

Semantic segmentation 성능 향상을 위한 병렬 어텐션 모듈

이도희¹, 이진혁², 김동우³, *신용구¹

¹ 한남대학교 AI 융합학과, ² 한남대학교 컴퓨터공학과, ³ 한남대학교 수학과

dlehgm154@gmail.com, jinlee382@gmail.com, mtkdw01@naver.com, *ygshin@hnu.kr

Parallel Dual Attention Module for Improving the Semantic Segmentation Performance

Do-Hee Lee, JinHyeok Lee, DongWoo Kim, and *Yong-Goo Shin
Hannam University

요 약

각 픽셀마다의 클래스를 구분하는 객체 분할 기술 (Semantic Segmentation) 기술은 다양한 분야에서 활발히 이루어지고 있다. 많은 연구자들이 Semantic segmentation 기술의 성능 향상을 위해 네트워크 구조 변화, 학습 기법, 어텐션 (Attention) 기법 등 다양한 방법을 제안한다. 본 논문은 다양한 기법들 중, Semantic Segmentation 성능 향상을 위한 새로운 Attention 모듈을 제안한다. 제안하는 기법은 병렬적으로 Attention 연산을 수행하여, 네트워크의 특성맵 (Feature Map)에서 중요한 부분을 효과적으로 강조한다. 제안하는 기법은 PASCAL VOC(2012) 데이터 세트에서 기존 Attention 기법 대비 IOU에서 0.58% 높은 성능을 보여준다.

1. 서 론

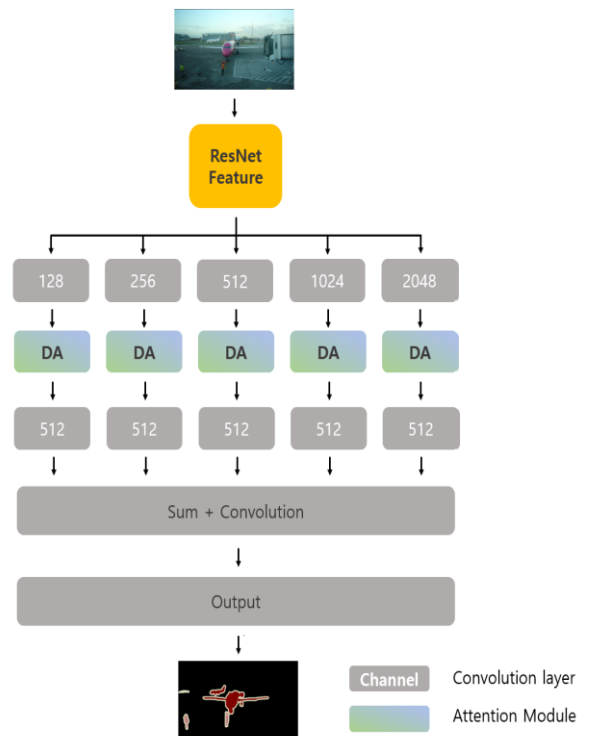
최근 자율주행과 암 진단 등에서 활용되는 시멘틱 세그멘테이션 작업에 다양한 네트워크가[1,2,3,4] 연구되고 있는데 이 중 사전 학습된 네트워크를 통과한 특성맵에 Attention을 적용하는 방법이[4,5,6] 많이 연구되고 있다. 이 중 Fu가 제안한 『Dual Attention Network for Scene Segmentation』[4] (DANet)에서 조금 더 발전된 네트워크를 제안한다.

본 논문에서는 사전 학습된 네트워크를 통하여 얻은 특성맵을 합성곱을 통하여 다양한 채널로 전개하여 각 특성맵에 DANet에서 제안한 Attention 모듈을 사용하여 모델에 복잡성과 비선형성을 추가하여 모델이 더욱 많은 정보를 학습하게끔 수정하여 PASCAL VOC(2012) 데이터 세트의 검증 세트에서 DANet에 비교하여 0.58% 정도 높은 IOU를 기록하였다. 사전 학습된 네트워크를 합성곱을 통해 다양한 채널로 변경하는 과정에서 네트워크에서 얻을 수 있는 특징을 늘리고 늘어난 특징에 Attention을 적용하여 더욱 큰 효과를 얻을 수 있다.

2. 본론

2.1 네트워크 구조

DANet에서 더욱 성능을 개선할 방법으로 Attention을 하나의 특성맵이 아닌 여러 특성맵에 적용하여 더 많은 특징들을 Attention하고 네트워크를 더욱 복잡하게 만들기 위해 Dual Attention을 병렬로 연결한 Parallel Dual Attention Network(PDANet)을 제안한다. [그림 1]과 같이 ResNet에서 나온 특성맵을 (1*1) 합성곱을 통하여 (128, 256, 512, 1024, 2048) 채널로 분리한 특성맵에 Attention을 적용한다. 이후 다시 합성곱을 통하여 채널 크기를 512로 맞추고 더해주며 간단한 합성곱을 적용한 후 출력한다. 기대효과로 다양한 채널에서 더 많은 특징을 Attention하고 네트워크가 깊어짐에 따라 비선형성



[그림 1] PDANet의 구조

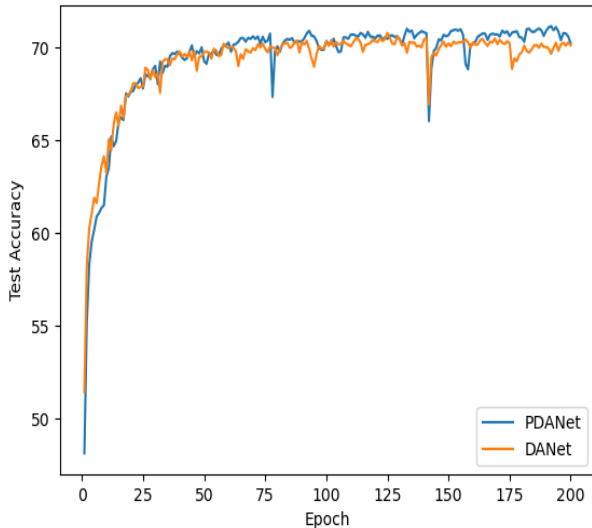
증가로 네트워크가 더 깊어진다.

2.2 실험

네트워크 실험에 앞서 실험의 제한점을 설명하고자 한

| 성능(IoU) | DANet | PDANet |
|-----------|--------|--------|
| 훈련 세트 정확도 | 87.27% | 87.96% |
| 검증 세트 정확도 | 70.56% | 71.14% |

[표 1] PDANet과 DANet의 검증 세트 성능 비교



[그림 2] 에폭 수에 따른 검증 세트 정확도 비교

다. 기존 모델은 Resnet 3번째 블록과 4번째 블록을 Dilated Convolution을 사용하여 입력 이미지를 1/8로 만들었지만 하드웨어 여건상 4번째 블록만 Dilated Convolution을 적용하여 입력 이미지를 1/16으로 실험하였다.

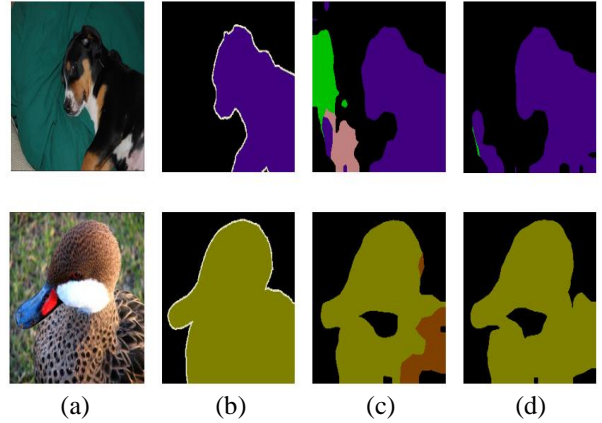
네트워크 학습할 때 DANet과 PDANet의 공정한 비교를 위해 같은 하이퍼 파라미터를 사용하였다. PASCAL VOC(2012) 데이터 셋에서 훈련 데이터 1,464장과 검증 데이터 1,449장에 대하여 미니 배치 학습으로 진행하였으며 입력 이미지 크기는 너비와 높이를 224로 고정하였다. 배치 사이즈는 8로 설정하였고 학습률은 0.00001로 고정하였다. 각각 200 에폭 훈련하였으며 옵티마이저로는 Adam 옵티마이저를 사용하였다. 안정적인 학습을 위해 모든 합성곱 신경망 뒤에 Batch Normalization[7]을 사용하였다.

2.3 실험 결과

200 에폭 훈련하였을 때 검증 세트에 대하여 [표 1]을 참고하였을 때 DANet이 70.56%의 IoU를 기록했고 PDANet이 71.14%의 IOU를 기록하여 0.58%정도 높은 IOU를 기록하였다. [그림 2]에서는 PDANet이 검증 세트에 대하여 더 잘 학습하는 것을 볼 수 있다. [그림 3]에서 PDANet가 실제 타겟을 다른 클래스로 오인하는 일이 줄어들어 네트워크 성능이 좋아진 것을 시각적으로 확인할 수 있다.

3. 결론

본 논문에서는 새로운 Attention (Parallel Dual Attention) 기법을 제안하였다. 제안하는 기법은 특성맵을 여러 채널로 분리하여 Attention 연산을 수행하였고, 실험결과에서 볼 수 있듯이, 기존 Attention 기술 대비 좋은 성능을 보이는 것을 알 수 있었다. 향후 더 다양한 데이터 셋과 Base Network에서 제안하는 기법의 우수성을 입증할 계획이다.



[그림 3] 결과 영상 비교. (a) 입력 영상 (b) 정답 영상 (c) DANet (d) PDANet

ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No.2022R1G1A1004001)

참고 문헌

- [1] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation." *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015.
- [2] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *International Conference on Medical image computing and computer-assisted intervention*, 2015.
- [3] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network." *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017.
- [4] J. Fu, J. Liu, H. Tian, Y. Li, Y. Bao, Z. Fang, and H. Lu, "Dual attention network for scene segmentation." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019.
- [5] Z. Lin, M. Feng, C. Nogueira dos Santos, M. Yu, B. Xiang, B. Zhou, and Y. Bengio, "A structured self-attentive sentence embedding." *arXiv preprint arXiv:1703.03130* (2017).
- [6] T. Shen, T. Zhou, G. Long, J. Jiang, S. Pan, and C. Zhang, "Disan: Directional self-attention network for rnn/cnn-free language understanding." *Proceedings of the AAAI conference on artificial intelligence. Vol. 32. No. 1*, 2018.
- [7] S. Ioffe, and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift." *International conference on machine learning. PMLR*, 2015.