Accepted Manuscript

Image super-resolution via a novel cascaded convolutional neural network framework

Fu zhang, Nian Cai, Guandong Cen, Feiyang Li, Han Wang, Xindu Chen

PII: S0923-5965(18)30076-6
DOI: https://doi.org/10.1016/j.image.2018.01.009
Reference: IMAGE 15324

To appear in: *Signal Processing: Image Communication*

Received date : 30 July 2017
Revised date : 18 December 2017
Accepted date : 27 January 2018

Image Super-Resolution Via a Novel Cascaded Convolutional Neural Network Framework

Fu zhang[a], Nian Cai[a]*, Guandong Cen[a], Feiyang Li[a], Han Wang[b], Xindu Chen[b]

[a] *School of Information Engineering, Guangdong University of Technology, Guangzhou 510006, China*
[b] *School of Electromechanical Engineering, Guangdong University of Technology, Guangzhou 510006, China*

A R T I C L E I N F O

Keywords:
Image super-resolution
Cascaded convolution neural network
Multi-scale feature mapping
Residual learning
Gradient clipping

A B S T R A C T

Due to excellent self-learning ability, deep convolutional neural networks (CNNs) are successfully employed in the field of single image super-resolution (SR) compared with interpolation methods and sparse coding methods. To implement the multi-scale image SR task with a single trained model and to further improve the performance of image SR, we propose a novel cascaded CNN framework with three stages, which are feature extraction, detail prediction and reconstruction. At the stage of feature extraction, we propose a scheme of multi-scale feature mapping to extract the inherent features via the low-resolution image. At the stage of detail prediction, the lost details of the original low-resolution image are predicted via the network cascade. Finally, a high-resolution image is achieved by the scheme of residual learning, which is implemented by superimposing the lost details and the original low-resolution image. To avoid gradient explosions, we use gradient clipping to train the proposed cascaded CNN framework. Comparison results indicate that our proposed cascaded CNN framework for image SR is superior to many state-of-the-art methods.

1. Introduction

Single image super-resolution (SR) is a classical problem in computer vision and widely used in many applications such as video surveillance, satellite imaging and medical imaging [1]. Many researchers and engineers devote themselves to image SR. Simple understanding of image SR, it aims to overcome resolution limitation of the sensors by restoring high resolution (HR) images from low resolution (LR) images with complementary information.

Many SR methods have been studied in the computer vision community. Early methods include interpolation such as bicubic interpolation, Lanczos resampling [2], some methods utilizing statistical image priors [3, 4] or internal patch recurrence [5], and neighbor embedding [6, 7]. Recently, learning-based methods attract more attentions in the field of image SR. Sparse coding methods [8-12] are widely studied among these learning-based methods, which use learned compact dictionaries based on sparse signal representation. Compare with the early methods, learning-based methods perform excellently in image SR.

Especially, more and more researchers pay attention to deep CNNs for image SR [13-17, 36, 39] due to the excellent learning ability. Dong et al. first employed a CNN to learn a mapping from LR to HR in an end-to-end manner [14]. Their method, termed SRCNN, does not require any engineered features that are typically necessary in sparse coding methods. Experimental results indicated that their method demonstrated excellent performance in image SR compared with the previous methods. However, some limitations exist in SRCNN: 1) training converges too slowly; 2) the network only works for a single scale; 3) it requires a large number of training data. Lately, some deep CNNs are proposed to overcome some of the shortcomings of SRCNN and show more outstanding performance [13, 15-17]. In order to improve the performance of the SRCNN in computation time and image SR quality, Dong et al. reduced the parameters of the network and increased the depth of the network [13]. They called the new network as FSRCNN. Shi et al. proposed a new method named ESPCN for image and video frame SR [17], which is on the basis of SRCNN. ESPCN introduces an efficient sub-pixel convolution layer that learns an array of upscaling filters to upscale the final LR feature maps into the HR output. Guo et al. took part in the Ntire 2017 challenge on single image super-resolution [40] and designed a deep CNN to predict the "missing details" of wavelet coefficients of the low-resolution images for image SR, called Deep Wavelet Super-Resolution (DWSR) [39]. DWSR is trained in the wavelet domain, whose inputs and outputs are 4 sub-bands of the low-resolution wavelet coefficients and residuals (missing details) of 4 sub-bands of high-resolution wavelet coefficients, respectively. Although those deep CNNs achieved fairly good performance in image SR, these deep CNNs are actually single scale SR systems. That is to say, they can only achieve single scale SR image when the systems are trained. If a new scale SR image is on demand, the systems should be retrained according to the new scale. To cope with multi-scale image SR, Wang et al. [15] extended the sparse coding model by using several key ideas from deep learning. Then, they made a thorough research [16] based on their previous work in [15]. Their methods indicated that domain expertise is complementary to large

* *Corresponding author.*
E-mail address: cainian@gdut.edu.cn

This work was supported in part by the National Natural Science Foundation of China (Grant Nos. 61001179 and 51305084), Guangdong Natural Science Foundation (Grant No. 2015A030312008), Guangdong Science and Technology Plan (Grant Nos. 2015B010104006, 2015B010102014, 2015B090903017 and 2015B010124001), and Guangzhou Science and Technology Plan (Grant No. 201604016064).