Skip Connection Variant of Modified U-Net Architecture for Satellites Imagery

Yelim Lee
School of Electronic Engineering
Soongsil University
Seoul, Korea
yllee4806@soongsil.ac.kr

Jin Won Jung
School of Electronic Engineering
Soongsil University
Seoul, Korea
jinwonj@soongsil.ac.kr

Yoan Shin*
School of Electronic Engineering
Soongsil University
Seoul, Korea
yashin@ssu.ac.kr

(*Corresponding author)

***Abstract*— In this paper, we introduce an innovative model that builds upon the traditional U-Net architecture, specifically elevating the semantic segmentation performance for satellite imagery. Our architectural modification capitalizes on a concatenate block to effectively integrate the feature maps derived from each block. This strategic integration aids in mitigating the information loss from the extracted features and facilitates their equal distribution among numerous decoder blocks. The methodology underpins the capacity to augment semantic segmentation performance pertinent to satellite imagery, inherently marked by its intricate characteristics and wide-ranging features.**

Keywords — U-Net, Concatenate Block, Semantic-Segmentation, Satellites Image

I. Introduction

The existing U-Net architecture [1] has demonstrated exceptional results in various class segmentation tasks, notably in object segmentation for satellite imagery encompassing a wide variety of classes, such as buildings, roads, and water bodies. A distinctive feature of this architecture is the implementation of skip connections between the encoder and the decoder, facilitating the direct transmission of feature information between high-resolution and low-resolution feature maps. However, while the U-Net's skip connections present an advantage in reusing low-level input features, they are somewhat limited in learning high-level features. Specifically, in tasks with a variety of classes requiring high-level feature learning, such as the semantic segmentation of satellite imagery, this approach may result in the loss of feature information and potentially diminish segmentation performance [2].

This paper proposes a modification to the existing U-Net architecture aimed at overcoming these issues and preventing the loss of feature information. Further, it seeks to validate performance enhancement compared to existing semantic segmentation models, and aims to achieve more accurate semantic segmentation by reducing the risk of feature information loss.

II. Proposed U-Net Variant Model

A. Modified U-Net Structure

The U-Net architecture, composed of an intricate combination of encoders and decoders, serves as a deep learning-based segmentation network [1]. The cornerstone of this architecture resides in the encoder's capacity to transform the input image into a low-resolution feature map, which is subsequently forwarded to the decoder, resulting in the generation of a high-resolution output. In the standard U-Net structure, skip connections between the encoder and decoder are instituted to facilitate the reuse of low-level features.

As depicted in Fig. 1, our model diverges from the conventional U-Net structure, employing VGG16 [3] in the encoder block and introducing a concatenate block. This revised model allows for the integration of feature information extracted from multiple encoder blocks through the concatenate block. The integrated results are conveyed to each decoder block, culminating in an image output via the softmax operation. This mechanism allows for the holistic processing of feature information across various depths, thereby ensuring more effective transmission of feature information.

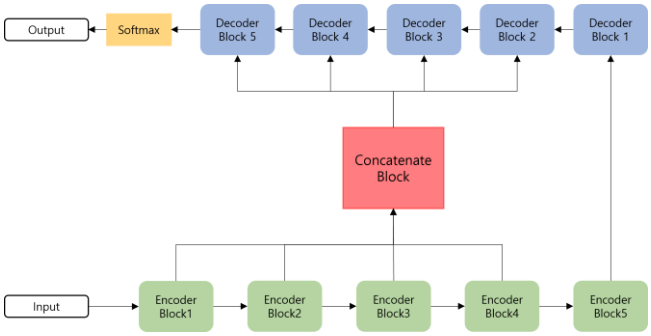

Fig. 1 The proposed model structure

B. Concatenate Block Details

The proposed novel concatenate block structure, as depicted in Fig. 2, allows for a detailed explanation. This structure adheres to a hierarchical design, repeatedly executing the merging process of feature maps from two encoder blocks at each stage. The merging process utilizes the concatenate operation [4], connecting feature maps from multiple stages and generating integrated feature maps with richer information.

In the initial stage, the first and second encoder blocks are merged through the “Concatenate1” operation, forming the first integrated feature map. Simultaneously, the third and fourth encoder blocks are merged through the “Concatenate2” operation, producing the second integrated feature map. Subsequently, the two integrated feature maps generated at each stage undergo further merging in the subsequent stage.

This process, also known as the “tournament-style” approach, ultimately forms integrated feature maps encompassing expanded depth and width, incorporating diverse feature information. An essential requirement for the concatenate operation is to ensure that the size of feature maps

This research was supported by the MSIT, Korea, under the ITRC support program (IITP-2023-2018-0-01424) supervised by the IITP.

extracted from each encoder block remains consistent. To achieve this, appropriate upsampling and downsampling processes have been employed, unifying the size of feature maps while concurrently facilitating fair processing of feature information at various levels.

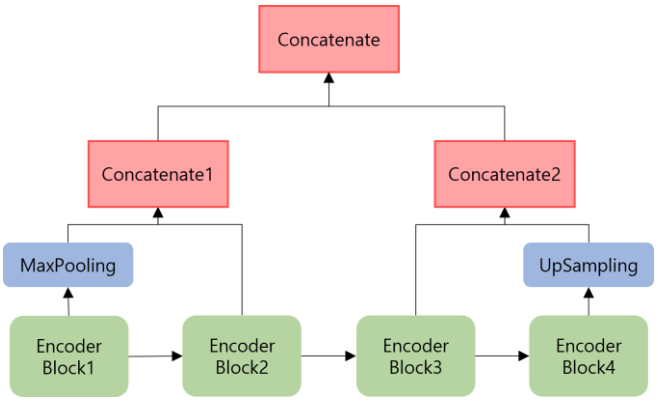

Fig. 2 Detailed structure of the concatenate block

III. Evaluation of Model Performance

A. Performance Metrics

Firstly, the experiments were conducted using the aerial imagery dataset of Dubai, collected through the utilization of satellites from Mohammed Bin Rashid Space Centre (MBRSC) [5]. To evaluate the performance of the proposed modified U-Net architecture, a comparison is made with the baseline U-Net[1] and LinkNet [6] models using performance metrics and results. The widely adopted evaluation metrics for semantic segmentation performance, namely the accuracy, the mean intersection over union(IoU) [7], and Dice coefficient(DSC) [8] are employed. The mean IoU and the DSC are defined as follows.

$$\text{Mean IoU} = \frac{|X \cap Y|}{|X \cup Y|} . \quad (1)$$

$$\text{DSC} = 2 \cdot \frac{|X \cap Y|}{(|X| + |Y|)} . \quad (2)$$

Here, X represents the set of actual pixels, while Y represents the set of predicted pixels. Metrics indicate higher agreement with the ground truth as they approach 1, and conversely, lower agreement as they approach 0.

B. Experimental Results

The results of this study are demonstrated in Table 1 and Fig. 3, confirming the contribution of the proposed concatenate block structure to the improved performance of semantic segmentation, considering the diverse natures of satellite image data.

TABLE I. Performance Comparison

Model	**Proposed**	U-Net	LinkNet
Accuracy	**90.2%**	88.5%	88.1%
Mean IoU	**77.6%**	70.2%	68.0%
Dice Coeff	**87.1%**	85.6%	84.6%

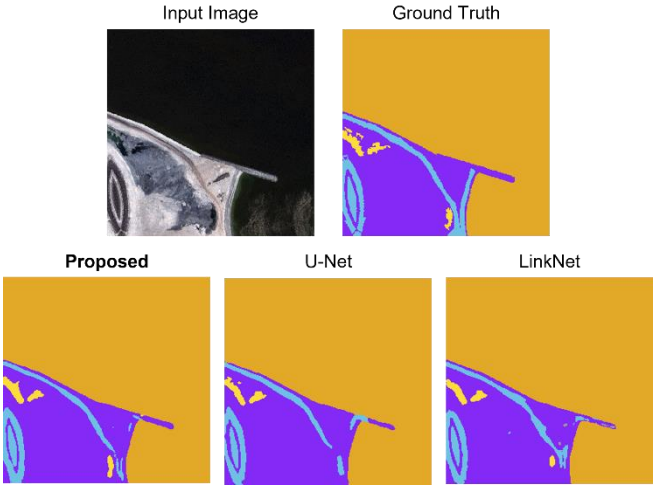

Fig. 3 Sample images of the experiment results

IV. Conclusion

In this study, we propose a novel approach to enhance the performance of semantic segmentation for satellite image data through the U-Net architecture and the proposed concatenate block structure. The proposed concatenate block effectively integrates feature information extracted at multiple depths between the encoder and decoder, aiming to minimize the challenges in learning high-level features and prevent feature information loss. Our future research will focus on maximizing the value of information obtained from complex satellite image data using deep learning algorithms, with the ultimate goal of continuously improving object detection and semantic segmentation performance.

References

[1] O. Ronneberger, A. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," *Lecture Notes in Comp. Sci*., vol. 9351, pp. 234–241, 2015.

[2] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," *Proc. IEEE CVPR 2015*, pp. 3431-3440, Boston, USA, June 2015.

[3] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition", *Proc. ICLR 2015*, San Diego, USA, Mav, 2015.

[4] G. Huang, Z. Liu, L. van der Maaten, and K.Q. Weinberger, "Densely connected convolutional networks," *Proc. IEEE CVPR 2017*, pp. 4700-4708, Honolulu, USA, July 2017.

[5] Humans in the Loop, "Semantic segmentation dataset," 2023, [Online]. Available: https://humansintheloop.org/resources/datasets/semantic-segmentation-dataset-2/ [Accessed: May 21, 2023].

[6] A. Chaurasia and E. Culurciello, "LinkNet: Exploiting encoder representations for efficient semantic segmentation", *Proc. IEEE VCIP 2017*, pp. 1-4, St. Petersburg, Russia, Dec. 2017.

[7] H. Rez., N. Tsoi, J. Gwak, A Sad., I. Reid, and S. Sav, "Generalized intersection over union: A metric and a loss for bounding box regression," *Proc. IEEE/CVF CVRR 2019*, pp. 658-666, Long Beach, USA, Feb. 2019.

[8] L. R. Dice, "Measures of the amount of ecologic association between species," *Jour. Ecological Soc. Amer*., vol. 26, no. 3, pp. 297-302, July 1945.