식생지수를 이용한 다채널 농작물 이미지의 의미론적 분할

김용태*, 박성준*, 황승준*, 김희영**, 백중환*
*한국항공대학교, **(주)링크투
kyt10192006@kau.kr, jjunny9410@kau.kr, fogfog2@kau.kr, hykim@linkto.co.kr, jhbaek@kau.ac.kr

Semantic Segmentation of Multi-channel Crop Image Using Vegetation Index

Kim Yong Tae*, Park Sung Jun*, Hwang Seung Jun*, Kim Hee Yeong**, Baek Joong Hwan*
*Korea Aerospace Univ., **LinkTo Co.

요 약

최근 국내 농경지 데이터 구축과 스마트 농업에 관한 연구가 활발히 진행되고 있다. 특정 작물의 지속적인 생육 정보에 대한 수집을 위해서는 의미론적 분할이 기초적인 작업으로 수행된다. 의미론적 분할은 현재 다양한 최신기술이 적용되며 성능을 개선하기 위한 연구가 진행되고 있지만, 여전히 데이터의 수와 특성에 따라 높은 성능을 달성하지 못하는 경우가 발생한다. 이에 본 논문에서는 다채널 농작물 이미지로부터 식생지수 맵을 추출하고 다양한 딥러닝 모델을 적용하여 최적의 모델을 도출한다. 실험 결과 Deeplabv3plus 모델로부터 배경, 작물, 잡초에 대해 F1 스코어 0.945, 0.814, 0.499의 성능을 달성하였다.

Ⅰ. 서 론

최근 농업종사자가 줄어들며 작물 분석과 효율적인 생산을 위해 스마트 농업 및 정밀 농업에 관한 적극적인 연구가 필요하다. 잡초를 제거하고 원하는 작물에 대한 분석과 불법 작물 탐지 등에 주로 사용되는 의미론적 분할은 객체에 대한 분류를 이미지의 픽셀 단위로 하는 고수준의 분류 작업이다. 컴퓨터 비전에서 의미론적 분할은 다양한 응용 서비스의 기초가 되는 매우 중요한 분야이며 최근 멀티 모달 네트워크를 이용한 다운스트림 테스크로의 적용과 Transformer를 이용한 연구에서 높은 성능을 달성하며 활발한 연구가 이뤄지고 있다.[1] 하지만 객체의 특성과 데이터 수에 따라 각 모델의 성능 차이가 두드러지게 나타나고 작은 객체에 대한 정확도와 폐색, 블러 등 여전히 해결하기 힘든 문제들이 도전적인 과제로 남아있다. Transformer를 적용한 최신 모델은 전역 특징에 대한 학습 능력을 높여 높은 성능을 달성했지만, 농작물과 같이 크기가 작은 객체는 지역 특징에 대한 학습 능력 또한 성능에 큰 영향을 미친다. 따라서 본 논문에서는 농작물 이미지에 적합한 딥러닝 모델을 실험을 통해 제시한다.

Ⅱ. 본론

본 논문에서 사용하는 식생지수 NDVI (Normalized Difference Vegetation Index)는 근적외선 (NIR)과 적색광 (RED)의 차이를 측정하여 식생을 정규화한 지수로 아래 식 (1)과 같이 계산된다.

$$NDVI = \frac{NIR - RED}{NIR + RED} \quad (1)$$

건강한 작물일수록 엽록소로 인해 근적외선과 녹색광 반사와 적색광 흡수가 강하기 때문에 사람의 눈에 녹색으로 보인다. 따라서 NDVI 데이터는 녹색 작물에서 높은 값을 가지고 작물이 아닌 배경에서 0에 가까운 값을 가져 배경을 미리 분류하는 효과를 가진다. 본 논문에서는 이를 이용하여 대표적인 의미론적 분할 딥러닝 모델인 PSPNet, Deeplabv3, Deeplabv3plus, Beit 모델에 학습하였다.[2-5] 여기서 농작물 데이터에 가장 적합한 모델을 찾고 최적의 학습 매개변수를 실험적으로 확인하였다. PSPNet, Deeplabv3, Deeplabv3plus는 CNN 기반의 모델로 학습 시 같은 매개변수를 사용하였다. PSPNet은 백본 네트워크에서 추출된 특징을 평균 풀링을 이용해 주변 환경을 고려할 수 있도록 하여 전역 컨텍스트 정보를 학습하도록 설계하였다. 하지만 지역 특징 정보가 손실되어 작물 영역 내부와 주위에서 지역적으로 발생하는 잡초에 대한 분류 시 성능 저하가 발생할 수 있다. Deeplabv3는 백본 네트워크에서 추출된 특징에 확장 비율을 다르게 적용한 Atrous 콘벌루션을 수행하여 다중 스케일 컨텍스트 정보를 학습할 수 있도록 설계하였다. 출력 stride를 유지해 특징 해상도가 작아지는 것을 방지한 ASPP (Atrous Spatial Pyramid Pooling)를 통해 지역 특징과 전역 특징을 효과적으로 학습할 수 있도록 하였다. 하지만 최종 출력을 입력 이미지 해상도에 맞추기 위해 업샘플링 과정에서 객체에 대한 경계값 정보가 손실될 수 있다. Beit는 Transformer 기반의 모델로 CNN 기반 모델의 학습 매개변수와 다르게 설정하여 실험을 진행하였다. Beit는 가장 최근에 발표된 모델로 의미론적 분할의 대표적인 벤치마크인 ADE20K에서 최고 성능을 달성하였다.[6] 하지만 Beit는 입력 이미지를 패치로 나눈 후 일부 패치를 임의로 마스킹하고 비주얼 토큰을 예측하는 방식으로 학습되기 때문에 일반적인 객체 또는 큰 객체에 유리하다. 따라서 부분적인 폐색이 많고 작은 크기의 객체를 픽셀 단위로 분류해야 하는 농작물 데이터에 적합하지 않다.

Deeplabv3plus는 CNN 기반의 모델로 WeedNet을 학습시켰을 때 가장 높은 성능을 달성하였다. Deeplabv3plus는 기존 Deeplabv3에서 사용한 ASPP와 공간축과 채널축을 분리하여 연산하는 depthwise separable 콘벌루션을 결합하였다. ASPP는 다중 스케일 컨텍스트 정보를 학습하므로 의미론적 분할의 성능에 크게 관여한다. Depthwise separable 콘벌루션은 특징 맵의 각 채널 단위 콘벌루션 수행 후 1x1 콘벌루션을 통해 채널 축 콘벌루션 수행을 한다. 이렇게 두 축을 분리하여 연산을 수행하게 되면 채널 축의 학습 매개변수와 연산량을 효과적으로 줄일 수 있다.

Deeplabv3plus의 Atrous 콘벌루션으로 인해 풀링 레이어가 대체되어 정보의 손실이 최소화되고 수용영역이 넓어진다. ASPP 모듈에서 추출된 특징 맵에 확장 비율을 다르게 적용하여 다양한 수용영역을 확인할 수 있도록 하였다. 그리고 Deeplabv3의 업샘플링 과정에서 발생하는 세밀한 경계값들의 정보를 유지하기 위해 디코더를 설계하여 업샘플링 시에 콘벌루션 연산을 통해 정보 손실을 최소화하였다.

실험을 위해 다채널 농작물 학습 데이터로는 오픈 데이터 셋 WeedNet을 채택하였다.[7] WeedNet은 사탕무 작물과 잡초에 대한 이미지를 4채널 Sequoia 카메라가 장착된 UAV로 촬영한 데이터 셋이며 465장의 멀티스펙트럼 이미지로 구성되어 있다. 객관적인 성능 분석을 위해 성능 평가 지표는 F1 스코어를 채택하였으며 아래 식 (2)로 계산된다. 표 1에 각 모델에 대한 성능을 보이며 가장 성능이 높았던 Deeplabv3plus 모델에 NDVI 이미지 입력 시 추론 결과 예시를 아래 그림 1에 보인다.

$$F_1 = 2 \cdot \frac{Precision \times Recall}{Precision + Recall} \quad (2)$$

표 1. 최신 모델 성능 분석

Model	Background	Crop	Weed
PSPNet	0.945	0.764	0.471
Deeplabv3	0.918	0.753	0.443
Beit	0.942	0.764	0.481
Deeplabv3plus	**0.945**	**0.814**	**0.499**

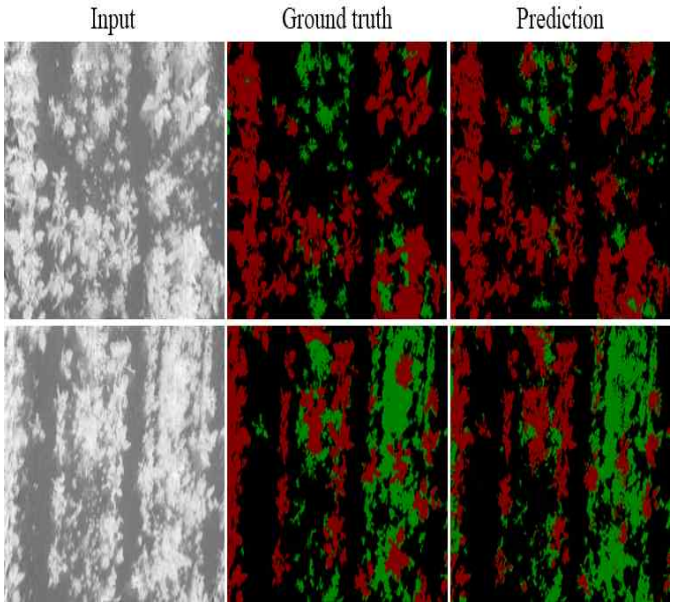

그림 1. NDVI 이미지 추론 결과 예시

입력 이미지는 NDVI 이미지로 녹색 식물이 있는 부분은 0.5~0.6의 값을 가져 흰색으로 보이며 배경 클래스에 해당하는 토양 부분은 검은색으로 보인다. NDVI 이미지 기반으로 의미론적 분할을 수행하기 때문에 RGB 이미지를 기반으로 사람이 라벨링을 수행한 ground truth보다 배경에 대해 더 세밀한 결과를 확인할 수 있다. 또한 넓은 수용영역으로 인해 이미지의 가장자리에서도 의미론적 분할이 큰 성능 저하 없이 수행되는 것을 확인할 수 있다.

Ⅲ. 결론

본 논문에서는 다채널 농작물 데이터 셋에서 효과적인 의미론적 분할을 수행하기 위한 최적의 모델을 도출한다. PSPNet, Deeplabv3, Deeplabv3plus, Beit 모델에 WeedNet 데이터 셋을 학습하고 성능 분석을 진행하였다. 농작물 데이터는 아주 작은 픽셀단위로 객체 간 폐색이 많이 발생하고 객체의 크기가 작다. 이러한 특성을 가진 데이터를 학습했을 때 CNN 기반의 모델에서 더 높은 성능을 보였으며 본 논문에서 제시한 실험을 통해 Deeplabv3plus 모델에서 배경 0.945, 작물 0.814, 잡초 0.499로 가장 높은 F1 스코어 성능을 달성하였다.

다채널 이미지는 채널별로 다른 특성을 가지기 때문에 병렬적으로 학습시켜 특징을 융합하는 연구가 필요하다. 향후 데이터 전처리 및 특징 융합에 관한 연구와 모델 내부 매개변수 및 구조 변경에 관한 연구를 통해 농작물 데이터에 대한 의미론적 분할 성능을 개선하고자 한다.

ACKNOWLEDGMENT

This research is supported by the GRRC program of Gyeonggi province [GRRC Aviation2017-B04, Development of Intelligent Interactive Media and Space Convergence Application System].

참 고 문 헌

[1] Vaswani, A., et al. “Attention is all you need,” Advances in neural information processing systems 30, 2017.

[2] Zhao, H. et al. “Pyramid scene parsing network,” In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2881-2890, 2017.

[3] Chen, L., et al. “Rethinking atrous convolution for semantic image segmentation,” arXiv preprint arXiv:1706.05587, 2017.

[4] Chen, L., et al. “Encoder-decoder with atrous separable convolution for semantic image segmentation,” In Proceedings of the European conference on computer vision (ECCV), pp. 801-818, 2018.

[5] Bao, H., et al. “Beit: Bert pre-training of image transformers,” arXiv preprint arXiv:2106.08254, 2021.

[6] Zhou, B., et al. “Semantic understanding of scenes through the ade20k dataset,” International Journal of Computer Vision, Vol. 127, Issue. 3, pp. 302-321, 2019.

[7] Sa, I., et al. “weednet: Dense semantic weed classification using multispectral images and mav for smart farming,” IEEE robotics and automation letters, Vol. 3, Issue. 1, pp. 588-595, 2017.