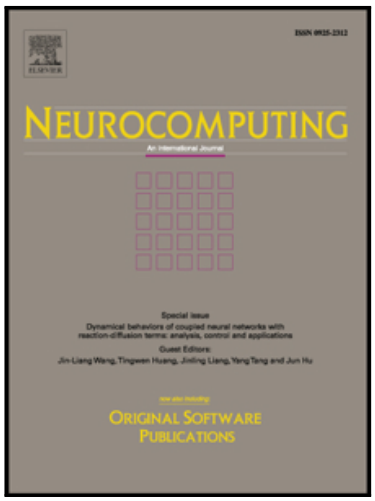


# Visualizing and Analyzing Convolution Neural Networks with Gradient Information

Ruigang Fu, Biao Li, Yinghui Gao, Ping Wang

PII: S0925-2312(18)30247-9
DOI: 10.1016/j.neucom.2018.02.080
Reference: NEUCOM 19385

To appear in: *Neurocomputing*

Received date: 8 December 2016
Revised date: 22 January 2018
Accepted date: 23 February 2018

Please cite this article as: Ruigang Fu, Biao Li, Yinghui Gao, Ping Wang, Visualizing and Analyzing Convolution Neural Networks with Gradient Information, *Neurocomputing* (2018), doi: 10.1016/j.neucom.2018.02.080

# Visualizing and Analyzing Convolution Neural Networks with Gradient Information

Ruigang Fu[a,*], Biao Li[a], Yinghui Gao[a], Ping Wang[a]

[a]*Deya load 109, Changsha, Hunan, China*

**Abstract**

Recently, convolution neural networks (CNNs) have demonstrated state-of-the-art results on image classification. However, our understanding of how they work is still limited. This paper expands the gradient-based visualization method by emphasizing the ability of CNNs in extracting local features, visualizes and analyzes the internal behavior of trained CNNs based on computing the gradient of the class score with respect to every layer of CNNs. Besides, we use a quantitative evaluation named prediction difference to verify the proposed visualization method. The saliency maps constructed from different layers demonstrate what CNNs care about in input images and it turns out that when CNNs classify an image, they identify the location of the object. Finally, we compare AlexNet with VGGnets, all of them have the same focus on the input images and VGGnets are more sensitive to pixels.

*Keywords:* CNNs, visualization, gradient, receptive field, prediction difference

## 1. Introduction

Convolution neural networks (CNNs) have improved the performance of computer vision systems greatly for a large number of high-level problems, including image classification [1, 2, 3, 4, 5], object detection [6, 7, 8, 9, 10], sentence modelling [11], etc. However, our understanding of CNNs' internal behavior has

*Corresponding author
*Email address:* furuigang08@nudt.edu.cn (Ruigang Fu)