Article

A Multi-Task Learning and Knowledge Selection Strategy for Environment-Induced Color-Distorted Image Restoration

Yuan Ding and Kaijun Wu *

School of Electronics and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China; 11210739@stu.lzjtu.edu.cn
* Correspondence: wkj@mail.lzjtu.cn; Tel.: +86-130-9922-0306

Abstract: Existing methods for restoring color-distorted images in specific environments typically focus on a singular type of distortion, making it challenging to generalize their application across various types of color-distorted images. If it were possible to leverage the intrinsic connections between different types of color-distorted images and coordinate their interactions during model training, it would simultaneously enhance generalization, address potential overfitting and underfitting issues during data fitting, and consequently lead to a positive performance boost. In this paper, our approach primarily addresses three distinct types of color-distorted images, namely dust-laden images, hazy images, and underwater images. By thoroughly exploiting the unique characteristics and interrelationships of these types, we achieve the objective of multitask processing. Within this endeavor, identifying appropriate correlations is pivotal. To this end, we propose a knowledge selection and allocation strategy that optimally distributes the features and correlations acquired by the network from the images to different tasks, enabling a more refined task differentiation. Moreover, given the challenge of difficult dataset pairing, we employ unsupervised learning techniques and introduce novel Transformer blocks, feedforward networks, and hybrid modules to enhance context relevance. Through extensive experimentation, we demonstrate that our proposed method significantly enhances the performance of color-distorted image restoration.

Keywords: image enhancement; multi-task knowledge allocation; transformer; unsupervised learning; image denoising; color correction

1. Introduction

The issue of image color distortion caused by varying capture environments has been a pivotal research topic in the field of image enhancement. Prominent cases include various similar environmental conditions such as haze, dust, underwater scenarios, rainy weather, and snowfall. The color distortion stemming from environmental factors results in an overall degradation of image quality, thereby adversely affecting computer vision tasks reliant on image data, including applications like surveillance systems, autonomous driving, underwater exploration, and more [1]. Existing methods for enhancing color-distorted images often focus solely on addressing individual cases, such as haze removal, dust reduction, and underwater image enhancement. However, introducing an algorithm or allocation strategy that effectively harnesses the intrinsic relationships among different types of color-distorted images for multi-task learning (MTL) has the potential to tap into the untapped opportunities in this domain, leading to significant improvements in processing outcomes.

In the current domain, the limitations of traditional algorithms mainly stem from their reliance on prior knowledge and parameters, rendering them unable to meet the demands of multi-task scenarios. The previously proposed RGB color balance algorithm [2] demonstrated a degree of effectiveness in enhancing dust and underwater images, yielding

favorable results. Nevertheless, due to the absence of non-uniform RGB channel adjustments in haze images, its performance remains suboptimal for such cases. In recent years, deep learning has exhibited substantial competitiveness in the field of computer vision. Some algorithmic models address multi-task challenges by employing distinct pre-trained weights, as exemplified by [3,4]. Other network models introduce gating mechanisms, enabling differentiated task handling across diverse network branches, thereby achieving improved results. However, the underlying principle of these approaches still hinges on accentuating the disparities between tasks to realize multi-task processing. This methodology fundamentally overlooks the inherent relationships among distinct types of images. Effective information exchange is lacking between different tasks, resulting in each task being influenced solely by other isolated tasks. This deficiency hampers the collaborative advancement of each task, leading to models of this nature often underperforming compared to models specifically tailored for individual tasks [3].

This study delves into a comprehensive exploration of haze, dust, and underwater images, all of which suffer color distortion due to environmental factors. Through the application of neural networks, we engage in feature exploration of these images with the aim of unearthing their intrinsic connections. Our focus lies in the deep-level features, where we design a strategy encompassing multi-task learning and knowledge allocation.

Our main contributions in this paper are as follows:

- In general, we have developed a multi-task learning and knowledge allocation strategy tailored for enhancing three categories of images: haze, dust, and underwater images. This strategy has been applied to our network model. By uncovering the interconnections and unique characteristics of these three image categories, we aim to mutually enhance the overall model performance and the enhancement effects for individual tasks. This approach also mitigates the performance degradation caused by the interplay between different tasks;
- We have introduced a novel approach called the Frequency Domain Similarity-Gated Selection Vision Transformer (FSGS-ViT) and a Mixed-Scale Frequency Domain Feedforward Network (MDFN). In these methods, we have incorporated the Discrete Cosine Transform (DCT) operator for self-attention value decomposition and similarity calculation. This enables adaptive selection of relevant self-attention values during training and incorporates gating operations to mitigate feature redundancy;
- We have devised a Hybrid-Scale Knowledge Selection Module (MKSM) to explore the retention of low-frequency and high-frequency information within a multi-scale representation. This module aims to enhance the accuracy of knowledge allocation and the potential restoration of clear images by determining which information to retain at various scales;
- We have introduced an Adaptive Gate Mixed Module (AGMM), which employs gating to selectively retain appropriate shallow features and then adaptively integrates information from both deep and shallow layers;
- Extensive experimental results on various tests demonstrate that our approach outperforms state-of-the-art (SOTA) methods, showcasing strong performance.

2. Related Works

2.1. Enhancement of Haze Images

In the early stages of research, Tan et al. [5] introduced the concept of Markov Random Fields, aiming to enhance haze images by maximizing local contrast. Subsequently, He et al. [6] proposed the widely acclaimed dark channel prior estimation method, which gained significant popularity. Following this, dark channel-based dehazing techniques underwent continuous improvements, such as the fast image dehazing method based on the dark channel introduced by Wu et al. [7], the sky-constrained dark channel prior-based dehazing method presented by Xiao et al. [8], and the remote sensing image cloud detection algorithm based on dark channel by Yang et al. [9]. Fattal et al. [10] introduced the color-line method based on the one-dimensional distribution of image blocks in the RGB color

channels. Despite the notable achievements of traditional algorithms in dehazing, there still exist several limitations, including issues related to poor robustness.

DehazeNet [11] was the pioneering model that utilized deep learning for dehazing, employing Convolutional Neural Networks (CNNs) for feature extraction, followed by the application of the Atmospheric Scattering Model (ASM) to obtain clear images through transmission map estimation. On the other hand, AOD-Net [12] aimed to eliminate the complex transmission map and atmospheric light estimation steps in previous models, achieving simultaneous estimation of transmission map and atmospheric light through a redesigned ASM. However, these ASM-based models tend to exhibit color distortion issues when handling dust and underwater images, prompting a shift towards end-to-end restoration approaches. Qin et al. [13] introduced the FFA-Net with attention modules, achieving promising results. Singh et al. [14] employed multiple UNet networks to output different scale features and then combined them for multi-scale learning. Das et al. [15] proposed the Fast Deep Multi-Patch Hierarchical Network, aggregating multiple features from various spatial regions with fewer parameters to restore non-uniform hazy images. Yu et al. [16] used a dual-branch neural network to address different aspects of the dehazing process, enhancing different aspects of hazy image restoration with distinct network branches. Additionally, several other models [17,18] have been dedicated to enhancing the clarity of haze images, yielding commendable outcomes.

2.2. Enhancement of Sand Dust Images

The primary task of enhancing dust-laden images is to address issues such as severe color shifts and decreased contrast. Some commonly used color correction algorithms like the Retinex algorithm [19], Grey World algorithm [20], Histogram Equalization [21], and Wavelet Transform [22] were found ineffective in directly enhancing dust-laden images based on experimentation. Yan et al. [23] attempted a combination of global fuzzy enhancement and constrained histogram equalization, which improved the contrast after enhancing dust-laden images, but the color correction effect was not ideal. Shi et al. [24] proposed a method that employs constrained contrast adaptive histogram equalization to enhance contrast in dust-laden images, followed by gamma correction for normalization, and finally obtains the enhanced image based on the Grey World principle. While this method exhibited good results, it struggled with color restoration in severely degraded images. Another category of methods relies on image restoration and frequently depends on the Atmospheric Scattering Model (ASM). In the domain of image dehazing, the ASM model has been widely employed. In the context of enhancing dust-laden images, many methods still rely on ASM. Gao et al. [25] proposed a dust-laden image enhancement algorithm based on ASM that reverses the blue channel. This method performed well in removing dust, but its effectiveness diminished when handling images heavily contaminated with dust.

Currently, dust-laden image enhancement algorithms based on deep learning have garnered significant attention in the academic community. However, the development of deep learning in the field of dust-laden image enhancement has not reached an ideal state due to the lack of large-scale publicly available standard datasets for dust-laden images. Currently, the primary approach involves training networks using synthesized dust-laden images and then testing them using real captured dust-laden images. Nonetheless, this method has evident drawbacks as synthetic images cannot fully capture the distinctive features of real dust-laden images. Some researchers have attempted to address the data scarcity issue through methods such as transfer learning [26] and unsupervised training [27], achieving some degree of success.

2.3. Enhancement of Under Water Images

In the realm of traditional methods, Li et al. [28] devised a piecewise linear function for histogram equalization, resulting in images with more details in the RGB channels. The CLAHE color model [29] employed a mixed contrast adaptive histogram equalization

approach in both RGB and HSV color spaces, effectively enhancing underwater image quality. However, this method sometimes led to over-enhancement or under-enhancement issues. Tang et al. [30] introduced a novel underwater image enhancement algorithm based on adaptive feedback and the Retinex algorithm, enhancing color saturation, richness, local contrast, and clarity. Nevertheless, it suffered from the problem of post-enhancement blurring. Nicholas et al. [31] proposed using the attenuation differences between three color channels underwater to estimate scene depth and subsequently perform image enhancement. Drews et al. [32] introduced an approach based on observing the absorption rates of the R channel in a significant number of underwater images to restore a high-quality image using the underwater dark channel prior (UDCP). However, this method often requires a substantial number of physical parameters and underwater optical characteristics, making its generalizability limited.

Anwar et al. [33] introduced an end-to-end underwater image enhancement strategy, achieving favorable results by optimizing a combination of MSE and SSIM loss functions. Li et al. [34] devised a multi-term loss function and proposed a weakly supervised color transfer method to rectify color, attaining state-of-the-art (SOTA) performance. Their approach successfully addressed the weakening of model generalization caused by the disregard of wavelength-dependent attenuation or assumptions about specific spectral profiles by most algorithms and models. UGAN [35] employed Cycle-GAN as an image degradation process to train a paired dataset and utilized the Pix2Pix model to enhance underwater image quality. Liu et al.'s MLFcGAN [36] utilized global features to enhance local features at each scale for color correction and image detail preservation. While MLFcGAN exhibited some enhancement in synthesized underwater images, its effectiveness was limited when dealing with severely degraded images. Naik et al. [37] proposed a light model named Shallow-UWNet for underwater image enhancement, which achieved enhancement performance comparable to the then SOTA model while employing fewer parameters.

2.4. Multi-Task Learning Strategy

Multi-task learning methods are generally classified into two categories based on parameter sharing strategies. The first category is hard parameter sharing, where all tasks share the same backbone encoder, and different tasks obtain their respective target results through branch decoders. This type of method is widely used in current multi-task learning, including [38,39]. The second category is soft parameter sharing, where each task has an independent network branch. Different branches obtain different parameter assignments through a designer-defined parameter sharing mechanism, such as [40,41]. Parameter sharing methods employ a single network for multi-task training, where different tasks can selectively choose different neural network functional layers to compose their own execution network paths.

3. Proposed Method

This paper primarily focuses on the multi-task enhancement of degraded images captured in hazy, dusty, and underwater environments. The aforementioned three image categories share several common characteristics: (1) all three categories belong to environmentally induced color-shifted images due to light refraction; (2) they all suffer from reduced clarity caused by environmental medium occlusion; (3) they exhibit similar distribution characteristics in their environmental backgrounds; (4) all three fall within the research domain of image color restoration. Despite these shared traits, there are also distinct differences: (1) the influencing environmental mediums for these categories significantly differ, with dust and haze involving particle refraction, where dust particles are larger than haze particles, while underwater images are color-shifted due to water molecule absorption; (2) variations in medium lead to different image clarity; (3) in the RGB color space, the attenuation patterns of the R, G, and B channels differ, as shown in the figure. Beyond these general distinctions and commonalities, this paper predominantly exploits neural networks to investigate both the differences and commonalities in deep

feature domains among the three categories, with the aim of enhancing the enhancement performance within this domain.

In this section, the overall network architecture employs Cycle-GAN for unsupervised learning, a structure previously applied in various unsupervised contexts [27], which we will not elaborate on extensively. The focus is on introducing the design of our generator model, as shown in Figure 1. Firstly, we provide a description and introduction to the details of FSGS-ViT and MDFN. These two modules serve as the backbone feature extraction and learning units within the network's main structure. Secondly, the Mixed-Scale Knowledge Selection Module (MKSM) is introduced, which serves as a pivotal design element within the multi-task neural network model. Lastly, we present the Adaptive Gate Mixed Module (AGMM).
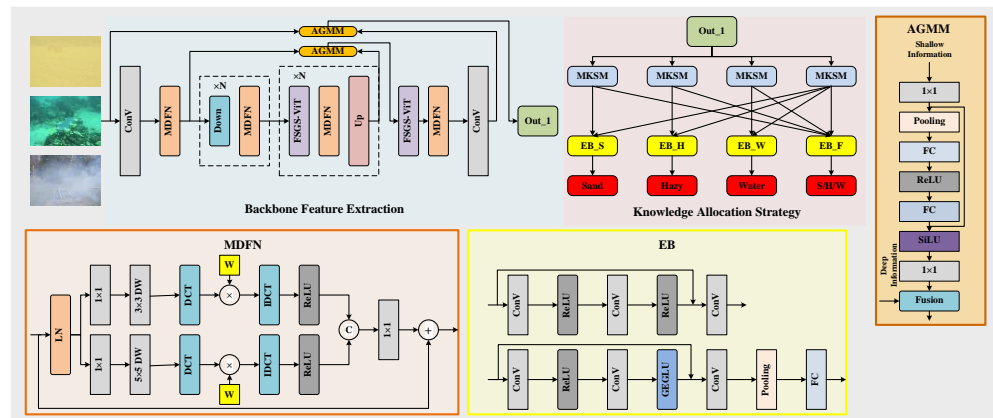

Figure 1. The overall architecture of the network generator proposed for image enhancement in this paper is depicted. It mainly comprises MDFN, FSGS-ViT, and EB, along with MKSM and AGMM. The detailed structures of MDFN, EB, and AGMM are presented beside the overall architecture.

3.1. Frequency Domain Similarity Gated Selection Vision Transformer

In the field of image restoration, noise interactions among unrelated features can have a negative impact on image restoration. The conventional Transformer employs global computation of self-attention to handle all token features, which is not an ideal operation for image restoration. The emergence of the sparse Transformer [42] effectively addresses this issue. It selects the top-k contributing scores with the aim of retaining the most significant elements and removing unnecessary parts. This is achieved by computing the similarity between all queries and keys in terms of pixel-wise similarity and masking out irrelevant elements with lower attention weights in the transposed attention matrix. This dynamic selection transforms the attention in the Transformer from dense to sparse, effectively mitigating the impact of unrelated features. The outputs of the standard Transformer and sparse Transformer are generally represented as follows:

$$\text{SelfAtt}(Q, K, V) = \text{softmax}(\frac{QK^T}{\sqrt{d_k}})V \tag{1}$$

$$\text{SpaAtt}(Q, K, V) = \text{softmax}(\text{TopK}(\frac{QK^T}{\sqrt{d_k}}))V \tag{2}$$

where $\text{SelfAtt}(Q, K, V)$ represents the output representation of the standard Transformer, Q is a matrix composed of a set of query sequences, K is a matrix composed of a set of key sequences, V is a matrix composed of a set of value sequences, and d_k is the dimension of the Query and Key vectors. $\text{SpaAtt}(Q, K, V)$ represents the output representation of the sparse Transformer. TopK denotes the learnable top-*k* selection operator.

The sparse Transformer provides a valuable insight by discarding irrelevant features during the similarity calculation, thereby reducing computational complexity while increas-

ing the utilization of useful features. Inspired by this idea, we have designed a novel block specifically for multi-task learning called the Frequency Domain Similarity-Gated Selection Vision Transformer block (FSGS-ViT). In our FSGS-ViT, we incorporate two branches, an upper branch containing the Frequency Domain Adaptive Attention Block (FAB), and a lower branch consisting of a Gate Selection Block (GSB). This design is guided by the notion of enhancing the utilization of essential features while efficiently managing computational resources, and the specific structures of both components are illustrated in Figure 2.
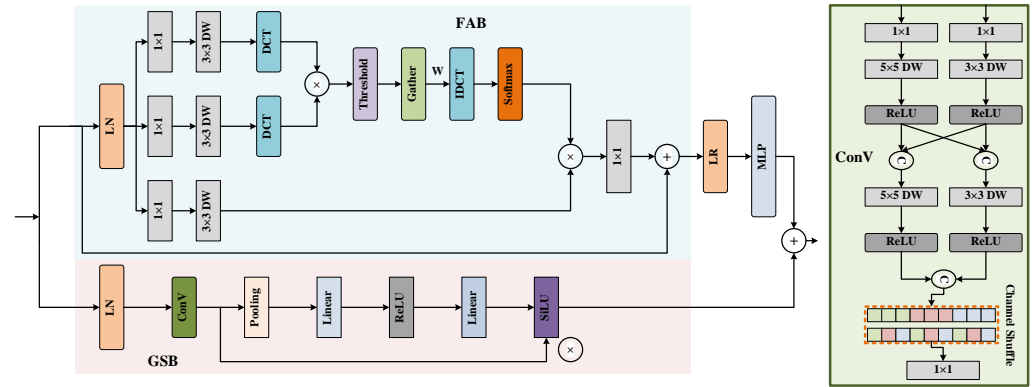

Figure 2. The overall architecture of the proposed FSGS-ViT in this paper is depicted. It primarily comprises the FAB and the GSB. The detailed structure of the ConV Block within GSB is illustrated adjacent to the overall architecture.

3.1.1. FAB

We start with an input feature *X*, which undergoes a three-branch process involving 1×1 convolutions and 3×3 Depthwise Separable (DW) convolutions to obtain *Q*, *K*, and *V*. Within the self-attention mechanism, our initial step involves applying Discrete Cosine Transform (DCT) to *Q* and *K*, thereby transforming them into the frequency domain. This transformation results in frequency representations, namely Q_{dct} and K_{dct}.

$$K_{dct}/Q_{dct} = D\{ConV_{3\times3}[ConV_{1\times1}(X)]\} \qquad (3)$$

where *D* represents Discrete Cosine Transform (DCT) in the frequency domain conversion.

Next, we compute the correlation scores between Q_{dct} and K_{dct}. Let us assume we use a matrix *M* to represent the correlation matrix, where $M(i, j)$ indicates the correlation score between query position *i* and key position *j*. We achieve this by element-wise multiplication of these correlation scores.

$$M = K_{dct} \otimes \overline{Q_{dct}} \qquad (4)$$

Next, we utilize the Gather function to collect or extract the relevant correlation scores from the frequency-domain representation *M* obtained after the DCT transformation. Firstly, we need to set a threshold value *T* for filtering attention weights. The threshold value *T* can be a fixed number or a parameter dynamically adjusted based on the characteristics of tasks and data. With the threshold value *T*, we compute an index array $ijdx$, which contains the indices of all positions exceeding the threshold *T*. The specific calculation is as follows:

$$ijdx = [i, j \text{ for } i, j \text{ in range}(N) \text{ if } M[i, j] > T] \qquad (5)$$

For each index $ijdx$, the Gather function selects the element at position $[i, j]$ from the input tensor *M* and keeps it unchanged. For the elements that were not selected, a Gather function is used to replace their probabilities with 0 at the given indices. Additionally, a

learnable parameter matrix is introduced to appropriately process or adjust the collected correlation scores M_{gather}.

$$M_{gather}[i,j] = \begin{cases} M[i,j] & M[i,j] > T \\ 0 & otherwise \end{cases} \tag{6}$$

$$M_{adjusted} = M_{gather} \otimes W_{randn} \tag{7}$$

where W represents the learnable parameter matrix.

Utilizing the adjusted correlation scores, M_{gather}, the final attention weights A are computed by applying the inverse DCT and the softmax function.

$$A = \sigma\left[D^{-1}\left(M_{adjusted}\right)\right] \tag{8}$$

where σ represents the softmax function, and D^{-1} represents the inverse DCT frequency domain transformation. Normalizing the correlation scores into attention weights ensures they lie within the range of (0, 1) and sum up to 1. Finally, applying the attention weights A to V yields the ultimate self-attention output. By implementing correlation computation and adaptive adjustment of attention weights, the model's perception of crucial features and details within the image is enhanced, thereby boosting the performance and effectiveness of image processing tasks. The overall output expression of FAB can be represented as:

$$\begin{aligned} FAB\_OUT = \quad & MLP(LN(X + (ConV_{1 \times 1}(V \otimes \\ & \sigma(D^{-1}(W_{randn}(Gather(D(Q) \otimes \\ & D(K))))))))) \end{aligned} \tag{9}$$

where *LN* represents Layer Normalization.

3.1.2. GSB

In the lower branch, we introduce the SiLU activation function as a gate for each feature channel. The inclusion of ConV blocks enables adaptive feature selection, better capturing detailed information within images, and avoiding feature redundancy in tasks. Inside the ConV block, we adopt two kernel sizes, 3×3 and 5×5, for mixed-scale processing, and the specific structures are presented in Figure 2 as shown. After concatenation, we apply Channel Shuffle for mixed feature handling, addressing inter-group feature communication.

SiLU, functioning as a gate for each feature channel, is applied to the feature vectors of each channel. It acts as a gating mechanism, activating elements within the feature vector to adjust the importance of each channel. This leads to saturation of feature vectors for each channel as they approach 0, enhancing non-linearity and adaptively adjusting the significance of feature channels based on their element values. Larger element values amplify the importance of the corresponding channel, focusing more on useful feature mappings, while smaller element values reduce the weight of the corresponding channel, minimizing the impact of irrelevant feature mappings on the model. Rejecting irrelevant features prevents feature redundancy. Elements with values below 0, after passing through the SILU function, are mapped to values close to 0. This implies rejection of feature mappings for that channel. For input feature X, the branch is represented as:

$$F_{3 \times 3/5 \times 5} = \text{ReLU}(ConV_{3 \times 3/5 \times 5}(LN(X))) \tag{10}$$

$$F_{Conc} = CS(Concat(F_{3 \times 3}, F_{5 \times 5})) \tag{11}$$

$$\begin{aligned} GSB\_OUT = \quad & F_{Conc} \otimes \text{SiLU}(FC(\\ & \text{ReLU}(FC(Pool(F_{Conc}))))) \end{aligned} \tag{12}$$

where $F_{3 \times 3/5 \times 5}$ represents the feature vector obtained through mixed-scale processing, *CS* stands for Channel Shuffle operation, and *FC* represents the fully connected layer.

Taking into consideration the introduction of both the *FAB* and the *GSB*, the output of our FSGS-ViT can be represented as follows:

$$FSGS\_OUT = FAB\_OUT + GSB\_OUT \tag{13}$$

3.2. Mixed-Scale Frequency Domain Feedforward Network

Historically, the research and design of Feedforward Neural Networks (FNNs) often focused on enhancing performance by introducing intricate convolutional structures. However, this approach overlooked the significance of correlations between multiscale image features and their advantageous impact on image restoration. Contemporary studies have substantiated the effectiveness of multiscale approaches in image processing tasks. Beyond multiscale features, correlations between frequency domains also exert a noticeable influence on the image restoration process. Analogous to the principles of attention mechanisms, not all high-frequency and low-frequency information in the frequency domain proves beneficial for image restoration. Therefore, considering the amalgamation of these concerns, we introduce the Mixed-Scale Frequency Domain Feedforward Network (MDFN), and the specific structures are presented in Figure 1 as shown.

According to the depiction of MDFN in Figure 1, given an input tensor X, subsequent to being processed through parallel branches of 3×3 and 5×5 convolutions, the image is transformed into the frequency domain. Then, an adaptable parameter matrix is incorporated to dynamically retain useful frequency domain information while discarding irrelevant frequency domain information. This process achieves efficient information filtering and retention. Taking the 3×3 branch as an example, the operational principle of MDFN is defined as follows:

$$F_{3\times3} = \text{ConV}_{3\times3}(\text{ConV}_{1\times1}(\text{LN}(X))) \tag{14}$$

where $F_{3\times3}$ represents the feature tensor after undergoing 1×1 and 3×3 convolutions. Subsequently, the DCT is applied to transform it into the frequency domain, utilizing a learnable parameter matrix to control the retention of information:

$$F_{3\times3}^{dct} = \text{ReLU}(D^{-1}(W \otimes D(F_{3\times3}))) \tag{15}$$

The 5×5 branch in the parallel processing follows a similar process to the previously mentioned 3×3 branch. After obtaining the results from these two parallel branches, we proceed to obtain the final output of the network module through the following steps:

$$MDFN\_OUT = \text{ConV}_{1\times1}(Concat(F_{3\times3}^{dct}, F_{5\times5}^{dct})) + X \tag{16}$$

3.3. Knowledge Allocation Strategy

As a multi-task and knowledge allocation system, a well-designed knowledge allocation strategy is crucial for achieving effective multi-task learning after the completion of the backbone feature extraction phase. In our study, we employed a mixed dataset training approach, combining dust images, haze images, and underwater images into a comprehensive dataset. The purpose of this approach is to solely rely on the network itself to differentiate between the features of the three categories and shared features. Through this strategy, we aim to perform multi-task learning within a unified framework, allowing the network to fully leverage the acquired knowledge and better address the challenges of multiple tasks. Such a knowledge allocation strategy enhances the model's generalization capabilities and enables more efficient utilization of existing data resources.

To effectively distinguish between the features extracted by the backbone feature extraction and classify them into the three distinct categories and shared features, we employed four independent parallel Mixed-Scale Knowledge Selection Modules (MKSM) for knowledge selection. Subsequently, knowledge allocation was performed using the approach provided by [43], where the features belonging to each of the three categories

were allocated to their respective Enhancement Modules (EB), while the shared features were allocated to each Enhancement Module. Through this approach, we were able to enhance the enhancement effects for the three categories of images.

Firstly, the backbone features were individually fed into the MKSM for the following knowledge selection process:

$$X_S = \text{MKSM}_S(X_B) \tag{17}$$

$$X_H = \text{MKSM}_H(X_B) \tag{18}$$

$$X_W = \text{MKSM}_W(X_B) \tag{19}$$

$$X_F = \text{MKSM}_F(X_B) \tag{20}$$

where X_B represents the backbone features, X_S represents the independent features required for dust image enhancement, X_H represents the independent features required for haze image enhancement, X_W represents the independent features required for underwater image enhancement, and X_F represents the shared features. The knowledge allocation strategy is carried out as follows:

$$R_S = EB_S(X_S, X_F) \tag{21}$$

$$R_H = EB_H(X_H, X_F) \tag{22}$$

$$R_W = EB_W(X_W, X_F) \tag{23}$$

$$R_F = EB_F(X_S, X_H, X_W) \tag{24}$$

where EB represents the respective enhancement modules, R represents the results after enhancing the three types of images, and R_F represents the recognition results of the three types of images. The structure of EB is illustrated in Figure 1 as shown.

The above constitutes the fundamental principle of the entire multi-task knowledge allocation. By utilizing the knowledge selection capability of MKSM, an adaptive multi-task processing is achieved, effectively leveraging the variations and connections among different tasks. This approach facilitates the collaborative enhancement of task performance through the interplay of relationships between different tasks, ultimately leading to a balanced performance across tasks.

3.3.1. Mixed-Scale Knowledge Selection Modules

We employed four independent parallel MKSM modules for knowledge selection, with all four modules sharing the same structure, as shown in Figure 3. The principle behind knowledge selection is achieved through the gating property of the SiLU activation function. During the feature transformation process, we utilized 3×3 and 5×5 depthwise convolutions to enhance the extraction of multi-scale local information, thus improving the effectiveness of knowledge selection. The specific procedure is illustrated as follows:

$$X_{3\times3} = \text{ReLU}(\text{ConV}_{3\times3}(\text{ConV}_{1\times1}(X_B))) \tag{25}$$

$$X_{5\times5} = \text{ReLU}(\text{ConV}_{5\times5}(\text{ConV}_{1\times1}(X_B))) \tag{26}$$

$$X_{3C5} = Concat(X_{3\times3}, X_{5\times5}) \tag{27}$$

where X_B represents the backbone features. The subsequent operations are identical. Taking the 3×3 branch as an example:

$$X_{3\times3}^2 = X_B + \text{ConV}_{1\times1}(\text{ReLU}(\text{ConV}_{3\times3}(X_{3C5}))) \tag{28}$$

$$X_{3\times3}^3 = \text{ReLU}(FC(Pool(X_{3\times3}^2)) \tag{29}$$

Obtain $X^3_{5\times5}$ using the same method, resulting in the final output of MKSM.

$$X^2_{3C5} = Concat(X^3_{3\times3}, X^3_{5\times5}) \tag{30}$$

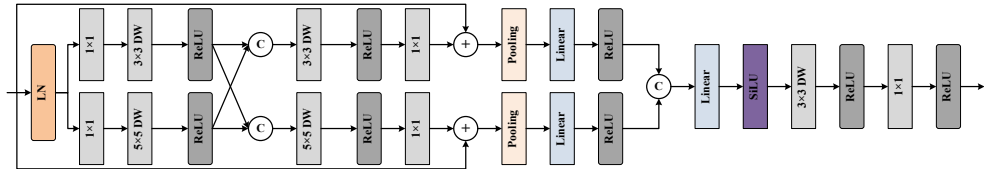$$MKSM\_OUT = \text{ConV}(\text{SiLU}(FC(X^2_{3C5})) \tag{31}$$

Figure 3. The overall architecture of the proposed MKSM in this paper is depicted.

3.3.2. Enhancement Modules

We employed four EB structures to achieve the final decoding function, with each EB module dedicated to the recovery of one of the three image categories. The $EB_{S/H/W}$ modules corresponding to the three image categories are implemented with a three-layer stacked convolutional structure.

$$EB_{S/H/W} = \text{ConV}_3(X_{S/H/W}, X_F) \tag{32}$$

where ConV_3 represents a three-layer convolutional structure.

The corresponding EB_F module, which shares features, also employs a three-layer stacked convolutional structure. However, what distinguishes it is the utilization of the GEGLU (Gated Exponential Linear Units) activation function. The GEGLU function combines the ELU function with the Sigmoid function, employing a gating mechanism to modulate activation values, aiming to enhance the neural network's modeling capability for different features. The gating mechanism allows the network to dynamically adjust activation values, adapting to various features in the input data. This combined design contributes to the improvement of the neural network's expressive capacity, making it more suitable for a variety of complex tasks.

$$EB_F = \text{FC}(\text{Pool}(\text{ConV}(\text{GEGLU}(\text{ConV}_3(X))))) \tag{33}$$

3.4. Adaptive Gate Mixed Module

To address the issue of shallow information loss in neural networks, conventional residual connections lack the ability to selectively retain useful information adaptively. To tackle this problem, we propose an Adaptive Gate-Mixing Module (AGMM), as shown in Figure 1. This module achieves adaptive preservation of shallow-level information through gate-based selection and a unique fusion strategy, enabling it to better aid the deep-level features in fitting target objects.

Firstly, through network processing, we achieve gate-based selection of shallow-level features using the SiLU activation function:

$$X_G = \text{SiLU}(FC(\text{ReLU}(FC \atop (Pool(\text{ConV}_{1\times1}(X_{SI})))))) \tag{34}$$

where X_{SI} represents shallow-level features, and X_G represents the features after gate selection. Subsequently, the processed shallow-level features are adaptively fused with deep-level features through a fusion process, which can be represented as:

$$AGMM\_OUT = \begin{aligned}&X_G \otimes \exp\text{end\_as}(\rho, X_G) +\\&X_{DI} \otimes (1 - \exp\text{end\_as}(\rho, X_{DI}))\end{aligned} \tag{35}$$

where X_{DI} represents deep features, ρ is a learnable scaling factor that can be obtained through training, and exp end_as is the expend_as() function used to match the tensor size.

3.5. Backbone Feature Extraction

In the context of image restoration tasks, the goal of designing the backbone feature extraction network is to ensure that the model can effectively capture and represent the crucial information in the images, enabling high-quality restoration in subsequent tasks. In this regard, we meticulously crafted the arrangement sequence of various modules.

It is noteworthy that in the encoder part, a deliberate decision was made to abstain from adopting the FSGS-ViT module. This decision was influenced by the perspective presented in [44], which posits that shallow-level features extracted by the encoder tend to be relatively ambiguous compared to the deep-level features extracted by the decoder. Given the critical importance of accurate similarity estimation in image restoration tasks, the inclusion of ambiguous features in the encoder could potentially lead to inaccurate restoration results. Hence, we chose to avoid integrating the FSGS-ViT module in the encoder, ensuring that the features we extract possess clarity and precision.

Contrarily, the FSGS-ViT module was exclusively employed in the decoder part. By introducing this module in the decoder, we could leverage deep-level clear features for more precise similarity estimation. This asymmetric encoder–decoder structure was intentionally designed to meet the specific requirements of image restoration tasks.

The objective of this design decision is to optimize the network architecture to better suit the demands of image restoration tasks. By emphasizing the critical role of deep-level features in similarity computation, we aim to enhance the model's understanding of the internal structure and context of images, thereby achieving superior performance in image restoration tasks. This clever structural arrangement contributes to achieving more accurate and reliable image restoration outcomes.

4. Experiments

In this section, we evaluate our method through experiments and compare it with state-of-the-art approaches using publicly available benchmark datasets.

4.1. Datasets and Experimental Settings

Datasets. For dust images, due to the lack of publicly available standard datasets in the academic community, other related studies often rely on synthetic datasets. To address this issue, we collected and curated a real dust image dataset named "D-Sand." This dataset consists of high-quality dust images that exhibit strong dust effects. The "D-Sand" dataset consists of a total of 900 high-quality dust images, primarily collected through a combination of web scraping and manual photography. The manually captured images were taken in the northwestern regions of China, including locations such as Dunhuang and Lanzhou. The dataset encompasses diverse content, featuring natural landscapes (mountains, sand dunes), urban scenes (buildings, roads, vehicles), and human subjects.

For haze images, we utilized multiple datasets, including "RESIDE" [45], "O-Haze" [46], "Dense-Haze" [47], and "NTIRE 2020" [48] datasets. In the underwater image domain, we selected the "SUIM" [49], "UIEB" [50], and "RUIE" [51] datasets. To ensure a sufficient and balanced number of samples for different categories in the mixed dataset, we ensured that each class of images contained 10,000 images in the training set. To address the scarcity of data, we excluded test images, split the remaining images into non-overlapping image blocks, and applied appropriate image augmentation techniques to generate 10,000 training images. Therefore, the mixed dataset contains a total of 30,000 pairs of training images. Additionally, we employed collected paired datasets for network pretraining to enhance the effectiveness of subsequent unsupervised training.

Training details. We employed the same loss function as in [52] to constrain the network. The image size was fixed at 256 × 256. During the training process, we utilized 3 NVIDIA A100 GPUs in conjunction with the FSDP training mode proposed by Facebook.

We employed the Adam optimizer for local parameter updates [10] and used default parameters to accelerate training efficiency. The initial learning rate was set to 1×10^{-4}, and after 100 K iterations, we employed a cosine annealing strategy to gradually reduce the learning rate. The minimum value of the learning rate was set to 1×10^{-6}.

Comparison methods. To validate the effectiveness of our method, we conducted comprehensive comparisons with both traditional and learning-based approaches. In the realm of traditional methods, techniques like DCP [6], RCP, MSRCR, RGHS, ACE, and Park [53] have maintained a certain status in the field of image enhancement due to their good results and simple processing pipelines. In our experiments, we evaluated these traditional methods on the three types of images based on their respective application domains.

For learning-based methods, we selected TBNN [16], USDR-Net [54] for sandstorm image enhancement comparison, TBNN [16], DehazeC [55], FSAD [15] for haze image enhancement comparison, and FUnIEGAN [56], UWNet [37], NUˆ2NET [57], and PUIE-NET [58] for underwater image enhancement comparison. For models that did not provide pre-trained parameters, we retrained the models with the provided code. For models with existing pre-trained parameters, we performed fair comparisons using their online code and parameters.

Evaluation metrics. We employed PSNR and SSIM as the evaluation metrics for the aforementioned benchmarks. Following prior research practices, we computed the PSNR and SSIM metrics for the Y channel of the YCbCr color space. For images without paired ground truth values, we utilized non-reference metrics such as e-score, r-score, UIQM, and UCIQE algorithms.

4.2. Comparison with the State-of-the-Art

Quantitative evaluation. For dust-laden images, Table 1 displays the e-score and r-score values of various enhancement methods. It can be observed that both the traditional methods, MSRCR and Park, as well as the learning-based methods TBNN and USDR-Net, exhibit remarkable performance in terms of evaluation metrics. Park, as a traditional method, not only boasts a notable advantage in runtime but also secures top-three rankings in the evaluation scores. USDR-Net achieves second-best results through unsupervised adversarial learning, showcasing impressive generalization capability. However, in comparison, our method attains the best performance with significant improvements. This underscores the relatively gradual progress in the field of dust-laden image enhancement and demonstrates the accomplishments stemming from our exploration of correlations and uniqueness among similar images.

Table 1. Quantitative comparison of real-word sand dust images.

Datasets	DCP e-Score/ r-Score	MSRCR e-Score/ r-Score	RGHS e-Score/ r-Score	Park e-Score/ r-Score	TBNN e-Score/ r-Score	USDR-Net e-Score/ r-Score	Our e-Score/ r-Score
D-Sand	0.1214/1.2073	0.7192/1.8406	0.4307/1.3385	0.6832/1.5354	0.5204/1.6448	0.8643/1.8511	1.2308/2.2216

Moving to the haze dataset as seen in Table 2, our approach showcases outstanding performance across various haze image datasets. It achieves the highest PSNR and SSIM indices in the challenging Dense-Haze dataset. TBNN not only demonstrates superior performance compared to other contrast methods but also maintains consistent performance across datasets.

Table 2. Quantitative comparison of haze images.

Datasets	DCP PSNR/SSIM	MSRCR PSNR/SSIM	ACE PSNR/SSIM	TBNN PSNR/SSIM	DehazeC PSNR/SSIM	FSAD PSNR/SSIM	Our PSNR/SSIM
RESIDE	17.6384/0.8544	16.3526/0.8096	16.9427/0.7925	36.6923/0.9819	30.5733/0.9736	27.3653/0.9573	37.3959/0.9912
O-Haze	13.1259/0.4843	12.9329/0.4893	13.2128/0.5102	25.3586/0.7883	19.5411/0.6823	18.3945/0.5730	27.8472/0.8265
Dense-Haze	11.3957/0.4691	10.7468/0.4135	9.6284/0.4409	16.2190/0.5722	12.5832/0.4782	12.2284/0.4937	17.2514/0.6342
NTIRE 2020	12.4926/0.4463	12.1688/0.4274	12.1039/0.4392	20.9754/0.7164	17.4927/0.6055	16.4856/0.5853	21.2485/0.7182

Table 3 presents the performance of different methods on underwater image datasets. Our method outperforms the comparative algorithms on all three underwater image datasets. Compared to the second-best overall performer, FUnIEGAN, significant enhancements are achieved in UIQM and UCIQE metrics.

Table 3. Quantitative comparison of underwater images.

Datasets	RCP UIQM/UCIQE	Park UIQM/UCIQE	FUnIEGAN UIQM/UCIQE	UWnet UIQM/UCIQE	NU^2NET UIQM/UCIQE	PUIE-NET UIQM/UCIQE	Our UIQM/UCIQE
SUIM	4.1746/0.3526	4.4736/0.3380	4.9362/0.3865	4.5729/0.3702	4.8352/0.4117	4.8794/0.4125	4.9746/0.4115
UIEB	4.1182/0.3163	4.3283/0.3374	4.7045/0.3610	4.4973/0.3479	4.6281/0.4013	4.5310/0.3825	4.7493/0.4037
RUIE	4.4931/0.3237	4.7209/0.3645	5.5312/0.4183	5.4581/0.4105	5.2938/0.4166	5.3832/0.3941	5.6841/0.4130

Qualitative evaluation. Figure 4 showcases the training results for real dust-laden images. Among traditional methods, only the MSRCR and Park algorithms exhibit effective dust removal, while other traditional methods fail to eliminate dust and introduce color distortions. Learning-based methods show varying degrees of enhancement, with our method demonstrating the best visual effect by achieving comprehensive dust removal and color restoration.
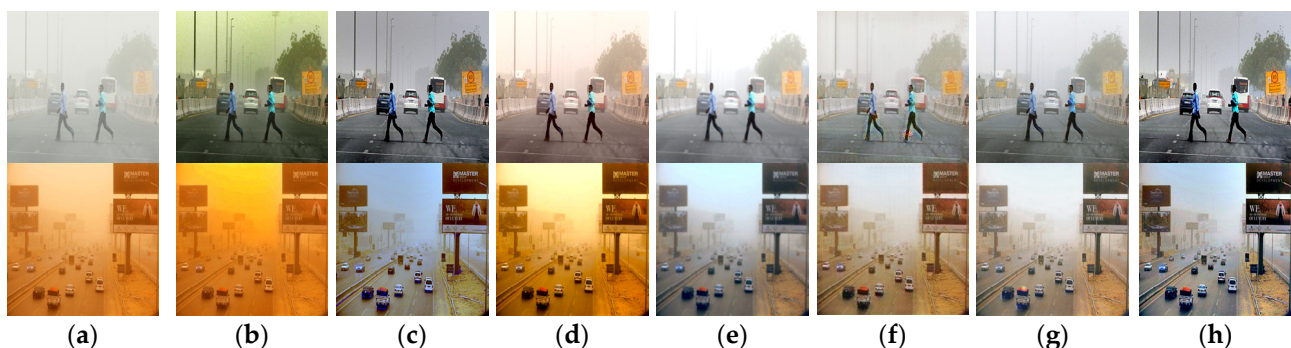
Figure 4. Qualitative comparisons with SOTA methods on two real sand dust images. (**a**) Input; (**b**) DCP; (**c**) MSRCR; (**d**) RGHS; (**e**) Park; (**f**) TBNN; (**g**) USDR-Net; (**h**) Ours.

Moving to Figure 5, we present enhancement results for synthetic and real haze images. Most methods display good defogging effects on synthetic haze images, but residual haze remains evident in enhancing real haze images. TBNN attains remarkable visual quality. In comparison, our method removes more haze and restores clearer image details.

Figure 6 vividly displays the enhancement outcomes for underwater images. In contrast to other methods, our approach successfully mitigates underwater light refraction-induced occlusion and color bias issues, effectively restoring image details. By delving into the connections and characteristics among the three image categories, we employ a multi-task knowledge processing mechanism to avoid interference between different tasks. This allows us to fully exploit these connections, enhancing the performance of individual tasks and ultimately achieving outstanding results.
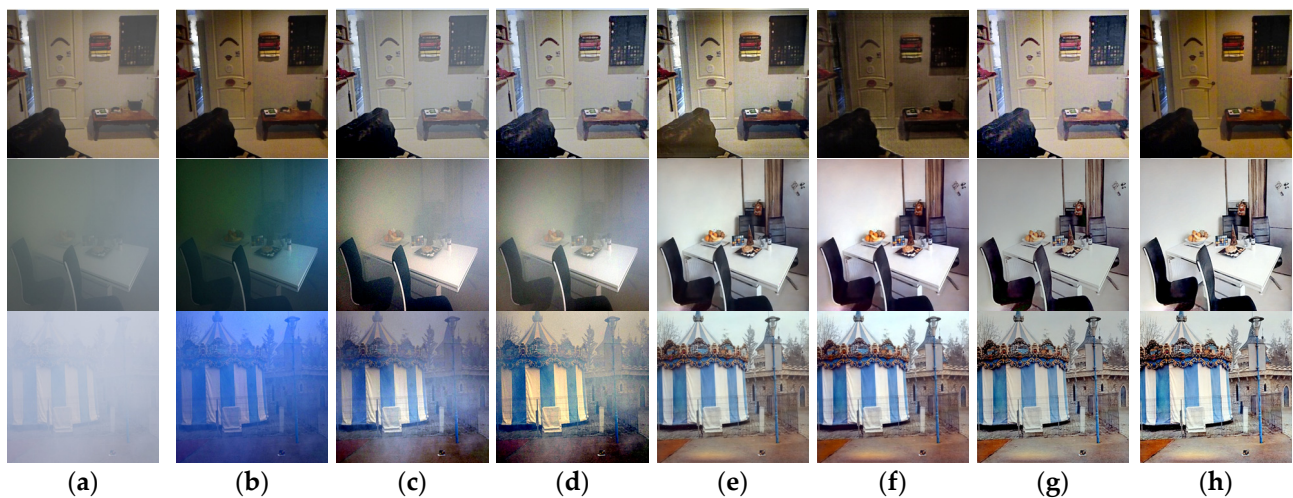
Figure 5. Qualitative comparisons with SOTA methods on three hazy images. (The upper image is a synthesized hazy image, and the two lower images are authentic simulated hazy images.) (**a**) Input; (**b**) DCP; (**c**) MSRCR; (**d**) ACE; (**e**) TBNN; (**f**) DehazeC; (**g**) FSAD; (**h**) Ours.
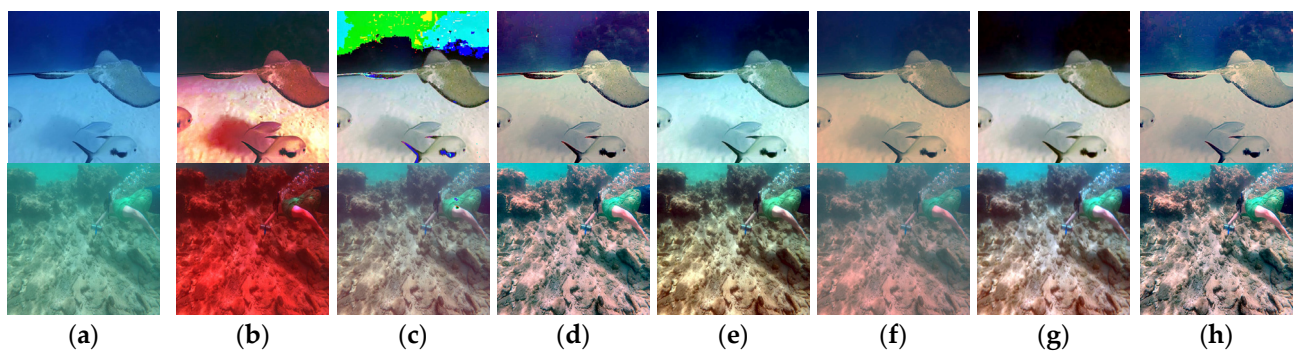

Figure 6. Qualitative comparisons with SOTA methods on two real underwater images. (**a**) Input; (**b**) RCP; (**c**) Park; (**d**) FUnIEGAN; (**e**) UWnet; (**f**) NUˆ2NET; (**g**) PUIE-NET; (**h**) Ours.

5. Analysis and Discussion

In the previous section, we demonstrated that the proposed model's characteristics can yield results comparable to advanced methods. In this section, we will delve into the analysis of the proposed approach through ablation experiments and showcase the effects of key modules and designs. The data presented in the table are based on dust-laden image data, serving as a representative example to illustrate the changes in evaluation metrics.

5.1. Effectiveness of FSGS-ViT

To validate the effectiveness of several key designs in FSGS-ViT, we conducted a series of ablation experiments to explore the impact of different designs on performance. Specifically, we performed the following ablation experiments: (1) ablating the entire FSGS-ViT and using the traditional ViT structure as a replacement; (2) ablating the lower branch GSB in the FSGS-ViT structure; (3) removing the Conv block design within GSB.

In Figure 7, we present the results without using FSGS-ViT. It is evident that the image evaluation metrics are lower without the use of FSGS-ViT. Additionally, the results of the ablation experiments for GSB and its Conv block are shown in Table 4, illustrating varying degrees of degradation in evaluation metrics. This robustly demonstrates the effectiveness of these key designs.
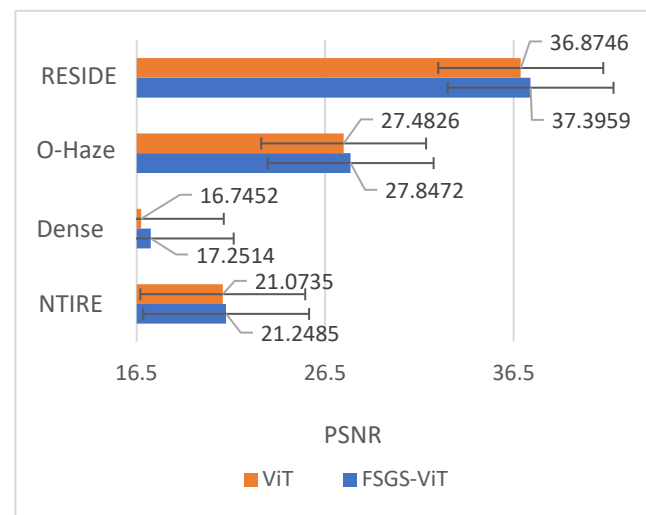
Figure 7. Ablation analysis for ViT on the benchmarks.

Table 4. In regard to the ablation experiments for the FSGS-ViT.

Datasets	(2) e-Score/r-Score	(3) e-Score/r-Score	FSGS-ViT e-Score/r-Score
D-Sand	1.1329/2.0283	1.1938/2.2039	1.2308/2.2216

5.2. Effectiveness of MDFN

To validate the effectiveness of the proposed MDFN method, we compared it with three other feedforward neural networks: (1) Traditional Feedforward Network (FN) [59], (2) Gated Deep Feedforward Network (GDFN) [60], and (3) Multi-Scale Feedforward Network (MSFN) [42]. We conducted quantitative analysis on the dataset, and the results are presented in Table 5. GDFN introduces a gating mechanism to achieve performance advantages, but it still does not fully consider the learning of multi-scale knowledge. MSFN improves image restoration performance by integrating and fusing local features from different scales, but the impact of irrelevant information is not fully addressed.

Table 5. In regard to the ablation experiments for the MDFN.

Datasets	(FN 1) e-Score/r-Score	(GDFN 2) e-Score/r-Score	(MSFN 3) e-Score/r-Score	Mdfn Our e-Score/r-Score
D-Sand	1.1993/2.2029	1.2241/2.2173	1.2284/2.2203	1.2308/2.2216

Our proposed MDFN, by incorporating parameter matrices and a frequency–domain combination on top of multi-scale features, achieves a superior gating-like information selection mechanism. As seen in Table 5, MDFN achieves the best performance in evaluation metrics, achieving significant numerical gains compared to MSFN.

5.3. Effectiveness of MKSM

In the design process of MKSM, we made several critical design decisions. We conducted crucial investigations through ablation experiments on the following two aspects: (1) Ablating MKSM and the knowledge allocation method, using the traditional Multi-Task Learning (MTL) mechanism, and (2) removing the multi-scale structure design from MKSM. The results of these ablation experiments are presented in Table 6 on the dataset.

We observed that the network's performance declined to varying degrees after ablating the mentioned structures individually. Particularly, after ablating MKSM and the knowledge allocation method, the network's performance notably decreased, underscoring

the efficacy of our analysis of inter-class relationships and uniqueness. Our design yields the current optimal enhancement results.

Table 6. In regard to the ablation experiments for the MKSM.

Datasets	(MTL 1) e-Score/r-Score	(2) e-Score/r-Score	MKSM Our e-Score/r-Score
D-Sand	1.1204/2.0741	1.2205/2.1843	1.2308/2.2216

5.4. Effectiveness of AGMM

To validate the effectiveness of the proposed AGMM method, we conducted further ablation experiments comparing the AGMM module with other methods, including: (1) Traditional Hybrid Residual Connection (HRC), and (2) Adaptive Mixing Module (AMM) [51]. The experimental results are presented in Table 7.

Table 7. In regard to the ablation experiments for the AGMM.

Datasets	(HRC 1) e-Score/r-Score	(AMM 2) e-Score/r-Score	AGMM Our e-Score/r-Score
D-Sand	1.1946/2.1388	1.2241/2.1774	1.2308/2.2216

HRC achieved certain results by enhancing the neural network's utilization of shallow and deep information, but it did not consider the exclusion of redundant information. AMM better utilized shallow information by adaptively fusing information from downsampling and upsampling layers, but it uniformly adapted the fusion of all information without prior information selection.

In contrast, our AGMM method efficiently uses shallow information through two steps: information selection and adaptive fusion. In the experiments, compared to AMM, our method achieved a performance gain.

6. Conclusions

In this paper, we propose a multi-task learning strategy for three types of environmentally biased color images. Within this context, we introduce several innovative key designs, primarily focused on information selection and the exploration of interconnections and distinct characteristics between different tasks. We integrate multi-task learning and knowledge allocation strategies into the network model, harnessing the relationships between these three image categories and their respective attributes. This complementary approach enhances the overall model performance and individual task enhancement effects. By doing so, we avoid performance degradation caused by mutual interference among different tasks, allowing private and shared information to synergistically boost network performance. Furthermore, to further enhance performance through information filtering, we introduce multiple innovative network structures, including the Frequency Domain Similarity-Gated Selection Vision Transformer and the Mixed-Scale Frequency Domain Feedforward Network. These structures leverage gating operations, parameter matrices, and frequency domain fusion methods to effectively integrate shallow and deep information, thereby enhancing the network's modeling capabilities for both remote and local patterns. Extensive experiments are conducted on benchmark datasets for the three image categories. The results demonstrate the significant efficacy of our proposed model and methods in tasks such as dust removal, haze removal, and underwater image enhancement.

Limitations. While our proposed method outperforms the compared methods in terms of performance, there are still some unresolved issues in terms of image enhancement effects. Our model has limitations in terms of image color restoration. In a significant number of experiments, we observed that the color of some images did not fully recover to the same level as the real images. We speculate that this could be related to the unsupervised training process we adopted, and the lack of supervised data also limits our ability to address this

issue. Nonetheless, due to the flexibility of unsupervised learning, these shortcomings are acceptable in cases where paired datasets are incomplete, and overall data volume is limited. In the future, we will focus on this issue and attempt to introduce effective methods to address it. We plan to explore new training strategies, data augmentation techniques, or the incorporation of richer supervised information to further improve our model's performance in image color restoration. This will help our method exhibit even better performance in a broader range of application scenarios.

Author Contributions: Y.D.: Project administration, Supervision, Conceptualization, Code, Writing—review and editing. K.W.: Funding acquisition, Methodology, Investigation, Software, Writing—original draft. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request. The data are not publicly available due to privacy.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

Abbreviation	Full Form
MTL	multi-task learning
FSGS-ViT	Frequency Domain Similarity-Gated Selection Vision Transformer
MDFN	Mixed-Scale Frequency Domain Feedforward Network
DCT	Discrete Cosine Transform
MKSM	Hybrid-Scale Knowledge Selection Module
AGMM	Adaptive Gate Mixed Module
SOTA	state-of-the-art
CNNs	Convolutional Neural Networks
ASM	Atmospheric Scattering Model
UDCP	underwater dark channel prior
FAB	Frequency-domain Adaptive Attention Block
GSB	Gate Selection Block
DW	Depthwise Separable
FNNs	Feedforward Neural Networks
MKSM	Mixed-Scale Knowledge Selection Modules
EB	Enhancement Modules
GEGLU	Gated Exponential Linear Units
FN	Traditional Feedforward Network
GDFN	Gated Deep Feedforward Network
MSFN	Multi-Scale Feedforward Network
HRC	Traditional Hybrid Residual Connection
AMM	Adaptive Mixing Module

References

1. Irani, B.; Wang, J.; Chen, W. A localizability constraint-based path planning method for autonomous vehicles. *IEEE Trans. Intell. Transp. Syst.* **2019**, *20*, 2593–2604. [CrossRef]
2. Ding, Y.; Wu, K. Sand-dust image enhancement using RGB color balance method. *Opt. Precis. Eng.* **2023**, *31*, 1053–1064. [CrossRef]
3. Chen, W.-T.; Huang, Z.-K.; Tsai, C.-C.; Yang, H.; Ding, J.-J.; Kuo, S.-Y. Learning multiple adverse weather removal via two-stage knowledge learning and multi-contrastive regularization: Toward a unified model. In Proceedings of the CVPR, New Orleans, LA, USA, 19–24 June 2022.

4. Li, B.; Liu, X.; Hu, P.; Wu, Z.; Lv, J.; Peng, X. All-in-one image restoration for unknown corruption. In Proceedings of the CVPR, New Orleans, LA, USA, 19–24 June 2022.
5. Tan, R.T. Visibility in bad weather from a single image. In *Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, USA, 23–28 June 2008*; IEEE: Piscataway, NJ, USA; pp. 1–8.
6. He, K.; Sun, J.; Tang, X. Single image haze removal using dark channel prior. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *33*, 2341–2353.
7. Wu, D.; Ge, X.Y. Research and implementation of image haze removal method based on dark channel prior. *J. Shenyang Norm. Univ. Nat. Sci. Ed.* **2018**, *36*, 82–86.
8. Xiao, J.S.; Gao, W.; Zou, B.Y.; Yao, Y.; Zhang, Y.Q. Image dehazing based on sky-constrained dark channel prior. *Acta Electron. Sin.* **2018**, *47*, 346–352.
9. Yang, H.; Cui, Y. Image defogging algorithm based on opening dark channel and improved boundary constraint. *Acta Photonica Sin.* **2018**, *47*, 244–250.
10. Fattal, R. Dehazing using color-lines. *ACM Trans. Graph. TOG* **2014**, *34*, 1–14. [CrossRef]
11. 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. [CrossRef]
12. Li, B.; Peng, X.; Wang, Z.; Xu, J.; Feng, D. AOD-Net: All-in-One Dehazing Network. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 4780–4788. [CrossRef]
13. Qin, X.; Wang, Z.; Bai, Y.; Xie, X.; Jia, H. FFA-Net: Feature Fusion Attention Network for Single Image Dehazing. *Proc. AAAI Conf. Artif. Intell.* **2020**, *34*, 11908–11915. [CrossRef]
14. Singh, A.; Bhave, A.; Prasad, D.K. Single images dehazing for a variety of haze scenarios using back projected pyramid network. In Proceedings of the Computer Vision—ECCV 2020 Workshops, Glasgow, UK, 23–28 August 2020; pp. 166–181.
15. Das, S.D.; Dutta, S. Fast Deep Multi-patch Hierarchical Network for Nonhomogeneous Image Dehazing. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, WA, USA, 14–19 June 2020; pp. 1994–2001. [CrossRef]
16. Yu, Y.; Liu, H.; Fu, M.; Chen, J.; Wang, X.; Wang, K. A Two-branch Neural Network for Non-homogeneous Dehazing via Ensemble Learning. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Nashville, TN, USA, 19–25 June 2021; pp. 193–202. [CrossRef]
17. Wu, K.; Ding, Y. Image Dehazing via Double-layer Vision and Multi-scale Attention Fusion. *J. Hunan Univ. Nat. Sci.* **2023**, *50*, 40–51.
18. Wu, K.; Ding, Y. Multi focus image fusion Model based on Deep unsupervised learning. *J. Hunan Univ. Nat. Sci.* **2023**, 1–10.
19. Land, E.H. The retinex theory of color vision. *Sci. Am.* **1978**, *237*, 108–128. [CrossRef]
20. Liu, C.; Chen, X.; Wu, Y.R. A modified grey world method to detect and restore colour cast images. *IET Image Process.* **2019**, *13*, 1090–1096. [CrossRef]
21. Zuiderveld, K. *Contrast Limited Adaptive Histogram Equalization*; Academic Press Professional, Inc.: Cambridge, MA, USA, 1994; pp. 474–485.
22. Xiao, J.F.; Yu, X.Y.; Chen, B.; Xun, X.G.; Liu, B. Image enhancement algorithm based on wavelet transform and fuzzy set theory. *J. Proj. Rocket. Missiles Guid.* **2010**, *30*, 183–186.
23. Yan, T.; Wang, L.; Wang, J. Method to enhance degraded image in dust environment. *J. Softw.* **2014**, *9*, 2672–2677. [CrossRef]
24. Shi, Z.; Feng, Y.; Zhao, M.; Zhang, E.; He, L. Normalised gamma transformation-based contrast-limited adaptive histogram equalisation with colour correction for sand-dust image enhancement. *IET Image Process.* **2020**, *14*, 747–756. [CrossRef]
25. Gao, G.X.; Lai, H.C.; Jia, Z.H.; Liu, Y.Q.; Wang, Y.L. Sand-dust image restoration based on reversing the blue channel prior. *IEEE Photonics J.* **2020**, *99*, 1–16. [CrossRef]
26. Ding, Y.; Wu, K. Sand dust degradation images enhancement algorithm via multi-branch restoration network. *Comput. Eng. Appl.* **2023**, *59*, 1–13. [CrossRef]
27. Chaitanya, B.S.N.V.; Mukherjee, S. Chaitanya and Snehasis Mukherjee, Single image dehazing using improved cycleGAN. *J. Vis. Commun. Image Represent.* **2021**, *74*, 103014. [CrossRef]
28. Li, C.; Tang, S.; Kwan, H.K.; Yan, J.; Zhou, T. Color Correction Based on CFA and Enhancement Based on Retinex with Dense Pixels for Underwater Images. *IEEE Access* **2020**, *8*, 155732–155741. [CrossRef]
29. Hitam, M.S.; Awalludin, E.A.; Yussof, W.N.H.; Bachok, Z. Mixture contrast limited adaptive histogram equalization for underwater image enhancement. In *Proceedings of the International Conference on Computer Applications Technology, New York, NY, USA, 20–22 January 2013*; IEEE: Piscataway, NJ, USA; pp. 1–5.
30. Tang, Z.; Jiang, L.; Luo, Z. A new underwater image enhancement algorithm based on adaptive feedback and Retinex algorithm. *Mult Imedia Tools Appl.* **2021**, *80*, 28487–28499. [CrossRef]
31. Carlevaris-Bianco, N.; Mohan, A.; Eustice, R.M. Initial results in underwater single image dehazing. In *Proceedings of the OCEANS 2010 MTS/IEEE SEATTLE, New York, NY, USA, 20–23 September 2010*; IEEE: Piscataway, NJ, USA; pp. 1–8.
32. Drews, P.L.; Nascimento, E.R.; Botelho, S.S.; Campos, M.F. Underwater Depth Estimation and Image Restoration Based on Single Images. *IEEE Comput. Graph. Appl.* **2016**, *36*, 24–35. [CrossRef]
33. Anwar, S.; Li, C.Y.; Porikli, F. Deep Underwater Image Enhancement. In *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018*; IEEE: Piscataway, NJ, USA; pp. 1–12.

34. Li, C.; Guo, J.; Guo, C. Emerging from Water: Underwater Image Color Correction Based on Weakly Supervised Color Transfer. *IEEE Signal Process. Lett.* **2018**, *25*, 323–327. [CrossRef]

35. Fabbri, C.; Islam, M.J.; Sattar, J. Enhancing Underwater Imagery Using Generative Adversarial Networks. In *Proceedings of the 2018 IEEE International Conference on Robotics and Automation (ICRA), Los Alamitos, CA, USA, 21–25 May 2018*; IEEE Computer Society: Piscataway, NJ, USA; pp. 7159–7165.

36. Liu, X.; Gao, Z.; Chen, B.M. MLFcGAN: Multilevel Feature Fusion-Based Conditional GAN for Underwater Image Color Correction. *IEEE Geosci. Remote Sens. Lett.* **2020**, *17*, 1488–1492. [CrossRef]

37. Naik, A.; Swarnakar, A.; Mittal, K. Shallow-UWnet: Compressed Model for Underwater Image Enhancement. In Proceedings of the 35th AAAI Conference on Artificial Intelligence, Vancouver, BC, Canada, 2–9 February 2021; AAAI: Washington, DC, USA; Volume 18, pp. 15853–15854.

38. Li, R.; Robby; Tan, T.; Cheong, L.-F. All in one bad weather removal using architectural search. In Proceedings of the CVPR. Seattle, WA, USA, 16–18 June 2020.

39. Ranjan, R.; Vishal; Patel, M.; Chellappa, R. Hyperface: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *41*, 121–135. [CrossRef]

40. Sun, X.; Panda, R.; Feris, R.; Saenko, K. Adashare: Learning what to share for efficient deep multi-task learning. In Proceedings of the NeurIPS, Virtual, 6–12 December 2020.

41. Gao, Y.; Ma, J.; Zhao, M.; Liu, W.; Yuille, A.L. Nddr-cnn: Layerwise feature fusing in multi-task cnns by neural discriminative dimensionality reduction. In Proceedings of the CVPR, Long Beach, CA, USA, 16–20 June 2019.

42. Chen, X.; Li, H.; Li, M.; Pan, J. Learning A Sparse Transformer Network for Effective Image Deraining. In Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 17–24 June 2023; pp. 5896–5905.

43. Wang, Y.; Ma, C.; Liu, J. SmartAssign: Learning A Smart Knowledge Assignment Strategy for Deraining and Desnowing. In Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 17–24 June 2023; pp. 3677–3686.

44. Kong, L.; Dong, J.; Ge, J.; Li, M.; Pan, J. Efficient Frequency Domain-based Transformers for High-Quality Image Deblurring. In Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 17–24 June 2023; pp. 5886–5895. [CrossRef]

45. Li, B.; Ren, W.; Fu, D.; Tao, D.; Feng, D.; Zeng, W.; Wang, Z. Benchmarking single-image dehazing and beyond. *IEEE Trans. Image Process.* **2018**, *28*, 492–505. [CrossRef] [PubMed]

46. Ancuti, C.O.; Ancuti, C.; Timofte, R.; De Vleeschouwer, C. O-HAZE: A dehazing benchmark with real hazy and haze-free outdoor images. *arXiv* **2018**, 754–762.

47. Ancuti, C.O.; Ancuti, C.; Sbert, M.; Timofte, R. Dense-haze: A benchmark for image dehazing with dense-haze and haze-free images. In *Proceedings of the 2019 IEEE International Conference on Image Processing, Taipei, China, 22–25 September 2019*; IEEE: Piscataway, NJ, USA; pp. 1014–1018.

48. Ancuti, C.O.; Ancuti, C.; Vasluianu, F.A.; Timofte, R. NTIRE 2020 challenge on NonHomogeneous dehazing. In *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, WA, USA, 14–19 June 2020*; IEEE: Piscataway, NJ, USA; p. 20292044.

49. Islam, M.J.; Edge, C.; Xiao, Y.; Luo, P.; Mehtaz, M.; Morse, C.; Enan, S.S.; Sattar, J. Semantic Segmentation of Underwater Imagery: Dataset and Benchmark. In Proceedings of the 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Las Vegas, NV, USA, 24 October–24 January 2020; pp. 1769–1776. [CrossRef]

50. Li, C.; Guo, C.; Ren, W.; Cong, R.; Hou, J.; Kwong, S.; Tao, D. An Underwater Image Enhancement Benchmark Dataset and beyond. *IEEE Trans. Image Process.* **2020**, *29*, 4376–4389. [CrossRef] [PubMed]

51. Liu, R.; Fan, X.; Zhu, M.; Hou, M.; Luo, Z. Real-world underwater enhancement: Challenges, benchmarks, and solutions under natural light. IEEE Trans. *Circuits Syst. Video Technol.* **2020**, *30*, 4861–4875. [CrossRef]

52. Jaisurya, R.S.; Mukherjee, S. Attention-based Single Image Dehazing Using Improved CycleGAN. In Proceedings of the 2022 International Joint Conference on Neural Networks (IJCNN), Padua, Italy, 18–23 July 2022; pp. 1–8. [CrossRef]

53. Park, T.H.; Eom, I.K. Sand-Dust Image Enhancement Using Successive Color Balance with Coincident Chromatic Histogram. *IEEE Access* **2021**, *9*, 19749–19760. [CrossRef]

54. Ding, B.; Chen, H.; Xu, L.; Zhang, R. Restoration of Single Sand-Dust Image Based on Style Transformation and Unsupervised Adversarial Learning. *IEEE Access* **2022**, *10*, 90092–90100. [CrossRef]

55. Engin, D.; Genc, A.; Ekenel, H.K. Cycle-Dehaze: Enhanced CycleGAN for Single Image Dehazing. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT, USA, 18–23 June 2018; pp. 938–9388. [CrossRef]

56. Islam, M.J.; Xia, Y.; Sattar, J. Fast Underwater Image Enhancement for Improved Visual Perception. *IEEE Robot. Autom. Lett.* **2020**, *5*, 3227–3234. [CrossRef]

57. Guo, C.; Wu, R.; Jin, X.; Han, L.; Zhang, W.; Chai, Z.; Li, C. Underwater Ranker: Learn Which is Better and How to Be Better. *arXiv* **2022**, arXiv:2208.06857. [CrossRef]

58. Fu, Z.; Wang, W.; Huang, Y.; Ding, X.; Ma, K.K. Uncertainty Inspired Underwater Image Enhancement. *Lect. Notes Comput. Sci.* **2022**, *13678LNCS*, 465–482. [CrossRef]

59. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An image is worth 16x16 words: Transformers for image recognition at scale. In Proceedings of the ICLR, Vienna, Austria, 4 May 2021.

60. Zamir, S.W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F.S.; Yang, M.H. Restormer: Efficient transformer for high-resolution image restoration. In Proceedings of the CVPR, New Orleans, LA, USA, 19–24 June 2022; pp. 5728–5739.