Boucle d'indexation interactive basée sur les gains d'information : application à un jeu de données expert

Solène VILFROY[1,2], Thierry URRUTY[2], Philippe CARRÉ[2], Lionel BOMBRUN[1]

[1]Laboratoire IMS, Université de Bordeaux, CNRS, UMR 5218, Talence, France,

[2]Laboratoire Xlim, Université de Poitiers, CNRS, UMR 7252, Poitiers, France

solene.vilfroy@u-bordeaux.fr, thierry.urruty@univ-poitiers.fr
philippe.carre@univ-poitiers.fr, lionel.bombrun@u-bordeaux.fr

Résumé – De nombreuses applications d'apprentissage supervisé dans la classification de données nécessitent une quantité considérable de données pour pouvoir entraîner un modèle. Malheureusement, avoir accès à de grandes bases de données d'images étiquetées n'est pas toujours possible dans tous les domaines d'application. L'étiquetage de ces images peut être compliqué et/ou coûteux car il nécessite l'intervention d'experts. Dans ce contexte, certaines études se concentrent sur l'apprentissage adaptatif à une petite quantité de données, tandis que d'autres se concentrent sur une stratégie d'étiquetage efficace et peu coûteuse. Ce papier propose une méthode de recherche d'images par similarité tout en minimisant l'intervention de l'expert. Le cadre d'apprentissage dit d'*active learning* proposé consiste en une boucle d'indexation interactive basée sur les gains d'information. Les gains d'informations sont mis à jour de manière itérative par un réseau de neurones entièrement connectés, sélectionnant automatiquement les images requêtes les plus utiles.

Abstract – Many supervised learning applications in data classification require a considerable amount of data to be able to train a model. Unfortunately, having access to large databases of tagged images is not always possible in all application domains. Labeling these images can be complicated and/or expensive as it requires the intervention of experts. In this context, some studies focus on adaptive learning to a small amount of data, while others focus on an efficient and inexpensive labeling strategy. This article proposes a method for searching images by similarity while minimizing the intervention of the expert. The proposed *active learning* learning framework consists of an interactive indexing loop based on information gains. The information gains are iteratively updated by a *fully connected* neural network, automatically selecting the most useful query images.

1 Introduction

Depuis l'essor du deep learning ces dernières années, les approches d'apprentissage supervisé ont connu une certaine popularité dans le domaine du traitement d'images. La conception de bases de données cohérentes et labellisées destinées à l'apprentissage supervisé représente aujourd'hui l'un des principaux défis dans le domaine du machine learning. Compliquée ou coûteuse, l'acquisition de telles bases de données nous pousse à développer de nouvelles stratégies pour optimiser la labellisation des bases de données ou pour adapter les algorithmes d'apprentissage à des bases de données composées d'un nombre limité d'images. Dans cet article, notre objectif n'est pas d'introduire une nouvelle stratégie d'apprentissage supervisé pour des bases de données limitées mais plutôt d'optimiser la labellisation de ces bases de données. Pour cela, des processus d'indexation interactifs basés sur le principe des méthodes d'active learning peuvent être envisagés [1]. L'idée originale derrière l'active learning est qu'un modèle de machine/deep learning peut atteindre de meilleures performances avec moins d'images d'entraînement si nous sélectionnons soigneusement les données à partir desquelles il apprend [2]. L'active learning interroge interactivement l'expert (l'oracle) en lui demandant d'annoter un ensemble d'images. Une fois annotées, ces images sont utilisées pour entraîner un algorithme de classification supervisée. Les images les plus proches/similaires de ces images requêtes sont alors supposées appartenir à la même classe que ces images requêtes. Les images confirmées par l'oracle sont ensuite ajoutées au sous-ensemble d'images annotées. Toutes ces étapes sont répétées au cours de nombreuses itérations. Pour mettre en place un algorithme d'apprentissage actif efficace, il convient de répondre à différentes questions. Comment le modèle est-il mis à jour à chaque itération en incorporant de nouvelles informations ? Quelles images doivent être montrées à l'oracle ? Quelles stratégies de sélection des images requêtes sont efficaces ? Pour y répondre, une approche de recherche d'image par le contenu est proposée dans ce papier. Cette dernière consistent classiquement en trois principales étapes : l'extraction des signatures des images [3], la sélection d'images requêtes et les annotations de l'expert. L'originalité scientifique de ce papier réside dans le développement d'un cadre d'apprentissage actif utilisant les gains d'information (poids attribués aux signatures pour les discriminer selon les classes) dans une boucle d'indexation interactive.

Le document est structuré de la façon suivante. La partie 2 détaille les différentes étapes de la boucle d'indexation interactive proposée. La partie 3 présente brièvement la base de données utilisée pour valider le cadre proposé, et propose un analyse de sensibilité afin d'évaluer l'influence des différents paramètres et une analyse des performances de classification permet d'identifier les forces et les faiblesses de la méthode proposée. Enfin, la partie 4 donne quelques conclusions et perspectives sur ce travail.

2 Approche proposée

L'approche proposée est une approche itérative et interactive, au cours de laquelle l'expert (l'oracle) est invité à interagir à chaque itération. Il s'agit d'une méthode d'indexation de base de données basée sur une mesure de similarité entre les images. Le principe général est illustré dans la figure 1. Partant d'un base de données totalement non étiquetée, il s'agit d'un processus itératif composé de plusieurs étapes : l'extraction des signatures des images, la sélection des images requête, les annotations de l'expert et la mise à jour des gains d'information (GI). Chacune de ces étapes est détaillée dans les sous-parties suivantes.

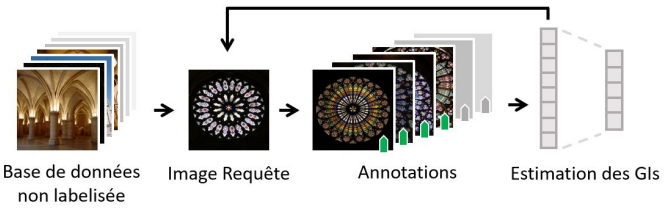

FIGURE 1 – Principe général de la méthode d'indexation interactive proposée.

2.1 Extraction des signatures

Tout d'abord, chaque image se voit attribuer une signature qui encode ses caractéristiques visuelles. Traditionnellement, ces signatures étaient des attributs extraits "manuellement" des images comme les descripteurs SIFT, SURF, etc. Devant le succès grandissant des réseaux de neurones, et en particulier des réseaux de neurones convolutifs (CNN) pour le traitement d'images, l'approche présentée ici propose d'exploiter les couches profondes de ces réseaux. Mais comme initialement, la base de données est non étiquetée, ces modèles ne peuvent pas être entraînés sur la base de données considérée. Pour pallier à cela, une approche de transfert d'apprentissage est mise en place. Elle consiste à utiliser un réseau de neurones comme un extracteur de caractéristiques. Ces réseaux pré-entraînés sur la base ImageNet sont par exemple les modèles VGG16, VGG19 [4], ResNet50 [5], AlexNet[6] etc. Dans la suite du papier, chaque image i est représentée par sa signature $s(i) \in \mathbb{R}^N$ contenant les coefficients obtenus en sortie de l'avant dernière couche entièrement connectée de ces réseaux.

2.2 Sélection des images requêtes

Chaque itération commence par une sélection de nouvelles images requêtes qui viennent s'ajouter aux images déjà annotées. Initialement, cet ensemble est vide et grandit à chaque itération. Dans ce papier, nous proposons de sélectionner les nouvelles images requêtes selon le principe de l'apprentissage actif. L'objectif est de choisir les images qui vont permettre d'améliorer les performances de classification de la base de données, en sollicitant le moins possible l'expert. Typiquement, deux grands types d'approches sont utilisés : l'exploration et l'exploitation. La stratégie d'exploration consiste à sélectionner des images requêtes à la frontière des classes, tandis que la stratégie d'exploitation consiste à sélectionner des images requêtes au centre de chacune des classe. Partant de ce contexte, nous proposons d'utiliser un classifieur SVM [7] entraînés sur les signatures de l'ensemble des images annotées. Dans ce papier, 4 stratégies de requêtes sont étudiées :

— **Random** - Sélection aléatoire d'un nombre n d'images, indépendamment des classes définies.
— **Oracle Random** - Sélection aléatoire d'une image par classe. En pratique, cette stratégie ne peut pas être employée puisque les classes sont inconnues. Cette stratégie servira par la suite de référence par rapport aux autres approches.
— **SVM Var** - La méthodologie d'exploration *SVM Variance* consiste en la sélection des images de la requête parmi les images aux frontières des autres classes. De cette façon, les images les plus déroutantes sont annotées et appliquées au processus d'étiquetage.
— **SVM Max** - Stratégie d'exploitation, la sélection des images requêtes s'effectue au cœur de chaque classe. Nous choisissons donc une image qui, en théorie, peut être un bon représentant pour la majorité des images de la classe.

A la première itération, comme l'ensemble d'images annotées est vide, la stratégie "Random" est utilisée, puis ensuite l'une de ces 4 stratégies est mise en place.

2.3 Annotations de l'expert

Le système d'annotation consiste en deux étapes. Dans un premier temps, les k images les plus proches de chacune des images requêtes se voient attribuées la même classe que l'image requête. La notion de plus proche voisin est défini ici au sens de la distance euclidienne pondérée par les gains d'information (GI). Si i_q est une image requête appartenant à la classe c, alors la distance entre les images i et i_q est calculée ainsi :

$$D(i_q, i) = \sqrt{\sum_{n=1}^{N} GI_{c_n} \times \left(s(i_q)_n - s(i)_n\right)^2} \qquad (1)$$

où $s(\cdot)$ est la fonction permettant d'extraire la signature de l'image, et l'indice n correspond au $n^{\text{ème}}$ élément du vecteur.

Lors de ce calcul de distance, la classe de l'image i étant inconnue, chaque signature est pondérée par les gains d'information estimé pour la classe c de l'image i_q, soit IG_c.

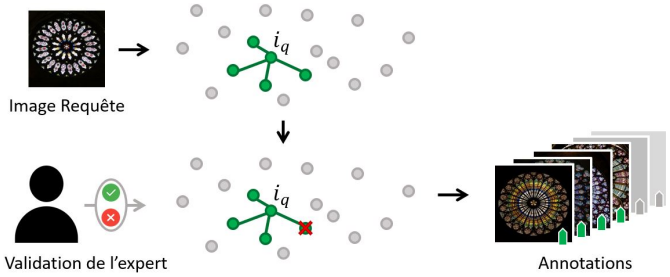

FIGURE 2 – Principe de création des images annotées.

Dans un second temps, comme illustré sur la Figure 2, les k images les plus proches de l'image requête i_q sont présentées à l'expert. Ce dernier valide ou non l'attribution de la classe de i_q à ces images. Les images validées par l'expert rejoignent l'ensemble des images annotées. Celles-ci servent à l'apprentissage du classifieur SVM utilisé dans la sélection des images requêtes (partie 2.2), ainsi qu'à l'estimation des gains d'information détaillés dans la partie suivante.

2.4 La mise à jour des gains d'information

On désigne par le terme gain d'information (GI), les poids attribués aux signatures pour les discriminer selon les différentes classes. Pour une signature de longueur N, les GI sont des vecteurs de taille N contenant les poids pour chacune des classes. Ces GI sont estimés à partir de l'ensemble des images annotées, à l'aide d'un réseau de neurones entièrement connecté composé uniquement de la couche d'entrée contenant les signatures des images et d'une couche de classification de sortie, équipée de la fonction d'activation sigmoïde. Dans la structure neuronale, le poids constitue le facteur qui détermine l'importance accordée à chaque élément de la signature. Les gains d'information par classes sont donc définis comme les poids du réseau. Par exemple, pour la classe 1, les gains d'information sont : $IG_1 : \{W_{1,n}, n \in [1, N]\}$, les N poids associés aux neurones de la couche de classification correspondant à la classe 1.

Ces 3 étapes de sélection des images requêtes, d'annotations de l'expert et de mise à jour des GI sont répétées plusieurs fois formant le processus d'indexation de la base de données. L'intervention de l'expert au sein de cette boucle permet d'affiner les GI afin qu'ils soient de plus en plus discriminants.

3 Résultats des expérimentations

Les expérimentations sont menées sur la base de données *Architectural Heritage Elements* (AHE) [8] (figure 3), composée d'une grande diversité d'éléments architecturaux, acquis selon différents points de vue. Alors que certaines des 10 catégories sont facilement identifiables par un public non initié, certaines peuvent impliquer plus de difficultés.

Autel	Abside	Clocher	Colonne	Dôme Intérieur	Dôme Extérieur	Contre-fort	Gargouille	Vitrail	Voute
829	505	1058	1919	595	1177	407	1571	1032	1108

FIGURE 3 – Exemples d'images de la base AHE, de gauche à droite : *Colonne, Contrefort, Gargouille, Vitrail*. Le tableau donne répartition des 10 201 images dans les 10 classes.

Dans cette partie, nous nous analysons l'influence des principaux paramètres du processus d'indexation proposé, comme le choix du modèle pour l'extraction des signatures, le nombre d'images annotées par l'expert, ainsi que leur répartition parmi les classes.

Pour évaluer les performances, nous sélectionnons et classons les k plus proches voisins de chaque image au sens de la distance euclidienne pondérée par les GI. La précision est égale au nombre d'éléments correctement classifiés sur le nombre total d'images récupérées. Nous utiliserons également comme valeur de comparaison, la précision *optimale*, qui correspond au résultat hypothétique dans un contexte où l'entièreté de la base de données aurait été annotée pour l'apprentissage des GI. Cette valeur correspond à la borne maximale théorique de performance.

Influence du modèle permettant d'extraire la signature Dans cette partie, nous comparons les différentes combinaisons de signatures. Pour cela, nous utilisons la précision *optimale* expliquée précédemment. Les signatures étant créées comme décrit dans la partie 2.1, nous avons testé plusieurs réseaux de neurones, mais aussi plusieurs combinaisons de ces réseaux. Après avoir comparé l'utilisation de plusieurs CNN profonds pour notre méthode, nous nous sommes limités aux réseaux suivants : VGG16, VGG19 et ResNet50 qui ont montré les meilleures performances. Ces expérimentations montrent que l'influence des réseaux choisis est limitée avec une précision rapidement bornée comme le montre le tableau 1.

VGG16	71.63
VGG19	70.15
ResNet50	69.93

ResNet50-VGG16	73.39
ResNet50-VGG19	73.13
VGG16-VGG19	72.68
VGG16-VGG19-ResNet50	73.38

TABLE 1 – Précision *optimale* pour différentes signatures

On constate que les modèles VGG16 et ResNet50 permettent d'obtenir les meilleurs résultats de classification 73.39%. En concaténant les signatures de ces réseaux, un gain de 2 à 3% est observé par rapport à utilisation individuelle.

Influence de la stratégie de requête Le second paramètre qui influence les performances de classification est la méthode de sélection des images requêtes. Les 4 méthodes décrites dans la partie 2.2 sont testées à savoir : Random, Random Oracle, *SVM Var* et *SVM Max*. A chaque itération, le nombre de nouvelles images requêtes sélectionnées reste le même quelle que soit la méthode sélectionnée. En revanche, selon la méthode de requête, l'expert peut être amené à rejeter plus ou moins d'images d'où une différence sur le nombre d'images annotées. Après 10 itérations, toutes les méthodes aboutissent à une proportion d'environ 18% de la base de données annotée par l'expert, dont on peut voir la répartition sur la figure 4. Aucune des stratégies testées ne permet de garantir l'équilibre entre les classes. Par exemple, la méthode d'exploration *SVM Var* va chercher à annoter beaucoup d'images des classes *Clocher* et *Vitrail* et peu d'images des classes *Autel* et *Dôme intérieur*, tandis que c'est l'inverse pour la stratégie d'exploitation *SVM Max*. La méthode de requêtage influence fortement la sélection des images requêtes et donc la distribution des images annotées.

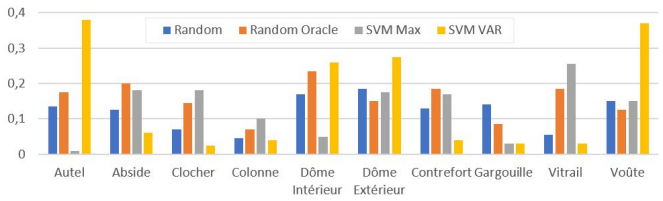

FIGURE 4 – Répartition des images annotées dans les 10 classes après 10 itérations.

La figure 5 montre l'évolution des performances sur 10 itérations, selon les différentes méthodes de sélection des images requêtes. L'itération 0 correspond aux performances obtenues sans les GI, puisqu'ils sont initialisés à 1. Comme on peut le voir sur la figure, l'utilisation des gains d'informations dans la boucle d'indexation interactive permet d'améliorer significativement les performances. La ligne rouge correspond à l'*précision optimale*. Cette performance ne peut être atteinte en pratique puisqu'elle nécessite l'annotation de toute la base de données. Cependant, on peut s'apercevoir qu'avec un pourcentage restreint d'apprentissage ($\simeq$18% pour 10 itérations), on s'approche très près de cette valeur. De plus, si toutes les méthodes arrivent à peu près aux mêmes performances après 10 itérations, la méthode d'exploration basée sur le stratégie *SVM Var* permet d'atteindre cette performance plus rapidement, c'est à dire avec un pourcentage d'images annotées plus faible.

4 Conclusion et perspectives

Dans ce papier, nous avons proposé une méthode itérative et interactive d'indexation d'images. La principale contribution de notre proposition consiste dans l'utilisation des gains d'informations (GI) pour chercher les images plus similaires des images requêtes. Ces GI sont mis à jour à chaque itération

FIGURE 5 – Évolution de la *précision* pour les différentes stratégies de sélection des images requêtes

en entraînant un réseau de neurones entièrement connectés sur l'ensemble des images annotées. Les expériences sur la base Architectural Heritage Elements ont validés l'apport de ces GIs dans le modèle. De plus, les résultats ont montré sur cette base de données que l'approche d'apprentissage actif basée sur le principe d'exploration (modèle *SVM Var*), permet d'augmenter plus rapidement la précision de la classification que la méthode basée sur le principe d'exploitation (*SVM Max*). Les travaux futurs concerneront l'étude de la complémentarité de ces deux types de stratégies afin d'améliorer les performances.

Références

[1] D. Picard at al.. *Challenges in content-based image indexing of cultural heritage collections.* IEEE Signal Processing Magazine, 32(4), 95-102, 2015.

[2] B. Settles, *Active learning literature survey*, 2009.

[3] D. Michaud. *Indexation bio-inspirée pour la recherche d'images par similarité* Thèse de doctorat, Université de Poitiers, 2018.

[4] C. Szegedy, et al. *Going deeper with convolutions.*. Proceedings of the IEEE conference on computer vision and pattern recognition, 2015.

[5] K. He, et al. *Deep residual learning for image recognition.* Proceedings of the IEEE conference on computer vision and pattern recognition, 2016.

[6] A. Krizhevsky, et al. *Image*N*et classification with deep convolutional neural networks*. Advances in neural information processing systems, 25 :1097-1105, 2012.

[7] C. Corinna, and V. Vapnik. *Support-vector networks.* Machine learning 20.3 :273-297, 1995.

[8] J. Llamas. *Datahub, Architectural HeritageElements image Dataset.*
URL : *https ://old.datahub.io/dataset/architectural-heritage-elements-image-dataset*, 2021.