

## Face Parsing を用いた顔認証モデルの解釈 Interpretation of face recognition models using face parsing

河合 洋弥\*      神津 岳志\*      伊藤 康一\*      陳 煥宗†      青木 孝文\*  
Hiroya Kawai      Takashi Kozu      Koichi Ito      Hwann-Tzong Chen      Takafumi Aoki

キーワード 顔認証, Face Parsing, CNN, バイオメトリクス

### あらまし

顔の情報をを用いて個人を認証する顔認証は、最も有望な個人認証技術の一つとして近年注目されている。顔認証で用いる顔画像は、標準的なカメラを用いて離れた位置から撮影できるため、指紋認証や虹彩認証と比較して、顔認証は利便性やコストの面で優れる。スマートフォンのような携帯端末におけるログイン認証だけでなく、決済サービスや犯罪捜査等において広く利用されている。ポーズや表情、照明等の変化により認証精度が低下するため、認証精度の向上が顔認証分野における主な研究課題となっている。近年では、深層学習を用いた手法 [1, 2] が多数提案されている。数百万枚以上の顔画像からなる大規模データセットを用いて学習された Convolutional Neural Network (CNN) を用いることで、顔認証の認証精度が飛躍的に向上した。一方で、CNN は学習に基づいて自動で最適化された多数の重みパラメータを用いて推論を行うため、CNN の推論過程を解釈することは極めて困難である。CNN を用いる顔認証においてどのような顔のパーツや特徴が重要であるかを明らかにすることは、顔認証モデルの信頼性や認証精度の向上のために重要である。

CNN の推論根拠の可視化は、顔認証に限らず、CNN を用いる画像処理分野全体において重要な課題である。CNN が推論時に重視した入力画像中の領域をヒートマップとして可視化する手法がこれまでに多数提案されている [3, 4]。これらの手法は、対象物体が入力画像中のランダムな領域に出現する物体認識モデルに対しては有効

であるが、顔が入力画像中の大部分を占め、かつ画像の中央に位置合わせされている顔認証モデルに対しては有効に機能しない可能性がある。

そこで、本稿では、Face Parsing を用いた顔認証モデルの可視化手法を提案する。Face Parsing は、目、鼻、口、髪等の顔のパーツに関するピクセル単位のラベルマップを求める手法である。Face Parsing により得られたラベルに基づいてパーツごとのバイナリマスクを作成し、入力画像に乗算して異なる一部の領域のみを含む画像を複数生成する。パーツ画像を CNN に入力して得られる確率スコアに基づいてバイナリマスクの重み付け和を求めることで、顔認証モデルの推論根拠を可視化する。特定のパーツを含む画像のみで学習された顔認証モデルに対する可視化実験を通して、一般的な CNN の可視化手法より提案手法の方が顔認証モデルの推論根拠を正確に可視化できることを実証する。

### 参考文献

- [1] F. Schroff, D. Kalenichenko, and J. Philbin, “FaceNet: A Unified Embedding for Face Recognition and Clustering,” Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp.815–823, June 2015.
- [2] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, “ArcFace: Additive Angular Margin Loss for Deep Face Recognition,” Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp.4685–4694, June 2019.
- [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,” Proc. IEEE Int 1 Conf. Computer Vision, pp.618–626, Oct. 2017.
- [4] H. Wang, Z. Wang, M. Du, F. Yang, Z. Zhang, S. Ding, P. Mardziel, and X. Hu, “Score-CAM: Score-Weighted Visual Explanations for Convolutional Neural Networks,” Proc. IEEE Conf. Computer Vision and Pattern Recognition Workshops, pp.111–119, June 2020.

\* 東北大学 大学院情報科学研究科, 〒 980-8579 仙台市青葉区荒巻字青葉 6-6-05. Graduate School of Information Sciences, Tohoku University, 6-6-05, Aramaki Aza Aoba, Aoba-ku, Sendai-shi, Miyagi 980-8579, Japan.

† Department of Computer Science, National Tsing Hua University, 101, Section 2, Kuang-Fu Road, Hsinchu 300044, Taiwan.