



Article A Lightweight Neural Network for the Real-Time Dehazing of Tidal Flat UAV Images Using a Contrastive Learning Strategy

Denghao Yang ¹, Zhiyu Zhu ^{1,*}, Huilin Ge ¹, Haiyang Qiu ², Hui Wang ² and Cheng Xu ¹

- ¹ Automation College, Jiangsu University of Science and Technology, Zhenjiang 212100, China; 211210301120@stu.just.edu.cn (D.Y.)
- ² School of Naval Architecture and Ocean Engineering, Guangzhou Maritime University, Guangzhou 510725, China
- * Correspondence: zhuzy@just.edu.cn

Abstract: In the maritime environment, particularly within tidal flats, the frequent occurrence of sea fog significantly impairs the quality of images captured by unmanned aerial vehicles (UAVs). This degradation manifests as a loss of detail, diminished contrast, and altered color profiles, which directly impact the accuracy and effectiveness of the monitoring data and result in delays in the execution and response speed of monitoring tasks. Traditional physics-based dehazing algorithms have limitations in terms of detail recovery and color restoration, while neural network algorithms are limited in their real-time application on devices with constrained resources due to their model size. To address the above challenges, in the following study, an advanced dehazing algorithm specifically designed for images captured by UAVs over tidal flats is introduced. The algorithm integrates dense convolutional blocks to enhance feature propagation while significantly reducing the number of network parameters, thereby improving the timeliness of the dehazing process. Additionally, an attention mechanism is introduced to assign variable weights to individual channels and pixels, enhancing the network's ability to perform detail processing. Furthermore, inspired by contrastive learning, the algorithm employs a hybrid loss function that combines mean squared error loss with contrastive regularization. This function plays a crucial role in enhancing the contrast and color saturation of the dehazed images. Our experimental results indicate that, compared to existing methods, the proposed algorithm has a model parameter size of only 0.005 M and a latency of 0.523 ms. When applied to the real tidal flat image dataset, the algorithm achieved a peak signal-tonoise ratio (PSNR) improvement of 2.75 and a mean squared error (MSE) reduction of 9.72. During qualitative analysis, the algorithm generated high-quality dehazing results, characterized by a natural enhancement in color saturation and contrast. These findings confirm that the algorithm performs exceptionally well in real-time fog removal from UAV-captured tidal flat images, enabling the effective and timely monitoring of these environments.

Keywords: unmanned aerial vehicle (UAV); tidal flat monitoring; image dehazing; attention mechanism; contrastive learning

1. Introduction

In examining tidal flats [1], unmanned aerial vehicles (UAVs) have many advantages, such as a wide field of view, excellent perspective, strong mobility, and rapid response, along with a wide patrol area. These capabilities allow a UAV to effectively monitor and track targets, facilitating the efficient execution of surveillance tasks within the intricate and dynamic settings of tidal flats [2–5]. Additionally, UAV aerial photography has been widely utilized in disaster monitoring and response [6,7], wildlife monitoring and conservation [8–10], urban planning and management [11], and various detection applications such as wetland feature detection [12], power facility inspection [13], forest fire detection [14], and environmental monitoring [15]. However, all of these applications



Citation: Yang, D.; Zhu, Z.; Ge, H.; Qiu, H.; Wang, H.; Xu, C. A Lightweight Neural Network for the Real-Time Dehazing of Tidal Flat UAV Images Using a Contrastive Learning Strategy. *Drones* 2024, *8*, 314. https:// doi.org/10.3390/drones8070314

Academic Editor: Diego González-Aguilera

Received: 27 May 2024 Revised: 5 July 2024 Accepted: 7 July 2024 Published: 10 July 2024



Copyright: © 2024 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 (https:// creativecommons.org/licenses/by/ 4.0/). are susceptible to reduced visual effects, as well as a decreased efficiency and accuracy of recognition and monitoring due to fog. Particularly in areas near tidal flats where fog is frequent, the images captured by UAVs can be severely degraded, leading to a loss of detail, reduced contrast, and color distortion. This degradation compromises the subsequent extraction of image information, highlighting the need to effectively process distorted images during tidal flat monitoring.

Image dehazing algorithms [16–19] can be divided into image dehazing algorithms based on traditional methods [20,21] and deep learning-based image dehazing algorithms [22–30]. Traditional image dehazing algorithms are generally based on the atmospheric scattering physics model to obtain the accurate atmospheric light value and transmission coefficient. One of the most representative algorithms is that of He et al., who based their design on the dark channel prior (DCP) [20]. The dehazing technique in question leverages the dark channel for estimating the transmission map and applies an atmospheric scattering model to reconstruct a fog-free image, thereby facilitating the removal of fog. While this approach yields a relatively precise transmission map, the algorithm's complexity in terms of both time and computational space is excessive, making it impractical for real-world use. As the field of artificial intelligence progresses, deep learning-based dehazing algorithms have emerged as a popular area of research. The most representative fog algorithm is that of Cai et al., who proposed the convolutional neural network model DehazeNet [23]. They took the fog image as input, the learned parameters such as transmission as output, reversed the degradation process through the atmospheric model, and finally obtained the fog image. Although the subjective vision of the human eye can intuitively see the obvious effect, the fidelity of the detail texture structure is not high enough, and the effect of fog hinders image quality. Qin et al. proposed a residual network combined with the attention mechanism (feature fusion attention network, FFA-Net [26]) to determine the residual features of fog and output the image after removing the fog. Although this network does not rely on the atmospheric scattering model, the robustness of the model is not strong, and it is prone to incomplete fog removal and noise problems. A simple and efficient feature fusion convolutional neural network, AOD-Net [24], was proposed by Li et al. as a dehazing algorithm; as the time consumption is low, it is suitable for practical application. However, the algorithm is prone to image detail loss, color deviation, and other phenomena.

Tidal flats are dynamic landscapes that change with the tides, and their surroundings typically include silt land, water, and plants, which contribute to the complexity and rich texture of tidal topography images. The features around a beach exhibit a variety of colors. Consequently, when processing tidal terrain images to perform dehazing, it is important to ensure the recovery of image contrast and saturation, as well as the maintenance of discernible and monitorable feature information within the foggy tidal flat images. Existing neural network methods often encounter challenges such as incomplete dehazing, loss of detail information, color distortion, and high computational time costs when processing tidal flat images. To address these issues, we propose a lightweight convolutional neural network with a contrastive learning strategy [31–33] that balances performance and efficiency. The main contributions of the present study can be summarized as follows:

• In current dehazing network research, there is often a trade-off between efficiency and performance. Traditional models primarily use single-scale convolutional kernels, which limit their ability to capture multi-scale features within images. This paper introduces a method employing multi-scale convolution, enabling the network to recognize different scale features more comprehensively in the image, thereby enhancing the indepth understanding of image semantics. At the same time, to address the deficiencies of traditional dehazing networks in information transmission and reuse, this study incorporates dense and residual connections, which optimize the flow of information, reduce parameter redundancy, and accelerate training speed. Through this approach, we ensure that the model can significantly enhance the dehazing performance for tidal flat images while maintaining real-time image processing capabilities.

- Due to the vast shooting range of UAVs, remote sensing images captured by drones often cover a complex and diverse range of scenes and changes. Traditional dehazing networks rely on fixed attention distributions or simple regional weighting. To enhance the network's processing capability for tidal flat terrain remote sensing images, the attention mechanism was improved in the presented study. The improved mechanism is able to adaptively adjust the network's focus based on the differences in feature information across various regions in the tidal flat images, guiding the network to focus on critical areas that are crucial to the dehazing effect. Through this, it not only preserves the texture details of the image more effectively, improving visual quality, but also enhances the stability and reliability of the dehazing process.
- The design of a loss function typically relies on simple loss functions such as mean squared error (MSE). However, the contribution of the present study lies in the design of a new composite loss function that combines contrastive learning strategies. This loss function, compared to a single loss function, is able to avoid the overfitting phenomenon of the network model and more effectively reduce the differences between the generated image and the clear image, thereby achieving more realistic color restoration while removing fog. The above provides a foundation for the subsequent effective monitoring of tidal flat terrain.

2. Preparation Knowledge

2.1. Atmospheric Scattering Physics Model and Conversion Formula

The atmospheric light scattering model [34,35] is a description model for classically generated haze images, expressed as follows:

$$I(x) = J(x)t(x) + A(1 - t(x))$$
(1)

where I(x) is the haze image, J(x) is the scene radiation (namely the clear image to be recovered), and x represents the pixel index; there are two key parameters: A representing the atmospheric light value and t(x) representing the transmission matrix, and the expression is as follows:

$$e(x) = e^{-\beta d(x)} \tag{2}$$

where β is the scattering coefficient of atmospheric light, and d(x) is the distance between the object and the camera.

Below is the clear image generation model:

$$J(x) = \frac{1}{t(x)}I(x) - A\frac{1}{t(x)} + A$$
(3)

Since t(x) and A are unknown, evaluating them separately will cause the accumulation or even magnification of errors. Inspired by AOD-Net [24], the core idea of the present study is to unify the two parameters t(x) and A into a formula, which is K(x) in Formula (4), and directly minimize the pixel domain reconstruction error. To achieve this, the Formula (3) is reformulated as the following transformation formula:

$$K(x) = \frac{\frac{1}{t(x)}(I(x) - A) + (A - b)}{I(x) - 1}$$
(4)

$$J(x) = K(x)I(x) - K(x) + b$$
(5)

2.2. Physics-Aware Dehazing Neural Network

Based on prior information about the physical model, the designed neural network structure was divided into two main parts. The first part is the K-estimation module, whose core task is to accurately estimate the depth and haze degree K(x) from the input image I(x) through a convolutional neural network. The second part is the clear image generation module, which consists of element-wise multiplication layers and multiple element-wise

addition layers and calculates the restored clear image J(x) through the computation of Formula (5). Figure 1 shows the overall structure diagram of this network.



Figure 1. Physics-aware dehazing neural network.

3. Modeling and Extension

3.1. Overall Structure of Multi-Scale Dense Residual Convolution Networks

To ensure the real-time fog removal in tidal beach images, it is necessary to design the overall network architecture with the lightweight neural network model so that the network model can establish a better balance between efficiency and performance.

First, to better capture the learning features of different scales, the image transformation of different scales is processed, and the multi-scale features of the images are leveraged. It is necessary to design a multi-scale network model to improve the feature representation ability, generalization ability, and performance of the network model. Traditional multi-scale network architectures, such as Feature Pyramid Networks (FPNs), typically incorporate multiple residual blocks, with each block using convolutional kernels of the same size to extract features. This architecture is relatively complex and often requires a substantial number of computational resources. In contrast, the network model designed in the present study extracts multi-scale features from images using convolutional kernels of different sizes, and each size of the convolutional kernel is applied only once, thus requiring fewer computational resources to achieve multi-scale feature extraction. Furthermore, dense connections can greatly reduce the number of network parameters, enhance feature reuse, propagation, and gradient flow, make the network structure more compact and simpler, improve the training efficiency of the model, and improve the processing ability of the network to image fog. Therefore, we connected these convolution kernels of different sizes through a dense connection. The dense connection mathematical expression is as follows:

$$x_i = H([x_1, x_2, \dots, x_{i-1}])$$
 (6)

Parameter description: H represents the nonlinear transformation, and x_i represents the output characteristics of i convolutional layers.

To fulfill the model's lightweight requirement, we strategically integrated only two residual connections at pivotal points within the network architecture. One connection performs a residual linkage between the input image and the output features, with the purpose of preserving the original image characteristics. The second connection is positioned before and after the attention mechanism, aiming to enhance the network's expressive and capturing capabilities of features during critical processing stages, thereby ensuring the efficient flow and accurate mapping of feature information within the network.

The detailed architecture of the specific network model is shown in Figure 2. For a given input foggy image, the feature mapping relationship eventually recovers into a fog-free image. This process is mathematically expressed as follows:

$$F = concat(x_1, x_2, x_3, x_4, x_5)$$
(7)

$$x_1 = Relu\left(f^{1 \times 1}(x_0)\right) \tag{8}$$

$$x_2 = Relu(f^{3\times3}(x_1)) \tag{9}$$

$$x_3 = Relu\left(f^{5\times 5}(x_1, x_2)\right) \tag{10}$$

$$x_4 = Relu(f^{7 \times 7}(x_1, x_2, x_3))$$
(11)

$$x_5 = Relu(f^{9 \times 9}(x_1, x_2, x_3, x_4))$$
(12)

$$F' = F + F \otimes T(F) \tag{13}$$

$$C = f^{3\times3} \left(x_0 + \operatorname{Relu}\left(f^{3\times3}(F') \right) \right)$$
(14)



Figure 2. The improved network structure shown in the present paper.

In the formula, x_0 represents the foggy image; x_1, x_2, x_3, x_4, x_5 represent the feature maps' output after different convolutional layers; $f^{1\times 1}$, $f^{3\times 3}$, $f^{5\times 5}$, $f^{7\times 7}$, $f^{9\times 9}$ represent the convolutional kernels with sizes of 1, 3, 5, 7, and 9, respectively; *C* represents the recovered clear image; and T(F) represents the feature extracted by the attention mechanism. The internal network structure and mathematical expression will be introduced in detail below.

3.2. Channel and Spatial Attention Module (CSAM)

Tidal flats often have many different shape areas; as such, tidal flat images contain rich texture detail information. An attention mechanism can provide better attention and utilize multiple scale features in an image, better capture and employ the details in the image, meaning that the details of the image can be better retained, and eliminate redundant information to improve network detail retention and fog ability; in light of the above, we used the residual network in the process of designing the network and attention mechanism module. The attention mechanism introduced includes two parts: the improved channel attention module (CAM) and the improved spatial attention module (SAM). The specific connection mode of the two attention modules and the network is shown in Figure 3. This process is mathematically represented as follows:

$$F' = F + F \otimes T(F) \tag{15}$$

$$T(F) = \sigma(T_C(F) + T_S(F))$$
(16)

where \otimes is the Kronecker product, $T_C(F)$ represents the feature after channel attention module processing, $T_S(F)$ represents the feature extracted by the spatial attention module, and σ represents the Sigmoid activation function. Before addition, the output of the two branches is adjusted to $\mathbb{R}^{C \times H \times W}$.



Figure 3. The connection mode of the channel attention module and spatial attention module.

3.2.1. Improved Channel Attention Module (CAM)

To effectively calculate channel attention, traditional methods often employ average pooling to decrease the spatial dimensions of the input feature maps. Research indicates that the maximum pooling layer can capture diverse object features, thus generating more refined attentional features. Therefore, in our subsequent research, we integrated an enhanced channel attention module that combines average and maximum pooling techniques to reduce the spatial dimensions of the image. This integrated strategy significantly improved the network's ability to express image features, thereby enhancing its dehazing performance. After applying average pooling and max pooling techniques, we introduced a convolutional layer with shared parameters, which simplified the network architecture and effectively reduced the model's parameter count. We then accumulated the attention weights for each channel to enhance the precision of the feature response. Following this, batch normalization was applied to keep the attention weights within a reasonable range, improving the network's stability and allowing for a more effective capturing and utilization of key information within the channels.

The specific internal expression of the CAM is shown in Figure 4, and the mathematical expression is as follows:

$$T_{C}(F) = BN(f^{1\times1}(Relu(f^{1\times1}(Avgpool(F))))) + f^{1\times1}(Relu(f^{1\times1}(Maxpool(F))))) = BN(w_{1}(w_{0}Avgpool(F)) + b_{0}) + b_{1} + w_{1}(w_{0}Maxpool(F) + b_{0}) + b_{1}$$
(17)



Channel Attention Module

Figure 4. The internal network structure of the channel attention module.

3.2.2. Spatial Attention Module (SAM)

When extracting spatial information from different regions of a feature map, using both max pooling and average pooling captures the unique detailed features of each area. Therefore, in the process of calculating spatial attention, we apply both pooling techniques along the channel axis to generate two-channel feature maps, which in turn construct feature descriptors with richer spatial details. The leveraging of contextual information plays a decisive role in extracting the spatial location. It is necessary for us to have different larger receptive fields to better utilize the contextual information so as to improve the accuracy of spatial attention. Compared with the traditional convolution, the dilation convolution not only increases the receptive field but also does not affect the size of the feature map, which is more beneficial when constructing an effective spatial feature map. Therefore, we employ three layers of the expansion convolution layer (the expansion rate r is set to 1, 2, and 3) to generate a more accurate spatial attention mapping $T_P(F)$ and emphasize or suppress the features of different spatial locations. Lastly, the batch normalization layer is added to ensure that reasonable spatial attention weights are generated and more attention is paid to the feature details.

The specific structure of the SAM is shown in Figure 5. The mathematical expression is as follows:

$$T_{\mathcal{S}}(F) = BN\left(f_3^{3\times3}\left(f_2^{3\times3}\left(f_1^{3\times3}([Avgpool(F); Maxpool(F)])\right)\right)\right)$$
(18)

where $f_1^{3\times3}$, $f_2^{3\times3}$, $f_3^{3\times3}$ represents a dilated convolution with an expansion rate of 1, 2, and 3.



Spatial Attention Module

Figure 5. The internal network structure of the spatial attention module.

3.3. The Loss Function Improved through Comparison with Regularization

Traditionally, processing foggy images with a single mean square error loss function often results in color distortion and loss of detail. This can cause deviations in monitoring tidal flat terrain and identifying salt marsh plants. Contrast learning optimizes the training process by helping to restore images closer to their true state through comparison with clear and blurry images. In addition, in the process of image dehazing, semantic information can be leveraged as important prior knowledge to help the network mitigate the effects of haze at the location in question and preserve image details and color by exploiting within-target semantic correlations. Contrast learning can make better utilization of this semantic information and improve the fog effect. Therefore, to solve this problem, the network uses the combined loss function obtained by combining the mean square error loss and contrast regularization as the constraint terms to train the network.

The mean square error loss function is an important measure of image quality, which is obtained by averaging the square of the difference between the foggy image and the fog-free image. When the loss function in the defogged image approaches the original image, gradient descent becomes slower, which has the advantage of improving structural similarity. Its expression is as follows:

$$L_{MSE} = \frac{1}{N \times W \times H} \sum_{i=1}^{N} ||J_i - J'_i||^2$$
(19)

where *N* is the number of channels to generate the image, *W* is the width of the image, *H* is the height of the image, J_i is the fog-free image, and J'_i represents the fog-free image generated using the fog network.

Regarding the role of the contrast regularization of the neural network in the fog removal training process, the principle of contrast learning in image fog removal training can be observed in Figure 6.



Contrastive Regularization

Figure 6. Principle of contrastive learning with and without images.

Anchor represents the defogged image after the restoration of the dehazing network, which is a clear image, and negative is the foggy image and multiple fuzzy images that are inconsistent with the clear image. The goal of the regularized R is to minimize the L 1 distance of the anchor to the clear image and to maximize the distance of the anchor to the foggy image. The formula is given as follows:

$$R = \sum_{i=1}^{n} \xi_{i} \frac{\|F_{i}(J) - F_{i}(J')\|_{1}}{\sum_{q=1}^{r} \|F_{i}(U_{q}) - F_{i}(J')\|_{1} + E_{i}}$$
(20)

In Formula (20), J' represents the haze-free image generated after passing through the dehazing network, F_i represents the *i*-th hidden feature map extracted from the trained VGG-19 model, U_q represents the blurred image inconsistent with the clear image, and ξ_i represents the hyperparameter.

Ultimately, our total loss function consists of a combination of mean squared error loss and contrast regularization.

$$L = L_{MSE} + \lambda R \tag{21}$$

where λ represents the penalty parameter, whose setting aims to achieve a balance during the model training process. Compared with the previous single loss function, the composite loss function can avoid the overfitting phenomenon of the network model and more effectively reduce the generated graph and clear graph and the difference in color, contrast, and details.

4. Experimental Results and Analysis

4.1. Experimental Environment Configuration and Dataset

The required environment of the network training runtime described in the present paper was configured as Python = 3.7.11 and torch = 1.10.0, with the learning rate of the training network set to 0.0001, the training steps set to 200,000 steps, and the number of steps set to 500 steps. We empirically set the penalty parameters λ to 0.2. We followed CR [33] in setting the L1 distance in Formula (20), which is based on the latent features of the 1st, 3rd, 5th, 9th, and 13th layers from the fixed pre-trained VGG-19 and their corresponding weights ξ_i , i = 1,, 5 to $\frac{1}{32}$, $\frac{1}{16}$, $\frac{1}{8}$, $\frac{1}{4}$, and 1, respectively. The dataset employed for network training is the common dataset of image dehazing REISDE (188), and the indoor training set (ITS) and outdoor training set (OTS) were leveraged to respectively train the network to produce different corresponding indoor and outdoor network weights. The pictures in the comprehensive target test set (SOTS) were selected, and the aerial images were added to test the network.

4.2. Comparison of Experimental Results Using Publicly Tested Datasets

Our proposed algorithm has been compared with traditional algorithms, including DCP [20], CAP [21], MSCNN [22], DehazeNet [23], AOD-Net [24], GCANet [25], FFA-Net [26], and C2PNet [27], on the test dataset (SOTS). In this study, we used two metrics, the peak signal-to-noise ratio (PSNR) and structural similarity (SSIM), to quantify and compare the advantages of different algorithms. The PSNR, a widely used objective image quality assessment index, evaluates image quality based on the error between corresponding pixels. However, it is important to note that the PSNR is less sensitive to contrast changes in low spatial frequency and may not align with the subjective assessments of human vision. A higher PSNR value indicates that the dehazed image is closer to the original fog-free image. The SSIM index measures the similarity of an image in terms of luminance, contrast, and structure. A higher SSIM value suggests that the image has less distortion and retains more details.

Figure 7 shows a qualitative comparison of the subjective visual effects of indoor, outdoor, and aerial fog images after fog removal using each dehazing algorithm.



Figure 7. A comparison of the dehazing effects of the different algorithms on the test datasets.

According to the results presented in Figure 7, the application of the DCP [20] algorithm results in the overall darkening of the image, with color distortions evident in the sky and river areas. The CAP [21] algorithm exhibits residual fog in the depth-of-field areas,

and its fog reduction effect is subpar. The MSCNN [22] algorithm can preserve the original color of the image to a high degree but fails to completely remove the fog. The DehazeNet algorithm achieves a high level of image tone recovery yet suffers from a significant loss of image texture detail. The AOD-Net [24] algorithm over-saturates the image, leading to the loss of original color information, the overall darkening of the image, and noticeable residual fog. The GCANet [25] algorithm produces images that are overly bright, with fog residue clearly visible at the junction between the sky and buildings. The FFA-Net [26] and C2PNet [27] algorithms perform poorly on aerial images, with a high degree of fog residue and unsatisfactory tone and detail recovery. In contrast, the improved algorithm presented in this paper yields ideal results, with a high degree of image color recovery. It naturally avoids image distortion and clearly retains image details. We quantitatively compared the improved algorithm with DCP [20], CAP [21], and deep learning-based defog algorithms such as MSCNN [22], DehazeNet [23], AOD-Net [24], GFN [29], FFA-Net [26], PMNet [28], and C2PNet [27] using the test dataset and averaged the values of the PSNR and SSIM. The experimental results are shown in Table 1.

Method	SOTS—Indoor		SOTS—Outdoor	
	PSNR	SSIM	PSNR	SSIM
DCP [20]	16.61	0.8546	19.14	0.8605
CAP [21]	16.95	0.7942	19.82	0.8255
MSCNN [22]	16.89	0.7796	19.01	0.7931
DehazeNet [23]	19.82	0.8209	22.46	0.8514
AOD-Net [24]	20.51	0.8162	24.14	0.9198
GFN [29]	22.30	0.880	21.55	0.844
GCANet [25]	23.34	0.9025	26.14	0.8582
FFA-Net [26]	36.39	0.9886	33.57	0.9840
PMNet [28]	38.41	0.990	34.74	0.985
C2PNet [27]	42.56	0.9954	36.68	0.9900
Proposed algorithm	32.24	0.9422	31.09	0.9723

Table 1. A comparison of the experimental results of the different algorithms using the test dataset.

As shown in Table 1, the algorithm presented in this paper outperforms DCP [20], CAP [21], MSCNN [22], DehazeNet [23], and GCANet [25] when applied to both indoor and outdoor test sets. However, it falls short when compared to FFA-Net [26] and C2PNet [27]. Specifically, for aerial images, this algorithm demonstrates superior performance compared to other comparative algorithms. When benchmarked against the lightweight network AOD-Net [24] on the SOTS—indoor dataset, the PSNR value shows an improvement of 11.73 dB, and the SSIM value increases by 0.126. On the SOTS—outdoor dataset, the PSNR value improves by 6.95 dB, and the SSIM value improves by 0.0525. For the aerial image test dataset, the PSNR value is enhanced by 4.12 dB, and the SSIM value is enhanced by 0.0476.

4.3. A Comparison of Experimental Results for the Aerial Tidal Flats Dataset

The images utilized in the aerial tidal flats dataset were collected by an assembled fixed-wing drone equipped with an RGB camera featuring a 1/2.3-inch CCD image sensor. The camera has a maximum effective pixel count of 16 million and an image resolution of 3.5 cm per pixel. During data collection, the drone maintained a flight altitude of 140 m, ensuring a 50% forward overlap and a 16% side overlap to ensure the comprehensiveness and accuracy of the data.

We utilized an aerial tidal flats image dataset found online for a comprehensive quantitative and qualitative comparison during our research. The comparison was between the proposed network model and a range of traditional dehazing algorithms, including DCP [20] and CAP [21], as well as deep learning-based dehazing algorithms such as MSCNN [22], DehazeNet [23], AOD-Net [24], GCANet [25], FFA-Net [26], FSNet [30], and C2PNet [27].

Qualitative analysis: The images in Figure 8 show that our network acquires better visual results on the tidal beach dataset compared to previous state-of-the-art models. The DCP [20] algorithm shows severe image shifts due to the underlying prior assumptions; the CAP [21], MSCNN [22], DehazeNet [23], and AOD-Net [24] algorithms do not completely remove the fog, and there is still a considerable amount of fog on the edges of the images. For GCANet [25], the processing of high-frequency details, such as the texture, edge, and seawater in the image, is too poor, making the image seriously distorted. Although the large models such as FFA-Net [26], C2PNet [27], and FSNet [30] performed well when applied to the publicly tested dataset, a considerable amount of fog and fuzzy edge details appeared when they were applied to the tidal beach dataset. Compared with previous algorithms, this algorithm has better processing ability in terms of edge details, the image color information remains intact, there is no serious distortion, and its fog processing ability is also better than other algorithms.



Figure 8. Cont.



Figure 8. Cont.



Figure 8. A qualitative comparison of the tidal flat datasets. A zoomed-in image (red boxes) is shown to provide the best view.

Quantitative analysis: We compared the performance of this network with the most advanced algorithms at present. In order to more accurately prove the performance of each network in all aspects, we adopted four evaluation indicators, PSNR, RGB-SSIM, Gray-SSIM, and MSE, for quantitative analysis. The results are shown in Table 2; our model exceeds the previous algorithms in terms of PSNR, Gray-SSIM, and MSE; although lagging behind the traditional algorithm CAP [3] in terms of RGB-SSIM, it still maintains a leading position. We were therefore able to demonstrate the utility of our algorithm for the tidal beach dataset.

4.4. A Comparison of the Algorithms' Model Parameters

While enhancing the visual quality and effects of images after defogging, we also emphasize the real-time capability of the defogging process to address the needs of practical applications. Faced with the challenge of real-time image defogging, we must carefully consider the number of model parameters. Although adding convolutional layers and modules can improve the quality of defogged images, it may also decrease the model's operational efficiency. Therefore, improving image quality should not come at the cost of operational efficiency. Table 3 provides a comparison of the parameters of each model and the number of floating-point operations.

Method —	Tidal Flats				
	PSNR	RGB-SSIM	Gray-SSIM	MSE	
DCP [20]	19.73	0.9857	0.9356	95.27	
CAP [21]	22.32	0.9864	0.9462	85.42	
MSCNN [22]	19.99	0.9858	0.9167	93.83	
DehazeNet [23]	19.02	0.9693	0.8662	93.19	
AOD-Net [24]	22.02	0.9814	0.9064	86.40	
GFN [29]	21.33	0.9764	0.8973	94.18	
GCANet [25]	17.86	0.9784	0.8343	99.43	
FFA-Net [26]	20.39	0.9776	0.8807	88.47	
PMNet [28]	19.18	0.9624	0.8785	95.84	
C2PNet [27]	20.94	0.9521	0.9096	107.26	
FSNet [30]	19.19	0.9150	0.8530	118.32	
Proposed algorithm	25.07	0.9893	0.9618	75.80	

Table 2. A quantitative comparison of the different algorithms on the tidal beach dataset.

Table 3. A comparison of the network model parameters for the different algorithms.

Method	Param. (M)	FLOPs (G)	Latency (Ms)
DCP [20]			
CAP [21]			
MSCNN [22]	0.008	0.525	0.619
DehazeNet [23]	0.009	0.581	0.919
AOD-Net [24]	0.002	0.115	0.390
GFN [29]	0.499	14.94	3.849
GCANet [25]	0.702	18.41	3.695
FFA-Net [26]	4.456	287.8	55.91
PMNet [28]	18.9	81.13	27.16
C2PNet [27]	7.17	461.7	73.13
FSNet [30]	4.72	39.67	18.75
Proposed algorithm	0.005	0.351	0.523

As can be seen in Table 3, although slightly higher than the parameters and floatingpoint operations of AOD-Net [24], the complexity of our model is still several times or even hundreds of times lower than other algorithms. With some improvement, we can achieve the goal of the clearer real-time monitoring of tidal flats.

4.5. Comparison of Ablation Experiments

To demonstrate the effectiveness of our design improvement on network attention mechanisms, for ablation analysis, we employed SOTS—outdoor as a dataset for testing. A multi-scale dense residual network was utilized as the infrastructure. The attention mechanism with only an average pooling layer, the attention mechanism combining average pooling and maxi pooling, and the dilated convolution were added for the comparison experiment. The corresponding peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) were used to quantify the effects of the model, as well as specific aerial images to demonstrate the effectiveness of our improved network.

As shown in Table 4, by using the attention mechanism and dilated convolution, which combine average pooling with max pooling, the PSNR increases from 26.9741 to 31.092, and SSIM increases from 0.9247 to 0.9723. As can be seen in Figure 9, adding the improved CSAM module can aid in producing a more natural, more detailed, and clearer visual ideal effect with a more obvious edge dehazing effect.

Base (Densenet) CSAM (Avgpool) Avgpool and Maxpool Dilated rate	\checkmark		$\sqrt[]{}$ $\sqrt[]{}$	\checkmark \checkmark \checkmark
PSNR	26.97	29.36	29.84	31.09
SSIM	0.9247	0.94	0.9571	0.9723

Table 4. Quantitative results of the ablation experiments.



Figure 9. The dehazing effect of the ablation experiment. A zoomed-in image (red boxes) is shown to provide the best view.

4.6. Effectiveness of λ in Loss Function

Below, we present the results of a trade-off experiment for the mean squared error (MSE) loss and the contrast regularization loss, as shown in Figure 10. In the composite loss function, we believe that the MSE loss predominates in the network training process, while contrast regularization serves as an auxiliary role in model training. Therefore, we set the range of the hyperparameter λ from 0 to 1. By adjusting the value of the hyperparameter λ , we identified the optimal λ value to optimize network performance. We set the hyperparameter λ to 0, 0.1, 0.2, 0.3, 0.5, 0.8, and 1. The chart shows that network performance is at its peak when the hyperparameter λ is set to 0.2.



Figure 10. A study on the influence of λ .

5. Discussion

In tidal flat areas, UAV monitoring tasks are often disrupted by fog, leading to a reduction in the clarity of captured images. This not only affects the quality of monitoring data but also limits our ability to accurately analyze the tidal flat environment. To overcome this challenge, this study developed a novel and efficient defogging algorithm. The algorithm significantly improves image quality through multi-scale feature recognition, optimized information flow, and contrastive learning strategies, reducing image distortion and residual fog while maintaining a relatively small model size. It is highly practical in rapid decision support systems, achieving an optimal balance between efficiency and performance, effectively addressing the real-time fog monitoring issues commonly encountered in tidal flats. However, our method also has certain limitations. For instance, when applied to foggy images with strong light sources, the algorithm's performance is suboptimal, often leading to processed images becoming darker and affecting the algorithm's performance. Therefore, future research efforts will need to further optimize the algorithm to enhance its robustness under various weather conditions. Additionally, contrastive learning plays a critical role in the algorithm's performance, enhancing feature learning and improving model generalization, thereby improving the dehazing effect. This discovery not only provides a new perspective for tidal flat image processing but may also inspire advancements in image enhancement and restoration tasks in other fields.

6. Conclusions

In light of the challenges associated with aerial photography and the real-time monitoring of tidal beaches, in the present study, a qualitative and quantitative comparison was conducted comparing traditional algorithms and more advanced deep learning algorithms. The experimental results indicate that, with a relatively small model parameter size, our approach significantly enhances the quality of aerial images following fog removal. This enhancement in turn leads to an optimal balance between efficiency and performance, effectively addressing the real-time fog monitoring issues prevalent in tidal flats. However, its performance when applied to foggy images with strong light sources is not ideal, which limits its applicability in certain scenarios. Despite this limitation, when considering objective performance metrics, visual outcomes, and model complexity, this algorithm stands out as a notably effective real-time image dehazing network model.

Author Contributions: Methodology, Z.Z.; Software, D.Y.; Investigation, H.G.; Data curation, C.X.; Writing—original draft, D.Y.; Writing—review & editing, Z.Z.; Visualization, H.Q.; Supervision, H.W.; Project administration, Z.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the Key Research and Development Plan of Jiangsu Province (BE2022783) and the Graduate Research Practice Plan of Jiangsu Province (SJCX23_2133).

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

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

References

- Murray, N.J.; Phinn, S.R.; DeWitt, M.; Ferrari, R.; Johnston, R.; Lyons, M.B.; Clinton, N.; Thau, D.; Fuller, R.A. The global distribution and trajectory of tidal flats. *Nature* 2019, 565, 222–225. [CrossRef] [PubMed]
- Dai, W.; Li, H.; Gong, Z.; Zhang, C.; Zhou, Z. applications of UAV technology in the evolution of tidal beach landform. *Prog. Water Sci.* 2019, 30, 359–372.
- Kim, K.L.; Woo, H.J.; Jou, H.T.; Jung, H.C.; Lee, S.K.; Ryu, J.H. Surface sediment classification using a deep learning model and unmanned aerial vehicle data of tidal flats. *Mar. Pollut. Bull.* 2024, 198, 115823. [CrossRef] [PubMed]
- 4. Li, Y.-T. Study on the control of Spartina alterniflora by UAV application-taking the tidal flat along Northern Chongming as an example. *J. Anhui Agric. Sci.* 2023, *51*, 57–60.
- 5. Fan, Y.; Chen, S.; Zhao, B.; Yu, S.; Ji, H.; Jiang, C. Monitoring tidal flat dynamics affected by human activities along an eroded coast in the Yellow River Delta, China. *Environ. Monit. Assess.* **2018**, *190*, 396. [CrossRef] [PubMed]

- Chandran, I.; Kizheppatt, V. Multi-UAV Networks for Disaster Monitoring: Challenges and Opportunities from a Network Perspective. Drone Syst. Appl. JA 2024, 12, 1–28.
- Bashir, M.H.; Ahmad, M.; Rizvi, D.R.; El-Latif, A.A.A. Efficient CNN-based disaster events classification using UAV-aided images for emergency response application. *Neural Comput. Appl.* 2024, *36*, 10599–10612. [CrossRef]
- Bhatia, D.; Dhillon, A.S.; Hesse, H. Preliminary Design of an UAV Based System for Wildlife Monitoring and Conservation. In Proceedings of the International Conference on Aeronautical Sciences, Engineering and Technology, Muscat, Oman, 3–5 October 2023; Springer Nature: Singapore, 2023; pp. 51–63.
- Haq, B.; Jamshed, M.; Ali, K.; Kasi, B.; Arshad, S.; Kasi, M.; Ali, I.; Shabbir, A.; Abbasi, Q.; Ur Rehman, M. Tech-Driven Forest Conservation: Combating Deforestation With Internet of Things, Artificial Intelligence, and Remote Sensing. *IEEE Internet Things* J. 2024, 11, 24551–24568. [CrossRef]
- Ahmed, Z.E.; Hashim AH, A.; Saeed, R.A.; Saeed, M.M. Monitoring of Wildlife Using Unmanned Aerial Vehicle (UAV) with Machine Learning. In *Applications of Machine Learning in UAV Networks*; IGI Global: Hershey, PA, USA, 2024; pp. 97–120.
- 11. Iheaturu, C.; Okolie, C.; Ayodele, E.; Egogo-Stanley, A.; Musa, S.; Speranza, C.I. Combining Google Earth historical imagery and UAV photogrammetry for urban development analysis. *MethodsX* **2024**, *12*, 102785. [CrossRef]
- 12. Wang, C.; Pavelsky, T.M.; Kyzivat, E.D.; Garcia-Tigreros, F.; Podest, E.; Yao, F.; Yang, X.; Zhang, S.; Song, C.; Langhorst, T.; et al. Quantification of wetland vegetation communities features with airborne AVIRIS-NG, UAVSAR, and UAV LiDAR data in Peace-Athabasca Delta. *Remote Sens. Environ.* **2023**, *294*, 113646. [CrossRef]
- 13. Zhuang, W.; Xing, F.; Lu, Y. Task Offloading Strategy for Unmanned Aerial Vehicle Power Inspection Based on Deep Reinforcement Learning. *Sensors* 2024, 24, 2070. [CrossRef] [PubMed]
- 14. Shamta, I.; Batıkan, E.D. Development of a deep learning-based surveillance system for forest fire detection and monitoring using UAV. *PLoS ONE* **2024**, *19*, e0299058. [CrossRef] [PubMed]
- 15. Yuan, S.; Li, Y.; Bao, F.; Xu, H.; Yang, Y.; Yan, Q.; Zhong, S.; Yin, H.; Xu, J.; Huang, Z.; et al. Marine environmental monitoring with unmanned vehicle platforms: Present applications and future prospects. *Sci. Total Environ.* **2023**, *858*, 159741. [CrossRef]
- An, S.; Huang, X.; Cao, L.; Wang, L. A comprehensive survey on image dehazing for different atmospheric scattering models. *Multimed. Tools Appl.* 2024, 83, 40963–40993. [CrossRef]
- 17. Gui, J.; Cong, X.; Cao, Y.; Ren, W.; Zhang, J.; Zhang, J.; Cao, J.; Tao, D. A comprehensive survey and taxonomy on single image dehazing based on deep learning. *ACM Comput. Surv.* 2023, 55, 279. [CrossRef]
- 18. Zhang, K.; Wang, A.; Xiong, Y.; Liu, Y. Survey of Transformer-Based Single Image Dehazing Methods. J. Front. Comput. Sci. Technol. 2024, 18, 1182.
- 19. Zheng, F.; Wang, X.; He, D.; Fu, Y.Y.; Yuan, S.X. Review of single image dehazing algorithm research. *Comput. Eng. Appl.* **2022**, *58*, 1–14.
- He, K.M.; Sun, J.; Tang, X.O. Single image haze removal usingdark channel prior. *IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI)* 2011, 33, 2341–2353.
- Zhu, Q.S.; Mai, J.M.; Shao, L. A fast single image haze removal al-gorithm using color attenuation prior. *IEEE Trans. Image Process.* 2015, 24, 3522–3533.
- Ren, W.Q.; Liu, S.; Zhang, H.; Pan, J.S. Single image dehazing viamulti-scale convolutional neural networks. In Proceedings of the2016 European Conference on Computer Vision (ECCV), Amsterdam, The Netherlands, 11–14 October 2016; Springer: Berlin/Heidelberg, Germany, 2016; pp. 154–169.
- Cai, B.L.; Xu, X.M.; Jia, K.; Qing, C.M.; Tao, D.C. DehazeNet: An end-to-end system for single image haze removal. *IEEE Trans. Image Process.* 2016, 25, 5187–5198. [CrossRef]
- Li, B.Y.; Peng, X.L.; Wang, Z.Y.; Xu, J.Z.; Feng, D. AOD-Net: All-in-one dehazing network. In Proceedings of the 2017 IEEE In-ternational Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 4780–4788.
- Chen, D.D.; He, M.M.; Fan, Q.N.; Liao, J.; Zhang, L.H.; Hou, D.D.; Yuan, L.; Hua, G. Gated context aggregation network for image dehazing and deraining. In Proceedings of the 2019 IEEE Winter Conferenceon Applications of Computer Vision (WACV), Waikoloa, HI, USA, 7–11 January 2019; pp. 1375–1383.
- Qin, X.; Wang, Z.L.; Bai, Y.C.; Xie, X.D.; Jia, H.Z. FFA-Net: Feature fusion attention network for single image dehazing. In Proceedings of the Association for the Advance of Artificial Intelligence, Hilton Midtown, NY, USA, 7–12 February 2020; pp. 11908–11915.
- Zheng, Y.; Zhan, J.; He, S.; Dong, J.; Du, Y. Curricular contrastive regularization for physics-aware single image dehazing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 5785–5794.
- 28. Ye, T.; Jiang, M.; Zhang, Y.; Chen, L.; Chen, E.; Chen, P.; Lu, Z. Perceiving and Modeling Density is All You Need for Image Dehazing. *arXiv* 2021, arXiv:2111.09733. [CrossRef]
- Ren, W.; Ma, L.; Zhang, J.; Pan, J.; Cao, X.; Liu, W.; Yang, M.H. Gated Fusion Network for Single Image Dehazing. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018. [CrossRef]
- Cui, Y.; Ren, W.; Cao, X.; Knoll, A. Image restoration via frequency selection. *IEEE Trans. Pattern Anal. Mach. Intell.* 2023, 46, 1093–1108. [CrossRef] [PubMed]

- Chen, T.; Kornblith, S.; Norouzi, M.; Hinton, G. A simple framework for contrastive learning of visual representations. In Proceedings of the 37th International Conference on Machine Learning, Virtual, 13–18 July 2020; pp. 1597–1607.
- He, K.; Fan, H.; Wu, Y.; Xie, S.; Girshick, R. Momentum contrast for unsupervised visual represen-tation learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 9729–9738.
- Wu, H.; Qu, Y.; Lin, S.; Zhou, J.; Qiao, R.; Zhang, Z.; Xie, Y.; Ma, L. Contrastive learning for compact single image dehazing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 19–25 June 2021; pp. 10551–10560.
- 34. McCartney, E.J. Optics of the Atmosphere: Scattering by Molecules and Particles; John Wiley and Sons, Inc.: New York, NY, USA, 1976; 421p.
- 35. Narasimhan, S.G.; Nayar, S.K. Vision and the atmosphere. Int. J. Comput. Vis. 2002, 48, 233–254. [CrossRef]

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.