

Un modèle de fondation multi-résolution pour les images de documents historiques : Application à la détection d'objets et la recherche d'images

Laboratoire/Entreprise : LITIS – UR 4108, Université de Rouen

Durée : 3 ans (du 01/10/2026 au 30/09/2029)

Contact : Laurent.Heutte@univ-rouen.fr

Date limite de publication : 26 mai 2026

Contexte :

Les bibliothèques, archives et musées conservent des millions de documents anciens numérisés : manuscrits, livres imprimés, registres, cartes ou chartes. Ces documents contiennent de nombreux motifs graphiques – lettrines décorées, symboles, ornements, sceaux ou illustrations – qui sont essentiels pour comprendre l'histoire, les styles artistiques et les pratiques culturelles d'une époque. Aujourd'hui, retrouver automatiquement ces motifs dans de vastes collections reste une tâche difficile, car les documents sont souvent abîmés, très variés visuellement et peu annotés.

Ce projet de thèse vise donc à développer un modèle d'intelligence artificielle généraliste, capable d'apprendre à reconnaître et à retrouver ces motifs graphiques dans des documents historiques, sans avoir besoin de grandes bases de données annotées. Le modèle sera entraîné à partir de très nombreux extraits d'images issus de documents anciens, afin d'apprendre de manière autonome à repérer des formes, des textures et des structures graphiques, même lorsqu'elles apparaissent sous des styles ou des tailles différentes.

L'objectif est de permettre à un utilisateur – archiviste, historien ou chercheur – de fournir un simple exemple de motif (par exemple une lettrine ou un symbole) et de retrouver automatiquement toutes ses occurrences dans une collection de documents. Les résultats pourront être visualisés sous forme de zones mises en évidence dans les images.

À terme, ce travail contribuera à mieux valoriser le patrimoine culturel, à faciliter la recherche historique et à offrir de nouveaux outils numériques pour l'exploration et l'analyse des archives et des collections patrimoniales.

Sujet :

Les documents historiques — manuscrits, imprimés anciens, chartes, enluminures ou registres — contiennent de nombreux motifs graphiques (lettrines, symboles, ornements, sceaux, figures iconographiques) dont l'identification et la localisation constituent un enjeu majeur pour les humanités numériques, les archives et l'histoire de l'art. La recherche automatique de ces motifs (*pattern spotting*) permet notamment l'indexation, la comparaison stylistique et l'analyse iconographique à grande échelle. L'objectif de cette thèse est de concevoir le premier modèle de fondation multirésolution spécifiquement dédié aux documents historiques, et capable d'apprendre de manière auto-supervisé, à partir d'images de tels documents, des représentations visuelles robustes face aux variations d'échelle, de style graphique et aux dégradations propres aux fonds patrimoniaux, afin de permettre une recherche de motifs graphiques efficace et performante. Les approches existantes, issues du *symbol spotting* ou du *word spotting*, restent toutefois limitées par la variabilité stylistique extrême, la multiplicité des échelles, les dégradations physiques (tâches, lacunes, bleed-through) et la rareté des annotations, ce qui freine leur généralisation à des collections hétérogènes. L'émergence récente des *foundation models* visuels auto-supervisés (MAE, DINOv2, iBOT, CLIP,...) offre une opportunité scientifique majeure pour apprendre des représentations génériques, robustes et transférables à partir de grandes masses de données non annotées. Toutefois, ces modèles n'ont pas été conçus spécifiquement pour les documents historiques ni pour la recherche de motifs graphiques multi-échelle. Cette thèse vise à combler ce manque.

Un premier objectif consistera à développer un backbone auto-supervisé multi-échelle, pré-entraîné sur des millions de patches extraits de collections historiques hétérogènes, afin de capturer conjointement des informations locales fines (traits, contours, structures graphiques) et des caractéristiques plus globales liées aux textures et aux styles. À partir de ce backbone, la thèse ambitionne de concevoir un module de *pattern spotting* en zero-shot et few-shot, permettant, à partir d'une simple image requête, de produire des cartes de similarité denses assurant la localisation automatique des occurrences d'un motif dans un document, sans recourir à un module de détection explicite. Un autre objectif majeur sera d'établir une évaluation standardisée du *spotting* de motifs historiques, en définissant des protocoles combinant recherche, localisation et robustesse, applicables à plusieurs collections de référence. Enfin, le projet prévoit le développement d'un démonstrateur logiciel interactif, destiné aux archivistes et aux chercheurs en humanités numériques, facilitant l'exploration visuelle des documents et l'analyse des motifs détectés.

Les principaux verrous scientifiques de cette thèse concernent tout d'abord l'apprentissage de représentations réellement multi-résolution, capables de rester cohérentes entre des niveaux de détail très fins et des structures graphiques globales, dans un contexte où les documents présentent une forte hétérogénéité visuelle et un bruit important. La variabilité stylistique extrême des motifs graphiques, liée aux époques, aux techniques de dessin ou d'impression et aux supports matériels, constitue un défi majeur pour la généralisation des modèles. À cela s'ajoutent les dégradations physiques inhérentes aux documents anciens — tâches, lacunes, trous, décolorations ou artefacts de numérisation — qui perturbent fortement les méthodes classiques de correspondance visuelle. Un autre verrou clé réside dans la capacité à assurer une localisation fine des motifs sans recourir à des détecteurs explicites, en s'appuyant uniquement sur des mécanismes d'attention ou de similarité dense. Enfin, le modèle devra démontrer une forte capacité de généralisation entre collections et périodes historiques distinctes, condition indispensable pour un usage réel dans des contextes patrimoniaux variés et peu annotés.

La méthodologie proposée pour cette thèse s'appuiera d'abord sur la constitution d'un large corpus multirésolution, obtenu par l'extraction automatique de millions de patches à différentes échelles à partir de collections de documents historiques, enrichi par des augmentations spécifiques simulant des dégradations réalistes. Sur cette base, un *foundation model* auto-supervisé multi-échelle sera développé, en explorant des approches de type MAE, DINOv2 ou iBOT adaptées au contexte des documents patrimoniaux, ainsi que des architectures potentiellement équivariantes afin de garantir la cohérence des représentations face aux transformations géométriques. Un soin particulier sera apporté à l'alignement des représentations entre échelles par des fonctions de coût dédiées. À partir de ce backbone, un module de pattern spotting sera conçu, reposant sur des mécanismes de cross-attention permettant de générer, à partir d'une requête visuelle, des cartes de similarité denses assurant à la fois la recherche et la localisation des motifs. L'évaluation du système s'appuiera sur la mise en place d'un benchmark complet, combinant des tâches de retrieval, de localisation, de robustesse et de few-shot spotting, et permettant des comparaisons rigoureuses avec les méthodes existantes, classiques ou profondes. Enfin, les contributions de la thèse seront valorisées à travers le développement d'un démonstrateur logiciel interactif, facilitant l'exploration des documents, la visualisation des heatmaps et l'exploitation des résultats par les acteurs des humanités numériques.

Les travaux de thèse s'appuieront sur plusieurs jeux de données de référence issus de fonds patrimoniaux numérisés, en particulier le jeu de données DocExplore¹ [En et al., 2017], dédié aux motifs médiévaux et aux lettrines, ainsi que sur des corpus tels que le corpus HORAE [Boillet et al., 2019] provenant du fonds numérisé Gallica (BnF)² et d'autres fonds numérisés des grandes bibliothèques européennes comme la British Library³, ou encore d'autres corpus comprenant des imprimés anciens, des enluminures et des registres numérisés. Des fonds d'archives départementales seront également exploités, notamment pour l'étude de sceaux et d'ornements graphiques, de même que des collections muséales offrant des iconographies variées de manuscrits et de symboles. Il est à noter que tous ces corpus sont bien souvent non annotés, ou alors uniquement pour certaines tâches.

Enfin, un jeu de données complémentaire sera constitué au cours de la thèse selon une approche de semi-supervision, afin d'enrichir les motifs répétés et de renforcer la capacité de généralisation du modèle à des collections hétérogènes. Nous avons déjà initié cette réflexion et avons constitué un premier jeu de données exploitable pour ces expérimentations dans le cadre des thèses de Zacarias Curi Filho [Curi Filho, 2023] et de Joseph Assaker [Assaker, 2025].

Programme de travail :

La première année de thèse sera consacrée à une revue approfondie de l'état de l'art sur les *foundation models*, en particulier les foundation models dédiés aux images de documents, le pattern spotting et l'analyse de documents historiques, ainsi qu'à la mise en place du pipeline d'extraction multi-échelle et à l'entraînement d'un premier backbone auto-supervisé. Ces travaux initiaux donneront lieu à une première valorisation scientifique dans des conférences ou revues du domaine (ICCV, ICPR, DAS, ICDAR).

La deuxième année portera sur le développement du modèle multirésolution complet, l'intégration du module de pattern spotting basé sur des cartes de similarité query-driven, ainsi que sur la définition des protocoles d'évaluation, la construction des benchmarks et les comparaisons avec l'état de l'art, avec pour objectif des publications dans des conférences et revues de référence (CVPR, ICPR, IJDA).

La troisième année sera dédiée aux approfondissements méthodologiques, notamment l'étude d'architectures équivariantes et des scénarios few-shot, au développement du démonstrateur logiciel

¹ <https://spotting.univ-rouen.fr>

² <https://gallica.bnf.fr/>

³ <https://www.bl.uk/>

complet, ainsi qu'à la rédaction et à la soutenance de la thèse, avec une valorisation finale visée dans une revue internationale de premier plan (Pattern Recognition, TPAMI).

Références:

[Assaker, 2025] J. Assaker. Unsupervised Learning-based Information Retrieval Applied to Spot Patterns in Historical Document Images. Thèse de l'Université de Rouen Normandie, France, 2025.

[Boillet et al., 2019] M. Boillet, ML. Bonhomme, D. Stutzmann, Christopher Kermorvant. HORAE: an annotated dataset of books of hours. *the 5th International Workshop on Historical Document Imaging and Processing*, Sep 2019, Sydney, Australia. pp.7-12, 2019.

[Curi Filho, 2023] Z. Curi Filho. Image Retrieval and Pattern Spotting in Historical Documents using Fully Convolutional Approaches. PhD manuscript, PUCPR, Curitiba, Brazil, 2023

[En et al., 2017] En, S., Nicolas, S., Petitjean, C., Jurie, F., Heutte, L. New public dataset for spotting patterns in medieval document images. *Journal of Electronic Imaging*, vol. 26, no. 1, 2017.

Profil du candidat :

Master 2 en informatique, mathématiques appliquées, ou école d'ingénieur

Le candidat devra impérativement envoyer son CV et ses relevés de notes (L3, M1 et année en cours, avec les classements), ainsi qu'une lettre de motivation, à Laurent HEUTTE et Stéphane NICOLAS (cf coordonnées ci-après), au plus tard pour le 22 mai 2026.

Formation et compétences requises :

Le candidat recherché doit être titulaire d'un Master (ou équivalent) dans le domaine de l'Informatique avec une dominante Traitement du Signal et des Images ou Sciences des Données. Il doit avoir de solides connaissances en apprentissage et classification, notamment en Deep Learning, et éventuellement déjà des compétences en Image Retrieval.

Adresse d'emploi :

Equipe d'accueil:

Equipe Apprentissage, laboratoire LITIS (UR 4108), Université de Rouen Normandie

<http://www.litislab.fr/equipe/docapp/>

Encadrement :

Laurent HEUTTE (directeur), laurent.heutte@univ-rouen.fr, (+33) 2 32 95 50 14

Stéphane NICOLAS (co-encadrant), stephane.nicolas@univ-rouen.fr, (+33) 2 32 95 52 14

Financement :

Allocation établissement Université Rouen