Signal Processing: Image Communication 68 (2018) 88–100

Contents lists available at ScienceDirect

Signal Processing: Image Communication

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

Deep unsupervised learning for image super-resolution with generative adversarial network

Guimin Lin [a,b], Qingxiang Wu [a,*], Liang Chen [a], Lida Qiu [a,b], Xuan Wang [a], Tianjian Liu [b], Xiyao Chen [b]

[a] Key Laboratory of Optoelectronic Science and Technology for Medicine of Ministry of Education, College of Photonic and Electronic Engineering, Fujian Normal University, Fuzhou 350007, PR China
[b] Department of Physics and Electronic Information Engineering, Minjiang University, Fuzhou 350108, PR China

ARTICLE INFO

Keywords:
Super-resolution
Deep unsupervised learning
Sub-pixel convolution
Regularizer
Generative adversarial network

ABSTRACT

The aim of Image super-resolution (SR) is to recover high-resolution images from low-resolution ones. By virtue of the great success in numerous computer vision tasks achieved by the convolutional neural networks (CNNs), it is a nice direction to tackle the SR problem using CNNs. Despite progress in accuracy of SR using deeper CNNs, those models are almost trained base upon supervised way. In this paper, we propose a deep unsupervised learning approach for SR with a Generative Adversarial Network (GAN) framework, which is composed of a deep convolutional generator network with dense connections and a discriminator. A sub-pixel convolutional layer is operated on the top of the generator to upscale the inputs, and the standard convolutions are all implemented in the LR space, which leads to a fast restoration. The generator is trained to directly recover the high-resolution image from the low-resolution image. Strided convolution and ReLU activations are employed in the discriminator to distinguish the HR images from the produced HR images. The generator model is optimized with a combination of a data error, a regular term and an adversarial loss, which ensures local–global contents consistency and pixel faithfulness. Note that no labeled training data is employed during the training. Comparisons with several state-of-the-art supervised learn-based methods, experimental results demonstrate that the proposed model achieves a comparable result in terms of both quantitative and qualitative measurements, and it also implies the feasibility and effectiveness of the proposed unsupervised learning-based single-image super-resolution algorithm.

1. Introduction

High-resolution (HR) images often preserve more details and critical information that play a key role in numerous fields, such as medical imaging, surveillance, astronomical imaging and face recognition. Traditional approaches of obtaining HR images mainly depend on increasing the chip size or reducing the pixel size. Nevertheless, increasing the chip size will be followed by a growth in capacitance, and reducing pixel size will lead to an increase in the shot noise. In addition, the high cost of high-precision optical devices and image sensors is also an important concern in many commercial applications regarding HR imaging. Consequently, as an effective technique, which can produce visually pleasing HR images from a low-cost imaging system and limited environmental condition, image super-resolution (SR) has received a lot of attention.

The aim of SR is recovering the original HR image from one or more low-resolution (LR) images by inferring all the missing high-frequency contents, based upon reasonable assumptions or prior knowledge about the imaging process. However, SR is inherently ill-posed because a given LR image may correspond with many HR images due to the degradation factors (e.g. blurring, noising and downsampling). Various methods have been proposed to tackle the ill-posed problem. Traditional SR methods are based on interpolation, such as bicubic interpolation and Lanczos resampling [1], are frequently exploited because of their computational simplicity, but due to use of low order polynomials, this type of SR approaches is lack of fine details in the produced HR images and is prone to produce blurring details in textures and edges. Then several algorithms based on reconstruction constraint or priors are widely studied, including edge-directed priors [2], gradient profile priors [3], and non local self-similarity priors [4]. This kind of reconstruction-based SR methods is particularly effective to preserve

* Corresponding author.
E-mail address: qxwu@fjnu.edu.cn (Q. Wu).

https://doi.org/10.1016/j.image.2018.07.003
Received 22 April 2018; Received in revised form 3 July 2018; Accepted 4 July 2018

geometric structure and to suppress ringing artifacts, however, it fails to insert sufficient novel high-frequency components to the reconstructed HR outputs and is limited in reconstructing the visual complexity of the nature image, especially at high magnification.

Lately, learning-based methods have been extensively explored to model mapping from LR to HR patches. This category of schemes supposes that there exist certain relationships between the LR images and their corresponding HR counterparts and that these relationships can be learned from millions of co-occurrence LR–HR image patches, before they are employed to produce a new HR image. Since the learning-based approaches exploit the information on training images effectively, they have the ability to recover the missing high-resolution details caused by the downsampling and are superior to other SR methods. For example, Yang et al. [5] learned a compact dictionary based on sparse signal representation to generate high-resolution image. Timofte et al. proposed an improved ANR framework (A+) [6] by combining the best qualities of ANR [7] and simple functions. Tang et al. [8] achieved convincing improvement in terms of the reconstruction quality and computational cost by merging an improved structured output regression machine(SORM) and sparse coding.

Recently, motivated by the great success achieved by deep learning [9–12] in various computer vision tasks, researchers begin to exploit convolutional neural networks (CNNs) [13] with deep architecture to improve the reconstruction accuracy for image SR. Recent state-of-the-art methods mostly adopt the CNN-based models, which have provided a new inspiration and direction for the SR problem. Dong et al. [14,15] used convolutional neural networks to address the SR problem (named SRCNN), which draws considerable attention due to its simple network structure and excellent restoration quality. Wang et al. [16] employed deep learning techniques with sparse coding and achieve notable improvement over the generic SC model. Lin et al. [17] proposed a cascade of dilated convolutional neural network(CDCNN), which benefits from the end-to-end training of deep network with a specially designed skip-connections and dilation rate. SRCNN and other methods based on CNNs [16–22] have shown exciting performance.

Although the image super-resolution methods based on CNNs have obtained great success, these methods are almost classified as supervised learning. Deep unsupervised learning for SR problem is seldom addressed. In this paper, we work on exploring the feasible approach to tackle the ill-posed SR problem via deep unsupervised learning. To obtain a well-posed solution, a crucial factor is obtaining an effective regularizer or constraint on a SR reconstruction algorithm. The representative regularization models, such as the Tikhonov regularization [23], Total Variation (TV) [24,25] and bilateral TV (BTV) [26] are effective in removing image noise and outstanding for attractive edge preserving ability. Inspired by the exciting performance of Generative Adversarial Networks (GANs) [11] in unsupervised representation learning, we design a dense-connected CNN generator architecture and a discriminator network, and develop a new regularizer motivated by BTV smoothness prior [26] to well pose the SR problem via unsupervised learning. Our proposed model is competitive with some state-of-the-art methods based upon supervised learning in terms of both recovery accuracy and human perception, as presented in Fig. 1. The model SRCNN-915 [15], trained on 91 images [7] , is composed of 3 convolutional layers, and the kernel size of each layer is 9, 1 and 5 respectively. SRCNN-955 [15] also consists of 3 convolutional layers, however, it is trained on ILSVRC 2013 ImageNet and the kernel size in each layer is 9, 5 and 5 respectively. The images illustrated in the first column are recovered by classical bicubic method and the corresponding region of interesting (ROI), and the number under the images are the PSNR(dB) index. The followed three columns are obtained from SRCNN-915, SRCN-955, and our proposed model, respectively. The last column is the original images. It shows that our proposed approach generates sharp edges with rare artifacts and is most close to the ground truth in terms of subjective visual evaluation even the PSNR index is inferior to SRCNN-955 for butterfly image. In addition, the computational speed of our algorithm is faster than the two SRCNN models.

In short, the contributions of this paper include:

- We propose a novel approach to handle the ill-posed SR problem with deep unsupervised learning. Dense connections are employed in our generator model to combine the local texture information and the global abstraction information. A sub-pixel convolution is used at the top of our generator to upscale the inputs and results in a fast restoration. The generator model is optimized with non-label training data, and achieves a pleasing SR performance.
- Inspired by the BTV regularization, we develop a simple regularizer which combines of the first and second order derivatives of image with separate weight, thus it implies the smoothness of image continual section and preserves the sharp edge.
- We detail the configuration of the discriminator employed in our GAN framework, and demonstrate that deep unsupervised learning is feasible for the problem of image super-resolution, and can achieve good quality.

The rest of this paper is organized as follows. The related work is firstly reviewed in Section 2. In Section 3, the proposed method and some key components are presented. The databases used for evaluation and the experiment results are demonstrated in detail in Section 4. Finally, the main conclusions are presented in Section 5.

2. Related work

There have been numerous publications over the last few years employing learning-based policies on the SR tasks. Compared with conventional SR methods, which depend on hand-crafted features, learning-based approaches may further boost the performance, especially CNN-based methods with deep learning technique.

2.1. Image super-resolution

The goal of image super-resolution is reconstructing an HR image from the LR one. To improve the reconstruction accuracy for image SR, recent state-of-the-art methods mostly employ the supervised learning-based approaches. This type of methods exploits a set of HR images and their corresponding downsampled LR ones to learn dictionaries, regression functions or end-to-end nonlinear mapping between the two. The dictionary-based techniques attempt to create a correspondence map between the LR and HR images by space transformation. Searching in this type of dictionaries is performed via approximate nearest neighbors, as exhaustive search would be prohibitive time cost. In addition, dictionaries quickly grow in size with the amount of training data. Yang et al. [5] proposed a technique to obtain a sparse "compact dictionary" from the training data to tackle the problem of growing dictionary sizes. He et al. [27] utilized the beta process prior to learn the over-complete dictionary pairs for adding more consistent and accurate mapping between two feature space. Ahmed and Shah [28] learned multiple coupled dictionaries, each containing features along a different direction. The high-resolution patch is reconstructed using a set of directional clustered dictionaries which gives the least sparse representation error. Due to hard capture the statistical variability of face images by only exploiting fixed l_1 norm penalty, Wang et al. [29] proposed a weighted adaptive sparse regularization (WASR) method for face hallucination reconstruction. They also introduced a neighbor embedding (NE) from the low- and high-resolution image manifolds simultaneously and proposed a coupled-layer NE (CLNE) [30] for very low resolution facial image restoration. Rasti et al. [31] just employed separate dictionaries for LR and HR patches and proposed a low complexity approach to generate HR images.

The regression-based methods typically establish regression models to reveal the relationships between the features of LR patches and a single HR patch. Dictionaries can also be leveraged together with regression based approaches to compute projection matrices to produce