

TSMD-Net: A two-stage mixed dehazing network with feature fusion and multi-window self attention on Jetson board

Nuthi Raju, Kankanala Srinivas*

School of Electronics Engineering, VIT-AP University, Inavolu, Beside AP Secretariat, Amaravati, 522237, AP, India



ARTICLE INFO

Keywords:

Single image dehazing
Two-stage mixed dehazing network
(TSMD-Net)
Feature fusion (FF)
Multi-window self attention (MWSA)
Jetson Nano board

ABSTRACT

Single image dehazing is a pivotal endeavor focused on enhancing the quality of images by removing haze-induced noise. Achieving this goal requires a delicate equilibrium between retaining spatial intricacies and contextual understanding, particularly crucial in real-world scenarios where complexity is amplified. In response to this challenge, we present the Two-Stage Mixed Dehazing Network (TSMD-Net), designed to reconcile these conflicting objectives by leveraging the foundational U-Net architecture. Our approach entails a dual-phase strategy: initially assimilating contextual comprehension through encoder-decoder frameworks, followed by amalgamating it with high-resolution branches to uphold spatial integrity. To streamline the analysis and comparison of this architecture, we opt to substitute or eliminate nonlinear activation functions with multiplication operations, effectively mitigating system complexity. Furthermore, to foster seamless information exchange across hierarchical layers of the encoder-decoder at varying scales, we introduce the Feature Fusion (FF) mechanism, facilitating the integration of insights from higher layers progressively down to the base layer. This amalgamated data significantly enhances the fidelity of the original image transformation by harmonizing the inclusion of contextual feature maps with the preservation of spatial details, resulting in noise-free, high-quality images. Additionally, we propose the “Multi-Window Self Attention” (MWSA) mechanism, characterized by linear time complexity, as the central block of the encoder-decoder. This module mitigates the confined receptive fields inherent in convolutional neural networks (CNNs), enabling the aggregation of more comprehensive feature-map data. Subsequently, our TSMD-Net undergoes rigorous training and testing on both synthetic and real-world datasets, showcasing superior performance compared to current state-of-the-art (SOTA) networks while demanding fewer processing resources. To validate the robustness and practical utility of our methodology, we implement TSMD-Net on a low-end Jetson Nano board, confirming its effectiveness and efficiency in real-world settings.

1. Introduction

In hazy environments, images often suffer from reduced clarity, leading to visual degradation. This degradation poses challenges for modern computer vision (CV) applications such as object tracking, autonomous vehicle driving, and intelligent surveillance, which require clear images to perform accurate tasks. However, hazy images are typically not conducive to achieving optimal performance in these applications. Consequently, there has been a growing interest among researchers in single image dehazing, aiming to restore clear images from their hazy counterparts. As a result, single image dehazing algorithms have emerged in recent years [31,9]. Nevertheless, image dehazing remains a sig-

nificant challenge in restoring haze-free images. It constitutes a fundamental low-level image restoration operation often considered as a pre-processing stage for high-level vision tasks.

Earlier, the Atmospheric Scattering (AS) model has been a classical framework for image dehazing approaches, formulated as:

$$J(n) = \hat{J}(n) \times T(n) + A \times (1 - T(n)) \quad (1)$$

where $J(n)$ represents the hazy image, $\hat{J}(n)$ is the dehazed image, $T(n)$ denotes the transmission map and A represents the light illumination. The transmission map is calculated as

$$T(n) = e^{-\beta d(n)} \quad (2)$$

* Corresponding author.

E-mail address: srinivas.kankanala@vitap.ac.in (K. Srinivas).

URL: <https://www.vitap.ac.in> (K. Srinivas).

where, β , $d(n)$ are scattering coefficient and depth of the image respectively. The scattering coefficient controls the haze density, the haze degradation in an image varies at each pixel level and depends on the depth $d(n)$ scene point of the camera.

Classic image dehazing methods are grounded in the restoration of hazy images based on atmospheric models, as established by techniques outlined in [23] and [25]. Various approaches have emerged to estimate the transmission map and light illumination, crucial for effective image dehazing. Among these, the Atmospheric Scattering (AS) model serves as the cornerstone for feature techniques in this domain.

In the early stages, image dehazing techniques were predominantly viewed as optimization tasks relying on priors. He et al. [12] introduced a seminal single image dehazing approach utilizing the Dark Channel Prior (DCP), which later became a benchmark in the field. Yeh et al. [40] further improved results by incorporating a pixel-based DCP. In contrast, [24] introduced a contextual regularization technique by exploring the intrinsic boundary constraint of the transmission function to estimate transmission in unknown scenes and remove haze from single images. Zhu et al. [49] proposed a fast haze removal algorithm based on attenuation prior. While these approaches have made significant advancements, conventional methods may falter in recovering hazy images when assumptions or priors are limited. In pursuit of improved dehazing performance, [34] introduced a learning framework tailored for single image dehazing, which was meticulously crafted based on comprehensive research into haze-related image attributes. This approach underscores the effectiveness of learning-based techniques, which have been successfully applied across a range of perceptual tasks. As highlighted in [15], [13], and [29], these techniques have shown notable success in areas such as object detection, tracking, and image understanding. Specifically addressing challenges associated with color distortion in DCP, Chen et al. [8] integrated the dark channel prior into a generative adversarial network, thereby strengthening dehazing capabilities.

CNN-based architectures, as detailed in [5], [35] and [30] are designed for end-to-end mapping of hazy input images to produce clear, restored images. Wang et al. [35] innovatively deviated from traditional approaches by creating a convolutional neural network that directly generates haze-free images via residual image computation. Meanwhile, Shao et al. [30] addressed the challenge of domain adaptation between synthetic and real images by employing a two-way translation network for image translation. To further optimize network performance, several advancements have been proposed. Li et al. [19] incorporated a residual channel attention module along with a multi-layer fusion technique. Qin et al. [26] introduced adaptive learning features in conjunction with a distinctive feature attention module. There has been a notable increase in the utilization of deep learning techniques for image dehazing, aiming to address challenges associated with inaccurate parameter estimation and handcrafted features. Li et al. [17] proposed a benchmark All-in-One Dehaze network (AOD-net), which is a lightweight end-to-end network capable of producing clear images. Additionally, Ren et al. [28] developed a multi-scale deep network (MSCNN). This approach first utilizes a coarse-scale network to predict the transmission map and subsequently refines the network to generate clearer images, demonstrating a comprehensive approach to improving dehazing outcomes.

Recent efforts in image restoration, deraining, and dehazing have increasingly turned to multistage networks, as emphasized in [11], [27], [19]. These approaches typically employ a uniform architecture throughout the stages, with options including a single-scale pipeline or an encoder-decoder design. The single-scale pipeline is particularly adept at maintaining spatial accuracy, yet it may not excel in capturing nuanced semantic details. On the other hand, the encoder-decoder architecture offers a broader context by incorporating more extensive contextual information, albeit potentially at the expense of spatial detail in the resulting images.

To address the intricate challenges of image restoration comprehensively, we developed a novel approach by integrating two distinct architectural frameworks, leading to the creation of the Two-Stage Mixed

Dehaze Network (TSMD-Net). This advanced network showcases remarkable proficiency across a broad spectrum of general image restoration tasks. Within the TSMD-Net architecture, the initial stage is meticulously engineered to prioritize the extraction of multi-scale contextual information. This is accomplished through the deployment of an encoder-decoder framework, which establishes a robust foundation for capturing detailed contextual nuances crucial for achieving accurate image restoration. Building on this foundation, the subsequent stage of the network focuses on preserving fine spatial details at higher resolutions, ensuring that the restored images retain their fidelity and intricate details. To enhance the network's capability for nuanced feature extraction and global information capture, we integrated advanced components such as Multi-Window Self-Attention (MWSA) and Simplified Attentive Channel Module (SACM). These components play a pivotal role in capturing global information and enhancing the network's capacity to extract and process features with greater precision and depth.

To further optimize the stability and efficiency of the training process, Layer Normalization (LN) is implemented consistently across the network architecture. This normalization technique aids in mitigating common training issues and ensures smoother convergence during the training phase. Additionally, to facilitate the seamless propagation and fusion of multi-scale features, Feature Fusion (FF) techniques are incorporated into the network architecture. These techniques enhance the network's adaptability and overall performance by promoting effective feature integration and propagation throughout the network.

Moreover, to refine the feature representation and extract both spatial and channel-specific information, a Dual Attention Module (DAM) is employed. Finally, to elevate the feature resolution and optimize the quality of the restored images, a pixel-shuffle mechanism is incorporated as a concluding processing step within the TSMD-Net framework. Fig. 1 shows a visual representation of the proposed method output image with input hazy image.

The salient contribution of this paper are:

1. Our proposed approach merges two distinct architectures to effectively generate dehazed images, leveraging extensive spatial and contextual information
2. Multi-window self-attention is incorporated in the central block of the encoder-decoder to consolidate the feature map data post-convolution.
3. Spatial information of the features is enriched through the introduction of feature fusion, pixel shuffle, and dual attention modules, enabling our network to capture a greater amount of information.
4. A hybrid loss function is introduced, that combines two distinct loss functions: The Charbonnier loss and the edge loss. This fusion is designed to harness the strengths of both loss functions, thereby enhancing the overall performance of the network.
5. In the hardware implementation of our proposed network for image dehazing, we utilize a Jetson board, a novel application to the best of our knowledge, which has not been explored previously.

The subsequent sections of the paper are structured as follows: Section 2 provides a succinct overview of prevailing image dehazing techniques documented in the literature. Section 3 delves into the intricate exposition of the TSMD-Net architecture proposed herein, along with comprehensive insights into its hardware implementation. Subsequently, Section 4 showcases experimental results, encompassing a comparative analysis with cutting-edge networks and an ablation study. Finally, Section 5 encapsulates the paper's conclusions.

2. Related work

Image dehazing presents a considerable challenge, prompting researchers to explore various approaches underpinned by different assumptions to restore hazy images. Broadly, dehazing methods can be

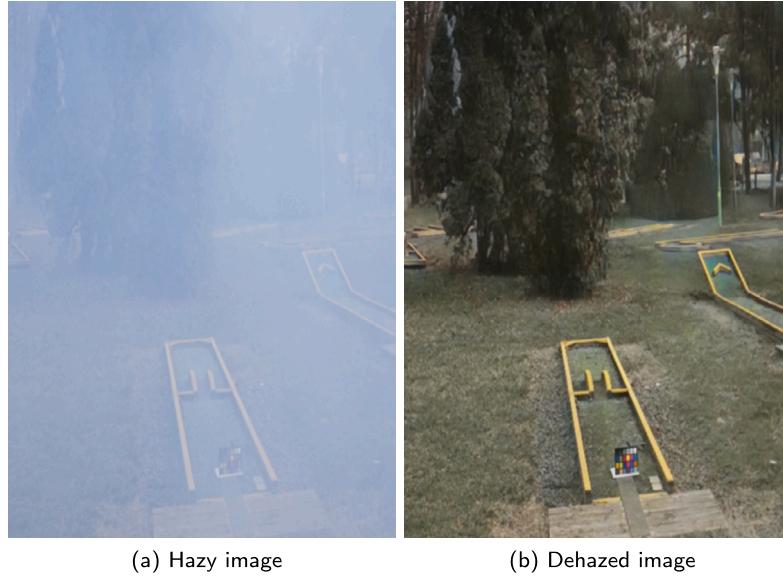


Fig. 1. Visual comparison of input haze and our Output images.

categorized into three main types: prior based, enhancement-based and deep learning-based algorithms tailored for image dehazing.

Image Prior-based dehazing algorithms tackle the dehazing problem by evaluating the air scattering process and investigating the underlying physical causes of image degradation. These algorithms are particularly effective in addressing hazy conditions, especially under cloudy weather. For instance, He et al. [12] proposed the benchmark Dark Channel Prior (DCP) algorithm, which estimates the atmospheric light by leveraging the dark channel. However, this method assumes the presence of at least one color channel in the image, with values approaching zero, thereby excluding single-channel monochrome images from processing. Another notable approach is the Color Attention Prior (CAP) introduced by Zhu et al. [49], which employs a supervised learning technique to compute correlation coefficients and construct a linear model based on the relationship between pixel saturation and the depth of hazy images. While CAP yields superior results compared to DCP, it is constrained by several assumptions from previous works and relies on an ideal atmospheric scattering model. Consequently, its applicability is limited in handling complex hazy images encountered in real-world scenarios.

Image enhancement-based dehazing algorithms aim to restore hazy images by employing techniques such as enhancement, contrast adjustment, noise reduction, among others. For instance, some methods like those discussed in [10] and [50] utilize gamma correction to correct underexposed hazy images. Additionally, the Artificial Multiple-Exposure Image Fusion (AMEI) method proposed in [10] applies laplacian blending on a set of multiply exposed images to produce a dehazed image. Similarly, Zhu et al. [50] introduce the Artificial Multi-Exposure Image Fusion (AMIF) technique, which incorporates brightness adjustment, luminance adjustment, and pixel-wise weight maps. However, these methods may not be universally effective across all lighting conditions. In contrast, Zheng et al. [20] present the Region Adoptive Defogging Enhancement (RADE) algorithm, which leverages plug-in region segmentation. Nonetheless, this method is restricted to hazy images with gray and light backgrounds. Furthermore, to enhance visibility and contrast, Liu et al. [22] propose the Contrast Enhancement and Exposure Fusion (CEEF) technique, which integrates an adaptive histogram equalization method to preserve original colors. However, it may lead to increased processing time during the restoration process.

Continuing the trajectory of advancements, the Deeply Connected Pyramid Dehazing Network (DCPDN) emerged, as detailed in [43] and [46]. This network architecture comprises two distinct sub-networks:

one dedicated to estimating ambient light, while the other focuses on estimating the transmission map. However, these methods often fail to accurately estimate parameters and do not demonstrate significant improvements compared to prior methods based on different principles. In order to solve this issue, researchers developed various CNNs to produce a residual mapping between input hazy image and its relevant output haze free images. For instance, Chen et al. [6] introduced the end-to-end Gated Context Aggregation Network (GCA-Net), which employs a fusion model to combine multilevel features and utilizes novel smooth dilated convolution techniques to eliminate artifacts.

Subsequently, attention-based methods have emerged within the realm of deep learning-based techniques for image dehazing. Notably, the Areal Image Dehazing (AID) transformer [16] introduces a spatial attention offset extractor aimed at recovering dark regions in hazy images. Lin et al. [21] propose the Multi-Scale Attention Feature Fusion Network (MSAFF-Net), which leverages channel attention and multi-scale spatial attention modules to enhance network performance. Chen et al. [9] introduced a novel approach termed Detail Enhanced Convolution and Content-Guided Attention (DEA-net). This method effectively captures spatial importance maps from individual channels by leveraging content-guided attention mechanisms. This enables the recovery of uneven haze distributions with enhanced precision. Additionally, Yi et al. [41] present the Priors-Assisted Dehazing Network (PAD-Net), designed to enhance the extraction of non-uniform feature distributions from hazy images by incorporating relevant image priors through attention monitoring and detail preservation mechanisms. However, it's important to note that these advanced techniques may demand significant computational resources due to their intricate attention mechanisms and feature fusion processes.

3. Proposed method

In this section we introduced a two stage mixed dehazing network (TSMD-Net) for image dehazing. This network consists with encoder-decoder subnetwork, attentive channel subnetwork (ACS-Net) and DAM.

Fig. 2 depicts the structure of our proposed TSMD-Net for single image dehazing. The primary goal of our network is to establish a robust architecture that maintains the performance of existing networks while minimizing computational costs. The network comprises two hierarchical structures. The first structure adopts an encoder-decoder architectural style, allowing it to capture contextual features with a large

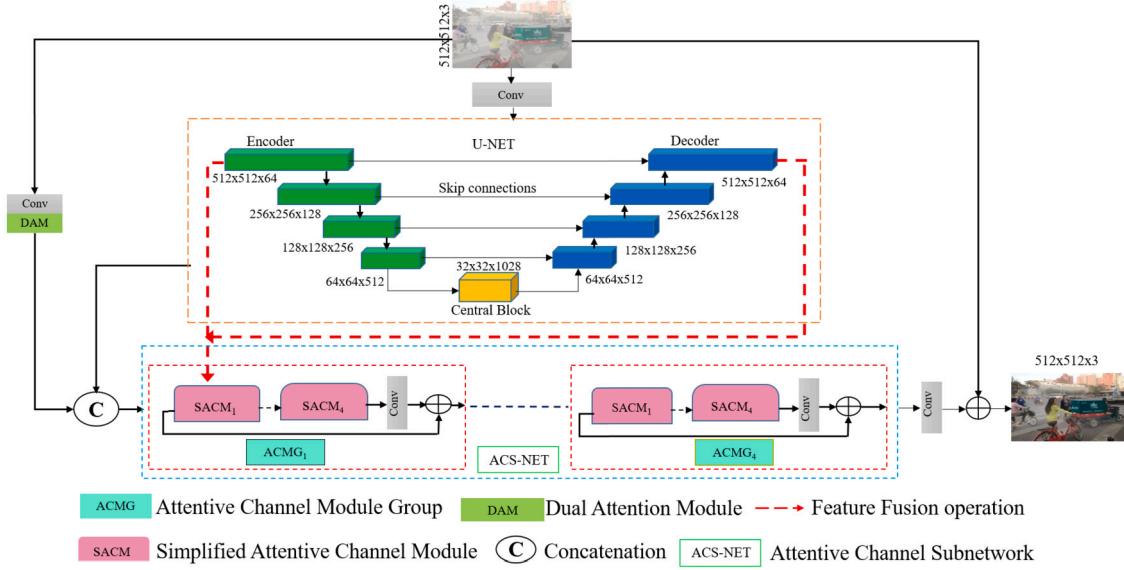


Fig. 2. TSMNet architecture for image dehazing.

context. The second structure focuses on preserving local feature map information by operating at the resolution of the original input image.

Based on the findings from research [7], we opted to replace the U-net architecture [29] with a simpler approach, where the nonlinear activation function is substituted by a basic multiplication operation. Additionally, we replaced the conventional convolution function with a MWSA within the central block. This hybrid architecture amalgamates more feature map information and concentrates on lower resolutions by employing convolutions for spatial down-sampling. Furthermore, we introduced a feature fusion approach to seamlessly blend the hierarchical structures of the network.

Initially a 3×3 convolution is applied on a hazy image $J \in R^{H \times W \times 3}$ to extract low-level features $\chi_0 \in R^{H \times W \times C}$. An encoder-decoder network processes which produce encoder features $E[\chi_{e1}, \chi_{e2}, \chi_{e3}, \chi_{e4}]$ and decoder features $D[\chi_{d1}, \chi_{d2}, \chi_{d3}, \chi_{d4}]$. These features are then integrated through skip connections to reconstruct a dehazed image. Both spatial and channel dimensions within the feature tensor are shared to facilitate feature fusion. Shallow feature χ_{sh} is obtained through convolution and a DAM prior to feature fusion. Deep features $\chi_{de} \in R^{H \times W \times C}$ are then obtained by the feature χ_2 after concatenation of χ_1 and χ_{d4} features are then fed to ACS-Net. The feature fusion module is introduced between encoder-decoder and ACS-Net. Furthermore, a 3×3 convolution is applied to the refined features to obtain the residual image $\chi \in R^{H \times W \times 3}$. The dehazed image is $\hat{J} = \chi + J$ obtained by adding the hazy image with the residual image.

3.1. Encoder-decoder subnetwork

In the proposed method, a classic U-shaped architecture with skip connections is adopted, as illustrated in Fig. 2. Skip connections in neural networks, especially for image processing, enable direct transmission of images from the contracting to the expanding path, preserving both high and low-level features. They facilitate direct information flow, improve gradient propagation, and enhance feature reuse, thus improving training performance and output quality. Several components are added or replaced in the encoder-decoder subnetwork

3.1.1. Layer normalization (LN)

LN has become increasingly popular in various techniques and has shown significant improvements when used with transformers in computer vision applications. Its ability to stabilize the training process has led to its incorporation into the U-Net architecture.

3.1.2. Multi-window self-attention (MWSA)

One of the main challenges in using CNNs for image dehazing is their limited receptive field. However, applying self-attention in vision tasks poses certain challenges, such as increased training complexity. Due to the quadratic complexity resulting from global calculations with respect to the number of tokens, as shown in Eq. (3), traditional self-attention is not well-suited for representing high-resolution images.

$$O_{MSA} = 4HWC^2 + 2(HW)^2C \quad (3)$$

To accomplish this objective, MWSA is employed as the central block in the encoder-decoder network, replacing the traditional CNNs. This design choice reduces training costs by enabling attention to focus on smaller resolutions. Additionally, to mitigate time complexity, the spatial dimension is substituted with MWSA across channels. This adjustment helps streamline computational processes, contributing to improved efficiency in the network.

The **query**, **key** and **value** matrices $q, k, v \in R^{H \times W \times C}$ are reshaped to $\hat{q}, \hat{k}, \hat{v} \in R^{(HW) \times (\frac{C}{a}) \times a}$, here a is number of aspects. The self-attention process computes the attention matrix, which is displayed in Eq. (4).

$$\text{attention}(\hat{q}, \hat{k}, \hat{v}) = \text{SoftMax}\left[\left(\frac{\hat{q}, \hat{k}}{\varsigma}\right)\right] \cdot \hat{v} \quad (4)$$

Here ς is learning scaling parameter.

Before using softmax function ς is used to modify magnitude of the dot product of \hat{q} and \hat{k} . At the end the attention matrix is reconfigured to have the original dimensions of $R^{H \times W \times C}$. Additionally, the computation switched from quadratic complexity O_{MSA} to linear complexity O_{MWSA} .

$$O_{MWSA} = 5HWC^2 + HWC \quad (5)$$

3.2. Attentive channel subnetwork (ACS-Net)

Image recovery necessitates a delicate balance between spatial features and contextualized information. To retain small details from the hazy image in the restored image, we introduce the ACS-Net. It comprises four ACMGs, each containing SACMs. Fig. 3 illustrates the architecture of an ACMG.

Let consider $\chi_{ei} \in R^{\frac{H}{l^2} \times \frac{W}{l^2} \times l^2 \times C}$ and $\chi_{di} \in R^{\frac{H}{l^2} \times \frac{W}{l^2} \times l^2 \times C}$ are the outputs of encoder and decode at i th level respectively ($i = 1, 2, 3, 4$).

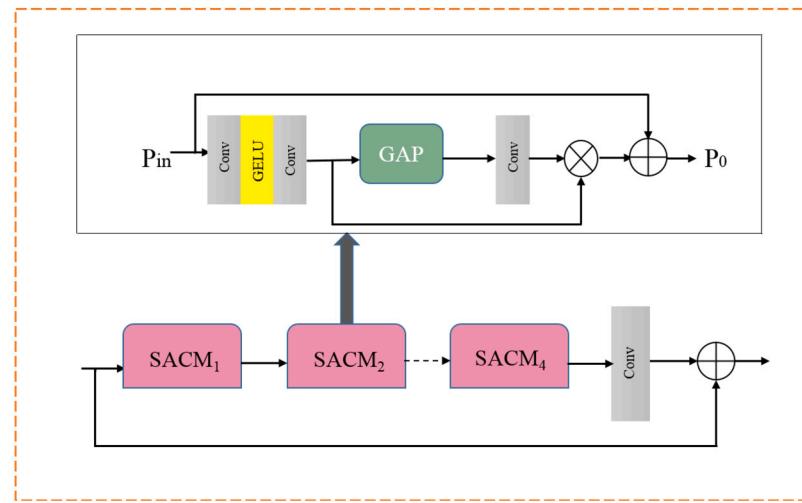


Fig. 3. The architecture of attentive channel module group (ACMG). Each ACMG further contains multiple SACMs.

The feature fusion information F_i at each block of ACS-Net is expressed as

$$F_i = ACMG_i(F_{i-1}) + \chi_{ei} + \chi_{di} \quad (6)$$

$$F_0 = DAM(Conv(J)), \chi_{d4} \quad (7)$$

Where J is hazy image, $ACMG(\cdot)$ consists of multiple attentive channel modules and a 3×3 convolution, and $DAM(\cdot)$ is the dual attention module operation.

3.2.1. Simplified attentive channel module (SACM)

The SACM [7] is integrated to extract features at multiple scales, thereby conserving computational resources and enhancing computation performance. To enhance feature capture and accelerate convergence, we utilized the GELU activation function and employed two convolution layers, as depicted in Fig. 3. To obtain the output from an input we perform the following steps:

$$P_0 = P_{in} + P_{in1} \otimes Conv(GAP(P_{in1})) \quad (8)$$

$$P_{in1} = Conv(GELU(Conv(P_{in}))) \quad (9)$$

$$GELU(\tau) = \tau \times \Phi(\tau) \quad (10)$$

Where GAP is global average pooling and Conv is the convolution operation.

3.2.2. Feature fusion

Feature fusion mechanism [42] is introduced in our network aimed at enhancing feature quality by minimizing feature loss during the up-down sampling process in the encoder-decoder architecture. This mechanism facilitates the incorporation of additional information into the next hierarchical level, thereby contributing to the stability of training models. In Fig. 4, the feature fusion is represented.

3.3. Dual attention module (DAM)

In earlier low-level vision techniques [37], [48] and [47], attention mechanisms played a significant role. In our proposed method, the ACS-Net is utilized to extract spatial information by integrating the encoder-decoder based feature transmission. Additionally, the DAM is employed to capture both spatial and channel information in the low-level features. DAM, as depicted in Fig. 5, comprises two components: spatial attention (SA) [38] and channel attention (CA) [14].

Spatial attention involves performing global max pooling and global average pooling simultaneously on the input feature. These outputs are concatenated and then passed through a convolutional layer followed

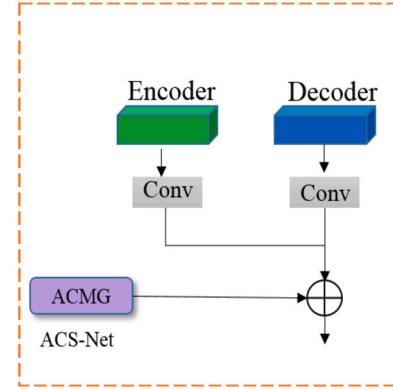


Fig. 4. Feature fusion between an encoder-decoder subnetwork and ACS-Net.

by a sigmoid activation function. Channel attention, on the other hand, employs global average pooling to extract global context, which is then convoluted twice and passed through a sigmoid activation function. The output features from these two channels are fused to the next level for feature enhancement.

Mathematical representation as follows:

$$D(Z_0) = Z_0 + Conv(Z_s + Z_c) \quad (11)$$

$$Z_s = Z_1 \otimes SA(Z_1) \quad (12)$$

$$Z_c = Z_1 \otimes CA(Z_1) \quad (13)$$

$$Z_1 = Conv(GELU(Conv(Z_0))) \quad (14)$$

$$SA(Z_1) = \sigma(Conv([GAP(Z_1), GMP(Z_1)])) \quad (15)$$

$$CA(Z_1) = \sigma(Conv(GELU(Conv(GAP(Z_1))))) \quad (16)$$

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (17)$$

3.4. Loss function

To optimize the performance of our network, we have adopted a novel hybrid loss function that combines the Charbonnier loss [4] and edge loss. It can be expressed as follows:

$$L = L_{char}(\chi, I) + \lambda L_{edge}(\chi, I) \quad (18)$$

Where χ is residual image, I is ground truth image. In eq (19) L_{char} is a Charbonnier loss defined as:

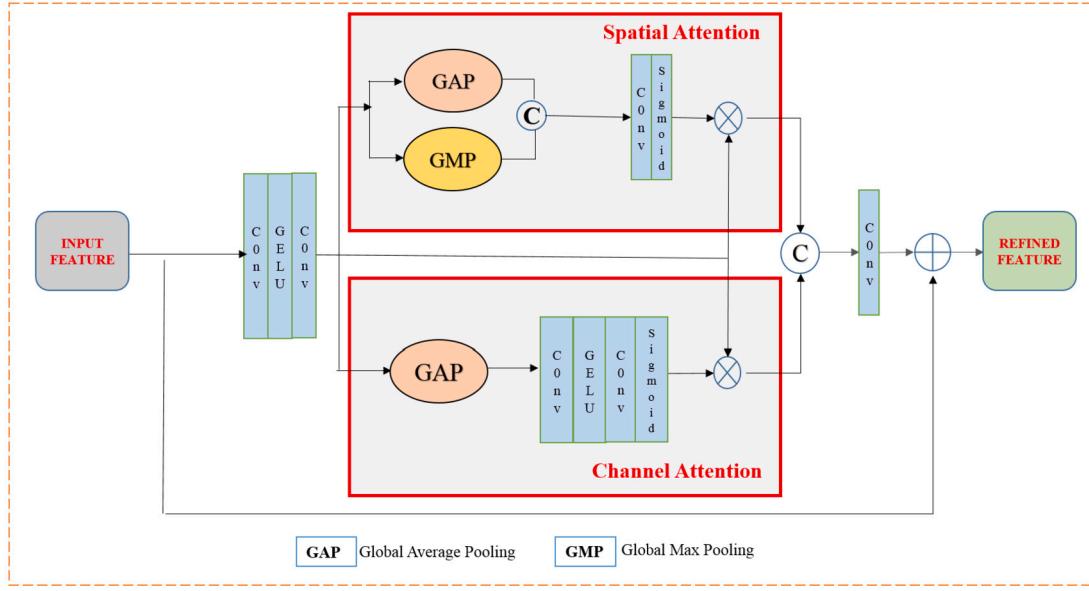


Fig. 5. Feature extraction with channel attention and spatial attention in Dual attention module.

Algorithm 1

Algorithm: During the training of our network at each epoch:
Input: $J(n) \Rightarrow$ input image, $T(n) \Rightarrow$ transmission map, from RESIDE-6k dataset
number of epochs $\Rightarrow e$.

Output: Predicted residual map

- Weight W is randomly initialized
- while $i < e$ (here $i = 1, 2, 3, \dots, e$) do
- Select random haze image $J(n)$ and corresponding transmission map $T(n)$
- Calculate predicted transmission map $T'(n) = TSMnet(T(n))$
- Find $Loss(I) =$ Hybrid loss (T, T')
- Weights are updated using Adom optimizer
- increment i
- end while.

$$L_{char}(\chi, I) = \sqrt{\|\chi - I\|^2 + \epsilon^2} \quad (19)$$

L_{edge} is the edge loss and it is presented as:

$$L_{edge}(\chi, I) = \sqrt{\|\Delta(\chi) - \Delta(I)\|^2 + \epsilon^2} \quad (20)$$

here Δ is laplacian operation.

4. Results and discussion

4.1. Dataset

We evaluated the performance of our proposed TSMD-net against other state-of-the-art networks using the RESIDE [18] dataset. This dataset is a widely recognized benchmark for dehazing and includes both real and synthetic images. It comprises several subsets: ITS, OTS, HSTS, RTTS, HTTS, and URHI. To manage the time cost of training, we randomly selected 6,000 pairs of hazy/clean images from the ITS, OTS, and HSTS subsets and labeled them as RESIDE-6K. Out of these, 5,000 pairs were used for training and 1,000 pairs for testing to improve the generalization ability of TSM-Net. Additionally, we assessed the robustness of our TSMD-Net using the I-Haze [1], O-Haze [2], Dense Haze [3] and real-world datasets.

4.2. Experiment setup

The proposed network was trained on Google Colab pro+ with single Tesla V100 GPU and implemented by using Pytorch 1.10.0. At each level

Table 1

Hyperparameters used in TSMD-Net.

Parameters	Values
Patch size	256×256
Optimizer	Adam ($\beta_1 = 0.9, \beta_2 = 0.9999$)
Batch size	12
Weight Decay	0.0004
Learning rate	0.0001 to 0.000001
Epoches	300

we utilize the following configurations as: [1,1,1,28] encoder blocks, [2,2,1,0] ASM blocks, [1,1,1,1] decoder blocks and one central block as mediator between encoder-decoder structure in our proposed network. To improve the robustness of our network we also applied various data augmentation techniques such as vertical and horizontal flips. Our network was trained from scratch in an end-to-end manner and the hyperparameters used for our network training are adopted from DEA-net [9], is shown in Table 1.

4.3. Evaluation metrics

It is a challenging yet important criterion to evaluate the quantity and quality of dehazed images. No single metric can accurately assess the quality of a dehazed image, even if similar images may receive varying quality ratings across different dimensions from a human perspective. Therefore, this study considers several metrics, including reconstruction quality, contrast, similarity, gradients, saliency, and brightness, to comprehensively assess various aspects of image quality.

The performance of the proposed network was evaluated using various quality assessment metrics, including Peak-Signal to Noise Ratio (PSNR) [33], Structural Similarity Index Metric (SSIM) [36], Feature Similarity Index Metric (FSIM) [45], Gradient Magnitude Similarity Deviation (GMSD) [39], Visual Saliency-based Index (VSI) [44], and Absolute Mean Brightness Error (AMBE) [32]. The selected parameters provide a well-rounded evaluation by covering different aspects of image quality, including noise reduction (PSNR), structural similarity (SSIM), feature preservation (FSIM), edge preservation (GMSD), brightness consistency (AMBE), and perceptual relevance (VSI). Besides, parameters like SSIM, FSIM, and VSI are closely aligned with human visual perception, making them highly relevant for assessing the visual quality of images. The higher values of PSNR, SSIM, FSIM, VSI and lower value of GMSD and AMBE will provide better output dehazed image.

Table 2

Quantitative Analysis of Dehazing Performance section delineates the comparative performance of our network against several SOTA networks on the RESIDE-6K [18] dataset. Enhanced performance outcomes are emphasized using bold formatting for clarity and emphasis.

Method	PSNR↑	SSIM↑	FSIM↑	AMBE↓	GMSD↓	VSIE↑
AOD-Net [17]	15.73	0.772	0.902	31.30	0.092	0.965
AMEI-Net [10]	18.64	0.789	0.927	22.25	0.074	0.977
AMIF-Net [50]	18.72	0.822	0.935	16.33	0.064	0.977
RADE [20]	16.08	0.715	0.832	19.96	0.146	0.939
CEEF [22]	17.60	0.730	0.895	22.28	0.089	0.967
MSCNN [28]	18.48	0.824	0.947	18.07	0.057	0.979
AID-Net [16]	12.69	0.621	0.820	45.96	0.158	0.921
DEA-Net [9]	16.77	0.795	0.918	20.56	0.076	0.976
PAD-Net [41]	22.93	0.903	0.961	11.56	0.044	0.989
Ours	25.31	0.924	0.972	07.84	0.032	0.993

Table 3

Quantitative analysis of Dehazing Performance section delineates the comparative performance of our network against several SOTA networks on the I-Haze [1] dataset. Enhanced performance outcomes are emphasized using bold formatting for clarity and emphasis.

Method	PSNR↑	SSIM↑	FSIM↑	AMBE↓	GMSD↓	VSIE↑
AOD-Net [17]	116.01	0.752	0.845	20.48	0.151	0.933
AMEI-Net [10]	15.01	0.724	0.838	21.10	0.144	0.940
AMIF-Net [50]	15.54	0.702	0.890	15.55	0.096	0.955
RADE [20]	11.83	0.561	0.854	37.96	0.117	0.932
CEEF [22]	15.51	0.724	0.901	20.34	0.086	0.956
MSCNN [28]	15.73	0.744	0.866	19.59	0.127	0.946
AID-Net [16]	14.54	0.670	0.824	15.70	0.175	0.936
DEA-Net [9]	13.57	0.582	0.764	15.70	0.223	0.907
PAD-Net [41]	15.97	0.745	0.835	15.77	0.154	0.929
Ours	22.29	0.849	0.922	07.46	0.083	0.976

4.4. Quantitative analysis

We conducted a comprehensive benchmarking study of our TSMD-Net against ten prominent dehazing networks: AOD-Net [17], AMEI-Net [10], AMIF-Net [50], RADE [20], CEEF [22], MSCNN [28], AID-Net [16], DEA-Net [9] and PAD-Net [41]. To ensure a rigorous and fair comparison, we retrained all the existing SOTA networks on Google colab Pro+ using an identical training dataset and conducted visual assessments. The objective evaluation average results on the RESIDE-6k [18] dataset are presented in Table 2. Remarkably, our TSMD-Net consistently outperformed all the compared state-of-the-art networks, showcasing significant advancements in dehazing performance.

On the SOTS-mixed dataset [18], our TSMD-Net achieved a PSNR of 25.31 dB and SSIM score of 0.924. When compared to other leading state-of-the-art networks, our model demonstrated superior performance. Particularly noteworthy is the comparison with AID-Net [16], AOD-Net [17], RADE [20] and CEEF [22], where our TSMD-Net exhibited improvements of 12.62 dB, 9.85 dB, 9.23 dB, and 7.7 dB, respectively. These results clearly indicate that our network produces higher-quality dehazed images.

Furthermore, the other prominent dehazing networks, including AMEI-Net [10], AMIF-Net [50] and MSCNN [28], achieved PSNR values around 18 dB. The recent technique PAD-Net [41] also demonstrated notable improvements in dehazing metrics. In addition to PSNR and SSIM, metrics such as AMBE and GMSD were evaluated. Our TSMD-Net achieved the lowest scores in these metrics, indicating that our network effectively maintains the average brightness value of the image and preserves the gradients in the output dehazed image.

In our study, we evaluated the performance of our network using the I-haze [1] and O-haze [2] datasets, which encompass data from both indoor and outdoor settings. Tables 3 and 4 present the various metric mean values of SOTA networks across these datasets. A comprehensive analysis of these tables reveals that our network consistently delivers

Table 4

Quantitative Analysis of Dehazing Performance section delineates the comparative performance of our network against several SOTA networks on the O-Haze [2] dataset. Enhanced performance outcomes are emphasized using bold formatting for clarity and emphasis.

Method	PSNR↑	SSIM↑	FSIM↑	AMBE↓	GMSD↓	VSIE↑
AOD-Net [17]	16.63	0.676	0.840	20.61	0.116	0.924
AMEI-Net [10]	14.24	0.738	0.926	49.08	0.070	0.977
AMIF-Net [50]	16.15	0.825	0.901	34.31	0.072	0.962
RADE [20]	13.18	0.491	0.594	20.02	0.209	0.827
CEEF [22]	12.22	0.490	0.756	57.25	0.144	0.901
MSCNN [28]	16.36	0.795	0.912	37.05	0.066	0.973
AID-Net [16]	16.68	0.672	0.835	19.77	0.111	0.922
DEA-Net [9]	17.01	0.684	0.845	19.86	0.112	0.927
PAD-Net [41]	18.26	0.702	0.856	12.41	0.130	0.948
Ours	19.11	0.715	0.861	11.96	0.113	0.941

Table 5

Quantitative Analysis of Dehazing Performance section delineates the comparative performance of our network against several SOTA networks on the Dense-Haze [3] dataset. Enhanced performance outcomes are emphasized using bold formatting for clarity and emphasis.

Method	PSNR↑	SSIM↑	FSIM↑	AMBE↓	GMSD↓	VSIE↑
AOD-Net [17]	12.15	0.471	0.537	41.27	0.276	0.798
AMEI-Net [10]	10.48	0.473	0.607	59.93	0.252	0.824
AMIF-Net [50]	12.84	0.497	0.607	34.95	0.251	0.837
RADE [20]	09.72	0.374	0.710	56.62	0.224	0.864
CEEF [22]	12.48	0.472	0.661	39.06	0.231	0.848
MSCNN [28]	10.22	0.464	0.643	63.65	0.239	0.838
AID-Net [16]	13.45	0.470	0.561	30.59	0.271	0.822
DEA-Net [9]	15.95	0.533	0.754	26.29	0.223	0.901
PAD-Net [41]	12.15	0.466	0.553	40.69	0.272	0.817
Ours	17.47	0.631	0.799	16.92	0.191	0.924

superior quality results compared to other leading SOTA networks. This underscores our network's effectiveness in effectively mitigating haze across varying intensities.

Additionally, Table 5 provides a comparative analysis on the Dense-haze [47] dataset, where our network demonstrates significant outperformance relative to other SOTA networks. From Table 5 it is observed that our proposed network achieves PSNR 17.47 dB and SSIM 0.631 scores and lower values of AMBE, GMSD and higher values of FSIM and VSIE.

4.5. Qualitative analysis

In Fig. 6, we present a visual comparison of exemplar results from SOTA networks alongside our proposed method, as evaluated on the RESIDE [18] dataset. AOD-Net [17] exhibits a notable dark region in the sky area, highlighted by a red box, attributed to inaccurate transmission map estimation. Furthermore, it fails to effectively remove haze from detailed regions. Both AMEI [10] and AMIF [50] introduce color distortions in their outputs. In contrast, our proposed approach leverages deep residual learning to better preserve color information.

In Fig. 7, RADE [20] and AID-Net [16] display uneven dehazing, leaving behind residual haze and lacking in detail fidelity in the background. MSCNN [28] offers improved visual results; however, it falls short in completely eliminating haze particles in detailed areas, as evident in the second example. CEEF [22] introduces high color contrast and artifacts into the dehazed images thereby undermining their overall quality and visual fidelity. Both DEA-Net [9] and PAD-Net [41] grapple with color saturation issues, with DEA-Net also yielding lower PSNR values across both examples.

In Fig. 8 and 9, we present visual sample results comparing our proposed network with SOTA networks on the I-Haze [1] and O-Haze [2] datasets, respectively. A notable observation is that many of the existing mainstream networks face challenges in effectively removing haze.

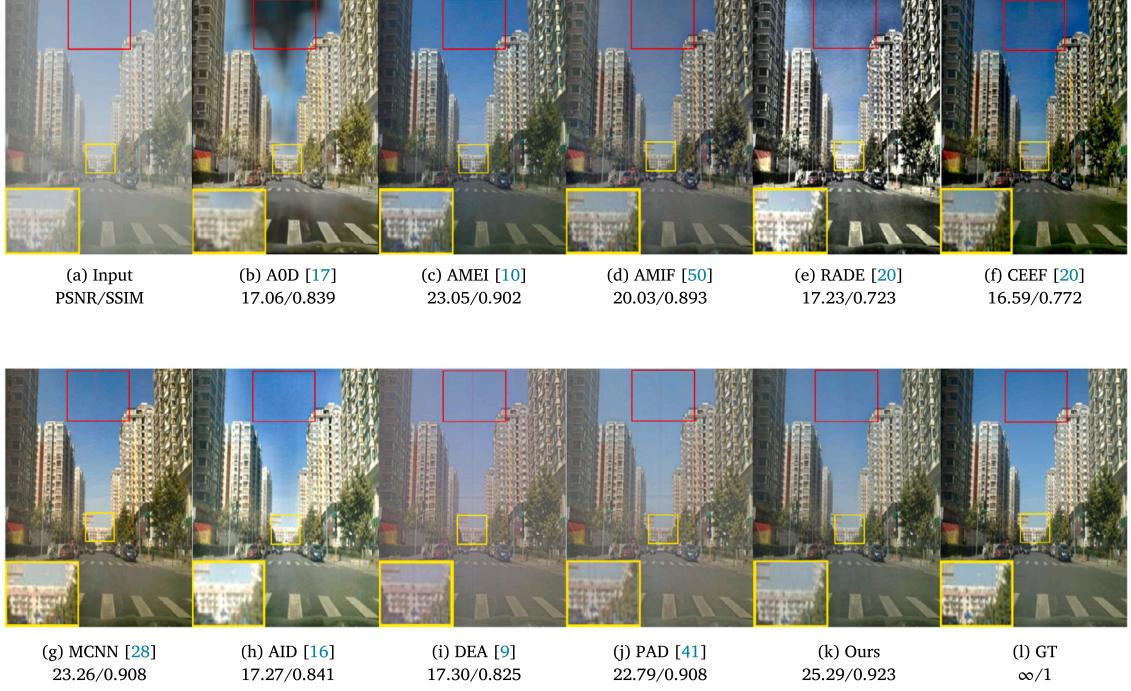


Fig. 6. The results of a qualitative comparison between our proposed method and ten leading SOTA, as evaluated on the RESIDE-6K dataset. Zoom in for better view.



Fig. 7. The results of a qualitative comparison between our proposed method and ten leading SOTA, as evaluated on the RESIDE-6K dataset. Zoom in for better view.

Specifically, the networks depicted in images b, f, and g across both figures tend to produce darkened backgrounds. Additionally, in images c of both datasets, the haze remains largely unaddressed, resulting in significant loss of image content and color fidelity. Images h and i exhibit lower contrast and color distortion, with image i further registering lower PSNR and SSIM values. While image j in both figures presents relatively better results, it still falls short in significantly reducing haze. In stark contrast, our network excels in nearly eliminating all haze, preserving intricate background details, and generating higher fidelity dehazed results. Furthermore, our approach consistently produces outputs that

closely align with the ground truth images, underscoring its superior performance in haze removal and image restoration.

To validate the effectiveness of our proposed image dehazing technique on Dense-Haze [3] we focused on specific regions within the dehazed images, as illustrated in Fig. 10. As indicated by the highlighted areas in Fig. 10 (denoted by yellow boxes), our suggested approach surpasses the SOTA method in several key aspects. Notably, our technique excels in reconstructing Dense-Haze with enhanced detail retention, superior color restoration, and improved contrast. These comparative analyses serve to underscore the superior performance and capabilities of our proposed dehazing method.

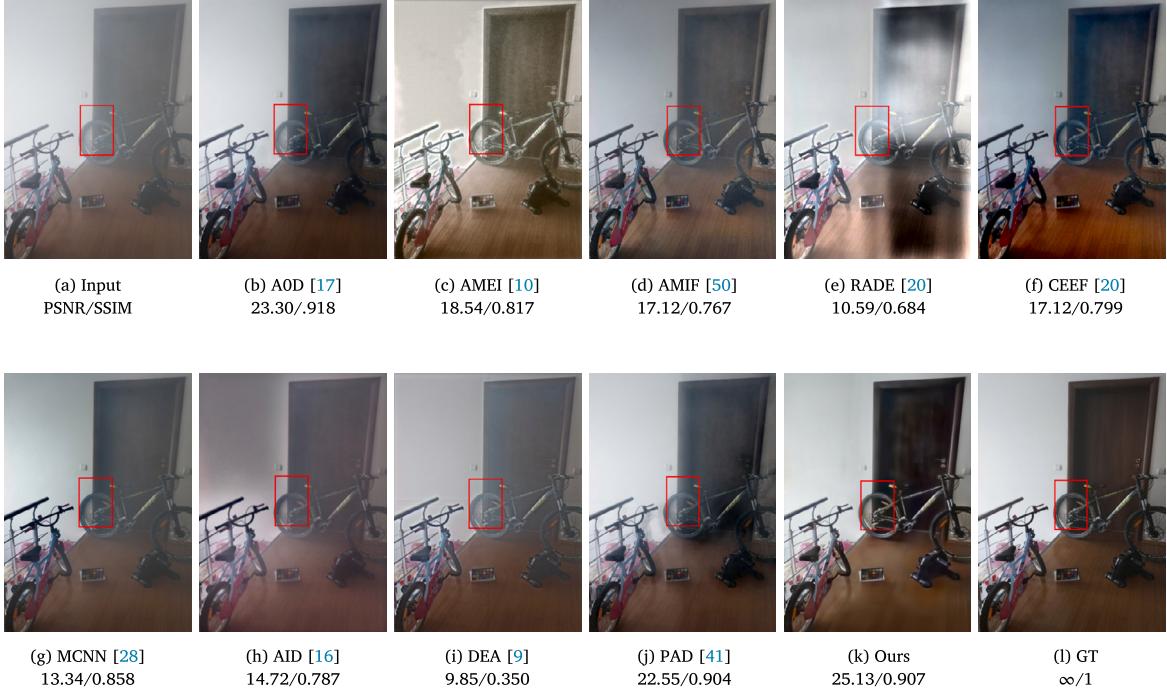


Fig. 8. The results of a qualitative comparison between our proposed method and ten leading SOTA, as evaluated on the I-Haze dataset. Zoom in for better view.



Fig. 9. The results of a qualitative comparison between our proposed method and ten leading SOTA, as evaluated on the O-Haze dataset. Zoom in for better view.

The comparative analysis shows superiority of our method over SOTA methods, but real-world and synthetic hazy images differ significantly from one another. Therefore, more verification is required to confirm the progress in real-world image recognition. We conduct tests on the real-world dataset collected from internet in order to achieve this goal. Fig. 11 illustrate that we are only able to perform a qualitative comparison since the ground truth images that correlate to the hazy images are not accessible in this dataset.

To assess the efficacy of our network, we conducted experiments using real-world images, comparing the outcomes with SOTA networks, as shown in Fig. 11. Image b exhibits clear residual images, yet retrieving distant objects poses a challenge. Images c and f accentuate darkening

in the background, particularly noticeable among closely positioned elements. Conversely, the results from images d, e, and g yield indistinct residuals and fail to retain some background details, exhibiting blocky artifacts. Image h illustrates a loss of detail and insufficient dehazing, possibly due to complex scenery with dense haze in specific regions. Image i produces a blurred output with scattered light, particularly evident in the cropped section, struggling to eliminate distant haze in the mid-region. Meanwhile, image j suffers from color contrast and hue issues in the dehazed output. Our proposed dehazing method yields visually realistic results with enhanced contrast, while maintaining the integrity of fine-grained texture and color information.

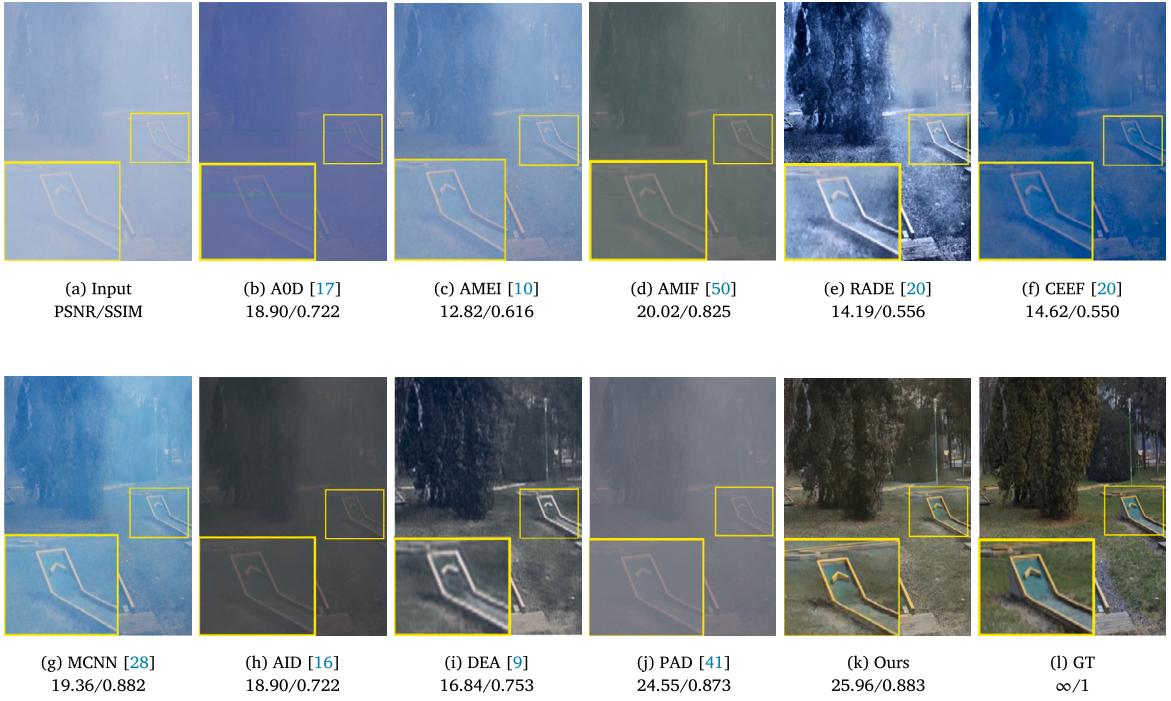


Fig. 10. The results of a qualitative comparison between our proposed method and ten leading SOTA, as evaluated on the Dense-Haze dataset [3]. Zoom in for better view.



Fig. 11. The results of a qualitative comparison between our proposed method and ten leading SOTA methods, as evaluated on the Real-Haze dataset.

4.6. Hardware implementation

The aim of the hardware implementation is to make dehazed images accessible to all systems without requiring high-performance CPUs and to reduce the power requirements of surveillance systems. Moreover, it could be integrated into portable devices such as smart cameras and smartphones. For practical purposes, real-time image dehazing is necessary, and a significant delay would render the system inefficient. Hence, this study introduces an effective image dehazing approach.

To showcase the practical efficacy of our network, we deployed the proposed TSMD-Net on a low-end NVIDIA Jetson Nano board. The Nvidia Jetson Nano board is designed for developing embedded systems with high processing power requirements, particularly in the fields of Computer Vision, deep learning, machine learning, and image/video

processing applications. It boasts a quad-core ARM A57 CPU operating at 1.43 GHz, coupled with 4 GB of LPDDR4 RAM, 128 CUDA cores on a Maxwell GPU, and a processing power of 472 GFLOPS. With its low power consumption of less than 5 watts, integrated GPU cores, and relatively lower cost compared to other boards, the Jetson Nano board has become a favored choice for implementing real-time low-level computer vision algorithms. Fig. 12 presents the experimental configuration deployed on the Jetson Nano board.

4.6.1. Real time implementation on Jetson Nano board

Real-time image dehazing has become increasingly crucial for a variety of practical applications. In response to this need, we have introduced a novel approach, the TSMD-Net, designed to offer both effectiveness and efficiency in dehazing processes. A significant aspect of our

Table 6

Comparison of time complexity on Tesla V100 GPU and CPU, in terms of the number of parameters, multiply and accumulate (MAC) operations, and runtime measured on 512×512 input images from the Dense Haze test dataset [3].

Method	Parameters (M)	MAC (G)	Runtime (sec) on GPU	Runtime (sec) on CPU
AOD-Net [17]	0.56	62	0.983	10.813
MSCNN [28]	2.63	208	0.427	3.843
AID-Net [16]	49.3	309	0.756	8.316
DEA-Net [9]	4.45	1154	0.864	10.368
PAD-Net [41]	5.8	69	0.208	2.288
Ours	16.3	57	0.146	1.898



Fig. 12. Practical setup employed for implementing the TSMD-Net on the Jetson Nano hardware.

investigation involved assessing the performance of our network on edge computing devices, specifically the Jetson Nano. By employing the same Python code utilized on our workstation, we conducted comprehensive evaluations of our network's capabilities on edge devices characterized by limited computational resources. Through rigorous testing using real-world hazy images, we scrutinized the functionality of our network. The findings, as depicted in Fig. 13, illustrate the outcomes of our dehazing process on real-haze images. Notably, these results underscore the reliability of our network in producing haze-free images while operating seamlessly on the Jetson Nano platform. This validation highlights the practical feasibility of our approach in addressing the challenges posed by haze in real-time scenarios, particularly in resource-constrained environments.

4.6.2. Overhead

In Table 6, we present a comparison of our method with SOTA networks, detailing the number of parameters required, the multiply and accumulate (MAC) operations performed, and the average runtime on the Dense Haze [3] test dataset using 512×512 images for single-image dehazing. Our network demonstrates increase in the number of parameters due to the replacement of some non-linear activation functions with simple multiplication operations. This design choice results in the lowest MAC operations among the compared methods, which optimizes hardware selection, reduces computational load, and minimizes energy consumption. Additionally, our method achieves exceptionally low runtimes on both GPU (0.146 seconds) and CPU (1.898 seconds), ensuring superior efficiency and performance.

4.7. Ablation study

In our evaluation, we introduced specific components into our TSMD-Net architecture and assessed their impact on performance compared to the basic U-Net architecture, using the RESIDE dataset.

a. Effect of basic components:

The MWSA module was found to significantly enhance performance, increasing PSNR from 14.37 dB to 21.56 dB. MWSA facilitates the capture of global feature data by providing large receptive fields. Additionally, the inclusion of the SACM yielded a moderate improvement in evaluation metrics, with PSNR increasing by approximately 1 dB and SSIM by around 0.1 score. Furthermore, integrating the Feature Fusion (FF) module across all levels of the encoder-decoder structure resulted in substantial improvements in both PSNR (25.31 dB) and SSIM (0.924) illustrated in Table 7. Which underscoring its significant contribution to enhancing dehazing performance within our network.

b. Effect of loss function:

In our evaluation, we examined the performance of our network using different loss functions, namely Charbonnier loss, edge loss, and hybrid loss (combining Charbonnier and edge losses), as outlined in Ta-

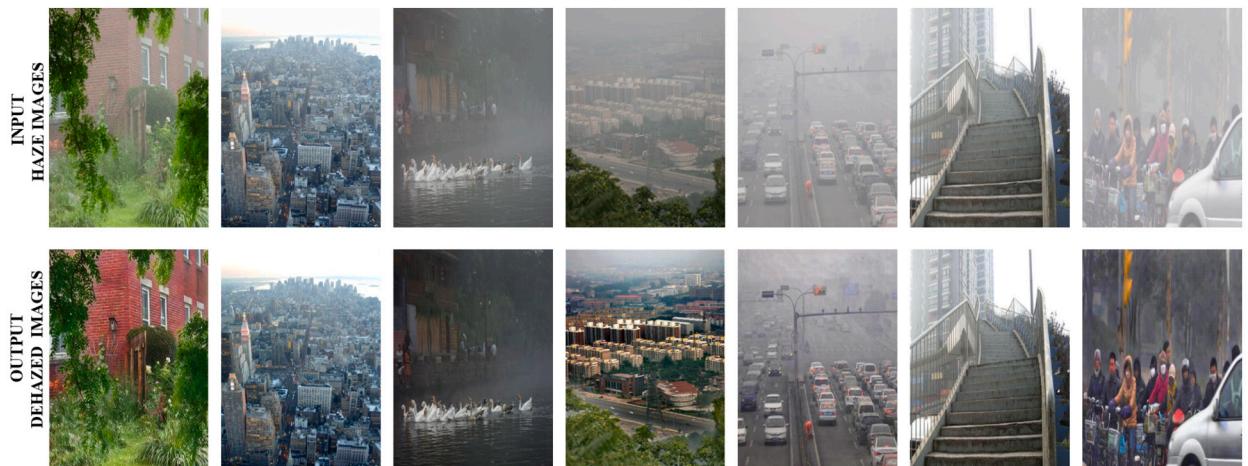


Fig. 13. Real-time test results for image dehazing on Jetson Nano board.

Table 7

Impact of individual components used in TSMD-Net.

Component	PSNR	SSIM
U-Net	14.37	0.765
U-Net + MWSA	21.56	0.781
U-Net + MWSA + SACM	22.48	0.895
U-Net + MWSA + SACM + FF	25.31	0.924

Table 8

Impact of loss functions.

Loss function	PSNR	SSIM
Charbonnier	24.96	0.895
Edge	25.27	0.902
Hybrid	25.31	0.924

ble 8. Through an ablation study aimed at identifying the most suitable loss function for training our network, we ultimately selected the hybrid loss function. This decision was informed by its ability to produce superior PSNR and SSIM scores compared to other alternatives. The adoption of the hybrid loss function further validates its efficacy in optimizing the performance of our network, underscoring its importance in achieving superior image restoration outcomes.

5. Conclusion

In this paper, we introduced a two-stage mixed dehazing network to capture multi-scale feature-map information effectively, that strikes a compromise between computational efficiency and picture restoration quality. Our main goal was to maximize the interaction between contextual feature-map data and spatial features while restoring images. A Feature Fusion Module was integrated to enable smooth information transmission across the hierarchical tiers of the encoder-decoder architecture. This module successfully combines data from neighboring layers to produce a complete representation at the initial image alteration stage. In order to enhance CNNs' receptive fields and obtain more comprehensive global feature-map data, we unveiled the MWSA. Our model was optimized for computational performance by replacing unnecessary non-linear activation functions with basic multiplicative operations, which led to a lightweight design. Comprehensive experimental assessments validated the enhanced effectiveness of our suggested network on many benchmark datasets, surpassing current SOTA networks in terms of computational efficiency and runtime.

Additionally, we tested our TSMD-Net's real-time resilience by deploying it on an Nvidia Jetson Nano platform, where it produced noise-free images in an incredibly short processing time of 2.56 seconds. The thorough experimental findings demonstrate our proposed network's effectiveness, generalizability, and real-time application on both synthetic and real haze datasets.

CRediT authorship contribution statement

Nuthi Raju: Conceptualization, Methodology, Software, Writing – Original draft preparation.

Kankanala Srinivas: Data curation, Visualization, Investigation, Supervision, Validation, Writing – Reviewing and Editing.

Declaration of competing interest

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.

Data availability

Data will be made available on request.

References

- [1] C. Ancuti, C.O. Ancuti, R. Timofte, C.D. Vleeschouwer, I-haze: a dehazing benchmark with real hazy and haze-free indoor images, in: Proc. Int. Conf. Adv. Concepts Intell. Vis. Syst., Springer, 2018, pp. 620–631.
- [2] C.O. Ancuti, C. Ancuti, R. Timofte, C. De Vleeschouwer, O-haze: a dehazing benchmark with real hazy and haze-free outdoor images, in: Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW), 2018, pp. 754–762.
- [3] C.O. Ancuti, C. Ancuti, R. Timofte, C. De Vleeschouwer, Dense-haze: a benchmark for image dehazing with dense-haze and haze-free images, in: ICIP, 2019, pp. 1014–1018.
- [4] J.T. Barron, A general and adaptive robust loss function, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 4331–4339.
- [5] B. Cai, et al., Dehazenet: an end-to-end system for single image haze removal, IEEE Trans. Image Process. 25 (2016) 5187–5198.
- [6] D. Chen, M. He, Q. Fan, J. Liao, L. Zhang, D. Hou, L. Yuan, G. Hua, Gated context aggregation network for image dehazing and deraining, in: 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), IEEE, 2019, pp. 1375–1383.
- [7] L. Chen, X. Chu, X. Zhang, J. Sun, Simple baselines for image restoration, in: European Conference on Computer Vision, Springer, 2022, pp. 17–33.
- [8] W.T. Chen, et al., Pmhld: patch map-based hybrid learning dehazenet for single image haze removal, IEEE Trans. Image Process. 29 (2020) 6773–6788.
- [9] Z. Chen, Z. He, Z.M. Lu, Dea-net: single image dehazing based on detail-enhanced convolution and content-guided attention, IEEE Trans. Image Process. (2024).
- [10] A. Galdran, Image dehazing by artificial multiple-exposure image fusion, Signal Process. 149 (2018) 135–147.
- [11] H. Gao, D. Tang, Mixed hierarchy network for image restoration, arXiv preprint, arXiv:2302.09554, 2023.
- [12] K. He, J. Sun, X. Tang, Single image haze removal using dark channel prior, IEEE Trans. Pattern Anal. Mach. Intell. 33 (2010) 2341–2353.
- [13] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
- [14] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7132–7141.
- [15] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, Adv. Neural Inf. Process. Syst. 25 (2012).
- [16] A. Kulkarni, S. Murala, Aerial image dehazing with attentive deformable transformers, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2023, pp. 6305–6314.
- [17] B. Li, X. Peng, Z. Wang, J. Xu, D. Feng, Aod-net: all-in-one dehazing network, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 4770–4778.
- [18] B. Li, et al., Benchmarking single-image dehazing and beyond, IEEE Trans. Image Process. 28 (2019) 492–505.
- [19] X. Li, Z. Hua, J. Li, Attention-based adaptive feature selection for multi-stage image dehazing, Vis. Comput. 39 (2023) 663–678.
- [20] Z. Li, X. Zheng, B. Bhanu, S. Long, Q. Zhang, Z. Huang, Fast region-adaptive defogging and enhancement for outdoor images containing sky, in: 2020 25th International Conference on Pattern Recognition (ICPR), IEEE, 2021, pp. 8267–8274.
- [21] C. Lin, X. Rong, X. Yu, Msaff-net: multiscale attention feature fusion networks for single image dehazing and beyond, IEEE Trans. Multimed. (2022).
- [22] X. Liu, H. Li, C. Zhu, Joint contrast enhancement and exposure fusion for real-world image dehazing, IEEE Trans. Multimed. 24 (2021) 3934–3946.
- [23] E. McCartney, Optics of the Atmosphere: Scattering by Molecules and Particles, Wiley, New York, 1976.
- [24] G. Meng, Y. Wang, J. Duan, S. Xiang, C. Pan, Efficient image dehazing with boundary constraint and contextual regularization, in: Proceedings of the IEEE International Conference on Computer Vision, 2013, pp. 617–624.
- [25] S. Narasimhan, S. Nayar, Proceedings ieee conference on computer vision and pattern recognition, cvpr 2000 (cat. no.pr00662), in: IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2000, pp. 598–605.
- [26] X. Qin, Z. Wang, Y. Bai, X. Xie, H. Jia, Ffa-net: feature fusion attention network for single image dehazing, in: Proceedings of the AAAI Conference on Artificial Intelligence, 2020, pp. 11908–11915.
- [27] D. Ren, W. Zuo, Q. Hu, P. Zhu, D. Meng, Progressive image deraining networks: a better and simpler baseline, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 3937–3946.
- [28] W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, M.H. Yang, Single image dehazing via multi-scale convolutional neural networks, in: Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part II 14, Springer, 2016, pp. 154–169.
- [29] O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation, in: Medical Image Computing and Computer-Assisted

- Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III 18, Springer, 2015, pp. 234–241.
- [30] Y. Shao, L. Li, W. Ren, C. Gao, N. Sang, Domain adaptation for image dehazing, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 2808–2817.
- [31] Y. Song, et al., Vision transformers for single image dehazing, *IEEE Trans. Image Process.* 32 (2023) 1927–1941.
- [32] K. Srinivas, A.K. Bhandari, Context-based novel histogram bin stretching algorithm for automatic contrast enhancement, *ACM Trans. Multimed. Comput. Commun. Appl.* 19 (2023), <https://doi.org/10.1145/3597303>.
- [33] K. Srinivas, A.K. Bhandari, P.K. Kumar, A context-based image contrast enhancement using energy equalization with clipping limit, *IEEE Trans. Image Process.* 30 (2021) 5391–5401.
- [34] K. Tang, J. Yang, J. Wang, Investigating haze-relevant features in a learning framework for image dehazing, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 2995–3000.
- [35] C. Wang, et al., Deep residual haze network for image dehazing and deraining, *IEEE Access* 8 (2020) 9488–9500.
- [36] Z. Wang, A. Bovik, H. Sheikh, E. Simoncelli, Image quality assessment: from error visibility to structural similarity, *IEEE Trans. Image Process.* 13 (2004) 600–612, <https://doi.org/10.1109/TIP.2003.819861>.
- [37] Z. Wang, J. Chen, S.C.H. Hoi, Deep learning for image super-resolution: a survey, *IEEE Trans. Pattern Anal. Mach. Intell.* 43 (2019) 3365–3387.
- [38] S. Woo, J. Park, J.Y. Lee, I.S. Kweon, Cbam: convolutional block attention module, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 3–19.
- [39] W. Xue, L. Zhang, X. Mou, A.C. Bovik, Gradient magnitude similarity deviation: a highly efficient perceptual image quality index, *IEEE Trans. Image Process.* 23 (2013) 684–695.
- [40] C.H. Yeh, et al., Haze effect removal from image via haze density estimation in optical model, *Opt. Express* 21 (2013) 27127–27141.
- [41] W. Yi, et al., Priors-assisted dehazing network with attention supervision and detail preservation, *Neural Netw.* 106165 (2024).
- [42] S.W. Zamir, A. Arora, S. Khan, M. Hayat, F.S. Khan, M.H. Yang, L. Shao, Multi-stage progressive image restoration, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 14821–14831.
- [43] H. Zhang, V.M. Patel, Densely connected pyramid dehazing network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 3194–3203.
- [44] L. Zhang, Y. Shen, H. Li, Vsi: a visual saliency-induced index for perceptual image quality assessment, *IEEE Trans. Image Process.* 23 (2014) 4270–4281.
- [45] L. Zhang, L. Zhang, X. Mou, D. Zhang, Fsim: a feature similarity index for image quality assessment, *IEEE Trans. Image Process.* 20 (2011) 2378–2386.
- [46] X. Zhang, et al., Pyramid channel-based feature attention network for image dehazing, *Comput. Vis. Image Underst.* 197 (2020) 103003.
- [47] Y. Zhang, K. Li, K. Li, B. Zhong, Y. Fu, Residual non-local attention networks for image restoration, arXiv preprint, arXiv:1903.10082, 2019.
- [48] Y. Zhang, et al., Image super-resolution using very deep residual channel attention networks, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 286–301.
- [49] Q. Zhu, J. Mai, L. Shao, Single image dehazing using color attenuation prior, in: BMVC, Citeseer, 2014, pp. 1–10.
- [50] Z. Zhu, H. Wei, G. Hu, Y. Li, G. Qi, N. Mazur, A novel fast single image dehazing algorithm based on artificial multiexposure image fusion, *IEEE Trans. Instrum. Meas.* 70 (2020) 1–23.

Kankana Srinivas is an Assistant Professor at VIT-AP University, Amaravathi, India. He completed his B.Tech with a specialization in Electronics and Communication Engineering in 2009 from JNTU Hyderabad. In 2013, he attained his M.Tech in VLSI System Design from VNR VJIET, Hyderabad. Later, in 2020, he was awarded a Ph.D. from the National Institute of Technology Patna, specializing in Digital Image Processing. His research areas span various domains, including hardware implementations, digital image processing, soft computing, and array signal processing. His expertise and contributions in these fields have been recognized through his academic pursuits and research endeavors.

Nuthi Raju received the B.Tech. Degree from JNTUH, Telangana, India, and the M.Tech. Degree from JNTUH, Telangana, India. He is currently pursuing a Ph.D. degree in Electronics and Communication Engineering at VIT-AP University, Andhra Pradesh, India. His research interests include machine learning, deep learning, and image processing.