


A Novel Image Super-Resolution Reconstruction Framework Using the AI Technique of Dual Generator Generative Adversarial Network (GAN)


Loveleen Kumar

(JECRC University, Jaipur, India)

 <https://orcid.org/0000-0002-6532-4220>, loveleentak@gmail.com

Manish Jain

(JECRC University, Jaipur, India)

 <https://orcid.org/0000-0002-3535-0030>, halomanish@gmail.com

Abstract: Image superresolution (SR) is the process of enlarging and enhancing a low-resolution image. Image superresolution helps in industrial image enhancement, classification, detection, pattern recognition, surveillance, satellite imaging, medical diagnosis, image analytics, etc. It is of utmost importance to keep the features of the low-resolution image intact while enlarging and enhancing it. In this research paper, a framework is proposed that works in three phases and generates superresolution images while keeping low-resolution image features intact and reducing image blurring and artifacts. In the first phase, image enlargement is done, which enlarges the low-resolution image to the 2x/4x scale using two standard algorithms. The second phase enhances the image using an AI-empowered Generative adversarial network (GAN). We have used a GAN with dual generators and named it EffN-GAN (EfficientNet-GAN). Fusion is done in the last phase, wherein the final improved image is generated by fusing the enlarged image and GAN output image. The fusion phase helps in reducing the artifacts. We have used the DIV2K dataset to train the GAN and further tested the results on the images of Set5, Set14, B100, Urban100, Manga109 datasets with ground truth of size 224x224x3. The obtained results were compared with the state-of-the-art superresolution approach based on important image quality parameters, namely, Peak signal-to-noise ratio (PSNR), Structural similarity index (SSIM), Visual information fidelity (VIF) image quality parameters. The results show that the proposed framework for generating super-resolution images from 2x/4x resolution downgraded images improves the aforementioned mentioned image quality parameters significantly.

Keywords: Industrial-image enhancement, image superresolution, image analytics, AI-empowered Imaging system, dual Generator, Generative Adversarial Network (GAN)

Categories: J.3, H.3.2, H.5.1, I.4.6, M.7

DOI: 10.3897/jucs.94134

1 Introduction

A single low-resolution (LR) image or multiple LR images of the same scene can be used to create superresolution images. However, generating the superresolution using a single image is still a challenging task. It has an ill-posed problem as a single pixel of LR image takes part in the prediction of multiple pixels in the superresolution image. This problem becomes more complex when we reconstruct superresolution on higher

scales; this can be resolved using statistical analysis of low-level features of image patches [Dong et al., 2022] [Timofte et al. 2014].

Low-resolution (LR) image has a smaller number of pixels; the super-resolution technique adds and enhances the quality of the pixels to make it equivalent to the high-resolution (HR). An image can be reconstructed from a single LR image or from sequences of LR images. A series of LR images can be taken from one or more cameras. Only one LR image is used to construct an equivalent HR image in Single image super-resolution (SISR). That is a tedious task to perform. SISR algorithms can be represented by mainly the below-mentioned three classes, these are based on interpolation, reconstruction, and learning methods [Yang et al. 2019].

An interpolation-based method is one of the simplest methods to enlarge and enhance the image. Bilinear interpolation, nearest neighbor interpolation, bicubic interpolation, and Simple Replication are among the popular interpolation-based super-resolution-based methods [Patil and Bormane 2007], which calculates the new pixel value by interpolating the existing pixels, but it is not able to generate the super-resolution with high accuracy. Although Bicubic interpolation gives images with smooth edges with fewer artifacts [Keys 1981], sometimes reconstruction-based SR methods [Jian et al. 2008] [Marquina and Osher 2008] [Qing et al. 2015] [Shengyang et al. 2009] are also not able to deliver sharp results. Learning-based algorithms use the prior data to predict the value of the new pixel. All machine learning and deep learning models fall under this category, which has better accuracy than the two other methods. Generative models like Generative Adversarial Networks (GANs) are also popular for generating super-resolution (SR) images [Ledig et al. 2017].

In the proposed framework, we have used EffNet-GAN, which is a Novel GAN, which uses the layers of EfficientNet [Sandler et al. 2018] along with enlargement algorithms in preprocessing and image fusion in postprocessing to generate a final superresolution image.

Two neighbor embedding approaches, ANR [Timofte et al. 2014] and SF [Jian et al. 2008] for image enlargement and enhancement, were introduced to generate high-quality SR images without reducing the image quality. Some essential features are taken from Anchored Neighborhood Regression (A+) SR algorithm, which has shown better results over the other neighbor embedding and sparsity-based methods [Timofte et al. 2014].

In 2014, Goodfellow invented the generative adversarial network (GAN), a sort of generative model, which has a combination of two networks that work in adversarial mode. The objective of this network is to generate new data that should fit into the input variable. It has a generator and a discriminator network, and both of which attempt to deceive the other. Training of the discriminator is being done in high-resolution and generated images. The generator gets trained through the discriminator's feedback. The training task gets completed once the model converges on the training dataset, which is when the generator successfully fools the discriminator, whose task is to discriminate the generated data from the original data [Goodfellow et al. 2014].

Training any deep learning network on these images takes a lot of effort and costs both time and resources. Efficient resources are not accessible at a low cost, so we need some techniques or frameworks to get improved results with the available resources. Transfer Learning provides us the flexibility to work with less datasets when we transfer the learning from one related task to another related task. The pretrained model approach is one of the most common and practical methods in transfer learning. In this

approach, to overcome the challenge, previously trained models are employed to extract features from the images [Goodfellow et al. 2017]. Complex networks don't always lead to higher accuracy, but systematic scaling does. The convolution neural network works on one area of the network among depth, width, or resolution. For the best results, all three dimensions should be scaled uniformly with a fixed ratio. EfficientNet has a CNN which is the most effective network to do so using compound scaling [Tan and Le 2019]. EfficientNet balances all three dimensions with the help of Depth-Wise Convolution,

Pointwise Convolution, Inverted Residual Block (MBConv Block), and Linear bottleneck [Jian et al. 2008].

EfficientNet works well with large widths, depth, and resolution. EfficientNet has variations from B0 to B7 to save time and computational power transfer learning can be used.

Image Fusion is the process of creating a single image with valuable and essential information from multiple images. It takes two or more registered images as an input and gives an output an image that shows better image quality parameters and human visual perception. Low-Level Feature-based fusion preserves the significant features of the input images and reduces the blurring and artifacts [Zhang et al. 2018].

In this study, a perceptual loss helps to generate the superresolution image from the network. This loss is calculated by combining adversarial loss, content loss, and EfficientNet loss, which is effective in human perception [Shahidi 2021]. Natural images are taken from DIV2K dataset [Agustsson and Timofte 2017] [Ignatov et al. 2019] to prepare the network and the performance was evaluated using image quality parameters like Peak signal-to-noise ratio (PSNR), Structural Similarity Index (SSIM), visual information fidelity (VIF) image quality parameters.

We offer a framework that takes an LR image as input and produces a superresolution image, which is an improved version of the original. Apart from image enhancement, we did preprocessing and postprocessing on the image. The following are the research's overall contributions:

In preprocessing of the input image, we applied bicubic interpolation (BI) and A+ image enlarging algorithms are used to keep the edges clear, remove the artifacts, and keep the detailed information of the low-resolution image. With a narrow-wide-narrow strategy, we suggested the EffN-GAN architecture with an Inverted Residual Block (MBConv Block), which can increase the network's overall performance.

We designed a dual generator GAN, one with a shallow and another with a deep network, to get both low and high level features. To generate the superresolution image, we applied adversarial loss, content loss, and EfficientNet loss. In postprocessing of the image, two images are fused using the proposed GAN and image enlargement algorithm to get a superresolution image after removing the blur and artifacts of the image. For the training of the EffN-GAN, we used the DIV2K dataset and the Set5, Set14, B100, Urban100, and Manga109 datasets for testing.

This preprocessing, EffN-GAN and postprocessing of the LR image have significantly improved the image quality and generated an image of 224x224 size.

The paper is arranged as follows: In Section 2, we survey the literature on the superresolution framework. In Section 3, the proposed methodology with the framework has been described. Section 4 contains the results and analysis. We have concluded the paper with the future work in Section 5. Finally, a list of references is provided.

2 Related Work

In the image super-resolution image get enlarged to 2x, 3x, 4x, or any other scale. This increment in pixels plays a very crucial role as other than adding pixels in the image, pixel enhancement should also be taken into consideration.

[Sung et al. 2003], have defined a few SR image reconstruction algorithms named based on the Nonuniform Interpolation, Regularized SR Reconstruction, Frequency Domain, and iterative back-projection approaches. The interpolation approach became famous due to the reduction in artifacts. These methods are being applied to many datasets to enlarge the images and enhance the images. Our proposed system also uses Bicubic Interpolation (BI) but with more postprocessing after applying this interpolation algorithm to the low-resolution image.

Moreover, in [Timofte et al. 2014], the authors have introduced one dictionary-based algorithm to enlarge and enhance the image, which has performed well using the appropriate neighborhood size and dictionary. This algorithm doesn't require the image prior to processing, which has increased the speed of processing. A Convolutional Neural Network (CNN) is a learning-based model to get the end-to-end superresolution image from low to high resolution images. CNN based superresolution gave good results [Chen et al. 2019], [Fan et al. 2017], [Lim et al. 2017], [Zhang et al., 2018(a)], [Zhang et al. 2018(b)] in a less blurry superresolution images.

Moreover, [Shahidi 2021] proposed a framework for wide-attention SRGAN (W-GAN) and displayed the findings for Quality Index based on Local Variance (QILV), another important image quality measure, that is more reliable for local variance. In this framework, the features are extracted using a VGG-19 pretrained model, which was trained using Perceptual Loss and enhanced Wasserstein Gradient Penalty.

Moreover, the authors in [Zhu et al. 2020] have used GMGAN with quality loss and WGAN-GP to polish up the results of super-resolution images.

To extract the features in the proposed system, we have used layers of the EfficientNet convolutional network that work on the height, width, and resolution parts of the image. We have not used GAN to enlarge the image, which reduces the extra work of the GAN, and it concentrates only on enhancing the input image.

[Dong et al. 2014] proposed SRCNN, a CNN-based technique for reconstructing a SR image from a LR image that optimizes all layers simultaneously. This model has preprocessed the LR image to increase its number of pixels, and then SRCNN was used to enhance this image. Furthermore, [Dong et al., 2016(b)] have improved the CNN named FRCNN with a deconvolutional layer for upscaling the image instead of the image enlargement algorithm. Deeper convolutional neural networks with high accuracy were introduced by [Ledig et al. 2017] in the year 2017. To enhance and enlarge the low-resolution image, they employed a generative adversarial network. The network was trained using perceptual loss. PSNR, SSIM, and mean-opinion score were used to assess the quality of the created superresolution image (MOS). We have used EfficientNet layers in the GAN and VIF as one of the image quality assessment parameters, which checks the image quality by considering human visibility. In the process of convergence of GAN training, the generator in the back-propagation should produce variation in the results. To get the diverse influences from the GAN, [Zareapoor et al. 2019] have used multiple generators. In this proposed model, the perceptual loss function is used to handle the diverse samples using two generators, which are also used to extract the low-resolution image's attributes.

3 Methodology

Our proposed framework uses three phases to reconstruct the superresolution image from a single LR image. In the first phase, also called preprocessing, two algorithms, namely, Bicubic Interpolation (BI) and A+, are used to enlarge the LR images. These algorithms increase the number of pixels in a certain ratio by taking care of the image quality. This phase takes a low-resolution image of size 112x112 and gives the 2x enlarged image. The enlarged image from BI is further passed to the dual generator GAN in the second phase. This proposed EffN-GAN has two generators, and the final output image is formed by concatenating the outputs of the generators. The dual generators structure in the GAN works on the low and high-level features of the image. Due to the enlargement of the LR image in the first phase, GAN needs to concentrate on the improvement of the image quality only, not on the image enlargement. Finally, essential and valuable information from the GAN output image and A+ algorithm passed images is extracted using the image fusion process. Fig.1. showing all phases of the framework to reconstruct the superresolution image.

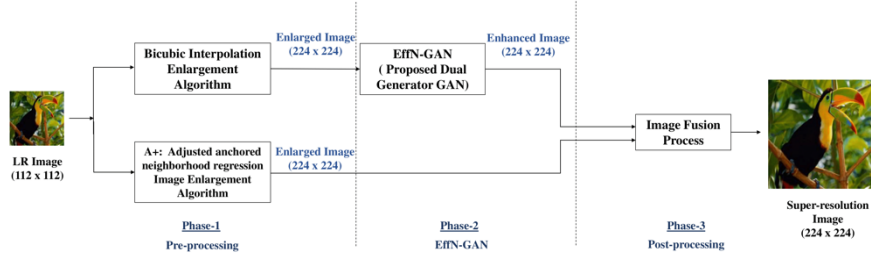


Figure 1: The framework of reconstruction of a super-resolution image from a low-resolution image with a scale factor of 2x

The complete components of the proposed framework are described below:

A. Datasets

We have used the DIVERse 2K resolution image dataset (DIV2K) [Agustsson and Timofte et al. 2017] to train the GAN. Further, Set5, Set14, B100, Urban100, Manga109 datasets are used to test the results. The DIV2K dataset has high-resolution images with 2K resolution [Agustsson and Timofte 2017] [Ignatov et al. 2019]. We have changed the size of the images to 112x112 and 56x56 using bicubic interpolation; this will work as a LR image. Our objective is to generate a superresolution image of 224x224 pixels.

GAN Structure	PSNR	SSIM	VIF
Single Generator	29.33	0.8954	0.9341
Dual Generator (EffN-GAN) (Proposed)	32.28	0.9344	0.9308

Table 1: Image Quality Parameter Comparison with Single and Dual Generator

B. Enlarging Algorithms

After experimenting with several algorithms, we have used two enlarging algorithms, the bicubic interpolation algorithm and the A+ algorithm, to enlarge the low-resolution image. Both have their distinct capabilities for dealing with low-resolution images. The

bicubic interpolation algorithm is one of the best algorithms for smoothing the edges of natural images. A+ algorithm preserves the features of the LR image. This is the preprocessing of the image before it is input to the GAN. Because the GAN does not have to enlarge the image as a result of this preprocessing, it requires less effort. This preprocessing of the low-resolution was also done in some recent work [Dong et al. 2014] [Kim et al. 2016] [Lyu et al. 2020].

C. Perceptual Loss

Perceptual loss identifies the difference between two images that can be perceptually identifiable. Perceptual loss that is a combination of EfficientNet loss, adversarial loss, and a content loss is used to train the EffN-GAN, as in SRGAN [Ledig et al. 2017].

1) *Adversarial Loss*: We produced the image using Adversarial Loss and fooled the discriminator into thinking it was a real image. Here, a modified minimax function is used to minimize the $-\log[D(G(I^r))]$ rather than of $\log[1 - D(G(I^r))]$ [Goodfellow et al. 2014] [Ledig et al. 2017].

2) *Content Loss*: To get a better image quality, we have used the popular MSE loss function that calculates the values of the function basis on the pixels. This is one of the most popular and effective methods for creating high-quality super-resolution images. [Dong et al. 2016(a)] [Ledig et al. 2017].

3) *EfficientNet Loss*: Sometimes, MSE is not able to generate superresolution on the natural image due to the high frequencies. To deal with this issue, we have used the loss function which uses the EfficientNet layer to produce the superresolution image. We have defined EfficientNet loss based on LeakyReLU activation, squeeze, and excitation blocks used in the architecture, as shown in Equation 1.

$$l_{eff} = \frac{1}{W_{m,n}H_{m,n}} \sum_{i=1}^{W_{m,n}} \sum_{j=1}^{H_{m,n}} (\varphi_{m,n}(I^r)_{i,j} - (\varphi_{m,n}(G_{\theta_G}(I^{hr}))_{i,j})^2 \quad (1)$$

Here:

- $W_{m,n}$ and $H_{m,n}$ are the EfficientNet feature maps' dimensions.
- $\varphi_{m,n}$ is the obtained feature map from the EfficientNet.
- I^{hr} is the high-resolution reference image and I^r is the low-resolution image.
- The distance between a reconstructed image's feature representations $G_{\theta_G}(I^{hr})$ and the original image I^{hr} .

Finally, the perceptual loss of the GAN is calculated and shown in Equation 2.

$$L = l_{pixel}^{SR} + 10^{-3} \cdot l_{gen}^{SR} + 6 * 10^{-3} l_{eff} \quad (2)$$

D. GAN Architecture

The GAN in this framework uses two generators, one with a shallow network and another with a deep network, that work on the low and high-level features, respectively. Furthermore, the outputs of these generators are concatenated to give the input to the discriminator for training purposes.

We have tried the output of the GAN with a single generator with a mix of shallow and deep networks, but that is still unable to generate the quality with the preserved features of the low-resolution image. Table 1 shows the improvement in image quality parameters with the same image passing through the GAN with a single generator and a dual generator.

The proposed GAN has two generators with shallow and deep networks and is responsible for low and high-level features, respectively, as shown in Fig.3. The same image is being passed to both generators, and the outputs are concatenated to generate the enhanced image. A 224x224 size image is supplied to the generator that is already of a size equivalent to the ground truth, which means we have GAN only for the enhancement task, not for the enlargement. This reduces the additional task of the generator. Proposed generators have squeeze and excitation blocks; these are responsible for re-calibrating the network channel interdependencies and improving the efficiency of the network.

In the generator, we have used Conv2D, MBConv Block, and squeeze and excitation blocks. Conv2D has 2D convolution with K11-S1, which means the size of kernel 11 and 1 stride; after applying batch normalization, Relu activation function is used. Fig.2 shows the layout of the squeeze and Excitation Block; where H, W, and C denote the height, width, and channels, respectively, of the input and output image. In Squeeze, the input is reshaped to 1D, and dense-activation-dense-activation layers are used for excitation. Layers of the B0 variant of EfficientNet [Tan and Le 2019] were used for feature extraction and to perform compound scaling to improve the image quality generated by the generator. The B0 variant is lightweight that takes less time in training with adam optimizer. The second generator has a deeper network than the first generator to perform compound scaling.

MBConv Block from the EfficientNet-B0 variant has been used with a distinct kernel size, Fig.3 is showing the complete architecture. For the training of EffN-GAN without overfitting with different types of images from the dataset, we have used batch normalization with $\text{momentuma} = 0.6$. This value of momentum is useful as we are using a bigger batch size.

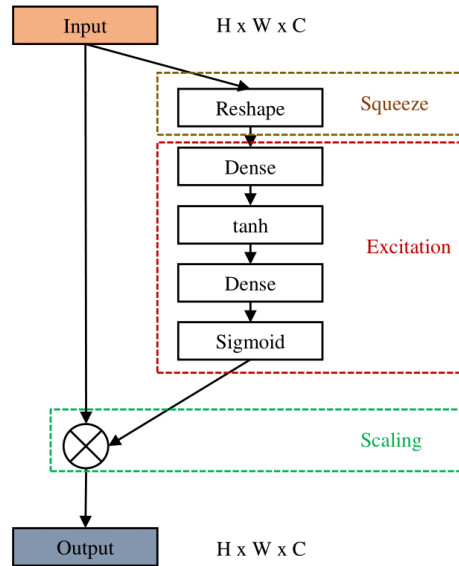


Figure 2: Squeeze and Excitation Blocks used in the generator.

The generated image is supplied to the discriminator along with the training image so that it can discriminate against the real image. If the discriminator has caught the image as a generated image, then a back-propagation generator is further trained. This training of the EffN-GAN stops when the generator fools the discriminator. Once the training is completed, we only need generators with the final weights to generate the enhanced images.

The architecture of the EffN-GAN's discriminator is shown in Fig.4. We have used a discriminator that is very much similar to SR-GAN [Ledig et al. 2017], the first Conv2D block has the structure of K3-N64-S1, which means it has 3 kernels, 64 channels and 1 stride. In the architecture, the kernels increase from 64 to 256 with stride 1 or 2, which reduces the image resolution. We have replaced the dense layer with one layer of global averaging pooling as the dense layer is more prone to overfitting [Samek et al. 2021]. Finally, the sigmoid function is employed to decide whether or not the image is genuine.

E. Image Fusion

The image received from the EffN-GAN can be further enhanced to remove the blur and artifacts. As the image quality assessment parameter values in Table 2, there may also be scope to improve the parameters after phase-2. To improve the image quality, we have fused images received from EffN-GAN and A+ enlarging algorithm. To get the final improved image preserving low-level features, we have used base layer and detailed layer fusion [Zhang et al. 2021] of two images received from the EffN-GAN and A+ enlarging image algorithm.

In this fusion process, initially, a fast guided filter (FGF) is applied to the EffN-GAN output image, FGF is an edge-preserving smoothing technique that assists in the reduction of the noise and keeping the image detail while preserving the image's overall low-level features [Zhang et al. 2021]. A Laplacian filter is also on the same input image to get the saliency map, which is further optimized on the FGF output image. To decompose the image into a basic and detailed layer, we applied local binary pattern (LBP) operators. These decomposed images are further used to get the final fused image I_1 . The same process is also performed on the A+ output image to get another image I_2 . Finally, I_1 and I_2 are XORed to get the final fused image.

F. Image Quality Assessment (IQA)

In the proposed framework, we have checked the quality of the output superresolution image using the following Image Quality Assessment parameters:

1) *Peak signal-to-noise ratio (PSNR)*: This is a well-known parameter to check the image quality. Its the ratio of a signal's greatest possible power to the corrupting noise's power. The quality of the images improves as the ratio increases.

In equation (6), a formula for calculating PSNR in terms of mean square error (MSE) is given. [Hore et al. 2010] [Setiadi 2020] [Vidhya and Ramesh 2017]. Given a noise-free $m \times n$ monochrome image I with its noisy approximation K

$$MSE = \frac{1}{ab} \sum_{m=0}^{a-1} \sum_{n=0}^{b-1} [I(m, n) - K(m, n)]^2 \quad (3)$$

$$PSNR = 10 \cdot \log_{10} \left(\frac{Maximum_I^2}{MSE} \right) \quad (4)$$

$$PSNR = 20 \cdot \log_{10} \left(\frac{Maximum_I}{\sqrt{MSE}} \right) \quad (5)$$

$$PSNR = 20 \cdot \log_{10} Maximum_I - 10 \cdot \log_{10} MSE_I \quad (6)$$

where,

- *Maximum*: pixel value of the image with maximum possibility

2) *Structural Similarity Index (SSIM)*: The perceptual difference between the ground truth and the formed image is checked with this parameter. Its maximum value is one if both images are identical. The formula to calculate the SSIM is given in equation (7) [32].

$$SSIM(a, b) = \frac{(2\mu_a\mu_b + M_1) + (2\sigma_{ab} + M_2)}{(\mu_a^2 + \mu_b^2 + M_1) + (\sigma_a^2 + \sigma_b^2 + M_2)} \quad (7)$$

where,

- μ_a and μ_b are denoting the average values of a and b, respectively;
- The variances of a and b are presented by σ_a^2 and σ_b^2 ;
- Covariance of a and b is denoted by σ_{ab} ;
- A weak denominator is used to balance the division using variables $M_1 = (s_1 R)^2$ and $M_2 = (s_2 R)^2$;
- R is the pixel values' dynamic range;
- $s_1 = 0.01$ and $s_2 = 0.03$, these are the default values.

3) *Visual Information Fidelity (VIF)*: This image quality metric is linked to the image's visual information, demonstrating the link between image information and visual quality. Its a straightforward ratio of the reference image's and reconstructed image's two information metrics. VIF will be unity in the reference image, and if it is greater than unity, it indicates better visual quality. Equation (8) is showing the formula to calculate VIF [Sheikh and Bovik 2006] [Sheikh and Bovik 2005].

$$VIF = \frac{\sum_{j \in \text{subbands}}^n I(\vec{C}^{N,j}; \vec{F}^{N,j} | \vec{S}^{N,j})}{\sum_{j \in \text{subbands}}^n I(\vec{C}^{N,j}; \vec{E}^{N,j} | \vec{S}^{N,j})} \quad (8)$$

- $\vec{C}^{N,j}$: the N blocks in the wavelet decomposition's subband;
- $\vec{C}^{E,j}$ and $\vec{C}^{N,j}$: the N blocks in the wavelet decomposition of the reference and test images' subbands, respectively;
- $\vec{S}^{N,j}$: the most probable estimate from N blocks in subband j ;

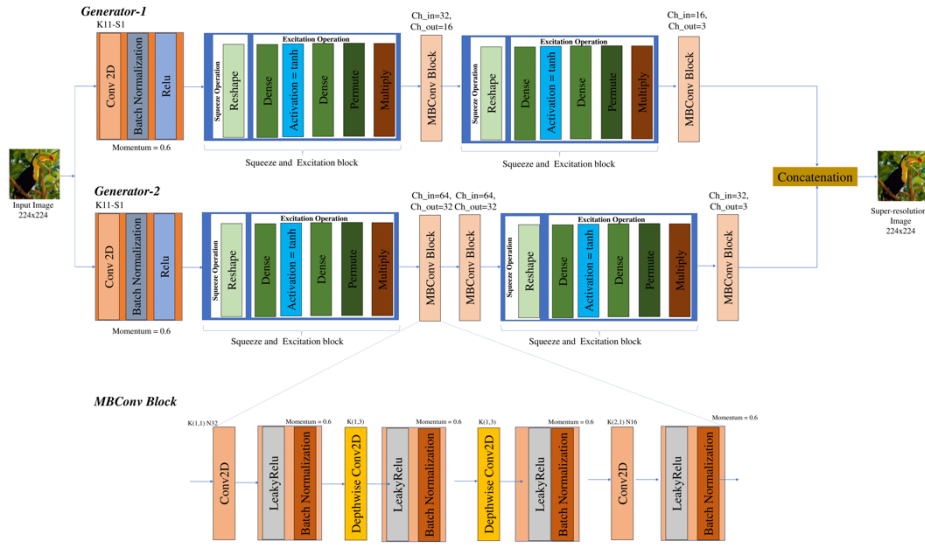


Figure 3: Dual Generator Architecture of proposed EffN-GAN.

G. Experiment Setup

The output of the framework is 224x224x3, which is a superresolution image equivalent to the ground truth. Patches with the same size as the ground truth and some with 112x112x3 dimensions have been created using DIV2K images. We have performed all experiments on Google Colaboratory (Colab) with the Pro version. Using Colab Pro, we could use GPUs or TPUs, such as the T4 or P100 with 12GB (upgradable to 26.75GB) high-memory virtual machine RAM.

In the preprocessing, the image was enlarged through BI and A+ algorithms, and we used the available code in MATLAB. In postprocessing, image fusion has also been coded in MATLAB. Furthermore, EffN-GAN has trained on Google colab with different batch sizes, but batch size 5 has given better results. Finally, using a batch size of 5, we trained EffN-GAN for 200 epochs and achieved a minimum discriminator loss of 0.166 on epoch no 194. The loss has started to increase after 194 epochs. Fig.5 depicts the output loss graph of the proposed EffN-GAN.

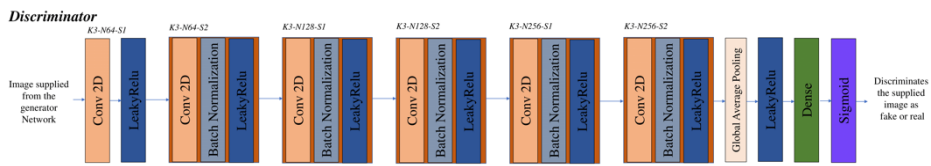


Figure 4: Discriminator Architecture of proposed EffN-GAN.

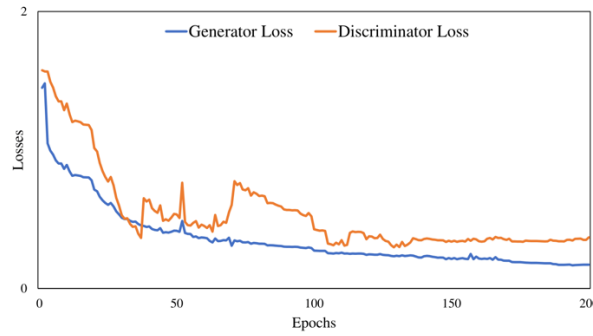


Figure 5: Proposed EffN-GAN Output loss graph.

4 Results and Analysis

EffN-GAN has been trained on DIV2K dataset images and the final results have been tested on the images of Set5, Set14, B100, Urban100, Manga109 datasets.

In our experiments, we used image quality assessment (IQA) measures such as PSNR, SSIM, and VIF to assess image quality.

In Table 2, we have shown the PSNR, SSIM, and VIF image quality parameters after each phase of the proposed framework calculated on the images of Set5, Set14, B100, Urban100, and Manga109 databases. A+ enlarging algorithm produces a better image with high PSNR and VIF than bicubic interpolation, with bicubic interpolation having better SSIM in Phase 1. Phase 2 has GAN with a dual generator that enhances all image parameters. In Phase 3, we have used an image fusion process that further enhances the image assessment parameters. In the proposed framework, we have enlarged the image in phase 1. We have not used convolutional neural networks to enlarge the image, which helps reduce the checkerboard representation in the generated image. The final enhanced image is generated through the fusion of images generated by A+ and EffN-GAN, which has a better value of PSNR, SSIM, and VIF.

Image Dataset	Enlargement Algorithm (112x112 to 224x224)						Output image from EffN-GAN (224x224)			Final output image after Image Fusion (224x224)		
	Bicubic Interpolation			A+			PSNR	SSIM	VIF	PSNR	SSIM	VIF
Set5	28.44	0.8821	0.8821	30.22	0.8749	0.9041	32.28	0.9344	0.9308	33.58	0.9318	0.9848
Set14	24.25	0.8016	0.8322	25.47	0.7955	0.8797	27.33	0.8628	0.8987	29.88	0.8841	0.9219
B100	24.22	0.8454	0.8814	26.11	0.8314	0.9111	27.65	0.8171	0.9314	28.55	0.8188	0.9541
Urban100	23.41	0.8811	0.8911	23.19	0.8746	0.8809	27.47	0.8814	0.9144	29.88	0.8934	0.9357
Manga109	33.54	0.9089	0.9129	32.41	0.9011	0.9044	37.85	0.9489	0.9177	39.99	0.9304	0.9041

Table 2: Image Quality Parameters Values after Each Phase of the Proposed Framework with A Scale Factor of 2x

Furthermore, we have compared the results of the proposed framework with those of the established superresolution frameworks SRCNN [Dong et al. 2014], SRGAN

[Ledig et al. 2017], GMGAN [Zhu et al. 2020], on the basis of PSNR, SSIM and VIF image quality assessment parameters with 2x and 4x image scale. As shown in Table 3, our proposed framework has better image quality assessment parameters than the SRCNN, SRGAN, and GMGAN with the datasets Set5 and Manga109. For 2x scaling, the SSIM value of all images of the test datasets has been increased. Apart from this value of PSNR, VIF from Set5, VIF from Set14, VIF from B100, PSNR from Urban100, PSNR, and VIF from the Manga109 dataset have been increased. While testing the PSNR from Set14 and B100, VIF from Urban100 has been identified less than other methods, as shown in Table 3.

Our proposed model has shown better results on a 4x scale. PSNR, SSIM, and VIF values for Set5, Set14, and Urban100 are the best among the compared methods. The VIF value of B100, PSNR, and SSIM value of Manga109 are showing the best. While the values of PSNR, SSIM for B100, and VIF for Manga109 dataset are slightly lower, Table 3 shows the complete values of the image quality assessment parameters.

Fig. 6 provides visual examples of the output superresolution images. Fig. 6(a) is the high resolution of the corresponding dataset, Fig. 6(b) shows the LR image of 112x112 size. Fig. 6(c), (d), (e), (f) are the generated superresolution from SRCNN, SRGAN, GMGAN, and the proposed framework, respectively, with a size of 224x224. Fig. 6(g) is the ground truth taken from the high-resolution image shown in Fig. 6(a).

Dataset	Image Quality Assessment Parameters	SRCNN	SRGAN	GMGAN	Proposed Framework	SRCNN	SRGAN	GMGAN	Proposed Framework
		2x				4x			
Set5	PSNR	30.51	32.35	33.06	33.58	26.97	28.39	28.19	29.19
	SSIM	0.9014	0.9014	0.9128	0.9318	0.8698	0.8803	0.8815	0.8944
	VIF	0.9219	0.9219	0.9319	0.9848	0.8729	0.9099	0.9014	0.9111
Set14	PSNR	26.44	29.99	28.86	29.92	24.71	25.64	24.51	26.25
	SSIM	0.8714	0.8714	0.8811	0.8841	0.8059	0.8261	0.8314	0.8412
	VIF	0.8914	0.8914	0.9022	0.9219	0.87	0.8998	0.9024	0.9125
B100	PSNR	25.41	27.99	28.91	28.55	23.08	27.16	24.58	26.89
	SSIM	0.8152	0.8211	0.8054	0.8388	0.7726	0.7878	0.8041	0.7999
	VIF	0.9458	0.9458	0.9351	0.9541	0.8895	0.8941	0.9014	0.9253
Urban100	PSNR	27.12	28.94	27.84	29.88	24.72	23.33	24.14	27.64
	SSIM	0.8517	0.8841	0.8854	0.8934	0.8416	0.8469	0.8579	0.8651
	VIF	0.8927	0.9393	0.9149	0.9357	0.9167	0.9311	0.9296	0.9328
Manga109	PSNR	32.91	37.84	35.98	39.99	32.86	34.99	34.87	36.41
	SSIM	0.9021	0.9248	0.9121	0.9304	0.9021	0.8979	0.9049	0.9179
	VIF	0.8721	0.9	0.9011	0.9041	0.8647	0.8891	0.8739	0.8846

Table 3: Image Quality Parameters of the proposed Super-Resolution Framework Compared to SRCNN, SRGAN, and GMGAN With 2x and 4x Scale Factors

5 Conclusion

We have described a framework for single image superresolution with enlarging algorithms as preprocessing and fusion as postprocessing. In between these processes, the proposed GAN named EffN-GAN is used. In EffN-GAN, an EfficientNet pretrained model is used that works on the width, depth, or resolution of an image. EffN-GAN was trained on the DIV2K dataset and testing was done on five different datasets, namely, Set5, Set14, B100, Urban100, Manga109. We have also visualized the results after generating superresolution images from one image from each dataset. The whole framework is divided into three phases, and the image quality gets enhanced after each phase. We have compared the performance of the framework with the established superresolution methods, SRCNN, SRGAN, and GMGAN, based on PSNR, SSIM, and VIF. In our experiment results, we have identified that this framework is better than the aforesaid methods with sufficient margins.

As a part of future work, the percentage of improvement in the value of image assessment parameters can be further improved by experimenting with and deploying the deep network in the framework. EffN-GAN was trained on DIV2K and the complete framework has been tested on Set5, Set14, B100, Urban100, and Manga109 datasets. Going forward, it can be tested on other relevant datasets also. For more improvement of the framework, the output can be tested with additional image quality assessment parameters. Through experimentation, this framework can be applied to the higher upscaling factor as well. Initial results obtained in this direction are encouraging.

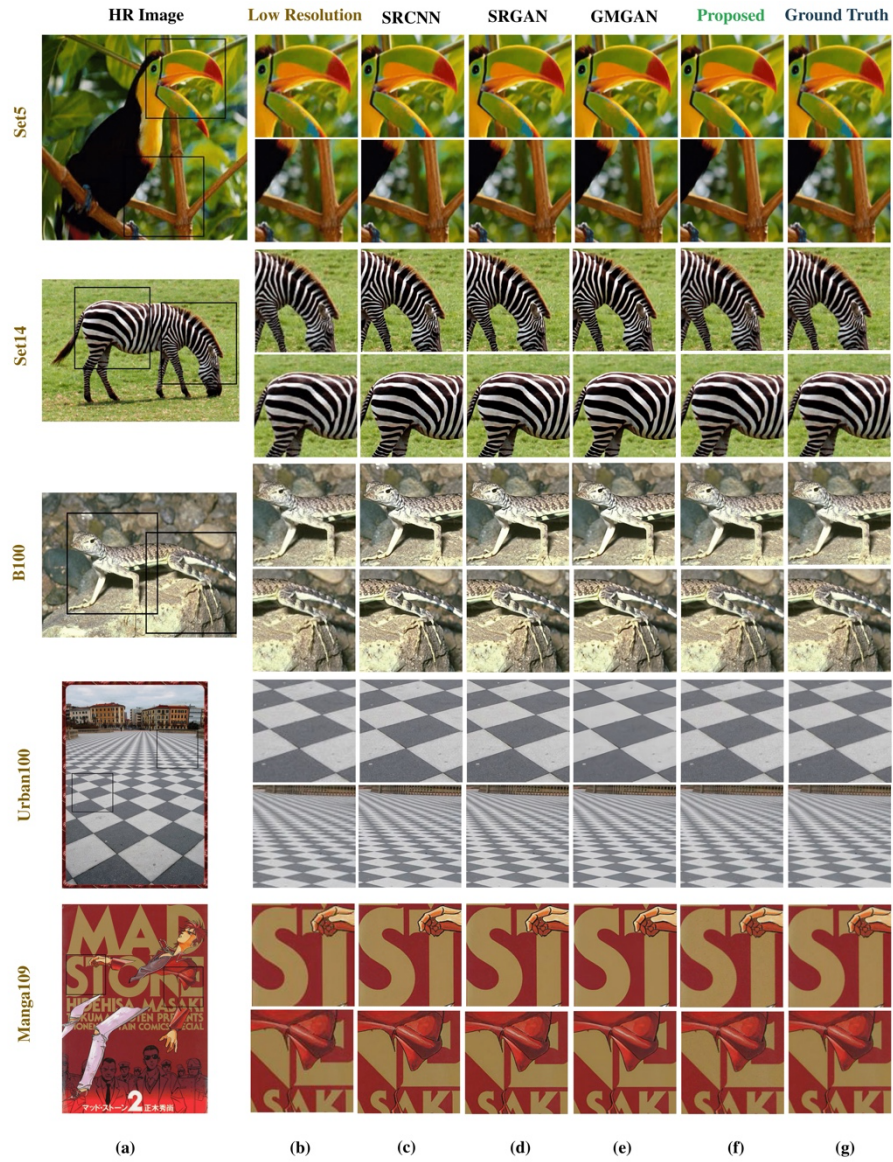


Figure 6: Reconstructed Super-resolution images for 2x upscaling using the proposed framework from the test dataset. (a) High-resolution images of the dataset, (b) Low-resolution images, images reconstructed using (c) SRCNN [Dong et al. 2014], (d) SRGAN [Ledig et al. 2017], (e) GMGAN [Zhu et al. 2020] (f) proposed framework (g) Ground truth (2x)

References

- [Agustsson and Timofte 2017] Agustsson, E., Timofte, R.: “NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study”; 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), (2017). DOI: <http://dx.doi.org/10.1109/cvprw.2017.150>
- [Chen et al. 2019] Chen, L., Pan, J., Li, Q.: “Robust Face Image Super-Resolution via Joint Learning of Subdivided Contextual Model”; IEEE Transactions on Image Processing, 28, 12 (2019), 5897-5909. DOI: <https://dx.doi.org/10.1109/tip.2019.2920510>
- [Dong et al. 2014] Dong, C., Loy, C., He K., Tang, X.: “Learning a Deep Convolutional Network for Image Super-Resolution”; Computer Vision – ECCV 2014, (2014), 184-199. DOI: https://dx.doi.org/10.1007/978-3-319-10593-2_13
- [Dong et al. 2016(a)] Dong, C., Loy, C., He K., Tang X.: “Image Super-Resolution Using Deep Convolutional Networks”; IEEE Trans Pattern Anal Mach Intell, 38, 2 (2016), 295-307. DOI: <https://dx.doi.org/10.1109/tpami.2015.2439281>
- [Dong et al. 2016(b)] Dong, C., Loy, C., Tang, X.: “Accelerating the Super-Resolution Convolutional Neural Network”; Computer Vision – ECCV 2016, (2016), 391-407. DOI: https://dx.doi.org/10.1007/978-3-319-46475-6_25
- [Dong et al. 2022] Dong, R., Zhang, L., Fu H.: “RRSGAN: Reference-Based Super-Resolution for Remote Sensing Image”; IEEE Transactions on Geoscience and Remote Sensing, 60 (2022), 1-17. DOI: <https://dx.doi.org/10.1109/tgrs.2020.3046045>
- [Fan et al. 2017] Fan, Y., Shi, H., Yu, J. et al.: “Balanced Two-Stage Residual Networks for Image Super-Resolution”; 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), (2017). DOI: <https://dx.doi.org/10.1109/cvprw.2017.154>
- [Goodfellow et al. 2014] Goodfellow, Ian, et al.: “Generative adversarial nets”; Advances in neural information processing systems, 27 (2014) DOI: <https://doi.org/10.48550/arXiv.1406.2661>
- [Goodfellow et al. 2017] Goodfellow, I., Bengio, Y., Courville, A.: “Deep Learning”; Cambridge, Mass: The MIT Press, (2017)
- [Hore et al. 2010] Hore, A., Ziou, D.: “Image Quality Metrics: PSNR vs. SSIM”; 2010 20th International Conference on Pattern Recognition, (2010). DOI: <https://dx.doi.org/10.1109/icpr.2010.579>
- [Ignatov et al. 2019] Ignatov A, Timofte R, Van Vu T et al.: “PIRM Challenge on Perceptual Image Enhancement on Smartphones: Report”; In: Leal-Taixé, L., Roth, S. (eds) Computer Vision – ECCV 2018 Workshops. ECCV 2018. Lecture Notes in Computer Science, (2019), 315-333. DOI: https://dx.doi.org/10.1007/978-3-030-11021-5_20
- [Jian et al. 2008] Jian Sun, Zongben Xu, Heung-Yeung Shum: “Image super-resolution using gradient profile prior”; 2008 IEEE Conference on Computer Vision and Pattern Recognition, (2008). DOI: <https://dx.doi.org/10.1109/cvpr.2008.4587659>
- [Keys 1981] Keys R.: “Cubic convolution interpolation for digital image processing”; IEEE Trans Acoust, 29, 6 (1981), 1153-1160. DOI: <https://dx.doi.org/10.1109/tassp.1981.1163711>
- [Kim et al. 2016] Kim J, Lee J, Lee K.: “Accurate Image Super-Resolution Using Very Deep Convolutional Networks”; 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), (2016). DOI: <https://dx.doi.org/10.1109/cvpr.2016.182>

- [Ledig et al. 2017] Ledig C, Theis L, Huszar F et al.: “Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network”; 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), (2017). DOI: <https://dx.doi.org/10.1109/cvpr.2017.19>
- [Lim et al. 2017] Lim, Bee, et al.: “Enhanced deep residual networks for single image super-resolution”; Proceedings of the IEEE conference on computer vision and pattern recognition workshops. (2017). DOI: <https://doi.org/10.1109/CVPRW.2017.151>
- [Lyu et al. 2020] Lyu Q, Shan H, Wang G.: “MRI Super-Resolution With Ensemble Learning and Complementary Priors”; IEEE Trans Comput Imaging, 6 (2020), 615-624. DOI: <https://dx.doi.org/10.1109/teci.2020.2964201>
- [Marquina and Osher 2008] Marquina A, Osher S.: “Image Super-Resolution by TV-Regularization and Bregman Iteration”; J Sci Comput, 37, 3 (2008), 367-382. DOI: <https://dx.doi.org/10.1007/s10915-008-9214-8>
- [Patil and Bormane 2007] Patil V, Bormane D.: “Interpolation for Super Resolution Imaging”; Innovations and Advanced Techniques in Computer and Information Sciences and Engineering, (2007), 483-489. DOI: https://doi.org/10.1007/978-1-4020-6268-1_85
- [Qing et al. 2015] Qing Yan, Yi Xu, Xiaokang Yang, Nguyen T.: “Single Image Superresolution Based on Gradient Profile Sharpness”; IEEE Transactions on Image Processing, 24, 10 (2015), 3187-3202. DOI: <https://dx.doi.org/10.1109/tip.2015.2414877>
- [Samek et al. 2021] Samek W, Montavon G, Lapuschkin S, Anders C, Muller K.: “Explaining Deep Neural Networks and Beyond: A Review of Methods and Applications”; Proceedings of the IEEE, 109, 3 (2021), 247-278. DOI: <https://doi.org/10.1109/JPROC.2021.3060483>
- [Sandler et al. 2018] Sandler M, Howard A, Zhu M, Zhmoginov A, Chen L.: “MobileNetV2: Inverted Residuals and Linear Bottlenecks”; 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, (2018). DOI: <https://doi.org/10.1109/CVPR.2018.00474>
- [Setiadi 2020] Setiadi D.: “PSNR vs SSIM: imperceptibility quality assessment for image steganography”; Multimed Tools Appl, 80, 6 (2020), 8423-8444. DOI: <https://dx.doi.org/10.1007/s11042-020-10035-z>
- [Shahidi 2021] Shahidi, F.: “Breast Cancer Histopathology Image Super-Resolution Using Wide-Attention GAN With Improved Wasserstein Gradient Penalty and Perceptual Loss”; IEEE Access, 9 (2021), 32795-32809. DOI: <https://doi.org/10.1109/ACCESS.2021.3057497>
- [Sheikh and Bovik 2005] Sheikh H, Bovik A.: “Information Theoretic Approaches to Image Quality Assessment”; Handbook of Image and Video Processing, (2005), 975-989. DOI: <https://dx.doi.org/10.1016/b978-012119792-6/50120-0>
- [Sheikh and Bovik 2006] Sheikh H, Bovik A.: “Image information and visual quality”; IEEE Transactions on Image Processing, 15, 2 (2006), 430-444. DOI: <https://dx.doi.org/10.1109/tip.2005.859378>
- [Shengyang et al. 2009] Shengyang Dai, Mei Han, Wei Xu, Ying Wu, Yihong Gong, Katsaggelos A.: “SoftCuts: A Soft Edge Smoothness Prior for Color Image Super-Resolution”; IEEE Transactions on Image Processing, 18, 5 (2009), 969-981. DOI: <https://dx.doi.org/10.1109/tip.2009.2012908>
- [Sung et al. 2003] Sung Cheol Park, Min Kyu Park, Moon Gi Kang: “Super-resolution image reconstruction: a technical overview”; IEEE Signal Process Mag, 20, 3 (2003), 21-36. DOI: <https://dx.doi.org/10.1109/msp.2003.1203207>

- [Tan and Le 2019] Tan M, Le Q.: “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks”; Proceedings of the 36th International Conference on Machine Learning, ICML, (2019). <http://proceedings.mlr.press/v97/tan19a.html> .
- [Timofte et al. 2014] Timofte R, De Smet V, Van Gool L.: “A+: Adjusted Anchored Neighborhood Regression for Fast Super-Resolution”; Computer Vision -- ACCV 2014, (2015), 111-126. DOI: https://dx.doi.org/10.1007/978-3-319-16817-3_8
- [Vidhya and Ramesh 2017] Vidhya G, Ramesh H.: “Effectiveness of Contrast Limited Adaptive Histogram Equalization Technique on Multispectral Satellite Imagery”; Proceedings of the International Conference on Video and Image Processing, (2017). DOI: <https://dx.doi.org/10.1145/3177404.3177409>
- [Wang et al. 2004] Wang Z, Bovik A, Sheikh H, Simoncelli E.: “Image Quality Assessment: From Error Visibility to Structural Similarity”; IEEE Transactions on Image Processing, 13, 4 (2004), 600-612. DOI: <https://dx.doi.org/10.1109/tip.2003.819861>
- [Yang et al. 2019] Yang W, Zhang X, Tian Y, Wang W, Xue J, Liao Q.: “Deep Learning for Single Image Super-Resolution: A Brief Review”; IEEE Trans Multimedia, 21, 12 (2019), 3106-3121. DOI: <https://dx.doi.org/10.1109/tmm.2019.2919431>
- [Zareapoor et al. 2019] Zareapoor M, Celebi M, Yang J.: “Diverse adversarial network for image super-resolution”; Signal Processing: Image Communication, 74 (2019), 191-200. DOI: <https://dx.doi.org/10.1016/j.image.2019.02.008>
- [Zhang et al. 2021] Zhang Y, Guo C, Zhao P.: “Medical Image Fusion Based on Low-Level Features”; Comput Math Methods Med, (2021), 1-13. DOI: <https://dx.doi.org/10.1155/2021/8798003>
- [Zhang et al. 2018(a)] Zhang Y, Tian Y, Kong Y, Zhong B, Fu Y.: “Residual Dense Network for Image Super-Resolution”; 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, (2018). DOI: <https://dx.doi.org/10.1109/cvpr.2018.00262>
- [Zhang et al. 2018(b)] Zhang, Yulun, et al.: “Image super-resolution using very deep residual channel attention networks”; Proceedings of the European conference on computer vision (ECCV), (2018). DOI: https://dx.doi.org/10.1007/978-3-030-01234-2_18
- [Zhu et al. 2020] Zhu X, Zhang L, Zhang L, Liu X, Shen Y, Zhao S.: “GAN-Based Image Super-Resolution with a Novel Quality Loss”; Math Probl Eng. 2020, (2020), 1-12. DOI: <https://dx.doi.org/10.1155/2020/5217429>