IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 10 June 2023, accepted 19 June 2023, date of publication 26 June 2023, date of current version 17 July 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3289591

RESEARCH ARTICLE

MACGAN: An All-in-One Image Restoration Under Adverse Conditions Using Multidomain Attention-Based Conditional GAN

MARIA SIDDIQUA[1], SAMIR BRAHIM BELHAOUARI[2], NAEEM AKHTER[1], ANEELA ZAMEER[1], AND JAVAID KHURSHID[1]

[1]Department of Computer and Information Sciences, Pakistan Institute of Engineering and Applied Sciences, Nilore, Islamabad 45650, Pakistan
[2]Division of Information and Communication Technologies, College of Science and Engineering, Hamad Bin Khalifa University, Education City, Doha, Qatar

Corresponding author: Samir Brahim Belhaouari (sbelhaouari@hbku.edu.qa)

This work was supported in part by Hamad Bin Khalifa University, and in part by the Qatar National Library.

ABSTRACT Various vision-based tasks suffer from inaccurate navigation and poor performance due to inevitable problems, such as adverse weather conditions like haze, fog, rain, snow, and clouds affecting ground and aerial navigation, as well as underwater images being degraded with blue-green tones and mud affecting marine navigation. Existing techniques in the literature typically focus on restoring specific degradations using separate models, leading to computational inefficiency. To address this, an all-in-one Multidomain Attention-based Conditional Generative Adversarial Network (MACGAN) is proposed to improve scene visibility for optimal ground, aerial, and marine navigation, using the same set of parameters across all domains. The MACGAN is a lightweight network with four encoder and decoder blocks and multiple attention blocks in between, which enhances the image restoration process by focusing on the most important features. To evaluate the effectiveness of MACGAN, extensive qualitative and quantitative comparisons are conducted with state-of-the-art image-to-image translation models, all-in-one adverse weather removal models, and single-effect removal models. The results highlight the superior performance of MACGAN in terms of scene visibility improvement and restoration quality. Additionally, MACGAN is tested on real-world unseen image domains, including smog, dust, fog, rain, snow, and lightning, further validating its generalizability and robustness. Furthermore, an ablation study is conducted to analyze the contributions of the discriminator and attention blocks within the MACGAN architecture. The results confirm that both components play significant roles in the effectiveness of MACGAN, with the discriminator ensuring adversarial training and the attention blocks effectively capturing and enhancing important image features.

INDEX TERMS Restoration, multidomain, adverse weather, navigation, aerial, marine, generative networks, attention mechanism.

I. INTRODUCTION

Various high-level vision-based tasks like object detection [1], [2], recognition, tracking [3], classification, localization, segmentation [4], scene understanding, analysis, manipulation, and many others have a high dependence on clear images for its optimal performance, due to which intensive research in the field of image restoration [5], [6], [7], [8], [9] can be found in the literature. Some naturally occurring causes of image degradations include adverse weather such as haze, fog, rain, snow, cloud, and blue-green color distortion problem in underwater images, the muddy water also hinders the scene visibility.

Weather effects such as fog and mist are usually caused by moisture in the air forming a layer that hinders the scene visibility. The light absorption and scattering of tiny particles in the polluted air cause haze degrading the color and contrast of what we perceive. Fog and smog last longer than haze, and

The associate editor coordinating the review of this manuscript and approving it for publication was Mostafa M. Fouda.

the scene is almost invisible in extreme weather conditions. The haze and fog are modeled as shown in (1):

$$Z = A\left(1 - e^{-\beta d}\right) + Ie^{-\beta d} \tag{1}$$

where Z is the hazy or foggy image, A represents the atmospheric light, β is the atmospheric scattering coefficient, d denotes the distance of an object from the camera, and I represents the effect-free image.

Rain streaks and snowflakes obscure the scene, making it difficult for vision-based applications to perform their tasks efficiently, such as object detection and tracking. The model for rain is shown in (2):

$$Z = I\,(1 - R) + bR \tag{2}$$

where Z is the rainy image, I is the rain-free image, R represents the rain mask, and b denotes the image blur.

Similarly, snow is modeled as shown in (3):

$$Z = I\,(1 - S) + cS \tag{3}$$

where Z is the snowy image, I is the snow-free image, S represents the snow mask, and c denotes the chromatic aberration map.

On the other hand, the images captured underwater are affected by a blue-green tone as water absorbs the red light and scatters the blue light. The dust particles present in water form a brownish layer, reducing the perceptual quality of images and making it challenging to explore the marine world to its full potential.

Autonomous vehicles perceive the scene through the camera but when the environment is degraded by these effects their performance decreases significantly resulting in poor decision-making. When the vehicle is navigating on ground level, haze [10], [11], fog [12], [13], [14], [15], rain [16], [17], and snow [18], [19] are the main causes of its poor performance. In aerial navigation, the drones [20], [21] are affected by fog [22], [23], [24], rain [25], [26], snow [27], [28], and the cloud reduces the scene visibility while the satellite [29] is remotely sensing the ground level information. The robots designed for marine navigation function inefficiently when the scene is affected by the mud and the blue-green color distortion [30], [31], [32], [33]. Vision-based tasks involve working with images; therefore, better image quality and proper scene visibility are crucial for their optimal performance.

In literature, there exist many techniques to remove the image degradations such as model-based in which the effect process is physically modeled, a transmission map is estimated, and then a reverse process of effect is performed to achieve the effect-free image. Model-based methods are outdated and do not provide up-to-the-mark outcomes as their performance highly depends on the correct transmission map estimation which is quite challenging in severely degraded conditions. Recently, with the advent of deep neural networks, the focus of research is shifted toward it. The deep learning-based methods are data-driven in which the data is passed to the deep network, the features are automatically extracted and then the model restores the degraded images. For better image restoration, the attention mechanism is widely used as it puts attention in the form of weights to those features which are crucial for generating better-restored images. Some of the noticeable works based on attention-based deep neural networks are described in the next sections.

A. HAZE REMOVAL

Due to haze, the image contrast and color are highly degraded. Yun et al. [34], [35], [36] developed a nighttime haze removal model by first adding various degradations within the hazy images, divided them into reflectance, illumination, and noise components, then improved the illumination using the prior-based method, improved reflectance contrast in the gradient domain, and finally combined them to get dehazed images. Zhao et al. [37] presented a dehazing model consisting of three parts namely global, local, and adaptive weighted network. In a global network, the image semantic context is captured. The local network preserves the local details of the image by embedding a boosted attention mechanism within it, finally, the global and local networks are adaptively assigned weights based on their significance in dehazing. Yongli et al. [38] incorporated channel and domain attention mechanisms into the cycleGAN model to preserve image contrast and color. The channel attention keeps or rejects channels based on the importance of different features; the channels selected are then passed to the domain attention block where only those feature maps survive where the haze is present. Xiaoqin et al. [39] restored the hazy images by progressively applying attention to different channel scales keeping the trade-off between the low and high-level features to preserve the structure and color. Similarly, Nian et al. [40] extracted local and global features recurrently to enhance feature representation while paying attention to dense hazy regions. Most of the models do not treat spatial and channel attention separately due to which the degradation is not properly restored. This problem was solved by Shibai et al. [41] using spatial and channel attention blocks in parallel to highlight significant feature positions and the correlation of features within channels, respectively. Yitong et al. [42] proposed an unsupervised cycleGAN model embedded with attention-guided modules and total variation loss to limit the noise caused by sea waves for dehazing remote sensing images. Zixuan et al. [43] boosted the dehazing performance of the network by using detail-enhancing convolution and content-guided attention blocks within the encoder block to increase the generalization capacity. Recently, Gang et al. [44] optimized the haze removal process by introducing a UNet model with different level feature fusion based on attention-dense residual learning to preserve local and global features required for better dehazing.

B. FOG REMOVAL

Images degraded by dense fog make the scene hard to visualize, resulting in poor performance of vision-based

tasks such as object detection, tracking, and segmentation. For defogging, Yan et al. [45] introduced a model that first extracts features, then an attention layer is used to keep significant features by assigning more weight to it and rejecting the rest, the atmospheric light and transmission map is computed using an estimating module with unified parameters, after that a restoration module is added which removes the fog based on the atmospheric scattering equation to get the better-defogged results at the end. Qingyi et al. [46] used an attention network in GAN, first the foggy image is passed to the attention network consisting of residual blocks, LSTM, and convolution layers to generate an attention map with the most discriminating features. The attention map is then passed to a convolutional autoencoder for generating a fog-free image. Similarly, the attention map is used in the discriminator as well, to enhance its performance for guessing whether the generated image is real or fake. With the help of an attention map, the most promising features are considered, thereby lowering the computational resources and generating better-defogged images. Zhiqin et al. [47] proposed an autoencoder that first extracts the shape and edge level features using residual octave convolution, then a dual self-attention module is used to enhance the extracted features, and finally, the features are fused with the decoding layers to reconstruct optimal defogged images. Wei et al. [48] developed an unsupervised cyclic GAN to generate fog to fog-free images and vice versa by fusing three separate modules in it namely, fog removal, color texture recovery, and fog synthetic. The fog removal module is an autoencoder that generates fog-free images, the fog-free image is then passed to the color texture recovery module which restores its color and texture based on its multiple derived inputs. To complete the cycle, the fog-free image is converted into a foggy image using a fog synthetic module which adds fog into it using the atmospheric scattering model. Du et al. [49] constructed a global defogging model by combining multiscale convolution and local residual layers along with a feature attention mechanism that consists of channel and pixel attention. Shengmin et al. [50] solved the problem of uneven distribution of fog by using an attention mechanism with an atmospheric scattering model to distinguish thick fog from light fog and assign attention according to the fog density, by doing so, the defogging is handled properly.

C. RAIN REMOVAL

Removing rain is a challenging task as the distribution of rain streaks is not uniform and can severely degrade the scene's visibility. Most of the methods focus on removing the rain streaks ignoring the background recovery. Xiao et al. [51] put forward a model capable of removing the rain streaks by identifying and assigning attention to significant low-level features such as shape and edges, the background is recovered by a network focused on contextual information. Qiang et al. [52] proposed a deraining model which extracts and applies attention to features by finding similarities from coarse to fine levels on multiple scales, finally the rain-free image is constructed. Similarly, Kui et al. [53] explored the feature similarity across multiple scales of rainy images and used the attention mechanisms for feature fusion according to their correlation and weighted attention to get rain-free images. Haitao et al. [54] introduced a multiscale UNet architecture with attention in attention technique which extracts features progressively by applying dynamic weights generated by exploiting the channel and spatial information. To remove rain streaks and restore the background details properly, Yeachan et. al [55] used a UNet architecture with multi-level features of encoder retrieved via channel-wise attention connected to every decoding block recursively and divided the images into wide horizontal patches to explore better rain-free background details. Chih-Yang et al. [56] presented an effective deraining model consisting of three modules connected sequentially namely the residual dense block which extracts the features, the sequential dual attention block which retains the most significant spatial and channel-wise features, and the multiscale feature aggregation module which combines the features for the reconstruction of rain-free images. Likewise, Pengcheng et al. [57] used residual dense blocks for extracting multiscale features which were refined by the attention mechanism and the depth-wise separable convolution of different receptive fields with different scales to reconstruct clean background images.

D. SNOW REMOVAL

Removing snowflakes is essential in many applications like object tracking and surveillance. Ting et al. [58] modified the UNet architecture by combining the self-attention transformer block with residue spatial attention along with residue channel loss to preserve global and local semantic information required for better snow removal. Aiwen et al. [59] removed snow from images by embedding an attention module in GAN which shortlists and encodes significant features required for decoding images without degradation. Sixiang et al. [19] introduced an efficient multi-head cross-attention mechanism that performs local-global context interaction between scale-aware snow queries and local patch embeddings to restore snowy images. Bodong et al. [60] proposed an efficient snow removal model consisting of three modules namely mask-net which has self-pixel and cross-pixel attention to capture significant features and their accurate location to predict the snow mask, guidance-fusion network which adaptively guides the model for snow removal, and finally, the reconstruct-net which removes the snow and generates a snow-free image. Junhong et al. [61] presented a lightweight model that first generates a coarse snow mask, then snow is removed from it using an encoder-decoder. Sixiang et al. [62] designed a desnowing model which understands various snow degradation features in a multi-path manner, a local capture module is connected in parallel to consider local details, and using self-attention the scene context information is integrated for generating a clean image.

E. CLOUD REMOVAL

When sensing information remotely by satellite, the images suffer from poor scene visibility due to the clouds. Peiyang et al. [63] presented a cloud removal method that uses convolution in the initial layers to extract simple features and self-attention with local positional encoding in the later layers to extract complex features at large with different position encodings for different inputs. Hao et al. [64] integrated spatial details in a sentinel-2 clean image with spectral patterns in the sentinel-3 image as spatiotemporal guidance to generate missing regions in the cloudy sentinel-2 image by introducing a spatiotemporal attention network consisting of self-attention mechanism, residual learning, and high-pass filters to enhance feature extraction from the multisource data. Xue et al. [65] combined channel attention with residual learning constructing an encoder-decoder-based residual channel attention network for cloud removal. Meng et al. [66] used an attentive recurrent network to extract feature maps of the cloudy images, then reconstructed the clean image by feeding those feature maps to the attentive residual network. Runhan et al. [67] improved the performance by using ghost convolution in the multi-head self-attention module to eliminate the redundant feature maps decreasing the computational cost. Further, a feature fusion module is used to combine high and low-level features to extract enough features for better cloud removal. Ran et al. [68] removed the cloud by adding a spatial attention mechanism in the unsupervised cycleGAN. Linjian et al. [69] proposed a two-stage generative network for cloud removal, the first stage uses an attention block to get a cloud mask, and the second stage uses an autoencoder conditioned by partial convolution to denoise and in-paint the occluded image patches.

F. UNDERWATER IMAGE ENHANCEMENT

Color distortion is a major problem in marine images, and if the water is murky then the visibility is further degraded making the exploration and exploitation of the marine world difficult. Zhen et al. [70] proposed an encoder-decoder consisting of a channel and shifted pixel-based self-attention transformer for enhancing the local and global texture and color details. Jing et al. [71] used class-conditioned GAN having channel and spatial attention block. The channel attention block aggregates the feature maps across the channels of all layers whereas, the spatial attention block combines the feature maps pixel-wise from the initial to final layers. Yang et al. [72] presented a multi-scale grid CNN that aggregates the multiple kernels with different scales to boost the receptive field for attentive maps, it adaptively focuses on the feature maps of degraded regions to enhance the underwater image patches consistently. Claudio et al. [73] put forward a self-supervised learning model which uses an attention mechanism to focus on saturated areas to avoid color channel imbalances. Shiben et al. [74] utilized the adaptive learning strategy by using three modules namely a multiscale fusion block to combine different spatial details, a parallel attention module focusing on the channel and pixel-wise attention, and finally, an adaptive learning module to dynamically learn the significant features for enhancing the degraded images. To bring diversity in the receptive fields, Xiaohong et al. [75] used an attention-guided dynamic multi-branch network having dynamic feature selection and a multiscale channel module. The features are selected dynamically and then weighted accordingly. Bo et al. [76] introduced a residual two-fold attention network with non-local and channel attention combined to enhance the features required for proper denoising and color correction.

G. ALL-IN-ONE IMAGE RESTORATION

Some noticeable works focusing on the restoration of multiple degradations using a single model can be found in the literature. Ruoteng et al. [77] proposed an all-in-one model for fog, rain, and snow removal using a neural architectural search technique, they designed domain-specific encoders to handle each type of degradation which increased the number of parameters and computational cost. Similarly, Jeya et al. [78] introduced a transformer-based single encoder and decoder model to remove fog, rain, and snow from the images. The encoder consists of intra-patch transformer blocks that work on small patches to remove small degradations. Ashutosh et al. [79] presented a much lighter transformer-based model for the removal of haze, rain, and snow. Their model consists of two parallel modules for restoration: first original resolution transformer for extracting fine-level features from the original size of the input, second multi-level feature aggregation for extracting features of variable sizes, then finally reconstructed the clean images. Recently, Ozan et al. [80] used a denoising diffusion probabilistic model for snow, combined rain and haze, and raindrop removal. First, the noise is added gradually into the images and then the model is designed to denoise the image until the degradation is completely removed. For the removal of haze, rain, and snow, Wei-ting et al. [81] proposed a two-stage knowledge learning-based model that uses student and teacher networks, where the student networks learn the effect removing ability from the teacher networks trained on specific weather effects. All these works perform well but are confined to the removal of few weather effects, contrary to MACGAN which can restore seven different types of degradations using a single model.

In the literature, no model can be found that restores ground and aerial-level weather-degraded images, and color-distorted and mud-affected marine images using a single model. The block diagram of the proposed work can be seen in Fig. 2.

H. CONTRIBUTION

Considering this research gap, an all-in-one Multidomain Attention-based Conditional Generative Adversarial Network (MACGAN) is introduced in this paper, which is capable to improve scene visibility using a single model with the same set of parameters across the adverse domains.

FIGURE 1. Given a degraded image, our all-in-one unified model MACGAN generates its corresponding restored image.

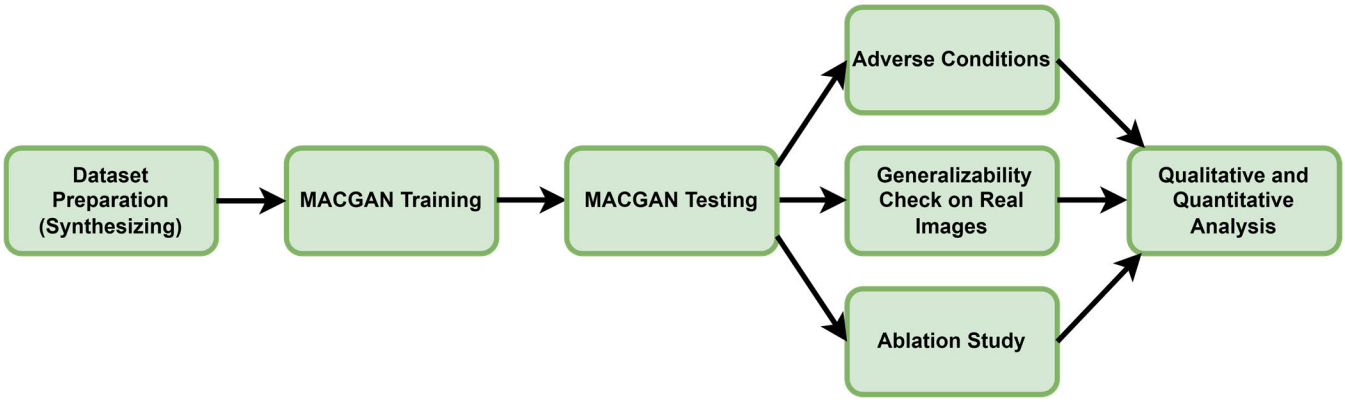


FIGURE 2. Block diagram of the proposed work.

MACGAN is a supervised learning-based conditional autoencoder that takes a degraded image along with its corresponding clean image to generate outcomes similar to the ground truth image with minimal reconstruction error. The addition of three attention blocks between the lightweight four-block encoder-decoder is responsible for generating optimal quality results. The degraded areas are assigned more attention so that the degradation is completely removed with improved scene visibility. Our all-in-one proposed model can restore seven different degraded domains such as haze, fog, rain streaks, snowflakes, thin and thick clouds, underwater and muddy underwater. A synthetic dataset with adverse scenes is created for training and testing MACGAN. The proposed model is quantitatively and qualitatively compared with various state-of-the-art models and tested on a few real unseen domains to check its generalizability and the results proved the success of our model. Moreover, an ablation study is also carried out to check the significance of discriminator and attention blocks in MACGAN architecture.

In short, the main contribution of this paper can be stated as:

1. An all-in-one multidomain restoration model is introduced capable of removing seven adverse degradations from images for facilitating ground, aerial, and marine navigation.
2. Synthetic fog, rain, snow, and cloud effect datasets are prepared for ground and aerial images.
3. Novelty in encoder-decoder architecture with the addition of three attention blocks.
4. Quantitative and qualitative comparison of MACGAN with various image-to-image translation models, all-in-one adverse weather removal models, and single-effect removal models are carried out confirming the optimal performance of MACGAN.
5. Checked the generalizability of MACGAN by testing it on real unseen domains such as smog, dust, fog, rain, snow, and lightning.
6. Carried out ablation study to further validate the outclass performance of the proposed MACGAN architecture with and without discriminator and multiple attention blocks.

The remaining paper is organized as follows: a comprehensive explanation of our proposed architecture is provided in section II. In section III, the detailed quantitative and qualitative experimental results are reported. Section IV includes the ablation study. Finally, the paper is concluded with a brief discussion of the proposed work along with its future scope.

II. METHODOLOGY

The objective of this paper is to restore images degraded under adverse conditions such as haze, fog, rain, snow, cloud, underwater, and muddy underwater for better ground, aerial, and marine navigation as shown in Fig. 1. As the input and

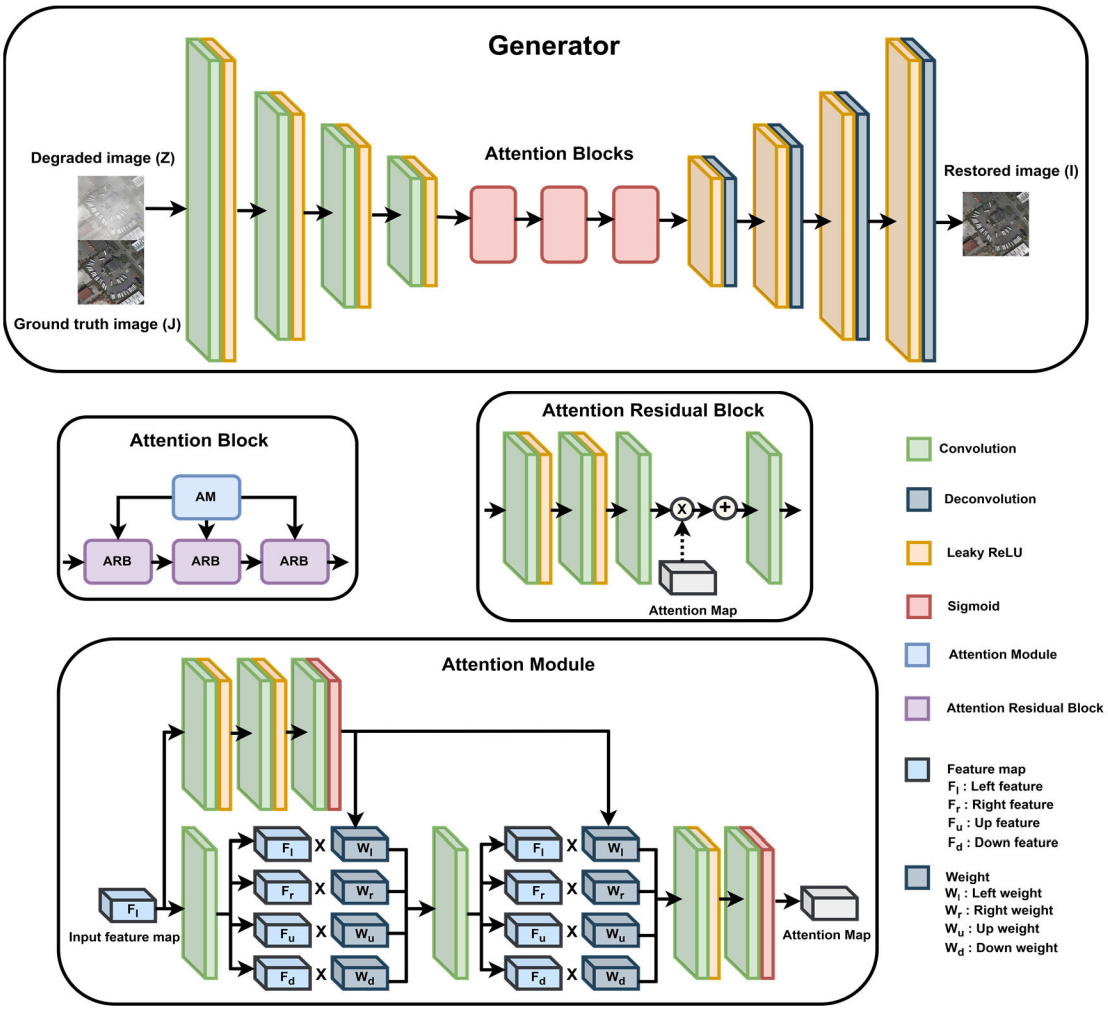


FIGURE 3. Proposed generator architecture with multiple attention blocks.

output both are images, it is considered as an image-to-image translation task, where the degraded images are mapped to the effect-free domain. One such model found in literature is GAN which gets a mapping from the noise vector Z to the final generated image I as $G : Z \rightarrow I$.

$$L_{GAN}(G, D) = E_{x \sim p_{data(x)}}[log(D(x))] + E_{Z \sim p_Z(Z)}[log(1 - D(G(Z)))] \quad (4)$$

Here, x is the real image, p_{data} is the data distribution, $D(x)$ is the discriminator output, p_Z is the distribution of Z, and $G(Z)$ is the generator output.

In this paper, conditional GAN is used that learns a mapping from multiple degraded image domains Z to a single effect-free domain J generating a restored image I as $G : \{Z, J\} \rightarrow I$.

$$L_{cGAN}(G, D) = E_{J,I \sim p_{data}(J,I)}[\log D(J, I)] + E_{J \sim p_{data}(J), Z \sim p_Z(Z)} \times [\log(1 - D(J, G(J, Z)))] \quad (5)$$

The addition of the ground truth image at input along with the degraded image helps the model to generate better results. It learns the mapping irrespective of the degradation due to the constraint imposed on the model. Instead of learning a degradation-specific mapping function, the model learns a general mapping function. Let's say Z is the cloudy satellite image and J is its corresponding clean image, then generator G finds a mapping from effect domain Z to clean domain J, generating the restored effect-free image I. The discriminator D is tasked to guess whether the generated image is effect-free (real) or restored (fake). To fool the discriminator, the generator must output an image similar to the ground truth.

A. PROPOSED MODEL

The two main networks of MACGAN are the generator and discriminator. The generator is a conditional autoencoder that takes both the degraded and its corresponding ground truth images to generate a restored image. The ground truth acts as a constraint on the model so that the generated image is like the ground truth and not a degraded image. The generated images are input to the discriminator, which validates whether the generated image is restored or ground truth. Ideally, the generator output should be similar to the ground truth so that

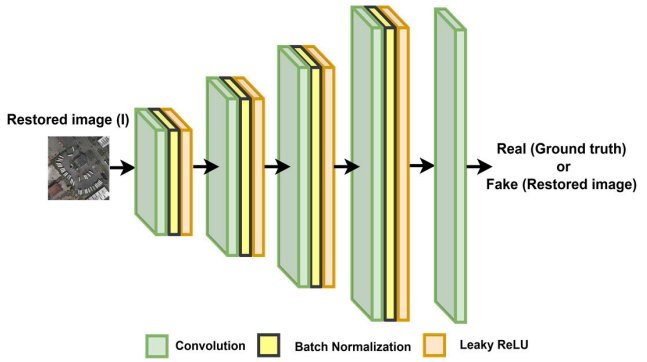


FIGURE 4. Discriminator architecture.

the discriminator gets confused and regards it as an original image with no degradation.

The MACGAN architecture is unified with one set of parameters valid for all input domains. It takes an image with any of the adverse conditions such as haze, fog, rain snow, cloud, underwater, or muddy underwater, and outputs a restored image free of any degradation.

1) GENERATOR

The generator is tasked to generate restored images using an encoder-decoder network with multiple attention blocks in between as shown in Fig. 3. The encoder compresses the image by extracting significant features with each successive block of convolution and activation layer. The compressed feature maps are given to the attention blocks for progressively identifying the degradation. Each attention block consists of three attention residual blocks with an attention module connected in parallel. The attention module generates the feature attention maps in two rounds and four directions (left, right, up, and down). In the first round, the local contextual information of pixels in four directions is encoded and in the second round, the global feature attention map is generated. The feature map generation in four directions results in the complete encoding of contextual and structural details. The attention map generated from the attention module is a two-dimensional matrix of continuous values, where the larger values indicate that more attention in the form of weights should be assigned to that pixel. The attention maps then guide the attention residual blocks in removing the degradation. The degradation is identified and removed progressively using three attention blocks, and then the image is decompressed using the deconvolutional layers until the image is upsized equal to the size of the input image. The addition of three attention blocks increased the reconstruction accuracy significantly. These blocks thoroughly learn the features and context of each input domain concisely by putting attention to the most significant features. As a result, the decoded images are free from each degradation and have minimum reconstruction error. Without attention blocks, the model is unable to properly remove the degradations from the images.

Several attempts were carried out to decide the optimal setting of architecture with and without multiple attention blocks. From the ablation study given in section IV, it is validated that the model generates better results with three attention blocks.

2) DISCRIMINATOR

The discriminator is a simple classifier consisting of convolutional, batch normalization, and leaky Rectified Linear Unit (ReLU) layers. The restored images generated by the generator are input to the discriminator which classifies whether the input image is real (ground truth image) or fake (restored image) as shown in Fig. 4. There is an adversary between the generator and the discriminator, each trying to outperform the other. The generator tries to fool the discriminator by generating an image similar to its corresponding ground truth image, and the discriminator is tasked to guess whether the generated image is real or fake. Better restored images are generated when the discriminator calls the restored image real. The discriminator's feedback in the form of a guess to the generator is crucial for the optimal restoration of the degraded images as validated by the ablation study in section IV.

The overall flow of the work is shown in Fig. 5.

3) LOSS FUNCTION

The loss function of MACGAN consists of three loss functions as shown in (6):

$$L_{MACGAN} = \arg\min_{G}\max_{D} L_{cGAN}(G, D) + L_{L1}(G) + L_{Att} \tag{6}$$

First is the conditional GAN loss as shown in (7):

$$\begin{aligned}\min_{G}\max_{D} L_{cGAN}(G, D) &= E_{J,I\sim p_{data}(J,I)}[\log D(J, I)] \\ &+ E_{J\sim p_{data}(J),Z\sim p_Z(Z)}[\log(1 - D(J, G(J, Z)))]\end{aligned} \tag{7}$$

Here, generator G and discriminator D play a min-max game, where G tries to minimize the above function and D maximizes it creating an adversary.

Second, is the L_1 loss which minimizes the difference between the ground truth and the restored image as shown in (8):

$$L_{L1}(G) = \frac{1}{CHW}\sum_{c=1}^{C}\sum_{v=1}^{H}\sum_{u=1}^{W}\lambda_C\left|I_{gt}^{(u,v,c)} - G(I_{in})^{(u,v,c)}\right|_1 \tag{8}$$

Here, C, W, and H represent the channel, width, and height respectively. λ_C is the weight of each channel's contribution to the loss, it is set to 1. I_{gt} is the ground truth image and $G(I_{in})$ is the restored image generated by the generator.

Third is the attention loss which minimizes the difference between the attention map and the binary image of the degradation as shown in (9):

$$L_{Att} = \|A - B\|_2^2 \tag{9}$$

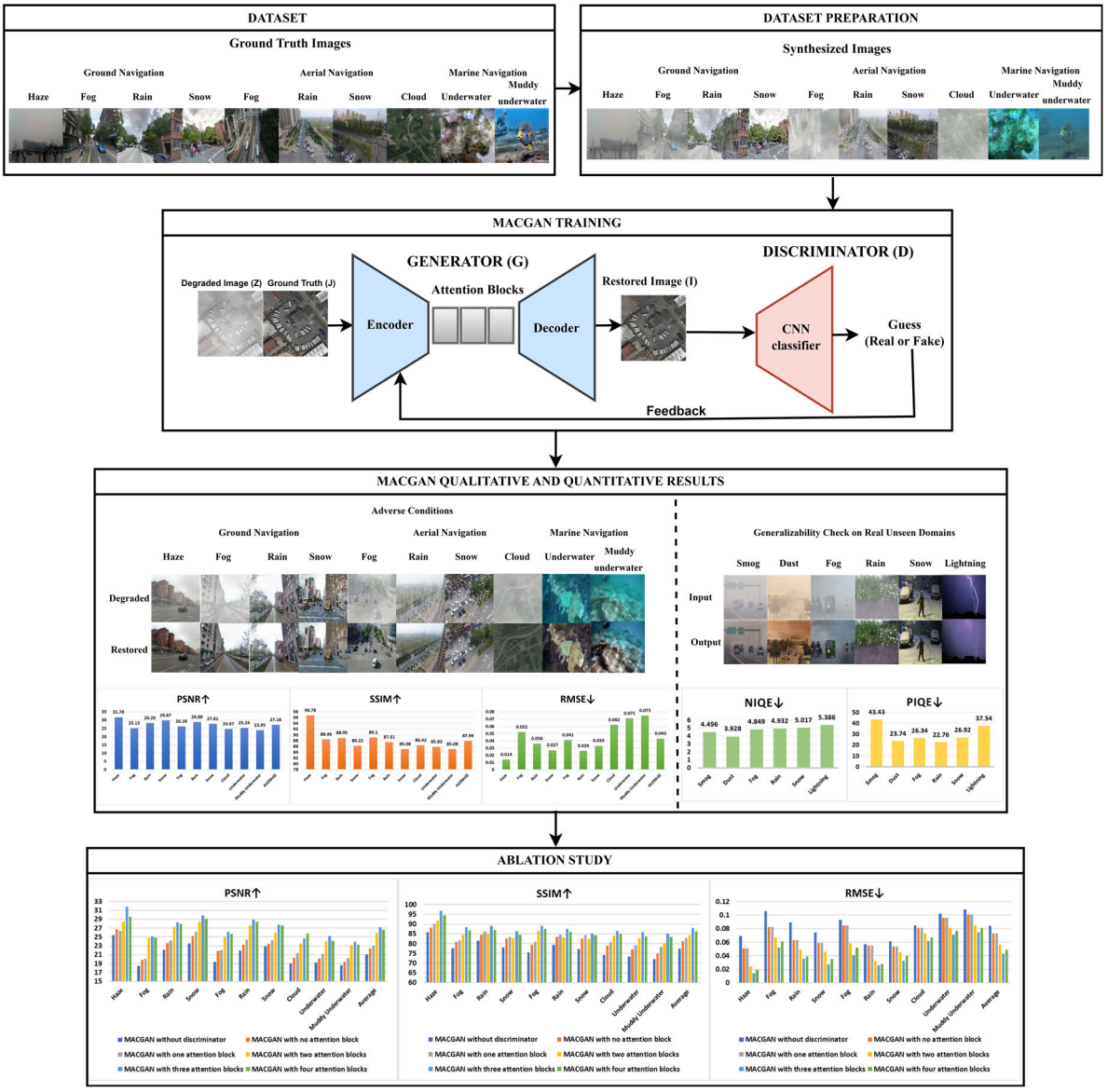


FIGURE 5. The overall flow of the work.

Here, A is the attention map generated by the attention module and B is the binary image calculated by the difference between the affected and effect-free image.

The standard conditional GAN loss function along with L_1 and attention loss combines the individual loss functions for each domain into a single objective function allowing the model to learn to perform well on all domains at the same time, rather than optimizing each domain separately. This helped in improving the overall performance and efficiency of the model by minimizing the difference between the restored and its corresponding ground truth images. This unified loss function is appropriate as the proposed model is unified and capable of restoring multiple degraded domains with the same set of parameters.

III. EXPERIMENTS

Detailed experiments are carried out for each input domain with five state-of-the-art image-to-image translation models: UNIT [82], pix2pix [83], cycle transformer [84], cycleGAN [85], CUT [86]; three all-in-one adverse weather removal models: UMWTransformer [79], Wei-Ting et al. [81], Transweather [78]; and six single-effect removal models: FFA-Net [87], Yeying et al. [88], MPRNet [89], DDMSNet [27], GLF-CR [90], and DGD-cGAN [91]. Their quantitative and qualitative results are provided in subsections C and D, respectively.

A. DATASET

In this work, six different image datasets as mentioned in Table 1 are used for ten input domains. The dataset consists of

TABLE 1. Dataset details.

Navigation	Domain	Dataset	$Train_A$	$Test_A$	$Train_B$	$Test_B$
Ground	Haze	Reside OTS [92]	1200	300	1200	300
	Fog	Google Street View [93]	1200	300	1200	300
	Rain		1200	300	1200	300
	Snow		1200	300	1200	300
Aerial	Fog	VisDrone [94]	1200	300	1200	300
	Rain		1200	300	1200	300
	Snow		1200	300	1200	300
	Cloud	DOTA [95]	1200	300	1200	300
Marine	Underwater	EUVP [96]	1200	300	1200	300
	Muddy Underwater	Raw890 [97]	680	210	680	210

28, 780 images divided equally into degraded domain A and clean domain B. The dataset is split into approximately 80 and 20 percent ratios for training and test images, respectively.

MACGAN and pix2pix are supervised models that utilize paired images from domain A only, whereas, the other models are unsupervised that are trained using both unpaired domains A and B. For a fair comparison, all the models are trained and tested on the same dataset.

Boyi et al. [92] introduced a uniform random haze generation process by adding atmospheric light within the range of [0.7, 1.0], along with a scattering coefficient between [0.6, 1.8]. To synthesize fog, rain, snow, and cloud effects in ground and aerial images, an image augmentation library [98] was utilized to randomly apply these effects. Fog images were created with a mean intensity range of [190, 255] and density range of [1, 0.1]. Cloud images were synthesized with densities ranging from [0.5, 1.5]. Rain streaks were added at various angles, with raindrop sizes ranging from (0.2, 0.5) and speeds between (0.01, 0.30). Snowflakes were added at different angles within the range of [-30, 30], with sizes varying between (0.70, 0.95) and speeds ranging from (0.001, 0.020). Md et al. [96] employed seven different cameras to capture underwater images, preparing unpaired images based on human perceptual preference and generating paired images using the cycleGAN model. Chongyi et al. [97] collected underwater images from open sources and self-captured videos, applying several underwater image enhancement methods to obtain paired images. All images were resized to 256×256 pixels.

B. TRAINING DETAILS

All the models are implemented using the PyTorch framework with GeForce GTX 1070 Ti GPU. The optimal hyperparameters for training the models are selected using grid search. Adam algorithm was used with an initial learning rate of 0.0004, momentum terms set to 0.5 and 0.999, weight decay 0.00001, and batch size 1. The batch size is set to 1 so that the weights are updated after each instance as multiple images from adverse domains are present in the dataset. All the models are trained up to 100 epochs until convergence.

C. QUANTITATIVE EVALUATION

For the quantitative evaluation of MACGAN, three performance metrics are used: peak signal-to-noise ratio (PSNR), structural similarity index measure (SSIM), and root mean square error (RMSE).

A PSNR computes the ratio to which the restored image is prone to get corrupted as shown in (10). A higher score indicates that the restored images are accurate having minimal noise.

$$PSNR = 10log_{10}\left(\frac{(2^n-1)^2}{MSE}\right) \tag{10}$$

where n is the number of bits which is 8 for grayscale images, and MSE is the mean squared error trying to minimize the difference between the generated restored image I and the ground truth image J as shown in (11):

$$MSE = \frac{1}{H*W}\sum_{x=1}^{H}\sum_{y=1}^{W}(I(x,y)-J(x,y))^2 \tag{11}$$

where H and W are the height and width of the images.

SSIM is a similarity comparison between the restored and its corresponding ground truth images. When both images are maximally similar in structure, a high SSIM score is achieved. SSIM is a product of brightness l, contrast c, and structure s as shown in (12):

$$SSIM = l(I,J).c(I,J).s(I,J) \tag{12}$$

$$l(I,J) = \frac{2\mu_I\mu_J + C_1}{\mu_I^3 + \mu_J^2 + C_1} \tag{13}$$

$$c(I,J) = \frac{2\sigma_I\sigma_J + C_2}{\sigma_I^2 + \sigma_J^2 + C_2} \tag{14}$$

$$s(I,J) = \frac{\sigma_{IJ} + C_3}{\sigma_I\sigma_J + C_3} \tag{15}$$

where μ and σ are the mean and variance, and σ_{IJ} is the covariance of the restored and its corresponding ground truth image. C_1, C_2, and C_3 are non-zero numerical constants to avoid divide by zero error.

RMSE measures the root mean squared difference between the restored image and its corresponding ground truth image as shown in (16), as it is an error measure, so a lower score is desirable.

$$RMSE = \sqrt{\frac{\sum_{k=1}^{N}(I(k)-J(k))^2}{N}} \tag{16}$$

The PSNR, SSIM, and RMSE values of MACGAN are compared with various image-to-image translation models in Table 2.

TABLE 2. MACGAN quantitative comparison of PSNR↑, SSIM ↑, and RMSE ↓ values with various image-to-image translation models.

	Domain	MACGAN (*ours*)			UNIT [82]			Pix2Pix [83]			Cycle Transformer [84]			CycleGAN [85]			CUT [86]		
		PSNR↑	*SSIM*↑	*RMSE*↓	*PSNR*↑	*SSIM*↑	*RMSE*↓	*PSNR*↑	*SSIM*↑	*RMSE*↓	*PSNR*↑	*SSIM*↑	*RMSE*↓	*PSNR*↑	*SSIM*↑	*RMSE*↓	*PSNR*↑	*SSIM*↑	*RMSE*↓
GROUND	*Haze*	31.78	96.78	0.014	29.40	93.57	0.027	29.17	92.78	0.034	27.09	89.52	0.046	25.39	86.97	0.060	24.64	84.23	0.081
	Fog	25.13	88.46	0.052	22.20	79.33	0.085	21.32	78.09	0.092	20.54	76.46	0.102	18.76	76.13	0.129	17.63	73.92	0.137
	Rain	28.29	88.95	0.036	25.59	77.87	0.056	25.29	76.95	0.055	24.37	76.35	0.063	23.82	76.30	0.068	21.99	74.60	0.082
	Snow	29.87	86.22	0.027	26.87	84.22	0.047	24.78	82.31	0.067	24.29	79.71	0.059	24.12	79.08	0.065	23.78	78.40	0.070
AERIAL	*Fog*	26.18	89.10	0.041	19.18	76.27	0.116	19.08	76.11	0.119	18.55	73.10	0.124	17.99	71.79	0.140	17.58	69.70	0.157
	Rain	28.88	87.51	0.026	23.88	77.56	0.066	23.82	74.51	0.068	23.35	73.63	0.069	22.51	73.56	0.079	21.38	72.38	0.088
	Snow	27.81	85.08	0.033	25.81	80.08	0.053	24.62	79.44	0.064	23.44	78.09	0.068	22.17	73.68	0.081	22.60	72.60	0.089
	Cloud	24.67	86.42	0.062	19.83	71.43	0.110	19.67	70.42	0.112	18.36	71.13	0.135	18.03	69.24	0.146	17.48	67.60	0.152
MARINE	*Underwater*	25.24	85.83	0.071	21.24	75.83	0.091	20.28	78.83	0.103	19.91	72.79	0.112	19.12	71.44	0.118	18.40	70.39	0.125
	Muddy Underwater	23.95	85.08	0.075	22.95	81.08	0.0759	20.85	81.56	0.095	20.10	75.24	0.101	19.48	73.30	0.102	18.34	72.09	0.127
	Average	**27.18**	**87.94**	**0.043**	23.69	79.72	0.073	22.88	79.10	0.081	22.003	76.60	0.088	21.14	75.15	0.099	20.38	73.59	0.111

TABLE 3. MACGAN quantitative comparison of PSNR↑, SSIM↑, and RMSE↓ values with various all-in-one adverse weather removal models.

	Domain	MACGAN (*ours*)			UMWTransformer [79]			Wei-Ting et al. [81]			Transweather [78]		
		PSNR↑	*SSIM*↑	*RMSE*↓	*PSNR*↑	*SSIM*↑	*RMSE*↓	*PSNR*↑	*SSIM*↑	*RMSE*↓	*PSNR*↑	*SSIM*↑	*RMSE*↓
GROUND	*Haze*	31.78	96.78	0.014	29.98	90.16	0.042	27.31	87.11	0.074	25.34	86.43	0.085
	Fog	25.13	88.46	0.052	22.72	81.41	0.072	20.35	78.47	0.109	19.92	77.39	0.122
	Rain	28.29	88.95	0.036	24.87	82.52	0.061	22.87	80.04	0.092	21.74	79.86	0.096
	Snow	29.87	86.22	0.027	26.69	81.72	0.059	24.52	78.21	0.083	24.2	77.72	0.089
AERIAL	*Fog*	26.18	89.1	0.041	21.95	80.26	0.084	19.98	77.62	0.118	18.11	76.45	0.13
	Rain	28.88	87.51	0.026	24.54	81.98	0.057	22.67	80.56	0.087	21.48	79.22	0.098
	Snow	27.81	85.08	0.033	26.31	81.87	0.061	23.01	79.64	0.089	22.39	78.17	0.092
	Cloud	24.67	86.42	0.062	22.83	81.97	0.075	20.29	79.02	0.117	19.52	78.67	0.147
MARINE	*Underwater*	25.24	85.83	0.071	23.81	81.05	0.101	20.34	76.22	0.126	19.93	75.95	0.133
	Muddy Underwater	23.95	85.08	0.075	21.7	79.96	0.099	20.06	73.11	0.135	19.67	71.87	0.141
	Average	**27.18**	**87.94**	**0.043**	24.54	82.29	0.071	22.14	79.00	0.103	21.23	78.17	0.113

UNIT [82] is an unsupervised image-to-image translation model that utilizes a shared latent space based on coupled GANs. Pix2pix [83] is a supervised image-to-image translation model based on conditional GAN. CycleGAN [85] is an unsupervised cyclic image-to-image translation model comprising of two generators and two discriminators. Cycle transformer [84] is a transformer-based CycleGAN [85]. CUT [86] is an unsupervised contrastive learning-based image-to-image translation model.

In comparison to other domains, the haze effect does not significantly hinder scene visibility, resulting in optimal haze metrics as demonstrated in Table 2. Our model considers both thin and thick foggy and cloudy images for restoration, achieving an average PSNR of 25, which outperforms other models. Our model also exhibits the second-best metrics for rain and snow. Additionally, it successfully addresses the blue-green color distortion issue in underwater images, attaining decent scores. Table 2 highlights that all other models underperform as they lack the attention modules required for effective image restoration. In contrast, our proposed model, MACGAN, exhibits superior and consistent scores across all ten input domains in comparison to other image-to-image translation models.

To evaluate the performance of MACGAN in comparison to other unified weather removal models, a comparative study is conducted, and their quantitative values are presented in Table 3.

TABLE 4. MACGAN quantitative comparison of PSNR↑, SSIM↑, and RMSE↓ values with various single-effect removal models.

Domain	*Model*	*PSNR↑*	*SSIM↑*	*RMSE↓*
Haze	FFA-Net [87]	24.56	87.32	0.057
	MACGAN (*ours*)	**31.78**	**96.78**	**0.014**
Fog	Yeying et al. [88]	20.66	79.74	0.105
	MACGAN (*ours*)	**25.13**	**88.46**	**0.052**
Rain	MPRNet [89]	23.02	82.51	0.079
	MACGAN (*ours*)	**28.29**	**88.95**	**0.036**
Snow	DDMSNet [27]	24.25	83.98	0.067
	MACGAN (*ours*)	**29.87**	**86.22**	**0.027**
Cloud	GLF-CR [90]	21.82	81.99	0.091
	MACGAN (*ours*)	**24.67**	**86.42**	**0.062**
Underwater	DGD-cGAN [91]	21.40	69.40	0.115
	MACGAN (*ours*)	**23.95**	**85.08**	**0.075**

From Table 3, it is evident that all models perform better in the haze domain compared to other domains. However, when it comes to the fog and cloud domain, the other models exhibit poor performance as they struggle to restore scene details in the presence of thick fog and cloud degradation. In terms of rain scores, UMWTransformer [79] and Wei-Ting et al. [81] achieve higher scores than Transweather [78]. PSNR values for the snow domain surpass those of the rain domain. Transweather [78] fails to deliver satisfactory results in underwater image restoration, as it struggles to enhance contrast and remove the muddy effect. Table 2 demonstrates that MACGAN outperforms other all-in-one weather removal models, with higher scores reported across various domains.

UMWTransformer [79] and Transweather [78] are computationally intensive transformer-based models. On the other hand, Wei-Ting et al. [81] adopts a student-teacher knowledge transfer architecture. Our model, MACGAN, is lightweight, comprising only four encoder and decoder blocks with three attention blocks in between. The performance of MACGAN is enhanced by the constraint imposed through ground truth images and the inclusion of attention mechanisms.

Lastly, the performance of MACGAN is compared to various single-effect removal models for haze, fog, rain, snow, cloud, and underwater domains, and their quantitative results are presented in Table 4.

FFA-Net [87] is a feature fusion attention network designed for haze removal, which combines channel and pixel features. Yeying et al. [88] introduced a structure representation network with uncertainty feedback learning for fog removal. MPRNet [89] is a multi-stage progressive model specifically developed for restoring degraded rain images. DDMSNet [27] is a deep dense multi-scale network that utilizes semantic and depth priors for snow removal. GLF-CR [90] is a global and local feature fusion network tailored for cloud removal. DGD-cGAN [91] consists of a dual generator-based conditional GAN model for improving degraded underwater images.

Analyzing Table 4, it is evident that FFA-Net [87] performs significantly worse than MACGAN in the haze domain. Regarding fog removal, the model proposed by Yeying et al. [88] struggles to properly remove fog, as illustrated in Fig. 10. MACGAN, on the other hand, successfully eliminates rain streaks and achieves a higher score compared to MPRNet [89]. Although DDMSNet [27] improves scene visibility, there are still a few remaining snowflakes visible in Fig. 10. GLF-CR [90] effectively removes thin clouds but underperforms when it comes to thick cloud removal. DGD-cGAN [91] manages to correct the blue-green tone, but it generates blurry images, resulting in a lower SSIM value. Instead of relying on a single-effect removal model, employing MACGAN for image restoration proves to be a better choice as it demonstrates optimal performance across various degraded domains and can handle multiple types of adverse degradations using a single model.

To visualize the performance metrics, the values reported in Tables 2, 3, and 4 are illustrated in the form of column charts, as depicted in Fig. 6.

From Fig. 6, it can be seen that MACGAN has optimal performance as compared to the various image-to-image translation, unified, and single-effect removal models proving its up-to-the-mark restoration performance quantitatively.

D. QUALITATIVE EVALUATION

The qualitative results of MACGAN for all ten input domains, each showcasing input, output, and ground truth images, are displayed in Fig. 7.

MACGAN demonstrates remarkable success in removing degradations across all ten input domains, as evident from Fig. 7. In the case of haze, MACGAN effectively enhances the color and contrast of the images, resulting in clear scenes. Fog removal leads to improved visibility, while rain streaks are completely eliminated. Snowflakes are not visible in the output images, indicating successful snow removal. Thick and thin clouds in satellite images are effectively removed, allowing for easy visualization of the scene. MACGAN also corrects the color of underwater images, eliminating the murky effect and improving overall scene visibility. These results demonstrate the proficiency of MACGAN in handling various degradations, establishing its qualitative success.

For comparative study, MACGAN is first compared qualitatively with various image-to-image translation models as shown in Fig. 8.

It is evident from Fig. 8 that MACGAN outperforms other models in restoring various degradations, producing results that closely resemble the ground truth images. In the case of haze, MACGAN achieves a better contrast that closely matches the ground truth. Other models struggle to completely remove fog, leaving remnants in their results. Rain streaks are visible in the outputs of cycleGAN [85] and CUT [86] models, indicating their failure in effectively eliminating rain. Similarly, other models fail to adequately inpaint regions occluded by snowflakes, resulting in their visibility in the outputs. MACGAN, however, successfully

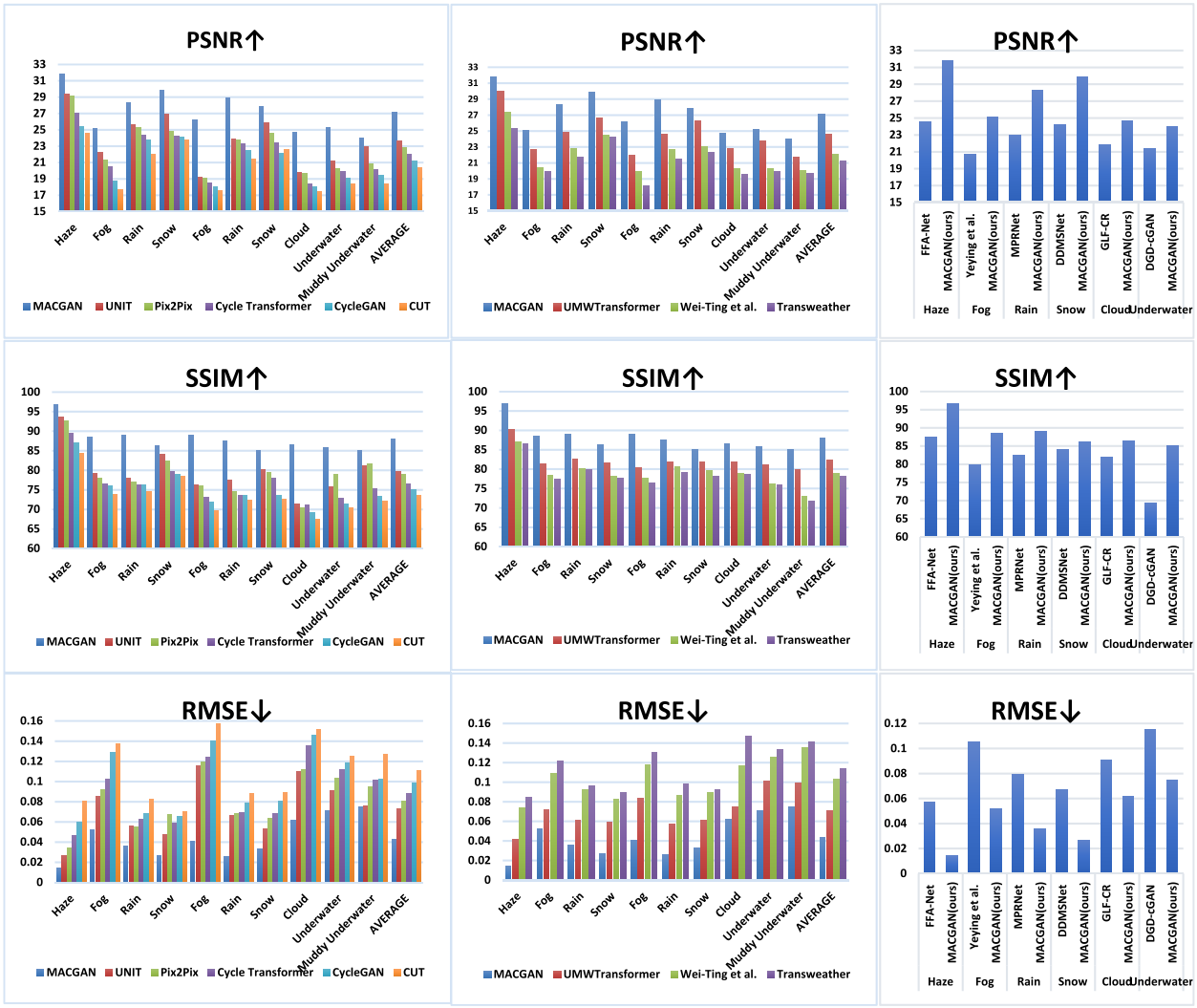


FIGURE 6. Quantitative comparison of PSNR↑, SSIM↑, and RMSE↓ values of various models using column charts.

removes clouds from satellite images and corrects the blue-green tone in underwater images, exhibiting scene clarity similar to the ground truth.

UNIT [82], pix2pix [83], cycle transformer [84], cycle-GAN [85], and CUT [86] models are trained until they reach minimum error convergence, yet they still struggle to restore images due to their lack of context encoding and effective feature learning ability. These limitations are overcome by MACGAN, which incorporates multiple attention blocks focused on learning local and global features for each degradation at precise pixel locations in all four directions. This enables MACGAN to encode the best features and generate superior effect-free images with minimal reconstruction error. The attention blocks play a vital role in MACGAN's optimal performance, while models like pix2pix [83], lacking an attention mechanism, fail to improve visibility across different input domains. The proposed MACGAN model's effectiveness lies in its ability to restore multiple degradations using a single model.

Further, MACGAN is compared qualitatively with various all-in-one adverse weather removal models as shown in Fig. 9.

From Fig. 9, it is evident that UMWTransformer [79] and Wei-Ting et al. [81] successfully remove the hazy effect from the images, while Transweather [78] still retains some patches of haze in certain regions. In the case of fog, UMWTransformer [79] effectively restores the fog, while the results of Wei-Ting et al. [81] and Transweather [78] appear distorted. Transweather [78] also exhibits visible rain streaks in its output. Wei-Ting et al. [81] and Transweather [78] models struggle to completely remove snowflakes. Thick clouds remain in the images generated by Transweather [78], and even UMWTransformer [79] and Wei-Ting et al. [81] show regions where restoration is not entirely successful. UMWTransformer [79] manages to remove the blue-green tone and enhance the contrast of muddy underwater images to some extent, but the outputs are still not comparable to the ground truth. There are minor changes in the

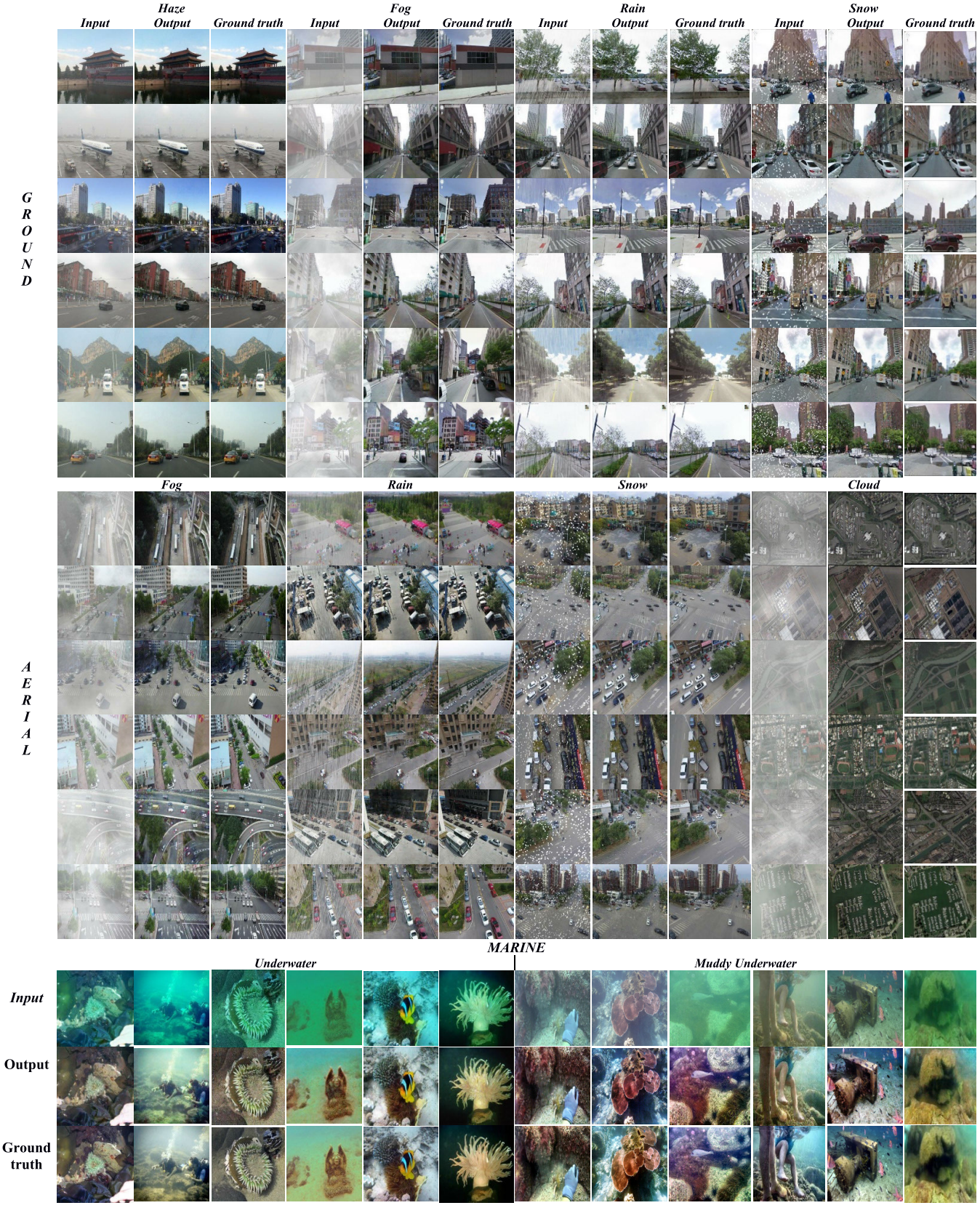


FIGURE 7. MACGAN qualitative results across ten adverse domains.

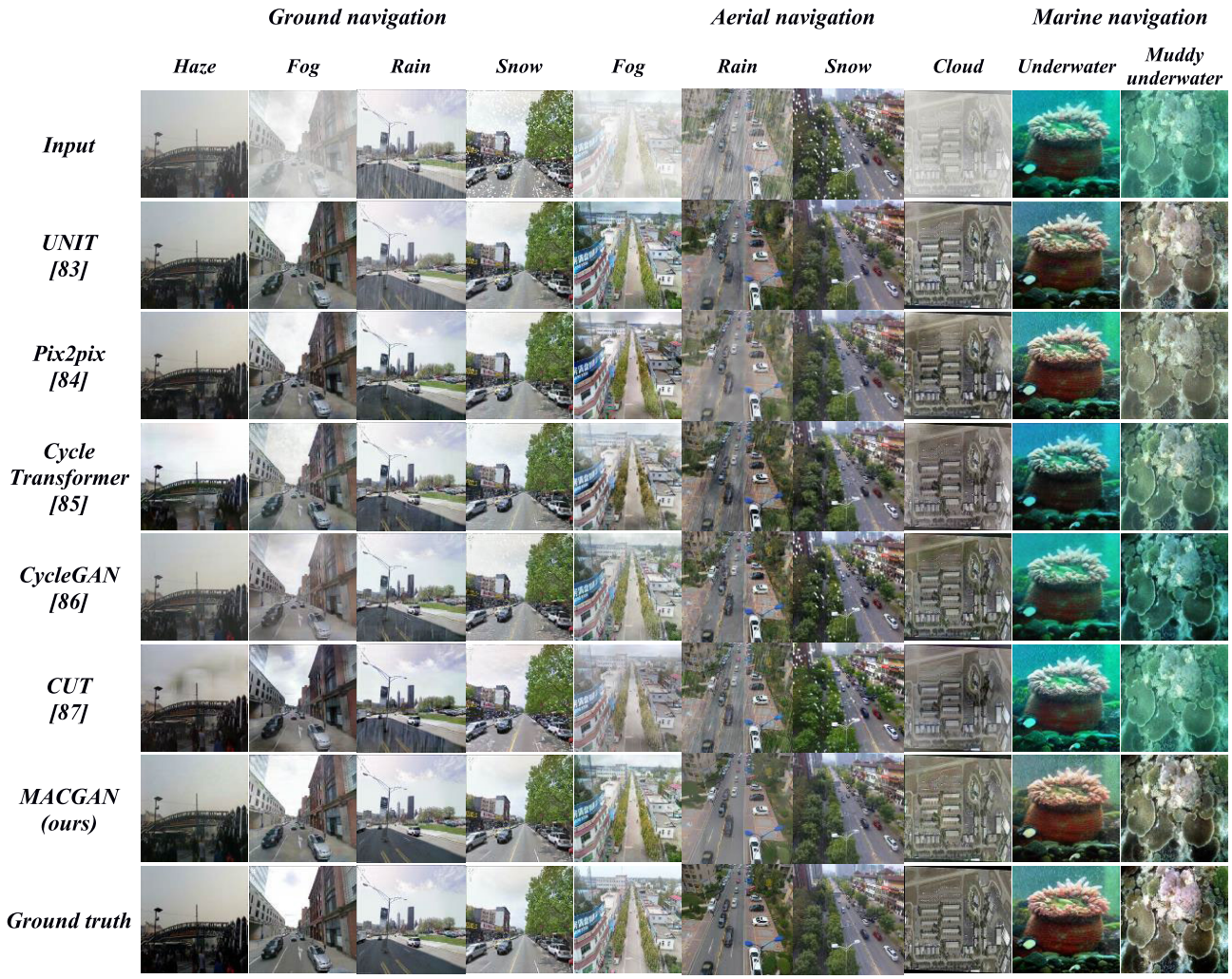


FIGURE 8. MACGAN qualitative comparison with various image-to-image translation models.

color and contrast of underwater images in the outputs of Wei-Ting et al. [81].

In comparison, MACGAN excels in restoring images across all domains. It successfully removes haze while retaining the contrast similar to the ground truth. MACGAN eliminates the distortion caused by fog, ensures rain streaks are not visible in the output images, completely removes snowflakes, properly restores images with thick cloud layers, corrects the blue-green color distortion in underwater images, eliminates the muddy effect, and enhances contrast. All outputs generated by MACGAN are optimally restored and closely resemble the ground truth images. Therefore, it can be concluded that MACGAN demonstrates the highest competence in removing multiple types of degradations, unlike other models that may perform well only in one or two specific domains.

Image features can be extracted locally and globally, MACGAN uses two rounds for attention map generation, the first round is utilized to capture local features in four directions (left, right, up, down); the second round captures the contextual, semantic, and textural features globally in four directions resulting in the full encoding of significant features, adding more rounds and eight directions (up-right, down-left, etc) might improve the feature encoding performance of our model but would make the model computationally heavy with longer training time. To keep a trade-off between the model complexity and accuracy, we proposed a lightweight model having only four encoding and decoding blocks with three attention blocks between them, serving our purpose of image restoration for multiple degraded domains using a single model.

Finally, the performance of MACGAN is qualitatively compared with various single-effect removal models as shown in Fig. 10.

In Fig. 10, it can be observed that MACGAN performs optimally when the degradation is uniform, as seen in the case of haze. Even as the thickness of the degradation increases in fog, MACGAN is still able to restore the images effectively. MACGAN successfully removes heavy rain streaks and numerous snowflakes, producing

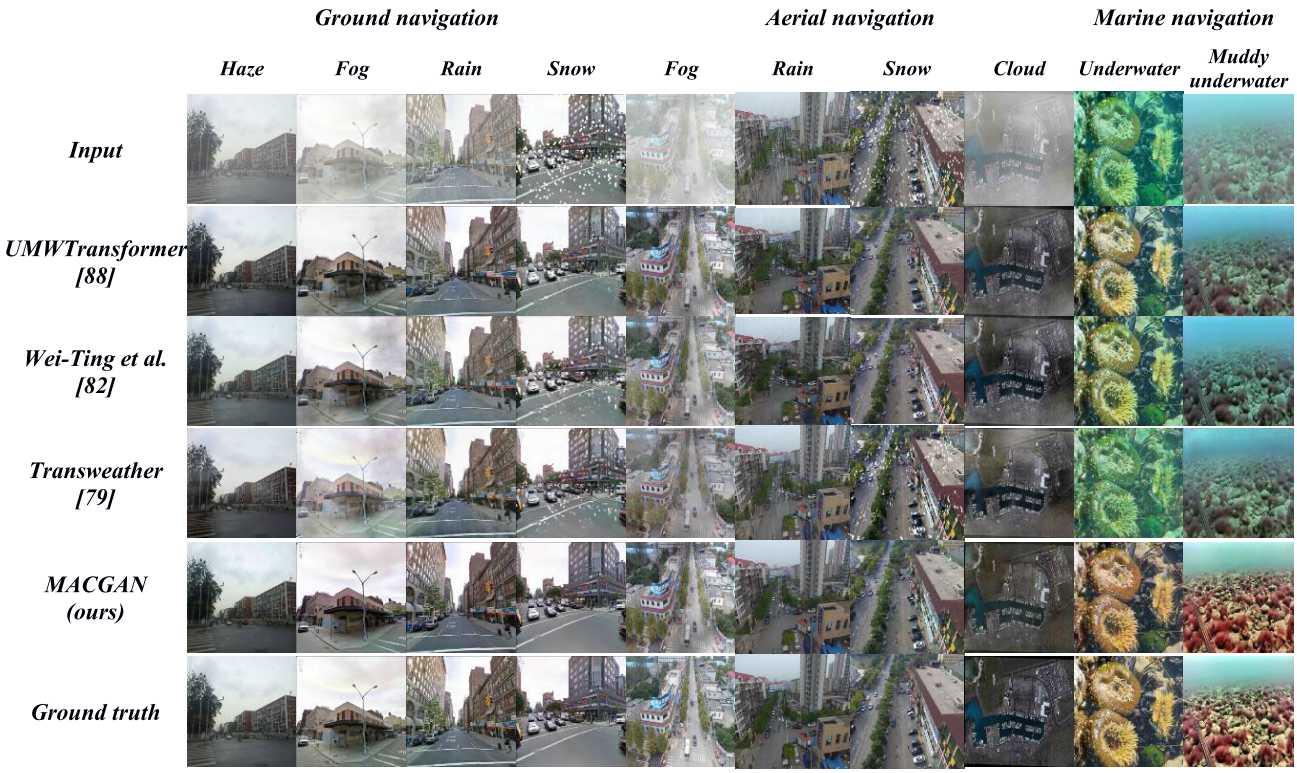


FIGURE 9. MACGAN qualitative comparison with various all-in-one adverse weather removal models.

clean, effect-free images. It demonstrates efficient removal of both thin and thick clouds. MACGAN significantly improves the visibility of underwater images by correcting the blue-green color distortion and eliminating the muddy effect. On the other hand, the outputs of FFA-Net [87], Yeying et al. [88], MPRNet [89], DDMSNet [27], GLF-CR [90], and DGD-cGAN [91] are not optimal, most of them failed to restore the degradation properly, whereas, MACGAN outputs showed up-to-the-mark restoration performance. MACGAN is a better choice for image restoration as it not only restores the single-effect degraded images but multiple degraded images as well using a single model.

To further ensure the effectiveness of MACGAN, its generalizability is tested on real unseen domains such as smog, dust, fog, rain, snow, and lightning. The 204 dust, 300 fog, 200 rain, and 323 snow images are taken from DAWN dataset [99]. Few smog and lightning images are taken from open sources. The qualitative outputs of MACGAN on real test images are shown in Fig. 11. For each input domain, the first row shows the input image, and the next row shows its corresponding output image generated by MACGAN.

Figure 11 reveals better-restored results on various real unseen domains. Although smog, dust, and lightning domains were not provided while training MACGAN, it still performed well in removing them. The contrast and visibility of fog and smog images are enhanced, the brownish dusty layer, snowflakes, lightning, and rain streaks are removed, thereby improving the scene clarity.

TABLE 5. NIQE↓ and PIQE↓ values of real test images.

Domain	*NIQE*↓	*PIQE*↓
Smog	4.496	39.27
Dust	3.928	23.74
Fog	4.849	26.34
Rain	4.932	22.76
Snow	5.017	26.92
Lightning	5.386	37.54

For quantitative evaluation of real images, the natural image quality evaluator (NIQE) and perception-based image quality evaluator (PIQE) are calculated as shown in Table 5.

According to the scores reported in Table 5, the naturalness and perception-based image quality are better for dust images.

IV. ABLATION STUDY

To understand the significance of discriminator and attention blocks in improving the overall performance of our proposed model MACGAN, an ablation study is carried out in which the model is trained without a discriminator network with L_1 and attention loss. By removing the discriminator, we are left with a simple autoencoder. Variational autoencoders

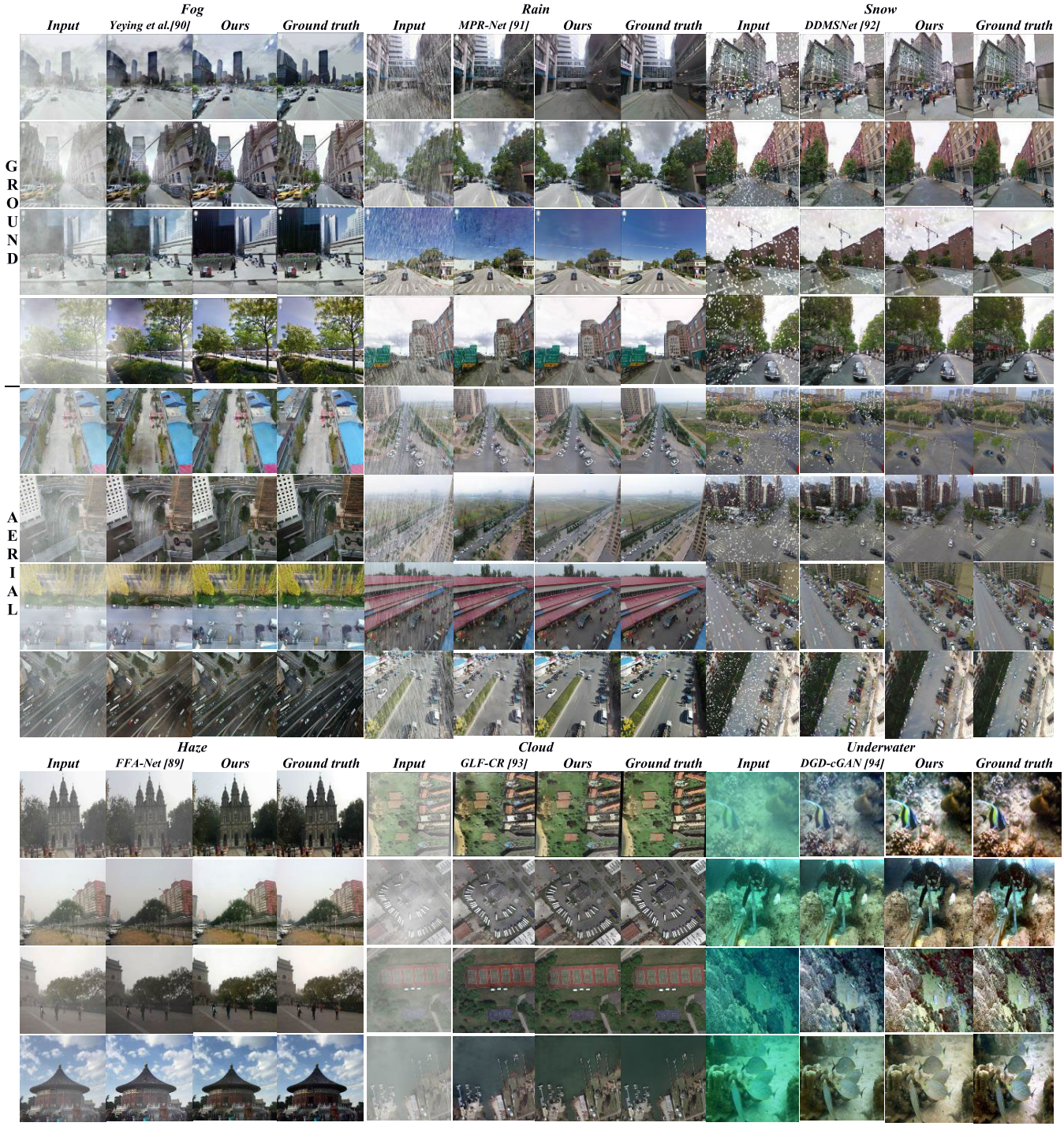


FIGURE 10. MACGAN qualitative comparison with various single-effect removal models.

were used before the advent of GANs but they had a main drawback in that their outputs are blurred and lacked significant image details.

Moreover, the optimal number of attention blocks is selected by training and testing MACGAN with and without multiple attention blocks. Without an attention block, MACGAN was trained using conditional GAN and L_1 loss. Further, the number of attention blocks was progressively increased for MACGAN performance evaluation. The quantitative analysis of the ablation study is given in Table 6.

From Table 6, it can be noticed that the performance metrics for MACGAN without discriminator are significantly low. The main reason to add a discriminator is to improve the performance of the generator, as they both work in adversary,

FIGURE 11. MACGAN qualitative outputs on various real unseen domains.

the discriminator tries to differentiate the generated image from the ground truth image, and the generator tries to fool the discriminator by generating a fake (restored) image similar to the ground truth image. The feedback in the form of a

TABLE 6. Quantitative analysis of ablation study.

	Domain	MACGAN without discriminator			MACGAN with no attention block			MACGAN with one attention block			MACGAN with two attention blocks			MACGAN with three attention blocks			MACGAN with four attention blocks		
		PSNR↑	*SSIM*↑	*RMSE*↓	*PSNR*↑	*SSIM*↑	*RMSE*↓	*PSNR*↑	*SSIM*↑	*RMSE*↓	*PSNR*↑	*SSIM*↑	*RMSE*↓	*PSNR*↑	*SSIM*↑	*RMSE*↓	*PSNR*↑	*SSIM*↑	*RMSE*↓
GROUND	*Haze*	25.39	85.82	0.069	26.72	88.23	0.051	26.34	90.13	0.049	28.46	91.85	0.024	31.78	96.78	0.014	29.56	94.32	0.019
	Fog	18.47	77.68	0.106	19.87	80.92	0.082	20.01	81.77	0.071	24.82	84.53	0.067	25.13	88.46	0.052	24.89	86.67	0.061
	Rain	22.11	81.54	0.089	23.63	84.47	0.063	24.23	86.18	0.052	27.34	85.24	0.049	28.29	88.95	0.036	27.93	86.79	0.039
	Snow	23.54	78.12	0.074	25.21	82.54	0.059	26.18	83.35	0.051	28.43	82.77	0.046	29.87	86.22	0.027	29.11	84.56	0.035
AERIAL	*Fog*	19.37	75.41	0.093	21.83	79.29	0.085	22.06	80.79	0.079	25.07	86.45	0.058	26.18	89.1	0.041	25.72	87.48	0.052
	Rain	21.96	79.43	0.057	23.29	83.38	0.055	24.36	84.45	0.047	27.58	83.34	0.032	28.88	87.51	0.026	28.47	85.95	0.028
	Snow	22.83	77.27	0.061	23.47	82.56	0.054	24.29	84.41	0.056	26.01	82.29	0.045	27.81	85.08	0.033	27.54	84.23	0.04
	Cloud	19.02	74.09	0.085	20.28	78.95	0.081	21.32	80.58	0.079	23.54	84.11	0.073	24.67	86.42	0.062	25.79	85.01	0.067
MARINE	*Underwater*	19.21	73.26	0.102	20.11	76.95	0.096	21.17	79.05	0.085	23.98	82.72	0.081	25.24	85.83	0.071	24.06	83.67	0.077
	Muddy Underwater	18.59	71.95	0.108	19.36	74.79	0.101	20.21	78.26	0.091	23.16	80.28	0.085	23.95	85.08	0.075	23.27	83.39	0.081
	Average	21.05	77.46	0.084	22.37	81.21	0.073	23.02	82.89	0.066	25.84	84.36	0.056	**27.18**	**87.94**	**0.043**	26.63	86.21	0.049

guess by the discriminator helps the generator to perform well, without the discriminator the results are not up-to-the-mark.

Further, the importance of attention blocks is studied by training MACGAN with and without multiple attention blocks. It is observed that when no attention block has been used the score for all performance metrics is low. When one attention block was added, the performance increased, therefore, the number of attention blocks was increased and tested till the optimal accuracy was achieved. When the fourth attention block was added, the performance started to decrease. After extensively testing the MACGAN variants, it is concluded that the PSNR, SSIM, and RMSE values are optimal when MACGAN is trained with a discriminator and the generator has three attention blocks.

V. CONCLUSION

The paper proposes an all-in-one, lightweight multidomain encoder-decoder architecture with multiple attention blocks to restore degraded scene visibility under adverse weather and underwater conditions. The addition of three attention blocks is found to enhance the performance of MACGAN, as demonstrated by the ablation study. MACGAN's performance is evaluated against various state-of-the-art image-to-image translation models, all-in-one adverse weather removal models, and single-effect removal models, with a detailed comparative analysis validating its performance both qualitatively and quantitatively, with an average of 27.18 PSNR, 87.94 SSIM, and 0.043 RMSE. Furthermore, MACGAN demonstrates exceptional generalizability to other real-world unseen image domains, making it a universal restoration model. The ablation study confirms the optimality of the proposed architectural design. In conclusion, MACGAN's image restoration capabilities can improve ground, aerial, and marine navigation, surveillance, traffic sign and lane mark detection, and other vision-based areas for better scene visualization.

ACKNOWLEDGMENT

The authors would like to thank the Data Science Laboratory, Department of Computer and Information Sciences, Pakistan Institute of Engineering and Applied Sciences, for providing an exceptional learning environment with advanced computational machines to carry out up-to-date research.

REFERENCES

[1] W. Peng, Y. Peng, W. Lien, and C. Chen, "Unveiling of how image restoration contributes to underwater object detection," in *Proc. IEEE Int. Conf. Consum. Electron.-Taiwan (ICCE-TW)*, Penghu, Taiwan, Sep. 2021, pp. 1–2, doi: 10.1109/ICCE-TW52618.2021.9602998.

[2] Y. Wang, X. Yan, K. Zhang, L. Gong, H. Xie, F. L. Wang, and M. Wei, "TogetherNet: Bridging image restoration and object detection together via dynamic enhancement learning," 2022, *arXiv:2209.01373*.

[3] D. Lee, G. Kim, D. Kim, H. Myung, and H.-T. Choi, "Vision-based object detection and tracking for autonomous navigation of underwater robots," *Ocean Eng.*, vol. 48, pp. 59–68, Jul. 2012.

[4] P. Drews-Jr, I. D. Souza, I. P. Maurell, E. V. Protas, and S. S. C. Botelho, "Underwater image segmentation in the wild using deep learning," *J. Brazilian Comput. Soc.*, vol. 27, no. 1, pp. 1–14, Dec. 2021.

[5] J. Su, B. Xu, and H. Yin, "A survey of deep learning approaches to image restoration," *Neurocomputing*, vol. 487, pp. 46–65, May 2022.

[6] M. Yang, J. Hu, C. Li, G. Rohde, Y. Du, and K. Hu, "An in-depth survey of underwater image enhancement and restoration," *IEEE Access*, vol. 7, pp. 123638–123657, 2019.

[7] A. M. Ali, B. Benjdira, A. Koubaa, W. El-Shafai, Z. Khan, and W. Boulila, "Vision transformers in image restoration: A survey," *Sensors*, vol. 23, no. 5, p. 2385, Feb. 2023.

[8] Z. Zhang, Y. Wei, H. Zhang, Y. Yang, S. Yan, and M. Wang, "Data-driven single image deraining: A comprehensive review and new perspectives," *Pattern Recognit.*, vol. 143, Nov. 2023, Art. no. 109740.

[9] T.-Y. Jia, L. Zhuo, J.-F. Li, and J. Zhang, "Research advances on deep learning based single image dehazing," *Acta Electonica Sinica*, vol. 51, no. 1, p. 231, 2023.

[10] N. Jiang, K. Hu, T. Zhang, W. Chen, Y. Xu, and T. Zhao, "Deep hybrid model for single image dehazing and detail refinement," *Pattern Recognit.*, vol. 136, Apr. 2023, Art. no. 109227.
[11] A. S. Parihar and A. Java, "Densely connected convolutional transformer for single image dehazing," *J. Vis. Commun. Image Represent.*, vol. 90, Feb. 2023, Art. no. 103722.
[12] H. Wu, Y. Qu, S. Lin, J. Zhou, R. Qiao, Z. Zhang, Y. Xie, and L. Ma, "Contrastive learning for compact single image dehazing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 10546–10555.
[13] H. Dong, J. Pan, L. Xiang, Z. Hu, X. Zhang, F. Wang, and M.-H. Yang, "Multi-scale boosted dehazing network with dense feature fusion," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Mar. 2020, pp. 2157–2167.
[14] C. Deqiang, Y. Yangyang, K. Qiqi, and X. Jinyang, "A generative adversarial network incorporating dark channel prior loss used for single image defogging," *Opto-Electron. Eng.*, vol. 49, no. 7, 2022, Art. no. 210448.
[15] J. Zhou, D. Zhang, P. Zou, W. Zhang, and W. Zhang, "Retinex-based Laplacian pyramid method for image defogging," *IEEE Access*, vol. 7, pp. 122459–122472, 2019.
[16] S. J. Lakshmi, R. Bhuvaneshwari, and P. Geetha, "TA-DNN-two stage attention based deep neural network for single image rain removal," *Signal, Image Video Process.*, vol. 17, pp. 3163–3171, 2023, doi: 10.1007/s11760-023-02538-7.
[17] Z. Fan, H. Wu, X. Fu, Y. Huang, and X. Ding, "Residual-guide network for single image deraining," in *Proc. 26th ACM Int. Conf. Multimedia*, Oct. 2018, pp. 1751–1759.
[18] W.-T. Chen, H.-Y. Fang, C.-L. Hsieh, C.-C. Tsai, I. Chen, J.-J. Ding, and S.-Y. Kuo, "ALL snow removed: Single image desnowing algorithm using hierarchical dual-tree complex wavelet representation and contradict channel loss," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 4176–4185.
[19] S. Chen, T. Ye, Y. Liu, E. Chen, J. Shi, and J. Zhou, "SnowFormer: Context interaction transformer with scale-awareness for single image desnowing," 2022, *arXiv:2208.09703*.
[20] E. Pahwa, A. Luthra, and P. Narang, "LVRNet: Lightweight image restoration for aerial images under low visibility," 2023, *arXiv:2301.05434*.
[21] N. Aswini and S. V. Uma, "Drone image de-noising and feature extraction," in *Proc. IEEE Int. Conf. Innov. Technol. (INOCON)*, Nov. 2020, pp. 1–6.
[22] F. Xia, H. Song, and H. Dou, "Fog removal and enhancement method for UAV aerial images based on dark channel prior," *J. Control Decis.*, vol. 10, no. 2, pp. 188–197, Apr. 2023.
[23] A. Kulkarni and S. Murala, "Aerial image dehazing with attentive deformable transformers," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2023, pp. 6294–6303.
[24] X. Sun, Y. Han, Y. Chen, Y. Duan, Y. Wang, and B. Su, "Aerial image defogging method based on nonlocal feature structure tensor by UAV cameras with three-channel RGB cameras," *J. Appl. Remote Sens.*, vol. 16, no. 4, Dec. 2022, Art. no. 046515.
[25] X. Fu, J. Huang, X. Ding, Y. Liao, and J. Paisley, "Clearing the skies: A deep network architecture for single-image rain removal," *IEEE Trans. Image Process.*, vol. 26, no. 6, pp. 2944–2956, Jun. 2017.
[26] B. Das, A. Saha, and S. Mukhopadhyay, "Rain removal from a single image using refined inception ResNet v2," *Circuits, Syst., Signal Process.*, vol. 42, no. 6, pp. 3485–3508, Jun. 2023.
[27] K. Zhang, R. Li, Y. Yu, W. Luo, and C. Li, "Deep dense multi-scale network for snow removal using semantic and depth priors," *IEEE Trans. Image Process.*, vol. 30, pp. 7419–7431, 2021.
[28] P. Li, M. Yun, J. Tian, Y. Tang, G. Wang, and C. Wu, "Stacked dense networks for single-image snow removal," *Neurocomputing*, vol. 367, pp. 152–163, Nov. 2019.
[29] B. Marhaba, "Satellite image restoration by nonlinear statistical filtering techniques," Theses, Dept. Autom. Control Eng., Université du Littoral Côte d'Opale, 2018. [Online]. Available: https://theses.hal.science/tel-02094292
[30] M. Zheng and W. Luo, "Underwater image enhancement using improved CNN based defogging," *Electronics*, vol. 11, no. 1, p. 150, Jan. 2022.
[31] C. O. Ancuti, C. Ancuti, C. De Vleeschouwer, and P. Bekaert, "Color balance and fusion for underwater image enhancement," *IEEE Trans. Image Process.*, vol. 27, no. 1, pp. 379–393, Jan. 2018.
[32] W. Zhang, P. Zhuang, H. Sun, G. Li, S. Kwong, and C. Li, "Underwater image enhancement via minimal color loss and locally adaptive contrast enhancement," *IEEE Trans. Image Process.*, vol. 31, pp. 3997–4010, 2022.
[33] W. Zhang, Y. Wang, and C. Li, "Underwater image enhancement by attenuated color channel correction and detail preserved contrast enhancement," *IEEE J. Ocean. Eng.*, vol. 47, no. 3, pp. 718–735, Jul. 2022.
[34] Y. Liu, Z. Yan, J. Tan, and Y. Li, "Multi-purpose oriented single nighttime image haze removal based on unified variational retinex model," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 33, no. 4, pp. 1643–1657, Apr. 2023, doi: 10.1109/TCSVT.2022.3214430.
[35] Y. Liu, Z. Yan, T. Ye, A. Wu, and Y. Li, "Single nighttime image dehazing based on unified variational decomposition model and multi-scale contrast enhancement," *Eng. Appl. Artif. Intell.*, vol. 116, Nov. 2022, Art. no. 105373.
[36] Y. Liu, Z. Yan, A. Wu, T. Ye, and Y. Li, "Nighttime image dehazing based on variational decomposition model," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2022, pp. 639–648.
[37] Z. Wang, F. Li, R. Cong, H. Bai, and Y. Zhao, "Adaptive feature fusion network based on boosted attention mechanism for single image dehazing," *Multimedia Tools Appl.*, vol. 81, no. 8, pp. 11325–11339, Mar. 2022.
[38] Y. Ma, J. Xu, F. Jia, W. Yan, Z. Liu, and M. Ni, "Single image dehazing using generative adversarial networks based on an attention mechanism," *IET Image Process.*, vol. 16, no. 7, pp. 1897–1907, May 2022.
[39] X. Zhang, T. Wang, J. Wang, G. Tang, and L. Zhao, "Pyramid channel-based feature attention network for image dehazing," *Comput. Vis. Image Understand.*, vols. 197–198, Aug. 2020, Art. no. 103003.
[40] S. Yin, Y. Wang, and Y.-H. Yang, "A novel image-dehazing network with a parallel attention block," *Pattern Recognit.*, vol. 102, Jun. 2020, Art. no. 107255.
[41] N. Wang, Z. Cui, Y. Su, and A. Li, "RGNAM: Recurrent grid network with an attention mechanism for single-image dehazing," *J. Electron. Imag.*, vol. 30, no. 3, Jun. 2021, Art. no. 033026.
[42] Y. Zheng, J. Su, S. Zhang, M. Tao, and L. Wang, "Dehaze-AGGAN: Unpaired remote sensing image dehazing using enhanced attention-guide generative adversarial networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5630413.
[43] Z. Chen, Z. He, and Z.-M. Lu, "DEA-Net: Single image dehazing based on detail-enhanced convolution and content-guided attention," 2023, *arXiv:2301.04805*.
[44] G. Jin, J. Zhai, and J. Wei, "CAA-Net: End-to-end two-branch feature attention network for single image dehazing," *IEICE Trans. Fundamentals Electron., Commun. Comput. Sci.*, vol. E106, no. 1, pp. 1–10, Jan. 2023.
[45] Y. Yang, C. Zhang, P. Jiang, and H. Yue, "Attention-based end-to-end image defogging network," *Electron. Lett.*, vol. 56, no. 15, pp. 759–761, 2020.
[46] Q. Zhang, C. Zhao, X. Zhang, F. Yuan, C. Li, and D. Hao, "The generative adversarial network based on attention mechanism for image defogging," in *Proc. 17th Int. Forum Digital TV Wireless Multimedia Commun. (IFTC)*, vol. 17, Singapore: Springer, 2021, pp. 12–25.
[47] Z. Zhu, Y. Luo, G. Qi, J. Meng, Y. Li, and N. Mazur, "Remote sensing image defogging networks based on dual self-attention boost residual octave convolution," *Remote Sens.*, vol. 13, no. 16, p. 3104, Aug. 2021.
[48] W. Liu, C. Chen, R. Jiang, T. Lu, and Z. Xiong, "Unpaired quad-path cycle consistent adversarial networks for single image defogging," 2022, *arXiv:2202.09553*.
[49] D. U. Fan, J. Pin-Qun, S. Shu-xiang, and X. Hai-ying, "Single image defogging algorithm based on attention mechanism," *Microelectron. Comput.*, vol. 38, no. 4, pp. 52–56, 2021.
[50] S. He, Z. Chen, F. Wang, and M. Wang, "Integrated image defogging network based on improved atmospheric scattering model and attention feature fusion," *Earth Sci. Informat.*, vol. 14, no. 4, pp. 2037–2048, Dec. 2021.
[51] X. Lin, Q. Huang, W. Huang, X. Tan, M. Fang, and L. Ma, "Single image deraining via detail-guided efficient channel attention network," *Comput. Graph.*, vol. 97, pp. 117–125, Jun. 2021.
[52] Q. Wang, G. Sun, H. Fan, W. Li, and Y. Tang, "APAN: Across-scale progressive attention network for single image deraining," *IEEE Signal Process. Lett.*, vol. 29, pp. 159–163, 2022.
[53] K. Jiang, Z. Wang, P. Yi, C. Chen, B. Huang, Y. Luo, J. Ma, and J. Jiang, "Multi-scale progressive fusion network for single image deraining," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 8343–8352.

[54] H. Yin and H. Deng, "RAiA-Net: A multi-stage network with refined attention in attention module for single image deraining," *IEEE Signal Process. Lett.*, vol. 29, pp. 747–751, 2022.

[55] Y. Park, M. Jeon, J. Lee, and M. Kang, "MCW-Net: Single image deraining with multi-level connections and wide regional non-local blocks," 2020, *arXiv:2009.13990*.

[56] C. Lin, Z. Tao, A. Xu, L. Kang, and F. Akhyar, "Sequential dual attention network for rain streak removal in a single image," *IEEE Trans. Image Process.*, vol. 29, pp. 9250–9265, 2020.

[57] P. Li and S. Gai, "Single image deraining using multi-scales context information and attention network," *J. Vis. Commun. Image Represent.*, vol. 90, Feb. 2023, Art. no. 103695.

[58] T. Zhang, N. Jiang, J. Lin, J. Lin, and T. Zhao, "DesnowFormer: An effective transformer-based image desnowing network," in *Proc. IEEE Int. Conf. Vis. Commun. Image Process. (VCIP)*, Dec. 2022, pp. 1–5.

[59] A. Jia, Z. Jia, J. Yang, and N. K. Kasabov, "Single-image snow removal based on an attention mechanism and a generative adversarial network," *IEEE Access*, vol. 9, pp. 12852–12860, 2021.

[60] B. Cheng, J. Li, Y. Chen, S. Zhang, and T. Zeng, "Snow mask guided adaptive residual network for image snow removal," 2022, *arXiv:2207.04754*.

[61] J. Lin, N. Jiang, Z. Zhang, W. Chen, and T. Zhao, "LMQFormer: A laplace-prior-guided mask query transformer for lightweight snow removal," 2022, *arXiv:2210.04787*.

[62] S. Chen, T. Ye, Y. Liu, T. Liao, J. Jiang, E. Chen, and P. Chen, "MSP-former: Multi-scale projection transformer for single image desnowing," 2022, *arXiv:2207.05621*.

[63] P. Wu, Z. Pan, H. Tang, and Y. Hu, "Cloudformer: A cloud-removal network combining self-attention mechanism and convolution," *Remote Sens.*, vol. 14, no. 23, p. 6132, Dec. 2022.

[64] H. Liu, B. Huang, and J. Cai, "Thick cloud removal under land cover changes using multisource satellite imagery and a spatiotemporal attention network," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5601218.

[65] X. Wen, Z. Pan, Y. Hu, and J. Liu, "An effective network integrating residual learning and channel attention mechanism for thin cloud removal," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[66] M. Xu, F. Deng, S. Jia, X. Jia, and A. J. Plaza, "Attention mechanism-based generative adversarial networks for cloud removal in Landsat images," *Remote Sens. Environ.*, vol. 271, Mar. 2022, Art. no. 112902.

[67] R. Shen, X. Zhang, and Y. Xiang, "AFFNet: Attention mechanism network based on fusion feature for image cloud removal," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 36, no. 8, Jun. 2022, Art. no. 2254014.

[68] R. Jing, F. Duan, F. Lu, M. Zhang, and W. Zhao, "Cloud removal for optical remote sensing imagery using the SPA-CycleGAN network," *J. Appl. Remote Sens.*, vol. 16, no. 3, Aug. 2022, Art. no. 034520.

[69] L. Sun, Y. Zhang, X. Chang, Y. Wang, and J. Xu, "Cloud-aware generative network: Removing cloud from optical remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 4, pp. 691–695, Apr. 2020.

[70] Z. Shen, H. Xu, T. Luo, Y. Song, and Z. He, "UDAformer: Underwater image enhancement based on dual attention transformer," *Comput. Graph.*, vol. 111, pp. 77–88, Apr. 2023.

[71] J. Wang, P. Li, J. Deng, Y. Du, J. Zhuang, P. Liang, and P. Liu, "CA-GAN: Class-condition attention GAN for underwater image enhancement," *IEEE Access*, vol. 8, pp. 130719–130728, 2020.

[72] Y. Li and R. Chen, "UDA-Net: Densely attention network for underwater image enhancement," *IET Image Process.*, vol. 15, no. 3, pp. 774–785, Feb. 2021.

[73] C. D. Mello, B. U. Moreira, P. J. D. de Oliveira Evald, P. J. L. Drews, and S. S. da Costa Botelho, "Underwater enhancement based on a self-learning strategy and attention mechanism for high-intensity regions," *Comput. Graph.*, vol. 107, pp. 264–276, Oct. 2022.

[74] S. Liu, H. Fan, S. Lin, Q. Wang, N. Ding, and Y. Tang, "Adaptive learning attention network for underwater image enhancement," *IEEE Robot. Autom. Lett.*, vol. 7, no. 2, pp. 5326–5333, Apr. 2022.

[75] X. Yan, W. Qin, Y. Wang, G. Wang, and X. Fu, "Attention-guided dynamic multi-branch neural network for underwater image enhancement," *Knowl.-Based Syst.*, vol. 258, Dec. 2022, Art. no. 110041.

[76] B. Fu, L. Wang, R. Wang, S. Fu, F. Liu, and X. Liu, "Underwater image restoration and enhancement via residual two-fold attention networks," *Int. J. Comput. Intell. Syst.*, vol. 14, no. 1, pp. 88–95, 2021.

[77] R. Li, R. T. Tan, and L. Cheong, "All in one bad weather removal using architectural search," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 3172–3182.

[78] J. M. J. Valanarasu, R. Yasarla, and V. M. Patel, "Transweather: Transformer-based restoration of images degraded by adverse weather conditions," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jan. 2022, pp. 2353–2363.

[79] A. Kulkarni, S. S. Phutke, and S. Murala, "Unified transformer network for multi-weather image restoration," in *Proc. Eur. Conf. Comput. Vis.* Cham: Springer Nature Switzerland, Oct. 2022, pp. 344–360.

[80] O. Özdenizci and R. Legenstein, "Restoring vision in adverse weather conditions with patch-based denoising diffusion models," *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 1–12, 2023.

[81] W.-T. Chen, Z.-K. Huang, C.-C. Tsai, H.-H. Yang, J.-J. Ding, and S.-Y. Kuo, "Learning multiple adverse weather removal via two-stage knowledge learning and multi-contrastive regularization: Toward a unified model," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 17632–17641.

[82] M.-Y. Liu, T. Breuel, and J. Kautz, "Unsupervised image-to-image translation networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–9.

[83] P. Isola, J. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5967–5976.

[84] N.-C. Ristea, A.-I. Miron, O. Savencu, M.-I. Georgescu, N. Verga, F. S. Khan, and R. T. Ionescu, "CyTran: A cycle-consistent transformer with multi-level consistency for non-contrast to contrast CT translation," 2021, *arXiv:2110.06400*.

[85] J. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2242–2251.

[86] T. Park, A. A. Efros, R. Zhang, and J.-Y. Zhu, "Contrastive learning for unpaired image-to-image translation," in *Proc. 16th Eur. Conf. Comput. Vis. (ECCV)*, vol. 16, Glasgow, U.K.: Springer, Aug. 2020, pp. 319–345.

[87] X. Qin, Z. Wang, Y. Bai, X. Xie, and H. Jia, "FFA-Net: Feature fusion attention network for single image dehazing," in *Proc. AAAI Conf. Artif. Intell.*, vol. 34, no. 7, 2020, pp. 11908–11915.

[88] Y. Jin, W. Yan, W. Yang, and R. T. Tan, "Structure representation network and uncertainty feedback learning for dense non-uniform fog removal," in *Proc. 16th Asian Conf. Comput. Vis.* Cham, Switzerland: Springer, 2023, pp. 155–172.

[89] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, M. Yang, and L. Shao, "Multi-stage progressive image restoration," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 14816–14826.

[90] F. Xu, Y. Shi, P. Ebel, L. Yu, G.-S. Xia, W. Yang, and X. X. Zhu, "GLF-CR: SAR-enhanced cloud removal with global–local fusion," *ISPRS J. Photogramm. Remote Sens.*, vol. 192, pp. 268–278, Oct. 2022.

[91] S. Gonzalez-Sabbagh, A. Robles-Kelly, and S. Gao, "DGD-cGAN: A dual generator for image dewatering and restoration," 2022, *arXiv:2211.10026*.

[92] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, and Z. Wang, "Benchmarking single-image dehazing and beyond," *IEEE Trans. Image Process.*, vol. 28, no. 1, pp. 492–505, Jan. 2019.

[93] A. R. Zamir and M. Shah, "Image geo-localization based on multiplen-earest neighbor feature matching usinggeneralized graphs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 8, pp. 1546–1558, Aug. 2014.

[94] P. Zhu, L. Wen, D. Du, X. Bian, H. Fan, Q. Hu, and H. Ling, "Detection and tracking meet drones challenge," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 11, pp. 7380–7399, Nov. 2022.

[95] G. Xia, X. Bai, J. Ding, Z. Zhu, S. Belongie, J. Luo, M. Datcu, M. Pelillo, and L. Zhang, "DOTA: A large-scale dataset for object detection in aerial images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3974–3983.

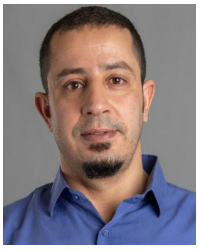
[96] M. J. Islam, Y. Xia, and J. Sattar, "Fast underwater image enhancement for improved visual perception," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 3227–3234, Apr. 2020.

[97] C. Li, C. Guo, W. Ren, R. Cong, J. Hou, S. Kwong, and D. Tao, "An underwater image enhancement benchmark dataset and beyond," *IEEE Trans. Image Process.*, vol. 29, pp. 4376–4389, 2020.

[98] *Image Augmentation for Machine Learning Experiments*. Accessed: Jan. 11, 2023. [Online]. Available: https://github.com/aleju/imgaug

[99] M. A. Kenk and M. Hassaballah, "DAWN: Vehicle detection in adverse weather nature dataset," 2020, *arXiv:2008.05402*.

MARIA SIDDIQUA received the Bachelor of Engineering degree (Hons.) in computer systems engineering from the Mehran University of Engineering and Technology, Jamshoro, Pakistan, in 2018, and the M.S. and Ph.D. degrees (Hons.) in computer science from the Pakistan Institute of Engineering and Applied Sciences, Islamabad, Pakistan, in 2020 and 2023, respectively. Her research interest includes data-driven image restoration. She is actively engaged in this field.



SAMIR BRAHIM BELHAOUARI received the master's degree in telecommunications and networks from Institut Nationale Polytechnique of Toulouse, France, in 2000, and the Ph.D. degree in mathematics from the Federal Polytechnic School of Lausanne-Switzerland, in 2006. He is currently an Associate Professor with the Division of Information and Communication Technologies, College of Science and Engineering, University of Hamad Bin Khalifa (HBKU), Qatar Foundation. His research interests include applied mathematics, statistics, data analysis to artificial intelligence, and image and signal processing due to both mathematics and computer science backgrounds.

NAEEM AKHTER received the bachelor's degree in agricultural engineering from the University of Agriculture, Faisalabad, Pakistan, in 2001, the master's degree in information technology from the Pakistan Institute of Engineering and Applied Sciences, Islamabad, Pakistan, in 2003, and the Ph.D. degree in computer vision from the Vienna University of Technology, Vienna, Austria, in 2012. He is currently an Assistant Professor with the Pakistan Institute of Engineering and Applied Sciences, Islamabad, Pakistan. His current research interests include data-driven image processing, image synthesis, and 3D modeling.

ANEELA ZAMEER received the M.Sc. degree in physics and the M.S. degree in nuclear engineering from Quaid-i-Azam University, Islamabad, Pakistan, in 1997 and 1999, respectively, the Ph.D. degree in device modeling from the University of West of Scotland, U.K., in 2004, and the Ph.D. degree from the University of Glasgow, U.K., in 2008. She is currently a full-time Professor with the Pakistan Institute of Engineering and Applied Sciences, Islamabad. Her research interests include modeling and simulation of physical systems, stochastic optimization, evolutionary computing, and machine learning.

JAVAID KHURSHID received the M.Sc. degree in electronics and the M.S. degree in system engineering from Quaid-i-Azam University, Islamabad, Pakistan, and the Ph.D. degree in computer application technology from the Harbin Institute of Technology, China. He is currently an Assistant Professor with the Pakistan Institute of Engineering and Applied Sciences, Islamabad. His research interests include artificial intelligence, mobile robots, and multi-robot systems.