

# 드론 영상에서 정규화 방법에 따른 딥러닝의 잘피 분류 성능 비교

전의익, 김성학  
(주)지오스토리

e-mail:dmlldr0323@geostory.co.kr

## Comparative Evaluation of Seagrass Classification in Drone Images of Deep Learning according to Normalization Methods

Eui-Ik Jeon, Sung-Hak Kim  
Geostory Inc.

### 요약

본 연구는 해양생태계에서 중요한 역할을 하는 잘피 서식지의 모니터링에 드론 영상과 딥러닝을 활용하기 위한 기초 연구이다. 그래서 본 연구에서는 연안해역의 잘피 서식지를 대상으로 획득된 드론 광학영상으로 의미론적 분할을 수행함에 있어, 영상 정규화가 미치는 영향을 분석하기 위한 실험을 하였다. 영상 정규화가 미치는 영향에서는 원본 광학영상과 광학영상에 Z-score, Min-Max 정규화 기법을 적용한 영상들을 사용하였다. 그리고 이들 영상들에 대한 의미론적 분할 실험은 Encoder-Decoder 형태를 가지는 U-Net 모델을 사용하였다. 실험 결과, 정규화 방법에 따라 의미론적 결과의 정확도가 다르게 도출되었으며, 정규화 방법에 따라 원본 광학영상을 사용한 결과보다 정확도가 낮은 경우도 존재하였다. 이에 따라 연안해역의 고해상도의 광학영상의 의미론적 분할에 있어 적절한 정규화 방법을 결정하기 위한 실험이 선행되어야 하는 것으로 판단된다.

### 1. 서론

잘피는 다양한 해양생물의 산란 및 서식지를 제공하고 지구 온난화의 주요 요인인 이산화탄소를 흡수한다. 또한, 광합성 작용을 통해 해양생물의 호흡에 필요한 산소를 생산하여 공급한다는 점에서 매우 중요한 생태적 기능을 하는 생물이다 [1]. 그러나 1970년대 이후 산업화에 따른 무분별한 개발과 환경오염으로 인해 잘피의 개체 수가 현저하게 감소함에 따라 2007년에 해양수산부에서 보호대상 해양생물로 지정하여 한국수산자원공단에 의해 관리되고 있다[2-3].

최근에는 상대적으로 기상 상황에 덜 민감하고 높은 공간해상도의 영상을 손쉽게 획득할 수 있는 드론을 활용하여 잘피 서식지 모니터링이 시도되고 있다[4]. 하지만 드론 영상에서 잘피 서식지의 면적을 정량적으로 파악하기 위해서는 화소 단위의 분류인 의미론적 분할이 수행되어야 한다. 의미론적 분할을 위한 방법에는 합성곱 신경망(Convolutional neural network) 기반의 딥러닝이 높은 정확도로 보이고 있다. 그러나 드론 영상에서 잘피 서식지를 모니터링하기 위해 딥러닝을 이용한 연구는 활발하게 이루어지지 않고 있다.

따라서 본 연구에서는 드론 영상에서 잘피 서식지의 의미론적 분할에 딥러닝을 활용하기 위한 기초 연구로써, 입력 자료

의 정규화 방법에 따른 의미론적 분할의 정확도 평가를 비교하였다.

### 2. 실험자료

#### 2.1 드론 영상

본 연구의 대상지역은 남해의 24개 지역을 대상으로 2019년 2월 ~ 2020년 9월 사이에 드론을 이용하여 고해상도 광학 영상을 획득하였다. 이때 드론은 DJI사의 Phantom4 Pro와 Mavic2 Pro를 사용하였으며 드론들의 주요 스펙인 센서 크기, 화소 수는 각각  $13.2 \times 8.8\text{mm}$ ,  $5472 \times 3648$ 로 동일하고 초점 거리는 8.8mm, 10.26mm이다.

획득된 영상은 Agisoft사의 Metashape 소프트웨어를 이용하여 위치 정보를 가지는 Geotiff 형태의 정사영상을 24개 생성하였다. 정사영상을 생성하는 과정에서 공간해상도는 잘피의 구분이 충분히 가능한 10cm로 통일하였다. 정사영상으로부터 실측 자료를 구축하기 위해, Blue Marble Geographics사의 Global mapper 소프트웨어를 이용하였다. 잘피 서식지의 구분은 육안과 더불어 현장 조사 결과를 참고하였다.

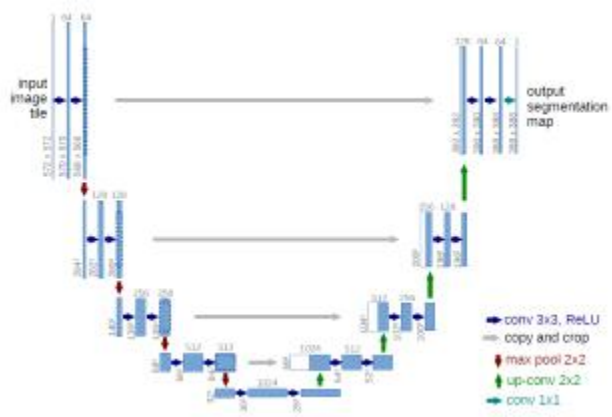
딥러닝 모델의 학습 및 평가에는 Google에서 제공하는 Colab Pro를 사용하였다. 다만, Colab Pro에서 제공하는 메모리의 크기가 최대 25.51GB로 제한적임에 따라 메모리를 최대한 활

용할 수 있도록 24개의 정사영상 중에서 16개를 학습에 사용하고, 나머지 정사영상 중 Validation dataset에 3개의 정사영상을 사용하였다. 그리고 학습된 모델의 객관적인 검증할 수 있도록 5개의 정사영상을 Test dataset으로 하였다.

### 2.2 딥러닝 모델

본 연구에서 사용할 딥러닝 모델인 U-Net은 ISBI 2015에서 바이오영상 분할 대회에서 우승을 차지한 모델로, ISBI 2014와 ISBI 2015의 학습 자료에 따라 IOU 기준으로 92%, 77.5%의 정확도로 다른 모델과 비교하여 광학 현미경 영상에서 불규칙한 크기와 형태를 가지는 헬라(HeLa) 세포의 경계를 가장 높은 정확도로 검출하였다[5].

U-Net 모델 구조는 [그림 1]과 같이 총 23개의 합성곱층으로 구성되며 영상 특징을 추출하는 부호화(encoder) 부분인 압축경로(contracting path)와 추출된 영상 특징을 원래의 입력으로 되도록 충실히 재현하는 것을 의미하는 복호화(decoder) 부분인 확장경로(expanding path)로 구분된다[6]. 압축경로는 합성곱과 풀링으로 입력 영상의 특징을 추출하면서 차원을 축소하는 과정으로 입력 영상의 가로, 세로 크기가 2칸씩 줄어드는 두 번의 3x3 합성곱으로 스트라이드(stride)가 2인 2x2 최대값 풀링(Max pooling)을 4번 반복하며, 각 다운샘플링 단계에서 특징 채널이 두 배로 증가하도록 하였다. 확장경로는 위치 정보에 대해 손실 없이 영상을 업샘플링하는 과정으로 2x2 합성곱을 사용하며 압축경로에서 생성된 특징 맵을 업샘플링한 영상과 접합(Concatenation)하여 지역화(localization)를 정교하게 하였다. 모델의 학습에서는 GPU 메모리를 최대한 이용하면서 오버헤드를 최소화하기 위해 배치보다 입력 타일의 크기를 크게 하는 방향으로 하였다.



[그림 1] U-Net 모델의 구조

### 2.3 영상 정규화

일반적으로 신경망에서는 학습 속도의 향상과 비용함수(cost function)의 최적화를 위해 입력 자료의 각기 다른 스케일을 일정한 범위로 변환하는 정규화가 이루어진다(Sammen et

al., 2018). 정규화의 대표적인 방법은 입력 자료가 정규 분포일 때 평균과 표준편차를 0과 1로 변환하는 Z-score 정규화가 있다. Z-score 정규화는 CNN의 초기 가중치가 0~1내에서 랜덤으로 선택될 때 비정상적인 기울기가 계산되는 것을 방지하며, activation과 gradient descent optimization을 개선한다[7].

$$z = \frac{x - \mu}{\sigma}$$

where z is Z-score normalized data,  $\mu$  and  $\sigma$  are the mean and standard deviation of x.

Z-score 정규화 이외에 입력 자료의 값의 범위를 0과 1 사이로 변환하는 Min-Max 정규화가 있으며, 원자료와 원자료의 최소치와 최대치를 이용한다. Min-Max 정규화는 원자료의 최소와 최대를 이용하기 때문에 데이터의 모든 관계를 보존하는 장점을 가지고 있으며, Z-score 정규화보다 표준편차의 값이 작고 이상치의 영향을 적게 해주는 것으로 알려져 있다[8].

$$X = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$$

where  $x_{\max}$ ,  $x_{\min}$  are maximum and minimum value of x. 방법에 관계없이 정규화를 하면 학습 시간의 단축과 학습 자료간의 범위를 동일하게하여 bias를 최소화하며, 분할 정확도가 향상되었다. 그러나 연구자들에 따라 8bit의 광학 영상의 화소 값의 범위는 0~255의 특정 범위로 고정되어 있는 점이 이미 정규화된 것으로 간주하였다[9]. 이에 따라 본 연구에서는 RGB 원본 영상과 Z-score, Min-Max 정규화 방법을 사용하였다.

### 3. 실험결과

정규화 유무와 방법이 신경망에 미치는 영향을 분석에 사용한 딥러닝 모델은 상대적으로 파라미터의 수가 적어 학습이 빠른 U-Net을 사용하였다. 학습 과정에서 accuracy, loss, precision, recall의 경향을 살펴보기 위해, 5회 반복한 결과의 평균에 가장 가까운 회차의 epoch에 따른 결과를 [그림 2]에 가시화하였다. 먼저 [그림 2]에서 학습자료의 그래프를 살펴보면, epoch가 200 이하에서는 RGB Z-score 영상이 다른 영상들보다 빠르게 높은 정확도와 낮은 loss 값에 도달했다. 그러나 epoch가 500회에 도달했을 때는 모든 그래프가 거의 동일한 값에 수렴하였다. Validation에 대한 결과에서도 Z-score 정규화가 적용된 영상의 정확도가 높고 loss가 낮다. 반면에 정규화가 적용되지 않은 영상들에 대해서는 반대 경향이 나타났다. 상대적으로 validation의 recall이 안정적으로 높게 나타나는 이유는 validation dataset의 영상에서 잘피 서

식지의 비율이 0.34% ~ 11.04%로 낮기 때문에 추측된다. 즉, 잘피 화소를 잘피가 아닌 화소로 판단한 FN의 절대적인 수가 적음을 의미한다. 전체적으로 정규화된 영상들의 결과가 좋음에 따라 테스트 영상들의 분할 평가에서도 원본 영상보다 정규화된 영상들의 결과가 좋을 것으로 예측할 수 있다. 각 학습이 종료될 때마다 테스트 영상의 타일에 대해 예측하고 예측된 타일을 다시 하나의 정사영상 형태의 predicted map으로 재구성하였다. 각각의 predicted map을 GTs와 비교하여 accuracy, precision, recall, IoU를 계산하는 과정을 5회 반복하고 그에 따른 결과를 도출하였다. 도출된 결과들의 테스트 영상별로 평균값을 계산하여 [표 1]에 명시하였다.

먼저, [표 1]에서 IoU를 기준으로 정확도가 높은 순서는 RGB Min-Max, RGB, RGB Z-score로 나타났다. RGB Min-Max의 결과가 다른 결과들보다 2%~15%의 높은 정확도를 보였다. 그러나 RGB와 RGB Min-Max의 IoU가 유사하다하더라도 precision, recall는 각각 81%, 85%, 87%, 82%로 반대 경향을 보인다. Precision과 recall의 차이가 계산식의 분모에서 FP와 FN에 의해서 결정되는 것에 따르면, RGB 영상으로 학습한 신경망이 RGB Min-Max로 학습한 신경망보다 상대적으로 FP가 높고 FN가 낮다. 즉, RGB 영상이 RGB Min-Max 보다 잘피 화소를 잘피가 아닌 화소로 오인한 비율이 높고, RGB Min-Max 영상이 잘피가 아닌 화소를 잘피 화소로 오인한 비율이 높은 것이다. 이렇게 상대적으로 precision이 낮고 recall 높게 나타나는 경향은 정규화를 하지 않은 영상에서 동일하게 나타났으나, 정규화를 수행한 영상에서는 precision이 높고 recall이 낮은 것으로 나타났다.

테스트 영상들의 결과에서도 차이가 있었는데, 이는 validation이 train과 test 을 대표하지 못했기 때문으로 판단된다. 그래서 학습 과정에서의 평가 수치를 신경망의 성능을 과소 또는 과대 평가할 수 있는 가능성을 최소화하기 위해서,

train과 test 영상들을 대표할 수 있는 적절한 validation의 선택이 이루어져야 할 것이다.

[표 1] 정규화 방법에 따른 테스트 영상들의 정확도

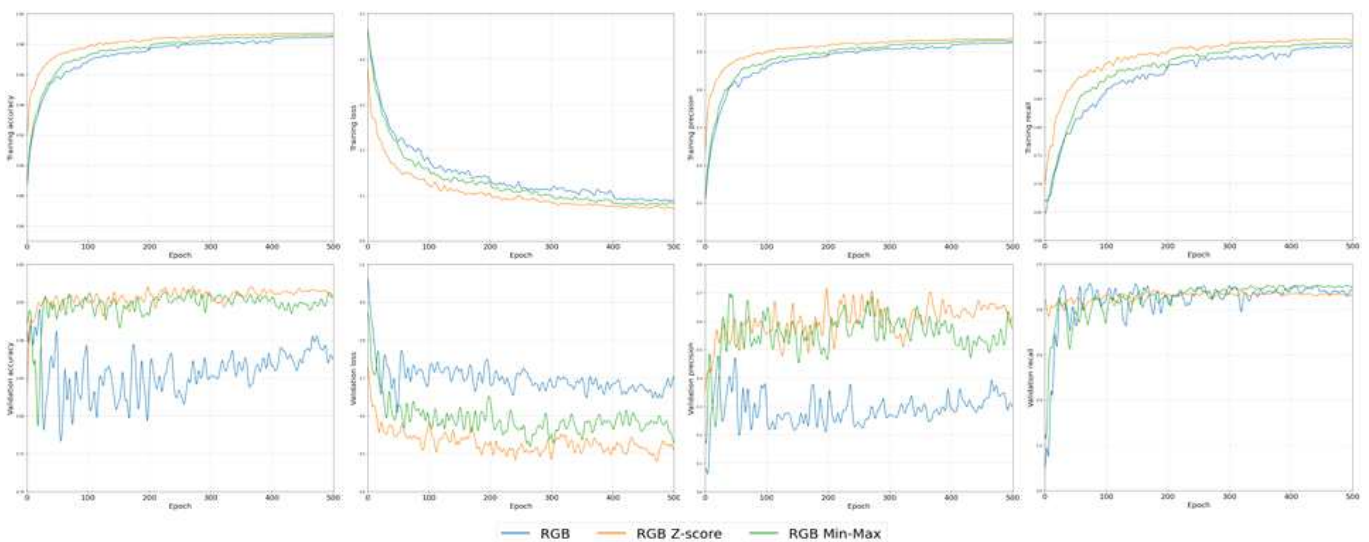
Test Image No.		1	2	3	4	5
RGB	Accuracy	0.95	0.97	0.88	0.87	0.96
	Precis.	0.85	0.87	0.52	0.89	0.93
	Recall	0.97	0.96	0.94	0.6	0.8
	IoU	<b>0.83</b>	0.84	0.5	<b>0.55</b>	0.75
RGB -zscore	Accuracy	0.96	0.95	0.91	0.8	0.97
	Precis.	0.97	0.92	0.57	0.93	0.98
	Recall	0.84	0.76	0.96	0.31	0.84
	IoU	0.82	0.71	0.56	0.31	<b>0.83</b>
RGB -minmax	Accuracy	0.95	0.98	0.95	0.83	0.97
	Precis.	0.86	0.92	0.7	0.91	0.97
	Recall	0.94	0.94	0.98	0.42	0.82
	IoU	0.81	<b>0.87</b>	<b>0.69</b>	0.4	0.8

#### 4. 결론

본 연구는 효율적인 잘피 서식지 모니터링을 위해 드론 광학 영상과 신경망 기반의 딥러닝을 활용하기 위해 영상 정규화가 미치는 영향을 분석하였다. 그 결과, 영상 정규화 방법에 따라 정확도 차이가 크게 나타났으며, 오히려 정규화를 하지 않는 것이 높은 정확도를 보이는 경우도 있었다. 따라서 본 연구를 포함하여 선행 사례가 부족한 영상의 의미론적 분할에서는 적절한 영상 정규화 방법을 선정하기 위한 실험이 선행되어야 할 것으로 판단된다.

#### 사사

본 연구는 해양수산부의 해양수산환경기술개발사업(과제번호:20170318)에서 지원받아 수행되었음.



[그림 32] 영상의 정규화 방법에 따른 학습 및 검증 자료의 정확도 비교

## 참고문헌

- [1] Thomas, F. I., & Cornelisen, C. D. (2003). Ammonium uptake by seagrass communities: effects of oscillatory versus unidirectional flow. *Marine Ecology Progress Series*, 247, 51-57.
- [2] Lee, K. S. (2003). 18 The seagrasses of THE REPUBLIC OF KOREA. *World Atlas of Seagrasses*
- [3] Park, J. I., Park, J. Y., & Son, M. H. (2012). Seagrass distribution in Jeju and Chuja islands. *Korean Journal of Environmental Biology*, 30(4), 339-348.
- [4] Duffy, J. P., Pratt, L., Anderson, K., Land, P. E., & Shutler, J. D. (2018). Spatial assessment of intertidal seagrass meadows using optical imaging systems and a lightweight drone. *Estuarine, Coastal and Shelf Science*, 200, 169-180.
- [5] Li, W., He, C., Fang, J., & Fu, H. (2018). Semantic segmentation based building extraction method using multi-source gis map datasets and satellite imagery. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (pp. 238-241).
- [6] Okatani, T.; Saito, M. *Deep learning*, Koudansha Ltd, Tokyo, Japan, 2015; pp. 101-121
- [7] Sameen, M. I., Pradhan, B., & Aziz, O. S. (2018). Classification of very high resolution aerial photos using spectral-spatial convolutional neural networks. *Journal of Sensors*, 2018.
- [8] Alickovic, E., & Subasi, A. (2019, May). Normalized neural networks for breast cancer classification. In *International Conference on Medical and Biological Engineering* (pp. 519-524). Springer, Cham.
- [9] Rakhlin, A., Davydov, A., & Nikolenko, S. (2018). Land cover classification from satellite imagery with u-net and lovász-softmax loss. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (pp. 262-266).
- 김갑수, “LCD 모듈 테스트 설계”, 산학기술학회논문지, 제 1권 1호, pp. 45-52, 1월, 1999년.