Republique Algerienne Democratique et Populaire
Ministère de l'Enseignement Superieur et de la Recherche Scientifique

Universite A/Mira de Bejaia
Faculte des Sciences Exactes
Departement d'Informatique

MEMOIRE DE MASTER RECHERCHE

En

Informatique

Option

intelligence artificielle

Thème

Estimation de l'âge à partir des images faciales par les Réseaux de Neurones Convolutifs (CNN)

Présenté par :
Mlle. AFENAI Lynda
Mlle. BOUBEKRI Katia

Soutenu le 6 juillet 2022 : devant le jury compose de

President	**Dr M.Moketfi**	**MCB**	**U. A/Mira Bejaia.**
Rapporteur	**Dr M.Khammari**	**MCA**	**U. A/Mira Bejaia.**
Examinateur	**Dr k.Souadih**	**Docteur**	**Sonatrach Bejaia.**

Bejaia, juillet 2022.

Remerciements

Avant toute chose, on remercie dieu tout puissant pour nous avoir aidé et éclairé le chemin pour la réalisation de ce mémoire.

À L'issue de ce modeste travail, on tient à exprimer nos sincères remerciements aux :

les membres du jury pour avoir accepté d'examiner et d'évaluer ce travail.

MR. Mohammed Khemmari pour son encadrement, sa grande disponibilité, sa confiance ainsi que son soutien, ses conseils et sa patience qu'il nous a accordés durant toute cette année.

Enseignants de la filière informatique pour les connaissances qu'ils nous ont transmises pendant toute la période de notre formation.

À nos familles et en particulier nos parents.

0

Dédicaces

Je dédié ce Modeste travail :

À l'homme de ma vie. mon exemple éternel, mon soutien moral et source de joie et de bonheur, celui qui s'est toujours sacrifié pour me voir réussir, qui éclaire mon chemin et m'illumine de douceur et d'amour, que dieu te garde pour nous PAPA.

À ma très chère mère aucune dédicace très chère maman, ne pourrait exprimer la profondeur des sentiments que j'éprouve pour vous, vos sacrifices innombrables et votre dévouement firent pour moi un encouragement. Vous avez guetté mes pas, et m'avez couvé de tendresse, ta prière et ta bénédiction m'ont été d'un grand secours pour mener à bien mes études.

À mes adorables sœurs : Sraya , Célia et Safia

Mes deux petits frères : Adem et Riad

À mes chers grands parents.

À toute la famille AFENAI, et la famille BENNAI, qui ont contribué de près ou de loin à ma réussite avec leurs conseils, aides et encouragements.

À mon encadreur Monsieur MOHAMMED KHEMARI pour son aide durant toute la période du travail et pour tout le soutien et l'orientation.

À tous mes enseignants durant mes années d'études avec lesquels j'ai beaucoup appris

À ceux qui j'aie beaucoup, qui m'ont toujours soutenus et étaient toujours à mes côtés, mes chers amis : Mounira, Imene, Sabine, Manel.

À ma binôme KATIA BOUBEKRI pour tous les encouragements qu'elle ma donnée et la façon dont elle crois toujours en moi.

Lynda

Dédicaces

Quand il y a la soif d'apprendre, tout vient à point a qui sait attendre

Je dédie se travail :

À mon père ; mon premier encadrant, aucune dédicace ne saurait exprimer l'amour, l'estime, et le respect que j'ai toujours eu pour toi que dieu te préserve.

À ma très chère mère ; tu représentes pour moi un symbole de bonté par excellence, la source de la tendresse et l'exemple de dévouement qui n'a pas cessé de m'encourager et de prier pour moi. Ce travail est le fruit des sacrifices que tu as consentis pour mon éducation et ma formation.

À ma grande sœur Sara et mon grand frére Nabil que j'aime plus que tout, que dieu vous préserve.

À ma grand-mère et à toute la famille BOUBEKRI et MEDDAS.

À mon encadrant monsieur MOHAMMED KHEMMARI pour son aide ainsi que son soutien et sa grande disponibilité.

À tous mes amis et en particulier : Yasmine, Katy, Tchin, Celia, Manel, Sabine et Imene ainsi qu'à mes camarades de promotion.
A mon Sidou qui a toujours était ma source de motivation que dieu le préserve.
À toi ma binôme AFENAI LYNDA pour ta motivation et tes éfforts, je souhaite beaucoup de réussite et de bonheur, que dieu te garde.

Katia

Table des matières

Table des figures

Liste des tableaux

Notations et symboles

A	AAM	Modèle Apparence Active.
	$AGES$	AGing pattErn Subspace.
	$AFAD$	Asian Face Age Dataset.
	$AdaBoost$	Adaptive Boosting.
B	BIF	Biologically Inspired Features.
	$BSIF$	Binarized Statistical Image Features.
C	CNN	Convolutional Neural Network.
	CEA	Conformal Embedding Analysis.
	CS	Cumulative Score.
	CCA	Canonical Correlation Analysis.
	$CACD$	Cross Age Celebrity Dataset.
	CV	Computer Vision.
D	DCT	Discrete Cosine Transform.
	DL	Deep Learning.
	DNN	Deep Neural Networks.
E	ELM	Extreme Learning Machine.
F	$FRGC$	Face Recognition Grand Challenge.
	FCN	Fully Convolutional Network.
H	HSV	Hue Saturation Value.
	HOG	Histogram Oriented Gradients.
I	IA	Intelligence Artificielle
	IBR	Iteration Bayesian Reweighed .
	$IFDB$	Iranian Face DataBase.
	ICA	Independant Component Analysis.
	IoU	Intersection over Union.
K	KRR	Kernel Ridge Regression.
	KNN	K Nearest Neighbors.
	KL	Kullback Leibler
L	LBP	Local Binary Patterns.
	LDA	Linear Discriminant Analysis.
	$LARR$	Locally Adjusted Robust Regression.
	LPQ	Local Phase Quantisation.
M	MAE	Mean Absolute Error.
	ML	Machine Learning.
	MLR	Multiple Linear Regression
	MLP	Multilayer Perceptron
	$MTCNN$	Multi Task Cascaded CNN .
N	NMS	Non maximu Suppression.
O	$ODFL$	Ordinal Deep Feature Learning.
	$O-Net$	Output Network.
P	PCA	Principal Component Analys
	PLS	Partial Least Squares.
	$P-Net$	Proposal Network.
R	$RESNet$	Remote Sensing Neural Netw
	RNN	Recurrent Neural Networks.
	$RCNN$	Regions Convolutional Neura
	$R-Net$	Refinement network.
	ROI	Region of Interest.
S	SVR	Support Vector Regression.
	SOM	Self Organizing Map.
	SPF	Spatially Patch Flexible.
	$SURF$	Speeded Up Robust Features
	SVM	Support Vecteur Machine.
	SSH	Single Stage Headless.
	SSD	Single Shot multibox Detect
	$SIFT$	Scale Invariant Feature Tran
U	UV	UltraViolet.
V	VGG	Visual Geometry Group.
W	$WIT-DB$	Waseda Interaction Technolo
Y	$YOLO$	You Only Look Once.
	YGA	Yahama Gender et Age

Introduction générale

Le visage humain contient de nombreuses informations sur une personne : la taille et la géométrie du menton, les lèvres, le nez, les sourcils et d'autres composants du visage peuvent être utilisés pour distinguer le sexe et la race humaine, tandis que les plis, les lignes, et les rides peuvent révéler des indices sur l'âge. La majorité des gens sont capables de reconnaître facilement des traits humains tels que les états émotionnels. Ils sont également capables de dire, en regardant une personne, si elle est un adulte, un adolescent, un enfant ou une personne âgée [102]. Cependant, il est souvent difficile de connaître l'âge exact d'une personne simplement en regardant des photos, et ceci represente un plus grand défi pour un ordinateur et cette technique s'appelle la vision par ordinateur.

La vision par ordinateur est un domaine en pleine expansion consacré à l'analyse, à la modification et à la compréhension des images. Son objectif est d'utiliser cette compréhension pour contrôler un ordinateur ou pour fournir de nouvelles images plus informatives que les originales. Et afin que cette compréhension soit possible l'apprentissage automatique a été mis en place.

L'apprentissage automatique est le domaine d'étude qui donne aux ordinateurs la capacité d'apprendre. Comme son nom l'indique, l'apprentissage donne aux machines une caractéristique qui le rend plus semblable aux humains [28]. Parmi les algorithmes qui permettent à la machine d'apprendre par elle-même, la simulation des neurones du corps humain. Pour cela des architectures de deep learning basées sur les réseaux de neurones sont apparues.

L'objectif de l'estimation automatique de l'âge est de juger si l'âge estimé après un apprentissage est aussi proche que possible de l'âge réel. Ces systèmes ont connu une croissance rapide ces dernières années en raison de leurs modules importants et de leurs utilisations bénéfiques pour de nombreuses applications de vision par ordinateur [102], notamment l'interaction homme-machine, les systèmes de sécurité et la surveillance visuelle. Par exemple, l'estimation automatique de l'âge est actuellement utilisée par les hôtels, les aéroports, les gares routières, les bâtiments publics, les universités, les hôpitaux, les cinémas ..., et cela afin d'augmenter le niveau de sécurité et de faire face à toute menace ou déficience éventuelle. Outre les applications de sécurité, les techniques d'estimation de l'âge sont également utilisées dans les systèmes de soins de santé, la recherche d'informations, les recherches universitaires et les systèmes de gestion électronique de la relation client, etc.

Bien que l'estimation de l'âge puisse être réalisée à l'aide de différentes caractéristiques biométriques, notre mémoire se concentre sur l'estimation de l'âge facial qui repose sur les caractéristiques biométriques extraites du visage. Ce cadre utilise les réseaux de neurones convolutifs (CNN).

Pour présenter correctement notre modeste travail, nous avons divisé notre mémoire en quatre chapitres principaux comme suit :

Dans le premier chapitre, nous avons donné une définition de l'estimation d'âge, l'intérêt de détection de visages, quelques problèmes que rencontre notre thématique, les applications et une vue d'ensemble du domaine de l'estimation de l'âge facial. Nous avons également présenté quelques travaux connexes de l'estimation d'âge à partir des images faciales.

Dans Le deuxième chapitre , nous avons cité quelques méthodes de détection de visages ainsi que des méthodes d'extraction de caractéristiques où nous avons détaillé celles que nous avons utilisées dans notre projet et les méthodes d'apprentissage.

Le troisième chapitre, expose le principe de fonctionnement de chaque méthode utilisée dans notre travail tels que les méthodes de détection du visage : les filtres de HAAR [113], YOLO [97] et MTCNN [121]) et la méthode d'extraction de caractéristique LBP [8], ainsi que deux méthodes d'apprentissage avec les réseaux de neurones convolutifs.

Dans le quatrième chapitre, nous avons présenté notre application, défini les bases de données utilisées et exposé les résultats obtenus.

Chapitre 1

Chapitre 1.Estimation faciale d'âge

1.1 Introduction

Au cours des dernières décennies, avec la nécessité croissante d'automatiser les systèmes de reconnaissance et de surveillance, les recherches sur le traitement et l'analyse numérique de visages humains (y compris la détection de visages, la reconnaissance de visages, la classification de genre et la reconnaissance de l'expression du visage) ont attirées une attention particulière dans les communautés de la vision par ordinateur et de la reconnaissance des formes.

L'âge et le sexe jouent un rôle fondamental dans les interactions sociales. Malgré le rôle essentiel que jouent ces attributs dans notre vie quotidienne, la capacité de les estimer automatiquement de manière précise et fiable à partir d'images de visages est encore loin pour répondre aux besoins des applications commerciales. En lien avec ces recherches, la prédiction de l'âge d'une personne à partir des images faciales est un sujet relativement nouveau. L'estimation de l'âge par analyse numérique du visage trouve de nombreuses applications pratiques dans le monde réel telles que la collecte des statistiques démographiques, le profilage client, l'optimisation de la recherche dans les grandes bases de données et l'aide des systèmes de la biométrie. Cette estimation est beaucoup plus lente en raison de la difficulté de collecter et d'étiqueter de grands ensembles de données. L'attribut de l'âge pourrait être également exploité dans la vérification du visage et de la récupération des données pour améliorer par exemple les outils utilisés dans les enquêtes policières. De manière générale, l'estimation automatique de l'âge par une machine est utile dans les applications où l'objectif est de déterminer l'âge d'un individu sans l'identifier précisément. Plusieurs méthodes d'estimation de l'âge ont été proposées pour différentes applications. Malgré les avantages de ces méthodes, elles souffrent de plusieurs limitations dûes à plusieurs défis rencontrés lors de leur développement.

Dans ce qui suit nous allons définir l'estimation faciale d'âge à partir des images faciales.

1.2 Estimation faciale d'âge

L'âge est un nombre réel qui signifie le nombre d'années écoulées depuis la naissance jusqu'à un certain point dans la vie. L'estimation de l'âge est le processus d'estimation de l'âge réel à l'aide d'artefacts visuels sur le visage.

L'estimation de l'âge est une tâche importante dans la classification des images faciales. Elle est définie comme l'âge d'une personne en fonction de ses caractéristiques biométriques, précisément sur la base d'images faciales bidimensionnelles [43]. Les points caractéristiques du visage (les yeux, le nez et la bouche) peuvent être définis comme des points de référence standard sur le visage humain, utilisés par les scientifiques afin de reconnaître le visage d'une personne ou, dans ce cas, d'estimer l'âge de la personne [43]. L'âge d'un individu peut être déterminé de plusieurs façons mais dans notre mémoire nous nous concentrons sur l'estimation d'âge a partir des images faciales. L'estimation automatique de l'âge à partir d'images faciales est l'une des tâches utilisées mais difficiles. La recherche sur l'estimation de l'âge a suscité beaucoup d'intérêt ces dernières années, des définitions et des termes de base sont données avec la publication annuelle de nombreux articles de journaux et de conférences, ainsi que la soutenance de thèses de maîtrise et de doctorat.

L'estimation de l'âge est une technique d'étiquetage automatique du visage humain avec un âge ou un groupe d'âge exact. Cet âge peut être l'âge réel, d'apparence, perçu ou l'âge estimé. L'âge réel est le nombre d'années qu'une personne a accumulées depuis sa naissance jusqu'à aujourd'hui, exprimé sous la forme d'un nombre réel. L'âge d'apparence et l'âge perçu sont estimés sur la base des informations visuelles sur l'âge du visage, tandis que l'âge estimé est l'âge du sujet estimé par une machine à partir de l'apparence visuelle du visage. L'âge d'apparence est supposé correspondre à l'âge réel, bien qu'il existe des variations dues à la nature stochastique du vieillissement chez les individus [10].

Dans notre travail la détection du visage représente la première étape du systéme d'estimation faciale d'âge, dans ce qui suit nous allons détailler l'intérêt de cette étape.

1.3 Détection de visages

La détection des visages dans l'image est une étape essentielle et cruciale. Elle consiste à rechercher dans cette dernière la position des visages et de les extraire comme un ensemble d'images pour faciliter le traitement ultérieur. Le concept de base de la détection de visages serait celui du K-Plus Proches Voisins (K-PPV), qui consiste à parcourir l'image avec une fenêtre, puis à comparer chaque imagette qui est une partie de l'image extraite avec une série de visages types et de définir un visage comme étant tout résultat dont la distance à l'une des images de la base soit suffisamment faible et inférieure à un certain seuil. Un visage est considéré comme correctement détecté si la taille de l'image extraite ne dépasse pas 20% de la taille réelle de la région du visage et qu'elle contient principalement les yeux, le nez et la bouche [102].

La détection du visage a de très nombreuses applications directes en vidéo surveillance, biométrie, robotique, commande d'interface homme-machine, photographie, indexation d'images et de vidéos, recherche d'images par le contenu, etc. Elle permet également de faciliter l'automatisation complète d'autres processus comme la reconnaissance de visage où la reconnaissance d'expressions faciales. Parmi les applications directes, la plus connue est sa présence dans de nombreux appareils photo numérique, où elle sert à effectuer la mise au point automatique sur les visages. C'est également une technique importante pour les interfaces homme-machine évoluées, afin de permettre une interaction plus naturelle entre un humain et un ordinateur [76].

La détection du visage est aussi utilisée en indexation d'images et à la recherche d'information, où elle peut être utilisée pour rechercher des images contenant des personnes, associées automatiquement à un visage ou à un nom dans une page web et identifier les principales personnes dans une vidéo par clustering. La détection du visage peut aussi servir à déterminer l'attention d'un utilisateur, par exemple face à un écran dans l'espace public, qui peut également, une fois le visage détecté, déterminer le sexe et l'âge de la personne afin de proposer des publicités ciblées. Cela peut également servir à savoir si une personne est bien présente devant une télévision allumée, et dans le cas contraire mettre l'appareil en veille ou réduire la luminosité afin d'économiser de l'énergie. De façon plus indirecte, la détection du visage est la première étape vers des applications plus évoluées, qui nécessitent la localisation du visage.

Les figures 1.1 montre la détection d'un visage dans une image.

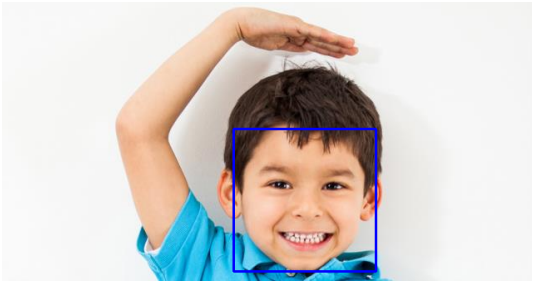

FIGURE 1.1 – Exemple d'image comportant un seul visage.

Après la détection de visages, nous allons cité les différents problèmes qu'on peut rencontré lors d'une estimation faciale d'âge.

1.4 Problèmes rencontrés dans l'estimation d'âge

L'estimation de l'âge présente de nombreux problèmes rencontrés dans d'autres tâches typiques d'interprétation d'images de visage telles que la détection de visages, la reconnaissance de visages, la reconnaissance d'expressions faciales et la reconnaissance de genre. L'orientation du visage, Les déformations de l'apparence du visage causées par différentes expressions tels que le profil de la

tête, l'éclairage, l'occlusion, la moustache, la barbe, le maquillage, la coiffure et le vieillissement [10] .

On a deux types de problèmes qui sont :

en premier ce sont les problèmes rencontrés lors de la détection du visage : si le visage n'est pas détecté l'estimation faciale d'âge ne peut pas se faire.

en second lieu ce sont les problèmes rencontrés lors de l'estimation de l'âge : si le visage est correctement détecté mais l'estimation faciale d'âge ne peut pas se faire.

1) Problèmes rencontrés lors de la détection de visage :

• **Profil de la tête :** Le taux de détection du visage baisse quand des variations de pose sont présentes dans les images. La variation de pose est considérée comme un problème majeur pour les systèmes de détection faciale. La figure 1.3 montre un exemple de visage en profil.

FIGURE 1.2 – Exemple de profil de la tête [69].

• **Expression faciale :** La déformation du visage qui est due aux expressions faciales est localisée principalement sur la partie inférieure du visage. L'identification d'un visage à partir d'une expression faciale est un problème difficile qui n'est toujours pas résolu. La figure 1.4 montre un exemple de variation d'expressions faciales.

FIGURE 1.3 – Exemple de variation d'expressions. [85]

• **Eclairage :** Le problème de l'éclairage est un vieux problème dans la vision de la machine. L'intensité et la direction de l'éclairage pendant la prise de vue affectent l'apparence du visage. Ces changements dans l'éclairage peuvent révéler des ombres qui mettent en évidence ou cachent certaines caractéristiques du visage. Par exemple, un visage vu sous une lumière bleue est totalement différent d'un visage vu sous une lumière rouge [10]. La figure 1.5 montre un exemple de changement d'éclairage.

FIGURE 1.4 – Exemple de changement d'éclairage

• **Occlusion :** Un visage peut être partiellement masqué par des objets ou par le port d'accessoires tels que les lunettes,le chapeau, l'écharpe,le livre et autres accessoires. Cela affecte l'extraction et la reconnaissance des caractéristiques d'un visage [10]. La figure 1.6 montre des exemples d'occlusion.

FIGURE 1.5 – Exemples d'occlusion

• **Moustache et barbe :** La présence d'éléments structurels tels que la barbe et la moustache peut modifier les caractéristiques du visage telle que la forme, la couleur ou la taille. En d'autres termes, ces composants peuvent masquer les caractéristiques faciales de base des traits du visage, entraînant l'échec du système de reconnaissance. La figure 1.7 montre un exemple d'éléments structurels du visage.

2) Problèmes rencontrés lors de l'estimation d'âge

• **Maquillage et coiffure :** L'utilisation de techniques de maquillage et le type de coiffure pour effets spéciaux a déjà induit en erreur autant les humains que les machines lors de la reconnais-

FIGURE 1.6 – Exemple d'éléments structurels du visage [5]

sance faciale ainsi que l'estimation faciale d'âge. La figure 1.8 montre un exemple de changement d'âge apparance avec du maquillage.

FIGURE 1.7 – Exemple de changement d'âge apparance avec maquillage

• **Vieillissement :** Les principaux facteurs qui influencent le vieillissement qui est le changement de la texture de la peau du visage sont la gravité, l'exposition aux rayons ultraviolets (UV) du soleil, la maturité des tissus mous, la restructuration des os et l'activité musculaire du visage. Ces facteurs entraînent des variations dans l'apparence du visage [10]. De plus, les hommes et les femmes peuvent vieillir différemment. La figure 1.9 montre un exemple de veillissements.

En effet, deux personnes différentes vieillissent très différemment, puisque le processus de vieillissement est déterminé non seulement par les gènes de la personne, mais aussi par de nombreux facteurs externes, tels que la santé, le style de vie, l'environnement et les conditions météorologiques. De l'enfance à la puberté, l'évolution la plus significative se manifeste par la croissance craniofaciale (changement de la forme du crâne). Dans l'ensemble, la taille du visage s'agrandit progressivement au cours de la croissance. Le changement de la forme se poursuit mais de façon moins marquée.

On peut aussi souligner le défi que présente la collecte de données d'apprentissage. En effet, contrairement aux problèmes de la détection et de l'identification pour lesquelles les images faciales peuvent être collectées aisément en laboratoire ou avec une fouille sur les sites internet, la collecte des images faciales avec vérité terrain pour l'âge est très fastidieuse. Souvent, les bases de données d'images faciales avec vérité terrain souffrent d'un certain déséquilibre en terme de représentativité de tous les âges. En effet, les âges avancés sont représentés souvent par un petit nombre d'images.

FIGURE 1.8 – Exemple de deux veillissements differents [5]

L'estimation faciale d'âge trouve de nombreuses applications pratiques dans le monde réel.

1.5 Application d'estimation d'âge

L'estimation automatique de l'âge à partir d'images faciales est récemment apparue comme une technologie aux multiples applications interressantes dont l'objectif est de déterminer l'âge spécifique ou la tranche d'âge d'une personne à partir d'une image faciale donnée. Ces applications peuvent aider à rendre les opérations, les transactions et la vie quotidienne plus sûres et plus pratiques telles que :

Applications commerciales : Comme exemple une application qui aide les agences de publicité qui cible des publics spécifiques en terme de tranche d'âge afin de développer des panneaux d'affichage intelligents, qui grâce à l'estimation faciales d'âge des personnes qui passent le panneau adapte le contenu affiché [10] ; par exemple un magasin de vêtements qui affiche les vêtements adaptés à la tranche d'âge des personnes qui passent ; ou encore un restaurant qui affiche les plats populaires pour chaque tranche d'âges, etc.

Applications de contrôle de sécurité et surveillance : les questions de contrôle de sécurité et de surveillance deviennent de plus en plus cruciales dans la vie quotidienne. Une application d'estimation d'âge humain peut générer un signal sonore lorsque des mineurs entrent dans des institutions à plus de 18 ans, ou celles qui contrôlent l'accès des mineurs à des distributeurs automatiques de produits sensibles tels que l'alcool et le tabac.

En ce qui concerne les applications de surveillance on cite l'exemple de la détection des enfants sans surveillance à des heures et lieux inhabituels [10], Ou encore celles qui contrôlent l'accès à internet afin d'adapter le contenu des sites en fonction de l'âge et de protéger les enfants sur les réseaux sociaux [34].

Application légale : la synthèse de l'âge du visage est la technique la plus utile dans ce domaine ; telle qu'une application utilisée dans la détection des crimes qui aide les experts médico-légaux confrontés à la nécessité d'identifier et d'estimer l'âge d'un cadavre et également d'identifier l'âge des terroristes et les enfants disparus, etc [10].

Application de santé et soins les applications d'estimation de l'âge peuvent également être

utilisées dans les systèmes de soins de santé, comme une infirmière robotisée pour déplacer, porter, communiquer avec les patients de différents groupes d'âges et même de sauvgarder leurs dossiers médicaux ou encore une unité de soins intensifs intelligentes pour des services personnalisés, etc [10].

Applications gouvernementales : l'estimation de l'âge est un type de biométrie qui peut être utilisé pour compléter les caractéristiques biométriques primaires afin d'améliorer les performances d'un système biométrique. Par exemple l'âge est une notion fondamentale en démographie, ne peut être directement mesuré pour la plupart des populations du passé car elles ne connaissaient pas leurs état civil, on peut seulement l'estimer d'après des indicateurs biologiques ou encore les contrôles de passeport en cas de grand écart d'âge entre l'image du passeport et la personne en question[10], Un bon exemple de cet usage est l'aéroport de Francfort (Allemagne) où on l'utilise pour automatiser le contrôle des passagers,le contrôle des frontières, etc.

Application juridique : un système armé d'une unité d'estimation d'âge décente peut être très pratique pour filtrer les suspects possibles de manière plus précise et efficace, telle que l'identification des personnes qui se déclarent être mineures à l'étrangers non accompagnées (en exil). Cette estimation est nécessaire car le droit définit les statuts des mineurs (moins de 10, 13, 15, 18 ans) pour en tirer des conséquences juridiques adaptées à l'âge. La détermination de l'âge des enfants et adolescents étrangers est donc indispensable pour définir les droits et protections dont ils peuvent ou doivent bénéficier en fonction de ce statut.

Application d'emploi : Certains emplois gouvernementaux comme l'armée et la police considèrent l'âge d'une personne comme une exigence. Les systèmes d'estimation de l'âge peuvent être utilisés pour déterminer l'âge des recrues pendant le processus de recrutement. Plusieurs gouvernements ont également pour politique de mettre les employés à la retraite après avoir atteint un certain âge. Les systèmes d'estimation de l'âge pourraient également jouer un rôle important pour déterminer si une personne a atteint l'âge de la retraite [10].

De nombreuses méthodes d'estimation de l'âge à partir d'images faciales sont apparues ces dernières années. Dans le titre suivant nous allons présenté quelques travaux connexes.

1.6 Travaux connexes

Les méthodes d'estimation de l'âge comprennent la classification, la régression, le classement et la combinaison. Ces méthodes d'estimation de l'âge peuvent être utilisées en conjonction avec diverses méthodes d'extraction de caractéristiques pour bien mener la tâche d'estimation de l'âge.

Au stade d'extraction des caracteristique, Kwon et al.[68] ont d'abord utilisé des modèles anthropométriques pour extraire les caractéristiques d'âge du visage et sur la base de la théorie du développement cranio-facial et des caractéristiques des rides de la peau. La base de données complète contient 47 visages comprenant des bébés, des adultes jeune/moyen d'âge et des per-

sonnes âgées, les images zoomées n'ont été obtenues que pour 15 visages. Pour ces 15 visages les classifications étaient correctes à 100%.

Kanno et al.[63] ont proposé une méthode de classification par groupe d'âge des jeunes hommes à partir de leurs images faciales. L'étude n'a pris en compte que les jeunes hommes, car ils ont une période plus longue pendant laquelle la forme du visage est un facteur déterminant dans l'estimation de l'âge. La classification de l'âge a été réalisée à l'aide de réseaux neuronaux artificiels. 440 images faciales sont utilisé dans l'expérience, composées de 4 images photographiques différentes prises aux âges de 12, 15, 18 et 22 ans de 110 jeunes hommes. Deux méthodes de classification de l'âge ont été utilisées, chacune employant des caractéristiques différentes extraites des images faciales, à savoir les "caractéristiques de la mosaïque" et les "caractéristiques KL (Kullback-Leibler)". Le taux de classification obtenu est d'environ 80% avec les caractéristiques en mosaïque et un taux légèrement inférieur avec les caractéristiques KL.

Lanitis et al.[71] ont été les premiers à appliquer le modèle d'apparence active (AAM : Active Appearance Model) à l'étude de l'estimation de l'âge des visages. Sur une base de données contenant 500 images progressives de 45 individus. Sur la base des caractéristiques de l'AAM, une fonction de régression quadratique a été utilisée pour l'estimation de l'âge et on obtenu un résultat de 66%.

Iga et al.[57] ont développé des fonctions d'extraction d'une région candidate au visage avec des informations de couleur de la peau avec le système de couleur HSV, mais aussi afin de détecter la position exacte du visage et de ses parties en détectant les points caractéristiques dans la région candidate les filtres de Gabor, arrangement géométrique et la texture sont utilisé. Ensuite ils ont utilisé des classificateurs SVM qui ont été formés avec la base de données HOIP avec 300 personnes japonais. Les résultats expérimentaux ont montré des taux de réussite de 58,4% pour l'âge.

Lanitis et al.[70] proposent une méthode pour générer des modèles statistiques à partir d'un ensemble d'exemples d'apprentissage . En effectuant une analyse en composantes principales (PCA) la tâche d'estimation de l'âge est effectuée sur une base de donnée de 400 image en utilisant quatre types de classifieurs différents : un classifieur basé sur des fonctions quadratiques, un classificateur de la plus courte distance, un perceptron Multicouche (MLP : Multilayer Perceptron) et une carte auto-organisatrice (SOM : a Self-Organizing Map) avec l'erreur absolue moyenne en année (MAE : Mean Absolute Error) est de 5.04, 5.65, 4.78, 4.9 respectivement au classifieur.

Zhou et al. [122] ont présenté un algorithme IBR général de régression basée sur l'apparence (l'image) qui est applicable à de nombreux problèmes de vision. Le régresseur proposé cible un paramètre à sorties multiples, et il a appris en utilisant la méthode de boosting pour la sélection des caractéristiques pertinentes des images de la base de donnée FG-NET [69] avec 1002 images. Le régresseur est évaluer en utilisant trois tâches difficiles : l'estimation de l'âge, la détection de tumeurs et la localisation de la paroi endocardique. La MAE rapportée en estimation d'age est de 7,48.

Ueki et al. [112] ont proposé sur une approche basée sur l'apparence, c'est une approche en

deux phases pour la classification des groupes d'âge en utilisant des images faciales dans différentes conditions d'éclairage. Première phase 2DLDA (2-Dimensional Linear Discriminant Analysis) : pour la redimention et la seconde phase LDA (Linear Discriminant Analysis) : pour l'extraction des caracteristiques ; et cela en utilisant des images faciales dans différentes conditions d'éclairage, sur WIT-DB (Waseda human-computer Interaction Technology - DataBase) [112] avec environ 2500 images de femmes et 3 000 hommes. Le taux de precision obtenu es de 50% pour les hommes et de 43% pour les femmes.

Yan et al .[119] ont proposé un système d'estimation de l'âge basé sur le classement et ont affiché les caractéristiques d'âge d'un rang faible à un rang élevé avec des étiquettes indéfinies. Ils ont expérimenté sur les bases de données YGA (Yahama Gender et Age) [119] et bases de données FG-NET [69]. La performance de leur modèle a été estimée à 5,33 MAE sur FG-NET et un CS (Cumulative Score) de 79% avec un MAE de 6,95 sur la base de données YGA.

Geng et al.[39] ont proposé une méthode d'estimation automatique de l'âge appelée AGES (AGing pattErn Subspace). Ce modèle est basé sur un modèle d'apparence et a proposé un concept de sous-espace des modèles de vieillissement (AGES). les bases de données FG-NET [69] et MORPH [100] sont utilisé dans l'expérience une MAE de 6.22 et un CS de 80% sont obtenu sur FG-NET et une MAE de 8.07 et un CS de 70% sur MORPH.

Guo et al.[47] ont introduit l'apprentissage par collecteur dans l'estimation de l'âge des images de visage, et a mis en correspondance des ensembles de données de visage à haute dimension avec des collecteurs à faible dimension, c'est-à-dire qu'il a transformé les images de visage en une caractéristique d'âge à faible dimension. Le travail est effectué sur 2 bases de donnée : UIUC-IFP-Y [37] avec 800 images de femme et 800 images d'homme et ont obtenu une MAE de 5.25 et 5.30 et un CS 83% (H),81% (F) respectivement, d'autant plus sur FG-NET [69] une MAE de 5.07. une fonction de régression quadratique a été utilisée pour l'estimation de l'âge.

Peu de temps apres aussi Guo et al. [46] ont proposé un algorithme de régression robuste ajustée localement (LARR : Locally Adjusted Robust Regression) qui combine le SVM (Support Vector Machine) et le SVR (Support Vector Regression), pour l'estimation de l'âge. D'abord le SVR pour effectuer une estimation exacte de l'âge et le SVM pour estimer une plage d'âge globale. Ces experiences sont realiser sur les bases de données UIUC-IFP-Y [37] Le meilleur résultat LARR en termes de MAE est de 5,25 ans pour les femmes avec 64 classes, tandis qu'il est de 5,30 ans pour les hommes quand la plage est de 32 classes. et pour FG-NET [69] la MAE est de 5.07 quand la plage est de 4 et 8 classes.

Gunay et al. [44] ont appliqué le modèle binaire local (LBP) à l'estimation de l'âge, sur la base de données FERET [94] avec 350 image et ont obtenu de bons résultats avec 80%, le classificateur de voisinage le plus proche est utiliser. Depuis lors, de nombreuses méthodes améliorées d'estimation de l'âge des visages basées sur le LBP sont apparues.

Suo et al. [106] ont proposé un modèle graphique hiérarchique de visage quatre types de

caractéristiques sont extraits de cette représentation graphique : topologie représente l'indice des styles de cheveux, géométrie L'AAM global au premier et second niveau donnent une localisation précise des points de repère du visage, photométrie décrit l'apparence du visage : la couleur, les intensités basse fréquence et les intensités haute fréquence des cheveux , rides et la peau et configuration comporte un ensemble de mesures de distance. Ce qui concerne l'estimation de l'âge la méthode proposée suit quatre types de régressions : âge régression linéaire (ALR : Age Linear Regression), perceptron multicouche (MLP), SVR et régression logistique (Adaboost multi-classes). Parmi les régressions utilisés, MLP donne la meilleure performance avec une MAE de 4,68 ans sur sa base de données collectées et le taux d'estimation atteint 91,6% et aussi la MAE est de 5,9 sur FG-NET [69].

Yan et al. [118] ont utilisé le descripteur de caractéristiques patch flexible dans l'espace (SPF :Spatially Flexible Patch) pour l'extraction de caractéristiques locales et leurs positions spécifiques sur le visage, d'autres aspects tels que la pose de la tête et l'occlusion ont également pu être traités efficacement. Ils ont fait des expériences sur la base de données Yamaha [118], les performances de leur système ont montré une MAE de 7,82 et un CS de 75% pour les hommes et une MAE de 8,53 et un CS de 70% pour les femmes.

Fu et al.[37] ont défini la formulation du collecteur par l'analyse d'emboîtement conforme (CEA : Conformal Embedding Analysis) pour représenter un sous-espace de faible dimension. L'âge statistique final a été estimé en utilisant régression linéaire multiple (MLR : multiple linear regression). Ils ont obtenu une MAE de 5-6 ans sur l'ensemble de données collecté.

Guo et al.[48] ont également proposé une caractéristique bio-inspirée (BIF : Biologically Inspired Features), qui a attiré l'attention sur l'estimation de l'âge en raison de meilleurs résultats expérimentaux ce qui est de l'estimation de l'age l'approche utilisé est le SVM linéaire pour la classification de l'âge et le SVR pour la regression de l'âge. Les technologies d'apprentissage profond telle que CNN, ont été progressivement appliquées à l'estimation de l'âge. Deux bases de données sont utiliséeé YGA (Yahama Gender et Age) [119]qui contient 8000 images et FG-NET [69]. les résultats obtenus est une MAE de 3.91 pour femme et 3.47 pour homme sur YGA, et une MAE de 4.77 sur FG-NET.

Hajizadeh et al [50] ont utilisé HOG pour l'extraction de caractéristiques et ont classé les images en 4 groupes d'âge en utilisant un réseau neuronal probabiliste. Les performances de leur système ont atteint une précision de 87,025% sur le jeu de données IFDB (Iranian Face Database) [14] avec 3600 images.

Choi et al.[22] ont proposé une approche d'estimation de l'âge utilisant des classificateurs hiérarchiques avec des caractéristiques faciales locales et globales. Les filtres de Gabor ont été utilisés pour les rides tandis que le LBP a été utilisé pour l'extraction des caractéristiques de la peau. Ils ont classé les images de visage dans des groupes d'âge en utilisant des machines à vecteurs de support. Pour une estimation précise de l'âge, la classification des groupes d'âge doit être robuste, ce qui peut être obtenu en utilisant un ensemble de classificateurs. La performance

de leur système a atteint une MAE de 4,7 et un CS de 73% sur la base de données FG-NET [69] et une MAE de 4,3 avec un CS de 70% sur la base de données PAL [84].

Bekhouche et al. [15] la méthode proposé est appliquée en plusieurs étapes. Premièrement ils ont effectué un traitement sur l'image où toutes les images de visage en couleur sont converties en niveaux de gris. Deuxièmement l'extraction des caractéristiques par les méthodes LBP et BSIF, les caractéristiques de l'histogramme sont obtenus en un vecteur composé de 13 histogrammes pour BSIF et LBP par la suite les combiner pour obtenir 26 caractéristiques. Troisièmement l'estimation de l'âge qui est traitée comme un problème de régression, deux algorithmes sont utilisés : LIBSVM pour SVR et SimpleR pour KRR (kernel Ridge Regression). Le travail est effectué sur deux bases de données PAL[84] avec 1046 images et FG-NET. ce qui est des performances de l'estimation de l'âge sur le PAL, SVR+BSIF+LBP et KRR+BSIF+LBP donnent la plus petite 6.25 et 6.38 respectivement. sur FG-NET [69] KRR+BSIF et SVR+BSIF+LBP donnent le plus petit MAE 6.28 et 6.34 respectivement.

Wang et al.[114] ont appliqué le CNN à l'extraction des caractéristiques de l'âge du visage. Cependant, le CNN n'est utilisé que pour l'extraction de caractéristiques, puis il est introduit dans un modèle de classification ou de régression distinct pour l'estimation de l'âge. Dans ce travail, les SVM pour la classification de l'âge et les SVR, PLS (Partial Least Squares), et CCA (Canonical Correlation Analysis) pour la régression de l'âge sont utilisés. Les expériences ont été menées sur deux jeux de données MORPH avec 5475 images et FG-NET. La MAE obtenue sur MORPH [100] est de 4.77 et la MAE de FG-NET [69] est de 4,26.

Bouchrika et al. [17] ont proposé une approche basée sur la vision pour l'estimation de l'âge d'un individu par le biais des caractéristiques du visage. L'opérateur de modèle binaire local (LBP) est appliqué pour dériver un ensemble hybride de caractéristiques comprenant caractéristiques locales et globales du visage. La sélection hiérarchique des caractéristiques est décrite pour le processus de classification où les tranches d'âge sont classées de manière arborescente. les experiences ont été menées sur BW kennedy [64] avec 180 images, en exécutant le processus hiérarchique de classification le taux de classification correcte atteint 91.1%.

Huerta et al [55] ont proposé un système basé sur la classification qui fusionne les descripteurs d'apparence et de texture locale HOG, SURF(Speeded Up Robust Features) et AAM qui a donné de très bons résultats d'estimation de l'âge lorsqu'il a été expérimenté sur les ensembles de données MORPH [100] et FRGC (Face Recognition Grand Challenge) [55]. Ils ont obtenu une MAE de 4,25 ans avec un CS de 71,17% sur le jeu de données MORPH et 4,17 ans avec un CS de 76,24% sur le jeu de données FRGC.

Cai et al. [19] ont proposé un impressionnant système d'estimation de l'âge basé sur la découverte du collecteur à faible dimension. Les caractéristiques associées ont été extraites à l'aide de l'histogramme double LBP et testées sur les ensembles de données FG-NET [69] et MORPH [100]. Ils ont obtenu une MAE de 4,64 ans avec un CS de 72,1% sur l'ensemble de données sur le vieillissement FG-NET et de 4,66 ans avec un CS de 62,2% sur l'ensemble de données MORPH.

Niu et al.[90] ont mis en œuvre une méthode d'apprentissage de bout en bout qui utilise des réseaux neuronaux convolutifs CNN profonds pour effectuer simultanément l'apprentissage de caractéristiques et la modélisation de régression. Les repères faciaux des images de visage sont localisés par l'AAM. Les experiences sont menées sur deux bases de données : MORPH II [100] contient 55 608 images de visages et AFAD (Asian Face Age Dataset) [90] avec 160k images. La performance de la methode est mesurée par MAE, sur la base de donnée MORPH II la MAE est de 3,27 et sur AFAD la MAE est de 3,24.

Feng et al. [36] sur leur système d'estimation de l'âge basé sur le classement ont proposé une autre approche AAM consistant à classer chacune des étiquettes d'âge en estimant leur importance pour l'image faciale. Ils ont rapporté une MAE de 4,35 ans sur FG-NET [69], 4,59 ans sur MORPH et 6,03 ans sur le jeu de données Webface [89].

Pontes et al.[95] ont fusioné l'AAM pour l'extraction de caractéristiques globales avec des méthodes basées sur la texture, principalement LBP, GW et LBP, GW et LPQ pour extraire les caractéristiques locales, cette structure hiérarchique flexible l'ont proposé comme approche pour l'estimation de l'âge. Leur système était un hybride de classification et de régression, le SVM étant utilisé pour classer l'âge en groupes d'âge tandis que le SVR estime l'âge exact final. Ils ont obtenu une MAE de 4,50 ans sur l'ensemble de données de vieillissement FG-NET, et de 5,85 ans sur l'ensemble de données MORPH.

Hu et al. [54] ont proposé un système basé sur la classification par apprentissage qui repose sur la différence d'âge. Afin de tirer profit des données faiblement étiquetées, ils ont utilisé des CNN GoogLeNet ainsi que la divergence de Kullback-Leibler pour localiser l'information de différence d'âge entre les paires d'images. Ils ont mené leurs expériences sur les jeux de données MORPH-II [100] et FG-NET [69]. Ils ont rapporté une MAE de 2,78 ans sur MORPH-II et de 2,8 ans sur FG-NET.

Qawaqneh et al.[96] a utilisé les réseaux neuronaux profonds appelés VGG-face formé pour la reconnaissance des visages sur une très grande base de données Adience [33]. Premièrement il a proposé une architecture CNN qui s'appuie sur un CNN de reconnaissance faciale très profond en capable d'extraire les caractéristiques faciales de manière distincte. Ensuite les images d'entrées sont redimensionnées puis recadrées ; l'optimisation du réseau est effectué en utilisant la méthode de descente du gradient stochastique qui minimise la prédiction pour l'estimation de l'âge. Finalement une classification est faite. En comparaisons avec d'autres résultats précèdent, cette méthode surpasse significativement les résultats de l'état de l'art en termes de précision exacte avec 59.6%.

Chen et al.[21] ont proposé un cadre de réseau CNN d'estimation de l'âge : ranking-CNN, qui comprend une série de réseaux CNN de base, chacun d'entre eux entraînant une étiquette, leurs sorties binaires sont agrégées pour la prédiction finale de l'âge. Le ranking-CNN surpasse les autres extracteurs de caractéristiques et estimateurs d'âge sur des jeux de données de référence, MORPH [100], FG-NET [69] et Adience [33] avec ses 26,580 images, et en obtenant la plus faible MAE sur les 3 bases de donnée avec la MAE de 2.92 , 4.13 et 4.4 respectivement.

Liu et al.[77] ont proposé une nouvelle méthode d'apprentissage de caractéristiques d'estimation de l'âge du visage, qui est une approche d'apprentissage de caractéristiques profondes ordinales (ODFL : ordinal deep feature learning) pour apprendre des descripteurs de visage pour l'estimation de l'âge du visage. L'ODFL est evalué sur quatre jeux de données de référence, notamment MORPH II [100], FG-NET [69], FACES [32] et l'estimation de l'âge apparent du visage [35]. La méthode proposer obtient une MAE, 3.12 sur MORPH, 3.89 sur FG-NET, 4.12 sur ÂGE APPARENT, et des MAE les plus faible sur six expressions du visage.

Li et al. [75] ont proposé une nouvelle méthode qui utilise AlexNet pour l'extraction de caractéristiques avec une couche cachée cumulative qui a pour principal avantage de surmonter le problème du déséquilibre de l'échantillon en apprenant indirectement par les visages d'âges voisins. Une couche supplémentaire a été ajoutée comme une amélioration de leur système d'estimation de l'âge basé sur la régression afin d'effectuer un classement comparatif dans le but de de faciliter l'apprentissage des caractéristiques de vieillissement et d'améliorer ainsi la performance globale. Grâce à leurs expériences, ils ont rapporté une MAE de 3,06 ans sur l'ensemble de données MORPH-II [100] et de 6,04 ans sur l'ensemble de données Webface [89].

Duan et al [31] ont proposé un autre modèle qui utilise un réseau d'apprentissage profond pour l'extraction de caractéristiques et a ensuite transmis la sortie à l'apprentissage automatique extrême (ELM :Extreme learning machine) pour la classification des groupes d'âge, puis a régressé la valeur finale de l'âge via le régresseur ELM. Le résultat final obtenu était une MAE de 2,61 ans sur la base de données MORPH-II [100].

Rothe et al. [101] ont proposé une solution d'apprentissage profond pour l'estimation de l'âge à partir d'une seule image de visage sans utiliser de repères faciaux. Ils ont amélioré les performances de leur système en affinant les CNN VGG-16 sur la base de données IMDB-WIKI sans contrainte, puis ont testé le système basé sur la régression sur les bases de données MORPH-II [100], FG-NET [69] et CACD(Cross Age Celebrity Dataset) [20]. Ils ont rapporté une MAE 2,68 ans sur MORPH-II, 3,09 ans sur FG-NET et 6,521 sur le jeu de données CACD.

Liu et al. [79], ont développé un réseau CNN léger (ShuffeNetV2), basé sur le mécanisme d'attention mixte (MA-SFV2) qui transforme la couche de sortie, qui façonne l'estimation de l'âge comme un problème de classification (qui classe l'âge comme une étiquette distincte), un problème de régression (qui classent l'âge du visage humain selon un ordre particulier) et l'apprentissage de la distribution (qui prend en compte l'âge du visage humain dans un ordre particulier) et l'apprentissage de la distribution (qui prend en compte la corrélation entre les âges adjacents). Le modèle comprend un prétraitement de l'image qui réduit l'effet des vecteurs de bruit et une méthode d'augmentation des données comme le filtrage, l'accentuation, l'amélioration de l'histogramme, ... qui augmente la taille de l'image et atténue l'adaptation excessive du réseau. Il combine des algorithmes de classification, de régression et d'apprentissage distribué pour la tâche d'estimation de l'âge. Les résultats expérimentaux obtenus sont une MAE de 2,68 ans sur MORPH-II [100] et de 3,81 ans sur la base de données FG-NET [69].

Micheala et al.[83] ont proposé une méthode de représentation par l'utilisation des CNN et ELM. Le CNN est utilisé pour extraire les caractéristiques des images d'entrée tandis que ELM classifie les résultats efficacement en raison de ses bonnes performances en matière de généralisation, de sa rapidité d'apprentissage et de son faible niveau d'intervention humaine. l'experience est faite sur la base de données Adience [33] pour l'estimation de l'âge et du sexe. Les résultats expérimentaux montrent que cette architecture surpasse les autres études en présentant une amélioration significative des performances en termes de précision qui atteint les 90.2%.

Nous avons également dans la suite de ce paragraphe un tableau récapitulatif des méthodes existantes

1.7 Tableau récapitulatif des méthodes existantes

dans cette partie, nous avons récapitulé ces méthodes dans le tableau 1.1 :

Publication	discripteur	algorithme	Database	Performance		
				MAE	CS	Accu
Kwon et al. 1999 [68]	Modèle anthropométrique	Classification	dataset privée 15 images	N/A	N/A	100%
Kanno et al. 2001 [63]	Modèle apparence	Classification	dataset privée 440 images	N/A	N/A	80%
Lanitis et al. 2002[71]	Forme 2D, valeurs de pixels brutes	Régression	dataset privée 500 images	N/A	N/A	66%
Iga et al. 2003[57]	Modèle HSV, filtres de gabor, forme 2D	Classification	HOIP avec 300 personnes japonaise	N/A	N/A	58.4%
Lanitis et al. 2004[70]	Modèle active apparence AAM	Classification régression	dataset privée 40 images	5.04, 5.65, 4.78, 4.9	N/A	N/A

Zhou et al. 2005 [122]	Modèle anthropométrique	Régression	FG-NET	7,48	N/A	N/A
Ueki et al. 2005 [112]	Modèle anthropometrique	Classification	WIT-DB	N/A	N/A	50% (H) 43% (F)
Yan et al. 2007 [119]	Modèle apparence	Régression	YGA, FG-NET	6.95, 5.33	79%	N/A
Geng et al. 2007 [39]	AGES	Régression	FG-NET et MORPH	6.22, 8.07	80%, 70%	N/A
Guo et al. 2008 [47]	AAM	Régression	UIFP-Y et FG-NET	5.07, UFy : 5.25, 5.30	UFY : 83% 81% (F)	N/A
Guo et al. 2009 [46]	Distributeur d'âge, Bif	Régression	UIFP-Y, FG-NET	5.25(F) , 5.30(H), 5.07	N/A	N/A
Gunay et al. 2008 [44]	LBP	Classification	FERET avec 350 image	N/A	N/A	80%
Suo et al. 2008 [106]	Apparence photométrique	Régression	dataset privée et FG-NET	4.68, 5.9	81.9	N/A%
Yan et al. 2008 [118]	Modèle apparence, SPF	Régression	YGA	7,82(H) 8,53(F)	75%(H), 70%(F)	N/A
Fu et al. 2008 [37]	CEA	Régression	YGA	5-6	N/A	N/A

Guo et al. 2009 [48]	Apparence d'âge, BIF	Classification et Régression	YGA, FG-NET	3.91(F), 3.47(H), 4.77	N/A	N/A
Xiao et al. 2009 [117]	AAM	Régression	FG-NET	4.93	N/A	N/A
Takimoto et al. 2009 [110]	model apparence	Régression	HOIP	5.	N/A	N/A
Hajizadeh et al. 2011 [50]	HOG	Classification	IFDB	N/A	N/A	87,02 %
Choi et al. 2011 [22]	Gabor, LBP	Classification	PAL et FG-NET	4.3, 4,7	70% , 73%	N/A
Wu et al. 2012 [116]	Collecteur Grassmann de forme 2D	Hybride	FGn-NET	8,84	N/A	N/A
choobeh 2012 [24]	Gabor 2D, PCA	Classification	FG-NET	4.85	N/A	N/A
Geng et al. 2013 [38]	AAM, étiquette distribution	Classification	FG-NET	5.77	N/A	N/A
Bekhouche et al. 2014 [15]	BSIF, LBP	Régression	PAL et FG-NET	6.25 et 6.28	N/A	N/A
Wang et al. 2015 [114]	CNN	Classification, Régression	MORPH et FG-NET	4.77, 4.26	N/A	N/A
Bouchrika et al. 2015 [17]	LBP	Classification	BW kennedy	N/A	N/A	91.1%

Huerta et al. 2015 [55]	HOG,LBP, SURF	Classification	MORPH et FRGC	4.25, 4.17	71,17% et 76,24%	N/A
Cai et al. 2016 [19]	HLBP	Régression	MORPH et FG-NET	4.66, 4,64	62,2% , 72,1%	N/A
Niu et al. 2016 [90]	CNN	Régression	MORPH II, AFAD	3,27, 3,24	N/A	N/A
Feng et al. 2016 [36]	AAM	Classification, Régression	Webface, MORPH et FG-NET	6,03, 4.35, 4.59	N/A	N/a
Pontes et al. 2016 [95]	AAM,LBP, LPQ,Gabor	Hybrid	FG-NET et MORPH-II	4.50, 5.85	N/A	N/A
Hu et al. 2016 [54]	GoogLeNet CNN	Classification	MORPH, FG-NET	2,78, 2,8	N/A	N/A
Qawaqneh et al. 2017 [96]	VGG-Face	Classification	Adience	N/A	N/A	59.6%
Chen et al. 2017 [21]	Ranking-CNN	Régression	MORPH, FG-NET et Adience	2.92, 4.13, 4.46	N/A	N/A

Liu et al. 2017 [77]	ODFL	Classification	MORPH II, FG-NET, FACES	3.12, 3.89, 4.12	N/A	N/A
Li et al. 2017 [75]	AlexNet	Régression	web face, MORPH II	6.04, 3.06	N/A	N/A
Duan et al. 2018 [31]	ELM	Classification	MORPH II	2,61	N/A	N/A
Rothe et al. 2018 [101]	VGG-16	Régression	MORPH-II, FG-NET et CACD	2.68, 3.09, 6.521	N/A	N/A
Liu et al. 2020 [79],	ShuffeNetV2	Classification Régression	MORPH et FG-NET	2.68, 3,8	N/A	N/A
Micheala et al. 2021 [83],	CNN	Classification	Adience	N/A	N/A	90.2%

TABLE 1.1 – Tableau récapitulatif des travaux connexes

1.8 Conclusion

Dans ce chapitre, nous avons défini l'estimation d'age, ensuite nous avons cité les différents problèmes rencontrés ainsi que les divers domaines d'application de l'estimation de l'âge dans le monde réel. Enfin, nous avons présenté quelques travaux connexes de l'estimation d'âge à partir des images faciales.

Chapitre 2

Chapitre 2.Aspect théorique des méthodes utilisées

2.1 Introduction

Le processus d'estimation d'âge d'un visage ne pourra jamais devenir intégralement automatique s'il n'a pas été précédé par une étape de détection efficace et pour cela plusieurs méthodes de détection de visage on été étudiées. Pour une meilleure exploitation de données, le cœur du système d'estimation d'âge est l'extraction des caractéristiques qui consiste à effectuer le traitement de l'image dans un autre espace de travail plus simple.

Dans ce chapitre, nous présentons quelques méthodes utilisée pour la détection de visage, ainsi que l'extraction des caractéristiques et leurs aspects théoriques.

2.2 Détection de visage

Le sujet de la détection de visage humain est très ancien, en raison de son importance pratique et théorique, il reste toujours un centre important de recherche. Cela est motivé par la multiplicité et la variété des champs d'applications tels que télésurveillance, haute sécurité, contrôle d'accès, etc. Les chercheurs ont montré que les humains reconnaissent un visage grâce à ces différentes caractéristiques, qui varient entre la texture, la géométrie et la couleur des différentes zones du visage. Grâce à cette remarque, plusieurs études ont été menées pour voir s'il est possible de modéliser ce comportement de manière informatisée. Cependant, pour assurer de bons résultats dans l'étape de reconnaissance, le problème de détection de visage est un processus essentiel [105].

Le visage est considéré comme une donnée biométrique, cette dernière est une donnée qui permet l'identification d'une personne sur la base de ce qui est caractéristique physiologique, comportementale ou morphologique. La détection de visage est un domaine de vision par ordinateur qui consiste à trouver les coordonnées spatiales qui définissent un visage dans une image ou une vidéo. En termes simples, cela revient à trouver le carré qui définit le mieux le visage visible sur l'image [76].

La détection de visage est un sujet difficile, notamment dû à la grande variabilité d'apparence des visages dans des conditions sans contraintes. Il est important de définir la qualité d'une détection qui se caractérise par trois types :

Le premier type est détection positive qui pointe un visage. Le deuxième est détection fausse positive dont la zone détectée ne correspond pas à un visage. Le troisième est non-détection dont il existe un ou plusieurs visages non détectés dans l'image.

Depuis quelques années la détection des visages est prise comme domaines de recherches par plusieurs personnes et sociétés, ce qui amène à l'existence de plusieurs méthodes pour la détection des visages dans une image.

2.3 Méthodes de détection de visage

Il existe plusieurs méthodes différentes pour la détection de visage telles que des méthodes basées sur les CNN's et des méthodes basées sur l'apparence.

2.3.1 Modèles de détection par Convolutional Neural Networks CNN

Il existe deux types de modèles de détection de visage par les architectures CNN. Le premier type de détection en deux coups est basé sur la proposition de région et comprend des modèles tels que R-CNN, Fast R-CNN, Faster R-CNN, R-CNN, Mask, MTCNN et le second type la détection à un coup est basé sur la régression et comprend SSH, YOLO, SSD, etc. [88]

— Détection en deux coups : Comme son nom l'indique, cette méthode comporte deux étapes. La première est la proposition de régions, puis, dans la deuxième étape, la classification de ces régions et le raffinement de la prédiction de l'emplacement.

— Détection à un coup saute l'étape de la proposition de région et produit la localisation finale et la prédiction du contenu en une seule fois [37]. Il existe plusieurs modèles de détection de visage :

2.3.1.1 R-CNN : Regions with CNN

Le model de détection de visage R-CNN a été proposé par Ross Girshick et all en 2014 [42], ce modèle se compose de trois modules :

• Génération de propositions régionales : indépendantes de la catégorie, qui définit l'ensemble des détections candidates disponibles pour notre détecteur.

• Extraction de caractéristiques : le deuxième module est un grand réseau neuronal convolutionnel qui extrait un vecteur caractéristique de longueur fixe de chaque région.

• Classification et localisation : Le troisième module est un ensemble de SVM linéaires spécifiques à chaque classe.

Le R-CNN atteint une précision moyenne (mAP) de 53,7% sur PASCAL VOC 2010[42].

2.3.1.2 Fast R-CNN

Ross Girshick en 2015 [41] a proposer un nouvel algorithme d'apprentissage qui corrige les inconvénients du R-CNN et du SPPNet, tout en améliorant leur vitesse et leur précision. Cette méthode est appelé le R-CNN rapide (Fast R-CNN) car elle est comparativement rapide à entraîner et à tester. Ce réseau prend en entrée une image entière et un ensemble de propositions d'objets. Le réseau traite d'abord l'image entière avec plusieurs couches convolutionnelles (conv) et de mise en commun maximale pour produire une carte de caractéristiques. Ensuite, pour chaque proposition d'objet, une couche de mise en commun des régions d'intérêt (RoI) extrait un vecteur de caractéristiques de longueur fixe de la carte de caractéristiques. Chaque vecteur de caractéristiques est introduit dans une séquence de couches entièrement connectées (FC : Fully Connected) qui se ramifient finalement en deux couches de sortie sœurs :

• Une couche qui produit des estimations de probabilité softmax sur K classes d'objets plus une classe de "fond".

• Une couche qui produit quatre nombres à valeur réelle pour chacune des K classes d'objets. Chaque ensemble de 4 valeurs code les positions raffinées de la boîte de liaison pour l'une des K classes.

Le Fast R-CNN obtient un résultat de 66.1% sur le VOC2010 et le meilleur résultat sur le VOC12 avec un a mAP de 65,7 % (et 68,4 % avec des données supplémentaires) [41].

2.3.1.3 Faster R-CNN

Ce réseau proposé par Shaoqing Ren et al en 2016 [99], il se décompose en deux modules principaux :

• Le premier module est un réseau convolutif profond qui crée la carte de caractéristiques convolutives qui utilise un module RPN (Réseau de Proposition de Région) qui prend cette carte de n'importe quelle taille et produit un ensemble de propositions d'objets rectangulaires, chacune avec un score de précision.

• Le second module est le détecteur Fast R-CNN qui utilise les régions proposées [99] comme nous avons vu dans l'architecture fast R-CNN.

Le résultat obtenu par ce model avec la base de données mscoco (COCO val) avec mAP c'est 41.5% et on COCO test-dev avec mAP@[0.5] c'est 42.7% [99].

2.3.1.4 G-CNN Iterative Grid Based Object Detector

Najibi et al en 2016 [87] G-CNN entraîne un CNN à déplacer et à mettre à l'échelle une grille fixe à plusieurs échelles de boîtes englobantes vers des objets.

Le G-CNN définit le problème de la détection d'objets comme une recherche itérative dans l'espace de toutes les boîtes de délimitation possibles. Le G-CNN part d'une pyramide spatiale multi-échelles fixe de boîtes. Le but de l'apprentissage est d'entraîner le réseau de manière à ce qu'il puisse déplacer cet ensemble de boîtes initiales vers les objets de l'image en plusieurs étapes, de manière itérative ; Ce comportement itératif est essentiel pour le succès de l'algorithme. [87]. Le résultat obtenu par ce modèle avec la base de donnée COV 2007 avec mAP est de 57.2%

2.3.1.5 Mask R-CNN

Kaiming He et al en 2017 [52] ont proposé une autre architecture appelé Mask R-CNN, étendu de Faster R-CNN en ajoutant une branche pour la prédiction d'un masque d'objet en parallèle avec la branche existante pour la reconnaissance de la boîte englobante.

Le Mask R-CNN masqué est conceptuellement simple : Le Faster R-CNN a deux sorties pour chaque objet candidat, une étiquette de classe et un décalage de la boîte englobante ; à cela, une troisième branche qui produit le masque de l'objet a été ajouté.

La branche masque est un petit FCN (Fully Convolutional Network) appliqué à chaque RoI (Region of Interest), prédisant un masque de segmentation d'une manière pixel à pixel, ce principe d'alignement pixel à pixel c'est la pièce manquante du Fast/Faster R-CNN. Le résultat obtenu sur la base de données MS COCO + fine est 36.4 mAP sur la validation en utilisant ResNet-50-FPN [52].

2.3.1.6 SSH : Single Stage Headless Face Detector

Najibi et al en 2017 [88] ont conçu le SSH pour diminuer le temps d'inférence, avoir une faible mémoire, et être invariable à l'échelle. SSH est un détecteur d'objet à une seule étape, c'est-à-dire qu'au lieu de diviser la tâche de détection en deux parties : boîte englobante et en classification, il effectue la classification en même temps que la localisation à partir de l'information globale extraite des couches convolutives. Il a été montré que le SSH peut supprimer la "tête" de son réseau sous-jacent tout en obtenant une précision de détection de visage a la point de la technologie son réseau sous-jacent tout en atteignant une précision de détection de l'état de l'art. De plus, SSH est invariant à l'échelle par conception et peut incorporer le contexte efficacement[88].

2.3.1.7 SSD : Single Shot MultiBox Detector

L'approche de la SSD est basée sur un réseau convolutif feed-forward qui produit une collection de boîtes englobantes de taille fixe et des scores pour la présence d'instances de classes d'objets dans ces boîtes [78], suivi d'une étape de suppression non maximale pour produire les détections finales Les premières couches du réseau sont basées sur une architecture standard utilisée pour la classification d'images de haute qualité (VGG-16) (tronquée avant toute couche de classification), qui s'appelle réseau de base. Ensuite une structure auxiliaire au réseau pour produire des détections avec les caractéristiques clés suivantes :

- Cartes de caractéristiques multi-échelles pour la détection
- Prédicteurs convolutifs pour la détection
- Boîtes et aspect par défaut

Le résultat obtenu du modèle SSD512 entrainer sur COCO trainval35k puis affiner en pascal voc 2007+2012 c'est le meilleur résultat : 81,6% mAP, et 68.0% mAP avec le model SSD300 apprend sur voc 2007, et le résultat obtenu des modelés SSD300 et SSD512 entrainer sur la base de données COCO trainval35k est 41.2% mAP et 46.5% mAP.[78].

Dans ce qui suit nous allons définir les méthodes que nous avons utilisé dans notre travail pour la détection de visage : MTCNN, YOLO et les filtres de Haar.

2.3.1.8 MTCNN : Multi-Task Cascaded CNN

MTCNN un cadre développé comme solution pour la détection des visages. C'est un réseau de neurones qui détecte les visages et les repères faciaux sur les images(les yeux, le nez et la bouche). basée sur CNN proposée par Zhang et al [121]. En particulier, Lors de la prédiction des visages et des marqueurs, MTCNN est l'un des outils de détection de visage les plus populaires et les plus précis aujourd'hui [76].

MTCNN intègre les tâches de détection et d'alignement des visages à l'aide de CNN unifiés en cascade par apprentissage multi-tâches[81]. Le MTCNN lui même est composé de trois réseaux. Le premier réseau, appelé Réseau de Propositions (P-Net : Proposal Network), obtient principalement des fenêtres candidates et leur vecteur de régression de boîte limite et utilise la suppression non maximale (NMS) pour fusionner les boîtes qui se chevauchent fortement. Le deuxième réseau, connu sous le nom de Réseau de Raffinement (R-Net : Refinement network), est utilisé pour filtrer un grand nombre de faux candidats à partir du P-Net et calibre le rectangle de délimitation par régression. Enfin, le dernier réseau, appelé réseau de sortie (O-Net : Output Network), est utilisé pour filtrer un grand nombre de faux candidats, fournit en sortie les fenêtres de candidats finaux et les positions de cinq points de repère faciaux avec un réseau [76].

Une image est redimensionnée souvent à différentes échelles pour construire une pyramide d'images comme entrées du cadre en cascade à trois étages suivants.

Il y a trois tâches à effectuees pour entraîner les réseaux P-Net, R-Net et O-Net, qui sont la classification des visages, la régression de la boîte de délimitation et la localisation des points de repère faciaux [81]. Ces taches sont effectuer sur les bases de données FDDB [58] et WIDER FACE [120].

Son principe consiste à détecter des boites englobantes de visage ("faces bouding box") dans une image et cinq points de repère qui sont les deux point des yeux , un point du nez et deux point d'extremitée de la bouche, ces points sont appelés "Landmarks". La figure 2.1 montre un exemple de détection de visage avec MTCNN.

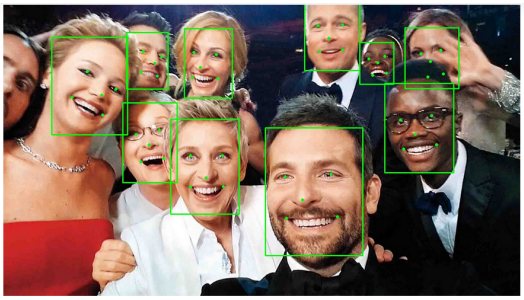

FIGURE 2.1 – Exemple de détection de visage avec MTCNN.

2.3.1.9 YOLO : You Only Look Once

You Only Look Once ou YOLO est l'un des algorithmes populaires de détection d'objets utilisé par les chercheurs du monde entier. Il a été décrit pour la première fois en 2016 dans l'article de Joseph Redmon et al [97]. Le mécanisme de l'algorithme fait appel à un seul réseau neuronal qui prend une photographie comme entrée et tente de prédire les boîtes englobante et les étiquettes de classe pour chaque boîte englobante directement.

L'architecture de ce réseau est inspirée du modèle GoogLeNet pour la classification d'images et a été évalué sur le jeu de données de détection de COV PASCAL avec une mAp de 63.4% [97]. Il se compose d'un total de 24 couches convolutionnelles suivies de 2 couches entièrement connectées. Les couches sont séparées par leur fonctionnalité de la manière suivante :

1. Les 20 premières couches convolutionnelles suivies d'une couche de mise en commun des moyennes et d'une couche entièrement connectée est pré-entraînée sur l'ensemble de données de classification ImageNet 1000 classes [97].

2. Le pré-entraînement po ur la classification est effectué sur un ensemble de données avec une résolution de 224×224 [97].

3. Les couches comprennent des couches de réduction 1×1 et des couches convolutives 3×3 [97].

4. Les 4 dernières couches convolutives suivies de 2 couches entièrement connectées sont ajoutées pour former le réseau à la détection d'objets [97].

5. La détection d'objets nécessite des détails plus granulaires, c'est pourquoi la résolution de l'ensemble de données est augmentée à 448×448 [97].

6. La dernière couche prédit les probabilités de classe et les boîtes englobantes[97]. La figure 2.2 montre un exemple de détection de visage par YOLO.

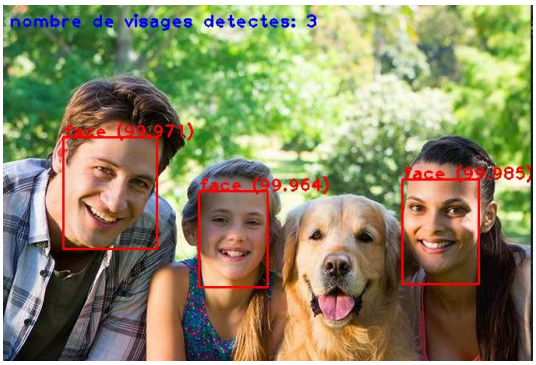

FIGURE 2.2 – exemple de détection de visage par Yolo

L'ensemble du système peut être divisé en deux composants principaux : L'extracteur de caractéristiques et le détecteur ; tous deux sont multi-échelles. Lorsqu'une nouvelle image arrive, elle passe d'abord par l'extracteur de caractéristiques afin d'obtenir des incorporations de caractéristiques à trois échelles différentes (ou plus). Ensuite, ces caractéristiques sont introduites dans trois branches (ou plus) du détecteur pour obtenir des boîtes englobantes et des informations sur la classe.

Il y a eu 6 versions du modèle jusqu'à présent, chaque nouvelle version améliorant la précédente en termes de vitesse et de précision. Dans notre travail nous avons utilisé le modèle YOLOv3.

Modèle YOLOv3

Joseph Redmon et Ali Farhadi en 2018 [98] on aussi fait une amélioration progressive dans la version précèdent, où ils ont utilisé comme réseau de base Darknet53(qui est une framework open source hautes performances pour la mise en œuvre de réseaux de neurones) pour extraction de caractéristique, ainsi que pour calculer un score de confiance de chaque boîte de délimitation en utilisant une régression logistique. Pour les prédictions de classe ils ont utilisés la perte d'entropie croisée binaire. Dans la version de YOLOv2 il y a un problème pour la détection des petits objets mais dans cette version la représentation des boîtes a 3 échelles différentes.

YOLO applique une grille sur l'image qui ne sert pas à segmenter l'image, mais plutôt pour analyser séparément chaque portion, puis utilise ces régions comme données d'entrée pour un CNN pour générer un certain nombre de boîtes englobantes (un rectangle sur l'image) qui permet au réseau de détecter plusieurs objets. La première étape, est de se débarrasser de toutes les boites englobantes qui ont une faible probabilité qu'un objet soit détecté. Cependant, même après un tel filtrage, on se retrouve avec de nombreuses boites englobantes pour chaque objet détecté.et enfin

il prend en entrée deux boîtes englobantes et comme son nom l'indique, il calcule le rapport de l'intersection et de l'union des deux.

2.3.2 Méthode basée sur l'apparence

Les modèles sont ici appris à partir d'un ensemble d'images d'apprentissage qui doivent permettre de caractériser la variabilité de l'apparence d'un visage. Ces méthodes se basent sur des techniques telles que l'analyse statistique et l'apprentissage automatique pour trouver les caractéristiques appropriées des images de visage et de non-visage.

L'une des méthodes les plus performantes est celle proposée par Viola et Jones[113], cette méthode est aussi utilisé dans la détection en temps réel, et cela est du grâce a son taux élevée de détection et au temps qu'elle prend pour l'exécution.

2.3.2.1 Viola et Jones : les filtres de Haar

La détection d'objets à l'aide de classificateurs en cascade, basée sur les fonctions de Haar est une méthode efficace de détection d'objet proposée par les chercheurs Paul Viola et Michael Jones [113]. Il s'agit d'une approche ou la fonction cascade est formée à partir de nombreuses images positives (images de visages) et négatives (images sans visages, qui est ensuite utilisée pour détecter des objets dans d'autres images. La méthode de Viola et Jones[113] est l'une des méthodes les plus connues et les plus utilisées, en particulier pour la détection de visages et la détection de personnes.

La méthode de Viola et Jones est la plus performante à l'heure. Ce qui la différencie des autres est notamment :

- l'utilisation d'images intégrales qui permettent de calculer plus rapidement les caractéristiques.
- la sélection par boosting des caractéristiques.
- la combinaison en cascade de classifieurs boostés, apportant un net gain de temps d'exécution.

La figure 2.3 montre un exemple de detéction de visage avec les filtres de Haar.

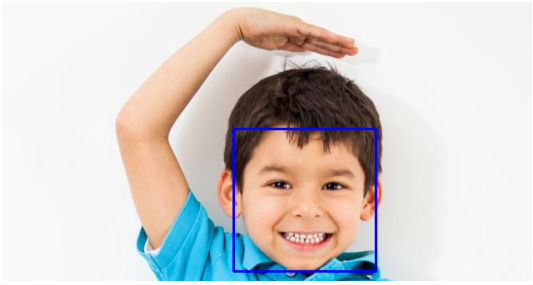

FIGURE 2.3 – Exemple de détection de visage avec les filtres de Haar.

L'étape de l'extraction des caractéristiques représente l'étape la plus essentielle du système d'estimation d'âge.

2.4 Extraction des caractéristiques

Cette étape consiste à des transformations mathématiques calculées sur les pixels d'une image numérique. Les caractéristiques nous permettent généralement de mieux se rendre compte de certaines propriétés de l'image.

Mais avant de faire l'étape d'extraction des caractéristiques et aprés la détction de visage. Le visage doit être aligner, effectivement l'alignement du visage est une tâche essentielle dans notre systhème. Cette étape a pour objectif de localiser des points caractéristiques sur le visage et cela permet de faire correctement l'étape d'extraction de caractéristique de visage.

2.4.1 Alignement

L'alignement des visages à partir d'images est un sujet de recherche actif, depuis les années 1990 [60]. L'alignement des visages joue un rôle important dans nombreuses applications telles que la reconstruction de visages, la reconnaissance des visages et la détection de visage, car il est souvent utilisé comme une étape de prétraitement.

L'alignement des visages est une technologie de vision par ordinateur permettant d'identifier la structure géométrique des visages humains dans les images numériques. Compte tenu de l'emplacement et de la taille d'un visage, il détermine automatiquement la forme des composants du visage tels que les yeux et le nez. Un programme d'alignement de visage fonctionne généralement en ajustant de manière itérative un modèle déformable, qui code la connaissance préalable de la forme ou de l'apparence du visage, pour prendre en compte les preuves d'image de bas niveau et trouver le visage présent dans l'image [60].

Récemment, l'alignement des visages a fait des progrès significatifs en théorie et en pratique. Bien que les approches basées sur l'AAM et les approches basées sur la régression fonctionnent bien pour les images de visages avec des poses réduites, elles ne peuvent généralement pas traiter

les images de visages de profil, car elles ne tiennent pas compte de la visibilité des points de repère [121]. Ces dernières années, plusieurs méthodes ont introduit le modèle morphale pour l'alignement des visages et obtiennent de meilleurs résultats. En utilisant un modèle de visage pour calculer la visibilité et la position des points de repère, ces méthodes peuvent gérer des cas difficiles avec de grandes variations de pose. Cependant, la précision de reconstruction de ces méthodes est souvent insuffisante [60]. La figure 2.4 montre un exemple d'alignemet d'un visage tel que (a) est le visage avant l'alignement et (b) le visage aprés l'alignement.

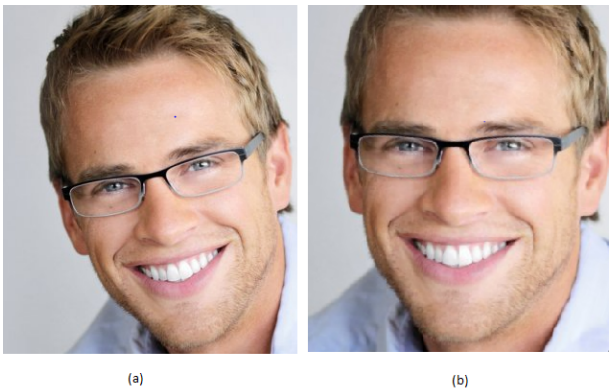

FIGURE 2.4 – Exemple d'alignemet d'un visage tel que (a) est le visage avant l'alignement et (b) le visage aprés l'alignement.

Les méthodes classiques d'alignement des visages dont le modèle actif de forme et le Active Apparence Model (AAM) , simulent le processus de génération d'image et réalise l'alignement du visage en minimisant la différence entre l'apparence du modèle et l'image d'entrée. Ces méthodes permettent d'obtenir des résultats de reconstruction précis, mais nécessitent un grand nombre de modèles de visage des correspondances détaill ées et précises, ainsi qu'un coût de calcul élevé pour l'ajustement des paramètres [60].

Après l'alignement de visage l'extraction des caractéristiques se fait de différentes méthodes.

2.4.2 Méthodes d'extraction des caractéristiques

Parmis les méthodes d'extraction de caractéristique :

2.4.2.1 SIFT : Scale-Invariant Feature Transform

Scale-Invariant Feature Transform que l'on peut traduire par transformation de caractéristiques visuelles invariante à l'échelle , consiste à calculer ce que l'on appelle les descripteurs SIFT des images à étudier. C'est une approche pour la détection et l'extraction de descripteurs de caractéristiques locales qui sont raisonnablement invariantes aux changements dans l'éclairage, le bruit d'image, la rotation, la mise à l'échelle et de petits changements de points de vue, autrement dit invariant aux informations qui caractérisent le contenu visuel de cette image de la façon la

plus indépendante possible. Aussi, deux photographies d'un même objet auront toutes les chances d'avoir des descripteurs SIFT similaires, Il a été développé en 1999 par le chercheur David Lowe [80] selon lui la base de l'algorithme SIFT se compose de cinq étapes :

- détection d'extrema dans l'espace des échelles,

- localisation précise de points clés,

- affectation d'orientation,

- calcul de descripteurs de points-clés,

- correspondance.

La figure 2.5 (a) SIFT correspondante entre les visages d'apprentissage et de test appartenant la même identité et (b) des identités différentes

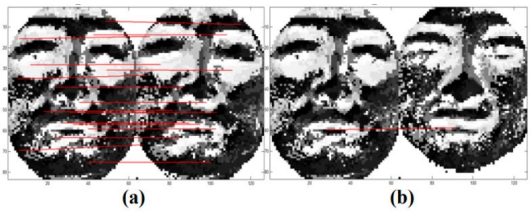

FIGURE 2.5 – Exemple détection Points-clés avec le descripteur SIFT [93]

2.4.2.2 Filtres de Gabor

Filtre de Gabor est un filtre linéaire dont la réponse impulsionnelle est une sinusoïde modulée par une fonction gaussienne (également appelée ondelette de Gabor). Il porte le nom du physicien anglais d'origine hongroise Dennis Gabor. Les filtres de Gabor sont des filtres passe-bande utilisés en traitement d'images pour l'extraction de caractéristiques, l'analyse de texture et l'estimation de disparité stéréo, etc [61].

Les filtres de Gabor sont connus comme un moyen d'analyse espace-fréquence très robuste ; tel qu'un ensemble de filtres de Gabor qui est utilisé avec 5 fréquences spatiales et 8 orientations distinctes, ce qui donne 40 filtres de Gabor différents. Cette spécificité a fait des filtres de Gabor un moyen puissant d'analyse de textures et de classification. Les filtres de Gabor analysent la texture d'un objet suivant différentes résolutions et différents angles dans le domaine spatial. La représentation de Gabor d'une image de visage est obtenue par la convolution de l'image avec la famille des filtres de Gabor. Cette convolution est définie par IG (r,o)= I*G(r,o)

I*G(r,o) est le résultat de la convolution de l'image par le filtre de Gabor à une certaine résolution r et à une orientation o.

La figure 2.6 représente le résultat de la convolution d'une image avec une famille de filtres de Gabor de 4 orientations (horizontales) et 4 résolutions (verticales). Les réponses en amplitude (a) et en phase (b) sont représentées.

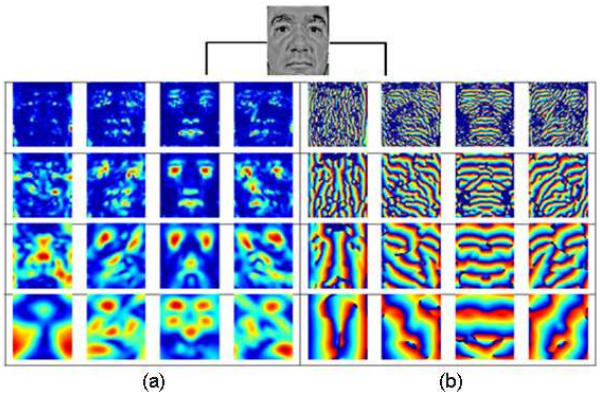

FIGURE 2.6 – Résultat de la convolution d'une image avec une famille de filtres de Gabor [61].

2.4.2.3 HOG : Histogram of Oriented Gradients

Afin d'explorer de nouvelles méthodes d'extractions de caractéristiques dans l'objectif d'améliorer les performances de reconnaissance de visages et détection de visage , un descripteur HOG est une technique accomplit une représentation de visages à l'aide d'une analyse en histogrammes des gradients présents dans l'image. La technique calcule des histogrammes locaux de l'orientation du gradient sur une grille dense, c'est-à-dire sur des zones régulièrement réparties sur l'image. Elle possède des points communs avec les SIFT. Plus particulièrement, la région d'intérêt où le visage est détecté est subdivisée en blocs de tailles égales et ces derniers sont également subdivisés à leur tour en cellules. Pour chacune des cellules, une analyse des gradients des pixels est accomplie afin de former un histogramme de gradient à neuf bandes. Plusieurs techniques de recombinaison des histogrammes en vecteurs existent dans la littérature [27].

2.4.2.4 PCA : Analyse en Composantes Principales

Une méthode très populaire, basée sur la technique d'analyse des composants, est la méthode eigenface introduite en 1991 par Turk et Pentland [111]. Son principe est le suivant : étant donné un ensemble d'images de visages exemples, il s'agit tout d'abord de trouver les composantes principales de ces visages. Ceci revient à déterminer les vecteurs propres de la matrice de covariance formée par l'ensemble des images exemples. Chaque visage exemple peut alors être décrit par une combinaison linéaire de ces vecteurs propres. Pour construire la matrice de covariance, chaque image de visage est transformée en vecteur. Chaque élément du vecteur correspond à l'intensité lumineuse d'un pixel, PCA construit un sous-espace pour représenter de manière optimale seulement l'objet [111].

2.4.2.5 DCT : Discrete Cosine Transform

la méthode DCT « Transformée de cosinus discrète » est beaucoup plus rapide que la PCA concernant l'extraction de vecteur caractéristique.

La méthode est simple chaque image de visage est représentée par un vecteur composé des premiers coefficients de la transformée DCT. Et lorsqu'un visage est présenté sa transformée est calculée et un certain nombre de coefficients est retenu pour comparaison avec ceux de la base de données. Et pour chacune des images de la base de donnée sa transforme est calculé en cosinus discrète de l'image normalisée et les premiers coefficients de la DCT son extrait afin de former un vecteur unifié, puis des représentations sont sauvegardé. Donc le processus d'apprentissage est réalisé sur chaque image indépendamment contrairement aux techniques ACP. On remarque que l'un des avantages de cette méthode repose sur sa grande flexibilité en cas d'ajouts d'images ou de personnes [49].

2.4.2.6 ICA : Independant Component Analysis

L'Analyse en composantes indépendantes (ou ICA pour Independant Component Analysis) est une généralisation de PCA. En outre, ICA permet une meilleure caractérisation des données dans un espace à n dimensions et les vecteurs de base trouvés par ICA ne sont pas nécessairement orthogonaux afin qu'ils réduisent également l'erreur de reconstruction. Deux architectures différentes pour l'ICA sont fournies : une première architecture (ICA I) qui construit une base d'images statistiquement indépendantes et une deuxième architecture (ICA II) qui fournit une représentation en code factoriel des données. Pour plus de détails sur la théorie et les applications possibles de l'analyse en composantes indépendantes, on peut se référer à [56].

2.4.2.7 LPQ : Local Phase Quantisation

La quantification de la phase locale ou le descripteur LPQ a été désigné pour la première fois par Ojansivu et Heikkilä [91] pour l'utiliser dans la classification de textures pour les images floues. Le descripteur LPQ est construit de façon à ne retenir dans une image que l'information locale invariante à un certain type de flou [91].

La quantification de phase locale (LPQ) est basée sur le calcul de la transformée de Fourier à court terme (STFT : short term Fourier transform) sur la fenêtre d'image locale. A chaque pixel, le coefficient local de Fourier est calculé pour quatre points de fréquence. Ensuite, les signes de la partie réelle et imaginaire de chaque coefficient sont quantifiés à l'aide d'un quantificateur scalaire binaire, afin de calculer l'information de phase. Le coefficient binaire de huit bits qui en résulte est ensuite représenté sous forme d'entiers par codage binaire. Au final un vecteur de caractéristiques à 256 dimensions est obtenu [91].

La fonction correspondante est définie comme suit :

$$F(u,x) = \sum_{y} f(x-y) e^{2\pi j y u^T} = w_u f_x \tag{2.1}$$

Où u est la fréquence, et x est une fonction de fenêtre définissant le voisinage N.

2.4.2.8 BSIF : Binarized Statistical Image Features

Kannala et Rahtuen 2012 [62] ont proposé un nouveau descripteur local appelé BSIF : Caractéristiques d'image statistiques binarisées.

Cett méthode construit des descripteurs d'images locaux qui encodent efficacement l'information de texture et qui conviennent à la représentation des régions de l'image par histogramme. Cette dernière calcule un binaire pour chaque pixel en projetant linéairement des patchs d'images sur un sous-espace, dont les vecteurs de base sont appris à partir d'images naturelles via une analyse en indépendants, et en binarisant les coordonnées dans cette base par seuillage. La longueur de la chaîne de code binaire est déterminée par le nombre de vecteurs de base. Les régions de l'image peuvent être représentées de manière pratique par des histogrammes des codes binaires des pixels.

En bref l'idée derrière BSIF est d'apprendre automatiquement un ensemble fixe de filtres à partir d'un petit ensemble d'images naturelles. Elle est définit de telle sorte etant donné une patch d'image X de taille l × l pixels et un filtre linéaire Wi de la même taille, la réponse du filtre s_i

$$s_i = \sum_{u,v} W_i(u,v)X(u,v) = w_i^T x \tag{2.2}$$

où la notation vectorielle est introduite dans la dernière étape, c'est-à-dire que les vecteurs w et x contiennent les pixels de Wi et de X. La caractéristique binarisée bi est obtenue en fixant bi=1 si ¿ 0 et bi = 0 sinon.

Dans la partie suivante nous avons décrit l'opérateur LBP qu'on a utilisé dans la partie extraction des caractéristiques dans notre projet

2.4.2.9 LBP : Local Binary Patterns

L'opérateur LBP où "Local Binary Patterns" a été proposé initialement par Ojala et al [51] en 1995. Dans le but de caractériser la texture d'une image. Il a été largement appliqué avec divers algorithmes de systèmes de reconnaissance de visage comme une méthode d'extraction de caractéristiques locales [16]. En raison de son pouvoir discriminant et la simplicité de calcul, le descripteur est devenu une approche très populaire dans diverses applications de vision par ordinateur. Par exemple pour la détection et l'analyse des visages, la biométrie, l'analyse d'images médicales, le mouvement et l'analyse de l'activité et la récupération des bases de données d'images ou vidéo. Plusieurs chercheurs en reconnaissance de visage sont devenus intéressés par LBP. Ahonen et al en 2006 [8] ont montré son grand succès dans la reconnaissance des visages. Selon [18], la LBP n'est pas considérée seulement comme un opérateur de texture simple, mais constitue le fondement d'une nouvelle direction de recherche importante pour les descripteurs binaires locaux sur l'image et la vidéo.

Le concept du LBP est simple, il propose d'assigner un code binaire à un pixel en fonction

de son voisinage. Ce code décrivant la texture locale d'une région est calculé par seuillage d'un voisinage avec le niveau de gris du pixel central.

Le LBP de base est défini par :

$$LBP(x_c, y_c) = \sum_{p=0}^{p-1} S(g_p - g_c) * 2^p \tag{2.3}$$

Où :

g_c est le niveau de gris du pixel central de coordonnées (x_c , y_c)

g_p(p= 0,1,...,7) est le niveau de gris de chaque pixel voisin.

Avec : S(x) une fonction définie comme suit :

$$S(x) = \begin{cases} 1 & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}$$

Afin de générer un motif binaire. La méthodologie de LBP a été développée récemment avec un grand nombre de variations pour l'amélioration des performances dans différentes applications. Ces variations portent sur différents aspects de l'opérateur LBP original :

1) l'amélioration de sa capacité discriminatoire.

2) l'amélioration de sa robustesse .

3) la sélection de son voisinages.

les deux avantages de la méthode d'analyse d'images de texture basée sur le LBP sont la rapidité et faible occupation de la mémoire [102].

2.5 Apprentissage profond : Deep Learning

Le Deep Learning (DL), ou apprentissage profond, est la principale technologie du ML et d'IA. Le DL repose sur la technologie des réseaux de neurones artificiels. Ils ont constitué de plusieurs couches artificielles interconnectés. Plus le nombre de couches cachées est élevé plus le réseau est profond d'où vient le nom "Deep".

Le Deep Learning est un nouveau domaine de recherche du ML, qui a été introduit dans le but de concevoir une machine procède de la même façon de celle d'un être humain. Ils peuvent apprendre plusieurs niveaux de représentation dans le but de modéliser des relations complexes entre les données. Il s'agit d'algorithmes inspirés par la structure et le fonctionnement du cerveau.

L'apprentissage profond repose sur l'utilisation d'une nouvelle technologie des réseaux de

neurones artificiels pour exploiter et gérer des quantités massives de données tout en ajoutant des couches au réseau. Ces multiples couches permettent au réseau de neurones DL d'extraire des caractéristiques complexes à partir des données brutes à l'aide de multiples transformations linéaires et non linéaires à travers les couches multiples [28].

L'avantage primordial du Deep Learning réside dans ses performances énormes lorsqu'il s'agit de grande quantité de donnée, au contraire des algorithmes ML classiques qui sont très limité avec les grandes quantités des données. En effet, DL s'adapte bien avec une très large quantité de données, et il peut aller jusqu' à dépasser la performance d'un être humain dans plusieurs domaines tels que le traitement d'image.

Avantages et inconvénients

• **Avantages :**

o Caractéristiques automatiquement déduites et optimisées pour le résultat souhaité.

o Certains problèmes se prêtent mieux à l'apprentissage en profondeur que d'autres applications.

o Modèles plus simples peuvent être suffisants pour certains domaines de problèmes.

o Les modèles sont très difficiles à expliquer par rapport aux autres algorithmes ML.

• **Inconvénients :**

o Technique difficile et ambigüe avec un calcul intensif.

o Certains problèmes se prêtent mieux à l'apprentissage en profondeur que d'autres applications.

o Modèles plus simples peuvent être suffisants pour certains domaines de problèmes.

o Les modèles sont très difficiles à expliquer par rapport aux autres algorithmes ML.

Dans ce qui suit nous présenterons les réseaux de neuronnes existant dans le domaine.

2.5.1 RNN : Recurrent Neural Networks

Un réseau de neurones récurrents (RNN) est un réseau de neurones artificiels présentant des connexions récurrentes a fin de ce souvenir des évènements passés. Ce reseau est constitué de neurones interconnectées interagissant non-linéairement et pour lequel il existe au moins un cycle dans la structure. En d'autre termes c'est un réseau dont les neurones s'envoient des signaux de rétroaction les uns aux autres. Les réseaux neuronaux récurrents conviennent aux données d'entrée de taille variable. Ils sont particulièrement utiles pour l'analyse des séries temporelles. Ils sont utilisés pour la reconnaissance automatique de la parole ou de l'écriture, aussi plus généralement pour la reconnaissance des formes. Lorsqu'ils sont déroulés, ils sont comparables aux réseaux neuronaux classiques, mais avec des restrictions sur l'égalité entre les poids du réseau. Mais ces réseaux de

neurones sont confrontés au problème de la disparition des gradients lorsqu'ils apprennent à se souvenir des événements passés [82].

2.5.2 DNN : Deep Neural Networks

Les réseaux neuronaux profonds (DNN) apprennent dans des couches hiérarchiques de représentation à partir d'entrées afin d'effectuer les tâches de classification nécessaires. Récemment, ces architectures d'apprentissage profond ont démontré des résultats impressionnants et parfois des résultats compétitifs pour l'homme dans de nombreuses applications. Un DNN se compose de couches convolutionnelles, de couches de max-pooling et de couches entièrement connectées. Tous les paramètres réglables sont conjointement optimisés conjointement par la minimisation de l'erreur de mauvais classement sur l'ensemble d'apprentissage.

Un DNN est un réseau neuronal artificiel de type action anticipative qui possède plus d'une couche d'unités cachées entre les entrées et les sorties. Ces couches cachées multiples peuvent s'avérer utiles pour résoudre des problèmes de classification avec des données complexes et chaque couche peut apprendre des caractéristiques à un niveau d'abstraction différent [108].

2.5.3 CNN : Convolutional Neural Network

Les réseaux neuronaux convolutifs (CNN), également appelés ConvNets, sont un type de réseaux neuronaux à anticipation bien adaptés aux tâches liées au domaine de la vision par ordinateur, notamment à la reconnaissance d'objets. L'architecture de deep learning la plus populaire est les réseaux de neurones convolutifs (CNN) qui sont une catégorie de réseaux de neurones qui se sont avérés très efficaces dans des domaines tels que la reconnaissance et la classification d'images. Les CNN ont réussi à identifier les visages, les objets, panneaux de circulation, etc [74], c'est la raison pour laquelle nous avons décidé de travailler avec cette structure pour notre projet. Les réseaux de neurones convolutifs ont une méthodologie similaire à celle des méthodes traditionnelles d'apprentissage supervisé : ils reçoivent des images en entrée, détectent les features automatiquement de chacune d'entre elles, puis entraînent un classifieur dessus. Donc les CNN réalisent eux-mêmes tout le boulot fastidieux d'extraction et description de features. Le but d'un CNN est d'apprendre des fonctionnalités d'ordre supérieur dans les données via des convolutions. Ils sont bien adaptés à la reconnaissance d'objets et de classification d'images.

Une architecture CNN typique comprend généralement des couches alternées de convolution et de mise en commun, suivies d'une ou plusieurs couches entièrement connectées à la fin. Dans certains cas, une couche entièrement connectée est remplacée par une couche de mise en commun de la moyenne globale. En plus des différentes fonctions de mappage, différentes unités de régulation telles que la normalisation et le dropout des lots sont également incorporées pour optimiser les performances du CNN [74]. La disposition des composants du CNN joue un rôle fondamental dans

la conception de nouvelles architectures et l'obtention de meilleures performances. Cette section aborde brièvement le rôle de ces composants dans une architecture CNN.

couche Convolution : La couche de convolution est parfois appelée couche d'extraction de caractéristiques, car les caractéristiques de l'image sont extraites dans cette couche. Tout d'abord, une partie de l'image est connectée à la couche convolution pour effectuer une opération de convolution et calculer le produit scalaire entre le champ récepteur (c'est une région locale de l'image d'entrée ayant la même taille que celle du filtre. Le résultat de l'opération est un entier unique du volume de sortie. Ensuite, nous faisons glisser le filtre sur le champ récepteur suivant de la même image d'entrée par une foulée et refaisons la même opération. Cette opération est répétée par le même processus encore et encore jusqu'à ce que toute l'image soit parcourue [67].

Couche de mise en commun (Pooling) : Ce type de couche est souvent placé entre deux couches de convolution : elle reçoit en entrée plusieurs feature maps, et applique à chacune d'entre elles l'opération de pooling. L'opération de pooling consiste à réduire la taille des images, tout en préservant leurs caractéristiques importantes. Pour cela, on découpe l'image en cellules régulière, puis on garde au sein de chaque cellule la valeur maximale.

La méthode utilisée consiste à imaginer une fenêtre de 2 ou 3 pixels qui glisse au-dessus d'une image, comme pour la convolution. Mais, cette fois-ci, nous faisons des pas de 2 pour une fenêtre de taille 2, et des pas de 3 pour 3 pixels [67]. La taille de la fenêtre est appelée « kernel size » et les pas s'appellent «strides ». Pour chaque étape, nous prenons la valeur la plus haute parmi celles présentes dans la fenêtre et cette valeur constitue un nouveau pixel dans une nouvelle image. Ceci s'appelle Max Pooling.

La couche de pooling permet de réduire le nombre de paramètres et de calculs dans le réseau. On améliore ainsi l'efficacité du réseau et on évite le sur-apprentissage. Ainsi, la couche de pooling rend le réseau moins sensible à la position des features : le fait qu'une feature se situe un peu plus en haut ou en bas, ou même qu'elle ait une orientation légèrement différente ne devrait pas provoquer un changement radical dans la classification de l'image [67].

Couche entièrement connecté (Fully Connected) : constitue toujours la dernière couche d'un réseau de neurones. Ce type de couche reçoit un vecteur en entrée et produit un nouveau vecteur en sortie. Pour cela, elle applique une combinaison linéaire puis éventuellement une fonction d'activation aux valeurs reçues en entrée.La dernière couche fully-connected permet de classifier l'image en entrée du réseau : elle renvoie un vecteur de taille N , où N est le nombre de classes dans notre problème de classification d'images. Chaque élément du vecteur indique la probabilité pour l'image en entrée d'appartenir à une classe [67].

Pour calculer les probabilités, la couche fully-connected multiplie donc chaque élément en entrée par un poids, fait la somme, puis applique une fonction d'activation (logistique si N=2 , softmax si $N \geq 2$) : Ce traitement revient à multiplier le vecteur en entrée par la matrice

contenant les poids. Le fait que chaque valeur en entrée soit connectée avec toutes les valeurs en sortie explique le terme fully connected [67].

2.5.4 Quelques réseaux convolutifs célèbres

Il existe plusieurs réseaux convolutifs célèbre tels que :

2.5.4.1 LeNet

Les premières applications réussies des réseaux convolutifs ont été développées par Yann LeCun dans les années 1990. Parmi ceux-ci, le plus connu est l'architecture LeNet utilisée pour lire les codes postaux, les chiffres, etc [73].

2.5.4.2 AlexNet

Le premier travail qui a popularisé les réseaux convolutifs dans la vision par ordinateur était AlexNet, développé par Alex Krizhevsky, Ilya Sutskever et Geoff Hinton en 2012 [66]. Ce CNN été soumis au défi de la base ImageNet et a nettement surpassé ses concurrents. Le réseau avait une architecture très similaire à LeNet, mais était plus profond, plus grand et comportait des couches convolutives empilées les unes sur les autres (auparavant, il était commun de ne disposer que d'une seule couche convolutifs toujours immédiatement suivie d'une couche de pooling) [66].

2.5.4.3 Overfeat

Overfeat : est un classificateur d'image basé sur un réseau convolutionnel et un extracteur de fonctionnalités. Il a été formé sur le jeu de données Image Net et a participé au concours Image Net 2013. [103].

2.5.4.4 Inception V3

Ce type d'architecture, introduit en 2016 par Szegedy et al [109] utilise des blocs avec des filtres de différentes tailles qui sont ensuite concaténés pour extraire des caractéristiques à différentes échelles [109].

2.5.4.5 Xception

Cette architecture a été proposée par François Chollet en 2017 [23] et la seule chose qu'il apporte à Inception est qu'il effectue de manière optimale les circonvolutions pour qu'elles prennent moins de temps. Ceci est réalisé en séparant les convolutions 2D en 2 convolutions 1D [23].

2.5.4.6 ResNet : Residual Neural Network

ResNet est une architecture de réseau profond avec 152 couches [53]. Ceci a été développé par Microsoft en 2016. Il a remporté l'ILSVRC 2016 avec un taux d'erreur de 3,6 % [53], ce qui est considéré comme meilleur que la précision au niveau humain. Les couches résiduelles dans ResNet calculent les changements dans l'entrée. Ceci est ensuite ajouté à l'entrée pour produire la sortie. ResNet-50 est l'un des premiers à adopter la normalisation des batchs. Le ResNet est constitué de deux blocs CONV et Identité.

2.5.4.7 VGG-16

Ren Simonyan et Andrew Zisserman de Oxford Vision Geometry Group (VGG) [104] ont proposé le VGG-16 qui a 13 couches de convolution et 3 couches entièrement connectée. VGGNet utilise des filtres de taille 3x3 comparant de 11x11 de AlexNet et 7x7 de ZFNet. Les auteurs donnent l'intuition derrière cela qu'avoir deux filtres de taille 3x3 consécutifs donne un champ récepteur efficace de 5x5, et trois séries de filtres de taille 3x3 donnent un champ. Récepteur de filtres 7x7, mais en utilisant cela, nous pouvons utiliser un nombre beaucoup moins élevé d'hyper-paramètres. Ils ont aussi proposé une variante plus profonde VGG-19.

2.5.4.8 GoogleNet

Google Net : a 22 couches de profondeur, et presque 12 fois moins de paramètres donc plus rapide et moins que Alex Net et beaucoup plus précis. Il réduit le nombre de paramètres de 60 millions (Alex Net) à 4 millions. Leur idée était de créer un modèle qui pourrait également être utilisé sur un Smartphone (conserver un budget de calcul d'environ 1,5 milliard de multiplications par prévision) [107].

Couche de lancement : l'idée de la couche de lancement est de couvrir une plus grande surface, mais aussi de conserver une résolution fine pour les petites informations sur les images. L'idée est donc de convoler en parallèle différentes tailles, des tailles plus précis (1*1) à un plus grand (5*5). Le moyen le plus simple d'améliorer les performances en matière d'apprentissage en profondeur consiste à utiliser plus de couches et plus de données.

Google Net utilise 9 modules de démarrage. Le problème est que plus de paramètres signifient également que votre modèle est plus enclin à sur- adapter. Ainsi, pour éviter une explosion de paramètres sur les couches initiales [107].

Dans tout système d'intelligence artificielle, chaque décision prise ou chaque action effectuée est le résultat d'un long processus de traitement de données. Cependant, cela ne serait pas possible sans une parfaite compréhension des données. En ce qui concerne la vision par ordinateur, cette compréhension commence avant tout par l'étiquetage des composants de l'image, autrement dit, la classification d'image.

2.6 Classification

la classification automatique des images consiste à attribuer automatiquement une classe à une image à l'aide d'un système de classification. On retrouve ainsi la classification d'objets, de scènes, de textures, la reconnaissance de visages, d'empreintes digitale et de caractères. La classification des images consiste à répartir systématiquement des images selon des classes établies au préalable, classer une image lui fait correspondre une classe, marquant ainsi sa parenté avec d'autres images. En général reconnaître une image est une taché aisée pour un humain au fil de son existence, il a acquis des connaissances qui lui permettent de s'adapter aux variations qui résultent de conditions différents d'acquisition. Il lui est par exemple relativement simple de reconnaître un objet dans plusieurs orientations partiellement caché par un autre de près ou de loin et selon diverses illuminations [29].

L'objectif de la classification d'images est d'élaborer un système capable d'affecter une classe automatiquement à une image. Ainsi, ce système permet d'effectuer une tâche d'expertise qui peut s'avérer coûteuse à acquérir pour un être humain en raison notamment de contraintes physiques comme la concentration, la fatigue ou le temps nécessité par un volume important de données images. Les applications de la classification automatique d'images sont nombreuses et vont de l'analyse de documents à la médecine en passant par le domaine militaire. Ainsi on retrouve des applications dans le domaine médical comme la reconnaissance de cellules et de tumeurs, la reconnaissance d'écriture manuscrite pour les chèques les codes postaux. Dans le domaine urbain comme la reconnaissance de panneaux de signalisation la reconnaissance de piétons la détection de véhicules la reconnaissance de bâtiments pour aider à la localisation. Dans le domaine de la biométrie comme la reconnaissance de visage, d'empreintes, d'iris. Le point commun à toutes ces applications est qu'elles nécessitent la mise en place d'une chaîne de traitement à partir des images disponibles composée de plusieurs étapes afin de fournir en sortie une décision [29].

2.6.1 CaffeNet model

Les modèles Caffe sont des systèmes d'apprentissage automatique de bout en bout. Un réseau typique commence par une couche de données qui charge à partir du disque et se termine par une couche de perte qui calcule l'objectif pour une tâche telle que la classification [59] .

Une couche Caffe est une couche de réseau neuronal : elle prend en entrée un ou plusieurs blobs, et produit en sortie un ou plusieurs les couches ont deux responsabilités essentielles pour le fonctionnement du réseau dans son ensemble : une passe avant qui prend les entrées et produit les sorties, et une passe arrière qui prend le gradient par rapport à la sortie, et calcule les gradients par rapport aux paramètres et aux entrées, qui sont à leur tour rétro-propagés vers les couches précédentes [59].

2.7 Regression

L'analyse de règression rèpond à des questions sur la dépendance d'une variable de réponse à un ou plusieurs prédicteurs, y compris la prédection des valeurs futures d'une réponse. L'informatique à haut débit, bon marché et largement disponible, A changé les régles d'extraction de ces questions. Les concurrents modernes comprennent la régression non paramétrique, les réseaux neuronaux, les machines à vecteur de support (SVM) et les méthodes basées sur les arbres, entre autres un nouveau domaine de l'informatique, appelé apprentissage automatique, ajoute de la diversité, et de la confusion, au mélange. Grâce à la disponibilité des logiciels, l'utilisation d'un réseau neuronal ou de l'une de ces autres méthodes semble être tout aussi facile que d'utiliser la régression linéaire[115]. Ainsi, on considère que les problèmes de prédiction d'une variable quantitative sont des problèmes de régression. La régression linéaire est un modèl qui vise à établir une relation linéaire entre une variable et une ou plusieurs variables dites explicatives.

La construction de modèles de réseaux neuronaux profonds étant de plus en plus facile en utilisant des cadres qui fournissent des modules prêts à l'emploi, par exemple Keras.

2.7.1 keras

Keras est une API de réseaux de neurones de haut niveau, écrite en Python et capable de fonctionner sur TensorFlow ou Theano. Il a été développé en mettant l'accent sur l'expérimentation rapide. Être capable d'aller de l'idée à un résultat avec le moins de délai possible est la clé pour faire de bonnes recherches. Il a été développé dans le cadre de l'effort de recherche du projet ONEIROS (Open-ended Neuro-Electronic Intelligent Robot Operating System), et son principal auteur et mainteneur est François Chollet, un ingénieur Google. En 2017, l'équipe TensorFlow de Google a décidé de soutenir Keras dans la bibliothèque principale de TensorFlow. Chollet a expliqué que Keras a été conçue comme une interface plutôt que comme un cadre d'apprentissage

l'architecture keras est diviser en 3 categorie : modèle, couches et modules de base.

Il existe deux façons de créer des modèles Keras :

Le modèle séquentiel , qui est très simple (une simple liste de couches), mais se limite à des piles de couches à entrée unique et à sortie unique (comme son nom l'indique).

L'API fonctionnelle , qui est une API facile à utiliser et complète qui prend en charge des architectures de modèles arbitraires. Pour la plupart des gens et la plupart des cas d'utilisation, c'est le modèle Keras "force de l'industrie".

Keras fournit de nombreuses couches de pré-construction afin que tout réseau de neurones complexe puisse être facilement créé. Certaines des couches importantes de Keras sont spécifiées : couche entrée (Input), couche convolution : pour recevoir et traiter l'entrée (image), couche mise en commun (pooling) : utiliser pour reduire la taille des entrées en supprimant les informations

inutile, couche entièrement connectée (fully connected) : est utilisé pour la classification de l'image , couche de sortie (outputs).

keras founit un grand nombre de fonction intégrée liée aux réseaux de neurones tels que les fonctions d'application, fonction de perte, optimiseur, metriques, initialiseurs, visualisation, regulariseurs ,etc.

FIGURE 2.7 – Exemple d'architecture de keras

2.8 Conclusion

Dans ce chapitre, nous avons définit la détection de visage et quelques méthodes existantes dans ce domaine, ainsi que les méthodes que nous avons utilisées dans notre système. Ensuite nous avons survolé certaines méthodes d'extraction de caractéristques du visage, De plus nous avons vu la structure des CNN's pour finir avec la classification et la regression.

Chapitre 3

Chapitre 3.Conception des méthodes proposées

3.1 Introduction

Le système d'estimation de l'âge est considéré comme un système de reconnaissance des formes qui comprend plusieurs étapes. Dans ce chapitre, nous allons citer les différentes étapes appliquées pour l'estimation de l'âge des visages à partir des images faciales comme suit : La détection du visage est utilisée comme entrée ; le prétraitement du visage détecté avec la normalisation, l'alignement et le recadrage de ce visage ; l'extraction des caractéristiques du visage avec le discripteur LBP. Ces étapes sont utilisées pour avoir à la sortie, un visage extrait et prêt pour l'estimation de l'âge en utilisant la regression avec keras. Notre travail est basé sur l'étude des performances du système d'estimation d'âge d'un visage par l'application des différents détecteur de visages ainsi qu'un descripteur d'extraction des caractéristiques.

La figure 3.1 montre un schéma récapitulatif du travail.

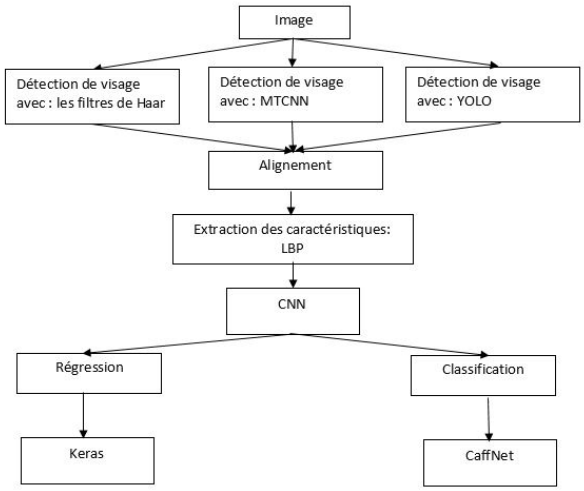

FIGURE 3.1 – Schéma récapitulatif de notre système

3.2 Méthodes de détection du visage utilisées

La première méthode utilisée pour la détection du visage est la méthode de Viola et Jones.

3.2.1 La méthode de Viola et Jones : les filtres de Haar

Elle a la particularité d'utiliser des caractéristiques très simples mais très nombreuses. Une première innovation de la méthode est l'introduction des images intégrales, qui permettent le calcul rapide de ces caractéristiques. Une deuxième innovation importante est la sélection de ces caractéristiques par boosting, en interprétant les caractéristiques comme des classifieurs. Enfin, la méthode propose une architecture pour combiner les classifieurs boostés en un processus en cascade, ce qui apporte un net gain en temps de détection.

3.2.1.1 Caractéristiques :

Une caractéristique est une représentation synthétique et informative, calculée à partir des valeurs des pixels. Les caractéristiques utilisées ici sont les caractéristiques pseudo Haar. Elle sont calculées par la différence des sommes de pixels de deux ou plusieurs zones rectangulaires adjacente [113].

Prenons un exemple : Voici deux zones rectangulaires adjacentes, la première en blanc,la deuxième en noire. Les caractéristiques seraient calculées en soustrayant la somme des pixels noirs à la somme des pixels blancs. Les caractéristiques sont calculées à toutes les positions et à toutes les échelles dans une fenêtre de détection de petite taille, typiquement de 24x24 pixels ou de 20x15 pixels. Un très grand nombre de caractéristiques par fenêtre est ainsi généré [113]. La figure 3.2 montre un exemple de caractéristiques pseudo-Haar.

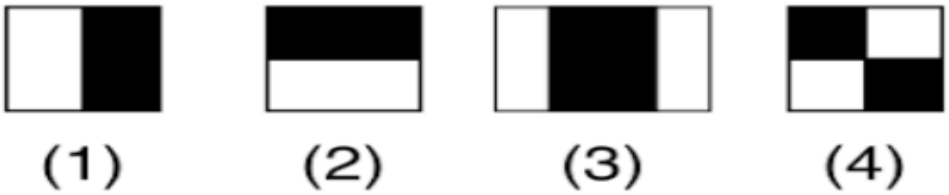

FIGURE 3.2 – Exemple de caractéristiques pseudo-Haar

Viola et Jones donnent l'exemple d'une fenêtre de taille 24 x 24 qui génère environ 160 000 caractéristiques.L'image précédente présente des caractéristiques pseudo-Haar à seulement deux caractéristiques mais il en existe d'autres, allant de 4 à 14, et avec différentes orientations [113].

Malheureusement, le calcul de ces caractéristiques de manière classique coûte cher en terme de ressources processeur, c'est là qu'interviennent les images intégrales .

3.2.1.2 Image intégrale :

L'image intégrale est une matrice de la même dimension que l'image a traiter, ou l'amplitude de chaque élément de coordonnée (x,y) dans l'image intégrale représente la somme des amplitudes de tous les pixels au-dessus et à gauche de l'image y compris le pixel de coordonnée (x,y) d'où le nom image intégrale. Cette image sera donc une représentation intermédiaire dans le traitement pour diminuer le temps de calcul et cela pour calculer d'une manière efficace et rapide les caractéristiques pseudo-Haar(figure 3.3).

On peut également définir l'image intégrale ii par :

$$in(x,y) = \sum_{x' \leq x, y' \geq y} i(x',y') \tag{3.1}$$

Où In(x, y) est l'image intégrale et i(x, y) l'image originale. Comme nous utilisons cette nouvelle représentation pour diminuer le temps de calcul, on peut expliquer ses avantages. D'abord, elle peut être calculée par un moyen efficace en utilisant la paire des récurrences suivantes :

$$S(x,y) = S(x,y-1) + i(x,y) \tag{3.2}$$

$$In(x,y) = In(x-1,y) + s(x,y) \tag{3.3}$$

D'où s(x, y), s est la somme cumulative

$$\forall x, S(x,-1) = 0, et \forall y, In(-1,y) = 0 \tag{3.4}$$

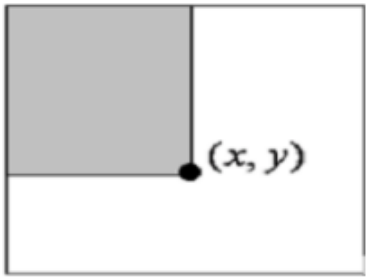

FIGURE 3.3 – La valeur de l'image intégrale au point (x,y)

L'utilisation de l'image intégrale réduit considérablement le nombre des opérations de somme. La somme de tous les éléments d'un rectangle peut être calculée en quatre opérations (voir la figure 3.3). Donc la différence entre les deux sommes rectangulaires peut être calculé en huit opérations, la caractéristique à deux rectangles définis ci-dessus, implique des sommes rectangulaires adjacentes qui peuvent être calculés en six opérations, huit dans le cas d'une caractéristique à trois rectangles,

et neuf pour la caractéristique à quatre rectangles. La figure 3.4 montre un exemple de Calcul de la somme du rectangle D avec l'image intégrale.

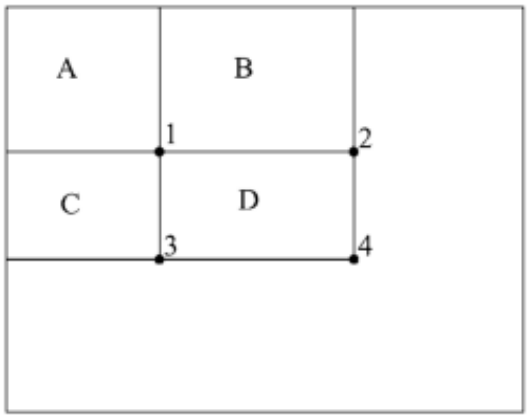

FIGURE 3.4 – Calcul de la somme du rectangle D avec l'image intégrale

La somme des pixels dans le rectangle D peut être calculée avec quatre opérations. La valeur de l'image intégrale au point 1 est la somme des pixels dans le rectangle A. La valeur à la position 2 est A + B, à la position 3 est A + C, et à la position 4 est A + B + C + D. La somme dans D peut être calculée comme 4 +1-(2 +3) Parce que la somme d'intensité dans une région rectangulaire ((x1 ; y1) ; (x2 ; y2)) est :

$$r = In(x, y-1) + In(x1, y2) - In(x2, y1) + In(x2, 2) \tag{3.5}$$

La figure 3.5 illustre un exemple de calcule de l'image intégrale.

Exemple :

1	1	1
1	1	1
1	1	1

Image initiale

1	2	3
2	4	6
3	6	9

image intégrale

FIGURE 3.5 – image intégrale

Par conséquent, la différence entre deux rectangles adjacents est obtenue à travers six références à l'image intégrale In (u, v).

Le deuxième élément clé de la méthode de Viola et Jones est l'utilisation d'une méthode de boosting afin de sélectionner les meilleures caractéristiques.

3.2.1.3 Sélection de caractéristiques par boosting :

Le boosting est un principe qui consiste à construire un classifieur « fort » à partir d'une combinaison pondérée de classifieur « faibles », c'est-à-dire donnant en moyenne une réponse meilleure qu'un tirage aléatoire. La valeur seuil de la caractéristique doit être trouvée pour l'apprentissage du classifieur faible qui va permettre de mieux séparer les exemples positifs des négatifs. Dans ce cas, le classifieur se réduit alors à un couple (caractéristique, seuil).

La méthode d'AdaBoost (Adaptive Boosting) est un algorithme de boosting qui permet de combiner plusieurs hypothèses pour créer une autre hypothèse plus performante (créer un ensemble dont la performance de chaque élément où ce dernier est booster). L'ensemble composé des hypothèses est défini comme suit :

$$f(x) = \sum_{t=1}^{T} \alpha_t h_t(x) \tag{3.6}$$

Avec a_t le poids qui est attribué à l'hypothèse h de l'ensemble. Les poids $_t$ et les hypothèses h_t doivent être entraînés pendant la procédure du Boosting. Les exemples d'apprentissage pondérés permettent une sélection itérative des coefficients a_t et des hypothèses h_t lors du Boosting. À chaque itération, les poids des exemples d'apprentissage sont recalculés de manière à attribuer une grande pondération aux exemples d'apprentissage mal classifiés et une faible pondération aux autres [113].

Adaboost : Pour la détection du visage, l'algorithme AdaBoost a été proposé avec une architecture de cascade sur laquelle on peut appliquer la détection du visage Pour les caractéristiques de l'algorithme, un classifieur faible est établi par une caractéristique rectangulaire, qui est l'équivalent d'une caractéristique faible, Cet algorithme a été proposé par Viola et Jones [113], son résultat est le plus abouti connu à ce jour sur la détection de visages, c'est l'algorithme AdaBoost.

Il est rapide quand il s'agit d'appliquer un nombre important de caractéristiques de rectangle a une petite région d'une fenêtre candidate, le classifieur fort est formé par quelques-unes qui sont choisies et combinées.

3.2.1.4 Cascade de classifieurs :

Une cascade de classifieurs est un arbre de décisions dégénéré où chaque étape est entrainée pour détecter un maximum d'objet intéressants tout en excluant une certaine fraction des objets non-intéressants [113]. La figure 3.6 montre une Illustration de l'architecture de la cascade : les fenêtres sont traitées séquentiellement par les classifieurs, et rejetées immédiatement si la réponse est négative (F).

Les fenêtres sont traitées séquentiellement par les classifieurs, qui prennent :

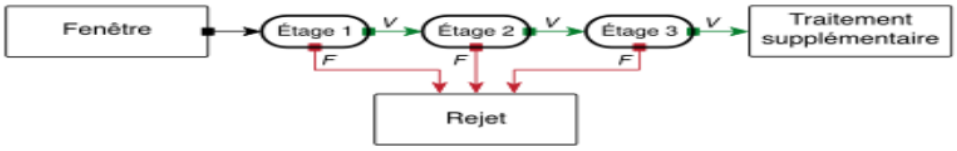

FIGURE 3.6 – Illustration de l'architecture de la cascade : les fenêtres sont traitées séquentiellement par les classifieurs, et rejetées immédiatement si la réponse est négative (F).

-une décision d'acceptation ; la fenêtre contient l'objet et l'exemple est alors passé au classifieur suivant.

-une décision de rejet ; la fenêtre ne contient pas l'objet et dans ce cas l'exemple est définitivement écarté [113].

La seconde méthode utilisée pour la détection du visage est YOLO.

3.2.2 YOLO : You Only Look Once

Comme mentionné précédemment, YOLO qui signifie "Vous ne regardez qu'une fois" est un algorithme de détection de tir unique, son nom donne une description parfaite de cet algorithme car il prédit les classes et les boîtes englobantes pour l'ensemble de l'image en une seule exécution de l'algorithme. Le mécanisme de l'algorithme utilise un seul réseau neuronal qui prend une photographie en entrée et tente de prédire les boîtes de délimitation et les étiquettes de classe pour chaque boîte de délimitation. Cette méthode permet d'atteindre des vitesses allant jusqu'à 45 images par seconde et jusqu'à 155 images par personne [97]

L'algorithme YOLO est divisé en 3 étapes :

La première étape est la division de l'image

3.2.2.1 Diviser l'image d'entrée en une grille S × S

la division de l'image en une grille de taille S x S donne N cellules au total. Cette cellule de la grille est responsable de la détection de l'objet. La figure 3.7 montre un exemple d'image divisé en cellule de taille 3x3 .

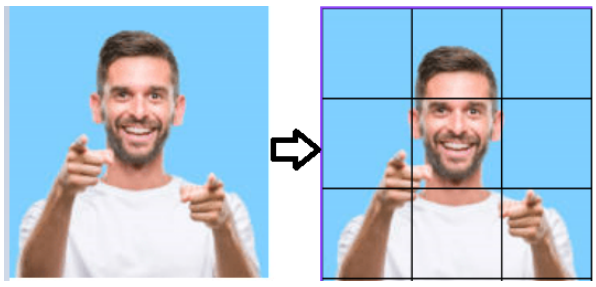

FIGURE 3.7 – Exemple d'image divisé en cellule de taille 3x3

La deuxième étape est la prédiction des boites englobantes

3.2.2.2 Chaque cellule prédit B boîtes englobantes :

Chaque cellule de la grille prédit B boîtes englobante (ou de délimitation) et des scores de confiance pour ces boîtes. Cette boite met en évidence le visage, chacune de ces boîtes englobantes B prédit 4 coordonnées pour chaque boîte englobante (bx,by,bw,bh). Les coordonnées (x, y) représentent le centre de la boîte ou le par rapport à la cellule correspondante. La valeur de bh est le rapport de la hauteur de la boîte englobante, et bw est le rapport de la largeur de la boîte englobante. Ces dernières sont prédites par rapport à l'image entière. Les prédictions de confiance représentent l'IOU (Intersection over Union) entre les boîtes prédites et toutes les boîtes de vérité [97].

Pour chaque cellule, le CNN prédit un vecteur Y qui contient :

• le score d'objectivité qui represente la probabilité que la cellule contienne un objet. Le score d'objectivité passe par une fonction sigmoïde pour être traité comme une probabilité avec une plage de valeurs comprise entre 0 et 1 [97].

• Les coordonnées bx, by, bh, bw prédit par les boites englobantes.

• La Probabilité que la cellule contienne un objet qui appartient à la classe visage étant donné que la cellule contient un objet.

La figure 3.8 montre un exemple de vecteur prédite par le CNN dans le cas d'un seul visage.

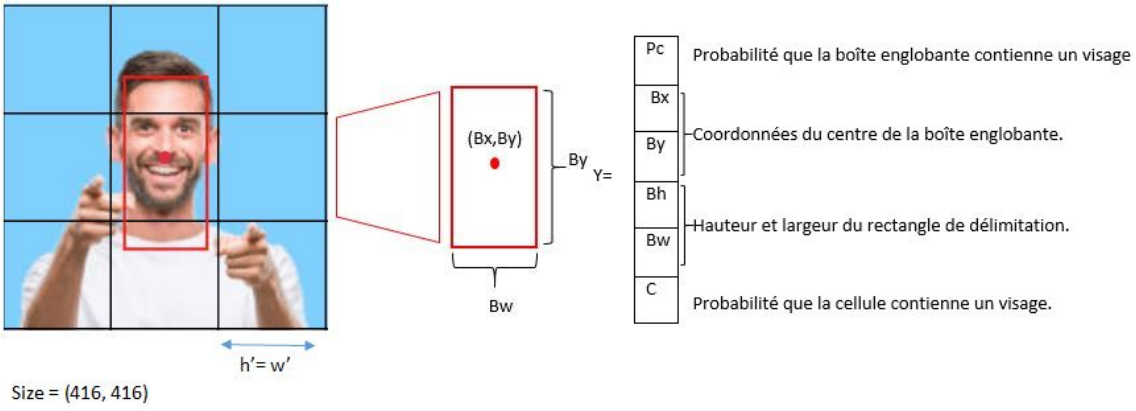

FIGURE 3.8 – Vecteur prédit dans le cas d'une seule boite englobante

Les valeurs du vecteur Y sont calculées comme suit :

P_c= la prédiction de confiance représente le IoU entre la boîte prédite et la boîte de vérité terrain.

b_x = (x - h')/h'

b_y = (y – w')/w'

b_h = h/416

b_w = w/416

• **Intersection sur Union (IoU) :** L'IoU est la métrique d'évaluation utilisée dans la détection d'objets. La métrique est utilisée pour déterminer les vrais positifs et les faux positifs dans un ensemble de prédictions. L'IoU compare la boîte prédite avec la boîte détectable. La figure 3.9 illuste un exemple de calcule de l'intersection sur l'union

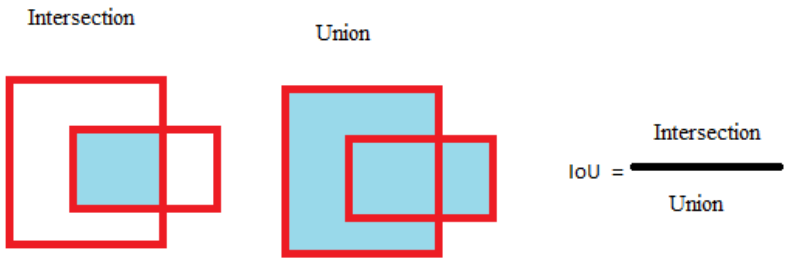

FIGURE 3.9 – Exemple de calcule de l'intersection sur l'union

La troisième étape est la suppression non maximale

3.2.2.3 Suppression non maximale (NMS) :

NMS ou Non maximum Suppression cette étape c'est la dernière étape dans l'algorithme de détection YOLO, elle consiste a supprimer les boîtes englobantes les moins probables et ne garder que la meilleure [97], le processus de cette technique se passe en 5 étapes :

1. La boite avec le score d'objectivité le plus élevé est sélectionnée

2. Le chevauchement calculé (intersection sur union) de cette boîte sera comparé avec d'autres boîtes.

3. Les boîtes englobantes dont le chevauchement (intersection sur union) est au-delà d'un certain seuil seront supprimées

4. Les étapes 2 et 3 vont se répéter jusqu'à ce qu'il n'y ait plus de cases avec un score inférieur à la case actuellement sélectionnée.

la figure 3.10 est une representation de la dernière étape dans la détection du visage avec YOLO.

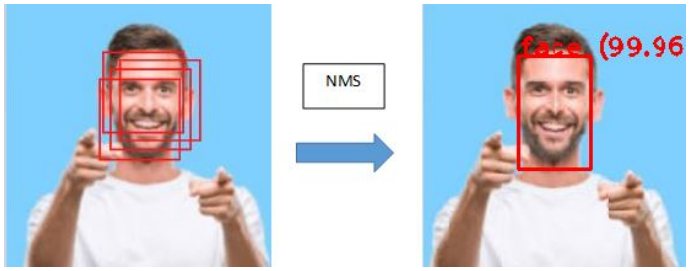

FIGURE 3.10 – Résultat de Suppression non maximale

La troisième méthode utilisée pour la détection du visage est le MTCNN

3.2.3 MTCNN : Multi-Task Convolutional Neural Network

Trois réseaux convolutifs (P-Net, R-Net et O-Net) sont mis en cascade pour sélectionner la meilleure fenêtre candidate a la présence des visages dans l'image. La figure 3.12 montre la composition de chaque bloc convolutif, on y retrouve les trois réseaux, qui effectuent une série de convolutions sur l'image de sorte que la valeur des quatre coins du cadre est générée par régression à la hiérarchie du visage. Ce dernier contient les visages potentiellement détectés et cinq points d'intérêt pour les visages (facial landmarks). Chaque CNN correspond aux étages traversés par un ensemble de trames choisi en fonction de la probabilité qu'ils contiennent un visage humain. La dernière étape produira le meilleur cadre qui contient le visage entier de l'image avec cinq coordonnées principales, qui representent les points d'intérêt du visage (ou repères faciaux) : deux pour les yeux, deux pour la bouche et un pour le nez.

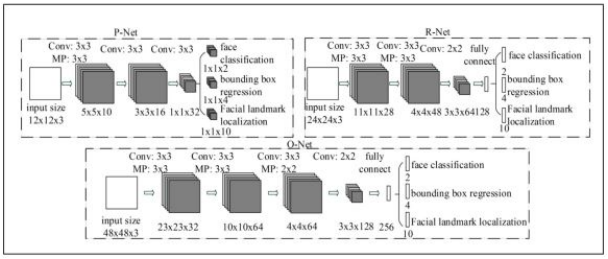

FIGURE 3.11 – Représentation architecturale du réseau en cascade MTCNN

3.2.3.1 Le fonctionnement du bloc P-Net

Le principe de cette étape de traitement consiste à créer une pyramide d'images à partir d'une image en entrée afin de détecter les visages de différentes tailles. En d'autres termes, différentes copies de différentes tailles d'une même image sont crées. Ceci afin de chercher des visages de différentes tailles dans l'image. Pour chaque copie du visage mise à l'échelle, un filtre de taille de noyau 12x12 parcourt toute l'image de manière incrémentale en recherchant les visages [121].

La figure 3.12 est un exemple d'une image sortie après le traitement du bloc P-net.

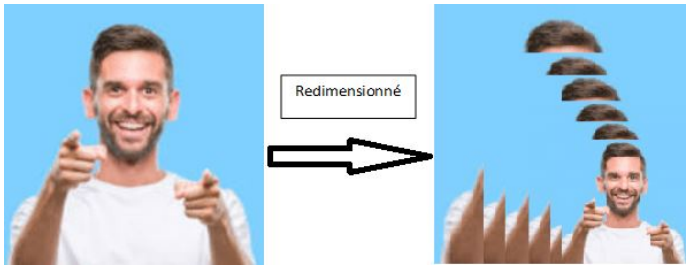

FIGURE 3.12 – Image piramide

Le traitement commence à partir de la zone d'image entre (0,0) et (12,12) situé dans le coin supérieur gauche comme le montre la figure. Ceci est passé au bloc PNet, qui renvoie les coordonnées de la boîte englobante s'il détecte un visage. Ensuite, il répète ce processus pour la région (0+2a,0+2b) à (12+2a,12+2b), En décalant les pixels d'un noyau 12x12 de 2 pixels vers la droite ou fond. Un décalage de 2 pixels s'appelle une foulée (foulée ou pas), ou Le pixel à chaque fois que le noyau bouge. La méthode MTCNN permet aussi l'alignement du visage [121].

La figure 3.13 le noyau du traitement a partir de la région d'une image.

FIGURE 3.13 – Noyau de la fenêtre

Chaque noyau serait plus petit par rapport à une image de grande taille en entrée, de sorte qu'il serait capable de trouver des visages plus petits. Plusieurs copies de l'image de différentes tailles sont créées et transmises au premier réseau de neurones P-Net.Les poids et les biais de P-Net ont été formés de manière à produire un cadre de sélection relativement précis pour chaque noyau de 12 x 12 [121].

il est nécessaire d'analyser la sortie P-Net pour obtenir une liste de niveaux de confiance pour chaque cadre de sélection et supprimer les cadres avec un niveau de confiance inférieur (c'est-à dire que les cases dont le réseau n'est pas tout à fait sûr de contenir un visage).les coordonnées des cadres de sélection sont normalisées en les convertissant en ceux de l'image réelle non mise à l'échelle. Cependant, il reste encore beaucoup de cadres de sélection, qui se chevauchent souvent. Le traitement basé sur la méthode suppression non maximale (NMS) est utilisé pour réduire le nombre de cadres de sélection. Le processus de NMS est effectué en triant d'abord les boites englobantes (et leurs noyaux respectifs 12 x 12) en fonction de leur confiance ou de leur score. Dans d'autres

modèles, le système NMS utilise la plus grande boite englobante au lieu de celle sur laquelle le réseau a le plus fiable score.

Les noyaux qui se chevauchent beaucoup avec le noyau très performant sont supprimés. Enfin, le traitement NMS renvoie une liste des boites englobantes survivantes. Le processus NMS est effectué une fois pour chaque image mise à l'échelle, puis une fois de plus avec tous les noyaux survivants de chaque échelle. Cela supprime les cadres de sélection redondants, ce qui permet de limiter la recherche à une boite précise par visage [121]. La figure 3.14 montre un réseau d'analyse par la sortie P-Net.

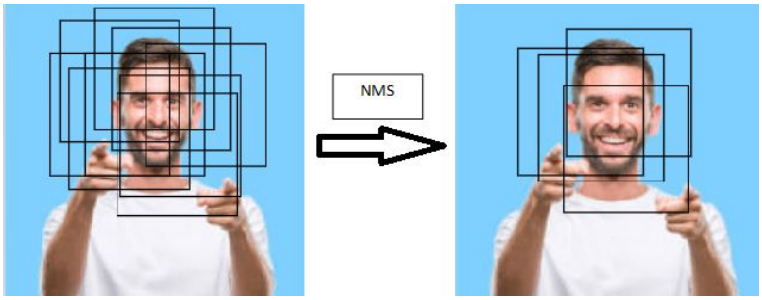

FIGURE 3.14 – Un réseau d'analyse par la sortie P-Net

3.2.3.2 Le fonctionnement du bloc R-Net

A la sortie du bloc P-Net, tous les candidats sont acheminés vers un autre CNN, appelé réseau raffiné (R-Net), qui rejette en outre un grand nombre de faux candidats et effectue un étalonnage avec une régression du cadre de sélection et une NMS [121].

Parfois, une image peut contenir uniquement une partie du visage qui s'observe du coté du cadre. Dans ce cas, le réseau peut renvoyer une zone de sélection partielle en dehors du cadre. Pour chaque cadre de sélection, un tableau de la même taille est créé et les valeurs de pixel (l'image dans le cadre de sélection) seront copiées dans le nouveau tableau. Si le cadre de sélection est en dehors des bords, uniquement la partie de l'image du cadre de sélection est copiée dans le nouveau tableau et le reste du tableau est complété par des zéros. Une fois que les tableaux des cadres de sélection sont remplis, ils sont redimensionnés à 24 x 24 pixels et normalisés à des valeurs comprises entre -1 et 1.

Une fois que tous tableaux d'images sont de taille 24 x 24 . La sortie de R-Net est similaire à celle de P-Net, elle inclut les coordonnées plus précises des nouveaux cadres de sélection, ainsi que le niveau de confiance de chacun de ces derniers. Une fois encore, les boites avec moins de confiance sont éliminées. Après normalisation des coordonnées, les boites englobantes sont transformées en un carré à transmettre au réseau O-Net [121].

La figure 3.15 est un exemple du réseau qui rejette les faux cadre condidats.

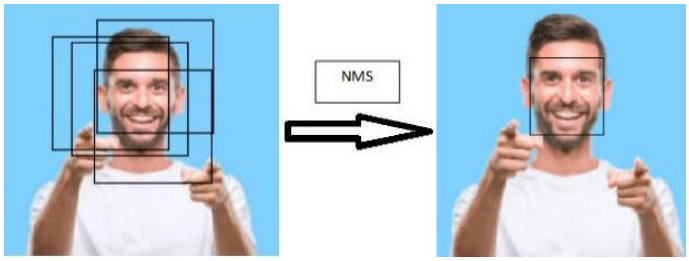

FIGURE 3.15 – Le réseau rejette un grand nombre de faux candidats.

3.2.3.3 Le fonctionnement du bloc O-Net

Cette étape est semblable à la deuxième étape, mais dans celle-ci, les régions de visage avec les cinq positions de points de repère faciaux sont identifiées. Avant de traiter les boites englobantes de R-Net, les boites qui sont hors limites sont remplies de zéros et sont redimensionnées à la taille 48 x 48 pixels, puis transmise au réseau O-Net. Le traitement effectué par le bloc O-Net fournit trois sorties : les coordonnées du cadre de sélection (sortie 0), les coordonnées des cinq points de repère faciaux (sortie 1) et le niveau de confiance de chaque cadre (sourtie 2).Une fois encore, les boites dont le niveau de confiance est faible sont éliminées et les coordonnées des boites englobantes et celles du repère facial sont standardisées. Enfin, le traitement NMS est appliqué. A ce stade, il ne doit exister qu'un seul cadre de sélection pour chaque visage de l'image. La toute dernière étape consiste à regrouper toutes les informations dans un dictionnaire comportant trois clés à savoir : (1) box, (2) confiance et (3) points-clés [121]. La figure 3.16 montre la détection du visage (cadre de sélection)avec les 5 points de repère faciaux.

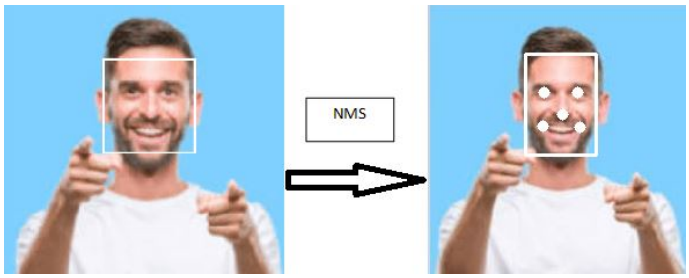

FIGURE 3.16 – Le cadre de sélection, les 5 points de repère faciaux.

3.3 Extractions des caractéristiques.

Après la détection de visages nous avons remarquée que plusieurs images ont besoin d'être aligner. Effectivement l'alignement de visage est une tache essentielle dans notre système d'estimation d'âge.

3.3.1 Alignement

L'alignement du visage est utilisé afin de normaliser le visage, pour cela les points centraux des yeux sont utilisés pour pivoter le visage et effectuer une rotation dans le sens des aiguilles d'une horloge, d'un angle θ comme le montre la figure 3.18 (La figure montre un exemple d'alignement). La figure 3.18(a) montre un exemple de visage non aligner et (b) et le resultat du visage aprés la rotation où le point central de l'image (Cx,Cy) est le centre de rotation ainsi que les nouvelles coordonnées.

$$\theta = tan^{-1}\left(\frac{R_y - L_y}{R_x - L_x}\right) \tag{3.7}$$

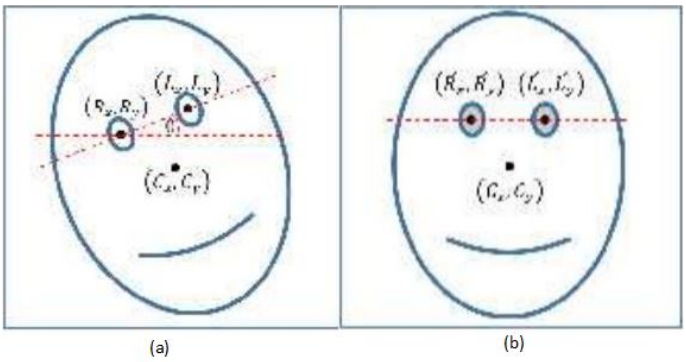

FIGURE 3.17 – Exemple de visage avant et après l'alignement [15].

Après rotation, les nouvelles coordonnées des points centraux des yeux sont données par :

$R'_x = C_y + (R_x - C_x).\sin(\theta) - (R_y - C_y).\cos(\theta)$ (3.8)

$L'_x = C_x + (L_x - C_x).\cos(\theta) - (L_y - C_y).\sin(\theta)$ (3.9)

$R'_y = C_y + (R_y - C_y).\sin(\theta) - (R_x - C_x).\cos(\theta)$ (3.10)

$L'_y = C_x + (L_y - C_y).\cos(\theta) - (L_x - C_x).\cos(\theta)$ (3.11)

Enfin, le ROI du visage est défini en utilisant la distance L entre les points centraux des yeux.

$L = \sqrt{(R'_x - L'_x)^2 + (R'_y - L'_x)^2}$

3.3.2 LBP : Local Binary Pattern

Le Local Binary Patterns est un descripteur de texture efficace et puissant. qui est largement utilisé dans le traitement des images pour supprimer les effets d'éclairage comme montre la figure 3.18 et dans les domaines de la vision par ordinateur en tant que caractéristique et représentation d'histogramme comme montre la figure 3.19.

Le principal mécanisme LBP est le suivant : l'image d'entrée est divisée en " N×N " régions locales, chaque région locale étant composée d'un voisinage 3×3 de chaque pixel par la valeur du

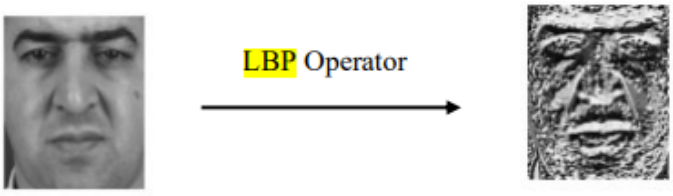

FIGURE 3.18 – L'image originale (gauche) traitée par l'opérateur LBP (droite)

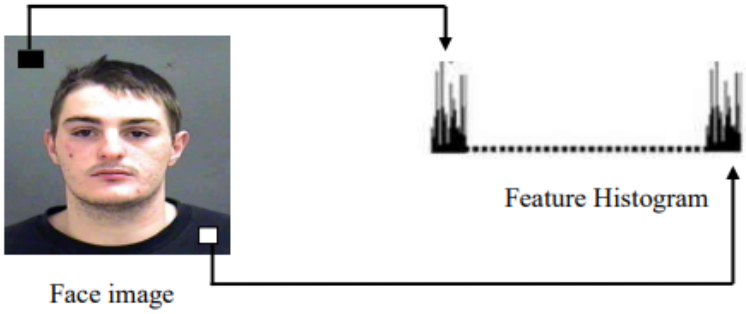

FIGURE 3.19 – Représentation par histogramme du visage basée sur le LBP.

pixel central. Ensuite, le type de motif binaire sera attribué une étiquette à chaque pixel en fonction de sa valeur d'intensité, où la distribution de ces motifs binaires dans chaque bloc représente les résultats avec un entier de 8 bits, où le calcul de ces motifs est utilisé comme représentation des caractéristiques [51].

En général, en supposant qu'un pixel se trouve à (Xc,Yc) les résultats LBP peuvent être calculés selon la formule suivante :

$$LBP(x_c, y_c) = \sum_{n=0}^{n-1} S(i_n - i_c) * 2^n \tag{3.4}$$

Ou i_n est un pixel voisin des N pixels autour du pixel central i_c.

Pour expliquer le mécanisme LBP (figure 3.21), nous supposons que l'image d'entrée a été divisée par l'algorithme LBP en 100 blocs. et qu'il traite actuellement le bloc numéro 29. Ce bloc est divisé en pixels de voisinage " 3×3 " (9 cellules), puis chaque pixel est codé en utilisant sa valeur d'intensité. Par seuil à partir de la valeur du pixel central (qui est 65 dans l'exemple sur la figure 3.21), LBP ordonne tous les pixels avoisinant le bloc, du coin supérieur gauche vers le coin droit. selon leur valeur d'intensité est supérieure ou inférieure à celle du pixel central (valeur supérieure = 1 ; valeur inférieure = 0). Enfin, nous obtiendrons un nombre binaire de 8 bits (c'est "11000011" dans la figure 3.21), qui peut être converti en un nombre décimal par la suite, par exemple (11000011)2 = (195)10,

où le calcul de ces nombres représente une valeur de un-demi. calcul de ces nombres représente un tableau unidimensionnel de motifs [51].

La figure 3.20 montre une description du visage basée sur le LBP.

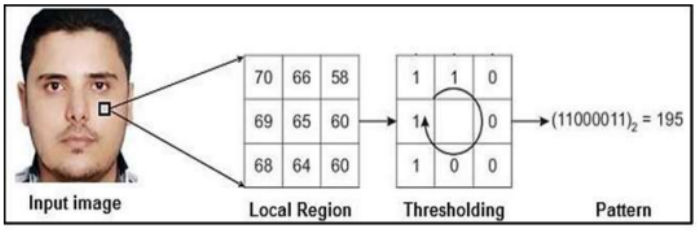

FIGURE 3.20 – Description du visage basée sur le LBP

La figure 3.21 montre un voisinage circulaire avec un rayon R différent et un nombre de voisins du point P.

La propriété la plus importante de LBP réside dans son invariance vis à vis des changements monotones de l'illumination causés par des variances d'éclairage pour les applications du monde réel. Une autre propriété aussi importante réside dans sa simplicité de calcul en temps réel [16].

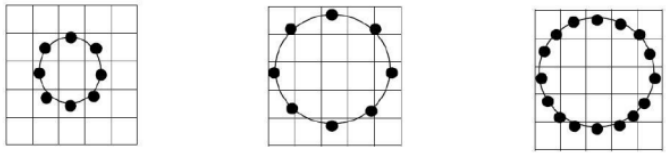

FIGURE 3.21 – Voisins symétriques circulaires pour différentes valeurs de p et r [16]

Le nombre total des valeurs des différentes sorties est 2^P qui sont générées par l'opérateur $LBP_{p,r}$ avec certaines valeurs correspondantes aux mêmes motifs suivant la rotation.

3.4 Réseau de neurone convolutif avec keras

Keras est une bibliothèque logicielle open source qui fournit une interface Python pour les réseaux de neurones artificiels et d'apprentissage automatique. Keras contient de nombreuses implémentations de blocs de construction de réseau de neurones couramment utilisés tels que des couches, des fonctions d'activation ..., et cela pour faciliter le travail avec les données d'image.

Keras fournit un cadre complet pour créer tout type de réseaux de neurones, il est innovant et très facile à apprendre. Il prend en charge un réseau de neurones simple à un modèle de réseau de neurones très vaste et complexe. Afin de créer un modèle dans Keras ces étapes sont essentielles :

• **Préparer des données :** Traiter, filtrer et selectionner les informations requises dans les données.

• **Division des données :** en ensemble de données d'entrainement et de test.

• **Compiler le modèle :** de sorte qui puisse être utilisé pour apprendre par entrainement

et faire de la prediction en utilisant des fonctions de perte (pour trouver l'erreur) et des fonctions d'optimisateur (afin que l'erreur soit minime).

- **Ajuster le modèle :** en utilisant un ensemble de données de formation.
- **Prédire le résultat :** pour une valeur inconnue prédir sa sortie.
- **Evaluer le modèle :** evaluer le modèle en predisant la sortie des données en comparant la prediction avec le resultat reel des données de test.

Les détails des réseaux de neurones CNN sont dans le chapitre2. Et la figure 3.22 représente le modèle du réseau de neurones utilisé.

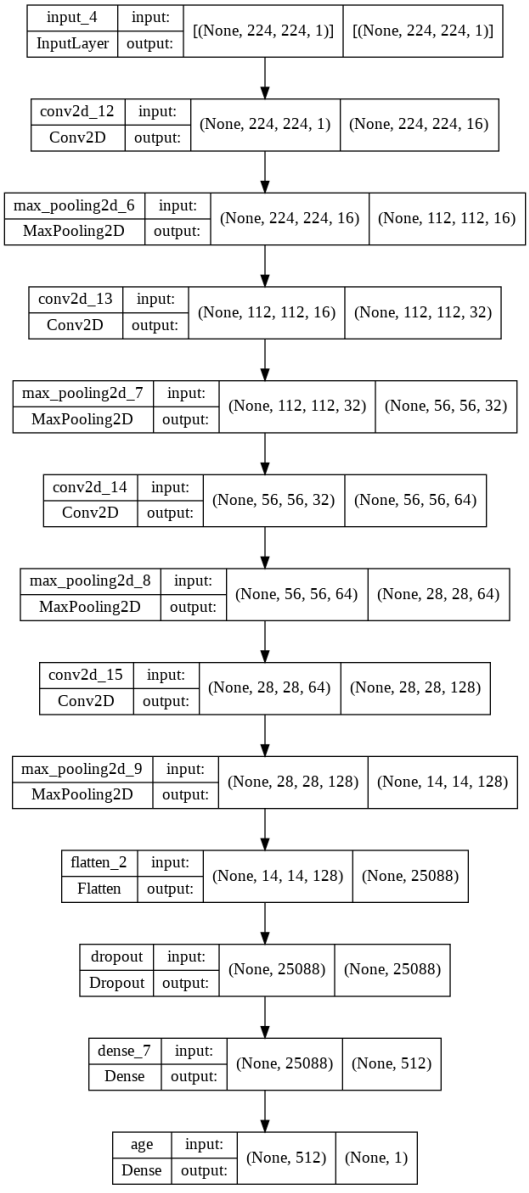

FIGURE 3.22 – Modèle du réseau de neurones utilisé

3.5 Conclusion

Dans ce chapitre nous avons détaillé le fonctionnemment des trois méthodes : la méthode de Viola et Jones, MTCNN et YOLO pour la détection du visage, ainsi que l'éxtraction des caractéristiques avec la fonction LBP, pour finir avec la regression par keras.

Chapitre 4

Chapitre 4.Tests et résultats

4.1 Introduction

Ce dernier chapitre est consacré aux testes et résultas de l'application qui permettra d'estimer l'âge des personnes à partir des images faciales avec les réseaux de neurones convolutifs (CNN).

Notre application permettra de détecter le visage des personnes à partir des images faciales après un apprentissage par les réseaux de neuronnes. Cette approche permet d'entrainer, prédire et évaluer les résultats expérimentaux obtenus.

4.2 Matériel utilisé

Pour réaliser notre application on a utilisé un micro portable qui a les spécifications suivantes :

FIGURE 4.1 – Caractéristique du matériel utilisé

4.3 Environnement de développement

Les outils d'apprentissage profond permettent aux sciences des données (Datas Scientists) de créer des programmes capables d'amener un ordinateur ou une machine à apprendre comme le cerveau humain et à traiter des données et des modèles avant d'exécuter des décisions.

La présentation suivante détaille certains d'outils les plus couramment utilisés et les plus importants pour le développement de notre approche basée sur le CNN.

4.3.1 Python

Python est un langage de programmation interprété, orienté objet, de haut niveau et à sémantique dynamique. La syntaxe de Python, est simple et facile à apprendre, privilégie la lisibilité et réduit donc le coût de la maintenance des programmes. Python prend en charge les modules et les packages, ce qui encourage la modularité des programmes et la réutilisation du code Python est un langage de programmation de haut niveau, polyvalent et très populaire.

4.3.2 OpenCV

OpenCV (Open Source Computer Vision Library) est une bibliothèque logicielle open source de vision par ordinateur et d'apprentissage automatique. OpenCV a été construit pour fournir une infrastructure commune pour les applications de vision par ordinateur et pour accélérer l'utilisation de la perception artificielle dans les produits commerciaux. Étant un produit sous licence BSD, OpenCV permet aux entreprises d'utiliser et de modifier facilement le code. La bibliothèque compte plus de 2500 algorithmes optimisés, ce qui inclut un ensemble complet d'algorithmes de vision par ordinateur et d'apprentissage automatique classiques et de pointe. Ces algorithmes sont spécialisé dans le traitement d'images en temps réel d'où ils peuvent être utilisés pour détecter et reconnaître des visages, identifier des objets, classer des actions humaines dans des vidéos, suivre les mouvements de la caméra, suivre des objets en mouvement, extraire des modèles 3D d'objets, etc.

Dans le processus de l'apprentissage nous avons utilisé la version 3.2.0.

4.3.3 NumPy

NumPy est le package de base pour le calcul scientifique en Python. Il s'agit d'une bibliothèque Python qui fournit un objet tableau multidimensionnel, divers objets dérivés (tels que des tableaux et des tableaux cachés) et une variété de procédures pour des opérations rapides sur les tableaux, y compris la forme, la logique, le contrôle, l'organisation, les changements, I/O, transformées de Fourier discrètes, algèbre variable linéaire de base, opérations statistiques de base, simulation stochastique et bien plus encore [3].

4.3.4 Matplotlib

Matplotlib est une bibliothèque complète permettant de créer des visualisations statiques, animées et interactives en Python. Matplotlib rend les choses faciles faciles et les choses difficiles possibles. Elle permet de crée des graphiques de qualité professionnelle, ainsi que des figures interactives pouvant être zoomées, panoramisées et actualisées tout en personnalisant le style visuel et la mise en page. Elle permet aussi de s'exporter vers de nombreux formats de fichiers [2].

4.3.5 TensorFlow

TensorFlow est une plateforme open-source pour la création d'applications d'apprentissage automatique. Il s'agit d'une bibliothèque de mathématiques symboliques qui utilise le flux de données et la programmation différentiable pour effectuer diverses tâches axées sur la formation et l'inférence de réseaux neuronaux profonds. Elle permet aux développeurs de créer des applications d'apprentissage automatique en utilisant divers outils, bibliothèques et ressources communautaires [6].

4.3.6 Google Colaboratory

Google Colab est un produit de Google, comme son nom l'indique. Il s'agit essentiellement d'un environnement de bloc-notes gratuit qui fonctionne entièrement dans le nuage informatique. Il dispose de fonctionnalités qui aident à modifier des documents de la même manière que travaille Google Docs. Colab prend en charge de nombreuses bibliothèques d'apprentissage automatique populaires et de haut niveau qui peuvent être facilement chargées dans votre notebook.

Google Colab nous offre trois types de runtime pour nos ordinateurs portables : CPUs, GPUs, et TPUs, Colab nous offre un total de 12 heures d'exécution continue. Après cela, toute la machine virtuelle est effacée et nous devons repartir de zéro. à cause de limite du l'utilisation des ressource google Colab.

Nous pouvons exécuter plusieurs instances CPU, GPU et TPU simultanément dans Google collab, mais les ressources sont partagées entre ces instances [4].

Pour notre apprentissage nous avons utilisé les ressources de Colab avec la GPU k80 avec RAM 13 GO et 12GO de VRAM.

4.4 Bases de données utilisées

Il existe plussieurs bases de données pour l'estimation de l'âge, nous avons travaillé avec deux bases de données FG-NET et UTKface, étant donné que ce sont les plus utilisées dans ce domaine [122], [39], [47], [22], [116], [24].

La première base de donnés utilisé est FG-NET :

4.4.1 FG-NET

FG-NET [69] contient 1002 images de 82 personnes et a été créé dans le but d'étudier l'âge réel. Les images sont à la fois en niveaux de gris et en couleur. Les âges vont de 0 à 69 ans. Chaque personne a une moyenne de 12 photos. La base de données fournit également des annotations pour différentes races humaines. Il existe une diversité de poses de tête et certaines expressions faciales ainsi qu'un certain éclairage sur les images. De plus, la base de données fournit 68 points de repère utiles pour la forme du visage la modélisation. L'ensemble de données est disponible en ligne.

La deuxième base de donnés utilisé est UTKface.

4.4.2 UTKFace

Le Dataset UTKFace est un Dataset de visages à grande échelle avec une longue fourchette d'âge allant de 0 à 116 ans. Elle se compose de plus de 20 000 images de visages avec des annotations d'âge, de sexe et d'ethnicité. Les images couvrent une grande variation dans la pose, l'expression faciale, l'illumination, l'occlusion, la résolution, etc. Ce jeu de données pourrait être utilisé pour une variété de tâches, par exemple, la détection de vi sages, l'estimation de l'âge, la progression/régression de l'âge, la localisation de points de repère, etc [5].

La Troisième base de donnés utilisé est Adience.

4.4.3 Adience

Le jeu de données Adience contient 26 580 photos à travers 2 284 sujets avec une étiquette binaire de sexe et une étiquette de huit groupes d'âge différents, partitionnés en cinq divisions. Le principe clé de l'ensemble de données est de capturer les images aussi proches que possible des conditions du monde réel, y compris toutes les variations d'apparence, de pose, de condition d'éclairage et de qualité d'image, pour n'en citer que quelques-unes[33].

4.5 Présentation de l'Application

Dans cette section on va présenter les différents aspects de notre application, elle contient 5 fenetres qu'on peut schematiser comme le montre la figure 4.2.

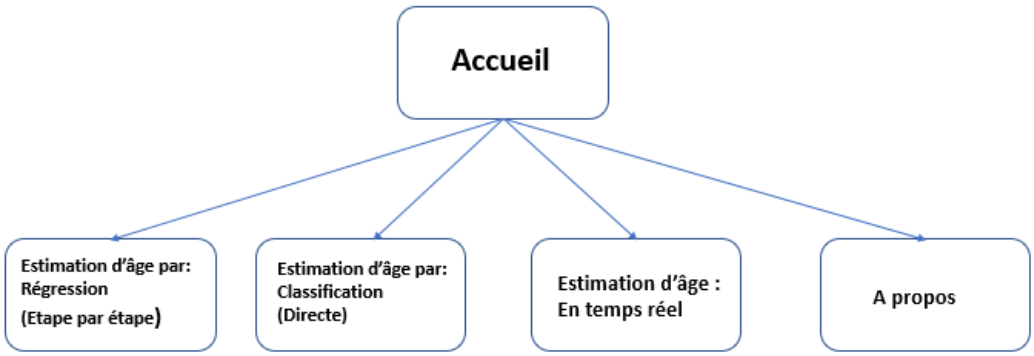

FIGURE 4.2 – Schéma de l'application

Fenêtre d'accueil

La fenêtre est montrée dans la figure 4.3.

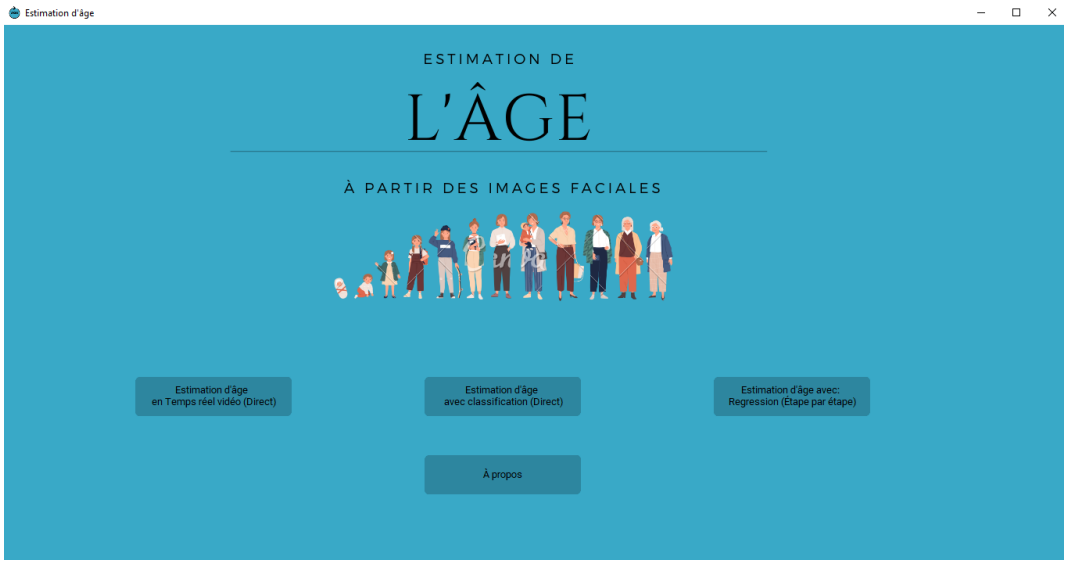

FIGURE 4.3 – Fenêtre d'accueil

Cette fênetre contient 4 boutons :

•Estimation d'âge avec Regression (Étape par étape) : c'est un bouton qui renvoi vers la fênetre de l'estimation d'âge en utilisant la régression étape par étape.

•Estimation d'âge avec classification (Direct) : C'est un bouton qui renvoi vers la fênetre de l'estimation d'âge en utilisant la classification Direct.

•Estimation d'âge en temps réel en vidéo Direct : C'est un bouton qui permet d'activé la cam et d'estimer l'âge.

•À propos : C'est un bouton qui renvoi vers la fênetre contenant les informations sur les développeurs de l'application.

• **Fênetre de l'estimation d'âge par la regression (étape par étape)**

La fenêtre est montrée dans la figure 4.4.

FIGURE 4.4 – Fenêtre d'estimation d'âge avec la regression

Cette fênetre contient 5 boutons et 3 boutons radio :

• Un bouton qui permet l'importation d'une image.

• Un bouton qui permet de faire la détection de visage. Ce dernier contient a son tour 3 boutons radio qui permettent de détecter le visage avec 3 algorithmes différent : les filtres de Haar, YOLO et MTCNN.

• Un bouton qui permet de faire l'alignement du visage détecté.

• Un bouton qui permet de faire l'extraction des caractéristiques.

• Un bouton qui fait l'estimation de l'âge.

• **Fênetre de l'estimation d'âge par la classification (Direct)**

Cette fênetre contient 2 boutons :

• Un bouton qui permet de faire l'importation d'une image.

• Un bouton qui permet de faire une détection de visage et une classification d'âge.

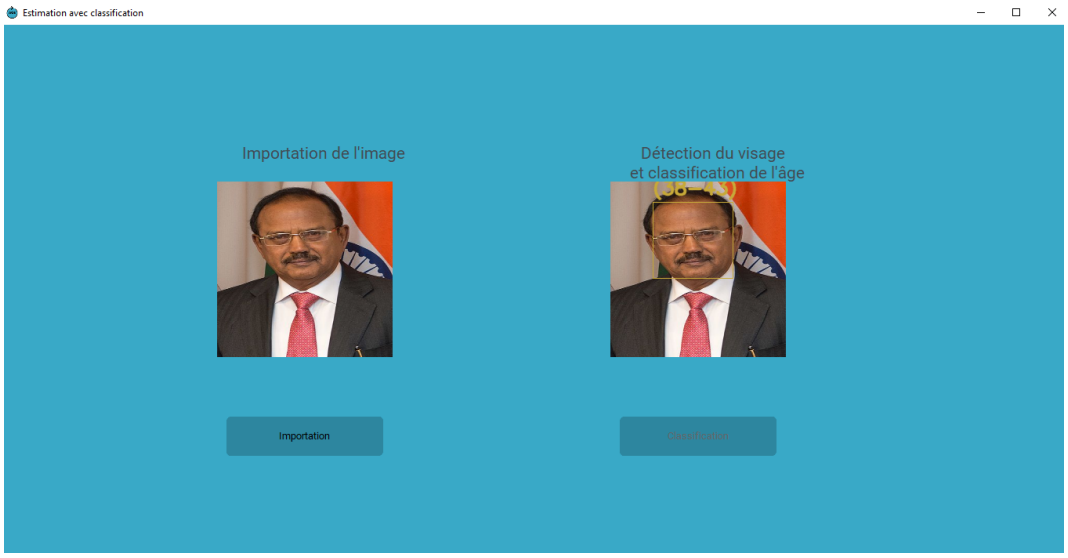

FIGURE 4.5 – Fenêtre d'estimation d'âge avec la classification

- **Fênetre A propos**

Est une fênetre qui contient les informations personnelles des personnes qui ont contribuer au développement de cette application.

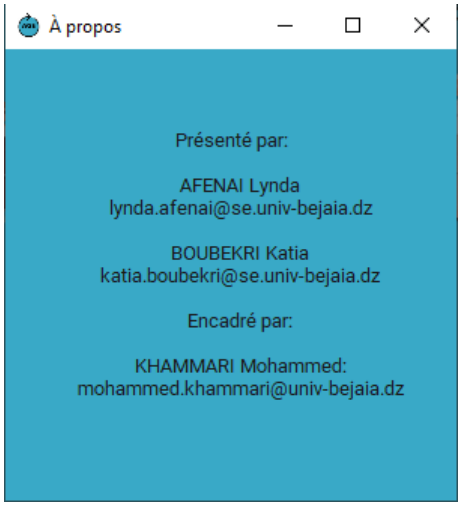

FIGURE 4.6 – Fênetre à propos

- **Fênetre de l'estimation d'âge en temps réel**

• Un bouton qui permet l'activation de la cam et l'estimation d'age en direct par la méthode de classification.

4.6 Test et Résultat

4.6.1 Pré-traitement

Afin d'effectuer l'apprentissage nous avons divisé les deux bases de données FGnet et UTKface en deux jeux de données. Le premier pour l'apprentissage avec 80% des images de FGNET et 80% des images de UTKface. Le second qui est le reste des 20% FG-NET et 20% UTKface est utilisé pour la validation du modèle aprés son apprentissage.

Les images ont été redimensionnées à la taille de 224x224 pixels et normalisées entre 0 et 1.

4.6.2 Apprentissage du modèle avec la régression

Préparer le modèle pour l'apprentissage

Le modèle est un réseau convolutif à 4 couches. Chaque couche se composant de 32 filtres doublant à chaque couche, puis est suivie d'une activation ReLU non linéaire puis d'un max pooling, sauf à la fin, qui a une activation linéaire afin de réaliser une regression. L'entraînement est lancé avec 100 epochs et un learning rate de 0.001.

La fonction côut utiliser est la MAE (MEAN ABSOLUTE ERROR). A la fin de l'apprentissage on a obtenu la courbe montré sur la figure 4.7.

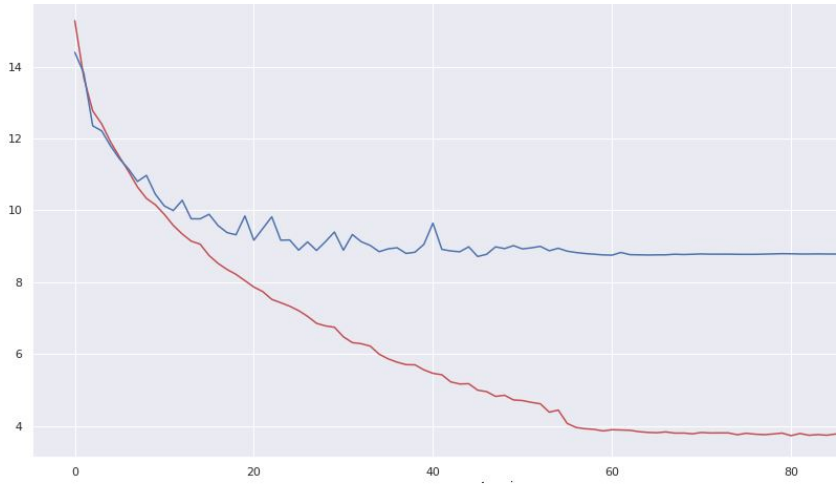

FIGURE 4.7 – Courbe de la MAE obtenue en fonction d'epochs

La courbe rouge représente la courbe obtenue lors de l'entrainement du modèle.

La courbe bleu représente la courbe obtenue lors de la validation du modèle.

Taux de réussite

Afin de déterminer un taux de réussite de l'apprentissage par la régression avec keras, au-

trement dit un taux de présision (Accuracy) nous avons donnée de differents seuils d'acceptation, cette notion permet de définir une marge d'erreur entre les résultats de l'estimation et l'âge réel.

Les résultatas obtenus sont dans le tableau 4.1.

Seuils par années	3	4	5	6	7	8	9	10
Résultats %	33.05	41.11	46.07	51.86	56.61	61.57	66.94	71.28

TABLE 4.1 – Tableau des résultats obtenus avec de différents seuils lors de la régression

Classes d'âge	[0-2]	[4-6]	[8-12]	[15-20]	[25-32]	[38-43]	[48-53]	[60-100]
Résultats %	70	57	55	23	61	29	14	35

TABLE 4.2 – Tableau des résultats obtenus lors de la classification

4.7 Comparaison avec l'état de l'art

Le tableau 4.2 représente une comparaison avec l'état de l'art en terme de MAE.

Méthodes	Bases données	MAE	accuracy
forme 2D [116]	FG-NET	8.84	/
Modèle anhropométrique [122]	FG-NET	7.48	/
AGES [39]	FG-NET MORPH	6.22, 8.07	/
Modèle apprence, SPF [118]	YGA	8.17	/
LBP [44]	FERET	/	80%
AAM [38]	FG-NET	5.77	/
BSIF, LBP [15]	PAl, FG-NET	6.25, 6.28	/
CNN [114]	MORPH FG-NET	4.77, 4.26	/
LBP [17]	BW kennedy	/	91.1%
ELM [31]	MORPH	2.61	/
ShuffleNet-V2 [79]	MORPH FG-NET	2.68, 3.8	/
Notre approche (Keras)	FG-Net, UTK-face	8.6	/
Notre approche (caffenet)	adience	/	91.6%

TABLE 4.3 – Tableau comparaison avec l'état de l'art en terme de MAE et de l'accuracy

4.8 Interprétation et discussion de résultat

En comparant les travaux connexes de l'estimation faciale d'âge, nous remarquons que notre approche en utilisant les réseaux de neurones convolutif avec Keras donne une moyenne d'erreurs absolue (MAE) de 8.6, qui sont meilleurs que les résultats de la méthode de Wu et al [116] de 0.24. De plus les résultats obtenus ne sont pas loin des résultats des méthodes de Yan et al [118], Zhou et al [122], ainsi que celle de Geng et al [39].

Les méthodes de Duan et al[31] et celle de Liu et al [79] ont obtenus les meilleurs résultats avec les plus faibles MAE.

On remarque que les auteurs qui ont utilisé le BSIF et le LBP converger vers de mêmes résultats que notre approche. L'extraction des caractéristiques avec LBP est une méthode pertinente alors on juge que notre approche est satisfaisante.

4.9 Conclusion

Dans ce chapitre, nous avons présenté en détail la conception de notre système, les algorithmes conçus, les méthodes mises en œuvre, tous les dispositifs utilisés, ainsi que les interfaces de notre application et quelques résultats de tests dans différents cas.

On peut dire que l'application mise en œuvre permet d'estimer l'âge des personnes en utilisant plusieurs algorithme de détection de visage. Les résultats obtenus est jugé satisfaisants.

Conclusion générale

Les systèmes d'estimation de l'âge ont connu une croissance rapide ces dernières années en raison de leurs modules importants et de leurs utilisations bénéfiques pour de nombreuses applications de vision par ordinateur.

Les chercheurs ont montré que l'être humain utilise pour estimer l'âge d'un visage ces différentes caractéristiques qui varient : les yeux, la bouche, le nez, le franc, les joues et le montant. Grâce à cette remarque, plusieurs études ont été développées afin de savoir s'il était possible de modéliser d'une manière informatique ce comportement.

Cependant, résoudre cette thématique n'est pas une tâche facile étant donné que les méthodes de prédiction ont une certaine complexité dans l'analyse des données et leurs généralisation, en dautres termes ces données ne sont pas faciles à généraliser, elles conviennent donc pas à tout le monde.

L'objectif de notre travail était de mettre en œuvre un système d'estimation de l'âge à partir d'une image faciale en utilisant les réseaux de neurones convolutifs.

Notre système comprend quatre étapes importantes : la détection du visage, l'alignement, l'extraction des caractéristiques et l'apprentissage par les réseaux de neurones convolutifs (CNN), pour obtenir l'âge estimé en sortie de notre système.

Dans la première étape trois méthodes de détection sont utilisées : les filtres de Haar, YOLO ainsi que MTCNN. Cette étape consiste à trouver les coordonnées spatiales qui définissent un visage dans une image. En termes simples, cela revient à trouver le carré qui définit le mieux le visage visible sur l'image.

La deuxième étape consiste à aligner le visage détecté. Cette méthode est utilisée comme une étape de prétraitement, elle permet identifier la structure géométrique des visages humains et pivoter le visage et effectuer une rotation de sorte à voir les yeux horizontal.

L'étape d'extraction des caractéristiques avec le descripteur LBP représente le cœur du système d'estimation d'âge, elle consiste à effectuer le traitement de l'image dans un autre espace de travail plus simple et qui assure une meilleure exploitation de données, et donc permettre l'utilisation seulement des informations utiles.

Finalement, l'étape d'estimation d'âge, deux méthodes sont utilisé la classification avec caffenet et la régression avec keras.

Les résultats était d'une précision de 91.6% avec la classification et d'une moyenne d'erreur absolue de 8.6. Les résultats obtenus est jugé satisfaisants.

Bibliographie

[1] https ://keras.io/about/, Consulter le : 23/06/2022.

[2] https ://matplotlib.org/, Consulter le : 23/06/2022.

[3] https ://numpy.org/doc/stable/user/whatisnumpy.html, consulter le : 23/06/2022.

[4] https ://research.google.com/colaboratory/faq.html, Consulter le : 23/06/2022.

[5] https ://susanqq.github.io/utkface/, Consulter le : 23/06/2022.

[6] https ://www.guru99.com/what-is-tensorflow.html, Consulter le : 23/06/2022.

[7] Ammari Ahmed. "facial age estimation using multidimensional teda method". *Mémoire de master, Université de Biskra*, pages 1–70, 2020.

[8] Timo Ahonen, Abdenour Hadid, and Matti Pietikainen. "face description with local binary patterns : Application to face recognition". *IEEE transactions on pattern analysis and machine*, 28 :2037–2041, 2006.

[9] John Ajala. Object detection and recognition using yolo : Detect and recognize url (s) in an image scene. pages 28–43, 2021.

[10] Raphael Angulu, Jules R Tapamo, and Aderemi O Adewumi. "age estimation via face images : a survey". *EURASIP Journal on Image and Video Processing*, pages 1–35, 2018.

[11] NADJETTE ASSADI. "mise au point d'une application de reconnaissance faciale". *Mémoire de master, Université de Biskra*, pages 1–66.

[12] NADJETTE ASSADI. Mise au point d'une application de reconnaissance faciale. *Mémoire de master, Université de Biskra*.

[13] Zineb Baghdadi and Besma Labandji. "système de vidéosurveillance avec la reconnaissance faciale". *Mémoire de master, Université de bouira*, pages 1–75, 2017.

[14] Azam Bastanfard, Melika Abbasian Nik, and Mohammad Mahdi Dehshibi. "iranian face database with age, pose and expression". In *2007 International Conference on Machine Vision*, pages 50–55. IEEE, 2007.

[15] Salah Eddine Bekhouche, Abdelkrim Ouafi, Abdelmalik Taleb-Ahmed, Abdenour Hadid, and Azeddine Benlamoudi. "facial age estimation using bsif and lbp". *ArXiv preprint*, pages 1–5, 2016.

[16] Halima Bouaka, Oum El kheir Bazzine, and Benlamoudi Azeddine. "estimation automatique de l'âge des patients à partir des images faciales". *Mémoire de master, Université de Ouargla*, pages 1–76, 2019.

[17] Imed Bouchrika, Nouzha Harrati, Ammar Ladjailia, and Sofiane Khedairia. "age estimation from facial images based on hierarchical feature selection". In *16th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering*, pages 393–397, 2015.

[18] Sheryl Brahnam, Lakhmi C Jain, Loris Nanni, Alessandra Lumini, et al. *book : "Local binary patterns : new variants and applications"*, volume 506. 2014.

[19] Lijun Cai, Lei Huang, and Changping Liu. "age estimation based on improved discriminative gaussian process latent variable model". *Multimedia Tools and Applications*, 75 :11977–11994, 2016.

[20] Bor-Chun Chen, Chu-Song Chen, and Winston H Hsu. "cross-age reference coding for age-invariant face recognition and retrieval". In *European conference on computer vision*, pages 768–783, 2014.

[21] Shixing Chen, Caojin Zhang, and Ming Dong. "deep age estimation : From classification to ranking". *IEEE Transactions on Multimedia*, 20 :2209–2222, 2017.

[22] Sung Eun Choi, Youn Joo Lee, Sung Joo Lee, Kang Ryoung Park, and Jaihie Kim. "age estimation using a hierarchical classifier based on global and local facial features". *Pattern recognition*, 44 :1262–1281, 2011.

[23] François Chollet. "xception : Deep learning with depthwise separable convolutions". In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1251–1258, 2017.

[24] Alireza Keshavarz Choobeh. "improving automatic age estimation algorithms using an efficient ensemble technique". *International Journal of Machine Learning and Computing*, 2 :118, 2012.

[25] Ammar Chouchane. "analyse d'images d'expressions faciales et orientation de la tête basée sur la profondeur". *Thèse de doctorat, Université Mohamed Khider de Biskra*, pages 1–143, 2016.

[26] RAOUNAK LILIA DAHAH. La détection de la colère chez le conducteur en utilisant le deep learning. 2020.

[27] Navneet Dalal and Bill Triggs. "histograms of oriented gradients for human detection". In *IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, pages 886–893, 2005.

[28] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. "imagenet : A large-scale hierarchical image database". In *IEEE conference on computer vision and pattern recognition*, pages 248–255, 2009.

[29] Chesne Desir. "classification automatique d'images, application à l'imagerie du poumon profond". *Thèse de doctorat, Université de Rouen*, pages 1–166, 2013.

[30] Abhinav Dhall, Akshay Asthana, Roland Goecke, and Tom Gedeon. "emotion recognition using phog and lpq features". In *IEEE International Conference on Automatic Face & Gesture Recognition (FG)*, pages 878–883, 2011.

[31] Mingxing Duan, Kenli Li, and Keqin Li. "an ensemble cnn2elm for age estimation". *IEEE Transactions on Information Forensics and Security*, 13 :758–772, 2017.

[32] Natalie C Ebner, Michaela Riediger, and Ulman Lindenberger. "faces—a database of facial expressions in young, middle-aged, and older women and men : Development and validation". *Behavior research methods*, 42 :351–362, 2010.

[33] Eran Eidinger, Roee Enbar, and Tal Hassner. "age and gender estimation of unfiltered faces". *IEEE Transactions on information forensics and security*, 9 :2170–2179, 2014.

[34] MYMH El Dib. "automatic facial age estimation". *Mémoire de master, Université de Cairo*, page $1_101, 2011$.

[35] Sergio Escalera, Junior Fabian, Pablo Pardo, Xavier Baró, Jordi Gonzalez, Hugo J Escalante, Dusan Misevic, Ulrich Steiner, and Isabelle Guyon. "chalearn looking at people 2015 : Apparent age and cultural event recognition datasets and results". In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 1–9, 2015.

[36] Songhe Feng, Congyan Lang, Jiashi Feng, Tao Wang, and Jiebo Luo. "human facial age estimation by cost-sensitive label ranking and trace norm regularization". *IEEE Transactions on Multimedia*, 19 :136–148, 2016.

[37] Yun Fu and Thomas S Huang. "human age estimation with regression on discriminative aging manifold". *IEEE Transactions on Multimedia*, 10 :578–584, 2008.

[38] Xin Geng, Chao Yin, and Zhi-Hua Zhou. "facial age estimation by learning from label distributions". *IEEE transactions on pattern analysis and machine intelligence*, 35 :2401–2412, 2013.

[39] Xin Geng, Zhi-Hua Zhou, and Kate Smith-Miles. "automatic age estimation based on facial aging patterns". *IEEE Transactions on pattern analysis and machine intelligence*, 29 :2234–2240, 2007.

[40] Markos Georgopoulos, Yannis Panagakis, and Maja Pantic. "modeling of facial aging and kinship : A survey". *Image and Vision Computing*, 80 :58–79, 2018.

[41] Ross Girshick. "fast r-cnn". In *Proceedings of the IEEE international conference on computer vision*, pages 1440–1448, 2015.

[42] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. "rich feature hierarchies for accurate object detection and semantic segmentation". In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 580–587, 2014.

[43] Petra Grd. "introduction to human age estimation using face images". *The Journal of Slovak University of Technology*, pages 24–30, 2013.

[44] Asuman Gunay and Vasif V Nabiyev. "automatic age classification with lbp". In *23rd international symposium on computer and information sciences*, pages 1–4, 2008.

[45] Gongde Guo, Hui Wang, David Bell, Yaxin Bi, and Kieran Greer. "knn model-based approach in classification". In *OTM Confederated International Conferences "On the Move to Meaningful Internet Systems"*, pages 986–996, 2003.

[46] Guodong Guo, Yun Fu, Charles R Dyer, and Thomas S Huang. "image-based human age estimation by manifold learning and locally adjusted robust regression". *IEEE Transactions on Image Processing*, 17 :1178–1188, 2008.

[47] Guodong Guo, Yun Fu, Thomas S Huang, and Charles R Dyer. "locally adjusted robust regression for human age estimation". In *IEEE Workshop on Applications of Computer Vision*, pages 1–6, 2008.

[48] Guodong Guo, Guowang Mu, Yun Fu, and Thomas S Huang. "human age estimation using bio-inspired features". In *IEEE conference on computer vision and pattern recognition*, pages 112–119, 2009.

[49] Ziad M Hafed and Martin D Levine. "face recognition using the discrete cosine transform". *International journal of computer vision*, 43 :167–188, 2001.

[50] Mohammad Ali Hajizadeh and Hossein Ebrahimnezhad. "classification of age groups from facial image using histograms of oriented gradients". In *7th Iranian Conference on Machine Vision and Image Processing*, pages 1–5, 2011.

[51] David Harwood, Timo Ojala, Matti Pietikäinen, Shalom Kelman, and Larry Davis. "texture classification by center-symmetric auto-correlation, using kullback discrimination of distributions". *Pattern Recognition Letters*, 16 :1–10, 1995.

[52] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. "mask r-cnn". In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017.

[53] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "deep residual learning for image recognition". In *Proceedings of the IEEE conference on computer vision and pattern recognition*, volume 2, pages 770–778, 2016.

[54] Zhenzhen Hu, Yonggang Wen, Jianfeng Wang, Meng Wang, Richang Hong, and Shuicheng Yan. "facial age estimation with age difference". *IEEE Transactions on Image Processing*, 26 :3087–3097, 2016.

[55] Ivan Huerta, Carles Fernández, Carlos Segura, Javier Hernando, and Andrea Prati. "a deep analysis on age estimation". *Pattern Recognition Letters*, 68 :239–249, 2015.

[56] Aapo Hyvärinen, Juha Karhunen, and Erkki Oja. "independent component analysis, adaptive and learning systems for signal processing, communications, and control". *John Wiley & Sons, Inc*, pages 11–14, 2001.

[57] Ryotatsu Iga, Kyoko Izumi, Hisanori Hayashi, Gentaro Fukano, and Tetsuya Ohtani. "a gender and age estimation system from face images". In *SICE 2003 Annual Conference (IEEE Cat. No. 03TH8734)*, pages 756–761, 2003.

[58] Vidit Jain and Erik Learned-Miller. "fddb : A benchmark for face detection in unconstrained settings". Technical report, UMass Amherst technical report, 2010.

[59] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. Caffe : Convolutional architecture for fast feature embedding. In *Proceedings of the 22nd ACM international conference on Multimedia*, pages 675–678, 2014.

[60] Boyi Jiang, Juyong Zhang, Bailin Deng, Yudong Guo, and Ligang Liu. "deep face feature for face alignment". *arXiv preprint arXiv :1708.02721*, 2017.

[61] J-K Kamarainen, Ville Kyrki, and Heikki Kalviainen. "invariance properties of gabor filter-based features-overview and applications". *IEEE Transactions on image processing*, 15 :1088–1099, 2006.

[62] Juho Kannala and Esa Rahtu. "bsif : Binarized statistical image features". In *Proceedings of the 21st international conference on pattern recognition (ICPR2012)*, pages 1363–1366, 2012.

[63] Tsuneo Kanno, Masakazu Akiba, Yasuaki Teramachi, Hiroshi Nagahashi, and Takeshi Agui. "classification of age group based on facial images of young males by using neural networks". *IEICE TRANSACTIONS on Information and Systems*, 84 :1094–1101, 2001.

[64] Kristen M Kennedy, Kelly Hope, and Naftali Raz. "life span adult faces : Norms for age, familiarity, memorability, mood, and picture quality". *Experimental aging research*, 35 :268–275, 2009.

[65] Adrian Kjærran, Christian Bakke Vennerød, and Erling Stray Bugge. "facial age estimation using convolutional neural networks". *ArXiv preprint*, pages 1–17, 2021.

[66] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. "imagenet classification with deep convolutional neural networks". *Advances in neural information processing systems*, 25 :1–9, 2012.

[67] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012.

[68] Young H Kwon and Niels da Vitoria Lobo. "age classification from facial images". *Computer vision and image understanding*, 74 :1–21, 1999.

[69] A Lanitis. "the fg-net aging database". *Available : www-prima. inrialpes. fr/FGnet/html/benchmarks. html*, 2, 2002.

[70] Andreas Lanitis, Chrisina Draganova, and Chris Christodoulou. "comparing different classifiers for automatic age estimation". *IEEE Transactions on Systems, Man, and Cybernetics*, 34 :621–628, 2004.

[71] Andreas Lanitis, Christopher J. Taylor, and Timothy F Cootes. "toward automatic simulation of aging effects on face images". *IEEE Transactions on pattern Analysis and machine Intelligence*, 24 :442–455, 2002.

[72] Than Le. "efficient post-contour correctness in object detection and segmentation". 2020.

[73] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. "gradient-based learning applied to document recognition". *Proceedings of the IEEE*, 86 :2278–2324, 1998.

[74] Haoxiang Li, Zhe Lin, Xiaohui Shen, Jonathan Brandt, and Gang Hua. "a convolutional neural network cascade for face detection". In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5325–5334, 2015.

[75] Kai Li, Junliang Xing, Weiming Hu, and Stephen J Maybank. "d2c : Deep cumulatively and comparatively learning for human age estimation". *Pattern Recognition*, 66 :95–105, 2017.

[76] Xiaochao Li, Zhenjie Yang, and Hongwei Wu. "face detection based on receptive field enhanced multi-task cascaded convolutional neural networks". *IEEE Access*, 8 :174922–174930, 2020.

[77] Hao Liu, Jiwen Lu, Jianjiang Feng, and Jie Zhou. Ordinal deep feature learning for facial age estimation. In *12th IEEE International Conference on Automatic Face & Gesture Recognition*, pages 157–164, 2017.

[78] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. "ssd : Single shot multibox detector". In *European conference on computer vision*, pages 21–37, 2016.

[79] Xinhua Liu, Yao Zou, Hailan Kuang, and Xiaolin Ma. "face image age estimation based on data augmentation and lightweight convolutional neural network". *Symmetry*, 12 :129–196, 2020.

[80] David G Lowe. "distinctive image features from scale-invariant keypoints". *International journal of computer vision*, 60 :91–110, 2004.

[81] Long-Hua Ma, Hang-Yu Fan, Zhe-Ming Lu, and Dong Tian. "acceleration of multi-task cascaded convolutional networks". *IET Image Processing*, pages 2435–2441, 2020.

[82] Larry R Medsker and LC Jain. "recurrent neural networks". *Design and Applications*, 5 :64–67, 2001.

[83] Anto A Micheala and R Shankar. Automatic age and gender estimation using deep learning and extreme learning machine. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, 12 :63–73, 2021.

[84] Meredith Minear and Denise C Park. "a lifespan database of adult facial stimuli". *Behavior research methods, instruments, & computers*, 36 :630–633, 2004.

[85] MERAMRIA Nabila. "reconnaissance de visages par analyse discriminante linéaire (lda)". *Mémoire de master, Université Badji Mokhtar Annaba*, pages 1–52, 2016.

[86] Muhammad Waqas Nadeem, Hock Guan Goh, Abid Ali, Muzammil Hussain, Muhammad Adnan Khan, and Vasaki a/p Ponnusamy. "bone age assessment empowered with deep learning : a survey, open research challenges and future directions". *Diagnostics*, 10 :781, 2020.

[87] Mahyar Najibi, Mohammad Rastegari, and Larry S. Davis. "g-cnn : An iterative grid based object detector". In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2369–2377, 2016.

[88] Mahyar Najibi, Pouya Samangouei, Rama Chellappa, and Larry S Davis. "ssh : Single stage headless face detector". In *Proceedings of the IEEE international conference on computer vision*, pages 4875–4884, 2017.

[89] Bingbing Ni, Zheng Song, and Shuicheng Yan. "web image and video mining towards universal and robust age estimator". *IEEE Transactions on Multimedia*, 13 :1217–1229, 2011.

[90] Zhenxing Niu, Mo Zhou, Le Wang, Xinbo Gao, and Gang Hua. "ordinal regression with multiple output cnn for age estimation". In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4920–4928, 2016.

[91] Ville Ojansivu and Janne Heikkilä. "blur insensitive texture classification using local phase quantization". In *International conference on image and signal processing*, pages 236–243, 2008.

[92] Hassiba Ouakkaf and M Berkane. "reconnaissance automatique des expressions faciales par support vector machine". *Mémoire de master, Université d'Oum El Bouaghi*, pages 1–60, 2017.

[93] Abdelmalik Ouamane. "reconnaissance biométrique par fusion multimodale du visage 2d et 3". *Thèse de doctorat, Université Mohamed Khider de Biskra*, pages 1–194, 2015.

[94] P Jonathon Phillips, Harry Wechsler, Jeffery Huang, and Patrick J Rauss. "the feret database and evaluation procedure for face-recognition algorithms". *Image and vision computing*, 16 :295–306, 1998.

[95] Jhony K Pontes, Alceu S Britto Jr, Clinton Fookes, and Alessandro L Koerich. "a flexible hierarchical approach for facial age estimation based on multiple features". *Pattern Recognition*, 54 :34–51, 2016.

[96] Zakariya Qawaqneh, Arafat Abu Mallouh, and Buket D Barkana. "deep convolutional neural network for age estimation based on vgg-face model". *ArXiv preprint*, pages 1–8, 2017.

[97] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. "you only look once : Unified, real-time object detection". pages 779–788, 2016.

[98] Joseph Redmon and Ali Farhadi. "yolov3 : An incremental improvement". *arXiv preprint arXiv :1804.02767*, 2018.

[99] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. "faster r-cnn : Towards real-time object detection with region proposal networks". *Advances in neural information processing systems*, 28, 2015.

[100] Karl Ricanek and Tamirat Tesafaye. "morph : A longitudinal image database of normal adult age-progression". In *7th international conference on automatic face and gesture recognition (FGR06)*, pages 341–345, 2006.

[101] Rasmus Rothe, Radu Timofte, and Luc Van Gool. "deep expectation of real and apparent age from a single image without facial landmarks". *International Journal of Computer Vision*, 126 :144–157, 2018.

[102] Mohamed Rouili. "automatic facial age estimation". *Thèse de doctorat, Université de Tebessa*, pages 1–700, 2020.

[103] Pierre Sermanet, David Eigen, Xiang Zhang, Michaël Mathieu, Rob Fergus, and Yann LeCun. "overfeat : Integrated recognition, localization and detection using convolutional networks". *arXiv preprint arXiv :1312.6229*, 2013.

[104] Karen Simonyan and Andrew Zisserman. "very deep convolutional networks for large-scale image recognition". *arXiv preprint arXiv :1409.1556*, pages 1–14, 2014.

[105] Ali SOUFI, Ismail ADDOU, Mohamed KOHILI, et al. *Réalisation d'un système de détection des visages en utilisant la matrice PSSM–positionspecific-scoring-matrix*. PhD thesis, University Ahmed Draia-ADRAR, 2018.

[106] Jinli Suo, Tianfu Wu, Songchun Zhu, Shiguang Shan, Xilin Chen, and Wen Gao. "design sparse features for age estimation using hierarchical face model". In *8th IEEE International Conference on Automatic Face & Gesture Recognition*, pages 1–6, 2008.

[107] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. "going deeper with convolutions". In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

[108] Christian Szegedy, Alexander Toshev, and Dumitru Erhan. "deep neural networks for object detection". *Advances in neural information processing systems*, 26, 2013.

[109] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016.

[110] Hironori Takimoto, Hironobu Fukai, Yasue Mitsukura, and Minoru Fukumi. "an analysis of the influence of facial feature for apparent age estimation". In *Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, pages 728–731, 2009.

[111] Matthew A Turk and Alex P Pentland. Face recognition using eigenfaces. In *Proceedings. 1991 IEEE computer society conference on computer vision and pattern recognition*, pages 586–587, 1991.

[112] Kazuya Ueki, Teruhide Hayashida, and Tetsunori Kobayashi. "subspace-based age-group classification using facial images under various lighting conditions". In *7th International Conference on Automatic Face and Gesture Recognition*, pages 6–pp, 2006.

[113] Paul Viola and Michael Jones. "robust real-time object detection". *International journal of computer vision*, pages 34–47, 2001.

[114] Xiaolong Wang, Rui Guo, and Chandra Kambhamettu. "deeply-learned feature for age estimation". In *2015 IEEE Winter Conference on Applications of Computer Vision*, pages 534–541, 2015.

[115] Sanford Weisberg. *"Applied linear regression"*, volume 528. 2005.

[116] Tao Wu, Pavan Turaga, and Rama Chellappa. "age estimation and face verification across aging using landmarks". *IEEE Transactions on Information Forensics and Security*, 7 :1780–1788, 2012.

[117] Bo Xiao, Xiaokang Yang, Yi Xu, and Hongyuan Zha. "learning distance metric for regression by semidefinite programming with application to human age estimation". In *Proceedings of the 17th ACM international conference on Multimedia*, pages 451–460, 2009.

[118] Shuicheng Yan, Ming Liu, and Thomas S Huang. "extracting age information from local spatially flexible patches". In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 737–740, 2008.

[119] Shuicheng Yan, Huan Wang, Thomas S Huang, Qiong Yang, and Xiaoou Tang. "ranking with uncertain labels". In *2007 IEEE international conference on multimedia and expo*, pages 96–99, 2007.

[120] Shuo Yang, Ping Luo, Chen-Change Loy, and Xiaoou Tang. "wider face : A face detection benchmark". In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5525–5533, 2016.

[121] Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, and Yu Qiao. "joint face detection and alignment using multitask cascaded convolutional networks". *IEEE signal processing letters*, 23 :1499–1503, 2016.

[122] Shaohua Kevin Zhou, Bogdan Georgescu, Xiang Sean Zhou, and Dorin Comaniciu. Image based regression using boosting method. volume 1, pages 541–548, 2005.

RESUME

L'âge humain, en tant que caractéristique personnelle importante, peut être directement déduit par des modèles de l'apparence du visage. Grâce aux progrès rapides de l'informatique et de la vision par ordinateur, l'estimation automatique de l'âge via les visages est devenue un sujet révolutionnaire en raison des applications émergentes dans le monde réel.

Nous nous concentrons ici sur l'estimation de l'âge dont l'objectif est de déterminer l'âge spécifique ou la tranche d'âge d'un sujet à partir d'une image faciale.

Nous proposons dans notre système une estimation d'âge à partir des images faciales, établie sur 4 étapes : la détection du visage (les filtre de Haar, MTCNN, YOLO), l'alignement, l'extraction des caractéristiques(LBP) pour arrivé a une estimation de l'âge a partir des réseaux de neurones convolutifs (Keras). L'apprentissage a été réaliser sur deux bases de données : FG-Net et UTKface et nous avons obtenu des résultats satisfaisant.

Mots cles : détection de visages ; MTCNN ; YOLO ; les filtres de Haar ; alignement ; LBP ; CNN ; Keras ; CaffNet.

ABSTRACT

Human age, as an important personal characteristic, can be directly inferred by models of facial appearance. With rapid advances in computer science and computer vision, automatic age estimation via faces has become a revolutionary topic due to emerging real-world applications.

Here, we focus on age estimation whose objective is to determine the specific age or age range of a subject from a facial image.

We propose in our system an age estimation from facial images, based on 4 steps : face detection(Haar filter, MTCNN, YOLO), alignment, feature extraction (LBP) to arrive at an age estimation from neural networks (Keras). The learning was carried out on two databases : FG-Net and UTKface and we obtained satisfactory results.

Key words : Face detection ; Haar filter ; MTCNN ; YOLO ; feature extraction ; LBP ; Keras ; CNN ; CaffeNet.

الـمـلـخص

يمكن الاستدلال على عمر الإنسان، باعتباره خاصية شخصية مهمة، مباشرة من خلال نماذج مظهر الوجه. بفضل التقدم السريع في الحوسبة ورؤية الكمبيوتر، أصبح التقدير التلقائي للعمر عبر الوجوه موضوعًا ثوريًا بسبب التطبيقات الناشئة في العالم الحقيقي.

نركز هنا على تقدير العمر الذي يهدف إلى تحديد العمر أو الفئة العمرية المحددة لموضوع ما من صورة الوجه.

نقترح في نظامنا تقديرًا للعمر من صور الوجه، تم إنشاؤه على 4 خطوات: اكتشاف الوجه (مرشحات Haar، MTCNN، YOLO)، المحاذاة، استخراج الميزات (LBP) للوصول إلى تقدير العمر من الشبكات العصبية التلافيفية (Keras).