Computer Engineering and Applications Vol. 10, No. 1, February 2021

Evaluation of Deep Convolutional Neural Network with Residual Learning for Remote Sensing Image Super Resolution

Rika Sustika[1,2,*], Andriyan B. Suksmono[1]

[1]*School of Electrical and Informatics Engineering (STEI), Institut Teknologi Bandung (ITB)*
[2]*Research Center for Informatics, Indonesian Institute of Sciences (LIPI)*
**rika002@lipi.go.id*

ABSTRACT

Remote sensing images generally have low spatial resolution because of the limitations of sensing devices, bandwidth transmission, or storage capacity. An effective way to improve spatial resolution with low cost is by using algorithm based approach, known as super resolution (SR). In recent years, deep learning is super resolution technique that received special attention because it gave better performance than traditional method. In this research, we evaluated two simple deep learning architectures and explored parameters setting of deep convolutional neural network with residual learning, to achieve the trade-off between performance and speed or computational complexity, for implementation on remote sensing image super resolution. Results from the experiment show that deeper network with smaller number of filter gives faster model than shallow network with bigger number of filter, without sacrificing the performance.

Keywords: Convolutional Neural Network, Deep Learning, Image, Remote Sensing, Super Resolution.

1. INTRODUCTION

Remote sensing is the process of obtaining information about targeted objects or areas by measuring its reflected and emitted radiation from a distance. Remote sensing imaging is used in many applications such as environmental monitoring, agriculture, land use mapping, meteorology, surveillance, target detection, etc. For solving such problems with high performance, it important to have high resolution remote sensing image datasets. Unfortunately, sensor resolution to capture remote sensing images is limited by the cost of hardware and acquisition technology. The higher resolution requires higher cost. A solution for increasing image resolution at lower cost is by using algorithm based approach, known as super-resolution (SR).

For many years ago, a number of traditional methods were used for super-resolution, such as bilinear and bicubic interpolation, sparse coding, iterative back projection, neighborhood embedding, and many other methods [1]. The simplest and fastest way is interpolation method that projects the initial low resolution (LR) image onto a high resolution (HR) grid and the missing pixel values are estimated using an interpolation function. Learning based super-resolution, especially deep learning with convolutional neural network (CNN) architecture gets considerable attention nowadays. The pioneering CNN model for SR is proposed by Dong et.al, known as super-resolution convolutional neural network (SRCNN) [2][3]. It consists only 3 convolutional layers but provides better performance than traditional methods. On the next research, Dong et al. proposed fast SRCNN (FSRCNN) to accelerate SRCNN, which could process images in real time [4]. Improvement from SRCNN has been actively explored by many researchers such as make deeper or wider networks to obtain better performance. VDSR (very deep super resolution) is one of the deep learning methods that has deeper and wider network than SRCNN [5]. VDSR used residual learning for training the super resolution model, and the result of this approach outperform SRCNN by a large margin [5].

In this work, we evaluated these two deep learning architectures for remote sensing super resolution purposes. We compared the performance of these deep learning methods with bicubic interpolation method. We also examined parameter settings of this deep convolutional neural network to get deep learning super resolution model with better performance but faster testing stage.

This paper is arranged as follows. In Section II, we present principle about deep learning for super-resolution. Section III explains about deep learning architectures that evaluated on this paper. In Section IV, we present about experiment detail. Results and discussion are given on Section V and Section VI is conclusion.

2. DEEP LEARNING SUPER RESOLUTION PRINCIPLE

Figure 1 is the diagram of deep learning approach for super-resolution on this research. There are two main phases on this method, i.e. training and prediction. On the training phase, deep learning model is trained to analyze statistical relationship between the LR (low resolution) and its corresponding HR (high resolution) image from a dataset training. On prediction stage, the trained SR model is then used to predict HR images from a set of LR images.

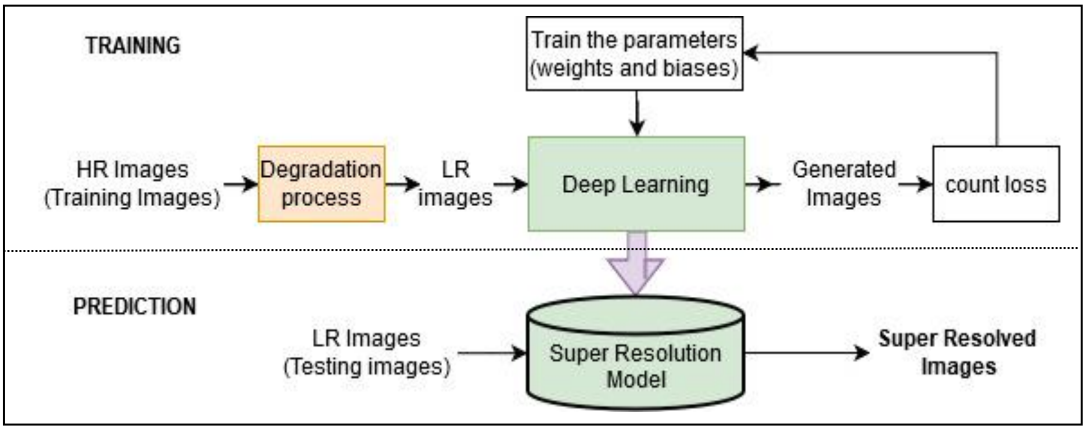

FIGURE 1. Deep Learning Super-resolution Diagram

The LR image in the experiment is modeled as output of the degradation process from HR image as can be stated by the following equation:

$$X = \mathfrak{D}(Y, \delta) \tag{1}$$

where Y is the HR image, while X is the LR image. $\mathfrak{D}$ represents the degradation mapping function from HR to the LR image and δ is parameter of the degradations process. Degradation process can be affected by various factor such as blurring,

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

noise, artefact, defocusing, etc. Super-resolution technique try to recover HR image $\tilde{Y}$ from LR image X. $\tilde{Y}$ is expected to be as similar as possible with Y.

3. EVALUATED ARCHITECTURE

At present, among various deep learning techniques, convolutional neural network (CNN) is the most widely used for super-resolution. There were many architectures proposed for image super-resolution based on convolutional neural network. The simplest architecture is super-resolution convolutional neural network (SRCNN) that proposed by Dong et al. [3]. Other simple method is very deep super resolution (VDSR) [5] that used residual learning for training the super resolution model. On this research, we evaluated the performance of these methods for super resolution of remote sensing image. We also evaluated different layers and filters on VDSR structure to see the tradeoff between performance and computational complexity.

3.1 SUPER RESOLUTION CONVOLUTIONAL NEURAL NETWORK (SRCNN)

The implementation of convolutional neural network (CNN) for super-resolution was first developed by Dong et al. with the method known as super resolution convolutional neural network (SRCNN) [2]. The SRCNN block diagram for remote sensing single image super-resolution can be seen in Figure 2.

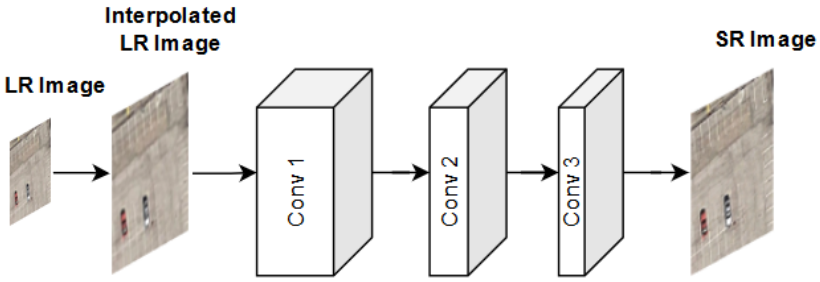

FIGURE 2. Super Resolution Convolutional Neural Network (SRCNN) Structure

SRCNN consists of three convolutional layers. The first layer is feature extractor layer. It has 64 filters with 9x9 kernel size. The second layer is non-linear mapping, consist of 32 filters with 1x1 kernel size, and the last layer is reconstruction layer that has 3 filters with 3x3 kernel of size [2].

3.2 VERY DEEP SUPER RESOLUTION (VDSR)

VDSR (very deep super resolution) is a deep learning approach for super-resolution with deeper and wider network than SRCNN. The diagram of VDSR network can be seen on Figure 3 [5].

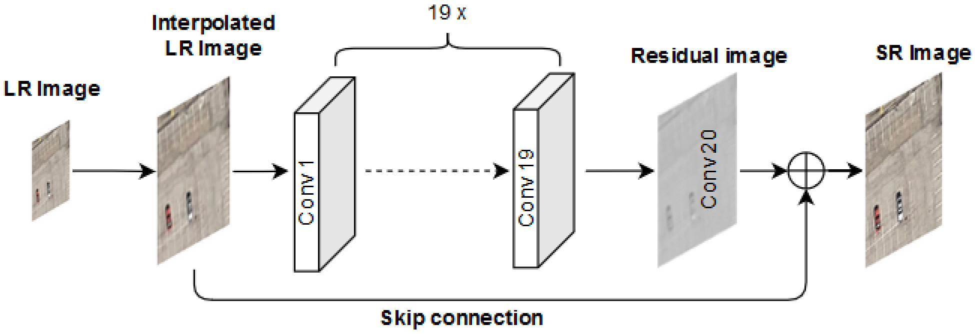

FIGURE 3. Very Deep Super Resolution (VDSR) Structure

Similar with input for SRCNN, input for the VDSR network is also interpolated LR image ($\tilde{X}$). VDSR consists of 20 convolutional and ReLU layers, followed by residual layer. Each convolutional layer consists of 64 filters with the size of 3x3. The VDSR network is learning the residual error (r) between the output (Y) and network input ($\tilde{X}$), instead of learning the HR output directly. On the residual layer, LR image is added to the output of the network to get back the HR image, and the loss function become $L = \frac{1}{2}\left\|r - \tilde{Y}\right\|^2$, where $r = Y - \tilde{X}$, and $\tilde{Y}$ is an estimate of the target HR image. In the residual layer, the output from the network is added with the LR image to get back the HR image.

4. EXPERIMENT

4.1 DATASET

We used BDG dataset as training data. BDG dataset is own created dataset for this super-resolution experiment purposes. BDG dataset consists of 6000 images, with dimension of each image patch is 256x256 pixel. This dataset was made from 3 large remote sensing images for Bandung area with each picture size is 10,000x10,000 pixel. We used 80% of data for training and 20% of them used for validation. Samples of training data can be seen on Figure 4.

FIGURE 4. Example of Dataset Training

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

This training dataset is consider as high resolution image. Degradation process to get LR image is done by using downscale process using bicubic interpolation with 2 and 4 scale factors.

Testing data is used to evaluate the performance of deep learning models. For testing, we use Indo dataset that collected from Google earth imagery for Indonesian area with size 1,500x750 pixel. This dataset is divided into smaller patches with size 256x256 pixel.

4.2 EXPERIMENT DETAILS

On this experiment, we first investigate the comparison of SRCNN and VDSR performances for implementation on remote sensing single image super resolution. We then examine the impact of number of layer and filter number towards the super resolution performance. All the experiment use batches of size 32. Mean square error (MSE) is used as loss function and Adam algorithm is used as optimization solver. Learning rate is fixed and we set it to 0,00005. All experiments was trained over 30 epochs.

On testing stage, LR images are super-resolved using trained SR model, for 2 and 4 upscaling factor. Performance of SR model are evaluated by comparing ground truth images (HR images) with super-resolved (SR) images using quality metric specified for measuring image quality. Four quality metrics were used: PSNR (peak signal to noise ratio), SSIM (structural similarity), ERGAS (Erreur Relative Globale Adimensionnelle de Synthese), and SAM (spectral angle mapper). The higher value of the PSNR and SSIM, and lower value of ERGAS and SAM indicate better performance of super-resolution.

5. RESULTS AND DISCUSSION

On the first experiment, we evaluated two deep learning architectures for super resolution: SRCNN and VDSR. On training phase, comparison between SRCNN and VDSR using PSNR metric on every epoch is shown on Figure 5.

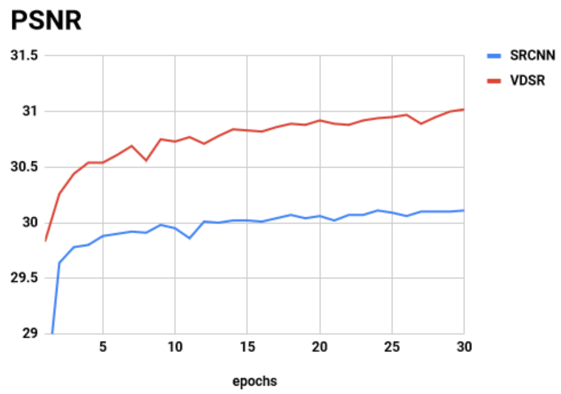

FIGURE 5. PSNR of Training Stage on Every Epoch

We notice that VDSR outperforms SRCNN approach by a large margin (0.91 dB on 30th epoch). We also compared the two methods with bilinear and bicubic interpolations, by testing these methods using Indo dataset as testing data, and the results can be seen on Table I, meanwhile the visualization result can be seen on Figure 6.

TABLE 1
The Average PSNR of SR using Deep Learning and Interpolation Methods

Upscaling factor	Methods	PSNR (dB)	ERGAS	SAM	SSIM
2x	Bilinear	32.90	4.15	0.21	0.87
	Bicubic	35.68	3.07	0.15	0.92
	SRCNN	35.61	3.00	0.23	0.93
	VDSR	36.35	2.76	0.15	0.93
4x	Bilinear	29.14	6.43	0.46	0.74
	Bicubic	30.48	5.51	0.28	0.79
	SRCNN	30.63	5.37	0.33	0.79
	VDSR	31.17	5.00	0.27	0.81

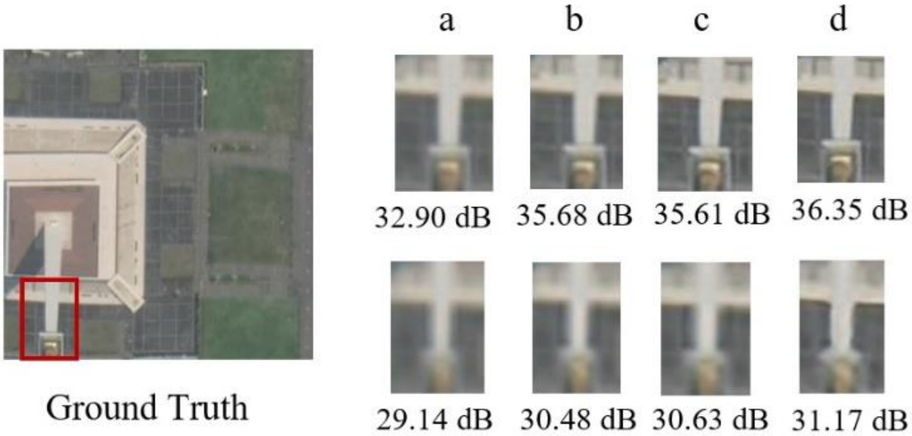

FIGURE 6. Super-resolution Result: (Top) 2x upscaling factor and (bottom) 4x upscaling factor. a. bilinear interpolation, b bicubic interpolation, c SRCNN, and d. VDSR

From Table 1 and Figure 6 we can see that on the two up scaling factors, the lowest SAM and ERGAS and the highest value of PSNR and SSIM were achieved by VDSR. Its mean that VDSR has superior performance compare with SRCNN and interpolation methods.

Unfortunately, there is always a tradeoff between performance and the computation complexity. The deeper and wider network means the larger number of parameters that have to be trained and make computation complexity increased. Table II show the evaluation of training time and testing time of VDSR based method.

Table 2 shows that training time of VDSR is more than 5x of SRCNN, and SRCNN also run faster than VDSR on testing stage with around 3x speed. Ideally, we want to have deep learning model with good performance with high speed or low

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

computation complexity. For SRCNN method, Dong et al. has studied the impact of number of layer, filter size, and number of filter toward the SR performance.

TABLE 2
The Training and Testing Time of SR using Deep Learning Methods

Methods	Number of Parameters	Training Time (second)	Testing time (second)
SRCNN	20,099	56.52	1.19
VDSR	667,073	294.65	3.14

The results show that deeper networks do not always result the better performance. Larger filter size could significantly improve the performance, and superior performance could be achieved by increasing the width [6]. We then evaluate the impact of number of layers (L) and number of filters (N) on VDSR architecture and the results is shows on Table 3 and Table 4.

TABLE 3
Number of Parameter on Different Layers

Number of layers (L)	Number of parameters	
	N=64	*N=32*
10	297,793	75,169
15	482,433	121,409
20	667,073	167,649
25	851,713	213,889
30	1,036,353	260,129

According to Table 3, reducing filter number from 64 to 32 can decrease number of parameter more significant than decreasing number of layer from 20 to 10. It also give impact to the training and testing time of the deep learning super resolution model, such as shown on Table 4.

TABLE 4
The Average Result of SR using Deep Learning and Interpolation Methods

Number of layers	PSNR (dB)		Training time (S)		Testing time (S)	
	N=64	*N=32*	*N=64*	*N=32*	*N=64*	*N=32*
10	30.63	30.47	162.26	76.53	2.99	2.89
15	30.71	30,63	215.59	99.83	3.05	2.92
20	30.75	30.61	294.65	128.25	3.14	2.99
25	30.77	30.77	358.24	150,07	3.73	3.21
30	30.91	30.95	404.97	168.17	3.86	3.44

Table 4 shows that decreasing the width of the network can reduce computation complexity which is seen on decreasing the training and testing time. For example, on 30 convolutional layers with 32 filters, training time is faster than 64 filters with significant improvement. Training time decrease more than 50%. On this experiment, reducing the width doesn't give significant impact to the performance, moreover, in 30 convolutional layers, performance of 32 filters is better than 64 filters. Therefore, increase the number of layer but decrease the filter number maybe become good choice for remote sensing single image super resolution. Optimum parameter setting for this remote sensing super resolution is by using 30 convolutional layers with 32 filter number. On this setting, performance of the SR, denoted by PSNR, is better than VDSR performance (0.2 dB increased), meanwhile testing time increase 9.5% for each image.

6. CONCLUSION

In this paper, we evaluated the effect of deeper and wider deep learning network toward the performance of remote sensing super resolution on very deep super resolution architecture. The results show that deeper architecture give better performance but significantly increase the training and testing time. There is a trade-off between performance and computational complexity. Increasing the number of layer with decreasing the filter number become good choice for remote sensing single image super resolution on our experiment.

In the future work, we could evaluate the robustness of the model against blurring effect, noisy data, and also other scaling factors.

ACKNOWLEDGEMENTS

The experiment on this research are conducted on High Performance Computing (HPC) facility in Research Center for Informatics, Indonesian Institute of Sciences (LIPI).

REFERENCES

[1] R. Fernandez-beltran, P. Latorre-carmona, and F. Pla, “Single-Frame Super-Resolution in Remote Sensing : a Practical Overview,” *International Journal of Remote Sensing*, vol. 38, no. 1, pp. 314–354, 2017.

[2] C. Dong, C. C. Loy, K. He, and X. Tang, “Learning a Deep Convolutional Network for Image Super-Resolution,” in *Computer Vision – European Conference on Computer Vision 2014*, 2014, pp. 184–199.

[3] C. Dong, C. C. Loy, and K. He, “Image Super-Resolution Using Deep Convolutional Networks,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 2, pp. 295–307, 2016.

[4] C. Dong, C. C. Loy, and X. Tang, “Accelerating the Super-Resolution Convolutional Neural Network,” in *European Conference on Computer Vision (ECCV)*, 2016, pp. 391–407.

[5] J. Kim, J. K. Lee, and K. M. Lee, “Accurate Image Super-Resolution Using Very Deep Convolutional Networks,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 1646–1654.

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)