Neurocomputing 205 (2016) 287–300

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

A versatile sparse representation based post-processing method for improving image super-resolution☆

Jun Yang, Jun Guo, Hongyang Chao*

School of Data and Computer Science, Sun Yat-sen University, Guangzhou 510006, China

ARTICLE INFO

Article history:
Received 23 November 2015
Received in revised form
28 February 2016
Accepted 13 April 2016
Communicated by Liang Lin
Available online 11 May 2016

Keywords:
Image super-resolution
Sparse representation
Principal component analysis
Post-processing
Iterative fine-tuning and approximation (IFA)

ABSTRACT

The objective of this work is single image super-resolution (SR), in which the input is specified by a low-resolution image and a consistent higher-resolution image should be returned. We propose a novel post-processing procedure named iterative fine-tuning and approximation (IFA) for mainstream SR methods. Internal image statistics are complemented by iteratively fine-tuning and performing linear subspace approximation on the outputs of existing external SR methods, helping to better reconstruct missing details and reduce unwanted artifacts. The primary concept of our method is that it first explores and enhances internal image information by grouping similar image patches and then finds their sparse or low-rank representations by iteratively learning the bases or primary components, thereby enhancing the primary structures and some details of the image. We evaluate the proposed IFA procedure over two standard benchmark datasets and demonstrate that IFA can yield substantial improvements for most existing methods via tweaking their outputs, achieving state-of-the-art performance.

1. Introduction

We consider a classical ill-posed problem of single image super-resolution (SR), in which the goal is to reconstruct a high-resolution image from a low-resolution version. The technique for this problem is expected to overcome certain inherent resolution limitations of low-cost imaging sensors (e.g., smart phones or surveillance cameras) and allow for better utilization of the increasing capacity of high-resolution displays (e.g., high-definition LEDs). Such resolution-enhancing technology may also prove to be essential in medical imaging, face recognition [1], image classification [2] and satellite imaging [3–5], where diagnosis or analysis from low-quality images can be extremely difficult.

Mainstream state-of-the-art methods for SR attempt to learn mapping functions from low- and high-resolution external exemplar pairs [6–14]. Such methods collect abundant low-/high-resolution patch pairs from multiple training images and then learn a compact dictionary, a manifold space or other complex models to relate these pairs using low-resolution patches as source data and the corresponding high-resolution patches as targets. Although they can exploit external image information across the entire dataset, external methods are challenged by the contradiction between "correctness" (the down-sampled version of the restored image is consistent with the given low-resolution image) and "comfortability" (reconstructing missing details and avoiding the introduction of artifacts, such as jaggies). More precisely, due to the ill-posedness of image super-resolution, it is inherently difficult to independently select a good trade-off between correctness and comfortability for each input using available information, typically resulting in high-resolution images that are overly "correct" (coarse) or "comfortable" (blurry). For example, there are coarse or fake textures in the smooth regions in the zoomed green box in Fig. 2(b), whereas in Fig. 1(b), the hairs are unclear and blurry. More image information is required to obtain better results.

In recent years, the sparse learning methods [15,16] and AndOr graph model [17,18] show the power in terms of the descriptive and discriminative abilities, which are efficiently used in image restoration [19], classification [20] and recognition [21,22]. In this paper, we also introduce a sparse representation based post-processing procedure named iterative fine-tuning and approximation (IFA) to explore the internal image statistics to provide more prior information. Rather than competing with existing external methods, our work is a complement to almost all of them. Based on the outputs of these methods, we run an iterative procedure for improvement. Each iteration consists of two in-place steps. In the first step, we attempt to fine-tune an output (a high-resolution image) such that its down-sample would produce the closest original low-resolution image. This step guarantees the "correctness". Then, we reduce the artifacts by approximating its

☆This work was partially supported by the NSF of China under Grant 61173081 and the Guangdong Natural Science Foundation, China, under Grant S2011020001215.

* Corresponding author. Tel./fax: +86 20 39943152.
E-mail address: isschhy@mail.sysu.edu.cn (H. Chao).

http://dx.doi.org/10.1016/j.neucom.2016.04.038
0925-2312/

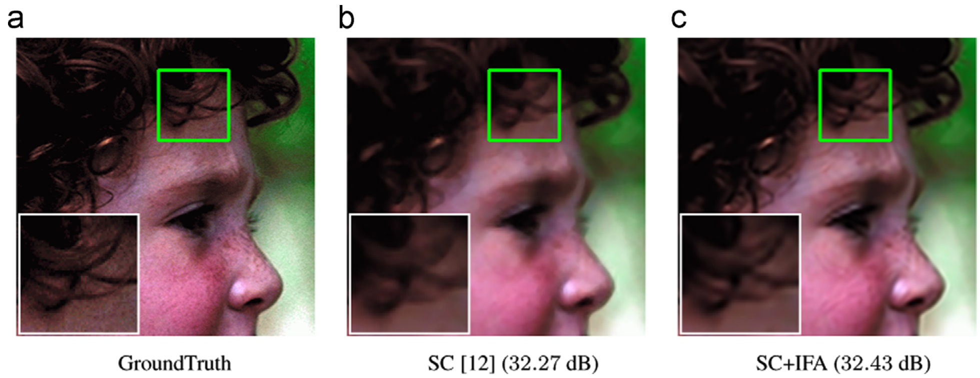

Fig. 1. “Head” from Set5 with an up-scaling factor of 4. The reconstructed images are followed with the corresponding PSNR values in the brackets.

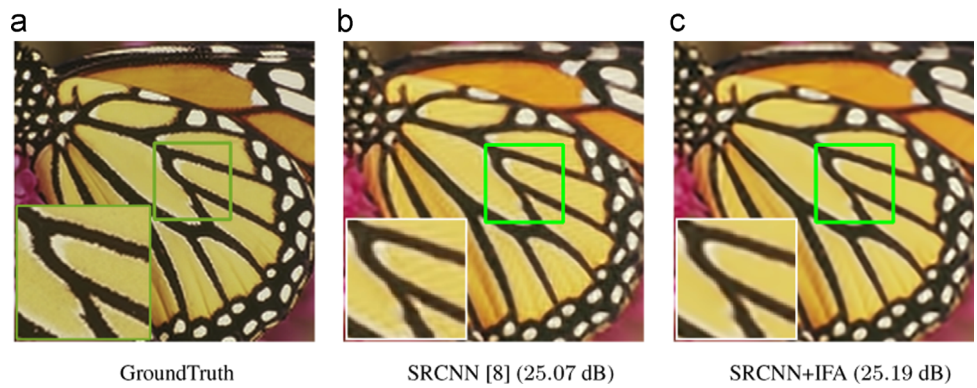

Fig. 2. “Butterfly” from Set5 with an up-scaling factor of 4. The reconstructed images are followed with the corresponding PSNR values in the brackets.

patch space through a union of lower-dimensional linear subspaces. This step guarantees the “comfortability”. In this step, our method first explores and enhances the internal image information by grouping similar image patches and then finds their sparse or low-rank representations by iteratively learning the bases or primary components, which enhances the primary structures and some details of the image. Iteratively performing these two steps in-place gradually improves the “correctness” and “comfortability”. Figs. 1 and 2 present two examples to illustrate the appealing properties of our work. In Fig. 1, after tweaking the output of the SC [12] by IFA, the hairs become clearer, yielding a more “correct” image, whereas in Fig. 2, fake textures introduced by SRCNN [8] are removed with the help of IFA, making the result more “comfortable”.

The main contributions of this paper are as follows. First, an effective projection approach named Approximation (A-step) was proposed to explore the internal image statistics by sparsely representing each group of similar patches. The internal statistics come from two aspects: one is that the grouped patches with similar contents and structures from distant regions of the image, which enhances the main structures of the patch or pixels considered, and the other is that all of the reconstructed values of pixels are the means of several approximately restored values, obeying a Gaussian distribution. Second, a joint projection approach, the A-step combined with the fine-tuning step (F-step), was established to reduce the artifacts, enhance the details of the image and improve the correctness. Finally, a versatile post-processing method named IFA was proposed to improve the results of the image super-resolution.

The remainder of this paper is organized as follows. In Section 2, we review some mainstream works for SR. Our proposed IFA is introduced in Section 3. Extensive experiments in Section 4 prove that the performance of existing external methods can be substantially improved by simply “plugging in” our work.

2. Related work

Among the approaches to address the problem of image super-resolution, interpolation based approaches (i.e., bilinear or bicubic interpolation [23,24]) are the conventional and fundamental methods. However, these methods tend to generate overly smooth images with ringing and jagged artifacts, and thus, they are limited in modeling the visual complexity of real images.

The leading state-of-the-art methods for SR are mostly example-based approaches, where the core is to train a dictionary or model to map an input image from low to high resolution. Among them, external methods that train models over a large training set are the most popular. The sparse-coding-based methods are one representative type. In the pioneering work of Yang et al. [12], two patches of a low-/high-resolution pair were encoded by a low-/high-resolution dictionary, respectively, with the constraint that they shared the same sparse coefficients. After extracting patch pairs from a dataset and training the dictionaries, at testing time, low-resolution sparse coefficients were passed into the high-resolution dictionary for high-resolution reconstruction. Zeyde et al. [14] employed different training approaches for constructing the dictionary pair (the k sparse singular value decomposition (K-SVD) [25], for the low-resolution dictionary and pseudo-inverse for the high-resolution dictionary), performing dimensionality reduction on the patches through principal component analysis (PCA) and using orthogonal matching pursuit [26] for sparse coding. They showed improvements in quality with less artifacts and a higher average peak signal-to-noise ratio (PSNR) compared to the results of Yang et al. [12]. Timofte et al. [11] made some subsequent improvements by anchoring the neighborhood embedding of a low-resolution patch to the nearest atom in the dictionary and achieved improved quality and one or two orders of magnitude speed improvements. Rather than sparse coding, Dong et al. [8] trained a deep convolutional neural network as a direct