A Survey on Image Super-Resolution

Jingxi Yu

Computer Science Department, UCLA, Los Angeles, California, United States, jingxi@g.ucla.edu

ABSTRACT

Image Super-Resolution (SR) is a classic task in the field of computer vision and image processing. There are various super-resolution applications in surveillance, medical imaging etc. With the recent breakthrough of deep learning techniques, there are major improvements of SR techniques as well. This survey briefly introduces history of super-resolution progress and discusses several representative approaches in detail. A benchmark among these methods is presented.

KEYWORDS

Super-resolution, Deep learning, SRGAN, Reference-based SR

1 INTRODUCTION

Image super-resolution reconstructs high-resolution images from low resolution images. The idea of image super-resolution was first mentioned in medical imaging processing field in 1978. Since then, super resolution techniques have displayed great value in the field of satellite and aerial imaging, surveillance, automated mosaicking, infrared imaging, text improvement, fingerprint image enhancement, facial images and medical imaging. With the development of streaming and video platform, there is also increasing demand of recovering vintage videos to high resolution.

Early super-resolution approaches range from signal processing perspective to machine learning perspective. Surveys focused on traditional methods like [1][2] states that the problem with these traditional approaches are lack of robustness and inefficiency. Before the increase in computing power and rise in deep learning, example-based approaches became more popular than others. Freeman et al. first presented example-based and learning-based super-resolution framework in 2002 [3]. Early learning-based methods also suffers from over- or under-fitting and inefficiency in time. Yang et al. proposed a sparse representation to address these problems and improve the performance [4].

Since convolutional neural networks (CNN) became more popular in last decades, Dong et al. [5] first proposed to use a CNN to model similar pipeline as a sparse model and achieved impressive results. To achieve better result, Dong et al. designed a slightly deeper architecture called FSRCNN [6] based on SRCNN. It has been proven that deeper architectures such as deeply-recursive CNN and deep recursive residual network applied to solve super-resolution tasks could achieve better results. However, there are certain bottlenecks of deeper architectures such as optimization during training process. Besides, these models tend to blur details to avoid mean square error penalty during training process. To restore more details, Legid et al. proposed to use a generative adversarial network which uses a loss function that access respect to perceptually characters [13].

GAN-based super resolution approaches seem to out-performed other deep learning approaches with realistic details. However, GAN tends to over compensate on details, such as adding details and textures to where they do not belong. To improve the quality of high-frequency details, SFTGAN proposed a GAN model with a SFT layer which can add features based on the semantic class information [15]. There are already segmentation networks pretrained could help segmentation and classification. There are also more work combining deep learning and example-based methods like [16][17].

In this survey, the evaluation is mainly based on basic metrics like peak signal-to-noise ratio (PSNR), structural similarities (SSIM) as well as human perception in general sense. For specific fields like medical microscopy and MRI images, the quality of results is less intuitive thus these metrics might not be as useful for specific field.

2 DEEP LEARNING BASED APPROACHES

In recent years, deep learning based methods have achieved great results in various fields including super resolution. Recent surveys such as [7][8] intend to summarize and categorize recent works based on their model framework and network design. Networks could be mainly divided into linear networks, residual networks, recursive networks, densely connected networks, GAN and others. Beside mainstream single image super-resolution, there are also reference-based networks. This survey focuses on analyzing these methods case by case. The overview of methods covered in this survey is shown in table 1.

Table 1. Overview of super-resolution methods

Method	Network Design	Year	Depth	Loss Function
SRCNN	Linear	2014	3	L2
FSRCNN	Linear	2016	81	L2
DRCN	Recursive	2016	5(Recursive)	L2
LapSRN	Progressive sampling	2017	24	Charbonnier
SR-DenseNet	Densely Connected	2017	64	L2
SRGAN	GAN model	2017	33	Perceptual loss, MSE
SRFeat	GAN model	2018	54	L_d, MSE
SFTGAN	GAN model	2018	33	L_d, MSE
SRNTT	Ref-Based	2019	63	L1, L_d, L_p

2.1 Deep Architectures

2.1.1 SRCNN

Super-Resolution Convolutional Neural Network (SRCNN) [5] is considered the first successful deep learning based approach in super-resolution which provides great insight for later work. The architecture is quite shallow compare to later SISR models. The first step of the pipeline is up-sampling the input to output size using bicubic interpolation. Then the structure consists of three convolutional layers and two rectified linear unit (ReLU) layers stacked together. As the paper states, the goal of three convolutional layers are feature extraction, non-linear mapping and reconstruction by aggregating feature maps. The visualized structure of SRCNN is shown in Figure 1. The cost function used in the training process to minimize the difference between reconstructed output and the ground truth is mean square error (MSE).

2.1.2 FSRCNN

Fast Super Resolution Convolutional Neural Networks (FSRCNN) [6] is an improved version of SRCNN proposed by the same research team. FSRCNN has four convolutional layers and one deconvolutional layer.

Each of the layers serves the purpose of feature extraction, shrinking, mapping, expanding and deconvolution. Unlike SRCNN which pre-upsampling the input image using bicubic interpolation before feeding to the network, FSRCNN does not upsample until the deconvolutional layer. Even though there are more layers in this model, post-upsampling still make FSRCNN a faster model than SRCNN. The cost function is MSE, same as SRCNN. And data augmentation is used to increase the training data. The structure of FSRCNN in shown in Figure 1.

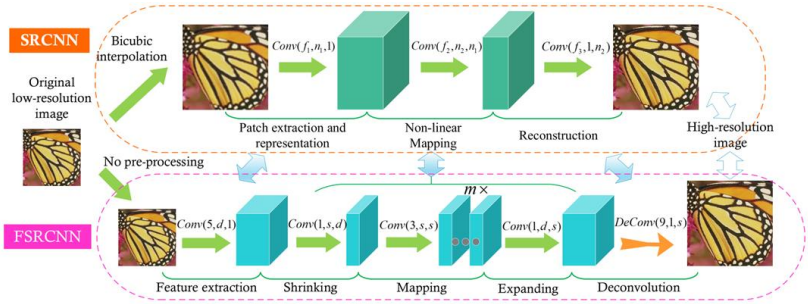

Figure 1. The visualized structure comparison of SRCNN and FSRCNN provided in [6]

2.1.3 DRCN

Deep-Recursive Convolutional Network for Image Super-Resolution [9] (DRCN) is considered the first model that applied Recursive Neural Network in super-resolution. The DRCN contains three parts. The first is embedding network, which is equivalent to feature extraction. The second part is called inference network which performs super resolution by using a single recursive layer and followed by ReLU. And the last part, reconstruction network, is used to reconstruct result from features as the name indicated. This work efficiently reuses weight parameters and ease the difficulty in training compared to its previous work.

2.1.4 LapSRN

Deep Laplacian Pyramid Networks for Fast and Accurate Super-Resolution (LapSRN) [10] is proposed by Lai et al. in 2017. The article states that pre-upsampling step increase unnecessary computational cost and post-upsampling does not work well with large scaling factors, thus LapSRN progressively reconstructs the sub-residuals of high-resolution images. Another issue mentioned is that l2 loss would make output images blurry. To address this issue, LapSRN uses a robust Charbonnier loss function. The more high-level comparison between above methods are shown in Figure 2.

Laplacian pyramid has been a classic image processing technique and is used in image compression. It is very insightful to combine traditional image processing algorithm with deep learning. The cascading architecture and robust Charbonnier loss function make sure this method achieves satisfactory visual result while using less time than SRCNN and DRCN.

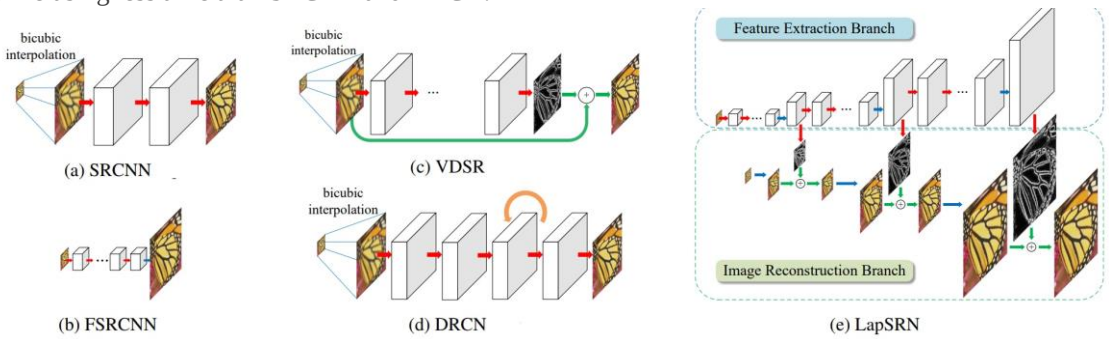

Figure 2. Network architectures comparison provided by the paper. [10]

2.1.5 SR-DenseNet

Image Super-Resolution Using Dense Skip Connections (SR-DenseNet) [11] is employed based on dense connected convolutional networks (DenseNet) [12]. The DenseNet connects each layer to every other layer in a feed-forward way. Such architecture alleviates the vanishing gradient problem, avoids redundant features and reuses feature maps. Deconvolution layers were integrated to recover details and speed up the process. The process is consisting of four parts, low level features, high level features (dense blocks), deconvolution layers and reconstruction layer. The author proposed 3 architectures with slightly different design. Among three architectures, concatenate features from all layers proves to be better than just using partial features.

2.2 Generative Adversarial Network

2.2.1 SRGAN

Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network [13] presented two architectures, SRResnet and SRGAN, a generative adversarial network for super-resolution. The task of super-resolution is ill posed and the mainstream loss function MSE could not fully represent the visual effect. MSE and PSNR are popular metrics to evaluate super-resolution algorithms. However, they are defined pixel-wised, not perceptually. High MSE or PSNR score does not guarantee best perceptual result. To address this issue, SRGAN proposed a loss function calculated on feature maps of VGG network [14], which is more invariant to change per pixel.

The major contribution of this work is to bring GAN to the table of current super-resolution methods. The general idea is to train a generative model with the goal of fooling a discriminator that is trained to distinguished super-resolved images from real images. The architecture of generator and discriminator is shown below.

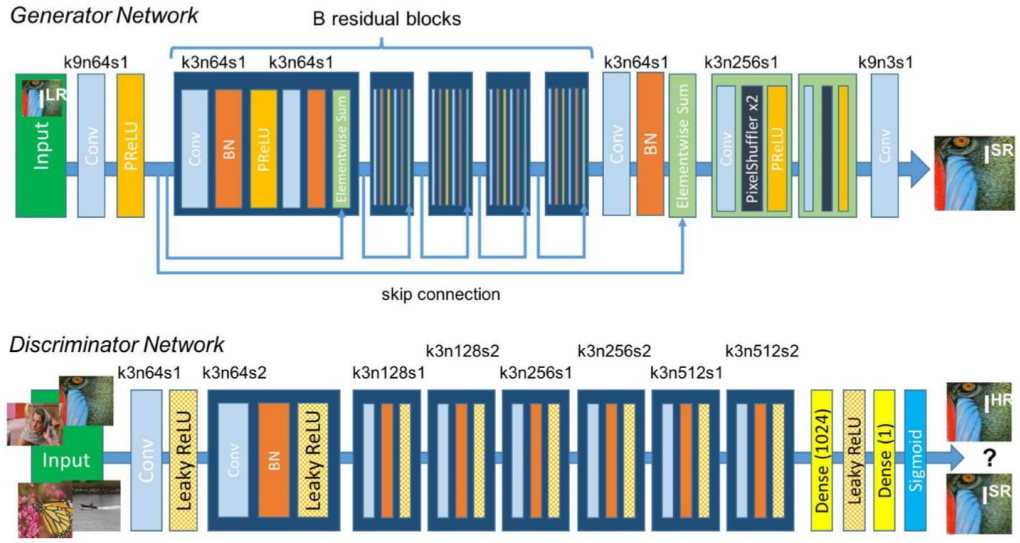

Figure 3. Architecture of generator network and discriminator network [13]

2.2.2 SRFeat

SRFeat [14] is also a GAN-based SISR method. The paper states that GAN-based super-resolution methods tend to produce less meaningful high-frequency noise in output images because the generator is trying to mimic high frequency details, so the discriminator would be fooled. In SRGAN [13], a perceptual loss that minimizes MSE of VGG features is used. Like MSE on pixels, MSE on VGG features would not be enough to fully represent perceptual effect. To improve these issues, SRFeat employs two different discriminators, the image discriminator and the feature discriminator. The feature discriminator includes both high-frequency components and structural components.

SRFeat also proposed a new generator network with long-range skip connections so that information in distant layers could be concatenate together efficiently. The provided results seem realistic with details with the feature discriminator and long-range skip connection.

2.2.3 SFTGAN

Previous works have shown great examples of reconstructed images. However, texture details remain an important issue. In this paper, Recovering Realistic Texture in Image Super-resolution by Deep Spatial Feature Transform (SFTGAN) [15], the authors believe that categorical prior which characterizes the sematic class of segmented region plays an important rule in recovering texture details. Multiple categories could co-exist in one image, which make this more challenging.

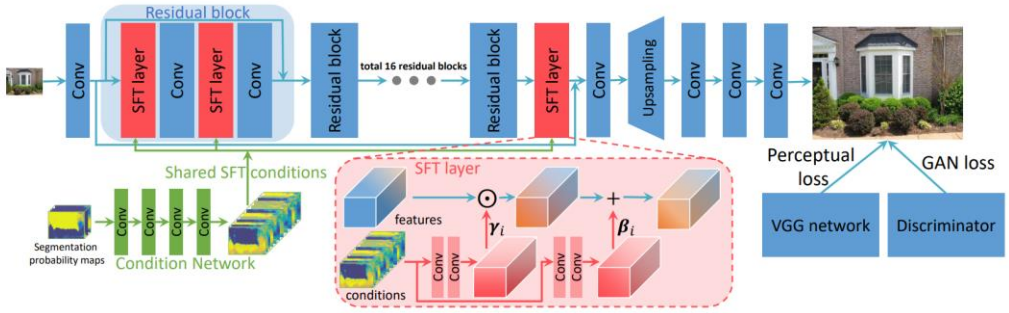

Figure 4. Generator Architecture of SFTGAN [15]

The overall framework used is based on adversarial learning like SRGAN [13] which consists of one discriminator and one generator. Besides, SFTGAN proposed a spatial feature transform (SFT) layer which could change the behavior of SR network. The SFT layer is conditioned on sematic segmentation probability maps and is capable of both feature-wise and spatial-wise transformation. As shown in Figure 4, the SFT layer is a single pass forward through the network which is far more efficient than training separate SR model for each sematic class. The segmentation network is trained separately on COCO dataset and the SFT layer could be trained end to end with the SR network. For discriminator, SFTGAN uses a VGG-style network.

A comparison of popular SR methods reconstructed output is provided by the paper, shown in Figure 5. As we could see, earlier networks tend to be blurry and not much improvement from bicubic interpolation in texture details. And GAN-based model tends to create high frequency unrealistic details. The image SFTGAN reconstructed is slightly different from original HR image but look real by itself.

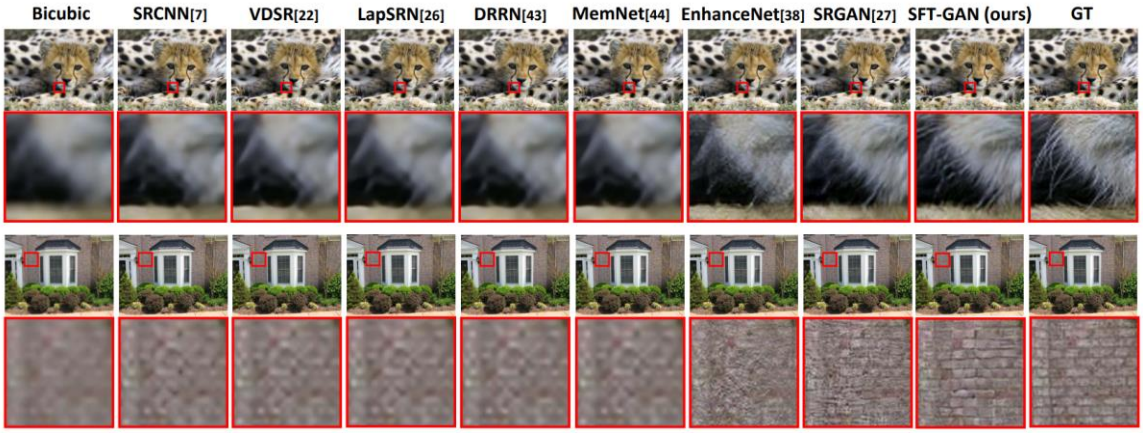

Figure 5. Restored images comparison [15]

2.3 Reference-Based Super-Resolution

2.3.1 SRNTT

Image Super-Resolution by Neural Texture Transfer (SRNTT) [16] is reference-based super-resolution (RefSR) approach. Reference-based methods are first mentioned by Freeman et al. in the early 2000s [3]. Recent SISR methods could not make major breakthroughs in upscaling large factors due to extremely lack of information. SFTGAN [15] proposed one way to add textures to input image based on current segmentation algorithms. SRNTT diverts from traditional SISR and continues to explore more of reference-based methods. Unlike previous reference-based, SRNTT does not depend on good alignment to achieve good result. The main framework of SRNTT consists of texture matching and transfers matched. Texture matching is performed in the feature space which is considered more invariant to color and lighting. Since the LR image is blurry so the paper sequentially applied up-sampling and down-sampling on the LR and

reference image. Then the structural and textural similarity is measured by an inner product between neural features. The neural texture transfer is designed by swapped texture feature maos into a base deep generative network at differen feature layers. The texture transfer model is shown in Figure 6. The residual block and skip connections are similar to [9] [13].

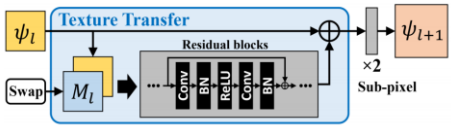

Figure 6. Texture Transfer Model [16]

Training with MSE loss, SRNTT could achieve top result in PSNR/SSIM compare to other SISR methods. With other loss function, the results seem better detailed while losing PSNR score. SRNTT's PSNR/SSIM result is the state of the art. With the demonstrated results, SRNTT outperforms the state of the art SISR method SRGAN [13] and Ref-based method CrossNet [17].

3 BENCHMARKS

Here in is a cross comparison of provided PSNR/SSIM result among different benchmark datasets shown in Table 2. The results are for 4x scale.

Table 2. A benchmark among paper introduced in this survey

Datasets	Set 5	Set 14	BSD 100	Urban100
Bicubic	28.42 / 0.810	26.00 / 0.704	25.96 / 0.669	23.14 / 0.674
SRCNN	30.48 / 0.862	27.49 / 0.754	26.91 / 0.712	24.41 / 0.738
FSRCNN	30.71 / 0.865	27.70 / 0.756	26.97 / 0.714	25.14 / 0.760
DRCN	31.35 / 0.884	28.04 / 0.770	27.24 / 0.724	24.26 / 0.735
LapSRN	31.54 / 0.885	28.19 / 0.772	27.32 / 0.728	25.14 / 0.751
SR-DenseNet	32.02 / 0.893	28.50 / 0.778	27.53 / 0.733	25.21 / 0.755
SRGAN	32.05 / 0.891	28.53 / 0.780	27.57 / 0.735	26.07 / 0.783
SRFeat	32.27 / 0.983	28.71 / 0.783	27.64 / 0.737	-
SFTGAN	29.82 / 0.840	26.13 / 0.694	25.33 / 0.651	-
SRNTT	-	-	-	25.50 / 0.783

4 DISCUSSION AND CONCLUSION

Deep networks have shown great progress over recent years. However, single image super-resolution seems reached its bottleneck. With little improvement in PSNR/SSIM score, the reconstructed images tend to be overly smoothed. The deeper architectures have made progresses but not a breakthrough. Generative adversarial networks help bring more details to the reconstructed images as well as high-frequency noises. On the high level, using semantic segmentation information [15] to help with the lack of information with low-resolution images is very insightful. Reference-based methods are rather non-mainstream but have demonstrated great potential in the future. Combining early image processing super-resolution techniques and deep learning could be very inspiring and might give better results than just using deep network by itself.

5 REFERENCE

[1] J. Yang and T. Huang. Image super-resolution: historical overview and future challenges. In P. Milanfar, editor, Super-resolution imaging, chapter 1. CRC Press, 2010.

[2] J. Van Ouwerkerk, “Image super-resolution survey,” Image and Vision Computing, vol. 24, 2006.
[3] W. T. Freeman, T. R. Jones, and E. C. Pasztor, “Example-based super -resolution,” IEEE Computer Graphics and Applications, vol. 22, 2002.
[4] Yang, J., Wright, J., Huang, T.S., Ma, Y.: Image super-resolution via sparse representation. IEEE Transactions on Image Processing 19(11), 2861–2873 (2010).
[5] C. Dong, C. C. Loy, K. He, and X. Tang, “Image super-resolution using deep convolutional networks,” TPAMI, 2016.
[6] C. Dong, C. C. Loy, and X. Tang, “Accelerating the super-resolution convolutional neural network,” in ECCV, 2016.
[7] Z. Wang, J. Chen, and S. C. Hoi. Deep Learning for Image Super-resolution: A Survey. arXiv preprint arXiv:1902.06068, 2019.
[8] S. Anwar, S. Khan and N. Barnes. A Deep Journey into Super-resolution: A Survey. arXiv:1904.07523, 2019.
[9] J. Kim, J. Kwon Lee, and K. Mu Lee, “Deeply-recursive convolutional network for image super-resolution,” in CVPR, 2016
[10] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, “Deep laplacian pyramid networks for fast and accurate super-resolution,” in CVPR, 2017.
[11] T. Tong, G. Li, X. Liu, and Q. Gao, “Image super-resolution using dense skip connections,” in ICCV, 2017.
[12] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in CVPR, 2017.
[13] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, ´ A. Acosta, A. P. Aitken, A. Tejani, J. Totz, Z. Wang et al., “Photorealistic single image super-resolution using a generative adversarial network.”
[14] S.-J. Park, H. Son, S. Cho, K.-S. Hong, and S. Lee, “Srfeat: Single image super-resolution with feature discrimination,” in ECCV, 2018.
[15] X. Wang, K. Yu, C. Dong, and C. C. Loy, “Recovering realistic texture in image super-resolution by deep spatial feature transform,” in CVPR, 2018.
[16] Z. Zheng, Z. Wang, Z. Lin and H. Qi, “Image super-resolution by neural texture transfer” in CVPR, 2019.
[17] H. Zheng, M. Ji, H. Wang, Y. Liu, and L. Fang. CrossNet: An end-to-end reference-based super resolution network using cross-scale warping. In ECCV, 2018.