

Single Image Dehazing using Alternate Pooling Fused Transformer Block with Attention Network

Abstract

Haze decreases contrast and limit sight in both outdoor and indoor images. Each pixel's deterioration is unique and is influenced by how far the scene point is from the camera. The transmission coefficients, which regulate the scene attenuation and degree of haze in each pixel, express this dependence. Previous techniques used a variety of patch-based priors and transformers to solve the single image dehazing problem. Although various researches had demonstrated the effectiveness of vision Transformers, our image dehazing method has been able to surpass the state-of-the-art image dehazing networks. As a result, we proposed a novel image dehazing network named Alternate Pooling Fused Transformer Network (APF_TRANS_NET) with Locally Grouped Self Attention. Compared to earlier deep learning-based methods, it performs far better. The proposed approach enhances the ability of vision transformer in Dehazing progress with an efficient transformer along with the dual weighted deep channel and spatial attention mechanism. To show the efficiency of our model, we trained it on five different datasets, including i-Haze dataset, O-Haze dataset, SOTS dataset, RESIDE-6K, and RS-Haze. The proposed our immense model outperforms the prior state-of-the-art techniques, with a significant improvement in its performance.

Keywords: Single Image Dehazing, Transformer ,Attention, Deep Learning.

1. Introduction

Nearly all research in vision is based on the presumption that the observer is submerged in a transparent medium (air). The reflected light rays from scene objects are believed not to change or weaken as they travel to the observer. This assumption holds that a scene point's brightness alone determines the brightness of an image point. Simply told, current algorithms and visual sensors are only intended to function during "clear" days. A trustworthy vision system, however, needs to be able to operate in all types of weather, including haze, fog, rain, hail, and snow [1].Haze removal, generally known as dehazing, is highly wanted in consumer/computational image and applications for computer vision. First, removing the haze will improve visibility and correct any color shift brought on by the air light. In general, the image without haze is more aesthetically pleasing. Second, most computer vision algorithms assume that the scene radiance is the input image after radiometric calibration, from low-level image analysis to high-level object detection. Many vision algorithms, including feature detection, filtering, and photometric analysis, will dependably perform poorly due to the biased and low-contrast scene illumination. Lastly, haze removal enables a variety of advanced image-editing tools and vision algorithms to access depth information. Fog can be a helpful depth cue for interpreting a scene. It is feasible to

utilize a poor, fuzzy image[2].The amount of image deterioration brought on by haze rises with distance from the camera because air light magnitude rises while scene radiance falls. As a result, hazy images can be represented as a per-pixel convex mixture of a haze-free image and the ambient light. They recovered the RGB values and transmission—the convex combination coefficient for each pixel in the haze-free image [3].

Light is scattered by aerosols, which are clouds of minute liquid or dust particles hanging in the atmosphere. The straight scene transmission that was earlier dispersed by this light deflection is replaced with air light, generally known as veiling light. Because of this, images captured in foggy or dusty conditions, as well as those captured at a distance on mostly clear days, often have poor contrast and have restricted scene visibility. Imaging underwater has a similar obstacle. In order to eliminate the haze layer, the bulk of image dehazing approaches restore the direct scene radiance. These methods are based on a physical picture generation concept where the hazy image is a convex mix of scene radiance and ambient light[4].Haze removal is a difficult task since the transmission of haze depends on the undetermined depth, which fluctuates at different positions. Several image enhancing techniques have been used to address the problem of reducing haze from a single image, notably histogram-based [5], contrast-based [6], and saturation-based [7]. There have also been suggested methods that leverage several pictures or depth information. Polarization-based approaches, for example, [8] remove the haze effect by taking many photos with varied degrees of polarization. Several applications of multi-constraint-based algorithms are made on various photographs of the same scene that were captured in various weather circumstances in [9]. For some depth information, depth-based approaches [10] need human input or 3D models that are well-known[11].

Hazy images have two main sources that need to be addressed. The first is the color cast that the airborne light causes. The second is the reduced vision brought on by attention. First, without assuming any restrictions on scene transmission or ambient light, Wenqi et al.,[12] presented a deep, adaptable neural network that can recover crisp images from start to finish. Second, they showed the value and effectiveness of a gated fusion network for single-image dehazing using the derived inputs from an initial hazy image. Finally, they used a multi-scale training technique for the suggested model to get rid of the halo artefacts that prevent image dehazing. GCANet, a novel end-to-end gated context aggregation network, for image dehazing is suggested byChen et al., [13]. In order to prevent gridding artefacts, this network employs a smoothed dilated convolution, and a gated sub network to aggregate features from various layers. Studies showed that GCANet beats the entire prior state-of-the-art image dehazing algorithms, both qualitatively and quantitatively. They also provide in-depth ablation studies to illustrate the value and necessity of each component. Additionally, they apply GCANet to the image de raining challenge, which demonstrates its generalizability and surpasses earlier cutting-edge image de raining techniques. The evident deterioration in object appearance and contrast in visual quality is largely caused by haze. Input images taken in hazy scenes have a big impact on high-level computer vision tasks like object detection [14, 15] and scene understanding [16, 17]. As a result, a lot of research has been done on image restoration using image dehazing to help develop effective computer vision systems[18].Zhao et al., [40] extract texture-level information for shallow features by stacking pixel and channel attention approaches. The multi-head self-attention (MHSA) methodology of the method is used to capture high-level features. MHSA improves dehazing performance by mining the dependencies of a wide variety of abstract data.

The superior feature extraction skills of the transformer architecture are strengthened by convolutions and cascaded attention approaches.

Deep learning has become very popular in computer vision in recent years, and many image dehazing techniques have been suggested by academics [18–21]. These techniques can outperform prior-based approaches with a large enough set of synthetic picture pairs. Many improved architectures [22–25] have since been put forth, and Transformer is now challenging CNNs' hegemony in complex vision tasks. ViT [26] used a simple Transformer architecture and outperformed practically all CNN architectures in high-level vision tests. Song et al., [27] proposed the Swin Transformer-inspired DehazeFormer, an image dehazing Transformer. It includes a number of enhancements, including an altered normalization layer, an improved activation function, and a new method for aggregating spatial information. Parihar et al., [28] put forth an effective architecture known as DCCT Densely Connected Convolutional Transformer for dehazing a single image where local dependencies and a multi-head Performer are combined. In high-level vision tests, the Prioritized Air light and Transmittance Extraction (PATE) model using Dual Weighted Deep Channel and Spatial Attention surpassed virtually all CNN architectures earlier. Although numerous studies have shown how effective vision Transformers are, there is yet no Transformer-based picture dehazing technique that can outperform cutting-edge image dehazing networks. In this work, we proposed a dehazing transformer for images called Alternate Pooling Fused Transformer with Dual weighted deep channel and Spatial attention. It performs far better than earlier deep learning based methods.

2. Related Work

Li et., [29] proposes AOD-Net, an all-in-one pipeline that directly reconstructs haze-free images using an end-to-end CNN. They compared AOD-Net against a number of state-of-the-art methods on photos with both manufactured and natural haze using both objective (PSNR, SSIM) and subjective assessments. Extensive experiment findings support the excellence, robustness, and effectiveness of AOD-Net. Also, they offer the first research of its type on how AOD-joint Net's pipeline tuning might enhance object recognition and detection on photos with natural haze. The approach taken in the study by Zhang et al., [30] to end-to-end learning for image dehazing is distinct. An image dehazing architecture known as the Densely Connected Pyramid Dehazing Network (DCPDN) allows for simultaneous estimation of the transmission map, ambient light, and picture dehazing while adhering to the image degradation model. Embedding Equation enables comprehensive learning, to put it another way. The network receives math operation modules directly through the deep learning framework in step 1. But training a network to do three complex jobs that are separate from one another is incredibly challenging. To hasten the training process and hasten network convergence, they used a stage-wise learning technique in which they first gradually optimize each component of the network before jointly optimizing the entire network. A new edge-preserving loss function to ensure that the estimated transmission map keeps sharp edges and avoids halo artefacts while dehazing, based on the discovery that gradient operators and the first few layers of a CNN structure can serve as edge extractors. A multilayer pooling module encoder-decoder network with dense connections is also recommended for the purpose of predicting the transmission map. To benefit from the structural connection between the transmission map and the dehazed image, a joint discriminator-based generative adversarial network (GAN) is suggested. The joint discriminator determines if a pair of estimated transmission map and dehazed image is genuine or not.

A trainable Grid DehazeNet Convolutional Neural Network (CNN) for single image dehazing was proposed by Yongrui et al., [31]. The pre-processing, backbone, and post-processing modules make up the GridDehazeNet. In comparison to derived inputs created by hand-selected pre-processing methods, the trainable pre-processing module can produce learned inputs with greater diversity and more useful properties. On a grid network, the backbone module implements a novel attention-based multi-scale estimation that can successfully address the bottleneck problem that the traditional multi-scale approach frequently runs into. The post-processing module aids in lowering the output artefacts. For both simulated and real-world photos, experimental results show that the GridDehazeNet performs better than state-of-the-art techniques, and this hazing solution does not rely on the atmospheric scattering model. To make sure that the ambient light can also be optimized throughout the structure, a U-net [32] is employed to estimate the homogeneous atmospheric light map. An end-to-end dehazing approach based on deep learning that can simultaneously optimize the transmission map, atmospheric light, and dehazed image is designed. To do this, the general optimization framework directly incorporates the atmospheric image degradation model. To effectively estimate the transmission map, a unique densely connected encoder-decoder structure with a multi-level pooling module and an original edge-preserving loss is proposed. The approach also incorporates a joint discriminator-based GAN framework to enhance the specifics and benefit from the structural mutual connection between the estimated transmission map and the dehazed image.

Estimating a medium transmission map for an input hazy image is essential for achieving haze removal. For medium transmission estimate, Cai et al., [11] presented DehazeNet, a trainable end-to-end system. A hazy image is input into DehazeNet, which then produces a medium transmission map that is utilised to recover a haze-free image using an atmospheric scattering model. DehazeNet uses a deep architecture built on Convolutional Neural Networks (CNN), whose layers are specifically created to represent the stated presumptions/priors in picture dehazing. For feature extraction, layers of Max out units are specifically chosen since they can produce practically all haze-relevant information. Also, they introduced the Bilateral Rectified Linear Unit (BReLU), a novel nonlinear activation function in DehazeNet that can enhance the quality of a recovered haze-free image. A powerful PFDN for image dehazing is provided by Dong et al., [33]. The key element PFDB is an FDU with a residual learning architecture. The FDU is made to completely study the useful characteristics for picture dehazing based on the physics model of the haze process. To make deep neural network training easier and more precise, the FDU makes use of the residual learning architecture. The PFDB is integrated as a complete backbone into an encoder-decoder network architecture for the purpose of image dehazing. They have investigated the impact of the planned PFDN on image dehazing. It performs better than the alternatives in both the quantitative and qualitative domains.

For the reduction of single picture haze, He et., [2] suggested a unique prior dark channel prior. The dark channel prior is based on data from haze-free outdoor photos. They discovered that some pixels, referred to as black pixels, frequently have extremely low intensity in at least one colour (RGB) channel in the majority of the local regions that do not cover the sky. In hazy photos, the air light is mostly responsible for the intensity of these dark pixels in that channel. Consequently, an exact evaluation of the haze transmission can be immediately provided by these dark pixels. A high-quality haze-free image can be recovered, and a useful depth map can be created, by combining a haze imaging model and a soft matting

interpolation technique. Their method is physically sound and capable of handling far-off objects in photos with significant haze. They don't rely on surface shading or a lot of variation in transmission. There are little halo artefacts in the outcome. In order to remove haze from a single input image, Kaiming et al., [34] suggested a straightforward yet effective image prior—dark channel prior. The previous dark channel contains statistics of clear outdoor photos. It is founded on an important finding: the majority of local patches in outdoor, haze-free photos contain some pixels with very low intensity in at least one-color channel. The haze thickness can directly be estimated and recover a high-quality haze-free image by using this prior in conjunction with the haze imaging model. Results on several blurry photos show how effective the proposed prior is. Furthermore, haze removal might also result in the creation of a superior depth map.

Zhu et al., [35] proposed a novel color attenuation prior for single image dehazing. This robust yet straightforward prior allows us to construct a linear model for the scene depth of the blurry image. The linear model's parameters are learned under supervision, and this process creates the useful link between the foggy image and the depth map that corresponds to it. With the recovered depth information, a single fuzzy image may be deleted without difficulty. They proposed a novel color attenuation prior for single image dehazing. This robust yet straightforward prior allows us to construct a linear model for the scene depth of the blurry image. The linear model's parameters are learned under supervision, and this process creates the useful link between the foggy image and the depth map that corresponds to it. With the recovered depth information, a single blurry picture may be simply eliminated. A brand-new multi-scale picture dehazing network was put forth by Deng et al., [36]. Instead of directly calculating the transmission map and atmospheric light intensity, it adaptively integrates the local location data with global atmospheric brightness, guided by the acquired haze-aware maps for each channel. Extensive testing on both synthetic and real-world blurry photos shows the effectiveness of the technology. Along with removing homogeneous haze from photos, this technique is also excellent at removing dense non-homogeneous haze.

Dai et al. [37] addressed the problem of hybridizing convolution and attention from two fundamental features of machine learning, generalization and model capacity. Their research shows that attention layers have a greater model capacity and can profit from bigger datasets because to their high prior of inductive bias, although convolutional layers often have better generalization and converging speeds. Combining the attention and convolutional layers can improve capacity and generalization, but doing so successfully to obtain better trade-offs between accuracy and efficiency is a significant challenge. They looked at two important conclusions: First, they discovered that straightforward relative attention can successfully combine layers of popular depth-wise convolution with layers of attention. Secondly, they discovered that properly stacking the layers of attention and convolution may be all that is needed to increase capacity and generalization. Based on these realizations, they presented the CoAtNet network architecture, which is simple yet effective and combines the benefits of Transformers and ConvNets.

Qin et al., [38] proposes and illustrates the use of the Fusion Attention Network for single-image dehazing. The FFA-Net structure is much better than earlier state-of-the-art approaches, despite its simplicity. Because of its great advantage in the restoration of picture detail and color integrity, it is envisaged that the network would tackle other low-level vision tasks including de raining, super-resolution, and denoising. FFA and other efficient modules in

FFA-Net play a vital part in the image restoration methods. Kulkarni et al., [39] proposed a piece of work that presents a novel spatially attentive offset extraction based deformable multi-head attention-based solution for aerial picture dehazing. In order to recreate fine level texture in the restored image, deformable multi-head attention is introduced here. In order to concentrate on pertinent contextual information, the also included a spatially attentive offset extractor in the deformable convolution. In order to efficiently convey edge features from shallow levels to deeper layers of the network, edge boosting skip connections are also implemented. comprehensive ablation investigation and comprehensive experimentation on synthetic and real-world data show that this method outperforms the existing studies on aerial image dehazing. An end-to-end multi-stage dehazing algorithm is created for the single image dehazing problem by Zhao et al., [40]. The algorithm has two separate components for extracting characteristics. By stacking pixel and channel attention techniques, texture-level information is extracted for shallow features. High-level features are captured using the suggested method's multi-head self-attention (MHSA) technique. By mining the dependencies of a wide variety of abstract data, MHSA enhances dehazing performance. Convolutions and cascaded attention techniques enhance the supremacy of the transformer architecture's feature extraction capabilities. The context information is equalized during the decoding stage using multilayer perceptron (MLP). In addition, a contrastive loss function is developed that includes numerous negative samples as well as correction terms. When dealing with various dehaze concentrations, the correction term is formed based on the difference between the exact and blurry images, which can improve the training impact. This loss function's training outcome helps the model approximate crisp images while avoiding blurry ones.

Song et al., [27] suggested DehazeFormer, a set of enhancements that includes a modified normalization layer, an activation function, and a spatial information aggregation strategy. In order to assess the method's capacity to eliminate highly non-homogeneous haze, they also collected a sizable realistic remote sensing dehazing dataset. For single image dehazing, Parihar et al., [28] presented the Densely Connected Convolutional Transformer (DCCT). DCCT is a productive architecture that combines the local dependencies and the multi-head Performer. They suggested a learnable connection layer that is used to fuse features at various levels throughout the entire design in order to prevent information loss between features at various levels. DCCT training using a joint loss that takes into use supervised metric learning is directed, which enables us to take into account both positive and negative characteristics for a multi-image perceptual loss. Through ablation trials, it confirms the design decisions and the suggested DCCT's efficacy. They determine that the suggested DCCT is highly competitive with the state of the art by comparison with the representative approaches.

3. Proposed Methodology for Image Dehazing using Alternate Pooling Fused Transformer Network with Attention

Based on DehazeFormer's Network [27], the proposed work makes a number of modifications to achieve our goal of developing an enhanced dehazing model for images using efficient proposed transformer block with our recently proposed Dual weighted deep channel attention [46]. The transformer block fetch the input from the alternate pooling module and that feature map is further fine-tuned by the locally grouped self-attention module instead of traditional self-attention in Vision Transformer. The output from the both the transformer is further

concatenated and send into once again next level of transformer block operation with pooling swapped module. The following Figure 1 shows the intended proposed work's architecture.

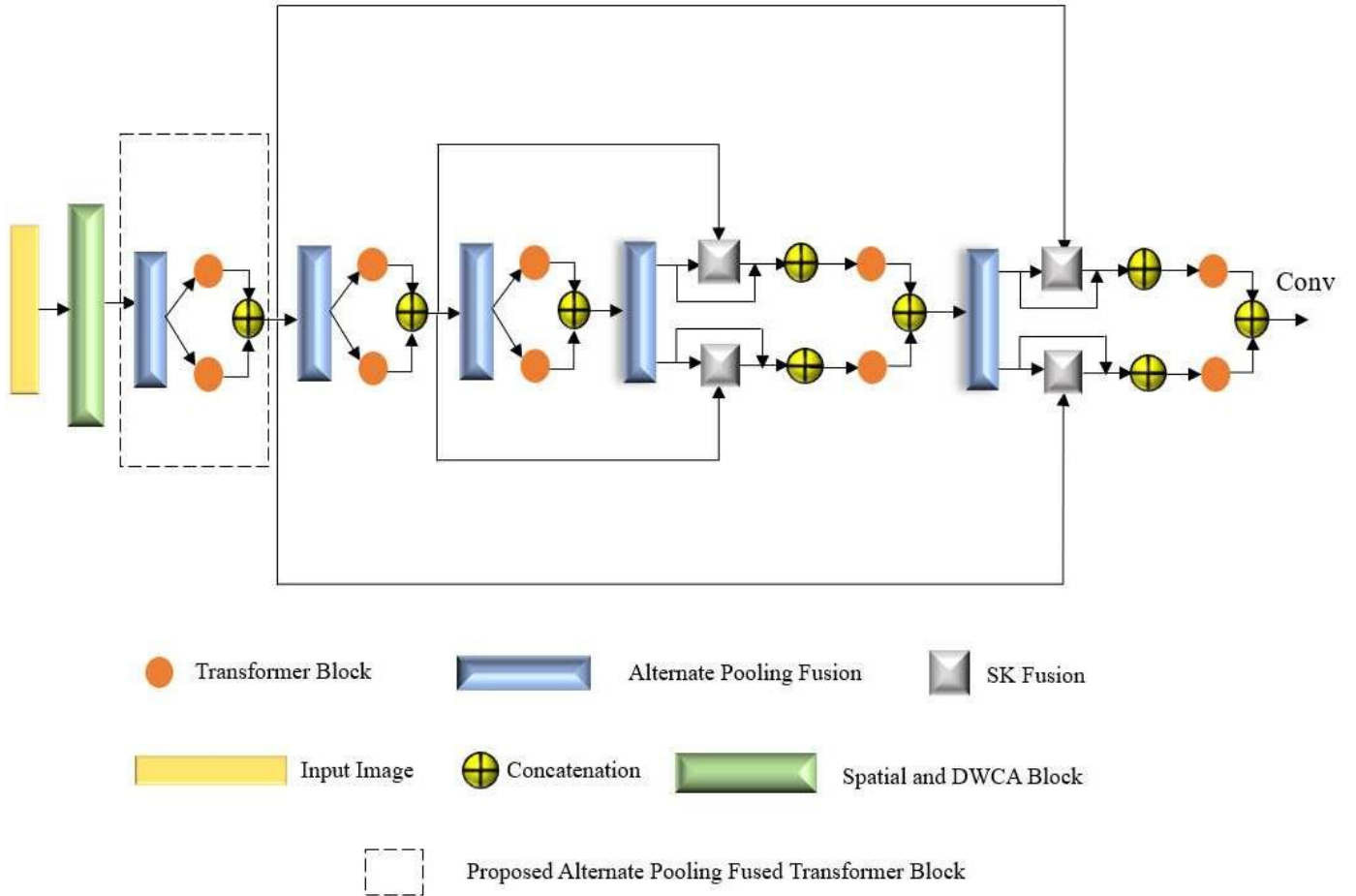


Fig 1.Architecture of the proposed Alternate Pooling Fused Transformer Block

3.1 Spatial and DWCA Block

3.1.1 Spatial Attention Model

Many research papers employ the attention model [41-48] to enforce the attention for a certain feature in various applications. Figure 2 displays a spatial attention model based on pixels. This makes it easier to extract the most important information from a spatial coordinate. The input patch, u , is subjected to average pooling, that will draw attention to the characteristics that are contributing most to the applications. The convolution layer applies this pooled output. By multiplying the sigmoid function on this input patch, u , the spatial attention model output, u' , is created.

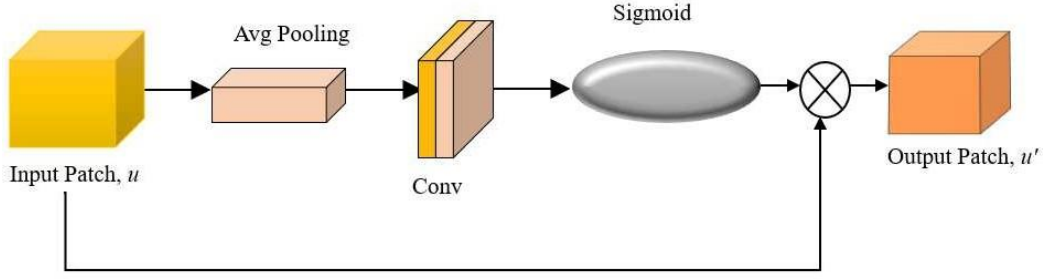


Fig 2. Spatial Attention Model

3.1.2 Double Weighted Channel Attention (DWCA)

Figure 3 shows the architecture of Dual Weighted Deep Channel Attention (DWCA). The channels' interactions provide a positive trait that might be regarded as unique. Methods to "channel attention" [49, 50] aid in highlighting these characteristics. First, channel-specific spatial information is used as the channel descriptor via global average pooling [51].

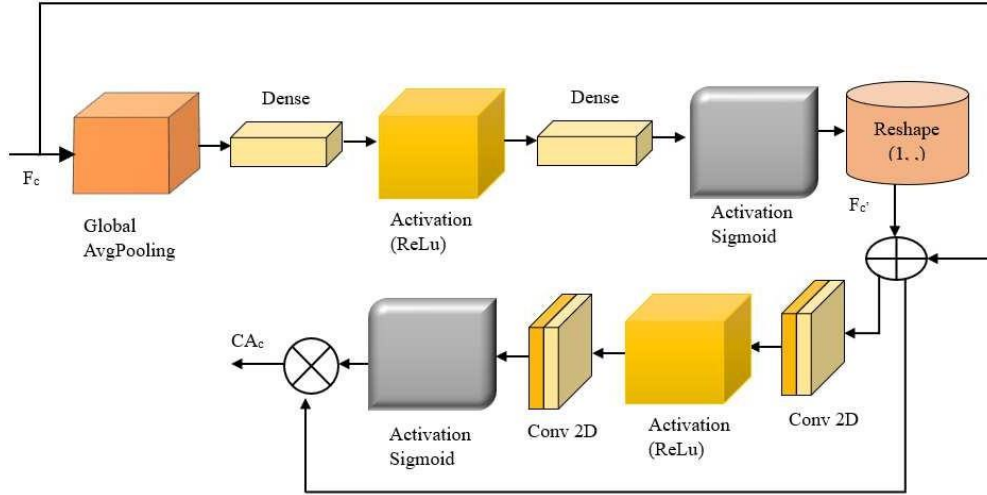


Fig 3. Double Weighted Deep Channel Attention (DWCA)

This study introduces a novel channel attention technique called DWCA (Figure 3), in which the links between the channels are discovered through in-depth learning. Let D_c be the input patch, which is fed into the global average pooling layer to enhance the patch's ability to represent data. The output of this layer is provided to the dense layer, which learns all of the information from the previous layer, and to the activation layer, which identifies the fired neurons. It is then delivered into the dense layer and a subsequent activation layer. This layer's output is modified and multiplied by D_c and the resulting D_c^* output is sent into the convolution layer before being supplied into the activation layer (Relu). The output of DWCA as EA_c is obtained by feeding this through the convolution layer and activation layer (sigmoid), which are then multiplied by the intermediate output D_c^* .

$$K_c = Gl_p(D_c) = \frac{1}{H_i \times W_i} \sum_{i=1}^{H_i} \sum_{j=1}^{W_i} Y_c(i, j) \quad \dots (I)$$

Gl_p provides the global pooling function, while Y_c at position (i, j) , which represents the value of the c -th channel, provides $Y_c(i, j)$. The feature map's original shape, $C \times H_i \times W_i$, is modified to $C \times 1 \times 1$. For the purpose of determining the weights of the various channels, features are passed through two dense layers. ReLu and sigmoid activation processes follow this.

$$D'_c = \delta (Den (\sigma (Den(K_c)))) \quad \dots(2)$$

Where the sigmoid function σ and the ReLu function δ are present. After that, multiply the inputs D_c & D'_c by element-wise multiply the input D_c and D'_c .

$$D_c^* = D'_c \otimes D_c \quad \dots(3)$$

The two convolution layers are sent with D_c^* along with ReLu and sigmoid activation functions as in equation (6)

$$D_c^{**} = \delta (Conv (\sigma (Conv(D_c^*)))) \quad \dots(4)$$

Then, D'_c is multiplied element-wise with D_c^{**} to get the outcome of DWCA, EA_c .

$$EA_c = D'_c \otimes D_c^{**} \quad \dots(5)$$

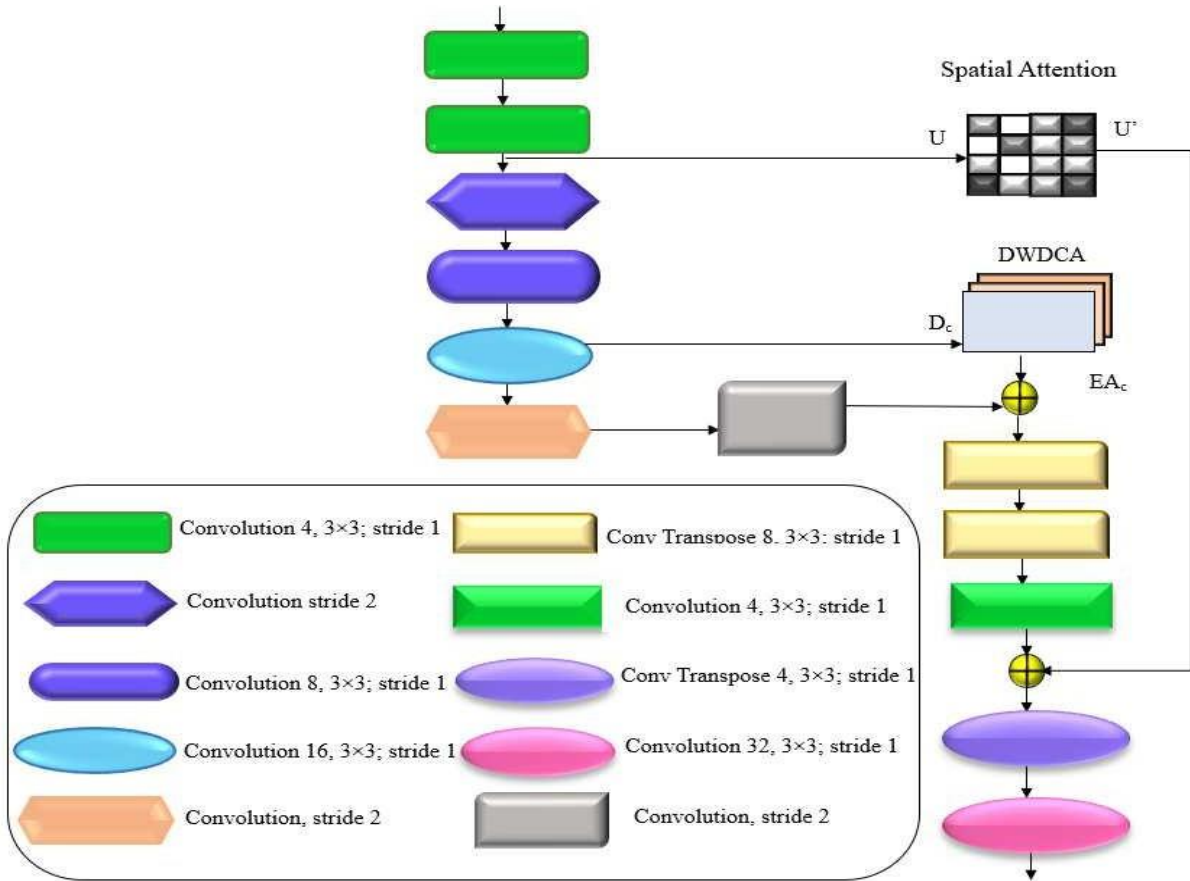


Fig 4. Architecture of the Spatial and DWCA block

The architectural representation of Spatial and DWCA block is depicted in figure 4. In this block, an input patch is fed into two convolution layers with a filter size 3×3 , stride 1 with a number of filters 4. The output of this convolution is input into two more convolutions, which have 8 filters per layer with stride 2 and stride 1 respectively. The output is then applied to two other convolution layers with strides 1 and 2 and with 16 filters per layer, respectively. The output of fifth convolution layer with 16 filter is applied with DWCA whose output is concatenated along with the output of convolution transpose layer of 8 filter with stride 2 which used the output of sixth convolution layer of 16 filters with stride 2 as input. Two convolution transpose layers, each with 8 filters, receive the output of concatenation for processing. The output of spatial attention is applied to the output of the second convolution layer with stride 1, which has 4 filters, and the output after passing through a convolution transpose layer with stride 1 and having 4 filters is concatenated. After concatenation, the concatenated output is sent into two other convolution transpose layers of 4 filters and another convolution transpose layer with 1 filter respectively.

3.2 Proposed Alternate Pooling Fused Transformer Block

3.2.1 Alternate Pooling Fusion

The architectural representation of alternate pooling fusion is depicted in figure 5. In the proposed block, the input is fed into two convolution layers. After the convolution layer, first convolution is sent into two separated pooling layers such as horizontal pooling and vertical pooling. On the other hand, another convolution is sent into two other separated pooling layers such as horizontal pooling and vertical pooling. Then, the result of first horizontal pooling and second vertical pooling is concatenated and yielded result. Similarly, the outcome of first vertical pooling and second horizontal pooling is concatenated and yielded result.

With a fixed filter size, the alternate pooling fusion will be repeated five times, with a different number of filters applied each time. A 3×3 kernel size, padding 1, and 24 filter sizes for the left convolution and a 7×7 kernel size, padding 3, and 24 filter size for the right convolution will be sent for the first alternate pooling fusion process. For the second alternate pooling fusion process, 3×3 kernel size, stride 2, and 48 filter sizes for the left convolution and a 5×5 kernel size, stride 2, pooling 2 and 48 filter size for the right convolution will be sent. A 3×3 kernel size, padding 1, stride 2 and 96 filter size for the left convolution and a 4×4 kernel size, padding 3, stride 2 and 96 filter size for the right convolution will be sent for the third alternate pooling fusion process. For the fourth alternate pooling fusion process, 2×2 kernel size, stride 2, and 48 filter size for the left convolution and the same 2×2 kernel size, stride 2, 48 filter size for the right convolution will be sent. A 2×2 kernel size, stride 2 and 24 filter size for the left convolution and a 2×2 kernel size, stride 2 and 24 filter size for the right convolution will be sent for the fifth alternate pooling fusion process.

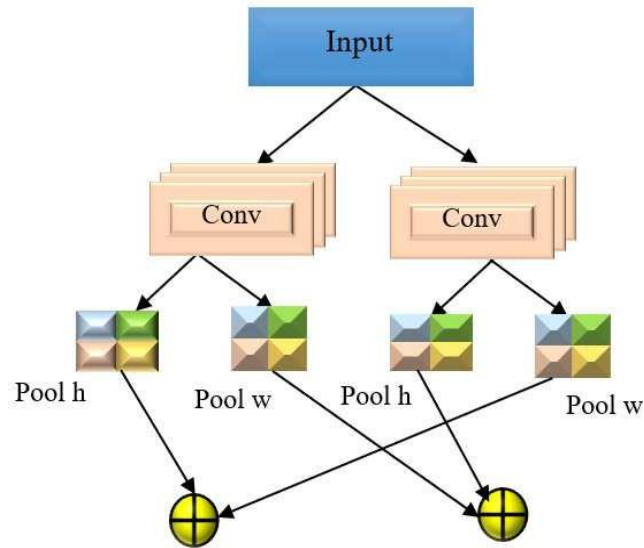


Fig 5. Architecture of Alternate Pooling Fusion

3.3 Transformer Block with Locally Grouped Self Attention

Since they may be extensively parallelized, transformers [38] were initially developed in the field of Natural Language Processing (NLP).

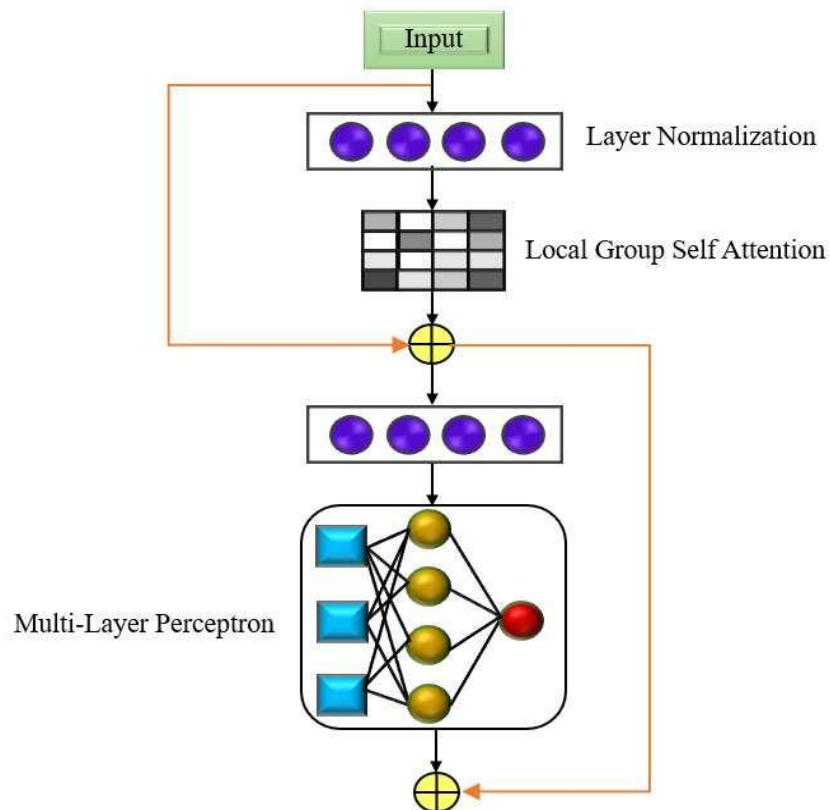


Fig 6. Architecture of Transformer Block with Locally Grouped Self Attention

Images must first be separated into patches before each patch is subjected to the calculation in order to use Transformers and other approaches successfully. At the moment, a lot of research is being done on patch communication. Before introducing Transformer to the classification problem, Vision Transformer (ViT) first uses the Transformer Encoder to extract features [39]. Transformers will soon be used for several computer vision tasks, including detection and low-level vision. It was recommended to embed Transformers within Transformers to model the internal data of patches. Small and lightweight transformers are also becoming more and more common, such as the Lite-Transformer [40]. Recent research, including CvT [41] and CMT [42], focuses on fusing CNN and Transformer to take advantage of both architectures' benefits. Figure 6 shows the architecture of transformer.

3.3.1 Locally Grouped Self Attention (LSA)

LSA and globally sub-sampled attention (GSA) are ideas of spatially separable self-attention (SSSA), originally put forward by Chu et al. [44]. It was influenced by the often-used separable depth-wise convolutions. While LSA manages the absolutely fine and close-by data, GSA handles the far-reaching and global information. With just a few lines of code and some matrix multiplications, we were able to apply LSA attention in our innovative method. Therefore, all of our architectures are easy to implement and have a wide range of applications.

Figure 7 shows a schematic representation of locally group self-attention. They equally divide the 2D feature maps into sub-windows inspired by the subject design in depth wise convolutions for effective inference in order to restrict self-attention communications to within each sub-window. A multi-head design resonates similarly to this in self-attention, where communication only occurs inside the channels of the same head.

The feature maps are arranged into sub-windows with precise sizes of $m \times n$. Presumably, $H \% m$ and $W \% n$ were set to be 0. Here, $O(\frac{H^2 W^2}{m^2 n^2} d)$ is the estimation load of the self-attention, and $O(\frac{H^2 W^2}{m^2 n^2} d)$ is the total cost, with elements for each group of $\frac{HW}{mn}$. When $K_1 = \frac{H}{m}$ and $K_2 = \frac{W}{n}$ are fixed, the expense can be estimated using the very efficient formula $O(k_1 k_2 H W d)$, where $k_2 \ll W$, $k_1 \ll H$, respectively. In the absence of HW , it climbs.

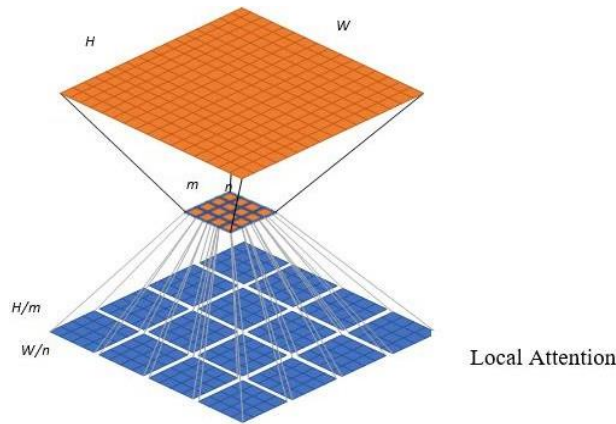


Fig 7.Diagram of locally group self-attention (LSA)

Despite the fact that the LSA technique is operationally effective, the image is divided into non-intersecting secondary windows. A system is therefore required to make it easier for different sub-windows to communicate with one another. If not, the data can only be used for domestic consumption, which has a very small receptive field and much worse performance. Similar to how depth-wise convolutions can't totally replace all of the ordinary convolutions in CNNs.

3.4 SK Fusion with Multi Level APF_Transformer

The SK fusion layer, which draws inspiration from SKNet [61], uses channel attention to merge several branches. Two independent transformer blocks receive the output of the proposed APF block and the outcome of those two transformer blocks are further concatenated and then given as input for the next level of proposed transformer block process. This process is repeated for further subsequent five levels. The process will be repeated five times with a set filter size, but each time a new number of filters will be used. But in the case of fourth level operation the output from the second transformer block is fed into SK fusion layer which is further concatenated with the output from the APF module of the fourth block and that output only given to transformer block. This same concept is carried out in the fifth level where as the output from the first level transformer block is fed into SK fusion layer and concatenated with the actual outcome from the proposed APF module of the fifth level of processing. Finally, the output is concatenated and fed into the convolution layer, which produces a four filtered feature maps. This is further processed by the Soft Reconstruction layer as described in [27] to produce the haze free image.

4. Result and Discussion

4.1 Metrics

4.1.1 SSIM

A perceptual metric called the Structural Similarity Index (SSIM) measures the loss of image quality brought on by data transmission losses or other processing steps like data compression. It is a complete reference metric that needs two images such as reference image and a processed image from the same image capture. Usually, the processed image is compressed. There is a lot of information detailing the theory of SSIM because it has been there since 2004, but very few resources go into great detail, especially for a gradient-based implementation because SSIM is frequently employed as a loss function.

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (6)$$

4.1.2 PSNR

The most popular method for estimating quality loss caused by various codecs and image compression is known as peak signal-to-noise ratio (PSNR). In this method, the original image serves as the signal, and the noise represents compression errors. When analyzing image improvement techniques like Super resolution, PSNR is frequently used. In this scenario, the signal is the original/ground truth image, and the noise is the error that the model was unable to recover. Despite being a logarithm-based metric, PSNR is MSE-based.

$$PSNR=10\log_{10}(\frac{R^2}{MSE}) \quad (7)$$

4.1.3 CIEDE2000

The link between the color difference value ΔE and the differences in brightness $\Delta L'$, hue $\Delta H'$, and chroma $\Delta C'$ is shown by the CIEDE2000 color difference formula. It's described as,

$$\Delta E^*_{ab} = [(\Delta L^*)^2 + (\Delta a^*)^2 + (\Delta b^*)^2]^{1/2} \quad (8)$$

4.2. i-Haze dataset

35 image pairs make up the dataset I-Haze (DB1) [48]. They were collected indoors in pairs of haze-free and hazy images. This dataset's hazy images were produced using a professional haze machine that actually creates haze. The MacBethcolor checker was utilized in each scene to enhance evaluations of performance and verify color alignment. Both of haze-free and hazy images are captured in a controlled environment with the same lighting conditions. The collection includes indoor scenes with various color schemes and specs that measure 54563632 and have a 24-bit depth.

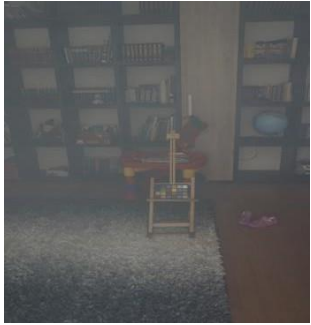

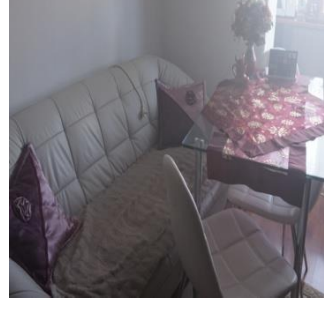
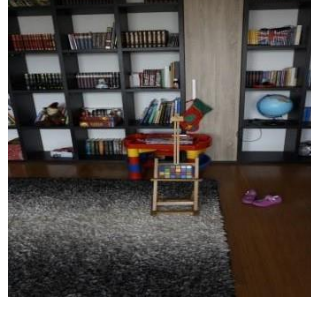

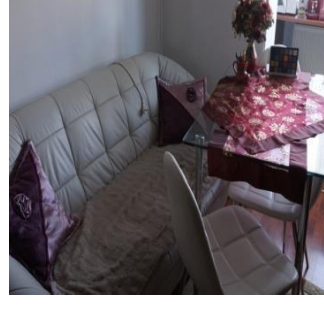
<p>Hazy images</p> <p>(a)</p>			
<p>Ground Truth images</p> <p>(b)</p>			

Fig8. Example images of DB1 dataset (a) hazy and (b) matching ground truth images

The sample images from the DB1 dataset are shown in Figure 8. Figure 8(a) displays the hazy images, while figure 8(b) displays the equivalent ground truth images.

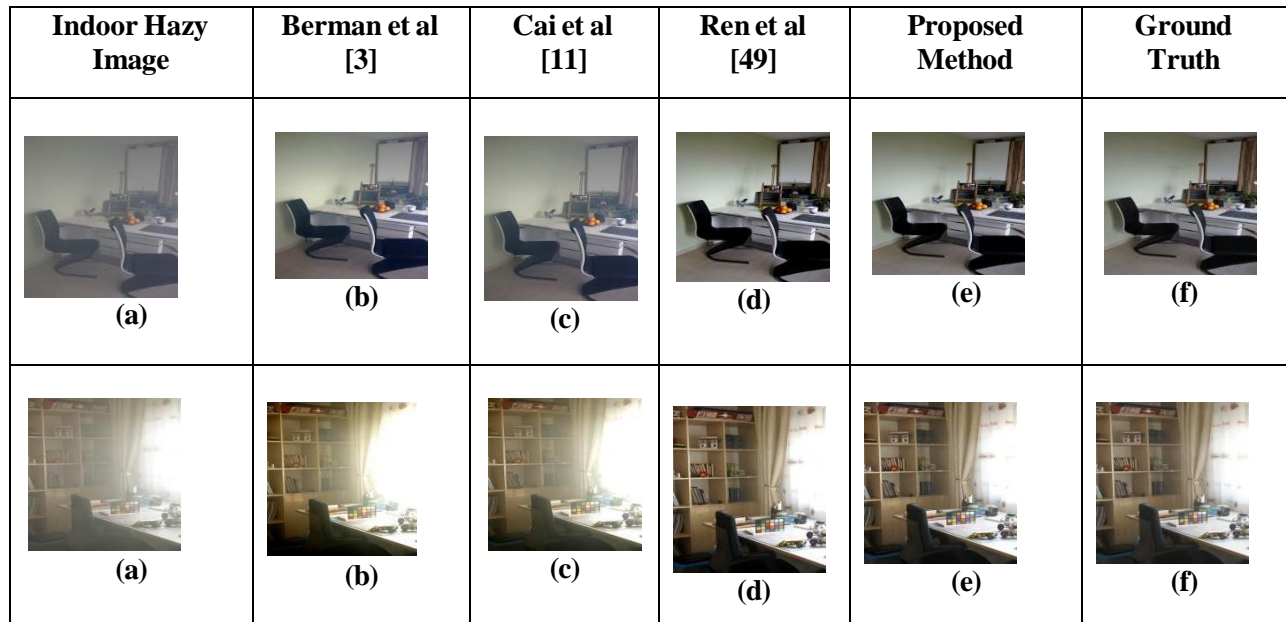


Fig. 9(a - f): Input-output pairs and corresponding ground truth for two example images of DB1. (a) Hazy input image (b) output of Berman et al [3] (c) output of Cai et al [11] (d) output of Ren et al [49] (e) output of proposed method and (f) ground truth image

The results of different methods on the i-haze dataset are displayed in Fig. 9. In particular than Ren et al. and Cai et al., Berman et al. clears the haze. Comparing the two results, Cai et al.'s appears to be superior. By retaining brightness and reducing haze, the proposed method produces an output that is almost identical to the actual situation. The performance comparison of this work with the prior efforts on the DB1 dataset is shown in Table 1. Figures 10, 11, and 12 show the analysis of the DB1 dataset using PSNR and SSIM CIEDE2000.

Table 1: Performance analysis of proposed work with existing works in DB1 dataset

	SSIM	PSNR	CIEDE2000
He et al. [2]	0.71	15.29	17.17
Meng et al. [51]	0.75	14.57	16.83
Fattal[52]	0.57	12.42	21.39
Cai et al. [11]	0.77	16.98	12.99
Ancuti et al. [53]	0.77	16.63	14.43
Berman et al. [3]	0.77	15.94	14.63
Ren et al. [45]	0.79	17.28	12.74
Suganthi et al.[46]	0.83	20.54	10.26
APF_TRANS_NET	0.84	22.54	9.57

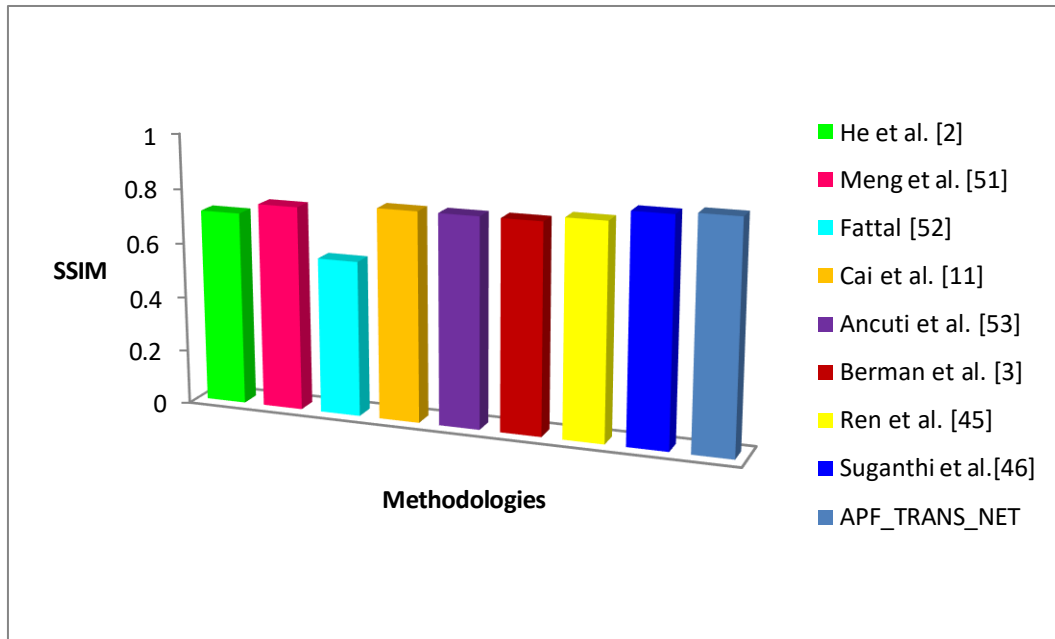


Fig10. Performance analysis in terms of SSIM in DB1 dataset

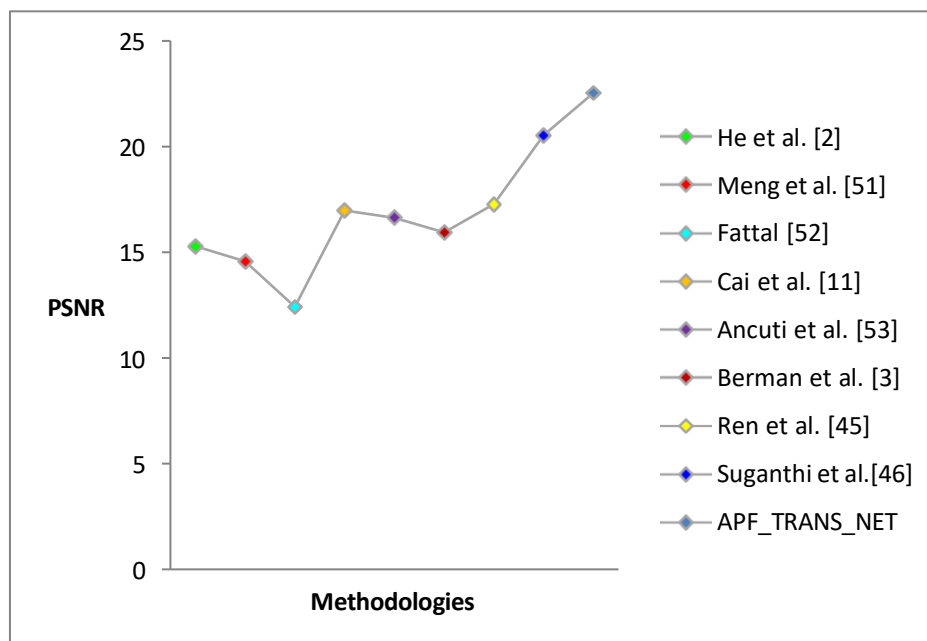


Fig 11. Performance analysis in terms of PSNR in DB1 dataset

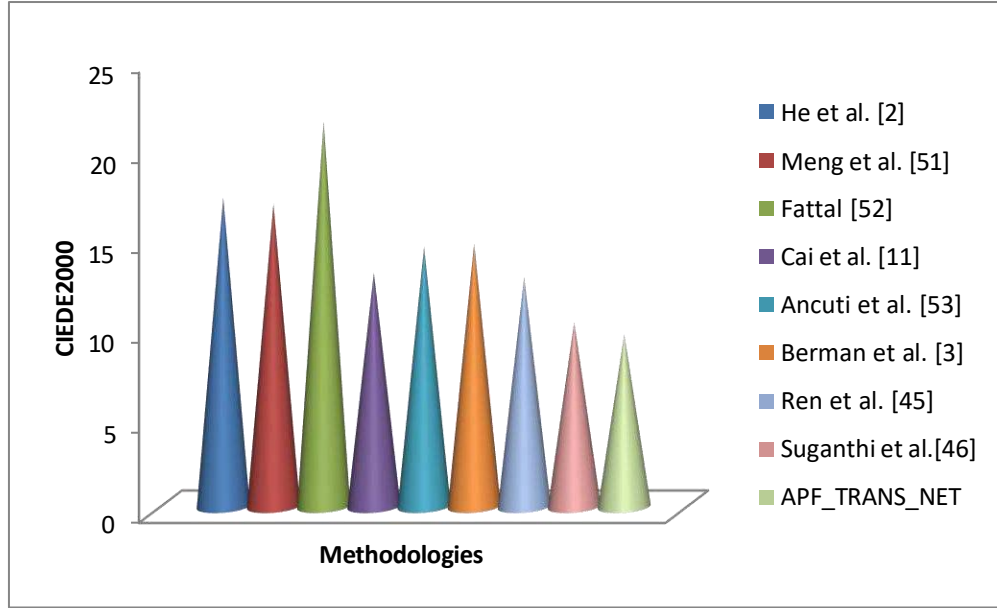


Fig 12. Performance analysis in terms of CIEDE2000 in DB1 dataset

The proposed work outperforms other existing works, according to the study based on Table 1, Figure 10, 11, and 12. This demonstrates the performance improvement of +0.73 for SSIM, +15.58 for PSNR, and -15.74 for CIEDE2000.

4.3 O-Haze dataset

The dataset O-Haze (DB2) [34] contains 45 image pairs. They were captured outside in both clear and cloudy pairs. The hazy images in this collection were created using a professional haze machine in real-world haze. The same illumination conditions are used to capture photos with and without haze.





Fig 13. Example images of DB2 dataset (a) hazy image (b) corresponding ground truth images















Outdoor Hazy Image	Berman et al [3]	Cai et al [11]	Ren et al [45]	Mondal et al [47]	Proposed Method	Ground Truth
 (a)	 (b)	 (c)	 (d)	 (e)	 (f)	 (g)
 (a)	 (b)	 (c)	 (d)	 (e)	 (f)	 (g)

Fig 14 (a - g): Input-output pairs and corresponding ground truth for two example images. (a) Hazy input image (b) output of Berman et al [3] (c) output of Cai et al [11] (d) output of Ren et al [45] (e) output of Mondal et al [47] (f) output of proposed method and (g) ground truth image

The results of various methods on the o-haze dataset are displayed in Figure 14. In comparison to Berman et al [3], Cai et al [11], and Ren et al [45], Mondal et al [47] produce better results. More haze is removed by Berman et al. [3] than by Ren et al. [45] and Cai et al., [11] although the brightness is different from the actual image. By retaining brightness and reducing haze, the suggested method produces an output that is almost identical to the actual situation. The performance comparison of this study with earlier efforts on the DB2 dataset is shown in Table 2. The examination of the SSIM, PSNR, and CIEDE2000 in the DB2 dataset is shown in Figures 15, 16, and 17.

Table 2: Performance analysis of proposed work with existing works in DB2 dataset

	He et al. [2]	Meng et al. [47]	Fattal [48]	Cai et al. [11]	Ancuti et al. [53]	Berman et al. [3]	Ren et al. [45]	Suganthi [46]	APF_TRANS_NET
SSIM	0.74	0.75	0.71	0.67	0.75	0.75	0.77	0.80	0.82
PSNR	16.59	17.44	15.64	16.21	16.86	16.61	19.07	22.26	23.57

CIEDE 2000	20.75	16.97	19.85	17.35	16.43	17.09	14.67	11.56	10.29
-----------------------	-------	-------	-------	-------	-------	-------	-------	-------	-------

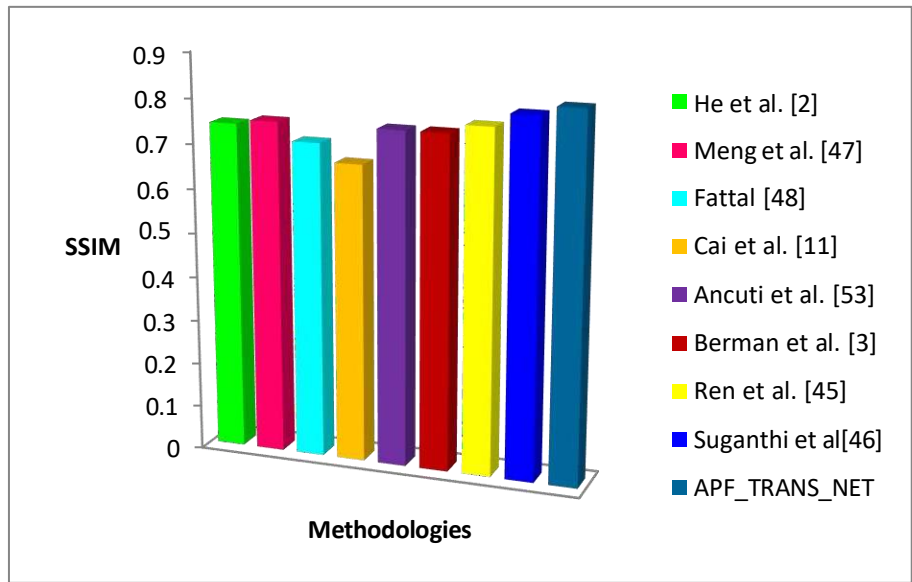


Fig 15. Performance analysis in terms of SSIM in DB2 dataset

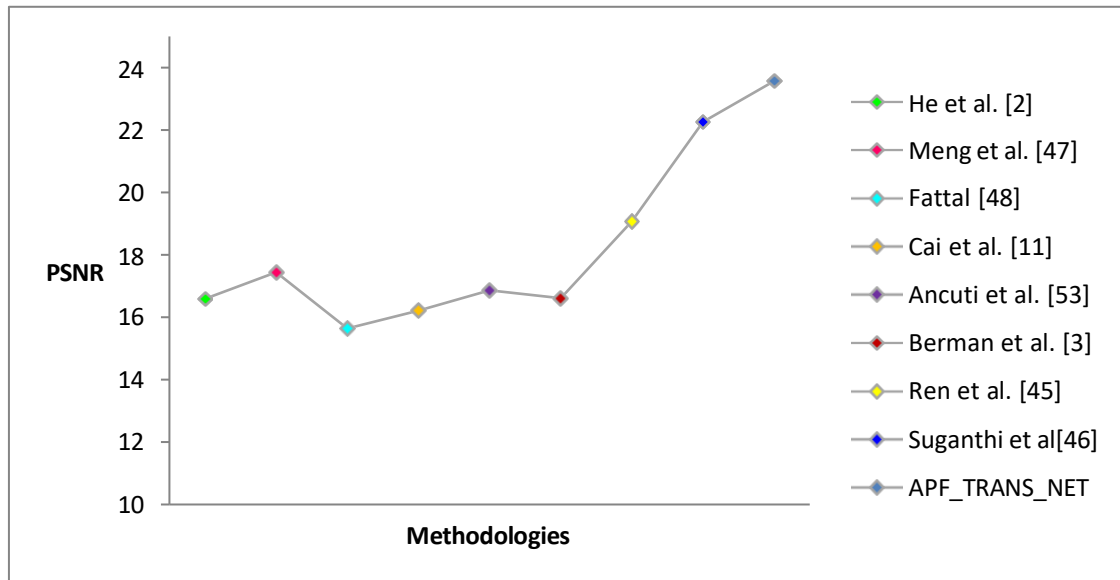


Fig 16. Performance analysis in terms of PSNR in DB2 dataset

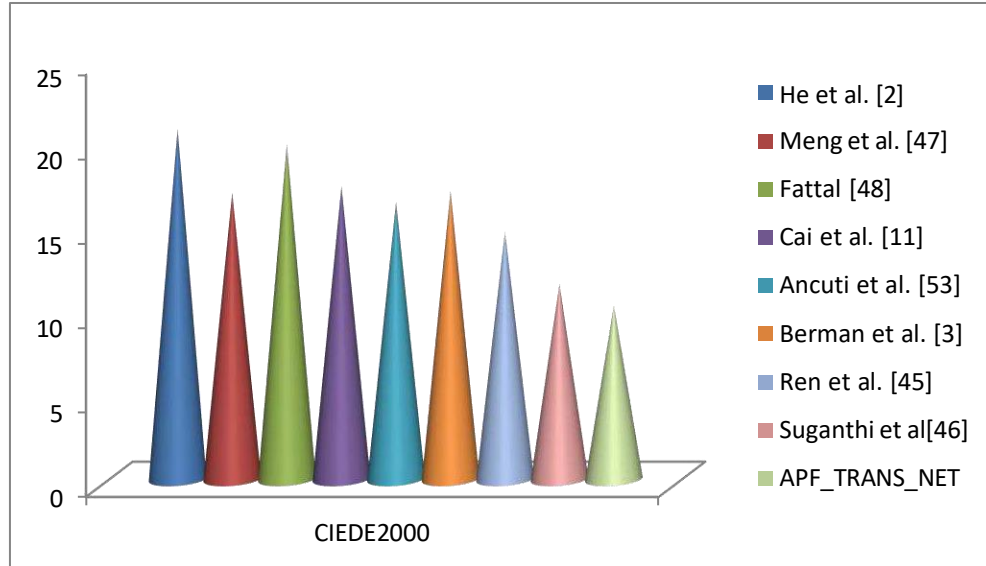


Fig 17. Performance analysis in terms of CIEDE2000 in DB2 dataset

The proposed work outperforms other existing efforts, according to the study based on Table 2, Figure 15, 16, and 17. It displays an improvement in performance of +0.73 for SSIM, +16.91 for PSNR, and -17.58 for CIEDE2000.

4.4 SOTS dataset

A dataset called Synthetic Objective Testing Set (SOTS) (DB3) [58] includes 500 images of the outside and 500 images of the interior. They have a variety of contents and include both genuine and artificial images. Corresponding images are shown in the following figure 18 and 19.

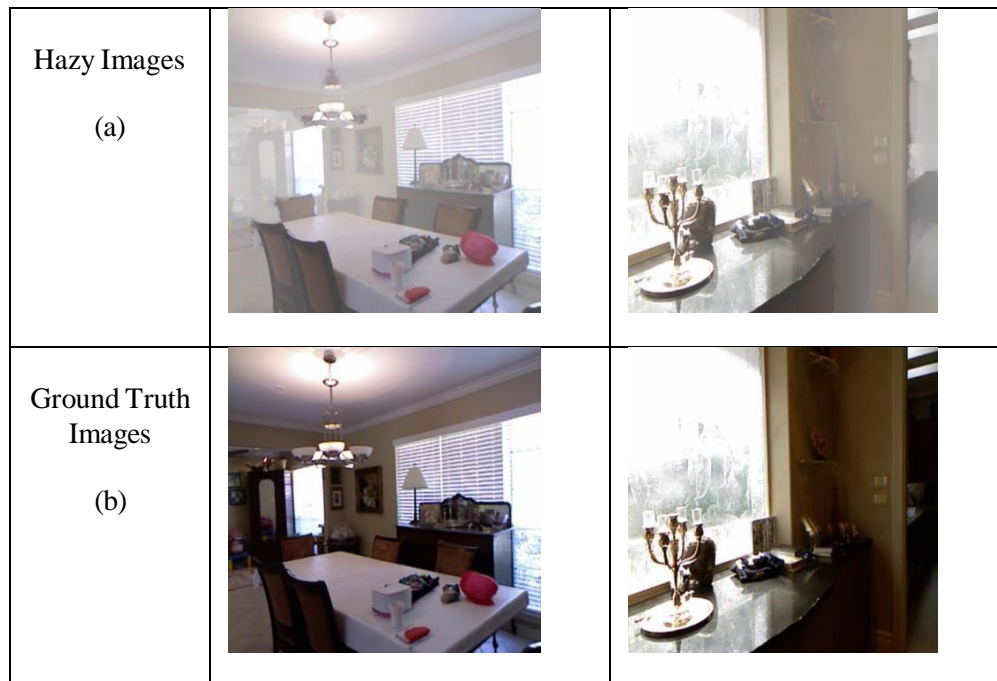


Fig18. Example images of DB2 dataset for indoor images (a) hazy image (b) corresponding ground truth images

















Hazy Images (a)		
Ground Truth Images (b)		

Fig 19. Example images of DB2 dataset for outdoor images (a) hazy image (b) corresponding ground truth images

	Indoor Images		Outdoor Images	
Input Image				
AOD-Net [29]				
GCANet [13]				

PFDN [59]				
FFA-Net [38]				
DehazeFormer-S [27]				
APF_TRANS_NET				
Ground Truth Image				

Fig 20.Input-output pairs and corresponding ground truth for two example images of DB3 for indoor and outdoor images.

The results of several approaches on the SOTS dataset are shown in the above Fig. 20.

Table 3: Performance analysis of proposed work with existing works in DB3 dataset

Methods	SOTS - Indoor		SOTS - Outdoor	
	PSNR	SSIM	PSNR	SSIM
DCP[2]	16.62	0.818	19.13	0.815
DehazeNet[11]	19.82	0.821	24.75	0.815
MSCNN[55]	19.84	0.833	22.06	0.908
AOD-Net[29]	20.51	0.816	24.14	0.92
GFN[12]	22.3	0.88	21.55	0.844
GCANet[13]	30.23	0.98	-	-
GridDehazeNet[31]	32.16	0.984	30.86	0.982

MSBDN[58]	33.67	0.985	33.48	0.982
PFDN[56]	32.68	0.976	-	-
FFA-Net[38]	36.39	0.989	33.57	0.984
AECR-Net[18]	37.17	0.99	-	-
DehazeFormer-T [27]	35.15	0.989	33.17	0.982
DehazeFormer-S [27]	36.82	0.992	34.36	0.983
DehazeFormer-B [27]	37.84	0.994	34.95	0.984
DehazeFormer-M [27]	38.46	0.994	34.29	0.983
DehazeFormer-L [27]	40.05	0.996	-	-
BCCR [54]	16.88	0.79	15.49	0.78
NLD [59]	17.29	0.75	18.07	0.8
CAP [35]	19.05	0.84	22.3	0.91
MSCNN [45]	17.11	0.81	19.56	0.86
Golts et al. [60]	19.25	0.83	24.08	0.93
Suganthi et al.[46]	39.56	0.99	36.25	0.99
APF_TRANS_NET	41.59	0.998	38.61	0.991

The performance comparison of the proposed work with the prior studies in the DB3 dataset is shown in Table 3. Figure 21 displays the performance analysis for indoor and outdoor images in DB3 in terms of SSIM. The PSNR performance study for indoor and outdoor images in DB3 is shown in Figure 22.

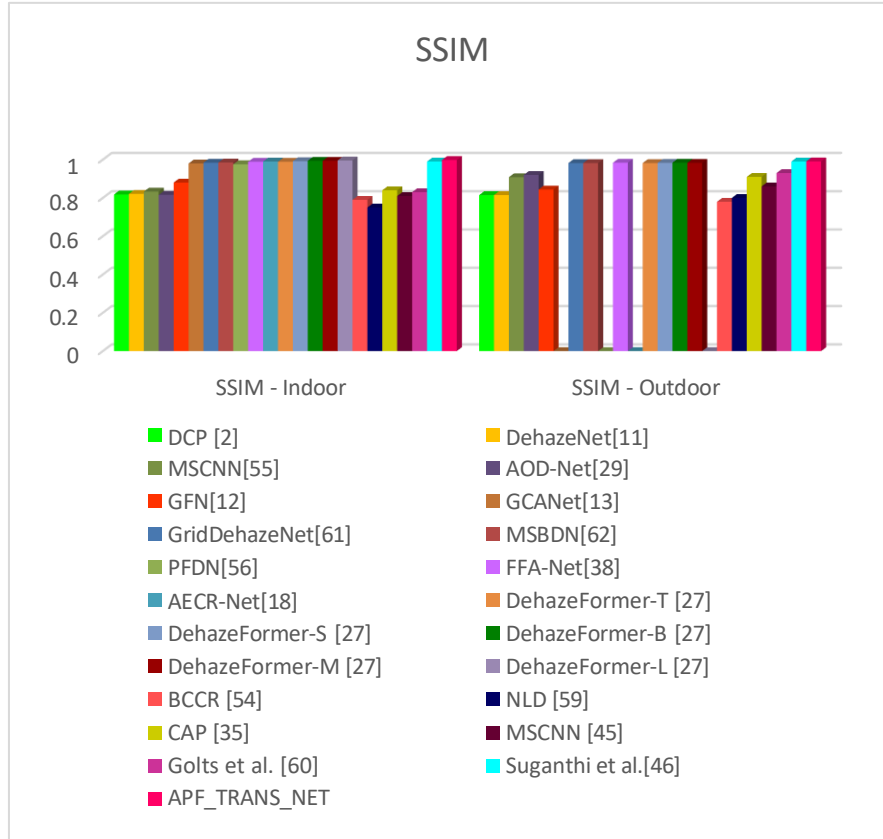


Fig 21.Performance analysis in terms of SSIM in DB3 dataset for indoor and outdoor images.

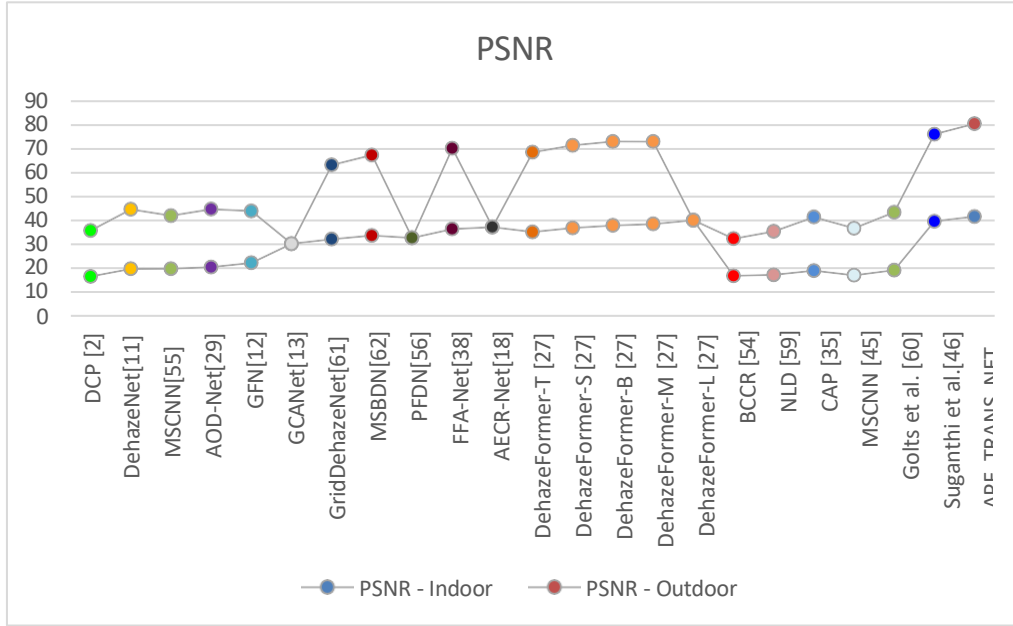


Fig 22.Performance analysis in terms of PSNR in DB3 dataset for indoor and outdoor images.

According to the analysis based on Table 3, Figures 21 and 22, the proposed work performs better than other studies that have already been done using the DB3 dataset. For indoor pictures, this demonstrates a minimum performance improvement of +27.58 in terms of PSNR and +0.87 in terms of SSIM. The performance gain for outdoor images is +26.22 in terms of PSNR and +0.90 in terms of SSIM. Our proposed technique significantly outperforms state-of-the-art works for the three datasets (i-Haze, O-Haze, and SOTS) in terms of PSNR, SSIM, and CIEDE2000.

4.5 RESIDE-6K

On the mixed dataset, models are tested and trained. We employ an experimental set-up that differs significantly from the RESIDE-Full from DA [38]. The images in its training set are all scaled to 400 400 and include 3,000 ITS image pairings and 3,000 OTS image pairs.

Table 4: Performance analysis of proposed work with existing works in DB4 dataset

Methods	RESIDE-6k	
	SOTS-mix	
	PSNR	SSIM
DCP[2]	17.88	0.816
DehazeNet[11]	21.02	0.87
MSCNN[55]	20.31	0.863
AOD-Net[29]	20.27	0.855
GFN[12]	23.52	0.905
GCANet[13]	25.09	0.923

GridDehazeNet[31]	25.86	0.944
MSBDN[58]	28.56	0.966
PFDN[59]	28.15	0.962
FFA-Net[38]	29.96	0.973
AECR-Net[18]	28.52	0.964
DehazeFormer-T[27]	30.36	0.973
DehazeFormer-S[27]	30.62	0.976
DehazeFormer-B [27]	31.45	0.98
DehazeFormer-M [27]	30.89	0.977
PATE[46]	31.34	0.97
APF_TRANS_NET	33.45	0.985

The performance comparison of the proposed work with the prior studies in the DB4 dataset is shown in Table 4. Figure 23 displays the performance analysis for images in DB4 in terms of SSIM. The PSNR performance study for images in DB4 is shown in Figure 24.

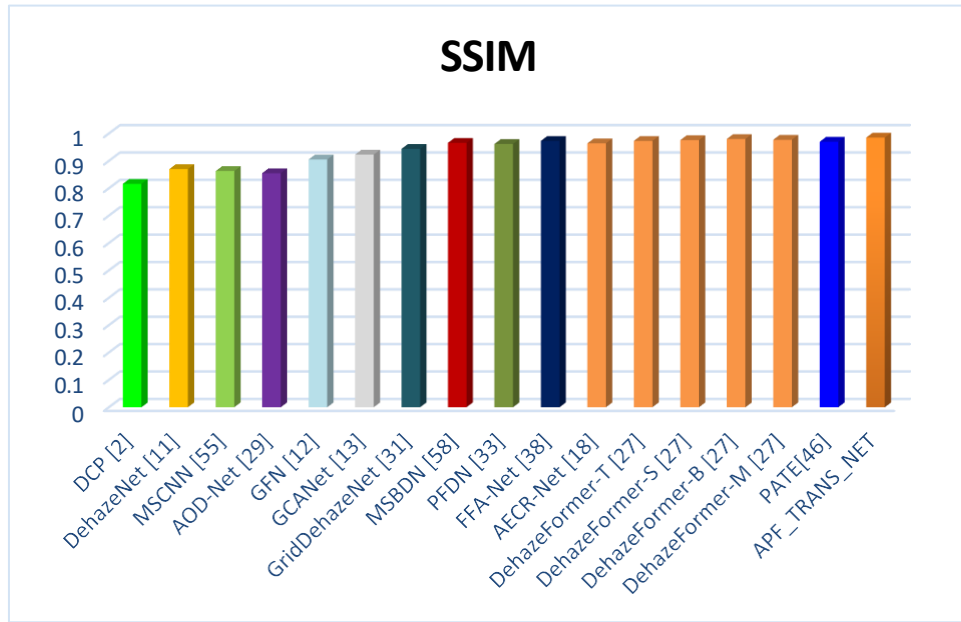


Fig 23.Performance analysis in terms of SSIM in DB4 dataset.

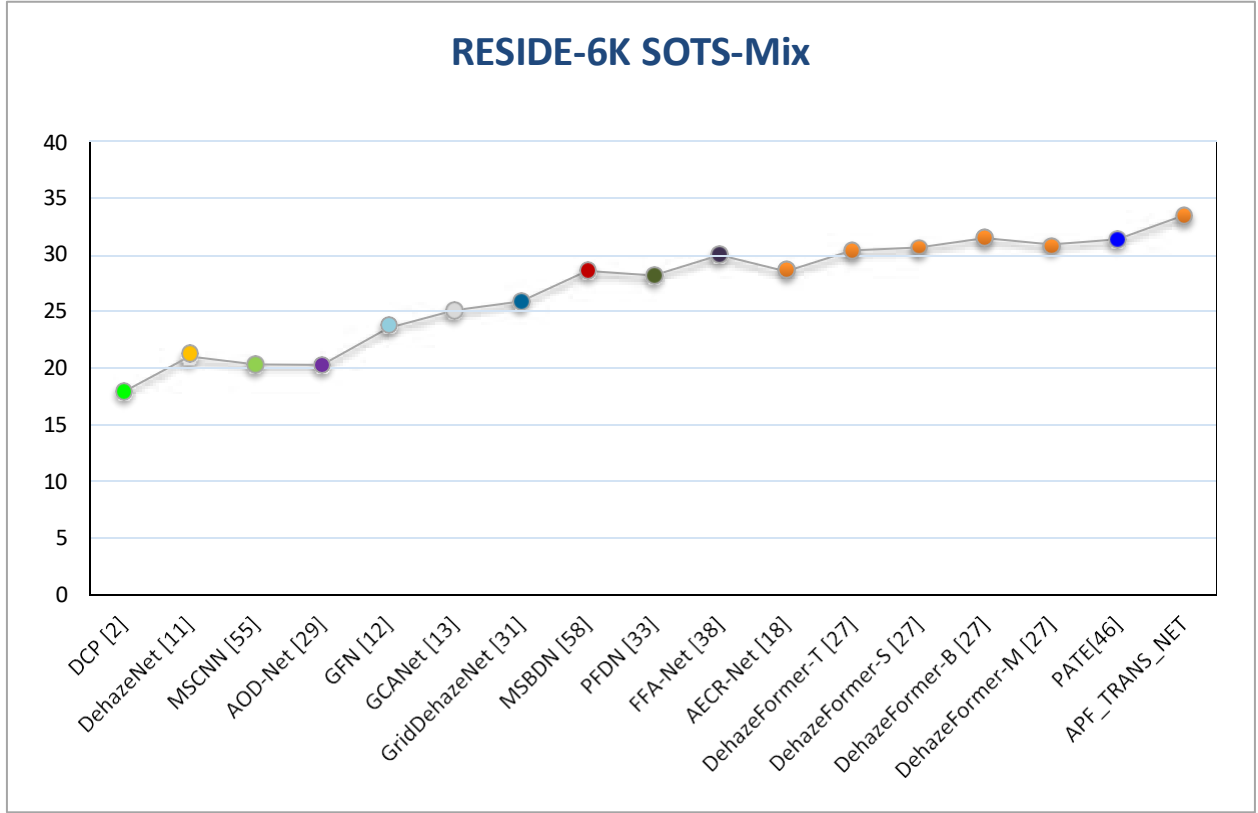


Fig 24.Performance analysis in terms of PSNR in DB4 dataset.

The analysis based on Table 4, Figure 23 and Figure 24 shows that the proposed work outperforms other existing works in DB4 dataset. This shows the performance improvement of +26.13 in terms of PSNR and +0.92 in terms of SSIM. Table 4 shows the significant improvement in our proposed method compared to the state-of-art works in terms of PSNR, SSIM for the dataset RESIDE-6k.

4.6 RS-Haze

RS-Haze is used to train models. We employ 8-bit gamma-corrected RGB images for training and testing in the standard experimental setup. All models are trained using L1 loss for 150 epochs with the same additional parameters as RESIDE-6K. We employ 16-bit linear images for training and testing in MS image dehazing. In order to dehaze images, it analyses the characteristics of MS and RGB images. Keep in mind that during testing, we compute PSNR and SSIM on the gamma-corrected RGB images.

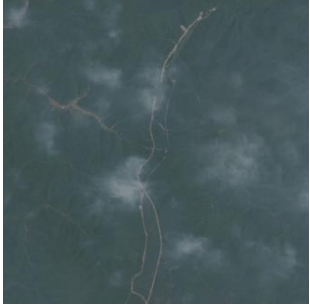

























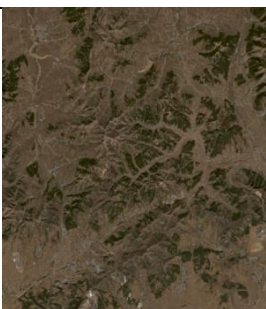
Hazy Images (a)			
Ground Truth Images (b)			

Fig 25. Example images of DB5 dataset (a) Hazy image (b) Corresponding ground truth images

Input Image (a)			
AOD-Net [29] (b)			

<p>GCANet [13]</p> <p>(c)</p>			
<p>PFDN [59]</p> <p>(d)</p>			
<p>FFA-Net [38]</p> <p>(e)</p>			
<p>DehazeFormer – S [27]</p> <p>(f)</p>			
<p><i>APF_TRANS_NET</i></p> <p>(g)</p>			

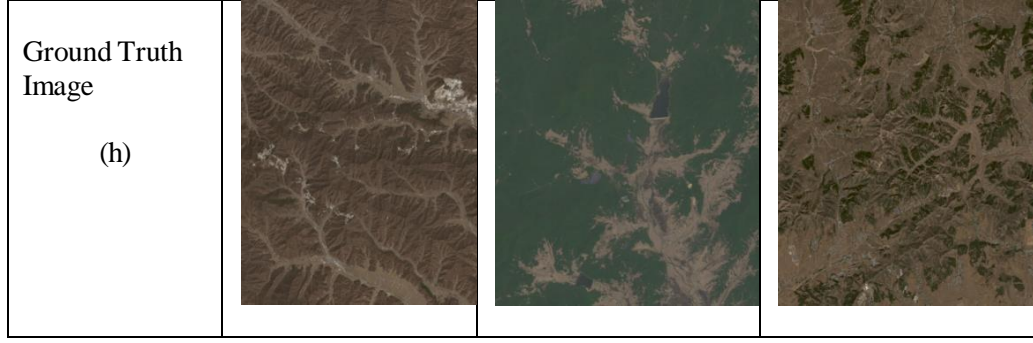


Fig 26.(a - h): Input-output pairs and corresponding ground truth for three example images. (a) Hazy input image (b) AOD-Net [29] (c) output of GCANet [13] (d) output of PFDN [59](e) output of FFA-Net [38] (f) output of DehazeFormer – S [27] (g) output of proposed method(h) ground truth image.

Table 5: Performance analysis of proposed work with existing works in DB5 dataset

Methods	RS-Haze	
	SOTS-mix	
	PSNR	SSIM
DCP [2]	17.86	0.734
DehazeNet [11]	23.16	0.816
MSCNN [55]	22.8	0.823
AOD-Net [29]	24.9	0.83
GFN [12]	29.24	0.91
GCANet [13]	34.41	0.949
GridDehazeNet [61]	36.4	0.96
MSBDN [58]	38.57	0.965
PFDN [33]	36.04	0.955
FFA-Net [38]	39.39	0.969
AECR-Net [18]	35.69	0.959
DehazeFormer-T [27]	39.11	0.968
DehazeFormer-S [27]	39.57	0.97
DehazeFormer-B [27]	39.87	0.971
DehazeFormer-M [27]	39.71	0.971
PATE[46]	38.67	0.97
APF_TRANS_NET	41.56	0.988

The performance comparison of the proposed work with the prior studies in the DB5 dataset is shown in Table 5. Figure 27 displays the performance analysis for images in DB5 in terms of SSIM. The PSNR performance study for images in DB5 is shown in Figure 28.

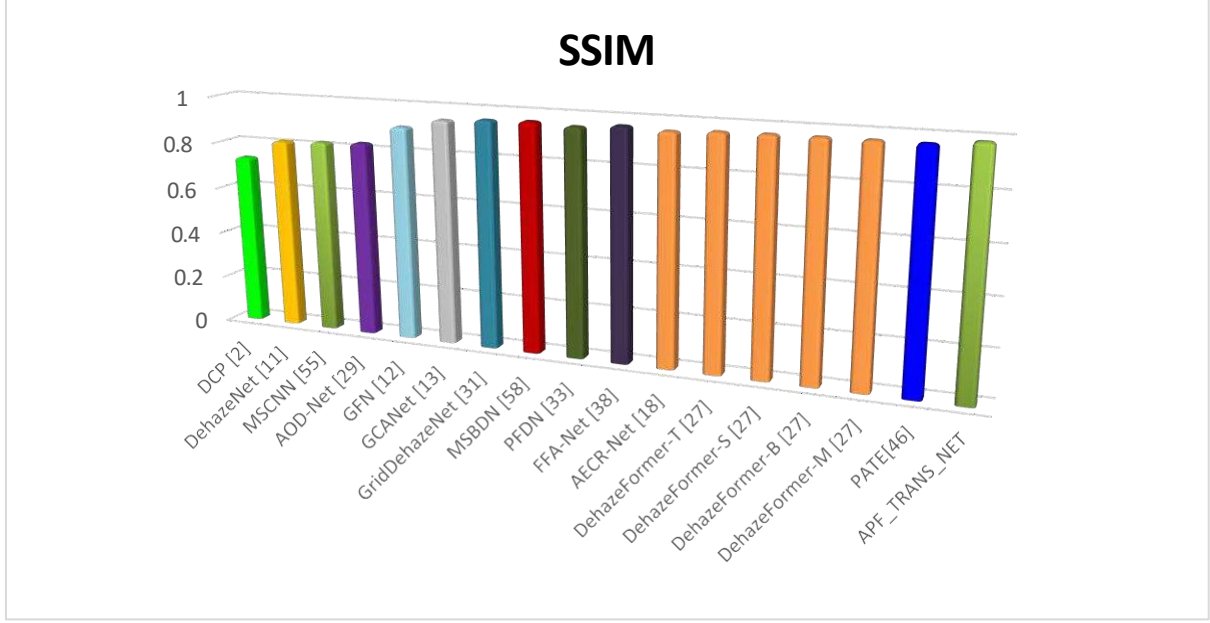


Fig 27.Performance analysis in terms of SSIM in DB5 dataset.

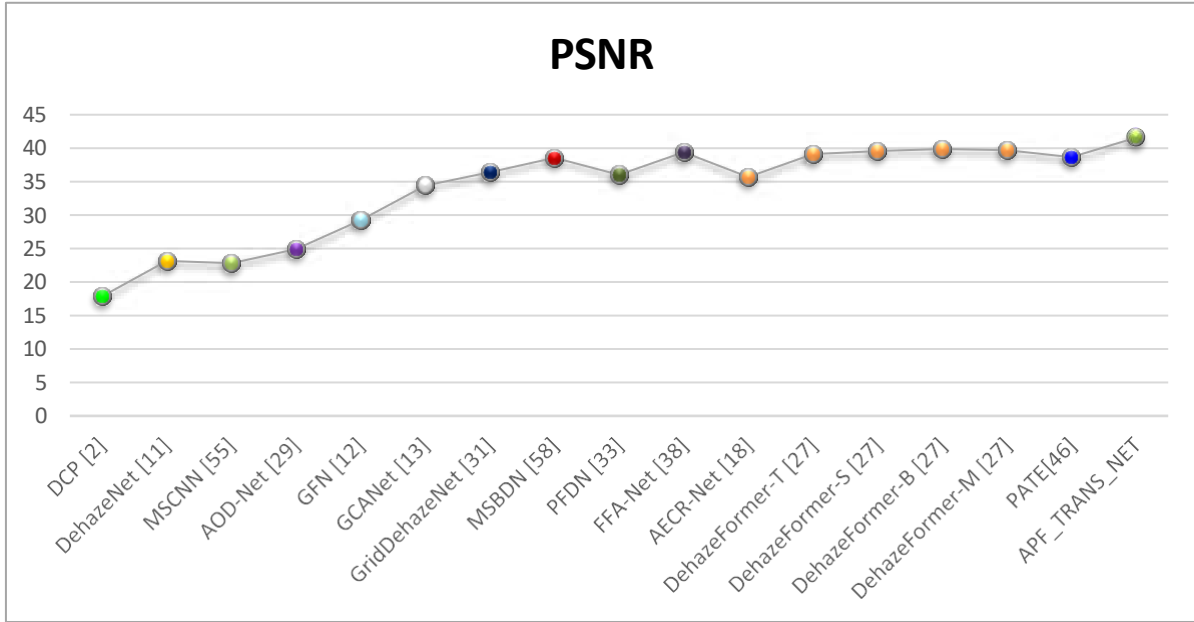


Fig 28.Performance analysis in terms of PSNR in DB5 dataset.

The analysis based on Table 5, Figure 27 and Figure 28 shows that the proposed work outperforms other existing works in DB4 dataset. This shows the performance improvement of +33.11 in terms of PSNR and +0.91 in terms of SSIM. Table 5 shows the significant improvement in our proposed method compared to the state-of-art works in terms of PSNR, SSIM for the dataset RS-Haze.

Conclusion

The concept of vision transformer used for image dehazing is improved in this work in a number of ways, and the proposed method outperforms it on a number of datasets. The proposed model introduce a new module before the actual layer normalization process of transformer with the name Alternate Pooling Fused block, where this block extract the dominant features in different scale which are swapped in vertical and horizontal direction. To sum up, we suggest using locally group self-attention-based transformer, which are typically utilized, in order to prevent some detrimental impacts that are necessary for low-level vision tasks but not important for high-level vision tasks. The five level of transformer network model achieves better result in large-scale remote sensing hazed RS-Haze dataset and achieves 41.56% PSNR and 98.8% in terms of SSIM for RS-Haze dataset. The proposed method also performs admirably in this test. The dual weighted channel and spatial attention block in the proposed dehazing model effectively conserve the data and enhance image visibility.

Abbreviations:

APF_TRANS_NET- Alternate Pooling Fused Transformer Network

MHSA- Multi Head Self Attention

PATE- Prioritized Air light and Transmittance Extraction

DCPDN- Densely Connected Pyramid Dehazing Network

CNN- Convolutional Neural Networks

MLP-Multi Layer Perceptron

ViT-Vision Transformer

GAN-Generative Adversarial Network

DWCA-Dual Weighted Deep Channel Attention

NLP- Natural Language Processing

Declarations:

Availability of data and materials

Only open datasets were used in the paper

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

This work was not supported by any funding scheme or organization

Authors' contributions (individual contributions)

M.Suganthi corresponding author of this manuscript did the process of concept, design, implementation, validation and manuscript writing.

The author Dr. C. Akila contribution is supervision, investigation and suggestion on manuscript writing.

Acknowledgements

There is no funding from any Research or Funding Agency. The authors would also like to thank the Dept. of Computer Science, Kongunadu college of Engineering and Technology, Tiruchy for the continuous support. The authors would like to thank the reviewers and the editor of the journal for their helpful comments which have greatly improved this paper.

References

- [1] S. K. Nayar and S. G. Narasimhan, Vision in bad weather, *Proceedings of the Seventh IEEE International Conference on Computer Vision*, vol2, pp. 820-827 ,1999, doi: 10.1109/ICCV.1999.790306.
- [2] K. He, J. Sun and X. Tang, Single Image Haze Removal Using Dark Channel Prior, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 12, pp. 2341-2353, Dec. 2011, doi: 10.1109/TPAMI.2010.168.
- [3] D. Berman, T. Treibitz and S. Avidan, Non-local Image Dehazing, *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, pp. 1674-1682, 2016,doi: 10.1109/CVPR.2016.185.
- [4] Fattal, Raanan. (2014). Dehazing Using Color-Lines. *ACM Transactions on Graphics*. Vol.34,pp.1-14,2015,doi:10.1145/2651362.
- [5] Tae Keun Kim, Joon Ki Paik and Bong Soon Kang, Contrast enhancement system using spatially adaptive histogram equalization with temporal filtering, in *IEEE Transactions on Consumer Electronics*, vol. 44, no. 1, pp. 82-87, Feb. 1998, doi: 10.1109/30.663733.
- [6] J. A. Stark, Adaptive image contrast enhancement using generalizations of histogram equalization, in *IEEE Transactions on Image Processing*, vol. 9, no. 5, pp. 889-896, May 2000, doi: 10.1109/83.841534.
- [7] Eschbach, Reiner, and Bernd W. Kolpatzik. "Image-dependent color saturation correction in a natural scene pictorial image." U.S. Patent No. 5,450,217. 12 Sep. 1995.

- [8] Schechner, Yoav & Narasimhan, S.G. & Nayar, S.K. Instant dehazing of images using polarization. *Proc. IEEE Conf. Computer Vision and Pattern Recognition*. Vol.1. pp.1-325, 2001, doi:10.1109/CVPR.2001.990493.
- [9] Narasimhan, S.G. & Nayar, S.K. Contrast restoration of weather degraded images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*. Vol.25. pp.713- 724. ,2003, doi:10.1109/TPAMI.2003.1201821.
- [10] Kopf, Johannes & Neubert, Boris & Chen, Billy & Cohen, Michael & Cohen-Or, Daniel & Deussen, Oliver & Uyttendaele, Matt & Lischinski, Dani. Deep Photo: Model-Based Photograph Enhancement and Viewing. *ACM Transactions on Graphics (TOG)*. Vol.27, pp.116. 2008, doi:10.1145/1409060.1409069.
- [11] B. Cai, X. Xu, K. Jia, C. Qing and D. Tao, DehazeNet: An End-to-End System for Single Image Haze Removal, in *IEEE Transactions on Image Processing*, vol. 25, no. 11, pp. 5187-5198, Nov. 2016, doi: 10.1109/TIP.2016.2598681.
- [12] Wenqi Ren, Lin Ma, Jiawei Zhang, Jinshan Pan, Xiaochun Cao, Wei Liu, Ming-Hsuan Yang, Gated Fusion Network for Single Image Dehazing, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3253-3261, 2018.
- [13] D. Chen *et al.*, "Gated Context Aggregation Network for Image Dehazing and Deraining," *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1375-1383, 2019, doi: 10.1109/WACV.2019.00151.
- [14] Li, Boyi & Peng, Xiulian & Wang, Zhangyang & Xu, Jizheng & Feng, Dan. End-to-End United Video Dehazing and Detection. *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, 2017, doi:10.1609/aaai.v32i1.12287.
- [15] Y. Chen, W. Li, C. Sakaridis, D. Dai and L. Van Gool, "Domain Adaptive Faster R-CNN for Object Detection in the Wild," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3339-3348, 2018, doi: 10.1109/CVPR.2018.00352.
- [16] Dai, Dengxin & Sakaridis, Christos & Hecker, Simon & Van Gool, Luc, Curriculum Model Adaptation with Synthetic and Real Data for Semantic Foggy Scene Understanding, 2019.
- [17] Sakaridis, C., Dai, D. & Van Gool, L. Semantic Foggy Scene Understanding with Synthetic Data. *Int J Comput Vis* vol.126. pp.973–992 ,2018, doi:10.1007/s11263-018-1072-8
- [18] H. Wu *et al.*, Contrastive Learning for Compact Single Image Dehazing, *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10546-10555, 2021, doi: 10.1109/CVPR46437.2021.01041.
- [19] Wang, Chao & Shen, Hao-Zhen & Fan, Fan & Shao, Ming-Wen & Yang, Chuan-Sheng & Luo, Jian-Cheng & Deng, Liang-Jian, EAA-Net: A novel edge assisted attention network for single image dehazing. *Knowledge-Based Systems*, Vol.228, pp.107279., 2021, doi:10.1016/j.knsys.2021.107279.

- [20] Shao, Yuanjie& Li, Lerenhan&Ren, Wenqi&Gao, Changxin& Sang, Nong.Domain Adaptation for Image Dehazing.pp.2808-2817.2020.
- [21] Li S, Yuan Q, Zhang Y, Lv B, Wei F, Image Dehazing Algorithm Based on Deep Learning Coupled Local and Global Features. *Applied Sciences*, vol 12.no.17,pp.8552,2022, doi:10.3390/app12178552.
- [22] Jeong, C.Y., Moon, K. & Kim, M. An end-to-end deep learning approach for real-time single image dehazing. *J Real-Time Image Proc* ,vol.20,no.12,2023,doi:10.1007/s11554-023-01270-2.
- [23] Nanfeng Jiang, Kejian Hu, Ting Zhang, Weiling Chen, YiwenXu, Tiesong Zhao, Deep hybrid model for single image dehazing and detail refinement, *Pattern Recognition*,vol.136,2023,doi: 10.1016/j.patcog.2022.109227.
- [24] Dong, P., Wang, B. TransRA: transformer and residual attention fusion for single remote sensing image dehazing. *Multidim Syst Sign Process* vol.33,pp. 1119–1138 2022,doi:10.1007/s11045-022-00835-x.
- [25] Singh A, Kalaichelvi V, DSouza A, Karthikeyan R. GAN-Based Image Dehazing for Intelligent Weld Shape Classification and Tracing Using Deep Learning. *Applied Sciences*,vol.12,no.14,pp.6860,2022,doi:10.3390/app12146860.
- [26] Vaswani, Ashish and Shazeer, Noam and Parmar, Niki and Uszkoreit, Jakob and Jones, Llion and Gomez, Aidan N and Kaiser, Łukasz and Polosukhin, Illia, Attention is All you Need, *Advances in Neural Information Processing Systems*,vol.30,pp. 5998–6008,2017,doi:10.48550/arXiv.1706.03762.
- [27] Yuda Song, Zhuqing He, HuiQian, Xin Du, Vision Transformers for Single Image Dehazing,{IEEE} Transactions on Image Processing,vol.32,pp.1927-1941,2023,doi:10.1109/tip.2023.3256763.
- [28] Parihar, Anil & Java, Abhinav,Densely connected convolutional transformer for single image dehazing, *Journal of Visual Communication and Image Representation*,vol. 90,pp. 103722,2023,doi: 10.1016/j.jvcir.2022.103722.
- [29] B. Li, X. Peng, Z. Wang, J. Xu and D. Feng,AOD-Net: All-in-One Dehazing Network,2017 *IEEE International Conference on Computer Vision (ICCV)*, pp. 4780-4788,2017, doi: 10.1109/ICCV.2017.511.
- [30] He Zhang and Vishal M. Patel, Densely Connected Pyramid Dehazing Network,pp.3194-3203,2018,doi: doi:10.48550/arXiv.1803.08396.
- [31] X. Liu, Y. Ma, Z. Shi and J. Chen, GridDehazeNet: Attention-Based Multi-Scale Network for Image Dehazing, in 2019 IEEE/CVF International Conference on Computer Vision (ICCV),pp.7313-7322,2019,doi: 10.1109/ICCV.2019.00741.
- [32] Ronneberger, O., Fischer, P., Brox, T, U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W., Frangi, A. (eds) *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. MICCAI 2015.

Lecture Notes in Computer Science, vol .9351,pp.5-9, 2015,doi:10.1007/978-3-319-24574-4_28.

[33] Dong, J., Pan, J, Physics-Based Feature Dehazing Networks. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, JM. (eds) Computer Vision – ECCV 2020. ECCV 2020. Lecture Notes in Computer Science, vol .12375,pp.188-204, 2020, doi:10.1007/978-3-030-58577-8_12.

[34] K. He, J. Sun and X. Tang, Single Image Haze Removal Using Dark Channel Prior, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 12, pp. 2341-2353, Dec. 2011, doi: 10.1109/TPAMI.2010.168.

[35] Q. Zhu, J. Mai and L. Shao, A Fast Single Image Haze Removal Algorithm Using Color Attenuation Prior,in *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3522-3533, Nov. 2015, doi: 10.1109/TIP.2015.2446191.

[36] Deng, Q., Huang, Z., Tsai, CC., Lin, CW, HardGAN: A Haze-Aware Representation Distillation GAN for Single Image Dehazing.,In: Vedaldi, A., Bischof, H., Brox, T., Frahm, JM. (eds) Computer Vision – ECCV 2020. ECCV 2020. Lecture Notes in Computer Science, vol 12351,2020, doi:10.1007/978-3-030-58539-6_43.

[37] Zihang Dai and Hanxiao Liu and Quoc V. Le and Mingxing Tan, CoAtNet: Marrying Convolution and Attention for All Data Sizes, *Neural Information Processing Systems*,pp.1-14,2021,doi: 10.48550/arXiv.2106.04803.

[38] Xu Qin and Zhilin Wang and Yuanchao Bai and Xiaodong Xie and Huizhu Jia, FFA-Net: Feature Fusion Attention Network for Single Image Dehazing, *Proceedings of the AAAI Conference on Artificial Intelligence*,vol.34,pp. 11908-11915,2020,doi: 10.1609/aaai.v34i07.6865.

[39] A. Kulkarni and S. Murala, Aerial Image Dehazing with Attentive Deformable Transformers,2023 *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 6294-6303, 2023,doi: 10.1109/WACV56688.2023.00624.

[40] Zhao, Yang & Wang, Yigang, Single Image Dehazing Based on Contrastive Learning and Transformer, *Journal of Physics: Conference Series*, Vol.2450,pp. 012085,2023,doi: 10.1088/1742-6596/2450/1/012085.

[41] Shikhar Vashishth, ShibSankar Dasgupta, Swayambhu Nath Ray, and Partha Talukdar, Dating Documents using Graph Convolution Networks, In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics* ,vol.1, pp. 1605–1615, 2018,doi:10.48550/arXiv.1902.00175.

[42] Jiangming Liu and Yue Zhang. 2017. Attention Modeling for Targeted Sentiment. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Vol. 2*, pp.572–577, 2017,doi:10.18653/v1/E17-2091.

[43] Thang Luong, Hieu Pham, and Christopher D. Manning, Effective Approaches to Attention-based Neural Machine Translation. In *Proceedings of the 2015 Conference* ,pp.1412-1421,doi:10.48550/arXiv.1508.04025.

- [44] Ancuti, C., Ancuti, C.O., Timofte, R., De Vleeschouwer, C, I-HAZE: A Dehazing Benchmark with Real Hazy and Haze-Free Indoor Images, In: Blanc-Talon, J., Helbert, D., Philips, W., Popescu, D., Scheunders, P. (eds) *Advanced Concepts for Intelligent Vision Systems. ACIVS 2018. Lecture Notes in Computer Science*, vol. 11182, 2018 ,doi: 10.1007/978-3-030-01449-0_52.
- [45] Ren, W., Liu, S., Zhang, H., Pan, J., Cao, X., Yang, MH, Single Image Dehazing via Multi-scale Convolutional Neural Networks, In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds) *Computer Vision – ECCV 2016. ECCV 2016. Lecture Notes in Computer Science*, vol. 9906, pp.154-169, 2016, doi:10.1007/978-3-319-46475-6_10.
- [46] Suganthi, M., Akila, C. Prioritized air light and transmittance extraction (PATE) using dual weighted deep channel and spatial attention based model for image dehazing. *Pattern Anal Applic* ,vol.**26**, pp.969–985 ,2023, doi: 10.1007/s10044-023-01187-3.
- [47] G. Meng, Y. Wang, J. Duan, S. Xiang and C. Pan, Efficient Image Dehazing with Boundary Constraint and Contextual Regularization, *2013 IEEE International Conference on Computer Vision*, pp. 617-624, 2013, doi: 10.1109/ICCV.2013.82.
- [48] Fattal, Raanan., Single image dehazing, *ACM Trans. Graph*, vol.27, 2008, doi:10.1145/1399504.1360671.
- [49] X. Wang, R. Girshick, A. Gupta and K. He, Non-local Neural Networks, in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7794-7803, 2018, doi: 10.1109/CVPR.2018.00813.
- [50] Wang, Yequan & Huang, Minlie & Zhu, Xiaoyan & Zhao, Li, Attention-based LSTM for Aspect-level Sentiment Classification, pp.606-615, 2016, doi:10.18653/v1/D16-1058.
- [51] Qin, Xu & Wang, Zhilin & Bai, Yuanchao & Xie, Xiaodong & Jia, Huizhu, FFA-Net: Feature Fusion Attention Network for Single Image Dehazing, 2019, doi: 10.48550/arXiv.1911.07559.
- [52] Chu, Xiangxiang & Tian, Zhi & Wang, Yuqing & Zhang, Bo & Ren, Haibing & Wei, Xiaolin & Xia, Huaxia & Shen, Chunhua, Twins: Revisiting Spatial Attention Design in Vision Transformers, pp.1-12, 2021, doi: 10.48550/arXiv.2104.13840.
- [53] Ancuti, C.O., Ancuti, C., Hermans, C., Bekaert, P, A Fast Semi-inverse Approach to Detect and Remove the Haze from a Single Image. In: Kimmel, R., Klette, R., Sugimoto, A. (eds) *Computer Vision – ACCV 2010. ACCV 2010. Lecture Notes in Computer Science*, vol. 6493, pp.501-514,, 2011, doi: 10.1007/978-3-642-19309-5_39.
- [54] R. Mondal, S. Santra and B. Chanda, Image Dehazing by Joint Estimation of Transmittance and Airlight Using Bi-Directional Consistency Loss Minimized FCN, *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1033-10338, 2018, doi: 10.1109/CVPRW.2018.00137.
- [55] B. Li *et al*, Benchmarking Single-Image Dehazing and Beyond, in *IEEE Transactions on Image Processing*, vol. 28, no. 1, pp. 492-505, Jan. 2019, doi: 10.1109/TIP.2018.2867951.

- [56] Dong, J., Pan, J, Physics-Based Feature Dehazing Networks, In: Vedaldi, A., Bischof, H., Brox, T., Frahm, JM. (eds) Computer Vision – ECCV 2020. ECCV 2020, Lecture Notes in Computer Science, vol 12375, pp-188-204, 2020, doi: 10.1007/978-3-030-58577-8_12.
- [57] J. Li, G. Li and H. Fan, Image Dehazing Using Residual-Based Deep CNN, in *IEEE Access*, vol. 6, pp. 26831-26842, 2018, doi: 10.1109/ACCESS.2018.2833888.
- [58] H. Dong *et al.*, Multi-Scale Boosted Dehazing Network With Dense Feature Fusion, 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2154-2164, 2020, doi: 10.1109/CVPR42600.2020.00223.
- [59] I. Tal, Y. Bekerman, A. Mor, L. Knafo, J. Alon and S. Avidan, NLDNet++: A Physics Based Single Image Dehazing Network, 2020 *IEEE International Conference on Computational Photography (ICCP)*, pp. 1-10, 2020, doi: 10.1109/ICCP48838.2020.9105249.
- [60] A. Golts, D. Freedman and M. Elad, Unsupervised Single Image Dehazing Using Dark Channel Prior Loss, in *IEEE Transactions on Image Processing*, vol. 29, pp. 2692-2701, 2020, doi: 10.1109/TIP.2019.2952032.
- [61] Xiang Li and Wenhai Wang and Xiaolin Hu and Jian Yang, Selective Kernel Networks, In *CVPR*, pp. 510-519, 2019, doi: 10.48550/arXiv.1903.06586.