Haze-Free Imaging Through Haze-Aware Transformer Adaptations

Patricia L. Suárez and Angel Sappa

Abstract This paper presents a novel method to remove non-uniform haze in real images. The process involves complex phases: initial extraction of image features, haze removal, and final image reconstruction. To address this complex challenge, the proposed approach implements a transformer-based architecture. By extending the SwinIR image restoration framework, the method includes modifications to the deep feature extraction module and also implements adaptive tokenization and learnable position embeddings. The results show significant advances with respect to existing models, validating the effectiveness of the proposed strategy to remove non-homogeneous haze within the images.

Keywords Non-homogeneous haze · Feature encoding · Adapting tokenization

1 Introduction

In the area of image processing, the challenge of mitigating blurring effects induced by non-uniform haze remains an unsolved problem. The presence of non-uniform haze introduces complexities that traditional image processing techniques can find difficult to overcome. Because it introduces varying levels of atmospheric interference in different regions of an image. This spatial variability requires an adaptive and nuanced approach to effectively restore scene clarity. Addressing this problem improves the clarity and fidelity of the scenes captured in the images. This challenge

http://www.espol.edu.ec.

P. L. Suárez (✉) · A. Sappa
ESPOL Polytechnic University, Guayaquil, Ecuador
e-mail: plsuarez@espol.edu.ec

A. Sappa
e-mail: asappa@espol.edu.ec; asappa@cvc.uab.es

A. Sappa
Computer Vision Center, Barcelona, Spain

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2024
S. Roy et al. (eds.), *Innovations in Computational Intelligence and Computer Vision*, Lecture Notes in Networks and Systems 1116,
https://doi.org/10.1007/978-981-97-6995-7_3

has generated a lot of research of possible methods in the image processing field. Several works have been proposed, which have made fundamental contributions (e.g., [3–6, 15]). These proposals focus on developing effective methodologies to reduce the impact of non-uniform haze in images.

Traditional methods for removing haze from a single image often have limited effectiveness in complex scenes. Those approaches lack the complete information needed to accurately model and mitigate the effects of haze in scenes with diverse visual elements, such as non-homogeneous haze. The absence of additional perspectives or contextual information hinders the algorithm's ability to make accurate estimates.In this work, we propose a non-homogeneous haze removal process utilizing adaptive tokenization, learnable position embeddings, and dynamic learning rate scheduling techniques. Our approach aims to process a single input image with different haze levels to more accurately estimate and remove the haze distribution within the scene.

Lately, different strategies have been implemented to mitigate the impact of atmospheric interference. These approaches focus on enhancing image characteristics and reconstructing scenes to restore the visual quality of images affected by haze. Understanding what makes haze in real scenarios is complicated. There is a common equation to explain it:

$$I(x) = J(x)t(x) + (1 - t(x))A. \quad (1)$$

This formula defines a hazy picture $I(x)$, what the scene really looks like $J(x)$, how light goes through the haze $t(x)$, and the light that's already in the air A. This formula helps to figure out how the haze changes what we see and guides us in finding ways to fix it. The provided formula encapsulates the essence of understanding and modeling the phenomenon of haze in images. Also, it delineates how the observed haze in an image results from the combination of the underlying scene, the transmission of light through the atmospheric medium, and the presence of ambient atmospheric illumination.

The manuscript is organized as follows. Section 2 presents works related to the haze removal problem and some basic concepts of transformer networks. Section 3 presents the proposed haze removal architecture. Experimental results and comparisons with different implementations are given in Sect. 4. Finally, conclusions are presented in Sect. 5.

2 Related Work

Non-homogeneous haze removal is a critical task in computer vision and image processing, aiming to enhance the visibility and perceptual quality of images captured in hazy or foggy conditions. Traditional methods for haze removal often assume a uniform haziness across the entire image, ignoring the spatially variant nature of haze, which can lead to inaccurate results. Non-homogeneous haze removal techniques seek to address this limitation by considering varying haze densities and colors across different regions of the image, leading to more accurate and visually pleasing results.

One of the pioneering solutions to this problem is the Dark Channel Prior (DCP) that has been presented by he et al. [7]. This approach is based on the usage of the prior knowledge that in most haze-free outdoor images, some pixels (usually found in shadows, colorful items, etc.) have very low intensity in at least one color channel. They use this observation to estimate the thickness of the haze and then remove it from the image. Their method works well for a wide range of outdoor images; however, it may fail to accurately recover the scene depth in scenarios with non-homogeneous haze. Subsequently, with the advent of advanced computational models, non-homogeneous haze models started to emerge. Dong [4] proposes a non-local low-rank matrix recovery approach, which exploits the inherent low-rank property of haze-free latent images and the sparsity structure of haze-overlain images. Their method demonstrates improved resilience in dealing with spatial variation of haze and has been shown to perform impressively on both synthetic and real-world hazy images. In Zhu et al. [16] the authors introduce a novel, fast single-image haze removal algorithm using color attenuation prior. This method contends that the depth map should correlate with the saturation and brightness of the image and its performance is verifiable superior to prior arts, especially under real-world, non-homogeneous haze scenarios.

Despite these advancements, non-homogeneous haze removal remains a challenging task. The primary concern focuses on the difficulty of precisely estimating the haze's spatial distribution, especially given the minimal information commonly provided by single images. In [1], NH-HAZE is a new image dehazing benchmark that provides non-homogeneous haze and haze-free images for evaluation. It has been used in several competitions on methods for removing non-homogeneous haze and determining the state-of-the-art approach. In [10] a novel approach has been presented to non-homogeneous haze removal based on real-world image dehazing with enhancement-and-restoration fused convolutional neural networks (CNNs). In [14], the authors explore the usage of ensemble learning for improving the performance of haze removal methods in non-homogeneous haze environments. The approach combines multiple haze removal networks to enhance the accuracy of the removing process, especially in situations where the haze is highly variable and non-uniform. Another non-homogeneous haze removal approach is introduced in [12], where the authors present an innovative attention-based architecture designed specifically for eliminating non-homogeneous haze from images. The model focuses on capturing the most essential image characteristics in each learning cycle. It achieves this through adaptive attention modules combined with a residual learning convolutional network, leveraging the Res2Net model.

Non-homogeneous haze removal is important in a wide range of applications, particularly due to the increasing reliance on clear image-based data and automation in various processes. Therefore, research into new, more advanced and robust methods must continue.

3 Proposed Approach

Removing image haze is a common challenge in computer vision approaches, particularly because of the complexity present in scenarios where images are affected by atmospheric haze. In order to address this challenge, the proposed approach see Fig. 1, leverages adaptive tokenization, which allows the system to dynamically determine region sizes based on the specific content of the input image. This adaptive tokenization see Fig. 2 enables the representation of complex visual features at varying scales, thereby significantly enhancing the quality of clear images. We define the adaptive mechanisms used to determine region sizes or scales in an image through a formula that dynamically adjusts the region size based on certain image characteristics or features. This is achieved through the following formula:

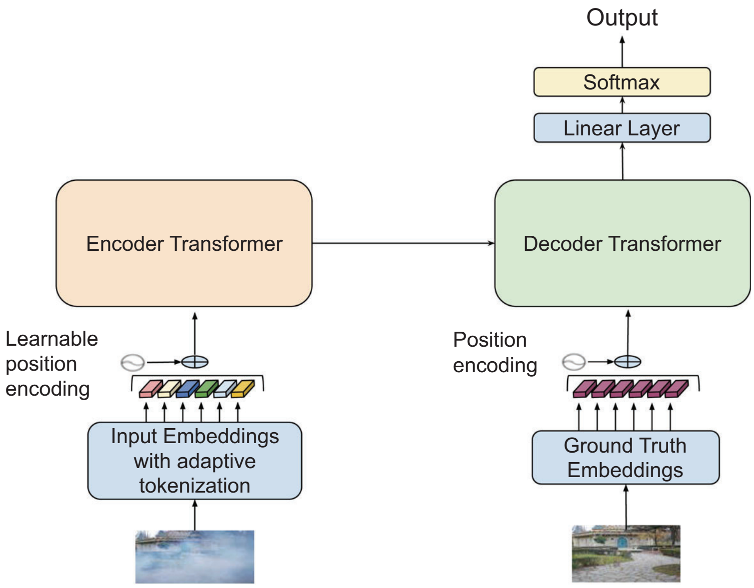

Fig. 1 Transformer proposed architecture

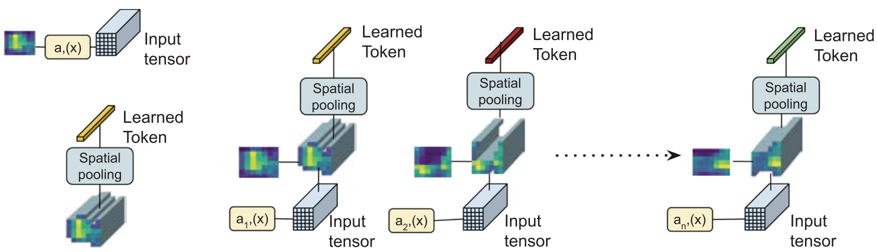

Fig. 2 Proposed adaptive tokenization

$$S_i = f(C_i) \tag{2}$$

where S_i is the size of the ith region; C_i represents specific content or features within the ith hazy region; and $f(C_i)$ is a function that determines the region size based on the content or features identified in the region to identify hazy patterns within the region and replace them with the corresponding pattern of the related clear regions. The function defines that the size of each region (S_i) is determined by a function $f(C_i)$ that takes into account the content within that region. This function has been defined based on the intensity of the pixel as:

$$C_i = k \times I_i \tag{3}$$

where C_i are the features of the ith region; I_i represents the mean intensity of pixels in the i th region, and k is a scaling factor empirically defined.

Within the context of non-homogeneous haze, $f(C_i)$ represents the function capturing the correlated spatial relationships between tokens affected by varying haze densities and patterns. This functionality allows the model to better understand and address the irregular distortions caused by non-uniform haze, ultimately improving the efficacy of haze removal across different regions of the image. The adaptability of $f(C_i)$ is crucial in enabling the model to interpret and respond to the irregular spatial distortions caused by non-uniform haze, thereby contributing significantly to the model's capability in accurately and comprehensively removing haze-induced artifacts while preserving image details and quality.

Another consideration about of learnable position embeddings is inspired by the concept of dense connectivity in convolutional neural networks as proposed by [8], which has shown the effectiveness of enhancing feature propagation and information flow within deep networks. Our approach integrates learnable position embeddings to enhance the model's spatial awareness and representation of visual features. This enables the model to dynamically capture spatial relationships between tokens during training, enhancing its spatial awareness and representation of visual features. In the context of non-homogeneous haze removal from images, this technique becomes particularly valuable.

Non-homogeneous haze introduces varying degrees of distortion across different regions of an image, causing complex and irregular degradation. Learnable position embeddings play a pivotal role in enabling the model to adapt to these varying distortions by capturing specific spatial relationships affected by different levels and patterns of haze. The learnable position embeddings are computed as follows:

$$\text{Position embedding} = sr(\text{sprela}) \tag{4}$$

where sr(sprela) represents the spatial relationship implemented by convolution operations designed for extracting haze-specific features from hazy image regions and can be defined as a function that operates on hazy image regions to extract features specific to haze. This function can be represented as:

$$\text{sprela} = (\text{Softmax}(\text{LeakyReLU}(\text{Conv}(\text{ft}))) \tag{5}$$

here, (sprela) denotes a function designed to process hazy image regions represented by the features tensors(ft), extracting representations that are indicative of haze-affected patterns. This operation is strategically incorporated into the network architecture to facilitate the task of dehazing. Feature extraction with a specific emphasis on characteristics associated with regions affected by haze. By discerning and highlighting these unique features, the operation acts as a specialized filter within the network. It effectively identifies patterns and attributes that are indicative of the presence of haze in certain areas of the image.

Moreover, the utilization of a one-cycle learning rate schedule has been incorporated into our approach to introduce dynamic adjustments to the learning rate during training. This strategy is intended to potentially improve model convergence speed and facilitate the identification of improved model optima. This dynamic learning rate schedule is represented by:

$$\text{Learning rate} = h(t) \tag{6}$$

where $h(t)$ represents the dynamic function that adapts the learning rate t during training. This technique draws inspiration from the work of Smith [11], which emphasizes the importance of adjusting learning rates to promote faster convergence and facilitate the exploration of diverse regions in the optimization landscape.

In summary, the proposed approach integrates adaptive tokenization, learnable position embeddings, and dynamic learning rate scheduling to significantly enhance the representation of visual features and improve the efficacy of non-homogeneous haze removal.

4 Experimental Results

This section firstly describes the data set that has been used to validate the proposed approach. Then, training settings details are provided. Finally, results from the proposed approach and comparisons with state-of-the-art technique are provided, both quantitative and qualitative.

4.1 Datasets

In order to evaluate the proposed approach we have used a realistic and diverse dataset containing pairs of real non-homogeneous hazy and haze-free images introduced in [2]. This dataset was utilized in the NTIRE 2020 dehazing challenge and an extended version was employed in the NTIRE 2021 dehazing challenge, comprising 55 and 25 pairs of images respectively, along with their corresponding haze-free images.

The NH-Haze dataset is the first non-homogeneous haze dataset, offering ground truth images. According to Ancuti et al. [2], this non-homogeneous haze has been generated utilizing a sophisticated haze generator, that simulates the real conditions of haze scenes.

To conduct our experiments, we have selected 10 images from each of the NH-Haze 2020 and NH-Haze 2021 datasets and used them to calculate the metrics for evaluating the results. Out of the total 80 image pairs, we designated 60 pairs for training, 10 for testing, and 10 for validation.

4.2 Pre-procesing and Training Settings

In previous approaches [13] the authors proposed a transformer-based methodology, particularly a tokenization process, which is employed as a dataset pre-processing step. Also segment the input images into 32×32 regions, known as visual tokens, and embed them into fixed-dimensional vectors. The patch positions are also embedded along with the image regions, forming the input for the transformer model, to enhance the feature extraction. In the current work the combination of adaptive tokenization and learnable position embeddings is proposed in order to enhance the tokenization and training process. With adaptive tokenization, the system determines the region sizes based on the content of each image, allowing for better representation of complex visual features. Additionally, by integrating learnable position embeddings, the system gains the capability to learn the spatial relationships between different tokens during training. This combination enables the system to adaptively capture visual features at varying scales and spatial relationships, potentially enhancing its ability to generate high-quality haze-free images with improved spatial awareness and representation of visual features. Also, we introduce a dynamic learning rate scheduling technique, a one-cycle learning rate schedule, rather than using a static learning rate decay. This can help in potentially finding better model optima and improving convergence speed. Our model utilizes an initial learning rate of 0.00036. Evaluation is quantitatively performed using PSNR and SSIM metrics.

4.3 Comparisons

The proposed approach has been evaluated and compared with the original model based on SwinIR by liang et al. [9] and the approach presented in [13]. Tables 1 and 2 present the results obtained from all approaches for each dataset. Figures 3 and 4 show qualitative results with all the approaches evaluated with the corresponding datasets. The proposed method demonstrates superior haze removal capabilities on all tested data sets. This corroborates the effectiveness in addressing the challenges of dehazing

Table 1 Results from the validation datasets (NH-Haze 2020)

Approaches	NH-Haze 2020	
	PSNR	SSIM
Liang et al. [9]	17.9170	0.6065
Suárez et al. [13]	18.4038	0.6371
Ours	**18.6721**	**0.6856**

Best results in **bold**

Table 2 Results from the validation datasets (NH-Haze 2021)

Approaches	NH-Haze 2021	
	PSNR	SSIM
Liang et al. [9]	19.4660	0.8000
Suárez et al. [13]	20.1259	0.8273
Ours	**21.6069**	**0.8421**

Best results in **bold**

Fig. 3 Experimental results: 1st. row depicts input hazy images; 2nd. and 3rd. rows show results of state-of-the-art approaches; 4th. row shows results of the proposed approach; 5th. row shows ground truth images from NH-Haze 2020

Fig. 4 Experimental results: 1st. row depicts input images; 2nd. and 3rd. rows show results of state-of-the-art approaches; 4th. row shows results of the proposed approach; 5th. row shows ground truth images from NH-Haze 2021

in non-homogeneous images, especially with the NHAZE 2021 datafet, where the visual results, see Fig. 4 and the quantitative results, see Table 2 demonstrate that the proposed method is effective in haze removal tasks.

5 Conclusions

The current work demonstrates the efficacy of the proposed enhancements, including adaptive tokenization and learnable position embeddings, in improving the spatial self-attention mechanism for haze removal. By integrating these techniques, we have achieved notable advancements in the quality and efficiency of haze-free image generation using the SwinIR model. This opens avenues for further exploration and refinement of transformer-based models in image processing tasks, providing a promising foundation for future research and development in this domain.

Acknowledgements This work has been partially supported by the Grant PID2021-128945NB-I00 funded by MCIN/AEI/10.13039/501100011033 and by "ERDF A way of making Europe"; and by the ESPOL project CIDIS-12-2022. The authors acknowledge the support of the Generalitat de Catalunya CERCA Program to CVC's general activities, and the Departament de Recerca i Universitats from Generalitat de Catalunya with reference 2021SGR01499.

References

1. Ancuti CO, Ancuti C, Timofte R (2020) Nh-haze: an image dehazing benchmark with non-homogeneous hazy and haze-free images. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops, pp 444–445
2. Ancuti CO, Ancuti C, Vasluianu FA, Timofte R (2020) Ntire 2020 challenge on nonhomogeneous dehazing. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops, pp 490–491
3. Ancuti C, Ancuti CO, Timofte R (2018) Ntire 2018 challenge on image dehazing: Methods and results. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp 891–901
4. Dong H, Pan J, Xiang L, Hu Z, Zhang X, Wang F, Yang MH (2020) Multi-scale boosted dehazing network with dense feature fusion. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 2157–2167
5. Engin D, Genç A, Kemal Ekenel H (2018) Cycle-dehaze: enhanced cyclegan for single image dehazing. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp 825–833
6. Fattal R (2008) Single image dehazing. ACM Trans Graphics (TOG) 27(3):1–9
7. He K, Sun J, Tang X (2010) Single image haze removal using dark channel prior. IEEE Trans Pattern Anal Mach Intell 33(12):2341–2353
8. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ (2017) Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4700–4708
9. Liang J, Cao J, Sun G, Zhang K, Van Gool L, Timofte R (2021) Swinir: image restoration using swin transformer. arXiv preprint arXiv:2108.10257
10. Liu C, Ye S, Zhang L, Bao H, Wang X, Wu F (2022) Non-homogeneous haze data synthesis based real-world image dehazing with enhancement-and-restoration fused CNNs. Comput Graphics 106:45–57
11. Smith LN (2017) Cyclical learning rates for training neural networks. In: 2017 IEEE winter conference on applications of computer vision (WACV). IEEE, pp 464–472
12. Suárez PL, Carpio D, Sappa AD (2021) Non-homogeneous haze removal through a multiple attention module architecture. In: Advances in visual computing: 16th international symposium, ISVC 2021, virtual event, October 4–6, 2021, Proceedings, Part II. Springer, pp 178–190
13. Suárez PL, Carpio D, Sappa AD, Velesaca HO (2022) Transformer based image dehazing. In: 2022 16th international conference on signal-image technology and internet-based systems (SITIS). IEEE, pp 148–154
14. Yu M, Cherukuri V, Guo T, Monga V (2020) Ensemble dehazing networks for non-homogeneous haze. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops, pp 450–451

15. Zhu JY, Park T, Isola P, Efros AA (2017) Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE international conference on computer vision, pp 2223–2232
16. Zhu Q, Mai J, Shao L (2015) A fast single image haze removal algorithm using color attenuation prior. IEEE Trans Image Process 24(11):3522–3533