Étude préliminaire de reconnaissance d'écriture sur des documents historiques

Adeline Granet, Emmanuel Morin, Harold Mouchère, Solen Quiniou, Christian Viard-Gaudin

Étude préliminaire de reconnaissance d'écriture sur des documents historiques

Adeline Granet* — **Emmanuel Morin*** — **Harold Mouchère*** — **Solen Quiniou*** — **Christian Viard-Gaudin***

* *1. UMR 6004, LS2N, Université de Nantes, 44000, Nantes, France*

RÉSUMÉ. *Ce travail s'intéresse à l'extraction d'informations dans les registres comptables de la Comédie-Italienne du XVIII^{e} siècle. Ces derniers renferment des informations précieuses pour des chercheurs en sciences humaines et sociales qui travaillent sur l'acculturation des acteurs italiens de cette époque. L'extraction d'informations, dans des documents anciens non encore étudiés, est un processus long et complexe qui demande une expertise à chaque étape : détection et segmentation en blocs, lignes ou mots, extraction de caractéristiques, reconnaissance d'écriture manuscrite. Les réseaux de neurones récurrents, de type BLSTM, avec un décodage CTC constituent une des méthodes les plus prometteuses en reconnaissance d'écriture, pour réaliser l'étiquetage d'une séquence donnée en entrée et produire un résultat de reconnaissance. Cet article présente une étude préliminaire de l'utilisation de ce type de réseau de neurones pour une première tâche : la reconnaissance des titres des pièces de théâtre, dans des documents historiques multilingues (français et italien) utilisant un vocabulaire fermé et essentiellement composé d'entités nommées.*

ABSTRACT. *This work cares about information retrieval in accounting registers of Italian comedy of the 18th century. These documents contain precious information for human and social science researchers interested in the integration of the Italian actors during this century. Information retrieval in old documents which have never been studied before, is a long and difficult process. Each step asks an expertise : detection and segmentation into blocs, lines or words; extraction efficient features; and handwriting recognition. The BLSTM recurrent neural network with CTC decoding is the most popular solution which outperforms others for alignment between a transcription and an input sequence. This paper explains a preliminary investigation using this kind of recurrent neural network for the following task : identify the play's titles in multilingual historical documents using closed vocabulary that mainly contains named entities.*

MOTS-CLÉS : *Reconnaissance d'écriture manuscrite$_1$, BLSTM-CTC$_2$, documents anciens$_3$*

KEYWORDS: *Handwriting recognition$_1$, BLSTM-CTC$_2$, old documents$_3$.*

1. Introduction

Le projet ANR CIRESF[1] s'est donné comme objectif de parfaire notre connaissance du théâtre italien du XVIII^{e} siècle. Les chercheurs en sciences humaines et sociales souhaitent révéler l'acculturation mise en œuvre par les acteurs de la Comédie-Italienne, pour finalement les a mener à fusionner avec l'Opéra en devenant l'Opéra-Comique en 1762. Ce projet vise à proposer des outils informatiques pour permettre aux spécialistes du théâtre Italien d'accéder efficacement aux informations. La tâche d'exploration dans les documents de cette période, pour y trouver les informations pertinentes, est complexifiée par la quantité de ressources disponibles. La ressource la plus importante, pour la Comédie-Italienne, est un ensemble de registres comptables qui répertorient les comptes journaliers, mensuels et annuels à travers plus de 28 000 pages, de 1716 à 1783. Ces registres ont été numérisés pour faciliter leur consultation tout en les préservant.

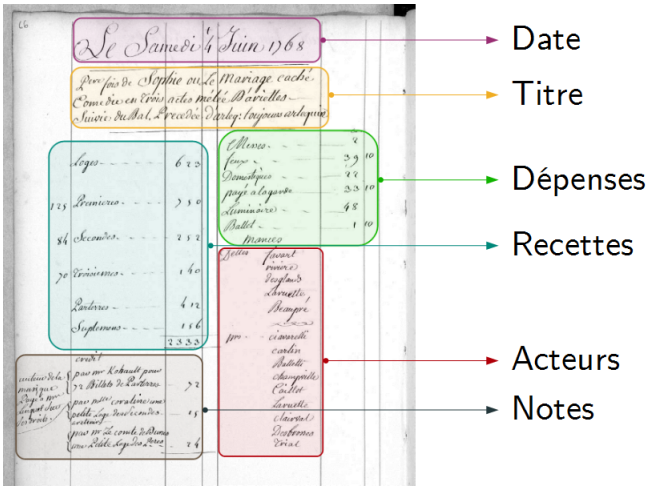


Figure 1 – Exemple de compte journalier de la Comédie-Italienne et ses informations

Cette étude s'intéresse tout particulièrement aux informations contenues dans les comptes journaliers. Sur la figure 1, une page, illustrant les comptes journaliers, est présentée. Dans la partie supérieure de ce type de page, il est possible de lire la date suivie des titres des pièces jouées. Dans la colonne de gauche, les recettes sont détaillées ; dans la colonne de droite, ce sont les dépenses qui sont présentées, suivies du nom des acteurs du jour. Au fil des registres et des saisons, une évolution de la langue est notable, jusqu'en 1730 : les registres étaient écrits jusque là en italien et laissent ensuite place au français. La mise en page varie également au fil des saisons.

Dans le cadre d'une analyse automatique avec reconnaissance d'écriture, plusieurs difficultés peuvent être identifiées : l'évolution de la langue de rédaction des registres

[1] Contrainte et Intégration : pour une Réévaluation des Spectacles Forains et Italiens sous l'Ancien Régime

ainsi que le siècle étudié. En effet, l'écriture longiligne et italique est très caractéristique du XVIII^{e} siècle. Nous pouvons mentionner une singularité de l'époque qui est la forme longue du "s". Quelques exemples illustrent cela en figure 2. La synergie entre la reconnaissance d'écriture et le traitement du langage à travers l'extraction d'information sera déterminante pour faire face au multilinguisme des données et à leurs singularités.

Figure 2 – Exemple de la forme longue de "s" dans des documents anciens.

Notre étude vise à mettre en place un système de reconnaissance d'écriture à partir de réseaux BLSTM-CTC. Ces derniers permettent de s'affranchir des contraintes de segmentation en caractères ou en mots, pour pouvoir étiqueter une séquence. L'apprentissage est effectué sur des corpus comparables, de la même période, mais de langues différentes. Rares sont les études qui ont entrepris de faire de la reconnaissance d'écriture multilingue ou multi-époque avec un unique système (Frinken et al., 2010). L'estimation de la séquence d'étiquettes obtenue en sortie du système BLSTM-CTC pour la zone de titre sera évaluée et corrigée grâce à un alignement sur une liste de titres de pièces candidats.

Cet article présente une étude préliminaire autour de la reconnaissance d'écriture manuscrite ancienne, à l'aide de réseaux de neurones. Nous commencerons par une étude bibliographique centrée sur les différentes approches utilisant les réseaux de neurones. Dans la section 3, nous décrirons les expériences mises en oeuvres, suivie par les résultats (section 4). Avant de conclure, nous présenterons une évaluation réalisée sur une transcription manuelle des titres de la Comédie-Italienne en section 5.

2. État de l'art

2.1. *Extraction des caractéristiques*

Nous avons identifié trois approches différentes. La première se base principalement sur la valeur, la position et l'agencement des pixels dans une fenêtre d'observation donnée. La largeur de cette dernière peut varier d'une colonne de pixels jusqu'à une image entière. La deuxième approche s'appuie sur le calcul des caractéristiques directionnelles. Les deux méthodes les plus utilisées sont SIFT (Lowe, 1999) et HOG (Terasawa et Tanaka, 2009). Elles construisent un histogramme de l'orientation des gradients sur un bloc d'image. La $3me$ approche, plus récente, utilise les auto-encoders à variation ou à convolution (Toledo et al., 2016 ; Chen et al., 2015 ; Masci et al., 2011) pour réaliser une extraction non-supervisée des caractéristiques.

2.2. *Techniques de reconnaissance de l'écriture*

Parmi les méthodologies de reconnaissance de l'écriture manuscrite, nous pouvons faire ressortir en première approximation trois types de systèmes. À la fin des

années 90, les HMMs (Hidden Markov Models) se sont imposés en surclassant par leur capacité d'apprentissage sur des séquences les approches structurelles qui prévalaient alors (Bunke et al., 1995 ; Park et Lee, 1996 ; Fine et al., 1998). Ensuite, le pouvoir discriminant des réseaux de neurones a permis, à travers des systèmes hybrides neuro-markovien, de mieux modéliser le caractère local et global de l'écriture (Koerich et al., 2002 ; Fischer et al., 2012). Depuis quelques années, les architectures de type BLSTM, définies ci-dessous, intègrent encore mieux cette capacité à mixer du local et du global, avec des effets de contexte, pour optimiser une décision sur une séquence complète (Grosicki et El Abed, 2009 ; Fischer et al., 2009).

Dernièrements, les réseaux de neurones récurrents (RNN) ont révolutionné le domaine. Ces méthodes discriminatives présentent un grand nombre d'avantages : ils sont robustes au bruit et ils n'ont pas besoin de connaissances *a priori*. Pour augmenter encore les performances déjà satisfaisantes, les RNN sont devenus des systèmes bidirectionnels (BRNN) afin de prendre en compte le contexte passé et futur de chaque position dans la séquence d'entrée. Différentes études, comme celle de (Hochreiter et al., 2001), ont montré une limite dans ces méthodes : il s'agit de la perte du gradient durant la rétropropagation sur les séquences très longues, ce qui favorisent le contexte à court terme. Pour pallier ce problème une cellule appelée *Long Short Term Memory* a été proposée (Hochreiter et Schmidhuber, 1997). L'utilisation de cette cellule, dans un réseau récurrent unidirectionnel, converge là où d'autres systèmes ne le pouvaient pas. Finalement, le réseau bidirectionel *Bidirectional Long Short Term Memory* (BLSTM), qui est une combinaison d'un système BRNN avec des LSTM, est devenu la méthode de référence ces dernières années (Graves, 2012 ; Fischer et al., 2009), car il montre de meilleures performances. En effet, la combinaison du passé et du futur rend le système plus stable.

En reconnaissance d'écriture, les réseaux présentés ici, bien qu'ils soient tous très performants, ne permettent pas de réaliser un alignement direct entre une séquence d'entrée et les étiquettes correspondantes, de façon automatique. En effet, ils demandaient, jusque-là, une pré-segmentation en caractères afin d'identifier leur localisation dans la séquence et un post-traitement afin d'étiqueter. Dans le but de résoudre ce problème d'étiquetage de séquences, sans avoir connaissance de l'alignement entre l'entrée et la sortie du système, Graves et al. (2012) ont proposé un système appelé CTC (*Connectionist Temporal Classification*). Ce système a été, dans un premier temps, utilisé en reconnaissance de la parole, puis appliqué en reconnaissance d'écriture (Graves, 2012).

3. Description du système de reconnaissance

Avant de pouvoir réaliser un alignement entre les titres contenus dans les registres manuscrits et une liste non-exhaustive de titres de pièces jouées à cette époque[2], un réseau BLSTM-CTC doit être mis en place pour fournir la séquence la plus probable

[2] "Le répertoire de la Comédie-Italienne de Paris (1716-1762)" réalisé par E. de Luca, référence 934 premières de pièces en indiquant la langue utilisée, la date et un résumé de la pièce.

pour une ligne donnée. La segmentation des titres en lignes, est effectuée par l'outil DMOS proposé par (Coüasnon et Camillerapp, 2002) et qui est dédié à la tâche de détection et segmentation automatique de documents anciens[3].

3.1. *Prétraitements*

La partie extraction des caractéristiques a une grande influence sur le succès (ou non) de la tâche. Nous souhaitons définir expérimentalement la méthode ainsi que la configuration la plus adaptée à nos données historiques, sans pour autant multiplier le nombre de caractéristiques. Pour les extraire, nous avons testé deux types de méthode, chacune utilisant un système de fenêtre glissante avec ou sans chevauchement.

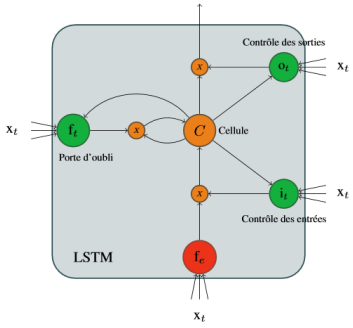
La première méthode extrait les profils haut et bas, le nombre de transitions entre l'encre et le fond, ainsi que la somme des pixels pour chaque colonne de pixels. Cela donne uniquement 4 caractéristiques locales. La seconde méthode, HOG, a été testé en faisant varier la largeur de la fenêtre d'observation de 1 à 20 pixels sans chevauchement. Pour évaluer l'impact de la taille de la cellule sur les informations extraites, nous avons défini des formats de cellules dont : la largeur de la cellule est égale à celle de la fenêtre et varie de 4 à 15 ; la hauteur de la cellule varie entre la hauteur de l'image (120 pixels) et 15 pixels. La dernière cellule expérimentée est de largeur 10 pixels sur 30 avec une fenêtre glissante de 20 pixels de large, et un pas de 2 pixels entre deux fenêtres consécutives. Le nombre de caractéristiques extrait ne dépasse pas le nombre de neurones contenus dans chaque couche cachée.

3.2. *Système BLSTM-CTC*

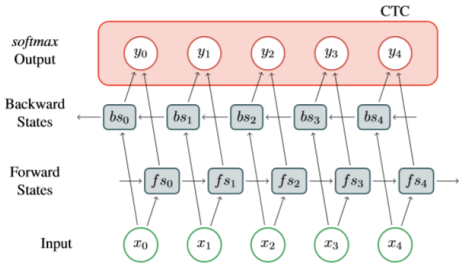
Le réseau BLSTM est constitué de deux couches cachées *forward* et *backward*, composées de blocs mémoires LSTM remplaçant les neurones récurrents de départ. Ces deux couches sont indépendantes sur la phase d'apprentissage, l'une donnant accès au passé et l'autre, au futur. Les blocs LSTM (voir figure 3a) gèrent l'influence des informations qui sont diffusées dans le réseau. Puis la somme pondérée des activations des deux couches cachées est fournie en entrée de la couche de sortie pour chaque instant t. L'algorithme CTC est placé en sortie du réseau BLSTM (voir figure 3b). Il permet d'apprendre l'étiquetage dans le cadre d'un alignement non-connu. Le nombre de neurones dépend du nombre de caractères de l'alphabet auquel un label *blank* est ajouté. Il permet au système de ne pas prendre de décision à chaque instant.

La phase d'apprentissage du BLSTM-CTC est réalisée par une descente de gradient à travers le réseau. Cela préserve l'intégrité du système en évitant de passer en phase de sur-apprentissage. Le CTC nécessite une phase d'apprentissage interne en adaptant l'algorithme *forward-backward* afin d'intégrer le label *blank* dans l'ensemble des étiquettes (au début, à la fin et entre chaque caractère). Il prend ainsi en considération l'ensemble des séquences possibles pour créer chaque étiquette.

[3] Cela ne sera pas développé dans cet article.



(a) Cellule mémoire BLSTM



(b) Réseau de neurones bidirectionnel à deux couches cachées avec cellule LSTM et CTC

Figure 3 – (a) montre la construction d'une cellule mémoire LSTM et (b) montre un réseau BLSTM.

Pour pouvoir décoder les sorties du CTC, l'algorithme le plus simple est *le meilleur chemin*. À chaque instant *t*, le neurone le plus actif est conservé. À la séquence obtenue, il faut supprimer les caractères *blank* et les lettres doublées (et non séparées par un autre caractère). L'utilisation d'un dictionnaire peut être utilisé afin de contraindre le décodage des labels.

4. Expérimentations

4.1. *Ressources utilisées*

Les ressources annotées du XVIII^{e} sont assez rares, et celles en français inexistantes, à notre connaissance. C'est pour cela que nous utilisons une base de données en anglais, Georges Washington (Fischer et al., 2012) pour réaliser la première étape d'apprentissage. La base de données Georges Washington (GW) est constituée de 20 pages de correspondances, avec deux scripteurs connus mais dont les écritures sont similaires. Un exemple d'une page est fourni en figure 4. La base fournie est constituée de 656 lignes de textes et de 4 894 instances de mots annotées. Chaque image a été normalisée avec une hauteur de 120 pixels, après avoir corrigé l'angle d'inclinaison des caractères. Nous avons utilisé des répartitions de validation croisée, en prenant 19 pages pour l'apprentissage et la validation, et 1 page pour les tests.

4.2. *Configuration du système*

Le nombre de neurones dans la couche d'entrée du réseau BLSTM dépend du nombre de caractéristiques extrait à chaque instant. Cela varie ainsi de 4 à 64 neurones suivant la méthode utilisée. Les couches cachées sont pourvues de 100 cellules LSTM comme cela a été défini à l'origine dans (Graves, 2012). La couche de sortie CTC est constituée de 75 neurones type *softmax*. Cela correspond aux caractères de l'alphabet minuscule et majuscule, les digits et différents signes de ponctuation et symboles particuliers comme la signature de Georges Washington. La phase d'appren-

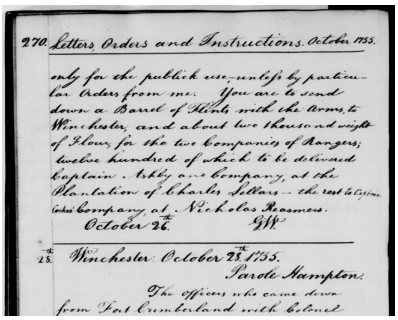


Figure 4 – Exemple de page extrait de la base Georges Washington.

tissage est interrompue dans le cas où le nombre d'itérations maximum est atteint ou si le coût de la fonction de perte n'a pas diminué sur l'ensemble de validation durant 5 itérations. Nous évitons ainsi le sur-apprentissage.

4.3. *Résultats sur la reconnaissance des lignes de GW*

Les tests que nous avons réalisés ont été effectués uniquement sur GW car les images de lignes de la Comédie-Italienne (CI) ne sont pas encore disponibles. Notre but étant de réaliser un transfert de connaissances de GW vers CI, la première étape consiste à mettre en place un système efficace sur GW. La mesure de référence utilisée est la précision sur les caractères correctement reconnus. Nous prenons ainsi en compte le nombre de caractères insérés, supprimés ou substitués entre la référence et la transcription, normalisé par la taille de la plus grande des deux séquences comparées. Nous incluons également la mesure *Word Accuracy* qui calcule la précision de mots correctement reconnus. Nous proposons, à titre indicatif, une comparaison avec des systèmes BLSTM-CTC dans la partie haute du tableau 1. Ces études utilisent des ressources beaucoup plus fournies, que sont IAM[4] et RIMES[5] : la base IAM possède 20 fois plus de lignes que GW, et RIMES, le double de mots. Cela leur permet de proposer des évaluations directement sur les lignes, tandis que nous devons opérer une étape préliminaire sur les mots. Dans la partie centrale du tableau 1, les résultats des expérimentations sur la reconnaissance des mots de GW sont présentés. Finalement, la dernière partie du tableau 1 correspond aux résultats obtenus sur la reconnaissance des lignes de GW à partir des lignes elles-mêmes ou par progression de l'apprentissage des mots vers lignes.

L'ensemble des systèmes que nous avons testés, fournit une bonne précision sur la reconnaissance de caractères, et même meilleure que celle de (Graves, 2012). Cela

[4]IAM est constitué de 1 539 lettres en anglais avec 650 scripteurs différents. `http://www.fki.inf.unibe.ch/databases/iam-handwriting-database`

[5]RIMES est constituée de 12 000 pages annotées en français. `http://www.rimes-database.fr`

peut s'expliquer par le peu de variabilité dans la forme des caractères, dans notre corpus, où les deux scripteurs ont un style tellement proche qu'il est d'usage de n'en considérer qu'un seul, contre 650 pour IAM. La capacité du système à reconnaître des mots semble plus limité mais similaire à (Morillot et al., 2013). Nos systèmes font au moins une erreur de caractère sur la moitié des mots du corpus de test. La comparaison avec les deux autres méthodes montre l'influence que peut avoir la quantité de données. La couverture du vocabulaire dans le corpus de notre apprentissage peut être une des raisons de ces erreurs car des mots du corpus de test peuvent ne pas avoir été vus au préalable.

Tableau 1 – Résultats obtenus et comparaison avec l'état de l'art en reconnaissance d'écriture avec des réseaux utilisant le BLSTM et CTC.

Système	Ressource	Methode	# Caract.	Char. Acc	Word acc.
(Graves, 2012)	IAM	Locale	9	81,8 %	74,1 %
(Morillot et al., 2013)	RIMES	Locale	56	-	43,2 %
BLSTM-CTC	GW (mots)	Locale	4	69.62 %	38,65 %
BLSTM-CTC		HOG	64	71,15 %	37,83%
BLSTM-CTC	GW (lignes)	HOG	64	10,07%	-
BLSTM-CTC	GW (mots,lignes)	HOG	64	77,31%	39,00%

La quantité d'images de lignes étant très limitée sur GW, deux expériences ont été menées afin d'évaluer la capacité du système à apprendre des lignes (voir la partie inférieure du tableau 1). Dans un premier temps, durant 100 itérations, le système a tenté d'apprendre les lignes directement. Puis, les poids du réseau, entraînés sur les mots de GW, ont été utilisés pour poursuivre l'apprentissage sur les lignes. Cette solution augmente de 67 % la précision sur la reconnaissance de caractères. Dans la première phase, le système s'est attaché à reconnaître les caractères puis il a cherché à identifier le caractère espace. Finalement, cette expérience nous montre qu'en reconnaissance de ligne, il est préférable d'effectuer l'apprentissage des caractères progressivement.

Dans le contexte de reconnaissance d'écriture sur GW, l'étude de (Lavrenko et al., 2004) avec un système de HMM, atteint une précision de 46.9% de mots reconnus en excluant les mots hors vocabulaire (de l'apprentissage) et avec 19 pages pour réaliser l'apprentissage. Nous obtenons une précision de mots correctement reconnus de 38,65% en nous plaçant dans les mêmes conditions. Une fois de plus, nous pouvons attribuer ce faible résultat au problème de quantité de données pour l'apprentissage du réseau de neurones, qui est plus gourmand en terme de ressources qu'un système HMM.

5. Évaluation manuelle de l'alignement des titres de la Comédie-Italienne

Dans cette section, nous souhaitons définir la limite supérieure de notre étude. Nous partons du postula que le système BLSTM-CTC ne pourra pas nous fournir une transcription avec 100 % de précision sur les caractères. Nous avons ainsi transcrit 155 images de zone de titre manuellement représentant cet idéal. Un exemple de transcription est fourni en figure 5. À cette transcription, nous avons associé le titre original de

la pièce tel qu'il *peut* être référencé dans une liste non-exhaustive de 1 200 titres de pièces de la Comédie-Italienne que nous avons constitué.

La distance entre les deux permet de montrer la difficulté d'aligner les titres écrits dans les registres avec les titres références. La distance de Levenshtein, qui mesure les insertions, les suppressions et les substitutions, a été appliquée entre la transcription et son titre. La précision au niveau caractère est de 76,72 %. Un quart des caractères n'a pas pu être reconnu car il y a des abréviations fortes comme "&c." qui correspond à une substitution avec plus de 4 mots. Puis, chaque titre transcrit a été comparé avec l'ensemble des titres référencés afin de trouver le plus proche. Seulement 65,4 % des transcriptions manuelles sont associées en première position avec le bon titre de référence. Une seconde méthode d'alignement a été testée sur nos données. Cette méthode, PHOC (Almazán et al., 2014), est une représentation pyramidale des occurrences des caractères contenus dans chaque titre. Pour un titre transcrit manuellement, une recherche par la méthode des plus proches voisins est utilisée. Nous obtenons une précision de 60,43 % de titres correctement reconnus en première position.

Il apparaît clairement que l'alignement entre les titres écrits et ceux référencés dans notre liste de référence ne pourra pas être résolu avec les méthodes standards uniquement. De plus, il faudra pouvoir détecter les différents titres présents sur une ligne utilisant des conjonction de coordinations comme séparateur (mais pouvant faire partie intégrante d'un titre) ou ceux coupés sur plusieurs lignes.

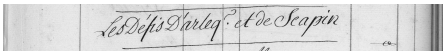


Figure 5 – Exemple d'une image de titre avec la transcription "Les Défis D'arleq et de Scapin" et pour le titre de référence "Défis d'Arlequin et de Scapin".

6. Conclusion

L'alignement entre une transcription "idéale" et les références de titres, nous donne une précision au niveau caractère de 76,72 %. Cela nous oblige à améliorer la méthode d'alignement car il faut qu'elle pallie les erreurs induites par le BLSTM-CTC. Cependant, les tests que nous avons réalisés pour amorcer la transcription des titres des pièces de théâtre de la Comédie-Italienne, prouvent que nous avons réussi à atteindre nos objectifs par rapport à l'état de l'art. En effet, nous avons égalé des systèmes fonctionnant avec deux fois plus de données d'apprentissage, l'objectif étant de produire une estimation de transcription la plus fiable possible. La prochaine étape consiste à ajouter des bases de données à notre apprentissage, pour couvrir un large spectre de variations d'écritures et de langues, et sur différentes époques.

Remerciements

Remerciement pour les implémentations de CTC en theano par Mohammad Peyeshki sur Github.

7. Bibliographie

Almazán J., Gordo A., Fornés A., Valveny E., « Word spotting and recognition with embedded attributes », IEEE trans. on PAMI, vol. 36, nº 12, p. 2552-2566, 2014.

Bunke H., Roth M., Schukat-Talamazzini E. G., « Off-line cursive handwriting recognition using hidden Markov models », Pattern Recognition, 1995.

Chen K., Seuret M., Liwicki M., Hennebert J., Ingold R., « Page segmentation of historical document images with convolutional autoencoders », ICDAR, 2015.

Coüasnon B., Camillerapp J., « DMOS, une méthode générique de reconnaissance de documents : évaluation sur 60 000 formulaires du XIXe siecle », CIFED, p. 225-234, 2002.

Fine S., Singer Y., Tishby N., « The hierarchical hidden Markov model : Analysis and applications », Machine Learning, 1998.

Fischer A., Keller A., Frinken V., Bunke H., « Lexicon-free handwritten word spotting using character HMMs », Pattern Recognition Letters, 2012.

Fischer A., Wüthrich M., Liwicki M., Frinken V., Bunke H., Viehhauser G., Stolz M., « Automatic transcription of handwritten medieval documents », VSMM, IEEE, 2009.

Frinken V., Fischer A., Bunke H., Manmatha R., « Adapting blstm neural network based keyword spotting trained on modern data to historical documents », ICFHR, IEEE, 2010.

Graves A., « Supervised sequence labelling », Supervised Sequence Labelling with Recurrent Neural Networks, Springer, 2012.

Grosicki E., El Abed H., « ICDAR 2009 handwriting recognition competition », ICDAR, IEEE, p. 1398-1402, 2009.

Hochreiter S., Bengio Y., Frasconi P., Schmidhuber J., Gradient flow in recurrent nets : the difficulty of learning long-term dependencies, A field guide to dynamical recurrent neural networks. IEEE Press, 2001.

Hochreiter S., Schmidhuber J., « Long short-term memory », Neural Computation, 1997.

Koerich A. L., Leydier Y., Sabourin R., Suen C. Y., « A hybrid large vocabulary handwritten word recognition system using neural networks with hidden Markov models », ICFHR, IEEE, p. 99-104, 2002.

Lavrenko V., Rath T. M., Manmatha R., « Holistic word recognition for handwritten historical documents », DIAL, p. 278-287, 2004.

Lowe D. G., « Object recognition from local scale-invariant features », CV, Ieee, 1999.

Masci J., Meier U., Ciresan D., Schmidhuber J., « Stacked convolutional auto-encoders for hierarchical feature extraction », ANN, Springer, 2011.

Morillot O., Likforman-Sulem L., Grosicki E., « New baseline correction algorithm for text-line recognition with bidirectional recurrent neural networks », Journal of Electronic Imaging, vol. 22, nº 2, p. 023028-023028, 2013.

Park H.-S., Lee S.-W., « Off-line recognition of large-set handwritten characters with multiple hidden Markov models », Pattern Recognition, 1996.

Terasawa K., Tanaka Y., « Slit style HOG feature for document image word spotting », ICDAR, IEEE, 2009.

Toledo J. I., Cucurull J. et al., « Election Tally Sheets Processing System », DAS, IEEE, 2016.