

# 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-

呉 強† 浜田 宏一† 荒井 正之†

帝京大学大学院理工学研究科†

## 1. はじめに

新型コロナウイルスの感染拡大により、マスク着用が日常化したことで、生活の各方面に不便をもたらしている。その1つは、会話中に相手の表情が読み取れず、コミュニケーションが難しくなる可能性があるということである。我々は、マスク着用画像から表情を認識する検討を進めている。

本稿では、顔画像データセット(FER+, RAF-DB)からマスク着用画像を生成し、表情認識実験を行った結果について考察する。

## 2. マスク着用画像データの作成

### 2.1 マスク着用処理手順

図1に本研究のマスク着用処理の流れを示す。まず、マスク無しの顔画像をFAN(Face Alignment Network)[1]に inputs する。FANは、Hourglass ネットワーク[2]をベースにして作られたネットワークであり、図1に示すように顔画像から68個のランドマークを検出することができる。本研究では、68個のランドマークのうち14個(赤い点の部分)を使用し、マスクを描画した。

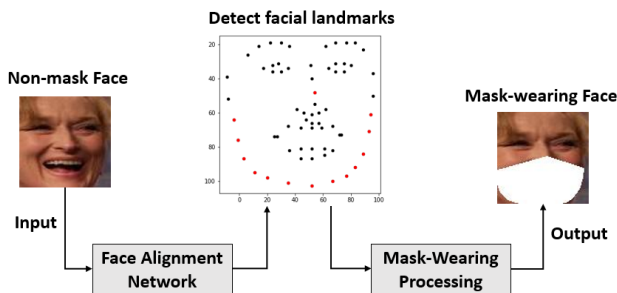


図1 A flowchart for adding a mask on a non-mask face image.

### 2.2 マスク着用処理結果

マスク着用処理の結果を表1に示す。FER+ [3]には合計28127枚、そのうち795枚(2.83%)、またRAF-DB [4]には合計15339枚の画像があり、そのうち114枚(0.74%)に対してマスク着用処理ができなかった。

マスク着用処理ができなかった画像の一例を図2に示す。

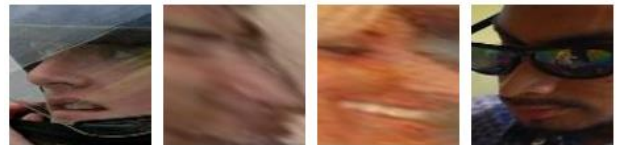


図2 Example images of which the landmarks cannot be detected.

## 3. 実験結果と考察

表2はSCN[5]を認識器とした場合の実験結果である。マスク着用処理により、顔の下半分が見えない画像でも正解率の低下が10ポイント程度であることがわかった。

クラス別の正解率を図3,4に示す。「Disgust」の正解率が最も低かった。表1に示すように「Disgust」の画像は全画像数の0.5%(FER+)と5.7%(RAF-DB)と少数であるため、「Disgust」の特徴がうまく学習できていない可能性がある。また、正解率の高いクラスは「Happiness」、「Neutral」、「Surprise」であった。これら3つのクラスの画像枚数が全体の7割以上を占めており、学習に多くのデータを利用できたことが影響していると考えられる。

RAF-DBにおけるマスク着用画像に対する混同行列(Confusion Matrix)を図5に示す。

表2 Overall accuracy in the FER+ and RAF-DB.

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

Facial Expression Recognition for Mask-Wearing Faces - Comparison of Expression Recognition between Non-mask Faces and Mask-wearing Faces-

†Qiang Wu, Kouichi Hamada, Masayuki Arai, Graduate School of Science and Engineering, Teikyo University

表 1 Total number of images by class and number of images that failed the masking process.

FER+	Surprise	Fear	Disgust	Happiness	Sadness	Anger	Neutral
total	3564	594	143	8152	3365	2387	9922
failure	113	19	0	134	238	61	230
failure ratio (%)	3.2	3.2	0	1.6	7.1	2.6	2.3

RAF-DB	Surprise	Fear	Disgust	Happiness	Sadness	Anger	Neutral
total	1619	355	877	5957	2460	867	3204
failure	11	4	2	21	29	13	34
failure ratio (%)	0.7	1.1	0.2	0.4	1.2	1.5	1.1

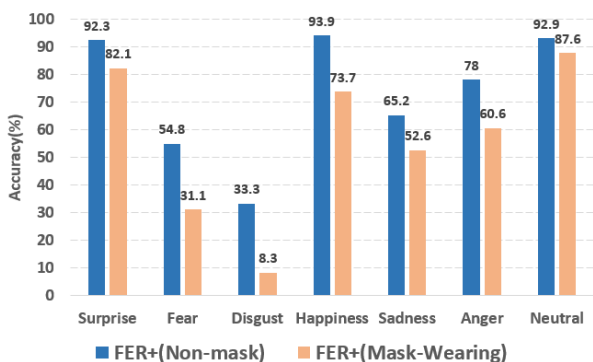


図 3 Recognition accuracy of each class for the FER+.

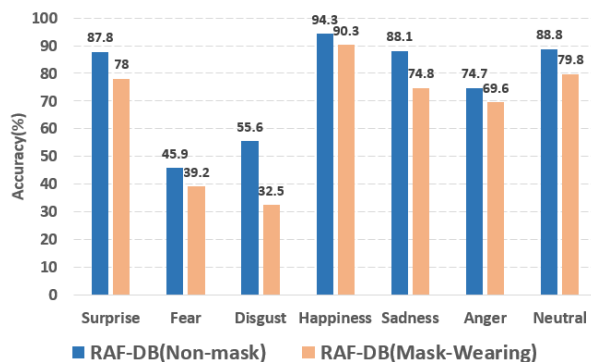


図 4 Recognition accuracy of each class for the RAF-DB.

同一クラスに分類される確率が最も高いが、「Fear」の多くが「Surprise」へ、また「Disgust」の多くが「Neutral」へ誤認識されていることがわかった。これらのクラスの表情の重要な特徴が口の周りにあるため、マスクによってそれらの特徴が失われ、モデルが十分に学習できていないことが一因と考えられる。

マスク着用処理を施し、それらを用いてマスク無し画像と表情認識性能の比較を行った。その結果、マスク着用画像はマスク無し画像に対して正解率が約 10 ポイント程度低下することがわかった。

今後は、混同行列のさらなる分析、認識結果に大きく影響を与える箇所同定、その他性別、人種などの属性による分析などを行う。その後、認識手法について検討を行う。

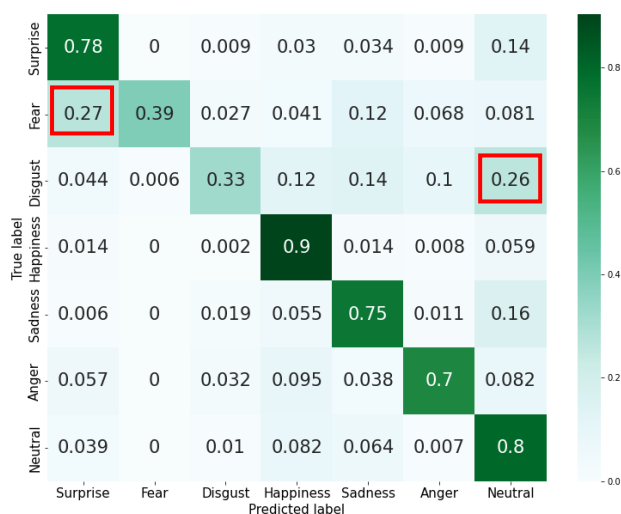


図 5 Confusion matrix on masked RAF-DB.

#### 4. おわりに

2 つのデータセット(FER+, RAF-DB)対して、

#### 参考文献

[1] A. Bulat, et al. "How far are we from solving the 2d & 3d face alignment problem?(and a dataset of 230,000 3d facial landmarks)." In ICCV2017, pp. 1021-1030.  
 [2] A. Newell, K. Yang, and J. Deng. "Stacked hourglass networks for human pose estimation." In ECCV2016, pp. 483-499.  
 [3] Barsoum Emad, et al. "Training deep networks for facial expression recognition with crowd-sourced label distribution." In Proc. 18th ACM Int. Conf. Multimodal Interact, pp. 279-283. 2016.  
 [4] Li Shan, et al. "Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild." In CVPR2017, pp. 2852-2861.  
 [5] Wang Kai, et al. "Suppressing uncertainties for large-scale facial expression recognition." In CVPR2020, pp. 6897-6906.