

Convolutional Recurrent Neural Network による方向指示器の認識

Turn Signal Recognition by Convolutional Recurrent Neural Network

萩 亮太^{*1}
Ryota Hagi

米陀 佳祐^{*1}
Keisuke Yoneda

菅沼 直樹^{*1}
Naoki Suganuma

^{*1} 金沢大学
Kanazawa University

Autonomous vehicle is effective in reducing traffic accidents. A behavior prediction for surrounding vehicle is one of the important functions for autonomous vehicle. This behavior prediction is generally based on moving and posture information which are recognized using range sensors such as laser and radar sensors. This study focused on the behavior prediction about lane changes and intersection driving for surrounding vehicles. When the surrounding vehicles change lane and turn left-or- right, they turn on their turn signal. To adopt turn signal information lead to predict such behavior in advance. This paper describes the method to recognize blinking turn signal using Convolutional Neural Network (CNN) and Recurrent neural networks(RNN). Two types of network models are investigated and evaluated the performance for daytime dataset during public load driving. As a result, one of networks operate with 6.25[FPS] and obtain almost 90[%] of recognition accuracy.

1. 緒言

近年、センサやカメラを使用した衝突回避システムなどの自動車の安全に対する技術の向上により交通事故の件数が減少している。さらなる交通事故件数の減少を目的として自動運転自動車の開発が行われている。

自動運転自動車に使用されている要素技術として、周辺環境認識が挙げられる。周辺環境認識において周囲の移動物体の位置と運動を考慮しながら自動運転自動車の走行軌道を最適化するにあたり、周辺車両の行動予測が重要な技術である。その周辺車両の行動予測において従来の手法では、レーザレーダやミリ波レーダなどの距離センサの情報から周辺車両の速度や加速度及び姿勢角などを認識し、この運動情報を用いて数秒先の状態を算出している。本研究は方向指示器という運転者が周囲に行動を示す手段を従来手法に取り入れることにより、より早い行動予測を目指す。方向指示器の情報を使用することで、路肩停止中の車両や対向車がいる場合での交差点での右折待ちの車両に対して効果が期待できる。

方向指示器認識の先行研究として [Fröhlich 2014]がある。Fröhlich らはステレオカメラの画像から車両位置を特定し、差分画像を用いて方向指示器の点灯を認識する。その認識結果を時系列に収集したものに高速フーリエ変換(FFT)を行うことで周波数のスペクトルを算出する。算出した周波数スペクトルに対して AdaBoost[Freund 1997]に適用することで方向指示器の点滅を認識していた。AdaBoostを学習するにあたり、車両画像からの周波数スペクトルだけでは方向指示器の認識をするのには学習データ数として不足である。そのため、収集した周波数スペクトルとは異なる、ランダムに生成した周波数スペクトルを学習に使用することで精度の向上を行っている。一方で、この手法ではステレオカメラを使用しており、システムが大規模になる課題もある。そこで先行研究[Yoneda 2017]では、単眼カメラを使用した方向指示器の認識を行ってきた。

先行研究[Yoneda 2017]では、車に搭載された単眼カメラの画像から手動で切り出した車両画像を Deep Neural

Network(DNN)に投入し点灯判定を行い、その結果を収集したものに FFT を行うことにより周波数推定を行っていた。[Yoneda 2017] この手法では時系列の特徴を考慮せず画像 1 枚からの点灯情報のみで判断するため、FFT による周波数推定において誤認識につながりやすい。そこで、本研究ではフレームからの情報を増やし、直近 10 フレームの情報を Recurrent Neural Network(RNN)を使用したネットワークに通して方向指示器の認識を行うことで精度の向上を目指す。

2. 先行研究

先行研究[Yoneda 2017]では、以下のような手順により方向指示器の認識を行っていた。

- 1) 左右に分割した車両画像を DNN に投入し方向指示器の点灯を判定
- 2) 直近 10 フレームでの点灯判定に対し FFT を行い、方向指示器の周波数を推定
- 3) 推定結果が方向指示器の周波数である約 1.5[Hz]であるかより方向指示器の点滅を判断

この手法において車両画像を左右に分割することにより、左右どちらが点灯しているかを判断する必要をなくし、認識が複雑になることを防いでいる。また、このアルゴリズムにおいて、方向指示器の点灯認識に使用する DNN には AlexNet[Krizhevsky 2012]の構造において各層のサイズを変更したものを使用し、概略図を図 1 に示す。また、方向指示器の点灯認識を行うために出力クラスを方向指示器の点灯/消灯の 2 クラスに変更して学習を行っている。この方向指示器認識ネットワークは手動で切り出された車両画像においては 90[%]近い認識精度が得られる。しかし、3.1 節で述べる DNN により検出した車両画像では認識精度が 72[%]であり、更なる精度の向上が求められる。この精度低下の原因として、各フレームでの情報が点灯/消灯のみであり、1 フレームの誤認識が FFT での周波数推定に大きな影響をもたらすことが挙げられる。そのため、本研究では各フレームでの情報を増やし、直近 10 フレームの情報を RNN を使用したネットワークに投入することで方向指示器認識を行い、精度の向上を目指す。

連絡先: 萩亮太, 金沢大学, 〒920-1192 石川県金沢市角間町,
Email:hagi@stu.kanazawa-u.ac.jp

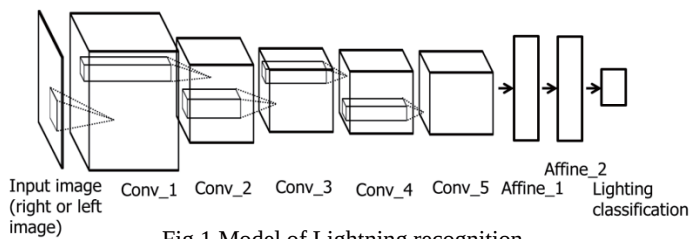


Fig.1 Model of Lightning recognition

3. 提案手法

本研究では、方向指示器の認識に先行研究と同様に車両の連続画像を使用する。そのため、カメラの画像から車両画像を取得する必要がある。先行研究ではこの部分を手動で行っていたため、本研究では認識精度の向上とともに、認識アルゴリズム全体の自動化を目指す。

3.1 方向指示器認識ネットワークの自動化

この節ではカメラ画像を取得してから車両の連続画像を取得するまでの流れを示す。この車両の連続画像を取得するために以下の作業を行う。

- カメラ画像における車両位置の特定
- フレーム間における同一車両の特定

まず、車両画像を切り出すために、カメラ画像内での各車両の位置を特定する必要がある。そのため、本研究では一般物体認識のネットワークである Shot Single MultiBox Detector(SSD) [Liu 2016]を使用し、車両位置の特定を行う。

また、方向指示器の認識には各車両の連続画像が必要となる。そのため、SSD で取得した各車両の位置情報を次のフレームと照らし合わせ、領域の重複度を Intersection over Unit(IoU) から計算する。この重複度が閾値より高いものを同一車両として判断する事で車両ごとにデータの収集を行う。

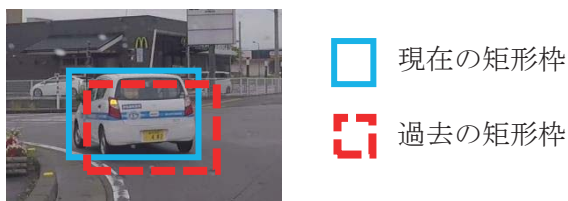


Fig.2 Judgement of same vehicle

3.2 方向指示器認識モデル

この節では、得られる連続画像に対して、RNN を使用した方向指示器の認識ネットワークについて述べる。作成するモデルを図 3, 4 に示す。

本研究ではフレームごとの特徴を[Yoneda 2017]で方向指示器の点灯認識に使用した AlexNet の途中の層の出力から得る。AlexNet の途中の層の出力を使用することで各フレームでの情報量を増やし、精度の向上を目指す。作成するネットワークには、図 1 において画像内での位置に関係した特徴量を持つ Conv_5 から特徴を取り出したモデル、方向指示器の点灯認識に使用している Affine_2 から特徴を取り出したモデルの 2 種類を作成し検証を行う。

作成する 2 種類のモデルはそれぞれ取り出した特徴量 10 フレーム分を収集し、RNN を含んだネットワークに入力し方向指示器の認識を行う。本研究では RNN に Grid Recurrent Unit(GRU)[Cho 2014]を使用する。

(1) Affine Layer の出力を使用したモデル

このモデルでは、本研究の先行研究において、点灯を認識するために重要な Affine Layer の情報を使用することで、各フレームでの情報量を増やし、方向指示器の認識を行う。

(2) Convolution Layer の出力を使用したモデル

このモデルでは、左右に分割した車両画像をそれぞれ AlexNet に入力し、Convolution Layer から特徴を取得する。この特徴は Affine Layer から取り出す特徴よりも浅い層からの特徴であるため、方向指示器の点灯との関係性の高くない特徴も含まれている可能性がある。しかし、位置情報を持つ特徴を取り出すことができるため Convolution Layer から取得した特徴を左右で結合することで車両画像全体での特徴を作成し方向指示器の点滅の認識に使用する。この特徴を 10 フレーム分取得し、3 次元に畳み込みを行った後、GRU に入力することで方向指示器の点滅を認識する。3 次元での畳み込みは入力に使用している車両の連続画像内で方向指示器の座標の変化を吸収するために使用する。

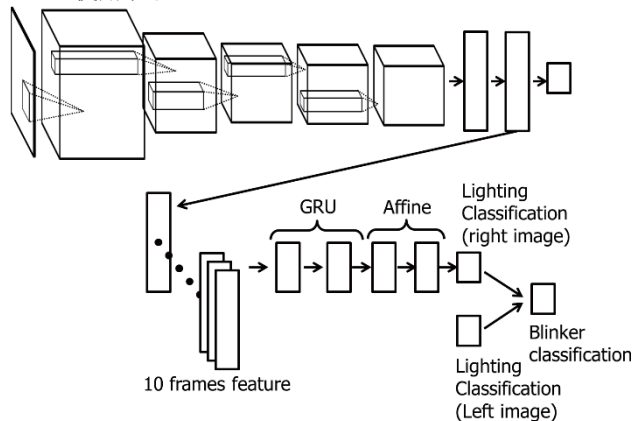


Fig.3 Affine model

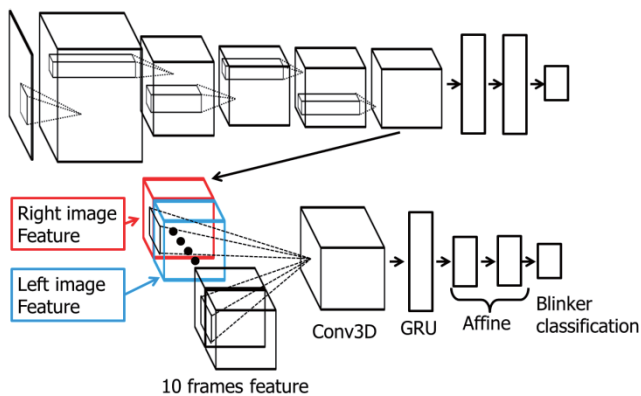


Fig.4 Conv model

3.3 データセット

本研究で行う方向指示器の認識用のデータセットは存在していないため、自作する必要がある。まず、車両の連続画像の切り出しにおいては 3.1 節で述べた SSD による車両画像位置の特定と IoU による同一車両の特定を行うことにより収集を行う。

また正解ラベルを付ける作業において、すべてを手作業で行うには効率が悪いので、先行研究[Yoneda 2017]を使用して方向指示器認識を行う。この方法では間違っただけのラベルが付いたデータも含まれているため、その部分においては目視で修正を行う。この収集方法により方向指示器の認識に必要な 10 フレーム

の連続画像を No signal:587 本, Right:457 本, Left:361 本, Hazards:2 本の合計 1407 本収集した。

しかし、これだけでは学習に必要なデータ数にはとどかないため、データを拡張することで不足を補う。本研究では、拡張方法として、基本的な方法である、

- 回転
- 明るさ
- コントラスト

の変更に加え、方向指示器の認識精度向上のため下記の対策を行うことで過学習を抑え、精度の向上を試みる。

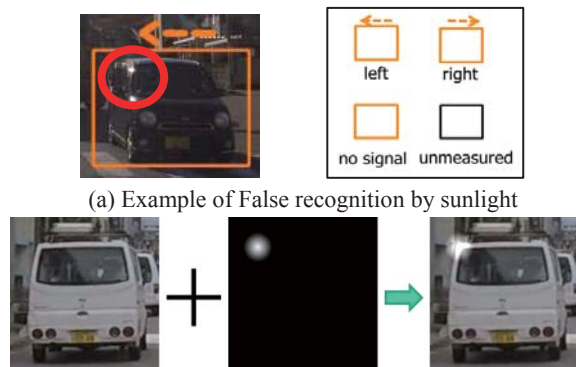
- 太陽光の対策
- 異なる周波数の対策

画像の基本的な拡張結果は図 4 のようになる。回転は、道路のうねりなどで実際に傾く可能性がある範囲である ± 4 度回転させたものを使用する。これに明るさとコントラストそれぞれ元画像を含む、3 倍に拡張させたものを組み合わせることで収集データ 1407 本を 27 倍の 37989 本に拡張を行った。



Fig.4 Example images for data augmentation

上記の基本的な拡張方法を行ったデータにおいて学習を行ったモデルにおいて、図 5(a)の赤丸で囲った部分に見られるような太陽光への反射を方向指示器の点灯として誤認識しているものが見られた。この対策として、学習データにランダムで図 5(b)のような白色円の入った画像を重ね合わせたものを消灯として学習を行った。



(b) Method of Sunlight measure
Fig. 5 Sunlight measure

また、方向指示器の点滅を認識するにあたり、0[Hz]付近は消灯、方向指示器の点滅周波数である 1.5[Hz]は点滅として学習を行うが、それ以外の周波数において学習を行っていないため、誤認識につながる。この対策に、Fröhlich らが AdaBoost の学習時に行っていた拡張と同様のデータ拡張を行う。Fröhlich らは、車両の差分画像から点灯を認識し、FFT の出力周波数スペクトルに加えてランダムに周波数スペクトルを作成することで広範囲の周波数への対策を行った。本研究では図 6 に示すように点灯画像と消灯画像をランダムに並び替えた連続画像を作成し、消灯として学習を行うことで誤認識の減少を目指す。本研究では点灯画像と消灯画像を 10:0, 9:1, 8:2 の割合で並び変えた連続画像 5824 本と回転などの拡張により得た 37989 本を合わせた 43813 本の連続画像で学習を行った。

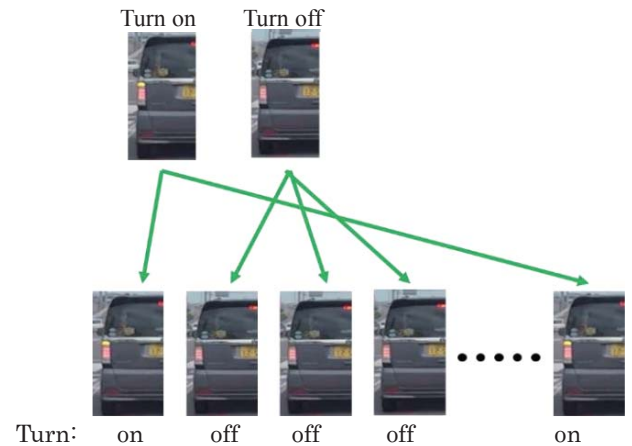


Fig.6 Frequency measure

4. 実験

4.1 学習と評価方法

学習時のハイパーパラメータの条件を表 1 に示す。また、GPU は NVIDIA の GeForce GTX 1080 Ti を 1 台使用して学習と評価を行った。

Table 1 Learning condition

学習方式	ミニバッチ学習
最適化関数	SGD
損失関数	Categorical crossentropy
学習率	0.01
バッチサイズ	8
エポック	30
トレーニングデータ	43813

評価に使用するデータとしては日中の 11 動画 1334 フレームを使用し、no signal: 924 回, right: 783 回, left:425 回の合計 2132 回での精度の比較を行う。今回使用しているテストデータでは出現回数の低い hazards を収める事ができなかったため、No signal, Right, Left の 3 クラスでの精度とクラス精度の平均精度 (Accuracy) の 4 種類において精度の比較を行う。評価方法として Accuracy を使用することでクラスごとの検出数の差に影響されにくい精度の比較を行う。

4.2 各ネットワークでの精度の比較

テストデータにおける本研究で作成したモデルと先行研究 [Yoneda 2017] のモデル (FFT model) での精度を表 2 に示す。Accuracy を比較すると Conv model が最も高い精度を示してお

り、約 90[%]近くを示している。また、各クラスでの精度の比較を行うと、FFT model での精度は Right と Left での精度が 50~65[%]であるが、提案手法ではどちらも 80[%]を上回っていることから、20~35[%]の精度の向上が得られた。また、No signal の精度も 90[%]を超えていることから、全体的な精度の向上を行えた。

Table 2 Accuracy evaluation of networks

	FFT model	Conv model	Affine model
No signal	0.989	0.904	0.929
Right	0.659	0.831	0.890
Left	0.500	0.825	0.864
Accuracy	0.716	0.852	0.894

4.3 学習データの拡張方法での精度の比較

Affine Layer の特徴を使用したネットワークにおいて、基本的な拡張である、回転、明るさ、コントラストの拡張を行ったもの(Base)、太陽光への対策を行ったもの(Shine)、周波数への対策を行ったもの(Freq)、太陽光と周波数の両方の対策を行ったもの(Shine+Freq)の 4 種類において精度の比較を行ったものを表 3 に示す。Accuracy において、太陽光と周波数両方の対策を行ったものもともと精度が高くなっている。各クラスでの精度を見ると、太陽光への対策を行うことで No signal の精度を向上しており、周波数の対策を行うことで、Right, Left の精度の向上が見られた結果両方の対策を行ったものがどのクラスにおいても基本の拡張より精度の向上が見られていることがわかる。

Table 3 Accuracy evaluation of data augmentation method

	Base	Shine	Freq	Shine+Freq
No signal	0.879	0.940	0.885	0.904
Right	0.810	0.736	0.839	0.831
Left	0.741	0.769	0.830	0.825
Accuracy	0.810	0.815	0.851	0.852

4.4 処理速度

今回使用しているカメラは、7.5[FPS]で動作しており、各フレームでリアルタイムに処理するための最大時間は $1/7.5[\text{FPS}] = 133[\text{msec}]$ となる。本研究で作成したモデルの平均処理時間は約 160[msec]でフレームレートは 6.25[FPS]となり、リアルタイム処理を行うには改善が必要である。また、各フレームでの処理時間の内訳において、方向指示器の認識に 41[%]と半分近くかかっているため、この部分のネットワークの軽量化を行うことで処理時間の向上を目指していきたい。

5. 結言

単眼カメラによる方向指示器の認識において、本研究では方向指示器の認識システム全体の自動化と方向指示器の認識精度の向上のため、RNN を使用したアルゴリズムの作成を行い、日中において識別精度約 90[%]を得ることができた。

今後の展望としては、下記のことを目指す。

- ネットワークの軽量化
- 精度の向上
- 夜間での方向指示器の性能評価

参考文献

- [Fröhlich 2014] Fröhlich, B., Enzweiler, M., and Franke, U.: Will this car change the lane?-Turn signal recognition in the frequency domain, in *Intelligent Vehicles Symposium Proceedings*, IEEE, pp.37-42 (2014)
- [Freund 1997] Freund, Y., and Schapire, R. E.: A decision-theoretic generalization of on-line learning and an application to boosting, in *Journal of computer and system sciences*, pp.119-139, (1997)
- [Yoneda 2017] Yoneda, K., Hagi, R., Kuramoto, A., Aldibaja, M., Yanase R., and Suganuma, N.: Convolutional Neural Network Based Vehicle Turn Signal Recognition, in *Journal of Advanced Control, Automation and Robotics(JACAR)*, pp. 102-106 (2017)
- [Krizhevsky 2012] Krizhevsky, A., Sutskever, I., and Hinton, G. E: ImageNet Classification with Deep Convolutional Neural Networks, In *Advances in neural information processing systems(NIPS)*, pp. 1097-1105 (2012)
- [Liu 2016] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., and Berg, A. C.: SSD: Single Shot MultiBox Detector, In *European conference on computer vision(ECCV)*, pp. 21-37 (2016)
- [Cho 2014] Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., and Bengio, Y.: Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation, in *Empirical Methods in Natural Language Processing(EMNLP)*, pp. 1724-1734 (2014)