論　文

重み付き条件を用いたGenerative Adversarial Networksによる認識に有効な顔画像生成

足立　浩規[†a)]　　平川　翼[†b)]　　山下　隆義[†c)]　　藤吉　弘亘[†d)]

Effective Facial Image Generation for Recognition by Generative Adversarial Networks Using Weighted Conditions

Hiroki ADACHI[†a)], Tsubasa HIRAKAWA[†b)], Takayoshi YAMASHITA[†c)], and Hironobu FUJIYOSHI[†d)]

あらまし　本研究では，認識に有効な顔画像生成を目的とし，重み付き条件によって段階的な条件反映が可能な conditional Generative Adversarial Network (cGAN) を提案する．cGAN の応用手法は，Discriminator に対する条件の与え方を改善することで高品質な画像生成を達成しているが，Generator に対する条件の入力方法は従来手法からほとんど進展がない．しかしながら，従来の方法では深いネットワーク設計のとき，出力層付近で条件が消失することが報告されている．この問題に対処するために，全層に条件を与える手法が提案されているが，我々は顔画像生成に焦点をおいているため改善が必要である．一般に，顔画像は 1 サンプルに対して複数の属性が付与されるため，各属性に適した層があると考える．そこで，我々は畳み込み処理を用いて各属性に重み付けをする Weighted conditional layer (Wc-layer) を導入する．Wc-layer は各属性の最適な位置だけでなく，その層において最適な強度で属性を反映できる．実験で，Wc-layer を用いることの優位性を定量的及び定性的に示す．

キーワード　敵対的画像生成，重み付き条件，画像生成，クラス識別

1. まえがき

画像内に映る人物がどのような属性（性別，髪色など）であるかを認識及び理解することは，顔画像から人物を特定するようなセキュリティシステム開発のために重要な問題である．近年では，深層学習の急速な発展によって高精度な認識を実現している．しかしながら，学習に利用するデータの属性に偏りが大きい場合，少量サンプルの属性の認識率は顕著に低下することが知られている．人手によってサンプル数を増幅することで性能向上は期待できるが，データの収集やアノテーションは非常に高い人的，時間的コストを要する．

深層生成モデル [1]〜[4] は，データの生成が可能であるため，人手を必要としないデータ増幅への利用が期待できる．特に，近年，最も注目を集めている手法が，Goodfellow らによって提案された Generative Adversarial Networks (GAN) [4] である．GAN は，データを生成する Generator と入力データを識別する Discriminator の二つを競合させながら学習する手法である．GAN は潜在変数から画像を生成するものと，潜在変数と条件（クラスラベル，文章など）から意図した画像を生成するものの二つに大別できる．前者は先進的に研究がされており，実画像と大差ないクォリティの画像生成が可能な手法が数多く提案されている [5]〜[16]．更に，潜在変数の代わりに画像を入力することで，画像のスタイル変換 [17]〜[21] や，超解像生成 [22] も可能である．後者は，conditional GAN (cGAN) [23] をはじめとして，幾つかの手法 [24]〜[32] が提案されているが，前者と比較すると発展途上だといえる．以降，誤解がない限り本論文では，cGAN の応用手法全てを単に cGAN と呼ぶ．本研究では，顔属性の認識率向上に貢献する顔画像を cGAN によって生成することを目的とする．

† 中部大学，春日井市
Chubu University, 1200 Matsumoto, Kasugai-shi, 487–8501 Japan
a) E-mail: ha618@mprg.cs.chubu.ac.jp
b) E-mail: hirakawa@mprg.cs.chubu.ac.jp
c) E-mail: takayoshi@isc.chubu.ac.jp
d) E-mail: fujiyoshi@isc.chubu.ac.jp
DOI:10.14923/transinfj.2021JDP7031

cGAN の Discriminator に対する条件の与え方は様々な工夫がされているにもかかわらず，Generator は潜在変数と同時に条件を与える方法が主流である．従来の Generator は入力層のみに条件を与えているため，深いネットワーク構造であるほど条件が考慮されづらくなり，出力層付近で条件消失が生じると考えられる．実際，この問題は [29], [30] で主張されており，Generator の入力層以外にも条件を与えることで問題に対処している．我々も，条件を反映した高品質な顔画像生成をするために，Generator を先行研究と同様に設計する．しかしながら，通常，顔画像は一つ以上の属性が定義されるため，ポジディブになる属性が複数個存在する．本論文では，このようなラベルをマルチラベルと呼称する．マルチラベルを用いて顔画像を生成するときは，各属性に対して適した層があると考える．この考えはヒトが似顔絵を描くときに全てのパーツを同時に描き始めるのではなく，大局的な情報から徐々に詳細な情報を描く感覚に類似している．この疑問を解消するための最も愚直な方法は，人の感覚によって与える層や強度を区別することであるが，人手による最適な位置の探索は高コストかつ厳密な基準がないため決定が困難である．そこで本研究では，条件を与える前に畳み込み処理を用いて重み付けをする Weighted conditional layer (Wc-layer) を導入する．Wc-layer は入力した属性に重み付けができるため，最適な入力位置を Generator 自身が吟味しながら画像生成をすることができる．Discriminator は，Auxiliary classifier GAN (AC-GAN) [26] と同様に Discriminator 内部で入力画像のクラス分類を行うようマルチタスク化する．以降，クラス分類を行うブランチを Recognition branch，従来と同様に実画像か生成画像か判別するブランチを Adversarial branch と呼称する．Wc-layer とマルチタスク Discriminator を組み合わせて学習することで，低解像度のときにグローバルな属性，高解像度になるにつれてローカルな属性のような段階的な反映が実現できる．

評価実験では，客観的評価及び主観的評価により提案手法の有効性を示す．基本的に，客観評価では生成画像が条件を満たしているかどうかを含めた評価はできない．そこで，被験者に画像を提示して画質及び条件を考慮した評価が可能な主観評価を利用する．更に，Convolutional neural network (CNN) で顔属性認識学習をする際に，提案手法で認識率の低い属性を生成して追加データとして利用したときの精度を調査する．しかしながら，生成画像全てが高解像度で条件を満たした画像である可能性は低いと考える．したがって，生成画像全て学習に用いた実験だけでなく，Active learning によって選別したデータを用いたときのパフォーマンスも調査する．Active learning は推論結果を基に，精度向上に貢献するデータを人手によって選択，ラベル付与を行い追加データとして学習することでモデルの性能を向上させる機械学習のテクニックである．生成画像に対して Active learning を行うことで，生成画像全てを利用した学習結果より精度が見込める．最後に考察として，Wc-layer の畳み込み処理の重みパラメータから，各層における属性の寄与率の調査をする．これにより，提案手法を用いることで段階的に条件を反映できることを示す．

2. 関 連 研 究

GAN は，任意の潜在変数から画像を生成する Generator と，実画像または生成画像のどちらが入力されたかを正確に判別する Discriminator の二つのネットワークで構成されている．Generator は Discriminator を惑わすために，リアリスティックな画像を生成することがモチベーションである．一方，Discriminator は生成画像と実画像を正確に分類することをモチベーションとしている．Generator と Discriminator を敵対的に学習することによって Variational Autoencdoer (VAE) [1] のように，任意の確率分布との近似計算なしで高精細な画像生成が可能である．GAN の目的関数は，真のデータ分布 $P_{data}(\mathbf{x})$ からサンプリングした実画像を $\mathbf{x}$，潜在変数の事前分布 $P_{\mathbf{z}}(\mathbf{z})$ からサンプリングした潜在変数 $\mathbf{z}$ を用いて Generator で生成した画像を $G(\mathbf{z})$ とすると以下に示す式で表現できる．

$$\mathcal{L}_{gan} = E_{\mathbf{x}\sim P_{data}(\mathbf{x})}[\log D(\mathbf{x})] + E_{\mathbf{z}\sim P_{\mathbf{z}}(\mathbf{z})}[\log(1 - D(G(\mathbf{z})))] \quad (1)$$

2.1 高解像度な画像生成

GAN では，ネットワーク構造を Convolutional Neural Network (CNN) をベースとして設計することによって Vanilla GAN より高解像度の画像生成が可能である．Radford らの提案した Deep Convolutional GAN (DCGAN) [5] は，Discriminator へ畳み込み処理，Generator へ逆畳み込み処理を用いることで，高解像度な画像生成を可能とした．しかしながら，DC-GAN は学習初期から所望の解像度までアップサンプリングするため，高精細に生成できる画像サイズに限界があった．Karras

ら [6] は，この問題への対処方法として段階的な画像生成をする Progressive Growing GAN (PG-GAN) を提案した．PG-GAN では，学習の進捗とともに Generator と Discriminator のネットワークの層数を追加することによって，1024×1024 [pixels] の画像生成を達成している．

高解像度な画像を高精細に生成するためには，PG-GAN のようにプログレッシブな画像生成方法が必要であるとされていたが，Brock らは独自の画像生成法により高解像度の画像生成が可能な BigGAN [9] を提案した．BigGAN は Residual Network (ResNet) [33] をベースに構築されており，サンプリングした潜在変数を分割して，条件ベクトルと合わせて任意の次元数へ投影したのちに各層の Batch Normalization のアフィンパラメータとして与えている．BigGAN では，最適な潜在変数の次元数や潜在空間として最適な確率密度関数について論じられており，120 次元の潜在変数を $U[-1,1]$ または $N(0,1)$ からサンプリングする事が最適であると明記されている．Zhang らの提案した Self-Attention GAN (SA-GAN) [7] も PG-GAN のような学習過程ではなく，Self attention を用いることで生成画像の大域的な依存関係を学習することが可能となり，高精細な画像生成を実現した手法の一つである．

また，Vanilla GAN を含めた多くの手法が Generator へ任意の潜在空間からサンプリングした潜在変数を入力する事で画像生成を行う．一方，Karras らの提案した Style-GAN [8] は，学習初期に決定した定数を入力する事で画像を生成している．サンプリングした潜在変数は，複数の全結合層で構築される Mapping network を介して異なる潜在空間へマッピングした後，Adaptive Instance Normalization (AdaIN) [34] のパラメータとして使用することで，生成画像のスタイルを操作する．Style-GAN は cGAN とは異なり，クラスラベルを条件として明示的に与えていないが，潜在変数を混ぜることで，生成画像のスタイルを意図的に変化することを可能としている．

2.2 conditional GAN

GAN を用いて意図した画像を生成するための最もシンプルな方法として，Mirza らの提案した conditional GAN (cGAN) [23] が有名である．cGAN は，Generator と Discriminator の双方にクラスラベルや文章等の条件を与えることによって画像生成をする．cGAN の応用手法は，Discriminator への条件の与え方に焦点を置いた手法が数多く提案されている．Reed ら [24] は，文章から画像の生成を前提として，Discriminator の中間層に Embedded した文章を条件として結合する手法を提案した．Odena らが提案した AC-GAN [26] は，Discriminator への条件入力を省く代わりに，Discriminator 内部で入力画像をクラス識別することで，意図した画像かつ認識しやすい画像の生成を実現した．Wan らの提案した手法 [35] は，マルチラベル（性別，年齢，エスニティシ）を用いて顔画像を生成するために AC-GAN と同様のアプローチを使用している．Wang ら [36] は DC-GAN を 3 次元に拡張することで，表情変化を捉えた GAN による画像生成法を提案した．また，Miyato らは，任意の次元数へ投影した条件を Discriminator の出力値へ加算することによって，高解像度な画像生成が可能な projection Discriminator [27] を提案した．一方，提案手法は，Generator 及び Discriminator 双方の条件入力方法に焦点をおいて高解像度な画像生成を図る．

Discriminator を改良することで高解像な画像生成を図った手法とは異なり，Generator を改良した手法も提案されている．Fused-GAN [37] は Generator の中間層で条件付きと条件なしに分岐することで，条件の詳細な情報を反映した画像生成を実現した．Yuan ら [38] は{0, 1}で表現した顔属性をグローバルとローカルなベクトルにそれぞれに分割して，StackGAN [25] を参考にした多重解像度の画像生成手法を用いることで，入力した顔属性の条件を満たした鮮明な画像生成を達成した．Sage らの提案した手法 [29] は，Generator の各層へ one-hot 表現した条件ベクトルを一様に与えている．Poly-GAN [30] は，条件となる画像と骨格情報を畳み込み処理を介して各層の特徴マップのサイズに合わせた後に結合する手法である．これらの手法により，出力層付近での条件消失を予防できる．一方，提案手法は条件を出力層まで均等に反映させつつ，条件の各要素に重み付けをすることで高解像度かつ条件を満たした画像を生成することを目的とする．

我々の手法は，入力条件が出力層付近での消失を予防するという観点で Sage らの手法と同じ設計をしているが，学習を通して最適な条件の入力位置を決定しない点が異なる．Poly-GAN も同様に条件の消失を予防するために条件を層ごとに入力するが，条件に画像を用いているため，各層で与える属性の強さや種類などは考慮されていない．また，条件は入力画像の特徴抽出をする Encoder の各層にのみ与えており，画像の生成に直結する Decoder 側に条件の入力はされていな

い．したがって，我々の手法は任意の層で反映する属性の種類や強度を考慮している点が新規性である．

3. 提案手法

本章では，提案手法の詳細を述べる．まず，**3.1** で本研究の問題設定を述べて，以降の節で手法の詳細を述べる．

3.1 問題設定

本研究では，cGAN の学習中に条件として与える各顔属性に重み付けをすることで，高精細かつ認識に有効な顔画像を生成することを目的とする．Generator は，$N(0,1)$ からサンプリングした d 次元の潜在変数 $\mathbf{z} \in \mathbb{R}^d$ と n 種類の顔属性を含んだ条件ベクトル $\mathbf{y} \in \{0,1\}^n$ を用いて顔画像を生成する．$\mathbf{y}$ は重み付けをして Generator の全層の特徴マップと結合する．Discriminator は実画像 $\mathbf{x}$，または生成画像 $\hat{\mathbf{x}} := G(\mathbf{z}|\mathbf{y})$ を入力して真偽判定及びクラス識別をする．

3.2 Weighted conditional layer の導入

本研究の提案である条件へ重み付けする層を Weighted conditional layer (Wc-layer) と呼称する．Wc-layer を介して Generator の各層へ条件を与えるメリットは二つある．1) 出力層付近での条件の消失を予防できる．2) 重み付けにより各層で異なる強度の条件を与えることができる．

まずはじめに，一つ目のメリットに関して述べる．Pandey ら [30] も述べているように Generator の入力層のみに条件を与えた場合，ネットワークの深い層で特徴が消失することが考えられる．この状況に陥らないために，提案手法も先行研究と同様に Generator の全層へ条件を与えて，特徴マップをチャネル方向に結合する．これにより，ネットワーク全体で一様に条件を反映する事ができるため，条件が消失する問題へ対処できる．

次に，二つ目のメリットに関して述べる．顔画像のような 1 サンプルに対して複数の属性が付与される場合，各属性で適した入力位置や強度があると考えられる．この問題を紐解くための最も愚直な方法は人手による重み付けである．しかしながら，人手によって各要素へ重み付けすることは時間的コストが高いため，非現実的である．また，人と生成モデルの知覚表現は反する可能性も高い．これらを対処するために，提案手法の Wc-layer では図 1 (c) に示すように，カーネルサイズが 1×1 の畳み込み処理と Sigmoid 関数による正規化によって重み付けをする．カーネルサイ

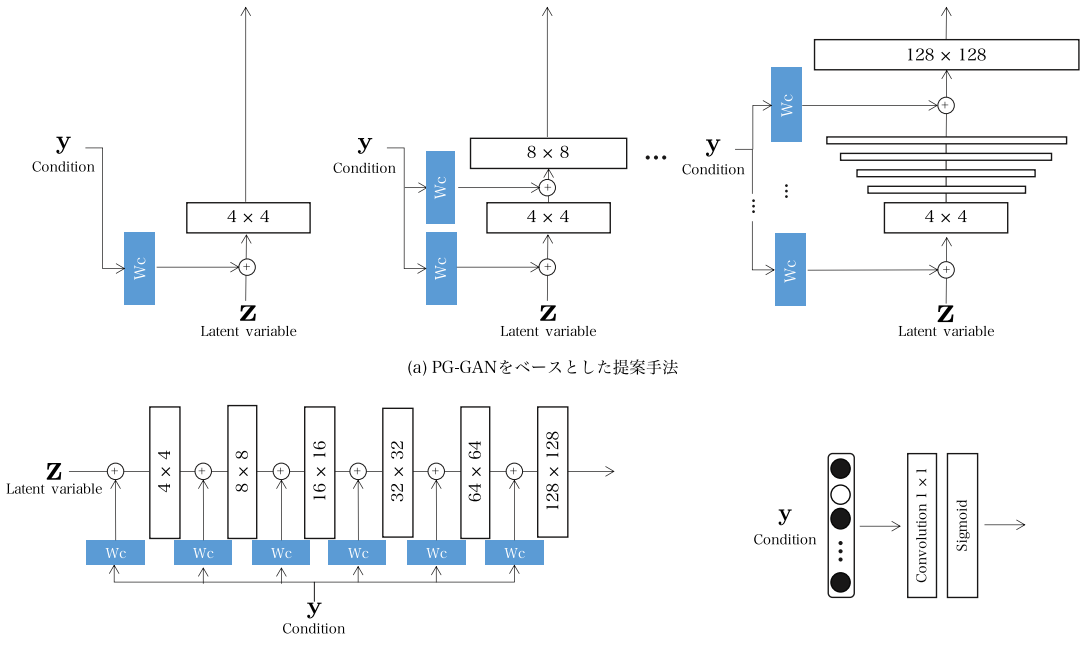


図 1　提案手法を導入した Generator の構造．図中の ⊕ は，重み付けした条件を特徴マップと結合する処理を表している．また，Wc は Weighted conditional layer である．

ズは 1×1 なので確実に条件の各要素に重み付けが可能になる．また，重み付けをしたことにより，条件が $(0,1)$ の範囲を超える可能性があるため，入力の値を $(0,1)$ で修めることが可能な Sigmoid 関数が活躍する．Wc-layer は，σ を Sigmoid 関数，重み行列 W_{conv} とバイアス $\mathbf{b} \in \mathbb{R}^m$ をパラメータにもつ畳み込み処理 $\phi(\cdot; W_{conv}, \mathbf{b})$ を用いて，以下の式で表すことができる．

$$\mathbf{h} = \sigma\left(\phi\left(\mathbf{y}; W_{conv}, \mathbf{b}\right)\right) \quad \text{s.t. } \phi : \{0,1\}^n \mapsto \mathbb{R}^m \quad (2)$$

ここで，$\mathbf{h}$ は Wc-layer を適用した後の m 次元の条件ベクトルである．一見すると，Poly-GAN と提案手法の処理は等価と捉えることができるが，Poly-GAN の畳み込み処理は条件として与えるデータをダウンサンプルするために利用している点で異なる．

図 1 (a) に示すように，提案手法のベースネットワークに PG-GAN を用いたときは，層を追加すると同時に Wc-layer も追加する．提案手法のベースネットワークを DC-GAN としたときは，図 1 (b) に示すように学習初期から全ての Wc-layer を導入して画像生成を行う．このとき，Wc-layer は重みを共有していないため，各層で各要素に対して最適な重みを学習することが可能となる．Wc-layer が出力する重み付けした条件は，与える層の特徴マップのサイズに合わせて拡大することで結合する．

3.3 マルチタスク Discriminator

本研究は認識に有効な画像を生成することが目的であるため，AC-GAN と同様に Discriminator でクラス分類をする．AC-GAN は，Discriminator の出力層を N+1 ユニットとすることでクラス分類をする．ここで，N はクラス数を表しており，残りの 1 ユニットは敵対的なゆう度の出力ユニットである．しかしながら，正確なクラス分類の効果を得るためには，クラス分類と敵対的なゆう度の出力を一つの全結合から出力することは相応しくないと考える．

したがって，本研究では Discriminator の出力層を個別のタスクに分割する．提案手法のベースネットワークを DC-GAN としたときの Discriminator を図 2 に示す．ベースネットワークが PG-GAN のときは，Generator のとき同様で層を順次追加して学習する．入力画像のクラス分類結果は Recognition branch から，実画像か生成画像かのゆう度は Adversarial branch から出力する．Recognition branch は，二つの全結合層で構成されており，クラス数分の確率 $p(\mathbf{x})$ を出力して教師信号との誤差計算をする．

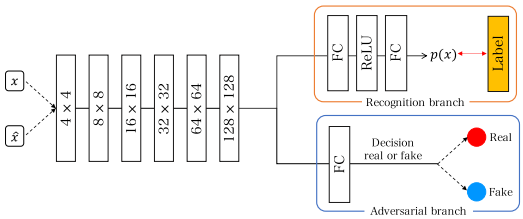


図 2　提案手法の Discriminator の構造

ポジティブな要素が一つでない顔属性は，Softmax 関数と Cross entropy による計算ができない．そこで，クラス分類誤差は Sigmoid cross entropy を用いて計算する．Sigmoid cross entropy は，教師ラベルを $\mathbf{y}$，サンプル数を N として以下に示す式で表すことができる．

$$\mathcal{L}_{cls} = -\frac{1}{N}\sum_{i=1}^{N} \mathbf{y}_i^{\top} \log \sigma(\mathbf{x}_i) \quad (3)$$

一方，Adversarial branch は，一つの全結合層で構成されており，実画像か生成画像かを判定するために一つのスカラー値を出力する．敵対的誤差は，ベースネットワークが DC-GAN のとき，式 (1) に式 (3) を加算したものを用いる．ここで，提案手法の Generator は条件入力を含むため，式 (1) 中の $G(\mathbf{z})$ が $G(\mathbf{z}|\mathbf{y})$ となる．また，ベースネットワークが PG-GAN のときは，式 (4) に式 (3) を加算したものを学習する．

$$\begin{aligned}\mathcal{L}_{gan} = {} & E_{\hat{\mathbf{x}}\sim P_{\mathbf{z}}(\mathbf{z})}[D(\hat{\mathbf{x}})] - E_{\mathbf{x}\sim P_{data}(\mathbf{x})}[D(\mathbf{x})] \\ & + \lambda E_{\tilde{\mathbf{x}}\sim P_{\tilde{\mathbf{x}}}}[(\|\nabla_{\tilde{\mathbf{x}}} D(\tilde{\mathbf{x}})\|_2 - 1)^2] \\ & + \alpha E_{\mathbf{x}\sim P_{data}(\mathbf{x})}[D(\mathbf{x})^2] \end{aligned} \quad (4)$$

ここで，$\tilde{\mathbf{x}} = \epsilon\mathbf{x} + (1-\epsilon)\hat{\mathbf{x}}$，$\epsilon$ は $U[0,1]$ からサンプリングした値，$\lambda = 10$，$\alpha = 0.001$ である．

一般的に GAN の学習は不安定であるため，生成画像のバリエーションが消失するモード崩壊に陥りやすいと言われている．そこで，PG-GAN で使用された Minibatch standard deviation (Minibatch stddev) [6] を最後の畳み込み処理前に導入することでモード崩壊を回避する．Minibatch stddev は，特徴マップに関して Minibatch 内の標準偏差を計算した標準偏差マップを特徴マップと合わせて畳み込むことで Minibatch 内の多様性を保証できる手法である．これはベースネットワークが DC-GAN と PG-GAN のどちらであっても導入する．

4. 評価実験

提案手法を導入した各 GAN の手法により顔画像を生成し，定量的な画質評価として主観評価及び客観評価を用いて従来法と比較する．また，重み付き条件を Generator に入力する効果とマルチタスク化した Discriminator の効果を Ablation study により示す．最後に，我々の手法で生成した画像を追加して学習した顔属性認識モデルの性能を示す．このとき，生成した画像をそのまま使用する方法と，人が介入してデータを選別する方法の二つを使用する．一般に，GAN の生成画像は表情などが読み取りづらい崩れた画像となることがあるため，学習に悪影響を及ぼす可能性が高い．したがって，人手によるデータ選別が性能向上に貢献すると考える．

4.1 実験概要

本実験では，条件付きに拡張した DC-GAN 及び PG-GAN を従来手法とする．この二つの手法を本実験中は，conditional DC-GAN (cDC-GAN) と conditional PG-GAN (cPG-GAN) と呼称する．提案手法を導入した DC-GAN 及び PG-GAN は，それぞれ WcDC-GAN と WcPG-GAN と定義する．

学習に使用するデータセットは，CelebA データセット [39] とする．CelebA データセットは，20 万枚を超える顔画像を保有するデータセットであり，約 1 万人の画像が含まれている．各顔画像データに対して，40 属性の顔属性ラベルの付与，五つの顔器官点 (両目，鼻頭，口先) のアノテーションがされている．顔属性は，ポジティブな属性に 1，ネガティブな属性に –1 が付与されている．本実験では問題を簡単化するために 40 種類中五つの属性を使用する．使用する属性は，比較的変化が分かりやすい，Male (性別)，Eyeglasses (メガネ)，Smiling (笑顔)，Goatee (ヒゲ)，Bangs (前髪) を使用する．Generator に与える潜在変数の次元数は，DC-GAN をベースとしたとき 100 次元，PG-GAN をベースとしたとき 512 次元とする．

定量的な画質評価は，主観評価及び客観評価によって従来手法との比較をする．主観評価では，21 人の被験者により画質と顔属性を考慮した評価をする．主観評価の詳細は，**4.2** で述べる．客観評価には，Inception score (IS) [11] と Fréchet inception distance (FID) [14] を用いて評価をする．IS 及び FID の詳細は，**4.3** で述べる．

顔属性認識の実験では，101 層の ResNet を 300 epoch 学習をする．最適化関数には momentum Stochastic Gradient Descent (momentum SGD) を学習率 0.1，momentum を 0.9 として使用する．学習率は，{150, 225} epoch で 1/10 に減衰させる．認識性能は，Baseline と Active learning の有無を比較する．Baseline は，CelebA データセットから 9 割の画像を用いて学習したモデルの認識率である．追加データは 40 属性全てを用いて学習した提案手法によって顔画像を生成する．データを追加する属性は，Baseline で認識率が 90% を下回る顔属性を対象として 4,000 枚の顔画像を加える．画像生成のとき，対象クラスは満たすべき条件であるため 1 として，それ以外の 39 属性は{0, 1}をランダムに定義する．この処理は対象クラスをランダムに変更しながら，追加データを収集する．Active learning 無しは，提案手法の生成画像を全て用いて Baseline の結果から追加学習した際の認識率である．Active learning 有りは，性能向上に寄与すると考えられる画像を人手で選別して追加学習した認識率である．このとき，生成画像の合計が 4,000 枚になるまで生成と選別を繰り返す．生成画像に対する教師信号は，画像生成時に与えた条件とする．

4.2 生成画像の主観評価

IS や FID などの客観評価は，主に生成画像の品質を重視して数値的に評価するため，生成画像が条件を満たしているかの評価は困難である．したがって，人により品質と条件を満たしているかの判定をする主観評価が必要である．

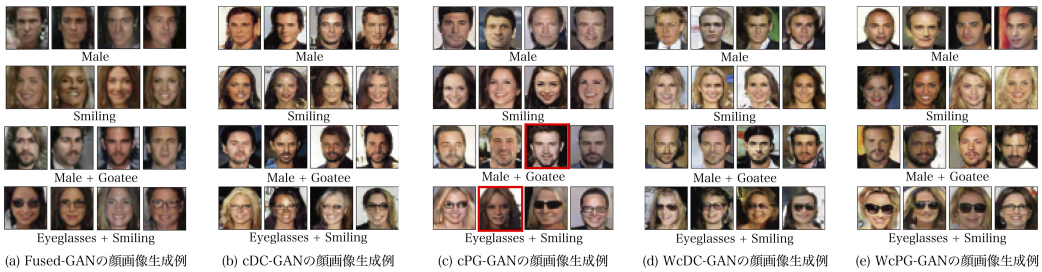
cDC-GAN と WcDC-GAN を例に挙げて主観評価の説明をする．まず，cDC-GAN と WcDC-GAN それぞれで生成した画像を 1 枚づつ被験者に提示する．被験者は，提示された画像がどちらの画像が高品質で条件を満たしているかを選択する．この工程を 150 枚の画像に対して行う．最終的なスコアは，以下に示す式で計算する．

$$S = \frac{n}{150} \times 100 \tag{5}$$

ここで，n は cDC-GAN または WcDC-GAN それぞれの選択数を示している．cPG-GAN と WcPG-GAN も同様にして評価する．

4.3 生成画像の客観評価

IS 及び FID は共に，一般物体認識のデータセットである ImageNet [40] を Inception network [41] を用いて 1,000 クラス分類した際の事前学習済みモデルを用いて計算をする．



(a) Fused-GANの顔画像生成例 (b) cDC-GANの顔画像生成例 (c) cPG-GANの顔画像生成例 (d) WcDC-GANの顔画像生成例 (e) WcPG-GANの顔画像生成例

図 3　各手法の顔画像生成例 [128×128 pixels]

Inception score (IS): IS は，入力した画像がどれほど識別しやすいかと入力画像のクラス間のバリエーションの二つの観点から評価を行うことが可能な生成モデルの指標の一つである．IS における識別のし易さは，識別器に入力する画像及びクラスラベルをそれぞれ $\mathbf{x}_i$，y とすると，条件付き分布 $p(y|\mathbf{x}_i)$ のエントロピーが小さくなることと等価であるとみなせる．

入力画像の多様性に関しては，条件付き分布 $p(y|\mathbf{x}_i)$ を x について積分を行い周辺化した分布 $p(y)$ のエントロピーが大きくなることに等しい．そこで，Kullback leibler divergence (KL ダイバージェンス) を用いて，入力データ数を N とすると，以下に示す式で表せる．

$$IS = \exp\left(\frac{1}{N}\sum_{i=1}^{N} KL[p(y|\mathbf{x}_i)\|p(y)]\right) \tag{6}$$

二つの分布の間の距離が大きい値であるほど良い生成画像となるため，IS は式 (6) のスコアが大きいほど優良データとなる．

Fréchet inception distance (FID): IS は，実画像または生成画像のどちらかのデータ群のみを入力してスコアを計算するため，実画像の分布を考慮した計算過程でない．そこで FID では，IS の弱点を改善するために実画像と生成画像の分布間の距離を測っている．FID は，Inception モデルの中間層の特徴マップが，多変量正規分布に従うと仮定して実画像及び生成画像の分布間を Fréchet distance を用いて求める．FID の式は，以下に示す式で表される．

$$FID = \|\mathbf{m} - \mathbf{m}_r\|_2^2 + tr\left(\mathbf{C} + \mathbf{C}_r - \sqrt{2\mathbf{C}\mathbf{C}_r}\right) \tag{7}$$

ここで，$\mathbf{m}_r$，$\mathbf{C}_r$ はそれぞれ実画像に関する特徴マップの平均と共分散行列，$\mathbf{m}$，$\mathbf{C}$ は生成画像に関する平均と共分散行列である．FID は，式 (7) のスコアが低いほど優良な画像データであることを示している．

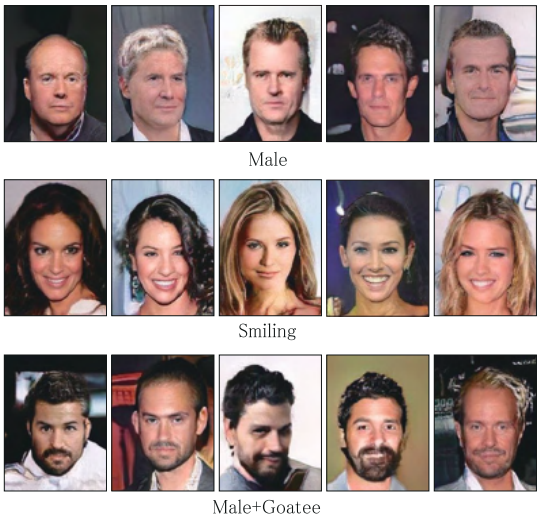


図 4　WcPG-GAN で 192 × 256 [pixels] の顔画像生成例

4.4 実験結果

まず，各手法で生成した顔画像を図 3 に示す．目視による判断では，全ての手法において，顔と判断可能な画像が生成されており，入力された条件を反映した顔画像が生成されていることが確認できる．cDC-GAN 及び WcDC-GAN は，提案手法を導入することによって画質が微小に劣化した．一方，cPG-GAN と WcPG-GAN は，提案手法を導入しても画質が劣化することなく顔画像生成ができた．cPG-GAN の生成画像である図 3 (c) に着目すると，赤枠で囲った顔画像は上手く条件が反映されていない．つまり，最終層に至るまでに条件が消失している．このことから，DC-GAN 程度のシンプル，かつ浅いネットワークであれば，通常通り条件を与えることが適しているといえる．

WcPG-GAN は図 4 に示すように，更に高解像度な画像を生成した場合でも，条件を満たしつつ自然で高品質な顔画像の生成ができる．高解像度の顔画像を生

成することで，顔の詳細な部位まで鮮明に生成することが可能である．

次に，生成画像を定量的に評価した結果を表 1 に示す．生成した画像サイズは，全ての手法 128×128 [pixels] とした．cDC-GAN と WcDC-GAN の比較では，IS は同程度のスコアであり FID は 0.5 ポイント良いスコアで，主観評価は従来手法が高くなる．言い換えれば，求めた結果を得ることができなかった．これは，Wc-layer を各層に導入して画像生成するには Generator の層が少なすぎたためだと考える．一方，cPG-GAN と WcPG-GAN の比較では，IS は同等であるが，主観評価及び FID は WcPG-GAN が高いスコアとなっている．特に FID は cPG-GAN と比べて 5 ポイント以上良いスコアを達成した．定性的な評価からも確認できるとおり，PG-GAN をベースにすることで DC-GAN より画質が改善される．これは，PG-GAN が DC-GAN と比較して倍以上の畳み込み層で構成されているためであると言える．しかしながら，表 1 の全ての定量評価において，cPG-GAN は WcPG-GAN のスコアより悪いスコアであることから，深いネットワーク構造が仇となり入力層のみへ与えた条件が消失していると言える．よって，深い構造の Generator で条件を満たしつつ高精細な顔画像生成をするためには，提案手法を導入することが有効であると言える．

また，AC-GAN の IS 及び FID は WcDC-GAN のスコアと同程度である．このことから，Discriminator の改良は生成画像の品質にあまり影響しないと考えられるが，AC-GAN の結果は WcDC-GAN より悪いスコアであるため，わずかながら有効性があると言える．

4.5 Ablation study

本節では，提案手法を用いて Wc-layer 及び Recognition branch の効果を検証するための実験をする．Wc-layer に関する実験は以下に示す 3 種類で比較する．

- Input only: 入力層のみに条件をそのまま与えるモデル
- All layers: 全ての層に条件をそのまま与えるモデル
- All layers + Wc-layer: Wc-layer で重み付けした条件を全ての層に与えるモデル

このとき，全てのモデルは Recognition branch を使用して学習する．したがって，Input only は AC-GAN と同じ構造である．Recognition branch に関する実験は，Recognition branch の有無で比較する．このとき，Generator の Wc-layer は切除しないものとする．

4.5.1 Wc-layer の効果の検証

表 2 に異なる条件入力方法で生成した画像に対する定量的評価結果を示す．Wc-layer を用いた提案手法は，従来手法と同様に入力層のみに条件を与えたときより IS と FID 共に優れていることが確認できる．また，提案手法から Wc-layer を除去して Sage らの手法 [29] と同じ設計にした場合も，Wc-layer を用いたときに劣る結果である．特に FID のスコアに着目すると，条件を単純に全ての層に与えたモデルの生成画像は入力層のみに条件を与えた結果より劣化したにもかかわらず，Wc-layer を導入することでスコアが飛躍的に改善していることが確認できる．このことから，全ての層に条件を与えて画像生成をする場合は，Wc-layer によって重み付けした条件を用いて画像生成することが有効であるといえる．

4.5.2 Recognition branch の効果の検証

表 3 に Recognition branch の有無で学習したモデルの生成画像の定量的評価結果を示す．提案手法から Recognition branch を切除することによって，IS のスコアは大差ないが FID のスコアが劣化した．これは，Recognition branch の誤差を最小化することが，Generator が鮮明な画像を生成することを促進しているといえる．したがって，Recognition branch は鮮明な画像生成には必要な要素だといえる．

4.6 生成画像を用いた顔属性認識

顔属性識別へ生成画像を使用した際の結果を表 4 に

表 1　定量的画質評価結果

Method	IS ↑	FID ↓	主観評価 (21 人) ↑
Real image	3.21	-	-
AC-GAN	1.59	17.19	-
cDC-GAN	1.70	17.22	53.1
WcDC-GAN (Proposed)	1.67	**16.70**	46.9
cPG-GAN	1.68	11.90	44.5
WcPG-GAN (Proposed)	**1.73**	**6.10**	**55.5**

表 2　異なる条件入力方法の定量的評価

	IS ↑	FID ↓
Input only	1.59	17.19
All layers	1.62	19.23
All layers+Wc-layer	**1.73**	**6.10**

表 3　Recognition branch の有無の定量的評価

Recognition branch	IS ↑	FID ↓
	1.65	10.82
✓	**1.73**	**6.10**

表 4　顔属性認識の結果（%）．w/o AL は active learning 無し，w/ AL は active learning 有りを指している．太字は，Baseline から精度が向上したことを表している．

	5 o Clock Shadow	Arched Eyebrows	Attractive	Bags Under Eyes	Bald	Bangs	Big Lips	Big Nose	Black Hair	Blond Hair	Blurry	Brown Hair	Bushy Eyebrows	Chubby	Double Chin	Eyeglasses	Goatee	Gray Hair	Heavy Makeup	High Cheekbones
Baseline	94.6	84.0	81.9	84.8	98.8	95.7	71.6	84.8	89.8	95.8	96.1	88.0	92.1	95.1	96.3	99.6	97.5	97.9	91.6	87.3
Fused-GAN	94.3	82.9	80.6	83.7	**98.9**	95.5	70.5	82.9	88.7	95.2	95.6	86.7	91.9	**95.2**	96.0	**99.7**	97.4	**98.2**	91.0	86.2
cPG-GAN (w/ AL)	94.3	83.0	80.6	83.7	**98.9**	95.4	70.6	83.1	88.9	95.4	95.8	86.4	91.9	95.1	95.9	**99.7**	97.3	**98.1**	91.0	86.1
ours (w/o AL)	**94.8**	83.6	81.4	84.6	98.8	**95.8**	71.4	84.0	89.7	**95.9**	95.5	**88.5**	**92.5**	**95.3**	96.1	**99.7**	97.5	**98.2**	91.3	87.3
ours (w/ AL)	**94.8**	83.2	80.6	84.7	**99.0**	95.7	71.2	83.9	89.8	95.8	96.0	**88.5**	**92.4**	**95.3**	96.3	**99.7**	97.5	**98.2**	91.4	**87.4**

	Male	Mouth Slightly Open	Mustache	Narrow Eyes	No Beard	Oval Face	Pale Skin	Pointy Nose	Receding Hairline	Rosy Cheeks	Sideburns	Smiling	Straight Hair	Wavy Hair	Wearing Earrings	Wearing Hat	Wearing Lipstick	Wearing Necklace	Wearing Necktie	Young
Baseline	98.2	93.7	97.0	86.6	96.2	75.9	97.1	76.1	93.4	94.3	97.6	92.9	83.0	83.5	90.6	99.0	94.1	87.6	96.8	86.8
Fused-GAN	**98.4**	93.6	96.8	**86.8**	96.0	73.2	96.7	74.1	93.2	**94.5**	**97.7**	92.4	82.2	82.8	89.9	99.0	93.2	86.4	96.7	**87.2**
cPG-GAN (w/ AL)	**98.4**	93.3	96.7	**86.8**	96.0	72.1	96.6	74.6	93.1	**94.6**	**97.7**	92.2	82.2	82.5	89.9	**99.1**	93.1	86.6	96.6	**87.2**
ours (w/o AL)	**98.4**	**93.9**	97.0	**87.4**	96.2	73.9	97.0	75.4	**93.6**	**95.2**	**97.8**	92.8	**83.8**	**83.9**	90.6	**99.1**	93.2	87.3	**96.9**	**87.3**
ours (w/ AL)	**98.5**	93.7	96.9	**87.1**	96.1	74.7	97.0	**76.4**	**93.6**	**94.9**	**97.9**	92.7	**83.5**	**83.6**	90.4	**99.1**	93.6	87.4	**97.0**	**87.5**

示す．任意の属性の認識率 Acc_{attr} は以下に示す式によって求める．

$$Acc_{attr} = \frac{1}{N}\sum_{i=1}^{N} s_i \times 100, \tag{8}$$

$$s = \begin{cases} 1 & \arg\max\left(p(x)\right) = t \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

ここで，$p(x)$ は ResNet-101 へ画像 x を与えて出力するクラス確率，N は評価データの総数，t は ground truth のインデックスである．

まず，データ選別なしの提案手法の結果に着目すると，19 個の顔属性が Baseline から精度向上していることが確認できる．特に，“Narrow Eyes”，“Rosy Cheeks”，“Straight Hair”，“Young” は 0.5 ポイント以上の精度向上を達成した．一方，データ選別なしの Fused-GAN は，精度向上した顔属性が 9 個であった．また，cPG-GAN はデータ選別をしたにもかかわらず，Fused-GAN と変わらない結果であることが確認できる．Fused-GAN や cPG-GAN の結果は，データを追加しても大幅に精度向上する属性が少ないことがわかる．この結果は，生成画像が学習データとして適切でないか，入力した条件を満たせていない画像が多いことが原因だと考える．

次に，データ選別ありの提案手法の結果に着目すると，精度向上した属性数はデータ選別なしのときと同様であった．しかしながら，“Pointy Nose” の認識精度はデータ選別なしのときより 1 ポイント向上しており，Baseline よりも高い精度である．これは，データ選別により鮮明な生成画像を収集したことで，モデルが細かいテクスチャを捉えることができた恩恵だといえる．

以上の結果より，提案手法は生成画像の選別しない場合でも，Baseline や他手法の結果よりも精度向上に寄与するが，鮮明な生成画像を多く学習に用いることで特徴が捉えづらい属性の認識精度向上に寄与することがわかった．また，定義が曖昧な属性はデータ選別によって鮮明な画像を追加したとしても，精度が向上しないこともわかった．

4.7 考　　察

提案手法で条件を満たした顔画像生成ができる要因として，重み付き条件により各層で最適な顔属性を反映していると考えられる．そこで，Generator の条件の寄与率を可視化する．寄与率は，各 Wc-layer の畳み込み層の重みフィルタを用いて算出する．寄与率 C_t は，重みフィルタ数を N，属性数を M，重みフィルタを W，寄与率を求める属性を t とすると式 (10) となる．

$$C_t = \frac{1}{N}\sum_{n=1}^{N} \frac{|W_{t,n}|}{\sum_{m=1}^{M} |W_{m,n}|} \tag{10}$$

WcDC-GAN 及び WcPG-GAN の各層における各属性の寄与率を図 5 及び図 6 にそれぞれ示す．ここで，WcPG-GAN では，各解像度の中間生成画像が取得できるため，図 6 中にそれぞれ示した．中間生成画像は，Blond Hair+No Beard+Smiling を生成したときの顔画像である．

まず，WcDC-GAN の寄与率に着目すると，1 層目の性別の寄与率が高いが，基本的に全ての層において寄与率に大きなばらつきがない事が確認できる．したがって，全ての層に均等な強度で属性を与えることが重要であると言える．

次に，図 6 の WcPG-GAN の寄与率に着目すると，Male (性別) は，入力層付近での寄与率が高く出力層に近づくにつれて低くなる．よって，中間層で性別が決定していると考えられる．中間生成画像においても，性別の寄与率が高くなる層から顔の輪郭が鮮明になることが確認できる．Eyeglasses (メガネ) は出力層付近で高い寄与率であり，入力層付近では，ほとんど条件としてされていないことが確認できる．そのため，高解像度時に Eyeglasses に最も着目し，顔画像を生成していると考えられる．Smiling (表情) は，中間層で寄与率が増加し，入力及び出力層では寄与率が低い傾向があることが確認できる．中間生成画像でも，中間層に近づくにつれ顔の表情が鮮明になる．No Beard (ヒゲ) は，各層で寄与率に大きな差はない．Blond Hair (金髪) は，入力層で多く寄与しており出力層へ近づくにつれ寄与率が低くなる．Male と Eyeglasses に着目すると，ネットワーク全体を通して寄与率が反比例していることが確認できる．よって性別が決定した後に装飾品等の詳細な顔属性が生成されると言える．このように，顔属性が各層で異なる寄与率であることから，学習を通じて Generator が最適な条件の反映位置を決定できたと言える．

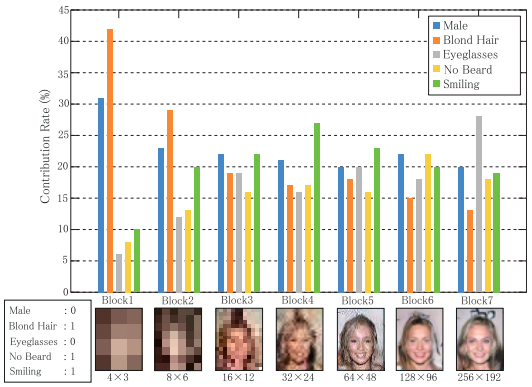


図 5　WcDC-GAN における層ごとの顔属性の寄与率

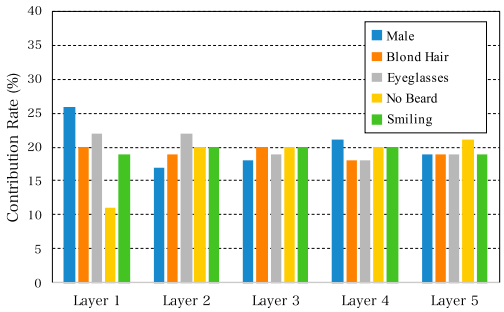


図 6　WcPG-GAN における顔属性の寄与率と中間生成画像

5. むすび

本論文では，Weighted conditional layer (Wc-layer) を導入することで段階的に条件を反映することが可能な Weighted conditional GAN (Wc-GAN) を提案した．Wc-layer は，1×1 の畳み込み処理を{0, 1}で定義された条件ベクトルに適用することで重み付けをする機構である．Wc-layer を Generator の各層へ導入することで，顔属性ごとに最適な入力位置を自動で決定することを可能とした．

定量的な画質評価より，ベースネットワークとして PG-GAN のような深いネットワークを用いた際に，提案手法の貢献度が大きいことがわかった．Ablation study より，マルチタスク化した Discriminator より Wc-layer が生成画像の品質向上には寄与していることを確認した．各層に導入した Wc-layer で反映される属性を調査した結果，低解像度時には性別などの大局的な属性，高解像度になるにつれて装飾品などの詳細な属性が反映される傾向があることが判明した．顔属性認識へ生成画像を使用した実験では，認識率の低い顔属性のサンプル数を増加して学習することで，Baseline より認識率が向上した．更に，生成画像の中からデータ選別をして学習データに追加することで，幾つかの顔属性がデータ選別しないときよりも精度向上した．このことから，認識に有効な顔画像を学習して提供できたと言える．

しかしながら，GAN の学習の特性上，サンプル数が顕著に少ない属性を鮮明に生成することが困難である．これは，提案手法を含めた多くの GAN を用いた手法に共通する問題である．そのため，今後は，この問題点を解消するための手法の提案を目的とする．

文　献

[1] D.P. Kingma and M. Welling, “Auto-encoding variational bayes,” Int. Conf. Learning Representations, pp.1–14, 2014.

[2] D.P. Kingma and P. Dhariwal, “Glow: Generative flow with invertible 1x1 convolutions,” Advances in Neural Information Processing Systems, pp.10215–10224, 2018.

[3] A. van denOord, N. Kalchbrenner, O. Vinyals, L. Espeholt, A. Graves, and K. Kavukcuoglu, “Conditional image generation with pixelcnn decoders,” Advances in Neural Information Processing Systems, pp.4790–4798, 2016.

[4] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets,” Advances in Neural Information Processing Systems, eds. by Z. Ghahramani, M. Welling, C. Cortes, N.D. Lawrence, and K.Q. Weinberger, pp.2672–2680, 2014.

[5] A. Radford, L. Metz, and S. Chintala, “Unsupervised representation learning with deep convolutional generative adversarial networks,” Int. Conf. Learning Representations, pp.1–16, 2016.

[6] T. Karras, T. Aila, S. Laine, and J. Lehtinen, “Progressive growing of gans for improved quality, stability, and variation,” Int. Conf. Learning Representations, pp.1–26, 2018.

[7] H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena, “Self-attention generative adversarial networks,” Int. Conf. Machine Learning, vol.97, pp.7354–7363, 2019.

[8] T. Karras, S. Laine, and T. Aila, “A style-based generator architecture for generative adversarial networks,” IEEE/CVF Conf. Computer Vision and Pattern Recognition, pp.4401–4410, 2019.

[9] A. Brock, J. Donahue, and K. Simonyan, “Large scale gan training for high fidelity natural image synthesis,” Int. Conf. Learning Representations, pp.1–35, 2019.

[10] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, “Analyzing and improving the image quality of stylegan,” IEEE/CVF Conf. Computer Vision and Pattern Recognition, pp.8110–8119, 2020.

[11] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, “Improved techniques for training gans,” Advances in Neural Information Processing Systems, pp.2234–2242, 2016.

[12] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, “Improved training of wasserstein gans,” Advances in Neural Information Processing Systems, pp.5767–5777, 2017.

[13] M. Arjovsky, S. Chintala, and L. Bottou, “Wasserstein generative adversarial networks,” Int. Conf. Machine Learning, vol.70, pp.214–223, 2017.

[14] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, “Gans trained by a tow time-scale update rule converge to a local nash equilibrium,” Advances in Neural Information Processing Systems, pp.6626–6637, 2017.

[15] X. Mao, Q. Li, H. Xie, R.Y. Lau, Z. Wang, and S.P. Smolley, “Least squares generative adversarial networks,” IEEE Int. Conf. Computer Vision, pp.2794–2802, 2017.

[16] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida, “Spectral normalization for generative adversarial networks,” Int. Conf. Learning Representations, pp.1–26, 2018.

[17] J.-Y. Zhu, T. Park, P. Isola, and A.A. Efros, “Unpaired image-to-image translation using cycle-consistent adversarial networks,” IEEE Int. Conf. Computer Vision, pp.2223–2232, 2017.

[18] A. Shrivastava, T. Pfister, O. Tuzel, J. Susskind, W. Wang, and R. Webb, “Learning from simulated and unsupervised images through adversarial training,” IEEE Conf. Computer Vision and Pattern Recognition, pp.2107–2116, 2017.

[19] P. Isola, J.-Y. Zhu, T. Zhou, and A.A. Efros, “Image-to-image translation with conditional adversarial networks,” IEEE Conf. Computer Vision and Pattern Recognition, pp.1125–1134, 2017.

[20] Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, and J. Choo, “Stargan: Unified generative adversarial networks for multi-domain image-to-image translation,” IEEE Conf. Computer Vision and Pattern Recognition, pp.8789–8797, 2018.

[21] X. Huang, M.-Y. Liu, S. Belongie, and J. Kautz, “Multimodal unsupervised image-to-image translation,” European Conference on Computer Vision, pp.172–189, 2018.

[22] C. Ledig, L. Thesi, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, “Photo-realistic single image super-resolution using a generative adversarial network,” IEEE Conf. Computer Vision and Pattern Recognition, pp.4681–4690, 2017.

[23] M. Mirza and S. Osindero, “Conditional generative adversarial nets,” arXiv preprint arXiv:1411.1784, 2014.

[24] S. Reed, Z. Akata, X. Yan, L. Logeswaran, B. Schiele, and H. Lee, “Generative adversarial text to image synthesis,” Int. Conf. Machine Learning, vol.48, pp.1060–1069, 2016.

[25] H. Zhang, T. Xu, H. Li, S. Zhang, X. Wang, X. Huang, and D.N. Metaxas, “Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks,” IEEE Int. Conf. Computer Vision, pp.5907–5915, 2017.

[26] A. Odena, C. Olah, and J. Shlens, “Conditional image synthesis with auxiliary classifier GANs,” Int. Conf. Machine Learning, vol.70, pp.2642–2651, 2017.

[27] T. Miyato and M. Koyama, “cgans with projection discriminator,” Int. Conf. Learning Representations, pp.1–23, 2018.

[28] H. Zhang, T. Xu, H. Li, S. Zhang, X. Wang, X. Huang, and D.N. Metaxas, “Stackgan++: Realistic image synthesis with stacked generative adversarial networks,” IEEE Trans. Pattern Analysis & Machine Intelligence, vol.41, no.8, pp.1947–1962, Aug. 2019.

[29] A. Sage, E. Agustsson, R. Timofte, and L.V. Gool, “Logo synthesis and manipulation with clustered generative adversarial networks,” IEEE Conf. Computer Vision and Pattern Recognition, pp.5879–5888, 2018.

[30] N. Pandey and A. Savakis, “Poly-gan: Multi-conditioned gan for fashion synthesis,” Neurocomputing, vol.414, no.13, pp.356–364, 2020.

[31] T. Chen, M. Lucic, N. Houlsby, and S. Gelly, “On self modulation for generative adversarial networks,” Int. Conf. Learning Representations, pp.1–18, 2019.

[32] H. Adachi, H. Fukui, T. Yamashita, and H. Fujiyoshi, “Facial image generation by generative adversarial networks using weighted conditions,” Int. Joint Conf. Computer Vision, Imaging and Computer Graphics Theory and Applications, vol.4, pp.139–145, 2019.

[33] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” IEEE Conf. Computer Vision and Pattern Recognition, pp.770–778, 2016.

[34] X. Huang and S. Belongie, “Arbitrary style transfer in real-time with adaptive instance normalization,” IEEE Int. Conf. Computer Vision, pp.1501–1510, 2017.
[35] L. Wan, J. Wan, Y. Jin, Z. Tan, and S.Z. Li, “Fine-grained multi-attribute adversarial learning for face generation of age, gender and ethnicity,” Int. Conf. Biometrics, pp.98–103, 2018.
[36] Y. Wang, A. Dantcheva, and F. Bremond, “From attribute-labels to faces: face generation using a conditional generative adversarial network,” European Conf. Computer Vision Workshops, pp.692–698, 2018.
[37] N. Bodla, G. Hua, and R. Chellappa, “Semi-supervised fusedgan for conditional image generation,” European Conf. Computer Vision, pp.669–683, 2018.
[38] Z. Yuan, J. Zhang, S. Shan, and X. Chen, “Attributes aware face generation with generative adversarial networks,” Int. Conf. Pattern Recognition, pp.1657–1664, 2020.
[39] Z. Liu, P. Luo, X. Wang, and X. Tang, “Deep learning face attributes in the wild,” IEEE Int. Conf. Computer Vision, pp.3730–3738, 2015.
[40] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A.C. Berg, and L. Fei-Fei, “Imagenet large scale visual recognition challenge,” Int. J. Computer Vision, vol.115, pp.211–252, 2015.
[41] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions,” IEEE Conf. Computer Vision and Pattern Recognition, pp.1–9, 2015.

（2021 年 6 月 16 日受付，9 月 24 日再受付，
12 月 13 日早期公開）

足立　浩規

2020 中部大学大学院修士課程了．同年よりロボット理工学専攻博士後期課程に在籍．生成モデル，敵対的学習の研究に従事．

平川　　翼

2017 広島大学大学院博士課程後期課程情報工学専攻了．2017〜2019 中部大学研究員，2019 同大学特任助教，2021 同大学講師．2014 独立行政法人日本学術振興会特別研究員 DC1．2014〜2015 ESIEE Paris 客員研究員．コンピュータビジョン，パターン認識，医用画像処理の研究に従事．2019 画像の認識・理解シンポジウム (MIRU) フロンティア賞．2019，2020 画像の認識・理解シンポジウム (MIRU) 長尾賞．博士 (工学).

山下　隆義　（正員）

2002 奈良先端科学技術大学院大学博士前期課程了．2002 オムロン株式会社入社，2011 中部大学大学院博士後期課程了（社会人ドクター），2014 中部大学講師，2017 同大学准教授，2021 同大学教授．人の理解に向けた動画像処理，パターン認識・機械学習の研究に従事．2009 画像センシングシンポジウム高木賞．2013 本会情報・システムソサイエティ論文賞．2013 本会 PRMU 研究会研究奨励賞．2016 画像センシングシンポジウム最優秀学術賞．2019 像の認識・理解シンポジウム (MIRU) フロンティア賞．2019，2020 画像の認識・理解シンポジウム (MIRU) 長尾賞．2020 電子情報通信学会論文賞．博士 (工学).

藤吉　弘亘　（正員：フェロー）

1997 中部大学大学院博士後期課程了．1997〜2000 米カーネギーメロン大学ロボット工学研究所 Postdoctoral Fellow．2000 中部大学講師，2004 同大准教授を経て 2010 年より同大教授．2005〜2006 米カーネギーメロン大学ロボット工学研究所客員研究員．計算機視覚，動画像処理，パターン認識・理解の研究に従事．2005 ロボカップ研究賞．2009 情報処理学会論文誌コンピュータビジョンとイメージメディア優秀論文賞，2009 山下記念研究賞．2010，2013，2014 画像センシングシンポジウム優秀学術賞．2013 本会情報・システムソサイエティ論文賞．2019 画像の認識・理解シンポジウム (MIRU) フロンティア賞．2019，2020 画像の認識・理解シンポジウム (MIRU) 長尾賞．博士 (工学)．情報処理学会，IEEE 各会員．