https://doi.org/10.7840/kics.2023.48.10.1219

이중 입력을 이용한 개선된 U-Net 모델 기반 위성 영상의 의미론적 분할

이 예 림•, 정 진 원*, 신 요 안°

Semantic Segmentation of Satellite Images Based on an Improved U-Net Model with Dual Inputs

Yelim Lee•, Jin-won Jung*, Yoan Shin°

요 약

본 논문은 고해상도 위성 영상을 이용한 의미론적 분할을 수행하는 새로운 딥러닝 기반 모델을 제안한다. 현존하는 딥러닝 기반 의미론적 분할 모델들은 인코더에서 얻어진 특징 정보가 제한적이므로, 이 정보가 디코더로 전달되어 정확도가 손상되는 문제를 가지고 있다. 이러한 한정된 특징 추출과 정확도의 저하는 예측의 비효율성을 초래하며, 이는 부정확한 결과를 유발한다. 본 연구에서는 이와 같은 한계를 극복하고자, 다양한 크기의 입력을 동시에 처리 가능하게 설계하여 기존 모델보다 향상된 인코더-디코더 구조를 제안한다. 이를 통해, 다양하고 풍부한 특징 정보를 효과적으로 추출하며 디코더로 더욱 효율적으로 정보들을 전달할 수 있고 의미론적 분할의 정확도를 향상시켜 성능을 높일 수 있다.

Key Words : semantic segmentation, deep learning, satellite images, U-Net, dual inputs

ABSTRACT

This paper proposes a new deep learning-based model for performing semantic segmentation using high-resolution satellite images. Because the feature information obtained from the encoder is limited in existing deep learning-based semantic segmentation techniques, they have a problem in that this information is transmitted to the decoder and accuracy is impaired. This limited feature extraction and degradation in accuracy causes inefficiency in prediction, which leads to inaccurate results. To overcome these limitations, this work proposes an encoder-decoder structure that is improved over existing architectures by designing inputs of various sizes to be processed at the same time. This effectively extracts a variety of rich feature information, enables more efficient delivery to decoders, and improves the accuracy of semantic segmentation.

Ⅰ. 서 론

영상의 의미론적 분할 (Semantic Segmentation)은 컴퓨터 비전의 한 분야로, 영상의 픽셀을 특정 클래스에 할당하는 과정이다[1]. 특히 위성 영상 기반의 의미론적 분할은 지리 정보, 농업 관리, 자연 재해 예측 및 감시 등에 널리 활용이 가능하고 광범위한 정보를 제공한다. 현재 딥러닝 알고리즘에 기반하여 인코더 (Encoder)와 디코더 (Decoder) 형태로 의미론적 분할을 수행하기 위한 모델들이 많이 존재하며, 대표적으로 U-Net[2]이 있다. U-Net은 인코더와 디코더 간의 스킵 연결 (Skip Connection)을 활용하여 특징 (Feature) 정보를 효율적으로 전달하는 장점을 가지고 있지만, 인코더에서 특징 정보를 충분히 추출하지 못하는 한계가 존재한다. 이렇게 인코더가 추출하는 정보의 양이나 질이 충분치 않을 경우, 픽셀 별 정확도가 낮아져 성능이 크게 저하된다.

이러한 한계를 극복하기 위해, 본 연구는 U-Net의 핵심 장점인 스킵 연결을 유지함과 동시에 인코더에 두 가지 서로 다른 크기의 영상, 즉 이중 입력 영상을 추출하여 입력 구조를 확장하는 새로운 모델을 제안한다.

※ 본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터지원사업의 연구결과로 수행되었음 (IITP-2023-2018-0-01424)
◆ First Author : (ORCID:0009-0009-9049-2750) School of Electronic Engineering, Soongsil University, yllee4806@soongsil.ac.kr, 학생(석사), 학생회원
° Corresponding Author : (ORCID:0000-0002-4722-6387) School of Electronic Engineering, Soongsil University, yashin@ssu.ac.kr, 정교수, 종신회원
* (ORCID:0000-0003-3026-0594) School of Electronic Engineering, Soongsil University, jinwonj@soongsil.ac.kr, 학생(석사), 학생회원
논문번호 : 202307-005-A-LU, Received July 6, 2023; Revised August 1, 2023; Accepted August 1, 2023

Ⅱ. 제안하는 이중 입력 U-Net 모델

2.1 이중 입력 모델 구성

제안하는 이중 입력 모델은 기존의 U-Net 기반 의미론적 분할 모델과 다르게 서로 다른 크기의 영상을 수용할 수 있으며, 이는 그림 1과 같이 구성된다. 첫 번째는 256×256×3의 크기, 두 번째는 512×512×3의 크기로 입력된다. 이 영상들에서 각각 독립적인 인코더 블록을 통해 특징이 추출되며, 이후 각 단계에서 Concatenate 연산을 통해 통합된다. 이렇게 통합된 특징 정보는 기존 U-Net의 스킵 연결 구조를 이용해 디코더로 전달되고, 이 풍부한 정보를 바탕으로 각 픽셀에 대한 세분화된 분류가 가능하다. 이러한 방식은 보다 풍부한 특징 정보를 활용하여 더욱 정밀한 의미론적 분할을 달성할 수 있다.

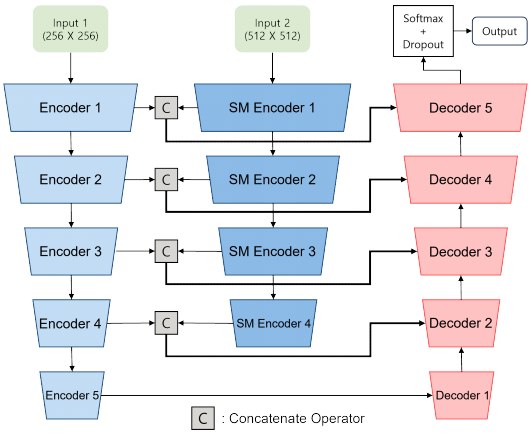

그림 1. 제안하는 이중 입력 U-Net 모델의 구조
Fig. 1. Structure of the proposed U-Net structure with dual inputs

2.2 Shape Matching Encoder 블록

제안된 모델에서 특징을 추출하는 모든 블록은 InceptionResNetV2[3] 아키텍처를 이용한다. 두 개의 입력 부분에 대해 각각의 특징 추출 과정에서 생성되는 출력의 크기가 일치하지 않는다. 이로 인해, 그림 2에서와 같이 첫 번째 입력인 Input 1과 두 번째 입력인 Input 2 각각의 인코더 블록 구조는 다르게 구성된다. 여기서, Input 1은 기존의 특징 추출 과정을 유지하여 첫 번째 특징 추출 결과가 128×128×64크기인 반면, Input 2는 초기 입력의 크기가 512×512×3이므로 기존 것을 사용하면 256×256×64크기의 특징을 추출한다. 하지만, Concatenate 연산을 수행하기 위해 모든 대상 블록의 크기가 같아야 한다. 따라서, Input 2는 Input 1의 인코더 블록에서 MaxPooling2D 연산을 추가하여, 256×256×64크기를 128×128×64크기로 변환 후 Concatenate 연산을 수행하고 디코더로 전달한다. 또한, 인코더 블록에서 MaxPooling2D 연산을 추가한 블록을 모양의 크기를 동일하게 맞춰준다는 의미로 Shape Matching Encoder라고 지칭한다.

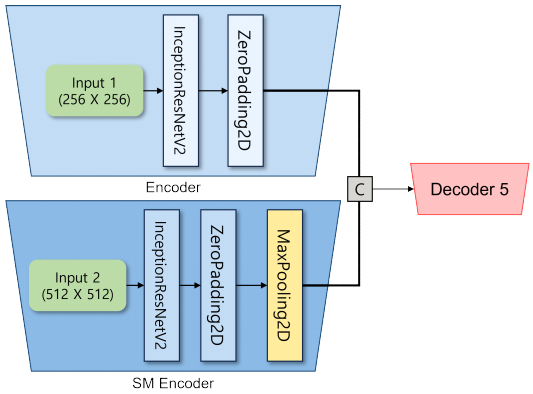

그림 2. Shape Matching Encoder 블록 구조
Fig. 2. Structure of shape matching encoder block

Ⅲ. 성능 평가

3.1 성능 지표

제안 모델에 대한 의미론적 분할 실험을 수행하고, 이의 성능을 평가하기 위해 Dice Coefficient[4], Mean Intersection over Union (MIoU)[5], 그리고 Frequency Weighted Intersection Union (FWIoU)[6] 세 가지 성능 지표를 활용하였다. 우선, Dice Coefficient는 예측 결과와 실제 결과 간의 유사성을 평가하며, 다음과 같이 정의된다[7].

$$Dice\ Coefficient = \frac{2(X \cap Y)}{|X| + |Y|}, \quad (1)$$

여기서 X와 Y는 각각 예측 결과와 실제 결과를 나타낸다. 분모의 $|X|$와 $|Y|$는 각각 실제 값과 예측 값이 1인 픽셀 값, 즉 Positive인 픽셀의 총 수이고 분자는 실제 값과 예측 값이 모두 1인 픽셀의 수이다. 따라서, 이 지표는 0과 1 사이의 값을 가지며, 값이 1에 가까울수록 두 영역 사이의 유사성이 높다는 것을 나타낸다.

MIoU는 분할된 각 클래스에 대한 IoU의 평균이며, IoU는 예측 영역과 실제 영역이 얼마나 겹치는지를 측정하는 지표로서 다음과 같다[5].

$$IoU = \frac{|X \cap Y|}{|X \cup Y|}, \tag{2}$$

$$MIoU = \frac{1}{n}\sum_{i=1}^{n} IoU_i, \tag{3}$$

식 (2)의 분모는 두 영역을 모두 포함하는 최소 영역의 면적을 의미하고 분자는 두 영역의 겹치는 부분의 면적이므로, 이 지표 역시 0에서 1 사이의 값을 가지게 되며 1에 가까울수록 두 영역이 겹쳐 좋은 성능임을 확인할 수 있다. 식 (3)에서 IoU_i는 i 번째 분할 클래스의 IoU 값, n은 전체 클래스 개수이다. 한편, FWIoU는 MIoU와 다르게 각 클래스 빈도 즉, 픽셀 빈도에 따라 가중치를 부여한다[6].

$$FWIoU = \frac{1}{\sum_{i=1}^{n} k_i}\sum_{i=1}^{n} IoU_i \cdot k_i, \tag{4}$$

여기서 k_i는 i 번째 분할 클래스가 차지하는 빈도 (비율)를 의미한다. FWIoU는 MIoU가 각 클래스의 해당되는 픽셀 비율을 고려하지 않아 발생하는 클래스 불균형 문제를 가중치를 이용하여 보완한다.

3.2 실험 결과

제안 모델의 성능을 기존의 대표적인 딥러닝 기반 의미론적 분할 모델인 U-Net 및 FPN (Feature Pyramid Network)[8]과 비교하였으며 제안한 모델을 제외한 비교 대상 모델들은 이중 입력이 아닌 512×512×3 크기의 입력 하나로 설정하였다.

실험은 MBRSC 위성에서 수집한 두바이 항공 위성 영상을 사용하였으며, 기본으로 제공하는 48장의 영상을 512×512×3 크기로, 총 1,036장의 영상으로 데이터 증강 기법을 포함하여 데이터 전처리를 한 후 모델 학습에 적용하였다[9]. 모든 모델들의 학습을 위해 학습 반복 횟수를 200으로 지정하였고, 과적합 방지를 위해 조기 종료 기법을 추가하고 배치 사이즈는 4로 하여 학습하였다.

표 1은 다양한 성능 지표 결과를 정리한다. 이 결과로부터, 제안 모델이 기존의 여러 모델들 보다 정확한 의미론적 분할을 수행하여 모든 성능 지표에 대해 우수한 성능을 보임을 알 수 있다. 특히, MIoU와 FWIoU는 기존 모델 대비 큰 향상이 있음을 확인할 수 있다.

한편, 그림 3은 대표적인 실제 실험 결과 영상을 도시하며, 제안된 모델이 기존 모델들에 비해 훨씬 정확한 의미론적 분할을 수행하였음을 보여준다.

표 1. 성능 지표 결과
Table 1. Performance metric results

Model	Performance metric results		
	Dice Coeff	Mean IoU	FWIoU
Proposed	**89.3%**	**79.9%**	**86.9%**
U-net[1]	88.4%	72.2%	83.3%
FPN[7]	85.4%	70.1%	80.3%

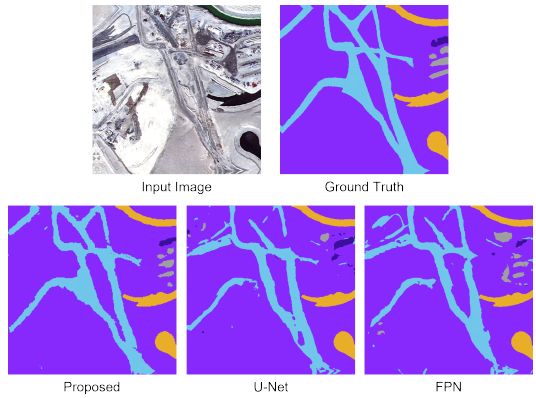

그림 3. 실험 결과에 대한 대표적인 예제 영상
Fig. 3. Representative example images of experimental results

Ⅳ. 결 론

본 논문은 위상 영상의 의미론적 분할을 위해 이중 입력을 이용하여 개선된 U-Net 구조를 제시하였다. 제안된 모델은 두 가지 다른 크기의 영상을 독립적으로 처리하여 풍부한 특징 정보를 추출하고, 이를 통합하여 효율적으로 전달한다. 이러한 접근 방식은 고해상도 위성 영상에서 기존 모델보다 성능 평가 지표에서 우수한 결과를 얻을 수 있었으며 높은 성능의 의미론적 분할을 가능하게 하였다.

References

[1] J. Jung and Y. Shin, “An extension of pre-trained deep learning model for sematic segmentation of gray-scale images,” *J. KICS*, vol. 48, no. 1, pp. 36-39, Jan. 2023. (https://doi.org/10.7840/kics.2023.48.1.36)

[2] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional networks for

biomedical image segmentation," *LNCS*, vol. 9351, Issue Cvd, pp. 234-241, Nov. 2015. (https://doi.org/10.1007/978-3-319-24574-4_28)

[3] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, Inception-ResNet and the impact of residual connections on learning," in *Proc. AAAI Conf. Artificial Intell. 2017*, pp. 4278-4284, San Francisco, USA, Feb. 2017. (https://doi.org/10.1609/aaai.v31i1.11231)

[4] K. H. Zou, S. K. Warfield, A. Bharatha, C. M. Tempany, M. R. Kaus, S. J. Haker, W. M. Wells III, F. A. Jolesz, and R. Kikinis, "Statistical validation of image segmentation quality based on a spatial overlap index," *Academic Radiology*, vol. 11, no. 2, pp. 178-189, Feb. 2004. (https://doi.org/10.1016/S1076-6332(03)00671-8)

[5] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized intersection over union: A metric and a loss for bounding box regression," in *Proc. IEEE/CVF CVRR 2019*, pp. 658-666, Long Beach, USA, Feb. 2019. (https://doi.org/10.1109/CVPR.2019.00075)

[6] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE/CVF CVPR 2015*, pp. 3431-3440, Boston, USA, Jun. 2015. (https://doi.org/10.1109/CVPR.2015.729865)

[7] L. R. Dice, "Measures of the amount of ecologic association between species," *J. Ecological Soc. Amer.*, vol. 26, no. 3, pp. 297-302, Jul. 1945. (https://doi.org/10.2307/1932409)

[8] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE CVPR 2017*, pp. 936-944, Honolulu, USA, Jul. 2017. (https://doi.org/10.48550/arXiv.1612.03144)

[9] *https://humansintheloop.org/resources/datasets/semanticsegmentation-dataset-2/*