

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 01/10/2018 au 30/09/2021)

Contact : Laurent.Heutte@univ-rouen.fr

Date limite de publication : 30 avril 2018

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 appuyons 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 efficiente.

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] ou [Seguin et al., 2016] 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.

[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.

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 30 avril 2018.

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 établissement Université Rouen Normandie (financement susceptible d'être attribué)