

## 딥러닝 기반 운전자 이상행동 탐지 모델 성능 비교에 관한 연구

임다빈, 윤수연\*

국민대학교, \*국민대학교

dabinim09@kookmin.ac.kr, \*1104py@kookmin.ac.kr

## A Study on the comparative performance of deep learning-based driver abnormal behavior detection models

Lim Da Bin, Yoon Soo Yeon\*

Kookmin Univ., \*Kookmin Univ.

## 요약

본 연구는 프레임 단위 운전자 이상행동 모니터링 환경에서 ResNet50 과 YOLO11n-cls 모델의 성능과 적용 가능성을 통일된 데이터 전처리 및 파이프라인과 일관된 평가 지표 하에 구현하고 비교 분석하였다. ResNet50 은 YOLO11n-cls 에 비해 정확도면에서 소폭 우세하였으나 YOLO11n-cls 는 연산 효율면에서 크게 앞서 실시간 이상행동 탐지에 적합함을 확인하였다. 본 연구는 실시간 운전자 모니터링 시스템 설계에서 모델 선정시 가이드라인을 제시한 연구로서 향후 실시간 환경에 적합한 모델에 대한 추가적 연구를 통해 다양한 분야에서 사고 예방과 운영 비용 효율화에 기여할 것으로 기대된다.

## I. 서론

경찰청이 발표한 통계에 따르면 최근 6 년 동안 (2018 년 ~ 2023 년) 교통사고의 주요 원인으로 안전운전 불이행이 꾸준히 높은 비율을 차지한다. [1] 이에 해당하는 유형으로는 졸음운전, 전방 주시태만, 휴대전화 사용, 차내 장치 조작 등이 있으며 교통사고의 직접 원인이자 교통안전에 위협하는 심각한 과제이다. [2]

이런 문제를 해결하기 위해 다양한 이상행동 모니터링 시스템 연구가 진행되어왔다. [3] 기존의 운전자 이상행동 모니터링 방법 중 프레임 기반 기법은 카메라로부터 입력된 개별 이미지를 분석하여 얼굴 표정, 시선 방향, 손 동작 등을 실시간으로 처리할 수 있으며, 2D CNN 기반의 분류 네트워크와 객체 탐지 네트워크, Vision Transformer 등이 제안되었다.

그러나 다양한 연구 간 성능 비교는 각기 다른 네트워크 구조와 전처리 방식, 평가 지표를 사용함으로써 동일한 조건에서의 공정한 비교 및 적합한 모델을 판별하기 어려운 문제가 있다. 실시간 운용 시 복잡한 딥러닝 모델의 연산 지연 문제는 안전 보조 시스템에 즉각 대응에도 한계가 있다. [4]

이에 본 연구에서는 프레임 기반 이상행동 탐지 모델을 통일된 데이터 전처리 및 학습 파이프라인과 일관된 평가 지표 하에 구현하고 비교함으로써 모델간 비교의 공정성을 확보하고 연산 지연 문제를 평가, 비교하여 실시간 운전자 모니터링 시스템 적용에 적합한 모델을 선정하고자 한다. 본 연구를 통해 운전자 이상행동 탐지 분야에서 객관적인 모델 비교 근거 마련 및 모델 선정 가이드라인을 제시하고 운전 보조 시스템의 상용화에 기여하고자 한다.

## II. 관련연구

## 2.1. 딥러닝 기반 이상행동 탐지 모델

## 2.1.1. ResNet50

ResNet50 은 50 개의 레이어로 구성된 네트워크로 깊어질수록 발생하기 쉬운 기울기 소실 문제를 잔차 학습 구조를 도입하여 완화한다. 각 블록마다 스킵 연결을 통해 이전 출력이 다음 블록으로 직접 전달되어 학습 안정성과 성능이 뛰어나다. [5] 운전자의 얼굴 표정 변화나 손동작 같은 정적 특징을 프레임 단위로 정확히 분류할 수 있다.

## 2.1.2. YOLO11n-cls

YOLO11n-cls 는 하나의 전역 특징 맵 위에서 바운딩 박스 좌표와 클래스 확률을 동시에 예측하여 높은 처리 속도와 우수한 검출 성능을 달성하는 객체 탐지 모델 YOLO11n 을 분류 버전으로 경량화한 모델이다. [6] YOLO11n 의 경량 백본을 공유되 바운딩 박스 헤드를 제거하고 분류 헤드를 연결한 설계로 운전 중 휴대전화 사용이나 핸들 조작과 같은 이상행동 프레임을 지연 없이 포착할 수 있다.

## III. 실험 설계

## 3.1 데이터셋 구성 및 실험환경

## 3.1.1 데이터셋 구성

본 연구에서 사용된 데이터는 AI-Hub 의 “운전자 및 탑승자 상태 및 이상행동 모니터링” 데이터로, 초당 30 프레임의 영상과 영상에서 추출된 5 개의 프레임 이미지에 각각 운전자의 29 가지 행동이 맵핑 되어 있으며 이 중 연구에서 분류한 주요 이상상태에 대한 행동 유형으로서 표 1 과 같이 정리하였다.

표 1. 운전자 이상행동 유형 분류

이상상태	행동 유형
졸음운전	꾸벅꾸벅졸다, 하품, 몸뭇가누기
휴대전화 사용	핸드폰키기, 핸드폰키에대기
전방 주시태만	허리굽히다
차내 장치 조작	중앙으로손을뻗다

또한 학습 데이터 분포를 균일하게 하기 위해 각 행동별로 15,000 개의 데이터로 다운샘플링 하여 학습을 진행하였으며 학습, 검증, 테스트 데이터는 각각 8:1:1 의 비율로 분할하였다.

### 3.1.2 실험환경

본 연구에서는 모델간 비교의 공정성을 달성하기 위해 표 2 에서 정리된 것과 같이 통일된 하이퍼파라미터 및 학습 파이프라인을 구성하였다.

표 2. 하이퍼파라미터 및 학습 파이프라인 구성

설정	값
Random Seed	42 (Python, numpy, dataloader, deterministic)
Image HxW	224 x 224
Epochs	100
Early stop patience	5
Learning Rate	1E-04
Optimizer	AdamW
Train augmentation	Rotation 10 -> ResizedCrop 0.5 -> Perspective 0.05 -> HorizontalFlip 0.5
Val augmentation	Resize (256/224) -> CenterCrop
Weight decay	0.2

## IV. 실험 결과 및 성능 평가 비교 분석

### 4.1. ResNet50 모델 실험 결과

그림 1 의 ResNet50 모델 실험 결과에서 학습은 epoch 84 에서 5 번 이상 정확도 향상이 이루어지지 않아 조기 종료되었으며 종료 시 검증 정확도는 91.9%로 측정되었다. 테스트 데이터셋을 이용해 실제 데이터의 클래스와 모델이 예측한 클래스로 혼동행렬을 구성한 결과 모델이 올바르게 예측한 클래스는 11,012 개로 전반적으로 분류가 잘 되고 있음을 확인하였다.

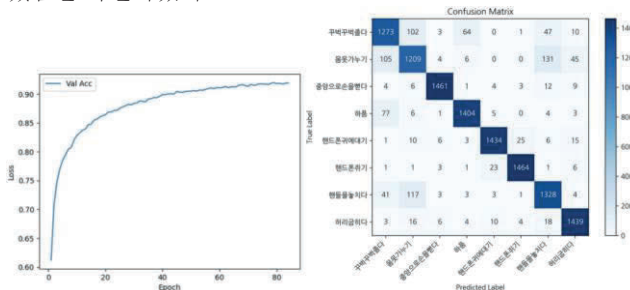


그림 1. ResNet50 학습 결과 (에포크별 정확도 <왼쪽>, 혼동행렬 <오른쪽>)

### 4.2. YOLO11n-cls 모델 실험 결과

그림 2 의 YOLO11n-cls 모델 실험 결과 또한 동일하게 epoch 84 에서 학습이 조기 종료되었으며 검증 정확도는 91.1%로 측정되었다. 혼동행렬을 구성한 결과 모델이 올바르게 예측한 클래스는 10,903 개로 ResNet50 모델에 비해 올바르게 예측한 클래스가 100 개 정도 적지만 YOLO 모델 또한 전반적으로 분류가 잘 되고 있음을 확인하였다.

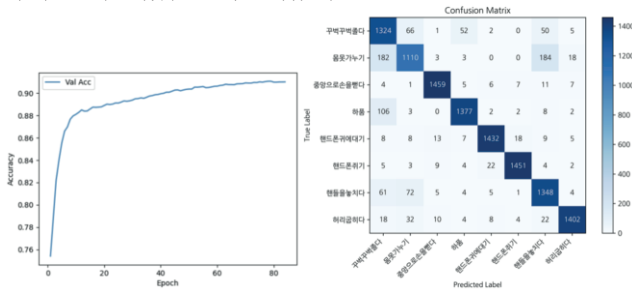


그림 2. YOLO11n-cls 학습 결과 (에포크별 정확도 <왼쪽>, 혼동행렬 <오른쪽>)

### 4.3. 모델 성능 평가 및 비교 분석

두 모델 간 성능 평가 및 비교 분석을 위해 분류 성능 지표로는 모델이 올바르게 예측한 비율을 나타내는 Accuracy 와 정밀도와 재현율 간의 균형을 고려한 F1-Score 수치를 비교하였으며 두 수치 모두 ResNet50 모델이 YOLO11n-cls 모델보다 약 1% 정도 높게 나타나 분류 성능 면에서는 소폭 우세한 것을 확인할 수 있다. 표 3 에서 이를 정리하였다.

표 3. 분류 성능 지표 비교

Metrics	ResNet50	YOLO11n-cls
Accuracy	0.9177	0.9086
F1-Score	0.9177	0.9088

또한, 실시간 적용 가능성 분석을 위한 연산 효율 지표로는 초당 분류 프레임 수(FPS), 프레임당 분류 시간(Latency) 그리고 파라미터 수를 비교하였으며 높은 gpu 자원을 활용하기 어려운 차량 내 상황을 고려하여 cpu 환경에서 측정하였다. ResNet50 모델은 초당 3 개 정도의 프레임을 분류할 수 있으며 파라미터 수가 23M 인 것에 비해 YOLO11n-cls 모델은 초당 59 개의 프레임의 분류가 가능하며 파라미터 수도 1.5M 으로 비교적 경량화 된 모델임을 증명한다. 표 4 에서 이를 정리하였다.

표 4. 연산 효율 지표(cpu)

Metrics	ResNet50	YOLO11n-cls
FPS	3.79 img/s	59.1 img/s
Avg Latency	263.54ms	16.92ms
Median Latency	265.81ms	15.15ms
파라미터 수	약 23M	약 1.5M

## V. 결론 및 시사점

본 연구에서는 통일된 조건에서 ResNet50 과 YOLO11n-cls 모델을 공정하게 비교한 결과 YOLO11n-cls 모델이 경량화된 모델로서 실시간 운전자 이상행동 모니터링 시스템에 더 적합함을 입증하였다. 경량화된 모델 선정을 통해 차량 내 엠티 디바이스에서 즉각적인 경고를 가능케 해 교통사고를 예방할 수 있고 유지관리 및 업데이트에 소요되는 비용을 절감할 수 있다. 이외에도 산업 현장에서 작업자의 피로 및 부주의 상태를 탐지하여 안정성을 확보할 수 있으며 의료현장에서는 환자의 낙상 및 급격한 자세 변화를 조기에 포착하여 응급 대응 시간을 대폭 단축할 수 있다. 로봇틱스 및 드론 분야에서는 주변환경 탐지를 통한 충돌 회피 등에 사용될 수 있다. 이처럼 본 연구에서 확인한 실시간 환경에 적합한 경량화된 이상행동 탐지 모델은 다양한 분야에서 실시간 위험 알림과 자동 제어 기능을 통해 사고 예방과 운영 비용 효율화에 기여할 것으로 기대된다.

## 참 고 문 헌

- [1] 경찰청 2024 교통사고통계
- [2] 한국도로교통공단 신호등
- [3] 고병철, “자율주행차량 운전자 모니터링 연구 동향” 전자통신동향분석, 대한전자공학회, 2022
- [4] 유민우, 차대웅, 윤영진 외 3 인, 임베디드 환경에서 딥러닝을 이용한 실시간 운전자 모니터링 시스템, 한국통신학회, 2022
- [5] K. He, X. Zhang, S. Ren, J. Sun, “Deep Residual Learning for Image Recognition,” CVPR, 2016.
- [6] Khanam, R., & Hussain, M. (2024). YOLOv11: An Overview of the Key Architectural Enhancements. arXiv:2410.17725.