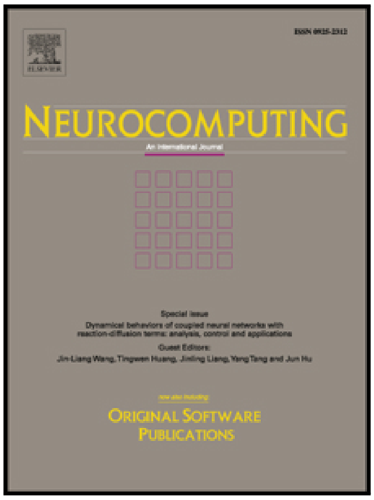
Communicated by Jun Yu

Accepted Manuscript



A Joint Residual Network with Paired ReLUs activation for Image Super-Resolution

Zhimin Tang, Linkai Luo, Hong Peng, Shaohui Li

PII:	S0925-2312(17)31384-X
DOI:	10.1016/j.neucom.2017.07.061
Reference:	NEUCOM 18774

To appear in:	*Neurocomputing*

Received date:	9 March 2017
Revised date:	29 July 2017
Accepted date:	31 July 2017

Please cite this article as: Zhimin Tang, Linkai Luo, Hong Peng, Shaohui Li, A Joint Residual Network with Paired ReLUs activation for Image Super-Resolution, *Neurocomputing* (2017), doi: 10.1016/j.neucom.2017.07.061

A Joint Residual Network with Paired ReLUs activation for Image Super-Resolution

Zhimin Tang, Linkai Luo*, Hong Peng, Shaohui Li

Department of Automation, Xiamen University, Xiamen 361005, China

Abstract

Recently, single image super-resolution (SR) models based on deep convolutional neural network (CNN) have achieved significant advances in accuracy and speed. However, these models are not efficient enough for the image SR task. Firstly, we find that generic deep CNNs learn the low frequency features in all layers, which is redundant and leads to a slow training speed. Secondly, rectified linear unit (ReLU) only activate the positive response of the neuron, while the negative response is also worth being activated. In this paper, we propose a novel joint residual network (JRN) with three subnetworks, in which two shallow subnetworks aim to learn the low frequency information and one deep subnetwork aims to learn the high frequency information. In order to activate the negative part of the neurons and to preserve the sparsity of activation function, we propose a paired ReLUs activation scheme: one of the ReLUs is for positive activation and the other is for negative activation. The above two innovations lead to a much faster training, as well as a more efficient local structure. The proposed JRN achieves the same accuracy of a generic CNN with only 10.5% training iterations. The experiments on a wide range of images show that JRN is superior to the state-of-the-art methods both in accuracy and computational efficiency.

Keywords:
Image super-resolution, Deep learning, Image restoration, Convolutional neural network

1. Introduction

Single image super-resolution (SR) aims to recover a high-resolution (HR) image from a given low-resolution (LR) image [40]. It is widely used in many fields such as surveillance, medical imaging, satellite imaging and face recognition. SR is typically a highly ill-posed problem since there are a lot of solutions. In general, strong prior information is needed in ill-posed problem.

Recently, learning-based methods have achieved great success in SR task. The methods based on neighbor embedding [2, 4] make use of the local linear embedding to generate HR patches under the assumption that the LR and its HR patches lie on low-dimensional nonlinear manifolds with similar local geometry. Sparse-coding-based methods [41, 42, 49] jointly train two dictionaries for the LR-HR patch pairs through enforcing the similarity of sparse representations. Therefore, the LR patch and the corresponding HR patch can share the same sparse representation over their own dictionaries. Based on the learned iterative shrinkage and thresholding algorithm [10], researchers [39, 25] extend the conventional sparse coding model [42] to the sparse-coding-based network, hence sparse prior can be encoded in the network. By combing sparse coding and neighbor embedding, Timofte *et al.* proposed a anchored neighborhood regression (ANR) method [34], in which the learned dictionary atoms are used as anchor points and then the regressors for each anchor points are learned. Instead of learning the regressors on the dictionary, ANR's improved version A+ [35] reaches the state-of-the-art quality by using the full training data.

More recently, neural networks [32, 26, 29, 21] have been widely used in many applications, such as image recognition [20], speech recognition, natural language processing [21], human pose recovery [14], image privacy protection [45], big multimedia analysis [43], image ranking [44] and image super-resolution [6]. Especially, convolutional neural networks (CNN) [20, 21] have achieved great success in machine learning and computer vision. Dong *et al.* [6, 7] proposed a super-resolution convolutional neural network (SRCNN) based on a fully convolutional neural network . SRCNN and other methods based on CNN [17, 18, 8, 16, 31, 38] have shown impressive performance.

Although the image super-resolution methods based on CNN obtain great success, we find some limitations.

Firstly, these CNNs are designed with generic deep architectures. Generic deep CNNs are designed for image recognition task and some other high-level vision problems, not specialized for the image SR problem. In the task of SR, the low frequency information needs to be propagated from the input layer to the output layer, each middle layer must learn the low frequency features. We find that generic deep CNNs learn the low frequency features in all layers. We think that it is redundant and leads to a slow training speed.

Similar to Taylor expansion, we can decompose an image into different components in which some are low frequency components and some are high-frequent components. The low

*Corresponding author
Email address: `luolk@xmu.edu.cn` (Linkai Luo)