March - April 2020
ISSN: 0193-4120 Page No. 4409 - 4415

Super Resolution Method Using Channel Attention for High Frequency Feature Emphasis

Dong Woo Lee[1], Sang Hun Lee[*2], Jin Soo kim[3]
[1]Master Student, Dept. of Plasma Bio Display, KwangWoon University, Korea
[*2]Associate Professor, Ingenium College of Liberal Arts, KwangWoon University, Korea
[3]Professor, IDP System., Co, Ltd, Korea
led0121@kw.ac.kr[1], leesh58@kw.ac.kr[*2], cwicc@hanmail.net[3]

Article Info
Volume 83
Page Number: 4409 - 4415
Publication Issue:
March - April 2020

Article History
Article Received*: 24 July 2019*
Revised*: 12 September 2019*
Accepted*: 15 February 2020*
Publication*: 26 March 2020*

Abstract
In this paper, we proposed a method that applied Channel Attention to emphasize features in the process of CNN-based single image super resolution. The existing Pre-upsampling method using deep learning uses an image extended to Bicubic Interpolation, so that the High amount of calculation, and distortion and meaningless values are added during the extension process. Further, in the existing super-resolution method using CNN, high-frequency components such as contours and textures are not emphasized in the process of feature extraction, and there is a problem that the contours are blurred or distorted. To solve these problems, a low-resolution image was used as input, and a Channel Attention structure with Residual Block was used. These Channel Attentions effectively extract features through emphasizing high frequency components and limiting low frequency components. This emphasized feature map was extended with sub-pixel convolution instead of deconvolution for super resolution. As a result, unnecessary duplication operations were reduced, and various features were extracted through many convolutions. Through the experiment, contour and texture expression were improved compared to Bicubic Interpolation and VDSR.

Keywords: *Super Resolution, Residual Block, Concatenation Layer, Channel Attention, CNN*

1. Introduction

With the development of image processing technology, the field of object detection and recognition[1-5] has been used, and high-resolution images have been required for smooth image processing. Accordingly, research on a super-resolution method for reconstructing a low-resolution image into a high-resolution image has been advanced. However, in general, a low-resolution image is reconstructed into a high-resolution image, and complete restoration is difficult. Recently, a super-resolution method using CNN(Convolutional Neural Networks) has been applied.

Super-resolution methods include Interpolation and learning-based methods. Interpolation is a method of restoring surrounding pixel values using a function similar to a discrete data distribution, based on an empty pixel generated in the process of enlarging a low-resolution image into a high-resolution image. Learning-based super resolution methods include dictionary-based methods and deep learning methods using CNN. A method based on pre-learning is to set a low-resolution image patch and a high-resolution image patch to a data set, and then expand to a high-resolution image patch if the input low-resolution patch and the low-resolution patch of the data set are similar Is the way. The deep learning-based method is a method of analyzing the relationship between a low-resolution patch and a high-resolution patch using a convolution layer. Methods using CNN include SRCNN(Super

March - April 2020
ISSN: 0193-4120 Page No. 4409 - 4415

Resolution Convolutional Neural Network)[6] and VDSR(Very Deep Super Resolution)[7].

2. Related Work

2.1 Channel Attention

Attention plays an important role in recognizing an object. Generally, an object to be recognized has a certain range rather than being distributed over the entire image. Taking advantage of these points, Channel Attention is applied to improve the performance at the recent recognition stage. Channel Attention is a method that emphasizes the channel of one feature map in a typical Attention Mechanism[8]. The purpose of GAP (Global Average Pooling)[9] is to find the average of each feature map channel and reduce it to three-dimensional to one-dimensional features. A one-dimensional feature map was constructed using the GAP of equation (1). Thereafter, the GAP value is readjusted using the Fully Connected Layer of the equation (2), and the final Attention value is generated using the Sigmoid function. The Attention value generated by equation (2) is multiplied by the input feature map as shown in equation (3), and the Attention value is advanced.

$$H_{GP}(x) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} x(i,j) \qquad (1)$$

Where, H and W mean the height and width of each feature map.

$$s = \delta\left(W_d \times f\left(W_d \times H_{GP}(x)\right)\right) \qquad (2)$$

Where, δ and $f(\cdot)$ mean Sigmoid and ReLU (Rectified Linear Unit) activation functions, respectively, and W_U is the weight of Fully Connected Layer.

$$CA(x) = x \times s \qquad (3)$$

Where, x is an input, and s is the weight of Channel Attention. Channel Attention proceeds through equation (3).

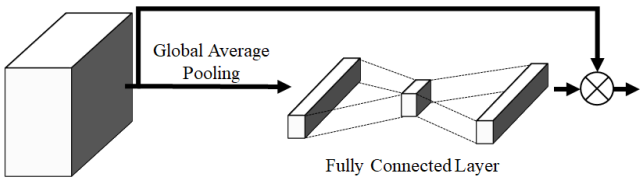
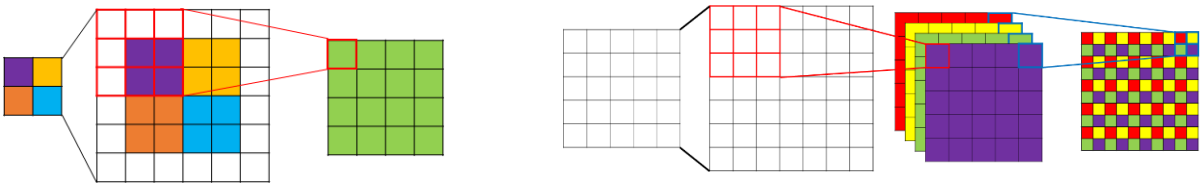


Figure 1. Channel Attention.

2.2 . Sub-pixel Convolution

Sub-pixel Convolution[10] extends the size of the feature map by using Pixel-shuffle in the Convolution layer having the same function as the Deconvolution layer. The Deconvolution process performs a Convolution operation after expanding the feature map to a target size. These processes are inefficient because there are many pixels that are overlap-calculated through the Convolution process after expanding the feature map using nearest neighbor interpolation. In contrast, since expansion is performed through rearrangement after the Convolution process, there is no pixel that is redundantly calculated, efficient expansion is possible, and various feature extractions are possible with many Convolution operations.



(a)Deconvolution (b) Sub-pixel Convolution

Figure 2. Compared Deconvolution with Sub-pixel Convolution

$$f^{sp}(x) = PS(W * x + b) \qquad (4)$$

Equation (4) is a Sub-pixel Convolution equation. Here, PS is a rearrangement formula, W is

March - April 2020
ISSN: 0193-4120 Page No. 4409 - 4415

Convolution, and b is bias. In the Convolution process, the number of channels is a multiple of the square of the magnification factor.

$$PS(T)_{x,y,z} = T_{\lfloor x/r \rfloor \lfloor y/r \rfloor, C \cdot r \cdot mod(y,r) + C \cdot mod(x,r) + c} \quad (5)$$

Where, x and y mean height and width, and z means channel. The feature maps of a plurality of channels are pixel-shuffled through equation (5).

3. Proposed Method

In this paper, we proposed a super-resolution method that combined Channel Attention structure and Residual Block [11] structure combination to emphasized high frequency features and Sub-Pixel Convolution to increase feature map size. The proposed method consisted of three major steps. A shallow feature extraction step using two Residual Blocks composed of 5×5 size Convolution and 3×3 size Convolution, a step of extracting deep features by combining Residual Block and Channel Attention, and Performed at the stage of expanding the feature map via Pixel Convolution.

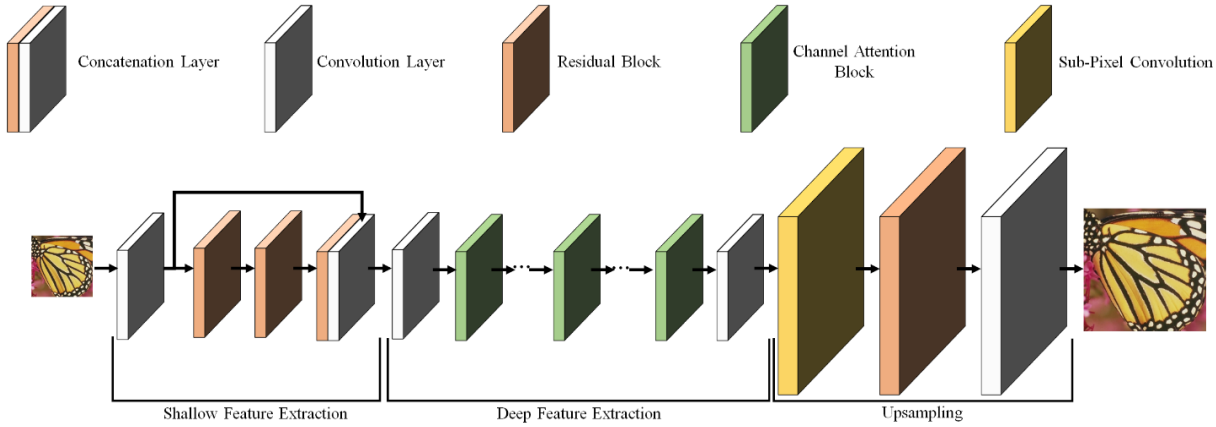


Figure 3. Proposed Method Network

3.1 Shallow Feature Extraction

This is a stage consisted for the purpose of extracting shallow features from an input image. It consists of Convolution Layer and Residual Block, and the feature map of Convolution and Residual Block is combined with Concatenation Layer[12]. All Convolution sizes in Shallow Feature Extraction were 5×5. A feature map was generated from the input low-resolution image via the Convolution Layer. In addition, a low-dimensional feature map was extracted via the Residual Block. Through the combination of the feature map of the Convolution Layer and the feature map of the Residual Block through the Concatenation Layer, it played a role in effectively utilizing various features. Figure 4 shows Shallow Feature Extraction structure.

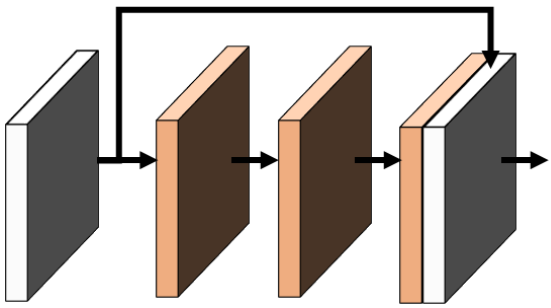


Figure 4. Shallow Feature Extraction Structure

Figure 5 shows the Residual Block. It was a structure connected to Skip Connection that adds

March - April 2020
ISSN: 0193-4120 Page No. 4409 - 4415

an input feature map and two feature maps of Convolution results. Such a structure was a Gradient Vanishing/Exploding problem in which the difference between the input and output feature was learned during the learning process, and a deep network was formed. Can be effectively solved, and high-speed stable learning can be performed.

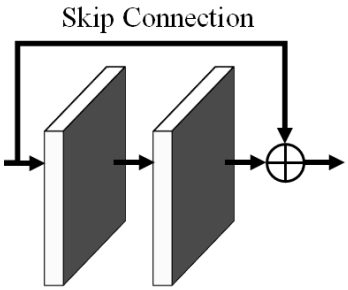


Figure 5. Residual Block

$$F_{Re}^{i}(x) = H_2(H_1(x)) + x \qquad (6)$$

Equation (6) is a Residual Block equation. Here, x is the input feature map of the Residual Block, and H_1 and H_2 are the first and second Convolutions, respectively.

3.2 Deep Feature Extraction

This is a stage for high-level feature extraction and enhancement from features extracted from Shallow Feature Extraction. This structure combines residual block, concatenation layer for feature transfer, and channel attention for feature emphasis. All Convolutions of Deep Feature Extraction consisted of 3×3, and the size of Fully Connected Layer consisted of 16 and 64. The feature map was extracted to the Residual Block, and the high-dimensional features for Attention were utilized through the Concatenation Layer. Attention weight was generated by GAP and Fully Connected Layer using Convolution to match the size of the extracted feature map with the input feature map. Channel Attention was advanced using the subsequent input feature map and multiplication operation, and Attention was performed using the last Convolution.

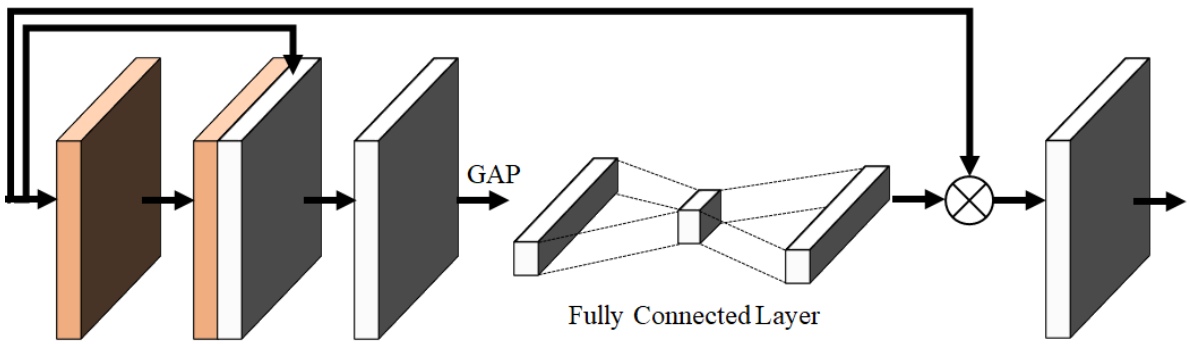


Figure 6. Channel Attention Block Structure

3.3 Upsampling

This is the stage for expanding the size of high-dimensional emphasized features with Deep Feature Extraction. It consisted of Sub-Pixel Convolution, Residual Block, and Convolution. All Convolutions in the Upsampling stage consisted of 3×3. When expanding using Deconvolution, the feature map of Deep Feature Extraction was expanded to Sub-Pixel Convolution for preventing Checkerboard Artifact, which causes plaid noise due to pixels that are duplicated operation. For effective expansion, we used sub-pixel convolution with different structure by scale factors as shown in Figure 7. Then, in order to generate a final image, the

March - April 2020
ISSN: 0193-4120 Page No. 4409 - 4415

expanded feature map was reconstructed into a Residual Block and a Convolution.

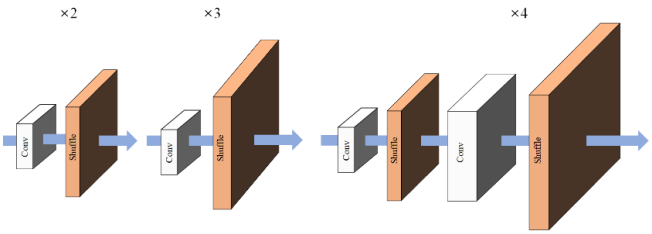


Figure 7. Sub-pixel Convolution by Magnification

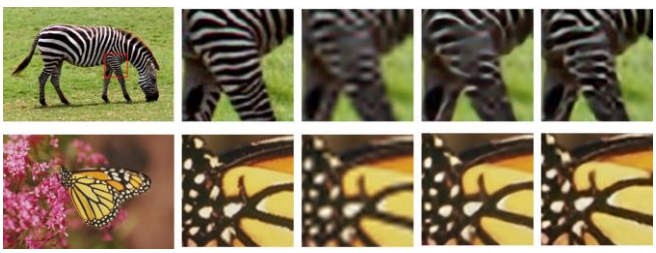
3.4 Loss Function

The network is learning to minimize the result of the loss function. In this paper, we used *L1-normalization loss*. In equation (7), x represents the original image, $x`$ represents the image reduced to Bicubic, and $f(\cdot)$ represents the network of the proposed method.

$$L_1(x, f(x`)) = \|x - f(x`)\| \qquad (7)$$

4. Experimental Results

In this paper, we used 800 DIV2K datasets composed of high-resolution images for learning the network. Patches were divided into 48×48 sizes randomly cut from the training image. ADAM Optimizer[13] was used, and the learning rate was set to 10^{-4}, $\beta_1 = 0.9$, $\beta_2 = 0.999$. Experiments were performed on Set5, Set14, B100, and Urban100 data sets. The experimental results were compared with Bicubic Interpolation and VDSR.



(a) (b) (c) (d)

Figure 8. Super Resolution Result (×4) of Set14 'zebra', 'monarch'

(a) Original Image (b) Bicubic Interpolation Image (c) VDSR (d) Proposed Method

Figure 8 is an image of the super-resolution result expand four times with the "zebra", "monarch" image of Set14. In Bicubic Interpolation, the outline of zebra stripes cannot be sharpened from the image, and the texture was blurred and distorted. In the case of VDSR, the results were improved compared to the Bicubic Interpolation image. However, since an image extended to the Bicubic Interpolation was used as input, distortion occurred in the stripes generated by the Bicubic Interpolation. The proposed method had a low distortion, which is different from Bicubic Interpolation or VDSR, because it takes low-resolution images as input.

(a) (b) (c) (d)

Figure 9. Super Resolution Result (×3) of B100 '8023', '102061'

(a) Original Image (b) Bicubic Interpolation Image (c) VDSR (d) Proposed Method

Figure 9 is an image of the super-resolution result three times expand from the "8023", "102061" image of B100. In Bicubic Interpolation, the stair phenomenon occurs in a straight line from the feathers of the image, and the texture was blurred. VDSR showed better results than Bicubic Interpolation image, but the staircase phenomenon still occurred. The proposed method did not cause the staircase phenomenon of feathers and showed clear results.

March - April 2020
ISSN: 0193-4120 Page No. 4409 - 4415

(a) (b) (c) (d)

Figure 10. Super Resolution Result (×3) of Urban100 'img060', 'img039'

(a) Original Image (b) Bicubic Interpolation Image (c) VDSR (d) Proposed Method

Figure 10 is an image of the super-resolution result three times expand from the "img060", "img039" image of Urban100. In Bicubic Interpolation, a small object is distorted, and a linear pattern is blurred. The VDSR showed better results than the Bicubic Interpolation image, but there was still small object distortion, and the "t" and background pattern were blurred and restored. The proposed method showed no distortion of small objects and 't' and the background was clear.

Table 1: Average PSNR/SSIM for scale factor ×2, ×3, ×4. Highlighted indicates the best performance.

Dataset	Scale	Bicubic PNSR/SSIM	VDSR PNSR/SSIM	Proposed Method PNSR/SSIM
Set 5	×2	34.532/0.956	36.634/0.968	**37.014/0.970**
	×3	24.780/0.792	32.401/0.930	**32.793/0.932**
	×4	23.083/0.706	29.698/0.883	**30.317/0.895**
Set 14	×2	30.445/0.903	31.831/0.918	**32.102/0.929**
	×3	27.072/0.812	28.329/0.841	**28.541/0.845**
	×4	25.219/0.734	26.316/0.772	**26.537/0.783**
Urban 100	×2	28.071/0.893	30.136/0.922	**30.634/0.927**
	×3	24.780/0.792	26.578/0.839	**27.128/0.848**
	×4	23.083/0.706	24.510/0.765	**25.148/0.786**
B 100	×2	31.013/0.904	31.958/0.913	**32.148/0.914**
	×3	27.723/0.804	28.685/0.829	**28.752/0.835**
	×4	26.078/0723	26.916/0.755	**27.134/0.763**

Table 1 shows the results of various data sets compared with the conventional method. The Set5 dataset consisted of 5 natural images, and the Set14 dataset consisted of 14 chapters. The B100 dataset was composed of 100 images of various situations, and the Urban100 dataset was composed of 100 various buildings. The results showed that PSNR and SSIM performance were improved over Bicubic Interpolation and VDSR.

5. Conclusion

In this paper, we proposed a super-resolution method combining Channel Attention and Residual Block for feature enhancement. The input image used was a low-resolution image. As a result, the possibility distortion and the meaningless expansion that occur in the Pre-upsampling method using Bicubic Interpolation and the like are prevented, and incorrect feature extraction is prevented. First, feature maps of Convolution and Residual Block were combined using Concatenation Layer for shallow feature

March - April 2020
ISSN: 0193-4120 Page No. 4409 - 4415

extraction. Thereafter, in order to emphasize high-frequency features such as contours and textures with shallow features, deep features were extracted by combining Residual Block and Channel Attention. Sub-Pixel Convolution instead of Deconvolution, the feature map size expand by reconstructing various features without duplication. The results show that the contour distortion and blur are improved through such a structure. Future research will need to study deeper network configurations and how to combine various features.

6. Acknowledgment

The present Research has been conducted by the Research Grant of Kwangwoon University in 2020.

References

[1] Lee DW, Lee SH, Han HH, Chae GS. Improved Skin Color Extraction Based on Flood Fill for Face Detection. Korea Convergence Society [Internet]. 2019 Jun 28;10(6):7–14. DOI: 10.15207/JKCS.2019.10.6.007

[2] Pyo S-K, Lee G, Park Y-S, Lee S-H. A license plate detection method based on contour extraction that adapts to environmental changes. Korea Convergence Society [Internet]. 2018 Sep 28;9(9):31–9. DOI: 10.15207/JKCS.2018.9.9.031

[3] Kim H-J, Park Y-S, Kim K-B, Lee S-H. Modified HOG Feature Extraction for Pedestrian Tracking. Korea Convergence Society [Internet]. 2019 Mar 28;10(3):39–47. DOI:10.15207/JKCS.2019.10.3.039

[4] Kim DI, Lee GS, Han GH, Lee SH. A Study on the Improvement of Skin Loss Area in Skin Color Extraction for Face Detection. Korea Convergence Society [Internet]. 2019 May 28;10(5):1–8. DOI:10.15207/JKCS.2019.10.5.001

[5] Lee DW, Lee SH, Han HH, Chae GS. Improved Skin Color Extraction Based on Flood Fill for Face Detection. Korea Convergence Society [Internet]. 2019 Jun 28;10(6):7–14. DOI: 10.15207/JKCS.2019.10.6.007

[6] Dong C, Loy CC, He K, Tang X. Image Super-Resolution Using Deep Convolutional Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence [Internet]. 2016 Feb 1;38(2):295–307. Available DOI:10.1109/TPAMI.2015.2439281

[7] Kim J, Lee JK, Lee KM. Accurate Image Super-Resolution Using Very Deep Convolutional Networks. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) [Internet]. IEEE; 2016. DOI:10.1109/CVPR.2016.182

[8] Park, J., Woo, S., Lee, J.Y., Kweon, I.S.: Bam: Bottleneck attention module. In: Proc. of British Machine Vision Conference (BMVC). (2018)

[9] Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A. Learning Deep Features for Discriminative Localization. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) [Internet]. IEEE; 2016. Available from: http://dx.doi.org/10.1109/CVPR.2016.319

[10] Shi W, Caballero J, Huszar F, Totz J, Aitken AP, Bishop R, et al. Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE; 2016. doi:10.1109/cvpr.2016.207.

[11] He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) [Internet]. IEEE; 2016. DOI:10.1109/CVPR.2016.90

[12] Huang G, Liu Z, Maaten L van der, Weinberger KQ. Densely Connected Convolutional Networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) [Internet]. IEEE; 2017. Available from: http://dx.doi.org/10.1109/CVPR.2017.243

[13] Kingma, D. P., Ba, J. Adam: A method for stochastic optimization. In: Conference paper at the 3rd International Conference for Learning Representations (ICLR). arXiv; 2015.