DeepLabv3+ 네트워크의 경량화 기법

김태준[1], 정인수[1], 강주완[1], 조승준[1], 문병인[2]
[1]경북대학교 대학원 전자전기공학부 석사과정
[2]경북대학교 전자공학부/대학원 전자전기공학부 교수

ktj016669@knu.ac.kr, insu8944@knu.ac.kr, jwkang1231@knu.ac.kr, jo73012@knu.ac.kr, bihmoon@knu.ac.kr

A Lightweighting Method for the DeepLabv3+ Network

Taejun Kim[1], Insu Jeong[1], Joowan Kang[1], Seungjun Jo[1], Byungin Moon[1, 2]
[1]Graduate School of Electronic and Electrical Engineering, Kyungpook National University
[2]School of Electronics Engineering, Kyungpook National University

요 약

최근 딥러닝 기반 영상 분할 방법은 정확도 측면에서 높은 성과를 보이고 있지만, 방대한 연산량과 복잡한 구조로 인하여 임베디드 플랫폼에서의 실시간 처리 및 구현에는 한계가 있다. 이러한 문제를 해결하기 위해 MobileNet 과 같은 경량 신경망을 활용하는 등의 다양한 시도가 이어지고 있으나, 하드웨어 자원이 제한된 환경에서는 신경망이 가지는 많은 매개변수로 인해 구현에 어려움이 있다. 이에 본 논문에서는 DeepLabv3+에 MobileNetV3-Large 를 백본 신경망으로 사용하면서, 매개변수와 연산량을 줄이기 위한 방법을 제안한다. 제안하는 방법을 통해 구현된 모델은 기존 모델에 비해 성능이 소폭 감소하였으나, 매개변수를 절반 이상 절감함으로써 자원 제약 환경에서의 효율성을 높일 수 있음을 확인하였다.

1. 서론

최근 자율 주행, 로봇 비전의 발전 등에 따라 영상 분할 기술의 수요가 크게 증가하고 있으며, 실시간으로 영상 분할을 수행할 수 있는 방법에 대한 연구가 활발히 진행되고 있다. 딥러닝 기반 영상 분할 방법은 정확도 측면에서 높은 지표를 보이지만, ANN(artificial neural network)의 복잡한 구조와 방대한 연산 비용으로 인해 실시간성을 확보하기 어려우며, 특히 자원이 제한된 임베디드 플랫폼 적용하기에 많은 제약이 따른다.

영상 분할을 위한 신경망은 대부분 인코더-디코더 구조로 이루어져 있다. 인코더(encoder)에서는 기존의 CNN(convolutional neural network) 모델을 백본(backbone) 신경망으로 사용하여 특징(feature)을 추출하고, 이 특징을 기반으로 디코더(decoder)에서 분할 지도(segmentation map)를 만든다. 이러한 구조적 특성으로 인해, 영상 분할 모델의 경량화는 백본 신경망으로 사용되는 CNN 모델의 경량화를 통해 수행할 수 있다. 이에 따라 영상 분할 신경망의 백본 신경망에 MobileNet 과 같은 경량 신경망을 사용하여 임베디드 플랫폼에 구현하는 방법에 대한 관심이 높아지고 있다. MobileNet 은 2017 년에 Google 에서 제안한 경량 CNN 모델로, depthwise separable convolution 을 기반으로 설계하여 매개변수(parameter)와 연산량을 크게 줄였다[1]. 2018 년에 제안된 MobileNetV2 는 inverted residual block 을 도입함으로써 기존의 residual block 과 반대로 비선형 함수에 의한 정보 손실을 줄이고, 채널 확장을 통해 특징의 표현력을 향상시켰다. 또한 적은 채널의 압축된 정보만 저장하면 되므로 메모리 비용도 줄일 수 있었다[2]. 이후 2019 년에 제안된 MobileNetV3 는 기존의 depthwise separable convolution 과 inverted residual block 을 그대로 채택하면서, 계산 비용이 많이 드는 계층에 대해서 최적화를 수행하여 경량화와 성능 개선을 동시에 달성하였다[3].

MobileNet 은 모바일 기기에서도 동작할 수 있도록 설계된 경량 신경망이지만, 여전히 매개변수와 연산량이 많기 때문에 자원 제약적 환경에 구현하기 어려우며 이를 해결하기 위한 노력이 이어지고 있다[4]. 이에 본 논문에서는 매개변수와 연산량을 줄이기 위해 MobileNetV3-Large 의 마지막 확장 계층(expansion layer)을 생략하여 DeepLabv3+의 백본 신경망으로 결합함으로써 매개변수를 절반 이상 절감하였으며, 확장 계층 생략에 따른 성능을 분석하였다.

2. 제안하는 방법

본 논문에서는 DeepLabv3+에 MobileNetV3-Large 를 백본 신경망으로 사용하면서 매개변수와 연산량을 줄일 수 있는 방법을 제안한다.

기존의 영상 분할 신경망은 Xception, ResNet 과 같은 높은 정확도의 CNN 모델들을 백본 신경망으로 사용하는 경우가 많지만, 이는 매개변수와 연산량이 많기 때문에 임베디드 플랫폼에 적합하지 않다. 따라서 MobileNet 과 같은 경량 신경망 사용이 필수적이다.

MobileNetV3 는 특징의 표현력을 높이기 위해 특징 맵(feature map)의 채널을 늘리는 확장 계층이 있다. 영상 분할, 분류 문제에서 특징 맵의 채널 수를 늘릴수록 특징의 표현력이 좋아지며, 일반적으로 DeepLabv3+의 백본 신경망으로 사용될 때도 이 확장 계층이 모두 포함된다. 그 결과, DeepLabv3+는 ASPP(atrous spatial pyramid pooling) 구조에서 더 풍부한 의미론적 정보(semantic information)를 얻을 수 있다. 하지만 ASPP 는 입력 특징 맵에 대해 4 개의 convolution 블록과 1 개의 GAP(global average pooling) 연산을 수행하므로 입력되는 채널 수가 많아질수록 연산 비용이 증가하여 자원 제약 환경에는 부적합할 수 있다.

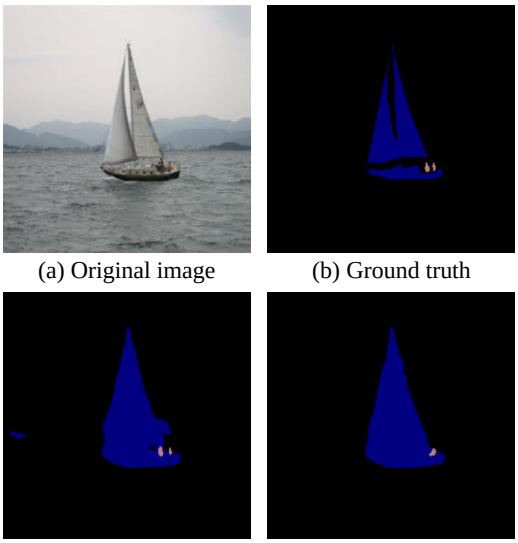
이에 본 논문에서는 ASPP 에 입력되는 채널 수를 줄이기 위해 MobileNetV3-Large 의 마지막 확장 계층을 생략하였으며, 확장 계층을 생략한 모델을 DeepLabv3+의 백본 신경망으로 결합함으로써 생략하지 않은 기존 모델에 비해 매개변수를 절반 이상 절감하였다.

3. 실험 및 결과

본 논문에서 제안하는 방법으로 구현된 DeepLabv3+와 기존 방법(conventional method)으로 구현된 DeepLabv3+의 매개변수와 mIoU 는 표 1 과 같다. 두 모델의 성능을 비교하기 위해 MS COCO 데이터세트(dataset)로 훈련하였으며, Pascal VOC 2012 데이터세트로 검증을 진행하였다. 표 1 에서 볼 수 있듯이 MobileNetV3-Large 의 마지막 확장 계층이 포함된 기존 모델의 경우 매개변수가 11,745,643 개이고, 생략된 모델은 5,632,933 개로, 마지막 확장 계층 생략을 통해 매개변수의 수를 약 52% 절감하였다. 또한 두 모델의 성능을 검증한 결과를 보면, 마지막 확장 계층을 생략한 모델의 mIoU 는 기존 모델에 비해 2.7%p 감소하였다. 비록 약간의 성능 저하가 발생하였으나, 이는 경량 모델이 요구되는 임베디드 시스템이나 모바일 환경에서 실용적인 대안이 될 수 있다.

<표 1> 기존 방법과 제안하는 방법으로 구현된 모델 비교

Model	Parameters	mIoU (%)
Conventional method	11,745,643	62.24
Proposed method	5,632,933	59.54



(a) Original image (b) Ground truth

(c) Conventional method (d) Proposed method

(그림 1) 기존 방법과 제안하는 방법으로 구현된 모델의 추론 결과

그림 1 은 두 모델의 추론(inference) 결과를 보여준다. 그림 1.(a)는 원본 이미지, 그림 1.(b)는 정답 이미지(ground truth), 그림 1.(c)는 마지막 확장 계층을 포함하는 기존 방법으로 구현된 모델의 추론 결과, 그림 1.(d)는 제안하는 방법으로 구현된 모델의 추론 결과이다.

4. 결론

본 논문에서는 자원 제약 환경에서도 DeepLabv3+를 구현할 수 있도록 MobileNetV3-Large 를 백본 신경망으로 사용하면서 매개변수를 절반 이상 절감할 수 있는 방법을 제안하고, 성능을 확인하였다. MobileNetV3 의 특징인 확장 계층 중 마지막 확장 계층을 포함하지 않는 것으로 매개변수의 수를 52% 절감하면서도 성능은 2.7%p 만 감소하여 자원 제약 환경에서 실용적인 대안이 될 수 있음을 확인하였다.

ACKNOWLEDGMENT

이 논문은 2025 년도 정부(산업통상자원부)의 재원으로 한국산업기술진흥원의 지원을 받아 수행된 연구임(RS-2024-00415938, 2024 년 산업혁신인재성장지원사업)

이 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단-시스템반도체융합전문인력육성사업의 지원을 받아 수행된 연구임(RS-2020-NR047144)

참고문헌

[1] Howard, Andrew G., et al. "Mobilenets: Efficient convolutional neural networks for mobile vision applications." *arXiv preprint arXiv:1704.04861* (2017).

[2] Sandler, Mark, et al. "Mobilenetv2: Inverted residuals and linear bottlenecks." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018.

[3] Howard, Andrew, et al. "Searching for mobilenetv3." *Proceedings of the IEEE/CVF international conference on computer vision*. 2019.

[4] Morì, Pierpaolo, et al. "Accelerating and pruning cnns for semantic segmentation on fpga." *Proceedings of the 59th ACM/IEEE Design Automation Conference*. 2022.

[5] Chen, Liang-Chieh, et al. "Encoder-decoder with atrous separable convolution for semantic image segmentation." *Proceedings of the European conference on computer vision (ECCV)*. 2018.