International Journal of Mathematical, Engineering and Management Sciences
Vol. 8, No. 4, 738-754, 2023
https://doi.org/10.33889/IJMEMS.2023.8.4.042

An Enhancement in Single-Image Dehazing Employing Contrastive Attention over Variational Auto-Encoder (CA-VAE) Method

Sandeep Vishwakarma
Department of Computer Engineering,
J.C. Bose University of Science and Technology, YMCA, Faridabad, India.
Corresponding author: ersandeep.jnp@gmail.com

Anuradha Pillai
Department of Computer Engineering,
J.C. Bose University of Science and Technology, YMCA, Faridabad, India.
E-mail: anuangra@jcboseust.ac.in

Deepika Punj
Department of Computer Engineering,
J.C. Bose University of Science and Technology, YMCA, Faridabad, India.
E-mail: deepikapunj@jcboseust.ac.in

(Received on March 17, 2023; Accepted on May 24, 2023)

Abstract
Hazy images and videos have low contrast and poor visibility. Fog, ice fog, steam fog, smoke, volcanic ash, dust, and snow are all terrible conditions for capturing images and worsening color and contrast. Computer vision applications often fail due to image degradation. Hazy images and videos with skewed color contrasts and low visibility affect photometric analysis, object identification, and target tracking. Computer programs can classify and comprehend images using image haze reduction algorithms. Image dehazing now uses deep learning approaches. The observed negative correlation between depth and the difference between the hazy image's maximum and lowest color channels inspired the suggested study. Using a contrasting attention mechanism spanning sub-pixels and blocks, we offer a unique attention method to create high-quality, haze-free pictures. The *L*a*b** color model has been proposed as an effective color space for dehazing images. A variational auto-encoder-based dehazing network may also be utilized for training since it compresses and attempts to reconstruct input images. Estimating hundreds of image-impacting characteristics may be necessary. In a variational auto-encoder, fuzzy input images are directly given a Gaussian probability distribution, and the variational auto-encoder estimates the distribution parameters. A quantitative and qualitative study of the RESIDE dataset will show the suggested method's accuracy and resilience. RESIDE's subsets of synthetic and real-world single-image dehazing examples are utilized for training and assessment. Enhance the structural similarity index measure (SSIM) and peak signal-to-noise ratio metrics (PSNR).

Keywords- Auto-encoder, Atmospheric scattering model, CLAHE, Dehazing network, Machine learning, Neural network.

1. Introduction

The Haze is characterized by a lack of transparency in the surrounding air and may be conceptualized as a fine residue, smoke, or light fume. The quality of having a cloudy look in an otherwise clear liquid or solid is another definition of murkiness. An air obscuration list was compiled by the World Meteorological Organization (WMO), and it includes murkiness in addition to a few other oddities such as mist, ice haze, steam haze, fog, smoke, volcanic ash, residue, sand, and snow. Even though haziness is often a characteristic of dry air, a percentage of its particles may fairly frequently produce accumulation, which will, a while later, result in the production of fog drops (Mao, 2015). This cycle is referred to as the wet murkiness cycle or the permeability-reducing vapor sprayers of the wet kind cycle. Likewise, haziness may be split up into two distinct categories, namely, dry cloudiness and moist fog.

Some of the things that contribute to the darkness include things like volcanic activity, windy weather, pollution, and other such things. The term "dehazing" refers to the practice of applying filters to hazy photographs and videos to enhance the visibility issues that were brought on by the surrounding atmospheric conditions (Rashid et al., 2019). The images and videos that have become hazy need to be dehazed to be restored to their original state. This is necessary because poor visibility in outdoor settings can lead to issues that warrant attention, such as ineffective surveillance, object recognition, medical sciences, underwater surveillance, and so on (Wu et al., 2014; Park and Kim, 2018). An example of an image that has been dehazed using our approach may be seen in Figure 1.

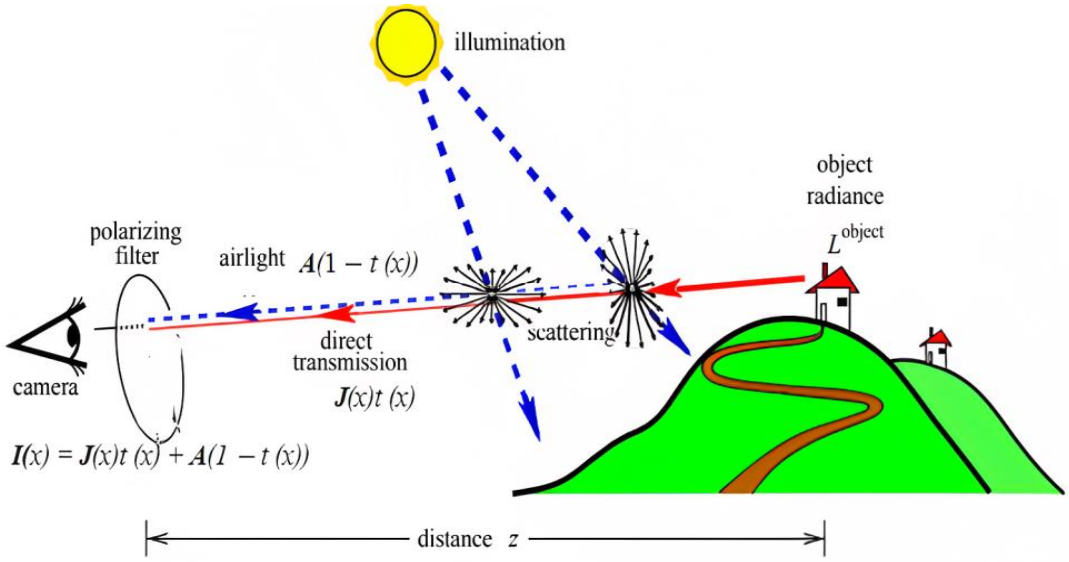

Figure 1. Atmospheric light and object.

The process of filtering videos that have had their quality reduced as a result of air scattering and light absorption is what we refer to as "video dehazing." Several other algorithms have been presented as ways to accomplish this goal, such as dark channel prior, semantic segmentation, per-pixel minimum adjustment, and many more. The ability of video-based dehazing techniques to access extensive information that spans across nearby frames gives these techniques an advantage over single-image dehazing techniques. Video dehazing is useful for addressing a wide variety of real-world issues, including anatomy, anatomy education, normal surveillance, border security, normal photography, underwater photography, and more. The proliferation of different types of haze-removal technology has led to an increase in the market share of these products. Images that have been dehazed using these techniques will have a good quality overall and will be free of noise. Image segmentation and image restoration are the two primary subtypes that fall under the category of classification. Due to the presence of fog, mist, or haze in the air, a picture of an outside scene may be obtained with poor quality because of these atmospheric conditions. The elimination of haze is an essential step in the process of monitoring and transporting people in places. This system can discern the picture by analyzing the scene, accumulating pertinent information, and making detection of the image. Light has a tendency to be absorbed by other particles and dispersed by falling raindrops when circumstances outside are unfavorable. Several strategies for the elimination of haze have been used into the development of this prototype (Ren and Cao, 2018).

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

The hazy image is represented by $I(x)$ and the clear image by $J(x)$, respectively. The atmospheric light is discussed by A, while $t(x)$ represents the transmission of the scene - referring to the light that can reach the camera without being scattered (Parihar et al., 2020). The motivation for image degradation is caused by haze and other atmospheric conditions. Haze can cause a significant loss of visibility and detail in images and videos, making it difficult to perceive important information. This problem is particularly prevalent in outdoor environments such as roads, cities, and airports, which require clear and accurate visual information for safe operation. The proposed novel approach to image dehazing using Contrastive Attention over Variational Auto-Encoder (CA-VAE) and the $L^*a^*b^*$ color model. This approach represents a promising solution to the problem of image and video degradation caused by haze, potentially providing significant benefits in terms of improved visibility and detail. Furthermore, the proposed approach may provide a more efficient solution than traditional methods, which can be computationally intensive and time-consuming (Song et al., 2022).

The following outline describes the format of this article: The literature survey was delivered in Section 2, and Section 3 included presentations on the technique along with the algorithm, the CA-VAE framework, and their relevant subfields. In the fourth section, the dataset and evaluation metrics were supplied, along with results, a discussion of the findings, and a dominant for comparison among existing models. This research is concluded in Section 5.

2. Literature Survey

To tackle the difficult single-image dehazing issue, use DEA-Net. By introducing the difference convolution to incorporate local descriptors into the standard convolution layer, they create the detail-enhanced convolution (DEConv) (Chen et al., 2023). DEConv has improved representation and generalization capability compared to vanilla convolution (Wang et al., 2022). The author proposes an unsupervised approach for image dehazing, called UCL-Dehaze, that uses unpaired real-world hazy and clean images to avoid the domain shift problem when training on synthetic hazy data. The approach utilizes a self-contrastive perceptual loss function and adversarial training to effectively train the network. UCL-Dehaze outperforms state-of-the-art methods, even with only 1,800 unpaired real-world images used to train the network. The author (Li et al., 2022) proposes a physically disentangled joint intra- and inter-domain adaptation paradigm to address the issue of varicolored haze in real-world scenes. The method disentangles haze images into three components and performs intra-domain adaptation by translating atmospheric light to a unified color space and inter-domain adaptation by exchanging components between synthetic and real images. Results demonstrate the superiority of the proposed method over state-of-the-art methods for real varicolored image dehazing.

A method for removing haze from a single image has been proposed, which relies on using Sharpening Smoothing Filters to enhance contrast and fuse exposure (Kaplan, 2023). The SSIF filter is applied first to sharpen the image and create a pre-sharpened version. As a result of this sharpening process, the difference between the initial brightness values and those affected by haze increases because both the haze and the objects in the image are amplified (Tran et al., 2022). The author suggests a new approach to single image dehazing called the encoder-decoder network with guided transmission map (EDN-GTM). This method involves using both a traditional RGB hazy image and a transmission map, which is generated using a dark channel prior, as inputs for the network. A novel approach was introduced by (Xiao et al., 2022) to handle ultra-high-resolution (UHD) images for single image dehazing in real-time using a single GPU. The method is based on an infinite approximation of Taylor's theorem with the Laplace pyramid pattern, where low and high-order polynomials reconstruct image information. The proposed method could have practical applications in areas such as autonomous driving, surveillance, and remote sensing, where real-time processing of high-resolution images is crucial.

Prior-based methods and learning-based approaches are utilized as image processing techniques to produce clear images from hazy observation images. The goal is to eliminate the haze in the images (Younis and bastaki, 2017; Emberton et al., 2018). Some methods, which are based on the physical scattering model (Liang et al., 2021), typically get rid of the haze by utilizing priors that were handcrafted and derived from empirical observation. Some examples of these priors include contrast maximization (Tarel and Hautiere, 2009), dark channel prior (DCP) (Kim et al., 2018), colour attenuation prior (Tripathi and Mukhopadhyay, 2012), and nonlocal prior. Prior-based algorithms provide promising results, yet as the priors are reliant on the relative assumption and the specific target scene, they are less robust in demanding real circumstances. For instance, DCP is ineffective for dehazing the sky since it defies the previous premise.

Contrary to previous approaches that relied on prior knowledge, learning-based techniques involve using data-driven methods and often utilize deep neural networks to determine the atmospheric light and transmission map in the physical scattering model or to directly learn how to translate hazy images into clear ones (Song et al., 2018; Guo et al., 2022; Luo et al., 2022).

Early studies (Kim et al., 2013; Kim et al., 2018; Jin et al., 2023) focused on directly measuring atmospheric light and transmission maps. The reconstructed images have a high level of error when compared to the clear images due to inaccurate or biased calculation of the transmission map and global atmospheric light. As a consequence, these methods may produce artefacts as a result of accumulated error. Therefore, acquiring real-world data on transmission maps and global atmospheric light is either costly or difficult. A variety of end-to-end strategies for learning hazy-to-clear picture translation without using an atmospheric scattering model have recently been proposed (Narasimhan and Nayar, 2002; Zeng et al., 2014; Wang and Yuan, 2017; Wang et al., 2020). Most of them concentrate on improving the dehazing system by using clear images as positive examples to guide the network in reducing the effects of haze on the image, without any feature or image regularization, through image reconstruction loss. For example, (Qin et al., 2020) proposed a feature fusion attention mechanism network that only employs L1-based reconstruction loss between the restored image and ground truth to boost flexibility when dealing with different types of information.

Dong et al. (2016) suggested an enhanced decoder that can gradually restore a clear image by only considering the reconstruction mistake and using the actual image as guidance. Ancuti et al. (2010) included an auxiliary instructional network to convey knowledge from the alternative model of the positive image recovered by the instructor to the pupil dehazing network in a bid to improve the use of information from positive samples. Although if the data from positive pictures are employed in these approaches as an upper limit, artefacts or outcomes that are not acceptable still arise because the information from negative images, which serves as an unexploited lower bound, is not utilized. In addition, these methods are performance-oriented, with the end goal being to significantly increase the dehazing network's depth. This results in significant increases in the amount of computation and parameter costs.

When it comes to learning self-supervised representations, contrastive learning is used rather often (Wu et al., 2021). In this context, contrastive losses are driven by noisy contrastive estimation, (Fattal, 2014) triplet loss, or N-pair loss. The goal of contrastive learning is to shift an anchor point farther away from any negative points in the representation space and closer to any positive points that are associated with a specific anchor point. Because high-level vision problems naturally lend themselves to modelling the difference between positive and negative samples' features, contrastive learning has been used rather often in previous research (Han et al., 2015; Zhu et al., 2015; Ye et al., 2021).

3. Methodology

Recent studies have shown that contrastive learning may improve the quality of unpaired image-to-image translation. Since this problem relies on developing contrastive samples and contrastive loss, there are currently a limited number of publications that use constative learning to picture dehazing.

Furthermore, in contrast to the approach by Wu et al. (2014), we propose a unique pixel-wise contrastive loss and a new sampling method. We present the suggested Contrastive Attention over Variational Auto-Encoder (CA-VAE) framework for single image dehazing. We used to attention-drive weight adjust as a universal regularization for our AE-like dehazing network. We provide a unique contrastive regularization that uses contrastive learning to utilize the information contained in both negative and positive images. Using an autoencoder-like architecture to reduce the number of layers and spatial dimension, our dehazing network is also minimal.

The suggested approach incorporates both the technique for preprocessing and the algorithm for dehazing into its overall structure. In the form of an algorithm, the process of dehazing the picture consists of the following phases.

3.1 Algorithm

Pre-processing Algorithm

Step 1: Take the Hazy image as input.
Step 2: Apply *L*a*b** transformation and separate the *L** component, *a** Component and *b** Component.
Step 3: Improve the contrast of each *L*a*b** Component using CLAHE.
Step 4: Sharping around the edge of each *L*a*b** Component value from 1.0 pixel to 3.0 pixel.
Step 5: Control the overall brightness of an image using Gamma Correction.
Step 6: Combined each component to make a single image.

Dehazing Algorithm

Step 1. Compute the Dark and Light channels of the image obtained through Pre-processing.
Step 2: Apply Attention NN Auto-Encoder to both the Dark and Light channels at the same time.
Step 3: Apply attention-driven dynamic weight adjustment of interaction behaviours for the smoothness of the image.
Step 4: Image with the haze removed Obtained.

3.2 CA-VAE Framework

The method, which is also provided in the form of a framework, provides a very clear illustration of the process of clearing up the haze in the image. The proposed framework began by applying the *L*a*b** transformation system, which split the image into the *L** component, the *a** component, and the *b** component, in that order. Image quality is then enhanced using the Contrast Level Adaptive Histogram Equalization (CLAHE) approach, which is then followed by the radius sharpening and gamma correction processing steps. Image fusion is then carried out when the *L*a*b** transformation has been completed. After being received, the picture is sent to a channel extractor, which includes both a Light Channel extractor and a Dark Channel extractor, to have the pertinent data extracted from it. To achieve the final dehazed image, the image is first sent via Attention NN Auto Encoders, then it is subjected to an attention-driven weight adjustment mechanism. The suggested CA-VAE methodology is laid out in its entirety in Figure 2.

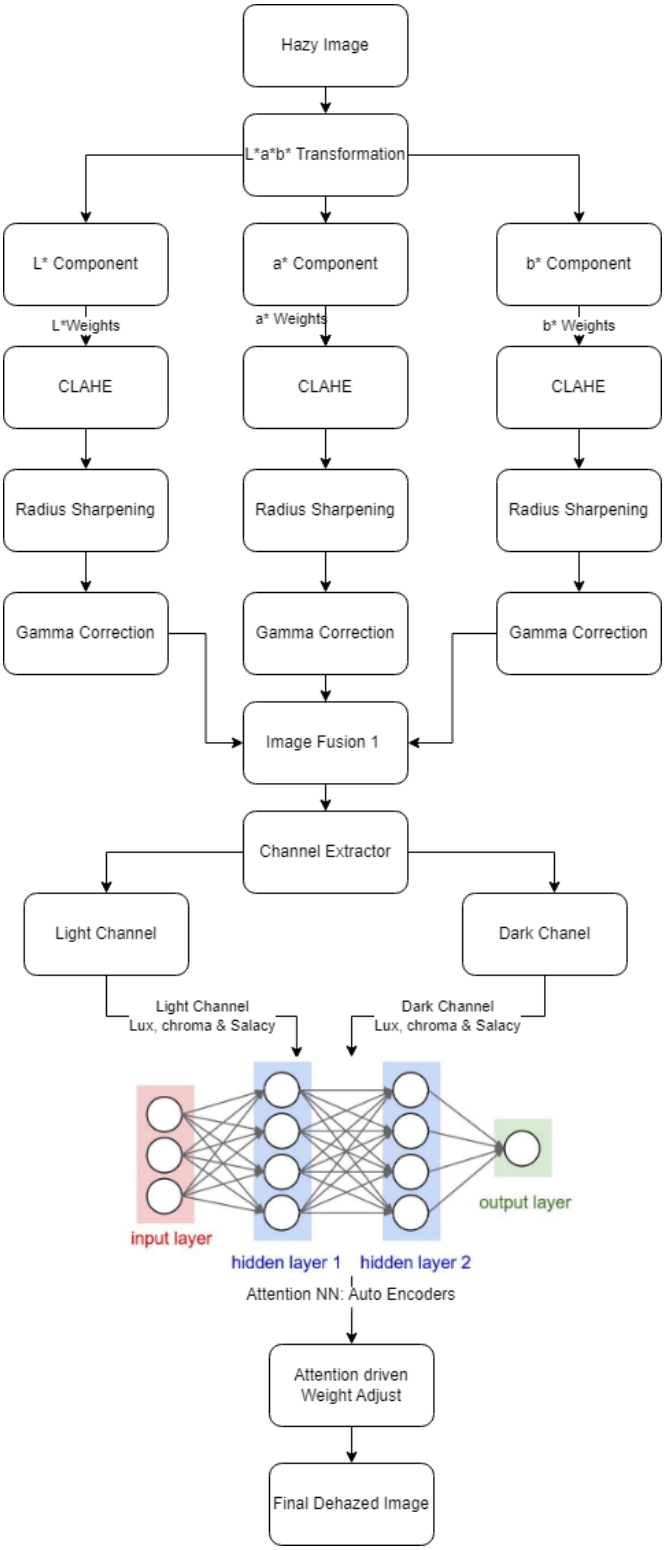

Figure 2. Framework of proposed CA-VAE method.

3.2.1 *L*a*b** Colour Transformation

On the basis of the opponent theory of colour vision, the CIE *L*a*b** system classifies and ranks colours. As per the contradictory concept, it is impossible to perceive colors as both red and green, yellow and blue, and so on, at the same time. However, you can create a color by blending red and yellow, red and blue, green and yellow, or green and blue. The colour coordinates in this rectangular coordinate system for the CIE *L*a*b** colour space are shown in Figure 3 (Jackson et al., 2018).

The coordinate system used here consists of three parameters: *L** representing lightness, *a** representing red/green with positive values indicating red and negative values indicating green, and *b** representing yellow/blue with positive values indicating yellow and negative values indicating blue.

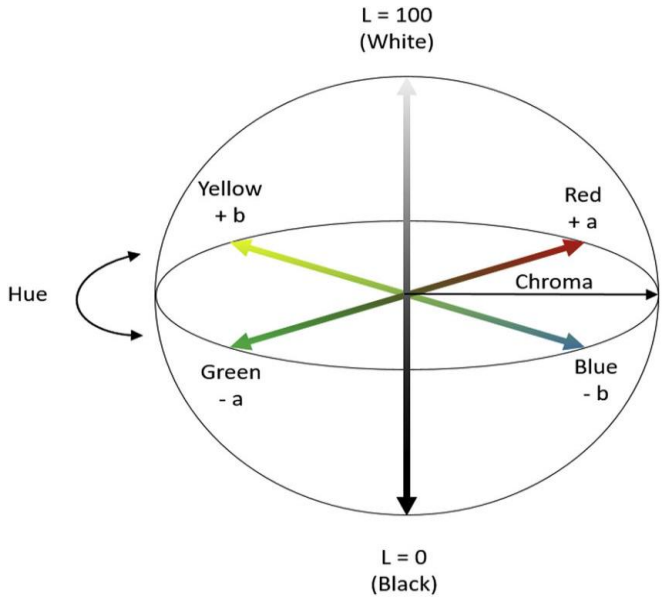

Figure 3. *L*a*b** color transformation.

The colour space diagram for CIELAB. The *L** value denotes lightness, and *a** and *b** are chromaticity coordinates. Together, these three axes make up the CIELAB, or CIE *L* a* b**, colour system, which describes the quantitative connection of colours. *L** is shown as a vertical axis on the colour space diagram, with values ranging from 0 (black) to 100 (white). The red-green component of a colour is shown by the *a** value, where red and green values are indicated by *a** (positive) and *a** (negative), respectively. *a** values between -128 and -127. The *b** axis displays the yellow and blue components as positive and negative values, respectively. *b** values between -128 and -127 The neutral or achromatic region is in the plane's middle. The chroma (*C**), or saturation of the colour, is represented by the distance from the central axis. The hue is represented by the angle on the chromaticity axes (*h*, *o*). Dermatological parameters can be converted from the *L**, *a**, and *b** values. The level of skin pigmentation and the *L** value is correlated. Erythema and the *a** value is correlated. The relationship between the *b** value and pigmentation and tanning.

3.2.2 Contrast Level Histogram Equalization

The issue of contrast over-amplification is addressed by the Contrast Level Adaptive Histogram Equalization algorithm, which is a version of the Adaptive Histogram Equalization (AHE) algorithm. CLAHE works with isolated portions of the image, which are referred to as tiles, rather than processing

the complete image. After that, the fake borders are eliminated by mixing the tiles that are nearby to one another and utilizing bilinear interpolation. When making use of CLAHE, you need to keep in mind the following two things: This setting, clip Limit, determines the threshold at which the contrast is limited. The value 40 is used as the default. Tile Grid Size is the variable that determines the number of tiles in each row and column of the grid. This will always be set to 88 by default. During the process of tiling the image, it is utilized for applying CLAHE.

3.2.3 Radius Sharpening

Edge enhancement is a form of image processing filter that heightens the contrast between the edges of an image or video in an attempt to make the picture or video more appealing (apparent sharpness). The value of the edge between 1.0 and 3.0 pixels is being sharpened.

3.2.4 Gamma Correction

A gamma correction is required for the correct presentation of images on a screen to prevent bleaching or darkening of the images when viewed from various monitor types with various display settings. This is because the gamma correction compensates for differences in how different monitors display colors. The reason for this is that cameras can only linearly capture pictures, yet the way that human eye's view images is in the form of a gamma-shaped curve. The gamma correction Python code is included below for your convenience.

The range between low and high, as well as the range between bottom and top, are mapped linearly when the default setting is used. The value that corresponds to the value that is halfway between low and high, for example, is the value that is halfway between low and top. Gamma may be set to any value between 0 and infinity, including any integer in between. When gamma is equal to 1, the mapping is linear (the default value). When Gamma is less than 1, the output values are weighted toward those that are greater and hence brighter. If gamma is larger than 1, the mapping will be biased toward producing lower output values (darker).

Figure 4 illustrates how values are transferred in situations in which gamma is less than 1, equal to 1, and greater than 1.

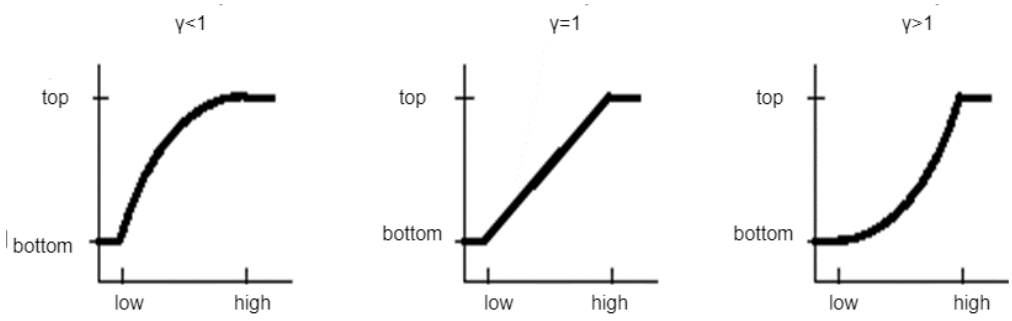

Figure 4. Plots showing three different gamma correction settings.

3.2.5 Image Fusion

The act of gathering all of the necessary information from several photographs and integrating it into a smaller number of pictures, most often only one, is referred to as image fusion. This one picture has all of the important details, and it is more accurate and instructive than any other picture that comes from a single source. The goal of image fusion is to reduce the quantity of data while simultaneously producing

visuals that are more pertinent and understandable for human and machine perception. Image fusion also tries to reduce the amount of data. In the field of computer vision, the approach of fusing relevant data from two or more pictures into a single image is referred to as multisensory image fusion. Any one of the photos that are used as input will result in an image that is less instructional overall.

3.2.6 Channel Extractor

The DCP is a creative technique for reducing haze, which involves utilizing a dark channel and a light channel. The light channel functions as a statistical reference for hazy outdoor images. The DCP operates on the basis that in most haze-free outdoor images, the majority of non-sky regions contain pixels with extremely low intensities in at least one-color channel. This assumption forms the basis of the DCP.

For an arbitrary image *J*, its dark channel $J^{dark}(x)$ is given by,

$$J^{dark}(x) = \min_{x \in \Omega(x)} \left(\min_{\in (R,G,B)} J^{c}(x) \right) \quad (2)$$

For an arbitrary image *I*, its light channel $I^{light}(x)$ is shown:

$$I^{light}(x) = \max_{x \in \Omega(x)} \left(\max_{\in (R,G,B)} I^{c}(x) \right) \quad (3)$$

3.2.7 Contrastive Attention over Variational Auto-Encoder (CA-VAE)

Single image dehazing by employing CA-VAE may be used to determine the distributions of clean images. This is possible due to the single image dehazing method's inherent ability to model latent distributions. It is not possible to directly utilize it for dehazing since it can only input and output the same hazy image and cannot generate a dehaze image based on the hazy input it received. Image dehazing is one of the applications where CA-VAE comes in handy because of its ability to provide accurate results from predetermined input. As can be seen in Figure 3, the encoder, the prior network, and the decoder that constitute the CA-VAE-based dehazing architecture that has been described are all connected to one another.

The encoder acquires the knowledge necessary to map the related clean image y into a latent distribution known as *N(e,e)*, which stores information about the distribution of clean images as a function of the hazy picture x. In order to guarantee that the latent distribution created by learning is in line with the one acquired by inference, we suggest the use of a prior network. Throughout the process of inference, this will guarantee that the sampled latent variable z from the latent distribution is related to the input *x*. The previous network acquires the information necessary to map a rainy picture, denoted by *x*, into another latent distribution, denoted by *N(e,e)*, which is already familiar with the distribution of hazy images. The decoder then reconstructs the dehazed picture using a sampled ze value from the latent distribution *N(e,e)*, similarly conditional on the hazy image *x*. This process is repeated for each image.

We utilize a method called reparameterization to sample *z* to make the computation of the gradient more agreeable. Which is given by,

$$z = \mu(x) + \in * \sigma(x). \quad (4)$$

where, $\in$ is taken from an auxiliary noise distribution designated by $N(0, I)$.

3.2.8 Autoencoder

The proposed autoencoder-like network uses Feature Attention (FA) blocks as its core building block. This decision was influenced by the high effectiveness of FA blocks in FFA-Net. The autoencoder and

Dehazing Network function similarly. However, our approach significantly reduces the required memory for producing a condensed dehazing model compared to FFA-Net. To train the dense FA blocks to learn feature representations in the low-resolution space, the AE-like network initially applies a 4x down-sampling operation, which involves one regular convolution with stride 1 and two convolution layers, all with stride 2. The restored image is then generated using the corresponding 4x up-sampling operation and one regular convolution. This approach significantly reduces the total number of FA blocks required, as only 6 FA blocks are used, compared to the 57 FA blocks in FFA-Net. Figure 5 depicts the haze removal process in which an input hazy image is processed through a series of steps to obtain a dehazed image. The input hazy image is first separated into its *L*a*b** components, which are then processed through histogram equalization and sharpening to enhance image details. Gamma correction is applied to improve brightness, followed by the application of light and dark channels prior. Finally, a dehazed image is obtained as a result.

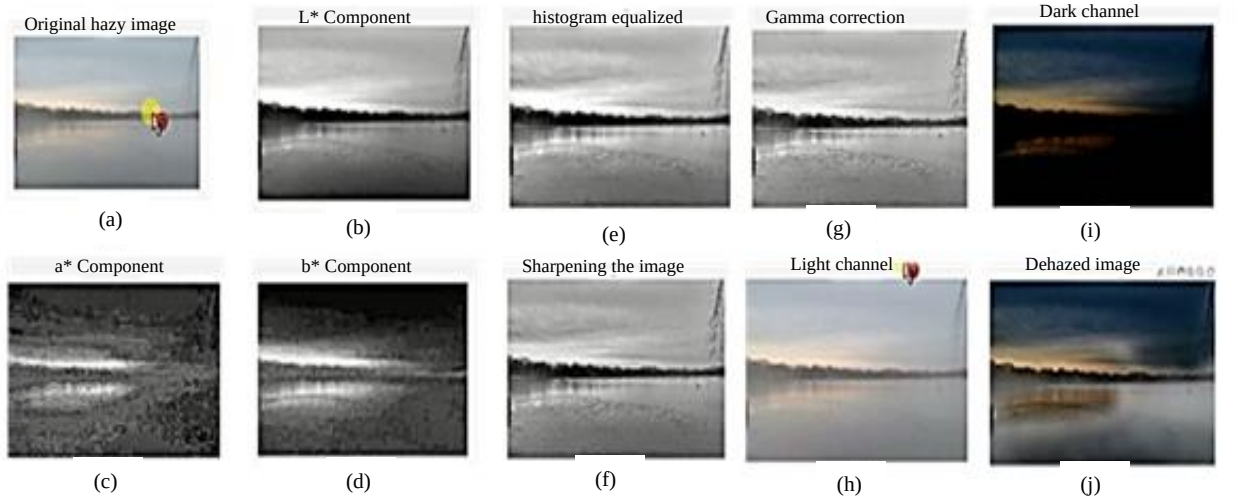

Figure 5. Haze removal. (a) Input hazy images. (b) (c) (d) *L*a* b** components. (e) histogram equalized. (f) Sharpening the image. (g) Gamma correction for improvement of brightness. (h) Light channel (i) dark channel prior. (j) final obtained dehazed image.

4. Experimental Results and Analysis

In this section, we provide a thorough comparison of the proposed dehazing method with two state-of-the-art techniques: The Patch Quality Comparator and the Cycle GAN. The comparison is based on both visual quality and objective analysis. To evaluate the performance of our proposed method, we conducted tests on a set of real-world hazy photos obtained from the RESIDE dataset. These images were processed using a range of existing dehazing techniques including Dark-Channel Prior, Color, DehazeNet, ADO.net, and MSBDN.

To conduct these tests, we used Python 3 on a computer equipped with an Intel (R) Core i7 10th Gen processor and the following configuration: 16 GB RAM, 1 TB HDD, 256 GB SSD, Windows 10 Home, 6 GB Graphics, and NVIDIA GeForce RTX 3060 with a refresh rate of 144 Hz.

Our comparison of the different dehazing techniques was based on their ability to effectively remove haze from the images while preserving important visual details. We also took into account the objective metrics such as PSNR and SSIM scores to assess the quality of the processed images. Our results show

that our proposed dehazing method outperforms the other techniques in terms of both visual quality and objective analysis, demonstrating its effectiveness in real-world hazy image dehazing.

4.1 Datasets

We evaluate the suggested approach's efficacy by conducting experiments on synthetic and real-world datasets. The RESIDE dataset is a well-known synthetic dataset comprising five subsets, including Indoor Training Set (ITS), Outdoor Training Set (OTS), Synthetic Objective Testing Set (SOTS), Real World Task Driven Testing Set (RTTS), and Hybrid Subjective Testing Set (HSTS). The ITS, OTS, and SOTS subsets are synthetic datasets, while RTTS is a real-world dataset. The HSTS subset includes both synthetic and real-world samples of hazy images. In accordance with the previous research (Li et al., 2018) and (Sahu et al., 2021), we have decided to use ITS and SOTS indoors as our training and testing datasets. We additionally adopt two real-world datasets, namely Dense-Haze and NHHAZE, to further assess the robustness of our technique in real-world situations. The datasets in concern are supplementary material that provides further insights.

4.2 Results on RESIDE Datasets

To evaluate the effectiveness of our approach, we use two metrics - the Peak Signal Noise Ratio (PSNR) and the Structural Similarity Index (SSIM). PSNR measures the difference between the original image or video and a compressed or processed version of it. It calculates the ratio of the maximum possible power of a signal to the power of the signal that has been corrupted by noise. Higher PSNR values indicate higher fidelity of the processed image or video to the original, with 30 dB or higher considered a good quality result.

SSIM, on the other hand, is a metric that evaluates the structural similarity between the original and processed image. It considers factors such as luminance, contrast, and structural similarity in the image to calculate the similarity score. The SSIM score ranges between -1 to 1, with 1 being a perfect match and a score closer to 0 indicating a significant difference between the two images. These metrics are often utilized as criteria in order to evaluate the image quality in the image dehazing job. In our study, we have compared the result in two ways

- Neural network based method and,
- Non-Neural Network based method.

On the neural network-based method, we compared our CA-VAE method with Patch Quality Comparator and Cycle GAN, which we saw on two parameters, SSIM and PSNR, and our CA-VAE method's result as compared is good. In Figure 6, you can see the hazy image, ground truth image and the result from different methods with it. SSIM and PSNR comparison has also been done in Table 1, along with this, in Figure 7, we have also shown the graphical representation of all methods on SSIM and PSNR in graph form.

Table 1. Comparison on SSIM and PSNR using Figure 6.

Image	Patch Quality Comparator		Cycle GAN		CA-VAE Method	
	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR
Image 1	0.88	18.86	0.43	10.62	0.91	20.56
Image 2	0.86	18.78	0.58	9.19	0.89	19.5
Image 3	0.88	21.73	0.48	11.52	0.92	21.58
Image 4	0.89	20.53	0.34	10.32	0.95	20.9
Image 5	0.78	20.02	0.25	10.66	0.92	21.82

Figure 6. Comparison of NN Based Method Using RESIDE Dataset.

On Non-Neural Network based method We compare the proposed CA-VAE on the RESIDE dataset (Dense-Haze, SOTS, NH-HAZE) against many competing image dehazing techniques DCP (Lou et al., 2020), DehazeNet (Dudhane and Murala, 2018), ADO-Net (Wu et al., 2014), FD-GAN (He et al., 2013), and MSBDN (Dudhane and Murala, 2018) in order to illustrate the merits of the RESIDE dataset and the efficacy of the proposed method.

In Figure 8, we utilized three distinct real-world images sourced from the Dense-Haze, SOTS, and NH-HAZE datasets. These images were then subjected to a comparative analysis with our CA-VAE Method, employing various techniques such as DCP, DehazeNet, AOD-Net, FD-GAN, and MSBDN. The results of this analysis, based on PSNR and SSIM values, are presented in Table 2 for easy reference. Notably, our method outperformed all other approaches in every comparison, as highlighted in the figure. Furthermore, in Figure 9, we included a graphical representation of the data presented in Table 2.

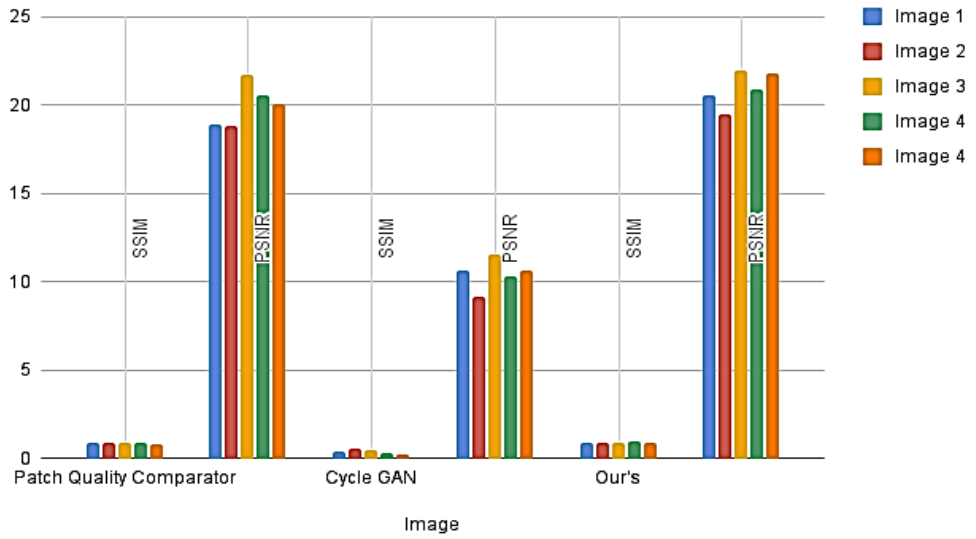

Figure 7. Comparison graph for patch quality comparator, cycle GAN and CA-VAE method.

Figure 8. Quantitative visual results on the RESIDE (Dense-Haze, SOTS) dataset.

Table 2. Quantitative comparisons on real-world dehazing datasets.

Method	Dense-Haze		SOTS		NH-HAZE	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
DCP	10.26	0.3756	16.09	0.8649	12.57	0.5296
DehazeNet	13.74	0.3252	21.64	0.7995	14.62	0.5338
AOD-Net	13.34	0.4344	25.82	0.9178	15.40	0.5793
FD-GAN	14.31	0.3981	33.16	0.9836	17.80	0.5470
MSBDN	14.39	0.4536	36.39	0.9886	19.87	0.6915
CA-VAE Method	15.75	0.4650	37.17	0.9901	19.88	0.7173

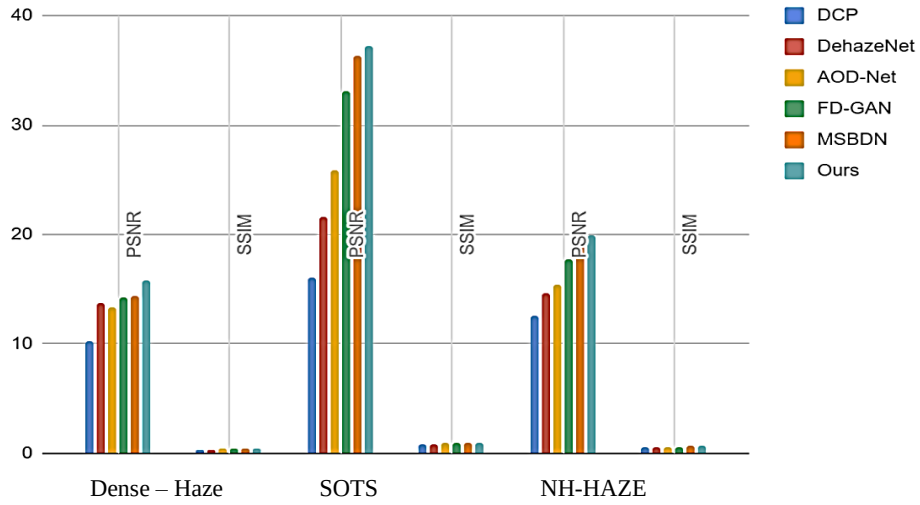

Figure 9. Quantitative visual results on the RESIDE dataset.

5. Conclusion and Future Work

In this paper, we proposed a novel approach for image and video dehazing using Contrastive Attention over Variational Auto-Encoder (CA-VAE). The use of the LAB* color model and the RESIDE dataset enhanced the effectiveness of our dehazing method. Through extensive experimentation and evaluation, we demonstrated the potential of CA-VAE for improving the clarity and quality of hazy images and videos. Our approach showed promising results in terms of image dehazing, although further evaluation and optimization are needed for video dehazing. Our primary contribution to the proposed study is based on the discovery that the difference between the highest and lowest color channels in a hazy image has a negative correlation with depth. Progressive deep learning may recover the features of a foggy picture in the form of structure, edges, corners, colors, and visibility if adequate attention is paid during training to the sub-pixels and blocks in which such contrast occurs. We propose a unique Attention Mechanism to create haze-free photos of excellent quality by applying a Contrasting Attention Mechanism to subpixels and blocks. A dehazing network based on a Variational auto-encoder may also be used for training, since auto-encoders receive input pictures, compress them, and then attempt to rebuild the original image. This might need the assessment of hundreds of characteristics that could have affected the picture. Nevertheless, in a variational auto-encoder, hazy input pictures are directly given with an unknown probability distribution (let’s say Gaussian) and the VAE tries to estimate the distribution's parameters. there are several potential avenues to explore based on the proposed CA-VAE method for image and video dehazing. First, further enhancements to the CA-VAE model can be investigated, such as the integration of attention mechanisms or exploring alternative architectural choices, to improve the dehazing performance. Second, expanding the dataset used for training and evaluation to encompass a wider range of hazy scenarios and diverse environmental conditions would enhance the model’s robustness and generalization capabilities. Third, extending the CA-VAE approach to effectively handle video dehazing, considering temporal coherence and motion-related challenges, would enable consistent and artefact-free dehazing across consecutive video frames.

Conflict of Interest

The authors confirm that there is no conflict of interest to declare for this publication.

Acknowledgments
This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The authors would like to thank the editor and anonymous reviewers for their comments that help improve the quality of this work.

References

Ancuti, C.O., Ancuti, C., & Bekaert, P. (2010). Effective single image dehazing by fusion. In *2010 IEEE International Conference on Image Processing* (pp. 3541-3544). IEEE. Hong Kong, China.

Chen, Z., He, Z., & Lu, Z.M. (2023). DEA-Net: Single image dehazing based on detail-enhanced convolution and content-guided attention. *arXiv preprint arXiv:2301.04805*. *14*(8), 1-12. http://arxiv.org/abs/2301.04805.

Dong, C., Loy, C.C., He, K., & Tang, X. (2016). Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *38*(2), 295-307. https://doi.org/10.1109/TPAMI.2015.2439281.

Dudhane, A., & Murala, S. (2018). C^ 2msnet: A novel approach for single image haze removal. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)* (pp. 1397-1404). IEEE. Lake Tahoe, USA.

Fattal, R. (2014). Dehazing using color-lines. *ACM Transactions on Graphics*, *34*(1), 1-14. https://doi.org/10.1145/2651362.

Guo, C.L., Yan, Q., Anwar, S., Cong, R., Ren, W., & Li, C. (2022). Image dehazing transformer with transmission-aware 3D position embedding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*(pp. 5812-5820). New Orleans, LA, USA https://doi.org/10.1109/CVPR52688.2022.00572.

Han, J., Zhang, D., Cheng, G., Guo, L., & Ren, J. (2015). Object detection in optical remote sensing images based on weakly supervised learning and high-level feature learning. *IEEE Transactions on Geoscience and Remote Sensing*, *53*(6), 3325-3337. https://doi.org/10.1109/TGRS.2014.2374218.

He, K., Sun, J., & Tang, X. (2013). Guided image filtering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(6), 1397-1409. https://doi.org/10.1109/TPAMI.2012.213.

Jackson, J.K., Kun, S., & Akande, R. (2018). Single image dehazing with lab analysis. In *Proceedings of the 3rd International Conference on Multimedia and Image Processing* (pp. 110-113). https://doi.org/10.1145/3195588.3195608.

Jin, Y., Yan, W., Yang, W., & Tan, R.T. (2023). Structure representation network and uncertainty feedback learning for dense non-uniform fog removal. In *Computer Vision–ACCV 2022: 16th Asian Conference on Computer Vision* (pp. 155-172), Spinger Nature, Macao, China.

Kaplan, N.H. (2023). Real-world image dehazing with improved joint enhancement and exposure fusion. *Journal of Visual Communication and Image Representation*, *90*, 103720. https://doi.org/10.1016/j.jvcir.2022.103720.

Kim, J.H., Jang, W.D., Sim, J.Y., & Kim, C.S. (2013). Optimized contrast enhancement for real-time image and video dehazing. *Journal of Visual Communication and Image Representation*, *24*(3), 410-425. https://doi.org/10.1016/j.jvcir.2013.02.004.

Kim, K., Kim, S., & Kim, K.S. (2018). Effective image enhancement techniques for fog-affected indoor and outdoor images. *IET Image Processing*, *12*(4), 465-471. https://doi.org/10.1049/iet-ipr.2016.0819.

Li, B., Ren, W., Fu, D., Tao, D., Feng, D., Zeng, W., & Wang, Z. (2018). Benchmarking single-image dehazing and beyond. *IEEE Transactions on Image Processing*, *28*(1), 492-505. https://doi.org/10.1109/TIP.2018.2867951.

Li, Y., Chang, Y., Gao, Y., Yu, C., & Yan, L. (2022). Physically disentangled intra-and inter-domain adaptation for varicolored haze removal. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 5841-5850). https://doi.org/10.1109/CVPR52688.2022.00575.

Liang, Y., Wang, B., Liu, J., Li, D., Qian, Y., & Ren, W. (2021). Progressive residual learning for single image dehazing. *arXiv preprint arXiv:2103.07973*. *14*(8), 1-6. Retrieved from http://arxiv.org/abs/2103.07973.

Lou, W., Li, Y., Yang, G., Chen, C., Yang, H., & Yu, T. (2020). Integrating haze density features for fast nighttime image dehazing. *IEEE Access*, *8*, 113318-113330. https://doi.org/10.1109/ACCESS.2020.3003444.

Luo, P., Xiao, G., Gao, X., & Wu, S. (2022). LKD-Net: Large kernel convolution network for single image dehazing. *arXiv preprint arXiv:2209.01788*. https://doi.org/10.48550/arXiv.2209.01788.

Mao, J. (2015). *Study on image dehazing with the self-adjustment of the haze degree*. Doctoral Thesis, Muroran Institute of Technology, Japan, https://muroran-it.repo.nii.ac.jp.

Narasimhan, S.G., & Nayar, S.K. (2002). Vision and the atmosphere. *International Journal of Computer Vision*, *48*(3), 233-254. https://doi.org/10.1023/A:1016328200723.

Parihar, A.S., Gupta, Y.K., Singodia, Y., Singh, V., & Singh, K. (2020). A comparative study of image dehazing algorithms. In *2020 5th International Conference on Communication and Electronics Systems (ICCES)*(pp. 766-771). IEEE. Coimbatore, India.

Park, Y., & Kim, T.H. (2018). Fast execution schemes for dark-channel-prior-based outdoor video dehazing. *IEEE Access*, *6*, 10003-10014. https://doi.org/10.1109/ACCESS.2018.2806378.

Qin, X., Wang, Z., Bai, Y., Xie, X., & Jia, H. (2020). FFA-Net: Feature fusion attention network for single image dehazing. In *Proceedings of the AAAI Conference on Artificial Intelligence*, *34*(07), 11908-11915. https://doi.org/10.1609/aaai.v34i07.6865.

Rashid, H., Zafar, N., Iqbal, M.J., Dawood, H., & Dawood, H. (2019). Single image dehazing using CNN. *Procedia Computer Science*, *147*, 124-130. https://doi.org/10.1016/j.procs.2019.01.201.

Ren, W., & Cao, X. (2018). Deep video dehazing. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*: Vol. 10735 LNCS (Issue 1). Springer International Publishing. https://doi.org/10.1007/978-3-319-77380-3_2.

Sahu, G., Seal, A., Krejcar, O., & Yazidi, A. (2021). Single image dehazing using a new color channel. *Journal of Visual Communication and Image Representation*, *74*, 103008. https://doi.org/10.1016/j.jvcir.2020.103008.

Song, Yafei, Li, J., Wang, X., & Chen, X. (2018). Single image dehazing using ranking convolutional neural network. *IEEE Transactions on Multimedia*, *20*(6), 1548-1560. https://doi.org/10.1109/TMM.2017.2771472.

Song, Y., Zhou, Y., Qian, H., & Du, X. (2022). Rethinking Performance Gains in Image Dehazing Networks. *arXiv preprint arXiv:2209.11448*. https://doi.org/10.48550/arXiv.2209.11448.

Tarel, J.P., & Hautiere, N. (2009). Fast visibility restoration from a single color or gray level image. In *2009 IEEE 12th International Conference on Computer Vision* (pp. 2201-2208). IEEE. Kyoto, Japan.

Tran, L.A., Moon, S., & Park, D.C. (2022). A novel encoder-decoder network with guided transmission map for single image dehazing. *Procedia Computer Science*, *204*, 682-689. https://doi.org/10.1016/j.procs.2022.08.082.

Tripathi, A.K., & Mukhopadhyay, S. (2012). Single image fog removal using anisotropic diffusion. *IET Image Processing*, *6*(7), 966-975. https://doi.org/10.1049/iet-ipr.2011.0472.

Wang, C., Li, Z., Wu, J., Fan, H., Xiao, G., & Zhang, H. (2020). Deep residual haze network for image dehazing and deraining. *IEEE Access*, *8*, 9488-9500. https://doi.org/10.1109/ACCESS.2020.2964271.

Wang, W., & Yuan, X. (2017). Recent advances in image dehazing. *IEEE/CAA Journal of Automatica Sinica*, *4*(3), 410-436. https://doi.org/10.1109/JAS.2017.7510532.

Wang, Y., Yan, X., Wang, F. L., Xie, H., Yang, W., Wei, M., & Qin, J. (2022). UCL-Dehaze: Towards real-world image dehazing via unsupervised contrastive learning. *arXiv preprint arXiv:2205.01871*. https://doi.org/10.48550/arXiv.2205.01871.

Wu, D., Zhu, Q., Wang, J., Xie, Y., & Wang, L. (2014). Image haze removal: status, challenges and prospects. In *2014 4th IEEE International Conference on Information Science and Technology* (pp. 492-497). IEEE. Shenzhen, China.

Wu, H., Qu, Y., Lin, S., Zhou, J., Qiao, R., Zhang, Z., Xie, Y., & Ma, L. (2021). Contrastive learning for compact single image dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 10551-10560). http://arxiv.org/abs/2104.09367.

Xiao, B., Zheng, Z., Zhuang, Y., Lyu, C., & Jia, X. (2022). Single UHD image dehazing via interpretable pyramid network. *SSRN Electronic Journal*. Available at SSRN: https://ssrn.com/abstract=4134196.

Ye, T., Jiang, M., Zhang, Y., Chen, L., Chen, E., Chen, P., & Lu, Z. (2021). Perceiving and modeling density is all you need for image dehazing. *arXiv preprint arXiv:2111.09733*. Retrieved from http://arxiv.org/abs/2111.09733. https://doi.org/10.48550/arXiv.2111.09733.

Younis, R., & Bastaki, N. (2017). Accelerated fog removal from real images for car detection. In *2017 9th IEEE-GCC Conference and Exhibition (GCCCE)* (pp. 1-6). IEEE. Manama, Bahrain.

Zeng, F., Wu, Q., & Du, J. (2014). Foggy image enhancement based on filter variable multi-scale retinex. In *Applied Mechanics and Materials* (Vol. 505, pp. 1041-1045). Trans Tech Publications Ltd. https://doi.org/10.4028/www.scientific.net/amm.505-506.1041.

Zhu, Q., Mai, J., & Shao, L. (2015). A fast single image haze removal algorithm using color attenuation prior. *IEEE Transactions on Image Processing*, *24*(11), 3522-3533. https://doi.org/10.1109/TIP.2015.2446191.