

GAN 과 확산 모델을 사용한 Labeled Data 생성 Generating Labeled Datasets with GAN and Diffusion Models

윤준영 · 강대열 · 김필은
Junyoung Yun · Daeyeol Kang · Pileun Kim

한국항공대학교
(ij04095@kau.kr)

ABSTRACT

To train deep learning models, a massive amount of data is required. But annotating those data one by one is exhausting and expensive work to do. We suggest a generative model based on SPADE to generate synthetic images from the segmented label. For an experiment, we trained a model generating aerial photographs with segmented labels, and further, applied a diffusion model to generate even segmented labels. We expect robust weights for existing models with fewer resources using datasets generated from this model.

Key Words : Generative model, GAN, Diffusion

1. 서론

객체 인식 모델이나 의미적 분할 등의 딥러닝 모델을 학습하기 위하여 큰 규모의 데이터가 요구된다. 데이터셋 제작하기 위해 이미지를 수집한 후 수집된 이미지를 라벨링하는 작업은 사람이 수행하게 되는데 이는 많은 시간과 비용이 소모된다. 생성 모델은 순수한 잡음에서 데이터를 생성하여 이미지를 수집하는 비용을 줄일 수 있지만 생성되는 이미지를 학습에 활용하기 위해선 라벨링 작업은 다시 수행해 주어야 한다. 따라서 본 연구에서는 생성 모델을 사용하여 이미지의 라벨과 이미지 자체를 모두 생성하는 모델을 제안한다. 본 연구에서는 항공사진과 라벨로 학습한 비조건부 확산 모델을 이용하여 새로운 라벨을 생성하고, 같은 데이터로 학습한 GAN 모델을 이용하여 생성된 라벨을 통해 이미지를 생성하였다.

2. 데이터셋 생성 모델

2.1 모델 구조

본 연구에서 사용한 생성 모델은 확산 모델을 통한 라벨 생성, 라벨 후처리, GAN 모델을 통한 이미지 생성의 3단계로 이루어져 있다.

2.1.1 확산 층

확산 모델은 데이터 분포를 확산시켜 순수 노이즈로 만든 뒤 다시 역으로 데이터 분포를 복구하는 과정을 학습하는 모델이다. 확산 모델은 기존 생성 모델에 비해 높은 성능을 보여주며, 생성된 데이터 다양성 측면에서도 높은 지표를 보여준다. 확산 층에서는 원본 데이터셋의 라벨을 학습하여 GAN 모델이 이미지를 생성하게 하기 위한 라벨 이미지를 먼저 생성해 준다. 본 연구에서는 DDPM^[1] 모델을 사용하여, 비조건부 확산으로 라벨 데이터를 생성한다.

2.1.2 라벨 후처리 층

확산 모델에서 생성한 라벨 이미지를 그대로 GAN 모델에 입력하였을 경우 여러 가지 문제점이 발생하였다. 먼저 라벨 섬(주위 라벨과 다르게 생성된 작은 라벨 픽셀)이 발생하여 작은 픽셀의 어색한 이미지가 생성되는 것을 확인할 수 있었다. 또한 서로 다른 라벨 사이의 램프 에지가 형성되어 두 라벨 사이에 의도하지 않은 다른 중간값을 가진 라벨이 생성되었다. 라벨 이미지의 경우 중간값의 라벨은 두 라벨의 중간의 의미가 아닌 전혀 다른 의미를 가지게 되어 어색한 이미지가 생성되었다. 이를 해결하기 위하여 열림 연산과 해당 현상이 발생한 부분을 최빈 이웃값으로 대체하여 라벨 섬과 램프 에지를 제거한 후 업스케일링을 통해 원본 라벨 이미지의 크기와 같게 설정하였다.

2.1.3 GAN 층

GAN 모델은 적대적 신경망 모델로 생성기와 판별기로 이루어져 있다. 생성기는 판별기가 진짜 데이터인지 생성된 데이터인지 구별하지 못하게, 판별기는 데이터가 생성기가 생성한 데이터인지 진짜 데이터인지 판별하도록 적대적으로 학습하게 된다. GAN 모델이 확산 모델만큼 높은 이미지 품질을 보여주지는 않지만, 학습 수렴 속도와 생성 속도가 빨라 많은 데이터 생성에 적절하다고 판단하였다. GAN 층은 pix2pix^[2] 모델 기반의 SPADE^[3] 모델을 사용하여 조건부 적대적 학습한 후, 확산 모델에 의해 생성되고 후처리를 거친 라벨 데이터를 이용하여 라벨에 맞는 이미지를 생성한다.

2.2 실험

2.2.1 실험 결과

학습은 토지 피복지도 항공 이미지 데이터로 건물,

주차장, 도로 등 8개의 클래스로 분할된 512x512 크기의 이미지 12719장을 사용하였다.

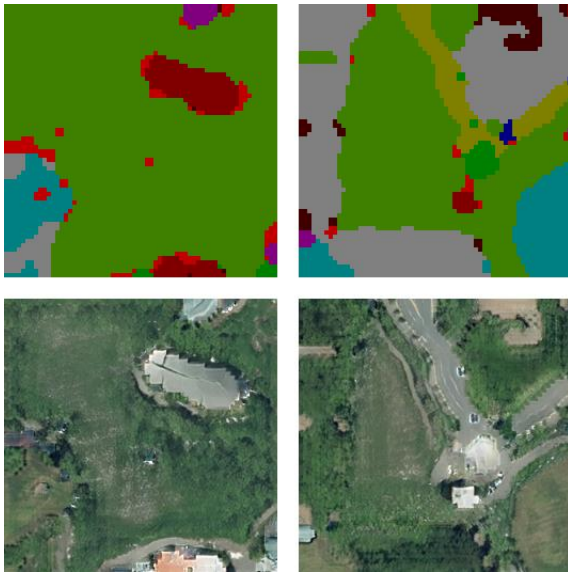


Fig. 1. Generated Label and Image

Fig. 1은 확산 층에서 생성된 라벨 이미지와 그를 통해 GAN 층에서 생성된 이미지이다. 라벨 후처리 층에 의해 경사 에지가 감소된 것은 확인할 수 있었으나 라벨 점은 완전히 제거하지 못하였다.

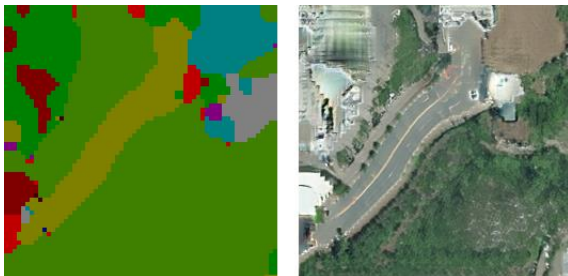


Fig. 2. Limitation of GAN model

Fig. 1과 Fig. 2를 보면 숲이나 도로, 밭은 자연스럽게 표현하는 것을 확인할 수 있으나 건물이나 주차장의 경우는 잘 학습되지 않아, 해당 이미지가 실체가 아닌 생성된 이미지라는 것을 쉽게 확인할 수 있었다.

2.2.2 평가

생성된 이미지에 대한 평가는 데이터셋의 규모를 맞추기 위해 원본 데이터셋에서 일부를 생성 데이터로 대체한 후, 기존의 분할 모델을 통해 학습 성능을 평가하였다. 만약 성능이 크게 감소하지 않는다면 생성된 데이터셋을 분할 모델에서 사용하여도 큰 문제가 없다는 것을 알 수 있다. 다음 표는 원본 이미지 일부(10%)를 생성된 이미지로 대체한 후 같은 모델로 학습하였을 때의 결과이다. 평가는 원본 데이터셋에서 나누어 놓은 테스트셋을 이용하여 진행하였다.

Table 1. mIoU[%] Score Trained from Datasets

Dataset	BiSeNet V2 ^[4]	PPLiteSeg ^[5]
Original	53.53	72.46
Original+ Ours	53.65	72.72

Table. 1을 보면 두 모델 모두 원본 데이터로 학습하였을 때보다 성능이 오차 범위 내로 소폭 증가한 것을 확인할 수 있었다. 따라서 생성된 이미지가 모델의 성능을 감소시키지 않으며 오히려 성능 향상을 기대할 수 있음을 알 수 있다.

3. 결론

본 연구에서는 확산 층을 통해 라벨을 생성하고, 라벨 후처리 층을 통과하여 GAN 층을 통해 라벨링된 이미지 데이터셋을 생성하였다. 또한 생성된 데이터셋을 원본 데이터셋과 섞어 분할 모델을 통해 평가해 보았을 때 비슷한 수준의 성능을 나타내는 것으로 데이터셋으로 충분히 사용 가능하다는 것을 확인하였다. 본 연구를 통해 적은 비용과 시간으로 기존에 사용하던 데이터셋의 양을 증가시켜 이미지 모델의 실질적 성능 향상을 기대할 수 있다.

Acknowledgement

이 논문은 2023년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(NRF-2022R1F1A1073799)

References

- [1] J. Ho, et. al., "Denoising Diffusion Probabilistic Models," Advances in Neural Information Processing Systems 33, pp. 6840-6851, December, 2020.
- [2] P. Isola, et. al., "Image-to-Image Translation with Conditional Adversarial Networks," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1125-1134, July, 2017
- [3] T. Park, et. al., "Semantic Image Synthesis with Spatially-Adaptive Normalization," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2337-2346, June, 2019.
- [4] C. Yu, et. al., "BiSeNet V2: Bilateral Network with Guided Aggregation for Real-time Semantic Segmentation," International Journal of Computer Vision, Vol. 129, No. 11, pp. 3051-3068, 2021.
- [5] J. Peng, et. al., "PP-LiteSeg: A Superior Real-Time Semantic Segmentation Model," arXiv:2204.02681, 2022.