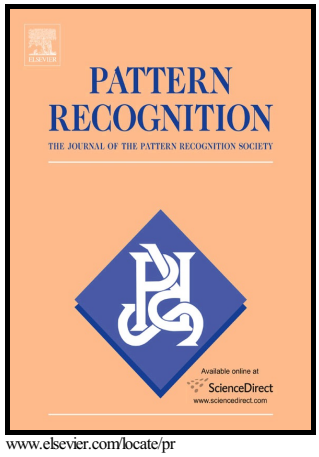
Towards Better Exploiting Convolutional Neural Networks for Remote Sensing Scene Classification

Keiller Nogueira, Otávio A.B. Penatti, Jefersson A. dos Santos



www.elsevier.com/locate/pr

PII: S0031-3203(16)30150-9
DOI: http://dx.doi.org/10.1016/j.patcog.2016.07.001
Reference: PR5787

To appear in: *Pattern Recognition*

Received date: 1 February 2016
Revised date: 1 June 2016
Accepted date: 1 July 2016

Cite this article as: Keiller Nogueira, Otávio A.B. Penatti and Jefersson A. dos Santos, Towards Better Exploiting Convolutional Neural Networks for Remote Sensing Scene Classification, *Pattern Recognition* http://dx.doi.org/10.1016/j.patcog.2016.07.001

Towards Better Exploiting Convolutional Neural Networks for Remote Sensing Scene Classification

Keiller Nogueira[a,*], Otávio A. B. Penatti[b], Jefersson A. dos Santos[a]

[a]*Departamento de Ciência da Computação, Universidade Federal de Minas Gerais (UFMG), Av. Presidente Antônio Carlos, 6627, Belo Horizonte, MG, CEP 31270-901, Brazil*
[b]*Advanced Technologies Group, SAMSUNG Research Institute, Av. Cambacica, 1200, Building 1, Campinas, SP, CEP 13097-160, Brazil*

Abstract

We present an analysis of three possible strategies for exploiting the power of existing convolutional neural networks (ConvNets or CNNs) in different scenarios from the ones they were trained: full training, fine tuning, and using ConvNets as feature extractors. In many applications, especially including remote sensing, it is not feasible to fully design and train a new ConvNet, as this usually requires a considerable amount of labeled data and demands high computational costs. Therefore, it is important to understand how to better use existing ConvNets. We perform experiments with six popular ConvNets using three remote sensing datasets. We also compare ConvNets in each strategy with existing descriptors and with state-of-the-art baselines. Results point that fine tuning tends to be the best performing strategy. In fact, using the features from the fine-tuned ConvNet with linear SVM obtains the best results. We also achieved state-of-the-art results for the three datasets used.

Keywords: Deep Learning, Convolutional Neural Networks, Fine-tune, Feature Extraction, Aerial Scenes, Hyperspectral Images, Remote Sensing

1. Introduction

Encoding discriminating features from visual data is one of the most important steps in almost any computer vision problem, including in the remote sensing domain. Since manual extraction of these features is not practical in most cases, during years, substantial efforts have been dedicated to develop automatic and discriminating visual feature descriptors [1]. In the early years, most of such descriptors were based on pre-defined algorithms independently of the underlying problem, like color histograms and correlograms [1, 2]. Then, descriptors based on visual dictionaries, the so-called Bag of Visual Words (BoVW), attracted the attention and have become the state-of-the-art for many years in computer vision [3, 4, 5, 6, 7, 8, 9]. Although the aforementioned visual feature extraction techniques have been successfully applied in several domains [10], due to the specificities of remotely sensed data, many of these

*Corresponding author at +55-31-3409-5860 (phone/fax), Departamento de Ciência da Computação, Universidade Federal de Minas Gerais (UFMG), Av. Presidente Antônio Carlos, 6627, Belo Horizonte, MG, CEP 31270-901, Brazil

Email addresses: keiller.nogueira@dcc.ufmg.br (Keiller Nogueira), o.penatti@samsung.com (Otávio A. B. Penatti), jefersson@dcc.ufmg.br (Jefersson A. dos Santos)