REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE
UNIVERSITE ECHAHID HAMMA LAKHDAR D'EL-OUED
FACULTE DES SCIENCES EXACTES
DEPARTEMENT: INFORMATIQUE

Mémoire de Fin D'étude

Présenté pour l'obtention du Diplôme de

MASTER ACADEMIQUE

Domaine : Mathématique et Informatique

Filière : Informatique

Spécialité : Systèmes Distribués et Intelligence Artificielle

Présenté par :

- BERGUIGA ASSIA
- DHEQUIR ANISSA

Thème

La détection de textes d'images de scènes naturelles à l'aide de l'apprentissage profond

Soutenue le: ..... Devant le jury:

Professeure HAMMOUD MERIEM..... **Président**

Professeure RTIMA FARIDA.....**Rapporteur**

Professeure GUIA SANA SAHAR..... **Encadreure**

Année Universitaire: 2021/2022

REMERCIEMENT

Nous tenons remercier avant tout DIEU qui nous a donné la force, la volonté ,

le courage, et la patience de pouvoir réaliser ce modeste travail.
On tient aussi remercier Proffessure Guia Sana Sahar , notre encadreure qui a su orienter notre travail, aussi pour sa disponibilité nous prodiguer des conseils, pour sa confiance et pour sa précieuse aide, on le remercie du fond du cœur.
Nous sinc res remerciements sont adress s tous les membres du jury qui ont accept de juger notre modeste travail, Nous, remercions infiniment tout le staff enseignant et administratif au département d'Informatiques, et surtout ceux et celles qui nous a enseignés durant la période d'études.
On remercie également tous ceux qui ont participé de près ou de loin à élaborer ce travail

Dédicace

A l'âme de mon père

À ma très cher mère adorée qui m'a aidé, grâce à leur prière et à leur bénédiction.

A tous mes frères ,a mes trés chères sœurss , A ma grande famille et tous mes amis surtout ma meilleure amie rahima benmeriem. À tous les étudiants de la faculté Informatique surtout les étudiants de la 2éme année master promotion 2022

Assia

Dédicace

A mes trés chèrs parents

pour tous les soins et le suivi dont ils font preuve depuis ma naissance et au long de mes tudes pour leurs soutien et surtout leurs conseils et amour

A mes trés chères sœurs

A mes trés chèrs frères

A ma grande famille

A tous mes amis

A tous ceux qui m'aiment et que j'aime

Anissa

RÉSUMÉ :

La détection et la reconnaissance de texte est un champ d'étude qui a longtemps intéressé la communauté des chercheurs. Le problème de la détection et de la reconnaissance d'un texte a de nombreuses solutions en utilisant différentes techniques. L'une des solutions utilisées et développées récemment est l'utilisation de l'apprentissage profond (Deep Learning) qui est un type d'intelligence artificielle dérivé de l'apprentissage automatique (Machine Learning) où la machine est capable d'apprendre par elle-même, il a permis d'obtenir d'excellents résultats et une grande précision dans plusieurs champ de la vision par ordinateur. Le Deep Learning s'appuie sur un réseau de neurones artificiels s'inspirant du cerveau humain Ce réseau est composé de dizaines voire de centaines de « couches » de neurones, chacune recevant et interprétant les informations de la couche précédente.

Dans ce travail, nous avons effectué la détection de texte dans les images de scènes naturelles en utilisant des méthodes de traitement d'images basées sur le réseau de neurones convolutifs (CNN: Convolutional neural network), la méthode proposée s'appuie sur le principe de la segmentation sémantique afin de localiser les régions de l'image contenant du texte dans une scène naturelle, où le texte peut être dans différentes langues, couleurs, polices, tailles, orientations et formes.

Mots clés: Localisation de Texte, Apprentissage profond, Réseau de neurones convolutifs, Détectionde texte de scène, Segmentation sémantique.

ABSTRACT :

Text detection and recognition is an active research area that has attracted increasing attention in the research community of computer vision. The problem of text detection and recognition has manysolutions using different techniques.

Deep Learning is a type of artificial intelligence technique derived from machine learning approaches where the machine is able to learn by itself. Exploring the powerful Deep Learning technology is one of recently solutions used and developed in computer vision which has achieved excellent results and high precision. Deep Learning based on artificial neurons network inspired by the structure and function of human brain. This network consists of multi-layer networks of neurons, each one receiving and interpreting information from the previous layer.

In this work, we performed text detection in images of natural scenes using image processing methods based on Convolutional Neural Network (CNN), the proposed method relies on the principle of semantic segmentation in order to localize the region of scene text with different languages,colors, fonts, including various shapes, scales, and multi-oriented text.

Keywords: Text Localization, Deep Learning, Convolutional Neural Network, Scene Text Detection, Semantic Segmentation.

ملخص:

يعد إكتشاف النص و التعرف عليه مجالا للدراسة طالما أثار إهتمام الباحثين.حيث أن لمشكلة الكشف على النص و التعرف عليه العديد من الحلول بإستخدام تقنيات مختلفة.أحد الحلول التي تم إستخدامها و تطويرها مؤخرا هو التعلم العميق (Deep Learning) و هو نوع من الذكاء الإصطناعي مشتق من التعلم الآلي (Machine Learning) حيث تكون الآلة قادرة على التعلم بنفسها و قد حققت نتائج ممتازة و دقة عالية في مختلف تطبيقات الرؤية الحاسوبية . يعتمد التعلم العميق على شبكة من الخلايا العصبية الإصطناعية المستوحاة من الدماغ البشري .تتكون هذه الشبكة من عشرات أو حتى مئات "طبقات" الخلايا العصبية ،كل منها يتلقى و يفسر المعلومات من الطبقة السابقة .

في هذا العمل، أجرينا إكتشاف النص في صور المشاهد الطبيعية بإستخدام طرق معالجة الصور المعتمدة على الشبكة العصبية التلافيفية (Convolutional Neural Network : CNN) ، و تعتمد الطريقة المقترحة على مبدأ التجزئة الدلالية من أجل تحديد مناطق الصورة التي تحتوي على النص في مشهد طبيعي ، حيث يمكن أن يكون النص بلغات و ألـــوان وخطوط و أحجام و إتجاهات و أشكال مختلفة.

الكلمات المفتاحية : توطين النص، التعلم العميق،الشبكة العصبية التلافيفية ، إكتشاف نص المشهد ، التجزئة الدلالية .

TABLE DES MATIÈRES

TABLE DES FIGURES

INTRODUCTION GÉNÉRALE

Dans un monde numérique, les informations sont stockées, traitées, indexées et recherchées par des systèmes informatiques, ce qui réalise des tâches rapides et pas cher. Le texte dans la scène transmet généralement des informations sémantiques précieuses. Ainsi, la détection de texte dans des images naturelles a récemment attiré une attention croissante dans la communauté de la vision par ordinateur, car la perception d'informations est un élément essentiel de l'intelligence artificielle. En tant qu'élément indispensable des systèmes de reconnaissance optique de caractères (OCR Optical Character Recognition), la détection de texte de scène est essentielle à la reconnaissance de texte ultérieure. Cependant, Cette tâche, est difficile en raison des attributs variables des images naturelles, tels que le degré de flou de l'image, les conditions d'éclairage ...etc

Avec l'avènement de l'apprentissage en profondeur (Deep Learning), en combinaison avec du matériel de l'Intelligence Artificielle et des GPU , des performances exceptionnelles peuvent être obtenues sur les tâches de vision par ordinateur. Par conséquent, l'apprentissage en profondeur a apporté de grands succès dans l'ensemble du domaine de la reconnaissance d'images, de la reconnaissance faciale et les algorithmes de classification d'images atteignant des performances supérieures au niveau humain et au niveau de la détection d'objets en temps réel. à cause d'avancement dans le deep learning et le besoin croissant d'automatisation des systèmes a affecté la détection et de la reconnaissance de texte à partir d'images dans une large mesure , la variétés des entrées que nous pouvons avoir tell que des informations non textuelles rendre la localisation et la séparation du texte, constituent le plus gros problème dans la procédure d'océration , allant d'une simple image d'un livre dans un éclairage et des circonstances parfaites à une image d'une scène naturelle avec tant d'imperfections.

Plusieurs approches basées deep learning sont proposées dans la littérature.

Dans ce travail, nous proposons une approche qui définie un modèle deep learning basée sur la segmentation sémantique et l'attention visuelle.

En va suivre la structure suivante pour répondre a notre objectif :

Introduction général.

Chapitre 1 : Détection de texte et reconnaissance

Dans ce chapitre nous soulignons à l'exploration de la détection et de la reconnaissance de texte, nous donnons un aperçu de ce qu'est ce problème et nous représentons les différentes techniques et approches proposées.

Chapitre 2 : Les approches basée sur l'apprentissage en profondeur(deep Learning)

Dans ce chapitre est défini le principe de deep learning et représenté les différentes techniques et approches deep learning proposées pour la détection de texte de scène naturelle

Chapitre 3 : Méthodologie de la detection de texte dans les scénes naturelles

Nous abordons la tache de segmentation et plus précésiement la segmentation sémantique

Chapitre 4 : Réalisation de systéme

Les différents étapes de code et les résultats obtenus sont présentés

Puis nous terminons par **Conclusion général.**

CHAPITRE 1

DÉTECTION DE TEXTE ET RECONNAISSANCE

1.1 Introduction

Le besoin d'extraire des informations textuelles de différentes sources s'est accru dans une large mesure, les études récentes de la vision par ordinateur nous permettent de faire de grands progrès en allégeant le fardeau de la détection de texte et d'autres analyses et compréhension de documents.dans Computer Vision, la méthode de conversion du texte présente dans les images ou les documents numérisés en un format lisible par machine qui peut ensuite être modifié, recherché et utilisé pour un traitement ultérieur est connue sous le nom de reconnaissance optique de caractères (OCR) qui utilisé pour la récupération d'informations et saisie automatique de données et joue un rôle très important pour de nombreuses entreprises et institutions qui ont des milliers de documents à traiter, analyser et transformer pour effectuer les opérations quotidiennes. Par exemple : dans les aéroports, tandis que le passeport vérifiant les informations peut également être extrait à l'aide de l'OCR, factures, formulaires, relevés, contrats,. . .etc.

1.2 Objectif et Motivation

1.2.1 Objectif :

La détection de textes des images de scènes naturelles est un problème fondamental qui a de nombreuses applications. Pour une scène naturelle donnée, le problème consiste à :

- Localiser la région textuelle dans l'image par un cadre englobant (La détection des objets).
- Reconnaitre les textes localisées qui peuvent être de n'importe quelle langue. (La génération automatique de phrase descriptive).

Le diagramme de flux de travail général est présenté par le diagramme ci-dessous :

FIGURE 1.1 – Le diagramme de flux.

En entrée une image contient un texte et produit un texte en sortie,cette tâche était presque impossible, même pour les chercheurs les plus avancé en vision par ordinateur avant le développement récent des réseaux de neurones profonds, avec le Deep Learning (L'apprentissage profond) , si nous disposons du jeu de données requis le problème est résolu facilement.

1.2.2 Motivations :

Afin de bien comprendre l'importance de ce problème dans les scénarios du monde réel, Il y a quelques applications où ce problème utilise une solution tell que :

-Nous pouvons atteindre un système de conduite autonome, si nous pouvons bien décrire la scène autour de la voiture. la voiture autonome utilise l'OCR pour reconnaître les panneaux de signalisation et ainsi prendre des mesures en conséquence. Nous pouvons créer des applications pour les aveugles qui les guideront sur les routes sans l'aide de personne.

Nous pouvons le faire en convertissant d'abord la scène en texte, puis le texte en son.

- La description automatique peut aider à rendre Google Recherche d'images aussi efficace que Google Recherche, car chaque image peut d'abord être convertie en description textuelle, puis la recherche peut être effectuée sur la base de cette description.

1.3 Définition de problème :

Nous pouvons présenter le probléme de ce travail sous forme d'une question : **Qu'elle est le texte affiche sur l'image ci-dessous ?** .

FIGURE 1.2 – Exemple d'image.

Dans notre travail, la détection et la reconnaissance de texte à partir d'image de scènes naturelles seront abordées.

Est-ce que nous pouvant écrire un programme informatique ou nous utilisons en entrée n'importe quelle image ou scène naturelle (pas particulièrement des documents) et atteindre une segmentation de ce texte en sortie ?

1.3.1 La détection de texte :

Un texte est une association de caractères appartenant à un alphabet, réunis dans des mots d'un vocabulaire donné ,pour détecte un texte il faux retrouver ces caractères puis les reconnaître.

1.3.2 L'importance :

- Enrichissement de la navigation dans le viewer.
- Enrichissement des index textuel pour la recherche de lieux par mot-clés.
- Utiliser la detection automatique dans les application de :
- Annotation automatique de bases de données d'images.
- Aide aux personnes malvoyantes.
- Navigation en riche en ville.

1.3.3 Localisation de la zone de texte :

Localiser la boîte englobante ou la région de chaque instance de texte.

1.3.4 Défis-lors-de-la-détection-et-de-la-reconnaissance-de-texte-dans-la-scène :

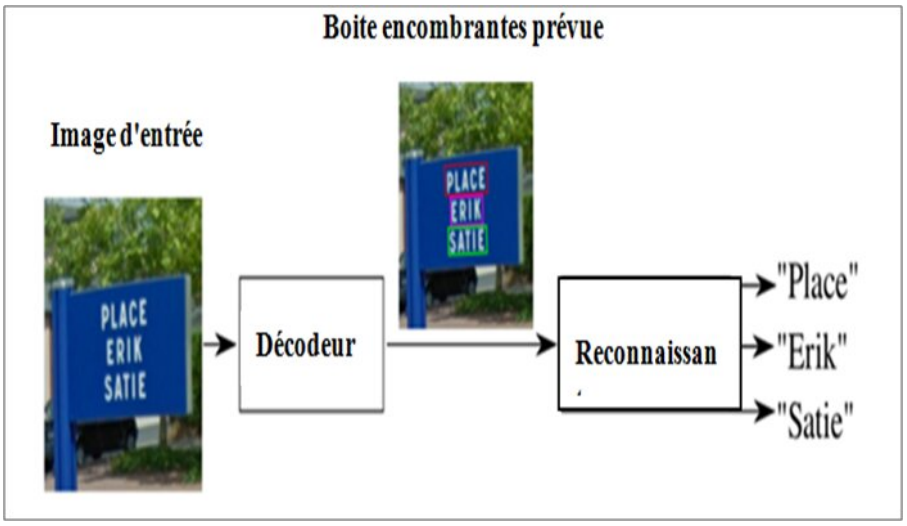


FIGURE 1.3 – détection et la reconnaissance de texte

Le besoin de vision par ordinateur dans la détection de texte et la reconnaissance d'une image ou d'une vidéo devient très populaire ces jours-ci, Parce que le texte doit diffuser ou acquérir des informations de manière fiable et efficace dans le temps et dans l'espace.
Cette approche peut être utilisée pour la reconnaissance de l'écriture manuscrite, la détection et la reconnaissance de texte de scène naturelle, la détection et la reconnaissance de numéro de véhicule, et bien d'autres.
Plusieurs défis peuvent encore être rencontrés lors de la détection et de la reconnaissance du texte dans la scène.

• Images floues :

Le flou dans les images est un phénomène essentiellement convolutif. Il est en grande partie, dû au fait que la profondeur de champ d'un appareil photographique ne peut être infinie.
De plus, les cameraphones sont souvent fabriqués avec des focales fixes, seule la vitesse d'obturateur est commandée. En outre, les capteurs utilisés dans les caméraphones sont très limités dans des conditions de luminosité faible. Dans ces conditions, on récupère souvent des images bruitées et floues.
Il existe plusieurs sortes de floue :
- Défocalisation lorsque le plan de l'image ne coïncide pas avec le plan du capteur .
- Flou de bougé si c'est le mouvement de l'utilisateur qui crée le flou dans l'image.
- Turbulences atmosphériques ou le flou soit causé par des (température, du vent, humidité . . .ect).

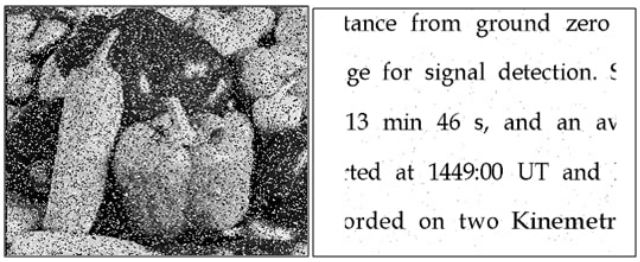


FIGURE 1.5 – images bruyante.

Le flou dans une image est complètement caractérisé par sa PSF,ces PSF ou noyauxde flou sont décrits par une ligne pour les flous de bougé, un disque pour les flous dedé focalisation et enfin par un noyau gaussien pour les flous atmosphériques.

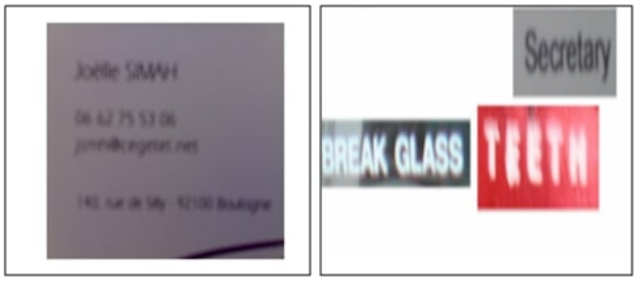


FIGURE 1.4 – images floues

- **- Images buitées :**

Le bruit est un signal aléatoire désigne les pixels de l'image dont l'intensité est très différente de celles des pixels voisins.

Le bruit peut provenir de causes suivants :

- Environnement lors de l'acquisition ,de numérisation, de transmission.

- La qualité du capteur.

- La qualité de l'échantillonnage.-

Ce que dégrade la qualité de l'image du document et gène les algorithmes de reconnaissances des formes intégrés dans les moteurs de l'OCR .

Le principe de la detection est basé sur l'analyse de la dynamique des niveaux de gris des K régions locales homogènes de l'image observe.la nature du bruit est impulsionnelle si la distribution est uniforme et proche de 255 (image codée sur 8 bits) sinon, il est additif ou multiplicatif. Cela suppose que la distribution du bruit est uniforme et prend ses valeurs sur l'intervalle de la dynamique des luminances de l'image degrade.[1]

- **Grande variété des polices de caractéres :**

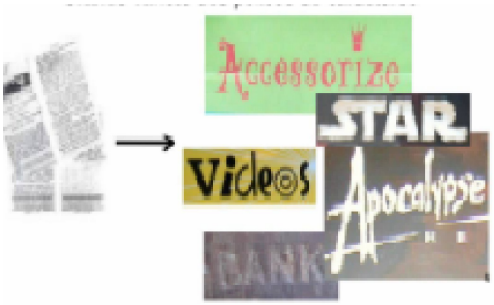


FIGURE 1.6 – la variétée des polices

- **Variations de luminosité :**

FIGURE 1.7 – la variétée de luminosité

- **Multi-orientées :**

Le sens de l'écriture est supposé horizontal pour la plupart des OCRs.

Une tolérance est souvent admise pour gérer des documents imprimés ou numérisés légèrement en oblique,certains documents comportent des textes écrits sur l'axe vertical, notamment dansle cas de tableaux ou de légendes.

Si l'OCR est paramétré pour être capable de détecter des textes verticaux, il sera de toute façon perturbé par la gestion d'hypothèses de blocs de texte d'orientation différentes.

Le traitement de l'OCR est facile lorsque les lignes de l'écriture sont horizontales.Si ce n'est pas le cas, il faut estimer l'inclinaison du texte. Ce traitement supplémentaire n'est pas toujours présent dans l'OCR. Par ailleurs, ci ce traitement n'est pas parfait, la reconnaissance sera dégradé.

La reconnaissance doit être invariante par rotation et changement d'échelle des symboles et caractères , En effet, de nombreuses chaînes de caractèreset de symboles sont présentes en de multiples orientationset en plusieurs tailles tell que :[2]

- **Pivotée**

FIGURE 1.8 – texte pivotéé

- **Courbe**

FIGURE 1.9 – texte courbe

- **Shape**

FIGURE 1.10 – texte circulaire

1.4 Les approches de détection de textes :

De nombreuses méthodes ont été, depuis plusieurs décennies, proposées pour analyser les documents et extraire les objets relatifs ‘a la composition des pages. Les objectifs de cette analyse ont également été différents et les objets extraits en conséquence (Blocs ou lignes de

texte, mots ou caract‘eres, analyse des objets entourant le texte tels que logos,tableaux ou figures).

1.4.1 Approches ascendantes :

Elles se basent sur l'analyse des composantes connexes. commencent par le niveau le plus bas et remontent d'un niveau à un autre jusqu'à compléter la page, les composants sont obtenues en scannant une image pixel par pixel et en regroupant les pixels en des composants en se basant sur la connexité des pixels (4 ou en 8 voisins). Principe :

- Fusionner du plus bas niveau, en formant les mots à partir des composantes connexes

- Remontent à un niveau supérieur en fusionnant les mots en lignes, les lignes en blocs, etc... jusqu'à ce que la page soit complètement reconstituée.[3]

Les techniques ascendantes de détection de lignes de texte groupent de petits éléments pour former les lignes de texte. Elles se décomposent en deux familles :

- Famille de filtrages :

Les approches ascendantes appliquées ‘a la segmentation des lignes de texte dans des documents peuvent commencer avec les techniques de morphologie mathématique, ces techniques sont également appelées smearing. L'algorithme RLSA (Run-Length Smoothing Algorithm) en est un bon exemple puisqu'il remplit horizontalement les espaces entre pixels noirs proches.

Les objets détectes sont relatifs aux composantes connexes obtenues. Ces objets peuvent être (des mots, des lignes , des paragraphes) de texte en fonction du filtre choisi pour la dilatation, cela revient ‘a dire que l'on groupe les pixels noirs suffisamment proches les uns des autres. D'autres filtres que le filtre horizontal ont été utilisés pour des taches différentes.

Des techniques similaires ont utilisé des successions d'opérations morphologiques de type ouverture et fermeture.

- Famille de Composantes connexes :

Se basent sur les composantes connexes comme élément de base et groupent ces composantes connexes pour former les lignes de texte, cela peut être fait à l'aide d'heuristiques basées sur la position relative des composantes connexes, leurs tailles et surfaces respectives ou sur les directions données par leur lignes de base permettent de travailler avec des images historiques fortement endommagées en utilisant directement des images en niveaux de gris.

Les composantes connexes sont obtenues à différents seuils et sont conservées si elles ont une forme correcte avant d'être groupées.

1.4.2 Approches descendantes :

Commencent par le niveau le plus élevé à savoir la page et descend d'un niveau à un autre jusqu'à arriver au niveau des composantes connexes ou au niveau pixel .
Elles requièrent généralement des connaissances a priori plus ou moins précises sur la structure des documents à traiter.[4]
Les approches descendantes, quant à elles, prennent comme ´élément de base la page entière et divisent celle-ci progressivement, elles voient leur apparition avec des méthodes de type X-Y cut pour lesquelles on cherche des vallées sans écritures, alternativement de manière horizontale et verticale, pour séparer les objets, avec une approche analogue, cherche dans l'image les rectangles blancs de taille maximale. ces blocs, en fonction de leurs formes et de leurs tailles, définissent des inter paragraphes, des inter-lignes, des inter-mots ou des inter-caractères. Les méthodes à base de projections de profils , pour les quelles un histogramme de la présence des pixels est construit horizontalement permettent de s'adapter à d'éventuels bruits de bancarisation ou à des caractères se touchant.
Ces techniques ont en commun de chercher un chemin entre les lignes et sont donc adaptées à de la segmentation de paragraphes ou de pages avec une seule colonne. Cependant, elles rencontrent des problèmes pour la segmentation pleine page de documents avec des mises en page plus complexes.[4]

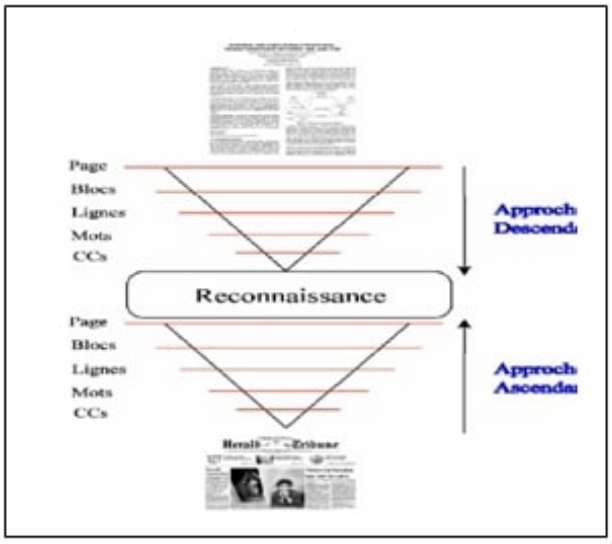


FIGURE 1.11 – Approche descendante et ascendante

[4]

1.5 L'apprentissage automatique :

1.5.1 définition :

L'apprentissage automatique (ou machine learning) peut être définie :

• Est un processus qui crée un modèle à partir d'un jeu de données d'entraînement, dans le but, par exemple, de classifier,mesurer ou prendre des décisions sur de nouvelles données.

• L'apprentissage ou entraînement, est une partie importante du système de vision par ordinateur. Le classificateur étant généralement une fonction paramétrique, l'apprentissage va permettre d'optimiser les paramètres du classificateur pour le problème à résoudre, en utilisant des données d'entraînement. [4]

Lorsque les données d'entraînement sont préalablement classées, l'apprentissage est dit :

Supervisé :qui consiste à analyser un jeu de données empiriques, appelée base de données d'entraînement. Ces données sont étiquetées ou labellisées, c'est à dire associées à une description ou information, afin de pou-voir caractériser un nouveau jeu de données inconnu. L'algorithme apprend ainsi à partir de milliers ou de millions d'exemples étiquetés : il cherche la relation qui permet de relier les données aux labels.[5]

Sinon il est non supervisé :où l'algorithme doit opérer en l'absence d'annotations.

• Est un sous-domaine de l'informatique qui s'intéresse à la construction d'algorithmes qui, pour être utile, s'appuie sur un ensemble d'exemples de certains phénomènes. Ces exemples peuvent provenir de la nature, être fabriqués à la main par l'homme ou générés par un autre algorithme.

• Le processus de résolution d'un problème pratique par la collecte d'un ensemble de données, et la construction algorithmique d'un modèle statistique basé sur cet ensemble de données. Ce modèle statistique est censé être utilisé d'une manière ou d'une autre pour résoudre le problème pratique.

• Une définition un peu plus générale :L'apprentissage automatique est le domaine d'étude qui donne aux ordinateurs la possibilité d'apprendre sans être explicitement programmés. Et une autre plus technique : on dit qu'un programme informatique apprend de l'expérience E en ce qui concerne une tâche T et une mesure de performance P, si sa performance sur T, mesurée par P, s'améliore avec l'expérience . [6]

1.5.2 Approches d'apprentissage automatique :

Les algorithmes qui utilisent des équations statistiques et mathématiques pour dériver les relations dans les données de formation relèvent de cette catégorie. Ces algorithmes sont également appelés algorithmes d'apprentissage automatique statistique. Il a l'avantage de la capacité d'explication (la capacité d'expliquer la raison de certaines prédictions pour l'entrée donnée). par exemple, K-means, arbres de décision, forêt aléatoire, machine à vecteur de support (SVM), etc.[4] La méthodologie de l'apprentissage par l'expérience est de plus en plus appliquée à des problèmes où une modélisation mathématique appropriée est impossible. Par exemple, on ne sait pas mathématiquement modéliser la relation entre un bloc de pixels pour distinguer le texte et les arrière-plans. Nous allons introduire deux approches d'apprentissage supervisé qui sont introduits(les perceptrons multicouches et les machines à vecteurs de support).

- **Perceptrons multicouches :**

Perceptron est la première et la plus simple forme de réseau d'anticipation proposée par Rosenblatt en 1962, un perceptron est composé d'une couche d'entrée et d'une couche de sortie de neurones, dans lesquelles un le nœud de sortie calcule la somme pondérée des nœuds d'entrée. Un modèle de perceptrons multicouches (MLP) est une extension multicouche du modèle perceptron, qui se compose d'une couche d'entrée, d'une sortie couche, et une ou plusieurs couches cachées de neurones.[7]

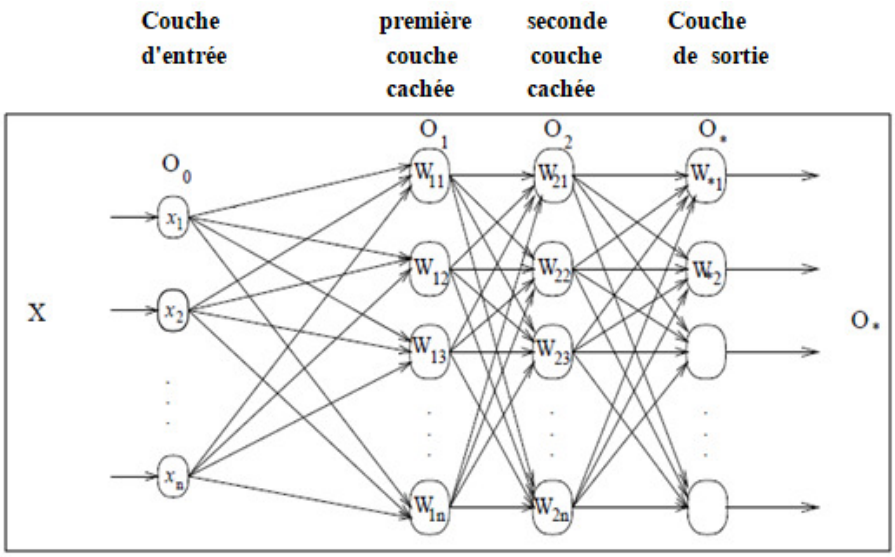


FIGURE 1.12 – Réseau perceptron multicouche avec deux couches cachées [7]

Les fonctions du MLP consistent en une opération de classification et une opération de formation. ou un vecteur d'entrée est présenté dans la couche d'entrée et chaque neurone calcule

une somme pondérée des sorties des neurones de la couche précédente, suivie d'une activité fonction jusqu'à ce qu'un ensemble de sorties soit finalement obtenu au niveau de la couche de sortie,puisqu'un seul neurone a une valeur de sortie à la fois, nous pouvons simplement écrire la sortie d'une couche sous forme de vecteur. L'un des résultats théoriques importants sur MLP est qu'une configuration avec un seul couche de neurones s'est avérée capable d'approximer n'importe quelle fonction continue. Dans le pratique, plusieurs perceptrons ont été appliqués dans de nombreuses applications telles que la reconnaissance vocale reconnaissance optique de caractères , détection de visage.[7]

• Soutenir la machine vectorielle :

Support Vector Machine (SVM) est une technique motivée par la théorie de l'apprentissage statistique et a été appliqué avec succès à de nombreuses tâches de classification. L'idée clé est de transformer l'entrée données dans un espace de caractéristiques de grande dimension et des classes séparées avec une surface de décision dans cet espace. Contrairement à l'algorithme empirique de minimisation du risque tel que MLP, qui minimise l'erreur sur l'ensemble de données, SVM est un algorithme de minimisation du risque structurel, qui vise à minimiser un borne sur l'erreur de généralisation d'un modèle dans un espace de grande dimension.

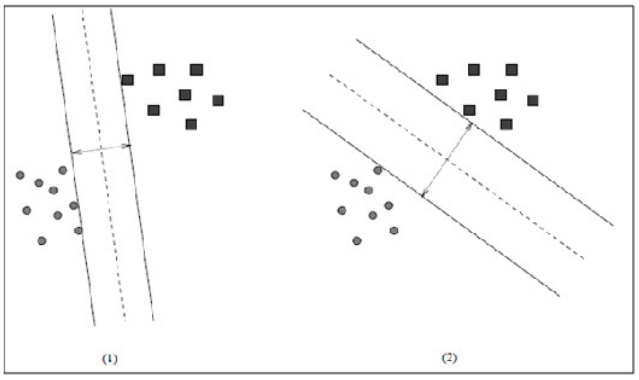


FIGURE 1.13 – Résolution de problèmes de classification avec des hyperplans.(1) petite marge(2)plus grande marge

[7]

• Marge douce :

- Cas linéaire non séparable :

Considérant que l'on cherche toujours une surface séparatrice linéaire, mais qu'un tel hyperplan séparateur n'existe pas, nous introduisons un hyperplan à marge souple pour résoudre le problème d'apprentissage,la marge souple a un ensemble de variables qui sont des variables d'écart positives et mesurent les pénalités proportionnel au nombre de violations de contraintes. La

tâche d'apprentissage implique maintenant la minimisation.

- Surfaces de décision non linéaires :

Cette méthode peut être facilement généralisée au cas non linéaire,les surfaces de décision les plus complexes peut être construit en cartographiant les exemples de formation xi dans une autre dimension supérieure espace ø(xi) , appelé espace des caractéristiques, et en travaillant avec une classification linéaire dans cet espace. Constatant que la formation de SVM ne dépend que d'exemples via des produits internes, cette cartographie peut être donnée implicitement en choisissant un noyau, K(xi, xj) = ø(xi). ø(xj) La tâche d'apprentissage implique maintenant la maximisation.

1.6 reconnaissance de texte :

La reconnaissance de texte est l'une des applications les plus populaires de l'analyse d'image de documents. Elle a connu ces dernières années de grands progrès, et les succès des travaux de recherches ont donné lieu à de nombreuses applications industrielles, dans plusieurs domaines, citons par exemple la lecture automatique de formulaires, de chèques ou d'adresses postales.
La reconnaissance de l'écriture suppose une localisation préalable des entités textuelles qui peuvent êtres mots ou bien caractères , deux modes d'écritures existent (imprimée,manuscrite).

1.7 Les dommaines d'application :

1.7.1 Image/vidéo compréhension :

La présence des images dans l'environnement quotidien n'est plus à démontrer et, contrairement à une opinion trés répandue, regarder et saisir le sens de ce qu'on voit n'est une activité ni naturelle, ni évidente pour les apprenants en langue : l'image n'est jamais immédiatement décodable, elle doit être objet d'étude. Il est important d'aider les apprenants étrangers á « décoder », à comprendre, la vidé o est le moyen de susciter chez l'apprenant des réactions.
Ils focaliseront leur attention sur un support encore relativement peu usité, et bien plus attractif ,toute vidéo peut être utilisée, étant entendu que ce n'est pas tant le degré de difficulté linguistique du document qui compte que la complexité de la tâche que l'on demande à l'apprenant lors du visionnement de la séquence.[8]

1.7.2 Compréhension de scéne :

La compréhension de scénes vise à répondre à la question : comment construire un modéle d'une région du monde réel afin d'y agir et d'y interagir ? Il s'agit donc d'extraire la sémantique et la géométrie des données disponibles : images, nuages de points 3D, etc.
Dans ce but, plusieurs approches d'apprentissage machine sont présentées : elles diffèrent par la proportion d'a priori de conception et d'apprentissage introduits tout au long des algorithmes,les premiers travaux visent à la compréhension du contenu sémantique des images (la classification, la détection d'objets et la segmentation sémantique). Puis, plusieurs approches d'apprentissage sont proposés pour l'observation de la Terre et la télédétection, notamment pour l'apprentissage interactif, la classification sémantique multimodale et la détection de changements sémantiques. Enfin, l'accent est mis sur la vision 3D, avec l'estimation de la profondeur à partir d'une seule image et la classification de nuages de points 3D par des réseaux de neurones, ces approches variées reposent sur des mécanismes sous-jacents communs qui prennent une importance croissante. Elles réalisent une analyse multimodale pour bénéficier des données complémentaires disponibles, issues de capteurs différents mais aussi de sources et métadonnées hétérogènes. Symétriquement, l'optimisation jointe d'objectifs multiples permet de régulariser l'apprentissage de modèles performants. Surtout, elles ont de plus en plus recours à une multiplicité des points de vue sur la scène pour relier, tant en apprentissage qu'en inférence, des invariances spatiales qui servent une analyse locale et une reconstruction sémantique globale. Cela est rendu possible par une intégration croissante de l'apparence et de la structure 3D, et conduit à une meilleure compréhension sémantique de la scène.[9]

1.7.3 Recherche visuelle :

Permet désormais de retrouver et d'identifier ce que vous voyez ! Il vous suffit de fournir un visuel et l'algorithme se charge du reste. Très intuitive d'utilisation, la recherche visuelle tend à se démocratiser et intègre peu à peu nos expériences de recherche, trouver des informations sur un objet ou encore un lieu ne prend plus que quelques secondes. La « Visual Search » se base sur des fonctionnements complexes et évolutifs tels que l'IA et notamment le Deep Learning. La recherche visuelle consiste quant à elle à rechercher une « information » via un visuel. C'est en prenant une photo ou en fournissant une capture d'écran que la recherche sera effectuée et les résultats fournis. De plus, les deux technologies ne se basent pas entièrement sur le même fonctionnement. La recherche visuelle fonctionne davantage sur le principe d'intelligence artificielle qui va effectuer un travail complexe et minutieux de reconnaissance, d'association et

de « tri » d'informations.[10]

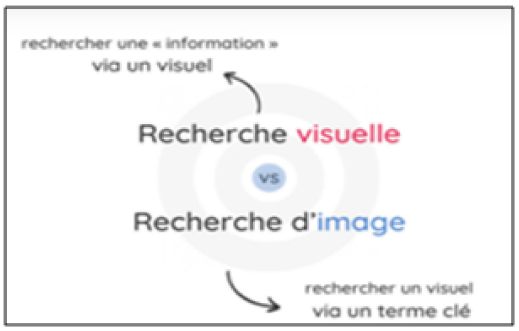


FIGURE 1.14 – recherche visuelle.

[10]

1.7.4 Interaction homme-machine, auxiliaire aveugle :

L'Interaction Homme-Machine est la discipline consacrée a la conception, la mise en œuvre et a l'évaluation de systèmes informatiques interactifs destines à des utilisateurs humains ainsi qu'a l' étude des principaux phénomènes qui les entourent , un système interactif est un système dont le fonctionnement dépend d'informations fournies par un environnement externe qu'il ne contrôle pas les systèmes interactifs sont également appelés ouverts, par opposition aux systèmes fermés dont le fonctionnement peut être entièrement décrit par des algorithmes[11]

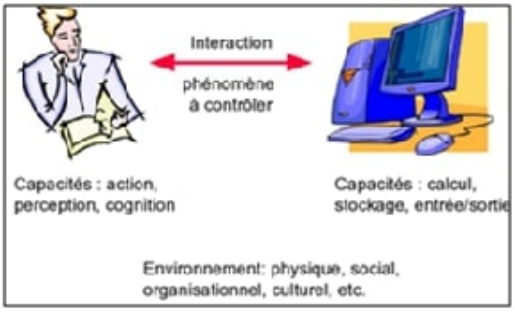


FIGURE 1.15 – Interaction homme-machine

[11]

1.7.5 Conduite automatique (assistance automobile) et récupération d'images :

Le système d'aide à la conduite (automobile) ou système avancé d'aide à la conduite est un système de sécurité active d'information ou d'assistance du conducteur pour :

- Eviter l'apparition d'une situation dangereuse risquant d'aboutir à un accident .

- Libérer le conducteur d'un certain nombre de tâches qui pourraient atténuer sa vigilance.

- Assister le conducteur dans sa perception de l'environnement (détecteurs de dépassement, de risque de gel, de piéton,.. etc.)

- Permettre au véhicule de percevoir le risque et de réagir de manière anticipée par rapport aux réflexes du conducteur.[12]

FIGURE 1.16 – Equipements et assistance des automobiles

[12]

1.8 Les domaines où l'OCR est le plus utilisé :

Le monde est devenu un village mondial grâce à la numérisation rapide, mais il a également ouvert la porte à de nombreux fraudeurs pour qu'ils interviennent et terrifient les gens.
Les organisations de tous les secteurs ne sont pas en sécurité en raison de l'augmentation des ransomwares et des violations de données. Compte tenu du nombre croissant de fraudes, les entreprises optent pour des systèmes de vérification robustes avec la technologie OCR pour n'intégrer que les clients légitimes. Ces systèmes permettent aux entreprises de filtrer les fraudeurs avant de devenir un problème pour les clients et l'entreprise.

1.8.1 Domaine bancaire :

Le secteur bancaire fonctionne comme une mine d'or pour les fraudeurs et fait face à d'énormes pertes en raison du blanchiment d'argent, du vol d'identité et de plusieurs autres fraudes les gouvernements de différents États ont également imposé des réglementations strictes en matière de connaissance de vos clients et de lutte contre le blanchiment d'argent Se conformer à ces réglementations est un défi sans un système de vérification solide. Pour la même raison, les organisations ajoutent la technologie OCR pour une extraction efficace des données. Se conformer au fardeau réglementaire sans cesse croissant et à l'intégration transparente des clients devient désormais plus simple pour les banques.[13]

L'utilisation la plus courante de l'OCR est la saine gestion des chèques :

- Le chèque manuscrit est numérisé.
- Ses détails sont transformés en texte numérique.
- La signature est validée.
- Le chèque est approuvé en temps réel.

Aujourd'hui, seule la vérification de la signature nécessite la validation avec une valeur résidente dans une base de données préexistante.[14]
Une diminution du temps de traitement des chèques est un avantage financier pour tout le monde : le débiteur, la banque et le créditeur.

FIGURE 1.17 – Échantillon de cas d'utilisation bancaire de l'OCR
[14]

1.8.2 Monde légal :

La numérisation, le stockage, la conservation en base de données accessible à la recherche sont désormais possibles pour tous les documents imprimés : affidavits, jugements, déclarations, avis, testaments,..etc. L'OCR est également disponible pour des documents en chinois, en arabe et en orthographes pour les langues ayant une autre écriture que celles de type « romaine ». L'accès rapide aux documents juridiques provenant de millions de cas antérieurs est certainement un avantage pour une industrie qui s'appuie fortement sur le passé.

1.8.3 Santé :

Il est possible de numériser tout l'historique médical d'un patient : rapports de santé, radiographies, historique de maladies, suivi des traitements, diagnostics, dossiers hospitaliers, couverture d'assurance, paiements. Après numérisation, toutes ces informations sont disponibles et consultables en un seul endroit. Le fait que l'ensemble du dossier patient soit stocké numériquement représente un avantage majeur pour l'épidémiologie et pour la logistique (maintien des niveaux de médicaments en pharmacies, équipements et autres produits de santé, etc.)
Une fois numérisés, tous les dossiers forment une énorme base de données qui peut être utile

d'étudier dans son ensemble pour fournir des insights aux législateurs et aux réseaux de santé partout dans le monde.[13]

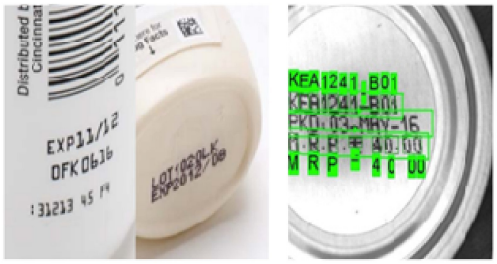


FIGURE 1.18 – OCR dans les produits pharmaceutiques

[13]

1.8.4 Chaîne d'approvisionnement :

Certains articles doivent être localisés dans la chaîne d'approvisionnement à tout moment, et fournir une documentation claire de leur origine et de leur emplacement. Bien que le suivi des produits soit souvent géré grâce aux code-barres ou aux puces de type « Near Field Communication (NFC) », l'OCR a malgré tout une utilité. Il permet de lire les instantanément codes des lots, les dates d'expiration et les numéros de série. Ces informations améliorent le suivi d'un produit à toutes les étapes du cycle d'emballage, de l'étiquetage à la mise du produit final sur les tablettes. L'OCR peut être également utile pour comparer le texte actuel avec la chaîne prévue définie dans la base de données, et signaler un numéro de série hors séquence ou manquant. Les code-barres et l'OCR sont souvent utilisés de pairs pour maximiser l'exactitude de la collecte d'informations.[13]

FIGURE 1.19 – Le code-barre

[13]

1.8.5 Assurances :

Le secteur de l'assurance est basé sur des documents, mais peu d'assureurs ont automatisé le traitement des documents importants, nécessaires pour les devis, les souscriptions, l'onboarding, le règlement des sinistres et la conformité. Cela ralentit les processus de cotation, de souscription et de gestion des sinistres et, au final, cela a un impact négatif sur l'expérience client. Comprendre le flux d'informations dans vos processus – et où les retards et interruptions se situent – est la première étape d'une automatisation intelligente dans l'assurance.[14]

1.8.6 L'expertise-comptable :

Il est désormais possible d'automatiser une grande partie de sa comptabilité : gestion et récupération des factures reçues, classement automatique, imputation comptable, rapprochement bancaire.[15]

1.8.7 Le domaine de la documentation :

Ou toutes les branches qui touchent au caractére manuscrit.
d'autres cas d'utilisation peuvent être :

• **Saisie automatique de données** pour des documents d'entreprise, par exemple : formulaires papier, factures, reçus, etc .

• **Reconnaissance automatique des plaques d'immatriculation**.

• **Reconnaissance des passeports** de voyageurs dans un aéroport et l'extraction de l'information importante.

• **Extraction automatique** d'informations clés dans des documents d'assurance.

• **Extraction des informations de carte d'affaires .**

• **Numérisation de gros documents imprimés**, par exemple des livres .[16]

1.9 Conclusion :

Dans ce chapitre nous avons bien défini le problème de détection de texte dans les scènes naturelle, et les différents techniques traditionnelles utilisées qui restent limitées et ne peuvent pas résoudre le problème qui présente plusieurs challenges. Le chapitre suivant, définit la technologie deep learning et étudie les solutions apportées par rapport aux techniques traditionneles.

CHAPITRE 2

LES APPROCHES BASEE SUR L'APPRENTISSAGE EN PROFONDEUR(DEEP LEARNING) :

2.1 Introduction

Un sous-ensemble de l'apprentissage automatique. C'est un domaine basé sur l'apprentissage et l'amélioration par lui-même en examinant les algorithmes informatiques. le deep learning fonctionne avec des réseaux de neurones artificiels, conçus pour imiter la façon dont les humains pensent et apprennent.

2.2 La définition de Deep Learning :

Jusqu'à récemment, les réseaux de neurones étaient limités par la puissance de calcul et étaient donc limités en complexité. Cependant, les progrès de l'analyse du Big Data ont permis des réseaux de neurones plus vastes et sophistiqués, permettant aux ordinateurs d'observer, d'apprendre et de réagir à des situations complexes plus rapidement que les humains. Deep learning a aidé à la classification des images, à la traduction de la langue et à la reconnaissance vocale. Il peut être utilisé pour résoudre tout problème de reconnaissance de formes et sans intervention humaine.[17]

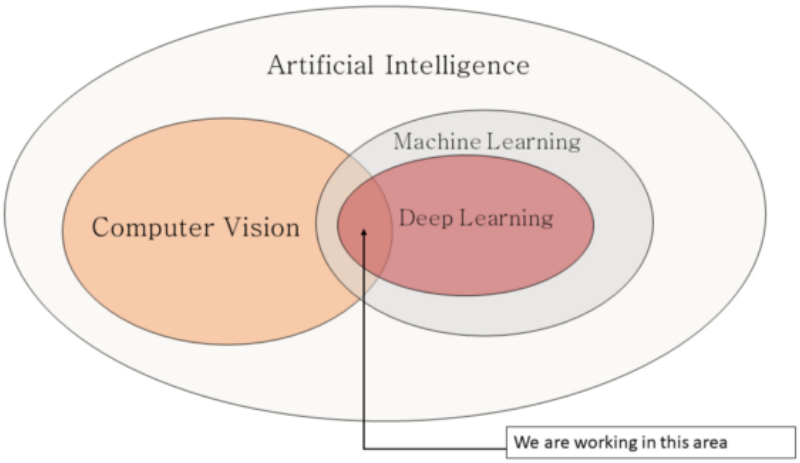
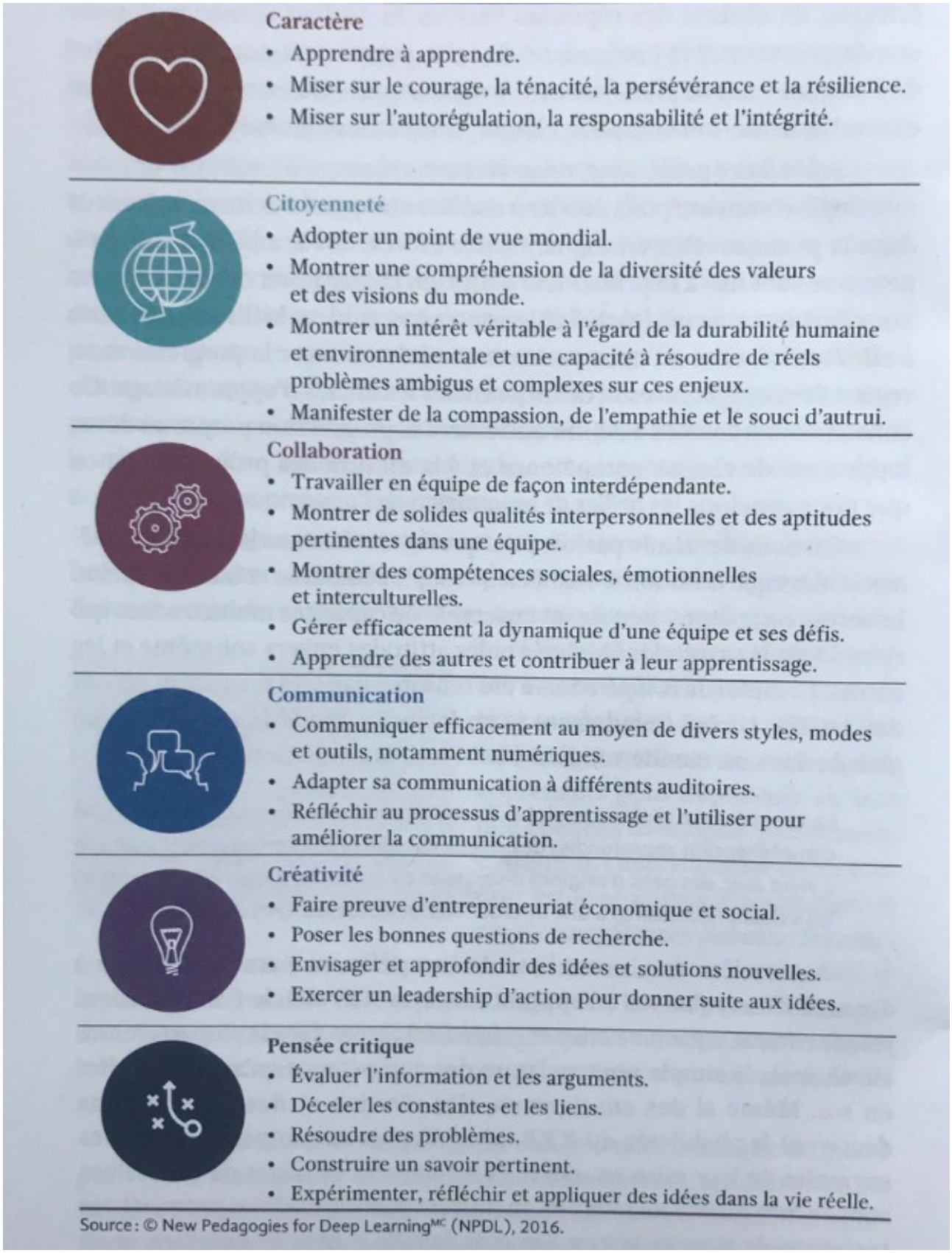


FIGURE 2.1 – Deep Learning

On peut le définis aussi que c'est le processus d'acquisition de six compétance globale le caractére- la sétoyenneté- la collaboration- la communication -la créativité et la pensé critique, présenté dans la figure 2.2.



Caractère

- Apprendre à apprendre.
- Miser sur le courage, la ténacité, la persévérance et la résilience.
- Miser sur l'autorégulation, la responsabilité et l'intégrité.

Citoyenneté

- Adopter un point de vue mondial.
- Montrer une compréhension de la diversité des valeurs et des visions du monde.
- Montrer un intérêt véritable à l'égard de la durabilité humaine et environnementale et une capacité à résoudre de réels problèmes ambigus et complexes sur ces enjeux.
- Manifester de la compassion, de l'empathie et le souci d'autrui.

Collaboration

- Travailler en équipe de façon interdépendante.
- Montrer de solides qualités interpersonnelles et des aptitudes pertinentes dans une équipe.
- Montrer des compétences sociales, émotionnelles et interculturelles.
- Gérer efficacement la dynamique d'une équipe et ses défis.
- Apprendre des autres et contribuer à leur apprentissage.

Communication

- Communiquer efficacement au moyen de divers styles, modes et outils, notamment numériques.
- Adapter sa communication à différents auditoires.
- Réfléchir au processus d'apprentissage et l'utiliser pour améliorer la communication.

Créativité

- Faire preuve d'entrepreneuriat économique et social.
- Poser les bonnes questions de recherche.
- Envisager et approfondir des idées et solutions nouvelles.
- Exercer un leadership d'action pour donner suite aux idées.

Pensée critique

- Évaluer l'information et les arguments.
- Déceler les constantes et les liens.
- Résoudre des problèmes.
- Construire un savoir pertinent.
- Expérimenter, réfléchir et appliquer des idées dans la vie réelle.

Source : © New Pedagogies for Deep Learning^{MC} (NPDL), 2016.

FIGURE 2.2 – Les six compétances globales de Deep Learning

2.3 L'importance de Deep Learning :

-L'apprentissage automatique fonctionne uniquement avec des ensembles de données structurées et semi-structurées, tandis que l'apprentissage en profondeur fonctionne avec des données structurées et non structurées.
- Les algorithmes d'apprentissage en profondeur peuvent effectuer efficacement des opérations complexes, tandis que les algorithmes d'apprentissage automatique ne le peuvent pas.
- Les algorithmes d'apprentissage automatique utilisent des échantillons de données étiquetés pour extraire des modèles, tandis que l'apprentissage en profondeur accepte de gros volumes de données en entrée et analyse les données d'entrée pour extraire les caractéristiques d'un objet.
-Les performances des algorithmes d'apprentissage automatique diminuent à mesure que le nombre de données augmente , donc pour maintenir les performances du modèle, nous avons besoin d'un apprentissage en profondeur.

2.4 Le fonctionnement de Deep Learning :

Les réseaux de neurones sont des couches de nœuds, tout comme le cerveau humain est composé de neurones, les nœuds au sein des couches individuelles sont connectés aux couches adjacentes. On dit que le réseau est plus profond en fonction du nombre de couches dont il dispose, un seul neurone du cerveau humain reçoit des milliers de signaux d'autres neurones. Dans un réseau neuronal artificiel, les signaux voyagent entre les nœuds et attribuent des poids correspondants, un nœud pondéré plus lourd exercera plus d'effet sur la couche de nœuds suivante. La couche finale compile les entrées pondérées pour produire une sortie.
Les systèmes d'apprentissage en profondeur nécessitent un matériel puissant car ils traitent une grande quantité de données et impliquent plusieurs calculs mathématiques complexes, cependant, même avec un matériel aussi avancé, les calculs de formation en apprentissage profond peuvent prendre des semaines.
Les systèmes d'apprentissage en profondeur nécessitent de grandes quantités de données pour renvoyer des résultats précis , en conséquence, les informations sont alimentées sous la forme d'énormes ensembles de données. Lors du traitement des données, les réseaux de neurones artificiels sont capables de classer les données avec les réponses reçues à partir d'une série de questions binaires vraies ou fausses impliquant des calculs mathématiques très complexes. Par exemple, un programme de reconnaissance faciale fonctionne en apprenant à détecter et à re-

connaître les arêtes et les lignes des visages, puis les parties les plus significatives des visages et, enfin, les représentations globales des visages. Au fil du temps, le programme s'entraîne et la probabilité de réponses correctes augmente. Dans ce cas, le programme de reconnaissance faciale identifiera avec précision les visages avec le temps.[17]

2.5 Applications du deep Learning :

Le deep Learning est utilisé dans de nombreux domaines :

- Reconnaissance d'image.
- Traduction automatique
- Voiture autonome
- Diagnostic médical
- Recommandations personnalisées
- Modération automatique des réseaux sociaux
- Prédiction financière et trading automatisé
- Identification de pièces défectueuses
- Détection de malwares ou de fraudes
- Chatbots (agents conversationnels)
- Exploration spatiale
- Robots intelligents [18]

2.6 Types de modèles utilisant des architectures Deep Learning) :

L'apprentissage en profondeur offre une grande précision dans les ensembles de données encombrés. Les algorithmes d'apprentissage en profondeur préfèrent dans les applications de prédiction, de classification et d'identification.
L'algorithme d'apprentissage en profondeur a été utilisé avec plusieure nombre de réseaux tellque :

2.6.1 Réseaux neuronaux convolutifs (CNN) :

Appelés ConvNets, les CNN sont constitués d'une multitude de couches chargées de traiter et d'extraire les caractéristiques des données, de manière spécifique, les réseaux neuronaux convolutifs sont utilisés pour l'analyse et la détection d'objets. Ils peuvent donc servir par exemple à reconnaître des images satellites, traiter des images médicales, détecter des anomalies ou prédire des séries chronologiques.

2.6.2 Réseaux neuronaux récurrents (RNN) :

Possèdent des connexions qui constituent des cycles dirigés. Cela permet aux sorties du LSTM d'être exploitées comme entrées au niveau de la phase actuelle. La sortie du LSTM se transforme en une entrée pour la phase actuelle. Elle peut donc mémoriser les entrées précédentes à l'aide de sa mémoire interne. Dans la pratique, les RNN sont utilisés pour le sous-titrage d'images, le traitement du langage naturel et la traduction automatique.

2.6.3 Réseaux de mémoire à long et court terme (LSTM) :

Les LSTM sont des dérivés de RNN. Ils peuvent apprendre et mémoriser des dépendances sur une longue durée. Les LSTM conservent ainsi les informations mémorisées sur le long terme. Ils sont particulièrement utiles pour prédire des séries chronologiques, car ils se rappellent des entrées précédentes. Outre ce cas d'utilisation, les LSTM sont également utilisés pour composer des notes de musique et reconnaître des voix.

2.6.4 Réseaux de fonction de base radiale (RBFN) :

Ils exploitent des fonctions de base radiales en tant que fonctions d'activation. Ils sont constitués d'une couche d'entrée, d'une couche cachée et d'une couche de sortie. Généralement, les RBFN sont utilisés dans la classification, la prédiction des séries temporelles et la régression linéaire.

2.6.5 Réseaux adversariaux génératifs (GAN) :

Les GAN créent de nouvelles instances de données qui s'apparentent aux données d'apprentissage profond. Ils possèdent deux principaux composants : un générateur et un discriminateur. Si le générateur apprend à produire des informations erronées, le discriminateur, quant à lui, ap-

prend à exploiter ces fausses informations. Les GAN sont généralement utilisés par les créateurs de jeux vidéo pour améliorer les textures 2D.

2.6.6 Machines de Boltzmann restreintes (RBM) :

Sont des réseaux neuronaux stochastiques constitués de deux couches : unités visibles et unités cachées. Ces réseaux artificiels sont capables d'apprendre en partant d'une distribution de probabilité sur un ensemble d'entrées. Néanmoins.[19]

2.7 Définition de réseaux de neurones convolutionnels (CNN) :

- Est un type de réseau neuronal artificiel utilisé dans la reconnaissance et le traitement d'images et spécifiquement conçu pour traiter les données de pixels.
Il sont de puissants systèmes de traitement d'images, d'intelligence artificielle (IA) qui utilisent un apprentissage approfondi (deep learning) pour effectuer des tâches à la fois génératives et descriptives, souvent à l'aide de Machine Vision qui inclut la reconnaissance d'images et de vidéos, ainsi que des systèmes de recommandation et le traitement du langage naturel.[20].
- Est un type de réseaux de neurones artificiels acycliques dans lequel le motif de connexion entre les neurones est inspiré par le cortex visuel des animaux, très utilisés dans le domaine de la vision par ordinateur. et ont aussi beaucoup de succès dans les reconnaissances faciales, la détection d'objets très utilisée dans les robots et les voitures automatiques. En gros, tout ce qui concerne la vision par ordinateur et les images.
De plus on peut utiliser les CNN dans tous les problèmes ayant en entrée une matrice. Par exemple, Gehring a utilisé une matrice de texte dans une tache de traduction automatique de langue.
On note que le terme "convolutional" vient de l'opération de convolution de matrices utilisée dans le traitement de signal.
Deux nouveaux types de couche ont été ajoutés dans le réseau : la couche convolution (convolutional layer) et la couche de mise en commun (pool layer).[21]

2.8 Les couches CNN :

Un CNN utilise un système semblable à un perceptron multicouche qui a été conçu pour des besoins de traitement réduits.

Les couches d'un CNN se composent d'une couche d'entrée, d'une couche de sortie et d'une couche cachée qui comprend plusieurs couches convolutionnelles, des couches de regroupement, des couches entièrement connectées et des couches de normalisation. La suppression des limitations et l'augmentation de l'efficacité pour le traitement des images aboutissent à un système beaucoup plus efficace, plus simple à former, et spécialisé pour le traitement des images et le traitement du langage nature.

Les perceptrons multicouches « **Multi Layer Perceptron** » est une technique d'apprentissage pour la classification de données, capables d'approximer des fonctions non–linéaires complexes afin de traiter des données de grande dimension.

les approches de classification d'images sont :

• Extraire des caractéristiques directement des données par un algorithme choisi par l'utilisateur.

• Présenter l'image en entrée d'un réseau de neurones ou l'image vectorisée,(la dimension est égale au nombre de pixels de l'image).

Parmie les défauts des MLP :

• La valeur d'un neurone d'une couche n va dépendre des valeurs de tous les neurones de la couche (n – 1) (connéction complète et peut être très grand).

• Une application à des images est qu'ils sont peu ou pas invariants à des transformations de l'entrée, ce qui arrive très souvent avec des images (légères translations, rotations ou distorsions).

• La reconnaissance de formes exige le prennent en compte la corrélation entre pixels d'une image ce qui né pas réalisée.

Une extension des MLP nommée **CNN « Convolutional Neural Network »** permettant de répondre efficacement aux principaux défauts des MLP. Ils sont conçus pour extraire automatiquement les caractéristiques des images d'entrée, sont invariants à de légères distorsions de l'image.

2.8.1 Les types de couches CNN :

Une architecture CNN est formée par un empilement de couches de traitement indépendantes :

•Couche de convolution(CONV) :

Traite les données d'un champ récepteur.c'est le bloc de construction de trois paramètres permettent de dimensionner le volume de la couche de convolution :

-Profondeur de la couche : 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 ' :la dimension spatiale du volume de sortie. En particulier, il est parfois souhaitable de conserver la même surface que celle du volume d'entrée.

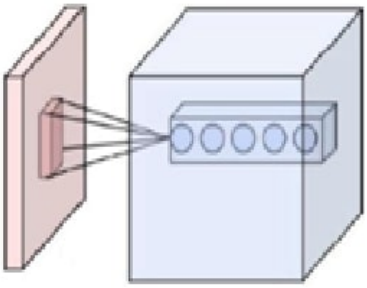


FIGURE 2.3 – Ensemble de convolution (bleu). Ils sont liés à un même champ récepteur (rouge)

• Couche de pooling (POOL) :

Permet de compresser l'information en réduisant la taille de l'image intermédiaire (souvent par sous-échantillonnage).C'est est une forme de sous-échantillonnage de l'image. **-**L'image d'entrée est découpée en une série de rectangles de n pixels de côté ne se chevauchant pas (pooling).

-Chaque rectangle peut être vu comme une tuile.

-Le signal en sortie de tuile est défini en fonction des valeurs prise par les différents pixels de la tuile.

-Le pooling réduit la taille spatiale d'une image intermédiaire, la quantité de paramètres et de calcul dans le réseau. pour contrôler l'overfitting Il est donc fréquent d'insérer périodiquement une couche de pooling entre deux couches convolutées successives d'une architecture CNN (sur aprentissage). La couche de pooling fonctionne indépendamment sur chaque tranche de profondeur de l'entrée et la redimensionne uniquement au niveau de la surface. La forme la plus courante est une couche de mise en commun. avec des tuiles de taille 2x2 (largeur/hauteur) et comme valeur de sortie la valeur maximale en entrée. On parle dans ce cas de Max-Pool 2x2.Il est possible d'utiliser d'autres fonctions de pooling que le maximum.

valeurs du patch d'entrée), du « L2-norm pooling».

Il est aussi possible d'éviter la couche de pooling mais cela implique un risque sur-apprentissage

plus important.

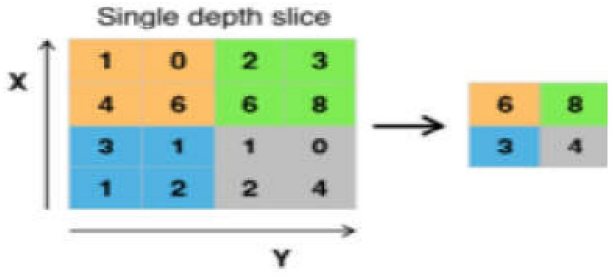


FIGURE 2.4 – Pooling avec un filtre 2x2 et un pas de 2

[22]

•Couches de correction (RELU) :

Souvent appelée par abus 'ReLU'en référence à la fonction d'activation (Unité de rectification linéaire).

Il est possible d'améliorer l'efficacité du traitement en intercalant entre les couches de traitement une couche qui va opérer une fonction mathématique (fonction d'activation) sur les signaux de sortie. la fonction F(x) = max (0,x) force les neurones à retourner des valeurs positives.

•Couches entièrement connectée (FC) :

Qui est une couche de type perceptron. le raisonnement de haut niveau dans le réseau neuronal se fait via des couches entièrement connectées après plusieurs couches de convolution et de max-pooling. Ces couches ont des connexions vers toutes les sorties de la couche précédente. La calcule de leurs fonctions d'activations se fait avec une multiplication matricielle suivie d'un décalage de polarisation.[21]

2.9 Etat de l'art des méthodes :

Les catégories de classemment du sont :méthodes récentes de détection de texte de scéne sont :

2.9.1 Méthode basée sur la régression :

Sont une série de modéles qui régresser directement les boîtes englobantes des instances de texte et basées sur la régression apprécient généralement des algorithmes de post-traitement simples (exemple : suppression).

Cependant, la plupart d'entre eux sont limités pour représenter des boîtes englobantes précises

pour des formes irréguliéres, comme les formes courbes.

L'intérêt de ses méthodes est qu'elle vise à atténuer les inconvénients de la multicolinéarité importante qui se manifeste dans beaucoup de problémes de régression multiple.

Reference	Dataset	résumé
[23]	ICDAR 2015 COCO-Text images	ils codent d'abord l'image d'entrée dans une séquence de caractéristiques puis appliquent des décodeurs tels que RNN (Hochreiter et Schmidhuber 1997) et CTC(Graves et coll. 2006) pour décoder la séquence cible Ces méthode produire de bons résultats lorsque le texte l'image est horizontal[24]
[25]	ICDAR'11 ICDAR'13	modifié les ancres et les échelles des noyaux convolutifs basée sur SSD (Single Shot detector)
[26]	ICDAR 2013 COCO-Text	propose une décomposition dimensionnelle du réseau de proposition de région pour gérer le problème de détection de texte de scène

TABLE 2.1 –

2.9.2 Méthode basée sur la segmentation :

Des méthodes de segmentation de texte sont appliquées sur les régions de texte extraites pour supprimer l'arrière-plan caractères de texte environnants, supposent généralement que la distribution en niveaux de gris est bimodal et que les caractères correspondent a priori soit à la partie blanche soit à la partie noire. De gros efforts sont donc consacrés à réaliser une meilleure binarisation, combinant seuillage global et local ou simple lissage Pour éliminer le non-caractère [27]
Combinent généralement algorithmes de prédiction et de post-traitement pour obtenir la limite des boites.

Segmentation : cette étape regroupe les pixels de l'image qui sont voisins et ont une forte probabilité d'appartenir au même objet. Un ensemble de régions est alors obtenu.[28]
Chacune de ces régions est définie par un ensemble d'attributs, parmi lesquels des statistiques relatives aux réponses spectrales ou à la forme (superficie, élongation) de la région.[29]

Reference	Dataset	résumé
[30]	MSRA-TD500, ICDAR2013 ICDAR2017-RCTW , ICDAR2017-MLT	La bordure de texte est utilisée pour diviser les instances de texte
[31]	SynthText ICDAR2013	détecté des instances de texte de forme propose une échelle progressive expansion en segmentant les instances de texte avec différents noyau d'échelle.
[32]	SynthText CTW1500 MSRA-TD500 Total-Text ICDAR 2015	texte multi-orienté détecté par segmentation sémantique et algorithmes basés sur MSER[33].

TABLE 2.2 –

2.9.3 Méthode de détection rapide de texte de scéne :

Se concentrent à la fois sur la précision et la vitesse d'inférence,plus d'entre eux ne peuvent pas traiter des instances de texte de formes irrégulières,comme la forme incurvée. Par rapport

à la scéne rapide précédente détecteurs de texte, notre méthode fonctionne non seulement plus rapidement, mais peut également détecter les instances de texte de formes arbitraires.[28]

Reference	Dataset	résumé
[34]	SynthText ICDAR 2015 MSRA-TD500	décomposer le texte en deux éléments détectables localement à savoir les segments et les liens.
[35]	ICDAR 2015 MSRA-TD500 COCO-Text	a proposé de appliquer PVAN et(Kim et al. 2016) pour améliorer sa vitesse

TABLE 2.3 –

2.10 Conclusion

Les différents approches basées sur le deep learning donnent des améliorations par rapport les solutions traditionnelles, cependant pas de détection parfaite, c'est un champ de recherche dans lequel les chercheurs proposent toujours de nouveaux modèles et approches afin d'arriver à des meilleurs résultats.

Le chapitre suivant définie la segmentation de façon générale et plus pécisemment la segemntation sémantique.

CHAPITRE 3

METHODOLOGIE DE LA DETECTION DE TEXTE DANS LES SCENES NATURELLES

3.1 Itroduction :

La segmentation est une étape essentielle en traitement d'image et reste un problème complexe, généralement c'est la première étape de l'analyse d'image qui vient après le prétraitement. C'est une transformation très utile en vision artificielle. Son but est d'extraire les informations pertinentes en regard de l'application concernée.

3.2 Segmentation des images :

Il ya plusieurs définitions de la segmentation tels que :

• Un processus de la vision par ordinateur, elle consiste à regrouper les pixels de ces images qui partagent une même propriété pour former des régions connexes. [4]

• Le procédé qui conduit à un découpage de l'image en un nombre fini de régions (ou segments) bien définies qui correspondent à des objets, des parties d'objets ou des groupes d'objets qui apparaissent dans une image.[4]

• Le partitionnement d'une scène en ses composants et elle la tâche de trouver des groupes de pixels qui « s'associent ». Par exemple, nous segmentons une scène de rue en ses objets et parties, tels que les voitures, les bâtiments, les piétons, les trottoirs. [36]

• C'est la tâche consistant à partitionner une image en objets d'intérêt,soit en traçant des contours, soit en classifiant chaque pixel. La segmentation sémantique implique d'identifier les objets par leur nature au moyen d'une étiquette numérique, aussi appelée classe. En cas d'occurrences multiples d'un même type d'objet, l'étiquetage peut être adapté pour séparer les éléments distincts d'un même type (segmentation d'instances multiples).[37]

Une erreur dans la segmentation de la forme à reconnaître augmente forcément le risque d'une mauvaise reconnaissance. Essentiellement, l'analyse de l'image fait appel à la segmentation où l'on va tenter d'associer à chaque pixel de l'image un label en s'appuyant sur l'information portée (niveaux de gris ou couleur), sa distribution spatiale sur le support image, des modèles simples.

• Les régions correspondent, au contraire, à la projection de zones homogènes en niveau de gris ou en couleur des objets.

3.3 Types de Segmentation :

Segmenter une image consiste à extraire cette information. Lorsque l'on parle de segmentation en points d'intérêt, contours, régions, mouvement, couleur , on parle en général des algorithmes permettant d'extraire l'information à Niveau Intermédiaire : c'est-à-dire des attributs numériques. Ceux-ci seront les entrées des procédures de plus haut niveau de filtrage sur attributs permettant la reconnaissance de forme et l'identification d'objets recherchés.

3.3.1 Segmentation de texte en lignes :

En général, le sous-titres ou le texte graphique ce compose de plusieurs lignes, en reconnaissance de texte, le texte se segmentant en lignes séparées. Il existe certaines méthodes utilisées à cet effet, telles que la projection horizontale.

3.3.2 Segmentation de lignes en caractères :

Il s'agit ici de la segmentation de lignes en caractères individuels. Les points de segmentation sont identifiés à la fin d'un caractère et au début de la suivante.

3.4 Les principes de la segmentation :

De nombreux travaux ont été réalisés sur ce sujet, dans des domaines aussi variés que le domaine médical, militaire, industriel, géophysique, etc... c'est toujours un sujet d'actualité et un problème qui reste ouvert où l'on retrouve de très nombreuses approches visent à l'extraction des caractéristiques.[4]

• La segmentation par regions :

Les points d'intérêt et contours correspondent à la projection dans l'image des bords externes ou internes (dans le cadre d'un objet comportant plusieurs zones de niveau de gris ou de couleur) des objets.La primitive point d'intérêt est très utilisée dans de nombreuses applications, citons : le recalage d'images, la création de panoramas, la reconnaissance d'objets, le suivi de cibles et même la reconnaissance de caractères déformés (CAPTCHA) ou manuscrits.[38]

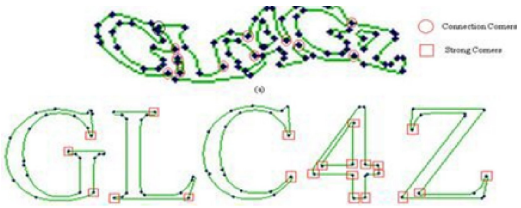


FIGURE 3.1 – Application du détecteur de coins de R.Al Nachar à la reconnaissance de caractères.

[38]

FIGURE 3.2 – segmentation par regions.

[38]

- **La segmentation par seuillage :**

Les méthodes de seuillage d'histogrammes sont souvent considérées comme des méthodes de segmentation d'images privilégiées, ces méthodes reposent sur l'exploitation de l'histogramme monodimensionnel de l'image qui caractérise la distribution des niveaux de gris. Cependant les performances de ces méthodes se dégradent rapidement lorsque les images à seuiller sont trop bruitées.[39]

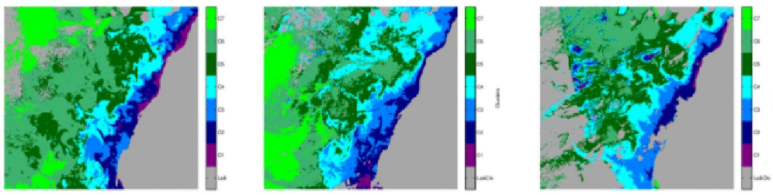


FIGURE 3.3 – segmentation par seuillage.

- **La segmentation par contours :**

Les contours correspondent aux bords externes et internes des objets. Ils sont filiformes c'est-à-dire d'épaisseur un pixel. En fonction des applications, plusieurs niveaux de « finition » peuvent être employés :

– **- Les points de contours :** par exemple pour colorier automatiquement l'intérieur des formes .les contours, c'est-à-dire une liste ordonnée de points de contour connexes.

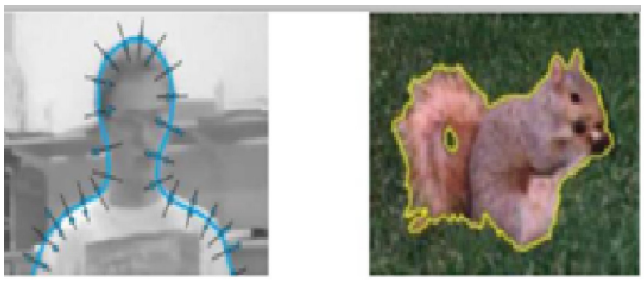


FIGURE 3.4 – segmentation par contour

[4]

– **- Une modélisation des contours :**

◦ Sous forme de segments de droites : on parlera d'approximation polygonale .

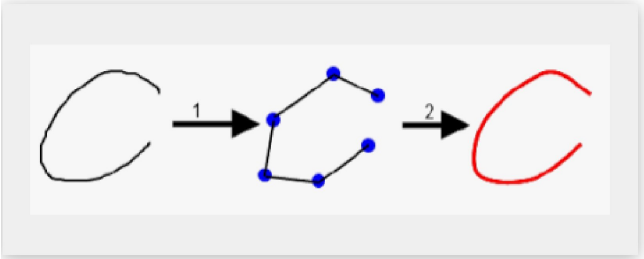


FIGURE 3.5 – segmentation par polygone

◦ Sous forme plus globale : de droites, de cercles, d'ellipses, par exemple en utilisant la Transformée de Hough :

Se base sur l'utilisation d'un espace paramétrique, appelé espace de Hough, permettant de simplifier le problème complexe de détection globale de formes dans l'espace image. En effet, dans cet espace paramétrique, la détection est locale et donc plus simple.[40]

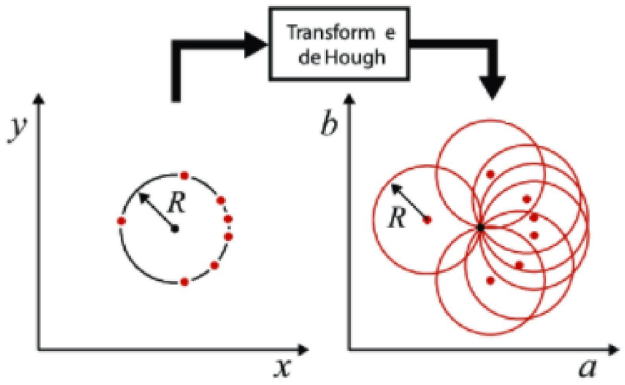


FIGURE 3.6 – la Transforme de Hough

[40]

• **La segmentation par étiquetage en composantes connexes :**

Les algorithmes d'ECC (l'étiquetage en composantes connexes) sont :

◦ Majoritairement séquentiels, irréguliers et utilisent une structure de graphe pour représenter les relations d'équivalences entre étiquettes ce qui rend complexe leur parallélisation.

◦ Mermet à partir d'une image binaire, de regrouper sous une même étiquette tous les pixels connexes.

◦ Le pont entre les traitements bas niveaux tels que le filtrage et ceux de haut niveau tels que la reconnaissance de forme ou la prise de décision. Il est donc impliqué dans un grand nombre de chaînes de traitements qui nécessitent l'analyse d'image segmentées. [41]

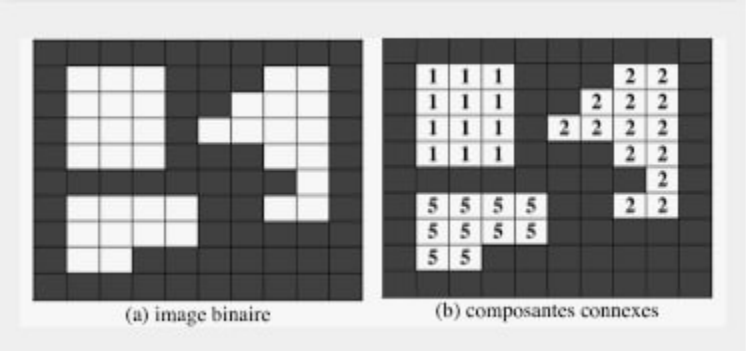


FIGURE 3.7 – a segmentation par étiquetage en composantes connexes

[41]

• **La segmentation par LPE(lignes de partage des eaux) :**

Définit par rapport à un processus d'inondation, qui consiste à partitionner l'image en différentes zones homogènes appelées **"bassins versants"**.applique a le gradient de l'image. l'image peut être perçue comme une surface (un relief) topographique,contenant des montagnes, des plateaux

et des vallées.Les pixels sombres correspondent donc aux vallées et bassins du relief alors que les pixels clairs correspondent aux collines et lignes de crêtes.[42]

3.5 Les methodes de la segmentation :

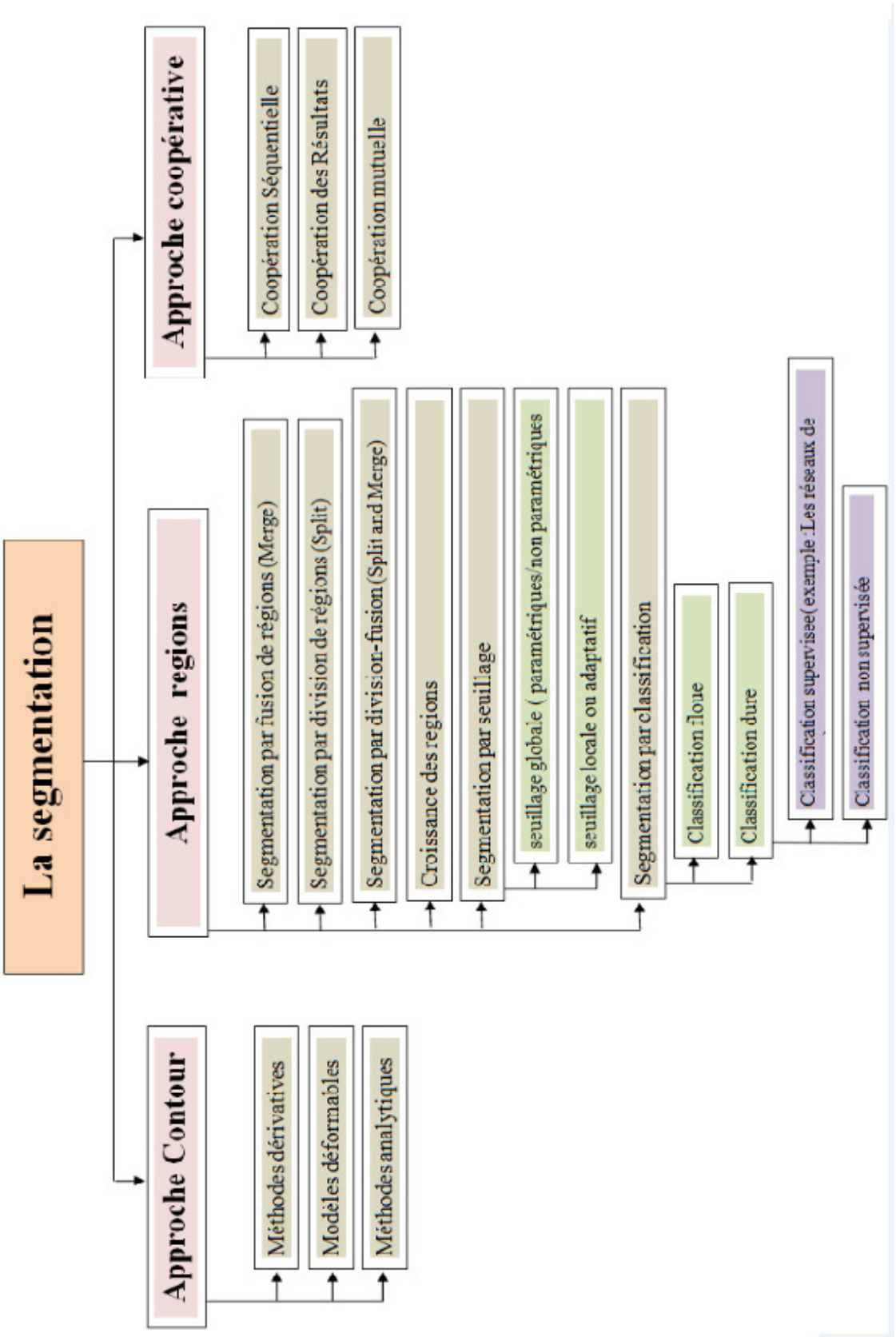


FIGURE 3.8 – Les methodes de la segmentation

Le Principes du modèle de segmentation entre des approches de segmentation qui ne prennent pas du tout en compte la structure des documents (segmentation en une suite de mots) alors qu'il s'agit d'une caractéristique forte des documents techniques, et celles qui ne proposent que cela nous devons parvenir à un équilibre.

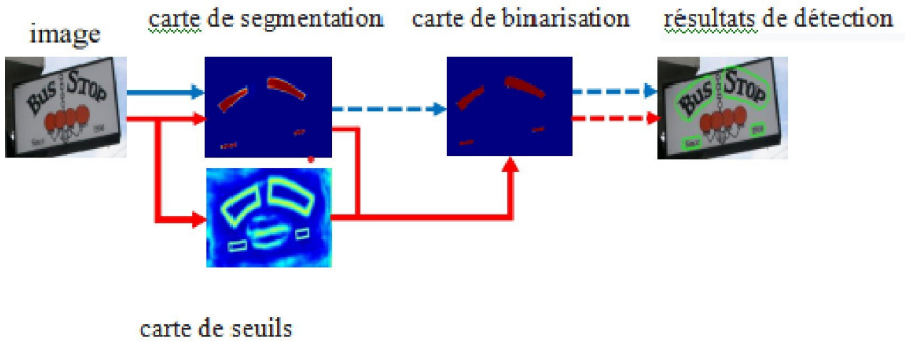


FIGURE 3.9 – exemple de modèle de segmentation sémantique

[43]

Ces dernières années, la lecture de texte dans des images de scène est devenue un domaine de recherche actif, en raison de ses larges applications pratiques telles que la compréhension d'images/vidéos, la recherche visuelle, la conduite et auxiliaire aveugle.
En tant qu'élément clé de la lecture du texte de la scène, le texte de la scène la détection qui vise à localiser la boîte englobante ou la région de chaque instance de texte est toujours une tâche difficile, car le texte de la scène est souvent à différentes échelles et formes, y compris texte horizontal, multi-orienté et incurvé. La détection de texte de scène basée sur la segmentation a attiré beaucoup d'attention récemment, car il peut décrire le texte de différentes formes, en profitant de ses résultats de prédiction au niveau du pixel. Cependant, la plupart des méthodes basées sur la segmentation nécessitent des post-traitement pour regrouper les résultats de la prédiction au niveau du pixel dans des instances de texte détectées.

3.6 La définition :

Egalement appelée classification au niveau des pixels,La segmentation sémantique associe une étiquette ou une catégorie à chaque pixel d'une image, elle permet de reconnaître un ensemble de pixels qui forment des catégories distinctes.

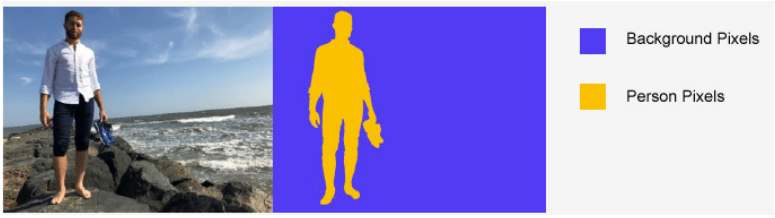


FIGURE 3.10 – Image et étiquette des pixels

[44]

La segmentation sémantique peut être réalisée avec ou sans supervision :

Supervisé : dans ce format, les images d'entraînement se présentent sous forme de paires d'images, ce processus d'apprentissage est donc analogue à la tâche de classification, c'est pourquoi nous l'appelons supervisé,on peut aussi appeler cette segmentation top-down. car l'information est d'abord intégrée de manière hiérarchique, et dans un second temps dépliée à la taille de l'image, à savoir une carte de segmentation.

Non supervisé : aucun apprentissage n'a lieu - contrairement à la segmentation supervisée, c'est pourquoi on l'appelle aussi segmentation ascendante.

3.6.1 Comment la segmentation sémantique est-elle utilisée ?

La segmentation sémantique étiquette les pixels d'une image c'est ce qui la rend utile dans des applications de divers domaines :

- Conduite autonome : pour identifier un parcours conduisible pour les véhicules en distinguant la route des obstacles tels que les piétons, trottoirs, poteaux et autres véhicules.

- Contrôles industriels : pour détecter les défauts dans des matériaux, comme le contrôle des composants électroniques.

- Imagerie satellite : pour identifier les montagnes, les rivières, les déserts et autres terrains.

- Imagerie médicale : pour analyser et détecter les anomalies cancéreuses dans les cellules.

- Vision robotique : pour identifier les objets et le terrain et s'y déplacer.[44]

3.6.2 Comment fonctionne la segmentation sémantique :

Le processus de formation d'un réseau de segmentation sémantique pour classer les images suit ces étapes :

- Analysez une collection d'images étiquetées au pixel près.

- Créer un réseau de segmentation sémantique.

- Entraînez le réseau à classer les images en catégories de pixels.

- Évaluer la précision du réseau.

3.6.3 Comprendre l'architecture :

- Une approche courante de la segmentation sémantique consiste à créer un SegNet, qui est basé sur une architecture de réseau neuronal convolutionnel (CNN).
Une architecture CNN typique qui classe l'image entière dans l'une des nombreuses catégories prédéfiniesest illustrée à la figure 3.12.

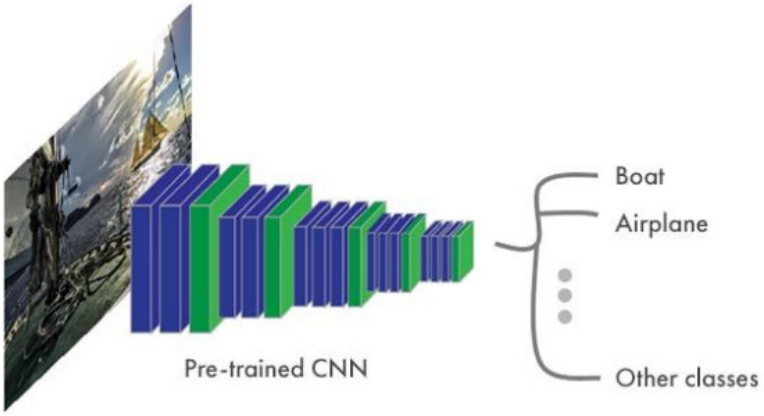


FIGURE 3.11 – Structure typique de CNN

[44]

- Pour classer au niveau du pixel au lieu de l'image entière, nous pouvons ajouter une implémentation inverse d'un CNN. Le processus de suréchantillonnage est effectué le même nombre de fois que le processus de sous-échantillonnage pour garantir que l'image finale a la même taille que l'image d'entrée. Enfin, une couche de sortie de classification de pixels est utilisée, qui mappe chaque pixel à une certaine classe, cela forme une architecture encodeur-décodeur, qui permet une segmentation sémantique.

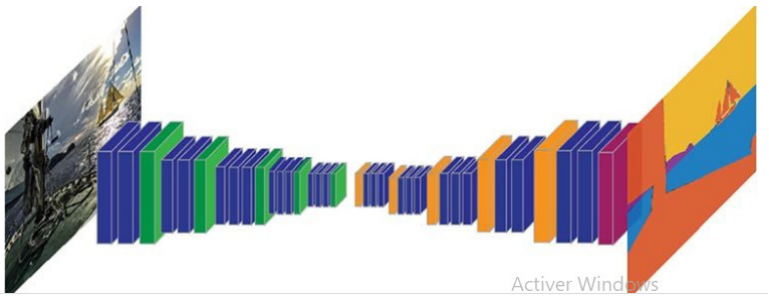


FIGURE 3.12 – encodeur-décodeur

[44]

3.7 Conclusion

Dans ce chapitre nous avons présenté la principe de segmentation commençant par les approches classiques de segmentation arrivant aux approches basées sur le deep learning. L'approche proposée et les détails d'implémentation sont abordées dans le chapitre suivant.

CHAPITRE 4

REALISATION DU SYSTEME

4.1 Introduction :

L'objectif de ce chapitre est de présenter les étapes de l'implémentation de l'approche proposée dans le cadre d'un système de La détection et la reconnaissance de textes d'images de scènes naturelles à l'aide de l'apprentissage profond. On commence tout d'abord par la présentation des ressources, du langage et de l'environnement de développement utilisé dans notre travail. Puis les étapes de la réalisation du modèle et on termine par les tests effectués.
Ce chapitre est composé de deux parties, l'implémentation du système et les résultats expérimentaux des testes.

L'approche proposée est basée sur la segmentation qui combinent généralement des prédictions effectuées par la segmentation sémantique et des post-traitements pour obtenir la limite des boites englobantes. Dans ce travail nous effectuons l'étape de prédiction par une approche proposée basée sur la segmentation sémantique et l'attention visuelle.

4.2 Environnement de développement :

4.2.1 Environnement matériels(Hardware) :

Nous présentons les expériences détaillées et les étapes d'évaluation pour tester l'efficacité du notre modèle, nos expériences étaient basées sur un ensemble de données d'images contienant du texte (Dataset Total-Text[45]). Toutes les expériences ont été réalisées sur un PC HP équipé d'un processeur Intel(R) Celeron(R) CPU N3060 @ 1.60GHz 1.60 GHz et une RAM de 8 Go et un GPU Intel(R)HD Graphics.

4.2.2 Environnement logiciel(Software) :

Python :

Python est un langage de programmation de haut niveau interprété pour la programmation à usage général. Créé par Guido van Rossum, et publié pour la première fois en 1991.Il repose sur une philosophie de conception qui met l'accent sur la lisibilité du code,notamment en utilisant des espaces significatifs. Il fournit des constructions permettant une programmation claire à petite et grande échelle.Python propose un système de typage dynamique et une gestion automatique de la mémoire. Il prend en charge plusieurs paradigmes de programmation, notamment orienté objet,impératif, fonctionnel et procédural, et dispose d'une bibliothèque standard étendue et complète. Python est un langage de programmation open-source et de

haut niveau, développé pour une utilisation avec une large gamme de systèmes d'exploitation. Il est qualifié de langage de programmation le plus puissant en raison de sa nature dynamique et diversifiée. Python est facile à utiliser avec une syntaxe super simple très encouragente pour les apprenants débutants, et très motivante pour les utilisateurs chevronés.contient plusieurs bibliothèques de l'intelligence artificiel populaires : (Keras, Tensorflow, Numpy, Pandas, OPENcv, Pytorch, etc.. . .).[46]

FIGURE 4.1 – Logo Python

[?]

TensorFlow :

TensorFlow est une plate-forme open source de bout en bout pour la création d'applications d'apprentissage automatique. Il s'agit d'une bibliothèque mathématique symbolique qui utilise un flux de données et une programmation différentiable pour effectuer diverses tâches axées sur la formation et l'inférence de réseaux de neurones profonds. Il permet aux développeurs de créer des applications d'apprentissage automatique à l'aide de divers outils, bibliothèques et ressources communautaires.

Actuellement, la bibliothèque d'apprentissage en profondeur la plus célèbre au monde est TensorFlow de Google. Google utilise l'apprentissage automatique dans tous ses produits pour l'optimisation des moteurs de recherche, la traduction, les légendes d'images ou les recommandations.

TensorFlow offre plusieurs niveaux d'abstraction afin que vous puissiez choisir celui qui convient le mieux à vos besoins. Créez et entraînez des modèles à l'aide de l'API Keras de haut niveau, ce qui facilite la mise en route de TensorFlow et l'apprentissage automatique[47].

FIGURE 4.2 – Logo TensorFlow

Keras :

Keras est une API de réseaux de neurones de haut niveau, écrite en Python et capable de fonctionner sur TensorFlow ou Theano. Il a été développé en mettant l'accent sur l'expérimentation rapide. Être capable d'aller de l'idée à un résultat avec le moins de délai possible est la clé pour faire de bonnes recherches. Il a été développé dans le cadre de l'effort de recherche du projet ONEIROS (Openended Neuro-Electronic Intelligent Robot Operating System), et son principal auteur et mainteneur est François Chollet, un ingénieur Google. En 2017, l'équipe TensorFlow de Google a décidé de soutenir Keras dans la bibliothèque principale de TensorFlow. Chollet a expliqué que Keras a été conçue comme une interface plutôt que comme un cadre d'apprentissage end to end. Il présente un ensemble d'abstractions de niveau supérieur et plus intuitif qui facilitent la configuration des réseaux neuronaux indépendamment de la bibliothèque informatique de backend. Microsoft travaille également à ajouter un backend CNTK à Keras aussi.[48].

FIGURE 4.3 – Logo Keras

Google colab :

Colab est un environnement de notebook Jupyter gratuit qui s'exécute entièrement dans le cloud., il ne nécessite pas de configuration et les blocs-notes que vous créez peuvent être modifiés simultanément par les membres de votre équipe . Colab prend en charge de nombreuses bibliothèques d'apprentissage automatique populaires qui peuvent être facilement chargées dans un ordinateur portable. Pour quoi Google Colab ?. A l'aide de Google Colab , vous pouvez effectuer les opérations ci-dessous :

- Ecrire et exécuter du code en Python.
- Documentez votre code qui prend en charge les équations mathématiques.
- Créer / télécharger / partager des blocs-notes.
- Importer / enregistrer des blocs-notes depuis / vers Google Drive.
- Importer / publier des blocs-notes depuis GitHub.
- Importez des ensembles de données externes, par ex. depuis Kaggle.
- Intégrer PyTorch, TensorFlow, Keras, OpenCV.

- Service Cloud gratuit avec GPU gratuit.[49]

FIGURE 4.4 – Logo Google colab

4.2.3 Les modèles utilisés :

- CNN (Convolutional Neural Network) :

Pour la classification des lésions bénignes et malignes nous avons utilisé les réseaux de neurones convolutive CNN. Caractérisées par les paramètres suivantes :

- L'apprentissage par transfert : En utilisant le principe de l'apprentissage par transfert, nous pouvons tirer parti des connaissances (caractéristiques, poids, etc.) Modèles déjà formés pour former des modèles plus récents et même résoudre des problèmes comme avoir moins de données pour la nouvelle tâche. L'apprentissage par transfert nous permet d'utiliser les connaissances des tâches apprises précédemment.

- Encodeur-Décodeur : Est un réseau de neurones , c'est un modèle de Machine Learning composé de deux réseaux de neurones en la même structure. La premiere sera utilisée normalement mais le deuxième fonctionnera de manière inversé.le premier réseau de neurones est un encodeur et le deuxième neurones est un décodeur.

4.3 L'Architecture de modéle :

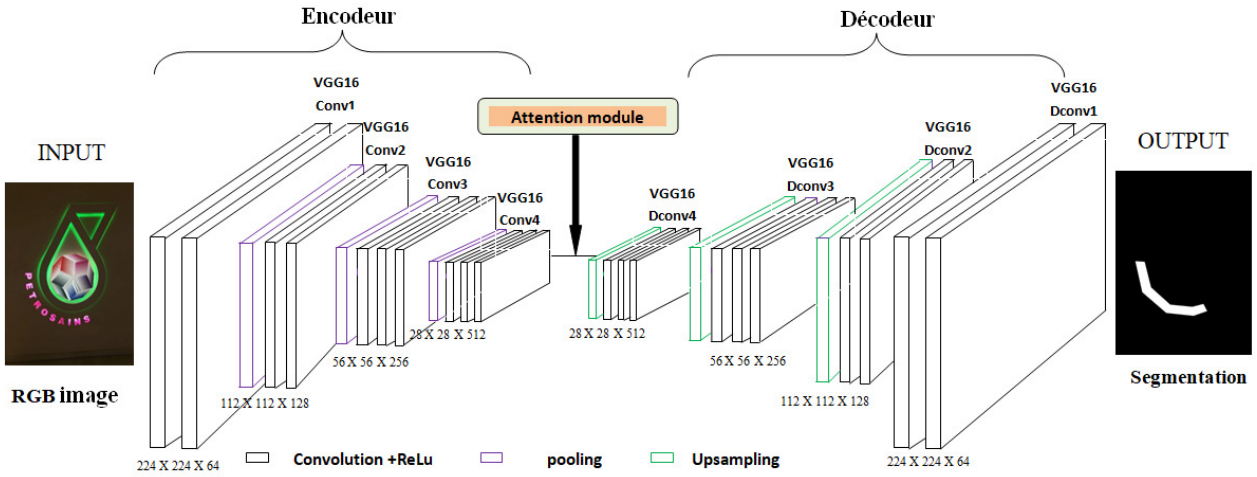


FIGURE 4.5 – architecture de modéle

Durant l'apprentissage du modèle l'architecture globale du modèle proposé se compose de :

– **Image entrée :**(image scéne naturelle contien un texte).

– **Les couches de modéle** :

Encodeur (VGG16) :

Est un réseau neuronal convolutif de 16 couches de profondeur est un algorithme connu en Computer Vision très souvent utilisé par transfert d'apprentissage pour éviter d'avoir à le réentraîner et résoudre des problématiques proches sur lesquelles VGG a déjà été entraîné. Il fonctionne avec un système imbriqué de 3*3 couches convolutives empilées les unes sur les autres.

Notre modéle Compose de 04 blocs ou chaque une contient des couches de convolution (extraction des caractéristiques) Bloc1 : (224 X 224 X 64),Bloc2 : (112 X 112 X 128),Bloc3 : (56 X 56 X 256) ,Bloc4 : (28 X 28 X 512)suivi par une classificateur (fonction d'activation Relu),Ces couches de convolution s'accompagnent de couche de Max-Pooling, chacune de taille 2×2, pour réduire la taille des filtres au cours de l'apprentissage.

Chaque couche de la couche d'extraction de caractéristiques prend en entrée la sortie de la couche qui la précède immédiatement, et sa sortie est transmise comme entrée aux couches suivantes.

La Convolution :est une opération mathématique simple généralement utilisée pour le traitement et la reconnaissance d'images. Sur une image, son effet s'assimile à un filtrage.

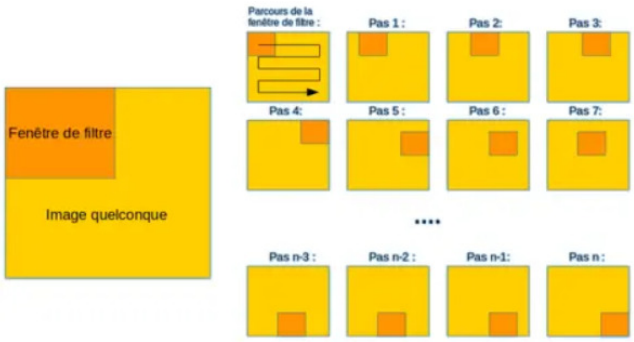


FIGURE 4.6 – convolution Layer

• On définit la taille de la fenêtre de filtre située en haut à gauche.

• La fenêtre de filtre, représentant la feature, se déplace progressivement de la gauche vers la droite d'un certain nombre de cases défini au préalable (le pas) jusqu'à arriver au bout de l'image.

• À chaque portion d'image rencontrée, un calcul de convolution s'effectue permettant d'obtenir en sortie une carte d'activation ou feature map qui indique où est localisées les features dans l'image : plus la feature map est élevée, plus la portion de l'image balayée ressemble à la feature.

Attention module :

Pooling layer (Max pooling- avrg Pooling) suivi par leur Concaténation .

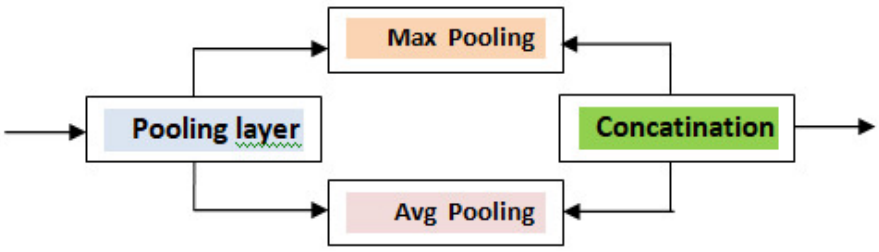


FIGURE 4.7 – Attention module

(Les couches de regroupement) fournissent une approche pour sous-échantillonner les cartes d'entités en résumant la présence d'entités dans des parcelles de la carte d'entités. Deux méthodes de mise en commun courantes sont la mise en commun moyenne (Avrg-Pooling) et la mise en commun maximale (Max-Pooling) qui résument respectivement la présence moyenne d'une fonctionnalité et la plus grande présence d'une fonctionnalité.

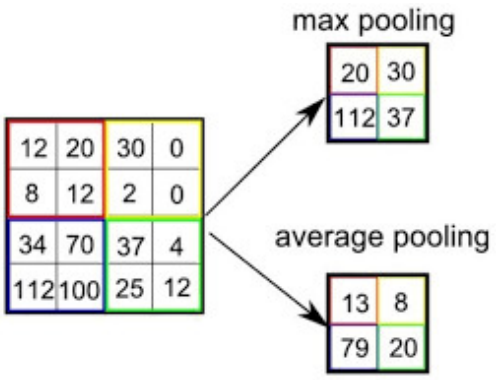


FIGURE 4.8 – Pooling Layer

et aprés nous faisons une Concaténation entre les deux couches (Avrg-Pooling) et (Max-Pooling).

Decodeur :

Fonctionne a l'Inverse de 1 er partie (Encodeur) nomé **Upsampling** ou le suréchantillonnage d'une image est tout le contraire de la convolution. Nous pouvons rendre l'image de plus en plus petite par convolution normale, tandis que le suréchantillonnage peut également rendre l'image de plus en plus grande par convolution, et enfin redimensionner à la même taille que l'image d'origine. il est Composé de 04 blocs ou chaque une contient une couche de Upsamlping et des couches de convolution Bloc4 : (28 X 28 X 512),Bloc3 : (56 X 56 X 256) ,Bloc2 (112 X 112 X 128) :,Bloc1 (224 X 224 X 64).

– **Image Sortie :**

Carte de prédiction que nous pouvons la représentée par une image niveau de gris contient la prédiction de chaque pixel par une valeur entre 0 et 1

Si la valeur de pixel = 0 c'est l'arrière plan

Si la valeur de pixel = 1 c'est le texte détecté

Les valeurs de pixels entre 0 et 1 sont des prédictions (prédiction de texte detecté), nous pouvons appliquer une binarisation ou postraitement afin de améliorer la détection.

4.3.1 Les étapes du code :

Dans cette partie nous décrirons l'ensemble des données et le code source utilisée dans ce travail.

Dataset :

Total-Text[45] est également un ensemble de données récemment publié pour la détection de texte de courbe. Cet ensemble de données comprend des instances de texte horizontales, multi-orientées et courbés et se compose de 1255 images d'entraînement et de 300 images de test.

- Importation des données : illustre les instructions de téléchargement des images à partir de Drive à Google Colab.

```
from google.colab import drive
drive.mount('/content/drive')
!rm -rf /content/sample_data
!unzip /content/drive/MyDrive/groundtruth_textregion.zip
!unzip /content/drive/MyDrive/totaltext.zip
```

FIGURE 4.9 – Code Python pour Importation des données

```
  inflating: groundtruth_textregion/__MACOSX/Text_Region_Mask/Train/._img359.png
  inflating: groundtruth_textregion/Text_Region_Mask/Train/img1536.png
  inflating: groundtruth_textregion/Text_Region_Mask/Train/img1500.png
  inflating: groundtruth_textregion/Text_Region_Mask/Train/img1519.png
  inflating: groundtruth_textregion/__MACOSX/Text_Region_Mask/Train/._img417.png
  inflating: groundtruth_textregion/__MACOSX/Text_Region_Mask/Train/._img445.png
  inflating: groundtruth_textregion/__MACOSX/Text_Region_Mask/Train/._img375.png
Archive:  /content/drive/MyDrive/totaltext.zip
  inflating: totaltext/__MACOSX/._Images
  inflating: totaltext/__MACOSX/Images/._Train
  inflating: totaltext/__MACOSX/Images/._Test
  inflating: totaltext/__MACOSX/Images/Test/._img898.jpg
  inflating: totaltext/__MACOSX/Images/Test/._img94.jpg
  inflating: totaltext/__MACOSX/Images/Test/._img996.jpg
  inflating: totaltext/__MACOSX/Images/Test/._img998.jpg
  inflating: totaltext/__MACOSX/Images/Test/._img91.jpg
```

FIGURE 4.10 – Télécharger l'ensemble de données pour entraîner le modèle

- La liste des bibliothèques utilisées :

```python
import os
import cv2
import numpy as np
from tqdm import tqdm
import tensorflow as tf
import matplotlib.pyplot as plt

from sklearn.utils import shuffle
from sklearn.model_selection import KFold
from sklearn.metrics import classification_report
from sklearn.model_selection import train_test_split

from keras.models import *
from keras.layers import *
from keras import backend as keras
from keras.preprocessing.image import ImageDataGenerator, image
from keras.callbacks import ModelCheckpoint, LearningRateScheduler, EarlyStopping, ReduceLROnPlateau
from keras.applications.vgg16 import preprocess_input
```

FIGURE 4.11 – la liste des bibliothèques

- Le modéle CNN :

```python
baseModel = tf.keras.applications.VGG16(weights = "imagenet" , include_top = False, input_tensor = Input(shape = (224, 224, 3)))
conv1 = baseModel.get_layer('block1_conv2').output

conv2 = baseModel.get_layer('block2_conv2').output

conv3 = baseModel.get_layer('block3_conv2').output

conv4 = baseModel.get_layer('block4_conv3').output
src = baseModel.get_layer('block4_pool').output

#Attention module
pool = MaxPooling2D(pool_size = (2, 2), strides=1, padding = 'same')(src)
avrg = AveragePooling2D(pool_size=(2, 2), strides=1, padding = 'same')(src)
conc = concatenate([pool, avrg], axis = -1)

#Upsampling
conv = Conv2D(1024, (7,7), activation = 'relu', padding = 'same')(conc)
conv = Conv2D(1024, (7,7), activation = 'relu', padding = 'same')(conv)
up1 = concatenate([Conv2DTranspose(512, (2, 2), strides = (2, 2), padding = 'same')(conv), conv4], axis = 3)

#Upsampling
conv5 = Conv2D(512, (7,7), activation = 'relu', padding = 'same')(up1)
conv5 = Conv2D(512, (7,7), activation = 'relu', padding = 'same')(conv5)
up2 = concatenate([Conv2DTranspose(256, (2, 2), strides = (2, 2), padding = 'same')(conv5), conv3], axis = 3)
```


FIGURE 4.12 – Code Python pour Model CNN

```
#Upsampling
conv6 = Conv2D(256, (7,7), activation = 'relu', padding = 'same')(up2)
conv6 = Conv2D(256, (7,7), activation = 'relu', padding = 'same')(conv6)
up3 = concatenate([Conv2DTranspose(128, (2, 2), strides = (2, 2), padding = 'same')(conv6), conv2], axis = 3)


#Upsampling
conv7 = Conv2D(128, (7,7), activation = 'relu', padding = 'same')(up3)
conv7 = Conv2D(128, (7,7), activation = 'relu', padding = 'same')(conv7)
up4 = concatenate([Conv2DTranspose(64, (2, 2), strides = (2, 2), padding = 'same')(conv7), conv1], axis = 3)


conv8 = Conv2D(64, (7,7), activation = 'relu', padding = 'same')(up4)
conv8 = Conv2D(64, (7,7), activation = 'relu', padding = 'same')(conv8)
conv10 = Conv2D(1, (1, 1), activation = 'sigmoid')(conv8)
model = Model(inputs = baseModel.input, outputs = conv10)
img = cv2.resize(cv2.imread("/content/drive/MyDrive/IMG_0001.jpg"), (224, 224))
x = image.img_to_array(img)
x = np.expand_dims(x, axis=0)
x = preprocess_input(x)
b4 = model.predict(x)
```

FIGURE 4.13 – Suite Model CNN

```
Downloading data from https://storage.googleapis.com/tensorflow/keras-applications/vgg16/vgg16_weights_tf_dim_ordering_tf_kernels_notop.h5
58892288/58889256 [==============================] - 0s 0us/step
58900480/58889256 [==============================] - 0s 0us/step
```

FIGURE 4.14 – télechagement

- Impression le image de sortie avec 1 chanel avec les méme dimention de l'image d'entree 224x224 :

```
[ ]  print(b4.shape)
     #print(b4)

     (1, 224, 224, 1)
```

FIGURE 4.15 – image output

- Changement de la dimension de dataset :

Importé les images originals et change leurs taille et inséré dans la listes des images imgs.

Importé les images résultas et change leurs taille et inséré dans la listes des images msks.

Transfére les image en array et ensuite en float .

```
def trainData(imgs_path, msks_path):
  imgs = []
  msks = []
  for imgname in tqdm(os.listdir(imgs_path)):
    img = cv2.resize(cv2.imread(imgs_path + imgname), (224, 224))#[..., 0]

    imgs.append(img)

  for imgname in tqdm(os.listdir(msks_path)):

    msk = cv2.resize(cv2.imread(msks_path + imgname), (224, 224))[..., 0]

    msks.append(msk)
  return shuffle((np.array(imgs).reshape(len(imgs), 224, 224, 3) / 127.0) - 1.0, (np.array(msks).reshape(len(msks), 224, 224, 1) > 127).astype(np.float32))
imgs, msks = trainData("/content/drive/MyDrive/totaltext/Images/Train/", "/content/drive/MyDrive/groundtruth_textregion/Text_Region_Mask/Train/")
print(imgs.shape, msks.shape)
```

FIGURE 4.16 – changement de la dimension de dataset

```
100%|██████████| 1255/1255 [00:54<00:00, 22.85it/s]
100%|██████████| 1255/1255 [00:28<00:00, 43.68it/s]
(1255, 224, 224, 3) (1255, 224, 224, 1)
```

FIGURE 4.17 – résultet de changement de la dimension de dataset

- Sauvgarder les resultats :

Création des Callbacks à appeler aprés chaque epoch pour sauvgarde les résultat (Savgarder les poids (weights) qui donne les bon résultats en cour de l'entrainement) et s'il n'est un pas une bonne amelioration apré 10 fois il faux arréter.

```
[10] checkpoint = ModelCheckpoint("/content/drive/MyDrive/weight.hdf5", monitor = 'val_loss', verbose = 1,
                                  save_best_only = True, mode = 'min', save_weights_only = True)
     """
     reduceLROnPlat = ReduceLROnPlateau(monitor = 'val_loss', factor = 0.5,
                                        patience = 3, verbose = 1, mode = 'min',
                                        min_delta = 0.0001, cooldown = 2, min_lr = 1e-6)
     """
     early = EarlyStopping(monitor = "val_loss", mode = "min", patience = 10)

     callbacks_list = [checkpoint, early]#, reduceLROnPlat]
```

FIGURE 4.18 – sauvgarder les poids (weights)

- Compilation :

Après avoir exécuté la ligne ci-dessus, le modèle commencera à s'entraîner et vous commencerez à voir la précision et la perte d'entraînement/validation. 20 % pour le teste , 80 % pour l'entainemment ,batch-size=8 ,epochs=20.

```
[11] model.compile(optimizer = tf.keras.optimizers.Adam(learning_rate = 1e-3),
                   loss = [dice_coef_loss], metrics = [dice_coef, 'categorical_crossentropy'])

     trn_imgs, tst_imgs, trn_msks, tst_msks = train_test_split(imgs, msks, test_size = 0.2)

     trn_imgs, val_imgs, trn_msks, val_msks = train_test_split(trn_imgs, trn_msks, test_size = 0.2)

     loss_history = model.fit(x = trn_imgs,
                              y = trn_msks,
                              batch_size = 8,
                              epochs = 20,
                              validation_data = (val_imgs, val_msks),
                              callbacks = callbacks_list)
```

```
Epoch 1/20
101/101 [==============================] - ETA: 0s - loss: -0.1550 - dice_coef: 0.1560 - categorical_crossentropy: 0.0000e+00
Epoch 1: val_loss improved from inf to -0.14850, saving model to /content/drive/MyDrive/weight.hdf5
101/101 [==============================] - 201s 2s/step - loss: -0.1550 - dice_coef: 0.1560 - categorical_crossentropy: 0.0000e+00 - val_loss: -0.1485 - val_dice_
Epoch 2/20
101/101 [==============================] - ETA: 0s - loss: -0.1551 - dice_coef: 0.1551 - categorical_crossentropy: 0.0000e+00
Epoch 2: val_loss did not improve from -0.14850
101/101 [==============================] - 179s 2s/step - loss: -0.1551 - dice_coef: 0.1551 - categorical_crossentropy: 0.0000e+00 - val_loss: -0.1485 - val_dice_
Epoch 3/20
```


FIGURE 4.19 – Compilation

- Evaluation de modéle :

Une fonction loss est une mesure de la précision avec laquelle notre modèle est capable de prédire le résultat attendu.

Une valeur élevée pour la perte (loss) signifie que notre modèle a très mal fonctionné. Une faible valeur de perte (loss) signifie que notre modèle a très bien fonctionné.

Dans notre modèle la fonction loss (score[0]) a donné une faible valeur = 0.0411 comme mentionné dans la figure 4.26

Pour l'évaluation de modèle nous avons utilisé les métrique de segmentation d'image définie par Keras (score[1]) : meanIuO : Calcule la métrique moyenne d'Intersection-Over-Union. Dis -coef.

```
score = model.evaluate(val_imgs, val_msks, verbose = 0)
print(score[0])
print(score[1])
```

```
0.041183408349752426
0.4590834379196167
```

FIGURE 4.20 – Evaluation de modéle

- Résumer l'histoire de la perte :

```
[ ]  import matplotlib.pyplot as plt


     # summarize history for loss
     plt.plot(loss_history.history['loss'])
     plt.plot(loss_history.history['val_loss'])
     plt.title('model loss')
     plt.ylabel('loss')
     plt.xlabel('epoch')
     plt.legend(['Train', 'Validation'], loc='upper left')
     plt.show()
```

FIGURE 4.21 – code python résumer l'histoire de la perte

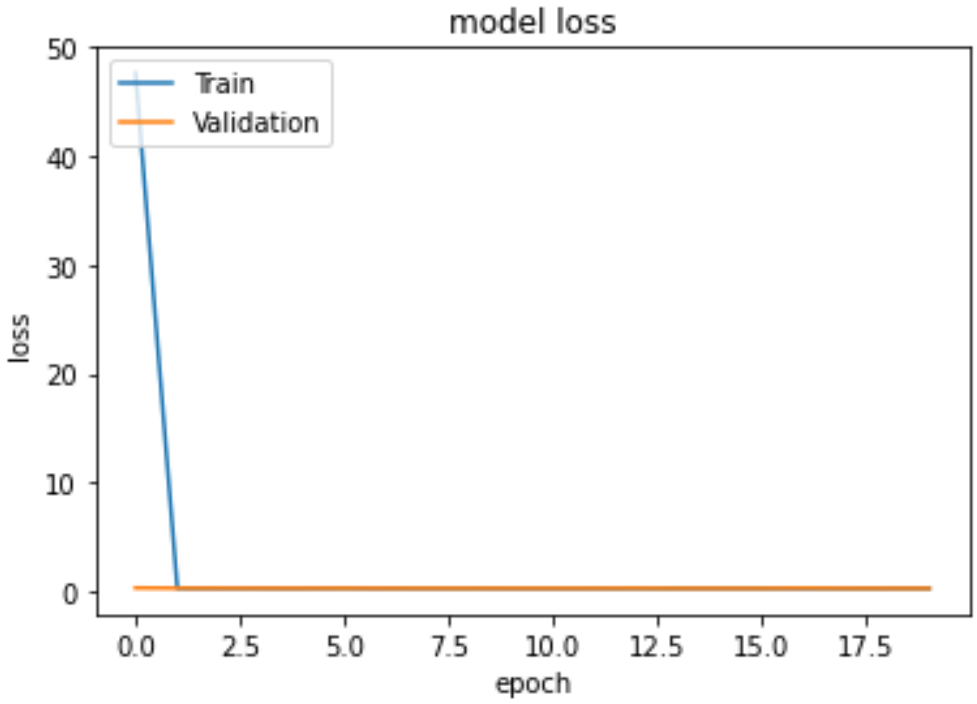


FIGURE 4.22 – Résumer l'histoire de la perte

- Précision binaire du modèle :

```
[ ]  # summarize history for loss
     plt.plot(loss_history.history['binary_accuracy'])
     plt.plot(loss_history.history['val_binary_accuracy'])
     plt.title('model binary_accuracy')
     plt.ylabel('binary_accuracy')
     plt.xlabel('epoch')
     plt.legend(['Train', 'Validation'], loc='upper left')
     plt.show()
```

FIGURE 4.23 – code python précision binaire du modèle

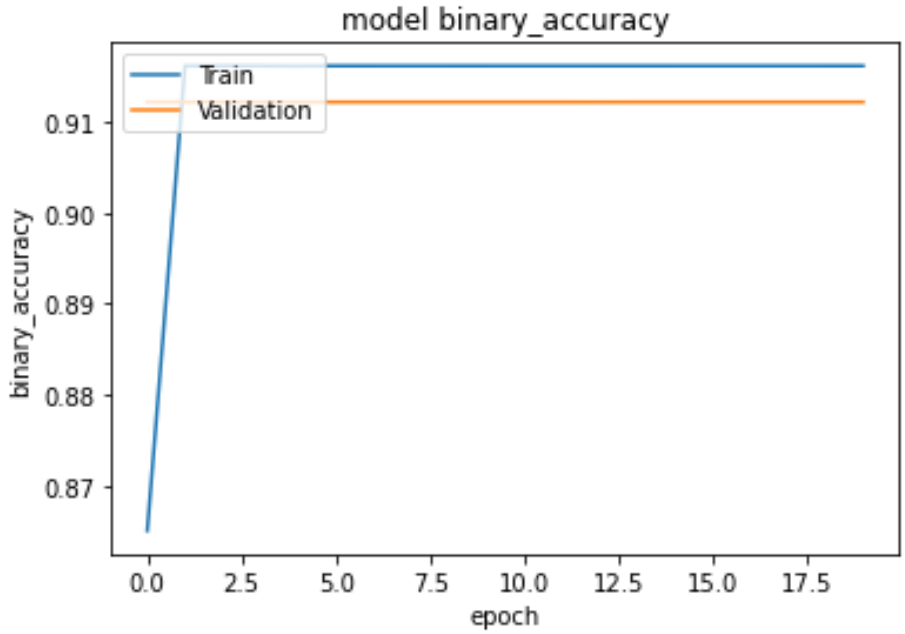


FIGURE 4.24 – précision binaire du modèle

- Coeffessions du modèle :

```
[ ]  # summarize history for loss
     plt.plot(loss_history.history['dice_coef'])
     plt.plot(loss_history.history['val_dice_coef'])
     plt.title('model dice_coef')
     plt.ylabel('dice_coef')
     plt.xlabel('epoch')
     plt.legend(['Train', 'Validation'], loc='upper left')
     plt.show()
```

FIGURE 4.25 – code python Coeffessions du modèle

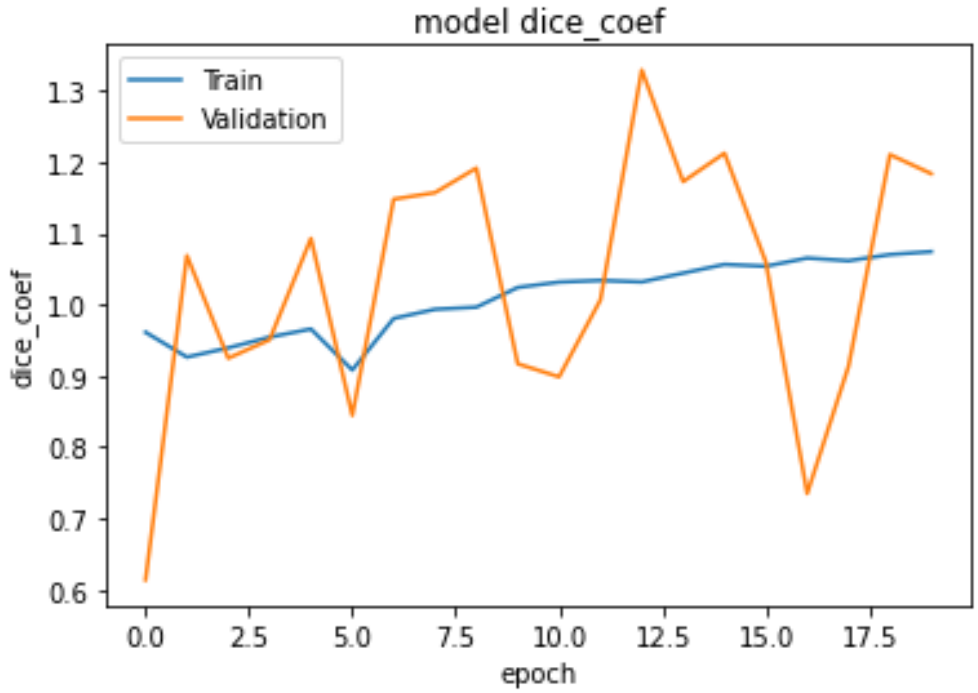


FIGURE 4.26 – Coeffessions du modèle

- Prédiction de modéle :

```
[ ]  #img = cv2.resize(cv2.imread("/content/drive/MyDrive/Images/Test/img1.jpg"), (224, 224))
     img = cv2.imread("/content/drive/MyDrive/Images/Test/img1.jpg")
     #im1 = model.predict(img)
     from skimage import io

     io.imshow(img)
     img = cv2.resize(cv2.imread("/content/drive/MyDrive/Images/Test/img1.jpg"), (224, 224))
     x = image.img_to_array(img)
     x = np.expand_dims(x, axis=0)
     x = preprocess_input(x)
     b4 = model.predict(x)
     print(b4.shape)
     io.imshow(img)
```

FIGURE 4.27 – Prédiction de modéle

Résultats et expériences :

Dans ce qui suit, le résultat obtenue après l'apprentissage par transfert de la segmentation sémantique de notre modéle .

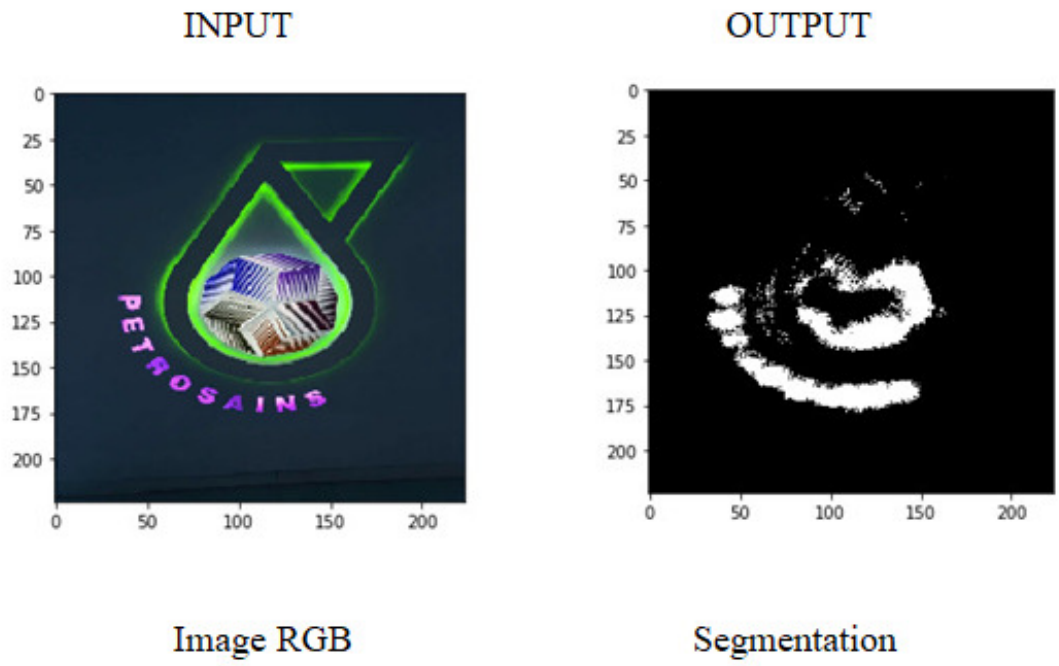


FIGURE 4.28 – Prédiction de modéle

D'autres résultats

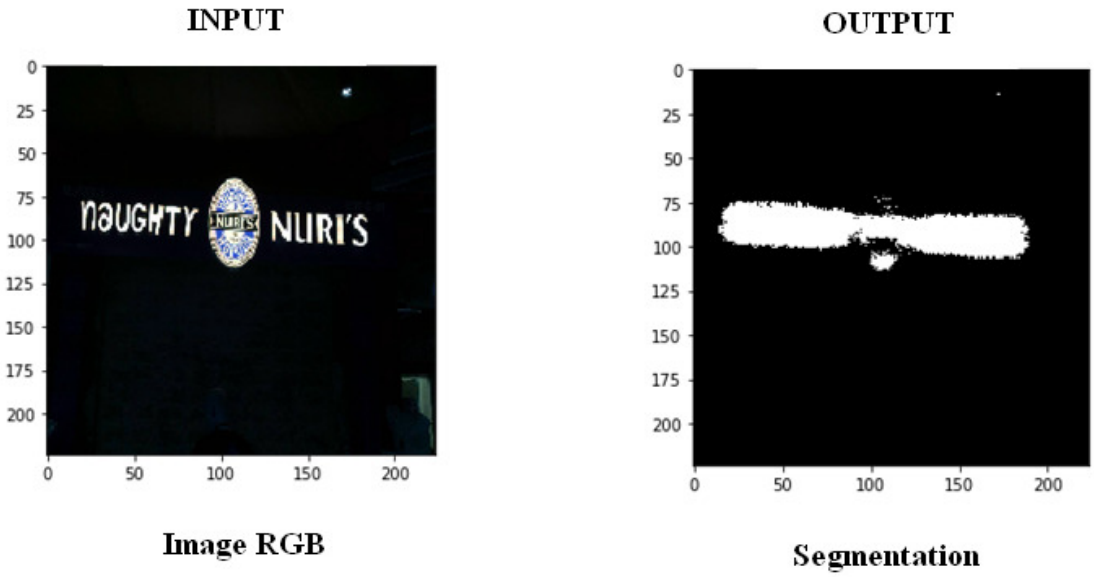


FIGURE 4.29 – Exemple d'image de Test

un autre exemple :

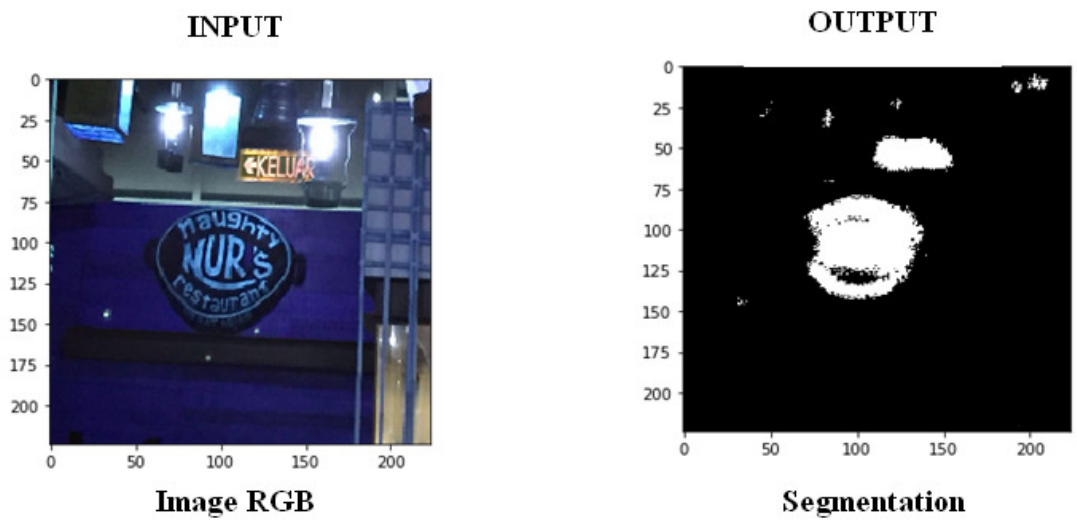


FIGURE 4.30 – Autre image de Test

4.4 Conclusion

Dans ce chapitre nous avons décrit le modèle CNN pour la segmentation sémantique afin d'effectuer la prédiction des zones de textes dans les images, les résultats obtenus sont acceptable et prouve l'efficacité de l'approche proposée.

CONCLUSION GÉNÉRALE :

La segmentation sémantique d'images est une tâche importante dans le domaine de la vision par ordinateur, la reconnaissance de texte et l'apprentissage automatique. Grâce à l'apprentissage en profondeur (Deep Learning) , l'avenir de l'intelligence artificielle dans le développement de grandes applications est très rapides.

Dans ce projet nous avons présenté une des opérations de traitements des images qui est la détection de textes d'images de scènes naturelles à l'aide de l'apprentissage profond en utilisant un ensembles de données des images contenant du textes horizontales, multi-orientées et courbés.

Nous avons proposé un modèle deep learning basé sur la segmentation sémantique augmenté par le méchanisme d'attention visuelle qui a montré ses performances ces dernières années.

Les résultats obtenus sont acceptable sur l'ensemble de données utilisé.

Dans le futur nous pouvons améliorer les résultats obtenus en entrainant le modèle sur plusieurs autres ensemble de données disponibles et populaires.

.

BIBLIOGRAPHIE

[1] K. CHEHDI, B. VOZEL, C. KERMAD, and V. PITURESCU, "Système aveugle de filtrage d'images numériques," in *17° Colloque sur le traitement du signal et des images, FRA, 1999.* GRETSI, Groupe d'Etudes du Traitement du Signal et des Images, 1999.

[2] G.-A. B. Salah-Nicolas, R.-K. Ait-Mohand, and T. Paquet, *Etat de l'art sur la caractérisation d'un document à OCRiser*, 2012.

[3] H. M. Jalil and B. Soumeya, *Recherche de mots dans les images de documents arabes*, 2011.

[4] H. O. EDDINE, *Une approche pour la détection du texte dans les images*, 2021.

[5] A. Gide, "Chapitre 3 l'apprentissage automatique en télédétection," *rat*, p. 63, 2017.

[6] H. BELLAHMER, *Implémentation et évaluation d'un modèle d'apprentissage automatique pour l'estimation de la valeur marchande de propriétés immobilières*, 2020.

[7] D. Chen, "Text detection and recognition in images and video sequences," Tech. Rep., 2003.

[8] J.-M. DUCROT, *l'utilisation de la vidéo en classe de français langue étrangère .*

[9] universite paris saclay, *Machine pour la Compréhension de Scènes bertrand le saux.*

[10] blog seo, *Egrimaud articles récents*, 07 2021.

[11] [Online]. Available : https://www.lri.fr/~chapuis/ihm-polytech/intro.pdf

[12] Le système d'aide à la conduite. [Online]. Available : http://fr.wikipedia.org/wiki/Systeme

[13] *La technologie OCR basée sur l'IA révolutionne le secteur bancaire*, 11 2020.

[14] [Online]. Available : https://moov.ai/fr/blog/reconnaissance-optique-de-caracteres-ocr/

[15] [Online]. Available : https://www.abbyy.com/fr/solutions/insurance

[16] [Online]. Available : https://www.abbyy.com/fr/solutions/insurance

[17] M. ZERROUGUI and S. HAMADENE, "Détection de la tumeur cérébrales dans l'image irm par l'apprentissage en profondeur." Ph.D. dissertation, Université Mohamed el-Bachir el-Ibrahimi Bordj Bou Arréridj Faculté de ..., 2021.

[18] [Online]. Available : https://www.futura-sciences.com/tech/definitions/intelligence-artificielle-deep-learning-17262

[19] Les algorithmes de deep learning. [Online]. Available : https://www.jedha.co/formation-ia/algorithmes-deep-learning

[20] [Online]. Available : https://actualiteinformatique.fr/intelligence-artificielle/definition-convolutional-neural-network

[21] N. Malki and T. Guerram, *Classification automatique des textes par Les réseaux de neurones à convolution*, 2019.

[22] [Online]. Available : https://fr.m.wikipedia.org

[23] M. Liao, B. Shi, and X. Bai, "Textboxes++ : A single-shot oriented scene text detector," *IEEE transactions on image processing*, vol. 27, no. 8, pp. 3676–3690, 2018.

[24] M. Liao, J. Zhang, Z. Wan, F. Xie, J. Liang, P. Lyu, C. Yao, and X. Bai, "Scene text recognition from two-dimensional perspective," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, 2019, pp. 8714–8721.

[25] L. G. L. Decker1a, A. Pinto, J. L. F. Campana, M. C. Neira, A. A. dos Santos, J. S. Conceiçao, M. A. Angeloni, L. T. Li *et al.*, "Mobtext : a compact method for scene text localization," 2020.

[26] L. Xie, Y. Liu, L. Jin, and Z. Xie, *DeRPN : Taking a further step toward more general object detection*, 2019.

[27] T. LELORE, *Segmentation d'image*, 2007.

[28] C. Pimm, *Real-time Scene Text Detection with Differentiable Binarization*, 2019.

[29] ——, "Quelle plus-value linguistique pour la segmentation automatique de texte ?'," in *Actes du Colloque international'Discours et Document'(ISDD'06)*, 2006, pp. 85–94.

[30] C. Xue, S. Lu, and F. Zhan, "Accurate scene text detection through border semantics awareness and bootstrapping," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 355–372.

[31] P. Lyu, M. Liao, C. Yao, W. Wu, and X. Bai, "Mask textspotter : An end-to-end trainable neural network for spotting text with arbitrary shapes," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 67–83.

[32] X.-O. Zhang, R. Dong, Y. Zhang, J.-L. Zhang, Z. Luo, J. Zhang, L.-L. Chen, and L. Yang, "Diverse alternative back-splicing and alternative splicing landscape of circular rnas," *Genome research*, vol. 26, no. 9, pp. 1277–1287, 2016.

[33] Z. Zhang, C. Zhang, W. Shen, C. Yao, W. Liu, and X. Bai, "Multi-oriented text detection with fully convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 4159–4167.

[34] B. Shi, X. Bai, and S. Belongie, "Detecting oriented text in natural images by linking segments," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2550–2558.

[35] X. Zhou, C. Yao, H. Wen, Y. Wang, S. Zhou, W. He, and J. Liang, "East : An efficient and accurate scene text detector," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[36] M. ABDELOUAHEB, *Utilisation de méthodes de Deep learning pour l'extraction de texte dans les images*.

[37] S. M.-S. Leclerc, "Automatisation de la segmentation sémantique de structures cardiaques en imagerie ultrasonore par apprentissage supervisé," Ph.D. dissertation, Université de Lyon, 2019.

[38] [Online]. Available : https://patrick-bonnin.developpez.com/cours/vision/apprendre-bases-traitement-image/partie-3-introduction-differents-types-segmentation

[39] O. Abdelli, "Segmentation d'images par seuillage d'histogrammes bidimensionnels." Ph.D. dissertation, Université Mouloud Mammeri, 2011.

[40] N. Poncelet and Y. Cornet, "Transformée de hough et détection de linéaments sur images satellitaires et modèles numériques de terrain," *Bulletin de la Société Géographique de Liège*, vol. 54, pp. 145–156, 2010.

[41] L. Cabaret, "Algorithmes d'étiquetage en composantes connexes efficaces pour architectures hautes performances," Ph.D. dissertation, Université Paris-Saclay (ComUE), 2016.

[42] F. M. K. TAGZIRT, *Segmentation d'images par la méthode de la Ligne de Partage des Eaux*, 2020.

[43] M. Liao, Z. Wan, C. Yao, K. Chen, and X. Bai, "Real-time scene text detection with differentiable binarization," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 07, 2020types methodes, pp. 11 474–11 481.

[44] [Online]. Available : https://www.mathworks.com/solutions/image-video-processing/semantic-segmentation.html

[45] C. K. Ch'ng and C. S. Chan, "Total-text : A comprehensive dataset for scene text detection and recognition," in *2017 14th IAPR international conference on document analysis and recognition (ICDAR)*, vol. 1. IEEE, 2017, pp. 935–942.

[46] Y. Derfoufi, *Programmation en langage Python*, 2019.

[47] A. MHAMMEDI, I. YAKOUB, A. OUAHAB *et al.*, "La détection de covid-19 par l'apprentissage profonde (deep learning)," Ph.D. dissertation, UNIVERSITE AHMED DRAIA-ADRAR, 2021.

[48] A. BENDJAAFER, T. MEDDAH *et al.*, "Classification d'image à l'aide d'un réseau totalement convolutionnelle," Ph.D. dissertation, university of M'sila, 2021.

[49] A. BELFARHI, *automatisation de la sélection des paramètres optimisés de filtre de Gabor pour la détection des expressions faciales.*, 2020.