Deep Learning Based Medical Image Super Resolution Reconstruction

Jing Xu 36920221153133,[1] Jieai Mai 36920221153107,[1] Yajun Jian 36920221153086, [1] Yi Zheng 36920221153154, [1] Dandan Shan 36920221153075 [1]

[1] Affiliation 1
[2] Affiliation 2
firstAuthor@affiliation1.com, secondAuthor@affilation2.com, thirdAuthor@affiliation1.com

Abstract

The clarity of medical images directly affects clinical diagnosis. Due to the limitations of imaging equipment and environmental factors, high-resolution images cannot be obtained directly. In recent years, deep neural networks have achieved excellent results in the field of super-resolution reconstruction. Therefore, a deep neural network model is proposed to reconstruct high-resolution medical images.

Introduction

In recent years, image super-resolution (SR) has attracted more and more attention in the research field. Super resolution aims to convert a given low resolution image with coarse details into a corresponding high resolution image with better visual quality and fine details(Yang et al. 2019)(Anwar, Khan, and Barnes 2020). Image super-resolution is also referred to as other names, such as image scaling, interpolation, upsampling, scaling, and magnification. You can use a single image or multiple images to perform the process of generating raster images with higher resolution. For practical reasons, this paper focuses on single image super resolution (SISR), which has been widely studied because of its challenges. Super resolution is a classical problem. For several reasons, it is still regarded as a challenging and open research problem in computer vision. First, SR is an ill posed inverse problem, namely, underdetermined case. For the same low resolution image, there are multiple solutions, rather than a single solution. To limit the solution space, reliable prior information is often required. Secondly, the complexity of the problem increases with the increase of the upgrade factor.

Superresolution methods can be roughly divided into two categories: traditional methods and depth learning methods. Traditional super-resolution methods include interpolation-based, reconstruction-based, and learning-based methods. Interpolation-based super-resolution reconstruction methods, as the basic methods of super-resolution reconstruction, mainly include linear interpolation and nonlinear interpolation. The interpolation algorithm is to calculate the value of the relevant pixel point of the HR map from the known value of the pixel point on the LR map according to the interpolation formula. Linear interpolation includes nearest neighbor interpolation, bilinear interpolation, and bicubic interpolation. In the reconstructed SR, the complex preconditions are used, the possible solution space is limited, and it has good flexibility. The performance of many reconstruction-based methods degrades dramatically when the scaling factor is increased, and such methods are generally very time-consuming(Sun, Xu, and Shum 2008). Learning-based methods have been widely studied due to their excellent performance(Chang, Yeung, and Xiong 2004). Usually, machine learning algorithms are used to analyze the statistical relationship between LR and corresponding HR from a large number of training examples for image reconstruction. Based on the learning method, a dataset is usually produced, and then feature learning is performed on the dataset, and image reconstruction is performed using the learned parameters.

The classical algorithm has existed for decades, but it performs better in similar algorithms based on deep learning. Therefore, the latest algorithms rely on data-driven depth learning models to reconstruct the details required for accurate superresolution.

Deep learning (DL) (LeCun, Bengio, and Hinton 2015) is a branch of machine learning algorithm, which aims to learn the hierarchical representation of data. Deep learning shows outstanding advantages over other machine learning algorithms in many aspects such as computer vision (Krizhevsky, Sutskever, and Hinton 2017), speech recognition (Hinton et al. 2012) and natural language processing (Collobert and Weston 2008). In general, DL's strong ability to handle large amounts of unstructured data can be attributed to two major contributors: the development of efficient computing hardware and advanced algorithms. In general, DL's strong ability to handle large amounts of unstructured data can be attributed to two major contributors: the development of efficient computing hardware and advanced algorithms.Through literature review, it can be known that the methods of image super-resolution based on deep leaning can be divided into four categories, namely classical methods, supervised learning-based methods,unsupervised learning-based methods and domain-specific DR methods. Some of the more novel and well-known models are as follows : the enhanced deep SR network (EDSR), cycle-

Copyright © 2023, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

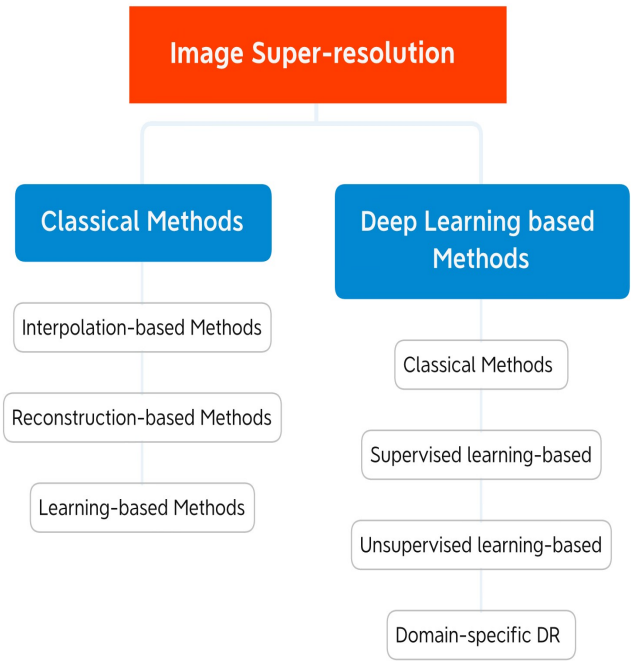

Figure 1: Classification of super-resolution methods

in-cycle GAN (CinCGAN), multiscale residual network (MSRN), meta residual dense network (Meta-RDN), recurrent back-projection network (RBPN),etc.

Considering the characteristics of data, we focuses on the methods based on supervised learning. The first is upsampling which is essential in deep learning-based SR methods such as its positioning, and the method performed for upsampling has a significant impact on the training and test performance of the model. The network design and advancements in design architecture are recent trends in deep learning. Some learning frameworks worth considering include Recursive learning,Residual learning, Dense connection-based learning, Multi-path learning,Advanced convolution-based learning, Attention-based learning, etc. In addition, unsupervised learning methods include Weakly-supervised super-resolution, Cyclic weakly-supervised SR and so on.

Related Work

Nowadays researchers have proposed a variety of super-resolution models with deep learning(Wang, Chen, and Hoi 2021). Compared with traditional methods, deep learning-based models show significant performance improvement in SISR tasks. Dong et al. (Dong et al. 2014),(Dong et al. 2015) first adopt the pre-upsampling SR framework and propose SRCNN to learn an end-to-end mapping from interpolated LR images to HR images. These models can take interpolated images with arbitrary sizes and scaling factors as input, and give refined results with comparable performance to single-scale SR models (Kim, Lee, and Lee 2016). Thus it has gradually become one of the most popular frameworks (Tai et al. 2017), (Tai et al. 2017), (Tai et al. 2017), (Tai et al. 2017). However, the predefined upsampling often introduce side effects (e.g., noise amplification and blurring), and since most operations are performed in high-dimensional space, the cost of time and space is much higher than other frameworks (Dong, Loy, and Tang 2016), (Shi et al. 2016).

In order to improve the computational efficiency and make full use of deep learning technology to increase resolution automatically, researchers propose to perform most computation in low-dimensional space by replacing the predefined upsampling with end-to-end learnable layers integrated at the end of the models (Dong, Loy, and Tang 2016), (Shi et al. 2016). Although post-upsampling SR framework has immensely reduced the computational cost, it still has some shortcomings. On the one hand, the upsampling is performed in only one step, which greatly increases the learning difficulty for large scaling factors (e.g., 4, 8). On the other hand, each scaling factor requires training an individual SR model, which cannot cope with the need for multi-scale SR. To address these drawbacks, a progressive upsampling framework is adopted by Laplacian pyramid SR network (LapSRN) (Lai et al. 2017).By decomposing a difficult task into simple tasks, the models under this framework greatly reduce the learning difficulty. However, these models also encounter some problems, such as the complicated model designing for multiple stages and the training stability, and more modelling guidance and more advanced training strategies are needed.

For the purpose of better capturing the mutual dependency of LR-HR image pairs, an efficient iterative procedure named back-projection (Irani and Peleg 1991) is incorporated into SR (Timofte, Rothe, and Van Gool 2016). This SR framework, namely iterative up-and-down sampling SR , tries to iteratively apply back-projection refinement, i.e., computing the reconstruction error then fusing it back to tune the HR image intensity. Specifically, Haris et al. (Haris, Shakhnarovich, and Ukita 2018) exploit iterative up-and-down sampling layers and propose DBPN, which connects upsampling and downsampling layers alternately and reconstructs the final HR result using all of the intermediately reconstructions. Similarly, the SRFBN (Li et al. 2019) employs a iterative up-and-down sampling feedback block with more dense skip connections and learns better representations. And the RBPN (Haris, Shakhnarovich, and Ukita 2019) for video super-resolution extracts context from continuous video frames and combines these context to produce recurrent output frames by a back-projection module.

Method

FMEN

Considering the goal of SR is to recover the lost highfrequency details (e.g., edges, textures), this paper propose a highfrequency attention block (HFAB) which learns an attention map with special focus on the high-frequency area. Specifically, the attention branch is designed in HFAB from local and global perspectives. This paper stack highly efficient operators like 3×3 convolution and Leaky ReLU layers sequentially for modeling the relationship between local signals. Batch Normalization (BN) is injected into HFAB to capture global context during training, while merged into

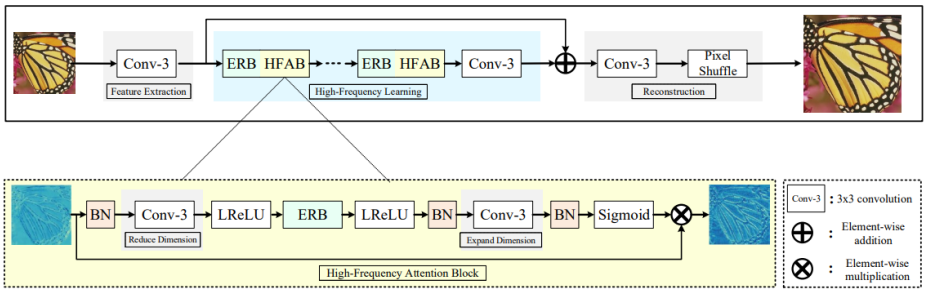

Figure 2: The overall architecture of FMEN

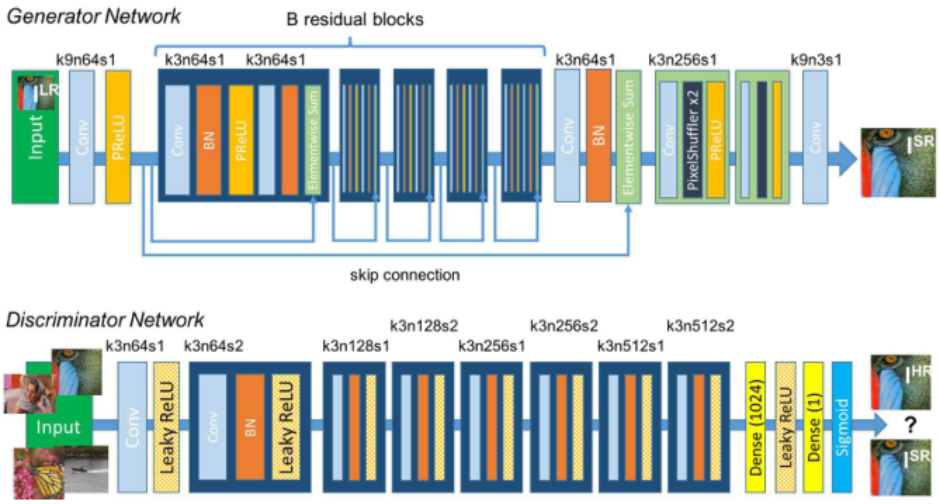

Figure 3: The overall architecture of SRGAN

convolution during inference.

Furthermore, they tailor the residual block (RB) and introduce an enhanced residual block (ERB), where the features are extracted in higher dimensional space during training and the skip connections are removed during inference using structural reparameterization technique.

By applying ERB and HFAB in a sequential and alternative way, this paper construct an efficient network, namely fast and memory-efficient network (FMEN), which demonstrates the clear advantage over existing EISR methods in terms of runtime and peak memory consumption when maintaining the same level of restoration performance, and the overall architecture is shown in Figure 2.

SRGAN

Ledig et al. proposed SRGAN (GANforSR) (Ledig et al. 2017) to recover high-frequency details of 4-fold upsampling factor images, and is the first GAN model of SRResNet as a generative network for hyper-segmentation, focusing on human perception visually high-resolution (photo-realistic). As shown in Figure 3, SRGAN consists of G (Generator) net and D (Discriminator) net.G net is a deep residual network that trains the generator to generate HR images from the input LR images, and since the deep network is difficult to train, skipconnection is introduced between different modules to improve the accuracy of the network. The D-network determines whether the input image is generated by the G-network or the real image in the database.The D-network uses LeakeyReLU as the activation function and avoids the use of max-pooling. The number of layers in the network ranges from 64 to 512, followed by two fully connected layers and a sigmoid layer, which are used to determine the probability of whether it is the same image. The network can be used for SR.1 when the game equilibrium is reached between the G and D networks.

SDSR

(Kim, Lee, and Lee 2016) proposed a high precision single image super-resolution method. Inspired by the image classification network VGGNet, the author uses a very deep convolution network to apply to the super-resolution field, and the final neural network model depth reaches 20 layers. The author found that by multiple convolutions in the deep network structure, the receptive field in the training process can be effectively expanded, and the context information in large image areas can be effectively used. The author introduces residual learning to improve the slow convergence speed of deep network by only learning residual and using

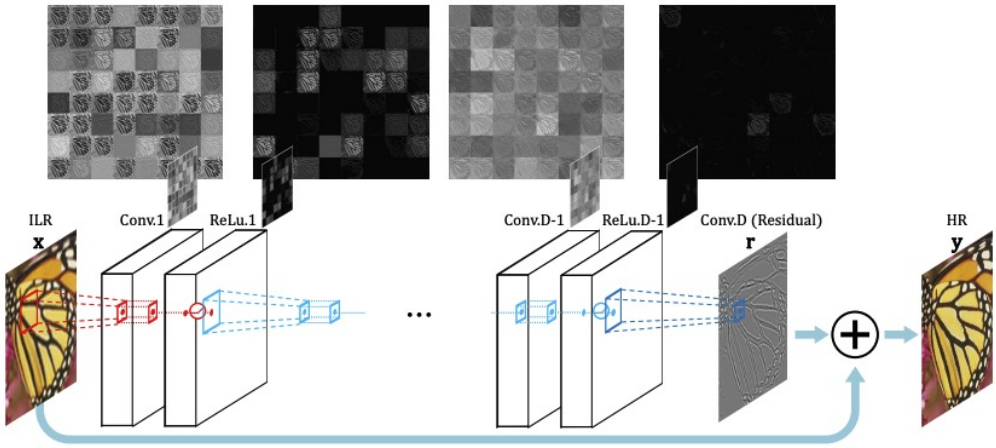

Figure 4: The Architecture of SDSR

high learning rate. The proposed method is better than the existing methods in both accuracy and visual improvement.

The author uses convolution with input channel of 1, output channel of 64, convolution core size of 3, input channel of 64, output channel of 1, and convolution core size of 3 at the beginning and end respectively. The same block with 18 layers is used in the middle layer. The block includes convolution layer with input and output channels of 64, convolution layer with convolution core size of 3, and RELU layer. That is to say, the number of channels of the feature map in the whole learning process is 64.

Because the size of the image will become smaller after the convolution operation when the image is not filled, the author uses padding=0 to fill the image to ensure that the size of the output image is consistent with the size of the input image, which solves the problem that the output image of SRCNN is smaller than the input image.

The author introduces residual learning. First of all, LR and HR share a lot of basic information (low-frequency structural information), so we only need to learn the difference between LR and HR (high-frequency information), which is called residual learning. It is obvious that the traditional learning method (SRCNN) is characterized by learning the complete result graph (SR) directly from LR, which is significantly stronger than learning only part of high-frequency information (i.e. residual learning), so residual learning is superior to previous models in terms of difficulty and learning time cost. Finally, the learned high-frequency information and LR (low-frequency information) can be integrated to obtain a result map SR close to the target HR.Figure 4 shows the architecture of SDSR.

Experiments

Datasets

Due to the limitation of hardware conditions, we choose BSD200 and General100 as the training set, which contain a relatively small number of images. BSD200 has 200 RGB three-channel images. General100 includes 100 RGB color three-channel images. The test data set uses Set5, which contains 5 color three-channel images.

Metric

We used two commonly used metrics, peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM).

- PSNR. The peak signal-to-noise ratio, which is the ratio of the energy of the peak signal to the average energy of the noise, is usually expressed as the log of dB.The formula is shown below:

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}\|I(i,j) - K(i,j)\|^2 \quad (1)$$

$$PSNR = 10 \cdot \log_{10}\left(\frac{MAX_I^2}{MSE}\right) \quad (2)$$

  Where I and K are clean image and noisy image with size m×n respectively. MAX_I is the maximum possible pixel value of the picture.
- SSIM. The formula evaluates the similarity of two images X and Y based on the following three scales.
  - Brightness. Measured as average gray, it can be calculated by averaging the value of all pixels.

$$l(X,Y) = \frac{2\mu_X\mu_Y + c_1}{\mu_X^2 + \mu_Y^2 + c_1} \quad (3)$$

  - Contrast. Measured by standard deviation of gray scale.

$$c(X,Y) = \frac{2\sigma_X\sigma_Y + c_2}{\sigma_X^2 + \sigma_Y^2 + c_2} \quad (4)$$

  - Structure. Measured by correlation coefficient.

$$s(X,Y) = \frac{\sigma_{XY} + c_3}{\sigma_X\sigma_Y + c_3} \quad (5)$$

    Where μ_X, μ_Y is the mean value of image X and Y, respectively. σ_X^2, σ_Y^2 is the variance of image X and Y,

respectively. σ_{XY} is the covariance of X and Y. c_1 = $(k_1 L)^2$ and c_2 = $(k_2 L)^2$ are two constants, where L is the range of pixel values, k_1 is set to 0.01 and k_2 is set to 0.03 by default. c_3 = c_2 / 2.

The formula of SSIM is shown below:

$$SSIM(X,Y) = \left[l(X,Y)^{\alpha} \cdot c(X,Y)^{\beta} \cdot s(X,Y)^{\gamma}\right] \tag{6}$$

Where α, β and γ represent the proportion of different features in SSIM. When α, β and γ are all 1, we have the common formula below:

$$SSIM(X,Y) = \frac{(2\mu_X\mu_Y + c_1)(2\sigma_{XY} + c_2)}{(\mu_X^2 + \mu_Y^2 + c_1)(\sigma_X^2 + \sigma_Y^2 + c_2)} \tag{7}$$

Results

Table 1 shows the size of input images and output images of the three neural networks during training. Table 2 shows the number of parameters of VDSR, SRGAN and FMEN networks. Figure 7 is a test of two images from Set5, both of which are downsampled to x4. These two down sampled images are used as input images. Figures 5 and 6 show the results of super resolution reconstruction of LR images by three kinds of neural networks. It can be seen from the results that as a high efficiency super resolution method proposed in 2022, the network parameters of FMEN are only half of those of VDSR and less than 1/10 of those of SRGAN. However, the quality of its reconstructed high resolution images is not inferior to that of VDSR and SRGAN.

Figure 5: baby

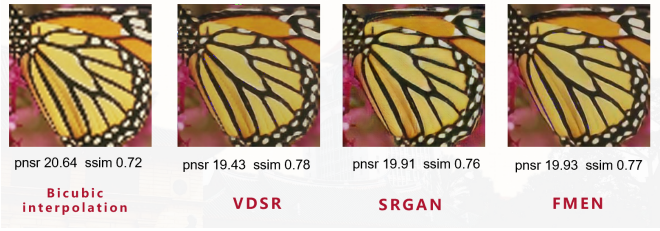

Figure 6: butterfly

Conclusion

The experimental hardware of this experiment is limited, and it is not trained in the commonly used large DIV2K dataset. The training dataset used is less, and the reconstruction results are also different from the experimental results of related papers. We also used very few test datasets. But we are very happy that we have completed this assignment and learned a lot of new knowledge. Our future work is to train on a larger dataset, use more test data, explore more effective super-resolution methods and innovate.

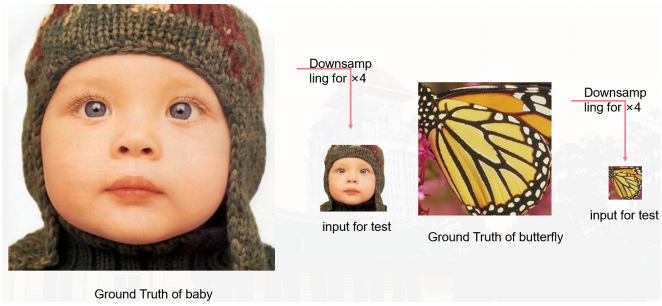

Figure 7: Ground truth and inputs

Table 1

Network	input size	output size
FMEN	32 × 32 × 3	128 × 128 × 3
VDSR	41 × 41 × 3	41 × 41 × 3
SRGAN	24 × 24 × 3	96 × 96 × 3

References

Anwar, S.; Khan, S.; and Barnes, N. 2020. A deep journey into super-resolution: A survey. *ACM Computing Surveys (CSUR)*, 53(3): 1–34.

Chang, H.; Yeung, D.-Y.; and Xiong, Y. 2004. Super-resolution through neighbor embedding. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, volume 1, I–I. IEEE.

Collobert, R.; and Weston, J. 2008. A unified architecture for natural language processing: Deep neural networks with multitask learning. In *Proceedings of the 25th international conference on Machine learning*, 160–167.

Dong, C.; Loy, C. C.; He, K.; and Tang, X. 2014. Learning a deep convolutional network for image super-resolution. In *European conference on computer vision*, 184–199. Springer.

Dong, C.; Loy, C. C.; He, K.; and Tang, X. 2015. Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2): 295–307.

Dong, C.; Loy, C. C.; and Tang, X. 2016. Accelerating the super-resolution convolutional neural network. In *European conference on computer vision*, 391–407. Springer.

Haris, M.; Shakhnarovich, G.; and Ukita, N. 2018. Deep back-projection networks for super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1664–1673.

Table 2

Network	parameterss
FMEN	341,066
VDSR	667,008
SRGAN	5,949,644

Haris, M.; Shakhnarovich, G.; and Ukita, N. 2019. Recurrent back-projection network for video super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3897–3906.

Hinton, G.; Deng, L.; Yu, D.; Dahl, G. E.; Mohamed, A.-r.; Jaitly, N.; Senior, A.; Vanhoucke, V.; Nguyen, P.; Sainath, T. N.; et al. 2012. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal processing magazine*, 29(6): 82–97.

Irani, M.; and Peleg, S. 1991. Improving resolution by image registration. *CVGIP: Graphical models and image processing*, 53(3): 231–239.

Kim, J.; Lee, J. K.; and Lee, K. M. 2016. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1646–1654.

Krizhevsky, A.; Sutskever, I.; and Hinton, G. E. 2017. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6): 84–90.

Lai, W.-S.; Huang, J.-B.; Ahuja, N.; and Yang, M.-H. 2017. Deep laplacian pyramid networks for fast and accurate super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 624–632.

LeCun, Y.; Bengio, Y.; and Hinton, G. 2015. Deep learning. *nature*, 521(7553): 436–444.

Ledig, C.; Theis, L.; Huszár, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z.; et al. 2017. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4681–4690.

Li, Z.; Yang, J.; Liu, Z.; Yang, X.; Jeon, G.; and Wu, W. 2019. Feedback network for image super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 3867–3876.

Shi, W.; Caballero, J.; Huszár, F.; Totz, J.; Aitken, A. P.; Bishop, R.; Rueckert, D.; and Wang, Z. 2016. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1874–1883.

Sun, J.; Xu, Z.; and Shum, H.-Y. 2008. Image super-resolution using gradient profile prior. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, 1–8. IEEE.

Tai, Y.; Yang, J.; Liu, X.; and Xu, C. 2017. Memnet: A persistent memory network for image restoration. In *Proceedings of the IEEE international conference on computer vision*, 4539–4547.

Timofte, R.; Rothe, R.; and Van Gool, L. 2016. Seven ways to improve example-based single image super resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1865–1873.

Wang, Z.; Chen, J.; and Hoi, S. C. H. 2021. Deep Learning for Image Super-Resolution: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(10): 3365–3387.

Yang, W.; Zhang, X.; Tian, Y.; Wang, W.; Xue, J.-H.; and Liao, Q. 2019. Deep learning for single image super-resolution: A brief review. *IEEE Transactions on Multimedia*, 21(12): 3106–3121.