A DAU-Net-ConvLSTM Model for Daytime Sea Fog Segmentation

Xiaokang Hu, Taorong Qiu, Yiqi Liao, Jing Wang, and Liangwei Lin*

Abstract—**Sea fog poses risks to coastal activities, necessitating effective monitoring and early warning systems. This study introduces a deep learning approach tailored for sea fog segmentation, considering its nonlinear multiscale variability, textural patterns, and temporal aspects. The U-Net model serves as the foundational network, enhanced by asymmetric multi-scale convolution modules to create the DAU-Net. This improved model effectively identifies sea fog features in images. Integrating the DAU-Net with ConvLSTM results in the DAU-Net-ConvLSTM model, which uses bidirectional ConvLSTM for processing temporal sequence data and refining segmentation outcomes. Comparative testing against seven segmentation models on augmented sea fog datasets revealed our model's superiority, achieving a 90.4% Kappa score and 86.4% MIOU. It outperforms existing CNN models like U-Net, U-Net++, Deeplab v3, and temporally-focused models like RNN, STGRU, 3D CNN-LSTM. This highlights its robust segmentation capabilities and potential for real-world applications.**

Index Terms—**Sea fog image; image segmentation; U-Net; DAU-Net; ConvLSTM**

I. Introduction

SEA fog introduces considerable hazards to maritime operations, including industrial and tourist activities, potentially causing extensive harm to human life and property. Therefore, it is of utmost importance to research the real-time monitoring and predictive warning of sea fog by utilizing advanced and efficient technologies from two aspects practical demands and safety requirements. Contrary to land fog detection, sea fog detection is predominantly dependent on remote sensing satellites, due to geographical limitations. These satellite-acquired remote sensing images of maritime areas are then analyzed, either manually or technologically, to confirm the presence of sea fog during specific intervals. If sea fog is identified, subsequent region segmentation is necessary for further analysis. Conventional methods involving manual inspection and segmentation of remote sensing images suffer from evident inefficiencies, struggling to fulfill real-time requirements and provide effective warning and prediction services. However, with the development of machine learning, particularly deep learning, methods based on deep learning have been widely applied in image detection, classification, and segmentation. These methods show remarkable competence in addressing non-linear issues inherent to semantic segmentation, allowing for a superior grasp of sea fog features, and consequently enabling more precise predictions and segmentation. Therefore, exploring the implementation of deep learning techniques for sea fog segmentation in remote sensing images is not only viable and cutting-edge but also offers substantial academic and practical value.

Liu [1] introduced the Multidimensional Attention and Feature Enhancement (MA-FE) method, which significantly improves the accuracy and feature representation in remote sensing image scene classification by integrating multidimensional attention and feature enhancement modules. Kim [2] employed VGG19 and ResnNet50 as pre-trained models, with the training and testing dataset extracted from six days of GOCI images of the coastal region of the Korean Peninsula in March 2015. This model adopted a transfer learning method, moving convolutional neural networks pre-trained on natural image datasets to maritime datasets, achieving a match accuracy rate of 96.3% with both VGG19 and ResNet50 for CNN-TL training data. Addressing the challenges posed by the limited resolution and spectral information in RGB preview images for cloud detection in remote sensing, a recent study introduced GANet[3], a novel system with an encoder-decoder architecture that efficiently fuses semantic and spatial features, demonstrating competitive performance on multiple datasets. The Ran team [4] introduced a fog detection method based on a deep learning algorithm, called DDF-Net, which leverages digital elevation model (DEM) data as auxiliary information to separate fog and low-level clouds, integrating squeeze-and-excitation networks (SE-Net) to optimize information extraction under different solar zenith angles (SZA), eliminating spectral feature differences within large regions. This study used the advanced Himawari 8 imager (AHI) data from the Himawari 8 (H8) satellite as the primary data source and compared the proposed model with the traditional threshold-based brightness temperature difference (BTD) method. Results revealed that the DDF-Net method achieved an average POD, FAR, and CSI of 84.0%, 16.4%, and 72.0%, respectively, during dawn and dusk, as well as 83.7%, 15.8%, and 72.6%, outperforming the BTD method. However, due to the characteristics of passive sensor images, it is challenging to detect fog below the clouds; additionally, due to the lack of ground observation points, it is difficult to thoroughly evaluate different types of fog. Chunyang et al. [5] employed the U-Net deep learning model to construct a sea fog detection model for MODIS multispectral images, proving to be more flexible and intelligent than traditional threshold methods, improving the accuracy of sea fog detection. The kappa score on the test set reached 0.972, and the

Manuscript received January 10, 2024; revised July 22, 2024.

This work was Supported by Fujian Provincial Natural Science Foundation of China, China Grant(No. 2022J011079).

Taorong Qiu is a Professor of the School of Mathematics and Computer Science, Nanchang University, Jiangxi, China (e-mail: qiutaorong@ncu.edu.cn).

Liangwei Lin is a researcher of Zhangzhou Meteorological Bureau, Fujian, China (Corresponding author, e-mail: 13709301618@qq.com).

Yiqi Liao is a graduate student of Nanchang University, Jiangxi, China (e-mail: 2943949376@qq.com).

Xiaokang Hu is a postgraduate student of the School of Mathematics and Computer Science, Nanchang University, Jiangxi, China (e-mail: 826901412@qq.com).

Jing Wang is a graduate of the School of Mathematics and Computer Science, Nanchang University, Jiangxi, China (e-mail: wj_20221@163.com).

overall accuracy was 0.98. However, only visible light images were used, and the atmospheric information contained in the infrared channels was not effectively utilized. The method of processing small blocks of images and simple splicing led to obvious mosaic traces, and the judgment accuracy of sea fog and cloud edges still needs improvement. In 2022, Chen et al. [6] conducted a study on sea fog detection in the Arctic region based on CALIOP and MODIS data, aiming to solve how to improve the accuracy of sea fog detection in polar regions. They used the Deep Neural Network (DNN) model to invert the cloud top height and judged whether it was sea fog based on the cloud height. Mean absolute error (MAE) and root mean square error (RMSE) were used as evaluation indicators and compared with the inversion results of MODIS's cloud height product and BP neural network. The results showed that the MAE of using DNN to invert cloud top height was about 701.140 m, better than the result of MODIS cloud height product (lower by about 1774.280 m), better than the result of BP neural network (lower by about 781.005 m), indicating that using DNN model can better invert cloud top height, improving the accuracy of sea fog detection. However, due to the time and location differences of CALIOP and MODIS observations, erroneous matching data may be introduced into the training dataset, affecting the model's generalization ability.Zhuo Li et al. [7] developed the MRBU-Net-WD model, an enhanced version of U-Net that effectively segments lung nodules by integrating residual 3D convolutions and multiscale dense connections, addressing pixel imbalance with a weighted Dice loss function, and evaluated it using the LUNA-16 dataset, showing superior performance over existing models.

In summary, while deep learning methods have been applied to remote sensing images for sea fog detection and segmentation research, the study of sea fog detection and segmentation based on deep learning is still in its infancy. Many issues require further exploration and resolution. For example, there is a lack of a shared dataset for evaluating sea fog segmentation performance. There is a need for effective construction of deep learning models to segment sea fog images during the day, night, and in all weather conditions. Questions also exist around how to construct models that can effectively segment clouds and sea fog, how to detect sea fog obscured by clouds, and how to fully utilize the spatiotemporal properties of sea fog images. It's clear that while some progress has been made, much work remains in this area of study.

This study explores deep learning models for sea fog segmentation in remote sensing images, aiming at the characteristics of sea fog data. In this study, after comprehensive consideration of several relatively advanced models based on deep learning, the U-Net model is chosen as the foundation model to be improved into the proposed sea fog segmentation model. The selected U-Net model is then optimized by introducing asymmetric multi-scale convolution modules to enhance the model's feature representation capability and enable more effective extraction of multi-scale features of sea fog. Finally, in order to make full use of the temporal characteristics of sea fog, a sea fog segmentation model based on deep learning is proposed by combining the optimized U-Net model and ConvLSTM.

II. Remote Sensing Sea Fog Image Datasets

Firstly, we collect an original dataset of remote sensing images of sea fog. These images are sourced from the geosynchronous orbit meteorological satellite "Himawari-8," capturing images from 117°E to 128°E longitude and 29°N to 41°N latitude. To reduce the workloads of manual annotations and ensure the precision of labels, we first use the SLIC superpixel segmentation algorithm [8] to segment the pseudo-color images, and then manually annotate the images based on the segmented superpixels. The labeled sea fog dataset contains only 2,562 images, including 1,501 foggy and 1,061 non-foggy images.

Secondly, we enhance the original dataset through data augmentation. Considering the relatively small number of original images, traditional data augmentation methods and DCGAN[9] (Deep Convolutional Generative Adversarial Networks) are utilized to enhance the generalizability of the model trained on this limited dataset. By traditional methods, 1,501 foggy and 1,061 non-foggy images are added to the original dataset, and by the DCGAN method, 1,365 foggy and 1,085 non-foggy images are generated.

Finally, the training dataset for the model consists of the original images, the generated images by traditional data augmentation methods, and the DCGAN method, a total of 7,574 images. There are 4,367 foggy and 3,207 non-foggy images.

III. A Daytime Sea Fog Image Segmentation Model Incorporating ConvLSTM Mechanism in DAU-Net

A. The DAU-Net Model and the ConvLSTM Model

1) The DAU-Net Model: The U-Net architecture has become a prominent solution in the field of image segmentation, widely acclaimed for its robust generalization capabilities and elegantly simplistic design. Despite these strengths, U-Net still has challenges, notably including limitations in feature extraction capacity and suboptimal utilization of multi-scale features. The appearance of DAU-Net addresses these issues. By integrating principles from residual learning[10], dense learning[11], hybrid attention mechanisms[12], and multi-scale feature handling, the DAU-Net redesigns the down-sampling and up-sampling branches of the model to achieve a richer representation and deeper understanding of image features. Furthermore, the application of unique training techniques for segmentation has contributed to a noticeable boost in segmentation accuracy. This approach contributes to the superior performance of DAU-Net compared to other traditional segmentation models which demonstrates its potential as a general tool applied in complex image analysis tasks. Figure 1 presents the detailed structure of DAU-Net which maintains U-Net's foundational architecture, and consists of an encoder and a decoder. In the encoder, four down-sampling layers are performed, including asymmetric multi-scale convolution functions alongside 2x2 max-pooling operations. This design aids in not only the effective extraction of relevant features but also the spatial compression of the image.

Conversely, the decoder performs four corresponding up-sampling layers, employing a unique fusion of asymmetric multi-scale convolution with deconvolution techniques. This

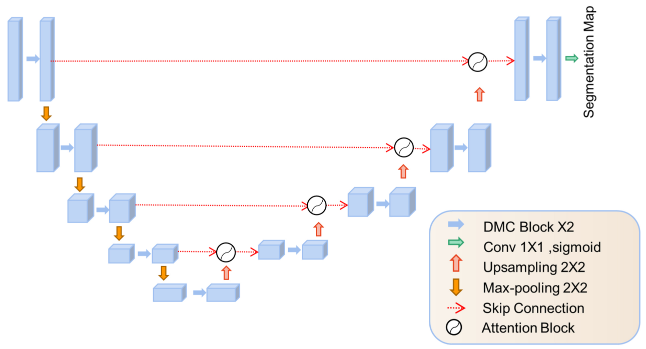

Fig. 1. DAU-Net Network Structure.

creative approach facilitates the precise restoration of the image's original resolution.

A further innovation lies in the incorporation of attention modules within the skip connections that mirror each layer between the decoder and encoder. Feature maps input into these attention modules are from the current down-sampling layer and the preceding up-sampling layer. The resulting attention weight coefficients are then multiplied by the features obtained from the preceding up-sampling layer. The merging of these processed features results in a refined feature map, optimized by the strategically implemented attention mechanism. This multifaceted construction shows the DAU-Net's advanced capability to efficiently analyze and interpret complex image data.

2) The ConvLSTM Model: ConvLSTM (Convolutional Long Short-Term Memory) represents an innovative model that synergistically integrates convolutional neural networks (CNNs) and long short-term memory (LSTM) networks. This fusion serves the specialized purpose of processing image data within temporal sequences.

By embedding convolutional layers within the LSTM framework[13], ConvLSTM exhibits an enhanced capability to discern and interpret spatial features in images. The structure of ConvLSTM is marked by a complex arrangement of input gates, forget gates, output gates, and convolutional layers. In the input gate, a convolutional layer is harnessed to confirm the relationship between the current and previous time steps' inputs and hidden layers. This dynamic correlation actively guides and regulates the information intake of the current time step.

Similarly, the forget gate employs a convolutional layer to evaluate the correlation between the current input and the hidden layer of the preceding time step, effectively determining which prior information is retained or forgotten. In the output gate, a convolutional layer is conducted for exploring the relationship between inputs of the current time step and information of the preceding hidden layer, which in turn determines the transmission of information of the current hidden layer.

In this way, ConvLSTM overcomes conventional limitations and exhibits the flexibility to concurrently analyze spatial and temporal features within time-sequenced image data. The unique design of ConvLSTM shows its potential applications in various areas that needs the integration of spatial and temporal data and presents its significance in contemporary machine learning research.

The computational process for each ConvLSTM unit can be delineated as follows:

$$i_t = \sigma(x_t * W_{xi} + h_{t-1} * W_{hi} + b_i) \tag{1}$$

$$f_t = \sigma(x_t * W_{xf} + h_{t-1} * W_{hf} + b_f) \tag{2}$$

$$c_t = c_{t-1} \odot ft + it \odot Tanh(xt * Wxc + ht - 1 * Whc + bc) \tag{3}$$

$$o_t = \sigma(x_t * W_{xo} + h_{t-1} * W_{ho} + b_o) \tag{4}$$

$$h_t = o_t \odot Tanh(ct) \tag{5}$$

In this framework, the symbol * signifies convolution, while $\odot$ refers to element-wise multiplication, and $\sigma(\cdot)$ is indicative of the Sigmoid activation function. The variables i_t,f_t,c_t, and o_t correspond to the input gate (i), forget gate (f), cell state (c), and output gate (o), respectively. The notation W and b are used to represent the convolution kernel and the biases associated with each gate, whereas x_t and h_t denote the input and output feature maps, respectively. The equation formulates that the output at a given time point t, expressed as h_t, is ascertained by the current input x_t, in conjunction with the preceding states c_(t-1) and h_(t-1). This mechanism allows ConvLSTM to leverage historical data during the prediction phase.

B. Integration of ConvLSTM Mechanism with DAU-Net for the Segmentation of Daytime Sea Fog Images

In remote sensing images, sea fog exhibits significant variations in scale and diverse textural shapes, and its formation and evolution follow specific temporal patterns. In a short time, the changes in appearance and motion of sea fog in preceding moments can provide a valuable reference, assisting in determining the current location. To achieve optimal segmentation effects in the developed model based on deep learning for sea fog, it is imperative to consider strategies for the efficient extraction of sea fog features and the comprehensive utilization of its temporal information.

DAU-Net, by leveraging asymmetric multi-scale convolution and attention mechanisms, optimizes the U-Net model, proficiently addressing the challenges of the inadequate extraction of sea fog features and the need to concentrate on particular sea fog image attributes. Through the application of ConvLSTM, it becomes possible to effectively extract inter-sequence features from a series of images. In light of the inherent characteristics of sea fog data, this study introduces the DAU-Net-ConvLSTM model, built upon the DAU-Net as the primary architecture and integrated with ConvLSTM, aiming to enhance the segmentation performance of sea fog images.

In the DAU-Net-ConvLSTM, the DAU-Net model is first utilized for the initial segmentation of the sea fog image. Subsequently, ConvLSTM, as a post-processor, extracts inter-regional sea fog feature information from pre-segmented images of adjacent fog areas. Ultimately, the DAU-Net-ConvLSTM model leverages temporal sequence information, studying the dynamic variations of sea fog and evolutionary

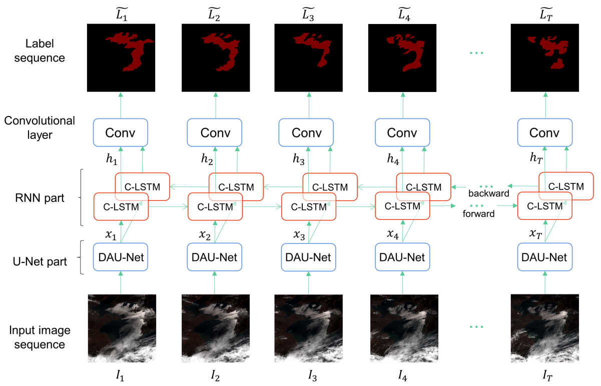

Fig. 2. DAU-Net-ConvLSTM Model Network Structure.

patterns to enhance the accuracy of the segmentation of the target area. More specifically, ConvLSTM is designed to capture contextual information within the time series, synthesizing both past and future image features during predictions. This integration serves to increase the accuracy of the segmentation model across the time series. Prior to the final activation function within the backbone model, pre-segmentation outputs are retrieved from the feature maps, functioning as input data within the DAU-Net-ConvLSTM model. These feature maps consist of two channels: channel 1 corresponds to the background label, and channel 2 is associated with the sea fog label. Consequently, each pixel within the feature map can be interpreted as representing the possibility of belonging either to the background or sea fog, depending on the respective channel values. For instance, a higher value in channel 1 implies that the model perceives the pixel as a more likely part of the background, whereas a greater value in channel 2 signifies a higher probability of belonging to the sea fog region.

As depicted in Figure 22, the DAU-Net-ConvLSTM model architecture receives an input comprising the image sequence at time point t denoted as $I = I_t|t = 1, 2, \ldots t$,and produces an output of the predicted label mapping sequence $\widetilde{L} = \widetilde{L}_t|t = 1, 2, \ldots t$.This method integrates two core components: DAU-Net and ConvLSTM. During prediction, DAU-Net meticulously assesses the spatial features of each input image I_t to generate segmented result images x_t, which are subsequently used as inputs for ConvLSTM that capitalizes on both time-series information and pre-segmented images from adjacent sea fog regions. This integration allows for precise extraction of inter-fog area features, as well as subsequent refinement of the target area's segmentation results. A bidirectional ConvLSTM is deployed here for its ability to concurrently harness both past and future information during prediction. This design enhances the capture of dynamic fluctuations and context in the time series. The forward and backward flows within the model demonstrate distinct directional roles, with the former facilitating information passage from past to future and the latter inversely transmitting information from future to past. This bidirectional scheme enables ConvLSTM to delve deeper into the temporal series features, augmenting both the precision and robustness of sea fog segmentation.

In a more detailed view, the output of ConvLSTM consists of per-pixel feature maps h_t, at each time point t. To formulate the probabilistic label mapping $\widetilde{L}_t$ the outputs from both the forward and backward ConvLSTM flows are concatenated, followed by a convolution operation, culminating in the final segmentation result. This approach ensures a nuanced representation of the image sequence, leading to a more accurate model.

IV. Model Training, Comparative Experiment, and Result Analysis

A. Model Training Parameter Configuration and Segmentation Effect Evaluation Metrics

1) Parameter Configuration: Hardware Environment: The processor used in the experiments of this paper is AMD Ryzen 5 5600X 6-Core Processor, with a clock frequency of 3.7 GHz, RAM of 32GB, and GPU of GTX 1080Ti.

Software Environment: The operating system is Windows 10, and the training environment is Python3.7.

Throughout the training process in this study, we employed the gradient-based optimizer, Adam[14], a tool that dynamically adjusts the learning rate for each parameter by leveraging estimates of the gradient's first and second moments. Considering time constraints, the total number of epochs was established at 100 for each iteration. We set an initial learning rate of 0.001 and standardized the batch size to 210, meaning that each training batch consisted of 210 individual images. Given that the dataset selected 7 images of sea fog a day, and each batch contained the number of images for a month, the value of the temporal window was configured to 7 which corresponded to the quantity acquired during a continuous daily interval.

Given that the challenge of sea fog segmentation pertains to pixel-level segmentation, the current study employs the Dice loss function[15] instead of the more conventional cross-entropy loss function. The mathematical representation of the Dice loss function is articulated below:

$$\mathrm{DICE}\left(y, y'\right) = 1 - \frac{2\sum_i^N y_i y_i'}{\sum_i^N y_i^2 + \sum_i^N \left(y_i'\right)^2} \tag{6}$$

where y denotes the true value labels, $y\prime$ symbolizes the predicted value labels, and N corresponds to the total number of pixel points.

2) Evaluation Metrics: In order to evaluate the performance of the model for sea fog segmentation, this study establishes a set of criteria as follows:

(1) Kappa

A method for evaluating the performance of image segmentation tasks involves measuring the agreement between predicted and actual classifications beyond chance. The Kappa score quantifies the level of agreement between the predicted segmentation and the true segmentation, accounting for the agreement that could occur by random chance. The mathematical definition for this metric is articulated below:

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

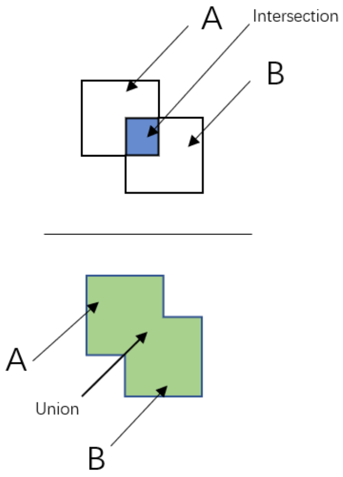

Fig. 3. Intersection over Union

$$\kappa = \frac{P_o - P_e}{1 - P_e}$$

Within this context, TP denotes true positives, FP signifies false positives, TN stands for true negatives, and FN designates false negatives. Additionally, P_o is the observed agreement between the predicted and actual classifications, while P_e is the expected agreement by chance. By employing the Kappa score, one can encapsulate the outcomes derived from both observed agreement and chance agreement, thereby arriving at a more balanced and comprehensive metric for evaluating the effectiveness of segmentation.

(2) Mean Intersection over Union
Among the prevalent metrics employed in the area of image segmentation, the Mean Intersection over Union (MIOU) stands as a prominent measure[16].

$$IOU = \frac{|A \cap B|}{|A \cup B|}$$

$$MIOU = \frac{1}{K} \sum_{c \in \mathbb{C}} IOU_c$$

Within the given equation, K represents the total number of distinct classes contained within the set $\mathbb{C}$, while IOU_c refers to the Intersection over Union for the specific class c.

B. Ablation Study and Analysis of Results for the Model

To assess the influence of the enhancements introduced in this study on the overall performance of the model, ablation experiments were conducted on three integral components: the asymmetric multi-scale convolution module, the attention mechanism module, and the temporal module. The results of these ablation studies, including Precision, Recall, Kappa Score, and Mean Intersection over Union (MIOU) for each modified model, are detailed in Table 1. Within this context, 'DI' is indicative of the asymmetric multi-scale convolution, while 'A' symbolizes the attention mechanism.

An examination of Table 1 reveals a 4.7% enhancement in precision and a 2.3% increase in MIOU for the U-Net model equipped with an attention mechanism. This demonstrates that the incorporation of the attention mechanism can augment the U-Net's capability to precisely segment crucial targets within an image, while meanwhile minimizing wrong segmentation within background regions. The U-Net model employing asymmetric multi-scale convolution exhibits a 2.1% improvement in the Kappa score (from 81.2 to 83.0) and a 1.4% rise in MIOU, signifying that the integration of the asymmetric multi-scale convolution module facilitates the model's capacity to detect features across multiple scales in the image, thus enhancing segmentation accuracy and robustness. Moreover, the hybrid U-Net model, which blends asymmetric multi-scale convolution with the attention mechanism, has shown enhancements in all evaluated metrics, illustrating that the synergistic optimization of these two aspects significantly contributes to the model's augmented performance in image segmentation tasks.

Aiming at the temporal information in sea fog images, the novel DAU-Net+ConvLSTM model, integrating ConvLSTM and DAU-Net, has manifested marked improvements across all four assessment indicators. Compared to the original U-Net model, the Kappa score and MIOU increased by 9.2% (from 81.2 to 90.4) and 5.1% respectively; compared to the non-temporally sensitive DAU-Net, these gains were 2.3% and 1.5% respectively. This illustrates that combining ConvLSTM and DAU-Net not only assimilates temporal information but also improves the capacity for sea fog segmentation. This boost stems from the DAU-Net model's superior extraction of sea fog features coupled with ConvLSTM's efficient exploitation of the time-series information embedded within the sea fog images. Such a fusion allows for a more precise partition of the sea fog region, enhancing both the accuracy and reliability of sea fog segmentation. Figure 4 illuminates the sea fog segmentation results derived from testing a sea fog image using the above five disparate models. The superior segmentation efficacy of the proposed DAU-Net+ConvLSTM model is apparent.

C. Comparative Experiments and Results Analysis of Different Models

To assess the effectiveness of the DAU-Net+ConvLSTM model, we performed comparative experiments with seven leading image segmentation models on a sea fog dataset, using five evaluation metrics. The models were divided into two groups: traditional image segmentation models (CNN, U-Net++, DeepLab v3, R2U-Net) and models incorporating temporal data for image sequence segmentation (RNN, STGRU, 3D CNN-LSTM). The first group was trained on individual images, while the second group processed daily image sequences. Results comparing DAU-Net+ConvLSTM with the first and second groups are detailed in Tables 2 and 3, respectively.

Table 2 demonstrates the superior performance of the DAU-Net+ConvLSTM model in sea fog segmentation. It exceeds the CNN model by 16.9% in Kappa score (from 73.5 to 90.4) and 12.6% in MIOU, and surpasses the highly effective R2U-Net by 4.8% in Kappa score (from 85.6 to 90.4) and 3.3% in MIOU[17]. These results emphasize the benefits of incorporating temporal information for improved segmentation accuracy.

TABLE I
ABLATION STUDY RESULTS.

Methods	Precision	Recall	Kappa	MIOU	Accuracy
U-Net	84.5%	85.6 %	81.2%	81.3%	85.2%
U-Net+A	89.2%	88.3%	83.1%	83.6%	89.1%
U-Net+DI	87.4%	86.8%	83.9%	85.7%	86.4%
DAU-Net	90.6%	89.6%	88.1%	85.9%	87.1%
DAU-Net+ConvLSTM	92.5%	91.6%	90.4%	86.4%	87.5%

TABLE II
THE EXPERIMENTAL RESULTS OF THE COMPARISON WITH THE FIRST CATEGORY OF MODELS.

Methods	Precision	Recall	Kappa	MIOU	Accuracy
CNN	78.6%	79.3 %	73.5%	73.8%	77.6%
U-Net++	86.5%	85.1%	81.8%	82.6%	80.3%
Deeplab v3	82.7%	83.4%	80.5%	80.5%	86.4%
R2U-Net	85.5%	86.7%	85.6%	83.1%	87.9%
DAU-Net+ConvLSTM	92.5%	91.6%	90.4%	86.4%	87.5%

TABLE III
PRESENTS THE EXPERIMENTAL RESULTS OF THE COMPARISON WITH THE SECOND CATEGORY OF MODELS.

Methods	Precision	Recall	Kappa	MIOU	Accuracy
RNN	79.1%	79.8 %	78.4%	75.1%	78.4%
STGRU	87.2%	86.9%	83.5%	83.6%	82.6%
3D CNN-LSTM	88.6%	89.8%	85.2%	85.3%	86.4%
DAU-Net+ConvLSTM	92.5%	91.6%	90.4%	86.4%	87.5%

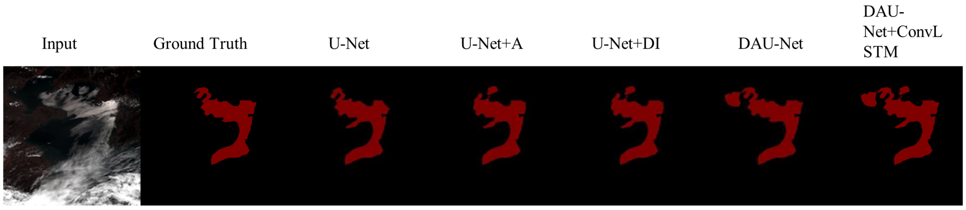

Fig. 4. Comparison of Segmentation Results in the Ablation Study.

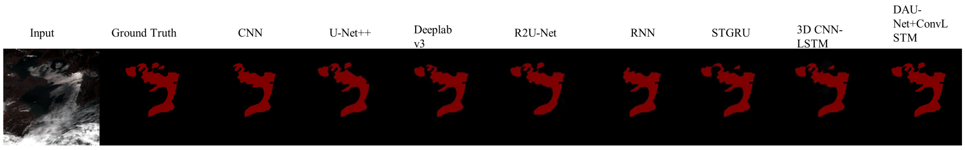

Fig. 5. Comparison of the testing segmentation results on the same hazy image by different models

A comparison of the evaluation metrics in Table 3 with those in Table 2 reveals that on the sea fog dataset, image sequence segmentation models display better performance than traditional single-image segmentation models. This improvement is attributed to the ability to utilize the relationships between individual images in a sequence, effectively capturing the temporal information of sea fog, thereby enhancing the model's accuracy. Therefore, employing the temporal information of sea fog images is essential and can provide more effective results for sea fog detection.

Compared to the STGRU[18] and 3D CNN-LSTM[19], which also effectively utilize the temporal information of sea fog images, the DAU-Net+ConvLSTM model proposed in this paper exhibits a higher Kappa score by 6.9% (from 83.5 to 90.4) and 5.2% (from 85.2 to 90.4), and MIOU by 2.8% and 1.1%, respectively. This indicates that the DAU-Net+ConvLSTM model delivers superior sea fog segmentation results when temporal information is similarly considered, fully validating the effectiveness of integrating asymmetric multi-scale convolution and attention mechanisms for improving the U-Net. Figure 5 shows the sea fog segmentation effects of the proposed model compared to the other seven models using a test sea fog image. The results visibly demonstrate that the DAU-Net+ConvLSTM model proposed in this paper achieves better segmentation performance.

V. CONCLUSIONS

Drawing on the unique characteristics of sea fog, including many scale variations, diverse texture patterns, and temporal dynamics, this study conducts research on a deep learning method for sea fog segmentation. We propose a DAU-Net model to extract features across variable scales and hierarchies by employing asymmetric multi-scale convolution modules and attention mechanisms into the backbone U-Net model. Then, the ConvLSTM is integrated into the DAU-Net model to leverage spatio-temporal information of sea fog for obtaining more accurate detection and prediction results of sea fog phenomena. Our proposed DAU-Net-ConvLSTM model demonstrates improved performance in sea fog segmentation compared to other deep learning methods, with a 90.4% Kappa score and 86.4% MIOU. In the future, the study needs to be further perfected from the expansion of the dataset scale, the enhancement of the generalization capability of the model, the research on sea fog detection for nighttime or all-weather, and the improved design for feature extraction and segmentation of sea fog.

REFERENCES

[1] C. Liu, H. Dai, S. Wang, and J. Chen, “Remote sensing image scene classification based on multidimensional attention and feature enhancement.” *IAENG International Journal of Computer Science*, Vol. 50, No. 4, pp. 1337–1346, 2023.

[2] H.-K. Jeon, S. Kim, J. Edwin, and C.-S. Yang, “Sea fog identification from goci images using cnn transfer learning models,” *Electronics*, Vol. 9, No. 2, pp. 311–319, 2020.

[3] Y. Tang, P. Yang, Z. Zhou, and X. Zhao, “Daytime sea fog detection based on a two-stage neural network,” *Remote Sensing*, Vol. 14, No. 21, pp. 5570–5583, 2022.

[4] Y. Ran, H. Ma, Z. Liu, X. Wu, Y. Li, and H. Feng, “Satellite fog detection at dawn and dusk based on the deep learning algorithm under terrain-restriction,” *Remote Sensing*, Vol. 14, No. 17, pp. 4328–4344, 2022.

[5] Z. Chunyang, W. Jianhua, L. Shanwei, S. Hui, and X. yanfang, “Sea fog detection using u-net deep learning model based on modis data,” 2019, pp. 1–5.

[6] W. D. CHEN Biao, “Arctic sea fog detection using caliop and modis,” *Journal of Atmospheric and Environmental Optics*, Vol. 17, No. 2, pp. 267–277, 2022.

[7] Z. Li, X. Zhang, and B. Zhang, “Segmentation of pulmonary nodules based on mrbu-net-wd model,” *IAENG International Journal of Computer Science*, Vol. 50, No. 2, pp. 673–682, 2023.

[8] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, “Slic superpixels compared to state-of-the-art superpixel methods,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 34, No. 11, pp. 2274–2282, 2012.

[9] A. Radford, L. Metz, and S. Chintala, “Unsupervised representation learning with deep convolutional generative adversarial networks,” arXiv 2016 arXiv:1511.06434.

[10] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[11] X. Wang, R. Zhang, C. Shen, T. Kong, and L. Li, “Dense contrastive learning for self-supervised visual pre-training,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2021, pp. 3024–3033.

[12] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, “Cbam: Convolutional block attention module,” in *Proceedings of the European Conference on Computer Vision (ECCV)*, September 2018, pp. 3–19.

[13] K. Kawakami, “Supervised sequence labelling with recurrent neural networks,” Ph.D. dissertation, Technical University of Munich, 2008.

[14] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” arXiv 2017 arXiv:1412.6980.

[15] X. Li, X. Sun, Y. Meng, J. Liang, F. Wu, and J. Li, “Dice loss for data-imbalanced NLP tasks.” Association for Computational Linguistics, Jul. 2020, pp. 465–476.

[16] L. Chen, G. Papandreou, F. Schroff, and H. Adam, “Rethinking atrous convolution for semantic image segmentation,” *CoRR*, arXiv 2017 arXiv:1706.05587.

[17] M. Z. Alom, M. Hasan, C. Yakopcic, T. M. Taha, and V. K. Asari, “Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation,” arXiv 2018 arXiv:1802.06955.

[18] D. Nilsson and C. Sminchisescu, “Semantic video segmentation by gated recurrent flow propagation,” in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 6819–6828.

[19] T. Akilan, Q. J. Wu, A. Safaei, J. Huo, and Y. Yang, “A 3d cnn-lstm-based image-to-image foreground segmentation,” *IEEE Transactions on Intelligent Transportation Systems*, Vol. 21, No. 3, pp. 959–971, 2020.