonction d'activation Softmax est la suivante:

$$F(x) = \frac{e^{x_i}}{\sum_{\geq 0} e^{x_p}} \tag{2.4}$$



(a) Sigmoid activation function (b) Hyperbolic tangent activation (c) Rectified linear unit activation

FIGURE 2.6 – Les fonctions d'activations

2.2.3 Convolutional Neural Network:

Un réseau de neurones convolutif (CNN), également connu sous le nom de Convolutional Neural Network, est un type de réseau neuronal artificiel qui utilise des perceptrons multicouches. Dans ce type de réseau, les couches cachées sont souvent des couches de convolution qui sont utilisées pour la reconnaissance et le traitement d'images [LeCun et al., 1989]. La figure 2.7 illustre un exemple de CNN qui peut être utilisé pour la classification des images, telles que des visages.

Un CNN est formé de:

- **Couche de convolution:** est un élément fondamental des réseaux neuronaux convolutifs utilisés pour les tâches de vision par ordinateur. Cette couche applique un ensemble de filtres pouvant être appris aux données d'entrée, en effectuant des opérations de convolution pour extraire des caractéristiques significatives. Les convolu-

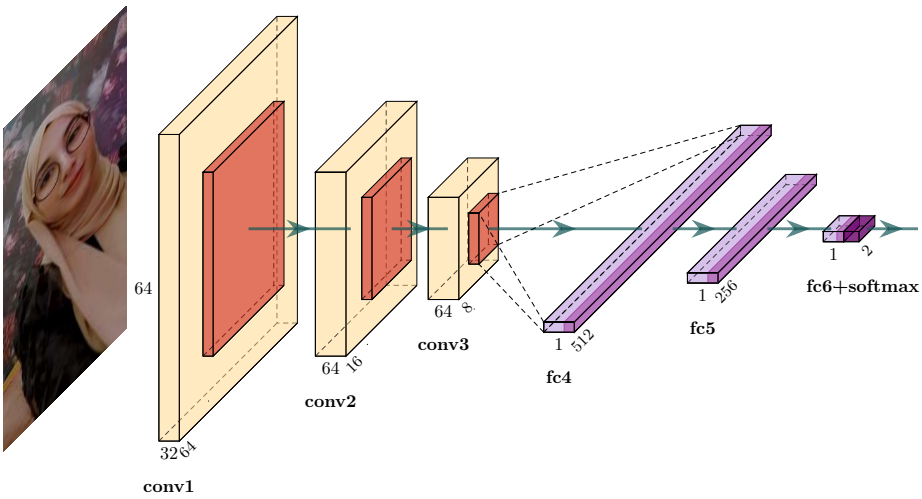


FIGURE 2.7 – Exemple du réseau de neurones convolutifs CNN 2D

tions sont effectuées sur les dimensions spatiales de l'entrée, telles que la largeur et la hauteur. La figure 2.8 illustre un exemple simple du calcul dans un CNN.

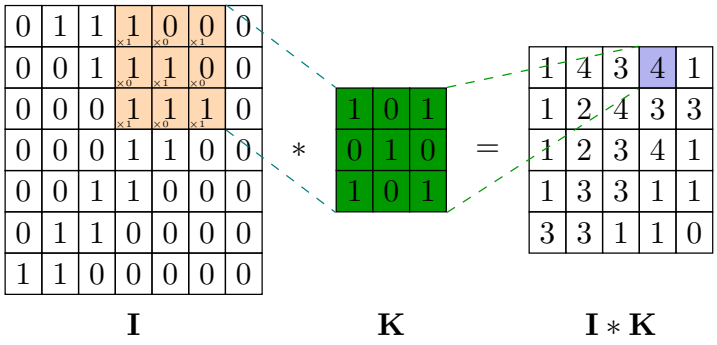


FIGURE 2.8 – Couche de convolution 2D

- **Couche de pooling):** Elle réduit la dimension par sous-échantillonnage en appliquant conventionnellement un noyau plus petit qui extrait la valeur souhaitable lorsqu'il est appliqué. Le pooling maximal renvoie la plus grande valeur dans le noyau de sous-échantillonnage, alors que le pooling moyen renvoi la valeur moyenne. La couche de pooling réduit l'image tout en conservant les informations les plus importantes. La figure 2.9 représente une opération simple de réduction de la dimension d'une carte d'activation à l'aide de pooling maximal et moyen, respectivement.
- **Fully connected layer (couche entièrement connectée):** Cette couche relie

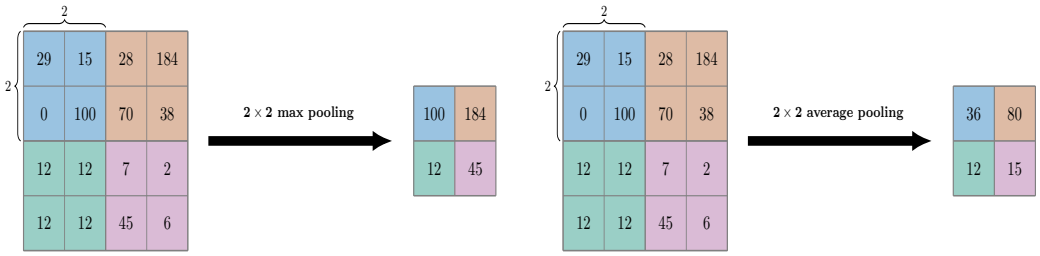


FIGURE 2.9 – Illustration des différents types de pooling

tous les neurones de la couche précédente à tous les neurones de la couche suivante, à l'aide d'une matrice de poids et d'un vecteur de biais, et produit la sortie finale du réseau [Liu et al., 2018]. La figure 2.10 donne un exemple d'une couche entièrement connectée.

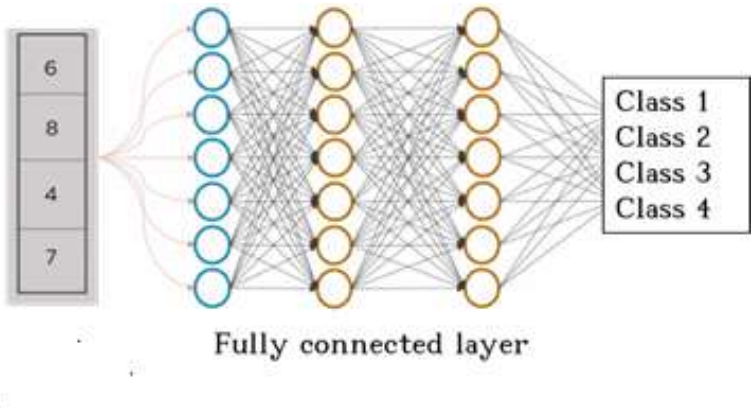


FIGURE 2.10 – Exemple d'une couche entièrement connectée

- **Dropout:** Le terme « dropout » désigne les unités d'abandon dans un réseau neuronal. Par abandon d'une unité, nous entendons sa suppression temporaire du réseau, ainsi que toutes ses connexions entrantes et sortantes, comme illustré à la figure 2.11. Le choix des unités à déposer est aléatoire Il empêche le sur-ajustement.
- **Flattening Layer (couche d'aplatissement) :** La couche aplatie convertit les données de la matrice en un tableau unidimensionnel comme l'indique la figure 2.12, qui peut être utilisé dans la couche entièrement liée. Lorsque l'on considère CNN, les deux dernières étapes sont des couches aplaties et entièrement connectées. Il est converti en un tableau 1D en préparation de la prochaine couche entièrement connectée de classification d'image [Tazin et al., 2021].

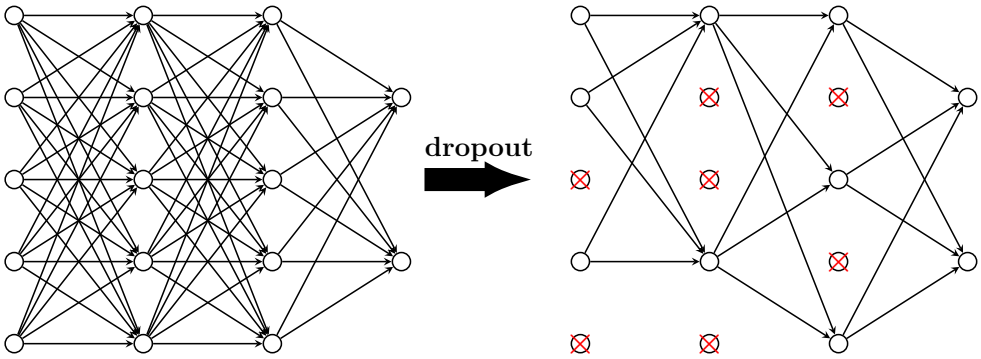


FIGURE 2.11 – la différence entre original networks et dropout networks

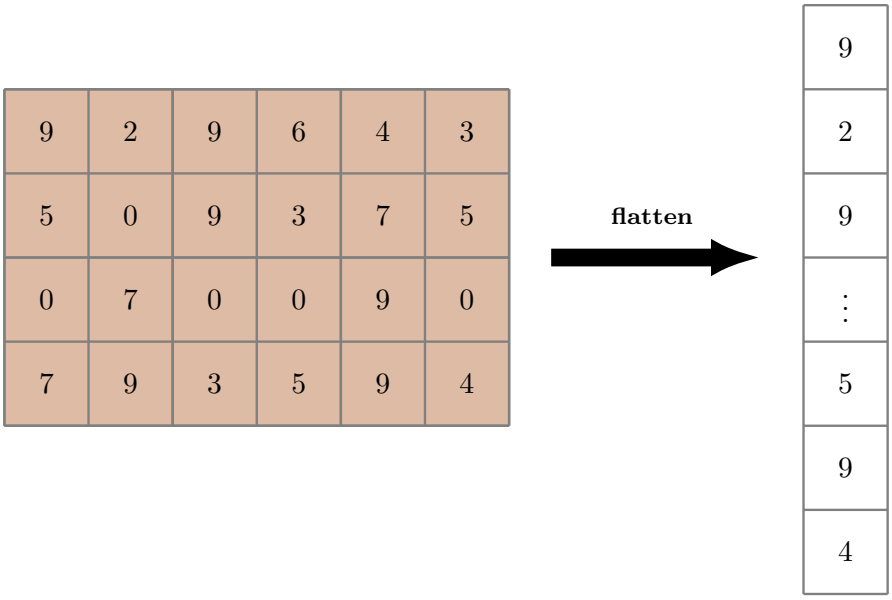


FIGURE 2.12 – Exemple de flattening 2D

La forme conventionnelle d'un CNN est la suivante :

$$\texttt{Input} \rightarrow [[\texttt{ConvLayer} \rightarrow \texttt{ReLU}] * N \rightarrow \texttt{PoolLayer}?] * M \rightarrow [\texttt{FC} \rightarrow \texttt{ReLU}] * K \rightarrow \texttt{FC}$$

où $*$ indique une répétition, et `PoolLayer`? indique une couche pooling facultative. En outre, $0 \leq N \leq 3$, $M \geq 0$ et $0 \leq K < 3$ sont des choix courants.

2.2.4 Apprentissage d'un réseau de neurones convolutifs (Neuron learning) :

On peut définir le processus d'ajustement des valeurs des poids comme le "training" ou l'apprentissage du réseau neuronal. Initialement, un CNN démarre avec des poids aléatoires. Au cours de l'apprentissage du CNN, le réseau neuronal est alimenté avec un grand ensemble de données d'images étiquetées avec leurs étiquettes de classe correspondantes (chats, chiens, chevaux, voitures, personnes, etc.). Pour chaque image, le réseau CNN attribue des valeurs aléatoires et compare les résultats avec l'étiquette de classe de l'image d'entrée. Si la sortie ne correspond pas à l'étiquette de classe, le réseau CNN procède à un petit ajustement des poids de ses neurones CNN afin que la sortie corresponde correctement à l'étiquette de classe de l'image.
Le processus d'apprentissage des réseaux neuronaux se déroule en deux étapes distinctes : Forward et Backward.
L'étape Forward implique la prédiction faite par le réseau neuronal pour une entrée donnée, suivie du calcul de l'erreur en comparant la sortie prédite avec la valeur de référence fournie par les données d'apprentissage. L'étape Backward, également appelée rétropropagation, est utilisée pour mettre à jour les poids des neurones en fonction de l'erreur commise par le réseau. Cette étape est effectuée à l'aide d'un algorithme d'optimisation choisi par l'utilisateur.
En outre, il existe différentes fonctions de coût qui peuvent être utilisées pour calculer l'erreur et guider le processus de mise à jour des poids telles que :
- Erreur quadratique moyenne:

$$L = \frac{1}{n}(\sum_{i=1}^{n} \left(y^i - Y^i)^2 \tag{2.5}$$

- Entropie croisée:

$$L = \frac{1}{n}\sum_{i=1}^{n}[y^i log(Y^i) + (1 - y^i) log(1 - Y^i)] \tag{2.6}$$

La prédiction Y du neurone est comparée à la valeur de référence y des données d'apprentissage, et cela est effectué n fois pour chaque prédiction.
L'algorithme de descente de gradient est utilisé pour optimiser une fonction objectif, qui dans le domaine de l'apprentissage automatique correspond à l'erreur d'apprentissage de dimension d. Ici, d représente le nombre de paramètres du modèle et l'objectif est de

minimiser ou de maximiser cette fonction en ajustant les paramètres du modèle.

Lorsque les données à traiter sont volumineuses, la descente de gradient devient difficile à appliquer.
Ainsi, pour remédier à cette situation, les données sont divisées en plusieurs mini-batchs, et pour chaque mini-batch, on calcule la perte et le gradient pour mettre à jour les paramètres.
Cette technique connue sous le nom de Stochastic Gradient Descent (SGD) est couramment utilisée de nos jours, car elle offre un bon compromis entre la robustesse et la rapidité d'évolution du réseau. Outre l'algorithme SGD, d'autres algorithmes tels que l'algorithme ADAM et l'algorithme RMSprop sont également utilisés pour minimiser la perte et mettre à jour les paramètres.
Pour améliorer la précision du réseau neuronal, il est entraîné pendant plusieurs itérations, appelées époques, sur l'ensemble des données.

L'un des défis majeurs du processus d'apprentissage est de choisir l'algorithme d'apprentissage approprié (optimiseur). L'algorithme Adam est une technique d'optimisation qui remplace la méthode classique de descente de gradient stochastique. Il utilise une approche itérative pour mettre à jour les poids du modèle au cours de l'apprentissage. Cette méthode est simple à implémenter et efficace en termes de calcul. De plus, elle consomme moins de mémoire et est performante.

Le terme "epoch" se réfère au nombre de fois que l'algorithme d'apprentissage parcourt l'ensemble de données d'entraînement. Au cours de chaque époque, tous les échantillons de données sont présentés au réseau neuronal une fois, permettant ainsi une mise à jour des paramètres du modèle interne.

La taille du lot (batch size) est un hyperparamètre qui détermine le nombre d'échantillons d'un ensemble de données qui seront traités simultanément avant de mettre à jour les paramètres du modèle interne.

Le réglage des hyperparamètres est un processus crucial pour obtenir les meilleures performances d'un algorithme d'apprentissage. Il consiste à sélectionner un ensemble de valeurs optimales pour les hyperparamètres avant le début du processus d'apprentissage. Les hyperparamètres sont des paramètres qui contrôlent le comportement de l'algorithme d'apprentissage, tels que la taille du lot, le taux d'apprentissage et le nombre de couches cachées. En général, les paramètres de poids des nœuds sont ajustés pendant le processus

d'apprentissage.

Il existe deux stratégies courantes pour le réglage des hyperparamètres : la recherche par grille et la recherche aléatoire. La recherche par grille consiste à spécifier un ensemble de valeurs pour chaque hyperparamètre, puis à tester toutes les combinaisons possibles pour déterminer la meilleure configuration. La recherche aléatoire, quant à elle, consiste à échantillonner aléatoirement des valeurs pour chaque hyperparamètre et à les évaluer pour trouver la meilleure combinaison.

2.3 Conclusion

Ce chapitre nous a permis de comprendre les bases de l'intelligence artificielle, du deep learning et du modèle de réseau de neurones convolutifs. Ces connaissances nous seront indispensables pour la suite de notre travail, où nous allons explorer plus en détail les applications spécifiques du deep learning, en mettant l'accent sur la reconnaissance faciale et ses implications dans des domaines tels que la sécurité et la surveillance.

Chapitre 3

Détection et Reconnaissance Automatique des Visages pour la Surveillance dans une Maison Intelligente

Sommaire

3.1 Introduction

Le domaine de la vidéo surveillance est un sous-ensemble du secteur de la sécurité physique. Il implique l'utilisation de caméras pour surveiller à distance les espaces publics et privés. Bien que ces caméras soient facilement disponibles et très peu coûteuses, le coût du maintien d'une présence humaine vigilante est élevé. Par conséquent, le flux vidéo émanant de ces caméras fait l'objet d'une surveillance modérée, voire inexistante. Elles sont souvent utilisées à des fins d'archivage ou comme moyen de communication après un incident.

3.2 Généralités sur la vidéo-surveillance

Aujourd'hui, les caméras sont omniprésentes dans tous les pays. La protection des lieux publics (aéroports, gares, banques, centres commerciaux, stades) la sécurité domestique (détection des vols, détection des incendies), la surveillance des personnes âgées (analyse de l'activité, détection des chutes), la sécurité routière (estimation des flux, détection des accidents) et la détection d'événements inhabituels ne sont que quelques exemples des systèmes de vidéo surveillance disponibles.
Il existe aujourd'hui plusieurs types d'équipements de vidéo surveillance tels que:

— Les caméras de sécurité utilisant la technologie analogique qui capturent des images analogiques stockées sur des enregistreurs numériques analogiques (DVR). Il existe des caméras câblées et sans fil, mais les caméras analogiques ont souvent une résolution d'image inférieure à celle des caméras numériques.

— Les caméras IP (Internet Protocol) sont utilisées dans les systèmes de vidéo surveillance numériques, qui collectent et transmettent les images numériquement sur un réseau IP. La résolution des caméras numériques est souvent meilleure que celle des systèmes analogiques, et elles peuvent être câblées ou sans fil. Les NVR, ou "Network Video Recorders", sont les dispositifs utilisés pour stocker les images.

— Les caméras IP utilisées dans un système de vidéo surveillance basé sur le cloud computing téléchargent leurs images sur un serveur distant. Tout appareil disposant d'une connexion internet peut être utilisé pour voir et enregistrer des images à distance. Ces systèmes sont utilisés dans les foyers et les petites entreprises.

— Dans les systèmes de vidéo surveillance sans fil, les caméras peuvent envoyer des images sans fil à une station de réception. Ils sont simples à mettre en place car

aucun câble ne doit être posé, mais des interférences et une faible portée peuvent survenir.

— Les systèmes de vidéo surveillance intelligents utilisent des technologies avancées telles que la détection de mouvement, la reconnaissance faciale et d'autres fonctions pour accroître la sécurité et l'efficacité.

Comme le montre la figure 3.1 , les systèmes de vidéo surveillance courants se composent généralement des étapes suivantes : acquisition, compression, transmission, décompression et traitement.

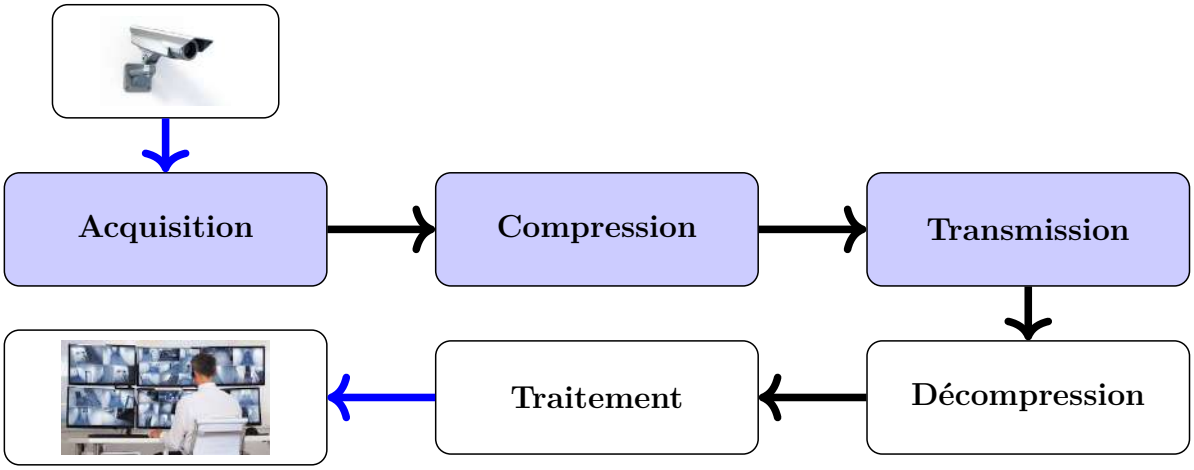


FIGURE 3.1 – Fonctionnement d'un système de vidéo-surveillance

— **L'acquisition:** Une caméra de sécurité enregistre la zone d'intérêt. Il existe de nombreux modèles de caméras pour répondre à un large éventail d'exigences en matière de surveillance. Elles peuvent être équipées ou non de moteurs et peuvent être analogiques ou numériques.

— **La compression:** La séquence vidéo numérique est un ensemble de données volumineux qui doit être transféré et traité, et qui doit donc être encodé (compressé). Cela nécessite une grande capacité de transfert et de stockage des données. Comme cela n'est pas toujours possible, la compression vidéo est utilisée pour réduire la taille du fichier en éliminant les données superflues telles que les images dupliquées et la pixellisation.

— **La transmission:** La séquence vidéo enregistrée par les caméras de sécurité doit être envoyée à un centre de commande. Il existe plusieurs moyens de transmission.

— **Le traitement:** Les flux vidéo peuvent être capturés, visualisés, analysés et recherchés une fois qu'ils atteignent l'unité de contrôle, parmi d'autres actions de traitement possibles en fonction de l'utilisation prévue du système de surveillance. Pendant une courte période, certaines infrastructures se contentent de sauvegarder les clips vidéo dans une archive. Nous ne regardons les enregistrements que lorsque nous en avons vraiment besoin. Dans d'autres, des opérateurs humains surveillent un réseau de centaines de caméras. La séquence vidéo des scènes transmises est automatiquement analysée en temps réel par des systèmes de vidéo surveillance intelligents, qui avertissent l'opérateur en cas de suspicion.

3.2.1 Reconnaissance de visages

Le système de reconnaissance faciale permet d'identifier ou d'authentifier des individus à partir d'une image numérique, d'une vidéo ou d'images capturées en direct. d'individus à partir d'une image numérique ou d'une vidéo ou d'images en direct capturées par une caméra. caméra. Différentes techniques sont utilisées dans les systèmes de reconnaissance faciale, mais tous les systèmes suivent le même processus qui comprend la détection des visages, l'extraction des caractéristiques faciales et la comparaison des caractéristiques faciales sélectionnées avec les visages présents dans notre base de données. et la comparaison des caractéristiques faciales sélectionnées avec les visages présents dans notre base de données. Le système de reconnaissance faciale Le système de reconnaissance faciale peut également être décrit comme un système de sécurité biométrique capable d'identifier les propriétés physiques d'un individu, qui sont uniques pour tout être humain et ne peuvent être modifiées. et qui ne peuvent être modifiées.
Bien qu'un système biométrique basé sur la reconnaissance des empreintes digitales ou de l'iris soit plus précis que la reconnaissance faciale. Cette dernière est largement utilisée car le processus de reconnaissance faciale ne nécessite aucune coopération de la part de l'individu, c'est-à-dire qu'il n'y a aucun contact entre le sujet et le système.

La communauté scientifique a largement exploré le traitement de la reconnaissance faciale en temps réel en raison de son importance dans les systèmes de surveillance et de garantie de la sécurité. Les approches traditionnelles et modernes d'apprentissage profond constituent la grande majorité des méthodes proposées dans ce domaine. Le principal objectif de Gupta et al. [Gupta et al., 2016] dans leur travail était d'examiner la possibilité de mettre en œuvre un système de reconnaissance faciale basé sur Raspberry Pi. Ils

ont l'intention d'utiliser des techniques conventionnelles de détection et de reconnaissance des visages telles que la détection Haar [Mita et al., 2005] et l'analyse en composantes principales (ACP) [Abdi and Williams, 2010]. L'objectif était de faire progresser la reconnaissance faciale jusqu'à ce qu'elle puisse remplacer les mots de passe et les cartes RFI pour l'accès aux systèmes et aux bâtiments hautement sécurisés. En utilisant le kit Raspberry Pi, ils se sont efforcés de créer un système rentable et convivial aux performances exceptionnelles.

Nagpa et al.[Nagpal et al., 2018] ont présenté une configuration pour reconnaître les individus suspects, en utilisant un système composé d'un Raspberry Pi Zero lié à un module de caméra Raspberry Pi, un capteur tactile capacitif et un écran OLED. Dans ce système, ils ont utilisé un classificateur Haar Cascade [Wilson and Fernandez, 2006] pour détecter les visages dans une image, suivi par Local Binary Pattern Histogram for facial recognition[Ahonen et al., 2006] (LBPH), mis en œuvre à l'aide d'OpenCV.

Saypadith et al. [Saypadith and Aramvith, 2018] ont présenté un cadre pour la reconnaissance de visages multiples, mis en œuvre dans un système GPU embarqué pour réaliser une reconnaissance en temps réel de plusieurs visages. Ils ont formé une architecture de réseau qui a réduit les paramètres du réseau et incorporé une technique de suivi dans le cadre pour minimiser le temps de traitement tout en maintenant un taux de reconnaissance acceptable. Ils ont utilisé la méthode de reconnaissance faciale FaceNet, qui vise à minimiser l'intégration des images d'entrée de la même personne tout en maximisant la distance entre les intégrations de différents individus. Pour réduire les paramètres du réseau et économiser la mémoire du GPU, ils ont utilisé l'architecture SqueezeNet.

Sajjad et al. [Sajjad et al., 2020] ont présenté un cadre qui combine Raspberry Pi et la technologie cloud pour augmenter les services d'application de la loi dans les villes intelligentes en utilisant la reconnaissance faciale. Pour capturer un flux vidéo, une caméra compacte et sans fil est fixée à l'uniforme d'un agent de police, et ce flux est ensuite dirigé vers un Raspberry Pi pour la détection et la reconnaissance des visages. L'approche proposée consiste à utiliser un sac de mots pour extraire les caractéristiques orientées du test de segment accéléré (FAST) [Viswanathan, 2009] et les points Binary Robust Independent Elementary Features (BRIEF) tournés du visage détecté [Calonder et al., 2010], puis à utiliser une machine à vecteurs de support (SVM) [Cortes and Vapnik, 1995] pour identifier les suspects potentiels. En raison des ressources limitées du Raspberry Pi en termes d'espace de stockage, de mémoire et de puissance de traitement, le classificateur proposé est stocké et entraîné dans le cloud. La méthode est mise en œuvre sur un Raspberry Pi 3 modèle B utilisant Python 2.7 et est testée de manière approfondie sur divers ensembles

de données établis.

Kanagaraj et al. [Kanagaraj et al., 2022] ont présenté un nouveau concept, connu sous le nom de robot espion basé sur Raspberry Pi avec reconnaissance faciale, qui sert d'outil de surveillance applicable dans divers endroits. La mise en œuvre utilise la technologie Wi-Fi pour une surveillance transparente de la zone et un contrôle à distance des mouvements du véhicule robotisé. Le robot est intelligemment construit pour permettre un contrôle sans fil par l'utilisateur via le Wi-Fi, tout en incorporant des capteurs et des capacités de reconnaissance faciale pour améliorer son efficacité en matière de surveillance. Un capteur infrarouge passif (PIR) est utilisé pour détecter les individus ou les obstacles. L'intégration d'une caméra permet au système robotique de capturer des images faciales pour l'identification des intrus. L'algorithme de détection faciale utilisé est le classificateur en Haar cascade, tandis que l'algorithme de reconnaissance faciale est le LBPH.

Kommaraju et al. [Kommaraju et al., 2022] ont proposé un système de reconnaissance de masque facial connecté à une caméra Raspberry Pi, qui enregistre des vidéos à l'entrée principale. Dans un premier temps, la détection des visages est effectuée sur la vidéo enregistrée. Pour détecter les visages, un modèle Caffe, plus précisément un détecteur de visage par réseau neuronal profond (DNN) dans OpenCV, est utilisé. Ensuite, la détection des masques n'est effectuée que lorsqu'un visage est détecté. Une architecture MobileNet [Howard et al., 2017] est utilisée pour développer un modèle qui détermine si une personne porte un masque ou non. Le fonctionnement de la porte est déterminé en fonction du résultat de la classification. Si la classification indique que la personne porte un masque facial, la barrière s'ouvre, permettant l'accès à la zone publique. À l'inverse, si la personne ne porte pas de masque facial, la barrière reste fermée. La technologie de l'internet des objets (IoT) est utilisée pour réguler le fonctionnement du portail en fonction du résultat de la classification.

Une application web a été développée par Coronel et al. [Coronel et al., 2022] pour permettre le contrôle sans contact de l'accès du personnel à une zone de travail, ce qui la rend particulièrement utile pour lutter contre l'urgence sanitaire Covid-19. Des techniques d'apprentissage profond et de vision par ordinateur ont été utilisées pour la détection et la reconnaissance des visages dans sa mise en œuvre. Le système comprend quatre phases distinctes. La phase initiale se concentre sur la détection et l'alignement des visages à l'aide d'algorithmes d'apprentissage profond. La deuxième phase consiste à extraire les caractéristiques du visage pour permettre la reconnaissance de différents individus. La troisième phase intègre un module qui détecte l'usurpation d'identité, empêchant ainsi les attaques potentielles sur le système en faisant la distinction entre les vrais et les faux

visages. La dernière phase comprend la conception et le développement d'une interface web, qui facilite la communication entre les algorithmes, les utilisateurs et l'administration.

Meddeb et al. [Meddeb et al., 2023] ont proposé un prototype de robot de surveillance mobile à faible coût basé sur Raspberry Pi 4, adapté à l'intégration dans des environnements industriels. Ce robot intelligent utilise l'IoT et la technologie de reconnaissance faciale pour détecter la présence d'intrus. Équipé d'un capteur PIR et d'une caméra, il capture des vidéos et des photos en direct, qui sont transmises à la salle de contrôle via l'IoT. Les algorithmes de reconnaissance faciale intégrés au système permettent de différencier le personnel de l'entreprise des intrus. Le traitement des images utilise le classificateur Haar Cascade et l'algorithme LBPH pour détecter et identifier les individus. Dès qu'un étranger est détecté, des notifications d'alerte et des courriels contenant l'image capturée sont envoyés à la salle de contrôle via l'IoT. Pour faciliter le contrôle à distance du robot, une interface web est développée, permettant une connectivité WiFi. Cette application permet de surveiller une zone étendue et d'améliorer le système de perception du robot. L'efficacité et la robustesse des algorithmes de reconnaissance faciale utilisés dans ce cadre ont été évaluées par le biais de tests de performance.

Un nouvel algorithme de reconnaissance des visages a été proposé par Sharifisoraki et al. [Sharifisoraki et al., 2023], qui a utilisé des algorithmes de détection de repères faciaux basés sur l'apprentissage profond pour extraire des caractéristiques clés des images. En utilisant les points de repère obtenus à partir d'un réseau neuronal profond, trois caractéristiques distinctes ont été extraites à l'aide de valeurs spécifiques (SV) et utilisées pour la reconnaissance des visages. La nouveauté de ce travail réside dans l'utilisation des valeurs spécifiques pour la reconnaissance faciale. Les valeurs spécifiques, notamment la distance cosinusoïdale, les angles et les surfaces, ont été dérivées des coordonnées des points de repère. Les taux de reconnaissance des visages utilisant les points de repère extraits et les SV comme caractéristiques proposées ont été évalués par de multiples expériences. L'effet de l'augmentation du nombre de points de repère sur le taux de reconnaissance des visages a été étudié à l'aide de l'algorithme de maillage des visages MediaPipe.

3.3 La détection de visages et la reconnaissance de visages

La conception d'un système de vidéo surveillance intelligent implique l'intégration de deux éléments primordiaux: la détection et la reconnaissance de visages.

De manière générale, un tel système se compose de trois modules comme décrits dans la figure 3.2: détection du visage, utilisation du modèle de deep learning pour l'extraction des caractéristiques et enfin la reconnaissance des visages.

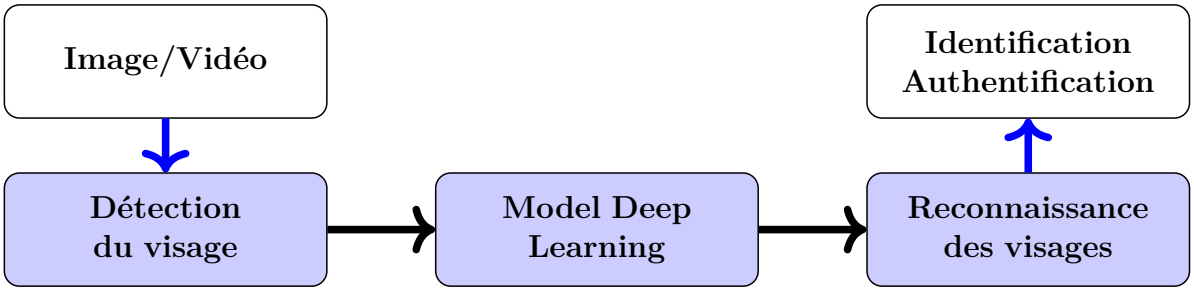


FIGURE 3.2 – Fonctionnement d'un système de reconnaissance de visages.

3.3.1 Détection de visages

La détection de visages est une tâche primordiale dans les domaines de la surveillance et de la sécurité. La première étape consiste à identifier la zone de l'image ou de la vidéo qui contient un visage, ainsi qu'à localiser précisément la position de chaque visage. Cette étape permet ensuite de convertir les données en fenêtres qui contiennent chaque visage en tant qu'image d'entrée. Pour y parvenir, les systèmes de détection de visages utilisent des caractéristiques extraites de points de repère faciaux pour segmenter l'image. En général, ces systèmes extraient uniquement une partie du visage, ce qui permet d'éliminer les zones non pertinentes comme l'arrière-plan ou les cheveux.
En 2015, Liu et son équipe ont développé le SSD (Single Shot MultiBox Detector) [Liu et al., 2016], également connu sous le nom de "Single Shot MultiBox Detector". Ce détecteur est à la fois rapide et précis, comme l'indique son nom. La figure 3.3 illustre un exemple de détection de visage réussie, avec une vue de face d'un visage humain on utilisant le SSD.

— **Single Shot:** Le terme "Single Shot" signifie que les tâches de classification et de localisation des objets sont effectuées en une seule passe dans le réseau, sans nécessiter plusieurs étapes séparées.

— **MultiBox:** MultiBox fait référence à une technique de régression utilisant des boîtes englobantes.

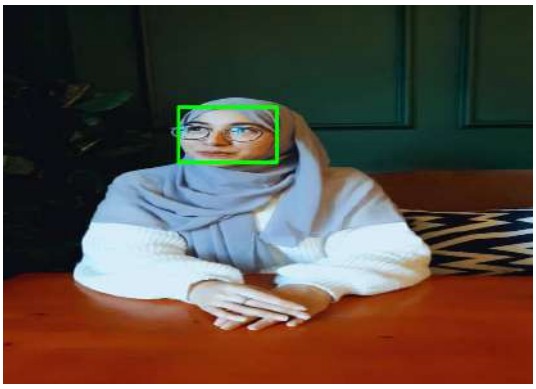


FIGURE 3.3 – Détection de visage dans une image utilisant le SSD

— **Detector:** le terme Detector indique que le réseau est capable de détecter les objets présents dans une image et de les classer en même temps.

Le réseau SSD est composé de trois parties principales. Tout d'abord, l'image d'entrée est introduite dans un réseau classifieur appelé "réseau de base". Dans ce travail, le réseau VGG 16 est utilisé à cette fin, mais les couches entièrement connectées de l'architecture standard sont remplacées par des couches de convolution. Ensuite, la sortie du réseau de base modifié est introduite dans une série de couches de caractéristiques supplémentaires qui diminuent en taille, permettant ainsi la prédiction de détections à plusieurs échelles. Enfin, les sorties de ces couches de caractéristiques supplémentaires, ainsi qu'une couche sélectionnée du réseau de base, sont utilisées comme prédicteurs de détection, chaque couche pouvant produire un ensemble fixe de prédictions à l'aide d'un nouvel ensemble de couches convolutives.

Le MultiBox utilise un ensemble fixe de boîtes d'ancrage par défaut comme références pour les prédictions, puis tente d'approximer les boîtes de délimitation de la vérité terrain en utilisant ces références. Pour chaque couche de caractéristiques, des boîtes d'ancrage par défaut sont calculées pour un ensemble donné de rapports d'aspect et d'échelles. Le nombre de boîtes d'ancrage pour chaque couche est indiqué dans le tableau 3.1.

3.4 La reconnaissance de visages par transfert learning

Une fois les visages détectés, les images faciales sont extraites puis utilisées pour connaître l'identité d'une personne en utilisant l'algorithme d'apprentissage Convolutional

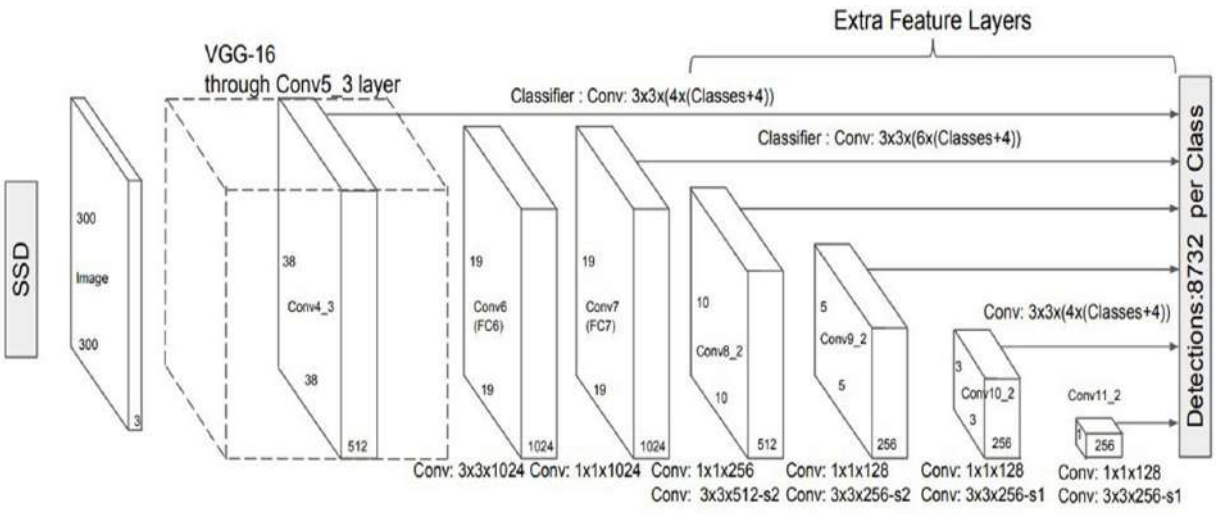


FIGURE 3.4 – Architecture du réseau SSD 300

TABLE 3.1 – Nombre de boites pour le réseau SSD 300

Couche	**Height**	**Width**	**# boites**
`conv4_3`	38	38	4
`conv7`	19	19	6
`conv8_2`	10	10	6
`conv9_2`	5	5	6
`conv10_2`	3	3	4
`conv11_2`	1	1	4

Neural Network (CNN).

3.4.1 Le MobileNet pour la reconnaissance faciale

L'objectif principal de l'architecture de réseau neuronal convolutif connue sous le nom de MobileNet est d'améliorer la qualité des applications de vision mobiles et embarquées. Cette conception permet non seulement de produire des réseaux compacts visant à améliorer les performances du réseau tout en réduisant le nombre de paramètres.

La figure 3.5 montre une vue d'ensemble du réseau de neurones convolutifs que nous avons utilisé.

L'architecture de MobileNetV2 se compose d'une série de blocs et de couches organisés

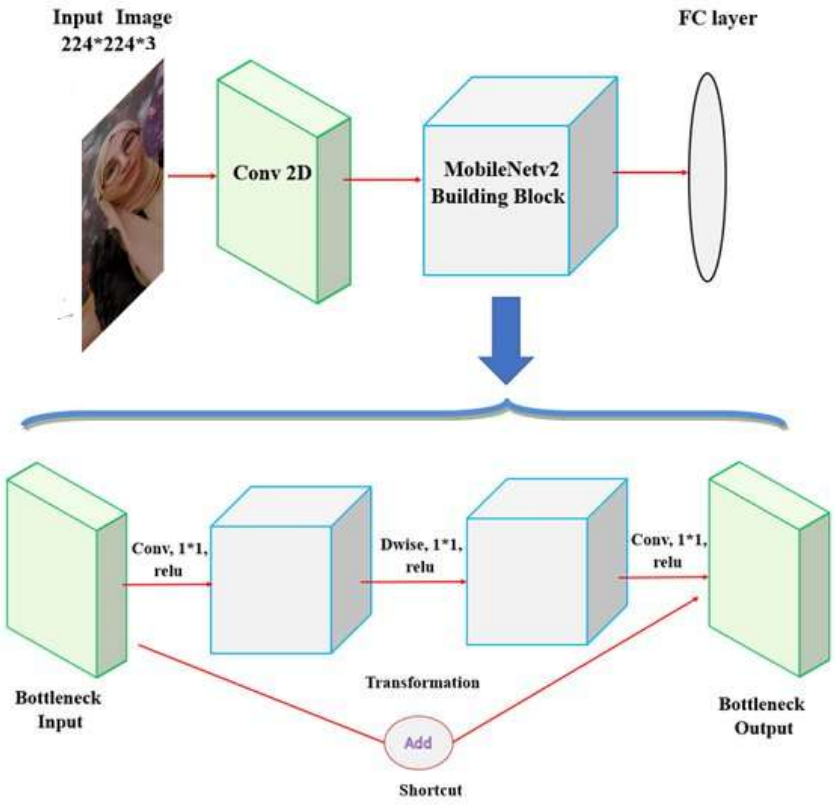


FIGURE 3.5 – Réseau de neurones MobilenetV2 [Hashmi et al., 2020]

de manière séquentielle. Voici un aperçu de l'architecture :

— Input Layer: MobileNetV2 prend en entrée une image de taille fixe.

— Initial Convolutional Layer: L'image d'entrée passe par une couche convolutive standard avec un pas de 2 pour réduire les dimensions spatiales de l'image.

— Inverted Residual Blocks: MobileNetV2 se compose de plusieurs blocs résiduels inversés, qui sont les principaux éléments constitutifs du réseau. Chaque bloc résiduel inversé suit une structure spécifique :

1. Pointwise Convolution: Une convolution ponctuelle 1x1 est appliquée pour réduire le nombre de canaux d'entrée.
2. Depthwise Convolution: Une convolution séparable en profondeur est effectuée sur les canaux réduits. Elle implique une convolution en profondeur (convolution au sein de chaque groupe de canaux) suivie d'une convolution ponctuelle (convolution 1x1 pour mélanger les informations entre les canaux).
3. Linear Bottleneck: Une couche de goulot d'étranglement linéaire est ajoutée

pour permettre un flux d'informations efficace et un apprentissage résiduel. Elle utilise une fonction d'activation linéaire (pas de non-linéarité) pour préserver l'information.

4. Skip Connection: Une connexion de saut (connexion résiduelle) est ajoutée de l'entrée du bloc à la sortie pour faciliter le flux d'informations.

— Feature Pyramid: MobileNetV2 incorpore plusieurs blocs résiduels inversés avec différentes résolutions et multiplicateurs de largeur pour créer une pyramide de caractéristiques. Cela permet au réseau de capturer des caractéristiques à plusieurs échelles, depuis les détails les plus fins jusqu'aux représentations de haut niveau.

— Final Layers: La pyramide des caractéristiques est suivie d'une couche de mise en commun de la moyenne globale qui réduit les dimensions spatiales des cartes des caractéristiques à une taille fixe. Ensuite, une couche entièrement connectée est appliquée pour générer les prédictions de classification finales.

— Output Layer: La couche de sortie utilise une fonction d'activation softmax pour produire la distribution de probabilité sur les différentes classes.

MobileNetV2 inclut également des considérations architecturales supplémentaires telles que le multiplicateur de largeur et le multiplicateur de résolution, qui permettent aux utilisateurs de contrôler la taille du modèle et l'efficacité des calculs. En ajustant ces multiplicateurs, le nombre de canaux et la résolution de l'image d'entrée peuvent être adaptés pour répondre à des besoins spécifiques.
Dans l'ensemble, l'architecture de MobileNetV2 est conçue pour trouver un équilibre entre la précision et la taille du modèle, ce qui la rend adaptée aux appareils mobiles et embarqués dotés de ressources informatiques limitées.

3.4.2 Environnement de travail

Pour assurer la mise en place de l'environnement de travail de notre projet, nous recommandons l'utilisation de Google Colaboratoryn également connu sous le nom de Google Colab. Ce service cloud gratuit prend en charge le GPU gratuit pour évaluer les compétences en langage de programmation Python. Google Colab est un environnement de cahier de notes Jupyter, composé de cellules qui peuvent contenir du code, du texte et des figures. Il ne nécessite aucune installation ou configuration préalable et qui offre un accès gratuit aux GPU et TPU.

3.4.3 Base de données et protocole d'évaluation

Dans ce projet, le modèle de base, qui est le réseau MobileNetV2, a été pré-entraîné à l'aide de l'ensemble de données ImageNet composé d'environ 1000 classes d'objets, 1281167 images d'apprentissage, 50000 images de validation et 100000 images de test [Deng et al., 2009].

Le processus du transfer learning commence par la collecte des données. Dans notre cas, nous avons utilisé Pinterest Faces (PINS), une base de données utilisée pour la reconnaissance faciale. 17534 images de visages de 105 célébrités enregistrées sur la plateforme de réseaux sociaux Pinterest constituent l'ensemble de données PINS avec des variation de pose, de condition d'éclairage et d'émotion. La figure 5.3 montre quelques exemples de cette base. Toutes les images sont redimensionnées à 224×224 de manière à ce que

FIGURE 3.6 – Exemple de visages de la base de données PINS

les visages aient la même taille que l'entrée requise par le modèle utilisé. L'ensemble de

données est normalisé en soustrayant les valeurs moyennes des pixels et en les divisant par l'écart-type.

Une fois le modelé téléchargé et la base de données répartie, des paramètres d'initialisation tels que la taille du batch, le nombre d'époques, le taux d'apprentissage, doivent être ajustés pour obtenir de meilleures performances. La taille du batch est le nombre d'images dans un lot (batch). Le nombre d'époques se définie comme étant le nombre de fois que toutes les données d'apprentissage sont passées par le réseau. Le taux d'apprentissage est un paramètre très sensible qui pousse le modèle vers la convergence. La recherche de sa meilleure valeur se fait par un processus expérimental. Dans ce travail, l'étape d'apprentissage est effectuée avec des hyper paramètres définis comme Batch_size = 32 et Epochs = 200. Enfin, le modèle est compilé en appelant la fonction **compile()**, qui a comme entrées :

- **optimiser** utilisé pour optimiser le réseau neuronal, tel que Adagrad et Adam.

- **loss** qui fait référence à la valeur que le modèle doit minimiser, comme l'entropie croisée et l'erreur quadratique moyenne.

- **Metrics** qui évaluent la précision des problèmes de classification.

Le modèle est entraîné à l'aide de la méthode **fit()**. Cette méthode permet au modèle d'itérer sur les données et de trouver le réseau neuronal le plus optimal pour ces données. La figure 3.7 montre l'évolution du taux d'apprentissage et de l'erreur en fonction du nombre d'époques.

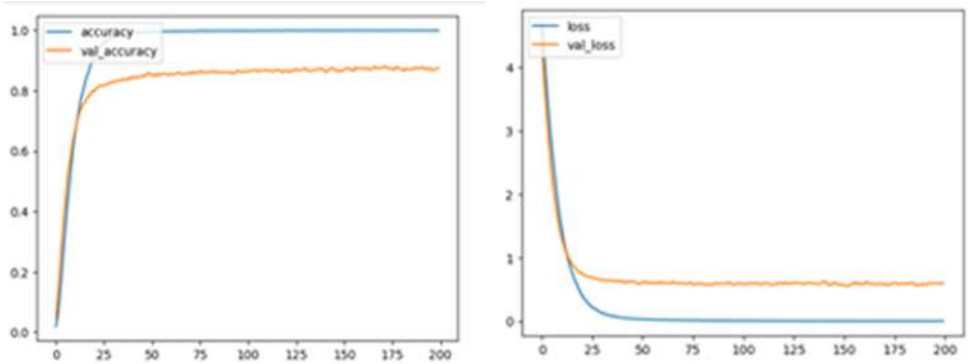


FIGURE 3.7 – Le taux d'apprentissage et l'erreur en fonction du nombre d'époques

Quelques résultats de reconnaissance sont présentés dans la figure 3.8

FIGURE 3.8 – Exemple de visages connu et inconnu

3.5 Conclusion

Dans ce chapitre, nous avons présenté les modèles basés sur les réseaux de neurones convolutifs que nous avons mis en oeuvre.
L'utilisation du modèle SSD pour la détection et de MobileNet V2 pour la reconnaissance nous a fourni des résultats précis et efficaces. Ces techniques offrent de vastes possibilités d'applications et sont essentielles pour le développement de systèmes de vision par ordinateur performants.

Deuxième partie

Partie Pratique

Chapitre 4

Environnement de Développement

Sommaire

4.1 Introduction

L'objectif de notre étude est de mettre en place un système de détection et de reconnaissance par vidéo-surveillance. Dans le chapitre précédent, nous avons effectué une analyse théorique. Dans ce chapitre, nous allons utiliser des outils matériels et logiciels pour mettre en œuvre notre système.

4.2 Environnement hardware

4.2.1 Présentation du Raspberry Pi

Le Raspberry Pi a été délibérément construit comme un ordinateur très flexible et puissant à une fraction des coûts d'un PC traditionnel à utiliser par quiconque pour résoudre des problèmes de manière créative. Son grand nombre d'actifs l'emporte facilement sur ses limites et font du Raspberry Pi un excellent outil de recherche qui peut être utilisé pour presque tout. Cela peut aller de la surveillance environnementale interactive et autonome à l'enregistrement vidéo d'expériences en laboratoire, aux stations de mesure de terrain à long terme et aux dispositifs avancés en boucle fermée capables de lire diverses entrées, de déclencher d'autres actions (par exemple, pour allumer et éteindre des lumières ou des servomoteurs et de traiter automatiquement les données et d'envoyer des messages d'avertissement.[Jolles, 2021]

4.2.1.1 Historique

Le Raspberry Pi est un Single Board Card (SBC) à faible coût développé par le Raspberry Pi Fondation (raspberrypi.org), une organisation caritative basée au Royaume-Uni. Depuis sa première sortie en 2012, plusieurs générations d'ordinateurs Raspberry Pi ont été publiées, qui peuvent être classées en trois modèles : Le Raspberry Pi A, B et Zero (un quatrième modèle, le Compute Module, est principalement utilisé dans les applications industrielles). Les fondamentaux de ces trois modèles (désormais 'Raspberry Pi's') sont très similaires, chacun étant doté d'un système sur puce composé d'un processeur intégré (unité centrale de traitement) et d'une unité de traitement graphique (GPU) on chip, d'une mémoire intégrée et d'une alimentation entrée de 5 V DC. Tous les modèles disposent également d'un port pour connecter une caméra dédiée, ainsi que d'un ensemble de broches d'entrée/sortie (GPIO) à usage général qui peuvent être utilisées pour communiquer avec une large gamme d'appareils électroniques, des LEDs et boutons aux servomoteurs et aux

relais de puissance et une vaste gamme de capteurs, les cartes d'extension spéciales qui se connectent aux broches GPIO, appelées Hardware Attached on Top (HAT), peuvent fournir d'autres fonctionnalités, allant de la gestion de l'alimentation, à l'identification par radiofréquence (RFID), contrôleurs de moteur et enregistrement audio de haute qualité.La plupart des modèles disposent également d'une connexion Ethernet et la connectivité sans fil (Wi-Fi et Bluetooth), qui, en combinaison avec les ports GPIO, donnent au Raspberry Pi une grande polyvalence.

Le Raspberry Pi possède toutes les fonctionnalités d'un ordinateur standard. En tant que tel, vous pouvez connecter une souris, un clavier et un écran sans aucune configuration et avoir le contrôle sur un environnement de bureau Linux facile à utiliser ou d'autres systèmes d'exploitation populaires, y compris Windows 10 IoT et Android. Le Raspberry Pi peut également être utilisé comme une unité sans tête (pas de clavier, de souris et d'écran connectés) qui peut être contrôlé et programmé à distance pour exécuter des scripts de manière autonome en utilisant un large éventail de langages informatiques.Le Raspberry Pi est différent d'un microcontrôleur, comme l'Arduino ou le récent a publié Raspberry Pi Pico, qui peut être programmé pour exécuter un seul programme écrit par l'utilisateur et communiquer avec des capteurs et d'autres composants électroniques.[Jolles, 2021]

La figure 4.1 présente les différentes générations du Raspberry Pi.

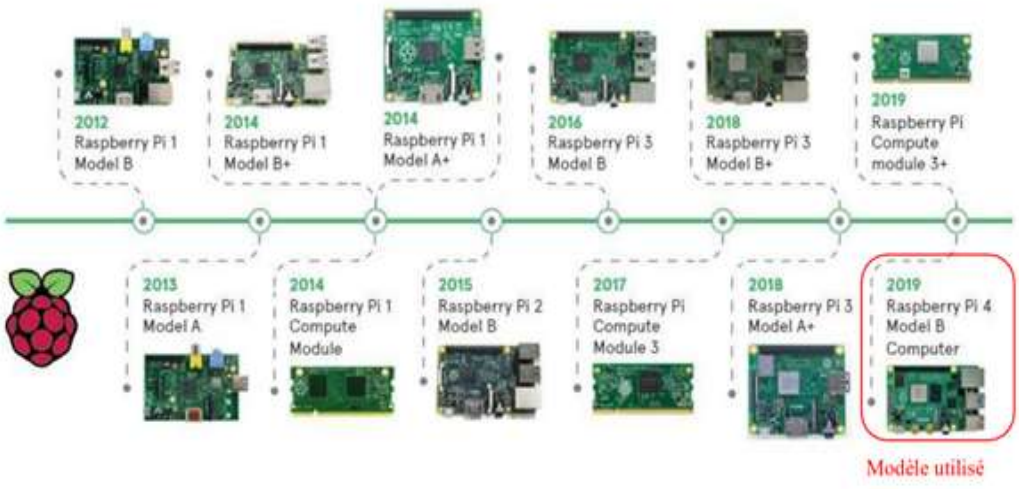


FIGURE 4.1 – Evolution du Raspberry Pi

4.2.1.2 Composant de base

Le Raspberry Pi 4 Model B+ est le plus récent ajout à la gamme populaire d'ordinateurs Raspberry Pi. Il représente une avancée significative par rapport à la génération précédente, le Raspberry Pi 4 Model B, grâce à des améliorations majeures dans la vitesse du processeur, les performances multimédia, la mémoire et la connectivité. De plus, il conserve une compatibilité ascendante avec les modèles précédents et une consommation d'énergie similaire. En raison de ces améliorations, le Raspberry Pi 4 Model B+ offre des performances de bureau comparables à celles des PC x86 d'entrée de gamme.

4.2.1.3 Spécifications

- **processeur:** Broadcom BCM2711, Quad core Cortex-A72 (ARM v8) SoC 64 bits @ 1,5 GHz
- **Mémoire:** 8 Go de mémoire SDRAM LPDDR4-3200
- **Connectivité:**
  - ✓ Wi-Fi : IEEE 802.11ac 2,4 GHz et 5,0 GHz sans fil
  - ✓ Bluetooth: Bluetooth 5.0, BLE
  - ✓ Ethernet: Gigabit Ethernet
  - ✓ USB: 2 ports USB 3.0; 2 ports USB 2.0
  - ✓ **GPIO :** connecteur GPIO standard à 40 broches (entièrement rétro compatible avec les cartes précédentes)
- **Vidéo et son:**
  - ✓ HDMI : 2 ports micro-HDMI × (jusqu'à 4kp60 pris en charge)
  - ✓ Port d'affichage MIPI DSI à 2 voies
  - ✓ Port caméra MIPI CSI à 2 voies
  - ✓ Port audio stéréo 4 pôles et port vidéo composite
- **Multimédia:**
  - ✓ H.265 (décodage 4Kp60)
  - ✓ H.264 (décodage 1080p60, codage 1080p30)
  - ✓ Graphisme OpenGL ES, 3.0
- **Support de carte SD :** Fente pour carte Micro SD pour le chargement du système d'exploitation et le stockage des données.

- **Alimentation en entrée :**
  - ✓ 5V DC via le connecteur USB-C (minimum 3A)
  - ✓ 5V DC via l'embase GPIO (minimum 3A)
  - ✓ Compatible avec l'alimentation par Ethernet (PoE) (nécessite un adaptateur PoE séparé)
- **Dimensions :**85.6mm × 56.5mm
- **Environnement :** Température de fonctionnement 0–50ºC.

Le Raspberry Pi a initialement son propre système d'exploitation précédemment appelé Raspbian basé sur Linux. Dans le monde émergent du logiciel, il y a peu de non-Linux options de système d'exploitation disponibles sur le marché .Les systèmes d'exploitation préférés pour le Pi sont la distribution Linux (Debian, Puppy Linux, Arch Linux, Fedora Remix et OpenELEC) car ils sont facilement disponibles gratuitement, mais principalement en raison de leur capacité à fonctionner sur le processeur ARM du Raspberry Pi[Ghael et al., 2020]. La figure 4.2 donne une description détaillée des diffèrent composants d'une carte Raspberry Pi 4 model B+.

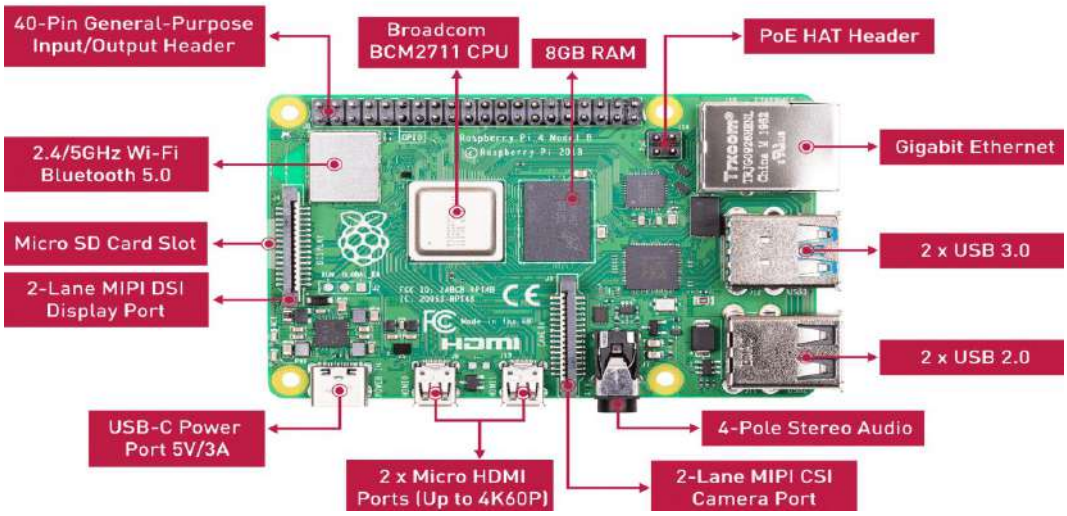


FIGURE 4.2 – Ecran TFT LCD 7" pour Raspberry

4.2.2 Ecran TFT LCD 7

Un écran à cristaux liquides à transistor à couche mince (TFT LCD) est une variante d'un écran à cristaux liquides qui utilise la technologie des transistors à couche mince pour améliorer les qualités d'image telles que l'adressabilité et le contraste. Un écran LCD TFT est un écran LCD à matrice active, contrairement aux écrans LCD à matrice passive ou aux écrans LCD simples à commande directe (c'est-à-dire avec des segments directement connectés à l'électronique en dehors de l'écran LCD) avec quelques segments.
Les écrans LCD TFT sont utilisés dans les appareils tels que les téléviseurs, les écrans d'ordinateur, les téléphones mobiles, appareils portables, systèmes de jeux vidéo, assistants numériques personnels, systèmes de navigation, projecteurs et tableaux de bord dans les automobiles(voir la figure 4.3)

FIGURE 4.3 – Ecran TFT LCD 7" pour Raspberry

4.2.3 La caméra Pi

Le module caméra haute résolution, compatible avec tous les modèles Raspberry Pi, permet de capturer des images à haute sensibilité et à faible bruit dans un design ultra petit et léger. Il est connecté à la carte Raspberry Pi via un connecteur d'interface série de caméra (CSI), conçu spécialement pour se connecter à des caméras capables de débits de données très élevés et pour fournir exclusivement des données de pixel au processeur(voir la figure 4.4).

4.2.3.1 Spécifications

- Résolution 8 mégapixels
- Résolution des images fixes 3280 x 2464
- Taux de transfert d'image maximum 1080p: 30fps (codage et décodage) 720p: 60 images par seconde
- Connexion au Raspberry Pi Câble plat 15 broches, à l'interface série dédiée MIPI 15 broches pour caméra
- Balance des blancs automatiques
- Filtre de bande automatique
- Détection automatique de la luminance 50/60 Hz
- Calibrage automatique du niveau de noir
- Plage de température de fonctionnement : -20º à 60º
- Image stable : De -20º à 60º
- Taille de l'objectif 1/4"
- Dimensions 23,86 x 25 x 9mm
- Poids 3g.

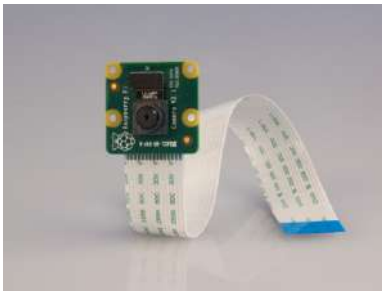


FIGURE 4.4 – Raspberry Pi Camera Module 2

4.2.4 Carte Micro SD

Une carte SD (Secure Digital) est un petit support de stockage amovible utilisé pour stocker des données numériques. Les cartes SD sont largement utilisées dans les appareils électroniques, tels que les appareils photo numériques, les smartphones, les tablettes et les ordinateurs portables.

Dans le contexte de l'informatique embarquée, les cartes SD sont souvent utilisées pour stocker le système d'exploitation et les programmes sur des appareils tels que les Raspberry Pi et les Arduino. Les cartes SD sont pratiques car elles sont compactes, amovibles et peuvent être facilement échangées entre les appareils(voir la figure 4.5).

FIGURE 4.5 – Carte micro SD

Il est important de choisir une carte SD de qualité pour garantir des performances optimales et une fiabilité à long terme. Les cartes SD ont des capacités de stockage variables, allant de quelques gigaoctets à plusieurs téraoctets. Il est recommandé de choisir une carte SD avec une capacité de stockage suffisante pour répondre aux besoins spécifiques de projet.

4.3 Configuration matérielle

Pour ce projet, nous avons choisi d'utiliser un Raspberry Pi 4, qui utilise une carte mémoire microSD pour le démarrage et le chargement du système d'exploitation. Dans les paragraphes suivants, nous allons détailler les étapes nécessaires pour l'installation initiale et les mises à jour du système d'exploitation.

4.3.1 Téléchargement de l'image de distribution

Pour commencer, nous téléchargeons gratuitement l'image de distribution Raspberry Pi OS à partir du site officiel du fabricant Raspberry Pi sur notre ordinateur.Nous avons téléchargé la version de distribution du système d'exploitation Raspberry Pi : Raspberry Pi OS (64-bit) avec bureau .

4.3.2 Enregistrement du contenu de l'image sur la carte SD

Pour enregistrer le contenu de l'image sur la carte SD, nous utilisons un logiciel spécialement conçu à cet effet. Nous utilisons l'application Raspberry Pi Imager pour assurer une compatibilité optimale.

En cliquant sur le bouton "CHOOSE OS", nous sélectionnons le dossier contenant l'image du système d'exploitation du Raspberry Pi, en précisant le fichier spécifique.

Ensuite, en utilisant le bouton "CHOOSE STORAGE", nous indiquons l'unité où se trouve la carte SD que nous voulons utiliser.

Enfin, en cliquant sur le bouton "WRITE", nous lançons le processus d'enregistrement.

Une fois le processus terminé, un message s'affiche pour indiquer que l'opération a été effectuée avec succès. La carte SD peut alors être retirée et insérée dans l'emplacement prévu à cet effet sur le Raspberry Pi.

4.3.3 Configuration initiale

Par défaut, le Raspberry Pi est configuré avec le nom d'hôte "raspberrypi", qui est visible sur le réseau. En tant que première étape, nous avons renommé notre Raspberry Pi. Pour ce faire, veuillez vous référer à la figure 4.6 illustrative qui montre la procédure :

1. connecter au terminal .
2. Exécutez la commande "sudo raspi-config".
3. Sélectionnez l'option "System Options".
4. Choisissez "Hostname" et appuyez sur Entrée. Si vous recevez un avertissement vous indiquant de ne pas utiliser de caractères spéciaux dans le nom d'hôte du Raspberry Pi, sélectionnez simplement "Ok".
5. Saisissez le nouveau nom d'hôte et appuyez sur "Ok".

Le nom de notre Raspberry Pi 4 est "IkhlaseRadjaa FACI M2INSTRUMENTATION2023 UAT".

Par mesure de sécurité, SSH est désactivé par défaut sur le Raspberry Pi, même s'il est installé. Dans notre configuration, la deuxième étape consiste à activer SSH. Veuillez vous référer à la figure 4.7 illustrative pour effectuer cette opération.

1. connecter au terminal.
2. Exécutez la commande "sudo raspi-config".
3. Sélectionnez l'option "Interface Options".

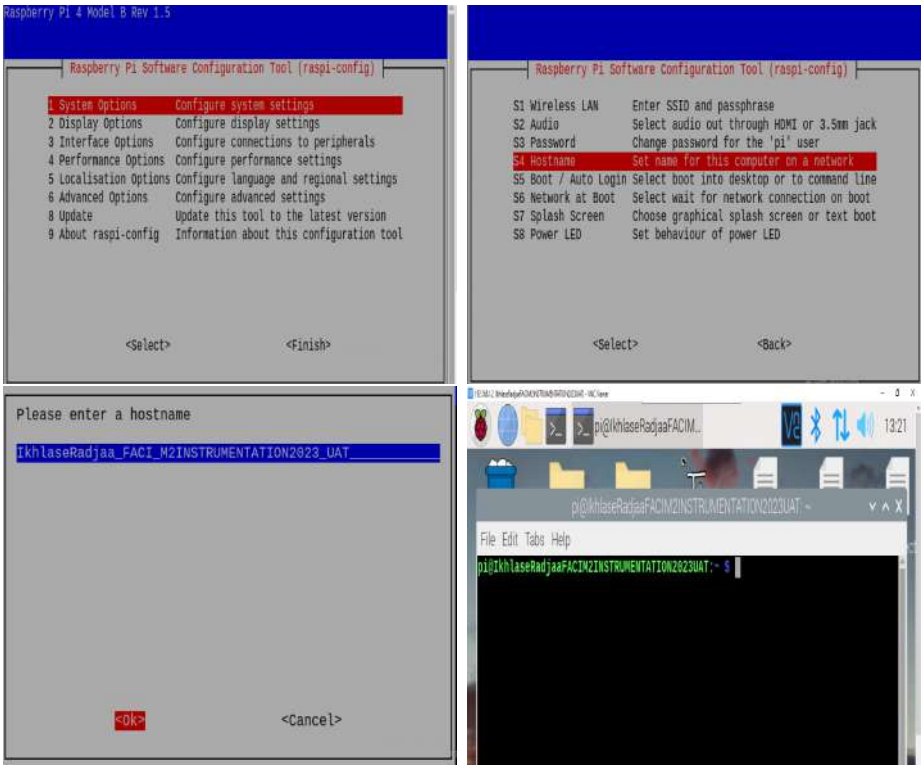


FIGURE 4.6 – Modification du nom d'hôte

4. Dans le menu déroulant, choisissez "SSH".
5. Cliquez sur "Enable" pour activer la fonctionnalité.
6. Utilisez le bouton "Ok" pour confirmer le message.
7. Utilisez le bouton "Finish" pour terminer la configuration.

Après avoir activé SSH sur notre Raspberry Pi, il est nécessaire de télécharger et d'installer un logiciel client SSH sur notre ordinateur. Étant donné que notre ordinateur fonctionne sous Windows, la solution recommandée consiste à télécharger et installer PuTTY. PuTTY est un logiciel open source développé et maintenu par une communauté d'utilisateurs. Une particularité de VNC est sa capacité à contrôler un ordinateur distant tout en visualisant son bureau. Grâce à cela, nous pouvons observer en temps réel ce qui se passe sur notre Raspberry Pi sans avoir à le connecter à un écran physique. Comme illustré dans la figure 4.8, l'activation de VNC se réalise de la même manière que pour SSH, à la différence près qu'il faut sélectionner VNC dans le menu "Interface Options".

La dernière étape consiste à activer la caméra Raspberry Pi (Raspicam) en suivant les étapes ci-dessous (voir la figure 4.9)

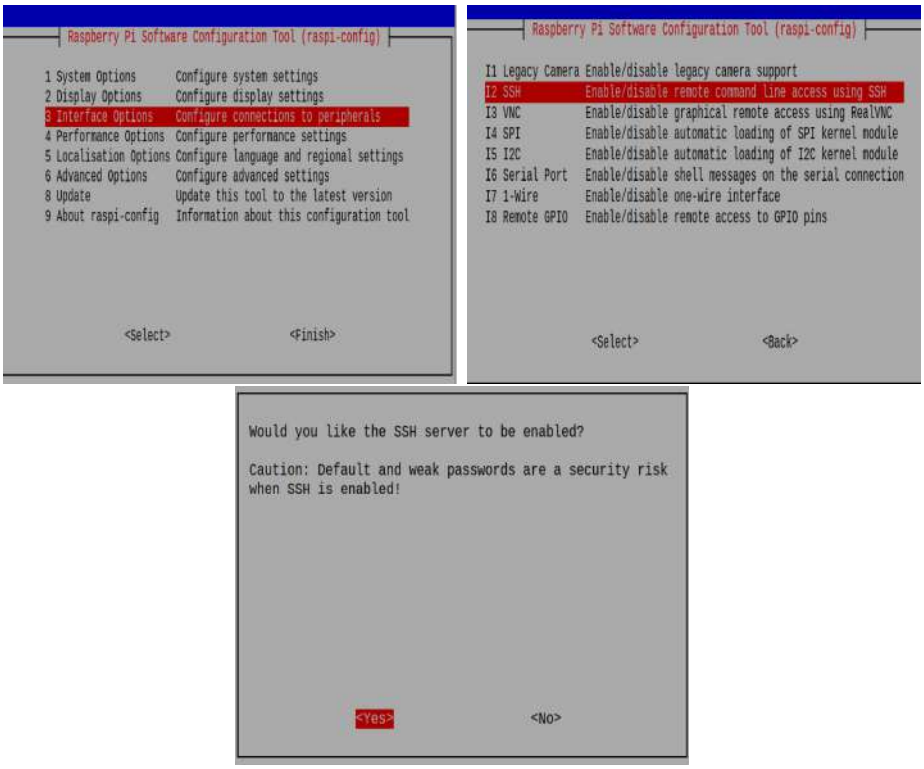


FIGURE 4.7 – Activation du SSH

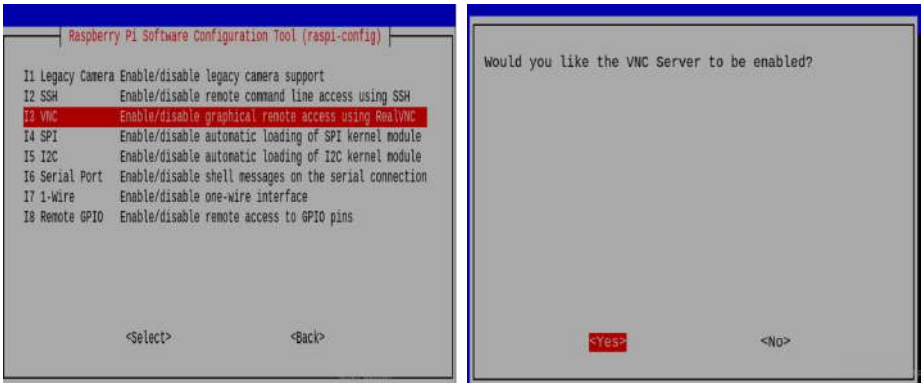


FIGURE 4.8 – Activation du VNC

1. Connecter au terminal.
2. Exécutez la commande "sudo raspi-config".
3. Sélectionnez l'option "Interface Options".
4. Dans le menu déroulant, choisissez "Legacy Camera" (Caméra classique).

5. Cliquez sur "Enable" pour activer la fonctionnalité.
6. Utilisez le bouton "Ok" pour confirmer le message.
7. Utilisez le bouton "Finish" pour terminer la configuration.

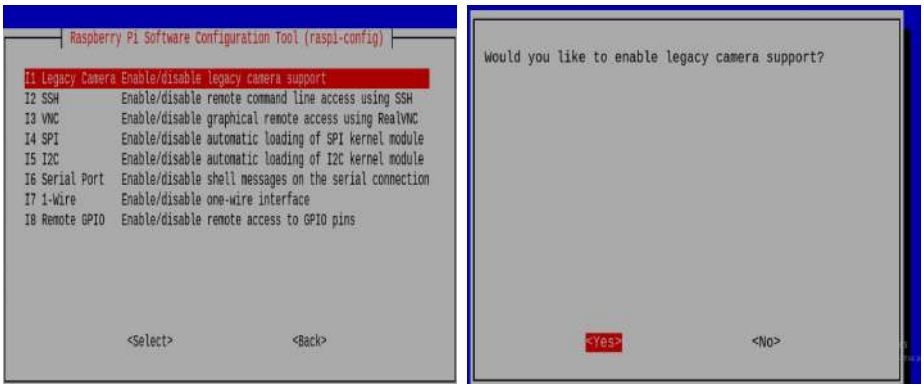


FIGURE 4.9 – Activation de le caméra

Il est essentiel et nécessaire de maintenir le système à jour pour s'assurer que tous les pilotes fonctionnent dans de bonnes conditions. Cette opération peut être effectuée très simplement en ouvrant le terminal et en tapant les commande sudo apt get update. La Raspberry Pi se connectera au référentiel hébergeant les paquets (fichiers, logiciels) de la distribution Raspbian installée, afin de récupérer les dernières versions.

4.4 Environnement de travail

Une illustration du Desktop de la Raspberry Pi 4 est montrée dans la figure 4.10

4.4.1 Langage Python

Python est un langage de programmation destiné aux programmeurs débutants. Il est facile à apprendre grâce à sa structure simple et à sa syntaxe bien définie, et il est facile à lire grâce à son code clairement défini. Il est facile à maintenir et à mettre à jour, et dispose d'une vaste bibliothèque standard qui est portable et compatible avec Windows, UNIX et Mac. C'est un excellent langage de programmation pour les débutants, car il prend en charge un large éventail d'applications, du traitement de texte simple à la programmation de jeux.

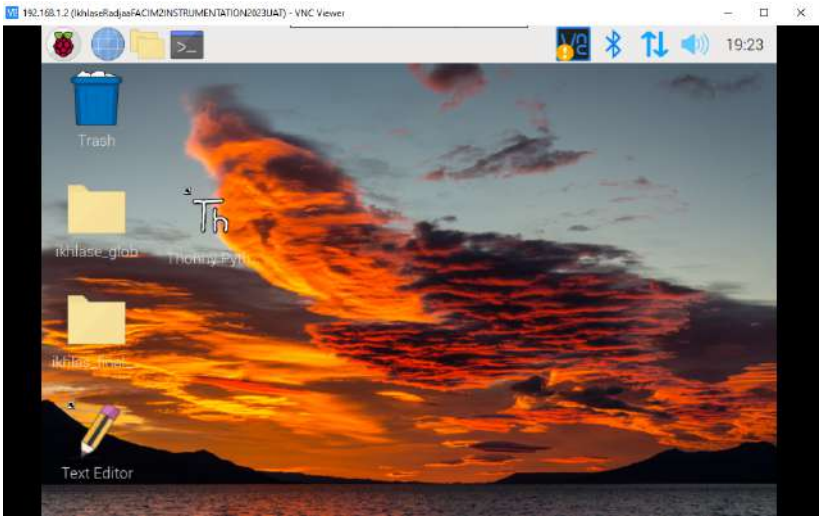


FIGURE 4.10 – illustration du Desktop de la Raspberry Pi

4.4.2 L'IDE Thonny

Thonny est un nouvel IDE Python pour l'apprentissage et l'enseignement de la programmation qui peut rendre la visualisation de programme naturelle fait partie du flux de travail des débutants. Parmi ses principales caractéristiques figurent différentes façons de parcourir le code, l'évaluation d'expression étape par étape, la visualisation intuitive de la pile d'appels et le mode d'explication des concepts de références et de tas. Thonny est un IDE Python utilisé pour développer des programmes Python (voir la figure 4.11). Vous pouvez écrire des scripts Python et les enregistrer dans des fichiers Python.Thonny fournit également un outil en cours d'exécution afin que vous puissiez voir la sortie du programme à partir d'un IDE. L'IDE Thonny est installé par défaut sur le bureau Raspbian Scratch. Vous pouvez le trouver dans le menu Programmation.

4.4.3 Bibliothèques utilisées

4.4.3.1 Math

Le module mathématique est un composant intégré standard de Python qui est toujours accessible. Afin d'utiliser les fonctions mathématiques fournies par ce module, on importe la bibliothèque en utilisant la déclaration "import math".

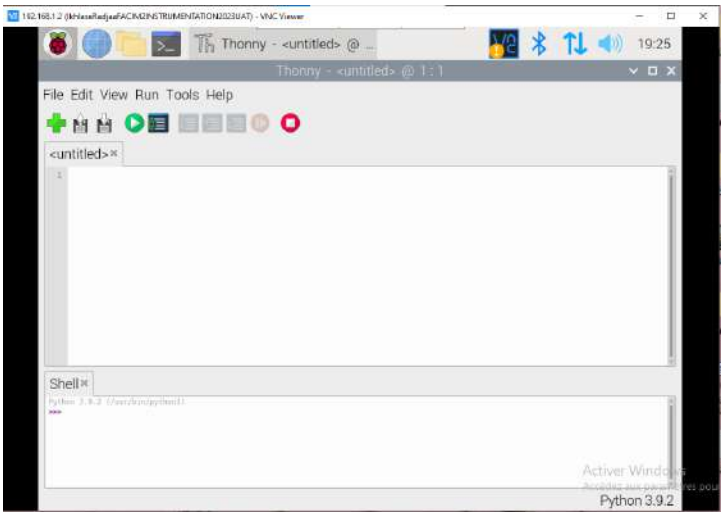


FIGURE 4.11 – L'IDE Thonny

4.4.3.2 Numpy

NumPy, abréviation de Numerical Python, est une bibliothèque Open Source de base pour le calcul scientifique en Python De nombreuses bibliothèques utilisent les tableaux NumPy comme entrées et sorties de base. NumPy introduit des objets pour les tableaux multidimensionnels et les matrices, ainsi que des routines qui permettent d'exécuter des fonctions mathématiques et statistiques avancées sur ces tableaux avec le moins de code possible.

4.4.3.3 Mathplotlib

Matplotlib est la bibliothèque Python standard pour la création de tracés et de graphiques en 2D. Elle est plutôt de bas niveau, ce qui signifie qu'il faut plus de commandes pour générer de beaux graphiques et de belles figures qu'avec certaines bibliothèques plus avancées.

4.4.3.4 OpenCV

OpenCV [OpenCV] est une bibliothèque de vision par ordinateur open source .la bibliothèque est écrite en C et C++ et fonctionne sous Linux, Windows et Mac OS X. il y a un développement actif sur les interfaces pour Python, Ruby, Matlab et d'autres langages. OpenCV a été conçu pour l'efficacité informatique et avec un fort accent sur les applications en temps réel. L'un des objectifs d'OpenCV est de fournir une infrastruc-

ture de vision par ordinateur simple à utiliser Cela aide les gens à créer rapidement des applications de vision assez sophistiquées. La bibliothèque OpenCV contient plus de 500 fonctions qui couvrent de nombreux domaines de la vision, notamment l'inspection des produits en usine, l'imagerie médicale, la sécurité, l'interface utilisateur, l'étalonnage de la caméra, la vision stéréoscopique et la robotique. Parce que la vision par ordinateur et l'apprentissage automatique vont de pair, OpenCV contient également une bibliothèque d'apprentissage automatique (MLL) complète et polyvalente.

4.4.3.5 TensorFlow

TensorFlow est une bibliothèque logicielle flexible et évolutive pour les calculs numériques à l'aide de graphiques de flux de données. Cette bibliothèque et les outils associés permettent aux utilisateurs de programmer et d'entraîner efficacement des réseaux neuronaux profonds et d'autres modèles d'apprentissage automatique.

4.4.3.6 Time

Le module time de Python est couramment utilisé pour mesurer le temps écoulé en secondes. En l'incorporant dans notre code, nous pouvons facilement calculer la durée d'exécution d'une section spécifique.. La figure 4.12 montre L'IDE Thonny avec l'importation des bibliothèques utilisées.

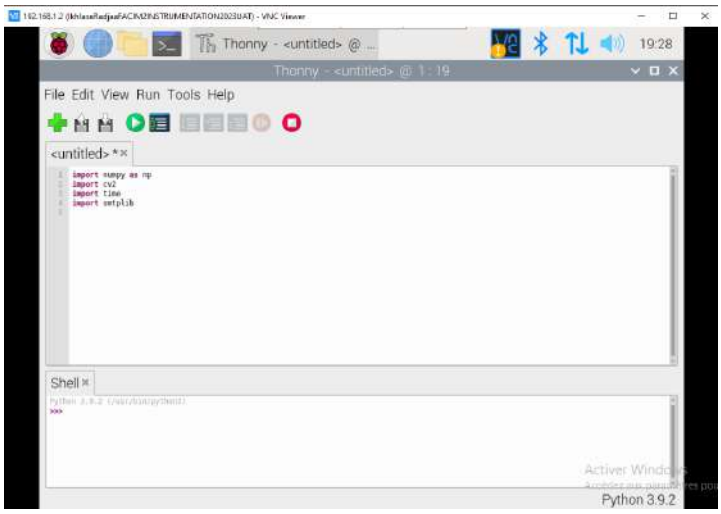


FIGURE 4.12 – L'IDE Thonny avec l'importation de quelques bibliothèques

4.5 Conclusion

Au cours de ce chapitre, nous avons exploré l'environnement matériel du Raspberry Pi 4. Nous avons examiné en détail l'environnement de programmation ainsi que les bibliothèques associées.

Le Raspberry Pi offre une performance de programmation considérable en fonction des exigences de l'utilisateur. Il offre la possibilité de concevoir et de mettre en œuvre des systèmes embarqués complexes en utilisant une puce unique compacte mais puissante.

Chapitre 5

Implémentation sur Raspberry Pi 4

Sommaire

5.1 introduction

Le domaine de la vidéo surveillance est en constante évolution, et l'un des outils les plus populaires pour le prototypage de systèmes de vidéo surveillance est le Raspberry Pi. Ce prototypage pratique nécessite une combinaison de compétences en programmation, en électronique et en conception de systèmes, ainsi qu'une solide compréhension des principes fondamentaux de la vidéo surveillance. Ce projet est une introduction pratique à la création de solutions de surveillance personnalisées, ouvrant la voie à une variété d'applications, telles que la sécurité domestique, la surveillance industrielle et bien plus encore. Ce chapitre contient les différentes expériences qui ont été jugées intéressantes pour déterminer la fiabilité et la robustesse du système proposé dans le cadre de ce travail.

5.2 Conception du système

L'objectif de ce projet est de fournir un système de haute sécurité utilisant la reconnaissance faciale sur la carte Raspberry Pi et d'envoyer une alerte à une personne autorisée via e-mail de toute personne inconnue en vue d'une reconnaissance ultérieure.
Le Raspberry Pi 4 et la caméra constituent l'infrastructure matérielle de base du système de vidéo surveillance basé sur la détection, la reconnaissance et la classification des visages sur le Raspberry Pi 4. La liste des composants matériels utilisés pour réaliser ce travail est comme suit:

- Le Raspberry Pi 4 Model B+ est utilisé comme dispositif principal; il est économique, portable et performant par rapport à d'autres systèmes embarqués.
- La caméra Raspberry Pi v2 est utilisée comme dispositif de capture vidéo.
- Un écran LCD.

L'ensemble de la conception est léger, pesant environ 140 g (0,31 lbs), y compris l'écran, le Raspberry Pi et la caméra Pi. Une vue de prés du prototype est présentée dans la figure 5.1.

Trois expériences sont proposées: une pour la détection de visages, une pour la reconnaissances de visages, une pour la reconnaissance d'inconnus et enfin une pour l'envoi d'un mail. La figure 5.2 présente le processus sous la forme d'un schéma fonctionnel.

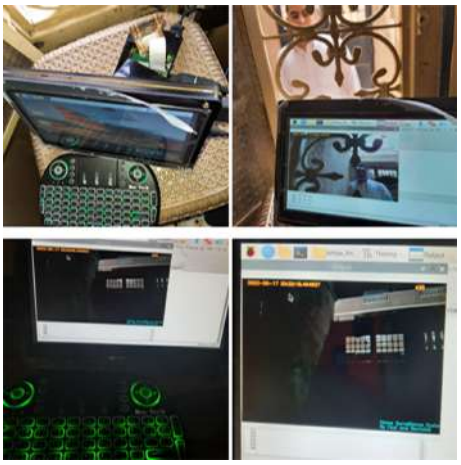


FIGURE 5.1 – Prototype de vidéo-surveillance proposé

5.3 Étapes de l'application

Le système de reconnaissance faciale cherche à identifier le visage humain. Ce visage peut changer d'apparence en fonction de l'éclairage et de l'expression faciale. Pour accomplir cette tâche, le système de reconnaissance des visages effectue trois étapes:

5.3.1 Création de l'ensemble de données

L'ensemble de données est composé d'images de personnes qui devraient être reconnues par le système. Dans cet ensemble, chaque personne est représentée par un sous-dossier, le nom de cette personne servant d'étiquette, comprenant environ 10 photos de la même personne sous différents angles et expressions faciales. Nous utilisons ici des Jpegs avec une résolution de 64k par 480px et un espace de couleur RGB.

L'objectif de cette étape est de trouver des visages dans une image et de les extraire afin qu'ils soient utilisés par le modèle de reconnaissance faciale. Nous commençons cette étape en important les modules spécifiés. Le module cv2 est déployé pour le traitement des images et Numpy est utilisé pour convertir les images en équivalents mathématiques. Le modèle de détection du visage et du masque sont également transfères du cloud Colab vers le Raspberry Pi 4.

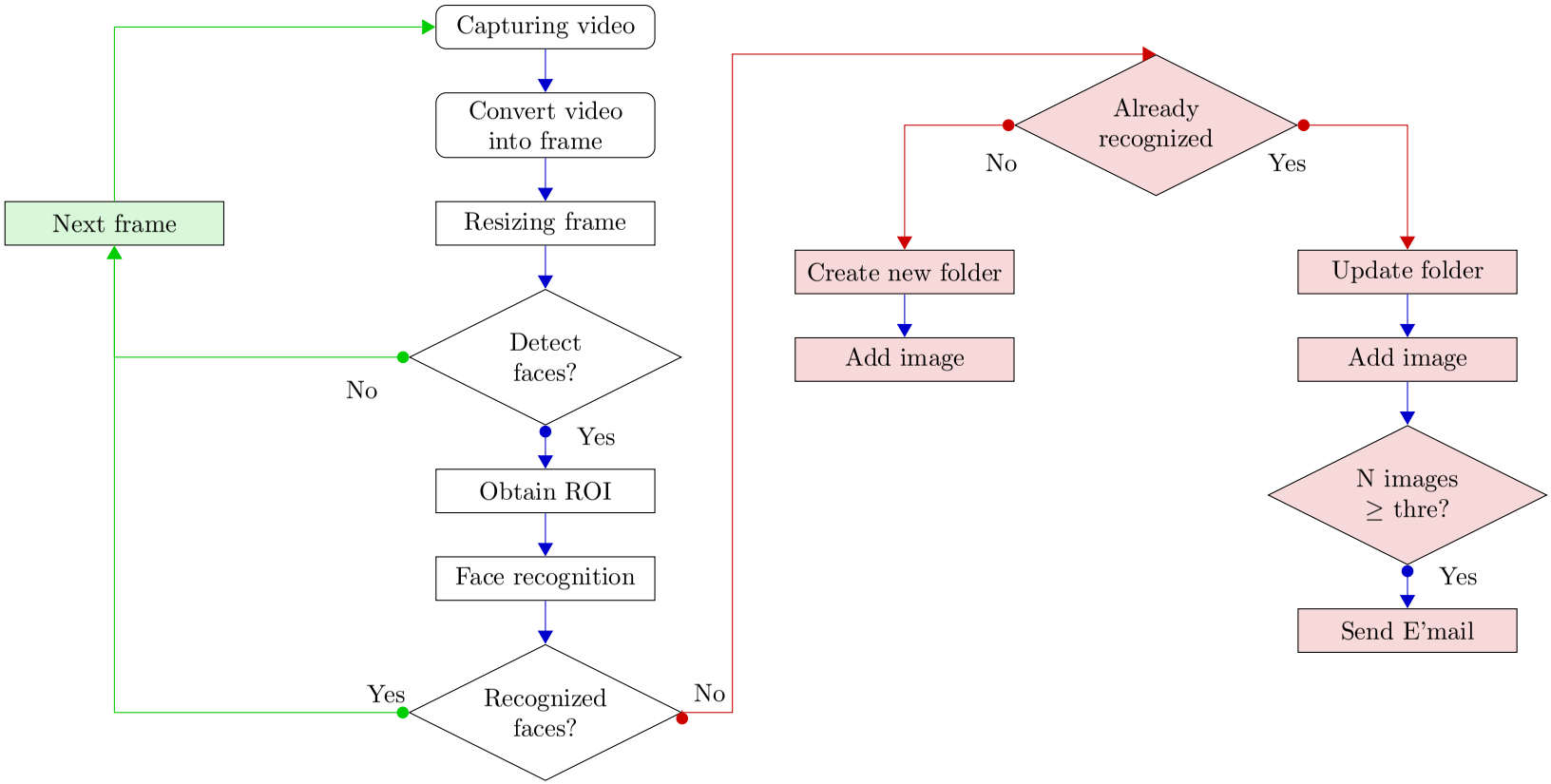


FIGURE 5.2 – Organigramme du processus général de vidéo surveillance

FIGURE 5.3 – Exemple de visages des personnes

5.3.2 Détection de visage

Après avoir allumé l'appareil, le modèle de détection de visages basé sur l'apprentissage profond est chargé dans le Raspberry Pi. Une vidéo en temps réel est capturée en permanence à l'aide de la caméra du Raspberry Pi. Le modèle utilisé détecte les visages et le système dessine un carré autour du (des) visage(s) détecté(s).

L'un des principaux défis auxquels sont confrontées les techniques actuelles de reconnaissance faciale réside dans la gestion des différentes poses et les conditions d'éclairage. Ce type de situation est souvent rencontré dans les systèmes de vidéo surveillance qui traitent avec des individus non coopératifs. Le détecteur SSD offre des résultats remarquables dans de telles situations comme le montre la figure 5.4.

5.3.3 Reconnaissance de visages

Cette étape consiste à utiliser l'image détectée dans la phase précédente pour reconnaître son identité en comparant ses caractéristiques uniques à toutes les caractéristiques des personnes stockées dans la base de données.
Le transfert learning, technique qui implique l'utilisation de modèles préalablement en-

FIGURE 5.4 – Détection de visages

traînés sur de vastes bases de données d'images faciales, est employée pour effectuer la tâche spécifique de reconnaissance sur le Raspberry Pi. Cette approche permet l'utiliser les connaissances déjà acquises par ces modèles pré-entraînés. Les modèles pré entraînés sont souvent complexes et demandent une puissance de calcul élevée, mais grâce à la capacité de traitement du Raspberry Pi, ces modèles peuvent être adaptés et déployés sur des systèmes embarqués.

Tout d'abord, le modèle de reconnaissance pré entraîné est transféré au Rasperry Pi et est utilisé pour extraire les caractéristiques des visages détectés. Cette étape implique de passer ces sous-images à travers le modèle et de récupérer les activations de la dernière couche, qui correspond aux caractéristiques spécifiques aux visages permettant ainsi de les distinguer et de les comparer.

Pour calculer la distance entre les caractéristiques extraites des visages, on utilise généra-

lement une mesure de similarité telle que la distance Euclidienne :

$$d(A, B) = \sqrt{\sum_{i=1}^{n}(A_i - B_i)^2} \tag{5.1}$$

ou A_i et B_i représentent les éléments des vecteurs A et B respectivement, pour i allant de 1 à n. Une distance Euclidienne plus petite indique une plus grande similarité entre les visages. La figure 5.5 en est une démonstration.

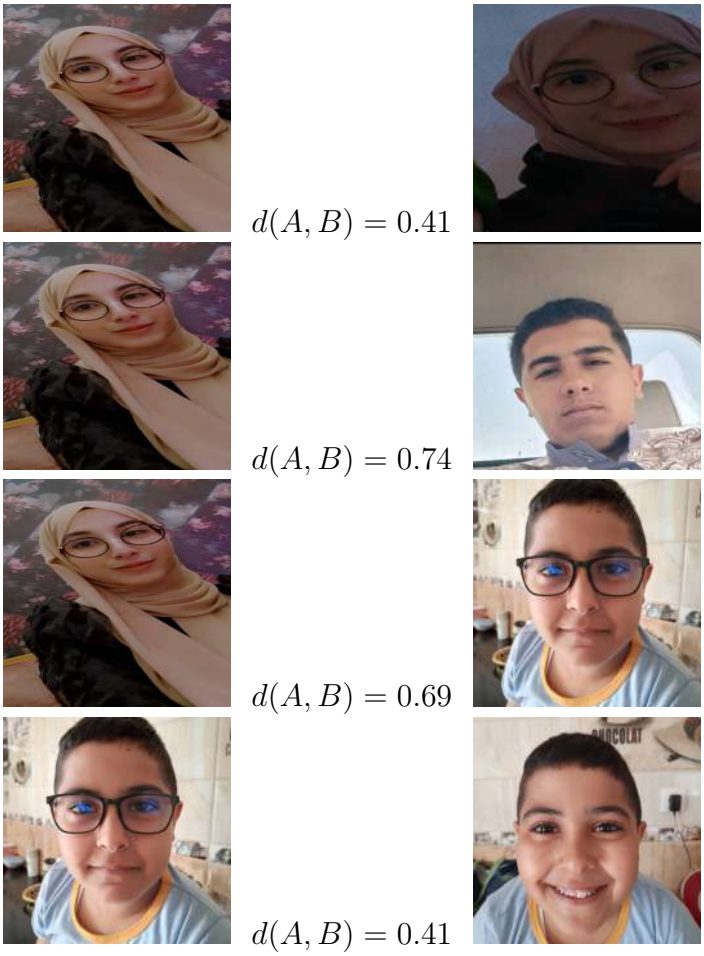


FIGURE 5.5 – Distance entre quelques visages

Lorsqu'un visage est identifié avec succès, son identité associée est affichée; si le visage ne correspond à aucune personne connue dans la base de données, il sera classé comme "inconnu" comme le montre la figure 5.6. Néanmoins, les visages inconnus sont également stockés pour une reconnaissance ultérieure et la création de dossiers d'identification. Le stockage des visages inconnus permet d'archiver les données pour une utilisation future. Si

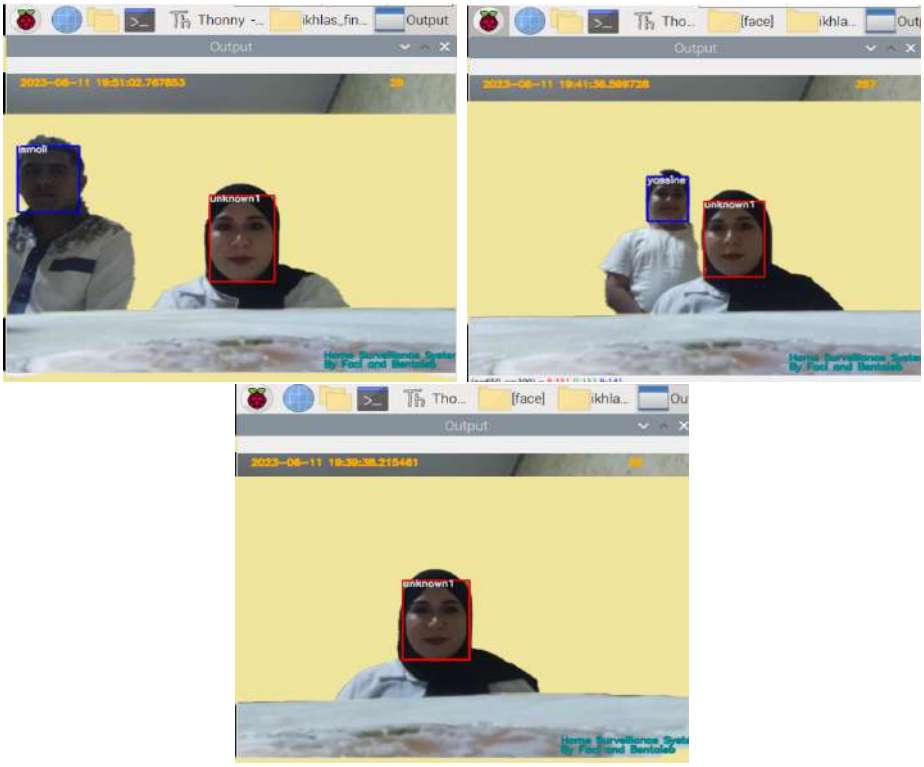


FIGURE 5.6 – Reconnaissance de visages

une correspondance est trouvée ultérieurement, les visages inconnus peuvent être reconnus et identifiés. Cela peut être particulièrement utile.

Ces dossiers contiennent des informations pertinentes, telles que l'image du visage, la date et l'heure de la détection, ainsi que les emplacements où le visage a été enregistré. Ces dossiers peuvent être consultés et analysés ultérieurement pour des besoins de sécurité ou d'investigation.

5.4 Dépassement du seuil et envoi de l'e-mail

Après chaque capture de l'image de sujet inconnu, le Raspberry Pi compare le nombre d'images stockées dans son dossier avec un seuil prédéfini *thre*. Si le nombre d'images dépasse le seuil, nous passons à l'étape qui consiste à l'envoi d'e-mail.(figure 5.8)

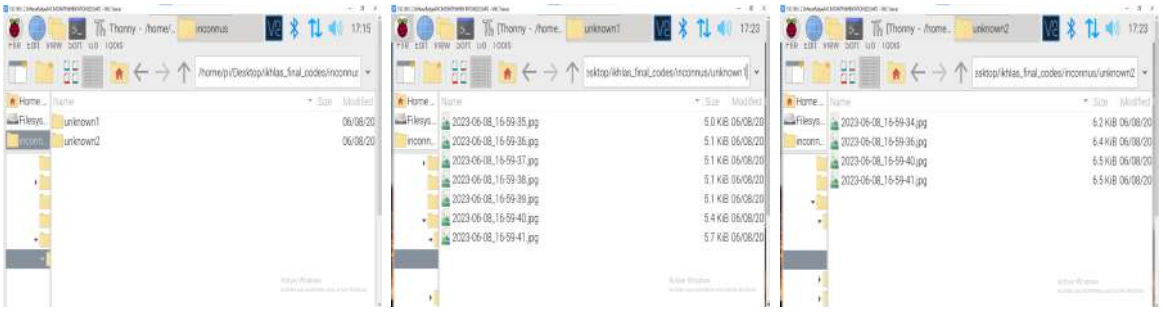


FIGURE 5.7 – Création de dossier et de fichier

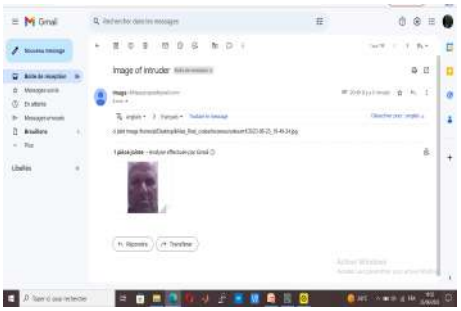


FIGURE 5.8 – Envoi d'un e-mail

5.5 Conclusion

La conception du système de reconnaissance faciale utilisant le Raspberry Pi permet de le rendre plus petit, plus léger et moins consommant d'énergie, ce qui le rend plus pratique que le système de reconnaissance faciale basé sur un PC. Nous avons utilisé des techniques basées sur le Deep Learning pour la detection et la reconnaissance des visages. Un message d'alerte de sécurité est également envoyé aux services publics autorisés. Le système développé est bon marché, rapide, très fiable et suffisamment flexible pour répondre aux exigences de différents systèmes.

Conclusion Générale

Dans ce projet, nous avons préposé une approche approfondie de la détection et de la reconnaissance faciales à l'aide du Raspberry Pi, une plateforme de développement de systèmes embarqués largement utilisée et peu coûteuse. Notre méthode est un moyen léger, peu coûteux et efficace d'exécuter ces fonctions sur des dispositifs embarqués, en utilisant la bibliothèque OpenCV et des algorithmes d'apprentissage approfondi.
Le modèle SSD a permis de détecter les visages de manière fiable et précise dans les images et les vidéos. En outre, l'application du transfer learning a donné de bon résultats pour la reconnaissance de visages grâce au modèle MobilenetV2.
La mise en œuvre pratique du Raspberry Pi a prouvé qu'en dépit de ses faibles ressources, la plateforme est capable d'effectuer certaines tâches efficacement. De nombreux autres domaines, tels que la sécurité, la surveillance, l'interaction homme-machine, etc. pourraient en bénéficier.

En conclusion, notre approche offre une méthode pratique et accessible pour effectuer la détection et la reconnaissance faciales à l'aide d'un Raspberry Pi. Les performances obtenues sont suffisantes pour une utilisation dans des applications embarquées malgré les limitations matérielles. Pour améliorer encore les performances de détection et de reconnaissance faciale, notre étude ouvre la voie à de futures recherches et développements dans le domaine de la vision par ordinateur sur les systèmes embarqués, en capitalisant sur les capacités du Raspberry Pi et en explorant de nouvelles approches de deep learning.

Bibliographie

[FPS,] What does fps mean? https://clipchamp.com/en/definition/what-does-fps-mean-frame-rates-explained/.

[Abdi and Williams, 2010] Abdi, H. and Williams, L. J. (2010). Principal component analysis. *Wiley interdisciplinary reviews: computational statistics*, 2(4):433–459.

[Ahonen et al., 2006] Ahonen, T., Hadid, A., and Pietikainen, M. (2006). Face description with local binary patterns: Application to face recognition. *IEEE transactions on pattern analysis and machine intelligence*, 28(12):2037–2041.

[Andrews et al., 2013] Andrews, H. C., Billingsley, F., Fiasconaro, J., Frieden, B., Read, R., Shanks, J., and Treitel, S. (2013). *Picture processing and digital filtering*, volume 6. Springer Science & Business Media.

[Bishop and Nasrabadi, 2006] Bishop, C. M. and Nasrabadi, N. M. (2006). *Pattern recognition and machine learning*, volume 4. Springer.

[Calonder et al., 2010] Calonder, M., Lepetit, V., Strecha, C., and Fua, P. (2010). Brief: Binary robust independent elementary features. In *Computer Vision–ECCV 2010: 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5-11, 2010, Proceedings, Part IV 11*, pages 778–792. Springer.

[Castleman, 1996] Castleman, K. R. (1996). *Digital image processing*. Prentice Hall Press.

[Chen et al., 2023] Chen, Z., Liang, Q., Wei, Z., Chen, X., Shi, Q., Yu, Z., and Sun, T. (2023). An overview of in vitro biological neural networks for robot intelligence. *Cyborg and Bionic Systems*, 4:0001.

[Coronel et al., 2022] Coronel, F., Barreno, N., Muñoz, P., Zabala-Blanco, D., Onofa, N., and Flores-Calero, M. (2022). Web-based personal access control system using facial

recognition with deep learning techniques. In *2022 IEEE Colombian Conference on Communications and Computing (COLCOM)*, pages 1–6. IEEE.

[Cortes and Vapnik, 1995] Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20:273–297.

[Deng et al., 2009] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee.

[Ghael et al., 2020] Ghael, H. D., Solanki, L., and Sahu, G. (2020). A review paper on raspberry pi and its applications. *International Journal of Advances in Engineering and Management (IJAEM)*, 2(12):4.

[Goodfellow et al., 2016] Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep learning*. MIT press.

[Goodfellow et al., 2020] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM*, 63(11):139–144.

[Gupta et al., 2016] Gupta, I., Patil, V., Kadam, C., and Dumbre, S. (2016). Face detection and recognition using raspberry pi. In *2016 IEEE international WIE conference on electrical and computer engineering (WIECON-ECE)*, pages 83–86. IEEE.

[Haenlein and Kaplan, 2019] Haenlein, M. and Kaplan, A. (2019). A brief history of artificial intelligence: On the past, present, and future of artificial intelligence. *California management review*, 61(4):5–14.

[Hashmi et al., 2020] Hashmi, M. F., Katiyar, S., Keskar, A. G., Bokde, N. D., and Geem, Z. W. (2020). Efficient pneumonia detection in chest xray images using deep transfer learning. *Diagnostics*, 10(6):417.

[Howard et al., 2017] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., and Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*.

[Jolles, 2021] Jolles, J. W. (2021). Broad-scale applications of the raspberry pi: A review and guide for biologists. *Methods in Ecology and Evolution*, 12(9):1562–1579.

[Kanagaraj et al., 2022] Kanagaraj, R., Amsaveni, M., Binsha, S., and Chella Keerthana, S. (2022). Raspberry pi-based spy robot with facial recognition. In *Proceedings of Third International Conference on Intelligent Computing, Information and Control Systems: ICICCS 2021*, pages 29–40. Springer.

[Kommaraju et al., 2022] Kommaraju, R., Humanvitha Sai Dharani, C., Mansoor, S., and Pujitha, S. (2022). Real-time face mask detection at gateway using opencv and iot. In *Proceedings of Third International Conference on Intelligent Computing, Information and Control Systems: ICICCS 2021*, pages 519–531. Springer.

[Krizhevsky et al., 2017] Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90.

[LeCun et al., 2015] LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *nature*, 521(7553):436–444.

[LeCun et al., 1989] LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., and Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. *Neural computation*, 1(4):541–551.

[Liu et al., 2018] Liu, K., Kang, G., Zhang, N., and Hou, B. (2018). Breast cancer classification based on fully-connected layer first convolutional neural networks. *IEEE Access*, 6:23722–23732.

[Liu et al., 2016] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., and Berg, A. C. (2016). Ssd: Single shot multibox detector. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*, pages 21–37. Springer.

[Ma et al., 2019] Ma, Y., Zhu, W., Benton, M. G., and Romagnoli, J. (2019). Continuous control of a polymerization system with deep reinforcement learning. *Journal of Process Control*, 75:40–47.

[Meddeb et al., 2023] Meddeb, H., Abdellaoui, Z., and Houaidi, F. (2023). Development of surveillance robot based on face recognition using raspberry-pi and iot. *Microprocessors and Microsystems*, 96:104728.

[Mita et al., 2005] Mita, T., Kaneko, T., and Hori, O. (2005). Joint haar-like features for face detection. In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, volume 2, pages 1619–1626. IEEE.

[Murphy, 2012] Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*. MIT press.

[Nagpal et al., 2018] Nagpal, G. S., Singh, G., Singh, J., and Yadav, N. (2018). Facial detection and recognition using opencv on raspberry pi zero. In *2018 Internatio-*

nal Conference on Advances in Computing, Communication Control and Networking (ICACCCN), pages 945–950. IEEE.

[Petrou and Petrou, 2010] Petrou, M. M. and Petrou, C. (2010). *Image processing: the fundamentals*. John Wiley & Sons.

[Sajjad et al., 2020] Sajjad, M., Nasir, M., Muhammad, K., Khan, S., Jan, Z., Sangaiah, A. K., Elhoseny, M., and Baik, S. W. (2020). Raspberry pi assisted face recognition framework for enhanced law-enforcement services in smart cities. *Future Generation Computer Systems*, 108:995–1007.

[Saypadith and Aramvith, 2018] Saypadith, S. and Aramvith, S. (2018). Real-time multiple face recognition using deep learning on embedded gpu system. In *2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pages 1318–1324. IEEE.

[Sharifisoraki et al., 2023] Sharifisoraki, Z., Amini, M., and Rajan, S. (2023). A novel face recognition using specific values from deep neural network-based landmarks. In *2023 IEEE International Conference on Consumer Electronics (ICCE)*, pages 1–6. IEEE.

[Su and Yang, 2022] Su, J. and Yang, W. (2022). Artificial intelligence in early childhood education: A scoping review. *Computers and Education: Artificial Intelligence*, page 100049.

[Tazin et al., 2021] Tazin, T., Sarker, S., Gupta, P., Ayaz, F. I., Islam, S., Monirujjaman Khan, M., Bourouis, S., Idris, S. A., Alshazly, H., et al. (2021). A robust and novel approach for brain tumor classification using convolutional neural network. *Computational Intelligence and Neuroscience*, 2021.

[Thévenaz et al., 2000] Thévenaz, P., Blu, T., and Unser, M. (2000). Image interpolation and resampling. *Handbook of medical imaging, processing and analysis*, 1(1):393–420.

[Viswanathan, 2009] Viswanathan, D. G. (2009). Features from accelerated segment test (fast). In *Proceedings of the 10th workshop on image analysis for multimedia interactive services, London, UK*, pages 6–8.

[Wilson and Fernandez, 2006] Wilson, P. I. and Fernandez, J. (2006). Facial feature detection using haar classifiers. *Journal of computing sciences in colleges*, 21(4):127–133.

[Xiao, 2019] Xiao, Y. (2019). Vehicle detection in deep learning. *arXiv preprint arXiv:1905.13390*.

[Zur et al., 2009] Zur, R. M., Jiang, Y., Pesce, L. L., and Drukker, K. (2009). Noise injection for training artificial neural networks: A comparison with weight decay and early stopping. *Medical physics*, 36(10):4810–4818.