Weather-Impaired Autonomous Driving Scene Regeneration using Local Self-Attention

Ghazala Rafiq, Ingyu Lee, Gyu Sang Choi, Yong Wan Park

Department of Information & Communication Engineering
Yeungnam University, Gyeongsan 38541, South Korea

ghazala@ynu.ac.kr, inleeatyu@yu.ac.kr, castchoi@ynu.ac.kr, ywpark@yu.ac.kr

Abstract

Autonomous vehicles (AVs) stand at the brink of transforming transport, offering enhanced efficiency and a promising horizon. Yet, interpreting scenes and making decisions under adverse weather conditions like fog, rain, haze, and snow remains a significant challenge. Drawing on the achievements of transformer architecture in vision tasks, this paper presents a method for regenerating weather-impaired images using local self-attention, aimed at enhancing scene understanding and decision-making in autonomous vehicles. The dataset is composed of images from the BDD100k collection, augmented with a variety of synthetically introduced noises to simulate adverse weather scenarios. The hierarchical transformer architecture utilizes a shifted windowing scheme for learning representations, enhancing efficiency and ensuring linear computational complexity relative to the size of the image. Comprehensive testing has shown the potential of degraded image regeneration to significantly improve the safety and reliability of autonomous vehicles in harsh weather conditions.

Ⅰ. Introduction

The ability to perceive the environment accurately is crucial for safe navigation in the autonomous vehicles (AV) technology. The navigation becomes particularly challenging when adverse weather conditions, such as fog, haze, rain, or snow, are there, causing significant degradation in the quality of images captured by onboard cameras. These weather conditions typically lead to issues like reduced contrast and color fidelity, and increased noise, quantitatively evident as alterations in pixel intensity distributions and increased mean squared error (MSE) values when compared to clear weather images.

Various weather conditions have been modelled in research by considering the physical phenomenon they exhibit. This image degradation can be modelled as in equation 1.

$$I_{WC} = I_{CLEAR}.\ T(x) + A(1-T(x)) \text{ ------- } (1)$$

Where I_{WC} is the weather-challenged or degraded image, I_{CLEAR} is the base clear image, $T(x)$ is the transmission map describing the portion of light reaching the camera, and A is the atmospheric light. The proposed model presents a clear avenue for the image regeneration through estimating the $T(x)$ and A and regeneration of I_{CLEAR}.

Current vision transformer [1] like methodologies primarily concentrate on global self-attention using resulting in quadratic computational complexity. Whereas, focusing on self-attention locally results in linear computational complexity within non-overlapping windows that partition an image. *Figure 1* presents a selection from our specially curated dataset, illustrating clear images alongside their counterparts that have been synthetically modified with various types of weather degradation overlays. This visual comparison showcases the range and quality of images used for training and testing in our study. In this work, we proposed a transformer architecture to regenerate and understand scenes in all types of adverse weather conditions for autonomous driving.

Figure 1. Dataset curation: showcasing base images alongside their degraded counterparts with four distinct noise variants.

Ⅱ. Method

We trained the network on 15k clear images from the BDD100k dataset and enhanced them with synthetic weather overlays or noises to train our model. These noises include gaussian, poison, perlin, random, and gradient noise. This approach resulted in a total of approximately 60K images, encompassing both clear and weather-affected conditions.

The weather-impaired image is divided into non-overlapping patches akin to the approach in Vision

Transformer (ViT) [1]. By employing local self-attention within the shifted window transformer block, similar to the methodology in Swin Transformer [2], multi-level features are extracted. This technique of layering attention scores has proven to be efficacious in reconstructing images, enhancing the quality and accuracy of the regenerated images. *Figure 2* demonstrates the transformer architecture used for image regeneration.

The network is trained employing a composite loss, consisting of structural similarity loss between generated and ground truth images, and perceptual loss as in [3]. The perceptual loss measures the features level discrepancies between the weather-impaired and clear ground truth image. Total loss can be formulated as formulated in equation 2.

$$Loss = SSIM\ Loss + \lambda\ Perceptual\ Loss \quad \text{-------} \quad (2)$$

Where λ is responsible to control the contribution of both loss components.

Ⅲ. Results

We evaluated our methodology using a dataset composed of synthetically weather-degraded images. For assessing the performance, we employed two metrics: PSNR (Peak Signal to Noise Ratio) and SSIM (Structural Similarity Index Measure), which together provide a comprehensive understanding of both the fidelity and the structural integrity of the regenerated images compared to the original ones. The quantitative results are displayed in *Table 1*. The network is trained for 200 epochs with a batch size of 32, using ADAM optimizer and a learning rate of 0.0002. Rest of the implementation detail is same as [2]. We compare the performance with our previously proposed encoder-decoder (ED) based architecture [4] for the same and noticed better performance both in terms of qualitative and quantitative evaluation. The reason being the transformer's ability to capture long-range dependencies and detailed context, which is essential in understanding and correcting complex patterns like those found in degraded weather conditions. Qualitative results are shown in *Figure 3*.

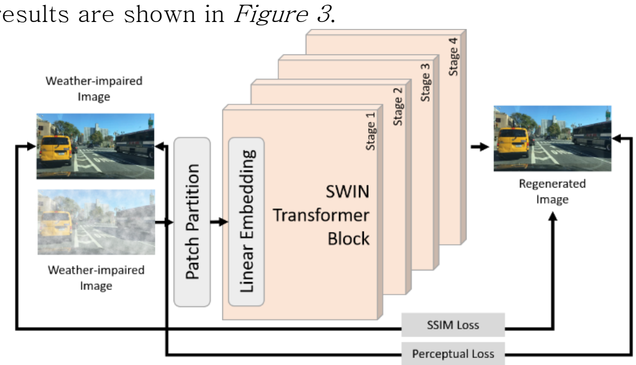

Figure 2. The Swin Transformer [2] is utilized for regenerating weather-impaired images, focusing on restoring quality and detail. The network is optimized using SSIM for structural similarity and Perceptual losses, which help maintain the visual fidelity of the regenerated images compared to the original ones. This approach ensures that the restored images are not only accurate in a pixel sense but also maintain perceptual quality, crucial for autonomous driving decisions.

Ⅳ. Conclusion

We proposed a transformer-based architecture utilizing local self-attention specifically designed for regenerating images impaired by adverse weather, enhancing autonomous driving capabilities. The local self-attention mechanism effectively mitigates degradations, facilitating precise scene comprehension and accelerated decision-making. Our extensive testing demonstrated improved performance and speed, marking a considerable advancement in autonomous vehicles' proficiency and safety across various weather conditions.

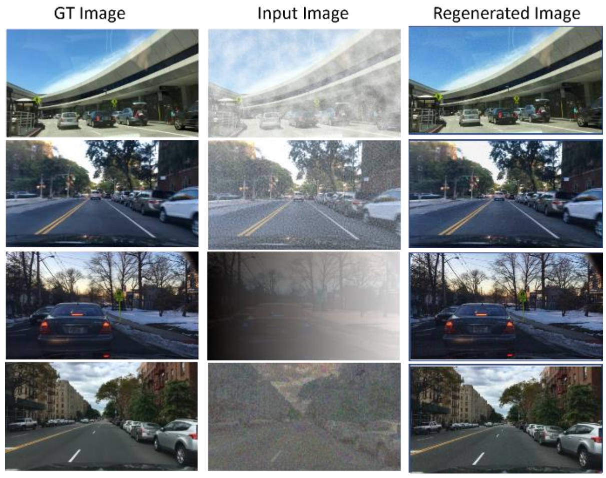

Figure 3. Qualitative Analysis: A visual comparison for evaluation, with column 1 displaying the clear images serving as the Ground Truth (GT), column 2 showing the weather degraded images used as input, and column 3 featuring the regenerated images, showcasing the effectiveness of our method in restoring image quality and detail in various weather conditions.

Table 1. Quantitative Results

Model	PSNR	SSIM
Current	32.97	0.9507
ED [4]	30.21	0.9401

ACKNOWLEDGMENT

This work was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2021R1A6A1A03039493).

REFERENCES

[1] A. Dosovitskiy et al., "An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale." Conference paper at ICLR 2021, [Online]. Available: https://arxiv.org/pdf/2010.11929.pdf

[2] Z. Liu et al., "Swin Transformer: Hierarchical Vision Transformer using Shifted Windows." [Online]. Available: https://openaccess.thecvf.com/content /ICCV2021/papers/ Liu_Swin_Transformer_Hierarchical _Vision_Transformer_Using_Shifted_Windows_ICCV_2021_paper.pdf

[3] J. Maria, J. Valanarasu, R. Yasarla, and V. M. Patel, "TransWeather: Transformer-based Restoration of Images Degraded by Adverse Weather Conditions." [Online]. Available: https://github.com/jeya-maria-jose/TransWeather.

[4] Ghazala Rafiq, Hiba Faisal, Gyu Sang Choi "Encoder-Decoder Based Hazy Image Restoration for Enhanced Autonomous Vehicle Perception" IEMEK 2023 Fall Conference, Jeju Island, South Korea.