

運転手の姿勢と表情に対する CNN を用いた危険運転の推定

2019SC051 廖永存

指導教員：河野浩之

1 はじめに

近年、自動車の数が増え、道路の運転環境が複雑になり、交通事故の可能性が高まる。警察庁交通局による交通事故のデータをもとに、公益財団法人交通事故総合分析センターがまとめた運転手の人的要因が交通事故の重要な原因であることが示されている。安全運転義務違反が73.7% 占めている [1]。その中に漫然運転と脇見運転などの運転手が運転することに集中していないのが大きく占めている。

深層学習の発展に伴い、危険な運転行動の特定に様々な技術を適用する研究者がますます増えている。その中で、顔表情、頭の姿勢の推定などの分野での深層学習の使用を通じて、多くの研究が運転手の顔を識別し、目の開閉状態、口の開閉状態などの情報を使用して、運転手が疲労運転、あくびなどの危険な運転行動をしているかどうかを判断できる。

2 姿勢と表情の認識に関する研究

表 1: 先行研究

研究者	課題	手法
西銘ら [2]	人の表情識別	CNN モデル
山賀ら [3]	運転姿勢の識別	距離データを SVM
西行ら [4]	運転手動作と姿勢推定	マルチタスク学習

姿勢と動作の認識に関する研究は表 1 に示す。西銘ら [2] は人の怒り、恐怖などの7つ表情のデータセットを使用して、構築した CNN モデルで実験を行った。全体で約 57% の認識精度が得られた。恐怖、怒りの精度が最小 30% 未満。喜び、驚きの精度が最大 70% になった。

山賀ら [3] は、搭乗者の姿勢に応じて複数の測距センサーで取得できる距離情報の分布から識別を行なう手法を提案し、取得したデータを機械学習することで、実験参加者個々の運転姿勢の識別が 90% 以上精度になった。

西行ら [4] は、ドライバーの姿勢と動作のマルチタスク学習を用いた高精度かつ演算量が少ない動作認識を提案した。姿勢推定を CNN、動作認識を RNN を用いてドライバーの運転状態を推定する。精度が動作認識には 82%、姿勢推定には 80% 以上になった。

3 危険運転推定の提案

3.1 危険運転の定義

危険運転はアルコールなどの影響で正常な運転ができない状態、制御できない高速度での運転、信号を無視する運転などがよく指している。

本研究では、運転手の顔を着目して、目や口の状態を検出し眠気を推定して危険運転を判定する。まず、顔の向きを検出して、前方に向いたら次に目と口の状態を検出して目が閉じたらまたは口があくびしたら運転手が運転に集中していないと推定でき、危険運転に判定する。

3.2 提案手法

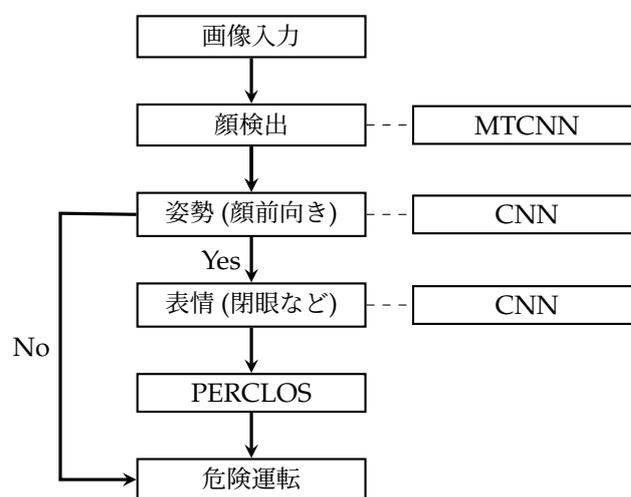


図 1: 危険運転推定フロー

本研究では、運転手の顔に対して、ディープラーニングに属する畳み込みニューラルネットワークを用いて危険運転の推定を提案した。提案手法の概要図を図 1 に示す。まずは MTCNN 手法を用いて顔である確率、顔の領域と顔のランドマーク座標を出力する。目が二つと口が検出したら顔が前向きだと考えられる。次に、目の部分と口の部分だけ含まれているデータセットを用意し、構築した CNN モデルに入力し学習させ、開いてるか閉じてるかを分類する。最後にはビデオや PC カメラを使ってリアルタイムに顔の領域、目と口の状態を出力する。

3.3 危険運転の推定方法

眼球の動きと瞬きは疲労を反映する重要な特徴とされ、瞬き頻度、平均開閉時間は直接疲労の検出に用いることができる。PERCLOS (Percent of Eye Closure) は現在最も応用可能性の高いリアルタイム疲労検出方法とされている。

本研究では、リアルタイムに目の状態を出力するために、

運転手の頭部は静止ではなく、よく動かしている、また、人によって目のサイズも違うから、眼球を覆うまぶたの面積を計算するのが難しいだと考えられる。この PERCLOS の考え方に従って、次のように変更する。目が閉じたから時間を計算し、ある時間を超えても目が閉じたままだったら、眠気による危険運転だと判定する。また、運転手が喋っているとき口が開いたり閉じたりするのをあくびだと誤認識しないため、目の部分と同じように口が開いたままである時間を超えたら、あくびによる危険運転だと判定する。

4 実験

4.1 データセット

使用するデータセットは三つがある。一つ目は WIDER FACE(<http://shuoyang1213.me/WIDERFACE/>) であり、61 のイベントクラスに基づいて編成され、クラスごとに数百以上のラベル付き jpg 画像で構成されている。二つ目は Drowsiness Detection Dataset(<https://www.kaggle.com/datasets/prasadvpatil/mrl-dataset>) であり、二クラスに分けられ、Closedeyes と Openeyes がそれぞれ 2000 枚の png 画像で構成されている。三つ目は Yawn Dataset(<https://www.kaggle.com/datasets/davidvazquezcic/yawn-dataset>) であり、二クラスに分けられ、yawn が 2528 枚、no yawn が 2591 枚合計 5119 枚の jpg 画像で構成されている。

4.2 実験環境の構築

本研究では Windows10、メモリが 16GB のカメラ付き PC を使って、Anaconda での Jupyter Notebook に Python 環境を構築し、フレームワークに Tensorflow や OpenCV を使用して実験を行う。学習には CPU を使用する。

4.3 実験と結果

まず、顔の検出については P-Net、R-Net、O-Net 三つのネットワークを構築し、顔の検出を行う。画像データを先に P-Net に入力し学習する、P-Net の出力が R-Net の入力とする、R-Net の出力が O-Net の入力として三回学習をし、顔である確率、顔の領域、ランドマーク座標を出力した。

次に、目の部分と口の部分だけ含まれているデータセットを用意し、構築した CNN モデルに入力し学習させ、開いているか閉じているかを分類した。結果は表 2 に示す。

最後に、PC カメラを使って目の方が閉じてからフレーム数の計算を始め、20 フレームを超えたら、ウォーニングがでる。口の方が開いてからフレーム数の計算を始め、10 フレームを超えたらウォーニングが出る。評価するためには、目と口両方が開いている、目と口両方が閉じている、目が閉じている口が開いている、目が開いている口が閉じている四つの様子をそれぞれ 10 回実験を行った。その結果、ウォーニングが出る回数がそれぞれ 6 回、7 回、7 回、2 回になった。危

険運転だと考えられる場合はほぼウォーニングが出たが、そうではない場合も 2 回の誤認識があった。30 回の危険運転の状態、ウォーニングが 20 回出て、危険運転推定の正解率が 66.7% になった。

表 2: 顔と表情検出学習結果

Net	エポック	バッチサイズ	学習率	精度
P-Net	30	100	0.001	92.8%
R-Net	22	100	0.001	96.1%
O-Net	22	100	0.001	96.9%
目の状態	20	22	0.01	100%
口の状態	20	22	0.01	98.0%

5 むすび

本研究では MTCNN 手法で二つの目と口の座標を正しく出力し、顔が前向きだと判断できる。次に、目と口の画像を CNN モデルに学習して Closed と Open の二クラスに分類できた。PC カメラを使って、目と口の状態をリアルタイムに正しく出力できてから、改善した PERCLOS 方法を導入し目と口の状態から被験者の眠気検出ができて、危険運転推定の正解率が 66.7% になった。

一方、本研究では目と口の状態がポイントであるため、サングラスやマスクを着用すると目と口を遮って、状態の分析ができなくなり、危険運転の推定には難しいだと考えられる。

参考文献

- [1] 三井住友海上, “自動車事故の発生状況,” https://www.ms-ins.com/special/rm_car/accident-data/, 2022 年 8 月.
- [2] 西銘大喜, 遠藤聡志, 當間愛晃, 山田孝治, 赤嶺有平, “畳み込みニューラルネットワークを用いた表情表現の獲得,” 人工知能学会論文誌, 32 巻 5 号, pp.F-H34.1-8, 2017.
- [3] 山賀瑛斗, 浅野直生, 杉浦裕太, 杉本麻樹, “Time-of-flight 測距センサアレイと機械学習を用いた運転手の姿勢識別,” 計測自動制御学会論文集, 55 巻 11 号, pp.772-779, 2019.
- [4] 西行健太, 日向匡史, 田崎博, 木下航一, 長谷川友紀, 山下隆義, 藤吉弘巨, “ドライバ姿勢と動作のマルチタスク学習による高速かつ高精度なドライバ動作認識,” 人工知能学会論文誌, 36 巻 2 号, pp.1-10, 2021.
- [5] Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, Senior Member, Yu Qiao, Senior Member, “Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks,” IEEE Signal Processing Letters, Volume:23, Issue:10, pp.1499-1503, 2016.