photonics

MDPI

Article

A Multi-Depth Enhanced Adaptive Fusion Polarization-Guided Network for Image Dehazing

Yiting Li [1], Jun Chang [1,*], Qiang Fu [2], Hua Cai [2] and Xianzhu Liu [2]

[1] National Key Laboratory on Near-Surface Detection, School of Optics and Photonics, Beijing Institute of Technology, Beijing 100081, China; 3120205311@bit.edu.cn
[2] National and Local Joint Engineering Research Center of Space Optoelectronics Technology, Changchun University of Science and Technology, Changchun 130022, China; fuqiang@cust.edu.cn (Q.F.); caihua@cust.edu.cn (H.C.); liuxianzhu@cust.edu.cn (X.L.)
* Correspondence: optics_chang@126.com

Abstract

To address insufficient feature representation and the insufficient utilization of auxiliary information in images degraded by complex haze, this paper proposes a Multi-depth Enhanced Adaptive Fusion Polarization-guided Network (MEAF-P-Net). The network enhances dehazing feature representation through multi-depth feature fusion and introduces a Polarization-Guided Fusion Module (PGFM) that utilizes polarization features at different resolutions to assist image restoration. Experiments are conducted on a synthetic visible-light dehazing dataset and a synthetic polarization dehazing dataset constructed from PolarLITIS. For RGB image dehazing, the MEAF backbone achieves a PSNR of 31.40 dB and an SSIM of 0.933, which are 1.89 dB and 0.0201 lower than those achieved by ED-Former, respectively. For polarization-based dehazing, MEAF-PGFM achieves a PSNR of 29.56 dB, an SSIM of 0.9592, and an MAE of 0.0332. Compared with DuRP, MEAF-PGFM improves PSNR by 6.01 dB and SSIM by 0.1193, while reducing MAE from 0.0673 to 0.0332. Furthermore, ablation experiments involving different fusion strategies and polarization features at different resolutions demonstrate the effectiveness of PGFM.

Keywords: polarization image dehazing; multi-depth feature fusion; polarization-guided feature fusion; adaptive fusion

1. Introduction

Haze is a typical atmospheric scattering phenomenon that can significantly degrade the imaging quality of optoelectronic systems. In hazy environments, light from the scene is absorbed and scattered by atmospheric particles during propagation, resulting in reduced contrast, blurred edges, color distortion, and loss of detail. These degradations reduce the discriminability of object features and further compromise the accuracy and reliability of downstream visual tasks, such as object detection, image recognition, and scene understanding. Therefore, image dehazing is essential for reliable visual sensing under complex weather conditions.

Traditional image dehazing methods are generally based on the atmospheric scattering model and estimate the transmission map and atmospheric light using handcrafted priors, such as the dark channel prior and the color attenuation prior. These methods can effectively improve image contrast in certain scenarios. The Dark Channel Prior (DCP) method proposed by He et al. [1] is one of the most representative traditional methods in this field. It estimates scene transmission based on the statistical characteristics of natural

Check for updates

Received: 17 September 2026
Revised: 14 November 2026
Accepted: 19 November 2026
Published: 23 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

images to achieve single-image dehazing. However, handcrafted priors rely on specific scene assumptions; when images contain complex depth variations, non-uniform haze distributions, and illumination changes, the restored results may suffer from color shifts, detail loss, and residual haze in distant regions. To address these limitations, CNN-based image dehazing methods have gradually become mainstream with the development of deep learning. Cai et al. proposed DehazeNet [2], which uses a convolutional neural network to learn the transmission mapping. Li et al. proposed AOD-Net [3], which reformulates the atmospheric scattering model by integrating multiple intermediate parameters into a unified end-to-end framework, thereby reducing the error accumulation caused by conventional stepwise estimation. Subsequently, various strategies, including attention mechanisms, generative adversarial training, contrastive learning, and domain adaptation, have been introduced into dehazing tasks. For example, FFA-Net [4] enhances effective feature representation through feature attention mechanisms, while GAN-based methods [5] improve the visual quality of restored images through adversarial learning. These strategies effectively improve image reconstruction quality, texture detail restoration, and generalization performance in complex scenes.

To further improve the restoration capability of dehazing models in complex hazy environments, extensive studies have been conducted on multi-scale feature extraction and fusion. Conventional multi-scale dehazing methods mainly perform feature modeling based on different receptive fields, spatial resolutions, or frequency components to capture both local textures and large-scale scene information. Huang et al. [6] proposed MAPF-Net, which employs multi-scale large-kernel convolutions to enlarge the receptive field and combines a dual-residual parallel attention mechanism to enhance multi-scale feature representation. MWA-Net [7] extracts image features from different frequency components through wavelet decomposition and improves image restoration by exploiting the complementarity between low-frequency structural information and high-frequency texture information. MCA-GAN [8] integrates multi-scale context attention with a generative adversarial network to enhance the modeling capability for multi-scale scene information in remote sensing images. A Transformer-based multi-scale adaptive dehazing method [9] further incorporates information from different image regions to enhance the network's restoration capability for the overall scene. These methods mainly focus on spatial scales, receptive fields, and frequency components to improve the network's ability to restore objects of different sizes, preserve texture details, and handle regions with non-uniform haze distributions.

Unlike multi-scale feature extraction, multi-depth feature fusion focuses more on the complementary information between features from different network levels. In deep networks, shallow features usually contain rich edge and texture information, while deep features exhibit stronger semantic representation capabilities. Ronneberger et al. [10] proposed U-Net, which uses skip connections between the encoder and decoder to combine shallow spatial details with deep semantic information, thereby providing an effective structure for integrating features from different levels. Subsequently, Zhou et al. [11] proposed U-Net++, which uses nested dense skip connections to further enhance information transfer between features at different network depths. For image dehazing, Wen et al. [12] proposed a multi-depth adaptive fusion dehazing generative network that extracts features at different network depths within a U-Net++-style structure and fuses them via adaptive weighting to coordinate information across different levels, allowing both shallow details and deep semantic information to contribute to image restoration.

In addition to multi-scale feature extraction and multi-depth feature fusion, researchers have introduced attention mechanisms to enhance the network's representation capability for key regions and informative features. MixDehazeNet [13] introduces a multi-scale paral-

lel large-kernel convolution module and an enhanced parallel attention module to improve restoration performance in complex hazy environments by expanding the receptive field and strengthening effective feature responses. FAM-DGAN [14] proposes a frequency-guided feature enhancement module that uses frequency-domain information to highlight structure-related features and suppress noise interference. Dual-Net [15] combines multi-axis feature learning, multi-scale attention mechanisms, and cross-spatial feature interaction strategies to enhance network representation across channel, spatial, and scale dimensions. Dong et al. [16] combine a hybrid self-attention mechanism with a generative adversarial network to improve image restoration quality in complex scenes. DEA-Net [17] integrates detail enhancement strategies with attention mechanisms to improve structural and detail restoration in dehazed images. Chen et al. [18] proposed ED-Former, which integrates frequency-aware hierarchical sampling with an attention-adaptive feed-forward network to enhance image detail restoration while reducing model complexity. Overall, existing RGB image dehazing methods primarily rely on intensity, color, and texture information for image restoration, achieving good performance in improving image contrast and recovering scene details. However, in dense haze, atmospheric scattering causes substantial degradation of scene structures and textures, making restoration based solely on RGB information challenging. Therefore, incorporating auxiliary information that reflects different imaging characteristics offers a promising research direction for image restoration in complex hazy environments.

Polarization imaging provides additional physical information for image restoration in hazy environments. By analyzing variations in polarized light intensity across different polarization directions, polarization information related to atmospheric scattering and target radiance can be obtained, thereby providing additional evidence for clear scene recovery. Early studies mainly relied on polarization-based physical models, Schechner et al. [19] analyzed the influence of atmospheric scattering using polarization imaging characteristics and proposed a method for recovering clear scene information based on polarization information, laying the foundation for polarization-based dehazing research. Shwartz et al. [20] further investigated the application of polarization information and scene constraints in image dehazing and achieved single-image dehazing. Fang et al. [21] considered the polarization characteristics of both target radiance and atmospheric scattered light, refining the simplified assumption regarding the polarization state of target light in traditional polarization dehazing methods. Liang et al. [22] improved image restoration performance in dense haze conditions by analyzing the distribution of polarization angles. Zhang et al. [23] combined the dark channel prior with polarization imaging to improve the estimation of atmospheric light and associated polarization parameters, achieving automated dehazing. Subsequently, Zhang et al. [24] combined multispectral and polarization imaging information to investigate dehazing performance under different observation modes, further expanding the application scope of polarization-based dehazing.

Meanwhile, deep learning-based polarization dehazing methods have also gradually emerged. Zhou et al. [25] proposed Learning to Dehaze with Polarization (LDP), which establishes a more general polarization haze imaging model and integrates deep neural networks for image restoration. This method reduces the dependence of traditional approaches on the polarization state of transmitted target light and specific parameter estimation conditions, promoting the integration of polarization physical models with data-driven methods. Bi et al. [26] proposed an unsupervised dehazing network based on polarization information and deep prior learning, which improves the contrast and visibility of hazy images without requiring paired clear reference images. Suo et al. [27] combined polarization characteristics, self-supervised learning, and multi-scale attention mechanisms to improve transmission estimation accuracy and enhance the brightness

and color consistency of restored images. Wang et al. [28] proposed a polarization dehazing network based on pseudo-3D convolution, which combines spatial feature attention with polarization feature attention to enhance the network's representation capability for different polarization directions and haze regions. Yang et al. [29] proposed DuRP, a dual-stage physics-embedded learning framework that incorporates a generalized polarization imaging model to jointly restore scene radiance and polarization information.However, existing methods mainly focus on polarization imaging models, transmission estimation, and polarization feature extraction and fusion, each emphasizing a particular aspect of polarization cues. In contrast, effectively integrating polarization information with the intermediate features of dehazing networks, thereby allowing it to directly participate in the feature restoration process, remains worthy of further investigation.

In multimodal information fusion, researchers have further investigated methods that dynamically adjust network features using auxiliary information. Perez et al. [30] proposed Feature-wise Linear Modulation (FiLM), which generates scaling and shifting parameters from conditioning information to adaptively adjust network features. Wang et al. [31] proposed Spatial Feature Transform (SFT), which modulates image features using semantic priors to improve texture restoration performance in super-resolution reconstruction. Park et al. [32] proposed Spatially Adaptive Normalization (SPADE), which adjusts internal network features through semantic information to improve the detail representation and structural consistency of generated images. These studies indicate that feature modulation can enable more targeted interaction between different sources of information by adjusting feature responses, providing new insights for utilizing polarization information in image dehazing.

In summary, existing studies have made significant progress in multi-scale feature extraction, multi-depth feature fusion, attention mechanisms, and polarization-based dehazing. Multi-scale and multi-depth feature modeling enhance the restoration of structural information and fine details, while polarization imaging provides additional physical information for image dehazing. Although previous studies have explored various approaches to utilizing polarization information and fusing features, how to effectively integrate polarization information with image features at different levels and further exploit its auxiliary role in image restoration still requires further investigation. To address this issue, this paper proposes a Multi-depth Enhanced Adaptive Fusion Polarization-guided Network (MEAF-P-Net). The proposed network adopts MEAF as the dehazing backbone and employs a Multi-depth Feature Extraction Module (MFEM) and an Adaptive Weight Fusion Module (AWFM) to extract and adaptively fuse image features from different levels. Furthermore, a Polarization-Guided Fusion Module (PGFM) is introduced to integrate multi-scale polarization features with intermediate backbone features through feature concatenation and residual fusion, facilitating effective fusion between polarization information and image features.

The main contributions of this work are summarized as follows:

1. A Multi-depth Enhanced Adaptive Fusion Polarization-guided Network (MEAF-P-Net) is proposed. Based on MEAF, the network integrates multi-depth feature extraction and adaptive fusion to effectively utilize image features from different semantic levels, thereby providing a network framework for polarization-assisted image restoration.
2. A Polarization-Guided Fusion Module (PGFM) is designed. This module extracts high- and low-resolution features from polarization information and introduces them into different feature levels of the backbone network through feature concatenation and residual fusion, enabling effective fusion between polarization information and image features.

3. Experiments are conducted on a synthetic RGB dehazing dataset and the PolarLITIS polarization dehazing dataset. MEAF-S_0, MEAF-Concat, and MEAF-PGFM are compared under the same backbone framework to analyze the effects of different polarization information utilization strategies on image restoration performance.
4. Polarization-guided ablation experiments are performed to verify the effectiveness of PGFM in utilizing polarization information during image restoration. Experiments under different haze densities further indicate that polarization guidance yields more pronounced performance gains under medium and high haze density conditions.

2. Method

The proposed MEAF-P-Net is designed for polarization-guided image dehazing in complex hazy environments. It adopts a multi-depth encoder–decoder architecture, enabling the interaction, enhancement, and adaptive fusion of features from different network depths to jointly restore global image structures and local details.

MEAF-P-Net mainly consists of a Multi-depth Feature Extraction Module (MFEM), an Adaptive Weight Fusion Module (AWFM), and a Polarization-Guided Fusion Module (PGFM). MFEM extracts features at different semantic levels, AWFM adaptively adjusts fusion weights according to the responses of features from different depths, and PGFM introduces polarization information into the restoration process through feature-level fusion with the backbone features.

Through the collaboration of these modules, MEAF-P-Net further incorporates polarization guidance into multi-depth feature restoration, allowing polarization information to participate in image restoration in the feature space rather than being simply concatenated with the main image input. The remainder of this section is organized as follows. We first introduce the overall network architecture, then describe each component module and its implementation in detail, and finally present the loss functions used for network training.

2.1. Overall Network Architecture

MEAF-P-Net adopts a conditional generative adversarial network framework consisting of a generator and a discriminator. The overall architecture is shown in Figure 1. The generator takes the hazy total-intensity component S_0^h as the main restoration input and the corresponding polarization component S_2^h as auxiliary guidance, and outputs the restored clear total-intensity component $\hat{S}_0^c$. The mapping is expressed as

$$\hat{S}_0^c = G\left(S_0^h, S_2^h\right). \tag{1}$$

Here, $G(\cdot)$ denotes the generator, which consists of MEAF and PGFM. MEAF performs visual feature extraction and adaptive fusion, while PGFM integrates polarization information with the backbone features, enabling effective interaction between polarization information and the restoration process. Through the collaboration of these modules, the generator can fully exploit image content features and polarization information, improving image restoration in complex hazy environments.

During backbone feature extraction, the hazy total-intensity component S_0^h is taken as input to extract multi-level features for image restoration. Meanwhile, S_2^h is processed by an independent polarization branch to extract high- and low-resolution polarization features, which are then concatenated with the backbone features at the corresponding resolution levels and fused through residual connections via PGFM. In this way, auxiliary polarization information participates in the image restoration process.

The multi-depth features produced by MFEM are adaptively fused by AWFM, and the restored clear total-intensity component is generated through the output layer. During

training, a conditional PatchGAN discriminator imposes adversarial constraints on the generated results. The discriminator takes the hazy total-intensity component as conditional information and performs local discrimination between real and generated samples.

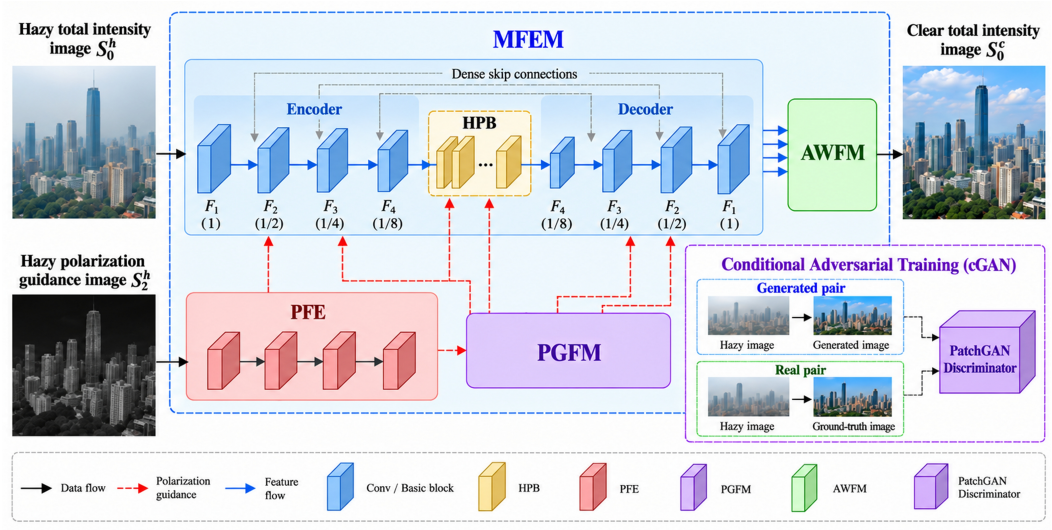

Figure 1. Overall architecture of the proposed MEAF-P-Net.

2.2. Multi-Depth Feature Extraction Module

To fully utilize image structure and texture information at different network depths, MFEM is constructed as the main feature extraction structure of the MEAF-P-Net generator. It adopts a U-Net++-based multi-level encoder–decoder architecture with dense skip connections to facilitate information exchange among features from different network depths.

During encoding, the input features are progressively downsampled to form representations at different spatial scales. Shallow features retain relatively high spatial resolution and preserve image edges, textures, and local intensity variations. As the network depth increases, the receptive field expands, allowing the network to capture broader scene structures and contextual information. Therefore, features from different depths are highly complementary for image restoration.

During decoding, feature resolution is progressively restored through upsampling. Dense skip connections combine encoder features with intermediate decoder features at the same scale, enabling information exchange across different network depths. Compared with the conventional encoder–decoder skip connections in U-Net, this structure enhances information interaction among features at different levels and allows shallow details and deep contextual information to be more fully integrated during decoding.

To further enhance the local representation capability of features at each level, a Haze Perception Block (HPB) is introduced into the feature mapping process of MFEM. HPB consists of a 3×3 depthwise separable convolution, a 3×3 standard convolution, and a residual connection. Local spatial features are first extracted by the depthwise separable convolution and then further transformed by the standard convolution. The transformed features are combined with the input features through a residual connection:

$$F_{\mathrm{HPB}} = F + \mathcal{H}_2(\mathcal{H}_1(F)). \tag{2}$$

Here, $\mathcal{H}_1(\cdot)$ denotes a feature transformation composed of depthwise separable convolution, batch normalization, and ReLU activation, while $\mathcal{H}_2(\cdot)$ denotes a feature mapping composed of standard convolution, batch normalization, and ReLU activation. The residual connection preserves the original feature information while enabling HPB to extract local

texture and structural features, thereby enhancing the feature representation capability at each level of MFEM. In MFEM, each feature mapping node consists of two sequentially connected HPBs. When the input and output channel dimensions differ, a 3×3 convolution is first applied for channel mapping.

Let $X^{i,j}$ denote the feature at the i-th scale and the j-th decoding level. The feature propagation process is expressed as

$$X^{i,j} = \mathcal{H}_{i,j}\left(\left[X^{i,0}, X^{i,1}, \ldots, X^{i,j-1}, \mathcal{U}\left(X^{i+1,j-1}\right)\right]\right). \tag{3}$$

Here, $\mathcal{H}_{i,j}(\cdot)$ denotes the feature mapping operation at the current node, $\mathcal{U}(\cdot)$ denotes upsampling, and $[\cdot]$ denotes feature concatenation along the channel dimension. These dense connections allow information from different network depths to participate in subsequent decoding stages, producing restoration features with different effective receptive fields and feature levels.

MFEM produces restoration features from different network depths, as shown in Figure 2. These features are subsequently fused by AWFM for image reconstruction, allowing information from different depths to contribute to image restoration.

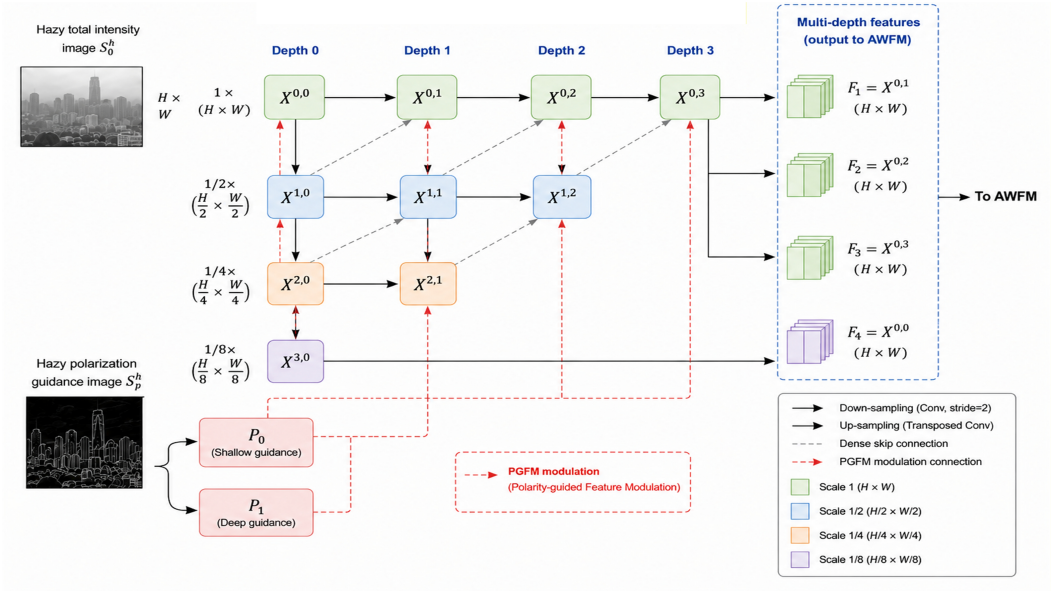

Figure 2. Structure of the Multi-depth Feature Extraction Module.

2.3. Adaptive Weight Fusion Module

Features produced by MFEM at different depths represent information at different levels and emphasize different aspects of the scene. Shallow features retain more local details, whereas deeper features provide broader contextual information. To effectively select and fuse features from different depths, AWFM is designed to adaptively generate fusion weights according to the global responses of these features, the flowchart of AWFM is shown in Figure 3. Let the N features produced by MFEM be

$$\{F_1, F_2, \ldots, F_N\}. \tag{4}$$

Here, $F_n \in \mathbb{R}^{C \times H \times W}$ denotes the feature from the n-th depth branch. Global average pooling is first applied to each feature to obtain its global descriptor:

$$g_n = \mathrm{GAP}(F_n) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} F_n(i,j). \tag{5}$$

Here, $g_n \in \mathbb{R}^{C\times1\times1}$ denotes the global response of the n-th feature branch. The descriptors from all branches are concatenated along the channel dimension:

$$g = [g_1, g_2, \ldots, g_N]. \tag{6}$$

The concatenated descriptor g contains global information from all depth branches. Two 1×1 convolutions with a ReLU activation between them are used to model the joint responses among features from different depths, and a Softmax function generates the corresponding fusion weights:

$$W = \text{Softmax}(\phi_2(\delta(\phi_1(g)))). \tag{7}$$

Here, $\phi_1(\cdot)$ and $\phi_2(\cdot)$ denote two 1×1 convolution mappings, and $\delta(\cdot)$ denotes the ReLU activation function. $W = [w_1, w_2, \ldots, w_N]$ represents the adaptive fusion weight vector, where w_n denotes the weight assigned to the n-th depth feature F_n. The Softmax function normalizes the fusion weights across different depth branches, ensuring that all weights are non-negative and sum to one:

$$w_n \geq 0, \qquad \sum_{n=1}^{N} w_n = 1. \tag{8}$$

The fused feature is obtained by weighted summation:

$$F_{\text{fuse}} = \sum_{n=1}^{N} w_n F_n. \tag{9}$$

In this way, AWFM adaptively determines the contribution of each branch according to the joint global responses of all depth features, so that restoration features containing different levels of structural and detail information are adaptively fused, and subsequently passed through the feature enhancement module and output layer to generate the final image restoration result.

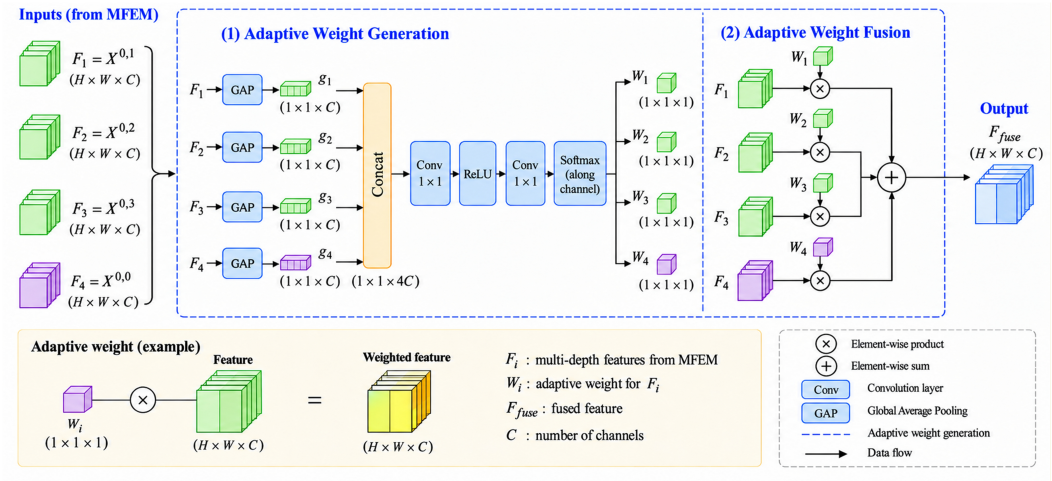

Figure 3. Structure of the Adaptive Weight Fusion Module.

2.4. Polarization-Guided Fusion Module

In MEAF-P-Net, a Polarization-Guided Fusion Module (PGFM) is designed to effectively utilize auxiliary polarization information and improve the image restoration capability of the network, as shown in Figure 4. PGFM extracts polarization features at different resolutions from S_2^h through a polarization feature encoder and introduces them into

the corresponding feature levels of the backbone network through feature concatenation and residual fusion.

The polarization feature encoder takes S_2^h as input and extracts high- and low-resolution polarization features through two consecutive convolutional stages, denoted as P_0 and P_1, respectively:

$$P_0 = E_0(S_2^h). \tag{10}$$

$$P_1 = E_1(P_0). \tag{11}$$

Here, E_0 and E_1 denote the two convolutional encoding stages. E_0 consists of two 3×3 convolutional layers, each with 32 output channels and a stride of 1, producing the high-resolution polarization feature P_0 with the same spatial resolution as the input. Similarly, E_1 consists of two 3×3 convolutional layers, each with 64 output channels. The first layer has a stride of 2, while the second has a stride of 1, producing the low-resolution polarization feature P_1 with half the spatial resolution of P_0. Each convolutional layer is followed by a ReLU activation function.

PGFM concatenates P_0 with the backbone features F at the high-resolution shallow encoding nodes and concatenates P_1 with the backbone features F at the low-resolution encoding nodes, along the channel dimension. A 1×1 convolution is then used to perform feature fusion. To preserve the original backbone features while incorporating auxiliary polarization information, a residual connection is employed to introduce the fused features into the backbone network. This process is expressed as

$$F' = F + \alpha \, \mathrm{Conv}_{1\times1}([F; P]). \tag{12}$$

Here, α is a learnable coefficient that controls the contribution of the fused features to the backbone features, with an initial value of 0.1. The weights and biases of the fusion convolutional layer are initialized to zero, allowing the network to gradually learn the fusion relationship between auxiliary polarization features and backbone features during training.

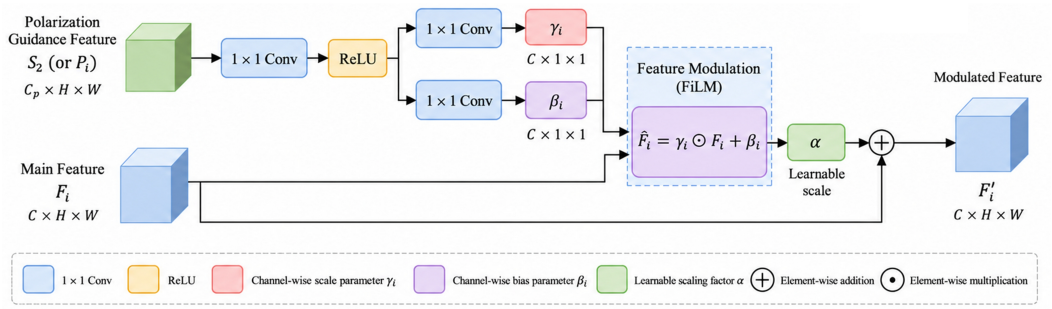

Figure 4. Structure of the PGFM.

2.5. Loss Functions

2.5.1. Pixel Reconstruction Loss

To constrain the differences between the generated result and the clear target in the pixel space, the L_1 loss is employed as the basic reconstruction constraint. Let $\hat{S}_0^c$ denote the predicted clear total-intensity component and S_0^c denote the corresponding ground-truth clear target. The pixel reconstruction loss is defined as

$$\mathcal{L}_1 = \left\| \hat{S}_0^c - S_0^c \right\|_1. \tag{13}$$

The L_1 loss directly constrains the pixel-wise differences between the predicted image and the ground-truth target, providing basic reconstruction supervision while helping preserve local structural information.

2.5.2. Perceptual Loss

The L_1 loss alone cannot fully characterize the differences in texture details between generated and ground-truth images. Therefore, a perceptual loss based on a pre-trained VGG16 network is further introduced. Features of the generated and ground-truth images are extracted at different levels of VGG16, and their L_1 distances are calculated as

$$\mathcal{L}_{\text{per}} = \sum_{l \in \mathcal{V}} \left\| \phi_l(\hat{S}_0^c) - \phi_l(S_0^c) \right\|_1. \tag{14}$$

Here, $\phi_l(\cdot)$ denotes the feature mapping at the lth level of VGG16, and $\mathcal{V}$ denotes the set of feature layers used to calculate the perceptual loss. Four feature layers, namely relu1_2, relu2_2, relu3_3, and relu4_3, are used to constrain image feature representations at different levels.

2.5.3. Adversarial Loss

To further improve the local structural quality of the generated results, WGAN-GP is adopted for adversarial training of the generator and discriminator. Following the conditional discrimination scheme, the discriminator takes the hazy total-intensity component S_0^h as conditional information and assigns local realism scores to the generated results and ground-truth targets.

The adversarial loss for the generator is defined as

$$\mathcal{L}_{\text{adv}}^G = -\mathbb{E}\left[D\left([S_0^h, \hat{S}_0^c] \right) \right]. \tag{15}$$

Here, $D(\cdot)$ denotes the PatchGAN discriminator.

For the gradient penalty, random linear interpolation is performed between the ground-truth clear image and the generated image:

$$\tilde{S}_0 = \epsilon S_0^c + (1-\epsilon)\hat{S}_0^c, \qquad \epsilon \sim U(0,1). \tag{16}$$

The gradient penalty is defined as

$$\mathcal{L}_{\text{gp}} = \mathbb{E}_{\tilde{S}_0}\left[\left(\left\| \nabla_{\tilde{S}_0} D\left([S_0^h, \tilde{S}_0] \right) \right\|_2 - 1 \right)^2 \right]. \tag{17}$$

Accordingly, the discriminator objective is expressed as

$$\mathcal{L}_D = \mathbb{E}\left[D\left([S_0^h, \hat{S}_0^c] \right) \right] - \mathbb{E}\left[D\left([S_0^h, S_0^c] \right) \right] + \lambda_{\text{gp}} \mathcal{L}_{\text{gp}}. \tag{18}$$

Here, the gradient penalty weight is set to $\lambda_{\text{gp}} = 10.0$. By incorporating the gradient penalty, the gradient norm of the discriminator is constrained in the interpolation region between real and generated samples, thereby improving the stability of adversarial training.

2.5.4. Generator Loss

Finally, the overall generator objective is formulated as a weighted combination of the pixel reconstruction loss, perceptual loss, and adversarial loss:

$$\mathcal{L}_G = \lambda_1 \mathcal{L}_1 + \lambda_{\text{per}} \mathcal{L}_{\text{per}} + \lambda_{\text{adv}} \mathcal{L}_{\text{adv}}^G. \tag{19}$$

The loss weights used during training are set to

$$\lambda_1 = 1.0, \qquad \lambda_{\text{per}} = 0.1, \qquad \lambda_{\text{adv}} = 0.01. \tag{20}$$

Through the joint optimization of these loss terms, the generator is encouraged to achieve accurate pixel-level reconstruction while maintaining feature consistency and satisfying the adversarial learning objective.

3. Experimental Results and Analysis

3.1. Experimental Settings

Datasets. To separately evaluate the dehazing capability of the MEAF backbone and the effectiveness of the polarization guidance mechanism, experiments are conducted on two datasets.

The first dataset is a synthetic visible-light dehazing dataset based on Cityscapes [33], which is used to evaluate the restoration performance of the MEAF backbone in visible-light image dehazing. Cityscapes urban street images are used as clear references. Following the synthesis procedure of Zhou et al. [25], hazy images and the corresponding polarization components are generated using the atmospheric scattering model. The dataset contains 2975 training image pairs and 500 validation image pairs. During training, images are randomly cropped to 240×240, while validation images are cropped to 240×496.

The second dataset is a synthetic polarization dehazing dataset constructed from real PolarLITIS [34] polarization data, which is used to evaluate the effectiveness of the PGFM module. PolarLITIS contains polarization images of real road scenes acquired using a monochrome polarization camera. Using the clear polarization images in this dataset, hazy observations of the total-intensity component S_0 and the second linear Stokes component S_2 are synthesized with the atmospheric scattering model, forming paired hazy and clear samples. Here, $S_0 = I_0 + I_{90}$ represents the total intensity of the scene, while $S_2 = 2I_{45} - S_0$ represents the intensity difference between the 45° and 135° linear polarization components. Compared with using S_0 alone, S_2 provides additional polarization information that can help the network distinguish scene structures from scattering-induced degradation. PGFM incorporates this information into the feature restoration process to improve the utilization of polarization observations. Four haze densities, $\beta \in \{0.6, 1.0, 1.4, 1.8\}$, are considered to simulate different degradation levels ranging from light to dense haze. The transmission model used to control the degree of haze degradation is defined as

$$t(x) = \exp[-\beta d_{\text{norm}}(x)]. \tag{21}$$

Here, $d_{\text{norm}}(x)$ denotes the normalized depth obtained using a monocular depth estimation network. In the conventional atmospheric scattering model, the transmission is determined by the atmospheric extinction coefficient and the actual propagation distance. The extinction coefficient, measured in m^{-1}, incorporates contributions from scattering and absorption and is related to meteorological visibility. Since the normalized depth employed in this study is not calibrated to physical distances, β is treated as a dimensionless parameter controlling the synthetic haze density.

The dataset is divided by scene into training, validation, and testing sets containing 1640, 420, and 508 scenes, respectively. Each scene contains paired hazy and clear samples corresponding to the four haze densities, with an original resolution of 496×496. To reduce computational cost, 410, 105, and 105 scenes are sampled from the training, validation, and test sets, respectively. During training, images are uniformly cropped to 240×496.

Implementation Details. MEAF-P-Net is implemented using the PyTorch (version 2.2.1) framework and trained on an NVIDIA GeForce RTX 3090 GPU. Both the generator

and discriminator are optimized using Adam, with initial learning rates of 1×10^{-4} and 4×10^{-4}, respectively. The Adam exponential decay rates β_1 and β_2 are set to 0.5 and 0.999, respectively. A cosine annealing learning rate schedule is employed. The network is trained for 100 epochs with a batch size of 2. The parameters of VGG16 are kept fixed during training.

3.2. Comparison on Visible-Light Image Dehazing

The MEAF backbone is first evaluated on the synthetic Cityscapes-based visible-light dehazing dataset to assess its feature extraction and image restoration capability in standard RGB dehazing tasks. Several representative methods are selected for comparison, including the prior-based DCP [1], DehazeNet [2], AOD-Net [3], FFA-Net [4], GridDehazeNet [35], LDP [25], EPDN [5], which is based on a generative adversarial network, and more recent representative methods, including DEA-Net [17] and ED-Former [18], are included for comparison. All methods are evaluated using the same dataset and evaluation metrics, and are compared with the proposed MEAF to analyze their restoration performance on the synthetic visible-light dehazing task.

As shown in Table 1, MEAF achieves a PSNR of 31.40 dB and an SSIM of 0.933 on this dataset. It outperforms FFA-Net and EPDN in both metrics and achieves a higher PSNR than GridDehazeNet (30.75 dB), although its SSIM is slightly lower (0.933 versus 0.935). Its performance is close to that of DEA-Net (31.50 dB, 0.934) but lower than that of ED-Former (33.29 dB, 0.9531). Among these methods, DehazeNet and DCP, which are based on early deep-learning techniques and hand-crafted priors, respectively, exhibit relatively lower restoration performance.

Table 1. Quantitative comparison of different dehazing methods on the synthetic visible-light dehazing dataset.

Method	PSNR (dB)	SSIM
DehazeNet	15.46	0.615
DCP	16.33	0.654
AOD-Net	21.45	0.761
EPDN	24.95	0.891
LDP	27.81	0.913
FFA-Net	28.74	0.912
GridDehazeNet	30.75	0.935
DEA-Net	31.50	0.934
ED-Former	33.29	0.953
MEAF	31.40	0.933

The visual results show that both MEAF and ED-Former can effectively recover the structures of buildings, roads, and vehicles in typical street scenes, achieving broadly comparable visual quality. However, in scenes with complex distant structures, both methods exhibit varying degrees of distortion in the restored images. Specifically, ED-Former exhibits local blurring and structural distortions, whereas MEAF produces some texture artifacts. Overall, the MEAF backbone demonstrates competitive performance in standard visible-light dehazing, but there is still room for improvement in detail restoration in complex scenes. These results provide a foundation for the subsequent introduction of polarization guidance.

To evaluate computational efficiency, FFA-Net, DEA-Net, and ED-Former are selected for comparison with MEAF, and the results are reported in Table 2. Under the same input resolution and testing environment, MEAF has 14.41 M parameters, a computational cost of 238.30 Gops, and an average inference time of 50.21 ms. Compared with FFA-Net, MEAF

requires fewer computational operations and achieves a shorter inference time. However, both its computational cost and inference time remain higher than those of DEA-Net and ED-Former, indicating that the proposed network still has room for further optimization in computational efficiency.

Table 2. Computational efficiency comparison of different dehazing methods.

Method	Params (M)	Computation (Gops)	Inference Time (ms)
FFA-Net	4.4559	522.2771	87.462
DEA-Net	3.6534	58.5439	16.661
ED-Former	0.8659	13.3423	25.818
MEAF	14.4141	238.2971	50.209

3.3. Comparison with Polarization Dehazing Methods

To evaluate the effectiveness of the proposed method for polarization-based image dehazing, we compare MEAF-PGFM with two existing polarization dehazing methods, LDP [25] and DuRP [29], on the synthetic polarization dehazing dataset constructed from PolarLITIS. The results are reported in Table 3. LDP achieves an overall MAE of 0.1446, a PSNR of 16.97 dB, and an SSIM of 0.6750, whereas DuRP achieves an overall MAE of 0.0673, a PSNR of 23.55 dB, and an SSIM of 0.8399. In comparison, MEAF-PGFM achieves an overall MAE of 0.0332, a PSNR of 29.56 dB, and an SSIM of 0.9592. These results show that MEAF-PGFM improves both image reconstruction accuracy and structural similarity, demonstrating the effectiveness of the polarization feature guidance mechanism for image dehazing.

Table 3. Quantitative comparison of polarization dehazing methods on the PolarLITIS test set.

Method	$\beta = 0.6$	$\beta = 1.0$	$\beta = 1.4$	$\beta = 1.8$	Overall MAE	PSNR (dB)	SSIM
LDP	0.0976	0.1350	0.1626	0.1831	0.1446	16.97	0.6750
DuRP	0.0576	0.0614	0.0657	0.0845	0.0673	23.55	0.8399
MEAF-PGFM	0.0297	0.0235	0.0333	0.0465	0.0332 ± 0.0046	29.56	0.9592

A possible explanation is that the training and test sets of PolarLITIS are collected from different scenes, resulting in a certain distribution shift in their overall imaging conditions. In particular, the training set contains relatively few high-brightness scenes, whereas the test set includes a large number of such scenes. Since scene brightness may reflect variations in imaging conditions across different real-world acquisition scenarios, we further divide the training and test sets into groups according to the mean brightness of the ground-truth (GT) images and analyze the restoration performance of LDP across different brightness intervals.

The results for LDP reveal a substantial performance difference between the PolarLITIS training and test sets. Its MAE is 0.008 on the training set but increases to 0.1446 on the test set. As shown in Table 4, LDP achieves high restoration quality across all brightness intervals in the training set, with PSNR values ranging from 37.90 to 40.11 dB. However, its restoration performance decreases considerably in the corresponding brightness intervals of the test set, with PSNR gaps ranging from 15.16 to 22.33 dB. These results indicate substantial performance degradation when LDP is applied to different real-world acquisition scenarios, suggesting certain limitations in its cross-scene generalization capability.

Table 4. PSNR comparison of LDP on the PolarLITIS training and test sets across different scene brightness intervals.

GT Mean Brightness	Train Scenes	Train PSNR (dB)	Test Scenes	Test PSNR (dB)	PSNR Gap (dB)
[0.0, 0.2)	269	40.11	52	24.95	15.16
[0.2, 0.3)	110	38.76	48	22.17	16.59
[0.3, 0.4)	25	38.17	56	18.34	19.83
[0.4, 0.5)	6	37.90	352	15.57	22.33

3.4. Dehazing Experiments on Real Hazy Scenes

To further evaluate the dehazing capability of the proposed method in real hazy scenes, qualitative experiments are conducted using real hazy polarization images, as shown in Figure 5. The results show that both MEAF-S_0 and MEAF-PGFM can improve image contrast to some extent and mitigate haze-induced obscuration of scene structures. Compared with MEAF-S_0, MEAF-PGFM incorporates polarization information for feature guidance, yielding clearer vehicle contours and more distinguishable structural details in certain foreground regions. These observations indicate that polarization information can assist image restoration in real hazy scenes.

However, residual haze persists in densely hazy regions, indicating that image restoration under complex scattering conditions remains challenging. Since strictly paired haze-free reference images are unavailable for these real scenes, the results are evaluated qualitatively through visual comparison rather than with full-reference metrics such as PSNR and SSIM. These experiments provide a preliminary validation of the applicability of the proposed method under real hazy conditions, although its generalization capability in more complex scenes requires further investigation.

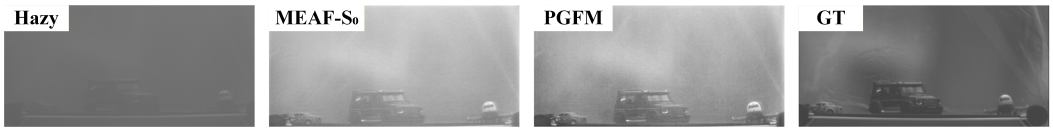

Figure 5. Dehazing results on real hazy scenes.

3.5. Ablation Experiments

This section conducts ablation experiments on the MEAF backbone to analyze the effects of different architectural components on image restoration performance.

3.5.1. Visible-Light Backbone Ablation

Module ablation experiments are conducted on the synthetic visible-light dehazing dataset constructed from Cityscapes. A U-Net++ generative adversarial network is adopted as the baseline architecture (Base). On this basis, the Haze Perception Block (HPB) and the Adaptive Weight Fusion Module (AWFM) are introduced sequentially to analyze their effects on the restoration performance of the backbone network. All experiments follow the same data split, training strategy, and evaluation protocol, with the random seed, batch size, and number of training epochs fixed to reduce the influence of training randomness on experimental comparisons. The results are presented in Table 5.

Table 5. Ablation results of the backbone network modules.

Experimental Setting	PSNR (dB)	SSIM
Base	31.137	0.9277
Base + HPB	31.258	0.9267
Base + HPB + AWFM	31.457	0.9287

As shown in Table 5, the baseline network achieves a PSNR of 31.137 dB and an SSIM of 0.9277. After introducing HPB, PSNR increases to 31.258 dB, an improvement of 0.121 dB over the baseline, while SSIM remains at a comparable level. With the further addition of AWFM, PSNR increases to 31.457 dB, an additional gain of 0.199 dB over the previous configuration, and SSIM reaches 0.9287. Compared with the baseline, the final backbone improves PSNR by approximately 0.32 dB in total. These results demonstrate that incorporating HPB and AWFM into the U-Net++ restoration framework can further improve image restoration performance. In particular, AWFM adaptively weights restoration features from different depths, helping to improve the utilization efficiency of multi-level feature information.

3.5.2. Ablation Study on Polarization Feature Fusion

In this section, ablation experiments are conducted on the polarization feature fusion module in MEAF-P-Net. With the backbone architecture and training settings kept consistent, different polarization feature fusion strategies and fusion configurations at different depths are compared to analyze their effects on image restoration performance and to verify the effectiveness of the proposed fusion module and multi-depth fusion design.

First, to investigate the effect of polarization information on image restoration performance, MEAF-S_0, which uses only the hazy total-intensity component S_0^h, is adopted as the baseline and compared with MEAF-PGFM, which incorporates the polarization-guided fusion module. Meanwhile, MEAF-Concat, which directly concatenates polarization information at the network input, is included as an additional comparison to examine the effects of different strategies for incorporating polarization information.

Furthermore, three ablation configurations, namely PGFM-Gamma, PGFM-Delta, and PGFM-Affine, are constructed to compare image restoration performance under different feature modulation strategies. Specifically, Gamma denotes multiplicative modulation, which adaptively adjusts the response magnitudes of backbone features at different spatial locations and channels using scaling coefficients generated from polarization features. Delta denotes additive modulation, which supplements and corrects backbone features through polarization-derived offsets. PGFM-Affine combines multiplicative and additive modulation to adjust both feature response magnitudes and offsets, allowing the effect of jointly applying these two modulation strategies to be investigated.

In addition, PGFM-P0 and PGFM-P1 are constructed to analyze the roles of polarization features at different resolutions during feature fusion. PGFM-P0 retains only the fusion of high-resolution polarization features, whereas PGFM-P1 retains only the fusion of low-resolution polarization features. These configurations are used to evaluate the effects of polarization features at different resolutions on image restoration performance.

Moreover, to assess the influence of random initialization on the restoration performance of the model, independent training runs are conducted using random seeds 0, 1, and 2. With the data split and training parameters kept consistent, variations in model performance are analyzed by comparing the MAE values obtained under different random seeds, together with their mean and standard deviation. The complete ablation results are presented in Table 6.

As shown in Table 6, different polarization feature fusion strategies exhibit variations in image restoration performance. PGFM-Gamma, PGFM-Delta, and PGFM-Affine achieve overall MAE values of 0.0469, 0.0446, and 0.0466, respectively. In comparison, the proposed PGFM employs feature concatenation and residual fusion, reducing the overall MAE to 0.0283, which represents a reduction of approximately 39.3% compared with PGFM-Affine. This result indicates that residual feature fusion can achieve better overall restoration performance.

Table 6. Quantitative results of the polarization feature fusion ablation experiments.

Configuration	Seed	$\beta = 0.6$ MAE	$\beta = 1.0$ MAE	$\beta = 1.4$ MAE	$\beta = 1.8$ MAE	Overall MAE	PSNR (dB)	SSIM
MEAF-S_0	0–2	0.0240	0.0333	0.0501	0.0755	0.0457 ± 0.0025	27.57	0.9448
MEAF-Concat	–	0.0338	0.0352	0.0502	0.0677	0.0467	27.17	0.9429
PGFM-Gamma	0	0.0506	0.0522	0.0441	0.0405	0.0469	26.58	0.9494
PGFM-Delta	0	0.0512	0.0365	0.0386	0.0522	0.0446	26.97	0.9485
PGFM-Affine	0	0.0518	0.0452	0.0422	0.0472	0.0466	26.81	0.9490
PGFM-P0	0	0.0150	0.0284	0.0383	0.0450	0.0317	29.92	0.9592
PGFM-P1	0	0.0149	0.0232	0.0346	0.0466	0.0298	30.19	0.9582
PGFM	0	0.0214	0.0228	0.0331	0.0361	0.0283	30.48	0.9639
PGFM	1	0.0472	0.0260	0.0302	0.0463	0.0374	28.51	0.9559
PGFM	2	0.0204	0.0217	0.0366	0.0572	0.0340	29.68	0.9577
PGFM (Mean ± Std.)	0–2	0.0297	0.0235	0.0333	0.0465	0.0332 ± 0.0046	29.56	0.9592

Further comparisons are conducted to evaluate the effects of fusing polarization features at different resolutions. PGFM-P0 and PGFM-P1 achieve overall MAE values of 0.0317 and 0.0298, respectively. By jointly fusing high- and low-resolution polarization features, the complete PGFM achieves the lowest overall MAE of 0.0283, together with a PSNR of 30.48 dB and an SSIM of 0.9639, outperforming both single-resolution fusion configurations. These results indicate that jointly incorporating polarization features at different resolutions contributes to improved overall image restoration performance.

In addition, under random seeds 0, 1, and 2, PGFM achieves overall MAE values of 0.0283, 0.0374, and 0.0340, respectively. The mean MAE is 0.0332, with a standard deviation of 0.0046, while the mean PSNR and SSIM are 29.56 dB and 0.9592, respectively. These performance variations across different random seeds indicate that network initialization, training sample order, and random cropping may influence the optimization process, resulting in certain fluctuations in restoration performance.

Finally, the model parameter counts and computational costs of different polarization fusion configurations are compared, as shown in Table 7. Compared with MEAF-S_0, MEAF-PGFM introduces approximately 0.17 million additional parameters, corresponding to an increase of about 1.2%. Its computational cost increases by approximately 18.6 GMACs, corresponding to an increase of about 7.3%. In contrast, MEAF-Concat has nearly the same parameter count and computational cost as MEAF-S_0. Combined with the preceding restoration results, these comparisons indicate that PGFM can effectively improve polarization-assisted image dehazing performance with relatively small increases in model parameters and computational cost.

Table 7. Comparison of model parameters and computational costs.

Method	Parameters (M)	GMACs
MEAF-S_0	14.41	255.86
MEAF-Concat	14.41	255.92
MEAF-PGFM	14.58	274.48

3.6. Visual Results

Representative scenes from the synthetic visible-light dehazing and polarization dehazing datasets are selected for qualitative comparison.

3.6.1. Visible-Light Dehazing Results

Figure 6 presents a visual comparison of different methods on the synthetic visible-light dehazing dataset using two representative urban street scenes. For each scene, the figure includes the hazy input image, the restoration results obtained by DCP, AOD-Net,

DehazeNet, EPDN, LDP, FFA-Net, GridDehazeNet, DEA-Net, ED-Former and MEAF, as well as the clear reference image (GT).

The compared methods exhibit noticeable differences in haze removal and detail restoration. DCP enhances scene contrast but introduces color deviations and over-enhancement in certain regions. DehazeNet exhibits limited haze removal and detail recovery. AOD-Net and EPDN reduce haze to some extent, but blurring remains in distant regions and fine textures. LDP removes much of the haze, although some image details remain insufficiently recovered.

FFA-Net, GridDehazeNet, DEA-Net, and ED-Former achieve better restoration of overall scene structure and brightness. Among them, ED-Former demonstrates good detail restoration performance. MEAF effectively removes most of the haze in both representative scenes while preserving scene structures, brightness, and local textures. Its visual results are broadly comparable to those of ED-Former and show good consistency with the clear reference images. These observations are generally consistent with the quantitative comparison in Table 1, indicating that the MEAF backbone exhibits strong image restoration capability in standard visible-light dehazing tasks.



Figure 6. Visual comparison of dehazing results obtained by different methods on the synthetic visible-light dehazing dataset.

3.6.2. Polarization Dehazing Results

Figure 7 presents a visual comparison of different polarization fusion strategies on the polarization dehazing dataset. From left to right, the columns display the hazy total-intensity image S_0^h, the polarization guidance component S_2^h, the restoration results obtained by MEAF-S_0, MEAF-Concat, PGFM-Affine, DuRP, PGFM-Delta, PGFM-Gamma, PGFM-P0, PGFM-P1, MEAF-PGFM, and LDP, followed by the clear reference image S_0^c.

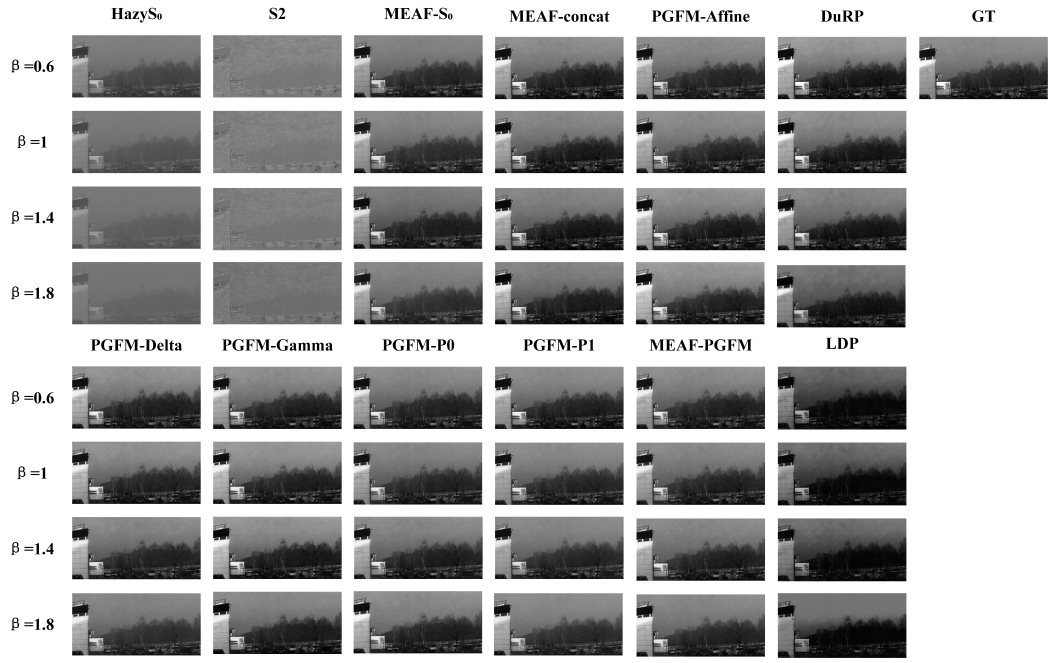

Figure 7. Visual comparison of different polarization fusion strategies on the polarization dehazing dataset.

As the haze density β increases, the scene structures in the hazy total-intensity image S_0^h become increasingly affected by haze, resulting in a noticeable reduction in contrast and texture information in distant regions. It can be observed that all dehazing methods recover tree contours, vehicle boundaries, and road details to varying degrees under different haze densities. Among these methods, MEAF-PGFM consistently maintains effective detail recovery and good overall visual quality across different haze densities. Particularly at $\beta = 1.4$ and $\beta = 1.8$, it achieves better restoration of scene structures and texture details, producing visual results that more closely resemble the ground-truth image S_0^c. Consistent with the quantitative results reported in Table 6, MEAF-PGFM outperforms the other ablation configurations in terms of average MAE, PSNR, and SSIM, demonstrating that PGFM can more effectively exploit polarization information to improve image restoration performance under complex haze conditions.

4. Discussion

The experimental results demonstrate that multi-depth feature fusion and the incorporation of polarization information can improve image restoration performance in different ways. In this study, PSNR, SSIM, and MAE are used to quantitatively evaluate image restoration quality; these metrics reflect reconstruction accuracy, structural similarity, and pixel-wise error, respectively. However, they cannot directly characterize the physical fidelity of the recovered information. For the MEAF backbone, the ablation results of HPB and AWFM indicate that effectively utilizing restoration features from different network depths improves image reconstruction quality. Multi-depth feature fusion integrates local details and scene structural information contained in features at different depths, thereby improving the overall quality of restored images.

In polarization-assisted dehazing, the experimental results further indicate that both the fusion strategy and the feature levels at which polarization information is introduced affect the final restoration performance. Comparison of different fusion strategies shows that polarization information can provide effective guidance for image restoration, but an appropriate fusion strategy is necessary to fully exploit its benefits. The results of PGFM-P0 and PGFM-P1 show that polarization features at a single resolution can provide supplementary information, whereas the complete MEAF-PGFM jointly incorporates polarization features at different resolutions and achieves better results on the overall evaluation metrics. This suggests that polarization features at different resolutions play complementary roles in image restoration, and fusing them jointly further improves overall dehazing performance.

The experimental results further show that the benefit of polarization information varies with haze density. Since the MEAF backbone itself has strong image restoration capability, under light haze conditions, where scene structures and textures are relatively less affected by haze, the backbone alone can achieve satisfactory restoration. Consequently, the additional performance gain provided by PGFM may be limited. As haze density increases, the degradation of scene structures and textures becomes more pronounced. In this case, polarization information can provide supplementary cues to the backbone, making the benefit of PGFM more prominent under medium and high haze densities.

The experimental evaluation in this study has certain limitations. The current polarization dehazing experiments are primarily conducted on synthetic hazy data constructed from real polarization images. The atmospheric scattering model used for synthesis does not fully account for real-environment factors such as non-uniform haze distributions, spatial variations in atmospheric light, wavelength-dependent scattering, and sensor noise. At present, qualitative dehazing experiments on real hazy scenes have only been conducted under laboratory conditions. Future work will further validate the proposed method in

real outdoor hazy environments to assess its applicability and generalization capability under complex natural scattering conditions.

5. Conclusions

This paper proposes an image dehazing method that combines multi-depth feature fusion with polarization guidance. An MEAF backbone is developed, incorporating the Haze Perception Block (HPB) and Adaptive Weight Fusion Module (AWFM) to enhance feature extraction and adaptive fusion across different depths. Building on this backbone, a Polarization-Guided Fusion Module (PGFM) is introduced to incorporate the polarization component S_2 into the restoration process through multi-scale feature extraction, feature concatenation, and residual fusion.

Experimental results show that MEAF achieves a PSNR of 31.40 dB and an SSIM of 0.933 on the synthetic visible-light dehazing dataset. On the synthetic polarization dehazing dataset constructed from PolarLITIS, MEAF-PGFM achieves a mean PSNR of 29.56 dB and a mean SSIM of 0.9592 across three training runs. Compared with the MEAF-S_0 baseline, the overall mean MAE decreases from 0.0457 to 0.0332, demonstrating the contribution of polarization guidance to image restoration. Qualitative experiments on real hazy scenes provide preliminary evidence of the method's applicability, although its generalization capability under natural outdoor haze conditions requires further validation. Overall, the proposed method enables the joint utilization of intensity and polarization information, providing an effective approach to polarization-assisted image dehazing.

Author Contributions: Conceptualization, Y.L. and J.C.; methodology, Y.L.; software, Y.L.; validation, Y.L.; formal analysis, Y.L.; investigation, Y.L.; resources, J.C.; data curation, Y.L.; writing—original draft preparation, Y.L.; writing—review and editing, J.C., Q.F., H.C. and X.L.; visualization, Y.L.; supervision, J.C.; project administration, J.C.; funding acquisition, J.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Innovation Foundation of CAST [WY-YY-CASTJY1100].

Data Availability Statement: The Cityscapes dataset used in this study is available from the official Cityscapes website upon registration and subject to its license terms. The PolarLITIS dataset is publicly available on Zenodo (DOI: 10.5281/zenodo.5547760). The hazy data used in the experiments were generated from these datasets using the atmospheric scattering model described in this study. The processed data are subject to the terms and conditions of the original datasets.

Acknowledgments: The authors gratefully acknowledge the support provided by the Innovation Foundation of CAST.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

MEAF-P-Net	Multi-depth Enhanced Adaptive Fusion Polarization-guided Network
MEAF	Multi-depth Enhanced Adaptive Fusion
MFEM	Multi-depth Feature Extraction Module
AWFM	Adaptive Weight Fusion Module
PGFM	Polarization-Guided Fusion Module
HPB	Haze Perception Block
DCP	Dark Channel Prior
LDP	Learning to Dehaze with Polarization
PSNR	Peak Signal-to-Noise Ratio

SSIM	Structural Similarity Index
MAE	Mean Absolute Error
WGAN-GP	Wasserstein Generative Adversarial Network with Gradient Penalty
GAP	Global Average Pooling
GT	Ground Truth

References

1. He, K.; Sun, J.; Tang, X. Single Image Haze Removal Using Dark Channel Prior. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *33*, 2341–2353. https://doi.org/10.1109/TPAMI.2010.168.
2. Cai, B.; Xu, X.; Jia, K.; Qing, C.; Tao, D. DehazeNet: An End-to-End System for Single Image Haze Removal. *IEEE Trans. Image Process.* **2016**, *25*, 5187–5198. https://doi.org/10.1109/TIP.2016.2598681.
3. Li, B.; Peng, X.; Wang, Z.; Xu, J.; Feng, D. AOD-Net: All-In-One Dehazing Network. In *Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017*; IEEE: New York, NY, USA, 2017; pp. 4770–4778. https://doi.org/10.1109/ICCV.2017.511.
4. Qin, X.; Wang, Z.; Bai, Y.; Xie, X.; Jia, H. FFA-Net: Feature Fusion Attention Network for Single Image Dehazing. In *Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020*; AAAI Press: Palo Alto, CA, USA, 2020; Volume 34, pp. 11908–11915. https://doi.org/10.1609/aaai.v34i07.6865.
5. Qu, Y.; Chen, Y.; Huang, J.; Xie, Y. Enhanced Pix2pix Dehazing Network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019*; IEEE: New York, NY, USA, 2019; pp. 8152–8160. https://doi.org/10.1109/CVPR.2019.00835.
6. Huang, G.; Zhang, J. MAPF-Net: Lightweight Network for Dehazing via Multi-Scale Attention and Physics-Aware Feature Fusion. *J. Supercomput.* **2025**, *81*, 560. https://doi.org/10.1007/s11227-025-07089-x.
7. Wu, Y.; Chen, Z.; Yi, W.; Yang, Y. MWA-Net: Multi-Scale Wavelet-Guided Attention Network for Single Image Dehazing. *Complex Intell. Syst.* **2025**, *11*, 449. https://doi.org/10.1007/s40747-025-02076-4.
8. Zhang, S.; Zhang, Y.; Yu, Z.; Yang, S.; Kang, H.; Xu, J. MCA-GAN: A Multi-Scale Contextual Attention GAN for Satellite Remote-Sensing Image Dehazing. *Electronics* **2025**, *14*, 3099. https://doi.org/10.3390/electronics14153099.
9. Zhang, D.; Ma, L.; Du, W.; Lin, X. Multi-Scale Adaptive Transformer-Based Image Dehazing. *J. Comput.-Aided Des. Comput. Graph.* **2025**. https://doi.org/10.3724/SP.J.1089.2025-00200.
10. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015*; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2015; Volume 9351, pp. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28.
11. Zhou, Z.; Rahman Siddiquee, M.M.; Tajbakhsh, N.; Liang, J. UNet++: A Nested U-Net Architecture for Medical Image Segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2018; Volume 11045, pp. 3–11. https://doi.org/10.1007/978-3-030-00889-5_1.
12. Wen, B.; Peng, S.; Yang, C.; Shen, Y.; Li, H. Multi-Depth Adaptive Fusion Dehazing Generation Network. *J. Jilin Univ. (Eng. Technol. Ed.)* **2025**, *55*, 2103–2113. https://doi.org/10.13229/j.cnki.jdxbgxb.20230813.
13. Lu, L.; Xiong, Q.; Xu, B.; Chu, D. MixDehazeNet: Mix Structure Block for Image Dehazing Network. In *Proceedings of the 2024 International Joint Conference on Neural Networks, Yokohama, Japan, 30 June–5 July 2024*; IEEE: New York, NY, USA, 2024; pp. 1–10. https://doi.org/10.1109/IJCNN60899.2024.10651326.
14. Zhu, X.; Ning, J.; Wu, X.; Duo, B. Frequency Attention-Based Multi-Scale Dehazing GAN. *Digit. Signal Process.* **2026**, *170*, 105747. https://doi.org/10.1016/j.dsp.2025.105747.
15. Cao, B.; Dong, Z.; Zhang, Z.; Wang, T. Remote Sensing Defogging Network with Physical-Level Internal Fusion under Dual Attention. *Spacecr. Recovery Remote Sens.* **2025**, *46*, 83–97. https://doi.org/10.3969/j.issn.1009-8518.2025.06.008.
16. Dong, W.; Xu, T.; Ji, M.; Luo, G.; Li, G.; Zheng, Z. Hybrid Self-Attention Aided Generative Adversarial Network for Hazy Image Restoration. *Optik* **2026**, *345*, 172611. https://doi.org/10.1016/j.ijleo.2025.172611.
17. Chen, Z.; He, Z.; Lu, Z.M. DEA-Net: Single Image Dehazing Based on Detail-Enhanced Convolution and Content-Guided Attention. *IEEE Trans. Image Process.* **2024**, *33*, 1002–1015. https://doi.org/10.1109/TIP.2024.3354108.
18. Chen, J.; He, Y.; Wang, X.; Guo, Z.; Chen, J.; Rao, Z.; Xiang, Y. ED-Former: Efficient Dehazing Transformer with Attention-Adaptive Feed-Forward Network. *Signal Process. Image Commun.* **2026**, *148*, 117634. https://doi.org/10.1016/j.image.2026.117634.
19. Schechner, Y.Y.; Narasimhan, S.G.; Nayar, S.K. Instant Dehazing of Images Using Polarization. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Kauai, HI, USA, 8–14 December 2001*; IEEE: New York, NY, USA, 2001; Volume 1, pp. I-325–I-332. https://doi.org/10.1109/CVPR.2001.990493.

20. Shwartz, S.; Namer, E.; Schechner, Y.Y. Blind Haze Separation. In *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, New York, NY, USA, 17–22 June 2006*; IEEE: New York, NY, USA, 2006; Volume 2, pp. 1984–1991. https://doi.org/10.1109/CVPR.2006.71.
21. Fang, S.; Xia, X.; Huo, X.; Chen, C. Image Dehazing Using Polarization Effects of Objects and Airlight. *Opt. Express* **2014**, *22*, 19523–19537. https://doi.org/10.1364/OE.22.019523.
22. Liang, J.; Ren, L.; Ju, H.; Zhang, W.; Qu, E. Polarimetric Dehazing Method for Dense Haze Removal Based on Distribution Analysis of Angle of Polarization. *Opt. Express* **2015**, *23*, 26146–26157. https://doi.org/10.1364/OE.23.026146.
23. Zhang, W.; Liang, J.; Ju, H.; Ren, L.; Qu, E.; Wu, Z. Study of Visibility Enhancement of Hazy Images Based on Dark Channel Prior in Polarimetric Imaging. *Optik* **2017**, *130*, 123–130. https://doi.org/10.1016/j.ijleo.2016.11.047.
24. Zhang, Y.; Luo, L.Y.; Zhao, H.J.; Qiu, R.D.; Ying, Y.S. Image Dehazing Based on Multispectral Polarization Imaging Method in Different Detection Modes. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2020**, *XLIII-B1-2020*, 615–620. https://doi.org/10.5194/isprs-archives-XLIII-B1-2020-615-2020.
25. Zhou, C.; Teng, M.; Han, Y.; Xu, C.; Shi, B. Learning to Dehaze with Polarization. In *Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Red Hook, NY, USA, 2021; Volume 34, pp. 11487–11500.
26. Bi, P.; Wang, D.; Chen, W.; Yang, L.; Liang, J.; Li, G.; Zhang, F.; Wang, Z.; Zhang, X. Image Dehazing Based on Polarization Information and Deep Prior Learning. *Optik* **2022**, *267*, 169746. https://doi.org/10.1016/j.ijleo.2022.169746.
27. Suo, K.; Lv, Y.; Yin, J.; Yang, Y.; Huang, X. Image Dehazing Combining Polarization Properties and Deep Learning. *J. Opt. Soc. Am. A* **2024**, *41*, 311–322. https://doi.org/10.1364/JOSAA.507892.
28. Wang, X.; Fu, W.; Yu, H. Polarization-Based Image Dehazing Network with Pseudo-3D Convolution. *Pattern Recognit. Lett.* **2025**, *196*, 156–161. https://doi.org/10.1016/j.patrec.2025.05.023.
29. Yang, Z.; He, Q.; Liu, Z.; Fan, B.; Tian, J. DuRP: Dual-Stage Physics-Embedded Learning for Joint Radiance and Polarization Restoration. In *Proceedings of the 43rd International Conference on Machine Learning (ICML)*; Proceedings of Machine Learning Research; PMLR: Cambridge, MA, USA, 2026; Volume 306, pp. 144683–144704. Available online: https://proceedings.mlr.press/v306/yang26u.html (accessed on 18 November 2026).
30. Perez, E.; Strub, F.; de Vries, H.; Dumoulin, V.; Courville, A. FiLM: Visual Reasoning with a General Conditioning Layer. In *Proceedings of the AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018*; AAAI Press: Palo Alto, CA, USA, 2018; Volume 32, pp. 3942–3951. https://doi.org/10.1609/aaai.v32i1.11671.
31. Wang, X.; Yu, K.; Dong, C.; Loy, C.C. Recovering Realistic Texture in Image Super-Resolution by Deep Spatial Feature Transform. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018*; IEEE: New York, NY, USA, 2018; pp. 606–615. https://doi.org/10.1109/CVPR.2018.00070.
32. Park, T.; Liu, M.Y.; Wang, T.C.; Zhu, J.Y. Semantic Image Synthesis with Spatially-Adaptive Normalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019*; IEEE: New York, NY, USA, 2019; pp. 2337–2346. https://doi.org/10.1109/CVPR.2019.00244.
33. Cordts, M.; Omran, M.; Ramos, S.; Rehfeld, T.; Enzweiler, M.; Benenson, R.; Franke, U.; Roth, S.; Schiele, B. The Cityscapes Dataset for Semantic Urban Scene Understanding. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016*; IEEE: New York, NY, USA, 2016; pp. 3213–3223. https://doi.org/10.1109/CVPR.2016.350.
34. Blin, R.; Ainouz, S.; Canu, S.; Mériaudeau, F. The PolarLITIS Dataset: Road Scenes Under Fog. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 10753–10762. https://doi.org/10.1109/TITS.2021.3095658.
35. Liu, X.; Ma, Y.; Shi, Z.; Chen, J. GridDehazeNet: Attention-Based Multi-Scale Network for Image Dehazing. In *Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019*; IEEE: New York, NY, USA, 2019; pp. 7314–7323. https://doi.org/10.1109/ICCV.2019.00741.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.