

深層学習を利用した顔画像に対する年齢および性別推定モデルにおいてマスク着用および人種の違いが同時に与える影響

Simultaneous effects of wearing face mask and variation of race to facial age and gender estimation model using deep learning

窪寺 健太[†] 秋葉 将和[‡] 堀田 忠義[‡]
Kenta Kubotera Masakazu Akiba Tadayoshi Horita

1. はじめに

近年の AI 技術の進歩により、人の顔画像から年齢や性別といった顔属性の推定を行う技術が、様々な分野で用いられている。代表的な顔属性推定技術の活用例は、小売店舗におけるマーケティング[1]である。一方で、近年の COVID-19 パンデミックの影響により、マスクの着用率が増加しているため、顔画像から年齢や性別を推定する顔属性推定システムの、精度低下が懸念されている。

先行研究においては、人種の違いが年齢推定に与える影響に関する研究として、例えば文献[2]がある。またマスク着用が年齢推定に与える影響に関する研究として、例えば文献[3]がある。しかしながら、人種の違いとマスク着用の有無という 2 つの要因が、同時に年齢推定に与える影響を検討した他者研究は、見当たらない。

そこで本研究の目的は、画像分類モデル Xception[4]を畳み込みベース（特徴抽出機能のベース）とした、顔画像に対する年齢および性別推定モデルにおいて、人種の違いおよびマスク着用の有無という 2 つの要因が同時に存在する事が、推定結果に与える影響について報告することである。

2. 顔属性推定モデル

図 1 に、本研究で使用した推定モデルを示す。本研究で用いる学習モデルは、Xception モデルを畳み込みベースとして利用する転移学習モデルである。Xception モデルとは、画像分類で効果を発揮する Inception モデル[5]を改良したものである。ImageNet データセット[6]を用いた画像分類では、Xception モデルは Inception V3 モデル[5]よりも優れた性能を発揮するため、本研究の畳み込みベースとして採用する。

3. 訓練用データセット

本研究の訓練用データセット作成に際し、UTKFace データセット[7]を利用する。この UTKFace は、以下の特徴をもつ。

- 0 歳から 116 歳の人物の顔画像を含む。
- 合計で 23,708 枚の顔画像を収録している。
- 各画像は、タグ情報として、顔の特徴点座標(目、鼻、口等の位置情報)、年齢、性別および人種のデータを持つ。
- 人種ラベルは、白人、黒人、アジア人、インド人、その他（ヒスパニック系、ラテン系、中東系等）の 5 種類である。

UTKFace の画像すべてに対し、図 2 のように、顔の特徴点座標情報（緑色の点）を参照して、画像の口部分周辺の画素の色を機械的に変更する事により、疑似的なマスクあり顔画像を生成する。この生成されたマスク画像を、後述

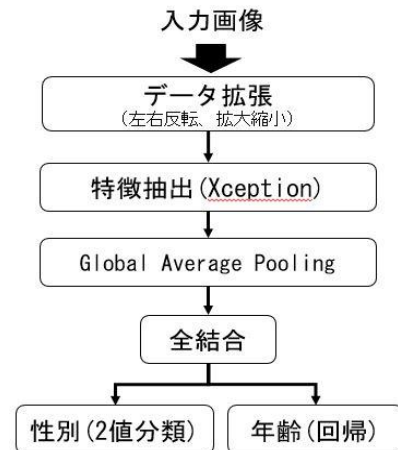


図 1 顔属性推定モデル

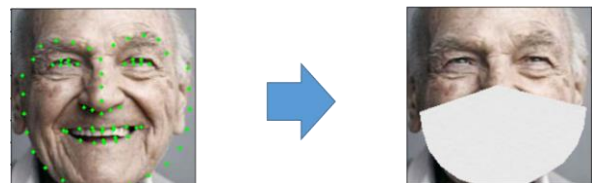


図 2 マスク顔画像の作成方法

の「マスクあり」訓練用データセットの画像として利用する。

なお UTKFace において、性別についての写真枚数に関する偏りは、ほぼない。しかしながら、人種および年齢については、以下に述べるような偏りが存在する。

- 他の人種と比べ、白人の写真枚数が最多であり、全体の 40%以上を占める。
- 年齢層については、いずれの人種においても、20～30 歳の枚数が、他の年齢層の枚数よりも多い。
- 0～5 歳の年齢層に注目すると、黒人の枚数は、それ以外の人種に比べて極端に少ない。

このような偏りがあるため、本研究では、白人以外の 1 つの人種を選び、それと白人の場合とを比較する、というスタイルで実験を進める。例えば選んだ人種が黒人の場合、以降では、「白人と黒人の組み合わせ」あるいは「白人—黒人」のように表記する。

[†] 静岡職業能力開発促進センター Shizuoka Polytechnic Center

[‡] 職業能力開発総合大学校 Polytechnic University

「白人—黒人」の場合に使用される 4 つの訓練用データセット（「白人マスクなし A」, 「白人マスクあり A」, 「白人マスクなし A」および「黒人マスクあり A」）について、ここで説明する。

- 「白人（または黒人）マスクなし A」訓練用データセットの写真は、UTKFace そのままの写真の中から、人種情報が白人（または黒人）のもので、かつ年齢情報が 20～50 歳のものを、ランダムに 3500 枚選択する。
- 「白人（または黒人）マスクあり A」訓練用データセットの写真は、図 2 のように作成したマスク有り顔画像集合の中から、同様に選択する。

加えて、「白人—インド人」の場合に使用される 4 つの訓練用データセット（「白人マスクなし B」, 「白人マスクあり B」, 「インド人マスクなし B」および「インド人マスクあり B」）について、ここで説明する。

- 「白人（またはインド人）マスクなし B」訓練用データセットの写真は、UTKFace そのままの写真の中から、人種情報が白人（またはインド人）のもので、かつ年齢情報が 0～2 歳のものから 300 枚、22～27 歳のものから 1100 枚、および 33～55 歳のものから 1200 枚（これらの合計は 2600 枚）を、それぞれランダムに選択する。
- 「白人（またはインド人）マスクあり B」訓練用データセットの写真は、図 2 のように作成したマスク有り顔画像集合の中から、同様に選択する。

「白人—黒人」および「白人—インド人」の場合で、訓練用データセットにおける年齢層区分及び写真枚数が異なる理由は、前述の UTKFace データセットにおける「偏り」による。

また上記に挙げた以外の人種の組み合わせとしては、「白人—アジア人」および「白人—その他人種」がある。前者の場合の 4 つの訓練用データセットを構成しようとする場合、同様の理由により、データセット 1 つ当たりの画像枚数は最大で 1396 となる。加えて、後者の場合のその枚数は、これ以下となる。このような枚数では、信頼できる検証結果が得られないと判断されるため、本研究ではこれらの場合の結果を示さない。

4. 顔属性推定の計算機実験

本研究で用いた計算機環境を、表 1 に示す。

各場合における 4 つのデータセット各々につき、図 1 に示す構造のモデルを、それぞれ学習する。ただし、各モデルの学習および評価においては、層化 k 分割交差検証[8]（分割数 5）を用いる。

こうして得られた各モデルの年齢推定誤差および性別推

表 1 計算機環境

OS	Windows 10
GPU	GeForce RTX-3060
CPU	Inter® Core™ i7-11700
RAM	32GB
ライブラリ	Python 3.9.12
	Keras 2.8.0
	Tensorflow 2.8.0

表 2 「白人—黒人」の場合のデータ

訓練用データセット名	年齢誤差(年)			性別 (%)
	平均	SD	マスク差	
白人マスクなし A	4.833	3.848	0.512	96.3
白人マスクあり A	5.345	4.273		94.4
黒人マスクなし A	4.415	3.799	0.352	95.1
黒人マスクあり A	4.767	3.939		93.4

表 3 「白人—インド人」の場合のデータ

訓練用データセット名	年齢誤差(年)			性別 (%)
	平均	SD	マスク差	
白人マスクなし B	4.904	4.538	0.329	91.4
白人マスクあり B	5.233	5.178		89.9
インド人マスクなし B	4.516	4.629	0.088	92.0
インド人マスクあり B	4.604	5.295		90.9

定精度に関するデータを、表 2 および 3 に示す。これらの表における各ラベルの意味は、以下の通りである。

- 「年齢誤差」は、年齢推定絶対誤差を意味し、入力画像に対する(モデルによる年齢の推定値)と(年齢属性情報)間の差の絶対値により算出する。
 - 「平均」は、「年齢誤差」の平均値である。
 - 「SD」は、「年齢誤差」の標準偏差値である。
 - 「マスク差」は、同じ人種についての、(マスクありの場合の「年齢誤差」平均値)と(マスクなしの場合の「年齢誤差」平均値)間の差の絶対値である。
 - 「性別」は、性別推定正解率であり、入力画像に対してモデルが推定した性別が、性別情報と一致している割合を示す。

また表 2 および 3 の「年齢誤差」部分をグラフ化したものを、図 3 に示す。

表 2, 表 3 および図 3 より、以下が確認できる。

- 「白人—黒人」および「白人—インド人」両方の場合において、マスクありおよびなし両方の場合で、他の人種よりも白人の場合の方が、年齢誤差が大きい。この主な理由は、肌の色の明るさの影響であると考えられる。つまり、肌の色が明るくなればなるほど、光の反射率が高くなるため、顔の特徴の代わりにその反射が画像に写ってしまう(つまり、その分だけ顔の特徴量が画像から失われる)から、と考えられる。
- 「白人—黒人」および「白人—インド人」両方の場合において、マスクありおよびなし両方の場合で、人種間の性別推定正解率の差はほぼない。つまり、この部分については、前述の肌の色の明るさの影響はほぼない。

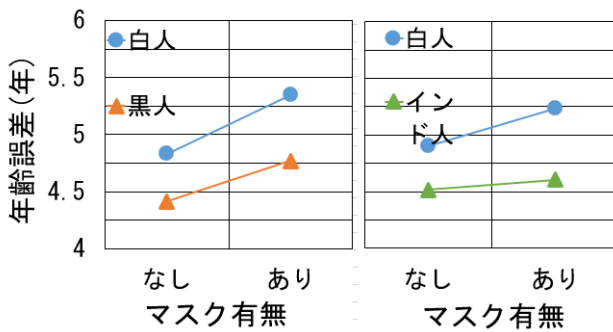


図 3 年齢誤差データ(白人—黒人, 白人—インド人)

- 「白人—黒人」および「白人—インド人」両方の場合において、他の人種の場合よりも白人の「マスク差」の値が大きい。
- 容易に予想される結果ではあるが、本研究で行った全ての場合において、マスクありの場合の方の誤差（および精度）が、マスクなしの場合に比べて大きく（および低く）なっている。

さらに、上記の層化 k 分割交差検証の過程において、最も年齢誤差が小さかった各場合のモデルについて、人種の違いおよびマスク有無に関して、以下のような条件で二元配置分散分析を行っている。

- マイクロソフトエクセルの「分散分析」ツールを使用する。
- そのモデルを学習した時と同じ検証用データを、この解析においても使用する。
- 帰無仮説として、「各要因（人種の違い、マスク有無）で年齢誤差の違いはない」を設定する。
- 対立仮説として、「各要因（人種の違い、マスク有無）で年齢誤差の違いがある」を設定する。
- 有意水準は 0.05 とする。

このような分散分析の結果、以下が確認された。

- 「白人—黒人」の年齢誤差の比較について、人種の違いおよびマスク有無それぞれの主効果が有意であった ($F(1,2796)=3.84, p<.005$)。また交互作用は認められなかった。
- 「白人—インド人」の年齢誤差の比較について、マスク有無の主効果が有意傾向 ($F(1,2076)=3.85, 05<p<.01$)であった。また人種の違いの主効果および交互作用は認められなかった。

上記の通り、「白人—黒人」および「白人—インド人」両方の場合において、人種の違いおよびマスク有無の交互作用は、有意ではなかった。しかしながら、前述の『他の人種の場合よりも白人の「マスク差」の値が大きい』ことは、他の人種と比較して、白人の年齢の特徴が口付近に表れやすい、という可能性を示している。

5. おわりに

本研究では、画像分類モデル Xception を畳み込みベースとした、顔画像に対する年齢および性別推定モデルにおいて、人種の違いおよびマスク着用の有無という 2 つの要因が同時に存在する事が、推定結果に与える影響について調査している。

「白人—黒人」および「白人—インド人」という 2 つの場合について、他の人種の場合よりも白人の推定年齢誤差の「マスク差」が大きいことが分かった。また分散分析の結果からは、その両要因の年齢推定に対する交互作用は認められなかった。しかしながら、特に「白人—黒人」の場合では、「人種の違い」および「マスク有無」の主効果がそれぞれ有意であった。

今後の課題として、「白人の年齢の特徴が口付近に表れやすいという可能性」について、さらに詳しく調査する。またこの調査結果より、人種の違いとマスク有無という 2 つの要因が同時に推定精度に与える影響について、より詳しい報告を行う。

参考文献

- [1] 三和コンピュータ株式会社：“顔認識技術のマーケティング分野での活用方法”，https://www.sanwa-comp.co.jp/article/column/face_market, 参照日：2023-08-19.
- [2] Andraž Puc, Vitomir Štruc and Klemen Grm：“Analysis of Race and Gender Bias in Deep Age”，Proceedings of 28th European Signal Processing Conference (EUSIPCO 2020), pp.830-834(2021)
- [3] 大野 朝香, 金城 光：“男性顔画像と女性顔画像の部分遮蔽が年齢推定に与える影響”，日本顔学会誌, Vol.23, No.2, pp.51-59 (2023).
- [4] François Chollet：“Xception: Deep Learning with Depthwise Separable Convolutions”，arXiv:1610.02357, pp. 1-8 (2017).
- [5] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens and Zbigniew Wojna：“Rethinking the Inception Architecture for Computer Vision”，arXiv:1512.00567, pp.1-9(2015)
- [6] Stanford Vision Lab：“IMAGENET”，<https://www.image-net.org/>, 参照日：2024-05-14
- [7] GitHub Pages：“UTKFace Large Scale Face Dataset”，<https://susanqq.github.io/UTKFace>, 参照日：2023-08-19.
- [8] scikit-learn developers：“3.1. Cross-validation: evaluating estimator performance”，https://scikit-learn.org/stable/modules/cross_validation.html