

الجمهورية الجزائرية الديمقراطية الشعبية

République Algérienne Démocratique et Populaire

Ministère de l'enseignement supérieur et de la recherche scientifique

Université de 8 Mai 1945 – Guelma -

Faculté des Mathématiques, d'Informatique et des Sciences de la matière

Département d'Informatique



Mémoire de Fin d'études Master

Filière : Informatique

Option : Science et technologie de l'information et de la communication

Thème :

Détection de la couleur de peau par Réseaux de Neurones à Convolution

Encadré Par :

**Dr. BENCHERIET Chemesse
Ennehar**

Présenté par :

MABREK Zahia

Octobre 2020

Remerciement

Tout d'abord, je remercie Dieu Tout-Puissant de m'avoir donné la volonté et la persévérance pour réaliser ce travail.

Je voudrais saisir cette occasion et adresser mes sincères remerciements et appréciation à **Madame chemmse ennahar BENCHRIET** et **Madame yamina BORDJIBA**, mes encadreurs de mémoire, Pour leur précieux soutien et leurs conseils tout au long de mes recherches.

Je tiens également à remercier les membres du jury pour l'intérêt qu'ils portent à mes recherches en acceptant d'examiner mon travail et de l'enrichir de leurs propositions.

Je tiens à adresser mes plus sincères remerciements aux personnes qui m'ont apporté leur aide et qui ont contribué au développement de cette mémoire ainsi qu'à la réussite de cette année universitaire.

Je remercie tous nos enseignants pour leur dévouement, leur patience et leur contribution à notre formation.

Enfin, j'adresse mes plus sincères remerciements à tous mes proches et amis, qui m'ont toujours encouragé lors de la préparation de ce mémoire.

Merci à tous et à toutes.

Dédicaces

Je dédie ce travail à ma mère, qui a fait de l'amour du savoir une vraie croyance pour nous, et elle a toujours été un véritable soutien.

Je dédie ce travail à mon père, "rahimahou allah", qui m'a recommandé d'étudier jusqu'au dernier jour de sa vie, et c'était le meilleur exemple pour moi dans la vie. Quoique je décrive, les mots ne remplissent pas mes remerciements aux parents.

À mes sœurs qui m'ont constamment encouragé et qui ont cru en moi.

À mon petit frère Qaussay, Ishaq et Afnan.

À Hasna, qui m'a aidé à terminer mon travail.

À mes meilleurs amis et aux meilleurs jours universitaires.

Résumé

La couleur de la peau humaine a attiré beaucoup d'attention par rapport à d'autres caractéristiques humaines, en raison de son importance et de son efficacité dans de nombreux domaines de vision artificielle dans divers applications tel que l'interaction homme-machine, la récupération des données dans les bases multimédias, la sécurité, la médecine, ...etc.

L'objectif de notre projet est de concevoir un système intelligent capable de détecter dans une image arbitraire RVB toutes les régions ayant les caractéristiques de la couleur de la peau humaine.

Aucune condition n'a été posé sur les images testées soit : éclairage variable, environnement fermé ou ouvert, background a complexité variable, nombres d'individu présent dans les images variables et à différentes échelles.

Notre système est fondé principalement sur les réseaux de neurones convolutionels, principalement Lenet-5 et VGG-16 un choix justifié par la comparaison entre un réseau pré-entraîné utilisant le 'transfert learning' (VGG-16) et un réseau dont l'entraînement est effectué par nos soins du début à la fin.

Les trois réseaux ont été entraîné sur 8944 échantillons de la base SFA où Nous avons obtenu des précisions de 93 % avec le réseau Lenet-5 ,86% avec le réseau VGG-16 et précision de 96 % avec le VGG-16 pré-entraîné. Les tests finaux ont été réalisés sur des images réelles , avec un environnement à complexité variable , issue des bases de données BAO et SFA . Nous avons obtenu des résultats très satisfaisants avec le réseau Lenet-5 , avec un taux de bonne détection de 95% parmi les images de test de la base BAO.

Mots-clés : détection de peau, CNN, LeNet5, VGG-16, transfer learning.

Abstract

Human skin color has attracted a lot of attention compared to other human characteristics, because of its importance and effectiveness in many areas of artificial vision in various applications such as human-computer interaction, data retrieval in multimedia databases, security, medicine, ...etc.

The objective of our project is to design an intelligent system capable of detecting in an arbitrary RGB image all the regions having the characteristics of the human skin color.

No conditions were placed on the images tested : variable lighting, closed or open environment, background with variable complexity, number of individuals present in the images variable and at different scales. Our system is mainly based on convolutional neural networks, mainly Lenet-5 and VGG-16, a choice justified by the comparison between a pre-trained network using transfer learning (VGG-16) and a network which is trained by us from scratch.

The three networks were trained on 8944 samples from the SFA database, where we obtained accuracies of 93% with the Lenet-5 network, 86% with the VGG-16 network and an accuracy of 96% with the pre-trained VGG-16. The final tests were performed on real images, with variable complexity environment, from BAO and SFA databases. We have obtained very encouraging results with the Lenet-5 network, achieving a good detection rate of 95% among the test images in the BAO database.

Keywords : détection de peau, CNN, LeNet5, VGG-16, transfer learning.

ملخص

جذب لون بشرة الإنسان الكثير من الاهتمام مقارنة بخصائص الإنسان الأخرى ، وذلك لأهميته وفعالته في العديد من مجالات الرؤية الاصطناعية في تطبيقات مختلفة مثل التفاعل بين الإنسان والحاسوب ، واسترجاع البيانات في قواعد بيانات الوسائط المتعددة ، والأمن ، والطب ، ... إلخ.

الهدف من مشروعنا هو تصميم نظام ذكي قادر على الكشف في صورة RGB عشوائية عن جميع المناطق التي لها خصائص لون جلد الإنسان. لم يتم وضع شروط على الصور المختبرة: إضاءة متغيرة ، بيئة مغلقة أو مفتوحة ، خلفية ذات تعقيد متغير ، عدد الأفراد الموجودين في الصور متغير وبمقاييس مختلفة

يعتمد نظامنا بشكل أساسي على الشبكات العصبية التلافيفية ، و بشكل أساسي على Lenet-5 و VGG-16 ، وهو خيار مبرر من خلال المقارنة بين شبكة مُدربة مسبقاً باستخدام نقل التعلم (VGG-16) وشبكة يتم تدريبها بواسطتنا من البداية. تم تدريب الشبكات الثلاث على 8944 عينة من قاعدة بيانات SFA ، حيث حصلنا على دقة تبلغ 93% مع شبكة Lenet-5 و 86% مع شبكة VGG-16 ودقة 96% مع VGG-16 المدربة مسبقاً. تم إجراء الاختبارات النهائية على صور حقيقية ، مع بيئة متغيرة التعقيد ، من قواعد بيانات و SFA. لقد حصلنا على نتائج مشجعة للغاية مع شبكة Lenet-5 ، وحققنا معدل اكتشاف جيد بنسبة 95% بين صور الاختبار في قاعدة بيانات BAO.

كلمات مفتاحية: .transfer learning ، VGG-16 ، LeNet5 ، CNN ، détection de peau

Table des matières

Introduction générale	1
Chapitre 1 : État de l'art	3
1 Introduction	3
2 La peau humaine	3
3 Application de la détection de la peau	4
4 Définition	4
5 Défis de la détection de la couleur de la peau	4
6 Benchmarks utilisés en détection de peau	9
6.1 Base de données Compaq	9
6.2 Base de données ECU	10
6.3 Base de données MCG	10
6.4 Base de données schmugge	11
6.5 Base de données HGR	11
6.6 Base de données UCI	12
6.7 Base de données SFA	12
6.8 Base de données UCD	12
6.9 Base de données BAO	13
6.10 Base de données Caltech	13
7 Méthodes de détection de la peau	14

8	Travaux récents sur la détection de la couleur de la peau	15
8.1	Travaux basés sur les méthodes classiques	16
8.2	Travaux basés sur le deep learning	18
9	Conclusion	21
Chapitre 2 : Réseaux de Neurones Convolutionnels		22
1	Introduction	22
2	Neurone biologique	22
3	Réseaux de neurones artificiels	22
3.1	Composants d'un Réseaux de neurones artificiels	23
3.2	Sructure d'interconnexion	24
4	Apprentissage automatique	25
5	Apprentissage profond	26
6	Types d'approches d'apprentissage profond	26
6.1	Apprentissage supervisé	27
6.2	Apprentissage semi-supervisé	27
6.3	Apprentissage non supervisé	28
6.4	Apprentissage par transfert	28
7	Réseau de neurones a convolution (CNN)	28
7.1	Structure globale d'un CNN	29
7.2	complexité du modèle	34
7.3	Difficultés des CNN	34
7.4	Architectures CNN	35
8	Conclusion	38
Chapitre 3 : Conception		39
1	Introduction	39

2	Architecture générale du système	39
3	Chargement de l'image	41
4	Prétraitement	41
5	Balayage de l'image	41
6	Redimensionnement	42
7	Réseau CNN	43
7.1	Architecture de réseau Lenet-5	44
7.2	Architecture VGG-16	44
7.3	Apprentissage du réseauVGG-16 avec transfert	46
8	Seuillage	46
9	Post-traitement	48
10	Extraction de l'image peau	49
11	Conclusion	49
Chapitre 4 : Implémentation		50
1	Introduction	51
2	Environnement de développement	51
2.1	Google colaboratory	51
2.2	Type d'exécution GPU	51
3	Langage de programmation et bibliothèques	52
3.1	Python	52
3.2	bibliothèques utilisés	52
4	Base d'apprentissage	53
5	Bases de test	53
6	Implémentation de l'architecture	54
6.1	Module apprentissage	54

6.2	Module segmentation	58
7	Interface de système	64
8	Conclusion	66
	Conclusion générale	67

Table des figures

- 1.1 La peau humaine 3
- 1.2 Détection de la peau 5
- 1.3 Exemple de variation d'éclairage 5
- 1.4 Exemple de diversité ethnique 6
- 1.5 Exemple d'image avec arrière plans complexe 6
- 1.6 Exemple d'image avec maquillage 7
- 1.7 Exemple d'image avec Montage 7
- 1.8 Le vieillissement 8
- 1.9 Les problèmes de la détection de la peau 8
- 1.10 La faible qualité des images dans compaq 9
- 1.11 Exemples de la base ECU 10
- 1.12 Exemples de la base MCG 10
- 1.13 Exemples de la base schmugge 11
- 1.14 Exemples de la base HGR 11
- 1.15 Exemples de la base SFA 12
- 1.16 Exemples de la base UCD 13
- 1.17 Exemples de la base BAO 13
- 1.18 Exemples de la base Caltech 14
- 1.19 Les méthodes de détection de la peau 15

2.1	Neurone biologique	23
2.2	Mise en correspondance neurone biologique/neurone artificiel	23
2.3	Types des réseaux de neurones	25
2.4	Apprentissage profond dans le contexte de l'intelligence artificielle	27
2.5	Réseau de neurones avec de nombreuses couches convolutives	29
2.6	Exemple explicative sur l'opération de convolution	30
2.7	pooling moyen	32
2.8	pooling maximal	32
2.9	Fonctions d'activation	33
2.10	Taxonomie des architectures CNN	35
3.1	Détection de peau	39
3.2	Architecture de notre système	40
3.3	Image avec amélioration d'éclairage	41
3.4	Illustration de l'effet du pas de la fenêtre sur la qualité de segmentation	42
3.5	Balayage et redimensionnement de l'imagette a l'issue de la fenêtre d'analyse	43
3.6	Architecture générale du réseau lenet-5	44
3.7	Architecture générale du réseau VGG-16	45
3.8	Apprentissage par transfert	46
3.9	Affichage de l'image résultante de la segmentation	48
3.10	Amélioration de l'image binaire par fermeture	48
3.11	Amélioration de l'image binaire par ouverture	49
3.12	Affichage de l'image peau	49
4.1	Illustration sur les résultats précédents	56
4.2	Résultats du réseau Lenet-5	57
4.3	Résultats du réseau VGG-16	57

4.4	Résultats du réseau VGG-16 pré-entraîné	58
4.5	Image d'une personne avec arrière-plan simple	59
4.6	Image d'une personne avec arrière-plan simple	60
4.7	Image d'une personne avec arrière-plan simple	61
4.8	Image d'une personne avec arrière-plan complexe	62
4.9	Image d'une personne avec mauvaise qualité	62
4.10	Image de plusieurs personnes	63
4.11	Interface initiale de notre application	64
4.12	Fenêtre principale de détection	65
4.13	Fenêtres des paramètres de traitements	66

Liste des tableaux

1.1	Banchmarks de peau	14
1.2	les travaux récents sur la détection de la peau	20
2.1	Types de padding	31
2.2	Complexité du modèle	34
3.1	Caractéristiques de la fenêtre de balayage	42
3.2	Tableau détaillé sur les du réseau Lenet-5	45
3.3	Tableau représentatif de la configuration du VGG16	47
4.1	La répartition de la base SFA	54
4.2	Les résultats du test final	64

Introduction générale

Introduction générale

La peau étant l'une des parties les plus importantes du corps humain, il est donc logique de la voir comme élément principal à détecter dans de nombreux systèmes de la vision artificielle opérant sur les êtres humains tels que : la médecine pour la détection et la reconnaissance de maladies, la sécurité pour la détection d'intrusions, l'identification de personnes, la reconnaissance faciale, l'analyse des gestes, le suivi des mains ...etc.

La tâche de détection de la peau semble très facile et simple à l'œil humain mais elle reste une opération d'une grande complexité pour la machine en raison de nombreux facteurs tel que : les conditions d'éclairage où l'image a été capturée, cela dépend également des caractéristiques de la caméra telles que la résolution et la qualité, la variation de l'arrière-plan et le chevauchement entre le couleur de la peau dans le cas de personnes trop blanches ou trop noires ... etc.

De nombreuses recherches ont été effectuées pour améliorer le processus de détection de la peau, mais cela nécessite une grande précision en raison du bruit et de la complexité des images. Ces méthodes ne parviennent pas à résoudre définitivement ces problèmes. Pour cela il existe plusieurs méthodes dans ce domaine parmi lesquelles des travaux, basés sur des méthodes classiques telles que : entropie floue, modèle de couleur de peau gaussienne ... etc. qui sont basées sur la couleur et la texture de la peau. Il existe également des méthodes basées sur l'intelligence artificielle telles que les réseaux de neurones, la machine à vecteur de support et l'apprentissage profond.

L'objectif principal de notre travail est de concevoir un modèle avec une architecture d'apprentissage en profondeur, et d'implémenter un modèle de réseau de neurones convolutionnel pour la détection de la peau, donc notre travail basé sur deux architectures CNN, l'une de ces architectures est moins profonde et l'autre plus profonde, puis utilise la technique d'apprentissage par transfert, alors que ces modèles doivent se renforcer avec des pré et post-traitements, et évaluer les classificateurs obtenus avec un algorithme de segmentation .

Notre mémoire est organisée en quatre chapitres comme suit :

Chapitre 1 : État de l'art

Dans ce chapitre nous présentons les notions de base et les difficultés de détection de la peau, nous décrivons les recherches les plus récentes et les bases de données les plus utilisés dans ce domaine.

Chapitre 2 : Réseaux de neurones convolutionels

Dans ce chapitre, nous nous sommes concentrés sur notre approche de recherche pour laquelle est l'apprentissage en profondeur et les réseaux de neurones convolutionels ainsi que les types et architectures les plus connus.

Chapitre 3 : Conception

Dans ce chapitre nous avons détaillé l'architecture globale de notre système de détection de la peau et les principales phases de sa conception.

Chapitre 4 : Implémentation

Dans ce chapitre, nous avons abordé l'aspect implémentation de notre application, qui consiste à affiner les concepts précédemment développés ainsi que l'environnement de développement, les détails des tests, les résultats obtenus et leur interprétation.

Chapitre 1

État de l'art

1 Introduction

La peau est le plus grand organe chez l'homme, la couleur de la peau est la caractéristique la plus importante qui distingue une personne des autres choses, et c'est ce qui la rend importante dans de nombreux domaines. La détection de la peau est une étape essentielle dans de nombreux systèmes d'interaction homme-machine tel que : le télé-enseignement, la sécurité, la communication,...etc, elle consiste en l'extraction des zones contenant la peau dans une image numérique. Dans ce contexte, des recherches ont été menées sur des moyens efficaces afin d'obtenir les résultats les plus corrects possibles et de réduire au minimum les erreurs.

2 La peau humaine

La peau humaine est le revêtement extérieur du corps et est le plus grand organe du système tégumentaire. La peau contient jusqu'à sept couches de tissu ectodermique et protège les muscles, les os, les ligaments et les organes internes sous-jacents. La peau humaine varie dans les tons de couleur du brun foncé au rose clair ou blanc. Elle varie à cause d'un certain nombre de facteurs comme l'exposition au soleil ou la différence entre les races humaines. La rugosité de la peau varie entre les hommes et les femmes. La couleur de la peau varie également selon la région géographique et les changements climatiques.

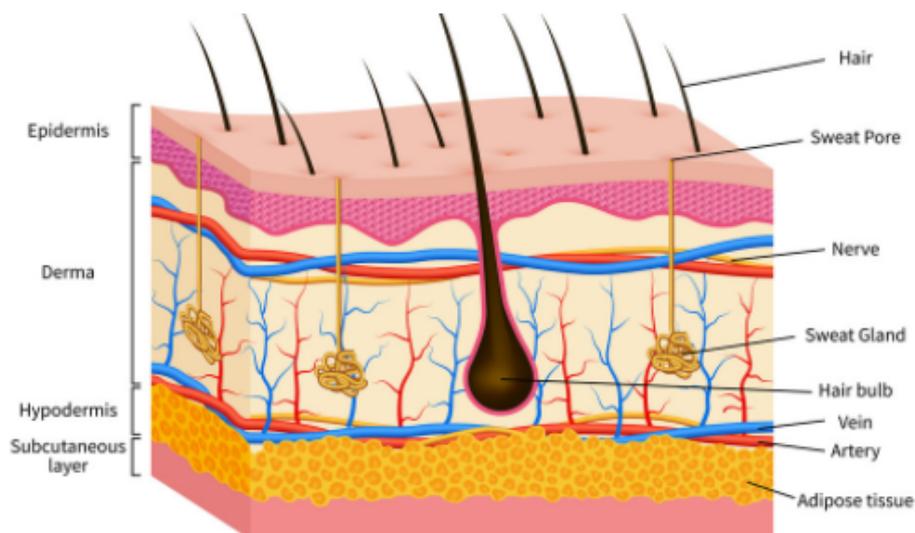


Figure 1.1 – La peau humaine

3 Application de la détection de la peau

La détection de la peau est une étape importante et primordiale dans plusieurs domaines de la vision par ordinateur :

- C'est une première étape pour Connaître les zones de présence humaine
- la reconnaissance faciale
- le détections des mouvements de personnes dans une vidéo
- l'identification des maladies de la peau
- la reconnaissance des races et des âges pour la classification
- la reconnaissance des empreintes digitales
- ... etc

4 Définition

La détection de la peau est le processus de recherche de pixels et de régions de couleur de la peau dans une image ou une vidéo. Ce processus est généralement utilisé comme étape de prétraitement pour trouver des régions qui sont potentiellement des visages et des membres humains dans les images. Plusieurs approches de vision par ordinateur ont été développées pour la détection de la peau. Un détecteur de peau transforme généralement un pixel donné en un espace colorimétrique approprié [Bencheriet et al., 2007], puis utilise un classificateur de peau pour étiqueter le pixel, qu'il s'agisse d'un pixel de peau ou non. Un classificateur de peau définit une limite de décision de la classe de couleur de peau dans l'espace colorimétrique sur la base d'une base de données d'apprentissage de pixels de couleur de peau [Bencheriet, 2019][Elgammal et al., 2009].

5 Défis de la détection de la couleur de la peau

Ce processus est exposé à des problèmes difficiles et nombreux, ils résultent de la conversion d'images à partir de la réalité continue au monde numérique discret qui provoque la perte d'informations, et avec la diversité des types d'images et des sources, la couleur de la peau hu-

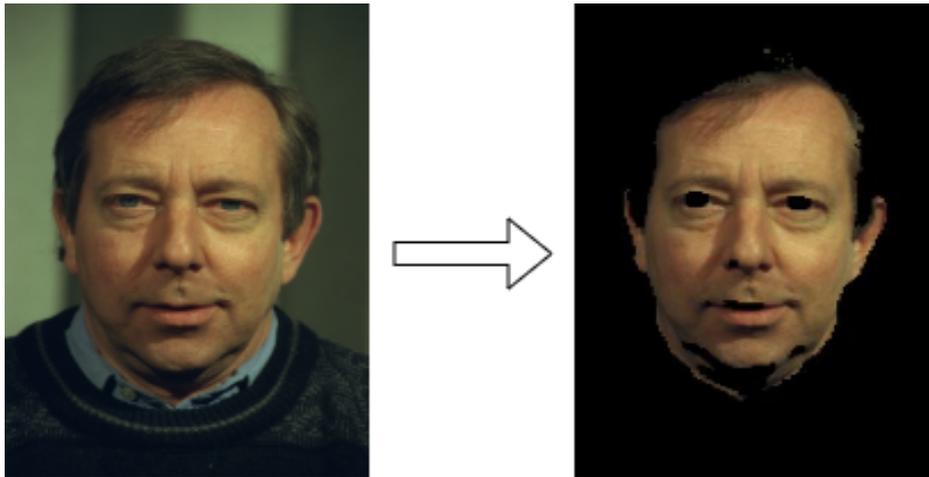


Figure 1.2 – Détection de la peau

maine peut varier considérablement en apparence qui fait de la détection précise de la peau une tâche difficile, parmi lesquels nous citons : Manque ou augmentation de l'éclairage, Faiblesse de la résolution. Environnement complexe, le bruit, le contraste ,...etc.[Naji et al., 2019]

- Variations d'éclairage : la variation d'éclairage est le problème le plus important qui dégrade sérieusement les performances de la segmentation. La figure 1.3 montre la différence d'éclairage.



Figure 1.3 – Exemple de variation d'éclairage

- Différents groupes ethniques (races) : L'apparence de la couleur de la peau varie d'une personne à l'autre en raison des différences physiques entre les groupes raciaux humains. La figure 1.4 montre les différents races humaines



Figure 1.4 – Exemple de diversité ethnique

- Arrière-plan complexe : La diversité des arrière plans ,par exemple, meubles, vêtements, cheveux blonds, cailloux, etc. La diversité des arrière-plans est pratiquement illimitée. Cela provoque le détecteur de peau à produire de fausses détections comme le montre la figure 1.5.



Figure 1.5 – Exemple d'image avec arrière plans complexe

- Maquillage :Il affecte l'apparence de la couleur de la peau.La figure 1.6 montre une personne avec et sans maquillage



Figure 1.6 – Exemple d'image avec maquillage

- Montage et reproduction d'images : Certaines images ont déjà été capturées avec l'utilisation de filtres de couleur comme le montre la figure 1.7. Cela rend le traitement des informations sur les couleurs encore plus difficile.



Figure 1.7 – Exemple d'image avec Montage

- Vieillessement : La peau humaine varie d'une peau fraîche et élastique à une peau sèche et rugueuse avec des rides. La figure 1.8 montre la différence de la peau dépend de l'âge.



Figure 1.8 – Le vieillissement

- Conditions d'imagerie : des facteurs tels que les caractéristiques de l'appareil photo (réponse du capteur, lentilles) affectent l'apparence de la peau. En général, différentes caméras couleur ne produisent pas nécessairement les mêmes apparences de couleur pour la même scène sous le même conditions d'imagerie

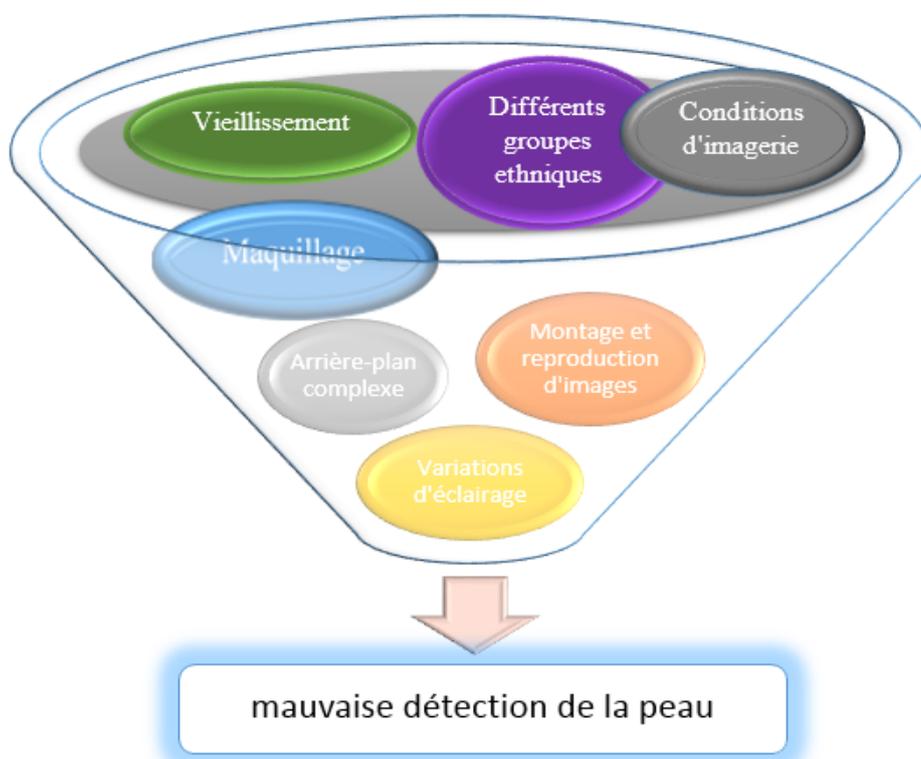


Figure 1.9 – Les problèmes de la détection de la peau

6 Benchmarks utilisés en détection de peau

Afin de surmonter autant que possible les problèmes de détection de peau, une grande variété de méthodes a été développée au cours des deux dernières décennies. En examinant différentes enquêtes sur la classification de la peau, l'un des principaux points conclus est que non seulement il est impossible d'évaluer précisément les performances des détecteurs eux-mêmes, mais il est également impossible de les comparer équitablement en raison de l'indisponibilité d'une base de données standard et unifié.

La plupart des méthodes de détection de la peau nécessitent un grand nombre d'échantillons de peau et non-peau pour les phases d'entraînement et de test. En général, il est recommandé d'utiliser une base de données de test standard pour que les chercheurs puissent comparer directement les résultats. Alors que la plupart des recherches ont utilisé leurs propres base de données (c'est-à-dire collecter des images à partir de différentes sources), il existe des bases de données d'images de peau actuellement très connues :

6.1 Base de données Compaq

cette base de données Se compose de 13 640 photos. Ces photos sont classées en deux groupes, à savoir : photos avec peau et non-peau. Compaq est la première grande base de données de peau et peut-être la base de données la plus largement utilisée, comprenant 9731 images contenant des pixels de peau et 8965 images sans pixels de peau. la base de données entière comprend environ 2 milliards de peau et non-peau collectés par l'exploration de pages Web. cette base de données n'est plus disponible pour un usage public. compaq est une base de données relativement ancienne et la qualité de ses images est trop faible comme le montre la figure 1.10.[Jones and Rehg, 2002]



Figure 1.10 – La faible qualité des images dans compaq

6.2 Base de données ECU

La base de données de peau et de visage de l'ECU est compilé sur la base de près de 4000 images couleur de haute qualité et de données de vérité terrain relativement précises pour l'analyse comparative directe des algorithmes de segmentation de la peau. Les images assurent la diversité des scènes d'arrière-plan, des conditions d'éclairage et des types de peau [Phung et al., 2005]. La figure 1.11 montre quelques exemples de cette base de données.



Figure 1.11 – Exemples de la base ECU

6.3 Base de données MCG

La base de données MCG-skin contient 1000 images échantillonnées au hasard à partir de sites Web de réseaux sociaux capturés dans une lumière ambiante variable, des arrière-plans déroutants, la diversité des races humaines ainsi que diverses résolutions et une qualité visuelle commele montre la figure 1.12. Les images de vérité au sol ne sont pas étiquetées avec précision, car les yeux, les sourcils, et même les bracelets sont également considérés comme les pixels de la peau autour des bords ne sont pas également marqués.[Huang et al., 2011]

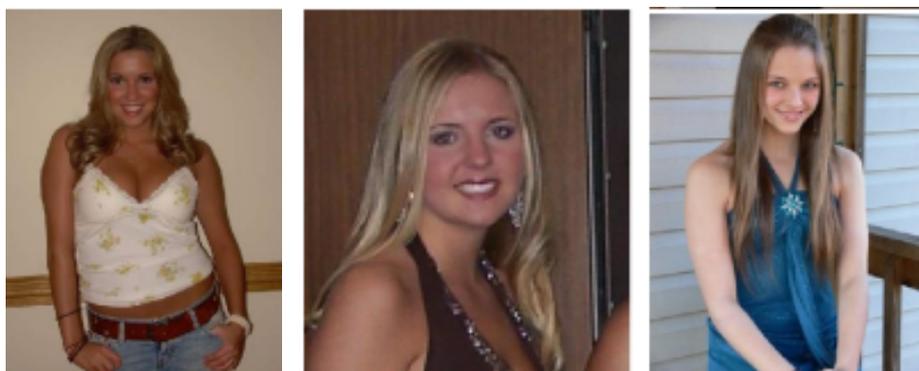


Figure 1.12 – Exemples de la base MCG

6.4 Base de données schmugge

La base de données de peau fourni par [Schmugge et al., 2007] se compose de 845 images, cette base de données est très générale car elle contient des images avec différentes expressions faciales, niveaux d'éclairage et calibrages de caméra. Cet ensemble de données est trop petit pour être utilisé pour les étapes de formation et d'évaluation de la plupart des méthodes. Des exemples de cette base sont présenté dans la figure 1.13



Figure 1.13 – Exemples de la base schmugge

6.5 Base de données HGR

Cette base de données est développée principalement pour la reconnaissance des gestes de la main. Elle est organisée en trois séries acquises dans des conditions différentes et comprenant au total 1558 images. Cette base de données ne contient que des images de mains et les arrière-plans ne sont pas du tout complexes.[Kawulok et al., 2013]



Figure 1.14 – Exemples de la base HGR

6.6 Base de données UCI

Le référentiel d'apprentissage automatique UCI se compose de pixels de peau collectés par échantillonnage aléatoire des valeurs R, V, B à partir d'images de différents groupes d'âge, races et sexes dérivés des bases de données FERET et PAL. le nombre des échantillons d'apprentissage total est 245057. Cette base de données n'est applicable qu'à des fins de formation et uniquement pour des méthodes particulières.[Bay et al., 2000]

6.7 Base de données SFA

A été construit sur la base d'images de visage de bases de données FERET (876 images) et AR (242 images), à partir desquelles des échantillons de peau et non-peau et les différents couleurs de la peau ont été récupérés différemment de 18 dimensions différentes(figure 1.15) [Casati et al., 2013].



Figure 1.15 – Exemples de la base SFA

6.8 Base de données UCD

Cette base de données contient deux parties, la première partie contient 100 images en couleurs de visages avec des variations en arrière-plan, composants structurels du visage, poses, orientation, expressions faciales, conditions d'imagerie zones occluses, âge, sexe, race et taille. la partie 2 de la base de données des visages UCD contient des résultats segmentés à la main pour chaque image de la base de données [Sharma and Reilly, 2003]. La figure 1.16 montre des exemples de cette base de données.



Figure 1.16 – Exemples de la base UCD

6.9 Base de données BAO

Cette base de données est destinée à la détection des visages, collectant des images couleur avec un degré élevé de variabilité capturées à l'aide d'appareils photo CCD, d'appareils photo numériques, du World Wide Web, d'images de sites Web de journaux et de collections de photos personnelles.

Ces images recueillies sont dans des conditions d'éclairage variables et avec des arrière-plans complexes. En outre, les images contiennent plusieurs visages avec des variations de couleur [Wang et al., 2008], comme le montre la figure 1.17.

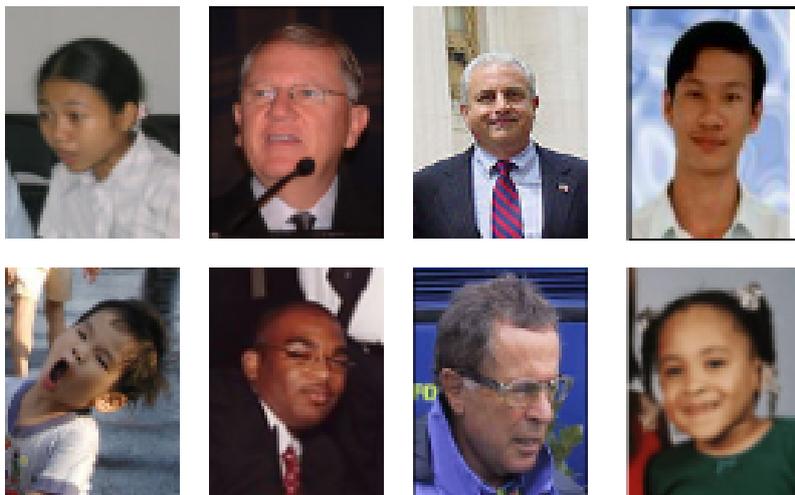


Figure 1.17 – Exemples de la base BAO

6.10 Base de données Caltech

La base de données contient des images de personnes collectées sur le Web. L'ensemble de données contient 10 524 visages humains de différentes résolutions et dans différents contextes, des images de portrait, des groupes de personnes [Angelova et al., 2005]. La figure monte des exemples de la base caltech.



Figure 1.18 – Exemples de la base Caltech

La table 1.1 présente un récapitulatif des différentes bases de données de la peau décrites dans ce mémoire.

la base de données	année	taille(image)	référence
COMPAQ	2002	13 640	[Jones and Rehg, 2002]
ECU	2005	4000	[Phung et al., 2005]
MCG	2011	1000	[Huang et al., 2011]
Schmugge	2007	845	[Schmugge et al., 2007]
HGR	2012	1558	[Kawulok et al., 2013]
UCI	2000	245057(RGB valeur)	[Bay et al., 2000]
SFA	2013	1118	[Casati et al., 2013]
UCD	2003	299	[Sharma and Reilly, 2003]
BAO	2008	370	[Wang et al., 2008]
Caltech	2005	10000	[Angelova et al., 2005]

Table 1.1 – Banchmarks de peau

7 Méthodes de détection de la peau

les méthodes de détection de la couleur de la peau sont classées sur la figure 1.19, selon le type de segmentation et l'approche utilisée. Cette classification est proposé par [Naji et al., 2019]

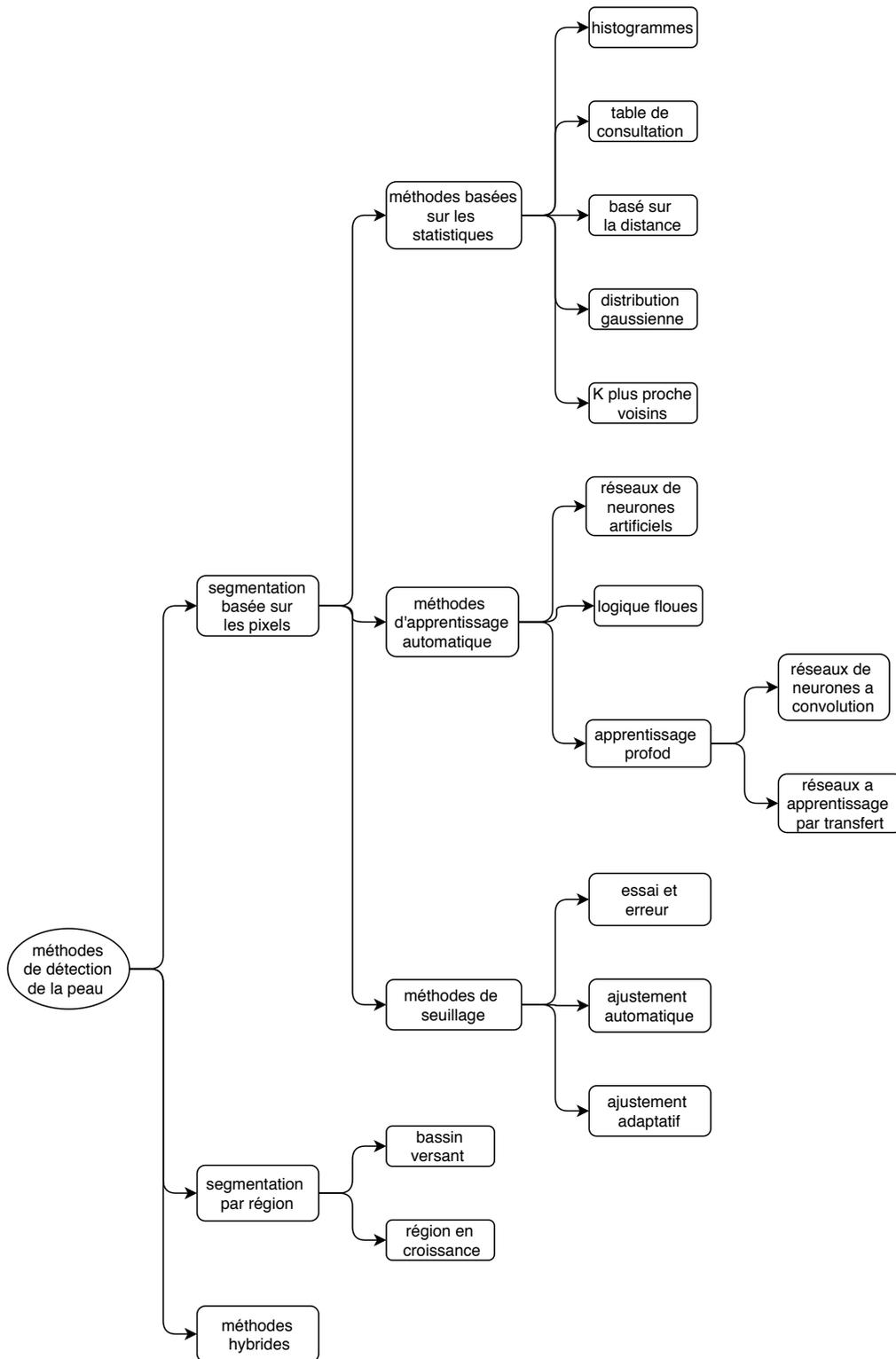


Figure 1.19 – Les méthodes de détection de la peau [Naji et al., 2019]

8 Travaux récents sur la détection de la couleur de la peau

L'étude bibliographique que nous avons élaborée sur les travaux les plus récents de détection de la peau a été partagée en deux parties : les travaux utilisant les méthodes classiques (voir figure

1.19) et les travaux basés sur l'apprentissage en profondeur (deep-learning).

8.1 Travaux basés sur les méthodes classiques

Nous présentons dans cette partie un bref état de l'art sur les travaux les plus récents utilisant les méthodes classiques :

“Face Detection using Skin Color Modeling and Geometric Feature”

[Verma et al., 2014]

Les régions de la peau sont extraites à l'aide du modèle de couleur de peau gaussienne dans l'espace Cb-Cr et la méthode du rapport de vraisemblance est utilisée pour créer un masque binaire. Le modèle de couleur de peau a été conçu en utilisant une combinaison de deux bases de données différentes, pour englober des teintes de peau plus grandes. Un total de 165 images faciales de la base de données Caltech ont été sélectionnées au hasard pour évaluer les performances de la méthode proposée ; une précision de 95% y a été obtenue.

“Pixel-based skin detection based on statistical models”

[Nadian-Ghomsheh, 2016]

les auteurs de ce travail ont proposé d'appliquer les distributions Gamma, Beta et Laplace pour modéliser les pixels de couleur de peau dans des espaces de chromaticité arbitraires utilisés pour la détection paramétrique de peau. Étant donné que les distributions proposées ne tiennent pas compte de manière inhérente de la corrélation entre les composants de chromaticité, une méthode pour éliminer la corrélation entre les informations de chrominance de la peau est également proposée. Cela permet la modélisation de la peau sans se soucier de la corrélation des données. Ils modélisent les pixels de couleur de peau en appliquant les distributions proposées dans cinq espaces colorimétriques différents. La base de données Compaq [Jones and Rehg, 2002] a été utilisée pour évaluer les performances de la méthode proposée. La précision de la détection de la peau sur la base de données Compaq était de 88% et a montré une amélioration par rapport aux méthodes statistiques précédentes. Cette méthode n'a pas dépassé le reste des méthodes en termes d'exactitude et de réussite.

”A Multi-Scale Approach to Skin Pixel Detection ” [Roheda, 2017]

Cette recherche présente un algorithme pour détecter les pixels de la peau dans une

image. Chaque pixel est classé en tant que pixel peau ou non peau en fonction des caractéristiques extraites de son voisinage. L'algorithme présenté utilise un rapport de vraisemblance modifié pour la classification et utilise une approche multi-échelles pour classer le pixel en question. L'algorithme a été développé et évalué à l'aide du jeu de données ColorFERET [Phillips et al., 2000]. Il a atteint une précision de classification de 95,6%.

“Robust Skin Segmentation Using Color Space Switching ”

[Gupta and Chaudhary, 2016]

Cette recherche vise à améliorer considérablement la détection de la peau dans une image couleur numérique en utilisant la commutation automatisée de l'espace colorimétrique. Un système avec trois algorithmes robustes a été construit sur la base de différents espaces colorimétriques vers une classification automatique de la peau dans une image 2D. Ces algorithmes sont basés sur la moyenne statistique de la valeur des pixels de la peau dans l'image. Ils adoptent également des approches bayésiennes pour faire la distinction entre les pixels peau et non-peau pour éviter le bruit. Ce travail est testé sur environ 500 images qui ont été capturées manuellement dans des conditions d'éclairage variées à partir d'une webcam et d'un appareil photo numérique standard.

“Face Detection Based on Skin Color Segmentation Using Fuzzy Entropy”

[Pujol et al., 2017] Dans cette recherche, Francisco and Pujol ont proposé un système flou pour détecter la peau dans les images en couleur, de sorte que chaque tonalité de couleur est supposée être un ensemble flou. Les systèmes de couleurs rouge, vert et bleu (RVB), la teinte, la saturation et la valeur (HSV) et le YCbCr (où Y est la luminance et Cb, Cr sont les composantes de chrominance) sont utilisés pour le développement de leur conception floue. Ainsi, une approche d'entropie floue à trois partitions est utilisée pour calculer tous les paramètres nécessaires pour les systèmes flous, puis, une méthode de détection de visage est également développée pour valider les résultats de la segmentation. Les résultats des expériences montrent une détection correcte de la peau d'un taux compris entre 94% et 96% pour ces méthodes de segmentation floue, avec un taux de faux positifs d'environ 0,5% dans tous les cas.

8.2 Travaux basés sur le deep learning

Alors que les algorithmes d'apprentissage automatique traditionnels sont linéaires, les algorithmes d'apprentissage profond sont empilés dans une hiérarchie de complexité et d'abstraction croissantes. Il surmonte les limites des approches traditionnelles d'apprentissage automatique. [Indolia et al., 2018]

Avec des progrès dans tous les domaines des logiciels et du matériel et avec l'avènement du 'Big Data', il est devenu possible de travailler sur ce type d'algorithme et d'obtenir des résultats impressionnants. Dans ce qui suit nous présentons les travaux les plus importants dans le domaine de détection de la peau utilisant l'apprentissage a profond.

“Convolutional neural networks and training strategies for skin detection”

[Kim et al., 2017]

Cette recherche présente deux réseaux de neurones convolutifs (convolutional neural network) et leurs stratégies d'entraînement pour la détection de la peau . Le premier CNN, composé de 20 couches de convolution avec des filtres 3x3, est une sorte de réseau VGG. Le second est composé de 20 couches réseau en réseau (Network In Network) qui peuvent être considérées comme une modification de la structure de démarrage. Lors de l'apprentissage de ces réseaux pour la détection de la peau humaine, ils envisagent un apprentissage basé sur des patches et des images complètes. La première méthode se concentre sur les caractéristiques locales telles que la couleur et la texture de la peau, et la seconde sur les caractéristiques de forme liées à l'homme ainsi que la couleur et la texture. Les expériences montrent que les CNN proposés donnent de meilleures performances que les méthodes conventionnelles et également que les méthodes existantes basée sur l'apprentissage en profondeur. De plus, Ils ont constaté que la structure NiN présente généralement une précision plus élevée que la structure basée sur VGG. Les expériences montrent également que l'apprentissage basé sur l'image entière qui apprend les caractéristiques de la forme donne une meilleure précision que l'apprentissage basé sur les patches qui se concentre uniquement sur la couleur et la texture locales.

“Combining Convolutional and Recurrent Neural Networks for Human Skin Detection ” [Zuo et al., 2017]

Dans ce travail, les auteurs ont proposé d'intégrer des couches de réseaux neuronaux récurrents (RNN) dans des réseaux neuronaux entièrement convolutifs (FCN) et de dé-

velopper un réseau de bout en bout pour la détection de la peau humaine. En particulier, les couches entièrement connectés capturent les caractéristiques locales génériques, tandis que les couches RNN modélisent le réseau sémantique des dépendances contextuelles dans les images. Les résultats expérimentaux montrent que l'algorithme FCN et RNN proposé est plus performant que les méthodes standard testés sur les bases de données COMPAQ et ECU [Phung et al., 2005]. RNN améliorent efficacement la stabilité de l'algorithme de détection de la peau sous des arrière-plans complexes. Le taux de détection sur la base de données COMPAQ est 95.93 %, et sur ECU est 98.10 %.

"Skin Identification Using Deep Convolutional Neural Network"

[Oghaz et al., 2019]

Pour combler les lacunes des détecteurs de la peau classiques basés sur la texture, cette recherche propose un modèle d'identification de peau utilisant des réseaux de neurones convolutifs profonds. Les expériences ont été menées sur deux bases de données : FSD Benchmark [Mazumdar and Bora, 2016] ainsi qu'une base de données interne sur la texture de la peau. Les performances proposées pour le modèle d'identificateur de peau profond ont été comparées aux techniques classiques basées sur la texture, y compris le spectre de puissance de Fourier, les filtres de Gabor, la matrice de concurrence de niveau de gris, les modèles binaires locaux à plusieurs blocs et la transformation discrète du cosinus et ont produit des résultats supérieurs en termes de précision : Les résultats montrent que l'identifiant de peau proposé offre une précision maximale de 0,932 et une perte minimale de 0,224 et a surpassé les méthodes basées sur la texture existantes.

Le tableau 1.2 résume les méthodes et les travaux sur la détection de peau que nous avons étudié.

Référence	Méthode	BDD	Précision	Année	nombre de citations
méthodes classiques					
[Verma et al., 2014]	modele gaussienne	caltech	95%	2014	12
[Nadian-Ghomsheh, 2016]	distributions Gama, beta, laplace	COMPAQ	88%	2016	3
[Roheda, 2017]	caractéristiques des voisins	ColorFERET	95%	2017	3
[Gupta and Chaudhary, 2016]	bayésiennes multi espace couleur	500 image manuellement capturées	-	2016	12
[Pujol et al., 2017]	entropie floue	XM2VTS	94% - 96%	2017	31
Apprentissage profond					
[Kim et al., 2017]	(CNNxVGG-16) network in network	ECU	87%	2017	14
		Pratheepan	90%		
[Zuo et al., 2017]	RNN+CNN	COMPAQ	95%	2017	68
		ECU	98%		
[Oghaz et al., 2019]	CNN	FSD + base interne	93%	2019	-

Table 1.2 – les travaux récents sur la détection de la peau

9 Conclusion

La détection (segmentation) de la peau humaine est restée un sujet d'étude et de recherche non encore résolu à cause de la spécificité de cet organe qui est une combinaison à la fois de couleur et de texture.

L'utilisation des réseaux de neurones convolutifs et l'apprentissage profond s'est arrivée une solution optimale vu les résultats promoteurs obtenus.

Chapitre 2

Réseaux de Neurones Convolutionnels

1 Introduction

Une personne peut distinguer automatiquement beaucoup de choses, et cela est dû à la vitesse de collecte et de déduction des informations. Ces caractéristiques chez l'homme ont inspiré les scientifiques à créer un modèle similaire au réseau neuronal humain et à l'utiliser dans divers domaines. Les réseaux neuronaux ont souvent été appliqués pour traiter des problèmes de vision par ordinateur et de traitement d'image tels que la classification et la reconnaissance de formes d'une façon générale. Parmi les points forts de ces méthodes leur résistance au bruit et leur capacité à traiter des données de grande dimension. L'apprentissage des caractéristiques à partir de données est considéré comme une méthode plausible pour remédier aux limites des fonctionnalités artisanales.

2 Neurone biologique

Le système nerveux contient 10 à la puissance de 12 neurones interconnectés. Bien que les neurones ne soient pas tous identiques, leur forme et certaines caractéristiques leur permettent d'être divisés en quelques grandes classes. Il est également important de savoir que tous les neurones ne se comportent pas de la même manière en fonction de leur position dans le cerveau. la figure 2.1 présente un neurone ou cellule nerveuse spécialisé dans le traitement de l'information [Brette, 2003].

Un neurone biologique contient :

- Le corps cellulaire
- Les dendrites
- L'axone
- La synapse

3 Réseaux de neurones artificiels

Les réseaux de neurones biologique réalisent un certain nombre des tâches facilement tel que la reconnaissance des formes , le traitement de signal , l'apprentissage ...etc. Les réseaux de neurones artificiels sont des réseaux informatiques qui tentent de simuler, de manière grossière, le processus de décision dans les réseaux de cellules nerveuses (neurones) du système nerveux

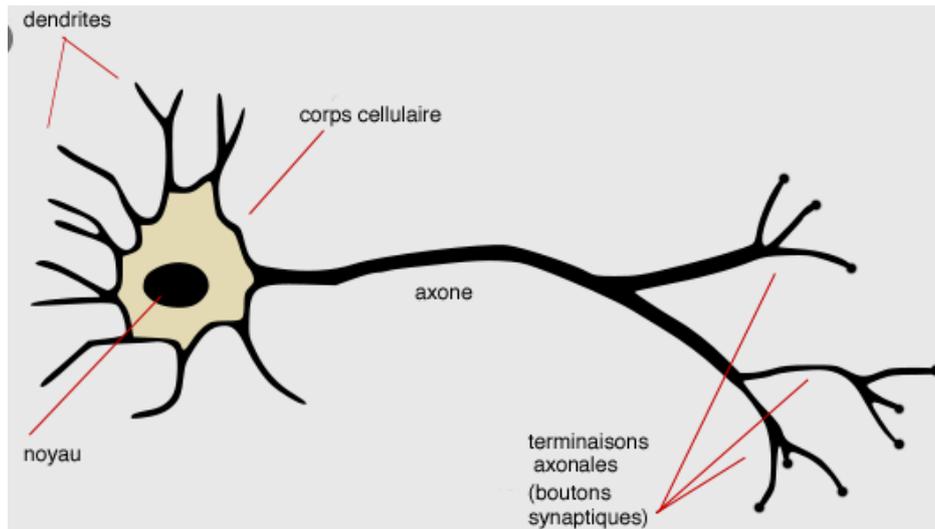


Figure 2.1 – Neurone biologique [3]

central biologique (humain ou animal) [Graupe, 2013](figure 2.2).

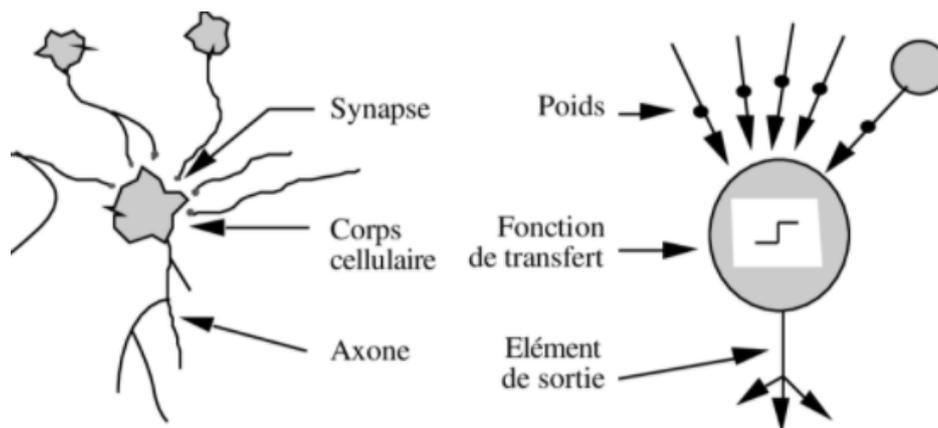


Figure 2.2 – Mise en correspondance neurone biologique/neurone artificiel [3]

3.1 Composants d'un Réseaux de neurones artificiels

Un neurone artificiel est une unité de calcul qui reçoit un certain nombre d'entrées directement de l'environnement ou de quantités de neurones. Lorsque l'information est émise, on associe des poids pour chaque entrée, il y a ensuite une fonction de somme entre ces paires (entrée \times poids) afin de décider de la sortie du neurone à l'aide d'une fonction d'activation [Yegnanarayana, 2009].

Un neurone artificiel est donc composé de :

- Un ensemble des valeurs d'entrée A_1, A_2, \dots, A_n

- Un ensemble des poids réels W_1, W_2, \dots, W_n
- Une fonction de sommation Σ qui calcule la somme des paires (A, W) .
- Une fonction d'activation f qui décide l'état du neurone (active/non active).

3.2 Structure d'interconnexion

Les connexions entre les neurones qui composent le réseau décrivent la topologie du modèle. Il y a plusieurs nouvelles architectures de réseaux de neurones qui apparaissent de temps en temps, mais le plus souvent il est possible de distinguer une certaine régularité. La figure 2.3 décrit les types les plus connus de connexions de réseaux neuronaux.

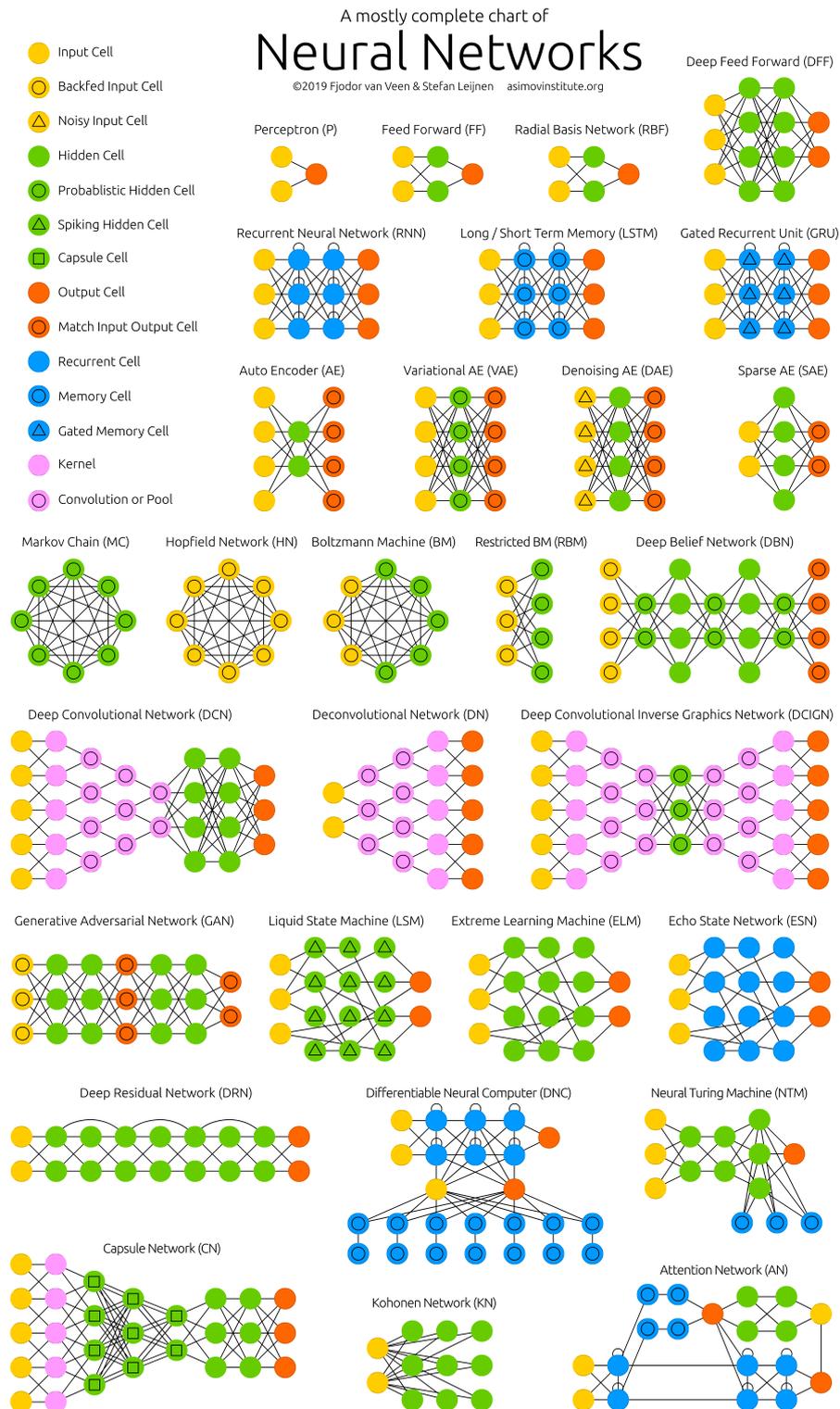


Figure 2.3 – Types des réseaux de neurones [15]

4 Apprentissage automatique

L'apprentissage automatique est l'étude des algorithmes informatiques qui s'améliorent automatiquement grâce à l'expérience. Il est considéré comme un sous-ensemble

de l'intelligence artificielle [Alpaydin, 2020].

La technologie d'apprentissage automatique alimente de nombreux aspects de la société moderne : des recherches sur le Web au filtrage de contenu sur les réseaux sociaux en passant par les recommandations sur les sites Web de commerce électronique, et elle est de plus en plus présente dans les produits de consommation tels que les appareils photo et les smartphones.

Les systèmes d'apprentissage automatique sont utilisés pour identifier les objets dans les images, transcrire la parole en texte, faire correspondre les actualités, les publications ou les produits avec les intérêts des utilisateurs et sélectionner les résultats de recherche pertinents.

Les techniques d'apprentissage automatique conventionnelles étaient limitées dans leur capacité à traiter les données naturelles sous leur forme brute [LeCun et al., 2015].

5 Apprentissage profond

L'apprentissage profond est une forme d'apprentissage automatique qui permet aux ordinateurs d'apprendre de l'expérience et de comprendre le monde en termes de hiérarchie de concepts. Parce que l'ordinateur recueille des connaissances à partir de l'expérience, il n'est donc pas nécessaire pour un opérateur informatique humain de spécifier formellement toutes les connaissances nécessaires à l'ordinateur. La hiérarchie des concepts permet à l'ordinateur d'apprendre des concepts complexes en les construisant à partir de concepts plus simples, un graphique de ces hiérarchies aurait plusieurs couches de profondeur [Goodfellow et al., 2016].

6 Types d'approches d'apprentissage profond

Tout comme l'apprentissage automatique, les approches d'apprentissage en profondeur peuvent être classées comme suit : supervisées, semi-supervisées ou partiellement supervisées et non supervisées. En outre, il existe une autre catégorie d'apprentissage appelée apprentissage par renforcement (RL) ou Deep RL (DRL) qui sont souvent abordées dans le cadre d'approches d'apprentissage semi-supervisé ou parfois non supervisé [Alom et al., 2018].

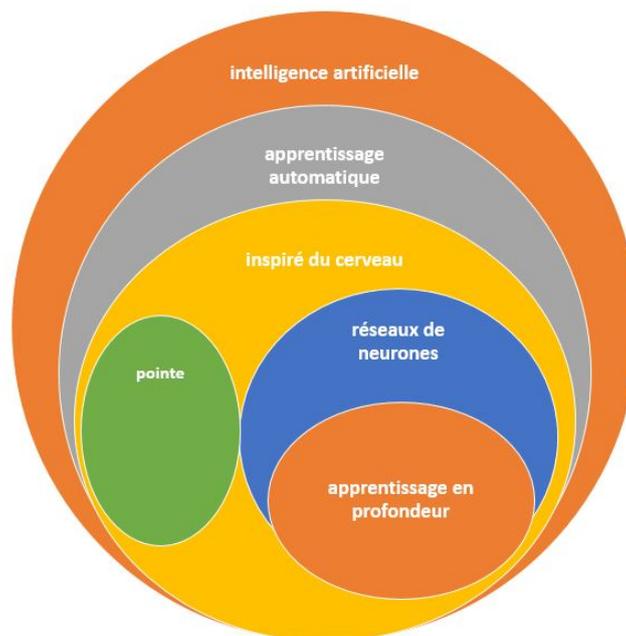


Figure 2.4 – Apprentissage profond dans le contexte de l'intelligence artificielle [Sze et al., 2017]

6.1 Apprentissage supervisé

L'apprentissage supervisé est une technique d'apprentissage qui utilise des données étiquetées. Dans le cas des approches DL supervisées, l'environnement dispose d'un ensemble d'entrées et de sorties correspondantes. Il modifiera alors itérativement les paramètres du réseau pour une meilleure approximation des sorties souhaitées.

Il existe différentes approches d'apprentissage supervisé pour le deep learning, notamment les réseaux de neurones profonds (DNN), les réseaux de neurones convolutifs (CNN), les réseaux de neurones récurrents (RNN), y compris la mémoire à long terme (LSTM) et les unités récurrentes fermées (GRU) [Alom et al., 2018].

6.2 Apprentissage semi-supervisé

L'apprentissage semi-supervisé est un apprentissage basé sur des ensembles de données partiellement étiquetés (souvent également appelés apprentissage par renforcement). Dans certains cas, les DRL et les Réseaux Adversaires Génératifs (GAN) sont utilisés comme techniques d'apprentissage semi-supervisé. De plus, RNN, y compris LSTM et GRU, est également utilisé pour l'apprentissage semi-supervisé [Alom et al., 2018].

6.3 Apprentissage non supervisé

Les systèmes d'apprentissage non supervisés sont ceux qui peuvent le faire sans la présence d'étiquettes de données. Dans ce cas, l'agent apprend la représentation interne ou des fonctionnalités importantes pour découvrir des relations ou une structure inconnues dans les données d'entrée.

Le regroupement, la réduction de la dimensionnalité et les techniques génératives sont souvent considérés comme des approches d'apprentissage non supervisé.

Il y a plusieurs membres de la famille du deep learning qui sont bons pour le clustering et la réduction de dimensionnalité non linéaire, y compris. Encodeurs automatiques (AE), Machines Boltzmann restreintes (RBM) et le GAN récemment développé. En outre, les RNN, tels que LSTM(Long-Short Term Memory neural network) et RL, sont également utilisés pour l'apprentissage non supervisé dans de nombreux domaines d'application - [Arulkumaran et al., 2017].

6.4 Apprentissage par transfert

Alors que les données sont un élément essentiel de la création du réseau, l'idée de l'apprentissage par transfert a contribué à réduire les demandes de données. L'apprentissage par transfert est le processus consistant à prendre un modèle pré-entraîné (les poids et les paramètres d'un réseau qui a été formé sur une grande base de données par quelqu'un d'autre) et à **affiner** le modèle avec notre propre base de données. L'idée est que ce modèle pré-entraîné agira comme un extracteur de fonctionnalités. nous supprimerons la dernière couche du réseau et la remplacerons par notre propre classificateur. nous figeons ensuite les poids de toutes les autres couches et formons le réseau normalement (geler les couches signifie ne pas changer les poids lors de la descente /optimisation du gradient) [Hon and Khan, 2017].

7 Réseau de neurones a convolution (CNN)

Les CNN sont l'un des meilleurs algorithmes d'apprentissage pour comprendre le contenu des images et ont montré des performances exemplaires dans les tâches liées à la segmentation, la classification, la détection et l'extraction d'images -[CireşAn et al., 2012] [Ciresan et al., 2012] [Liu et al., 2019].

Les CNN peuvent également être utilisés pour des applications deep learning dans les soins de santé, telles que l'imagerie médicale. Ainsi, CNNs sont des réseaux hiérarchiques multicouches, où chaque couche, utilisant une banque de noyaux convolutifs, effectue plusieurs transformations [LeCun et al., 2010].

L'opération de convolution aide à l'extraction de fonctionnalités utiles à partir de points de données corrélés localement. La sortie des noyaux convolutifs est ensuite affectée à l'unité de traitement non linéaire (fonction d'activation), qui non seulement aide à apprendre les abstractions, mais intègre également la non-linéarité dans l'espace des fonctionnalités. Cette non-linéarité génère différents modèles d'activations pour différentes réponses et facilite ainsi l'apprentissage des différences sémantiques dans les images [Khan et al., 2020].

7.1 Structure globale d'un CNN

La caractéristique intéressante de CNN est sa capacité à exploiter la corrélation spatiale ou temporelle des données. La topologie de CNN est divisée en plusieurs étapes d'apprentissage composées d'une combinaison des couches convolutives, des unités de traitement non linéaires et des couches de sous-échantillonnage [Jarrett et al., 2009]

[Khan et al., 2020]. La figure 2.2 montre la structure générale d'un réseau CNN.

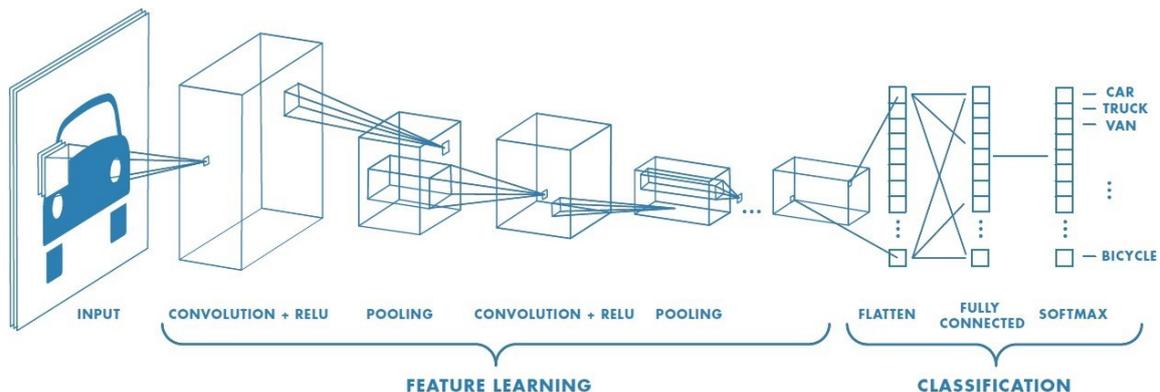


Figure 2.5 – Réseau de neurones avec de nombreuses couches convolutives [11]

Couche de convolution

La convolution est la première couche à extraire des entités d'une image d'entrée. La convolution préserve la relation entre les pixels en apprenant les caractéristiques de l'image à l'aide de petits carrés de données d'entrée. Il s'agit d'une opération mathématique qui prend deux entrées

telles qu'une matrice d'image et un filtre ou un noyau [Liu et al., 2013] [Indolia et al., 2018]. La figure 2.6 monte une opération simple d'un filtre pour une étape de convolution.

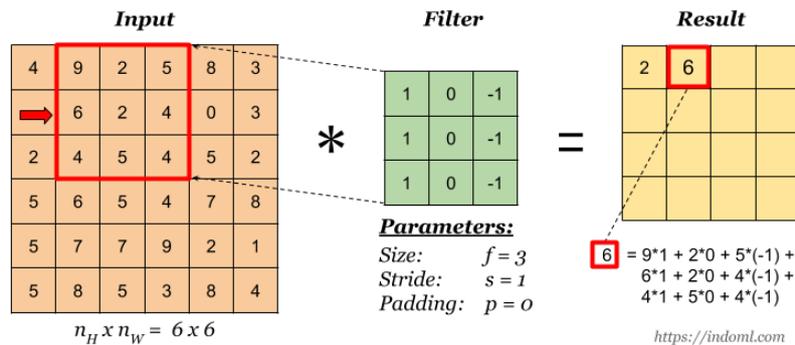


Figure 2.6 – Exemple explicative sur l'opération de convolution [13]

Paramètres du filtre

La couche convolutionnelle contient des filtres pour lesquels il est important de savoir comment ajuster ses paramètres.

- **Dimensions d'un filtre** : Un filtre de taille $F \times F$ appliqué à une entrée contenant C canaux est un volume de taille $F \times F \times C$ qui effectue des convolutions sur une entrée de taille $I \times I \times C$ et qui produit une carte de caractéristiques de sortie (aussi appelé activation map) de taille $O \times O \times 1$ (avec un seul filtre). appliquer K filtres de taille $F \times F$ engendre un feature map de sortie de taille $O \times O \times K$.
- **Le pas (Stride)** : Dans le contexte d'une opération de convolution ou de pooling, le stride S est un paramètre qui dénote le nombre de pixels par lesquels la fenêtre se déplace après chaque opération.
- **Zéro-remplissage (Zero-padding)** est une technique consistant à ajouter P zéros à chaque côté des frontières de l'entrée. Cette valeur peut être spécifiée soit manuellement, soit automatiquement par le biais d'une des configurations les types de padding sont dans le tableau 2.1

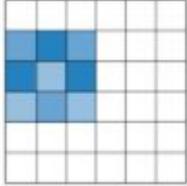
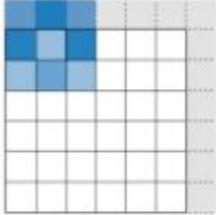
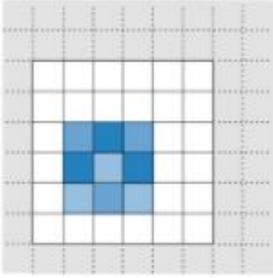
Configuration	Valide	Pareil	Total
Valeur	$P = 0$	$P_{\text{start}} = \left\lfloor \frac{S \left\lceil \frac{I}{S} \right\rceil - I + F - S}{2} \right\rfloor$ $P_{\text{end}} = \left\lceil \frac{S \left\lfloor \frac{I}{S} \right\rfloor - I + F - S}{2} \right\rceil$	$P_{\text{start}} \in [0, F - 1]$ $P_{\text{end}} = F - 1$
Illustration			
But	<ul style="list-style-type: none"> • Pas de padding • Enlève la dernière opération de convolution si les dimensions ne collent pas 	<ul style="list-style-type: none"> • Le padding tel que la feature map est de taille $\left\lceil \frac{I}{S} \right\rceil$ • La taille de sortie est mathématiquement satisfaisante • Aussi appelé 'demi' padding 	<ul style="list-style-type: none"> • Padding maximum tel que les dernières convolutions sont appliquées sur les bords de l'entrée • Le filtre 'voit' l'entrée du début à la fin

Table 2.1 – Types de padding [12]

Couche de pooling

La couche pooling généralement placée entre deux couches de convolution, cette couche permet de réduire le nombre de sous-échantillonnage lorsque les images sont trop grandes. Pooling également appelé “subsampling” ou “downsampling” réduit la dimensionnalité de chaque carte mais conserve des informations importantes [Khan et al., 2020]. C’est une opération simple qui consiste à remplacer un carré de pixels (généralement 2x2 ou 3x3) par une seule valeur selon le type de pooling, Il existe différents types de pooling :

- **pooling moyen** :il prend la moyenne de tous les pixels de la sélection figure 2.7.

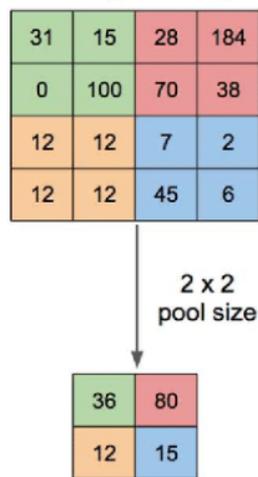


Figure 2.7 – pooling moyen

- **pooling maximal** :il prend le pixel qui a la valeur maximale entre tous les pixels de la sélection

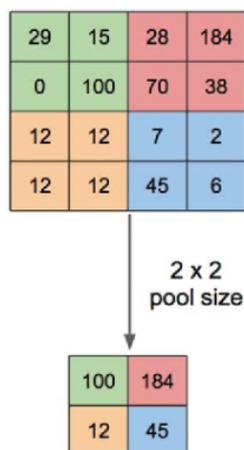


Figure 2.8 – pooling maximal

Couche fully-connected

La couche entièrement connectée est similaire au réseau entièrement connecté dans les modèles conventionnels. La sortie de la première phase (comprenant la convolution et la mise en commun répétitive) est introduite dans la couche entièrement connectée, et le produit scalaire du vecteur de poids et du vecteur d'entrée est calculé afin d'obtenir la sortie finale . [Indolia et al., 2018]

Couche de correction

Il existe une vaste littérature qui utilise la fonction d'activation sigmoïde dans les algorithmes classiques d'apprentissage automatique. Afin d'introduire la non-linéarité, l'utilisation de l'unité linéaire rectifiée (ReLU) s'est révélée meilleure que la première, en raison de deux facteurs majeurs. Une autre fonction d'activation largement utilisée est la tangens hyperbolicus, ou fonction hyperbolique tangente / tanh. La figure 2.9 montre les différentes fonctions d'activation [Indolia et al., 2018].

le rôle de cette couche est de définir la sortie du neurone, selon la fonction d'activation utilisée le neurone est actif et transmet sa valeur aux autres ou non.

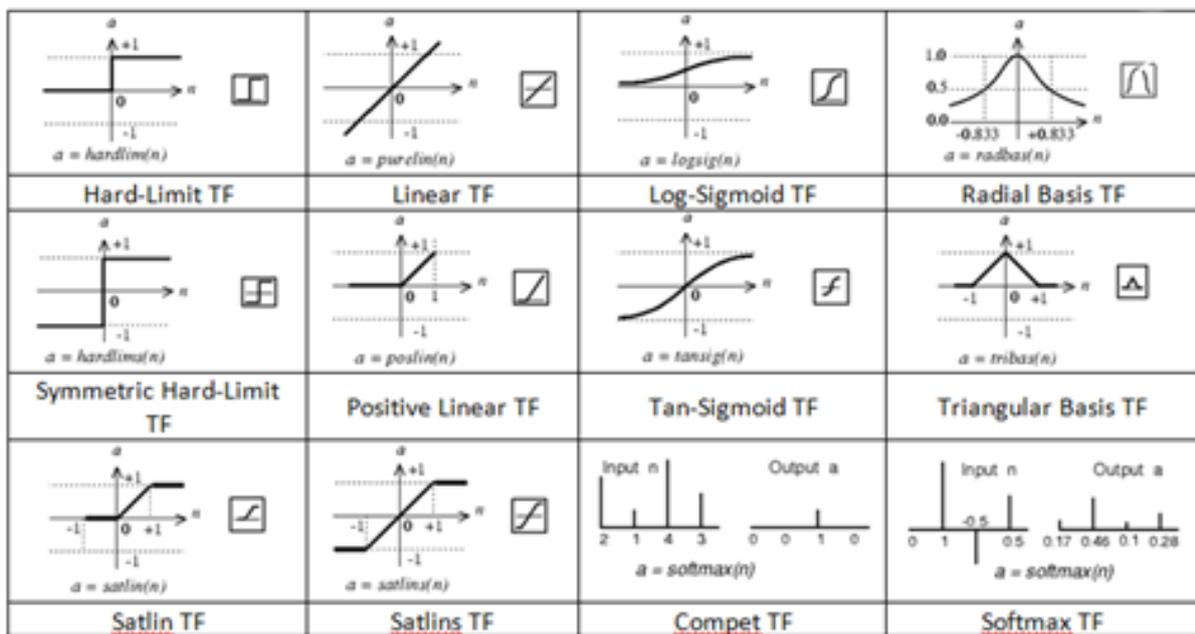


Figure 2.9 – Fonctions d'activation [9]

7.2 complexité du modèle

On définit la complexité du modèle par rapport au nombre de paramètres utilisés. le tableau 2.2 détermine les formules de calculs des paramètres dans chaque couche CNN.

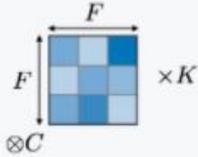
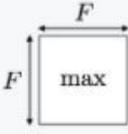
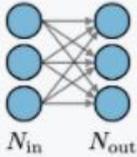
	CONV	POOL	FC
Illustration			
Taille d'entrée	$I \times I \times C$	$I \times I \times C$	N_{in}
Taille de sortie	$O \times O \times K$	$O \times O \times C$	N_{out}
Nombre de paramètres	$(F \times F \times C + 1) \cdot K$	0	$(N_{in} + 1) \times N_{out}$

Table 2.2 – Complexité du modèle [12]

7.3 Difficultés des CNN

L'utilisation des CNN a toujours posé un gros problème, car elle nécessitait des ensembles de données de grande taille, mais avec l'avènement du Big Data, nous avons bénéficié de sa variété et de sa taille. La recherche basée sur le CNN a été un succès retentissant, mais il fallait considérer certaines conditions importantes qui changent complètement les performances de ce modèle.

- L'absence de méthode systématique permettant de définir la meilleure topologie du réseau et le nombre de neurones à placer dans la (ou les) couche(s) cachée(s).
- Le choix des valeurs initiales des poids du réseau et le réglage du pas d'apprentissage, qui jouent un rôle important dans la vitesse de convergence.
- les données d'entraînement (apprentissage) et les données de test (Il y a un manque général de données) .

7.4 Architectures CNN

Différentes améliorations de l'architecture CNN ont été apportées de 1989 à ce jour. Ces améliorations peuvent être classées comme optimisation des paramètres, régularisation, reformulation structurelle, etc. Cependant, il est observé que l'essentiel de l'amélioration des performances de CNN provenait de la restructuration des unités de traitement et de la conception de nouveaux blocs.

La plupart des innovations dans les architectures CNN ont été faites en relation avec la profondeur et l'exploitation spatiale. Selon le type de modification architecturale, CNN peut être largement classé en sept classes différentes à savoir ; exploitation spatiale, profondeur, multi-trajets, largeur, exploitation de la carte des caractéristiques, amplification des canaux et CNN basés sur l'attention. [Khan et al., 2019]

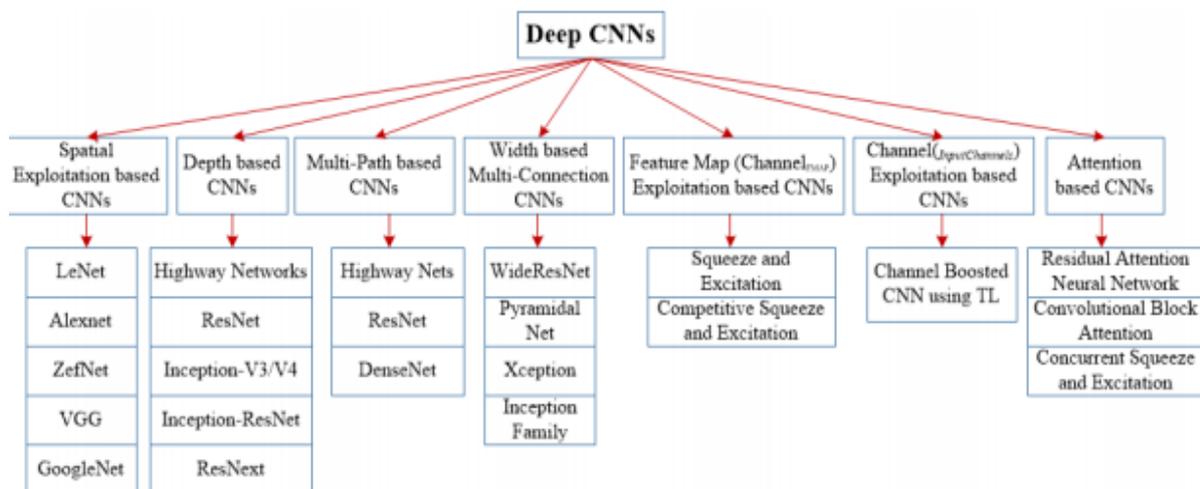


Figure 2.10 – Taxonomie des architectures CNN [Khan et al., 2019]

CNN basés sur l'exploitation spatiale

Comme l'opération convolutive considère le voisinage (localité) des pixels d'entrée, par conséquent, différents niveaux de corrélation peuvent être explorés en utilisant différentes tailles de filtre. Différentes tailles de filtres encapsulent différents niveaux de granularité ; généralement, des filtres de petite taille extraient des informations à grains fins et de grande taille. Par conséquent, au début des années 2000, les chercheurs ont exploité des filtres spatiaux pour améliorer les performances et ont exploré la relation d'un filtre spatial avec l'apprentissage du réseau.

Différentes études menées à cette époque suggèrent que grâce à l'ajustement des filtres, CNN peut bien fonctionner à la fois sur les détails grossiers et fins [Khan et al., 2019].

Malgré que VGG-16 classé dans cette section mais c'est aussi un CNN profond, ce qui lui permet d'être classé parmi les types basés sur la profondeur.

CNN basés sur la profondeur

Les architectures CNN profondes sont basées sur l'hypothèse qu'avec l'augmentation de la profondeur, le réseau peut mieux se rapprocher de la fonction cible avec un certain nombre de mappages non linéaires et des hiérarchies de fonctionnalités plus enrichies [Bengio, 2013]. La profondeur du réseau a joué un rôle essentiel dans le succès de la formation supervisée. Des études théoriques ont montré que les réseaux profonds peuvent représenter certaines classes de fonctions plus efficacement que les architectures superficielles [Montufar et al., 2014]. Inception et VGG, qui ont montré les meilleures performances lors du concours 2014, renforcent encore l'idée que la profondeur est une dimension essentielle dans la régulation de la capacité d'apprentissage des réseaux [Simonyan and Zisserman, 2014] [Szegedy et al., 2015] [Khan et al., 2019].

CNN multi-chemins

Pour l'apprentissage des réseaux profonds, le concept de connectivité multi-chemins ou inter-couches a été proposé. Plusieurs chemins ou connexions de raccourcis peuvent systématiquement connecter une couche à une autre en sautant certaines couches intermédiaires pour permettre le flux spécialisé d'informations à travers les couches. La connectivité croisée partitionne le réseau en plusieurs blocs. Ces chemins tentent également de résoudre le problème du gradient de fuite en rendant le gradient accessible aux couches inférieures. À cette fin, différents types de connexions de raccourci sont utilisés, tels que les connexions à zéro remboursé, basées sur la projection, les abandons, les sauts de connexions et les connexions 1x1, ... etc [Srivastava et al., 2015] [Larsson et al., 2016] [Khan et al., 2019].

CNN à connexions multiples basées sur la largeur

[Kawaguchi et al., 2019] ont indiqué que la largeur du réseau est également importante. Le perceptron multicouche a gagné l'avantage de mapper des fonctions complexes sur le perceptron en utilisant en parallèle plusieurs unités de traitement au sein d'une couche. Cela suggère que la largeur est un paramètre essentiel dans la définition des principes d'apprentissage avec la profondeur.

[Lu et al., 2017a] [Lu et al., 2017b] ont récemment montré que les réseaux de neurones avec la fonction d'activation ReLU doivent être suffisamment larges pour posséder une propriété d'approximation universelle avec une augmentation de la profondeur. De plus, une classe de fonctions continues sur un ensemble compact ne peut pas être arbitrairement bien approximée par un réseau arbitrairement profond, si la largeur maximale du réseau n'est pas supérieure à la dimension d'entrée .

CNN basés sur l'exploitation de Feature-Map (Channel Map)

CNN est devenu populaire pour les tâches de vision par ordinateur en raison de son apprentissage hiérarchique et de sa capacité d'extraction automatique de caractéristiques [LeCun et al., 2010]. La sélection des caractéristiques joue un rôle essentiel dans la détermination des performances des modules de classification, de segmentation et de détection. Dans CNN, les caractéristiques sont sélectionnées dynamiquement en ajustant les poids associés à un noyau également appelé masque. En outre, plusieurs étapes d'extraction d'entités sont utilisées, qui peuvent extraire divers types d'entités (appelées cartes d'entités ou canaux dans CNN).[Khan et al., 2019]

CNN basés sur l'exploitation des canaux (input)

Une bonne représentation de l'image est celle qui peut définir les caractéristiques saillantes d'une image à partir d'un code compact. Dans les tâches de vision par ordinateur, divers types de filtres conventionnels sont appliqués pour extraire différents niveaux d'informations pour un seul type d'image [Dollár et al., 2009]. Ces diverses représentations sont ensuite utilisées comme entrée du modèle pour améliorer les performances. Désormais, CNN est un apprenant de caractéristiques convaincant qui peut extraire automatiquement des caractéristiques discriminantes en fonction du problème . le concept de renforcement de canal (dimension de canal

d'entrée) utilisant des apprenants auxiliaires est introduit dans les CNN pour améliorer la représentation du réseau [Yang et al., 2019] [Khan et al., 2019].

CNN basés sur l'attention

Multiplés hiérarchies d'abstractions se concentrent sur des caractéristiques pertinentes pour le contexte joue également un rôle important dans la localisation et la reconnaissance des images. Dans le système visuel humain, ce phénomène est appelé attention.

Ce processus sert non seulement à focaliser sur des régions sélectionnées, mais déduit également différentes interprétations des objets à cet endroit et aide ainsi à capturer la structure visuelle d'une meilleure manière. Un type d'interprétabilité plus ou moins similaire est ajouté dans RNN et LSTM [Sundermeyer et al., 2012]

[Khan et al., 2019].

8 Conclusion

La classification d'images basée sur le contenu visuel est une tâche très difficile, en grande partie parce qu'il y a généralement une grande quantité de variabilité intra-classe, résultant de différentes conditions d'éclairage, d'un désalignement et de déformations non rigides.

Bien que les fonctionnalités de bas niveau puissent être conçues à la main avec un grand succès pour certaines données et tâches, la conception de fonctionnalités efficaces pour de nouvelles données et tâches nécessite généralement de nouvelles connaissances du domaine, car la plupart des fonctionnalités artisanales ne peuvent pas simplement être adaptées à de nouvelles conditions.

Et c'est là que réside l'importance de l'apprentissage à profond dans la possibilité d'extraire et d'utiliser la plupart des fonctionnalités sans avoir besoin d'experts.

Chapitre 3

Conception

1 Introduction

Dans les applications de vision par ordinateur nécessitant la détection de personnes, la détection de la peau est souvent une étape préliminaire et essentielle. Détecter la peau dans une image RVB revient à vérifier la présence ou non des régions de peau dans cette image et retourner les pixels exacts de la peau détectée si elle existe.

L'objectif de notre travail est d'utiliser les réseaux de neurones à convolutions pour détecter la peau dans un environnement sans conditions externes.

2 Architecture générale du système

L'objectif de notre projet est de concevoir un système intelligent capable de détecter dans une image arbitraire RVB toutes les régions ayant les caractéristiques de la couleur de la peau humaine (figure 3.1).

Aucune condition n'a été posé sur les images testées soit : éclairage variable, environnement fermé ou ouvert, background a complexité variable, nombre d'individu présent dans les images variables et à différentes échelles. Notre système est fondé principalement sur les réseaux de neurones convolutionels, principalement Lenet-5 et VGG-16. Un choix justifié par la comparaison entre un réseau pré-entraîné utilisant le 'transfert learning' (VGG-16) et un réseau dont l'entraînement est effectué par nos soins du début à la fin.

L'architecture globale de notre système est illustrée par la figure 3.2

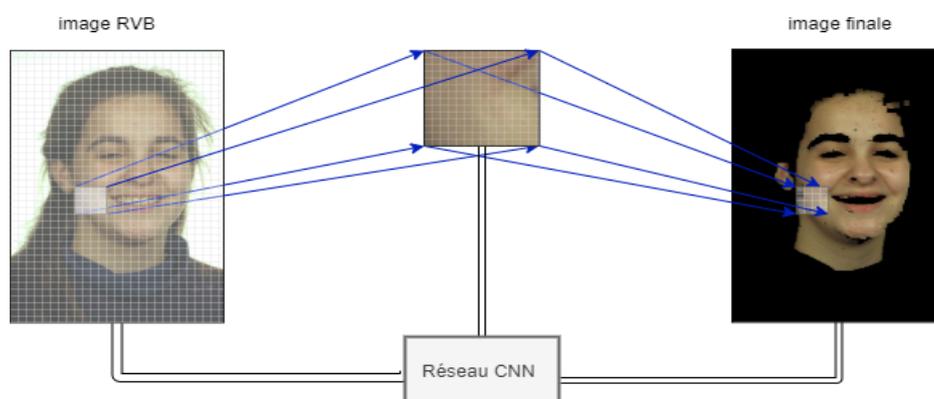


Figure 3.1 – Détection de peau

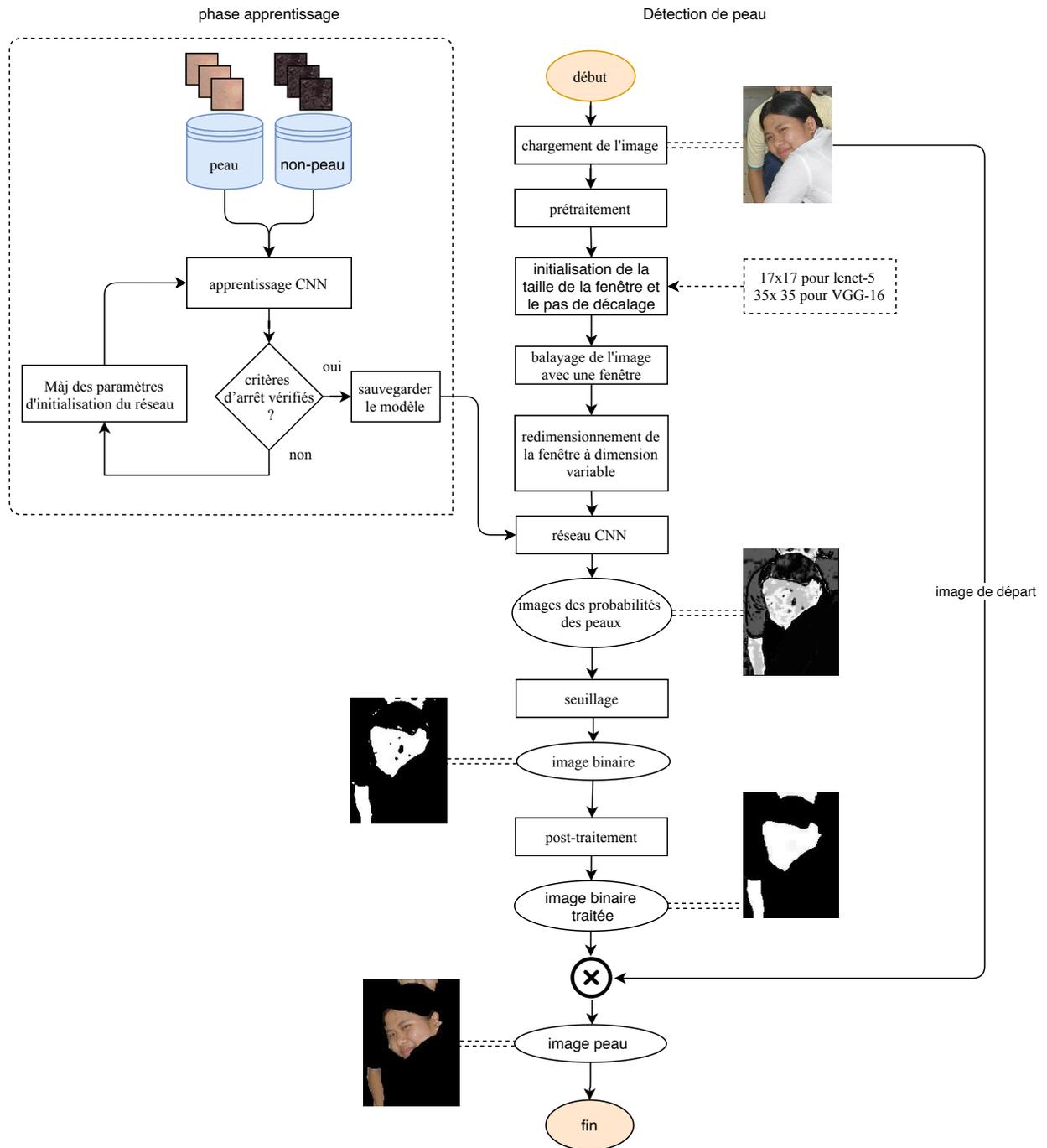


Figure 3.2 – Architecture de notre système

3 Chargement de l'image

Nous avons appliqué notre système de détection de la peau sur une image RVB de taille libre. Pour le cas d'image trop grande un redimensionnement peut être effectué, pour arriver à un temps d'exécution acceptable, le redimensionnement se fait par une interpolation bilinéaire ; la valeur du pixel résultat est la moyenne des pixels de voisinage les plus proches.

4 Prétraitement

A cause de l'éclairage qui peut affecter la couleur de la peau, nous avons prévu une phase de pré-traitement optionnelle afin d'améliorer la qualité des images trop sombre ou trop éclairées.



Figure 3.3 – Image avec amélioration d'éclairage

5 Balayage de l'image

Le réseau CNN choisi va procéder à la classification peau/non peau de l'image d'entrée par le biais d'une fenêtre de taille et de déplacement variable, qui va balayer l'image entière. Le choix de la taille de la fenêtre et le pas de décalage affecte considérablement la qualité de l'image classifiée (figure 3.4). Pour cela et pour garder une qualité acceptable de l'image résultante nous avons fixé les valeurs limites de ces deux paramètres (tableau 3.1).

Réseau CNN	Taille de fenêtre	décalage	Redimensionnement
Lenet-5	2x2 :17x17	1 :f-1	17
VGG16	5x5 :35x35	1 :f-1	35

Table 3.1 – Caractéristiques de la fenêtre de balayage

Où f :la taille de la fenêtre

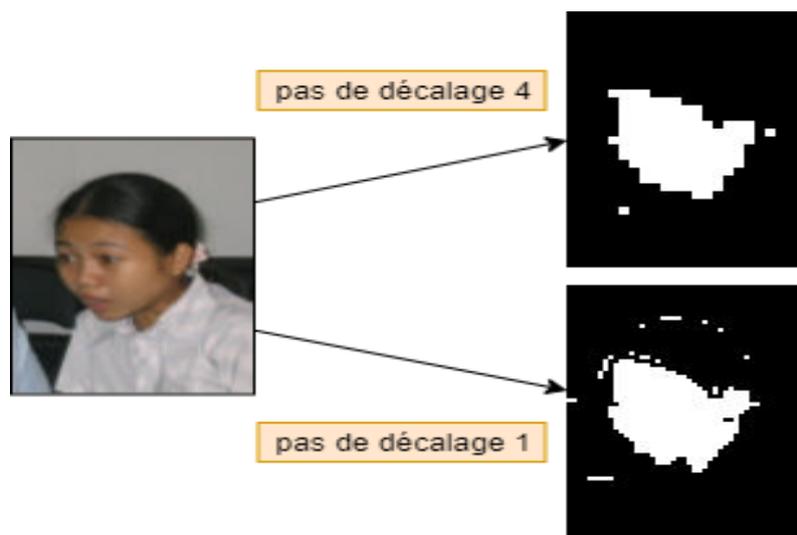


Figure 3.4 – Illustration de l'effet du pas de la fenêtre sur la qualité de segmentation

6 Redimensionnement

L'entrée du réseau CNN étant fixée au début de l'apprentissage, il est donc important de faire un redimensionnement de l'image issue de la fenêtre a la taille du réseau. La figure 3.5 explique le processus de balayage d'une image .

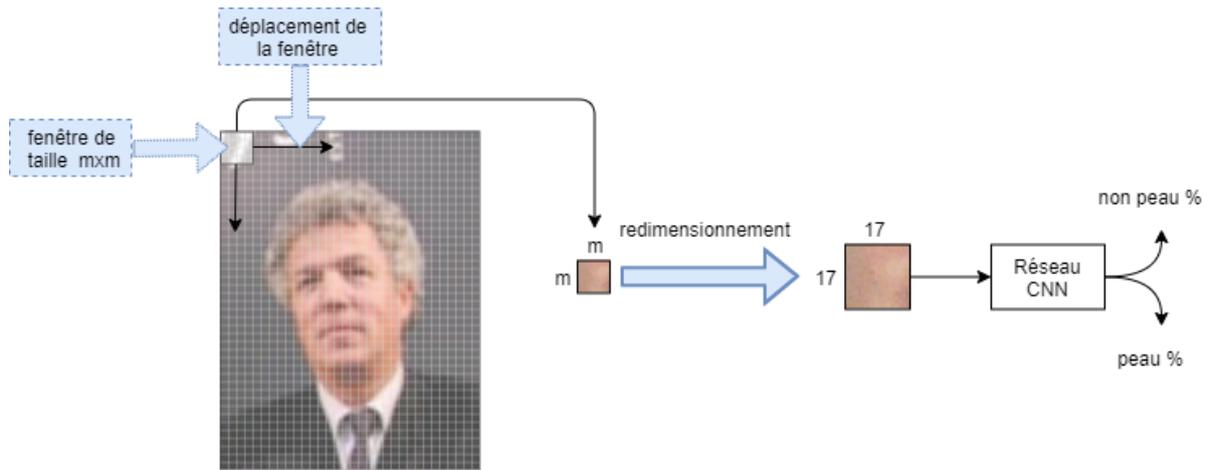


Figure 3.5 – Balayage et redimensionnement de l’image à l’issue de la fenêtre d’analyse

7 Réseau CNN

La phase d’apprentissage est effectuée pour atteindre un meilleur classifieur CNN qui prend en entrée une image RVB et renvoie le pourcentage de (peau / non peau). pour l’apprentissage il y a des étapes très précises :

- Préparation de la base de données utilisée pour l’apprentissage (redimensionnement, étiquetage, division apprentissage / test).
- Choix des paramètres d’entrée en fonction des expérimentations pour un meilleur apprentissage.
- Détermination de l’architecture CNN utilisée pour l’apprentissage.
- Sauvegarde du meilleur classifieur obtenu. La phase apprentissage est bien expliquée dans la figure 3.2.

Nous avons présenté dans le chapitre précédent, de nombreuses architectures fréquemment utilisées, dont certaines réussissent et d’autres qui ont obtenu des résultats presque excellents. Mais le succès d’un modèle dans un domaine ou une base de données spécifique n’en fait pas le choix approprié ou n’en fait pas un critère de réussite. Dans notre projet, nous avons choisi deux architectures : Lenet-5 classique et VGG-16, puis nous avons testé l’apprentissage par transfert en utilisant le célèbre modèle pré-entraîné VGG16. Dans cette section nous présentons en détail chacune de ces architectures.

7.1 Architecture de réseau Lenet-5

L'une des principales raisons du choix de "Lenet-5" est sa simplicité et cette fonctionnalité nous permet de préserver le maximum de caractéristiques possibles, car le grand nombre de couches selon notre expérience détruit les caractéristiques de base, que sont la couleur et la texture. La base de données contient des échantillons de petite taille, ce qui rend l'utilisation d'un grand nombre de couche convolution ou pooling inutile ou mauvaise. La figure 3.5 montre l'architecture CNN utilisée. Le tableau explique les détails de la configuration de l'architecture utilisée. La fonction d'activation utilisée dans cette architecture est 'tanh' dans toutes les couches sauf la dernière couche est 'softmax'. Le tableau 3.2 présente les détails du réseau.

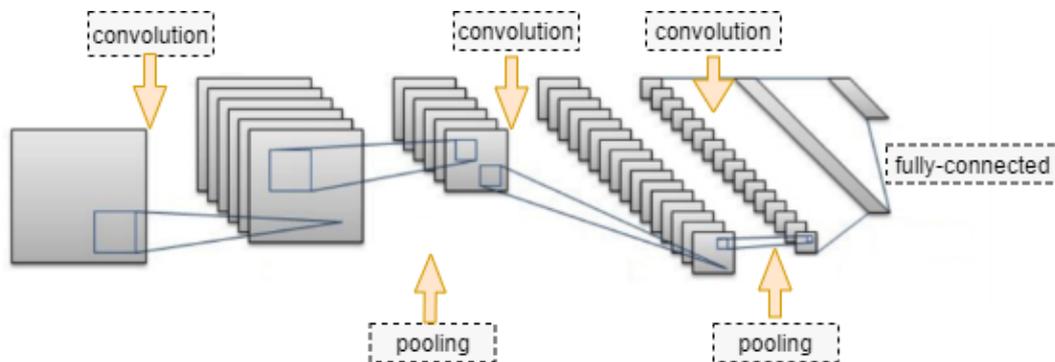


Figure 3.6 – Architecture générale du réseau lenet-5

7.2 Architecture VGG-16

VGG16 est un réseau convolutionnel très profond pour la reconnaissance d'images à grande échelle. Le modèle atteint une précision de test de 92,7 % top-5 dans ImageNet. Il apporte une amélioration par rapport à AlexNet en remplaçant les grands filtres de la taille d'un noyau (11 et 5 dans la première et la deuxième couche convolutive, respectivement) par plusieurs filtres dont la taille du noyau est de 3x3 l'un après l'autre.

La fonction d'activation utilisée dans cette architecture est 'relu' dans toutes les couches du modèle sauf la dernière couche est 'softmax'. La raison du choix de ce modèle est la comparaison entre un modèle très profond (VGG-16) et un modèle peu profond (Lenet-5) et également la comparaison entre l'apprentissage avec ou sans transfert (VGG-16). Le tableau 3.3 présente les détails du réseau.

couche	taille d'entrée	nombre de filtres	taille de filtre	stride	padding	taille de sortie	nombre de paramètres
convolution	17x17x3	6	5x5	1	pareil	17x17x6	456
pooling moyenne	17x17x6	-	5x5	1	valide	16x16x6	
convolution	16x16x6	16	5x5	1	valide	12x12x16	2416
pooling moyenne	12x12x16	-	2x2	2	valide	6x6x16	
convolution	6x6x16	120	5x5	1	valide	2x2x120	48120
Entièrement connecté	480	-	-	-	-	84	40404
Entièrement connecté	84	-	-	-	-	2	170

Table 3.2 – Tableau détaillé sur les du réseau Lenet-5

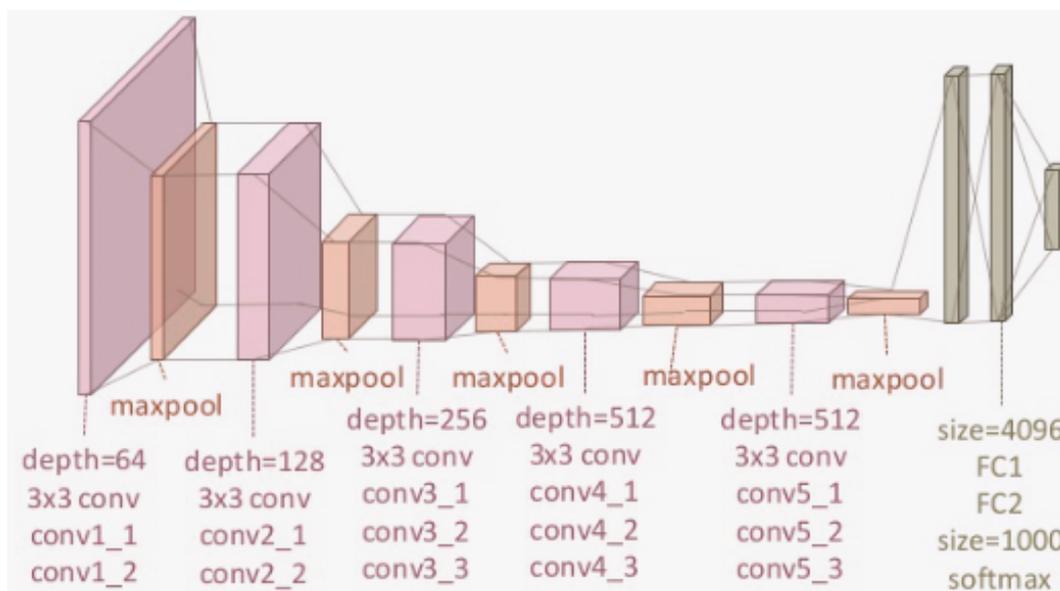


Figure 3.7 – Architecture générale du réseau VGG-16

7.3 Apprentissage du réseau VGG-16 avec transfert

Un problème majeur dans l'utilisation des CNN, c'est la taille de la base de formation qui doit être de l'ordre du million d'échantillons pour avoir un bon modèle, par contre notre base de formation ne dépasse pas 10 mille échantillons. L'une des solutions généralement utilisées est la formation par transfert. Alors que l'apprentissage par transfert offre la possibilité d'utiliser une plus petite base de données d'entraînement, le choix des meilleures données possibles pour l'entraînement est toujours essentiel au succès de la méthode globale. Nous avons donc utilisé une architecture VGG avec pré-entraînement basée sur la base de données Image-Net. La figure 3.8 montre le concept de l'apprentissage par transfert.

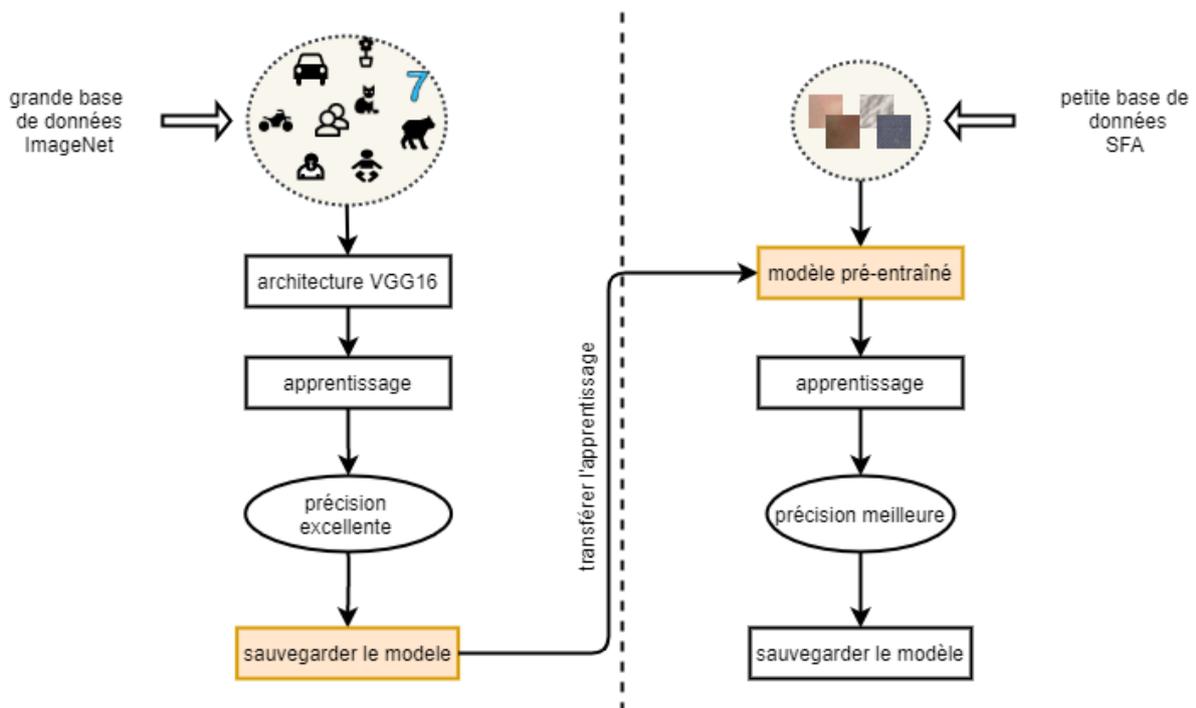


Figure 3.8 – Apprentissage par transfert

8 Seuillage

La sortie du réseau CNN étant des valeurs numériques représentant la probabilité de l'appartenance de l'imagette (image à l'issue de la fenêtre) à la classe peau ou non peau, l'image globale

couche	taille d'entrée	nombre de filtres	taille de filtre	stride	padding	taille de sortie	nombre de parametres
convolution	35x35x3	64	3x3	-	pareil	35x35x64	1792
convolution	35x35x64	64	3x3	-	pareil	35x35x64	36928
pooling maximal	35x35x64	-	2x2	2	pareil	17x17x64	0
convolution	17x17x64	128	3x3	-	pareil	17x17x128	73856
convolution	17x17x128	128	3x3	-	pareil	17x17x128	147584
pooling maximal	17x17x128	-	2x2	2	pareil	8x8x128	0
convolution	8x8x128	256	3x3	-	pareil	8x8x256	295168
convolution	8x8x256	256	3x3	-	pareil	8x8x256	590080
convolution	8x8x256	256	3x3	-	pareil	8x8x256	590080
pooling maximal	8x8x256	-	2x2	2	-	4x4x256	0
convolution	4x4x256	512	3x3	-	pareil	4x4x512	1180160
convolution	4x4x512	512	3x3	-	pareil	4x4x512	2359808
convolution	4x4x512	512	3x3	-	pareil	4x4x512	2359808
pooling maximal	4x4x512	-	2x2	2	-	2x2x512	0
convolution	2x2x512	512	3x3	-	pareil	2x2x512	2359808
convolution	2x2x512	512	3x3	-	pareil	2x2x512	2359808
convolution	2x2x512	512	3x3	-	pareil	2x2x512	2359808
pooling maximal	2x2x512	-	2x2	2	-	1x1x512	0
Entièrement connecté	512	-	-	-	-	256	131328
Entièrement connecté	256	-	-	-	-	128	32896
Entièrement connecté	128	-	-	-	-	2	258

sera donc une image de probabilité affichée en niveau de gris. Un seuillage est donc nécessaire pour afficher l'image binaire de peau (figure 3.9).

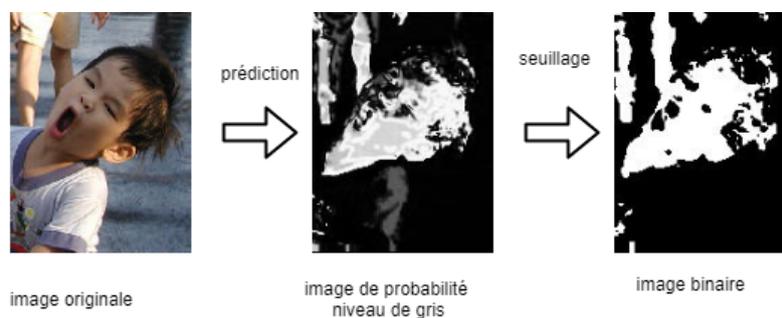


Figure 3.9 – Affichage de l'image résultante de la segmentation

9 Post-traitement

Les images binaires obtenues sont sûrement affectés de bruit causé par les fausses détections, il est donc préférable d'améliorer leur qualité avant d'afficher l'image peau.

Dans cette étape, nous avons appliqué deux opérations morphologiques qui sont la fermeture et l'ouverture pour supprimer le bruit présent sous forme de segments blancs ou de trous noirs.

Le but de l'opération de fermeture est donc de supprimer les petits trous noirs de l'image. La figure 3.10 montre un effet de cette opération.

Le but de l'opération d'ouverture est de supprimer les petits segments blancs de l'image. La figure 3.11 montre un effet de cette opération.



Figure 3.10 – Amélioration de l'image binaire par fermeture

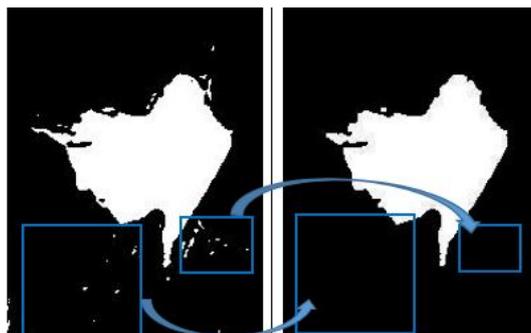


Figure 3.11 – Amélioration de l'image binaire par ouverture

10 Extraction de l'image peau

L'extraction ou l'affichage de l'image peau se fait par une simple multiplication entre l'image originale et l'image binaire, pour donner comme résultat une image RVB avec la peau seulement (figure 3.12).



Figure 3.12 – Affichage de l'image peau

11 Conclusion

Afin d'afficher une image de peau de bonne qualité, nous avons utilisé un système très compliqué malgré les étapes qui semblent claires et faciles, mais la modification de l'un des paramètres ou de la configuration affecte directement la qualité du résultat, donc notre système est très sensible aux modifications, pour cela nous avons fait beaucoup d'expérimentation afin de définir un classificateur acceptable et des traitements qui sont vraiment nécessaires pour améliorer les résultats finaux.

Chapitre 4

Implémentation

1 Introduction

Après avoir présenté dans le chapitre précédent la conception détaillée de notre application, nous présentons dans ce chapitre deux parties :

La première partie est l'implémentation de notre application, ainsi que l'environnement de développement avec les bibliothèques utilisées et les résultats de la précision de nos modèles.

La deuxième partie est la présentation du test des modèles et des résultats obtenus et l'interprétation de ces résultats.

2 Environnement de développement

2.1 Google colab

Colaboratory ou "Colab". Permet d'écrire et d'exécuter le code Python de son choix par le biais du navigateur. Offert par Google (gratuit), basé sur Jupyter Notebook et destiné à la formation et à la recherche dans l'apprentissage automatique. Cette plateforme permet d'entraîner des modèles de Machine Learning directement dans le cloud. [1]

Colab permet :

- D'améliorer les compétences de codage en langage de programmation Python.
- De développer des applications en Deep Learning en utilisant des bibliothèques Python populaires telles que Keras, TensorFlow, PyTorch et OpenCV.
- D'utiliser un environnement de développement (Jupyter Notebook) qui ne nécessite aucune configuration, Mais la fonctionnalité qui distingue Colab des autres services est l'accès à un processeur graphique GPU, totalement gratuit.

Mais la fonctionnalité qui distingue Colab des autres services est l'accès à un processeur graphique GPU, totalement gratuit.

2.2 Type d'exécution GPU

Les types de GPU disponibles dans Colab peuvent varier au fil du temps. Cette fluctuation est nécessaire pour maintenir un accès gratuit aux ressources de Colab. Les GPU disponibles dans Colab incluent souvent les modèles K80, T4, P4 et P100 de Nvidia.

- Accélérez les tâches de calcul telles que le machine learning et le HPC
- Un large choix de GPU adapté à une variété de besoins en termes de performances et de budgets
- Une tarification flexible et des personnalisations de machines en fonction de votre charge de travail .[2]

3 Langage de programmation et bibliothèques

3.1 Python

Python est python est devenu ces dernières années le langage de programmation le plus employé par les informaticiens. Ce langage s'est propulsé en tête de la gestion d'infrastructure, d'analyse de données ou dans le domaine du développement de logiciels. En effet, parmi ses qualités, Python permet notamment aux développeurs de se concentrer sur ce qu'ils font plutôt que sur la manière dont ils le font. Il a libéré les développeurs des contraintes de formes qui occupaient leur temps avec les langages plus anciens. Ainsi, développer du code avec Python est plus rapide qu'avec d'autres langages [5].

3.2 bibliothèques utilisés

TensorFlow

Nous avons utilisé cette bibliothèque pour définir les composants de base de l'architecture CNN. cette bibliothèque est destiné pour l'implémentation des algorithmes d'apprentissage automatique, elle offre aussi une grande flexibilité dans le cadre de l'utilisation pour le développement d'un réseau CNN [14].

Keras

Permet les bibliothèques utilisé avec TensorFlow est Keras , nous avons utilisé cette bibliothèque pour implémenter les couches les différents couches , les fonctions d'activation et la préparation de la base d'apprentissage [4] .

NumPy

Nous avons utilisé cette bibliothèque pour adapter le types d'entrée selon la configuration du modèles utilisés, destinée à manipuler des matrices ou tableaux multidimensionnels ainsi que des fonctions mathématiques opérant sur ces tableaux. Nous avons utilisé cette bibliothèque exactement dans le cas de balayage de l'image et l'extraction des fenêtres [7].

PIL et OpenCV

Nous avons utilisé ces bibliothèques pour le traitement d'images, nous permet d'ouvrir, de manipuler et d'enregistrer différents formats de fichiers graphiques. Dans notre cas est utilisées pour manipuler les images de test[10] [8].

Matplotlib

Matplotlib est une bibliothèque du langage de programmation Python destinée à tracer et visualiser des données sous formes de graphiques, nous avons utilisé cette bibliothèque pour visualiser note images sous formes de graphiques [6].

4 Base d'apprentissage

Dans la phase d'apprentissage nous avons utilisé un ensemble de 8944 patches, dont 3354 d'exemples positifs (peau) et 5590 d'exemples négatifs (non peau) de la base SFA disponibles avec 18 dimensions. Le tableau 4.1 résume le contenu de la base SFA.

La base divisée en 80% pour l'apprentissage et 20% pour le test.

5 Bases de test

Pour le test nous avons utilisé : La base de données BAO qui est une base de données open source destinée aux travaux sur le visage, elle contient 221 images à plusieurs visages et 149 images à un seul visage.

La base SFA que nous l'avons utilisé dans l'apprentissage, elle contient 1118 image a un seul

type d'image	quantité
images originales	1118
images binaires	1118
échantillons de peau	3354
échantillons de non-peau	5590
Total	163228

Table 4.1 – La répartition de la base SFA

visage. Ils sont Présenté dans la section 6 chapitre 1.

6 Implémentation de l'architecture

Notre application est constituée de deux modules : apprentissage et segmentation, nous présentons dans cette section chaque module avec leurs détails de réalisation :

6.1 Module apprentissage

Comme nous l'avons déjà expliqué dans le chapitre 3, nous avons testé deux architectures classiques Lenet-5 et VGG16, ainsi que l'apprentissage par transfert du VGG16 pré-entraîné Pour chaque modèle nous avons suivi les étapes :

- La préparation de l'environnement de l'exécution Google Colab pour un mode GPU.
- La préparation de la base la base d'apprentissage SFA comme nous avons montré dans la section de base d'apprentissage. Nous avons utilisé cette bibliothèque de Keras : `from keras.preprocessing.image import ImageDataGenerator`.
- L'importation des bibliothèques nécessaires Keras pour l'implémentation d'un modèle séquentiel. Puis, nous avons importé les couches principales du réseau : `From keras. Layers import Dense, Dropout, Activation, Flatten`.
- L'entraînement du modèle.
- le sauvegarde du modèle obtenue.

Pour améliorer notre résultat nous avons relancé l'apprentissage plusieurs fois afin de fixer les meilleurs paramètres, les paramètres principaux a fixé pour l'apprentissage sont batch-size,

epoch, steps-per-epoch, validation-steps :

- **Le batch-size** est le nombre d'échantillons utilisés dans une itération, la bonne taille de batch-size fournit une estimation suffisamment stable de ce que serait le gradient de la base de données complet. Pour cela nous avons fait plusieurs expérimentations dans ce contexte.
- **Epoch** est le nombre total d'itération d'apprentissage, il est défini comme un critère d'arrêts que ce soit les résultats, nous avons fixé un nombre d'epoch à 100 epochs.
- **Steps-per-epoch** est le nombre total d'étapes à produire du générateur avant de déclarer une époque terminée et de commencer la prochaine époque. Il doit généralement être égal à (nombre d'échantillons total / nombre de batch-size).

Nous utilisons les appellations Step-per-epoch pour la partie apprentissage et validation-steps pour la partie test.

Les résultats d'apprentissages précédents sont montrés dans la figure 4.1, la précision avec toutes les configurations précédentes n'a pas dépassé 85 %, et nous avons obtenu des résultats non stables.

Résultats et interprétation

Nous validons chaque modèle selon leur taux de précision, le taux d'erreur. Comme nous avons divisé la base en deux parties d'apprentissage et de test, donc pour chaque itération d'apprentissage, il y a le taux de précision pour l'apprentissage (accuracy) et le test (val-accuracy) et le taux d'erreur pour l'apprentissage (loss) et le test (val-loss). Les figures montrent les résultats obtenus pour chaque modèle où nous avons obtenu les précisions suivantes : 93% (figure 4.2) pour le réseau lenet-5, 86% (figure 4.3) pour le VGG-16 que nous avons entraîné par nos soins du début à la fin et une précision de 96% (figure 4.4) pour le VGG-16 pré-entraîné avec 'transfer learning'.

Nous pensons que le VGG-16 pré-entraîné a été renforcé par un entraînement sur les patches de la peau d'où le taux élevé de 96%, mais ces mêmes patches sont insuffisants pour entraîner le VGG-16 à zéro d'où le taux faible 86%, il est peut-être plus adéquat de l'entraînement sur une base d'images complètes (non pas de patches) et de taille considérable afin d'atteindre une précision acceptable.

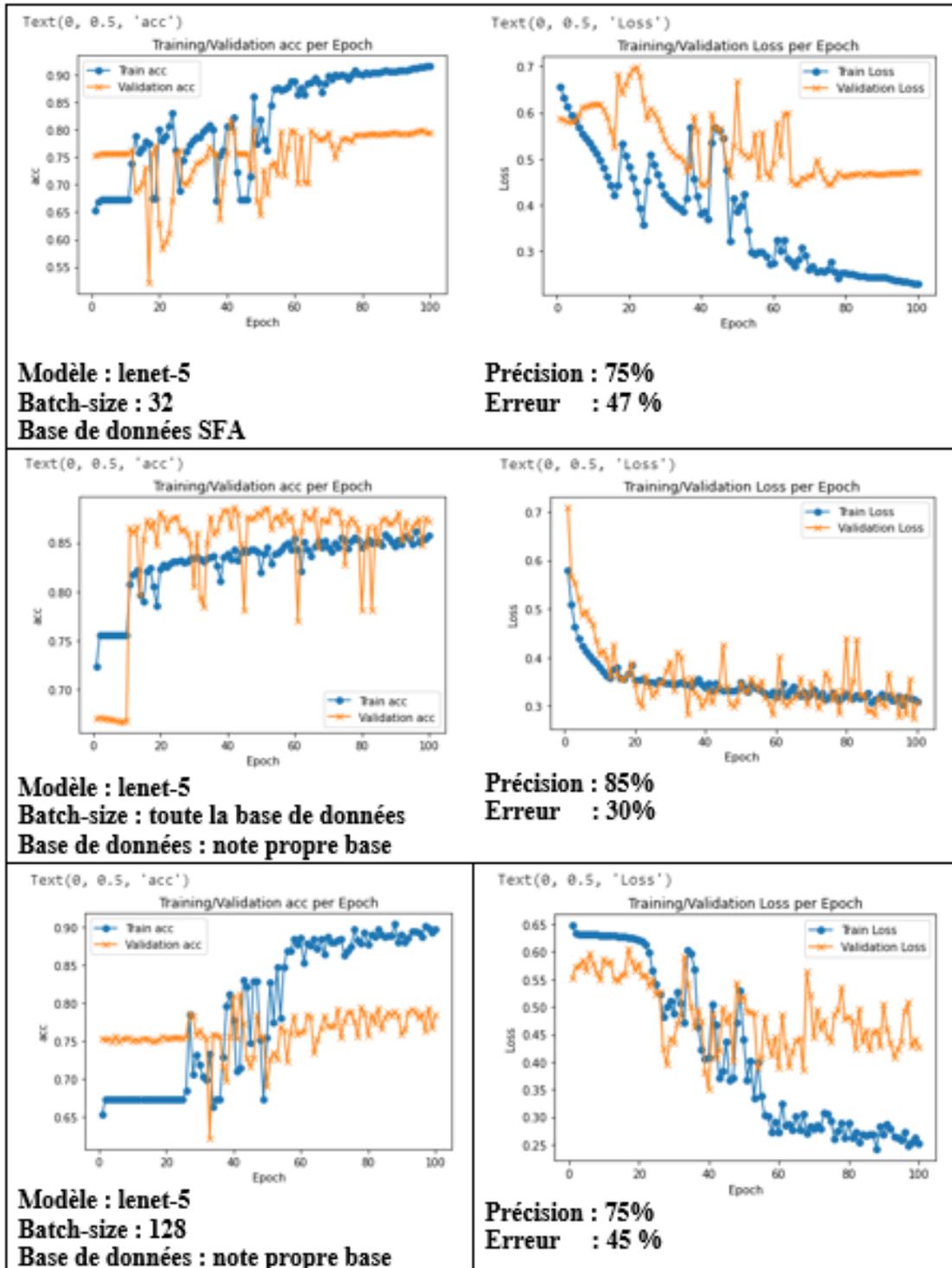
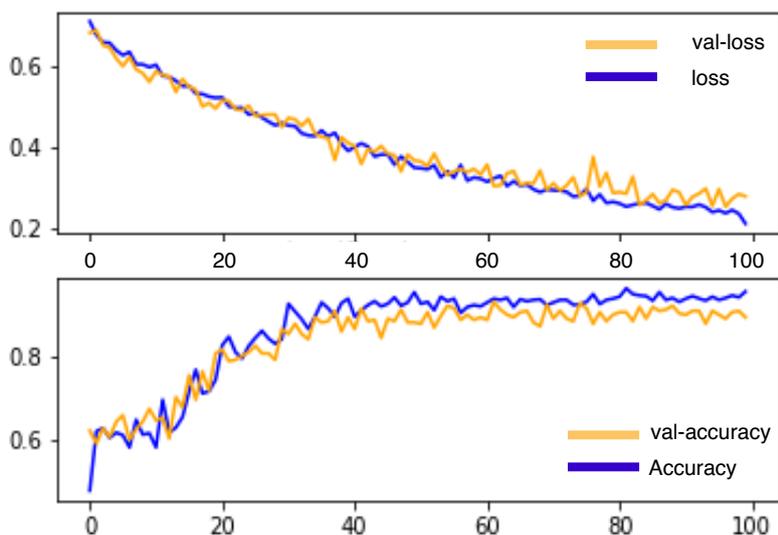


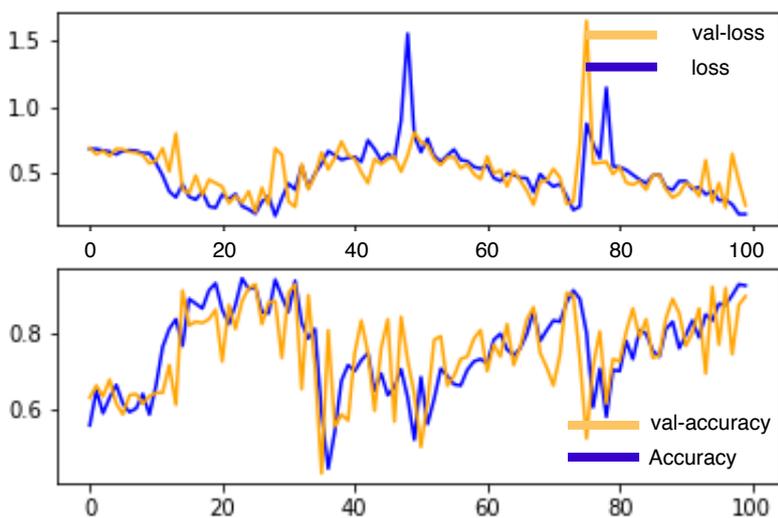
Figure 4.1 – Illustration sur les résultats précédents



Modèle : lenet-5 steps_per_epoch 10 Base d'apprentissage : SFA
 batch-size : 64 validation_steps : 5

:

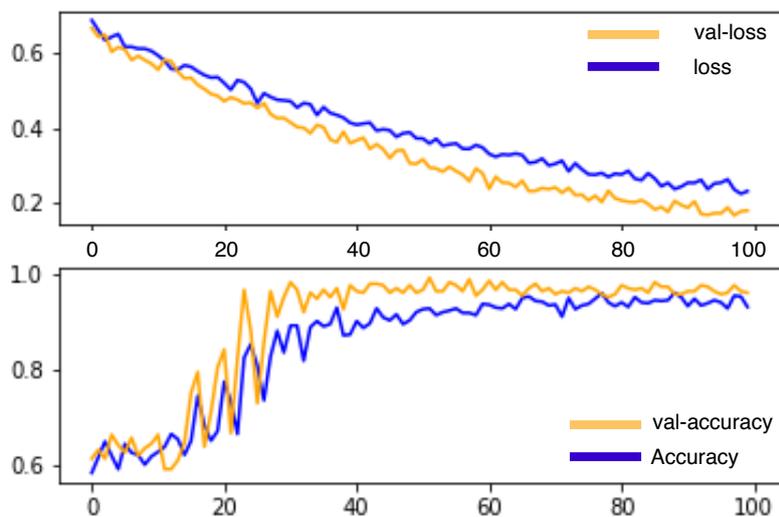
Figure 4.2 – Résultats du réseau Lenet-5



Modèle : VGG-16 steps_per_epoch 10 Base d'apprentissage : SFA
 batch-size : 64 validation_steps : 5

:

Figure 4.3 – Résultats du réseau VGG-16



Modèle : VGG-16 pré-entraîné steps_per_epoch 10 Base d'apprentissage : SFA
batch-size : 64 validation_steps : 5

Figure 4.4 – Résultats du réseau VGG-16 pré-entraîné

6.2 Module segmentation

Nous avons appliqué ce processus pour le but de tester le modèle obtenue sur une image complète, pour cela nous avons implémenté un algorithme de segmentation, accompagné par des étapes de prétraitement et post-traitements. Nous avons partagé nos tests selon le type et les caractéristiques des images utilisées :

Une personne avec arrière-plan simple, plusieurs personnes avec arrière-plan simple, une personne avec arrière-plan complexe, plusieurs personnes avec arrière-plan complexe, images de bonne qualité, images de mauvaise qualité.



Figure 4.5 – Image d’une personne avec arrière-plan simple

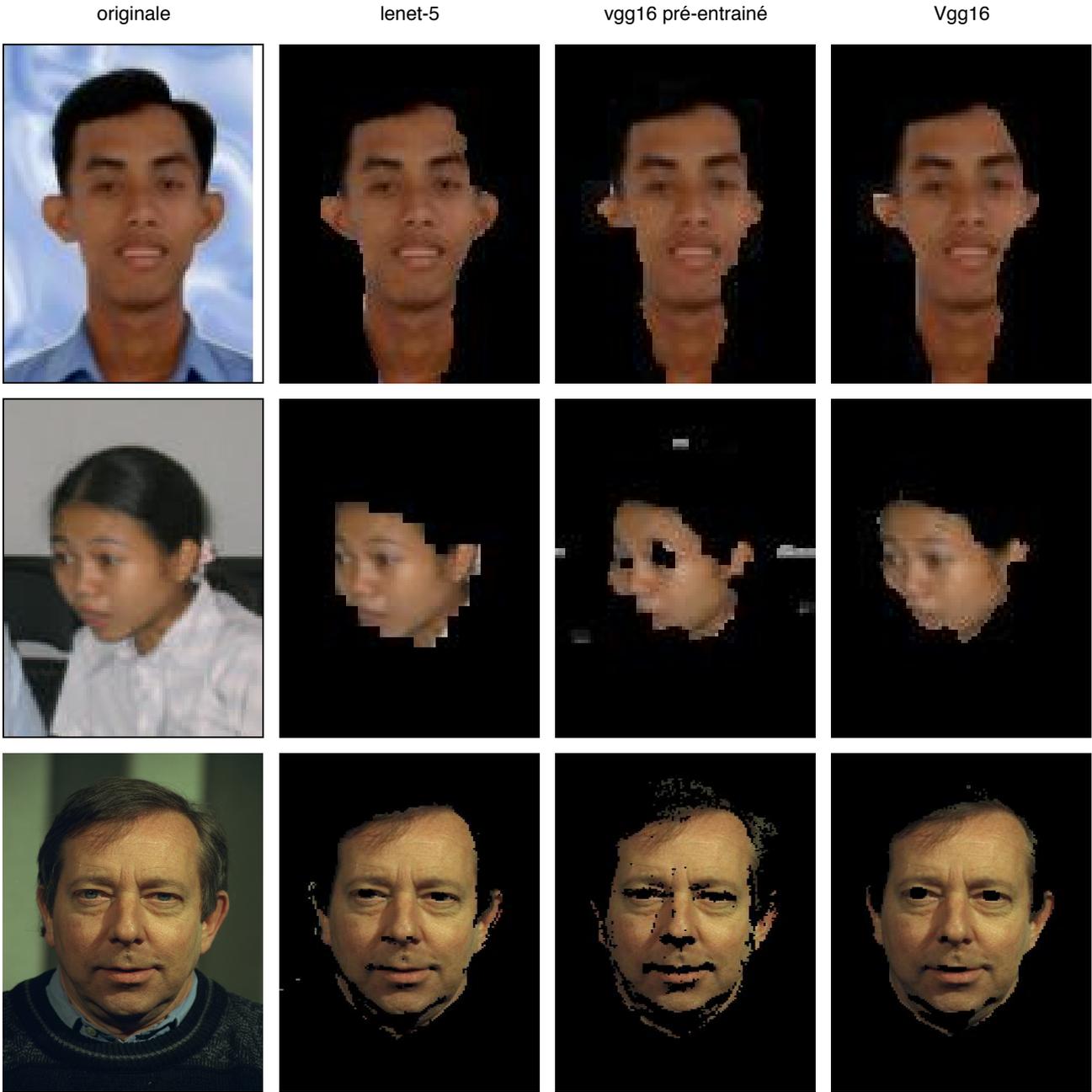


Figure 4.6 – Image d’une personne avec arrière-plan simple

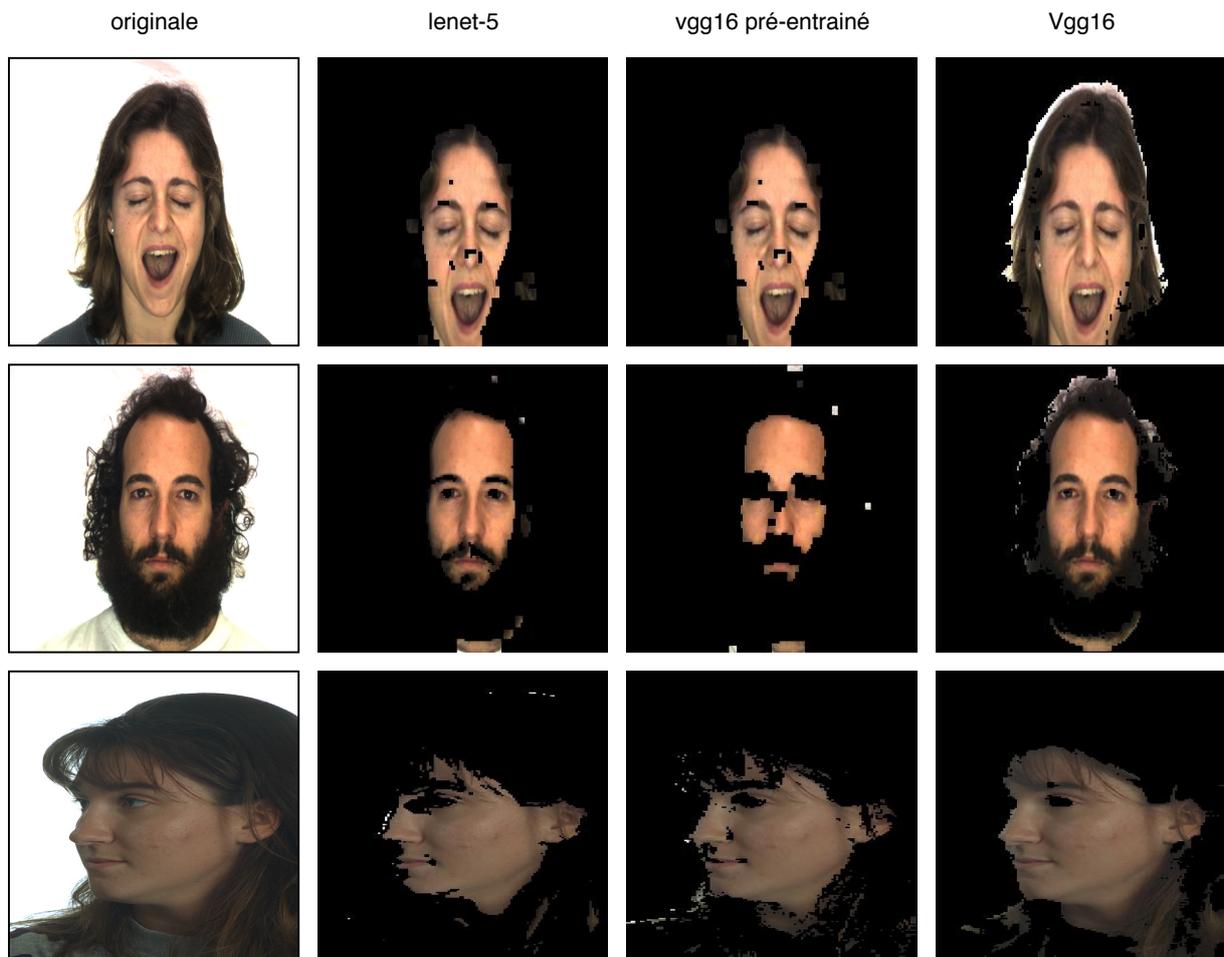


Figure 4.7 – Image d’une personne avec arrière-plan simple

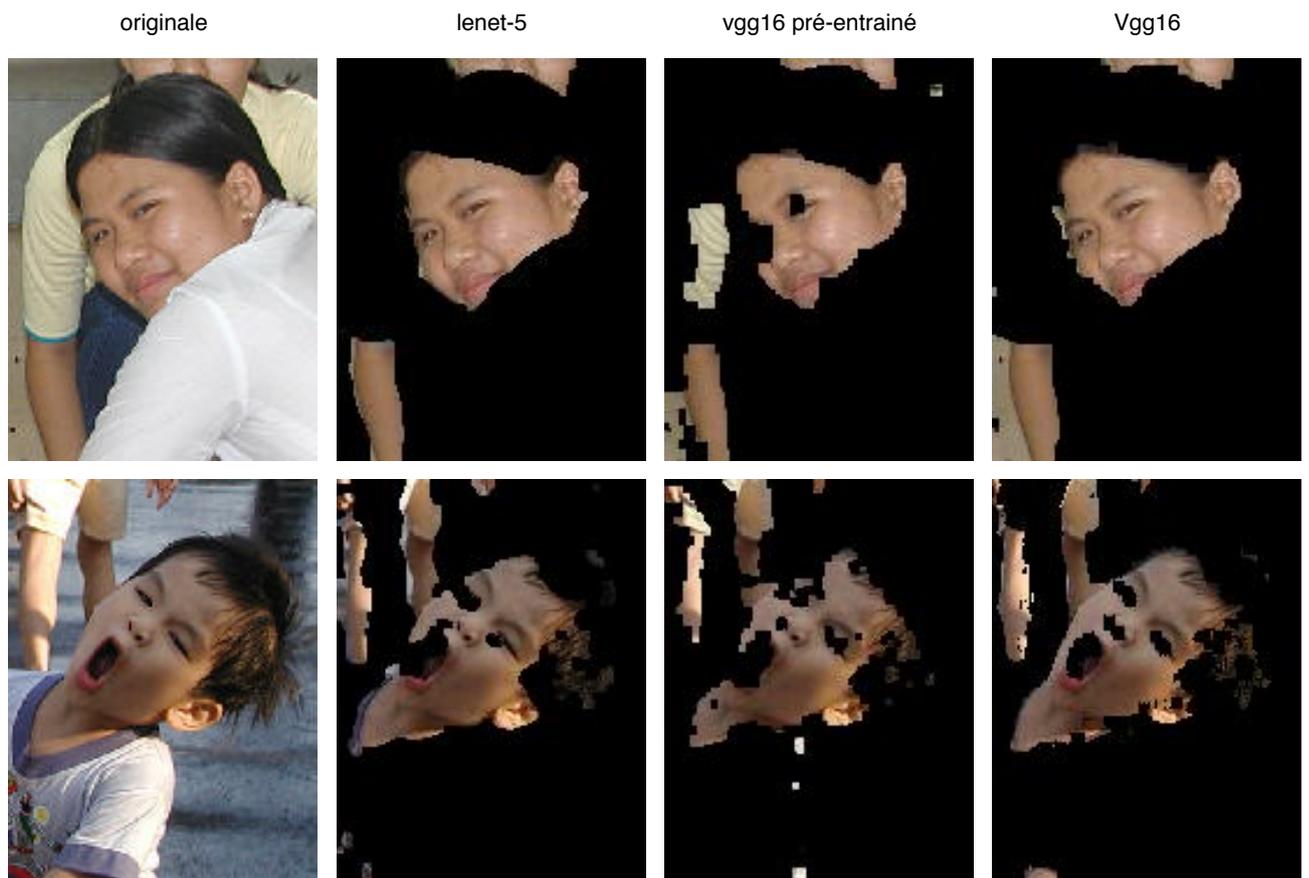


Figure 4.8 – Image d’une personne avec arrière-plan complexe

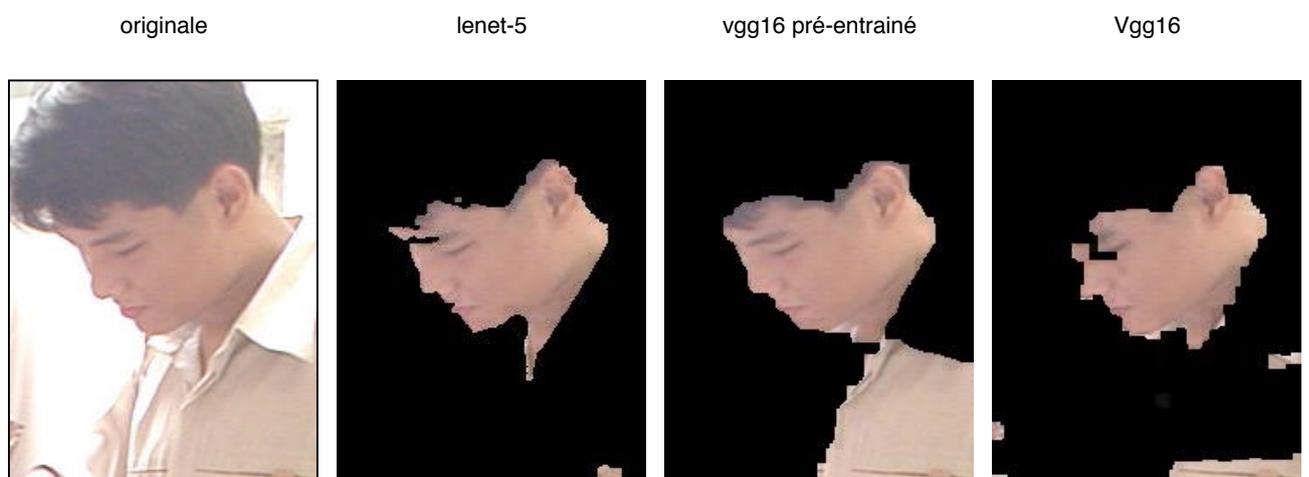


Figure 4.9 – Image d’une personne avec mauvaise qualité

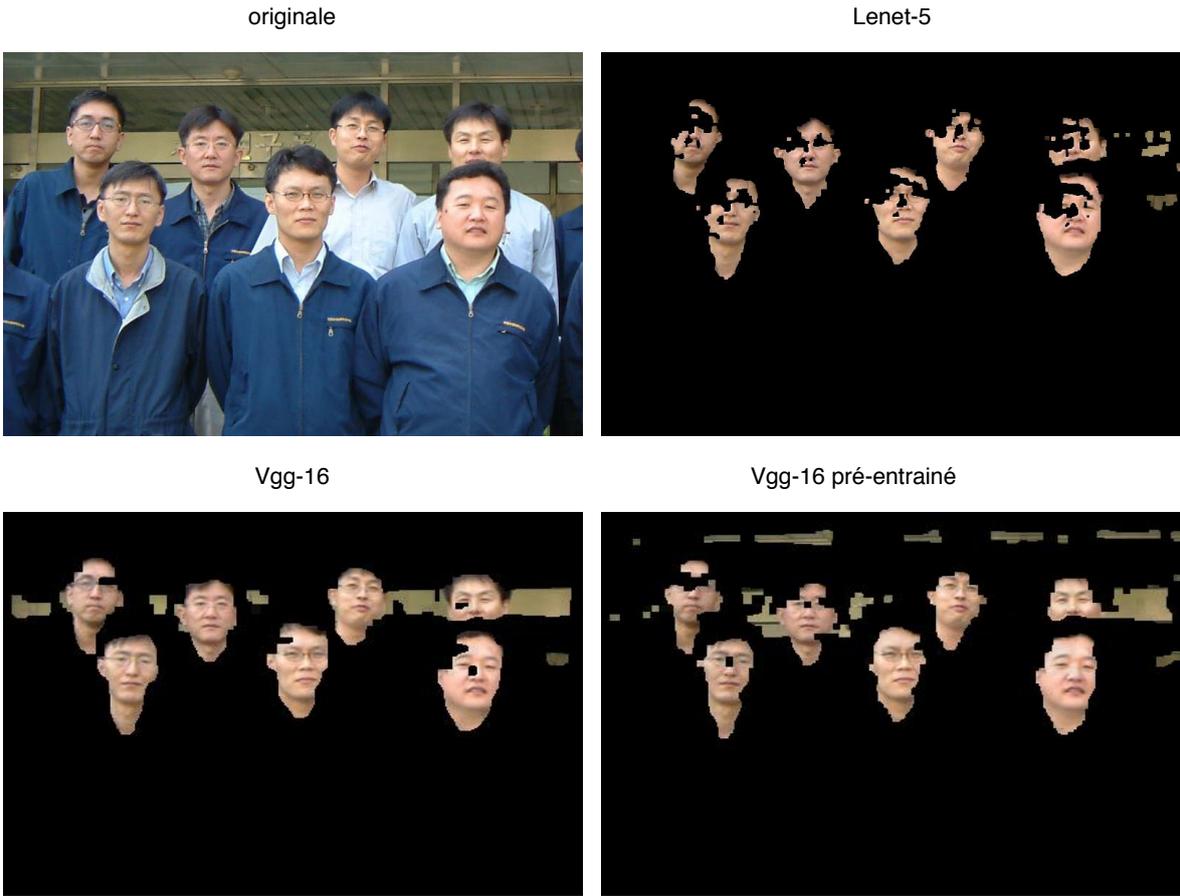


Figure 4.10 – Image de plusieurs personnes

Résultats finaux

Nous avons testé les trois réseaux CNN avec 20 images avec des caractéristiques aléatoires et nous avons obtenue pour chaque réseau des taux de bonne détection classé dans le tableau 4.2 :

	Lenet-5	VGG-16	VGG-16 pré-entraîné
Précision d'apprentissage	93%	86%	96%
Précision de test sur SFA	96%	80%	93%
Précision sur BAO	95%	80%	83%

Table 4.2 – Les résultats du test final

7 Interface de système

Fenêtre initiale

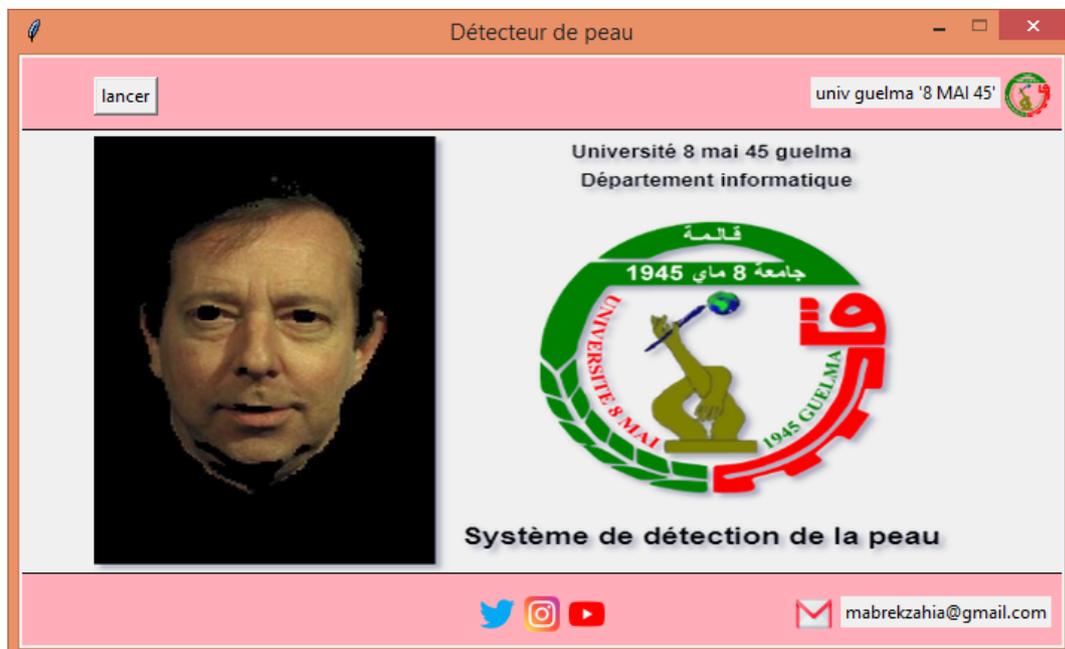


Figure 4.11 – Interface initiale de notre application

Bouton "Lancer" pour lancer la fenêtre principale de la détection

Fenêtre principale de détection



Figure 4.12 – Fenêtre principale de détection

Bouton "ouvrir image" permet de charger une image.

Bouton "prétraitement" permet de choisir les paramètres de prétraitements .

Bouton "segmentation" permet de choisir les paramètres de segmentation a fin de lancer l'opération de segmentation .

Bouton " post-traitement" permet de choisir les paramètres de post-traitement a fin de lancer l'opération de ouverture ou fermeture.

Fenêtres des traitements

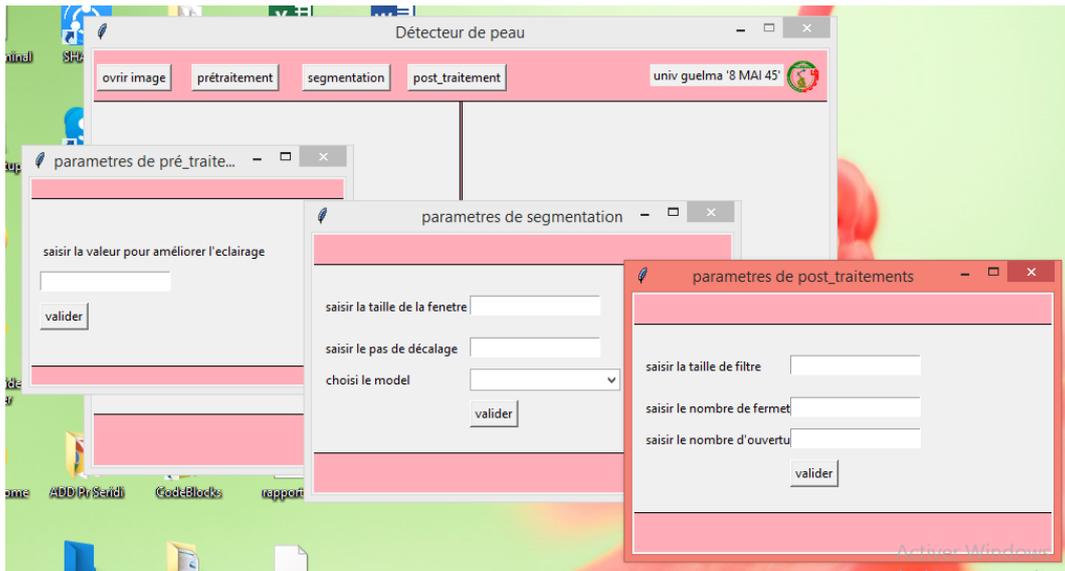


Figure 4.13 – Fenêtres des paramètres de traitements

8 Conclusion

Les tests effectués sur des images réelles avec des environnements à complexité variables montrent que le réseau Lenet-5 a dépassé en matière de précision le réseau VGG-16 pré-entraîné malgré la robustesse prouvée par ce type de réseau dans la classification des objets. Cela est dû à notre avis à la spécificité de l'objet à classifier (peau humaine) qui est une combinaison de texture et de la couleur et n'a aucune forme spécifique.

Conclusion générale

Conclusion générale

La détection de la peau est devenue très importante dans les applications de détection ou de reconnaissance des personnes là où elles sont utilisées dans les domaines de la sécurité et la médecine ...etc. Ces dernières années ce domaine a suscité un intérêt croissant pour les chercheurs en raison de l'augmentation des outils matériels et des techniques d'apprentissage automatique et l'intelligence artificielle.

Dans notre recherche nous avons étudiés les travaux les plus récents sur la détection de la peau et les méthodes qu'ils ont utilisées pour obtenir ou améliorer les résultats dans ce domaine.

La détection est réalisée par les réseaux de neurones à convolution, ils sont l'un des meilleurs algorithmes d'apprentissage pour comprendre le contenu des images et ont montré des performances exemplaires dans les tâches liées à la segmentation, la classification, la détection et l'extraction des objets.

Les difficultés dans le processus de détection de la peau est la grande complexité de la classe non peau et la diversité de la classe peau quelle peut être effectuer par : la diversité des types d'images et des sources, la couleur de la peau humaine, manque ou augmentation de l'éclairage, de la réflexion du soleil ou de la lumière sur la peau, mauvaise résolution, environnement complexe, le bruit, le contraste,...etc.

Les résultats de la détection sont fiables dans la plupart des cas, (testées sur des images réelles), seulement qu'il y a quelques images nécessitant certain traitements pour améliorer la qualité de détection.

En fin , il serait intéressant dans l'avenir de généraliser notre système en un système de détection et reconnaissance des personnes avec une combinaison de deux réseaux de neurones a convolution pour la détection de peau et pour l'identification des personnes dans la même image.

Bibliographie

- [Alom et al., 2018] Alom, M. Z., Taha, T. M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M. S., Van Esesn, B. C., Awwal, A. A. S., and Asari, V. K. (2018). The history began from alexnet : A comprehensive survey on deep learning approaches. *arXiv preprint arXiv :1803.01164*.
- [Alpaydin, 2020] Alpaydin, E. (2020). *Introduction to machine learning*. MIT press.
- [Angelova et al., 2005] Angelova, A., Abu-Mostafam, Y., and Perona, P. (2005). Pruning training sets for learning of object categories. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, volume 1, pages 494–501. IEEE.
- [Arulkumaran et al., 2017] Arulkumaran, K., Deisenroth, M. P., Brundage, M., and Bharath, A. A. (2017). A brief survey of deep reinforcement learning. *arXiv preprint arXiv :1708.05866*.
- [Bay et al., 2000] Bay, S. D., Kibler, D., Pazzani, M. J., and Smyth, P. (2000). The uci kdd archive of large data sets for data mining research and experimentation. *ACM SIGKDD explorations newsletter*, 2(2) :81–85.
- [Bencheriet, 2019] Bencheriet, C. (2019). New face features to detect multiple faces in complex background. *Evolving Systems*, 10(2) :79–95.
- [Bencheriet et al., 2007] Bencheriet, C., Boualleg, A., Tebbikh, H., Guerzize, B., and Belguidoum, W. (2007). Dtection de visages par méthode hybride couleur de peau et template matching. *SETIT'2007*.
- [Bengio, 2013] Bengio, Y. (2013). Deep learning of representations : Looking forward. In *International Conference on Statistical Language and Speech Processing*, pages 1–37. Springer.
- [Brette, 2003] Brette, R. (2003). *Modeles impulsionnels de réseaux de neurones biologiques*. PhD thesis, Université Pierre et Marie Curie-Paris VI.

- [Casati et al., 2013] Casati, J. P. B., Moraes, D. R., and Rodrigues, E. L. L. (2013). Sfa : A human skin image database based on feret and ar facial images. In *IX workshop de Visao Computational, Rio de Janeiro*.
- [Ciresan et al., 2012] Ciresan, D., Giusti, A., Gambardella, L. M., and Schmidhuber, J. (2012). Deep neural networks segment neuronal membranes in electron microscopy images. In *Advances in neural information processing systems*, pages 2843–2851.
- [CireşAn et al., 2012] CireşAn, D., Meier, U., Masci, J., and Schmidhuber, J. (2012). Multi-column deep neural network for traffic sign classification. *Neural networks*, 32 :333–338.
- [Dollár et al., 2009] Dollár, P., Tu, Z., Perona, P., and Belongie, S. (2009). Integral channel features.
- [Elgammal et al., 2009] Elgammal, A., Muang, C., and Hu, D. (2009). Skin detection-a short tutorial. *Encyclopedia of Biometrics*, 4 :1218–1224.
- [Goodfellow et al., 2016] Goodfellow, I., Bengio, Y., Courville, A., and Bengio, Y. (2016). *Deep learning*, volume 1. MIT press Cambridge.
- [Graupe, 2013] Graupe, D. (2013). *Principles of artificial neural networks*, volume 7. World Scientific.
- [Gupta and Chaudhary, 2016] Gupta, A. and Chaudhary, A. (2016). Robust skin segmentation using color space switching. *Pattern Recognition and Image Analysis*, 26(1) :61–68.
- [Hon and Khan, 2017] Hon, M. and Khan, N. M. (2017). Towards alzheimer’s disease classification through transfer learning. In *2017 IEEE International conference on bioinformatics and biomedicine (BIBM)*, pages 1166–1169. IEEE.
- [Huang et al., 2011] Huang, L., Xia, T., Zhang, Y., and Lin, S. (2011). Human skin detection in images by msr analysis. In *2011 18th IEEE International Conference on Image Processing*, pages 1257–1260. IEEE.
- [Indolia et al., 2018] Indolia, S., Goswami, A. K., Mishra, S., and Asopa, P. (2018). Conceptual understanding of convolutional neural network-a deep learning approach. *Procedia computer science*, 132 :679–688.
- [Jarrett et al., 2009] Jarrett, K., Kavukcuoglu, K., Ranzato, M., and LeCun, Y. (2009). What is the best multi-stage architecture for object recognition? In *2009 IEEE 12th international conference on computer vision*, pages 2146–2153. IEEE.

- [Jones and Rehg, 2002] Jones, M. J. and Rehg, J. M. (2002). Statistical color models with application to skin detection. *International Journal of Computer Vision*, 46(1) :81–96.
- [Kawaguchi et al., 2019] Kawaguchi, K., Huang, J., and Kaelbling, L. P. (2019). Effect of depth and width on local minima in deep learning. *Neural computation*, 31(7) :1462–1498.
- [Kawulok et al., 2013] Kawulok, M., Kawulok, J., Nalepa, J., and Papiez, M. (2013). Skin detection using spatial analysis with adaptive seed. In *2013 IEEE International Conference on Image Processing*, pages 3720–3724. IEEE.
- [Khan et al., 2019] Khan, A., Sohail, A., Zahoor, U., and Qureshi, A. S. (2019). A survey of the recent architectures of deep convolutional neural networks. *arXiv preprint arXiv :1901.06032*.
- [Khan et al., 2020] Khan, A., Sohail, A., Zahoor, U., and Qureshi, A. S. (2020). A survey of the recent architectures of deep convolutional neural networks. *Artificial Intelligence Review*, pages 1–62.
- [Kim et al., 2017] Kim, Y., Hwang, I., and Cho, N. I. (2017). Convolutional neural networks and training strategies for skin detection. In *2017 IEEE International Conference on Image Processing (ICIP)*, pages 3919–3923. IEEE.
- [Larsson et al., 2016] Larsson, G., Maire, M., and Shakhnarovich, G. (2016). Fractalnet : Ultra-deep neural networks without residuals. *arXiv preprint arXiv :1605.07648*.
- [LeCun et al., 2015] LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *nature*, 521(7553) :436–444.
- [LeCun et al., 2010] LeCun, Y., Kavukcuoglu, K., and Farabet, C. (2010). Convolutional networks and applications in vision. In *Proceedings of 2010 IEEE international symposium on circuits and systems*, pages 253–256. IEEE.
- [Liu et al., 2013] Liu, C., Sun, W., Chao, W., and Che, W. (2013). Convolution neural network for relation extraction. In *International Conference on Advanced Data Mining and Applications*, pages 231–242. Springer.
- [Liu et al., 2019] Liu, X., Deng, Z., and Yang, Y. (2019). Recent progress in semantic image segmentation. *Artificial Intelligence Review*, 52(2) :1089–1106.
- [Lu et al., 2017a] Lu, H., Li, B., Zhu, J., Li, Y., Li, Y., Xu, X., He, L., Li, X., Li, J., and Serikawa, S. (2017a). Wound intensity correction and segmentation with convolutional neural networks. *Concurrency and computation : practice and experience*, 29(6) :e3927.

- [Lu et al., 2017b] Lu, Z., Pu, H., Wang, F., Hu, Z., and Wang, L. (2017b). The expressive power of neural networks : A view from the width. In *Advances in neural information processing systems*, pages 6231–6239.
- [Mazumdar and Bora, 2016] Mazumdar, A. and Bora, P. K. (2016). Exposing splicing forgeries in digital images through dichromatic plane histogram discrepancies. In *Proceedings of the Tenth Indian Conference on Computer Vision, Graphics and Image Processing*, pages 1–8.
- [Montufar et al., 2014] Montufar, G. F., Pascanu, R., Cho, K., and Bengio, Y. (2014). On the number of linear regions of deep neural networks. In *Advances in neural information processing systems*, pages 2924–2932.
- [Nadian-Ghomsheh, 2016] Nadian-Ghomsheh, A. (2016). Pixel-based skin detection based on statistical models. *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, 8(5) :7–14.
- [Naji et al., 2019] Naji, S., Jalab, H. A., and Kareem, S. A. (2019). A survey on skin detection in colored images. *Artificial Intelligence Review*, 52(2) :1041–1087.
- [Oghaz et al., 2019] Oghaz, M. M. D., Argyriou, V., Monekosso, D., and Remagnino, P. (2019). Skin identification using deep convolutional neural network. In *International Symposium on Visual Computing*, pages 181–193. Springer.
- [Phillips et al., 2000] Phillips, P. J., Moon, H., Rizvi, S. A., and Rauss, P. J. (2000). The ferret evaluation methodology for face-recognition algorithms. *IEEE Transactions on pattern analysis and machine intelligence*, 22(10) :1090–1104.
- [Phung et al., 2005] Phung, S. L., Bouzerdoum, A., and Chai, D. (2005). Skin segmentation using color pixel classification : analysis and comparison. *IEEE transactions on pattern analysis and machine intelligence*, 27(1) :148–154.
- [Pujol et al., 2017] Pujol, F. A., Pujol, M., Jimeno-Morenilla, A., and Pujol, M. J. (2017). Face detection based on skin color segmentation using fuzzy entropy. *Entropy*, 19(1) :26.
- [Roheda, 2017] Roheda, S. (2017). A multi-scale approach to skin pixel detection. *Electronic Imaging*, 2017(4) :18–23.
- [Schmugge et al., 2007] Schmugge, S. J., Jayaram, S., Shin, M. C., and Tsap, L. V. (2007). Objective evaluation of approaches of skin detection using roc analysis. *Computer vision and image understanding*, 108(1-2) :41–51.

- [Sharma and Reilly, 2003] Sharma, P. and Reilly, R. B. (2003). A colour face image database for benchmarking of automatic face detection algorithms. In *Proceedings EC-VIP-MC 2003. 4th EURASIP Conference focused on Video/Image Processing and Multimedia Communications (IEEE Cat. No. 03EX667)*, volume 1, pages 423–428. IEEE.
- [Simonyan and Zisserman, 2014] Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv :1409.1556*.
- [Srivastava et al., 2015] Srivastava, R. K., Greff, K., and Schmidhuber, J. (2015). Highway networks. *arXiv preprint arXiv :1505.00387*.
- [Sundermeyer et al., 2012] Sundermeyer, M., Schlüter, R., and Ney, H. (2012). Lstm neural networks for language modeling. In *Thirteenth annual conference of the international speech communication association*.
- [Sze et al., 2017] Sze, V., Chen, Y.-H., Yang, T.-J., and Emer, J. S. (2017). Efficient processing of deep neural networks : A tutorial and survey. *Proceedings of the IEEE*, 105(12) :2295–2329.
- [Szegedy et al., 2015] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9.
- [Verma et al., 2014] Verma, A., Raj, S. A., Midya, A., and Chakraborty, J. (2014). Face detection using skin color modeling and geometric feature. In *2014 International Conference on Informatics, Electronics & Vision (ICIEV)*, pages 1–6. IEEE.
- [Wang et al., 2008] Wang, X., Xu, H., Wang, H., and Li, H. (2008). Robust real-time face detection with skin color detection and the modified census transform. In *2008 International Conference on Information and Automation*, pages 590–595. IEEE.
- [Yang et al., 2019] Yang, J., Xiong, W., Li, S., and Xu, C. (2019). Learning structured and non-redundant representations with deep neural networks. *Pattern Recognition*, 86 :224–235.
- [Yegnanarayana, 2009] Yegnanarayana, B. (2009). *Artificial neural networks*. PHI Learning Pvt. Ltd.
- [Zuo et al., 2017] Zuo, H., Fan, H., Blasch, E., and Ling, H. (2017). Combining convolutional and recurrent neural networks for human skin detection. *IEEE Signal Processing Letters*, 24(3) :289–293.

Webographie

- [1] Moov AI. Initiation au deep learning avec google colab. <https://moov.ai/fr/blog/deep-learning-avec-google-colab/>. Dernier accès : 04-10-2020.
- [2] Google Cloud. Gpu sur google cloud. <https://cloud.google.com/gpu>. Dernier accès : 04-10-2020.
- [3] Koetzel Gandia. Le cerveau artificiel. <http://tpe-robots-hommes-2012.e-monsite.com/pages/partie-1/b.html>. Dernier accès : 04-10-2020.
- [4] Keras. Introduction to keras for engineers. https://keras.io/getting_started/intro_to_keras_for_engineers/. Dernier accès : 04-10-2020.
- [5] Dave Kuhlman. *A python book : Beginning python, advanced python, and python exercises*. Dave Kuhlman Lutz, 2009.
- [6] matplotlib. Ai courses by opencv. <https://matplotlib.org/tutorials/index.html#introductory>. Dernier accès : 04-10-2020.
- [7] numpy. Learn numpy. <https://numpy.org/learn/>. Dernier accès : 04-10-2020.
- [8] OpenCV. Ai courses by opencv. <https://opencv.org/courses/>. Dernier accès : 04-10-2020.
- [9] J Pawan. Complete guide of activation functions. <https://stanford.edu/~shervine/1/fr/teaching/cs-230/pense-bete-reseaux-neurones-convolutionnels>, 2018. Dernier accès : 04-10-2020.
- [10] Pillow. Pillow. <https://pillow.readthedocs.io/en/stable/>. Dernier accès : 04-10-2020.
- [11] Sumit S. A comprehensive guide to convolutional neural networks — the eli5 way. <https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way--\3bd2b1164a53>. Dernier accès : 04-10-2020.

- [12] A Afshine A Shervine. Pense-bête de réseaux de neurones convolutionnels. <https://stanford.edu/~shervine/1/fr/teaching/cs-230/pense-bete-reseaux-neurones-convolutionnels>, 2019. Dernier accès : 04-10-2020.
- [13] Student. Student notes : Convolutional neural networks (cnn) introduction. <https://indoml.com/2018/03/07/student-notes-convolutional-neural-networks-cnn-introduction/>. Dernier accès : 04-10-2020.
- [14] TensorFlow. Introduction à tensorflow. <https://www.tensorflow.org/learn>. Dernier accès : 04-10-2020.
- [15] F. Van Veen and S. Leijnen. The neural network zoo. <https://www.asimovinstitute.org/neural-network-zoo>. Dernier accès : 04-10-2020.