令和 2 年度　特別研究報告書

PULSE を用いた
顔画像置換の有効性の検討

龍谷大学　理工学部　情報メディア学科

T170550　　和澤大介

指導教員　三好力　教授

内容概要

本論文では、google ストリートビュー上に映ったぼかしのかかった人物の顔を別の顔画像に置き換えるシステムの提案を行う。本研究では、Photo Upsampling via Latent Space Exploration（PULSE）という低解像度の画像を高解像度に変換する技術を用いて、低解像度の人の顔画像を高解像度に変換する。また、PULSE は学習済み Style GAN を用いており、元の人物画像 と変換後の人物画像の差異を検証し、プライバシーの観点でその有効性を検証する。
実験では男性と女性の画像から PULSE を用いて一度低解像度にリサイズしてから、学習済み Style GAN で高解像度の顔画像を生成し、PULSE を用いて生成した顔画像と元画像とを比較し、検討を行う。比較には、平均オピニオン評点(MOS)を用いる。また、Mac のプレビューを用いてトリミングし、低解像度にリサイズして Style GAN で高解像度画像を生成する。この方法で生成した画像と、PULSE を用いて生成した画像とを比較し PULSE の有効性を検証する。

目次

第 1 章 研究の背景と目的

2020 年は新型コロナウイルスの感染拡大に伴って、人々は外出を控える日々が続いた。そこで Google ストリートビューを用いたバーチャルトラベルが話題になった。Google ストリートビューとは、何百もの都市の 3D の街並みをブラウザやアプリから探索できるものである。家にいながら旅行している気分を味わえるため話題となった。

私はこの Google ストリートビューに映る人の顔に、ぼかしがかかっていることが、現実感や没入感を欠いてしまっているのではないかと感じた。VR 技術の発展も伴って、ユーザはよりリアルさを求めている。

本研究では、PULSE という低解像度の画像を高解像度に変換する技術を用いて、低解像度の人の顔画像を高解像度に変換する。また、PULSE は学習済み Style GAN を用いており、元の人物画像と変換後の人物画像の差異を検証し、プライバシーの観点でその有効性を検証する。

第 2 章 既存技術

2.1 Generative Adversarial Networks (GAN)

GAN は互いに敵対し合う生成器と判別器の 2 つのネットワークから成り立つ。生成器は Generator、判別器は Discriminator と呼ばれ、Generator は Discriminator を欺くようにデータを生成し、Discriminator は生成されたデータが生成されたデータであると暴くことができるように学習を行う。GAN の構造をまとめたものを図 2.1 に示す。

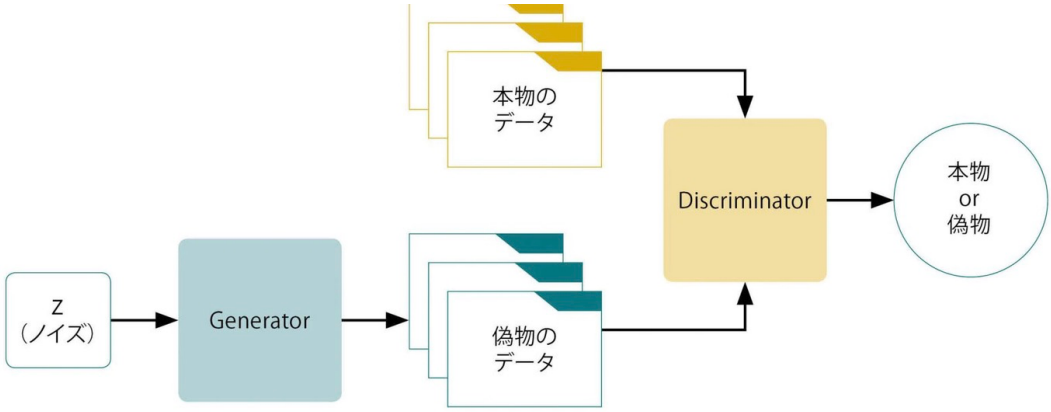


図 2.1 GAN の構造（[1]から引用）

Generator は教師データの特徴に沿ってデータを生成するように学習を行う。様々なパターンの画像を生成できるように、入力には乱数によって生成された値の配列を用いる。画像は転置畳み込み処理によって生成される。転置畳み込み処理は畳み込み処理の逆で、1 ピクセル毎に kernel との積を求め、stride 毎の和を取り、画像をアップサンプリングする処理である。

Discriminator は入力画像に対して、生成された画像なのか教師データなのかを判別する。どれだけ教師データに近いかを表す数値を 0 から 1 の範囲(1 が教師データ、0 が生成された画像)で出力する。

損失関数について以下のように用語を定義する。

・G : Generator
・D : Discriminator
・z: ノイズベクトル
・G(z) : Generator が生成したデータ

とすると、Generator の損失関数を式 2.1 に示す。

$$L_G = \frac{1}{M}\sum_{i=1}^{M}\log\Big(1 - D\big(G(z)\big)\Big) \quad (2.1)$$

Generator が生成したデータを Discriminator に入力し、教師データと判定されると最小になる。

Discriminator の損失関数を式 2.2.に示す。

$$L_D = \frac{1}{M}\sum_{i=1}^{M}[\log D(x) + log\left(1 - D\left(G(z)\right)\right)] \quad (2.2)$$

Discriminator が教師データを教師データ(出力 1)、生成画像を生成画像(出力 0)と判定すると最小になる。[1]

2.2 Style Generative Adversarial Networks (Style GAN)

Style GAN が GAN と異なる点は大きく 3 つ存在する。1 つ目は各転置畳み込み処理後に Style の調整を行う。2 つ目は細部の特徴はノイズによって生成されるということ。3 つ目は潜在変数 z を潜在空間 w に非線形変換する。この変更により高精度な画像の生成を可能としている。ジェネレーターのネットワーク図を図 2.2 に示す。

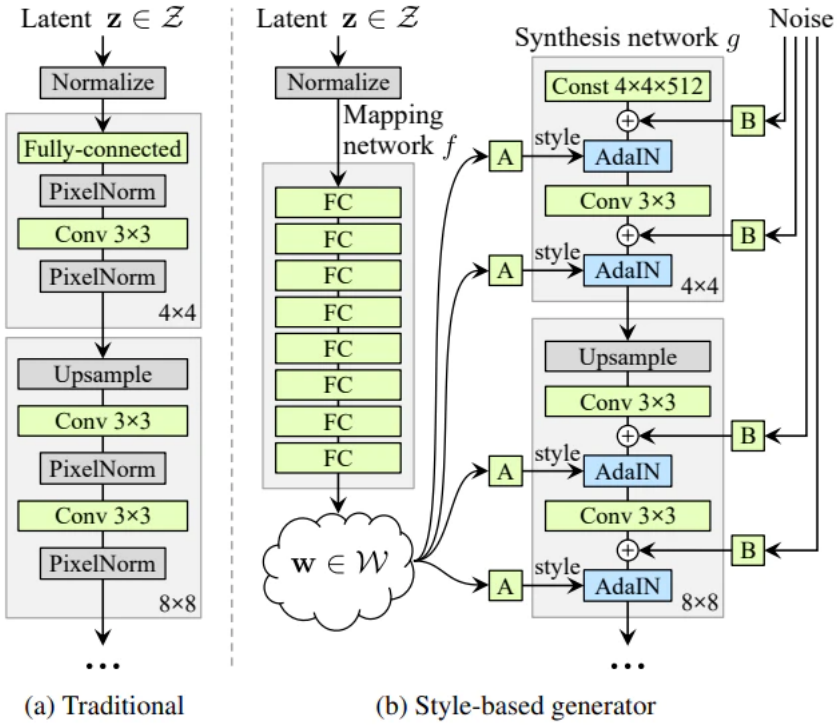


図 2.2 ジェネレーターのネットワーク図（[2]から引用）

左の図がこれまでの GAN(PG-GAN)、右の図が Style GAN である。

Style GAN は Mapping network と Synthesis network の 2 つのネットワークで構

成されている。GAN では潜在変数 z から直接画像を生成していたのに対し、Style GAN では 4×4×512 の固定のテンソルから画像を生成している。

Mapping network では 8 層の全結合層によって潜在変数 z を潜在空間 w に非線形変換している。これは入力時には情報的な意味を待たないただの乱数の数列(潜在変数 z)を多次元的なスタイル情報(年齢、性別、表情など)を表す空間に数値をマッピングし、スタイルの特徴を表すものに変換するという処理である。

Synthesis network は画像のアップサンプリング、スタイル及ノイズ情報のマージを行う。アップサンプリングを何度にも分け、その都度スタイル情報を挿入することには、画像のサイズによって持つ情報が異なるため大きな意味を持つ。7680×4320(8K)の画像では画像の特徴は非常に分散しており 1 ピクセルから得られる情報は極めて少なく、細かい。対して、1×1 の画像が表す 1 ピクセルの情報は画像最大の特徴を表し、一番アバウトな情報である。
つまり、画像サイズが大きくなるに従って、物体の色、物体の位置、物体の輪郭、物体の細かな模様と順番にアバウトな情報から細かい情報を表すものへと変化していく。よって、小さいサイズの画像にスタイル情報を挿入すれば全体の色味等が操作でき、大木サイズの画像にスタイル情報を挿入すれば表情や服の模様などを操作することができる。

AdalN は、Mapping network から得られた潜在空間 w を各解像度ごとにスタイル情報としてマージする部分である。マージの計算式を式 2.3 に示す。

$$AdalN(x_i, y) = y_{s,i}\frac{x_i - \mu(x_i)}{\sigma(x_i)} + y_{b,i} \quad (2.3)$$

Xi は特徴マップ、Ys,i、Yb,i は潜在変数 w をアフィン変換したものである。[2]

2.3 Photo Upsampling via Latent Space Exploration (PULSE)

これまでの手法では、生成画像と HR(High Resolution)画像間のピクセル距離の MSE(平均二乗誤差)を最小化するように訓練していた。しかしこの方法で生成した画像は、どうしてもぼやけてしまう。PULSE はこの MSE によるアプローチの代わりに自然な画像の manifold 上で、正しくダウンスケーリングされるような点を見つけ出すことを目標としている。manifold とは、ある画像に対し、回転・移動などの操作を加えた増幅画像群の特徴量の集合のことである。つまり manifold 上の点は元画像と同じ性質を持っている。PULSE は入力された LR(Low Resolution)画像に対し、事前訓練済み生成器の潜在空間によってパラメータ化された manifold に沿って、正しくダウンスケーリングできる領域を探し出す。

以下のように用語を定義する。

・ILR : 低解像度の入力画像（次元は R^(m×n)）

・ISR : 生成される超解像画像（次元は R^(M×N)）
・M : R^(M×N)である自然画像の manifold
・DS : ダウンスケーリング関数
・R : 正しくダウンスケーリングされる画像の集合
とすると、式 2.4 で表せる。

$$R = \{I \in \mathbb{R}^{N*M} : DS(I) = I_{LR}\} \quad (2.4)$$

従来の手法における目標は、ILR を用いて HR 画像を復元することだった。||HR 画像 − ISR||のノルムを最小化する最適化問題を解くというアプローチであった。しかしこのアプローチでは、最適な ISR は HR 画像の画素ごとの加重平均となり、詳細が欠如されるためうまくいかない。

そこで PULSE では ILR に対して、ISR∈Mを見つけることを目標とした。これは上記で定義した画像の集合 R を見つけることである。ダウンスケーリングすると ILR となる集合 I を見つける式を 2.5 に示す。

$$R_\epsilon = \{I \in \mathbb{R}^{N*M} : ||DS(I) - I_{LR}||_p^p \leq \epsilon\} \quad (2.5)$$

損失関数について従来の手法では、ISR が HR 画像にどれだけ近いかを計算していた。ISR と HR 画像の lp ノルムを最小化させると ISR は manifold に近づくが、M に近いという保証はない。PULSE では M 中のある ISR がどれだけ ILR に対応しているかを探しているため、M にどれだけ近いかという議論を回避している。ISR のダウンスケーリング画像が ILR から逸脱した場合にダウンスケーリング損失としてペナルティを与えることにした式を 2.6 に示す。

$$L_{DS}(I_{SR}, I_{LR}) \coloneqq \left|\left|DS(I_{SR}) - I_{LR}\right|\right|_p^p \quad (2.6)$$

manifold 上の探索について仮に、manifold の各パラメータが微分可能であれば、ダウンスケーリング損失の勾配を利用しながら、manifold に沿って ISR の検索をすることが可能となる。ここで得られる ISR は manifold 上の画像なので高解像度であることが保証され、かつダウンスケーリング損失によって正確であることも保証される。現実にはこのような manifold のパラメータ化は存在しないが、教師なし学習の手法を利用することで、パラメータ化を近似することができる。VAE や GAN は、ある潜在空間から関心領域の manifold へのマッピングを意識した作りとなっているため、これらの事前学習済みモデルを利用することは有効である。VAE や GAN などの生成器を G、潜在空間を L とすると、G による生成画像で M を近似し、以下の式 2.7 を満たす潜在ベクトル z∈L を見つけるタスクだと言い換えられる。

$$\left|\left|DS(G(z)) - I_{LR}\right|\right|_p^p \leq \epsilon \quad (2.7)$$

しかし、ほとんどの生成モデルでは z∈L であることだけで $G(z)$∈M を保証することはできない。 $G(z)$∈M を保証するためには、事前に選ばれた状態が高い確

率で L の領域にいる必要がある。具体的には事前に負の対数尤度の損失項を追加することである。事前分布がガウス分布に従う場合、これは L2 正規化の形をとる。しかしながら他変量ガウス分布の次元を上げると、分布はほとんどが超球面上に分布することが分かっている。つまり L2 正規化によって潜在ベクトルが 0 に向かうのは好ましくない。PULSE はこの点を工夫することで、潜在空間全体の勾配降下から超球面上への統計勾配降下へと置き換えている。[3]

第 3 章 提案手法

Google ストリートビューで、人の顔にぼかしが入っているのを取り除き、プライバシーの観点から本人とは別の人の顔画像を生成し、置換することを目的とする。

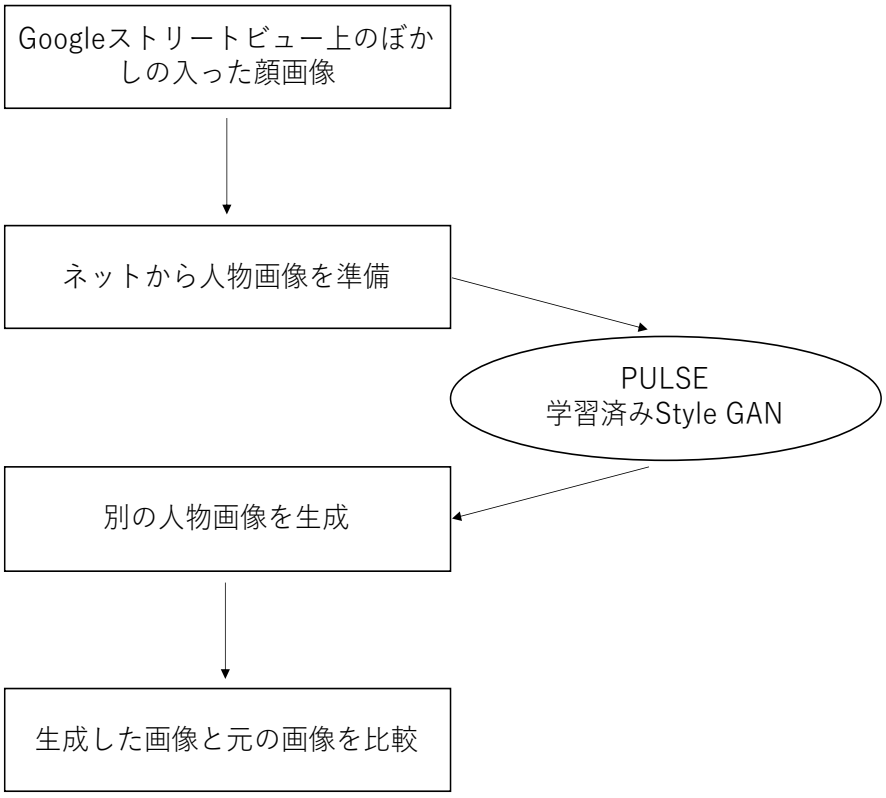
本研究では、PULSE という低解像度の画像を高解像度に変換する技術を用いて、低解像度の人の顔画像を高解像度に変換する。 ぼかしの入った顔画像から、学習済みの Style GAN を用いて別の人物の顔画像を生成する。この手法で、ぼかしの入った顔画像から別の人物の顔画像を生成する。

次に生成した画像と元の画像とを比較し、プライバシーの観点からも、Google ストリートビューでの有効性を検証する。

第 4 章 実験

4.1 実験目的

本実験では男性、女性、反転した女性の 3 種類の画像から PULSE を用いて一度低解像度に変更し、学習済み Style GAN を用いて高解像度画像を生成する。さらに、元の画像と生成した画像の比較は主観評価法を用いて行い、その結果から有効性を検証することを目的とする。



4.2 実験 1

女性画像と反転した女性画像、男性画像の 3 枚の画像を、PULSE を用いて 8 x 8 や 16 x 16、32 x 32、64 x 64 に低解像度に変更し、学習済み Style GAN を用いて 1024 x 1024 の画像を生成した。

4.2.1 実験結果 1

図 4.1 の女性の画像を顔の部分に合わせて切り取った結果を図 4.2 に、PULSE を用いて 8 x 8 や 16 x 16、32 x 32、64 x 64 に低解像度に変更した画像をそれぞれ図 4.3、4.4、4.5、4.6 に示す。

図 4.1 元の女性画像

図 4.2　顔の部分に切り取った女性画像

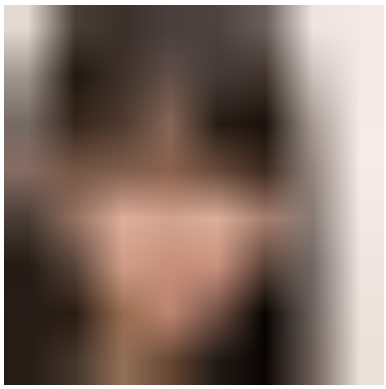


図 4.3　8 x 8 の女性画像

図 4.4　16 x 16 の女性画像

図 4.5　32 x 32 の女性画像

図 4.6　64 x 64 の女性画像

図 4.3 から生成した女性画像を図 4.7、図 4.4 から生成した女性画像を図 4.8、図 4.5 から生成した女性画像を図 4.9、図 4.6 から生成した女性画像を図 4.10 に示す。

図 4.7　8 x 8 から生成した女性画像

図 4.8　16 x 16 から生成した女性画像

図 4.9　32 x 32 から生成した女性画像

図 4.10　64 x 64 から生成した女性画像

図 4.11 の反転した女性画像を顔の部分合わせて切り取った結果を図 4.12 に、PULSE を用いて 8 x 8 や 16 x 16、32 x 32、64 x 64 に低解像度に変更した画像をそれぞれ図 4.13、4.14、4.15、4.16 に示す。

図 4.11 元の反転女性画像

図 4.12　顔の部分に切り取った反転女性画像

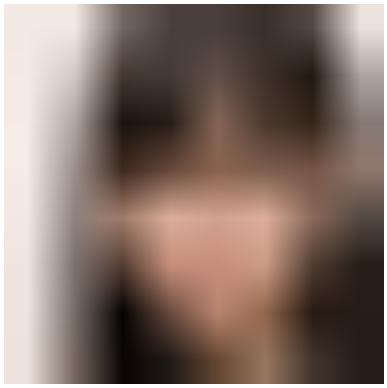


図 4.13　8 x 8 の反転女性画像

図 4.14　16 x 16 の反転女性画像

図 4.15　32 x 32 の反転女性画像

図 4.16　64 x 64 の反転女性画像

図 4.13 から生成した女性画像を図 4.17、図 4.14 から生成した女性画像を図 4.18、図 4.15 から生成した女性画像を図 4.19、図 4.16 から生成した女性画像を図 4.20 に示す。

図 4.17　8 x 8 から生成した反転女性画像

図 4.18　16 x 16 から生成した反転女性画像

図 4.19　32 x 32 から生成した反転女性画像

図 4.20　64 x 64 から生成した反転女性画像

図 4.21 の男性の画像を顔の部分に合わせて切り取った結果を図 4.22 に、PULSE を用いて 8 x 8 や 16 x 16、32 x 32、64 x 64 に低解像度に変更した画像をそれぞれ図 4.23、4.24、4.25、4.26 に示す。

図 4.21 元の男性画像

図 4.22　顔の部分に切り取った男性画像

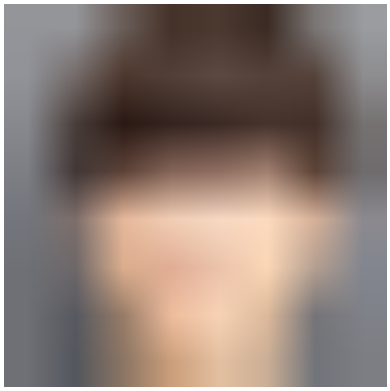


図 4.23　8 x 8 の男性画像

図 4.24　16 x 16 の男性画像

図 4.25　32 x 32 の男性画像

図 4.26　64 x 64 の男性画像

図 4.23 から生成した男性画像を図 4.27、図 4.24 から生成した男性画像を図 4.28、図 4.25 から生成した男性画像を図 4.29、図 4.26 から生成した男性画像を図 4.30 に示す。

図 4.27　8 x 8 から生成した男性画像

図 4.28　16 x 16 から生成した男性画像

図 4.29　32 x 32 から生成した男性画像

図 4.30　64 x 64 から生成した男性画像

4.3 実験 2

実験 1 で使用した図 4.1 の女性画像を Mac のプレビューを用いて顔の部分を切り取り、8 x 8 や 16 x 16、32 x 32、64 x 64 にリサイズし、学習済み Style GAN で 1024 x 1024 の画像を生成した。

4.3.1 実験結果 2

図 4.1 の女性画像を Mac のプレビューを用いて顔の部分を切り取った画像を図 4.31 に示す。図 4.31 を 8 x 8 や 16 x 16、32 x 32、64 x 64 に低解像度に変更し、学習済み Style GAN を用いて生成した画像をそれぞれ図 4.32、4.33、4.34、4.35 に示す。

図 4.31　プレビューを用いて切り取った女性画像

図 4.32　プレビューを用いて 8 x 8 から生成した女性画像

図 4.33　プレビューを用いて 16 x 16 から生成した女性画像

図 4.34　プレビューを用いて 32 x 32 から生成した女性画像

図 4.35　プレビューを用いて 64 x 64 から生成した女性画像

4.4 実験 3

大学生で 22 歳の男性 10 人を対象に主観評価法の平均オピニオン評点(MOS)用いて評価実験を行った。実験は Microsoft Teams の画面共有を用いてリモートで行った。質問は全部で 16 問用意し、はじめの 12 問は PULSE を用いて生成した画像と元の画像とを比較し、似ているかどうかを 1 から 5 の数字で問う内容となっている。1 から 5 の数字の内容は 1(非常に似ていない)、2(似ていない)、3(普通)、4(似ている)、5(非常に似ている)とする。
残りの 4 問は、Mac のプレビューを用いて生成した画像と PULSE を用いて生成した画像とを比較し、きれいかどうかを 1 から 5 の数字で問う内容となっている。今回の 1 から 5 の数字の内容は 1(きれいでない)、2(ややきれいでない)、3(普通)、4(ややきれい)、5(きれい)とする。

順序効果が作用するため画像提示の順番はランダムに行った。また、リモートで複数人同時に実験を行ったため、他の人の解答がわからないように解答方法は個別に LINE で送ってもらった。

また、全質問終了後に解答者に一言コメントをいただいた。

作成した質問のスライドを図 4.33 に示す。

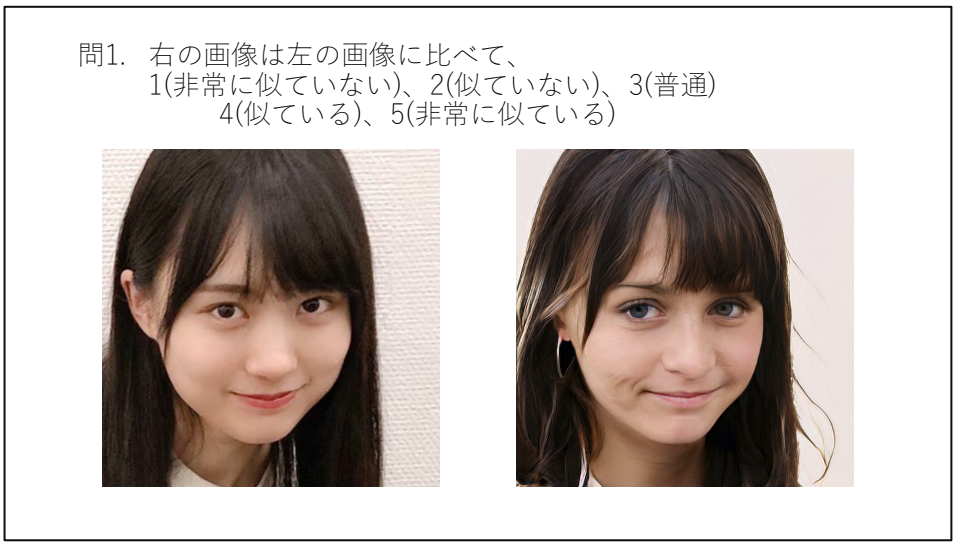


図 4.33　作成した質問のスライド例

4.4.1 実験結果 3

元の画像と生成した画像とが似ているかどうか比較した結果を表 4.1 に示す。解答は 1 から 5 の数字でそれぞれ平均を求めた。1 に近いほど似ておらず、5 に近いほど似ている。

Mac のプレビューを用いてリサイズし生成した画像と PULSE を用いて生成した画像との比較結果を表 4.2 に示す。1 に近いほどきれいでなく、5 に近いほどきれいである。

表 4.1　元の画像と生成した画像を比較した MOS による結果

	8x8	16x16	32x32	64x64
女性の画像	1.6	1.7	3.3	3.8
男性の画像	1.5	1.7	2.8	3
女性の反転画像	1.6	1.9	2.2	3.6

表 4.2　プレビューを用いて生成した画像と PULSE で生成した画像との比較結果

	8x8	16x16	32x32	64x64
Mac のプレビュー	1	1.1	1	1

表 4.1 より、64 x 64 にリサイズして生成した画像が、一番元の画像に似ているという結果となった。一方で 8 x 8 にリサイズして生成した画像が、一番元の画像に似ていないという結果となった。女性の画像と男性の画像、女性の反転画像を 64 x 64 にリサイズして生成した画像は全て 3 を超えており、女性の画像を 32 x 32 にリサイズして生成した画像も 3 を超える結果となった。

男性の画像は 8 x 8 のサイズで 1.5 の数値となり、64 x 64 で 3 の数値となった。その差は 1.5 とあまりサイズ間での差が出なかった。

女性の画像は 8 x 8 のサイズで 1.6 の数値となり、64 x 64 のサイズで 3.8 の数値となった。その差は 2.2 と男性の画像と比べてサイズ間での差がでた。

女性の反転画像は 8 x 8 のサイズで 1.6 の数値となり、64 x 64 のサイズで 3.6 の数値となった。その差は 2.0 とこちらも男性の画像と比べてサイズ間での差がでた。

全ての画像の 8 x 8 と 16 x 16 のサイズから生成した画像はどれも 2.0 を超えておらず、元の画像とあまり似ていないという結果となった。

表 4.2 より Mac のプレビューを用いてリサイズし生成した画像は、どのサイズも PULSE を用いて生成した画像に比べてきれいでないという結果になった。

解答者のコメントについて以下に記載しておく。

・質問内容をより具体的にすればよかったのではないか。
・似ている、似ていない意外にも質問を用意すればよかったのでは。
・似ている、似ていないの判断基準が人によって異なるのではないか。
・顔のどの部分を注視するかによっても似ている、似ていないが変わってくるのではないか

第 5 章 考察

今回、男性と女性、反転した女性の 3 種類の画像を PULSE を用いてリサイズし、高解像度画像を生成し、比較することを目的として実験を行った。結果は、64 x 64 にリサイズして生成した画像が一番元の画像に似ており、8 x 8 にリサイズして生成した画像が一番元の画像に似ていないという結果になった。PULSE で一度低解像度にリサイズする際に、より解像度を高くするほど元の画像に近い画像が生成できることが分かった。また、PULSE を用いず、Mac のプレビューでトリミングし、低解像度にリサイズしてから生成した画像はどのサイズもきれいな画像は生成できなかった。このことから、PULSE を用いて生成した画像の方が、google ストリートビューに映った人物の顔を置き換えることに有効だと考える。

また、表 4.1 の実験結果から 8 x 8 や 16 x 16 にリサイズして生成した画像は元の画像とあまり似ておらず、プライバシーの観点から、低解像度であるほど有効だと考える。

謝辞

本研究を進めるにあたって、多忙の中ご指導いただきました三好力教授に深く感謝いたします。また、研究室の皆様には多くの助言をいただきましたことを感謝いたします。

参考文献

[1] Style GAN とはなにか GAN
https://qiita.com/YasutomoNakajima/items/1e0153cfb598641f5c9b
,2020-10-27
[2] Style GAN とはなにか Style GAN
https://qiita.com/YasutomoNakajima/items/1e0153cfb598641f5c9b
,2020-10-27
[3]【PULSE】最大 64 倍の超解像アルゴリズム解説
https://medium.com/lsc-psd/pulse-最大 64 倍の超解像アルゴリズム解説-2e61938cf52b
,2020-10-26
[4] PULSE で低解像度の顔画像を高解像度に変換する
http://cedro3.com/ai/pulse/ , 2020-10-1
[5] Sachit Menon*, Alexandru Damian*, Shijia Hu, Nikhil Ravi, Cynthia Rudin Duke University , Durham, NC (2020) . "PULSE: Self-Supervised Photo Upsampling via Latent Space Exploration of Generative Models ". arXiv : 2003.03808v3
[6] 映像品質の主観評価法
https://www.ntt.co.jp/qos/technology/visual/01_5_1.html
,2020-11-22