A Composite Noise Removal Network Based on Multi-domain Adaptation

Fan Bai[1], Pengfei Li[2]*, Haoyang Sun[3], Hui Zhang[4]
School of Equipment Engineering, Shenyang Ligong University, Shenyang, China[1, 4]
Science and Technology on Electromechanical Dynamic Control Laboratory[2]
School of Mechanical and Electrical Engineering, Beijing Institute of Technology, Xi'an Beijing, China[2]
School of Mechanical and Electrical Engineering, Beijing Institute of Technology Beijing, China[3]

***Abstract*—Addressing the limitation of conventional single-scene image denoising algorithms in filtering mixed environmental disturbances, and recognizing the drawbacks of cascaded image enhancement algorithms, which have poor real-time performance and high computational demands, The composite weather adaptive denoising network (CWADN) is proposed. A Cascade Hourglass Feature Extraction Network is constructed with a visual attention mechanism to extract characteristics of rain, fog, and low-light noise from authentic natural images. These features are then transferred from their original real distribution domain to a synthetic distribution domain using a deep residual convolutional neural network. The generator and style encoder of the adversarial network work together to adaptively remove the transferred noise through a combination of supervised and unsupervised training, this approach achieves adaptive denoising capabilities tailored to complex natural environmental noise. Experimental results demonstrate that the proposed denoising network yields a high signal-to-noise ratio while maintaining excellent image fidelity. It effectively prevents image distortion, particularly in critical target areas. Additionally, it adapts to various types of mixed noise, making it a valuable tool for preprocessing images in advanced machine vision algorithms such as target recognition and tracking.**

Keywords*—Image denoising; domain adaptation; generative adversarial network; autoencoder

I. INTRODUCTION

Haze, rain, and low illumination are the three types of natural noises that have the greatest impact on the detection accuracy of machine vision. These noises will destroy the optical information in the original image through global blurring, superimposed noise, and information desalination, bringing a great challenge to all-weather target detection tasks.[1], [2]. Therefore, the denoising methods for the above three natural noises have become the key research directions of domestic and foreign scholars in the field of image denoising. Among them, the study [3] directly learns and estimates the mapping function between the noisy image and its noise-free counterpart and cooperates with the bilateral rectified linear unit (BReLU) to reduce the search space and improve the convergence, to realize the end-to-end training and interference process of the dehazing network. Based on the transmittance parameters and atmospheric light of the scattering model, the research in [4] directly learned the residual information between the haze image and the haze-free image by using smooth dilated convolution and threshold fusion sub-networks to realize image dehazing. The study in [5] pass a binary rain mask to the multi-task network for learning, and the negative rain layer generated by iteration is compared with the input, which reduces the effect of rain noise on the original image. Based on the reverse stacking denoising strategy, the study in [6] use a dataset marked with the rain size to train a multi-task network and realizes image denoising through the obtained rain noise features, [7] use Retinex theory, the reflectance image under ideal illumination is multiplied by the noisy low-illumination image, and the low-illumination noise is directly removed by guided filtering, which solves the problem that the traditional low-illumination denoising algorithm over-enhances or under-enhances some areas. Enhanced question in study [8] constructed an unsupervised network EnlightenGAN, which can be trained in a large number of imprecisely matched images to establish a mapping relationship, which overcomes the problem of low-illumination noise denoising accuracy when the dataset is insufficient. The study in [9] designed a dual-branch unit endowed with physics-aware, complemented by a course learning contrast regularization approach. This research underscores the significance of fine-tuning various negative samples within the contrast regularization process. These insights offer valuable concepts for leveraging multimodal contrastive regularization techniques to enhance image quality.

However, because the environmental noise in nature appears in the form of mixed accompaniment, that is the three kinds of noises of haze, rain and low illumination may be generated at the same time in different weather and will be mapped on the original image in the form of mixing in any proportion. The interference of noise on image information will become more complicated. The above algorithms all adopt the directional denoising strategy, and it is difficult to achieve an optical result when it comes to mixed noise in real scenes. Therefore, researchers gradually focus their research on the field of adaptive denoising that is more in line with actual needs and can integrate various physical models of weather noise. The research in [10] applied the strategy of Neural Architecture Search in reinforcement learning to image restoration to generate the most suitable denoising network structure, and at the vector level, Denoising the output results of Encoders), giving the algorithm the ability of adaptive filtering of compound noise; The study [11] introduced an attention mechanism (Spatial Attention Mechanism, SAM) and

cross-multi-stage feature fusion in the encoding and decoding process of the network. The mechanism (Cross-stage Feature Fusion, CSFF) avoids the loss of target feature information before and after denoising. It takes into account the functions of efficient denoising and target information transfer. Although these composite noise-denoising deep neural networks have good image processing capabilities, their overly complex structures lead to high demands on computing resources. The training difficulty and convergence speed are not ideal. At the same time, a large number of supervised learning links make this kind of network must be supported by abundant real noise datasets to obtain better training results. When the real noisy images of the actual scene are difficult to obtain, the noise reduction accuracy of this kind of network will be greatly improved. These problems will limit the versatility of denoising algorithms in real environments. Providing all-weather adaptive denoising capabilities for platforms such as space vehicles and ground-based photodetectors is difficult.

In order to solve the problems of the above algorithms and improve the adaptability and all-weather computing efficiency of the image denoising algorithm in the denoising task for complex natural environments, this paper proposes a denoising method for the free mixed environment noise of rain, haze and low illumination. Noise neural network, the innovations of this network are: (1) An end-to-end image denoising network based on domain transfer is proposed, which realizes rain, haze, low illumination and three kinds of mixed noise images under a single structure. (2) Integrate multi-stage autoencoder structure and multi-domain transfer strategy to achieve directional separation and targeted denoising of an unknown proportion of mixed noise. (3) Based on domain adaptive generative confrontation module, effectively reduce the difference between noisy synthetic data and real noisy data is eliminated, and the traditional denoising algorithm training process is free from the dependence on a large number of real noisy data sets. Based on the above methods and characteristics, the denoising network proposed in this paper achieves a high signal-to-noise ratio and image structure consistency in multi-type mixed noise filtering tasks and achieves high-quality denoising and information restoration for complex natural environment noise.

II. Proposed Methods

This paper draws extensive inspiration from the multi-level architecture of MPRNet [12], which strikes a balance between preserving local and global information. It introduces the concept of projecting any natural environmental image into multiple modes, followed by individual processing and subsequent integration. The composite weather adaptive denoising network (CWADN) designed in this paper is composed of a separation module, a denoising module and a conversion module, and its network structure is shown in Fig. 1.

CWADN comprises three integral components: the Multi-stage Progressive Separation Network (MPSN), the Multi-domain Translation Denoise Network (MTDN), and the Domain Adaptation Translation Network (DATN). In the denoising process, CWADN initially takes images containing complex real-world composite noise as input into the multi-level autoencoder of MPSN. Subsequently, MTDN leverages the noise distribution within the image space as a feature and transfers the three distinct noise images to the synthetic domain for generation. Finally, DATN restores these three transformed images, consolidates the acquired results, and accomplishes the denoising task. Through these methodologies, CWADN achieves adaptive and precise denoising, as well as the restoration of target image information for original images captured in real-world scenarios featuring natural compound noise.

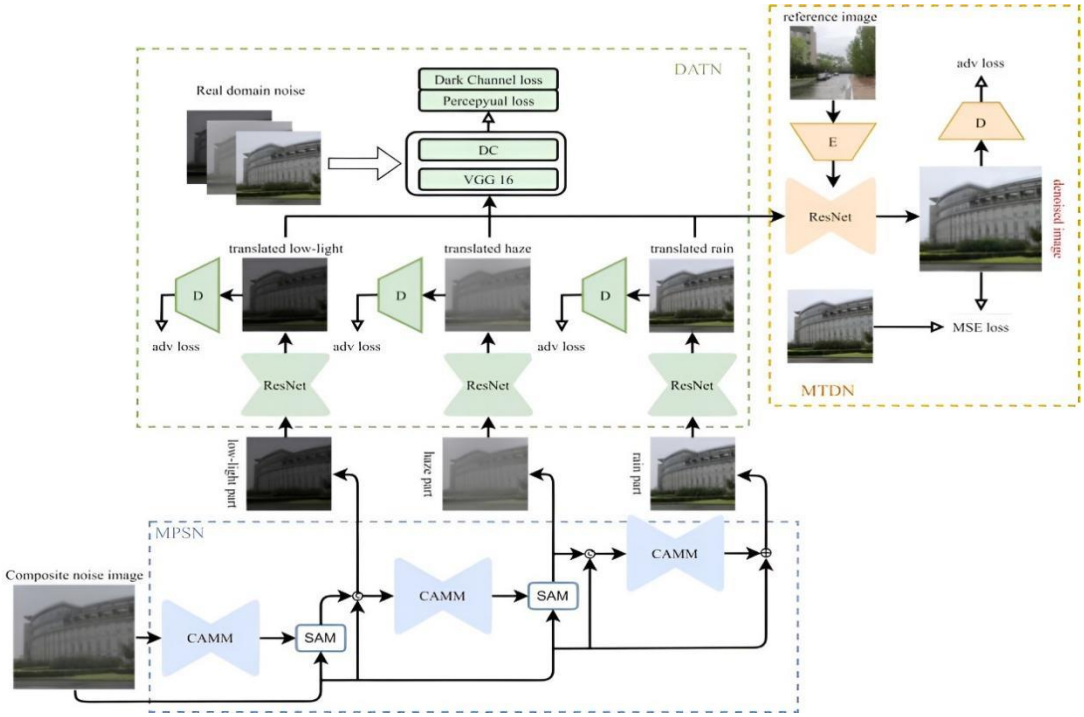

Fig. 1. The framework of our proposed network.

III. NETWORK DETAILS

A. Multi-Stage Progressive Separation Network

For any noisy image shot in a natural environment, it can be regarded as the result of the original noise-free image. I_{ori} is affected by three kinds of noise: rain, haze and low illumination, and different types of noise have different effects on I_{ori}. The image quality degradation process can be expressed as:

$$I_n = \lambda_{rain} N_{rain}(I_{ori}) \odot \lambda_{haze} N_{haze}(I_{ori}) \odot \lambda_{dark} N_{dark}(I_{ori}) \quad (1)$$

Where $\odot$ is the mixing operation of noise, N_i and λ_i are the noise degradation sub-function and the noise component weight coefficient under the conditions of rain, haze, and low illumination (i= {rain, haze, dark}), respectively. When there is no certain kind of noise, the trade-off λ_i equals to 0, and when the image is an ideal noise-free image, all λ are 0.

Due to the diversity and unpredictability of the combination of each proportional coefficient λ_i of the composite noise in the natural environment, it is difficult for the conventional denoising network to fit the image quality degradation function with a certain coefficient, and it is impossible to learn the mapping relationship between In and I_{ori} or complete image restoration. To solve this problem, we design a multi-stage progressive separation network (MPSN) to separate the composite noise and sequentially extract the low-illumination noise component N_{dark}, the haze noise component N_{haze} and the rain noise component N_{rain}, and the different noise components are limited to their own domain according to their characteristics. MPSN is composed of three cascaded Channel Attention Autoencoder Module (CAAM), and its network structure is shown in Fig. 2.

In the autoencoder network based on the full convolution layer, although the continuous convolution operation can enrich the semantic information of the feature map, it also causes the gradual loss of the texture information in the original image, which makes the deconvolution operation unable to correct the decoding process and accurately restored details of the image. Therefore, we add a Channel Attention Block (CAB) [13] to all convolution and deconvolution operations in all CAAM to reduce information loss in key areas of interest. The input feature map group of CAB is denoted as $X = [x_1, x_2, \ldots, x_n]$, where x_n is the feature map of the n channel with size $H \times W$. The frequency information of features is included in x_n, and its high-frequency features can better represent the edge and detail information in the image. Therefore, the global average pooling obtains the global feature frequency z_n by scaling the size of x_n to 1×1. Then, in order to extract the channel feature of z_n and obtain the weight coefficient through the activation operation, CAB will adjust the weight of x_n to obtain the final feature x_n * with the attention mechanism. This process can be expressed as:

$$\begin{cases} z_n = \dfrac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} x_n(i, j) \\ x_n^* = x_n \otimes \sigma(\phi(z_n)) \end{cases} \quad (2)$$

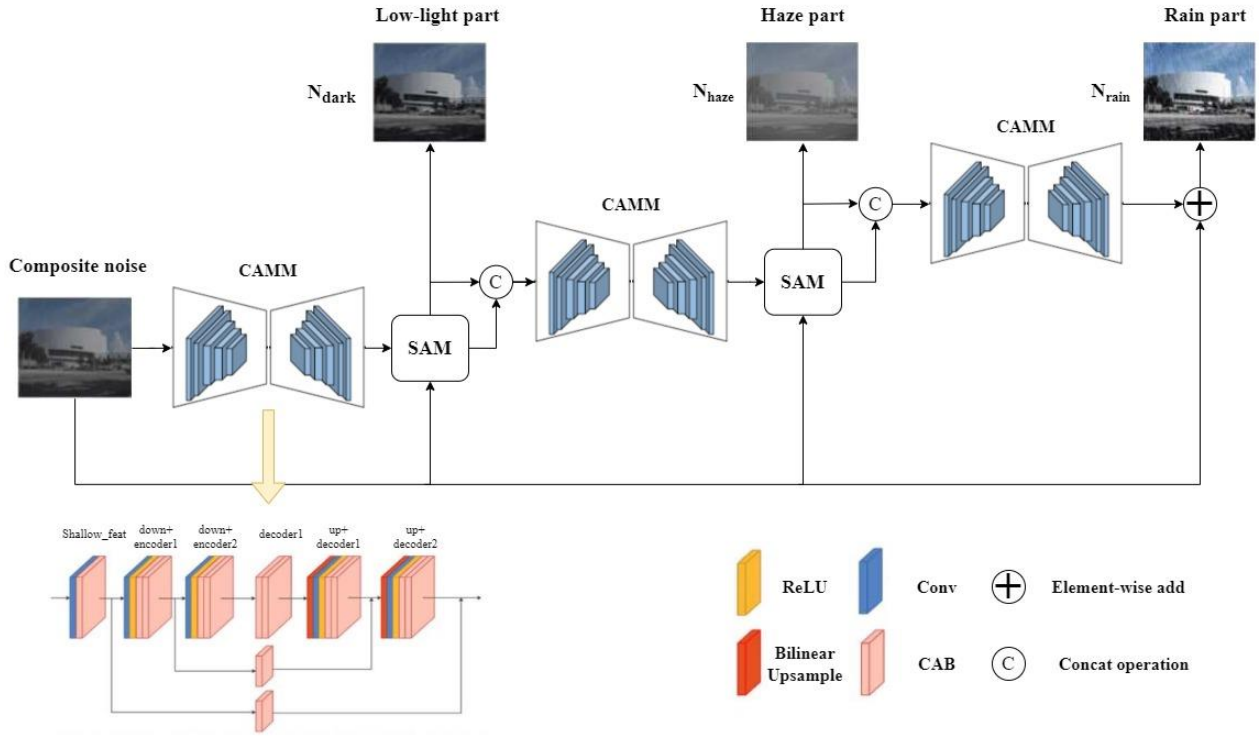

Fig. 2. Channel attention autoencoder Module for our network.

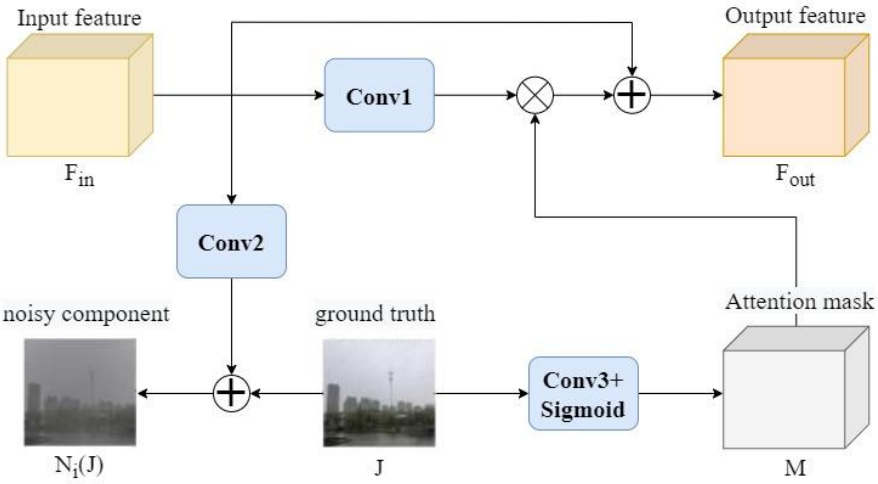

Fig. 3. Spatial attention module for our network.

Where *i* and *j* represent a pixel's horizontal and vertical position coordinates in the current image, respectively; σ and ϕ represent the sigmoid activation function and the channel feature extraction process, respectively, and $\otimes$ is the corresponding multiplication on pixel-wise. Through the above attention mechanism, MPSN can assign higher weights to feature channels with high-frequency information in the reconstruction process so as to preserve the texture details of the region of interest.

In addition, we introduce a Spatial Attention Module [14] between each two CAMM to enhance the transfer and fusion of information between different stages. As shown in Fig. 3, SAM obtains the noise component Ni by performing channel dimension reduction on the input feature map Fin and adding it pixel-by-pixel with the noise-free image Iori. At the same time, SAM extracts features from Iori and activates it to obtain the attention mask M and then linearly changes Fin to obtain the output feature map Fout as the input of the next stage. The process is expressed as:

$$\begin{cases} N_i(I_{ori}) = W_2(F_{in}) + I_{ori} \\ M = sigmoid(W_3(I_{ori})) \\ F_{out} = F_{in} + W_1(F_{in}) \times M \end{cases} \quad (3)$$

where, the W_1, W_2, and W_3 represent three convolution operations, respectively. After passing through the SAM module, F_{out} contains ground truth information, which can improve the texture feature representation ability of the subsequent separated images. Due to the constraint of the attention mask, the transmission of invalid information to the next stage is suppressed. After the original noisy image is separated by MPSN, three images containing only a single type of independent noise can be obtained, which effectively solves the problem that the composite noise image quality degradation function is difficult to fit and limits the scope of the solution space, and facilitates the subsequent denoising network.

B. Multi-Domain Translation Denoise Network

At present, the image restoration methods using image translation mainly regard the noisy image and the noise-free image as two independent domains and use the generator to learn the mapping relationship between them to realize image denoising. Still, this method can only remove a single type of noise. In order to realize the adaptive removal of multiple types of composite noise under a single model, inspired by StarGANv2 [15], we construct a multi-domain translation network (MTDN), including generator *G*, style encoder *E* and discriminator *D*, three main parts which are shown in Fig. 4.

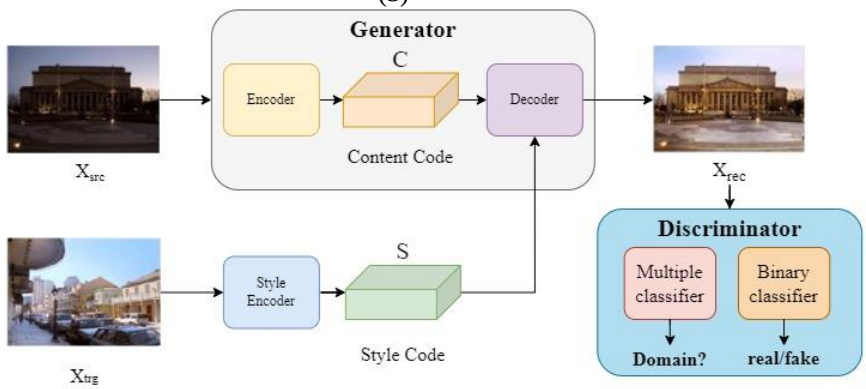

Fig. 4. The framework of multi-domain translation network.

We take the samples of noisy images, which include rain, haze and low illumination as the source domain X_{src}, that is, X_{src}= {N_{haze}, N_{rain}, N_{dark}}, and the noise-free images as the target domain X_{trg}=I_{org}, by learning the mapping function between X_{src} and X_{trg} to realize the removal of many different types of noises. E extracts features from the target domain image x_{trg} (x_{trg}∈X_{trg}), and obtains the style code s containing the high-dimensional feature information of the noise-free image. G extracts the high-dimensional semantic information c of the original noise image x_{src} (x_{src}∈X_{src}) through internal coding, integrates c and s in the decoder, and decodes the reconstructed image of x_{rec}. As a multi-task discriminator, D is composed of a binary classifier and a multi-classifier. The multi-classifier is used to determine whether the x_{rec} has the characteristics of a noise-free image, that is, whether the denoising process is complete; the binary classifier evaluates the image quality of the x_{rec}, that is, whether the reconstructed image has higher restoration. Through the above process, MTDN can achieve directional denoising for various types of noises and effectively retain the original image's texture, details and other information.

C. Domain adaptation translation network

Due to the inability to obtain a large number of real noisy-clean paired datasets, both MPSN and MTDN can only be trained on synthetic datasets. However, the spatial distribution and texture features of environmental noise in the synthetic dataset are different from those in the real dataset. This makes the model trained on the synthetic dataset generalize well to the real-world samples and reduces the denoising performance. In order to solve the above problems, we use DATN to improve the adaptability of the model in different datasets.

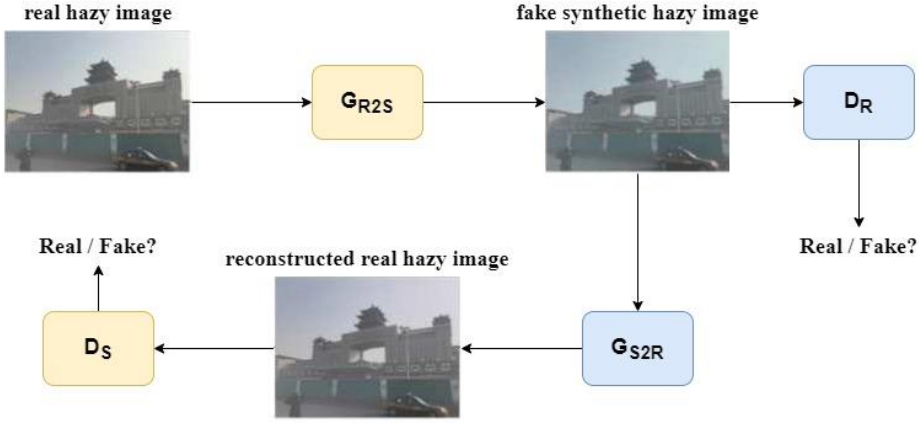

Fig. 5. The framework of domain adaptation translation network.

The structure of DATN is shown in Fig. 5, which is composed of three translation modules with the same structure, corresponding to three different types of noise, respectively. DATN takes CycleGAN [16] as the core framework, which contains two generators, G_{S2R}, G_{R2S} and two discriminators, D_R and D_S. The generator G_{S2R} takes the synthetic noise N_{syn} as input and uses the spatial distribution characteristics of noise in the real noise image N_{rea} as the learning object to reconstruct N_{syn}. The discriminator D_R compares the difference between the reconstructed image and N_{rea} and uses it as an indicator to constrain G_{S2R}. During training, G_{S2R} improves its domain adaptation ability by continuously increasing the similarity between the generated image and N_{rea}. Similarly, the same process will be used for training and calculation for G_{R2S} and D_R. Therefore, DATN can realize the mutual translation between real and synthetic noise images through a large number of unpaired datasets, thereby improving the generalization performance of CWADN in real noise samples.

D. Loss function

In order to obtain the best training effect of each functional module and improve the convergence ability and subsequent calculation accuracy of network denoising learning, this paper analyzes the loss of each sub-network in the proposed CWADN according to the task types and structural characteristics of different functional modules. The function has been specially designed.

Image separation Losses

We expect MPSN to separate three different kinds of noise N_i from composite noise. We train the MPSN in a supervised manner due to the synthetic training samples. Firstly, we calculate the mean square loss between MPSN result N_i^{out} and ground truth N_i^{gt}

$$L_{mse}^{MPSN} = \left\| N_i^{out} - N_i^{gt} \right\|_2^2 \quad (4)$$

Where i = {haze, rain, dark} represents the three noise conditions under haze, rain and low illumination, respectively. In order to prevent the noise separation process from destroying the information of non-noise areas in the original image, we use structural loss L_{ssim}^{MSPN} to constrain the distortion between N_i^{out} and N_i^{gt}, and judge the image quality after noise separation.

$$L_{ssim}^{MSPN} = 1 - SSIM(N_i^{out}, N_i^{gt}) \quad (5)$$

Where SSIM is the image similarity calculation rule, which is based on the distribution of image data in the mean, variance and covariance, compared the difference between N_i^{out} and

N_i^{gt} in lighting, contrast, and image structure, the computing process can be defined as:

$$SSIM(N_i^{out}, N_i^{gt}) = \frac{(2\mu_{out}\mu_{gt} + c_1)}{(\mu_{out}^2 + \mu_{gt}^2 + c_1)} \cdot \frac{(2\sigma_{out,gt} + c_2)}{(\sigma_{out}^2 + \sigma_{gt}^2 + c_2)} \quad (6)$$

Where μ_{out} and μ_{gt} represent the mean of N_i^{out} and N_i^{gt}, σ_{out}^2 and σ_{gt}^2 represent the variance of N_i^{out} and N_i^{gt}, $\sigma_{out,gt}$ is the covariance of N_i^{out} and N_i^{gt}. c_1 and c_2 are tiny decimals, L is the dynamic range of pixel values. When the SSIM value equals to 1, it represents that the MPSN has strong image noise separation and quality restoration capabilities. Therefore, the total loss of MPSN can be expressed as follows:

$$L^{MSPN} = \lambda_{mse} L_{mse}^{MSPN} + \lambda_{ssim} L_{ssim}^{MSPN} \quad (7)$$

Where λ_{mse} and λ_{ssim} represent the trade-off of L_{mse}^{MSPN} and L_{ssim}^{MSPN} respectively.

Image denoising Losses

For MTDN, unsupervised training can enable the network to achieve denoising by reconstructing images. Still, the reconstructed images lose texture details, affecting the accuracy of subsequent target recognition, tracking and other advanced machine vision [17]. To solve this problem, we combine supervised and unsupervised learning, adopting a semi-supervised learning method to train MTDN. The unsupervised training process uses a generative adversarial loss. The style encoder E_{dn} maps the target domain image x_{trg} to the corresponding style code $s = E_{dn}(x_{trg})$, and the generator G_{dn} integrates the original domain images x_{src} and s. After reconstruction and denoising, the image $G_{dn}(x_{src}, s)$ is combined with the discriminator G_{dn} to get the adversarial loss of the denoising network, which is defined as:

$$L_{adv}^{MTDN} = \mathbb{E}_{x_{src}}[\log(D_{dn}(x_{src}))] + \mathbb{E}_{x_{src}, x_{trg}}[\log(1 - D_{dn}(G_{dn}(x_{src}, s)))] \quad (8)$$

MTDN realizes the adaptive removal of various types of noise by learning the mapping function between x_{src} and x_{trg}. However, when only introduce an adversarial loss in the training process; it will cause the generator to confuse the discriminator to get a higher score, which will reduce the problem of the diversity of generated images, called mode collapse. In order to solve this problem, we add the cycle consistency loss; after translating image G_{dn} to the source domain again, the result should be consistent with source domain image x_{src}, that is $x_{src} = G_{dn}(G_{dn}(x_{src}, s), s^*)$, the above process can be expressed as:

$$L_{cyc}^{MTDN} = \mathbb{E}_{x_{src}, x_{trg}}[\| x_{src} - G_{dn}(G_{dn}(x_{src}, s), s') \|_1] \quad (9)$$

Where s* is the style code of the source domain image, that is, $s^* = E_{dn}(x_{src})$. In addition, in order to make the style code s better guide the image reconstruction process, we introduce the style reconstruction loss on the basis of the above, which is similar to the cycle consistency loss. The style code s of the noise-free image is obtained after G_{dn}; when it is encoded by E_{dn} again, the output result should be less different from s, that is, $s^* = E_{dn}(G_{dn}(x_{src}, s))$, the style reconstruction loss is defined as follows:

$$L_{sty}^{MTDN} = \mathbb{E}_{x,z}[\| s - E_{dn}(G_{dn}(x_{src}, s) \|_1] \quad (10)$$

Through the above unsupervised loss, MTDN can recover images affected by three different noises of rain, haze and low illumination, respectively, and can effectively preserve the content information of the images. In order to further retain the details of the denoised image, we introduce a supervised loss based on the unsupervised loss to calculate the mean square error between the denoised image and its noise-free counterpart Iori and preserve the underlying texture information of the image by minimizing the pixel-by-pixel difference between the two.

$$L_{mse}^{MTDN} = \| G_{dn}(x_{src}, E_{dn}(x_{trg})) - I_{ori} \|_2^2 \quad (11)$$

In summary, the overall loss function of MTDN Loss can be expressed as follows:

$$L^{MTDN} = L_{adv}^{MTDN} + \lambda_{sty} L_{sty}^{MTDN} + \lambda_{cyc} L_{cyc}^{MTDN} + \lambda_{mse} L_{mse}^{MTDN} \quad (12)$$

Where, λ_{sty}, λ_{cyc}, and λ_{mse} are trade-off weights. During the training process, MTDN minimizes loss to realize the targeted removal of rain, haze and low illumination on the basis of retaining the target texture information.

E. Image translation Losses

Noise Ni can be divided into real noise N_{rea}^i and synthetic noise N_{syn}^i that is Ni= {N_{rea}^i, N_{syn}^i}, The training datasets of MPSN and MTDN are all synthetic samples, due to the problem of domain shift between different datasets; the removal effect of the model for real noise has decreased, so the loss function of DATN mainly solves the difference between the real noise domain and the synthetic noise domain. DATN contains three parallel transformation sub-networks, each of which is composed of generators GS2R, GR2S and discriminators DS, DR. For the real noise image x_{rea} ($x_{rea} \in N_{rea}^i$), the GR2S(x_{rea}) transformed by the generator can be closer to the synthetic noise image x_{syn} ($x_{syn} \in N_{syn}^i$). The adversarial loss can be expressed as:

$$L^{MTDN} = L_{adv}^{MTDN} + \lambda_{sty} L_{sty}^{MTDN} + \lambda_{cyc} L_{cyc}^{MTDN} + \lambda_{mse} L_{mse}^{MTDN} \tag{13}$$

Similarly, the adversarial loss function of x_{syn} to x_{rea} translation can be expressed as:

$$L_{adv}^{DATN}(G_{S2R}, D_R) = \mathbb{E}_{x_{rea}} \| \log D_R(x_{rea}) \|_1 + \mathbb{E}_{x_{syn}} \| \log(1 - D_R(G_{S2R}(x_{syn}))) \|_1 \tag{14}$$

In addition to the same principle as the denoising module, DATN adopts cycle consistency loss to alleviate the mode collapse problem. For the real noisy image x_{rea}, after sequentially passing through GR2S and GS2R, the result should be close to the input image, that is, $x_{rea} \approx$ GR2S (GS2R (x_{rea})), then the cycle consistency loss of DATN can be defined as:

$$L_{adv}^{DATN}(G_{S2R}, D_R) = \mathbb{E}_{x_{rea}} \| \log D_R(x_{rea}) \|_1 + \mathbb{E}_{x_{syn}} \| \log(1 - D_R(G_{S2R}(x_{syn}))) \|_1 \tag{15}$$

In order to further improve the details of the generated image and enhance the texture feature information of the image, we introduce the perceptual loss to measure the distance of the image before and after transformation in the perceptual feature space; it will not only be limited to the pixel space and the feature extraction will be carried out on the two through the convolutional neural network, but the transformed image will be constrained from the high-dimensional space to make it more stylistically the target domain image.

$$L_{pect}^{DATN} = \mathbb{E}_{x_{rea}, x_{syn}} \frac{1}{CHW} [\| \phi(G_{R2S}(x_{rea})) - \phi(x_{syn}) \|_1 + \| \phi(G_{S2R}(x_{syn})) - \phi(x_{rea}) \|_1] \tag{16}$$

Where, C, H, and W represent the feature map's channel number, height and width, respectively; φ is the feature extraction network. In this paper, VGG16 is used as the network basis, and the high-dimensional perceptual information provided by perceptual loss is used to enhance the high-frequency information of the converted image so as to improve the reconstruction effect of image details.

In addition, the dark channel D_c (In) of the image with noise can represent the approximate location of the noise distribution in space [16], so this paper proposes the dark channel consistency loss, which limits the dark channel of the image before and after transformation. The L1 loss is used to ensure the consistency of the dark channels of the two so as to strengthen the network's ability to learn the law of noise distribution. Dark channel consistency loss is defined as:

$$L_{dc}^{DATN} = \mathbb{E}_{x_{rea}, x_{syn}} [\| D_c(G_{R2S}(x_{rea})) - D_c(x_{syn}) \|_1 + \| D_c(G_{S2R}(x_{syn})) - D_c(x_{rea}) \|_1] \tag{17}$$

Where D_c is the dark channel computing process that can be defined as:

$$D_c(I) = \min_{y \in W(x)} [\min_{c \in \{r,g,b\}} I^c(y)] \tag{18}$$

Where x, y represents the coordinates of the pixel point; $I^c(y)$ represents the colour channel of the image I; W(x) represents the sliding window where the pixel point x is located. Therefore, the overall loss of DATN is defined as:

$$L^{DATN} = L_{adv}^{DATN}(G_{R2S}, D_S) + L_{adv}^{DATN}(G_{S2R}, D_R) + L_{cyc}^{DATN}(G_{R2S}, G_{S2R}) + L_{pect}^{DATN} + L_{dc}^{DAYN} \tag{19}$$

In summary, DATN can alleviate the domain shift problem and improve the generalization of models trained on synthetic samples in real scenarios. So far, the design of all loss functions of CWADN has been completed. By training the network to achieve the best convergence of the loss function, the adaptive denoising and information restoration of composite noise in images captured in real scenes can be realized.

IV. IMPLEMENTATION

In order to verify the adaptive denoising and image information restoration capabilities of the designed CWADN, this paper uses the noisy dataset to train and test the network. It compares its computing performance with the current state-of-the-art denoising algorithms to prove the feasibility, accuracy and practicality of CWADN.

A. Dataset

Due to the lack of various types of natural weather noise image datasets so far, and it is impossible to get a real paired dataset, we design and build a synthetic natural noise dataset, namely Campus. The dataset contains 1009 images taken on Campus as noise-free samples. It augments the data through random cropping, inversion, etc., and according to the atmospheric physical model, by applying rain, haze and lower brightness on the noise-free samples. The method completes the construction of synthetic noise datasets. At the same time, in order to verify the self-adaptive denoising capability of CWADN for composite noise, we randomly generate three decimals with a sum of 1 as the proportional weight in the Campus dataset and mix the three noises of rain, haze and low illumination to generate different noises like Synthetic noise datasets with different types of noise and matching different natural environments. Finally, the obtained 2500 images are paired with the set noise type labels {clean, haze, rain, dark, compound}. Part of the paired data is shown in Fig. 6. In addition, we also randomly selected 1000 samples from the real datasets RESIDE, LOL, and SPA to further verify the computing power of the CWADN conversion module.

B. Training Details

The verification platform of this paper is a computer equipped with an NVIDIA GeForce RTX 3080 GPU. The algorithm is written in the Tensorflow framework and uses ADAM as the training optimizer. The batch size is set to 1, and the size of the input image is set to 256×256. The training process adopts two-stage training. In the first stage, MPSN, MTDN and DATN are trained, respectively. We train MPSN firstly, the epoch is set to 100, λ_{mse}=1.0, λ_{ssim}=0.5, and the learning rate is set to 2×10^{-4}; Then train the denoising network MTDN, with a total of 20k iterations, λ_{sty}, λ_{cyc}, λ_{mse} are 1.0, 1.0, 0.8, respectively, and the learning rate decays linearly from 10^{-}

^{4} to 10^{-6}; Finally, the conversion module DATN is trained, and the epoch is set to 150, The learning rate is set to 2×10^{-4}. The second stage uses a small amount of real data sets to fine-tune the network and supervises the results of the current mainstream denoising algorithms as paired real data. In this process, all the models trained in the first stage are imported, and then the transformation and denoising network parameters are frozen. Only the parameters of the MPSN are updated for 50 epochs. In the second training process, the real paired dataset we used is the results from the mainstream denoise algorithms.

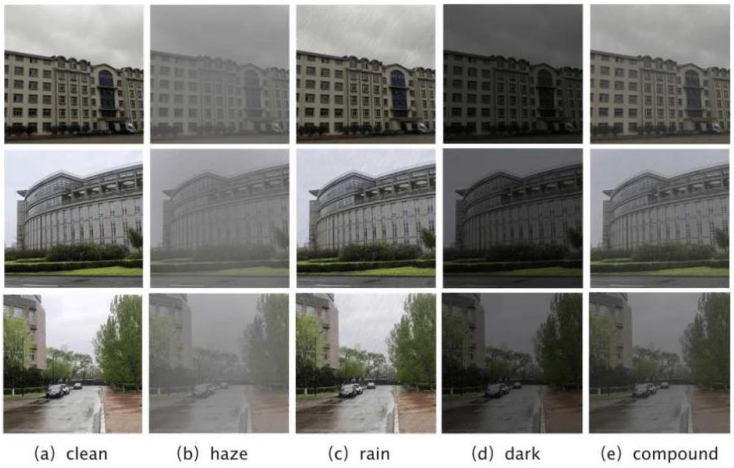

Fig. 6. Samples of the Campus dataset.

V. Results Analysis

A. Qualitative Experiment

In this paper, in the real rain, fog, and low illumination scenarios, CWADN and mainstream algorithms in various fields (refer to DCP [18], Cycle GAN, AOD-Net[19] for dehazing capability; reference RESCAN [20] for dehazing capability, PReNet[21], CycleGAN for deraining capability; low illumination enhancement capability is compared with CycleGAN, EnlightenGAN, and Zero-DCE[22]), and the algorithm designed in this paper will be compared from the perspectives of intuitive visual qualitative and image data quantification respectively. The denoising results and computing power are analyzed and judged.

Fig. 7, 8, and 9, respectively, show each algorithm's intuitive visual solution effects in processing images that are degenerated by real haze, rain, and low illumination. Fig. 7 shows the result of the haze removal test of the real sample, from which it can be seen that after DCP calculates the area where the pixel value is close to the atmospheric light value and the sky area with low contrast, the image after the haze removal will have colour spots and colour shift problems; The image processed by CycleGAN loses more detailed information, and the restored image has low clarity; the image after AOD-Net dehazing is dark as a whole, and the fog noise in some areas is not completely removed; compared with the above algorithm, the saturation and brightness of the image after denoising by CWADN designed in this paper are more natural, and the details of the image are better restored.

Fig. 8 shows the real sample's comparison test of the rain removal effect map. The test image will have rain noise, haze noise and mixed rain and haze noise at the same time. It can be seen from the test results that CycleGAN can filter rain noise and has a certain ability to remove fog noise, but the reconstruction of image texture information is relatively vague; Compared with RESCAN, the rain removal effect of PReNet is significantly improved, but it does not have the ability to remove rain and fog; compared with other algorithms, CWADN has the ability to remove composite noise, it also has a suppressive effect on haze while removing the rain.

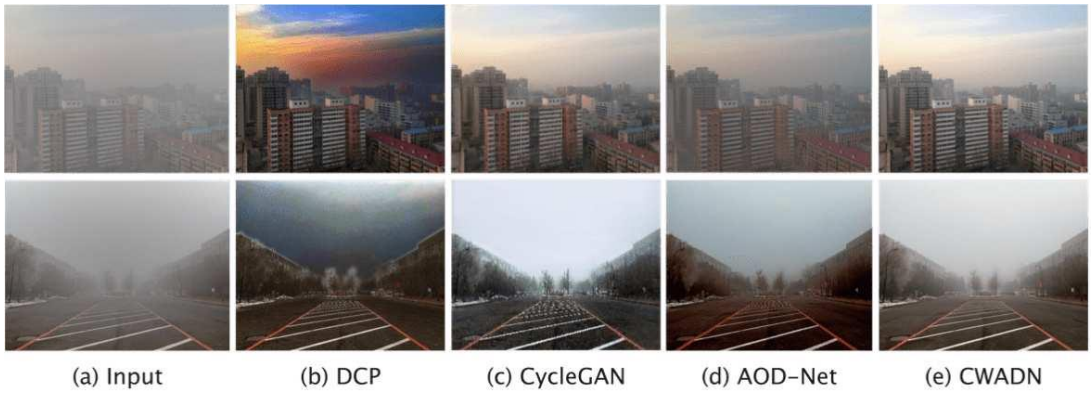

Fig. 7. Comparison of dehazing results on the real-world samples.

Fig. 8. Comparison of deraining results on real-world rain samples.

Fig. 9. Comparison of low-light enhancement results on real-world low-light samples.

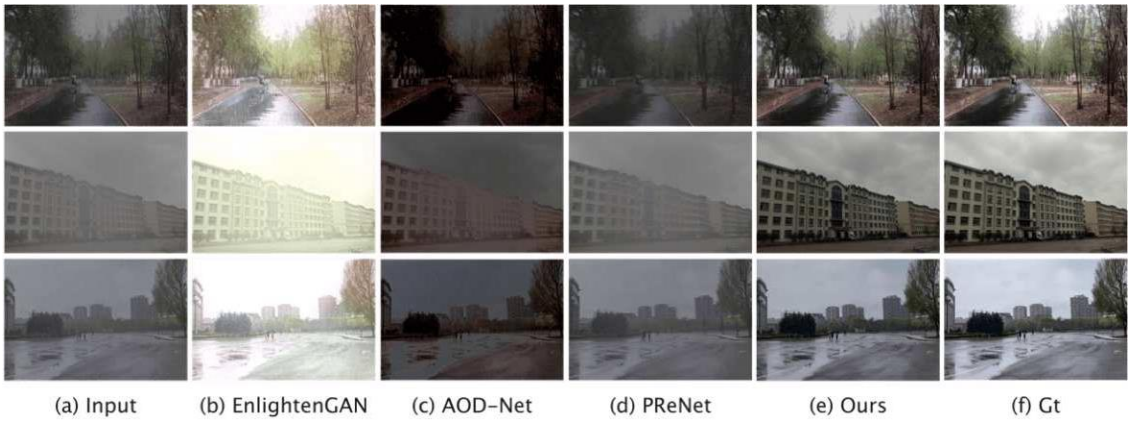

Fig. 10. Comparison of composite noise removal results on Campus dataset.

Fig. 9 shows the low-light enhancement test results under the real sample. It can be seen that the brightness improvement of CycleGAN in some areas is not obvious enough; the saturation of the image enhanced by Zero-DCE is improved, but the overall brightness is low; after EnlightenGAN and CWADN are enhanced. The image achieves good results in both saturation and brightness.

In addition, to further verify the composite noise removal capability of CWADN, we also use the composite noise images in the Campus dataset for testing. In this test, EnlightenGAN, AOD-Net and PReNet were selected for comparison, and the test results are shown in Fig. 10. It can be seen that the image restored by EnligtenGAN is overexposed; the overall brightness of the denoising result of AOD-Net is low, and the removal effect of dense haze is not ideal; PReNet can effectively remove the rain noise in the image, but there is still a lot of haze noise residual. The above three algorithms cannot effectively remove the image's compound noise. We propose that CWADN is closer to the labelled data image regarding overall visual effect and texture detail retention, which verifies CWADN's denoising ability on composite natural environment noise.

To sum up, through testing on real and synthetic natural noise samples, it can be seen that the CWADN proposed in this paper can remove various types of natural environment noise and various types of mixed noise in any proportion, and its noise adaptive denoising ability can reach even exceeding the effect of some mainstream algorithms, the total subjective

image performance verifies the mixed noise adaptive denoising capability of CWADN.

Intuitive visual qualitative analysis can only judge the denoising and image restoration capabilities of the algorithm from the perspective of macroscopic morphology, and this process is only for the denoising task with the human eye as the observation terminal. For advanced machine vision algorithms such as target recognition and tracking, subsequent calculations need to be performed from the pixel-level data direction after image denoising is completed. In order to further verify the denoising and information restoration capabilities of CWADN at the digital image level, we will further carry out quantitative analysis and evaluation of denoised image data on the basis of the above denoising results.

B. Quantitative Test

We use Peak Singal to Noise Ratio (PSNR) and Structural Similarity (SSIM) as evaluation indicators to quantitatively verify CWADN and comparison algorithms on three public datasets of SOTS, Rain100H and LOL. The results are shown in Table I, Table II and Table III, respectively. In addition, this paper also tested the effect of each algorithm in removing composite noise on the Campus dataset, and the results are shown in Table IV.

TABLE I. QUANTITATIVE COMPARISON OF DEHAZING RESULTS ON SOTS

Method	PSNR	SSIM
DCP	15.49	0.64
CycleGAN	14.65	0.48
AOD-Net	19.06	0.85
Our method	19.21	0.84

TABLE II. QUANTITATIVE COMPARISON OF DERAINING RESULTS ON RAIN100H

Method	PSNR	SSIM
CycleGAN	24.22	0.77
RESCAN	26.36	0.79
PReNet	26.77	0.86
Our method	26.72	0.86

TABLE III. QUANTITATIVE COMPARISON OF LOW-LIGHT ENHANCEMENT RESULTS ON LOL

Method	PSNR	SSIM
CycleGAN	7.83	0.15
Zero-DCE	14.86	0.59
EnlightenGAN	17.48	0.68
Our method	16.51	0.63

It can be seen from the results in the table that in the dehazing results on the SOTS dataset, the PSNR of CWADN is 19.21dB, which is the best, and the SSIM index is slightly lower than that of AOD-Net; in the rain removal experiment of Rain100H, CWADN and PReNet. The SSIM achieved the best value of 0.86 among all comparison algorithms; in the low-light enhancement comparison results of LOL, the two indicators of EnlightenGAN achieved the best results, and the two indicators of CWADN were slightly lower than those of EnlightenGAN. In the noise-free results, CWADN achieves the best results in both PSNR and SSIM, with 25.48dB and 0.88, respectively. In Table I, Table II, and Table III, although some indicators of the CWADN designed in this paper are slightly lower than the current mainstream algorithms, the gap between the indicators of CWADN and the mainstream optimal algorithms is small, and the intuitive visual performance and image data calculation are not consistent. No impact. At the same time, Table IV verifies that CWADN has achieved the best performance in the complex noise denoising task, which proves that the algorithm has the ability of all-weather adaptive denoising and is more practical, generalization and denoising than other mainstream algorithms, adaptability.

To sum up, the CWADN proposed in this paper shows good denoising and image restoration capabilities in quantitative experiments, some quantitative detection indicators of denoising exceed the current mainstream algorithms, and the rest performance is on par with the mainstream algorithms. In addition, the experiments on the Campus dataset show that the ability of CWADN to remove mixed noise breaks through the limitation that traditional algorithms are difficult to remove composite noise. Combined with the intuitive visual qualitative analysis results, it is proved that the algorithm in this paper has good noise adaptive filtering and information restoration capabilities in single-type noise directional denoising, multi-type noise denoising and mixed noise denoising tasks, which verifies the performance of the algorithm that is feasibility, practicality and efficiency.

C. Ablation Study

In order to verify the effectiveness of the components added in MPSN and the loss function added in DATN, a series of ablation experiments are established in this paper. Therefore, a series of ablation experiments are established to test different feature extraction strategies and loss function convergence methods. The effects on MPSN and DATN are, respectively, to verify network performance improvement by the method introduced in this paper.

The effectiveness of the components added in the separation network versus the loss function added in the transformation module.

For MPSN, this paper uses UNet as the basis and gradually increases CAB, SAM and multi-stage strategies. The performance impact of different feature extraction strategies on MPSN is shown in Table IV.

It can be seen from Table IV that after the introduction of CAB, SAM and multi-stage strategies in MPSN, the network performance in PSNR and SSIM is increased by 5.5% (1.555) and 0.43% (0.004), respectively, compared with the standard UNet, which proves that the effectiveness of the feature extraction enhancement strategy introduced in this paper.

TABLE IV. ABLATION RESULTS FOR SEPARATION NETWORK

Method	Multi-stage	SAM	PSNR	SSIM
UNet	-	-	28.649	0.936
UNet+CAB	-	-	28.661	0.821
UNet+CAB	√	-	28.685	0.852
UNet+CAB	√	√	30.206	0.94

TABLE V. ABLATION RESULTS FOR TRANSLATION NETWORK DATN

Loss	PSNR	SSIM
Ladv	13.012	0.463
Ladv +Ldc	13.961	0.496
Ladv +Lpect	23.389	0.923
Ladv +Ldc +Lpect	25.632	0.943

For DATN, based on the unsupervised adversarial loss Ladv, this paper adds the perceptual loss Ldc and the dark channel consistency loss Lpect designed in this paper. The test results are shown in Table V.

It can be seen from Table V that the two added loss functions are helpful to the improvement of network performance, and when the two loss functions of Ldc and Lpect are added at the same time, PSNR and SSIM are nearly doubled compared with the original network. To sum up, by introducing different feature extraction strategies and loss functions into MPSN and DATN, this paper effectively enhances the network denoising performance and convergence state and has a significant improvement effect.

Schemes follow another format (see Fig. 2). If there are multiple panels, they should be listed as (a) a Description of what is contained in the first panel; (b) a Description of what is contained in the second panel. Figures should be placed in the main text near the first time they are cited. A caption on a single line should be centered.

VI. CONCLUSIONS

In this paper, the composite weather adaptive denoising network (CWADN) is introduced to address the challenge posed by complex mixed weather interference on natural imaging. CWADN leverages cascaded autoencoders with an attention mechanism to effectively separate distinct noise components within mixed noise. Through the adoption of an image translation strategy, a multi-domain denoising network is constructed, enabling adaptive denoising across various noise types. The adaptive domain network structure narrows the gap between real and synthetic noise, thereby enhancing the model's generalization capability in real-world scenarios. This approach combines both supervised and unsupervised techniques during training, yielding excellent results. Experimental findings demonstrate the algorithm's effectiveness in removing single-type and multi-type random mixed noises in real shooting environments while preserving detailed information in key areas of interest. In comparison to traditional algorithms, CWADN exhibits significant improvements in both PSNR and SSIM metrics, highlighting its robust adaptive denoising and image information restoration capabilities.

VII. LIMITATIONS AND PROSPECTS

While the white-light image enhancement under complex weather conditions is explored in this paper, however，it has been hampered by the limited availability of diverse white-light images captured in varying weather conditions within the same scene. The constrained data volume and scene uniformity have resulted in limited generalization of the training outcomes. Future work aims to address these challenges through extensive data collection efforts, particularly during nighttime, to curate datasets essential for algorithms in this domain.

Furthermore, considering the difficulties associated with training GAN networks and the inherent blurriness and uncertainties in generated images, the intention is to replace them with state-of-the-art diffusion models in subsequent research. Additionally, the introduction of LLM and video comprehension models into the framework is expected to yield superior image generation results, potentially enhancing the quality of generated images and overall performance.

ACKNOWLEDGMENT

This work was supported by the project of Science and Technology on Electromechanical Dynamic Control Laboratory, China, No. 6142601220603.

CONFLICTS OF INTEREST

The authors declared that they have no conflicts of interest.

REFERENCES

[1] Y. Hu, Y. Shang, X. Fu, and H. Ding, “A low illumination video enhancement algorithm based on the atmospheric physical model,” in 2015 8th International Congress on Image and Signal Processing (CISP), IEEE, 2015, pp. 119–124.

[2] W. Chen, Z. Jia, J. Yang, and N. K. Kasabov, “Multispectral Image Enhancement Based on the Dark Channel Prior and Bilateral Fractional Differential Model,” Remote Sens (Basel), vol. 14, no. 1, p. 233, 2022.

[3] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, “Dehazenet: An end-to-end system for single image haze removal,” IEEE Transactions on Image Processing, vol. 25, no. 11, pp. 5187–5198, 2016.

[4] D. Chen et al., “Gated context aggregation network for image dehazing and deraining,” in 2019 IEEE winter conference on applications of computer vision (WACV), IEEE, 2019, pp. 1375–1383.

[5] W. Yang, R. T. Tan, J. Feng, J. Liu, Z. Guo, and S. Yan, “Joint rain detection and removal via iterative region dependent multi-task learning,” CoRR, abs/1609.07769, vol. 2, no. 3, pp. 1–12, 2016.

[6] H. Zhang and V. M. Patel, “Density-aware single image de-raining using a multi-stream dense network,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 695–704.

[7] C. Li, J. Guo, F. Porikli, and Y. Pang, “LightenNet: A convolutional neural network for weakly illuminated image enhancement,” Pattern Recognit Lett, vol. 104, pp. 15–22, 2018.

[8] Y. Jiang et al., “Enlightengan: Deep light enhancement without paired supervision,” IEEE transactions on image processing, vol. 30, pp. 2340–2349, 2021.

[9] Yu Zheng et al., “Curricular Contrastive Regularization for Physics-aware Single Image Dehazing” IEEE Conference on Computer Vision & Pattern Recognition. IEEE, 2023.

[10] R. Li, R. T. Tan, and L.-F. Cheong, “All in one bad weather removal using architectural search,” in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 3175–3185.

[11] S. W. Zamir et al., “Multi-stage progressive image restoration,” in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2021, pp. 14821–14831.

[12] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao, “ Multi-Stage Progressive Image Restoration.” IEEE Conference on Computer Vision & Pattern Recognition. IEEE, 2021.

[13] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, “Image super-resolution using very deep residual channel attention networks,” in Proceedings of the European conference on computer vision (ECCV), 2018, pp. 286–301.

[14] S. Waqas Zamir et al., “Multi-stage progressive image restoration,” arXiv e-prints, p. arXiv-2102, 2021.

[15] Y. Choi, Y. Uh, J. Yoo, and J.-W. Ha, "Stargan v2: Diverse image synthesis for multiple domains," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 8188–8197.

[16] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in Proceedings of the IEEE international conference on computer vision, 2017, pp. 2223–2232.

[17] H. Wang, Z. Yue, Q. Xie, Q. Zhao, Y. Zheng, and D. Meng, "From rain generation to rain removal," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 14791–14801.

[18] Kaiming He et al, "Single image haze removal using dark channel prior," IEEE Conference on Computer Vision & Pattern Recognition. IEEE, 2009.

[19] Boyi Li et al, "An All-in-One Network for Dehazing and Beyond," arXiv e-prints, p. arXiv:1707.06543,2017.

[20] Xia Li et al, "Recurrent Squeeze-and-Excitation Context Aggregation Net for Single Image Deraining," arXiv e-prints, p. arXiv:1807.05698,2018.

[21] Dongwei Ren et al, "Progressive Image Deraining Networks: A Better and Simpler Baseline," IEEE Conference on Computer Vision & Pattern Recognition. IEEE, 2019.

[22] Chunle Guo et al, "Zero-DCE: Zero-Reference Deep Curve Estimation for Low-Light Image Enhancement," IEEE Conference on Computer Vision & Pattern Recognition. IEEE, 2020.