2024 IEEE Students Conference on Engineering and Systems (SCES), June 21- 23, 2024, Prayagraj, India

Enhancing Traffic Visibility Using Self Attention based Generative Adversarial Network Model

Darshanbhai Maheshbhai Rathod
Computer Science & Engineering
MNNIT Allahabad
Prayagraj, India
darshanrathod4400@gmail.com

Rajitha B.
Computer Science & Engineering
MNNIT Allahabad
Prayagraj, India
rajitha@mnnit.ac.in

***Abstract*—Highway traffic monitoring systems often encounter difficulties caused by poorer image quality due to adverse weather conditions like rain, fog, insufficient lighting, and low visibility during nighttime. This paper introduces a fresh strategy to tackle these challenges by employing an advanced image enhancement algorithm centered on an enhanced Generative Adversarial Network (GAN) incorporating multiscale feature fusion. Our methodology integrates attention mechanisms within the generator network to effectively diminish noise and elevate focus on intricate details within high-frequency regions, thereby addressing issues such as poor resolution and blurred imagery. Furthermore, we refine the discriminator utilizing an improved PatchGAN methodology that employs a localized discrimination strategy, enhancing its ability to distinguish between generated and authentic images. By facilitating ongoing interaction between the generator and discriminator, we attain a state of Nash equilibrium, thereby preserving the integrity and authenticity of the restored images. Our proposed model delivers impressive results, boasting peak signal-to-noise ratio (PSNR) values of 32.5, 29.16, and 28.7, coupled with structural similarity index (SSIM) scores of 0.89, 0.85, and 0.84 for night, fog, and rain scenarios respectively. These outcomes underscore the efficacy of our approach in enhancing highway traffic imagery under adverse weather conditions, holding promising prospects for real-world traffic monitoring systems**

***Index Terms*—Image Enhancement, multiscale feature, Generative Adversarial Network, Channel Attention, Pixel Attention.**

I. Introduction

Nowadays, many organizations are actively engaged in the development of intelligent transportation systems, conducting studies encompassing image segmentation, image classification, object detection, speed measurement, target detection, and more. These studies heavily rely on traffic road image data collected from diverse locations. However, varying weather conditions such as rain, fog, insufficient illumination, and night lighting can lead to unclear images, impacting the performance of classification and target detection algorithms. Consequently, image enhancement for such challenging scenarios has emerged as an essential research topic.

In the past, numerous traditional methods have been proposed, categorized into model-based and non-model-based approaches. Non-model-based methods, including Gaussian filtering, spatial filtering, and light intensity-based algorithms like Retinex, have been explored to address these challenges. Despite the advancements in traditional methods, the complexities introduced by diverse weather conditions have prompted a shift toward modern techniques. These include deep learning-based super-resolution, adaptive contrast enhancement, and the integration of generative adversarial networks (GANs) [1]. These contemporary approaches aim to improve the accuracy of image processing tasks under adverse conditions, contributing to the ongoing refinement of image enhancement methods for intelligent transportation systems.

To address the limitations posed by adverse weather conditions on image clarity and accuracy in intelligent transportation systems, this paper introduces an attention-based GAN model. By leveraging attention mechanisms within the generator network and refining the discriminator through an enhanced PatchGAN methodology, the model aims to enhance image details and restore true colors effectively. This introduction sets the stage for a focused exploration of the proposed methodology and its potential impact on improving the reliability of traffic monitoring systems in challenging environmental conditions.

The paper is structured as follows: Section II presents the Existing Approaches in the image enhancement domain. Section III introduces the proposed model and architecture. Section IV displays the experimental results. Finally, Section V concludes the proposed model and its results.

II. EXISTING APPROACHES

This section present some Traditional and deep learning methods for image Enhancement.

Donglei Zhang [2] introduced an image enhancement technique based on histogram equalization tailored for various road conditions. Their method yielded structural similarity index (SSIM) values of 0.7703, 0.5273, and 0.3618 for night, fog-degraded, and rainy images, respectively. Gaussian Filtering (GF) [3] utilizes a sliding window with dimensions N × N to perform a mean operation on each pixel within a noisy image. Renowned for its efficacy, this method proves particularly adept at mitigating fog across the entirety of an image. Orest Kupyn [4]introduced a model-based algorithm for motion deblurring known as DeblurGAN, which utilizes a conditional Generative Adversarial Network (GAN). This method achieves

Identify applicable funding agency here. If none, delete this.

a structural similarity index (SSIM) of 0.86 and a peak signal-to-noise ratio (PSNR) of 24.42 when applied to fog-degraded images. Xin Cheng [5] proposed an enhanced Generative Adversarial Network (GAN)-based model augmented with additional perceptual loss. This method demonstrates notable improvements across various weather conditions.

An enhancement technique was introduced in [6] to improve the quality of low-resolution underwater images. This method utilized an enhanced generation countermeasure network, where the traditional CNN was replaced by the U-Net network. Additionally, jump connection technology was employed to integrate low-level and high-level feature image information, thereby enhancing image brightness and improving image details simultaneously. While effective for images with uneven lighting intensity, this approach had limitations in restoring details from overexposed areas.

Existing approaches have historically been employed to address challenges posed by adverse weather conditions like fog and rain. However, their effectiveness can be limited in preserving image details and accurately restoring visibility, especially in scenarios with significant degradation. HE often produces artifacts and exaggerated contrast, while DehazeNet [7], while specialized in fog removal, may struggle to adapt to varying weather conditions. These constraints underscore the necessity for more robust and adaptable image enhancement techniques. Hence, the proposed model presents a novel approach to overcome these limitations, leveraging advanced Generative Adversarial Networks (GANs) integrated with attention mechanisms for improved performance across diverse environmental factors

III. Proposed Methodology

The proposed model is based on GAN which has a generative block and discriminator block. The generative feature extractor architecture is designed with two parallel branches for Configuring different and crucial information from input images. Each of this branches have different attention blocks for extracting the features channel wise and pixel wise respectively. The channel wise feature extractor is represented in Fig. 1 and named as CWF (channel wise feature extractor) block. Similarly the pixel wise feature extractor block(PWF) is shown in Fig. 2. The overall process of proposed model is depicted in Fig. 4.

A. Generator Optimization Based on Attention Mechanism

Incorporating CWF (channel wise feature extractor) block [8] into the network enables adaptive learning of the varied weighted information across different channels within the image. This addition empowers the network to swiftly grasp the inter-channel relationships, thereby enhancing its feature extraction capabilities effectively. Similarly, Pixel Attention mechanisms [8] are integrated to further refine feature extraction by dynamically emphasizing pixel-level details, facilitating comprehensive understanding and representation of the image content.

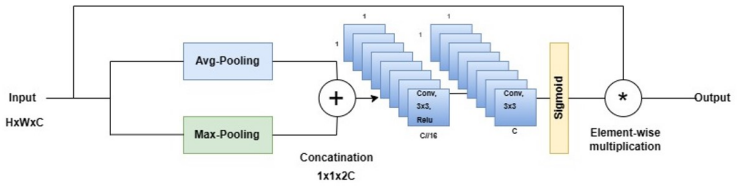

Fig. 1. CWF structure module.

The CWF structure module is shown in Fig. 1. The module initially employs both average pooling and max pooling operations to condense the input feature map F_c along the channel axis, reducing its dimensions from $H \times W \times C$ to $1 \times 1 \times C$. Following this, it applies a convolutional layer with C filters and a 1×1 kernel size, along with ReLU and sigmoid activation functions, to capture a holistic understanding of channel-level weight relationships. .

$$\text{CWF}_c = \delta\left(\text{Conv}_{\text{filters=C}}\left(\sigma\left(\text{Conv}_{\text{filters=C}}\left(\text{Concat}\left(\text{AP}(F_c), \text{MP}(F_c)\right)\right)\right)\right)\right) \quad (1)$$

Where, δ represents the activation function Sigmoid, σ represents the activation function ReLU, and Conv represents the convolution layer. The equation describes the Channel Attention mechanism. It begins by computing average and max pooled features from the input F_c, denoted as AP(F_c) and MP(F_c) respectively. These are concatenated and fed through two convolutional layers. Finally, an elementwise multiplication is conducted to generate new weighted features (F_c^*).

$$F_c^* = \text{CWF}_c \cdot F_c \quad (2)$$

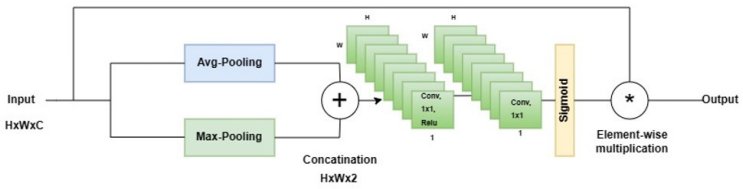

Fig. 2. PWF structure module.

The PWF structure module is shown in Fig. 2. This architecture shares similarities with Channel Attention but diverges in the number of filters and kernel window size. By compressing the input feature map from $H \times W \times C$ to $H \times W \times 1$, the focus shifts towards individual pixel features. This prioritization of pixel-level details over channel-level relationships enhances the model's sensitivity to local features. The reduction in dimensionality, facilitated by a convolutional layer with 1 filter and a 3×3 kernel size, enables a more granular analysis of

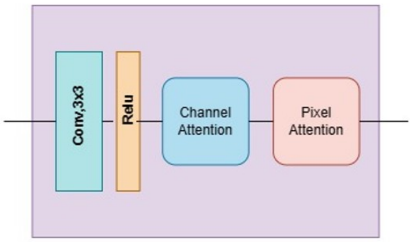

Fig. 3. CWPWAB structure module.

spatial information, aiding the network in capturing intricate pixel-level patterns within the data.

$$\mathrm{PWF}_c = \delta\left(\mathrm{Conv}_{\mathrm{filters=1}}\left(\sigma\left(\mathrm{Conv}_{\mathrm{filters=1}}\left(\mathrm{Concat}\left(\mathrm{AP}(F_c), \mathrm{MP}(F_p)\right)\right)\right)\right)\right) \quad (3)$$

$$F_p^* = \mathrm{PWF}_c \cdot F_p \quad (4)$$

Where, δ represents the activation function Sigmoid, σ represents the activation function ReLU, and Conv represents the convolution layer. The equation describes the Channel Attention mechanism. It begins by computing average and max pooled features from the input F_p, denoted as AP(F_p) and MP(F_p) respectively. These are concatenated and fed through two convolutional layers. Finally, an elementwise multiplication is conducted to generate new weighted features (F_p^*).

B. CWPWAB (Channel-wise Pix-wise Attention Block)

The CWPWAB Module combines Channel Attention and Pixel Attention Block functionalities (Fig. 3). Alongside the CWF and PWF modules, a convolution layer aids in local learning. Initially, the Convolution layer extracts features, followed by processing through the CWF module for enhanced inter-channel relationships. Subsequently, the PWF module emphasizes pixel-level details, enriching the model's sensitivity to local features and intricate pixel-level patterns.

With the integration of the Convolution layer, CWF module, and PWF module, the CWPWAB block orchestrates a sophisticated interplay of channel and pixel-level attention mechanisms, ensuring effective capture of both global and local features. This approach contributes to the network's ability to generate high-quality images while maintaining stability and convergence efficiency during training

C. Generator Architecture

The generator function defines a convolutional neural network (CNN) architecture tailored for image generation tasks. It begins with an input layer accepting images of size 256x256 pixels with three RGB color channels. The network bifurcates into two branches, each processing the input imagery differently. Within each branch, a series of convolutional layers followed by channel attention and pixel attention mechanisms extract and refine features. These mechanisms dynamically adjust feature responses and emphasize pixel-level details, aiding in understanding image content comprehensively. Each branch undergoes 5 iterations, refining feature extraction through successive convolutional operations. The resultant features from both branches are concatenated and further processed through additional convolutional layers. Finally, a convolutional layer generates the final output image with three RGB color channels. This architecture effectively captures both global and local features of the input imagery, facilitating the generation of high-quality images while ensuring stability and efficiency during training. Fig. 4 shows the Generator architecture.

D. Discriminator Architecture (PatchGAN)

The discriminator plays a crucial role in determining whether an input image is genuine. Similar to a binary classifier, the traditional global discriminator makes decisions based on its output value. Deviations from a probability of 0.5 suggest disparities between the generated and real images, prompting the provision of a descending gradient to the generator to refine its output. Conversely, a probability of 0.5 indicates indistinguishability between the clear and generated images, signifying the completion of training.

PatchGAN [9] revolutionizes image discrimination by dividing input images into smaller patches for detailed assessment. Each patch undergoes independent grading, allowing the discriminator to discern subtle differences between real and generated images.

The discriminator network consists of four convolutional blocks followed by a final convolutional layer and a sigmoid activation function. It takes pairs of input and target images, concatenates them, and processes them through the convolutional blocks. Each block includes a convolutional layer, optional instance normalization, and LeakyReLU activation. The final convolutional layer outputs a single value indicating the likelihood that the input images are real. Fig. 5 shows the Disciminator architecture.

E. Loss Function

The generator's total loss $\mathcal{L}_{\mathrm{Gen}}$ comprises two components: Binary Cross-Entropy ($\mathcal{L}_{\mathrm{adv}}$) and L1 Loss ($\mathcal{L}_{\mathrm{content}}$). Binary Cross-Entropy ensures effective deception of the discriminator, while L1 Loss preserves the fidelity of generated samples to the ground truth. A hyperparameter $\lambda = 10$ scales the influence of the content loss relative to the conditioning loss, balancing realism with content preservation [10].

$$\mathcal{L}_{\mathrm{Gen}} = \mathcal{L}_{\mathrm{adv}} + \lambda \cdot \mathcal{L}_{\mathrm{content}} \quad (5)$$

The discriminator's loss $\mathcal{L}_D$ encompasses three components: $\mathcal{L}_{\mathrm{real}}$, $\mathcal{L}_{\mathrm{generated}}$, and $\mathcal{L}_{\mathrm{GP}}$ (WGAN) [11]. $\mathcal{L}_{\mathrm{real}}$ ensures accurate classification of real samples, while $\mathcal{L}_{\mathrm{generated}}$ targets correct classification of fake samples. $\mathcal{L}_{\mathrm{GP}}$ penalizes large gradients to stabilize training. By combining these terms, $\mathcal{L}_D$

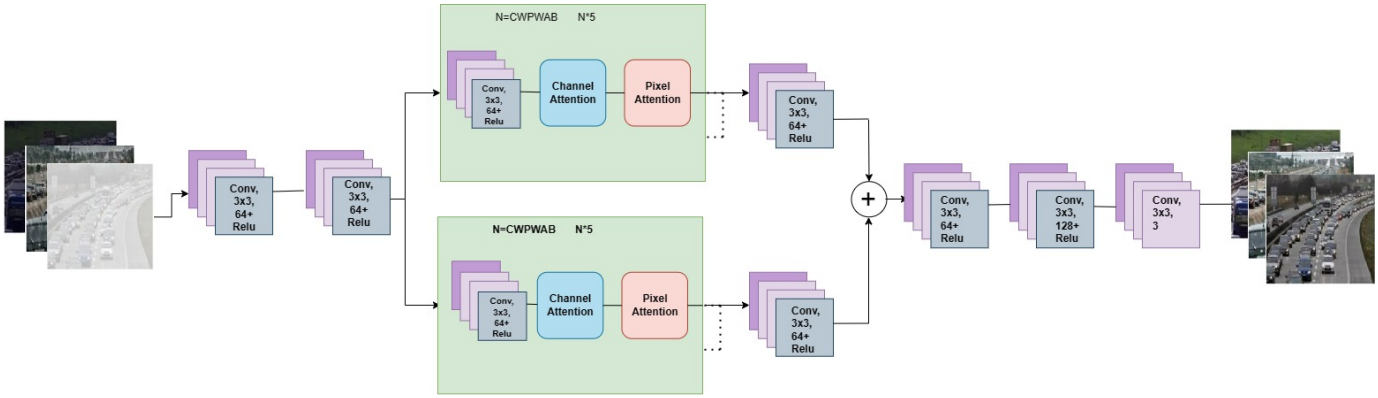

Fig. 4. Generator Block Architecture

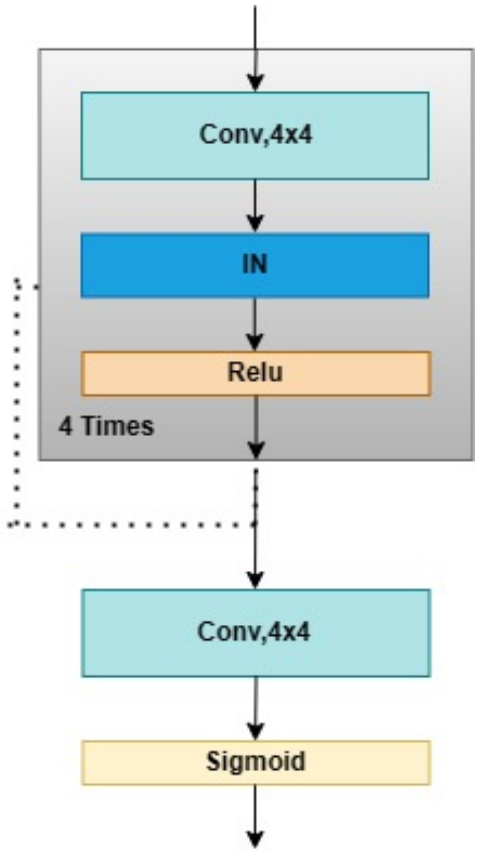

Fig. 5. Discriminator Block Architecture

guides the discriminator to effectively differentiate between real and generated samples while maintaining stability in the training process.

$$\mathcal{L}_{\mathrm{D}} = \mathcal{L}_{\mathrm{real}} + \mathcal{L}_{\mathrm{generated}} + (\lambda \times \mathrm{mean}(\mathcal{L}_{\mathrm{GP}})) \tag{6}$$

Fig. 6illustrates the optimization trends of the generator, offering insights into the evolution of image enhancement. Conversely, Fig. 7 monitors the discriminator's learning trajectory, facilitating comprehension of the adversarial training dynamics. These graphical representations play a crucial role in evaluating the stability and effectiveness of the proposed image enhancement algorithm

IV. Experimental Results

A. Experimental Data and setup

The research paper utilizes the TrafficNet dataset [12], comprising various types of traffic images such as dense traffic, sparse traffic, and accident scenarios, among others. Notably, the dataset consists of clear images depicting different traffic situations. To expand the dataset's diversity and assess the algorithm's efficacy, image augmentation techniques including Gaussian Blur, Contrast Enhancement, and other OpenCV techniques are employed to simulate different weather conditions such as fog, night, and rain. For evaluation purposes, a subset of 200 images are chosen for training and 50 images for testing.

The study utilizes the TensorFlow deep learning framework for training the model, with an experimental setup comprising an Intel Core i5-1235U processor, Intel UHD Graphics, and a Windows 10 operating system with 8GB of RAM. The training data consists of pairs of fog-degraded images and corresponding ground truth images, both with dimensions of 256x256 pixels. Optimization is performed using the Adaptive Moment Estimation algorithm with a learning rate set to 0.0002, and the training process is conducted over 250 epochs.

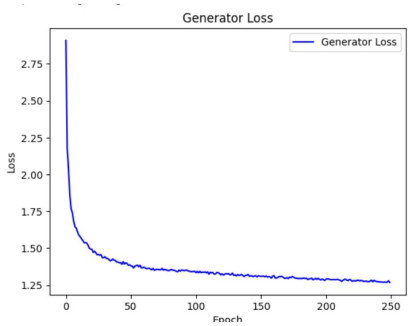

Fig. 6. Generator Loss(Rain)

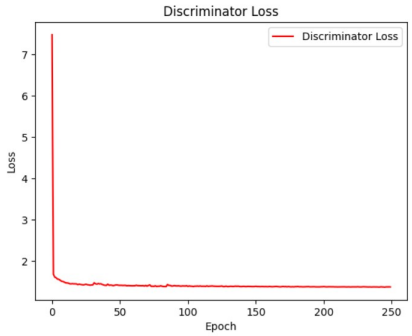

Fig. 7. Discriminator Loss(Rain)

Fig. 8. Results

B. Performance Evaluation

In this study, we employ PSNR (Peak Signal-to-Noise Ratio) and SSIM (Structural Similarity Index Measure) [13] simultaneously for quantitative assessment of the processed images, illustrating the advantages of image enhancement. PSNR measures the degree of image distortion or noise by computing the mean square error between two images. A smaller PSNR value indicates heightened degradation in the images and inferior image enhancement effects, while a larger PSNR value signifies greater image similarity, reduced distortion, and thus superior image enhancement.

$$\text{PSNR} = 20 \log_{10} \left(\frac{\text{MAX}}{\sqrt{\text{MSE}}} \right) \tag{7}$$

Where, MAX denotes the maximum value of pixel, for 8 bit binary value of MAX is 255. MSE denotes Mean Square Error.

SSIM evaluates image similarity considering brightness, contrast, and structure. It is computed based on the means, standard deviations, and covariance between the denoised and clear images. A higher SSIM value indicates a more similar pair of images with lower distortion levels, thus indicating better image enhancement. During computation, a sliding window is applied to the image, and the average SSIM value over the entire image is calculated.

$$\text{SSIM} = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \times \frac{2\sigma_{xy} + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \tag{8}$$

In the SSIM equation, μ_x and μ_y represent the mean values of the denoised and reference (clear) images, while σ_x and σ_y denote their respective standard deviations. Two constants, C_1 and C_2, are computed as $(k_1 L)^2$ and $(k_2 L)^2$, where L represents the range of pixel values. Specifically, k_1 and k_2 are predefined parameters set to 0.01 and 0.03, respectively. During the SSIM calculation process, a sliding $N \times N$ window is applied to the image, continuously shifting for computation. The resulting SSIM values are then averaged to obtain the global SSIM assessment. This methodology ensures a comprehensive evaluation of image similarity while considering local variations across the image.

TABLE I
PERFORMANCE OF PROPOSED MODEL FOR REMOVING RAIN,FOG AND NIGHT EFFECTS ON TRAFFICNET DATASET

	PSNR	SSIM	Cosine Similarity
Rain			
Img1	28.50	0.84	0.988
Img2	28.21	0.82	0.985
Img3	28.60	0.84	0.987
Img4	29.30	0.83	0.9907
Night			
Img1	35.63	0.943	0.998
Img2	29.19	0.79	0.984
Img3	35.56	0.967	0.998
Img4	34.14	0.93	0.997
Fog			
Img1	32.22	0.91	0.995
Img2	30.92	0.92	0.996
Img3	27.26	0.72	0.977
Img4	32.8	0.924	0.997

Figure 8 displays the outcomes achieved by our proposed method. In the first column, rain-degraded, night, and fog-degraded images are presented. In the second column, images

TABLE II
COMPARISON OF PSNR AND SSIM FOR DIFFERENT IMAGE ENHANCEMENT METHODS

Effect	Images	HE [2]		CLAHE [14]		Proposed model	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Rain	Img1	13.02	0.62	15.51	0.68	28.50	0.84
	Img2	14.9	0.62	17.59	0.70	28.21	0.81
	Img3	16.33	0.61	17.08	0.64	28.6	0.83
Fog	Img1	12.96	0.68	5.57	0.39	30.92	0.92
	Img2	15.77	0.82	8.86	0.70	32.22	0.91
	Img3	17.16	0.77	6.19	0.51	32.8	0.92
Night	Img1	17.31	0.84	13.10	0.87	33.28	0.92
	Img2	19.06	0.88	12.97	0.86	34.27	0.96
	Img3	15.74	0.79	19.80	0.93	35.37	0.94

generated by our model are showcased. The results demonstrate that prior to being processed by our model, the information within the images is not distinctly visible. However, after undergoing enhancement through our model, the images exhibit significant improvement in clarity and detail.

Table I shows Performance of Proposed model.The proposed model significantly improves image quality across various weather conditions in the TrafficNet Dataset. It achieves high PSNR values ranging from 28.21 to 35.63 dB for images affected by rain, fog, and nighttime conditions, indicating excellent preservation of image fidelity. SSIM values, ranging from 0.72 to 0.943, suggest strong similarity between enhanced and ground truth images. Additionally, cosine similarity values near 1 demonstrate a high degree of similarity in pixel distributions. Overall, the model effectively mitigates adverse weather effects, producing visually pleasing and perceptually similar images across different environmental conditions.

The table II provides a thorough analysis of PSNR and SSIM metrics comparing various image enhancement methods on images influenced by different weather conditions. In all cases, the proposed model surpasses both Histogram Equalization (HE) and Contrast Limited Adaptive Histogram Equalization (CLAHE), achieving significantly better PSNR and SSIM values. These findings underscore the efficacy of the proposed model in maintaining image fidelity and improving visual quality, especially in adverse weather scenarios. Moreover, the substantial enhancements demonstrated by the proposed model underscore its suitability for practical applications necessitating reliable image enhancement techniques.

V. CONCLUSION

This paper introduces an image enhancement algorithm utilizing a Generative Adversarial Network (GAN). The GAN incorporates Channel and Pixel attention mechanisms to reconstruct traffic images affected by various weather conditions into clear images. The channel and pixel attention blocks within the generator optimize both channel and pixel-wise aspects to produce clear images. The discriminator, designed to distinguish between generated and clear images, employs a patchGAN architecture, which considers patches across the entire image. The training of the GAN utilizes Binarycrossentropy, l1 loss, and WGAN-GP loss functions. Evaluation of the proposed model involves calculating Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM). The proposed model achieves PSNR scores of 32.5, 29.16, and 28.7, along with SSIM scores of 0.89, 0.85, and 0.84 for night, fog, and rain scenarios respectively.

This model verifies the effectiveness of image enhancement techniques under various weather conditions such as fog, night, and rain. However, it hasn't been tested on other weather conditions such as snow or extreme sunlight. To address this, the next step is to evaluate the model on a dataset containing diverse weather conditions. Furthermore, optimizing the generator's attention mechanism may enhance its capacity to extract temporal information from images. Additionally, the model hasn't been tested on pictures with mixed weather conditions yet, so that's something to consider for further experiments.In addition to these considerations, challenges also arise in training and processing large datasets, where computational resources and time constraints can slow down the scalability and efficiency of the proposed techniques.

REFERENCES

[1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, 2020.

[2] D. Zhang, "Research on road image enhancement based on histogram equalization," in *2023 IEEE 2nd International Conference on Electrical Engineering, Big Data and Algorithms (EEBDA)*, pp. 178–183, 2023.

[3] S. Selva Nidhyanandhan, R. Sindhuja, and R. S. Selva Kumari, "Double stage gaussian filter for better underwater image enhancement," *Wireless Personal Communications*, vol. 114, pp. 2909–2921, 2020.

[4] O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin, and J. Matas, "Deblurgan: Blind motion deblurring using conditional adversarial networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[5] X. Cheng, J. Zhou, J. Song, and X. Zhao, "A highway traffic image enhancement algorithm based on improved gan in complex weather conditions," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 8, pp. 8716–8726, 2023.

[6] B. Xu, D. Zhou, and W. Li, "Image enhancement algorithm based on gan neural network," *IEEE Access*, vol. 10, pp. 36766–36777, 2022.

[7] 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, p. 5187–5198, Nov. 2016.

[8] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European Conference on Computer Vision (ECCV)*, September 2018.

[9] U. Demir and G. Unal, "Patch-based image inpainting with generative adversarial networks," 2018.

[10] Z. Pan, W. Yu, B. Wang, H. Xie, V. S. Sheng, J. Lei, and S. Kwong, "Loss functions of generative adversarial networks (gans): Opportunities and challenges," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 4, no. 4, pp. 500–522, 2020.

[11] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of wasserstein gans," in *Advances in Neural Information Processing Systems* (I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, eds.), vol. 30, 2017.

[12] OlafenwaMoses, "Traffic-net," 2019. Release date: July 4, 2019.

[13] A. Horé and D. Ziou, "Image quality metrics: Psnr vs. ssim," in *2010 20th International Conference on Pattern Recognition*, pp. 2366–2369, 2010.

[14] I. Lashkov, R. Yuan, and G. Zhang, "Edge-computing-facilitated nighttime vehicle detection investigations with clahe-enhanced images," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 11, pp. 13370–13383, 2023.