

Architectures profondes pour la classification d'images

Stage Master 2 Recherche 2009-2010 - LIP6 / UPMC

1 Informations pratiques

Mots-clés : Sciences pour l'ingénieur, Apprentissage, Traitement des images, Indexation

Durée : 5 à 6 mois

Localisation : Laboratoire d'Informatique de Paris 6, 104 ave du President Kennedy, 75016 Paris

Encadrement : Nicolas Thome, Matthieu Cord

Contact : nicolas.thome@lip6.fr, Matthieu.Cord@lip6.fr

Rémunération LIP6 : 2000 Euros pour le stage

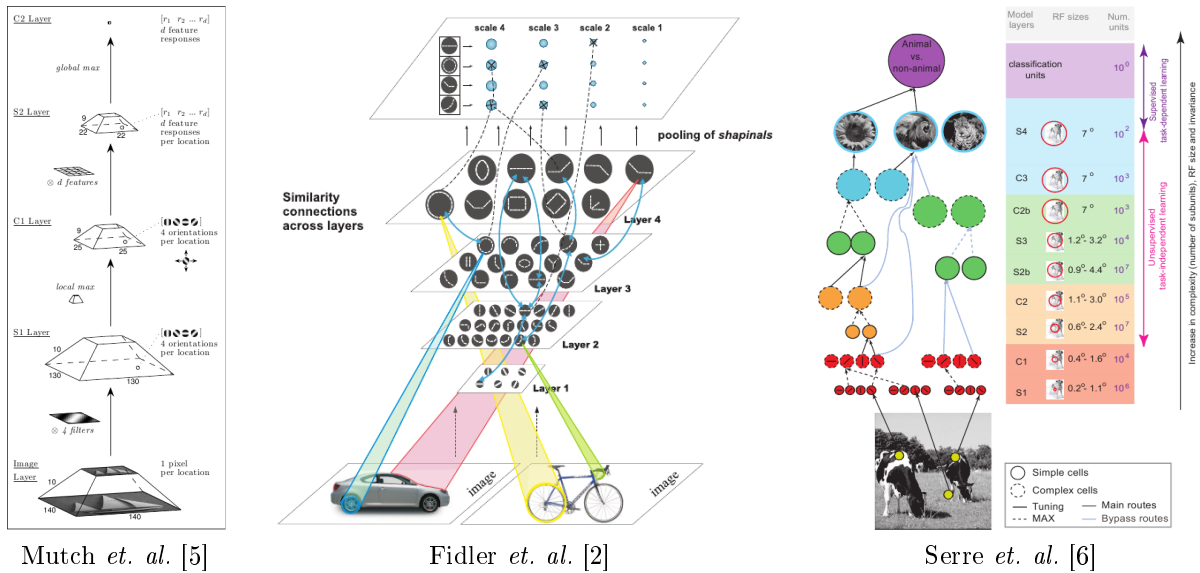


FIG. 1 – Architectures profondes pour l'apprentissage visuel de catégories sémantiques

2 Contexte

Dans la plupart des méthodes de la littérature, la recherche d'image par le contenu se décompose en deux étapes clés. Tout d'abord on cherche à construire des signatures le plus souvent vectorielles pour chaque image de la base traitée. Ensuite, ces signatures sont utilisées pour des tâches de recherche d'images par similarité, de classification en ligne ou supervisée. La première étape repose le plus souvent sur la construction d'un dictionnaire de mots visuels, la fréquence d'apparition des différents mots constituant le descripteur appris par le classifieur.

Ainsi, la majorité des approches actuelles utilise une architecture plate, avec une seule couche. Le bon niveau de performances de ces méthodes réside autant dans l'extraction de descripteurs spécifiques et discriminants pour un tâche donnée que dans le processus de classification. Quelques travaux récents ont proposé d'introduire des niveaux intermédiaires dans la représentation des données, sous forme de topics par exemple [4]. Cependant, ce type d'architecture à deux niveaux reste assez simpliste et rigide, et les performances des méthodes utilisant cet espace de description enrichi demeurent décevantes.

Dans un cadre de reconnaissance de catégories, des méthodes d'inspiration biologique ont été proposées [6, 2, 5] (figure 1). Ces travaux s'appuient sur des modèles d'architectures profondes datant des années 80, et qui ont récemment été revues par Hinton *et. al.* [3]. En particulier, les progrès réalisés en optimisation et en apprentissage statistique ont démontré qu'il était possible de réaliser un apprentissage efficace dans des architectures empilant de nombreuses couches.

3 Objectifs

Les approches d'inspiration biologique visent à établir des modèles computationnels reproduisant le fonctionnement du cortex cérébral. Elles s'appuient sur l'excellent niveau de performances des systèmes de vision des mammifères, qui surclassent de très loin les meilleurs algorithmes automatiques actuels. En particulier, les systèmes de vision biologique présentent une très grande robustesse à l'extrême variabilité des images réelles (point de vue, apparence, forme, *etc*) pour la reconnaissance de catégories sémantiques.

Le stage a pour objectif d'évaluer ce type d'approches biologiquement inspirées. Le sujet s'inscrit dans le cadre du projet ANR ASAP (2010-2013)¹, dans lequel le LIP6 est impliqué. Ces travaux pourront être poursuivis dans le cadre d'une thèse.

La première étape du stage consistera à développer une méthode originale par rapport aux méthodes de l'état de l'art (*e.g* [5, 6]), en proposant une architecture profonde adaptée à notre contexte. De ce point de vue, de nombreuses questions restent ouvertes quant à l'apprentissage global du réseau. Des choix pertinents pour l'apprentissage (non supervisé) des premières niveaux de la hiérarchie devront être effectués. Parallèlement, le candidat devra évaluer les algorithmes mis au point pour des tâches de classification de scènes et/ou de détection d'objets. On s'attachera en particulier à évaluer l'algorithme sur des bases jugées difficiles, *e.g.* PASCAL[1], ou dans le cadre de d'images urbaines (projet ITOWNS). Enfin, en fonction de l'avancement et des résultats obtenus, le candidat abordera la question de la mise en place de boucles de rétroaction (feedback) entre les différents niveaux de l'architecture profonde. En effet, les modèles actuels n'intègrent que des stratégies où l'apprentissage est effectué de manière ascendante ("Bottom-Up"), alors que les systèmes biologiques sont connus pour être très fortement bouclés.

4 Compétences requises

Des connaissances minimales en traitement d'images et en apprentissage statistique sont nécessaires. Un bon niveau de programmation est demandé (le développement se fera en Java). Le sujet requiert à la fois un intérêt pour la manipulation de modèles théoriques complexes, ainsi que leur application à des données réelles.

Références

- [1] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The PASCAL Visual Object Classes Challenge 2009 (VOC2009) Results. <http://www.pascal-network.org/challenges/VOC/voc2009/workshop/index.html>.
- [2] Sanja Fidler, Marko Boben, and Ales Leonardis. Similarity-based cross-layered hierarchical representation for object categorization. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 0 :1–8, 2008.
- [3] Geoffrey E. Hinton, Simon Osindero, and Yee-Whye Teh. A fast learning algorithm for deep belief nets. *Neural Comp.*, 18(7) :1527–1554, July 2006.
- [4] Diane Larlus. *Création et utilisation de vocabulaires visuels pour la catégorisation d images et la segmentation de classes d objets*. PhD thesis, INPG, nov 2008.
- [5] Jim Mutch and David Lowe. Object class recognition and localization using sparse features with limited receptive fields. *International Journal of Computer Vision*, 80(1) :45–57, October 2008.
- [6] T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber, and T. Poggio. Robust object recognition with cortex-like mechanisms. *IEEE Trans Pattern Anal Mach Intell*, 29(3) :411–426, 2007.

¹Apprentissage Statistique par une Architecture Profonde