Deep Learning Based Approach Implemented to Image Super-Resolution

Thuong Le-Tien, Tuan Nguyen-Thanh, Hanh-Phan Xuan, Giang Nguyen-Truong, and Vinh Ta-Quoc
Ho Chi Minh City University of Technology (HCMUT), 268 Ly Thuong Kiet Street, District 10, Ho Chi Minh City, Vietnam
Vietnam National University Ho Chi Minh City (VNU-HCM), Linh Trung Ward, Thu Duc District, Ho Chi Minh City, Vietnam
Email: {thuongle, nttuan}@hcmut.edu.vn, {phantyp, ntg2208, taquocvinhbk10}@gmail.com

***Abstract*—The aim of this research is about application of deep learning approach to the inverse problem, which is one of the most popular issues that has been concerned for many years about, the image Super-Resolution (SR). From then on, many fields of machine learning and deep learning have gained a lot of momentum in solving such imaging problems. In this article, we review the deep-learning techniques for solving the image super-resolution especially about the Generative Adversarial Network (GAN) technique and discuss other ways to use the GAN for an efficient solution on the task. More specifically, we review about the Enhanced Super-Resolution Generative Adversarial Network (ESRGAN) and Residual in Residual Dense Network (RRDN) that are introduced by 'idealo' team and evaluate their results for image SR, they had generated precise results that gained the high rank on the leader board of state-of-the-art techniques with many other datasets like Set5, Set14 or DIV2K, etc. To be more specific, we will also review the Single-Image Super-Resolution using Generative Adversarial Network (SRGAN) and the Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN), two famous state-of-the-art techniques, by re-train the proposed model with different parameter and comparing with their result. So that can be helping us understand the working of announced model and the different when we choose others parameter compared to theirs.**

***Index Terms*—image super-resolution, deep learning, inverse problems, Residual in Residual Dense Network (RRDN), Generative Adversarial Network (GAN), Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN)**

I. Introduction

The fundamental concept of Artificial Intelligence (AI), including Machine Learning and Deep Learning, has been implemented to various fields gained many achievements. Unlike analytical methods for which the problem is explicitly defined and domain-knowledge carefully engineered into solution, Deep Neural Networks (DNNs) do not benefit from such prior know instead make use of large data sets to learn the unknown solution to the inverse problem. The inverse problem has defined by Lucas *et al.* 2018 [1] that has a long history of research and development. These inverse problems are also known as recovery problems: restoration, deconvolution, inpainting, reconstruction from projections, compressive sensing, Super-Resolution (SR), etc. The solutions for inverse problems have been raised for many years with many different methods. But with the support of Neural Network recently, the solutions are more specific and efficient in the way that DNNs has gave us. In the deep-learning literature, such modules are commonly referred to as layers, and each layer is composed of multiple units or neurons, that makes the model easier to extract the feature of the input for solving the data that relate to ground truth. Super-resolution is the process of recovering a High Resolution (HR) image from a given Low Resolution (LR) image [2]. An image may have a "lower resolution" due to a smaller spatial resolution or due to a result of degradation. When we use the degradation, function is scale down the image resolution, we obtain the LR data from the HR. The inverse problem is the solution that make the LR back to HR as much similarly as possible. Deep Learning techniques have proven to be effective for Super Resolution by given a method that was trained times by times with the dataset concludes HR and LR that are downscaled from HR. There has been much studies in recent decades, with significant progress results with many methods. In the demand of human about the super-resolution image nowadays is very important because of the detail of information in many fields of science. A better detail of the images, the more feature that we get. Such as the television every day we watch required more and more detail with quality of vision about 4K or more than that. The better the image quality, the more comfortable our eyes will be and the less myopia will occur. Another vision that image super-resolution has the effect on is in medical, whether the doctor can recognize exactly the size of the tumor or other pathological. That need the exact detail in order to give the precise decision. Not only these fields, but also another like in agriculture, military and traffic for example. In China, recently, they have taken a picture 195 gigapixels from satellite by quantum technology which is from space can zoom scale up to the street in the Shanghai's city with every single particular detail about the sign of the cars, the numbers or

Manuscript received May 22, 2020; revised September 29, 2020.

doi: 10.12720/jait.11.4.209-216

even the human's faces. In Vietnam recently, by photographer Binh Bui had showed down the picture about Hanoi capital with 13 gigapixels [3]. With the crucial of the super-resolution images, many methods were established to generate image SR.

II. Related Works

The Image Super-resolution processing that are considered as two type: Single Image Super-Resolution (SISR) and Multiple Image or multi-frame images. In this paper, we focus on the SISR with techniques used DL Network that will be discussed in detail. For the construction image super-resolution methods, Zhu *et al.* [4] has introduced about the reconstruction image SR based on sparse representation via direction and edge, direction. With SISR the experimental setting that has achieved much better PSNR and SSIM index than bicubic and NCSR. With the same method Jiang *et al.* [5] applied for image in medical fields, CT images and proved that their method was effectively improve the resolution of a single CT image. In 2007, Sinh Nguyen L. H. *et al.* [6] made an experiment about the heterogeneous interpolation and filter to reconstruct image HR in several of interpolation methods. Another SISR that is about the combination of frequency domain and wavelet domain to reconstruct the image super-resolution of Thuong Le-Tien *et al.* [7] in 2009. On the other hand of reconstruction image SR, Deep Learning method has been experienced with many significant results. Recently, Yang *et al.* [8] has introduced about the SRCNN for upscaling LR image into HR image and gained good result, they also reported that the acceleration of deep models and extensive comprehension of deep models and the criteria for designing and evaluation the objective functions are their challenges for optimizing their model. Similarly, Kawulof *et al.* [9] has applied Deep learning for Multi-frame Image SR and gained better results. In the appearance and improvement of Generative Adversarial Network (GAN) by Goodfellow *et al.* [10], many GAN methods were raised and gained better results for SISR. In 2017, Ledig *et al.* [11] presented SRGAN – the first framework capable of inferring photo-realistic natural images for 4x upscaling factors. With the achievement that performance on different dataset, they made a comparison of their model with PSNR or SSIM and other metrics. However, their limitation is that PSNR and SSIM fail to capture and accurate assess image quality with respect to the human visual system and confirmed that SRGAN reconstructions for large upscale factor are, by a considerable margin, more photo-realistic than reconstructions obtained with state-of-the-art reference methods. Bingzhe Wu *et al.* [12] introduced about the SRPGAN that contains perceptual loss based on the discriminator of the built SRPGAN and used the Charbonnier loss function to build the content loss and combine it with the proposed perceptual loss and the adversarial loss. They had mad comparison with SRGAN and show that their model has a better job with the same scaling factor of 4x. The limitation of their method is constructed images have checkerboard artifacts at the pixel level. In March 2018 Zhang *et al.* [13] represented about the model that has the Residual Dense Block (RBD) for feature extraction capture and made Residual Dense Network to generate image and calculated loss. That extensive benchmark evaluation well demonstrates their model achieves superiority over state-of-the-art. In September of 2018, with the RDN, Xintao Wang *et al.* [14] made Residual in Residual Dense Network (RRDN) and Enhanced SR Network (ESRGAN) that gained state-of-the-art that generated better results than SRGAN. In this paper, we review about the RRDN technique and evaluate the result that experienced.

III. Method

A. Algorithm

Fig. 1 is the algorithm flow chart; the Low-resolution images will be fed over the generator to create the Super-resolution images output. Then the SR images will be distinguished with HR image from the ground truth to find whether the SR image is well generated closer to the ground truth or not.

The loss is calculated by the feature maps of the SR passed through the discriminator and the other is the HR image will be extracted the feature maps by the VGG model. The loss will be then updated for generator to adjust the SR images to be as the most similar as the HR. The losses of the generator and the discriminator are then back propagation for update the process. At the final, these losses will gain the value of the middle means that the discriminator cannot distinguish the generated images and the generator is also converged.

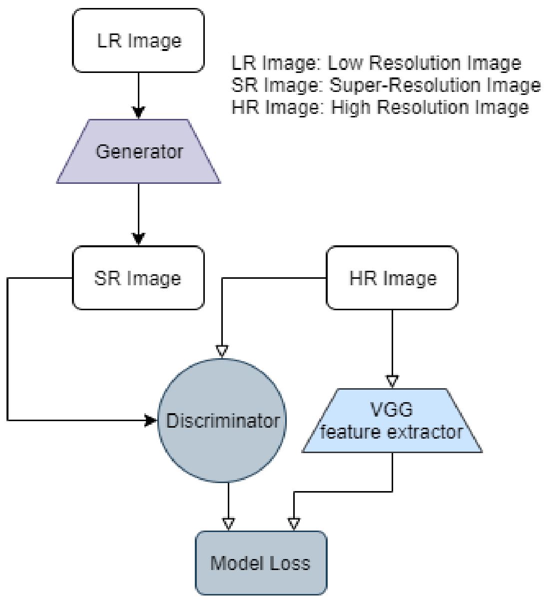

Figure 1. The algorithm flow chart for GAN process to generate the super-resolution images.

B. Network Architecture

In other to improve recovered image quality of SRGAN, Wang [14] had introduced about the Residual in Residual Dense Block (RRDB) with removal all BN layers that proved the increase performance and reduce computational complexity in different PSNR-oriented tasks including SR and deblurring. BN layers normalize the features using mean and variance in a batch during training and use estimated mean and variance of the whole training dataset during testing. As shown in Fig. 2, the RRDN mainly consists four parts: Shallow Feature Extraction (SFENet), Residual Dense Blocks (RDBs), Dense Feature Fusion (DFF) and finally the up-sampling net (UPNet). By replacing RDB block with RRDB block that Wang [14] has introduced in their method. The RRDB contains dense connected layers, Local Feature Fusion (LFF) and local residual learning, leading to a contiguous memory mechanism.

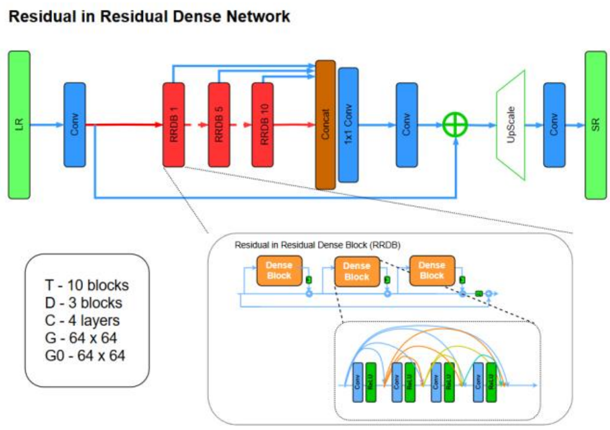

Figure 2. Residual in Residual Dense Block (RRDB) with BN removal [14].

C. Residual Dense Blocks

Local feature fusion: adaptively fuse the states from proceeding RRDB and whole Conv layers in current RRDB.

$$F_{d,LF} = H_{LFF}^{d}\left(\left[F_{d-1}, F_{d,1}, \dots, F_{d,c}, \dots, F_{d,C}\right]\right) \quad (1)$$

where H_{LFF}^{d} denotes the function of the 1 x 1 Conv layer in the d-th RRDB, $F_{d-1}, F_{d,1}, etc$ is input and output of the d-th RDB respectively.

Local residual learning: is introduced in RDB [11] to further improve the information flow. The final output of the d-th RDB can be obtained:

$$F_d = F_{d-1} + F_{d,LF} \quad (2)$$

D. Dense Feature Fusion

Global feature fusion: is proposed as shown in Fig. 3 to extract the global feature F_{GF} by using features from all the RDBs:

$$F_{d,LF} = H_{GFF}([F_1, \dots, F_D]) \quad (3)$$

where $[F_1, \dots, F_D]$refers to the concatenation of feature-maps produced by RDB 1,…,D.

Global residual learning: is then utilized to obtain the feature-maps before conducting up-scaling by:

$$F_{DF} = F_{-1} + F_{GF} \quad (4)$$

where F_{-1} denotes the shallow feature-maps and F_{GF} is further adaptive fused to form.

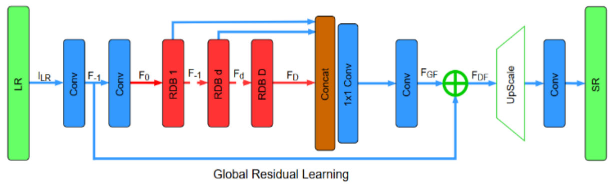

Figure 3. The architecture of our proposed Residual Dense Network (RDN) [11].

E. Relativistic Discrimination

Beside the improved structure of generator, Wang *et al.* [14] also enhance the discriminator based on the Relativistic GAN [15] that can be seen in Fig. 4.

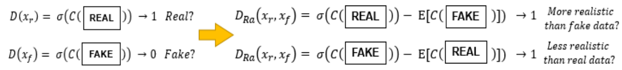

Figure 4. Difference between standard discriminator (left) and relativistic discriminator (right).

The discriminator loss is then defined as:

$$L_D^{Ra} = -E_{x_r}\left[\log\left(D_{Ra}\left(x_r, x_f\right)\right)\right] - E_{x_f}\left[\log\left(1 - D_{Ra}\left(x_f, x_r\right)\right)\right] \quad (5)$$

The adversarial loss for generator is in a symmetrical form:

$$L_G^{Ra} = -E_{x_r}\left[\log\left(1 - D_{Ra}\left(x_r, x_f\right)\right)\right] - E_{x_f}\left[\log\left(D_{Ra}\left(x_f, x_r\right)\right)\right] \quad (6)$$

where $x_f = G(x_i)$ and x_i stand for the input LR image. It observed that the adversarial loss for generator contains both x_r and x_f. Therefore, the generator benefits from the gradients from both generated data and real data in adversarial training.

F. Perceptual Loss

They had also developed a more effective perceptual loss $L_{perceptual}$ by constraining on features before activation rather than after activation as practiced in SRGAN.

The total loss for the generator is:

$$L_G = L_{percep} + \lambda L_G^{Ra} + \eta L_1 \quad (7)$$

where $L_1 = E_{x_i}\left\|G(x_i) - y\right\|_1$ is the content loss that evaluate the 1-norm distance between recovered image $G(x_i)$ and the ground truth y and λ, η are the coefficients to balance different loss terms.

G. Content Loss

The pixel-wise MSE loss is calculated as:

$$l_{MSE}^{SR} = \frac{1}{r^2WH}\sum_{x=1}^{rW}\sum_{y=1}^{rH}\left(I_{x,y}^{HR} - G_{\theta_G}(I^{LR})_{x,y}\right)^2 \quad (8)$$

where $G_{\theta_G}(I^{LR})$ is the reconstructed image and $I_{x,y}^{HR}$ followed by a down sampling operation with down sampling factor r. This is the most widely used optimization target for image SR on which many state-of-the-art approaches rely. As Ledig, instead of relying on pixel-wise losses, they had built the ideas that definition of VGG loss. VGG loss based on the ReLU activation layers of the pre-trained 19-layer VGG network. They defined the VGG loss as the Euclidean distance between the feature representations of a reconstructed image $G_{\theta_G}(I^{LR})$ and the reference I^{HR}:

$$l_{VGG-i,j}^{SR} = \frac{1}{W_{i,j}H_{i,j}}\sum_{x=1}^{W_{i,j}}\sum_{y=1}^{H_{i,j}}(\phi_{i,j}(I^{HR})_{x,y}-\phi_{i,j}(G_{\theta_G}(I^{LR}))_{x,y})^2 \quad (9)$$

IV. Results

For training data, we mainly use the DIV2K dataset. This dataset contains high quality dataset with 2K resolution for image restoration tasks, which are about 800 images with many different details like animals, human, nature views, etc. The whole dataset has variety of scaling factors that we can use 2x, 4x, 3x and 8x with scale down method is mainly bicubic. Besides, there are validation dataset about 200 images for the validation task to predict the process that corresponds to training set.

By applying model of Wang *et al.*, we can set the parameter: C: 4, D: 3, G: 64, G0: 64, T: 10 and the scale factor is 2x. Where the parameters were defined by idealo's team as C is the number of convolutional layers stacked inside a RDB, T is the number of Residual in Residual Dense Blocks, D is the number of Residual Dense Blocks inside each RRDN, G and G0 are the number of the feature maps of each convolutional layers inside RDBs and feature maps for convolutions outside of RDBs and of each RDB output respectively.

Figure 5. Train PSNR_Y.

The training process takes hours for training with learning rate: 5e-4 the decay factor is 0.5 and decay frequency is 30. We trained 80 epochs with step per epochs 500 iterations and batch size is 16. The metric for optimize is PSNR. After 80 epochs we got the validation PSNR is quite acceptable 32 and the train PSNR_Y was 44 at the peak as shown in Fig. 5 and Fig. 6.

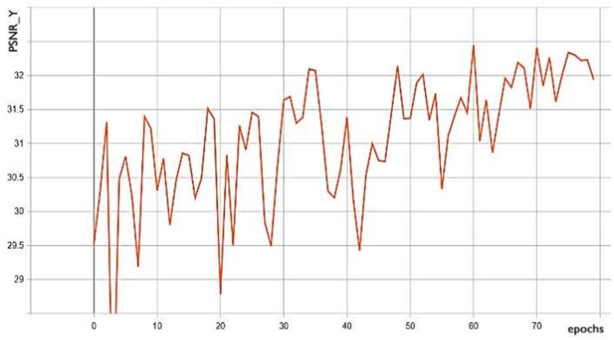

Figure 6. Valid PSNR_Y.

For that PSNR we use tensor-board that is supported by tensor-flow framework and set the smooth weight to 0.82 for better observation. The PSNR_Y increases showed that the generated SR images look closer and closer to the original HR images. We need to maximize the loss function of discriminator. After 40 thousand of steps or iterations, the discriminator started to be unchanged. The model started to converge and no overfit occurred. Our final training PSNR_Y gained 34.22 which is acceptable compared to current state-of-the-art in image super-resolution.

We can see from the Fig. 7 and Fig. 8 about the loss of the generator over the training and validation process. The training process loss is fluctuated, because the generated images fed to the discriminator, the fluctuation means that the discriminator recognized the fake images compared with the ground truth, then generator has to adjust to create better images.

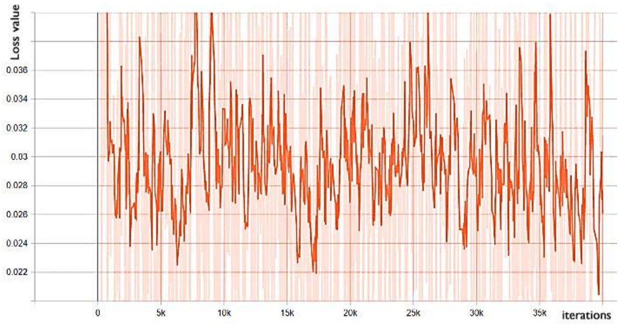

Figure 7. Train generator loss.

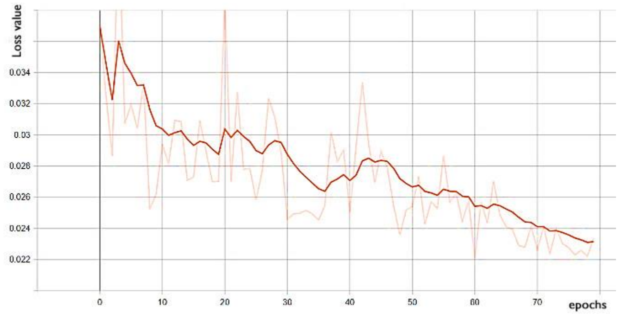

Figure 8. Valid generator loss.

The discriminator losses in both training and validation process in Fig. 9 and Fig. 10 are seen to be unstable. The training process loss trends to decrease to the middle value which is explained that the discriminator is going not to be able to recognize which is fake images from generator and the ground truth images.

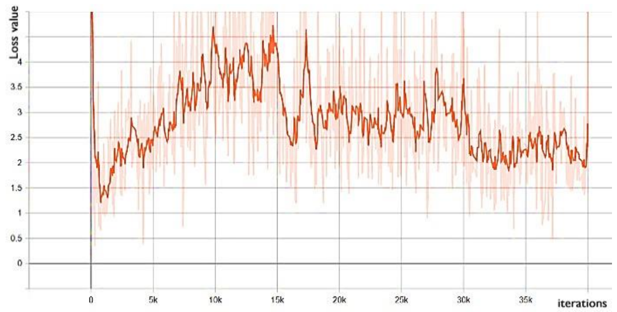

Figure 9. Train discriminator loss.

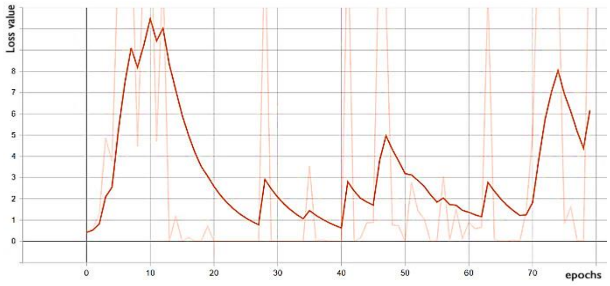

Figure 10. Valid discriminator loss.

We can see that the loss of the validation generator is reduce by epochs and the loss of the train generator is fluctuated by iterations due to the improvement of fake image generation in generator. And the discriminator loss is fluctuated in the between of the graph that is the discriminator is unable to classified whether the real image or the fake image generated by the generator. The train_d_fake_accuracy in Fig. 11 is close to 1 and train_d_fake_loss in Fig. 12 is close to 0 to show exactly the results of the model. With the similar of the accuracy and the loss, we can see that wherever the accuracy increases close to 1, the loss decreases close to 0 and oppositely that wherever the accuracy reduces, the loss also increases significantly.

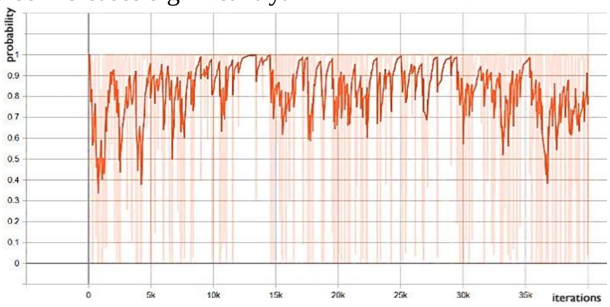

Figure 11. Train d_fake_accuracy.

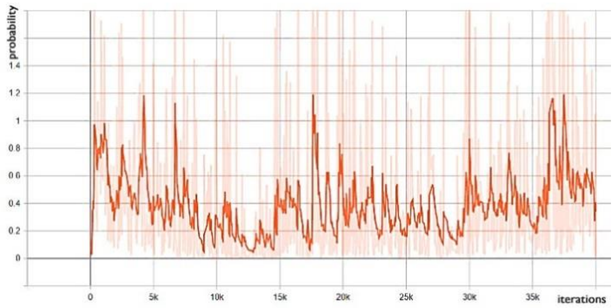

Figure 12. Train d_fake_loss.

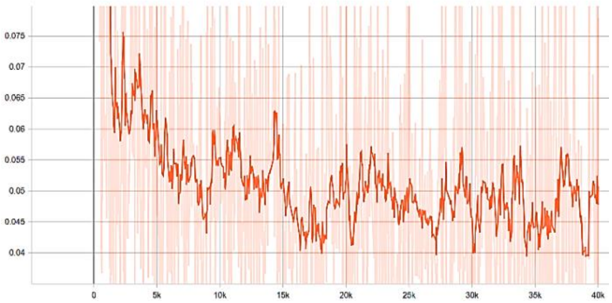

Figure 13. Train generator loss with scale factor x4.

The Fig. 13 showed us the generator loss with scale factor x4. We can see that the loss line is steadily decrease along to the 40k iterations. Thus, the operation of model is quite good as the expectation.

The dataset we trained and predicted is high quality with many subjects and variety of sceneries. With scale factor of 2x we put an image in validation set with size 192x255x3 and got the result of the picture with size 384x510x3 in detail. To make a comparison with the result, we use the bicubic scaling method to up-scale the input again to double of the size. Our prediction is the bicubic method will scale up the images with lossy detail and it is not as clear as the model's results. Then we tried to create the precise of the model with other inputs, which is not in the dataset to prove that whether the model can give the suitable results or not.

TABLE I. THE TRANING RESULTS OF TWO SCALING FACTORS WITH PSNR METRIC VALUES

	Scale	PSNR_x2	PSNR_x4
Bicubic down sampling	Bicubic	29.83	25.75
	SRGAN	29.54	21.62
	ESRGAN	30.13	24.64
Unknown down sampling	Bicubic	24.84	21.83
	SRGAN	24.48	20.19
	ESRGAN	24.62	20.84

Table I compares Average PSNR on validation data set with 100 Low-Resolution images which were down sample by "Bicubic down sampling method" and "Unknown down sampling method". In "Bicubic down sampling" data set, ESRGAN achieve highest average PSNR with x2 up sampling, even higher than the bicubic up sample itself. In contrast, SRGAN have lower in most of the time we test, but it is able to re-create the scaled Super-Resolution image. SRGAN have a low performance in most of the evaluation process, it shows that in order for the neural network to work well, we need to use a suitable loss function. In the x4 section is where the different between 3 methods is shown clearly, in the calculated number as well as in the demo pictures as we can see that, bicubic have higher PSNR and SSIM but the neural network to keep more details than and better visualization.

As we see from Table II, the parameters of two models are different to each other. The SRGAN has fewer parameter than ESRGAN due to the architecture of layers and algorithm. The ESRGAN has RRDB that is the continuous RDB in SRGAN. The scale factor does not affect so much on creating the weights and parameters in these models.

TABLE II. TRANING PARAMETERS OF TWO MODELS WITH TWO SCALING FACTORS

	SRGAN		ESRGAN	
Scale	X2	X4	X2	X4
Iterations	40k	40k	40k	40k
Parameters	1,381k	1,392k	16,644k	16,645k

It can be seen in Fig. 14, which are shown the results that generated by three methods we used, the first one is bicubic scale method, second is SRGAN used RDN and the last is ESRGAN with RRDN. According to the PSNR values of the test results, we can see that with RRDN method the PSNR will higher than the others due to the residual in residual blocks in network. Although our visual to these images does not recognize the difference of these images. However, in pixel level we can know the difference of these methods.

a) PSNR: 29.69

b) PSNR: 30.59

c) PSNR: 31.90

Figure 14. The comparison of results after trained by SRGAN (b), ESRGAN (c) and upscale bicubic method (a) with scaling factor x2.

The result is quite good compared with the bicubic scaling method to reconstruction the image that was scale down from the ground truth, the bicubic scale we can see it make the squirrel a little blurry and we got the detail with the result of the model shown in Fig. 14. Then we test the model's process with random input that is not in the training data or validation data that is shown in Fig. 15. As the result, the model that are developed by Wang *et al.* has proved its efficiency with the results to be very satisfied. The results above are corresponding to the input data which is not in the dataset. Therefore, we just applied the scaling up of model with scale x2 then the PSNR will not be calculated. To test the PSNR, first we have to scale down the input and generate by the model to get the image with the initial size and then find the PSNR metric value to compare with bicubic or other scaling methods.

Figure 15. The input (left) with RRDN scale (above right) and bicubic (bottom right).

As we can see from Table III, the differences in numbers of layers in the model will make the differences in metric values. The ideal model of previous related project had gained the significant results in our test set. With 100 images from our test set the 'idealo' team's model has gained the value of PSNR to approximate 31 and SSIM to 26 in RDN and RRDN respectively. Our model with differences in number of layers, deeper particularly, gained the results with the same value of metric PSNR and SSIM. However, due to the limitation of the training devices, we cannot train the model as long as the 'idealo' team did, but the gained results are approximately the same value as the ideal model. Thus, in our future work, with this expectedly results, we can make the results better by training them long enough to gain better results.

TABLE III. THE COMPARISON OF DIFFERENCES IN NUMBER OF LAYERS BETWEEN IDEALO TEAM'S AND OUR MODEL

	idealo's RDN	Our RDN	idealo's RRDN	Our RRDN	
C	3	4	4	4	4
D	10	3	3	3	5
G	64	64	32	64	64
G0	64	64	32	64	64
T	-	-	10	10	15
x	2	2	4	4	4
PSNR	30.91	29.54	26.03	24.64	23.2
SSIM	0.8016	0.8476	0.673	0.6755	0.6711

Furthermore, we also look to combine these results with image watermarking for practical applications, especially in medical imaging. The first approach is to evaluate and improve the robustness and capacity of the embedded information against the proposed image super-resolution processing. Another way is to obtain the good high-resolution images with high embedded information capacity.

V. Conclusion

As considering problem mentioned, the BN removal keeps the stability and consistence performance without artifacts. It does not decrease the performance but save the computational resources and memory usage. In this paper of research, we used the vast.ai server with 1080Ti with RAM is 16GB that a little bit limited about the batch size and step to train carefully. Our results have reached the same quality of the Wang's with PSNR reaching to 30 for validation test and gain good result for images that outside of the dataset and validation set. By applying RRDN model then we consider that the deeper the model is the more precise results we get. However, because the model is in the state-of-the-art of performance, then to get better results, our suggestion is about the data. Before getting into training process, dataset should be passed denoise filters in order not to generate and upscale the noisy much uncomfortable. And with the project of Jahidul *et al.* [16] about the underwater dataset, we suggest that the whole dataset should be classified by simple classes such as: underwater, land, animals, etc. There are many classification methods we can use like YOLO with image segmentation classification. With classes, we trained other files of weights, and before we use RRDN to test, classified the input to use the corresponding file of weight suitably. We figure out this method will give the generated images with better detail due to the similar detail they are in the class trained input. However, that will reduce the number of dataset and we need as much data as possible for each class.

Another idea is based on the trained network from smaller resolution for example: 128 to 248 to 456, etc. By increasing the resolution gradually, the network is continuously asked to learn a much simpler piece of the overall problem. The incremental learning process greatly stabilizes training. The low-to-high resolution trend also forces the progressively grown network to focus on high level structure first and fill in the details later. This improves the quality of the final image by reducing the likelihood that the network will get some high-level structure drastically wrong. The future work of this idea is started with downscale images with multiple factor. Then we will train 16x with 8x, save and load weight to train next 8x with 4x, continuously like that until we do the train with 2x factor. The results can obtain better quality even with upscale factor 8x or more. The expansion of these results for image watermarking in medical imaging is also considered in further research.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

All authors had contributed equally and approved the final version.

Acknowledgment

This study was partially supported by Office for International Study Programs – OISP, Department of Electrical Electronics Engineering, University of Technology, HCMUT, National University - Ho Chi Minh City (VNU).

References

[1] A. Lucas, M. Iliadis, R. Molina, and A. K. Katsaggelos, "Using deep neural networks for inverse problems in imaging," *IEEE Signal Processing Magazine*, January 2018.

[2] Z. Wang, J. Chen, S. C. H. Hoi, "Deep learning for image super-resolution: A survey," arXiv:1902.06068v2, Feb. 2020.

[3] Ha Noi 9 gigapixel. [Online]. Available: https://www.360view.vn/gallery/data/projects/vietnam/2015/hn9g/?fbclid=IwAR1eZ8F2COBm2KlgZC4OkcffmKOvWACUYSEiT4KaXizIDXU3gMtge5LoJ58

[4] X. Zhu, X. Wang, J. Wang, P. Jin, L. Liu, and D. Mei, "Image super-resolution based on sparse representation via direction and edge dictionaries," *Mathematical Problems in Engineering*, 2017.

[5] C. Jiang, Q. Zhang, R. Fan, and Z. Hu, "Super-resolution CT image reconstruction based on dictionary learning and sparse representation," Scientific Reports, 11 June 2018.

[6] T. L. Tien and N. L. H. Sinh, "Performing image resolution super by the method of irregular data interpolation and filter bank," The Minor Project on Superresolutions and Filter Banks (in Vietnamese), Ho Chi Minh University of Technology, Ho Chi Minh City, Vietnam, 2007.

[7] T. Le Tien and D. V. Nguyen, "Super-resolution image processing combining transform between the frequency domain and wavelets domains (In Vietnamese)," *Journal Research, Development and Application on Information & Communication Technology*, pp. 39-47, April 2009.

[8] W. Yang, X. Zhang, Y. Tian, W. Wang, J. Xue, and Q. Liao, "Deep learning for single image super-resolution: A brief review," *IEEE Signal Processing Magazine*, 2019.

[9] M. Kawulok, P. Benecki, S. Piechaczek, K. Hrynczenko, D. Kostrzewa, and J. Nalepa, "Deep learning for multiple-image super-resolution," *IEEE Signal Processing Magazine*, 2019.

[10] I. J. Goodfellow, *et al.*, "Generative adversarial nets," in *Proc. Neural Information Processing Systems Conference*, 2014.

[11] C. Ledig, *et al.*, "Photo-Realistic single image super-resolution using a generative adversarial network," arXiv:1609.04802, 2017.

[12] B. Wu, H. Duan, Z. Liu, and G. Sun, "SRPGAN: Perceptual generative adversarial network for single image," arXiv:1712.05927, 2017.

[13] Y. Zhang, Y. Tian, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," arXiv:1802.08797, 2018.

[14] X. Wang, *et al.*, "ESRGAN: Enhanced super-resolution generative adversarial networks," arXiv:1809.00219, 2018.

[15] A. Jolicoeur-Martineau, and L. D. Institute, "The relativistic discriminator: A key element missing from standard GAN," arXiv:1807.00734, 2018.

[16] M. J. Islam, S. S. Enan, P. Luo, and J. Sattar, "Underwater image super-resolution using deep residual multipliers," arXiv:1909.09437, 2019.

Thuong Le-Tien (MIEEE-96) was born in Saigon, Ho Chi Minh City, Vietnam. He received the Bachelor and Master Degrees in Electronics-Engineering from Ho Chi Minh City Uni. of Technology (HCMUT), Vietnam, then the Ph.D. in Telecommunications from the Uni. of Tasmania, Australia. Since May 1981 he has been with the EEE Department at the HCMUT. He spent 3 years in the Federal Republic of Germany as a visiting scholar at the Ruhr Uni. from 1989-1992. He served as Deputy Department Head for many years and had been the Telecommunications Department Head from 1998 until 2002. He had also appointed for the second position as the Director of Center for Overseas Studies since 1998 up to May 2010. His areas of specialization include: Communication Systems, Signal Processing and Electronic Circuits. He has published more than 180 scientific articles and the teaching materials for university students related to Electronic Circuits 1 and 2, Digital Signal Processing and Wavelets, Antenna and Wave Propagation, Communication Systems. Currently he is a full professor at the HCMUT.

Tuan Nguyen-Thanh was born in Ho Chi Minh City, Vietnam. He received B.Eng. and M.Eng. degrees from Ho Chi Minh City University of Technology (HCMUT), Vietnam, in 2002 and 2004, respectively, both in electrical engineering and telecommunications. He has been at the HCMUT since 2002. Currently, he is pursuing the Ph.D. degree at the HCMUT. His main research interests include watermarking, digital signal processing and communication systems.

Hanh Phan-Xuan got the B.Eng and M.Eng. Degrees from the Ho Chi Minh City University of Technology (HCMUT), Vietnam. His research relates to Image Signal Processing, Neural Networks and Deep Learning Techniques to solve problems in Computer Vision, Image Forgery Detection, Biometrics Signal Processing, and Autonomous Robotics. Currently, he is a Ph.D. Student at EEE Department of the HCMUT.

Giang Nguyen-Truong was born in Tay Ninh, Vietnam. He got the Bachelor degree in 1999 with the Advanced Program of the Office for International Student Program (OISP) in Telecommunication major, Electrical and Electronics Engineering Department, Ho Chi Minh City University of Technology, HCMUT, Vietnam. His research related to apply machine learning to help solving inverse problem in image processing, particularly in Image Super Resolution.

Vinh Ta-Quoc was born in Binh Phuoc Province, Vietnam. He got the Bachelor degree in 1999 with the Advanced Program of the Office for International Student Program (OISP) in Telecommunication major, Electrical and Electronics Engineering Department, Ho Chi Minh City University of Technology (HCMUT), Viet Nam. His research related to apply machine learning to help solving inverse problem in image processing, particularly in Image Super – Resolution.