平成27年度自動車安全運転センター交通安全等に関する調査研究

事業用自動車の安全対策に資するもの 報告書

ドライブレコーダ映像からの

映像情報を用いた危険度推定

平成 28 年 5 月

九州工業大学大学院

株式会社 堀場製作所

榎田 修一

石倉 理有

報告書要旨

近年,交通事故時の映像を記録する機器であるドライブレコーダが,事業用車両,一般 用車両問わず普及している.事業用自動車へ本機器を装備する際は,国土交通省自動車交 通局が提供するドライブレコーダによる事故防止マニュアル「映像記録型ドライブレコー ダ活用手順書」にも示されるように,交通事故の状況把握のみならず安全運転管理への活 用が期待される.しかし,加速度センサにより検知した急制動によって映像記録のトリガ をかける方式のドライブレコーダにより記録された映像の全てが安全運転の評価に利用で きることは少なく,一般的には道路の凹凸などによる車両の大きな振動に起因するトリガ により得られた環境データを多く含む.よって,安全運転管理におけるドラレコ活用の現 状は,オペレータによる目視による事故・ニアミス状況の確認が必要であることで莫大な コストを要する.さらに,危険と考えられるいくつかの運転状況,例えば一時停止不履行 や信号無視には急な加減速を含まないものも多く,安全運転管理の現場では常時映像を記 録することが要求されている.

本研究では、映像記録型ドライブレコーダが記録した映像情報に対し画像処理技術を活 用することで交通状況を自動認識することで運転危険度を推定し、事故類別や運転マナー の確認を効率良く行えるシステムの開発を目指す. 平成 27 年度の調査研究においては,目 標システムの実現に不可欠となる車両の検出、および道路標識の検出について調査を行っ た.本報告書では,車両検出には Deformable Part Model(DPM)に基づく物体検出アル ゴリズムを適用した.適用対象としては農工大 TLO が提供するヒヤリハットデータ(2013) 年度、および2014年度の約2万件)とし、タグ付けされた情報から追突事故に絞り、危険 度が中以上の 506 件に対して検証実験を行った.しかし,一般的に利用される DPM アル ゴリズムでは、ドライブレコーダ映像における車両検出精度が低く、活用にはさらなる改 善が必要であることを確認した.本報告書では、著者らにより提案されたアルゴリズムの 改善手法について紹介し,車両検出精度が向上したこともあわせて示ししている. 道路標 識検出には深層学習(Deep Learning)を用いた.一般に利用される SIFT マッチングに基 づく標識検出に比べ,頑健に標識の検出が実現できることを確認した. しかし,一時停止 の道路標識への応用を検討した場合、停止すべき場所と一時停止の標識の位置との関係は 自由度が高く、標識を検出するだけでは不十分であることも判明した。そこで、本報告書 では,停止すべき交差点映像そのものを Deep Learning により学習することで,上記の問 題を解決することを提案し,識別性能を評価した.ただし,後者の手法については,実用 にはさらなる精度の向上が望まれることを確認するに留まった.

本報告書では、1章に背景、2章にドライブレコーダ記録映像からの車両検出、3章にド ライブレコーダ記録映像からの道路標識検出について述べ、4章にまとめを述べている.

目 次

第1章	はじめに	3
第2章	ドライブレコーダ記録映像からの車両検出	5
2.1	Deformable Part Model	5
	2.1.1 Deformable Part Model における検出スコアの計算	5
2.2	Latent SVM による Deformable Part Model の構築	7
	2.2.1 ルートフィルタの初期化	8
	2.2.2 ルートフィルタの更新	9
	2.2.3 パートフィルタの初期化	9
	ルートフィルタ内のエッジ強度計算	10
	パートフィルタ生成領域の選択	10
	パートフィルタ内のエッジ強度更新..............	12
	2.2.4 モデルの更新	15
2.3	拡張 Deformable Part Model を用いた車両検出実験	15
	2.3.1 車両データセットを用いた定量評価実験	15
	評価アルゴリズム................................	16
	実験結果	18
	2.3.2 ドライブレコーダデータを用いた定性評価実験	20
	実験環境設定	20
	実験結果	20
	2.3.3 ドライブレコーダデータを用いた追突検出実験	22
	評価方法	22
	追突検出実験結果	23
	2.3.4 考察	25
2.4	まとめ	25
		~ -
第3章	ドライフレコータ記録映像からの標識検出	27
3.1	量み込みニューフルネットリーク	27
	3.1.1 ニューフルネットワーク	27
	活性化関数	27
	人ナツフ 関数	27
	シクモイド関数	28
	単純パーセブトロン	28
	多層パーセプトロン	29

		学習最適化	29
	3.1.2	畳み込みニューラルネットワークの概要::::::::::::::::::::::::::::::::	30
	3.1.3	畳み込み層::::::::::::::::::::::::::::::::	31
	3.1.4	プーリング層::::::::::::::::::::::::::::::::	32
3.2	畳み込	みニューラルネットワークを用いた道路標識検出実験:	33
	3.2.1	道路標識の検出実験::::::::::::::::::::::::::::::::	33
		定性評価実験 : : : : : : : : : : : : : : : :	34
	3.2.2	一時停止すべきシーンの検出実験:	35
		定性評価実験	37
	3.2.3	考察	37
	3.2.4	まとめ	38

第4章 まとめ

第1章 はじめに

近年,交通事故時の映像を記録する機器であるドライブレコーダが,事業用車両,一般用 車両問わず普及している.事業用自動車へ本機器を装備する際は,国土交通省自動車交通局 が提供するドライブレコーダによる事故防止マニュアル「映像記録型ドライブレコーダ活用 手順書」にも示されるように,交通事故の状況把握のみならず安全運転管理への活用が期待 される.しかし,加速度センサにより検知した急制動によって映像記録のトリガをかける方 式のドライブレコーダにより記録された映像の全てが安全運転の評価に利用できることは少 なく,一般的には道路の凹凸などによる車両の大きな振動に起因するトリガにより得られた 環境データを多く含む.よって,安全運転管理におけるドラレコ活用の現状は,オペレータ による目視による事故・ニアミス状況の確認が必要であることで莫大なコストを要する.さ らに,危険と考えられるいくつかの運転状況,例えば一時停止不履行や信号無視には急な加 減速を含まないものも多く,安全運転管理の現場では常時映像を記録することが要求されて いる.ただし,常時記録された長時間の映像を効率良く解析するためには,システムによる 自動切り出しが必要であるが,現状では加速度による判断にとどまるため,問題となる上記 の危険運転は抽出することが不可能である.

本研究では,映像記録型ドライブレコーダが記録した映像情報に対し画像処理技術を活用 することで交通状況を自動認識し,運転危険度を推定したのち事故類別や運転マナーの確認, 安全運転教育を効率良く行えるシステムの開発を目指す(図1参照).本システムでは,常 時記録された映像を加速度によるトリガ処理が行われた部分映像と,常時記録された全ての 映像データとに分けて処理することを検討している.加速度によるトリガ処理がなされた映 像前後では,急制動を引き起こす危険運転がなされたと判断し,回避対象を画像処理により



図 1.1 ドライブレコーダ記録映像による画像処理を活用した安全運転教育システム

検出することで事故・ニアミスを自動類別することが可能となると期待される.回避対象 は車両,歩行者等様々あり得るが,本報告書では車両の検出に焦点を当てて自動化の可能性 を検討した.第2章では,車両検出には Deformable Part Model (DPM)に基づく物体検 出アルゴリズムを適用した.適用対象としては農工大 TLO が提供するヒヤリハットデータ (2013 年度,および 2014 年度の約2万件)とし,タグ付けされた情報から追突事故に絞り, 危険度が中以上の 506 件に対して検証実験を行った.しかし,一般的に利用される DPM ア ルゴリズムでは,ドライブレコーダ映像における車両検出精度が低く,活用にはさらなる改 善が必要であるため,本報告書では,アルゴリズムの改善手法について紹介し,ドライブレ コーダ記録映像からの車両検出精度の比較実験を行う.

次に、常時記録された全ての映像データに関する画像解析について検討する.第2章で紹 介する車両検出の技術を応用することで車間距離の推定が可能となることが期待される.ま た、道路標識の認識を行うことで、一時停止の履行、不履行も評価可能となる.以上から、 急制動を含まない危険な運転について自動検出が可能となると期待される.第3章では道路 標識検出には深層学習(Deep Learning)を用いた.一般に利用されるSIFTマッチングに 基づく標識検出は解像度やダイナミックレンジが高いカメラで記録された映像において高い 検出能力を発揮するが、記録映像の質が必ずしも高くないドライブレコーダにより記録さ れた映像からの道路標識検出には性能が不十分であった.そこで、本報告書では道路標識検 出に Deep Learning を活用し、頑健な道路標識検出を目指す.しかし、一時停止の道路標 識検出を行った結果、停止すべき場所(停止線や歩道の手前)と一時停止の一路標 識検出を行った結果、停止すべき場所(停止なき場所で停止したかの判断には不十分 であることも懸念される.そこで、本報告書では、停止すべき交差点映像そのものを Deep Learning により学習することで、上記の問題を解決することを提案し、識別性能を評価する.

本報告書では、1章に背景、2章にドライブレコーダ記録映像からの車両検出、3章にド ライブレコーダ記録映像からの道路標識検出について述べ、4章にまとめを述べている.

第2章 ドライブレコーダ記録映像からの車両 検出

安全運転支援システムにおける物体検出の精度向上を目指すためには、主な検出対象とし て考えられる人や車両に対する検出精度が問題となる.人は様々な姿勢をとることがあり、 車両は車種により形状が異なる.そのような人の姿勢変化や、車両の車種による形状変化に 頑健な検出手法として、Deformable Part Model(DPM)[2][3]を用いた物体検出がある.

2.1 Deformable Part Model

Deformable Part Model(DPM)の最大の特徴は、検出対象全体を捉える一つのルートフィ ルタとルートフィルタ内の特徴的な局所領域を捉える複数のパートフィルタという2種類の フィルタによって構成され、検出時にパートフィルタが配置を変えることにより姿勢変化や 形状変化に頑健な検出を実現していることである. DPM による人物検出例を図 2.1 に示す. 検出には INRIA Person Dataset[1]を用いた. 図 2.1 中の水色矩形がルートフィルタ、黄色 矩形がパートフィルタを表す.



図 2.1 DPM を用いた INRIA Person Dataset[1] における人物検出例

2.1.1 Deformable Part Model における検出スコアの計算

本節では、DPM を用いた検出を行う際に計算するスコアについて詳述する. DPM では Histograms of Oriented Gradients(HOG) 特徴量 [1] が用いられている. DPM を用いた検 出では、パートフィルタはルートフィルタ内の特徴的な局所領域を捉える必要があるため、 パートフィルタではルートフィルタの2倍の解像度で HOG 特徴量を算出する.そこで,図 2.2 に示す HOG ピラミッドを定義する.HOG ピラミッドは,異なる解像度の画像と HOG 特徴量をピラミッドのように並べたもので,上層ほど低い解像度画像,下層ほど高い解像度 画像を用いる.



図 2.2 HOG ピラミッド (画像は VOC2007[9] より)

HOG ピラミッドを *H*,あるセル配置を p = (x, y, l)とする. *l*は HOG ピラミッドの階層 レベルを表す.すると,HOG ピラミッド *H* においてセル配置 p での HOG 特徴ベクトルは $\phi(H, p)$,フィルタ *F* のスコアは $F \cdot \phi(H, p)$ と表せる. ここで、モデルの*n*個のフィルタ配置を $z = (p_0, ..., p_n)$ とし、式 (2.1) に基づき検出スコ アを計算する.また、式 (2.1) 中の $(\tilde{x}_i, \tilde{y}_i)$ は式 (2.2) により計算される.図 2.3 にパートフィ ルタの理想位置からの移動量 \tilde{x}_i, \tilde{y}_i を示す. a_i はパートフィルタの移動方向、 b_i は移動距離 の係数を表す 2 次元ベクトルである.

$$score(\boldsymbol{p}_{0},...,\boldsymbol{p}_{n}) = \sum_{i=0}^{n} \boldsymbol{F}_{i} \cdot \boldsymbol{\phi}(H,\boldsymbol{p}_{i}) - \sum_{i=1}^{n} \{\boldsymbol{a}_{i} \cdot (\tilde{x}_{i}, \tilde{y}_{i}) + \boldsymbol{b}_{i} \cdot (\tilde{x}_{i}^{2}, \tilde{y}_{i}^{2})\}$$
(2.1)

$$(\tilde{x}_i, \tilde{y}_i) = \boldsymbol{v}_i + (x_i, y_i) - 2(x, y)$$

$$(2.2)$$



図 2.3 パートフィルタの移動量

図 2.3 から, $(\tilde{x}_i, \tilde{y}_i)$ は F_i の理想位置からフィルタ F_i までのパートフィルタの移動量を示 すべクトルである.ここで,式 (2.3)と式 (2.4)を定義する.

$$\boldsymbol{\beta} = (F_0, ..., F_n, a_1, b_1, ..., a_n, b_n)$$
(2.3)

$$\boldsymbol{\varphi} = (\phi(H, p_0), ..., \phi(H, p_n), (\tilde{x_1}, \tilde{y_1}), (\tilde{x_1}^2, \tilde{y_1}^2), ..., (\tilde{x_n}, \tilde{y_n}), (\tilde{x_n}^2, \tilde{y_n}^2))$$
(2.4)

すると,式 (2.1) は式 (2.3) と式 (2.4) の内積 $\beta \cdot \varphi(H,z)$ で表すことができる.

2.2 Latent SVM による Deformable Part Model の構築

本章では, Latent SVM による DPM の構築について詳述する. DPM の構築において, 学 習用データセットは PASCAL Visual Object Classes の VOC2007[9] が用いられている. 各 学習画像には annotation ファイルが用意されており,画像中にどのようなカテゴリの物体 がどの領域に写っているかという情報が記述されている. annotation ファイルの情報を可視 化した学習画像の例を図 2.4 に示す.画像中の物体全体を囲むように bounding box が表示 されている.



図 2.4 VOC2007[9] の例

2.2.1 ルートフィルタの初期化

ルートフィルタの初期化において, bounding box の情報をもとに学習を行う.まず,学 習を行うカテゴリの全ポジティブサンプルの bounding box 情報を取得する.車両カテゴリ における bounding box の例を図 2.5 に示す.



図 2.5 VOC2007[9] の車両カテゴリにおける bounding box の例



図 2.6 切り取ったポジティブ画像の例

その後, bounding box の情報から計算されたアスペクト比の統計値により, ルートフィ ルタのサイズを決定する.次に, 潜在変数を用いずに通常の SVM によってルートフィルタ の学習を行う.このとき, 学習画像から bounding box の領域を切り取った画像をポジティ ブ画像とし, ポジティブ画像を決定されたルートフィルタのサイズにリサイズしたものを使 用する.ネガティブ画像は, ネガティブサンプル画像からランダムに切り取った画像を使用 する.学習画像から bounding box の領域を切り取ったポジティブ画像の例を図 2.6 に示す.

2.2.2 ルートフィルタの更新

ルートフィルタの更新では、2.2.1 節により決定されたルートフィルタの再学習を行う.こ こではポジティブ画像のリサイズは行わず、リサイズ前のポジティブ画像を用いて学習を 行う.

2.2.3 パートフィルタの初期化

パートフィルタの初期化は、以下の (1)~(3) のステップで行う. p 個のパートフィルタ生 成時には (1)~(3) を p 回繰り返す.

- (1) エッジ強度の総和が最大となる局所領域を探索
- (2) (1) で選択された領域にパートフィルタを配置
- (3) パートフィルタ内のエッジ強度を0に更新

パートフィルタの初期化の流れを図2.7に示す.



パートフィルタ領域の再探索

図 2.7 パートフィルタの初期化

2.2.3 節ではステップ(1), 2.2.3 節ではステップ(3) における処理手順を詳述する.

ルートフィルタ内のエッジ強度計算

2.2.3 節に示す処理の前準備として,まず学習されたルートフィルタ内のエッジ強度計算 を行う.計算結果は学習するアスペクト比ごとに配列として出力され,エッジ強度の強い領 域ほど配列要素の値は大きくなる.学習するルートフィルタのサイズが4×11 セルのとき, パートフィルタ生成時にはルートフィルタの2倍の解像度でHOG特徴量を計算するため, 計算結果は8×22 セルの配列として出力される.実際に使用された学習画像例と,学習に よって計算された初期エネルギーのヒートマップを図2.8 に示す.





図 2.8 VOC2007[9] の学習画像例と初期エネルギーヒートマップの比較

エッジ強度が強い領域ほど赤に近い色で,エッジ強度が弱い領域ほど青に近い色で表示されている.図 2.8 より,タイヤや屋根付近,フロント部分など,車両として特徴的な領域が ヒートマップに反映されている.

パートフィルタ生成領域の選択

2.2.3 節に示した処理が終わると、次はエッジ強度の総和が最大となる局所領域を探索する. パートフィルタ生成領域の選択において、エッジ強度の総和が最大となる局所領域を探索するために、ルートフィルタ内の全ての領域に6×6セルのフィルタがかけられ、それぞれの領域におけるエッジ強度の総和計算が行われる. 従来手法では平均化フィルタが用いられ、フィルタ内のエッジ強度の配置によらず同じ重み付けが行われる. 提案手法では、従来手法における平均化フィルタをガウシアンフィルタに変更し、フィルタの中心付近に位置するエッジ強度ほど大きな重み付けを行った. 6×6 セルの平均化フィルタとガウシアンフィルタのヒートマップを図 2.9 に示す. 図 2.9 のガウシアンフィルタの標準偏差σは、フィルタサイズをsとしたとき 1.0s とした.



パートフィルタ内のエッジ強度更新

2.2.3 節で選択された領域においてはパートフィルタが配置され,エッジ強度の更新が行われる.ルートフィルタ内のエッジ強度の座標をx,n個目のパートフィルタ生成後のルートフィルタ内のエッジ強度 E_n , E_n 内でのエッジ強度の総和最大領域を R_{n+1} , R_{n+1} の中心座標をpとすると,従来手法におけるエッジ強度更新は式 (2.5)のように示される.また,式 (2.6)に示す関数を図 2.10に示す.

$$E_{n+1} = W(\boldsymbol{x}; \boldsymbol{p}) \times E_n \tag{2.5}$$

$$W(\boldsymbol{x};\boldsymbol{p}) = \begin{cases} 0 & \boldsymbol{x} \in R_{n+1} \\ 1 & otherwise \end{cases}$$
(2.6)



図 2.10 エッジ強度更新関数 (従来手法)

ここで,6×7セルのルートフィルタを例とし,初期エネルギーのヒートマップを図 2.11 に示す.また,従来手法によるエッジ強度の変化を図 2.12 に示す.



12



図 2.12 従来手法でのエッジ強度更新の様子

図2.12より,更新6回目でルートフィルタ内の全ての領域でエッジ強度が0に更新されている.ルートフィルタ内の全ての領域でエッジ強度が0に更新されてしまうと,次に生成されるパートフィルタがエッジ強度を考慮できなくなってしまう.そこで,提案手法ではエッジ強度の更新をガウス分布に基づくエッジ強度の削減に変更した.提案手法におけるエッジ 強度更新を式 (2.7)に示す.また,式 (2.9)に示す関数を図2.13に示す.

$$E_{n+1} = W(\boldsymbol{x}; \boldsymbol{p}, \sigma) \times E_n \tag{2.7}$$

$$W(\boldsymbol{x};\boldsymbol{p},\sigma) = \begin{cases} G(\boldsymbol{x};\boldsymbol{p},\sigma) & \boldsymbol{x} \in R_{n+1} \\ 1 & otherwise \end{cases}$$
(2.8)

$$G(\boldsymbol{x};\boldsymbol{p},\sigma) = 1 - \exp\left(-\frac{(\boldsymbol{x}-\boldsymbol{p})^2}{2\sigma^2}\right)$$
(2.9)



図 2.13 エッジ強度更新関数 (提案手法)

提案手法によるエッジ強度の変化を図 2.14 に示す. このとき使用したガウス分布の標準 偏差 σ は、生成されるパートフィルタのサイズを s としたとき 0.5s とした.



図 2.14 提案手法でのエッジ強度更新の様子

更新6回目のエッジ強度を比較すると、従来手法ではルートフィルタ内の全ての領域で エッジ強度が0に更新されているのに対し、提案手法ではエッジ強度が残っていることが確 認できる.このことから、提案手法におけるエッジ強度更新により、効果的に多くのパート フィルタが配置されることが期待される.

2.2.4 モデルの更新

モデルの更新で学習するパラメータは、フィルタのスコアに関わる重みベクトルβとパートフィルタの配置 *z* の二つである.この二つのパラメータを同時に学習することはできないため以下の手法を用いる.

*β*を固定し,最もスコアの高いパートフィルタの配置 *zi* を求める.

$$z_i = \max_{z \in Z(x_i)} \boldsymbol{\beta} \cdot \boldsymbol{\phi}(H(x_i), z)$$
(2.10)

次に、zを固定し、固定された配置で最も高いスコアとなる βを求める.

$$\boldsymbol{\beta}^{*}(D) = \arg\min_{\beta} \left\{ \lambda ||\boldsymbol{\beta}||^{2} + \sum_{i=1}^{n} \max(0, 1 - y_{i} f_{\beta}(x_{i})) \right\}$$
(2.11)

2.3 拡張 Deformable Part Model を用いた車両検出実験

本章では,従来手法と提案手法を用いた車両検出における定量評価実験,及び定性評価実 験について詳述する.また,実験に用いる従来手法と提案手法を表 2.1 にまとめる.

衣 2.1 従木子伝と従糸子伝							
	エッジ強度の総和計算フィルタ	エッジ強度更新					
従来手法	平均化フィルタ	0 に更新					
提案手法	ガウシアンフィルタ	ガウス分布に基づくエッジ強度削減					

表 2.1 従来手法と提案手法

提案手法において、ガウス分布の標準偏差 σ は生成されるパートフィルタのサイズをsとしたとき、実験的に決定した 0.5s, 0.7s, 1.0s の 3 通りで評価を行った.表 2.2 に、提案手法で用いる標準偏差 σ ごとの設定をまとめる.

表 2.2 提案手法の設定

手法名	標準偏差 σ
提案手法1	0.5s
提案手法2	0.7s
提案手法3	1.0s

2.3.1 車両データセットを用いた定量評価実験

本節では、車両データセットを用いた定量評価実験について詳述する. 評価用のポジティ ブデータセットとして Stanford cars Dataset[8] を用いた. Stanford cars Dataset の例を図 2.15 に示す.



図 2.15 Stanford cars Dataset[8] の例

実際のデータセット中には車両が複数台含まれる画像もあるが、本実験では車両が1台の み含まれる画像を評価用ポジティブ画像として用いた.また、評価用ネガティブ画像はVOC Dataset の VOC2007 から、車両が含まれない画像を用いた.定量評価実験において使用す るデータ数は、ポジティブ画像 1,054 枚、ネガティブ画像 1,054 枚とした.

評価アルゴリズム

ある評価画像1枚に対する評価アルゴリズムの流れを図2.16に示す.



図 2.16 評価アルゴリズムの流れ

評価を行う際、ポジティブ画像においては検出対象を正しく検出できているか確認を行 う必要がある. Stanford cars Dataset において、学習画像の VOC2007 と同様に bounding box の情報が提供されており、出力された検出矩形が正しいものであるか bounding box の 情報との照らし合わせを行う.本実験では、検出矩形と bounding box の重なり率を計算す ることで正しい検出であるか確認を行った.重なり率の計算方法を以下に詳述する.

画像全体, bounding box, 検出矩形 (ルートフィルタ), 重なり領域をそれぞれ図 2.17 の ように図示すると, 検出矩形と bounding box の関係は, 図 2.18 に示すようになる.



図 2.18 重なり領域

bounding box と検出矩形の重なりが図 2.18 の (c) または (d) に示すような関係になると, bounding box, もしくは検出矩形のどちらか一方における重なり率計算だけでは正しい判断 ができない場合がある.また, bounding box と検出矩形のサイズが大きく異なる場合,正 しく検出が行えたと判断するには不十分であると考える.そこで, bounding box と検出矩 形のそれぞれにおいて重なり領域が占める割合を計算し,その二つの値の F 値を重なり率 とし,重なり率がある閾値 $th_{overlap}$ より大きければ正しい検出であると判断した.重なり領 域の面積を $S_{overlap}$, bounding box の面積を S_{bbox} ,検出矩形の面積を S_{ds} とし,重なり領 域が bounding box に占める割合を O_{bbrate} ,重なり領域が検出矩形に占める割合を O_{dsrate} とすると, O_{bbrate} , O_{dsrate} はそれぞれ式 (2.12) と式 (2.13) により計算される.

$$O_{bbrate} = \frac{S_{overlap}}{S_{bbox}} \tag{2.12}$$

$$O_{dsrate} = \frac{S_{overlap}}{S_{ds}} \tag{2.13}$$

また, O_{bbrate} と O_{dsrate} の F 値である F_{overlap} は式 (2.14) により計算される.

$$F_{overlap} = \frac{2 \cdot O_{bbrate} \cdot O_{dsrate}}{O_{bbrate} + O_{dsrate}}$$
(2.14)

評価用ポジティブ画像を用いた検出において、 $th_{overlap} = 0.8$ とし、 $F_{overlap} > 0.8$ を満たさない検出矩形は誤検出であると判断した.

実験結果

2.3.1 節に示したアルゴリズムによる評価結果を,生成するパートフィルタ数 N_{parts} ごと に示す.





提案手法における標準偏差 σ の値によらず、実験を行った N_{parts} の全てにおいて、検出 精度の向上が確認できる.

ドライブレコーダデータを用いた定性評価実験 2.3.2

本節では、2.3.1節によって車両検出の精度向上が確認できた提案手法の定性評価実験に ついて詳述する.評価用データは、農工大TLO(http://www.tuat-tlo.com/)より提供さ れたドライブレコーダ記録映像を用いた.また,生成されるパートフィルタ数は8個とし, 従来手法と表 2.2 の提案手法1の比較を行った.

実験環境設定

本実験では、ドライブレコーダにより危険であると判断されたトリガー付近での車両検出 を行う. 検出範囲は, 図 2.22 の矩形で囲まれた領域に限定した.



図 2.22 フレーム画像中の検出範囲

実験結果

車両が検出されたフレーム画像に対して、図 2.23 に示すように検出された車両の向きに より検出矩形の色を変化させた.実験結果を図 2.24 に示す.



横向き

斜め向き



正面または後ろ向き

図 2.23 検出矩形の描写例



(a) 従来手法による検出例1

(b) 提案手法による検出例1



(d) 提案手法による検出例 2



- (e) 従来手法による検出例 3
- - (f)提案手法による検出例3



- (g) 従来手法による検出例 4
- (h) 提案手法による検出例 4

図 2.24 定性評価実験結果

2.3.3 ドライブレコーダデータを用いた追突検出実験

本節では、2.3.2 節において使用したドライブレコーダ記録映像から追突と分類されたデー タのみを用いて従来手法と表 2.2 の提案手法 1 の比較を行った.また、生成されるパートフィ ルタ数は 8 個とした.実験環境設定は 2.3.2 節と同様とした.ドライブレコーダ記録映像の 検索画面を図 2.25 に示す.

DRヒヤリハットデ	ータペース検索画面	自事動作の「対象のの」同田の時後「清路形法の」その後(ソー	ersion(160.1.(2015/0	2727) MCOR (Soil
分類選択 とマリハットレベル 「マオペて 「マオレベル 「 モレベル 「 低レベル 「 仮定レベル 「 軍故	対象の 「すべて 「東両 「オーリ54 「自転車 「歩行者 「単快 「その他	 ● □ 直通 ● □ 直通 ● 古市 ● 古市 ● 日前 ● 日前 ● □ 自用の目行状態 	r en fræmgen f		
自重動作0 ▽ すべて 「 直進 「 右折 「 左折 「 軍線変更 「 その他	- ビヤリハット類型(3) 「すべて 一 すべて 一 出合類 「 飛び出し 「 右道 「 すれちがい 「 その他				
道路形状@ (* すべて (* 文差点 (* 文差点以外	-地域 マ すべて 「 都内 「 毎回 「 毎回		自重動作の条件はの	R結合で検索しま	. 7 .
カメラ マ すべて 厂 1カメラ 厂 2カメラ	□ 秋田 □ 北海道 □ 愛知 □ その他	検索条件 検索条件保存 とマリハットレベルー 一 やしベル 中レベル 加 一 一 一 一 一 一 一 一 一 一 一 一 一			
ID 厂 ID検索を使う		1月197 年46月1日20月77 1日大 視聴な検索条件を指定すると、検索時間がかかることがあります			
		ヘルプ 旧詳細検索を使う 検索	実行の	17 総	7

図 2.25 DR ヒヤリハットデータベース検索画面

図 2.25 より, ヒヤリハットの度合いを示すヒヤリハットレベルは, 高, 中レベルとした. ヒヤリハットの類型は今回着目する追突とした. 対象は車両とした. 自車動作, 道路形状, カメラ, 地域についてはすべての場合とした. 選定の結果, 509 件のデータを取得した. し かし, そのうち3件は車両内の映像が映っていたため, 対象外のデータとし, 結果 506 件の データを実験対象とした.

評価方法

本実験では従来手法と提案手法を用いて検出を行い,結果の比較を行った.図 2.26 に評 価例を示す.

ID	時間帯	天気	先行車検出	検出比較	誤検出	特徴(色/車種)
36	昼	晴	0	0	0	
48	夜	晴	×	—	0	シルバー/普通
50	夜	晴	×	—	Δ	黒/軽
77	昼	晴	×	—	Δ	黒/普通
93	夜	晴	×	—		黒/タクシー
94	昼	晴	×	—	Δ	黒/タクシー
110	昼	同時	0	0	0	
239	昼	晴	0	\triangle	0	
259	夜	晴	×	—	0	シルバー/軽
288	夜	晴	×	_	_	白/タクシー
302	尼	量	Ó	Δ	Ó	

図 2.26 評価例

評価データには, ID, 時間帯, 天気, 先行者検出, 検出比較, 誤検出, 特徴の属性を記録する.時間帯は昼, 夜, 夕のいずれかを記入する. 天気は晴, 雨, 曇, などを記入する. 続いて, トリガー要因となった車両を検出できたかどうかを判定する. 先行車検出の欄は, 検出できた場合は〇, できなかった場合は×とした. このとき, ×であった場合は, 特徴の 欄に対象車両の色や車種を記入する. 次に, 従来手法と提案手法の検出精度の比較を行う. 従来手法より検出精度が向上した場合は〇, 同程度の場合は△, 精度が落ちた場合は×とした. 最後に, 誤検出の増減を調べる. 従来手法に比べて誤検出が減少した場合は〇, 変化な しの場合は△, 増加した場合は×とした.

追突検出実験結果

時間帯別の評価結果を図 2.27 に示す.検出可能な例について,検出比較と誤検出の項目が 共に○であった場合を「向上」,検出比較,誤検出のどちらかが○であり,片方が△であっ た場合を「一部向上」,どちらも△であった場合を「同精度」として分類を行う.

			昼	夜	タ	合計
	向上		19	2	0	21
	一部向上	検出比較〇 誤検出△	22	11	0	33
探山り		検出比較△ 誤検出〇	51	6	2	59
	同精度		170	39	7	216
検出不可			68	103	6	177
時間帯別合計			330	161	15	506

図 2.27 評価結果

また,「向上」として評価されたシーンに対して,検出例を図 2.28 に示す. 出力される検 出矩形の色は, 2.3.2 節の図 2.23 と同様とした.



- (a) 従来手法による検出例1
- (b) 提案手法による検出例1



(c) 従来手法による検出例 2



(d) 提案手法による検出例 2



(e) 従来手法による検出例 3



(f) 提案手法による検出例 3



(g) 従来手法による検出例4

(h) 提案手法による検出例 4

図 2.28 追突検出実験結果

2.3.4 考察

従来手法と提案手法を用いた車両検出における検出精度について考察する. 生成される パートフィルタ数を8個と固定したとき、定量評価実験より従来手法を用いた車両検出にお ける検出精度は、誤り受け入れ率が10%のとき検出率が99%以上と高精度であることが確 認できる.また,提案手法を用いた車両検出においては,誤り受け入れ率が 10%のとき従 来手法と比較して検出精度が約0.3%向上している.検出率が99%以上の手法に変更を加え, 検出精度の向上が確認できたことは,提案手法において拡張を行ったパートフィルタ生成ア ルゴリズムが検出精度に関係していることを示す.さらに,定性評価実験より,従来手法を 用いた車両検出では検出が確認できなかったフレームにおいても、提案手法を用いると検出 が行えたことを確認した、従来手法と提案手法での検出に差が生じた車両の特徴として、車 両の隠れや切れが生じていること、設置されたカメラの傾きにより検出対象である車両が傾 いて撮影されていること、車両の色が黒であるなどが挙げられる。また、検出時に出力され る矩形から、提案手法を用いた検出時に従来手法と比較し、本来検出してほしい車両の向き を正しく判定していることが確認できた.加えて.ドライブレコーダ記録映像を用いた追突 検出実験において,全 506 件のデータ中,先行者検出が可能なシーンは 329 件,検出不可な シーンは 177 件となり,全体的な検出率は約 65%となった.分類の結果,「向上」が 21 件, 「一部向上」が92件となり,検出可能なシーンのうち約34%において提案手法による検出性 能の向上が確認された.以上のことから,定量評価実験においてはわずかな精度向上の確認 に留まったが、ドライブレコーダ記録映像の解析においては定量評価実験の結果以上に検出 精度の向上が得られていると考えられる.

2.4 まとめ

本章では、DPM におけるパートフィルタ生成条件に注目した拡張 DPM を提案し、生成 されるパートフィルタの個数によらず車両検出精度の向上が確認できた.また、ドライブレ コーダ記録映像を用いた追突検出実験において、多くのシーンで従来手法と比較し検出精度 の向上を確認した.今後は、提供されたドライブレコーダ記録映像においての追突検出に関 して詳細な解析を進め、追突以外のタグ付けが行われているシーンに対しても同様の解析を 進めていきたい.

参考文献

- [1] N.Dalal, B.Triggs, "Histograms of oriented gradients for Human Detection", IEEE Conference on Computer Vision and Pattern Recognition, 2005.
- [2] P.Felzenszwalb, R.Girshick, D.McAllester, D.Ramanan, "Object Detection with Discriminatively Trained Part Based Models", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.32, No.9, Sep. 2010.
- [3] P.Felzenszwalb, D.McAllester, D. Ramanan, "A discriminatively trained, multiscale, deformable part model", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2008.
- [4] 山本国起, "Deformable Part Model におけるパラメータ設定の自動化に関する研究", 九州工業大学卒業論文, 2015.
- [5] Eduard Trulls, Stavros Tsogkas, Iasonas Kokkinos, Alberto Sanfeliu, Francesc Moreno-Noguer, "Segmentation-aware Deformable Part Models", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 168-175.
- [6] Yicong Tian, Rahul Sukthankar, Mubarak Shah, "Spatiotemporal Deformable Part Models for Action Detection", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2013, pp. 2642-2649.
- [7] Fidler, Sanja and Sven Dickinson and Urtasun, Raquel, "3D Object Detection and Viewpoint Estimation with a Deformable 3D Cuboid Model", Advances in Neural Information Processing Systems 25, 2012, pp. 611-619.
- [8] Jonathan Krause, Michael Stark, Jia Deng, Li Fei-Fei, "3D Object Representations for Fine-Grained Categorization" 4th IEEE Workshop on 3D Representation and Recognition, at ICCV 2013 (3dRR-13). Sydney, Australia. Dec. 8, 2013. http://ai.stanford.edu/~jkrause/cars/car_dataset.html
- [9] M.Everingham, L.Van Gool, C.K.I.Williams, J.Winn, A.Zisserman, "The PASCAL Visual Object Classes Challenge 2007" http://host.robots.ox.ac.uk/pascal/VOC/voc2007/

第3章 ドライブレコーダ記録映像からの標識検出

3.1 畳み込みニューラルネットワーク

畳み込みニューラルネットワーク (CNN) は,画像認識で用いられている多層 NN である. 本章では,CNN の基礎となるニューラルネットワーク (NN) の基本構造を説明し,パーセ プトロンと CNN の構造や畳み込み層,プーリング層について詳しく説明する.

3.1.1 ニューラルネットワーク

NNとは、人間の情報処理システムの中枢である脳神経系を模擬したアルゴリズムである. NNの研究は1940年代に始まり、1950年代から60年代にかけて盛んに研究が行われた. Minskyなどの人工知能で有名な人物も脳やニューラルネットワークの研究を行っている. NNの特徴は学習能力であり、学習を繰り返すことで高精度な識別が可能である. NNの基本的な構成であるニューロンモデルについて以下に説明する. ニューロンモデルとは、ニューロンと呼ばれる神経細胞が、シナプスと呼ばれる結合部位を介して多数結合してできたネットワーク構成である. ニューロンの人工的なモデルのことをユニットと呼び、NNは、ユニットを重みで繋げることによってネットワークを構成する. ユニット間の信号は一方方向で伝わり、入力 $x = \{x_1, ..., x_n\}$ に重み $w = \{w_1, ...w_n\}$ を与え、結合先ユニットに入力される. 入力された重みつきの信号 wxの総和を求め、活性化関数 f に与え、出力 y を求める. 出力 y は式 (3.1)のように表される.

$$y = f(\sum_{i=1}^{n} w_i x_i)$$
 (3.1)

活性化関数

活性化関数 f には様々な種類があり、ネットワーク構造や入力に応じて用いる活性化関数 は異なる.一般的に活性化関数は非線形の関数であり、以下に活性化関数について説明する.

ステップ関数

ステップ関数を式 (3.2) に示す.ステップ関数は初期のパーセプトロンの活性化関数とし て用いられている.式 (3.2) は入力 *x* に対して, *x* が 0 以上であれば出力 *y* は 1 であり, *x* が 0より小さければ出力 y は0となるような活性化関数である.

$$f(x) = \begin{cases} 1 & if \quad x \ge 0\\ 0 & if \quad x < 0 \end{cases}$$
(3.2)

シグモイド関数

シグモイド関数とは,式(3.3)に示すような関数である.

$$f(x) = \frac{1}{1 + e^{-\alpha x}}$$
(3.3)

シグモイド関数とは、どのような入力 *x* に対しても必ず 0 から 1 の値が出力される.式 (3.3)のαはゲインと呼ばれ、αはシグモイド関数の曲線の傾きを変更するパラメータであ る.αの値が大きいほどシグモイド関数の傾きは急になり、逆にαの値が小さいとシグモイ ド関数の傾きは緩やかになる.シグモイド関数がパーセプトロンの識別性能を大幅に向上さ せたためパーセプトロンの活性化関数として使用されている.識別性能を向上させた理由は、 式(3.3)よりシグモイド関数は、微分可能な関数であることが挙げられる.2.5節で説明す る誤差逆伝播法にうまく適用することができることから識別精度の向上につながっている.

単純パーセプトロン

単純パーセプトロンは 1950 年代の終わりに Rosenblatt によって提案されたモデルである. 単純パーセプトロンは,入力層,中間層,出力層の3つの層で構成される.本例のネット ワークでは,入力層のユニットが x_n,中間層のユニットが a_m からなり,出力層は1つのユ ニットで構成される.出力層が1つのユニットであるため分類できるクラス数は2つである. 単純パーセプトロンでは,入力層と中間層の重み w_{ij}を固定の重みとして与える.式 (3.4) に単純パーセプトロンの出力 y を示す.

$$y = f(\sum_{j=1}^{m} w_j a_j - \theta) \tag{3.4}$$

中間層 $a = \{a_1, ..., a_m\}$ に重み $w = \{w_1, ..., w_m\}$ を付与した値の加算値から,バイアス θ の 差を出力層のユニットに入力し活性化関数 f を通して出力 y を得る. 単純パーセプトロン の学習方法を,式 (3.5) に示す. 単純パーセプトロンの出力 y と教師信号 t を用いて学習を 行う.

$$w_j^{new} = \begin{cases} w_j^{old} & if \quad y = t \\ w_j^{old} + \eta(t - y)x_j & if \quad y \neq t \end{cases}$$
(3.5)

ηは学習係数と呼ばれ、重みの更新量を制御する係数であり、ηは $0 < \eta \le 1$ とする.単純パーセプトロンの学習は、指定された学習回数、または誤識別率が0%になるまで行われる.2クラス以上のクラス分類を行うために多層パーセプトロンがある.

多層パーセプトロン

多層パーセプトロンは,非線形の多クラス識別器である.単純パーセプトロンと同様に入 力層,中間層,出力層の3種類の層で構成される.多層パーセプトロンと単純パーセプトロ ンの違いは大きく2つある.一つ目に,単純パーセプトロンは中間層と出力層の重みを学習 していたが,多層パーセプトロンではネットワーク全体の重みに対して学習を行う.二つ目 に,単純パーセプトロンの出力層は1つのユニットであったが,多層パーセプトロンは多ク ラス分類を行うため出力層のユニットを分類したいクラス数だけ用意することである.

入力 $x = \{x_1, ..., x_n\}$ に対する出力 $z = \{z_1, ..., z_p\}$ は以下のように示すことができる.

$$\alpha_j = \sum_{i=1}^n w_{ij} x_i + \theta \tag{3.6}$$

$$y_j = f(\alpha_j) \tag{3.7}$$

中間層の出力 *y_j* を次の層の入力として式 (3.6) に与え,これを中間層の数 *N* まで繰り返 す.出力層の入出力関数への入力は式 (3.8) となり,活性化関数ではなく式 (3.9) の Softmax 関数を使用する.ネットワークの出力結果は式 (3.10) となる.

$$\gamma_k = \sum_{j=1}^l w_{jk} \beta_j + \tau \tag{3.8}$$

$$g(\gamma_k) = \frac{\exp(\gamma_k)}{\sum_{a=1}^p \exp(\gamma_a)}$$
(3.9)

$$z_k = g(\gamma_k) \tag{3.10}$$

 w_{ij} は、i番目の入力から中間層のj番目のユニットへの結合荷重であり、 w_{jk} は、中間層 Nのj番目のユニットから出力層のk番目のユニットへの結合荷重である。 θ , ρ , τ は、各中間層のユニットに対するバイアスである。また、fは、活性化関数であり、式 (3.3)のシグモイド関数がよく使われる。出力層のユニットの入出力関数は多項ロジスティック関数もしくは式 (3.9)の Softmax 関数が使用される。多層パーセプトロンの学習方法は、教師付き学習による誤差逆伝播法を用いた最急降下法で学習する [2].

学習最適化

パーセプトロンの学習則の欠点は出力層のユニットへの結合重みしか修正できないことで あった. 誤差逆伝播法は,出力層以前の中間層のユニットへの結合重みも修正することがで きる [3]. 微分可能な活性化関数 (例として式 (3.3) シグモイド関数)を用いて,出力層での 誤差評価 *R*(*W*)への任意の結合重み*w*の寄与を,偏微分係数 *∂R*(*W*)/*∂w* として計算でき るようにした.これによって最急降下法などの最適化手法により,出力と教師信号との誤差 を最小となるように全ての重みを修正可能とした.層の数を *N*, 第 *n* 層の第 *k* 番目のユニッ トから第n+1層の第j番目のユニットへの結合重みを $w_{k,j}^{n+1}$ としたとき,最急降下法 (確率的降下法)による結合重みの修正を式 (3.11)に示す.

$$w_{k,j}^{n+1} \leftarrow w_{k,j}^{n+1} - \frac{\eta \partial R(W)}{\partial w_{k,j}^{n+1}}$$

$$(3.11)$$

ここで, 誤差評価 R(W) を式 (3.12) のように二乗誤差であるとする.

$$R(W) = \frac{1}{2} \sum_{j} (t_j - y_j^{(N)})^2$$
(3.12)

この時の具体的な修正規則は式 (3.13) となる.

$$w_{k,j}^{n+1} \leftarrow w_{k,j}^{n+1} - \eta \delta_j^{(n+1)} y_k^{(n)}$$
 (3.13)

 t_j は教師信号, $y_j^{(N)}$ は第 N 層の第 j 番目のユニットの出力, $y_k^{(n)}$ は第 n 層の第 k 番目のユニットの出力とする. $\delta_j^{(n+1)}$ は合成関数の微分法より, $(t_j - y_j^{(N)})$ から計算される. $\delta_j^{(N)}$ を基準として $\delta_k^{(n)}$ を再帰的に計算することで求める.

$$\delta_j^{(N)} = -(t_j - y_j^{(N)})y_j^{(N)}(1 - y_j^{(n)})$$
(3.14)

$$\delta_k^n = \left\{ \sum_{j=1}^{K_{n+1}} \delta_j^{(n+1)} w_{k,j}^{(n+1)} \right\} y_j^{(N)} (1 - y_j^{(n)})$$
(3.15)

この計算過程が,出力層での誤差を前の層に伝播させていく形となっており誤差逆伝播法と 呼ばれる.この誤差逆伝播法を用いて多層パーセプトロンの学習を行い,重みとバイアスの 値を修正する.

3.1.2 畳み込みニューラルネットワークの概要

CNN は, Fukushima らが提案したネオコグニトロン [4] が基礎となっている.ネオコグ ニトロンの構造は,S層(単純細胞層)とC層(複数細胞層)を組合わせた2層の神経回路を 基本要素とする.このS層が入力に対して特徴的なパターンを検出する特徴抽出器の役割を 担っている.C層は,受容野内(S層の小領域)にある最大値を出力する.この処理はS層の 出力結果を集めてC層の出力とすることからプーリングと呼ばれる.CNNの構造は,S層 を畳み込み層とC層をプーリング層として2つの層を対に持つ構造をしている.この2つ の特殊な層を除くと,CNNは順伝播型NNと同様に扱うことができる.入力画像が与えら れたとき,畳み込み層(C1,C2)とプーリング層(P1,P2)で入力画像の特徴を取得し,全結合 ネットワーク(F)に取得した特徴を伝え,クラス識別を行う.



図 3.1 CNN の構造 (文献 [1] より出典)

3.1.3 畳み込み層

CNN を画像認識で用いる際, 畳み込み層は, 入力画像に対してフィルタの畳み込みを行う. これは, 一般的な画像処理での畳み込みであり, 画像をぼかす処理や, エッジを強調する処理と同様である. 畳み込み層でのフィルタの値とバイアス値は, 学習によって最適化される. フィルタのサイズはネットワーク設計時に決め, 学習によって変化しない. 畳み込み 層の処理を図 3.2 に示す.



図 3.2 i 番目に対する畳み込み処理(文献 [1] より出典)

入力を $W \times W$ サイズの画像 N 枚とする. この入力を以下では $W \times W \times N$ とする. 入 力層では,入力画像がグレイスケールの場合 N = 1となり,カラーであるなら RGB より N = 3となる. これ以降の中間層の数は直前の畳み込み層の出力 (フィルタの総数) と同じで ある. 具体的に畳み込み処理を考える.入力 $W \times W$ [pix] 解像度の画像 N 枚に $w \times w$ [pix] サイズのフィルタ N 枚を畳み込む. ここでフィルタを $h(p,q,k)(p \in [1,...,w], q \in [1,...,w])$, $k \in [1,...,N]$)とし,入力を $I(x,y,k)(x \in [1,...,W], y \in [1,...,W], k \in [1,...,N])$,出力を f(x,y,i)とすると,ある i 番目の出力結果は,式 (2.16) で示される.最後の係数 b_i はバイ アスである. b_i は一般的に $b_i = b_{xyi}$ とすることが多い.

$$f(x, y, i) = \sum_{k=1}^{N} \left[\sum_{p,q=1}^{w} I(x+p, y+q, k) h(p, q, k) \right] + b_i$$
(3.16)

3.1.4 プーリング層

プーリング層は畳み込み層と対になるように使用され,畳み込み層の出力がプーリング層 の入力となる.プーリング層の目的は,画像のどの位置でフィルタの応答が強かったかとい う情報を一部捨てることで,画像内に現れる特徴の位置や回転等の微小な変化に対する不変 性を実現することにある.プーリング層では,畳み込み層と違い,計算内容はネットワーク 設計時に決まっており,学習によって変化することはない.通常,プーリングの処理は画像 が持つ情報を間引く処理である.間引く量をスライド*s*とすると,*s* = 2のとき,出力は入 力の半分のサイズとなる.スライド幅*s*と小領域*w*×*w*によっては隣接する出力ユニット の受容野 P_{ij} が重なり合うこともある.プーリング層の入力(畳み込み層の出力)の一部の小 領域 P_{ij} について,この小領域内部のユニット(*p*,*q*) $\in P_{ij}$ の出力 y_{pq} を集約し一つの出力と する.図 3.3 に各プーリングアルゴリズムを示す.プーリングのアルゴリズムは複数ある.

最大プーリング (max pooling) とは,式 (3.17) のように *P*_{ij} に属するユニットの最大値を 選択する.

$$\tilde{y}_{ijk} = \max_{(p,q)\in P_{ij}} y_{pqk} \tag{3.17}$$

平均プーリング (average pooling) とは、式 (3.18) のように P_{ij} に属するユニットの出力 の平均を選択する.

$$\tilde{y}_{ijk} = \frac{1}{|P_{ij}|} \sum_{(p,q)\in P_{ij}} y_{pqk}$$
(3.18)

Lp プーリングとは式 (3.19) は最大プーリングと平均プーリングの中間的な存在として扱われている.

$$\tilde{y}_{ijk} = \left(\frac{1}{|P_{ij}|} \sum_{(p,q) \in P_{ij}} y_{pqk}^P\right)^{\frac{1}{P}}$$
(3.19)

LeCun らの CNN[5] をはじめとする初期の研究では,平均プーリングが主であったが, Serre らの HMAX モデル [6] や Ranzato らの研究 [7] をきっかけに最大プーリングが一般的 になった. [8].



図 3.3 プーリングのアルゴリズム

3.2 畳み込みニューラルネットワークを用いた道路標識検出実験

本章では、道路標識と、一時停止すべきシーンの二つの検出実験を畳み込みニューラル ネットワーク(CNN)を用いて行った結果を報告する.

3.2.1 道路標識の検出実験

CNNを用いて一時停止の標識をどの程度識別可能か実験を行った. CNN には一般物体認 識のコンテストである ILSVRC(ImageNet Large Scale Visual Recognition Challenge)で, 高い識別率を持つ Network In Network[10]を用いた. 学習に用いたデータは,東京農工大 学スマートモビリティ研究拠点より提供して頂いたドライブレコーダー映像から,一時停止 の標識を切り出した画像(図 3.4) 68,300 枚と,同じくドライブレコーダー映像から切り出 した背景画像(図 3.5) 72,725 枚枚の合計 141,025 枚を学習画像とした.テスト画像には学 習画像と同様に切り出した一時停止画像 10,000 枚と背景画像 10,000 枚の計 20,000 枚を用い てテストを行った.

学習の状況を表したグラフが図 3.6 で,繰り返し回数約 10,000 回を超えた辺りで正答率 が 99 %に収束していることが分かる.この結果から,学習が上手く進んだことが分かった. また,テストデータに対しての正答率は,学習が収束したモデルを用いて約 98.4 %という 結果になった.テストデータに対しても,高い識別率で識別が可能である事が分かり,CNN を用いた手法が,一時停止標識検出で高精度な検出が可能である事が分かった.



図 3.4 学習データ中の一時停止標識画像の例 図 3.5 学習データ中の背景画像の例

定性評価実験

学習済みのネットワークを用いて,実際のドライブレコーダー映像に一時停止標識の検出 実験を行った.従来,道路標識検出では SIFT 特徴量 [11] 等の画像特徴量がよく用いられて きた.そこで、実験では、一時停止標識の検出実験において、CNN を用いた手法と、SIFT 特徴量を用いた手法を比較した.



図 3.6 一時停止標識学習の結果

CNNを用いた手法と,SIFT特徴量を用いた手法を比較した結果,図3.7と図3.8より,+ 分な大きさで標識が写っている場合はどちらも検出が行えている.しかし,図3.9と図3.10 より,+分な大きさで標識が写っていない場合,SIFT特徴量を用いた手法は検出が困難で あったが,CNNを用いた手法では正しく検出が行える事が分かった.この結果から,CNN を用いた手法では,低解像度画像からも高精度な道路標識検出が行える事が分かった.

3.2.2 一時停止すべきシーンの検出実験

一般的に,公道に設置されている一時停止標識は,標識の位置と停止線が一意に定まって いない事が多い.速度制限や駐車禁止の場合は,標識を認識した区間から注意を払えば良 い.しかし,一時停止の場合は,標識を認識してから厳密に一時停止すべき場所を知るため



図 3.7 CNN を用いた検出結果



図 3.8 SIFT 特徴量を用いた検出結果





図 3.9 CNN を用いた検出結果

図 3.10 SIFT 特徴量を用いた検出結果



図 3.11 CNN を用いた道路標識検出結果

には、標識の認識だけでは不十分である.また、標識を検出する際は、一定の矩形領域をず らしながら画像をラスタスキャンする必要があり、検出のコストが高くなる.そのため、車 載カメラ映像中の一定の矩形領域内の映像をそのまま学習させ、道路標識の有無に関わら ず、停止すべき位置かどうかの識別を行った. CNN には Network In Network を用い、学 習データには、道路標識の検出実験と同じドライブレコーダー映像から一時停止すべきシー ンを切り出した画像(図 3.12) 12,900 枚と、それ以外のシーンを切り出した画像(図 3.13) 52,000 枚の合計 64900 枚を学習画像とした.テスト画像には学習画像と同様に切り出した 一時停止画像 500 枚とそれ以外の画像 500 枚の計 1,000 枚を用いてテストを行った.

学習の状況を表したグラフが図 3.14 で,繰り返し回数約 10,000 回を超えた辺りで正答率 が 99 %に収束していることが分かる.この結果から,学習が上手く進んだことが分かった. また,テストデータに対しての正答率は,学習が収束したモデルを用いて約 99 %という結 果になった.



図 3.12 一時停止画像の例

図 3.13 一時停止以外の画像の例



図 3.14 一時停止学習の結果

定性評価実験

学習済みのネットワークを用いて,実際のドライブレコーダー映像に一時停止すべきシーンの検出実験を行った.ネットワークへの入力画像には,ドライブレコーダー映像の中央部 256 × 256[pixel]を用いた.実験の結果は以下のようになった.

図 3.15 と図 3.16 の実験結果では、一時停止すべきシーンと検出された場合は赤い矩形、 それ以外は青い矩形で表示される.二つのドライブレコーダー映像に対して検出を行った結 果、一時停止の標識が矩形領域に入っている場合や、標識が写っていない場合でも、道路上 の停止線等に反応し、一時停止すべきシーンであると検出される事が分かった.しかし、図 3.17 から、一時停止以外の標識や、それ以外のシーンでも一時停止すべきシーンであると検 出されてしまう場合がある事が分かった.

3.2.3 考察

一時停止標識の検出実験では、テストデータに対して約 98.4 %の識別率となり、高精度な 一時停止標識検出が可能であることが分かった.定性評価実験では、CNN を用いた手法と



図 3.15 一時停止すべきシーンの検出結果



図 3.16 一時停止すべきシーンの検出結果

SIFT 特徴量を用いた手法を比較し,低解像度画像から一時停止標識を検出する場合の CNN の有効性を確認した.また,一時停止すべきシーンの検出実験では,テストデータに対して 約 99 %の識別率となり,定性評価を行い,一時停止すべきシーンの検出が正しく行える事 を確認した.しかし,一時停止以外の標識,電柱やミラーなどの柱状物体の影響で誤検出し てしまう可能性があり,今後,学習データを増やして再学習することで,より高精度な検出 が期待できる.

3.2.4 まとめ

本章では、CNNを用いた一時停止標識の検出実験と、一時停止すべきシーンの検出実験 の二つの実験を行った.実験には、Network In Networkを用い、ドライブレコーダー映像 から切り出した画像を用いて学習、評価を行った.一時停止標識の検出実験では、高精度な 一時停止標識検出が可能であることが分かった.また、一時停止すべきシーンの検出実験で



図 3.17 一時停止すべきシーン以外での誤検出

は、一時停止すべきシーンの検出が正しく行えるが、誤検出が起きる場合もあるので、今後、 他の標識や電柱などを学習データに加え、さらに時系列情報を用いることで、一時停止すべ きシーンの推定精度向上を目標とする.

参考文献

- [1] 岡谷貴之, 画像認識のための深層学習, 人工知能学会誌 28 巻 6 号, pp.962-974, 2013.
- [2] 麻生英樹, 多層ニューラルネットワークによる深層表現の学習,人工知能学会誌 28 巻 4 号, pp.649-659, 2013.
- [3] Rumalhart, D.E., Hinton, G.E. and Williams, R.J, Learning representations by backpropagating errors, Nature, Vol.323, No.9, pp.533-536, 1986.
- [4] Fukushima, K, Artificial Vision by multi-layered neural networks: Neocognitron and its advances, Neural Netowrks, Vol.337, pp.103-119, 2013.
- [5] LeCun, Y., Bottou, L., Bengio, Y. and Haffner, P., Gradient-based learning applied to document recognition, Proc. IEEE, 1998.
- [6] Serre, T., Wolf, L. and Poggio, T., Object recogniton with features inspired by visual cortex, Proc. CVPR, 2005.
- [7] Ranzato, M.A.,, Bouerau, Y.-L. and LeCun, Y., Sparse feature learning for deep delief networks, Proc.NIPS, 2007.
- [8] Ciresan, D.C., Meier, U., Masci, J., Gambardella, L.M. and Schmidhuber, J. Flexible, high performance convolutional neural networks for image classification, Proc. CVPR, 2012.
- [9] D. G. Lowe, Object recognition from local scale-invariant features, Proceedings of the International Conference on Computer Vision-Volume 2 - Volume 2, pp.1150-, IEEE Computer Society, 1999.
- [10] Min Lin, Qiang Chen, Shuicheng Yan, Network In Network, arXiv preprint arXiv:1312.4400, 2013.
- [11] David G. Lowe, "Distinctive image features from scale-invariant keypoints", Journal of Computer Vision, 60, 2, pp. 91-110, 2004.

第4章 まとめ

本研究は、映像記録型ドライブレコーダが記録した映像情報から運転危険度を推定し、事 故類別や運転マナーの確認,安全運転教育を効率良く行えるシステムの開発を目標とし実施 している.本システムでは、常時記録された映像を加速度によるトリガ処理が行われた部分 映像と、常時記録された全ての映像データとに分けて処理することを検討しており、加速度 によるトリガ処理がなされた映像前後では回避対象を画像処理により検出し、事故・ニアミ スを自動類別する.また,常時記録された全ての映像データに関する画像解析については, 車両検出技術の応用による車間距離推定,道路標識の認識を行うことで,急制動を含まない 危険な運転について自動検出行うことを目標としている.本報告では、これらのモジュール の機能を実現するために必要である車両検出技術,および道路標識技術の調査研究を行った. 車両検出には Deformable Part Model による物体検出を改善,応用し,道路標識検出には 深層学習(Deep Learning)を用いた、本報告書で紹介する技術によって、車両検出、およ び標識検出に関して一般的に活用される技術に対し,性能が改善したことが確認されたが, 夜間の映像等、ドライブレコーダが備えるカメラのダイナミックレンジ不足等で検出が困難 となる場面については、性能が低下することも確認している.また、一時停止の道路標識に ついては、道路標識の検出結果から一時停止すべき場所で停止したかの確認が困難であると 判明した.そこで,一時停止すべき交差点映像自体を Deep Learning によって学習すること で,一時停止不履行の検出を行うことを提案した.

今後の課題としては、車両の自動検出結果から事故を類別するアルゴリズムの検討が必要 である.また、車両検出結果を車間距離推定に応用する技術の開発も必要である.これらの 課題は、車両にあらかじめ取り付けられ十分に校正されたカメラであれば容易に解決可能で あるが、ドライブレコーダは車両に後付けされることも多く、取り付け工程での校正作業が 省かれるため走行映像からの自己校正技術が必要となる.他の課題としては、一時停止すべ き交差点の認識精度向上が挙げられる.本問題は、時系列情報を取り扱い可能な学習器の適 用が考えられる.

最後に、これらの技術が社会で活用されるためには、適用技術が実時間処理可能である必要があり、システムに特化したアルゴリズムの改善や情報技術の更なる応用が必要であると考える.また、画像処理技術の精度に特化して報告したが、今後はこれらの技術を活用する 立場である企業からの率直な意見を反映し、目標システムに対する必要精度を見積り、処理時間とのバランスや計算機環境の構成を含めプロトタイピングすることで、事業者用自動車の安全対策における早期の実利用を目指す.