Multi-Neighborhood Convolutional Networks

Elnaz Barshan University of Waterloo, ON, Canada
Paul Fieguth University of Waterloo, ON, Canada
Alexander Wong University of Waterloo, ON, Canada

Abstract

We explore the role of scale for improved feature learning in convolutional networks. We propose multi-neighborhood convolutional networks, designed to learn image features at different levels of detail. Utilizing nonlinear scale-space models, the proposed multi-neighborhood model can effectively capture fine-scale image characteristics (i.e., appearance) using a small-size neighborhood, while coarse-scale image structures (i.e., shape) are detected through a larger neighborhood. The experimental results demonstrate the superior performance of the proposed multi-scale multi-neighborhood models over their single-scale counterparts.

1 Introduction

In recent years the computer vision community has witnessed the success of "learned" image descriptors against engineered handcrafted features for image classification. These feature learning models are generally multi-layer architectures with different weight-sharing schemes between layers, where the Convolutional Neural Network (CNN) [2] is the most widely-used model of this type.

The conventional convolutional networks used for image classification are *single-neighborhood*, in the sense that for each pixel only a single-size neighborhood is considered, and *single-scale*, in the sense that only the input image at a single level of detail is used for extracting feature.

In this paper we present an extension of deep convolutional models which can more explicitly capture features at different levels of details through the use of *non-linear* scale-space models [1]. Our proposed multi-neighborhood model captures fine-scale image characteristics through a small-size neighborhood and coarse-scale image characteristics by exploring a wider range of dependencies over a larger neighborhood. Finally, we will demonstrate how one can train these models to get a higher accuracy without increasing training cost.

2 Multi-Neighborhood Architectures

In order to capture a richer set of features using a multi-scale convolutional architecture, it will be essential to have neighborhoods of different sizes, consequently we need a multi-neighborhood architecture with small/large neighborhoods at fine/coarse scales, respectively.

Based on nonlinear scale-space models [1], our purpose is to design a multi-scale multi-neighborhood convolutional architecture with feature detectors sensitive over a range of scales. Our baseline solution, the *Uniform Multi-Scale* architecture, is a set of parallel feature extractors each with the same architecture applied on images of the same size but at different levels of details. The final vector is obtained by concatenating the extracted features.

The second proposal, *Multi-neighborhood in Input* architecture, is based on down-sample the representation of the input image at coarser scales by a larger factor. The final strategy is to introduce aspects of scale in pooling. In the *Multi-neighborhood in Pooling* architecture, all of the input representations are of the same size, but the difference appears in the pooling ratio; larger pooling size for coarser scales.

The proposed multi-scale architectures are composed of a set of parallel feature extractors applied to the representations of the input image at an ensemble of scales. Our strategy to train a set of parallel single-scale networks is to train the network for scale 0 (i.e., the original input representation) and then to reuse its learned parameters for all of the single-scale networks. Therefore, the computational complexity of a multi-scale architecture is unchanged from single-scale networks.

CNN Architecture	No. of Scales	Accuracy %
Single Scale: scale 0	1	74.78
Uniform Multi-scale	5	74.98
Multi-neigh. Input	5	76.15
Multi-neigh. Pooling	5	75.77

Table 1: Performance for supervised feature learning.

3 Experiments

Given the scalable learning algorithm, we evaluate the effectiveness of the three proposed multi-scale architectures for image classification, compared to each other and to an equivalent single-scale architecture. For the experiments the standard CIFAR10 data set is used.

First we focus on unsupervised feature learning. Figure 1 shows the recognition performance of different architectures on the CIFAR10 data set as a function of the number of learned filters. Observe that, with the same training cost, all of the three multi-scale architectures outperform the single-scale counterpart, clearly demonstrating the advantage of extracting scale-dependent or detail-dependent image representations. Among the multi-scale models, both of the multi-neighborhood architectures (i.e., in input and in pooling) offer a better performance than the uniform model, especially for larger numbers of filters.

Also, we examine the proposed multi-scale architectures for supervised feature learning. In this case, at each scale a CNN with two layers of convolution is used. Table 1 illustrates the positive effect of extracting features at multiple scales. Among the proposed multi-scale models, the multi-neighborhood architectures generate more discriminative features due to exploring a wider range of dependencies over a larger neighborhood. Also, for multi-scale models observe how the scalable learning strategy substantially reduces the training cost (i.e., the number of learned filters) while the recognition rate is still superior to that of their single-scale counterpart. Note that these results are obtained using a simple base-line two-layer network and can be applied to the state of the art models without loss of generality.

References

[1] Wong, A. and Mishra, A. and Yates, J. and Fieguth, P. and Clausi, D. and Callaghan, J., Intervertebral disc segmentation and volumetric reconstruction from peripheral quantitative computed tomography imaging. *Biomedical Engineering, IEEE Transactions on* (2009).

[2] LeCun, Y. and Bottou, L. and Bengio, Y. and Haffner, P., Gradient-based learning applied to document recognition. *Proceedings of the IEEE* (1998).

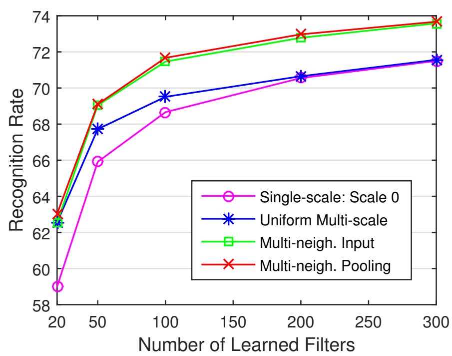

Fig. 1: Performance for unsupervised feature learning.