

# 위성 영상의 의미론적 분할을 위한 이중 입력 딥러닝 모델

정진원, 신요안\*

승실대학교 전자정보공학부

jinwonj@soongsil.ac.kr, yashin@ssu.ac.kr

## Dual Input Deep Learning Model for Semantic Segmentation of Satellite Images

Jin Won Jung and Yoan Shin\*

School of Electronic Engineering, Soongsil University

(\*Corresponding Author)

### 요약

본 논문에서는 위성 영상에서 의미론적 분할을 위해 이중 입력을 사용하는 구조의 새로운 딥러닝 모델을 제안한다. 기존의 딥러닝 기반 의미론적 분할 모델은 학습 데이터가 적거나 데이터의 질감에 따라 분할 정확도가 줄어들며 물체를 제대로 구별하지 못하고 인식 성능이 저하된다는 문제가 발생한다. 제안하는 이중 입력 모델은 이러한 문제를 개선하기 위해 서로 다른 두 개의 입력을 사용하여 학습한다. 이는 각기 다른 두 개의 특징 추출 모델을 거쳐 서로 다른 특징을 학습하고 중간에 모델을 연결해 한쪽 모델에서 학습하지 못한 질감에 대한 패턴을 찾는 데 도움을 준다. 또한, 데이터 전처리 과정에서 가우시안 블러링을 사용해 노이즈를 최소화한 입력을 사용한다. 실험 결과, 제안하는 이중 입력 모델이 입력을 두 개만 사용하는 기존의 의미론적 분할 모델보다 나은 인식 성능을 보여준다.

### I. 서론

위성 영상은 촬영할 수 있는 범위가 매우 넓어 많은 영상을 획득할 수 있으며 다양한 지역의 데이터 획득이 가능하나, 동일 영역을 다시 촬영하는 주기가 길어 같은 지역을 촬영한 데이터가 적고, 지표면을 관측하기 때문에 기상 상황이나 계절에 따라 같은 영역에서도 이미지의 품질이나 질감이 크게 차이난다. 이러한 위성 영상의 특징으로 인해 분할 정확도가 줄어들며 물체를 제대로 구별하지 못하고 인식 성능이 감소된다[1].

본 논문에서는 위성 영상의 의미론적 분할 (Semantic Segmentation)[2]에서 인식 성능이 떨어지는 문제를 해결하기 위해 이중 입력 딥러닝 모델을 제안한다. 입력을 한 개만 사용하는 기존의 딥러닝 기반 의미론적 분할 모델과 다르게, 제안하는 이중 입력 모델은 두 개의 입력 데이터를 사용하며, 두 개의 모델 특징을 가져와 연결한 구조를 만들어 더욱 다양한 특징을 학습하고 특징 정보의 손실을 개선한다.

그림 1에서 볼 수 있듯이 각기 다른 두 개의 인코더 블록으로 구성되어 있다. 인코더 블록은 백본으로서 CNN (Convolutional Neural Network) 딥러닝 모델의 일종인 ResNet101[3]을 사용해 특징을 추출하는 역할을 하며 각각 다른 입력을 받게 된다. 이중 입력 모델에서 사용하는 입력은 두 가지로 나뉘는데, “Input1”에는 학습 데이터의 원본이 들어가며, “Input2”에는 학습 데이터에 가우시안 블러링(Gaussian Blurring)[4]을 적용해 위성 이미지가 가진 노이즈를 제거하거나 경계선을 흐리게 만들어 학습에 사용한다. 이렇게 두 개의 다른 입력으로 각기 다른 두 개의 모델이 다르게 학습하며 이후 연결 블록에서 두 개의 특징맵을 연결해 준다. 이는 서로 다른 두 개의 모델을 학습해 한 개의 모델에서 학습할 때보다 더욱 다양한 특징 패턴을 학습하고, 학습 데이터의 품질이나 질감에 크게 다를 경우 생기는 인식 저하 문제를 최소화하며, 디코더가 더욱 정밀한 분할 결과를 구성하는 데 도움을 준다.

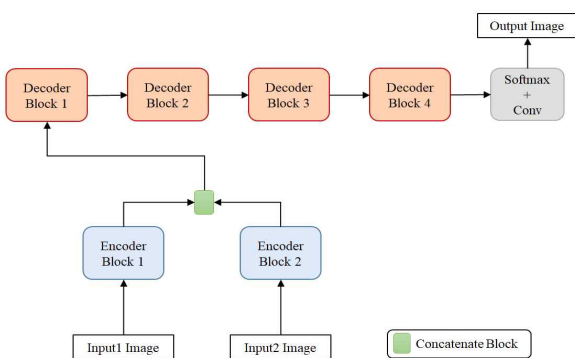


그림 1. 위성 영상의 의미론적 분할을 위해 제안하는 이중 입력 딥러닝 모델 구조

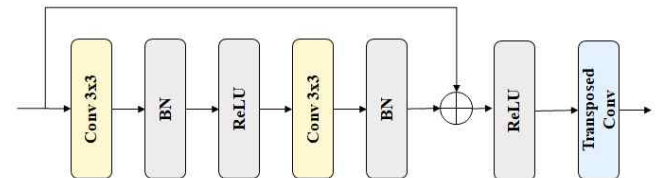


그림 2. 디코더 블록 구조

### II. 이중 입력 딥러닝 모델

제안하는 이중 입력 딥러닝 모델은 두 개의 입력을 가지는 모델로서,

그림 2는 이중 입력 모델의 디코더 구조를 나타낸다. 디코더 블록에서는 연결 블록을 통해 연결된 특징맵 (Feature Map)이 입력으로 사용된다. 이후 Convolution Layer (Conv)와 Batch Normalization (BN), Rectified Linear Unit (ReLU)가 연결된 구조를 지난다. 이 구조는 잔차학습 (Residual Learning)[5]을 적용하여 연결된 특징맵을 디코더의 뒷 부분에 더해준다. 이후 Transposed Convolution Layer (Transposed Conv)를 사용하여 특징맵의 크기를 커지게 만들어 준다. 이러한 잔차학습을 통해 특징맵의 추출과 압축 과정에서 생기는 특징 손실을 줄여준다.

### III. 실험 및 결과

실험을 위한 위성 영상 데이터 세트는 MBRSC 위성에서 얻은 두바이의 컬러 위성 영상으로서, 256x256의 크기로 총 1,136개의 영상을 전처리해 데이터 세트로 사용하였다[6]. 비교 대상 모델로서, CNN 기반 영상 분할 모델인 U-Net[7], 제안 모델에서 입력을 하나로 줄인 변형 구조, 그리고 제안하는 이중 입력 모델을 고려하였다. 성능평가 지표는 분할 정확도 (Accuracy) 및 다음 식과 같이 정의되는 Intersection over Union (IoU)의 평균인 Mean IoU를 사용하였다[8].

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

여기서,  $X$ 는 정답 픽셀,  $Y$ 는 예측 픽셀로서,  $X$ 와  $Y$ 의 두 영역이 같아질수록 1에 가까워지고 멀어질수록 0에 가까워진다.

표 1의 실험 결과를 보면 제안 모델의 Mean IoU가 79.6%로 물체를 잘 식별하였고, 정확도 또한 90.5%로 다른 모델에 비해 크게 높았다. 입력을 하나로 줄인 제안 모델의 경우 가장 결과가 좋지 않았는데, 이는 입력을 하나만 사용할 경우보다 가우시안 블러링을 적용한 입력을 하나 더 사용하는 게 의미론적 분할의 성능 향상에 큰 도움이 된다는 것을 알 수 있다. 그림 3은 샘플 위성 영상을 사용한 시각적 결과를 보여준다. U-Net과 입력을 하나만 이용한 제안 모델을 사용했을 때는 픽셀 노이즈가 나타난 것을 볼 수 있지만, 두 개 입력을 사용하는 제안하는 모델은 픽셀 노이즈 없이 거의 정확한 예측이 가능함을 보여준다.

표 1. 의미론적 분할 실험 결과. (A) U-Net, (B) 입력을 하나로 줄인 제안 모델, (C) 제안하는 이중 입력 모델.

Scheme	Backbone	Metrics	
		Mean IoU	Accuracy
(A)	ResNet101	71.4%	87.3%
(B)		67.7%	80.1%
(C)		<b>79.6%</b>	<b>90.5%</b>

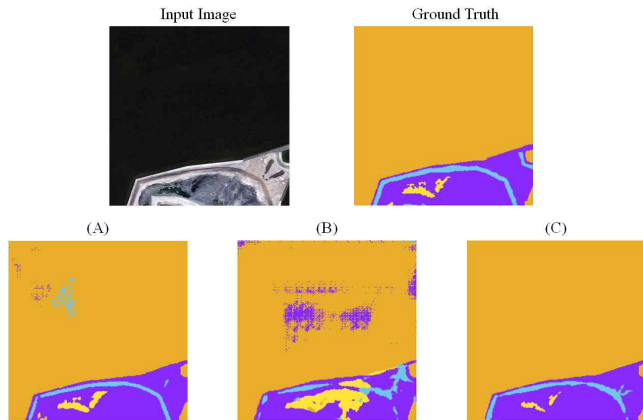


그림 3. 샘플 영상의 의미론적 분할 결과. (A) U-Net, (B) 입력을 하나로 줄인 제안 모델, (C) 제안하는 이중 입력 모델.

### IV. 결론

본 논문에서는 위성 영상의 의미론적 분할에서, 학습 데이터가 적고 이미지의 품질이나 질감이 크게 차이 나서 생기는 인식 성능 저하 문제를 해결하기 위해, 딥러닝 기반의 이중 입력 모델을 제안하였다. 제안하는 이중 입력 모델은, 원본 학습 데이터뿐만 아니라 가우시안 블러링을 사용

노이즈를 제거한 학습 데이터도 입력으로 넣어 두 개의 입력을 갖는 모델이다. 각기 다른 두 개의 특징 추출 모델을 거쳐 서로 다른 특징을 학습하고, 중간에 특징맵을 연결해 한쪽 모델에서 학습하지 못한 질감에 대한 패턴을 찾게 도움을 주어 효율적으로 학습하였다. 특히 입력을 하나만 사용하는 의미론적 분할 모델은 품질이나 질감이 크게 차이 나는 물체를 제대로 구별하지 못하고 성능이 좋지 못했지만, 제안하는 이중 입력 모델은 기존의 모델에 비해 크게 우수한 성능을 얻을 수 있음을 확인하였다.

### ACKNOWLEDGMENT

"본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터사업의 연구결과로 수행되었음" (IITP-2023-RS-2023-00258639)

### 참 고 문 헌

- [1] B. Neupane, T. Horanont, and J. Aryal, "Deep learning-based semantic segmentation of urban features in satellite images: A review and meta-analysis," *Remote Sensing*, vol. 13, pp. 808 - 849, Feb. 2021.
- [2] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," *Proc. IEEE/CVF CVPR 2015*, pp. 3431 - 3440, Boston, USA, June 2015.
- [3] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proc. IEEE/CVF CVPR 2016*, pp. 770 - 778, Las Vegas, USA, June 2016.
- [4] R. Wang, W. Li, and L. Zhang, "Blur image identification with ensemble convolution neural networks," *Signal Proc.*, vol. 155, pp. 73-82, Feb. 2019.
- [5] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," *Proc. ECCV 2016*, pp. 630-645, Amsterdam, Netherland, Oct. 2016.
- [6] <https://humansintheloop.org/resources/datasets/semantic-segmentation-dataset-2/>
- [7] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *Proc. MICCAI 2015*, pp. 234 - 241, Munich, Germany, Oct. 2015.
- [8] 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," *Proc. IEEE/CVF CVPR 2019*, pp. 658-666, Long Beach, USA, Feb. 2019.