

# 同一人物の正面の顔を用いた StyleGAN2 による 顔のマスク領域の補完

小池 泰景<sup>†</sup> 河合 紀彦<sup>†</sup>

大阪工業大学 情報科学部 情報メディア学科<sup>†</sup>

## 1. はじめに

近年、コロナウイルスが世界的に大流行しており、人物を撮影する際にはマスクを着用している機会が多くなっている。そのため、顔の下半分が隠されている写真をウェブページや SNS にアップロードした場合、撮影された人の魅力が伝わりにくくなっている。

このような問題を解決する方法として、機械学習の一種である Generative Adversarial Networks (以下, GAN) を用いたマスク領域の補完手法が提案されている<sup>[1]</sup>。この手法では2つの識別機を用いることによって顔の一貫性と欠落領域を段階的に学習している。しかし、学習に用いられた顔画像に依存した鼻や口の補完となっているため、本人の特徴を持つ顔画像となっているとは限らない。

そこで本研究ではマスクを着用している画像とは別に用意したマスクを着用していない同一人物の画像と、StyleGAN2<sup>[2][3]</sup>を用いて、本人の特徴を有した口や鼻の補完を行う。

## 2. 提案手法

### 2.1 概要

提案手法では、まずマスクを着用している画像とマスクを着用していない同一人物の画像の二種類を入力とする。前処理として、入力画像の口や鼻が所定の位置にくるように整列させる。次に、StyleGAN2 で前処理した画像の潜在変数を推定した画像を生成する。続いて StyleGAN2 の特徴である Style Mixing を用いてマスクを着用していない画像の顔の向きを、マスクを着用している顔の向きに合わせて変形させる。表情を変更したい場合は、追加で顔画像を一枚入力し、同様に Style Mixing を用いて、その顔画像の口の特徴を反映させる。最後に Poisson Image Editing<sup>[4]</sup>によって変形した画像を、マスクを着

用している画像に色調を合わせながら合成し、マスク部分を補完する。以下、提案手法の詳細な処理について述べる

### 2.2 顔画像の整列

今回使用した StyleGAN2 の学習モデルは大量の顔画像がまとめた FFHQ データセットであり、学習には顔のランドマークの位置が同じ位置に整列されたものが使われている。そのため、StyleGAN2 に入力する前処理として同様に整列しないと画像の生成品質が低くなってしまふ。そこで機械学習のライブラリである Dlib を用いて顔のランドマークを検知し、整列を行う。眼の高さが同じになるように回転し、目の位置と口の位置から中心座標を求め、正方形になるように画像を切り抜く。

### 2.3 潜在変数の推定

StyleGAN2 では、学習データの様々な要素が分布している潜在空間上の一点である潜在変数を入力することで、それに対応した画像が生成される。また、StyleGAN2 を用いて顔画像を編集するには、編集したい顔画像を StyleGAN2 で生成する必要がある。そのため、まず潜在空間の中から用意した画像に近い画像の潜在変数を推定し、推定した潜在変数によって画像生成を行う。潜在変数の推定は、ランダムな値を入力し、推定したい画像との差が無くなるように繰り返し探索することで行われる。

### 2.4 Style Mixing による顔の向きと表情の変換

StyleGAN2 はニューラルネットワークで 4×4 の画像から 1024×1024 の画像まで少しずつ解像度を上げて画像を生成している。このとき、各解像度の画像生成に潜在変数がスタイル情報として取り込まれるが、解像度ごとによって異なる影響を与えている。StyleGAN2 では解像度ごとに異なる潜在変数を入力することができ、複数の潜在変数を異なる解像度で入力することで複数の画像の特徴を混ぜ合わせた画像を生成できる。具体的には低解像度では主に顔の向きを、中解像度では口や目の形を、高解像度では髪や肌の色をコントロールする。

Completion of Facial Mask Region with StyleGAN2 Using Frontal Face of the Same Person

<sup>†</sup>Hiroaki Koike, Norihiko Kawai, Osaka Institute of Technology

本研究では、StyleGAN2 で生成したマスクを着用していない画像の顔の向きを、同様に生成したマスクを着用している画像の顔の向きに合わせて変更したいため、低解像度である4×4の画像と8×8の画像を生成するときにマスクを着用している画像を生成する潜在変数を入力し、他の解像度のときにはマスクを着用していない画像を生成する潜在変数を入力する。これによって、顔の向きや輪郭の生成にマスクを着用していない画像の特徴が反映される。

また、表情を変更した画像を生成したい場合は、追加で変更先の表情を示す顔画像を用意し、推定した潜在変数を、中解像度である16×16の画像を生成するときに入力する。これにより、変更先の口の特徴が反映され、表情が変化した画像が出力される。

### 3. 実験と考察

#### 3.1 実験

実験1では画像のマスク部分を補完した結果を確認し、マスクを取った実際の顔画像と比較した。図1に入力画像、図2に結果と比較画像を示す。実験2では感情を表現する図3の画像の潜在変数を推定し、図2(a)を生成するネットワークの中解像度で入力した時の効果を検証した。図4に喜楽・驚愕の表情を反映した補完結果を示す。

#### 3.2 考察

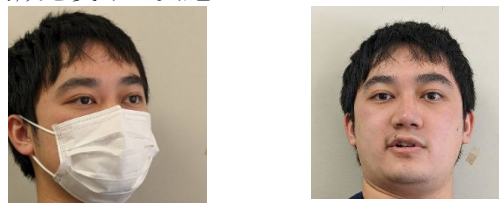
実験1では図2(a)からマスクを補完した箇所と合成先の顔の境界部分が目立たず、口の形や頬のふくらみ、髭といった特徴が反映されているのがわかる。しかし、ほくろといった特徴は実画像を潜在空間に投影する段階で表現されておらず、消失してしまっている。また、顎の長さが実際より長い印象を受ける。これは、マスクを着用している画像を潜在空間上で再現した画像はマスクの膨らみから顎が広くなりやすく、その画像の向きに合わせて style mixing した際に、顔の向きだけでなく顔の輪郭も影響を受けているからである。改善案としてあらかじめ他の画像修復方法でマスク箇所を補完しその画像を用いて style mixing することが挙げられる。

実験2では図3の表情に伴って変化しているのがわかる。しかし主に反映されているのは口の縦横の開き具合のみであり、付与できる表情はそれによって表現できる範囲に限られる。

### 4. おわりに

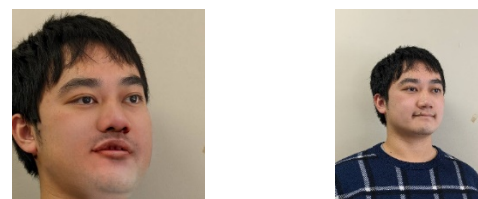
本研究では StyleGAN2 および対象画像本人のマスクをつけていない別の画像を用いたマスク領域の補完を試みた。今後の課題として他の画像修復方法の導入し、マスク画像をあらかじめ補完することが挙げられる。

謝辞 本研究の一部は、JSPS 科研費 (JP18H03273, JP18H04116, JP21H03483) の助成を受けて実施した。



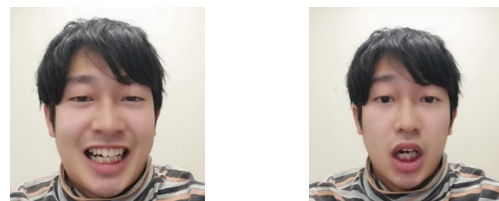
(a) マスク有り (b) 正面

図1 入力画像



(a) 補完結果 (b) 実際の画像

図2 補完結果と実際画像の比較



(a) 喜楽 (b) 驚愕

図3 感情を表す人物の画像



(a) 喜楽 (b) 驚愕

図4 感情を反映した補完結果

### 参考文献

- [1] Nizam Ud Din, Kamran Javed, Seho Bae, Juneho Yi, "A Novel GAN-Based Network for Unmasking of Masked Face", IEEE Access, Vol. 8, pp. 44276-44287, 2020
- [2] Tero Karras, Samuli Laine, Timo Aila, "A Style-Based Generator Architecture for Generative Adversarial Networks", Proc. CVPR, 2019
- [3] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, Timo Aila, "Analyzing and Improving the Image Quality of StyleGAN", Proc. CVPR, 2020
- [4] Patrick Pérez, Michel Gangnet, Andrew Blake, "Poisson Image Editing", ACM Transactions on Graphics (SIGGRAPH'03), Vol. 22, No. 3, pp. 313-318, 2003