

# 交通事故リスクを低減する車載カメラ画像認識技術

A scene recognition method using dashcams for reducing traffic accident risks

井上 俊明

Toshiaki Inoue

**要旨** 視線が広範囲に分散しやすい複雑な風景を検出して運転支援などに利用することで、視認負荷による交通事故リスクの低減が期待できる。そこで本研究では、人間の視覚的注意を定量化する顕著性マップを特徴量として用い、視線分布の大きさに応じて高精度に画像を分類する手法を提案する。顕著性マップの推定には、解像度の劣化を抑えたエンコーダ・デコーダ構造のCNNを、3つの損失関数で最適化して用いる。さらに顕著性マップに加え、エンコーダ出力も特徴量として用いてクラス分類構造のCNNを転移学習することで、効果的に分類精度を向上できる。SALICONデータセットを用いて評価環境の構築と分類実験を行い、提案手法の有効性を確認した。

**Summary** Reducing accident risks due to driver's visual workloads is widely expected in recent traffic environment. In this paper, we describe a novel method to classify complex traffic scenes tending to increase the workloads from scattering visual saliency. Classification accuracy is improved by key methods based on a visual saliency predictor and a scene classifier. The predictor is improved by training an encoder-decoder CNN model with reducing pooling layers and with three loss functions. The classifier is improved by training a CNN model using the encoder outputs along with the predicted saliency. In the scene classification tests using our experimental system and SALICON dataset, the proposed method is effective in comparison with baseline methods such as simple VGG classifiers.

**キーワード:** ディープラーニング, 視覚的顕著性, 車載カメラ, 画像認識, 交通事故リスク

## 1. はじめに

車外の複雑な風景や車内の過度な情報提示は、妨害刺激となって運転者の視認負荷を増大し、情報処理能力の逼迫による状況判断や運転操作の遅れをもたらすことが指摘されている。視認負荷の研究については、瞳孔径や心拍などの生体情報<sup>(1)(3)</sup>をはじめ、画像の光学的変化<sup>(4)</sup>、標識や信号など交通環境のアピランス<sup>(5)</sup>の解析に基づくもの、あるいは画像の物理的特徴由来の誘目性と運転時の視認行動を考慮したモデルに基づくもの<sup>(6)</sup>など数多く提案されており、注目度の高い分野として期待されている。そこで本研究では、交通事故リスクの低減に貢献することを目標とし、SALICONデータセット<sup>(7)</sup>を用いて画像からend-to-endで視認負荷を推定、分類問題への応用を提案する。

SALICONは、人間の視線に関する情報を収集した代表的なデータセットであり、スポーツ、動物、乗り物、料理など、さまざまなジャンルを含む屋内外の画像とともに、画像の自由視による実験で取得した視線データ、および視線データから生成した顕著性マップで構成される。顕著性マッ

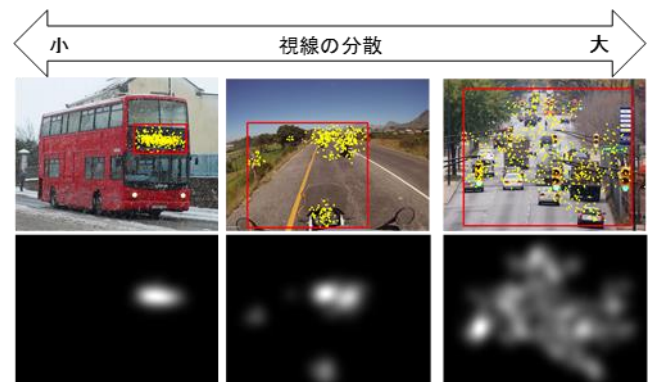


図1 SALICONデータセット

プは、人間の視覚的注意、すなわち視線の向きやすさを定量的に表現している。図1 はデータセットのサンプルを可視化したもので、上段は画像上に重ね合わせた視線データ、下段は顕著性マップである。データセットを観察すると、視線データ、すなわち顕著性マップが広く分散する画像は、視覚的に顕著な物体などが広く分散する傾向にあり、一方、顕著性マップが狭く集中する画像は、逆の傾向が

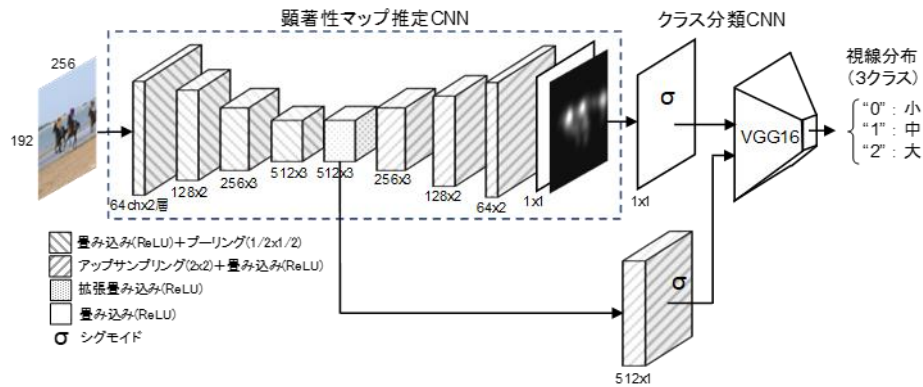


図2 提案手法の全体構成

みられる。このことから、たとえば特定物体を探すことを優先する運転タスクでは、顕著性マップが分散するほど妨害刺激が強くなり、風景全体を把握することを優先する運転タスクでは、顕著性マップが集中するほど妨害刺激が強くなる。いずれにおいても運転者の視認負荷が増大すると考えられる。

本研究では、人間の視線データ由来の顕著性マップを特徴量として用いて任意の風景での視線分布を推定し、その大きさに応じて高精度に分類する手法を提案する。SALICONデータセットを用いた実験により提案手法の有効性を示す。

## 2. 提案手法

### 2.1 画像分類手法の概要

画像を視線分布の大きさに応じて分類するため、顕著性マップ推定とクラス分類のそれぞれのタスクを実行するCNNを用いる。

図2に、提案手法の全体構成を示す。顕著性マップ推定CNNでは、3つの損失関数で最適化したエンコーダ・デコーダ構造を用いて、入力画像から顕著性マップを推定する。またクラス分類CNNでは、顕著性マップ推定CNNのエンコーダ出力と顕著性マップをあわせて畳み込むVGG16構造を用いて視線分布の大きさを推定することで、単純な構成で効果的に分類精度を向上する。

以下、顕著性マップ推定とクラス分類の高精度化について詳細に述べる。

### 2.2 顕著性マップ推定の高精度化

全体構成の前段に位置する顕著性マップ推定の精度は、後続のクラス分類に大きく影響することが考えられる。顕著性マップの推定手法としては、視覚的注意の外的要因である、画像の輝度、色、エッジ、方向などの低次特徴に注目した手法<sup>(8)</sup>をはじめ、物体領域に注目した手法<sup>(9)</sup>、オプティカルフローに注目した手法<sup>(10)</sup>など、さまざまな計算モデルが提案されている。一方、近年、人間の経験、知識、文脈理解などの内的要因を含めて高精度に推定できると

する、CNNを用いた手法<sup>(11)-(13)</sup>の提案が目立っており、公開ベンチマークであるSALICON Saliency Prediction Challenge<sup>(14)</sup>においても上位を占めている。そこで提案手法では、エンコーダ・デコーダ構造のCNNを用いて顕著性マップ推定を高精度化する(図2)。

エンコーダ部は、 $1/2 \times 1/2$ のプーリングを3回として解像度劣化を抑え、最終層で $5 \times 5$ の拡張畳み込み(dilated convolution)を行うことで大域特徴を抽出する<sup>(13)</sup>。またデコーダ部は、エンコーダ部と対称構造の畳み込み層と3回のアップサンプリングを行って解像度を徐々に復元することで、1回のアップサンプリングで復元する構造<sup>(13)</sup>に対して精度の高い顕著性マップを推定できることを、予備実験により確認した。さらにCNNを最適化する損失関数として、CC (Correlation Coefficient)、KLD (Kullback-Leibler Divergence)、およびNSS (Normalized Scanpath Saliency)の重みづけ線形和を用いることで、学習パラメータ数17.3Mの単純な構造ながら、(14)において、KLDで0.208(2位)、IG (Information Gain)で0.796(5位)に相当するスコアを獲得した(2020/3/3時点)。

### 2.3 クラス分類の高精度化

全体構成の後段に位置するクラス分類には、3クラスの出力層をもつVGG16構造<sup>(15)</sup>のCNNを用いる(図2)。画像分類の一手法としてよく知られているVGG16では、画像の低次特徴をもとに抽象度を高めた特徴量を用いて分類を行うが、提案手法では、顕著性マップを特徴量として用いて分類を行う。これにより視覚的注意が向きやすい重要物体のみを選択的に抽出するとともに、テクスチャの少ない空や地面、建造物の単調な壁面など、妨害刺激になる可能性の低い非重要物体を効率的に排除して、クラス分類を高精度化する。

また顕著性マップ推定において、デコーダの訓練誤差や汎化誤差も分類精度に影響することが考えられる。そこで提案手法では、顕著性マップ推定CNNのエンコーダ出力を取り出し、顕著性マップとあわせて特徴量として活用

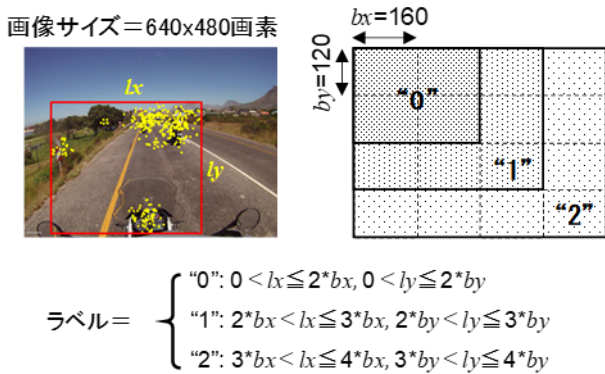


図3 ラベルの定義方法

することで、誤差の影響を相対的に抑制し、分類精度をさらに改善する。32x24x512chのエンコーダ出力をアップサンプリングするネットワークを追加した予備実験により、学習パラメータ数の増加は全体の0.4%と小さいながら、約8%の改善効果を確認した。

### 3. 実験

#### 3.1 データセット

SALICON データセットに含まれる視線データを用い、画像を視線分布の大きさに応じて3クラスに分類するためのラベルを新たに定義した。

図3 にラベルの定義方法を示す。初期的な分類実験を行うため、視野の形状を矩形とし、矩形内での視線分布の広がりのみに着目して単純化する。まず任意の画像に対応するすべての注視点の位置情報を参照し、それらを包含する最小の矩形 (lx, ly) を算出する。次に画像サイズを縦横にそれぞれ4分割して得られた長さ (bx, by) を用い、図中に示す定義式に従って付与するラベルを決定する。すなわち画像上の位置に依存しない、視線分布の大きさのみによるクラス分類とする。ここで“0”のラベルを付与された画像は最も視線が集中しやすく、“2”のラベルを付与された画像は最も視線が分散しやすいことを意味する。

実験に用いるデータセットは、SALICON データセットの訓練用および検証用のデータに対し、図3 の定義に従ってラベルを付与したものとする。

#### 3.2 実験条件

構築したデータセットには10000枚の訓練用データが含まれるが、そのままではクラス間のサンプル数の差異が顕著であったため、サンプル数の多いクラス“1”とクラス“2”については、それぞれランダムに1000枚のデータを選択し、サンプル数の最も少ないクラス“0”については、クラス間の移動が伴わない範囲で、学習時に幾何学的なデータ拡張を行って684枚に水増ししたものを、実験に用いる訓練用データとした。

表1 分類結果の比較

提案/VGG		“0”	“1”	“2”
推定結果	“0”	49/28	21/36	10/16
	“1”	196/167	633/544	356/474
	“2”	157/363	467/1100	3111/2272

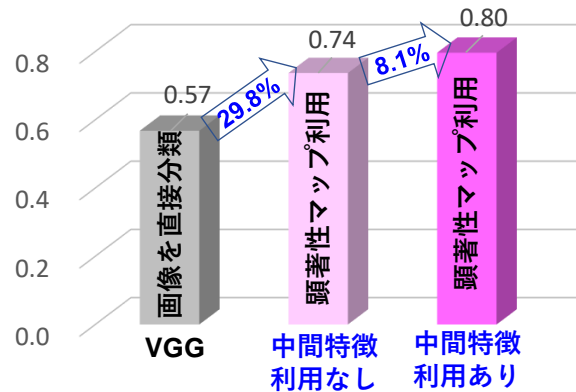


図4 分類精度の改善効果

画像分類の実験に先だち、まずSALICONデータセットの訓練用データをすべて用い、顕著性マップ推定CNNのパラメータを確定した。次にラベルを付与した訓練用データ2684枚を用い、顕著性マップ推定CNNの転移学習によりクラス分類CNNのパラメータを確定した。

分類精度の評価には、SALICONデータセットの検証用データ5000枚にラベルを付与したものをすべて用いた。

#### 3.3 結果

検証用データの画像5000枚の分類結果を表1 にまとめる。比較のため訓練用データの画像とラベルのみを用いて3クラス分類のVGG16を学習した場合の分類結果も併記する。提案手法の正解率は75.9% (VGG16は56.9%)となり、顕著性マップが効果的に分類精度の向上に貢献していることを確認した。また図4に分類精度の改善効果をまとめる。VGG16を用いて画像を直接分類する手法と比較し、顕著性マップの利用、さらには中間特徴の利用が有効であることを確認した。

図5 に正しく分類された画像の例を示す。スポーツ、交通、自然、および屋内の風景を代表的なカテゴリとし、各クラスに分類された画像、推定された顕著性マップ、および正解に相当するSALICONの視線データをまとめる。顕著性マップと視線データの関係に注目すると、定性的にも分布の一致性が高く、顕著性マップから正しく分類されていることがわかる。

図6に誤って分類された画像の例を示す。正解クラス“0”の画像がクラス“2”に誤分類されたサンプルにつ



カテゴリは上から「スポーツ」、「交通」、「自然」、「屋内」

図5 正しく分類された画像の例



図6 誤って分類された画像の例

いては、顕著性マップと視線データとの一致性が高い傾向にある。すなわち顕著性マップの推定は妥当であるため、クラス“0”の訓練用データを再考することで、さらなる分類精度の改善が考えられる。一方、正解クラス“2”の画像がクラス“0”に誤分類されたサンプルについては、顕著性マップと視線データとの一致性が低い傾向にある。すなわち顕著性マップの推定に大きな誤差が含まれているため、推定モデルを再考することで、さらなる分類精度の改善が考えられる。

#### 4. まとめ

本研究では、顕著性マップを用いて画像から視線分布の大きさを推定し、その大きさに応じてクラス分類する手法を提案した。

顕著性マップ推定には、解像度の劣化を抑えたエンコーダ・デコーダ構造のCNNを用い、3つの損失関数で最適化することで学習パラメータ数の増加を抑えつつ推定精度を高めた。さらにクラス分類には、顕著性マップとエンコ

ーダ出力を畳み込んで特徴量として活用するVGG16構造のCNNを用い、転移学習を行うことで分類精度を効果的に向上した。SALICONデータセットを用いて評価環境の構築と初期的な分類実験を行い、提案手法の有効性を確認した。

今後は、運転行動特性を含む運転者属性などを考慮した視線データの収集と特徴解析、視線分布や顕著性マップと視認負荷のより厳密な関係解析、上記知見に基づくデータセットの構築と運転者属性などに応じた分類手法の提案、および計算リソースの軽量化などに取り組んで実用性を高めていきたい。

#### 参考文献

- (1) S. P. Marshall, “The Index of Cognitive Activity: Measuring Cognitive Workload”, IEEE 7th Human Factors Meeting, 2002.

- (2) S. Tokuda, G. Obinata, E. M Palmer, and A. Chaparro, “Estimation of mental workload using saccadic eye movements in a free-viewing task”, Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Aug. 2011.
- (3) U. E. Manawadu, T. Kawano, S. Murata, M. Kamezaki, and S. Sugao, “Estimating driver workload with systematically varying traffic complexity using machine learning: Experimental design”, Proceedings of 1st International Conference in Intelligent Human Systems, pp. 106-111, Nov. 2018.
- (4) Y. Okafuji, T. Fukao, Y. Yokokohji, and H. Inou, “Design of a preview driver model based on opticalflow”, IEEE Transaction on Intelligent Vehicles, Vol. 1, No. 3, pp. 266-276, Sept. 2016.
- (5) N. Lyu, L. Xie, C. Wu, Q. Fu, and C. Deng, “Driver’s Cognitive Workload and Driving Performance under Traffic Sign Information Exposure in Complex Environments: A Case Study of The Highways in China”, International Journal of Environmental Research and Public Health, Vol. 14, Issue 2, 203, Feb. 2017.
- (6) 二反田, 坂東, 公文, 宮原, 玉津, 「走行環境における視認負荷量推定法」, MIRU2011, IS2-29, 2011年7月.
- (7) M. Jiang, S. Huang, J. Duan, and Q. Zhao, “SALICON: Saliency in Context”, IEEE Conference on Computer Vision and Pattern Recognition, Jun. 2015.
- (8) L. Itti, C. Koch, and E. Nieber, “A saliency-based visual-attention for rapid scene analysis”, IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 20, No. 11, pp. 1254-1259, 1998.
- (9) A. Borji, M. Cheng, H. Jiang, and J. Li, “Salient Object Detection: A Survey”, arXiv: 1411.5878, Nov. 2014.
- (10) M. Kodama, T. Kohama, and H. Yoshida, “A saliency based motion detection model for visual system considering visual adaptation properties”, 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp.6658-6661, Aug. 2015.
- (11) J. Pan, E. Sayrol, X. Giro-i-Nieto, “SalGAN: Visual Saliency Prediction with Generative Adversal Networks”, arXiv:1701.01081, 2017.
- (12) S. Jia, “EML-NET: An Expandable Multi-Layer NETWORK for Saliency Prediction”, arXiv:1805.01047, 2018.
- (13) M. Comia, L. Baraldi, G. Serra, and R. Cucchiara, “Predicting Human Eye Fixations via an LSTM-based Saliency Attentive Model”, IEEE Trans. On Image Processing, Vol 27, Issue 10, pp.5142-5154, Jun. 2018.
- (14) <http://salicon.net/challenge-2017/>
- (15) K. Simonyan and A. Zisserman, “A Deep Convolutional Neural Network for Large-Scale Image Recognition”, arXiv:1409.1556, 2015.

### 著者紹介

井上 俊明(いのうえ としあき)

パイオニア(株)変革推進本部に所属。

半導体回路・アーキテクチャ技術の研究, プラズマディスプレイ製品向けSoCの開発を経て, 現在, 画像認識アルゴリズムの研究に従事。