Accepted Manuscript

Understanding Convolutional Neural Networks with A Mathematical Model

C.-C. Jay Kuo

PII: S1047-3203(16)30226-7
DOI: http://dx.doi.org/10.1016/j.jvcir.2016.11.003
Reference: YJVCI 1888

To appear in: *J. Vis. Commun. Image R.*

Received Date: 15 September 2016
Accepted Date: 4 November 2016

Please cite this article as: C.-C. Jay Kuo, Understanding Convolutional Neural Networks with A Mathematical Model, *J. Vis. Commun. Image R.* (2016), doi: http://dx.doi.org/10.1016/j.jvcir.2016.11.003

Understanding Convolutional Neural Networks with A Mathematical Model

C.-C. Jay Kuo

Ming-Hsieh Department of Electrical Engineering
University of Southern California, Los Angeles, CA 90089-2564, USA

Abstract

This work attempts to address two fundamental questions about the structure of the convolutional neural networks (CNN): 1) why a nonlinear activation function is essential at the filter output of all intermediate layers? 2) what is the advantage of the two-layer cascade system over the one-layer system? A mathematical model called the "REctified-COrrelations on a Sphere" (RECOS) is proposed to answer these two questions. After the CNN training process, the converged filter weights define a set of anchor vectors in the RECOS model. Anchor vectors represent the frequently occurring patterns (or the spectral components). The necessity of rectification is explained using the RECOS model. Then, the behavior of a two-layer RECOS system is analyzed and compared with its one-layer counterpart. The LeNet-5 and the MNIST dataset are used to illustrate discussion points. Finally, the RECOS model is generalized to a multilayer system with the AlexNet as an example.

Keywords: Convolutional Neural Network (CNN), Nonlinear Activation, RECOS Model, Rectified Linear Unit (ReLU), MNIST Dataset.

1. Introduction

There is a strong resurging interest in the neural-network-based learning because of its superior performance in many speech and image/video understanding applications nowadays. The recent success of deep neural networks (DNN) [1] is due to the availability of a large amount labeled training data (e.g. the ImageNet) and more efficient computing hardware. It is called deep learning since we often observe performance improvement when adding more