Contrats doctoraux
en Intelligence Artificielle

Projet AHEAD
Artificial Intelligence for Health, Physics, Transportation and Defense

Deep learning pour la recherche visuelle par le contenu d'images de logos de marques

1- Contexte

La recherche visuelle par le contenu consiste à retrouver dans une base de données des images similaires à une requête. C'est une tâche de reconnaissance visuelle historique qui présente des applications dans un très large spectre de domaines, depuis la recherche mobile, la robotique, l'assistance médicale, *etc*.

Co-financement et collaboration

Cette thèse s'inscrit dans le cadre du projet AHEAD (Artificial Intelligence for Health, Physics, Transportation and Defense) porté par le Cnam Paris et financé par l'ANR dans le cadre des contrats doctoraux en Intelligence Artificielle. La thèse est co-financée par SWORD-GROUP, ESN déployée à l'international, qui est un acteur majeur du développement logiciel dans le secteur de la protection de la propriété industrielle (marques, brevets, dessins et modèles).
L'objectif applicatif de la thèse consiste à mettre en place des solutions logicielles d'analyse dans le cadre d'une recherche d'antériorité de marque. Dans le cas d'un dépôt de marque figurative, représentée par un logo, la recherche d'antériorité consiste à assurer que le logo de la marque envisagée n'est pas similaire à celui d'une marque existante (voir Figure 1).

Positionnement du sujet de thèse

Le domaine de la recherche visuelle par le contenu a considérablement évolué au cours des dernières années par le recours à des méthodes d'apprentissage profond (« deep learning »). SWORD édite un logiciel permettant de comparer un logo (image question) à différentes bases de données. La première mouture de ce logiciel était basée sur des techniques déterministes de reconnaissance de formes ; ses versions récentes intègrent des technologies de deep learning. En particulier, des réseaux convolutifs profonds entraînés à partir de bases de plusieurs millions d'images étiquetées selon une classification métier (la classification internationale de Vienne) constituent aujourd'hui le cœur de la chaîne de traitement pour le calcul de similarité sémantique dans le contexte de la tâche de recherche d'antériorités de marques.

Image Requête :

Résultats de la recherche :

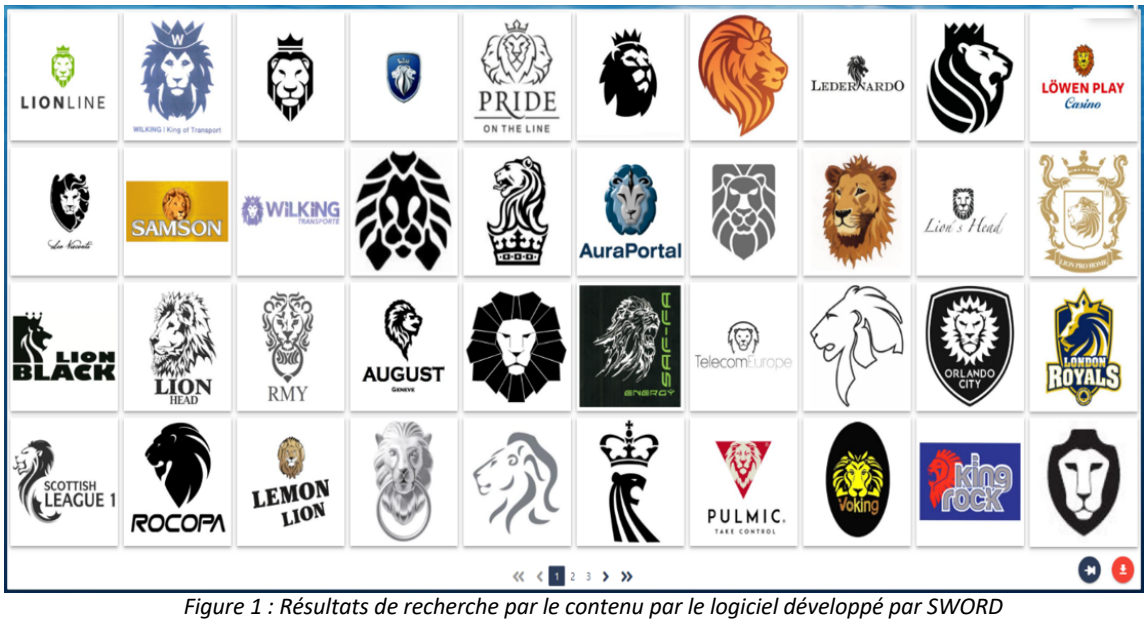

Figure 1 : Résultats de recherche par le contenu par le logiciel développé par SWORD

2- Objectifs

L'objectif général de la thèse est d'améliorer les méthodes de deep learning pour la recherche par le contenu d'images de logos. Il s'agit en particulier d'interroger la notion de similarité dans le contexte métier et d'améliorer la qualité de la mesure de similarité entre images, tout en respectant des contraintes de temps de calcul (traitement temps réel). Les axes de recherche envisagés pour y parvenir sont détaillés ci-dessous. A noter que la méthodologie mise en œuvre sera également évaluée sur des bases de données généralistes pour permettre la validation académique des travaux de thèse.

Approches proposées

Axe 1 : Représentations visuelles pour la recherche par le contenu.

En dépit du succès rencontré par les méthodes de deep learning dans le domaine de la recherche d'images depuis 2012 [1], les réseaux de neurones convolutifs profonds encodent une information locale des images relativement rigide, peu adaptée de fait à la problématique de recherche par le contenu d'images de logos.

Nous nous intéresserons à proposer des représentations permettant d'extraire de l'information visuelle de sous-régions d'une image, en s'appuyant sur les méthodes récemment proposées dans ce domaine de la recherche, e.g. MAC, R-MAC [2]. Nous explorerons des extensions de ces méthodes, notamment avec l'objectif d'introduire une mise en correspondance des régions plus explicites dans le calcul de similarité. Une attention

particulière sera portée à l'inclusion d'une étape de segmentation de régions d'intérêt dans la chaîne de traitement, visant à extraire de l'image de logo la/les région(s) contenant l'information utile. Nous étudierons la manière dont la segmentation d'image et les représentations locales peuvent être combinées, notamment à travers un apprentissage global du modèle, comme ceci est le cas pour des tâches de détection d'objets [3].

Axe 2 : Apprentissage profond pour la recherche par le contenu.

Le succès du deep learning repose avant tout sur des tâches de classification, e.g. le succès emblématique obtenu au challenge ILSVRC'12 [1]. Les réseaux convolutifs sur la base ImageNet constituent également des représentations visuelles très performantes et connues sous le nom de « Deep Features » [4], dont l'utilisation pour la recherche par le contenu a récemment montré des résultats très prometteurs [5].

Une première étape de cette thèse consistera à raffiner ces Deep Features (« fine-tuning »). SWORD dans le cadre de son développement logiciel a enrichi ces représentations visuelles en exploitant une grande masse de données annotées disponibles dans ses bases de données métiers. On évaluera le gain relatif de ce raffinement pour la recherche par similarité.

Une seconde étape consistera à aller au-delà des métriques de classification et de proposer des fonctions de coût d'apprentissage directement liées à la problématique finale de recherche par similarité. Nous nous appuierons sur les approches de l'état de l'art basées sur l'introduction de paires, triplets [6] ou quadruplets [7] d'exemples afin de définir des contraintes de distances relatives entre paires d'images similaires et dissimilaires. Nous adapterons ces approches à notre problématique en menant une réflexion pour leur passage à l'échelle, puisque le nombre de contraintes est quadratique ou cubique par rapport au nombre d'exemples. Nous explorerons en particulier des méthodes de recherche de contraintes actives pour sélectionner les exemples non pertinents.

Enfin, la dernière étape consistera à optimiser lors de l'entraînement des réseaux les métriques les plus en lien avec l'application finale, comme la précision moyenne (Average Precision), le Rappel à k ou d'autres métriques reliées (Precision at Recall, NDCG, etc.). Dans ce contexte, nous aborderons deux verrous à lever dans le cadre de la recherche de similarité par deep learning. L'enjeu sera d'abord de définir des variantes dérivables pour ces métriques, qui soient applicables dans un schéma de descente de gradient stochastique. Nous nous appuierons sur la définition de bornes supérieures spécifiquement pour des tâches spécifiques [8, 9, 10], ou des méthodes récentes permettant d'apprendre une mesure de similarité dérivable [11]. Un second aspect à prendre en compte dans le cadre de l'entraînement de réseaux de neurones profonds réside dans le fait que la plupart des métriques d'ordonnancement ne sont pas linéairement décomposables par rapport aux exemples d'entraînement. Nous nous inspirerons de certains travaux récents pour aborder le problème de l'optimisation globale [12], que nous adapterons au contexte de l'apprentissage de similarité par deep learning.

Axe 3 : Passage à l'échelle pour la recherche par le contenu

L'objectif de cette partie est de proposer des approches pour accélérer le calcul de similarité pour rendre possible le passage à l'échelle de la méthode. Dans le contexte applicatif retenu, il est impératif d'effectuer des recherches en un temps raisonnable dans des bases de données contenant plusieurs millions voire dizaines de millions d'images.

Une première méthode consistera à évaluer des stratégies pour compresser des représentations internes des réseaux profonds. Nous adapterons la méthode de référence

« Product Quantization » [13] pour les représentations internes d'images de logo. En particulier, nous ferons le lien avec les méthodes de « Hashing », qui permettent de calculer des signatures binaires entre représentations, rendant l'empreinte mémoire de ces approches très compacte. Nous analyserons également l'impact de la taille de l'espace de représentation sur la qualité de compression afin de définir des critères de seuillage adaptés.
La seconde étape consistera à introduire de la supervision dans ces approches de compression, qui historiquement ont été utilisées sur des descripteurs manuels, avec des schémas d'apprentissage non supervisés. Nous nous appuierons notamment sur les travaux récents menés pour effectuer l'apprentissage de métrique d'ordonnancement dans l'espace compressé [14]. Un enjeu consistera à étudier comment l'étape de compression et création de signature binaire peut être effectuée conjointement à l'apprentissage du modèle profond de recherche par le contenu.
Enfin, l'objectif final sera d'inclure les architectures convolutives pour les représentations visuelles locales mises en place dans l'axe 1 et les schémas d'apprentissage spécifiques à la recherche par le contenu de l'axe 2.

Échéancier

Les premiers mois de la thèse seront consacrés à une étude bibliographique des différents aspects de l'apprentissage profond qui sont au cœur du sujet de thèse, ainsi qu'à la prise en main des outils expérimentaux et des bases de données. L'exploration de l'axe 1 sera mené dès la première année de la thèse. L'axe 2 du programme de recherche sera ensuite entamé dans le courant de la première ou dans la seconde année en fonction de l'avancement. Ce n'est que dans la seconde moitié de la thèse que l'axe 3 devrait quant à lui être abordé.

3- Candidature

Master ou école d'ingénieur à dominante informatique ou mathématiques appliquées
Expérience en machine learning et deep learning, en particulier réseaux convolutifs
Très bonnes compétences en programmation, avec une expérience sur les librairies de deep learning (Tensorflow, Pytorch)
Bonne qualité de synthèse à l'écrit et à l'oral pour la présentation des travaux de recherche. Une expérience d'écriture d'un article serait un plus.

Envoyer un CV et une lettre de motivation à nicolas.thome@cnam.fr, xavier.bitot@sword-group.com

Bibliographie

[1] A. Krizhevsky, L. Sutskever and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems, pages 1097–1105, 2012.

[2] T. Giorgos R. Sicre and H. Jégou. Particular object retrieval with integral max-pooling of CNN. International Conference on Learning Representations,ICLR 2016

[3] S. Ren, K. He, R. Girshick and J. Sun, Jian. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. In Advances in neural information processing systems, 2015.

[4] Azizpour, H., Razavian, A. S., Sullivan, J., Maki, A., and Carlsson, S. Factors of transferability for a generic convnet representation. IEEE Trans. Pattern Anal. Mach. Intell., 38(9):1790–1802, 2016.

[5] A. Gordo, J. Almazán, J. Revaud, D. Larlus. End-to-End Learning of Deep Visual Representations for Image Retrieval. International Journal of Computer Vision (IJCV), 124 (2), pp. 237-254, 2017.

[6] Weinberger, K., & Saul, L.. Distance metric learning for large margin nearest neighbor classification. The Journal of Machine Learning Research (JMLR),10, 207–244, 2009

[7] Marc T. Law, Nicolas Thome, and Matthieu Cord. Learning a Distance Metric from Relative Comparisons between Quadruplets of Images. International Journal of Computer Vision (IJCV), 121:65 – 94, January 2017.

[8] Y. Yue, T. Finley, F. Radlinski, and T. Joachims, “A support vector method for optimizing average precision,” in SIGIR, 2007.

[9] Thibaut Durand, Nicolas Thome, Matthieu Cord. Exploiting Negative Evidence for Deep Latent Structured Models. IEEE Transactions on Pattern Analysis and Machine Intelligence, 41, 337-351 (2019).

[10] Efficient optimization for rank-based loss functions. P Mohapatra, M Rolinek, C Jawahar, V Kolmogorov. In, CVPR2018

[11] M Engilberge, L Chevallier, P Pérez, M Cord. SoDeep: a Sorting Deep net to learn ranking loss surrogates. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019

[12] Sanyal, A., Kumar, P., Kar, P. et al. Optimizing non-decomposable measures with deep networks. Mach Learn 107, 1597–1620 (2018).802.

[13] H. Jégou, M. Douze, and C. Schmid. Product Quantization for Nearest Neighbor Search. IEEE Trans. Pattern Anal. Mach. Intell. 33 (1): 117-128, 2011.

[14] Kun He, Yan Lu, Stan Sclaroff. Local Descriptors Optimized for Average Precision. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018