Apprentissage auto-supervisé pour la reconnaissance d'écriture manuscrite

Offre d'emploi de l'Institut de Recherche en Informatique et Systèmes Aléatoires (IRISA, Rennes)

Poste	Doctorat
Affectation	laboratoire IRISA, équipe SHADoc 263 avenue du Général Leclerc, 35000 Rennes (France)
Durée	3 ans
Date de début souhaitée	01/10/2024
Contacts	Bertrand Couasnon bertrand.couasnon@irisa.fr Yann Soullard yann.soullard@univ-rennes2.fr Denis Coquenet denis.coquenet@irisa.fr

Thématiques

> Apprentissage auto-supervisé, réseaux de neurones, reconnaissance d'écriture manuscrite, apprentissage automatique, adaptation à de nouveaux documents

Sujet

Cette thèse s'inscrit dans la thématique de la reconnaissance d'écriture manuscrite. On s'intéressera plus particulièrement à reconnaître du texte manuscrit présent dans des images de pages de documents. L'objectif est ainsi de faciliter l'accès à la transcription automatique de documents divers, potentiellement anciens. Les documents anciens présentent en effet des difficultés accrues liés à la langue ancienne, le style d'écriture, la dégradation des ouvrages ou les numérisations de faibles qualités qui rendent souvent difficile et fastidieux la tâche de transcription d'un ouvrage même pour un expert.

Les systèmes traditionnels, combinant successivement une tâche d'extraction de lignes de texte, un modèle optique reconnaissant les caractères dans les images, et un modèle de langues venant apporter des corrections linguistiques [Soullard et al., 2019], sont progressivement remplacés par des systèmes dit *end-to-end* réalisant ces différentes tâches simultanément [Yousef and Bishop, 2020, Coquenet et al., 2022, 2023]. Ces systèmes s'appuient sur des architectures totalement convolutives [Yousef and Bishop, 2020] ou sur des Transformers [Coquenet et al., 2023] qui s'imposent dans de nombreux domaines comme celui de la reconnaissance d'écriture grâce à l'efficacité des mécanismes

d'attention qu'elles incluent [Barrere et al., 2024, Kang et al., 2022].

Ces systèmes ont ainsi de grandes capacités à analyser des images et reconnaître du texte. Bien que très performants sur le ou les types de documents sur lesquels ils ont été appris, ces systèmes peuvent néanmoins souffrir d'un manque de capaciter à généraliser à de nouveaux documents. Ceci est notamment lié au manque de données supervisées et variés pouvant être exploitées pour l'entraînement de ces modèles.

L'objectif de cette thèse sera de s'intéresser à de l'apprentissage auto-supervisé afin de pouvoir tirer profit d'exemples non supervisés pour l'apprentissage de ces modèles. Ces dernières années, de nombreux travaux se sont intéressés aux approches auto-supervisées afin de pré-entraîner les réseaux. Partant d'une tâche prétexte ne nécessitant pas d'annotations manuelles, le modèle acquiert tout de même de fortes capacités de modélisation. Ces approches sont à la base de tous les modèles de fondation de la litérature, qu'ils soient basés sur du texte (GPT-3 [Brown et al., 2020]), de l'image (DINO [Caron et al., 2021]) ou les deux (CLIP [Radford et al., 2021]). Les tâches prétextes peuvent être diverses : prédiction du token suivant pour GPT-3, apprentissage contrastif pour CLIP et DINO. Plus récemment, des travaux [He et al., 2022] ont montré des résultats remarquables pour la tâche prétexte de reconstruction d'images partiellement masquée, dans le cadre de réseaux Transformers.

Ainsi, on s'intéressera à réaliser de l'apprentissage auto-supervisé pour une tâche de reconnaissance d'écriture manuscrite dans des images. L'objectif est ici double : a) améliorer les capacités en généralisation d'un système générique entraîné sur de nombreux exemples variés ; b) permettre de spécialiser un système sur un corpus non étiqueté pour renforcer les capacités de transcription de ce corpus. L'apprentissage auto-supervisé devra également permettre de réaliser ces objectifs de généralisation et de spécialisation en minimisant les quantités de données annotées nécessaires.

Références

Killian Barrere, Yann Soullard, Aurélie Lemaitre, and Bertrand Coüasnon. Training transformer architectures on few annotated data : an application to historical handwritten text recognition. In *International Journal on Document Analysis and Recognition*, pages –. Springer, 2024.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language models are few-shot learners. In *Advances in Neural Information Processing Systems*, volume 33, pages 1877–1901, 2020.

Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. In *International Conference on Computer Vision*, pages 9630–9640, 2021.

Denis Coquenet, Clément Chatelain, and Thierry Paquet. End-to-end handwritten paragraph text recognition using a vertical attention network. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1) :508–524, 2022.

Denis Coquenet, Clément Chatelain, and Thierry Paquet. Dan : a segmentation-free document at-

tention network for handwritten document recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross B. Girshick. Masked autoencoders are scalable vision learners. In *Conference on Computer Vision and Pattern Recognition*, pages 15979–15988, 2022.

Lei Kang, Pau Riba, Marçal Rusiñol, Alicia Fornés, and Mauricio Villegas. Pay attention to what you read : non-recurrent handwritten text-line recognition. *PR*, 129 :108766, 2022.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 8748–8763, 2021.

Yann Soullard, Wassim Swaileh, Pierrick Tranouez, Thierry Paquet, and Clément Chatelain. Improving text recognition using optical and language model writer adaptation. In *2019 International Conference on Document Analysis and Recognition (ICDAR)*, pages 1175–1180. IEEE, 2019.

Mohamed Yousef and Tom E Bishop. Origaminet : weakly-supervised, segmentation-free, one-step, full page text recognition by learning to unfold. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14710–14719, 2020.

Compétences attendues

- Formation souhaitée : Apprentissage automatique, réseaux de neurones
- Compétences techniques :
  - Programmation Python
  - Une expérience avec un environnement dédié au deep learning (Keras, TensorFlow, PyTorch)

Modalité de candidature

- CV
- Lettre de motivation
- Notes de master ou d'école d'ingénieur
- Lettres de recommandation (facultatif)

Equipe SHADoc

L'équipe Shadoc se concentre sur la modélisation des données manuelles de la communication écrite : écriture, geste (2D et 3D), et documents, sous différents aspects : analyse, reconnaissance, composition, interprétation.
L'objectif est de réaliser un continuum entre le papier et les documents numériques avec une certaine lisibilité. Nous nous concentrons principalement sur les sujets suivants :

- Reconnaissance intelligente de contenus manuscrits : documents, écrits, gestes ;
- Analyse du contenu sémantique/structurel : structure des documents, étapes de production des diagrammes, dessins, partitions musicales, croquis, plans d'architecture ;
- Conception de nouvelles IA, combinant reconnaissance et analyse : offrir des expériences enrichies pour les humanités numériques ou l'e-éducation.

La feuille de route de l'équipe Shadoc se situe à la frontière de plusieurs axes de recherche : Reconnaissance des formes, apprentissage automatique, intelligence artificielle, interaction homme-machine, usages et apprentissage numérique.
Nos recherches se caractérisent par l'hybridation de plusieurs approches d'IA : grammaires bidimensionnelles, apprentissage profond, systèmes d'inférence flous... Cette hybridation vise à garantir, au-delà de la performance, des aspects importants tels que : l'explicabilité, la généricité, l'adaptabilité, la frugalité des données.
Au-delà de l'hybridation, l'originalité de cette recherche est de mettre l'accent sur l'interaction avec l'utilisateur. Cette stratégie vise à répondre aux limites des approches actuelles qui sont basées sur des traitements non interactifs. Le concept est de renforcer les processus de décision en s'appuyant sur les validations implicites ou les corrections explicites d'un utilisateur afin d'éviter la propagation d'erreurs tout au long de l'analyse. Les notions d'interprétation, d'adaptation et d'apprentissage incrémental sont au cœur de cette recherche, l'objectif étant de concevoir des systèmes efficaces, robustes et auto-évolutifs.

L'institut de Recherche en Informatique et Systèmes Aléatoires

L'IRISA (Institut de Recherche en Informatique et Systèmes Aléatoires) est aujourd'hui l'un des plus grands laboratoires de recherche français (+ de 850 personnes) dans le domaine de l'informatique et des nouvelles technologies de l'information.
Structuré en sept départements scientifiques, le laboratoire constitue un pôle de recherche d'excellence avec des priorités scientifiques telles que : la bio-informatique, la sécurité des systèmes, les nouvelles architectures logicielles, la réalité virtuelle, l'analyse des masses de données et l'intelligence artificielle.
Implanté à Rennes, Lannion et Vannes, l'IRISA est au cœur d'un écosystème régional riche en recherche et en innovation et se positionne comme la référence en France avec une expertise reconnue à l'international au travers de nombreux contrats européens et de collaborations scientifiques internationales.
Tourné vers le futur de la science informatique et forcément axé vers l'international, l'IRISA se veut au cœur de la transition numérique de la société et de l'innovation au service de la cybersécurité, la santé, l'environnement et l'écologie, les transports, la robotique, l'énergie, la culture et l'intelligence artificielle.
L'IRISA est issu d'une volonté de collaboration entre neuf établissements tutelles pluridisciplinaires, par ordre alphabétique : CentraleSupélec, CNRS, ENS Rennes, IMT Atlantique, Inria, INSA Rennes, Inserm, Université Bretagne Sud, Université de Rennes. De cette pluridisciplinarité nait une force issue de femmes et d'hommes qui donnent le meilleur d'eux-mêmes pour la recherche fondamentale et appliquée, la formation, les échanges avec d'autres disciplines, les transferts de savoir-faire et de technologie et la médiation scientifique.