Université de Carthage

Faculté des Sciences de Bizerte

Département informatique

Extraction du texte à partir d'une image

Réalisé par :

Sidi Brahim Mohamed (mohamedsidibrahim11@gmail.com)

Soueidatt Mohamed (mohamedabdallahi388@gmail.com)

Mohamed Abderrahmane Ebnou (bouyaebnou@gmail.com)

Sidi M'Hamed Adel (adelsidimed@gmail.com)

Encadrante :

Dr. Neili Sameh

Année Universitaire 2021-2022

Table des matières :

Table des Figures :

Abstract :

Les données textuelles présentes dans les images contiennent des informations utiles pour l'annotation, l'indexation et la structuration automatiques des images.
L'extraction de ces informations implique la détection, la localisation, l'extraction et reconnaissance du texte d'une image donnée.
Cependant, les variations de texte dues à des différences de taille, de style, d'orientation et de l'alignement, ainsi que le faible contraste de l'image et l'arrière-plan complexe posent le problème de l'extraction automatique de texte extrêmement difficile.
Alors que des enquêtes complètes sur des problèmes connexes tels que la détection de visages, l'analyse de documents et l'indexation d'images et de vidéos peut être trouvée, le problème de l'extraction d'informations textuelles n'est pas bien étudié.

1. Introduction générale

1.1- **Introduction** :

Le volume de la base de données multimédia a augmenté de façon exponentielle en raison des progrès technologiques dans le domaine des processeurs informatiques et des périphériques de stockage. Malheureusement, ces grands référentiels multimédias ne sont pas indexés et ne sont accessibles que par balayage séquentiel de l'intégralité des archives multimédias. Naviguer ou parcourir une grande base de données multimédia est fastidieux et prend du temps. Les moteurs de recherche Web populaires tels que Google, Yahoo et AltaVista offrent aux utilisateurs un modèle de recherche basé sur le contenu afin d'accéder aux pages Web et multimédias du World Wide Web. Mais dans ce moteur de recherche textuel typique, les images et les vidéos sont annotées manuellement en identifiant un nombre limité de mots-clés qui décrivent leurs informations visuelles et leur contenu. Cependant, pour la récupération d'images et de vidéos, ce n'est pas une solution efficace. Par conséquent, besoin d'un système de navigation et de navigation efficace et réel basé sur un contenu ou un modèle, à travers lequel les utilisateurs pourront accéder au matériel multimédia d'intérêt.

Comme indiqué précédemment, dans les moteurs de recherche textuels, les images et les vidéos sont annotées manuellement en identifiant un nombre limité de mots-clés décrivant leurs informations visuelles et leur contenu. Certaines images peuvent être liées différemment par différentes personnes. Deuxièmement, il n'est pas toujours possible d'identifier tous les mots-clés souhaités par des descripteurs de texte manuels. Troisièmement, l'examen séquentiel de l'intégralité du contenu vidéo pour de grandes archives multimédias en pleine croissance est nécessaire pour identifier les mots-clés. Ce processus d'indexation manuelle du contenu des images par listes de documents sera de plus en plus fastidieux et chronophage. Cette méthode d'indexation manuelle n'est pas rentable et l'efficacité de l'indexation devient fortement dépendante de la qualité de la main-d'œuvre et enfin, elle dépend de la langue.

Le texte a des caractéristiques visuelles compactes et distinctives, c'est-à-dire un ensemble de symboles avec des caractéristiques géométriques et morphologiques distinctes. Deuxièmement, le texte peut être d'une police, d'une couleur ou d'une langue différente, il est généralement étroitement lié à son contenu sémantique et

au maintien d'un motif spécifique dans l'image. Par conséquent, le texte est souvent considéré comme un candidat solide pour une utilisation en tant que fonctionnalité d'indexation sémantique de haut niveau et de récupération basée sur le contenu. Le texte est utile pour effectuer une analyse de texte comme dans la diffusion, pour afficher le nom du programme, le nom de l'ancre, les introductions du programme, les annonces spéciales. Dans le nom d'un produit publicitaire, le nom des sociétés vendant les produits est affiché. Dans les prévisions météorologiques, la température et l'humidité de différents endroits sont affichées. Dans d'autres cas, les objets et les emplacements peuvent être identifiés par du texte à partir d'annotations textuelles implicites et explicites, comme dans un événement sportif, les joueurs peuvent être identifiés par leur nom et leur numéro sur leurs maillots, les véhicules peuvent être repérés par leur plaque d'immatriculation, une gare ou des rues ou les magasins peuvent être localisés par leurs panneaux d'affichage ou palissades.

1.2- **Problématique**

À quel point l'extraction de données est importante pour les organisations, commençons maintenant par ce qu'est l'extraction de données. En termes simples, l'extraction de données est le processus d'extraction de données capturées dans des sources semi-structurées et non structurées, telles que des e-mails, des documents PDF, des formulaires PDF, des fichiers texte, des réseaux sociaux, des codes-barres et des images. Comment se fait l'extraction de données non structurées ? Un outil d'extraction de données de niveau entreprise rend les données commerciales entrantes provenant de sources non structurées ou semi-structurées utilisables pour l'analyse et la création de rapports de données. Ils sont également appelés outil d'extraction de données.

Par exemple, une entreprise peut souhaiter extraire divers points de données, tels que le nom d'employeur, date de naissance et la poste. Ces accords se présentent généralement sous la forme de PDF non structurés, un mélange de texte libre et de données tabulaires. Alors la question devient comment extraire automatiquement les données d'un PDF ? Il est difficile d'extraire des informations à partir de données, en particulier de fichiers PDF, car les ensembles de données non structurés sont lisibles par l'homme et les machines ont besoin d'informations structurées pour les traiter numériquement en vue d'analyses supplémentaires ou d'une intégration avec d'autres applications informatiques. De plus, l'extraction de données non structurées doit être automatisée pour des informations plus rapides et efficaces. Par conséquent, des programmes d'extraction de données automatisés sont nécessaires pour rationaliser l'ensemble du processus du début à la fin.

1.3- Solution proposée

Ce projet consiste à concevoir et développer un modèle permettant d'extraire les informations utiles depuis une image en utilisant OCR.
L'utilisateur peut prendre une photo carte d'identité et extraire les données à l'aide d'un modèle machine learning. Le but de ce modèle est la reconnaissance automatique des caractères, la numérisation de l'image capturée, ainsi que la copie des données sous forme numérique.

Cela élimine le besoin de conserver les cartes d'identité et de les envoyer aux équipes de saisie des données de l'entreprise, et réduit également les erreurs. L'application permet aux entreprises d'économiser plus de temps de réclamation par rapport à la gestion manuelle des dépenses.
De plus, les entreprises qui embauchent des travailleurs peuvent bénéficier de l'application. Comment faire ? En automatisant le processus de recrutement. Avec notre application, les cartes d'identité peuvent être numérisées et les données extraites automatiquement. Après cela, les informations numériques sont automatiquement prêtes à être importées. Cela élimine le besoin de traitement personnel manuel.
Depuis ses humbles débuts en tant qu'aide pour les malvoyants, l'application est devenue un outil puissant. De la vie quotidienne aux opérations commerciales, notre application a été adaptée pour simplifier notre façon de vivre et de travailler.

2. Contexte générale du projet

2.1 Introduction

Extraire le texte d'une image c'est la conversion informatisée d'une image représentant du texte vers un format de données textuel.

L'extraction de texte à partir d'une image est un problème difficile à cause de l'image contient du texte en raison de la taille, du style, orientation, alignement, faible contraste, bruit et complexe structure de fond. Ce texte extrait ne contient que du noir texte sur fond blanc, c'est-à-dire qu'il peut être reconnu par n'importe quel système de reconnaissance. Extraire du texte d'une image ou d'une vidéo inclut dans différentes applications comme le traitement de documents, indexation d'images, résumé du contenu vidéo, récupération vidéo, compréhension vidéo, etc.

La détection de texte fait référence à la détermination de la présence de texte dans une image d'entrée donnée, fait en exploitant le discriminer les propriétés des caractères de texte tels que la verticale la densité des bords, la texture ou la variance de l'orientation des bords.

Une fois la partie reconnaissance de texte terminée, nous pouvons passer à l'extraction de texte. À la fin de cette étape, nous avons toujours une image non modifiable avec du texte plutôt que le texte lui-même. Pour résoudre ce problème, l'étape suivante consiste à extraire du texte d'une image. Juste après la reconnaissance de texte, le processus de localisation est effectué. Toutes les fonctionnalités liées à une image particulière sont rassemblées.

Extraction de texte : comment ça marche ?

L'extraction de texte, également connue sous le nom d'extraction de mots clés, repose sur l'apprentissage automatique pour analyser automatiquement le texte et extraire des mots et des phrases pertinents ou basiques à partir de données non structurées telles que des articles de presse, des sondages et des plaintes de support client.

Figure 1 :Numérisation de l'image

Source de l'image : https://addepto.com/wp-content/uploads/2019/12/Projekt-bez-tytulu-27-1.png

Les méthodes d'extraction et d'amélioration de texte sont appliquées à l'aide d'algorithmes d'apprentissage automatique. Et enfin, le texte extrait est collecté à partir de l'image et transféré vers l'application donnée ou un type de fichier spécifique. Il existe de nombreux types d'algorithmes et de techniques d'extraction de texte qui sont utilisés à diverses fins. Par conséquent, nous pouvons les diviser en cinq méthodes principales.

- **MÉTHODE RÉGIONALE**
  Cette méthode d'extraction de texte utilise une fenêtre coulissante pour détecter le texte de tout type d'image. Cette approche repose sur plusieurs facteurs, tels que les caractéristiques de couleur, d'arête, de forme, de contour et de géométrie.

- **MÉTHODE BASÉE SUR LA TEXTURE**
  Cette méthode utilise différents types de texture et ses propriétés pour extraire le texte d'une image.

- **TECHNIQUE HYBRIDE**
  C'est la combinaison des deux techniques précédentes. Premièrement, l'approche basée sur la région est utilisée pour détecter un texte. Ensuite, avec l'utilisation de la méthode basée sur la texture, toutes les caractéristiques sont extraites de la zone de texte.

- **MÉTHODE BASÉE SUR LES BORDS**
  Comme son nom l'indique, cette méthode est basée sur la détection des bords de chaque lettre et de chaque chiffre. Cette méthode est

utilisée pour développer un contraste de haut niveau entre le texte et l'arrière-plan.

- **MÉTHODE FONDÉE SUR LA MORPHOLOGIE**
  Cette méthode est utilisée pour extraire toutes les caractéristiques liées au texte de l'image traitée.

Figure 2 : Méthode fondée sur la morphologie

Source de l'mage : https://cloud.google.com/vision/docs/images/sign_text.png

2.2 Objectif

Chaque jour, 2.5 quintillions d'octets de données sont générés par les internautes. Un fait fascinant est qu'en 2021, chaque personne a généré 1,7 gigaoctet en une seule seconde. Commentaires sur les réseaux sociaux, critiques de produits, e-mails, articles de blog, requêtes de recherche, discussions, etc. Mais la question est de savoir comment l'extraction de texte à partir d'images peut aider en particulier entreprise à devenir plus efficace et à tirer pleinement parti du potentiel des données ?

- **Surveillance des médias sociaux**

  L’entreprise peut utiliser l'extraction de texte à partir d'images pour suivre les conversations sur les réseaux sociaux afin de mieux comprendre les clients, d'améliorer les produits ou de prendre des mesures rapides pour éviter une crise de relations publiques. L'extraction de texte à partir d'images peut offrir des exemples spécifiques de ce que les gens sur les réseaux sociaux disent de votre entreprise. De plus, vous pouvez découvrir des mots-clés et suivre les tendances grâce à l'extraction de texte à partir d'une image.

- **Service client avec extraction de texte**

  Un service client de qualité peut donner à votre entreprise un avantage concurrentiel. Après tout, lorsqu'il s'agit d'acheter quelque chose, 64% des clients préfèrent la qualité du service client au prix. En d'autres termes, l'extraction de texte à partir d'images permet au personnel du service client d'automatiser le processus de marquage des tickets, ce qui permet d'économiser des dizaines d'heures qui pourraient être consacrées à la résolution de problèmes réels. C'est donc la clé de la satisfaction client.

Figure 3 : Service client avec extraction du texte

Source de l’mage : https://addepto.com/wp-content/uploads/2019/12/Projekt-bez-tytulu-30-1.png

- **Intelligence d'affaires et extraction de texte à partir d'images**

  L'extraction de texte à partir d'images peut également être efficace dans les applications de business intelligence (BI) telles que les études de marché et l'analyse de la concurrence. Nous pouvons également obtenir des informations de diverses sources, notamment des critiques de produits et des réseaux sociaux, et participer à des discussions sur des sujets d'intérêt. De plus, vous pouvez comparer vos avis sur les produits avec ceux de vos concurrents en utilisant l'extraction de texte à partir d'images et d'autres outils d'analyse de texte. Cela aide à obtenir des informations qui vous aideront à prendre des décisions basées sur les données pour améliorer votre produit ou service

Figure 4 : Intelligence d'affaires et extraction du texte

Source de l'image : https://addepto.com/wp-content/uploads/2019/12/Projekt-bez-tytulu-31-1-1.png

❖ **Exemple du monde réel :**

Une autre application répandue de l'extraction du texte à partir d'e image est la reconnaissance des plaques d'immatriculation des voitures. Cela a

également de nombreuses applications possibles, des bases de données policières (données obtenues à partir des radars) aux parkings privés qui ouvrent la barrière après vérification d'une plaque d'immatriculation.
Lorsqu'un propriétaire de voiture donné veut quitter le parking, il doit se rendre à l'horodateur et choisir sa plaque d'immatriculation dans la liste. Juste après le paiement, le logiciel de gestion des barrières reçoit un signal indiquant que la voiture donnée peut quitter le parking. Lorsque la voiture s'approche de la barrière, sa plaque d'immatriculation est à nouveau numérisée et si le numéro numérisé correspond à la liste des numéros déjà payés, la barrière s'ouvre.

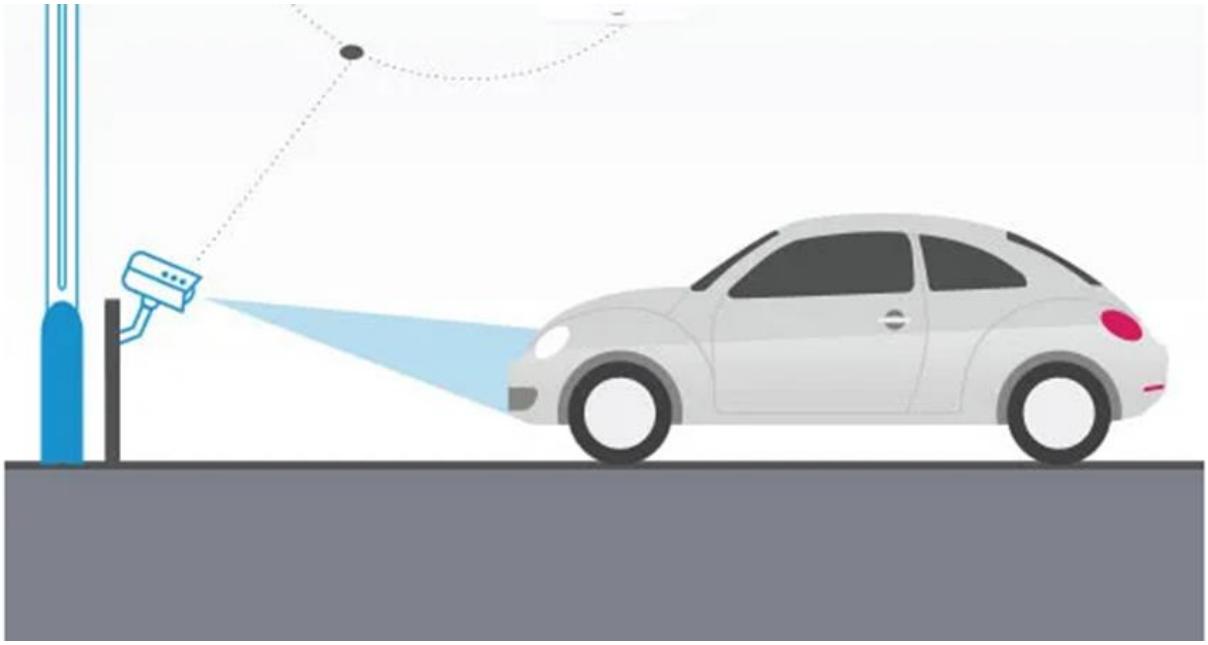

Figure 5 : Exemple d'extraction du texte à partir d'une image du monde réel

Source de l'image : https://parklio.com/assets/img/parking-solutions/lpr/parklio-anpr-parking-license-plate-recognition-lpr.webp

Pour résumer, il existe actuellement une demande croissante d'extraction de texte à partir d'images. De nombreuses techniques d'extraction pour récupérer les informations pertinentes ont été développées. Ainsi, pour utiliser avec succès l'extraction de texte à partir d'une image dans les entreprises, ils doivent identifier les objectifs commerciaux, analyser les données accessibles à partir de jeux de

données open source et privés. De plus, ils doivent décider si des mesures de sécurité supplémentaires sont nécessaires pour confirmer une défaillance de l'exactitude du mécanisme OCR.
L'extraction de texte à partir d'images à l'aide de l'apprentissage automatique est bénéfique pour l'entreprise et accélère son travail.

2.3 Système existant

L'extraction de texte est un ensemble de tâches de vision par ordinateur qui convertissent les images en texte lisible par machine. Il prend des images de documents, de factures et de reçus, y trouve du texte et le convertit dans un format que les machines peuvent mieux traiter.
Les convertisseurs d'image en texte, souvent intégrés en tant que sous-fonctionnalité dans les programmes de traitement d'images/documents, offrent un moyen pratique d'extraire du texte à partir d'images.
Des outils tels que Snagit et OneNote, entre autres, exploitent les capacités OCR de base pour extraire le texte des images. Et les convertisseurs en ligne comme Workbench ou img2text extraient également le texte des images avec OCR.
Bien que ces outils fassent du bon travail, le texte/les données extraits sont souvent présentés de manière non structurée, ce qui entraîne de nombreux post-traitements.

2.4 Système proposé

2.4.1 Introduction

Les entreprises de dématérialisation souhaitent déterminer si des documents semi-structurés tels qu'une carte d'identité, un ticket de train, une facture de téléphone, etc. sont présents sur une page numérisée. Ce que nous appelons "image" est une

page numérisée. Une image peut contenir un ou plusieurs documents de différentes natures. Si un document est présent dans une image à analyser, il doit être localisé précisément afin de pouvoir exploiter les informations qu'il contient, telles que le nom, le prénom, etc.
Toute organisation peut utiliser l'apprentissage en profondeur pour automatiser l'extraction d'informations sur les cartes d'identité, la saisie de données et les procédures de révision afin d'obtenir une plus grande efficacité et de réduire les coûts.
Le prototype qu'on a choisi est l'extraction du nom et l'identifiant à partir d'une carte d'identité.

2.4.2 Avantages de système proposé

- **Extraction d'informations :**
  Nous pouvons capturer toutes les informations fournies sur la carte d'identité et transmettre ces données en tant que source unique pour une utilisation ultérieure. Toutes les informations extraites de la carte d'identité capturée seront dans un format texte/numérique simple. Cela aide à maintenir les données de manière organisée et facilite toute sorte de processus de vérification ou d'enregistrement.

- **Plus de rapidité et d'efficacité :**
  La numérisation des cartes d'identité peut faire économiser beaucoup de temps et d'argent aux entreprises et aux organisations. Il suffit de quelques secondes pour numériser simplement une carte d'identité et en récupérer toutes les données.

- **Données sans erreur :**
  Avec les progrès des technologies et de la puissance de calcul, les machines sont désormais capables de capturer les données sans beaucoup d'erreurs. Les risques d'erreur humaine peuvent être réduits en automatisant les tâches

répétitives et en permettant aux humains d'examiner les informations des documents sur les étapes finales du pipeline d'extraction d'informations.

2.4.3 Architectures du système proposé

Considérons maintenant le problème de l'extraction des informations de la carte d'identité, imaginons que tous les nœuds de l'image (que nous utiliserons en entrée) sont associés à une étiquette. L'extraction des données qu'on a besoin se fait par deux méthodes soit détecter les informations précises et après on extraire les informations détectées ou bien extraire tous les caractères qui sont dedans l'image et filtre seulement les informations souhaiter à l'aide de l'expression régulière.

2.4.4 Type d'image

Le type de l'image est l'image de carte d'identité.
L'image peut être bruité et incliner, de toute façon elle va passer par des prétraitements pour éviter que la précision de notre sortie puisse baisser.
Cela inclut le redimensionnement, la binarisation, la suppression du bruit, le redressement, etc.

2.5 Bibliothèques utilisées

- **Pytesseract :** est un wrapper pour Tesseract-OCR Engine. Il est également utile en tant que script d'invocation autonome pour tesseract, car il peut lire tous les types d'images pris en charge par les bibliothèques d'imagerie Pillow et Leptonica, y compris jpeg, png, gif, bmp, tiff et autres.

- **OpenCV (Open source computer vision) :** est une bibliothèque de fonctions de programmation principalement destinées à la vision par ordinateur en temps réel. OpenCV en python permet de traiter une image

et d'appliquer diverses fonctions telles que le redimensionnement d'une image, la manipulation de pixels, la détection d'objets, etc.

3. La reconnaissance optique de caractères

3.1 Introduction

La reconnaissance optique de caractères (OCR) est l'identification électronique et l'encodage numérique de texte dactylographié ou imprimé au moyen d'un scanner optique et d'un logiciel spécialisé. L'utilisation d'un logiciel OCR permet à un ordinateur de lire des images statiques de texte et de les convertir en données modifiables et consultables. L'OCR implique généralement trois étapes : ouvrir et/ou numériser un document dans le logiciel OCR, reconnaître le document dans le logiciel OCR, puis enregistrer le document produit par OCR dans le format de votre choix.

L'OCR peut être utilisé pour une variété d'applications. Dans les milieux universitaires, il est souvent utile pour les projets d'exploration de texte et/ou de données, ainsi que pour les comparaisons textuelles. IL est également un outil important pour créer des documents accessibles, en particulier des PDF, pour les personnes aveugles et malvoyantes.

3.2 Pipline

Tout pipeline OCR d'apprentissage automatique typique suit les étapes suivantes :

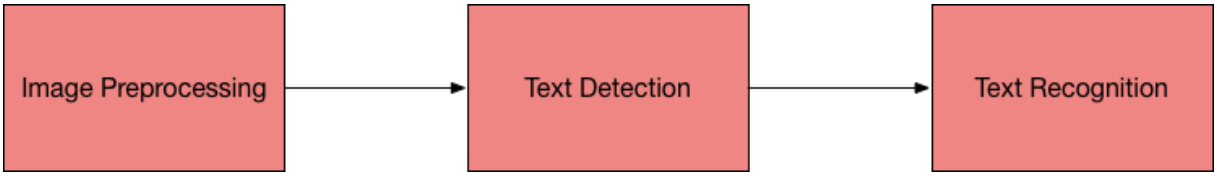

Figure 6 OCR Pipline

Source de l'image : https://nanonets.com/blog/content/images/2019/08/image-12.png

3.2.1 Pré-traitement de l’image

Pour éviter toutes les façons dont la précision de notre sortie tesseract peut baisser, nous devons assurer que l'image est correctement prétraitée .
Cela inclut le redimensionnement, la binarisation, la suppression du bruit, le redressement, etc.
Voici les fonctions les plus utilisées pour le pré-traitement d’images dans les OCR :

a. **Normalisation** : **La normalisation de l’image est** un processus dans lequel nous modifions la plage de valeurs d'intensité des pixels pour rendre l'image plus familière ou normale aux sens, d'où le terme de normalisation. La normalisation de l'image est souvent utilisée pour **augmenter le contraste,** ce qui aide à améliorer l'extraction des caractéristiques ou la segmentation de l'image.

   Souvent, la normalisation de l'image est utilisée pour supprimer le bruit de l'image (données). Avec l'aide de la normalisation de l'image, nous pouvons supprimer le bruit haute fréquence et le bruit très faible de l'image, ce qui est vraiment utile.

   Pour ce processus on a utilisé la méthode normalize( ) dans la bibliothèque OpenCV.

b. **Redimensionnement** : La *mise à l'échelle*, ou simplement le *redimensionnement*, consiste à augmenter ou à diminuer la taille d'une image en termes de largeur et de hauteur.
   Lors du redimensionnement d'une image, il est important de garder à l’esprit le *rapport hauteur/largeur*, qui est le rapport entre la largeur d'une image et sa hauteur. Ignorer le rapport hauteur/largeur peut conduire à des images redimensionnées qui semblent compressées et déformées.
   Généralement *diminuer* la taille d'une image plutôt que l’augmentation (des exceptions sont applicables, bien sûr). En diminuant la taille de l'image, nous avons moins de pixels à traiter (sans parler de moins de « bruit » à gérer), ce qui conduit à des algorithmes de traitement d'image plus rapides et plus précis.
   Gardez à l'esprit que si les images haute résolution sont visuellement attrayantes pour les yeux humains, elles nuisent aux pipelines de vision par ordinateur et de traitement d'images :

- Par définition, plus l'image est grande, plus il y a de données, et donc plus il faut de temps aux algorithmes pour traiter les données
- Les images haute résolution sont très détaillées - mais du point de vue de la vision par ordinateur/du traitement d'images, nous sommes plus intéressés par les composants structurels des images, pas tant par les détails très fins
- Les images à grande résolution sont presque toujours sous-échantillonnées pour aider les systèmes de traitement d'images à fonctionner plus rapidement et à être plus précis

Pour ce processus on a utilisé la méthode resize( ) dans la bibliothèque OpenCV.

c. **Conversion en niveau de gris :** Une image en niveaux de gris est une image dans laquelle un seul pixel représente la quantité de lumière ou ne contient que des informations sur l'intensité lumineuse. Il s'agit d'une image unidimensionnelle et présente uniquement différentes nuances de gris.
Comme les images en niveaux de gris sont unidimensionnelles, elles sont utilisées pour réduire la complexité d'apprentissage des modèles dans divers problèmes et dans des algorithmes.
La fonction cv2 mettra à disposition les fonctionnalités nécessaires pour lire l'image originale et la convertir en niveaux de gris.
Pour lire l'image originale, il suffit d'appeler la « imread ». Fonction du CV2 module, le passage en entrée le chemin d'accès à l'image, sous forme de chaîne.
image = cv2.imread('img.jpg').
Comme première entrée, cette fonction reçoit l'image originale. En deuxième entrée, il reçoit le code de conversion de l'espace colorimétrique. Puisque nous voulons convertir notre image originale de l'espace colorimétrique BVR en gris, nous utilisons le code « COLOR_BGR2GRAY ».

d. **Suppression du bruit :** L'objectif principal de l'étape de suppression du bruit est de lisser l'image en supprimant les petits points/patchs qui ont une intensité élevée par rapport au reste de l'image. La suppression du bruit peut être effectuée pour les images colorées et binaires.

Une façon d'effectuer la suppression du bruit en utilisant la fonction «medianBlur( )»

e. **Binarisation : La** binarisation d'une image consiste à convertir l'image colorée en format noir et blanc. La méthode la plus simple consiste à calculer une valeur seuil et à convertir en blanc tous les pixels dont la valeur est supérieure à la valeur seuil et à convertir le reste des pixels en noir. Cette étape aide le moteur à bien comprendre les données. La binarisation d'une image peut également aider à réduire la taille de l'entrée.

f. **Correction de l'inclinaison :** Lors de la numérisation d'un document, il peut arriver qu'il soit légèrement incliné (image alignée à un certain angle par rapport à l'horizontale). Lors de l'extraction des informations de l'image numérisée, la détection et la correction de l'inclinaison sont cruciales.
Plusieurs techniques sont utilisées pour la correction de l'obliquité.

- Méthode du profil de projection
- Méthode de transformation de Hough
- Méthode de la ligne supérieure
- Méthode de la ligne de balayage

Cependant, la méthode du profil de projection est la plus simple, la plus facile et la plus utilisée pour déterminer le ***skew*** dans les documents.
Dans cette méthode, nous prenons d'abord l'image binaire, puis la projeter horizontalement (en prenant la somme des pixels le long des lignes de la matrice de l'image) pour obtenir un histogramme des pixels le long de la hauteur de l'image, c'est-à-dire le nombre de pixels de premier plan pour chaque ligne.
Maintenant, l'image est tournée à différents angles (à un petit intervalle d'angles appelé Delta) et la différence entre les pics sera calculée (la variance peut également être utilisée comme une des mesures). L'angle pour lequel la différence maximale entre les pics (ou la variance) est trouvée, cet angle correspondant sera l'angle d'inclinaison de l'image.
Après avoir trouvé l'angle d'inclinaison, nous pouvons corriger l'inclinaison en faisant tourner l'image d'un angle égal à l'angle d'inclinaison dans la direction opposée à l'inclinaison.

3.2.2 Détection de texte

Détection de texte est le processus de la localisation où un texte d'image est.

Vous pouvez considérer la détection de texte comme une forme spécialisée de détection d'objets.

Dans la détection de texte, notre objectif est de calculer automatiquement les cadres de délimitation pour chaque région de texte dans une image.

Tesseract suppose que l'image du texte d'entrée est assez propre. Malheureusement, de nombreuses images d'entrée contiendront une pléthore d'objets et pas seulement un texte prétraité propre. Par conséquent, il devient impératif d'avoir un bon système de détection de texte capable de détecter du texte qui peut ensuite être facilement extrait.

Il existe plusieurs méthodes pour la détection de texte :

- Manière traditionnelle d'utiliser OpenCV
- Manière contemporaine d'utiliser les modèles d'apprentissage profond
- Construire votre propre modèle personnalisé

3.2.2.1 Détection de texte avec OpenCV

La détection de texte à l'aide d'OpenCV est la façon classique de faire les choses. On peut appliquer diverses méthodes dans le pré-traitement de l'image enfin de le bien nettoyer (voir la phase de pré-traitement) Une fois que ceci est faite, on peut utiliser la détection de contours OpenCV pour détecter les contours afin d'extraire des morceaux de données.
Enfin, on applique la reconnaissance de texte sur les contours qu'on a pour prédire le texte.
Les contours ne détectent pas le texte à chaque fois.

Mais, tout de même, faire de la détection de texte avec OpenCV est une tâche fastidieuse nécessitant beaucoup de jouer avec les paramètres. De plus, il ne réussit pas bien en termes de généralisation.

3.2.2.2 Modèle contemporain d'apprentissage en profond – EAST

EAST, ou Efficient and Accurate Scene Text Detector, est un modèle d'apprentissage en profondeur permettant de détecter du texte à partir d'images de scènes naturelles. Il est assez rapide et précis car il est capable de détecter des images 720p à 13,2 ips avec un F-score de 0,7820.

Le modèle se compose d'un réseau entièrement convolutif et d'une étape de suppression non maximale pour prédire un mot ou des lignes de texte. Le modèle, cependant, n'inclut pas certaines étapes intermédiaires telles que la proposition de candidat, la formation de régions de texte et la partition de mots qui étaient impliquées dans d'autres modèles précédents, ce qui permet un modèle optimisé.
EAST a un réseau en forme de U. La première partie du réseau est constituée de couches convolutives formées sur l'ensemble de données ImageNet. La partie suivante est la branche de fusion de caractéristiques qui concatène la carte de caractéristiques actuelle avec la carte de caractéristiques non regroupée de l'étape précédente.

Ceci est suivi par des couches convolutives pour réduire le calcul et produire des cartes de caractéristiques en sortie. Enfin, en utilisant une couche convolutive, la sortie est une carte de score montrant la présence de texte et une carte géométrique qui est soit une boîte pivotée, soit un quadrilatère qui couvre le texte. Cela peut être visuellement compris à partir de l'image de l'architecture qui a été incluse dans le document de recherche :

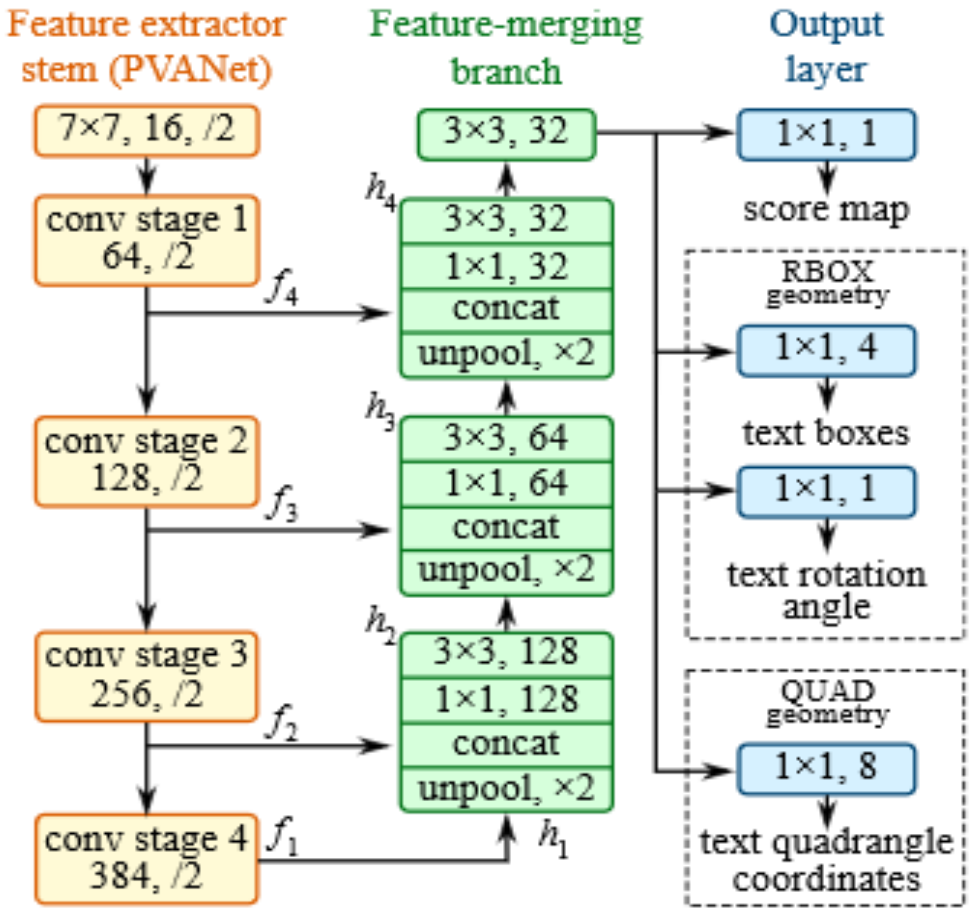

Figure 7:La structure du réseau de détection de texte EAST entièrement convolutif

Source de l'image : https://cdn.analyticsvidhya.com/wpcontent/uploads/2020/05/OCR_9.png

3.2.2.3 Modèle personnalisé utilisant l'API d'objets TenserFlow pour la détection de texte

La dernière méthode pour créer un détecteur de texte consiste à utiliser un modèle de détecteur de texte personnalisé à l'aide de l' API d'objet TensorFlow . Il s'agit d'un Framework open source utilisé pour créer des modèles d'apprentissage en profondeur pour les tâches de détection d'objets.
Pour construire un détecteur de texte personnalisé, nous aurons évidemment besoin d'un ensemble de données de quelques images, Ensuite, nous devons annoter ces images afin que le modèle puisse savoir où se trouve l'objet cible et tout apprendre à son sujet.
Enfin, nous pouvons choisir parmi l'un des modèles pré-entraînés, en fonction du compromis entre performances et vitesse, du zoo de modèles de détection de TensorFlow .
Maintenant, l'entraînement nécessite des calculs, c-à-d nous avons besoin d'un environnement de travail avec GPU ou TPU comme type d'exécution pour bien atteindre la performance et trouver le bien de cette méthode.
Après avoir toute ces méthodes c'est évidemment de choisir la méthode EAST car cette ci est la meilleure en termes de performance et du cout.

L'image ci-dessous fournie par les auteurs dans leur article comparant le modèle EAST avec d'autres modèles précédents :

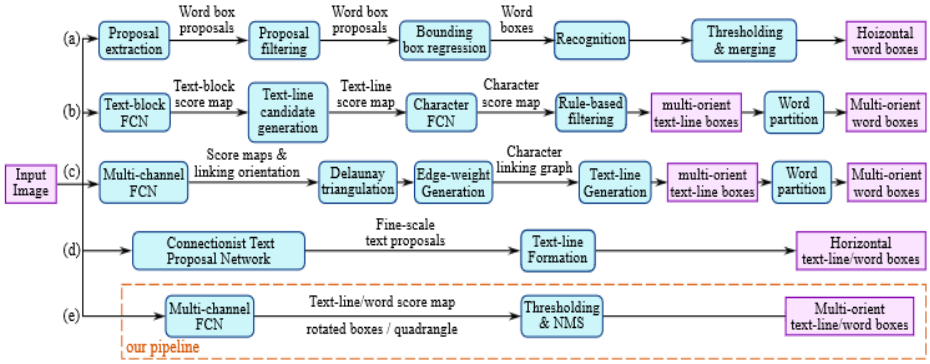

Figure 2. Comparison of pipelines of several recent works on scene text detection: (a) Horizontal word detection and recognition pipeline proposed by Jaderberg *et al.* [12]; (b) Multi-orient text detection pipeline proposed by Zhang *et al.* [48]; (c) Multi-orient text detection pipeline proposed by Yao *et al.* [41]; (d) Horizontal text detection using CTPN, proposed by Tian *et al.* [34]; (e) Our pipeline, which eliminates most intermediate steps, consists of only two stages and is much simpler than previous solutions.

Figure 8:comparaison de pipelines sur détection de texte de scène

Source de l'image : https://cdn.analyticsvidhya.com/wp-content/uploads/2020/05/OCR_8.png

Exemple d’utilisation du méthode EAST

Image avant la détection du texte :

Figure 9:Carte d'identification

Source de l’image : https://5.imimg.com/data5/EH/KI/MY-7724195/plastic-id-card-250x250.jpg

Après le passage par méthode EAST :

Figure 10:Detection du texte sur une carte d'identification

Source de l’image : manipulation personnelle du méthode EAST

3.2.3 Reconnaissance de texte

Une fois que nous avons détecté les cadres de délimitation contenant le texte, l'étape suivante consiste à reconnaître le texte.
Il existe plusieurs techniques pour reconnaître le texte.
Voici certaines des meilleures techniques les plus utilisées :

- CRNN
- Machine Learning OCR avec Tesseract

3.2.3.1 CRNN

Le réseau de neurones récurrents convolutifs (CRNN) est une combinaison de pertes CNN, RNN et CTC (classification temporelle connexionniste) pour les tâches de reconnaissance de séquences basées sur des images, telles que la reconnaissance de texte de scène et l'OCR.
L'architecture du réseau est issue d'une article publié en 2015.

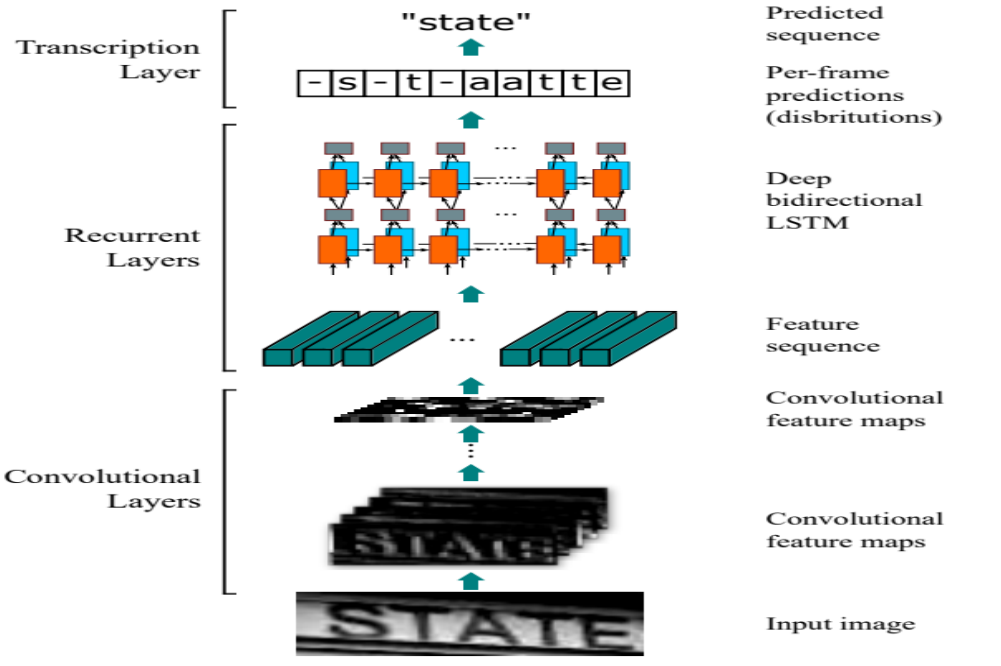

Figure 1. The network architecture. The architecture consists of three parts: 1) convolutional layers, which extract a feature sequence from the input image; 2) recurrent layers, which predict a label distribution for each frame; 3) transcription layer, which translates the per-frame predictions into the final label sequence.

Figure 11:Architecture du réseau CRNN

Source de l'image : https://arxiv.org/pdf/1507.05717.pdf

Cette architecture de réseau neuronal intègre l'extraction de caractéristiques, la modélisation de séquences et la transcription dans un cadre unifié. Ce modèle n'a pas besoin de segmentation des caractères. Le réseau de neurones à convolution extrait les caractéristiques de l'image d'entrée (région de texte détectée).

Le réseau de neurones récurrent bidirectionnel profond prédit la séquence d'étiquettes avec une certaine relation entre les caractères. La couche de transcription convertit la trame par image réalisée par RNN en une séquence d'étiquettes. Il existe deux modes de transcription, à savoir la transcription sans lexique et la transcription à base de lexique. Dans l'approche basée sur le lexique, la séquence d'étiquettes probable la plus élevée sera prédite.

3.2.3.2 Machine Learning OCR avec Tesseract

Tesseract a été développé à l'origine aux laboratoires Hewlett-Packard entre 1985 et 1994. En 2005, il a été open-source par HP. Selon Wikipédia- *En 2006, Tesseract était considéré comme l'un des moteurs OCR open source les plus précis alors disponibles.*
La capacité de Tesseract était principalement limitée aux données textuelles structurées. Il fonctionnerait assez mal dans du texte non structuré avec un bruit important. Le développement ultérieur de tesseract est sponsorisé par Google depuis 2006.

La méthode basée sur l'apprentissage en profondeur fonctionne mieux pour les données non structurées. Tesseract 4 a ajouté une capacité basée sur l'apprentissage en profondeur avec le moteur OCR basé sur le réseau LSTM (une sorte de réseau neuronal récurrent) qui se concentre sur la reconnaissance de ligne mais prend également en charge l'ancien moteur OCR Tesseract de Tesseract 3 qui fonctionne en reconnaissant les modèles de caractères. La dernière version stable 4.1.0 est publiée le 7 juillet 2019. Cette version est également nettement plus précise sur le texte non structuré.
Donc dans notre cas (texte structuré) tesseract avec un solide pré-traitement d'image permet d'atteindre les besoins.

4. Réalisation (implémentation)

4.1 Restructuration (post-traitement)

Le développement des technologies de l'information a de plus en plus changé les moyens d'échange d'informations, entraînant le besoin de numériser les documents imprimés.
De nombreuses organisations au cours de leurs procédures d'intégration, pour disposer d'une quantité adéquate d'informations sur leurs clients, exigent que les clients soumettent des documents qu'ils pourraient utiliser pour vérifier leur identité et obtenir des détails pertinents à leur sujet.
L'opérateur de service ou quelqu'un qui vérifie manuellement la carte d'identité est normalement géré par des humains. L'homme a besoin de se reposer, l'homme ne peut pas travailler 24 heures sans dormir et parfois les humains ont fait des erreurs lors de la saisie des données parce qu'il y avait trop de carte d'identité qui devrait être entrée ou pour écrire le contenu de la carte d'identité.
En raison d'erreurs parfois commises par l'homme, nous avons ensuite automatisé la saisie de données sur la carte d'identité à l'aide de la reconnaissance optique de caractères (OCR) et du post-traitement à l'aide du traitement du langage naturel (NLP).
On a donné comme une entrée l'image suivante :

Figure 12 l'image d'entrée

Ensuite lorsque l’image est traitée par le processus OCR afin d’extracter les données textuelles dans l’image, le résultat de l’extraction sera comme indiqué dans l'image ci-dessous :

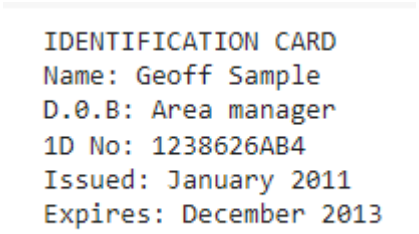

Figure 13 sortie de l'image

Après avoir obtenu le résultat de l'extraction de l'image (Figure 7), nous utilisons une expression régulière, l'expression régulière utilisé pour trouver une séquence de caractères qui définissent un modèle de recherche.

Dans notre exemple on a besoin seulement de Name et ID, nous avons divisez la phrase par " : " pour prendre chaque contenu de champ, car nous avons juste besoin du contenu et non de l'attribut. Par exemple « Name : Geoff Sample », nous avons juste besoin de « Geoff Sample » et non de « Name : ».

Figure ci-dessous montre le résultat après post-traitement (résultat finale) :

Figure 14 résultat finale

4.2 Interface graphique :

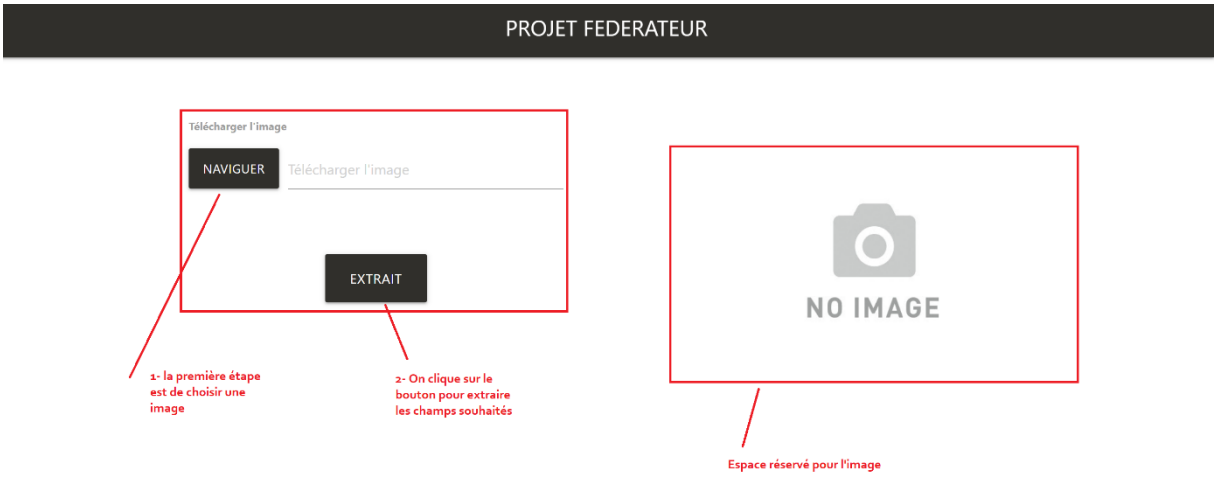

Figure 15:L'interface graphique

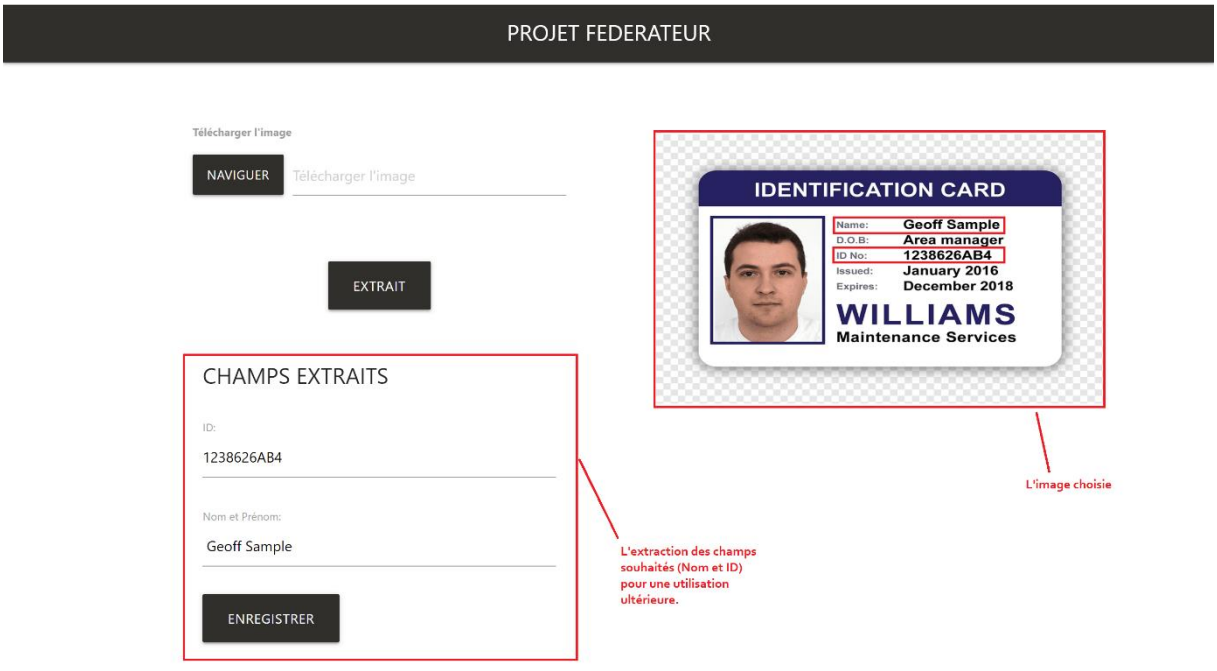

Figure 16:Exemple d'utilisation de l'interface

Conclusion et Perspective

Nous avons créé un extracteur (qui extrait le nom et l'ID) de carte d'identité à l'aide de l'OCR et du post-traitement à l'aide de TAL (traitement automatique du langue).
La numérisation des cartes d'identité réduit beaucoup de temps et d'efforts humains dans plusieurs organisations et modèles économiques. Grâce à l'intelligence artificielle, nous pouvons automatiser ce problème et déployer des solutions en temps réel sur différentes applications.
Ces modèles réduisent non seulement les coûts et les efforts, mais sont également évolutifs et hautement performants.
Ces modèles peut s'intègre facilement dans n'importe quel système :
Les solutions numérisées peuvent être facilement intégrées dans n'importe quel système. Par exemple, le modèle qui est formé pour identifier les informations d'une carte d'identité particulière peut être déployé sur un site Web où les utilisateurs téléchargent les images en masse, ou il peut être utilisé dans les téléphones mobiles où les utilisateurs cliquent sur les images et ainsi, les informations sont extraites.

Références :

https://addepto.com/wp-content/uploads/2019/12/Projekt-bez-tytulu

https://cloud.google.com/vision/docs/images

https://addepto.com/wp-content/uploads/2019/12

https://towardsdatascience.com/pre-processing-in-ocr-fc231c6035a7

https://datapeaker.com/fr/Big-Data/reconnaissance-optique-de caract%C3%A8res-ocr-avec-tesseract-opencv-et-python/

https://members.loria.fr/ABelaid/Enseignement/esstt/3-Pretraitement.pdf

https://github.com/UB-Mannheim/tesseract/wiki

https://www.analyticsvidhya.com/blog/2020/05/build-your-own-ocr-google-tesseract-opencv/#:~:text=OCR%2C%20or%20Optical%20Character%20Recognition,even%20a%20natural%20scene%20photograph.

https://ichi.pro/fr/ocr-avec-tesseract-opencv-et-python-231743215466598

https://gist.github.com/jaafar-benabderrazak/a949d1d85b8a8bd225232edcea99c5cd#file-show_images-py