

Détection d'arbres urbains à partir de données aériennes multi-sources (optique, infrarouge, MNS)

L. Pibre, M. Chaumont, G. Subsol, D. Ienco, and M. Derras.

TEMU'2017, Atelier Télédétection pour l'Etude des Milieux Urbains, Centre International de Conférences de Météo-France, Toulouse, France, 16-17 janvier, 2017. Présentation orale.

Introduction

En télédétection, beaucoup de méthodes de détection d'objets utilisent l'apprentissage automatique et combine l'extraction de descripteurs comme les HOG et un classifieur performant comme le SVM. Ces dernières années, sont apparus les Réseaux de Neurones Convolutionnels (CNNs) qui intègrent dans un même schéma d'optimisation ces deux étapes.

Dans le cas de données multi-sources (optique, proche infrarouge, Modèle Numérique de Surface), il n'est pas facile de combiner toutes ces informations car elles donnent des mesures qui peuvent être dans des dimensions ou des échelles très différentes. Il est donc nécessaire de normaliser et la méthode peut avoir une grande influence sur les résultats. Or les CNNs peuvent gérer ces différences et intégrer ces types de données puisqu'il va normaliser les valeurs d'entrée en même temps qu'il va créer son modèle d'apprentissage.

Dans cet article, nous souhaitons évaluer la pertinence d'utiliser et quantifier les performances des CNNs par rapport aux méthodes utilisant un descripteur avec un classifieur. Pour cela nous allons comparer deux Réseaux de Neurones Convolutionnels, AlexNet et GoogleNet, et deux méthodes d'apprentissage automatique, un SVM et un Random Forest avec un descripteur HOG. Nous prendrons comme exemple d'application la localisation des arbres urbains.

Méthode

La méthode que nous avons utilisée est la suivante: tout d'abord, on entraîne un CNN à discriminer la classe "Arbre" de la classe "Autre". La base d'entraînement est composée d'images ayant toutes la même taille. Pour la phase de test, une fenêtre glissante est appliquée sur les nouvelles images. Les images extraites par la fenêtre glissante sont envoyées au CNN afin d'obtenir leur probabilité d'appartenir à la classe "Arbre" ou "Autre". Puisque la fenêtre glissante est appliquée à plusieurs échelles, plusieurs prédictions sur une même zone vont être produites. Nous avons alors à fusionner toutes ces sorties afin d'obtenir une boîte englobante finale et précise des arbres dans les images. Dans nos expériences, nous avons appliqué deux méthodes de fusion, une méthode de fusion basée sur l'aire et une basée sur le chevauchement.

Résultats

Les résultats que nous avons obtenus avec les CNNs (AlexNet et GoogleNet) sont toujours supérieurs aux résultats obtenus avec les deux méthodes de télédétection utilisées (descripteur hog avec un SVM et descripteur hog avec un random forest). Les performances obtenus avec les deux méthodes d'apprentissage automatique sont extrêmement basses. Cela peut être dû au fait que les arbres sont souvent très proches les uns des autres ce qui les rend difficiles à différencier sur la base de leurs contours qui sont accentués par le descripteur HOG.