Photo-Realistic Single Image Super-Resolution

Hongzhi Liu

Jun 26, 2018

Abstract

As we all know, single image super-resolution break through in accuracy and speed using faster and deeper convolutional neural networks. However, there is one central problem remains largely unsolved that how to recover the finer texture details when we super-resolve at large upscaling factors. Today, I read a thesis written by Christian Ledig, who is from Twitter. His team introduce SRGAN, a generative adversarial network (GAN) for image super-resolution (SR). To our knowledge, it is the first framework capable of inferring photo-realistic natural images for 4× upscaling factors. An extensive mean-opinion-score (MOS) test shows hugely significant gains in perceptual quality using SRGAN. The MOS scores obtained with SRGAN are closer to those of the original high-resolution images than to those obtained with any state-of-the-art method.

1. Overview of SRGAN

Recent overview articles on image SR include Nasrollahi and Moeslund [6] or Yang *et al.* [8]. Prediction-based methods were among the first methods to tackle SISR. Besides, more powerful approaches aim to establish a complex mapping between low and high resolution image information and usually rely on training data. More powerful approaches aim to establish a complex mapping between low- and high-resolution image information and usually rely on training data. Many methods that are based on example-pairs rely on LR training patches for which the corresponding HR counterparts are known. And recently convolutional neural network (CNN) based SR algorithms have shown excellent performance. Subsequently, it was shown that enabling the network to learn the upscaling filters directly can further increase performance both in terms of accuracy and speed [2]. Johnson *et al.* [3] and Bruna *et al.* [1] rely on a loss function closer to perceptual similarity to recover visually more convincing HR images.

In this paper, Christian Ledig and his team propose a super-resolution generative adversarial network (SRGAN) for which they employ a deep residual network (ResNet) with skip-connection and diverge from MSE as the sole optimization target [4]. They define a novel perceptual loss using high-level feature maps of the VGG network [3] combined with a discriminator that encourages solutions perceptually hard to distinguish from the HR reference images. An example photo-realistic image that was superresolved with a 4× upscaling factor is shown in Fig. 1.

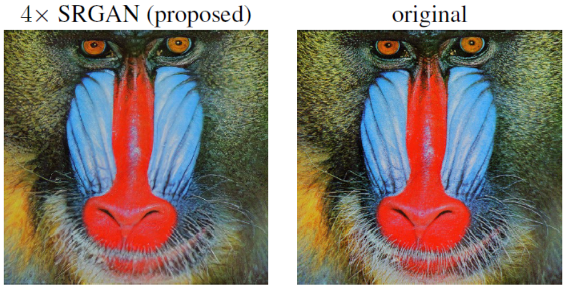

Figure 1. Super-resolved image (left) is almost indistinguishable from original (right).

2. Method of the Super-Resolution Generative Adversarial Network

In SISR the aim is to estimate a high-resolution, superresolved image l^{SR} from a low-resolution input image I^{LR}. The ultimate goal of Ledig is to train a generating function G that estimates for a given LR input image its corresponding HR counterpart. To achieve this, he trains a generator network as a feed-forward CNN G_{θ_G} parametrized by θ_G as Eq. 1:

$$\hat{\theta_G} = arg\min_{\theta_G} \frac{1}{N}\sum_{n=1}^{N} l^{SR}(G_{\theta G}(I_n^{LR}), I_n^{HR}). \qquad (1)$$

Here $\theta_G = \{W_{1:L}; b_{1:L}\}$ denotes the weights and biases of a L-layer deep network and is obtained by optimizing a SR-specific loss function I^{SR}. For training images I_n^{HR}, $n = 1, \ldots, N$ with corresponding I_n^{LR}, $n = 1, \ldots, N$. In this work, the team specifically design a perceptual loss I^{SR} as a weighted combination of several loss components that

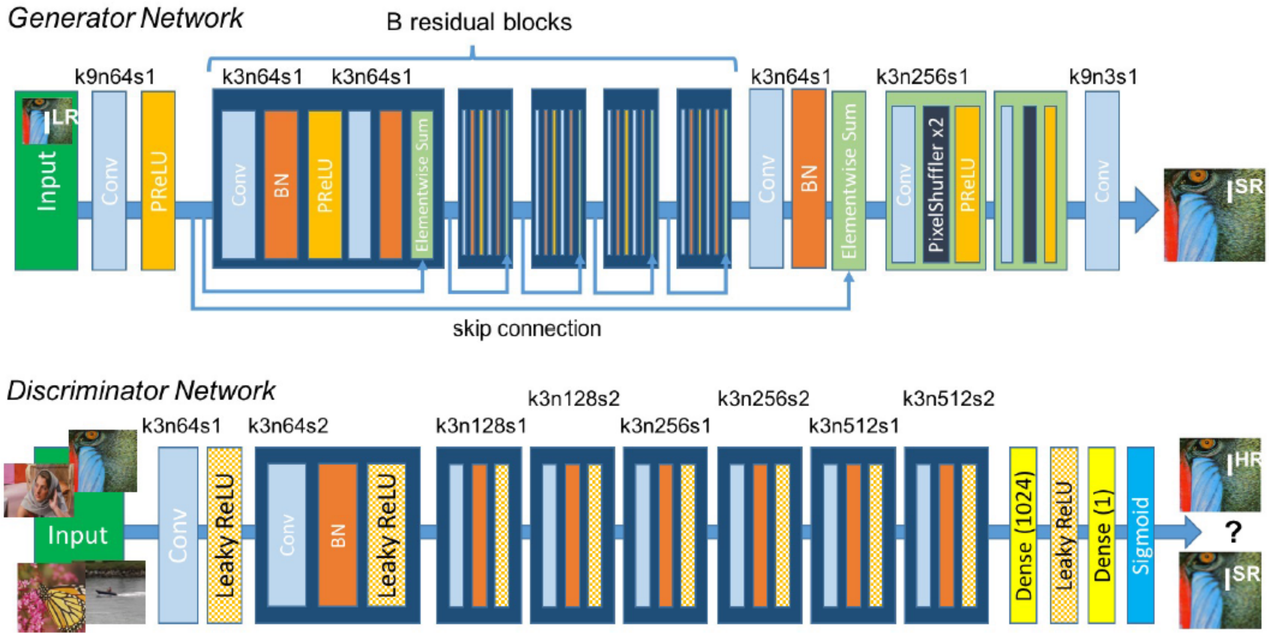

Figure 2. Architecture of Generator and Discriminator Network with corresponding kernel size (k), number of feature maps (n) and stride (s) indicated for each convolutional layer.

model distinct desirable characteristics of the recovered SR image.

2.1. Adversarial network architecture

Ledig and his team further define a discriminator network $D_{\theta D}$ which they optimize in an alternating manner along with $G_{\theta G}$ to solve the adversarial min-max problem:

$$\min_{\theta_G} \max_{\theta_D} \mathbb{E}_I^{HR} \sim P_{train}(I^{HR})[\log D_{\theta D} I^{HR}] + \\ \mathbb{E}_I^{LR} \sim P_G(I^{LR})[\log(1 - D_{\theta D}(G_{\theta G} I^{LR}))]. \quad (2)$$

The general idea behind this formulation is that it allows one to train a generative model G with the goal of fooling a differentiable discriminator D that is trained to distinguish super-resolved images from real images. With this approach their generator can learn to create solutions that are highly similar to real images and thus difficult to classify by D. This encourages perceptually superior solutions residing in the subspace, the manifold, of natural images. This is in contrast to SR solutions obtained by minimizing pixel-wise error measurements such as the MSE.

To discriminate real HR images from generated SR samples the team train a discriminator network. The architecture is shown in Fig. 2. They follow the architectural guidelines summarized by Radford *et al.* [7] and use LeakyReLU activation (α = 0.2) and avoid max-pooling throughout the network.

Table 1. Performance of different loss functions for SRResNet and the adversarial networks on Set5 and Set14 benchmark data. MOS score significantly higher ($p < 0.05$) than with other losses in that category.

	SRResNet-		SRGAN-		
Set5	MSE	VGG22	MSE	VGG22	VGG54
PSNR	32.05	30.51	30.64	29.84	29.40
SSIM	0.9019	0.8803	0.8701	0.8468	0.8472
MOS	3.37	3.46	3.77	3.78	3.58

2.2. Perceptual loss function

The definition of Ledig's perceptual loss function l^{SR} is critical for the performance of their generator network. They formulate the perceptual loss as the weighted sum of a content loss (l_X^{SR}) and an adversarial loss component as Eq. 3:

$$l^{SR} = l_X^{SR} + 10^{-3} l_{Gen}^{SR}. \quad (3)$$

While l^{SR} is commonly modeled based on the MSE [10, 48], he improves on Johnson *et al.* [3] and Bruna *et al.* [1] and design a loss function that assesses a solution with respect to perceptually relevant characteristics.

3. Empirical Validation of DCGANs Capabilities

The team perform experiments on three widely used benchmark datasets Set5, Set14 and BSD100, the testing set of BSD300 [5]. All experiments are performed with a scale factor of $4\times$ between low and high resolution images. The experimental results of the conducted MOS tests are summarized in Tab. 1.

Besides, the team also evaluate the performance of the generator network without adversarial component for the two losses l_{MSE}^{SR}(SRResNet-MSE) and $l_{VGG/2.2}^{SR}$ (SRResNet-VGG22). They refer to SRResNet-MSE as SRResNet and quantitative results are summarized in the table above.

References

[1] J. Bruna, P. Sprechmann, and Y. Lecun. Super-resolution with deep convolutional sufficient statistics. *Computer Science*, 2015. 1, 2

[2] C. Dong, C. L. Chen, and X. Tang. Accelerating the super-resolution convolutional neural network. In *ECCV*, 2016. 1

[3] J. Johnson, A. Alahi, and F. F. Li. Perceptual losses for real-time style transfer and super-resolution. In *ECCV*, 2016. 1, 2

[4] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, and Z. Wang. Photo-realistic single image super-resolution using a generative adversarial network. In *CVPR*, 2017. 1

[5] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *ICCV*, 2002. 3

[6] K. Nasrollahi and T. B. Moeslund. Super-resolution: A comprehensive survey. *Machine Vision and Applications*, 2014. 1

[7] A. Radford, L. Metz, and S. Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *Computer Science*, 2015. 2

[8] C. Y. Yang, C. Ma, and M. H. Yang. Single-image super-resolution: A benchmark. In *ECCV*, 2014. 1