RepBlock을 활용한 의미적 분할 모델 성능 및 실시간 처리 속도 향상 연구

박성준, 이하림*

국립금오공과대학교

darkjyuk@kumoh.ac.kr, *hrlee@kumoh.ac.kr

Improvement of Performance and Processing Speed in Semantic Segmentation Network Models Using RepBlock

Park Sungjun, Lee Harim*

Kumoh National Institute of Technology

요 약

자율주행 기술은 미래 사회를 이끌어가는 핵심 기술로서 인지, 제어, 판단 등으로 구성되며, 그중에서도 주변 환경을 정확히 인지하여 적절한 주행 결정을 내리는 인지 기술이 특히 중요하다. 본 연구에서는 기존 DeepLabV3+ 모델의 백본 네트워크인 ResNet의 콘볼루션 계층을 RepBlock으로 대체하여 모델을 설계 및 학습 하였다. 본 연구의 RepBlock을 활용한 ResNet50을 백본으로 하는 DeepLabV3+의 경우 기존 모델 대비 mAP는 62.23%에서 62.54%로 약 0.5%, mean ioU는 53.34%에서 53.41%로 약 0.13% 증가하였으며, 임베디드 시스템에서의 추론 속도 또한 19FPS에서 22FPS로 향상되었다. 또한, RepBlock 기반 ResNet101을 백본으로 사용할 경우, mAP는 62.34%에서 62.51%로 약 0.27%, mean ioU는 53.41%에서 53.85%로 약 0.82% 증가하였고, 추론 속도는 14 FPS에서 15 FPS로 개선되었다. 이를 통해 RepBlock을 활용하는 것은 인지 성능 및 실시간 처리을 동시에 향상할 수 있음을 검증하였다.

Ⅰ. 서 론

자율주행 기술은 인지, 제어, 판단 등으로 구성된 미래 사회의 핵심 기술중 하나로, 특히 주변 환경을 정확하게 인지하고 신속하게 판단하여 안정적인 주행 결정을 내리는 것이 중요하다. 이를 해결하기 위해 높은 정확도와 빠른 추론 속도를 동시에 갖춘 인지 모델이 필수적이다. 자율 주행의 인지 기술은 크게 물체 탐지(object detection)와 의미적 분할(semantic segmentation)과 같은 딥러닝 기반 기술들이 존재한다.

DeepLabV3+ 모델은 의미론적 분할 분야에서 가장 많이 활용되고 있으며 ASPP(Atrous Spatial Pyramid Pooling) 구조와 스킵 연결(Skip Connection)을 활용하여 다양한 크기의 정보를 효과적으로 포착한다 [1]. 해당 모델의 백본으로는 가장 많이 활용되는 ResNet을 사용하며 여기에 추가적으로 dilated 콘볼루션 계층을 적용하여 다양한 크기 정보를 추출한다.

해당 모델과 같이 추가적인 구조 및 계층을 활용하는 것은 의미적 분할 성능을 향상 시킬 수 있지만 추론 속도는 저하 시킬 수 있다. 하지만 실시간 자율주행의 경우 성능 뿐만 아니라 추론 속도 향상 역시 상황 인지 및 돌발 상황 대처에 있어 중요한 요소 중 하나이다. 따라서, 추론 속도 및 의미적 분할 성능을 동시에 향상시킬 수 있는 연구가 필요하다.

본 연구에서는 의미적 분할 모델의 성능 및 추론 속도를 향상시키는 것을 목표로 하며, 이를 위해 RepBlock기반 DeepLabV3+ 모델을 제안한다. 해당 모델은 DeepLabV3+ 모델의 백본 네트워크인 ResNet50과 ResNet101의 콘볼루션 계층에 RepBlock 기법을 적용 설계된다 [2, 3]. RepBlock의 경우 학습시에는 다중 브랜치 구조를 통한 잔차 학습을 적용하고 추론 시에는 단일 브랜치 구조로 변경하여 기존 컨볼루션 계층과 동일하게 동작한다. 따라서 성능 및 추론 속도를 동시에 향상시킬 수 있으며, 실험 결과에서도 알 수 있듯이 제안된 RepBlock 기반 DeepLabV3+ 모델은 기존 DeepLabV3+ 대비 의미적 분할 정확도와 임베디드 시스템에서의 초당 프레임 처리 속도을 동시에 향상시켰다.

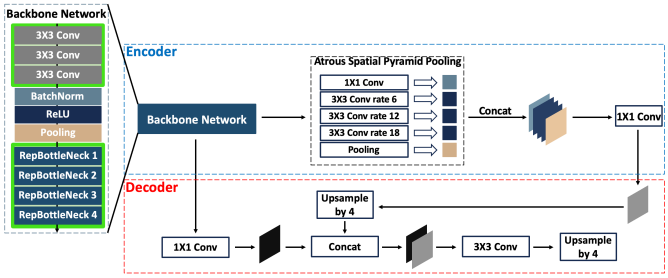

그림 1. RepBlock기반 DeepLabV3+ 구조.

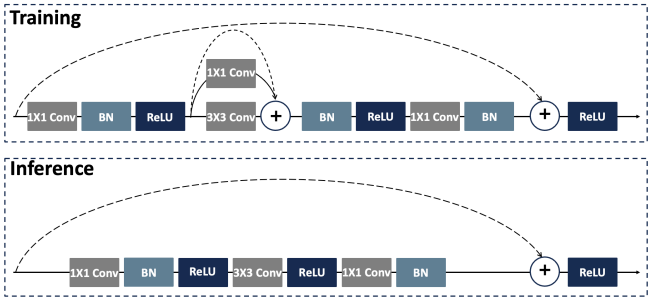

그림 2. RepBottleNeck 블럭의 학습 및 추론 구조.

Ⅱ. 분할 딥러닝 네트워크 기반 도로 상황 인지 임베디드 시스템

2.1 RepBlock기반 DeepLabV3+ 네트워크

의미적 분할 정확도와 실시간 처리 성능 향상을 목표로 설계된 RepBlock기반 DeepLabV3+ 네트워크 구조의 백본은 그림 1의 왼쪽의 Backbone Network 부분의 초록색 박스에서 볼 수 있듯이 3개의 RepBlock이 적용된 콘볼루션 계층들과 RepBlock구조가 적용된 RepBottleNeck 블럭들로 구성된다. RepBlock이 적용된 부분은 다중 브랜치 구조를 통해 풍부한 표현력을 확보하여 정확도를 향상 시키고 추론 시에는 단일 브랜치 재구성을 통해 모델의 성능은 유지하면서도 추론 속도를 개선할 수 있다.

Model	backbone	Image size	mAP	mioU	macro F1-score
FCN8s	-	256	62.00%	50.64%	59.20%
FCN16s	-	256	57.67%	47.56%	58.36%
FCN32s	-	256	51.19%	39.76%	57.12%
RepBlockFCN8s	-	256	63.81%	48.59%	63.72%
Pretrained RepBlockFCN8s	-	256	61.49%	48.07%	64.38%
DeepLabV3+	ResNet50	256	62.23%	53.34%	60.76%
	ResNet101	256	62.34%	53.41%	61.82%
RepBlock DeepLabV3+	ResNet50	256	62.54%	53.41%	58.69%
	ResNet101	256	62.51%	53.85%	61.25%

표 1. 의미적 분할 모델의 성능.

Model	backbone	Image size	FPS	
			FP32	FP16
FCN8s	-	256	8	12
FCN16s	-	256	8	12
FCN32s	-	256	5	12
RepBlockFCN8s	-	256	8	10
Pretrained RepBlockFCN8s	-	256	8	10
DeepLabV3+	ResNet50	256	19	21
	ResNet101	256	14	14
RepBlock DeepLabV3+	ResNet50	256	19	22
	ResNet101	256	14	15

표 2. Jetson Orin NX에서 초당 프레임 처리 속도.

그림 2는 RepBottleNeck 블럭의 내부 구조를 보여준다. 해당 블럭은 학습시 다중 브랜치 구조가 적용되며 추론시에는 3x3 콘볼루션, 1x1 콘볼루션, BatchNorm, 그리고 identity를 3x3 콘볼루션 단일 브랜치로 재구성함으로써 성능을 유지하면서도 연산을 최적화할 수 있도록 설계되었다.

2.2 자율 주행 인지 시스템을 위한 임베디드 시스템 구성

본 연구에서는 자율 주행 인지 시스템을 위한 임베디드 시스템을 카메라, NVIDIA Jetson Orin NX, 디스플레이 장치로 구성하였다. 시스템의 각 요소는 Robot Operating System(ROS)를 통해 상호 연결된다. 시스템은 카메라로 촬영한 영상을 의미적 분할 네트워크가 탑재된 Jetson Orin NX로 전송한다. Jetson Orin NX는 의미적 분할 네트워크를 이용해 영상의 각 픽셀에 해당 클래스 레이블을 할당하고, 이를 바탕으로 입력 영상과 분할 결과를 결합한 출력 이미지를 생성한다. 생성된 출력 이미지는 디스플레이 장치를 통해 실시간으로 시각화된다.

Ⅲ. 성능 평가 및 임베디드 시스템 구현 실증

표 1은 기존 연구에서 사용된 FCN 모델과 RepBlock을 적용시킨 RepBlockFCN8s, FCN8s의 가중치를 사용하여 학습에 적용한 Pretrained RepBlockFCN8s, DeepLabV3+ 그리고 RepBlock 기반 DeepLabV3+ 모델의 성능을 보여준다 [4]. 성능 지표로는 mAP, mean ioU, 그리고 Macro F1-score를 사용하였다.

제안한 모델은 RepBlock의 다중 브랜치 구조를 적용함으로써 기존 모델 대비 성능이 전반적인 성능 향상을 보였다. RepBlockFCN8의 모델은 기존 FCN 계열 모델에 비해 mAP와 Macro F1-score에서 향상된 성능을 기록하였으며, RepBlock 기반 DeepLabV3+ 모델 또한 기존 DeepLabV3+ 모델 대비 mAP와 mean ioU가 향상되는 결과를 얻었다.

표 2는 표1에서 비교 분석한 모델들을 2.2절의 임베디드 시스템에 적용하고 FP32 및 FP16 정밀도 환경에서 측정한 초당 프레임 처리 속도를 보여준다. 실험 결과, FP32 대비 FP16 정밀도로 변경하였을 때 모든 모델에서 초당 프레임 처리 속도가 증가한 것을 확인하였다. RepBlock 기반 DeepLabV3+ 모델은 FP32 대비 FP16 환경에서 가장 높은 초당 프레임 처리 속도를 보였다.

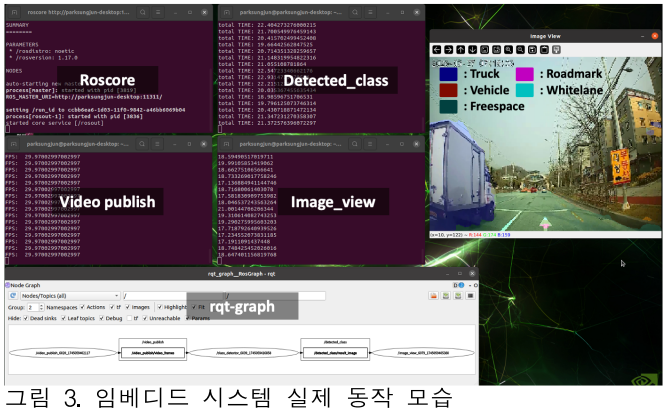

그림 3. 임베디드 시스템 실제 동작 모습

그림 3은 본 연구에서 구성한 자율 주행 인지 시스템에 RepBlock기반 DeepLabV3+를 탑재하여 실증하는 장면이다. 터미널들은 ROS의 각 노드 실행을 의미하며 왼쪽 하단은 노드간의 연결을 보여준다. 실증은 국립금오공과대학교 주변을 주행하여 얻은 블랙박스 영상으로 진행하였다. 그림 3의 우측 상단은 실증 영상의 한 장면을 보여주며 제안한 모델을 통해 차량 주변의 상황을 실시간으로 인지할 수 있음을 보여 준다. 해당 장면에서 분할되지 않은 영역은 학습데이터에는 없는 물체로써 추후 추가적인 데이터로 추가 학습하여 인지 범위를 더 확장하는 추가 연구를 진행할 계획이다.

Ⅳ. 결론

제안한 RepBlock 기반 DeepLabV3+모델은 제안한 임베디드 시스템에서 비교 모델 중 가장 높은 의미적 분할 성능 및 초당 프레임 처리 속도를 달성하였다. 해당 결과는 RepBlock을 적용하는 것은 기존 모델 대비 분할 성능 향상과 처리 속도 향상을 달성할 수 있음을 보여준다.

ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT: Ministry of Science and ICT) RS-2025-00557827.

참 고 문 헌

[1] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation,"in Proc. Eur. Conf. Comput. Vis. (ECCV), vol. 7, pp. 833-851. 2018.

[2] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun “Deep Rsidual Learning for Image Recognition”, CoRR vol. abs/1512.03385, 2015.

[3] X. Ding, X. Zhang, N. Ma, J. Han, G. Ding, and J. Sun, "RepVGG: Making VGG-Style ConvNets Great Again,"in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 13733-13742. 2021.

[4] Evan Shelhamer, Jonathan Long, and Trevor Darrell “Fully Convolutional Networks for Semantic Segmentation” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.39, no.4, pp. 640-651, 2017.