

Visualizing and Understanding Convolutional Networks

January 21st, 2019

Clémence Réda Xiaoqi Xu
École Normale Supérieure Paris-Saclay
61, avenue du Président Wilson, 94230 Cachan
{creda, xiaoqi.xu}@ens-paris-saclay.fr

Abstract

We want to design our own CNN for image classification, and to be able to justify its architecture and its filters by visualizing the feature maps and filters.

1. Introduction

The Convolutional Neural Networks (CNNs) have shown impressive capability in fields like computer vision and natural language processing. They are also widely used in a lot of domains to upgrade the state-of-the-art results of artificial intelligence. However, despite the ubiquitous application of CNNs, they are poorly understood in theory.

Recent years, some researchers have made the first attempts to decode CNNs by developing various visualizing techniques. Erhan et al.[1] have proposed the method of activation maximization, i.e. finding the patterns that maximize the activation of a given unit. This gives information about just one unit, but does not explain the unit's invariances. Simonyan Girshick

2. Method

2.1. Model

Deconvnet

Our model The architecture is standard: each convolutional layer consists of convolution, rectified linear function and max pooling; the top layer is fully connected and the final layer is a softmax classifier.



Figure 1. Example of caption. It is set in Roman so that mathematics (always set in Roman: $B \sin A = A \sin B$) may be included without an ugly clash.

2.2.

2.3. Visualizing activation

3. Experiments

3.1. Data

4. Results

5. Conclusion

References

- [1] A. C. Dumitru Erhan, Yoshua Bengio and P. Vincent. Visualizing higher-layer features of a deep network. Technical report, 2009.