Neurocomputing 000 (2017) 1–8

Contents lists available at ScienceDirect

Neurocomputing

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

Single image super-resolution using a deep encoder–decoder symmetrical network with iterative back projection

Heng Liu[a], Jungong Han[b,*], Shudong Hou[a], Ling Shao[c], Yue Ruan[a]

[a] School of Computer Science and Technology, Anhui University of Technology, 243032, China
[b] School of Computing & Communications, Lancaster University, LA1 4YW, UK
[c] School of Computing Sciences, University of East Anglia, NR7 Tj4, UK

ARTICLE INFO

Article history:
Received 30 August 2017
Revised 29 November 2017
Accepted 4 December 2017
Available online xxx

Communicated by Prof. X. Gao

Keywords:
Single image super-resolution
Deep encoder–decoder
Symmetrical network
Iterative back projection

ABSTRACT

Image super-resolution (SR) usually refers to reconstructing a high resolution (HR) image from a low resolution (LR) image without losing high frequency details or reducing the image quality. Recently, image SR based on convolutional neural network (SRCNN) was proposed and has received much attention due to its end-to-end mapping simplicity and superior performance. This method, however, only using three convolution layers to learn the mapping from LR to HR, usually converges slowly and leads to the size of output image reducing significantly. To address these issues, in this work, we propose a novel deep encoder–decoder symmetrical neural network (DEDSN) for single image SR. This deep network is fully composed of symmetrical multiple layers of convolution and deconvolution and there is no pooling (down-sampling and up-sampling) operations in the whole network so that image details degradation occurred in traditional convolutional frameworks is prevented. Additionally, in view of the success of the iterative back projection (IBP) algorithm in image SR, we further combine DEDSN with IBP network realization in this work. The new DEDSN–IBP model introduces the down sampling version of the ground truth image and calculates the simulation error as the prior guidance. Experimental results on benchmark data sets demonstrate that the proposed DEDSN model can achieve better performance than SRCNN and the improved DEDSN–IBP outperforms the reported state-of-the-art methods.

1. Introduction

With the rapid development of high definition (HD) display device, HR image or video becomes urgently in need. Traditional optical ways to get HR images is to improve image acquisition sensors or optical systems, namely to reduce the size of the imaging unit by adopting high precision imaging chips and other related devices [1]. However, due to the great cost and the hardware limitation, in practice people usually reconstruct HR image from LR one base on image processing way. Therefore, image SR technology naturally appears and holds a huge demand.

Depending on the number of input images, image SR methods can be classified as single image based SR and multiple images based SR. It is obvious that image SR can be realized in the frequency domain or the spatial domain based on classical signal processing techniques. Generally speaking, image SR can be implemented by three means, namely, interpolation based methods, reconstruction based methods and learning based ones. A simple and fast SR method is image interpolation, which is based on the smoothness assumption, including linear, bilinear and bi-cubic interpolation ways. However, the simple smoothness assumption unavoidably results in jaggy and ringing effects due to the discontinuities across image regions. For more effective image SR, more sophisticated prior information can be learned from image correspondences, such as sparsity constraints, self-similarity prior and exemplar prior.

Actually, interpolation and reconstruction based methods are merely signal processing while the learning based ones pay more attention to the understanding of the content and the structure of the image to be reconstructed. Learning based reconstruction methods, which are quite popular nowadays, make full use of the prior related to the image itself to provide strong constraints for better SR performance. Fig. 1 summarizes the taxonomy of image SR methods [1–3].

Aiming at obtaining an HR image based on one LR input image, single image SR becomes the most attractive way among all SR solutions. The LR image is usually generated based on a degenerated model, which consists of diverse degradation operations such as geometric deformation, blur, down sampling and additive

* Corresponding author.
E-mail address: jungong.han@lancaster.ac.uk (J. Han).

https://doi.org/10.1016/j.neucom.2017.12.014
0925-2312/© 2017 Published by Elsevier B.V.

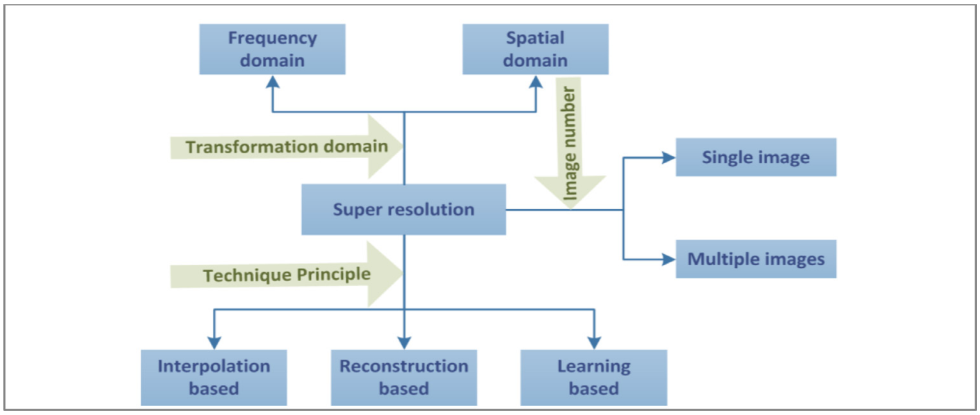

Fig. 1. The taxonomy of image SR methods.

noise. Super-resolving an LR image is an inverse process of the degradation that manages to recover the missing high-frequency details in the original HR image, such as edges and textures.

Compressive sensing theory was firstly introduced as a sparse representation method for image super-resolution in [4–6]. This approach assumes that HR patches hold the same representation coefficients with LR patches and usually requires a lot of training data to jointly learn a couple of dictionaries, which are utilized for the reconstruction of HR and LR images, respectively. In addition, the image local structures tend to recur within and across different image scales. Therefore, image SR can be realized based on self-similar examples instead of external data sets, some recent works of which can be found in [12–14]. Similar to the sparse representation based methods, exemplar images based SR ones also require large numbers of training data to learn the mapping relationship. It should be noted that for each scale down-sampling, all these image SR methods require to retrain their learning models [11,16] at corresponding scale.

Recently, deep learning, especially convolutional neural network (CNN), achieves great progress in many low-level computer vision tasks, such as image dehazing [29], image denoising [21,28] and image SR [8,18,27], etc. Benefitting from powerful end-to-end mapping learning, image SR based on convolutional neural network (SRCNN [7,8]) obtains significant improvement. Moreover, many other deep learning based image SR methods have also demonstrated their good performance in [9,17].

In this work, our main contributions are summarized as follows:

- We propose a deep encoder–decoder symmetrical network (DEDSN) for image SR. The network only consists of convolution and deconvolution layers. The convolution and deconvolution operations are purely symmetrical, which can ensure that the reconstructed image does not discard useful image details and can avoid the size reduction of the output image with respect to the input image.
- We combine DEDSN with the iterative back projection (IBP) algorithm to form a new model, named DEDSN–IBP. Each step of the IBP algorithm is converted to a specific layer in the new network, which is trained as a part of the whole network. The simulation error between the reconstructed LR image and the down-sampling of the ground truth image is iteratively utilized to update the final reconstructed HR image. The merit of this new model is the continuous prior introduction of the ground truth image in the SR procedure.

The remainder of this work is organized as follows. Section 2 provides an overview of the related work. Section 3 proposes the deep encoder–decoder symmetric network (DEDSN) in detail. Section 4 describes the improved DEDSN–IBP compositional model. Section 5 presents extensive experimental results with comparative discussions. Finally, Section 6 concludes the paper.

2. Related work

Yang et al. [4–6], propose a sparse coding based super-resolution (ScSR) approach. Given the LR/HR training patch pairs, ScSR can learn a couple of HR and LR dictionaries with the assumption that LR and HR patch pairs share the same sparse coding coefficients. Firstly, the input LR image is divided into overlapped patches and then the coefficients can be obtained through the learned LR dictionary. Then, the required HR patches can be estimated by the HR dictionary with acquired coding coefficients. Due to the division of overlapped patches, the sparse representation based SR method ignores the consistency of pixels in overlapped patches, which is a strong constraint for image reconstruction. Taking note of this problem, Gu et al. [17] raise a convolutional sparse coding (CSC) based method for image super-resolution. CSC first decomposes the input image into sparse LR feature maps by the learned LR filters to avoid dividing overlapped patches. Then the LR feature maps are mapped onto HR feature maps through the mapping function. Finally, the HR image can be estimated based on the convolution of HR feature maps with learned HR filters. In addition, by combining sparse representation with non-local similarity, Li et al. [33] present a self-learning image SR method, which aims to preserve the image details while keeping the lower computation load. For multi-frame SR, a double sparsity structure [31] has been recently proposed to overcome the separation problem of block matching and sparse coding in traditional procedure.

The pioneer work of deep learning based image SR was introduced by Dong et al. in [7,8]. In their work, SRCNN is designed to directly learn the non-linear mapping between LR and HR images, which consists of three convolution layers. SRCNN implements patch extraction and representation, non-linear mapping and image reconstruction, to simulate the processing of a sparse coding procedure to generate HR images. However, SRCNN ignores the image prior, which is a significant component for image recovery. Liang et al. [32] introduce Sobel edge prior so as to capture gradient information to accelerate the training convergence. In fact, the method does reduce the training time but the resultant reconstruction enhancement is limited. Wang et al. [9] directly extended the conventional sparse coding with the merits of deep learning. Based on the learned iterative shrinkage and thresholding algorithm, a feed-forward neural network is implemented. To exploit both external and self-similarities, Wang et al. [18] also produced a deep joint super-resolution model which contains complex fine-tuning operations and is not end-to-end.

Please cite this article as: H. Liu et al., Single image super-resolution using a deep encoder–decoder symmetrical network with iterative back projection, Neurocomputing (2017), https://doi.org/10.1016/j.neucom.2017.12.014