2024 IEEE International Conference on Robotics and Automation (ICRA 2024)
May 13-17, 2024. Yokohama, Japan

See Through the Real World Haze Scenes: Navigating the Synthetic-to-Real Gap in Challenging Image Dehazing

Shijie Chen†, Mohammad Mahdizadeh†, Chong Yu,
Jiayuan Fan*, *Member, IEEE,* Tao Chen, *Senior Member, IEEE*

***Abstract*— Dehazing real-world hazy images is challenging due to the complexity of natural haze, varying haze conditions, details preservation, and the risk of overexposure. Existing methods excel in synthetic hazy scenarios but struggle in the real world because they don't use all available features. Classical dehazing techniques primarily focus on low-level dehazing enhancements, whereas deep learning-based methods extract more intricate weather-related features. However, both of these approaches exhibit limitations in effectively addressing the real-world dehazing. To address these challenges, we introduce an innovative approach that combines the strengths of both modalities to dehaze and enhance the visibility of real-world hazy scenes. Firstly, we extract both low-level and deep features and then employ a pre-trained vector quantization GAN to create well-detailed data patches. A decoder, with a normalized module, effectively utilizes these high-quality features. Additionally, we introduce a controllable operation to improve feature matching. To further enhance dehazing and generalizability, the decoder's output undergoes a sequence of gamma-correction operations and generates a series of multi-exposure images that are combined to create a haze-free and higher-quality image. Our method effectively reduces haziness, enhances sharpness, preserves natural colors, and minimizes artifacts in challenging real-world scenarios. The approach surpasses five SOTA methods in both qualitative and quantitative evaluations across three key metrics, utilizing three synthetic and two real-world hazy datasets. Notably, it achieves a substantial improvement in real-world datasets over the second-best method, with 0.5702 and 0.129 in FADE metrics for the RTTS and Fattal datasets, respectively.**

I. INTRODUCTION

Image dehazing is a crucial task aimed at removing atmospheric fog and haze from images taken under adverse weather conditions in real-world scenarios. In such conditions, the presence of minute atmospheric particles scatters light in all directions [1], [2], severely reducing image visibility. In fact, real-world hazy images pose significant challenges for dehazing due to the variability in haze conditions, light scattering and absorption, multi-scattering, non-uniform haze distribution, color restoration, the tendency to capture excessive noise, and overexposure. These challenges not only make dehazing a highly demanding task but also affect its generalizability, whereas synthetic hazy images and their dehazing approaches do not account for these complex conditions. This limitation arises from neglecting to extract the most possible features including low-level and deep level from the images and process them in different shapes to get the most performance. Consequently, many image dehazing methods often struggle to perform effectively on real-world hazy images, highlighting their limitations in critical applications such as autonomous vehicles, drones, surveillance cameras, and photography [3].

Early dehazing methods, only are able to exploit low-level image features and process them as the main valuable data for dehazing including Dark Channel Prior (DCP) and Color Attenuation Prior (CAP) [4], [5], were pioneering. DCP involves calculating the dark channel of an RGB image to estimate scene transmission for generating haze-free results. CAP, on the other hand, employs depth and transmission map estimation techniques to recover haze-free images, demonstrating effectiveness across various haze types. However, CAP extensions have been proposed to address limitations under complex lighting conditions, incorporating adaptive weights, sparse representation, and learning-based approaches [4], [5], [6], [7], [8]. Nonetheless, these methods still exhibit drawbacks, such as they cannot be employed in all the foggy scenes, their suboptimal dehazing performance, sensitivity to color distribution, and high computational complexity [9], [10]. Significantly, they typically struggle to dehaze realistic hazy images, especially in the presence of the sky, which often leads to high color distortions and artifacts.

Deep learning methods have gained significant attention for their ability to learn intricate mappings between hazy and haze-free image pairs [11]. Methods like [12] and [13] have demonstrated high performance in terms of visual quality, noise and artifact resilience, and adaptability to diverse hazy conditions, including non-uniform haze. These methods also support real-time applications on low-power devices, making them suitable for a wide range of real-world uses [14]. However, addressing key challenges associated with deep learning-based approaches is crucial, especially concerning diverse hazy input images. Challenges include limited generalization, scarcity of ground truth data for training, and high computational complexity, all of which add complexity to this domain. More importantly, it has been

This work is supported by National Natural Science Foundation of China (No. 62071127, and 62101137), National Key Research and Development Program of China (No. 2022ZD0160100), Shanghai Natural Science Foundation (No. 23ZR1402900), Shanghai Municipal Science and Technology Major Project (No.2021SHZDZX0103). The computations in this research were performed using the CFFF platform of Fudan University.

Shijie Chen, Mohammad Mahdizadeh, and Tao Chen are with the School of Information Science and Technology, Fudan University, Shanghai 200433, China. Chong Yu, and Jiayuan Fan are with the Academy for Engineering and Technology, Fudan University, Shanghai 200433, China. E-mail: 22210720103@m.fudan.edu.cn, 20110720115@fudan.edu.cn, 21110860050@m.fudan.edu.cn, jyfan@fudan.edu.cn, and eetchen@fudan.edu.cn.

† These Authors Contributed Equally to This Work.

* Corresponding Author.

discovered that their performance is significantly inferior when operating in real-world haze conditions as compared to synthetic scenes. There might have several reasons, the real hazy images suffer from noise, overexposure, physical and dynamic hazy complexity, and the inability to extract intricate details concealed within a densely hazy image.

By considering both methods, we can unlock a wealth of benefits that would otherwise be neglected. Focusing solely on one method may result in a significant gap in our understanding and limit our potential for success. Considering the factors discussed above, we introduce a single-image dehazing framework that leverages deep learning and classical methods to tackle real-world hazy images, enhancing visibility and improving image quality. Our approach integrates Low-level and deep feature extraction techniques, utilizing a Vector Quantization GAN to obtain a discrete codebook of well-detailed data patches. These patches are used to mitigate the adverse effects of haze. The decoder component is enhanced with a normalized module to effectively utilize high-quality features, yielding pristine results. We introduce a controllable operation to refine the matching process and identify superior feature correspondences. To achieve comprehensive dehazing and adaptability to challenging real-world hazy scenarios, the decoder's output undergoes a sequence of gamma-correction operations, creating multiply exposed images. These images are seamlessly combined using a multi-scale Laplacian fusion strategy to produce haze-free, visually appealing, and high-quality results. The key contributions of our work are as follows:

- We propose a comprehensive and novel method for dehazing single RGB real-world images by leveraging both low-level and deep-level features. Firstly, it removes haze from the RGB foggy image. Then, by utilizing an exposure sequence generation, it enhances the dehazing procedure and reconstructs a highly appealing image. The proposed approach is capable of revealing finer details in real-world dense haze scenarios.
- We introduce a modular and adaptable approach that leverages complementary methods by utilizing images in various shapes and formats throughout the process. These divisions result in unique enhancements that cannot be achieved through a single-step procedure. The modules are flexible and can be applied to other methods.
- We conducted comprehensive experiments to assess the effectiveness of our method in dehazing three synthetic and two real-world datasets. Our proposed approach outperforms five SOTA methods in both qualitative and quantitative results, particularly in real-world scenes evaluated by FADE, BRISQUE, and Entropy metrics.

II. Related Works

A. Classical Methods

Early techniques for addressing haze removal are based on manually defined features and physical models [6], [19]. An initial strategy for single image dehazing introduced a method centered on DCP, to identify pixels that have lower intensity in at least one color channel, which can be considered as dark pixels. This approach exhibited promising outcomes and surpassed other SOTA techniques for haze removal [4]. Subsequent advancements, as detailed in [20], incorporated a physical model for light scattering and absorption, leveraging the image itself to estimate atmospheric parameters. Another variation, discussed in [21], initiated fusion attempts to enhance both color balance and contrast in underwater images. This method involves combining multiple images of the same scene captured with different color filters and utilizing a color balance algorithm to adjust the resulting image's colors. To enhance the versatility of these models, recent research documented in [22] introduced a perception-aware fusion technique based on a structural patch decomposition and fusion framework. This method enables the fusion of two complementary images in an isolated image space, allowing different components to be fused independently without any interference or loss of information. These methods suffer from the assumption of uniform haze, and lack of generalization. DCP algorithms frequently encounter challenges when it comes to dealing with color distortions and artifacts in the sky regions. The effectiveness of these methods in real-world image dehazing applications can be significantly impeded by their limitations.

Our method differs significantly from the above methods. Firstly, while the previous methods rely on handcrafted features and physical models, our approach employs a comprehensive and modular framework that combines classical and deep learning-based approaches for optimal results. This empowers us to achieve high-performance dehazing of real-world hazy images and enhance the resultant image. Secondly, we utilize the classical methods not as the primary dehazing approach but as a complementary technique to enhance overall image quality after a deep learning procedure.

B. Deep Learning-Based Methods

Recently, there has been a surge in methods harnessing the power of deep learning to address haze removal. Consequently, a multitude of works have employed CNNs for single image dehazing [23], [24]. One notable contribution, DehazeNet [14], introduces an end-to-end deep learning system based on a fully convolutional neural network. The model is extensively evaluated on multiple benchmark datasets and compared against various SOTA techniques, demonstrating its efficacy. Several variant methods, as proposed in [25] and [26], have explored the incorporation of multi-scale information and leveraged deep learning techniques to capture features across different scales. These networks typically adopt an encoder-decoder architecture with skip connections and integrate attention mechanisms to enhance their ability to handle complex haze conditions. Their performance superiority is evident in the evaluations conducted on various benchmark datasets, particularly in synthetic haze datasets. Additionally, an extension of encoder-decoder networks with attention mechanisms is presented in [24]. By employing attention mechanisms at various network levels, these meth-

ods can focus on pertinent image regions, improving their effectiveness in addressing challenging haze scenarios.

In current research endeavors [27], [28], the aim is to address the challenges posed by featuring non-uniform and dense haze. One such approach introduces a sliding self-attention wavelet network as an extension of the hierarchical encoder-decoder network. Furthermore, the spectral dual-channel encoding method [27] is proposed to tackle uniform haze scenarios by incorporating spectral dual-channel encoding into deep learning encoding techniques. To address nighttime dehazing, [29] presents a variational retinex model. Retinex-based dehazing methods leverage principles of color constancy and local image statistics to estimate and enhance the underlying scene radiance. These methods assume that haze-free images maintain a consistent color distribution across different regions, irrespective of atmospheric conditions.

Real-world haze exhibits significant variability in density, thickness, color, and distribution. Deep learning models may face challenges in effectively handling this complexity, as they often rely on simplifying assumptions that may not hold in all real-world scenarios. Moreover, they may prioritize haze removal at the cost of overall image quality. In real-world situations, achieving a balance between haze removal and preserving crucial image details, features, and overall image quality is often crucial.

Our approach to dehazing real-world hazy images stands out from conventional deep learning-based methods in several significant ways. Firstly, we go beyond haze removal by leveraging the power of both deep learning and classical approaches. What sets us apart is our commitment to enhancing the overall image quality. We achieve this by addressing issues like underexposure and overexposure through a multi-exposure sequence generation and fusion strategy. This innovative strategy seamlessly combines complementary classical methods, improving image quality and ensuring the suitability of our model for real-world dehazing scenarios. Secondly, we introduce a modular methodology that integrates various techniques, including image dehazing, fusion, and enhancement. This modular-based approach enhances the adaptability of our method for future advancements and makes it highly compatible for integration into other methodologies.

III. Proposed Method

Our goal is to enhance the visibility of images obscured by challenging real-world hazy weather conditions and restore lost details caused by atmospheric water droplets. To achieve this, we introduce an innovative framework that leverages Low-level and deep feature extraction. Specifically, the deep feature extraction module comprises multiple Residual Swin Transformer blocks (RSTB), each containing several Swin Transformer layers with a residual connection. Additionally, a Vector Quantization GAN is pre-trained on a large-scale, high-quality dataset to acquire a discrete codebook containing well-detailed data patches (WDDP). By replacing haze-induced distortions with WDDP, the decoder, equipped with a normalized feature alignment module, can effectively utilize superior features. To obtain clean results, we initially under-expose the resulting image through a series of gamma-correction operations. These under-exposed images are then integrated into a haze-free result using a multi-scale Laplacian fusion approach. This approach not only enhances the overall image quality but also avoids underexposure or overexposure in the resultant haze-free image.

A. Low-level and deep feature encoders

To provide a concise overview of Vector Quantization GAN's functioning, we begin with an input image patch denoted as P. This patch serves as the input for the Vector Quantization GAN encoder, which generates corresponding features represented as $\hat{f}$. Subsequently, each individual "pixel" $\hat{f}_{ij}$ within $\hat{f}$ is matched to the nearest well-detailed data patches (WDDP) in a codebook denoted as w, with dimensions $w \in \mathbb{R}^{K \times n}$. This matching process results in obtaining a discrete representation denoted as w_{ij}^q, which can be expressed as follows:

$$w_{ij}^q = \mathcal{M}\left(\hat{f}_{ij}\right) = \arg\min_{w_C \in \mathcal{W}} \left(\left\| \hat{f}_{ij} - w_C \right\|_2 \right), \quad (1)$$

C refers to the codebook's dimension, and the operation denoted as $M()$ signifies the matching process. Ultimately, the input P undergoes reconstruction via the function $\mathbf{Img}_{R1}$:

$$P' = \mathbf{Img}_{R1}\left(w^q\right) = \mathbf{Img}_{R1}\left(\mathcal{M}\left(\mathbf{E}_{R1}(P)\right)\right) \quad (2)$$

Our encoder, denoted as E, has been developed to achieve impressive feature extraction capabilities in image restoration. To elaborate, our encoder's initial stage involves a low-level feature extraction head composed of a series of residual layers, which efficiently downsample the features by a factor of 4. Subsequently, we incorporate four residual Swin Transformer blocks [31] to form the deep feature extraction module. To process a hazy image with specific dimensions of height (H), width (W), and number of input channels (C_{in}), $I_{LQ} \in \mathbb{R}^{H \times W \times C_{in}}$ we apply a 3x3 convolutional layer called to extract Low-level features. These features are represented as a tensor with the same height, width, and a reduced number of channels (C), which we denote as $F_0 \in \mathbb{R}^{H \times W \times C}$. The convolutional layer is particularly effective for initial visual processing, which results in more stable optimization and improved outcomes.

Introducing the decoder with a feature normalizer, we propose this technique to enhance the decoder's utilization of features reconstructed from well-detailed data patches (WDDP). One issue we address is the fidelity reduction in results caused by information loss during vector quantization in Vector Quantization GAN. Our solution involves mitigating this distortion by guiding the features before matching them with WDDP. More precisely, in this layer, we employ deformable convolution to align the features F_{R1}^i generated by $\mathbf{Img}_{R1}$ with the features F^i produced by Img, which can be expressed as follows:

$$F_w^i = \text{DCONV}\left(F_{R1}^i, \text{CONV}\left(\text{Concat}\left(F_{R1}^i, F^i\right)\right)\right) \quad (3)$$

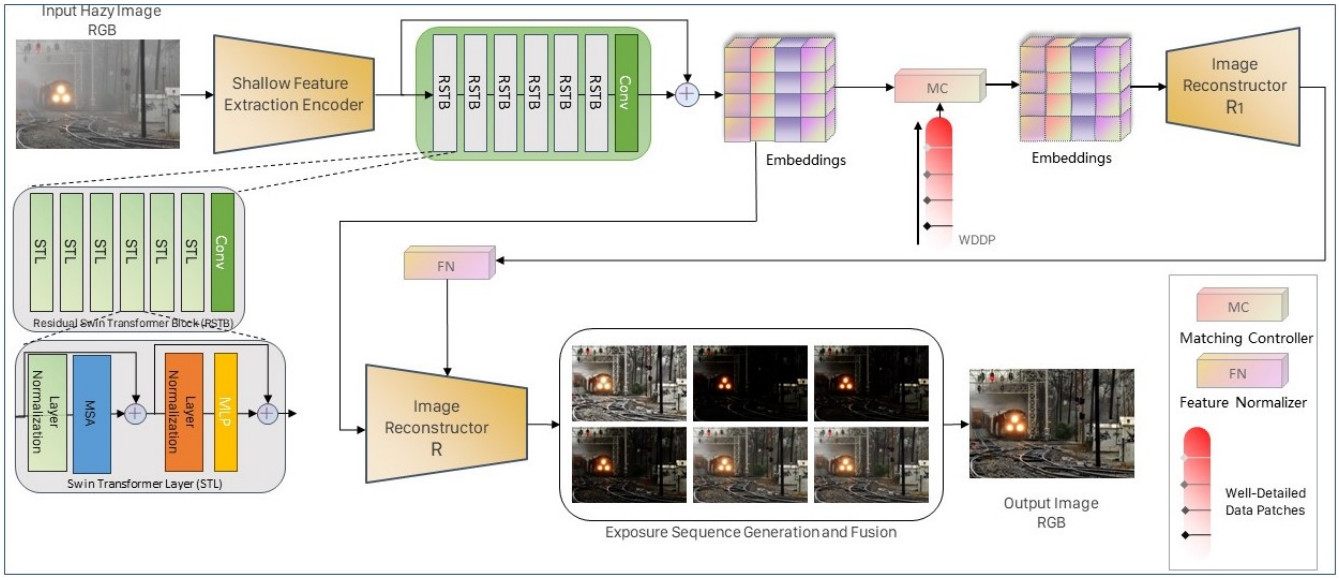

Fig. 1. The proposed pipeline for real-world image dehazing comprises of three main subsections. The first section is the Low-level and deep feature encoders network, which encodes the most possible features. Then, the image is reconstructed with normalized embedding, enhanced by the MC module and image reconstructors. Finally, a sequence of exposure is generated and fused to avoid under/over exposure, improve dehazing and overall image quality. This process involves Exposure Sequence Generation and Fusion.

where, F_w^i represents the features following the warping process, and DCONV represents the deformable convolutional layer. CONV denotes the convolutional layer responsible for generating offsets. Furthermore, we've observed that the values of F_w^i and F^i have an unstable ratio, leading to an insufficient combination. To address this, we equalize their contributions by ensuring they are within the same order of magnitude, which can be expressed as follows:

$$F^i = F^i + \frac{\sum F^i}{\sum F_w^i} F_w^i. \tag{4}$$

B. Adaptable WDDP and distance recalculation matching

To prevent diminished color saturation in challenging real-world data scenarios. the principal challenge lies in the difficulty of accurately identifying suitable WDDP, stemming from the domain gap between synthetic and real data. We randomly gathered 300 high-quality, clean images for input into the pre-trained Vector Quantization GAN and computed the activation frequencies $f_c \in \mathbb{R}^K$ for each code. Similarly, we processed 400 real hazy images through a dehazing network to calculate the frequency $f_h \in \mathbb{R}^K$. This analysis reveals a significant distribution shift, substantiating our claim that the inevitable domain gap results in divergent WDDP matching. Consequently, WDDP continues to possess untapped potential. We use two Voronoi diagrams to model the changes that occur in the high-dimensional space during feature matching. This helps us see how points originally in gray cells are now matched to colored cells after distances are recalibrated. This process helps identify the best WDDP for matching.

The divergence between the dehazing results and the clean domain is encapsulated by the distinction between these two probability distributions. Consequently, the challenge of real domain adaptation is reframed as the pursuit of an optimal parameter, denoted as $\hat{\alpha}$. This parameter's objective is to minimize the forward Kullback-Leibler Divergence between $P_c\left(x = w_k\right)$ and $P_h\left(x = w_k \mid \alpha\right)$, effectively representing the maximum likelihood estimation of α:

$$\arg\max_{\alpha} \prod_{i=1}^{K} P_c\left(x = w_i\right) P_h\left(x = w_i \mid \alpha\right) \tag{5}$$

C. Multi-exposure sequences generation via gamma correction transforms

One of the most straightforward techniques for adjusting exposure is gamma correction. This involves making global modifications to the intensities of an image by applying a power-function transform:

$$\mathbf{I}(x) \mapsto \alpha \cdot \mathbf{I}(x)^{\gamma} \tag{6}$$

Here, α and γ are real positive numbers. Gamma-corrected digital signals are quantized in a manner that employs wider quantization intervals in higher luminance ranges, where changes are less discernible.

$$\mathscr{C}(\Omega) = I_{\max}^{\Omega} - I_{\min}^{\Omega}\,, \tag{7}$$

Here, $I_{\max}^{\Omega}$ represents the maximum intensity value within the image, calculated as $I_{\max}^{\Omega} = \max\{I(x) \mid x \in \Omega\}$, where ω represents the image domain. Similarly, $I_{\min}$ represents the minimum intensity value within the image, calculated as and $I_{\min}^{\Omega} = \min\{I(x) \mid x \in \Omega\}$.

The vast majority of multiple exposure fusion can be categorized under a unified framework, which strives to determine the optimal weights Wk:

$$\mathbf{J}(x) = \sum_{k=1}^{K} \mathbf{W}_k(x)\mathbf{E}_k(x) \tag{8}$$

where K represents the count of differently exposed images denoted as $\mathbf{E}_k(x)$, and $J(x)$ represents a globally well-exposed image created by combining various correctly-exposed regions from Ek. The weights Wk are normalized to satisfy the condition $\sum_k \mathbf{W}_k(x) = 1 \forall x$. To prevent visual artifacts, we adopt a multi-scale approach for image fusion, which leverages the Laplacian pyramid. To seamlessly fuse different scales, we initially constructed a Gaussian pyramid for each weight map. We also construct a Laplacian pyramid for each E_k using the following recursive equation:

$$\mathbf{L}_k^i = \mathbf{E}_k^i - \mathrm{us}_2\left[\mathbf{E}_k^{i+1}\right] \tag{9}$$

Here, $us_2[\cdot]$ represents an operator that upsamples an image to twice its size. In the recursive formula provided above, we define $\mathbf{L}_k^N = \mathbf{E}_k^N$. As $\mathbf{L}_k^i(x)$ captures the frequency content of the original image at scale i, we can achieve a multi-scale combination of all $E_k(x)$ images by:

$$\begin{aligned}\mathbf{J}(x) = &\,\mathrm{us}_{(m,n)}\left[\mathbf{L}_1^1(x)\cdot\mathbf{W}_1^1(x) + \ldots + \mathbf{L}_K^1(x)\mathbf{W}_K^1(x)\right] + \\ &\,\mathrm{us}_{(m,n)}\left[\mathbf{L}_1^2(x)\cdot\mathbf{W}_1^2(x) + \ldots + \mathbf{L}_K^2(x)\mathbf{W}_K^2(x)\right] + \\ &\ldots \\ &+ \mathrm{us}_{(m,n)}\left[\mathbf{L}_1^N(x)\cdot\mathbf{W}_1^N(x) + \ldots + \mathbf{L}_K^N(x)\mathbf{W}_K^N(x)\right] \\ &= \sum_{i=1}^{N} \mathrm{us}_{(m,n)}\left[\sum_{k=1}^{K} \mathbf{L}_k^i(x)\cdot\mathbf{W}_k^i(x)\right]\end{aligned} \tag{10}$$

where, $us_{(m,n)}$ is the operator for upsampling any given image to the dimensions of E_k. when given a source image $\mathbf{E}^k(x) = \left(\mathbf{E}_k^R(x), \mathbf{E}_k^G(x), \mathbf{E}_k^B(x)\right)$, we measure contrast $\mathbf{C}_k(x)$ at each pixel x by calculating the absolute value of the response to a simple Laplacian filter. Additionally, we estimate saturation $S_k(x)$ for each pixel by computing the standard deviation across the color channels. Lastly, a haze map for each under-exposed image is generated by combining the contrast and saturation maps through a simple multiplication:

$$\mathbf{W}^k(x) = \mathbf{C}^k(x)\cdot\mathbf{S}^k(x) \tag{11}$$

IV. Experimental Results

Our model is trained on the Haze4k dataset [42], with a training set of 3000 image pairs and a test set of 1000 image pairs. Adam optimizer is utilized with $\beta_1 = 0.9$ and $\beta_2 = 0.99$. The learning rate is set to 0.0001, and the batch size is 16. To augment the training data, we apply random flipping, rotation, and cropping operations to the images, resizing them to 256×256. During the generation of exposure sequences, we set α to 1 and γ to $1, 2, 3, 4, 5$. To comprehensively evaluate the effectiveness of our method, we compared it with five STOA dehazing approaches, including DCP [32], MSBDN [33], FFA-Net [34], Dehamer [35], and C2PNet [36]. In addition, we use two real hazy image datasets. Fattal dataset [37], and RTTS dataset [38]. Three evaluation metrics including Fog Aware Density Evaluator (FADE) [39], Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) [40], and Entropy [41] are employed.

TABLE I

Quantitative evaluation on non-referenced real-world hazy Fettal [37], and RTTS [38]datasets. A comparison of average FADE [39], BRISQUE [40], and Entropy [41]. The best performance is in RED, and the second-best in BLUE.

	Fattal [37]			**RTTS** [38]		
	FADE ↓	**BRISQUE** ↓	**Entropy** ↑	**FADE** ↓	**BRISQUE** ↓	**Entropy** ↑
Hazy Images	1.0607	21.0797	7.0921	2.5742	37.0116	6.9414
DCP [32]	0.3498	19.8322	7.1332	0.9924	34.2984	6.9082
MSBDN [33]	0.5662	19.2355	7.2813	1.5783	29.9316	7.1665
FFA-Net [34]	0.6715	20.5973	7.2524	2.0734	34.4381	7.047
Dehamer [35]	0.6767	21.5404	7.2663	1.9211	34.5497	7.0835
C2PNet [36]	0.6362	19.9327	7.2575	2.0617	34.7967	7.0508
Ours	0.2208	19.7688	7.5175	0.4222	17.5412	7.6005

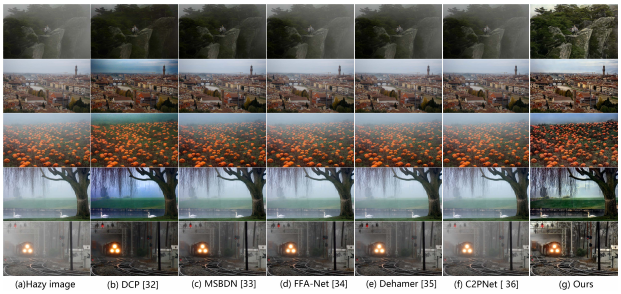

Fig. 2. Qualitative verification on the **Fattal dataset** [37] using state-of-the-art single image dehazing methods. Our method shows the best dehazing enhancement performance, with vibrant colors, balanced contrast, well-exposedness, and higher overall image quality.

A. Qualitative evaluations

Qualitative assessments should consider factors such as dehazing performance, ungraded color representation, vividness, preservation of textures and details, and overall image quality. Figures 2 and 3 show the qualitative results of our model and five SOTA methods on images from the Fattal and RTTS datasets. Our approach significantly improved dehazing, visibility, vividness, and overall quality. The resulting images had reduced or eliminated haze, with adjustments in brightness, contrast, color richness, sharpness, and texture while maintaining naturalness and visual appeal without amplifying noise or introducing artifacts.

Fig. 3. Qualitative verification on the **RTTS dataset** [38] using SOTA single image dehazing methods. Our method shows the best dehazing enhancement performance, with vibrant colors, balanced contrast, well-exposedness, and higher overall image quality.

TABLE II

QUANTITATIVE EVALUATION ON THREE SYNTHETIC REFERENCED DATASETS. A COMPARISON OF AVERAGE PSNR [43], SSIM [43], FADE [39], BRISQUE [40], AND ENTROPY [41]. THE BEST PERFORMANCE IS HIGHLIGHTED IN RED, AND THE SECOND-BEST IN BLUE.

	SOTS-indoor [19]					**SOTS-outdoor** [19]					**Haze4k** [42]				
	PSNR ↑	**SSIM** ↑	**FADE** ↓	**BRISQUE** ↓	**Entropy** ↑	**PSNR** ↑	**SSIM** ↑	**FADE** ↓	**BRISQUE** ↓	**Entropy** ↑	**PSNR** ↑	**SSIM** ↑	**FADE** ↓	**BRISQUE** ↓	**Entropy** ↑
Hazy images			1.7110	33.2919	7.1310			1.7610	20.4954	7.0084			1.5650	28.4805	7.1697
Ground truth	∞	1.0000	0.5016	34.7527	7.4142	∞	1.0000	0.7200	20.1623	7.4027	∞	1.0000	0.7012	27.9665	7.3462
DCP [32]	16.62	0.8179	0.5342	33.2664	7.1467	19.13	0.8148	0.5086	19.8272	7.1837	14.01	0.7604	0.5517	27.5699	7.2893
MSBDN [33]	32.85	0.9799	0.4948	32.4791	7.4419	34.49	0.9834	0.7143	18.868	7.4130	26.51	0.9561	0.8046	25.3660	7.3707
FFA-Net [34]	36.36	0.9879	0.4863	32.7938	7.4388	33.55	0.9843	0.7155	19.7558	7.4057	26.96	0.9503	1.2293	26.9592	7.2898
Dehamer [35]	36.63	0.9881	0.574	33.3093	7.5097	35.18	0.9860	0.7529	20.0784	7.4056	25.40	0.9659	0.8919	27.1457	7.361
C2PNet [36]	42.46	0.9946	0.5009	33.2681	7.4323	33.07	0.9583	0.7819	24.5441	7.3994	22.53	0.9412	1.1096	27.3691	7.3192
Ours	19.95	0.8078	0.2711	29.2923	7.6580	20.82	0.8180	0.3200	17.0903	7.6016	20.26	0.8315	0.3166	22.6369	7.5771

B. Quantitative evaluations

Table I presents a quantitative evaluation of SOTA methods using three metrics, including FADE [39], BRISQUE [40], and Entropy [41], across two publicly accessible non-referenced real hazy image datasets. Our method excels in reducing haze, evident in the lowest FADE scores on both datasets. Notably, on the Fattal dataset, we achieve 0.2208, surpassing the second-ranked method (DCP) at 0.3498. On the RTTS dataset, our method outperforms with a score of 0.4222, compared to 0.9924 by the second-best method (DCP). In terms of BRISQUE, our method consistently ranks second in the Fattal dataset, with a marginal difference in scores. However, on the RTTS dataset, we achieve the best BRISQUE score. Additionally, our method stands out by securing the highest Entropy scores on both datasets, indicating superior image clarity and information content.

Table II presents the quantitative evaluation of SOTA methods using five metrics, including PSNR [43], SSIM [43], FADE [39], BRISQUE [40], and Entropy [41], across three publicly accessible referenced synthetic hazy image datasets: SOTS-indoor [19], SOTS-outdoor [19], and Haze4k [42]. We provide the ground truth metrics as a reference for the performance. Our method performs well across all three datasets, achieving the best values for FADE, BRISQUE, and Entropy. On the other hand, we could not achieve good results in PSNR and SSIM due to achieving higher quality resultant images than the reference image.

C. Ablation study

To demonstrate the effectiveness of each component in our pipeline, we conducted a series of ablation experiments. Table III presents the model performance metrics on the Fattal and RTTS real-world haze datasets, with and without the use of well-detailed data patches (WDDP) and Exposure Sequence Generation and Fusion (ESGF). Both components improved the model performance: WDDP helped remove residual haze, while ESGF further revealed objects obscured by haze as shown in Figure 4.

TABLE III

ABLATION STUDY ON REAL-WORLD HAZY DATASETS. A COMPARISON OF AVERAGE FADE [39], BRISQUE [40], AND ENTROPY [41]

	Fattal [37]			**RTTS** [38]		
	FADE ↓	**BRISQUE** ↓	**Entropy** ↑	**FADE** ↓	**BRISQUE** ↓	**Entropy** ↑
Hazy Images	1.0607	21.0797	7.0921	2.5742	37.0116	6.9414
w/o WDDP&ESGF	0.4092	21.0891	7.4313	1.0305	23.2411	7.4575
w/o WDDP	0.2260	20.7481	7.4941	0.4743	19.3794	7.5876
w/o ESGF	0.4087	20.0602	7.4270	0.9174	21.3825	7.4803
Ours	**0.2208**	**19.7688**	**7.5175**	**0.4222**	**17.5412**	**7.6005**

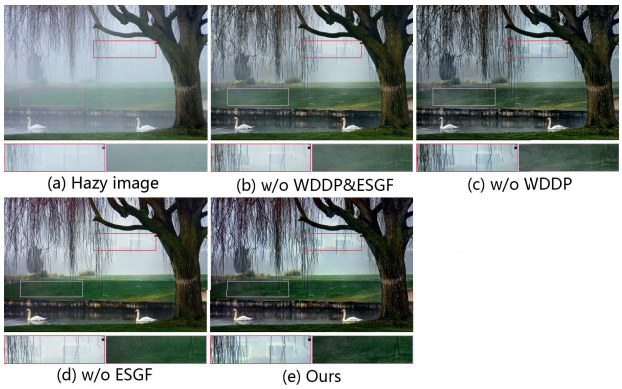

Fig. 4. Ablation results of WDDP and ESGF in our pipeline. The hazy image is the swan_input.png in **Fattal dataset** [37].

V. CONCLUSION

This paper presents a new method for improving visibility in real-world hazy conditions using a single RGB image. It not only improves real-world dehazing performance but also fulfills the gap of low dehazing performance of the other methods in real hazy scenes, and enhances the overall image quality. By employing both Low-level and deep feature extraction, applying dehazing to the extracted features, performing reconstruction, and generating a sequence of exposures, while leveraging the Laplacian pyramid to fuse them, our approach accomplishes the task of dehazing while enhancing the overall image quality. Our experimental evaluations, conducted on three synthetic and two real-world hazy datasets, firmly establish the superiority of our proposed method over existing SOTA dehazing techniques particularly in real-world scenarios. The results showcase substantial qualitative and quantitative improvements across three key metrics. Notably, our approach effectively mitigates haziness, enhances image sharpness, reveals finer details, maintains natural color fidelity, and minimizes artifacts. These qualities make the resulting images highly suitable for a wide range of image processing and machine vision applications, including surveillance, monitoring, and autonomous vehicles. For future work, we envisage exploring the potential of incorporating other modalities and refining the fusion process to achieve even more remarkable visibility enhancement.

REFERENCES

[1] X. Song et al., "TUSR-Net: Triple Unfolding Single Image Dehazing With Self-Regularization and Dual Feature to Pixel Attention," in *IEEE Transactions on Image Processing*, vol. 32, pp. 1231-1244, 2023, doi: 10.1109/TIP.2023.3234701.

[2] L. Schaul, C. Fredembach and S. Süsstrunk, "Color image dehazing using the near-infrared," *2009 16th IEEE International Conference on Image Processing (ICIP)*, Cairo, Egypt, 2009, pp. 1629-1632, doi: 10.1109/ICIP.2009.5413700.

[3] X. Xia and X. Liu, "Image dehazing technique based on polarimetric spectral analysis," *Optik*, vol. 127, no. 18, pp. 7350-7358, 2016, doi: 10.1016/j.ijleo.2016.05.071.

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

[5] Kaiming He, Jian Sun and Xiaoou Tang, "Single image haze removal using dark channel prior," 2009 *IEEE Conference on Computer Vision and Pattern Recognition*, Miami, FL, 2009, pp. 1956-1963, doi: 10.1109/CVPR.2009.5206515.

[6] Jiao Long, Zhenwei Shi and Wei Tang, "Fast haze removal for a single remote sensing image using dark channel prior," *2012 International Conference on Computer Vision in Remote Sensing*, Xiamen, 2012, pp. 132-135, doi: 10.1109/CVRS.2012.6421247.

[7] Lee, S., Yun, S., Nam, J.H. et al., "A review on dark channel prior based image dehazing algorithms," *Journal of Image and Video Processing*, vol. 4, 2016, https://doi.org/10.1186/s13640-016-0104-y.

[8] Q. Wang, L. Zhao, G. Tang, H. Zhao and X. Zhang, "single image Dehazing Using Color Attenuation Prior Based on Haze-Lines," *2019 IEEE International Conference on Big Data (Big Data)*, Los Angeles, CA, USA, 2019, pp. 5080-5087, doi: 10.1109/BigData47090.2019.9005603.

[9] Q. Zhu, J. Mai, and L. Shao, "Single Image Dehazing Using Color Attenuation Prior," in *Proceedings of the British Machine Vision Conference*, M. Valstar, A. French, and T. Pridmore, Eds. BMVA Press, 2014, pp. 114. doi: 10.5244/C.28.114.

[10] J. John and P. Sevugan, "Image Dehazing Through Dark Channel Prior and Color Attenuation Prior," in *Advances in Computing and Data Sciences*, M. Singh, V. Tyagi, P. K. Gupta, J. Flusser, T. "Oren, and V. R. Sonawane, Eds. Cham: Springer International Publishing, 2021, pp. 147-159, doi: 10.1007/978-3-030-88244-0_15.

[11] E. Wang, S. Shu and C. Fan, "CNN-based Single Image Dehazing via Attention Module," *2022 IEEE 5th International Conference on Automation*, Electronics and Electrical Engineering (AUTEEE), Shenyang, China, 2022, pp. 683-687, doi: 10.1109/AUTEEE56487.2022.9994347.

[12] Yu, H., Li, X., Feng, Y. et al. Underwater vision enhancement based on GAN with dehazing evaluation. *Appl Intell* 53, 5664–5680 (2023). https://doi.org/10.1007/s10489-022-03789-6

[13] S. Ki et al., "Fully End-to-End Learning Based Conditional Boundary Equilibrium GAN with Receptive Field Sizes Enlarged for Single Ultra-High Resolution Image Dehazing," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Salt Lake City, UT, USA, 2018, pp. 930-9307, doi: 10.1109/CVPRW.2018.00126.

[14] B. Cai, X. Xu, K. Jia, C. Qing and D. Tao, "DehazeNet: An End-to-End System for Single Image Haze Removal," in *IEEE Transactions on Image Processing*, vol. 25, no. 11, pp. 5187-5198, Nov. 2016, doi: 10.1109/TIP.2016.2598681.

[15] B. Li, X. Peng, Z. Wang, J. Xu and D. Feng, "AOD-Net: All-in-One Dehazing Network," *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, 2017, pp. 4780-4788, doi: 10.1109/ICCV.2017.511.

[16] W. Ren et al., "Gated Fusion Network for Single Image Dehazing," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, 2018, pp. 3253-3261, doi: 10.1109/CVPR.2018.00343.

[17] Zou, D., Yang, B., Li, Y., Zhang, X., & Pang, L. (2023). Visible and NIR Image Fusion Based on Multiscale Gradient Guided Edge-Smoothing Model and Local Gradient Weight. *IEEE Sensors Journal*, 23(3), 2783-2793. doi: 10.1109/JSEN.2022.3232150.

[18] Li, J., Liu, J., Zhou, S., Zhang, Q., & Kasabov, N. K. (2023). Infrared and visible image fusion based on residual dense network and gradient loss. *Infrared Physics & Technology*, 128, 104486. https://doi.org/10.1016/j.infrared.2022.104486.

[19] B. Li et al., "Benchmarking Single-Image Dehazing and Beyond," in *IEEE Transactions on Image Processing*, vol. 28, no. 1, pp. 492-505, Jan. 2019, doi: 10.1109/TIP.2018.2867951.

[20] R. T. Tan, "Visibility in bad weather from a single image," *2008 IEEE Conference on Computer Vision and Pattern Recognition*, Anchorage, AK, USA, 2008, pp. 1-8, doi: 10.1109/CVPR.2008.4587643.

[21] C. O. Ancuti, C. Ancuti, C. De Vleeschouwer and P. Bekaert, "Color Balance and Fusion for Underwater Image Enhancement," in *IEEE Transactions on Image Processing*, vol. 27, no. 1, pp. 379-393, Jan. 2018, doi: 10.1109/TIP.2017.2759252.

[22] Y. Kang, Q. Jiang, C. Li, W. Ren, H. Liu and P. Wang, "A Perception-Aware Decomposition and Fusion Framework for Underwater Image Enhancement," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 33, no. 3, pp. 988-1002, March 2023, doi: 10.1109/TCSVT.2022.3208100.

[23] Z. Ling, G. Fan, Y. Wang and X. Lu, "Learning deep transmission network for single image dehazing," *2016 IEEE International Conference on Image Processing (ICIP)*, Phoenix, AZ, USA, 2016, pp. 2296-2300, doi: 10.1109/ICIP.2016.7532768.

[24] S. Gao, J. Zhu and H. Xi, "Attention-Based Encoder-Decoder Network For Single Image Dehazing," *2021 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, Shenzhen, China, 2021, pp. 1-6, doi: 10.1109/ICMEW53276.2021.9455979.

[25] S. Chen, Y. Chen, Y. Qu, J. Huang and M. Hong, "Multi-Scale Adaptive Dehazing Network," *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Long Beach, CA, USA, 2019, pp. 2051-2059, doi: 10.1109/CVPRW.2019.00257

[26] W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, and M.H. Yang, "Single Image Dehazing via Multi-scale Convolutional Neural Networks," in *Computer Vision - ECCV 2016*, edited by B. Leibe, J. Matas, N. Sebe, and M. Welling, vol. 9906 of Lecture Notes in Computer Science, Springer, Cham, 2016, pp. 154-169, doi: 10.1007/978-3-319-46475-6-10.

[27] Z. Zhu, D. Zhang, Z. Wang, S. Feng and P. Duan, "Spectral Dual-Channel Encoding for Image Dehazing," in *IEEE Transactions on Circuits and Systems for Video Technology*, doi: 10.1109/TCSVT.2023.3264717.

[28] Y. Feng, X. Meng, F. Zhou, W. Lin and Z. Su, "Real-world Non-homogeneous Haze Removal by Sliding Self-attention Wavelet Network," in *IEEE Transactions on Circuits and Systems for Video Technology*, doi: 10.1109/TCSVT.2023.3256414.

[29] Y. Liu, Z. Yan, J. Tan and Y. Li, "Multi-Purpose Oriented Single Nighttime Image Haze Removal Based on Unified Variational Retinex Model," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 33, no. 4, pp. 1643-1657, April 2023, doi: 10.1109/TCSVT.2022.3214430.

[30] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, and R. Timofte, "SwinIR: Image restoration using Swin Transformer," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 1833-1844, 2021.

[31] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 10012–10022, 2021.

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

[33] H. Dong et al., "Multi-Scale Boosted Dehazing Network With Dense Feature Fusion," *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, 2020, pp. 2154-2164, doi: 10.1109/CVPR42600.2020.00223.

[34] X. Qin, Z. Wang, Y. Bai, X. Xie, and H. Jia, "FFA-Net: Feature Fusion Attention Network for Single Image Dehazing," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 07, pp. 11908-11915, 2020. doi: 10.1609/aaai.v34i07.6865.

[35] C. Guo, Q. Yan, S. Anwar, R. Cong, W. Ren and C. Li, "Image Dehazing Transformer with Transmission-Aware 3D Position Embedding,"*2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 2022*, pp. 5802-5810, doi: 10.1109/CVPR52688.2022.00572.

[36] Y. Zheng, J. Zhan, S. He, J. Dong and Y. Du, "Curricular Contrastive Regularization for Physics-Aware Single Image Dehazing," *2023 IEEE/CVF Conference on Computer Vision and Pattern Recog-*

nition (CVPR), Vancouver, BC, Canada, 2023, pp. 5785-5794, doi: 10.1109/CVPR52729.2023.00560.

[37] R. Fattal, "Dehazing Using Color-Lines," *ACM Trans. Graph.*, vol. 34, no. 1, art. 13, Nov. 2014. doi: 10.1145/2651362.

[38] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, and Z. Wang, "Benchmarking Single-Image Dehazing and Beyond," *IEEE Transactions on Image Processing*, vol. 28, no. 1, pp. 492-505, 2019.

[39] L. K. Choi, J. You, and A. C. Bovik, "Referenceless Prediction of Perceptual Fog Density and Perceptual Image Defogging,"*IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3888-3901, Nov. 2015

[40] A. Mittal, A. K. Moorthy and A. C. Bovik, "No-Reference Image Quality Assessment in the Spatial Domain," in *IEEE Transactions on Image Processing*, vol. 21, no. 12, pp. 4695-4708, Dec. 2012, doi: 10.1109/TIP.2012.2214050.

[41] D. Y. Tsai, Y. Lee, and E. Matsuyama, "Information Entropy Measure for Evaluation of Image Quality," *J. Digit. Imaging*, vol. 21, pp. 338-347, 2008. [Online]. Available: https://doi.org/10.1007/s10278-007-9044-5.

[42] Y. Liu, L. Zhu, S. Pei, H. Fu, J. Qin, Q. Zhang, L. Wan, and W. Feng, "From Synthetic to Real: Image Dehazing Collaborating with Unlabeled Real Data," in Proceedings of the 29th ACM International Conference on Multimedia (MM '21), Virtual Event, China, 2021, pp. 50-58, doi: 10.1145/3474085.3475331.

[43] Zhou Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13 no. 4, pp. 600–612, Apr. 2004,doi: 10.1109/TIP.2003.819861