SR: Super-Resolution, SISR: Single Image Super Resolution, SRCNN: Super-Resolution Convolution Neural Network, HR: High-Resolution, LR: Low-Resolution

Since the publication of this paper we discussed, a large number of improvement attempts have been made, such as new network designs with higher performance, learning strategies with different emphasis, evaluation metrics with the aim of reflecting the quality of the model reasonably.

There are some key ideas of novel network designs, which can achieve faster, more accurate and even lower-cost performance compared to the original one. Four representative networks are listed below.

a) Residual learning framework can ease the original convolutional neural network, and perform SISR accurately, efficiently [1]. With cascading residual learning it can even achieve a lightweight solver [2].
b) Using a generative adversarial network (GAN) to handle super resolution. The major advantage of it is to recover the finer texture details when we super-resolve at large upscaling factors [3].
c) Deeply recursive convolutional network (DRCN) can obtain larger receptive field and learn higher-level features. This network has a very deep recursive layer (up to 16 recursions). And increasing recursion depth can improve performance without introducing new parameters for additional convolutions [4].
d) Using dense skip connections in a very deep network to handle SISR. In this network, the feature maps of each layer are propagated into all subsequent layers, providing an effective way to combine the low-level features and high-level features to boost the reconstruction performance [5].

In brief, good network design not only determines a hypothesis space with great performance upper bound, but also helps efficiently learn data representations without excessive spatial and computational redundancy [6].

Besides, some new learning strategies are tried. The main learning strategy is the improvement of loss function (expert for loss function, there are other methods such as batch normalization [7]). In the super-resolution field, loss functions are used to measure the difference between generated HR images and ground truth HR images, and guide the model optimization. In early times, researchers usually employ the pixel-wise L2 loss, but later discover that it cannot measure the reconstruction quality very accurately [6]. Therefore, a large number of loss functions are adopted to better measure the reconstruction error. Three typical loss functions are listed below (Pixel loss, Content loss, Texture loss):

$$L_{pixel\_l2}(\hat{I}, I) = \frac{1}{hwc}\sum_{i,j,k}\left(\hat{I}_{i,j,k} - I_{i,j,k}\right)^2 \tag{1}$$

Where h, w, c denote the height, width and number of channels of the evaluated images, respectively. It is original loss function – L2 loss, i.e. mean square error. However, the pixel loss does not represent image quality (e.g. texture [8], perceptual quality [9]), since it tends to lack high-frequency details and overly smooth textures [3].

$$L_{content}(\hat{I}, I; \emptyset, l) = \frac{1}{h_l w_l c_l} \sqrt{\sum_{i,j,k} \left(\emptyset_{i,j,k}^{(l)}(\hat{I}) - \emptyset_{i,j,k}^{(l)}(I)\right)^2} \quad (2)$$

Where h_l, w_l and c_l are the height, width and number of channels of the extracted feature maps on layer l in the network $\emptyset$, respectively. The content loss is based on the perceptual quality, and introduced into super resolution [9], [10]. In contrast to the pixel loss, the content loss encourages the output image $\hat{I}$ to be perceptually similar to the target image I instead of forcing them to match pixels exactly.

$$L_{texture}(\hat{I}, I; \emptyset, l) = \frac{1}{c_l^2} \sqrt{\left(G_{i,j}^{(l)}(\hat{I}) - G_{i,j}^{(l)}(I)\right)^2} \quad (3)$$

According to [11], [12], the texture of an image is regarded as the correlations between different feature channels and defined as Gram matrix $G^{(l)} \in R^{c_l \times c_l}$, where $G_{i,j}^{(l)}$ is the inner product between the vectorized feature maps i and j on layer l. By applying texture loss, the SR model can create realistic textures and produce visually more satisfactory results [8]. Despite this, determining the size of the patch to match textures is still empirical.

However, evaluation metrics which are approved by most of researchers still does not come into an agreement. Metrics for super-resolution face many challenges and need more exploration. The most widely used metrics for super-resolution are PSNR and SSIM. However, the PSNR tends to result in excessive smoothing, and the results often vary wildly between almost indistinguishable images [6]. The SSIM fails to capture and accurately assess image quality with respect to the human visual system, though it performs evaluation in terms of luminance, contrast and structure [8], [13]. In addition, the Mean Opinion score (MOS) is closest to human visual response, whereas requires sufficient evaluators and evaluations and is highly cost [3], [8]. That is why more accurate metrics for evaluating reconstruction quality are urgently needed.

From my own perspective, finding a better evaluation metric is the topic most interesting to me. In general, the targets of super-resolution are to progress faster with more accurate results and lightweight cost in terms of efficiency, to produce more high-quality high-resolution images in terms of performance. For the former, there has been a majority of novel solutions to improve the efficiency. However, for what the high-quality high-resolution image is, there is still not a well-behaved and unified evaluation metric approved by most of researchers, though there have been many tries. Finding a better evaluation metric is of great importance, which could quantitatively evaluate a model and thus benefits to guide the research of super resolution in the future.

Reference:

[1] HE, Kaiming, et al. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016. p. 770-778.

[2] AHN, Namhyuk; KANG, Byungkon; SOHN, Kyung-Ah. Fast, accurate, and lightweight super-resolution with cascading residual network. In: *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018. p. 252-268.

[3] LEDIG, Christian, et al. Photo-realistic single image super-resolution using a generative adversarial network. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017. p. 4681-4690.

[4] KIM, Jiwon; KWON LEE, Jung; MU LEE, Kyoung. Deeply-recursive convolutional network for image super-resolution. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016. p. 1637-1645.

[5] TONG, Tong, et al. Image super-resolution using dense skip connections. In: *Proceedings of the IEEE International Conference on Computer Vision*. 2017. p. 4799-4807.

[6] WANG, Zhihao; CHEN, Jian; HOI, Steven CH. Deep learning for image super-resolution: A survey. *arXiv preprint arXiv:1902.06068*, 2019.

[7] IOFFE, Sergey; SZEGEDY, Christian. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015.

[8] SAJJADI, Mehdi SM; SCHOLKOPF, Bernhard; HIRSCH, Michael. Enhancenet: Single image super-resolution through automated texture synthesis. In: *Proceedings of the IEEE International Conference on Computer Vision*. 2017. p. 4491-4500.

[9] JOHNSON, Justin; ALAHI, Alexandre; FEI-FEI, Li. Perceptual losses for real-time style transfer and super-resolution. In: *European conference on computer vision*. Springer, Cham, 2016. p. 694-711.

[10] DOSOVITSKIY, Alexey; BROX, Thomas. Generating images with perceptual similarity metrics based on deep networks. In: *Advances in neural information processing systems*. 2016. p. 658-666.

[11] GATYS, Leon; ECKER, Alexander S.; BETHGE, Matthias. Texture synthesis using convolutional neural networks. In: *Advances in neural information processing systems*. 2015. p. 262-270.

[12] GATYS, Leon A.; ECKER, Alexander S.; BETHGE, Matthias. Image style transfer using convolutional neural networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016. p. 2414-2423.

[13] TODERICI, George, et al. Full resolution image compression with recurrent neural networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017. p. 5306-5314.