IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 30, 2020, accepted July 14, 2020, date of publication July 23, 2020, date of current version August 19, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3011580

Residual Forward-Subtracted U-Shaped Network for Dynamic and Static Image Restoration

HO MIN JUNG[1], BYEONG HAK KIM[2], AND MIN YOUNG KIM[1,3], (Member, IEEE)
[1]School of Electronics Engineering, Kyungpook National University, Daegu 41566, South Korea
[2]Hanwha Systems Company, Gumi 39376, South Korea
[3]Research Center for Neurosurgical Robotic System, Kyungpook National University, Daegu 41566, South Korea

Corresponding author: Min Young Kim (minykim@knu.ac.kr)

This work was supported in part by the Brain Korea 21 (BK21) Plus Project funded by the Ministry of Education, South Korea, under Grant 21A20131600011, in part by the Institute for Information and Communications Technology Promotion (IITP) funded by the Korea Government Ministry of Science and ICT (MSIT) through the Development of Intelligent Interaction Technology Based on Context Awareness and Human Intention Understanding under Grant 2016-0-00564, and in part by the Korea Institute for Advancement of Technology (KIAT) funded by the Korea Government Ministry of Trade, Industry and Energy (MOTIE) through the Multichannel Telecommunications Control Unit and Associated Software under Grant P0000535.

ABSTRACT Advanced image sensors with high resolution are now being developed for specially purposed electro-optical systems, with research focused on robust image quality performance in terms of super resolution and noise removal under various environmental conditions. Recently, machine-learning and deep-learning methods have been studied as the best practical techniques for restoration to improve the deteriorated image quality of sensors. However, these methods show limitations and side effects of image degradation such as image non-uniformity. In this paper, we analyze and randomly generate additive white Gaussian noise, non-uniform line noise, and dark saturation as representative image degradations. We then propose an advanced U-net model based on global and local residual learning in order to restore complexly deteriorated images. The proposed method shows unparalleled performance compared to alternative models and previous studies. In particular, various complex noise components are minimized and improved with equal quality so that variation between sequential images is minimized. These findings leverage mutual corroboration of quantitative and qualitative evaluation metrics. In the future, the proposed model is expected to contribute to a wide range of field applications such as defense, surveillance, and video media for image quality enhancement technologies.

INDEX TERMS Restoration, multi-type noises, image denoising, image enhancement, convolutional neural network, residual learning.

I. INTRODUCTION

As image system technology advances, image restoration for high-resolution and special-purpose images is among the key technologies for application in fields such as multimedia, intelligent vehicles, defense, surveillance, and reconnaissance. Image restoration comprises the task of restoring image damage such as noise, motion blur, focus error, excessive light, and insufficient light. The success of this task is determined by how completely and efficiently the clean image (prediction image or restored image) is restored from the degraded image, in comparison to the ground truth (GT) image. Denoising, one of the most substantial tasks in image restoration, is the task of finding noise components in the input component and restoring clean images that come as close to the original as possible. As an example of such work, Fig. 1 shows (a) the noise component corresponding to Fig. 1 (b), which is analyzed in the generation of a clean image as shown in Fig. 1 (c).

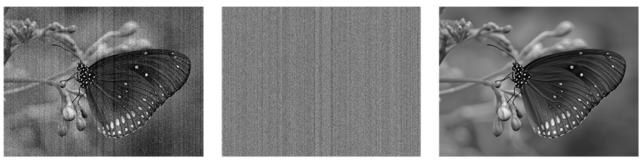

FIGURE 1. Example of correlation between noise image and GT image. (a) noise image, (b) noise component, (c) GT image.

In order to perform the image denoising task, it is necessary to analyze the causes and types of image degradation. A classification of these types is as follows:

The associate editor coordinating the review of this manuscript and approving it for publication was Qiangqiang Yuan.

1) Signal noise of photon detector and non-uniformity from dead or bad pixels of image sensors
2) Vertical line noise from a sensor that realizes 2D images using vertical scanning by a high-performance 1D sensor (line scan camera)
3) Dynamic noise with characteristics that change in units of frames in successive images
4) Black and white low dynamic images due to lack of sensor signal strength or environmental influence

As examples of these types, Fig. 2 shows image degradation due to complex noise in a thermal infrared (TIR) image or a 1D scan-based hyper-spectral (HS) image of a special design structure [1]. Image deterioration due to image degradation and noise can be inconvenient for users and causes the loss of important image information.

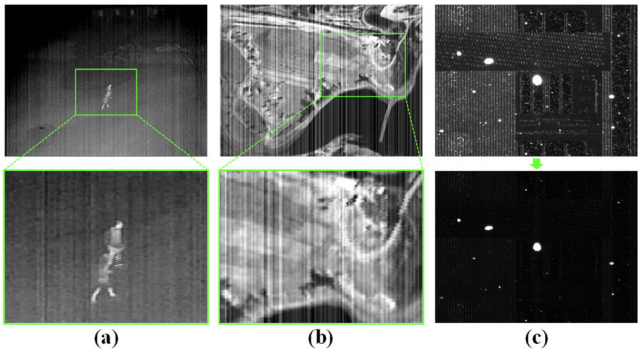

FIGURE 2. Example of noise and saturation image. (a) line and white gaussian noise in infrared image, (b) line noise in hyper-spectral image, (c) dark saturation image in Automated Optical Inspection (AOI) system.

In particular, techniques used in application systems such as object detection, tracking, and segmentation can show various performance limitations when using deteriorated images. The background registration-based adaptive noise filtering (BRANF) algorithm has been studied to solve both static and dynamic noise issues, And also general machine learning-based algorithms and the cross fusion-based adaptive contrast (CFACE) have been studied to solve these problems, but they show limitations in effectively removing noise with high uniformity and speed [2]–[5].

Recent advances in convolutional neural networks (CNNs), software, and hardware have shown that deep learning [6]-based algorithms (VDSR [7], DnCNN [8], FFDNet [9], MWCNN [10], Noise2Noise [11]) rank high in state-of-the-art (SOTA)-based paperwithcode.[1] However, side effects of these techniques include image distortion, and limitations are found in the removal of complex noise components, as well as in overall speed. In the present paper, we propose the performance of image improvement by constructing a network based on global residual learning (GRL) and local residual learning (LRL).

Also, as mentioned in MWCNN [10], In this paper, we consider the correlation between computing cost and performance caused by a sufficient acceptance field. As the receptive field grows, the consumption of computing resources increases, leading to a trade-off between efficiency and performance. However, we also argue that a sufficient receptive field is helpful for image reconstruction, since the noise is unpredictable and appears throughout the range. In addition, we establish a receptive field of sufficient size (192×192 pixels) in the trade-off process of performance and resources and consider this a problem to be solved with the development of software and hardware. Although there is a risk of information loss, the pooling layer is used in this paper to effectively cope with the distribution of noise according to the image scale. As a result, our method shows excellent noise removal and low variation in performance compared to other algorithms.

[1] [Online]. Available: http://paperswithcode.com/sota

II. RELATED WORK

The task of image denoising is representative works in the field of image restoration research. As a traditional method for removing noise, basic image processing algorithms such as average filter, median filter, and Gaussian filter have been studied [12]. Image restoration has also been developed into machine learning-based algorithms such as Block-Matching and 3D filtering (BM3D) [3] and Weighted Nuclear Norm Minimization (WNNM) [4]. However, BM3D [3] and WNNM [4] are very slow in real-time application and show poor image restoration performance. Recently, with advances in software and hardware, CNN-based algorithms (VDSR [7], DnCNN [8], FFDNet [9], MWCNN [10], Noise2noise [11], etc.) are dramatically developed and shown outstanding performance and high speed. Among them, Multi-level Wavelet-CNN (MWCNN) [10], one of the best-performing algorithms, applies discrete wavelet transform (DWT) and inverse wavelet transform (IWT) operations between networks based on enlarged receptive field. Complementing the deep learning method that produces somewhat blurry results, it shows excellent resilience by utilizing textural detail and sharp structures. In addition, most existing denoising tasks require both a noise image and a clean image as supervised learning methods, but in Noise2Noise [11], restoration was performed with only noisy images without clean data, suggesting a new paradigm for denoising tasks. Also, in this task, research on video denoising is actively underway based on the Recurrent Neural Network (RNN) algorithm [13].

However, most denoising algorithms aim to remove AWGN, one of the noise types, and show poor performance when applied to the complex and various types of noise generated in software and hardware for a wide range of multimedia applications. Research is also being conducted using a scene-based non-uniform correction (SBNUC) method [1] to remove line noise that is often observed in HS systems. Also, several CNN-based methods have been studied to solve these problems [14], [15], [16]. Among them, the two-stream wavelet enhanced U-net (TSWEU) [16] method analyzed noises of various line patterns and has shown observable performance.

Image enhancement, one of the other image restoration fields, is a technique that corrects values when an undesirable brightness value is obtained from an image due to errors or malfunctions such as lighting, exposure time, aperture value, sensor, etc. For example, in Fig. 2 (c), the defect is not correctly found due to the abnormal operation of the lighting, and Fig. 3 (a) shows that detection is not performed properly in the dark image. Fig. 3 (b) shows successful detection after yolo-v3 tiny [17] restoration as part of the method proposed below.

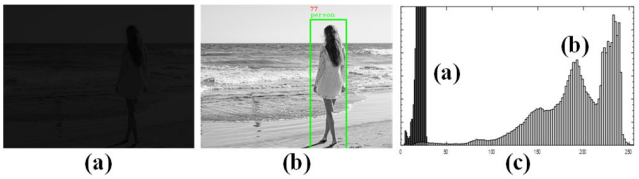

FIGURE 3. Example of image restoration and enhanced detection performance. (a) sample of dark saturation image, (b) enhanced image and result applied detection algorithm (yolo-v3 tiny model), (c) histograms of pixel brightness for (a) and (b).

In order to resolve these issues, one of the most common and popular methods is general histogram equalization (GHE). However, in images with dynamic brightness, GHE shows poor reconstruction performance when comparing the restored image to the ground truth. Engineers and researchers have attempted to solve this problem using vision-based algorithms such as contrast-limited adaptive histogram equalization (CLAHE) as well as machine learning-based algorithms (SVD-DWT [18], AGCWD [19], CegaHe [20], etc.). Also, under study is restoration using the generative adversarial network (GAN) algorithm [21], [22] using a deep learning technique. In the future, this technique may be applied to technologies such as high dynamic range (HDR).

Yet another restoration task is the improvement of old or degraded images and videos through technologies such as Single-Image Super Resolution, JPEG artifact removal, deblur, and defocus [23]–[26]. Research is actively underway for this task within image restoration. Image restoration tasks such as image denoising and enhancement are essential for viewing outdated or deteriorated images or videos, as well as for other applications. Reflecting this, ETH Zurich's computer vision laboratory in Switzerland has been leading the field of image restoration by holding the NTIRE (New Trends in Image Restoration and Enhancement workshop[2]) challenge every year since 2016.

In the restoration task proposed in this paper, practicality is considered for application in various applications. Unlike the conventional method of removing only AWGN (a single noise type), we propose a method to simultaneously correct multi-type noise based on Global Residual Learning (GRL) and Local Residual Learning (LRL). GRL is a process of subtracting noise features and inputs, which are the final output of the network, and LRL is a process of subtracting for each scale of images (max-pooling, up-sampling). Designed to reflect the characteristics of noise, these techniques are expected to improve network performance and will lead to advanced image reconstruction. Furthermore, this model is a general restoration model that can show excellent performance by applying the re-learning method as a dataset for improving contrast.

[2][Online]. Available: http://www.vision.ee.ethz.ch/ntire19/

III. METHOD

A. DATA GENERATION AND AUGMENTATION

This section describes data generation and augmentation for network learning as shown in the flowchart in Fig. 4 (a).

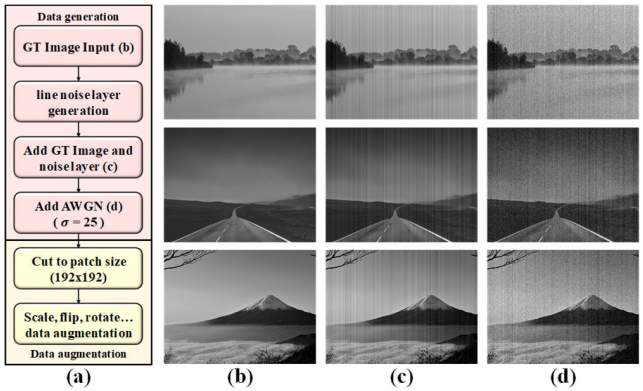

FIGURE 4. Data generation and augmentation. (a) algorithm flowchart, (b) GT image, (c) line noise image, (d) line + AWGN image.

Since the existing single-type noise imag3e denoising task generates noise based on the sigma level in the input image, it is limited to various types of noise generated in the application system. We focus on multi-type noise image denoising, which removes complex noise by generating line noise and AWGN, which is one of the noise types that appear frequently in IR and HS systems.

However, there are limitations in obtaining a real dataset from non-uniform noise observed in IR and HS systems. In particular, in the IR system, there are functions such as fixed pattern noise (FPN) correction for removing uniform noise, but it is difficult to completely remove non-uniform noise [27]. Therefore, it is not easy to build a dataset because of the characteristic of CNN, where input (Noise image) and label (GT image) are simultaneously needed. In order to overcome the limitations, this study analyzes the non-uniformity of real IR systems and generates at an equivalent level of real IR noises.

The dataset used for training and validation is the DIV2K [28] dataset (training images: 800, validation images: 100). Using this dataset, noise components are generated as similar as possible, as shown in Fig. 5.

First, white and black layers (WL, BL) of the same size as the input in the prepared dataset were randomly generated as

$$WL, BL = \sum_{n=1}^{NL} round\left\{\frac{(rand\ (1) \times 255)}{Int}\right\} \quad (1)$$

FIGURE 5. An example of noise that is generated at real IR systems. (a) an example of real IR systems, (b) a feature of real noise component, (c) a feature of generated noise component.

where NL is number of lines, Int is the intensity of the generated lines, and *rand* $(*)$ is the function that randomly generates $*$ values from 0 to 1.

Next, a Gaussian filter, one of the smoothing operations, was applied to create a line noise layer that looked as equally as possible as follows:

$$WL_\sigma, BL_\sigma(p_i, p_j) = \frac{1}{2\pi\sigma^2} e^{-\left(\frac{p_i^2+p_j^2}{2\sigma^2}\right)} \tag{2}$$

where σ is a parameter for determining blurring and is called a scale parameter, and p_i, p_j are the width and height of each pixel.

Adding each dynamic noise layer to the input image produced a GT + line noise image as shown in Fig. 4 (c). Finally, by adding AWGN ($\sigma = 25$), a dataset with multi-type noise was created as shown in Fig. 4 (d).

Then, in the augmentation process, the image was divided into patches, and rotation (180°) and flip (up/down, left/right) were randomly performed based on the scale change ($\times 0.6 \sim 1.2$). The key point of this process was to focus on the scale change by reflecting the noise characteristics appearing randomly overall, and to exclude the 90° and 270° reorientations during rotation because the purpose is to remove the vertical line noise. Finally, in this process, a dataset for network learning was generated by augmenting noise characteristics equivalent to those of reality.

B. NETWORK ARCHITECTURE DESIGN

This section discusses the overall network architecture design. We explain why we designed the network architecture and why we decided upon this particular architecture.

1) NETWORK ARCHITECTURE

The correlation between GT image (x), noise image (y), and noise ingredient (n) can generally be defined as $x = y - n$. In general, most algorithms predict the x component based on y. In our network, however, the clean image (x) is restored by predicting the noise feature (n) component from the residual learning perspective. In fact, existing residual learning has been introduced in that deeper networks are likely to run into gradient vanishing or exploding problems, and deep learning cannot be performed well, for example increased training error. Reference [29]. However, referring to the method used in DnCNN [8] in an attempt to approach from a different perspective than the existing residual learning, we applied the new residual learning method which is modified overall to the network. This residual learning method removes the clean image components from the noisy image and then calculates the residual features (n) to finally derive them as

$$\hat{x} = y - R(y) \tag{3}$$

where y is the input (noise image), $R(y)$ is the residual feature through the network, and $\hat{x}$ is the final prediction image (clean image).

The biggest difference from previous research is the application of Local Residual Learning (LRL) and Global Residual Learning (GRL) in the U-shaped [30] network structure. Residual operations are applied to both global and local parts to calculate the residual features effectively. In this way, the noise image is ultimately restored to a clean image (prediction image).

First, in this network, LRL is a residual operation applied to each layer immediately after the max-pooling or interpolation-based resize up-sampling. This operation is added to correspond to the scale of the input before obtaining the final feature. In fact, pooling operations tend to lose information, so researchers and engineers prefer not to use them. However, we added this process to deal with noise of various distributions at various scales. As will be mentioned again in the next section, checkerboard artifacts that occur when using the transposed convolution operation degrade the image restoration. Therefore, to prevent the artifact, we use an interpolation-based up-sampling operation instead. As shown in Fig. 6 (b), LRL is a combination of convolution layer + batch normalization [31] + ReLU [32], and four sets are connected to each group. Finally, the residual operation is performed through a subtract operation in groups 1 and 4.

GRL is the network output, and the final clean image is obtained by calculating the difference between the final residual image and the input. The distinguished two residual learnings (GRL, LRL) are the key point of this network, since noise is additional information not wanted by users, and eventually added unwanted features coming out of the network can be removed by difference calculations. The overall network is designed as shown in Fig. 6 (a).

For the validation and analysis of the effectiveness of residual learning on the network, we have confirmed through the ablation study. In the ablation study, quantitative and qualitative evaluations were performed. The equations used for quantitative evaluation are peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) [33].

$$\text{PSNR}\,(I_x, I_{\hat{x}}) = 10 \times \log\left(\frac{I_{max}^2}{\text{MSE}\,(I_x, I_{\hat{x}})}\right) \tag{4}$$

where I_x is the GT image, and $I_{\hat{x}}$ is the predicted image. I_{max} is the max dynamic range of input and output images. I_{max} is 255 for 8-bit image. The MSE is the mean square error of the output image in comparison with the original image.

$$\text{SSIM}\,(I_x, I_{\hat{x}}) = \frac{2\mu_x\mu_{\hat{x}} + C_1}{\mu_x^2 + \mu_{\hat{x}}^2 + C_1} \cdot \frac{2\sigma_{x\hat{x}} + C_2}{\sigma_x^2 + \sigma_{\hat{x}}^2 + C_2} \tag{5}$$

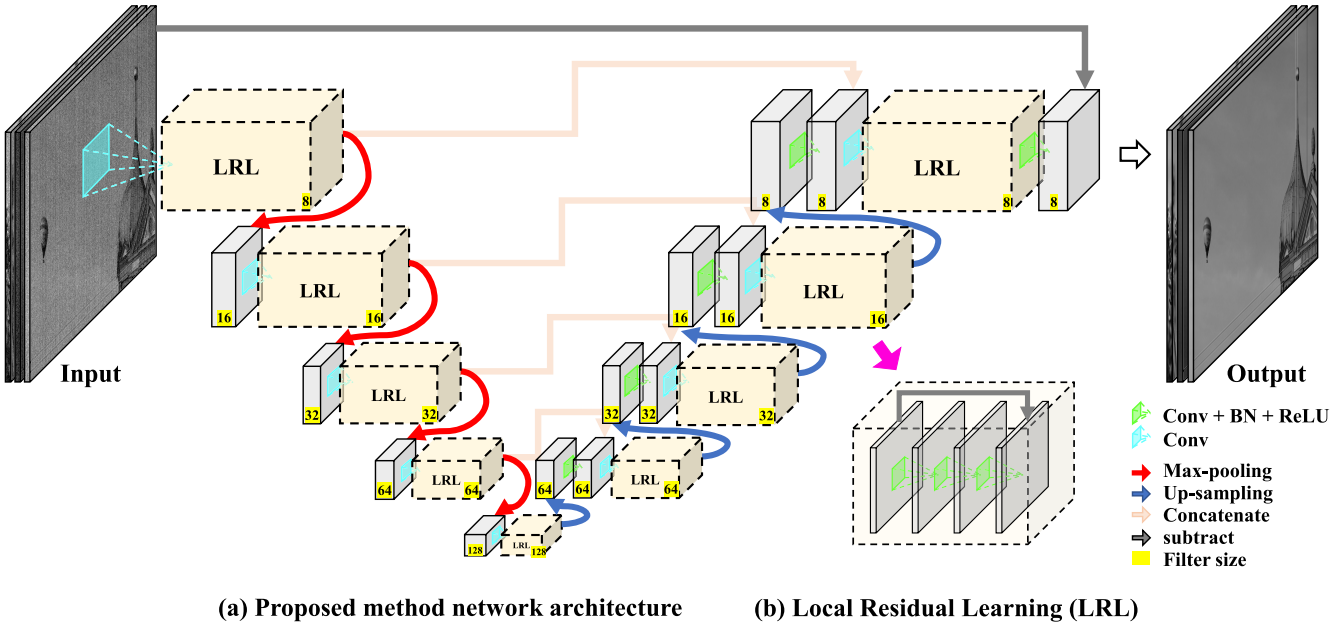

FIGURE 6. Proposed network architecture. This network consists of (a) and (b). the black line (subtraction operation) corresponds to GRL and LRL.

where μ_x and $\mu_{\hat{x}}$ are the respective average values of input and output images. σ_x and $\sigma_{\hat{x}}$ are the respective variances of input and output images. $C_1 = k_1 I_{max}^2$, and $C_2 = k_2 I_{max}^2$. $k_1 = 0.01, k_2 = 0.03$.

As shown in Fig. 7, in the validation process of training, both methods are applied to LRL and GRL, that are shown the improved performance compare to the method no applied residual learning.

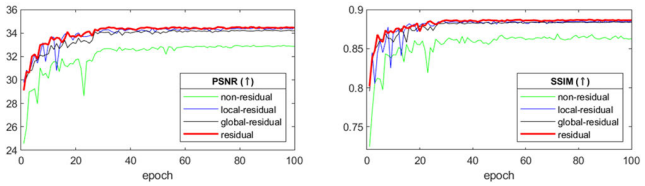

FIGURE 7. The ablation study with PSNR,SSIM results during validation of training.

Also, in the inference, as shown in Table 1, it is analyzed that the quantitative evaluation showed a performance difference.

TABLE 1. The comparisons of the ablation experiments.

method	SSIM (↑)	PSNR (↑)
Non-residual learning	0.9169	34.08
Only applied LRL	0.9282	35.01
Only applied GRL	0.9302	34.90
All applied	**0.9309**	**35.14**

In addition, quantitative evaluation was performed based on the restored image. Most of the participants in the experiment responded that the results were good when they used LRL and GRL together.

2) INTERPOLATION-BASED UP-SAMPLING

The commonality between the transposed convolution (deconvolution) operation and the up-sampling operation is mainly used to enlarge the image reduced by operations such as pooling. The difference between the two operations (based on the Keras[3] API) is that the up-sampling is an interpolation-based image resize, while the transposed convolution operation enlarges the size of the image based on the learned filters. In this process, as mentioned in Distil's blog [34], when using transposed convolution operation, uneven overlap may occur depending on filter size and stride, causing checkboard artifacts.

Similarly, artifacts such as those in Fig. 8 were also often observed in this study when transposed convolution is used. Therefore, we modified the network with up-scaling in order to reduce the effects of incorrect restoration. The number of

[3][Online]. Available: https://keras.io

FIGURE 8. Comparison using interpolation-based resizing with transposed convolution. When using up-sampling (c, d), these phenomena disappeared (a, b).

hyper-parameters in the network is slightly increased, but this is not significantly affecting the performance.

C. NETWORK TRAINING

For learning we used the ADAM [35] algorithm for the optimizer (learning rate = 0.001, $\beta_1 = 0.9$, $\beta_2 = 0.999$, epsilon = None), and a mini batch randomly shuffled based on the maximum size that the graphics processing unit could accommodate (batch size: 128). In order to select the loss function, we referred to [36], which analyzed the field of image restoration in detail, and [37], which analyzed which loss function should be used in the image restoration task. Through various experiments, we selected the loss function (L) according to the mean absolute error (MAE):

$$L(h) = \frac{1}{N}\sum_{i=1}^{N} |R(y_i; h) - (y_i - x_i)| \tag{6}$$

where h is the network parameter learned through this proposed network, and N is the number of pairs of clean-noise images (x, y) during training (patch). Experiments were selected by comparison with a total of two control groups. The first control was a mean square error (MSE)-based loss function:

$$L(h) = \frac{1}{2N}\sum_{i=1}^{N} ||R(y_i; h) - (y_i - x_i)||^2 \tag{7}$$

The second control was the loss function of the combination of MAE and SSIM [33]. This expression can be redefined as in (8), reflecting the characteristics of the CNN:

$$L^{\text{SSIM}} = 1 - \text{SSIM}(I_x, I_{\hat{x}}) \tag{8}$$

Finally, this control is combined as

$$L^{Mix} = \alpha \cdot L^{MAE} + \beta \cdot L^{SSIM} \tag{9}$$

where α, β are hyperparameters for the ratio of loss function.

As shown in Fig. 9, and in Table 2, control groups showed good performance in learning when the loss function was used.

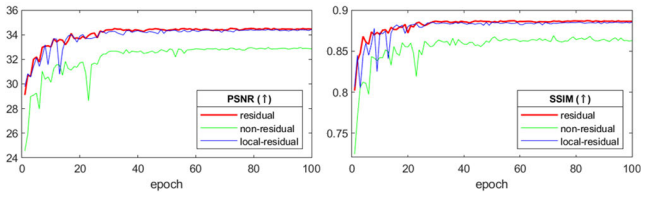

FIGURE 9. The comparison of PSNR,SSIM results of loss functions during validation of training.

TABLE 2. Quantitative comparison of loss functions.

Loss function	MAE (↓)	SSIM (↑)	PSNR (↑)
Mix (MAE+SSIM ratio)	3.07	2.25	35.09
MSE	3.13	2.36	35.04
MAE	**3.04**	**2.21**	**35.14**

However, in the evaluation, control groups showed better performance quantitatively and qualitatively when MAE was used. We confirmed that control groups show outstanding performance throughout the experiment. In fact, we found it best to combine the necessary loss functions depending on the task.

In addition, loss of information after learning may result from the relatively high depth and relatively large receptive field (192 × 192) of this network. Therefore, we verified the image that activated after passing the convolution layer of each layer. From this, we could see that the network was working correctly as shown in Fig. 10.

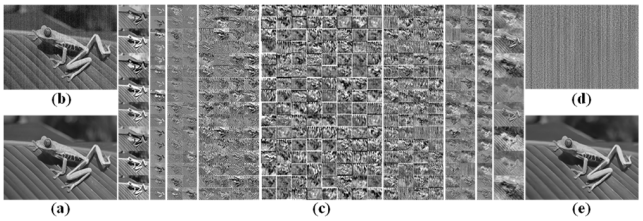

FIGURE 10. Activated Network layer. (a) GT image, (b) Noise image, (c) activated images of convolutional layer, (d) residual image (predicted image), (e) output image (clean image). The noise image shows that the noise component has been calculated through the network.

The software used in this paper was Keras (2.2.4, backend TensorFlow[4]: 1.13) based on cuda 10 and cudnn 7.5. The server computer specification used for network learning consisted of an Intel (R) Core (TM) i9-7900X with a 3.3 GHz CPU, 64 GB of RAM, and NVIDIA Titan V 2way system. The evaluated computer specification was 99900KF with a 3.6 GHz CPU, 32 GB of RAM, and NVIDIA GeForce RTX 2080Ti. Due to sufficient hardware specifications, we quickly performed large batches of learning, inference, and evaluation.

IV. EXPERIMENTS

A. DATASET FOR EVALUATION

For the evaluation of the multi-type noise image denoising task, we used 100 randomly selected videos from pixabay.com,[5] a website that provides copyright-free photos and videos. This dataset was regenerated into an evaluation dataset using the data generation process of section III.A. These datasets were edited to 30 fps, up to 10 seconds long, because of image frame length deviation. The reason why video is used for multi-type noise image denoising task evaluation is that video is composed of frames. Based on this factor, it serves to evaluate how effectively different dynamic noise is removed for each frame. In addition, for evaluation of σ the image enhancement task, 155 images randomly selected from pixabay.com were darkened, and evaluation was performed on a total of 2635 images.

[4][Online]. Available: https://www.tensorflow.org

[5][Online]. Available: https://pixabay.com

TABLE 3. Quantitative comparison of algorithmic performance on multi-type noise image denoising task.

100 videos in total, Noise type: AWGN (Noise level σ: 25) + random line noise (white & black)						
Algorithm	MAE (↓)	Variation (↓)	RMSE (↓)	MS-SSIM (↑)	PSNR (↑)	SSIM (↑)
BM3D *	4.73	3.14	6.81	0.9380	31.82	0.8915
WNNM *	5.22	2.81	8.25	0.9288	30.44	0.8683
VDSR	3.40	2.87	5.06	0.9571	34.41	0.9181
DnCNN	3.41	2.98	5.02	0.9591	34.46	0.9211
FFDNet	3.37	2.59	5.11	0.9578	34.37	0.9232
MWCNN	3.06	2.27	**4.65**	0.9709	**35.19**	**0.9334**
Noise2Noise	3.26	2.56	4.86	0.9682	34.78	0.9328
RFSUNET	**3.04**	**2.21**	4.69	**0.9717**	35.15	0.9309

(*) is none training algorithm (ML-based). Proposed method shows that var, MAE, and MS-SSIM values are good, and RMSE, PSNR, and SSIM values are relatively lower than other algorithms. Observing Fig. 10 and Table II together, which show that the relation between image restoration quality and numerical improvement is relatively small.

The degraded dataset for evaluation was created dynamically for each frame.

B. QUANTITATIVE MEASUREMENT

Machine learning-based algorithms such as BM3D [3] and WNNM [4] did not need to be trained to evaluate multi-type noise image denoising tasks, but deep learning-based algorithms needed to be trained anew. We proceeded implementation of training each algorithm based on the paper in which it was referenced. Based on the same dataset (DIV2K [28] + random multi-type noise in each frame), training and validation were conducted in the same way as for the proposed method. After applying the algorithm to the dataset, evaluation proceeded, using a total of six formulas as quantitative indicators. The indicators were MAE, frame-to-frame variation, root-mean-square error (RMSE), PSNR, SSIM [33], and multi-scale SSIM (MS-SSIM) [38], as respectively quantified in (10), (11), (12), (13), (4), (5), and (14):

$$\mathrm{MAE}\,(p) = \frac{1}{N}\sum\nolimits_{p\in P} |x(p) - \hat{x}(p)| \tag{10}$$

where x is the GT image, and $\hat{x}$ is the clean image (prediction image or restored image), and p is the pixel value.

$$\mathrm{AE}\,(p) = |x(p) - \hat{x}(p)| \tag{11}$$

$$\mathrm{Variation} = \frac{1}{N}\sum_{p\in P} AE(p)_{f_n} - AE(p)_{f_n-1} \tag{12}$$

where AE is absolute error between pixels, and f_n is the n^{th} frame; variation is calculated as the difference between the $(n-1)^{\mathrm{th}}$ frame and the n^{th} frame.

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{p\in P}(x\,(p) - \hat{x}(p))^2} \tag{13}$$

where RMSE is an operation that takes the square root of the mean square error (MSE).

$$\mathrm{MS-SSIM}\,(p) = l_M^{\alpha}(p)\prod_{j=1}^{M} cs_j^{\beta_j}(p) \tag{14}$$

where MS-SSIM [37] is based on (5). It calculates the SSIM score at multi-scale (M) and then weights it to obtain a final score. l is luminance, c is contrast, and s is structure. These parameters are the basic building blocks of SSIM [33]. For convenience, we used the default values as $A = B = 1$, for $j = \{1, \ldots, M\}$.

These formulas were used to evaluate how properly the noise was removed in successive frames and how completely the image was restored in comparison with the GT image. The supplementary data include some videos and graphs with quantitative indicator results, presented frame by frame.

Table 3 lists the average values of quantitative indicators for 103 videos. In terms of numerical values, the RMSE, PSNR, and SSIM [33] methods showed lower (worse) values than MWCNN [10] and Noise2Noise [11], while MAE, variation, and MS-SSIM performed well. In some images, the features in the video were slightly blurry, but in terms of video footage or frame-by-frame, AWGN and line noise were clearly removed, demonstrating excellent image restoration quality. As shown in Fig. 11, three algorithms showed high performance: MWCNN [10], Noise2Noise [11], and the proposed method.

We observed that networks with relatively large receptive fields showed excellent image restoration performance. There were distortions in the videos when algorithms other than the proposed method were applied; for these same algorithms, line noise was not properly removed, remaining faint. Also, as shown in Fig. 12, it is clearly confirmed that non-uniform noises when the histogram equalization is applied to the restored image.

As mentioned in the SRGAN [26] paper, the high values for PSNR and SSIM do not necessarily mean that the video is visually appealing or that the noise is removed well; rather, they reflect the characteristics of deep learning optimization based on the loss function. As shown in (4), the use of MSE loss would produce high values in the evaluation metrics (PSNR), but at the potential cost of blurry output or improper noise removal. Therefore, in accordance with [39], which analyzes the distortion measure and the human perceptual measure, we also adopted the mean opinion score (MOS) method and used it as a quantitative indicator.

Table 4 provides a quantitative average index of the image enhancement task based on 2635 images (dark degraded

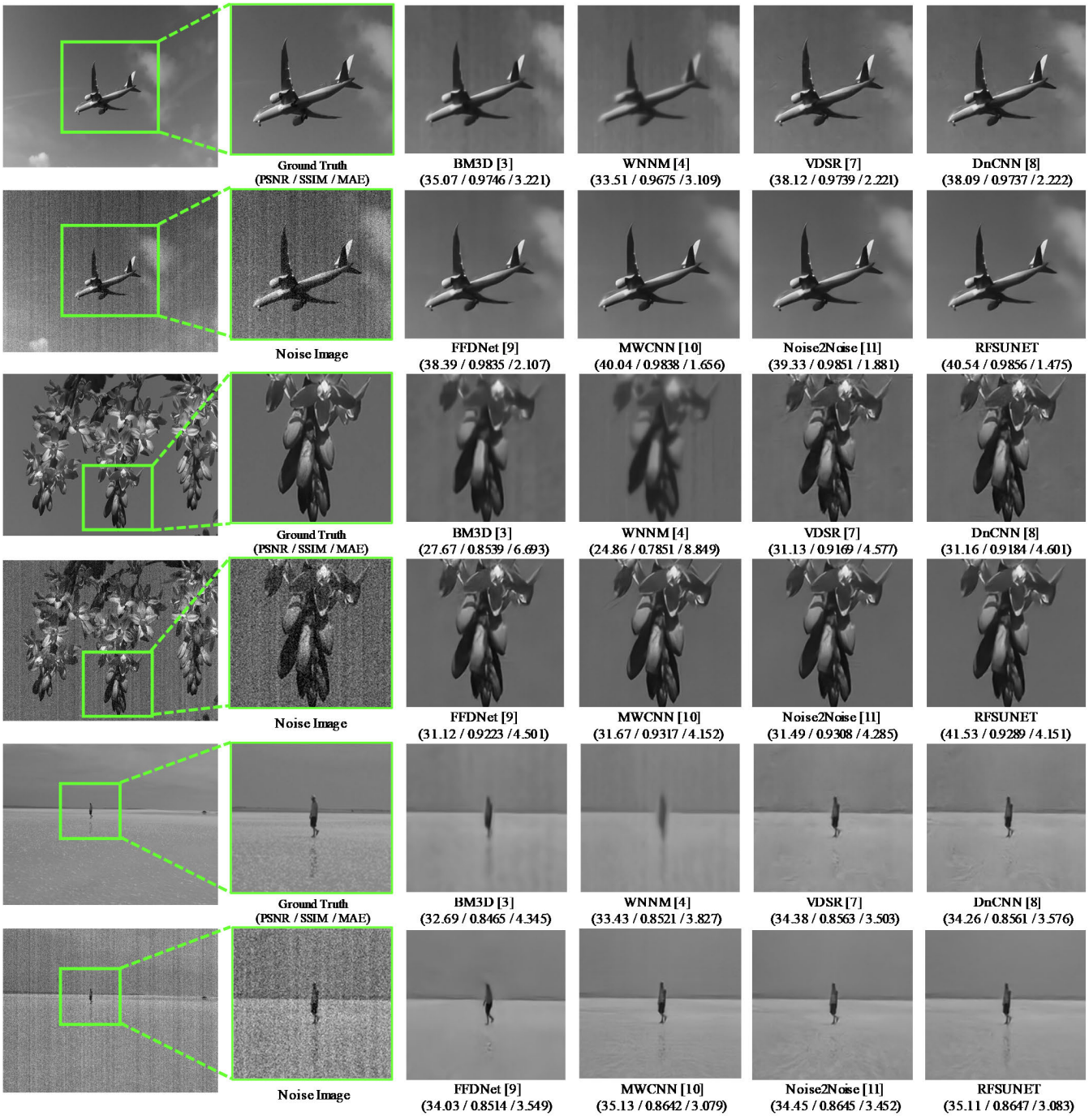

FIGURE 11. Quantitative evaluation of multi-type noise denoising task by various algorithms (PSNR, SSIM, MAE). In the proposed method, multi-type noises are effectively removed. However, from a numerical point of view, it can be observed that there is a difference in image restoration quality.

images). Indicators used for quantitative evaluation were MAE, PSNR, SSIM [33], MS-SSIM [38], and additionally Natural Image Quality Evaluator (NIQE) [40], which does not require a reference (GT image) to evaluate image quality.

As shown in Fig. 13, three algorithms performed properly: AGCWD [19], CegaHe [20], and the proposed method. Particularly numerically, the proposed method showed the image restoration closest to GT images by improving dark saturation. However, among the algorithms with good performance (AGCWD [19], CegaHe [20]), there were cases in which the quantitative evaluation was bad even though the quality of the restored image was good. This suggests that evaluation using qualitative indicators is necessary.

C. QUALITATIVE MEASUREMENT

As mentioned above, we conducted the MOS test because there appeared to be a limit to the usefulness of quantitative indicators in the evaluation of algorithm performance. For MOS testing, we asked a human evaluator to assign to

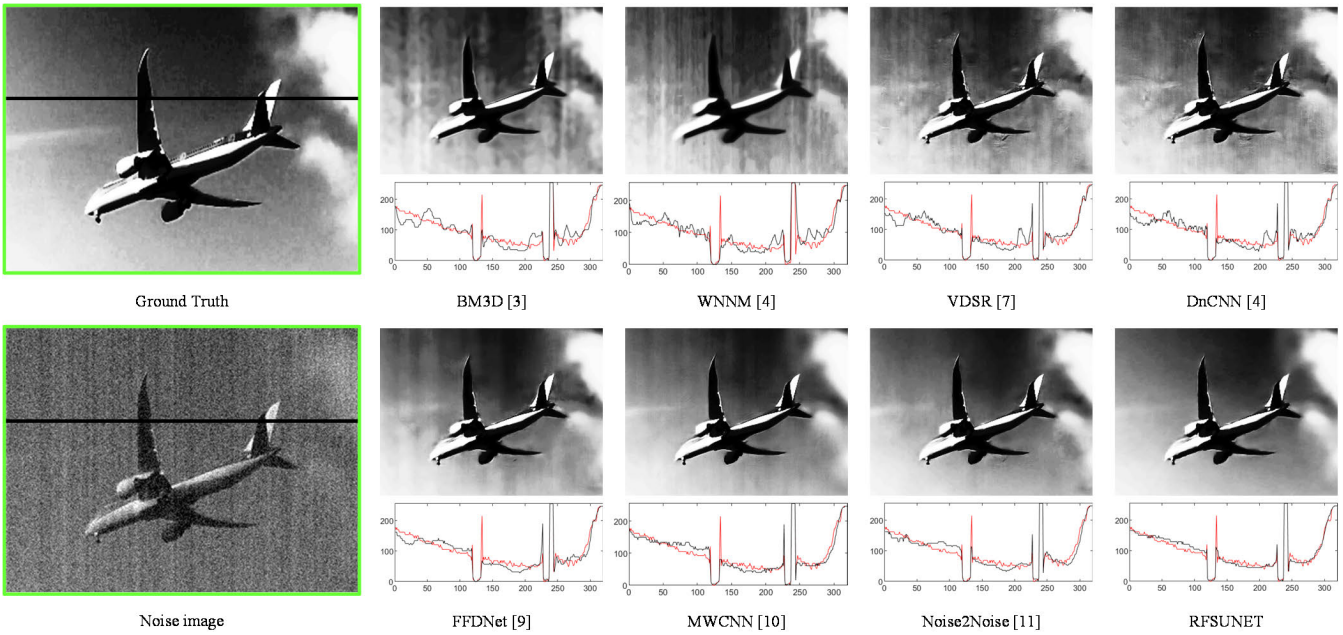

FIGURE 12. The result of applying histogram equalization to the restored image. It can be confirmed that it is incorrectly restored, or noises remain. Also, looking at the line profile (GT: black, each algorithm result: red line) at the bottom of each figure, it can be confirmed that the proposed method is most similar to the graph of GT, and that the fluctuation range of values is small.

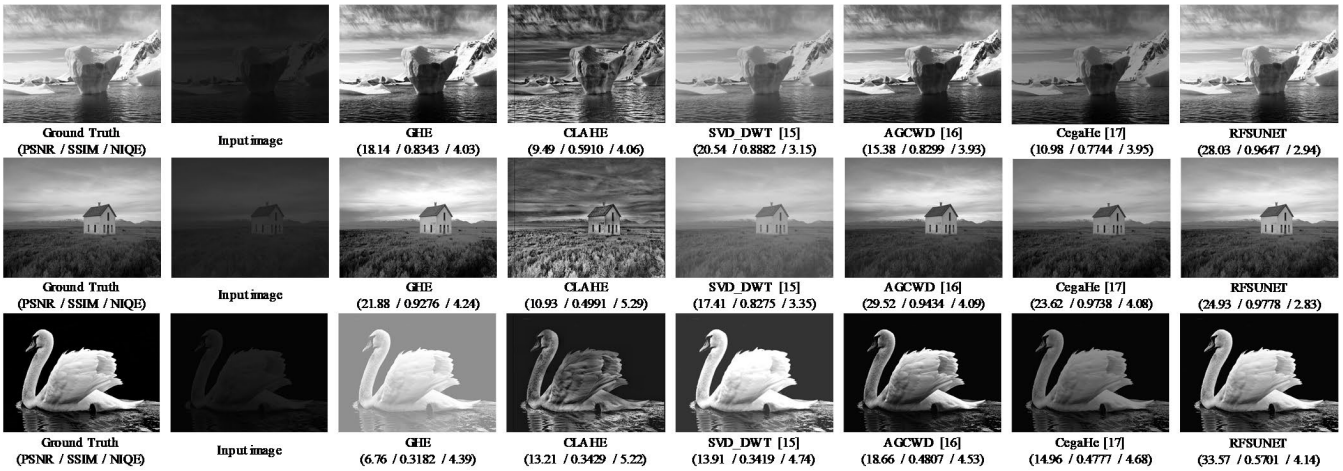

FIGURE 13. Quantitative evaluation of algorithms (PSNR, SSIM, NIQE) on an image enhancement task. Compared to other algorithms, the image derived from the proposed method is most similar to the GT image.

TABLE 4. Quantitative comparison of algorithmic performance on image enhancement task.

2635 images in total, brightness is randomly dark					
Algorithm	MAE (↓)	NIQE (↓)	PSNR (↑)	SSIM (↑)	MS-SSIM (↑)
GHE	48.88	5.72	14.21	0.7064	0.8014
CLAHE	56.42	5.08	12.71	0.6234	0.8292
SVD_DWT	39.40	4.61	16.90	0.8106	0.9208
AGCWD	22.83	5.40	21.36	0.8771	0.9569
CegeHe	62.54	5.34	13.74	0.6963	0.8909
RFSUNET	**8.25**	**4.41**	**32.89**	**0.9436**	**0.9847**

each reconstructed image a point value from 1 (the worst quality) to 5 (excellent quality).

Table 5 provides a qualitative index of the MOS for the multi-type noise image denoising task. We showed 10 selected videos to 21 different study subjects. For each, we asked in a questionnaire survey whether the noise was properly removed so that there were no inconveniences in viewing the images, i.e., if the videos were smooth and natural. Most of the participants who conducted the evaluation presented the following common opinions.

1) Quality was considered worse when noise remained in successive scenes or when the transition between frames was unnatural.
2) Quality was considered better when the transition between frames was natural and clear.
3) The usefulness of quantitative algorithm evaluation metrics was considered quite limited (based on subjects' comparisons of their qualitative scores with quantitative evaluation metrics provided to them afterward).

TABLE 5. Qualitative comparison of algorithmic performance on multi-type noise image denoising task.

Video ID.	BM3D	WNNM	VDSR	DnCNN	FFDNet	MWCNN	Nose2Noise	RFSUNET
10	1.22	1.00	2.03	2.05	2.47	3.87	3.65	**4.21**
25	1.33	1.00	2.23	2.31	3.11	3.76	3.42	**4.05**
38	1.18	1.00	2.27	2.26	2.44	3.88	3.72	**3.97**
51	1.05	1.00	1.67	1.83	2.32	3.51	3.38	**3.84**
56	1.11	1.00	2.15	2.16	2.56	3.96	3.78	**4.23**
70	1.03	1.00	1.98	2.01	2.37	3.44	3.26	**3.78**
85	1.23	1.00	2.18	2.19	2.96	3.84	3.72	**4.31**
97	1.26	1.00	2.34	2.31	2.87	3.91	3.65	**4.26**
Total Avg	1.18	1.00	2.11	2.14	2.64	3.77	3.57	**4.08**

FIGURE 14. Quantitative (variation, MAE, PSNR, SSIM, MS-SSIM) and qualitative (MOS) evaluation of algorithmic performance on a multi-type noise image denoising task.

MWCNN [10], Noise2Noise [11], and the proposed method received good qualitative evaluations. According to the opinions of evaluators, the proposed method not only effectively removed multi-type noise, but also minimized the variation among frames, resulting in a natural image reconstruction with uniform quality. In the denoising task, the restoration of features and backgrounds needed to be harmonious and natural. Although numerical evaluation is important for quantitative measurement, the qualitative evaluation of the proposed method demonstrates that the perceptual measure is important in terms of quantitative measurement [39].

Table 6 shows the MOS of the image enhancement task. We showed 35 selected images to 21 different study subjects. For each, we asked in a questionnaire survey how closely it was restored to the GT, i.e., how natural the images were.

TABLE 6. Qualitative comparison of algorithmic performance on image enhancement task.

	GHE	CLAHE	SVD_DWT	AGCWD	CegaHe	RFSUNET
Avg	1.41	1.00	2.23	3.71	3.46	**4.37**

Through this quantitative measurement, three methods received good qualitative evaluations: AGCWD [19], CegaHe [20], and the proposed method. In the image enhancement task, the restoration of dynamic brightness needed to be harmonious and natural. Among these three winning methods, the proposed method was evaluated to give the most natural and harmonious restoration. Table 5 shows that the proposed method also provided the best performance for this task in terms of quantitative metrics.

D. TOTAL MEASUREMENT

In this section, quantitative and qualitative evaluations are combined and evaluated.

Fig. 14 uses an integrated indicator to show that the performance of the proposed method for the multi-type noise image denoising task is excellent. The performance of the proposed method is also good the image enhancement task.

Our study once again emphasizes that both quantitative and qualitative indicators should be used to determine how well image restoration has been performed.

E. REAL EXPERIMENT

In this section, actual datasets are created and evaluated to demonstrate the diversity of the proposed methods.

In IR systems, various noises caused by scene changes or heat generation can be observed frequently. For this reason, NUC method is essential to obtain high quality images. NUC method works to address common multi-type noise (AWGN, line noise) and temperature compensation.

However, as mentioned by FLIR [41], it takes about 20 minutes to warm up for accurate temperature measurements. During preheating, the NUC method continues to

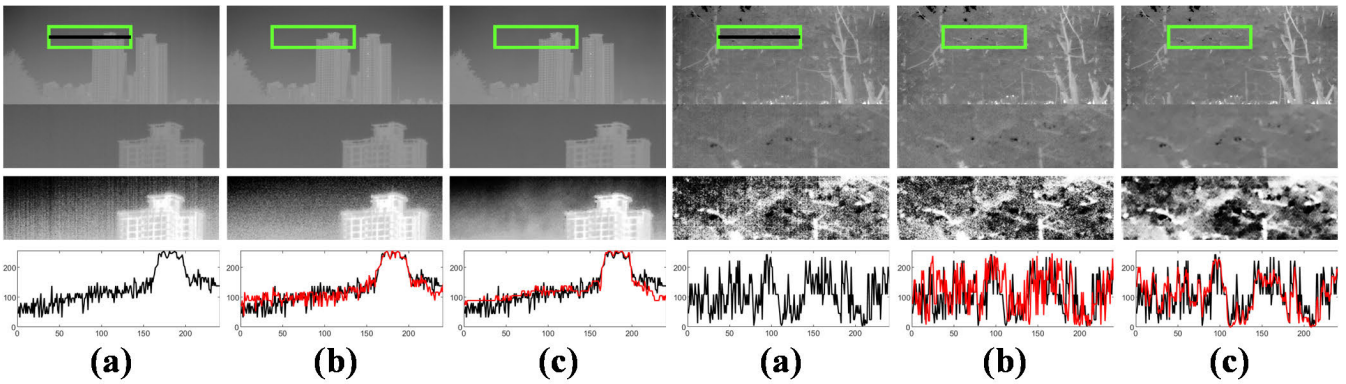

FIGURE 15. Real image comparison using proposed method with NUC method. (a) an examples of real IR systems, (b) results of applying NUC method (c) results of applying proposed method. Looking at the normalized image, it can be seen that the proposed method removes noise better than the NUC method. In addition, it can be observed from the line profile that it is well restored in harmony based on the uniformity of the brightness value.

work for 20 minutes, and no image can be obtained during operation.

Because of these limitations, the proposed method is used to replace the NUC method, and the following Fig. 15 is the result.

As shown in Fig. 15, It can observe that the NUC method is not always perfect. Because of the characteristic of the noise, it is not easy to obtain an image (GT Image) that is perfectly and harmoniously removed. The proposed method not only shows results equivalent to images processed with the NUC method, but also shows the same and excellent level of restoration quality without any deviation between images, and better results depending on the specific scene.

V. CONCLUSION

This paper presented a residual learning based RFSUNET architecture for image restoration, which consists of GRL and LRL. For multi-type noise image denoising tasks, actual multi-type noises such as TIR images and HS images were analyzed and used to generate equivalent evaluation datasets. The performance of RFSUNET in restoration tasks was evaluated through quantitative and qualitative means. In particular, the proposed method was found to effectively remove various types of noise in comparison with conventional algorithms and networks. In addition, the results show low variation and uniform quality in both dynamic and static images. Also, we performed comparative analysis experiments on how effectively to remove multi-type noises, which are frequently observed in real IR systems, using the NUC method and the proposed method. Through these experiments, we were able to prove the excellence of our research once again.

However, most CNN-based algorithms, including this proposed method, show slightly blurring results due to the deep learning structure. This flaw produces admittedly imperfect restoration results compared to the GT image.

In future work, we will create a network that sharpens and blurs the somewhat blurry image once again. As a two-stage network, we aim to design a CNN that enhances the performance of restoration as close as possible to the GT image.

REFERENCES

[1] B.-L. Hu, S.-J. Hao, D.-X. Sun, and Y.-N. Liu, "A novel scene-based non-uniformity correction method for SWIR push-broom hyperspectral sensors," *ISPRS J. Photogramm. Remote Sens.*, vol. 131, pp. 160–169, Sep. 2017.

[2] B. Kim, M. Kim, and Y. Chae, "Background registration-based adaptive noise filtering of LWIR/MWIR imaging sensors for UAV applications," *Sensors*, vol. 18, no. 2, p. 60, Dec. 2017.

[3] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.

[4] S. Gu, Q. Xie, D. Meng, W. Zuo, X. Feng, and L. Zhang, "Weighted nuclear norm minimization and its applications to low level vision," *Int. J. Comput. Vis.*, vol. 121, no. 2, pp. 183–208, Jan. 2017.

[5] B. H. Kim, C. Bohak, K. H. Kwon, and M. Y. Kim, "Cross fusion-based low dynamic and saturated image enhancement for infrared search and tracking systems," *IEEE Access*, vol. 8, pp. 15347–15359, 2020.

[6] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, p. 7553, May 2015.

[7] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1646–1654.

[8] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, Jul. 2017.

[9] K. Zhang, W. Zuo, and L. Zhang, "FFDNet: Toward a fast and flexible solution for CNN-based image denoising," *IEEE Trans. Image Process.*, vol. 27, no. 9, pp. 4608–4622, Sep. 2018.

[10] P. Liu, H. Zhang, K. Zhang, L. Lin, and W. Zuo, "Multi-level wavelet-CNN for image restoration," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2018, pp. 773–782.

[11] J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and T. Aila, "Noise2Noise: Learning image restoration without clean data," 2018, *arXiv:1803.04189*. [Online]. Available: http://arxiv.org/abs/1803.04189

[12] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. New York, NY, USA: Person, 2018.

[13] X. Chen, L. Song, and X. Yang, "Deep RNNS for video denoising," *Proc. SPIE*, vol. 9971, Sep. 2016, Art. no. 99711T.

[14] Y. Chang, L. Yan, and S. Zhong, "Transformed low-rank model for line pattern noise removal," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 1735–1743.

[15] Y. Chang, L. Yan, L. Liu, H. Fang, and S. Zhong, "Infrared aerothermal nonuniform correction via deep multiscale residual network," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 7, pp. 1120–1124, Jul. 2019.

[16] Y. Chang, M. Chen, L. Yan, X.-L. Zhao, Y. Li, and S. Zhong, "Toward universal stripe removal via wavelet-based deep convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3235–3247, Jun. 2016.
[17] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*. [Online]. Available: http://arxiv.org/abs/1804.02767
[18] H. Demirel, C. Ozcinar, and G. Anbarjafari, "Satellite image contrast enhancement using discrete wavelet transform and singular value decomposition," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 2, pp. 333–337, Apr. 2010.
[19] S.-C. Huang, F.-C. Cheng, and Y.-S. Chiu, "Efficient contrast enhancement using adaptive gamma correction with weighting distribution," *IEEE Trans. Image Process.*, vol. 22, no. 3, pp. 1032–1041, Mar. 2013.
[20] C.-C. Chiu and C.-C. Ting, "Contrast enhancement algorithm based on gap adjustment for histogram equalization," *Sensors*, vol. 16, no. 6, p. 936, Jun. 2016.
[21] Y.-S. Chen, Y.-C. Wang, M.-H. Kao, and Y.-Y. Chuang, "Deep photo enhancer: Unpaired learning for image enhancement from photographs with GANs," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6306–6314.
[22] Y. Jiang, X. Gong, D. Liu, Y. Cheng, C. Fang, X. Shen, J. Yang, P. Zhou, and Z. Wang, "EnlightenGAN: Deep light enhancement without paired supervision," 2019, *arXiv:1906.06972*. [Online]. Available: http://arxiv.org/abs/1906.06972
[23] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2016.
[24] M. S. M. Sajjadi, B. Scholkopf, and M. Hirsch, "EnhanceNet: Single image super-resolution through automated texture synthesis," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 4501–4510.
[25] K. Zhang, W. Zuo, S. Gu, and L. Zhang, "Learning deep CNN denoiser prior for image restoration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3929–3938.
[26] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4681–4690.
[27] P. A. Coelho, J. E. Tapia, F. Pérez, S. N. Torres, and C. Saavedra, "Infrared light field imaging system free of fixed-pattern noise," *Sci. Rep.*, vol. 7, no. 1, p. 13040, Dec. 2017.
[28] E. Agustsson and R. Timofte, "NTIRE 2017 challenge on single image super-resolution: Dataset and study," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 1122–1131.
[29] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[30] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2015, pp. 234–241.
[31] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.
[32] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 2012, pp. 1097–1105.
[33] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[34] A. Odena, V. Dumoulin, and C. Olah. (2016). *Deconvolution and Checkerboard Artifacts*. Distill. [Online]. Available: http://distill.pub/2016/deconv-checkerboard
[35] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent.*, 2015, pp. 1–41.
[36] S. Anwar, S. Khan, and N. Barnes, "A deep journey into super-resolution: A survey," 2019, *arXiv:1904.07523*. [Online]. Available: http://arxiv.org/abs/1904.07523
[37] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss functions for image restoration with neural networks," *IEEE Trans. Comput. Imag.*, vol. 3, no. 1, pp. 47–57, Mar. 2017.
[38] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. 37th Asilomar Conf. Signals, Syst. Comput.*, vol. 2, 2003, pp. 1398–1402.
[39] Y. Blau and T. Michaeli, "The perception-distortion tradeoff," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6228–6237.
[40] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Mar. 2013.
[41] FLIR. (2020). *What is a Non-Uniformity Correction (NUC)?*. [Online]. Available: https://www.flir.com/discover/professional-tools/what-is-a-non-uniformity-correction-nuc/

HO MIN JUNG is currently pursuing the M.S. degree with the School of Electronic Engineering, Kyungpook National University (KNU), Daegu, South Korea. His current research interests include deep-learning based image restoration, visual object tracking, and autonomous vehicle.

BYEONG HAK KIM received the Ph.D. degree from the School of Electronic Engineering, Kyungpook National University (KNU), Daegu, South Korea, in 2020. He is currently a Senior Engineer with the Department of Optronics, HANWHA Systems Company, Gumi, South Korea. His current research interests include infrared image enhancement, visual object tracking, deep-learning object auto detection, 3D laser radar, and counter drone systems.

MIN YOUNG KIM (Member, IEEE) received the B.S., M.S., and Ph.D. degrees from the Korea Advanced Institute of Science and Technology, South Korea, in 1996, 1998, and 2004, respectively. He was a Senior Researcher with Mirae Corporation, from 2004 to 2005. He was also a Chief Research Engineer in artificial vision systems for intelligent machines and robots with Kohyoung Corporation, from 2005 to 2009. Since 2009, he has been with the School of Electrical Engineering and Computer Science, Kyungpook National University, as an Assistant Professor. He was a Visiting Associate Professor with the Department of Electrical and Computer Engineering and the School of Medicine, Johns Hopkins University, from 2014 to 2016. He is currently an Associate Professor with the School of Electronics Engineering, Kyungpook National University. He is also a Deputy Director with the KNU−LG Convergence Research Center and the Director of the Research Center for Neurosurgical Robotic Systems. His research interests include visual intelligence for robotic perception, recognition of autonomous unmanned ground, and aerial vehicles.