



République Algérienne Démocratique et Populaire



Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université des Sciences et de la Technologie Houari Boumediene

**Faculté d'Electronique et d'Informatique
Département Informatique**

Mémoire de Licence

Filière: Informatique

**Spécialité : Ingénierie des Systèmes d'information et des
Logiciels**

Thème :

Analyse des sentiments basée sur la vidéo.

**Sujet Proposé par :
Mme DAHMANI Djamila.**

Soutenu le : ../../....

Devant le jury composé de:

Présenté par :

- Mlle MOKHBAT Selma
- Mr AMOURA Youcef

Mr LARABI Slimane

Président

Mme BENSEBAA Amina

Membre

Binôme n° : ISIL 024 / 2022

Remerciements

On tient à exprimer nos plus vifs remerciements à notre promotrice **Madame Djamila Dahmani** pour son aide, pour ses précieux conseils constructifs ainsi que ses critiques et orientations qu'elle nous a prodigués en contribuant à l'élaboration de ce modeste travail.

Nous tenons à remercier les membres du jury, **Monsieur Slimane Larabi** et **Madame Amina Bensebaa** pour l'honneur qu'ils ont bien voulu nous faire par leur présence.

Nous exprimons notre profonde reconnaissance à nos parents, à nos frères et sœurs qui nous ont soutenus et aidés durant la période de notre formation et surtout à la réalisation de ce modeste travail.

Que nos amis de la promotion, trouvent l'expression de notre profonde reconnaissance pour leurs aides qui nous ont été très utiles.

Enfin, nous remercions nos enseignants de **L'USTHB** pour la formation qu'ils nous ont dispensées.

Selma et Youcef

Résumé

Les expressions faciales constituent l'une des plus importantes manières de communiquer. En effet, les expressions des visages reflètent le comportement d'un individu. Dans ce projet, un système de reconnaissance d'émotions faciales est conçu, mis en œuvre et testé. Une méthode de classification basée sur l'apprentissage profond - CNN - est intégrée. Notre système est testé sur une base de données contenant plus de deux milles photos et nous avons obtenu des résultats plus qu'acceptables pour reconnaître les émotions dans les limites prévus.

Mots clés : expressions faciales, visage, émotions, apprentissage profond, CNN.

Liste des abréviations

IA: Intelligence Artificielle

ML: Machine Learning

DL: Deep Learning

ANN: Artificial Neural Network

CNN: Convolutional Neural Network

DNN: Deep Neural Network

BDD: Base de données

SSD: Single Shot-multibox Detector

Tables des matières

| | |
|--|----------|
| Introduction générale..... | 1 |
| Chapitre I : Analyse des sentiments | 2 |
| 1- Introduction : | 2 |
| 2- Les besoins de connaître les sentiments des autres : | 2 |
| 3- L'intelligence artificielle(IA) : | 2 |
| 3.1- Une petite histoire sur l'intelligence artificielle (IA) : | 2 |
| 3.2- Définition de L'IA (Intelligence Artificielle) : | 4 |
| 4- L'approche d'apprentissage automatique : | 4 |
| 5- L'apprentissage profond : | 4 |
| 5.1-Définition : | 4 |
| 5.2-Domaine d'utilisation du Deep Learning : | 5 |
| 6- Domaine et tâches d'analyse des sentiments : | 5 |
| 7- Les méthodes d'analyse des sentiments : | 5 |
| 8- Les différentes étapes pour l'analyse visuelle des sentiments : | 5 |
| 9- L'analyse des sentiments à travers une image/Vidéo : | 6 |
| 9.1- Analyse des sentiments à l'aide d'images : | 6 |
| 9.1.1- Les réseaux de neurones artificiels (ANN) : | 6 |
| 9.1.2- Les réseaux de neurones convolutifs (CNN) : | 7 |
| 9.2- Analyse de sentiments à partir d'une vidéo : | 8 |
| 10- Conclusion : | 8 |
| Chapitre II : Conception..... | 9 |
| 1- Introduction : | 9 |
| 2- Schéma global pour la reconnaissance des émotions : | 9 |
| 3- Détection du visage : | 9 |
| 3.1- Détecteur Facial basé sur l'apprentissage profond : | 9 |
| 4- Détection des yeux et de la bouche : | 10 |
| 4.1- Points de repères du visage : | 10 |
| 5- Méthode d'apprentissage : | 11 |
| 5.1- L'apprentissage par transfert : | 11 |
| 6- L'approche multimodale : | 12 |
| 7- Caractéristique du modèle: | 12 |
| 7.1- Choix du modèle : | 12 |
| 7.1.1- Construction du premier modèle : | 12 |
| 7.1.2- Construction du deuxième modèle : | 14 |
| 7.2- Entraînement du modèle : | 16 |

| | |
|--|-----------|
| 8- Conclusion : | 17 |
| Chapitre III : Tests et implémentations | 18 |
| 1- Introduction : | 18 |
| 2- Création de la base de données : | 18 |
| 2.1- Manipulation de la base de données : | 18 |
| 2.2- Extraction des caractéristiques de la base de données : | 19 |
| 2.3- Etiquetage des images : | 19 |
| 3- Description du Matériel, Logiciel et Bibliothèques utilisés : | 19 |
| 3.1- Matériels : | 19 |
| 3.2- Logiciel : | 20 |
| 3.2.1- définition du langage de programmation python : | 20 |
| 3.3- Bibliothèques : | 20 |
| 4- Environnement de développement : | 21 |
| 4.1- Google Colaboratory : | 21 |
| 4.2- Visual Studio Code: | 22 |
| 5- Modèles entraînés : | 22 |
| 6- Quelques Essaies sur les modèles : | 22 |
| 6.1- Modèle du visage : | 22 |
| 6.2- Modèle des yeux : | 23 |
| 6.3- Modèle de la bouche : | 23 |
| 7- Tests et Résultats : | 23 |
| 7.1- Notre algorithme : | 23 |
| 7.2- Résultats obtenus : | 23 |
| 8- Conclusion : | 27 |
| Conclusion générale | 28 |
| Bibliographies : | i |

Liste des figures

| | |
|---|----|
| Figure 1-1 : Réseaux de neurones. | 5 |
| Figure 1-2 Système de reconnaissance des expressions faciales. | 6 |
| Figure 1-3 : réseaux neuronaux classiques pour reconnaissance d'image. | 7 |
| Figure 1-4 : l'architecture des réseaux de neurones convolutifs. | 7 |
| Figure 1-5 : réseau CNN. | 8 |
| Figure 2-1 : Conception générale du système. | 9 |
| Figure 2-2: détection du visage en utilisant un réseau de neurones profonds. | 10 |
| Figure 2- 3: Points de repères du visage..... | 10 |
| Figure 2- 4: Détection des yeux..... | 11 |
| Figure 2- 5: Détection de la bouche. | 11 |
| Figure 2- 6: L'apprentissage par transfert. | 12 |
| Figure 2- 7: Architecture de base du modèle inception-resnet-v1..... | 14 |
| Figure 2- 8: Les composants du stem. | 14 |
| Figure 2- 9: Architecture du modèle faceNet-pytorch. | 14 |
| Figure 2- 10: Architecture de base du modèle MobileNetV..... | 15 |
| Figure 2- 11: Architecture du modèle des yeux..... | 16 |
| Figure 2-12: Architecture du modèle de la bouche. | 16 |
| Figure 3-1 : étiquetage des images. | 19 |
| Figure 3-2 : logo Python..... | 20 |
| Figure 3-3 : logo Numpy. | 20 |
| Figure 3-4 : logo Pandas..... | 20 |
| Figure 3-5: logo OpenCV. | 21 |
| Figure 3-6 : logo PyTorch. | 21 |
| Figure 3-7 : logo Matplotlib. | 21 |
| Figure 3-8: logo Google Colab..... | 22 |
| Figure 3-9: logo Visual Studio Code..... | 22 |
| Figure 3-10 : Essai sur quelques visages..... | 22 |
| Figure 3-11 : Essai sur les yeux..... | 23 |
| Figure 3-12 : Essai sur la bouche. | 23 |
| Figure 3-13: Résultat sur deux personnes. | 27 |

Liste des tableaux

| | |
|--|----|
| Tableau 3.1-Différentes machines utilisées..... | 19 |
| Tableau 3.2-Présentation des modèles entraînés. | 22 |
| Tableau 3.3- Résultat des tests sur notre base de données du modèle du visage..... | 24 |
| Tableau 3.4- Résultat des tests sur notre base de données du modèle des yeux..... | 24 |
| Tableau 3.5- Résultat des tests sur notre base de données du modèle de la bouche. | 25 |
| Tableau 3.6- Neutre..... | 25 |
| Tableau 3.7- Heureux. | 25 |
| Tableau 3.8- Triste..... | 26 |
| Tableau 3.9- Colère..... | 26 |
| Tableau 3.10- Peur. | 26 |
| Tableau 3.11- Dégout. | 26 |
| Tableau 3.12- Surpris..... | 27 |

Introduction générale

L'analyse des sentiments comprend l'évaluation des opinions, des appréciations, des émotions ou des attitudes envers un sujet, une personne ou une entité. Les expressions peuvent être classées comme positives, négatives ou neutre. Grâce aux progrès récents de l'apprentissage profond, la capacité des machines à analyser les émotions s'est considérablement améliorée. L'utilisation créative des techniques d'intelligence artificielle avancées peuvent être un outil efficace pour effectuer des recherches approfondies.

Le cadre d'analyse visuelle des sentiments peut prédire le sentiment d'une image en analysant son contenu. De nos jours, les gens téléchargent des millions d'images sur internet. Ces images jouent un rôle crucial dans l'expression des émotions des utilisateurs en ligne, par conséquent, l'analyse des sentiments d'images est devenue importante dans le domaine de la recherche multimédia en ligne sur les métadonnées. Plusieurs travaux de recherche se concentrent sur l'analyse des sentiments des contenus textuels. Cependant, peu de recherches ont été menées pour développer des modèles capables de prédire le sentiment d'un contenu visuel.

De ce fait, le but de notre travail est de décrire les étapes permettant d'analyser les sentiments à partir d'une vidéo.

Pour cerner ces étapes, notre approche va tourner autour de la question suivante :

- Comment procède-t-on pour analyser les sentiments à partir d'un contenu visuel ?

Pour ce faire, ce travail sera divisé en trois chapitres :

- Dans le premier chapitre, nous décrirons l'analyse des sentiments, l'intelligence artificielle, l'approche d'apprentissage automatique, l'apprentissage profond, et les types de réseaux neuronaux.
- Dans le deuxième chapitre, nous décrirons les différentes étapes suivies pour étudier les sentiments.
- Dans le troisième chapitre, nous présenterons notre étude de cas. Il sera question de la création de la base de données, ainsi que les tests et les résultats de notre travail.

Chapitre I : Analyse des sentiments

Chapitre I : Analyse des sentiments

1- Introduction :

Dans la littérature, l'analyse des sentiments est connue sous le nom «Opinion Mining» et elle est récemment devenue un domaine en plein développement en raison de ses nombreuses applications et ses utilisations comme : la consultation des avis sur les produits, la détection de spam, l'analyse et la surveillance des opinions pour améliorer les produits (matériels ou intellectuels) ou l'étude de marché...

Dans ce chapitre, nous présenterons l'analyse des sentiments, l'intelligence artificielle, l'apprentissage automatique, l'apprentissage profond et les types des réseaux neuronaux.

2- Les besoins de connaître les sentiments des autres :

Connaître l'opinion des autres a toujours été un élément d'information important durant le processus de décision. Avant de prendre des décisions, les gens s'intéressent énormément aux avis des autres personnes dans différents domaines. Ils consultent les opinions des autres constamment avant d'effectuer un achat par exemple. Aujourd'hui, nous pouvons découvrir les opinions et les expériences de très grand nombre de personnes qui ne sont ni nos amis, ni les experts de domaines, mais des gens qui peuvent avoir les mêmes goûts que nous et donc leurs avis peuvent être très utiles pour nous avant d'avoir notre propre avis sur un sujet donné.

En outre, l'expression faciale est une source de communication très importante dans les relations interpersonnelles. Les expressions faciales et autres gestes véhiculent des indices non verbaux dans les interactions en face à face. Ces derniers peuvent aider l'auditeur à comprendre le sens des mots prononcés. Selon Mehrabian[1], dans la communication en face à face entre les personnes, 7% du message concerne les sentiments et les attitudes dans les mots prononcés, 38% dans la façon dont les mots sont prononcés et 55% dans les expressions faciales. Ces dernières aident beaucoup à montrer les émotions pour une meilleure communication.

3- L'intelligence artificielle(IA) :

3.1- Une petite histoire sur l'intelligence artificielle (IA) :

L'histoire de l'intelligence artificielle (IA) est passée par des étapes si importantes, en voici quelques une [2]:

La période 1943-1955, appelée gestation de l'IA, était considérée comme les débuts de l'intelligence artificielle (même si le terme n'existait pas encore). Les travaux de McCulloch et Pitts ont introduit un modèle de neurones artificiels en 1943. Quelques années après, Hebb proposa une règle pour modifier des connexions entre neurones [3]. Minsky et Edmonds [4] construisirent le premier réseau de neurones. Ce fut aussi durant cette

Chapitre I : Analyse des sentiments

période que Turing publia son fameux article dans lequel, il introduit le test de Turing[5]. Durant l'année 1956, qui est considérée comme l'année de naissance de l'IA, a permis à un petit groupe d'informaticien intéressés par l'étude de l'IA, de se réunir pour une conférence sur ce thème. Cette conférence permit de poser les fondements de l'IA.

La période 1952-1969, fut une période très active pour le domaine de l'IA, car un grand nombre de programmes furent développés pour résoudre des problèmes d'une grande diversité. McCarthy [6] publia un article devenu célèbre dans lequel, il traite des programmes qui ont du sens commun. La recherche sur les réseaux de neurones fut également poursuivie. Ce fut aussi l'époque du Shakey, le premier robot à être capable de raisonner sur ses propres actions.

La période 1966-1973, de grandes déceptions se produisirent durant ces années. Des chercheurs réalisèrent un traducteur automatique, mais se sont vite rendu compte que leur approche purement syntaxique n'étaient pas suffisante (pour bien traduire un texte, il faut d'abord le comprendre). Cet échec a provoqué l'annulation en 1966 de tout le financement du gouvernement américain pour les projets de traduction automatique. De grandes déceptions se produisirent également lorsque les chercheurs en IA essayèrent d'appliquer leurs algorithmes aux problèmes de grande taille, et découvrirent alors qu'ils ne fonctionnaient pas, par manque de mémoire et de puissance de calcul. Ce fut une des critiques adressée à l'IA dans le rapport de Lighthill de 1973, qui provoqua l'arrêt du financement de la quasi-totalité des projets en IA de Grande Bretagne. Et comme si cela ne suffisait pas, Minsky et Papert prouvèrent dans leur livre « Perceptrons » de 1969[7] que les réseaux de neurones de l'époque ne pouvaient pas calculer certaines fonctions pourtant très simples, ce qui mit en cause toute la recherche en apprentissage automatique, entraînant une crise dans cette branche de l'IA.

La période 1969-1979, fut la période des Systèmes Experts. Le premier système expert, appelé DENDRAL, fut créé en 1969 pour la tâche spécialisée consistant à déterminer la structure moléculaire d'une molécule étant donné sa formule et les résultats de sa spectrométrie de masse. DENDRAL, comme tous les systèmes experts, est basé sur un grand nombre de règles élaborées par des experts humains. Après le succès du DENDRAL, d'autres Systèmes Experts furent créés, notamment le système MYCIN, qui réalisait un diagnostic des infections sanguines. Avec 450 règles, MYCIN réussissait à diagnostiquer à un niveau proche des experts humains et considérablement meilleur que celui des jeunes médecins.

Au début des années 80 du siècle dernier, l'IA est rentrée dans l'industrie. En effet l'entreprise DEC Commença à utiliser un système expert d'aide à la configuration de systèmes informatiques, ce qui leur permit d'économiser des dizaines de millions de dollars chaque année. Beaucoup de grandes entreprises commencèrent alors à s'intéresser à l'IA et à former leurs propres équipes de recherche. Les Etats-Unis et le Japon financèrent de gros projets en IA, et la Grande Bretagne relança son programme de financement.

Chapitre I : Analyse des sentiments

Au milieu des années 80 du siècle dernier, fut le retour des réseaux de neurones. Effectivement quatre groupes de chercheurs découvrirent indépendamment la règle d'apprentissage « back-propagation » [8] qui permit le développement de réseaux de neurones capables d'apprendre des fonctions très complexes. Depuis, l'apprentissage automatique est devenu l'un des domaines les plus actifs de l'IA, et a été appliqué avec succès à de nombreux problèmes pratiques (comme par exemple la fouille de données). À partir de 1987, L'intelligence artificielle est devenue au fil du temps une matière scientifique de plus en plus rigoureuse et formelle. La plupart des approches étudiées aujourd'hui sont basées sur des théories mathématiques ou des études expérimentales plutôt que sur l'intuition, et sont appliquées plus souvent aux problèmes issus du monde réel.

3.2- Définition de L'IA (Intelligence Artificielle) :

L'intelligence Artificielle ou IA, c'est la pratique qui tend à faire penser les machines. C'est un domaine scientifique qui cherche à résoudre des problèmes logiques ou algorithmiques et plus généralement elles constituent des dispositifs imitant ou remplaçant les êtres-humains.

4- L'approche d'apprentissage automatique :

L'apprentissage machine (ML : Machine Learning) [9] est un sous-domaine de l'intelligence artificielle, qui donne à un système une capacité de compréhension grâce à ses algorithmes. Il est basé sur l'idée de faire apprendre des algorithmes à partir de données et de faire des prédictions avec ces données et par cela les ordinateurs apprennent à résoudre des tâches spécifiques, sans avoir besoin de les programmer. Il existe 3 types ou catégories de l'apprentissage machine. Le premier s'appelle l'apprentissage supervisé où, après avoir présenté les données et les résultats souhaités aux ordinateurs, ils auront la capacité de faire des prédictions pour de nouvelles données d'entrée. Le second est l'apprentissage non supervisé où l'on ne donne à l'ordinateur que les données et où il doit trouver une structure avec un sens par lui-même sans l'intervention d'une supervision extérieure. Elle dépend principalement du clustering. Le processus d'apprentissage se fait par l'observation et non par l'exemple. Le troisième est l'apprentissage par renforcement où la machine se comporte comme un agent qui apprend de son environnement d'une manière interactive jusqu'à ce qu'il découvre les comportements qui produisent des récompenses.

5- L'apprentissage profond :

5.1-Définition :

Deep Learning ou l'apprentissage profond [10] est une méthode d'IA issue du concept de machine learning. Cette méthode d'apprentissage profond se base plus spécifiquement sur la notion de réseau de neurone artificiel.

L'apprentissage profond est un ensemble de méthodes d'apprentissage automatique tentant de modéliser avec un haut niveau d'abstraction des données grâce à des architectures articulées de différentes transformations non linéaires.

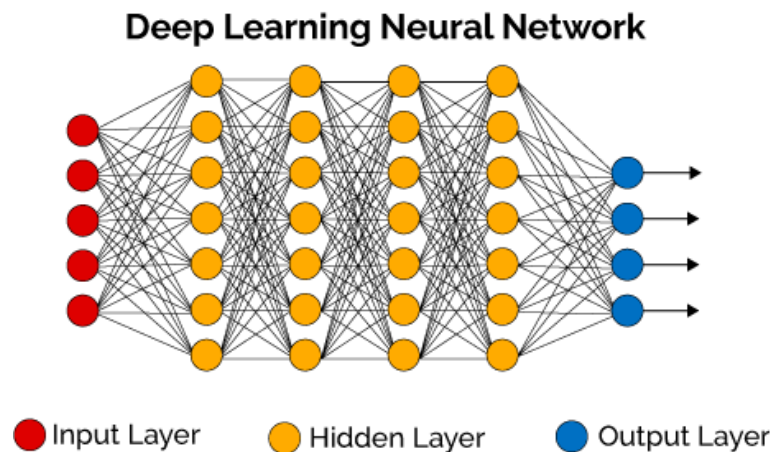


Figure 1-1 : Réseaux de neurones.

5.2-Domaine d'utilisation du Deep Learning :

Après avoir été préalablement entraîné sur une base d'exemples, un réseau de neurone artificiel pourra être utilisé dans la reconnaissance d'images, pour identifier l'humeur par exemple. Le deep learning est également très utilisées dans le traitement automatique du langage (NLP), la cyber sécurité, la bio-informatique, le traitement d'images, diagnostic médical, traduction automatique... etc.

6- Domaine et tâches d'analyse des sentiments :

L'analyse des sentiments a de nombreuses applications dans plusieurs domaines de recherche et de développement, tels que la détection du mensonge, l'analyse du comportement, les systèmes de surveillance, les transports, les études de satisfaction des utilisateurs pour la diffusion et les systèmes de détections de l'humeur des conducteur... etc.

7- Les méthodes d'analyse des sentiments :

Il existe deux grandes catégories :

- **L'analyse lexicale** : ou l'approche lexicque, définit un ensemble de règles dans un type de langage de programmation qui identifie la subjectivité, la polarité ou le sujet d'une opinion.
- **Apprentissage automatique** : reposent sur des techniques de machine learning. La tâche d'analyse des sentiments est généralement modélisée comme un problème de classification dans lequel un classificateur est alimenté avec un texte et renvoie la catégorie correspondante.

8- Les différentes étapes pour l'analyse visuelle des sentiments :

Elle comporte 3 étapes :

Chapitre I : Analyse des sentiments

- **Etape 1 :** la détection du visage. Cette phase consiste à détecter le visage à partir d'une image/vidéo en entrée et il existe plusieurs méthodes. C'est la première étape dans un système de reconnaissance des émotions faciales.
- **Etape 2 :** l'extraction des caractéristiques, elle consiste à extraire les caractéristiques des images de visage détectées. Cette étape représente un visage avec un ensemble de vecteurs de caractéristiques qui décrit les traits du visage tel que la bouche, les yeux et le nez avec leurs distributions géométriques.
- **Etape 3 :** la classification est effectuée à la dernière étape d'un système de reconnaissances des expressions faciales en utilisant un classifieur.

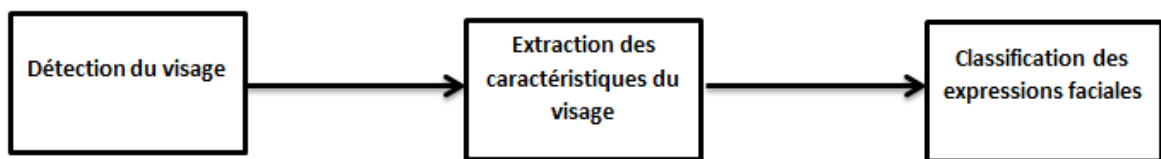


Figure 1-Error! No text of specified style in document.-2 Système de reconnaissance des expressions faciales.

9- L'analyse des sentiments à travers une image/Vidéo :

Les informations textuelles ne sont jamais suffisantes pour détecter l'avis d'une personne, surtout en cas de sarcasme ou d'ambiguïté... Par exemple, il est impossible de reconnaître le sentiment d'une phrase sarcastique "Belle chemise" comme un sentiment négatif en considérant uniquement l'information textuelle. Cependant, si le système peut accéder à la modalité visuelle, il peut facilement détecter les gestes sarcastiques du locuteur et les classer avec la polarité du sentiment négatif.

9.1- Analyse des sentiments à l'aide d'images :

L'analyse des sentiments à l'aide d'images consiste à déterminer si une image donnée est positive, négative ou neutre, il existe plusieurs méthodes de classification dans le domaine de la reconnaissance des expressions faciales. On retrouve :

9.1.1- Les réseaux de neurones artificiels (ANN) :

Les réseaux de neurones artificiels [11] sont tout simplement une imitation des fonctions d'un neurone dans le cerveau humain pour résoudre des problématiques d'apprentissages de la machine.

Les réseaux de neurones artificiels permettent de créer un classifieur qui est le résultat obtenu d'une première phase d'apprentissage faite. Ils peuvent prendre des formes

Chapitre I : Analyse des sentiments

différentes selon l'objet de la donnée qu'il traite et selon sa complexité et la méthode de traitement de la donnée.

Les réseaux neuronaux sont remarquables par leur capacité d'adaptation, ils se modifient eux-mêmes en fonction de l'entraînement initial et les exécutions suivantes leur apportent encore plus d'informations sur le monde qui les entourent.

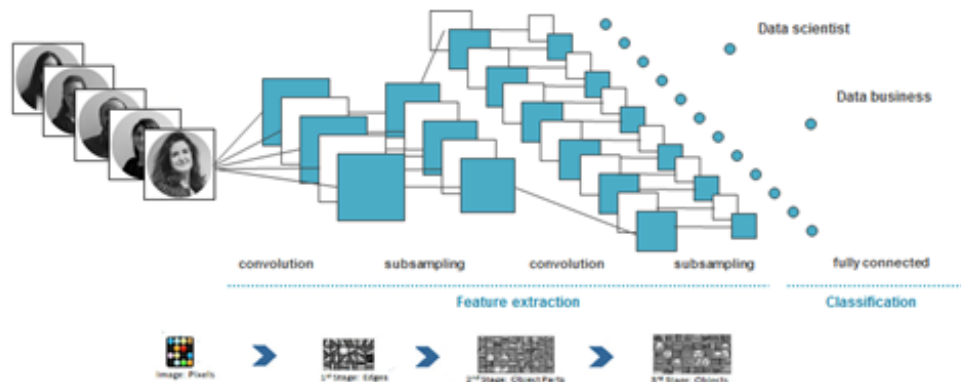


Figure 1-3 : réseaux neuronaux classiques pour reconnaissance d'image.

9.1.2- Les réseaux de neurones convolutifs (CNN) :

Le CNN (Convolutional Neural Network) [12] est un réseau de neurones spécialisé pour le traitement d'images. Son architecture est alors plus spécifique : elle est composée de deux parties. Une première partie que l'on appelle la partie convolutive du modèle et la seconde partie, que l'on va nommer la partie classification du modèle.

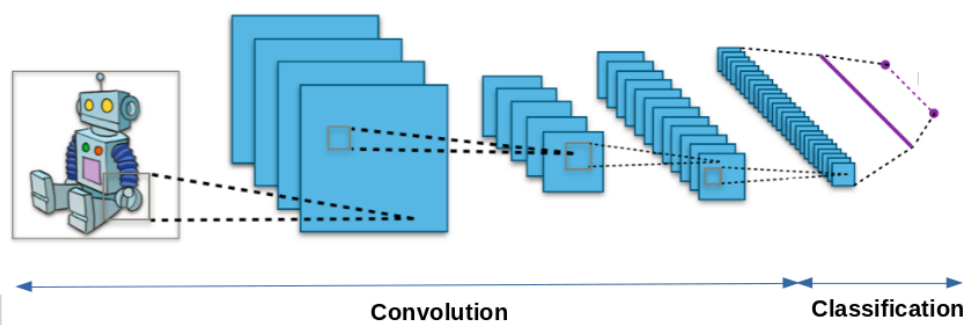


Figure 1-4 : l'architecture des réseaux de neurones convolutifs.

« Le CNN compare les images fragment par fragment. Les fragments qu'il recherche sont appelés les caractéristiques. En trouvant des caractéristiques approximatives qui se ressemblent à peu près dans 2 images différentes. Le CNN et en particulier sa partie convolutive permet de diminuer le nombre de poids à calculer dans les modèles. En effet, la convolution va avoir pour effet de réduire la dimension de la

Chapitre I : Analyse des sentiments

carte de caractéristique que l'on obtient après convolution. Si l'on répète ce processus plusieurs fois, en prenant comme nouvelle entrée la sortie de la convolution précédente, nous allons diminuer de plus en plus la taille de la carte de caractéristiques et donc nous diminuons aussi le nombre de calculs. »

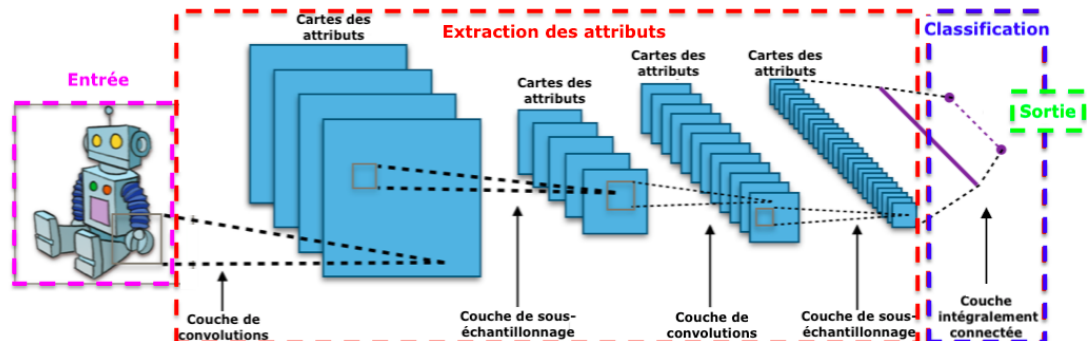


Figure 1-5 : réseau CNN.

9.2- Analyse de sentiments à partir d'une vidéo :

L'analyse à l'aide d'une vidéo ne diffère pas beaucoup de l'analyse à partir d'une image, puisqu'une vidéo est tout simplement une série d'images, recadrer le visage par exemple puis le passer au modèle pré-entraîné qui est construit de la même manière dans [9.1] pour prédire si l'image est positive, négative ou neutre.

10- Conclusion :

Dans ce chapitre, nous avons parlé des performances de l'apprentissage profond. Nous avons présenté aussi les différentes utilisations du Deep Learning ainsi que l'utilité des réseaux de neurones. Nous avons également montré l'importance de l'analyse des sentiments, d'autant plus les domaines, tâches ainsi que les méthodes utilisés par cette dernière.

Afin de pouvoir identifier les émotions de chaque personne nous allons devoir les classer suivant des méthodes qu'on va expliquer dans le chapitre prochain.

Chapitre II : Conception

Chapitre II : Conception

1- Introduction :

Ce chapitre présentera l'ensemble des méthodes qui ont été utilisées lors de la réalisation de ce projet, à savoir les méthodes suivies pour la détection du visage, des yeux et de la bouche. Nous avons aussi expliqué les modèles de la détection de l'expression du visage, ainsi que quelques justifications de choix qui vont être discuté.

2- Schéma global pour la reconnaissance des émotions :

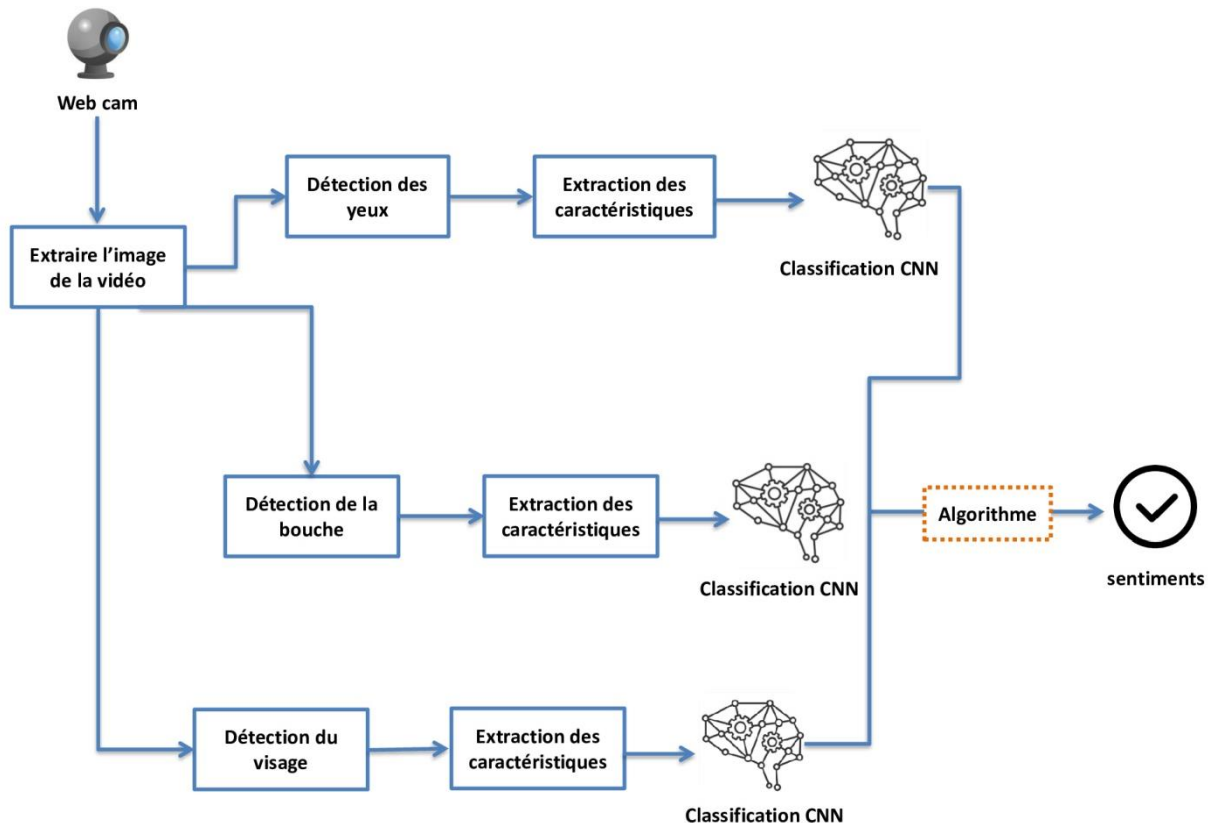


Figure 2-1 : Conception générale du système.

3- Détection du visage :

Il existe un très grand nombre de méthodes permettant d'accomplir la 1ère étape de la détection du visage sur les images ou vidéos telles que les machines à vecteurs de support, les classificateurs Naïve Bayes [13]. Cependant, nous avons utilisé une méthode basée sur l'apprentissage profond (DNN).

3.1- Détecteur Facial basé sur l'apprentissage profond :

En effet, pour que les réseaux de neurones profonds puissent détecter un ou plusieurs visages sur une image, il est nécessaire de les entraîner. Pour cela, il faut collecter ou bien, constituer une base de données visuelle qui va être utile pour l'entraînement.

Chapitre II : Conception

Le but final de cet entraînement est que l'algorithme puisse réaliser des prédictions après avoir analysé l'image en entrée, en d'autres mots, indiqué si le visage est présent ou pas. Dans notre cas, nous avons utilisé un modèle entraîné sur la base de données ImageNet [14].

Le détecteur de visage DNN est une approche d'apprentissage profond qui est basée sur le détecteur Single Shot-Multibox (SSD) [15]. Il utilise l'architecture ResNet-10 comme colonne vertébrale. Cela nous a permis de détecter plusieurs visages à la fois comme le montre la **Figure 2-2**. Il nous a également donné la possibilité de détecter des visages partiellement visibles.

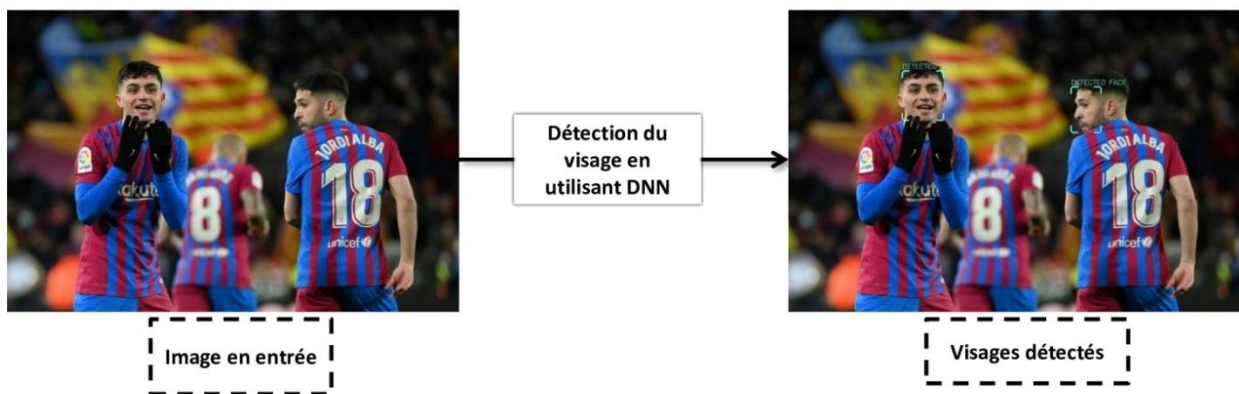


Figure 2-2: détection du visage en utilisant un réseau de neurones profonds.

4- Détection des yeux et de la bouche :

Dans cette partie, nous avons opté pour une méthode basée sur les points de repères du visage.

4.1- Points de repères du visage :

Les points de repères du visage sont un ensemble de points clés sur les images du visage. Ces points sont définis par leurs coordonnées réelles (x,y) comme le montre la Figure 2-3. Nous pouvons accéder aux yeux, à la bouche, au nez... etc, en utilisant ces points.

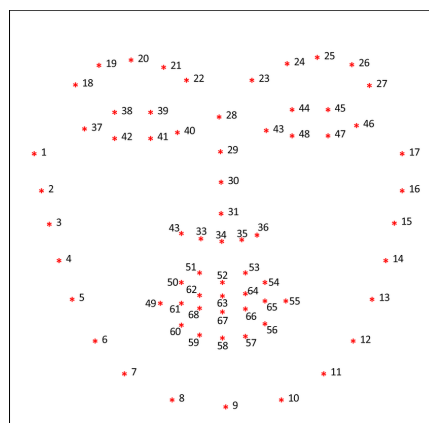


Figure 2- 3: Points de repères du visage.

Chapitre II : Conception

Dans notre cas, nous avons utilisé un point de repère facial qui se compose de 68 points clés. Nous avons donc été capables d'accéder aux yeux en utilisant les points clés de 36 à 48.



Figure 2- 4: Détection des yeux.

Et à la bouche simplement en utilisant les points clés de 49 à 68.



Figure 2- 5: Détection de la bouche.

5- Méthode d'apprentissage :

Pour l'apprentissage, nous allons suivre la méthode de l'apprentissage par transfert.

5.1- L'apprentissage par transfert :

L'apprentissage par transfert ou transfer learning [16], est une méthode d'apprentissage automatique qui désigne l'ensemble des méthodes qui permettent de transférer les connaissances acquises à partir de la résolution de problèmes données pour traiter un autre problème.

Les modèles utilisés dans l'apprentissage profond (Deep learning) nécessitent des temps de calcul élevés et des ressources importantes. Or, en utilisant des modèles pré-

Chapitre II : Conception

entraînés comme point de départ, l'apprentissage par transfert (Transfer learning) permet de développer rapidement des modèles performants et résoudre efficacement des problèmes complexes en vision par ordinateur [17].

En appliquant l'apprentissage par transfert [16] à une nouvelle tâche, on peut obtenir des performances nettement plus élevées qu'en s'entraînant avec une petite quantité de données seulement.

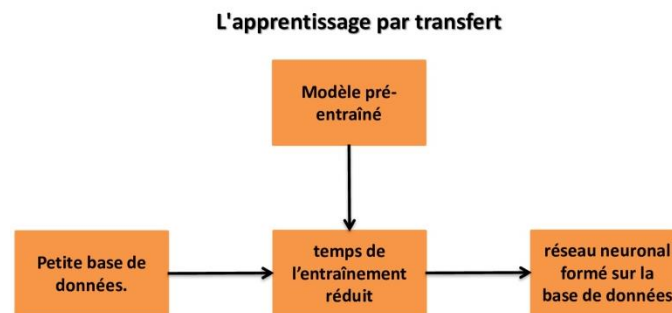


Figure 2- 6: L'apprentissage par transfert.

6- L'approche multimodale :

L'intelligence artificielle multimodale (IA multimodale) [18] est un type d'intelligence artificielle qui consiste à combiner plusieurs modèles d'IA pour la même tâche afin d'obtenir une meilleure précision.

7- Caractéristique du modèle:

7.1- Choix du modèle :

Dans cette étape, nous avons suivis une approche multimodale. En effet, nous allons implémenter trois modèles de sorte qu'ils puissent reconnaître les émotions fondamentales.

En plus, du modèle de visage principal qui reconnaît les sept émotions fondamentales à savoir «Heureux, Triste, Colère, Surpris, Dégoût, Peur et Neutre», nous avons choisi deux autres modèles pour obtenir une meilleure précision. Le premier est basé sur la bouche pour reconnaître les six émotions suivantes «Heureux, Triste, Surpris, Dégoût, Colère et Neutre». Le deuxième modèle est basé sur les yeux pour reconnaître les quatre émotions suivantes «Triste, Colère, Surpris et Neutre». Ces trois modèles constituent notre solution multimodale pour la tâche de reconnaissance des émotions faciales.

7.1.1- Construction du premier modèle :

Nous avons choisi le modèle faceNet-pytorch [19], comme modèle pour le visage, pour sa relativité à notre tâche puisqu'il s'agit d'un modèle de reconnaissance faciale.

Le faceNet-pytorch [19] utilise l'architecture inception-resnet-v1 [20] qui est un réseau hybride inspiré à la fois d'Inception et des performances de Resnet. Le modèle

Chapitre II : Conception

faceNet-pytorch [19] est pré-entraîné sur la base de données vggface2 [21] pour la reconnaissance faciale, il a atteint une précision de 99,65% sur la base de données LFW [22].

Le réseau Inception-ResNet-v1 [20] est constitué de six grands blocs et d'une couche d'entrée au début et d'une couche de sortie à la fin. En commençant par l'entrée, le modèle prend un vecteur de la forme de (3, 299, 299) ce qui signifie qu'il prend des images colorées RVB qui ont au moins 299 de hauteur et 299 de largeur. La couche d'entrée est suivie du bloc stem (**Figure 2-8**), le bloc stem est ensuite suivi de cinq couches de inception-resnet-A (**Figure 2-7**), la sortie du dernier inception-resnet-a est ensuite alimentée par un bloc reduction-a. Ensuite, ils viennent 10 couches de inception-resnet-b suivies d'un bloc reduction-b. Enfin, nous arrivons au dernier bloc qui consiste en 5 couches d'inception-resnet-c. Les dernières couches de inception-resnet-c donnent une sortie de (8x8x1792) cette sortie est alimentée à la couche de sortie (classificateur).

La sortie par défaut de l'inception-resnet-v1 [20] est un Sous-échantillonnage par valeur moyenne (Average Pooling) qui est une technique de sous-échantillonnage qui consiste à remplacer tous les éléments inclus dans une fenêtre d'observation (fenêtre de calcul, ou fenêtre de voisinage) par la moyenne des valeurs des éléments dans cette fenêtre. La sortie est suivie d'un dropout puis d'un dernier linéaire, ce qui donne une sortie de 1000.

Le modèle facenet-pytorch [19] prend une entrée de (3x160x160), c'est-à-dire la taille des images dans la base de données vggface2 sur lequel le modèle est pré-entraîné.

Nous avons appliqué l'apprentissage par transfert au modèle facenet-pytorch [18] et avons donc apporté quelques modifications à la couche de sortie, qui est la couche de sortie de notre modèle :

- Adaptive Sous-échantillonnage par valeur moyenne (Average Pooling) conservé de l'architecture originale.
- Dropout avec la probabilité de 0,8 également conservé de l'architecture originale.
- Linéaire avec 1792 caractéristiques d'entrée et 1036 caractéristiques de sortie, l'entrée est aplatie (**flatten()**) avant d'entrer dans le linéaire et la sortie est normalisée (**normalize()**).
- Linéaire avec 1036 caractéristiques d'entrée et 512 caractéristiques de sortie.
- Fonction d'activation Relu.
- Linéaire avec 512 caractéristiques d'entrée et 7 caractéristiques de sortie (nombre d'émotions prédites par le modèle).
- Fonction d'activation Softmax.

Chapitre II : Conception

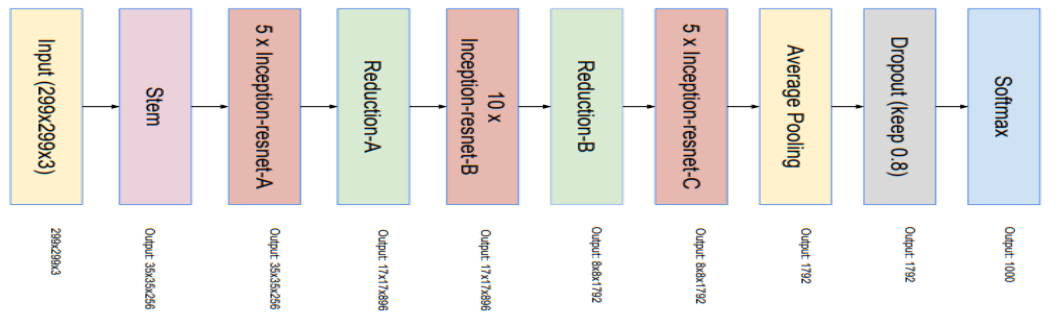


Figure 2- 7: Architecture de base du modèle inception-resnet-v1.

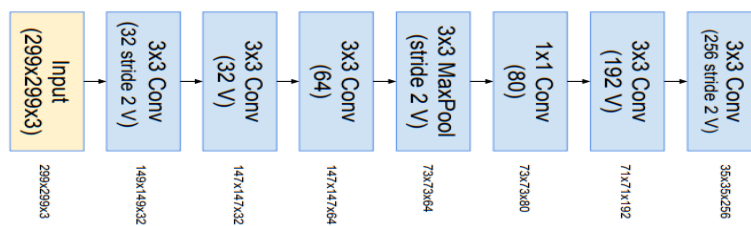


Figure 2- 8: Les composants du stem.

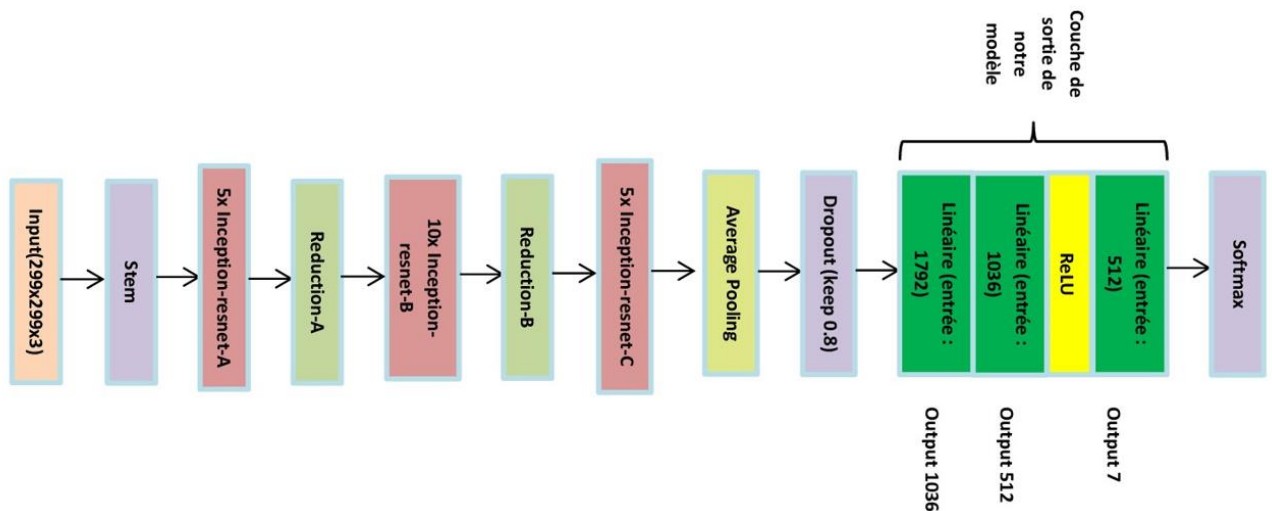


Figure 2- 9: Architecture du modèle faceNet-pytorch.

7.1.2- Construction du deuxième modèle :

Comme notre programme de reconnaissance des émotions sera exécuté sur des webcams (en direct) ou des vidéos enregistrées, nous avons dû prendre en compte le temps d'inférence du modèle car il affecte directement le FPS.

Ceci nous a amené à rechercher un modèle léger qui répond à nos besoins. Nous avons finalement opté pour le MobileNetV2 [23] comme modèle pour les yeux et la bouche.

Le MobileNetV2 [23] est un modèle qui se compose de 53 couches de convolution et d'un pool moyen. Nous pouvons décomposer l'architecture du MobileNetV2 [23] en 3 parties.

Chapitre II : Conception

Tout d'abord, nous avons l'entrée qui est suivie par la couche de convolution. Ensuite, 17 couches du bloc de goulot d'étranglement, ces dernières, peut être soient un bloc résiduel inversé [24] ou un bloc résiduel de goulot d'étranglement [25] ou un bloc Stride 1 ou Stride 2 (**Figure 2-11**). La dernière partie est une couche de convolution suivie d'un pool de moyenne et d'une autre couche de convolution.

MobileNetV2 [23] est pré-entraîné sur la base de données ImageNet [14], ce qui signifie qu'il prend en entrée des images RVB colorées (3x224x224) et produit 1000 classes.

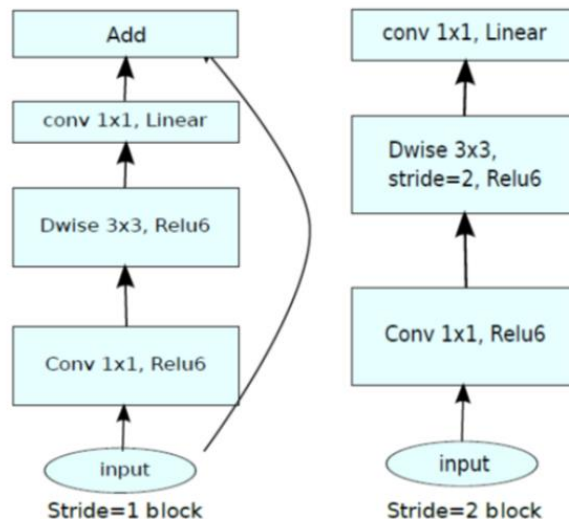


Figure 2- 10: Architecture de base du modèle MobileNetV.

Pour appliquer l'apprentissage par transfert, nous avons remplacé le classificateur par défaut du modèle par notre propre classificateur comme ceci :

Classifieur des yeux :

- Dropout avec une probabilité de 0.2 conservé de l'architecture originale.
- Couche linéaire avec 1280 caractéristiques d'entrée et 512 caractéristiques de sortie.
- Fonction d'activation ReLU
- Dropout avec une probabilité de 0,4.
- Couche linéaire avec 512 éléments entrants et 4 éléments sortants (nombre d'émotions considérées pour les yeux).
- Fonction d'activation Softmax.

Chapitre II : Conception

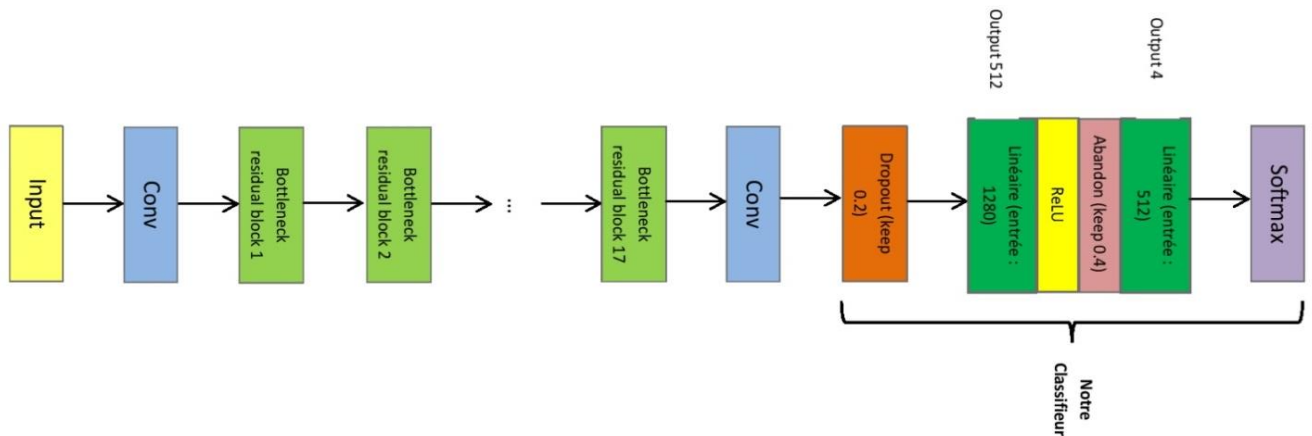


Figure 2- 11: Architecture du modèle des yeux.

Classifieur de la bouche :

- Dropout avec une probabilité de 0,2 conservé de l'architecture originale.
- Couche linéaire avec 1280 caractéristiques d'entrée et 512 caractéristiques de sortie.
- Fonction d'activation ReLU
- Dropout avec une probabilité de 0,4.
- Couche linéaire avec 512 caractéristiques d'entrée et 6 caractéristiques de sortie (nombre d'émotions considérées pour la bouche).
- Fonction d'activation Softmax.

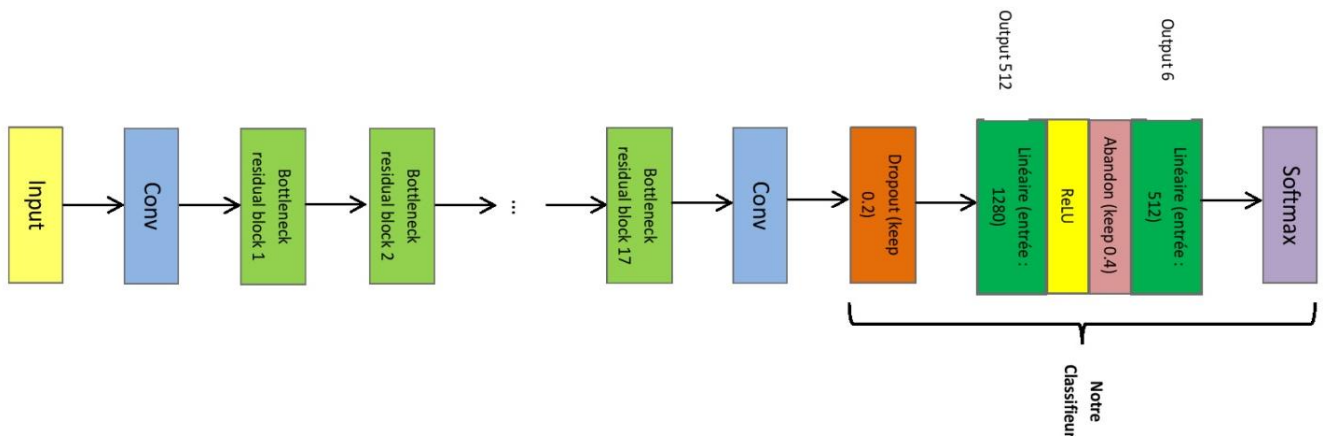


Figure 2-12: Architecture du modèle de la bouche.

7.2- Entraînement du modèle :

L'entraînement a été effectué deux fois :

Le premier entraînement nommé **Extraction des primitives (Feature Extractraction)**, pour extraire les caractéristiques des images, il ne se fait que sur la partie ajoutée à notre modèle, sans toucher la partie haute vu que les poids sont déjà entraînés.

Le deuxième entraînement nommé **Réglage affiné (Fine-tuning)**, a été effectué sur la totalité du modèle créé, pour mieux calibrer les poids du modèle.

Chapitre II : Conception

8- Conclusion :

Dans ce chapitre, nous avons décrit les différentes étapes utilisées pour la reconnaissance faciale des expressions, la méthode de l'apprentissage qu'on a suivie ainsi que les modèles avec leurs caractéristiques.

Dans le prochain chapitre, nous allons présenter la création de notre base de données, les résultats ainsi que tests de nos modèles.

Chapitre III : Tests et implémentations

Chapitre III : Tests et implémentations

1- Introduction :

Dans le chapitre précédent nous avons présenté les méthodes qui ont été considérées lors de la réalisation de ce projet. Ce chapitre présentera la création de notre base de données, les différentes approches ainsi qu'aux différents résultats de tests de notre projet.

2- Création de la base de données :

En exploitant trois bases de données existantes (RAVDESS, Google facial expression comparison et iSAFE), nous avons généré des ensembles d'images de visages, yeux et bouches correspondant comme ensembles de données d'entraînement.

La base de données audio-visuelle Ryerson de la parole émotionnelle (RAVDESS) (Livingstone et Russo, 2018) [26] est une base de données audio et visuelle de discours et de chants émotionnels. L'ensemble de données a été collecté de 24 acteurs professionnels et comprend huit catégories d'émotions : neutre, calme, heureux, triste, en colère, peur, dégoût et surprise. Toutes les séquences de l'ensemble de données sont disponibles en format visage et voix, visage seul et voix seule. Cependant nous n'utilisons que le format visage seul.

La base de données Google facial expression comparison [27], qui est un ensemble de données à grande échelle sur l'expression faciale, composé de triplets d'images de visage accompagnés d'annotations humaines qui précisent quels sont les deux visages de chaque triplet qui forment la paire la plus similaire en termes d'expression faciale. Chaque triplet de cette base de données a été annoté par six évaluateurs humains ou plus. Cette base de données est destinée à aider les chercheurs qui travaillent sur des expressions, le résumé d'albums photo basé sur les expressions, la classification des émotions... etc.

La base de données iSAFE [28] est une base de données d'expressions faciales contenant des expressions spontanées de participants masculins et féminins d'origine indienne. La base de données est constituée de 428 clips vidéo segmentés des expressions faciales spontanées de 44 volontaires âgés de 17ans à 22ans.

2.1- Manipulation de la base de données :

Dans notre cas, les images vidéo sont extraites toutes les 2 secondes du flux vidéo d'entrées de la base de données (BDD) RAVDESS, pour ensuite les utilisées comme input (entrée) pour la détection. Même chose pour la base de données (BDD) iSAFE, sauf que les images vidéo sont extraites toutes les 1 secondes du flux vidéo d'entrées.

En ce qui concerne la base de données (BDD) Google facial expression comparison, on utilisera directement ces photos comme input.

Chapitre III : Tests et implémentations

2.2- Extraction des caractéristiques de la base de données :

L'extraction de caractéristiques est une composante du traitement de l'image. En effet, on aura besoin d'extraire le visage, les yeux ainsi que la bouche.

2.3- Etiquetage des images :

Une fois les caractéristiques des images sont extraites, nous passons à l'étiquetage. Nous avons utilisé the Face API [29] en premier (l'étiquetage automatique), ceci consiste à décomposer les photos en sept principales parties correspondant à chacune des sept émotions abordées à savoir : neutre, heureux, colère, peur, dégoût, surpris et triste. Ensuite, nous avons étiqueté quelques images manuellement.

Sachant que, the Face API [29] peut effectuer une détection d'émotion pour détecter les émotions. Il est important de noter, cependant, les expressions faciales seules ne représentent pas nécessairement les états internes des personnes.

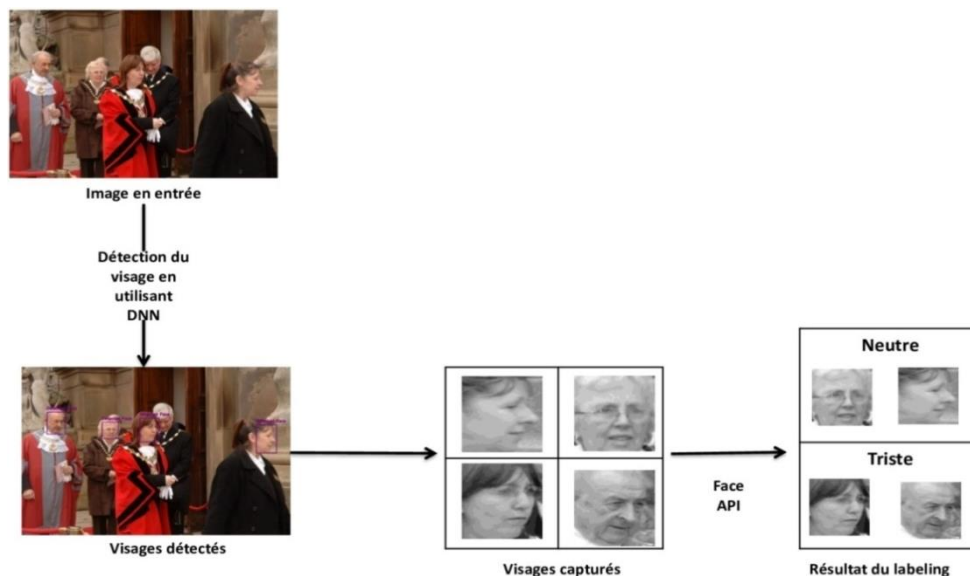


Figure 3-1 : étiquetage des images.

3- Description du Matériel, Logiciel et Bibliothèques utilisés :

3.1- Matériels :

| | Carte Graphique | Marque | Système d'exploitation | CPU | RAM |
|------|------------------------|--------|----------------------------------|------------------|------|
| PC 1 | Intel HD graphics 620 | HP | Windows 10 | Intel i3 7th gen | 4 GO |
| PC 2 | Intel UHD graphics 620 | ASUS | Windows 11 & Linux mint cinnamon | Intel i5 8th gen | 8 GO |

Tableau 3.1-Différentes machines utilisées.

3.2- Logiciel :

Langage de programmation utilisé est : **python**.

3.2.1- définition du langage de programmation python :

Le langage Python [30] est un langage de programmation open source multiplateformes et orienté objet. Grâce à des bibliothèques spécialisées, Python s'utilise pour de nombreuses situations comme le développement logiciel, l'analyse de données, ou la gestion d'infrastructures. Il permet d'exécution du code sur n'importe quel ordinateur.



Figure 3-2 : logo Python.

3.3- Bibliothèques :

- **Numpy** : [31] est le paquetage fondamental pour le calcul scientifique en Python. Il s'agit d'une bibliothèque Python qui fournit un objet tableau multidimensionnel, divers objets dérivés (tels que les tableaux masqués et les matrices), ainsi qu'un assortiment de routines permettant d'effectuer des opérations rapides sur les tableaux, notamment des opérations mathématiques, logiques, de manipulation de formes, de tri, de sélection, d'E/S, de transformées de Fourier discrètes, d'algèbre linéaire de base, d'opérations statistiques de base, de simulation aléatoire et bien plus encore.



Figure 3-3 : logo Numpy.

- **Pandas** : [32] est un package Python conçue pour la manipulation et l'analyse de données. Il est à la fois performant, flexible et simple d'utilisation. De manière générale, Pandas excelle aussi pour traiter les données structurées sous forme de tableaux, de matrices ou de séries temporelles.



Figure 3-4 : logo Pandas.

Chapitre III : Tests et implémentations

- **OpenCV** : [33] (Open Source Computer Vision Library) est une bibliothèque logicielle open source de vision par ordinateur et d'apprentissage automatique. OpenCV a été construit pour fournir une infrastructure comme pour les applications de vision par ordinateur et pour accélérer l'utilisation de la perception automatique dans les produits commerciaux.

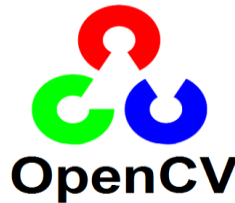


Figure 3-5: logo OpenCV.

- **PyTorch** : [34] est une bibliothèque logicielle Python open source d'apprentissage machine qui s'appuie sur Torch (en anglais). PyTorch permet d'effectuer les calculs tensoriels nécessaires notamment pour l'apprentissage profond (deep learning). Ces calculs sont optimisés et effectués soit par le processeur (CPU) soit, lorsque c'est possible, par un processeur graphique (GPU) supportant CUDA.



Figure 3-6 : logo PyTorch.

- **Matplotlib** : [35] est une bibliothèque Python open source permettant de créer des visualisations de données. Matplotlib proposant plusieurs fonctions simples pour ajouter des éléments tels que des lignes, des images ou des textes aux axes d'un graphique.



Figure 3-7 : logo Matplotlib.

4- Environnement de développement :

4.1- Google Colaboratory :

Google colab [36] est un produit de Google Research. Colab permet à n'importe qui d'écrire et d'exécuter le code Python de son choix par le biais du navigateur. C'est un environnement particulièrement adapté à la machine learning, à l'analyse de données et à l'éducation. En termes plus techniques, Colab est un service hébergé de notebooks Jupyter qui ne nécessite aucune configuration et permet d'accéder gratuitement à des ressources informatiques, dont des GPU.



Figure 3-8: logo Google Colab.

4.2- Visual Studio Code:

Visual Studio Code [37] est un éditeur de code open-source développée par Microsoft supportant un très grand nombre de langages grâce à des extensions.



Figure 3-9: logo Visual Studio Code.

5- Modèles entraînés :

| Modèle pré entraînés | Base de données | Taux de précisions |
|----------------------------------|-----------------|--------------------|
| Facenet-pytorch (pour le visage) | Vggface2 | 0.67 |
| MobileNetV2 (pour les yeux) | ImageNet | 0.83 |
| MobileNetV2 (pour la bouche) | ImageNet | 0.77 |

Tableau 3.2-Présentation des modèles entraînés.

6- Quelques Essaies sur les modèles :

6.1- Modèle du visage :

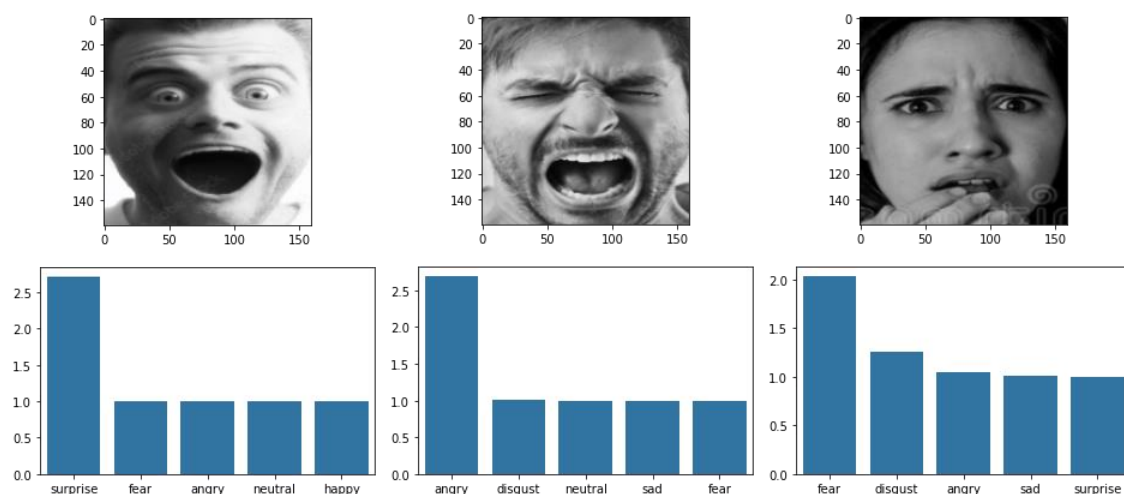


Figure 3-10 : Essai sur quelques visages.

6.2- Modèle des yeux :

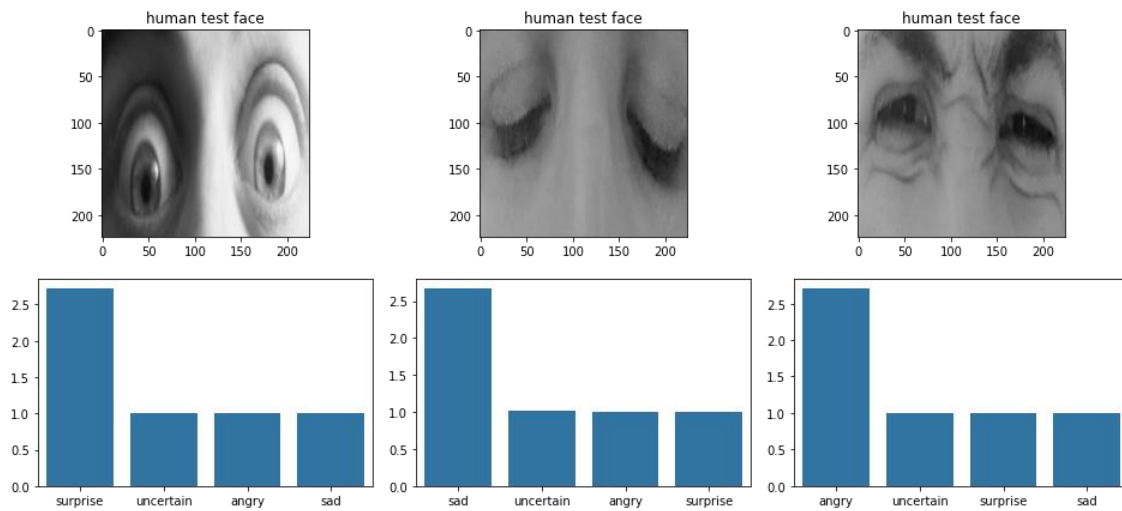


Figure 3-11 : Essai sur les yeux.

6.3- Modèle de la bouche :

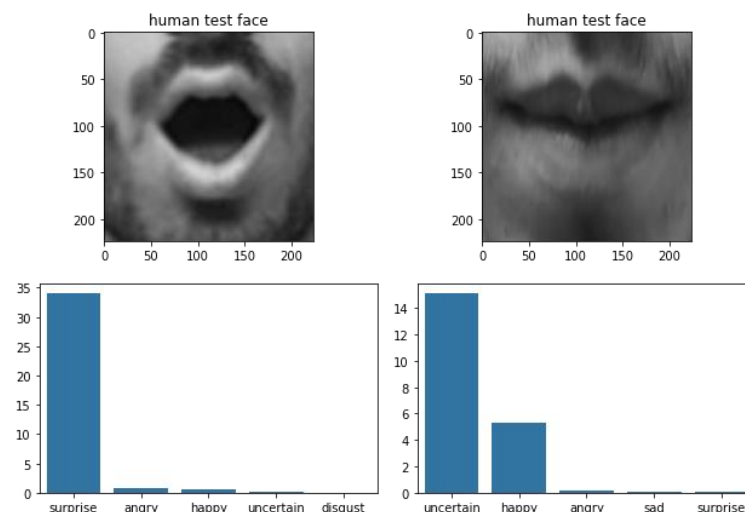


Figure 3-12 : Essai sur la bouche.

7- Tests et Résultats :

7.1- Notre algorithme :

Nous avons développé un algorithme de sorte qu'il puisse combiner les résultats des 3 entrées qu'on a (Visage, Yeux et Bouche) pour ensuite nous donner comme sortie l'émotion.

7.2- Résultats obtenus :

Nous présentons les résultats de nos modèles appliqués sur 2407 images de notre base de données.

Chapitre III : Tests et implémentations

Modèle du visage :

| Emotions | Nombre de support | Taux de précision de chaque émotion |
|----------|-------------------|-------------------------------------|
| Neutre | 369 | 0.89 |
| Heureux | 443 | 0.94 |
| Triste | 284 | 0.53 |
| Colère | 220 | 0.67 |
| Surpris | 200 | 0.69 |
| Dégout | 114 | 0.63 |
| Peur | 134 | 0.37 |

Tableau 3.3- Résultat des tests sur notre base de données du modèle du visage.

- Nous remarquons que l'émotion « neutre » a un taux de précision de 89%
- Nous remarquons que l'émotion « heureux » a un taux de précision de 94%
- Nous remarquons que l'émotion « triste » a un taux de précision de 53%
- Nous remarquons que l'émotion « colère » a un taux de précision de 67%
- Nous remarquons que l'émotion « surpris » a un taux de précision de 69%
- Nous remarquons que l'émotion « dégout » a un taux de précision de 63%
- Nous remarquons que l'émotion « peur » a un taux de précision de 37%

Modèle des yeux :

| Emotions | Nombre de support | Taux de précision de chaque émotion |
|----------|-------------------|-------------------------------------|
| Neutre | 40 | 0.85 |
| Triste | 38 | 0.76 |
| Colère | 41 | 0.95 |
| Surpris | 21 | 0.76 |

Tableau 3.4- Résultat des tests sur notre base de données du modèle des yeux.

- Nous remarquons que l'émotion « neutre » a un taux de précision de 85%
- Nous remarquons que l'émotion « triste » a un taux de précision de 76%
- Nous remarquons que l'émotion « colère » a un taux de précision de 95%
- Nous remarquons que l'émotion « surpris » a un taux de précision de 76%

Chapitre III : Tests et implémentations

Modèle de la bouche :

| Emotions | Nombre de support | Taux de précision de chaque émotion |
|----------|-------------------|-------------------------------------|
| Neutre | 104 | 0.73 |
| Heureux | 133 | 0.84 |
| Triste | 38 | 0.78 |
| Colère | 76 | 0.84 |
| Surpris | 105 | 0.81 |
| Dégout | 67 | 0.61 |

Tableau 3.5- Résultat des tests sur notre base de données du modèle de la bouche.

- Nous remarquons que l'émotion « neutre » a un taux de précision de 73%
- Nous remarquons que l'émotion « heureux » a un taux de précision de 84%
- Nous remarquons que l'émotion « triste » a un taux de précision de 78%
- Nous remarquons que l'émotion « colère » a un taux de précision de 84%
- Nous remarquons que l'émotion « surpris » a un taux de précision de 81%
- Nous remarquons que l'émotion « dégoût » a un taux de précision de 61%

Nous allons présenter les différents résultats obtenus de notre système de reconnaissance des émotions faciales sur différentes personnes :



Tableau 3.6- Neutre.

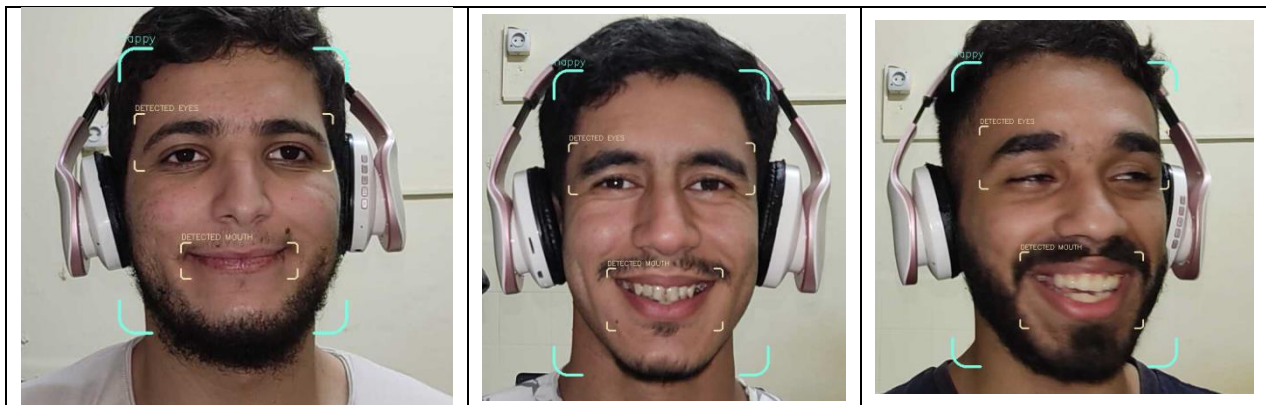


Tableau 3.7- Heureux.

Chapitre III : Tests et implémentations

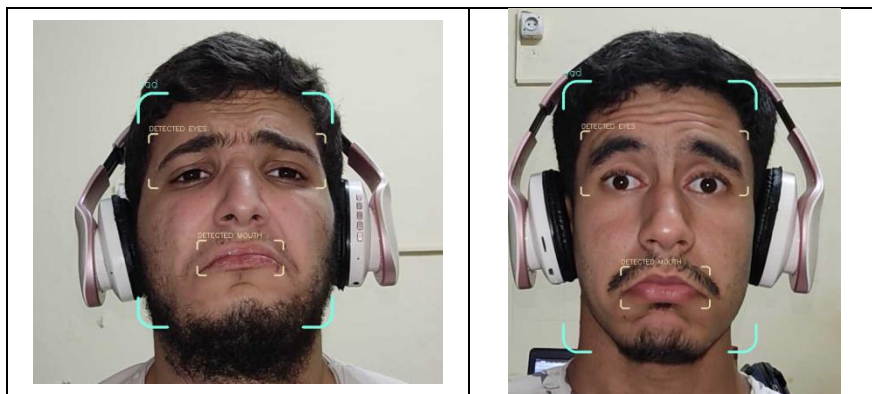


Tableau 3.8- Triste.

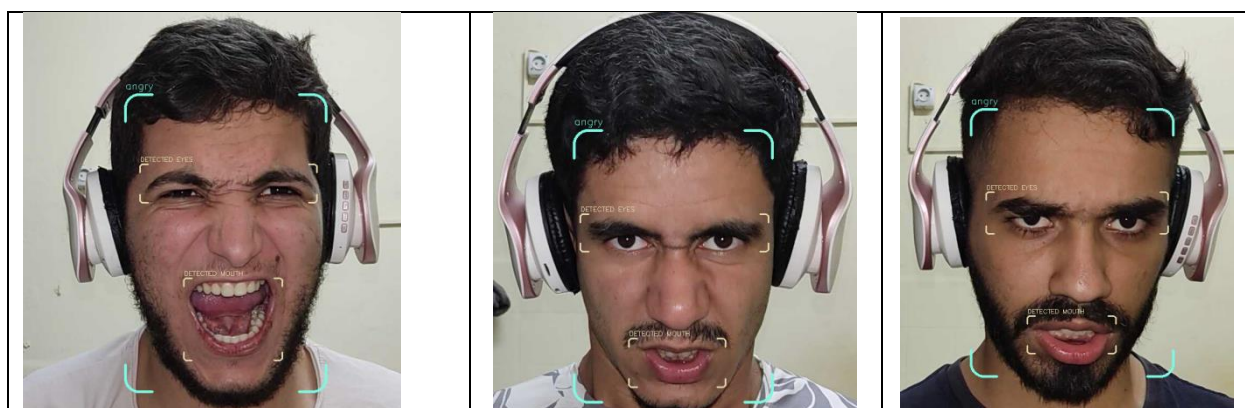


Tableau 3.9- Colère.

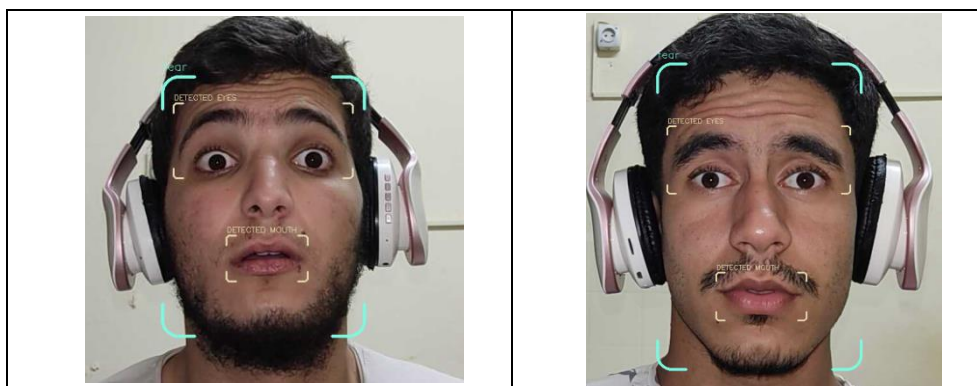


Tableau 3.10- Peur.

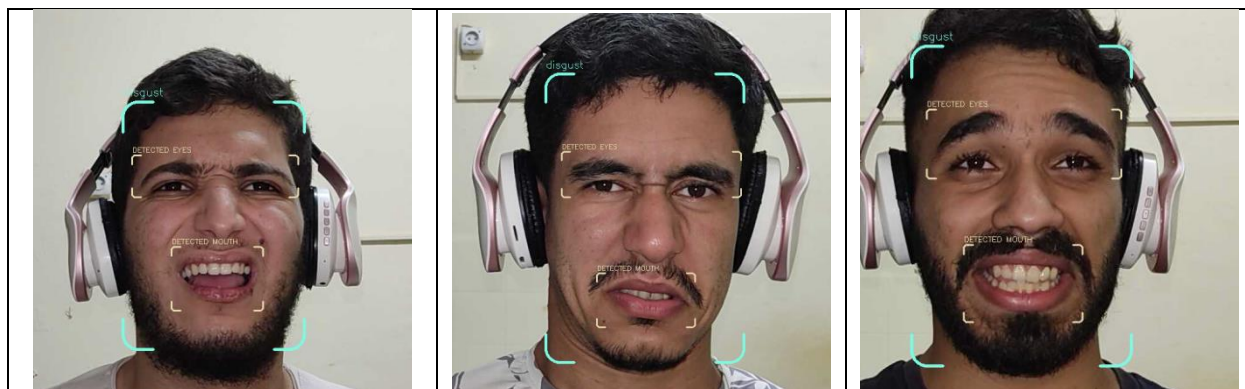


Tableau 3.11- Dégout.

Chapitre III : Tests et implémentations

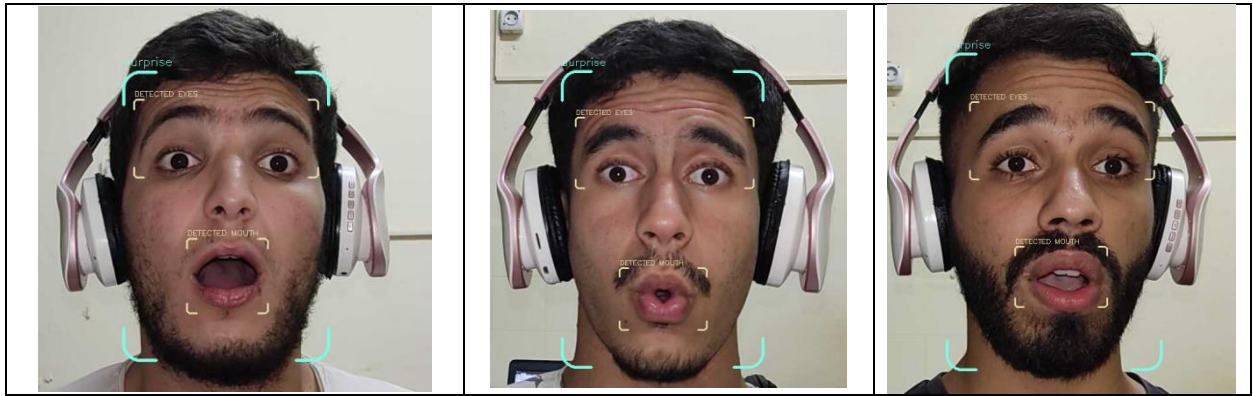


Tableau 3.12- Surpris.



Figure 3-13: Résultat sur deux personnes.

8- Conclusion :

Dans ce chapitre, nous avons d'abord présenté les différentes étapes suivies lors de la création de notre base de données pour l'entraînement de nos modèles. Ensuite, nous avons cité les différentes technologies utilisées pour la réalisation de notre projet. Pour finir, nous avons présenté et discuté les différents résultats obtenus par notre système de reconnaissance d'émotions ainsi que les critères d'évaluation.

Conclusion générale

Les travaux réalisés dans notre thème de mémoire ont porté sur l'analyse des sentiments basés sur la vidéo, en utilisant la méthode d'apprentissage par transfert. Nos recherches à travers les documents consultés, nous ont permis de déceler et de comprendre les diverses facettes concernant la reconnaissance des émotions à partir d'un contenu visuel. Ceci a permis de valider notre réponse sur la problématique formulée dans le cadre de notre mémoire de fin d'études.

Pour cadrer le contexte de notre étude, nous avons proposé une solution composée de deux phases. En premier, la phase de la détection du visage en utilisant une approche d'apprentissage profond (DNN), la détection des yeux en utilisant une méthode de Viola et Jones et la détection de la bouche en utilisant une méthode basée sur les points de repères. En second, nous avons opté pour une méthode d'approche multimodale, afin d'obtenir une meilleure précision en combinant nos trois modèles.

Dans le but d'évaluer notre système, nous avons créé notre base données en combinant trois base de données, nous avons également utilisé trois paramètres d'entrées (Visage, Yeux et Bouche). Ainsi, nous avons conclu qu'en les utilisant nous obtenons des résultats satisfaisants, nous pouvons donc dire que nos objectifs ont été atteint.

Pour finir, plusieurs perspectives sont envisageables, nous citons entre autre :

- Effectuer des évaluations plus approfondies de notre système, en considérant des données plus volumineuses, autrement dit, augmenter le nombre d'images dans notre base de données.
- Ajouter un autre paramètre en entrée (des sourcils), à notre programme, pour obtenir des résultats plus précis.

Bibliographies :

- [1] : Mehrabian, A.: Nonverbal Communication. Transaction Publish-ers, Los Angeles (1972)
- [2] : S.Russell et P.Norvig, "Introduction à l'Intelligence Artificielle", Ces notes ont été préparées pour un cours du niveau L3 offert en télé-Enseignement par l'Université Aix Marseille 1 en 2007-2008 et 2008-2009.
- [3]: D.O. Hebb, The organization of behavior, New York, Wiley, 1949.
- [4]: Minsky et Edmonds sur :
«https://www.pearson.fr/resources/titles/27440100534680/extras/7455_chap01.pdf»
- [5] : Alan Turing, « Computing machinery and intelligence », Oxford University Press.
- [6] : J. McCarthy, Programs with Common Sense [archive]. In Proceedings of the Teddington Conference on the Mechanization of Thought Processes, 756-91. London: Her Majesty's Stationery Office, 1959.
- [7]: Marvin Minsky et Seymour Papert, Perceptrons, 1969, 258 p. (ISBN 0-262-13043-2).
- [8] : **Back-Propagation** Algorithm Which Varies the Number of Hidden Units." Neural Networks, Vol. 4, No. 1, pp.61-66, 1991.
- [9]: « Machine Learning For Absolute Beginners: A Plain English Introduction (Second Edition) » par Oliver Theobald.
- [10]: L'apprentissage profond avec Python de François Chollet. PDF
- [11]: Léon Personnaz et Isabelle Rivals, **Réseaux de neurones** formels pour la modélisation, la commande et la classification, CNRS Éditions, 2003.
- [12]: « Convolutional Neural Networks (LeNet) – DeepLearning 0.1 documentation » [archive], sur DeepLearning 0.1, LISA Lab.
- [13] : Domingos, Pedro & Michael Pazzani (1997), " On the Optimality of the Simple Bayesian Classifier under Zero-One Loss. Pedro Domingos & Michael Pazzani."An empirical study of the *naïve Bayes* classifier".
- [14] : Base de données ImageNet, sur « www.image-net.org »
- [15]: Single-Shot multibox Detector (littéralement, en anglais: détection multi-boîte à passe unique), un type de détecteur employé dans le domaine de la vision par ordinateur.
- [16]: Sinno Jialin Pan et Qiang Yang, « A Survey on Transfer Learning », IEEE Transactions on Knowledge and Data Engineering.
- [17] : L'apprentissage par transfert sur : « <https://datascientest.com/transfer-learning> »

Bibliographie

- [18] : Intelligence artificielle multimodale sur :
« <https://www.aimesoftware.com/fr/multimodalai.html> »
- [19] : Modèle facenet-pytorch sur « <https://github.com/timesler/facenet-pytorch> »
- [20]: Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning.
- [21]: Q. Cao, L. Shen, W. Xie, O. M. Parkhi, A. Zisserman. VGGFace2: A dataset for recognising face across pose and age, International Conference on Automatic Face and Gesture Recognition, 2018. PDF
- [22]: Base de données LFW Labeled faces sur : « <http://vis-www.cs.umass.edu/lfw/> »
- [23]: Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, Liang-Chieh Chen, Google Inc, MobileNetV2: Inverted Residuals and Linear Bottlenecks.
- [24]: bloc résiduel invers sur:
« <https://paperswithcode.com/method/inverted-residual-block> »
- [25]: bloc résiduel de goulot d'étranglement sur :
« <https://paperswithcode.com/method/bottleneck-residual-block> »
- [26]: Livingstone & Russo (2018). The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English.
- [27]: Base de données Google facial expression comparison sur :
« <https://research.google/tools/datasets/google-facial-expression/> »
- [28]: Singh, Shivendra; Benedict, Shajulin (2020). Thampi, Sabu M.; Hegde, Rajesh M.; Krishnan, Sri; Mukhopadhyay, Jayanta; Chaudhary, Vipin; Marques, Oge; Piramuthu, Selwyn; Corchado, Juan M. (eds.). "Indian Semi-Acted Facial Expression (iSAFE) Dataset for Human Emotions Recognition". Advances in Signal Processing and Intelligent Recognition Systems. Communications in Computer and Information Science.
- [29] : FaceAPI sur : « <https://docs.microsoft.com/en-us/connectors/faceapi/> »
- [30]: Introduction de la première édition du livre Programming Python de Mark Lutz [archive], Guido van Rossum 199.
- [31]: Stefan Van Der Walt, S. Chris Colbert et Gaël Varoquaux, « *The NumPy array: a structure for efficient numerical computation* », Computing in Science & Engineering.
- [32]: Andreas C. Müller, Sarah Guido, *Introduction to Machine Learning with Python*, O'Reilly, 2016.

Bibliographie

[33]: Site officiel d'Opencv sur : «<http://opencv.org>».

[34]: Jeremy Howard, Sylvain Gugger, *Deep Learning for Coders with fastai and PyTorch*, O'Reilly, 2020.

[35]: « *Matplotlib for Python Developers - About dependencies* », novembre 2009

[36]: Google Colab sur : «<https://colab.research.google.com/>».

[37]: Frederic Lardinois, « Microsoft Launches Visual Studio Code, A Free Cross-Platform Code Editor For OS X, Linux And Windows »