HU-Net: A Hybrid U-Net for Single Image Dehazing using Composite Attention Module

Banala Revanth
Department of computer science
Babasaheb Bhimrao Ambedkar University
Lucknow, India
revanthbanala302@gmail.com

Manoj Kumar
Department of computer science
Babasaheb Bhimrao Ambedkar University
Lucknow, India
mkjnuiitr@gmail.com

Sanjay kumar Dwivedi
Department of computer science
Babasaheb Bhimrao Ambedkar University
Lucknow, India
skd200@yahoo.com

***Abstract*— Image dehazing plays a crucial role in computer vision applications encompassing categorization, object identification, and picture restoration in adverse weather circumstances. Dehazing is a problem that lacks a clear solution, and to address it, many methods such as transformers and convolutional neural networks (CNNs) have been suggested. The transformer-based and CNN-based methods employ distinct approaches to extract features and yield disparate representations of features for a particular image. The performance of transformer and CNN-based approaches vary due to differences in their properties. This study introduces a composite attention block for Single-image dehazing, which aims to integrate the performance of both approaches. The attention block is a deep learning method that selectively focuses on specific areas of the input to enhance accuracy and computational efficiency. In this case, we employ a composite attention block that combines CNN-based and vision-based attention models to extract features. The entirety of our network is referred to as Hybrid U-Net (HU-Net), and it is constructed using the U-Net design. The HU-Net consists of an encoder and decoder, both equipped with skip connections. Additionally, it incorporates a fusion model for feature fusion. The HU-Net consists of blocks that utilize a composite attention module instead of the Multi-Head Attention Network to extract features. The HU-Net is trained using a hybrid loss function (HLF) on three different picture datasets: O-Hazy, I-Hazy, and NH-Hazy. The suggested network demonstrates superior performance concerning PSNR and SSIM when compared to other established techniques.**

Keywords*— *Attention, U-Net, CNN, Vision Transformers, Hybrid Attention, Image Fusion.

I. Introduction

Image dehazing is the method of clearing atmospheric haze, such as fog, smoke, or rain, from an image captured during bad weather conditions. Dehazing is a crucial process in computer vision applications, including recognizing objects, recognizing patterns, and image enhancement, etc., particularly when working with photos captured in a cloudy environment. The obscured image is produced utilizing the atmospheric scattering model (ASM), as formulated by Narasimhan et al. [1]

$$Haze\ (u) = Clear\ (u)\tau\ (u) + \kappa\ (1 - \tau\ (u)) \tag{1}$$

$$\tau\ (u) = e^{-\lambda d(u)} \tag{2}$$

Where, *Haze(u)* is the haze image, *Clear(u)* is the clean image, *τ(u)* is the transmission map, *κ* is the global atmospheric light factor, $d(u)$ is the distance from object to camera, λ is the haze coefficient. To estimate *Clear(u),* dehazing methods aim to estimate the unknowns *τ(u)* and *κ*, which can be rewritten as:

$$Clear(u) = \frac{Haze(u) - \kappa}{\tau(u)} + \kappa \tag{3}$$

Many image dehazing (ID) approaches are offered to employ priories, CNN and GANs-based methods give good results, however owing to a lack of attention to specific regions of the hazy image (HI), the generated output suffers from blurring or reduced dehazing in some portions of the dehazed image. The Vision Transformer (ViT) is an attention-based neural network designed for image-processing tasks across several applications. ViTs, or Vision Transformers, are a computer vision adaptation of transformers, which are commonly employed for natural language processing tasks. Typically, transformers possess the capability to employ self-attention, where the entire input is treated as an attention map. The ViT possesses the capacity to generate long-range dependencies, which are employed for computer vision applications. Several applications, including classification, segmentation, and image reconstruction, have utilized various ViT-based approaches

The Vision Transformer (ViT) partitions the input image into many patches. Every patch contains a position, which is a class embedding that may be learned. The vision encoder receives image patches, coupled with location embedding, as its input. The visual encoder comprises three components: layer norm (LN), multi-head attention network (MSP), and multi-layer perceptron (MLP).

The primary component of the ViT encoder for image reconstruction is the MSP, responsible for extracting features for image processing. Various investigators have concluded that, in certain cases, combining hybrid techniques to extract features instead of using MSP can provide favorable results. Some approaches employ convolutional neural network (CNN)-based attention networks, while others consist of various types of vision transformers. The U-Net is a deep learning network designed in the shape of a "U" that utilizes convolutional neural networks (CNNs) for image processing applications. The U-Net architecture has made significant contributions to the field of picture dehazing, with multiple variations of the U-Net being utilized. The U-Net consists of three main components: the encoder, decoder, and skip-connections. The encoder is the first element of the network, comprising one or more CNNs or CNN

979-8-3503-7471-1/24/$31.00 ©2024 IEEE

blocks to extract visual features from the input image. The decoder is the counterpart of the encoder in the U-Net architecture, operating in the reverse manner. It generates the output image by utilizing image attributes as input. The encoder and decoder in U-Net are linked together by skip connections, which form the third component of the architecture. The skip connections are established between the encoder and decoder in a layer-wise manner to facilitate the transmission of information, therefore preserving spatial information while enhancing accuracy.

The contributions of the author are of three folds:

1) We have developed an Composite Attention Module for effective features extraction.

2) The Composite Attention Module is designed to extract two distinct characteristics using two separate sub-modules (CNN based and Vision based), which are then combined using the fusion module to effectively extract features for single image dehazing.

3) We created a new end-to-end trainable vision transformer-based U-Net for image dehazing.

The Composite Attention Module specifically intends to extract characteristics for single image dehazing. This attention module combines two distinct modules: one based on CNN and the other based on vision. The CNN utilizes a custom-developed sub-module that extracts features using depth-wise convolution layers and depth-wise dilation convolution layers. The Vision-based sub-module is developed utilizing linear transformers. Finally, the fusion module has successfully combined the two sub-modules. The fusion module combines the distinct features extracted from two independent modules in order to make an end-to-end trainable network.

II. RELATED WORK

A. Priorie based methods

Several techniques have been suggested for image dehazing by making assumptions about the underlying principles of haze production in adverse weather circumstances. Various priority-based techniques have been suggested for image dehazing. He et al. [2] Proposed a method that employs the dark channel prior and atmospheric scattering model. The dark channel prior postulates that inside a Haze Image, the local patches encompass pixels exhibiting diminished intensities in at least one of the color channels. The proposed method utilizes the dark channel before to estimate the transmission map and atmospheric light, then employs an atmospheric scattering model to recover the image without haze. Bui et al. [3] Implemented a new color line beforehand to improve the sharpness of an individual image by minimizing haze. The color line prior assumption states that in haze images, tiny groups of pixels exhibit a distribution that can be represented in a single dimension in the RGB color space. Ju et al. [4] introduced an approach that utilizes gamma correction prior to image dehazing. In this approach, the haze image undergoes gamma correction before estimating the virtual transformation of the haze image. The depth ratio of the haze image is determined by comparing the haze image with its estimated virtual transformation. The clear image is computed using a global approach and a visual indicator. The clear image produced by the prioritized-based approaches are affected by excessive saturation and halo artifacts, which result from the assumptions made about the haze images.

B. Learning-based methods

The learning-based techniques have significantly transformed the field of real-time image dehazing, with numerous ways being suggested. Revanth et al. [5] proposed a network specifically designed for the dehazing of non-homogeneous hazy images. The network is designed with two sub-networks to independently estimate the global atmospheric light factor and the refined transmission map from the provided haze image. It then utilizes a modified atmospheric scattering model, known as the haze formation model, to recover the haze-free image. Revanth et al. [6] introduced a technique that combines the dark channel prior method and Generative Adversarial Network employing a discrete wavelet transform-based image fusion algorithm for image dehazing. Zhang et al. [7] developed a dual-path recurrent network designed specifically for image dehazing. The dual-path recurrent network simultaneously acquires fundamental content and picture particulars. The network consists of a number of convolutional LSTM blocks, with each block containing CNN layers. Liu et al. [8] introduced a comprehensive generative network for dehazing that uses residual learning. Engin et al. [9] presented Cycle-Dehaze, a method for dehazing that utilizes a Generative Adversarial Network (GAN) that combines cycle consistency with perceptual losses. Raj et al. [10] proposed an enhanced version of the conditional Generative Adversarial Network (CGAN) by incorporating the U-Net model and a patch discriminator for image dehazing. Ebenezer et al. [11] introduced the Wasserstein Generative Adversarial Network as a method for image dehazing. The network employs the Wasserstein loss function, which incorporates gradient penalties to enforce the Lipschitz constraint. Malav et al. [12] developed a DHSGAN that utilizes a Generative Adversarial Network to rebuild fog and smoke images. Li et al. [13] introduced a network model for image dehazing that utilizes a recurrent neural network. The model divides the network into two phases: The first phase of the process involves taking a hazy image as input and generating a transmission map as output. while the subsequent phase takes ratio of the haze image and the transmission map as input and utilizes a residual network to provide a dehazed image as output. Liu et al. [14] employed the CycleGAN network to perform image dehazing. They utilized haze photos and haze-free images from MRFID datasets as training data. Ullah et al. [15] introduced LD-Net is a novel approach for image dehazing that integrates the transmission map and atmospheric light into a unified entity. Mehra et al. [16] created TheiaNet, a dehazing technology that utilizes free-form intermediate computation to calculate the transmission map. TheiaNet incorporates a spatial cleaning bottleneck block to facilitate effective feature extraction. Xiao et al. [17] introduced a comprehensive network that utilizes convolutional neural networks to estimate haze layers. These haze layers are essentially graphical representations that depict the correlation between a hazy image and a clear image. Mei et al. [18] introduced a U-Net model that utilizes progressive feature fusions to accurately convert a hazy image into a clear image by extracting relevant features.

Gandelsman et al. [19] defined Deep-image-Prior (DIP) as a method for picture dehazing. Santra et al. [20] introduced a network that is trained end-to-end for image dehazing. The network receives a hazy picture as input and simultaneously predicts the transmission map and atmospheric light to remove haze from the image.

C. Vision Transformers (ViTs)-based methods

ViTs elevate the performance of vision applications in comparison to CNN. Several techniques were suggested for identification utilizing Vision Transformers (ViTs). Tu et al. [21] suggest replacing the standard MLP in the ViT with a multi-axis MLP-based architecture dubbed MAXIM. Maxim is a network that uses the U-Net to perform image dehazing. Ji et al. [22] introduce a U2-Former model that is built around a Vision Transformer (ViT). The U2-Former is a neural network that utilizes the U-Net architecture and is specifically developed for dehazing. Dong et al. [23] introduce TransRA, a hybrid model that merges the Transformer and residual attention techniques for image dehazing. TranRA is constructed using the U-Net model as its foundation. Zhang et al. [24] defined a comprehensive network for picture dehazing that utilizes the Residual Mixed-convolution Attention Module (RMAM).

Prior-based models utilize predetermined priors to estimate the unobserved element of the Atmospheric Scattering Model (ASM) and produce a clear image from a hazy one. Conversely, learning-based methodologies like GANs and U-Net directly convert a hazy picture into a clear one by removing the haze. Although learning-based methods surpass apriori-based methods, the pivotal factor is in the intrinsic components of the methods, which are important for extracting visual information. The technique of feature extraction can vary depending on whether it is based on prior information or learning methodologies. Novel fusion-based methodologies were introduced to amalgamate image feature extraction techniques to integrate this image. Fusion-based solutions outperform alternative approaches. We introduced a hybrid U-Net architecture that integrates feature fusion between CNN and ViT-based approaches.

III. PROPOSED WORK

A hybrid network capable of substituting the attention mechanism (MSP) in ViTs with a corresponding attention block composed of generic CNNs. A composite attention module is established for single picture dehazing in which a CNN and a vision transformer-based attention module are combined in each block of the U-Net. The suggested U-Net architecture is referred to as Hybrid U-Net (HU-Net). In this architecture, a hazy image is inputted into the HU-Net, which consists of BLOCK layers. The result produced by the HU-Net is an image that is free from haze. The configuration of HU-Net is depicted in Fig. 1(a). HU-Net comprises an encoder and decoder that are provided with skip connections.

The provided haze image is processed by the encoder and decoder to produce a haze-free image. The encoder is composed of six levels. The first level is a Convolutional Neural Network (CNN), while the second level consists of a composite attention module block called a BLOCK. Furthermore, there are downsampling layers positioned at the third and fifth levels. The BLOCK layer is present in the fourth and sixth layers. The input image is processed by the encoder network, which consists of six levels: a CCN layer, a BLOCK layer, a downsample layer, another BLOCK layer, another downsample layer, and a final BLOCK layer. These layers work together to extract the features of the image. The extracted characteristics are transmitted through the decoder network to generate the resulting image.

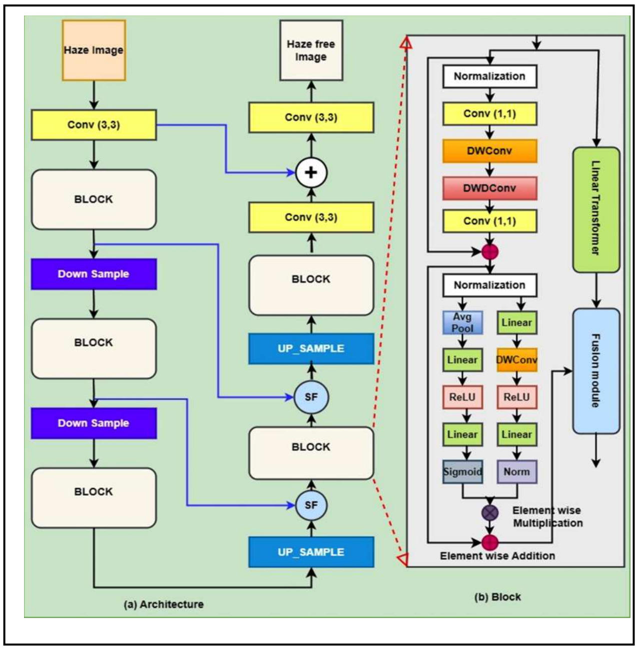

Fig. 1. Hybrid U-Net (HU-Net) for single image dehazing

The decoder with skip connections comprises nine stages. The first level is the up-sampling layer, followed by the self-attention fusion layer ($SF\ layer$), which is an adapted version of SK Fusion [5] at the second level. The BLOCK layers are present at both the third and sixth levels. The fifth level comprises the self-attention fusion layer. The fourth level has an up-sampling layer, while the seventh level consists of CNN block layers. The eighth level, on the other hand, has a summation layer. The ninth levels incorporate CNN layers to perform image reconstruction. The skip connections refer to the connections established between the relevant levels of the encoder and decoder. Specifically, the layers involved in these skip connections are the BLOCK layers in the encoder and the BLOCK layers in the decoder. The subsequent subsections explain the BLOCK and self-attention fusion layer.

A. Transformer Block

Computer vision applications often utilize Vision Transformers (ViTs) which consist of encoder and decoder blocks, each containing attention blocks (MSP). We developed a transformer block that incorporates a composite attention module instead of the MSP to attain equivalent or superior outcomes for image dehazing tasks. The transformer block is denoted as BLOCK in Fig. 1(a). The BLOCK has three sub-modules: a sequence of Convolutional Neural Networks (CNNs) referred to as

CNN module [25], a linear transformer [26] , and. a fusion module [27]. The CNN module and linear transformer operate concurrently to acquire the characteristics, while the fusion module integrates these characteristics. The BLOCK within the HU-Net is illustrated in Fig. 1(b). The image or features are input for into the BLOCK, where the input is duplicated to create two identical inputs. The same inputs are sent to two sub-modules simultaneously, where each module is responsible for extracting the features. The retrieved features are combined using the third sub-module to obtain effective features for single image dehazing. The sub-modules are further explained as follows.

The CNN module comprises two residual blocks specifically intended for feature acquisition. The first residual block is intended to extract intricate features from the input, whereas the subsequent residual block aims to improve the extracted intricate features. The first block consists of five layers, while the second block consists of eleven layers. In both residual blocks, the first layer is a normalization layer, which is responsible for normalizing the data with regard to the batch size in order to enhance performance. The next four layers in the first block are CNN layers. The network consists of four layers: the first layer is composed of convolution layers, the second layer is made up of depth-wise convolution layers, the third layer consists of depth-wise dilation convolution layers, and the fourth layer is composed of convolution layers. The CNNs in the first residual block are responsible for extracting intricate and specific features. The intricate features are sent via the summing layer for residual learning. The remaining ten layers in the second block are divided into two sub-blocks, each consisting of five levels that are linked in parallel to the normalization layer (first layer). The features from the normalization layer are processed via the first sub-block, which includes an average pooling layer, a linear layer, a ReLU layer, another linear layer, and a sigmoid layer. This sub-block is responsible for generating an effective attention map. The second sub-block comprises a linear layer, depth-wise convolution layers, ReLU layer, linear layer, and a norm layer for the purpose of feature extraction. The first sub-block and subsequent sub-block underwent an element-wise multiplication layer, followed by a summing layer for residual learning.

The Self-attention is an essential element of Vision Transformer architectures, enabling models to capture distant relationships in images. The Linformer model devised a modification of the self-attention mechanism to attain linear computing complexity in relation to the input size, hence enhancing efficiency and scalability. The main concept of Linformer is to mimic the conventional self-attention mechanism by constraining the attention matrix to a low-rank structure, hence minimizing the number of parameters and computations needed. Linformer achieves a large reduction in memory and computing needs by using low-rank approximations, all while preserving the expressive capabilities of self-attention. This enables the training of Transformer models on more extensive image datasets without compromising performance or necessitating exorbitant computer resources.

A linear transformer is a type of transformer that has a time complexity of $O(n)$, where n represents the size of the input. The linear transformer is characterized by three variables, namely $V, K,$ and Q. These variables represent characteristics that have been extracted from the input picture. The Scaled Dot Product is utilized to acquire discerning characteristics. Multiple scaled dot product layers are combined and then processed through the CNN layers to facilitate feature learning. The Linear transformer extracts features from the input image and generates the features as output, as depicted in Fig. 2 (b).

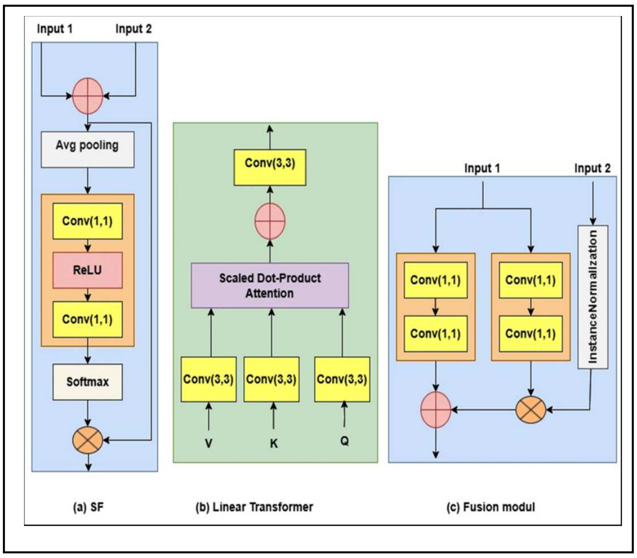

Fig. 2. The constituent elements of the Hybrid U-Net (HU-Net)

The fusion module is a network that integrates the features of the CNN module with the Linear transformer. The fusion module accepts two inputs and generates a single output, as depicted in Fig. 2(c). Input 1 and Input 2 refer to the CNN module and linear transformer characteristics, respectively. The features of the CNN module are fed into two parallel CNN blocks, each consisting of two CNN layers. The output of one CNN block is multiplied by the linear transformer features after going through an instance normalization layer. The output of another CNN block and the multiplied output is fed into the summing layer to extract features.

B. Self-Attention Fusion Layer ($SF\ layer$)

Typically, the U-Net architecture consists of an encoder, a decoder, and a skip connection (SC) linking the encoder and decoder. The skip connection is employed to directly input the features from the encoder to the decoder, hence mitigating the training cost. The skip connection can be adjusted in several ways depending on the applications. One approach is to utilize the general addition layer, while another approach involves employing CNNs to incorporate both the encoder and decoder layer outputs as input to the skip connection layer. The output of the skip connection layer is passed as input to the top decoder layer. The proposed network replaces the skip connection layer with a self-attention fusion layer (SF layer) to enhance feature extraction. The self-attention fusion layer is a layer that combines and extracts features from both input layers, resulting in efficient features as output. The fusion layer consists of five layers, as depicted in Fig. 2(a). The initial layer is the concatenation layer, succeeded by average pooling layers and a CNN block as the second and third layers. The SoftMax layer is located at the fourth layer, whereas the element-wise multiplication layer is positioned at the fifth layer for learning residuals.

IV. EXPERIMENT AND DISCUSSION

A. Datasets and Experimental setting

The HU-Net is evaluated using the I-HAZE, O-HAZE, and NR-Indoor datasets. The I-HAZE dataset consists of 30 haze image (HI) and an equal number of clean images (CI). O-HAZE consists of 45 haze images, together with corresponding clean images and indoor (NR-Indoor) datasets have 1346 images containing CI and HI. The HU-Net is tested on the Google Colab platform with the following parameters: the input picture size is 224×224, the training is done for 200 epochs, the optimizer used is Adam with learning rates $\beta_1 = 0.9$ and $\beta_2 = 0.999$, and the loss function employed is the hybrid loss function (HLF). The NR-Indoor dataset is divided into two separate datasets: the training dataset and the testing dataset. The HU-Net is trained using the NR-Indoor training dataset and assessed using the I and O HAZE datasets, together with the NR-Indoor testing dataset. HLF is a loss function that merges the Mean Squared Error (MSE) with perceptual loss.

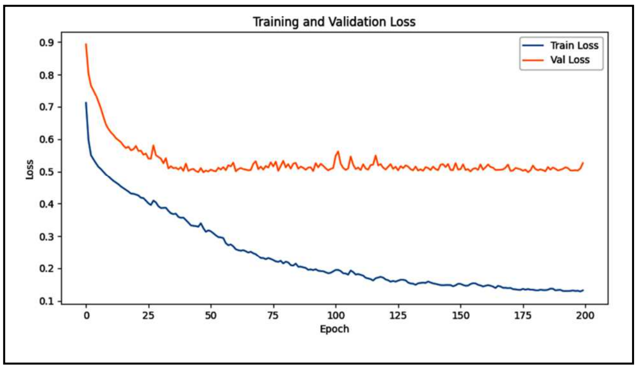

Fig. 3. Training and Validation loss graph with respective to epochs

The Fig. 3 shows the training and validation loss graph with respect to epochs. The graph represents that as epochs increase, training and validation losses decrease. The performance of our HU-Net is sufficiently better when we consider the number of epochs 200 and the constrained dataset size. By increasing the total number of epochs and images employed for training, the performance of HU-Net improves.

B. Results

The performance assessment of the proposed HU-Net based on the I-HAZE, O-HAZE, and NR-Indoor datasets are compared with general approaches such as DCP [2], CEP [3], IDRBD [13], SMID [28], IDE [29], ICAP [30], and RADE [31] in TABLE I. The re-implementation and testing of these approaches on the datasets are conducted to compare them with the proposed HU-Net in terms of PSNR and SSIM. The RADE achieved the best performance on the I-HAZE dataset, as measured by PSNR and ICAP is the best in terms of SSIM. ICAP achieved the highest score in terms of PSNR while SMID performed very well in SSIM on the O-HAZE dataset. Among the NR-INDOOR dataset, CEP obtained the best score concerning PSNR, while DCP achieved the highest result about SSIM. HU-Net obtained the second-best results in PSNR and third best score in terms SSIM on I-HAZE dataset. On O-HAZE dataset our HU-Net achieved the fourth best score in terms of PSNR and second best score in terms of SSIM. On the other hand, HU-Net obtained the second highest score in PSNR and fifth best score in terms of SSIM on NR-INDOOR dataset.

TABLE I. PSNR AND SSIM RESULTS FOR THE I-HAZE. O-HAZE, AND NR-INDOOR DATASETS.

Methods	**I-HAZE**		**O-HAZE**		**NR-INDOOR**	
	PSNR	***SSIM***	***PSNR***	***SSIM***	***PSNR***	***SSIM***
DCP	12.14	0.523	10.68	0.389	11.85	**0.754**
CEP	12.30	0.454	12.29	0.397	**12.70**	0.682
IDRBD	12.37	0.290	14.28	0.270	6.99	0.168
SMID	13.33	0.532	12.36	**0.456**	10.26	0.655
IDE	12.74	0.355	13.59	0.300	11.88	0.699
ICAP	13.93	**0.568**	**14.33**	0.430	11.50	0.569
RADE	**14.66**	0.551	11.70	0.411	10.24	0.523
HU-Net	14.45	0.537	13.46	0.447	12.38	0.628

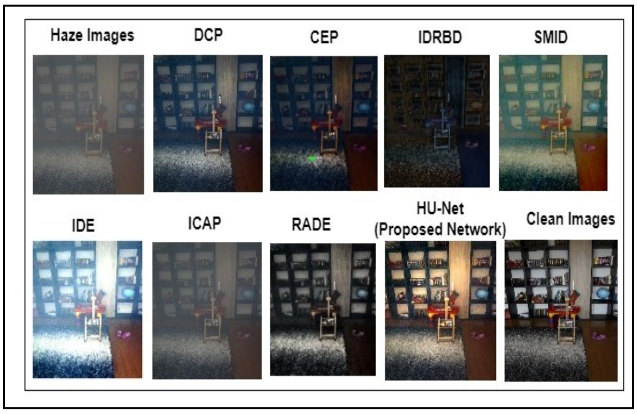

Fig. 4. Visual comparision of proposed network along with exiting network

From the TABLE I and Fig. 4 on average our HU-Net yields superior outcomes compared to the others models for single image dehazing. Therefore, the HU-Net model performs second best with I-Haze and has the forth-best efficiency with O-haze, and second best performance than other approaches when evaluated on the NR-indoor dataset. It can be concluded that the suggested HU-Net yields superior outcomes in comparison to alternative approaches. HU-Net is trained on a small dataset, while deep learning-based techniques yield superior results when provided with a large dataset for training, together with an increased number of training iterations and an effective optimization algorithm coupled with an appropriate loss function. The visual results of proposed method and other existing network are given in Fig. 4 along with input image (haze image) and corresponding haze-free image.

V. CONCLUSION

The HU-Net is a newly developed network that replaces the Multi-Head Attention Network in ViTs-based U-Nets with composite attention module. The HU-Net is trained using hybrid loss function on the NR-indoor dataset. The performance of HU-Net is assessed using PSNR and SSIM metrics. The proposed strategy yields satisfactory results comparable to the conventional methods. The results presented here are based on a training period of just 200 epochs. It is vital to remember that increasing the number of epochs may lead to improved outcomes. Future research may explore the utilization of the knowledge distillation method to develop image dehazing.

REFERENCES

[1] M S. G. Narasimhan and S. K. Nayar, “Contrast restoration of weather degraded images,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 25, no.6, pp. 713–724, Jun. 2003.

[2] He, Kaiming, Jian Sun, and Xiaoou Tang. ”Single image haze removal using dark channel prior.” IEEE transactions on pattern analysis and machine intelligence 33, no. 12 (2010): 2341-2353.

[3] Bui, Trung Minh, and Wonha Kim. ”Single image dehazing using color ellipsoid prior.” IEEE Transactions on Image Processing 27, no. 2 (2017):999-1009.

[4] Ju, Mingye, Can Ding, Y. Jay Guo, and Dengyin Zhang. ”IDGCP: Image dehazing based on gamma correction prior.” IEEE Transactions on Image Processing 29 (2019): 3104-3118.

[5] Revanth, Banala, Manoj Kumar, and Sanjay K. Dwivedi. ”Non-Homogeneous Haze Image Formation Model Based Single Image Dehazing.” In 2023 3rd International Conference on Innovative Sustainable Computational Technologies (CISCT), pp. 1-6. IEEE, 2023.

[6] Revanth, Banala, Sanjay K. Dwivedi, and Manoj Kumar. ”A FrameworkFor Single Image Dehazing Using DWT Based Cross Bilateral Filter Fusion of Generative and ASM Models.” In 2022 2nd International Conference on Innovative Sustainable Computational Technologies (CISCT),pp. 1-6. IEEE, 2022.

[7] Zhang, Xiaoqin, Runhua Jiang, Tao Wang, and Wenhan Luo. ”Single image dehazing via dual-path recurrent network.” IEEE Transactions on Image Processing 30 (2021): 5211-5222.

[8] Liu, Zheng, Botao Xiao, Muhammad Alrabeiah, Keyan Wang, and Jun Chen. ”Generic model-agnostic convolutional neural network for single image dehazing.” arXiv preprint arXiv:1810.02862 (2018).

[9] Engin, Deniz, Anil Genc¸, and Hazim Kemal Ekenel. ”Cycle-dehaze: Enhanced cyclegan for single image dehazing.” In Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp. 825-833. 2018.

[10] Raj, N. Bharath, and N. Venketeswaran. ”Single image haze removal using a generative adversarial network.” In 2020 International Conference on Wireless Communications Signal Processing and Networking (WiSPNET), pp. 37-42. IEEE, 2020.

[11] Ebenezer, Joshua Peter, Bijaylaxmi Das, and Sudipta Mukhopadhyay. ”Single image haze removal using conditional wasserstein generative adversarial networks.” In 2019 27th European Signal Processing Conference (EUSIPCO), pp. 1-5. IEEE, 2019.

[12] Malav, Ramavtar, Ayoung Kim, Soumya Ranjan Sahoo, and Gaurav Pandey. ”DHSGAN: An end to end dehazing network for fog and smoke.” In Asian conference on computer vision, pp. 593-608. Springer, Cham, 2019.

[13] Li, Jinjiang, Guihui Li, and Hui Fan. ”Image dehazing using residual based deep CNN.” IEEE Access 6 (2018): 26831-26842.

[14] Liu, Wei, Xianxu Hou, Jiang Duan, and Guoping Qiu. ”End-to-end single image fog removal using enhanced cycle consistent adversarial networks.” IEEE Transactions on Image Processing 29 (2020): 7819-7833.

[15] Ullah, Hayat, Khan Muhammad, Muhammad Irfan, Saeed Anwar, Muhammad Sajjad, Ali Shariq Imran, and Victor Hugo C. de Albuquerque. ”Light-DehazeNet: a novel lightweight CNN architecture for single image dehazing.” IEEE transactions on image processing 30 (2021): 8968-8982.

[16] Mehra, Aryan, Pratik Narang, and Murari Mandal. ”TheiaNet: Towards fast and inexpensive CNN design choices for image dehazing.” Journal of Visual Communication and Image Representation 77 (2021): 103137.

[17] Xiao, Jinsheng, Mengyao Shen, Junfeng Lei, Jinglong Zhou, Reinhard Klette, and HaiGang Sui. ”Single image dehazing based on learning of haze layers.” Neurocomputing 389 (2020): 108-122.

[18] Mei, Kangfu, Aiwen Jiang, Juncheng Li, and Mingwen Wang. ”Progressive feature fusion network for realistic image dehazing.” In Computer Vision–ACCV 2018: 14th Asian Conference on Computer Vision, Perth, Australia, December 2–6, 2018, Revised Selected Papers, Part I 14, pp. 203-215. Springer International Publishing, 2019.

[19] Gandelsman, Yosef, Assaf Shocher, and Michal Irani. ”” Double-DIP”: unsupervised image decomposition via coupled deep-image-priors.” In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11026-11035. 2019.

[20] Santra, Sanchayan, Ranjan Mondal, Pranoy Panda, Nishant Mohanty, and Shubham Bhuyan. ”Image Dehazing via Joint Estimation of Transmittance Map and Environmental Illumination.” In 2017 Ninth International Conference on Advances in Pattern Recognition (ICAPR), pp.1-6. IEEE, 2017.

[21] Tu, Zhengzhong, Hossein Talebi, Han Zhang, Feng Yang, Peyman Milanfar, Alan Bovik, and Yinxiao Li. ”Maxim: Multi-axis mlp for image processing.” In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5769-5780. 2022.

[22] Ji, Haobo, Xin Feng, Wenjie Pei, Jinxing Li, and Guangming Lu. ”U2- former: A nested u-shaped transformer for image restoration.” arXiv preprintarXiv:2112.02279(2021).

[23] Dong, Pengwei, and Bo Wang. ”TransRA: transformer and residual attention fusion for single remote sensing image dehazing.” Multidimensional Systems and Signal Processing 33, no. 4 (2022): 1119-1138.

[24] Zhang, Xiaoqin, Tao Wang, Wenhan Luo, and Pengcheng Huang. ”Multi-level fusion and attention-guided CNN for image dehazing.”IEEE Transactions on Circuits and Systems for Video Technology 31,no. 11 (2020): 4162-4173.

[25] Luo, Pinjun, Guoqiang Xiao, Xinbo Gao, and Song Wu. ”LKD-Net: Large Kernel Convolution Network for Single Image Dehazing.” arXiv preprint arXiv:2209.01788 (2022).

[26] Wang, Sinong, Belinda Z. Li, Madian Khabsa, Han Fang, and Hao Ma. ”Linformer: Self-attention with linear complexity.” arXiv preprint arXiv:2006.04768 (2020).

[27] Guo, Chun-Le, Qixin Yan, Saeed Anwar, Runmin Cong, Wenqi Ren, and Chongyi Li. ”Image dehazing transformer with transmission-aware 3D position embedding.” In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5812-5820. 2022.

[28] Van Nguyen, Thuong, Truong Thanh Nhat Mai, and Chul Lee. ”Single maritime image defogging based on illumination decomposition usingtexture and structure priors.” IEEE Access 9 (2021): 34590-34603.

[29] Ju, Mingye, Can Ding, Wenqi Ren, Yi Yang, Dengyin Zhang, and Y. Jay Guo. ”IDE: Image dehazing and exposure using an enhanced atmospheric scattering model.” IEEE Transactions on Image Processing30 (2021): 2180-2192.

[30] Ngo, Dat, Gi-Dong Lee, and Bongsoon Kang. "Improved color attenuation prior for single-image haze removal." Applied Sciences 9, no. 19 (2019): 4011.

[31] Li, Zhan, Xiaopeng Zheng, Bir Bhanu, Shun Long, Qingfeng Zhang, and Zhenghao Huang. "Fast region-adaptive defogging and enhancement for outdoor images containing sky." In 2020 25th International Conference on Pattern Recognition (ICPR), pp. 8267-8274. IEEE, 2021.