FIRN:Fast Invertible Rescaling Net

Junjae Lee
qwopqwop200@gmail.com

Abstract

Invertible Rescaling Net (IRN) [1] modeled the downscaling and up-scaling process using Invertible Neural Networks (INN) [1,22,26] instead of upscaling to the traditional Single-image super resolution (SISR) method. As a result, it showed significantly improved performance than the previous method. However, apart from its high performance, IRN requires a lot of computation. hence, to improve this, we replace the existing dense block [2] with Pixel Attention Distillation Block (PADB). In addition, we use Charbonnier loss [43] instead of Mean Absolute Error (MAE) for the existing reconstruction loss. Through these improvements, we trade off the high performance and speed of the existing architecture and achieve higher performance than the lightweight SR model using the conventional method. In addition, by improving the perceptual loss and adversarial loss. we achieve perceptually satisfactory results than the model using the IRN+ method.

Key words: Deep Learning, Invertible Neural Networks, image super resolution, image processing

1 Introduction

SISR is a work of approximating high-resolution (HR) images based on low-resolution (LR) images. And since SRCNN [4], it has been evolved in the direction of using deep convolution neural network (CNN). In addition, SISR has been evolved in the direction of increasing Peak Signal-to-Noise Ratio (PSNR) [9,12,44,45,46]. However, there was a limit to the method of approximating HR image based on the LR image reduced by the traditional (e.g. bicubic [5]) method, and to solve this problem, methods of integrating the upscaling and downscaling of the image were proposed. [1,6,7,8] And IRN [1] succeeded in achieving high performance through an attempt to capture the characteristics of lost information by connecting two processes in this way. However, this model has the disadvantage of high computational cost apart from high performance.

Our first goal is to keep the high performance of the IRN as much as possible while reducing the computational cost. To do this, we replace the existing IRN with dense block [2] with PADB. In addition, we improve the existing restoration loss. Our second goal Is to produce perceptually satisfactory results than the model using the IRN+ method. For this purpose, perceptual loss and adversarial loss are improved

2 Related Work

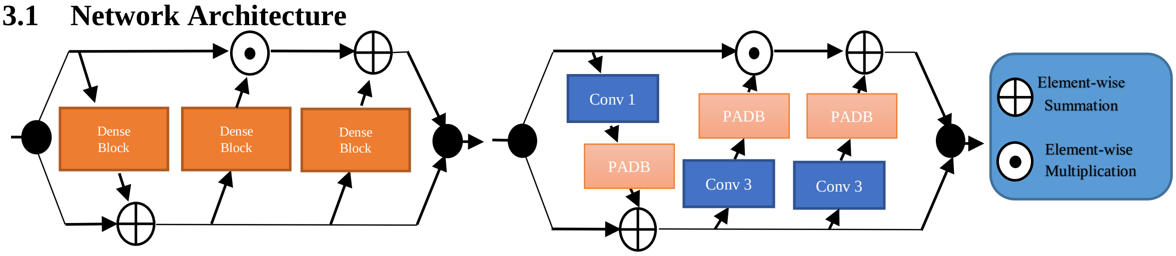
SRCNN [4] first proposed SISR through deep learning. And SRCNN outperforms the traditional method. various models that improve SRCNN appear. VDSR [9] improves performance by using deeper model, SRResNet [10] uses residual block [44] and SR–Densenet [11] used a dense block [2]. and RDN [12] showed high performance by combining the residual block and the dense block. However, these methods had the disadvantage of high computational load apart from the high performance. And while maintaining such high performance as much as possible., there have been many attempts to reduce computational load [13,14,15,16,17,38,43]. Among them, IMDN [15] showed high performance compared to low computational load. RFDN [16], which recently improved IMDN, and PAN [17] using pixel attention have been proposed.

Existing image downscaling methods tend to omit details and make images smooth. In order to solve this problem, various downscaling methods have been proposed [18,19,20,21,8]. Among them, the method of training the existing SISR model through the content adaptive resampler achieved higher performance than the existing method., However still has limitations [1]. Recently, IRN[1] used INN [1,22,26] to overcome these limitations and achieve high performance.

Usually, SR models are trained with the goal of increasing PSNR. However, it does not mean that an image with a high PSNR is a perceptually good image [10]. Hence, the existing SR model produced a perceptually poor image, and to solve this, SRGAN [10] using Generative adversarial network (GAN) [27] and perceptual loss [28]. ESRGAN [29] introduces a deeper architecture consisting of residual-in-residual dense blocks. It also improves SRGAN by using the feature map before the activation function in the VGG19-conv54 [45] and using a RaGAN [30]. CESRGAN [23] improves ESRGAN using Cosine Contextual (CCX) loss [24].

3 Proposed Methods

3.1 Network Architecture



(a) InvBlock using Dense block **(b)InvBlock using PADB**

Fig. 1: **Left:** InvBlock used in IRN. **Right:** InvBlock used in the proposed FIRN

In order to create a lighter model than the existing IRN [1], the original block is replaced with the proposed PADB. The structure of the invertible neural network blocks (InvBlock) remains as shown on the left in Fig. 1, but the existing dense block [2] is replaced with a new block as shown on the right in Fig. 1.

The proposed new block is based on residual feature distillation block [16] as shown in Fig. 2, replaces contrast-aware channel attention layer [15] with pixel attention block [17]. In addition, a convolution layer is used before PADB. The new InvBlock effectively reduces the number of parameters of the existing IRN. Models using this new InvBlock have fewer parameters than the existing IRN while maintaining the existing performance as much as possible. The model using this new InvBlock is called FIRN.

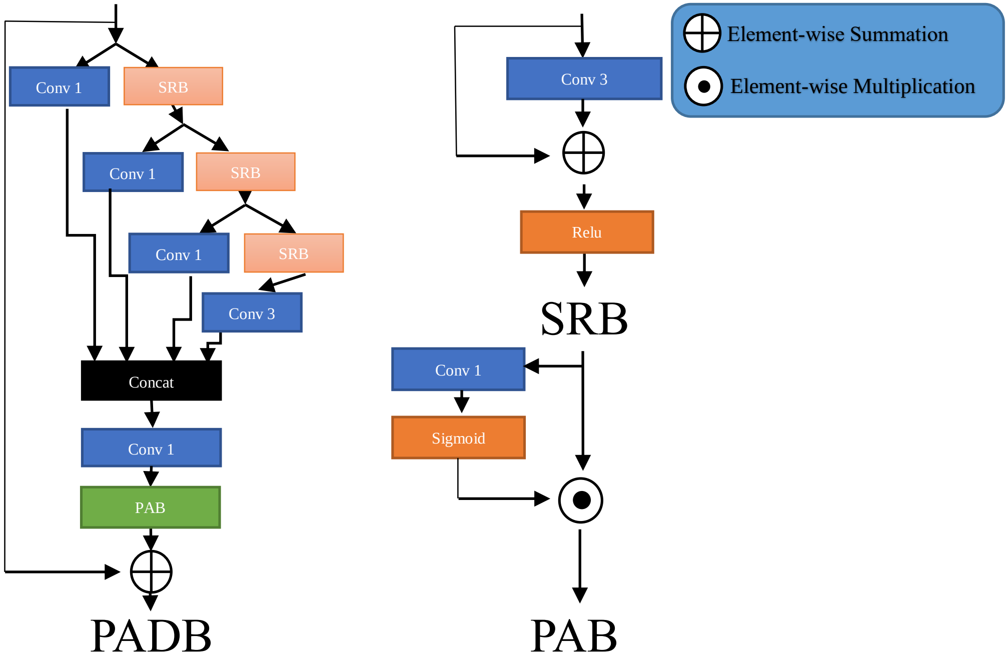


Fig. 2: PADB used in the proposed FIRN

3.2 Reconstruction loss

Previously, MAE was used as the restoration loss to measure the pixel-by-pixel similarity between the original image and the restored image, because the MAE achieved a higher PSNR than the mean square error.[1] However there is still room to improve the loss, so we use Charbonnier loss [43] as the reconstruction loss to achieve a higher PSNR. As shown in Tab. 2, when using the Charbonnier loss, a higher PSNR was achieved than MAE. The Charbonnier loss is defined as:

$$L_{recon}(\hat{y}, y) = \frac{1}{N}\sum_{i=1}^{N}\sum_{s=1}^{L}\rho\left(\hat{y}_s^{(i)} - y_s^{(i)}\right),$$
$$\rho(x) = \sqrt{x^2 + \varepsilon^2}, \tag{1}$$

where $\varepsilon = 10^{-6}$. In addition,y and $\hat{y}$ mean the original image and the generated image, respectively.

3.3 RaGAN

In the existing SR field, using GAN [27] tends to make the existing image more realistic [10]. And when using a RaGAN, it tends to make it more realistic than the existing GAN [29]. Therefore, FIRN+ uses the RaGAN [30]. The discriminator loss is then defined as:

$$L_D^{Ra} = -\mathrm{E}_y\left[\log\left(D_{Ra}(y,\hat{y})\right)\right] - \mathrm{E}_{\hat{y}}\left[\log\left(1 - D_{Ra}(\hat{y},y)\right)\right], \tag{2}$$

The adversarial loss for generator is then defined as:

$$L_G^{Ra} = -\mathrm{E}_y\left[\log\left(1 - D_{Ra}(y,\hat{y})\right)\right] - \mathrm{E}_{\hat{y}}\left[\log\left(D_{Ra}(\hat{y},y)\right)\right], \tag{3}$$

Where y and $\hat{y}$ mean the original image and the generated image, respectively. D and G mean discriminator and generator, respectively.

3.4 perceptual loss

perceptual loss [28] leads to more realistic images than existing content loss [10]. And, using feature maps pulled before the activation function in the VGG19-conv54 [45] generate more satisfactory results than after [29]. And using CCX loss [24] instead of existing perceptual loss generate more visually satisfactory results [23]. Therefore, for FIRN+, we use CCX loss. CCX loss is defined as:

$$L_{CCX}(\phi(\hat{y}),\phi(y)) = -\log\left(\frac{1}{N}\sum_j max_i \mathrm{A}_{ij}\right),$$
$$\mathrm{A}_{ij} = \frac{\mathrm{e}^{(1-d_{ij}^{'}/h)}}{\sum_k \mathrm{e}^{(1-d_{ik}^{'}/h)}},$$
$$d_{ij}^{'} = \frac{d_{ij}}{min_k d_{ik} + \varepsilon},$$
$$d_{ij} = \frac{(x_i - r) \cdot (y_j - r)}{||x_i - r||_2 \times ||y_j - r||_2}, \tag{4}$$

where $h > 0$ is a bandwidth parameter, $\varepsilon = 10^{-6}$,and $\mathrm{r} = \frac{1}{N}\sum_j y_j$.$y$ and $\hat{y}$ mean the original image and the generated image, respectively. ϕ denotes the feature mapping from before the activation function in the VGG19-conv54.

the overall loss of FIRN+ is defined as:

$$L_{FIRN+} = \lambda_1 L_{recon} + \lambda_2 L_{guide} + \lambda_3 L_G^{Ra} + \lambda_4 L_{CCX}, \tag{5}$$

Where$\lambda_1, \lambda_2, \lambda_3, \lambda_4$ are coefficients for balancing different loss conditions. In addition where L_{guide} is defined as:

$$L_{guide}(\hat{y}_{LR}, y_{LR}) = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_{LR} - y_{LR})^2, \tag{6}$$

Where$y_{LR}$and$\hat{y}_{LR}$are the original LR image and the generated LR image, respectively.

4 Experiments

4.1 Data

The training set used uses DF2K, which incorporates DIV2K [31] and Flickr2K [32]. DF2K contains 3450 high- quality 2k resolution images. In addition, four data set for evaluation (Set5[33],Set14 [34],BSD100[35],Urban100[36]) are used. Using this data set, PSNR and SSIM [20] are evaluate in the Y channel displayed in the YCbCR (Y, Cb, Cr) color space.

4.2 Training Details

We train and compare model using two downscaling modules. downscaling module has 8 InvBlocks. The size of the mini-batch is 16. We crop 144 x 144 HR sub images. The cropped image is augmented by random horizontal flips. When learning FIRN, the learning rate is repeated 50k times at 2×10^{-4}. FIRN+ is based pre-train FIRN, The learning rate is 1×10^{-4} and it repeats 400k times. Pre-training the discriminator 5k times. In addition, at [50k,100k,200k,300k], it reduces the learning rate by half. The discriminator is following to [1]. The model is optimized using Adam[37] with β1 = 0.9 and β2 = 0.999. Implement the model with the PyTorch framework and train it using NVIDIA Tesla v100 GPU.

4.3 Results

Table 1: Quantitative evaluation results of various downscaling and upscaling methods for image reconstruction for data set Set5, Set14, BSD100, and Urban100 (PSNR/SSIM): Red/Blue : Best/Second Best

Upscaling&Downscaling	Scale	Param	Set5	Set14	B100	Urban100
Bicubic&Bicubic	4x	/	28.42/0.8104	26.00/0.7027	25.96/0.667	23.14/0.6577
SRCNN&Bicubic	4x	0.05M	30.48/0.8628	27.50/0.7513	26.90/0.7101	24.52/0.7221
FSRCNN&Bicubic	4x	0.01M	30.71/0.8657	27.61/0.7550	26.98/0.7150	24.62/0.7280
SRResNet&Bicubic	4x	1.51M	32.17/0.8951	28.61/0.7823	27.59/0.7365	26.12/0.7871
EDSR&Bicubic	4x	40.7M	32.62/0.8984	28.94/0.7901	27.79/0.7437	26.86/0.8080
RCAN&Bicubic	4x	15.6M	32.63/0.9002	28.87/0.7889	27.77/0.7436	26.82/0.8087
ESRGAN&Bicubic	4x	16.3M	32.74/0.9012	29.00/0.7915	27.84/0.7455	27.03/0.8152
SAN&Bicubic	4x	15.8M	32.64/0.9003	28.92/0.7888	27.78/0.7436	26.79/0.8068
RDN&Bicubic	4x	22.2M	32.47/0.8990	28.81/0.7871	27.72/0.7419	26.61/0.8028
VDSR&Bicubic	4x	0.66M	31.35/0.8838	28.01/0.7674	27.29/0.7251	25.18/0.7524
DRCN&Bicubic	4x	1.77M	31.53/0.8854	28.02/0.7670	27.23/0.7233	25.14/0.7510
LapSRN&Bicubic	4x	0.81M	31.54/0.8850	29.19/0.7720	27.32/0.7280	25.21/0.7560
IDN&Bicubic	4x	0.55M	31.82/0.8903	28.25/0.7730	27.41/0.7297	25.41/0.7632
MemNet&Bicubic	4x	0.67M	31.74/0.8893	28.26/0.7723	27.40/0.7281	25.50/0.7630
IMDN&Bicubic	4x	0.71M	32.21/0.8948	28.58/0.7811	27.56/0.7353	26.04/0.7838
RFDN&Bicubic	4x	0.55M	32.24/0.8952	28.61/0.7819	27.57/0.7360	26.11/0.7858
PAN&Bicubic	4x	0.27M	32.13/0.8948	28.61/0.7822	27.59/0.7363	26.11/0.7854
EDSR&CAR	4x	52.8M	33.88/0.9174	30.31/0.8382	29.15/0.8001	29.28/0.8711
IRN	4x	4.35M	36.19/0.9451	32.67/0.9015	31.64/0.8826	31.41/0.9157
FIRN(ours)	4x	0.32M	33.55/0.9189	30.01/0.8452	29.56/0.8301	27.43/0.8406

We compare the proposed FIRN with various SR methods for upscaling factor x4, including SRCNN[4],FSRCNN[38],SRResNet[10],EDSR[39],RCAN[40],ESRGAN[29],SAN[41],RDN[12],VDSR[9],DRCN[42],LapSRN[43],IDN[13],MemNet[14],IMDN[15] ,RFDN[16],PAN[17], CAR[8] and IRN[1]. It can be seen from the Tab. 1 that the proposed FIRN has a high performance while using a parameter of 320k.

4.4 Ablation Study

Table 2: Performance comparison by type of restoration loss (MAE, Charbonnier), red indicates the best performance.

	Set5	Set14	B100	Urban100
MAE	33.51/0.9190	29.98/0.8454	29.55/0.8300	27.39/0.8399
Charbonnier	33.55/0.9189	30.01/0.8452	29.56/0.8301	27.43/0.8406

An experiment is conducted to analyze the performance according to the restoration loss. As shown in Tab. 1, FIRN has higher PSNR when using Charbonnier loss than when using MAE loss.

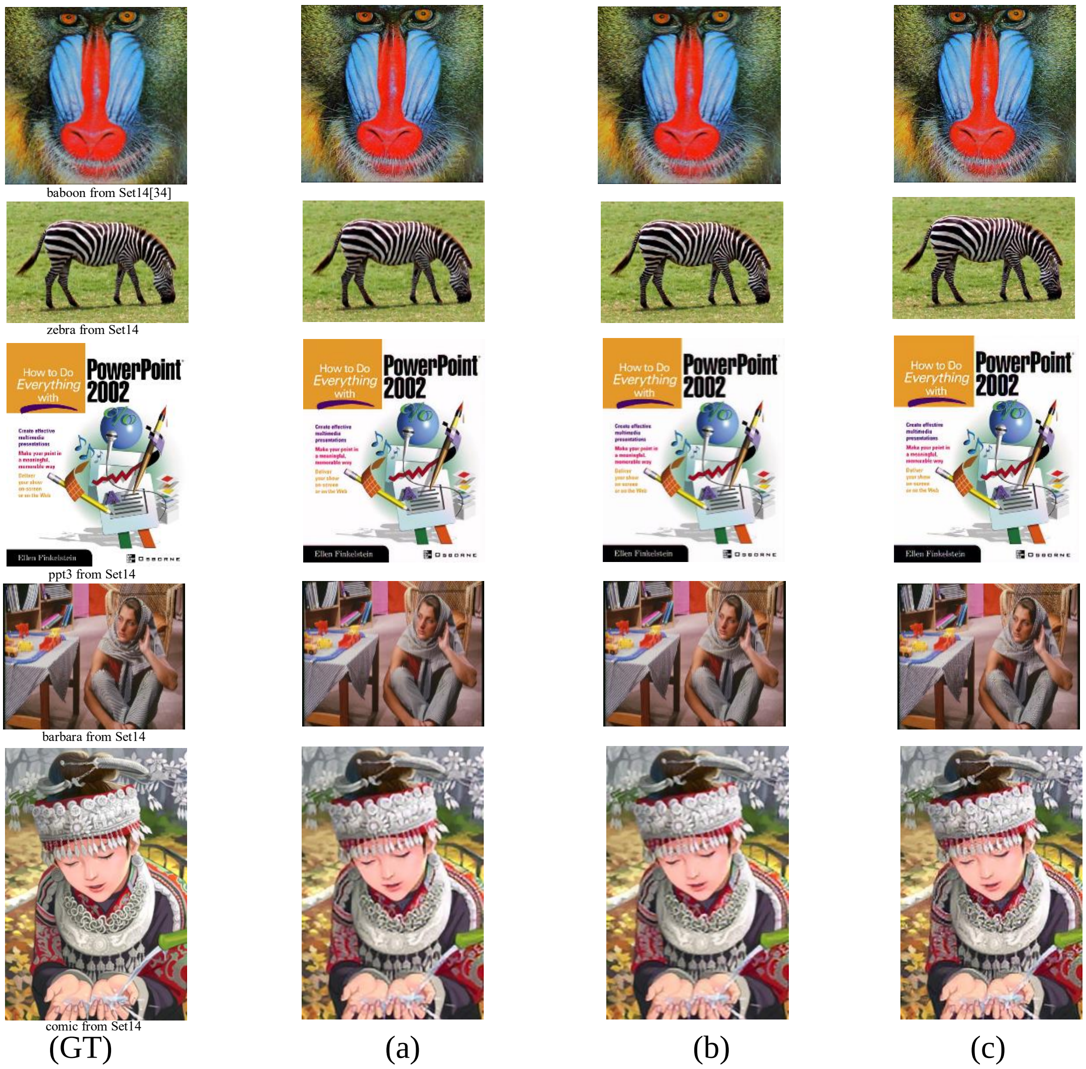
In order to study the effect of each component in the proposed FIRN+, we correct several losses in FIRN applied the method of IRN+ and compare the effects. The overall visual comparison is illustrated in Fig. 3.A detailed discussion is provided as follows.

If CCX was used instead of existing perceptual loss, the texture became sharper than before, as observed by CESRGAN, resulting in satisfactory results.

Similar to the observation result of ESRGAN by applying a RaGAN, the texture becomes sharper and more visually satisfactory than before.

Fig. 3: Overall visual comparison to show the effect of each component in FIRN. Each model was compared after 100k times training based on pre-trained FIRN.

Setting	(a)	(b)	(c)
perceptual loss?	L1	CCX	CCX
GAN?	Standard Gan	Standard Gan	RaGan



As you can see from Fig. 3, improving losses can give you more visually satisfactory results than before.

5 Conclusion

We propose a lighter FIRN while maintaining the performance of the existing IRN as much as possible. We introduced PADB to make the network lighter. In addition, Charbonnier loss is used as the reconstruction loss. we succeeded in creating a lighter network while maintaining the existing performance as much as possible.

FIRN+ improves the existing perceptual loss and adversarial loss. Through these improvements, it generate a more perceptually satisfactory image than FIRN using existing method.

6 References

1. Xiao, Mingqing, et al. "Invertible Image Rescaling." arXiv preprint arXiv:2005.05650 (2020).
2. Huang, Gao, et al. "Densely connected convolutional networks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.
3. Rakotonirina, Nathanaël Carraz, and Andry Rasoanaivo. "ESRGAN+: Further Improving Enhanced Super-Resolution Generative Adversarial Network." ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2020.
4. Dong, Chao, et al. "Image super-resolution using deep convolutional networks." IEEE transactions on pattern analysis and machine intelligence 38.2 (2015): 295-307.
5. Keys, Robert. "Cubic convolution interpolation for digital image processing." IEEE transactions on acoustics, speech, and signal processing 29.6 (1981): 1153-1160.
6. Kim, Heewon, et al. "Task-aware image downscaling." Proceedings of the European Conference on Computer Vision (ECCV). 2018.
7. Li, Yue, et al. "Learning a convolutional neural network for image compact-resolution." IEEE Transactions on Image Processing 28.3 (2018): 1092-1107.
8. Sun, Wanjie, and Zhenzhong Chen. "Learned image downscaling for upscaling using content adaptive resampler." IEEE Transactions on Image Processing 29 (2020): 4027-4040.
9. Kim, Jiwon, Jung Kwon Lee, and Kyoung Mu Lee. "Accurate image super-resolution using very deep convolutional networks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
10. Ledig, Christian, et al. "Photo-realistic single image super-resolution using a generative adversarial network." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.
11. Tong, Tong, et al. "Image super-resolution using dense skip connections." Proceedings of the IEEE International Conference on Computer Vision. 2017.
12. Zhang, Yulun, et al. "Residual dense network for image super-resolution." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.
13. Hui, Zheng, Xiumei Wang, and Xinbo Gao. "Fast and accurate single image super-resolution via information distillation network." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.
14. Tai, Ying, et al. "Memnet: A persistent memory network for image restoration." Proceedings of the IEEE international conference on computer vision. 2017.
15. Hui, Zheng, et al. "Lightweight image super-resolution with information multi-distillation network." Proceedings of the 27th ACM International Conference on Multimedia. 2019.
16. Liu, Jie, Jie Tang, and Gangshan Wu. "Residual Feature Distillation Network for Lightweight Image Super-Resolution." arXiv preprint arXiv:2009.11551 (2020).
17. Zhao, Hengyuan, et al. "Efficient Image Super-Resolution Using Pixel Attention." arXiv preprint arXiv:2010.01073 (2020).
18. Kopf, Johannes, Ariel Shamir, and Pieter Peers. "Content-adaptive image downscaling." ACM Transactions on Graphics (TOG) 32.6 (2013): 1-8.
19. Oeztireli, A. Cengiz, and Markus Gross. "Perceptually based downscaling of images." ACM Transactions on Graphics (TOG) 34.4 (2015): 1-10.
20. Wang, Zhou, et al. "Image quality assessment: from error visibility to structural

similarity." IEEE transactions on image processing 13.4 (2004): 600-612.

21. Weber, Nicolas, et al. "Rapid, detail-preserving image downscaling." ACM Transactions on Graphics (TOG) 35.6 (2016): 1-6.
22. Ardizzone, Lynton, et al. "Analyzing inverse problems with invertible neural networks." arXiv preprint arXiv:1808.04730 (2018).
23. Zhong, Sheng, and Shifu Zhou. "Optimizing Generative Adversarial Networks for Image Super Resolution via Latent Space Regularization." arXiv preprint arXiv:2001.08126 (2020).
24. Mechrez, Roey, Itamar Talmi, and Lihi Zelnik-Manor. "The contextual loss for image transformation with non-aligned data." Proceedings of the European Conference on Computer Vision (ECCV). 2018.
25. Gatys, Leon, Alexander S. Ecker, and Matthias Bethge. "Texture synthesis using convolutional neural networks." Advances in neural information processing systems. 2015.
26. Behrmann, Jens, et al. "Invertible residual networks." International Conference on Machine Learning. 2019.
27. Goodfellow, Ian, et al. "Generative adversarial nets." Advances in neural information processing systems. 2014.
28. Johnson, Justin, Alexandre Alahi, and Li Fei-Fei. "Perceptual losses for real-time style transfer and super-resolution." European conference on computer vision. Springer, Cham, 2016.
29. Wang, Xintao, et al. "Esrgan: Enhanced super-resolution generative adversarial networks." Proceedings of the European Conference on Computer Vision (ECCV). 2018.
30. Jolicoeur-Martineau, Alexia. "The relativistic discriminator: a key element missing from standard GAN." arXiv preprint arXiv:1807.00734 (2018).
31. Agustsson, Eirikur, and Radu Timofte. "Ntire 2017 challenge on single image super-resolution: Dataset and study." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2017.
32. Timofte, Radu, et al. "Ntire 2017 challenge on single image super-resolution: Methods and results." Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2017.
33. Bevilacqua, Marco, et al. "Low-complexity single-image super-resolution based on nonnegative neighbor embedding." (2012): 135-1.
34. Zeyde, Roman, Michael Elad, and Matan Protter. "On single image scale-up using sparse-representations." International conference on curves and surfaces. Springer, Berlin, Heidelberg, 2010.
35. Martin, David, et al. "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics." Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001. Vol. 2. IEEE, 2001.
36. Huang, Jia-Bin, Abhishek Singh, and Narendra Ahuja. "Single image super-resolution from transformed self-exemplars." Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.
37. Kingma, Diederik P., and Jimmy Ba. "Adam: A method for stochastic optimization." arXiv preprint arXiv:1412.6980 (2014).
38. Dong, Chao, Chen Change Loy, and Xiaoou Tang. "Accelerating the super-resolution convolutional neural network." European conference on computer vision. Springer, Cham, 2016.
39. 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.
40. Zhang, Yulun, et al. "Image super-resolution using very deep residual channel attention networks." Proceedings of the European Conference on Computer Vision (ECCV). 2018.
41. Dai, Tao, et al. "Second-order attention network for single image super-resolution." Proceedings of the IEEE conference on computer vision and pattern recognition. 2019.
42. Kim, Jiwon, Jung Kwon Lee, and Kyoung Mu Lee. "Deeply-recursive convolutional network for image super-resolution." Proceedings of the IEEE conference on computer vision and

pattern recognition. 2016.

43. Lai, Wei-Sheng, et al. "Fast and accurate image super-resolution with deep laplacian pyramid networks." IEEE transactions on pattern analysis and machine intelligence 41.11 (2018): 2599-2613.
44. Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014).
45. He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.