Proceedings of the 4th International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/46/20241164

# Enhancing medical image classification based on super-resolution techniques: A comparative study

**Xiangning Deng**

Department of Computer Science, Beijing Normal University-Hong Kong Baptist University United International College, Zhuhai, 519000, China

q030026027@mail.uic.edu.cn

**Abstract.** This paper explores the application of super-resolution techniques in the context of medical image classification, focusing on the impact of various super-resolution models on classification model performance. With the burgeoning influence of deep learning in the field of medical image analysis. This research discusses the potential of image super-resolution to enhance the precision of medical image categorization. This paper compares three super-resolution methods: bicubic interpolation, Super Resolution Residual network (SRResnet), and Super Resolution Generative Adversarial Networks (SRGAN), using the OCT2017 dataset for optical coherence tomography (OCT) images. The experiments reveal that deep learning-based super-resolution methods surpass bicubic interpolation in boosting classification model accuracy. Notably, SRResnet stands out as the preferred choice for enhancing classification accuracy, despite yielding less visually appealing results compared to SRGAN. This suggests that while Generative Adversarial Networks (GANs) and perceptual loss functions hold promise in super-resolution, their translation to improved classification model performance may vary. This paper's findings provide guidance for practitioners and researchers in the field, emphasizing that selecting an appropriate super-resolution method could optimize classification models for medical image analysis.

**Keywords:** Super Resolution, Image Classification, SRGAN, ResNet.

## 1. Introduction

With the ongoing advancements in neural network technologies, the application of deep learning has expanded across various domains, addressing a multitude of challenges [1-3]. A prominent instance is observed in the realm of modern medicine, where medical images play a pivotal role by furnishing clinicians and medical experts with critical insights crucial for diagnosis, treatment, and disease monitoring. The integration of neural networks empowers computers to effectively categorize medical images. Nonetheless, the realm of medical imaging grapples with challenges arising from equipment limitations and image acquisition procedures, leading to issues such as inadequate resolution and indistinct details. These factors can potentially compromise the precision and efficacy of clinical decision-making. Studies demonstrated that images marked by blurriness, low contrast, or other deficiencies inherently undermine the accuracy of classification models [4], casting a shadow of uncertainty over the machine's capability to accurately discern objects.

© 2024 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/46/20241164

One approach to reducing the negative impact of blurry images on the performance of classification models is to eliminate samples that are too blurry (i.e., low-quality samples) and only train with clearer samples. The core idea of this method is to establish indicators to evaluate the degree of image blurriness and to find a threshold for removing samples through experiments. In a pertinent study [5], the author employed edge-detection operators like Sobel and Scharr to identify boundaries within original X-ray images. The resultant images were then utilized to train a shallow Convolutional Neural Network (CNN) to distinguish between clear and poor-quality images. The classification model's accuracy experienced a notable upswing by removing low-quality samples. However, given the challenges in procuring medical image samples, an excessive purge of data could potentially exacerbate the risk of model overfitting, thereby undermining rather than augmenting model performance. The wholesale removal of low-quality samples might not constitute the optimal data pre-processing approach. Instead, this study posits that the realm of image super-resolution, a subfield of computer vision, could offer a more efficacious route for enhancing medical image classification. Recent years have witnessed a rapid evolution of deep learning techniques, providing potent tools for image super-resolution. Models like Super-Resolution Generative Adversarial Networks (SRGAN) [6] have harnessed adversarial training and perceptual loss functions to generate images with improved resolution and quality. The objective of this study centers on elevating the performance of medical image super-resolution models by embracing super-resolution methodologies in lieu of conventional pre-processing paradigms.

This paper embarks on an exploration of diverse super-resolution models' influence on the classification performance of medical images. Within the context of the SRGAN [6] model, the introduction of a perceptual-based loss function as a replacement for the conventional Peak Signal-to-Noise Ratio-based loss function holds significance. Although SRGAN [6] might trail behind the Super Resolution Residual network (SRResNet) [6] according to traditional image metrics, its visually appealing output captivates the human eye with enhanced clarity. This investigation endeavors to ascertain whether such visually refined images correlate with improved model performance in classification tasks. Additionally, a compelling avenue for exploration rests in evaluating whether generative adversarial networks outshine SRResNet, devoid of confrontational components, in the context of image pre-processing for classification tasks. The experimental structure this study promotes is explained as follows: As the cornerstone of medical image classification, the Residual network (ResNet) [7] is adopted. In the super-resolution paradigm, fidelity to the original author's directives prompts the utilization of bicubic interpolation to curtail dataset resolution. The approach entails training ResNet [7] using low-resolution images, subsequently generating high-resolution counterparts, and juxtaposing them with original images within the dataset for conclusive comparisons.

## 2. Method

### *2.1. Introduction to SRResNet*

The SRResNet shown in Figure 1 is a fully convolutional network designed for 4x super-resolution. As indicated in the name, it incorporates residual blocks with skip connections to increase the optimization ability of the network despite its significant depth [6].

DOI: 10.54254/2755-2721/46/20241164

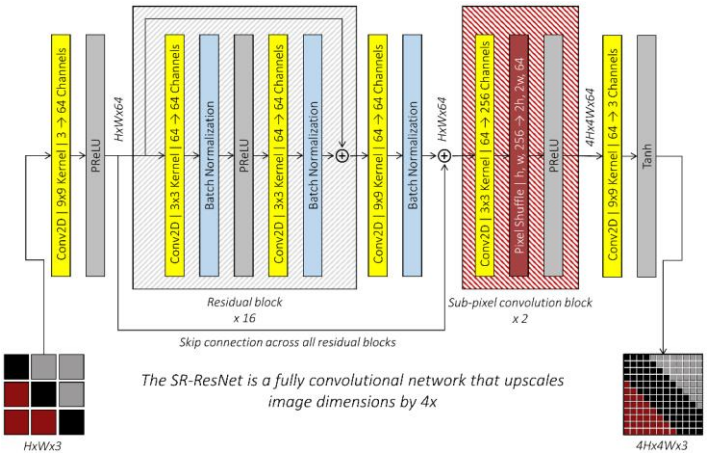



**Figure 1.** SRResNet Architecture.

*2.2. Introduction to SRGAN*

The reconstructed outputs generated by SRResNet exhibit a favorable signal-to-noise ratio, yet they lack high-frequency details and tend to display overly smoothed textures. In contrast, SRGAN [6] shown in Figure 2 has been crafted with the aim of producing high-resolution reconstructions that bear a strong resemblance to the original high-resolution images. It does so by focusing on replicating not only the low-level pixel details but also the high-level abstract features, capturing the overall essence and style of the images. This approach utilizes Generative Adversarial Networks (GANs) to provide a robust framework for creating lifelike images that are visually convincing and exhibit exceptional perceptual quality. The GAN methodology propels the reconstructions towards regions within the search space that have a high probability of containing photorealistic images, thus positioning them closer to the natural image manifold. The Generator is identical to the SRResNet in architecture, and the discriminator is a convolutional network that functions as a binary image classifier.

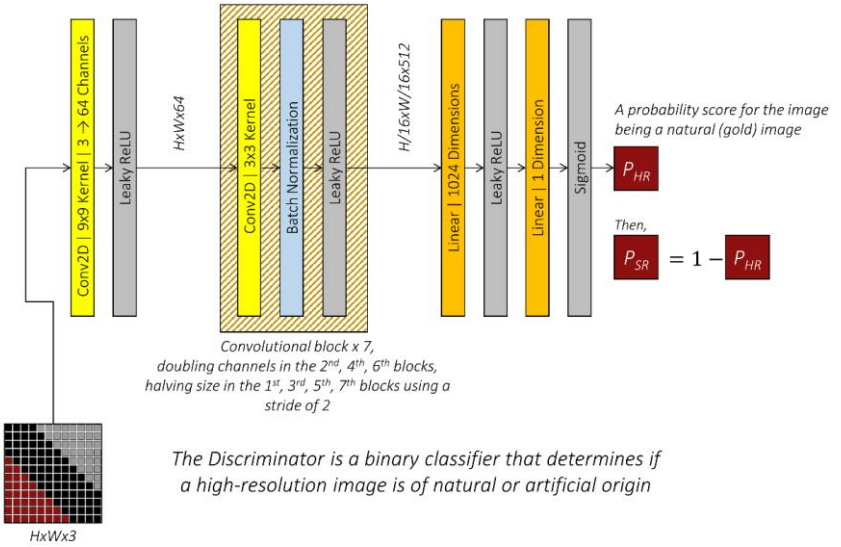



**Figure 2.** SRGAN Architecture.

*2.3. Perceptual loss function*

Most image super-resolution algorithms use mean squared error (MSE) as a loss function for optimization, which can result in images with a high signal-to-noise ratio but few high-frequency features [6].

DOI: 10.54254/2755-2721/46/20241164

$$l_{MSE}^{SR} = \frac{1}{r^2WH}\sum_{x=1}^{rW}\sum_{y=1}^{rH}(l_{x,y}^{HR} - G_{\theta_G}(I^{LR})_{x,y})^2 \quad (1)$$

Thus, SRGAN [6] integrates the VGG loss, which depends on the Rectified Linear Unit (ReLU) activation layer of the pre-trained 19-layer Visual Geometry Group Network (VGG network), as explained by Simonyan and Zisserman [7]. To be more precise, it utilizes the feature map produced by the $i^{th}$ convolution layer (after activation) that precedes the $i^{th}$ maximum pooling layer within the VGG19 network as a point of reference. The VGG loss is defined as the Euclidean distance between the reconstructed images $G_{\theta_G}(I^{LR})_{x,y}$ and the feature representation of the reference image $l_{x,y}^{HR}$. This loss function serves the purpose of preserving high-frequency details within the reconstructed images.

### *2.4. Introduction to ResNet*

ResNet, introduced by Kaiming He et al. in their ground-breaking paper [8] in 2015, aimed to tackle the challenge of training extremely deep neural networks. The motivation behind ResNet stemmed from the observation that as network depth increased, performance reached a saturation point and then deteriorated. This phenomenon, known as the vanishing gradient problem, arises when gradients become exceedingly small, hindering effective learning by the network.

To address the vanishing gradient problem, ResNet introduced a novel architectural paradigm centered around the concept of residual learning. The pivotal concept was the inclusion of skip connections, also known as shortcut connections, which facilitated the direct flow of information from earlier layers to deeper layers. These skip connections empowered the network to learn residual mappings rather than complete mappings, streamlining the optimization process and enabling the acquisition of deep representations. At the heart of residual connections lies the idea that, instead of explicitly learning the mapping from input to the desired output, the network learns the residual mapping—the disparity between the desired output and the input. By harnessing residual connections, the network is encouraged to focus on capturing the residual information, which can be perceived as the intricate details or alterations necessary to attain the desired output. A residual connection simply constitutes a direct link between the input of a block and the output of that block. Occasionally, these residual connections may include additional layers, but most often, they do not. The below Figure 3 shows an example block featuring an identity residual connection.

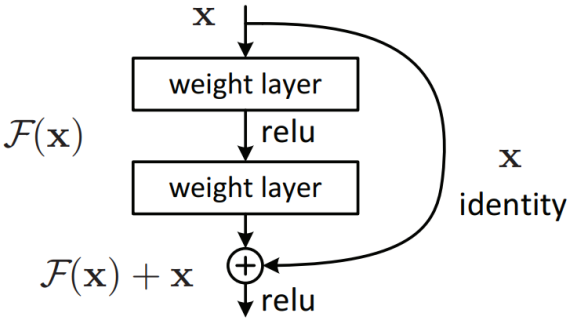



**Figure 3.** The skip connection.

ResNet configurations are commonly identified by their total layer count, labeled as ResNet18, ResNet34, ResNet50, ResNet101, and ResNet152.

Upon a more detailed examination of the table presented, both ResNet18 and ResNet34 initiate with an initial block consisting of two 3x3 convolutional layers, each of which is outfitted with 64 filters. The key distinction here is that ResNet18 includes two such blocks in its first layer, while ResNet34 integrates three of them. In contrast, ResNet50, ResNet101, and ResNet152 employ a distinct block structure known as “bottleneck” blocks. These bottleneck blocks involve an initial reduction in the number of channels within the input, followed by a subsequent expansion. The accompanying Figure 4

DOI: 10.54254/2755-2721/46/20241164

illustrates the standard "BasicBlock" (on the left), employed by ResNet18 and ResNet34, alongside the "Bottleneck" block used by ResNet50, ResNet101, and ResNet152.

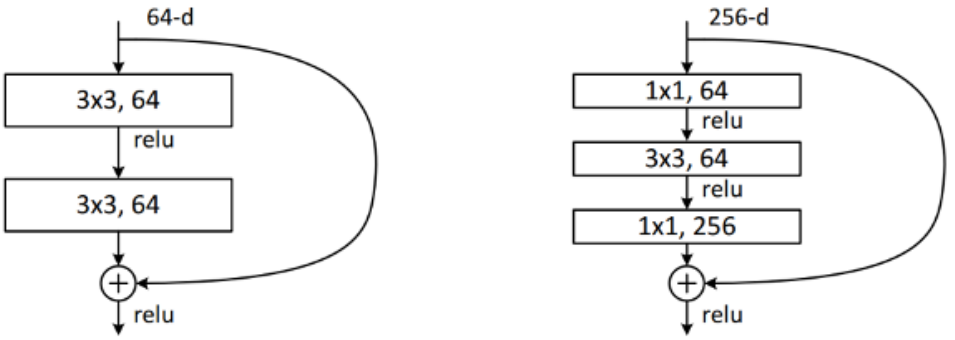



**Figure 4.** Basic Block and Bottleneck Block.

## 3. Results and discussion

### *3.1. Data Preprocessing*

This paper used OCT2017 [9] as the dataset, OCT Images are labeled as (disease)- (randomized patient ID)- (image number by this patient). Optical coherence tomography (OCT) images [9] were selected from retrospective cohorts of adult patients from the Shiley Eye Institute of the University of California San Diego, the California Retinal Research Foundation, Medical Center Ophthalmology Associates, the Shanghai First People's Hospital, and Beijing Tongren Eye Center between July 1, 2013 and March 1, 2017. There are 84,495 X-Ray images (JPEG), and split into 4 directories: CNV, DME, DRUSEN, and NORMAL.

This paper applies data augmentation as data pre-processing by conducting random rotations, horizontal flipping, and cropping. These techniques help increase the diversity of the training data and improve the model's robustness to variations in the input images [10].

This article first uses a pre-trained super-resolution model to perform 4x super-resolution operations on images in the OCT2017 dataset. The super-resolution model used includes bicubic differencing based on differencing methods, SRResnet based on deep learning, and SRGAN with added adversarial mechanisms.

This article uses the above three methods to infer the OCT dataset, and the inferred dataset is used for training the classification model (Resnet). After training, the accuracy of different super-resolution methods is compared to summarize the ability of different super-resolution models to improve the performance of the classification model.

### *3.2. Performance Comparison of Super-Resolution Models*

In this paper, the Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) evaluation metrics were employed for computing for the OCT2017 testing set shown in Table 1 and Table 2. All tests involved down-sampling and super-resolution magnification with a factor of four.

**Table 1.** PSNR of testing set.

| | CNV | DME | DRUSEN | NORMAL |
|---|---|---|---|---|
| **SRResNet** | 30.191 | 30.766 | 30.894 | 32.668 |
| **SRGAN** | 26.792 | 27.413 | 27.246 | 28.835 |

DOI: 10.54254/2755-2721/46/20241164

**Table 2.** SSIM of testing set.

| | CNV | DME | DRUSEN | NORMAL |
|---|---|---|---|---|
| **SRResNet** | 0.700 | 0.715 | 0.724 | 0.780 |
| **SRGAN** | 0.569 | 0.597 | 0.593 | 0.664 |

The indicators of SRResnet outperform those of SRGAN. The SRGAN paper [6] consistently highlights that PSNR and SSIM may not accurately represent the quality of super-resolved images.

Even though SRResNet generates images that are less realistic and overly smooth, it attains higher scores in these metrics when compared to SRGAN. Nevertheless, from a visual perspective, SRGAN yields more discernible results to the naked eye in contrast to SRResnet. The following images shown in Figure 5 offer a visual comparison of the final output results produced by different super-resolution methods.

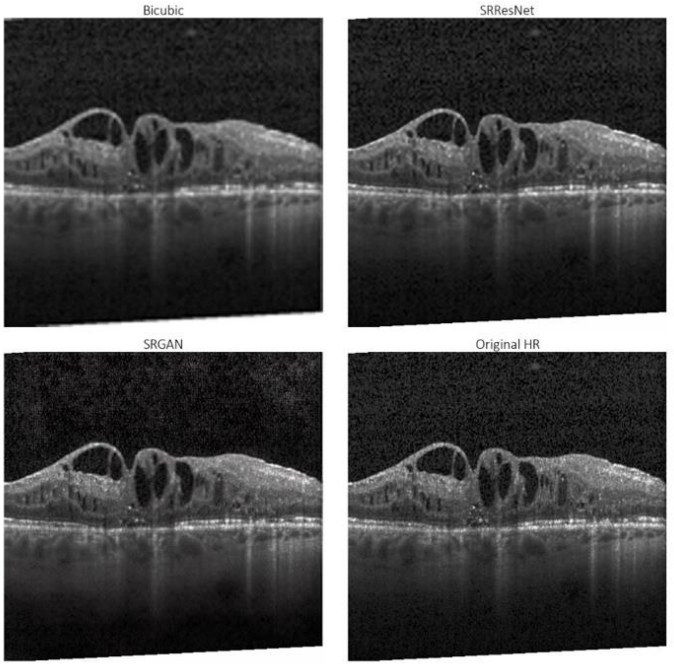



**Figure 5.** Comparison of different methods on a sample of DME category.

### *3.3. Performance Comparison of Classification Models*

This article uses Resnet-34 as the basic classification model, loads the pre-trained model based on the ImageNet weight, and optimizes hyperparameters for the OCT2017 dataset on this basis. The accuracy and loss during the training shown in Figure 6 and Figure 7.

DOI: 10.54254/2755-2721/46/20241164

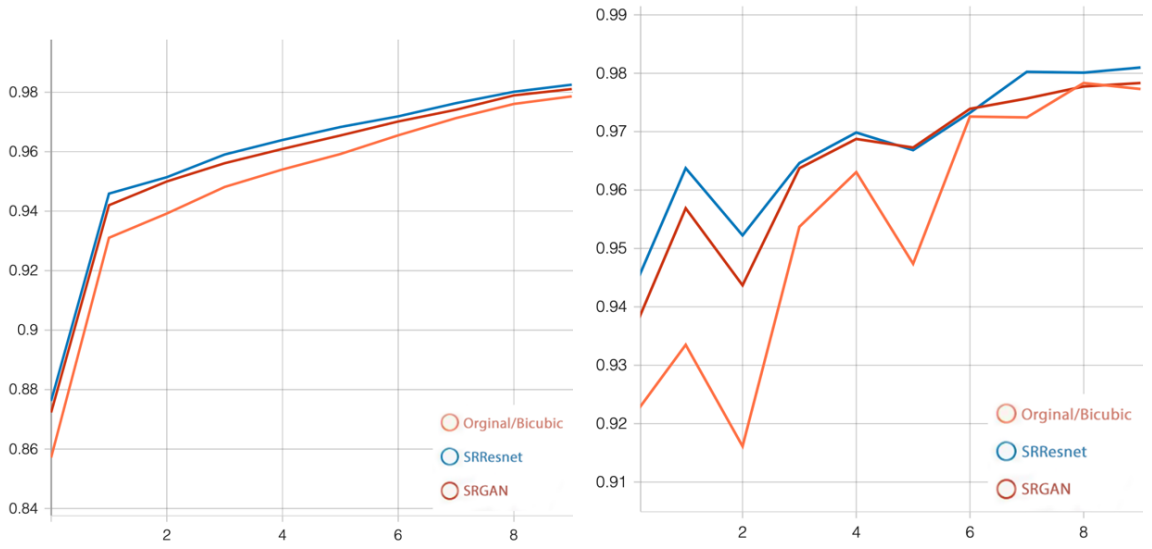



**Figure 6.** Accuracy during training.

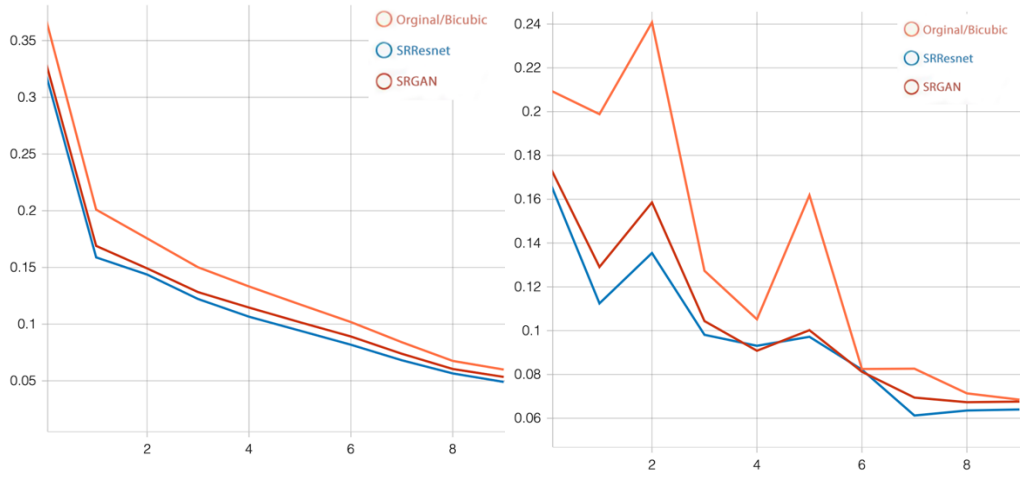



**Figure 7.** Loss during training.

In this study, super-resolution techniques based on interpolation and deep learning are examined to assess their impact on the performance of classification models. The findings reveal that bicubic interpolation does not yield significant improvements in classification model accuracy.

Conversely, an investigation into the effectiveness of two deep learning-based approaches, SRResnet and SRGAN, demonstrates that both methods have a positive influence on classification accuracy. It is noteworthy, however, that when considering visual clarity as an additional factor, SRGAN, despite achieving relatively lower PSNR and SSIM scores in comparison to SRResnet, does not surpass the latter in terms of enhancing classification model performance.

It can be inferred that employing Generative Adversarial Networks (GANs) and loss functions based on perception may not always lead to better results for enhancing classification model performance. Additionally, when evaluating these two deep learning approaches, SRResnet emerges as the preferred option. Therefore, under similar training conditions and iterations, SRResnet remains the preferred option for enhancing classification model performance while upholding visual quality standards. These

DOI: 10.54254/2755-2721/46/20241164

findings provide valuable guidance for practitioners and researchers seeking to optimize classification models through the implementation of image enhancement techniques.

## 4. Conclusion

In the era of advancing neural network technologies, particularly deep learning, their integration into medical image analysis holds great promise. This paper explored the use of diverse super-resolution models to enhance medical image classification. By comparing bicubic interpolation, SRResnet, and SRGAN, this study found that deep learning-based super-resolution methods outperformed interpolation in improving classification model accuracy. SRResnet demonstrated superior performance in terms of classification accuracy, even though SRGAN produced visually more appealing results. These findings suggest that while GANs and perceptual loss functions offer potential in super-resolution, they may not always translate to improved classification model performance. In practical applications, SRResnet remains a preferred choice for enhancing classification models while maintaining high visual quality standards.

## References

[1] Venkateswarlu Y et al 2022 An efficient outlier detection with deep learning-based financial crisis prediction model in big data environment. Computational Intelligence and Neuroscience

[2] Qiu Y et al 2022 Pose-guided matching based on deep learning for assessing quality of action on rehabilitation training Biomedical Signal Processing and Control 72 103323

[3] Liu T Siegel E and Shen D 2022 Deep learning and medical image analysis for COVID-19 diagnosis and prediction. Annual review of biomedical engineering 24 179-201

[4] Han Q Li Q and Niu X 2011 A Face Image Database for Evaluating Out-of-Focus Blur. In T. Barbu (Ed.), Face Analysis, Modeling and Recognition Systems. InTech Available from: http://www.intechopen.com/books/face-analysis-modeling-and-recognition-systems/a-face-image-database-for-evaluating-out-of-focus-blur

[5] Caseneuve G Valova I LeBlanc N and Thibodeau M 2021 Chest X-Ray Image Preprocessing for Disease Classification. Procedia Computer Science, 192, 658-665. https://doi.org/10.1016/j.procs.2021.08.068

[6] Ledig C et al 2017 Photo-realistic single image super-resolution using a generative adversarial network In CVPR 105–114

[7] Simonyan K and Zisserman A 2015 Very deep convolutional networks for large-scale image recognition. In International Conference on Learning Representations (ICLR).

[8] He K Zhang X Ren S and Sun J 2016 Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, 770-778.

[9] Kermany D Zhang K and Goldbaum M 2018 Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification. Mendeley Data, V2. doi:10.17632/rscbjbr9sj.2

[10] Shorten C and Khoshgoftaar T M 2019 A survey on Image Data Augmentation for Deep Learning. J Big Data, 6, 60. https://doi.org/10.1186/s40537-019-0197-0