ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

Semantic Segmentation for Waterbody Extraction Using Superpixels and Convolutional Neural Networks Classifier

Salim Iratni[1], Ferhat Attal[2], Yacine Amirat[2], Abdelghani Chibani[2] and Moussa Diaf[1]

1- LVAAS Laboratory, Mouloud Mammeri University of Tizi-Ouzou
2- LISSI Laboratory, Paris-Est Creteil University

Abstract. Waterbody extraction from satellite images is an important task for many applications, such as hydrological modeling, ecosystem monitoring and water reserve level tracking. To tackle this problem, several deep learning based approaches have been proposed in the literature. However, these approaches have difficulty delineating water bodies due to their variations in color, size and shape. To overcome this limitation, a novel deep learning-based approach that leverages the power of Convolutional Neural Networks (CNNs) and Superpixel technique using Simple Linear Iterative Clustering (SLIC) algorithm is proposed. The proposed method involves an initial over-segmentation of the input satellite image into homogeneous zones using the SLIC algorithm. These zones are then further processed to extract Regions Of Interest (ROI) that are classified as either water or non-water using a CNN model. Finally, each pixel within a homogeneous zone is assigned the predicted class of its associated ROI. The obtained results using Gaofen Image Dataset show the effectiveness of the proposed approach, while highlighting its superiority over state-of-the-art (SOTA) approaches.

1 Introduction

Water is life, making it the most important element for human vitality. However, in some cases, water can also cause natural disasters, such as floods. This is why monitoring water resources and reserves has always been a crucial task. Since water resources and reserves often cover vast areas, satellite imagery offers a fast and accurate solution for remotely monitoring these regions. Extracting water bodies from satellite images is a necessary step in various environmental remote sensing applications, including soil hydrological modeling, ecosystem monitoring, and water reserve level tracking [1]. In recent years, several deep learning based approaches for waterbody extraction have been proposed in the literature [2]. Although these approaches offer great potential, they often have difficulty in delimiting the boundaries between aquatic surfaces and the others. In this paper, to efficiently extract water bodies from satellite images, we propose a novel deep learning-based approach by combining CNN model and SLIC algorithm. The proposed approach includes an over-segmentation of the input satellite image, dividing it into homogeneous zones using the SLIC algorithm. The SLIC algorithm allows to effectively delimit water bodies. ROI are then extracted from these zones and subsequently classified as *water* or *non-water*

ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

using a CNN model. Finally, the predicted class of each ROI is assigned to all pixels in the corresponding zone, thus producing a semantic segmentation of water bodies in the image. The rest of the paper is organized as follows. Section 2 provides a detailed description of the proposed approach. Section 3 presents and discusses the obtained results. Finally, Section 4 concludes the paper and proposes future work.

2 Proposed approach

The framework of the proposed approach for waterbody extraction is presented in Figure 1. This framework includes three main steps, namely, over-segmentation, ROI extraction and ROI classification. The over-segmentation step consists of dividing input image into k superpixels using the SLIC algorithm. ROI are then extracted from each superpixel by identifying the largest square within the sub-region, resized to 30×30 pixels. The ROI classification step consists of classifying each resized ROI, into *water* or *not-water* surface using a CNN model. In addition, each pixel within sub-region k is assigned the predicted class of its associated ROI.

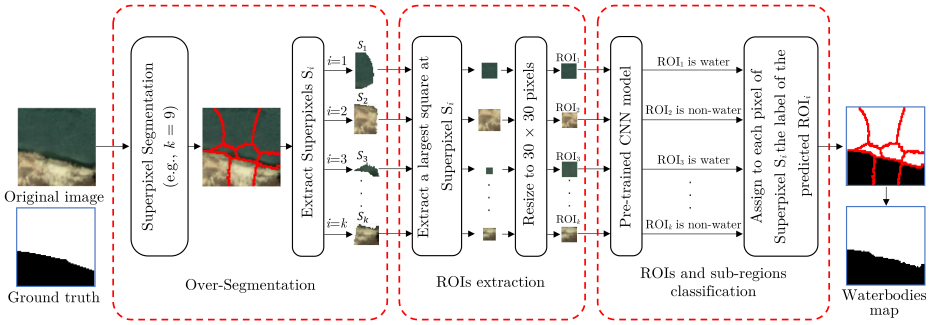

Fig. 1: Framework of the proposed approach for waterbody extraction

2.1 Superpixel segmentation

Superpixel segmentation involves the over-segmentation of an image by grouping similar adjacent pixels, resulting in homogeneous sub-regions. In this work, the SLIC algorithm [3], known as one of the most widely used methods for superpixel segmentation, is used to perform the over-segmentation of input images. This algorithm is based on a modified version of K-means, which clusters pixels based on their color similarity and spatial proximity. This step allows effectively delimits the aquatic zones from the others.

ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

2.2 ROI extraction

The ROI extraction consists of extracting the largest squares within the homogeneous sub-regions identified by SLIC, which are then resized to a uniform size (30×30). These largest squares are determined using an iterative algorithm that starts from the center of gravity of the sub-region and expands by one pixel in each direction: left, right, up, and down until a boundary pixel is reached.

2.3 ROI and sub-regions classification

To classify the ROI, a CNN classifier is used, as it remains among the most effective models for image classification tasks. The architecture of the CNN model used in this paper is illustrated in Figure 2. This architecture was selected after testing several models, starting with shallow architectures and gradually transitioning to deeper ones to improve performance. Given the small image size (30×30 pixels), the number of MaxPooling operations is limited to four to prevent excessive feature reduction. During testing, we found that adding more than two convolutional layers before each MaxPooling step did not yield significant performance improvements. The architecture employs kernel sizes of 3×3, a pool size of 2×2, a stride of 2 on both axes, and the ReLU activation function, chosen for its efficiency in non-linear feature mapping.

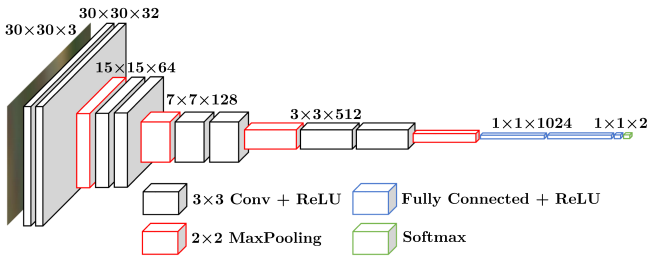

Fig. 2: Architecture of the proposed CNN model

3 Results and Discussion

3.1 Dataset description and pre-processing

To evaluate the performance of the proposed approach, the Gaofen Image Dataset (GID) [1] is used. GID consists of images captured by the Gaofen-2 (GF-2) satellite, launched in 2014. This dataset includes 150 multispectral images with a resolution of 6908×7300 pixels. These images are labeled into 24 classes, providing high intra-class diversity and low inter-class separability, which makes them widely used in several segmentation works. The annotated classes can be grouped into five main categories: built-up areas, agricultural land, forests,

[1] https://paperswithcode.com/dataset/gid

ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

grasslands, and waterbodies. In this work, we focus on two classes: all water-related categories (e.g., rivers and lakes), which are combined into a single class named *water*, while all other categories are grouped into a second class named *non-water*. Each image of the dataset is divided into 728 RGB images with a resolution of 256×256 pixels.

3.2 Evaluation metrics

Different metrics for assessing image segmentation quality have been used. These metrics are calculated from the components of the confusion matrix, namely, True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN). However, due to the imbalanced classes in the dataset (less than 5% of water areas), metrics based on True Negatives (TNs) can be misleading and do not accurately reflect segmentation quality. Therefore, metrics focusing on True Positives (TPs) provide a better performance evaluation [4]. Therefore, the evaluation metrics used in this study are: Accuracy (Acc), Precision (Pre), Recall (Rec), F1-Score (F1), and Intersection over Union (IoU) [2].

3.3 Results

In this sub-section, the performance of the proposed approach is presented and discussed. The model's performance is evaluated using hold-out validation, where the dataset is split randomly into 80% for training and 20% for testing. To maximize the classification performance of the proposed approach, a parameters tuning step is carried out. For the CNN model, the optimal architecture, presented in sub-section 2.3, is used. The training process is set to run for 50 epochs, with an initial learning rate of 0.01 and a batch size of 128. For the superpixel segmentation step, the only parameter to be tuned is the number of superpixels k. This number is set by varying k from 100 to 400 in steps of 10 for the two types of input data namely: RGB and the Hue (H) component of the uncorrelated Hue, Saturation, and Value (HSV) model. The optimal value of k, which yields the best performance, is found to be k=300.

Table 1 shows the performance of the proposed approach using RGB and H images as input data for the SLIC algorithm with k=300. It can be observed that the best performance across most evaluation metrics is achieved when using the H component. This can be explained by the fact that water bodies are distinctly characterized by their unique H values compared to other surface types, making the H component particularly useful in superpixel segmentation of water bodies.

The performance of the proposed approach has been compared to three widely used methods from the literature, namely UNet [5], SegNet [6], and DeepLabV3+ [7]. These methods are trained and evaluated using the same dataset. Table 2 shows that UNet with ResNet18 backbone achieves relatively

[2] $Acc = \frac{TP+TN}{TP+TN+FP+FN}$, $Pre = \frac{TP}{TP+FP}$, $Rec = \frac{TP}{TP+FN}$, $F1 = 2 \times \frac{Pre \times Rec}{Pre+Rec}$, $IoU = \frac{TP}{TP+FP+FN}$.

ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

Method	Metrics				
	Acc	*Pre*	*Rec*	*F1*	*IoU*
CNN + SLIC(RGB)	0.971	0.657	**0.825**	0.732	0.577
CNN + SLIC(H)	**0.977**	**0.736**	0.808	**0.777**	**0.627**

Table 1: Performance Comparison of CNN+SLIC using RGB and H images data for k=300

low Rec (0.506) and IoU (0.442). SegNet delivers good result, particularly with the VGG19 backbone, achieving the highest Rec (0.822), along with a competitive F1 (0.763) as well as IoU (0.601). However, with VGG16 backbone, it results in lowest Pre (0.695). DeepLabV3+ delivers hight performance in terms of Acc (0.966) and Pre (0.825); however, it shows poor performance in Rec (0.366), F1 (0.507), and IoU (0.328). It can also be observed that the proposed approach achieves the best overall balance across all metrics, including the highest Acc (0.977), F1 (0.777), and IoU (0.627), while maintaining competitive Pre (0.736) and Rec (0.808). Overall, the proposed approach outperforms state-of-the-art models, followed closely by SegNet with the VGG19. The performance of these

Method	Backbone	Metrics				
		Acc	*Pre*	*Rec*	*F1*	*IoU*
UNet [5]	ResNet18	0.971	0.820	0.506	0.626	0.442
SegNet [6]	VGG16	0.974	0.695	0.812	0.748	0.583
	VGG19	0.976	0.711	**0.822**	0.763	0.601
DeepLabV3+ [7]	ResNet50	0.966	**0.825**	0.366	0.507	0.328
CNN + SLIC(H)	Proposed CNN	**0.977**	0.736	0.808	**0.777**	**0.627**

Table 2: Performance Comparison Between the Proposed and SOTA Approaches

models are noticeably similar, which can be explained by the imbalance in the dataset. The predominance of the majority class contributes to a overall high performance, as the theses models tend to favor this class. However, state-of-the-art models struggle to accurately identify the boundaries between the two classes, limiting their effectiveness in this case. This phenomenon is illustrated in Figure 3, which shows visual results of semantic segmentation for selected images. It can be observed that the proposed approach outperforms all other methods, achieving the most accurate segmentation with minimal FP and FN.

4 Conclusion and future works

This paper presents a novel deep learning-based semantic segmentation approach for waterbody extraction from satellite images. This approach exploits the classification power of CNN to classify homogeneous regions of interest, obtained through a superpixel over-segmentation method based on the SLIC algorithm. Tests conducted on the Gaofen Image dataset show a significant improvement

ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

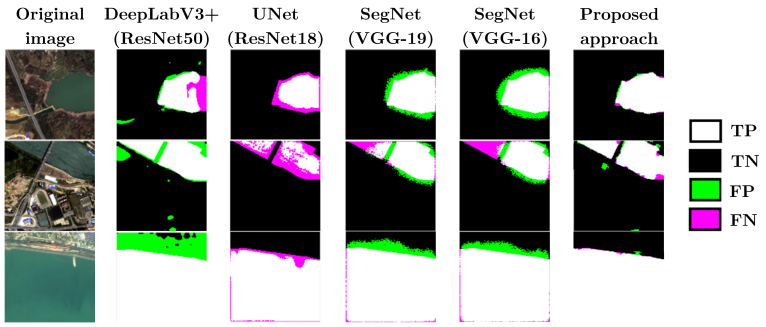

Fig. 3: Visual results of semantic segmentation

in performance compared to state-of-the-art semantic segmentation methods. Future work will focus on extending the proposed approach to multi-class segmentation, using larger and more diverse datasets, combining multiple classifiers, exploring alternative segmentation techniques, and addressing class imbalance.

References

[1] F. Wenqing, S. Haigang, H. Weiming, X. Chuan, and A. Kaiqiang. Water body extraction from very high-resolution remote sensing imagery using deep u-net and a superpixel-based conditional random field model. *IEEE Geoscience and Remote Sensing Letters*, pages 618–622, 2019.

[2] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos. Image segmentation using deep learning: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 3523–3542, 2022.

[3] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk. Slic superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 2274–2282, 2012.

[4] F. Kamalov A. Gonsalves F. Thabtah, S. Hammoud. Data imbalance in classification: Experimental evaluation. *Information Sciences*, pages 429–441, 2020.

[5] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention*, pages 234–241, 2015.

[6] V. Badrinarayanan, A. Kendall, and R. Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, pages 2481–2495, 2017.

[7] L. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Computer Vision - ECCV 2018*, pages 833–851. Springer.