

Class Activation Mapping を用いた視線推定手法の評価および有効性の確認

Evaluation and Validation of Gaze Estimation Method with Class Activation Mapping

杉山 秀治[†] 福田 大翔[†] 中山 光典[†] 渡辺 裕[†]

Hideharu SUGIYAMA[†] Hiroto FUKUTA[†] Kousuke NAKAYAMA[†] and Hiroshi WATANABE[†]

[†] 早稲田大学

[†] Waseda University

Abstract Gaze Estimation is a technique for estimating the direction of human gaze and is indispensable for understanding human behavior. In previous studies, each method has been quantitatively evaluated using average angular error. This study shows the effectiveness of applying Class Activation Mapping to Gaze Estimation methods.

1. はじめに

視線推定とは、人間の視線方向を推定する技術のことである。人間の視線は、非言語コミュニケーションにおいて重要な要素であり、人間の行動理解にあたり視線推定は欠かせない技術である。近年、AR や VR、自動運転などのアプリケーションで視線推定が活用されており、さまざまな産業領域で応用可能な技術として注目されている。これまで視線推定手法は、平均角度誤差 (MAE) により評価されてきたが、評価用データセットにおける誤差の平均値であることから、被験者と推定精度の傾向やモデルの分析に向かないという問題があった。そこで本研究では、Class Activation Mapping[1]を用いることで、視線推定手法を視覚的に評価、分析することを提案し、有効性を示す。

2. 関連研究

2.1. L2CS-Net

L2CS-Net[2]は、アピアランスベースの視線推定手法であり、CNN を用いた手法の中で優れた性能を示すものである。物体認識タスクで有名な ResNet をバックボーンに使用しており、一般的な手法とは異なり、視線ベクトルではなく視線角度を回帰することで精度の向上を達成している。

2.2. Class Activation Mapping

Class Activation Mapping (CAM) は、CNN の特定のクラスの予測に寄与する画像の領域を可視化するための技術である。この技術は、CNN のブラックボックスとなっている内部の動作の理解するために活用できる。しかし、CAM は特定のネットワーク構造に依存するため、より汎用的な手法として Grad-CAM[3]が登場した。

Grad-CAM は、CNN の勾配情報を用いるため、あらゆる CNN アーキテクチャに応用可能であり、本研究でも利用する。さらに本研究では、Grad-CAM を改良した手法である Grad-CAM++[4]も利用する。

3. 提案手法

視線推定モデルに対して、CAM を適用し、得られたヒートマップ画像を基に手法の分析することを提案する。まず、学習済みの視線推定モデルに対して、評価用データセットの顔画像を入力として、CAM により視線を推定する際の判断根拠の可視化をおこなう。次に、得られたヒートマップ画像から被験者ごとのアテンションの傾向やモデルの判断傾向などの分析する。また、ヒートマップ画像に基づいて、モデルが意図通りに設計されているかを判断する。本研究では、視線推定用のデータセットの中で大規模な Gaze360[5]データセットで学習させた L2CS-Net のモデルに対して、Grad-CAM と Grad-CAM++を適用し、ヒートマップ画像を生成する。

4. 実験

4.1. L2CS-Net への CAM の適用

まず、Gaze360 データセットで学習済みの L2CS-Net のモデルに対して、Grad-CAM と Grad-CAM++を適用する。得られたヒートマップ画像の例を図 1 に示す。各画像は左から入力画像、Grad-CAM の結果、Grad-CAM++の結果である。そして、L2CS-Net による各入力画像の視線推定の角度誤差を表 1 に示す。図 1 の(a), (b), (c)は、Attention の分散が小さく、目の領域に対して高い Attention があり、推定精度も良い。一方で(d), (e), (f)は、Attention の分散が大きく、Attention が目以

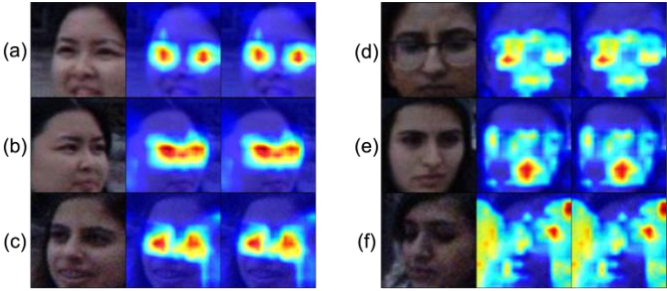


図 1 CAM による L2CS-Net の可視化画像の例

表 1 L2CS-Net による視線推定結果

入力画像	(a)	(b)	(c)	(d)	(e)	(f)
角度誤差	8.38	4.72	0.40	31.74	41.39	43.55

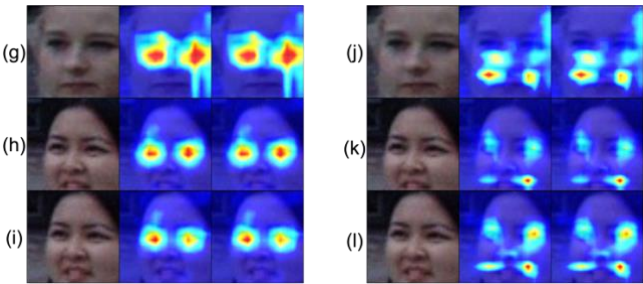


図 2 異なる手法に対する CAM の可視化画像の例

表 2 L2CS-Net と類似手法の視線推定結果

Method	MAE(°)
L2CS-Net	10.31
類似手法	10.97

外の領域で高く、推定精度も悪い。したがって、可視化画像と視線推定精度に相関があり、Attention の分散が小さく、目の領域への Attention が高い場合、視線推定の精度が良いことがわかる。

4.2. 異なる手法への CAM の適用

次に、異なる視線推定手法に対して Grad-CAM と Grad-CAM++ を適用し、可視化結果の違いを確認する。比較手法として、L2CS-Net の回帰対象を一般的な視線推定手法と同様の視線ベクトルとした手法を類似手法と呼ぶ。得られたヒートマップ画像の例を図 2 に示す。図 2 の (g), (h), (i) は L2CS-Net のモデルを可視化したものであり、(j), (k), (l) は同じ入力画像に対して類似手法のモデルを可視化したものである。また、従来手法 L2CS-Net と類似手法の MAE の値を表 2 に示す。ここで、図 2 の各画像に対する角度誤差は、類似手法より L2CS-Net の方が小さい。そして図 2 より、L2CS-Net のモデルの方が目の領域への Attention が高いことがわかり、角度誤差も小さいことから、異なる手法に対して CAM を適用しても、第 4.1 節で述べた Attention

と視線推定精度の相関が確認できる。さらに、表 2 より MAE の値が L2CS-Net の方が小さく、L2CS-Net のモデルの可視化画像の方が目の領域で Attention が高い傾向から、視線推定手法の良し悪しを CAM による可視化で判断できる可能性が示唆される。

5. まとめ

本研究では、視線推定手法における定性的な評価指標として CAM を使い、得られたヒートマップ画像を基に分析することを提案した。これまでの定量的評価指標である角度誤差と CAM による Attention の様子に相関があることを確認し、CAM によるモデル設計の良し悪しの判断が可能であることを示唆した。また、これまでの MAE の値では把握できなかった被験者に対する傾向やモデルの判断根拠が可視化画像により一目で確認できるため、モデルの詳細な分析に役立つと考えられる。一方で、現在の CAM による評価は可視化画像に基づく傾向分析であり、定量的評価ではないため、CAM での評価が可能であることを断定するためにはさらなる実験が必要である。

文 献

[1] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva and A. Torralba, "Learning Deep Features for Discriminative Localization," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2921-2929, (2016).

[2] A. A. Abdelrahman, T. Hempel, A. Khalifa, A. Al-Hamadi and L. Dinges, "L2CS-Net : Fine-Grained Gaze Estimation in Unconstrained Environments," 2023 8th International Conference on Frontiers of Signal Processing, pp. 98-102, (2023).

[3] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," 2017 IEEE International Conference on Computer Vision, pp. 618-626, (2017).

[4] A. Chattopadhyay, A. Sarkar, P. Howlader and V. N. Balasubramanian, "Grad-CAM++: Generalized Gradient-Based Visual Explanations for Deep Convolutional Networks," 2018 IEEE Winter Conference on Applications of Computer Vision, pp. 839-847, (2018).

[5] P. Kellnhofer, A. Recasens, S. Stent, W. Matusik and A. Torralba, "Gaze360: Physically Unconstrained Gaze Estimation in the Wild," 2019 IEEE/CVF International Conference on Computer Vision, pp. 6911-6920, (2019).

† 早稲田大学 基幹理工学研究科 情報理工・情報通信専攻
〒169-0072 東京都新宿区大久保 3-14-9 早大シルマンホール 401
TEL.03-5286-2509
E-mail: hide_garmat.1123@suou.waseda.jp