Vol.2023
2024/1/27

眼鏡やマスクが画像認識モデルの表情認識精度に与える影響の検討と評価

横田晋太朗† 武田敦志†

眼鏡やマスクにより顔の一部が見えない場合に表情認識精度が大きく減少することが明らかになっている．本稿ではネットワークへの入力の直前で隠れている部位に 0 の値を代入することにより，隠された部位のデータを除外する方法を提案し，その評価を行う．隠れた部位のデータを検証から除外することにより，表情認識精度の減少が穏やかになることが明らかになった．また，隠す部位や人種による認識精度の変化についても検証を行ったが，人種による有意な差はみられないという結果を得た．

Performance Evaluation of Facial Image Recognition Models when the Images contain Masks or Glasses

Shintaro Yokota† and Atsushi Takeda†

Our observations indicate that the accuracy of facial expression recognition significantly diminishes when certain parts of the face are obscured by glasses or masks. In this paper, we propose and assess a method to ignore data from concealed facial regions by replacing the obscured part with a value of 0 to the obscured part prior to feeding it into the network. The exclusion of data from concealed regions during validation results in a modest decline in the accuracy of facial expression recognition. We also investigated the variation in recognition accuracy based on the concealed area and race; however, no significant differences by race were found in these results.

† 東北学院大学教養学部情報科学科
Department of Information Science, Tohoku Gakuin University

1. はじめに

深層学習を用いた画像認識技術が急速に発展したことにより，顔画像からその人の表情を高い精度で認識することが可能となった．一方，新型コロナウイルスの流行にともないマスクを着用する場面が増えており，眼鏡（サングラス）やマスクなどによる顔の一部が隠れている状態であっても顔画像からその人の表情を高い精度で認識する技術が期待されている．一般的に，目や口が隠れた状態であれば，その顔画像から表情を認識することは簡単ではないと考えられる．また，マスクをした状態の顔画像から表情の認識を試みた研究では，マスクを着用することにより表情認識精度が低下することが報告されている[1]．

そこで，本稿では，眼鏡やマスクにより顔の一部が隠されている場合であっても正確に表情を認識する手法を検討する．具体的には，顔画像の眼鏡やマスクの部分について，白色や黒色で塗りつぶした場合と深層学習モデルへの入力値を 0 とした場合の表情認識モデルの性能を評価した．この評価結果より，眼鏡やマスクの部分については深層学習モデルへの入力値を 0 とすることにより表情認識性能を向上できることが実験を通じて確認できた．

さらに，目元と口元を隠した顔画像に対する表情認識モデルの性能を比較し，表情を認識するためには口元の情報の方が重要であることを確認した．また，東洋人と西洋人では表情を認知する部位が異なると言われることがあるため，東洋人と西洋人に分割したデータセットを用いて表情認識精度を検証することにより，東洋人と西洋人で表情が現れる部位に違いがあるかを検証した．この検証により，表情が現れる顔の部位については東洋と西洋で大きな違いはないことを確認した．

2. 関連研究

先行研究では，マスク非着用時の顔画像での表情認識の精度と比較して，マスク着用処理後の顔画像での表情認識の精度は 87.51%から 75.16%まで低下したという結果が出ている[1]．しかし，この研究では実際のマスク着用に近い条件となるよう，顔画像の口元を白で埋めて実験を行っていた．そのため隠された顔の一部を検証に使用しない場合の精度の変化は明らかになっていない．

表 1 先行研究での認識制度

Dataset	Non-mask	Mask-wearing
FER+	87.51%	75.16%
RAF-DB	87.19%	78.88%

3. 眼鏡やマスクをした顔画像に対する表情認識手法

顔画像から表情を識別するためのニューラルネットワークは大規模な顔画像データセットを用いて学習処理を実施する．一般的な顔画像データセットの場合，眼鏡やマスクで顔の一部が隠されている顔画像の割合は少ない．そのため，このデータセットを用いて学習した表情認識モデルは，眼鏡やマスクを含まない顔画像の表情認識精度に対して，眼鏡やマスクで顔の一部が隠されている顔画像の表情認識精度は低くなる傾向がある．眼鏡やマスクと顔の表情とは関係を持たない事象である．そのため，表情認識モデルに入力するデータから眼鏡やマスクに関する情報を削除すれば，表情認識モデルが眼鏡やマスクの情報の影響を受けなくなるため，より正確に顔画像の表情を認識できると考えられる．

表情認識モデルには画像データに対して正規化などの前処理を行ったデータが入力されるが，この入力データの一部の値を 0 とすることで，表情認識モデルが対象となる領域について計算することを抑制できる．そこで，眼鏡やマスクで顔の一部が隠された顔画像に対してその眼鏡やマスクの領域の入力データの値を 0 に設定することで，表情認識モデルが眼鏡やマスクの影響を受けずに推論することが可能となり，眼鏡やマスクによって顔の一部が隠された顔画像であっても正確に表情を認識できると考えられる．

4. 実験評価

4.1 実験方法

本研究では，マスク非着用状態の顔画像を対象に，ニューラルネットワークを用いて学習および検証を行った．まずは，先行研究[1]をベースラインとするため，この先行研究と同じデータセットと表情認識モデルを用いた実験評価を行った．データセットには，RAF-DB と FER+の中から個人でも利用可能な FER+を用いる[2]．このデータセットは，48x48 ピクセルの白黒の顔写真を 8 つの表情で分類するもので，各画像は各表情の度合いを表し，その合計が 10 になるように構成されている．ただし，曖昧な表情や低品質な画像における誤学習を防止するため，一部の不正確な画像データの再ラベリングを行った．また，表情認識モデルとして Self Cure Network[3]を実装した．この表情認識モデルには，学習過程で各画像の重要度に重みづけを行い，その重みをランク付けし正則化する仕組みがある．ただし，Self Cure Network はクラス分類のためのニューラルネットワークであり，FER+データセットのような重みの学習方法については明らかになっていない．そこで，それぞれの画像の表情ラベルの重みが最も大きいカテゴリーを代表クラスとし，この代表クラスを正解ラベルとして Self Cure Network の学習処理を実施した．

以上の実装を用いて表情認識モデルの学習を実施し，眼鏡やマスクを含まない顔画像の表情認識精度が 83%となる表情認識モデルを作成した．さらに，目元や口元を隠した画像データを用いて作成した表情認識モデルの表情認識精度を検証した．これは，眼鏡やマスクにより顔の一部が隠されている画像データの表情認識精度の検証を目的としている．顔画像の目元や口元を隠すため，Face Alignment Network を用いて顔画像のランドマークを検出し，ランドマークから目元や口元の領域を判別した．これらの領域を白色や黒色で塗りつぶすことにより，眼鏡（サングラス）やマスクにより顔の一部が隠された顔画像を生成する．

4.2 実験結果

提案手法により，サングラスやマスクを着用した状態の顔画像において，単純な白や黒，ノイズで隠す方法では精度が著しく低下することが確認されました．一方で，本論文で提案した隠す部位の入力データの値を 0 とする手法を用いると，目元を隠す場合は 74.11%，口元を隠す場合は 60.34%までの精度の低下に抑えられた．この結果から，顔の一部分がマスクやサングラスによって隠れている場合，その部分の入力データの値を 0 とすることにより，そのままの画像を入力した場合よりも正確に表情を認識できることが確認できた．

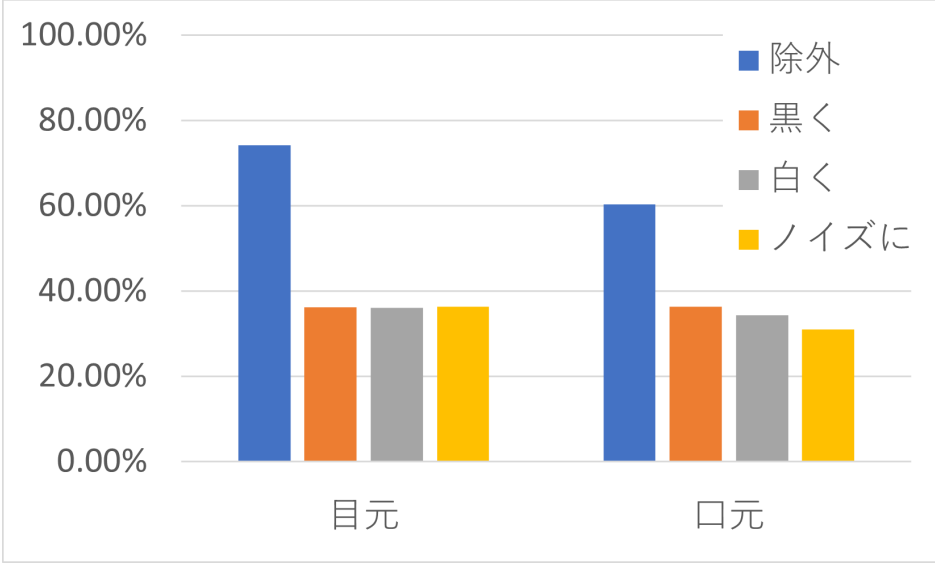

図 1　マスク着用処理方法ごとの認識精度

4.3 目元と口元を隠した際の精度の変化の比較

目元と口元を隠した場合の表情認識精度を比較した．先行研究[1]では口元を隠す実験のみが行われたが，ここでは目元及び口元を隠した場合の精度の差を検証した．ここで実装した表情認識モデルに対して目元を隠した画像を入力した場合は 74.11%の精度で表情を識別でき，口元を隠した画像を入力した場合は 60.34%の精度で表情を識別することができた．この結果は，表情認識において口元の情報が目元の情報よりも重要であることを示唆している．続いて，各表情における表情二錦精度を比較したところ，目元を隠した画像よりも口元を隠した画像の表情認識精度が低くなる傾向が見られた．図 3 に目元を隠した画像を入力したときの画像認識精度を示し，図 4 に口元隠した画像を入力したときの画像認識精度を示す．ここで，disgust や contempt の表情の画像データに対する表情認識精度が他と比較して低い結果となっているが，これらの表情については学習データの数が少なく，これらの表情について十分な学習でできなかったためと思われる．

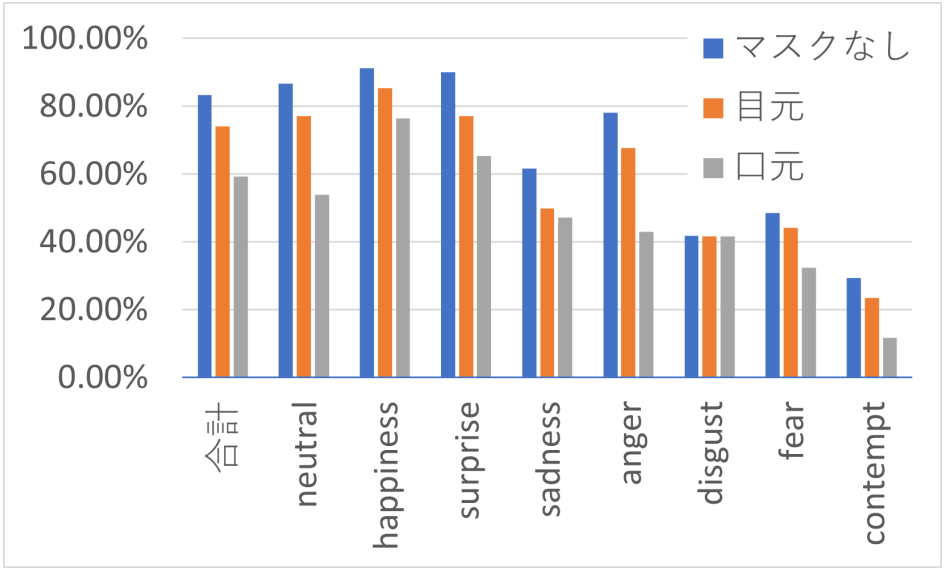

図 2　隠す部位とクラスごとの認識精度

予測クラス

正解クラス	neutral	happiness	surprise	sadness	anger	disgust	fear	contempt
neutral	0.774	0.068	0.025	0.098	0.029	0.001	0.002	0.004
happiness	0.053	0.867	0.010	0.035	0.031	0.000	0.003	0.001
surprise	0.112	0.036	0.753	0.026	0.052	0.000	0.021	0.000
sadness	0.324	0.097	0.018	0.484	0.055	0.005	0.016	0.000
anger	0.124	0.088	0.044	0.076	0.653	0.000	0.016	0.000
disgust	0.083	0.083	0.000	0.167	0.417	0.250	0.000	0.000
fear	0.118	0.118	0.206	0.088	0.088	0.015	0.368	0.000
contempt	0.353	0.235	0.000	0.118	0.000	0.059	0.000	0.235

図 3　目元を隠した際の各クラスの正答・誤答の傾向

予測クラス

正解クラス	neutral	happiness	surprise	sadness	anger	disgust	fear	contempt
neutral	0.539	0.337	0.023	0.077	0.022	0.000	0.002	0.000
happiness	0.154	0.765	0.031	0.031	0.019	0.001	0.000	0.000
surprise	0.128	0.159	0.654	0.013	0.005	0.000	0.042	0.000
sadness	0.268	0.200	0.008	0.471	0.037	0.000	0.016	0.000
anger	0.179	0.231	0.036	0.120	0.430	0.000	0.004	0.000
disgust	0.000	0.333	0.000	0.250	0.000	0.417	0.000	0.000
fear	0.118	0.147	0.294	0.074	0.044	0.000	0.324	0.000
contempt	0.294	0.353	0.000	0.118	0.059	0.000	0.059	0.118

図 4　口元を隠した際の各クラスの正答・誤答の傾向

4.4 日本と欧米の表情の文化差に着目

表情認識モデルを用いた実験結果より，東洋と西洋の表情文化の差について考察する．東洋人は目元で表情を認知する傾向があり，一方で西洋人は口元で表情を認知すると言われることがある．この文化的な差異から，円滑な表情の理解を促進するために，重要な部位で表情がより強く現れる可能性について実験的に考察する．具体的には，東洋人の顔画像データと西洋人の顔画像データを用意し，それぞれの顔画像の目元と口元を隠した場合の表情認識精度の低下を調べ，それぞれの顔画像の目元と口元に表情判断のための情報が強く現れているかどうかを検証する．

ここでは，顔画像データセットである FER+のテストデータを目視により東洋人と西洋人に分類し，東洋人は 451 枚，西洋人は 1925 枚の 2 つのデータセットを作成した．目元を隠した顔画像データを表情認識モデルに入力したところ，東洋人の顔画像は

72.06%，西洋人の顔画像 75.74%の精度で表情を識別できた．一方，口元を隠した顔画像データを表情認識モデルに入力したところ，東洋人は 57.21%，西洋人は 60.10%の精度で表情を識別できた．どちらのテストでも表情認識精度に有意差は見られず，目元よりも口元を隠した画像データの表情識別制度が低い結果となった. この結果より，東洋と西洋のどちらの文化圏においても表情に関する情報は口元に強く表現されると考えられる．

5. まとめ

本稿では，眼鏡やマスクなどによって顔の一部が隠れている状態でも高い精度で表情を認識する手法を検討し，眼鏡やマスクにより顔の一部が隠されている顔画像の場合，眼鏡やマスクの部分についてニューラルネットワークへの入力値を 0 とすることにより，そのままの画像データを入力する場合に比べて高い精度で表情を認識できることを確認した．また，表情認識においては目元に関する情報よりも口元に関する情報が重要であることが実験的に示された．さらに，東洋人と西洋人での表情の違いを検証した結果，どちらの文化圏においても表情に関する情報は口元に強く表現されることが示唆された．

参考文献

1) 呉 強, 浜田 宏一, 荒井 正之, "マスク着用画像を用いた表情認識に関する研究 -マスク無し顔画像との表情認識性能の比較-," 第 84 回全国大会講演論文集, 2022, pp. 593-594.
2) Emad Barsoum, Cha Zhang, Cristian Canton Ferrer, and Zhengyou Zhang, "Training deep networks for facial expression recognition with crowd-sourced label distribution," ICMI '16: Proceedings of the 18th ACM International Conference on Multimodal Interaction, 2016, pp. 279–283.
3) K. Wang, X. Peng, J. Yang, S. Lu and Y. Qiao, "Suppressing Uncertainties for Large-Scale Facial Expression Recognition," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 6896-6905.
4) A. Bulat and G. Tzimiropoulos, "How Far are We from Solving the 2D & 3D Face Alignment Problem? (and a Dataset of 230,000 3D Facial Landmarks)," 2017 IEEE International Conference on Computer Vision (ICCV), 2017, pp. 1021-1030.