

閉塞領域を含む顔画像からの表情認識に関する研究

2019SE014 門脇啓斗 2019SE024 小島圭人

指導教員：沢田篤史

1 はじめに

2019年新型コロナウイルスの影響で多くの人がフェイスマスクを着用することが義務付けられた。表情の閉塞によりコミュニケーションをとる際、感情を読み取るのが困難になった。特にネガティブ（恐怖、怒り、嫌悪、悲しみ）の感情は口と鼻の動きに大きく影響するので、認識精度が下がる。

本研究の目的はフェイスマスクによる閉塞下においても表情認識の精度を保つことのできる顔画像認識技術の確立である。そのために本研究では、次の2つの技術課題を設定する。

1. 閉塞下の表情認識精度を上げるためのデータ拡張方法を明らかにする
2. データ拡張の表情認識精度向上への効果を明らかにする

フェイスマスク閉塞下で表情認識する時に、フェイスマスクをした画像データが少ないので表情を豊かに認識ができないことが問題点として挙げられる。

本研究の具体的なアプローチは次の通りである。

1. 顔の特徴を強調したデータを拡張させる
2. データ拡張を行っていく上で顔の特徴点と可動量を表すアクションユニットを参考にする
3. 表情認識のためのCNNを構築し、拡張したデータにより学習させ、表情認識精度を評価する

データ拡張（DA）とは、訓練用のデータに「変換」を加えることでデータを水増しする手法である。元画像を回転させたり、拡大、縮小をすることで別の画像とする。このように訓練データを増やすことで、認識精度を改善することができる。西銘ら [1] の研究では、驚き、恐怖、嫌悪、怒り、喜び、悲しみの6つの感情ごとに顔の特徴点と可動量を表すアクションユニットを整理している。

本研究の場合、考えられるDA方法は2つあり、フェイスマスク部分（鼻、口）に変換を加えるか、フェイスマスク部分以外（目、眉）に変換を加えることである。本研究では高い認識精度を出すために2つの方法を考慮したデータ拡張を行う。また、4層のCNNを構築し、データ拡張したものを訓練データとして学習させる。

2 閉塞下での表情認識の課題

2.1 背景

2019年新型コロナウイルスの世界的な蔓延に伴い、公共の場でコミュニケーションをとる場合、フェイスマスクを着用することは感染を防ぐための一般的な行動になって

いる。フェイスマスクを着用した対人コミュニケーションは頻繁に行われているが、表情が分かりづらいことが問題であった。

2.2 表情認識における問題

フェイスマスクを着用した顔画像の表情認識は、ポジティブ（驚き、幸せ）の感情の場合、主に眉毛や目元が大きく影響するので閉塞の影響は少ない。一方で、ネガティブ（恐怖、嫌悪、怒り、悲しみ）の感情は閉塞された口や鼻などに大きく影響することから認識率が低くなってしまふことが課題である [2]。

2.3 関連研究

フェイスマスクによる閉塞を考慮した表情認識を対象とした研究として2つを取り上げる。

- 閉塞部分以外を抽出し表情認識行う方法
- 閉塞領域と非閉塞領域の重みづけを行う方法

以下、それぞれの研究について説明した後、問題点について議論する。

Giovanna ら [3] の研究では、フェイスマスク着用の有無を判別し、それを表情認識に活かす方法を提案している。フェイスマスクを着用していなければそのまま表情認識を行い、着用していればフェイスマスクで閉塞されていない部分だけを抽出して認識を行う方法をとっている。ポジティブな感情（幸せ、驚き、平均精度 79.00 %）は、ネガティブな感情（怒り、嫌悪、恐怖、悲しみ、平均精度 23.00 %）よりもよく認識されていることがわかった。

Yang ら [4] では、表情認識精度を向上させるために、フェイスマスクの有無を判別する分類器とフェイスマスクの領域と非領域で重み付けを行う重み付け器で構成される2段階のモデルを提案している。提案された方法は、自然の場合と人工的につけられたフェイスマスクの場合の両方で、他の最先端の表情認識方法よりも優れていた。

この2つの研究の成果は、フェイスマスクした顔画像をそのまま使用し学習器を構成するなどの他のアプローチに比べると認識精度は高かったが、実用的に十分とは言えないことが報告されている。特に閉塞部分をあまり考慮していないので、ネガティブの感情の認識精度が低いことが問題である。

3 本研究の目的と技術課題

3.1 目的

本研究はフェイスマスクによる閉塞下においても表情認識の精度を保つことのできる顔画像認識技術の確立することを研究目的とする。関連研究の問題点でも挙げている通

り、閉塞部分をあまり考慮していないことが認識精度の低下に繋がっている。閉塞下での表情認識の精度向上に繋げるには、閉塞部分からいかにして情報を得るかが鍵になる。閉塞部分の特徴をより多く、確実に捉えることで、表情認識の際に用いることのできる情報が増えると考えている。

3.2 技術課題

フェイスマスクによる閉塞領域を考慮した表情認識を行う上での技術課題は次の2つが挙げられる。

1. 閉塞下の表情認識精度を上げるためのデータ拡張方法を明らかにする
2. データ拡張の表情認識精度向上への効果を明らかにする

フェイスマスクによる閉塞下で表情認識する時に、フェイスマスクをした画像データが少ないので表情を豊かに認識できないことが問題点として挙げられる。

2章で述べた Giovanna ら [3] の研究と Yang ら [4] の研究ではデータが少ない問題に対し、顔画像に人工的なフェイスマスクをつけることで解決していた。本研究では認識向上に繋げるために自然にフェイスマスクをつけたデータを増やすにはどのような方法が考えられるかを明らかにする。また増やしたデータを CNN の学習器で用いたときに認識率がどのように変化するか明らかにする。具体的な方法は4章で述べる。

4 データ拡張と CNN を用いた表情認識手法

4.1 提案手法の概要

本研究では、精度の高い表情認識を達成するために顔の特徴を強調したデータを拡張 (DA) させ、CNN を用いた表情認識手法を提案する。提案する手法の概要を図1に示す。

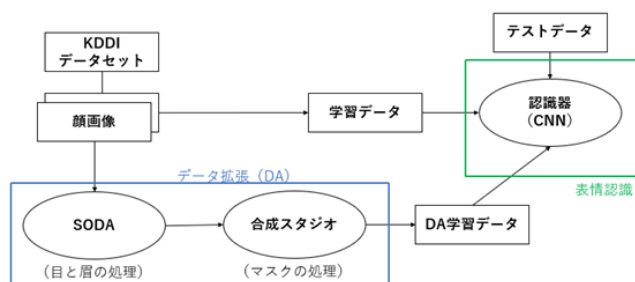


図1 提案手法の概要

4.2 データ拡張

データ拡張とは既存のデータに対し処理を施すことによって、データの量を人為的に増やす技術である。例えば、漁業における害鳥認識システムに用いられている [5]。害鳥を認識する課題に対し、AI は別のものと判断してしまう問題があるとすると、元画像を回転させたり、拡大、縮小をすることで別の画像とする。このように訓練

データを増やすことで、AI の認識精度を改善することができる。したがって、本研究に DA を採用する。

4.3 アクションユニット

西銘ら [1] の研究では、驚き、恐怖、嫌悪、怒り、喜び、悲しみの6つの感情ごとに顔の特徴点と可動量を表すアクションユニット (Action Unit, 以下 AU) を整理している。本研究では AU を参考に DA を行っていく。

4.4 DA の方法

本研究での効果的な結果が推測できる DA の方法は2つある。

- 閉塞部分以外 (目、眉) の処理
- 閉塞部分 (鼻、口) の処理

本研究では、両方を考慮した DA を行うことでより精度の高い表情認識ができると期待している。AU を参考にすると、本研究ではポジティブは喜び、ネガティブは怒り、ニュートラルは真顔とした。ニュートラルのアクションユニットはないため、目と眉の処理は行っていない。

ポジティブとネガティブの AU はそれぞれ3つずつに絞りまとめる。

ポジティブ・・・眉の内側を上げる、頬を上げる、唇両端を引き上げる

ネガティブ・・・眉の外側を上げる、眉を下げる、鼻のしわを寄せる

4.4.1 目と眉の処理

目と眉の処理は SODA*1 というアプリで手作業で行う。SODA は顔写真を読み込み、眉や目など顔のパーツの大きさや角度を自在に変えられるものである。度量は-100% から 100% となっている。

ポジティブ AU の眉の内側を上げるを強調するために、眉の角度を-100%、眉の位置を 100% とした。実際に変更したものを図3に示す。

ネガティブ AU の眉の外側を上げると眉を下げるを強調するために、眉の角度を 100%、眉の位置を-100% とした。また眉を下げることで目が吊り上がることから目の角度も 100% とした。実際に変更したものを図4に示す。

4.4.2 フェイスマスクの処理

フェイスマスクの処理は合成スタジオ*2 というアプリで手作業で行う。合成スタジオでは目と眉の処理を行った画像に対して、フェイスマスクをつける。実際に各表情の AU の特徴点と可動量を考慮したときに、フェイスマスクのしわと形は図2のようになる。これを目と眉の処理を行った画像に合成する。

ポジティブ AU の頬をあげるを考慮するとフェイスマス

*1 <https://apps.apple.com/jp/app/soda-ソーダ/id1437880869>

*2 <https://apps.apple.com/jp/app/合成写真-背景透過>

クが目の付近まで上がるので、そのように顔画像にフェイスマスク合成する。実際に変更したものを図3に示す。

ネガティブ AUの鼻にしわを寄せるを考慮するとフェイスマスクの真ん中の部分だけが上がるのでフェイスマスクは図2のような形になる。実際に変更したものを図4に示す。

ニュートラル ニュートラルには対応するAUがないので、図2のような普通のフェイスマスクを合成する。実際に変更したものを図5に示す。



図2 左: ポジティブ, 中: ネガティブ, 右: ニュートラル

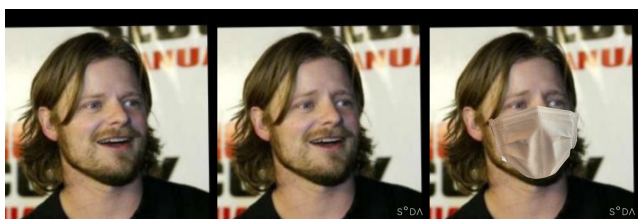


図3 左: 元画像 中: 目と眉処理 右: マスクの処理 (ポジティブ)



図4 左: 元画像 中: 目と眉処理 右: マスクの処理 (ネガティブ)



図5 左: 元画像 右: マスクの処理 (ニュートラル)

4.5 データセット

4.5.1 訓練データ

訓練データはDAを行わないものを行うものの2種類の顔画像を用いる。DAを行わない顔画像には自動でフェイスマスクを装着したデータセットのM-LFW-FERデータセット[6]の中からポジティブの顔画像700枚, ネガティブの顔画像700枚, ニュートラルの顔画像700枚, 計2100枚用いる。DAを行う顔画像はフェイスマスクを着用していない顔画像のデータセットであるLFW-FERデータセット[6]の中からポジティブの顔画像700枚, ネガティブの顔画像700枚, ニュートラルの顔画像700枚, 計2100枚を選びDAを行ったものをデータセットとした。

4.5.2 テストデータ

テストデータには、25名のRGBカラーの顔画像を用いる。事前処理として、画像の顔領域のみを抽出し、CNNの入力層に合わせ、画像サイズを320×240にする。参加者はフェイスマスクを着用した状態で、ポジティブの表情4枚, ネガティブの表情4枚, ニュートラルの表情1枚, 計25枚を撮影し、それをテストデータとして使用する。

4.6 開発環境

Pythonは機械学習のライブラリやフレームワークが多い。そのため、本研究の使用言語はPythonとする。ライブラリはTensorFlow, Keras, Sklearnを使用する。

4.7 畳み込みニューラルネットワーク (CNN)

[7]ではバッチサイズ128で畳み込み層はrelu関数を用いている。また, [8]では4層でバッチサイズ128で、畳み込み層にrelu関数を用いている。[9]では畳み込みにrelu関数を、全結合にはsoftmax関数を用いている。[10]では4層で畳み込み層にrelu関数を用いて重み付けを行っている。表情認識のCNNの構造はバッチサイズ128で、4層で畳み込み層にrelu関数を用いて、全結合にsoftmaxを用いて重み付けを行っている研究が多かったので本研究のCNNの構造は図6に示すように定義した。4層のネットワークで構成され、バッチサイズは128、畳み込み層はrelu関数を用い、全結合はsoftmax関数を用いて、epochはlossがepoch数200で収束したことから200として重み付けを行った。

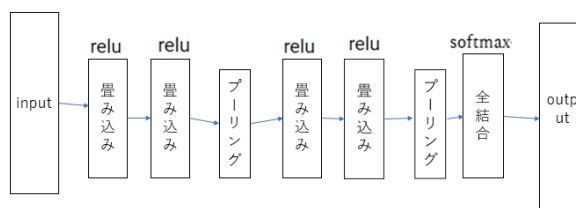


図6 CNN 構造

5 結果と評価

本研究の結果を表1に示す。ポジティブの認識率は8%

表1 実験結果

	DAなし	DAあり
ポジティブ	46%	38%
ネガティブ	21%	41%
ニュートラル	16%	24%

低下、ネガティブの認識率は20%増加、ニュートラルの認識率は8%増加とDAをすることでネガティブとニュートラルの認識率を上げることができた。一方で、DAありとDAなしとで認識率の差があまり見られない結果となった。

6 考察

認識率が変化していない原因として、訓練データ数の不足がある。本研究の訓練データはポジティブの顔画像700枚、ネガティブの顔画像700枚、ニュートラルの顔画像700枚、計2100枚であり、さらに枚数を増やすことで認識率の改善が見込まれる。また、データの解像度の低さも原因として挙げられる。LFWデータセット、M-LFWデータセットは解像度が低く画質の質が低かった。したがって、より高精細の画像を拡張することで精度の向上を見込むことができる。

Yangら[12]の研究では、自動顔面マスク装着法(AWFM)を提案している。この提案手法と本研究のマスクのしわや形を考慮した方法をうまく合わせることができれば、閉塞下での表情認識に特化したデータを増やすことができると考えている。

7 おわりに

2019年新型コロナウイルスの影響で多くの人が日常的にフェイスマスクを着用するようになった。ネガティブ(恐怖、怒り、嫌悪、悲しみ)の感情は口と鼻の動きに大きく影響するので、フェイスマスクで閉塞されている場合認識精度が下がる。

フェイスマスク閉塞下で表情認識する時に、フェイスマスクをした画像データが少ないので表情を豊かに認識できないことが問題点として挙げられる。本研究の具体的なアプローチは次の通りである。

1. 顔の特徴を強調したデータを拡張させる
2. データ拡張を行っていく上で顔の特徴点と可動量を表すアクションユニットを参考にする
3. 表情認識のためのCNNを構築し、拡張したデータにより学習させ、表情認識精度を評価する

ポジティブは8%低下、ネガティブは20%増加、ニュートラルは8%増加という結果になった。認識率が変わらなかった原因として訓練データ数の不足と画質の質が低かっ

たことが上げられる。高画質の画像でDAをして訓練データ数を増やすことができれば精度の向上を見込むことができる。

参考文献

- [1] 西銘大喜, et al. “ディープニューラルネットワークによる画像からの表情表現の学習” 人工知能学会全国大会論文集 第29回 2015. 一般社団法人人工知能学会, 2015.
- [2] Ligang Zhang, Brijesh, Verma Dian, Tjondronegoro, Vinod Chanrav, “Facial Expression Analysis under Partial Occlusion: A Survey”, ACM Computing Surveys, Vol. 51, No. 2, Article 25. Publication date: April 2018.
- [3] Giovanna Castellano, Berardina De Carolis, Nicola Macchiarulo, “Automatic Emotion Recognition from Facial Expressions when Wearing a Mask”, CHIItaly’ 21, July 11–13, 2021, Bolzano, Italy
- [4] Yang Bo, Wu Jianming, Gen Hattori, “Face mask aware robust facial expression recognition during the COVID-19 pandemic”, 2021 IEEE International Conference on Image Processing (ICIP). IEEE, 2021.
- [5] 北風裕教, et al. “Data Augmentation を用いた CNN 学習画像の増加による害鳥認識システムの認識率の改善”, 産業応用工学会論文誌 7.2 2019: 69-76.
- [6] Yang Bo, Wu Jianming, Gen Hattori, “Facial Expression Recognition with the advent of face masks”, MUM 2020, November 22–25, 2020, Essen, Germany
- [7] AmineHorseman “facial-expression-recognition-usingcnn”, 2019, <https://github.com/amineHorseman/facial-expression-recognition-using-cnn>
- [8] RishabhJain “Facial-Expression-Recognition”, 2018, <https://github.com/rishabhjainps/Facial-Expression-Recognition>
- [9] ShawDa “facial-expression-recognition” 2018, <https://github.com/ShawDa/facial-expression-recognition>
- [10] thoughtworksarts “EmoPy”, 2021, <https://github.com/thoughtworksarts/EmoPy>
- [11] Yang, Bo, Jianming Wu, and Gen Hattori, “Facial expression recognition with the advent of face masks”, 19th International Conference on Mobile and Ubiquitous Multimedia 2020.