

위성 영상의 의미론적 분할을 위한 확장된 심층 신경망 연결 모델

정진원*, 신요안^o

An Extended Deep Neural Network Connection Model for Semantic Segmentation of Satellite Images

Jin-won Jung*, Yoan Shin^o

요약

위성 영상에서 의미론적 분할 성능을 향상시키기 위한 인코더-디코더 구조의 새로운 심층 신경망 모델을 제안한다. 기존 의미론적 분할은 얇은 구조로 인해 특징 압축과 확장 중에 생기는 손실이 크다. 이에 따라 분할 정확도가 줄어들며 물체를 제대로 구별하지 못하는 문제가 발생한다. 제안하는 확장 연결 모델은 이러한 문제를 해결하기 위해, 기존 인코더-디코더 모델을 확장시켜 여러 계층의 특징을 가져와 디코더에 연결한 구조를 만들어 특징 정보의 손실을 개선한 모델이다. 이러한 확장 연결 모델은 특징 학습 과정에서 정보 손실을 개선하고 업샘플링 과정에서 잔차학습을 사용하여 의미론적 분할의 정확도를 향상시켰다. 실험결과, 제안하는 확장 연결 모델이 기존 의미론적 분할 모델보다 크게 향상된 성능을 보여준다.

키워드 : 위성 영상, 의미론적 분할, 사전학습 모델, 전이학습, 잔차학습

Key Words : satellite image, semantic segmentation, pre-trained model, transfer learning, residual learning

ABSTRACT

We propose a novel deep neural network model of encoder-decoder structures to improve the performance of semantic segmentation in satellite images. Existing semantic segmentation has a lot of losses during feature compression and expansion due to its shallow structure. This reduces the accuracy of segmentation and leads to the problem of not being able to distinguish objects properly. The proposed extended connection model improves the loss of spatial information by expanding the existing encoder-decoder model to solve this problem, taking multiple layers of features and creating a structure connected to the decoder. These extended connection models improve feature loss in the feature learning process and improve the accuracy of semantic segmentation by using residual learning for upsampling. Experimental results show that the proposed model achieves significantly better performance than the existing semantic segmentation model.

1. 서론

영상 처리 분야에서 주로 사용하는 의미론적 분할 (Semantic Segmentation)은 특정 범주의 객체 또는 영역에 따라 픽셀 수준 단위로 클래스를 분류하는 작업으로서, 종양 검출, 토지 포지 분류, 도로 포지판 감지 등 다양한 분야에서 연구가 진행되고 있다. 최근 의미론적 분할은 주로 컨볼루션 신경망 (Convolutional Neural Network) 계열의 심층 신경망 (Deep Neural Network)을 이용한 방법들이 연구되고 있으며, 대표적인 방법에는 U-Net이 있다¹⁾. U-Net은 스킵 연결을 통해 인코더 (Encoder)의 정보를 디코더 (Decoder)로 전달해 정보 손실을 최소화한 모델이다. 특히, 복잡한 특징 (Feature)에서 정밀한 특징을 얻게 도와주며 가려져 있는 물체를 분할하는 데도 탁월하다. 그러나 위성 영상에 대한 의미론적 분할의 경우 작은 정보오류에도 큰 문제가 될 수

* 본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터지원사업의 연구결과로 수행되었음 (IITP-2023-2018-0-01424)

• First Author : (ORCID:0000-0003-3026-0594) School of Electronic Engineering, Soongsil University, jinwonj@soongsil.ac.kr, 학생(석사), 학생회원

^o Corresponding Author : (ORCID:0000-0002-4722-6387) School of Electronic Engineering, Soongsil University, yashin@ssu.ac.kr, 정교수, 종신회원

논문번호 : 202306-128-A-LU, Received June 17, 2023; Revised July 7, 2023; Accepted July 7, 2023

있으며, 인코더에서 추출하는 특징맵의 정보가 부족하여 객체의 분할이 잘못되거나 실패하는 문제가 있다.

본 논문에서 제안하는 확장 연결 모델은 ResNet152^[2]의 사전 학습된 모델을 사용하고, 디코더에 잔차학습 (Residual Learning)을 사용하여 특징 정보의 전달을 최대한 활용한다. 또한 인코더의 특징 손실을 줄이기 위해, 여러 층의 특징 정보를 디코더의 부분과 연결하여 특징 정보의 손실을 최소화 한다.

II. 제안된 확장 연결 모델

2.1 ResNet152 Block과 Concatenate Block

제안하는 확장 연결 모델은 그림 1에 보이는 것처럼 ResNet152 Block 층을 각각의 인코더로 사용하며, 업샘플링 (Upsampling)을 개선하기 위해 여러 층 인코더의 특징맵을 디코더에 연결하는 스킵 연결 (Skip Connection)이 사용된다. 스킵 연결은 Concatenate Block 에서 이루어지며, 특징맵은 인코더의 압축 중에 손실된 원래 공간 정보를 포함하여 디코더가 더욱 정밀한 분할 결과를 구성하는 데 도움이 된다. 이후 디코더

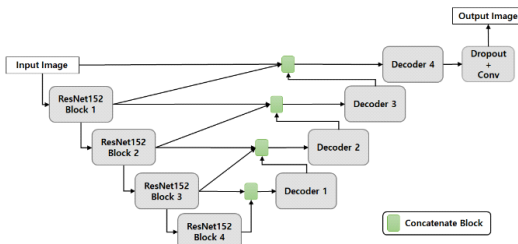


그림 1. 제안된 확장 연결 모델 구조
Fig. 1. Proposed extended connection model architecture

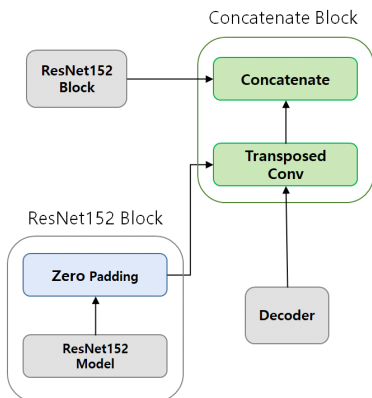


그림 2. ResNet152 Block과 Concatenate Block의 구조
Fig. 2. ResNet152 block and concatenate block architectures

를 지나 최종적으로 Dropout Layer와 Convolution Layer를 거쳐 출력을 얻는다.

그림 2는 ResNet152 Block과 Concatenate Block의 구조를 나타낸다. 인코더 역할을 하는 ResNet152 모델이 Zero Padding을 걸쳐 특징맵 정보를 Concatenate Block으로 전달한다. Concatenate Block은 Stride 2를 가진 Transposed Convolution으로 특징맵을 2배 확장한다. 그런 다음, 확장된 특징맵과 같은 해상도를 가지는 인코더 중간 특징맵과 디코더의 값을 각각 Concatenate하며, 다른 블록의 특징맵을 이어 붙이는 과정을 수행한다.

2.2 디코더

그림 3은 디코더의 구조를 나타낸다. x 는 디코더로 들어오는 특징맵이고, $A(x)$ 는 Convolution Layer 뒤에 Batch Normalization (BN), Rectified Linear Unit (ReLU)가 연결된 구조의 출력이다. 이 구조는 기존의 인코더에만 적용된 잔차학습을 디코더에도 활용하여 모델을 확장시킨다. Concatenate된 특징맵을 디코더의 마지막 부분에 더해준다. 식 (1)은 이러한 잔차학습 과정을 나타낸 것이다. 이 과정을 통해, 연속적인 Convolution의 특징 추출과 압축 과정에서 발생하는 특징 손실 및 네트워크가 깊어질수록 발생하기 쉬운 기울기 소실 문제를 최소화한다.

$$F(x) = A(x) + x. \quad (1)$$

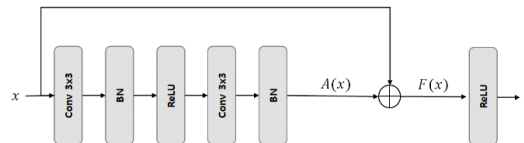


그림 3. 디코더 구조
Fig. 3. Decoder architecture

III. 실험결과 및 결론

실험은 기존의 의미론적 분할 방식인 U-Net^[1], Feature Pyramid Networks (FPN)^[3] 그리고 LinkNet^[4] 모델과 제안 방식의 성능을 각각 비교하였다. 데이터 세트는 MBRSC 위성에서 얻은 두바이의 컬러 위성 영상^[5]으로 구성되어 있으며, 256x256의 크기로 총 1,136 개의 영상을 전처리한 데이터 세트로 사용하였고 하이퍼 파라미터는 전부 동일하게 적용했다. 성능 평가 지표는 정확도 (Accuracy)와 다음 식과 같이 정의되는 모든 클래스에 대한 Intersection over Union (IoU)의 평균을

구한 값인 Mean IoU^[6] 그리고 Dice Coefficient (Dice)^[7]를 사용하였다.

$$IoU = \frac{|X \cap Y|}{|X \cup Y|}, \quad (2)$$

$$Dice = 2 \cdot \frac{|X \cap Y|}{(|X| + |Y|)}, \quad (3)$$

여기서, X 는 정답 픽셀, Y 는 예측 픽셀이며, 두 성능 지표 모두 X 와 Y 의 두 영역이 같아질수록 1에 가까워지고 떨어질수록 0에 가까워진다. 표 1에서 볼 수 있듯이, 제안하는 방법이 89.1%의 매우 높은 정확도를 보였으며, MIOU가 76.3%로 다른 비교 모델에 비해 물체를 보다 잘 식별하였다.

본 논문에서는 위성 영상에서 의미론적 분할 정확도를 향상시키기 위해, 전이학습, Concatenate과 잔차학습을 활용한 방법을 제안하였다. 디코더에도 잔차학습을 적용하는 방법을 통해 특징 손실을 줄이고 특징맵 정보를 효율적으로 학습하며, 제안하는 방법이 기존의 여러 모델에 비해 크게 우수한 성능을 얻을 수 있음을 확인하였다.

표 1. 실험결과
Table 1. Experimental results

Scheme	Metrics		
	Accuracy	Mean IoU	Dice Coff
Proposed	89.1%	76.3%	87.6%
U-Net[1]	86.8%	70.5%	82.7%
FPN[3]	83.5%	72.6%	83.5%
LinkNet[4]	84.5%	63.2%	79.4%

References

[1] 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)

[2] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE CVPR 2016*, pp. 770-778, Las Vegas, USA, Jun. 2016. (<https://doi.org/10.1109/CVPR.2016.90>)

[3] 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*, vol. 1, pp. 936-944, Honolulu, USA, Jul. 2017. (<https://doi.org/10.48550/arXiv.1612.03144>)

[4] A. Chaurasia and E. Culurciello, "LinkNet: Exploiting encoder representations for efficient semantic segmentation," in *Proc. IEEE VCIP 2017*, pp. 1-4, St. Petersburg, USA, Dec. 2017. (<https://doi.org/10.1109/VCIP.2017.8305148>)

[5] <https://humansintheloop.org/resources/datasets/semantic-segmentation-dataset-2/>

[6] 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>)

[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>)