

深層生成モデルを用いた顔分布の獲得と  
顔サンプルの生成

筑波大学

図書館情報メディア研究科

2021年03月

周東 拓哉

# 目次

<b>第1章 序章</b>	<b>1</b>
1.1 研究背景	1
1.2 研究の目的	1
1.3 本論文の構成	1
<b>第2章 関連研究</b>	<b>3</b>
2.1 本章の概要	3
2.2 顔の集団的特徴に関する研究	3
2.2.1 平均顔	3
2.2.2 基底顔を用いた顔空間	4
2.3 深層生成モデルに関する研究	6
2.3.1 GAN	6
2.3.2 DCGAN	8
2.3.3 PGGAN	10
2.3.4 StyleGAN	11
2.3.5 StyleGAN2	12
2.4 深層生成モデルの潜在変数空間に関する研究	14
2.4.1 InterFaceGAN	14
2.4.2 Image2StyleGAN	16
2.5 本章のまとめ	17
<b>第3章 提案手法</b>	<b>18</b>
3.1 本章の概要	18
3.2 顔分布の獲得	18
3.2.1 潜在変数群の獲得	19
3.2.2 次元圧縮	19
3.2.3 多変量正規分布の計算	21
3.3 顔サンプルの生成	21
3.3.1 顔分布からのサンプリング	21
3.3.2 次元復元	22
3.3.3 StyleGAN2による顔画像生成	23
3.4 本章のまとめ	24
<b>第4章 実験</b>	<b>25</b>
4.1 本章の概要	25
4.2 データセット	25
4.3 顔画像埋め込み結果	25
4.3.1 潜在変数の散布図	25
4.3.2 顔グループ間の分離可能性	27



4.4	潜在変数空間中の顔分布間の関連性 . . . . .	28
4.5	顔サンプル生成結果 . . . . .	30
4.5.1	平均顔 . . . . .	30
4.5.2	累積寄与率および主成分数ごとの生成結果 . . . . .	31
4.5.3	データセットの画像枚数ごとの生成結果 . . . . .	33
4.6	StyleGAN2 における InterFaceGAN の追実装 . . . . .	34
4.7	本章のまとめ . . . . .	37
<b>第 5 章</b>	<b>考察と議論</b>	<b>38</b>
5.1	本章の概要 . . . . .	38
5.2	顔分布における主成分の特性 . . . . .	38
5.3	顔分布の次元数と顔サンプルの関係 . . . . .	40
5.4	主成分分析における適切な累積寄与率および主成分数の選定 . . . . .	42
5.5	本章のまとめ . . . . .	43
<b>第 6 章</b>	<b>まとめ</b>	<b>44</b>
	謝辞	45
	参考文献	46

# 目次

2.1	平均顔の例	4
2.2	重ね合わせ法による平均顔の生成	4
2.3	顔空間上の顔画像合成	5
2.4	顔空間上における無表情から笑顔への遷移のイメージ	5
2.5	GAN の概念図	6
2.6	GAN による顔画像生成	7
2.7	潜在変数と生成画像の関係	8
2.8	DCGAN の生成器	9
2.9	潜在変数同士の算術演算	9
2.10	属性変換軸による属性変換の例	10
2.11	PGGAN の学習	10
2.12	StyleGAN のアーキテクチャ	11
2.13	Style Mixing	12
2.14	StyleGAN2 の生成器と識別器の構造	13
2.15	StyleGAN2 の顔画像生成結果	14
2.16	意味的境界となる超平面の計算	15
2.17	StyleGAN を用いた年齢属性の変換	15
2.18	直行しない属性ベクトルが存在する場合の属性ベクトルの修正	16
2.19	Image2StyleGAN による画像の再構成	17
3.1	提案手法のフロー	18
3.2	顔分布の計算	19
3.3	高次元多変量正規分布のサンプルの極座標表示	20
3.4	主成分分析による主成分の決定	20
3.5	顔分布からの顔サンプル生成フロー	21
3.6	3次元のギブスサンプリング	22
3.7	次元圧縮と次元復元の流れ	23
3.8	サンプル潜在変数の Synthesis Network への入力	24
4.1	3次元の主成分空間上の潜在変数群の散布	26
4.2	第一主成分と第二主成分で構成される平面	26
4.3	第一主成分と第三主成分で構成される平面	27
4.4	k-means クラスタリングの計算結果の可視化	27
4.5	第一主成分方向への遷移による生成画像の変化	29
4.6	第二主成分方向への遷移による生成画像の変化	29
4.7	第三主成分方向への遷移による生成画像の変化	30
4.8	重ね合わせ法による平均顔と提案手法による平均顔の比較	31
4.9	累積寄与率ごとの顔サンプル生成結果	32

4.10	累積寄与率 1.0 (112 次元) の主成分空間からのエラーサンプル例 . . . . .	32
4.11	$W_+$ 空間中でサンプリングを行った場合の顔サンプル例 . . . . .	33
4.12	データセットの枚数ごとの顔サンプル生成例 (累積寄与率 0.25) . . . . .	34
4.13	StyleGAN2 を用いた InterFaceGAN の追実装結果 . . . . .	36
5.1	男性アイドルの主成分軸方向の生成画像の変化 . . . . .	39
5.2	女性アイドルの主成分軸方向の生成画像の変化 . . . . .	39
5.3	男性格闘家の主成分軸方向の生成画像の変化 . . . . .	40
5.4	次元数増加による平均付近のサンプルの生成確率の変化 . . . . .	41
5.5	顔分布の次元数ごとの L2 ノルムと角度によるサンプルの極座標表示 . . . . .	42
5.6	最適化を用いる際の損失関数の特性 . . . . .	43

# 第1章 序章

## 1.1 研究背景

顔を理解するための研究は古くから行われており、特に顔の集団的特徴の理解は顔学における重要なテーマとして研究がなされている。顔が集団として持つ特徴を理解するための研究として、集団の代表値として平均を用いる手法である平均顔が知られている。画像を重ね合わせて平均値を取ることで、個別具体的な特徴を平均化処理によって消して共通の特徴のみを抽出することができる。また、顔の分布化を行う試みとして基底顔を用いた顔空間に関する研究 [1] がある。基底顔と呼ばれる互いに直交するような顔画像を重み付き加算によって重ね合わせることで一つの顔を再現する手法である。各基底顔に対する重みを座標として扱うことができるため、基底顔による顔空間は顔の分布化に成功した。このような顔の統計量を計算する手法が主流である一方で、これらの手法は抽出した顔特徴を視覚的に理解するために画像を生成しても、重ね合わせ法を用いることから画像がぼやけてしまうという問題点がある。したがって、ぼやけのない高品質な顔画像が生成できる空間での顔の統計量化が必要である。

## 1.2 研究の目的

近年の深層生成モデルは本物の顔を見分けのつかないほどの顔画像を生成することができ、重ね合わせ法を用いない画像生成を行うことからぼやけを発生させない。そのため、集団の顔画像に対応する潜在変数群を特定することで、集団の顔を分布化することが可能である。十分に訓練された深層生成モデルの潜在変数空間には顔の属性を表す属性軸が存在することがわかっているため [2]、分布に対して大きな分散を持つ軸を抽出することで集団を特徴付ける属性を特定することができると本研究では考えた。そこで、本研究では深層生成モデルの潜在変数空間中で顔の分布化を行うことでぼやけのない顔サンプルの生成の実現と、顔分布を用いた顔の集団的特徴の理解を目的とした。

## 1.3 本論文の構成

本論文では、深層生成モデルとして高解像度で多種多様な顔画像の生成が可能なモデルである StyleGAN2[3] を採用し、100 名程度の顔グループを対象とした顔分布・顔サンプルの生成結果を報告し、顔分布から得られる顔の集団的特徴を示す。以下の構成で提案手法の有効性について論じる。

- 2章 関連研究 顔の集団的特徴と顔画像生成モデルに関する従来の研究の紹介
- 3章 提案手法 顔分布の計算と顔サンプルの生成手法に関する具体的な説明
- 4章 実験 顔分布の計算結果と顔サンプルの生成結果

**5章 考察と議論** 実験結果に関する説明および問題点の説明

**6章 まとめ** 本論文の寄与と今後の課題

## 第2章 関連研究

### 2.1 本章の概要

本章では，本研究の研究背景の位置づけを明確にするために必要な関連研究の紹介を行う．まず従来の平均顔や任意の顔の再構成を，古典的な重ね合わせ法によって実現する手法とその問題点について述べる．次に近年の深層学習を用いた顔画像生成分野において代表的な深層生成モデルを紹介し，モデルと潜在変数空間の関係についてもその中で概要を述べる．深層生成モデルの中でも特に，著しい発展を遂げた Generative Adversarial Nets (GAN) の発展系であるモデルについても複数取り上げることで顔画像生成モデルの歴史を概観する．最後に，近年の深層生成モデルの潜在変数の解釈に関する研究を紹介し，本研究の提案手法の理解の一助とする．

### 2.2 顔の集団的特徴に関する研究

本節では，集団としての顔を説明する上で重要な概念である平均顔を説紹介するとともに，その発展系である顔空間について述べる．また，これらの古典的手法を用いた顔画像生成手法の問題点についても説明する．

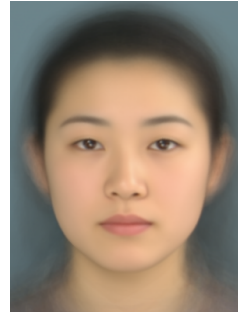
#### 2.2.1 平均顔

顔が集団として持つ特徴を理解するために有効な手法として平均顔が有名である．顔画像を重ね合わせて平均化することで，集団の固有の顔特徴を抽出することが目的とされている．図 2.1 に向田ら [4] による平均顔生成例を示す．平均顔の特徴としては，以下のよう  
なものが挙げられる．

- 左右対象
- シワ・ニキビ等がない
- 画像全体のぼやけ



(a) 日本人男性 60 人の平均顔



(b) 日本人女性 60 人の平均顔

図 2.1: 平均顔の例. [5] より引用.

これらの平均顔の特徴は、重ね合わせ法を用いた画像間の平均化処理によってもたらされる。実際の顔にはシワやニキビ、骨格の歪みなど個別具体的な特徴が存在する。しかし、画像間のパターンの平均化処理は集団に共通の特徴のみを抽出するため個別具体的な特徴は失われる。永田ら [6] は平均顔の特徴を利用し職業別の平均顔を生成することで、職業ごとに顔の集団的特徴に差があることを述べている。平均顔は重ね合わせ法と呼ばれる合成手法を採用しており、各画像の全ピクセル値に均等に重みをかけて足し合わせるという方法である。重ね合わせ法概念図を図 2.2 に示す。

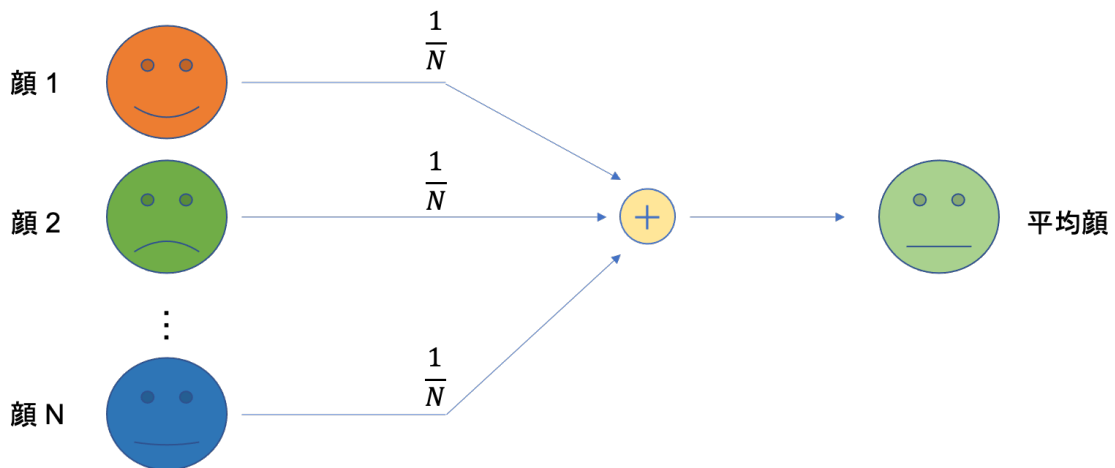


図 2.2: 重ね合わせ法による平均顔の生成

### 2.2.2 基底顔を用いた顔空間

平均顔が集団に固有の特徴を抽出できる一方で、顔は平均だけでは理解することができない。顔の集団の特徴のばらつきを空間上の分布として表現することで顔の分散を考慮した顔合成技術として顔空間 [1] がある。顔空間のコンセプトは、基底となるような複数の顔画像を重ね合わせ法で合成するというものである。ただし、平均顔の生成とは異なり基底顔画像にかける重みが単なる平均化処理ではなく、調節可能な点が異なる。図 2.3 に示すように、 $n$  次元の顔空間の場合、各基底顔にかかる重み  $c_1 \sim c_n$  の組み合わせによって顔空間は構成される。基底顔は顔空間の基底ベクトルを構成するため、それぞれ直交するような顔画像が採用される。顔空間の基底顔は正弦波の実現と似ており、人の顔に共通な成分である直流成分と各部位それぞれ異なる表情などの成分である交流成分で構成される。

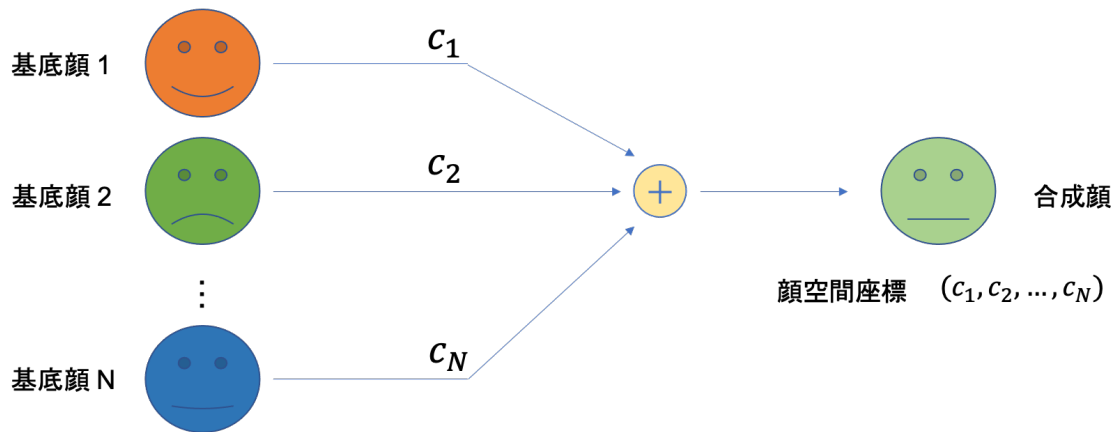


図 2.3: 顔空間上の顔画像合成

顔空間を定義することによって、顔空間上の座標で顔を表現することができる。顔空間を応用することで顔グループの座標の分布を得ることが出来るため、顔分布による集団の特徴の理解が可能である [7].

顔の属性の代表的なものには「年齢」・「表情」など連続的に変化するものが存在する。これらの顔の印象と顔空間上の顔の変化というのは必ずしも一致せず、一定の表情を表す顔空間上の軸なども曖昧である。図 2.4 に例として示すように、無表情から笑顔までの変化を顔空間上で再現すると、自然な変化を再現するには非線形な経路を辿り、線形に表情同士を結んでも不自然な表情変化になってしまうことがある。また、あくまでも重ね合わせ法を用いていることから基底顔の表現力への依存を避けられないことが課題となっている。

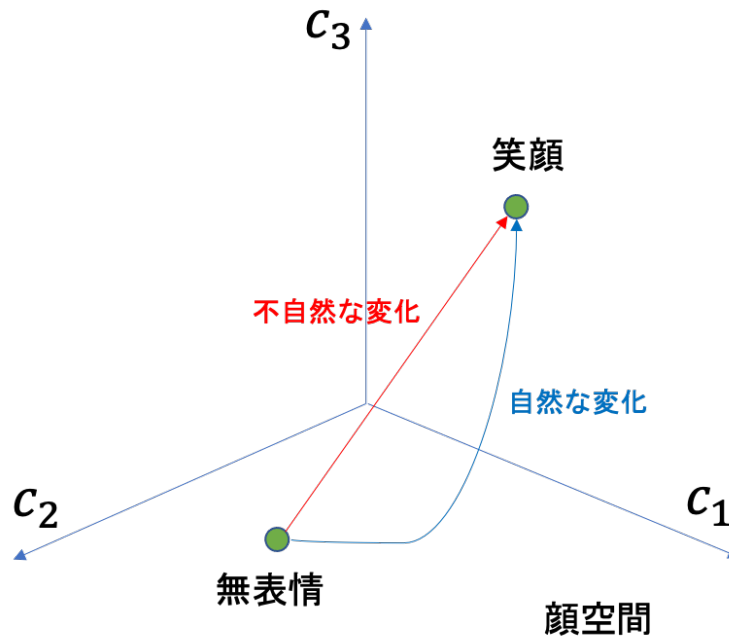


図 2.4: 顔空間上における無表情から笑顔への遷移のイメージ



## 2.3 深層生成モデルに関する研究

本節では、顔画像生成分野において代表的な画像生成モデルについて述べる。また、GAN (Generative Adversarial Nets) の発展の歴史を概観することでGANを用いた画像生成の仕組みを説明し、生成モデルと潜在変数空間の対応関係や近年の生成モデルの潜在変数空間の特性について述べる。

### 2.3.1 GAN

GANとは、Goodfellowら[8]によって提案された画像生成分野における深層生成モデルの学習手法の基礎になった手法である。図2.5に示すように、画像を生成する生成器(Generator)と画像を識別する識別器(Discriminator)を互いに競わせる形で学習が進む。生成器は潜在空間における潜在変数から画像を生成するネットワークになっている。識別器はその画像が偽造されたものかどうかを判別する役割を持っており、生成器は識別器をいかに騙すかを学習していく。図2.6にGANによる顔画像生成結果を示す。

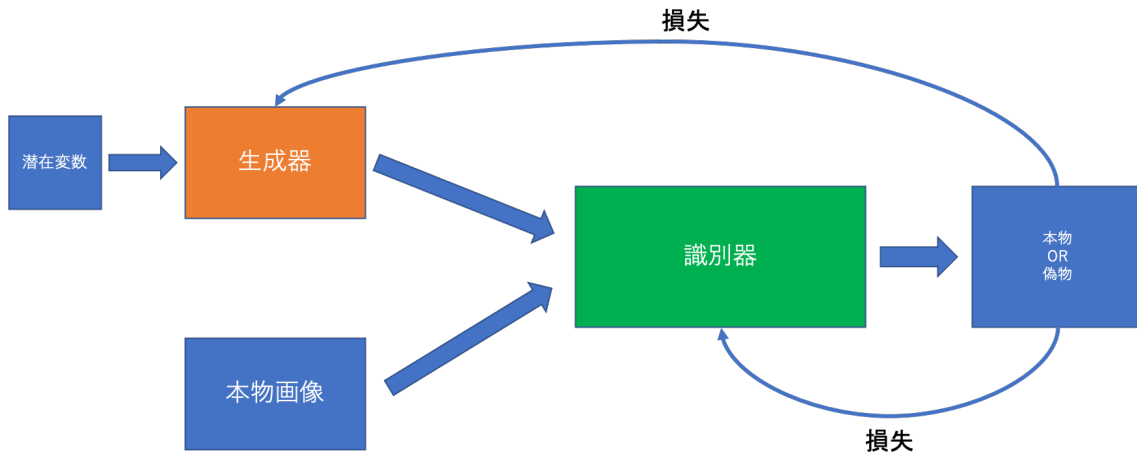


図 2.5: GAN の概念図



図 2.6: GAN による顔画像生成. [8] より引用.

識別器は事前に学習に用いるデータセットで学習させておき、ある程度の識別精度を確保しておいてから学習をスタートする. 学習段階では生成器と判別器を交互に学習させることでお互いに競わせるようにネットワークを学習させる. 損失関数には敵対的損失 (Adversarial Loss) を用いる. 潜在変数空間の位置ベクトル  $z$  から画像を生成する生成器  $G(z)$  のパラメータを  $\theta_G$ , 画像  $x$  が本物か偽物かを識別する識別器  $D(x)$  のパラメータを  $\theta_D$  とすると, それぞれの損失  $J_D$  と  $J_G$  はそれぞれ以下のように表せる. また  $p_d$  は訓練画像の確率分布であり,  $p_z$  はランダムノイズの確率分布とする.

$$J_D(\theta_G, \theta_D) = -\frac{1}{2} E_{x \sim p_d} [\log D(x)] - \frac{1}{2} E_{z \sim p_z} [\log(1 - D(G(z)))] \quad (2.1)$$

$$J_G(\theta_G, \theta_D) = -J_D(\theta_G, \theta_D) \quad (2.2)$$

識別器の損失関数  $J_D$  においては, 第一項で訓練画像の集合  $p_d$  から抽出した画像を用いた場合の損失を計算し, 第二項で生成器による生成画像から抽出した場合の損失を計算する.

GAN による画像生成においては, 図 2.7 に示すように潜在変数と生成画像が一对一の対応関係を持っている. ゆえに潜在変数空間は一種の顔空間であると言える. 潜在変数と生成画像の関係に関しては, GAN の発展系のモデルに関しても同様であり生成モデルごとに顔空間としての潜在変数空間を持っている.

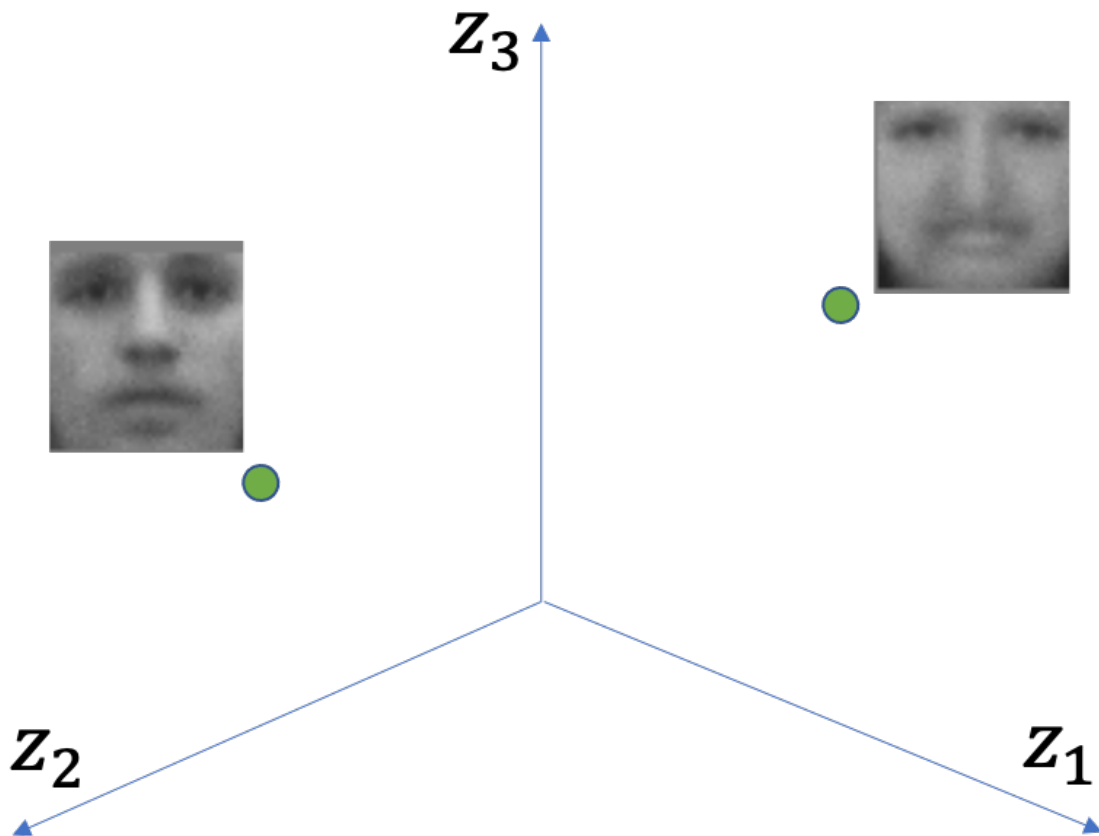


図 2.7: 3次元の潜在変数空間を例とした潜在変数と生成画像の関係. [8]より顔画像を引用・修正.

### 2.3.2 DCGAN

Radfordら[9]のDCGAN (Deep Convolutional GAN)はGANにCNN (Convolutional Neural Network)を導入した. GANでは画像内ピクセルの隣接関係や局所的な画像特徴までは再現できなかったが, 畳み込みおよび転置畳み込みを導入することでより鮮明な画像の生成を実現した. DCGANの生成器モデルの構造を図2.8に示す.

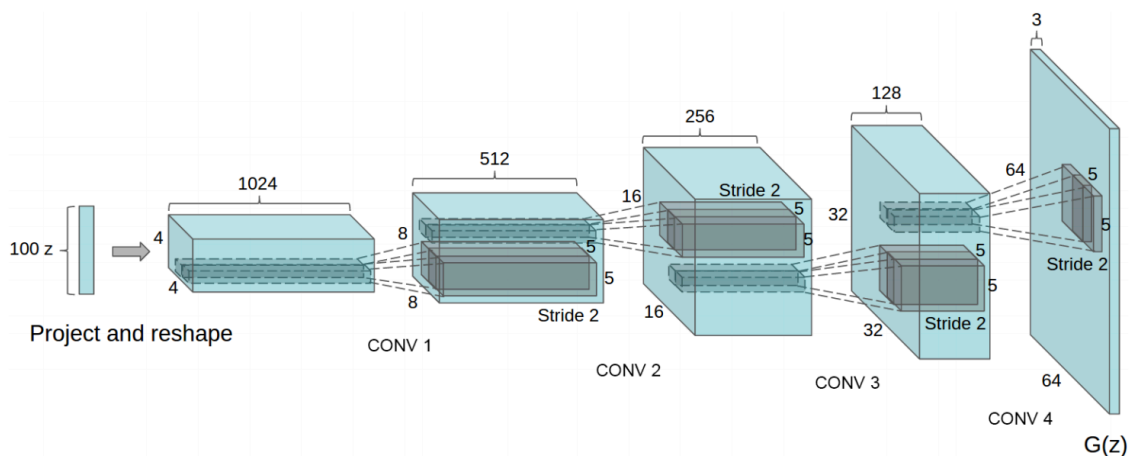


図 2.8: DCGAN の生成器. [9] より引用.

DCGAN のさらなる特徴として、潜在変数同士の算術演算と属性変換軸の発見が挙げられる。図 2.9 と図 2.10 に示すように、DCGAN の潜在変数同士の算術変換によって画像の意味的算術演算が可能であり、属性を変換することが可能な属性軸が存在する。これらの発見は潜在変数空間が意味的に定量化されていることを示唆した。

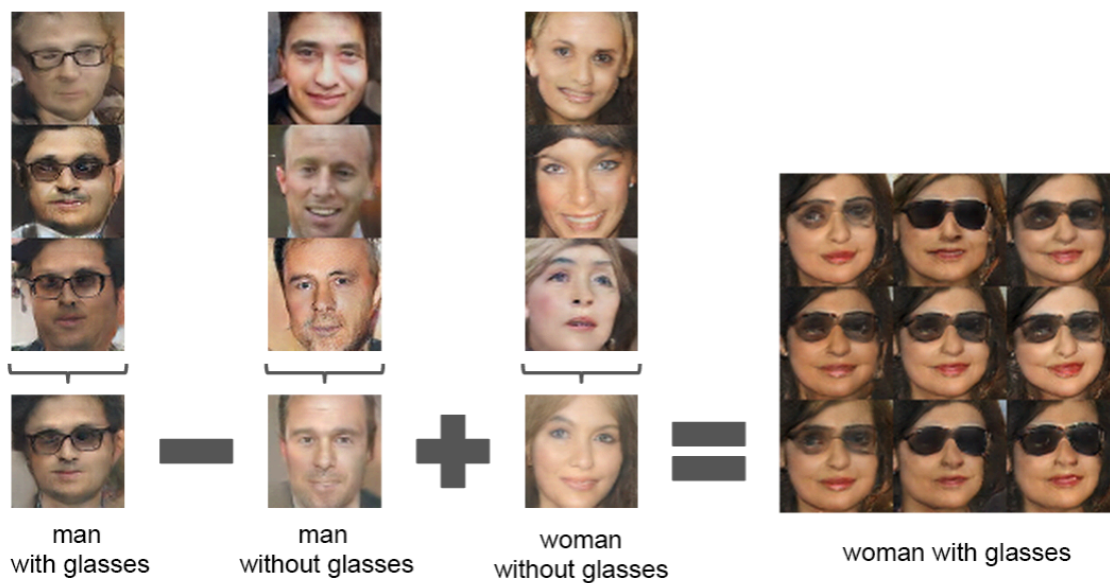


図 2.9: 潜在変数同士の算術演算. [9] より引用.



図 2.10: 属性変換軸による属性変換の例 (顔の向き). [9] より引用.

### 2.3.3 PGGAN

DCGAN はより鮮明な画像の生成を実現したが, その画像サイズは  $64 \times 64$  pixel 程度であり高解像と呼ぶには十分ではない. Karras ら [10] が提案した PGGAN (Progressive-Growing GAN) はネットワーク構造を段階的に成長させる Progressive-Growing を導入することで高解像度の顔画像生成を実現した. Progressive-Growing では, まず低解像度の画像で学習を行い, その解像度での精度が十分になったら画像サイズを倍にして学習済みネットワークに層を追加していく. 高解像度の画像の生成は生成器の勾配をとることが非常に困難で不安定なため, 段階的にネットワークを拡張していくことで学習の安定性を向上させた. 図 2.11 に PGGAN の学習過程を示す.

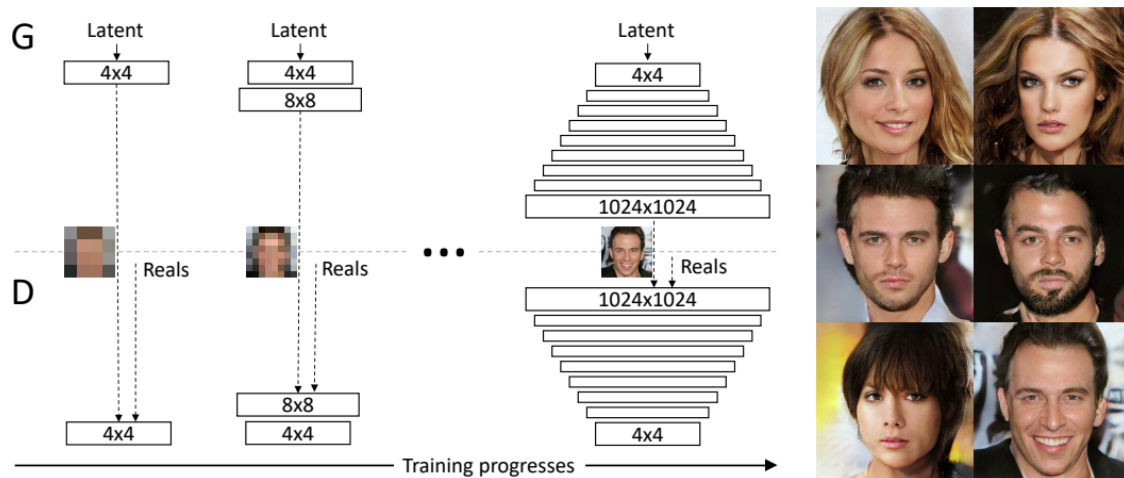


図 2.11: PGGAN の学習. [10] より引用.



PGGAN では 1024x1024pixel の高解像度顔画像の生成に成功したが、一方でシワやソバカスなどのテクスチャ情報の再現が難しいことや、訓練データの画像特徴への偏りが大きいことから DCGAN のような生成画像の属性変換が難しいという潜在変数空間上の問題点を抱えている。

### 2.3.4 StyleGAN

Karras らが提案した StyleGAN[11] は PGGAN に StyleTransfer[12] と呼ばれる画像処理を導入し、生成器の構造を Mapping Network と Synthesis Network の二つに分割することで、画像のスタイル操作に有利な潜在変数空間の学習を実現した。図 2.12 に示すように、Mapping Network により潜在変数空間に中間表現としての潜在変数空間を一つ増やすことで、データセットにおける特徴の偏りに依存する PGGAN の問題を解決した。また、Synthesis Network においてはスタイル変換の前にテンソルにノイズを加算することでシワやソバカスなど、細かなバリエーションの変化を実現した。

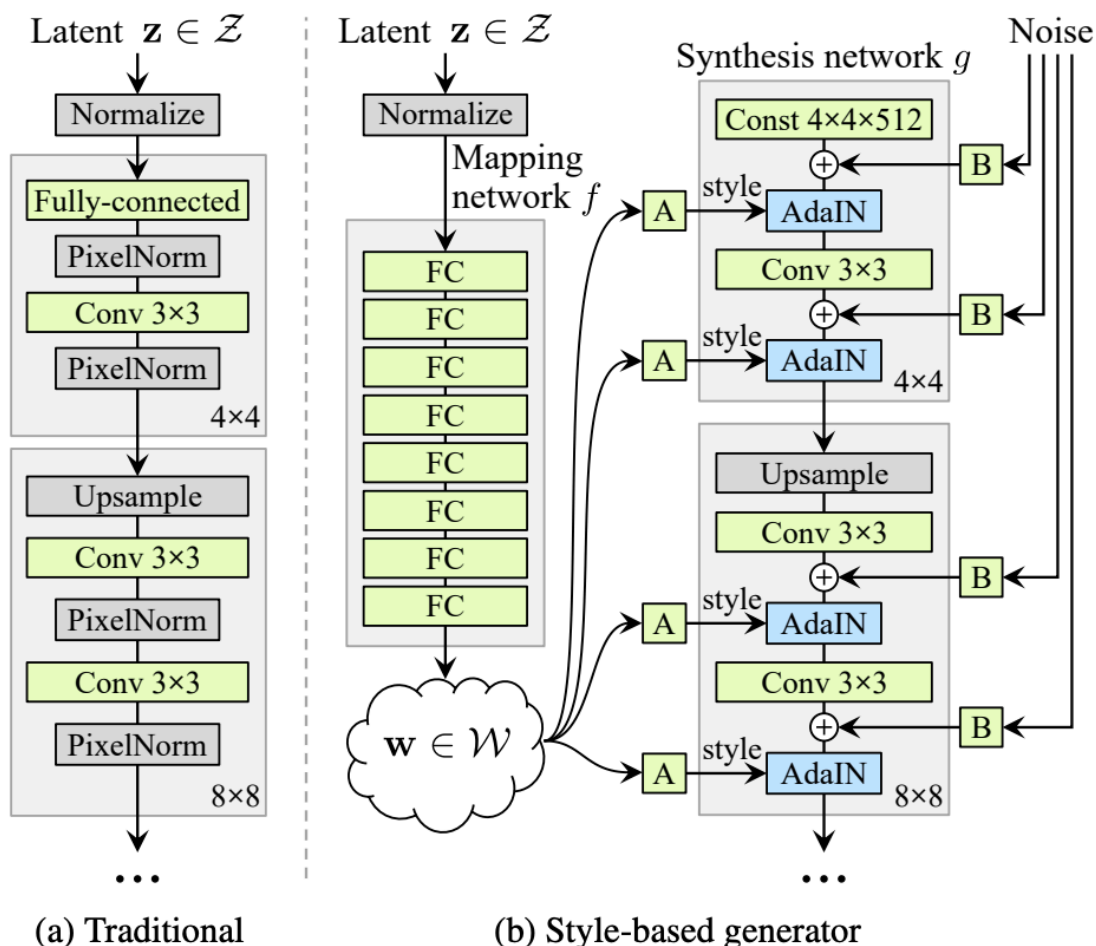


図 2.12: StyleGAN のアーキテクチャ. [11] より引用。

StyleGAN は図 2.13 に示す Style Mixing と呼ばれるスタイル情報の混合処理が可能なアーキテクチャであることが特徴の一つである。潜在変数空間の中間表現である  $W$  空間の潜在

変数は生成器の各層に入力可能であり、どの層にどの潜在変数を入力するかによって生成画像のスタイルを操作することができる。



図 2.13: Style Mixing. [11] より引用.

### 2.3.5 StyleGAN2

StyleGAN には Droplet と呼ばれる水滴状のノイズや不自然なモードが発生するという問題点が残っていた。Karras ら [3] は Style Transfer を CNN の重み正規化に置き換え、Progressive-Growing をやめることでこの問題を解決した。また、図 2.14 に示すように生成器と識別器の構造をそれぞれ Skip Network と Residual Network に変更することで、潜在変数空間の滑ら

かさを表す指標である PPL (Perceptual Path Length) と FID (Frechet Inception Distance) を改善している。StyleGAN2 による顔画像生成結果を図 2.15 に示す。

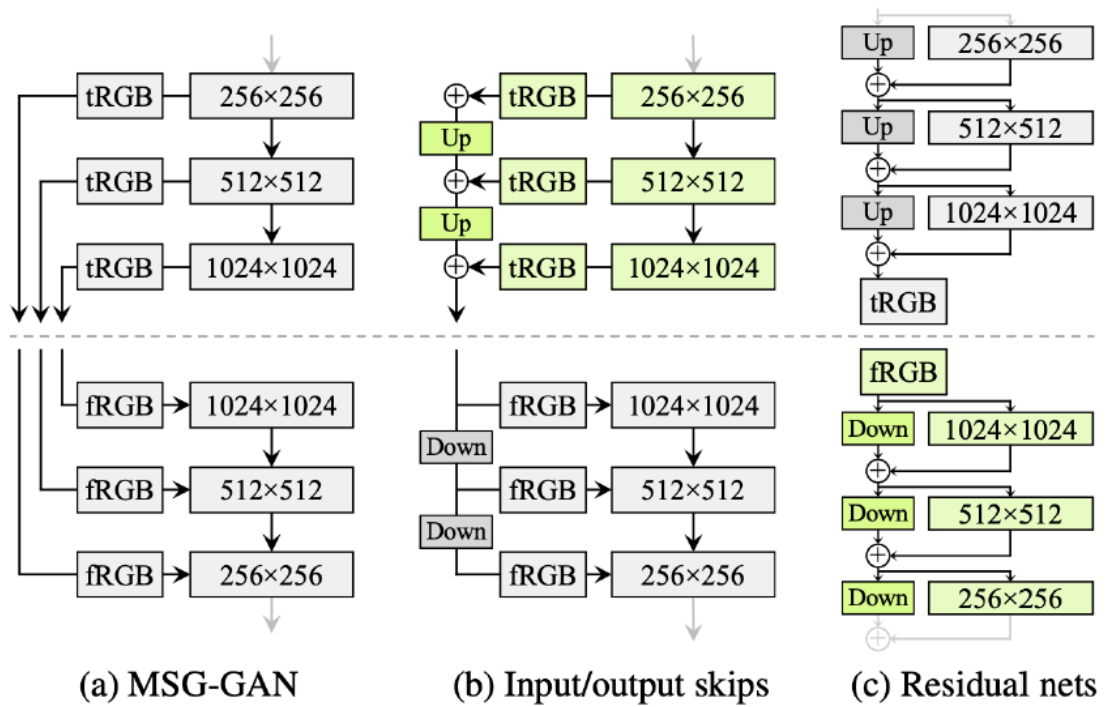


図 2.14: StyleGAN2 の生成器と識別器の構造. [3] より引用.





図 2.15: StyleGAN2 の顔画像生成結果. [3] より引用.

## 2.4 深層生成モデルの潜在変数空間に関する研究

本節では、本研究の提案手法で用いる顔画像に対する操作を理解するために必要な深層生成モデルの応用研究を紹介する。まず、深層生成モデルの潜在変数空間上の意味的定量化に関する根拠となる研究として InterFaceGAN を紹介する。次に、本研究の提案手法に 응용されている顔画像を潜在変数空間に埋め込む手法に関する研究として Image2StyleGAN を紹介する。

### 2.4.1 InterFaceGAN

2.3.2 項でも述べたように、DCGAN は深層生成モデルの潜在変数空間上に属性のとしての意味を持つ軸が存在することを示唆した。Shen ら [2] は StyleGAN を用いた生成画像の属性変換手法 InterFaceGAN を提案した。Shen らは潜在変数空間上にいくつかの意味的境界 (Semantic Boundary) があることを提唱した。実験では PGGAN および StyleGAN を採用し、大量のランダムな潜在変数空間上の座標から生成された顔画像に対して属性ラベルを付与し、属性値を線形分離するような超平面を線形 SVM (Linear Support Vector Machine) で求めている。図 2.16 に示すように、計算により求められた超平面に対して法線ベクトル方向に潜在変数空間中の座標を移動させることで属性変換を達成している。

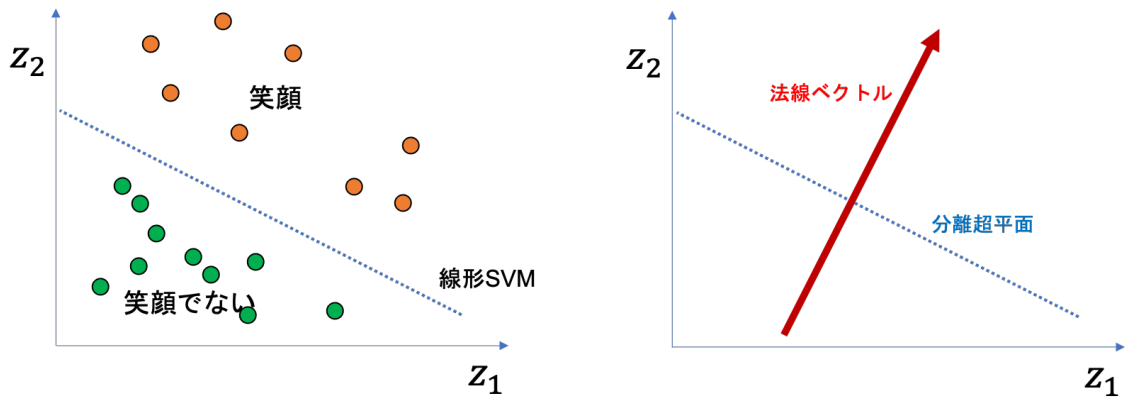


図 2.16: 意味的境界となる超平面の計算

InterFaceGANにより求められた属性ベクトルは必ずしも直交せず、互いに相関関係を持つことがわかっている。例えば、年齢ベクトルと眼鏡ベクトルは  $\cos$  類似度を計算すると近い値を示す。従って、図 2.17 に示すように、年齢ベクトル方向に潜在変数を移動させると生成画像上の顔に眼鏡が現れる。



図 2.17: StyleGAN を用いた年齢属性の変換。[2] より引用。

Shen らは属性ベクトルが互いに直交しない属性同士が同時に変化してしまう問題に対して、図 2.18 に示すように変更したい属性ベクトル  $n_1$  から、変更したくない属性ベクトル  $n_2$  の  $n_1$  方向の成分を引くベクトル演算を行うことで解決している。こうすることで、 $n_2$  と平行に移動させることができるため、特定の属性の変換を抑制することができる。

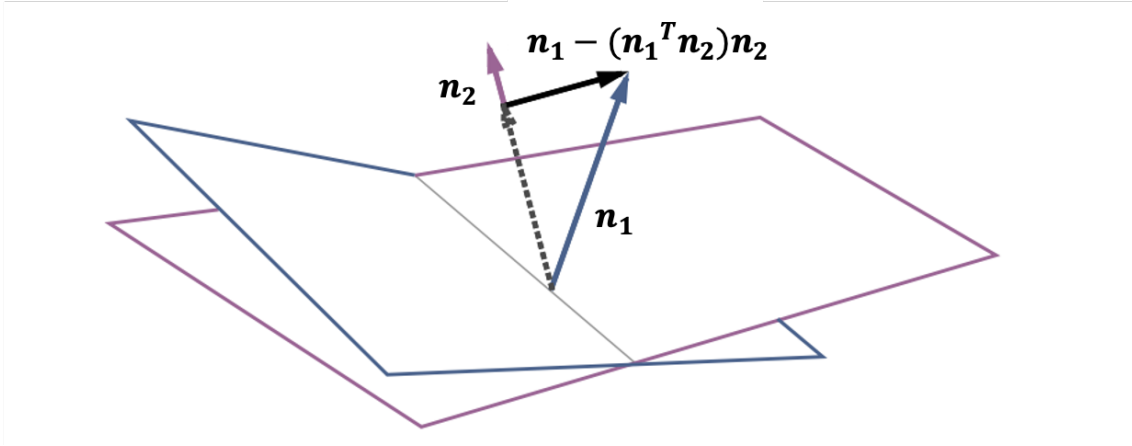


図 2.18: 直交しない属性ベクトルが存在する場合の属性ベクトルの修正. [2] より引用・一部修正.

この研究が示唆する通り、深層生成モデルは現実世界の属性の相関関係すら意味的定量化を行うようになっている。2.2.2 項で述べた基底顔による顔空間では顔の属性と空間上の座標の対応関係がついていないことが問題であったが、本研究では近年の深層生成モデルの潜在変数空間を用いることでこの問題は解決できると考えた。また、InterFaceGAN を用いることで潜在変数空間中における属性軸の推定を行うことが可能になるものの、顔の集団的特徴などの未知の属性を抽出するといった目的には利用することができない。従って、本研究で提案するような特定の集団にとって意味のある属性軸の抽出が必要である。

#### 2.4.2 Image2StyleGAN

2.2 節では、顔空間中に埋め込んだ顔の分布を考察することで集団の顔を理解する手法を紹介した。現実の顔を顔空間上に投影できなければ、現実の顔の特徴を分布化することはできない。Abdal ら [13] は、StyleGAN を用いて任意の画像を再構成するような潜在変数を特定する手法である Image2StyleGAN を提案した。Image2StyleGAN では、生成画像とターゲット画像の間の知覚損失 (Perceptual Loss) を最小化するような潜在変数を特定することで任意の顔画像を再構成する。知覚損失の計算には、Zhang ら [14] の ImageNet で学習済みの VGG16 モデルが使われており、物体認識モデルにより抽出された特徴量を比較することで損失が計算されるため、単なる画素値の比較とは違い知覚的な特徴を捉えることができる。図 2.19 に Image2StyleGAN による最適化処理の概要を示す。

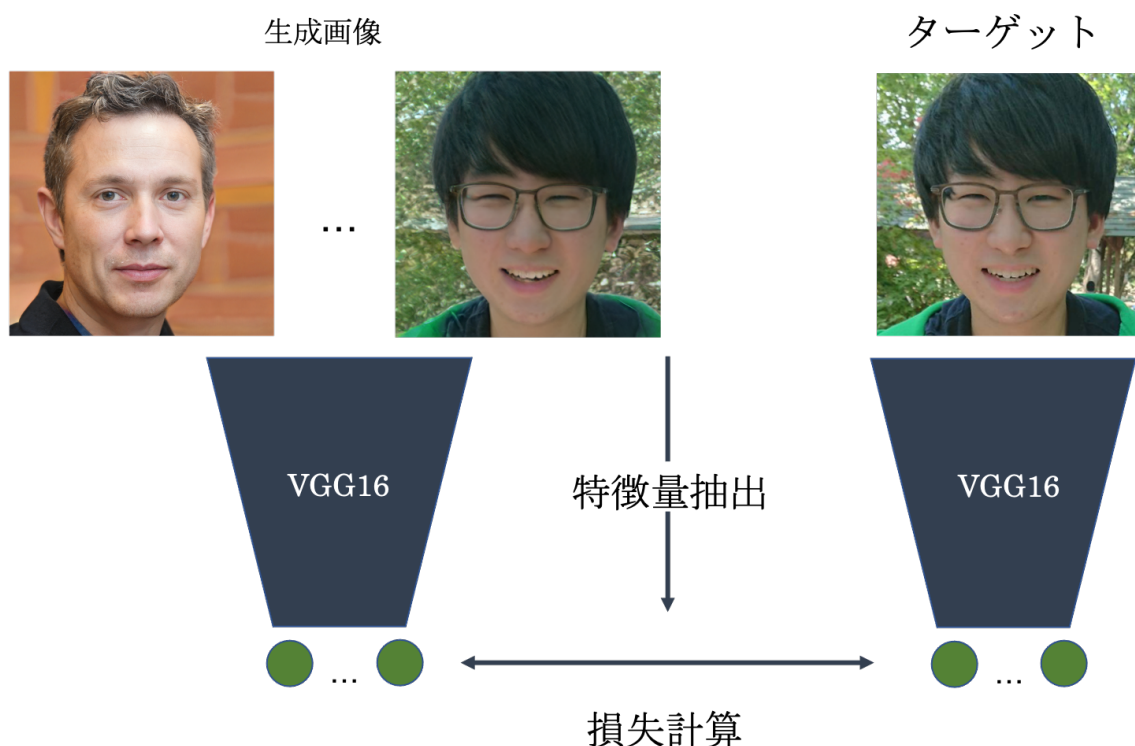


図 2.19: Image2StyleGAN による画像の再構成

2.3.4 項で述べたように StyleGAN には中間表現の潜在変数空間  $W$  が存在する。  $W$  空間の次元数は 512 次元であり、 Synthesis Network の 18 個の層にどの潜在変数を入力するかによって生成画像のスタイルを決定している。 そのため、 512 次元の  $W$  空間そのままでは StyleGAN の表現力を最大限に活用することができない。 そこで、 Abdal らは Synthesis Network の各層に入力される 18 個の  $W$  空間の潜在変数を結合した  $W_+$  空間上で最適化処理を行っている。  $W_+$  空間の次元数は 9216 ( $=18 \times 512$ ) であり、 この空間上で最適化処理が行われる。 本研究では、 StyleGAN2 の潜在変数空間を顔空間として扱うため Image2StyleGAN の手法を StyleGAN2 の  $W_+$  空間で追実装した上で応用している。

## 2.5 本章のまとめ

本章では、 本研究の提案手法の前提となる知識や近年の深層生成モデルの研究状況について紹介した。 また、 潜在変数空間が意味的定量化を施した顔空間である根拠となる InterFaceGAN と本研究の提案手法で応用している Image2StyleGAN について説明した。 本研究では顔の集団的特徴を表す属性軸を抽出する必要があるため、 従来の所望の属性軸の推定手法では対応することはできない。 次章では深層生成モデルの潜在変数空間中にて顔の集団的特徴を抽出し、 顔の分布化を行う手法を説明する。



## 第3章 提案手法

### 3.1 本章の概要

本章では、前章で述べた研究背景や関連研究の課題を踏まえて提案手法を説明する。従来の手法では所望の属性軸を推定することが可能だが、集団が持つ顔の分布を特徴付けるような属性軸の抽出は出来ない。そこで、本研究では潜在変数空間に集団の顔画像群に対応する潜在変数を推定して対応づけることで、集団の顔分布を計算する手法を提案する。顔分布は集団が持つ属性の分析と分布に従う顔サンプルに利用することが出来る。本提案手法では、まず顔分布の計算手法を最初に説明する。次に、顔分布に従う未知の顔サンプルの生成手法を説明する。本提案手法のフローを図 3.1 に示す。

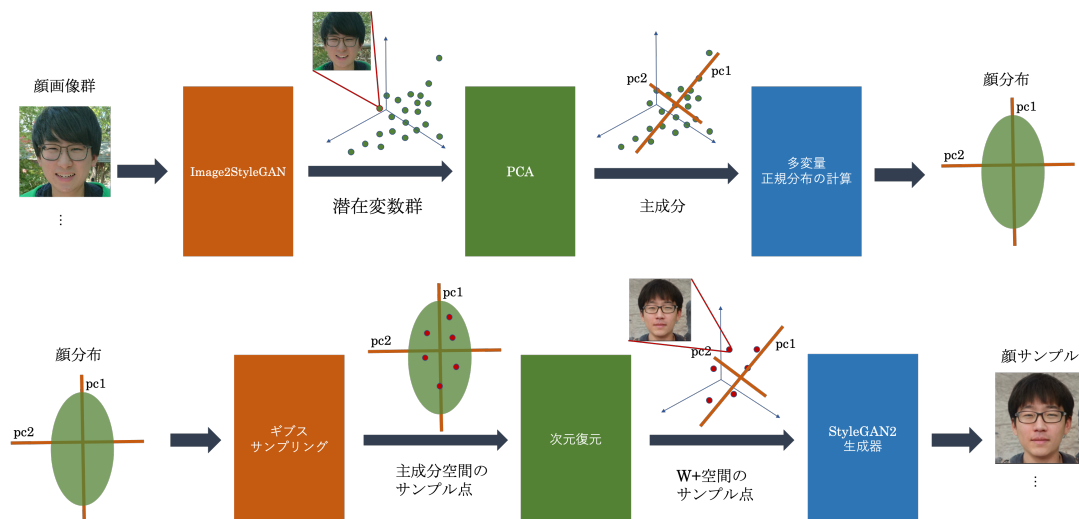


図 3.1: 提案手法のフロー

### 3.2 顔分布の獲得

本節では、ある集団の顔画像群からその集団の特徴を表現する顔分布の計算手法について三つの工程に分けて説明する。まず、データセット中の顔画像群を深層生成モデルの潜在変数空間に埋め込む最適化処理について説明する。次に、得られた潜在変数群を表現する上で重要な特徴成分を抽出するために主成分分析による次元圧縮を適用し、主成分空間上に射影する手法を説明する。最後に、主成分空間上で得られた潜在変数群から多変量正規分布を計算することで顔分布を獲得する手法を説明する。顔分布の計算フローを図 3.2 に示す。

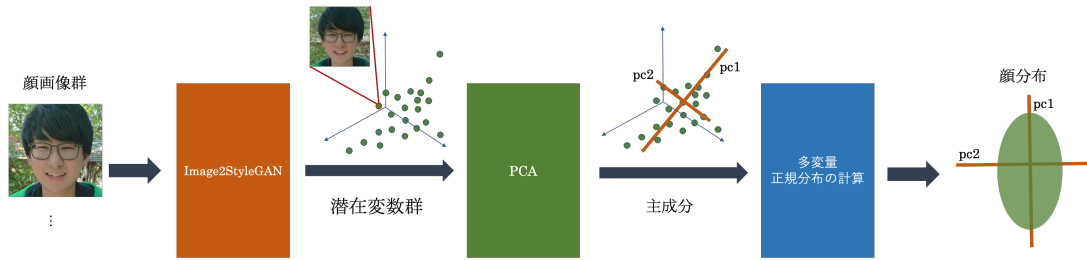


図 3.2: 顔分布の計算

### 3.2.1 潜在変数群の獲得

2.4.2項で述べたように、本研究では顔画像群の潜在変数空間への埋め込みには Image2StyleGAN を StyleGAN2 の  $W_+$  空間上で追実装することで実現する。StyleGAN と StyleGAN2 の  $W_+$  空間の扱いは同様であり、生成画像とターゲット画像の間の知覚損失を最小化するような  $W_+$  空間中の座標を最適化処理によって推定する。以下に顔画像の潜在変数空間への埋め込みに用いた最適化処理の概要を示す。また、知覚損失モデルに用いた VGG16 モデルは画像の入力サイズは 299x299pixel であり、知覚損失計算の前に画像のリサイズを施す必要がある。

**最適化手法** Adam[15]

**最適化処理の更新回数** 1000 回

**知覚損失モデル** ImageNet で学習済みの VGG16[14]

### 3.2.2 次元圧縮

前項の処理で得られた潜在変数群は  $W_+$  空間上の変数であり、9216 次元の高次元空間で表現される。これらの基底軸が表す意味は曖昧であり、このまま分布を計算すると非効率で高次元な分布になってしまう。本研究ではただ分布化するだけではなく、顔の集団的特徴を表す属性を抽出する必要がある。また、このような高次元空間上の多変量正規分布ではサンプルが超球面上に分布してしまう球面集中現象が発生する [16]。したがって高次元空間では、平均付近に密であり平均から離れると疎になる正規分布の性質を持つサンプルの生成が再現できなくなってしまう。この現象は、多次元空間における平均付近が全ての軸において平均付近であることによるものである。つまり、全軸中の一つの軸でも平均付近から外れてしまうと平均付近であるとは言えなくなってしまうのである。図 3.3 は、多変量正規分布から得られた大量のサンプルに対して L2 ノルムと角度を用いた極座標表示をした結果である。低次元空間では中心が密で外側が疎な確率密度関数と一致するサンプルの分布が得られる一方で、次元数  $n$  の高次元空間において、標準化された多変量正規分布では半径  $\sqrt{n}$  の超球面上に分布してしまう。

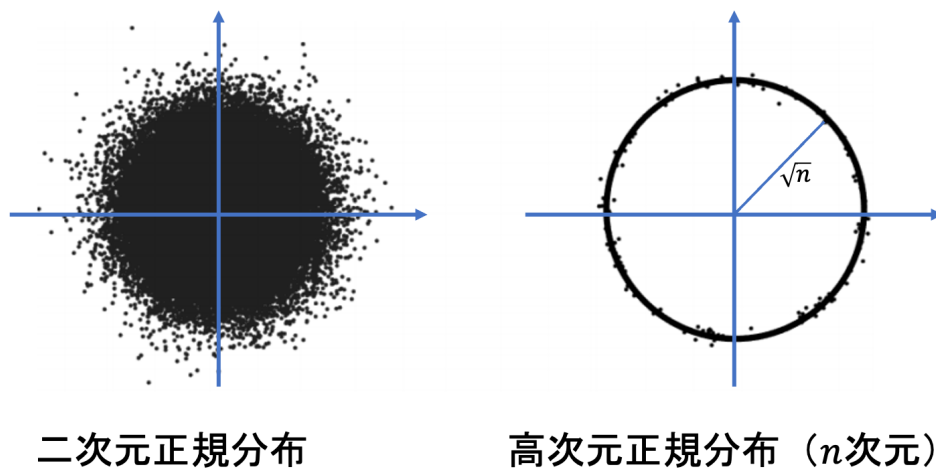


図 3.3: 高次元多変量正規分布のサンプルの極座標表示. [16] より引用・修正.

そこで、本研究では主成分分析 (Principal Component Analysis: PCA) による次元圧縮を施す。図 3.4 に示すように、主成分分析はデータに対して分散幅が大きくなるような主成分軸を複数選定し、主成分軸で構成される主成分空間上にデータを投影することでデータの次元数を圧縮をする手法である。データに含まれるばらつきの軸を重要と思われる順に採用するため、抽出される主成分には集団を特徴づける属性が現れる。また、結果として次元数を減らすこともできるため球面集中現象を軽減することが可能である。次元圧縮後のデータから復元できるデータの割合を表すのが累積寄与率であり、主成分空間の次元数が増加すればするほどカバーできる次元方向が増えるため、累積寄与率は高くなる。圧縮後の主成分数については、生成される顔サンプルの品質と多様性について考慮した上で決定する。主成分数の決定に関しては、4 章で累積寄与率ごとの顔サンプルの生成結果を示し、5.4 節で結果について議論する。

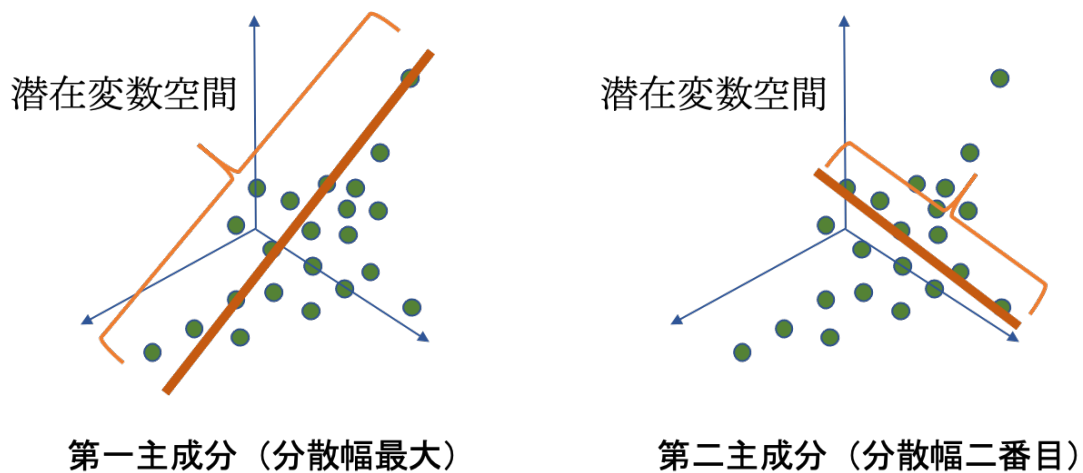


図 3.4: 主成分分析による主成分の決定.

### 3.2.3 多変量正規分布の計算

顔空間中に集団として顔が持つ特徴を表現するために、主成分分析により主成分空間中に投影された潜在変数群から顔分布を計算する。本研究では、顔分布は多変量正規分布に従うものと仮定する。こうすることで潜在変数群の平均と分散共分散により顔分布を表現することが可能になる。顔分布の平均を  $\mu$ ，分散共分散行列を  $\Sigma$  とすると  $n$  次元の顔分布の確率密度関数は以下のように表せる。

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{|\Sigma|}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})\right\} \quad (3.1)$$

顔分布における平均は集団内で共通の特徴を表し、主成分は集団内においてばらつきの多い特徴を表している。したがって、平均の顔を基準として各主成分が示す属性ごとに分散共分散を元にばらつきをもたせることで、集団の傾向を捉えた未知の顔サンプルが生成できる。

## 3.3 顔サンプルの生成

本節では、前節で述べた手法で得られる顔分布から、未知の顔サンプルを生成するための手法を三つの工程に分けて説明する。まず、多変量正規分布で表される顔分布からのサンプリング手法について説明する。次に、サンプル潜在変数を主成分空間から StyleGAN2 が生成可能な  $W_+$  空間への射影する手法について説明する。最後に、StyleGAN2 による顔サンプルの生成について説明する。顔サンプルの生成フローを図 3.5 に示す。

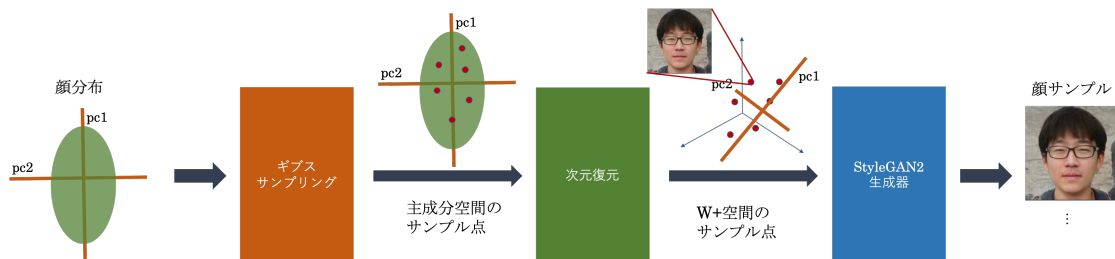


図 3.5: 顔分布からの顔サンプル生成フロー

### 3.3.1 顔分布からのサンプリング

顔分布からのサンプリングにはギブスサンプリングを用いる。ギブスサンプリングはマルコフ連鎖モンテカルロ法に分類される [17]。マルコフ連鎖モンテカルロ法は直前のサンプル点に依存して新しいサンプル点を決定する手法であり、サンプル同士の連鎖関係がマルコフ連鎖と名前がつく由来である。その中でもギブスサンプリングはある 1 つの軸を除くすべての軸を固定することで新しいサンプル点を求め、各軸に対して同様の処理を繰り返す [18]。まず初期値を設定し、条件付き確率に基づいて一つずつ軸に沿って値を更新する。サンプリングしたい分布が多変量正規分布の場合、ギブスサンプリングの座標の更新に用いられる条件付き確率はガウス分布になることが知られている。本研究では顔分布を多変量正規分布として扱っているため、効率的にサンプル点を生成することができる。また、サンプリング初期の段階で生成されたサンプルは初期値に近い値のため、初期値への依存を減らす



ための猶予期間としてバーンイン（Burn-in）を設ける必要がある．本研究では 10000 サンプルのバーンインを設けている．図 3.6 に 3 次元の場合のギブスサンプリングの例を示す．

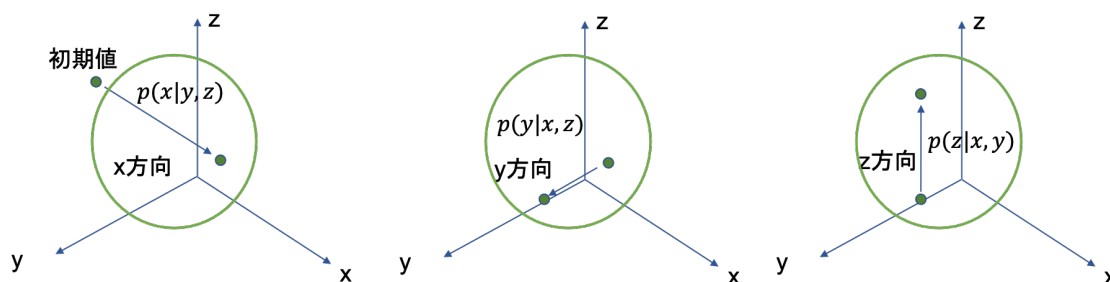


図 3.6: 3 次元のギブスサンプリング

### 3.3.2 次元復元

顔分布からサンプリングされたサンプルは主成分空間中にあるため， $W_+$  空間に写像を計算する必要がある． $W_+$  空間の写像であれば Synthesis Network に入力可能な形式で表現することが可能になる．主成分軸は  $W_+$  空間上に定義される軸であるため，主成分空間からの写像を容易に計算することができる．本提案手法では，次元圧縮に用いた主成分軸情報から  $W_+$  空間への基底変換による写像を取得することで次元復元を行う．次元復元と次元圧縮の流れを図 3.7 に示す．

本提案手法では顔分布を主成分空間上で定義しているため，顔分布からサンプリングされたサンプルからそのまま StyleGAN2 の Synthesis Network に入力することはできない．主成分空間から StyleGAN2 の  $W_+$  空間への投影が必要である．主成分軸は  $W_+$  空間上に定義される軸であるため，主成分空間からの写像を容易に計算することができる．本提案手法では，次元圧縮に用いた主成分軸情報から  $W_+$  空間への基底変換による写像を取得することで次元復元を行う．次元復元と次元圧縮の流れを図 3.7 に示す．

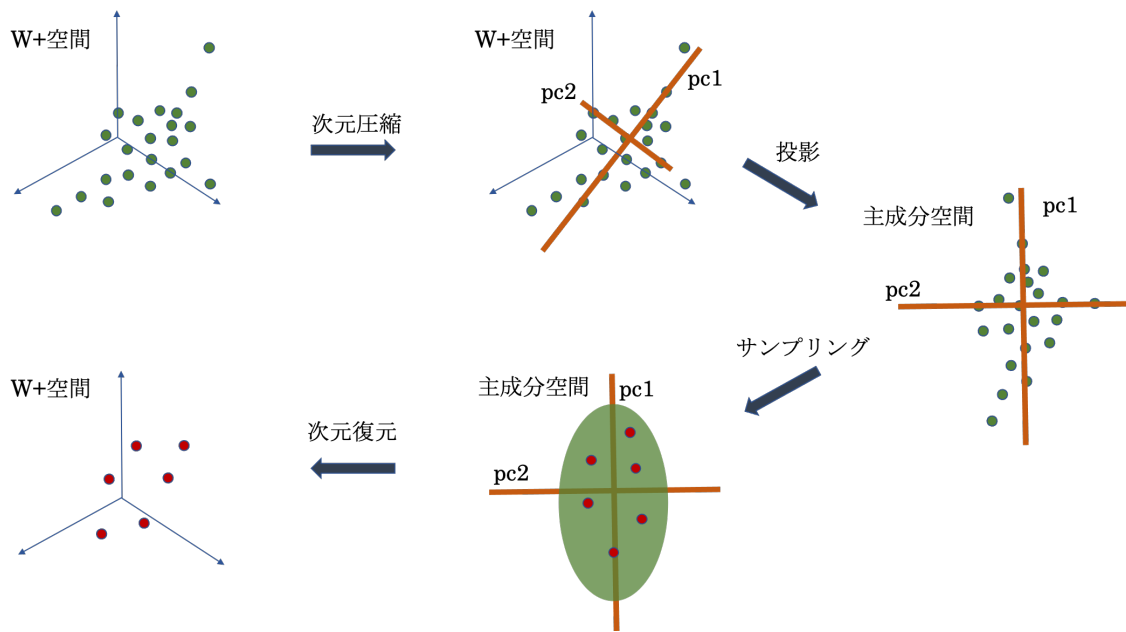


図 3.7: 次元圧縮と次元復元の流れ

### 3.3.3 StyleGAN2 による顔画像生成

次元復元により  $W_+$  空間に投影されたサンプル潜在変数を StyleGAN2 の Synthesis Network に入力することで顔サンプルを生成する。2.4.2 項で述べたように、 $W_+$  空間は Synthesis Network の各層に入力する  $W$  空間の潜在変数を統合したものである。したがって図 3.8 に示すように、サンプル潜在変数も同様に  $W$  空間の 18 個の潜在変数に分割されて生成に利用される。

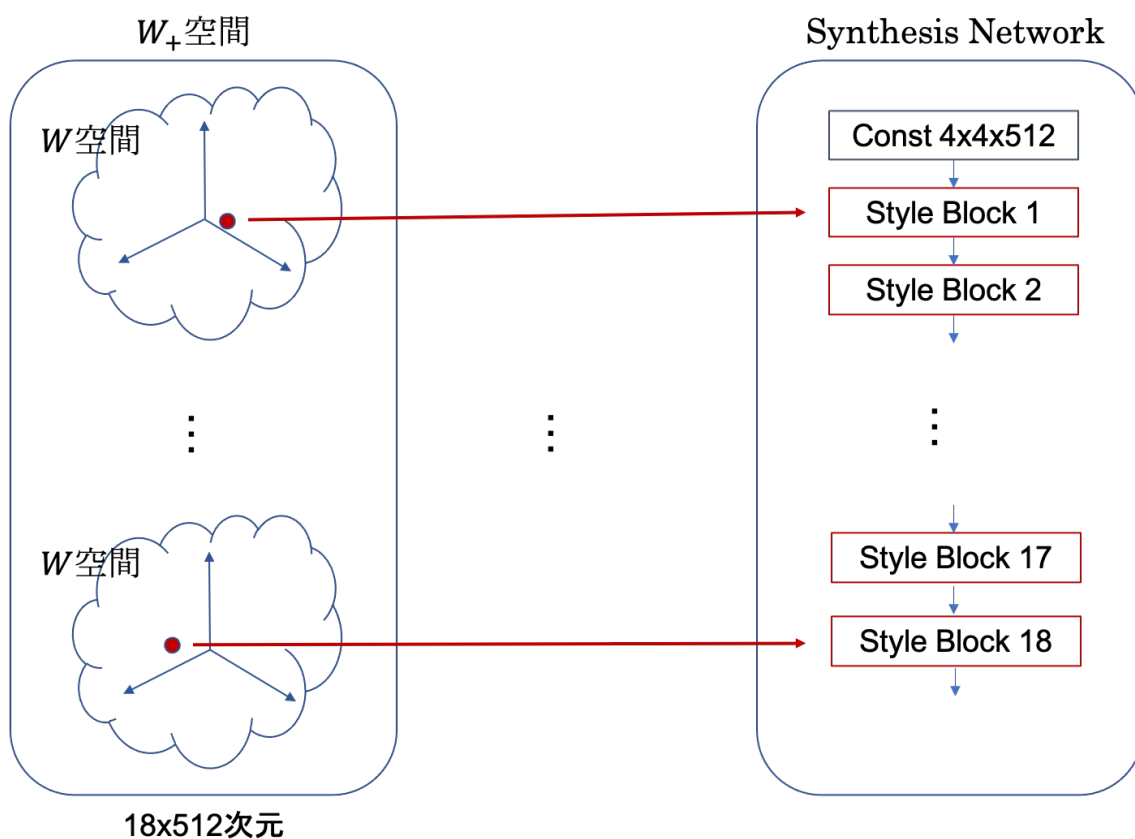


図 3.8: サンプル潜在変数の Synthesis Network への入力

### 3.4 本章のまとめ

本章では、深層生成モデルである StyleGAN2 を用いた顔分布の計算手法とその顔分布から得られる顔サンプルの生成手法について述べた。潜在変数空間においては主成分分析による集団を特徴づける属性軸の抽出を行うことで分布の次元数を下げ、球面集中現象を軽減する。こうすることで、主成分軸に対してのみを分布を計算することで効率的な顔分布を獲得することができる。次章では、提案手法によって得られる顔分布と顔サンプルに関する実験結果を紹介する。次元圧縮に用いる主成分数はチューニング可能なパラメータであるため、変化させた場合の生成結果への影響を報告する。

## 第4章 実験

### 4.1 本章の概要

本章では、前章で述べた提案手法の有効性を確かめるために行なった実験結果について述べる。まず、利用したデータセットやその収集方法について説明する。次に、データセットの顔画像群から得られる顔分布の計算結果と、顔分布同士の分離可能性を検証するための実験結果について述べる。それから、深層生成モデルから得られる平均顔や、顔分布から得られる顔サンプルの生成結果をチューニング可能なパラメータごとに紹介する。最後に、本研究の前提でもある潜在変数空間中の属性軸に関して、InterFaceGAN を StyleGAN2 で再現した結果を用いて説明する。

### 4.2 データセット

本研究では、実験のために3つの顔グループを用意した。顔グループはいずれも日本人で構成し、グループ内での性別と職業属性を統一した。また、顔グループ内での人物の重複はなく、一人につき一枚の顔画像が含まれる。以下に本件研究で作成した顔グループの詳細を示す。

日本人の男性総合家 114名

日本人の男性アイドル 103名

日本人の女性アイドル 104名

### 4.3 顔画像埋め込み結果

本節では、3.2.1項にて述べた顔画像の  $W_+$  空間への埋め込み技術を用いて得られたグループごとの潜在変数群について述べる。本実験では、顔グループごとに  $W_+$  空間上で異なる分布になることを視覚的に示し、 $W_+$  空間上での主成分による分離可能性への影響についてクラスタリング手法を用いて検証を行った結果を紹介する。

#### 4.3.1 潜在変数の散布図

実験ではデータセット中の全グループの顔画像群を  $W_+$  空間に投影し、3次元の主成分分析で可視化を行った。図4.1に示すように、顔グループごとに異なる分布を持っていることがわかった。また、図4.2からわかるように、第一主成分(PC1)のみで3つの顔グループのほとんどを区別可能である。また、第二主成分(PC2)は全グループの共通成分であることも読み取れる。図4.2では顔グループ間の境界が曖昧であるが、図4.3では境界が明瞭になることから、第三主成分(PC3)まででおおむね顔グループ間の区別が可能になると言える。

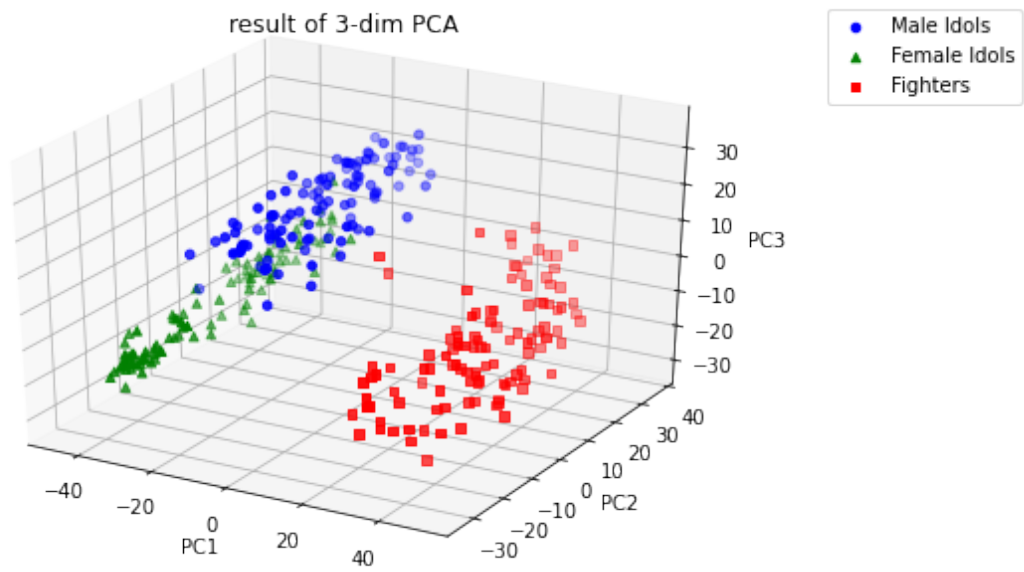


図 4.1: 3次元の主成分空間上の潜在変数群の散布

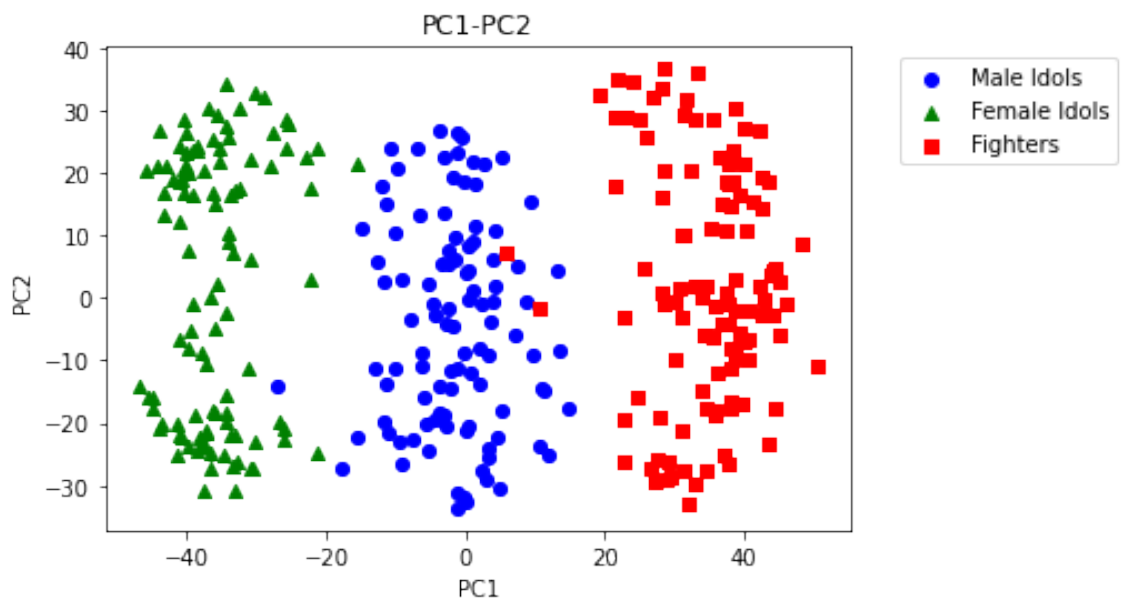


図 4.2: 第一主成分と第二主成分で構成される平面

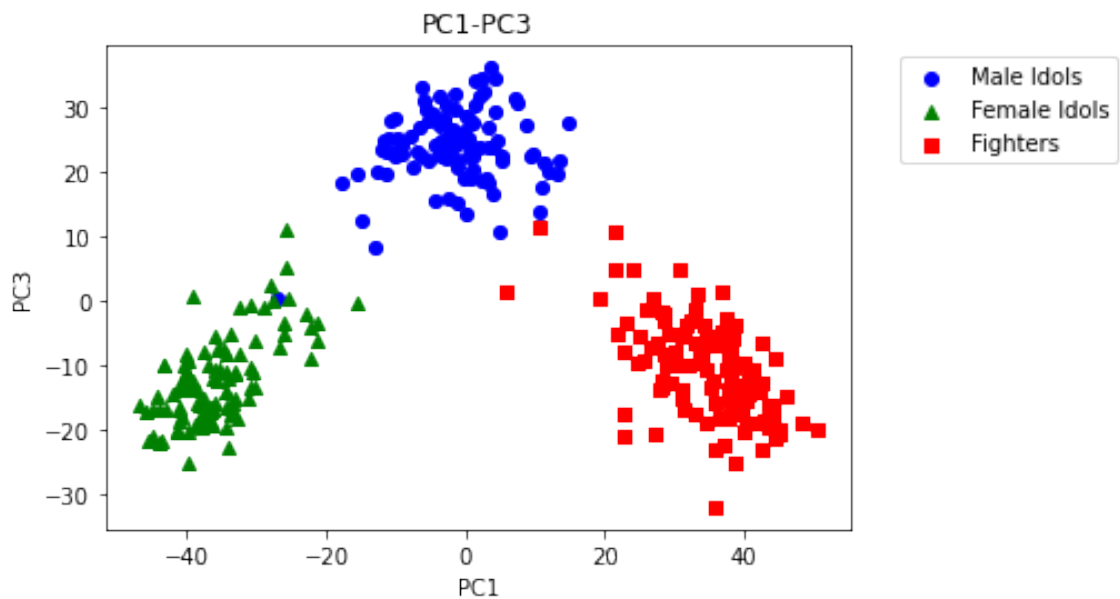
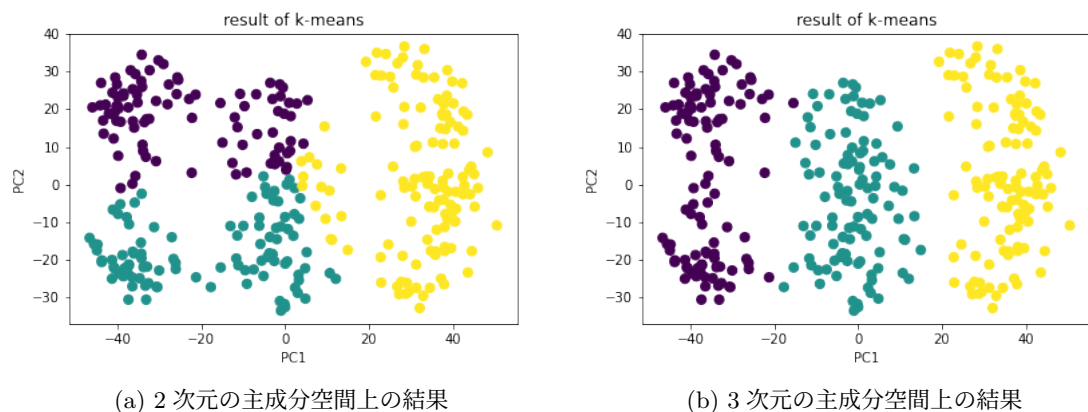


図 4.3: 第一主成分と第三主成分で構成される平面

#### 4.3.2 顔グループ間の分離可能性

4.3.1 項では 3 次元の散布図を用いてグループ間の潜在変数群の分布を可視化した。本研究では、より定量的に顔グループ間の分離可能性を検討するために、主成分空間において k-means 法によるクラスタリングを行い精度を計算する。評価には Accuracy, Recall, Precision, F1-Score を用いた。二次元 (PC1, PC2) と三次元 (PC1, PC2, PC3) の主成分空間における k-means 法によるクラスタリングの結果を二次元空間上 (PC1, PC2) で可視化した結果を図 4.4 に示す。前節でも述べたように、第三主成分を除外すると顔グループ間の境界が曖昧になり、特に男性アイドルと女性アイドルの分離精度が著しく低下する。図 4.4 と表 4.1 からわかるように、第三主成分を除外するとクラスタリング精度が低下してしまう。



(a) 2次元の主成分空間上の結果

(b) 3次元の主成分空間上の結果

図 4.4: k-means クラスタリングの計算結果の可視化

表 4.1: k-means クラスタリングの評価結果

評価手法	二次元空間			三次元空間		
	女性アイドル	男性アイドル	男性格闘家	女性アイドル	男性アイドル	男性格闘家
Accuracy	0.710			0.990		
Recall	0.552	0.548	1.000	0.982	1.000	0.990
Precision	0.630	0.548	0.897	1.000	0.990	0.980
F1-Score	0.588	0.548	0.946	0.991	0.995	0.985

#### 4.4 潜在変数空間中の顔分布間の関連性

4.3.1 項では、Image2StyleGAN を用いて  $W_+$  空間に埋め込まれた各顔グループの潜在変数群を低次元の主成分分析を用いて可視化した。図 4.2 と図 4.3 に示したように、それぞれの顔グループの潜在変数群は第二主成分方向に共通成分を持ち、第二主成分と第三主成分で分離されるような分布を示した。本研究で用いた 3 つの顔グループにおいては、第三主成分までで k-means 法によるクラスタリングを用いて 99.0% の精度で分離が可能であった。このことから、圧縮前の  $W_+$  空間における顔グループの分離可能性の高さは十分であると考えられる。また、それぞれの主成分軸方向が持つ意味に関してはその主成分方向に潜在変数を移動することで考察することが可能である。平均顔から第一成分方向に潜在変数を遷移させると生成画像は図 4.5 のように変化する。負の方向へ遷移すると性別が変化しており、正の方向へ遷移すると男性的特徴が増すことがわかる。平均顔から第二主成分方向へ遷移させると生成画像は図 4.6 のように変化する。第二主成分は全顔グループ間で共通の成分であり、顔の向きに関する情報を持っていることがわかる。平均顔から第三主成分方向へ遷移させると生成画像は図 4.7 のように変化する。男性アイドルのみが性別が変化しているように見えることから、第三主成分は第一主成分と組み合わせさせて性別を変化させる情報を持っていると推測できる。4.3.2 項でも述べたように、第三主成分が除外されると男性アイドルと女性アイドルの分離が難しくなることから、第一主成分と第三主成分の合成軸によって年齢属性軸を表していると推測できる。

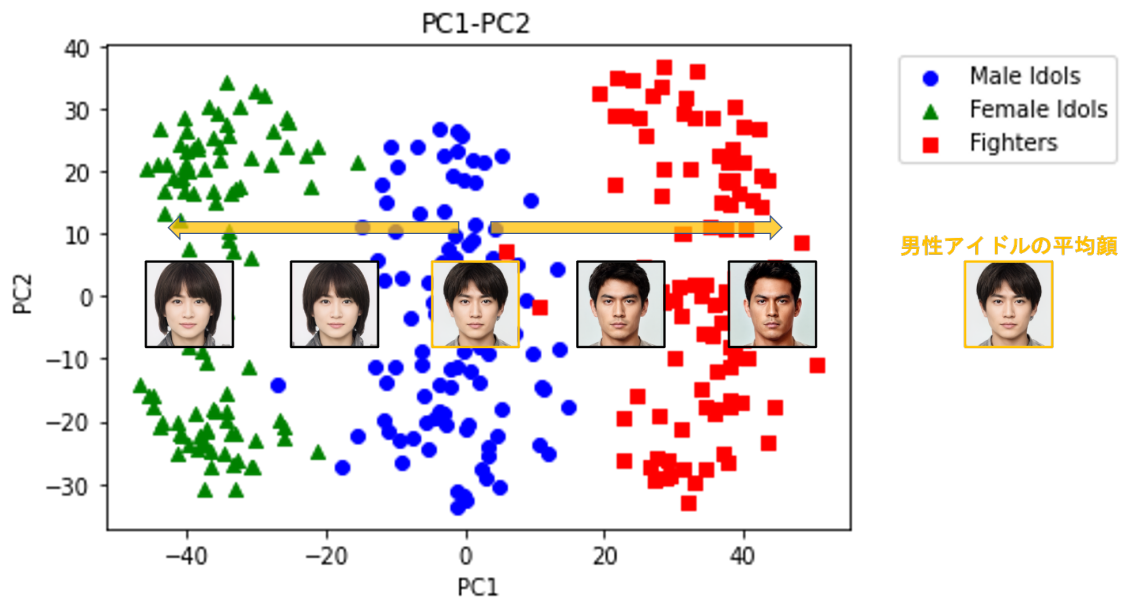


図 4.5: 第一主成分方向への遷移による生成画像の変化

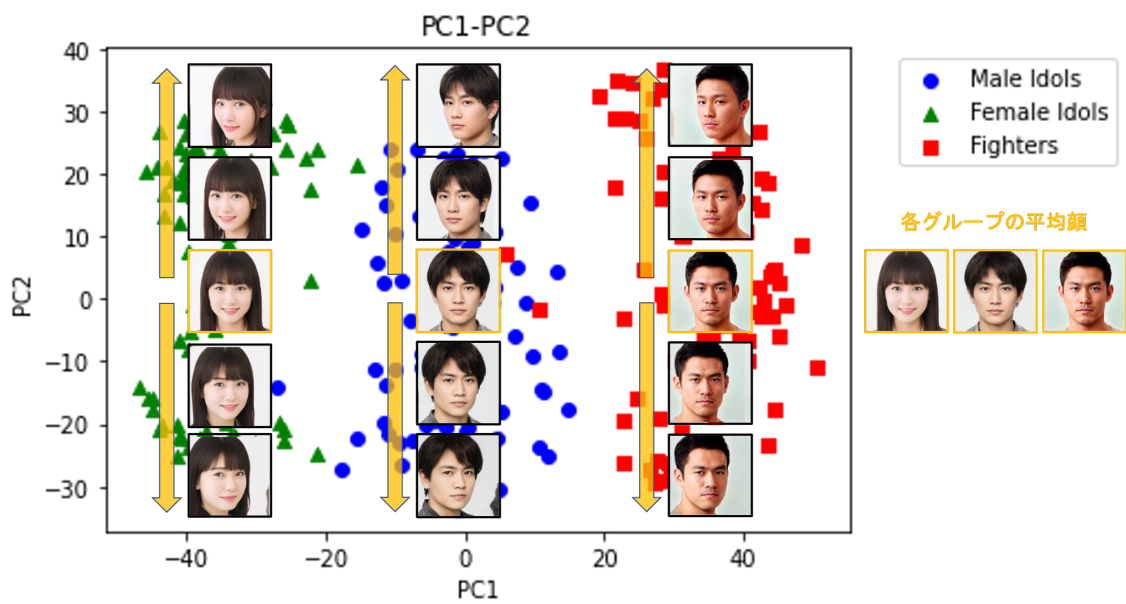


図 4.6: 第二主成分方向への遷移による生成画像の変化



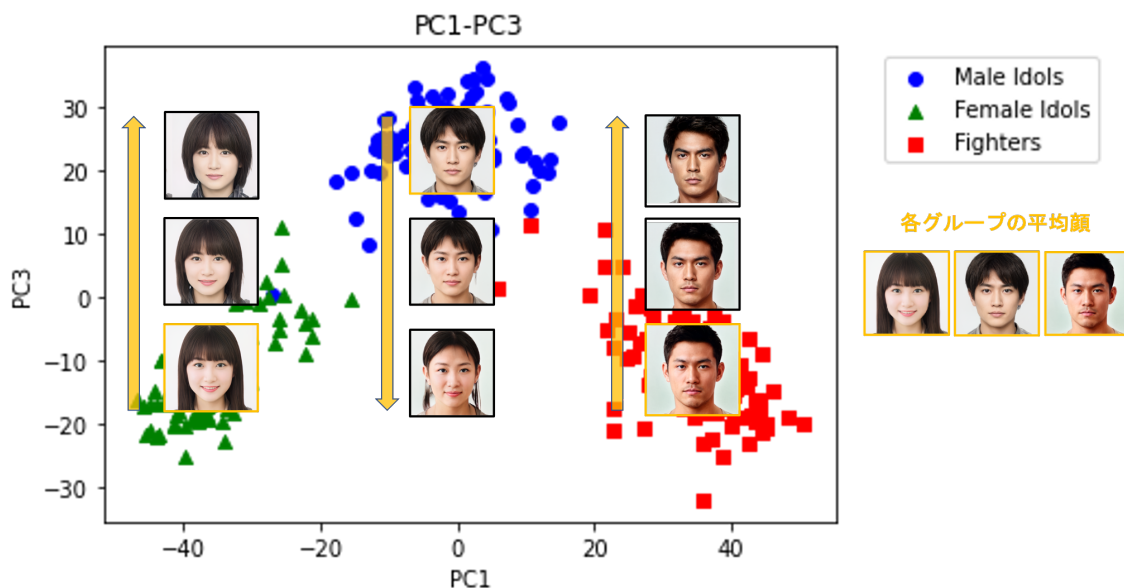


図 4.7: 第三主成分方向への遷移による生成画像の変化

## 4.5 顔サンプル生成結果

本節では、実験で得られた顔分布から生成できる顔サンプルの生成結果を示す。まず、顔が集団として持つ特徴を表す平均顔について、従来の重ね合わせ法を用いる方法と提案手法の潜在変数空間中で平均化する方法のそれぞれの生成結果を比較する。次に、主成分分析による次元圧縮に用いる主成分数と生成画像の関係を明らかにするために、主成分の累積寄与率ごとの生成結果を比較する。最後に、データセットのサンプルサイズごとの生成結果についても述べ、必要なデータ量について検討する。

### 4.5.1 平均顔

顔分布の平均  $\mu$  から生成した画像は、潜在変数空間中で定義される平均顔と言える。本項では、本手法における平均顔が従来の重ね合わせ法と同様に、顔の集団的特徴を示すかどうかを確認する。重ね合わせ法 [4] による平均顔と提案手法の深層生成モデルによる平均顔の比較を図 4.8 に示す。重ね合わせ法による平均顔は画像間の平均化処理を行うため画像全体のぼやけが発生する一方で、深層生成モデルから得られる平均顔は平均化処理を潜在変数空間中で行うためぼやけが発生せず鮮明な顔画像を生成することができた。

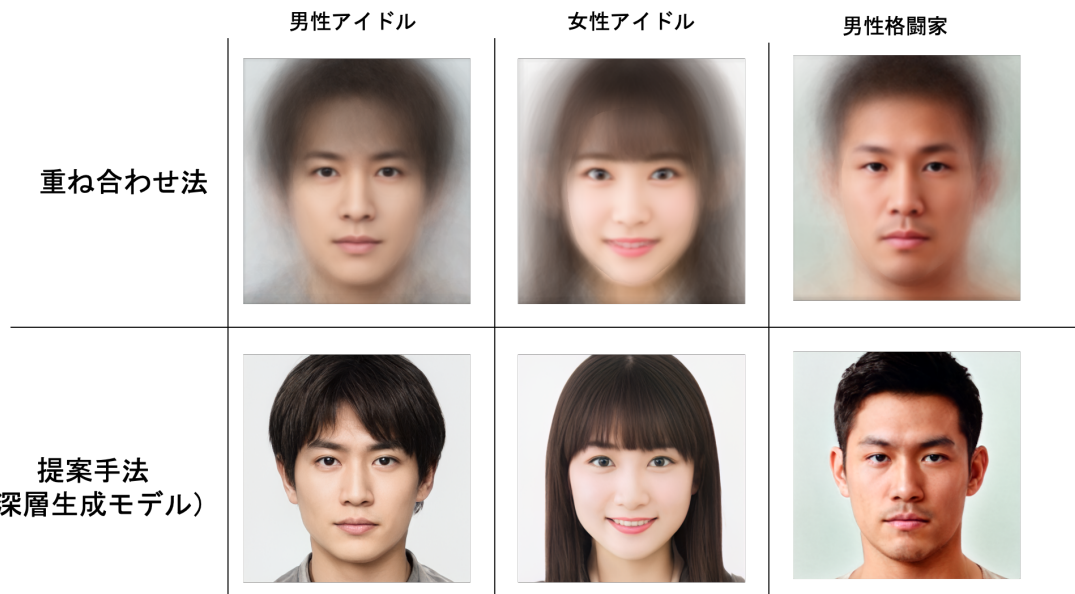


図 4.8: 重ね合わせ法による平均顔と提案手法による平均顔の比較

どちらの手法で生成した平均顔も、同じ顔グループであれば同様の特徴をとらえることが出来ている。男性アイドルは鼻が高く、鼻筋が通っている。女性アイドルは目が大きく、歯を見せて笑顔になっており、口角が上がっている。男性格闘家は重ね合わせ法でも消えることがなかった口髭がどちらにもあり、鋭い眼光と彫りの深さによる顔の影があり、口角は下を向き口が固く閉じられている。これらの従来法と同様の特徴を捉えることから、提案手法による平均顔は集団固有の特徴を捉えることができると考えられる。

重ね合わせ法では顔の特徴に分散が少ない領域では鮮明な画像が生成される一方で、顔の輪郭や髪型は分散が大きいためぼやけさせることでしか平均化ができない。提案手法においては、画像間の平均化を用いず意味的な属性値の平均化をするという特徴から輪郭と髪型にはボケが発生せず、特定の形状が再現出来るようになった。ぼらつきの大きな部位のボケのない平均化に成功したと言える。

#### 4.5.2 累積寄与率および主成分数ごとの生成結果

本項では第3章で述べた手法を用いて顔サンプルを生成した結果を次元圧縮時の主成分分析の累積寄与率ごとにまとめる。図 4.9 に示すように、累積寄与率が低いと平均値に近い顔サンプルのみが生成される一方で、累積寄与率が高いと平均とは異なるバリエーションに富んだ画像が生成される。また、累積寄与率が大きくなるにつれて図 4.10 に示すように顔が絵画のようになってしまったり、肌のテクスチャ情報に違和感のある顔サンプルが生成されるという現象も見られ、より小さな累積寄与率の顔サンプルに比べて鮮明さが失われることがわかった。生成結果に対する主観的評価にとどまるが、本研究では累積寄与率には顔サンプルにエラーが出ない範囲で最も高い値であった 0.25 を仮の最適な累積寄与率として以降の実験を行うこととした。

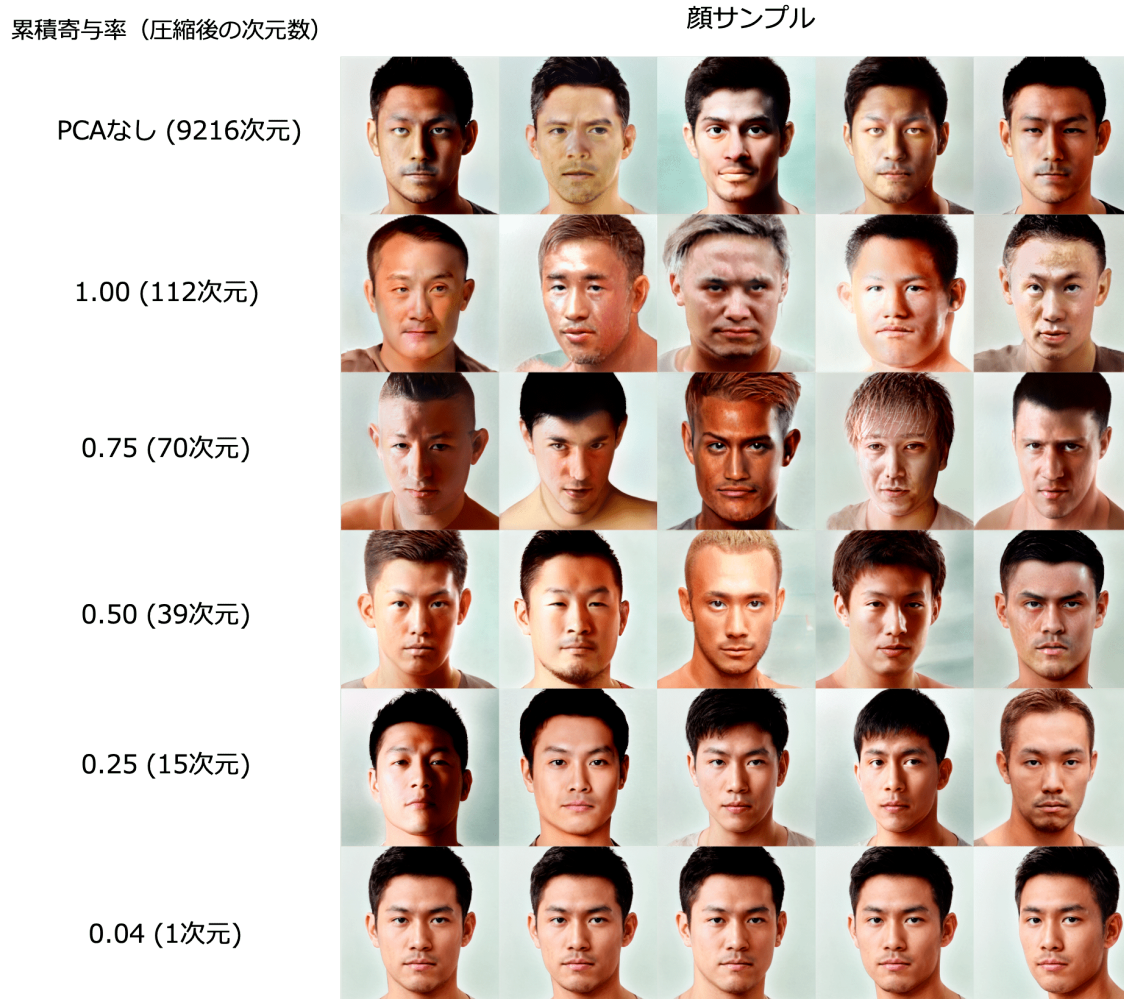


図 4.9: 累積寄与率ごとの顔サンプル生成結果

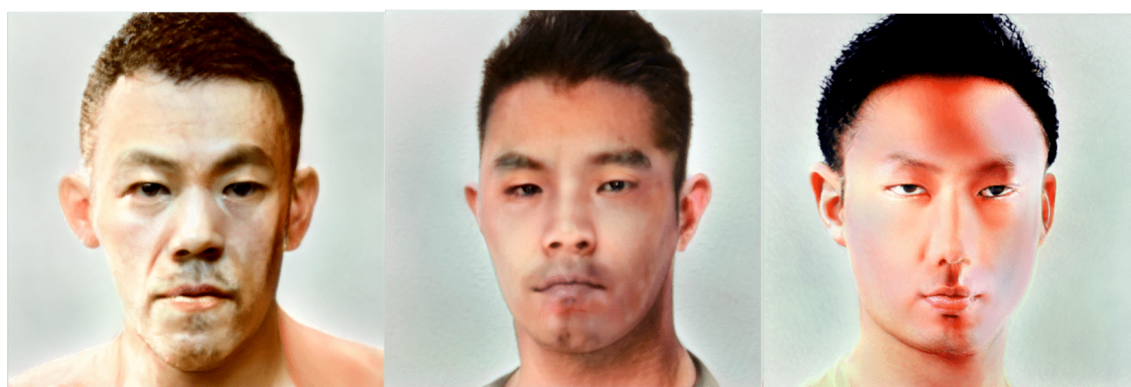


図 4.10: 累積寄与率 1.0 (112 次元) の主成分空間からのエラーサンプル例

3.2.2 項でも述べたように、主成分分析を行わずに高次元空間上で多変量正規分布からサンプリングを行うと、サンプルが超球面上に分布してしまい平均と離れたサンプルしか生成できなくなってしまう。その影響は顔サンプルにも現れており、図 4.11 に示すように主成分分析を介さず  $W_+$  空間から直接サンプリングを行ってしまうと平均から一定の距離だ



け離れたノルムの期待値付近に集中してしまい、画像の鮮明さも画像のバリエーションも中途半端なサンプルが生成されてしまう。



図 4.11:  $W_+$  空間中でサンプリングを行った場合の顔サンプル例

#### 4.5.3 データセットの画像枚数ごとの生成結果

データセットの画像枚数ごとの生成結果を図 4.12 に示す. 本研究では, データセットのサイズに対して不変なエラーの少ない累積寄与率として 0.25 を仮定して利用した. データセットの画像枚数が増加するにつれて, 顔サンプルのバリエーションは増加した. 一方で, 画像枚数が変化しても生成画像の鮮明さを損なうことがなかったため, 顔サンプルのテクスチャ情報や鮮明さがサンプルサイズに対してロバストな性質を持つことがわかった. 本手法では, 生成前に顔の集団的特徴を潜在変数空間中の顔分布に変換するため, 分布を構成する画像枚数が減ったとしても分散が小さくなるのみで画像の品質自体には影響しないと考えられる.

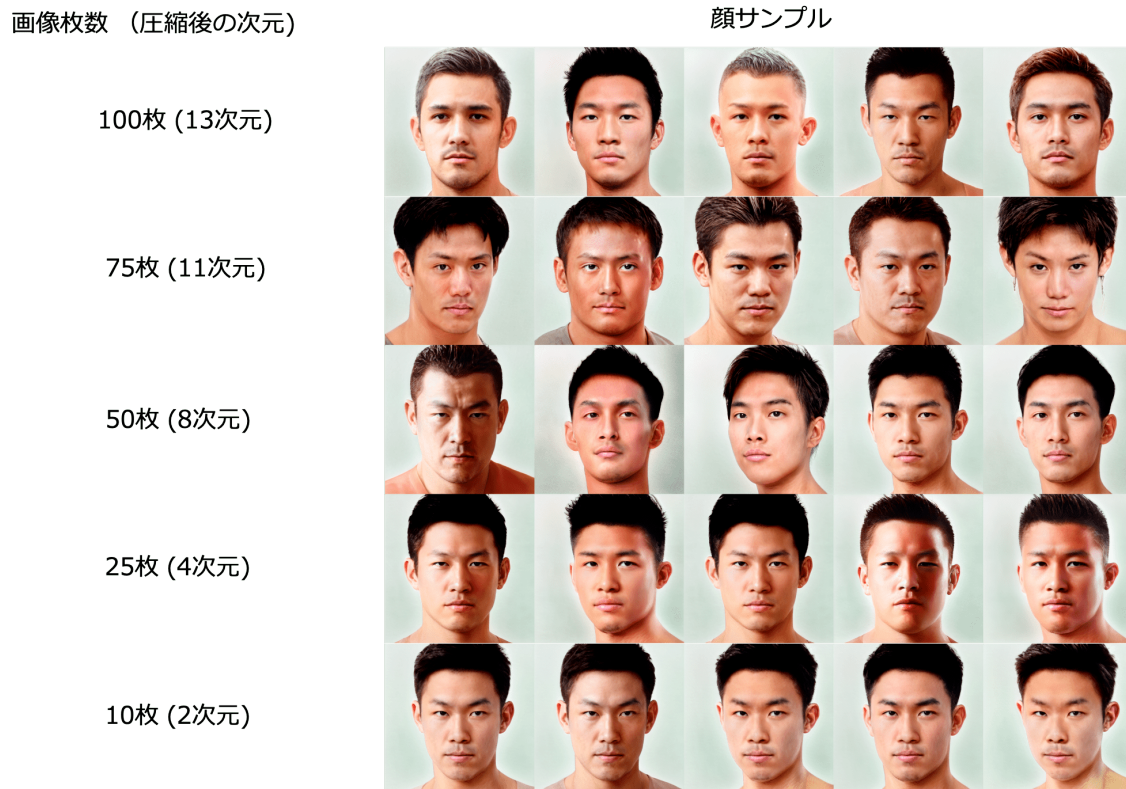


図 4.12: データセットの枚数ごとの顔サンプル生成例 (累積寄与率 0.25)

## 4.6 StyleGAN2 における InterFaceGAN の追実装

本研究では、深層生成モデルの潜在変数空間に属性変換を線形ベクトルで表せるような属性軸が存在することを前提としている。Shen らの InterFaceGAN では StyleGAN を用いて実験を行って属性軸の存在を確認したが、StyleGAN2 を用いた実験はまだ報告されていない [2]。そこで、本研究では生成に用いる StyleGAN2 の潜在変数空間中に属性軸が存在することを InterFaceGAN の再現実験を行うことで確認した。

2.4.1 項でも述べたように、InterFaceGAN では大量のサンプルをランダムに生成し、その生成画像に対して属性スコアを付与する必要がある。本研究では、サンプルの属性スコアリングには顔の属性認識モデルである Anzalone らの学習済みモデルを用いた [19]。Anzalone らのモデルでは、表 4.2 に示す CelebA データセット [20] にて顔画像に付与されている 37 種類の属性値を 90.95% の精度で推定することができる。本研究では、10000 件のランダムな潜在変数から生成された顔画像に対して属性スコアリングを行い、潜在変数空間中で線形 SVM を用いて分離超平面を推定した。推定された分離超平面の法線ベクトルを属性軸として採用し、その方向へ潜在変数を遷移させた際に生成画像に現れる変化を観察した。その結果、図 4.13 に示すように、ほとんどの属性軸による属性変換を適切に再現できた。一方で、帽子 (Wearing Hat) やネックレス (Wearing Necklace) などの装飾品などは完全な再現は出来ておらず、ネックレスの属性軸に遷移させると女性が生成されるなどの関連属性の変化のみが現れる結果となった。

表 4.2: 顔画像に付与できるラベル一覧

CelebA で定義される属性名	日本語訳
5 O Clock Shadow	青髭
Arched Eyebrows	三日月眉
Bags Under Eyes	目の下のたるみ
Bald	禿頭
Bangs	前髪
Big Lips	大きな唇
Big Nose	大きな鼻
Black Hair	黒髪
Blond Hair	金髪
Brown Hair	茶髪
Bushy Eyebrows	ゲジゲジ眉
Chubby	太っている
Double Chin	二重顎
Eyeglasses	眼鏡
Goatee	山羊髭
Gray Hair	白髪
Heavy Makeup	厚化粧
High Cheekbones	高い頬骨
Male	男性 (性別)
Mouth Slightly Open	口が少し空いている
Mustache	口髭
Narrow Eyes	細目
No Beard	髭がない
Oval Face	瓜実顔
Pointy Nose	とがった鼻
Receding Hairline	額が後退した
Rosy Cheeks	薔薇色の頬
Sideburns	もみあげ
Smiling	笑顔
Straight Hair	ストレートヘア
Wavy Hair	癖毛, パーマ
Wearing Earrings	イヤリング
Wearing Hat	帽子
Wearing Lipstick	口紅
Wearing Necklace	ネックレス
Wearing Necktie	ネクタイ
Young	若い, 年齢



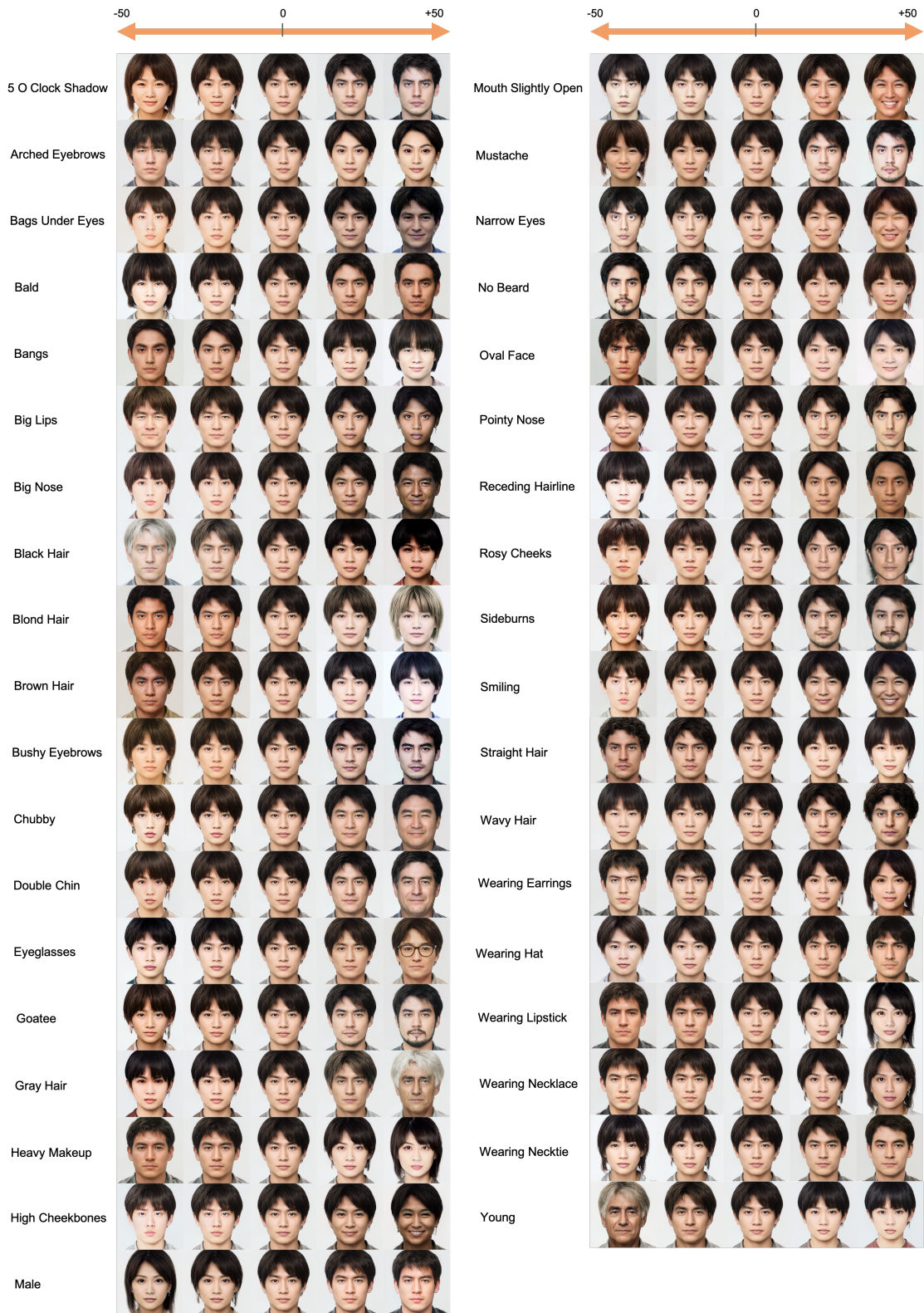


図 4.13: StyleGAN2 を用いた InterFaceGAN の追実装結果

## 4.7 本章のまとめ

本章では、本研究の提案手法による顔分布と顔サンプルの特性を知るための実験結果を示した。顔分布に関する実験では、 $W_+$  空間に埋め込まれた3つの顔グループの潜在変数群はk-means クラスタリングを用いると第三主成分までのデータを利用することで99.0%のAccuracyで分離可能であることがわかった。顔サンプルに関する実験では、累積寄与率によって顔サンプルのバリエーションと鮮明さ、顔のテクスチャ情報の品質に違いがあることがわかった。また、次元圧縮を介さない高次元多変量正規分布からの直接的なサンプリングに関しては、3.2.2項で述べたようにサンプルが超球平面上に偏ることから適切な画像生成が困難であることも確認できた。潜在変数空間中の属性軸の存在に関しても、InterFaceGANをStyleGAN2で追実装することで確認することができた。



## 第5章 考察と議論

### 5.1 本章の概要

本章では、前章で述べた実験実験をふまえて本研究から得られる知見を考察し、それに対する仮説および問題点について議論する。まず、顔分布が顔が集団として持つ特徴を捉える際に、どのような特性を示すのかを説明する。次に、顔サンプルの性質と次元圧縮による効果についてまとめる。最後に、本研究の問題点とその解決策について議論する。

### 5.2 顔分布における主成分の特性

4.4項では複数の顔分布間の関連性について述べた。本節では、単一の顔分布に対して代表的な主成分がどんな意味を持っているのかを検討する。各顔グループの平均顔から各主成分方向に遷移させた際に生成される顔画像から、その主成分が持つ意味を考察する。各顔グループの第三主成分方向までの変換結果を図5.1, 5.2, 5.3に示す。それぞれの顔グループにおける主成分には、共通の属性と集団固有の属性の2種類が含まれる。各顔グループに共通の属性としては、主に「顔の向き」「髪型」などが挙げられる。これらは現実世界でも集団によらず分散の大きな成分である。それぞれの顔グループに固有の属性としては、女性アイドルの「笑顔」や男性格闘家の「顎の引き方」などが挙げられる。女性アイドルにおいては、被写体として笑顔を求められることが多いことから複数の主成分において、色々な顔における「笑顔」属性の変換が見られたと考えられる。男性格闘家は分散が二番目に多い第二主成分において、「あごの引き方」の変換が見られた。顎を上げる行為は、弱点を見せたとしても相手には負けないという「自信」の現れを表す。また、顎を引く場合には上目遣いで相手を睨みつけるような顔になり闘争心を表す。男性格闘家においては、余裕のある自信満々な顔と闘争心をあらわにした緊迫した顔の間で大きな分散を持っていると考えられる。このように、顔分布は顔の特徴を分布化することによって集団固有の顔の特徴のばらつきを捉えることが出来ていると考えられる。したがって提案手法では、平均顔が集団で共通の特徴を抽出できることに加え、主成分方向の分散を用いて集団に固有の顔の特徴のばらつきまで再現することができる。

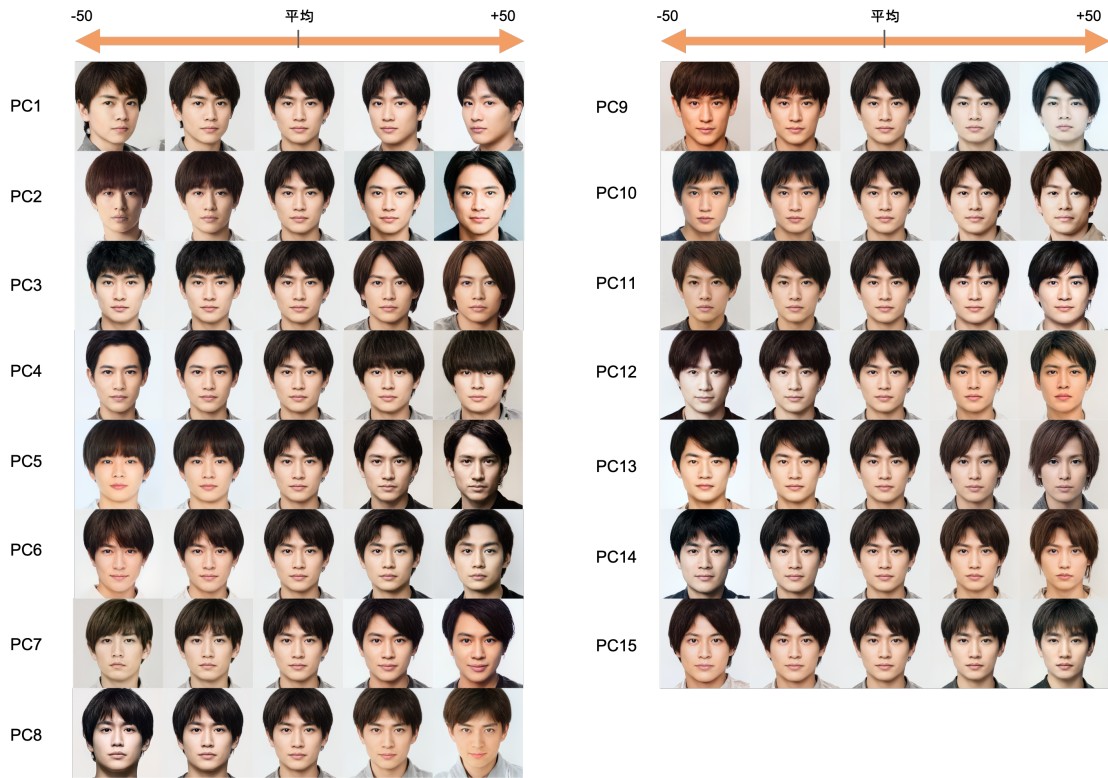


図 5.1: 男性アイドルの主成分軸方向の生成画像の変化



図 5.2: 女性アイドルの主成分軸方向の生成画像の変化

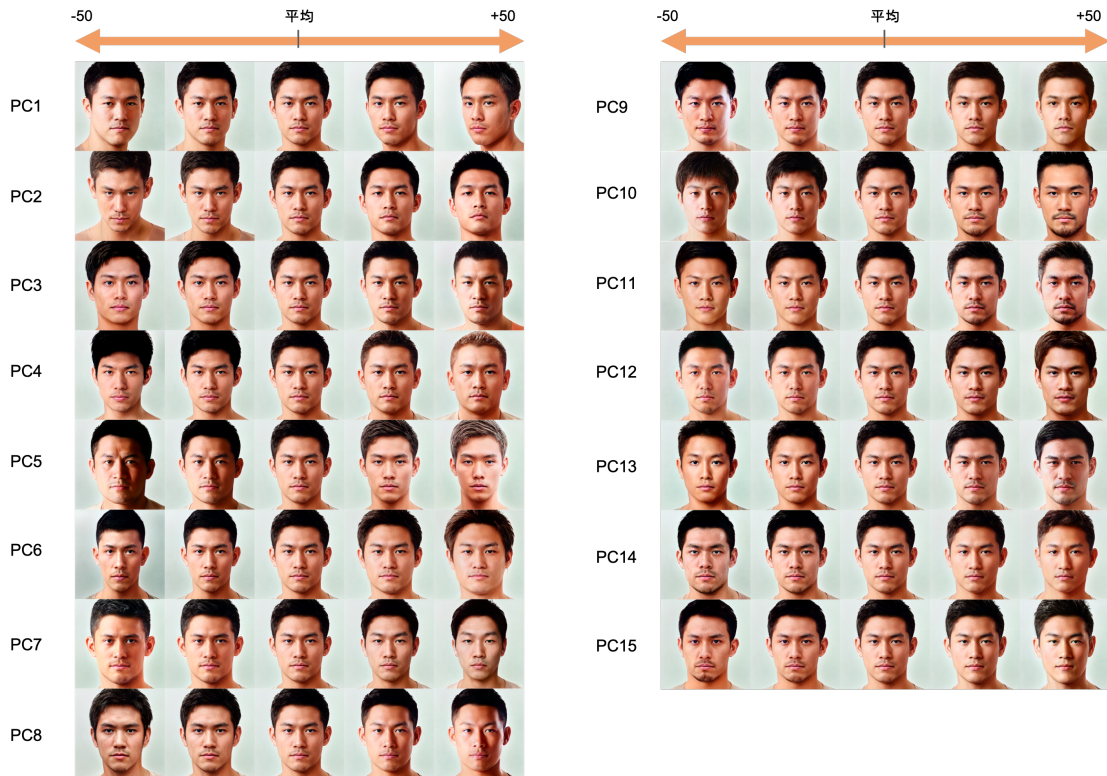


図 5.3: 男性格闘家の主成分軸方向の生成画像の変化

### 5.3 顔分布の次元数と顔サンプルの関係

3.2.2 項で述べたように、高次元の多変量正規分布からのサンプリングはサンプルの L2 ノルムが一定に集中する現象が起きる。この現象は正規分布の次元数が増えるにつれて平均付近のサンプルが生成される確率が低下し、より大きな L2 ノルムを持つサンプルばかりが生成されてしまうことが原因として考えられる。図 5.4 に、簡単のため各軸の正規分布が独立であるとして、平均付近のサンプルが生成される確率を仮に  $p = 0.4$  とすることで次元数増加による影響を示す。二次元の正規分布の場合、二つの軸どちらも平均付近の値が採用される確率は同時確率で表されるため  $p^2 = 0.16$  である。また、 $n$  次元の多変量正規分布においては  $p^n$  と一般化できることから、全ての軸上で平均付近の値が採用される確率は次元数の増加に依存して指数関数的に現象することがわかる。従って、次元圧縮を行わずに 9216 次元の  $W_+$  空間からサンプリングを行ってしまうと平均から離れた大きなノルムを持つサンプルばかりが生成されてしまう。平均顔は顔の集団的特徴を示すため、平均から大きく離れてしまうと顔が集団として持つ特徴を捉えることができなくなってしまうと考えられる。

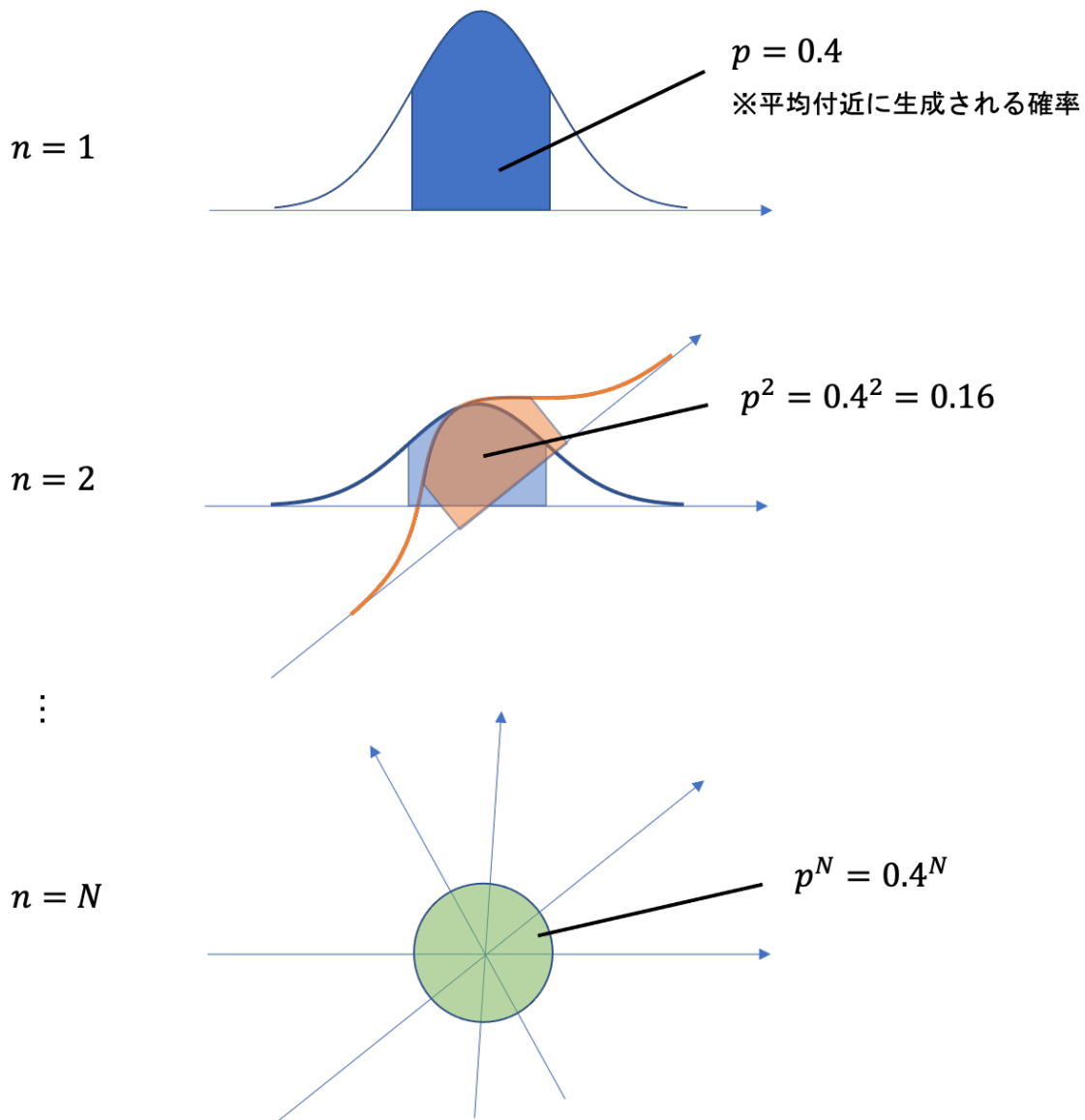


図 5.4: 次元数増加による平均付近のサンプルの生成確率の変化

顔分布のサンプルの L2 ノルムに対して、一様に生成される角度変数で極座標表示を行った結果を図 5.5 に示す。想定通り、顔分布の次元数が増加すると L2 ノルムが超球面上に分布するため極座標上でドーナツ型の分布を示す結果が得られた。4.10 や 4.11 では、顔のテクスチャ情報が損なわれるなどのエラーが見られた。これらは顔分布の次元数が大きいことからサンプルの L2 ノルムが平均から離れた位置に集中するために、 $W_+$  空間の高精細な描画範囲を外れてしまったのだと考えられる。StyleGAN2 の本来の顔画像生成プロセスでは、Mapping Network を用いた  $W$  空間への投影を介することで品質の高い描画可能領域を利用する。一方で、本研究では Mapping Network を用いずに  $W$  空間を直接利用していることから、極度に分布から距離が離れる領域では顔画像生成が正常に行われなかったと考えられる。



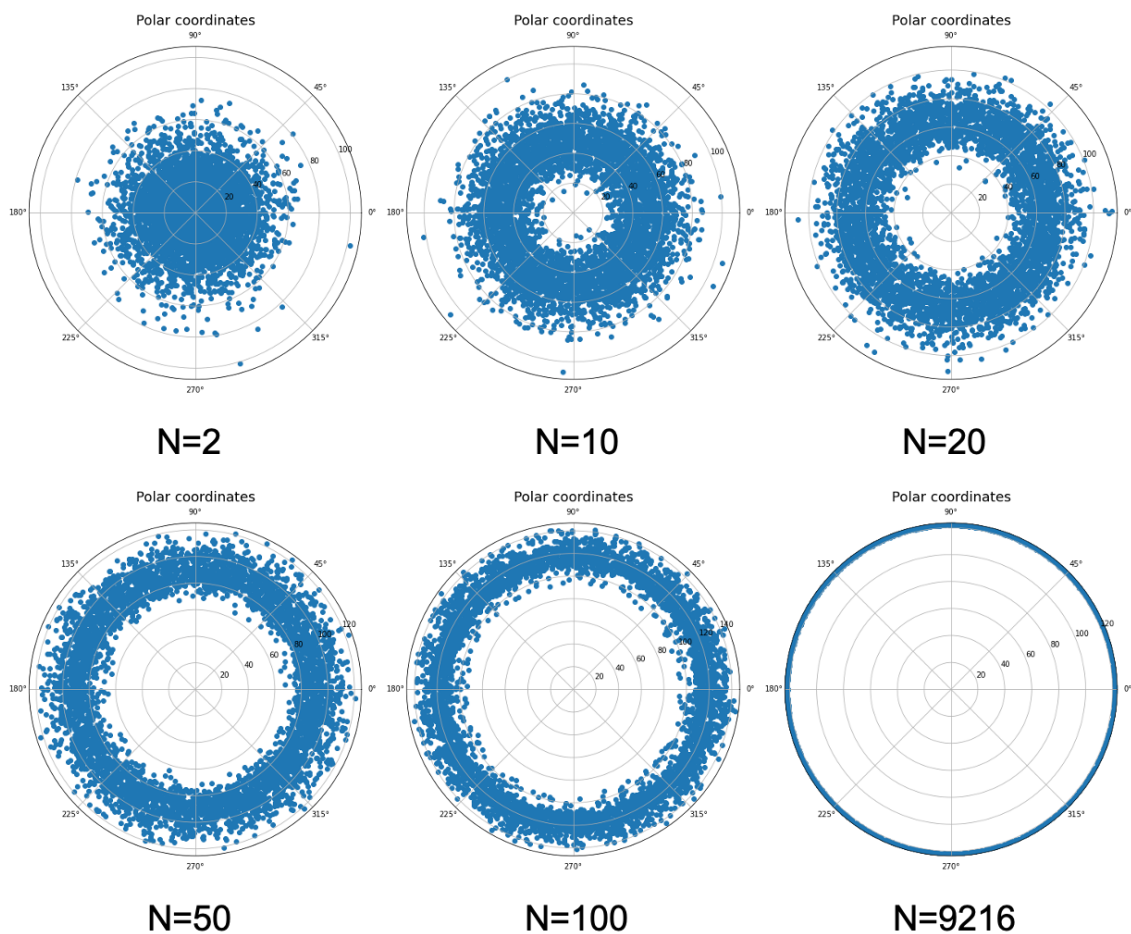


図 5.5: 男性格闘家の顔分布の次元数ごとの L2 ノルムと角度によるサンプルの極座標表示. 各 5000 件. 簡単のために, 角度は一律な一変数で表現されると仮定しランダムに生成した.

## 5.4 主成分分析における適切な累積寄与率および主成分数の選定

図 5.5 に示したように, 次元圧縮後の主成分空間の次元数に依存して顔分布は直感とは異なった振る舞いをする. 顔サンプルを生成する上では, どの次元数の主成分空間上でサンプリングを行うかによって生成される画像の性質が異なる. 顔が集団として持つ特徴を損なわず, 集団中の顔のばらつきを再現した鮮明な顔サンプルの生成のためには, 適切な累積寄与率の設定による主成分数の決定が必要になる. 本研究では, 適切な顔サンプル生成のために考えられる方策として大きく分けて以下の二つの方法を検討している.

- 累積寄与率および主成分数に対して, 顔サンプルの品質を評価する損失関数を用いて最適化処理をかける手法
- 次元数が異なる主成分空間から生成される顔サンプルの性質が異なることを利用して, 複数の主成分空間を組み合わせる方法

前者においては, 生成される顔サンプルの鮮明さとバリエーションを考慮した評価関数が必要になる. 4.5.2 項で述べた結果から, 顔サンプルに関しては累積寄与率および主成分数に依存して生成画像の鮮明さとバリエーションがトレードオフの関係にあることがわかる.

したがって、その両者を考慮した損失関数は図 5.6 のような特性を示す関数を採用する必要がある。このような特性の損失関数は勾配法等の最適化処理によって機械的に最適な解を求めることが可能であるため、顔サンプル生成システムの自動化に役立つ。後者では、図 5.5 に示した次元数増加による極座標系でのサンプルの特性の変化が手掛かりになる。正規分布に求められる振る舞いとしては、平均付近が密で外側が疎なサンプリングである。しかし、顔サンプルにばらつきを持たせるために次元数を増加させると平均付近のサンプルが生成されなくなり、平均から距離のあるサンプルばかりが生成されて顔サンプル生成が不安定になるという問題点がある。そこで、複数の主成分空間が分担して平均付近のサンプルとばらつきをもったサンプルを生成することで、擬似的に正規分布的振る舞いを再現することが可能になるはずである。実現のためには、エラーが発生し始める累積寄与率及び次元数の特定が必要である。エラーの定義がユーザにより異なるため、後者の手法は生成に幅を持たせることでユーザが定義するよりよい顔サンプルの作成に役に立つと考えられる。

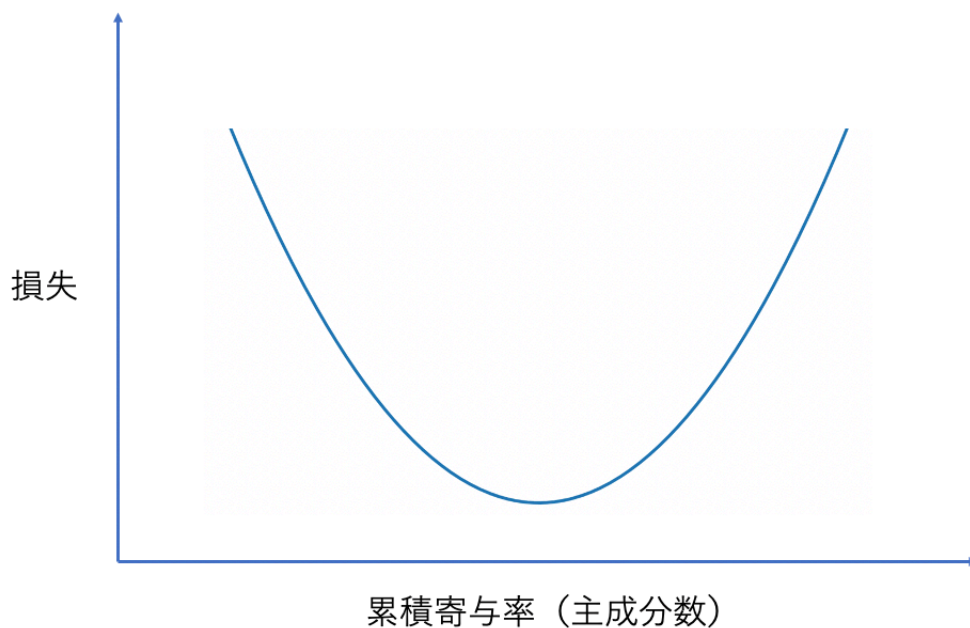


図 5.6: 最適化を用いる際の損失関数の特性

## 5.5 本章のまとめ

本章では、実験から得られた顔分布と顔サンプルに関する知見をまとめ、今後の課題となるパラメータ調整について述べた。本研究では、顔分布が定義される主成分軸はそれぞれ属性変換軸としての役割を持っており、顔グループ間で共通の属性と固有の特性にわけられることがわかった。また、高次元空間からのサンプリングにおいてサンプルの L2 ノルムの頻度分布が偏ってしまう現象により、顔サンプルの生成において次元圧縮に用いる主成分数に依存して正規分布の直感と反したサンプリングが行われるという問題点も明らかになった。本研究ではこの問題に対して、損失関数を用いた主成分数の最適化を行うことで一意に自動的に主成分数を決定する方法と、複数の主成分空間を組み合わせることで生成に幅を持たせる方法の二つを今後検討すべき解決策として紹介した。

## 第6章 まとめ

本論文では、深層生成モデルの潜在変数空間を活用した顔の分布化と顔サンプルの生成手法を提案した。顔の集団的特徴を理解し顔の違いや共通点を理解するためには、顔を何らかの統計量として扱う必要がある。本研究では深層生成モデルの潜在変数空間中で主成分分析を用いることで分布に対して寄与の大きい主成分を抽出し、その主成分空間中で分布を計算する。主成分分析を行うことで、その集団を特徴づける重要な主成分を抽出することができる。

顔分布が示す集団の特徴を理解するためには、平均値から生成される顔画像に含まれる属性と主成分が持つ属性を観察する必要がある。まず集団の間での平均値は、個別具体的な特徴のばらつきが違いに相殺されあって消えてしまうため、集団に共通の属性だけが抽出される。例えば、男性アイドルや女性アイドルの平均値では口角の上がった笑顔とも取れる表情が確認できる一方で、男性格闘家の平均値からは口角が固く結ばれた威圧的な表情が確認された。このような集団に固有で共通な特徴が平均値から分析することができる。次に、集団の間で重要な属性として抽出された主成分には、その集団において個別具体的な特徴な部分に対して分散の大きな属性だけが抽出される。例えば、アイドルには髪型に関する主成分が多く抽出されていたり、髪の色は彩度の低い範囲でのばらつきが大きいことなどが確認された。一方で、男性格闘家は彩度の高い派手な髪色に変化する主成分や髭などのアイドルとは異なる主成分が多く抽出されているのが確認された。これらのばらつきに関する情報は平均値だけでは失われてしまう情報であるため、分布として顔を理解することの重要性を示唆する結果となった。

顔分布を用いてサンプリングを行うことで、分布に従う顔サンプルを生成することができる。サンプリングを行う際には、主成分空間の次元数が生成されるサンプルの品質に影響を与える。高次元空間では分布からのサンプルのノルムが超球面上に偏る球面集中現象が発生するため、できるだけ少ない主成分数で構成される主成分空間でサンプリングを行う必要がある。一方で、主成分数を減らしてしまうと分布が捉えることができる顔の集団的特徴が減ってしまうため、生成されるサンプルが平均付近のものばかりになってしまう。この問題に対しては、本研究では損失関数を用いて最適な主成分数を最適化数で求める方法と複数の主成分空間の組み合わせを用いることの二種類を今後比較検討していく予定であり、今後の課題としたい。

今後は、本研究の課題である最適な主成分数の選定について検討を行いながら、顔サンプルを用いた顔画像データセットのオーグメンテーションや顔以外の画像を生成する深層生成モデルの潜在変数空間を用いた人間以外の顔分布の作成・分析などの本提案手法を応用した研究を進めていく。

## 謝辞

この研究を遂行するにあたり、終始暖かく見守って下さった三河正彦准教授と藤澤誠准教授に深く感謝いたします。また、学類生の加藤優一氏には私の研究に興味を持っていただき、研究に新しい発展をもたらしていただきました。本当にありがとうございました。最後になりますが、普段から励ましの言葉をかけあい共に学んだ研究室の皆様や、研究に前向きなコメントをいただいた日本顔学会の会員の皆様に感謝の意を表します。



## 参考文献

- [1] L. Sirovich, M. Kirby, "Low-dimensional procedure for the characterization of human faces," *Journal of the Optical Society of America*, 4, pp.519-524, 1987.
- [2] Yujun Shen, Ceyuan Yang, Xiaou Tang, Bolei Zhou, "InterFaceGAN: Interpreting the Disentangled Face Representation Learned by GANs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020.
- [3] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, Timo Aila, "Analyzing and Improving the Image Quality of StyleGAN," *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp.8110-8119 2019.
- [4] 向田茂, 蒲池みゆき, 尾田政臣, 加藤隆, 吉川左紀子, 赤松茂, 千原國宏, 「操作性を考慮した顔画像合成システム : FUTON : 顔認知研究のツールとしての評価」, *電子情報通信学会論文誌. A*, 10, 2002, pp.1126-1137.
- [5] 「FUTON が提供する基本合成機能」(最終閲覧日:2020年12月5日) , [https://search.ieice.org/data/d\\_data/j85-a\\_10\\_1126/data/data5/data1.htm](https://search.ieice.org/data/d_data/j85-a_10_1126/data/data5/data1.htm) .
- [6] 永田明德, 金子正秀, 原島博, 「平均顔を用いた顔印象分析」, *電子情報通信学会論文誌. J-80A*, 8, pp.1266-1272, 1997.
- [7] 永田明德, 岡崎透, 崔昌石, 原島博, 「主成分分析による顔画像の基底生成と空間記述」, *電子情報通信学会論文誌. D-2, 情報・システム. 2, パターン処理 / 電子情報通信学会編 79(7)*, p.1230-1235, 1996.
- [8] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio, "Generative adversarial nets," *Advances in Neural Information Processing Systems 27*, pp.2672-2680, 2014.
- [9] Alec Radford, Luke Metz, Soumith Chintala, "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks," *International Conference on Learning Representations*, 2016.
- [10] Tero Karras, Timo Aila, Samuli Laine, Jaakko Lehtinen, "Progressive Growing of GANs for Improved Quality, Stability, and Variation," *International Conference on Learning Representations*, 2018.
- [11] Tero Karras, Samuli Laine, Timo Aila, "A Style-Based Generator Architecture for Generative Adversarial Networks," *Conference on Computer Vision and Pattern Recognition*, 2018.
- [12] Xun Huang, Serge Belongie, "Arbitrary Style Transfer in Real-time with Adaptive Instance Normalization," *International Conference on Computer Vision*, 2017.

- [13] Rameen Abdal, Yipeng Qin, Peter Wonka, "Image2StyleGAN: How to Embed Images Into the StyleGAN Latent Space?," International Conference on Computer Vision, 2019.
- [14] Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, Oliver Wang, "The unreasonable effectiveness of deep features as a perceptual metric," Conference on Computer Vision and Pattern Recognition, 2018.
- [15] Diederik P. Kingma, Jimmy Ba, "Adam: A Method for Stochastic Optimization," International Conference on Learning Representations, 2015.
- [16] R. Vershynin, "High-Dimensional Probability: An Introduction with Applications in Data Science," Cambridge: Cambridge University Press, doi:10.1017/9781108231596, 2018.
- [17] Robert, Christian and Casella, George, "A Short History of Markov Chain Monte Carlo: Subjective Recollections from Incomplete Data," Institute of Mathematical Statistics, Statistical Science, vol. 26, no. 1 pp.102-115, 2011.
- [18] S. Geman and D. Geman, "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-6, no. 6, pp. 721-741, 1984.
- [19] Anzalone, Luca and Barra, Paola and Barra, Silvio and Narducci, Fabio and Nappi, Michele, "Transfer Learning for Facial Attributes Prediction and Clustering," Springer, International Conference on Smart City and Informatization, pp.105-117, 2019.
- [20] Liu, Ziwei and Luo, Ping and Wang, Xiaogang and Tang, Xiaoou, "Deep Learning Face Attributes in the Wild," International Conference on Computer Vision, 2015.