

マスク着用顔画像の表情認識を目的とした SCN-SAM の提案

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

顔認識における重要な研究の1つとして、表情認識はすでに生活の中で広く使われている。例えば、人間とマシンとのインタフェースでは、人間の表情を認識・分析し、処理に対して正しい結果を出すことができれば、マシンはより人間的になる。そのため、現在は研究対象として注目されている。

しかし、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ている。現在の表情認識手法は、顔全体の画像に対して非常に高い精度を実現することができる。しかし、マスクを付けた顔画像では、顔の特徴の一部が失われるため、表情認識が困難になることがある。また、一般的な顔認識問題では、大規模な画像データセットに画像の不鮮明さや画像の不正確なアノテーションにより、モデルの学習が難しくなり、認識性能の低下を招くことがある。

そこで、本研究では、マスクを付けた顔画像の表情認識を実現し、その精度を向上させるために重要度の調整モジュール (Self-Adjustment Module) を構築して SCN-SAM (Self-Cure Network with a Self-Adjustment Module) モデルを提案する。

2. マスク着用画像データの作成

本研究では、2つのデータセット RAF-DB [1]、FERPlus [2] に対して実験を行った。これら2つのデータセットから7種類の表情画像「Neutral (無表情・平常心)、Happiness (喜び)、Surprise (驚き)、Sadness (悲しみ)、Anger (怒り)、Disgust (嫌悪)、Fear (恐怖)」は、マスク着用後の画像に対しても同じラベルを使用して、表情認識を行う。

RAF-DB と FERPlus の画像はマスクなし顔画像であるため、その画像にマスクを付ける必要がある。マスク着用処理の流れを図1に示す。まず、マスクなし顔画像をニューラルネット

SCN-SAM: A Modified Self-Cure Network for Facial Expression Recognition Under Face Masks

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

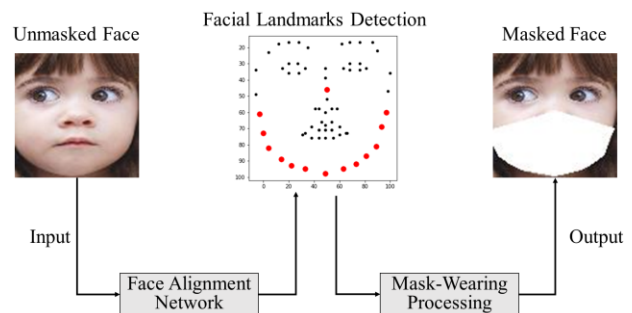


図1 A flowchart for adding a mask on an unmasked face image.

ワーク FAN (Face Alignment Network) に入力し、顔画像から68個のランドマークを検出する。本研究では、68個のランドマークのうち14個(赤い点)を使用し、マスクを描画した。

3. 提案手法

本研究で提案する SCN-SAM は、Wang ら[3]の研究 SCN (Self-Cure Network) に基づいている。SCN-SAM のフレームワークの詳細を図2に示す。SCN-SAM には深層学習を用いる。したがって、大規模のデータが必要であるため、データを拡張する。大規模な表情画像データセットから取得した画像の一括処理で、画像にある顔の部分にマスクを付加する。

次に、ニューラルネットワークを用いて、マスク着用顔画像の特徴量を抽出する。ここで利用されるニューラルネットワークのバックボーンは CNN である。CNN から得られた特徴を画像重要度の計算モジュール AIWM (Attention Importance Weighting Module) に入力する。AIWM は、シグモイド関数を用いて、抽出された各画像サンプルの特徴の重要度重みを計算することができる。曖昧な画像サンプルの重要度を下げるために、重要度の調整モジュール SAM (Self-Adjustment Module) を設計し、重要度によって画像サンプルを分ける。

正則化モジュール IWRM (Importance Weighting Regularization Module) では、得られた重要度の高い画像サンプルと低い重要度のサンプルを降順に選べ替え、2つのグループに分ける。また、重要度重み付け正則化損失 IWR-Loss 損失関数を用い、モデルの学習がその損失

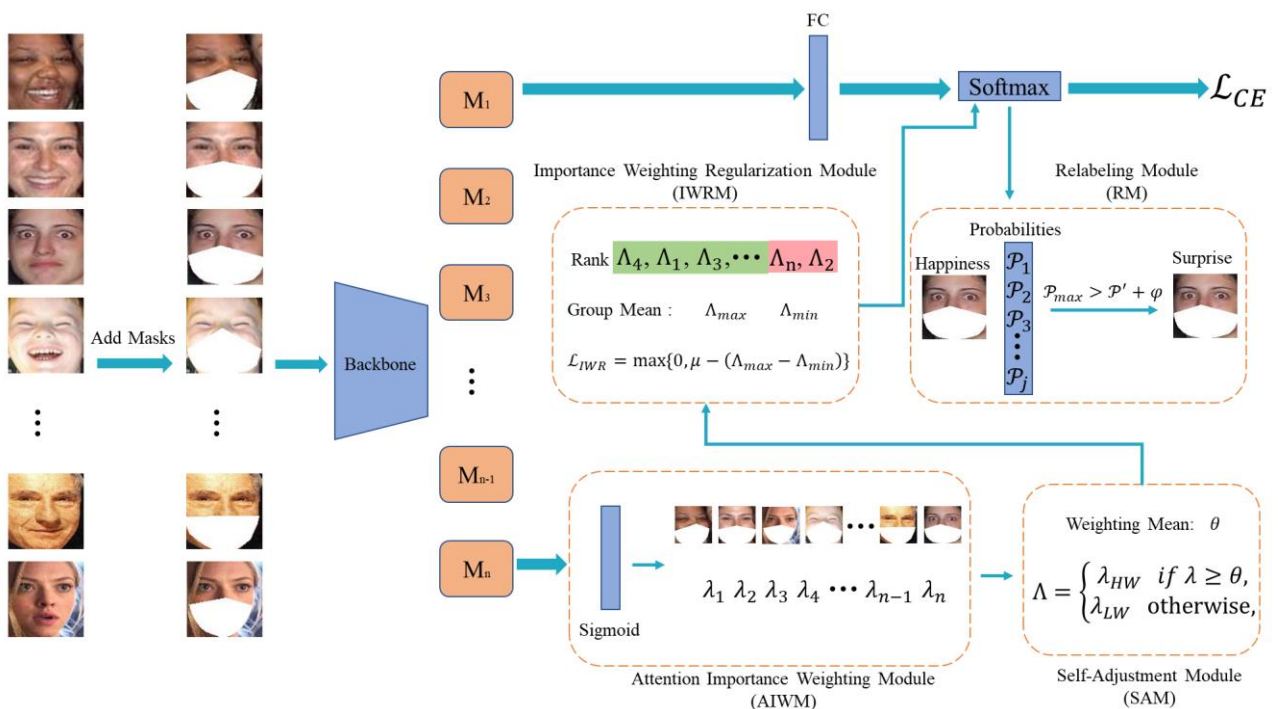


図 2 The structure of our SCN-SAM.

関数によって制約を受ける。最後に、ラベルの修正モジュール RM (Relabeling Module) を使い、一定回数で学習した後、重要度の低い画像サンプルの最大予測確率のラベルと元のラベルの確率を比較し、その差がある閾値を超えた場合に、そのラベルが最大予測確率のラベルへ修正する。

4. 実験結果と考察

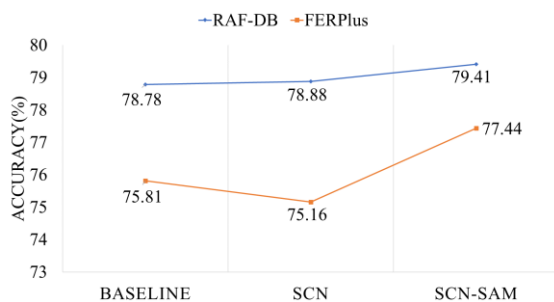


図 3 Comparison to SCN on masked RAF-DB and FERPlus.

マスク着用顔画像の表情認識結果を図 3 に示すように、SCN-SAM はベースライン (ResNet-18) と SCN を一貫して上回っている。RAF-DB に対して行った評価実験では、SCN はマスクによる曖昧なラベルにうまく適応できず、ベースラインとほぼ同じ性能になった。しかし、我々の SCN-SAM はマスクによる曖昧なラベルの影響に適応でき、SCN の精度よりも 0.53% 高くなっている。FERPlus の実験結果は更に明白であ

る。SCN がベースラインの精度よりも低く、負の性能を示している。SAM を用いることで、その精度は 2.28% 向上した。

不鮮明な画像の増加やマスクの存在による曖昧なラベリングの影響により、高重要度グループの画像数 δ 値は、非常に不安定な状態になる。SCN-SAM は、モデルのトレーニングにあたり、重要度の高いサンプルをマイニングし、高重要度のサンプルと低重要度のサンプルを分類することで、高重要度のサンプル数 δ 値が一定でなくなるので、この問題をうまく解決している。

5. おわりに

本研究では、マスク着用顔画像の表情認識の認識率を向上するため、SCN-SAM モデルを提案した。SCN-SAM では、SCN より優れた性能を発揮し、80% 弱の認識率を得ることができた。

今後の課題は、SCN-SAM モデルの汎用性を高めて、表情認識以外のコンピュータビジョン分野に適用することである。

参考文献

[1] Li Shan, et al. "Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild." In CVPR2017, pp. 2852-2861.
 [2] 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.
 [3] Wang Kai, et al. "Suppressing uncertainties for large-scale facial expression recognition." In CVPR2020, pp. 6897-6906.