Water-Body Segmentation from Remote Sensing Satellite Images Utilizing Hierarchical and Contour-Based Multi-Scale Features

R. S. M. Lakshmi Patibandla, Adusumalli Yaswanth, and Syed Inamulla Hussani

Abstract Satellite image photography with very high resolution (VHR) presents a significant problem in identifying water bodies. In this work, correlations between extracted features at each scale, which extract the whole target. Using data from several sources, including the immediate environment, a broader geographic area, and the relationships that exist between the various channels, display features. In addition, to better anticipate water bodies' delicate contours, use Fusion of many scales of prediction. In addition, feature semantic inconsistency is resolved. Encoder-decoder semantic fusion allows us to combine the encoding and decoding processes module for promoting the fusion of features. The outcome demonstrates that our approach is cutting-edge superior performance in the segmentation process compared to the most contemporary and traditional approaches. In addition, have offered methods that are reliable even when used in the most difficult water body extraction situations.

Keywords Satellite image · Contour-based · Multiscale features · Segmentation

1 Introduction

In monitoring water supplies, natural water bodies are critical. There is a need for catastrophe prediction and nature conservation. They are dependent on the measurement of the change in a water body. Recognizing the water body in detail A critical task is to be able to monitor changes in water bodies through remote sensing images. The objective of this research is to accurately discover waterbodies in strenuous and complex environments. High-resolution remote sensing footage was used to create scenarios [1–3]. The board has a variety of instruments. Remote sensing photography from satellites and airborne vehicles covers large-scale water areas. Sensing pictures might be difficult to interpret. Aquatic organisms are often to blame for such degradations [4]. The bank is blocked by vegetation, silts, and boats, as well as shadows

R. S. M. Lakshmi Patibandla (✉) · A. Yaswanth · S. I. Hussani
Department of IT, Vignan's Foundation for Science Technology and Research, Vadlamudi, Andra Pradesh, India
e-mail: patibandala.lakshmi@gmail.com

cast by the surrounding tall tree plants. Imagery conditions and water quality may all play a role in producing these unique hues and microbes are involved [5, 6]. Consequently, obtaining the shape of aquatic bodies is a major difficulty (Fig. 1).

VHR remote sensing imaging may be used correctly in complicated settings. Existing remote sensing image extraction approaches concentrate on the spectral features of each band and manually constructed algorithms to extract water bodies methodologies, such as band cutoff point methods, supervised classification-based methods, water and vegetation indices-based methods, and spectral interaction ways the techniques [7]. These approaches, on the other hand, do not pay a lot of attention to the geographic information (i.e., shape, size, texture, edge, shadow, and context

(a)

(b)
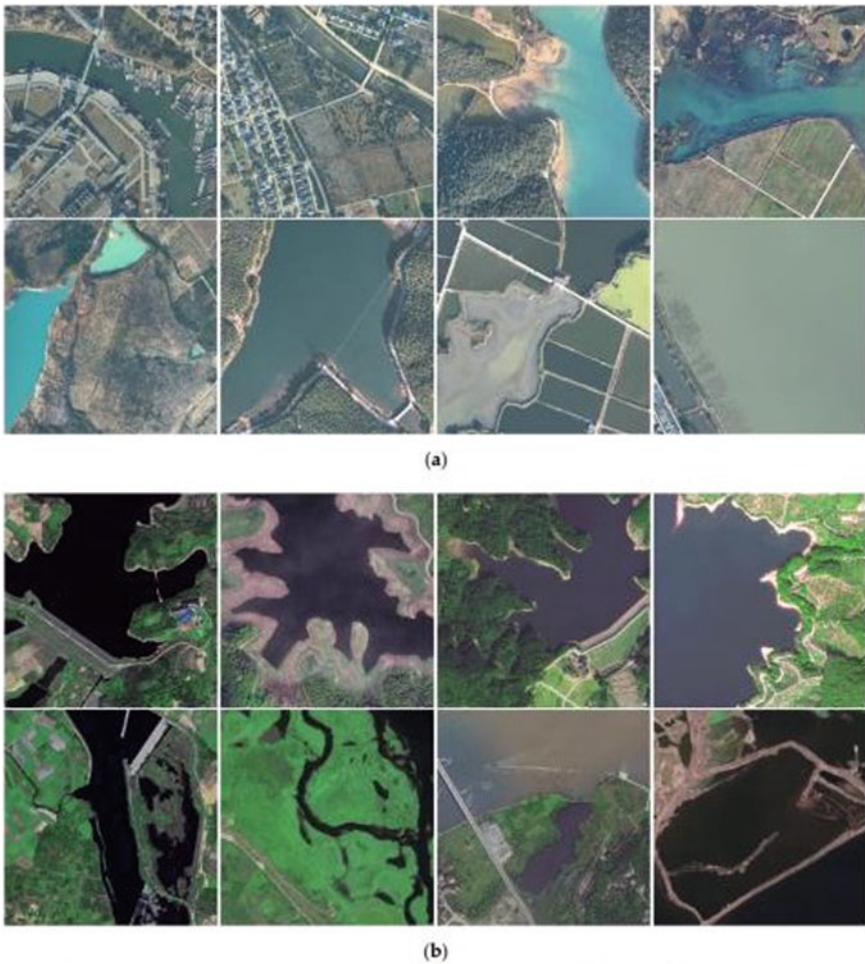
Fig. 1 Some typical water-body samples **a** in VHR aerial images and **b** Gaofen2 (GF2) satellite images

semantics) of the water bodies, which has a substantial impact on classification accuracy. The scarcity of automation in traditional approaches is also a barrier to large-scale remote sensing visuals. The tremendous convolutional capabilities of convolutional neural networks (CNNs) can indeed be attributed to image classification, target recognition, and semantic segmentation [8–13]. Long et al. [8] pioneered the thoroughly convolutional network (FCN), which replaces the last fully connected layers with convolutional ones for entire semantic segmentation. End-to-end FCNs are broadly utilized and well-developed in the realm of semantic segmentation, making them a mainstream technology.

Deep learning-based water-body segmentation using remote sensing images has triggered a lot of interest recently. The FCN-based method's feature fusion combines high-semantic features and features with exact locations, making it easier to identify waterbodies and extract waterbody borders with precision. Three parts of our technique are considered: feature extraction, prediction optimization, and the merging of shallow and deep layers.

2 Methodology

To begin, we'll go through our ideas for a MECNet architecture. A multi-feature extraction and combination (MEC) module is then described in order to get more diversified and richer features as well as enhanced semantics. This is why, to better anticipate the water-fine body's contour, we create an MPF module that combines prediction results from three separate levels. Once we've solved the issue of semantic inconsistency between encoding and decoder, propose an encoder- decoder feature fusion module (DSFF).

## 2.1 MECNet's Underlying Network Architecture

The MECNet is made up of three primary components. A first multi-feature extraction and combination modules are then built, which provides a more diversified set of encoded characteristics. Three alternative feature extraction sub-modules are suggested for the MEC module to simulate the spatial and channel interactions between feature maps in the proposed MEC module. Local feature extraction, bigger receptive field feature extraction, and between-channel feature extraction are the three sub-modules that make up this system. An encoder-decoder semantic feature fusion module is built to resolve the semantic discrepancy of features from the encoding stage and the decoding stage. Water-body segmentation contours may be generated using a simple multi-scale prediction fusion module that takes input from three distinct scales. The mask that offers a binary label to each pixel in our attention-guided, multi-scale image is derived from this input tensor. The encoder-decoder architecture of the proposed MECNet [9] is portrayed in Figs. 2, 3 and 4.
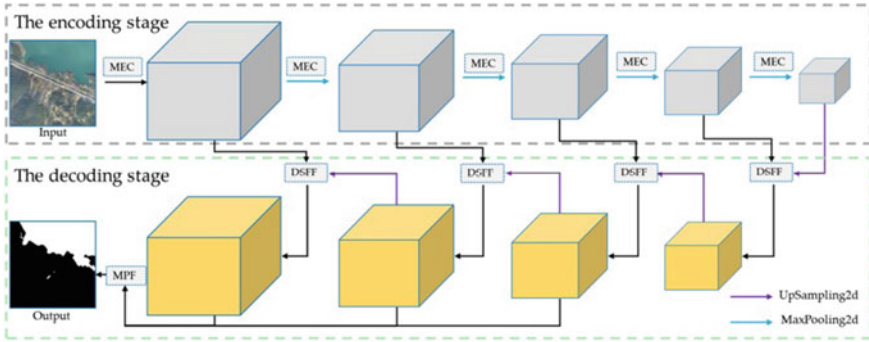
Fig. 2 An overview of our proposed Multi-feature Extraction and Combination Network (MECNet).MECNet has three modules: Multi-feature Extraction and Combination (MEC), Encoder and Decoder Semantic Feature Fusion (DSFF),and Multi-scale Prediction Fusion (MPF)
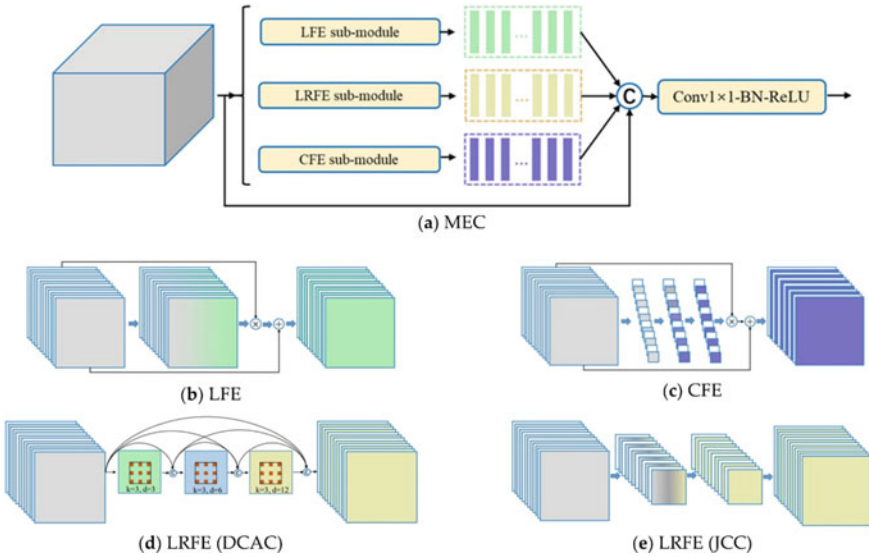
Fig. 3 The details of multi-feature extraction and combination module. **a** The Multi-feature Extraction and Combination (MEC) module consists of **b** a Local Feature Extraction (LFE) sub-module, **c** a between-channel feature enhancement module (CFE) and a longer receptive-field feature extraction sub-module (LRFE); **d** Densely Connected Atrous Convolutions (DCAC), and **e** JCC (Joint Conv7-S4-Conv3-S1, for the longer receptive field feature extraction.to properly forecast the water-body segmentation map, whereas DSFF combines distinct information from the encoding and decoding phases at the same scale
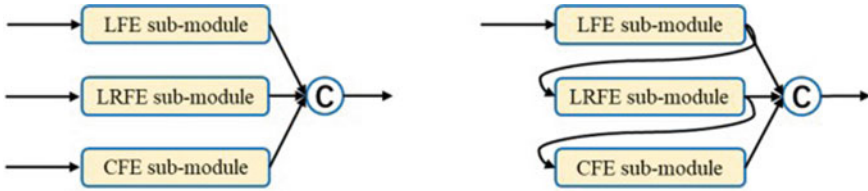
Fig. 4 Two ways to combine different feature sub-modules in the MEC (Multi-feature Extraction and Combination), Left: in a parallel way. Right: in a cascade way

Two methods in the MEC for combining various feature sub-modules (Multi-feature Extraction and Combination), Right: in a similar fashion. In a cascading fashion, that's correct.

2.2 Semantic Features Fusion Module for Encoder-Decoder

The DSFF module (Fig. 6) extends the 3D channel attention module described in our earlier work [28] to overcome the issue of semantic inconsistency in feature fusion at the decoding stage. To minimize the number of channels in the concatenated feature maps at the same scale from both the encoding and decoding stages, the DSFF first conducts a 1 1 convolution using BN and ReLU. The concatenated features are then used to construct the global context, which is then used to do 1 1 convolutions using BN, ReLU, and a Sigmoid function. As a guide for combining various semantic characteristics, it automatically learns how to link the channels of feature maps together semantically. The concatenated characteristics multiply and add the global context information. To finish, 3×3 convolutions with BNs and a ReLU are applied to the feature maps that were generated. To accomplish an effective fusion of distinct semantic features, the DSFF module is used on various scale characteristics at the decoding step. 2021, 03, ×8 of 19 Remote Sensing Multi-scale Prediction Fusion (MPF) is seen in Fig. 5.
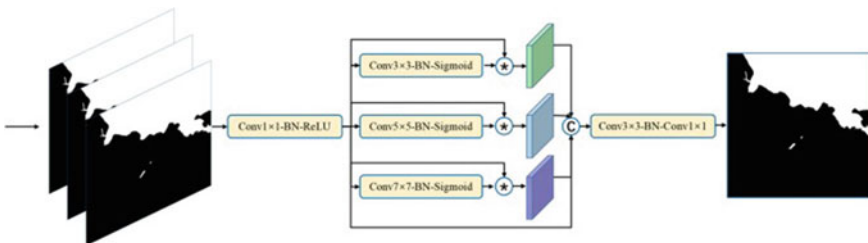
Fig. 5 MPF: Multi-scale prediction fusion module
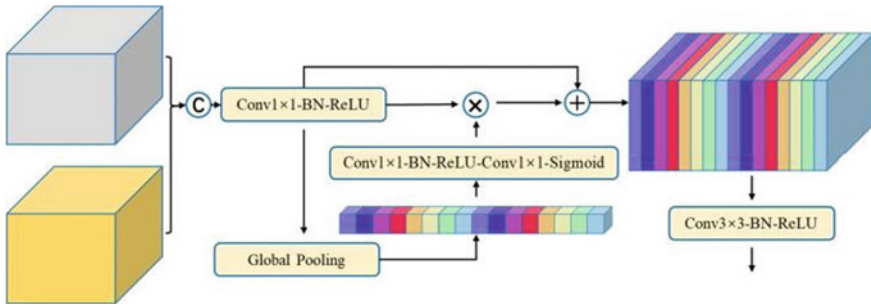
Fig. 6 Different semantic feature fusion module, DSFF

The Total Loss Function (TLF) The difficulty of training deep neural networks grows as the network's depth increases [20]. We implement a simple and effective output layer at each scale in the decoding step and apply loss restrictions between its result and the ground truth to train our proposed model more efficiently.

The total loss function and the cross-entropy function L are illustrated in the following manner. Figure 6, The DSFF module, which stands for Different Semantic Feature Fusion.

3   The Architecture of the Proposed Model

The proposed architecture mainly depends on four steps. The first is the Image Processing where all the goes through the geometric correction, i.e., all the color, texture, and shape are identified and produces the image immediately after analyzing which is knows as Image Fusion. Later, in the second step the image which is produced in the preprocessing is transformed to sample generation state in which the image is analyzed by pixel-by-pixel and forms two datasets one is the training dataset and the other is test dataset. Later, the training dataset values are compared with the test dataset. In the third process, the data gets water extraction where the image is predicted with the accurate position on the water content on the image, and the final step is the accuracy assessment where the percentage and the accuracy is evaluated and represented in the graphical format (Fig. 7).

4   Experimental Results

Extraction of characteristics at many scales, there are three main methods of feature extraction for the multi-featured integrated network: local (LFE), receptive longer features (RLFE), and channel-based feature extractions (RLFE) (CFE). In this study,
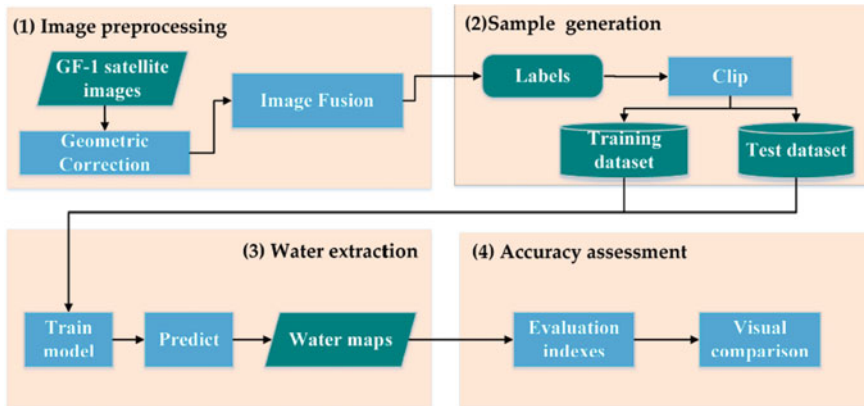
Fig. 7 Architecture of water-body segmentation

spatial feature relationships between linked features are characterized using LFE and RLFE, and the maximum features acquired across multiple channels are explored.

Optimization of the Contour Map, there are a variety of state-of-the-art methods for detecting contours in a picture that include localization information. Based on multi-scale globalization and semantic picture attributes such as texture, color, and form, we explore contour detection from a satellite image.

Multi-scale feature extraction using contours, multi-feature extraction using a contour-based approach is quite different from typical multi-scale extraction methods, and we employ the following modules to test its viability: local feature extraction (LFE), channel feature extraction, and long receptive field feature extraction (LRFFE) (CFE & LRFE). Submodules LFE and LRFE are used to identify regions with certain features, whereas CFE investigates the relationships between distinct feature maps.

Optimized water-body segmentation extraction, training is finished, and the weights of each pixel are evaluated using the proper neighbor pixel selection for each picture once the multi-feature extraction procedure is done. Raw images of the linked objects are used as input, and the probability maps derived from the multi-scale feature search approach are used to segment the water. A lot of pixels are involved in this work, hence this module includes a significant number of pixels from the picture. The most effective model for evaluating pixels with varied decoding variables is optimal water-body segmentation with multi scale-feature extraction (Figs. 8 and 9).

5　Conclusion

To enhance water-body contour identification from VHR remotely sensed photos, combining aerial and satellite pictures, we use the structure of the embedding. Our
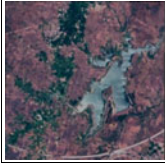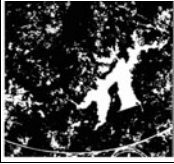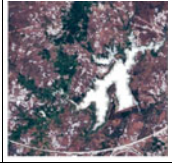
Original Image	True Mask	CNN Prediction	Accuracy	Duration
			72.355419	3.30596

Fig. 8 CNN prediction
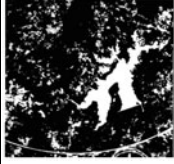
Original image	True Mask	MECNET prediction	Accuracy	Duration
			81.706900	3.0280 74

Fig. 9 MECNET prediction

approach relies on the following three components: A DSFF module solves the issue of semantic inconsistencies of features extraction between the encoding and decoding stages by automatically extracting richer and more diverse features in the encoding stage and obtaining more advanced semantic information for feature fusion in the decoding stage. On VHR aerial and satellite photos, our technique achieved the greatest accuracy as well as the best resilience under tough conditions, according to the results of our studies. In addition to feature extraction, this new design module may be used for semantic segmentation and object recognition. In this, we compared this project between CNN, MECNET, and MECNET-CMO among these we could find that CNN consumes more duration and results with the less efficiency where as MECNET consumes less duration and produces more efficiency which is proved in our project.

References

1. Miao Z, Fu K, Sun H, Sun X, Yan M (2018) Automatic water-body segmentation from high-resolution satellite images via deep networks. IEEE Geosci Remote Sens Lett 15(4):602–606. https://doi.org/10.1109/LGRS.2018.2794545
2. Li B, Zhang H, Xu F (2014) Water extraction in high resolution remote sensing image based on hierarchical spectrum and shape features. IOP Conf Ser Earth Environ Sci 17:012123. https://doi.org/10.1088/1755-1315/17/1/012123
3. Li K, Hu X, Jiang H, Shu Z, Mi Z (2020) Attention-guided multi-scale segmentation neural network for interactive extraction of region objects from high-resolution satellite imagery. Remote Sens 12:789. https://doi.org/10.3390/rs12050789

4. Patibandla RL, Narayana VL, Gopi AP, Rao BT (2021) Comparative study on analysis of medical images using deep learning techniques. In: Deep learning for biomedical applications. CRC Press, pp 329–345

5. Gautam VK, Gaurav PK, Murugan P, Annadurai M (2015) Assessment of surface water Dynamics in Bangalore using WRI, NDWI, MNDWI, supervised classification and KT transformation. Aquat Procedia 4:739–746

6. Tarakeswara Rao B, Lakshmi Patibandla RSM, Subramanyam K, Lakshman Narayana V (2021) Deep learning method for identification and selection of relevant features. Mater Today Proc. ISSN 2214-7853. https://doi.org/10.1016/j.matpr.2021.06.103

7. Zhao X, Wang P, Chen C, Jiang T, Yu Z, Guo B (2017) Waterbody information extraction from remote-sensing images after disasters based on spectral information and characteristic knowledge. Int J Remote Sens 38:1404–1422

8. Long J, Shelhamer E, Darrell T (2015) Fully convolutional networks for semantic segmentation. IEEE Trans Pattern Anal Mach Intell 39:640–651

9. Ronneberger O, Fischer P, Brox T (2015) U-net: convolutional networks for biomedical image segmentation. In: Proceedings of the international conference on medical image computing and computer-assisted intervention, Munich, Germany, 5–9 October 2015, pp234–241

10. Lin G, Milan A, Shen C, Reid I (2017) Refinenet: Multi-path refinement networks for high-resolution semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, Honolulu, HI, USA, 22–25 July 2017, pp 1925–1934

11. Yu Z, Feng C, Liu M-Y, Ramalingam S (2017) Casenet: deep category-aware semantic edge detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition, Honolulu, HI, USA, 21–26 July 2017, pp5964–5973

12. Bertasius G, Shi J, Torresani L (2015) Deepedge: A multi-scale bifurcated deep network for top–down contour detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition, Boston, MA, USA, 8–10 June 2015, pp4380–4389

13. Xie S, Tu Z (2015) Holistically-nested edge detection. In: Proceedings of the IEEE international conference on computer vision, Santiago, Chile, 13–16 December 2015, pp 1395–1403

14. Yu L, Wang Z, Tian S, Ye F, Ding J, Kong J (2017) Convolutional neural networks for water body extraction from Landsat imagery. Int J Comput Intell Appl 16:1750001

15. Miao Z, Fu K, Sun H, Sun X, Yan M (2018) Automatic water-body segmentation from high-resolution satellite images via deep networks. IEEE Geosci Remote Sens Lett 15:602–606

16. Li L, Yan Z, Shen Q, Cheng G, Gao L, Zhang B (2019) Water body extraction from very high spatial resolution remote sensing data based on fully convolutional networks. Remote Sens 11:1162

17. Duan L, Hu X (2019) Multiscale refinement network for water-body segmentation in high resolution satellite imagery. IEEE Geosci Remote Sens Lett 17:686–690

18. Guo H, He G, Jiang W, Yin R, Yan L, Leng W (2020) A multi-scale water extraction convolutional neural network (MWEN) method for GaoFen-1 remote sensing images. ISPRS Int J Geo Inf 9:189

19. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556

20. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, Las Vegas, NV, USA, 27–30 June 2016, pp 770–778