Computer Science and Information Systems 22(4):1617–1636 https://doi.org/10.2298/CSIS240810064A

A study on Multi-scale Attention dense U-Net for image denoising method

MingShou An , XuHang Zhao , Hye-Youn Lim and Dae-Seong Kang

Dong-A University, Dept. of Electronics Engineering, 37 Nakdong-daero 550 beon-gil
Saha-gu, Busan, Korea
anmingshou@xatu.edu.cn
dskang@dau.ac.kr

Abstract. Although many models that have applied learning exhibit good performance, the dataset or image generation and transmission process used for learning may contain noise, which cannot produce the expected results and performance. The representative image denoising technique using deep neural networks generates noisy images by forcibly adding special noise to the original image and learning to make it the same as the original image. However, the performance of deep neural networks depends on depth, and to improve performance, increasing only depth will reach a performance saturation state, which will encounter difficulties. In order to improve these issues, this article applies the Multi-scale Attention model to the representative denoising deep learning model U-Net, to suppress unnecessary information and provide functionality that only emphasizes important information. In a new modular approach, the given input value is divided into two parts based on its internal relationship: the part where the important parts are concentrated and the part where the important parts are concentrated through spatial information. The attention unit based Outburst structure, which combines the two parts after parallel execution, has been implemented, demonstrating better performance than existing models. Moreover, without adding too many parameters, more spatial feature maps than other models are generated by focusing on the effects of components, not only through PSNR and SSIM. The improved performance was also confirmed by removing noisy in images.

Keywords: deep learning, image denoising, Multi-scale Attention, U-Net, outburst structure.

1. Introduction

The existing image processing and analysis involve the entire field of digital image information preprocessing, feature extraction, image restoration and image compression, etc. At present, artificial modeling of human learning and reasoning abilities involves a field of artificial intelligence such as character recognition and image pattern analysis - deep learning. It is worth mentioning that deep learning has emerged in the field of image processing by directly utilizing image spatial information for feature extraction [1]. Representative deep learning techniques for image processing include image classification, object detection, etc. Image classification marks pre determined class information as image data for learning, and the learned model takes the image to be classified as input and outputs class information to distinguish the types of objects. Representative deep learning technologies in these fields include AlexNet [2], VGGNet [3], GoogLeNet [4], ResNet [5], and many other deep learning models based on CNN [6], which have achieved significant results. The performance of deep learning models may be affected by various noises contained in the input images used during learning. Moreover, in the process of image generation and transmission, there will inevitably be noise involved. Deep learning models that learn through input data also calculate noise during learning, so if the input data is heated by noise, it can lead to performance degradation. So in practical environments, denoising during evaluation is essential [7]. To improve these issues, image denoising techniques are needed. Image denoising technology has been widely applied in multiple fields such as restoring image details and accepting images as inputs. At present, in the field of image denoising, research on the application of deep neural networks is very active in order to improve performance [8].

In order to improve the performance of image denoising, simply increasing the depth of the deep neural network will result in too much computational complexity and may lead to difficulties in performance degradation or saturation [9]. To improve these issues, based on noise pattern prediction, the structure of the neural network was changed to a parallel form instead of increasing the depth of the neural network, thereby expanding the scope of the neural network. By improving the existing Attention units, a new module called multiscale Attention units has been implemented, which can extract more spatial feature information, suppress unnecessary information, and only emphasize important information. Multiscale Attention units are composed of parallel parts that concentrate what is important and where spatial information is important through internal relationships with inputs [10]. In terms of overall structure, noise can be effectively removed by using an Outburst structure [11] that surges feature information in the latter half of the neural network. This can be widely applied to systems that require denoising in image processing.

2. Related Work

2.1. DnCNN

DnCNN(Denoising Convolutional Neural Network) [12] is a deep learning technique that utilizes CNN to implement image denoising. Using Additive White Gaussian Noise (AWGN) as noise, and training the model to remove noise. The existing denoising techniques have long computation time, complex parameter settings, and require a large amount of computation and direct manual interference. The focus of DnCNN is not to directly remove noise from noisy images, but to separate noise from noisy images. The use of CNN ensures the flexibility of image dinoising and achieves performance improvement through residual learning [13] and Batch Normalization [14].

In addition, they also modified the network structure based on VGGNet(Visual Geometry Group) to achieve image denoising. In the technology used here, residual learning is used to improve accuracy, even if the model becomes deeper, it can maintain generalization well, rather than adding parameters, and the calculation is not complicated. And Batch Normalization, through its own generalization process, has low sensitivity and is not affected by parameter size during learning. It can greatly set the learning rate and achieve fast learning. The size of the each convolutional filter of DnCNN is 3x3. The

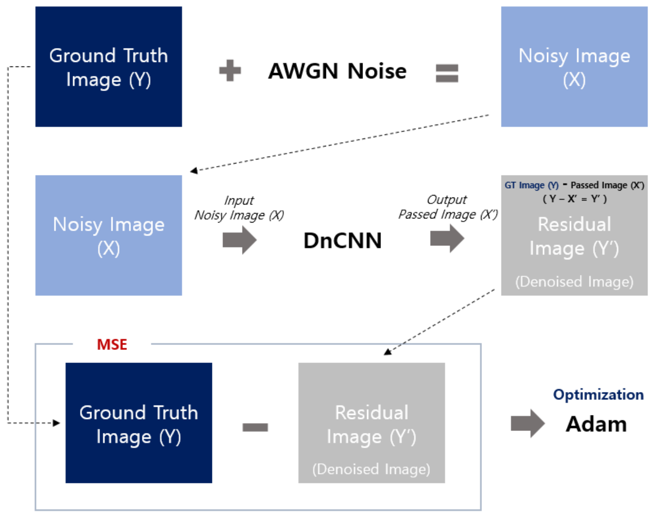


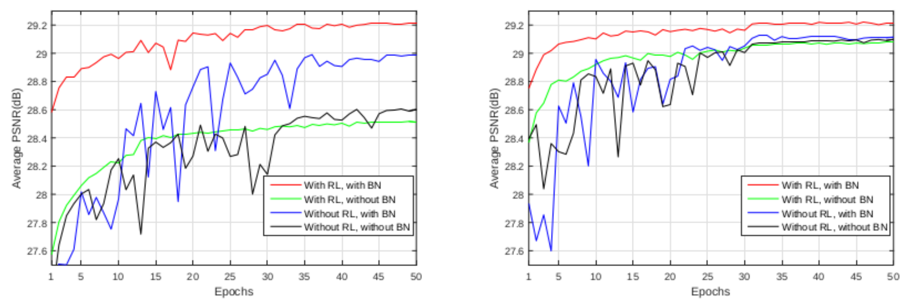
Fig. 1. The structure of DnCNN

model in the Denoising field determines the receptive field size based on the effective patch size. High noise levels typically require a larger effective patch size to capture more detailed features. DnCNN provides a noise level fixed at 25, analyzing the effective patch size of leading noising methods for depth design.The structure of DnCNN is shown in Fig.1, which creates a Noise Image (X) by obtaining the Ground truth image (Y) and adding AWGN noise. Create an image $(X^{'})$ using a CNN network in Noisy Image (X), subtract the resulting image from the original image (Y), and generate a Residual Image $(Y^{'})$. The final calculation of MSE for Y and $Y^{'}$ is reflected in the optimization of the CNN network. Fig.2 shows the average PSNR with and without batch normalization (BN) and residual learning (RL). Residual learning converges faster and more stably than original mapping learning. If residual learning and batch normalization are used simultaneously, it will converge faster than original mapping and exhibit better denoising performance, especially helping SGD and Adam achieve better performance results[15].

2.2. Attention Model

The research on network structure has developed in multiple aspects such as depth and breadth. So far, Attention[16] has focused on research in specific fields, and has not conducted much research in the field of imaging. Recently, in combination with Residual technology, various networks have been studied in the field of image related fields, and Attention as a component of the model is being developed. Attention establishes a complementary relationship and demonstrates meaningful results in improving deep learning performance. Multiple benchmark tests such as ImageNet classification, COCO detection, VOC detection, and multiple models such as ResNet, WideResNet, ResNext, and MobileNet have been validated [17].

Firstly, from the structure of the Attention module, convolutional features are extracted from the generalized Signmode state map and element wise product is performed.



(a) SGD (b) Adam

Fig. 2. The average of PSNR with BN and RL

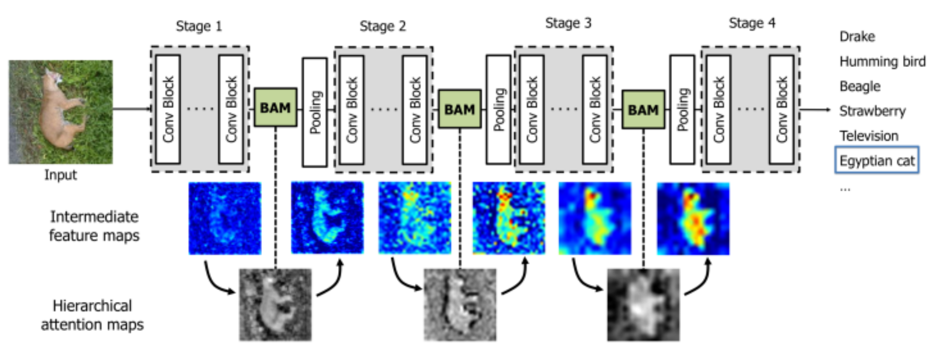


Fig. 3. A general deep learning model with BAM

The performance of Attention plays an important role in achieving greater performance improvement with less computation. The attention map is decomposed into channel wise/spatial wise for calculation. Typical Attention modules include Bottleneck Attention Module (BAM) and Convolutional Block Attention Module (CBAM). These consist of very simple pooling and convolutions. Modularize self tension to easily connect to any CNN. In addition, end-to-end training can be conducted together with existing networks. Fig.3 shows a general deep learning model with added BAM structure. Before the amount of information in the bottleneck interval in the neural network structure is reduced, by adding a BAM structure, can increase the information of important parts while reducing the information of unimportant parts. As a follow-up study, CBAM is a variant of the combination of pooling, spatial, and channel attention, which exhibits higher performance than BAM. The CBAM in Channel attention is different from the previous BAM, as it combines average pool and max pool. The max pool and avg pool values of the 3D feature map are used as meaningful feature maps in global attention. Two pooling features share values with the same meaning, so a shared MLP can be used while reducing the number of parameters. In spatial attention, it is also symmetric, and spatial attention is calculated using a convolution [18].

2.3. U-Net

U-Net [19] combines the concepts of ResNet and Autoencoder, and is a model constructed based on the end-to-end fully convolutional network (FCN) [20] proposed in the Biomedical field for image segmentation. This modifies the FCN structure to provide more accurate segmentation in situations with limited data. As shown in Fig.4, the network used for overall image flow and the network used for precise localization are composed in a symmetrical form.

U-Net has a U-shaped structure, which can be roughly divided into three parts based on the center. The first is to extract the semantic information of image pixels over a large range as the encoding role Contracting Path. The second is to match the semantic information with pixel position information as the decoding role Expansive Path. Finally, there is a conversion interval from the contraction path to the expansion path. This structure eliminates the existing problem of unnecessary repetition of overlapping patches, thereby improving performance. The Contracting Path involves repeating the 3x3 conv layer operation twice, which reduces the size of the feature map. The activation function uses ReLU. In each low sampling process, the size of the feature map will be reduced by half, but the number of channels will increase by twice. At each upsampling, the size of the feature map will double and the number of channels will decrease by half for the Expansive Path. Contrary to Contracting Path, we can use it to expand the size of feature maps. Then, the final layer is processed through a 1x1 conv layer.

3. Proposal Method

Noise may be generated during image generation, transmission, or processing, which can lead to performance degradation during image analysis. In addition, the presence of noise in images can also affect the performance of deep learning models. To improve these issues, this article proposes a Multi-scale Attention U-Net image denoising method. The

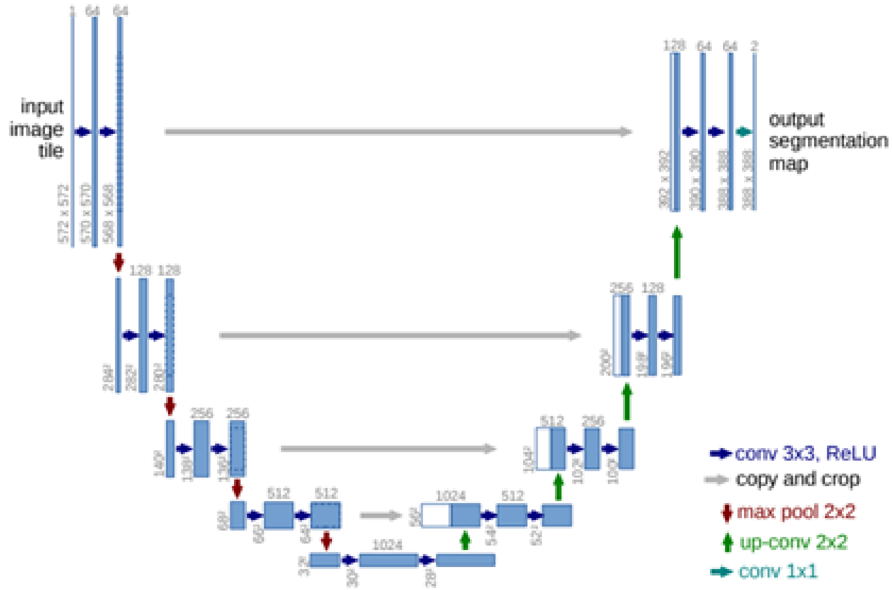


Fig. 4. The structure of U-Net

recommended approach is to follow the process shown in Figure 5 and apply multi-scale attention unit to U-Net based on the typical denoising deep learning model DnCNN, and convert it to an Outburst structure. Using the Attention model to suppress unnecessary information and only emphasize important information. In addition, the second half of the neural network can effectively achieve noise reduction by using the Outburst structure to rapidly increase feature information.

3.1. Improved Attention DnCNN

The performance of deep neural networks depends on depth, but to improve performance, simply increasing depth may reach a performance saturation state. In order to solve this problem, this article did not increase the depth of the neural network, but instead connected two neural networks in parallel, expanded the width of the network, and formed a multi-layer tension module that composed of channel and spatial attention unit. It extracted more scale feature maps and improved the performance saturation state. One of the two networks used for parallel connections is built based on DnCNN. Another approach is to add a Dilated Layer in the front and back half of the middle layer, giving the network scalability and helping to extract more complex feature maps.

The Multi scale Attention unit is configured to execute the Spatial function again after parallel execution of the Channel and Spatial regions and merging their respective outputs. In the field of Channel, what is more important for a given input value is focused on the information about the channel, rather than spatial information. In addition, as shown in Fig. 6, as all nodes are connected using a fully connected layer, the weights will be shared and output more prominent information together with the input value multiple.

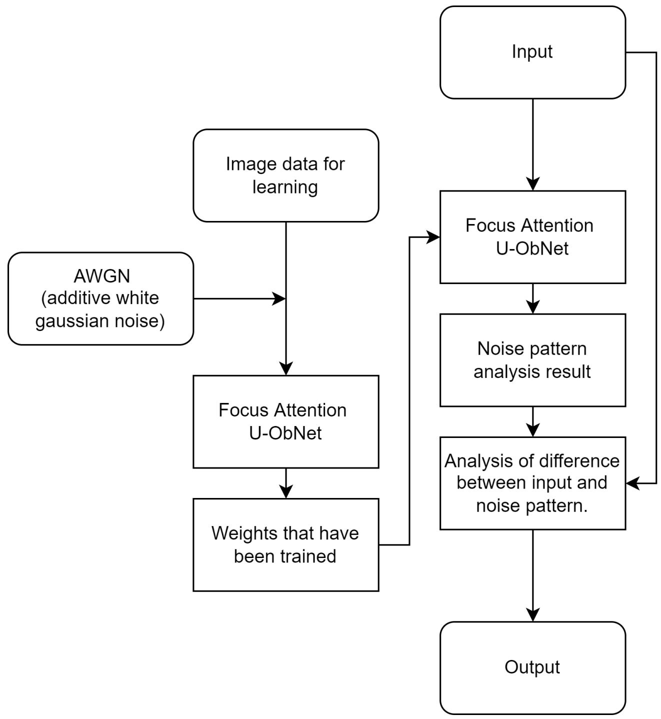


Fig. 5. The flowchart of our proposed method

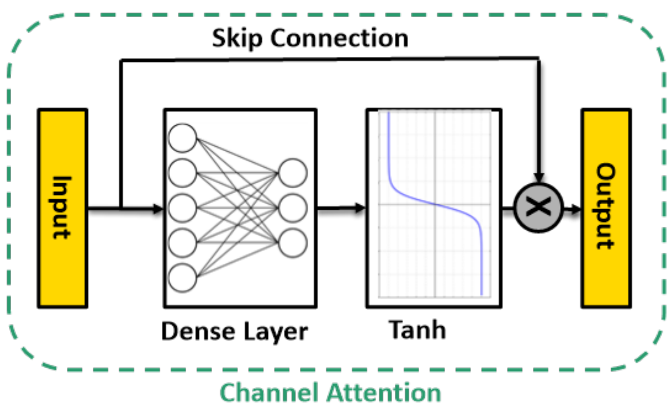


Fig. 6. The structure of Channel Attention

The Spatial region does not consider channel information and only focuses on the spatial importance of input values. As shown in Fig. 7, it is composed by reducing the number of channels and simply continuously arranging convolutional layers. When channels exist, they can cause a sharp increase in parameters during the operation process, which has a significant impact on learning speed. By using a 1x1 conv layer to reduce the computational complexity of initial input values and sharply reduce information about channels, the three-dimensional input is transformed into two-dimensional, thus concentrating only the spatial area. Secondly, after passing through the 1x1 conv layer, only the two-dimensional spatial layer as a channel can continuously apply the 3x3 conv layer, thereby outputting spatial features with only important features as more features.

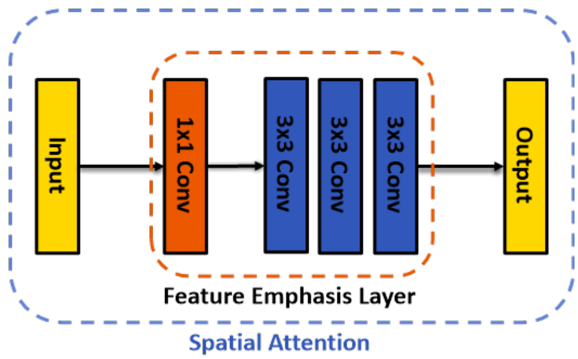


Fig. 7. The structure of Spatial Attention

Finally, merge the output of the Channel region with the output of the Spatial region. By combining channel information and spatial information, important information about channels and spaces can be grasped, and the necessary effective information can be extracted. From a broader perspective, it's like outputting only important and necessary information. Therefore, this output value is used as input in the Spatial module, highlighting spatial and channel elements as a whole, and extracting various important features. The last application of the Spatial module requires the use of both spatial and channel information. Unlike the Spatial module applied earlier, the 1x1conv layer used here serves to delete channel information, therefore the 1x1conv layer is composed in the form of deletion. As shown in Fig. 8, from a larger perspective, the three modules are combined in parallel and serial forms as a multi-scale salient module application.

When two networks(DnCNN and Multi-scale Attention) are connected in parallel, connect the Multi-scale Attention module to the input and output layers. Then, as shown in Fig. 9, the initial model DnCNN proposed in this paper using a Multi-scale Attention module was re-learned through Skip Connection between each network input and output.

In the initial model, it is not simply to increase the depth of the neural network, but to combine two neural network models in parallel, expanding the breadth of the model. In addition, by using ResNet's skip connections, more features can be extracted by referring to the compression function of Autoencoder and previous values. Based on the structure of U-Net and combined with Attention DnCNN, the performance of image denoising meth-

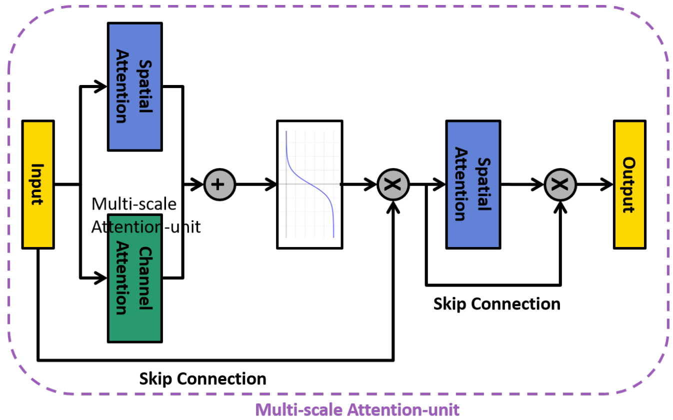


Fig. 8. The structure of Multi-scale Attention

ods can be improved. The existing U-Net adopts the core technologies of Autoencoder and ResNet shortcuts. Moreover, the existing U-Net undergoes four downsampling, but to prevent spatial feature loss at low resolutions, the recommended method only performs two times.

3.2. Outburst Structure

The initial model of this article is constructed in parallel rather than serial form. On the one hand, it adopts the structure of DnCNN, and on the other hand, it uses a dilated layer instead of ordinary convolutional layers to extract feature maps from multiple different levels, while attempting to reduce the number of parameters. Although the performance has been improved, the model also has a flaw.

If layers are added to achieve better performance, the number of parameters will actually increase, leading to a decrease in learning speed and performance. The model developed to compensate for its shortcomings utilizes the structure of U-Net. U-Net can also be seen as a model composed of a combination of the structure of an autoencoder and skip connections. Basically, autoencoders coordinate the spatial size of layers through downsampling and upsampling, reducing the number of parameters required for learning, which is an effective method. Although it is a model of the same length, the size of the kernel decreases and the computational load also decreases. Therefore, the learning speed has also been improved. The decoder has the advantage of having the same output size as the input image, and by extracting features from small-sized images, more feature maps can be obtained, thereby improving learning speed and performance. Every time downsampling is performed, the number of upsampling executions in the decoder region will increase equally. Jumping connections with kernels of the same size can prevent spatial loss caused by downsampling. The improved model is based on the fact that simply passing the original value in skip connections is inefficient. Therefore, for the scalability of the feature map, highlighted information will be passed when connecting the initial model. From the overall structure perspective, this method is also a parallel connection

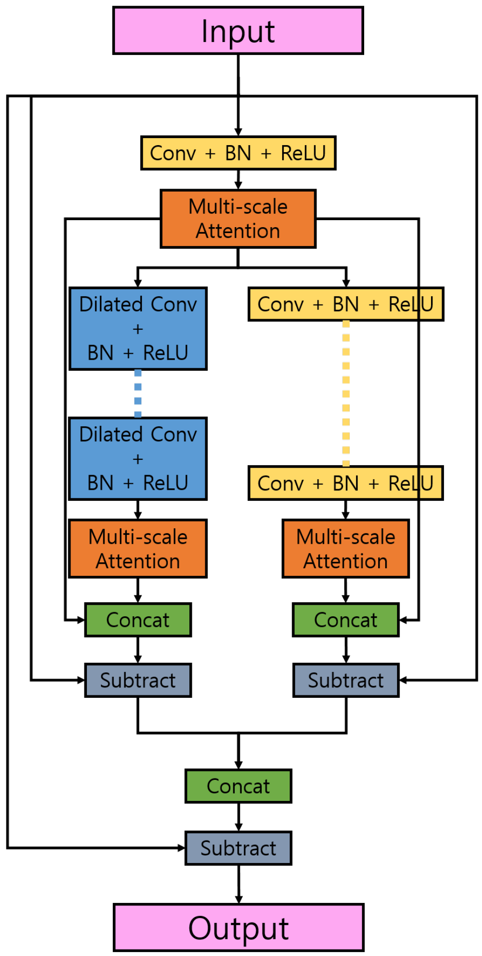


Fig. 9. The structure of Attention DnCNN

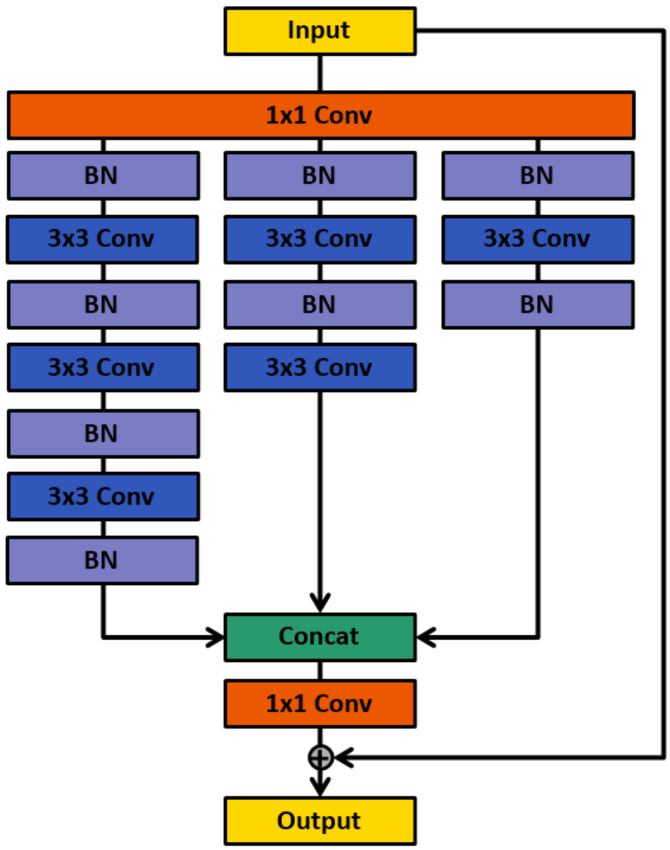


Fig. 10. The outburst structure of our method

rather than a serial connection, thus enhancing the scalability of the network. The performance of pattern analysis in deep learning basically depends on how many different feature maps are extracted. So far, if the model is improved based on parallel connections without increasing the number of parameters, the focus of this article is on how to effectively increase the extracted feature maps. A novel Outburst structure is proposed by explosively adding feature maps using the structure of xception.

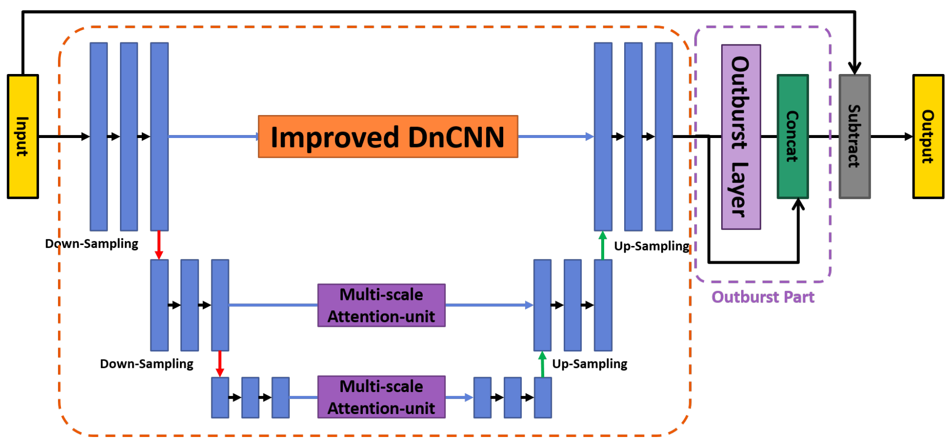


Fig. 11. The structure of our proposed method

The Outburst structure, as the name suggests, refers to an explosion, which refers to a rapid change. The sharp changes also mean the diversity and quantity of feature maps. Outburst can be roughly divided into two parts. If the current model is still part of a calm flow, then a sharp change is made in the latter half by adding the deformation structure of the xception shown in Fig.10. The calm part, also known as the preheating zone, plays a role in consolidating the feature map and extracting many features with sufficient diversity, so it is very important to maintain it well. After the bottleneck structure compression process, it enters the Outburst region and is endowed with diversity by various convolutional layers. Continuous convolutional layers make the features very prominent. The final output end will combine all of these to output. The output value does not remove noise, but rather captures the pattern of noise, indicating its association with the larger framework of DnCNN introduced earlier. Eliminating noise by having a difference in the output of the noise mode through Outburst with the input image containing noise.

3.3. The structure of our proposed method

The suggested methods can be divided into Multi scale Attention unit and Outburst structures, as shown in Fig.11. From the suggested model configuration, the first thing to see is the structural changes. Attention DnCNN uses the calm part as the preheating stage, while the Outburst part is associated with the input of the deformation structure that guides the xception. Effectively extract multiple feature maps from the calm part and endow the feature maps extracted from the Outburst part with scalability. Secondly, the Multi scale

Attention unit is used to assist in extracting various feature maps during the warm-up stage. Compared with traditional Attention modules, the number of layers has increased, allowing for more efficient feature extraction using channel and spatial information.

4. Experimental Results

4.1. Experimental method

In this paper, the learning dataset used 20000 images with dimensions of 180 * 180. The test dataset was tested using grayscale images Set12 and BSD68[21]. The loss function uses Adam, The performance of the epoch remains unchanged after more than 80 attempts, and if executed further, it will deteriorate. Therefore, only 80 attempts were made. For smooth learning, set the learning rate to 1e-3. At the beginning, use a higher learning rate to quickly reduce the value of the loss function before 80 iterations, and then set it to 1e-7 to adjust the safety and details of the loss function.

In order to conduct performance evaluation, experiments were conducted comparing PSNR (Peak Signal to Noise Ratio) and SSIM (Structural Similarity Index Map) considering human visual image quality differences with other models. PSNR evaluates the quality loss information of generated and compressed videos. The less the loss, the higher the price can be confirmed. If it is a lossless video, its MSE will be 0 and cannot be defined.

$$PSNR = 10log_{10}(\frac{R^2}{MSE}) \tag{1}$$

SSIM (Structural Similarity) is an indicator that measures the similarity between two images. This indicator was first proposed by the Laboratory for Image and Video Engineering at the University of Texas at Austin. Among the two images used by SSIM, one is an uncompressed and undistorted image, and the other is a distorted image.

$$SSIM = [l(x, y)]^\alpha \cdot [c(x, y)]^\beta \cdot [s(x, y)]^\gamma \tag{2}$$

Table 1 shows the results of the comparison between our method and other methods tested on BSD68 dataset. The test results indicate that the PSNR value of our method is better than others.

Tables 2 shows the PSNR and SSIM values tested using our method and other methods on datasets Set12 and BSD68. Our model tested results on the Set 12 dataset are PSNR=30.52/SSIM=0.9315, and on the BSD68 dataset are PSNR=29.43/SSIM=0.9106. The test results indicate that the PSNR and SSIM values tested using our method perform better than other models. From these results, it can be seen that if used for practical noise reduction, universality is feasible.

Table 3 shows the PSNR values of each method tested on 12 types of images in the SET12 dataset. Although it can be seen that the performance of our method is not as good as other models in some class images, it overall shows good performance, with the highest average PSNR value. It can be said that overall, our model showed the better performance than others.

Fig. 12 shows the denoising effect of Monarch, one of the Set12 images, which contains noise. By zooming in on specific parts of the butterfly wing texture, it can be seen that the recommended model has the best level of denoising.

Table 1. The comparison results of our method and other methods

Num	Models	PSNR(dB)
1	BM3D[22]	28.57
2	EPLL[23]	28.68
3	CSF[24]	28.74
4	WNNM[25]	28.83
5	TNRD[26]	28.92
6	IRCNN[27]	29.15
7	FFDNet[28]	29.19
8	ECDNet[29]	29.22
9	DnCNN[12]	29.23
10	ADNet[30]	29.25
11	Ours	29.43

Table 2. The average values of PSNR and SSIM of methods with SET12 and BSD68 datasets

Target	Dataset	BM3D	TNRD	DnCNN	IRCNN	Ours
PSNR	Set12	29.97	30.05	30.44	30.38	30.52
	BSD68	28.57	28.92	29.23	29.15	29.43
SSIM	Set12	0.8505	0.8515	0.8618	0.8601	0.9315
	BSD68	0.8017	0.8148	0.8278	0.8249	0.9106

Table 3. PSNR values of methods with 12 classes of SET12 dataset

Class	BM3D	MLP	TNRD	DnCNN	IRCNN	Ours
C.man	29.45	29.61	29.72	30.18	30.08	30.20
House	32.85	32.56	32.53	33.06	33.06	33.32
Peppers	30.16	30.30	30.57	30.87	30.88	30.96
Starfish	28.56	28.82	29.02	29.41	29.27	29.31
Monarch	29.25	29.61	29.85	30.28	30.09	30.49
Airplane	28.42	28.82	28.88	29.13	29.12	29.15
Parrot	28.93	29.25	29.18	29.43	29.47	29.52
Lena	32.07	32.25	32.00	32.44	32.43	32.54
Barbara	30.71	29.54	29.41	30.00	29.92	30.18
Boat	29.90	29.97	29.91	30.21	30.17	30.23
Man	29.61	29.88	29.87	30.10	30.04	30.13
Couple	29.71	29.73	29.71	30.12	30.08	30.16

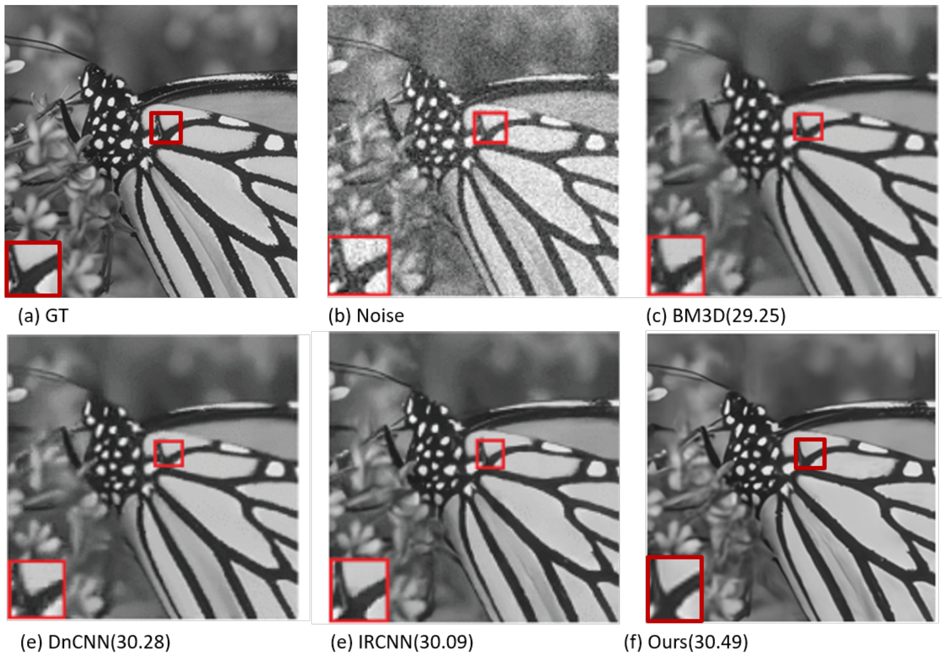


Fig. 12. The results of denoising by models (1)

Fig. 13 and Fig. 14 show the denoising results of one of the images from the BSD68 test dataset by models. After comparing 6 models with the our model, it can be confirmed that our model has the best PSNR values, which are 30.18/38.52, respectively.

5. Conclusions

Noise may be present during image generation, transmission, or processing, which can lead to performance degradation during image analysis. In addition, due to images containing noise, the performance of deep learning models will also decrease during deep learning. To improve these issues, this article proposed a Multi-scale Attention U-Net. The rapproach is to apply Multi-scale Attention unit to denoising network based on DnCNN and convert it into an Outburst structure for image processing. Attention Unit is a new modular approach that can suppress unwanted information and master the function of emphasizing only important information. By analyzing the internal relationships of a given input value, it is divided into two parts: the part that concentrates the important parts and the part that concentrates the important parts through spatial information. The two parts are executed in a parallel structure and then merged. And without adding too many parameters, under the action of Attention unit, more spatial feature maps were generated than other models, not only through PSNR SSIM. The improved performance was also confirmed by removing noisy images. In addition, as the feature part of the overall structure, the latter half of the neural network uses Xception's deformation of the Outburst structure to significantly increase feature information, endowing various feature maps with

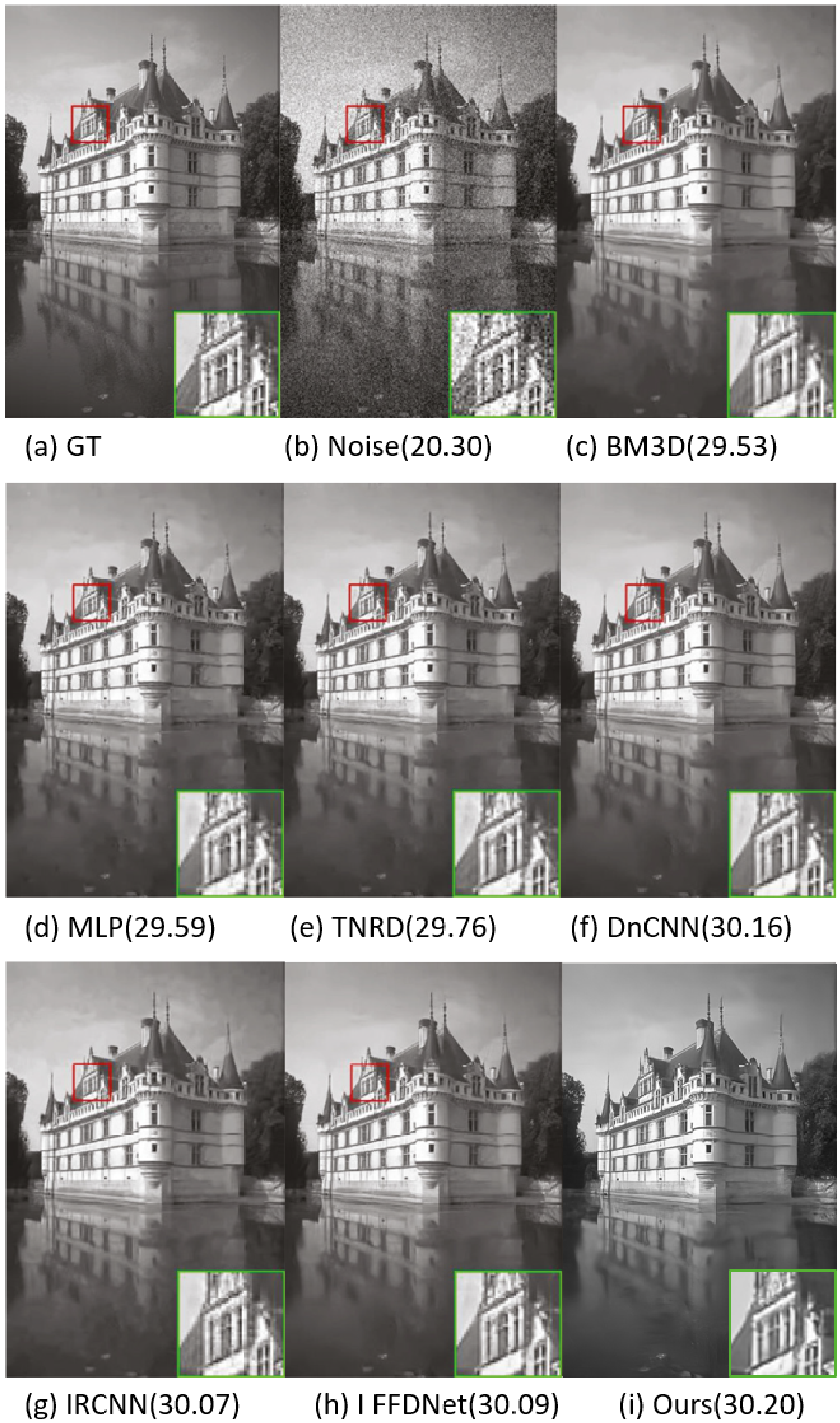


Fig. 13. The results of denoising by models (2)

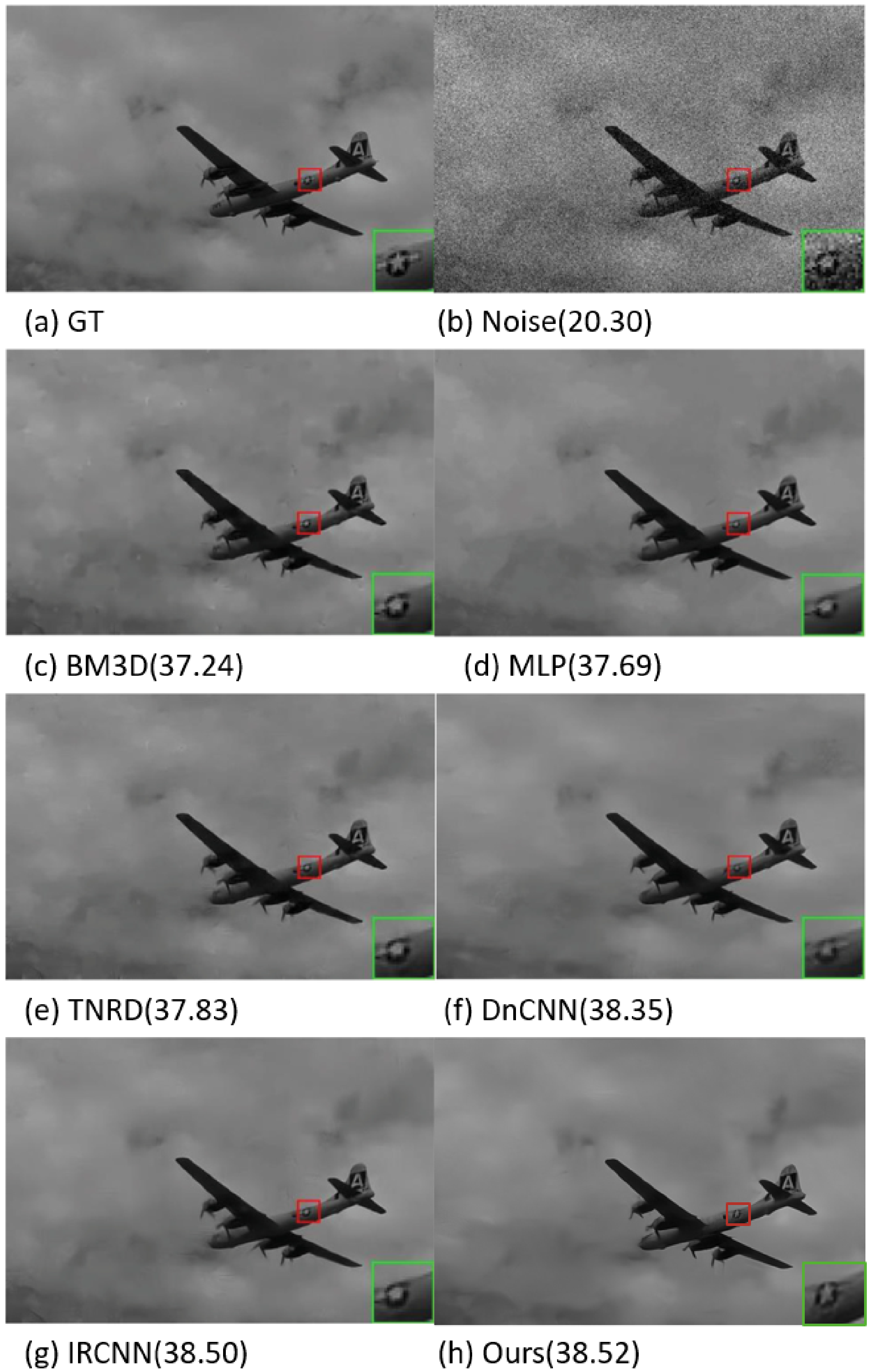


Fig. 14. The results of denoising by models (3)

scalability and demonstrating improvements in noise pattern analysis. The recommended method is to use Set12 and BSD68 test data, compare PSNR and SSIM values with existing models and early versions of the proposed model, and expand specific sections to make their differences more apparent. The experimental results showed that the overall denoising level improved the performance on average compared to other models, especially in terms of numerical results, which confirmed that the model proposed by SSIM for evaluating human visual image quality differences performed the best.

Acknowledgments. This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT)(No.RS-2023-00247045)

References

1. F. Emmert-Streib, Z. Yan, H. Feng, S. Tripathi, and M. Dehmer.: An Introductory Review of Deep Learning for Prediction Models With Big Data, Frontiers in Artificial Intelligence, Vol. 3, pp. 4, Feb. 2020.
2. Shanthi, T., and R. S. Sabeenian.: Modified Alexnet architecture for classification of diabetic retinopathy images, Computers and Electrical Engineering, Vol. 76, pp. 56-64, 2019.
3. Majib, Mohammad Shahjahan, et al.: Vgg-scnet: A vgg net-based deep learning framework for brain tumor detection on mri images." IEEE Access 9, pp. 116942-116952, 2021.
4. Tang, Pengjie, Hanli Wang, and Sam Kwong.: G-MS2F: GoogLeNet based multi-stage feature fusion of deep CNN for scene recognition, Neurocomputing, Vol. 225, pp. 188-197, 2017.
5. Wightman, Ross, Hugo Touvron, and Hervé Jégou.: Resnet strikes back: An improved training procedure in timm, arXiv preprint arXiv:2110.00476, 2021.
6. Arkin, Ershat, et al.: A survey: object detection methods from CNN to transformer, Multimedia Tools and Applications Vol. 82, No. 14, pp. 21353-21383, 2023.
7. C. Tian, Y. Xu, and W. Zuo.: Image denoising using deep CNN with batch renormalization, ISSN, Vol 121, pp. 461-473, Jan. 2020.
8. J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and T. Aila.: Noise2Noise: Learning Image Restoration without Clean Data, International Conference on Machine Learning, ICML, Vol. 7, No. 12, pp. 4620-4631, Jan. 2018.
9. Li, Zhuo, Hengyi Li, and Lin Meng.: Model compression for deep neural networks: A survey, Computers, Vol. 12, No. 3, pp. 60, 2023.
10. Yan, Puti, et al.: STDMANet: Spatio-temporal differential multiscale attention network for small moving infrared target detection, IEEE transactions on geoscience and remote sensing, Vol. 61, pp. 1-16.
11. Zhu, Junqi, et al.: Evaluation of deep coal and gas outburst based on RS-GA-BP, Natural Hazards, Vol. 115, No. 3, pp. 2531-2551, 2023.
12. Murali, Vineeth, and P. V. Sudeep.: Image denoising using DnCNN: An exploration study, Advances in Communication Systems and Networks: Select Proceedings of ComNet 2019, pp. 847-859, 2020.
13. Shafiq, Muhammad, and Zhaoquan Gu.: Deep residual learning for image recognition: A survey, Applied Sciences, Vol. 12, No. 18, pp. 8972, 2022.
14. Garbin, Christian, Xingquan Zhu, and Oge Marques.: Dropout vs. batch normalization: an empirical study of their impact to deep learning, Multimedia tools and applications, Vol. 79, No. 19, pp. 12777-12815, 2020.
15. danielseo, "[Computer Vision] DnCNN", gihyun.log, Jun 2021, https://velog.io/@danielseo/Computer-Vision-DnCNN.
16. Niu, Zhaoyang, Guoqiang Zhong, and Hui Yu.: A review on the attention mechanism of deep learning, Neurocomputing, Vol. 452, pp. 48-62, 2021.

17. Ren, Zhongzheng, et al.: Instance-aware, context-focused, and memory-efficient weakly supervised object detection, Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020.
18. Li, Xiaohui, Haiying Xia, and Lidan Lu.: ECA-CBAM: Classification of Diabetic Retinopathy: Classification of diabetic retinopathy by cross-combined attention mechanism, Proceedings of the 2022 6th international conference on innovation in artificial intelligence, 2022.
19. Siddique, Nahian, et al.: U-net and its variants for medical image segmentation: A review of theory and applications, IEEE Access, Vol. 9, pp. 82031-82057, 2021.
20. Wu, Jizhong, et al.: Fault detection based on fully convolutional networks (FCN), Journal of Marine Science and Engineering, Vol. 9, No. 3, pp. 259, 2021.
21. Yin, Haitao, and Siyuan Ma.: CSformer: Cross-scale features fusion based transformer for image denoising, IEEE Signal Processing Letters, Vol. 29, pp. 1809-1813, 2022.
22. Yahya, Ali Abdullah, et al.: BM3D image denoising algorithm based on an adaptive filtering, Multimedia Tools and Applications, Vol. 79, pp. 20391-20427, 2020.
23. Yu, Qiqiong, et al.: EPLL image denoising with multi-feature dictionaries, Digital Signal Processing, Vol. 137, pp. 104019, 2023.
24. Uetani, Hiroyuki, et al.: A preliminary study of deep learning-based reconstruction specialized for denoising in high-frequency domain: usefulness in high-resolution three-dimensional magnetic resonance cisternography of the cerebellopontine angle, Neuroradiology, Vol.63, pp. 63-71, 2021.
25. Averbuch, Amir, et al.: Cross-boosting of WNNM image denoising method by directional wavelet packets, arXiv preprint arXiv:2206.04431, 2022.
26. Jia, Nan, et al.: Background noise suppression using trainable nonlinear reaction diffusion assisted by robust principal component analysis." Exploration Geophysics, Vol. 51, No. 6, pp. 642-651, 2020.
27. Zou, Xueyan, et al.: Intelligent diagnosis method of bearing fault based on ICEEMDAN and Ghost-IRCNN, Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, Vol. 237, No. 13, pp. 3115-3130, 2023.
28. Cao, Like, Jie Ling, and Xiaohui Xiao.: Study on the influence of image noise on monocular feature-based visual slam based on ffdnet, Sensors, Vol. 20, No. 17, pp. 4922, 2020.
29. Qi, Huiqing, Shengli Tan, and Zhichao Li.: Anisotropic weighted total variation feature fusion network for remote sensing image denoising, Remote Sensing, Vol. 14, No. 24, pp. 6300, 2022.
30. Liu, Zhen, et al.: ADNet: Attention-guided deformable convolutional network for high dynamic range imaging." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021.

Mingshou An received his PhD in Electronics Engineering at Dong-A University. He is a Lecturer at the School of Computer Science and Engineering, Xi'an Technological University. His research interests include issues related to artificial intelligence, pattern recognition, deep learning, and industrial measurement. He is author of a great deal of research studies published at national and international journals, conference proceedings.

XuHang Zhao received his Master's in Computer Science and Engineering at Xi'an Technological University. His research interests include issues related to artificial intelligence, pattern recognition, deep learning, and industrial measurement. He is author of a great deal of research studies published at national and international journals, conference proceedings.

Hye-Youn Lim received her PhD in Electronics Engineering at Dong-A University. She is a Lecturer at the Electronics Engineering, Dong-A University. Her research interests include issues related to artificial intelligence, pattern recognition, and deep learning. She is author of a great deal of research studies published at national and international journals, conference proceedings.

Dae-Seong Kang received his PhD in Electrical Engineering at Texas A&M University. He is a Professor at the Electronics Engineering, Dong-A University. His research interests include issues related to artificial intelligence, pattern recognition, and deep learning. He is author of a great deal of research studies published at national and international journals, conference proceedings.

Received: August 10, 2024; Accepted: July 15, 2025.