Research on the Extraction Method of Wetland River Network from Remote Sensing Image Based on MobileNetV2

Zhenyu Wu [1, a], Jingyi Zhang [1, b], Shanshan Hong [2, c]

[1] School of Control Science and Engineering, Dalian University of Technology, Dalian, Liaoning Province, China;

[2] School of Construction Engineering, Dalian University of Technology, Dalian, Liaoning Province, China.

[a] zhenyuwu@dlut.edu.cn, [b] zhangjingyi@9876@163.com, [c] hongshanshan0727@163.com

Abstract. Estuarine wetlands serve as transitional zones between terrestrial river ecosystems and marine ecosystems, playing a significant role in ecological terms. Addressing the challenges in river network extraction from remote sensing images of estuarine wetlands, such as diverse target scales, high detail proportion, and substantial model computational load, this study proposes a multi-scale lightweight river network extraction method based on MobileNetV2. The method employs the MobileNetV2 backbone network to extract local features, incorporates a spectral attention mechanism to enhance the spectral response characteristics of water bodies, thereby improving the model's adaptability to complex spectral information. It also combines lightweight Depthwise Separable Atrous Spatial Pyramid Pooling to capture multi-scale contextual information, significantly reducing computational complexity while ensuring model accuracy. Finally, the method optimizes the feature fusion and upsampling process through a progressive feature fusion decoder, further enhancing the accuracy and detail restoration capability of river network segmentation. Experiments were conducted on a mixed dataset constructed from Sentinel-2 multispectral images and publicly available datasets. The results show that the method achieves a mean Intersection over Union (mIoU) of 94.5%, which is an improvement of 1.5% and 0.8% over DeepLabV3+ and DeepWaterMapV2, respectively. This method provides efficient and reliable technical support for ecological monitoring of estuarine wetlands and analysis of river network morphological evolution.

Keywords: Remote sensing; MobileNetV2; Sentinel-2; River network.

1. Introduction

Estuarine wetlands, serving as transitional zones between terrestrial river ecosystems and marine ecosystems, play a significant role in carbon-nitrogen biogeochemical cycles, biodiversity maintenance, global climate change mitigation, and blue carbon sequestration, among other aspects [1]. Numerous studies have shown a close connection between the geometric characteristics of river networks, such as density and curvature, and the hydrological connectivity of wetlands. Therefore, quantifying the hydrological connectivity of estuarine wetlands based on the morphological characteristics of river networks is one of the common methods currently in use, with the accurate extraction of river network morphological features being a crucial prerequisite for the application of this method.

Due to the unique geomorphological features of estuarine wetlands, traditional methods of collecting elevation geographic data using LiDAR DEM are not efficient for extracting topographic features of large-scale wetlands [2]. Moreover, they are time-constrained and cannot extract historical data. With the advancement of remote sensing technology, the extraction of river network morphology using optical remote sensing imagery has significantly progressed. Traditional methods mainly rely on the spectral characteristics of remote sensing images, index calculations, threshold segmentation, and manually designed rules, such as the NDWI index (Normalized Difference Water Index), OTSU (Otsu's method), and object-based methods (OBIA), among others [3-5]. In recent years, there has been further development in the use of data-driven algorithms like deep learning, enabling automated water body extraction. Deep learning technology, through the construction of end-to-end semantic segmentation models (such as U-Net, DeepLab) and multi-source data fusion

frameworks, has achieved high-precision intelligent extraction of water bodies from pixel to object level [6-7]. By autonomously learning spectral response differences, spatial texture features, and global contextual associations, these models have significantly improved generalization capabilities in complex scenarios, making breakthroughs in suppressing shadow interference, identifying fragmented water bodies, and continuous river networks.

Inspired by the aforementioned work and the needs of related research, we propose a multi-scale lightweight river network extraction method based on MobileNetV2. The contributions of this paper can be summarized as follows:

(1) A Spectral Channel Attention Module (SCAM) is introduced to enhance the spectral response characteristics of water bodies.

(2) Lightweight Depthwise Separable Atrous Spatial Pyramid Pooling (DASPP) is incorporated to capture multi-scale contextual information, improving the ability to extract features at different scales while reducing computational complexity without compromising model accuracy.

(3) The Progressive Feature Fusion Decoder (PFFD) is employed to optimize the feature fusion and upsampling process, further enhancing the accuracy and detail restoration capability of river network segmentation.

2. MobileNetV2 Network

MobileNetV2, introduced by the Google team in 2018, is a lightweight convolutional neural network. Its core structure comprises Inverted Residual Blocks and Linear Bottleneck layers, which significantly reduce the number of parameters and computational load through Depthwise Separable Convolution. This design maintains high accuracy while markedly improving the model's operational efficiency. The specific architecture of the network is illustrated in Figure 1.

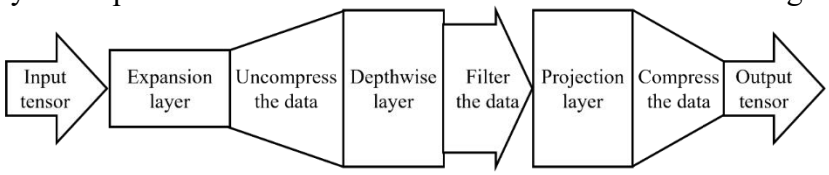

Fig.1 Network Structure of MobileNetV2

MobileNetV2 is a lightweight convolutional neural network whose core structure consists of multiple Inverted Residual Blocks, focusing on efficient feature extraction. The input image is first normalized and resized before entering the network. In the backbone network, each Inverted Residual Block expands the channel dimensions using a 1x1 convolution, then decomposes the computation into Depthwise Convolution and Pointwise Convolution through Depthwise Separable Convolution, significantly reducing computational costs. A Linear Bottleneck layer is applied at the output to prevent information loss in low-dimensional feature spaces. The network progressively extracts shallow (high-resolution details), intermediate (balanced semantics and resolution), and deep (high-semantic global information) features through staged Inverted Residual Blocks. These features can be directly used for downstream tasks or further enhanced by lightweight modules to fuse multi-scale information. The highlight of MobileNetV2 lies in the synergistic design of Inverted Residual Structures and Depthwise Separable Convolution, which achieves high accuracy while drastically reducing the number of parameters and computational complexity.

3. MobileNetV2 Network

3.1 Spectral Channel Attention Module

In this study, to address the task of river network extraction from remote sensing images, we adapted the input layer and feature extraction mechanism of MobileNetV2. To fully utilize the spectral information of multispectral satellite imagery, we expanded the original 3-channel RGB

input to a 6-channel multispectral input (RGB, NIR, and SWIR) to capture the reflectance characteristics of water bodies across multiple bands. To enhance the spectral response characteristics of water bodies, we introduced the Spectral Channel Attention Module (SCAM) [8] at the input layer. By incorporating the Normalized Difference Vegetation Index as prior knowledge, SCAM dynamically adjusts the contribution of each spectral channel using channel attention weights, emphasizing bands highly correlated with water bodies (e.g., near-infrared) while suppressing interference from vegetation-covered areas. Additionally, we adjusted the number of channels in the first convolutional layer of MobileNetV2 (from 3 to 6). The 6-channel feature tensor, refined by SCAM, serves as the input to the Inverted Residual Blocks, which further extract spatial contextual information. This synergy enhances the discriminative ability of the model for multi-scale water bodies. SCAM introduces only a small number of parameters, ensuring overall computational efficiency.

3.2 Lightweight Depthwise Separable Atrous Spatial Pyramid Pooling

Atrous Spatial Pyramid Pooling (ASPP) is a multi-scale feature extraction module initially proposed in the DeepLab series for semantic segmentation. Its core idea is to capture contextual information at multiple scales using Atrous Convolution, thereby improving the model's adaptability to variations in target size [9]. For large-scale remote sensing images, the computational load is significant. To maintain the model's lightweight nature, we improved it to DASPP, where standard convolutions are replaced with depthwise separable convolutions. The global pooling branch restores spatial resolution through bilinear interpolation, and atrous convolutions with different dilation rates are selected to cover varying receptive field sizes. Finally, the features are fused, and the output feature map is generated, reducing computational cost and parameter count while preserving multi-scale feature extraction capabilities.

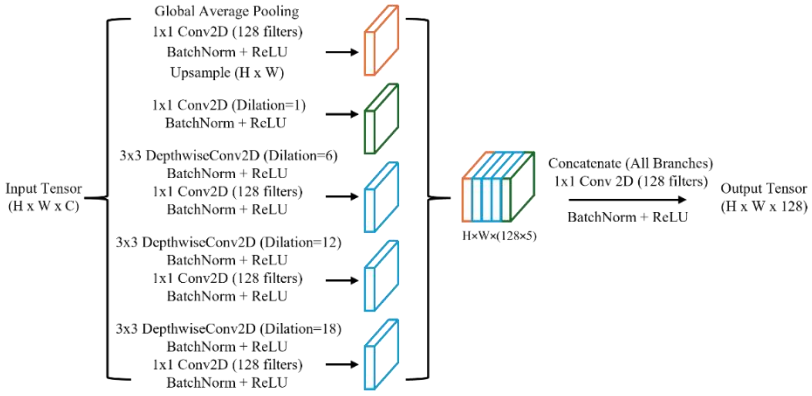

Fig. 2 DASPP Module

3.3 Progressive Feature Fusion Decoder

After extracting deep, intermediate, and shallow features using MobileNetV2, a progressive feature fusion strategy[10] is employed to fuse multi-level features. The first fusion step involves extracting multi-scale contextual information from deep features using the DASPP module, followed by concatenating these features with intermediate features along the channel dimension to produce a 1/4 resolution image. The concatenated features are then fused and upsampled to 1/2 resolution. The second fusion step concatenates the 1/2 resolution features obtained from the first fusion with the projected output of shallow features at 1/2 resolution along the channel dimension. A transposed convolution is applied to upsample the features to the original resolution. Finally, the number of channels is compressed to 2 to generate the segmentation probability map.

3.4 Overall Model Structure

During the final training phase of the model, this study addresses a binary classification problem. We selected the Adam optimizer and binary cross-entropy loss to measure the model's loss. Through experimentation, the initial learning rate was set to 0.005, and the batch size was set to 4. Data augmentation techniques such as mirroring, horizontal flipping, and vertical flipping were applied during the training process. The overall structure of the model is illustrated in the accompanying figure 3.

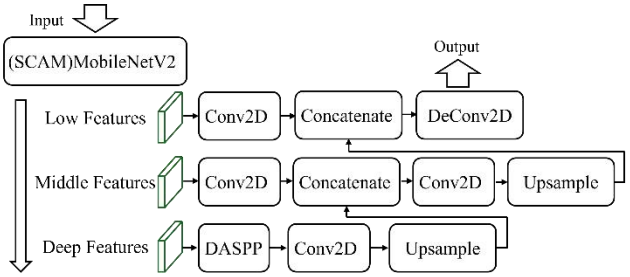

Fig. 3 Model Structure Diagram

4. Experiments and Analysis

4.1 Dataset and Experimental Environment

To evaluate the effectiveness of the proposed algorithm and enhance the extraction performance for the study area, we selected images from the Earth Surface Water Knowledge Base (ESWKB) built using Sentinel-2 imagery, which includes natural areas such as estuarine wetlands [11]. Additionally, manually digitized water body images of the study area from three historical years were cropped and mixed with the aforementioned images to form the training dataset. The relevant remote sensing data can be freely obtained from the official website of the European Space Agency.

The experiments were conducted using the TensorFlow deep learning framework. The model was deployed on a server equipped with multiple NVIDIA GeForce RTX 2080 Ti GPUs for training. The epoch was set to 200, the batch size was set to 4, and the initial learning rate was set to 0.005. The dataset images were cropped to 512×512 pixels for further processing.

4.2 Evaluation Metrics

We adopted the overall accuracy on the validation set and the mean Intersection over Union on the validation set as evaluation metrics to comprehensively assess the model's performance. ValOA reflects the proportion of correctly classified pixels in the validation set, measuring classification quality, while valMIoU focuses more on the model's prediction accuracy for river network boundaries by calculating the average Intersection over Union. MIoU is calculated as follows, where p and g represent the predicted and ground truth regions, respectively, and 0 and 1 denote non-target and target categories, respectively.

$$MIoU = \frac{1}{2}\sum_{i=0}^{1} \frac{p_i \cap g_i}{p_i \cup g_i} \tag{1}$$

4.3 Ablation Study

To verify the effectiveness of the main components of our proposed model, we designed an ablation study. The specific experimental results are shown in Table 1, where "Base" represents the original MobileNetV2 model, and " √ " indicates the added modules. To adapt to remote sensing imagery, we made appropriate adjustments to the original MobileNetV2 to ensure testing on the same dataset.

Table 1: Results of Ablation Study

Base	SCAM	DSAPP	PFFD	valMIoU
√				0.924
√	√			0.929
√		√		0.935
√			√	0.928
√	√	√		0.939
√	√		√	0.932
√		√	√	0.942
√	√	√	√	0.945

From the results in Table 1, it can be observed that when only the SCAM module was added, the MIoU improved by 0.5%; when only the DASPP module was added, it improved by 1.1%; and when only the PFFD module was added, it improved by 0.4%. When all three modules were added, the MIoU improved by 2.1%, indicating that each of the proposed modules contributed to varying degrees of improvement in MIoU, leading to a significant enhancement in the overall segmentation accuracy of the model.

4.4 Comparative Testing

We compared our method with traditional approaches such as NDWI and deep learning methods including DeepLabV3+ and DeepWaterMapV2 [12].

(a) Ours (b) NDWI +C lustering of means

Fig.4 Comparison of Results

Extraction tests were conducted on remote sensing images at different scales, and the results, as shown in Figure 5, met expectations.

Satellite image Our extraction results Targeted areas intercepted from our programs

Fig.5 Extraction Results

To ensure consistency in the experiments, we made appropriate modifications to DeepLabV3+ and DeepWaterMapV2 to match the current dataset for training and testing. All models were trained for 200 epochs.

Table 2: Comparative Test Results

	Ours	DeepLabV3+	DeepWaterMap
Miou	0.945	0.930	0.937

Under the same dataset and training epochs, our algorithm achieved a higher MIoU compared to other algorithms, demonstrating that the improved algorithm provides more accurate extraction of multi-scale river network targets from remote sensing images. Additionally, in terms of the number of parameters, DeepLabV3+ has 4.5M, DeepWaterMapV2 has 4.0M, and our model has 3.4M. Our model also outperforms the other two models in training time.

5. Conclusion

Aiming at the scene of extracting multi-scale river network in estuarine wetland from remote sensing image, the river network scale range is large and easy to be disturbed by vegetation and other problems, this paper proposes a multi-scale river network extraction model based on MobileNetV2. By incorporating the improved SCAM module, combined with DASPP and a progressive feature fusion decoder, the extraction performance is optimized. Both ablation experiments and controlled experiments demonstrate that our proposed algorithm outperforms the original MobileNetV2 algorithm and other typical target-related models, achieving superior results in segmenting multi-scale river network features from optical remote sensing images. In future work, we plan to further enhance the model's generalization capability to make it applicable to more diverse topographic regions.

References

[1] Elizabeth Mcleod, Gail L Chmura, Steven Bouillon, Rodney Salm, Mats Björk, Carlos M Duarte, Catherine E Lovelock, William H Schlesinger, Brian R Silliman. A blueprint for blue carbon: toward an improved understanding of the role of vegetated coastal habitats in sequestering CO2. Frontiers in Ecology and the Environment, 2011, 9(10): 552–560

[2] Jessica E. Chassereau, Joseph M. Bell, Raymond Torres. A comparison of GPS and lidar salt marsh DEMs. Earth Surface Processes and Landforms, 2011, 36(13): 1770–1775

[3] S. K. McFEETERS. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. International Journal of Remote Sensing, 1996, 17(7): 1425–1432

[4] Jianbo Tan, Yi Tang, Bin Liu, Guang Zhao, Yu Mu, Mingjiang Sun, Bo Wang. A Self-Adaptive Thresholding Approach for Automatic Water Extraction Using Sentinel-1 SAR Imagery Based on OTSU Algorithm and Distance Block. Remote Sensing, 2023, 15(10): 2690

[5] Kevan B. Moffett, Steven M. Gorelick. Distinguishing wetland vegetation and channel features with object-based image segmentation. International Journal of Remote Sensing, 2013, 34(4): 1332–1354

[6] Bo Liu, Shihong Du, Lubin Bai, Song Ouyang, Haoyu Wang, Xiuyuan Zhang. Water extraction from optical high-resolution remote sensing imagery: a multi-scale feature extraction network with contrastive learning. GIScience & Remote Sensing, 2023, 60(1): 2166396

[7] Anusha Ch, Rupa Ch, Samhitha Gadamsetty, Celestine Iwendi, Thippa Reddy Gadekallu, Imed Ben Dhaou. ECDSA-Based Water Bodies Prediction from Satellite Images with UNet. Water, 2022, 14(14): 2234

[8] Chuanhui Shan, Xinlong Geng, Chao Han. Remote sensing image road network detection based on channel attention mechanism. Heliyon, 2024, 10(18): e37470

[9] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, Hartwig Adam. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. Computer Vision – ECCV 2018, 2018, 11211: 833–851]

[10] Chongyang Zhang, Bin Wang. Progressive Feature Fusion Framework Based on Graph Convolutional Network for Remote Sensing Scene Classification. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2024, 17: 3270–3284

[11] Xin Luo, Xiaohua Tong, Zhongwen Hu. An applicable and automatic method for earth surface water mapping based on multispectral images. International Journal of Applied Earth Observation and Geoinformation, 2021, 103: 102472

[12] Leo F. Isikdogan, Alan Bovik, Paola Passalacqua. Seeing Through the Clouds With DeepWaterMap. IEEE Geoscience and Remote Sensing Letters, 2020, 17(10): 1662–1666