



ÉCOLE POLYTECHNIQUE DE L'UNIVERSITÉ DE NANTES  
DÉPARTEMENT D'INFORMATIQUE

RAPPORT DE RECHERCHE ET DÉVELOPPEMENT

# Détection et segmentation des expressions mathématiques des manuscrits de Leibniz

## *Rapport finale*

Mamisoa RANDRIANARIMANANA & Xianxiang ZHANG

25 Février 2024

encadré par Yejing XIE & Harold MOUCHÈRE

— Équipe IPI —

LABORATOIRE DES SCIENCES DU NUMÉRIQUES DE NANTES  
CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

coordinateur : Philippe LERAY



## **Avertissement**

Toute reproduction, même partielle, par quelque procédé que ce soit, est interdite sans autorisation préalable.

Une copie par xérographie, photographie, photocopie, film, support magnétique ou autre, constitue une contre-façon passible des peines prévues par la loi.

# Détection et segmentation des expressions mathématiques des manuscrits de Leibniz

## Rapport finale

Mamisoa RANDRIANARIMANANA & Xianxiang  
ZHANG

### Résumé

L'étude des manuscrits anciens constitue un domaine de recherche fascinant, offrant un aperçu précieux dans le passé intellectuel de l'humanité. Malgré les avancées significatives dans la reconnaissance de textes manuscrits, l'analyse de manuscrits historiques demeure un défi complexe. Notre projet vise à détecter et segmenter les expressions mathématiques contenues dans les manuscrits de Leibniz. Réalisé au sein de l'équipe IPI du LS2N, en synergie avec le laboratoire SPHERE Paris Diderot, notre projet se déploie en deux phases distinctes et complémentaires. La première phase de revue bibliographique nous a permis de cerner l'état actuel des connaissances et d'identifier les lacunes dans notre domaine. La seconde phase se concentre sur la conception et la mise en œuvre de solution basée sur un réseau neuronal convolutif profond spécifiquement adapté à la détection et à la segmentation fines des expressions mathématiques dans les textes de Leibniz. Confrontés à une limitation en termes de volume de données originales, nous avons également opté pour des stratégies d'augmentation de données, générant ainsi des ensembles d'images artificielles pour enrichir notre base de données.

La méthodologie adoptée a permis de former notre modèle sur un ensemble de 300 images générées, conduisant à une évaluation sur un corpus distinct, reflétant une précision moyenne à 60%. Cette recherche souligne la complexité de l'analyse de documents anciens, particulièrement pour les expressions mathématiques, ajoutant un niveau de difficulté crucial pour progresser dans l'étude des manuscrits historiques.

## **Remerciements**

Dans un premier temps, nous tenons à remercier nos encadrants, Harold MOUCHÈRE et Yeqing XIE pour nous avoir suivis et encadrés tout au long du projet et d'avoir proposé ce sujet profondément intéressant. Nous remercions aussi notre coordinateur Philippe LE-RAY. Nous tenons également les équipes du laboratoire SPHERE Paris Diderot notamment David RABOUIIN, qui dirige le projet d'édition et de commentaires des manuscrits mathématiques inédits de Leibniz, ainsi que le Leibniz Archiv de Hanovre pour leur contribution et de nous avoir mis à disposition des images des manuscrits de Lieibniz. Et pour finir, nous remercions toutes les personnes qui ont contribué de près ou de loin au projet.

# Table des matières

<b>1</b>	<b>Introduction</b>	<b>7</b>
1.1	Contexte . . . . .	7
1.1.1	Présentations de Leibniz . . . . .	7
1.1.2	Les manuscrits de Leibniz . . . . .	7
1.2	Présentation de la problématique . . . . .	8
1.3	Objectifs poursuivis . . . . .	9
1.4	Travail réalisé . . . . .	9
1.5	Contribution . . . . .	10
1.6	Plan de l'étude . . . . .	10
<b>2</b>	<b>État de l'art</b>	<b>12</b>
2.1	Approches de détection et de segmentation pour le repérage des expressions mathématiques dans les manuscrits historiques . . . . .	12
2.1.1	Approches de segmentation pour le repérage des expressions mathématiques dans les manuscrits historiques . . . . .	13
2.1.2	Méthodes de segmentation de structure U-net . . . . .	13
2.1.3	Approches de détection des expressions mathématiques dans les manuscrits historiques . . . . .	15
2.1.4	Récapitulatif . . . . .	21
2.2	Augmentation de donnée . . . . .	22
2.2.1	Déformation de donnée . . . . .	23
2.2.2	Le surechantillonnage . . . . .	24
2.2.3	Augmentation de données par "copier-coller" . . . . .	27
2.2.4	Analyse . . . . .	27
2.2.5	Récapitulatif . . . . .	28
2.3	Méthodes d'évaluation . . . . .	28
2.3.1	Intersection sur Union (IoU) . . . . .	28
2.3.2	La méthode par pixel . . . . .	29

2.3.3	Méthodes Basées sur la Perception Humaine . . . . .	30
<b>3</b>	<b>Proposition</b>	<b>32</b>
3.1	Augmentation de donnée . . . . .	32
3.2	Détection et Segmentation . . . . .	34
3.2.1	Sélection du modèle . . . . .	34
3.2.2	Architecture du modèle . . . . .	35
3.3	Conclusion . . . . .	36
<b>4</b>	<b>Expérimentation et résultat</b>	<b>40</b>
4.1	Préparation de données . . . . .	40
4.1.1	Annotation de donnée . . . . .	40
4.1.2	Augmentation de donnée et constitution de la base de donnée . . . . .	43
4.2	Entraînement . . . . .	47
4.2.1	Ajustement des paramètres . . . . .	47
4.2.2	Fonction de perte . . . . .	48
4.2.3	Processus de traitement d'images . . . . .	50
4.3	Les résultats du traitement d'image. . . . .	51
4.4	Evaluation . . . . .	54
4.4.1	Évaluation d'images synthétiques . . . . .	55
4.4.2	Évaluation d'images réelles . . . . .	58
4.5	Conclusion . . . . .	59
<b>5</b>	<b>Conclusion</b>	<b>61</b>
5.1	Résumé du travail effectué . . . . .	61
5.2	Enseignements . . . . .	62
5.3	Perspectives de recherche . . . . .	62
<b>A</b>	<b>Fiches de lecture</b>	<b>67</b>
<b>B</b>	<b>Planification</b>	<b>78</b>

<b>C Fiches de suivi</b>	<b>81</b>
<b>D Auto-contrôle et auto-évaluation</b>	<b>93</b>



---

# Introduction

Nous décrivons dans cette section le contexte de ce projet de recherche et développement.

## 1.1 Contexte

### 1.1.1 Présentations de Leibniz

« *Gottfried Wilhelm Leibniz, parfois francisé en Godefroid-Guillaume Leibniz, né à Leipzig le 1er juillet 1646 et mort à Hanovre le 14 novembre 1716, est un philosophe, scientifique, mathématicien, logicien, diplomate, juriste, historien, bibliothécaire et philologue allemand. Esprit polymathe, personnalité importante de la période Frühaufklärung, il occupe une place primordiale dans l'histoire de la philosophie et l'histoire des sciences, notamment des mathématiques et est souvent considéré comme le dernier « génie universel »* » [Wik]

Leibniz est surtout connu pour avoir développé le calcul infinitésimal indépendamment d'Isaac Newton, et les deux sont souvent crédités de manière indépendante

de la création du calcul différentiel et intégral. Leibniz a également introduit plusieurs notations et concepts clé, tels que le symbole "d" pour la dérivée et l'intégrale définie. Ses contributions à la philosophie incluent la monadologie, une théorie de la substance basée sur des unités simples.

Pour cela, les manuscrits de Leibniz sont une collection de documents essentiels pour comprendre son processus de pensée et la manière dont il a développé ses théories. Ces manuscrits sont principalement conservés au Leibniz Archiv de Hanovre. Ils font l'objet d'études de beaucoup de chercheurs en particulier les historiens. Parmi eux figurent les chercheurs du laboratoire SPHERE Paris Diderot où est dirigé le projet d'édition et de commentaire des manuscrits de Leibniz, un projet qui est en étroit lien avec ce PRED.

### 1.1.2 Les manuscrits de Leibniz

Les manuscrits de Leibniz sont caractérisés par leur diversité thématique et leur complexité. Ils abordent une

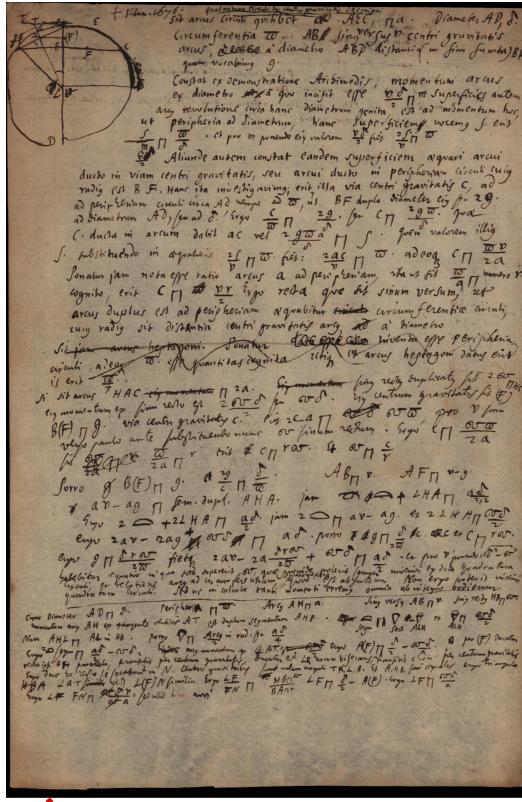


FIGURE 1.1 – Exemple d'image du manuscrit de Leibniz

large gamme de sujets et de poursuites intellectuelles, écrits en français, allemand, latin, etc., comprenant non seulement des formules mathématiques et des expositions, mais également des figures géométriques et des analyses, des études philosophiques, historiques et

linguistiques. Ils varient en longueur, allant de simples notes à des traités complets. Leibniz écrivait souvent en latin, en français, et parfois en allemand, reflétant ainsi son érudition et sa capacité à communiquer dans plusieurs langues académiques de son temps.

Les manuscrits de Leibniz sont d'une importance capitale car :

- ils démontrent plusieurs théories clé notamment le développement du calcul infinitésimal et d'autres concepts mathématiques, ainsi que des idées philosophiques novatrices.
- ils ont une importance historique en un aperçu unique de l'époque de Leibniz, y compris les interactions intellectuelles et les débats de son temps.
- ils ont eu un impact considérable sur les théories et pratiques ultérieures dans divers domaines, influençant d'autres générations, d'autres esprits comme Kant et les philosophes des Lumières ainsi que la génération future.

## 1.2 Présentation de la problématique

Gottfried Wilhelm Leibniz a laissé derrière lui une richesse de manuscrits couvrant un spectre de connaissances allant des mathématiques à la philosophie. Leur compréhension exhaustive nécessiterait des décennies de travail manuel par les historiens.

Cependant, les progrès en informatique et en apprentissage profond offrent une opportunité de réduire considérablement cette charge de travail. Afin d'éviter la trans-

cription manuelle, des outils de transcriptions automatiques ont déjà été mis en place et sont actuellement utilisés par les historiens chercheurs. Ils permettent la transcription du texte et l'édition numérique des manuscrits. Malheureusement, ils n'arrivent pas à reconnaître et à retranscrire les formules et expressions. Et l'entrelacement de textes et d'expressions dans les manuscrits rend la retranscription manuelle des expressions inévitable.

Il existe donc un besoin de pouvoir faire la transcription et l'édition automatique du manuscrit de Leibniz entièrement incluant textes et expressions.

Pour ce faire, il faudra d'abord détecter et segmenter les expressions mathématiques automatiquement. Le fait de juste détecter et de segmenter les expressions est une avancée considérable qui permettra d'alléger les travaux des transcripteurs manuels.

C'est dans cette optique qu'a vue le jour ce projet de recherche et développement.

Placé sous la supervision de Yeqing XIE et bénéficiant de la co-encadrement d'Harold Mouchère, ce projet de recherche et développement se concentre sur la détection et la segmentation des expressions présentes dans les manuscrits de Gottfried Wilhelm Leibniz.

### 1.3 Objectifs poursuivis

Notre recherche se focalise sur le développement de techniques avancées pour détecter et segmenter les expressions mathématiques dans les manuscrits historiques de Leibniz. L'objectif se limite précisément à la localisation et à la segmentation des expressions mathéma-

tiques, sans englober la reconnaissance ou l'interprétation de leur contenu. Dans une première phase, notre projet ambitionne de réaliser un état de l'art, permettant d'identifier les progrès et les défis actuels dans l'analyse des manuscrits anciens, en mettant un accent particulier sur les aspects mathématiques. Cette démarche vise à cerner les besoins spécifiques et les opportunités d'innovation dans notre domaine d'étude. La seconde phase vise à exploiter l'apprentissage profond via un réseau de neurones convolutif profond pour faire la détection et la segmentation. Cette approche sera évaluée sur des images synthétiques ainsi que des images de manuscrits authentiques de Leibniz pour vérifier son efficacité et sa contribution à la sauvegarde et à l'étude de son œuvre. Enfin, la préparation de nos modèles nécessitera une phase d'augmentation de données, étant donné que notre base initiale se compose uniquement d'images issues de dix pages décomposées qui sont divisé en deux sections chacune. Cette limitation quantitative nous oblige à enrichir ensemble de données pour assurer la robustesse et la fiabilité de nos analyses.

### 1.4 Travail réalisé

Dans un premier temps, nous avons investi du temps pour nous imprégner des données disponibles et comprendre la problématique à fond. Ce processus nous a permis de nous informer ensuite sur les approches actuelles en matière de détection et segmentation des expressions mathématiques dans les manuscrits anciens, aboutissant à la rédaction d'un état de l'art exposé dans la première section de notre rapport. Forts de ces connaissances, nous

avons formulé diverses stratégies afin de répondre à la problématique, exposées dans la seconde section du document. La troisième section est dédiée à nos travaux expérimentaux, où nous avons commencé par annoter nos données d'origine, enrichir notre base de données via des techniques d'augmentation, puis suivies de l'entraînement d'un réseau de neurones convolutif profond. Les détails de ces expériences ainsi que les résultats obtenus sont présentés dans le chapitre correspondant.

## 1.5 Contribution

La présente étude apporte une contribution significative à la compréhension et à l'analyse des manuscrits historiques, en mettant particulièrement l'accent sur les écrits de Leibniz. Notre travail constitue une approche innovante qui utilise un réseau neuronal convolutif profond pour détecter et segmenter les expressions mathématiques dans les manuscrits de Leibniz, une tâche qui reste complexe et peu explorée dans la littérature existante. D'un côté, notre projet a enrichi la compréhension des défis inhérents à l'analyse des manuscrits anciens, grâce à une revue bibliographique approfondie qui a permis d'identifier les lacunes actuelles et de positionner notre recherche au sein du domaine. Ensuite, en développant un modèle spécifiquement adapté à la complexité des textes de Leibniz, notre étude a démontré la faisabilité et l'efficacité de l'application de l'intelligence artificielle pour détecter et segmenter des expressions mathématiques dans un contexte où les données sont particulièrement ardues à traiter. La méthode d'augmentation de données que nous

avons adoptée a permis de pallier le manque de données originales, aboutissant à la création d'un modèle robuste, capable de traiter avec une précision de 60% des données qui imitent fidèlement les caractéristiques des manuscrits de Leibniz. Cette précision est encore plus élevée si elle est calculée au niveau des pixels en ne prenons en compte que les pixels sombres correspondant aux écritures. Enfin, notre recherche contribue à la méthodologie en proposant des techniques d'augmentation de données et des stratégies de segmentation spécifiques, offrant ainsi de nouvelles voies pour les futures recherches dans le domaine des manuscrits anciens.

## 1.6 Plan de l'étude

Le chapitre 2, dédié à l'état de l'art, expose nos investigations dans la littérature scientifique en lien avec notre sujet d'étude.

Dans le chapitre 3, nous exposons nos propositions inspirées par l'état de l'art, structuré également en deux segments. Le premier segment détaille notre approche initiale pour l'augmentation de donnée, tandis que le second discute des approches pour la détection et la segmentation des expressions mathématiques dans les manuscrits.

Le chapitre 4 se focalise sur nos expériences, implémentations et les résultats obtenus. Nous y décrivons les apprentissages initiales, les optimisations réalisées, et nos tests en évaluant et analysant la qualité des reconstructions en fonction des méthodologies appliquées.

Finalement, le chapitre 5 conclut en revisitant le projet pour mettre en lumière nos contributions, le travail ac-

compli, les apprentissages tirés, et les futures directions de recherche.

---

# État de l'art

L'état de l'art constitue un pilier essentiel de notre recherche, permettant de situer notre travail dans le contexte plus large des avancées scientifiques actuelles. Cette section vise à synthétiser et à évaluer les travaux antérieurs pertinents, offrant ainsi une base solide sur laquelle notre étude se construit et se justifie. Bien que l'état de l'art présenté dans cette section aspire à fournir une vue d'ensemble éclairée des avancées récentes dans notre domaine d'étude, il convient de souligner que notre revue n'est pas exhaustive, mais cible les principales tendances, innovations, et défis en lien direct avec la question de recherche abordée.

## 2.1 Approches de détection et de segmentation pour le repérage des expressions mathématiques dans les manuscrits historiques

On peut définir par détection la localisation d'une cible particulière. Pour notre cas, les cibles sont les expressions mathématiques. La détection de caractéristiques, dans le cadre de l'informatique, est un concept de vision par ordinateur et de traitement d'image. On utilise les ordinateurs pour extraire des informations sur l'image et déterminer si chaque point d'image appartient à une caractéristique de l'image. Le but est de diviser les points de l'image en différents sous-ensembles appartenant souvent à des points isolés, des courbes continues ou des zones continues.

De l'autre côté, la segmentation fait référence au processus de subdivision d'une image numérique en plusieurs sous-régions d'image (une collection de pixels).

Le but de la segmentation d'image est de simplifier ou de modifier la représentation d'une image pour la rendre plus facile à comprendre et à analyser. Elle est souvent utilisée pour localiser des objets et les délimiter. Plus précisément, la segmentation d'image est un processus d'étiquetage de chaque pixel d'une image afin que les pixels portant la même étiquette présentent des caractéristiques visuelles communes. Le résultat de la segmentation de l'image est un ensemble de sous-régions sur l'image (toutes ces sous-régions couvrent la totalité de l'image.), ou un ensemble de contours extraits de l'image (comme la détection des contours).

Nous allons en premier lieu voir les différentes approches pour la segmentation puis la détection.

### **2.1.1 Approches de segmentation pour le repérage des expressions mathématiques dans les manuscrits historiques**

Dans le cadre de ce projet, la segmentation des expressions mathématiques dans les manuscrits historiques consiste à détecter des zones d'expressions mathématiques dans des images et à représenter les zones d'expressions à l'aide de polygones.

Il existe plusieurs méthodes traditionnelles utilisé pour la segmentation dans le traitement d'images :

- La segmentation de seuil qui consiste à utiliser des seuils globaux ou locaux pour différencier le texte et les expressions. Cette méthode est simple mais sensible à l'éclairage et au bruit de fond.
- Les méthodes basées sur la détection des contours :

segmentez le texte et les expressions en identifiant les contours des images. Les algorithmes couramment utilisés incluent le détecteur de bord Canny.

- L'analyse des régions connectées : divisez l'image en plusieurs régions en fonction de la similarité des pixels pour distinguer le texte et les expressions.

Il y a également des méthodes basées sur l'apprentissage profond tels que :

- Le réseau neuronal convolutif (CNN) afin de reconnaître et segmenter le texte et les expressions dans les images en entraînant un modèle CNN. Cette méthode est généralement plus précise, mais nécessite de grandes quantités de données étiquetées pour la formation.
- Les réseaux de segmentation sémantique : tels que U-Net ou DeepLab, ces réseaux peuvent classer chaque pixel, permettant une segmentation précise des images.
- Les réseau de détection d'objets : tel que YOLO ou Faster R-CNN, qui peut être utilisé pour détecter des régions de texte et d'expression dans les images, puis les segmenter.

Nous allons voir plus en détails les réseaux de segmentation de structure U-Net.

### **2.1.2 Méthodes de segmentation de structure U-net**

L'architecture U-net est une structure de réseau de neurones convolutif conçue spécifiquement pour la segmentation d'images. Sa forme caractéristique en "U" se di-

vise en deux parties principales : un chemin de contraction pour l'encodage et un chemin d'expansion pour le décodage. Le chemin de contraction suit une architecture convulsive classique, réduisant la dimension de l'image tout en augmentant la profondeur des informations pour capturer le contexte. Inversement, le chemin d'expansion augmente la résolution spatiale des caractéristiques pour permettre une localisation précise, utilisant des connexions de saut pour fusionner les informations contextuelles et spatiales et assurer la précision des frontières segmentées.

Cette architecture peut être adaptée pour la segmentation des lignes de texte dans les documents historiques, comme illustré dans l'étude "Text Line Segmentation in Historical Document Images Using an Adaptive U-Net Architecture" par Olfa Mechi et al [MMIEBA19]. L'architecture modifiée conserve la structure fondamentale en U, mais avec des ajustements significatifs dans le chemin de contraction pour réduire la complexité et éviter le surajustement, et dans le chemin d'expansion pour améliorer la précision de la reconstruction des lignes de texte. Les connexions de saut adaptées facilitent une meilleure intégration des caractéristiques à différentes échelles, essentielle pour la segmentation précise du texte.

**Interêts de la proposition** L'adaptation de l'architecture U-net pour la segmentation des lignes de texte dans les documents historiques dans l'étude de Olfa Mechi et al [MMIEBA19] offre des avantages significatifs pour l'analyse des expressions mathématiques dans les manuscrits de Leibniz car elle est spécifiquement op-

timisée pour traiter les variations inhérentes aux documents anciens, telles que les altérations de l'encre et les styles d'écriture hétérogènes, garantissant la fiabilité même avec les manuscrits les plus dégradés. De plus, la robustesse de cette méthode face aux diverses langues et formats de mise en page assure son applicabilité aux documents de Leibniz, qui peuvent présenter un mélange complexe de textes et de calculs. Enfin, l'architecture propose un cadre adaptable, suggérant qu'avec des ajustements ciblés, elle pourrait être encore plus affinée pour isoler et analyser spécifiquement les symboles et structures mathématiques, offrant ainsi une voie prometteuse pour l'extraction et l'étude précises des contributions mathématiques de Leibniz.

**Limites de la proposition** L'architecture U-net adaptée, bien que performante pour la segmentation des lignes de texte dans les documents historiques, présente certaines limites qui la rendent moins idéale pour la détection et la segmentation des expressions mathématiques dans les manuscrits de Leibniz. Premièrement, cette architecture est optimisée pour le texte et non pour les symboles mathématiques, qui nécessitent une reconnaissance et une interprétation plus nuancées pour chaque composant unique. Les expressions mathématiques complexes, caractéristiques des travaux de Leibniz, avec leurs interactions et superpositions de symboles, dépassent la simple segmentation linéaire du texte et exigent une approche plus sophistiquée. De plus, l'efficacité de l'U-net dépend fortement de la qualité et de la spécificité des annotations dans les ensembles de données d'entraînement,

qui doivent être extrêmement détaillées pour les expressions mathématiques, souvent plus difficiles à obtenir que pour le texte standard. Les variations dans les mises en page et les contenus des manuscrits de Leibniz, tels que les annotations marginales ou les ajouts, requièrent une flexibilité et une adaptabilité que l’U-net actuel peut ne pas posséder sans modifications significatives. Enfin, la capacité de généralisation de l’architecture à des documents aux caractéristiques uniques comme ceux de Leibniz peut être limitée, soulignant le besoin potentiel de développements spécifiques pour traiter efficacement les notations mathématiques anciennes et complexes.

### 2.1.3 Approches de détection des expressions mathématiques dans les manuscrits historiques

Dans ce projet, l’image à traiter est le manuscrit de Leibniz, donc la détection fait référence à l’extraction des informations dans l’image du manuscrit, à la division des points de l’image en différents sous-ensembles et à la détection des pixels appartenant à la zone de points considérée comme expressions mathématiques. Nous allons par la suite voir différentes méthodes qui traitent cet objectif.

#### 2.1.3.1 Méthodes basées sur des modèles HMM profonds

La détection et la segmentation par mot-clé dans les images peuvent être réalisées sur la base du modèle de Markov caché (HMM), en particulier dans le traitement des données de séquence. En effet, le HMM a montré ses

avantages uniques. Dans l’application de la repérage de mots-clés d’image, l’image peut être considérée comme une série de séquences d’observation et les mots-clés correspondent à des états cachés.

Une nouvelle approche utilise un modèle d’apprentissage en profondeur intégré aux modèles de Markov cachés (HMM) pour identifier plusieurs mots-clés dans les documents manuscrits. Il vise à extraire des mots-clés arbitraires en prenant des décisions de segmentation et de reconnaissance au niveau des lignes. Tels que les résultats visibles dans la figure-2.1. Les expériences ont été menées à l’aide de la base de données RIMES, qui constitue une référence pour diverses tâches de reconnaissance d’écriture manuscrite.[TCH<sup>+15]</sup>

**Intérêts de la proposition** Ce système se distingue par sa grande flexibilité, étant capable d’identifier toute variété de mots-clés, et non pas seulement ceux qui ont été spécifiquement pré-entraînés. Cette caractéristique étend considérablement l’applicabilité du système. Une autre particularité de cette méthode est qu’elle reconnaît les mots-clés sans avoir besoin de segmenter le texte en mots individuels, ce qui simplifie le traitement et diminue les taux d’erreur. Par ailleurs, le système utilise un classificateur discriminant, ce qui lui permet d’offrir de meilleures performances par rapport au modèle standard de mélange gaussien (GMM). Le framework sous-jacent est conçu pour apprendre des caractéristiques discriminantes directement à partir des données brutes, sans dépendre de fonctionnalités élaborées manuellement, améliorant ainsi

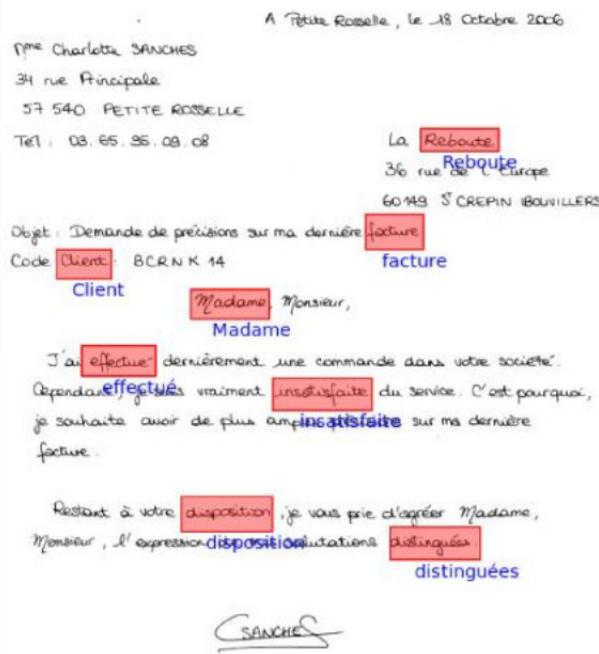


FIGURE 2.1 – Résultats de repérage de plusieurs mots-clés obtenus grâce au modèle HMM [Source : “A Deep HMM model for multiple keywords spotting in handwritten documents,” page 11]

la capacité de généralisation du système. Enfin, grâce à l'intégration des réseaux neuronaux profonds, le système apprend efficacement des fonctionnalités avancées d'écriture manuscrite, ce qui renforce la précision de la reconnaissance.

**Limites de la proposition** Le modèle présente une exigence significative en termes de volume de données de formation. En tant que modèle d'apprentissage profond, il nécessite de vastes ensembles de données pour atteindre des performances optimales. La qualité et la quantité des données d'entraînement sont cruciales : si elles sont limitées ou insuffisantes, les performances du modèle peuvent s'en trouver affectées. Par ailleurs, même si le modèle est conçu pour apprendre à partir de données brutes, il reste susceptible aux variations extrêmes dans le style d'écriture manuscrite ou aux documents fortement formatés. Ces facteurs peuvent compromettre sa capacité à généraliser et à s'adapter à de nouveaux exemples de données.

### 2.1.3.2 Méthodes basées sur le Transformer

Les modèles Transformer ont été initialement conçus pour résoudre des problèmes de traitement du langage naturel (NLP), mais à mesure que la technologie a évolué, ils ont été utilisés avec succès dans le domaine de la vision par ordinateur.

Il prétraite d'abord l'image. Contrairement aux réseaux de neurones convolutifs (CNN) traditionnels, les modèles Transformer nécessitent généralement de segmenter l'image en une série de petits patchs. Chaque patch est aplati et transformé en une séquence de vecteurs, qui servent d'entrée au Transformer.

Le composant principal pour le traitement de ces vecteurs est l'encodeur Transformer. Il utilise un

mécanisme d'auto-attention pour traiter l'intégralité de l'image et capturer la relation entre les différentes parties.

Après avoir traité l'image d'entrée, le modèle Transformer peut être connecté à une ou plusieurs couches entièrement connectées pour la classification ou la reconnaissance de mots-clés. Cette étape implique généralement d'extraire des informations significatives des fonctionnalités générées par le Transformer et de les mapper à différents mots-clés.

Il utilise un ensemble de données d'images annotées avec des mots-clés pour entraîner le modèle. Au cours du processus de formation, les performances du modèle peuvent être améliorées en ajustant ses paramètres, en utilisant différents algorithmes d'optimisation ou en modifiant l'architecture du modèle.

Il existe actuellement une nouvelle méthode d'identification par mot-clé (KWS) dans les documents manuscrits historiques utilisant l'apprentissage auto-supervisé appelée ST-KeyS. Cette méthode est basée sur une architecture de transformateur de vision pure sans aucune couche CNN.[\[JAS<sup>+</sup>23\]](#)

Elle est composée d'une phase de pré-entraînement et d'une phase de réglage fin.

Un encodeur automatique de récupération de masquage est utilisé pour apprendre des représentations utiles à partir des images de mots non étiquetées.

Dans la phase de réglage fin, une stratégie en deux étapes est utilisée pour extraire d'autres représentations promues

pour la tâche de repérage. Pour produire des fonctionnalités plus robustes et plus significatives, une approche siamoise est d'abord utilisé pour intégrer visuellement les images, suivies d'une approche d'alignement des attributs PHOC produits à partir du texte. Comme le montre la figure-[2.2](#).

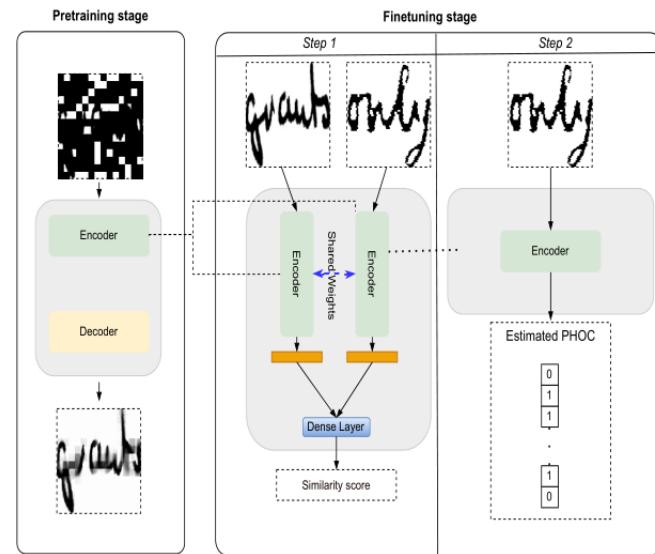


FIGURE 2.2 – Pipeline du cadiciel (framework) ST-KeyS proposé. Elle se compose de deux étapes : une étape de pré-formation et une étape de mise au point. [Source : “ST-KeyS : Self-Supervised Transformer for Keyword Spotting in Historical Handwritten Documents,” page 8]

**Intérêts de la proposition** Cette méthode adopte l'apprentissage auto-supervisé, éliminant ainsi le besoin de grandes quantités de données étiquetées, ce qui est particulièrement bénéfique pour les documents historiques souvent dépourvus d'annotations détaillées. Elle est spécifiquement optimisée pour relever les défis posés par les documents manuscrits anciens, qui incluent la variété des styles d'écriture et les problèmes liés à la dégradation des documents. En combinant le modèle d'auto-encodeur masqué du transformateur avec le réseau neuronal siamois, cette méthode représente une avancée significative par rapport aux technologies précédentes.

**Limites de la proposition** L'utilisation de transformateurs et d'encodeurs automatiques dans le traitement de documents peut exiger des ressources informatiques conséquentes, surtout pour les fichiers de grande taille ou de haute résolution. Bien que ces modèles puissent être efficaces en général, ils nécessitent souvent des ajustements et optimisations spécifiques pour s'adapter à divers types de documents ou styles d'écriture. Un inconvénient notable des modèles d'apprentissage profond, y compris ceux basés sur les transformateurs, est leur manque d'interprétabilité, compliquant la compréhension des mécanismes décisionnels sous-jacents. De plus, bien que ces modèles diminuent la dépendance aux données étiquetées, ils peuvent entraîner des coûts significatifs en termes de calcul et de temps lors des phases d'entraînement et d'ajustement fin.

### 2.1.3.3 Méthodes basées sur l'indexation probabiliste

La détection de mots-clés dans les images basée sur l'indexation probabiliste est un processus complexe qui combine la vision par ordinateur et apprentissage automatique. Cette approche implique généralement de convertir le contenu de l'image en une forme pouvant être traitée avec des modèles probabilistes, puis d'appliquer ces modèles pour identifier et indexer les informations clé de l'image.

Il existe déjà une technologie de détection de mots-clés basée sur l'indexation probabiliste pour améliorer la capacité d'accès rapide au texte des collections d'images de textes manuscrits et permettre la récupération de mots-clés dans des textes manuscrits à grande échelle. Cette approche propose un cadre appelé indexation probabiliste (PrIx) pour une récupération d'images et une indexation efficace de textes manuscrits non transcrits. Cette méthode nous permet de localiser rapidement des emplacements dans une image pouvant contenir des mots-clés spécifiques sans avoir à effectuer une transcription en texte intégral. Grâce à cette méthode, l'efficacité et la précision du traitement des collections de documents manuscrits peuvent être considérablement améliorées, en particulier pour les images de textes manuscrits présentant des styles variés et à grande échelle.[\[VTP23\]](#)

**Intérêts de la proposition** La méthode PrIx représente une avancée significative dans l'efficacité du traitement

de vastes collections d’images de texte manuscrit comparativement à la reconnaissance de texte manuscrit traditionnelle. Elle est conçue pour traiter efficacement des ensembles de données étendus et hétérogènes, facilitant ainsi l’accès et l’utilisation de ces documents historiques ou archivés. En outre, grâce à un modèle de compromis entre précision et rappel, PrIx permet aux utilisateurs d’ajuster finement la balance entre exactitude et exhaustivité des résultats de recherche, répondant ainsi de manière flexible à divers besoins et exigences spécifiques.

**Limites de la proposition** La performance de la méthode est fortement influencée par la qualité et la quantité des données traitées, indiquant que de meilleurs résultats nécessitent des données de haute qualité en volume suffisant. Malgré sa supériorité en termes d’efficacité par rapport aux techniques entièrement automatiques de reconnaissance de texte manuscrit, elle peut engendrer des coûts de calcul importants. En outre, bien que cette méthode améliore les capacités de recherche au sein de documents textuels manuscrits, elle ne garantit pas une précision absolue dans la reconnaissance de texte, soulignant une limite potentielle en termes de fidélité de transcription.

#### 2.1.3.4 Méthodes basées sur les réseaux de neurones convolutifs

La méthode de reconnaissance et de segmentation de manuscrits utilisant un réseau neuronal convolutif (CNN) se décompose en plusieurs étapes clés. Initialement, le

manuscrit est soumis à un prétraitement pour ajuster la taille de l’image, normaliser les valeurs des pixels et réduire le bruit, visant à accroître l’efficacité de l’apprentissage et la précision de reconnaissance. Ensuite, un modèle CNN est conçu pour identifier le texte et segmenter les éléments du manuscrit, utilisant des couches pour extraire les caractéristiques, réduire la dimension et associer les caractéristiques aux résultats de classification. Le modèle est formé avec des données annotées, passant par les phases de propagation avant et de rétropagation pour ajuster les paramètres. Après l’évaluation sur un ensemble de validation, le modèle est optimisé, par exemple, en ajustant le taux d’apprentissage ou la structure du réseau. Enfin, pour la segmentation, le CNN classe chaque pixel pour identifier différentes zones du texte, nécessitant une architecture bien conçue et des stratégies d’optimisation pour éviter le sur ou sous-apprentissage et améliorer la généralisation.

De plus, nous avons trouvé une méthode similaire pour séparer les chiffres et les lettres. Nous pensons que nous pouvons apprendre de cette méthode pour l’adapter à notre projet, qui vise à séparer le texte des formules.

Cet article [RJVGGM18] présente une méthode utilisant l’apprentissage par transfert pour reconnaître la structure de documents manuscrits non étiquetés en utilisant un réseau de neurones convolutif profond entraîné sur des documents générés artificiellement. Cette méthode se concentre sur la distinction au niveau des pixels entre les mots et les chiffres, en utilisant un modèle pré-entraîné sur des images annotées artificiellement pour prédire directement la position des structures au niveau

des pixels. Bien qu'il y ait des défis dans la distinction de l'écriture manuscrite et de certaines structures numériques, ce modèle affiche une haute précision dans la détection de structures sur des images générées artificiellement et montre une portabilité vers des documents manuscrits réels.

**Intérêts de la proposition** L'utilisation des réseaux de neurones convolutifs (CNN) pour détecter les expressions mathématiques dans les manuscrits de Leibniz présente des avantages distincts alignés sur cet objectif spécifique. La haute précision des CNN est cruciale pour identifier et segmenter avec exactitude les éléments mathématiques complexes, souvent entrelacés avec du texte manuscrit dans les documents de Leibniz. Cette précision assure que même les détails subtils des expressions mathématiques ne sont pas perdus ou mal interprétés. L'extraction automatique des caractéristiques par les CNN élimine la nécessité de définir manuellement des caractéristiques spécifiques aux expressions mathématiques, permettant au modèle d'apprendre des représentations optimales directement à partir des données. Cette capacité est particulièrement bénéfique étant donné la variété des symboles mathématiques et des styles d'écriture dans les manuscrits de Leibniz. La flexibilité et l'adaptabilité des CNN leur permettent de s'ajuster aux diversités des manuscrits, reconnaissant des expressions dans différents contextes et langues utilisées par Leibniz. Leur forte capacité de généralisation signifie que, une fois formés, les CNN peuvent identifier efficacement des expressions mathématiques dans une large gamme de documents, même

ceux qu'ils n'ont pas rencontrés durant la phase d'entraînement. Enfin, les capacités de traitement parallèle des CNN accélèrent significativement l'analyse des manuscrits, facilitant un traitement rapide et efficace des vastes corpus de documents de Leibniz, ce qui est essentiel pour avancer rapidement dans les recherches et analyses historiques et mathématiques.

**Limites de la proposition** Bien que les réseaux de neurones convolutifs (CNN) présentent des avantages significatifs pour la détection des expressions mathématiques dans les manuscrits de Leibniz, ils sont confrontés à plusieurs limites dans ce contexte spécifique notamment une grande quantité de données annotées pour atteindre une haute précision. En outre, la formation de modèles CNN pour des tâches aussi spécialisées exige des ressources informatiques considérables,. Un autre défi majeur est le risque de surajustement, particulièrement préoccupant dans le contexte des manuscrits historiques où les données d'entraînement sont intrinsèquement limitées. De plus, l'interprétabilité des modèles CNN reste une question ouverte ; leur nature de « boîte noire » rend difficile la compréhension des raisons sous-jacentes à leurs prédictions, un aspect crucial pour l'analyse et la validation scientifique des résultats obtenus sur des documents historiques. Enfin, le réglage des paramètres et de la structure du réseau pour les CNN est complexe et nécessite un grand nombre d'itérations expérimentales pour optimiser les performances, un processus qui peut être à la fois long et exigeant en termes de compétences spécialisées.

### 2.1.3.5 Analyse

On a pu voir différentes méthodes avec ses avantages et ses limites, notamment pour relever le défi du traitement d'un grand nombre de manuscrits de Leibniz en détectant et récupérant des régions d'expressions.

La méthode basée sur de modèle HMM profond permet d'identifier et de classer efficacement des mots ou des mots-clés dans un texte manuscrit, ce qui est crucial pour le traitement de manuscrits tels que Leibniz. Nous pouvons rechercher des expressions en définissant tous les symboles du dictionnaire de symboles mathématiques comme mots-clés.

De l'autre côté, la méthode basée sur le Transformer est particulièrement efficace pour gérer les variations de style d'écriture, qui sont courantes dans les documents historiques tels que le manuscrit de Leibniz. Cela implique également l'architecture du réseau neuronal siamois (SNN) et les intégrations PHOC, les rendant plus adaptables aux défis spécifiques du manuscrit de Leibniz, tels que la distinction entre le texte et les expressions. Néanmoins, le processus de formation à cette méthode est trop complexe et nécessite beaucoup de temps. En plus, étant donné qu'on étudiera les manuscrits d'un seul auteur, nous ne sommes pas confrontées aux défis de la variation de styles d'écritures.

Enfin, la méthode basée sur l'indexation probabiliste est particulièrement adaptée au traitement de grandes quantités de documents manuscrits non transcrits, fourni une méthode puissante d'indexation et de re-

cherche de manuscrits. Toutefois, son processus de formation est également très compliqué et demande beaucoup de temps.

La méthode basée sur les CNN offre des avantages significatifs pour la segmentation du texte et des équations dans les documents manuscrits. Elle utilise l'apprentissage par transfert et les réseaux de neurones convolutifs profonds pour identifier et différencier avec précision le texte et les informations numériques directement à partir des données de pixels. Cette technologie est particulièrement efficace pour traiter une variété de styles d'écriture manuscrite et de complexités, offrant ainsi une solution robuste pour l'identification automatique et la classification des différents éléments dans les documents manuscrits non structurés. L'efficacité de cette méthode est encore renforcée par sa capacité à adapter les modèles pré-entraînés à de nouvelles données non vues, démontrant ainsi son potentiel pour améliorer la précision et l'efficacité de l'analyse documentaire. Nous ne pouvons pas comprendre clairement quelles fonctionnalités le réseau neuronal de notre modèle a appris. Cela constituera un défi pour nous. Il est également important de noter que nous avons trouvé le modèle pré-entraîné des auteurs grâce aux liens fournis dans l'article, ce qui pourrait considérablement aider notre entraînement. Cependant, la formation CNN est un processus implicite.

### 2.1.4 Récapitulatif

Grâce à la comparaison ci-dessus, nous avons constaté que la méthode basée sur CNN peut réaliser une segmen-

tation de texte et d'équations de manière très précise et peut donc être considérée comme la meilleure méthode pour résoudre notre problème.

Nous pouvons apprendre de la méthode de formation décrite dans l'article «Transfer Learning for Structures Spotting in Unlabeled Handwritten Documents using Randomly Generated Documents» pour former le modèle de segmentation dont nous avons besoin pour réaliser la séparation des équations et du texte. Le français dans le journal peut segmenter les mots et les chiffres des paragraphes de texte, comme le montre la figure-2.3. Même si nous ne disposons pas de suffisamment d'ensembles de données pour mettre en œuvre un modèle de haute précision, nous pouvons continuer la formation basée sur le modèle pré-entraîné par l'auteur de l'article. Étant donné que les combinaisons et les équations numériques ont des caractéristiques très similaires et que le modèle pré-entraîné inclut déjà la reconnaissance de texte, cela sera très pratique pour notre projet.

## 2.2 Augmentation de donnée

Pour ce projet, nous avons à disposition un nombre de données assez faible par rapport aux besoins en termes de données d'entraînement ce qui peut facilement engendrer un problème de surajustement.

Nous avons exploré plusieurs méthodes afin de pouvoir faire la segmentation et la détection en utilisant peu de données, notamment avec les travaux de Solène Tarride, Aurélie Lemaitre, Bertrand Coüasnon et Sophie Tardivel : “Combination of deep neural networks and logical rules



FIGURE 2.3 – Exemple de repérage de structures réalisé sur des documents manuscrits réels. Les pixels (bleu), (vert) et (rouge) correspondent respectivement aux classes fond, nombre et mot. [Source : “Transfer Learning for Structures Spotting in Unlabeled Handwritten Documents using Randomly Generated Documents,” Page 10]

for record segmentation in historical handwritten registers using few examples”. Leur travaux comparent l'utilisation de réseaux de Détection d'Objets en utilisant le Mask R-CNN et le deep Syntax combinant des réseaux en U avec des règles logiques. Les résultats suggèrent une efficacité à partir de 25 documents d'entraînement uniquement si les données sont homogènes. En ce qui concerne la Deep Syntax, les expériences montrent que le symbole fiscal est bien appris à partir de 10 images, tandis que les premières lignes de texte sont bien apprises à partir de 25 images. [ST21]

Toutefois, considérant la diversité des éléments contenus dans le manuscrit de Leibniz pouvant entremêler le langage naturel, les notations mathématiques et les symboles dans les formules et expressions, ainsi que la multitude de combinaisons de symboles que peuvent constituer une expression donnée, l'homogénéité demandée par les techniques n'est pas respectée pour avoir une efficacité à une vingtaine de données d'apprentissage. Les données originales restent insuffisantes pour éviter le surajustement. De cette manière, procéder à une augmentation de données ou Data augmentation reste la meilleure approche. L'augmentation de donnée ou "data augmentation" est une technique utilisée en apprentissage automatique, notamment en vision par ordinateur, pour augmenter la taille d'un ensemble de données en appliquant diverses transformations aux exemples existants. Elle a pour but d'améliorer la généralisation du modèle en lui fournissant une variété d'exemples, ce qui peut aider à éviter le surajustement aux données d'entraînement limitées et de diversifier l'ensemble de données, tout en améliorant les performances du modèle. Considérant que nos manuscrits sont des images, nous allons nous pencher sur l'augmentation de donnée dans le traitement d'image. Nous pouvons classifier les techniques de Data augmentation en deux grandes classes, comme suggérées dans les travaux de Connor Shorten\* and Taghi M. Khoshgoftaar "A survey on Image Data Augmentation for Deep Learning" [Kho19] : la déformation des données (Data Warping) et le suréchantillonnage (Oversampling).

## 2.2.1 Déformation de donnée

La déformation de donnée est une méthode d'augmentation de donnée en dupliquant les données d'origines et en effectuant des transformations géométriques telles que la rotation, la translation et le recadrage, changement d'échelle, etc. ou le changement de couleur (contraste, luminosité, etc). On peut aussi considérer comme déformation, l'ajout de bruit ou encore les coupures aléatoires. Nous pouvons voir dans le tableau 2.1 quelques méthodes de suréchantillonage avec leurs avantages et leurs inconvénients. Notons que cette liste est non exhaustive et ne couvre pas toutes les méthodes existantes.

### 2.2.1.1 Intérêts

Il permet capturer les variations naturelles dans l'apparence tout en augmentant la variabilité des données et permet ainsi d'introduire des variations structurelles dans les données, ce qui peut rendre le modèle plus robuste face à des transformations similaires dans des données réelles. Pour le cas de manuscrit, il est utile pour simuler les variations dans la taille de l'écriture, la disposition et la présentation des expressions mathématiques.

Cette méthode est pertinente pour l'augmentation des données avec une structure spatiale claire, présentant une variabilité géométrique et dont la variation dans les données peut être modélisée par des changements de perspective, etc.

Pour notre cas, il permet de simuler des variations réalistes dans la présentation des expressions, augmenter la diversité des exemples sans avoir besoin de générer

de nouvelles instances. Il permet également de préserver leur sémantique afin de ne pas introduire d’altérations qui pourraient compromettre la signification des données. Cela peut être crucial lorsque le contenu des expressions est central pour la tâche. En plus, vu que les manuscrits ne suivent pas forcément un format de ligne ou d’espacement normalisé, elles peuvent présenter une variabilité structurelle importante dans l’écriture. Les déformations élastiques et les transformations de perspective peuvent être utilisées pour simuler ces variations, contribuant ainsi à un modèle plus robuste.

### 2.2.1.2 Limites

La déformation de donnée peut présenter un risque que de déformations excessives pouvant altérer la structure réelle des objets dans les images ou encore introduire des distorsions qui ne sont pas présentes dans les données d’origine. Elle nécessite une validation soigneuse afin d’éviter cette déformation excessive.

## 2.2.2 Le suréchantillonnage

Le suréchantillonnage est une méthode visant à augmenter la quantité de données en générant des données synthétiques à partir d’un modèle entraîné avec des données initiales. Cette approche implique la duplication d’échantillons existants avec des variations mineures. Le suréchantillonnage est largement utilisé comme technique d’augmentation de données pour équilibrer les classes d’un ensemble de données en augmentant la fréquence des instances sous-représentées. Cela est accom-

pli en reproduisant ou en générant synthétiquement de nouvelles instances, ce qui contribue à améliorer la performance du modèle, notamment dans le cas où certaines classes sont moins fréquentes. Nous pouvons voir dans le tableau 2.2 quelques méthodes de suréchantillonnage avec leurs avantages et leurs inconvénients. Notons que cette liste est non exhaustive et ne couvre pas toutes les méthodes existantes.

### 2.2.2.1 Intérêts de la proposition

Il permet de compenser le déséquilibre de classe . Il permet d’atténuer les problèmes liés à l’imbalance de classes, améliorant ainsi la performance des modèles, en particulier lorsqu’il y a une sous-représentation significative d’une classe spécifique. Dans notre cas, peut aider à surmonter les défis liés à l’absence fréquente d’exemples d’expressions spécifiques, en augmentant la présence d’expressions dans l’ensemble de données. Il est pertinent pour les données avec des classes sous-représentées, mais qui est important, ainsi que des données ayant une faible variabilité intra-classe par rapport aux autres classes.

### 2.2.2.2 Limites de la proposition

Il peut présenter des risques de surajustement lorsque la classe minoritaire est artificiellement augmentée de manière excessive. Cela peut conduire à une prédominance des caractéristiques de la classe minoritaire, affectant la généralisation du modèle.

Méthode	Avantages	Inconvénients
Rotation	Capture les variations structurelles. Simule des changements d'orientation (horizontale, verticale, oblique).	Risque de déformation excessive. Peut introduire des distorsions si utilisées de manière excessive.
Changement de perspective	Capture les variations structurelles. Simule des variations dans l'angle d'observation	Risque de déformation excessive. Peut introduire des distorsions si utilisées de manière excessive.
Translation déplacement de l'image dans l'espace.	Introduit des variations spatiales. Utile pour simuler des décalages dans les documents.	Risque de perdre des informations importantes.
Modification de l'échelle de l'image.	Simule des variations de taille. Pertinent pour représenter des changements d'échelle.	Risque d'altérer les proportions réelles des expressions. Peut nécessiter une gestion appropriée pour ne pas perdre d'informations importantes
Changement de couleur, contraste, de luminosité ou de teinte	Rend le modèle invariant aux couleurs	non approprié si la couleur est une caractéristique critique.
Ajout de Bruit	réduit le risque de surapprentissage en introduisant une certaine incertitude dans le modèle rend le modèle plus robuste	peut rendre l'interprétation du modèle plus difficile, car il peut apprendre à partir de données perturbées.
Coupure Aléatoire	simule des portions variées des données. réduire le risque de surapprentissage, aidant le modèle à généraliser mieux	si excessive, elle peut entraîner une perte d'informations cruciales, affectant la qualité de l'apprentissage.

TABLE 2.1 – Tableau comparatif des méthodes de déformations de données (Liste non exhaustive)

Méthode	Avantages	Inconvénients
Generative Adversarial Networks (GAN) : machine learning en apprentissage non supervisé avec deux réseaux neuronaux, un générateur des données synthétiques et un discriminateur qui fait l'évaluation	<p>Crée des données réalistes et diversifiées.</p> <p>Utile pour augmenter la variabilité des données.</p> <p>Risque de générer des données non pertinentes.</p>	<p>Risque de déformation excessive.</p> <p>Nécessite un modèle de génération bien formé</p> <p>nécessite beaucoup de données d'entraînement.</p>
Neural Style Transfer : utilise les réseaux neuronaux pour transférer le style d'une image à une autre image tout en préservant son contenu.	<p>Introduit des variations stylisées</p> <p>Ajoute des variations esthétiques aux expressions.</p>	<p>Risque de perturber la structure des expressions.</p> <p>nécessite un modèle bien entraîné</p>
Metalearning : Apprentissage à plusieurs niveaux, optimisation bayésienne	Adaptation aux nouvelles tâches avec peu de données ou des données spécifiques.	Nécessite des architectures complexes et ressources importantes.

TABLE 2.2 – Tableau comparatif des méthodes de suréchantillonage pour la génération synthétique de données (Liste non exhaustive)

### **2.2.3 Augmentation de données par "copier-coller"**

La technique d'augmentation de données par "copier-coller" ou "copy-paste data augmentation" consiste à augmenter la quantité et la diversité des données d'entraînement en copiant des objets ou des segments d'une image pour les coller dans d'autres images. Cette méthode permet de créer de nouvelles images combinant différents éléments de plusieurs sources, ce qui est particulièrement utile pour les tâches de détection d'objets ou de segmentation.

Le processus implique la sélection et la copie d'éléments d'une image source, leur éventuelle transformation (en termes de taille, orientation, couleur, etc.), et leur collage dans une nouvelle image de destination. Cette technique vise à diversifier les scénarios de données disponibles et à renforcer la robustesse des modèles d'apprentissage automatique, les rendant ainsi plus généralisables et précis.

Parmi les avantages du copy-paste data augmentation, on compte la réduction des coûts liés à la collecte et à l'annotation de nouvelles données, l'amélioration des performances des modèles grâce à la variété des données, et la flexibilité permettant de simuler des situations rares ou difficiles à capturer dans la réalité. Cependant, cette méthode présente également des inconvénients, tels que le risque de créer des images peu réaliste, l'introduction de biais si les éléments copiés ne sont pas bien intégrés ou représentatifs, et une certaine complexité nécessitant une compréhension approfondie des données pour éviter

des erreurs d'annotation ou des incohérences.

### **2.2.4 Analyse**

Considérant le fait que nous travaillons sur des documents historiques assez particulier, préserver la caractéristique propre des œuvres qu'on souhaite analyser est essentiels même dans les données qu'on va augmenter. Ainsi générer des données synthétiques peut être délicat, notamment pour le Neural Style transfert qui introduit de variation de style, et qui va à l'encontre de la préservation de l'identité qu'on souhaite garder.

Parallèlement, on souhaite pouvoir simuler des variations réalistes dans la présentation des expressions tout en préservant leur sémantique afin de ne pas introduire d'altérations qui pourraient compromettre le sens et la logique des expressions propre à la pensée de l'auteur. Le suréchantillonage comme le GAN peuvent parfois générer des données qui ne sont pas cohérentes avec le style d'écriture de l'auteur ou qui peuvent sembler artificielles. L'utilisation des transformations géométriques est la méthode avec le moins de risque, car il n'y a pas de création de nouveau contenu qui sont potentiellement bruyantes ou incohérents.

Sachant qu'on aura l'écriture d'un seul auteur avec une seule mode d'écriture et style homogène, le suréchantillonage par génération synthétique peut entraîner une répétition excessive de données existantes, ne fournissant pas une variété suffisante. En plus, les transformations géométriques sont souvent plus faciles à interpréter et à annoter, car elles modifient géométriquement des aspects vi-

suels. Enfin, la génération synthétique peut être coûteuse en termes de ressources computationnelles et de temps d'entraînement par rapport aux transformations géométriques. Bien que le but est d'augmenter les données, le suréchantillonage nécessite un grand ensemble de données d'entraînement pour produire des résultats de qualité, ce qui n'est pas notre cas dans cette étude, car nos données sont très limitées. Si ces ressources sont limitées, les transformations géométriques peuvent être une alternative plus efficace.

## 2.2.5 Récapitulatif

Après la comparaison ci-dessus, nous remarquons que les méthodes de déformation sont largement suffisantes pour augmenter nos données et correspondent aux besoins en termes de variabilité. Nous allons ainsi combiner cette méthode avec la méthode par "Copier Coller" afin d'obtenir des données variées, mais fidèle aux caractéristiques des manuscrits.

## 2.3 Méthodes d'évaluation

Dans cette section, nous abordons quelques métriques d'évaluation utilisées dans la littérature pour évaluer de la performance de la détection et la segmentation des expressions mathématiques résultant des méthodes évoquées dans les deux sections précédentes.

### 2.3.1 Intersection sur Union (IoU)

L'IoU est une métrique permettant évaluer la précision de la segmentation spatiale des expressions mathématiques en mesurant la précision de localisation et d'extension des expressions mathématiques détectées par rapport à un standard de vérité terrain (Ground truth) qui est annoté à la main. Elle quantifie le degré de superposition entre les zones identifiées par l'algorithme et les annotations de référence.

Ainsi, l'Intersection over Union (IoU) nécessite alors deux ensembles principaux de données. D'un côté, les annotations de Référence ou Vérité Terrain (Ground Truth) : ces données représentent la vérité terrain de la localisation et de la segmentation des expressions mathématiques dans les manuscrits. Elles sont créées manuellement par des experts qui annotent chaque expression mathématique dans les images des manuscrits. Ces annotations sont généralement sous forme de coordonnées de boîtes englobantes (bounding boxes) ou de masques de segmentation qui délimitent précisément chaque expression mathématique.

De l'autre coté, les prédictions du Système : ce sont les résultats produits par le système de détection et de segmentation. Pour chaque image de manuscrit, votre système doit générer des boîtes englobantes ou des masques de segmentation qui indiquent où il pense que les expressions mathématiques se trouvent.

L'IOU est calculé en prenant la proportion de la zone d'intersection entre la prédition du modèle et le masque de vérité terrain par rapport à la zone d'union entre les deux. La formule est la suivante :

$$\text{IoU} = \frac{\text{Area of Intersection}}{\text{Area of Union}}$$

En d'autres termes, plus l'IOU est proche de 1, meilleure est la correspondance entre la prédiction et la vérité terrain, ce qui indique une meilleure performance du modèle.

Un IOU de 0 signifierait une absence totale de chevauchement, indiquant une mauvaise prédiction.

**Analyse** En mesurant le rapport entre l'intersection et l'union des zones prédites et de référence, l'IoU fournit une indication claire et intuitive de la performance de segmentation. Toutefois, cette métrique présente des limites, notamment sa sensibilité aux variations de la taille des objets segmentés, ce qui peut entraîner une évaluation biaisée dans les cas où les expressions mathématiques occupent une petite partie de l'image. De plus, l'IoU ne prend pas en compte la structure interne ou la précision des symboles au sein des expressions mathématiques, ce qui signifie que deux segmentations ayant le même score d'IoU peuvent varier considérablement en termes de qualité structurelle.

### 2.3.2 La méthode par pixel

La méthode par pixel est une approche d'évaluation détaillée qui se concentre sur l'analyse des résultats de segmentation au niveau le plus granulaire possible, celui des pixels, permettant une évaluation précise et quantitative de la performance des algorithmes de segmentation. Le principe de la méthode repose sur la comparaison,

pixel par pixel, entre l'image segmentée par l'algorithme et l'image de référence (ground truth) annotée manuellement. Chaque pixel est classé comme un vrai positif (TP), faux positif (FP), vrai négatif (TN), ou faux négatif (FN), selon qu'il a été correctement identifié ou non. Sur cette base, des métriques telles que la précision, le rappel et la mesure F (F-Measure) sont calculés pour évaluer la performance de l'algorithme.

### Métriques utilisées

**Précision (Accuracy)** La précision (Accuracy) est la fraction des prédictions correctes parmi l'ensemble des cas analysés totale de prédiction. Elle mesure la performance globale du modèle sur l'ensemble des classes.

Sa formule est :

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

avec TP : le nombre de vrais positifs, TN le nombre de vrais négatifs, FP le nombre de faux positifs, et FN le nombre de faux négatifs.

**Précision (Precision)** La précision est la fraction des vrais positifs parmi tous les positifs prédits par le modèle. Elle mesure la qualité des prédictions positives du modèle. Sa formule est :

$$\text{Precision} = \frac{TP}{TP+FP}$$

avec TP : le nombre de vrais positifs et FP le nombre de faux positifs.

**Rappel (Recall)** Le rappel est la fraction des vrais positifs parmi tous les cas réellement positifs. Elle me-

sure la capacité du modèle à détecter les cas positifs. Sa formule est :

$$textRecall = \frac{TP}{TP+FN}$$

avec TP : le nombre de vrais positifs et FN le nombre de faux négatifs.

**F-mesure (F1 Score)** Le score F1 est la moyenne harmonique de la précision et du rappel. Il fournit un seul indicateur de performance qui équilibre à la fois la précision et le rappel, particulièrement utile lorsque les distributions de classe sont déséquilibrées. Sa formule est :

$$F1\ Score = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

**Analyse** La méthode par pixel permet de calculer avec précision des métriques telles que la précision, le rappel, et la mesure F, offrant ainsi une vue détaillée de la capacité de l'algorithme à identifier correctement chaque pixel comme faisant partie d'une expression mathématique ou non. Bien que cette méthode offre une évaluation extrêmement précise de la performance de segmentation et permette d'identifier spécifiquement où et comment les erreurs de segmentation se produisent, elle présente des limites, notamment une sensibilité accrue aux petites variations de segmentation qui peuvent ne pas affecter significativement la perception globale de la qualité. De plus, elle se concentre uniquement sur l'aspect spatial et ne prend pas en compte la structure sémantique ou la correcte interprétation des symboles mathématiques. En conclusion, bien que la méthode par pixel soit un outil d'évaluation puissant pour la segmentation d'expressions

mathématiques dans les manuscrits, elle doit être utilisée en complément d'autres évaluations pour fournir une analyse complète et nuancée de la performance des algorithmes.

### 2.3.3 Méthodes Basées sur la Perception Humaine

Les méthodes basées sur la perception humaine jouent un rôle complémentaire crucial aux évaluations algorithmiques, offrant des perspectives uniques sur la qualité et l'utilité des segmentations d'expressions mathématiques dans les manuscrits. Elles permettent d'intégrer les retours d'utilisateurs réels et d'experts dans le processus d'évaluation, assurant que les algorithmes ne répondent pas seulement à des critères techniques, mais aussi aux besoins et aux attentes des utilisateurs finaux dans des contextes réels d'utilisation.

#### 2.3.3.1 Études Utilisateurs

Les études utilisateurs constituent une approche fondamentale pour évaluer la segmentation d'expressions mathématiques dans les manuscrits en se basant sur la perception humaine. Cette méthode implique la participation directe d'utilisateurs, souvent des étudiants, des enseignants, ou des chercheurs en mathématiques, qui interagissent avec les expressions mathématiques segmentées par les algorithmes pour évaluer leur lisibilité, leur exactitude et leur utilité. Les participants peuvent être invités à accomplir des tâches spécifiques, telles que résoudre des

problèmes mathématiques utilisant les données segmentées, ou à exprimer leur degré de satisfaction concernant la clarté et la fidélité des expressions segmentées par rapport à leurs attentes. L'avantage de cette approche réside dans sa capacité à fournir des retours qualitatifs et quantitatifs précieux sur la performance des algorithmes du point de vue de l'utilisateur final, révélant des insights sur les aspects de la segmentation qui sont les plus importants pour une expérience utilisateur réussie. Toutefois, les études utilisateurs peuvent être limitées par leur coût, leur complexité de mise en œuvre et la nécessité d'une analyse approfondie des données recueillies.

### 2.3.3.2 Analyse Qualitative par des Experts

L'analyse qualitative par des experts est une autre méthode clé basée sur la perception humaine pour évaluer la segmentation d'expressions mathématiques dans les manuscrits. Cette approche fait appel à des experts en mathématiques, en traitement d'images, ou en reconnaissance de formes, qui examinent et évaluent les résultats de la segmentation en se fondant sur leur expertise spécifique. Les experts peuvent évaluer la qualité de la segmentation en termes de précision des contours, de fidélité des symboles mathématiques segmentés, et de la préservation de la structure logique des expressions. Ils peuvent également identifier des erreurs subtiles ou des défauts qui ne seraient pas évidents pour des utilisateurs non-experts. L'analyse qualitative par des experts offre une évaluation approfondie et nuancée de la qualité des algorithmes de segmentation, bénéficiant de l'expertise spé-

cialisée pour détecter des problèmes spécifiques et proposer des améliorations ciblées. Cependant, comme pour les études utilisateurs, cette méthode peut être coûteuse et chronophage, et les résultats peuvent varier en fonction de l'expertise et des perspectives individuelles des experts consultés.



---

# Proposition

À l'issue de ce travail de recherche bibliographique, nous avons pu voir plusieurs propositions qui peuvent servir de base à la résolution de notre problème. Ces propositions seront détaillées dans cette section.

## 3.1 Augmentation de donnée

Dans un premier temps, nous avons d'abord utilisé un algorithme en découpant aléatoirement des fragments d'image et nous avons appliqué des transformations géométriques notamment des translations ainsi que le brassage des contenus entre les différents contenus des images.

Cette méthode ne nécessite pas de donnée annotée.

Cependant, nous avons remarqué que la coupure aléatoire peut engendrer des données de mauvaise qualité, car il peut y avoir des coupures au milieu des contenus, engendrant des données erronées.

Afin d'éviter la coupure erronée des données et avoir des nouvelles données, nous avons annotés manuelle-

ment les données en différenciant les expressions mathématiques et les textes. Ainsi, nous avons pu constituer des données avec uniquement des expressions mathématiques, des textes et des éléments combinés. On a pu également appliquer aléatoirement des transformations, telles des translations et des brassages entre contenu de différentes images du manuscrit 1. Notre méthode permet également la vérification des collisions 2 fin de ne pas avoir de donnée transposée et d'avoir des images qui sont déjà annotées en adaptant les annotations en fonction des transformations.

En somme, l'annotation manuelle initiale dans notre processus d'augmentation de données garantit que chaque élément, qu'il s'agisse de texte ou d'expressions mathématiques, est identifié avec une précision optimale dès le départ. Cette précision des annotations est cruciale pour le développement ultérieur de masques via OpenCV, qui isolent ces éléments pour des manipulations détaillées, y compris les transformations géométriques et la vérification des collisions. On a pu alors créer des don-

---

**Algorithme 1** Génération d'images sans collisions

---

**Require:** Dossier source  $D$ , Chemin de l'image de fond  $C$ , Nombre d'images  $N$ , Éléments par image  $E$  et  $T$

**Ensure:** Images sans collisions dans le dossier de sortie  $S$

- 1: **for**  $i = 1$  à  $N$  **do**
- 2:   Préparer l'image de fond et les listes d'éléments équations et textes
- 3:   **for** chaque élément  $e$  **do**
- 4:     Initialiser les tentatives de placement sans collision
- 5:     **while** Tentatives < limite et Collision **do**
- 6:       Essayer un nouveau placement pour  $e$
- 7:       **if** pas de Collision **then**
- 8:         Placer  $e$  et mettre à jour les annotations
- 9:       **end if**
- 10:      **end while**
- 11:      **if** limite de Tentatives atteinte **then**
- 12:       Signaler l'échec de placement de  $e$
- 13:      **end if**
- 14:   **end for**
- 15:   Sauvegarder l'image et les annotations
- 16: **end for=0**

---

---

**Algorithme 2** Placement d'éléments sans collision

---

**Require:** Image de fond  $I$ , Liste d'éléments  $E$ , Zone d'image  $Z$

**Ensure:** Image  $I$  avec éléments  $E$  placés sans collision

- 1: **for** chaque élément  $e \in E$  **do**
- 2:   Calculer masque de  $e$ ,  $M_e$
- 3:   Initialiser collision à VRAI
- 4:   Tentatives  $\leftarrow 0$
- 5:   **while** collision ET Tentatives < limiteTentatives **do**
- 6:     Choisir position aléatoire  $p$  dans  $Z$
- 7:     Vérifier collision avec autres éléments dans  $I$  à  $p$
- 8:     **if** pas de collision **then**
- 9:       Placer  $e$  dans  $I$  à  $p$
- 10:      Mettre à jour  $Z$  pour inclure  $e$
- 11:      collision  $\leftarrow$  FAUX
- 12:     **end if**
- 13:     Tentatives  $\leftarrow$  Tentatives + 1
- 14: **end while**
- 15: **if** Tentatives = limiteTentatives **then**
- 16:    Échec de placement de  $e$ , considérer action alternative
- 17: **end if**
- 18: **end for=0**

---

nées d’entraînement diversifiées et complexes, renforçant l’efficacité de l’entraînement des modèles.

## 3.2 Détection et Segmentation

### 3.2.1 Sélection du modèle

Pour la tâche de détection et segmentation du texte et des expressions mathématiques dans les images, les réseaux de neurones convolutifs (CNN) présentent certains avantages par rapport aux méthodes telles que les modèles de Markov cachés (HMM), le transformateur et les méthodes d’indexation probabiliste. En effet, le CNN est efficace pour extraire des fonctionnalités des zones locales, ce qui est très important pour les tâches de segmentation d’images, car le texte et les expressions mathématiques sont généralement distribués à différents endroits de l’image. CNN peut automatiquement apprendre les caractéristiques des différentes zones de l’image, aidant ainsi à segmenter avec précision le texte et les expressions mathématiques. Ensuite, CNN peut être formé de bout en bout, c’est-à-dire de l’image d’entrée d’origine au résultat de segmentation final, ce qui simplifie l’ensemble du processus de formation et d’optimisation du modèle. En revanche, les modèles tels que les HMM et les transformateurs nécessitent souvent des étapes supplémentaires ou des fonctions de perte pour une formation de bout en bout. Le CNN fonctionne bien également sur des ensembles de données à grande échelle, tandis que les tâches de segmentation d’images nécessitent généralement de grandes quantités de données an-

notées pour entraîner le modèle. L’avantage de CNN est qu’il peut apprendre des caractéristiques plus générales à partir de données à grande échelle, améliorant ainsi les performances du modèle. En plus, le CNN peut conserver efficacement les informations spatiales des images, ce qui est particulièrement important pour les tâches de segmentation d’images. Le texte et les équations ont souvent des dispositions et des structures spatiales spécifiques, et CNN peut mieux capturer ces caractéristiques, réalisant ainsi une segmentation plus précise. Bien que des méthodes telles que HMM, Transformer et les méthodes d’indexation probabiliste puissent présenter certains avantages dans certains scénarios spécifiques, dans les tâches de segmentation d’images, CNN et ses variantes sont généralement des choix plus efficaces et couramment utilisés en raison de leurs caractéristiques locales et des informations spatiales ont de fortes capacités d’apprentissage et peut obtenir de meilleures performances grâce à une formation de bout en bout. Ainsi, pour la détection et la segmentation des expressions mathématiques et de textes, nous choisissons de concentrer nos recherches sur les méthodes de détection de mots-clés. L’utilisation de méthodes basées sur CNN peut servir de base au processus d’identification de mots-clés et également de base pour résoudre notre problème.

Nous définirons un réseau d’apprentissage profond pour apprendre les caractéristiques du texte et des équations. Et utilisez nos fonctionnalités apprises pour segmenter le texte et les expressions mathématiques à partir d’images.

### 3.2.2 Architecture du modèle

Lorsqu'il s'agit de tâches de segmentation de texte et d'expression mathématique dans des images, choisir d'utiliser un modèle pré-entraîné comme architecture sous-jacente présente des avantages évidents. Cette approche accélère non seulement le processus de formation, mais améliore également les performances du modèle. En tirant parti d'un modèle pré-entraîné qui a été entraîné avec succès, nous pouvons tirer parti des riches fonctionnalités qu'il a apprises sur des ensembles de données à grande échelle, permettant ainsi à notre modèle de s'adapter plus rapidement à notre ensemble de données. De plus, les modèles pré-entraînés fournissent une plate-forme flexible qui nous permet de les affiner et de les ajuster pour répondre aux besoins de tâches spécifiques. Cette méthode d'apprentissage par transfert améliore non seulement l'efficacité du modèle, mais améliore également l'adaptabilité du modèle aux tâches de segmentation de texte et d'équations qui nous concernent. Dans l'ensemble, l'utilisation de modèles pré-entraînés comme infrastructure fournit à notre recherche un cadre puissant et flexible qui nous aide à atteindre nos objectifs ultimes.

$X = \{x_1, x_2, \dots, x_N\}$  avec  $x_i \in \mathbb{R}$ , est l'image de  $N$  pixels d'un document manuscrit donné.  $S = \{s_1, s_2, \dots, s_N\}$ , avec  $s_i = (s_i^0 \ s_i^1 \ s_i^2)^T$ ,  $s_i^k \in \{0, 1\}$  correspond à la carte de classification pixel par pixel de vecteurs one-hot indiquant à quelle classe appartient chaque pixel de  $X$ . Dans notre contexte,  $k = 0$  correspond à la classe "arrière-plan",  $k = 1$  est la classe "équation" et  $k = 2$  est la classe "texte". L'objectif de notre modèle est de construire la carte  $Y = \{y_1, y_2, \dots, y_N\}$ , avec

$y_i = (y_i^0 \ y_i^1 \ y_i^2)^T$ ,  $y_i^k \in [0, 1]$  comme estimation la plus proche de  $S$ , à partir de l'image  $X$ . En utilisant un ensemble de  $M$  images  $\chi = \{X_1, X_2, \dots, X_M\}$ , l'ensemble correspondant de cartes de structures  $S$  et un modèle de réseau neuronal avec les paramètres  $\Theta$ , nous visons à apprendre une fonction de transformation  $\tau(X, \Theta) = Y$  en trouvant les paramètres  $\Theta^*$  qui minimisent la fonction de coût  $C$ ,

$$\Theta^* = \underset{\Theta}{\operatorname{argmin}} C(\tau(X, \Theta), S). \quad (3.1)$$

Comme ciblé  $Y = \tau(X, \Theta)$  devrait être l'estimation la plus proche de  $S$ , nous définissons la fonction de coût  $C$  comme l'entropie croisée pondérée entre  $Y$  et  $S$  :

$$C(Y, S) = - \sum_i^N \sum_{k=0}^2 \frac{1}{p_i^k} (s_i^k \cdot \log(y_i^k)), \quad (3.2)$$

où le coefficient de pondération  $p_i^k$  est la probabilité pour le pixel  $x_i$  d'appartenir à la classe  $k$ . Cela devrait réajuster le poids de l'erreur en fonction de la proportion de chaque classe au sein de chaque image  $X$ . Étant donné une image  $X$ , nous avons calculé  $p_i^k$  comme la proportion de pixels appartenant à la classe  $k$  au sein de l'image  $X$ . Considérant un ensemble d'images non étiquetées  $\chi^u$ , l'ensemble des cartes de vecteurs uniques  $S^u$  n'est pas disponible pour apprendre la transformation  $\tau^u(X^u, \Theta) = \gamma^u$ . L'objectif de l'apprentissage par transfert est donc d'apprendre une transformation  $\tau^l(X^l, \Theta) = \gamma^l$  en utilisant un ensemble d'images étiquetées  $X^l$  suffisamment similaire à  $X^u$  pour garantir que  $\tau^l(X^u, \Theta) \approx \gamma^u$ .

Le modèle est basé sur un réseau neuronal entièrement convolutif (FCNN) qui classe chaque pixel de l'image

d'entrée pixel par pixel en trois catégories : arrière-plan, nombre et mot. Une représentation graphique de l'architecture du FCNN entraîné est présentée dans la figure-3.1, avec des détails dans le Tableau-3.1. Puisqu'il est entièrement convolutif, le modèle fonctionne sur toutes les tailles d'image d'entrée, ce qui lui permet d'être appliqué à des documents non étiquetés de manuscrits de Leibniz de différentes tailles. Une représentation pyramidale à 5 niveaux de l'image d'entrée est utilisée comme entrée pour agrandir le champ de réception de chaque carte de caractéristiques et garantir que le réseau peut gérer la reconnaissance de structures de différentes tailles. FONDAMENTALEMENT, le réseau se compose de deux parties : l'extraction de caractéristiques et la construction de la carte des structures. La partie d'extraction des caractéristiques se compose de 5 couches de convolution (taille du noyau  $5 \times 5$ , complétées par des zéros) combinées avec un pooling maximal de  $2 \times 2$ , ce qui donne une forme de couche intermédiaire de  $48 \times 48 \times 256$ . La partie de construction du masque numérique se compose de 6 couches de convolution transposées (taille du noyau  $5 \times 5$  et rembourrage des deux côtés avec la moitié de la taille du filtre) et de deux couches d'upsampling. Elle reconstruit un tenseur de  $384 \times 384 \times 3$  avec un upscaling de  $2 \times 2$  (par répétition de l'échelle), finalement mis à l'échelle à  $1536 \times 1536 \times 3$ . À l'exception de la couche de sortie avec la fonction softmax, la fonction ReLU est sélectionnée comme non-linéarité de sortie de chaque couche. En plus de ne pas nécessiter de normalisation des entrées, il a été démontré que la fonction ReLU contribue à améliorer la vitesse d'entraînement des modèles FCNN. La fonction softmax

effectue une normalisation exponentielle sur chaque vecteur one-hot de la carte produite, donc  $\forall i, y_i^0 + y_i^1 + y_i^2 = 1$ , définie comme

$$\text{Softmax}(x)_j = \frac{e_j^x}{\sum_k e_k^x} \text{ avec } k = 0, 1, 2 \quad (3.3)$$

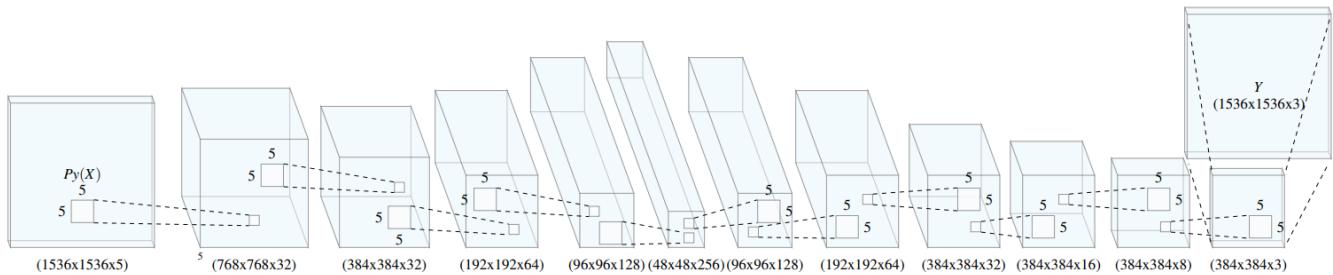
Le modèle a été construit et entraîné à l'aide de la bibliothèque Python Theano et du wrapper d'apprentissage profond Lasagne. Nous entraînons le modèle en minimisant la fonction objectif C (équation-3.2) à l'aide de l'algorithme de descente de gradient stochastique d'Adam. Notez que la sortie directe du réseau est  $Y = \{y_1, y_2, \dots, y_N\}$ , où  $y_i = (y_i^0, y_i^1, y_i^2)^T$  et  $y_i^k = [0, 1]$ . Pour évaluer les performances de classification de notre modèle sous sortie binaire, nous avons calculé la carte de classification résultante  $\hat{S} = \{\hat{s}_1, \hat{s}_2, \dots, \hat{s}_N\}$ , où

$$\hat{s}_i = \begin{pmatrix} \hat{s}_i^0 \\ \hat{s}_i^1 \\ \hat{s}_i^2 \end{pmatrix}, \hat{s}_i^k \begin{cases} 1 & \text{si } \max(y_i) = y_i^k \\ 0 & \text{sinon} \end{cases} \quad (3.4)$$

### 3.3 Conclusion

Dans ce chapitre, nous avons présenté diverses stratégies pour avancer notre étude. Nous allons procéder à l'annotation manuelle de nos données soignée pour maintenir la qualité des données, pour ensuite créer des données synthétiques en sélection des expressions mathématiques et des textes aléatoirement provenant de plusieurs images d'origines, suivi d'un brassage et un positionnement aléatoire. Ces étapes posent une fondation pour

les futurs travaux, notamment en approfondissant l'utilisation de réseaux neuronaux pour la segmentation. Ces étapes ont été soigneusement planifiées et validées en collaboration avec nos encadrants, posant les bases pour les phases expérimentales à venir.



**FIGURE 3.1 –** Représentation graphique de l’architecture du réseau neuronal entièrement convolutif.  $Py(X)$  correspond à la représentation pyramidale de l’image d’entrée  $X$  avec 5 niveaux de résolutions. Une taille de filtre de  $5 \times 5$  a été utilisée pour les couches de convolution et de convolution transposée. Chaque couche de convolution est associée à une couche de pooling maximum de  $2 \times 2$ . Les deux premières couches de convolution transposées sont associées à une couche haut de gamme voisine la plus proche de  $2 \times 2$ . [Source : “*Transfer Learning for Structures Spotting in Unlabeled Handwritten Documents using Randomly Generated Documents*,” Page 7]

TABLE 3.1 – Architecture of our model based on convolutional neural network

Layer type	Filter size	Output layer shape	Activation function
Input image	///	$1536 \times 1536$	///
Py(X)	///	$1536 \times 1536 \times 5$	///
Convolution+maxPool ( $2 \times 2$ )	$5 \times 5 \times 32$	$768 \times 768 \times 32$	ReLU
Convolution+maxPool ( $2 \times 2$ )	$5 \times 5 \times 32$	$384 \times 384 \times 32$	ReLU
Convolution+maxPool ( $2 \times 2$ )	$5 \times 5 \times 64$	$192 \times 192 \times 64$	ReLU
Convolution+maxPool ( $2 \times 2$ )	$5 \times 5 \times 128$	$96 \times 96 \times 128$	ReLU
Convolution+maxPool ( $2 \times 2$ )	$5 \times 5 \times 256$	$48 \times 48 \times 256$	ReLU
Convolution+maxPool ( $2 \times 2$ )	$5 \times 5 \times 128$	$96 \times 96 \times 128$	ReLU
Trans. Conv.+ upscale ( $2 \times 2$ )	$5 \times 5 \times 64$	$192 \times 192 \times 64$	ReLU
Trans. Conv.+ upscale ( $2 \times 2$ )	$5 \times 5 \times 32$	$384 \times 384 \times 32$	ReLU
Trans. Conv.	$5 \times 5 \times 16$	$384 \times 384 \times 16$	ReLU
Trans. Conv.	$5 \times 5 \times 8$	$384 \times 384 \times 8$	ReLU
Conv. + upscale ( $4 \times 4$ )	$5 \times 5 \times 3$	$1536 \times 1536 \times 3$	Softmax
Output map	///	$1536 \times 1536 \times 3$	///

Remarque : ReLU correspond à la fonction Unité Linéaire Rectifiée définie par

$$ReLU(x) = \max(0, x)$$

. Softmax correspond à la fonction exponentielle normalisée définie comme

$$\text{softmax}(x)_j = \frac{e_j^x}{\sum_k e_k^x} \text{ avec } k = 0, 1, 2$$

. La convolution transposée effectue le passage en arrière d'une convolution normale.

# 4

## Expérimentation et résultat

Dans cette partie, nous allons présenter les différentes expérimentations réalisées et résultats obtenus suite aux propositions faites précédemment. Nous présenterons aussi les différentes implémentations et stratégies utilisées, ainsi que les difficultés rencontrées pendant ces expérimentations. Les implémentations et les expérimentations ont été réalisées avec le langage de programmation Python. En premier, nous aborderons la préparation de donnée incluant les annotations, l'augmentation de donnée et la constitutions de la base de donnée. En second, nous parlerons des méthodes utilisées pour la détection et la segmentation.

### 4.1 Préparation de données

#### 4.1.1 Annotation de donnée

Nous avons eu à notre disposition dix images sous format jpg des manuscrits de Leibniz. Chaque jpg représente l'image d'une page qui est divisée en deux parties qu'on

considérera pour la suite comme deux images distinctes : gauche et droite. Ce qui nous a permis au final d'avoir 20 images de départ. Les manuscrits qu'on va étudier seront massivement en latin et en français qui sont assez proches, et ne concerne pas les textes écrits en allemand. Pour nos données, il n'y aura pas de style d'écriture différent, car on aura les images des écritures originales de Leibniz et non les recopies.

Pour annoter nos documents, nous avons employé LabelMe, un outil d'annotation graphique qui nous permet de marquer visuellement les différentes parties des manuscrits en attribuant des étiquettes spécifiques, dans notre cas, "Texte" et "Equation". Ce processus se fait en dessinant des polygones autour des lignes de texte ou des blocs d'équations, permettant une distinction claire, même lorsque ces éléments sont intercalés. Lorsqu'une ligne d'équation est interrompue par du texte, nous segmentons les annotations en polygones séparés pour chaque portion d'équation, veillant à isoler chaque élément selon sa nature. Ainsi, on aura trois Labels dis-

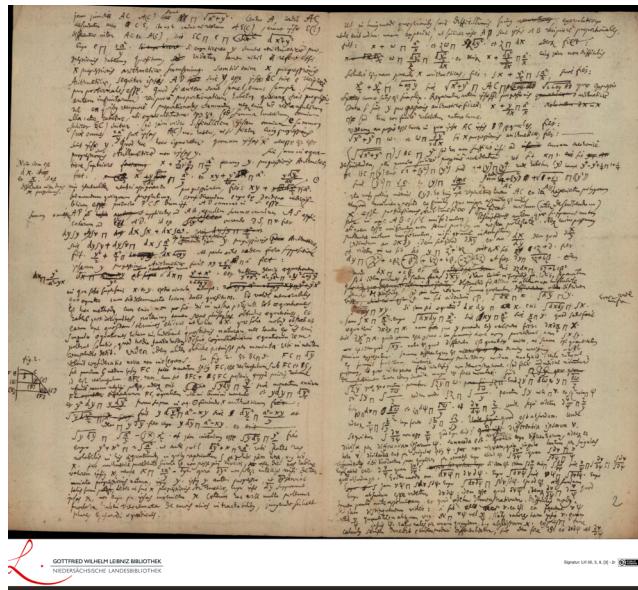


FIGURE 4.1 – Exemple d'image du manuscrit de Leibniz avec 2 parties gauche et droite pour chaque page

tincts : "Equation" pour les parties annotées comme équation, "Texte" pour les parties annotées comme texte et "Fond" pour les parties non annotées. Dans les cas ambigus, où la classification en tant que texte ou équation n'est pas évidente, les segments ne sont pas annotés pour éviter d'introduire des incertitudes dans le jeu de données. Ces données non catégorisées ne sont pas retenues lors de la génération de données synthétiques pour l'entraînement, afin de garantir la précision et la fiabilité des annotations utilisées. Les annotations réalisées avec LabelMe sont sauvegardées sous forme de fichiers JSON, qui offrent

une structure détaillée : pour chaque annotation, le fichier contient les coordonnées des points formant le polygone, l'étiquette associée (texte ou équation), ainsi que d'autres métadonnées utiles comme l'identifiant de l'annotation et des attributs supplémentaires si nécessaires. Ces fichiers JSON fournissent ainsi une représentation structurée et exploitable des annotations, essentielle pour la suite du traitement et de l'apprentissage automatique, où chaque élément annoté est précisément délimité et identifié. Nous avons annoté les 20 images de bases.

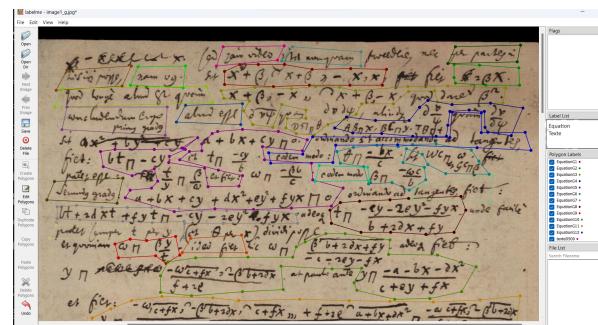


FIGURE 4.2 – Exemple d'annotation avec labelMe

L'ensemble des données des manuscrits montre 136 annotations textuelles et 387 annotations d'équations. Les images varient en nombre d'annotations, avec des instances spécifiques comme présentant jusqu'à 48 annotations d'équation 4.3, tandis que d'autres, contiennent majoritairement d'annotations textuelles. Certaines images ont une proportion relativement plus élevée d'équations par rapport au texte et vice-versa, ce qui indique une variété dans la nature du contenu

des pages. Dans l'ensemble de données, on observe également que certaines images sont densément remplies de contenu tandis que d'autres présentent de larges zones de fond vierge, dénuées de textes ou d'équations. Cependant, ces chiffres ne reflètent pas nécessairement la représentativité précise du contenu annoté. La taille des polygones utilisés pour les annotations varie; parfois, de grands blocs sont annotés en une seule instance, et d'autres fois, des annotations plus petites sont utilisées. Par conséquent, le nombre de pixels à l'intérieur de chaque polygone est un indicateur plus significatif de la quantité de contenu que le nombre d'annotations lui-même. Pour avoir une meilleure représentation, nous pouvons voir dans 4.2 les proportions en pixel des textes par rapport à la quantité de texte et équations combinés ainsi que les proportions en pixel des textes par rapport à la quantité de texte et équations combinés. Nous avons utilisé un algorithme qui analyse uniquement les pixels sombres pour extraire les textes et équations et ignorer les pixels blancs correspondant au fond 3

En prenant exemple avec l'image3\_d 4.3, on a eu comme résultat de l'annotation 1 texte et 14 équations, mais on peut voir dans la quantification par pixel que l'annotation de texte représente quand même 16% si on ne considère pas le fond.

---

**Algorithme 3** Analyse d'images par pixel pour détection de texte et d'équations

---

**Require:** Dossier des images  $D$

**Ensure:** Statistiques des proportions de texte, d'équation et de fond

- 1: Initialiser la liste des résultats  $R$
  - 2: **for** chaque fichier  $f$  dans  $D$  **do**
  - 3:   **if**  $f$  est une image **then**
  - 4:     Lire  $f$  en niveaux de gris
  - 5:     Initialiser comptes :  $pixels\_texte = 0$ ,  
 $pixels\_equation = 0$ ,  $pixels\_fond = 0$
  - 6:     **for** chaque pixel  $(x, y)$  dans l'image **do**
  - 7:       Classer  $(x, y)$  comme texte, équation ou fond
  - 8:       Incrémenter le compte correspondant
  - 9:     **end for**
  - 10:    Calculer les proportions de texte, d'équation, et de fond
  - 11:    Ajouter les proportions à  $R$
  - 12:   **end if**
  - 13: **end for**
  - 14: Calculer et retourner les moyennes pour  $R = 0$
-

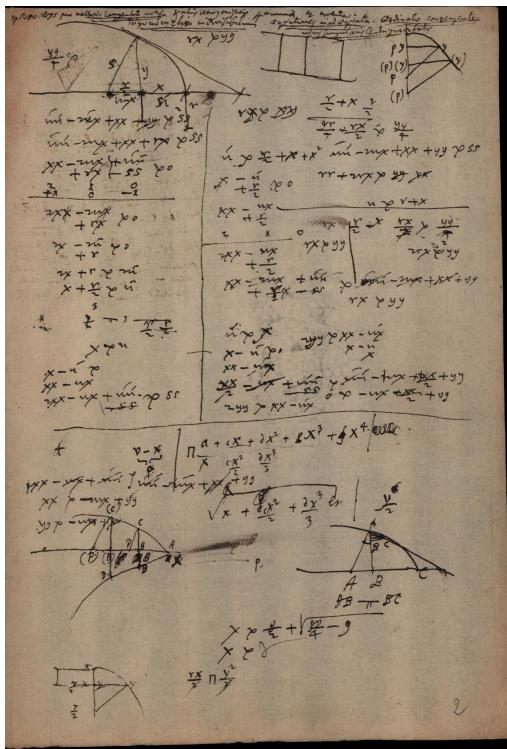


FIGURE 4.3 – Exemple d'image avec majoritairement des équations

#### 4.1.2 Augmentation de donnée et constitution de la base de donnée

L'augmentation de données s'est déroulée en deux étapes distinctes. Au début, l'approche a consisté à utiliser une technique de découpe aléatoire et d'appliquer

Image	Texte	Equation
image1_d	20	30
image1_g	8	12
image2_d	11	42
image2_g	8	22
image3_d	1	14
image3_g	8	11
image4_d	9	31
image4_g	2	5
image5_d	0	11
image5_g	6	30
image6_d	1	1
image6_g	0	13
image7_d	9	38
image7_g	12	48
image8_d	5	4
image8_g	8	16
image9_d	9	16
image9_g	8	24
image10_d	9	15
image10_g	2	4
<b>Total</b>	136	387

TABLE 4.1 – Quantification des annotations Texte et Equation dans les manuscrits.

des transformations géométriques comme des translations et un brassage de contenus entre différents segments d'images, ce qui a généré de nouvelles images sans nécessiter de données annotées, comme illustré dans la Figure 4.4. Cependant, cette méthode a parfois créé des

Image	Proportion Texte	Proportion Équation
image1_g.jpg	0.3099	0.6901
image1_d.jpg	0.7948	0.2052
image2_g.jpg	0.4764	0.5236
image2_d.jpg	0.6792	0.3208
image3_g.jpg	0.9818	0.0182
image3_d.jpg	0.1632	0.8368
image4_g.jpg	0.2791	0.7209
image4_d.jpg	0.6929	0.3071
image5_g.jpg	0.7159	0.2841
image5_d.jpg	0.1200	0.8800
image6_g.jpg	0.1300	0.8700
image6_d.jpg	0.7200	0.2800
image7_g.jpg	0.5488	0.4512
image7_d.jpg	0.6659	0.3341
image8_g.jpg	0.4223	0.5777
image8_d.jpg	0.2100	0.79
image9_g.jpg	0.8874	0.1126
image9_d.jpg	0.8184	0.1816
image10_g.jpg	0.7844	0.2156
image10_d.jpg	0.710	0.390
<b>Moyennes</b>	<b>0.55</b>	<b>0.45</b>

TABLE 4.2 – Quantification par pixel des textes et équations dans les manuscrits.

images de moindre qualité, avec des coupures à travers des éléments cruciaux, induisant des erreurs.

Pour améliorer la qualité, une seconde phase d'augmentation plus précise a été adoptée, utilisant des annotations manuelles pour distinguer entre les textes et les

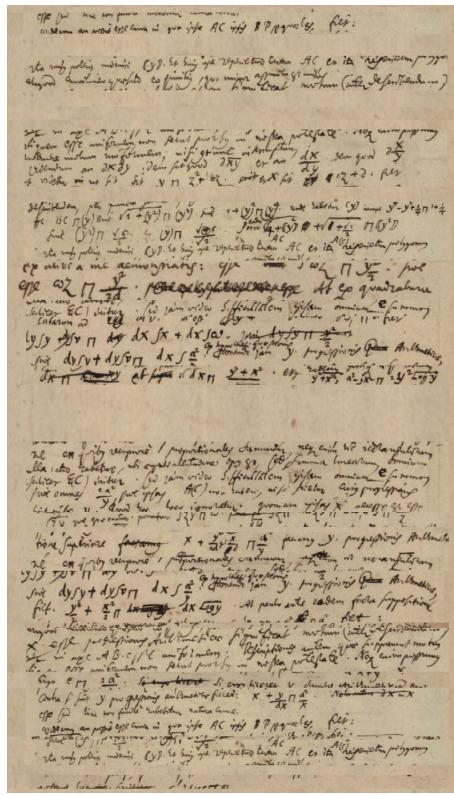


FIGURE 4.4 – Résultats de l'augmentation de donnée avec la coupure aléatoire.

expressions mathématiques, facilitant la création de datasets spécialisés. Cette phase incluait des transformations géométriques aléatoires et un contrôle des collisions pour maintenir l'intégrité des éléments après transformation, produisant ainsi des images de haute qualité et bien anno-

tées, comme le montre la Figure 4.5. Étant donné qu'on n'annote pas les cas ambigus, où la classification en tant que texte ou équation n'est pas évidente, on ne prend pas en compte ces cas dans la génération des données synthétiques. Ce qui garantit le fait que le modèle n'apprend pas de donnée erronée comme considérer du texte comme fond si on suit la logique qu'il n'y a que trois labels : Texte, Equation et Fond.

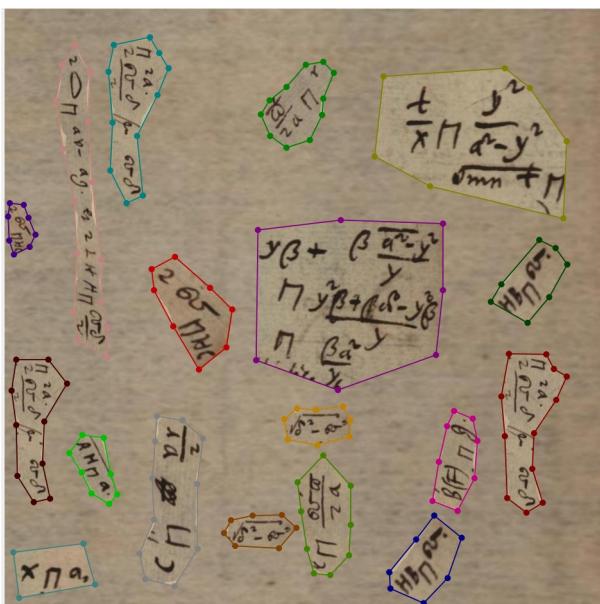


FIGURE 4.5 – Résultats de l'augmentation de donnée avec des données annotées en ne prenant que les expressions mathématiques

Pour constituer notre base de données, nous avons

segmenté les données en trois ensembles distincts : les données d'entraînement, composées de 300 images de chaque type (expressions seules, textes seuls et combinés), les données de test et les données de validation. Les données de test réelles et synthétiques comprennent des proportions variées de texte et d'équations, comme résumé dans le Tableau 4.3. Les images de ces ensembles sont accompagnées de leurs boîtes englobantes correspondantes, les "vérités terrain", qui ont été générées via l'outil PyLabelMe.

Afin de respecter l'équilibre des classes pour les données d'entraînement, nous avons fait en sorte d'avoir à peu près la même quantité de texte et d'équation dans les données, comme illustré dans la figure 4.6 sachant que l'algorithme prend en paramètre le nombre d'équations et le nombre de texte qu'on veut placer dans l'image à créer. Donc, le maintien de l'équilibre des classes peut être facilement contrôlé, ce qui permet d'éviter le surajustement.

Il est essentiel de noter que les totaux d'annotations dans le Tableau 4.3 ne sont pas pleinement représentatifs de la répartition et de la densité des contenus dans les manuscrits. La taille des polygones d'annotation varie, avec certains englobant de larges blocs et d'autres, plus petits. Par conséquent, le nombre de pixels sombres contenus dans chaque annotation offre une mesure plus fidèle de la proportion des contenus annotés dans les images. C'est ce qui est représenté pour la proportion de textes par rapport aux expressions mathématiques et vice-versa.

Donnée	Images d'origine	Détails
Donnée d'entraînement	image1_gauche, image1_droite image2_gauche, image2_droite image3_gauche, image3_droite	300 images avec équations et textes Proportion des expressions mathématiques par rapport aux textes : 50% Proportion d'équations : Proportion de textes par rapport aux expressions mathématiques : 50%
Donnée de validation	image4_gauche, image4_droite image5_gauche, image5_droite image6_gauche, image6_droite	300 images avec équations et textes Proportion des expressions mathématiques par rapport aux textes : 50% Proportion d'équations : Proportion de textes par rapport aux expressions mathématiques : 50%
Donnée de test synthétique	image7_gauche, image7_droite image8_gauche, image8_droite image9_gauche, image9_droite	Proportion des expressions mathématiques par rapport aux textes : 60% Proportion d'équations : Proportion de textes par rapport aux expressions mathématiques : 40%
Donnée de test sur image réel	image9_gauche, image9_droite image10_droite, image10_gauche	Proportion des expressions mathématiques par rapport aux textes : 20% Proportion de textes par rapport aux expressions mathématiques : 80%

TABLE 4.3 – Résumé des données

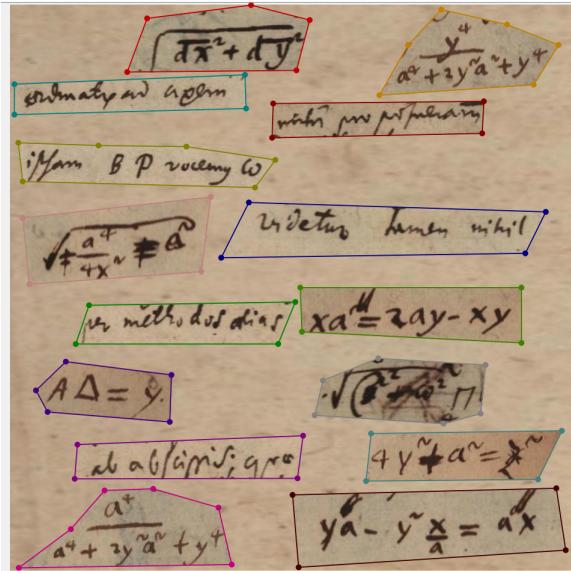


FIGURE 4.6 – Exemple d'image d'entraînement

## 4.2 Entraînement

L’entraînement d’un modèle de réseau neuronal est une étape fondamentale dans notre projet. Dans cette section, nous abordons le processus d’entraînement, en détaillant les différentes phases et approches adoptées pour optimiser notre modèle. Nous mettons l’accent sur la manière dont nous préparons, ajustons et testons le modèle pour qu’il réponde efficacement à nos besoins spécifiques, en soulignant l’importance de chaque étape pour atteindre notre objectif final : un modèle performant et fiable.

### 4.2.1 Ajustement des paramètres

L’ajustement des paramètres pour s’adapter au modèle pré-entraîné constitue une étape cruciale dans le processus d’entraînement de notre réseau neuronal. Cette phase intermédiaire entre la configuration initiale de l’entraînement et l’application concrète des connaissances acquises permet une transition fluide et efficace vers des tâches spécifiques. Dans le contexte de notre projet, nous exploitons un modèle pré-entraîné qui distingue initialement cinq couches de fonctionnalités pour saisir « l’arrière-plan », les « chiffres », les « combinaisons de chiffres », les « lettres » et les « mots ». Toutefois, nos besoins se concentrent sur la segmentation de l’arrière-plan, des équations et du texte. Ainsi, nous réajustons le modèle pour que ces couches de fonctionnalités correspondent à nos cibles spécifiques, en alignant les équations avec les chiffres et leurs combinaisons et le texte avec les lettres et les mots.

Afin de nous adapter à ces cinq couches de fonctionnalités, nous définissons les cinq couches respectivement comme « arrière-plan », « équation », « équation », « texte » et « texte », car les caractéristiques des « équations » sont plus proches des « chiffres » et des « combinaisons de chiffres », tandis que les caractéristiques du « texte » sont plus proches des « lettres » et des « mots ». Nous ferons ainsi correspondre les « équations » aux « chiffres » et « combinaisons de chiffres » d’un côté et nous ferons correspondre les « textes » aux « lettres » et aux « mots ».

Pour récupérer les positions des différents éléments, nous pouvons les localiser grâce aux coordonnées dans

le fichier csv. Nous devons localiser les positions des textes et les positions des équations. Les positions restantes sont les positions de l'arrière-plan. Comme la majorité des blocs d'éléments présents dans le fichier CSV sont de forme polygonale alors que notre traitement nécessite la lecture de coordonnées rectangulaires, il est indispensable de procéder à un prétraitement des coordonnées des éléments à l'aide d'un algorithme spécifique.[-4](#).

---

#### Algorithme 4 Calculate Bounding Rectangle

---

**Require:** *points\_list*

```

1: points  $\leftarrow$  AST.LITERAL_EVAL(points_list)
2: x_coords  $\leftarrow$  [point[0] for point in points]
3: y_coords  $\leftarrow$  [point[1] for point in points]
4: min_x, max_x            $\leftarrow$  MIN(x_coords),
   MAX(x_coords)
5: min_y, max_y  $\leftarrow$  MIN(y_coords), MAX(y_coords)
   =0

```

---

Lorsque nous obtenons les positions des différents éléments, nous pouvons définir la structure de masque des cinq caractéristiques que nous avons définies grâce aux positions de coordonnées des éléments. Nous définissons d'abord le masque de toutes les coordonnées de la première carte de caractéristiques (c'est-à-dire l'arrière-plan) sur 1, et les masques des quatre autres cartes de caractéristiques sur 0. Chaque fois que nous effectuons une lecture sur une plage de coordonnées dans le fichier csv, on ajoute le message que nous recevons à la carte des fonctionnalités. Par exemple, lorsque nous lisons la plage

de coordonnées d'une équation, nous devons mettre à jour le masque de la plage de coordonnées correspondante des deuxième et troisième cartes de caractéristiques (c'est-à-dire l'équation) à 1, et définir la plage de coordonnées correspondante dans la première, mis à jour à 0. Lorsque nous lisons la plage de coordonnées d'un texte, nous devons mettre à jour le masque de la plage de coordonnées correspondante des quatrième et cinquième cartes de caractéristiques (c'est-à-dire le texte) à 1, et changer la plage de coordonnées correspondante. La plage est mise à jour à 0 dans la première carte des fonctionnalités. Nous pouvons implémenter ce processus via l'algorithme[-5](#).

Grâce à la méthode décrite [5](#), nous pouvons localiser les caractéristiques de l'image et les saisir dans le réseau neuronal convolutif pour l'apprentissage.

#### 4.2.2 Fonction de perte

Dans Theano, le principe de calcul de la fonction de perte implique la construction d'une expression algébrique pour définir cette fonction de perte(Comme le montre la formule[-3.2](#)), puis l'utilisation de la capacité de différentiation automatique de Theano pour calculer les gradients de cette fonction par rapport aux paramètres du modèle. Concrètement, cela commence par la définition des représentations symboliques des variables d'entrée et de cible, suivie de la construction d'un chemin de propagation avant pour prédire la sortie. La fonction de perte, en tant que formule mathématique, décrit la différence entre la sortie prédite et les cibles réelles. Une fois la fonction de perte définie, Theano utilise sa puis-

---

**Algorithme 5** Read Image

---

**Require:** *image\_path, mask\_path*

```
1: thisImg ← IMAGE.OPEN(image_path)
2: csvData ← PD.READ_CSV(mask_path)
3: mask_structures ← NP.ZEROS((5, 1024, 1024))
4: for each row in csvData do
5:   bounding_rectangle ← CALCULATEBOUNDING-
   RECTANGLE
6:   if “equation” in row[‘Label’] then
7:     mask_structures[0, bounding_rectangle] ←
      0
8:     mask_structures[1, bounding_rectangle] ←
      1
9:     mask_structures[2, bounding_rectangle] ←
      1
10:    else if “texte” in row[‘Label’] then
11:      mask_structures[0, bounding_rectangle] ←
          0
12:      mask_structures[3, bounding_rectangle] ←
          1
13:      mask_structures[4, bounding_rectangle] ←
          1
14:    end if
15:  end for
16:  return input, mask_structures
   =0
```

---

sante capacité de différentiation automatique pour calculer automatiquement les gradients de la perte par rapport à chaque paramètre du modèle, sans nécessiter de code de calcul de gradient manuel. Ces gradients sont ensuite utilisés pour mettre à jour les paramètres du modèle, généralement via un algorithme d’optimisation tel que la descente de gradient, afin de minimiser la fonction de perte, améliorant ainsi les performances du modèle au cours de l’entraînement. Le cœur de ce processus repose sur des itérations répétées, optimisant continuellement les paramètres du modèle jusqu’à atteindre une condition d’arrêt, telle qu’une amélioration insignifiante de la valeur de la fonction de perte ou un nombre prédéfini d’itérations.

En mettant continuellement à jour les données de perte de formation et de perte de validation, nous pouvons obtenir la courbe de perte suivante.

Comme le montre la figure-4.7, il s’agit de la courbe de perte où nous utilisons uniquement nos propres données pour l’apprentissage ; la figure-4.8 est la courbe de perte où nous continuons à nous entraîner sur la base du modèle pré-entraîné.

Lors d’un entraînement sans utiliser de modèle pré-entraîné, la perte d’entraînement et la perte de validation commencent à une valeur plus élevée, mais diminuent très lentement, et sont toujours à une valeur plus élevée. On peut penser que ce modèle n’a appris aucune connaissance.

Lors de la formation utilisant le modèle pré-entraîné, nous pouvons voir que la perte de formation et la perte de validation chutent rapidement à partir d’une valeur très élevée. Après les premières itérations, les deux valeurs de

perte se stabilisent et présentent des tendances similaires. Cela signifie généralement que le modèle a bien appris à la fois sur les données d'entraînement et sur les données de validation invisibles, sans aucun signe significatif de surajustement. À mesure que le nombre d'itérations augmente, les valeurs de perte ont tendance à osciller, ce qui peut être dû à un taux d'apprentissage plus élevé ou au modèle essayant de s'adapter à certains bruits dans les données d'entraînement. Cependant, puisque la perte de validation reste proche de la perte de formation et fluctue en conséquence, cela indique que le modèle conserve sa capacité à généraliser à de nouvelles données.

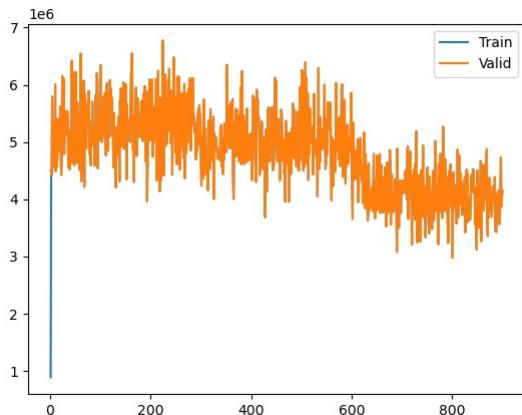


FIGURE 4.7 – La courbe de perte d'apprentissage sans utiliser de modèle pré-entraîné.

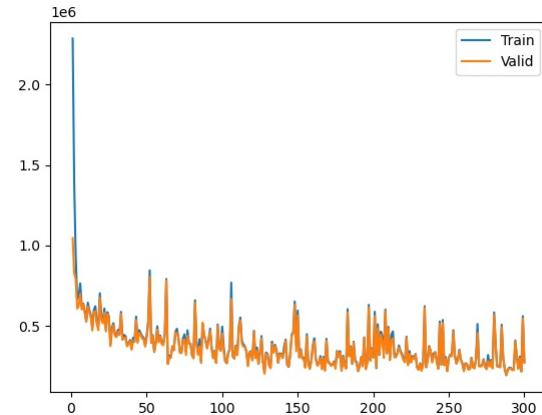


FIGURE 4.8 – La courbe de perte d'apprentissage du modèle pré-entraîné est utilisée.

### 4.2.3 Processus de traitement d'images

#### 4.2.3.1 Prétraitement des images

Les étapes de prétraitement comprennent la lecture de l'image, le redimensionnement (mise à l'échelle) de l'image, l'inversion des couleurs de l'image et le redimensionnement de l'image aux dimensions requises pour l'entrée du réseau neuronal.

#### 4.2.3.2 Construction du modèle d'apprentissage profond

Nous construisons un modèle basé sur le modèle précédemment formé, puis créer le même modèle d'apprentis-

sage profond présenté dans la proposition et illustré par la figure de la figure 3.1. L’architecture du modèle est un réseau neuronal entièrement convolutif avec une représentation pyramidale à cinq niveaux de l’image d’entrée, utilisant des filtres de convolution de taille 5x5, des couches de pooling maximal de 2x2, et des couches de convolution transposée avec un suréchantillonnage (upsampling) pour segmenter précisément chaque pixel de l’image selon différentes catégories.

#### 4.2.3.3 Chargement des poids pré-entraînés

Nous avons utilisé `np.load()` pour lire le fichier de poids pré-entraîné dans le chemin spécifié. Après avoir chargé les poids, nous avons utilisé la fonction `set_all_param_values` de la bibliothèque lasagne pour appliquer ces poids au modèle construit. Cette fonction accepte deux paramètres : le premier paramètre est la couche de sortie du modèle ou l’ensemble du modèle, et le deuxième paramètre est une liste de poids. Les poids de la liste de poids seront appliqués aux calques correspondants dans l’ordre des calques du modèle.

#### 4.2.3.4 Application du modèle

Nous recevons les données d’image ajustées via le modèle, puis utilisons le modèle d’apprentissage en profondeur pour prédire l’image d’entrée. Les résultats de prédiction incluent la distribution de probabilité de chaque pixel correspondant à chaque catégorie. Nous avons utilisé la bibliothèque Theano pour créer un graphique informatique afin d’effectuer le processus de propagation

vers l’avant du modèle et de générer les résultats de prédiction.

Sur la base de la probabilité générée par le modèle, nous attribuons une classe à chaque pixel en sélectionnant la probabilité la plus élevée. De cette manière, l’image originale est convertie en une carte de segmentation, où chaque pixel est étiqueté comme une catégorie spécifique (par exemple, arrière-plan, équation ou texte).

Nous avons créé trois masques correspondant aux zones d’arrière-plan, d’équation et de texte de l’image. L’image d’entrée est ensuite inversée pour rendre ces objets plus visibles. Nous avons utilisé le masque ci-dessus et l’image inversée pour générer un canal pour chaque catégorie : les zones de texte ont été mappées sur le canal rouge, les zones d’équation ont été mappées sur le canal vert et les zones d’arrière-plan ont été mappées sur le canal bleu. De cette manière, différentes catégories de prédiction sont présentées dans différentes couleurs dans les images générées pour faciliter la distinction et l’identification visuelle. Enfin, nous enregistrons l’image générée dans le chemin spécifié.

### 4.3 Les résultats du traitement d’image.

Lors de l’entraînement du modèle de réseau neuronal entièrement convolutionnel, la phase initiale d’entraînement n’a pas utilisé de modèle pré-entraîné, mais a utilisé notre propre ensemble de données pour initialiser l’entraînement. Malgré cela, les poids des paramètres du

réseau après l'entraînement n'ont pas permis d'obtenir de bons résultats de segmentation, mais ont au contraire produit des résultats complètement erronés (Comme le montre la figure-4.9). Pour résoudre ce problème, nous avons décidé de re-entraîner le modèle basé sur un modèle pré-entraîné.

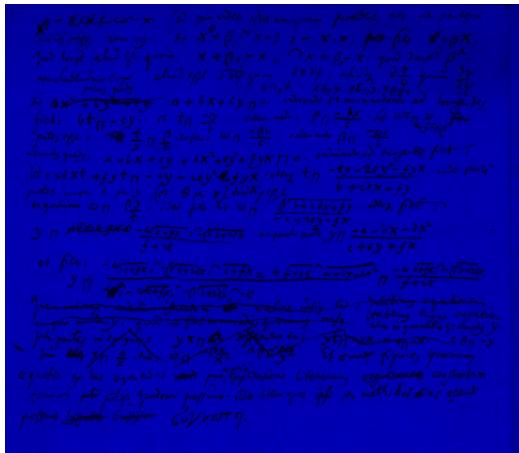


FIGURE 4.9 – Un résultat après entraînement sans utiliser le modèle pré-entraîné.

Dans cette phase d'entraînement, nous avons d'abord utilisé un modèle pré-entraîné sur un ensemble de données volumineux, car de tels modèles ont généralement une meilleure capacité de représentation des caractéristiques et de généralisation. Ensuite, nous avons utilisé notre ensemble de données avec ce modèle pré-entraîné, en ajustant les paramètres du modèle pour s'adapter à notre tâche spécifique. Pendant le processus de fine-

tuning, nous avons ajusté des hyperparamètres tels que la taille des lots d'entraînement pour garantir que le modèle puisse mieux s'adapter à notre ensemble de données.

En re-entraînant le modèle de cette manière, nous avons obtenu des résultats de segmentation plus précis, car le modèle pré-entraîné a déjà appris des représentations de caractéristiques plus riches, ce qui contribue à améliorer les performances et la capacité de généralisation du modèle. De plus, nous avons pu éviter de tomber dans des optimaux locaux pendant le processus de fine-tuning, car le modèle pré-entraîné nous a fourni un bon point de départ.

**Test de segmentation sur images synthétique :** Prenons un exemple de test de segmentation sur une image synthétique qui contient uniquement des équations. L'image de base est présentée à la figure 4.10 et l'image résultante à la figure-4.11.

La partie bleue de l'image représente l'arrière-plan, la partie verte représente l'équation et la partie rouge représente le texte. Dans les résultats, nous pouvons constater intuitivement que le modèle marque les positions des coordonnées de presque toutes les équations, mais il reste encore quelques erreurs dans l'image, telles que le marquage incorrect des équations sous forme de texte, comme le montre la figure 4.12.

Ensuite, prenons un exemple de test de segmentation sur une image contenant des équations et de texte : avec l'image originale à la figure-4.13, et l'image résultante à la figure : 4.14. Nous pouvons constater quelques erreurs

FIGURE 4.10 – Exemple d'image synthétique contenant uniquement des équations.

au niveau des blocs de textes. Mais le modèle arrive à

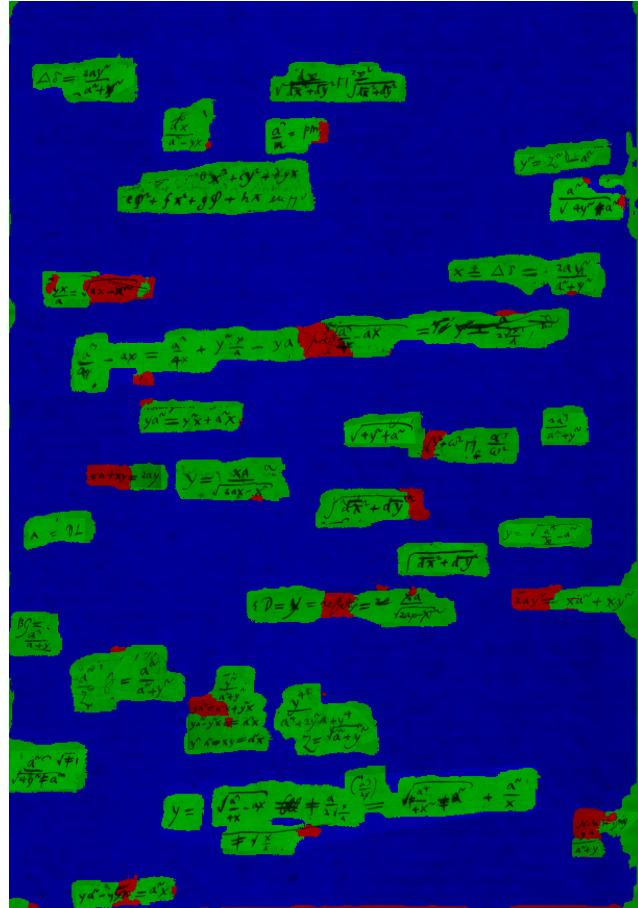


FIGURE 4.11 – Un exemple de segmentation d'une image contenant uniquement des équations.

déetecter toutes les équations qui sont séparées des blocs

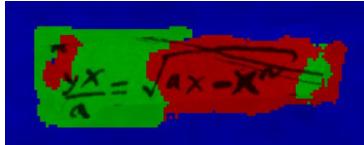


FIGURE 4.12 – Un exemple d'étiquetage incorrect d'une équation sous forme de texte.

de textes.

**Test de segmentation sur image de manuscrit original de Leibniz :** Exemple de résultat de segmentation d'image de manuscrit original de Leibniz : (image originale - figure-4.15, image segmentée - figure-4.16)

Après avoir terminé le fine-tuning, nous avons testé le nouveau modèle pour évaluer ses performances et l'avons comparé au modèle précédemment entraîné. De cette manière, nous avons pu déterminer si le fait de re-entraîner le modèle à l'aide d'un modèle pré-entraîné pouvait améliorer significativement nos résultats de segmentation, et ainsi améliorer les performances et la stabilité du modèle.

## 4.4 Evaluation

Dans un problème de classification à trois types (catégorie 0 - arrière-plan, 1 - équation, 2 - texte), il faut calculer les vrais exemples (TP), les vrais exemples négatifs (TN), les faux positifs (FP) et les faux positifs. exemples (FP) de chaque type. Les exemples négatifs (FN) doivent être considérés séparément pour chaque catégorie.

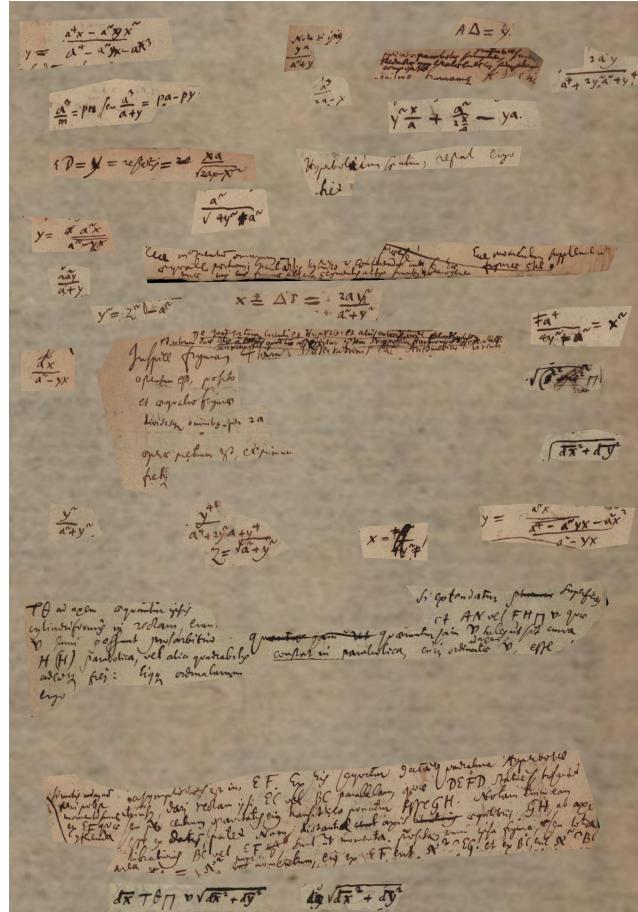


FIGURE 4.13 – Exemple d'image synthétique contenant des équations et de texte.

Sensibilité(Sensitivity) : également connu sous le nom

Catégorie	TP	TN	FP	FN
0 - Arrière-plan	Correct comme 0	Correct comme non-0	Non-0 comme 0	0 comme non-0
1 - Équations	Correct comme 1	Correct comme non-1	Non-1 comme 1	1 comme non-1
2 - Texte	Correct comme 2	Correct comme non-2	Non-2 comme 2	2 comme non-2

TABLE 4.4 – Résumé des métriques de classification par catégorie.

de taux de vrais positifs, il s'agit de la proportion de classes positives correctement identifiées.

$$Sensitivity = \frac{TP}{TP + FN}$$

Spécificité(Specificity) : également connu sous le nom de taux de vrais négatifs, il s'agit de la proportion de classes négatives correctement identifiées.

$$Specificity = \frac{TN}{TN + FP}$$

Précision(Accuracy) : il s'agit de la proportion de classification globalement correcte, calculée comme suit : (exemples vrais + exemples vrais négatifs)/nombre total d'échantillons.

$$Accuracy = \frac{TP}{TP + FN}$$

Précision(Precision) : il s'agit de la proportion d'échantillons prévus comme positifs qui le sont réellement.

$$Precision = \frac{TP}{TP + FN}$$

Score F1 : C'est la moyenne harmonique de la précision et de la sensibilité, utilisée pour considérer l'équilibre entre la précision et le rappel (sensibilité). Idéalement, plus le score F1 est élevé, mieux c'est.

$$F1 = \frac{2 \times TP}{2 \times TP + FP + FN}$$

Coefficient de corrélation de Matthews (MCC) : prenant en compte les quatre quadrants de la matrice de confusion, il s'agit d'une mesure équilibrée avec des valeurs allant de -1 (classification complètement erronée) à +1 (classification complètement correcte), 0 indiquant une supposition aléatoire.  
 $MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP) \times (TP+FN) \times (TN+FP) \times (TN+FN)}}$  (4.1)

#### 4.4.1 Évaluation d'images synthétiques

Pour l'image-4.11, nous pouvons utiliser la méthode ci-dessus pour analyser les résultats de la segmentation de l'image. Comme le montre le tableau-4.4.1. Ces indi-

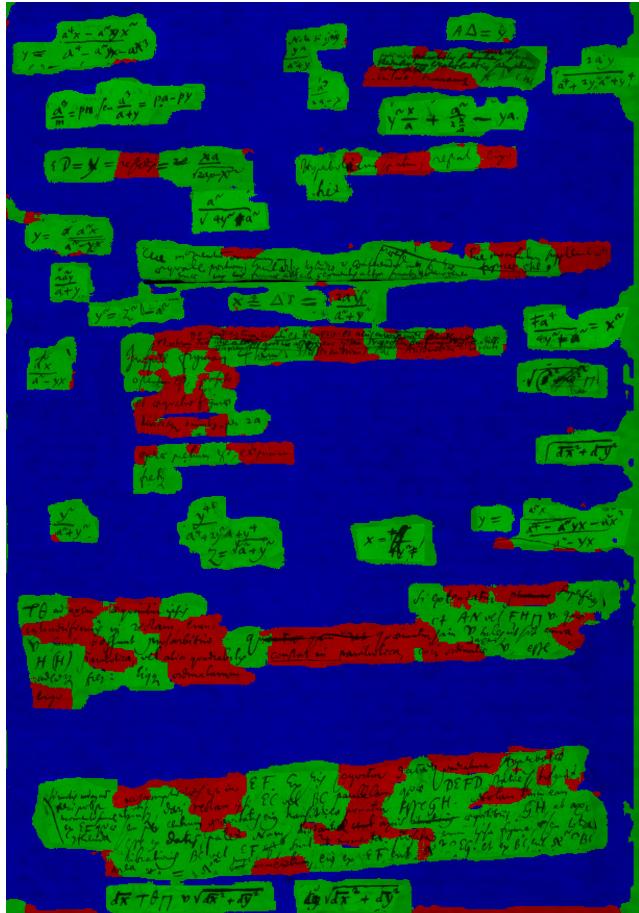


FIGURE 4.14 – Un exemple de segmentation d'une image synthétique contenant des équations et de texte.

ateurs de performance indiquent que le modèle de classi-

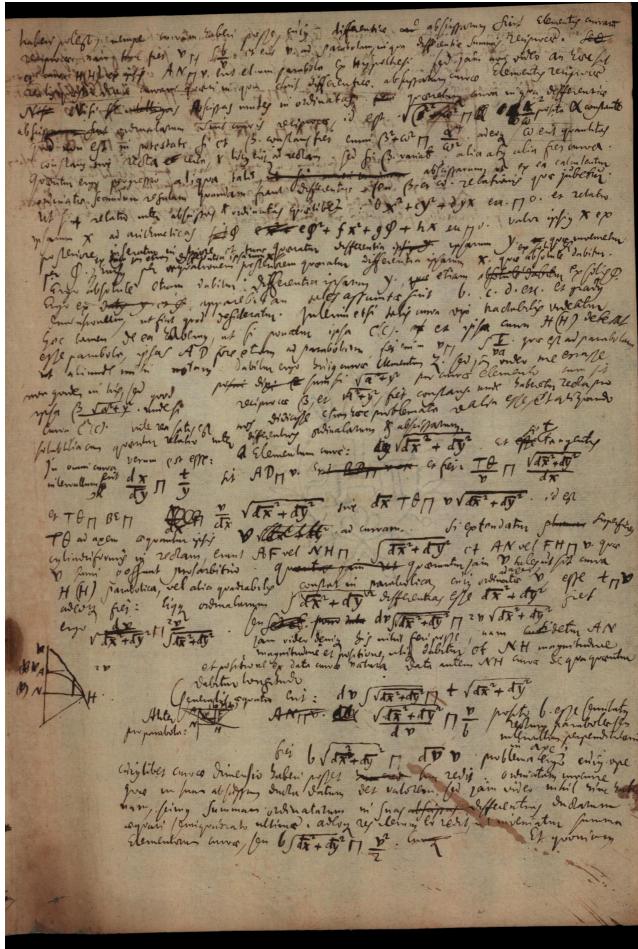


FIGURE 4.15 – Un exemple d'image originale.

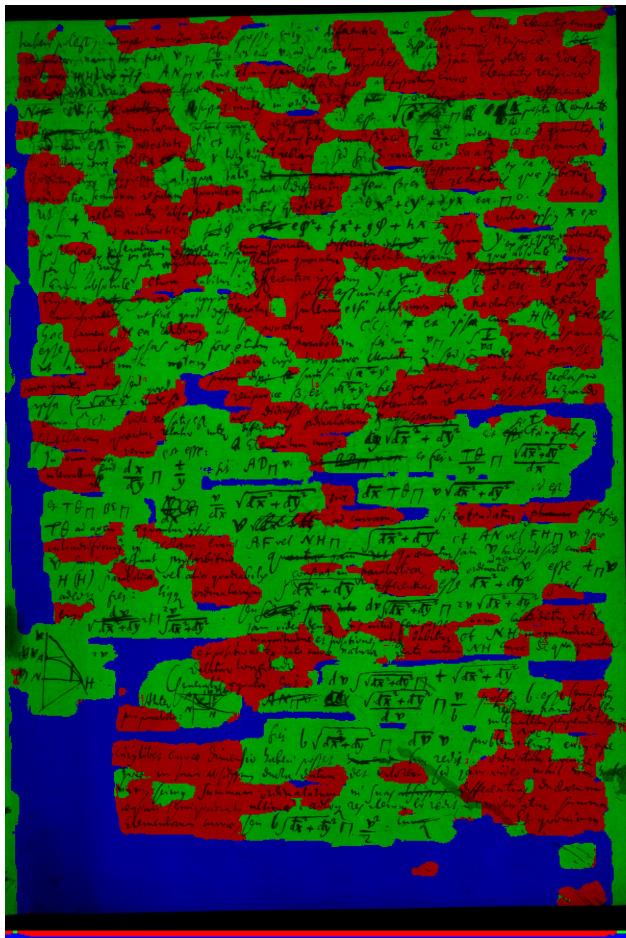


FIGURE 4.16 – Un exemple de résultat de segmentation d'image de manuscrit original de Leibniz.

Metric	Value
Sensitivity	0.7419058545338991
Specificity	0.8709529272669496
Accuracy	0.8279372363559327
Precision	0.7419058545338991
F1 Score	0.7419058545338991
MCC	0.6128587818008487
True Positive	63503.58
True Negative	149098.78
False Positive	22091.62
False Negative	22091.62

TABLE 4.5 – Données d'évaluation pour une image contenant uniquement des équations.

fication se comporte bien dans deux aspects importants : la sensibilité (également appelée rappel) est de 74,19 %, ce qui signifie qu'il peut correctement identifier la plupart des instances positives ; la spécificité est de 87,10 %, ce qui indique qu'il est également assez précis pour exclure les instances négatives. La précision globale atteint 82,79 %, ce qui signifie que le modèle est correct dans plus de trois quarts de toutes les prédictions. La précision et le score F1 sont les mêmes, à 74,19 %, ce qui indique que le modèle est précis et équilibré dans ses prédictions de classes positives. La valeur de MCC est de 0,61, ce qui indique qu'il existe une corrélation assez forte entre les prédictions du modèle et la réalité. Le nombre de vrais positifs et de faux positifs est égal, ce qui peut indiquer que le modèle a des limites dans la distinction entre les classes positives et négatives, surtout en

présence d'un nombre similaire d'instances positives et négatives. Dans l'ensemble, les performances du modèle sont satisfaisantes, mais il existe peut-être des possibilités d'amélioration, notamment en réduisant les faux positifs.

Pour l'image-4.14, nous pouvons utiliser la méthode ci-dessus pour analyser les résultats de la segmentation de l'image. Comme le montre le tableau-4.4.1. Ces indi-

Metric	Value
Sensitivity	0.5626611071648877
Specificity	0.7813305535824439
Accuracy	0.7084407381099252
Precision	0.5626611071648877
F1 Score	0.5626611071648878
MCC	0.34399166074733156
True Positive	48161.09
True Negative	133756.29
False Positive	37434.11
False Negative	37434.11

TABLE 4.6 – Données d'évaluation pour une image contenant des équations et du texte.

cateurs montrent les performances d'un modèle de classification pour une autre image. La sensibilité (ou rappel) est de 56,27 %, ce qui signifie que le modèle peut correctement identifier plus de la moitié des instances positives, mais qu'une partie importante des instances positives est négligée. La spécificité est de 78,13 %, ce qui signifie que le modèle est assez bon pour confirmer les échantillons négatifs, avec un faible taux d'erreur dans la classification des échantillons négatifs en tant que posi-

tifs. L'exactitude globale est de 70,84 %, ce qui indique que près des deux tiers des prédictions sont correctes. La précision (le ratio des échantillons prédits comme positifs qui sont réellement positifs) est identique à la sensibilité, à 56,27 %, ce qui peut se produire dans certaines situations, comme lorsque le nombre d'échantillons positifs et négatifs est équilibré dans un problème de classification binaire. Le score F1 est également de 56,27 %, étant la moyenne harmonique de la précision et du rappel, indiquant qu'il y a encore de la marge pour améliorer l'équilibre entre ces deux aspects du modèle. Le coefficient de corrélation de Matthews (MCC) est de 0,344, un niveau moyen, indiquant une certaine corrélation entre les prédictions du modèle et les résultats réels, mais loin d'une prédition parfaite.

#### 4.4.2 Évaluation d'images réelles

Pour l'image-4.16, nous pouvons utiliser la méthode ci-dessus pour analyser les résultats de la segmentation de l'image. Comme le montre le tableau-4.4.2.

Les métriques présentées révèlent un modèle qui éprouve des difficultés significatives dans sa performance prédictive : avec une sensibilité de 42,7 %, il a du mal à identifier la majorité des cas positifs, et sa spécificité de 71,3 % indique un nombre encore élevé de faux positifs. La précision globale est à 61,84 %, ce qui indique que les prédictions du modèle ne sont guère meilleures que des suppositions aléatoires mais suggère un modèle viable qui peut être amélioré pour une meilleure performance. Le coefficient de corrélation de Matthews (MCC)

Metric	Value
Sensibilité	0.42
Spécificité	0.71
Précision globale	0.61
Précision	0.42
Score F1	0.42
MCC	0.14
Vrai Positif	85037.73
Vrai Négatif	283895.53
Faux Positif	113820.07
Faux Négatif	113820.07

TABLE 4.7 – Données d'évaluation pour une image réelle.

positive de 0,14 implique que les prédictions du modèle sont corrélées avec les résultats réels, ce qui indique que la performance du modèle n'est pas du fait du hasard.

— Image synthétique avec uniquement des équations :

82%

— Image synthétique avec équations et textes mais qui sont éloignés :

70%

— Image réelle :

61%

Après l'exploration et l'évaluation ci-dessus, nous avons constaté que notre modèle avait une grande précision pour un traitement simple de segmentation d'images, mais pour des situations très complexes, notre modèle

n'est pas applicable et la précision ne peut atteindre qu'environ 60 %. Nous devons encore affiner et améliorer le modèle.

## 4.5 Conclusion

En conclusion, ce chapitre a détaillé l'ensemble du processus d'expérimentation et les résultats obtenus dans le cadre du projet de segmentation de texte et d'équations sur des manuscrits de Leibniz. Les méthodes et les stratégies adoptées ont été présentées étape par étape, de la préparation des données à l'entraînement et l'évaluation du modèle, en mettant l'accent sur les défis rencontrés et les solutions apportées.

L'annotation minutieuse des données a constitué la première étape cruciale, permettant de distinguer clairement le texte des équations dans les manuscrits. Le processus d'augmentation des données a ensuite été mis en œuvre pour enrichir la base de données, améliorant ainsi la robustesse et l'efficacité du modèle d'apprentissage.

L'entraînement du modèle a été réalisé en ajustant les paramètres pour s'adapter à un modèle pré-entraîné, ce qui a permis d'améliorer significativement la précision de la segmentation. Les résultats obtenus montrent que le modèle est capable de distinguer efficacement le texte des équations, même si quelques erreurs subsistent, ce qui indique des pistes d'amélioration. L'évaluation finale du modèle a mis en évidence sa capacité à généraliser sur des données nouvelles, confirmant ainsi l'efficacité des techniques employées. Les tests de segmentation sur images synthétiques et réelles ont prouvé que le modèle peut être

appliqué avec succès pour la segmentation de contenu varié dans les manuscrits de Leibniz.



# Conclusion

A l’issue de ce travail, nous conclurons notre rapport par un bref résumé du travail effectué tout au long de ce projet de recherche et de développement. Nous ferons ensuite un bilan plus personnel quant aux enseignements que nous pouvons tirer de ce projet. Et nous finirons par expliciter quelques perspectives de recherches et de développement lié au travail effectué.

## 5.1 Résumé du travail effectué

Nous avons débuté notre projet par une phase d’immersion, afin de nous approprié de la problématique et nous familiariser avec les données des manuscrits de Leibniz. Cette étape initiale a été suivie d’une revue bibliographique approfondie, permettant de saisir les avancées et défis dans la détection et la segmentation des expressions mathématiques. Nous avons ensuite conceptualisé et mis en œuvre une solution innovante, exploitant un réseau neuronal convolutif pour traiter précisément ces expressions. La phase d’augmentation de données a été essen-

tielle, enrichissant notre base de données pour améliorer l’apprentissage et la validation de notre modèle. Après l’élaboration et l’entraînement de notre méthode basée sur l’apprentissage profond, nous avons procédé à une phase d’expérimentation rigoureuse. Cette phase a inclus l’entraînement du modèle sur un ensemble étendu de données augmentées, suivi par une série de tests d’évaluation pour mesurer l’efficacité de la segmentation et de la détection. Les résultats ont révélé une précision moyenne significative au niveau des pixels, démontrant la viabilité de notre approche. Ces expérimentations ont non seulement confirmé la pertinence de notre méthode mais ont également fourni des perspectives précieux pour l’amélioration continue du processus d’analyse des manuscrits anciens, marquant un progrès notable dans la compréhension et la préservation du patrimoine intellectuel de Leibniz.

## 5.2 Enseignements

Ce projet de recherche et développement sur les manuscrits de Leibniz nous a enseigné plusieurs leçons importantes. Premièrement, il a illustré la complexité inhérente à l'analyse des documents historiques, soulignant la nécessité d'une approche minutieuse et détaillée. Nous avons appris l'importance de l'état de l'art pour comprendre le contexte et orienter notre recherche. Ensuite, la mise en œuvre pratique, de la collecte de données à l'expérimentation, nous a confrontés aux défis techniques réels, renforçant notre compréhension théorique par une application concrète, nous avons pu renforcer nos compétences en technologies de préparation de données, maîtrisant des outils tels que LabelMe pour l'annotation et OpenCV pour le traitement d'images. En outre, notre travail avec l'intelligence artificielle a affiné nos compétences en modélisation et en évaluation des algorithmes, nous préparant à relever de futurs défis technologiques avec une base solide et diversifiée. Par ailleurs, cette expérience a également amélioré notre capacité à gérer des projets de recherche complexes, depuis la conceptualisation jusqu'à l'évaluation, et a renforcé nos compétences en matière de travail d'équipe et de communication scientifique. Enfin, ce projet de recherche et développement nous a également préparés pour d'éventuelles poursuites académiques ou professionnelles, en nous dotant d'une expérience concrète dans la résolution de problèmes complexes et le monde de la recherche.

## 5.3 Perspectives de recherche

Pour la suite de la recherche, il serait intéressant d'approfondir l'entraînement sur des jeux de données plus conséquents. Durant notre projet, nous avons constaté la limitation due au faible volume de données ; ainsi, travailler avec un ensemble plus large permettrait d'affiner les modèles. Pour améliorer l'analyse, il sera intéressant de se concentrer uniquement sur les pixels représentant l'encre. Cette focalisation permettrait de minimiser le bruit de fond et d'améliorer la clarté de l'information extraite. En isolant uniquement les pixels correspondant à l'encre, les algorithmes peuvent se concentrer sur l'essence du contenu manuscrit, améliorant ainsi la détection et la segmentation. Une telle approche améliorerait la qualité de l'analyse textuelle et pourrait révéler des détails subtils, souvent masqués ou altérés par l'hétérogénéité du support. En ayant recours à l'utilisation des pixels sombres uniquement, les résultats des évaluations seront également supérieurs et seront plus précis. Concernant les manuscrits de Leibniz, la prochaine étape logique serait d'étendre l'étude à la reconnaissance et à la transcription automatique, passant de la simple segmentation à l'analyse sémantique. L'utilisation de méthodes d'apprentissage avancées et l'extension à d'autres corpus enrichiraient significativement la recherche, tout en offrant de nouveaux outils pour les spécialistes désireux d'explorer ces textes historiques.

# Bibliographie

- [JAS<sup>+</sup>23] Sana Khamekhem Jemni, Sourour Ammar, Mohamed Ali Souibgui, Yousri Kessentini, and Abbas Cheddad. St-keys : Self-supervised transformer for keyword spotting in historical handwritten documents. *arXiv preprint arXiv:2303.03127*, 2023. 17
- [Kho19] Connor Shorten Taghi M. Khoshgoftaar. A survey on image data augmentation for deep learning. *Journal of Big Data*, 2019. 23
- [MMIEBA19] Olfa Mechi, Maroua Mehri, Rolf Ingold, and Najoua Essoukri Ben Amara. Text line segmentation in historical document images using an adaptive u-net architecture. In *2019 International Conference on Document Analysis and Recognition (ICDAR)*, Sousse, Tunisia ; Fribourg, Switzerland, 2019. IEEE. 14
- [RJVGGM18] Geoffrey Roman-Jimenez, Christian Viard-Gaudin, Adeline Granet, and Harold Mouchère. Transfer learning for structures spotting in unlabeled handwritten documents using randomly generated documents. In *International Conference on Pattern Recognition Applications and Methods*, 2018. 19
- [ST21] Bertrand B. Coüasnon Sophie Tardivel Solène Tarride, Aurélie Lemaitre. Combination of deep neural networks and logical rules for record segmentation in historical handwritten registers using few examples. *International Journal on Document Analysis and Recognition*, 32 :98–107, 2021. 22
- [TCH<sup>+</sup>15] Simon Thomas, Clément Chatelain, Laurent Heutte, Thierry Paquet, and Yousri Kessentini. A deep hmm model for multiple keywords spotting in handwritten documents. *Pattern Analysis and Applications*, 18 :1003–1015, 2015. 15
- [VTP23] Enrique Vidal, Alejandro H Toselli, and Joan Puigcerver. Lexicon-based probabilistic indexing of handwritten text images. *Neural Computing and Applications*, pages 1–20, 2023. 18
- [Wik] Wikipédia. Gottfried Wilhelm Leibniz. [https://fr.wikipedia.org/wiki/Gottfried\\_Wilhelm\\_Leibniz](https://fr.wikipedia.org/wiki/Gottfried_Wilhelm_Leibniz). Consulté le 28 Novembre 2023. 7

# Table des figures

1.1	Exemple d'image du manuscrit de Leibniz . . . . .	8
2.1	Résultats de repérage de plusieurs mots-clés obtenus grâce au modèle HMM . . . . .	16
2.2	Pipeline du cadiciel (framework) ST-KeyS proposé. Elle se compose de deux étapes : une étape de pré-formation et une étape de mise au point. . . . .	17
2.3	Exemple de repérage de structures réalisé sur des documents manuscrits réels. Les pixels (bleu), (vert) et (rouge) correspondent respectivement aux classes fond, nombre et mot. . . . .	22
3.1	Représentation graphique de l'architecture du réseau neuronal entièrement convolutif. Py(X) correspond à la représentation pyramidale de l'image d'entrée X avec 5 niveaux de résolutions. Une taille de filtre de $5 \times 5$ a été utilisée pour les couches de convolution et de convolution transposée. Chaque couche de convolution est associée à une couche de pooling maximum de $2 \times 2$ . Les deux premières couches de convolution transposées sont associées à une couche haut de gamme voisine la plus proche de $2 \times 2$ .[Source : “Transfer Learning for Structures Spotting in Unlabeled Handwritten Documents using Randomly Generated Documents,” Page 7] . . . . .	38
4.1	Exemple d'image du manuscrit de Leibniz avec 2 parties gauche et droite pour chaque page . . . . .	41
4.2	Annotation avec labelMe . . . . .	41
4.3	Image avec equations majoritaires . . . . .	43
4.4	Résultats de l'augmentation de donnée avec la coupure aléatoire. . . . .	44
4.5	Résultats de l'augmentation de donnée avec des données annotées . . . . .	45
4.6	Exemple d'image d'entraînement . . . . .	47
4.7	La courbe de perte d'apprentissage sans utiliser de modèle pré-entraîné. . . . .	50
4.8	La courbe de perte d'apprentissage du modèle pré-entraîné est utilisée. . . . .	50
4.9	Un résultat après entraînement sans utiliser le modèle pré-entraîné. . . . .	52
4.10	Exemple d'image synthétique contenant uniquement des équations. . . . .	53
4.11	Un exemple de segmentation d'une image contenant uniquement des équations. . . . .	53
4.12	Un exemple d'étiquetage incorrect d'une équation sous forme de texte. . . . .	54

4.13 Exemple d'image synthétique contenant des équations et de texte. . . . .	54
4.14 Un exemple de segmentation d'une image contenant des équations et de texte. . . . .	56
4.15 Un exemple d'image originale. . . . .	56
4.16 Un exemple de résultat de segmentation d'image de manuscrit original de Leibniz. . . . .	57
B.1 Planification prévisionnelle . . . . .	79
B.2 Planification mise à jour . . . . .	80
D.1 Points à contrôler sur le rapport à l'issue de la phase I . . . . .	94
D.2 Points à contrôler sur le travail à l'issue de la phase I . . . . .	94
D.3 Points à contrôler sur la projection à l'issue de la phase I . . . . .	95
D.4 Points à contrôler à l'issue de la phase II . . . . .	96
D.5 Points à contrôler à l'issue de la phase II . . . . .	97
D.6 Points à contrôler à l'issue de la phase II . . . . .	98

# Liste des tableaux

2.1	Tableau comparatif des méthodes de déformations de données (Liste non exhaustive) . . . . .	25
2.2	Tableau comparatif des méthodes de suréchantillonage pour la génération synthétique de données (Liste non exhaustive) . . . . .	26
3.1	Architecture of our model based on convolutional neural network . . . . .	39
4.1	Quantification des annotations Texte et Equation dans les manuscrits. . . . .	43
4.2	Quantification par pixel des textes et équations dans les manuscrits. . . . .	44
4.3	Résumé des données . . . . .	46
4.4	Résumé des métriques de classification par catégorie. . . . .	55
4.5	Données d'évaluation pour une image contenant uniquement des équations. . . . .	57
4.6	Données d'évaluation pour une image contenant des équations et du texte. . . . .	58
4.7	Données d'évaluation pour une image réelle. . . . .	59
C.1	Avancement du projet par rapport au temps de travail théorique minimal (respectivement haut) . . . . .	91



---

## Fiches de lecture

**Article 1 : “ST-KeyS : Self-Supervised Transformer for Keyword Spotting in Historical Handwritten Documents”.** Auteur : Sana Khamekhem Jemnia, Sourour Ammara, Mohamed Ali Souibguid, Yousri Kessentinia,b, Abbas Cheddade

**Résumé.** Dans l'article "ST-KeyS : Self-Supervised Transformer for Keyword Spotting in Historical Handwriting Documents", une nouvelle méthode utilisant l'apprentissage auto-supervisé pour la détection de mots-clés (KWS) dans des documents manuscrits historiques est proposée, utilisant un transformateur visuel basé sur le masqué. Le modèle d'encodeur automatique élimine le besoin de données étiquetées lors de la phase de préformation.

En effet, la recherche explore l'utilisation de l'apprentissage automatique pour repérer des mots-clés dans des documents historiques numérisés. Face au manque d'annotations pour les manuscrits historiques, ils proposent ST-KeyS, un modèle d'auto-apprentissage qui extrait des in-

formations utiles sans annotations humaines. Ce modèle surpassé les méthodes existantes en combinant l'auto-apprentissage avec une phase de réglage fin, montrant des performances supérieures dans la recherche de mots-clés dans des ensembles de données historiques. Les caractéristiques principales sont

- la pré-formation utilisant des auto-encodeurs masqués et des transformateurs visuels.
- Mise au point à l'aide des réseaux de neurones siamois.
- Exploitation de l'intégration d'un histogramme pyramidal de caractères (PHOC).
- Efficacité démontrée sur trois ensembles de données de référence, surpassant les méthodes de pointe.

Un processus de reconnaissance de mots qui ne repose pas sur la reconnaissance complète du texte mais plutôt sur la mise en correspondance d'images de mots est proposé. Les transformateurs en PNL ont conduit à leur application dans des tâches de vision par ordinateur, mais ils

nécessitent de grands ensembles de données. Approche proposée est donc composé de :

- Phase de pré-formation : des modèles d'encodeurs automatiques masqués sont utilisés pour apprendre des représentations approfondies à partir de données non étiquetées.
- Étape de réglage fin : l'encodeur pré-entraîné est intégré dans un réseau neuronal siamois pour l'extraction de fonctionnalités et amélioré avec les intégrations PHOC.

**Analyse** Cet article est intéressant, car il évoque une solution pour le repérage de mots-clés (KWS) dans les documents historiques en prenant en compte le problème de la rareté des données, qui est le cas des manuscrits de Leibniz à notre disposition. Le principe repose sur l'utilisation de l'apprentissage auto-supervisé pour extraire des représentations utiles des données d'entrée sans recourir aux annotations humaines, puis utiliser ces représentations dans la tâche en aval.

**Article 2 : “A Deep HMM model for multiple keywords spotting in handwritten documents”.** Auteurs : Simon Thomas, Clément Chatelain, Laurent Heutte, Thierry Paquet, Yousri Kessentini

L'article démontre l'efficacité du Deep HMM model pour extraire des mots-clés dans des documents manuscrits, en particulier dans des styles d'écriture ambigus ou illisibles.

**Résumé.** Dans cet article, nous proposons un système de recherche de mots-clés par requête, capable d'extraire des mots-clés arbitraires dans des documents manuscrits, en prenant à la fois des décisions de segmentation et de reconnaissance au niveau de la ligne. Le système repose sur la combinaison d'un modèle de ligne HMM composé de modèles de mots-clés et non-mots-clés (de remplissage), avec un réseau neuronal profond (DNN) qui estime les probabilités d'observation dépendantes de l'état. Des expériences sont menées sur la base de données RIMES, une base de données de documents manuscrits sans contrainte utilisée pour comparer différentes tâches de reconnaissance d'écriture manuscrite. Les résultats obtenus montrent la supériorité du framework proposé sur les architectures hybrides GMM-HMM classique et HMM standard. Dans l'article « Handwriting Keyword Spotting Using Deep Neural Networks and Certainty Prediction », une méthode d'utilisation de réseaux neuronaux convolutifs profonds (CNN) pour détecter et identifier des mots-clés dans des documents manuscrits est proposée. Cette méthode convient aux scénarios dans lesquels des informations doivent être récupérées à partir de documents manuscrits, tels que l'analyse de documents historiques, la recherche de documents juridiques, etc.

**Innovation :** L'introduction du dropout de Monte-Carlo comme mesure de l'incertitude améliore non seulement les performances du modèle, mais améliore également la précision de la reconnaissance des mots clés. Cette approche peut être appliquée à la fois à la requête par exemple et à la requête par contexte.

Architecture d'apprentissage profond : le CNN utilisé dans l'article peut extraire efficacement des fonctionnalités d'un texte manuscrit. Les CNN se sont révélés très efficaces dans le domaine de la reconnaissance d'images et de textes, notamment lorsqu'il s'agit de motifs complexes ou irréguliers.

Mesure de l'incertitude : grâce à l'abandon de Monte-Carlo, le modèle est capable d'estimer l'incertitude lors de l'extraction de caractéristiques, ce qui est très important pour gérer des styles d'écriture de manuscrits divers et irréguliers.

**Analyse** Cet article est intéressant pour notre étude, car il propose un système de recherche de mots-clés par requête, capable d'extraire des mots-clés arbitraires dans des documents manuscrits, en prenant à la fois des décisions de segmentation et de reconnaissance au niveau de la ligne. Les mots ou les phrases dans le texte manuscrit peuvent être identifiés et classés efficacement, ce qui est crucial pour le traitement de manuscrits tels que celui de Leibniz. Cependant, il se peut qu'il ne soit pas explicitement conçu pour faire la différence entre les zones de texte et les zones d'expression.

**Article 3 : “Lexicon-based probabilistic indexing of handwritten text images” .** Auteurs : Enrique Vidal, Alejandro H. Toselli1, Joan Puigcerver

**Résumé.** Le Keyword Spotting (KWS) est ici considéré comme une technologie de base pour l'indexation probabiliste (PrIx) de grandes collections d'images de

texte manuscrites afin de permettre un accès textuel rapide au contenu de ces collections. Dans cette perspective, un cadre probabiliste pour les KWS basés sur le lexique dans les images textuelles est présenté. La présentation vise à fournir des informations formelles qui aident à comprendre les déclarations classiques de KWS (dont PrIx emprunte des concepts fondamentaux), ainsi que les défis relatifs qu'impliquent ces déclarations. Le développement du cadre proposé montre clairement que la reconnaissance ou la classification des mots soutient implicitement ou explicitement toute formulation de KWS. De plus, cela suggère que les mêmes modèles statistiques et méthodes de formation utilisés avec succès pour la reconnaissance de texte manuscrit peuvent également être avantageusement utilisés pour PrIx, même si PrIx ne nécessite généralement ni ne s'appuie sur aucun type de transcriptions d'images produites précédemment. Les expérimentations menées à partir de ces approches confortent la cohérence et l'intérêt général du cadre proposé. Les résultats sur trois ensembles de données traditionnellement utilisés pour l'analyse comparative KWS sont nettement meilleurs que ceux précédemment publiés pour ces ensembles de données. En outre, de bons résultats sont également rapportés sur deux nouveaux ensembles de données d'images de texte manuscrit plus grands (BENTHAM et PLANTAS), montrant le grand potentiel des méthodes proposées dans cet article pour l'indexation et la recherche textuelle dans de grandes collections de documents manuscrits non transcrits. L'idée principale de cet article est d'améliorer les capacités d'accès rapide au texte pour les collections d'images de texte

manuscrit à grande échelle en développant une technologie KWS (Keyword Spotting) basée sur les probabilités. L'article propose un cadre appelé indexation probabiliste (PrIx) pour la récupération et l'indexation efficaces d'images de textes manuscrits non transcrits. Cette méthode permet aux utilisateurs de localiser rapidement des emplacements dans une image pouvant contenir des mots-clés spécifiques sans effectuer de transcription du texte intégral. L'idée centrale de l'article est que grâce à cette approche innovante, l'efficacité et la précision du traitement des collections de documents manuscrits peuvent être considérablement améliorées, en particulier pour les styles diversifiés et à grande échelle d'images de textes manuscrits historiques.

#### Méthodes :

- Cadre probabiliste : L'article propose un cadre probabiliste pour la localisation de mots-clés (KWS) basée sur le vocabulaire dans les images textuelles, en particulier pour les images textuelles manuscrites.
- Indexation probabiliste (PrIx) : Un cadre de recherche et de récupération appelé indexation probabiliste (PrIx) a été développé pour traiter des images de textes manuscrits non transcrits.
- Technologie de localisation par mots clés : utilisez la technologie de localisation par mots clés pour identifier les emplacements dans une collection d'images pouvant contenir des mots de requête sans avoir besoin d'une transcription du texte intégral.
- Modèle de compromis précision-rappel : met en

œuvre un cadre de recherche et de récupération flexible qui permet d'obtenir différents compromis de précision et de rappel en définissant des seuils pour contrôler la probabilité que des mots-clés apparaissent dans les zones d'image.

Les résultats montrent une amélioration significative des performances : les expériences sur les ensembles de données traditionnellement utilisés pour les benchmarks KWS montrent que les résultats de cette méthode sont nettement meilleurs que les résultats publiés précédemment. Ils démontrent également un large potentiel d'application : de bons résultats sont également rapportés sur deux nouveaux ensembles de données d'images de texte manuscrit plus grands, démontrant le potentiel de cette approche dans l'indexation et la recherche de texte de grandes collections de documents manuscrits. Avantages :

- Efficacité améliorée : par rapport à la reconnaissance complète de texte manuscrit, la méthode PrIx améliore l'efficacité du traitement de grandes collections d'images de texte manuscrit.
- Forte applicabilité : Capable de gérer des collections vastes et diversifiées d'images de texte manuscrites, améliorant ainsi l'accessibilité et la convivialité de ces documents.
- Flexibilité : grâce au modèle de compromis précision-rappel, les utilisateurs peuvent ajuster l'exactitude et l'exhaustivité des résultats de recherche en fonction de leurs besoins.

#### Inconvénients :

- Dépendance à des données de haute qualité : la mé-

- thode proposée dans l'article peut ne pas fonctionner correctement sur des images de texte manuscrites de moindre qualité.
- Coût de calcul : bien que plus efficaces que les méthodes de reconnaissance de texte manuscrit entièrement automatiques, les coûts de calcul peuvent néanmoins être relativement élevés.
  - Limites de précision : même si elle offre des performances de recherche améliorées, cette méthode peut ne pas permettre une reconnaissance de texte totalement précise.

**Analyse** Cette méthode est particulièrement adaptée au traitement de grandes quantités de documents manuscrits non transcrits, offrant un moyen puissant d'indexer et de rechercher des mots-clés dans les manuscrits. Cependant, il semble se concentrer davantage sur l'indexation du texte plutôt que sur la distinction entre le texte et les expressions. Mais nous pouvons indexer l'expression uniquement pour les symboles qui apparaissent dans l'expression et obtenir son aire via le cadre de délimitation.

**Article 4 : A survey on Image Data Augmentation for Deep Learning .** Auteurs : Connor Shorten et Taghi M. Khoshgoftaar

**Résumé.** L'article explore de manière approfondie le domaine complexe de l'augmentation de données pour les images, mettant en lumière son rôle crucial dans

la formation des modèles d'apprentissage profond. Son principal objectif est de surmonter le défi du surajustement, l'équilibrage des classes et les problèmes d'imbalance fréquemment rencontré lorsqu'on dispose de jeux de données limités, et vise à utiliser l'augmentation de données pour simuler des ensembles de données plus vastes. Il aborde d'abord deux approches majeures d'augmentation de données : la déformation des données, consistant à altérer les images de manière contrôlée, et le suréchantillonnage, qui implique d'augmenter la fréquence des données d'une classe spécifique. Il examine l'impact de la résolution des images sur la performance des modèles, soulignant la balance entre la qualité des données et les contraintes computationnelles. Pour la partie la déformation des données, l'article analyse en détail des techniques telles que la rotation, le recadrage aléatoire, et le mélange de pixels. Il explore également la combinaison de différentes techniques de déformation pour améliorer la diversité des données, offrant ainsi des perspectives sur la création d'ensembles de données plus riches. De l'autre côté, En ce qui concerne le suréchantillonnage, l'article présente des méthodes basées sur l'apprentissage profond, y compris l'entraînement adversarial et la rétropropagation. Il compare les avantages et les inconvénients de ces techniques, offrant une évaluation approfondie de leur pertinence dans le contexte de l'augmentation de données. Il analyse l'impact de la résolution des images sur la performance des modèles, proposant des comparaisons entre différentes résolutions pour former des modèles d'ensemble. La résolution des images s'avère être un facteur déterminant. Les images

haute résolution nécessitent plus de ressources, tandis que la baisse de résolution peut entraîner une perte d'informations cruciales. Les chercheurs ont constaté que combiner des modèles formés sur des résolutions différentes donne de meilleurs résultats. L'article aborde également la question cruciale de la taille finale de l'ensemble de données après augmentation, en considérant les compromis entre mémoire et calcul. Enfin, l'article conclut sur une note optimiste, soulignant que l'avenir de l'augmentation de données est prometteur. Il appelle à des recherches futures sur la création de référentiels, l'amélioration de la qualité des données générées par GANs, et l'extension des principes d'augmentation à d'autres types de données que les images. En résumé, l'article offre une perspective approfondie sur l'importance, les techniques et les défis à venir de l'augmentation de données pour les images, sans omettre de souligner son rôle crucial dans le domaine de l'apprentissage profond.

**Analyse** Cet article est intéressant, car il explore et compare les différentes techniques d'augmentation de données avec les avantages et inconvénients. Ce qui facilite nos recherches afin de nous orienter dans les recherches de solution adéquates afin d'augmenter nos données.

**Article 5 : “Combination of deep neural networks and logical rules for record segmentation in historical handwritten registers using few examples” .** Auteurs : Sôlène Tarride, Aurélie Lemaitre, Bertrand Coüasnon et Sophie Tardivel

**Résumé.** Cet article se concentre sur l'analyse de la mise en page de registres manuscrits historiques, spécifiquement ceux consignant des cérémonies religieuses locales. L'objectif principal est de délimiter chaque enregistrement dans ces registres. Deux approches sont présentées dans l'article : 1. Réseaux de Détection d'Objets : Trois architectures de pointe sont comparées, avec une attention particulière portée à Mask R-CNN, qui offre les meilleures performances dans les expérimentations. Cependant, cette approche montre des limitations, notamment en termes de généralisation sur des documents hétérogènes. 2- Deep Syntax : Une approche hybride originale combinant des réseaux en U avec des règles logiques. Deep Syntax s'appuie sur des motifs récurrents tels que les frontières de page, les lignes de texte initiales et les signatures pour délimiter chaque enregistrement. Malgré sa complexité, Deep Syntax offre des résultats solides, même avec un ensemble d'entraînement limité.

Les résultats suggèrent en effet que les deux systèmes deviennent efficaces à partir de 25 documents d'entraînement, ce qui correspond à environ 250 enregistrements. Cette rapide acquisition de connaissances peut s'expliquer par l'homogénéité des enregistrements, car il n'y a qu'un seul rédacteur sur une courte période. En ce qui concerne la Deep Syntax, les expériences montrent que le symbole fiscal est bien appris à partir de 10 images, tandis que les premières lignes de texte sont bien apprises à partir de 25 images.

**Évaluation :** Les deux approches sont évaluées sur 3708 registres français (16-18ème siècles) et sur une base de données espagnole contenant 25 documents d'entraî-

nement, ce qui correspond approximativement à 253 registres (17ème siècle). Alors que les deux systèmes performant bien sur des documents homogènes, Deep Syntax dépasse Mask R-CNN sur des documents hétérogènes, notamment lorsque l'entraînement est effectué sur un sous-ensemble non-représentatif. Deep Syntax montre une meilleure capacité de généralisation, produisant 15% de configurations correspondantes supplémentaires et réduisant l'erreur de surface de la ZoneMap de 30%, même avec un ensemble d'entraînement trois fois plus petit.

**Analyse** Les réseaux de Détection d'Objets démontrent une performance sur documents homogènes comme démontrés par les résultats sur la base de données Esposalles. Ils fournissent une détection précise des objets grâce à l'utilisation des boîtes englobantes. Cependant, on démontre une sensibilité à la Variabilité des Documents qui montre une baisse significative de performance sur des documents hétérogènes, notamment lorsqu'entraîné sur un sous-ensemble non représentatif. Le modèle requiert un ensemble d'entraînement substantiel et représentatif, ce qui peut être difficile à obtenir pour des documents historiques. Parallèlement, le Deep Syntax (Approche Hybride) montre une capacité de généralisation élevée avec un ensemble d'entraînement réduit, soulignant son adaptabilité à des documents variés. Il exploite des motifs structurels récurrents tels que les frontières de page et les lignes de texte, simplifiant la tâche de détection. Cependant, la méthode Deep Syntax est décrite comme complexe, ce qui peut rendre son application et sa compréhension plus difficiles. En plus, les boîtes en-

globantes générées par Deep Syntax sont contraintes par des règles logiques et peuvent ne pas s'adapter à la spécificité de chaque enregistrement.

Par rapport à notre étude sur les manuscrits de Leibniz, les deux méthodes auraient pu être intéressantes si on avait un peu plus de données d'entraînement. La complexité est un facteur important à prendre en compte également que ce soit en termes de coût computationnel ou encore en terme de temps car le PRED reste un projet académique dont le temps de développement dure 9 semaines au plus.

**Article 6 : “Transfer Learning for Structures Spotting in Unlabeled Handwritten Documents using Randomly Generated Document” .** Auteurs : Geoffrey Roman-Jimenez, Christian Viard-Gaudin, Adeline Granet, Harold Mouchère

**Résumé.** Malgré les avancées dans la reconnaissance de texte manuscrit, l'analyse des documents manuscrits historiques reste difficile en raison du besoin de vastes bases de données annotées. Pour remédier à cette lacune, l'étude explore une approche novatrice basée sur le transfert de connaissances pour traiter de nouvelles collections de documents non annotés.

Le défi de localiser les structures au niveau des mots, en distinguant les mots des chiffres dans des documents manuscrits non annotés, est au cœur de cette étude. L'approche adoptée repose sur le transfert transductif d'un réseau neuronal convolutionnel profond pré-entraîné sur des images artificielles générées aléatoirement. Le

modèle, un réseau neuronal entièrement convolutionnel (FCNN), réalise une classification pixel par pixel en trois classes : fond, chiffre et mot. L’entraînement du modèle a été effectué sur 100 000 images artificielles étiquetées générées de manière aléatoire. Elles sont construites avec des patchs de chiffres/mots de la base de données IRO-nOff et des arrière-plans de pages de la base de données RECITAL.

La phase d’évaluation se déroule en trois étapes cruciales. Tout d’abord, en considérant l’ensemble d’images artificielles générées et les cartes de classification correspondantes. Ensuite, en évaluant la détection des chiffres parmi les patches de la base de données RIMES, avec une attention particulière à la distinction entre chiffres et mots. L’évaluation inclut des mesures telles que la précision, le rappel et l’exactitude, avec l’extension du coefficient de corrélation de Matthews pour la classification multiclasse.

Les résultats démontrent la transférabilité réussie du modèle vers des documents réels, soulignant sa capacité à distinguer des structures spécifiques sans annotations préalables. En conclusion, l’étude propose une approche prometteuse basée sur le transfert d’apprentissage pour la détection de structures dans des documents manuscrits non annotés. Ces résultats encourageants ouvrent la voie à des applications potentielles dans l’analyse de documents historiques, tandis que des améliorations futures pourraient inclure une plus grande variabilité dans la génération de documents artificiels.

**Analyse** La méthode présentée dans ce document se concentre sur la détection de structures dans des documents manuscrits non étiquetés en utilisant un paradigme d’apprentissage par transfert avec un réseau neuronal convolutionnel profond. Elle est particulièrement adaptée pour distinguer les mots des chiffres dans des documents sans étiquette en générant artificiellement des images avec des structures de mots et de chiffres. Cette méthode pourrait être bien adaptée pour la détection d’expressions mathématiques en raison de sa capacité à identifier des structures spécifiques et à généraliser à partir d’exemples synthétiques. Cependant, elle pourrait ne pas être optimale pour les expressions mathématiques complexes qui nécessitent la compréhension de relations spatiales ou de symboles mathématiques spécifiques non représentés dans le jeu de données d’entraînement, en raison de la diversité et de la complexité des expressions mathématiques par rapport aux mots et chiffres simples.

**Article 7 : “Text Line Segmentation in Historical Document Images Using an Adaptive U-Net Architecture”.**  
Auteurs : Olfa Mechi, Maroua Mehri, Rolf Ingold†, Najoua Essoukri Ben Amara

**Résumé.** Ce papier parle d’une architecture U-Net adaptative pour la segmentation des lignes de texte dans les images de documents historiques. La méthode proposée vise à résoudre des défis liés à l’efficacité computationnelle et à la gestion des tailles d’image variables. L’architecture U-Net, reconnue pour son succès dans les tâches de segmentation sémantique, est adaptée et opti-

misée pour les besoins spécifiques de l'analyse de documents historiques.

L'architecture U-Net adaptative implique des changements dans le chemin de contraction, en réduisant notamment le nombre de filtres pour optimiser les paramètres et résoudre des problèmes tels que le surajustement. Les modifications proposées entraînent une réduction significative du nombre de paramètres par rapport à l'architecture U-Net originale, ce qui se traduit par une meilleure efficacité mémoire, un temps de traitement réduit et une complexité numérique améliorée pendant la phase d'entraînement.

L'évaluation de la méthode proposée implique l'utilisation de divers ensembles de données de référence, notamment READ1, cBAD2, DIVA-HisDB3 et un ensemble de données privé provenant de ANT4. La définition de la vérité terrain est basée sur la représentation de la hauteur X, choisie pour ses avantages dans les tâches de reconnaissance de texte. Les métriques d'évaluation des performances comprennent la précision (P), le rappel (R), la mesure F (F), l'intersection sur l'union (IoU) et l'aire sous la courbe (AUC).

Le protocole expérimental inclut une pré-formation sur un grand ensemble de données privé et un ajustement fin sur l'ensemble de données cBAD. L'architecture U-Net est mise en œuvre à l'aide du framework Keras et d'un GPU TITAN X. Les résultats mettent en évidence l'efficacité de l'architecture U-Net adaptative, avec une amélioration des mesures F et IoU par rapport à l'U-Net original. L'évaluation inclut également des comparaisons avec une méthode de pointe de Renton et al., démontrant

des performances supérieures en termes de précision et de mesure F sur l'ensemble de données cBAD.

La conclusion met en avant l'efficacité qualitative et quantitative de la méthode proposée, soulignant sa robustesse pour différentes langues et mises en page dans les images de documents historiques. Des travaux futurs sont envisagés, notamment une évaluation plus poussée sur un ensemble plus important d'images de documents manuscrits arabes et des efforts en cours pour étendre le cadre à la segmentation de sous-mots et à la reconnaissance de mots.

Dans l'ensemble, le document offre une exploration approfondie de l'architecture U-Net adaptative proposée, étayée par des résultats expérimentaux approfondis et une feuille de route claire pour des améliorations futures.

**Analyse** L'adaptation de l'architecture U-Net pour la segmentation des lignes de texte dans les images de documents historiques offre des avantages prometteurs pour la détection et la segmentation des expressions mathématiques notamment l'efficacité améliorée et la capacité à gérer des images de tailles variables. La réduction du nombre de filtres et l'optimisation des paramètres contribuent à une meilleure efficacité mémoire et un temps de traitement réduit, ce qui est crucial pour la segmentation précise des expressions mathématiques complexes. De plus, la robustesse de l'architecture face à différentes langues et mises en page indique son potentiel pour être appliquée à des documents historiques contenant des expressions mathématiques variées, insérées dans des contextes textuels ou des mises en page

complexes. Cependant, l'architecture U-Net adaptative a été spécifiquement conçue pour la segmentation de lignes de texte, ce qui peut limiter son application directe aux expressions mathématiques. Les symboles et les équations mathématiques, avec leurs dispositions spatiales uniques, pourraient nécessiter des adaptations supplémentaires ou des mécanismes de reconnaissance spécifiques qui ne sont pas intégrés dans une approche principalement axée sur le texte. En outre, la segmentation efficace des expressions mathématiques peut exiger des annotations détaillées spécifiques aux symboles mathématiques, un défi non adressé par les méthodes basées sur la segmentation de texte. La nécessité de développer des ensembles de données annotés spécifiquement pour les expressions mathématiques souligne cette limitation. Enfin, bien que l'article montre que l'architecture est testée sur des documents en différentes langues, y compris l'arabe et le latin, la généralisation à des documents historiques non-latins ou contenant des notations mathématiques anciennes ou régionales pourrait représenter un défi supplémentaire.

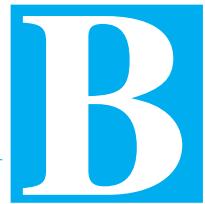
**Article 8 : “SpottingNet : Learning the Similarity of Word Images with Convolutional Neural Network for Word Spotting in Handwritten Historical Documents”.**

Auteurs : Zhuoyao Zhong, Pan Weishen, Harold Mouchère, Christian Viard-Gaudin, Jin Lianwen

**Résumé.** La détection de mots est un processus de récupération basé sur le contenu qui permet d'obtenir une liste classée d'images de mots candidats similaires

au mot recherché dans les images de documents numériques. Dans cet article, une approche innovante basée sur un réseau neuronal convolutif (CNN) pour la détection de mots par requête par exemple (QBE) dans des documents historiques manuscrits est présentée. Cette méthode intègre un apprentissage simultané des descripteurs d'images de mots et une évaluation de la mesure de similarité entre eux, sans recourir à des méthodes de reconnaissance ou à des informations préalables sur les catégories de mots. L'approche se distingue par l'intégration de modèles hybrides de classification et de régression d'apprentissage profond, une nouvelle technique de fusion de scores de similarité, et une méthode de génération d'échantillons basée sur la gigue de localisation pour créer un ensemble de données équilibré et vaste. L'efficacité de cette méthodologie est validée par des expériences sur l'ensemble de données George Washington (GW), où elle atteint une précision moyenne (mAP) remarquable de 80,03%, dépassant nettement les performances des approches précédentes. En explorant trois architectures CNN différentes et en ajustant les fonctions de perte, l'article démontre l'avantage du réseau à 2 canaux sur les réseaux siamois et pseudo-siamois, et l'adéquation de la fonction de perte softmax pour cette tâche spécifique. La fusion de scores proposée améliore significativement les performances, offrant une méthode robuste et efficace pour le repérage de mots dans les manuscrits historiques, en dépit de leur variabilité et confusion inhérentes.

**Analyse** La méthode décrite dans le document pour détecter des mots spécifiques dans des manuscrits utilise une approche basée sur le réseau neuronal convolutif (CNN) pour la recherche de mots par exemple (QBE) dans des documents historiques manuscrits. Cependant, cette méthode pourrait ne pas être adaptée pour la détection d'expressions mathématiques telles que les formules, les équations, les notations, les expressions littérales et les démonstrations, car ces éléments impliquent souvent une combinaison de symboles, de nombres et d'arrangements spatiaux qui sont significativement différents du contenu textuel. Les CNN entraînés pour la reconnaissance de mots sont optimisés pour reconnaître des motifs dans des données textuelles, tandis que les expressions mathématiques nécessitent de comprendre les relations entre les symboles et leur disposition spatiale, ce que ces modèles ne capturent pas efficacement.



---

## Planification

La figure B.1 présente le planning élaboré *a priori* qu'on essayera de suivre pendant ce projet de recherche et développement.

Activité	Temporalité																			
	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20
Choix du sujet																				
Acquisition et reformulation																				
Etude bibliographique																				
Analyse et comparaison des méthodes																				
Soutenance et Livrable 1																				
Conception détaillée																				
Validation du choix																				
Préparation de donnée																				
Augmentation des données																				
Implémentation de la solution																				
Tests																				
Implémentation de la méthode d'évaluation																				
Evaluation																				
Analyse des résultats																				
Finalisation du projet																				
Documentation																				
Soutenance et Livrable finale																				
Recette																				

FIGURE B.1 – Planification prévisionnelle

Activité	Temporalité																			
	S43	S44	S45	S46	S47	S48	S49	S50	S51	S52	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
Choix du sujet																				
Acquisition et reformulation																				
Etude bibliographique																				
Analyse et comparaison des méthodes																				
Soutenance et Livrable 1																				
Conception détaillée																				
Validation du choix																				
Annotation des données																				
Préparation de donnée																				
Implémentation de la méthode d'augmentation de donnée																				
Augmentation des données																				
Constitution de la base de donnée																				
Correction et amélioration de la méthode d'augmentation de donnée																				
Implementation de la solution pour la détection et la segmentation																				
Entrainement du modèle																				
Correction et amélioration de la solution																				
Tests																				
Implémentation de la méthode d'évaluation																				
Evaluation																				
Correction et amélioration de la méthode d'évaluation																				
Analyse des résultats																				
Finalisation du projet																				
Documentation																				
Soutenance et Livrable finale																				
Recette																				

FIGURE B.2 – Planification mise à jour



# Fiches de suivi

••• Cette annexe est *obligatoire*.

---

## Fiche de suivi de la semaine 1 du 23 Octobre 2023 au 29 Octobre 2023

---

Temps de travail de Mamisoa RANDRIANARIMA-NANA: 6 h 30 m

Temps de travail de Xianxiang ZHANG: 6 h 30 m

### Travail effectué.

- tâche 1 : Prise de connaissance du sujet ;  $t\%$  ; etc. ;
- tâche 2 : Réunion avec les commanditaire ;
- tâche 3 : Reformulation, acquisition du sujet.

### Échanges avec le commanditaire.

- Echange avec les encadrants ;
- Echange avec les historiens en charge du projet de recherche Edition des manuscrits de Leibniz ;
- éléments de clarification, compréhension ;
- choix, orientations, redéfinitions ;

### Planification pour la semaine prochaine.

- Exploration bibliographique ;
- Lecture de l'article « Multiple Document Datasets Pre-training Improves Text Line Detection With Deep Neural Networks » ;;

---

## Fiche de suivi de la semaine 2 du 6 Novembre 2023 au 12 Novembre 2023

---

Temps de travail de Mamisoa RANDRIANARIMA-NANA: 4 h 00 m

Temps de travail de Xianxiang ZHANG: 10 h 5 m

### Travail effectué.

- tâche 1 : lecture article 1 : “Multiple Document Datasets Pre-training Improves Text Line Detection With Deep Neural Networks”

- Auteurs : Melodie Boillet ,Christopher Kermorvant, Thierry Paquet.;
- tâche 2 : lecture article 2 : “Transfer Learning for Structures Spotting in Unlabeled Handwritten Documents using Randomly Generated Document”  
Auteurs : Geoffrey Roman-Jimenez, Christian Viard-Gaudin, Adeline Granet, Harold Mouchère
  - tâche 3 : lecture article 3 : Recognition of Online Handwritten Mathematical Expressions Using Convolutional Neural Networks  
Auteurs : Catherine Lu, Karanveer Mohan

#### **Travail non effectué.**

#### **Échanges avec le commanditaire.**

Mise à plat de l' Objectif : Detecter et segmenter automatiquement les expressions dans le manuscrit de Leibniz ;

Préparation des données : Mise à disposition d'une cinquantaine de donnée données sous formats Tex et pdf qui vont être préparer par David et Vincent ;

#### **Planification pour la semaine prochaine.**

- tâche 1 : lecture article "Deep CNN-based Segmentation of Handwritten Mathematical Expressions"  
Auteur(s) : S. Kumar, M. Saini, R. Singh
- tâche 2 : lecture article "Mathematical Expression Recognition Using a Hierarchical Classifier Based on Geometric and Structural Information"  
Auteur(s) : T. M. Rathore, A. G. Ramakrishnan, A. Agrawa
- tâche 3 : lecture article “Offline Recognition of Handwritten Mathematical Expressions Using

CNN and Xception”  
Auteur : Shreedevi Biradar1 , Dr.Manikamma2, Jyoti Neginal3 , Dr.Gajendran Malshetty4

- tâche 4 : Veille technologique Python

### **Fiche de suivi de la semaine 3 du 13 Novembre 2023 au 19 Novembre 2023**

Temps de travail de Mamisoa RANDRIANARIMANANA: 14 h 30 m

Temps de travail de Xianxiang ZHANG: 11 h 20 m

#### **Travail effectué.**

- tâche 1 : lecture article "Article : “Text Line Segmentation for Challenging Handwritten Document Images Using Fully Convolutional Network ”  
Auteurs : Berat Barakat, Ahmad Droby, Majeed Kassis and Jihad El-Sana
- tâche 2 : lecture article “ Segmentation Methods for Hand Written Character Recognition ”  
Auteurs : Namrata Dave
- tâche 3 : lecture article“ Lexicon-based probabilistic indexing of handwritten text images”  
Auteurs : Enrique Vidal, Alejandro H. Toselli, Joan Puigcerver
- tâche 4 : lecture article “Keyword spotting in histo-

- rical handwritten documents based on graph matching ” Auteurs : Michael Stauffe, Andreas Fischer, Kaspar Riese
- tâche 5 : lecture article "A survey on Image Data Augmentation for Deep Learning"  
Auteurs : Connor Shorten and Taghi M. Khoshgoftaar

### **Échanges avec le commanditaire.**

Orientation de l'étude bibliographique vers l'augmentation de donnée et la détection de mot clé Nous avons divisés les taches respectives : Xianxiang fera des recherches orientés vers les “Keywords Spotting” et Mamisoa sur le “Data Augmentation”

### **Planification pour la semaine prochaine.**

- tâche 1 : lecture article " A Survey of Data Augmentation Approaches for NLP"  
Auteurs : Steven Y.Feng , Varun Gangal, Jason Wei† , Sarath Chandar, Soroush Vosoughi,Teruko Mitamura, Eduard Hovy1"
- tâche 2 : lecture article "Mathematical Expression Recognition Using a Hierarchical Classifier Based on Geometric and Structural Information"  
Auteur(s) : T. M. Rathore, A. G. Ramakrishnan, A. Agrawa
- tâche 3 : lecture article “Offline Recognition of Handwritten Mathematical Expressions Using CNN and Xception”  
Auteur : Shreedevi Biradar1 , Dr.Manikamma2, Jyoti Neginal3 , Dr.Gajendran Malshetty4

- tâche 3 : lecture article "Combination of deep neural networks and logical rules for record segmentation in historical handwritten registers using few examples"  
Auteurs : Solène Tarride, Aurélie Lemaitre, Bertrand Coüasnon, Sophie Tardivel
- tâche 4 : Analyse et comparaison des différentes méthodes
- tâche 5 : Rédaction du livrable

### **Fiche de suivi de la semaine 4 du 20 Novembre 2023 au 26 Novembre 2023**

Temps de travail de Mamisoa RANDRIANARIMANA: 12 h 30 m

Temps de travail de Xianxiang ZHANG: 11 h 00 m

### **Travail effectué.**

- tâche 1 : lecture article " A Survey of Data Augmentation Approaches for NLP"  
Auteurs : Steven Y.Feng , Varun Gangal, Jason Wei† , Sarath Chandar, Soroush Vosoughi,Teruko Mitamura, Eduard Hovy1"
- tâche 2 : lecture article "Mathematical Expression Recognition Using a Hierarchical Classifier Based on Geometric and Structural Information"

- Auteur(s) : T. M. Rathore, A. G. Ramakrishnan, A. Agrawa
- tâche 3 : lecture article “Offline Recognition of Handwritten Mathematical Expressions Using CNN and Xception”  
Auteur : Shreedevi Biradar1 , Dr.Manikamma2, Jyoti Neginal3 , Dr.Gajendran Malshetty4
  - tâche 3 : lecture article "Combination of deep neural networks and logical rules for record segmentation in historical handwritten registers using few examples"  
Auteurs : Solène Tarride, Aurélie Lemaitre, Bertrand Coüasnon, Sophie Tardivel
  - tâche 4 : Analyse et comparaison des différentes méthodes
  - tâche 5 : Rédaction du livrable

#### **Travail non effectué.**

#### **Échanges avec le commanditaire.**

Discussion sur les différents méthodes trouvés, ceux à retenir, ceux à enlever et ceux à approfondir.

#### **Planification pour la semaine prochaine.**

- tâche 1 : Finalisation de l'état de l'art : comparaison, synthèse et récapitulatif des différents méthodes
- tâche 2 : Rédaction du livrable
- tâche 3 : Préparation de la soutenance

---

#### **Fiche de suivi de la semaine 5 du 27 Novembre 2023 au 3 Décembre 2023**

---

Temps de travail de Mamisoa RANDRIANARIMANA: 18 h 00 m

Temps de travail de Xianxiang ZHANG: 18 h 30 m

#### **Travail effectué.**

- tâche 1 : Finalisation de l'état de l'art : comparaison, synthèse et récapitulatif des différents méthodes
- tâche 2 : Rédaction du livrable
- tâche 3 : Préparation de la soutenance

#### **Échanges avec le commanditaire.**

Discussion sur le contenu du livrable et de la présentation

#### **Planification pour la semaine prochaine.**

- tâche 1 : Soutenance
- tâche 2 : Validation des choix avec les commanditaires.
- tâche 3 : Conception détaillée.
- tâche 4 : Installation de l'environnement de travail.

---

#### **Fiche de suivi de la semaine 6 du 04 Décembre 2023 au 10 Décembre 2023**

---

Temps de travail de Mamisoa RANDRIANARIMA-NANA: 15 h 00 m

Temps de travail de Xianxiang ZHANG: 15 h 00 m

### **Travail effectué.**

- tâche 1 : Préparation de la soutenance
- tâche 2 : Soutenance
- tâche 3 : lecture article "Transfer Learning for Structures Spotting in Unlabeled Handwritten Documents using Randomly Generated Documents"
- tâche 4 : lecture article "SpottingNet : Learning the Similarity of Word Images with Convolutional Neural Network for Word Spotting in Handwritten Historical Documents"
- tâche 5 : Analyse et comparaison des différentes méthodes
- tâche 6 : Correction de l'état de l'art

### **Travail non effectué.**

### **Échanges avec le commanditaire.**

Plusieurs remarques ont été remontées lors de la soutenance notamment l'enrichissement de l'état de l'art ainsi que l'approfondissement sur les méthodes. D'autre méthodes et papiers ont également été suggéré

### **Planification pour la semaine prochaine.**

- tâche 1 : Correction et enrichissement état de l'art
- tâche 2 : lecture article "Text Line Segmentation in Historical Document Images Using an Adaptive U-Net Architecture"
- tâche 3 : Validation des choix avec le commandi-

taire

- tâche 3 :Début conception

---

### **Fiche de suivi de la semaine 7 du 11 Décembre 2023 au 17 Décembre 2023**

---

Temps de travail de Mamisoa RANDRIANARIMA-NANA: 15 h 00 m

Temps de travail de Xianxiang ZHANG: 13 h 30 m

### **Travail effectué.**

- tâche 1 :Validation des choix avec le commanditaire. Nous avons discuté avec Yejing des méthodes utilisées pour la suite du projet, la proposition pour le Data Augmentation a été validé, tandis qu'une approfondissement et d'autre recherches sont à faire pour la détection et segmentation
- tâche 3 : lecture article "Text Line Segmentation in Historical Document Images Using an Adaptive U-Net Architecture"
- tâche 4 : lecture article "SpottingNet : Learning the Similarity of Word Images with Convolutional Neural Network for Word Spotting in Handwritten Historical Documents"
- tâche 5 : Correction et enrichissement de l'état de l'art

## **Échanges avec le commanditaire.**

Discussion sur les différents propositions et objectifs qu'on va poursuivre pour la suite. Il faudra faire plus de recherche et approfondir sur les méthodes de detection et segmentation.

## **Planification pour la semaine prochaine.**

- tâche 1 : Conception détaillée
- tâche 2 : Validation des choix avec les commanditaires.
- tâche 3 : Installation de l'environnement de travail.
- tâche 4 : Début des développements

in Historical Document Images Using an Adaptive U-Net Architecture"

- tâche 4 : lecture article "SpottingNet : Learning the Similarity of Word Images with Convolutional Neural Network for Word Spotting in Handwritten Historical Documents"
- tâche 5 : Correction et enrichissement de l'état de l'art

## **Échanges avec le commanditaire.**

Discussion sur les différents propositions et objectifs qu'on va poursuivre pour la suite. Il faudra faire plus de recherche et approfondir sur les méthodes de detection et segmentation.

## **Planification pour la semaine prochaine.**

- tâche 1 : Conception détaillée
- tâche 2 : Rectification des propositions et validation des choix avec les commanditaires.
- tâche 3 : Installation de l'environnement de travail.
- tâche 4 : Début des développements

---

### **Fiche de suivi de la semaine 8 du 11 Décembre 2023 au 17 Décembre 2023**

---

Temps de travail de Mamisoa RANDRIANARIMA-NANA: 15 h 00 m

Temps de travail de Xianxiang ZHANG: 15 h 00 m

## **Travail effectué.**

- tâche 1 : Validation des choix avec le commanditaire. Nous avons discuté avec Yeqing des méthodes utilisées pour la suite du projet, la proposition pour le Data Augmentation a été validé, tandis qu'une approfondissement et d'autre recherches sont à faire pour la détection et segmentation
- tâche 3 : lecture article "Text Line Segmentation

---

### **Fiche de suivi de la semaine 9 du 18 Décembre 2023 au 24 Décembre 2023**

---

Temps de travail de Mamisoa RANDRIANARIMA-NANA: 5 h 00 m

Temps de travail de Xianxiang ZHANG: 12 h 30 m

### **Travail effectué.**

- tâche 1 : Correction et enrichissement de l' état de l'art
- tâche 2 : Conception détaillé de la méthode d'augmentation des données  
Validé par le commanditaire
- tâche 2 : Conception détaillé de la méthode de détection des expressions mathématiques  
Validé par le commanditaire
- tâche 4 : Mise en place de l'environnement de travail
- tâche 5 : Préparation de donnée - début annotation

### **Planification pour la semaine prochaine.**

- tâche 1 : Préparation des données
- tâche 2 : Développement des méthodes d'augmentation de donnée
- tâche 3 : Développement des méthodes detection des expressions mathématiques
- tâche 4 : Début des développements des méthodes d'évaluation avec L'IOU, ou Intersection over Union,

Temps de travail de Mamisoa RANDRIANARIMA-NANA: 15 h 00 m

Temps de travail de Xianxiang ZHANG: 12 h 30 m

### **Travail effectué.**

- tâche 1 : Préparation des données : annotation
- tâche 2 : Implementation et test des méthodes d'augmentation des données
- tâche 3 : Implementation et test des méthodes d'évaluation avec L'IOU, ou Intersection over Union sur des données de test.
- tâche 4 : Implementation et test de la détection par mot clé

### **Planification pour la semaine prochaine.**

- tâche 1 : Préparation des données : suite annotation et constitution de la base de donnée
- tâche 2 : Suite du développement des méthodes d'augmentation de donnée : correction et amélioration du méthode
- tâche 3 : Développement du méthode detection des expressions mathématiques : correction et amélioration du méthode

NANA: 13 h 00 m

Temps de travail de Xianxiang ZHANG: 13 h 30 m

#### **Travail effectué.**

- tâche 1 : Préparation des données : annotation
- tâche 2 : Implementation et test des méthodes d'augmentation des données
- tâche 3 : Implementation et test des méthodes d'évaluation avec L'IOU, ou Intersection over Union sur des données de test.
- tâche 4 : Implementation et test de la détection par mot clé

#### **Planification pour la semaine prochaine.**

- tâche 1 : Préparation des données : suite annotation et constitution de la base de donnée
- tâche 2 : Suite du développement des méthodes d'augmentation de donnée : correction et amélioration de la méthode
- tâche 3 : Développement de la méthode de détection des expressions mathématiques : correction et amélioration du méthode

---

#### **Fiche de suivi de la semaine 12 du 15 Janvier 2024 au 21 Janvier 2024**

---

Temps de travail de Mamisoa RANDRIANARIMA-NANA: 12 h 00 m

Temps de travail de Xianxiang ZHANG: 12 h 30 m

#### **Travail effectué.**

- tâche 1 : Préparation des données : suite annotation et constitution de la base de donnée -
- tâche 2 : Suite du développement des méthodes d'augmentation de donnée : correction et amélioration de la méthode
- tâche 3 : Développement du méthode detection des expressions mathématiques : correction et amélioration du méthode

#### **Planification pour la semaine prochaine.**

- tâche 1 : Préparation des données : constitution de données qui ne regroupent que les expressions mathématiques , de données qui ne regroupent que les textes, de données qui ne regroupent que les images de fond
- tâche 2 : Test, Évaluation et adaptation du méthode pour la détection des expressions mathématiques

---

#### **Fiche de suivi de la semaine 13 du 22 Janvier 2024 au 28 Janvier 2024**

---

Temps de travail de Mamisoa RANDRIANARIMA-NANA: 14 h 00 m

Temps de travail de Xianxiang ZHANG: 15 h 30 m

#### **Travail effectué.**

- tâche 1 : Préparation des données : suite annotation et constitution de la base de donnée -
- tâche 2 : Suite du développement des méthodes d'augmentation de donnée : correction et amélioration de la méthode
- tâche 3 : Adaptation des méthodes d'augmentation de donnée pour filtrer en fonction du type de données (expressions mathématiques, texte, etc.)
- tâche 4 : Correction de la méthode pour la détection des expressions mathématiques suivi de test et évaluation.

### **Planification pour la semaine prochaine.**

- tâche 1 : Enrichissement de la base de donnée en fonction des spécifications de la méthode de détection et de segmentation
- tâche 2 : Test, Évaluation et correction éventuelle de la méthode pour la détection des expressions mathématiques

---

### **Fiche de suivi de la semaine 14 du 29 Janvier 2024 au 04 Fevrier 2024**

---

Temps de travail de Mamisoa RANDRIANARIMA-NANA: 6 h 00 m

Temps de travail de Xianxiang ZHANG: 10 h 30 m

**Travail effectué.**

- tâche 1 : Enrichissement de la base de donnée en fonction des spécifications de la méthode de détection et de segmentation : pixel 1024\*1024 , uniformisation des fonds
- tâche 3 : Correction de la méthode pour la détection des expressions mathématiques suivi de test et évaluation.

### **Planification pour la semaine prochaine.**

- tâche 1 : Prépartition de donnée : suite annotation et augmentation du nombre de données
- tâche 2 : Test, Évaluation et correction éventuelle de la méthode pour la détection des expressions mathématiques

---

### **Fiche de suivi de la semaine 15 du 04 Fevrier 2024 au 11 Fevrier 2024**

---

Temps de travail de Mamisoa RANDRIANARIMA-NANA: 10 h 00 m

Temps de travail de Xianxiang ZHANG: 15 h 30 m

**Travail effectué.**

- tâche 1 : Préparation de donnée : suite annotation et augmentation du nombre de données d'entraînement et de validation - Enrichissement de la base de donnée en fonction des spécifications
- tâche 2 : Entrainement du modèle de réseau neu-

- ronal
- tâche 3 : Correction de la méthode pour la détection des expressions mathématiques suivi de test et évaluation.
  - tâche 4 Entrainement avec les nouvelles données du modèle, Test et Évaluation

#### **Planification pour la semaine prochaine.**

- tâche 1 : Préparation de donnée : suite annotation et augmentation du nombre de données d'entraînement et de validation
- tâche 2 : correction éventuelle de la méthode pour la détection des expressions mathématiques,suivi de test, Évaluation

---

#### **Fiche de suivi de la semaine 16 du 12 Fevrier 2024 au 18 Fevrier 2024**

---

Temps de travail de Mamisoa RANDRIANARIMA-NANA: 20 h 00 m

Temps de travail de Xianxiang ZHANG: 20 h 00 m

#### **Travail effectué.**

- tâche 1 : Expérimentation et analyse des résultats
- tâche 2 : Enrichissement des bases de données d'entraînement, de validation et de test
- tâche 3 : Améliorer les modèles de réseaux neuro-naux, suivi d'autre expérimentations et d'analyse

des résultats

#### **Planification pour la semaine prochaine.**

- tâche 1 : Expérimentation et analyse des résultats
- tâche 2 : Rédaction du rapport final
- tâche 2 : Préparation de la soutenance

---

#### **Fiche de suivi de la semaine 17 du 19 Fevrier 2024 au 25 Fevrier 2024**

---

Temps de travail de Mamisoa RANDRIANARIMA-NANA: 17 h 00 m

Temps de travail de Xianxiang ZHANG: 17 h 30 m

#### **Travail effectué.**

- tâche 1 : Tests et évaluations finaux
- tâche 2 : Rédaction du rapport final
- tâche 3 : Préparation de la soutenance

#### **Planification pour la semaine prochaine.**

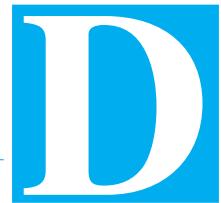
- tâche 1 : Soutenance
- tâche 2 : Préparation de la recette

Le tableau C.1 récapitule le taux d'avancement du projet. Rappelons que le temps de travail théorique *mini-*

Semaine	Temps prévu		Mamisoa RANDRIANARIMANANA				Xianxiang ZHANG		
	bas	haut	hebdo.	$\Sigma$	%	hebdo.	$\Sigma$	%	
	h : m	h : m	h : m	h : m		h : m	h : m		
1	10 : 00	12 : 30	6 : 30	6 : 30	65 (52)	6 : 30	6 : 30	65 (52)	
2	20 : 00	25 : 00	4 : 00	10 : 30	52 (42)	10 : 5	16 : 35	82 (66)	
3	30 : 00	37 : 30	14 : 30	25 : 00	83 (66)	11 : 20	27 : 55	93 (74)	
4	40 : 00	50 : 00	12 : 30	37 : 30	93 (75)	11 : 00	38 : 55	97 (77)	
5	50 : 00	62 : 30	18 : 00	55 : 30	111 (88)	18 : 30	57 : 25	114 (91)	
6	60 : 00	75 : 00	15 : 00	70 : 30	117 (94)	15 : 00	72 : 25	120 (96)	
7	70 : 00	87 : 30	15 : 00	85 : 30	122 (97)	13 : 30	85 : 55	122 (98)	
8	80 : 00	100 : 00	15 : 00	100 : 30	125 (100)	15 : 00	100 : 55	126 (100)	
9	90 : 00	112 : 30	5 : 00	105 : 30	117 (93)	12 : 30	113 : 25	126 (100)	
10	100 : 00	125 : 00	15 : 00	120 : 30	120 (96)	12 : 30	125 : 55	125 (100)	
11	110 : 00	137 : 30	13 : 00	133 : 30	121 (97)	13 : 30	139 : 25	126 (101)	
12	120 : 00	150 : 00	12 : 00	145 : 30	121 (97)	12 : 30	151 : 55	126 (101)	
13	130 : 00	162 : 30	14 : 00	159 : 30	122 (98)	15 : 30	167 : 25	128 (103)	
14	140 : 00	175 : 00	6 : 00	165 : 30	118 (94)	10 : 30	177 : 55	127 (101)	
15	150 : 00	187 : 30	10 : 00	175 : 30	117 (93)	15 : 30	193 : 25	128 (103)	
16	160 : 00	200 : 00	20 : 00	195 : 30	122 (97)	20 : 00	213 : 25	133 (106)	
17	170 : 00	212 : 30	17 : 00	212 : 30	125 (100)	17 : 30	230 : 55	135 (108)	

TABLE C.1 – Avancement du projet par rapport au temps de travail théorique minimal (respectivement haut)

*mal* correspond au temps indiqué sur la maquette pédagogique auquel on ajoute un strict minimum de 20 % correspondant au travail personnel hors emploi du temps. La partie « haute » de la fourchette correspond à 50 % de temps supplémentaire au titre du travail personnel.



---

## Auto-contrôle et auto-évaluation

Les figures D.1 , D.2, D.3 permettent d'énumérer un certain nombre de points importants dans les trois composantes du travail :

1. rapport représenté par la figure D.1 ;
2. présentation orale représenté par la figure D.3 ;
3. travail de fond représenté par la figure D.2;

ainsi que d'évaluer notre niveau de satisfaction à l'issue de la phase I, composée de trois étapes :

1. étude préalable ;
2. étude bibliographique ;
3. conception générale.

• • • La figure D.4, D.5, D.6 permettent d'énumérer un certain nombre de points importants dans les trois composantes du travail ainsi que d'évaluer notre niveau de satisfaction à l'issue de la phase II, constituée de :

1. la conception détaillée D.1 ;

PRD	Prénom	Nom	Notation								
Binome 1	Mamisona	RANDRIANARIMANANA	A. Maîtrise dans l'application du savoir-faire requis								
Binome 2	Xianxiang	ZHANG	B. Application du savoir-faire requis								
PROPOSITION DE NOTE (I)	#DIV/0!		C. Insuffisances, lacunes à corriger dans l'application du savoir-faire requis				D. Insuffisances flagrantes, inacceptables, voire travail absent				
PROPOSITION DE NOTE (II)	#DIV/0!										
NOTE PROJET	#REF!		A	B	C	D	Remarque / Note / Commentaire				
Phase I : Etude préalable, étude bibliographique et conception générale											
Rapport	Organisation	Plan	Equilibre	X	bien équilibré		1	1	1	1	0,75
			Coherence	X	on a suivi le plan fourni sur moodle		1	1	2	2	1,5
		Fluidité	Introductions (partielles)	X	introduction en chaque début de partie et sous-partie		1	1	2	2	1,5
			Transitions	X	ok		1	1	2	2	1,5
			Conclusions (partielles)	X	conclusion en fin de chaque partie		1	1	2	2	1,5
	Tableaux, figures	Numérotées		X	ok		1	1	1	1	0,75
		Légendées		X	ok		1	1	1	1	0,75
		Référencés (non "en ligne")		X	ok		1	1	1	1	0,75
Rédaction	Orthographe	Couilles		X	verification		1	1	1	1	0,75
		Fautes évitables		X	verification		1	1	2	2	1,5
		Franglais, jargon		X	il existe des mots qu'on n'a pas su traduire		1	1	3	2,25	1,5
	Rédaction	Aléas		X	ok, reformulation et rédaction simple et claire		1	1	4	3	2
		Absence de plagiat !		X			1	1	1	1	0,75
Bibliographie	Références	Suffisantes (nombre, intérêt)		X	ok		1	1	3	3	2,25
		Pénèttes		X			1	1	1	1	0,75
		Complètes (auteurs, pages..)		X	Auteur et source ok		1	1	1	1	0,75
		Conséquentes (volume)		X			1	1	2	1,5	1
		Références dans le texte		X			1	1	2	1,5	1
											32
		Proposition de note haute			16,88						
		Proposition de note basse			11,88						
		Proposition de note du jury									

FIGURE D.1 – Points à contrôler sur le rapport à l'issue de la phase I

Projection	Organisation	Plan	X	Discussion en amont fait avec les encadrants	1	1	1	2	2	1,5	
		Liaisons	X	ok	1	1	1	2	2	1,5	
		Numérotation	X	ok	1	1	1	1	1	0,75	
	Contenu	Informatif	X	discussion en amont fait avec les encadrants	1	1	1	2	2	2,25	
		Concise	X	ok	1	1	1	1	1	0,75	
		Clair	X	ok	1	1	1	2	2	1,5	
		Orthographe	X	ok	1	1	1	1	1	0,75	
		Illustrations	X	ok	1	1	1	2	2	1,5	
Oral	Présentation	Assurance	X	à améliorer	1	1	1	2	2	1,5	
		Tenue	X	à améliorer			1	1	1	0,5	
		Articulation, compréhension	X	à améliorer			1	1	1	0,75	
	Durée	Respect	X	ok	1	1	1	2	2	1,5	
		Temps de parole équilibré	X	ok	1	1	1	2	2	1,5	
	Réponses	Pertinence	X	ok	1	1	1	2	2	1,5	
		Argumentation	X	ok	1	1	1	3	2,25	1,5	
											27
		Proposition de note haute		16,85							
		Proposition de note basse		12,04							
		Proposition de note du jury									

FIGURE D.2 – Points à contrôler sur le travail à l'issue de la phase I

Travail	Etude	Bibliographie	Adequate suffisante	X				1	1	1	3	2,25	1,5	
		Cahier des charges	Clair	!!! X		D'autres méthodes à ajouter		1	1	1	2	1,5	1	
			Formalisé	!!! X		Valider par les encadrants		1	1	1	2	2	1,5	
		Hypothèses envisagées	Nombre	X		ok		1	1	1	1	0,75	0,5	
		Pertinence	!!! X	X		Ok, amélioration au fur et à mesure. Résultat final bien		1	1	1	2	1,5	1	
		Analyse a priori		X				1	1	1	2	1,5	1	
		Validation	Tableau comparatif	!!! X		ok		1	1	1	2	2	1,5	
			Choix argumenté(s)	!!! X		ok		1	1	1	2	2	1,5	
		Faisabilité	!!! X	X		ok		1	1	1	2	2	1,5	
		Complexité	Temps consacré	X				1	1	1	1	1	0,75	
Annexes		Résultats obtenus		X				1	1	1	2	1,5	1	
		Difficulté	Intrinsèque	!!! X				1	1	1	3	2,25	1,5	
			Vis-à-vis du binôme	X				1	1	1	1	0,75	0,5	
		Fiches d'avancement	Régularité	!!! X		ok		1	1	1	1	1	0,75	
			Détailées	!!! X		Réunion hebdomadaire avec les encadrants.		1	1	1	2	1,5	1	
	Gantt		Gantt	!!! X		ok		1	1	1	2	1,5	1	
		Prévisionnel et justifications		!!! X				1	1	1	2	1,5	1	
		Effectifs, erreurs, corrections		!!! X		ok		1	1	1	2	1,5	1	
		Proposition note haute			16,67							33		
		Proposition note basse			11,67									
		Proposition de note du jury												

FIGURE D.3 – Points à contrôler sur la projection à l'issue de la phase I

FIGURE D.4 – Points à contrôler à l’issue de la phase II

FIGURE D.5 – Points à contrôler à l’issue de la phase II

FIGURE D.6 – Points à contrôler à l’issue de la phase II