

物体特徴量に基づく危険運転状況の推定

山本 修平¹ 倉島 健¹ 松林 達史¹ 戸田 浩之¹

概要：ドライブレコーダで記録された交通事故やそれに近い危険な運転状況に関する映像やセンサデータは、ドライバーの安全運転教育や法人車両の運行管理サービスに利用される有益な情報である。このようなドライブレコーダのデータは、車に急な挙動の変化があったことがトリガーとなり記録されるが、段差を乗り越えた際の衝撃等で危険運転ではないデータも数多く記録され混在している。また危険運転を含むとしても、多様なデータが存在するため、内容に応じて類型化されることが望ましい。本論文では、ドライブレコーダデータに対して、危険運転の発生対象ラベルの自動推定のタスクに取り組む。著者らはこれまでも同様のタスクに取り組んできたが、本論文では特に、異なる環境で記録されるデータに対しても頑健な推定を実現するため、前方映像に対して物体検出技術を活用して得られる物体検出結果に着目する。深層学習に基づく既存手法によって物体検出結果を特徴ベクトルに変換するにあたり、本論文では2つの拡張点を提案する。1つ目は物体検出結果から得られる境界領域をもとに、物体の重要度を考慮できる特徴量を算出する。2つ目はその特徴量に基づいて重要物体を優先的に選出し、危険度の高い物体の取りこぼしの可能性を低くする。実際のドライブレコーダデータを用いた評価実験の結果、2つの拡張点を用いた提案手法が、既存手法に比べて高い推定性能を示すことを明らかにした。

Estimating Dangerous Driving Scene based on Object Features

SHUHEI YAMAMOTO¹ TAKESHI KURASHIMA¹ TATSUSHI MATSUBAYASHI¹ HIROYUKI TODA¹

1. はじめに

ドライブレコーダとは、車内に設置され、前方映像、加速度、ウィンカー操作、ブレーキ操作などの車両運行状況を記録する装置である^{*1}。走行中の全時刻の情報を記録することは、記憶容量などの問題もあり困難なため、実際には加速度に一定の閾値を設け、車両に何らかの衝撃があったと考えられるタイミングを検知し、その前後十数秒の映像とセンサ情報を記録することが多い。このように記録されたデータを、本論文ではイベントデータと呼ぶ。イベントデータの内、交通事故やそれに近い危険な状況（ヒヤリハット）が記録されたデータは、交通事故の際の証跡として活用されているだけでなく、実際に映像をドライバーが視聴して、自身の運転を振り返ったり [27]、危険予知訓練を行うなど [2]、安全運転教育に活用されている。また、法人運行車両などを対象に、走行中の運行車両から危険運転

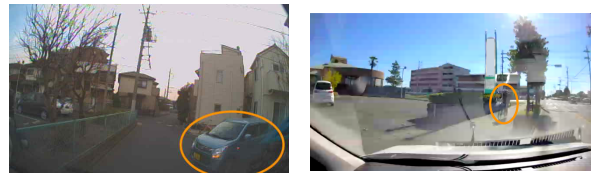


図 1 ドライブレコーダで記録された危険運転シーンの例。左図では、三叉路に侵入する車に対して、右図では、歩道を走る自転車に対して急ブレーキし、危険運転を起こしている。

シーンを検出し、頻繁に危険運転をするドライバーの事故を未然に防ぐことにも利用されている [31]。現在は個人の経験や知識を元に人手で作成している「ヒヤリハットマップ」^{*2}のようなサービス^{*3} ^{*4}も、イベントデータを大量に蓄積し危険運転の発生箇所を集計していくことによって、自動的に作成できると考えられる。実際の危険運転シーンとなった前方映像の画像例を図 1 に示す。

^{*2} 交通事故が起りやすい危険性の高い場所を記録した地図

^{*3} <https://www.city.saitama.jp/kita/001/001/003/p020992.html>

^{*4} <https://www.city.matsuyama.ehime.jp/kurashi/kurashi/seibi/keikaku/childmap/kodomomesen.html>

¹ 日本電信電話株式会社 NTT サービスエボリューション研究所

^{*1} http://www.honda.co.jp/safetyinfo/global/safetyinfo_2016_E.pdf

一方、ドライブレコーダで記録される全てのイベントデータが交通事故や危険運転に関するものではない。東京農工大学スマートモビリティ研究拠点の調査によると、例えば、道路の激しい起伏が原因で車両に大きな衝撃が加えられ記録されたものや、狭い道路での方向転換のために行う小刻みな加速が原因で記録されたものなど、危険シーンを含まないイベントデータも全体の70%近く含まれている [25]。また、実際の安全運転教育や運行管理の現場では、例えばドライバーの苦手な運転状況やヒヤリハットマップの詳細化のために、危険運転の発生対象（例えば、歩行者と衝突しそうになったなど）を付与することも望まれている。しかしながら、収集された膨大なイベントデータを人手で抽出し、更なる発生対象をラベリングすることは、十数秒の映像を膨大な件数閲覧して判定する作業が必要になるため、その判定者に多くの労力や注意力を要する。このため、これらの蓄積されたイベントデータの中から、自動的に危険運転シーンを含むデータを抽出しその発生対象をラベリングすることは、安全運転教育や運行管理にかかるコストを低減し、詳細な情報を提供できることから、それを促進する効果が期待される。そこで本論文では、イベントデータの集合に対して、危険運転が起こった際の発生対象のラベルを自動的に付与するタスクに取り組む。

このタスクに対して、著者らはこれまでも、前方映像とセンサ情報を活用した教師あり深層学習に基づく手法を提案している [28], [29]。ここでの提案手法は、ある程度統一された環境で記録されたイベントデータを想定したものであった。しかし、運行管理のような現場では、様々なドライブレコーダの機種や車種、走行地域でイベントデータが記録されることが想定される。このような状況では、例えばドライブレコーダの機種によっては前方映像の解像度や画角が異なったり、ドライブレコーダの設置位置や設置する車種によってセンサが異なる傾向を持って記録されるため、個々の異なりを吸収するための前処理が必要となる。個々の環境に合わせた前処理の方式を確立することは困難であり、上記の提案手法では多様な環境で記録されたイベントデータに対しては精度が低下すると考えられる。

本論文では、前方映像に対して物体検出技術を適用し、得られた物体情報に着目する。物体検出は、画像中に映る物体の名称を境界領域と共に検出する技術である [15]。物体情報はイベントデータの記録環境に依存しない情報として獲得できるため、特別な前処理を必要としない利点がある。さらに、本論文では映像中の物体検出結果を扱う既存技術として、Dynamic Spatial Attention (DSA) [1] に着目し、イベントデータ中の前方映像に適用するための拡張手法を提案する。DSAは深層学習に基づく手法で、前時刻の画像を考慮しながら、画像中の物体検出結果を適切な特徴量に変換する機構である。本論文での主な拡張点は、物体検出結果から得られる境界領域に基づく特徴量を算出

する点と、その特徴量に基づいて画像中の重要と思われる物体を優先的に抽出する点である。この拡張により、DSAでこれまで捉えていなかった、自車と前方映像中の物体との距離や角度などの位置関係を簡易に扱えるようになり、さらに重要物体の取りこぼしの可能性を低くでき、危険運転の発生対象推定タスクにおける精度向上が期待できる。

以下、本論文の構成を示す。2章では危険運転シーンの検出や分類の関連研究について述べる。3章では本論文で扱うイベントデータの構造やその前処理について述べ、DSAについて簡単に紹介する。4章では本論文で提案するDSAの拡張点と、手法の全体像について説明する。5章では実際のドライブレコーダで記録されたイベントデータを用いて、実験によって提案手法の有効性を評価する。6章では本論文のまとめと今後の課題について述べる。

2. 関連研究

本章では、ドライブレコーダからの危険運転シーンの検出を扱う関連研究と、ドライブレコーダを用いたその他の関連研究について述べ、本研究の位置付けを整理する。

ドライブレコーダの前方映像データを用いた危険運転シーンを対象とする研究として、そのシーンを検出・分類する研究 [6], [19] と、危険運転シーンが将来発生するかを予測する研究 [1], [20] に別れる。いずれも画像解析はConvolutional Neural Network (CNN) [3]、時系列モデリングはRecurrent Neural Network (RNN) [13] を用いている。特に Chan らの研究 [1] では、車載カメラで記録される前方映像中の移動体間の交通事故を予測するため、画像中の物体検出結果を入力に用い、DSAによる特徴変換手法を提案している。Suzuki らの研究 [20] でも、DSAを用いて物体検出結果を特徴変換し、より正確な予測結果を得るための目的関数の提案をしている。本論文で取り組むタスクにおけるDSAの有効性については、提案手法と共に5章の実験によって評価する。ドライブレコーダのセンサデータを対象とした危険運転シーンを扱う研究として、速度と加速度に経験的なルールを設定し、閾値調整によって危険運転を含むイベントデータを検出する研究 [26] や、センサデータを適切な特徴量に変換して右折時のヒヤリハットの発生形態を分析し、右折時のヒヤリハット検出に取り組む研究 [30] がある。本研究では自車の危険運転の発生対象の識別タスクに取り組み、前方映像のみを入力に扱っていることから、これらの研究とは異なる。

危険運転の検出以外にも、ドライブレコーダは様々なタスクで用いられている。センサデータに含まれるドライバの運転操作を用いたものでは、危険運転の起こりやすい潜在リスクを持つ道路の推定をする研究 [11], [32] や、頻繁に危険運転を起こすドライバーの分類に関する研究 [23], [24] がある。車両運行中の前方映像データは、自動車の自動運転に関する研究で幅広く利用されている。自動運転タス

クでは、前方映像に基いてハンドルの操舵角を予測する研究 [9], 車の直進, 左折, 右折, 車線変更, 停止などの運転操作を予測する研究 [21], 同様のタスクに車内映像も用いた研究 [5] などがある. これらはいずれも深層学習に基づく手法を提案しており, 画像解析は CNN, 時系列モデリングは RNN において有効性の知られている, Long Short-Term Memory (LSTM) [4] を用いている. 本論文でも CNN や LSTM を用いたモデル化に取り組むものの, タスクが異なる.

3. 準備

3.1 物体検出の適用

本論文では前方映像のフレーム画像系列 I_1, I_2, \dots, I_T に対し物体検出処理を行い, その結果を活用する. ここで, T はイベントデータの総フレーム数である. 時刻 t の画像 I_t から得られる物体検出結果は, N_t 個の検出されたオブジェクトを持ち, 各オブジェクトの ID を総物体種類数 V 次元の 1-of-K 表現で持つベクトル $\hat{\mathbf{x}}_{t,n}^O$, 各オブジェクトの検出された領域情報を持つベクトル $\mathbf{l}_{t,n}$, 及び各オブジェクトの検出信頼度 $p_{t,n} \in (0, 1]$ からなる ($n = 1, 2, \dots, N_t$). ここで, 領域情報ベクトル $\mathbf{l}_{t,n}$ は, その左端, 上端, 右端, 下端の 4 つの画像中の座標を持つ ($\mathbf{l}_{t,n} = \{x_{t,n}^{\text{lef}}, y_{t,n}^{\text{top}}, x_{t,n}^{\text{rig}}, y_{t,n}^{\text{bot}}\}$).

3.2 Dynamic Spatial Attention

Dynamic Spatial Attention (DSA) [1] は, ドライブレコーダなどで記録される映像中の交通事故を予測するため, 画像中の複数の物体に関する領域情報を DNN によって特徴変換する手法である. 各画像の物体情報を, 時系列情報も考慮しながら特徴変換することができ, 交通事故の予測タスクで高い精度を達成している. 本論文で取り組むイベントデータに対する危険運転状況の推定においても, 前方映像中の対象物体が重要な特徴であることから, その効果を期待できる. 本節では, DSA の特徴変換の方法について簡単に説明する.

DSA は, まず, 時刻 t の画像 I_t から検出された各物体 j に対し, 物体特徴ベクトル $\hat{\mathbf{x}}_{t,j}$ を算出する. $\hat{\mathbf{x}}_{t,j}$ は, 対象物体の領域のみ抽出した画像を, ImageNet データセット [16] で Pre-training した VGG[17] によってエンコードすることで得られる. 時刻 t の画像から検出された物体集合 $\{\hat{\mathbf{x}}_{t,j}\}_{j=1}^J$ に関する特徴ベクトル \mathbf{x}_t^L は, Self-Attention 機構 [22] によって計算される.

$$\mathbf{x}_t^L = \sum_{j=1}^J \alpha_{t,j} \hat{\mathbf{x}}_{t,j}, \quad \alpha_{t,j} = \frac{\exp(e_{t,j})}{\sum_{j'=1}^J \exp(e_{t,j'})}. \quad (1)$$

ここで, J は入力物体数をコントロールする閾値で, $e_{t,j}$ は Softmax 関数で正規化される前の Attention 重みであり, 次の式で計算される.

$$e_{t,j} = \mathbf{w}^T \tanh(\mathbf{W}_e \mathbf{h}_{t-1} + \mathbf{U}_e \hat{\mathbf{x}}_{t,j} + \mathbf{b}_e). \quad (2)$$

ここで, $\mathbf{w}, \mathbf{W}_e, \mathbf{U}_e, \mathbf{b}_e$ はニューラルネットワークのモデルパラメータ, \mathbf{h}_{t-1} は, 1 時刻前の入力画像 I_{t-1} に対して DSA がエンコードした特徴ベクトルであり, この後の式 (3) によって得られる. これにより, 前時刻の画像中の物体集合と現在時刻の特定の物体との関係性を捉えた重み $e_{t,j}$ を算出する.

これとは別に, 画像全体の大まかな傾向を考慮した特徴ベクトル \mathbf{x}_t^F を, 画像 I_t そのものを上記と同様の学習済みの VGG によるエンコードで得る. この画像全体の特徴ベクトル \mathbf{x}_t^F と, 画像中の物体集合に関して得られた特徴ベクトル \mathbf{x}_t^L を共に考慮するため, 2 つの特徴ベクトルの結合 $\mathbf{x}_t = [\mathbf{x}_t^F; \mathbf{x}_t^L]$ を行う. この \mathbf{x}_t を LSTM に入力し, その出力として特徴ベクトル \mathbf{h}_t を得る.

$$(\mathbf{h}_t, \mathbf{c}_t) = \text{LSTM}(\mathbf{x}_t, \mathbf{h}_{t-1}, \mathbf{c}_{t-1}). \quad (3)$$

ここで, \mathbf{c}_t は LSTM が系列状態を考慮するためのメモリセルであるが, 本論文では説明を省略する. この処理を系列データの最終時刻 T まで行い得られる系列 $\{\mathbf{h}_t\}_{t=1}^T$ が, DSA で得られる特徴ベクトルである.

4. 提案手法

本章では, 前方映像中の物体検出結果に基づく, イベントデータに対する危険運転の発生対象の推定手法について述べる. 本論文で提案する手法は, 3.2 節で述べた DSA による物体検出結果の特徴変換を採用している. DSA を本論文で取り組むタスクに適用するにあたって, 考えられる課題について 4.1 節で整理し, それらの課題に対する本論文でのアプローチを 4.2 節以降で述べる.

4.1 既存手法適用の課題

イベントデータに対する危険運転の発生対象の推定に DSA を適用するにあたり, 以下 2 つの課題が考えられる.

課題 1: 危険な運転状況とは, 主に自車と対象物体との位置関係によって判定できるが, 既存の DSA は自車と前方映像中の物体との位置関係を考慮していないこと.

課題 2: 入力物体数 J が, 時刻 t の画像 I_t から検出された物体数 N_t を超えると, 推定に重要な危険対象の物体を取りこぼしてしまう可能性があること.

課題 1 に関しては, 自車に対して遠い距離にある物体よりも, 近い距離にある物体のほうが, 危険運転の対象として考慮すべきと考えられる. しかし, 2 章でも述べたように, DSA は前方映像中の自車を含まない移動体間の交通事故を予測するため, 画像特徴 $\hat{\mathbf{x}}_{t,j}$ のみを入力として想定しており, 自車との位置関係を明確に考慮していない.

課題 2 の素朴な解決策として, J をデータセット中の N_t と比較しながら十分に大きな値とすることで, 重要な物体を取りこぼす可能性を低くすることが考えられる. ただし, DSA は入力物体に関して別途訓練済みの CNN へ入力

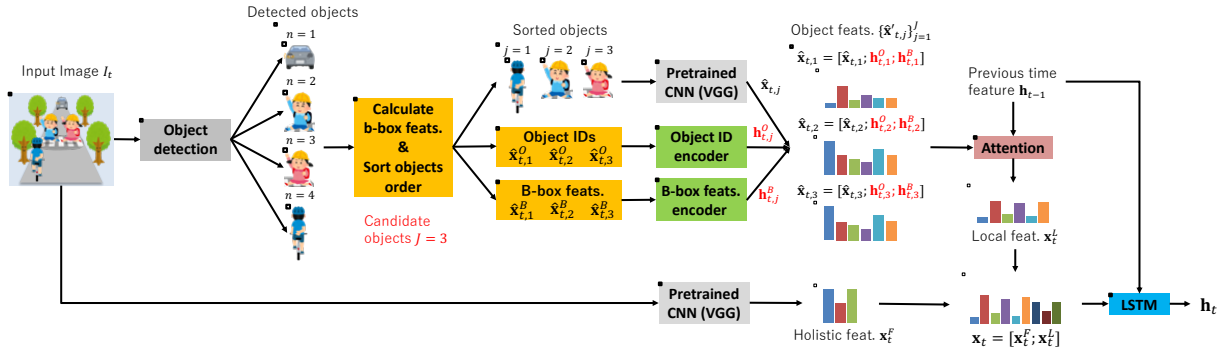


図 2 DSA を拡張した提案手法の概要図

して画像特徴 $\hat{x}_{t,j}$ を得る必要がある。また $\hat{x}_{t,j}$ の次元数は数千程度が想定されるため、 J の増加に伴い多くの計算処理とメモリを要する。以上の理由から、データセット中の物体数に関わらず、計算量の観点から J は可能な限り小さい値とすることが望ましい。

4.2 本論文のアプローチ

これらの課題を解決するために、本論文では、前方映像上の自転車と物体との位置関係を考慮可能な特徴量を算出し、さらにその特徴量に基づいて、画像中の重要な（危険な）物体を優先的に J 個選出する手法を DSA の拡張として提案する（図 2）。具体的なアプローチの説明をする前に、これらの拡張に関する重要な知見として、Lee ら [8] のシミュレーションデータを用いた自動車事故映像の分析結果を簡単に紹介する。ここでは、カーシミュレーターを用いて自動車走行中の前方映像のデータセットを収集・分析し、数秒後に自転車と交通事故を起こしやすい車は、以下 3 つの傾向を持ちやすいことを報告している。

- 車輪の角度:** その車の車輪が走行中の車線と平行でない。
- 車体の向き:** その車の車体方向が自転車と同じ向きでない。
- 距離:** その車が自転車と近い距離を走行している。

「車輪の角度」と「車体の向き」は、自転車が走行中の道路上で右左折をしようとしている車や、その道路へ侵入・横切るなどの動きをしている車などが該当する。

このような傾向は車だけでなく、例えば自転車と距離が近い歩行者や、自転車が走行中の道路を横切る自転車は危険性が高いなど、別の物体に関しても該当すると考えられる。これらの要素を含む特徴を算出できれば、自転車と物体の位置関係について考慮できるだけでなく、前方映像中の重要な（危険な）物体の特定ができ、同時に自転車の危険な運転状況の判定にも活用できると考えられる。一方、上記の論文では、物体の向きや距離といった詳細な情報をシミュレータから参照して得ているが、本論文で扱うイベントデータでは、前方映像からこれらの要素を推定する必要がある。ドライブレコーダのような、自転車の走行によって撮影視点も動いている状況で、前方映像中の対象物との実距離や角

度を厳密に算出するためには、カメラの焦点距離、設置位置の高さ、画角などのパラメータが必要になる [7]。これらのパラメータはドライブレコーダの機種や車種、ドライブレコーダの設置位置などによって大きく異なり、様々な環境で収集される大量のイベントデータに対してこれらを設定するのは困難である。

このような理由から、本論文では物体検出によって得られる画像中の物体の境界領域に基づく、境界領域特徴量を算出する（課題 1 に対応）。これは、「車体の向き」や「自転車との距離・角度」を境界領域から計算可能な特徴量である。更に、境界領域特徴ベクトルから重要物体を選出し、優先的に J 個選出して入力物体とする（課題 2 に対応）。それぞれの手法について、4.3 節と 4.4 節で説明した後、DSA を拡張した提案手法の全体像について 4.5 節で述べる。

4.3 境界領域特徴量の算出

本節では、物体検出結果として得られている領域情報ベクトル $\mathbf{l}_{t,n}$ から、境界領域特徴量を算出する手法について説明する。計算方法の簡単な説明を図 3 に示す。まず、特徴量の算出にあたって、画像の縦幅 H と横幅 W を用意する。画像中の座標 (x, y) について、原点を画像の左上端 $(0, 0)$ 、画像の右下端を (W, H) とする。本論文では、自転車の撮影位置を下端中央 $(\frac{W}{2}, H)$ に設定する。準備として、 n 番目の物体の画像中の座標 $(x_{t,n}^{\text{cen}}, y_{t,n}^{\text{cen}})$ と、境界領域の横幅 $w_{t,n}$ と横幅 $h_{t,n}$ を次のように計算する。

$$x_{t,n}^{\text{cen}} = \frac{x_{t,n}^{\text{lef}} + x_{t,n}^{\text{rig}}}{2}, \quad y_{t,n}^{\text{cen}} = \frac{y_{t,n}^{\text{top}} + y_{t,n}^{\text{bot}}}{2}. \quad (4)$$

$$w_{t,n} = x_{t,n}^{\text{rig}} - x_{t,n}^{\text{lef}}, \quad h_{t,n} = y_{t,n}^{\text{bot}} - y_{t,n}^{\text{top}}. \quad (5)$$

このとき、自転車と n 番目の物体の画像上の距離 $d_{t,n}$ と角度 $r_{t,n}$ を次のように計算する。

$$\Delta x_{t,n} = x_{t,n}^{\text{cen}} - \frac{W}{2}, \quad \Delta y_{t,n} = H - y_{t,n}^{\text{cen}}, \quad (6)$$

$$d_{t,n} = \sqrt{\Delta x_{t,n}^2 + \Delta y_{t,n}^2}, \quad r_{t,n} = \arctan\left(\frac{\Delta y_{t,n}}{\Delta x_{t,n}}\right). \quad (7)$$

また、物体の形状や大きさに関する情報を得るため、物体

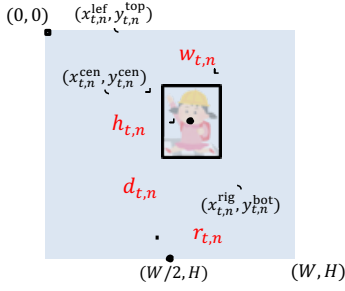


図 3 境界領域特徴量の算出方法

の境界領域のアスペクト比 $a_{t,n}$ と、画像全体に対する面積比 $s_{t,n}$ を計算する.

$$a_{t,n} = \frac{w_{t,n}}{h_{t,n}}, \quad s_{t,n} = \frac{w_{t,n} \times h_{t,n}}{W \times H}. \quad (8)$$

これにより、例えば前方映像中の対向車が右折しているとき、その車は横向きなため横長の境界領域で物体検出され、アスペクト比 $a_{t,j}$ も正面を向く車に比べて大きな値で算出される。また、自車から近い距離にいる物体は $d_{t,n}$ も小さな値となるが、物体も大きな境界領域で物体検出されるため $s_{t,n}$ は大きな値で算出される。すなわち、4.2 節で述べた事故を起こしやすい車の傾向について、「車体の向き」に該当するものは $a_{t,n}$ 、「距離」に該当するものは $d_{t,n}, a_{t,n}, s_{t,n}$ である。このようにして得られる $\hat{\mathbf{x}}_{t,n}^B = \{d_{t,n}, a_{t,n}, r_{t,n}, s_{t,n}\}$ を境界領域特徴ベクトルとする。

4.4 境界領域特徴量に基づく重要物体の選出

得られた境界領域特徴を用いて、画像から J 個の重要物体を選出する。ここでは一般化のため、4 つの境界領域特徴量の重要度合いを考慮するためのパラメータを導入する。それぞれの境界領域特徴量とそれを考慮するパラメータを、 $\{(d_{t,n}, \beta_d), (r_{t,n}, \beta_r), (a_{t,n}, \beta_a), (s_{t,n}, \beta_s)\}$ と対応付け、次のような線形和によって各物体の重要度 $f_{t,n}$ を計算する。

$$f_{t,n} = \beta_d d_{t,n} + \beta_r r_{t,n} + \beta_a a_{t,n} + \beta_s s_{t,n}. \quad (9)$$

時刻 t の画像から得られている N_t 個の物体に関する重要度 $\{f_{t,n}\}_{n=1}^{N_t}$ について、 N_t 個の中で降順で並び替えた後、上位から J 個の物体を選出し、境界領域特徴ベクトル $\{\hat{\mathbf{x}}_{t,j}^B\}_{j=1}^J$ 、ID 情報 $\{\hat{\mathbf{x}}_{t,j}^O\}_{j=1}^J$ を得た後、該当する物体画像を訓練済みの CNN に入力して $\{\hat{\mathbf{x}}_{t,j}\}_{j=1}^J$ を得る。

本論文の実験では、危険運転の発生対象の推定にあたって、自車からより近い距離にある物体が重要であるという仮説のもと、パラメータを $\{\beta_d, \beta_r, \beta_a, \beta_s\} = \{-1, 0, 0, 0\}$ として物体の選出をした。すなわち、 $d_{t,n}$ が小さい物体ほど降順に並び替えた際に上位になるように設定した。

4.5 DSA の拡張点

本節では、4.3 節と 4.4 節で得られた境界領域特徴量と、それに基づく重要物体の選出を DSA に導入した際のモデ

ルの拡張点について述べる。

入力特徴量. 訓練済みの CNN によって得られていた $\hat{\mathbf{x}}_{t,j}$ だけでなく、境界領域特徴ベクトル $\hat{\mathbf{x}}_{t,j}^B$ と物体の ID 情報 $\hat{\mathbf{x}}_{t,j}^O$ を加えた入力に変更する。ここでは、1 層の全結合ニューラルネットワークを 2 つ導入し、それぞれの特徴ベクトルを線形変換することで、 $\mathbf{h}_{t,j}^B$ と $\mathbf{h}_{t,j}^O$ を得る。この 2 つを $\hat{\mathbf{x}}_{t,j}$ と結合した、 $\hat{\mathbf{x}}_{t,j}^I = [\hat{\mathbf{x}}_{t,j}; \mathbf{h}_{t,j}^B; \mathbf{h}_{t,j}^O]$ を用意し、式 (1) と (2) の $\hat{\mathbf{x}}_{t,j}$ と置き換える。

出力層. DSA は物体検出結果と画像全体の画像特徴をエンコードし、LSTM によって時間方向の状態変化を考慮した特徴ベクトルの集合 $\{\mathbf{h}_t\}_{t=1}^T$ を得るまでの手法である。本論文で取り組む危険運転の発生対象の推定は、発生対象の数をクラス数とする分類問題と捉えられる。この分類ラベルを効果的に得るため、まず、各時刻 t の重要度を考慮し得られる特徴ベクトル \mathbf{h}_t を次の式で示す Self-Attention によって得る。

$$\mathbf{h}_\tau = \sum_{t=1}^T \alpha_t \mathbf{h}_t, \quad (10)$$

$$\alpha_t = \text{softmax}(\mathbf{u}_t^T \mathbf{u}_\tau), \quad (11)$$

$$\mathbf{u}_t = \tanh(\mathbf{W}_\tau \mathbf{h}_t + \mathbf{b}_\tau). \quad (12)$$

$\mathbf{u}_\tau, \mathbf{W}_\tau, \mathbf{b}_\tau$ は、ニューラルネットワークのモデルパラメータである。この後、 \mathbf{h}_τ を推定対象のクラス数に合わせた特徴ベクトルに変換するため、1 層の全結合ニューラルネットワークを導入し、更に Softmax 関数によって正規化して、出力ベクトル \mathbf{y} を得る。このベクトルの中で最大値を持つ要素に対応するラベルを出力する。

5. 評価実験

本章では、実際のイベントデータを用いて定量的に提案手法の有効性を評価する。その後、定性評価によって提案手法の実際の推定結果を確認する。

5.1 データセットとパラメータ設定

実験で用いるデータセットには、日本カーソリューションズ株式会社*5から提供して頂いたイベントデータを用いた。各イベントデータは加速度トリガが反応した時刻を中心に、前後十数秒の前方映像とセンサ系列からなり、映像とセンサ共に 30fps で記録されている。また、各イベントデータは訓練された人間による危険運転（ヒヤリハット）の有無と、危険運転の発生対象のラベルが付与されている。本論文では、{非ヒヤリハット, 車, バイク, 歩行者} の 4 つのラベルを対象に実験を行った。以下、ラベル名を { (非ヒヤリハット, No near miss), (車, Car), (バイク, Bicycle), (歩行者, Pedestrian) } とする。各ラベルの訓練データと評価データの件数を表 1 に示す。実験では、用意したデータの 70% の 3,429 件を訓練データとし、30% の 1,469 件を

*5 <https://www.ncsol.co.jp>

表 1 使用したデータセットの内訳

Label	Train	Test	Total
No near miss	2,191	924	3,115
Car	699	337	1,036
Bicycle	315	127	442
Pedestrian	224	81	305
Total	3,429	1,469	4,898

評価データとした。

前処理として、各イベントデータから危険運転シーンに関連するフレームを抽出する。各イベントデータ d の総フレーム数を T_d としたとき、その中心フレーム番号 $\frac{T_d}{2}$ を起点に T フレームを抽出する。すなわち、各イベントデータについて、区間 $(\frac{T_d}{2} - \frac{T}{2}, \frac{T_d}{2} + \frac{T}{2})$ を分析対象のフレームとし、本論文の実験では $T = 100$ として抽出した。各フレームの画像データは、RGB 形式で 640×400 ($W = 640, H = 400$) の解像度で記録されている。物体検出は YOLO[15] を用いて元画像に対して直接行い、検出された順番で物体を記録した。先行研究 [1] に倣い、検出された物体の内、交通に関する物体として、{car, person, bicycle, motorbike, bus, truck, train} を対象に抽出した。すなわち、物体の種類数 $V = 7$ である。画像特徴抽出は、画像を 224×224 に線形変換した後に、ImageNet で Pre-training 済みの VGG19 に入力し、出力層の手前の層 (fc8 と呼ばれる) が変換した特徴ベクトル (4,096 次元) を抽出して \mathbf{x}_t^f を用意した。物体の画像特徴 $\hat{\mathbf{x}}_{t,j}$ も、YOLO で得られた物体の境界領域を切り抜いて、同様の処理によって用意した。

深層学習部分のパラメータについては、各 Fully-Connected Layer の出力ユニット数を 256 とし、出力された特徴ベクトルを ReLu 関数 [14] で非線形変換処理をし、更に過学習を抑制する技術として知られる Dropout[18] を $p = 0.5$ に設定して入れた。ネットワークの学習では、誤差逆伝搬法で求められた目的関数の勾配に基づいて、Adam[10] によって最適化した*6。出力層の誤差は交差エントロピーによって計算した。

5.2 評価尺度

危険運転の発生対象分類タスクにおいて、提案手法の有効性を議論するため、推定結果がどれくらい正解しているかという正確性と、推定結果が全ての正解のうちどれだけ網羅しているかという網羅性の、2つの観点から評価する。本論文では、正確性を適合率 (Precision)、網羅性を再現率 (Recall)、またこれらの評価指標の調和平均である F 値 (F1-score) によって提案手法の分類性能を評価する [12]。これらの評価指標を計算するため、評価データの集合と推定結果の集合に関して、次のような 3 種類のパラメータを定義する。

*6 最適化のハイパーパラメータは、 $\alpha = 0.001, \beta_1 = 0.9, \beta_2 = 0.999, \epsilon = 10^{-8}$ とした

TP : 正例に対して推定結果が正であったデータの件数

FP : 負例に対して推定結果が正であったデータの件数

FN : 正例に対して推定結果が負であったデータの件数

このとき、Precision, Recall, F1-score は次のように計算される。

$$\text{Precision} = \frac{TP}{TP + FP}, \text{Recall} = \frac{TP}{TP + FN}, \quad (13)$$

$$\text{F1-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (14)$$

なお、今回用意したデータセットでは、ラベルごとのデータ数の偏りが大きく、全体の TP, FP, FN から得られるマイクロ平均だけではデータが少ないラベル (Bicycle, Pedestrian) の正解度合いが評価されにくいことから、ラベル別に TP, FP, FN を用意し上記の評価指標を算出した後に平均するマクロ平均も評価した。これにより、データ数が少ないラベルでも一度、Precision, Recall, F1-score が計算されるため、各ラベルの評価値が全体の評価値に影響を与えるよう計算される。

5.3 実験結果

比較手法として DSA[1] を用意し、物体の並び替えの効果と境界領域特徴量を入力に追加する効果をそれぞれ評価するため、境界領域特徴量を算出し物体を並び替えた後、入力に境界領域特徴量を用いない手法 (DSA+Sorted) と、境界領域特徴量を算出し物体を並び替えず、入力に境界領域特徴量を用いる手法 (DSA+B-feats) を用意した。Proposed は本論文の提案手法である。物体を並び替えない場合は、画像に対して YOLO が出力した順に上位 J 件を抽出した。

表 2 に、提案手法と比較手法の実験結果を示す。実験では入力物体数 J に関して、 $J = 5$ と $J = 10$ の条件で評価した。それぞれの評価値 (列) について最大値であったものを太字で示している。Proposed がマイクロ平均、マクロ平均共にいずれの評価値でも最大値を示した。特に、 $J = 5$ におけるマクロ平均の F 値に着目すると、それぞれの拡張手法で、DSA に比べて精度が約 7% 程度精度向上していることから、本論文で提案した 2 つの拡張点それぞれに有効性が確認できる。また、Proposed はそこから更に約 9% 精度向上していることから、それぞれの拡張点を組み合わせることの有効性も確認できる。 $J = 5$ と $J = 10$ の間で比較すると、それぞれの手法の評価値について $J = 5$ に比べて $J = 10$ の方が上回っているものが多い。これは、入力物体数を増やすことで、推定に必要な重要物体を入力に含められる可能性が高くなったためと考えられる。

$J = 5$ の条件での DSA と Proposed の分類結果を Confusion Matrix として図 4 に示す。図では、各ラベルを簡単のため表 1 の先頭文字で表記しており、横方向に正解ラベル、縦方向に推定ラベルを示している。各セルの中の数

表 2 各手法の推定性能. 太字は各評価 (列) での最高値を示す.

Method	$J = 5$						$J = 10$					
	Micro Avg.			Macro Avg.			Micro Avg.			Macro Avg.		
	Pre.	Rec.	F1	Pre.	Rec.	F1	Pre.	Rec.	F1	Pre.	Rec.	F1
DSA [1]	82.03	82.03	82.03	63.39	58.36	59.63	84.34	84.34	84.34	69.59	64.22	65.73
DSA+Sorted	84.07	84.07	84.07	69.71	65.42	66.98	85.23	85.23	85.23	74.47	68.44	70.63
DSA+B-feats	85.23	85.23	85.23	70.03	63.41	65.15	85.57	85.57	85.57	70.08	67.39	67.84
Proposed	86.86	86.86	86.86	77.35	72.75	74.69	86.18	86.18	86.18	78.11	72.87	75.20

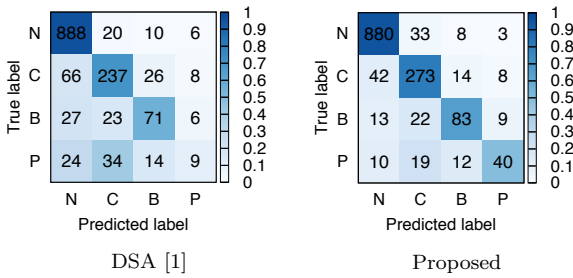


図 4 $J = 5$ の条件での, 既存手法 [1] と提案手法の混合行列. “N”, “C”, “B”, “P” はそれぞれ, No near miss, Car, Bicycle, Pedestrian を示す.

字は, そのラベルを推定した件数を示しており, 左上から右下にかけての対角の数字が高いほど, 正しく推定できていることを表している. また, 各セルの色の濃淡は各ラベルの正解データ数に対して推定件数が多いほど濃くなっている. DSA に比べて, Proposed は No near miss 以外のラベルで正解の件数が大きく増えている. このように, DSA では推定できていなかった少ないラベルの危険運転対象 (Bicycle, Pedestrian) などを, Proposed では正しく推定できるようになったと考えられる. このことは, 表 2 でのマイクロ平均に比べてマクロ平均の数値の向上具合が大きいためからも確認できる.

最後に, ある危険運転事例の画像に対する物体検出結果と, DSA 及び Proposed が算出した Attention スコア $\alpha_{t,j}$ の可視化結果を図 5 に示す. 画像に対する物体検出結果が左, DSA と Proposed の Attention スコア $\alpha_{t,j}$ の可視化結果が, それぞれ中央と右の画像に対応する. 各物体の境界領域の赤の濃淡が濃いほど, Attention スコアが大きいことを表す. このデータに対する危険運転の発生対象は Bicycle で, 自車が左折しようとした際に, 画像中の左側を走行する自転車と衝突しそうな例である. このデータに対して DSA は誤って Car, Proposed は正しく Bicycle と推定した. この画像に対して物体検出が誤って推定した物体も他にあるが, 可視化のため誤検出した物体は除いて表示している. Attention スコアを見ると, DSA は右側にある車 (Car) に対して高いスコアを割り当て, Proposed は手前の自転車 (Bicycle) に対して高いスコアを割り当てている. Proposed は物体の並び替えだけでなく, 境界領域特徴として物体との距離や角度, アスペクト比などの特徴量を入力に追加しており, これらの特徴量が有効に活用

され, この例に対して Proposed は正しい推定をしたと考えられる.

6. おわりに

本論文では, ドライブレコーダで記録されたイベントデータに対して, 危険運転の発生対象を推定することを目的に, 映像に対して物体検出処理をして得られる物体情報を入力に用いる手法を提案した. DSA という深層学習を元にした既存手法によって, 物体情報を特徴ベクトルに変換する際に, 提案手法は 2 つの拡張点から構成される. 1 つ目は物体検出で得られる物体の境界領域を元にした境界領域特徴量の算出, 2 つ目はその特徴量によって重要物体を優先的に選出し, 危険な物体の取りこぼしの可能性を低くする. 実際のドライブレコーダデータを用いた評価実験の結果, 2 つの拡張点を用いた提案手法が, 既存手法に比べて高い推定性能を示すことを確認した.

今後は, 物体特徴量の算出に関して各物体の時系列上の軌跡などを追跡した特徴量に拡張し, より精密に物体の動きを捉えた推定を可能にすることを検討している.

謝辞

本研究は, 日本カーソリューションズ株式会社から提供して頂いたデータを利用した. ここに記して謝意を示す.

参考文献

- [1] Chan, F.-H., Chen, Y.-T., Xiang, Y. and Sun, M.: Anticipating Accidents in Dashcam Videos, *ACCV*, pp. 136–153 (2016).
- [2] Driving Safety Promotion Center Honda Motor Co. Ltd.: Honda Driving Safety Promotion Activities 2016, , available from (http://www.honda.co.jp/safetyinfo/global/safetyinfo_2016_E.pdf) (accessed 2017-10-30).
- [3] Fukushima, K.: Neocognitron : A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position, *Biological Cybernetics*, Vol. 36, pp. 193–202 (1980).
- [4] Hochreiter, S. and Schmidhuber, J.: Long Short-term Memory, *Neural computation*, Vol. 9, pp. 1735–1780 (1997).
- [5] Jain, A., Koppula, H. S., Raghavan, B., Soh, S. and Saxena, A.: Recurrent Neural Networks for Driver Activity Anticipation via Sensory-Fusion Architecture, *ICRA*, pp. 3118–3125 (2016).
- [6] Kataoka, H., Suzuki, T., Oikawa, S., Matsui, Y. and



物体検出結果

DSA が計算した Attention の可視化

Proposed が計算した Attention の可視化

図 5 画像に対する物体検出結果 (左) と DSA (中央), Proposed (右) が計算した, 物体に対する Attention スコア $\alpha_{t,j}$ の可視化例. 各物体の境界領域の赤の濃度が濃いほど, Attention スコアが大きいことを表す. 危険運転の発生対象は Bicycle で, DSA は誤って Car, Proposed は正しく Bicycle と推定した.

- Satoh, Y.: Drive Video Analysis for the Detection of Traffic Near-Miss Incidents, *ICRA*, pp. 1–8 (2018).
- [7] Ke, R., Lutin, J., Spears, J. and Wang, Y.: A Cost-effective Framework for Automated Vehicle-pedestrian Near-miss Detection through Onboard Monocular Vision, *In CVPR workshop*, pp. 25–32 (2017).
- [8] Kim, H., Lee, K., Hwang, G. and Suh, C.: Crash To Not Crash: Learn to Identify Dangerous Vehicles using a Simulator, *AAAI* (2019).
- [9] Kim, J. and Canny, J. F.: Interpretable Learning for Self-Driving Cars by Visualizing Causal Attention, *ICCV*, pp. 2942–2950 (2017).
- [10] Kingma, D. P. and Ba, J.: Adam: A Method for Stochastic Optimization, *ICLR* (2015).
- [11] Lam, H. T.: A Concise Summary of Spatial Anomalies and Its Application in Efficient Real-time Driving Behaviour Monitoring, *In SIGSPATIAL*, pp. 30:1–30:9 (2016).
- [12] Manning, C. D., Raghavan, P. and Schütze, H.: *Introduction to Information Retrieval*, Cambridge University Press (2008).
- [13] Mikolov, T., Karafát, M., Burget, L., Cernocký, J. and Khudanpur, S.: Recurrent neural network based language model, *In INTERSPEECH 2010*, Vol. 2, pp. 1045–1048 (2010).
- [14] Nair, V. and Hinton, G. E.: Rectified Linear Units Improve Restricted Boltzmann Machines, *ICML*, pp. 807–814 (2010).
- [15] Redmon, J. and Farhadi, A.: YOLO9000: Better, Faster, Stronger, *CVPR*, pp. 6517–2525 (2017).
- [16] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C. and Fei-Fei, L.: ImageNet Large Scale Visual Recognition Challenge, *IJCV*, Vol. 115, No. 3, pp. 211–252 (2015).
- [17] Simonyan, K. and Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition, *CoRR*, Vol. abs/1409.1556 (online), available from <http://arxiv.org/abs/1409.1556> (2014).
- [18] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R.: Dropout: A Simple Way to Prevent Neural Networks from Overfitting, *Journal of Machine Learning Research*, Vol. 15, pp. 1929–1958 (2014).
- [19] Suzuki, T., Aoki, Y. and Kataoka, H.: Pedestrian Near-Miss Analysis on Vehicle-Mounted Driving Recorders, *MVA*, pp. 416–419 (2017).
- [20] Suzuki, T., Kataoka, H., Aoki, Y. and Satoh, Y.: Anticipating Traffic Accidents With Adaptive Loss and Large Scale Incident DB, *CVPR*, pp. 3521–3529 (2018).
- [21] Xu, H., Gao, Y., Yu, F. and Darrell, T.: End-to-end Learning of Driving Models from Large-scale Video Datasets, *In CVPR*, pp. 2174–2182 (2017).
- [22] Yang, Z., Yang, D., Dyer, C., He, X., Smola, A. J. and Hovy, E. H.: Hierarchical Attention Networks for Document Classification, *HLT-NAACL* (2016).
- [23] Yokoyama, D. and Toyoda, M.: Do Drivers’ Behaviors Reflect Their Past Driving Histories? - Large Scale Examination of Vehicle Recorder Data, *In the BigData Congress 2016*, pp. 361–368 (2016).
- [24] Yokoyama, D., Toyoda, M. and Kitsuregawa, M.: Understanding Drivers’ Safety by Fusing Large Scale Vehicle Recorder Dataset and Heterogeneous Circumstantial Data, *In PAKDD*, pp. 734–746 (2017).
- [25] スマートモビリティ研究拠点: ドライブレコーダデータセンター, 東京農工大学大学院工学研究院 (オンライン), 入手先 (<http://web.tuat.ac.jp/~smrc/drcenter.html>) (参照 2017-6-12).
- [26] 菊池理人, 日景由華, 御室哲志: ドライブレコーダデータからの自動分別の試み, 計測自動制御学会東北支部第 290 回研究集会 (2014).
- [27] 警察庁交通局: ドライブレコーダの活用について, 警察庁 (オンライン), 入手先 (https://www.npa.go.jp/bureau/traffic/anzen/drive_recorder.html) (参照 2017-6-12).
- [28] 山本修平, 遠藤結城, 戸田浩之: 映像とセンサ信号を用いたドライブレコーダデータからのヒヤリハット検出手法, *TOD (テクニカルノート)*, Vol. 10, No. 4, pp. 26–30 (2017).
- [29] 山本修平, 倉島健, 戸田浩之: ドライブレコーダデータに対するヒヤリハット発生対象分類, *DICOMO2018*, pp. 542–553 (2018).
- [30] 森村哲郎, 谷澤悠輔, 山崎慎也, 井手剛: 統計的機械学習を用いたプローブカーデータからのヒヤリハット発生形態の推定, 自動車技術会秋季学術講演会 (2011).
- [31] 日本カーソリソリューションズ株式会社: NTT ドコモの通信機能と GPS 機能を備えたドライブレコーダーを車両に搭載したテレマティクスサービスを展開, *NCS (オンライン)*, 入手先 (<https://www.ncsol.co.jp/corporation/service/dr.html>) (参照 2019-5-9).
- [32] 豊田正史, 横山大作, 伊藤正彦: 運転状況を考慮したドライブレコーダデータからの潜在リスク交差点検出手法, *DEIM Forum* (2017).