The Journal of The Institute of Internet, Broadcasting and Communication (IIBC)
Vol. 22, No. 5, pp.185-191, Oct. 31, 2022. pISSN 2289-0238, eISSN 2289-0246

https://doi.org/10.7236/JIIBC.2022.22.5.185
JIIBC 2022-5-27

# 국소부위 패턴 표현을 위한 샘플링 기반 초해상도 U-Net

# Sampling-based Super Resolution U-net for Pattern Expression of Local Areas

**이교석*, 갈원모**, 임명재***

Kyo-Seok Lee*, Won-Mo Gal**, Myung-Jae Lim*

**요 약** 본 연구에서는 U-Net, 잔차 신경망, 서브 픽셀 컨볼루션을 기반으로 새로운 초해상도 신경망을 제안한다. U-Net의 최대 풀링으로 인해 세부적인 정보의 손실이 일어나는 것을 막기 위해 서브 픽셀 컨볼루션을 활용한 다운 샘플링 그리고 연결을 제안한다. 이는 필터 안의 최대 값만으로 새로운 피처맵을 만드는 최대 풀링과 다르게 필터 안의 모든 픽셀을 사용한다. 2x2 크기의 필터가 지나가면서 왼쪽 위, 오른쪽 위, 왼쪽 아래, 오른쪽 아래의 픽셀들로만 이루어진 피처맵을 만든다. 이를 통해 크기가 절반이 되고, 피처맵이 개수가 4배가 된다. 그리고 연산량을 줄이기 위해 두 가지 방법을 제안했다. 첫 번째는 U-Net의 업 컨볼루션 대신 연산량이 없고, 성능이 더 좋은 서브 픽셀 컨볼루션을 사용한다. 두 번째는 U-Net의 연결 층 대신 두 피처 맵을 더하는 층을 사용한다. 밴치 마크 데이터 세트로 실험한 결과 스케일 2의 set5 데이터를 제외하고 모든 스케일 및 벤치마크 데이터 세트에서 더 나은 PSNR 값을 보여주고, 국소부위의 패턴을 명확하게 표현할 수 있었다.

**Abstract** In this study, we propose a novel super-resolution neural network based on U-Net, residual neural network, and sub-pixel convolution. To prevent the loss of detailed information due to the max pooling of U-Net, we propose down-sampling and connection using sub-pixel convolution. This uses all pixels in the filter, unlike the max pooling that creates a new feature map with only the max value in the filter. As a 2x2 size filter passes, it creates a feature map consisting only of pixels in the upper left, upper right, lower left, and lower right. This makes it half the size and quadruple the number of feature maps. And we propose two methods to reduce the computation. The first uses sub-pixel convolution, which has no computation, and has better performance, instead of up-convolution. The second uses a layer that adds two feature maps instead of the connection layer of the U-Net. Experiments with a banchmark dataset show better PSNR values on all scale and benchmark datasets except for set5 data on scale 2, and well represent local area patterns.

**Key Words :** Down-sampling, Single Image Super Resolution(SISR), Sub-pixel convolution, U-Net

*준회원, 을지대학교 의료IT학과
**정회원, 을지대학교 보건환경안전학과
접수일자 2022년 8월 10일, 수정완료 2022년 9월 10일
게재확정일자 2022년 10월 7일

Received: 10 August, 2022 / Revised: 10 September, 2022 / Accepted: 7 October, 2022
*Corresponding Author: lk04@eulji.ac.kr
Department of Medical IT, Eulji University, Korea

## Ⅰ. Introduction

Single Image Super-Resolution (SISR) is the extraction of high resolution (HR) images from low resolution (LR) images. It is widely used in applications such as medical imaging, satellite imaging, security, and surveillance, where high-frequency details are greatly desired.

SISR using deep learning generally utilizes the residual structure. Input and output have similar information, and make deep network learning differences easy and efficient[3][4][5][7]. It goes through the process of adding the input image to the last feature map.

In this study, a U-Net structure that connects the shallow feature map and the deep feature map was used. SISR proposed Down Sampling And Concatenate (DSAC) to reduce the size and maintain detailed information in place of max-pooling of the U-Net because detailed information is important.

In SISR, deeper models perform better. Therefore, previous studies have used various methods to reduce computation while deepening the model. We used sub-pixel convolution with excellent performance and little computation instead of up-convolution[2]. At this time, the number of feature maps decreased to a quarter, so the number of feature maps coming from the contracting path of the U-Net was reduced by half. In addition, the number of feature maps was reduced by performing an add operation instead of concatenation.

## Ⅱ. Related work

### 1. DRRN

DRRN and earlier SR models (SRCNN, VDSR, DRCN) have the same input size and output size. When these models preprocess images, they reduce the high resolution image by scale factor through bicubic interpolation to create low resolution images, and then grow it to the same size as HR through bicubic interpolation. After scaling, convert format RGB into YUV, of which Y is used for training and testing. When training, the low resolution image is cut into patches. when testing, four bench mark data sets (Set5, Set14, BSD100, Urban100) are used without a patch. And this is evaluated through SSIM and PSNR values. PNSR is a method of comparing absolute input pixel values. This method can be visually bad because it uses only the difference of each pixel. It is SSIM that has improved this shortcoming, and it was created to evaluate the characteristics of a person visually. It consists of Luminance, contrast, and structure.

### 2. U-Net

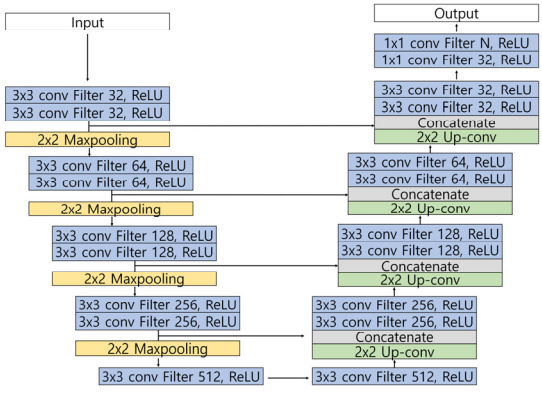


**그림 1. U-Net의 구조.**
**Fig. 1. U-Net architecture.**

U-Net is a fully-convolutional network(FCN) based model that is proposed for image segmentation in the biomedical field. As shown in figure1, it has an end-to-end scheme and consists of a contracting path, expansive path, and skip architecture. The Contracting path has four blocks, each block consists of two convolutions, rectified linear unit(ReLU) activation function and 2x2 max-pooling. The expansive path has four blocks, each block consists of up-convolution , two convolutions, a ReLU activation layer. The last block additionally

has two 1x1 convolution layers for nonlinear prediction. Skip architecture concatenates the feature map before each max-pooling layer of the contracting path with the feature map from each up convolution layer of the expansive path[1].

### 3. ESPCN

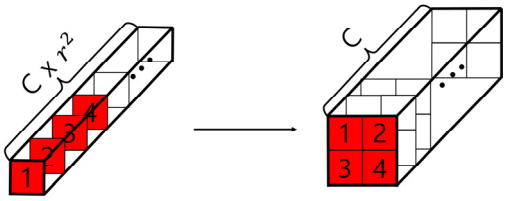


**그림 2. r=2일 때 서브 픽셀 컨볼루션.**
**Fig. 2. Sub-pixel convolution when r=2.**

Previous SISR methods use input images after increasing their size through cubic interpolation. At this time, the amount of calculation increases due to the increased input value[3][7].

To avoid this disadvantage, the process of increasing the size of the image is placed at the end. To increase the size of the image, it proposes an efficient sub-pixel convolution layer (ESPCN) that increases the size by making several feature maps into one feature map, rather than interpolation and up-convolution. Increase the number of feature maps by $r^2$ than LR, and then combine the feature maps to create an HR image. This means that the model implicitly learns the preprocessing process required for SR through the layer. Therefore, the network can learn to switch from LR to HR more precisely without separate interpolation.

## Ⅲ. Proposed network

Previous SR uses LR image as an input image by increasing the size through the bicubic interpolation[3][4][5][7]. The LR image is treated as a blurred HR image, and the SR process is treated as if reconstructing the blurred HR image. The Contracting path of U-Net extract key information and the expansive path of U-Net generate a better feature map with a feature map from skip architecture. It has a structure that connects a shallow feature map and a deep feature map through skip architecture to deliver the input value to the output value[1]. So we think that U-Net architecture is appropriate for the reconstruction method.

### 1. Network architecture

U-Net consists of contracting path, expanding path, and skip architecture. The contracting path captures the context of the image, and the expansive path expands the feature map to provide accurate localization. Skip architecture extracts features of the shallow layer of the CNN network that are local and detailed, while the deep layer extracts features that are general and abstract[1]. It combines these two layers that extract different features, allowing both local and global information to be included. We used this structure.

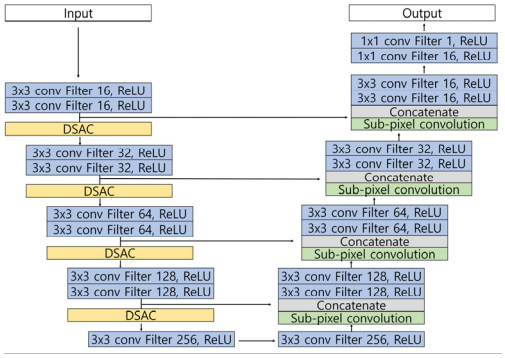


**그림 3. 제안한 네트워크 구조.**
**Fig. 3. Proposed network architecture.**

The proposed network architecture consists of a contracting path, expanding path, and skip architecture, as shown in the U-Net. The Contracting path consists of four blocks. Each block consists of two 3x3 convolution layers, a

ReLU activation function, and a down sampling that quadruples the number of feature maps and halves the size. The convolution layer of the first block has and generates 16 feature maps. The feature map doubles for each block. The expansive path consists of four blocks. Each block consists of an upsampling using an ESPCN method that doubles its size by combining four feature maps into one feature map, two 3x3 convolution layers, and a ReLU activation function. The last block additionally has two 1x1 convolution layers for nonlinear prediction. The convolution layer of the first block generates 256 feature maps. The feature map halves for each block. Skip architecture adds the feature map before each downsampling layer of the contraction path with the feature map from each upsampling layer of the expansive path. Between the contracting path and expansive path, there are two 3x3 convolution and ReLU activation function layers, which are called bottlenecks.

## 2. Down sampling and concatenation

The contracting path of the U-Net captures the context of the image and extracts local and detailed information as the size of the feature map halves[1].

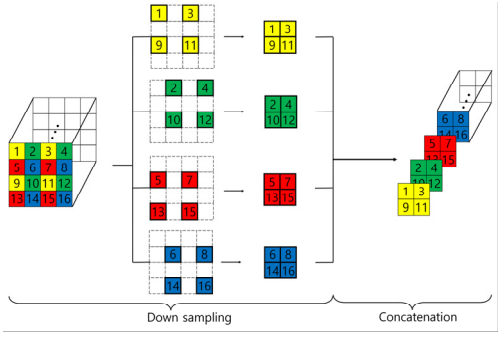


**그림 4. 제안한 네트워크의 다운 샘플링과 연결 구조.**
**Fig. 4. Down sampling and concatenation of proposed network.**

In this process, in order to reduce the size of the feature map, a max-pooling layer with a stride of 2 and a filter size of 2x2 is used. Using this, information loss occurs because the feature map is made using only the largest value of the four pixels in the filter. The loss of detailed information in SR is a problem. In this paper, using all four pixels, unlike the max-pooling layer, to prevent the loss of detailed information. As shown in Fig. 5, As a filter with a 2x2 size and a stride of 2 passes by, pixels are extracted to make a feature map with half size. In this case, the created feature map is made of pixels corresponding to the same position of the filter.

## 3. Sub-pixel convolution

The expansive path of the U-Net extends the feature map and provides accurate localization. In this process, the up-convolution layer is used to double the size of the feature map. The up-convolution, which is used to increase size, has a large amount of computation but can restore a lot of information. However, it is not possible to restore information properly by using methods such as bicubic interpolation to reduce computation. In this paper, we use ESPCN layers to reduce computation while properly restoring information. As shown in Figure 3, a feature map with $2^2$ feature maps is created that is doubled in size. Concatenate the increased size feature map and feature map from skip architecture.

## 4. Add layer

Skip architecture of the U-Net concatenate the feature map before each max-pooling layer of the contraction path with the feature map from each up-convolution layer of the expansive path. Between the contracting path and expansive path.

This method increases the number of feature maps to be processed. This is one of the issues raised in the SISR. To reduce parameters and maintain performance, add operations are used as in the residual structure. This keeps the

performance intact and reduces the number of parameters. Using sub-pixel convolution with r=2, the number of feature maps is reduced by a quarter, making the feature map produced by the second convolution of each block of the continuing path half the feature map produced by the first convolution.

## Ⅳ. Experiment & Result

### 1. Datasets

For training, we used 291 images. 200 images were taken from the BSD dataset, and 91 images were taken from T91. For validation, 200 images were taken from the BSD dataset. It uses benchmark datasets Set5, Set14, BSD100, and Urban100[6], which are widely used for testing.

### 2. Implement details

As shown in Tabel1, We conducted training and testing on three scales (2x, 3x, and 4x). Training images are split into 128x128 patches, with the stride of 83, by considering our downsampling method. We set the mini-batch size of SGD to 128, momentum parameter to 0.9, and learning rate 1e-1. Training our model roughly takes 30 minutes with Titan X GPU.

### 3. Comparison with other models

Figure 5 shows the performance comparison between other models and our models through four benchmark datasets. Except for Set5 data on scale x2, we can confirm that our model performs better for all scales and all datasets. And as the scale increases, the performance differences of the other models stand out.

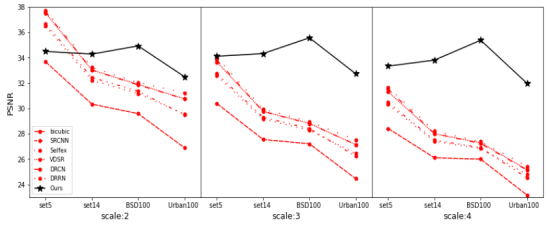


**그림 5. 스케일 팩터가 x2, x3, x4인 4개의 벤치마크 데이터 세트에 대한 정성적 비교.**
**Fig. 5. Qualitative comparison for four benchmark datasets with scale factors of x2, x3, and x4.**

However, when comparing SSIM, it can be seen that it produces a performance similar to that of cubic interpolation, as shown in Figure 6. In other words, the pixel value is properly expressed, but it is not good when viewed from a human eye.

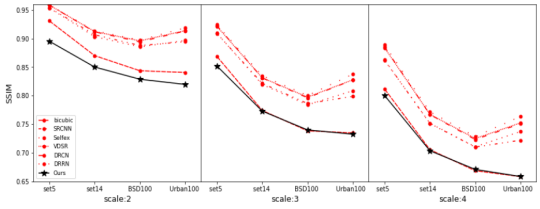


**그림 6. 스케일 팩터가 x2, x3, x4인 4개의 벤치마크 데이터 세트에 대한 정성적 비교.**
**Fig. 6. Qualitative comparison for four benchmark datasets with scale factors of x2, x3, and x4.**

We compare our model with other models through two images of Urban100. As shown in Figures 7 and 8, when we zoom in on some of the images, we can see that our model is good at representing patterns of small areas of large images.

**표 1. 실험 파라미터**
**Table 1. Simulation Parameters**

| Parameter | |
|---|---|
| Learning rate | 1e-1 |
| Loss function | MSE |
| Optimizer | SGD |
| Patch size | 128 |
| Epoch | 400 |

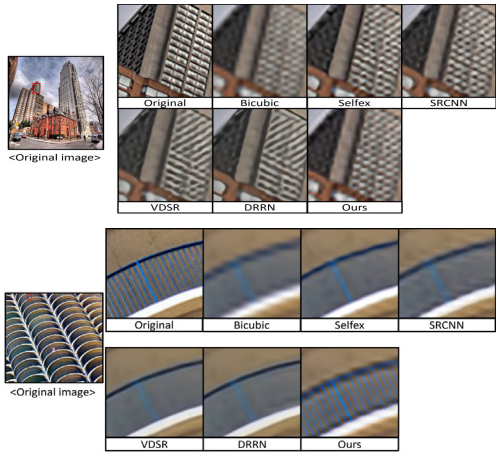


**그림 7. Urban100 데이터의 "IMG20", "IMG100"대한 각 모델별 출력.**

**Fig. 7. Output for each model for "IMG20", "IMG100" of Urban100 data.**

## V. Conclusion

In this paper, we proposed a SISR algorithm-based U-Net. The lost information was minimized by using down-sampling and concatenation instead of max-pooling of the U-Net. In addition, instead of the up-convolution, sub-pixel convolution with a small computation amount and good performance was used. In order to reduce the amount of computation, the number of second convolution filters in each block in the contracting path of the U-Net was half the number of first convolution filters. In addition, add was used instead of concatenating in the expansive path. As a result, it showed the best performance in all scale factors and all benchmark datasets except for the set5 data of scale2.

## References

[1] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. MICCAI 2015 Part III LNCS 9351, pp. 234-241.
DOI: https:// doi.org/10.1007/978-3-662-54345-0_3

[2] Wenzhe Shi, Jose Caballero, Ferenc Huszar´, Johannes Totz, Andrew P. Aitken, Rob Bishop, Daniel Rueckert, Zehan, (2016). Real-Time Single Image and Video Super-Resolution Using an Efficient SubPixel Convolutional Neural Network, CVPR2016, 1874-1884.
DOI: https://doi.org/10. 1109/cvpr.2016.207

[3] Jiwon Kim, Jung Kwon Lee and Kyoung Mu Lee, (2016). Accurate Image Super-Resolution Using Very Deep Convolutional Networks, CVPR2016, 1646-1654.
DOI: https://doi.org/10.1109/cvpr.2016. 182

[4] Jiwon Kim, Jung Kwon Lee and Kyoung Mu Lee, (2016). Deeply-Recursive Convolutional Network for Image Super-Resolution, CVPR2016, 1637-1645.
DOI: https://doi.org/10.1109/cvpr.2016. 181

[5] Ying Tai, Jian Yang, and Xiaoming Liu, (2017), Image Super-Resolution via Deep Recursive Residual Network CVPR2017, 3147-3155.
DOI: https://doi.org/10.1109/cvpr.2017.298

[6] Jun-Jie Huang, Tianrui Liu, Pier Luigi Dragotti, and Tania Stathaki, (2015). SRHRF+: Self-Example Enhanced Single Image Super-Resolution Using Hierarchical Random Forests, CVPR2015, 71-79.
DOI: https://doi.org/ 10.1109/cvprw.2017.144

[7] Chao Dong, Chen Change Loy, Kaiming He, Xiaoou Tang, Image super-resolution using deep convolutional networks. In CVPR, 2016, 295-307.
DOI: https://doi.org/10.1109/TPAMI.2015.2439281

[8] Lee Ju Hee, Kang Bong soon. Improving Performance of Machine Learning-Based Algorithms with Adaptive Learning Rate. The Journal of KIIT, Vol. 18, No. 10, pp. 9-14, 2020.
DOI: https://doi.org/10.14801/jkiit.2020.18.10.9

[9] Joohyun Song, Deokwoo Lee. Classification of Respiratory States based on Visual Information using Deep Learning. Journal of the Korea Academia-Industrial cooperation Society(JKAIS), Vol. 22, No. 5, pp. 296-302, 2021.
DOI: http://dx.doi.org/10.5762/KAIS.2021.22.5.296

[10] Myung-Jae Lim, Jae-Ju An, So-Hee Jun and Young-Man Kwon4*.(2020). Efficient algorithm for malware classification: n-gram MCSC. International Journal of Computing and Digital Systems, March(2). 179-185.
DOI: http://dx.doi. org/10.12785/ijcds/090204

[11] Myung-Jae Lim, So-Hee Jun, Won-Mo Gal, and Young-Man Kwon.(2020). THE ENHANCED VERSION OF TF-IDF FEATURE VECTOR FOR MALWARE DETECTION. International Journal of Heat and Mass Transfer. specialissue, 161-172.
DOI:http://dx.doi.org/10.17654/ HMSI20161

## 저 자 소 개

**이 교 석(준회원)**

• 을지대학교 성남캠퍼스 의료IT학과 학부생

**갈 원 모(정회원)**

• Professor of Health and Environmental Safety department at Eulji University

**임 명 재(정회원)**

• 을지대학교 바이오 융합 대학 의료IT학과 교수