IEICE TRANSACTIONS

on Information and Systems

DOI:10.1587/transinf.2023EDP7191

Publicized:2024/05/14

This advance publication article will be replaced by the finalized version after proofreading.

A PUBLICATION OF THE INFORMATION AND SYSTEMS SOCIETY

The Institute of Electronics, Information and Communication Engineers

Kikai-Shinko-Kaikan Bldg., 5-8, Shibakoen 3 chome, Minato-ku, TOKYO, 105-0011 JAPAN

PAPER

Remote Sensing Image Dehazing Using Multi-Scale Gated Attention For Flight Simulator

Qi LIU[†], **Bo WANG**[†], **Shihan TAN**[†], **Shurong ZOU**[†a)], ***Nonmembers, and*** **Wenyi GE**[††,†,†††], ***Member***

SUMMARY For flight simulators, it is crucial to create three-dimensional terrain using clear remote sensing images. However, due to haze and other contributing variables, the obtained remote sensing images typically have low contrast and blurry features. In order to build a flight simulator visual system, we propose a deep learning-based dehaze model for remote sensing images dehazing. An encoder-decoder architecture is proposed that consists of a multiscale fusion module and a gated large kernel convolutional attention module. This architecture can fuse multi-resolution global and local semantic features and can adaptively extract image features under complex terrain. The experimental results demonstrate that, with good generality and application, the model outperforms existing comparison techniques and achieves high-confidence dehazing in remote sensing images with a variety of haze concentrations, multi-complex terrains, and multi-spatial resolutions.

key words: *remote sensing images dehazing, multi-scale fusion, gated attention, flight simulator*

1. Introduction

The advancement of computer simulation technology supports in the advancement of national defense, business, and other areas, particularly in the sector of aerospace, where flight simulators are quickly evolving due to their reliance on simulation technology. A flight simulator is a type of simulation flight training tool that can simulate an aircraft's flying condition in flight and provide pilots the same operational experience, visual feedback, and audio feedback as a real aircraft. Flight simulators are frequently used in place of actual aircraft throughout the pilot training process as they provide significant benefits over real aircraft in terms of safetya, economics, and other factors [1]. In addition to the pilot's flight training, military and civil aircraft flying tests, as well as the pilot's free recovery training, the flight simulator is indispensable. Flight simulator consists of visual system, motion system, navigation system and other systems, of which the visual system is an important part of the flight simulator, the visual system can provide pilots with the use of real aircraft training with the same real-time dynamic environment outside the cockpit and inside the cabin, which is an important guarantee of the authenticity of the flight simulator.

The dynamic scene simulation of the flight simulator visual system [2] is mainly through the real-time scheduling of the visual database, which affects the pilot's judgment of the external environment. The terrain database is usually generated with remote sensing images [3], and the quality of the remote sensing images directly affects the quality of the terrain database, which in turn affects the quality of the visual system, and ultimately affects the quality of the entire flight simulation training. Although more remote sensing image data has become available in recent years due to advances in sensor technology and the expansion of remote sensing platforms, this is despite the fact that remote sensing images are primarily obtained through the observation of electromagnetic wave information from the sun, making them extremely vulnerable to weather and other factors. For example, more cloudy skies or haze will cause the edge to be blurred, and the color of the distortion and other problems. Since it is challenging to directly apply these haze-affected remote sensing images to the creation of the terrain database, it is necessary to dehaze the haze images in order to enhance the terrain database's overall data quality and the functionality of the detailed features, to enhance the quality of the visual system, and ultimately to guarantee the quality of the pilot's training.

Most of the image dehazing methods are based on the atmospheric scattering model starting from the estimation of the atmospheric transmission map to realize image dehazing:

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

where I is a haze image, $J(x)$ is a haze-free image, A is the global atmospheric light, and $t(x)$ is a medium transmission map. The transmission map $t(x)$ can be further expressed as $t(x) = e^{-\beta d(x)}$, where β is the atmospheric scattering coefficient and $d(x)$ is the scene depth. Although this approach is effective for image dehazing, it is less applicable to remote sensing images due to the fact that the imaging range of remote sensing images is wide and the distribution of haze is also inhomogeneous leading to the global atmospheric light A is inhomogeneous, it ought to be a variable, makes it distinct from natural images [4]. Additionally, the remote sensing images has a variety of spatial resolutions and topography, which making it essential to estimate various atmospheric transmission maps based on various remote sensing images when dehazing multiple remote sensing images. This has a significant impact on efficiency.

Many image dehazing networks based on deep learning

[†]The author is with the college of Computer Science, Chengdu University of Information Technology, Chengdu, 610225, China

[††]The author is with the Chengdu Institute of Computer Applications Chinese Academy of Sciences. Chengdu, 610041, China

[†††]The author is with the Sichuan Jiuzhou Investment Holding Group Co., Ltd., Mianyang, 621000, China

a) E-mail: zousr@cuit.edu.cn

have emerged in recent years due to the rapid development of deep learning and neural network technology in the field of computers. These dehazing networks work to remove haze from images by estimating the residuals between hazy and clear images. However, the application of these methods in the remote sensing image dehazing is less effective. The reason is the imaging range of remote sensing images is broad, a remote sensing image contains a lot of information about the landscape features, presenting a complex and varied topography to the senses; at the same time, the spatial resolution of remote sensing images is varied, and remote sensing images with different spatial resolutions of the same location contain different amounts of information, sensory presentation of similar in parts but different in whole. These characteristics make it very easy for underfitting or overfitting to occur when applying deep learning modeling methods for dehazing. The use of these underfitted or overfitted modeled dehazing images to construct the terrain database will result in terrain blurring, loss of saturation and contrast leading to poor realism of the visual system and thus affecting the quality of pilot training. The paper proposes an end-to-end remote sensing image dehazing model that may be used for a variety of spatial resolutions and complicated terrains based on the aforementioned issues. We propose a Multi-Scale Fusion(MSF) module to extract image features through multiple dilation convolutions with different dilation sizes, which can be applied to different spatial resolutions to obtain heterogeneous scale correlations, in order to address the issue of multi-spatial resolution of remote sensing images. We propose the Gated Large Kernel Attention (GLKA) module, which introduces adaptive attention to improve the feature extraction capability of the model under multiple landscape features information and multiple complex terrain, with a focus on the characteristics of remote sensing images with many landscape features information and complex terrain. Additionally, we created a dataset of remote sensing images with various spatial resolutions and varied terrain attributes. We employed both qualitative and quantitative evaluation to assess the effectiveness of various image dehazing networks. Peak signal-to-noise ratio (PSNR), structural similarity (SSIM), and learned perceptual image patch similarity (LPIPS) are used in the quantitative evaluation to measure the dehazing effectiveness of the computational model. Experiments show that our suggested model produces favorable outcomes. The following is a summary of the contributions made by this paper:

• We design a remote sensing image dehazing model based on encoder-decoder structure suitable for the construction of flight simulator visual system, which is capable of dehazing remote sensing images with multi-complex terrains and multi-spatial resolutions, and propose the MSF module and the GLKA module for feature extraction as well as feature fusion. Also proposed is a remote sensing images dataset we refer to as DMRSI.

• The MSF module enables the combination of shallow semantic information and deep local information, which can efficiently reduce information loss during the convolution process and improve the stability of the model. Additionally, the combination of multi-path convolution can ensure that the model avoids overfitting and affects the performance of the dehazing when dehazing remote sensing images with various spatial resolutions.

• The GLKA module consists of a gating mechanism and a large kernel of attention. By using a large kernel convolution and depth expansion convolution, these two techniques ensure the adaptability of the attention and the establishment of long-range dependence. Pure convolution, on the other hand, avoids a significant amount of computational and memory overhead, improving performance and efficiency. The gating method makes sure that the model doesn't lose local information while creating long-range dependencies, which ensures that the model may be applied to remote sensing images of numerous complex terrains.

• DMRSI comprises remote sensing images with a range of spatial resolutions, including from 512 meters to 1 meter, in a halved stepwise distribution. It also includes a variety of landscape features, such as cities, coasts, deserts, farmlands, forests, and mountains. A realistic simulation of haze in nature is provided by DMRSI's two types of haze states, mist and hazy.

2. Related Work

The three primary categories of remote sensing images dehazing methods now in use are as follows: the first is based on image enhancement, which does not take into account the physical model of image deterioration but instead enhances image quality by boosting contrast. The most representative of these is histogram equalization [5], Retinex algorithm [6] and homomorphic filtering [7] method. However, these methods are typically used for single image dehazing and have poor generalizability. The histogram equalization method, for example, is only applicable to images with heavy haze, the Retinex algorithm has high complexity, and homomorphic filtering is not applicable to images that are too bright or dark.

The second category is based on physical a priori approaches, which are often based on the atmospheric scattering model. According to extensive observation and statistical analysis of outdoor clear images, He et al. [8] discovered that most non-sky patches contain some pixels in at least one of the color channels that have very low intensities. Image dehazing is accomplished by utilizing this low pixel intensity for use in the estimation of atmospheric transmission inputs. Li et al. [9] proposed a simple and effective single-image dehazing method based on an improved bright-channel prior and dark-channel prior, which divides sky and non-sky regions by particle swarm optimization and estimates the atmospheric transport map by a bright and dark-channel prior. He et al. [4] proposed an image haze removal algorithm for visible light based on a non-uniform atmospheric light prior and a side-window filter, which presents a side-window filter-based transmission estimation algorithm to suppress the block effect in the transmission map due to the large

window of the smallest filter used in the dark channel algorithm, and also combines it with a simple estimation of the non-uniform atmospheric light to achieve image dehazing. Li et al. [10] obtained accurate haze transmittance by introducing Gaussian-weighted image fusion, and also used an unsharpened mask method to correct the dehazed image to solve the problem of image color distortion, which achieved good results in both outdoor and remote sensing images.

While the above methods (augmented, a priori) have shown high performance on specific datasets, such methods are usually only applicable to specific scenarios or specific datasets and require physical knowledge to back them up, and thus are not applicable to flight simulators that require a large number of remote sensing images.

The third category is machine learning based or deep learning based methods. Cai et al. [11] proposed an image dehazing method based on CNN architecture, which takes a hazy image as input and outputs its transmission mapping for image dehazing. Ren et al. [12] proposed a multi-scale CNN for image dehazing, the model is mainly divided into two parts: coarse scale network and fine scale network, the coarse scale will estimate the transmission map of the input hazy image at the coarse scale and send the result to the fine scale network, the fine scale network will refine the transmission map to realize the image dehaze. Li et al. [13] proposed an end-to-end trainable dehazing model that can recover clear images directly from hazy images without relying on any intermediate parameter estimation. Zhang et al. [14] proposed a Densely Connected Pyramid Dehazing Model (DCPCN), which directly embeds the atmospheric scattering model into the network, and directly learns the projected map and atmospheric light to realize the image dehazing through the encoding and decoding network with edges keeping the pyramids densely connected. Chen et al. [15] proposed GCANet, which employs smooth dilation convolution instead of the original dilation convolution, solves the mesh artifacts induced by the dilation convolution, and utilizes a gated sub-network to fuse high and low dimensional features to improve the dehazing effect. Guo et al. [16] proposed RSDehazeNet, introducing both local and global residual learning and using a channel attention module to achieve fast convergence of the model. Wu et al. [17] proposed AECRNet, a compact image dehazing method based on contrast learning by mining negative sample information. Ge et al. [18] proposed a U-Net based image dehazing method, which has achieved good results in both natural and remote sensing image fields. Chen et al. [19] proposed an end-to-end hybrid high-resolution learning network framework called H2RL-Net utilizing a parallel cross-scale fusion module to aggregate information from multiple scales and perform dynamic feature recalibration of channel features to produce better dehazing results. He [20]et al. proposed to fuse the features of visible and infrared bands to utilize the strong penetration ability of infrared band for the dehazing of remote sensing images. He et al. [21] proposed an end-to-end convolutional neural network based on an attention mechanism that contains a residual block structure, which combines channel and spatial attention mechanisms. Li et al. [22] proposed M2SCN, an end-to-end image dehazing network consisting of a multi-model joint estimation module with enhanced generalization capability and a self-correction module with enhanced blurring capability. Wei et al. [23] proposed a self-supervised remote sensing (RS) image dehazing network based on zero-sample learning by combining a priori knowledge with deep learning, where the self-supervision process is able to reduce the data requirements while the learning-based structure is able to refine the artifacts caused by the complex real-world environment.

Constructing terrain database requires a large number of remote sensing images, so the dehazing methods based on image enhancement [5–7]and based on physical prior knowledge [4, 8–10]are difficult to satisfy the demand, and the dehazing methods based on deep learning can quickly realize the dehazing of large-volume images, so the research focus of this paper is on the dehazing methods based on deep learning. Although many remote sensing image dehazing methods based on deep learning have been proposed, there are some problems that are difficult to be applied to the construction of the flight simulator visual system terrain database. The reason is that the existing deep learning-based remote sensing image dehazing methods can not be directly applied to the construction of the flight simulator visual system terrain database, Through Table 1 we can see that there are two main directions of existing methods, one is generalized dehazing methods [11–14, 17, 18, 22], and the other is model optimization based on one main problem, such as the use of multi-scale, etc. to fuse low- and high-dimensional [15, 16, 19] information or the introduction of attention and residual networks [20, 21], or self-supervision [23] to reduce the problem of missing training samples., while the construction of the flight simulator visual system terrain database requires multiple spatial resolution, multiple terrain remote sensing images, and the existing methods can not meet the multiple spatial resolution, multiple terrain remote sensing image. However, the construction of the flight simulator visual system terrain database requires remote sensing images with multiple spatial resolutions and multiple terrains, and the existing methods are unable to meet the demand for dehazing remote sensing images with multiple spatial resolutions and multiple complex terrains. The characteristics of multi-spatial resolution and multi-complex terrain of remote sensing images in flight simulators challenge the existing dehazing methods, and enhancement-based methods such as [4, 8–10] are unable to process quickly and in large quantities, and [11–23]are unable to satisfy the requirements of multi-spatial resolution and multi-complex terrain, and our method is based on deep learning, which realizes fast dehazing for multi-spatial resolution and multi-complex terrain.

3. Methodology

In this section, we will present three main parts, one is the specific details of the proposed model. The basic architecture

Table 1 Studies relating to images dehazing

Paper	Year	Highlight
DCPCN	2018	Direct learning of projection maps and atmospheric light for image de-fogging by edge-keeping pyramid densely connected codec networks.
GCA	2019	Introducing smooth dilation convolution to reduce artifacts and gating subnetworks to fuse high and low dimensional features.
RSDehazeNet	2020	Introducing both local and global residual learning and using a channel attention module to achieve fast convergence of the model.
AECR	2021	Contrast learning based dehazing method by mining negative sample information.
U-net based	2021	Unet-based dehazing method for airports and other scenarios.
H2RL-Net	2021	Utilizing a parallel cross-scale fusion module to aggregate information from multiple scales and perform dynamic feature recalibration of channel features.
VFF	2022	Fuse the features of visible and infrared bands to utilize the strong penetration ability of infrared band for the images dehazing.
ARCNN	2022	Introducing Attention Mechanisms and Residual Blocks.
M2SCN	2022	Multi-model joint estimation module with enhanced generalization capability and self-correction module with enhanced fuzzy processing capability.
IBDCP	2023	Estimation of atmospheric transport maps via bright-channel priori and dark-channel priori.
NUALP	2023	Side-window filters suppress block effects while simply estimating non-uniform atmospheric light.
GWIF	2023	Introducing Gaussian-weighted image fusion to obtain accurate haze transmittance.
SSRS	2023	Self-supervised networks based on zero-sample learning by combining prior knowledge with deep learning.

of the proposed model is shown in Fig.1. The second is the design intent and specific module design details of the Multi-Scale Fusion module, and the third is the design intent and module design details of the Gated Large Kernel Attention module, including the use of Large Kernel Attention and the gating mechanism.

3.1 Network Architecture

The fundamental design of our proposed model is based on U-Net [24], which is a very classical and very successful Encoder-Decoder architecture. As shown in the Fig.1, the model proposed in this paper consists of two main modules, namely the Multi-Scale Fusion (MSF) module and the Gated Large Kernel Attention (GLKA) module. First, there are four stages in the encoder stage, each of which includes a 3×3 convolution, an MSF module, and a GLKA module. The input image I will first go through a 3×3 convolution to extract the original image features, and then the MSF module and the GLKA module, which are termed F_{mf} and F_{ga}, will further extract the high dimensional feature image. The stage can be formulated as:

$$f_i(x) = F_{ga}(F_{mf}(Conv(f_{i-1}))) \tag{1}$$

where $f_i(x)$ represents the feature map at the end of the stage and $f_0(x)$ represents the input image I.

In the Decoder stage, there are 3 stages, each stage consists of upsampling, MSF module with GLKA module and will establish residual connections with different stages of the encoder stage. The upsampling uses PixelShuffle to restore the feature information to the size of the original image. The last stage additionally includes a 3×3 convolution to recover the initial size. The stage can be formulated as:

$$f_i(x) = F_{ga}(F_{mf}(PS(f_{i-1}(x)) \oplus f_{N-i+1}(x))) \tag{2}$$

where N represents the number of all stages including Encoder and Decoder.

3.2 Multi-Scale Fusion(MSF)

Different spatial resolutions at the same location typically imply a high degree of similarity of the feature maps for remote sensing images with multiple spatial resolutions. While it is possible to change the extraction of feature images by adjusting the size of the convolution kernel according to the size of the various spatial resolutions, this greatly reduces the generalizability of the model and runs the risk of overfitting, which is against our original intention. We think that by incorporating dilated convolution, the computational approach of using various rate convolution kernels to extract and fuse the local semantic information with the global semantic information can more effectively reduce the information loss while also improving the model robustness, and can effectively realize the generalization to remote sensing images with various spatial resolutions. Therefore we propose to use the MSF module to extract and fuse the dilation convolutions with different dilation rates to process the feature maps. The MSF module is shown in Fig.2. The input feature maps will first be normalized and undergo a pointwise convolution, after which feature maps with different sizes will be extracted by 3×3 convolution kernels with rates of 1, 3, and 5 respectively, and will be summed up to perform pointwise convolution with ReLU activation operation. The stage can be formulated as:

$$f_{lp}(x) = PWConv(LN(f(x))) \tag{3}$$

$$\begin{aligned}(f_{d1}(x), f_{d3}(x), f_{d5}(x)) = (&Conv_{3,rate=1}(f_{lp}),\\ &Conv_{3,rate=3}(f_{lp})), Conv_{3,rate=5}(f_{lp}))\end{aligned} \tag{4}$$

$$\begin{aligned}f_{msf} = ReLU(&PWConv(f_{d1}(x)+\\ &f_{d3}(x) + f_{d5}(x)))\end{aligned} \tag{5}$$

where $f(x)$ represents the input feature map, $f_{lp}(x)$ represents the feature map by normalization with pointwise convolution, f_{d1}, f_{d3} and f_{d5} represents the feature maps after convolution by dilation with convolution kernel size 3 and rate 1, 3, 5, respectively. f_{msf} represents the final output feature map of the module.

3.3 GLKA(Gated Large Kernel Attention)

The investigation of human vision led to the creation of Attention Mechanism (AM). According to cognitive science, humans selectively focus on a subset of all information while ignoring other observable information because of information processing bottlenecks. Numerous earlier visual tasks have demonstrated the potent effectiveness of the attention mechanism. These visual tasks typically include the use of attention, including self-attention (S-A), channel attention

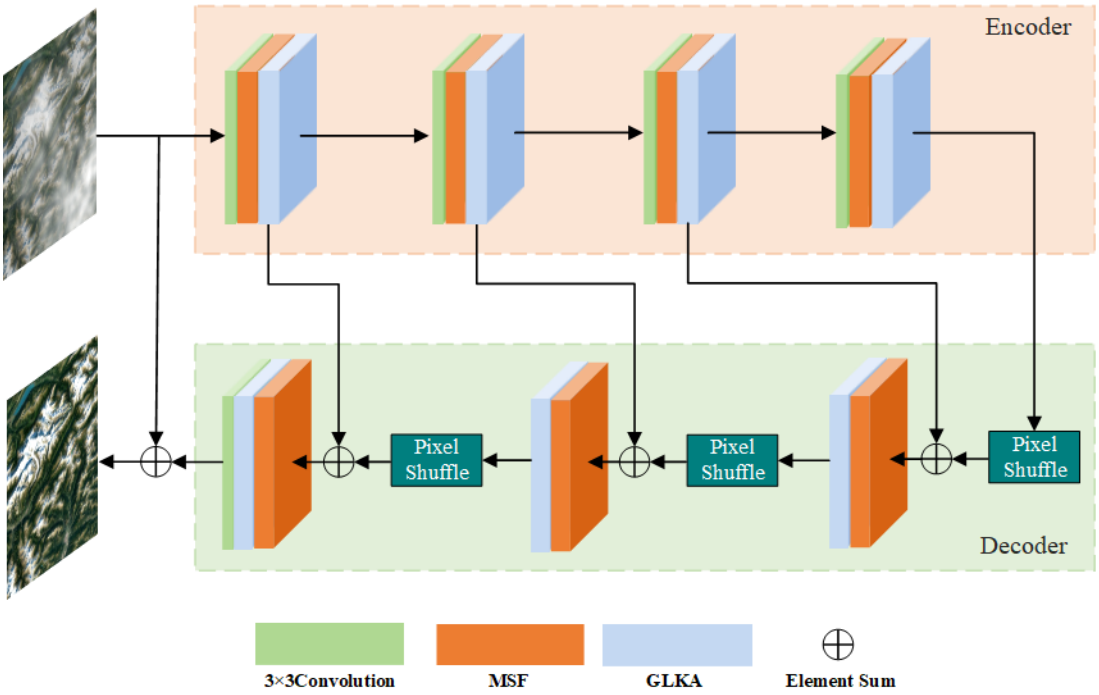


Fig. 1 Overall structure of our proposed dehazing network

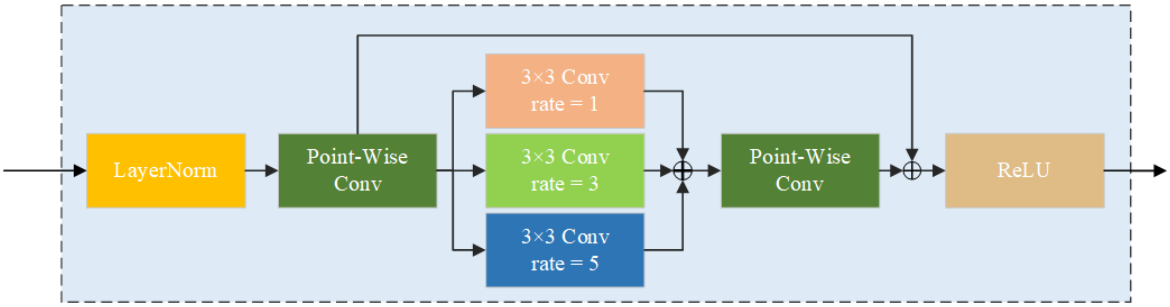


Fig. 2 Details of MSF module

(CA), spatial attention (SA), and convolutional block attention (CBA). The global contextual information is easily ignored by CA, SA, and CBA, which usually only focus on local information. S-A can establish the dependency between two global pixels while disregarding the local information and requiring a significant amount of processing. It is challenging to figure out how to combine attention with adaptivity, establish long-range dependencies, and model local information in remote sensing images with complex scenes and terrains where multiple features and landscapes need to focus on different goals. Literature [25] proposes a new convolutional attention capable of achieving self-attention adaptivity and long-range correlation while avoiding large computational and memory overheads, based on which we propose gated large kernel attention (GLKA) as shown in Fig. 3, which combines the gating mechanism with the large kernel attention to ensure adaptivity as well as the establishment of long-range dependencies through the large kernel attention, and the gating mechanism ensures that no local information is lost while long-range dependencies are established.

Large Kernel Attention

The large kernel convolutional attention decomposes the K×K convolution into three parts, which are depth convolution, depth dilation convolution, and channel convolution. For given feature map $X \in R^{c*w*h}$, following the determination of the dilation d, a (2d-1)×(2d-1) depth convolution, a (k/d)×(k/d) depth-wise dilation convolution, and a 1×1 channel convolution are carried out . By breaking down the convolution, the long-range link between pixel properties is recorded. The formula representation is:

$$LKA = PWConv(DWDConv(DWConv(X))) \quad (6)$$

where $DWConv$ represents depth convolution, $DWDConv$ represents depth dilation convolution, and $PWConv$ represents pointwise convolution.

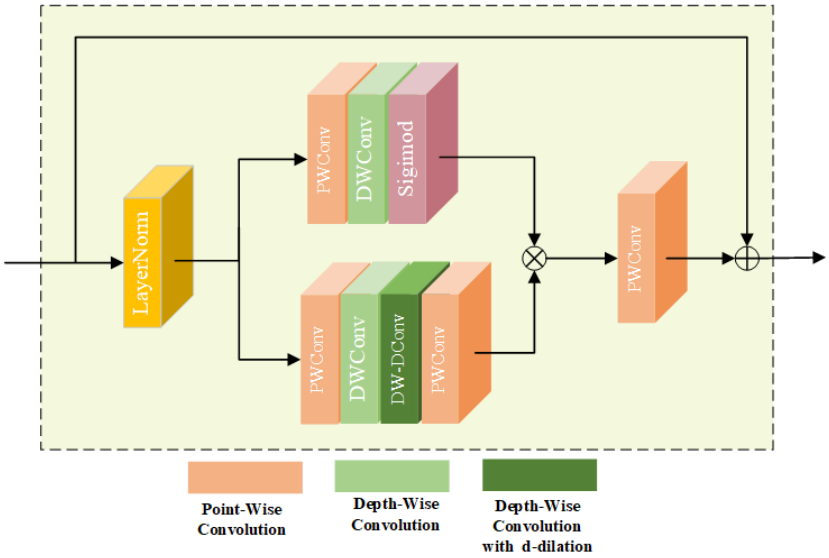


Fig. 3 Details of GLKA module

Gated Mechanism

We materialize the gating mechanism as a composite of two parallel paths, one of which introduces the large-kernel convolutional attention to establish long-range dependencies and the other of which only uses deep convolution to encode spatially adjacent pixel location information to aid in learning and recovering local image structure. For the input feature map $f(x)$, the formalization is expressed as:

$$\begin{aligned} X_1 &= LKA(PWDConv(LN(f(x)))) \\ X_2 &= DWConv(PWDConv(LN(f(x)))) \end{aligned} \tag{7}$$

$$f_{ga} = PWConv(X_1 \otimes X_2) \oplus f(x) \tag{8}$$

where X_1, X_2 represent the feature maps obtained through different paths, respectively, f_{ga} represents the final output feature map of the module.

4. Experiments

4.1 Datasets

Remote sensing images have a broad imaging range, and different remote sensing satellites have different spatial resolutions, which also results in the same location, different resolutions of remote sensing images containing different geographic feature information. And under the premise of the same visual source, remote sensing images taken by remote sensing satellites with the same spatial resolution and at different locations are frequently significantly different. For instance, the Alps, which fall under the category of natural landscape sources, are constantly covered in glaciers and snow. Remote sensing images typically show these features on the east-west oriented main mountain ranges, while the washed plains, lowlands, and hills are typically covered in forest vegetation. The intuitive color is typically composed of white glacier and snow interspersed with green forest vegetation. The east-west trending Kunlun Mountains also have snow, but much less than the Alps, and remote sensing imagery often shows the east-trending main mountain range scattered with snow, and the rest are granite, clastic and sand deposits. The intuitive color is usually a scattering of white snow covered by yellow granite and other rocks. A comparison of the two is shown in the Fig.4.

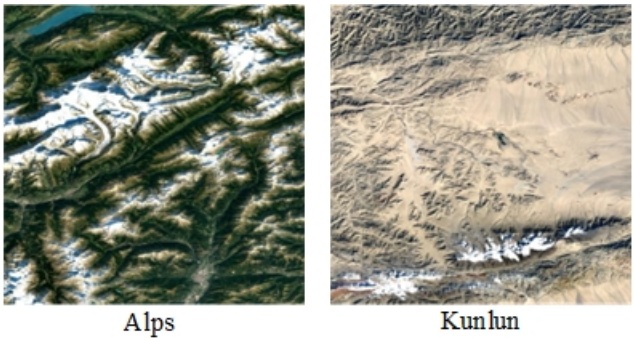


Fig. 4 Comparison of mountain ranges

We note that there are fewer existing publicly available remote sensing imagery datasets. And these datasets typically suffer from the following problems: **1**.The haze is too evenly distributed. Although haze is isotropic from a human perspective, meaning that the distribution of haze is uniform, this is due to the human perspective being too limited and the information received being restricted to what can be seen. When the entire spatial territory inside the hazy region is considered, the distribution of haze within it has a tendency to be anisotropic, meaning it is not spread evenly in space. The greater the area that is considered as a whole, the more

Table 2 Composition of different areas with the same landscape features

	Area1	Area2	Area3	Area4	Area5	Area6	Total
city	Chengdu	London	Moscow	New York	Paris	Tokyo	1800
coast	Austria	France	Seychelles	South Africa			1125
desert	Badanjeirin	Sahara	Taklamakan				1600
farmland	Crimea	Hubei(China)	Jilin(China)	Mediterranean	Nigeria		1500
forest	Amazon	Congo	Costa Rica	Yunnan(China)			1600
mountain	Alps	Andean	Himalayas	Kunlun	Rocky	Tanggula	1455

Table 3 Composition of different spatial resolutions with the same landscape features

	512	256	128	64	32	16	8	4	2	1
city				✓	✓	✓	✓	✓	✓	
coast	✓	✓	✓	✓	✓	✓	✓	✓	✓	
desert	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
farmland			✓	✓	✓	✓	✓	✓	✓	✓
forest		✓	✓	✓	✓	✓	✓	✓	✓	
mountain		✓	✓	✓	✓	✓				

apparent the anisotropy is. **2**.Remote sensing images with less spatial resolution and a single scene. Multiple spatial resolutions, or scales, are typically present in remote sensing images. As a result, the prominence and detail of the landscape vary depending on the spatial resolution. For instance, a remote sensing image with a 32-meter spatial resolution and a scale of 1:72220 frequently contains exceptionally rich landscape information, which is expressed in the variety of landscape feature information and apparent color variations. While the remote sensing image with 2 m spatial resolution and a scale of 1:4510 has significantly less information about the landscape than the remote sensing image with 32 m spatial resolution, the color change is not immediately apparent, but the latter contains significantly more comprehensive information about landscape objects. Starting from the above problems, this paper constructs a new remote sensing image dataset named DMRSI(Double Multi Remote Sensing Images).

Composition of the dataset

DMRSI consists of six types of landscape features, such as forests, deserts, farmlands, cities, coasts, and mountains, etc. As mentioned before, the information contained in remote sensing images of different regions and resolutions under the same landscape feature is not the same, so we selected remote sensing images of the same landscape feature, with different geographic regions and resolutions, to constitute the dataset. The composition of landscape features and areas is shown in Table 1 and the spatial resolution composition is shown in Table 2. Examples of the composition of different landscape features in the dataset are shown in Fig.5, and examples of the composition of the same location at different spatial resolutions are shown in Fig.6.

Remote sensing images hazing algorithms

This research uses the atmospheric scattering model as the foundation for adding haze to the clear images that have been acquired from the various places listed above in order to make hazy images. The formula states that we already have a clear image $J(x)$, and that all we need to do is compute $t(x)$ and A to obtain the hazy image. The Berlin function [26]is frequently used to mimic natural textures, which closely resemble the dispersion of haze. As a result, in this paper, we mimic the creation of the atmospheric transport map $t(x)$ using the Berlin function, which we indicate by $pl(x)$, where $pl(x) \in (0, 1)$. As for the atmospheric light A, the literature [Remote Sensing Image Dehazing Using Heterogeneous Atmospheric Light Prior] considers that the atmospheric light received at different locations is different due to the influence of haze and makes some changes to the atmospheric scattering model with the formula of

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

and also proposed to use a fixed-size window to separate the hazy image into non-overlapping patches before identifying the color of the pixel with the highest intensity in each patch as the local atmospheric light. We believe that the composition of local atmospheric light in a haze image should be affected by two aspects, one is the color of the pixel with the highest intensity in the haze image that is divided into window patches of different sizes, and the local atmospheric light with and without haze is different, here we delimit each pixel point as a window patch, and the search for the local atmospheric light of the window patch becomes the search for the atmospheric light of each pixel point. And the pixel point with haze should normally be white, and its pixel point value expressed in RGB is (255,255,255), and the pixel point without haze is itself, and its pixel point value expressed in RGB is $(r(x), g(x), b(x))$. The second is the projection ratio of the atmospheric light, which determines the intensity performance of the atmospheric light under the influence of the medium, which we denote by $K(x)$. Now we assume that the whole remote sensing image is affected by haze, then the atmospheric light $A(x)$ should be White, but due to the irregular distribution of haze and therefore the distribution of atmospheric light is also irregular, the formula for calculating $A(x)$ can be expressed as

$$A(x) = White * K(x) \tag{9}$$

We have already mentioned that we generated the atmospheric transmission map $pl(x)$ with the Berlin function, and we already know that the atmospheric transmission map is used to measure the ratio between the radiation received through the atmospheric medium and the initial scene radiance, the higher the ratio, the less it is affected by the atmospheric medium, i.e., the less it is affected by the haze, and the closer the atmospheric light at that point is to itself.

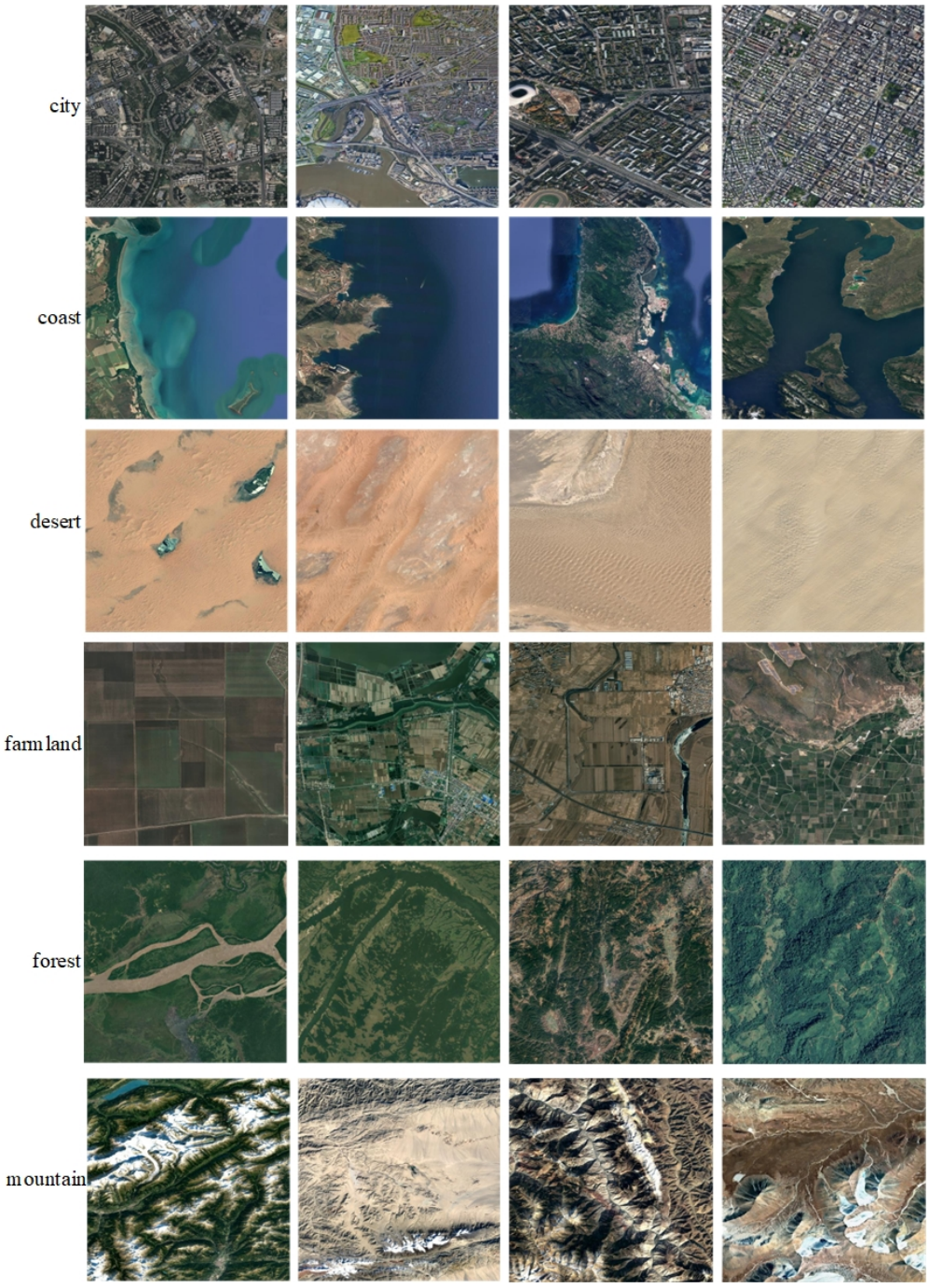


Fig. 5 Examples of different landscape feature compositions

Thus $K(x)$ can be fomulated as

$$K(x) = 1 - pl(x) \tag{10}$$

According to Formula 9 and Formula 10, the atmospheric light $A(x)$ can finally be fomulated as

$$A(x) = White * (1 - pl(x)) \tag{11}$$

After obtaining the atmospheric transmission map $pl(x)$ and atmospheric light $A(x)$, the image hazing algorithm formula can be obtained:

$$I(x) = J(x)pl(x) + White * (1 - pl(x))^2 \tag{12}$$

Dataset generation

We downloaded 36 multispectral images of urban areas, 23 coastal areas, 32 desert areas, 29 farmlands, 32 forests, and 29 mountains with resolutions ranging from 1565×862 to 32938×15220 via Google Earth. Each multispectral image is randomly cropped into 512×512 images 10 (some

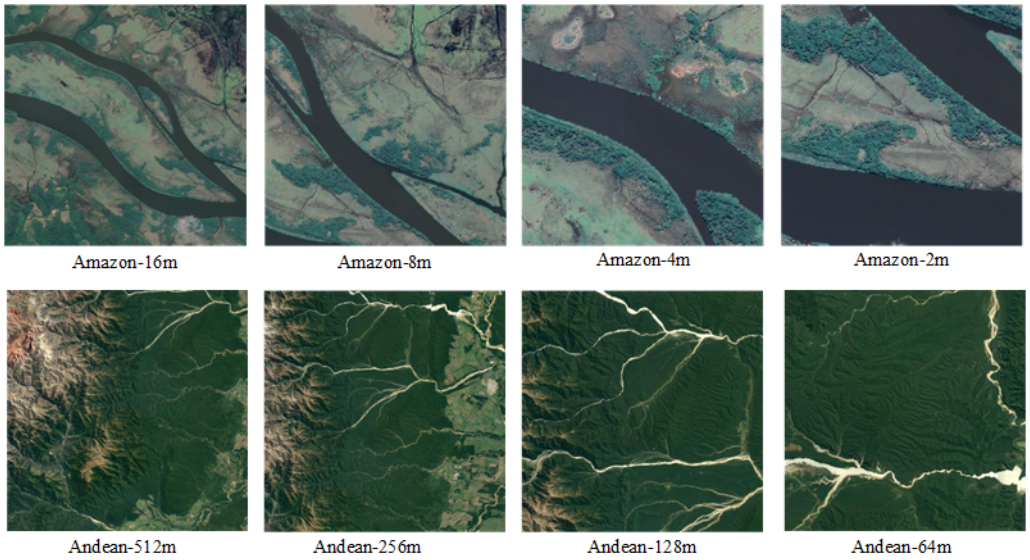


Fig. 6 Examples of different spatial resolution compositions at the same location

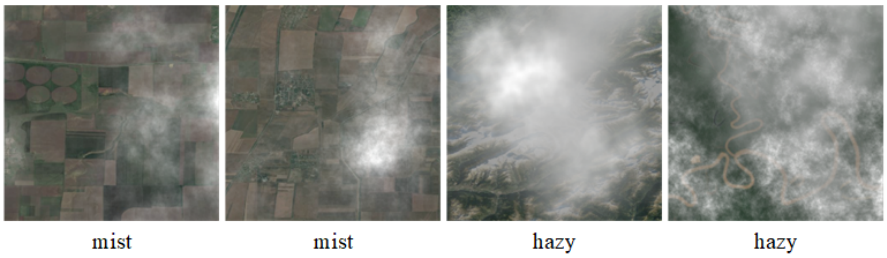


Fig. 7 Comparison of mist and hazy

Table 4 Results of quantization for different spatial resolutions

Spatial Resolution	Aod-Net			DehazeNet			DCPDN			GCA			AECR			proposed		
	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
512m	21.80	0.835	0.208	20.11	0.848	0.180	26.27	0.875	0.127	28.88	0.924	0.066	33.84	0.954	0.042	**35.03**	**0.961**	**0.035**
256m	19.84	0.825	0.197	20.41	0.837	0.165	23.65	0.876	0.124	27.14	0.941	0.051	31.85	0.966	0.031	**33.49**	**0.972**	**0.024**
128m	20.06	0.828	0.196	20.62	0.838	0.168	23.98	0.875	0.121	27.42	0.939	0.053	31.85	0.961	0.035	**33.39**	**0.967**	**0.029**
64m	20.32	0.812	0.214	20.50	0.830	0.177	24.06	0.867	0.132	27.46	0.936	0.055	30.49	0.958	0.036	**32.45**	**0.965**	**0.030**
32m	20.30	0.812	0.217	21.08	0.833	0.177	24.20	0.867	0.132	27.54	0.928	0.059	30.76	0.955	0.041	**32.26**	**0.961**	**0.035**
16m	20.42	0.802	0.200	21.12	0.824	0.173	24.03	0.859	0.133	27.43	0.925	0.065	30.09	0.948	0.045	**31.36**	**0.954**	**0.040**
8m	20.08	0.801	0.185	20.96	0.827	0.165	23.87	0.859	0.137	27.61	0.928	0.060	29.92	0.950	0.039	**31.18**	**0.956**	**0.035**
4m	18.71	0.759	0.206	20.14	0.801	0.180	22.45	0.827	0.163	26.93	0.913	0.072	29.54	0.942	0.047	**30.57**	**0.947**	**0.044**
2m	17.47	0.734	0.218	18.68	0.778	0.188	21.82	0.819	0.169	26.66	0.912	0.073	28.93	0.940	0.049	**29.63**	**0.946**	**0.046**
1m	17.76	0.721	0.277	18.05	0.744	0.255	22.40	0.763	0.244	26.31	0.884	0.113	28.43	0.908	0.086	**28.37**	**0.914**	**0.088**

areas of multispectral image cropping more), each image to add five different features of the Berlin function to generate a 9080 group of mist training set, and the second addition of haze to generate a 9080 group of hazy training set, a total of 18,160 pairs of images together to form a model training set, the contrast of mist and haze as shown in Fig.7. And from each of the above multispectral images again randomly cropped 6 512×512 images, 5/6 applying 1 different features of the Berlin function to generate 905 pairs of mist test set, 1/6 the second addition of haze to generate 181 pairs of hazy test set, a total of 1086 pairs of images together to form the test set of the model test.

4.2 Experimental setup and evaluation metrics

Experimental setup

The dataset used for the experiments is performed on our proposed DMRSI, we use Pytroch to implement the model and train the model on a single NVIDIA RTX A4000 GPU. For training, the input image is 512×512 and is randomly cropped to 256×256. we use AdamW optimizer to optimize the training results with initial values of 0.9 and 0.999 for β_1 and β_2 respectively and an initial learning rate of 1e-4.

Table 5 Results of quantization for different terrains

Terrain	Aod-Net			DehazeNet			DCPDN			GCA			AECR			proposed		
	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
city	18.59	0.769	0.207	20.59	0.827	0.159	22.34	0.866	0.127	27.07	0.943	0.045	28.80	0.960	0.031	**30.25**	**0.964**	**0.025**
coast	19.73	0.783	0.194	20.98	0.811	0.178	23.57	0.847	0.139	26.03	0.892	0.081	30.25	0.949	0.037	**31.63**	**0.957**	**0.032**
desert	20.14	0.823	0.242	18.98	0.845	0.194	24.93	0.849	0.173	28.91	0.936	0.067	33.47	0.953	0.045	**34.17**	**0.959**	**0.041**
farmland	20.16	0.780	0.217	20.52	0.791	0.211	23.77	0.820	0.166	27.27	0.902	0.085	29.29	0.927	0.066	**30.11**	**0.933**	**0.063**
forest	20.68	0.786	0.192	21.76	0.808	0.178	25.55	0.857	0.141	28.57	0.928	0.062	32.14	0.952	0.045	**33.93**	**0.958**	**0.040**
mountain	18.43	0.811	0.190	19.23	0.812	0.156	20.81	0.862	0.120	25.24	0.933	0.051	27.51	0.953	0.038	**28.95**	**0.959**	**0.032**

Table 6 Results of quantization for mluti-resolution and multi-terrain

Method	AOD-Net	DehazeNet	DCPCN	GCANet	AECR	proposed
PSNR	19.60	20.33	23.50	27.27	30.27	**31.54**
SSIM	0.792	0.817	0.851	0.900	0.950	**0.955**
LPIPS	0.208	0.179	0.144	0.064	0.043	**0.039**

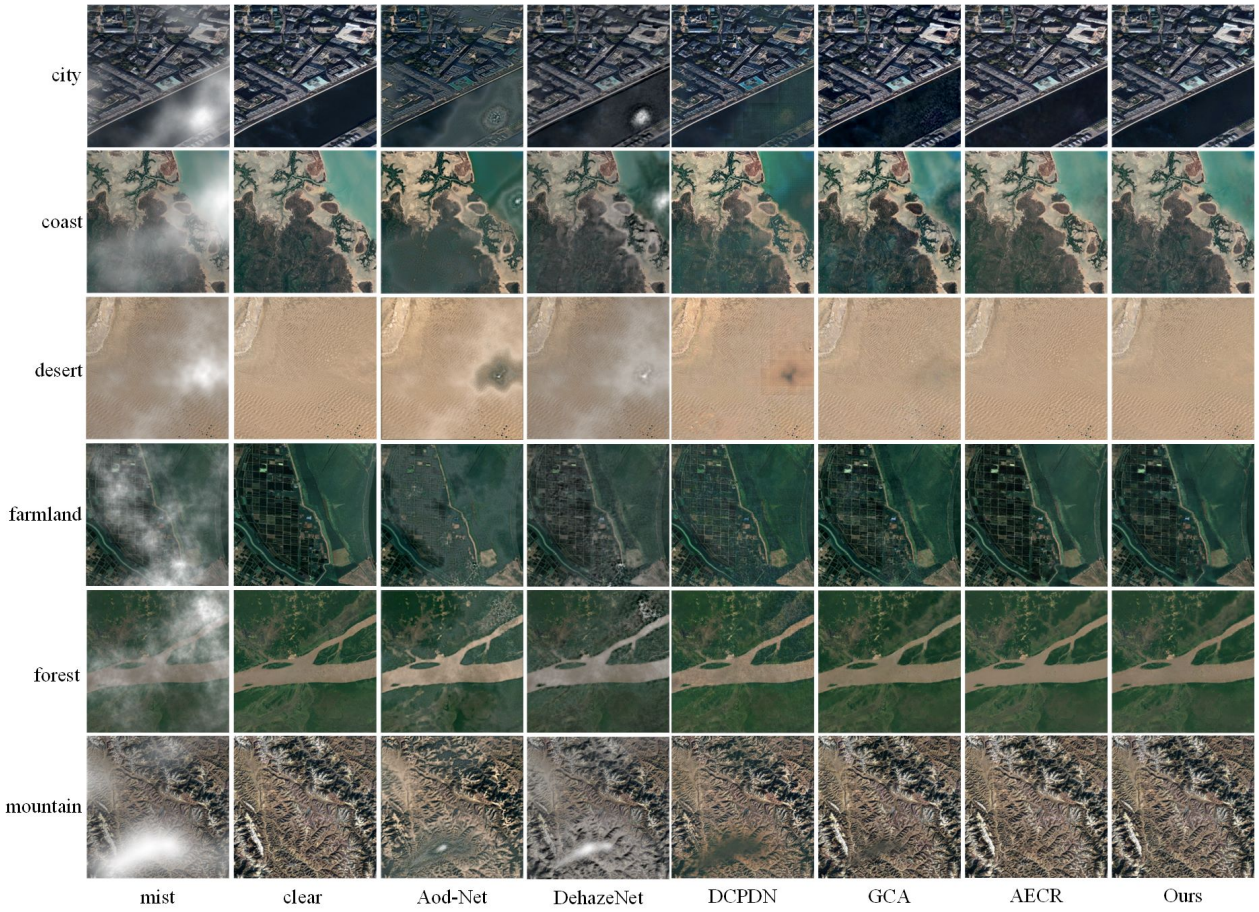


Fig. 8 Comparison of different dehazing methods for mist

Evaluation metrics

We use three quantitative metrics for quantitative evaluation, which are peak signal-to-noise ratio(PSNR) [27], structural similarity(SSIM) [28] and learned perceptual image patch similarity(LPIPS) [29]. PSNR is a reference value that measures the image quality between the maximum signal and the background noise, and the larger the value is, the better the quality of the image is. SSIM is a metric that quantifies the structural similarity between the two images, and the closer the value is to 1, the more similar the images are. SSIM is an index that quantifies the structural similarity between two images, the closer the value is to 1, the more similar the images are. LPIPS is standard to learn the inverse mapping of a generated image to Ground Truth enforces the generator to learn the inverse mapping of a reconstructed real image from a fake image and prioritizes the perceived similarity between them, where a lower value indicates that the two images are more similar.

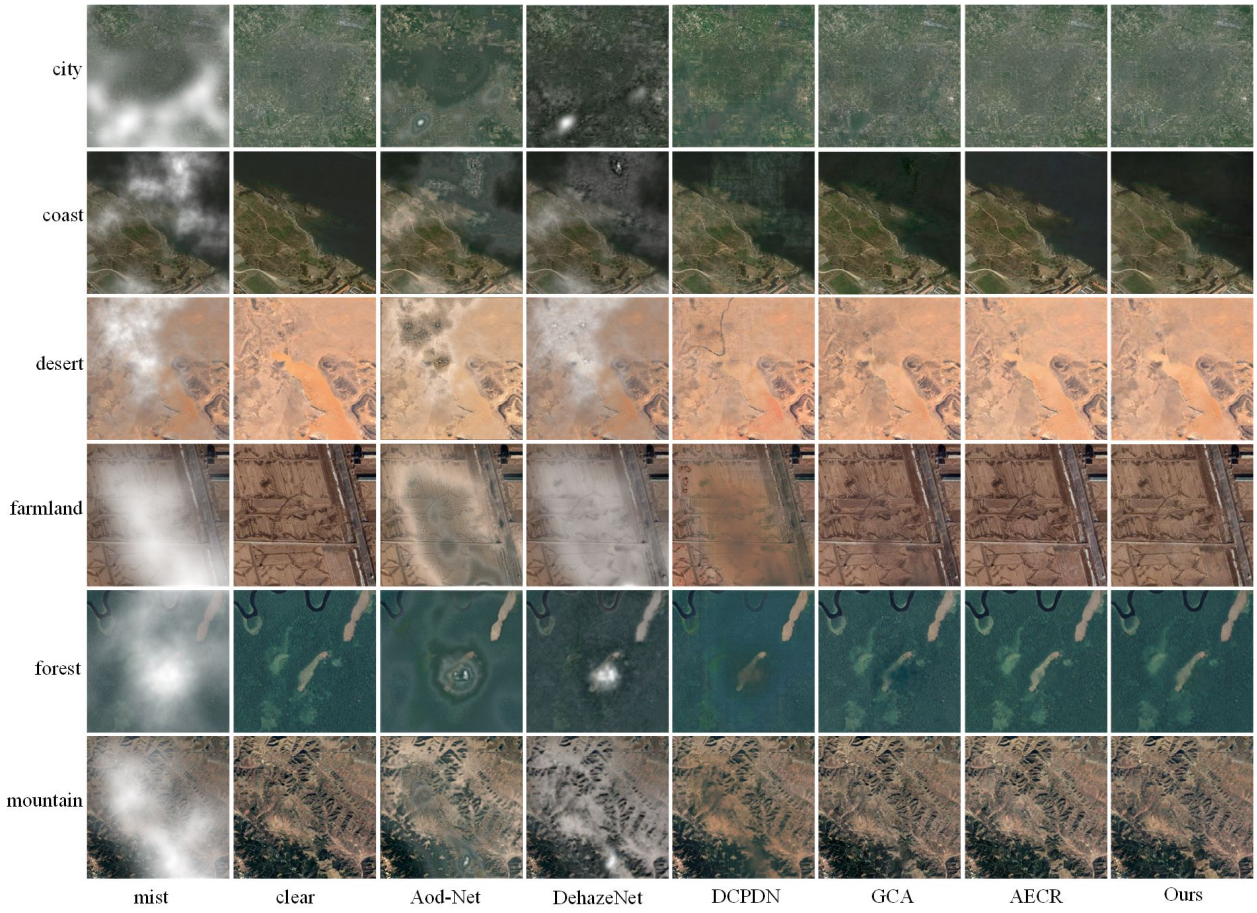


Fig. 9 Comparison of different dehazing methods for hazy

4.3 Experimental result and discussion

We applied the dehazing model to the proposed DMRSI dataset and compared our model with AOD-Net, DehazeNet, DCPDN, GCA and AECR in three scenarios of different spatial resolutions, different terrains, and mixed multi-spatial resolution-multi-terrain, and the results of quantization for different spatial resolutions are shown in Table 3, the results of quantization for different terrains in Table 4, and the results of quantization for mixed multi-spatial resolution-multi-terrain are shown in Table 5. According to Table 3, as spatial resolution increases, the metrics of AOD-Net show an overall decreasing trend, whereas the metrics of DehazeNet and DCPDN are generally more stable, with the exception of some spatial resolutions, while the metrics of GCA are generally more stable. Despite this overall decreasing trend, our method continues to outperform all previously mentioned methods at all spatial resolutions. Table 5 shows that DCPCN and GCA perform better across most terrains, with the exception of mountains, while AOD-Net and DehazeNet have poor overall dehazing metrics. In all terrain, the method we proposed performs better than every way previously mentioned. According to Table 3, our proposed method outperforms other methods in all metrics, with PSNR reaching 31.54 dB, SSIM reaching 0.955, and LPIPS dropping to 0.039, Compared to the best resultant network AECR, our network PSNR performance improves up to 3.44%, SSIM improves up to 0.5%, and LPIPS improves up to 9.3%.

We also made visual qualitative comparisons between the proposed method and various dehazing methods, and the comparison graphs of the mist in different terrains are shown in Fig.8, and the comparison graphs of the hazy are shown in Fig.9. In the mist scenarios, DehazeNet suffers from color distortion problems and has poor dehazing ability, otherwise, most of the methods can achieve good results, but AOD-Net has a very obvious halo phenomenon and serious color distortion problems; DCPDN and GCA do not have serious color distortion problems, but there is a blurring of detail phenomenon. In hazy scenarios, the dehazing effect of AOD-Net and DCPCN is poor, and GCA is better than the first two methods but still suffers from the problems of incomplete dehazing and poor detail recovery, AECR is still inferior to our proposed method, although it achieves better visualization results than the previous methods, whereas our proposed method outperforms the above mentioned methods in terms of both dehazing ability and detail recovery, whether

in the conditions of mist or hazy, and exhibits good dehazing ability, and achieves good results in the aspects of color contrast and detail recovery.

4.4 Ablation experiments

We discuss the influence of two crucial modules in our proposed model, the MSF and GLKA, on the performance of the network in order to assess the efficacy of various modules in our suggested model. We add the MSF and GLKA modules as models 1 and 2, respectively, and replace the MSF and GLKA modules with regular convolutional blocks as the baseline. The quantization results are displayed in Table 4.

Table 7 Comparison of quantitative results of different modules

Method	MSF	GLKA	PSNR	SSIM	LPIPS
Baseline			31.20	0.950	0.044
Model1	✓		28.50	0.931	0.063
Model2		✓	31.26	0.953	0.040
proposed	✓	✓	**31.54**	**0.955**	**0.039**

The results indicate that adding various modules can improve or worsen network performance to varying degrees. The PSNR of the model with the addition of the MSF module alone falls by 2.7 dB or 8.654% compared to BASELINE, the SSIM falls by 0.019 or 2%, and the LPIPS rises by 0.019 or 43.182%, this is owing to the fact that, although it can enhance the number of feature maps, the inclusion of the MSF module alone does not add a logical This is due to the fact that, despite the fact that adding an MSF module alone can increase the number of feature maps, it does not provide a suitable feature map selection process, which lowers the model's overall resilience. The model with the GLKA module alone improves by 0.06dB or 0.192% compared to the baseline, while SSIM improves by 0.003 or 0.316% and LPIPS decreases by 0.004 or 9.091%. This improvement in model performance is made possible by the attention mechanism of the GLKA module.

5. Conclusion

In this paper, a remote sensing images dehazing model with an encoder-decoder structure is proposed for the construction of flight simulator visual system, the end-to-end architecture enables the model to directly achieve image dehazing by learning the residuals and recovering the image characteristics, which addresses the issue of the lack of stability and robustness of conventional image dehazing techniques. The MSF module and GLKA module are designed for feature extraction as well as feature fusion for remote sensing photos with multi-complex terrain and multi-spatial resolution, improving the stability of the model. In addition, this paper also collects and constructs a remote sensing image dataset with different concentrations of inhomogeneous haze to evaluate the proposed method. The experiment proves that compared with other methods, the proposed method shows strong performance in image dehazing and image recovery, regardless of mist or hazy, and verifies the effectiveness of the proposed model in remote sensing images with multiple spatial resolutions and complex terrains to remove haze, which is suitable for flight simulator visual system.

Funding

This study was co-supported by the Research on Simulator Three-Dimensional View Modeling Technology and Database Matching and Upgrading Methods, and the Key Laboratory of Civil Aviation Flight Technology and Flight Safety(FZ2022KF08).

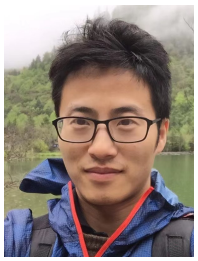
References

[1] J.M. Rolfe and K.J. Staples, Flight simulation, Cambridge University Press, 1988.
[2] W. Ge, Z. Wang, G. Wang, S. Tan, and J. Zhang, "Remote sensing image super-resolution for the visual system of a flight simulator: Dataset and baseline," Aerospace, vol.8, no.3, p.76, 2021.
[3] Y. Cao, "Design and implementation of certain type flight test simulation platform visual system," Proceedings of the 2020 4th International Symposium on Computer Science and Intelligent Control, pp.1–5, 2020.
[4] Y. He, C. Li, and X. Li, "Remote sensing image dehazing using heterogeneous atmospheric light prior," IEEE Access, vol.11, pp.18805–18820, 2023.
[5] E. Hadjidemetriou, Use of histograms for recognition, Columbia University, 2002.
[6] E.H. Land and J.J. McCann, "Lightness and retinex theory," Josa, vol.61, no.1, pp.1–11, 1971.
[7] R. Fries and J. Modestino, "Image enhancement by stochastic homomorphic filtering," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol.27, no.6, pp.625–637, 1979.
[8] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," IEEE transactions on pattern analysis and machine intelligence, vol.33, no.12, pp.2341–2353, 2011.
[9] C. Li, C. Yuan, H. Pan, Y. Yang, Z. Wang, H. Zhou, and H. Xiong, "Single-image dehazing based on improved bright channel prior and dark channel prior," Electronics, vol.12, no.2, p.299, 2023.
[10] C. Li, H. Yu, S. Zhou, Z. Liu, Y. Guo, X. Yin, and W. Zhang, "Efficient dehazing method for outdoor and remote sensing images," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2023.
[11] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, "Dehazenet: An end-to-end system for single image haze removal," IEEE transactions on image processing, vol.25, no.11, pp.5187–5198, 2016.
[12] W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, and M.H. Yang, "Single image dehazing via multi-scale convolutional neural networks," Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14, pp.154–169, Springer, 2016.
[13] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, "Aod-net: All-in-one dehazing network," Proceedings of the IEEE international conference on computer vision, pp.4770–4778, 2017.
[14] H. Zhang and V.M. Patel, "Densely connected pyramid dehazing network," Proceedings of the IEEE conference on computer vision and pattern recognition, pp.3194–3203, 2018.
[15] D. Chen, M. He, Q. Fan, J. Liao, L. Zhang, D. Hou, L. Yuan, and G. Hua, "Gated context aggregation network for image dehazing and deraining," 2019 IEEE winter conference on applications of computer vision (WACV), pp.1375–1383, IEEE, 2019.
[16] J. Guo, J. Yang, H. Yue, H. Tan, C. Hou, and K. Li, "Rsdehazenet: Dehazing network with channel refinement for multispectral remote

sensing images," IEEE Transactions on geoscience and remote sensing, vol.59, no.3, pp.2535–2549, 2020.
[17] H. Wu, Y. Qu, S. Lin, J. Zhou, R. Qiao, Z. Zhang, Y. Xie, and L. Ma, "Contrastive learning for compact single image dehazing," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.10551–10560, 2021.
[18] W. Ge, Y. Lin, Z. Wang, G. Wang, and S. Tan, "An improved u-net architecture for image dehazing," IEICE TRANSACTIONS on Information and Systems, vol.104, no.12, pp.2218–2225, 2021.
[19] X. Chen, Y. Li, L. Dai, and C. Kong, "Hybrid high-resolution learning for single remote sensing satellite image dehazing," IEEE Geoscience and Remote Sensing Letters, vol.19, pp.1–5, 2021.
[20] Z. He, C. Gong, Y. Hu, F. Zheng, and L. Li, "Multi-input attention network for dehazing of remote sensing images," Applied Sciences, vol.12, no.20, p.10523, 2022.
[21] Z. He, C. Gong, Y. Hu, and L. Li, "Remote sensing image dehazing based on an attention convolutional neural network," IEEE Access, vol.10, pp.68731–68739, 2022.
[22] S. Li, Y. Zhou, and W. Xiang, "M2scn: Multi-model self-correcting network for satellite remote sensing single-image dehazing," IEEE Geoscience and Remote Sensing Letters, vol.20, pp.1–5, 2022.
[23] J. Wei, Y. Cao, K. Yang, L. Chen, and Y. Wu, "Self-supervised remote sensing image dehazing network based on zero-shot learning," Remote Sensing, vol.15, no.11, p.2732, 2023.
[24] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18, pp.234–241, Springer, 2015.
[25] M.H. Guo, C.Z. Lu, Z.N. Liu, M.M. Cheng, and S.M. Hu, "Visual attention network," Computational Visual Media, pp.1–20, 2023.
[26] K. Perlin, "Improving noise," Proceedings of the 29th annual conference on Computer graphics and interactive techniques, pp.681–682, 2002.
[27] M.M. Petrou and C. Petrou, Image processing: the fundamentals, John Wiley & Sons, 2010.
[28] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE transactions on image processing, vol.13, no.4, pp.600–612, 2004.
[29] R. Zhang, P. Isola, A.A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," Proceedings of the IEEE conference on computer vision and pattern recognition, pp.586–595, 2018.

Qi Liu is currently a master's student at the college of Computer Science, Chengdu University of Information Technology. His research interests include remote sensing image enhancement and three dimensional reconstruction of large scenes.

Bo Wang is currently a master's student at the college of Computer Science, Chengdu University of Information Technology. His research interests include semantic segmentation of remote sensing images and three dimensional semantic segmentation



Shihan Tan is currently an assistant research professor at the college of Computer Science, Chengdu University of Information Technology. His research interests relate to computer graphics and VR development of real-time rendering systems for flight simulation.

Shurong Zou is currently a professor at the college of Computer Science, Chengdu University of Information Technology. Her research interests include graphics and image processing, artificial intelligence.

Wenyi Ge is currently an assistant research professor at the college of Computer Science, Chengdu University of Information Technology. His research interests include graphics and image processing, remote sensing image enhancement.