CHC-UNet: Cascaded Heterogeneous Convolution UNet Framework for Adaptive Image Dehazing

CH. Mohan Sai Kumar* and R. S. Valarmathi
Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, India
Email: mohansaikumarchadalavada@gmail.com* (CH.M.S.K.); atrmathy@gmail.com (R.S.V.)

***Abstract*—Haze is a phenomenon of degrading the visual quality of outdoor images by wet aerosols and particles suspended in the atmosphere under poor weather conditions. The incident light on to the object will be scattered and causes the distortion of color, contrast reduction and generate the halo artifacts. In this article, a Cascaded Heterogeneous Convolution UNet (CHC-UNet) with re-defined Squeeze-and-Excitation (SE) mechanism framework is designed to reduce the adverse hazes' effect in the images. The Heterogeneous Convolution kernels with size 3×3 and 1×1 are used in each block of UNet to extract the features at various image resolutions and to overcome excessive computation load. The adaptive residual connections preserve the original image features, prevent from overfitting problem and make the dehazing process more thorough. The SE Algorithm is re-defined to better represent the channel-wise image features with Local Feature Fusion (LFF) module. The qualitative and quantitative findings of the ablation experiments show that the suggested framework outperforms state-of-the-art methodologies in the examined literature using the benchmark I-Haze, O-Haze, and D-Hazy datasets. The proposed framework has an average improvement from 17.95 dB to 27.52 dB in Peak Signal-to-Noise Ratio (PSNR), from 0.56 to 0.87 in Structural SIMilarity (SSIM) index on retrieved images, from 4.98 to 3.12 in Naturalness Image Quality Evaluator (NIQE). Furthermore, the framework has varying degrees of improvement in the subjective visual effects of the dehazed output, and it also has the added advantage of optimum running time in comparison with the existing dehazing methods.**

***Index Terms*—Aerosols, halo artifacts, squeeze and excitation, heterogeneous convolution, computation load, residual connection, overfitting**

I. Introduction

The real-world image processing algorithms and automated systems often require high contrast images with good scene quality to increase the accuracy of computer vision-related tasks. In general, the natural imaging systems construct low contrast image scenes in adverse weather conditions due to light absorption and scattering by the tiny and floating particles present in the atmosphere [1] is shown in Fig. 1. This phenomenon of degrading the scene quality causes the fog and haze formation in the scene acquisition that leads to degrading effects and performance reduction in free space detection, low light image enhancement [2], scene classification, target object detection [3, 4], scene understanding [5], semantic segmentation, video surveillance and other advanced machine vision tasks. Hence, it becomes significant in research to design and develop defogging and dehazing algorithms for scene restoration. The image dehazing is the worthwhile procedure of restoring the scene from adversely effected image by preserving the original color and contrast. Dehazing, however, is a challenging and non-trivial task because it requires hard collection of haze free ground truth images and non-availability of high performance computing hardware for processing with optimum running time.

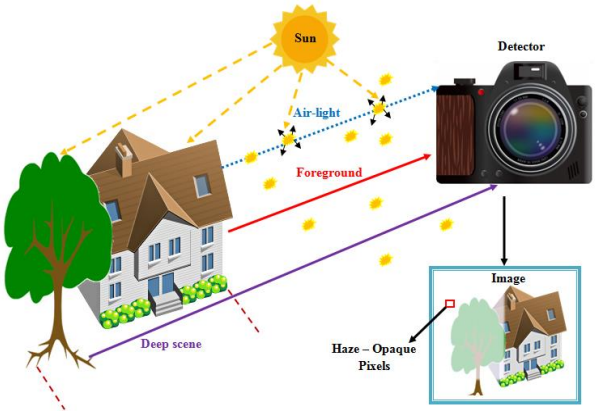

Fig. 1. Haze formation.

The multiple images of the same landscape taken in various climatic conditions can improve the dehazing performance, which would be practically impossible. Therefore, the single image dehazing methods [6]-[10] is designed using a priori of physical scattering model [11] defined by

$$I_{(p,q)} = J_{(p,q)} T_{(p,q)} + A\left(1 - T_{(p,q)}\right) \quad (1)$$

where I is the hazy input image captured with image sensor and J is the corresponding haze free clear image at pixel location (p, q) with the global atmospheric light intensity A, the scene transmission map T is defined by

Manuscript received April 8, 2023; revised May 4, 2023; accepted May 9, 2023.
*Corresponding author: CH. Mohan Sai Kumar

$$T_{(p,q)} = e^{-\beta d(p,q)} \tag{2}$$

where β is the light scattering parameter, the scene depth d defined as the distance from the scene to image capturing device. The concentration of haze depends on the scene depth which is practically unavailable. These prior-based dehazing techniques include distortion and undesired artefacts since they rely on effective model parameters estimation.

The robustness of image dehazing increases with the accurate estimation of A and T, described by

$$\begin{aligned} J_{(p,q)} &= \frac{I_{(p,q)} - A\left(1 - T_{(p,q)}\right)}{T_{(p,q)}} \\ &= V_{(p,q)} I_{(p,q)} - V_{(p,q)} + b \\ &= V_{(p,q)} \left(I_{(p,q)} - 1\right) + b \end{aligned} \tag{3}$$

$$V_{(p,q)} = \frac{\frac{1}{T_{(p,q)}}\left(I_{(p,q)} - A\right) + (A - b)}{I_{(p,q)} - 1} \tag{4}$$

where both A and T are unified into a single variable V, b is the deviation with a default constant value of 1.

With the advancements in the recent data driven models, the learning based algorithms employs the datasets with image pairs for the scene transmission map [12, 13] estimation or for the joint estimation of global atmospheric light intensity and scene transmission map [14, 15]. The current methods do not learn the inherent relationship between atmospheric light and transmission map, limiting the performance of dehazing.

Zhang [14] proposed to utilize the joint estimation of A and T as a single variable and thus optimizes the network. However, the model-free methods do not rely on the traditional atmospheric model, but, performs image dehazing using end-to-end deep convolutional neural networks. Recently, an end-to-end dehazing networks for scene restoration is designed using encoder-decoder framework [16, 17]. This results in degeneration of spatial information with the downsampling operations and missing concatenation of information from non-adjacent layers.

In order to overcome the missing spatial information, a re-defined Squeeze-and-Excitation (SE) residual learning mechanism with Local Feature Fusion (LFF) module to recalibrate the channel-wise features is designed.

The predominant contributions made in this research work are summarized as follows:

- Cascaded Heterogeneous Convolution Kernel based UNet (CHC-UNet) Framework: An UNet framework with heterogeneous convolution kernels in each block of encoder and decoder stages with LFF enabled SE mechanism in skip connections and bottleneck co-evolutionary atrous convolution block is designed.
- ShuffleNet model for Transmission map estimation: A light weight network model with cross-channel information communication and optimum number of channels without increasing the computation cost to estimate the transmission map of the hazy image is proposed.
- Re-defined Gating SE mechanism: A LFF module enabled attention based SE mechanism for recalibration of features and better channel interdependencies is recommended.
- Extensive experimentation and exhaustive ablation analysis on the quantitative results are performed to illustrate the effectiveness of the proposed framework.

The structure of the article is organized as follows. The recent research works related to the image dehazing are briefly reviewed in Section II. In section III, the image dehazing framework of the proposed UNet using Heterogeneous convolution kernel and re-defined SE mechanism is presented. Section IV includes the information of experimental setup, ablation study of qualitative and quantitative results. Section V provides the conclusion and future work of the proposed framework.

II. Related Works

The concept of image dehazing aims to reduce the effect of haze's adverse impacts on the quality of image and boosts the scene visibility suitable for image processing, analysis and understanding. It has received a great deal of interest in the research field of many computer vision applications that employ natural indoor and outdoor atmospheric imagery. This section provides the related research work of image dehazing as discussed below.

1) Image Enhancement Based Dehazing Methods

These methods restore the image features, emphasize the target scenario, and necessitate no prior knowledge of an image's physical degradation. Ghani [18], He [19], Li [20] and Liu [21] have employed the histogram modification, weighted histogram, fusion based contrast enhancement and dynamic image enhancement techniques respectively for dehazing. The persistent oversaturation problem prevails in the dehazed image along with texture preservation issue. Additionally, it improves the contrast level to a certain extent and it is not appropriate for large non homogeneous haze gradient effected images.

2) Priori Based Dehazing Methods

These methods rely on estimating the depth map and atmospheric light using the haze's prior information or atmospheric scattering model. He [11] proposed a dark channel prior model for calculating the air light and scene depth using local mask. The large size of the local mask causes oversaturation problem whereas small mask size introduces halo and gradient reversal artifacts. Fattal [22] and Tan [23] have proposed to estimate the scene reflectance by analyzing image reflectivity and local contrast maximization algorithm using Markov random field model for maximizing the local contrast to achieve effect dehazing. Zhu [24] used a color attenuation prior model for transmission and Tang [25] proposed to learn the correlation between transmission map and features using random forest regression. Lai [26] designed an intuitive method to estimate the optimal transmittance map using atmospheric scattering for image dehazing. Additionally, the atmospheric scattering prior based

methods suffer from white sky-region, degradation of edges, color distortion, and artifacts etc.

3) Learning Based Dehazing Methods

With the introduction and rapid advancements in neural network concept, the design of novel Convolution Neural Networks (CNN) to restore the image scene from haze based on deep learning have been continuously emerged. Instead of using a traditional technique and hand-crafted priors, this type of approach optimizes deep neural networks on large-scale data sets to discover high-level transmission patterns. To determine the transmission map, Cai [27] proposed a DehazeNet, Ren [28] developed a multi-scale deep neural network, Li [29] designed AOD-Net, Ren [30] suggested a Gated Fusion Network (GFN), Wang [31] designed a haze concentration adaptive network with feature extractor, enhancement and attention module for different haze conditions. Zhu [32] proposed Generative Adversarial Network (GAN) and achieve promising results in image dehazing. Zhao [33] proposed a prior guided generator-discriminator network to extract the features and enhance by the attention aggregation block. The method combines the prior based knowledge and CNN to balance between the artifacts caused by each other.

4) Data-driven Based Non-physical Model Dehazing Methods

Data-driven deep learning algorithms have been used to learn features, which have overcome the drawbacks of hand-made features. Dong [34] designed a dense feature fusion module using UNet framework based on the principles of boosting decoder and back projection feedback. Shao [35] designed an image domain transformation framework for dahazing. To increase the contrast without introducing dehazing artifacts, Shin [36] proposed the use of triple convolutional networks, which combine dehazing, enhancement, and concatenating sub-networks. The computational complexity varies with the combined and separated sub-networks. Huang [37] suggested the self-filtering in dehazing network to increase the learned feature representation abilities and improve the performance in restoring the scene content and details.

Xiao [38] mimic the frequency content of images using Taylor's infinite approximation theorem with a Laplacian pyramid network. Tran [39] proposed to utilize the dark channel priors, spatial pyramid pooling module as an encoding–decoding network to determine the transmission maps. Alenezi [40] suggested utilizing the features and color channels of RGB images in the underwater image dehazing network to improve overall usability. Yang [41] designed a self-enhanced dehazing framework with unpaired hazy and clear images in the training network to restore the depth information, scattering coefficient and haze free images.

Song [42] proposed the improved structure of Swin Transformer [43] and a gUNet based on UNet compact network for image dehazing. Song [44] proposed vision transformers for haze removal using swin transformer and UNet and conducted experimentation on the synthetic and remote sensing hazy dataset. The computational complexity involved in using shifted window partitioning is very high with reflection padding and reduces the method's operating efficiency.

Hence, to solve the problem of capturing incomplete global feature information by the CNN based image dehazing algorithm, a heterogeneous convolution kernel based UNet framework is proposed in this article. The UNet encoder and decoder sections consist of a convolution branches with Heterogeneous convolution kernel to improve the computational cost of the network, and a re-defined Squeeze-and-Excitation (SE) mechanism with Local Feature Fusion (LFF) module to increase the image representation capability of the network. Finally, the dehazed image free from distortion and other artifacts is obtained with rich contextual features as in the haze free image.

III. Methodology

A novel end-to-end UNet framework is designed to improve the image dehazing effect using the strong image representation capability of deep learning. The framework of the proposed Cascaded Heterogeneous Convolution kernel based UNet (CHC-UNet) framework with refined Squeeze-and-Excitation (SE) mechanism and bottleneck co-evolutionary atrous convolution block is shown in Fig. 2.

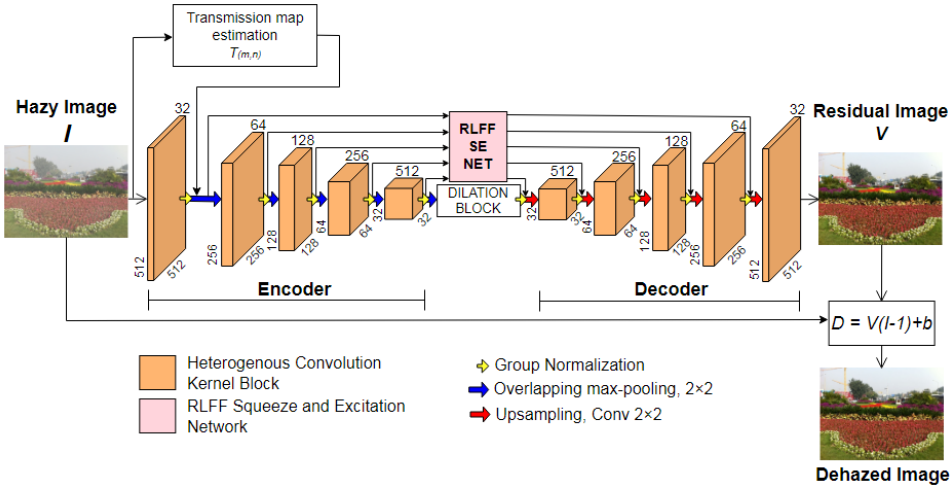

Fig. 2. Cascaded heterogeneous convolution kernel based UNet framework.

The encoder and decoder section serves as the base modules for UNet framework. The encoder module consists of five cascaded convolution blocks of varying depth to extract the spatial features from the input image via heterogeneous kernel. After each convolution block, the feature map size is halved while the number of feature maps is doubled.

The bottleneck atrous convolution unit captures both the local and global contextual characteristics of the input image from the heterogeneous convolution blocks. The decoder module receives the output of the atrous convolution and uses deconvolution to reconstruct the deep feature maps from the encoded features. The decoder module has the same structure as that of the encoder module, but differs in the sampling operation. The Low-resolution Edge Separation (LES) scheme of upsampling is used in the decoder module to remove the checkerboard artifacts if any introduced in the dehazed image.

The Local Feature Fusion (LFF) module enabled SE network captures and weights the attentive features in the skip connections. The skip connections help in upsampling the image back to the original resolution. Additionally the Group Normalization (GN) is used instead of Batch Normalization (BN) and Funnel Rectified Linear Unit (FReLU) non-linear activation function is used instead of Rectified Linear Unit (ReLU) activation function. The heterogeneous convolution blocks, the interaction information among different resolution scales and efficient reuse of features are the key characteristics of proposed framework.

A. Transmission map Estimation Model

Designing Convolution Neural Network (CNN) architecture with deep and large number of layers is the recent research trend in many depth scene recognition and other computer vision applications. The transmission map estimation network uses the shufflenet architecture which is most computationally efficient model for small neural networks is shown in Fig. 3. This architecture is accurate within the given complexity constraint and greatly reduces the computation cost. In its basic model, the dense 1×1 convolutions of neural networks are replaced with pointwise group convolutions to improve the computational efficiency and number of feature channel connections. The flow of cross-group information between the various group convolution layers is now made possible by the newly proposed differentiable channel shuffle operation, which strengthens the network structure.

Both the pointwise group convolutions and bottleneck-like structures increases the number of feature channels without increasing the computational load. But the excessive group convolutions will greatly affect the degree of parallelism in the network which leads to degradation of individual convolution filter and increase in memory access cost M during the training. Hence neither the dense 1×1 convolutions nor the large group convolutions can be used to achieve the model accuracy, efficiency and maintain the large number of equal wide channels. The relation between memory access cost M and computational cost C for 1×1 group convolution is expressed by

$$M = wh\left(C_1 + C_2\right) + \frac{C_1 C_2}{g} = whC_1 + \frac{Cg}{C_1} + \frac{C}{wh} \tag{5}$$

where w and h represents the feature map spatial size, C_1 and C_2 are the number of feature channels at the input and output, g represents the group number. The memory access cost M reaches the optimum value with the equal number of input and output channels.

A channel split operator is used to split the input feature channels into two branches. There are three 1×1 pointwise convolutions and two 3×3 depthwise separable convolutions with the equal number of input and output feature channels for each convolution. The pointwise 1×1 convolutions are used for feature pooling and dimensionality reduction. In order to keep the channel size consistent with the shortcut connection, a second 1×1 pointwise convolution is used in the residual connection.

The 3×3 depthwise separable convolution is performed individually with a single filter on separate input channels with a stride factor of 2, combining the input channels linearly at the output. As a result, the depthwise separable convolution will effectively optimize the model's size and parameterization. The depthwise separable convolution $G_{fm}(v,u,i)$ and pointwise convolution $O(v,u,n)$ can be mathematically expressed in (6) and (7) respectively.

$$G_{fm}\left(v,u,i\right) = \sum_{x=1}^{k}\sum_{y=1}^{k} K(x,y,i) I_{fm}(v+x-1, u+y, i) \tag{6}$$

$$O\left(v,u,n\right) = \sum_{m=1}^{C_1} G_{fm}\left(v,u,i\right) P(m,n) \tag{7}$$

where G_{fm} and I_{fm} are the output and input feature maps, respectively, and K is the depthwise separable convolution kernel size with $K \times K \times C_1$. Using (7), the pointwise convolution operation reconstructs the new feature map with $P(m,n)$ which is $1 \times 1 \times C_1 \times C_2$ kernel size.

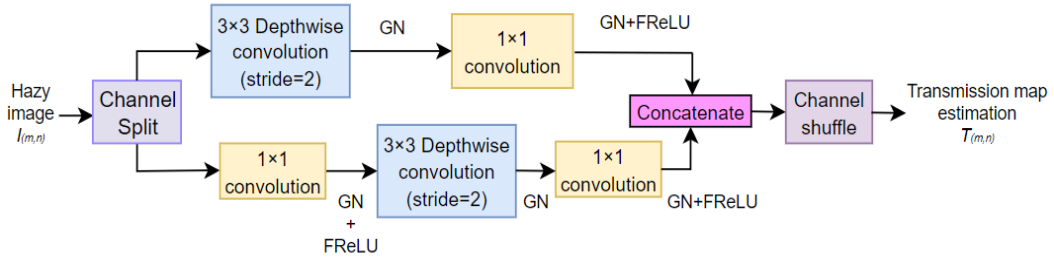

Fig. 3. Transmission map estimation model.

After convolutions, channel-wise concatenation is used in lieu of the element-wise addition operation to make equal number of input and output channels with minimum additional computational cost. The GN activation function [45] and FReLU activation function [46] are used in conjunction with the convolutions, but after 3×3 depthwise separable convolution, however, the activation function is not preferred. With the nature of the channel ShuffleNet, it can be integrated in small-scale network structures for end-to-end training to improve the performance as it encodes enormous feature information by allowing more feature map channels.

B. Heterogeneous Convolution Kernel Block

In this research, the regular convolution block of UNet is replaced with a novel heterogeneous convolution kernel to improve the efficiency and computational cost of convolutions. The homogeneous kernel is made up of all 3×3 or 5×5 convolutions, whereas, the heterogeneous kernel has few kernels of size $K\times K$ and the others are of size 1×1. The Computation Load (CL) of repeated convolution kernel in a particular layer is expressed by:

$$\mathrm{CL} = K \times K \times N \times M \times N_o \times N_o \tag{8}$$

where the $K\times K$ kernel size, spatial width and height $N_o\times N_o$ of the feature map, $N\times M$ input and output channels determine the FLoating point Operations Per second (FLOPs) of a convolutional layer.

Using the heterogeneous kernel, the repeated convolutional kernels in a part P are replaced by the fraction 1/P kernels of size $K\times K$ and $(1-1/P)$ kernels of size 1×1, so as the computational cost is reduced by P times. The computational costs incurred by using $K\times K$ and 1×1 kernels with in a part P is expressed in Eqs. (9), (10) respectively.

$$\mathrm{CL}_k = (K \times K \times N \times N_o \times N_o) \times \frac{M}{P} \tag{9}$$

$$\mathrm{CL}_1 = \left(N \times N_o \times N_o\right) \times \left(M - \frac{M}{P}\right) \tag{10}$$

The Computation Cost (CC) is effectively reduced in comparison with the repeated convolution and is expressed by:

$$\mathrm{CC} = \frac{\mathrm{CL}_k + \mathrm{CL}_1}{\mathrm{CL}} = \frac{1}{P} + \frac{(1-1/P)}{K^2} \tag{11}$$

Different values of P are tested during the experimentation and finally fixed with 2 and 4 to maintain the tradeoff between efficiency and FLOPs with zero latency. The spatial range of the filter reduces with the optimum values of P. The heterogeneous convolution filter with P=4 is shown in Fig. 4. The operation of heterogeneous convolution kernel at each block of UNet is shown in Fig. 5.

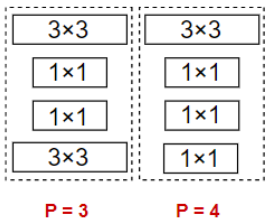

Fig. 4. Heterogeneous kernel convolution filters.

The GN is used after the convolution block as a replacement to batch normalization to overcome the disadvantage of inaccurate estimation of the batch size. The batch normalization usually works fine with larger batch sizes and normalizes the feature maps with the computed mean and variance of a batch. The large batch size is not suitable for hardware specifications and leads to dramatic increase in model error.

In order to normalize the features, the GN computes the absolute mean and variance by dividing the convolution feature channels into smaller groups. The calculation of GN, in general, is independent of batch sizes and outperforms in accuracy over a various batch sizes. It performs the feature computation x by a layer with index i as:

$$\hat{x}_i = \frac{1}{\sigma_i}\left(x_i - \mu_i\right) \tag{12}$$

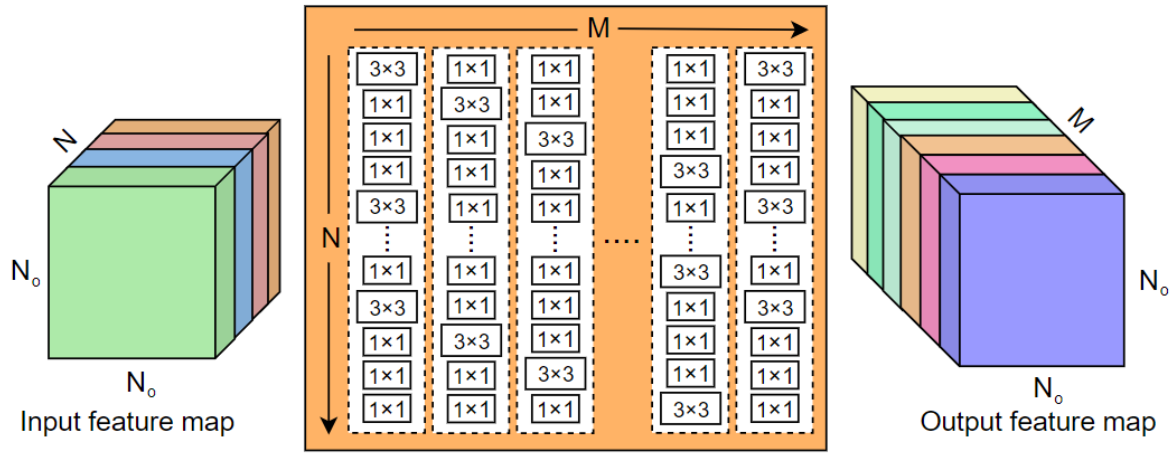

Fig. 5. Heterogeneous convolution kernel block of UNet.

The absolute mean and variance using group normalization is expressed as:

$$\mu_i = \left| \frac{1}{n} \sum_{k \in S_i} x_k \right| \tag{13}$$

$$\sigma_i = \sqrt{\frac{1}{n} \sum_{k \in S_i} (x_k - \mu_i)^2 + \varepsilon} \tag{14}$$

where S_i is collection of image coordinates with size n and ε is constant.

The input 4D tensor for a 2D image is represented by (N, C, H, W) which denotes the batch size, channel number, spatial height and width respectively. The Batch Normalization (BN) computes the mean and variance along (C, H, W) axes, whereas the GN computes along the (H, W) axes which do not require the batch dimensions in the computation.

C. Local Feature Fusion Enabled Squeeze and Excitation Mechanism

In order to enhance the image representation capability of deep learning framework, an efficient LFF enabled SE network structure is designed and achieved satisfactorily results. It is necessary to model the channel interdependencies and perform the feature recalibration for the network to learn to use global information to emphasize the most useful features and selectively suppress less useful ones.

Fig. 6 shows the re-defined SE mechanism with LFF module, overlapping max-pooling layer, excitation operation, sigmoid activation function and rescaling the feature maps to produce the output which can be concatenated to succeeding layers.

The LFF module extracts and aggregates the different level features by exploring the suitable fusion method. It combines the input features and the feature activated by the activation function. The structure of the residual LFF module is shown in Fig. 7.

The feature maps from the convolution blocks at each stage will be combined with the feature maps produced by the activation function and second convolution layer. The operation can be mathematically expressed by:

$$F_{\text{LLF}} = C_{\text{LLF}}\left(\left[F_i F_\sigma F_{\text{conv2}}\right]\right) \tag{15}$$

where C_{LLF} is the function of $N \times N$ convolution layer, F_i denotes the heterogeneous convolution block output and F_σ and F_{conv2} are the feature maps generated by the activation function and second convolution layer. The Table V shows the effect of using LFF module in the SE block.

The SE block is used to explicitly model the channel interdependencies and adaptively recalibrate the channel feature maps. This type of explicit modeling and adaptive feature recalibration is regarded as channel attention mechanism. The attention mechanism is added after the LFF module but before the summation and allows emphasizing the selective informative features among the channels.

The channel attention mechanism's input feature maps have a size of $N \times N \times C$, where C is the channel dimension. To extract the most global spatial information from each feature map, overlapping max-pooling layers are used, which results in the generation of $1 \times 1 \times C$ channel feature descriptors. The squeezing function F_{sq} is expressed as:

$$F_{\text{sq}}(y_c) = \frac{1}{N \times N} \sum_{m=1}^{N} \sum_{n=1}^{N} y_c(m,n) \tag{16}$$

where y_c represents the local descriptors. The transformation output Y in Fig. 6 denotes the collection of local descriptors.

In order to completely capture the channel-wise dependencies, the information gathered from the squeeze operation will then be transferred to the excitation operation. To reduce the channel dimensionality by a value of C/r, the feature descriptors are concatenated to 1×1 kernel size convolutional layer, with the reduction ratio (r) set to 4 during the experimentation. To enhance the nonlinearity of the channel responses, a FReLU nonlinear activation function is used. A subsequent 1×1 kernel size convolutional layer is then used to produce C feature tensors by expanding the channel dimension.

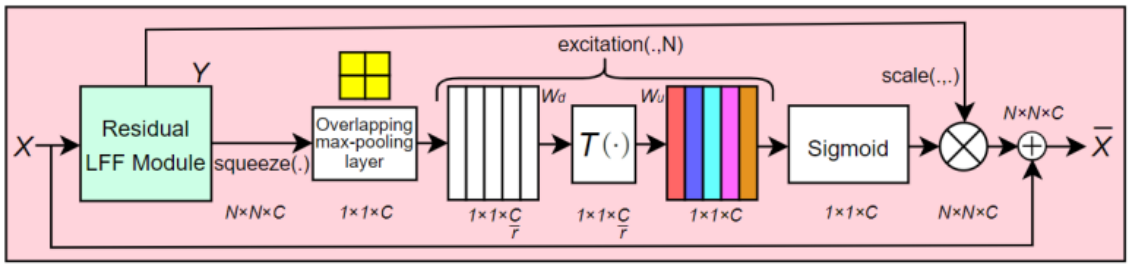

Fig. 6. Re-defined squeeze and excitation mechanism.

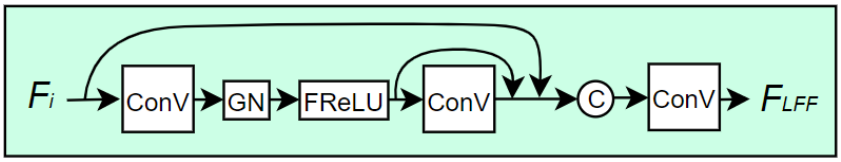

Fig. 7. Residual local feature fusion.

To recalibrate the features, a flexible sigmoid function is adopted. The function learns the non-mutual-exclusive relationship among the multiple channels using a gating mechanism, given as:

$$F_{ex}\left(F_{sq},W\right)=\sigma\left[W_d T\left(W_u F_{sq}\right)\right] \quad (17)$$

where σ refers to sigmoid function, T denotes the FReLU non-linear activation function, $W_d \in \mathbb{R}^{(C/r)C}$ is the dimensionality reduction layer and W_u is the dimensionality increasing layer after being activated by FReLU function with a reduction ratio (r), and F_{sq} is the input squeezed signal.

The feature maps of LFF module is multiplied channel-wise with gating mechanism output to provide the rescaled $N\times N\times C$ feature maps at the attention mechanism output and achieve better recalibrated features.

The conceptually simple Funnel Rectified Linear Unit (FReLU) activation function which is an effective extension of regular Rectified Linear Unit (ReLU) activation function is used in the residual multiscale block of UNet. A spatial contextual feature extractor $T(x)$ is added as a negligible computational overhead to the ReLU function to achieve pixel-wise modeling capacity. The ReLU and FReLU activation functions are expressed in Eqs. (18) and (19) respectively.

$$f\left(x\right)=\max(x,0) \quad (18)$$

$$f\left(x\right)=\max\left(x,T(x)\right) \quad (19)$$

where x is the input to activation function, $T(x) = x^{\omega}\cdot c^{\omega}$ is the funnel condition, x^{ω} denotes the pooling window centered on the input pixel x, c^{ω} denotes the window coefficient. The normal activation function generates spatial condition in the convolution layer and performs separate non-linear transformation. In contrast, the FReLU function generates the spatial dependencies and performs the separate non-linear transformations simultaneously.

D. Co-evolutionary Atrous Convolution

The co-evolutionary usage of atrous convolutions with four different dilation factors that increase by an order of one is a theoretically simple and computation effective. The structures of atrous convolution with different dilation factors and our co-evolutionary atrous convolution block are shown in Fig. 8 and Fig. 9.

The atrous convolution kernel position can be expressed as:

$$z(m)=\sum_{k} y\left(m+dk\right)w[k] \quad (20)$$

where y and z are the input and output of respective layer with a kernel w, m is the kernel position, and d is the dilation factor. The atrous convolution expands the receptive field which is the area where the convolution operation is performed.

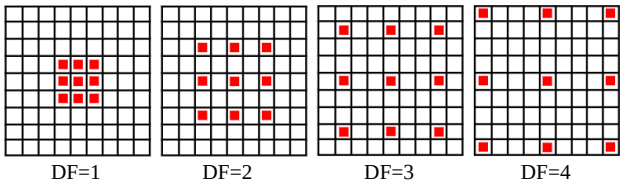

Fig. 8. Atrous Convolution with Dilation Factors (DF=1, 2, 3, 4).

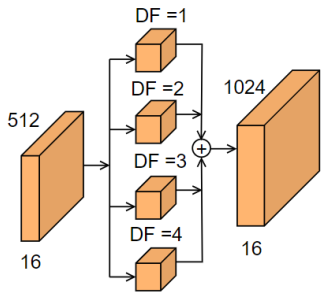

Fig. 9. Co-evolutionary atrous convolution block.

E. Loss Function

To restore the details in the scene, the loss function of the proposed framework depends on Mean Square Error (MSE) and Mean Absolute Error (MAE). Firstly, the model learns MSE function with the residual variation as described by:

$$\begin{aligned} \text{Res} &= \left|I_{\text{Ground truth}} - I\right| \\ L_{\text{MSE}} &= \left\|\text{CHC-UNet}(I) - \text{Res}\right\|^2 \end{aligned} \quad (21)$$

where CHC-UNet(I) is the residual image obtained using CHC-UNet, and Res is the residual variation among the ground truth haze free image $I_{\text{Ground truth}}$ and the hazy input image I.

Second, the Mean Absolute Error (MAE) function is sensitive to sparse features of the hazy input image that can have an apparent dehazing effect, described as:

$$L_{\text{MAE}} = \frac{1}{CWH}\sum_{C}\sum_{W}\sum_{H}\text{CHC-UNet}(I) - I_{\text{Ground truth}} \quad (22)$$

where C, W, H denote the respective channel count, spatial width and height of the training image.

$$L_{\text{Total}} = L_{\text{MSE}} + L_{\text{MAE}} \quad (23)$$

The optimal solution obtained with the sum of MSE function and MAE function as final loss function described by Eq. (23) with continuous testing and calculation.

IV. Experiments and Result Analysis

This section provides the concise information of datasets used for training and testing the proposed framework, experimental setup and image assessment metrics. In addition, the ablation analysis of qualitative and quantitative results and comparative evaluation with the existing methods of dehazing including He [11], Fattal [22], Ren [28], Zhu [32], Song [47], Liu [48], Tangsakul [49], and Kim [50] is provided. We use synthetic images for quantitative analysis and both

synthetic and real-world images for qualitative analysis. The respective parameters were optimized in the available techniques used for comparison in accordance with the corresponding references.

To verify the performance of newly designed architectural framework for image dehazing with transmission map estimation model, a series of experiments using synthetic images from publicly available datasets (D-Hazy, I-Haze and O-Haze) and real-world images with varying image resolutions and haze densities have been conducted.

A. Experiment Setup and Dataset Collection

The network training phase and testing phase of the proposed framework is carried out in the Python platform using a personal computer with Intel® Core i7 Processor at 4.80 GHz and 16 GB RAM. The NVIDIA Graphical Processing Unit (GPU) is used during the experimentation for fast network training.

The D-Hazy dataset is a synthetic dataset with pairs of 1449 synthesized hazy images and 1449 image pairs of ground truth reference images and their associated depth maps generated using a Microsoft Kinect camera. An assumption of uniform haze density and atmospheric light is used to synthesize the hazy images.

Both the I-Haze and O-Haze are hazy image datasets that include corresponding 30 indoor and 45 outdoor hazy/haze free images scenes respectively. In this datasets, the hazy images were acquired under various haze densities that mimic real-world haze circumstances. The lighting parameters of the hazy and haze-free image acquisition are considered to be same with high spatial resolution.

These datasets contain a wide variety of image resolutions, haze densities and scene depths, and are significantly suitable for evaluating the efficiency of different image dehazing models. In addition, 200 sets of real-world hazy and haze free images have been collected, and used during experimentation for qualitative evaluation.

B. Network Training and Testing details

To ensure the convergence, the proposed framework has been trained and validated on Nvidia GPU using Pytorch for 180 epochs and 10 numbers of iterations per epoch. The batch size is fixed to 8 images, and the learning rate is initially set at 0.0002 for the first half of the epochs before decaying linearly to zero for the second half of the epochs, Stochastic Gradient Descent (SGD) algorithm is used for the loss function minimization in Eq. (23). The images are resized to scale of 512×512 during network training. In our proposed architectural framework, 3×3 kernel size, and 16 filters are used. As all the input images are in RGB color space, the filter channels are set to 3.

C. Qualitative Evaluation and Comparison on Synthetic and Real-world Images

For qualitative evaluation of the proposed framework, several indoor haze, dense haze, outdoor haze images from the datasets and real-world hazy images were selected during experimentation and the results obtained are shown in Fig. 10, Fig. 11, Fig. 12 and Fig. 13 respectively. The hazy input is shown in first column and the outputs of He [11], Fattal [22], Ren [28] has over enhancement in the white region of image. The outputs obtained with Zhu [32], Song [47], Liu [48] suffer from color artifacts and poor contextual features. The outputs of Tangsakul [49], and Kim [50] has similar visual quality with the proposed framework but our framework restores the scene with rich contextual features and optimum running time.

D. Quantitative Evaluation and Comparison

The five full-reference quality metrics: Peak-Signal-to-Noise-Ratio (PSNR), Structural SIMilarity (SSIM) [51], Feature SIMilarity extended to color image (FSIMc) [52], Mean Absolute Error (MAE) and Visual Information Fidelity (VIF) [53], and one no-reference quality metrics [54]: Naturalness Image Quality Evaluator (NIQE) are used in the ablation analysis of the proposed framework comparing with existing methods of dehazing. The Table I, Table II and Table III provides the quantitative evaluation of the proposed framework on I-Haze, D-Hazy and O-Haze datasets.

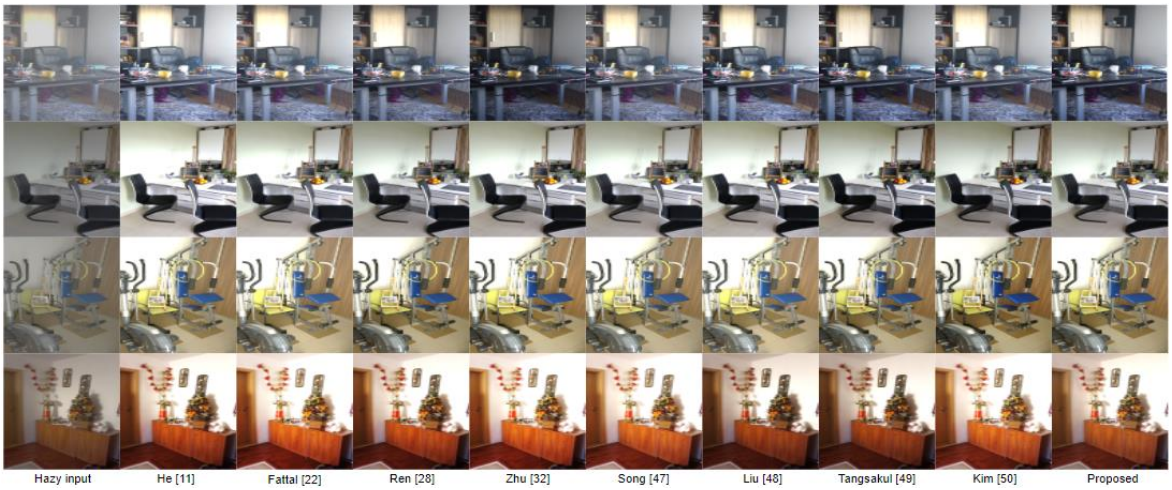

Fig. 10. Qualitative comparison of dehazed outputs on I-Haze dataset (a) Hazy input (b) He [11] (c) Fattal [22] (d) Ren [28] (e) Zhu [32] (f) Song [47] (g) Liu [48] (h) Tangsakul [49] (i) Kim [50] (j) Our proposed framework.

Fig. 11. Qualitative comparison of Dehazed outputs on D-Hazy dataset (a) Hazy input (b) He [11] (c) Fattal [22] (d) Ren [28] (e) Zhu [32] (f) Song [47] (g) Liu [48] (h) Tangsakul [49] (i) Kim [50] (j) Our proposed framework.

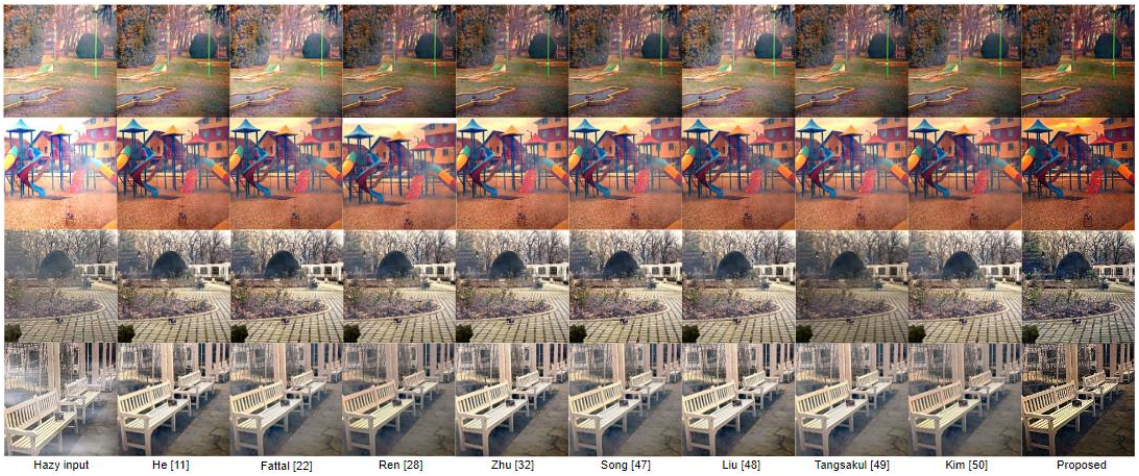

Fig. 12. Qualitative comparison of Dehazed outputs on O-Haze dataset (a) Hazy input (b) He [11] (c) Fattal [22] (d) Ren [28] (e) Zhu [32] (f) Song [47] (g) Liu [48] (h) Tangsakul [49] (i) Kim [50] (j) Our proposed framework.

Fig. 13. Qualitative comparison of Dehazed outputs on real-world aerial images with proposed framework Hazy image input (first row), Dehazed image output (second row).

TABLE I: AVERAGE QUANTITATIVE ANALYSIS BETWEEN PROPOSED FRAMEWORK AND OTHER EXISTING TECHNIQUES ON I-HAZE DATASET

Algorithm/ Metric	He [11]	Fattal [22]	Ren [28]	Zhu [32]	Song [47]	Liu [48]	Tangsakul [49]	Kim [50]	Proposed
SSIM	0.73	0.68	0.79	0.75	0.82	0.44	0.78	0.81	**0.85**
PSNR (dB)	17.27	17.69	21.54	18.74	25.17	18.81	22.73	24.36	**27.38**
FSIMc	0.762	0.815	0.823	0.885	0.853	0.936	0.896	0.866	**0.895**
VIF	0.624	0.592	0.629	0.611	0.637	0.694	0.713	0.782	**0.805**
MAE	0.74	0.68	0.69	0.65	0.57	0.59	0.64	0.54	**0.463**
NIQE	4.504	4.914	4.328	5.237	4.819	4.764	5.187	4.557	**3.281**

TABLE II: AVERAGE QUANTITATIVE ANALYSIS BETWEEN PROPOSED FRAMEWORK AND OTHER EXISTING TECHNIQUES ON D-HAZY DATASET

Algorithm/ Metric	He [11]	Fattal [22]	Ren [28]	Zhu [32]	Song [47]	Liu [48]	Tangsakul [49]	Kim [50]	Proposed
SSIM	0.65	0.63	0.78	0.71	0.80	0.78	0.77	0.83	**0.86**
PSNR (dB)	18.02	18.53	19.55	19.16	20.51	21.67	22.52	24.58	**27.58**
FSIMc	0.729	0.835	0.882	0.888	0.8748	0.897	0.906	0.914	**0.924**
VIF	0.615	0.616	0.672	0.638	0.646	0.715	0.761	0.796	**0.812**
MAE	0.69	0.68	0.71	0.67	0.65	0.66	0.63	0.59	**0.47**
NIQE	4.786	4.652	4.769	4.824	4.443	3.981	4.114	3.982	**3.727**

TABLE III: AVERAGE QUANTITATIVE ANALYSIS BETWEEN PROPOSED FRAMEWORK AND OTHER EXISTING TECHNIQUES ON O-HAZE DATASET

Algorithm/ Metric	He [11]	Fattal [22]	Ren [28]	Zhu [32]	Song [47]	Liu [48]	Tangsakul [49]	Kim [50]	Proposed
SSIM	0.77	0.72	0.76	0.65	0.64	0.66	0.79	0.77	**0.89**
PSNR (dB)	18.57	18.94	21.66	19.57	19.02	20.14	22.57	23.61	**27.59**
FSIMc	0.842	0.837	0.859	0.855	0.869	0.876	0.882	0.835	**0.967**
VIF	0.632	0.549	0.617	0.594	0.582	0.628	0.657	0.684	**0.796**
MAE	0.74	0.69	0.63	0.62	0.57	0.55	0.58	0.51	**0.4**
NIQE	4.168	4.879	3.751	4.716	4.182	4.379	3.863	3.418	**3.142**

TABLE VII: AVERAGE RUNNING TIME (IN SECONDS) ANALYSIS

Algorithm	He [11]	Fattal [22]	Ren [28]	Zhu [32]	Song [47]	Liu [48]	Tangsakul [49]	Kim [50]	Proposed
I-Haze	1.51	1.74	1.28	0.91	0.89	0.83	0.62	0.57	**0.32**
D-Hazy	1.47	1.69	1.18	1.23	0.81	0.67	0.59	0.36	**0.27**
O-Haze	1.84	1.89	1.38	1.25	0.72	0.64	0.51	0.49	**0.34**
Real-world	7.52	6.49	5.71	6.26	5.34	4.83	4.37	4.28	**2.12**
Average	3.085	2.953	2.388	2.413	1.94	1.743	1.524	1.425	**0.764**

TABLE IV: IMPACT OF HETEROGENEOUS CONVOLUTION ON CHC-UNET

CHC-UNet	Homogeneous Kernel	Heterogeneous Kernel	
		P=3	*P*=4
SSIM	0.784	0.829	0.86
PSNR	23.57	25.53	27.58
NIQE	4.438	3.647	3.727

TABLE V: IMPACT OF LFF MODULE AND SE NETWORK ON CHC-UNET

SE / LFF	SSIM	PSNR	NIQE
× / ×	0.618	20.17	4.5778
√ /×	0.789	22.97	4.218
× / √	0.674	23.82	4.174
√ / √	0.86	27.58	3.727

TABLE VI: IMPACT OF GROUP NORMALIZATION (GN) AND FRELU ACTIVATION IN LFF MODULE ON CHC-UNET

CHC-UNet	BN			GN		
	ReLU	PReLU	FReLU	ReLU	PReLU	FReLU
SSIM	0.612	0.694	0.742	0.638	0.756	0.865
PSNR	21.62	22.41	26.82	23.74	24.36	27.58

E. Ablation Analysis

It is vital to analyze the performance of each block in the validation of proposed framework. Firstly, the heterogeneous convolution filter in each convolution block UNet is ablated on D-Hazy dataset. The heterogeneous convolution kernels with two different values of *P* are used during the experimentation and its impact on the framework is shown in Table IV w.r.t. homogeneous convolution kernel.

The influence of LFF module in the squeeze and excitation mechanism has a better impact to recalibrate the features among different channels. The Table V shows the impact of SE network and LFF module on the framework.

The use of GN+FReLU non-linear activation function has improved the framework performance compared with BN and other variations of ReLU functions and its impact is shown in Table VI. Similar improved performance is observed on the other datasets and real-world images.

F. Running Time Analysis

The performance of the most recent deep learning based image dehazing research has attracted a great deal of interest, while the resource-constrained platform concerns are often ignored. Our framework was designed with the goal of achieving great performance while also meeting the demands of real-time processing. The Table VII provides the average running time analysis of the existing methods in comparison with proposed framework. It is observed that the methods, e.g., He [11], Fattal [22], Ren [28], Zhu [32], Song [47], Liu [48], Tangasakul [49], and Kim [50] requires extra processing time, whereas, our proposed framework is thought to be able to meet the demands for real-time processing since it achieves improved computational efficiency with the least amount of running time.

V. CONCLUSION

In this article, a Cascaded Heterogeneous Convolution UNet (CHC-UNet) framework with re-defined Squeeze-and-Excitation (SE) mechanism for adaptive image dehazing is proposed. Firstly, a light weight ShuffleNet model is designed to estimate the scene transmission map with a special feature of cross-channel information communication and without increasing computation cost. Then heterogeneous convolution kernels are used to mitigate the computing cost of repeated convolution in the UNet framework and an improvement is observed in SSIM and PSNR from 0.784 to 0.865 and 23.57 to 27.58 respectively. The computation of GN along with FReLU activation function is independent of batch sizes, generates the spatial dependencies and performs the separate non-linear transformations simultaneously. The performance improvement in SSIM and PSNR is

observed from 0.612 to 0.865 and from 21.62 to 27.58 with BN+ReLU and GN+FReLU respectively. For better image representation capability of network learning, a Local Feature Fusion (LFF) enabled SE mechanism is designed to model the channel interdependencies and perform feature recalibration. The performance with SE mechanism and LFF module is improved from 0.618 to 0.86 in SSIM, 20.17 to 27.58 in PSNR and 4.577 to 3.727 in NIQE. The average running time for validation of test images is significantly improved from 3.085s to 0.764s with 512×512 pixel size images. The proposed framework can be further extended with meta-heuristic optimization algorithms to optimize the hyper parameters in the network training to obtain a better dehazed output with reduced execution time for high resolution images and optimum hardware requirements. The dehazed images can be used for aerial image change detection applications, autonomous driving systems, night time surveillance and outdoor vision systems.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

AUTHOR CONTRIBUTIONS

The authors confirm their contribution to the article as follows: CH. Mohan Sai Kumar, Dr. R.S. Valarmathi have conceived the idea; CH. Mohan Sai Kumar has developed the theory and conducted research and analyzed the data; Dr. R.S. Valarmathi has supervised the findings; CH. Mohan Sai Kumar, Dr. R.S. Valarmathi have drafted and finalized the article.

REFERENCES

[1] C. M. S. Kumar, R. S. Valarmathi, and S. Aswath, "An empirical review on image dehazing techniques for change detection of land cover," in *Proc. of 2021 Asian Conference on Innovation in Technology*, Pune, India, 2021, pp. 1-9.

[2] N. Jiang, W. Chen, Y. Lin, T. Zhao, and C. W. Lin, "Underwater image enhancement with lightweight cascaded network," *IEEE Trans. on Multimedia*, vol. 24, pp. 4301-4313, 2022, doi: 10.1109/TMM.2021.3115442.

[3] Y. Chen, W. Li, C. Sakaridis, D. Dai, and L. V. Gool, "Domain adaptive faster R-CNN for object detection in the wild," in *Proc. of 2018 IEEE/CVF Conf. on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, 2018, pp. 3339-3348.

[4] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, "End-to-end united video dehazing and detection," in *Proc. of the AAAI Conf. on Artificial Intelligence*, New Orleans, LA, USA, 2018, https://doi.org/10.1609/aaai.v32i1.12287.

[5] D. Dai, C. Sakaridis, S. Hecker, and L. V. Gool, "Curriculum model adaptation with synthetic and real data for semantic foggy scene understanding," *International Journal of Computer Vision*, vol. 128, pp. 1182–1204, 2020, https://doi.org/10.1007/s11263-019-01182-4.

[6] D. Berman, T. Treibitz, and S. Avidan, "Non-local Image Dehazing," in *Proc. of 2016 IEEE Conf. on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 2016, pp. 1674-1682.

[7] Y. Jiang, C. Sun, Y. Zhao, and L. Yang, "Image dehazing using adaptive bi-channel priors on superpixels," *Computer Vision and Image Understanding*, vol. 165, pp. 17-32, Dec. 2017.

[8] M. Ju, Z. Gu, and D. Zhang, "Single image haze removal based on the improved atmospheric scattering model," *Neurocomputing*, vol. 260, pp. 180-191, Oct. 2017.

[9] P. W. Hsieh and P. C. Shao, "Variational contrast-saturation enhancement model for effective single image dehazing," *Signal Processing*, vol. 192, #108396, Mar. 2022.

[10] H. Zhou, H. Xiong, C. Li, *et al.*, "Single image dehazing based on weighted variational regularized model," *IEICE Trans. Inf. Syst.*, vol. 104, pp. 961–969, 2021.

[11] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 33, no. 12, pp. 2341-2353, 2011.

[12] Z. Ling, G. Fan, Y. Wang, and X. Lu, "Learning deep transmission network for single image dehazing," in *Proc. IEEE Int. Conf. Image Process.*, 2016, pp. 2296–2300.

[13] W. Ren, S. Liu, H. Zhang, *et al.*, "Single image dehazing via multi-scale convolutional neural networks," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 154–169.

[14] H. Zhang and V. M. Patel, "Densely connected pyramid dehazing network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 3194–3203.

[15] R. Mondal, S. Santra, and B. Chanda, "Image dehazing by joint estimation of transmittance and airlight using bi-directional consistency loss minimized FCN," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, 2018, pp. 920–928.

[16] Y. W. Lee, L. K. Wong, and J. See, "Image dehazing with contextualized attentive U-NET," in *Proc. of 2020 IEEE International Conf. on Image Processing*, Abu Dhabi, United Arab Emirates, 2020, pp. 1068-1072.

[17] R. Thomas, L. Thampi, S. Kamal, *et al.*, "Dehazing underwater images using encoder decoder based generic model-agnostic convolutional neural network," in *Proc. of 2021 International Symposium on Ocean Technology*, Kochi, India, 2021, pp. 1-4.

[18] A. S. A. Ghani and N. A. M. Isa, "Automatic system for improving underwater image contrast and color through recursive adaptive histogram modification," *Computers and Electronics in Agriculture*, vol. 141, pp. 181-195, 2017.

[19] S. He, Q. Yang, R. W. H. Lau, and M. H. Yang, "Fast weighted histograms for bilateral filtering and nearest neighbor searching," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 26, no. 5, pp. 891-902, 2016.

[20] H. Li, P. Zhuang, W. Wei, and J. Li, "Underwater image enhancement based on dehazing and color correction," in *Proc. of 2019 IEEE Intl. Conf. on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking*, Xiamen, China, 2019, pp. 1365-1370.

[21] M. Liu, Q. Jiang, and J. Hu, "Detection of highway lane lines and drivable regions based on dynamic image enhancement algorithm under unfavorable vision," *Computers & Electrical Engineering*, vol. 89, #106911, 2021.

[22] R. Fattal, "Dehazing using color-lines," *ACM Transaction on Graphics*, vol. 34, no. 13, pp. 1-14, 2014.

[23] Y. Tan and G. Wang, "Image haze removal based on superpixels and Markov random field," *IEEE Access*, vol. 8, pp. 60728-60736, 2020.

[24] Q. Zhu, J. Mai, and L. Shao, "A fast single image haze removal algorithm using color attenuation prior," *IEEE Trans. on Image Processing*, vol. 24, no. 11, pp. 3522 – 3533, 2015.

[25] K. Tang, J. Yang, and J. Wang, "Investigating haze-relevant features in a learning framework for image dehazing," in *Proc. of 2014 IEEE Conf. on Computer Vision and Pattern Recognition*, Columbus, OH, USA, 2014, pp. 2995-3002.

[26] Y. S. Lai, Y.L. Chen, and C. T. Hsu, "Single image dehazing with optimal transmission map," in *Proc. of the 21st International Conf. on Pattern Recognition*, Tsukuba, Japan, 2012, pp. 388-391.

[27] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, "Dehazenet: An end-to-end system for single image haze removal," *IEEE Trans. on Image Processing*, vol. 25, no. 11, pp. 5187-5198, 2016.

[28] W. Ren, J. Pan, H. Zhang, X. Cao, and M. H. Yang, "Single image dehazing via multi-scale convolutional neural networks with holistic edges," *International Journal of Computer Vision*, vol. 128, pp. 240–259, 2020, https://doi.org/10.1007/s11263-019-01235-8.

[29] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, "AOD-Net: All-in-one dehazing network," in *Proc. of 2017 IEEE International Conf. on Computer Vision*, Venice, Italy, 2017, pp. 4780-4788.

[30] W. Ren, L. Ma, J. Zhang, *et al.*, "Gated fusion network for single image dehazing," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, 2018*, 2018, pp. 3253-3261.

[31] T. Wang, L. Zhao, P. Huang, X. Zhang, and J. Xu, "Haze

concentration adaptive network for image dehazing," *Neurocomputing*, vol. 439, pp. 75-85, 2021,https://doi.org/10.1016/j.neucom.2021.01.042.
[32] J. Zhu, L. Meng, W. Wu, D. Choi, and J. Ni, "Generative adversarial network-based atmospheric scattering model for image dehazing," *Digital Communications and Networks*, vol. 7, no. 2, pp. 178-186, 2021.
[33] Y. Z. Su, Z. G. Cui, C. He, *et al.*, "Prior guided conditional generative adversarial network for single image dehazing," *Neurocomputing*, vol. 423, pp. 620–638, 2021.
[34] H. Dong, J. Pan, L. Xiang, *et al.*, "Multi-scale boosted dehazing network with dense feature fusion," in *Proc. of 2020 IEEE/CVF Conf. on Computer Vision and Pattern Recognition*, Seattle, WA, USA, 2020, pp. 2154-2164.
[35] Y. Shao, L. Li, W. Ren, C. Gao, and N. Sang, "Domain adaptation for image dehazing," in *Proc. of 2020 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, 2020, pp. 2805-2814.
[36] J. Shin, H. Park, and J. Paik, "Region-based dehazing via dual-supervised triple-convolutional network," *IEEE Trans. on Multimedia*, vol. 24, pp. 245-260, 2022, doi: 10.1109/TMM.2021.3050053.
[37] P. Huang, L. Zhao, R. Jiang, T. Wang, and X. Zhang, "Self-filtering image dehazing with self-supporting module," *Neurocomputing*, vol. 432, pp. 57-69, Apr. 2021.
[38] B. Xiao, Z. Zheng, X. Chen, *et al.*, "Single UHD image dehazing via interpretable pyramid network," *arXiv*, arXiv:2202.08589, 2022.
[39] L. A. Tran, S. Moon, and D. C. Park, "A novel encoder-decoder network with guided transmission map for single image dehazing," *Procedia Computer Science*, vol. 204, pp. 682-689, 2022. https://doi.org/10.1016/j.procs.2022.08.082.
[40] F. Alenezi, A. Armghan, and K. C. Santosh, "Underwater image dehazing using global color features," *Engineering Applications of Artificial Intelligence*, vol. 116, #105489, 2022. https://doi.org/10.1016/j.engappai.2022.105489.
[41] Y. Yang, C. Wang, R. Liu, *et al*., "Self-augmented unpaired image dehazing via density and depth decomposition," in *Proc. of 2022 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, LA, USA, 2022, pp. 2027-2036.
[42] Y. Song, Y. Zhou, H. Qian, and X. Du, "Rethinking performance gains in image dehazing networks," *arXiv*, arXiv:2209.11448, 2022.
[43] Z. Liu, Y. Lin, Y. Cao, *et al.*, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. of 2021 IEEE/CVF International Conf. on Computer Vision (ICCV)*, Montreal, QC, Canada, 2021, pp. 9992-10002.
[44] Y. Song, Z. He, H. Qian, and X. Du, "Vision transformers for single image dehazing," *IEEE Trans. on Image Processing*, vol. 32, pp. 1927-1941, 2023, doi: 10.1109/TIP.2023.3256763.
[45] Y. Wu and K. He, "Group normalization," *Proc. European Conf. on Computer Vision (ECCV)*, 2018, pp.3–19.
[46] N. Ma, X. Zhang, and J. Sun, "Funnel activation for visual recognition," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 351-368.
[47] Y. Song, J. Li, X. Wang, and X. Chen, "Single image dehazing using ranking convolutional neural network," *IEEE Trans. on Multimedia*, vol. 20, no. 6, pp. 1548-1560, 2018.
[48] X. Liu, Y. Ma, Z. Shi, and J. Chen, "GridDehazeNet: Attention-based multi-scale network for image dehazing," in *Proc. of 2019 IEEE/CVF International Conf. on Computer Vision (ICCV)*, Seoul, Korea (South), 2019, pp. 7313-7322, 2019.
[49] S. Tangsakul and S. Wongthanavasu, "Single image haze removal using deep cellular automata learning," in *IEEE Access*, vol. 8, pp. 103181-103199, 2020.
[50] G. Kim, S. W. Park, and J. Kwon, "pixel-wise wasserstein autoencoder for highly generative dehazing," *IEEE Trans. on Image Processing*, vol. 30, pp. 5452-5462, 2021, doi: 10.1109/TIP.2021.3084743.
[51] J. Peng, C. Shi, E. Laugeman, *et al.*, "Implementation of the structural SIMilarity (SSIM) index as a quantitative evaluation tool for dose distribution error detection," *Med Phys.*, vol. 47, no. 4, pp. 1907-1919, 2020, doi: 10.1002/mp.14010.
[52] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. on Image Processing*, vol. 20, no. 8, pp. 2378-2386, 2011.
[53] K. Ding, K. Ma, S. Wang, and E. P. Simoncelli, "Comparison of full-reference image quality models for optimization of image processing systems," *International Journal of Computer Vision*, vol. 129, pp. 1258–1281, 2021.
[54] Y. H. Liu, K. F. Yang, and H. M. Yan, "No-reference image quality assessment method based on visual parameters," *Journal of Electronic Science and Technology*, vol. 17, no. 2, pp. 171-184, 2019.

CH. Mohan Sai Kumar is a research scholar at Department of ECE, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai. He has received B.Tech. degree from Acharya Nagarjuna University, Guntur and M.Tech. degree from JNTU, Kakinada in 2013 and 2015 respectively. He is presently employed as assistant professor in the Department of Electronics and Communication Engineering at Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai. He has 8+ years of teaching experience in engineering. He published 14 research articles in International & National Journals and Conferences in the fields of Image Processing, Networks and Deep Learning.

R. S. Valarmathi has received B.E. degree and M.E. degree in Applied Electronics from Bharathiar University, Coimbatore in April 1989 and January 2000, respectively. She received her Ph.D. degree from Anna University, Chennai, in the area of Biometrics in 2009. She is presently employed as a professor in the Department of Electronics and Communication Engineering at Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai. She has 29+ years of teaching experience in various engineering colleges and 19 years of research experience. She has produced 12 Ph.D. graduates till date and published 129 research articles in international and national journals, 105 research articles in international and national conferences in the fields of image processing, biometrics, networks, VLSI, and cloud computing. She is the life member in Indian Society for Technical Education (ISTE), life member in Institution of Electronics and Telecommunication Engineers (IETE), member in Institution of Engineers (IE) and member in Institution of Engineering and Technology (IET).