



جامعة عبد المالك السعدي
 ٢٠٠٨٠٤٢٢ | ٢٠٠٨١٢٢٢ ٢٢٢٢٢٢
 Université Abdelmalek Essaadi

Faculté des sciences AbdelMalek Essaâdi
Département de mathématique

Reconnaissance optique des caractères

Realise par:
CHAMA El habib
MANSSAR Islam
KAMROUN Saad

Superviseur: Pr. Abdeltif EL OUISSARIA

May 17, 2025

Contents

1	Introduction	1
1.1	Contexte et motivation	1
1.2	Objectifs du projet	1
2	Introduction a la reconnaissance de caractere optique	2
2.1	C'est quoi la reconnaissance de caractere optique (OCR) ?	2
2.2	OCR classique	2
2.3	OCR moderne	3
3	Application web pour l'extraction du text	6
3.1	Introduction	6
3.2	Conception du model CRNN	6
3.3	Test et Resultats	8
3.4	Conception de notre site	8
3.4.1	digramme de cas d'utilisation	9
3.4.2	outils et Frameworks utiliser	9
4	Conclusion et perspectives	10
4.1	Résumé du projet	10
4.2	Travaux futurs	10
5	Annexes	12

List of Figures

2.1	Les étapes de l'OCR classique	3
2.2	Les étapes de l'OCR moderne cas du CRNN	4
3.1	Architecture du premier reseau (CRAFT)	7
3.2	Architecture du deuxieme reseau CRNN (VGG ou Resnet)	7
3.3	Pérfomences de CRNN(VGG/ResNet) sur les jeux de données	8
3.4	Digramme de cas d'utilisation	9
3.5	outils de developement	9
4.1	site web ESI-OCR pour l'extraction du text	10

Chapter 1

Introduction

1.1 Contexte et motivation

L'objectif du clustering est de segmenter et diviser un ensemble de données en sous-groupes, en utilisant une mesure de similarité, afin de regrouper des observations similaires pour tenter de reconnaître des motifs ou révéler des informations cachées à partir des données.

1.2 Objectifs du projet

Dans ce projet nous avons pour objectif d'implémenter la reconnaissance des caractères optiques dans une application web afin de simplifier l'extraction du texte de différents types de documents (image, pdf ...)

Chapter 2

Introduction a la reconnaissance de caractere optique

2.1 C'est quoi la reconnaissance de caractere optique (OCR) ?

La reconnaissance optique de caractères (OCR) est la technologie qui transforme les images de texte en **texte numérique lisible** par machine, permettant aux ordinateurs de traiter le contenu de documents ou de scènes contenant du text.

En pratique, l'OCR prend un document numérisé, une photographie ou toute image contenant des documents imprimés ou texte manuscrit et convertit les caractères en données modifiables et consultables. Ce processus offre la possibilité de numériser des documents imprimés (livres, factures, relevés bancaires, etc.) afin qu'ils puissent être édité, recherché et stocké électroniquement.

Avantage

- **Medical** extraction d'informations à partir d'ordonnances ou de rapports.
- **Administrative** Numérisation de documents anciens ou manuscrits.
- **Social** Traduction du text en temp reel.

2.2 OCR classique

L'OCR classique fait référence aux méthodes traditionnelles utilisées pour reconnaître le texte. Ces systèmes s'appuyaient fortement sur la reconnaissance de formes et règles élaborées. Le pipeline général d'un système OCR classique comportait plusieurs étapes :

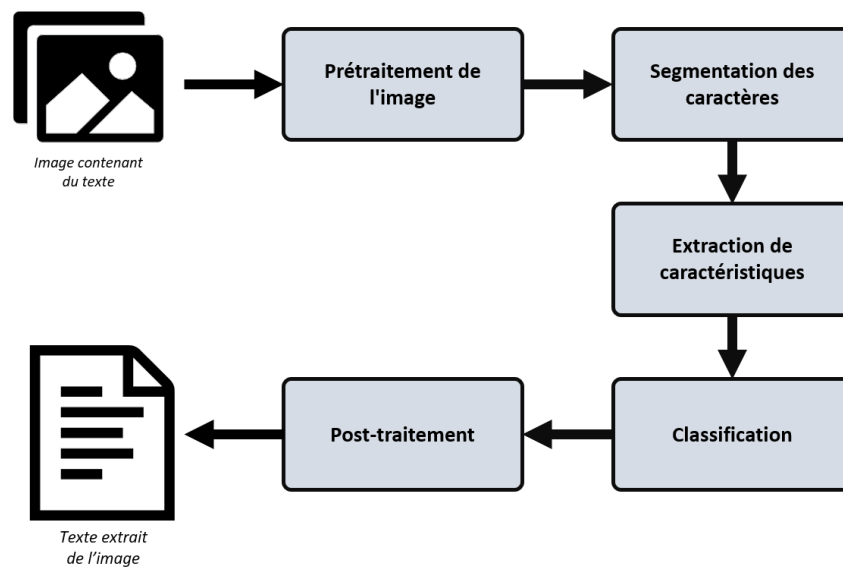


Figure 2.1: Les étapes de l'OCR classique

- **Prétraitement d'image:** Souvent ca implique de transformer l'image en noir et blanc, la correction d'inclinaison du texte filtrage de bruit et normalisation
- **Extraction de caractéristiques et correspondance de modèles:** Séparer les lignes, mots, puis caractères via Histogrammes horizontaux/verticaux ,Méthodes de projection et Analyse de contours.
- **Classification des caracteres:** Souvent implique des modeles de machine learning comme SVM,random forest,KNN...
- **Post-traitement avec des règles ou des dictionnaires:** Appliquer des regle linguistique pour verifier la coherence des mot et phrase. Les methodes OCR classiques étaient souvent adaptés à des polices ou à des formats spécifiques, ce qui les rend efficaces sur du texte propre et imprimé, mais moins adaptables nouvelles variantes(style,orientation,polices...).

2.3 OCR moderne

L'OCR moderne s'est considérablement améliorée en adoptant **l'apprentissage automatique (ML)** et en particuliers le **Deep Learning**. Au lieu de s'appuyer sur des modèles rigides et des règles faites à la main, les systèmes modernes apprennent reconnaître le texte à partir des données – ils s'entraînent sur des exemples d'images étiquetées par le texte correspondant, ce qui leur permet de développer automatiquement les fonctionnalités et les règles de décision nécessaires à la reconnaissance.

Ce qui rend l'OCR moderne beaucoup plus flexible et précis, en particulier pour les images diverses et complexes que nous rencontrons aujourd'hui.

Un exemple notable est l'utilisation d'un réseau neuronal pour lire les codes postaux pour le courrier postal – LeNet (un réseau de neurones convolutifs) de Yann LeCun a été utilisé par *U.S. Postal Service* pour la reconnaissance des chiffres manuscrits, cette exemple illustre que même les premiers CNN pouvaient surpasser les systèmes basés sur des règles. Ces systèmes basés sur le ML suivaient encore souvent les étapes des systèmes classiques (prétraitement

→ segment → classifie ensuite chaque caractère avec le modèle appris), mais l'habilité du classifieur à s'entraîner sur les données, ce qui le rend plus adaptable à de nouvelles polices et à de nouveaux styles d'écriture sans reprogrammation explicite. Cependant, ils nécessitent de grands ensembles de données pour l'entraînement à cette époque.

Deep Learning (CNN et RNN): Le réseau CRNN a révolutionné l'OCR dans les années 2010. Les systèmes OCR modernes utilisent les réseaux neuronaux convolutifs (CNN) et les réseaux neuronaux récurrents Réseaux (RNN) (souvent en combinaison) pour atteindre une grande précision.

- **CNN** excellent dans l'extraction caractéristiques visuelles à partir d'images ; dans l'OCR, les couches CNN peuvent apprendre à détecter les traits, les courbes et les motifs de lettres automatiquement à partir des données d'entraînement. Au lieu de définir manuellement le bord ou le coin à rechercher, un CNN apprend les fonctionnalités optimales pour la reconnaissance de caractères (par exemple, il peut avoir des filtres qui activent en boucle pour un « O » ou des barres transversales pour un « T », sans qu'on le lui dise explicitement).
- **RNN**, en particulier Les réseaux LSTM (Long Short-Term Memory) sont utilisés pour modéliser des dépendances séquentielles, qui sont très importantes pour le texte. Les lettres des mots ne sont pas indépendantes ; un RNN peut apprendre des séquences communes ou Utiliser les caractères précédents pour prédire le suivant. Les architectures OCR modernes combinent souvent ces éléments dans ce qu'on appelle un CRNN (Convolutional Recurrent Neural Network) : l'image (éventuellement une ligne entière de texte) est traitée par les couches CNN pour produire une séquence de vecteurs de caractéristiques, qui sont ensuite introduits dans un RNN qui génère une séquence de caractères. Cela permet au système de reconnaître un mot ou une ligne entière sous forme de séquence sans avoir besoin de segmenter parfaitement chaque personnage.

Ces modèles gèrent efficacement les textes complexes, notamment manuscrits, grâce à leur capacité à apprendre les variations d'écriture. Ils intègrent souvent un modèle de langage en sortie, qui assure la cohérence linguistique du texte, remplaçant ainsi les anciennes règles dictionnaires par un apprentissage automatique sur de larges corpus. La figure en dessous montre les étapes du traitement des images

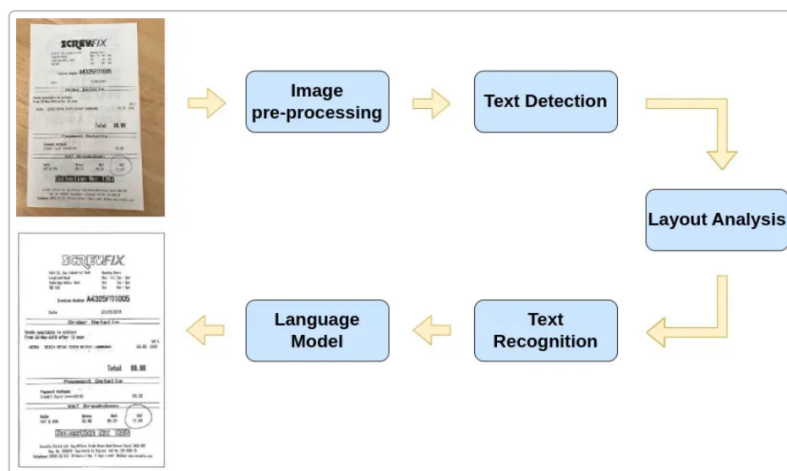


Figure 2.2: Les étapes de l'OCR moderne cas du CRNN

OCR basé sur transformer: Dans la dernière génération de technologie OCR, les chercheurs ont introduit Transformers, une architecture d'apprentissage profond popularisée à l'origine dans le traitement du langage naturel pour Tâches OCR. Les détails techniques de ces modèles dépassent le cadre de ce rapport.

Chapter 3

Application web pour l'extraction du text

3.1 Introduction

Dans le cadre de notre projet nous avons crée un site web qui propose une solution avancée de reconnaissance optique de caractères (OCR), permettant de transformer des contenus visuels — tels que des documents imprimés, manuscrits ou des images — en texte numérique exploitable.

Reposant sur des modèles CRNN, notre système est conçu pour offrir une reconnaissance précise, multilingue et robuste, capable de s'adapter à la diversité des supports et à la variabilité de l'écriture. Il s'intègre facilement dans des flux de travail professionnels et vise à améliorer la productivité, la fiabilité de données textuelles.

3.2 Conception du model CRNN

L'entraînement de modèles CRNN est très coûteux en ressources matérielles, en particulier en GPU. Pour cette raison, nous avons opté pour l'utilisation d'un modèle préentraîné, entraîné sur une large base de données de plus de +60 000 images.

Ce système repose sur deux étapes complémentaires :

- Le premier réseau, basé sur une architecture CNN (CRAFT), est chargé de détecter l'emplacement des mots dans l'image.
- Le second réseau, un modèle de type CRNN on utilise souvent (VGG ou ResNet + BLSTM + CTC), est ensuite utilisé pour reconnaître le texte dans les zones détectées.

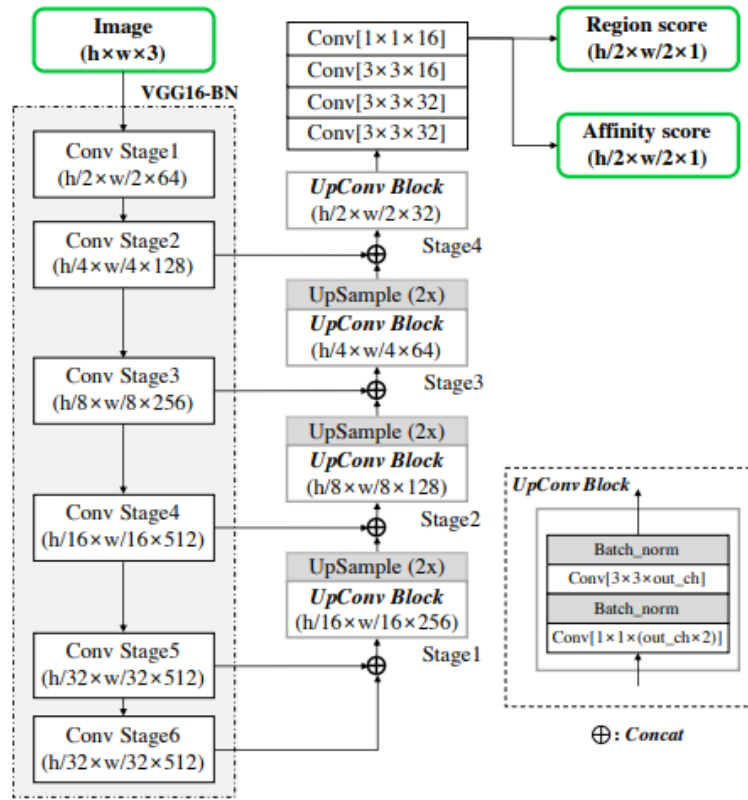


Figure 3.1: Architecture du premier réseau (CRAFT)

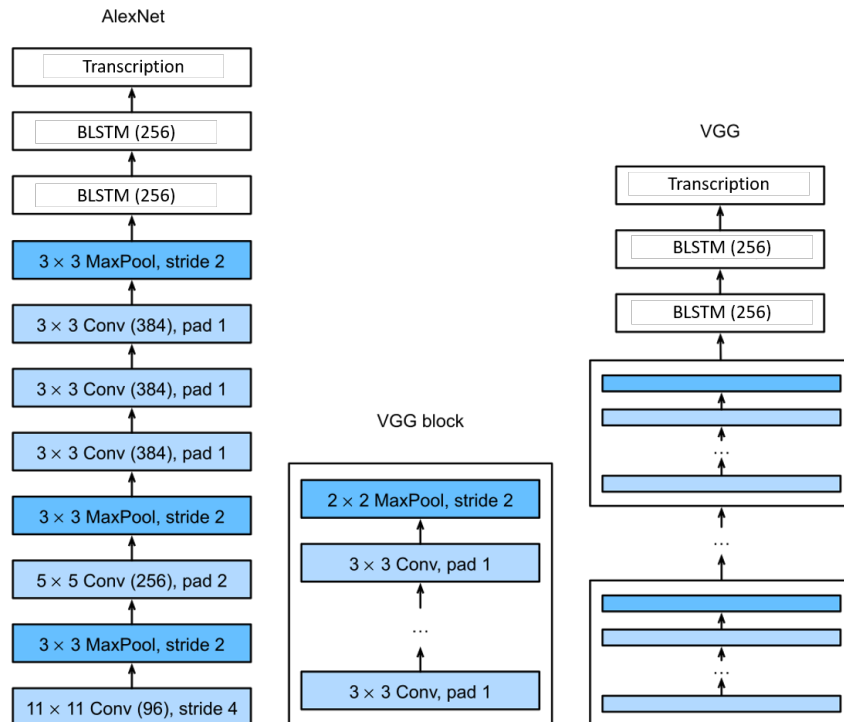


Figure 3.2: Architecture du deuxième réseau CRNN (VGG ou Resnet)

On utilise "*CTCLoss*" a cause de la nature séquentielle et non alignée des données. Elle permet d'apprendre la correspondance automatiquement, sans avoir besoin d'alignement explicite entre les séquences d'entrée (features extraits de l'image) et la sortie (texte).

Cette approche modulaire améliore à la fois la précision de détection et la qualité de la reconnaissance, tout en réduisant les coûts d'entraînement.

3.3 Test et Resultats

Afin de mesurer les performances des modèles nous avons choisi différents jeux de données.

- **IIIT 5K (IIIT5K-Words)**: contient 5 000 images collectés à partir de Google Image Search
- **Street View Text (SVT)**: 1 286 images extraits de Google Street View.
- **ICDAR03/13/15**: Jeux de données publiés lors des compétitions ICDAR (International Conference on Document Analysis and Recognition), contenant du texte dans des scènes naturelles (enseignes, panneaux, etc.)
- **CUTE80 (Curved Text in the Wild)**: Des images contenant du texte courbe et déformé (ex. logos, enseignes artistiques)

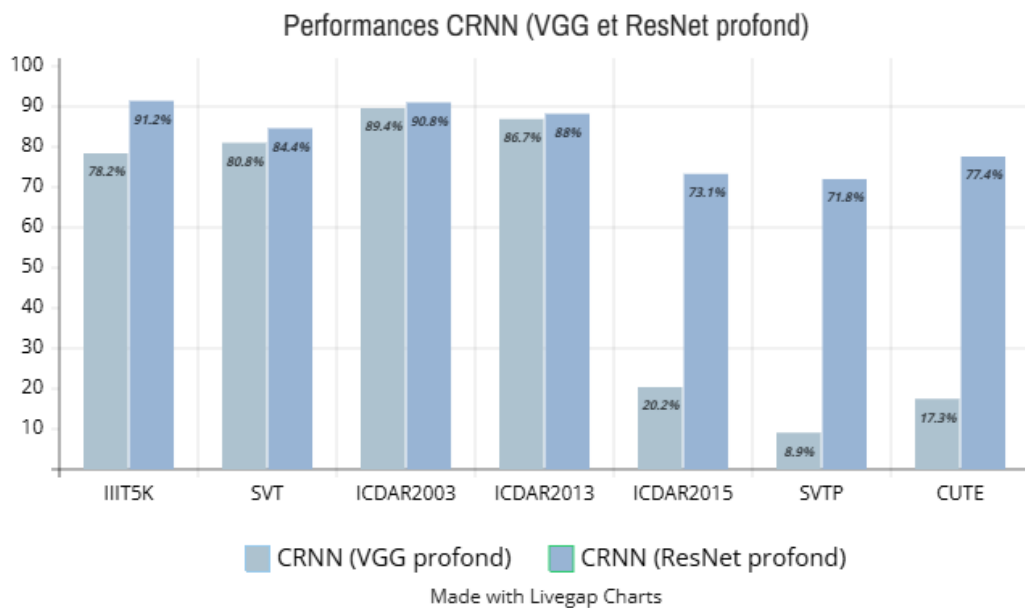


Figure 3.3: Performances de CRNN(VGG/ResNet) sur les jeux de données

Analyse et remarque

On peut remarquer que l'utilisation d'architectures profondes telles que ResNet permet un gain de 5 à 10% de précision sur certains jeux de données, tout en offrant des performances robustes et fiables, même sur des datasets complexes ou bruités.

3.4 Conception de notre site

Afin de faciliter la compréhension de notre site nous proposons ces différents diagrammes d'UML

3.4.1 digramme de cas d'utilisation

Cas d'utilisation: L'utilisateur commence par importer un document (JPG/PNG/PDF...) directement sur le site web. Ce document est alors automatiquement envoyé au modèles qui génère un texte. Le texte est ensuite restitué à l'utilisateur de manière claire et structurée.

Une fois le texte récupéré, l'utilisateur peut appliquer des filtres pour extraire des données précises tel que le numéro de téléphone, email, organisation, etc...

Le Diagramme représente un cas d'utilisation

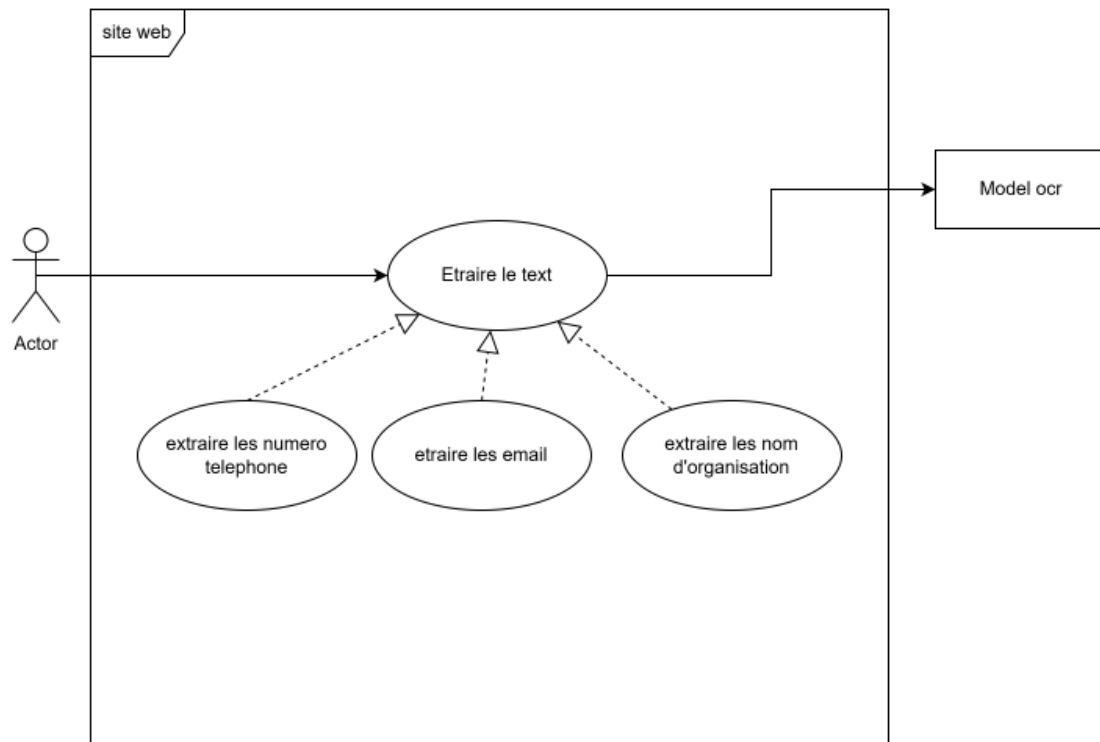


Figure 3.4: Digramme de cas d'utilisation

3.4.2 outils et Frameworks utiliser



Figure 3.5: outils de developpement

Nous avons choisi ces outils afin de faciliter l'implémentation et l'efficacité de notre modèle.

Chapter 4

Conclusion et perspectives

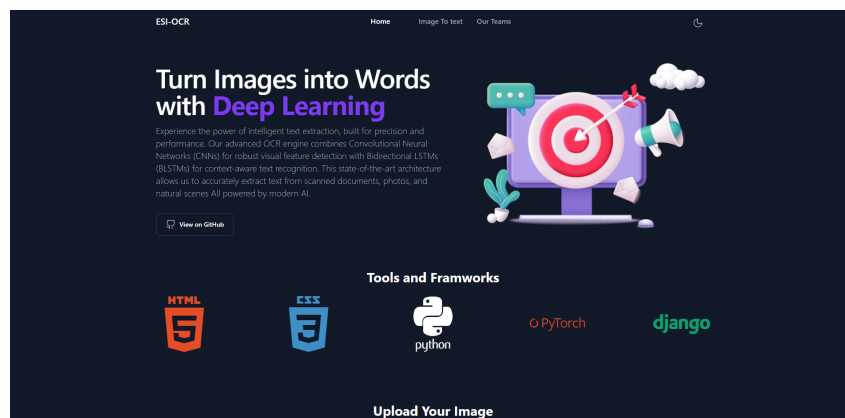


Figure 4.1: site web ESI-OCR pour l'extraction du text

4.1 Résumé du projet

L'OCR a parcouru un long chemin : d'une technologie de niche soumise à de nombreuses contraintes, elle est devenue un outil intelligent et omniprésent de notre quotidien .

Les méthodes classiques d'OCR ont posé les bases en démontrant qu'une machine pouvait lire du texte imprimé à l'aide de règles et de correspondances de formes. Cependant, ces systèmes étaient rigides, peu flexibles, et dépendaient d'un environnement très contrôlé.

Aujourd'hui, les approches modernes s'appuient sur l'apprentissage automatique et les réseaux de neurones profonds pour atteindre des niveaux de précision et de polyvalence qui permettent de lire pratiquement n'importe quel texte, quelle que soit sa source.

Grâce à l'intégration de cette technologie avancée, nous avons conçu une plateforme web capable de traiter automatiquement différents types de documents et d'en extraire le contenu textuel de manière précise, rapide et efficace. Ce système repose sur des modèles de deep learning, combinant réseaux convolutifs (CNN) et réseaux récurrents (BLSTM), offrant ainsi une solution robuste pour la reconnaissance de texte dans des contextes variés.

4.2 Travaux futurs

- Optimisation des performances du modèle.
- Intégration d'une API pour les applications externes.

- Couplage de l'OCR avec des techniques de traitement automatique du langage naturel (NLP).

Chapter 5

Annexes

Bibliographie

- Shaip. What is OCR (Optical Character Recognition)? – Definition and examples of OCR technology
- An End-to-End Trainable Neural Network for Image-based Sequence Recognition and Its Application to Scene Text Recognition- *Baoguang Shi, Xiang Bai and Cong Yao School of Electronic Information and Communications Huazhong University of Science and Technology, Wuhan, China*
- Character Region Awareness for Text Detection- *Youngmin Baek, Bado Lee, Dongyoon Han, Sangdoo Yun, and Hwalsuk Lee Clova AI Research, NAVER Corp.*
- Scene Text Recognition Based on Improved CRNN - *Shmuel Tomi Klein*