

Apprentissage de représentations pour le pattern spotting et le pattern discovery dans les images de documents historiques

Laboratoire/Entreprise : LITIS- EA 4108, Université de Rouen

Durée : 3 ans (du 01/10/2021 au 30/09/2024)

Contact : Laurent.Heutte@univ-rouen.fr

Date limite de publication : 7 mai 2021

Contexte :

Ce sujet de recherche s'inscrit dans le cadre d'une collaboration entre informaticiens, historiens et archivistes initiée en 2009 par le projet DocExplore 2009-2013 (<http://www.docexplore.eu>), projet du Programme de Coopération Transfrontalière Franco-Britannique Interreg IVa France (manche) – Angleterre. Cette collaboration avec historiens et archivistes a été poursuivie au niveau régional dans le cadre du projet PlaIR 2.0 soutenu par le GRR TL-TI de 2013 à 2016, puis étendue à d'autres acteurs dans le cadre du projet PlaIR2018 soutenu par le FEDER et la Région Normandie de 2017 à 2020. Cette collaboration vise à l'élaboration d'une plateforme logicielle pour l'étude et la valorisation de documents historiques anciens, en particulier médiévaux, dans le but de faciliter le travail des historiens qui étudient ces documents et des conservateurs qui cherchent à les valoriser. Cette plateforme doit leur offrir des fonctionnalités avancées d'analyse d'images et de l'écriture, et de recherche d'information par indexation automatique (<http://spotting.univ-rouen.fr>).

Du point de vue fondamental, les travaux proposés dans cette thèse s'inscrivent dans le thème "Apprentissage conjoint représentation/décision" de l'équipe Apprentissage du LITIS et concernent plus particulièrement un des points importants développés dans l'équipe à savoir l'apprentissage automatique de représentation pour des tâches de détection.

Sujet :

L'objectif de cette thèse est de développer des techniques robustes de détection de patterns (pattern spotting) et de découverte de motifs (pattern discovery) dans les images de documents, en s'appuyant sur les avancées récentes en Deep Learning. Le pattern spotting permet de rechercher et de localiser précisément, dans l'image d'un document, les occurrences d'un « objet » graphique, c'est à dire une forme plus ou moins complexe telle par exemple qu'un logo, une signature, une lettrine, un symbole, une croix, un blason, ... la requête étant formulée en désignant dans l'image un exemple de l'objet à rechercher (requête image). L'intérêt du pattern spotting est de faciliter la recherche d'information dans des bases de documents historiques numérisés relativement complexes comme des documents médiévaux par exemple. Le pattern discovery permet quant à lui d'identifier automatiquement dans les grandes bases d'images de documents, des catégories de motifs graphiques, ou plus généralement des objets, de manière non supervisée, c'est à dire sans connaissance a priori sur les classes d'objets, ni même sur le nombre de classes possibles. L'objectif est de pouvoir découvrir dans les images de documents des structures graphiques qui se répètent ou qui sont similaires lorsqu'elles sont analysées à un certain niveau d'abstraction. Ces deux modes d'utilisation, en recherche et en découverte, d'un tel système d'indexation pourraient être d'une grande utilité pour les historiens, afin de trouver de manière efficace des motifs spécifiques dans des grandes bases d'images de documents hétérogènes, ou de découvrir des relations entre des motifs similaires présents dans des manuscrits différents et présentant des variations de style de représentation plus ou moins importantes.

Nous nous appuyerons pour cela sur les travaux menés dans le cadre de la thèse de Sovann En (soutenue en 2016) où nous avons proposé un système complet de recherche d'images et de localisation des objets graphiques de petite taille dans des images de documents médiévaux [En et al., 2016]. Ce système est basé sur une première extraction/indexation des régions d'intérêt dans l'image (region proposal / Binarized Normed Gradients), d'une caractérisation de ces régions par des descripteurs ad-hoc (Vector of Locally Aggregated Descriptors et Fisher Vector), et d'une recherche par similarité à la requête intégrant des techniques de compression et d'approximation (Inverted File, Product Quantization et Asymmetric Distance Computation). Si ce système a montré de bonnes performances sur le corpus d'images de documents étudié [En et al., 2017], il souffre toutefois d'un certain nombre de faiblesses qui rendent ce système peu adaptable à d'autres types d'images de documents (l'information couleur n'est actuellement pas exploitée par exemple), très sensible aux variations de taille, de forme, de couleur et plus généralement de style, des motifs à détecter. D'autre part, ce système supporte difficilement le passage à l'échelle et nécessite des post-traitements pour une localisation fine des objets dans les régions d'intérêt, à l'aide par exemple de méthodes classiques de matching. Enfin, le mode d'interrogation supporté suppose que l'utilisateur puisse présenter au système un exemple graphique visuellement ressemblant de l'objet qu'il souhaite rechercher. Cette

condition est très forte, et difficile en pratique à réaliser. Il serait plus pratique pour l'utilisateur de pouvoir fournir au système de recherche, une description sémantique des objets qu'il recherche, ou bien qu'il puisse en donner une description graphique plus sommaire (par exemple à partir d'un schéma ou d'un dessin à main levée). Il faut donc que le système d'indexation et de recherche soit plus tolérant aux variations de représentation (ou de style graphique) d'un même objet, et qu'il permette de lier une description sémantique de haut niveau à de multiples représentations graphiques d'un même objet, ce qui suppose d'être dans un contexte supervisé pour apprendre des modèles d'objets préalablement identifiés. Or il n'est pas possible de savoir a priori sur quel type d'objet va porter la recherche de l'utilisateur. Un moyen de contourner ce problème peut être alors d'apprendre de manière non supervisée lors de l'indexation quelles sont les structures similaires présentes dans les données (corpus indexé) à différents niveaux de représentation.

L'objectif de la thèse est donc d'explorer les techniques d'apprentissage de représentation (deep learning) récemment proposées dans la communauté « object detection » pour contourner ces difficultés. Dans un premier temps, il s'agira d'étendre les capacités du système de spotting pour le rendre moins sensible aux variations de représentation (en termes de taille, de forme ou encore de couleur). Nous pourrions nous appuyer pour cela sur les techniques de type Faster R-CNN [Ren et al., 2017] qui devront être étudiées et adaptées pour remplacer avantageusement l'extraction de régions d'intérêt (region proposal) basée sur BING. De même, VLAD et Fisher Vector supportent mal la caractérisation des petites régions et des textures couleur ; on pourra s'inspirer des deep features, comme celles proposées par exemple par [Zhou et al., 2016] ou [Babenko et al., 2015], pour une meilleure caractérisation des régions. Enfin, les techniques de Deep Supervised Hashing, comme celles proposées récemment dans [Liu et al., 2016] ou [Jiang and Li, 2017], devraient permettre de faire face au passage à l'échelle pour une recherche par similarité plus efficace. Cette première partie de la thèse fera également suite à plusieurs travaux réalisés dans le cadre de collaborations internationales entre l'équipe Apprentissage du LITIS et d'autres équipes de recherche [Wiggers et al., 2018], [Wiggers et al., 2019], [Ubeda et al., 2019] et [Ubeda et al., 2020].

Dans un deuxième temps, il s'agira d'étudier l'application de ces modélisations profondes à la découverte de motifs, dans un cadre non supervisé, dans de grands corpus d'images de documents pour permettre une indexation plus fine de ces corpus à différents niveaux de représentation, autorisant ainsi des exploitations de ces contenus indexés qui doivent mieux correspondre aux attentes de l'utilisateur (recherche sémantique de haut niveau, recherche de similarité graphiques, recherche de similarité sémantique). On pourra s'inspirer par exemple de techniques récentes telles que celles proposées dans [Doersch et al., 2015], [Seguin et al., 2016] ou [Shen et al. 2019], pour apprendre des représentations adaptées au cadre non supervisé.

L'équipe Apprentissage mettra à disposition pour la réalisation de ce travail de nombreuses collections d'images de documents, données acquises et annotées dans le cadre du projet DocExplore [En et al., 2016] et qui ont fait l'objet d'une convention signée entre l'Université de Rouen et la Bibliothèque Municipale de Rouen. Ces données, annotées au niveau pattern, permettront de conduire une réelle évaluation expérimentale, car en grandeur nature, du travail de recherche qui, par l'importance du sujet et l'originalité des approches proposées, pourra être valorisé par des publications dans des revues internationales de haut niveau et par l'intégration de nouvelles fonctionnalités dans la plateforme PlaIR et la suite logicielle DocExplore.

Références:

[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.

[En et al., 2016] En, S., Petitjean, C., Nicolas, S., Heutte, L. A scalable pattern spotting system for historical documents. *Pattern Recognition*, vol. 54, pp. 149-161, 2016.

[Ren et al., 2017] S. Ren, K. He, R. Girshick, J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks", *Pattern Analysis and Machine Intelligence IEEE Transactions on*, vol. 39, pp. 1137-1149, 2017

[Zhou et al., 2016] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, A. Torralba. Learning Deep Features for Discriminative Localization. *CVPR2016*, pp. 2921-2929, 2016.

[Babenko et al., 2015] Babenko, V. Lempitsky. Aggregating Local Deep Features for Image Retrieval. *ICCV 2015*, pp. 1269-1277, 2015.

[Liu et al., 2016] H. Liu, R. Wang, S. Shan, X. Chen. Deep Supervised Hashing for Fast Image Retrieval; *CVPR 2016*, pp. 2064-2072, 2016.

[Jiang and Li, 2017] Q.Y Jiang, W.J. Li. Asymmetric Deep Supervised Hashing. arXiv preprint arXiv:1707.08325, 2017.

[Ubeda et al., 2020] I. Ubeda, J. Saavedra, S. Nicolas, C. Petitjean, L. Heutte. Improving pattern spotting in historical documents using feature pyramid networks. Pattern Recognition Letters, vol. 131, pp. 398-404, 2020.

[Ubeda et al., 2019] I. Ubeda, J. Saavedra, S. Nicolas, C. Petitjean, L. Heutte. Pattern spotting in historical documents using convolutional models. 5th International Workshop on Historical Document Imaging and Processing, HIP@ICDAR 2019, Sydney, NSW, Australia, pp. 60-65, 2019.

[Wiggers et al., 2018] K. Wiggers, A. Britto, L. Heutte, A. Koerich, L. Oliveira. Document image retrieval using deep features. 2018 International Joint Conference on Neural Networks, IJCNN2018, Rio de Janeiro, Brazil, pp. 1-8, 2018.

[Wiggers et al., 2019] K. Wiggers, A. Britto, L. Heutte, A. Koerich, L. Oliveira. Image retrieval and pattern spotting using siamese neural network. International Joint Conference on Neural Networks 2019, IJCNN2019, Budapest, Hungary, pp. 1-8, 2019.

[Doersch et al., 2015] Doersch, A. Gupta, A. Efros. Unsupervised visual representation learning by context prediction. ICCV2015, pp. 1422–1430, 2015.

[Seguin et al., 2016] Seguin, C. Striolo, I. di Lenardo, F. Kaplan. Visual link retrieval in a database of paintings. ECCV2016, pp. 753–767, 2016.

[Shen et al., 2019] X. Shen, A. Efros, M. Aubry. Discovering Visual Patterns in Art Collections With Spatially-Consistent Feature Learning. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 9278-9287, 2019.

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 7 mai 2021.

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 des compétences en Image Retrieval.

Adresse d'emploi :

Equipe d'accueil:

Equipe Apprentissage, laboratoire LITIS (EA 4108), Université de Rouen

<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 IA ou allocation établissement Université Rouen Normandie (financement susceptible d'être attribué)

Context :

This research topic is part of a collaboration between computer scientists, historians and archivists initiated in 2009 by the DocExplore 2009-2013 project (<http://www.docexplore.eu>), a project of the Interreg IVa France (Channel) - England Cross-Border Cooperation Programme. This collaboration with historians and archivists was continued at the regional level within the framework of the PlalR 2.0 project supported by the TL-TI RWG from 2013 to 2016, then extended to other actors within the framework of the PlalR2018 project supported by the ERDF and the Normandy Region from 2017 to 2020. This collaboration aims to develop a software platform for the study and enhancement of ancient historical documents, in particular medieval ones, with the aim of facilitating the work of historians who study these documents and curators who seek to enhance them. This platform should offer them advanced functionalities for image and handwriting analysis, and for information retrieval by automatic indexing (<http://spotting.univ-rouen.fr>).

From a fundamental point of view, the work proposed in this thesis is part of the "Joint representation/decision learning" theme of the LITIS Learning team and concerns more particularly one of the important points developed in the team, namely automatic representation learning for detection tasks.

Subject :

The objective of this thesis is to develop robust pattern spotting and pattern discovery techniques in document images, based on recent advances in Deep Learning. Pattern spotting allows to search and locate precisely, in the image of a document, the occurrences of a graphical "object", i.e. a more or less complex shape such as a logo, a signature, a letter, a symbol, a cross, a coat of arms, ... the request being formulated by designating in the image an example of the object to be searched (image request). The interest of pattern spotting is to facilitate the search for information in databases of relatively complex digitised historical documents such as medieval documents for example. Pattern discovery allows the automatic identification of categories of graphic patterns, or more generally objects, in large document image databases in an unsupervised manner, i.e. without any a priori knowledge of the classes of objects, or even the number of possible classes. The objective is to be able to discover in document images graphical structures that are repeated or similar when analysed at a certain level of abstraction. These two modes of use, in search and discovery, of such an indexing system could be of great use to historians, in order to efficiently find specific patterns in large databases of heterogeneous document images, or to discover relationships between similar patterns present in different manuscripts and presenting more or less important variations in representation style.

For this, we will build on the work carried out in Sovann En's thesis (defended in 2016) where we proposed a comprehensive system for image retrieval and localization of small graphic objects in medieval document images [En et al., 2016]. This system is based on a first extraction/indexation of regions of interest in the image (region proposal / Binarized Normed Gradients), a characterization of these regions by ad-hoc descriptors (Vector of Locally Aggregated Descriptors and Fisher Vector), and a search by similarity to the query integrating compression and approximation techniques (Inverted File, Product Quantization and Asymmetric Distance Computation). While this system has shown good performance on the corpus of document images studied [En et al., 2017], it nevertheless suffers from a certain number of weaknesses that make this system not very adaptable to other types of document images (colour information is not currently exploited, for example), very sensitive to variations in size, shape, colour and more generally style, of the patterns to be detected. On the other hand, this system does not easily support scaling and requires post-processing for a fine localisation of objects in the regions of interest, using for example classical matching methods. Finally, the supported query mode assumes that the user can present the system with a visually similar graphical example of the object he/she wishes to search. This condition is very strong, and difficult in practice to achieve. It would be more convenient for the user to be able to provide the search system with a semantic description of the objects he is looking for, or else to be able to give a more summary graphical description of them (for example, from a diagram or freehand drawing). The indexing and retrieval system must therefore be more tolerant of variations in the representation (or graphical style) of the same object, and must be able to link a high-level semantic description to multiple graphical representations of the same object, which presupposes that it is in a supervised context to learn models of previously identified objects. However, it is not possible to know a priori on which type of object the user will search. A way to circumvent this problem can be to learn in an unsupervised way during the indexing what are the similar structures present in the data (indexed corpus) at different levels of representation.

The objective of the thesis is therefore to explore the deep learning techniques recently proposed in the object detection community to overcome these difficulties. The first step will be to extend the capabilities of the spotting system to make it less sensitive to variations in representation (in terms of size, shape or colour). We will be able to rely on Faster R-CNN type techniques [Ren et al. 2017] which will have to be studied and adapted to replace advantageously the extraction of regions of interest (region proposal) based on BING. Similarly, VLAD and Fisher Vector do not support well the characterization of small regions and colour textures; one could be inspired by deep features, such as those proposed by [Zhou et al., 2016] or [Babenko et al., 2015], for a better characterization of regions. Finally, Deep Supervised Hashing techniques, such as those recently proposed in [Liu et al., 2016] or [Jiang and Li, 2017], should help to cope with the scaling up for a more efficient similarity search. This first part of the thesis will also follow up on several works carried out in international collaborations between the LITIS Learning team and other research teams [Wiggers et al., 2018], [Wiggers et al., 2019], [Ubeda et al., 2019] and [Ubeda et al., 2020].

In a second step, we will study the application of these deep models to pattern discovery, in an unsupervised framework, in large corpora of document images in order to allow a finer indexing of these corpora at different levels of representation, thus authorising exploitations of these indexed contents that should better correspond to the user's expectations (high-level semantic search, graphical similarity search, semantic similarity search). Recent techniques such as those proposed in [Doersch et al., 2015], [Seguin et al., 2016] or [Shen et al. 2019] could be used to learn representations adapted to the unsupervised framework.

The Learning team will make available for this work numerous collections of document images, data acquired and annotated in the framework of the DocExplore project [En et al., 2016] and which were the subject of an agreement signed between the University of Rouen and the Municipal Library of Rouen. These data, annotated at the pattern level, will enable a real experimental evaluation, because in real size, of the research work which, by the importance of the subject and the originality of the proposed approaches, could be valued by publications in high-level international journals and by the integration of new functionalities in the PlAIR platform and the DocExplore software suite.

Candidate profile :

Master 2 in computer science, applied mathematics, or engineering school.

The candidate must send his/her CV and transcripts (L3, M1 and current year, with rankings), as well as a letter of motivation, to Laurent HEUTTE and Stéphane NICOLAS (see contact details below), by May 7, 2021 at the latest.

Education and required skills :

The desired candidate must have a Master's degree (or equivalent) in Computer Science with a major in Signal and Image Processing or Data Science. He/she should have a solid knowledge of learning and classification, especially in Deep Learning, and skills in Image Retrieval.

Job address :

Host team: Equipe Apprentissage, laboratoire LITIS (EA 4108), Université de Rouen
<http://www.litislab.fr/equipe/docapp/>

Supervision :

Laurent HEUTTE (director), laurent.heutte@univ-rouen.fr, (+33) 2 32 95 50 14
Stéphane NICOLAS (co-supervisor), stephane.nicolas@univ-rouen.fr, (+33) 2 32 95 52 14

Funding :

IA allocation or Université de Rouen Normandie institutional allocation (funding likely to be awarded)