Low Resolution Image Enhancement Using Res-Net GAN

Arun Sai Narla, Shalini Kapuganti, Hathiram Nenavath

Department of ECE, Vardhaman College of Engineering, Hyderabad, India

Corresponding author: Arun Sai Narla, Email: arunsaiknr@gmail.com

Though Deeper Convolutional Neural Networks performs better in terms of speed and accuracy of SISR (Single Image Super Resolution), one essential challenge faced is to regain texturing details that are finer while resolving at a greater up-scaling factor. The objective function determines the characteristics of optimization-based super-resolution algorithms. The mean squared reconstruction error has been a major focus of recent research. The resultant approximates had excellent (PSNR) peak signal-to-noise ratio, but they frequently lack great-frequency features and are conceptually unsatisfactory and fall short of fidelity expected at greater resolution. Res-Net GAN is a typical generative adversarial network for super resolution (SR) of image, is presented in the study. Res-Net GAN, a framework which is apt of concluding photo realistic images at a 4x upscaling factors. Here, we present perceptual loss function that comprises of the adversarial loss as well as content loss to achieve this. Using only a discriminator network that was trained to discern among super resolved pictures as well as actual photo realistic pictures, adversarial loss drives solution to natural image manifold. Furthermore, rather than pixel space resemblance, we apply content loss that has been driven by perceptual resemblance. On available standards, deep residual network is capable of recovering photo realistic details from highly down sampled photos. Res-Net GAN exhibits tremendously substantial improvements in imperceptibility in a comprehensive mean-opinion-score (MOS) test. MOS scores achieved with Res-NetGAN were nearer to those achieved with the actual high-resolution pictures.

Keywords: Generative Adversarial Network (GAN), Super Resolution (SR), Mean Opinion Score (MOS)

2024. In Saroj Hiranwal & Garima Mathur (eds.), *Artificial Intelligence and Communication Technologies*, 1143–1151. Computing & Intelligent Systems, SCRS, India. https://doi.org/10.52458/978-81-955020-5-9-108

1. Introduction

The Process Super Resolution cites to the daunting task of predicting a High Resolution (HR) picture from its Low Resolution (LR)equivalent. Meanwhile, Super Resolution has gained a great deal of interest from the computer vision research field, as well as it has many applications. The SR point is especially ill-posed for high up-scaling factors because texture details in the reconstructed SR pictures is generally lacking. Mean squared error among recovered HR picture as well as ground truth was frequently optimization aim of supervised SR algorithms. This was useful since lowering MSE further optimizes peak signal-to-noise ratio (PSNR), which was a standard metric for evaluating and contrasting SR techniques. Furthermore, PSNR and MSE are established dependent on pixel-wise picture differences. Their capacity to detect perceptually meaningful differences like significant texture detail, is severely limited. We come up with a generative adversarial network (GAN), where we use deep residual network i.e., ResNet having skip-connection as the primary optimization objective and depart from MSE. Using high-level feature maps and a discriminator, we design a unique perceptual loss that helps solutions that are perceptually difficult to differentiate from the HR reference pictures.

2. Literature Survey

Liu Jing, Gan Zongliang and Zhu Xiuchang [1] proposed image super-resolution using directional bicubic interpolation. Because of its minimal complexity as well as reasonably decent outcomes, bicubic interpolation is a typical approach in the picture interpolation area. However, it interpolates only in the horizontal as well as vertical directions, where edges are prone for blocking, blurring, and ringing.

Seonja Kim, Dongsan Jun and Hunjoo Lee [2] used 2 light- weight neural networks which holds hybrid residual with a dense connection structure. Under presented approaches, trade-off among complexity of network and quality enhancement performance and have been estimated. When compared to earlier approaches, the suggested methods can greatly lower both the memory requirement and the inference speed to hold parameters as well as intermediate feature maps while keeping equivalent picture quality.

Na Sun and Huina Li [3] proposed a new algorithm which is the combination of traditional algorithms and deep learning. To begin, deep learning's capability to extract features automatically is used to undertake deeper reconstructions of low-resolution images. Deep learning technique was then used for training and learning, combined with standard interpolation reconstruction findings, to provide high resolution rebuilt data. The suggested approach outperforms the standard interpolation algorithm.

Kun Zeng, Jun Yu, Ruxn Wang and Dachen Tao [4] used data-driven model holding deep auto encoder (CDA) for a single image super resolution (SISR). The CDA has strong representational ability as well as it is built on a revolutionary deep architecture. CDA grasps inherent portrayals of LR, HR image patches and a data driven function which correctly connects various representations. The CDA's design implies that the concept is basic and adaptable. The auto-encoders guarantee that the intrinsic representations match LR, HR pictures well.

Wei-Shen Lai, Jia-Bin Hang and Narendra Ahuja [5] utilized a deep Laplacian Pyramid Super Resolution network for a faster as well as an efficient image super resolution (SR). At several pyramid levels, the suggested network rebuilds sub-band residuals of the high-resolution pictures. The suggested technique extracts feature directly from low resolution input space, using little computing resources. Recursive layers are utilized to exchange parameters within as well as across pyramid levels, lowering number of parameters dramatically. Network is constructed dependent on the pyramid framework.

Wenzh Shi, Jose Cabalero and Feren Huszar [6] used convolutional neural networks (CNN) that is apt of real time Super Resolution with 1080p videos over a K2 GPU. The presented CNN consists of feature-maps which were taken out from low resolution space. The last LR feature-maps are up-scaled into HR output using an effective sub pixel convolution layer that acquires an array of up-scaling filters. Handcrafted bicubic filter in SR pipeline is replaced by a greater number of up-scaling filters which are trained specifically for each and every feature map. By doing so, complexity of the total SR operation is reduced.

Yulun Zhang, Kunpeng Li and Yun Fu [7] proposed deep residual channel attention networks. To develop deep network, residual in residual structure was initiated that holds numerous residual blocks with long skip connections. There are numerous residual blocks with brief skip- connections for every residual group. Furthermore, RIR permits a huge amounts of low frequency data to be routed across numerous skip connections, permitting main network to concentrate acquiring high frequency data. A channel attention technique is presented for adaptively rescaling channel-wise features.

Greg Shakhnarvich and Norimich Ukita [8] used Deep Back-Projection Networks which uses down-sampling and up-sampling layers by supplying an error feedback approach at every stage. Each of the interconnected up- and down- sampling steps demonstrate different kinds of picture deterioration and high-resolution components. Features concatenation across up sampling and down sampling stages permits for improved super-resolution reconstruction, resulting in better outcomes.

Bee Lim, Sanghun Son and Hewon Kim [9] used enhanced deep super-resolution network with a magnificent achievement. Significant productivity of model's gain is associated to enhancement of traditional residual networks with ignoring unnecessary modules. Though training technique was stable, performance was enhanced through enlarging size of model. Authors offer latest multi-scale deeper SR system as well as a training procedure for rebuilding HR pictures with various up-scaling factors in single pattern.

Herminio Roman, Volodymr Ponomartov, Richard Fabi [10] proposed a technique that was dependent on an extra edge-safeguarding process as well as common interpolation among high frequency sub-band picture and input LR image using discrete wavelet transform (DWT). Finally, using Inverse DWT, total sub-band pictures were integrated for creating a new HR image. Various simulation findings show the unique SR framework outperforms previous approaches with regards to objective metrics as well as more subjective assessment of person eyesight in various pictures, enhancing resolution significantly.

Justin Jhonson, Alexandre Alahi, Li Fei-Fei [11] proposed high quality photos could be produced by developing as well as optimizing perceptual loss functions depended on high-level characteristics collected through pre-trained networks. They suggest using loss functions to train feed- forward networks for image enhancement tasks, which combines advantages of both approaches. They presented outputs on picture style transfer, in which a feed-forward network was trained for handling Gatys et al. enhancement issue in real-time.

Rott Saham and Tali Dekel [12] proposed SinGAN, generative model which could be trained from single natural picture unconditionally. Model was trained for recognizing internal distribution of patches inside a picture, as well as it can subsequently provide high-quality, varied samples with the similar visual content as image. SinGAN is made up with pyramid of fully convolutional GANs, each of which is in charge of grasping patch distribution with different picture scale.

Chao Dong and Xiaou Tang [13] used compact hourglass- shape CNN structure for quicker as well as best SR to speed up present SRCNN. The SRCNN architecture is redesigned primarily in three areas. First, a deconvolution layer is added to the network's end, and later mapping was learnt straight from given low-resolution picture to HR image. Secondly, they restructured mapping layer through decreasing input feature dimension prior to mapping as well as later extending it again afterward. They use lower filter sizes and greater mapping layers in third step.

Sachit Menon and Alexandru Damian [14] proposed a new approach to SR challenge depended on the creation of realistic super-resolution pictures that downscale appropriately. We offer PULSE (Photo Up-sampling through Latent Space Exploration), a revolutionary super- resolution technique that creates HR, realistic pictures at resolutions hitherto encountered in literature. That does in a completely self-supervised manner, as well as it isn't limited to a single degrading operator utilized during training. PULSE searches the HR natural image manifold, pictures which downscales to original LR picture, rather than beginning with LR picture as well as gradually increasing texture.

Dmitry Ulyanov and Victor Lempitsky [15] showed that Prior to any learning, the topology of generator network was adequate to record large amount of low-level picture measures. In common inverse tasks like denoising, super resolution, and in painting, a randomly-initialized neural network may be utilised as a constructed prior with great results. The same prior may also be used to diagnose deep neural representations and recover pictures based on flash-no f lash input pairings using the same prior. Apart from its many uses, the method emphasises how ordinary generator network topologies capture inductive bias.

Mengyuchu, Youxie and Laura Leal-TAIXE [16] suggested a self-supervised method that is temporally self-contained. Temporal adversarial learning is essential for both problems in order to get time lucid solutions despite losing spatial textures. They proposed new Ping-Pong loss for increasing temporal consistency across time. It efficiently avoids recurrent networks from amassing artefacts over time while maintaining detailed characteristics. They also developed a first set of measures for evaluating the temporal evolution's correctness and perceptual quality objectively.

Syed Waqas Zamir and Salman Khan [17] proposed a unique architecture with aim of keeping spatially accurate HR representations across network as well as getting significant contextual data from LR representations. Approach was built around a multi-scale residual block that includes (a) concurrent multi-resolution convolution flows to retrieve multi-scale characteristics, (b) exchange of information across multi-resolution flows, (c) spatial as well as channel attention processes to acquire context data (d) attention-depended multi-scale feature accumulation. To be short, the method learns expanded collection of characteristics from different scales while keeping high- resolution spatial details.

Jiahui Yu, Yuchen Fan and Jianchao Yang [18] revealed that models with broader features before ReLU activation had much superior performance of single picture super- resolution with the same parameters and computational budgets (SISR). In each residual block, the resulting SR residual network features a thin identity mapping route with broader channels before activation. They integrated linear low-rank convolution into SR networks to expand activation even further without adding computational complexity, resulting in even superior accuracy-efficiency tradeoffs.

Takeru Miyato and Masanori Koyama [19] suggested an innovative, projection-based method for incorporating conditional information into the discriminator of GANs while respecting the conditional information's function in the underlying probabilistic model. This differs from the majority of conditional GAN frameworks currently in use, which incorporate conditional information by concatenating the (embedded) conditional vector to the feature vectors.

Xintao Wang and Ke Yu [20] developed a new Video Restoration framework based on Enhanced Deformable Convolutions, dubbed EDVR. To accommodate huge movements, they created a Pyramid, Cascading, and Deformable (PCD) alignment module, where frame arrangement was made at the feature level utilizing coarse- to-fine deformable convolutions. Secondly, they presented fusion module for Temporal and Spatial Attention, where attention was applied both temporally as well as spatially.

3. Proposed Methodology

Generative Adversarial Network primarily holds a Generator as well as a Discriminator. Generator acts just like a counterfeiter where it tries to produce fake copies, where the Discriminator network acts like a cop to check whether the generated copies are fake or not (stated in Figure 2). So, this way of approach is chosen to be the suitable one for generating the Super Resolution (SR) images where several cycles of Generator and Discriminator operations are employed in order to obtain better results. GANs offer a strong architecture for creating natural looking pictures that has excellent picture quality that were realistic. GAN process pushes reconstructions to travel toward searching space regions that has a greater likelihood of having photorealistic pictures, bringing them nearer to original image manifold. The very first deep-ResNet framework employing the notion of GANs to construct perceptual loss function for photo- realistic SISR is described in this paper. Res-Net GAN, a GAN-dependent network that is developed to new perceptual loss. Here, we substitute MSE-based content loss with loss computed on VGG network feature maps, which are most even to pixel space changes.

Res-Net GAN is the modern state of the art for estimate of photo-realistic SR pictures with higher upscaling factors (x4), according to an exhaustive mean opinion score test on photos from data. Goal of SISR was to compute super- resolved ISR, high-resolution picture from low-resolution input image ILR. ILR is the low-resolution equivalent of IHR, which is the high-resolution counterpart. During training, high-resolution photos are available. ILR was procured during training through applying Gaussian filter over IHR as well as down-sampling it with a factor of the primary objective was to train generator function G which estimates HR counterpart of given LR input image. Furthermore, we use feed-forward CNN G parametrized by G to train a generator network. Weights as well as Biases of an L-layer deep-network are processed by developing an SR-specific loss function ISR, G = {W1: L; b1:γL} indicating weights as well as biases of an L-layer convolutional network. Architecture's overall purpose is to build the generative model G that has a goal of deceiving a differentiable discriminator D which has been taught for discriminating super-resolved pictures to that of genuine pictures. Generator could adapt to develop solutions which were very close to $^{\theta}$genuine photos as well as challenging for categorization by D using this method.

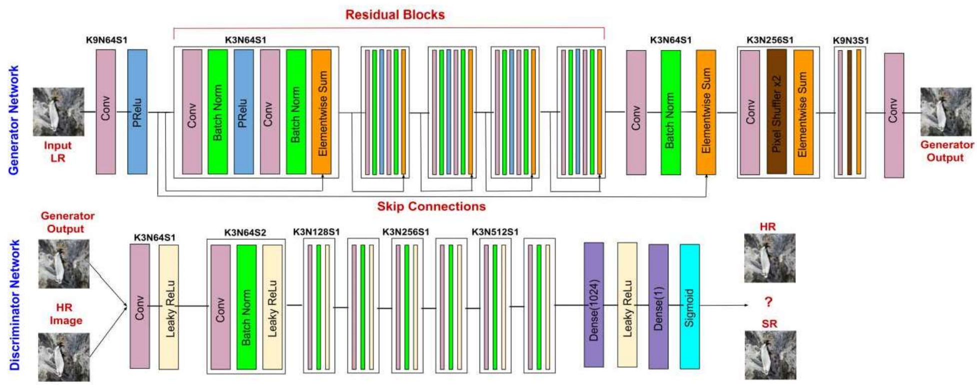

Fig. 1. Schematic diagram of Res-Net GAN

This promotes perceptually superior solutions to be found in the natural image's subspace, the manifold. In contradiction, SR solutions obtained through lessening pixel-wise error computations, including MSE, were obtained through reducing pixel-wise error computations. The Generator network is made up of identical residual blocks similar to [22], each of which has 2 convolutional layers, two batch normalization layers, one parametric ReLu, and an element wise adder (stated in Figure 1). We have employed 10 such residual blocks. The input Low Resolution (LR) image is first passed through convolutional layer (64 filters with size 9x9 and stride 1) and a parametric ReLu, which is a more sophisticated version of the leaky ReLu, before being sent through a sequence of residual blocks as described above. The convolutional layer of each residual block has 64 filters with a size of 3x3 and a stride of 1. Following that, two-pixel shufflers (x2) were implemented, thus quadrupling the input picture. Finally, a convolutional layer with three 3x3 filters and stride 1 is added. Discriminator network is trained to distinguish actual HR photos than that of produced SR data.

Throughout the network, we use LeakyαReLu (= 0.2) and prevent max pooling. To tackle maximizing problem, discriminator network was trained. As in VGG network, it consists of 8 convolutional layers by a rising number of 3x3 filter kernels, rising through factor of 2 from 64 to 512 kernels. When number of features were doubled, strided convolutions were used for lowering picture resolution. For producing a probability for sample classification, 512 feature maps were accompanied by 2 dense layers as well as a sigmoid activation function. The input for the Res-Net GAN network is a Low-Resolution image of size 96 x 96 where the output image size turns out into 4x up sampled one i.e., 384 x 384. Performance of the generator network was dependent on the perceptual loss function lSR, where Mean Square Error (MSE) is widely used to represent lSR. Sum of weights of content loss as well as adversarial loss component was utilised for calculating perceptual loss. ReLU activation layers of pre-trained 19-layer VGG model developed by Simonyan and Zisserman [21] are used to determine the VGGloss.

Size of various feature maps inside VGG model were demonstrated by Li:j and Bi:j. Furthermore, we add generative component of the Res-Net GAN with perceptual loss along with content losses as in [22], that pushes the model for favoring solutions which exist on manifold of original images through trying to deceive discriminator model.

Perceptual Loss:

$$l_{SR} = l_X^{SR} + 10^{-3} l_{Gen}^{SR} \quad (1)$$

Content Loss:

$$l_{VGG/i,j}^{SR} = \frac{1}{L_{i,j}B_{i,j}} \sum_{x=1}^{L_{i,j}} 1 \sum_{y=1}^{B_{i,j}} 1 \left(\Phi_{i,j}(I^{HR})_{x,y} - \Phi_{i,j}(G_{\Theta_G}(I^{LR}))_{x,y}\right)^2 \quad (2)$$

Adversarial Loss:

$$l_{Gen}^{SR} = \sum_{n=1}^{N} 1(-\log D_{\Theta_D}(G_{\Theta_G}(I^{LR}))) \quad (3)$$

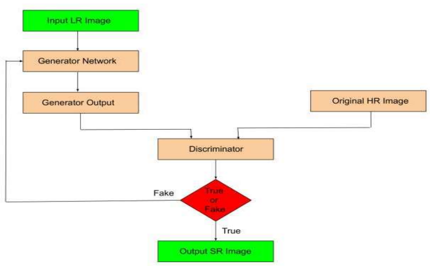

Fig. 2. Flow chart of the Res-Net GA

4. Experimental Results

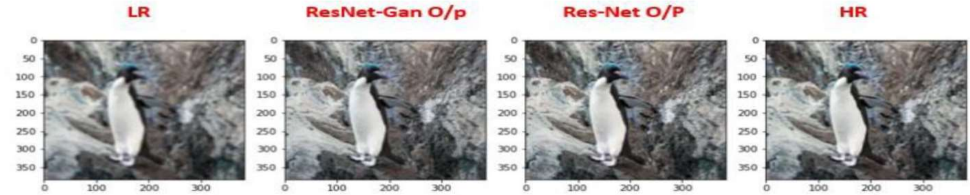

Fig.3. Results for an input image 1

Fig.4. Results for an input image 2

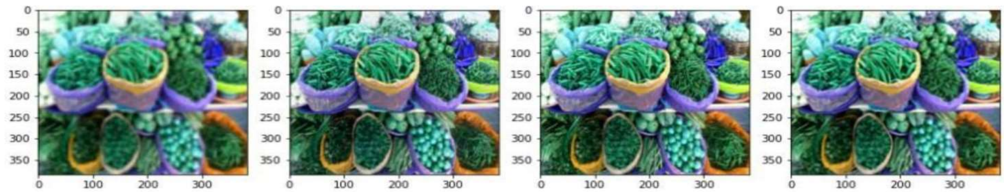

Fig.5. Results for an input image 3

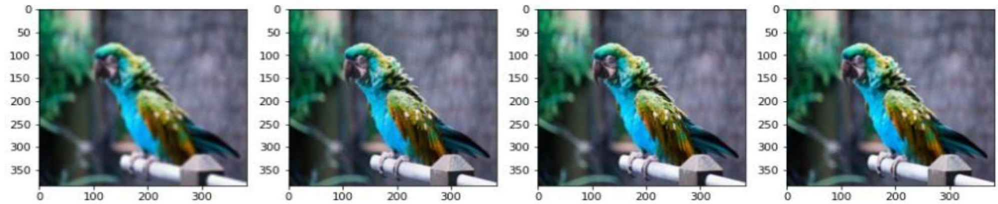

Fig.6. Results for an input image 4

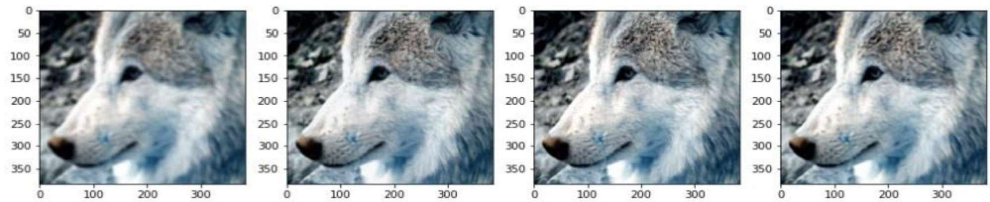

Fig.7. Results for an input image 5

Table 1: Comparison on 4 different parameters

S. No	PSNR		MSE		SSIM		MOS	
	Res-Net GAN	Res-Net	Res-Net GAN	Res-Net	Res-Net GAN	Res-Net	Res-Net GAN	Res-Net
1	28.942	29.625	0.830	0.709	0.525	0.681	3.535	2.357
2	29.885	30.735	0.668	0.549	0.689	0.773	3.268	2.949
3	28.740	29.327	0.869	0.759	0.604	0.712	4.019	3.269
4	33.609	35.553	0.283	0.182	0.863	0.926	3.943	2.782
5	29.817	30.770	0.678	0.545	0.652	0.749	3.678	2.545

A MOS on the generated images was used to evaluate the ability of various techniques to rebuild perceptually compelling pictures. We specifically requested 28 raters to award an integral score to the super-resolved pictures ranging from 1 (poor quality) to 5 (great quality). Here are the overall results.

5. Conclusion

We have described the Res-Net Generative Adversarial Network (GAN) and brought its advantage in the enhancement of a low-resolution image. The comparison is also made with Res-Net based on 4 different parameters to prove the superiority of Res-Net GAN. The perceptual quality of super-resolved pictures was the emphasis of this research rather than computational efficiency. Though Res- Net shows quite better performance when parameters PSNR, MSE and SSIM are considered, we concluded that Res-Net GAN reformations on large upscaling factors (4x) were much more photo-realistic (Cited in Fig. 3,4,5,6,7) compared to reconstructions generated with state-of-the-art approaches utilizing comprehensive MOS testing. For all the parameters compared above (Table 1), Res-Net GAN performs better over regular Res-Net.

References

[1] Jing, Zongliang and Xiuchang (2013).Directional Bicubic Interpolation - A New Method of Image Super-Resolution. ICMT-13.

[2] Kim, Jun and Lee (2021).Single Image Super-Resolution Method Using CNN- Based Lightweight Neural Networks. Applied Sciences.

[3] Na Sun and Huina Li (2019).Super Resolution Reconstruction of Images Based on Interpolation and Full Convolutional Neural Network and Application in Medical Fields, special section on deep learning algorithms for internet of medical things, IEEE Access.

[4] Zeng, Yu, Wang and Tao (2015). Coupled Deep Autoencoder for Single Image Super- Resolution Article, IEEE Transactions on Cybernetics.

[5] Lia, Hang and Ahuja (2018). Fast and Accurate Image Super-Resolution with Deep Laplacian Pyramid Networks, IEEE Transactions on Pattern Analysis and Machine Intelligence.

[6] Shi, Cabalero and Huszar (2016). Real-Time Single Image and Video Super-Resolution UsinganEfficientSub-PixelConvolutionalNeuralNetwork, Conference on Computer Vision and Pattern Recognition (CVPR).

[7] Zhang, Li and Fu (2018). Image Super-Resolution Using Very Deep Residual Channel Attention Networks, European Conference on Computer Vision (ECCV).

[8] Shakhnarvich and Ukita (2018). Deep Back-Projection Networks for Super-Resolution, Conference on Computer Vision and Pattern Recognition (CVPR).

[9] Lim, Son and Kim (2017). Enhanced Deep Residual Networks for Single Image Super-Resolution, IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW).

[10] Roman, Ponomartov and Fabi (2012). Image Super Resolution Using Interpolation and Edge Extraction in Wavelet Transform Space International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE).
[11] Jhonson, Alahi, Fei-Fei (2016). Perceptual Losses for Real-Time Style Transfer and Super-Resolution, CVPR.
[12] Saham and Dekel (2019). SinGAN: Learning a Generative Model from a Single Natural Image, CVPR.
[13] Dong and Tang (2016). Accelerating the Super-Resolution Convolutional Neural Network, CVPR.
[14] Menon and Damian (2020). PULSE: Self-Supervised Photo Up-sampling viaLatent Space Exploration of Generative Models, CVPR.
[15] Ulyanov and Lempitsky (2020). Deep Image Prior, CVPR.
[16] Mengyuchu and Touxie (2020). Learning Temporal Coherence via Self-Supervision for GAN-based Video Generation, CVPR.
[17] Zamir and Khan (2020). Learning Enriched Features for Real Image Restoration and Enhancement, ECCV.
[18] Yu, Fan and Yang (2018). Wide Activation for Efficient and Accurate Image Super-Resolution, CVPR.
[19] Miyato and Koyama (2018). CGAN's with projection discriminator, ICLR.
[20] Wang and Yu (2019). EDVR: Video Restoration with Enhanced Deformable Convolutional Networks, CVPR.
[21] Simonyan and Zisserman (2015). Very deep convolutional networks for large-scale image recognition.In International Conference on Learning Representations (ICLR).
[22] Ledig and Theis (2017). Photo-Realistic Single Image Super-Resolution Usinga Generative Adversarial Network, IEEE conference on Computer Vision and Pattern Recognition (CVPR).