

신호등 인식 성능 향상을 위한 쿠버네티스 기반의 프레임워크: YOLOv5와 Visual Attention을 적용한 C-RNN의 융합 Vision AI 시스템

조형서¹, 이민정², 한연지[†]

¹명지대학교 융합소프트웨어학부, ²중원대학교 컴퓨터공학과,

[†]한양대학교 컴퓨터·소프트웨어학과

hyoungseocho@gmail.com, minchocopie.lee@gmail.com, yeon9891@naver.com

Kubernetes-based Framework for Improving Traffic Light Recognition Performance: Convergence Vision AI System based on YOLOv5 and C-RNN with Visual Attention

Hyoung-Seo Cho¹, Min-Jung Lee², Yeon-Jee Han[†]

¹Dept. of Convergence Software, Myongji University

²Dept. of Computer Engineering, Jung-Won University

[†]Dept. of Computer Science, Hanyang University

요 약

고령화로 인해 65세 이상 운전자가 급증하며 고령운전자의 교통사고 비율이 증가함에 따라 시급한 사회 문제로 떠오르고 있다. 이에 본 연구에서는 객체 검출, 인식 모델을 결합하고 신호등을 인식하여 Text-To-Speech(TTS)로 알리는 쿠버네티스 기반의 프레임워크를 제안한다. 객체 검출 단계에서는 YOLOv5 모델들의 성능을 비교하여 활용하였으며 객체 인식 단계에서는 C-RNN 기반의 attention-OCR 모델을 활용하였다. 이는 신호등의 내부 LED 영역이 아닌 이미지 전체를 인식하는 방식으로 오탐지 요소를 낮춰 인식률을 높였다. 결과적으로 1,628장의 테스트 데이터에서 accuracy 0.997, F1-score 0.991의 성능 평가를 얻어 제안한 프레임워크의 타당성을 입증하였다. 본 연구는 후속 연구에서 특정 도메인에 딥러닝 모델을 한정하지 않고 다양한 분야의 모델을 접목할 수 있도록 하며 고령 운전자 및 신호 위반으로 인한 교통사고 문제를 예방할 수 있다.

1. 서론

인구의 고령화 현상이 심화됨에 따라 고령운전자의 수는 꾸준히 증가하고 있다. 도로교통공단 자료에 의하면 고령운전자의 사고 건수는 2021년 기준 31,814건으로 2017년 26,713건 대비 19.1퍼센트 증가했다. 뿐만 아니라 교통사고 분석 시스템의 자료에 따르면 법규 위반 별 교통사고 항목에서 신호 위반이 2위를 차지하고 있다[1]. 이처럼 신체능력의 저하 및 신호 위반에 따른 교통사고의 예방을 위해 신호등 인식이 절대적으로 필요한 상황이다.

본 연구에서는 One-Stage 기반의 YOLOv5 모델을 검출 단계에 활용하였고 신호등 인식을 위해 C-RNN 기반의 attention-OCR 모델을 활용하였다. C-RNN 기반의 모델은 주로 OCR 및 손글씨 인식

그리고 객체 검출 task에 활용되는 모델이지만 task에 모델을 한정하지 않고 신호등 인식을 위한 단계에서 활용하였다. 뿐만 아니라 신호등을 인식하는 기존 연구들 보다 많은 9개의 클래스를 설정하여 인식했다는 점에서 의의를 가진다.

2. 본론

2.1 데이터셋

본 연구에서는 한국전자통신연구원(ETRI)에서 제공하는 신호등 데이터셋을 이용하였다. 본 데이터셋은 개인정보 보호를 위해 비식별화 되어 있는 데이터셋으로 96,307장의 이미지가 2048x1536 사이즈의 픽셀로 구성되어 있다. 본 데이터셋 안에 총 147,201개의 신호등 정보가 제공되며 어노테이션 된 정보를 함께 제공한다. 검출 단계에서는 서울특별시, 세종특별자치시, 2개의 대전광역시 행정구역 각 4개의 도시에서 바운딩 박스의 크기가 큰 순서대로 25장, 작은

[†] 교신저자(Corresponding author): KT 융합기술원 한연지(yeonjee.han@kt.com)

순서대로 25장씩 총 200장의 학습 데이터를 이용하였고 각 도시마다 stratified하게 무작위 추출을 통해 80장의 validation 데이터 셋을 구성하였다. 인식 단계에서는 본 연구에서 목표로 하는 클래스를 추려 총 8142 장의 이미지를 가공하여 학습에 이용했다. 본 데이터를 8:2 비율로 나눠 6,514 장으로 학습을 진행하고 1,628장으로 테스트를 진행했다.

2.2 객체 탐지 (Object Detection)

본 연구의 검출 단계에서는 YOLOv5 모델들의 성능을 비교하였고 속도와 정확도를 고려해 YOLOv5 small 모델을 활용했다(그림 1). YOLOv5의 nano, small, medium, large 모델을 모두 같은 데이터와 같은 파라미터를 이용하여 학습하였다. large 모델의 경우 100 epoch가 넘어가면서부터 validation loss가 증가하는 overfitting이 발생하였고 medium 모델의 경우 이미지 한 장당 최대 3,000.0ms의 시간이 소요되어 영상에서의 사용성이 떨어졌다. 이에 따라 NVIDIA GeForce GTX 1080 Ti 11164MiB 환경에서 이미지 당 평균 9.2ms의 추론 시간을 갖는 small 모델을 이용하였으며 validation에서 accuracy 0.98, Recall 1.00, Precision 0.98, F1-Score 0.99의 성능을 보였다(표 1).

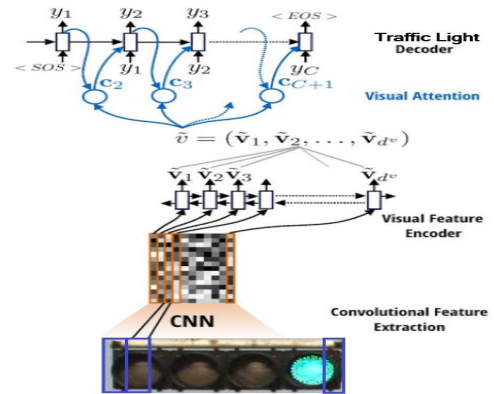


(그림 1) 신호등 탐지 결과

2.3 객체 인식 (Object Recognition)

본 연구의 인식 단계에서는 Attention-OCR 모델을 이용하였다[2]. Attention-OCR 모델은 인코더, 디코더 파트로 나누어있다(그림 2). 인코더 부분에서는 Sliding CNN을 feature extractor로 이용하여 입력으로 들어온 이미지에 대한 feature vector를 추출한다. 디코더 부분에서는 추출된 feature vector를 output layer가 아닌 LSTM 모델의 input으로 하여 신호등에 대한 sequence를 예측한다. 이후 이 예측한 sequence를 텍스트로 변환하는 과정을 거친다.

기존의 연구는 신호등 인식을 위해 내부의 원 영역을 추출하여 색을 검출하는 방식으로 이루어졌다. 하지만 본 연구에서는 C-RNN 방식을 이용해 신호등 전체 이미지로부터 신호를 인식하여 오탐지가 될 요소를 낮추었으며 그 결과 신호등 인식률을 높일 수 있었다. 뿐만 아니라, NVIDIA GeForce GTX 1080 Ti 11164MiB 환경에서 이미지 당 평균 20ms의 추론 시간으로 영상에서의 활용 가능성을 확보하였다.



(그림 2) Attention-OCR 모델의 아키텍처[3]

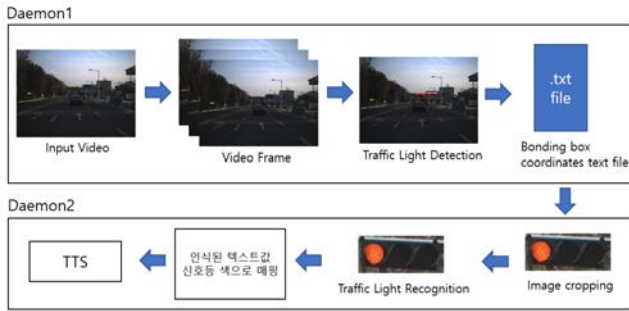
<표 1> 모델에 대한 성능 평가 점수

	Accuracy	Recall	Precision	F1-Score
YOLOv5	0.98	1.00	0.98	0.99
Attention-OCR (macro)	0.99	1.00	0.99	0.99
Attention-OCR (weighted)	0.99	1.00	1.00	1.00

2.4 프레임워크

2.4.1 통합 모델

본 프레임워크는 멀티 스레드 방식으로 객체 감지, 객체 인식 두 가지 모델을 연동하였으며 [그림 3]과 같이 동작한다. 먼저 Daemon Thread 1에서 openCV로 불러온 동영상의 프레임을 추출하여 각 이미지마다 객체 감지 모델을 통해 신호등을 탐지한다. 그리고 탐지 결과로 생성된 bounding box의 좌표와 이미지 파일의 위치를 텍스트 파일로 저장한다. 본 파일은 Daemon Thread 2에서 읽혀져 이미지가 cropping 되고 객체 인식 모델의 input으로 들어간다. 인식된 신호등에 따라 4자리의 텍스트 값이 출력되며, 이를 토대로 빨간불, 초록불 등 신호를 분류하여 TTS를 통해 알린다.

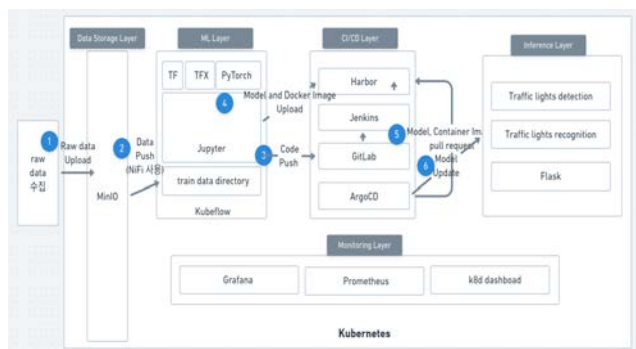


(그림 3) 통합 모델 아키텍처

2.4.2 쿠버네티스 기반 통합 시스템

머신러닝 모델을 기존 시스템과 결합하여 가치를 창출하는 활동이 증가하고 있으나 안정적인 업데이트와 효과적인 적용에는 어려움이 많다. 또한 변화하는 실세계에 적응하기 위하여 지속적인 재학습이 필수이다[4]. 본 연구에서는 쿠버네티스 기반의 데이터 수집 및 분석과 개발, 배포를 자동화하는 파이프라인을 구축해 향상된 모델의 신속한 배포와 안정적인 시스템 운영을 제공한다(그림 4).

수집된 raw data는 데이터 스토리지에 저장 후 flow data를 효율적으로 처리, 배포할 수 있는 시스템인 NiFi를 통해 추출된다. 해당 데이터로 모델의 학습 및 테스트가 이루어지며 도출된 모델과 소스코드는 GitLab Runner를 통해 구축한 파이프라인을 거친다. 테스트 및 빌드 Job을 통해 docker imager가 생성되며 harbor를 통해 구축한 private docker register에 업로드된다. 업로드 이후, GitLab과 연동된 ArgoCD에서 변경된 yml 파일을 감지하여 현재 쿠버네티스의 manifest와 새로 업데이트된 manifest를 비교하여 새로운 버전을 쿠버네티스에 배포한다. 배포된 모델은 지속적인 모니터링을 통해 성능 감지를 파악하여 재학습을 통한 정확도를 유지할 수 있게 한다.



(그림 4) System workflow

3. 결론

본 연구에서는 C-RNN 기반의 모델을 신호등 인식에 활용하여 유의미한 결과를 얻어냈다. 본 프레임워크에서는 두 개의 모델을 연동함으로써 기존에 없던 인사이트를 찾아내고 검증했다. 또, 머신러닝에 GitOps를 적용하여 CI/CD 파이프라인을 구축해 안정성과 확장성을 갖추었다. 향후 이를 토대로 고령 운전자 및 신호 위반으로 인한 교통사고를 예방할 뿐만 아니라 다양한 융합 모델을 이용한 후속 연구의 발전에 도움이 될 것이다.

사 사

본 논문은 과학기술정보통신부 정보통신창의인재 양성사업의 지원을 통해 수행한 ICT멘토링 프로젝트 결과물입니다.

본 연구는 과학기술정보통신부 및 정보통신기술진흥센터의 정보통신·방송 연구 개발사업의 일환으로 수행하였음. [2017-0-00068, 운전자 주행경험 모사기반 일반 도로환경의 자율주행4단계(SAE)를 지원하는 주행판단엔진 개발 / 2018-0-00327, 고정밀 맵 음영 환경의 완전자율주행 네비게이션 인공지능 기술개발]

참고문헌

- [1] “교통사고정보 인포그래픽”, TAAS 교통사고분석시스템, http://taas.koroad.or.kr/sta/acs/gus/selectStaInfoGraph.do?menuId=WEB_KMP_IDA_TAI
- [2] Baoguang S., Xiang B., Cong Y., An End-to-End Trainable Neural Network for Image-based Sequence Recognition and Its Application to Scene Text Recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 2298-2304
- [3] Yuntian Deng, Anssi Kanervisto, Jeffrey Ling, Alexander M. Rush, Image to Markup Generation with Coarse to Fine Attention, Proceedings of the 34th International Conference on Machine Learning, PMLR 70:980-989, 2017.
- [4] Sasu M., Henrik S., Eero L., Tommi M., Who Needs MLOps: What Data Scientists Seek to Accomplish and How Can MLOps Help?, IEEE/ACM 1st Workshop on AI Engineering - Software Engineering for AI (WAIN), 2021