

潜在空間解析による顔画像年齢操作の多様化と精度向上

伊藤 泰志 遠藤 結城 金森 由博
筑波大学

1 はじめに

人物の顔画像の年齢編集は、人物の同一性を維持しつつ年齢を変化させるタスクであり、科学捜査や映像制作分野での活用が期待されている。年齢変化による人物の見た目は、個人によって一意に決まるものではなく、その人の境遇や環境などにより皺や髪の量、髪色などに多様な変化が予想される。最先端手法の SAM [1] は、深層学習を用いてターゲット年齢への高品質な編集が可能だが、年齢や人物同一性を保ったまま、結果に多様性を持たせることは難しい。また、訓練データの偏りの問題から、低年齢の子供への編集精度は十分とはいえない。

そこで本研究では、顔画像の年齢編集において、結果に多様性を持たせるとともに編集精度を向上させることを目指す。提案手法では、出力多様化の際に、StyleGAN [2] の潜在空間の一つであり良く disentangle されている StyleSpace [3] (S 空間) を利用する。 S 空間を年齢と人物同一性について分析し、相関係数に基づいて髪の色や量、皺の量などの年齢依存属性を多様化する。その後、潜在変数の最適化によって多様化の際にターゲット年齢と人物同一性を保持する。さらに、ターゲット年齢周辺の推定情報から導かれる潜在空間内の経路を外挿することで、低年齢への編集精度を向上させる教師なしのアプローチを検討する。

2 ベースライン手法

本研究のベースとなる SAM [1] について説明する。SAM は、人物の顔画像とターゲット年齢を入力とし、入力を低次元の潜在空間 $\mathcal{W}+$ 内の潜在変数に符号化するエンコーダと、潜在変数から画像を出力するデコーダによって年齢編集画像を生成する。事前学習済み年齢分類器によって得られる出力画像の推定年齢とターゲット年齢が一致するようエンコーダを学習させることで、ターゲット年齢を直接指定可能な教師なし学習を実現している。しかし、SAM は基本的にターゲット年齢に対して単一の結果を得ることしか想定していない。デコーダである StyleGAN [2] に備わっている Style Mixing を用いれば出力の多様化も可能だが、ターゲット年齢や人物の同一性の維

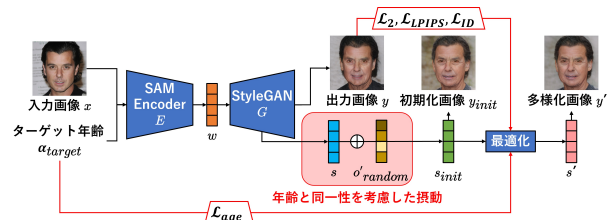


図 1: 本手法における多様化の流れ。

持は考慮されていない。また、編集精度は年齢分類器の精度に依存するため、年齢分類器の精度が低い低年齢に対する編集精度にも限界がある。

3 提案手法

3.1 多様性を考慮した年齢編集

本手法の流れを図 1 に示す。本手法では、人物画像 x と所望のターゲット年齢 α_{target} を入力とし、事前学習済みの SAM エンコーダ E から年齢編集済み潜在変数 $w \in \mathcal{W}+$ を出力する。 w は事前学習済み StyleGAN 生成器 G に入力され出力画像 y となる。ここで w は G の各レイヤでのアフィン変換によって $s \in S$ へ変換される。

本研究では S 空間における操作が良く disentangle されていることに着目し、この操作を多様化に活用する。 S 空間では、個別のチャンネルを操作することで、口や耳の形など、細かい属性を編集できることが知られている [3]。そこで提案手法では S 空間の分析を通して、年齢の変化とともに変動する属性(皺や髪など)に影響を与えるチャンネルを特定・操作することで年齢依存の多様化を図る。この際、なるべく人物同一性には影響を与えないチャンネルを対象とする。

具体的には、 S 空間の各チャンネルについて以下の操作を行う。ランダムサンプリングされた 100 個の S 空間潜在変数について、任意のチャンネルに微小な摂動を与える。摂動前後の潜在変数から生成される画像の推定年齢および人物同一性の変化量を記録し、摂動量との相関係数を計算する。これをもとに、年齢と高相関かつ人物同一性と低相関なチャンネルほど大きな重みを与えるマスク o_{mask} を作成する。ランダムなオフセット $o_{random} \in S$ に o_{mask} を適用した o'_{random} を s に加えることで、人物同一性を保持しつつ年齢依存属性を多様化する。

とはいえ、この時点では年齢や人物同一性が α_{target} や x から多少変化することが想定される。そのため、多様化した潜在変数を初期値 s_{init} としてさらに最適化する。最適化に用いる損失関数に

表 1: 各手法に関する出力多様化の定量的評価.

手法	LPIPS \uparrow	ID \downarrow	AGE_SD \downarrow
SAM	0.0495	0.253	4.20
提案手法	0.0549	0.0181	3.00

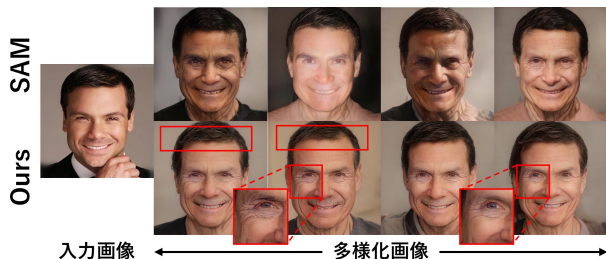


図 2: 多様化に関する定性的評価.

は, SAM の学習時と同じものを用いる. 以上により, ターゲット年齢と人物同一性を保持した多様な潜在変数 s' と多様化画像 y' が得られる.

3.2 低年齢への年齢編集精度向上

SAM の低年齢への編集では人物に十分な年齢変化が現れないが, これは SAM が該当の年齢方向へ潜在変数を移動させきれていないためであると考えられる. そこで, SAM の出力潜在変数を外挿することでこの問題に取り組む. 具体的には, SAM を用いて, ターゲット年齢から年齢中央方向へ一定間隔でいくつかのサンプル潜在変数を作成する. 実験により, 潜在変数の年齢間隔は 3 歳, サンプル点は 4 個とした. 例として, ターゲット年齢が 5 歳の場合, 5, 8, 11, 14 歳に対応する潜在変数を使用する. これを線形外挿することで潜在空間内の経路を作成する. 潜在変数から作成される画像の推定年齢がターゲット年齢となるまで, この経路上を反復的に探索する. 推定年齢の算出には, SAM の学習に用いられた事前学習済み年齢分類器を使用する. これにより, よりターゲット年齢に近い潜在変数が得られる.

4 実験

SAM の Style Mixing と提案手法による多様化について定量的評価を行った結果を表 1 に示す. CelebA-HQ [4] のテスト画像のうち 100 枚について, それぞれ 10 枚の多様な出力画像を作成し評価した. 評価の際, 背景は除去している. LPIPS は多様性を, ID は出力間の人物同一性を, AGE_SD は出力間の推定年齢の標準偏差を表す. 表の太字は最良の値を示す. 提案手法はすべての指標において SAM を上回る数値を示した.

また, 多様化の定性的評価を図 2 に示す. SAM の Style Mixing は肌や照明など年齢変化と関係のない属性が変化している様子が見られ, 一部画像は外見上の年齢がターゲット年齢 (80 歳) と離れている. 提案手法はターゲット年齢と人物同一性を保持しつつ, 皺や髪の色, 顔概形など年齢依存属性のみを多様化

表 2: 各手法に関する低年齢への編集の定量的評価.

α_{target}	手法	FID \downarrow	AGE_MAE \downarrow
5	SAM	149.1	17.7
	提案手法	109.3	7.75
8	SAM	119.0	15.5
	提案手法	113.0	13.9
12	SAM	110.4	12.6
	提案手法	109.3	11.6



図 3: 低年齢 (5 歳) への編集に関する定性的評価.

できている (赤四角参照).

低年齢への編集についても定量的評価を行った. 結果を表 2 に示す. CelebA-HQ のテスト画像 2,000 枚を入力とし, ターゲット年齢 α_{target} は 5, 8, 12 歳とした. AGE_MAE は各出力画像の推定年齢とターゲット年齢間の差の絶対値の平均, FID [5] は画像品質を表す. FID の評価には, FFHQ-Aging [6] データセットの 3-6 歳, 7-9 歳, 10-14 歳の年齢グループに属する画像を使用した. 提案手法はどのターゲット年齢においても SAM を上回る性能を示した. ターゲット年齢が低年齢になるほど改善率が上昇している.

低年齢への編集に関する定性的評価を図 3 に示す. 提案手法は SAM と比較して定性的にもターゲット年齢に近い編集ができているといえる.

5 まとめ

本研究では, S 空間解析と最適化による, ターゲット年齢と人物同一性の保持を考慮した既存年齢編集手法出力の多様化手法を提案した. また, 既存手法の年齢編集精度を潜在空間における外挿によって改善する手法を提案した.

参考文献

- [1] Alaluf *et al.* Only a Matter of Style: Age Transformation Using a Style-Based Regression Model. *ACM Trans. on Graph.*, Vol. 40, No. 4, pp. 1–12, 2021.
- [2] Karras *et al.* Analyzing and Improving the Image Quality of StyleGAN. In *CVPR*, pp. 8110–8119, 2020.
- [3] Wu *et al.* StyleSpace Analysis: Disentangled Controls for StyleGAN Image Generation. In *CVPR*, pp. 12863–12872, 2021.
- [4] Liu *et al.* Deep learning face attributes in the wild. In *ICCV*, pp. 3730–3738, 2015.
- [5] Heusel *et al.* Gans trained by a two time-scale update rule converge to a local nash equilibrium. pp. 6626–6637, 2017.
- [6] Or-Eli *et al.* Lifespan Age Transformation Synthesis. In *ECCV*, pp. 739–755, 2020.