République Algérienne Démocratique et Populaire
Ministère de l'enseignement supérieur et de la recherche scientifique

Université Laarbi Tébessi - Tébessa

Faculté des Sciences Exactes et des Sciences de la Nature

Département : Mathématiques et Informatique

Mémoire de fin d'études

Pour l'obtention du diplôme de MASTER
Domaine : Mathématiques et Informatique
Filière : Informatique

Option : Systèmes et Multimédias

Thème

Reconnaissance du visage, application aux dessins animés

Présenté Par : Bezza Khouloud

Devant le jury :

Dr.Laimech Lakhder	*MCA*	*Univérsité Larbi Tébessi*	*Examinateur*
Dr.Bennour Akrem	*MCA*	*Univérsité Larbi Tébessi*	*Président*
Dr.Ahmed Zeggari	*MCB*	*Univérsité Larbi Tébessi*	*Encadreur*
Dr.Tahar Gherbi	*MCB*	*Univérsité Larbi Tébessi*	*Co-Encadreur*

Date de soutenance :

11/07/2021

Résumé

Il existe de nombreuses techniques dans les avancées récentes technologies de détection et de reconnaissance de visage humain de l'apprentissage en profondeur (Deep Learning), mais « à quel point fonctionneraient-ils pour les dessins animés ? ». La reconnaissance du visage de dessins animés est un domaine qui est resté largement inexploré jusqu'à récemment, principalement en raison de l'indisponibilité de données à grande échelle et de l'échec des méthodes traditionnelles. Dans ce travail, nous présenterons un programme qui reconnaît le visage de dessin animé en utilisant l'apprentissage par transfert. Puisqu'il n'y a pas d'état des travaux pour les données qui peuvent valider notre travail, nous avons collecté des images de personnages de dessins animés. Le système proposé utilise le modèle Mask RCNN pour recadrer la zone du visage, et par conséquence nous aurons de nouvelles données pour des images qui ne contiennent que les visages, ces données sont utilisé pour entrainer le model inception V3 pour reconnaitre (classifier) le visage de caractère. Le système recadre correctement le visage de personnage et reconnaît le caractère conséquent avec un score de précision de 0.98.

Mots clés : dessin animés, reconnaissance du visage, Deep Learning, Apprentissage par transfer, Mask RCNN, InceptionV3

Abstract

There are many techniques in the recent advancements in deep learning human face detection and recognition technologies, but "how well would they work for cartoons?". Cartoon face recognition is an area that has remained largely unexplored until recently, mainly due to the unavailability of large-scale data and the failure of traditional methods. In this work, we will present a program that recognizes cartoon faces using transfer learning. Since there is no state of work reports for data that can validate our work, we have collected cartoon character images of 6 characters each has 150 images. The proposed system uses the Mask RCNN model to crop the face area, and as a result we will have new data for images that only contain faces, this data is used to train the inceptionV3 model to recognize (classify) the face. Character. The system correctly crops the character's face and recognizes the consistent character with an accuracy score of 0.98.

Keywords: cartoons, face recognition, Deep Learning, Transfer learning, Mask RCNN, InceptionV3.

ملخص

هناك العديد من التقنيات في التطورات الحديثة في تقنيات اكتشاف الوجه البشري والتعرف عليه ، ولكن "ما مدى نجاحها مع الرسوم المتحركة؟". التعرف على الوجوه الكرتونية هي منطقة ظلت غير مستكشفة إلى حد كبير حتى وقت قريب ، ويرجع ذلك أساسًا إلى عدم توفر البيانات على نطاق واسع وفشل الأساليب التقليدية. في هذا العمل ، سنقدم برنامجًا يتعرف على الوجوه الكارتونية باستخدام التعلم الانتقالي. نظرًا لعدم وجود أعمال سابقة للبيانات التي يمكننا استخدامها للتحقق من صحة عملنا ، فقد جمعنا صورًا لشخصيات كرتونية من 6 شخصيات لكل منها 150 صورة.

يستخدم النظام المقترح نموذج Mask RCNN لاقتصاص منطقة الوجه ، ونتيجة لذلك سيكون لدينا بيانات جديدة للصور التي تحتوي على وجوه فقط ، وتستخدم هذه البيانات لتدريب نموذج inceptionV3 للتعرف على (تصنيف) الوجه. حرف. يقوم النظام بقص وجه الشخصية بشكل صحيح ويتعرف على الشخصية المتسقة بدرجة دقة تبلغ 0.98.

الكلمات المفتاحية: الرسوم المتحركة ، التعرف على الوجوه ، التعلم العميق ، نقل التعلم ، Mask r-cnn ، InceptionV3.

Remerciements

"الشكر و الحمد لله"

Merci ***Allah*** *de m'avoir donné la capacité d'écrire et de réfléchir, la force d'y croire, la patience d'aller jusqu'au bout du rêve.*

Je tiens à remercier vivement mon encadreur,

Dr.Ahmed Zeggari *de m'avoir encadré pour réaliser ce travail, pour ses précieux conseils et de m'avoir donné le meilleur de son savoir et aide.*

Je remercie également les membres de jury de nous faire l'honneur de juger mon travail.

Je remercie profondément touts ***les enseignants de département d'informatique*** *et toutes les personnes qui ont contribuées à l'élaboration de ce travail.*

Finalement, je remercie ma grande famille, mes amies, mes collègues de l'université ***Laarbi Tébessi Tebessa*** *et toute la promotion 2021 de l'informatique.*

Liste de figure

Table de matières

Introduction générale

L'essor massif des médias numériques au cours des dernières années a suscité la tendance des thèmes animés, s'adaptant à un large éventail d'aspects de la vie tels que l'offre d'écoles à domicile pour les enfants, l'amélioration de l'enseignement et des résultats scolaires ainsi que la représentation de son opinion sur les pratiques de la société (par le biais de politiques dessins animés) et photojournalisme). Les graphiques qui apparaissent dans les animations ont des caractéristiques exagérées de manière à ce que ces visages s'écartent souvent des caractéristiques implicites d'un être humain (telles qu'une violation de la symétrie faciale, une couleur de peau anormale, une anomalie du contour du visage, etc.) que la plupart des détections standard et les techniques de reconnaissance supposent. Bien que ces technologies aient été largement utilisées pour les humains dans les appareils de tous les jours tels que les bio-scanners et les équipements de soins de santé, l'essor fulgurant de l'industrie de l'animation a amplifié le besoin de techniques similaires pour les visages d'animation avec certaines applications notables, notamment :

- Intégrer dans l'image moteurs de recherche pour rechercher sur le Web des dessins animés similaires.

- Intégration avec des lecteurs d'écran pour aider les malvoyants à comprendre les dessins animés.

- Aide le logiciel de contrôle de contenu à surveiller les dessins animés inappropriés sur les réseaux sociaux.

Notre travail vise à atteindre l'objectif mentionné ci-dessus en tirant parti de systèmes d'apprentissage en profondeur capables de reconnaître plus précisément les visages des dessins animés, notre système proposé pour la reconnaissance des visages de dessins animés indique que le modèle de mask r-cnn est utilisé pour recadrer les régions faciales afin d'obtenir des données d'image qui ne contiennent que le visage, le but de cette étape est de supprimer des caractéristiques d'objets supplémentaires (autres objets de l'image, arrière-plans ...), à la fin de cette étape, nous obtiendrons des données contenant uniquement des régions

faciales à utiliser comme données d'entraînement pour entraîner le modèle de classification dans la phase de reconnaissance.

Notre mémoire implique 04 chapitres

-Chapitre01 : le premier chapitre explique le principe de système de recherche d'image par le contenu (Content based image retrieval CBIR), avec une vision générale sur les images numériques qui forment les données de ce travail.

-Chapitre02 : ce chapitre parle de l'apprentissage automatique, Deep Learning et les réseaux de neurones, et parle en détail les CNNs.

-Chapitre03 : ce chapitre présente un état de l'art sur le système de la détection et la reconnaissance faciale qui est une combinaison entre le CBIR et Deep Learning, ensuite il parle des particularités de dessins animés.

-Chapitre04 : ce chapitre présente l'implémentation et montre les différents scenarios de notre système.

Chapitre01

La recherche d’image par le contenu CBIR

5. Introduction

La transition des images du dessin manuel aux images numériques conduit à une science spéciale contient des images et l'évolution et comment la recherche et la récupération en conjonction avec l'avènement de l'ordinateur. La recherche d'image par le contenu (Content based image retrieval CBIR) est l'un des défis fondamentale de la recherche qui largement étudiés dans la communauté multimédia depuis des décennies. À l'heure actuelle, CBIR est devenu l'un des domaines les plus avancés de l'informatique en raison de la disponibilité d'une énorme base de données qui contient une grande quantité d'images disponibles sur Internet, c'est pourquoi, il était très important d'extraire des fonctionnalités d'images pour prendre en charge la recherche d'images.

Dans ce chapitre, nous allons parler tout d'abord de les images numériques qui forment les données utilisés dans ce travail, après on va présenter le système de la recherche d'image par le contenu avec les différents concepts de base, et enfin on va parler à les techniques d'évaluation.

6. Les images numériques

6.1. Définition

Une image numérique est la représentation de toute image acquise à partir des convertisseurs analogiques-numériques située dans des dispositifs tels que (scanner, appareil photo…) en langage informatique (format binaire).

Cette dernière composée par des cellules appelées pixels qui sont visuellement des petits carrés juxtaposés en lignes (Largeur) et en colonnes (Hauteur) construisant une matrice bidimensionnelle. Ces pixels caractérisés par couleur, luminosité, brillance, ce qui permet de différencier les images.

6.2. Les caractéristiques d'image numérique

6.2.1. La résolution

La résolution d'une image correspond au niveau de détail qui va être représenté sur cette image. C’est le nombre de pixels par unité de longueur dans l'image à numériser. Elle est en dpi (dots per inch) ou en ppp (points par pouce). Plus le nombre de pixels est élevé par unité de longueur, plus la résolution est élevée.

6.2.2. La dimension :

Indique la taille de l’image, se présente sous forme de matrice dont les éléments sont des pixels. On peut calculer la dimension en multipliant le nombre de lignes de cette matrice par le nombre de colonnes, le résultat est le total de pixels dans une image qui représente la taille d’image.

6.2.3. La texture

Est la répétition d'un motif qui se traduit par une image visuellement cohérente. Plus précisément, la texture peut être vue comme un ensemble de pixels (niveaux de gris) disposés spatialement selon un ensemble de relations spatiales, aboutissant à une région homogène.

6.2.4. Les contours

Les contours représentent la frontière entre les objets image, ou la limite entre deux pixels où les niveaux de gris représentent une différence significative. La structure de ces objets est décrite par leurs textures. Le but de l'extraction de contour est de trouver les points dans une image qui séparent deux textures différentes.

Figure 1.1 : *Exemple de contour*

6.2.5. Contraste

Il s'agit du contraste entre deux régions d'une image, plus précisément entre les régions sombres et claires de cette image. Le contraste est déterminé par les luminances de deux zones d'image. Si L1 et L2 sont les niveaux de luminosité de deux zones d'image adjacentes A1 et A2, le contraste C est défini par :

$$C = \frac{\mathrm{L1} - \mathrm{L2}}{2\mathrm{L1} + \mathrm{L2}}$$

6.2.6. Le bruit

Un bruit (parasite) dans une image est considéré comme un phénomène de brusque variation de l'intensité d'un pixel par rapport à ses voisins, il provient de l'éclairage des dispositifs optiques et électroniques du capteur.

6.2.7. La luminance

C'est le degré de luminosité des points de l'image. Elle est définie aussi comme étant le quotient de l'intensité lumineuse d'une surface par l'aire apparente de cette surface, le mot luminance est substitué au mot brillance, qui correspond à l'éclat d'un objet.

7. La recherche d'image par le contenu (CBIR)

En 1992, kato *[Kato, 1992]* a introduit le terme la recherche d'image par le contenu en englais : content based image retrieval (CBIR), pour décrire ses expériences sur la récupération automatique d'images à partir d'une base de données par caractéristiques de couleur et de forme…, depuis lors, le CBIR est devenu comme un nouveau domaine de recherche.

7.1. Le principe de système CBIR

Selon [*Zhang et Lu, 2007]*, la recherche d'images par le contenu (CBIR) est le processus de recherche et de récupération d'images à partir d'une base de données sur la base de caractéristiques visuelles qui sont extraites de l'image elles-mêmes. D'une façon plus détaillé, le système fournit à l'utilisateur de chercher une image avec des requêtes exprimant ses désirs, ensuite il analyse ces requêtes et extrait les caractéristiques visuelles. La plupart des systèmes CBIR utilisent des caractéristiques visuelles de bas niveau extraites automatiquement à l'aide de méthodes de traitement d'image pour représenter le contenu brut de l'image .On peut

classer ces contenus en deux classes: Le contenu visuel général comprend les caractéristiques indépendantes de l'application telles que la couleur, la texture, la forme, etc. D'autre part, le contenu visuel spécifique au domaine comprend des caractéristiques dépendant de l'application telles que les visages humains, les empreintes digitales, etc. Après l'extraction des caractéristiques de requêtes le système fait une comparaison de contenu de requêtes avec les données enregistrées dans la base d'indexation pour récupérer les images qui ont des caractéristiques similaires aux requêtes comme résultats de la recherche.

7.2. Architecture générale d'un système d'CBIR

Le système d'CBIR s'exécute en deux phases [Lynda, 2009] :

a. **La phase d'indexation (hors-Ligne) :** Dans cette phase, des caractéristiques sont extraites d'une manière automatiquement à partir de base d'image et les enregistrées dans la base des index sous forme un vecteur numérique appelé descripteur visuel. Grâce aux techniques de la base de données, on peut stocker ces caractéristiques et les récupérer rapidement et efficacement.
b. **La phase recherche (On-line) :** Dans cette étape, le système analyse les requêtes d'utilisateur et les compare avec les descripteurs visuel existant dans la base d'index pour donner le résultat correspond en une liste d'images ordonnées, en fonction de la similarité entre leur descripteur visuel et l'image requête en utilisant une mesure de distance.

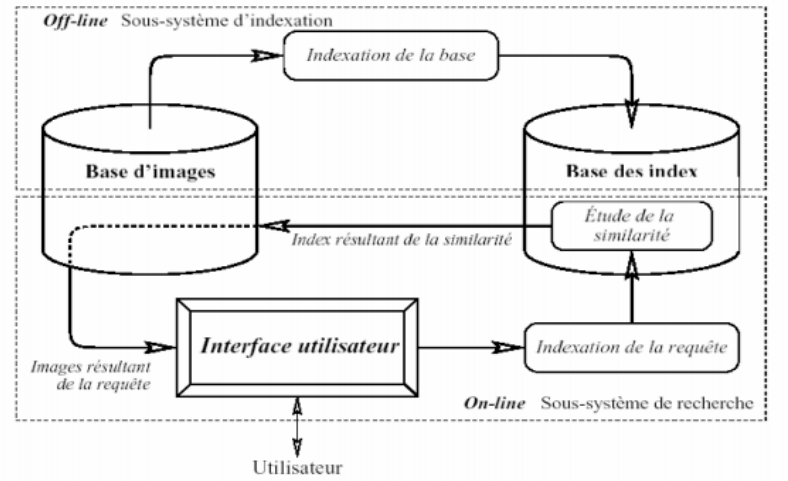


Figure 1.2 :*L'architecture de système CBIR*

7.2.1. La base d'image

La base d'image est les données principales du système CBIR. La différence entre les bases d'images peut identifier par :

- La taille : la plus part des systèmes sont conçues pour des bases de milliard (880 millions d'images) lorsque la base est constituée par les images collectées par des robots sur Internet.
- Le type d'image : le type d'image influe sur la conception globale du système, particulièrement sur les descripteurs de bas niveau calculés. D'une manière générale, plus la variabilité intra et inter images est importante, plus le système doit être riche et précis (et plus le problème d'indexer/rechercher ces images est difficile).
- Stabilité : le taux de changements (ajouts d'images, retraits, etc.) en fonction du temps. Faible pour une collection d'images représentant les œuvres d'un peintre ne créant plus, elle peut être très forte lorsque, par exemple, on s'intéresse aux images de la Toile ou à l'actualité.

7.2.2. L'indexation

L'indexation est les opérations qui construisent d'un index de l'image. Les images doivent être caractérisées par des informations à la fois discriminantes et invariables à certains paramètres. L'indexation peut être :

-fixe: les descripteurs calculés sont toujours les mêmes.

- évolutive: les descripteurs s'adaptent à l'utilisateur ou au contexte dans le temps, ce qui permet de renforcer l'adéquation système/utilisateur. *[Min, 1996]*

- générique (indexation de photographies diverses dans *[Fli, 1995]*), pouvant caractériser des collections hétérogènes, ou spécifique (indexation de peintures chinoises dans) adaptée à un type d'image particulier. Une collection hétérogène est par exemple constituée de photographies personnelles, mettant en scène diverses entités physiques dans des conditions de prise de vue variables. Indexer une telle collection impose l'usage de descripteurs suffisamment génériques (la couleur par exemple), c'est-à-dire qui caractérisent une propriété discriminante applicable à la plupart des entités physiques. A l'inverse, indexer une collection d'images très spécifiques (des empreintes digitales par exemple) requiert l'utilisation de

descripteurs également très spécifiques qui, par ailleurs, ne conviendraient probablement pas à une collection hétérogène.

-Inclut une étape de segmentation : pour caractériser des régions homogènes de l'image ou bien indexer l'image dans sa globalité. La segmentation de l'image précède généralement l'indexation individuelle des régions de l'image et cela permet, outre le fait d'accéder à des parties de l'image, de calculer des descripteurs de « forme ». [Smi, 1996]

7.3. Les composants du CBIR

7.3.1. Les requêtes

La requête est la première étape dans le système CBIR, elle représente le désir d'utilisateur, plusieurs type de requêtes sont proposées lesquels :

a. Requêtes par mot clé : Les images sont recherchées suivant un ou plusieurs critères, par exemple trouver les images contenant des arbres, ici le système se base sur l'annotation manuelle et textuelle d'images.

b. Requêtes par esquisse : dans ce cas, le système met à disposition de l'utilisateur des outils qui lui permettent de construire un croquis (dessin) en fonction de ses besoins. Les croquis fournis seront utilisés comme exemples de recherche. L'esquisse peut être le contour de la forme ou du contour de l'image entière, ou le contour de la couleur ou de la texture de la zone d'image. L'utilisateur choisira ses propres besoins et préférences en fonction de la bibliothèque d'images utilisée, et de l'une ou l'autre de ces représentations. Le principal inconvénient de cette technique est que malgré les outils fournis, il est parfois difficile pour les utilisateurs de fournir des croquis.

c. Requête par exemple : l'utilisateur, pour représenter ses désirs utilise une image (ou une partie d'image) qui est similaire aux images qu'il recherche. L'image exemple peut soit être fournie par l'utilisateur, soit être choisie par ce dernier dans la base d'images utilisée. Cette technique est simple et ne nécessite pas de connaissances approfondies pour manipuler le système.

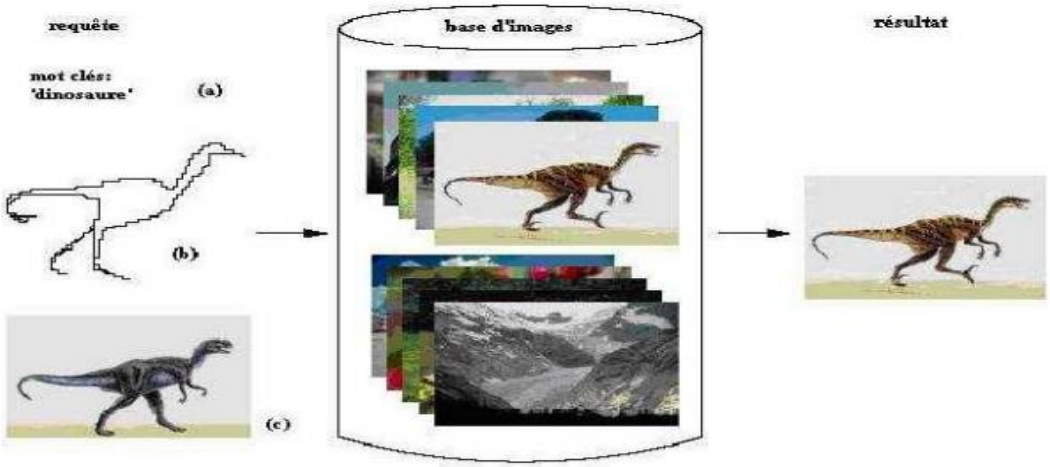


Figure 1.3 : *Les 03 grandes catégories de requete dans CBIR*

7.3.2. La similarité entre la requête et base d'index

Dans la recherche d'images, la calcule de la similarité entre les caractéristiques extraites de la requête et les caractéristiques de chaque image dans la base, cela aboutit généralement à une valeur de correspondance qui caractérise la pertinence d'une image par rapport à la requête. Cette mise en correspondance peut être simple (Histogramme) ou complexe (par exemple [Smi, 1996], avec une mise en correspondance qui tient compte de l'arrangement spatial des régions).

La phase de mise en correspondance peut également inclure une pondération des descripteurs. Pondérer les descripteurs permet d'éliminer une partie du bruit dans la mesure où les descripteurs les moins pertinents voient leur influence diminuer dans l'évaluation de la similarité requête/image.

La mise en correspondance peut également inclure un bouclage de pertinence. L'objectif est d'éliminer le bruit (augmenter la précision) en tentant de converger vers une précision maximale.

7.4. Extraction des caractéristiques

L'analyse se concentre généralement sur les attributs de bas niveau (couleur, texture et forme). L'extraction de ces attributs constitue la première étape de tout programme d'analyse d'images destiné à symboliser son contenu. Il existe deux méthodes principales pour extraire des caractéristiques. La première consiste à construire un descripteur global pour l'image entière. Dans ce cas, le problème est de fournir une observation de l'image entière. L'avantage du descripteur global est la simplicité de l'algorithme de mise en œuvre et la réduction du nombre d'observations obtenues. Cependant, le principal inconvénient de ces descripteurs est la perte des informations de localisation des éléments d'image. La deuxième méthode locale consiste à calculer les attributs d'une petite partie de l'image. L'inconvénient majeur de cette technique est que la quantité d'observations produite est très grande, ce qui implique un gros volume de données à traiter. Le choix des caractéristiques extraites est souvent guidé par la volonté d'invariance ou de robustesse par rapport à des transformations de l'image.

7.4.1. Les descripteurs de la couleur

La couleur est la première caractéristique employée pour la recherche d'images par le contenu grâce à son invariance par rapport à l'échelle, la translation et la rotation *[M. A. Bou, 2009]*. Ces trois valeurs font que son potentiel discriminatoire soit supérieur à la valeur en niveaux de gris des images. Une indexation couleur repose sur deux principaux choix: l'espace colorimétrique et le mode de représentation de la couleur dans cet espace *[G.Quellec, 2008]*.

7.4.1.1.L'espace de Couleur

Il existe plusieurs types d'espaces de couleur, une couleur est identifiée par trois composantes définissant un espace de couleurs. Plusieurs études ont été réalisées sur l'identification d'espaces colorimétriques le plus discriminants mais sans succès car il n'existe pas d'espace de couleurs idéal, c'es pour ça il faut sélectionner l'espace de couleur avant de choisir le type de description de contenu couleur.

a. **L'espace RGB :** est un espace qui implique trois couleurs fondamentaux (rouge, vert, bleu), il est sensible au changement de l'illumination à cause de la corrélation de ses trois composantes, un petit changement d'éclairage modifie les trois composants, et les objets dans la scène seront assombris ou éclairés. Pratiquement, la valeur de chaque canal est un entier avec le canal rouge entre 0 et NR, le canal vert entre 0 et NG, et le canal bleu entre 0 et NB. Par conséquent, chaque couleur appartient à un parallélépipède (cube de Maxwel

Figure 1.4). Le codage le plus couramment utilisé est NR = NG = NB = 255, qui permet à

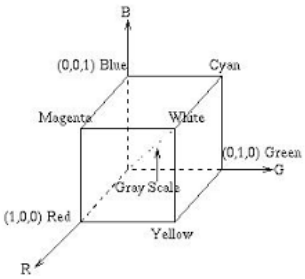


Figure 1.4 : *Espace de couleurs RGB représenté par Cube maxwel*

chaque composant d'être stocké sur un octet.

b. L'espace HVS : (Hue –Saturation –Value) l'idée derrière cet espace est de décrire les couleurs de manière plus intuitive et de se rapprocher de l'expérience humaine. Teinte, Saturation et Intensité sont les trois éléments qui composent cet espace. La sphère de la (figure 1.5) représente un modèle de cet espace. La couleur pure est représentée par une teinte qui est un angle de 0 degrés à 360 degrés dans ce modèle. La valeur de saturation indique la gamme de gris dans l'espace colorimétrique. Elle varie de 0% à 100%. Parfois, la valeur était calculée en utilisant une échelle de 0 à 1, quand la valeur est 0, la couleur est grise et lorsque la valeur est 1, la couleur est une couleur primaire. L'intensité d'une

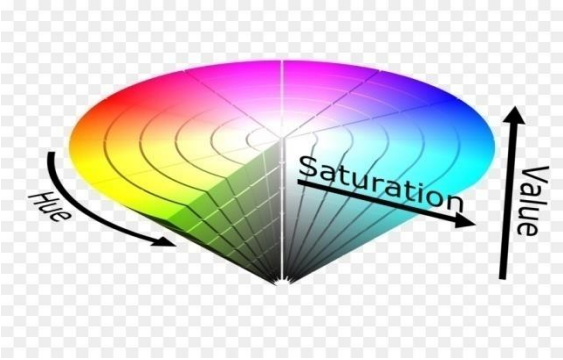


Figure 1.5 : *Représentation de l'espace couleur HVS*

couleur est une mesure de sa luminosité, qui peut varier entre le noir et le blanc.

Les avantages de HVS sont :

- La couleur est représentée par un certain nombre de groupes perceptifs différents.
- L'espace HSI est facilement quantifiable.

- L'information la plus importante pour la couleur est la teinte.

L'un des principaux inconvénients de cet espace est qu'il n'est pas uniforme, la distance calculée entre les couleurs visuellement proches dans cet espace peut être très importants.

c. L'espace CIE Lab : consiste à caractériser une couleur à l'aide d'un paramètre d'intensité L qui correspond à la luminance et de deux autres paramètres de chrominance qui définissent la couleur. Le composant a vous permet de naviguer dans la roue chromatique du rouge au vert, avec une gamme de couleurs allant du vert (-128) au rouge (+127), et le composant b navigue dans la roue chromatique du jaune au bleu, avec une gamme de couleurs du bleu (-128) au rouge (+127). L'espace Lab est perceptuellement uniforme, il a une bonne propriété de respecter les distances entre les couleurs visuellement proches, et il n'est pas affecté par le matériel utilisé. Cependant, lorsqu'il s'agit de calculer les caractéristiques chromatiques avec précision, les variations de couleurs sur les axes a ou b sont cinq fois moins visibles que les variations de luminance dans les applications de traitement d'images.

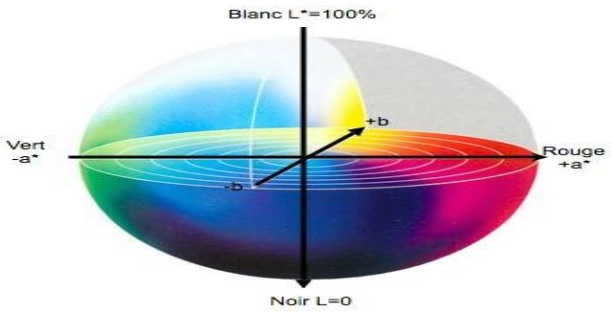


Figure 1.6 : *L'espace de couleur CIELab*

7.4.2. Descripteurs de texture

La texture est le deuxième attribut visuel largement utilisé dans un système CBIR. Elle permet de combler un vide que la couleur est incapable de faire, notamment lorsque les distributions de couleurs sont très proches. De la définition du texture, les recherches sur la modélisation des textures se sont portées sur la caractérisation de ces relations spatiales [L.Gue, 2008]. De nombreuses approches et modèle sont été développées pour la caractérisation de la texture tel que : la transformée en ondelettes de Gabor, sur les caractéristiques de Tamura, Word decomposition, Fractal…Dans cette partie on va parler de quelques approches les plus populaires.

7.4.2.1.Les matrices de co-occurrences

Dans les années 70 Haralick et al [Haralick et al, 1989] ont proposé une des premières méthodes de caractérisation de texture baptisée matrice de cooccurrence. La dépendance spatiale de la texture est explorée en établissant d'abord une matrice de cooccurrence basée sur la direction de l'image et la distance entre les pixels. Dans chacune de ces matrices, Haralick adéfinit 14 paramètres de texture caractéristiques, tels que le contraste, Contraste d'entropie ou de moment. Cette méthode est basée sur des paramètres correctement sélectionnés, notamment: La matrice dans laquelle la mesure est effectuée et la distance d entre les deux motifs de pixels. Ce type de méthode a récemment été appliqué dans la recherche d'image.

7.4.2.2.Filtre de Gabor

Le principe de cette approche est d'analyser s'il existe un contenu de fréquence spécifique dans l'image dans des directions spécifiques dans une région localisée autour du point ou de la région d'analyse. Les représentations de fréquence et d'orientation des filtres de Gabor sont revendiquées par de nombreux scientifiques contemporains de la vision comme étant similaires à celles du système visuel humain. [B.Ols, 1996] Ils se sont avérés être particulièrement appropriés pour la représentation et la discrimination de texture. Dans le domaine spatial, un filtre de Gabor 2D est une fonction noyau gaussienne modulée par une onde plane sinusoïdale.Certains auteurs affirment que de simples cellules du cortex visuel des cerveaux de mammifères peuvent être modélisées par des fonctions de Gabor [S.Mars, 1973]. Ainsi, l'analyse d'image avec les filtres de Gabor est considérée par certains comme similaire à la perception dans le système visuel humain.

7.4.2.3.Transfer en ondelettes

La transformée en ondelettes est à la base de nombreuses analyses de texture, la description de texture à base d'ondelettes est utilisée pour la recherche d'images. L'approche continue des ondelettes pour un signal 2D est trop complexe pour être applicable rapidement sur des images. Pour résoudre ce problème, Mallat [Mallat , 1989] considère l'analyse en ondelettes comme une décomposition du signal par une cascade de filtres, en utilisant une paire de filtres pour chaque niveau de résolution (un filtre passe-haut et un filtre passe-bas). Il propose ainsi la DWT (DiscreteWaveletTransform) qui permet d'obtenir une transformée rapide. Le choix de l'ondelette mère est alors remplacé par le choix du filtre. Pour calculer une transformée en ondelettes, on n'a alors besoin que des deux filtres : au lieu de calculer le produit scalaire de l'ondelette avec le signal, on réalise un produit de convolution du signal avec ces filtres. Une

des transformées en ondelettes les plus couramment employées en analyse d'images est la transformée de Haar, mais d'autres ondelettes sont aussi largement exploitées. Les filtres de Haar sont fréquemment employés en apprentissage pour obtenir la description d'un objet (comme un visage ou une personne).

7.4.3. Descripteurs de formes

Dans les images numériques la forme aide à l'identification des objets du monde réel. [*Zhang et Lu, 2004]* ont présenté une revue complet de l'application des caractéristiques de forme dans le domaine de la récupération et de la représentation d'images. Les principales classifications des caractéristiques de forme sont basées sur les régions et les contours *[P.Tian, 2013]*. La figure 2.5 présente un aperçu de base de la classification des éléments de forme.

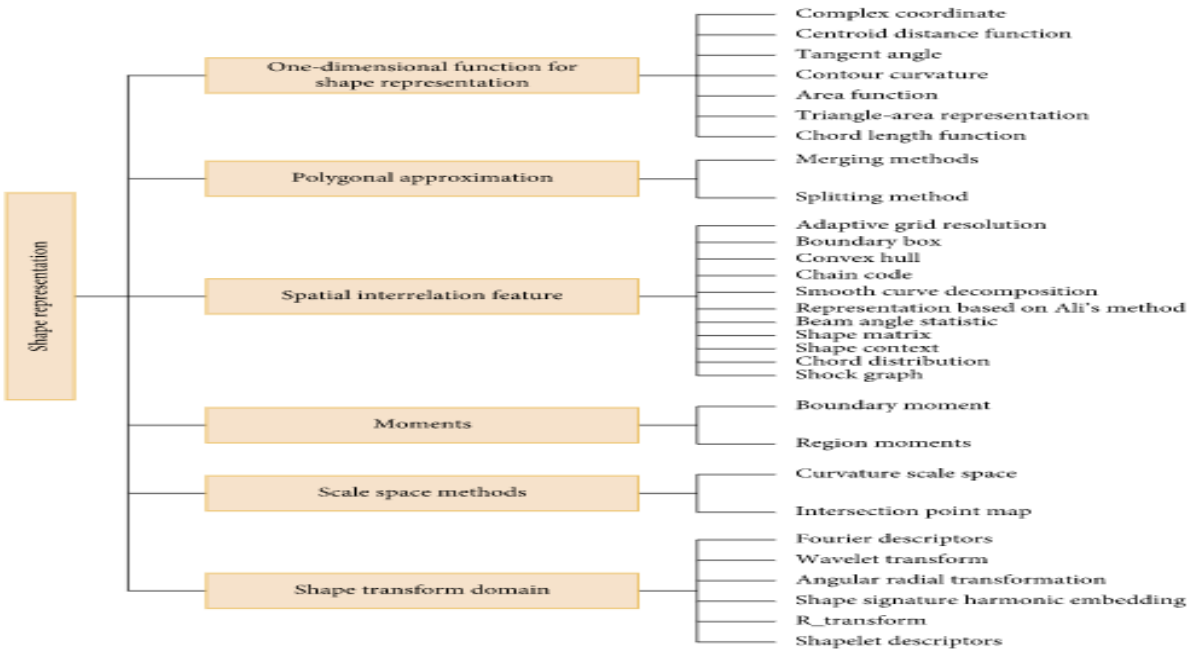


Figure 1.7 : *Aperçu de la classification de forme*

8. Les techniques d'évaluation

Il existe différents techniques d'évaluation des performances pour le CBIR et elles sont traitées dans une norme prédéfinie. Il est important de mentionner ici qu'il n'y a pas de technique standard unique pour évaluer la performance du CBIR. Il existe un ensemble de mesures communes qui sont rapportées dans la littérature. La sélection de toute mesure parmi les techniques mentionnées ci-dessous dépend du domaine d'application, des besoins de l'utilisateur et de la nature de l'algorithme lui-même.

Precision and Recall (Précision et rappel)

La précision (P) et le rappel (R) sont couramment utilisés pour l'évaluation des performances de la recherche CBIR. La précision est le rapport entre le nombre d'images pertinentes dans les k premiers résultats et le nombre total d'images récupérées et s'exprime comme suit: la précision (P) est équivalente au rapport entre les images pertinentes récupérées et le nombre total d'images récupérées (Ntr) :

$$P = \frac{\mathrm{tp}}{\mathrm{Ntr}} = \frac{\mathrm{tp}}{\mathrm{tp} + \mathrm{fp}'}$$

Où tp se réfère aux images pertinentes récupérées et fp 'se réfère au faux positif, c'est-à-dire aux images qui sont mal classées comme images pertinentes.

Recall

Rappel (R) est indiqué comme le rapport entre les images pertinentes récupérées et le nombre d'images pertinentes dans la base de données:

$$R = \frac{\mathrm{tp}}{\mathrm{Nri}} = \frac{\mathrm{tp}}{\mathrm{tp} + \mathrm{fn}'}$$

Où tp fait référence aux images pertinentes récupérées, Nri fait référence au nombre d'images pertinentes dans la base de données. Nri est obtenu sous la forme, tp + fn où fn se réfère au faux négatif, c'est-à-dire aux images qui appartenaient réellement à la classe pertinente, mais mal classées comme appartenant à une autre classe.

***F*-Measure**

C'est la moyenne harmonique de P et R; les valeurs de mesure F plus élevées indiquent une meilleure puissance prédictive:

$$F = 2\frac{\mathrm{P} * \mathrm{R}}{\mathrm{P} + \mathrm{R}'}$$

Où P et R se réfèrent respectivement à la précision et au rappel.

AveragePrecision

L'AveragePrecision (AP) pour une seule requête k est obtenue en prenant la moyenne sur les valeurs de précision à chaque image pertinente:

$$AP = \frac{\sum_{k=1}^{NRI}(\mathrm{P(K)} * \mathrm{R(K)})}{\mathrm{NRI}}$$

MeanAveragePrecision

Pour un ensemble de requêtes S, le MeanAvaregaPrecision (MAP) est la moyenne des valeurs AP pour chaque requête et est donnée par :

$$MAP = \frac{\sum_{q=1}^{S} \mathrm{AP(q)}}{\mathrm{S}}$$

Où S est le nombre de requêtes.

9. Conclusion

Dans ce chapitre nous avons présenté les images numériques et nous avons expliqué le principe des systèmes de recherche d'images par le contenu (Content-Based Image Retrieval systèmes) qui permettent de rechercher les images d'une base de données en fonction de leurs caractéristiques visuelles. Dans ces systèmes, la requête est une image similaire, mot clé, ou esquisse, et le résultat de la requête correspond à une liste d'images ordonnées en fonction de la similarité. Ensuite nous avons mentionné les descripteurs qui sont regroupés en trois catégories (couleur, texture, forme) pour l'extraction des caractéristiques d'images. En fin nous avons parlé des techniques d'évaluation les plus utilisables dans un système CBIR.

Chapitre 02

Apprentissage automatique, Deep Learning

1. Introduction

L'intelligence artificielle(IA) est un domaine scientifique informatique qui implique un ensemble des techniques et technologies développant des programmes capables de simuler l'intelligence humaine pour résoudre des problèmes à forte complexité logique ou algorithmique. L'apprentissage automatique ou bien Machine Learning (ML) est un sous-domaine de l'intelligence artificielle, basé sur des approches mathématiques permettant aux ordinateurs d'améliorer leurs performances à résoudre des processus sans être explicitement programmés. Par conséquent, l'apprentissage profond (en anglais Deep Learning) est un ensemble de méthodes d'apprentissage automatique (ML) tentant de modéliser avec un haut niveau d'abstraction des données grâce à des architectures articulées de différentes transformations non linéaires.

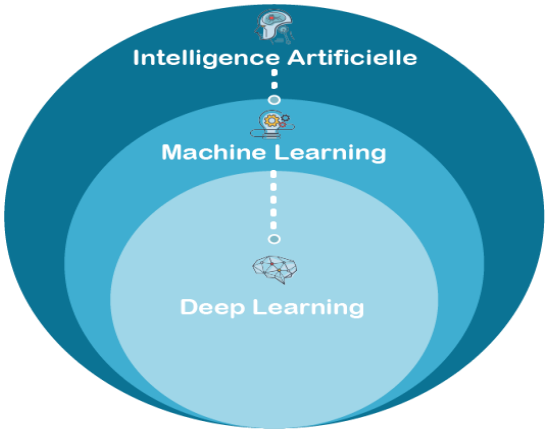


Figure2.1: *La relation entre IA, ML et*

Dans ce chapitre nous allons présenter tout d'abord sur d'apprentissage automatique, ensuite le Deep Learning et les notions de bases qui sont l'associées.

2. Apprentissage automatique (Machine Learning)

2.1. Définition

Machine Learning (ML) est un domaine de l'intelligence artificielle (IA) qui offre aux systèmes la possibilité d'apprendre et de s'améliorer automatiquement à partir de l'expérience sans être explicitement programmés. L'apprentissage automatique se concentre sur le

développement de programmes informatiques qui peuvent accéder aux données et les utiliser pour apprendre par eux-mêmes.

2.2. Types d'apprentissage (Types of Learning)

Il existe différentes manières de former des algorithmes d'apprentissage automatique. Pour comprendre les avantages et les inconvénients de chaque type d'apprentissage automatique, nous devons d'abord examiner le type de données qu'ils ingèrent. En ML, il existe deux types de données: les données étiquetées et les données non étiquetées.

Les données étiquetées ont à la fois les paramètres d'entrée et de sortie dans un modèle entièrement lisible par machine, mais nécessitent beaucoup de travail humain pour étiqueter les données, pour commencer. Les données non étiquetées n'ont qu'un seul ou aucun des paramètres sous une forme lisible par machine. Cela élimine le besoin de main-d'œuvre humaine mais nécessite des solutions plus complexes.

Dans ML on distingue trois types principaux d'apprentissage (figure1.2) :

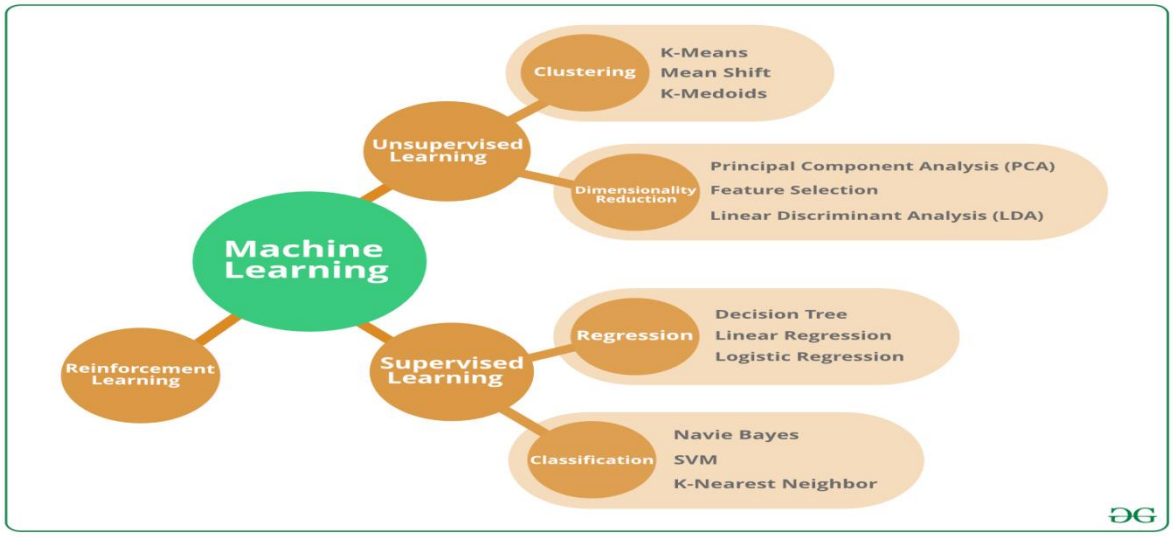
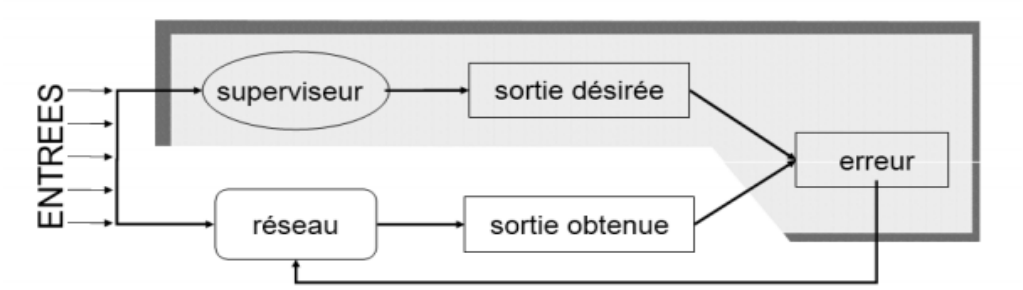


***Figure 2.2**: Méthodes et algorithmes d'apprentissage en ML*

2.2.1. Apprentissage supervisé

Une tâche consiste à apprendre une fonction qui mappe une entrée à une sortie basée sur des exemples de paires d'entrée-sortie *[J.Russ, 2010]* .Il déduit une fonction à partir de données d'apprentissage étiquetées constituées d'un ensemble d'exemples d'entraînement *[Mehri et al, 2012]*.Dans l'apprentissage supervisé, chaque exemple est une paire constituée d'un objet

d'entrée (généralement un vecteur) et d'une valeur de sortie souhaitée (également appelée signal de supervision). Un algorithme d'apprentissage supervisé analyse les données d'apprentissage et produit une fonction déduite, qui peut être utilisée pour cartographier de nouveaux exemples. Un scénario optimal permettra à l'algorithme de déterminer correctement



les étiquettes de classe pour les instances invisibles. Cela nécessite que l'algorithme d'apprentissage généralise les données d'apprentissage à des situations invisibles d'une manière raisonnable.

2.2.1.1. Classification :

Les algorithmes de classification sont utilisés lorsque la variable de sortie est catégorique, ce qui signifie qu'il existe deux classes telles que Oui-Non, Homme-Femme, Vrai-faux, etc. Voici quelques algorithmes de classification populaires qui font l'objet d'un apprentissage supervisé:

- Forêt aléatoire

***Figure 2.3** : Schéma d'un modèle supervisé.* [A. Metref, 2010]

- Arbres de décision
- Régression logistique
- Machines vectorielles de soutien

2.2.1.2. Régression :

Les algorithmes de régression sont utilisés s'il existe une relation entre la variable d'entrée et la variable de sortie. Il est utilisé pour la prédiction de variables continues, telles que les prévisions météorologiques, les tendances du marché, etc. Voici quelques algorithmes de régression populaires qui font l'objet d'un apprentissage supervisé:

- Régression linéaire

- Arbres de régression
- Régression non linéaire
- Régression linéaire bayésienne
- Régression polynomiale

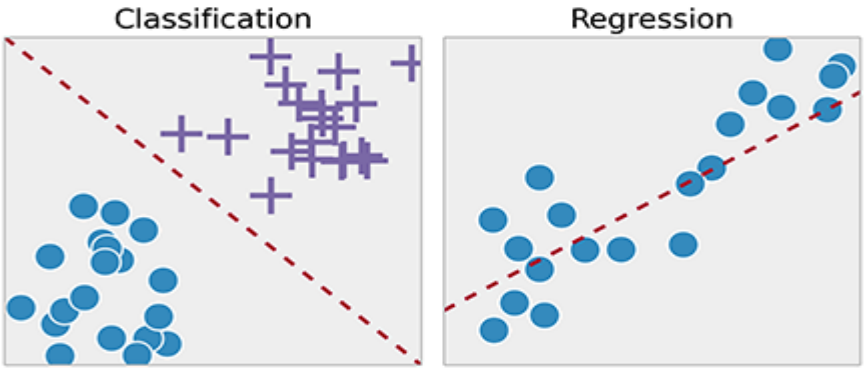


Figure 2.4: *Exemple de Classification et Régression supervisé*

2.2.2. Apprentissage non-supervisé

Une tache d'apprentissage consiste à apprendre des modèles à partir de données non étiquetées. L'espoir est que, par mimétisme, la machine est obligée de construire une représentation interne compacte de son monde puis de générer un contenu imaginatif. UL présente une auto-organisation qui capture des modèles sous forme de prédilections neuronales ou de densités de probabilité [H.Geo, 1999].Les deux grandes méthodes de l'UL sont les réseaux de neurones et les méthodes probabilistes.

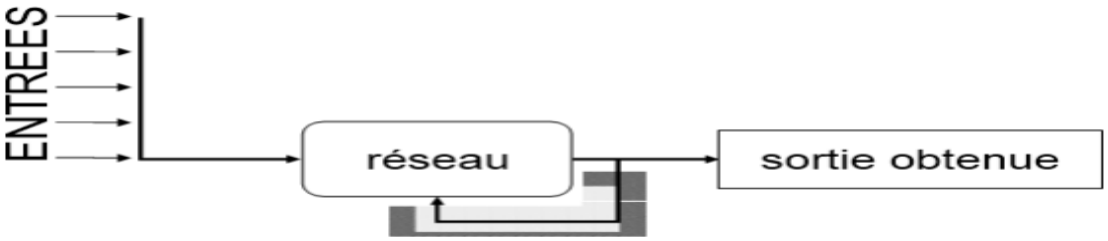


Figure 2.5:*Schéma d'un modèle non supervisé [A. Metref ,2010]*

2.2.2.1. Clustering

Clustering est une méthode de regroupement des objets en clusters de sorte que les objets présentant le plus de similitudes restent dans un groupe et présentent moins ou pas de similitudes avec les objets d'un autre groupe. L'analyse de cluster trouve les points communs entre les objets de données et les

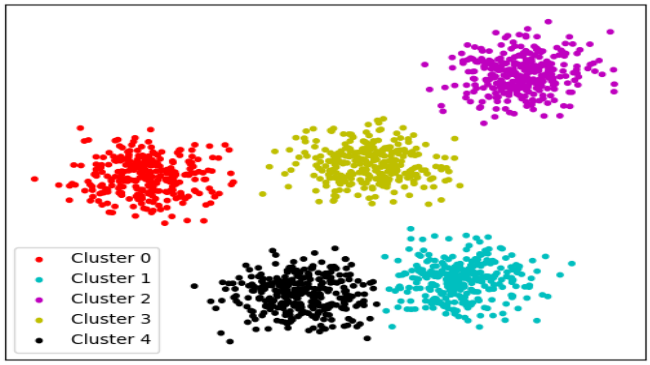


Figure 2.6:*Exemple de Clustering*

catégorise en fonction de la présence et de l'absence de ces points communs**.**

2.2.2.2. Réduction de la dimension

Le principe de cette technique consiste à transformer les données d'un espace de grande dimension en un espace de petite dimension afin que la représentation en basse dimension conserve certaines propriétés significatives des données d'origine, idéalement proches de sa dimension intrinsèque. Travailler dans des espaces de grande dimension peut être indésirable pour de nombreuses raisons; les données brutes sont souvent rares en raison de la malédiction de la dimensionnalité, et l'analyse des données est généralement insoluble d'un point de vue informatique. La réduction de la dimension peut être utilisé pour la réduction du bruit, la visualisation de données, l'analyse de grappes ou comme étape intermédiaire pour faciliter d'autres analyses.

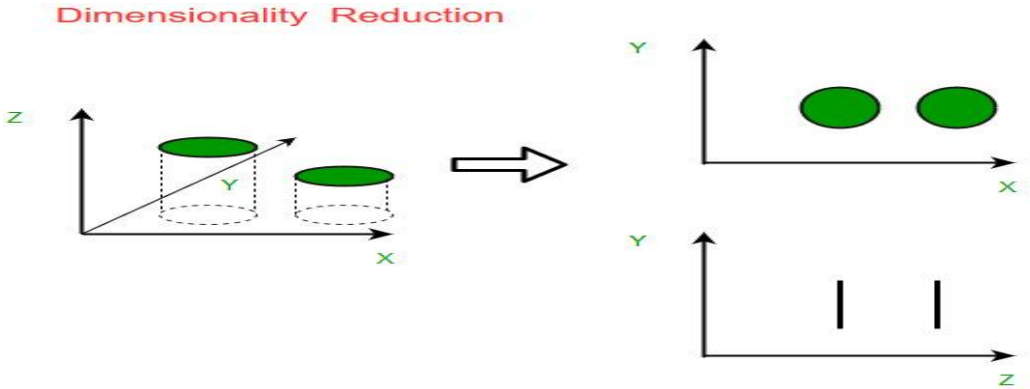


Figure 2.7 : La réduction dimensionnelle

2.2.3. Apprentissage par Renforcement

L'apprentissage par renforcement (RL) s'intéresse à la manière dont les agents intelligents doivent agir dans un environnement afin de maximiser la notion de récompense cumulative. L'apprentissage par renforcement ne nécessitant pas la présentation de paires (entrées / sorties) étiquetées et ne nécessitant pas d'actions sous-optimales pour être explicitement corrigées. Au lieu de cela, l'accent est mis sur la recherche d'un équilibre entre l'exploration (d'un territoire inexploré) et l'exploitation (des connaissances actuelles). *[Mehri et al, 2012]*

2.3. Les étapes de l'apprentissage automatique

L'apprentissage automatique n'est actuellement pas un ensemble d'algorithmes mais une succession d'étapes *[W, 2]* :

a. **Collection de données :** est une étape importante car la quantité et la qualité de données entrées déterminent la précision de modèle. Généralement cette étape est la représentation de données **d'entraînement.**
b. **Préparation des données :** les données doivent être traitées avant l'utilisation, et nettoyer ce qui peut exiger (supprimer les redondances, corriger les erreurs, traiter les valeurs manquantes, normalisation, conversions de types de données, etc.), après il faut randomiser les données pour effacer les effets de l'ordre particulier dans lequel nous avons collecté ou autrement préparé nos données. Cette étape est un bon moment pour visualiser des données pour aider à détecter les relations pertinentes entre les variables ou les déséquilibres de classe (alerte de biais!), Ou effectuez d'autres analyses exploratoires.
c. **Sélection ou construction un modèle** : il existe plusieurs choix d'algorithmes parmi les nombreux que les chercheurs et les data scientistes ont créés, cet algorithme doit être adapté au problème et aux données.
d. **Entrainement de modèle :** Le but de cette étape est de répondre à une question ou de faire une prédiction correctement aussi souvent que possible, chaque itération du ce processus est une étape d'entrainement.
e. **Evaluation de modèle :** consiste a :
   - Utiliser une métrique pour mesurer les performances objectives du modèle.

- Tester le modèle par rapport à des données qui n'ont jamais été auparavant vues et utilisées pour l'entrainement et sont supposées être représentatives de la façon dont le modèle pourrait fonctionner dans le monde réel.

f. Réglage des paramètres : Ajuster les paramètres du modèle pour des performances améliorées, les hyper-paramètres du modèle simple peuvent inclure: le nombre d'étapes d'entraînement, le taux d'apprentissage, les valeurs d'initialisation et la distribution, etc.

g. Prédiction : répondre aux questions en utilisant des prédictions. Il peut s'agir de toutes sortes de prédictions, allant de la reconnaissance d'image à la sémantique en passant par l'analyse prédictive.

3. Apprentissage profond (Deep Learning)

3.1. Définition

L'apprentissage profond ou Deep Learning est une sous-partie de l'intelligence artificielle, il implique un ensemble de techniques analysant des données de différents types dans le but de tirer un ensemble de règles qui permettront de tirer des conclusions sur de nouvelles données. L'apprentissage profond se base sur ce qu'on appelle les réseaux de neurones artificiels disposant de nombreuses couches cachées (c'est-à-dire les résultats d'une première couche servent d'entrée aux calculs d'une deuxième couche et ainsi de suite).

3.2. La différence entre Deep Learning et Machine Learning

La différence majeure entre Machine Learning et Deep Learning est l'extraction de caractéristiques, dans les algorithmes de Machine Learning traditionnelles l'extraction de caractéristiques est exécutée d'une manière manuelle, c'est une étape difficile et prend du temps et nécessite un spécialiste en la matière, au contraire en Deep Learning cette étape est faite d'une manière automatiquement par l'algorithme.

Une autre différence entre le Deep Learning et les algorithmes de ML qui se concentrent sur la quantité de données, est qu'il offre une bonne adaptabilité, plus il y a de données fournies, meilleures sont les performances des algorithmes de Deep Learning. Contrairement à de nombreux algorithmes ML classiques, qui limitent parfois la quantité de données reçues et parfois appelés «plateau de performance». Les modèles de Deep Learning n'ont pas de limites

théoriques, voire dépassent les performances des humains dans certains domaines. Par exemple, le traitement d'image.

3.3. Les réseaux de neurones artificiels (ANN)

Le réseau de neurones distribue la valeur de la variable à Automates (neurones). Ces unités sont chargées de combiner leurs informations pour déterminer la valeur du paramètre discriminant. C'est grâce à la connexion de ces unités que RN a la capacité de distinguer. Chaque neurone reçoit des informations numériques des neurones voisins. Chacune de ces valeurs est associée à un poids représentant la force de la connexion. Chaque neurone effectue des calculs localement, puis transmet ses résultats aux neurones en aval. *[A. Schm, 2016]*

Dans la littérature, nous avons parlé de perceptrons simples et de perceptrons multicouches.

- **Perception simple**

Il est lié à un neurone binaire, ce qui signifie que sa sortie est égale à 0 ou 1. Pour calculer cette sortie, le neurone effectue une somme pondérée de ses entrées (chaque entrée a un poids):$Y = f\ (W1 \times X1 + W2 \times X2)$, puis appliquer la fonction d'activation de seuil. Si le poid dépasse une certaine valeur, la sortie du neurone est 1, sinon elle vaut 0. Alors le perceptron simple ne fait que la classification, puis de la prédiction. [N.P.Rou, 2018]

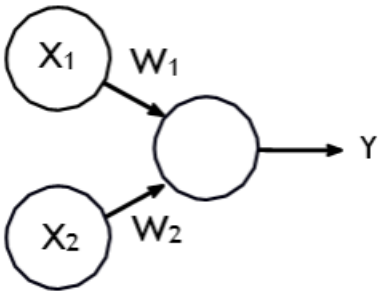


Figure 2.8: Schéma d'un perceptron simple

- **Perceptron Multicouche (PMC)**

La famille de réseaux de neurones la plus largement utilisée est le perceptron multicouche, ce type de réseau couvre à lui seul plus de 95% des applications scientifiques et industrielles.

Dans des circonstances normales, il contient des dizaines à des centaines de neurones, et pour les applications graphiques, il contient même des milliers de neurones.

PMC est un modèle de réseau de propagation par couche, les neurones sont organisés en couches successives: la couche d'entrée, la couche de sortie et entre deux ou plusieurs couches intermédiaires (également appelées couches cachées). Il n'y a pas de connexion entre les neurones de la même couche, mais chaque neurone d'une couche est connecté à tous les neurones de la couche suivante.

La « couche » d'entrée n'est pas une réelle couche de neurones car elle se contente de coder les variables d'observation. La couche de sortie code les variables d'identification. La valeur d'activité du neurone se propage à travers le réseau de l'entrée à la sortie sans retour en arrière. L'existence de la couche cachée permet de modéliser la relation non linéaire entre l'entrée et la sortie.

Théoriquement, une seule couche cachée est suffisante, mais avoir une deuxième couche cachée facilite la modélisation de fonctions discriminantes non continues. En fait, la plupart des problèmes peuvent être résolus avec un à deux niveaux (jusqu'à trois). *[A. Sc, 2016]*.

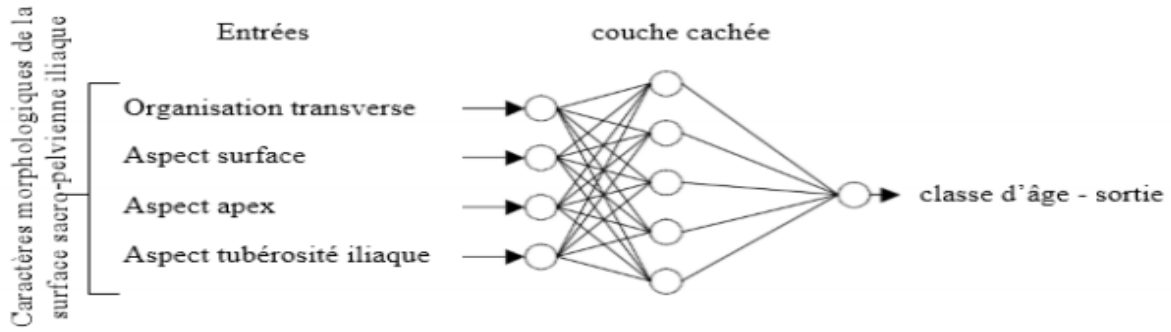


Figure 2.9: Schéma d'un perceptron multicouche

3.3.1. Les réseaux de neurones conventionnels CNNs

Les réseaux de neurones convolutifs sont un type spécial de réseaux de neurones multicouches. Comme presque tous les autres réseaux de neurones, ils sont formés avec une version de l'algorithme de rétro-propagation, là où ils diffèrent, c'est dans l'architecture.

Les réseaux de neurones convolutifs sont conçus pour reconnaître les modèles visuels directement à partir d'images de pixels avec un prétraitement minimal, ils peuvent reconnaître des motifs avec une extrême variabilité (tels que des caractères manuscrits) et avec une robustesse aux distorsions et aux transformations géométriques simples. [W, 3]

L'architecture du CNN dispose la partie convolutive et comporte par conséquent deux parties bien distinctes :

- *Une partie convolutive :* Consiste à extraire des caractéristiques propres à chaque image en les compressant pour réduire leur taille initiale. En bref, l'image fournie en entrée passera à travers une série de filtres tout en créant une nouvelle image appelée carte de convolution. Enfin, la concaténation de carte de convolution obtenue dans un vecteur de caractéristiques appelé code CNN.
- *Une partie classification :* Le code CNN obtenu en sortie de la partie convolution est fourni en entrée dans la seconde partie, qui se compose de couches entièrement connectées appelées perceptrons multicouches. La fonction de cette partie est de combiner les fonctionnalités du code CNN pour classer l'image. [W,4]

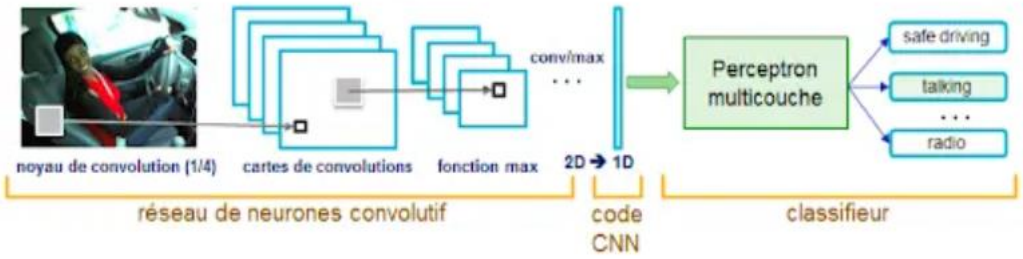


Figure 2.10 : *Fonctionnement d'un CNN*

3.3.2. L'architecture de CNNs

3.3.2.1. La couche de convolution (CONV)

Le rôle de cette couche est d'analyser les images fournies en entrée et de détecter la présence d'un ensemble de caractéristiques. On obtient en sortie de cette couche un ensemble de « featuresmaps ». [W, 1]

Dans cette couche il existe trois hyper-paramètres qui permettent de dimensionner son volume : la profondeur, le pas et la marge *[A. kri et I. Su, 2012]*.

- Profondeur : nombre de noyaux de convolution (ou nombre de neurones associés à un même champ récepteur).
- Le pas : contrôle le chevauchement des champs récepteurs. Plus le pas est petit, plus les champs récepteurs se chevauchent et plus le volume de sortie sera grand.
- La marge à 0 ou zero padding: parfois, il est commode de mettre des zéros à la frontière du volume d'entrée. La taille de ce 'zero-padding' est le troisième hyper-paramètre. Cette marge permet de contrôler la dimension spatiale du volume de sortie. En particulier, il est parfois souhaitable de conserver la même surface que celle du volume d'entrée.

3.3.2.2. Couche de pooling (POOL)

La couche de Pooling est une opération généralement appliquée entre deux couches convolutives. Celui-ci reçoit la carte de caractéristiques formée à la sortie de la couche convolutive en entrée, et son rôle est de réduire la taille de l'image tout en conservant ses caractéristiques les plus élémentaires. Parmi les méthodes les plus utilisées, nous avons retrouvé le pooling maximum ou pooling moyen susmentionné, dont le fonctionnement est de maintenir la valeur moyenne de la fenêtre de filtrage à chaque étape. Enfin, dans la sortie de cette couche de regroupement, nous obtenons le même nombre de cartes d'entités que l'entrée, mais avec une compression considérable.[W,1]

3.3.2.3. Couche de correction (ReLu)

En général, l'efficacité du traitement peut être améliorée en insérant une couche qui exécutera une fonction mathématique (fonction d'activation) sur le signal de sortie entre les couches de traitement. Nous avons notamment:

- Correction ReLU (abréviation d'unité de rectification linéaire): f (x) = max (0, x). Cette fonction est également appelée "fonction d'activation non saturée", ce qui peut augmenter la non-linéarité de la fonction de décision et de l'ensemble du réseau sans affecter le champ de réception de la couche convolutive.

- Corriger par tangente hyperbolique f (x) = tanh (x).

- Corrigé par tangente hyperbolique saturée: f (x) = | tanh (x) |.

- Correction par fonction sigmoïde. En général, la correction Relu est préférable car elle peut rendre l'entraînement du réseau neuronal plusieurs fois plus rapide sans impact significatif sur la précision de la généralisation.

3.3.2.4. Couche entièrement connectée (FC)

Ces couches sont situées à la fin de l'architecture CNN et sont entièrement connectées à tous les neurones de sortie (d'où le terme «entièrement connecté»). Après avoir reçu le vecteur d'entrée, la couche FC appliquera la combinaison linéaire à son tour, puis appliquera la fonction d'activation. Le but ultime est de classer l'image d'entrée. Enfin, il renvoie un vecteur de taille d, qui correspond au nombre de catégories, où chaque composante représente la probabilité que l'image d'entrée appartienne à une catégorie. [W, 1]

3.3.2.5. Couche de perte (LOSS)

La couche de perte spécifie comment le glissement du réseau pénalise l'écart entre les signaux attendu et vrai. Il s'agit généralement de la dernière couche du réseau. Il y a de nombreuses fonctions qui peuvent utiliser des pertes appropriées à différentes tâches ici, tel que : la perte "Soft max" pour prédire une seule classe parmi K classes mutuellement exclusives, la perte par entropie croisée sigmoïde est utilisé pour prédire K valeurs de probabilité indépendantes dans [0,1], et la perte euclidienne est utilisée pour revenir à la vraie valeur.

3.3.3. Choix des hyper-paramétres

CNN utilise plus d'hyper-paramètres que le ML standard, même si les règles habituellement adaptées au taux d'apprentissage et à la constante de régularisation. Généralement, le nombre de filtres, leur forme et le forme Max pooling doivent être pris en compte.

3.4.3.1. Nombre de filtres

Au fur et à mesure que la taille de l'image intermédiaire diminue avec l'augmentation de la profondeur de traitement, les couches proches de l'entrée ont tendance à avoir moins de filtres, tandis que les couches proches de la sortie peuvent avoir plus de filtres. Afin de rendre le calcul de chaque couche égal, le produit du nombre d'entités et du nombre de pixels traités est généralement sélectionné pour être approximativement constant parmi les couches. Afin de

conserver les informations d'entrée, il est nécessaire de conserver le nombre de sorties intermédiaires (le nombre d'images intermédiaires multiplié par le nombre de positions de pixels) d'une couche à l'autre (au sens large).

Le nombre d'images intermédiaires contrôle directement la fonction du système et dépend du nombre d'exemples disponibles et de la complexité du traitement.

3.4.3.2. Forme de filtres

La forme du filtre varie considérablement dans la littérature. Ils sont généralement sélectionnés en fonction de l'ensemble de données. Les meilleurs résultats sur les images MNIST (28x28) se situent généralement dans la plage 5x5 de la première couche, tandis que les ensembles de données d'images naturelles (généralement des centaines de pixels dans chaque dimension) ont tendance à utiliser une couche de filtre de 12x12 ou même 15x15. Par conséquent, le défi est de trouver le bon niveau de granularité afin de créer une abstraction adaptée à chaque situation à l'échelle appropriée.

3.4.3.3. Forme de Max pooling

Max Pooling est un processus de convolution où le noyau extrait la valeur maximale de la zone qu'il convolutionne. Max Pooling dit simplement au réseau de neurones convolutifs que nous ne transférerons que ces informations, si c'est la plus grande information disponible en termes d'amplitude.

Max-pooling sur un canal 4 * 4 (figure2.11) utilisant un noyau 2 * 2 et une foulée de 2: Comme nous sommes en convolution avec un noyau 2 * 2. Si nous observons le premier ensemble 2 * 2 sur lequel le noyau se concentre, le canal a quatre valeurs 8, 9, 7,6. Max-Pooling choisit la valeur maximale de cet ensemble qui est «9».

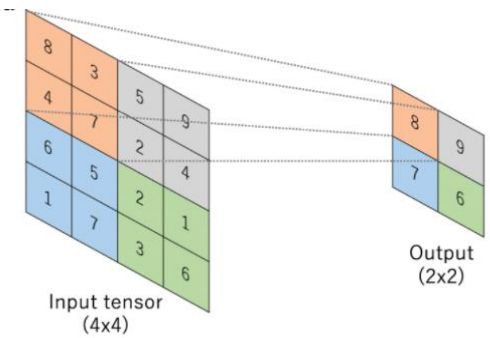


Figure 2.11 : Max Pooling dans les CNNs

4. Conclusion

Dans ce chapitre, nous avons présenté une vision générale sur le domaine de l'apprentissage automatique (définition, type d'apprentissage …). Aussi on a parlé de l'apprentissage profond (Deep Learning) qui se base sur les réseaux de neurones, par conséquence on a présenté en particulière les réseaux de neurones convolutionnels qui sont capables d'extraire et classifier des caractéristiques d'image présentés en entrées.

Chapitre03

Etat de l'art

Chapitre03 : Etat de l'art

1. Introduction

Le système automatique de la reconnaissance facial est une application dans le domaine de la vision par ordinateur « computer vision » qui sert à reproduire la vision humaine, ce système consiste à reconnaitre une personne à partir d'une image de son visage de manière automatique. Avec le développement de la technologie multimédia, de plus en plus de visages de dessins animés apparaissent un large éventail d'aspects de la vie tels que fournir des écoles à domicile pour les enfants, améliorer le processus d'enseignement et les résultats scolaires, ainsi que la représentation de son opinion sur les pratiques de la société (à travers les caricatures politiques et le journalisme de bande dessinée).

Par conséquent, la reconnaissance faciale des dessins animés est également devenue particulièrement importante. Bien que les approches récentes de reconnaissance faciale puissent traiter la plupart des visages humains, elles ne peuvent pas connaitre les visages de dessins animés avec précision avec de nombreux résultats manquants et faux positif, aussi a cause des traits exagérés d'une manière qui conduit souvent à la déviation de ces visages par rapport aux attributs humains implicites (par exemple violation de la symétrie faciale, teint artificiel, contour du visage anormal, etc.) présumé par la plupart des techniques de détection de référence, et de reconnaissance. En effet, de nombreux scénarios dans les dessins animés sont plus difficiles que le monde réel, par exemple, différents visages peuvent avoir des caractéristiques complètement différentes, de nombreux visages sont très similaires aux échantillons négatifs (les corps, l'arrière-plan, etc.).

Dans ce chapitre, nous allons tout d'abord présenter un état de l'art sur les techniques de détection et de reconnaissance de visage, ensuite nous allons voir les particularités de dessins animés dans la reconnaissance de visage.

2. Le fonctionnement d'un système de reconnaissance de visage

Un système automatique fonctionne par le même principe de CBIR, qui passe par deux modes:

- **Un mode d'enrôlement (hors-ligne):** C'est l'étape d'apprentissage de tout système automatisé qui consiste à extraire les caractéristiques d'une personne de

son image à enregistrer dans la base de données, puis à la convertir en un vecteur distinct. Cette inscription sera associée à une étiquette d'identification.

- **Un mode d'identification (enligne):** Il permet d'identifier l'individu à partir de sa photo en faisant une comparaison avec les modèles de toutes les personnes de la base de données. c'est à dire de retrouver l'identité associée à l'image.

Le système de reconnaissance faciale fonctionne principalement en trois étapes après l'acquisition d'image (figure3.1) :

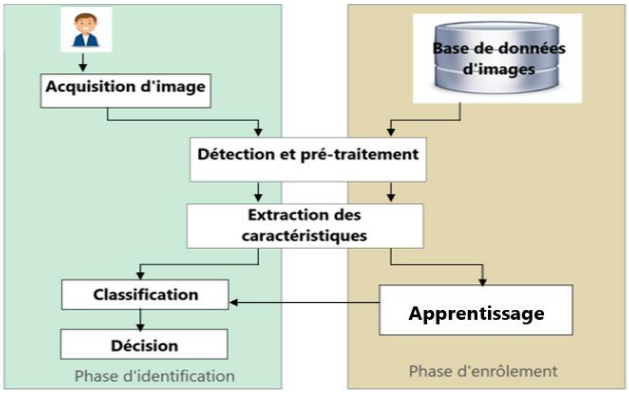


Figure 3.1 : Représentation d'un système de reconnaissance faciale

a. La détection des visages et le prétraitement

- **Détection :** L'étape de détection de visage est importante dans les systèmes de reconnaissance faciale, elle commence par capturer la ressource qui contient le visage puis extrait le visage de l'image capturée par l'une des méthodes de détection afin de localiser d'identifier la zone contenant les traits du visage (yeux, nez , bouche...).**Prétraitement** : le rôle de cette étape est d'éliminer les parasites causés lors de l'acquisition de l'image en entrée, dans le but de ne conserver que les informations essentielles et donc préparer l'image à l'étape suivante.

b. Extraction de caractéristiques

Cette étape est le cœur du système de reconnaissance, consiste à extraire de l'image les caractéristiques qui seront stockées dans le mémoire pour prendre une décision plus tard.

Le choix de ces informations utiles revient à établir un modèle pour le visage, elles doivent être discriminantes et non redondantes.

c. La reconnaissance de visage

Elle consiste à modéliser les paramètres extraits d'un ensemble de visages d'un individu en se basant sur leurs caractéristiques communes. Un modèle est un ensemble d'informations utiles, discriminantes et non redondantes qui caractérise un ou plusieurs individus ayant des similarités, ces derniers seront regroupés dans la même classe. Selon les caractéristiques extraites précédemment, les algorithmes de comparaison diffèrent. On trouve dans la littérature plusieurs approches dont la plus simple est le calcul de distance (recherche de similarité). L'apprentissage consiste donc à mémoriser les représentations calculées dans la phase d'analyse pour les individus connus.

3. Les approches de détection et de la reconnaissance faciale (état de l'art)

3.1. Les approches de détection

Selon *[Yang et al 2002]* les approches de détection de visage humain sont devisées en quatre catégories :

-Approche basée sur les connaissances acquises.

- Template-matching.

- Approche basée sur l'apparence.

- Approche basée sur des caractéristiques invariantes.

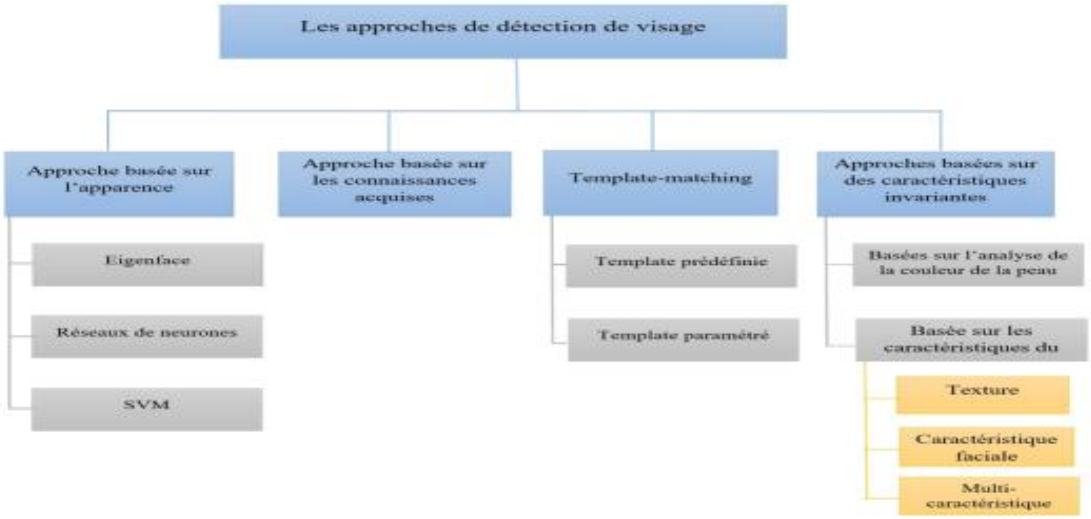


Figure 3.2 : Les approches de détection de visage

3.2. Approche basée sur les connaissances acquise

Cette approche s'intéresse aux différents composants du visage (la bouche, le nez et les yeux ...etc.), ensuite ces dernières sont mesurées classifiés en classe visage ou non visage. L'objectif principal de cette approche consiste à la localisation du visage.

Pour localiser les caractéristiques du visage, *[Kotro et Pitas, 1997]*ont utilisé une technique basée sur des règles bien définies à l'aide d'une méthode de projection proposée par *[kanade,1973]*pour détecter les contours de visage, le problème de cette méthode est que ne détecte pas le visage quand il existe sur un arrière plan complexe.

Quand la résolution d'une image d'un visage est réduite et faible, les traits macroscopiques duvisage disparaissent et le visage devient uniforme. A partir de cette observation [*Yang et Huang ,1994]* ont proposé une méthode pour la détection de visage qui se base sur la fonction de la résolution. Le processus commence par une image à faible résolution et un ensemble de règles, en déduit un ensemble de candidats de visage après l'application de l'ensemble des règles sur l'image à faible résolution, les candidats de visage permettent de vérifier l'existence des trais de visage grâce au calcul des minimas locaux. Malheureusement le nombre des fausses détections de cette méthode est grand.

3.2.1. Approches basées sur le « Template-matching »

Elle permet de détecter un visage ou une partie de visage à travers un apprentissage d'un exemple standard de visage. L'idée de cette approche est de calculer la corrélation entre les visages candidats (chaque partie de l'image d'entrée) et le Template. Les Template peuvent être définis manuellement ou paramétrés à l'aide des fonctions. Bien que touts les visages aient la même structure mais la distance, la position et taille de visage posent des problèmes robustesses pour cette approche qui sont liés aux variations de lumière et de l'échelle.

La procédure de détection du visage pour cette méthode se fait en deux étapes :

- Première étape : La détection des régions candidats du visage.
- Deuxième étape : s'occupe de les examinassions des détails pour la détermination des caractéristiques nécessaires du visage.

Selon [Sinha, 1994] l'utilisation d'un ensemble d'invariants afin de décrire un modèle de visage pour déterminer les invariants aux changements de luminosité permet de caractériser les différents composent du visage (les yeux, le front ... etc). L'algorithme utilisé, calcule le rapport de luminance entre les différentes régions du visage et il retient la direction de ces rapports.

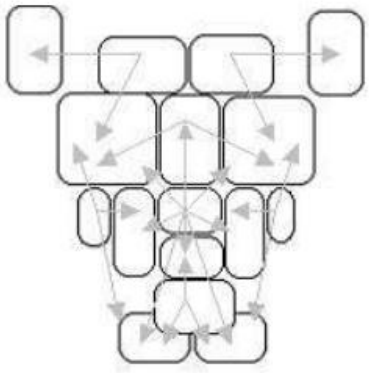


Figure 3.3:*Modèle d'un visage compose en 16 régions et 23 directions [Sinha, 1994]*

Ce modèle prédéfinie est décomposé en 16 régions et 23 relations, qui est divisé en deux groupes, un groupe se compose de 11 relation est qui représente les relations essentielles et l'autre groupe se compose de 12 relation qui représente les relations de confirmation. Chaque flèche dans le modèle représente une relation entre deux régions, une relation entre deux régions est vérifiée si et seulement si le degré de correspondance dépasse un seuil défini. Le visage ne peut pas être déterminé ni localisé sauf si et seulement si le nombre de relations essentielles et de confirmation lui aussi dépasse un seuil défini.

Contrairement à la technique de Sinha, [Yuille et al 1992] ont utilisé un Template déformable pour modéliser les caractéristiques de visage, ils ont créé un modèle de Template élastiqué adaptatif aux caractéristiques du visage comme les yeux, la bouche...etc. Le Template paramétré de cette technique permet de décrire les caractéristiques du visage. La définition de la fonction d'énergie est pour lier les contours, les sommets et les vallées dans l'image d'entrée aux paramètres correspondants dans le template.

D'après ces deux expériences on peut déduire deux techniques de détection de visage appartenant à l'approche de détection basée sur le Template matching qui sont :

- Template prédéfinie

- Template déformable ou élastique

3.2.2. Approches basées sur l'apparence

Les méthodes de cette approche se basent généralement sur les techniques d'apprentissage automatique, l'apprentissage des modèles qui sont utilisés plus tard pour la détection du visage se fait à l'aide ou plutôt par l'utilisation d'un ensemble d'images qui représente la variation de l'espace facial.

Le problème de la détection du visage pour cette approche est considéré comme un problème de classification entre deux classes : Classe visage et Classe non-visage.
Les méthodes de l'approche basée sur l'apparence se basent sur des techniques d'analyse statistique et l'apprentissage automatique pour trouver les caractéristiques appropriées des images de visage et des images de non-visage. Plusieurs techniques ont été utilisé pour cette approche telles que : (Eigenface, Réseaux de neurone et Support vecteur machine « SVM »).

3.2.2.1. Eigenface

[Tur et Pen, 1991] ont été les premiers qui ont développés la méthode, Eigenface qui sera ensuite l'une des méthodes les plus connue de la détection de visage. Le principe de cette méthode est de projeter une image dans un espace puis on calcule la distance euclidienne entre l'image original et sa projection, le codage d'une image dans un espace sert à dégrader l'information contenue dans l'image, Après l'évaluation de la distance que l'on compare à un seuil fixé a priori si la perte d'information est plus grand cela implique que l'image n'est pas bien représenter dans l'espace et elle ne contient pas une zone de visage: une classe non visage.

L'avantage de cette méthode est qu'elle donne des résultats très encourageants, mais le calcul prend beaucoup de temps.

3.2.2.2. Réseaux de neurones

Le principe de détection de visage par une classification basée sur les réseaux de neurones est d'utiliser deux ensemble d'images un pour les images de visage et autre pour les images non visage pour former le réseau de neurones, une fenêtre interchangeable balaye toute l'image en entrée. Cette fenêtre introduite au réseau sera classifiée en deux classes : classe visage et classe non-visage.

[Rowley et al, 1998] proposent un système de détection de visage utilisant la classification avec les réseaux de neurones. Les techniques de ce système sont divisées en deux étapes : la localisation des visages en utilisant un réseau de neurones et la vérification des résultats obtenus. La première étape consiste à balayer l'image

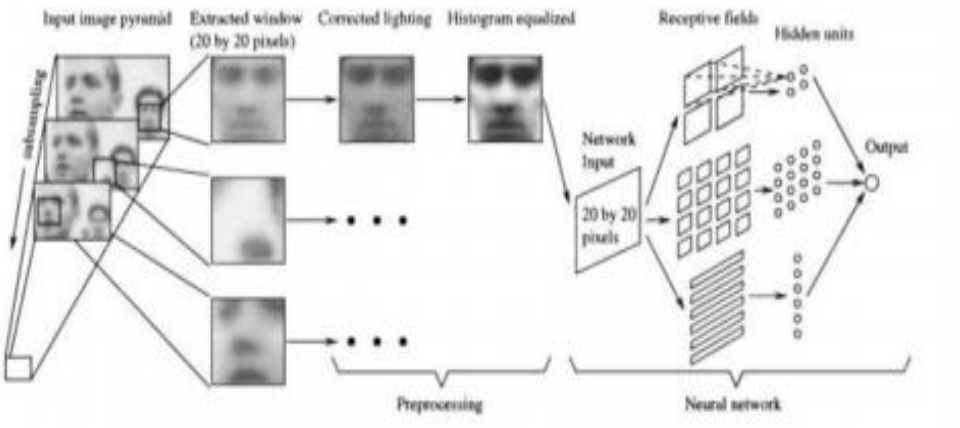


Figure 3.4 : Modèle de réseaux de neurones de Rowley et al [Row et al, 1998]

3.2.2.3. Support vector machin (SVM)

L'une des premières méthodes statistiques basée sur la théorie d'information pour la détection de visage, SVM est considéré comme un nouveau modèle de classifier d'apprentissage de fonction polynomial, réseau de neurones ou radial basis fonction (RBF).

La plupart des classifier d'apprentissage cités précédemment sont basés sur la minimisation de l'erreur d'apprentissage « l'erreur empirique », SVM opère avec un autre principe appelé « structural risk minimisation » qui a pour but de minimiser les sauts supérieurs sur les erreurs généralisées probables.

Pendant l'apprentissage pour chaque paire de pixels dans l'ensemble d'apprentissage un histogramme est utilisé pour créer des fonctions de probabilité pour les classes visage et les classes non visage parce que les valeurs des pixels dépendent des valeurs de leurs voisins.

Pour l'apprentissage [Col et Hu, 1998] ils ont utilisé un grand ensemble d'images de taille 11×11 pixels de visage et non visage, les résultats d'apprentissage forment un ensemble Look up Table (LUT) avec des rapports de probabilité, dans le but d'améliorer les performances.

3.2.3. Approches basées sur des caractéristiques invariantes

L'objectif principal de cette approche est la localisation de visage, Les algorithme de cette approche visent à trouver et chercher les caractéristiques structurales (de visage) dans des conditions variantes telles que l'éclairage, le changement de la position facile, changement d'expression…etc. Après ils étudient les caractéristiques invariantes pour localiser le visage humain, L'algorithme développé par De Silva et al [De.S, 1995], est un exemple typique des méthodes basées sur les caractéristiques.

Les Algorithmes de cette approche on peut les divisées en deux catégories qui sont :

- Basée sur les caractéristiques du visage
- Basée sur l'analyse de la couleur de la peau

3.2.3.1. Basée sur les caractéristiques du visage

Les algorithmes de cette catégorie utilisent en première étape une hypothèse sur la position du haut du visage ensuite l'algorithme de recherche parcourt le visage de haut en bas afin de trouver l'axe des yeux caractérisé par une augmentation soudaine de la densité de contours (mesurée par le rapport noir/blanc le long des plans horizontaux). La longueur entre le haut du visage et le plan de l'œil est alors utilisée comme une longueur de référence pour construire un Template. Ce Template couvrant des caractéristiques telles que les yeux et la bouche est initialisé à partir de l'image d'entrée. La forme initiale du Template est obtenue en utilisant la longueur anthropométrique en respectant la longueur de référence. Le Template flexible est alors ajusté par rapport aux positions finales des caractéristiques en utilisant un algorithme de réglage fin qui emploie une fonction de coût basée contour. Bien que ces algorithmes réussissent à détecter les caractéristiques d'ethnies différentes puisqu'ils ne se basent pas sur les informations de niveaux de gris et de couleur, ils n'arrivent pas cependant à détecter correctement ces caractéristiques si l'image du visage contient des lunettes ou bien si les cheveux couvrent le front.

Les techniques les plus connues pour cette catégorie sont : (Par Texture, Les caractéristiques faciales, Multi-caractéristiques)

3.2.3.1.1. Texture

La texture de l'être humain est distinctive et peut être utilisée pour séparer les visages par rapport à d'autres objets. [Auget al, 1993] ont développé une méthode de détection de visages sur une image en se basant uniquement sur la texture. Le calcul de la texture se fait en utilisant les caractéristiques de second ordre sur des sous- images de $16 * 16$ pixels. Dans

cette méthode trois types de caractéristiques sont pris en considération : la peau, les cheveux et le reste des composants de visage.

3.2.3.1.2. Les caractéristiques faciales

Cette technique utilise les plants d'arrêtes et des heuristiques pour supprimer tous les groupes d'arrêtes sauf celles qui représentent les contours du visage. Une ellipse est déduite comme frontière entre l'arrière-plan et le visage. Celle-ci est décrite comme étant formée des points de discontinuité dans la fonction de luminance (intensité) de l'image. Le principe de base consiste à reconnaitre des objets dans une image à partir de modèles de contours connus aux préalables. Pour réaliser cette tâche, deux méthodes seront présentées : la transformée de Hough (permettant d'extraire et de localiser des groupes de points respectant certaines caractéristiques, équation d'une forme bien déterminée), et la distance de Hausdorff (vise à mesurer la distance entre deux ensembles de points séparés).

3.2.3.1.3. Multi-caractéristiques

Nombreuses méthodes qui combinent plusieurs caractéristiques faciales pour localiser ou détecter des visages. La plupart utilisent des propriétés globales comme la couleur de la peau, la taille et la forme du visage pour trouver les candidats. Elles vérifient ensuite ces candidats en utilisant les caractéristiques locales telles que les sourcils, le nez, et les lèvres.

3.2.3.2. Les Méthodes Basées sur l'analyse de la couleur de la peau

Les méthodes de détection basées sur l'analyse de la couleur de la peau sont des méthodes efficaces et rapides. Elles réduisent l'espace de recherche de la région visage dans l'image. De plus, la couleur de la peau est une information robuste face aux rotations aux changements d'échelle et aux occultations partielles. Plusieurs espaces couleur peuvent être utilisés pour détecter la peau humaine, dans l'image les pixels qui ont la couleur de la peau, l'efficacité de la détection dépend essentiellement de l'espace couleur choisi. [A.Ver et al, 2014]

3.3. Les approches de la reconnaissance de visage

Selon [Tana et al, 2006], on peut deviser ou classer les approches de reconnaissance faciale en trois grandes catégories qui sont : les approches globales (holistiques), les approches locales, et les approche hybrides.

3.3.1. Méthodes Globales (holistiques)

Dit aussi l'approche holistiques, Le principe de cette approche selon [O'Toole et al, 1993] est de représenter l'image du visage par un seul vecteur de grande dimension n x m, en concaténant les niveaux de gris de tous les pixels du visage [A.Ver et al, 2014], il n'est pas nécessaire de repérer certains points caractéristiques du visage locaux comme les yeux, la bouche et le nez.

L'un des avantages de ces méthodes est qu'elle conserve implicitement toutes les informations de texture et de forme utiles pour reconnaitre le visage. Aussi, elle peut tenir compte des aspects d'organisation structurelle globaux du visage. Mais, l'inconvénient majeur réside dans la dimension très grande de l'espace de l'image ce qui affectera négativement sur la classification. En revanche, elles sont très sensibles aux variations d'éclairement, de pose et d'expression faciale.

Xiaoguang [X.Lu, 2003] a distingué deux types de techniques parmi les méthodes globales, les techniques linéaires et les techniques non linéaires.

Les techniques les plus populaires de l'indentification de visage de cette approche : ACP, LDA.

3.3.1.1. Analyse en composants principales (ACP) Présentation

Dite aussi Eigenfaces une méthode très populaire dans le domaine de reconnaissance proposée par [Tur et pen, 1991]. Une méthode mathématique utilisée pour simplifier et réduire les dimensions d'un ensemble de donnée et pour représenter des images de visage qui peuvent être reconstruite à partir d'un visage standard et un ensemble de points.

Le principe : Il s'agit de trouver l'ensemble des composants principaux du visage dans un ensemble d'images du visage.

- Chaque exemple de visage décrit par une combinaison linéaire des vecteurs propres.

- Transformation des visages en vecteurs.

- Détermination de la matrice de covariance.

-Détermination des vecteurs propres de la matrice de covariance formée par l'ensemble des images exemple.

- Chaque élément dans le vecteur correspond à l'intensité lumineuse d'un pixel.

L'ACP est une technique simple, populaire et rapide, propose de bons résultats dans les systèmes d'identification, ainsi que la projection est optimale pour la reconstruction d'une base de dimension réduite. Les problèmes ou plutôt les inconvénients de cette technique sont la sensibilité aux problèmes d'éclairage, expression facial et la pose.

3.3.1.2. Analyse discriminante linéaire (LDA)

Connu aussi sous le nom « Fisherfaces » [Belh et al, 1997] sont les premiers qui ont introduit cet algorithme, il effectue une séparation de classes et pour pouvoir l'utiliser il est nécessaire d'organiser une base d'apprentissage d'images en plusieurs classes, une classe par personne et plusieurs images par classe.

La 'LDA' détermine les directions de projection, et pour cela elle maximise les variations entre les images de différents individus « inter-classe » avec la minimisation des variations entre les images d'un même individu « intra-classe ». Si le nombre d'individus à traiter est plus faible que la résolution de l'image cela rend les performances de LDA faible par rapport à Eigeface [Mart et al, 2001] pour résoudre ce problème beaucoup d'autres méthodes basées sur LDA ont été développées tels que : U_LDA, O_LDA, N_LDA.

3.3.2. Méthode Locale

C'est une approche qui se base sur les caractéristiques locales du visage pour la reconnaissance faciale, tels que : le nez, la bouche et les yeux, dans cette approche contrairement à l'approche globale le visage est représenté par un ensemble de vecteurs de caractéristiques de faible dimension. (Rappel l'approche globale utilise un vecteur de très grande dimension).

L'approche locale peut être subdivisée en deux catégories :

- Catégorie 01 : Les méthodes basées sur les caractéristiques locales : L'extraction et la localisation des points caractéristiques.
- Catégorie 02 : Les méthodes basées sur l'apparence locale : La division de l'image visage en zones caractéristiques.

3.3.2.1. Les Méthodes basées sur les caractéristiques locales

Ces méthodes sont aussi subdivisées en deux groupes : Les techniques géométriques et les techniques basées sur les graphes.

3.3.2.1.1. Les Techniques géométriques

Ces techniques se basent sur l'extraction et la localisation des caractéristiques composantes du visage (les yeux, la bouche, le nez) ils utilisent principalement les coins et les points d'intérêt de ces caractéristiques.

Les techniques géométriques présentent des inconvénients sui sont :

- La difficulté d'extraction des caractéristiques géométriques dans les conditions complexes telles que la variation d'illumination.

- Les caractéristiques géométriques seules ne sont pas suffisantes pour la représentation du visage.

3.3.2.1.2. Les Techniques basées sur les graphes

C'est une représentation graphique des caractéristiques locales du visage, ces techniques formulent le problème de reconnaissance faciale comme un problème de mise en correspondance des graphes, [Man et al, 1992] a validé l'efficacité de cette technique sur une base de données de 86 images qui contient des variations des expressions faciales et de poses le résultat est représenté par un taux de reconnaissance de 90% en moyenne.

Une fois le graphe topologique construit il ne peut pas être changé, en revanche les images de visage en entré sont variantes en terme de changement d'expressions, de pose, etc. Pour résoudre ce problème plusieurs technique ont été développé tels que : Elastic Graph Matching (EGM), ElasticBuch Graph Matching (EBGM), … etc.

3.3.2.2. Les Méthodes basées sur l'apparence locale

Ces méthodes se basent principalement sur les différentes régions du visage, le modèle global est défini à partir de la combinaison des modèles locaux ce qui n'influence pas sur les régions faciales par les différentes variations, tels que : le sourire, le port des lunettes…etc. Il

existe deux paramètres pour définir les régions locales du visage : La Frome et La Taille et les caractéristiques de ces régions locales sont déterminées à l'aide d'une analyse des valeurs de niveaux gris. Cette dernière représente ou préserve les informations de texture.

3.3.3. L'Approche Hybride

Cette approche est un résultat d'une combinaison ou d'une fusion entre deux autres approches : Approche holistique ou Globale et l'approche Locale afin d'améliorer les performances des systèmes de reconnaissance, en effet les caractéristiques locales et globales sont complètement différentes, Mais chacune de ces méthodes ont ses inconvénients, mais l'une peut être complémentaire de l'autre dans le but d'améliorer la classification.

3.4. Les dessins animés dans la reconnaissance de visage

3.4.1. Les dessins animés

On dit images ou parties d'image matérielle photographique ou photos si elles ont été obtenues au moyen d'un appareil photographique (film ou fixe), par contre, on dit des images de dessins animés si elles ne contiennent aucun matériel photographique. Certaines caractéristiques particulières des dessins animés sont:

- **La couleur :** dans les dessins animés les couleurs sont peu et fortes par rapport aux scènes réelles, car l'abstraction de la transformation d'une scène du monde réel dans le monde du dessin animé conduit à une réduction des couleurs et à une exagération de la saturation.
- **La texture :** la texture dans les dessins animés est souvent présentée en manière uniforme.
- **Les contours :** sont le plus souvent des bords noir et puissants entourent des taches de couleur pour déterminer les organes de caractère de l'anime.
-

3.4.2. L'utilité de la reconnaissance de visage de dessins animés

- Problème de pirate de dessin animé : Certains pirates rééditent parfois même le dessin animé pour la publicité, créent un nouveau produit de dessin animé basé sur les personnages de dessins animés célèbres et publient leurs propres œuvres. La

reconnaissance de visage est une méthode pour résoudre le problème des pirates des dessins animés qui consiste à bloquer le chemin des pirates vers le site Web de partage, afin que le site Web de partage puisse être protégé de toute responsabilité légale pour la diffusion de dessins animés pirates. Par conséquent, il est nécessaire que le site Web de partage détecte puis rejette les dessins animés avec une déclaration de droit d'auteur.

- Amélioration des fonctionnalités de recherche : La reconnaissance de visage de dessins animés aide les animateurs à trouver des plans et des séquences spécifiques dans les archives, par exemple : si un animateur travaillant sur une nouvelle saison de programme spécifié veut trouver un référence pour faire quelque chose pour ce saison en cours, cette personne doit avoir passer des heures sur YouTube à regarder une vidéo parce que il ne peut pas trouver cela en regardant simplement les titres des épisodes. Mais avec l'aide de la technique de la reconnaissance de visage, l'animateur pourra simplement rechercher les métadonnées requises.
- Aide les logiciels de contrôle de contenu à censurer les images de dessins animés inappropriés sur les réseaux sociaux.
- Intégration dans les moteurs de recherche d'images pour rechercher sur le Web des dessins animés similaires.
- Intégration avec des lecteurs d'écran pour aider les malvoyants à comprendre les films d'animation.

3.4.3. La détection et la reconnaissance de visage de dessins animés

[Takayama et al, 2012] ont présenté les premiers travaux pertinents concernant la détection et la reconnaissance des visages de dessins animés. Pour la détection, ils ont premièrement extraitsla région de couleur de peau utilisant HVS et le contour utilisant la méthode de Canny [J.Canny, 1986], après ils ont utilisé le contour de la mâchoire et la symétrie comme deux critères pour évaluer si une région segmentée en fonction de la couleur de la peau et des bords est un visage ou non. Pour la reconnaissance de visage de dessins animés, une méthode consiste à extraire rois caractéristiques (couleur de la peau, couleur des cheveux et quantité de cheveux) de chaque image et ont distingué la classe d'images d'entrée correcte en fonction de la similitude des vecteurs de caractéristiques. Leur méthode se limite néanmoins à des images en couleur avec une couleur de peau proche de personnes réelles et cible principalement la posture frontale. Avant cela, des packages tels que *[AnimeFace, 2009]* utilisaient des architectures de perceptron simples formées sur des millions de données d'image pour juger si

les candidats de la région du visage pour les visages d'anime sont réellement des visages ou non.

[Nguyen et al, 2018] ont effectué une détection de personnages comiques en appliquant le modèle YOLOv2 *[J.Redmon, 2016]* pour prédire les coordonnées d'emplacement des cadres de délimitation par rapport à l'emplacement de la grille de cellules SxS formée sur l'image.

[B.Zhang et al, 2020] ont proposé un détecteur de visage de dessin animé asymétrique, nommé ACFD. Plus précisément, il se compose des modules suivants : une nouvelle épine dorsale VoVNetV3 composée de plusieurs modules d'agrégation à un coup asymétrique (AOSA), d'un réseau pyramidal asymétrique bidirectionnel (ABi-FPN), d'une stratégie de correspondance d'ancrage dynamique (DAM) et de la marge correspondante. En particulier, pour générer des caractéristiques avec divers champs récepteurs, les caractéristiques pyramidales à plusieurs échelles sont extraites par VoVNetV3, puis fusionnées et améliorées simultanément par ABi-FPN pour gérer les visages dans certaines poses extrêmes et avoir des rapports d'aspect disparates. En outre, DAM est utilisé pour faire correspondre suffisamment d'ancres de haute qualité pour chaque visage, et MBC est pour le fort pouvoir de discrimination.

[S.Jha et al, 2018] pour la détection des visages, ils ont intégré l'architecture MultitaskCascadedConvolutional Network (MTCNN) et la comparons aux méthodes conventionnelles. Pour la reconnaissance faciale, ils ont proposé deux contributions dont : (i) une approche d'apprentissage par transfert inductif combinant la capacité d'apprentissage des fonctionnalités du réseau Inception v3 *[C,Szegedy et al, 2016]* et la capacité de reconnaissance des fonctionnalités des machines à vecteurs de support (SVM), (ii) une proposition Cadre de réseau neuronal convolutif hybride (HCNN) formée sur une fusion de valeurs de pixels et de 15 points clés faciaux localisés manuellement.

4. Conclusion

Dans ce chapitre nous avons vu une vision générale sur le fonctionnement d'un système de reconnaissance de visage, après nous avons présenté un état de l'art sur les approches de la détection et la reconnaissance de visage. Ensuite nous avons parlé des particularités et l'utilité de la reconnaissance de visage de dessins animés afin de réaliser dans le chapitre suivant une implémentation de cette dernière.

Chapitre04

Conception et implémentation

Chapitre04 : *Conception et implémentation*

1. Introduction

La reconnaissance de visage de dessin animé est une tâche plus difficile que la reconnaissance de visage humain en raison de nombreux scénarios difficiles. Viser les caractéristiques des visages de dessins animés, telles que les énormes différences au sein des intra-faces.

Dans ce chapitre nous avons présenté notre projet qui est un système permet de reconnaitre le visage d'un caractère animé avec le Deep Learning. Dans ce qui suit, nous détaillerons les différentes étapes de la conception et de la réalisation de notre système, ainsi que les différents résultats obtenus et enfin nous allons présenter l'évaluation de ce système.

2. La conception de système proposé

Notre système proposé s'exécute en deux étapes principales :

a. Le recadrage du visage

Cette étape consiste à détecter le visage du personnage de dessin animé par dessiner un cadre entouré le visage. A partir de les cadres entourés le visage on peut recadrer les régions faciales du personnage. La méthode définit pour exécuter cette étape est expliquée comme suivante :

- Les images capturées sont prises à la même taille. De plus, toutes les images sont redimensionnées à des dimensions fixes et le rapport hauteur/largeur est entretenu.

- Après cela, l'image redimensionnée (H * L) est transmise comme entrée pour détecter les visages des personnages. Il renvoie ensuite des informations concernant le visage détecté : les coordonnées du cadre frontière (x_min, y_max, x_max , y_min,), qui sont générées autour du visage animé.

- Après avoir obtenu les cadres entourant le visage, nous recadrons la partie à l'intérieur de ces cadres afin d'obtenir des données contenant uniquement les visages à utiliser dans l'étape suivante.

L'objectif de cette étape est pour supprimer les traits supplémentaires de l'image comme (des autres objets existés dans l'image d'entrée, arrière plan…)

b. La reconnaissance du visage

L'étape de la reconnaissance du visage consiste à reconnaitre le visage d'un caractère animé avec le Deep Learning. Les images recadrées reçus en sortie de l'étape précédente sont redimensionnées et classées de sorte que chaque image se trouve dans un dossier nommé par le nom de personnage de l'image.

Nous avons utilisées les données (les images recadrées) pour entraîner le modèle DNN.

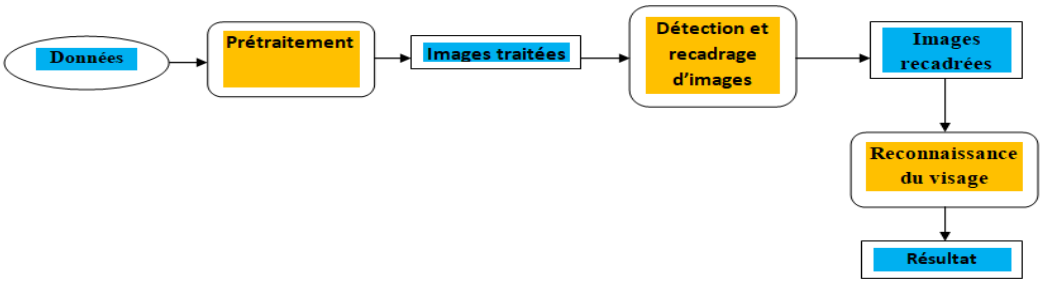


Figure 4.1: *Schéma de notre système*

3. Implémentation et résultat

a. Les données

i. La collection de données

Lorsque aucune travail ne fournit une grande quantité de données pour valider notre système, nous avons collecté un ensemble de données d'images et les annotées. Les images sont collectées à partir de google-chrome et de divers sites comme Disney.

ii. La préparation de données

Chaque image de données est augmentée avec un fichier JSON qui stocke le nom d'image, le et les coordonnées (x_min, y_max, x_max , y_min,) du visage de dessin animé correspondant. Les coordonnées (x_min, y_max, x_max, y_min,) du visage ont été marquées à l'aide de l'application labelImg (figure 4.1), qui nous donne un fichier xml correspondant à l'image annotée (figure 4.3).

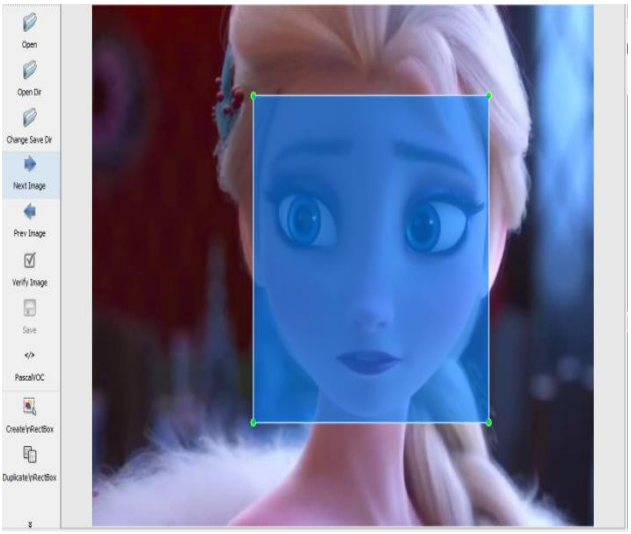


Figure4. 2 : *Exemple de l'utilisation de labelImg*

```
<annotation>
        <folder>Train_data</folder>
        <filename>image1.jpg</filename>
        <path>C:\Users\Administrateur\Desktop
        <source>
                <database>Unknown</database>
        </source>
        <size>
                <width>0</width>
                <height>0</height>
                <depth>3</depth>
        </size>
        <segmented>0</segmented>
        <object>
                <name>face</name>
                <pose>Unspecified</pose>
                <truncated>0</truncated>
                <difficult>0</difficult>
                <bndbox>
                        <xmin>135</xmin>
                        <ymin>89</ymin>
                        <xmax>431</xmax>
                        <ymax>315</ymax>
                </bndbox>
        </object>
</annotation>
```

Figure 4.3*:* *Exemple de fichier xml*

b. Le recadrage du visage avec Mask r-cnn

Mask RCNN est un réseau de neurones profonds visant à résoudre le problème de segmentation d'instances dans l'apprentissage automatique ou la vision par ordinateur. En d'autres termes, il peut séparer différents objets dans une image ou une vidéo. Nous lui donnons une image, il nous donne les cadres entourés les objets, les classes et les masques.

Dans notre travail Il y a deux étapes de Mask RCNN, génère d'abord des propositions sur les régions où il pourrait y avoir un visage en fonction de l'image d'entrée. Deuxièmement, il affine le cadre entouré le visage.

Après avoir obtenu les cadres de délimitation et les avoir affinés, le visage détecté est marqué d'un score de confiance qui signifie la confiance de la détection du visage. Comme le montre dans la figure 4.4 (l'image qui contient le caractère Jerry montre pourcentage du 99% de visage)

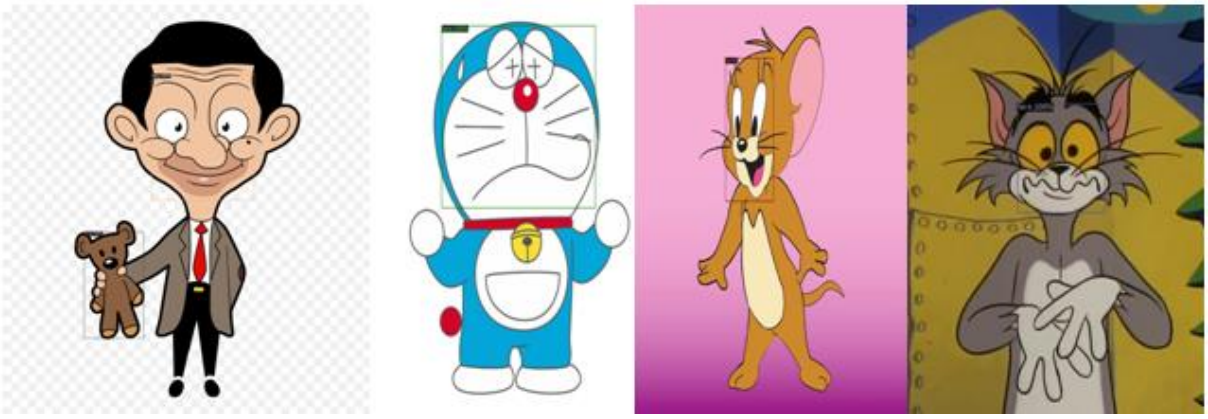


Figure 4.4 : Exemples de résultats de Mask RCNN

Après avoir obtenu les cadres entourant le visage, nous recadrons la partie à l'intérieur de ces cadres en utilisant la librairie OpenCV afin d'obtenir des données contenant uniquement les visages, où chaque image recadrée est placée dans la catégorie qui correspond au personnage de l'image. (Figure 4.5)

Figure 4.5: Images Recadrées

c. La reconnaissance du visage avec l'apprentissage par transfert

L'apprentissage par transfert utilise les poids et les connaissances acquises en résolvant un problème spécifique et en appliquant ces connaissances pour résoudre d'autres tâches similaires. Il aide à tirer parti des poids et des biais de différents algorithmes de pointe et l'utilise donc comme un avantage sans qu'il soit nécessaire d'avoir de grandes quantités de

données ou des capacités de calcul étendues. La dernière étape consiste à affiner le modèle en dégelant les parties spécifiques du modèle et en le ré-entraînant sur les nouvelles données avec un faible taux d'apprentissage. Le pipeline générique suivi pour la reconnaissance faciale des dessins animés est le suivant :

a- Prétraitement des images recadrées :

Les images recadrées (de taille 256 * 256) sont reçus en sortie de la phase de recadrage de visage à l'aide de Mask R-CNN. Ces images ont été redimensionnées à 224*224 pour la classification des noms de personnages à l'aide du modèle inceptionV3. Ces images sont ensuite converties en tenseurs. La valeur de ces tenseurs se situe dans la plage de 0 à 255, normalisée dans la plage de 0 à 1. Ensuite, des lots de données sont créés, chacun ayant 32 images à entrer dans le modèle de classification des émotions. La figure 10 montre un exemple de lot de données créé.

b- Formation et classification

Le réseau de neurones profonds inceptionV3 est entrainé à l'aide de l'apprentissage par transfert à partir des lots de données créés (figure 4.5). Un instantané des résultats obtenus par cette approche de bout en bout proposée est illustré à la (Fig 4.6). Par exemple, 0,95577 est un score de confiance représentant « pink panther ».

Figure 4.6: Résultat de système proposé

Chapitre04 : *Conception et implémentation*

4. Environnement de l'implémentation

4.1. **Environnement matériel :** Nous avons utilisé un ordinateur qui à les caractéristiques suivants :

- Type : PC/ Microsoft surface Pro 4.

- Processeur : Intel® Core™i3CPU @

-Mémoire installée (RAM) : 4.00Go.

-Type Système : Système d'exploitation 64bits, processeur x64.

4.2. Environnement logiciel

Pycharm Assistance intelligente pour Python PyCharm, fournit la saisie automatique de code intelligente, des inspections de code, la mise en évidence d'erreur à la volée et des correctifs rapides, en plus de refactorisations de code automatisées et de riches capacités de navigation.

Python Python est un langage de programmation qui vous permet de travailler plus rapidement et d'intégrer vos systèmes plus efficacement. L'index de package Python (PyPI) héberge des milliers de modules tiers pour Python.

Colab 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.

Keras est une API conçue pour les êtres humains, pas pour les machines. Keras suit les meilleures pratiques pour réduire la charge cognitive : il propose des API

cohérentes et simples, il minimise le nombre d'actions utilisateur requises pour les cas d'utilisation courants et il fournit des messages d'erreur clairs et exploitables. Il contient également une documentation complète et des guides de développement.

TensorFlow est une plate-forme Open Source de bout en bout dédiée à l'apprentissage automatique. Elle propose un écosystème complet et flexible d'outils, de bibliothèques et de ressources communautaires permettant aux chercheurs d'avancer dans le domaine de l'apprentissage automatique, et aux développeurs de créer et de développer facilement des applications qui exploitent cette technologie.

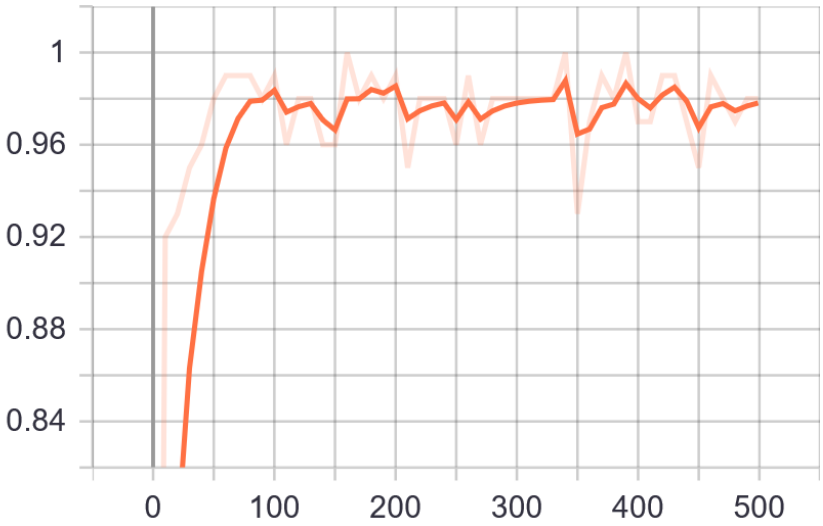
OpenCV (Open Source Computer Vision Library) est une bibliothèque de fonctions de programmation principalement destinées à la vision par ordinateur en temps réel. La bibliothèque est multiplateforme et gratuite pour une utilisation sous la licence open source Apache 2. À partir de 2011, OpenCV propose une accélération GPU pour les opérations en temps réel.

Tensorboard permet d'inspecter visuellement et d'interpréter vos graphes et exécutions TensorFlow. Il exécute un serveur web qui gère une page web afin d'afficher les visualisations TensorBoard et de vous permettre d'interagir avec elles.

5. Evaluation de la performance du modél

Pour l'évaluation du note modèle utilisé inceptionV3 pour classifier les visages des caractères de dessin animés nous avons appliqué deux fonctions de mesures :

- **Accuracy (Précision) :** La précision est une métrique qui décrit généralement les performances du modèle dans toutes les classes. Il est utile lorsque toutes les classes



sont d'égale importance. Il est calculé comme le rapport entre le nombre de prédictions correctes et le nombre total de prédictions. Pour notre modèle inceptionV3 la précision

Figure 4.7 : accuracy vs le nombre de pas

a donné 98% pour 500 pas. (figure 4.7).

-Cross entrorpy: est une mesure du domaine de la théorie de l'information, s'appuie sur l'idée d'entropie de la théorie de l'information et calcule le nombre de bits nécessaires pour représenter ou transmettre un événement moyen d'une distribution par rapport à une autre distribution. Lors de l'ajustement des poids du modèle pendant l'entraînement. L'objectif est de minimiser la perte, c'est-à-dire que plus la perte est petite, meilleur est le modèle. Un modèle parfait a une perte d'entropie croisée de 0. Normalement, il sert aux classifications multi-classes et multi-étiquettes. Pour notre modèle cross entropy a donné valeur 0.092.

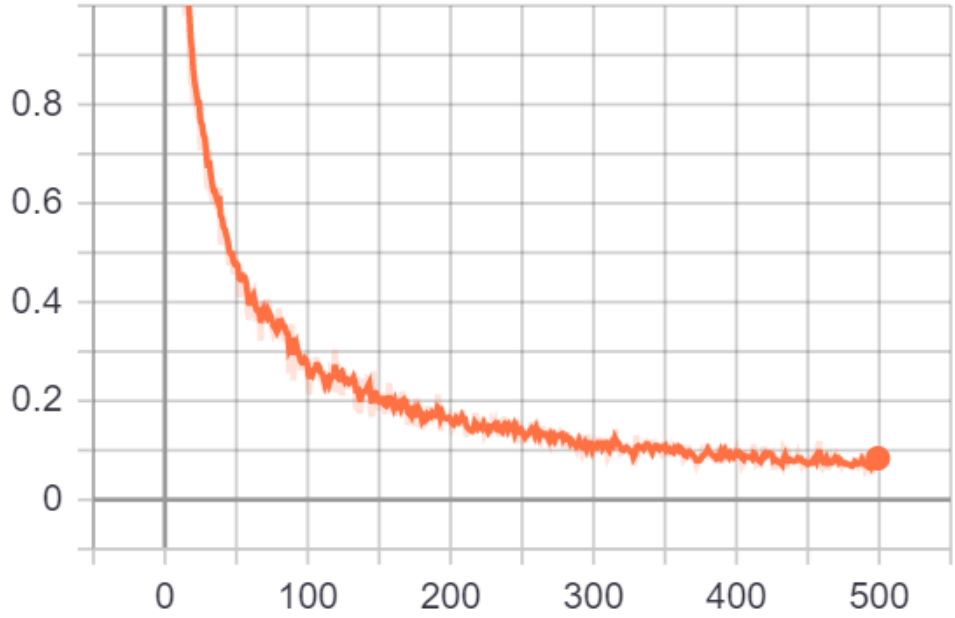


Figure 4.8 : cross entropy vs le nombre de pas

6. Conclusion

Dans ce chapitre nous avons parlé de la conception de notre système qui consiste à reconnaitre le visage de dessin animé en utilisant le deep Learning, ensuite nous avons détaillé l'implémentation de notre système qui utilise le modèle Mask r-cnn pour détecter la zone du visage, après nous avons encadré les régions des visages qui sont détectés avec Mask r-cnn pour obtenir des images contient que les visages afin de les utiliser pour entrainer le modèle DNN inceptionV3. Le résultat donnant par le système est le nom de caractère avec score de confiance. Pour l'évaluation nous avons appliqué la fonction de précision qui a donné 98%.

Conclusion générale

La reconnaître des visages autres que les humains est un problème intéressant et difficile. Bien que la littérature actuelle se soit efforcée de détecter et de reconnaître des objets, la reconnaissance de visages de dessin animes n'a pas été largement couverte. Par conséquent, dans ce travail, nous présentons une approche intégrée de réseau de neurones profonds (DNN) qui reconnaît avec succès les personnages d'images de dessins animés. Nous avons collecté un ensemble de données et l'avons renvoyé. L'approche DNN intégrée proposée a été formée sur l'ensemble de données et a correctement segmenté les masques faciaux, reconnaissant le caractère résultant avec une précision de 98%. L'approche Mask R-CNN de la découverte de personnages et un modèle d'apprentissage en profondeur moderne, InceptionV3, ont utilisé l'approche de la caractérisation.

L'ouvrage sera utile aux animateurs, illustrateurs et dessinateurs de presse. Il peut également être utilisé pour créer un système de recommandation permettant aux utilisateurs de choisir des personnages de dessins animés. L'étude de la reconnaissance faciale des dessins animés extrait également d'autres informations associées aux alliés, qui, combinées à l'intelligence artificielle, pourraient ouvrir un grand nombre d'opportunités.

Bibliographie

[A. kri et I. Su, 2012] A. Krizhevsky, I. Sutskever et G. E. Hinton. *"ImageNet Classification with Deep Convolutional Neural Networks ", Advances in neural Processing Systems de traitement. 2012*

[A. Metref, 2010] A. Metref, "*Contribution à l'étude du problème de synchronisation de porteuse dans le contexte de la Radio Intelligente*", thèse de doctorat, université de Rennes, 2010.

[AnimeFace, 2009] https://github.com/nagadomi/animeface-2009

[A.schm, 2016] A. Schmitt and B. Le Blanc, "Les réseaux de neurones artificiels," vol. 13, 2016.

[Aug et al, 1993] M.F.Augustejin et al "Identification of human faces through texture based feature recognition and neutral network technologie ", Proc. IEEE Conf. Neutral Network, pp. 392- 398, 1993.

[A .Verma et al, 2014] A .Verma et al, "Face detection using skin color modeling and geometric feature". International conference on informatics, electronics and vision (ICIEV). IEEE, pp 1–6,2014.

[Belh et al, 1997] P. Belhumeuret al, "fisherfaces: recognition using class specific linear projection", IEEE Trans. Pattern Anal. Mach. Intell. 19 (7) 711–720, 1997.

[B.Ols, 1996]Olshausen, B. A. & Field, D. J. (1996). "Emergence of simple-cell receptive-field properties by learning a sparse code for natural images". Nature. 381 (6583): 607–609.

[B.Zhang et al, 2020] B.Zhang et al, " ACFD: Asymmetric Cartoon Face Detector", Computer Vision and Pattern Recognition (cs.CV), 2020.

[Col et al, 1998] J .Colmenarez et al, "Pattern detection with information-based maximum discrimination and error bootstrapping, in Proc", Of International conference on pattern recognition, 1998.

[C.Szegedy et al, 2016] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the inception architecture for computer vision,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2818–2826.

[De.S, 1995] L. De Silva et al, "Detection and tracking of facial features by using a facial feature model and deformable circular template", IEICE Trans. Inform. Systems E78–D(9), 1195–1207, 1995.

[Fli, 1995]M. Flickher, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D.Steele, and P. Yanker, "*Query by Image and Video Content: The QBIC System*," IEEE Computer, vol. 28, no. 9, pp. 23-32, Sept. 1995.

[*Fuhui Long, 2003]*Fuhui Long, Hongjiang Zhang, David D. Feng: "*Fundamentals of Content-based Image retrieval*", in Multimedia Information Retrieval and Management – Technological Fundamentals and Applications, D. Feng, W.C. Siu, and H.J.Zhang. Springer, 2003.

[G.Quellec, 2008]GwénéléQuellec. "Indexation et fusion multimodale pour la recherche d’information par le contenu". Application aux bases de données d’images médicales. Université européenne Bretagne. Thèse de Doctorat, Septembre 2008.

[Haralick et all, 1973]R.M.Haralik, K. Shanmugam et I. Dinstein, "Textural features for images Classification". IEEE Transaction on System , Man, Cybernetics, 3,610-621, 1973.

[J. M. Guo et al, 2015] J. M. Guo, H. Prasetyo, and J. H. Chen, “*Content-based image retrieval using error diffusion block truncation coding features*,” *IEEE* Transactions on Circuits and Systems for Video Technology, vol. 25, no. 3, pp. 466–481, 2015.

[J.Canny, 1986] J. Canny, “A computational approach to Edge detection”, IEEE Trans. PAMI, Vol. 8, No. 6, pp. 679-698, 1986.

[J.Redmon, 2016] J. Redmon and A. Farhadi, “Yolo9000: better, faster, stronger,” arXiv preprint, vol. 1612, 2016.

[J.Russ, 2010] Stuart J. Russell, Peter Norvig (2010) "*Artificial Intelligence: A Modern Approach*". 2010

[J.Russ et al, 2010] Stuart J. Russell, Peter Norvig "Artificial Intelligence: A Modern Approach", 2010.

[kanade, 1973] T .Kanade, “Picture Processing by Computer Complex and Recognition of Human Faces,” PhD thesis, Kyoto Univ., 1973.

[Kato, 1992]Kato, andToshikayo., 1992.“*Database architecture for large image gallery management*"– IEEE Transactions on multimedia, 1992.

[Kotro et al, 1997] C. Kotropoulos et al, « Rule-Based Face Detection in Frontal Views. Proc. Int’l Conf. Acoustics, Speech and Signal Processing, vol. 4, pp. 25372540, 1997

[L.Gue, 2007] L. Gueguen. "*Extraction d’information et compression conjointes des séries temporelles d’images satellitaires*". Ecole Nationale Supérieure des Télécommunications Paris. Thèse de Doctorat, Octobre 2007.

[L.Gue, 2008]L. Gueguen. "*Extraction d’information et compression conjointes des séries temporelles d’images satellitaires*". Ecole Nationale Supérieure des Télécommunications Paris. Thèse de Doctorat, Octobre 2007.

[Lynda, 2019]Conférence:"Proceedings of the 2nd Conférence Internationale sur l'Informatique et ses Applications" (CIIA'09), Saida, Algeria, May 3-4, 2009

[M. A. Bou, 2009] M. A. Bourenane. "Un outil pour l’indexation des vidéos personnelles par le contenu". Université de Québec à trois- rivières. Thèse de Doctorat, 2009.

[Mallat , 1989] S.G.Mallat, A theory for multiresolution signal decomposition: the wavelet representation, IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 11, pp. 674-693, 1989.

[Man et al, 1992] B.S. Manjunath et al,"A feature based approach to face recognition", in: Proceedings, IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 373–378, 1992.

[Mar, 1980]Marčelja, S. (1980). "Mathematical description of the responses of simple cortical cells". *Journal of the Optical Society of America*. **70** (11): 1297–1300.

[Mart et al, 2001] A. Martinez,et al, "PCA versus LDA". IEEE Trans. Pattern Anal. Mach. Intell. 23 (2) 228– 233, 2001.

[Mehri et al, 2012] Mehryar Mohri, Afshin Rostamizadeh, Ameet Talwalkar "Foundations of Machine Learning", 2012.

[Min, 1996] T.P.MinkaAa Image Database Broxser that Learn from User Interaction, Master of Engineering Thesis, 1996.

[N.P.Rou] N. P. Rougier "*Perceptron simple Perceptron multi-couches*.", Cours en ligne pour les Master 2 - Sciences cognitives, université de Bordeaux.

[O'Toole et al, 1993] A.J. O'Toole et al. Low-dimensional representation of faces in higher dimensions of the face space, Opt. Soc. Am. 10 (3), 405–411,1993.

[P.Tian, 2013]D. Ping Tian, "A review on image feature extraction and representation techniques," International *Journal of Multimedia and Ubiquitous Engineering*, vol. 8, no. 4, pp. 385–396, 2013.

[Rowley et al, 1998] HA .Rowley et al "Neural Network based Face Detection", IEEE Trans, Pattern Anal. Mach, Intell, 23-38, 1998

[Sinha, 1994] P .Sinha, "Processing and Recognizing 3D Forms," PhD thesis, Massachusetts Inst. Of Technology, 1995.

[S.Jha et al, 2018] *S.Jha et al,* "Bringing Cartoons to Life: Towards Improved Cartoon Face Detection and Recognition Systems", 2018.

[Smi, 1996] J. R. Smith and S.-F. Chang, "*Querying by color regions using the VisualSEEk content-based visual query system*", In Intelligent Multimedia InformationRetrieval. IJCAI, pages 159-173, 1996.

[SO ,1995] M.A. Stricker and M. Orengo. "*Similarity of color images . In SPIE, Storage and Retrieval for image Video Databases* ", Pages 381-392, 1995

[S.Pan, Q.Yan, 2010] S. J. Pan and Q. Yang. *"A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering",* pages : *1345–1359, 2010.* and Retrieval of image Video Databases. Pages 381-392, 1995.

[Striet Ore, 1995]M.A.Striker and M.Orengo. "*Similarity of color images*". InSPIE, Storage

[Takayama et al, 2012] K. Takayama, H. Johan, and T. Nishita, "Face detection and face recognition of cartoon characters using feature extraction," in Image, Electronics and Visual Computing Workshop, 2012

[Tana et al, 2006] X. Tana et al, " Face recognition from a single image per person: A survey. Pattern Recognition", 2006

[Tur et Pen, 1991] M .Turk et al,"Eigenfaces for Recognition", J. Cognitive Neuroscience, vol. 3, no. 1, pp. 71-86, 1991.

[X. Lu, 2003] X.Lu " Image Analysis for Face Recognition ", Dept. Of Computer Science & Engineering Michigan State University, 2003.

[Yan et al 2002] Ming-Hsuan Yang, David J. Kriegman et Narendra Ahuja. " Detecting faces in images : A survey. " IEEE Transactions on Pattern Analysis and Machine Intelligence, volume 24(1), pages 34–58, 2002.

[Yang et Huang ,1994] G. Yang and T. S. Huang. "Human Face Detection in Complex Background," Pattern Recognition, vol. 27, no. 1, pp. 53-63, 1994

[Yuille et al, 1992]AL .Yuille et al, "Feature extractiong from faces using feformale templates".Int, J.Comput. Vis, 8,99-112,1992

[Zhang et Lu, 2004]D. Zhangand G. Lu, "*Review of shaperepresentation and description techniques*," Pattern Recognition, vol. 37, no. 1, pp. 1–19, 2004.

Les sites web

[W, 1] https://datascientest.com/convolutional-neural-network 25 juin 2020

[W, 2] *https://www.kdnuggets.com/2018/05/general-approaches-machine-learning-process.html* mai 2015

[W, 3] http://yann.lecun.com/exdb/lenet/ (consulté le 16 nov 2013)

[W, 4] *https://datascientest.com/convolutional-neural-network* 25 juin 2020