운전자 폭행 탐지를 위한 경량 CNN 아키텍처 기반

시간적 세그먼트 네트워크 기법[[1]](#footnote-2)

(Temporal Segment Network Technique based on

Lightweight CNN Architecture for Driver Assault Detection)

김 준[[2]](#footnote-3) 박 준 희[[3]](#footnote-4) 이 준 용[[4]](#footnote-5) 조 종 호§

(Joon Kim) (Junhui Park)   (Juneyong Lee) (Jongho Jo)

**요 약** 최근 택시와 같은 교통수단에서 운전자를 폭행하는 사건이 증가하고 있는데, 특히 주취자에 의한 폭행은 주로 늦은 시간에 발생해 신속한 대응이 어려운 상황이다. 이러한 문제에 대응하기 위해, 본 연구는 운전자 이외의 탑승자에 의한 운전자 폭행 상황을 신속하고 정확하게 탐지할 수 있는 경량 컨볼루션 신경망(CNN) 기반의 시간적 세그먼트 네트워크(TSN) 모델을 제안하고자 한다. 본 모델은 차량 내 운전자 모니터링 시스템(DMS)에 통합되어 제한된 자원 하에서도 효율적으로 작동할 수 있도록 기존의 ResNet 기반 TSN 모델을 개선하여, 자원 제약 조건 하에서도 정확하고 빠른 성능을 유지할 수 있는 경량 CNN 아키텍처를 도입하였다. 본 연구에서 제안하는 기법은 운전 중 발생할 수 있는 위험한 상황에 대한 신속한 대응 및 예방에 기여할 것으로 기대된다.[[5]](#endnote-2)

**키워드** : 시간적 세그먼트 네트워크, 경량 합성곱 신경망 아키텍처, 3D 비디오 RGB 기반 폭력 인지, 비정상적 행동

***Abstract*** Recently, there has been an increase in incidents of drivers being assaulted in transportation services like taxis, especially assaults by intoxicated passengers, which often occur late at night, making prompt response difficult. To address this issue, this study proposes a lightweight Convolutional Neural Network (CNN) based Temporal Segment Network (TSN) model capable of quickly and accurately detecting driver assaults by passengers other than the driver. This model is designed to be integrated into the in-vehicle Driver Monitoring System (DMS) and operates efficiently under limited resources by improving upon the existing ResNet-based TSN model. It introduces a lightweight CNN architecture that maintains accurate and fast performance even under resource-constrained conditions. The technique proposed in this study is expected to contribute to the rapid response and prevention of dangerous situations that can occur while driving.

**Key words :** Temporal Segment Network, Lightweight CNN Architecture, 3D Video RGB based Assault Recognition, Abnormal Behavior

**1. 서 론**

현대 사회에서 운전자 폭행 사건에 대한 보도가 끊임없이 나오고 있고, 21년 8월 전남 여수 택시기사 폭행 사건[1], 22년 12월 광주 서구 택시기사 폭행 사건 등 대대적으로 보도될 만큼 크고 작은 운전자 폭행 사건이 전국적으로 발생하고 있다.

경찰청에 따르면 지난 2022년 운전자 폭행 발생건수는 약 4300여건으로 매일 평균 약 12회 꼴로 운전자 폭행 사건이 발생하고 있다. 이에 따라 보호 격벽 설치 의무화[2] 등 운전자 폭행 방지를 위한 노력을 하고 있지만, 경찰청 통계 자료에 따르면 최근 5년간 운전자 폭행 건수는 줄어들기는 커녕 오히려 급증하는 추세이다.

스크린샷, 라인, 텍스트, 그래프이(가) 표시된 사진

자동 생성된 설명

그림 1 5년간 운전자 발생건수 (경찰청 통계)

Fig. 1 Number of Driver Incidents Over 5 Years (Police Agency Statistics)

본 연구는 택시와 같은 승용차 내에서의 운전자 폭행을 탐지할 수 있는 딥러닝 모델을 구축했고, 이를 통해 운전자들이 폭행을 당해 더욱 큰 피해로 이어지지 않을 수 있도록 빠른 연산을 지원하는 경량화 모델을 제안하였다.

**2. 관련 연구**

**2.1 관련 데이터셋**

운전자 폭행 탐지를 위한 데이터셋으로는 다음과 같은 다양한 데이터 셋들을 사용할 수 있다.

먼저 다양한 손동작에 중점을 둔 데이터 셋(Everal hand-focused datasets)으로 인간-기계 상호 작용을 위한 손 제스처 인식 데이터셋인 CVRR-HANDS 3D[3], VIVA-Hands[4], Driver MH[5], 손동작 및 자세 등을 통한 운전자 상태 식별 데이터셋인 Ohn-bar[6][7], 손동작 및 6DOF 머리 자세 특정 데이터셋인 DD-Pose[8], Drive A Head[9], 손동작 및 눈동자 추적 데이터셋인 A Ground-Truth Data[10], Dr(eye)VE[11], DADA [12]등이 있다.

그리고 운전자의 행동에 중점을 둔 데이터셋으로 운전자의 얼굴 데이터셋인 DriveFace[13], 측면으로 촬영된 운전자 행동 영상 데이터셋인 AUC Distracted Driver(AUC DD)[14], 운전자 행동 멀티 모달 영상 데이터셋인 Drive&act[15], 운전자 행동 이미지 데이터셋인 StateFarm[16]등이 있다.

위 데이터셋들은 운전자의 행동을 다루는 데이터셋이긴 하지만 대부분 운전자 개인의 행동을 분류하기 위한 데이터셋으로 운전자 폭행을 탐지하기 위한 데이터셋으로는 적합하지 않다.

최종적으로 본 연구에서는 AI Hub의 “운전자 및 탑승자 상태 및 이상행동 모니터링” 데이터셋을 채택하였다. 해당 데이터셋에는 운전자 개인의 이상행동 뿐만 아니라 운전자를 포함한 탑승자가 2명 이상이고, 탑승자가 운전자를 폭행하는 영상도 존재하므로 이를 이용해 운전자 폭행 탐지 모델을 학습시킬 수 있다.

**2.2 경량 CNN 아키텍처**

운전자 폭행 탐지를 위해서는 자동차 내부의 하드웨어를 사용해야 하므로 한정된 자원에서의 효율적인 자원활용이 불가피하다. 그렇기에 기존의 Resnet 기반의 아키텍처에 경량화 기법을 적용해 최소한의 자원으로 폭행을 탐지할 수 있는 다양한 경량 CNN 아키텍처를 사용할 수 있다.

SqueezeNet[18]은 가장 처음이자 가장 잘 알려진 아키텍처로, AlexNet 수준의 정확도를 50배 적은 매개 변수로 달성하기 위해 fire 모듈로 구성되어 있다. ShuffleNet[19]은 포인트와 그룹 컨볼루션을 함께 사용하기 위한 채널 셔플 작업을 제안한다. ShuffleNet V2[20]은 경량 아키텍처 설계에 효과적인 여러 원칙을 도입하여 기존의 ShffleNet에서 업그레이드를 하였다. Neural Architecture Search(NAS)[21], NASNet [22], FBNet[23]은 기본적인 block은 depth wise, pointwise layer로 factorization하였다. MobileNet [24]은 더 얇거나 더 넓은 네트워크를 달성하기 위해 너비 멀티플라이어 매개 변수와 함께 깊이 별 분리 컨볼루션을 포함한다. MobileNetV2[25]은 역방향 잔여 블록과 ReLU6 활성화 함수를 포함한다. MobileOne [26]은 skip connection를 제거하고 SE-block의 수는 적당히 조절한다. 기본적인 block은 depth wise, pointwise layer로 factorization하고, Rep-VGG에서 사용한 re-parameterizable skip connection을 사용한다.

**3. 경량 CNN 기반 TSN(Temporal Seg ment Network)**

운전자 폭행을 탐지하기 위해서는 행동을 인식해야 하므로 이를 위해 TSN(Temporal Segment Network) 모델을 도입하였다.

**3.1 Temporal Segment Network(TSN)**

Temporal Segment Network for detecting action (동작 인식을 위한 시간적 세그먼트 네트워크)는 시간적 정보를 적절하게 처리하여 정확하고 효과적으로 동작 인식을 수행하는데 특화된 모델로 제안된 딥러닝 아키텍처 중 하나이며, 이 모델은 비디오를 여러 세그먼트로 나누고 각 세그먼트에서 동작을 인식한 후 이를 병합해 결과를 도출한다.

Two Stream Model[27]은 공간적정보와 시간적정보를 별개의 stream으로 각각 학습한 후 마지막에 그 정보들을 융합하는 모델의 아키텍처이다. Spatial Stream에는 이미지가 입력되어 공간적 정보를 입력 받고, Temporal Stream에는 두 이미지 간의 차이를 의미하는 Optical Flow를 입력 받아 시간적 정보를 입력 받는다. Two Stream Model은 이러한 구조를 통해 시⋅공간적 정보를 통합하여 해당 영상이 어떤 영상인지 분류하게 된다.

텍스트, 스크린샷, 폰트이(가) 표시된 사진

자동 생성된 설명

그림 2 2-스트림 모델 아키텍처

Fig. 2 Two Stream Model Architecture

광학 흐름(Optical Flow)와 같이 다음프레임으로 넘어갈 때 각 픽셀이 어느 방향으로 이동했는지 벡터로 나타낸 것이다. 광학 흐름을 계산할 때 밝기 유지, 시간적 지속성, 공간적 위상이라는 3가지 기본 가정이 있다. 광학 흐름을 계산하여 시간적 stream에 활용한다.

스크린샷이(가) 표시된 사진

자동 생성된 설명

그림 3 시간적 세그먼트 네트워크 옵티컬 플로우

Fig. 3 TSN Optical Flow

Temporal Segment Networks[28]는 장기적인 시간 모델링을 위해 입력 영상을 K개의 Snippets 또는 Clips으로 나누어 K개의 two stream 모델에 입력한다. 아래의 그림 4에서는 입력Snippets이 3개이다.

텍스트, 스크린샷, 도표, 지도이(가) 표시된 사진

자동 생성된 설명

그림 4 시간적 세그먼트 네트워크 아키텍처

Fig. 4 Temporal Segment Network Architecture

본 연구에서는 총 8개의 Snippets으로 나누어 모델에 입력한다.

**3.2 Mobileone**

본 연구에서 선택하게 된 경량 CNN아키텍처는 Mobileone이다. 대다수의 파라미터 수가 높은 모델들이 latency가 낮게 측정되 것을 확인했고, MobileNet이 Transformer계열 모델들과 비슷한 FLOPs임에도 latency가 낮은 것을 확인했다. 런타임 성능에 영향을 주는 것은 메모리 접근과 병렬 구조정도라고 생각한다. 메모리 접근 비용은 각각의 branch에서 다음 텐서(tensor)까지 이동하는 것에서 저장을 하기 때문에, activation같은 multi-branch 구조에서 병목이 생기기 때문에 최대한 적게 하는 것이 성능에 좋아질 것이라고 생각한다. 또한, global pooling같은 경우의 동기화도 속도를 느리게 하는 원인이 되는 것이라고 파악한다. Inference할때 branch를 사용하지 않았고, Squeeze -Exite Block을 가장 큰 variant로 제한해서 사용하면서 성능을 올렸다고 제안했다.

이를 통해 Mobileone은 우리의 폭행데이터에 대해 효율적인 자원을 활용하여 폭행을 탐지하는 태스크를 효과적으로 수행할 수 있다. Apple에서 여러 경량 CNN아키텍처와 비교하며 가장 성능이 좋은 아키텍처로 연구를 통해 밝혔기에 본 연구에서는 Mobileone 아키텍처를 선정하였다.

**4.** **실험 및 평가**

**4.1 데이터셋 및 실험 환경**

2.1 관련 데이터셋에서 밝혔듯이 운전자 폭행 탐지에 가장 적합한 AI Hub의 “운전자 및 탑승자 상태 및 이상행동 모니터링” 데이터셋을 사용하였다. 이 데이터셋은 이상행동 7종(졸음운전, 음주운전, 물건 찾기, 통화, 휴대폰 조작, 차량 제어, 운전자 폭행)의 장면(Scene)이 영상 및 5개의 이미지로 이루어져 있다. 본 연구에서는 운전자 폭행 탐지를 위해 이상행동 7종의 항목 중 운전자 폭행 데이터만을 사용했고, 각 장면의 5개 이미지를 사용하였다. 또한 학습 시 데이터를 샘플링하는 샘플러(sampler)를 수정해 비슷한 장면은 한 에폭에서 한 번만 학습할 수 있도록 했다. 추가로 실제 환경의 테스트를 위해 유튜브(YouTube)에 업로드 되어 있는 운전자 폭행 영상을 파이썬 라이브러리 moviepy, pytube, editor를 이용하여 추출했고, 폭행구간과 정상 구간을 나누어 라벨링했다.

**4.2 모델에 맞는 데이터 전처리**

본 연구의 TSN은 8개의 Snippets을 입력 받는 영상 인식 모델이므로 이를 학습하기 위해서 8개 이미지로 이뤄진 정상/폭행 데이터가 필요하다.

폭행 데이터 구축을 위해 5개 이미지를 8개로 증강하였고, 정상 데이터 구축을 위해 5개의 이미지 중 운전자 폭행이 없는 이미지를 8개로 증가하였다. 이 과정에서 해당 장면의 흐름을 해치지 않도록 각 이미지들의 순서를 고려하여 증강하였다. 예를 들어 폭행 데이터는 5번째 이미지 후에 4번째, 3번째, 2번째 이미지가 올 수 있게 함으로써 장면의 흐름을 해치지 않으며 8개의 이미지로 증강시켰다. 정상 데이터는 폭행이 없는 연속된 이미지들을 찾고 이들을 장면의 흐름을 해치지 않도록 증강시켰다.

표 1 정상데이터와 폭행데이터 전처리 및 증강방법

Table 1 Preprocessing and Augmentation Meth ods for Normal and Assault Data

|  |  |  |
| --- | --- | --- |
| Scene | Assault | Normal |
| A\_1 | 1,2,3,4,5,4\*,3\*,2\* | 1,2,3,2\*,1\*,2\*,3\*,2\* |
| A\_2  ⋮ | 1,2,3,4,5,4\*,3\*,2\*  **⋮** | 1,2,3,4,3\*,2\*,1\*,2\*  **⋮** |
| E\_1 | 1,2,3,4,5,4\*,3\*,2\* | 1,2,1\*,2\*,1\*,2\*,1\*,2\* |
| E\_2  ⋮ | 1,2,3,4,5,4\*,3\*,2\*  **⋮** | 1,2,3,2\*,1\*,2\*,3\*,2\*  **⋮** |

\* 은 증강된 데이터를 뜻한다.

**4.3 실험결과**

신속한 폭행탐지를 위한 Mobileone 모델을 선정하였음을 3장에서 설명하였다. 여러 Mobileone 모델 중 우수한 성능을 내는 모델을 선정하기 위하여, 여러 실험을 거쳐 최종 모델을 선정하는 과정을 거쳤다. 표 2는 본 연구의 모델로 사용된 Mobileone s0부터 s4의 Params(M)와 Flops(G)의 차이를 비교한 표이다. 표 3에서 AI Hub의 기존 데이터를 테스트 데이터로 성능을 검증한 결과, Tsn\_mobileone\_s3가 기존의 Resnet 기반 TSN 및 다른 Mobileone 기반 TSN 보다 좋은 성능을 기록하였다. 또한 앞선 테스트에서 가장 좋은 성능을 기록한 Tsn\_mobileone\_s3와 기존의 방식인 Resnet기반 TSN을 유튜브(YouTube)에서 추출한 실제 폭행데이터셋으로 검증한 결과, 표4과 표5와 같이 Tsn\_mobileone\_s3가 더 우수한 성능을 보이는 것을 알 수 있다.

표 2 기존 ResNet기반 TSN모델과 본 연구의 제안된 경량 CNN아키텍처 기반 TSN모델 비교

Table 2 Comparison Between the Existing ResNet-Based TSN Model and the Proposed Lightweight CNN Architecture-Based TSN Model in this Study

텍스트, 스크린샷, 폰트, 번호이(가) 표시된 사진

자동 생성된 설명

표3 기존 Resnet기반 TSN모델과 본 연구의 제안된 경량 CNN아키텍처 기반 TSN의 운전자 폭행 탐지에 관한 테스트 데이터로 성능 평가표

Table 3 Performance Evaluation Table on Driver Assault Detection Using Test Data for the Existing ResNet-Based TSN Model and the Proposed Lightweight CNN Architecture-Based TSN Model in this Study

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | | **Precision** | **Recall** | **F1-score** | **Accuracy** |
| Tsn\_r50 | Assault  Normal | 0.88  0.62 | 0.43  0.94 | 0.58  0.75 | 0.69 |
| Macro avg | 0.75 | 0.69 | 0.67 |
| Tsn\_mobileone\_s0 | Assault  Normal | 0.80  0.62 | 0.47  0.88 | 0.83  0.71 | 0.68 |
| Macro avg | 0.71 | 0.68 | 0.66 |
| Tsn\_mobileone\_s1 | Assault  Normal | 0.78  0.64 | 0.51  0.86 | 0.62  0.73 | 0.68 |
| Macro avg | 0.71 | 0.68 | 0.67 |
| Tsn\_mobileone\_s2 | Assault  Normal | 0.75  0.63 | 0.50  0.83 | 0.60  0.72 | 0.67 |
| Macro avg | 0.69 | 0.67 | 0.66 |
| Tsn\_mobileone\_s3 | Assault  Normal | 0.83  0.66 | 0.53  0.89 | 0.65  0.76 | **0.71** |
| Macro avg | 0.74 | 0.71 | **0.70** |
| Tsn\_mobileone\_s4 | Assault  Normal | 0.81  0.65 | 0.53  0.88 | 0.64  0.75 | 0.70 |
| Macro avg | 0.73 | 0.70 | 0.69 |

표 4 기존 Resnet기반 TSN모델의 운전자 폭행 탐지에 관한 실제 데이터로 성능 평가표

Table 4 Performance Evaluation Table on Driver Assault Detection for the Existing ResNet-Based TSN Model using Real-World Data

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | | **Precision** | **Recall** | **F1-score** | **Accuracy** |
| Tsn\_r50 | Assault  Normal | 0.83  0.71 | 0.83  0.71 | 0.83  0.71 | 0.79 |
| Macro avg | 0.77 | 0.77 | 0.77 |
| Weighted avg | 0.79 | 0.79 | 0.79 |

표5 본 연구에서 가장 좋은 성능을 보인 경량 CNN아키텍처 기반 TSN의 운전자 폭행 탐지에 관한 실제 데이터의 성능 평가표

Table 5 Performance Evaluation Table on Driver Assault Detection using Real-World Data for the Lightweight CNN Architecture-Based TSN Model Showing the Best Performance in this Study

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | | **Precision** | **Recall** | **F1-score** | **Accuracy** |
| Tsn\_mobileone\_s3 | Assault  Normal | 0.91  0.75 | 0.83  0.86 | 0.87  0.80 | **0.84** |
| Macro avg | 0.83 | 0.85 | 0.83 |
| Weighted avg | 0.85 | 0.84 | 0.84 |

텍스트, 스크린샷, 라인, 그래프이(가) 표시된 사진

자동 생성된 설명

텍스트, 그래프, 스크린샷, 라인이(가) 표시된 사진

자동 생성된 설명

(a) (b)

그림 5 (a) 훈련 손실 그래프 (b) 학습률

Fig. 5 (a) train loss graph (b) Learning Rate

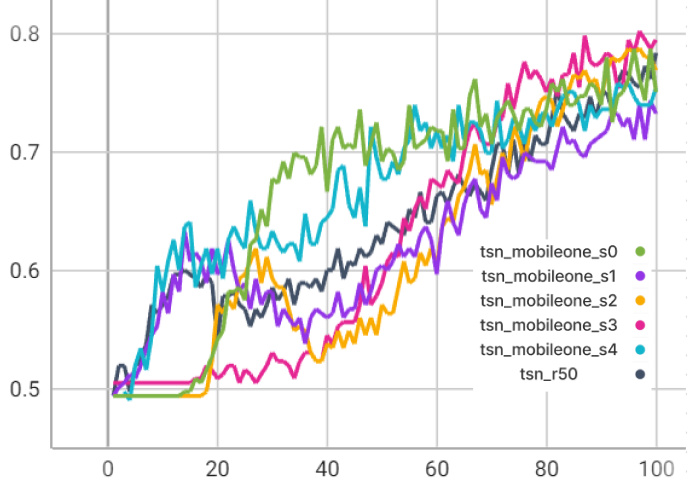


그림 6 Top1 정확도

Fig. 6 Top1\_accuracy

위 그래프는 훈련 손실 그래프로 Loss가 잘 줄어들고있음을 확인할 수 있다. 모든 모델은 100 에폭으로 학습하였지만, Tsn\_r50과 Tsn\_mobileone\_s0은 배치사이즈 16으로 훈련하였고 나머지 모델들은 배치사이즈 8로 훈련하였기에 x축의 step이 절반으로 줄었다. Learning rate의 경우 MultiStepLR Scheduler를 사용하여 40, 80에폭에서 학습률이 작아지도록 설정하여 모델이 고도화될 수 있도록 하였다.

마지막으로 제일 높은 정확도는 확률적 경사 하강법(SGD)으로 훈련시켰을 때 x축 에폭(epoch) 100에 대한 정확도이다. 위 그래프에서 제일 높은 정확도는 tsn\_mobileone\_s3모델의 97에폭임을 알 수 있다.

모델을 수행시켜본 결과, 운전자 폭행탐지를 위해 TSN 모델을 활용한 시스템이 tsn\_mobileone\_s3모델로 80%의 높은 정확도를 달성하였으며 효과적임을 확인하였다.

**5. 결론 및 향후연구**

개발된 모델은 택시 운전자가 당하는 폭행을 빠르고 정확하게 탐지하여 보호해 줄 수 있는 기반을 만들었으며 운전중인 운전자 폭행으로부터 발생하는 다른 차들과 충돌하는 교통사고와 같은 2차 피해를 예방할 수 있을 것으로 기대한다. 또한 본 연구는 다음과 같은 다양한 응용에 활용될 수 있다.

첫째, 보험사 및 법 집행기관에 정확한 사고 경위를 제공하여 공정하고 효과적인 조치를 취할 수 있는 방안으로도 활용될 수 있다.

둘째, 운전자 폭행 사건을 탐지하는 데이터의 수집을 도와 도로 위 운전자 폭행문제에 대한 통계를 제공할 수 있다.

셋째, 운전자 모니터링 시스템과 결합되어 택시나 버스와 같은 대중교통에서의 폭행을 탐지 뿐 아니라 신고까지 자동으로 이루어지도록 할 수 있다.

추후 연구로는 본 연구 결과를 활용하여 택시 뿐만 아니라 버스 등 다양한 교통수단에 확대하여 적용할 계획이다.

**References**

1. 김대우 등 3명, “택시기사 폭행 매년 느는데… ‘보호 칸막이’ 설치는 지지부진”, 문화일보, 2023.
2. 김태연, “줄지 않는 기사 폭행... '가림막' 없는 마을버스 기사는 더 서럽다”, 한국일보, 2023.
3. Eshed Ohn-Bar, Sujitha Martin, and Mohan Trivedi. Driver hand activity analysis in naturalistic driving studies: challenges, algorithms, and experimental studies. Journal of Electronic Imaging, 22(4):041119, 2013.
4. Nikhil Das, Eshed Ohn-Bar, and Mohan M Trivedi. On performance evaluation of driver hand detection algorithms: Challenges, dataset, and metrics. In 2015 IEEE 18th International Conference on Intelligent Transportation Systems, pages 2953–2958. IEEE, 2015
5. Okan Kop¨ ukl ¨ u, Thomas Ledwon, Yao Rong, Neslihan ¨ Kose, and Gerhard Rigoll. Drivermhg: A multi-modal dataset for dynamic recognition of driver micro hand gestures and a real-time recognition framework. arXiv preprint arXiv:2003.00951, 2020
6. Eshed Ohn-Bar, Sujitha Martin, and Mohan Trivedi. Driver hand activity analysis in naturalistic driving studies: challenges, algorithms, and experimental studies. Journal of Electronic Imaging, 22(4):041119, 2013.
7. Eshed Ohn-Bar and Mohan Trivedi. In-vehicle hand activity recognition using integration of regions. In 2013 IEEE Intelligent Vehicles Symposium (IV), pages 1034–1039. IEEE, 2013.
8. Markus Roth and Dariu M Gavrila. Dd-pose-a large-scale driver head pose benchmark. In 2019 IEEE Intelligent Vehicles Symposium (IV), pages 927–934. IEEE, 2019.
9. Anke Schwarz, Monica Haurilet, Manuel Martinez, and Rainer Stiefelhagen. Driveahead-a large-scale driver head pose dataset. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pages 1–10, 2017.
10. Ioannis Agtzidis, Mikhail Startsev, and Michael Dorr. 360- degree video gaze behaviour: A ground-truth data set and a classification algorithm for eye movements. In Proceedings of the 27th ACM International Conference on Multimedia, pages 1007–1015, 2019.
11. Andrea Palazzi, Davide Abati, Francesco Solera, Rita Cucchiara, et al. Predicting the driver’s focus of attention: the dr (eye) ve project. IEEE transactions on pattern analysis and machine intelligence, 41(7):1720–1733, 2018
12. Jianwu Fang, Dingxin Yan, Jiahuan Qiao, and Jianru Xue. Dada: A large-scale benchmark and model for driver attention prediction in accidental scenarios. arXiv preprint arXiv:1912.12148, 2019.
13. Katerine Diaz-Chito, Aura Hernandez-Sabat ´ e, and Anto- ´ nio M Lopez. A reduced feature set for driver head pose ´ estimation. Applied Soft Computing, 45:98–107, 2016.
14. Yehya Abouelnaga, Hesham M Eraqi, and Mohamed N Moustafa. Real-time distracted driver posture classification. arXiv preprint arXiv:1706.09498, 2017.
15. Manuel Martin, Alina Roitberg, Monica Haurilet, Matthias Horne, Simon Reiß, Michael Voit, and Rainer Stiefelhagen. Drive&act: A multi-modal dataset for fine-grained driver behavior recognition in autonomous vehicles. In Proceedings of the IEEE international conference on computer vision, pages 2801–2810, 2019.
16. State Farm. State farm distracted driver detection. https://www.kaggle.com/c/state-farm-distracted-driverdetection, accessed: 22-September-2020.
17. Neslihan Kose, Okan Kopuklu, Alexander Unnervik, and Gerhard Rigoll. Real-time driver state monitoring using a cnn based spatio-temporal approach. In 2019 IEEE Intelligent Transportation Systems Conference (ITSC), pages 3236–3242. IEEE, 2019.
18. Forrest N Iandola, Song Han, Matthew W Moskewicz, Khalid Ashraf, William J Dally, and Kurt Keutzer. Squeezenet: Alexnet-level accuracy with 50x fewer parameters and¡ 0.5 mb model size. arXiv preprint arXiv:1602.07360, 2016.
19. Xiangyu Zhang, Xinyu Zhou, Mengxiao Lin, and Jian Sun. Shufflenet: An extremely efficient convolutional neural network for mobile devices. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 6848–6856, 2018.
20. Ningning Ma, Xiangyu Zhang, Hai-Tao Zheng, and Jian Sun. Shufflenet v2: Practical guidelines for efficient cnn architecture design. In Proceedings of the European Conference on Computer Vision (ECCV), pages 116–131, 2018.
21. Barret Zoph and Quoc V Le. Neural architecture search with reinforcement learning. arXiv preprint arXiv:1611.01578, 2016.
22. Barret Zoph, Vijay Vasudevan, Jonathon Shlens, and Quoc V Le. Learning transferable architectures for scalable image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 8697–8710, 2018.
23. Bichen Wu, Xiaoliang Dai, Peizhao Zhang, Yanghan Wang, Fei Sun, Yiming Wu, Yuandong Tian, Peter Vajda, Yangqing Jia, and Kurt Keutzer. Fbnet: Hardware-aware efficient convnet design via differentiable neural architecture search. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 10734–10742, 2019.
24. Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861, 2017.
25. Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 4510–4520, 2018.
26. Pavan Kumar Anasosalu Vasu, James Gabriel, Jeff Zhu, Oncel Tuzel and Anurag Ranjan. “Mobile One : An Improved One millisecond Mobile Backbone”, pages 2-4, 2022,( https://arxiv.org/abs/2206.04040)
27. Karen Simonyan, Andrew Zisserman, “Two-Stream Convolutional Networks for Action Recognition in Videos”,2014,( https://arxiv.org/abs/1406.2199)
28. Limin Wang, Yuanjun Xiong, Zhe Wang, Yu Qiao, Dahua Lin, Xiaoou Tang, Luc Van Gool , “Temporal Segment Networks: Towards Good Practices for Deep Action Recognition”, 2016 (<https://arxiv.org/abs/1608.00859>)

김 준

2017~ 현재 한국외국어대학교 전자물리학과 학사과정 재학중.   
관심분야 Computer Vision, Object Detection, Natural Language Processing.

인간의 얼굴, 사람, 눈썹, 턱이(가) 표시된 사진

자동 생성된 설명박 준 희

2019~ 현재 한국외국어대학교 통계학과 학사과정 재학중.  
관심분야는 Datascience , Data Analysis , Computer Vision, Object Detection

이 준 용

2019~ 현재 한국외국어대학교 컴퓨터전자시스템공학과 학사과정 재학중. 관심분야는 Computer Vi sion, Object Detection, Image Seg mentation

조 종 호

2017~ 현재 한국외국어대학교 컴퓨터전자시스템공학과 학사과정 재학중. 관심분야는, Economics of IT and AI, Data Science, Deep learning

1. 본 연구는 한국외국어대학교 딥러닝 과목에서 AI 분야에 관심이 있는 학생들이 모여 자유롭게 주제를 탐색하여 연구해보고 Research Paper를 작성해보고 자 수행된 연구입니다. [↑](#footnote-ref-2)
2. 한국외국어대학교 전자물리학과 학생 [↑](#footnote-ref-3)
3. 한국외국어대학교 통계학과 학생 [↑](#footnote-ref-4)
4. 한국외국어대학교 컴퓨터 전자시스템공학 학생 [↑](#footnote-ref-5)
5. https://github.com/HUFSuperman/HUFS\_SavingDriver [↑](#endnote-ref-2)