5P-08

車載カメラ映像における危険の検知

北垣修[1,3]　ワン　ユ[2,3]　加藤ジェーン[2,3]　間瀬健二[2,3]

1. はじめに

現在自動車は社会に広く普及して生活必需品となっているが、その半面、交通事故が大きな問題になっている。そこで近年では安全を目指し、「未然防止」や「危険回避」を支援するシステム開発が進められている。[1]

H19 年に日本自動車連盟（JAF）が行った「これからの安全運転支援」に関するアンケート [2] では、もらい事故や自分のミスによる事故を防ぐことを期待している人は 70〜80%と多くいるが、警告や運転補助などの機能が必要という人は 36.9%と少ないことがわかった。この理由として、運転はなるべく自分の思い通りにしたい、緊急の際でも機械に頼るのは不安であるという、「頼りすぎて運転技量の低下、慢心を懸念。」「誤作動が心配。」など、機械に頼る不安や心配の意見があがっている。以上のことから、運転操作はドライバーを主体としたシステムの需要があるといえる。

本研究では、運転はドライバーに委ね、危険を発見することに重点を置き、ドライバーの前方不注意による事故を減らすために、前方を撮影した車載カメラ映像を、CNN を通して、平常時、注意が必要な時、ブレーキやハンドル操作による事故回避が必要な時の 3 つに分類することを目的とする。

2. 関連研究

最近、世界の自動車メーカーから種々の運転支援システムが次々に製品化され、安全性の向上に役立っている。15 年位前は運転支援の搭載機器といえばカーナビゲーション装置くらいしかなかったが、衝突防止装置（衝突被害軽減ブレーキ）・車庫入れ支援（自動車庫入れ）・居眠り運転検知（2 カメラ式ドライブレコーダ）・車線維持支援（白線検知）・横滑り防止装置など新型車での標準装備や 10 万円程度の費用で後付けできる装置などが製品化されている。[3]

これまで畳込みニューラルネットワーク（Convolutional Neural Network、CNN）[4] は様々な問題、特に静止画に適応されて成果を出してきたが、CNN による映像分類・認識タスクは、計算コストやデータ量の問題からあまり研究が盛んではなかったが、近年はデータも増えつつあり、活発に研究がされるようになってきている。Karpathy ら [5] がスポーツ認識用のデータセット UCF-101 を作成し、CNN でビデオフレームを統合する順序を検証したが、同時に複数フレームを入れるだけでは時間的ダイナミクスの効率的な抽出は難しいことを指摘した。Simonyan ら [6] は CNN にオプティカルフローも入力として与え、モーション情報を抽出する手法を提案し、先のデータセット UCF-101 で 88 %の精度を達成した。この研究では CNN を利用して分類を行う。

3. 提案手法

本研究は、車載カメラの映像において、自動車同士の事故危険度、すなわち事故リスクを 3 段階でレーティングすることを目的としている。危険度の分類としては、危険度１：安全（通常時）、危険度２：注意が必要（危険要因が出現）、危険度３：危険（ブレーキやハンドル操作が必要）の 3 段階に分類する。

3.1 データセット

動画サイト YOUTUBE に公開されている自動車事故の際の車載カメラ映像を収集したものに、危険が開始する時刻（危険度１→２になる時刻）、ブレーキやハンドル操作が必要だと思われる時刻（危険度２→３になる時刻）、事故の瞬間、または車同士が最も接近した時刻をラベリングし、動画の開始から危険が開始する時刻までを危険度１、その時刻からブレーキやハンドル操作が必要だと思われる時刻までを危険度２、さらにその時刻から事故の瞬間、または車同士が最も接近した時刻までを危険度３、としてこれらを分類する。図 1 に危険度 3 の状態の例を示す。この図は左から車が飛び出してきているシーンである。ここで映像において自車が動いていないもの、他車同士の事故で自車に影響があまりないと思われるものは使用しない。

この動画データから、10 フレームの動画をそれぞれの危険度から取り出し、それを 1 つのデータとして分類する。今回の実験で使用したデータ数は、危険度 1 が 3534 個、危険度 2 が 3436 個、危険度 3 が 3324 個である。

1　名古屋大学工学部電気電子・情報工学科 4 年
2　名古屋大学
3　間瀬研究室

Copyright ©2017 Information Processing Society of Japan.
All Rights Reserved.

図 1 危険度３の例

3.2 CNN モデル

本研究では、映像の分類に convolutional neural network(CNN) を使用する。使用するのは、ImageNet[7] という大規模画像データセットを用いた画像認識コンペティションにおいて有用な結果を出した通称 Alexnet[8] を,ImageNet を用いて学習した (pre-trained) ものを fine-tuning して使用する。元の Alexnet の入力次元は RGB カラー 3 チャンネルの 227*227*3 の画像であるが、今回の分類対象は画像ではなく動画なので、画像 10 フレームを 3 次元方向に連結して、227*227*30 の 30 チャンネルの画像として入力する。

4. 評価実験

Alexnet の ImageNet による pre-trained モデルを利用した fine-tuning する実験を行った。データセットの約 1/3 をテストデータ、残りを訓練データとして、Data Augmentation は、ランダムクロッピングとランダム左右反転を行った。エラー率の遷移は図 2 のようになった。最終的な分類精度は 50%程度であることがわかる。

4.1 実験考察

学習があまり進んでいないのは、動画の認識タスクなのに、入力は 10 フレームの画像であるため、動画としての特徴をあまり捉えられていないためだと思われる。また、train と val の精度の差が大きいのは、データ数の少なさが影響していると思われる。

5. まとめと今後の課題

本研究では車載カメラ映像において自動車の事故リスク (危険度) の分類を行った。今後の課題としては、オプティカルフローを用いてより時空間情報を学習できる CNN を使用する、使用する 10 フレームのデータの取り方を変えてデータ数を増やす、などして精度を向上させる必要がある。

6. 謝辞

本研究の一部は，科学研究費補助金基盤研究 (B)(No. 26280057)，挑戦的萌芽 (No. 16K12460)，国立研究開発法人科学技術振興機構 (JST) の研究成果展開事業「センター・オブ・イノベーション (COI) プログラム」の支援によるものである。

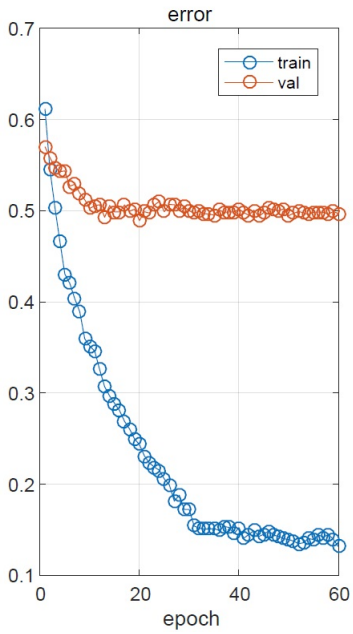

図 2 fine-tunig のエラー率の遷移
下の線が train、上が val

参考文献

[1] 国土交通省：ASV（先進安全自動車）｜自動車総合安全情報，入手先 ⟨http://www.mlit.go.jp/jidosha/anzen/01asv/index.html⟩
[2] JAF：「これからの安全運転支援」アンケート結果，入手先 ⟨http://www.jaf.or.jp/eco-safety/safety/data/info/0804enq/⟩
[3] 科学技術動向センター: 科学技術動向 2013 年 1・2 月号，科学技術政策研究所 科学技術動向研究センター
[4] Y. LeCun, B. Boser, J.S. Denker, D. Henderson, R.E. Howard, W. Hubbard, and L.D. Jackel : *Backpropagation applied to handwritten zip code recognition*, Neural computation, vol.1, no.4, pp.541 551, 1989.
[5] Andrej Karpathy, George Toderici, Sanketh Shetty, Thomas Leung, Rahul Sukthankar, Li Fei-Fei: *Large-scale Video Classification with Convolutional Neural Networks*, The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 1725-1732
[6] Karen Simonyan , Andrew Zisserman : *Two-Stream Convolutional Networks for Action Recognition in Videos*, NIPS2014
[7] ImageNet 入手先 ⟨http://www.image-net.org/⟩ :
[8] Alex Krizhevsky,Ilya Sutskever, Geoffrey E. Hinton : *ImageNet Classification with Deep Convolutional Neural Networks*, NIPS2012

Copyright ©2017 Information Processing Society of Japan.
All Rights Reserved.