트랜스포머를 활용한 차선 인식에서의 전역 및 지역 정보 통합 연구

이현종, 이재구*
국민대학교 일반대학원 컴퓨터공학과

*jaekoo@kookmin.ac.kr

Integration of Global and Local Information for Lane Detection Using Transformers

Hyunjong Lee, Jaekoo Lee*
College of Computer Science, Kookmin University

요 약

차선 인식은 도로 이미지에서 차선의 위치를 정확하게 분할하는 과업이다. 차선은 도로 이미지에서 길고 얇게 분포되어 있으며 수직 및 수평 관점에서 각각 다른 특성을 보인다. 하지만 이러한 특성을 효과적으로 모델링하는 연구가 부족한 상황이다. 따라서 본 논문에서는 차선 특징을 효과적으로 포착할 수 있는 트랜스포머 기반의 차선 인식 모델을 제안한다. 제안한 방법으로 Tusimple 데이터 집합에서 평가한 결과, 기존 방법보다 0.25% 더 높은 정확도를 얻을 수 있었다. 결과적으로 차선의 특징을 모델링하는 연구가 유효함을 입증했다.

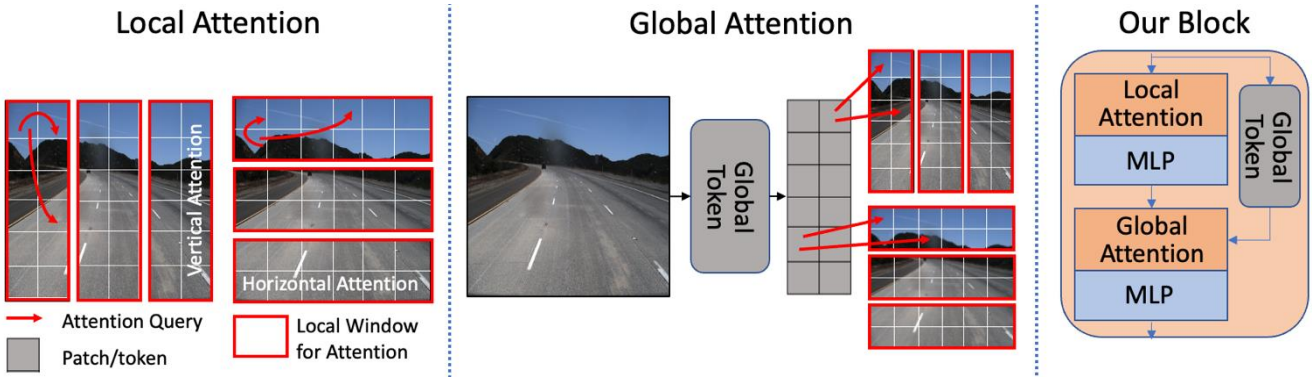

그림 1. 제안 어텐션 모듈

Ⅰ. 서론

차선 인식은 도로 이미지에서 차선의 위치를 분할하는 과업이다. 기존 차선 인식 모델은 합성곱 신경망(Convolutional Neural Network) 기반으로 활발히 연구되고 있다[1]. 하지만 합성곱 신경망은 이미지의 지역적인 정보를 추출하는 데 특화되어, 이미지 전반에 걸친 차선의 특징을 추출하는 데는 한계점을 가지고 있다.

최근에는 이러한 한계점을 극복하기 위해 이미지의 전역적인 정보를 추출할 수 있는 트랜스포머(Transformer) 기반 모델이 연구되고 있다[2]. 그러나 트랜스포머 기반 방식은 전역적 정보 추출에 특화된 대신, 지역적인 정보를 추출하는 능력이 부족하다. 따라서 이미지 내에서 차선을 정밀하게 예측하기 어려운 한계점이 있다.

본 논문에서는 이미지의 차선의 정보를 효율적으로 추출할 수 있는 트랜스포머 기반의 차선 인식 모델을 제안한다. 구체적으로 수직 및 수평 어텐션(Attention)을 사용하여 차선의 특징을 강화하였으며, 전역적인 토큰(Global Token)을 사용하여 이미지의 전역적인 정보를 추출했다. 또한 다양한 트랜스포머 기반 모델을 평가하여 제안 방식의 우수한 성능을 확인하였다. 결론적으로 차선의 특징을 종합하는 방식이 유의미한 방식임을 확인할 수 있었다.

Ⅱ. 본론

수평 차선은 이미지 내에서 연속적이며 길게 얇게 분포되어 있다. 이러한 특징으로 인해 트랜스포머 기반 모델의 어텐션 연산으로 이미지 전체에 희소하게 배치된 차선의 중요한 특징을 추출하기가 어렵다. 따라서 차선에 특화된 어텐션 연산이 요구된다.

본 논문에서는 수직 및 관점에서 차선의 특징을 분석하여 차선에 특화된 방법론을 제안한다. 수직 관점 분석에서는 차선이 길고 얇기 때문에 각 차선 픽셀(Pixel)이 서로 근접하여 배치된 특성을 고려할 수 있고, 수평 관점 분석은 서로 다른 차선 간의 관계를 파악하는 것이 용이하다[3]. 따라서 그림 1 에서 보이듯 긴 띠 형태의 수직 및 수평 어텐션을 적용하여 차선의 특징을 보강하여 추출한다. 추가로 효율성을 고려하여 동일한 특징 맵(Feature Map) 내에서 채널을 분할하여 수직 및 수평 어텐션을 적용하고 이를 합치는 전략을 사용한다[4].

하지만 수직 및 수평 어텐션은 차선이 가려진 경우 차선의 특징을 추출하기 어려운 차선 폐색(Occlusion) 문제가 있다. 차선이 폐색된 경우 이미지의 전역적인 정보를 추출하여 차선을 예측하는 작업이 필요하다. 따라서 본 논문에서는 이미지의 전체 정보를 주입할 수 있는 전역적인 토큰을 추가적으로 활용하는 것을 제안한다[5].

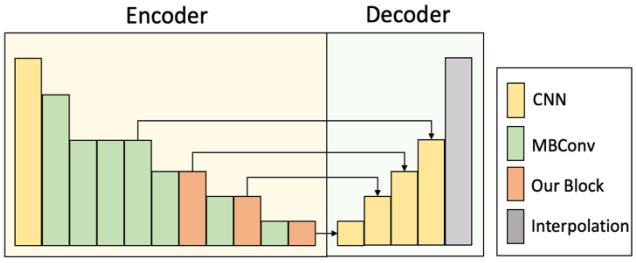

그림 2. 제안 모듈 개요

전역적인 토큰은 그림 1 에서 보이듯 이미지의 전체 정보를 대표하는 토큰이다. 이를 생성하기 위해 합성곱 신경망에 특징 맵을 통과시켜 하나의 토큰으로 재구성한다. 이후 토큰을 복제하여 원본 특징 맵의 크기로 만든 후 어텐션 연산을 수행한다. 전역적인 토큰을 어텐션 연산에 활용하여 모델이 차선 정보와 이미지의 전역적인 정보를 효과적으로 결합하여 추출할 수 있게 된다. 차선을 정밀하게 분할하기 위해서는 차선 정보 외에도 윤곽선, 색상과 같은 저수준의 시각적 정보도 필요하다. 따라서 본 논문에서는 이러한 다양한 정보를 함께 활용하기 위해 그림 2 에서 보이는 UNet 구조를 채택하여 Decoder 를 구성했다[6].

Ⅲ. 실험

본 논문에서는 트랜스포머 기반 모델을 평가하기 위해 Tusimple[7] 데이터 집합을 사용했다. Tusimple 데이터 집합은 도로 이미지에서 차선을 검출하기 위해 구성된 데이터 집합이다. 학습 이미지는 총 3,626 장, 검증 이미지는 총 358 장 테스트 이미지는 2,782 장이다.

실험 과정에서 정확도, 거짓 양성(False Positive), 거짓 음성(False Negative)을 지표로 차선 인식 모델의 성능을 평가했다. 실험은 기본적으로 UNet 구조를 기반으로 인코더(Encoder)부분을 MobileViT 프레임워크로 대체하여 어텐션 영역을 변경하며 수행했다. 사용한 Attention 영역으로는 Full Attention[8], MobileViT[9], Swin Transformer[10], GC-ViT[5], Cswin Transformer[4], 그리고 제안한 방식으로 구성했다.

표 1 을 살펴보면 CsWin Transformer 와 Swin Transformer 간에는 차선 인식 모델의 성능 차이가 명확하게 드러났다. 이 두 모델은 모두 트랜스포머를 활용하여 지역 정보 추출 능력을 강화한 모델이다. 따라서 차선의 수직 및 수평 관계를 모델링하여 어텐션을 수행하는 방식이 차선 인식에 적합함을 보여준다.

또한, Full Attention 과 MobileViT 를 사용하여 이미지의 전체 영역에 연산을 수행한 결과, Swin Transformer 보다 더 높은 성능을 보이지만, Cswin Transformer 보다는 성능이 낮았다. 이는 차선 인식 모델에서는 지역 정보보다는 전역 정보 추출의 중요성이 강조되나, 차선의 특징을 잘 파악하기 위해서는 지역 정보를 적절하게 반영하는 것이 필요하다는 것을 보여준다.

또한, Swin Transformer 에 전역적인 토큰을 통합한 GC-ViT 모델은 Cswin Transformer 와 유사한 성능을 보였다. 즉 지역적인 트랜스포머 모델에 전역 정보를 결합하는 방식이 차선의 전체적 특징과 정밀한 위치 예측을 고려하는데 효과적임을 확인했다.

이러한 실험 결과를 기반으로 본 논문에서는 차선에 특화된 수직 및 수평 어텐션 연산에 전역적인 토큰을 통합하여 전역 및 차선의 정보를 효과적으로 추출하도록 모델을 개선했다. 따라서 다른 트랜스포머 기반 방식보다 차선 인식에서 좋은 성능을 달성했다.

표 1. Tusimple 실험 결과

Metric / Method	Accuracy ↑	FP ↓	FN ↓
Full Attention	96.37	0.0590	0.0300
MobileViT	96.37	**0.0480**	0.0337
CsWin Trans	96.44	0.0586	0.0287
Swin Trans	96.27	0.0602	0.0346
GC-ViT	96.41	0.0483	0.0320
Ours	**96.62**	0.0513	**0.0281**

IV. 결론

본 연구에서는 트랜스포머 기반 차선 인식 모델을 새롭게 제안했다. 제안한 모델은 이미지 내에서 차선의 지역적 특징과 전역적인 특징을 동시에 잘 추출할 수 있는 특성이 있다. 더불어, 다양한 트랜스포머 기반 모델들과 비교 평가하여 제안한 모델의 우수성을 입증했다.

ACKNOWLEDGMENT

이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No.RS-2023-00212484, 복잡한 실제 주행환경에서 설명 가능한 움직임 예측).

참 고 문 헌

[1] Tang, Jigang, Songbin Li, and Peng Liu. "A review of lane detection methods based on deep learning." Pattern Recognition 111 (2021): 107623.

[2] Zhang, Han, et al. "Lane Detection Transformer Based on Multi-frame Horizontal and Vertical Attention and Visual Transformer Module." European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022.

[3] Han, Jianhua, et al. "Laneformer: Object-aware row-column transformers for lane detection." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 36. No. 1. 2022.

[4] Dong, Xiaoyi, et al. "Cswin transformer: A general vision transformer backbone with cross-shaped windows." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022.

[5] Hatamizadeh, Ali, et al. "Global context vision transformers." International Conference on Machine Learning. PMLR, 2023.

[6] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." Medical Image Computing and Computer-Assisted Intervention– MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Springer International Publishing, 2015.

[7] TuSimple lane challange, https://github.com/TuSimple/tusimple-benchmark

[8] Dosovitskiy, Alexey, et al. "An image is worth 16x16 words: Transformers for image recognition at scale." arXiv preprint arXiv:2010.11929 (2020).

[9] Mehta, S., & Rastegari, M. (2021). Mobilevit: light-weight, general-purpose, and mobile-friendly vision transformer. arXiv preprint arXiv:2110.02178.

[10] Liu, Ze, et al. "Swin transformer: Hierarchical vision transformer using shifted windows." Proceedings of the IEEE/CVF international conference on computer vision. 2021.