

THÈSE DE DOCTORAT DE

NANTES UNIVERSITÉ

ÉCOLE DOCTORALE N° 641

*Mathématiques et Sciences et Technologies du numérique
de l'Information et de la Communication*

Spécialité : *Informatique*

Par

Elmokhtar MOHAMED MOUSSA

Conversion d'écriture hors-ligne en écriture en-ligne et réseaux de neurones profonds

En vue de la soutenance de Thèse à Nantes, le 16 Janvier 2024

Unité de recherche : Laboratoire des Sciences du Numérique de Nantes

Rapporteurs avant soutenance :

Prof. Andreas FISCHER Professeur des universités à la Haute École spécialisée de Suisse occidentale (HEIA Fribourg)
Dr. Laurence LIKFORMAN Maître de conférences HDR à Telecom Paris

Composition du Jury :

Attention, en cas d'absence d'un des membres du Jury le jour de la soutenance, la composition du jury doit être revue pour s'assurer qu'elle est conforme et devra être répercutée sur la couverture de thèse

Président :	Prénom NOM	Fonction et établissement d'exercice (<i>à préciser après la soutenance</i>)
Examineurs :	Prof. Eric ANQUETIL	Professeur des universités à INSA Rennes
	Dr. Clément CHATELAIN	Maître de conférences HDR à INSA Rouen Normandie
Dir. de thèse :	Pr. Harold MOUCHERE	Professeur des universités à Nantes Université
Co-dir. de thèse :	Dr. Thibault LELORE	Ingénieur R&D à MyScript SAS

Titre : Conversion d'écriture hors-ligne en écriture en-ligne et réseaux de neurones profonds

Mot clés : écriture manuscrite, CNN, Transformer

Résumé : Cette thèse se focalise sur la conversion automatisée d'images statiques d'écriture hors-ligne en signaux temporels d'écriture en-ligne. Cette tâche est complexe en raison de l'ambiguïté dans les zones de l'écriture où la direction ou l'intention du mouvement du stylo ne sont pas claires, ainsi que de la variabilité individuelle dans l'ordre de tracé des traits, ce qui ajoute une couche de complexité supplémentaire. Les systèmes classiques proposés dans l'état de l'art sont conçus manuellement avec un ensemble de règles et d'heuristiques strictes qui sont difficiles à maintenir et ne se généralisent pas à suffisamment de langues et types de contenus. Récemment, de nombreuses approches basées sur des apprentissages automatiques de réseaux de neurones à partir de couples de données hors-ligne et en-ligne ont été proposées dans la littérature afin de résoudre ce problème. Cette thèse constitue un effort pour étendre ses réseaux de neurones au-delà de l'échelle de lettres isolés et de les étendre à d'autres types de contenus plus complexes.

La thèse explore deux approches distinctes basées sur des réseaux de neurones pour reconstruire la trajectoire du stylo à partir d'images d'écriture hors-ligne, avec l'objectif de généraliser au-delà des symboles isolés vers des contenus manuscrits plus vastes, y compris des mots, des lignes de texte et des équations mathématiques. La première approche est un réseau de neurones convolutif

UNet entièrement convolutif multitâche basé sur la méthode de [26]. Cette approche a démontré des bons résultats de squelettisation mais en revanche une extraction de trait problématique : les expériences réalisées sur les expressions mathématiques ont montré une tendance du modèle à sursegmenter les traits, en raison d'une part du contexte spatiale et temporel insuffisant et d'autres parts les limitations de modélisation temporelle intrinsèque à l'architecture CNN. La deuxième approche s'appuie sur les résultats de squelettisation précédent pour extraire les sous-traits du squelette et propose une modélisation au niveau sous-traits avec des Transformers constitué d'un encodeur de sous-trait pour extraire un vecteur caractéristique de sous-trait (SET) et d'un décodeur (SORT) pour ordonner les différents sous-traits à l'aide de leur embedding ainsi que la prédiction de lever de stylo. Cette approche s'est montrée meilleures que l'état de l'art sur les bases de données de lignes de textes et d'équations mathématiques et a permis de surmonter plusieurs limitations relevées dans la littérature.

Ces avancées ont permis d'étendre la portée de la conversion hors-ligne vers en-ligne pour inclure des lignes entières de texte et d'aborder un contenu bidimensionnel tel que les équations mathématiques. Ouvrant ainsi la voie à une généralisation à l'échelle d'une page entière d'écriture.

Title: Offline handwriting conversion to online and neural networks

Keywords: handwriting, CNN, Transformer

Abstract: This thesis focuses on the automated conversion of static images of offline handwriting into temporal signals of online handwriting. The complexity of the problem lies in the ambiguity that occurs in areas of writing where the direction or intention of pen movement is unclear. The individual variability in stroke order and orientation further complicates the issue. Classical systems proposed in the state of the art are manually designed with a set of strict rules and heuristics that are difficult to maintain and do not generalize well across different languages and types of content.

Recently, many approaches based on machine learning using neural networks from pairs of offline and online data have been proposed in the literature to address this problem. This thesis represents an effort to extend these neural networks beyond the scale of isolated letters and to apply them to more complex types of content.

The thesis explores two distinct neural network-based approaches to reconstruct the pen trajectory from images of offline handwriting, with the goal of generalizing beyond isolated symbols towards larger handwritten content, including words, text lines, and mathematical equations. The first approach is a

fully convolutional multitask UNet-based neural network, inspired by the method of [26]. This approach demonstrated good results for skeletonization but suboptimal stroke extraction. Experiments on mathematical expressions showed a tendency of the model to over-segment strokes, partly due to insufficient spatial and temporal context and inherent temporal modeling limitations within the CNN architecture.

The second approach builds on the previous skeletonization results to extract sub-strokes from the skeleton and proposes a sub-stroke level modeling with Transformers, consisting of a sub-stroke encoder to extract a sub-stroke embedding (SET) and a decoder (SORT) to order the different sub-strokes using their embeddings, as well as pen up predictions. This approach outperformed the state of the art on text lines and mathematical equations databases and addressed several limitations identified in the literature.

These advancements have expanded the scope of offline-to-online conversion to include entire text lines and generalize to bidimensional content, such as mathematical equations. This paves the way for a the generalization to the scale of a full page of handwriting.