Non-local Network Routing
for Perceptual Image Super-Resolution

Zexin Ji[1], Xin Dong[1], Zhendong Li[1,2(✉)], Zekuan Yu[3], and Hao Liu[1,2]

[1] School of Information Engineering, Ningxia University, Yinchuan 750021, China
[2] Collaborative Innovation Center for Ningxia Big Data and Artificial Intelligence
Co-founded by Ningxia Municipality and Ministry of Education,
Yinchuan 750021, China
`lizhendong13@mails.ucas.ac.cn, liuhao@nxu.edu.cn`
[3] Academy for Engineering and Technology, Fudan University,
Shanghai 200433, China
`yzk@fudan.edu.cn`

Abstract. In this paper, we propose a non-local network routing (NNR) approach for perceptual image super-resolution. Unlike conventional methods which generate visually-faked textures due to exiting hand-designed losses, our approach aims to globally optimize both procedures of learning an optimal perceptual loss and routing a spatial-adaptive network architecture in a unified reinforcement learning framework. To this end, we introduce a reward function to teach our objective to pay more attention on the visual quality of the super-resolved image. Moreover, we carefully design an offset operation inside the neural architecture search space, which typically deforms the receptive field on boundary refinement in a non-local manner. Experimentally, our proposed method surpasses the perceptual performance over state-of-the-art methods on several widely-evaluated benchmark datasets.

Keywords: Image super-resolution · Neural architecture search · Deep learning · Reinforcement learning

1 Introduction

Single image super-resolution aims to recover a high-resolution (HR) image (Fig. 1(a)) for a given low-resolution (LR) image (Fig. 1(b)), which plays a key role in image enhancement [8,12,29]. Although numerous image super-resolution approaches have been proposed recently [4,9–11,25], the performance still remains unsatisfied in practice. This is because the high-frequency information from the high-resolution image is excessively missing when it degrades due to extreme illumination conditions, motion blur, etc. Hence, this motivates us to develop a robust super-resolution approach to particularly recover the high-frequency information and enhance the *visual quality* for the super-resolved static images.
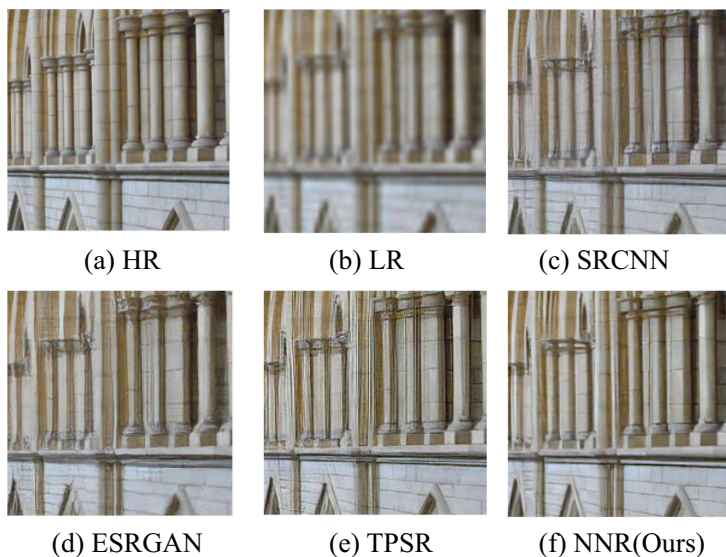
This is a student paper.

(a) HR   (b) LR   (c) SRCNN

(d) ESRGAN   (e) TPSR   (f) NNR(Ours)

Fig. 1. Image super-resolution results of different methods. The distortion-based method SRCNN [4] generates blurred textures. The perceptual-based methods including ESRGAN [25], TPSR [10] generate unnatural artifacts. Our method achieves to recover sharper boundary and finer textures visually compared with others.

Based on different types of optimization losses, the image super-resolution methods can be roughly divided into distortion-based [11,22,31] and perceptual-based [9,10,21,25]. Specifically, the distortion-based methods aim to generate high PSNR images which typically minimize the discrepancy between the super-resolved images and the ground truth images in a pixel-wise manner. One major issue in these methods is that the pixel-level reconstruction loss likely results in blurred textures, ignoring the high-frequency details (Fig. 1(c)). To address this issue, the perceptual-based methods have been proposed to improve the visual quality of the super-resolved images. For example, Wang *et al.* [25] developed a ESRGAN method, the generative models typically use perceptual loss and adversarial loss to improve the perceptual quality. However, these methods likely generate fake textures and unnatural artifacts when recovering super-resolved images (Fig. 1(d),(e)). The underlying reason is that the discriminator likely produces bias supervision signal during the optimization process, which hardly captures texture details accurately. Moreover, the existing loss functions (*e.g.*, perceptual loss [7], pixel-wise MSE loss) are hand-crafted which provide local perceptual supervision signals.

Besides a well-defined perceptual objective function, making full use of the self-similarity information in the image itself is also effective on improving the perceptual quality [18]. For example, Yang *et al.* [27] proposed to explicitly transfer similar high-frequency features from a given reference image, so that the produced textures are more reasonable rather than the conventional fake ones. How-

ever, the performance of this method is semantically sensitive to the reference, which degrades seriously when the irrelevant reference images are given. Besides, the local features are in fixed-location neighborhoods, which cannot adapt the spatial relevant textures. To fully exploit the global cues of the input image itself, we introduce an offset learning strategy, which takes in the non-local information by utilizing the self-similarity of the inputs. By doing this, we use feature similarity to propagate between non-local pixels to explore high-frequency information (such as edges). In parallel, it reduces the geometric distortions produced by GAN-based methods [9,25].

In this work, we argue to jointly optimize both procedures of learning an optimal perceptual loss function and searching a reliable network architecture, which can further improve the perceptual quality of the super-resolved images (Fig. 1 (f)). To achieve this, we propose a non-local network routing (NNR) method for perceptual image super-resolution. Specifically, we leverage the neural architecture search which optimizes using reinforcement-based algorithm. To improve the visual quality of the super-resolved images, we develop a learnable reward to optimize an optimal perceptual loss for image super-resolution. Moreover, we design an offset learning strategy to adaptively capture spatial boundary information in a non-local manner. Extensive experiments on the widely-used datasets demonstrate the effectiveness of the proposed method quantitatively and qualitatively.

2   Related Work

Single Image Super-Resolution. Low-resolution images are affected by many degradation factors during the imaging process, such as motion blur, noise and downsampling. Shallow single image super-resolution approaches can be roughly divided into two categories: interpolation-based [13,15], reconstruction-based [6, 19]. Interpolation-based methods recover high-resolution images by interpolation algorithm. For example, bicubic interpolation. However, these methods usually undergo accuracy shortcomings. To address this limitation, reconstruction-based methods have been proposed to adopt prior knowledge to restrict the possible solution space, which can restore sharp details. Nevertheless, these methods are usually time-consuming.

Recent years have witnessed that deep learning networks have been applied to address the nonlinear issue in image super-resolution [4,9,16,18,25], which learns a set of nonlinear mapping from low-resolution to high-resolution image in an end-to-end manner. The distortion-based methods aim to improve the fidelity of images which typically minimize the mean square error between the predicted pixel and the ground-truth pixel. For example, Dong *et al.* [4] proposed SRCNN, which is the first work that applies deep learning for image super-resolution. Mao *et al.* [16] proposed to use encoder-decoder design to super-resolve the image. Although these methods have achieved the promising performance, one major issue is that the pixel-wise loss results in smooth images due to a lack of high-frequency details. To address this issue, perceptual-based methods have

been proposed to improve the visual quality. For example, SRGAN [9] used an adversarial loss to restore the high-frequency details for perceptual satisfaction. However, the generative models likely produce geometric distortion textures. Besides, the hand-designed perceptual loss are not optimal for image perceptual evaluation and efficient training. To address these problems, our proposed method optimizes the procedures of learning an optimal perceptual objective function.

Neural Architecture Search. Recent trends have been seen that neural architecture search(NAS) [33] is gradually introduced to many computer vision applications. The coarse-grained tasks include image classification [20], object detection [34]. The fine-grained tasks include semantic segmentation [14], image super-resolution [3,10]. Auto-deeplab [14] proposed to search the network level structure in addition to the cell level structure, which aims to search the outer network structure automatically for semantic segmentation. Ulyanov *et al.* [23] argued that the structure of networks can be used as a structured image prior. Hence, Chen *et al.* [3] proposed to search for neural architectures which can capture stronger structured image priors for image restoration tasks. However, these methods mainly focus on searching for a network architecture and ignoring the image visual quality. Lee *et al.* [10] incorporated the NAS algorithm with GAN-based image super-resolution to improve the quality of perception while considering the computation cost. However, this method cannot fully exploit the global cues of image itself. To fully exploit the global cues of the input image itself, we exploit an offset learning strategy based on the self-similarity of images. Then we add the offset operation to the search space to further search for the perceptual-based super-resolution network. Besides, the GAN-based super-resolution method may likely produce fake textures duo to the unstable training. Thus our approach propose to optimize the perceptual loss function and perceptual-based super-resolution network simultaneously.

3   Methodology

In image super-resolution, we aim to restore a high-resolution image denoted by I^{SR} based on the given low-resolution input denoted by I^{LR}. As demonstrated in Fig. 2, we develop a non-local network routing method. Technically, we leverage reinforcement learning algorithms and incorporate neural architecture search with the image super-resolution task. Furthermore, we design a learnable perceptual reward as loss function to produce optimal supervision signal for efficient training. Besides, we develop a search space by introducing spatial-adaptive offset operation, which aims to reason a reliable network for perceptual image super-resolution.

3.1   Non-local Network Routing

Although traditional perceptual-based methods can significantly improve the perceptual quality of the super-resolved images, it will produce inconsistent artifacts and false textures. Moreover, the hand-designed perceptual loss function
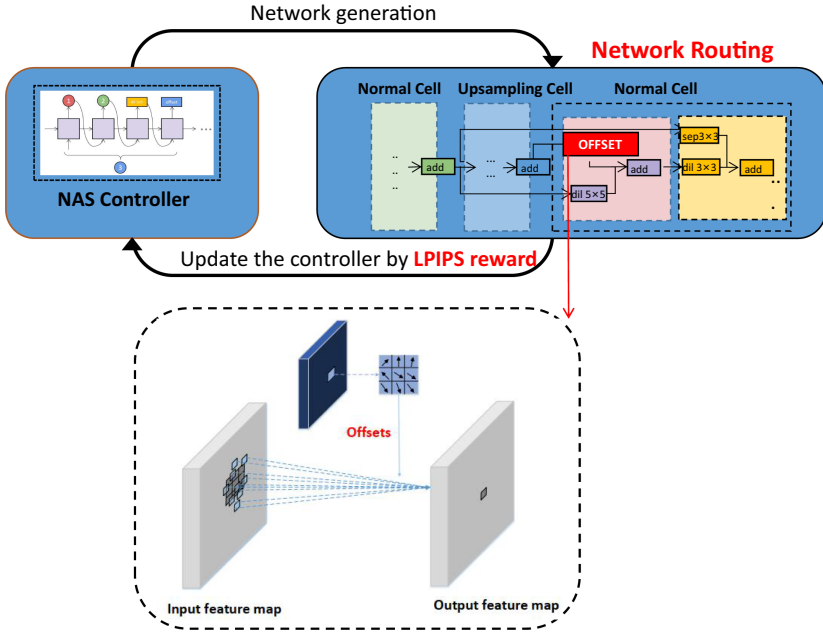
Fig. 2. Overall framework of our NNR method. We design a reward (LPIPS) and exploit neural architecture search algorithm (reinforcement learning-based with LSTM controller) beyond a search space. The offset operation is deployed in the search space to seek for a reliable network architecture.

is easy to fall into local optima and cannot be considered as a strong supervision signal to train the optimal super-resolution network. To address this, we introduce NAS into image super-resolution task. The search algorithm is mainly based on reinforcement learning, which incorporates with LSTM as the controller. The action a specifies the generation of a neural network architecture. The state s is defined by a set of observed network architecture. We design a learnable reward (LPIPS) to jointly optimize both procedures of learning an optimal perceptual loss and routing a reliable super-resolution network architecture denoted by $\boldsymbol{\omega}$. The LPIPS reward function is designed to measure the image patch similarity from feature space, which is defined as follows:

$$r(s^t, a^t) = -l_{\text{LPIPS}} \tag{1}$$

Specifically, we define the LPIPS function [30] by the following equation:

$$l_{\text{LPIPS}} = \sum_l \frac{1}{H_l W_l} \sum_{h,w} \|w_l \odot (\hat{I}_l^{\text{SR}} - \hat{I}_l^{\text{HR}})\|^2, \tag{2}$$

where $\hat{I}_l^{\text{SR}}, \hat{I}_l^{\text{HR}} \in \mathbb{R}^{H_l \times W_l \times C_l}$ is the feature from l layers of the pre-trained network. $w_l \in \mathbb{R}^{C_l}$ is used to scale the channel-wise activations .

The traditional PSNR is the distortion based metric, which is insufficient for estimating the visual quality of images. This is because the pixel-wise restraint results in over-smoothed results without sufficient high-frequency details. However, the LPIPS is a perceptual-based metric to measure the image patch similarity from feature space. It is mentioned in [30] that the perceptual similarity measurement of two images is more consistent with human judgment than PSNR. Therefore, we use the LPIPS as the perceptual reward to optimize an optimal perceptual loss.

Aside from learning an optimal perceptual loss function, we introduce an offset learning strategy to fully exploit the global cues of the input image itself. Moreover, the non-local feature representation is also effective on improving the perceptual quality of the super-resolved images. We explore the boundary information by the self-similarity of images. The captured high-frequency information such as spatial textures and edges reinforces the visual quality. In this way, the boundary information further resolves the geometric distortion. The offset strategy can be written as follows:

$$y(p_0) = \sum_{p_n \in R} w(p_n) \cdot x(p_0 + p_n + \triangle p_n), \tag{3}$$

where x is the input, y is the output feature map, and p_n enumerates the location in a regular grid R respectively. $\{\triangle p_n | n = 1, ..., N\}$ is the learnable offsets and the sampling performs on the offset locations $p_n + \triangle p_n$.

To automatically search for a network architecture with promising perceptual performance, we design to plug the offset operation inside the search space. The offset operation adaptively learns a set of offsets from the input image itself. Then our search space is developed to perform micro-cell approach and the normal cell can be regarded as a feature extractor. In our approach, we aim to obtain high-frequency feature representation which is crucial for perceptual image super-resolution. As a result, our model focuses on selecting for the best architecture of the normal cell. We show the candidate operation Op_normal of the normal cell search subspace as follows:

Op_normal = {Offset,
Dilated Conv (k, n) with $k = 3, 5$,
Separable Conv (k, n) with $k <= 3, 5$,
Residual Channel Attention Block(RCAB),
Identity }

For the normal cell, the search space is composed by the offset operation [32] and other several commonly used candidate operations including 3×3 and 5×5 dilated convolution, 3×3 and 5×5 separable convolution, residual channel attention block [31] and skip connection.

The upsampling cell is used to recover images with higher spatial resolution. We develop a search space with several upsampling operations. The candidate operation $Op\_upsampling$ of upsampling search subspace can be expressed as

---

Algorithm 1: NNR

Input: Training set: $(I_i^{\mathrm{LR}}, I_i^{\mathrm{HR}})_{i=1}^{N_1}$ with N_1 samples,
             Validation set: $(I_i^{\mathrm{LR}}, I_i^{\mathrm{HR}})_{i=1}^{N_2}$ with N_2 samples. $Inter = 300$.
Output: NNR model
1 **Require:** learning rate η, sharing parameters $\boldsymbol{\omega}$, controller parameters θ
2 //Initialize $\boldsymbol{\omega}$ and θ
3 **for** $i = 1$ *to Inter* **do**
4   **for** *each training sample* $i = 1, ... N_1$ **do**
5   //Sampling a group of network architectures;
6   //training $\boldsymbol{\omega}$ on training sample
7   Sampling $s \sim \pi(s, \theta\ )$;
8   Compute $l_{total}\ (s, \boldsymbol{\omega}\ )$ via (4)
9          $\boldsymbol{\omega} \leftarrow \boldsymbol{\omega}$-$\eta\ \bigtriangledown_{\boldsymbol{\omega}} l_{total}\ (s, \boldsymbol{\omega}\ )$
10   **end**
11   **for** *each validation sample* $j = 1, ... N_2$ **do**
12   //training controller parameters θ
13   //compute LPIPS reward on validation samples
14   Sampling $s \sim \pi(s, \theta\ )$;
15          $\theta \leftarrow \theta + \eta r(s)\ \bigtriangledown_\theta \log \pi(s; \theta\ )$
16   **end**
17 **end**

$$\text{Op\_upsampling} = \{\text{Pixel Shuffle Layer,}$$
$$\text{Deconvolution Layer}$$
$$\text{Nearest-neighbor Interpolation}$$
$$\text{Bilinear interpolation}\}$$

3.2   Model Learning

For our optimization, we specify a small scale of epochs and higher batch size as the proxy task. In detail, we first leverage the proxy task to search the optimal architecture. Then we exploit the weight sharing strategy, which uses the weights of step t to initialize the model at step $t + 1$. We evaluate the searched network architecture by computing the LPIPS reward between the ground truth image and the super-resolved image. With the learnable perceptual reward, we exploit the policy gradient [26] to train the LSTM controller. Based on the learned policy $\pi(\cdot)$, we obtain the best-performing network architectures and the optimal reward loss function simultaneously. Finally, we apply the full task to retrain the acquired best-performing super-resolution network architecture from scratch. We also use the LPIPS loss to train the searched super-resolution network. The previous works [2] mentioned that only using perceptual quality to constrain the network may produce undesirable artifacts. Hence, we incorporate ℓ_1 loss in our final optimization. The overall loss of training the searched network architecture

can be expressed as follows:

$$l_{total} = \alpha \frac{1}{N} \sum_{i=1}^{N} |I_i^{\text{SR}} - I_i^{\text{HR}}| + \beta l_{\text{LPIPS}}, \tag{4}$$

where α and β are the trade-off weights and N is the total number of images respectively. Specifically, we specified the parameters $\alpha=0.8$ and $\beta=0.2$ of our model.

Algorithm 1 details the training procedure of our NNR.

4   Experiments

4.1   Evaluation Dataset and Metric

In our experiments, we used the DIV2K dataset as the training data, and the commonly-used SR benchmarks, namely Set5 [1], Set14 [28], BSD100 [17] and Urban100 [5] as the testing datasets. All experiments were performed with a scale factor 4x between low-resolution and high-resolution images. For data augmentation, we used horizontal flip, verticle flip and rotation randomly.

We evaluated the trained model under the learned perceptual image patch similarity (LPIPS) and the peak signal-to-noise ratio (PSNR). Accordingly, we used LPIPS to measure the perceptual quality of the super-resolved images, where the lower the LPIPS value indicates better image visual quality. The PSNR is distortion-based measures that pays more attention to the fidelity of images. Obviously, the higher the PSNR value and the smaller the image distortion. Following the standard settings in [10], we evaluated PSNR and LPIPS on the Y channel and RGB image respectively.

4.2   Implementation Details

Our model was built based on the popular accelerated deep learning toolbox PyTorch[1]. We conducted all experiments on a NVIDIA Tesla V100 GPU with 300 epochs for searching network architectures and 300 epochs for training networks. The batch-size was set to 16. The ADAM optimizer was for searching and SGD for training. Moreover, we use sample entropy regularization for robust and fast convergence in our NAS controller.

### 4.3   Derived Architecture

Figure 3 shows the normal cell and upsampling cell searched via our method respectively. As the figure shows, each cell contains four intermediate nodes and every node has two operations from previous nodes. For each cell, the nodes represent the feature map, and the edge is the searched operation. It can be concluded that the cell structure selection is controlled by our proposed reward, which achieves the highest reward during the optimization iterations.

---

[1] https://github.com/pytorch/pytorch.

Table 1. Quantitative results of our method on different datasets including Set5, Set14, BSD100, and Urban100. Note that higher is better for PSNR, lower is better for LPIPS. Our method achieves compelling performance especially under the perceptual LPIPS metric.

Model	Set5 PSNR/LPIPS	Set14 PSNR/LPIPS	BSD100 PSNR/LPIPS	Urban100 PSNR/LPIPS
Bicubic	28.420/0.341	26.100/0.439	25.961/0.525	23.145/0.473
SRGAN [9]	29.168/0.088	26.171/0.166	25.459/0.198	24.397/0.155
ESRGAN [25]	30.454/0.075	26.276/0.133	25.317/0.161	24.360/0.123
SFTGAN [24]	29.930/0.089	26.223/0.148	25.505/0.177	24.013/0.143
NatSR [21]	30.991/0.094	27.514/0.176	26.445/0.211	25.464/0.150
TPSR [10]	29.600/0.076	26.880/0.110	26.230/0.116	24.120/0.141
Baseline	30.866/0.078	27.747/0.109	26.941/0.129	24.839/0.125
w/Reward	31.477/0.070	28.072/0.102	27.170/0.123	25.344/0.109
NNR (Ours)	31.427/0.065	28.074/0.095	27.183/0.117	25.248/0.108
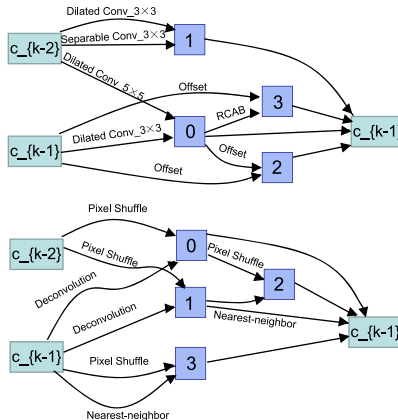
Fig. 3. Resulting routing of our NNR

4.4 Comparison with State-of-the-Art Methods

Quantitative Comparison. We compared our approach with folds of state-of-the-art perceptual driven super-resolution methods. In Table 1, we reported the PSNR, LPIPS on Set5, Set14, BSD100 and Urban100 under the evaluation setting. From the results, we made two-fold conclusion: (1) Our proposed learnable perceptual reward can capture sufficient high-frequency details which improves the visual quality of the super-resolved images. We also used the perceptual reward to route a reliable super-resolution network. In this manner, our method provides optimal supervision for texture recovery and trains the super-resolution network efficiently. However, the traditional perceptual loss is handcrafted which may be local for capturing high-frequency information. Meanwhile, the hand-designed super-resolution network architecture is redundant for

Fig. 4. Visualized results of our NNR versus conventional super-resolution approaches.

recovering visual quality. (2) Benefiting from the offset operation, our method can capture non-local similar feature representation and further improve the visual quality of the super-resolved images. Since the Urban100 dataset contains more building structure images, we observe that our NNR method generalizes better and gives performance gain 0.015 in LPIPS on the dataset. Unfortunately, some methods such as NatSR [21] and TPSR [10] which enforce local features, likely resulting in the inability to obtain global information. Thus, the LPIPS performance of these methods is poor than us. In addition, we figure out that our NNR achieves competitive performance on PSNR and better performance on LPIPS especially on the perceptual dimension.

Besides, we performed ablation study to validate the effectiveness for different components of our NNR method. Specifically, we only applied PSNR as the reward to search the super-resolution network as Baseline model. We first validated the importance of the designed learnable perceptual reward, which only used the LPIPS reward to constrain the network training in w/ Reward model. Then we added the offest operation to the search space based on the w/ Reward model, which further validates the effectiveness of the offset learning strategy. Table 1. presents the quantitative comparison of our ablation study.

We see that w/Reward has a significant improvement compared with Baseline model, which demonstrates the effectiveness of the learnable perceptual reward. The reason lies on that the optimal loss function efficiently dominates the optimal super-resolution network routing. Furthermore, our NNR method also indicates improvements especially on LPIPS performance over w/ Reward. With this, the offset operation adaptively captures the relevant features from complex images,

especially the spatial textures and edges. In summary, our results clarify the effectiveness integrated with both the learnable perceptual reward and offset learning strategy.

Qualitative Comparison. Finally, we compared our NNR method with other methods qualitatively. The traditional perceptual-based super-resolution methods produce inconsistent fake textures duo to bias supervision signals. As shown in Fig. 4, our method is more realistic than others. For example, for the stone statue's head, the compared methods are accompanied with unpleasant artifacts, while our method generates sharper textures. This is mainly due to the optimal perceptual loss and reliable super-resolution network architecture. Besides, the offset learning strategy achieves the non-local edge information, which reduces the geometric distortion and enhances the discriminability of the boundary and texture information.

5   Conclusion

In this paper, we have proposed a non-local network routing (NNR) method for perceptual image super-resolution. We have designed a learnbale reward to select a reliable super-resolution network architecture with an offset learning strategy. Quantitative and qualitative results have shown the effectiveness of our NNR. Differential convolutions in frequency domain with NAS is a desirable direction.

References

1. Bevilacqua, M., Roumy, A., Guillemot, C., Alberi-Morel, M.: Low-complexity single-image super-resolution based on nonnegative neighbor embedding. In: BMVC, pp. 1–10 (2012)
2. Chen, R., Xie, Y., Luo, X., Qu, Y., Li, C.: Joint-attention discriminator for accurate super-resolution via adversarial training. In: ACM MM, pp. 711–719 (2019)
3. Chen, Y., Gao, C., Robb, E., Huang, J.: NAS-DIP: learning deep image prior with neural architecture search. ECCV. **12363**, 442–459 (2020)
4. Dong, C., Loy, C.C., He, K., Tang, X.: Learning a deep convolutional network for image super-resolution. In: ECCV, vol. 8692, pp. 184–199 (2014)
5. Huang, J., Singh, A., Ahuja, N.: Single image super-resolution from transformed self-exemplars. In: CVPR, pp. 5197–5206 (2015)
6. Irani, M., Peleg, S.: Improving resolution by image registration. Graph. Models Image Process. **53**(3), 231–239 (1991)
7. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: ECCV, vol. 9906, pp. 694–711 (2016)

8. Kouame, D., Ploquin, M.: Super-resolution in medical imaging: an illustrative approach through ultrasound. In: ISBI, pp. 249–252 (2009)
9. Ledig, C., et al.: Photo-realistic single image super-resolution using a generative adversarial network. In: CVPR, pp. 105–114 (2017)
10. Lee, R., Dudziak, L., Abdelfattah, M.S., Venieris, S.I., Kim, H., Wen, H., Lane, N.D.: Journey towards tiny perceptual super-resolution. In: ECCV, vol. 12371, pp. 85–102 (2020)
11. Lim, B., Son, S., Kim, H., Nah, S., Lee, K.M.: Enhanced deep residual networks for single image super-resolution. In: CVPRW, pp. 1132–1140 (2017)
12. Lin, F., Fookes, C., Chandran, V., Sridharan, S.: Super-resolved faces for improved face recognition from surveillance video. In: Lee, S.-W., Li, S.Z. (eds.) ICB 2007. LNCS, vol. 4642, pp. 1–10. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-74549-5_1
13. Lin, T., RoyChowdhury, A., Maji, S.: Bilinear CNN models for fine-grained visual recognition. In: ICCV (2015)
14. Liu, C., et al.: Auto-deeplab: hierarchical neural architecture search for semantic image segmentation. In: CVPR, pp. 82–92 (2019)
15. Loop, C.T., Schaefer, S.: Approximating catmull-clark subdivision surfaces with bicubic patches. ACM Trans. Graph. **27**(1), 8:1–8:11 (2008)
16. Mao, X., Shen, C., Yang, Y.: Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. In: NIPS, pp. 2802–2810 (2016)
17. Martin, D.R., Fowlkes, C.C., Tal, D., Malik, J.: A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: ICCV, pp. 416–425 (2001)
18. Mei, Y., Fan, Y., Zhou, Y., Huang, L., Huang, T.S., Shi, H.: Image super-resolution with cross-scale non-local attention and exhaustive self-exemplars mining. In: CVPR, pp. 5689–5698 (2020)
19. Patti, A.J., Sezan, M.I., Tekalp, A.M.: Superresolution video reconstruction with arbitrary sampling lattices and nonzero aperture time. IEEE Trans. Image Process. **6**(8), 1064–1076 (1997)
20. Real, E., Moore, S., Selle, A., Saxena, S., Suematsu, Y.L., Tan, J., Le, Q.V., Kurakin, A.: Large-scale evolution of image classifiers. In: ICML, vol. 70, pp. 2902–2911 (2017)
21. Soh, J.W., Park, G.Y., Jo, J., Cho, N.I.: Natural and realistic single image super-resolution with explicit natural manifold discrimination. In: CVPR, pp. 8122–8131 (2019)
22. Tong, T., Li, G., Liu, X., Gao, Q.: Image super-resolution using dense skip connections. In: ICCV, pp. 4809–4817 (2017)
23. Ulyanov, D., Vedaldi, A., Lempitsky, V.S.: Deep image prior. In: CVPR (2018)
24. Wang, X., Yu, K., Dong, C., Loy, C.C.: Recovering realistic texture in image super-resolution by deep spatial feature transform. In: CVPR, pp. 606–615 (2018)
25. Wang, X., et al.: SRGAN: enhanced super-resolution generative adversarial networks. In: ECCVW, vol. 11133, pp. 63–79 (2018)
26. Williams, R.J.: Simple statistical gradient-following algorithms for connectionist reinforcement learning. Mach. Learn. **8**, 229–256 (1992)
27. Yang, F., Yang, H., Fu, J., Lu, H., Guo, B.: Learning texture transformer network for image super-resolution. In: CVPR, pp. 5790–5799 (2020)
28. Zeyde, R., Elad, M., Protter, M.: On single image scale-up using sparse-representations. Curves Surf. **6920**, 711–730 (2010)

29. Zhang, L., Zhang, H., Shen, H., Li, P.: A super-resolution reconstruction algorithm for surveillance images. Sig. Process. **90**(3), 848–859 (2010)
30. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: CVPR, pp. 586–595 (2018)
31. Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., Fu, Y.: Image super-resolution using very deep residual channel attention networks. In: ECCV, vol. 11211, pp. 294–310 (2018)
32. Zhu, X., Hu, H., Lin, S., Dai, J.: Deformable convnets V2: more deformable, better results. In: CVPR, pp. 9308–9316 (2019)
33. Zoph, B., Le, Q.V.: Neural architecture search with reinforcement learning. In: ICLR (2017)
34. Zoph, B., Vasudevan, V., Shlens, J., Le, Q.V.: Learning transferable architectures for scalable image recognition. In: CVPR, pp. 8697–8710 (2018)