

顔認識精度向上のための StyleGAN を用いた疑似訓練画像生成
Pseudo-training image generation
using StyleGAN to improve the accuracy of face recognition

佐藤 立基[†]
Sato Ritsuki

金 義鎮[†]
Kim Euijin

1. はじめに

顔認識において、対象人物の認識精度を高めるには、様々な撮影環境における大量訓練画像が必要である。しかし、対象人物の大量訓練画像を収集するのは容易ではなく、少ない訓練画像から最適パラメータを導き出さなければならない。この問題に対して、1枚の画像から複数の疑似訓練画像を生成する Data Augmentation という考えが知られている。

Data Augmentation では、主に元画像に対して幾何学変化 [1] や属性画像 (眼鏡, 髪型, 顔の向きごとの 3D モデル) の合成処理 [2] を施すことで疑似訓練画像を生成し、その有効性も報告されている。しかし、これらの手法では、顔の上下左右などの 3次元属性合成や色変化 (髪, 肌) が自然に反映された疑似訓練画像の生成までには至っていない。

一方、StyleGAN という高解像度画像生成法が提案されている [3]。StyleGAN では、対象人物の顔画像に別顔の属性 (顔の向き, 眼鏡, 髪色, 肌色) を転移させることで、顔の向きなどの 3次元属性合成や色変化 (髪, 肌) がより自然に反映された画像の生成が可能である。

そこで本研究では、StyleGAN を用いて従来手法では不自然であった 3次元属性合成や色変化をより自然に反映させた疑似訓練画像生成を試み、その結果を報告する。

2. 従来手法

ここでは、従来の幾何学変化や属性画像の合成処理における課題について述べる。

小林らは元画像に対して回転, 水平・垂直移動, 拡大などの 2次元幾何学変化を施す疑似訓練画像生成手法を提案している。しかし、この手法は農作物病原診断を目的としたものであり、顔画像に対しては十分な検証がなされていない。

次に、Jiang-Jing らは顔の輪郭, 目, 鼻の位置を検出し、属性として用意した眼鏡や髪型, 顔の向き の 3D モデルなどの合成処理を施す疑似訓練画像生成手法を提案している。しかし、自然な 3次元属性合成や色変化などに対応していない。

従って、従来手法では自然な 3次元属性合成や色変化に対応できない。この課題点をまとめて以下に示す。

- [課題1] 自然な 3次元属性合成に対応していない。
- [課題2] 色変化 (髪, 肌) に対応していない。

3. 提案手法

上述した課題を解決するために、本研究では StyleGAN の属性転移に着目した新たな疑似訓練画像生成手法を提案する。

[†] 東北学院大学 工学研究科 電気工学専攻 Graduate School of Engineering, Master's Program in Electrical Engineering, Tohoku Gakuin University

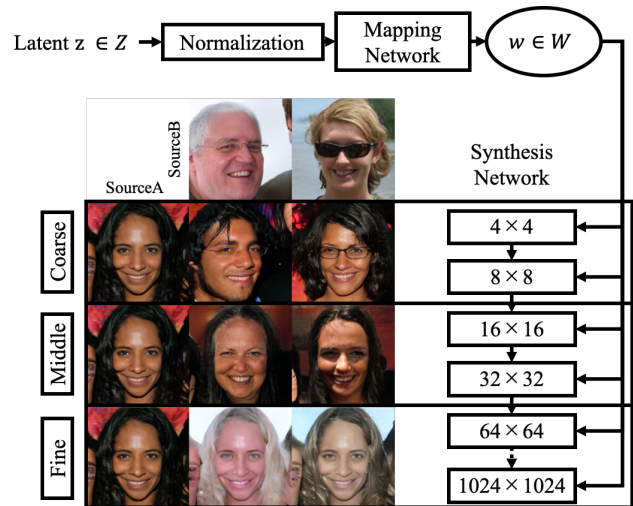


図 1 StyleGAN のネットワークと StyleMixing

StyleGAN の Generator は、Mapping Network と Synthesis Network で構成されている。図 1 に示すように、Mapping network では、入力である潜在変数 z に非線形変換を施すことで得られる新たな潜在変数 w を出力する。Synthesis Network では、この w を属性として入力することでより自然な画像の変形が可能である。その際、Synthesis Network では、SourceA と SourceB の属性を画像生成途中で切り替える StyleMixing 処理が行われる。StyleMixing は、SourceA と SourceB の属性を Coarse style, Middle style, Fine style の 3つに分かれた各層で切り替える処理である。

提案手法では、Middle style を除外し Coarse style (大まかな属性) と Fine style (微小な属性) のみを用いる。その理由として、Middle style では個人顔の不変な属性 (目, 鼻, 口) が大きく変化し、認識精度向上が目的である疑似訓練画像生成が逆効果になると予測されるからである。

また、提案手法では StyleMixing における各 style 内での別画像属性入力層数の変化が、出力画像に影響を与える点に着目した。図 1 に示すように、Style Mixing の Coarse style では自然な 3次元属性合成が行われる。しかし、合成する別画像によっては顔の不変の属性が大きく変化する場合がある。そこで提案手法では 4×4 と 8×8 の Coarse style の間に 6×6 層を追加することで、より正確な 3D 属性合成が可能になると考えられる。

以上から、提案手法の特徴は以下の 3つである。

- ① Coarse style を用いた 3次元属性合成。
 - ② Fine style, を用いた色変化 (髪, 肌)。
 - ③ 新たな層の追加による、より正確な 3次元属性合成。
- 従って、①と③を用いることで [課題 1] を、②を用いることで [課題 2] を解決できると考えられる。

実験 4.では Coarse style を用いた 3 次元属性合成と Fine style を用いた色変化を行い、それぞれ疑似訓練画像生成を行う。

4. 実験

提案手法の有効性を示すために、Coarse style を用いた 3 次元属性合成と Fine style を用いた色変化（髪，肌）を行い、それぞれ疑似訓練画像の生成した。

実験で用いた StyleGAN におけるネットワークは、Discriminator と Synthesis Network に 6×6 層を追加した。学習において、データセットとして CelebA データセットを用いた[4]。本研究では、データセットの画像の中心 178×178 で切り取り、 128×128 に縮小して利用した。学習では、10177 名の画像をランダムで各 1 枚ずつ取り出した 10177 枚と、眼鏡の属性を持った画像 5000 枚、4 種の髪色の属性を持った画像各 3000 枚ずつの計 27177 枚を用いた。この学習したネットワークを用い、Middle style を除いた Coarse style と Fine style を用い StyleMixing を行った。

また、本研究では高解像度生成が目的ではないので、1000000 枚の画像を用いてパラメータを更新し、疑似訓練画像の生成を行った。

まず、顔の向きの 3 次元属性合成では、男女 1 名ずつの画像に眼鏡属性を持った画像 4 枚の合成を行った。この StyleMixing では Coarse style の 4×4 、 $4 \times 4 \sim 6 \times 6$ 、 $4 \times 4 \sim 6 \times 6 \sim 8 \times 8$ の 3 つの範囲で合成したい画像の属性に切り替えた。この結果を図 2 に示す。

次に、色変化（髪，肌）では、男女 1 名ずつの画像に金髪，黒髪，茶髪，白髪の各髪色属性を転移した。この StyleMixing では Fine style の $64 \times 64 \sim 128 \times 128$ で合成したい画像の属性に切り替えた。この結果を図 3 に示す。

5. 考察

実験では、StyleGAN を用いた顔の向きの 3 次元属性合成や色変化（髪，肌）を施した疑似訓練画像の生成を行った。

図 2 に示すように、図 2(a)と図 2(b)を比較すると顔の向きが大きく変化しており、 6×6 層を追加したことで多くの属性が取り入れられている。図 2(b)と図 2(c)を比較すると、さらに大きな顔の向きの変化や眼鏡の合成が自然にできた。

また、図 3 に示すように、元画像の特徴を崩すことなく、色変化（髪，肌）を反映した疑似訓練画像の生成を行った。これにより、従来手法の課題点を解決できていると考えられる。

しかしながら、本研究では以下に示す 3 つの課題が残されている。

- ① 提案手法において潜在空間から任意の画像の潜在変数を取り出せないため、希望の画像に StyleMixing を行うことができていない。
- ② 3 次元属性合成と色変化（髪，肌）同時に行うことができていない。
- ③ 従来手法との認識精度の比較を行えていないため、実際に顔認識において提案手法が有効であるか証明できていない。

6. まとめ

本研究では StyleGAN を用いた新たな疑似訓練画像生成手法を提案した。実験結果より、提案手法では 3D 属性合成や色変化（髪，肌）自然な反映が可能であることがわか



(a) 4×4 の Coarse style



(b) $4 \times 4 \sim 6 \times 6$ の Coarse style



(c) $4 \times 4 \sim 6 \times 6 \sim 8 \times 8$ の Coarse style

図 2 3 次元属性合成を施した疑似訓練画像の一部

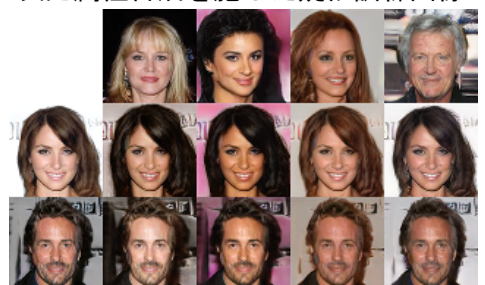


図 3 色変化を施した疑似訓練画像の一部

った。これは、従来手法において課題として残されていた点であり、提案手法の有効性が示されている。今後は、5 章で述べた課題点を解決していきたいと考えている。

参考文献

- [1] 小林 賢一, 辻 順平, 能登 正人, “ディープラーニングを用いた画像処理による農作物病害診断への Data Augmentation の応用”, 情報処理学会第 79 回全国大会講演論文集, pp.289-290 (2017).
- [2] Jiang-Jing Lv, Xiao-Hu Shao, Jia-Shui Huang, Xiang-Dong Zhou, Xi Zhou, “Data Augmentation for face recognition”, Neurocomputing, vol.230, Issue C, pp.184-196 (2017).
- [3] Tero Karras, Samuli Laine, Timo Aila, “A Style-Based Generator Architecture for Generative Adversarial Networks”, IEEE Conference on Computer Vision and Pattern Recognition, pp.4401-4410 (2019).
- [4] Ziwei Liu, Ping Luo, Xiaogang Wang, Xiaoou Tang, “Deep Learning Face Attributes in the Wild”, IEEE International Conference on Computer Vision, pp.3730-3738 (2015).