



«Apprentissage et exploitation de représentations sémantiques pour la classification et la recherche d'images»

Soutenance de thèse – Maxime BUCHER

27 novembre 2018

Palaiseau

Devant le jury composé de :

Mme Céline HUDELOT

M. Nicolas THOME

M. Frédéric PRECIOSO

M. Stéphane HERBIN

M. Frédéric JURIE

Rapporteur

Rapporteur

Examineur

Encadrant

Directeur de thèse

École Centrale de Paris

CNAM

Université Nice Sophia Antipolis

ONERA

Université Caen

Résumé :

Dans cette thèse nous étudions différentes questions relatives à la mise en pratique de modèles d'apprentissage profond. En effet malgré les avancées prometteuses de ces algorithmes en vision par ordinateur, leur emploi dans certains cas d'usage réels reste difficile. Une première difficulté est, pour des tâches de classification d'images, de rassembler pour des milliers de catégories suffisamment de données d'entraînement. C'est pourquoi nous proposons deux nouvelles approches adaptées à ce scénario d'apprentissage, appelé «classification zero-shot». L'utilisation d'information sémantique pour modéliser les classes permet de définir les modèles par description, par opposition à une modélisation à partir d'un ensemble d'exemples, et rend possible la modélisation sans donnée de référence. L'idée fondamentale du premier chapitre est d'obtenir une distribution d'attributs optimale grâce à l'apprentissage d'une métrique, capable à la fois de sélectionner et de transformer la distribution des données originales. Dans le chapitre suivant, contrairement aux approches standards de la littérature qui reposent sur l'apprentissage d'un espace d'intégration commun, nous proposons de générer des caractéristiques visuelles à partir d'un générateur conditionnel. Une fois générés ces exemples artificiels peuvent être utilisés conjointement avec des données réelles pour l'apprentissage d'un classifieur discriminant. Dans une seconde partie de ce manuscrit, nous abordons la question de l'intelligibilité des calculs pour les tâches de vision par ordinateur. En raison des nombreuses et complexes transformations des algorithmes profonds, il est difficile pour un utilisateur d'interpréter le résultat retourné. Notre proposition est d'introduire un «goulot d'étranglement sémantique» dans le processus de traitement. La représentation de l'image est exprimée entièrement en langage naturel, tout en conservant l'efficacité des représentations numériques. L'intelligibilité de la représentation permet à un utilisateur d'examiner sur quelle base l'inférence a été réalisée et ainsi d'accepter ou de rejeter la décision suivant sa connaissance et son expérience humaine.

Mots clés

Classification zero-shot, attributs, plongement sémantique, apprentissage de métrique, goulot d'étranglement sémantique, recherche d'images