

# 딥러닝을 이용한 실시간 운전자 얼굴 검출 시스템

유민우, 한동석

경북대학교 대학원 전자전기공학부

dshan@knu.ac.kr

## Real-time Driver Face Detection System Using Deep Learning in Embedded Environment

Min Woo Yoo, Dong Seog Han

School of Electronic and Electrical Engineering, Kyungpook National Univ.

### 요약

운전자 모니터링 시스템이란 운전자가 졸린 상태 또는 전방 주시 태만과 같은 운전자의 행동을 감지한다. 위험 행동이 감지된다면 경보음이나 진동으로 운전자에게 경고를 보낸다. 운전자 모니터링 시스템에서 가장 중요한 알고리즘은 운전자의 얼굴을 검출하는 알고리즘이다. 운전자의 얼굴에는 운전자의 상태정보가 매우 많으므로 운전자의 얼굴을 정확히 검출하는 것이 매우 중요하다. 그러나 얼굴 검출기의 정확도를 위해 고성능 네트워크를 사용한다면 전체 시스템이 느려지는 문제가 발생한다. 본 논문에서는 운전자 환경 분석을 통한 데이터 전처리 알고리즘과 최신 검출기인 YOLOv5 알고리즘을 이용하여 이러한 문제를 해결한다. 본 논문의 제안 방법은 연산량이 제한된 임베디드 보드에서 실시간 처리 문제점을 해결하는 방법이 될 것으로 기대한다.

### I. 서론

운전자 모니터링 시스템은 운전자의 상태를 모니터링하여 이상이 발견되면 운전자에게 경고하여 안전한 주행을 돕는 기능을 제공한다. 기존 운전자 모니터링 시스템은 영상처리 기법을 사용한다. 얼굴과 눈을 검출하기 위해 영상처리 방법인 Haarlike 방법을 사용한다[1]. 그러나 영상처리 기반 운전자 모니터링 시스템은 운전자의 다양한 행동에 강인하게 검출하지 못해 제한된 환경에서만 사용할 수 있다. 이러한 영상처리 기반 운전자 모니터링 시스템의 한계를 해결하기 위해 최근 딥러닝 기술을 많이 적용하고 있다. 그러나 딥러닝 알고리즘은 높은 계산 복잡도를 요구한다. 또한, 이러한 시스템은 낮은 연산자원을 가진 임베디드 보드에서 동작함으로써 실시간 처리에 어려움이 있다.

본 논문에서는 운전자 모니터링 시스템에서 가장 높은 연산자원을 요구하는 얼굴 검출기를 성능 저하 없이 최적화를 한다. 이를 위해 운전자 환경 데이터 분석과 객체 검출기인 YOLOv5[2]를 사용한다. 운전자의 얼굴을 검출하기 위해 WIDER FACE[3] 데이터를 사용한다. WIDER FACE 데이터는 약 40만 개의 얼굴이 포함되어있는 3만 장의 이미지가 있다. 객체 검출기는 1-stage 방법과 2-stage 방법이 있다. 2-stage 방법은 느린 추론속도 대신 정확도가 높은 특징이 있다. 그러나 실시간 처리를 위해서 2-stage를 적용하기 어렵다. 그러므로 빠른 추론속도를 가진 1-stage 방법인 YOLOv5 모델을 사용한다.

### II. 본론

운전자의 얼굴을 검출하기 위해 WIDER FACE 데이터를 사용한다. 그림 1은 WIDER FACE 데이터 중 하나로 매우 작은 얼굴이 많은 것을 알 수 있다. 일반적인 비전 센서에서 입력되는 운전자의 얼굴 크기와 다른 것을 알 수 있다. 그림 2는 운전자의 얼굴을 검출하기 위해 사용된 WIDER

FACE 데이터의 분석 결과다. 얼굴의 크기가 500픽셀 이하인 데이터가 93%를 차지한다. 그러므로 일반적인 비전 센서에서 입력되는 운전자의 얼굴 크기와 다른 얼굴을 모두 제거했다. 제거하는 방식은 얼굴이 작은 레이블링 값을 모두 지웠다. 만약 제거 후 이미지에 유효한 얼굴 레이블링이 없다면 이미지를 제거했다.



그림 1. WIDER FACE 데이터

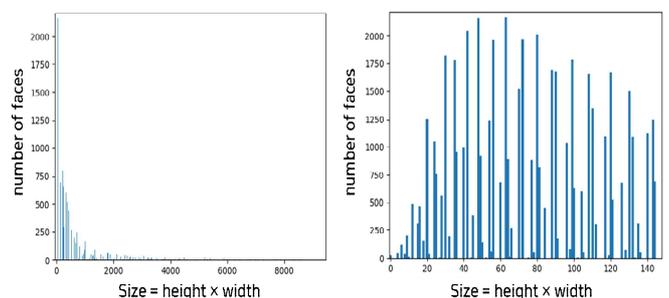


그림 2. WIDER FACE 데이터의 얼굴 크기 분포도

객체 검출기로 YOLOv5를 사용했다. YOLOv5 모델 구조는 입력 이미지의 특징 정보를 추출하는 Backbone과 Backbone에서 추출된 각기 다른 계층에서 추출된 특징 정보로 객체를 추론하는 Head로 구분된다. Backbone의 네트워크 깊이에 따라 YOLOv5\_N, YOLOv5\_S, YOLOv5\_M, YOLOv5\_L, YOLOv5\_X로 구분된다. 본 논문에선 n과 s 그리고 m 모델로 학습 후 성능을 비교했다.

그림 3은 작은 얼굴을 제거한 앵커 박스와 제거하지 않은 앵커 박스의 결과다. 그림 3 (a)는 제거한 앵커 박스의 결과로 제거하지 않은 그림 3 (b)보다 다양한 크기를 가지고 있다. 그러나 작은 얼굴을 제거하지 않은 그림 3 (b)는 매우 작은 얼굴을 검출하기 위해 앵커 박스가 생성되었다.

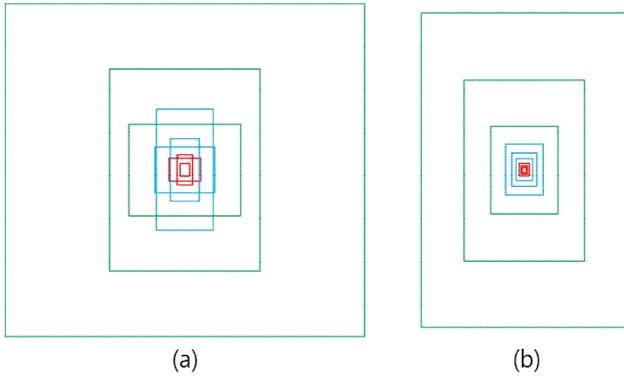


그림 3 YOLOv5의 각 계층에서 사용되는 앵커 박스 결과:  
(a) 제안한 데이터, (b) 기존 데이터의

표 1은 YOLOv5의 3가지 Backborn으로 학습한 결과다. 성능을 평가하기 위한 검증데이터는 큰 얼굴 데이터를 사용했다. 데이터 변경 전에는 M 모델이 가장 성능이 좋았다. 그러나 데이터 변경 후 3가지 모델 모두 비슷한 성능을 보였다.

표 1. YOLOv5 모델의 성능

	FPS	AP@.5	
		Original	Propose data
YOLOv5_N	100	0.758	0.934
YOLOv5_S	73	0.780	0.933
YOLOv5_M	68	0.801	0.935

### III. 결론

본 논문에서는 운전자를 모니터링하는 비전 센서의 프레임에 운전자의 얼굴 크기를 분석을 통해 WIDER FACE 데이터를 전처리했다. 또한, 전처리된 데이터와 전처리 되지 않은 데이터를 각각 YOLOv5로 학습하여 N 모델에서도 M과 성능이 유사한 모델을 만들었다.

### ACKNOWLEDGMENT

본 연구는 산업통상자원부와 한국산업기술진흥원이 지원하는 5G기반 자율주행 융합기술 실증 플랫폼 과제(과제고유번호 : 1415169669)와 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터지원사업 (IITP-2022-2020-0-01808\*)의 지원을 받아 수행된 연구임

### 참고 문헌

- [1] R. Lienhart and J. Maydt, "An Extended Set of Haar-Like Features for Rapid Object Detection." Proc. IEEE Int'l Conf. Image Processing Processing. vol. 1, pp. 900-903, 2002.
- [2] YOLOv5, [https://pytorch.org/hub/ultralytics\\_yolov5/](https://pytorch.org/hub/ultralytics_yolov5/).
- [3] Yang, Shuo, et al. "Wider face: A face detection benchmark." Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 5525-5533, 2016.