Adaptive Point-wise グループ化畳み込みを用いた小規模データセットからの画像の生成

武田 麻奈[1,a)]　柳井 啓司[1,b)]

概要

近年の Generative Adversarial Networks (GAN) の発展によって，高精度な画像生成が可能となっている．しかし，通常は大量の学習データと，長時間の学習が必要になるという問題点がある．それに対して，大規模データで学習した画像生成モデルを，少数の学習データで短時間だけファインチューニングする Few-shot GAN が研究されている．本論文では，Few-shot GAN に対して Adaptive Point-wise 畳み込みを用いた新しいモデルを提案する．実験によって，従来手法に比べて品質の高い画像の生成が可能となることを示した．

1. はじめに

深層学習モデルでは，一般的に学習に多くのデータを使用する．しかし，大規模データセットの構築には多大な労力が必要である．事前に学習されたモデルを用いた事前知識の伝達は，小規模データセットを用いた学習に効果的である．画像分類モデルについては，ImageNet などの入手が容易な大規模ラベル付きデータセットを使用して学習され，別の特定ドメインの小規模データセットを使用して別のドメインに転送される転移学習が広く研究されている．また，事前に学習されたモデルの重みからターゲットデータセットを学習する方法はファインチューニングと呼ばれる．事前に学習されたデータセットとは異なるドメインのデータセットを使用してモデルをファインチューニングする場合でも，精度は向上する傾向にある．これは，事前に学習されたモデルが，小さなターゲットデータセットでは取得できない一般的に有用な重みを取得するためである．

深層生成モデルについても，事前知識を別のデータセットに転送する方法が提案されている．Noguchi と Harada[10] は，小規模データセットからの画像生成を実現するために，事前に学習された生成モデルを転送するための新しい方法を提案した．事前知識を適応させるために，ジェネレータのバッチ統計量のスケールとシフトパラメータに焦点を当てた．畳み込み層のパラメータを固定して，畳み込み層のフィルタを重み付けするスケール/シフトパラメータを更新することで，事前に学習したモデルを少ないデータセットでターゲットドメインに適応させることができる．

本論文では，Noguchi と Harada と同様に，小規模データセットからの画像生成を実現するために，事前に学習された生成モデルを転送する手法を提案する．Noguchi と Harada と異なり，ジェネレータに Adaptive Point-wise 畳み込みを追加し，ジェネレータの隠れ層における最適な特徴チャネルの結合を学習することで事前知識を適応させる．小規模データセットを用いた実験を行い，提案手法は従来手法よりも品質が高い画像を生成でき，画像間の柔軟な補間が可能であることを示した．

[1] 電気通信大学 大学院情報理工学研究科 情報学専攻
[a)] takeda-m@mm.inf.uec.ac.jp
[b)] yanai@cs.uec.ac.jp

2. 関連研究

一般に，GAN は，高品質の画像を生成するために多くの学習サンプルを必要とする．Few-shot GAN では，事前学習のために ImageNet などの大規模な画像データセットを必要とするが，ファインチューニング時はより小さなデータセットを使用する．Few-shot GAN には 2 つのタイプがある: （1）事前学習済みドメインと Few-shot ドメインは同じであり，同じドメインの新しいカテゴリを追加するために Few-shot 学習に使用される画像は 3 つまたは 5 つだけである．（2）事前学習済みドメインと Few-shot ドメインが異なり，新しいドメインの Few-shot 学習には約 50 枚の画像が使用される．本研究では，（2）の場合に焦点を当てる．

(1) の主な研究には，Antoniou らの研究 [1]，Hong らの研究 [3], [4] と FS-GAN(Few-shot GAN)[12] がある．Antoniou らは，画像を条件とするジェネレータにランダムノイズを追加して，同じクラスからわずかに異なるサンプルを生成している．Hong らは，同じクラスの複数の条件付き画像からの情報を融合する手法を提案している．FS-GAN は，SVD を使用して事前学習済みモデルの畳み込み/全結合層の重みを因数分解し，適応のためのパラメータ空間を識別する．

（2）の代表的な研究は，Noguchi と Harada [10] の研究である．彼らは，事前に学習された生成モデルを様々なドメインのデータセットに適応させる方法を提案した．事前に学習された知識を効果的に使用するために，ジェネレータ内の畳み込み層のすべての重みは，ファインチューニング時に固定されたままになる．代わりに，バッチ正規化（BN）層のスケールおよびシフトパラメータのみが，事前学習に使用された学習データとは異なるドメインの小さなデータセットに適合される．SNGAN projection [8] で使用されるクラス条件付きバッチ正規化を使用して，スケールとシフトのパラメータが動的に変更され，小さな学習サンプルから様々な画像を生成できるようになった．BN パラメータのファインチューニングは，チャネルごとの特徴変調 [11] と見なされる．

本論文では，Noguchi と Harada と同様に，Few-shot の適応を通じて GAN のサンプル効率を改善することに焦点

を当てている．ただし，それらとは異なり，各チャネルのパラメータを調整するだけでなく，チャネル間の線形結合によってもチャネルを調整する．つまり，提案手法では，「Channel-wise」特徴変調を使用する従来研究と異なり，「Cross-channel」特徴変調を使用する．Cross-channel 特徴変調を有効にするために，ファインチューニング時に Adaptive Point-wise グループ化畳み込み層をジェネレータに導入し，すべての畳み込み層を固定してそれらのみを学習することを提案する．Channel-wise 変調を Cross-channel 変調に置き換えることで，事前に学習されたジェネレータを，小さなデータセットを持つ新しいドメインにさらに柔軟に適応させることができると考えられる．

3. 手法

本論文では，小規模データセットを使用して，事前に学習されたジェネレータを様々なドメインに適応させる方法を提案する．Noguchi と Harada [10] の研究を拡張し，より柔軟なドメイン適応を実現するために，クラス条件付きバッチ正規化の代わりに Adaptive Point-wise グループ化畳み込みを導入する．これは，Channel-wise 特徴変調の代わりに，Cross-channel 特徴変調を使用することを意味する．

3.1 Adaptive Point-wise 畳み込み

本論文では，チャネルの選択方法として，Depth-wise Separable 畳み込みの深さ方向の要素である Point-wise 畳み込みを使用する．Depth-wise Separable 畳み込みは，多くの効率的なニューラルネットワークアーキテクチャの重要な要素であり [5]，2 種類の畳み込み層で構成されている．最初の層は Depth-wise 畳み込み層であり，入力チャネルごとに 1 つの畳み込みフィルタを適用することによって軽量フィルタリングを実行する．2 番目の層は，Point-wise 畳み込みと呼ばれる 1×1 の畳み込み層であり，入力チャネルの線形結合を計算することによって新しい特徴を構築する．

チャネル選択の観点から，提案手法を簡単に分析する．pint-wise 畳み込みを適用し，入力チャネルの線形結合を計算することは，チャネル方向に全結合層を計算することと同等である．これは，次の畳み込み演算として表すことができる．

$$\mathbf{x}_{\text{Adapt}} = \mathbf{W}\mathbf{x} + \mathbf{b} \tag{1}$$

ここで，$\mathbf{x}$ は，入力特徴マップの特定のピクセル（または位置）上のすべてのチャネルにわたる特徴ベクトルを表し，$\mathbf{x}_{\mathbf{Adapt}}$ は対応するピクセルの出力ベクトルを表す．Point-wise 畳み込みの計算はピクセルの各ペア間で独立しているため，Point-wise 畳み込みは Point-wise 全結合層と見なすことができる．$\mathbf{W}$ と $\mathbf{b}$ は，それぞれ Point-wise 畳み込みの重み行列とバイアスベクトルを表す．入力チャネルの数が c_{in} で，出力チャネルの数が c_{out} の場合，それらは $\mathbf{W} \in \mathbb{R}^{c_{out} \times c_{in}}$ および $\mathbf{b} \in \mathbb{R}^{c_{out}}$ である．「adaptive」Point-wise 畳み込みの場合，$\mathbf{W}$ と $\mathbf{b}$ を，外部の全結合（FC）層で動的に生成することにより，適応的に変化させる．$\mathbf{W}$ を変更することは，チャネル要素の線形結合の重みを調整することを意味し，$\mathbf{b}$ を変更することは，Adaptive Point-wise 畳み込み層の直後の ReLU 活性化のしきい値を調整することを意味する．

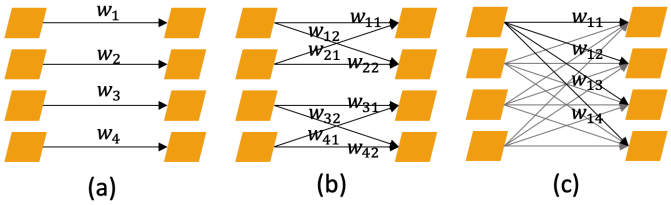

図 1 (a) チャネルごとの変調（BN）．(b) 制限された Cross-channel 変調．(c) 完全な Cross-channel 変調．

提案手法では，Adaptive Point-wise 畳み込みのパラメータ $\mathbf{W}$ と $\mathbf{b}$ は，潜在ベクトル z からの単一の全結合層によって生成される．これにより，新しいドメインに対してより柔軟なモデルの適応が可能になる．

3.2 学習パラメータの削減

Adaptive Point-wise 畳み込みにより，Noguchi と Harada [10] の研究で使用されるクラス条件付きバッチ正規化（BN）と比較して，チャネルの線形結合を使用したより複雑な表現が可能になる．ただし，重み $\mathbf{W}$ のパラメータ数は $c_{in} \times c_{out}$ であり，条件付き BN の重みの数 c_{out} よりもはるかに大きくなる． これにより，小規模データセットで学習する場合に過剰適合が発生する可能性がある．したがって，パラメータ数を減らす方法として，グループ化畳み込みのアイデアを Adaptive Point-wise 畳み込みに適用する．グループ化畳み込みでは，入力特徴マップがチャネル方向にグループ化され，各グループ間で畳み込み演算が適用される．グループ化されるチャネルの数が畳み込み層のチャネルの数と等しい場合，Depth-wise 畳み込みを表すことができる．Point-wise 畳み込みを Depth-wise Point-wise 畳み込みにすると，バッチ正規化層の γ の重みに相当するチャネルごとの重みを意味する．図 1 は，条件付き BN，Point-wise グループ化畳み込み，および Point-wise 畳み込みの違いを示しており，それぞれ，チャネルごとの変調 [11]，制限された Cross-channel 変調，および完全な Cross-channel 変調に対応する．本論文では，パラメータ数と特徴変調の柔軟性のバランスをとるために，制限付き Cross-channel 変調を採用している．

本論文では，$c_{in} \times c_{out}$ 重み行列を必要とする通常の Point-wise 畳み込みの代わりに，Point-wise グループ化畳み込みを使用することを提案する．Point-wise グループ化畳み込みの学習パラメータの数を，クラス条件付きバッチ正規化の 2，4，または 8 倍の学習パラメータに制限する．

3.3 学習

GAN の場合，ディスクリミネータは実際の学習画像と偽の生成画像を区別し，ジェネレータは敵対的学習を実行することによって現実的な画像を生成する．ただし，この方法は，学習サンプルがその分布を密に満たすことができるという事実に基づいている．学習サンプルの数が少ない場合，小規模データセットへの過剰適合が発生し，学習が不安定になる．したがって，VAE などの教師あり学習フレームワークで学習することが望ましい．教師あり学習でジェネレータを学習するために，Noguchi と Harada [10] に従って，学習画像にも対応する潜在ベクトル z を最適化する．提案するネットワークを図 2 に示す．学習中に，ターゲット画像からの距離関数としてモデル化された損失関数 L が最適化される．損失関数 L も，ピクセルレベルの距離である L1 損失と，セマンティックレベルの距離である知覚損失を使用するという点で，Noguchi と Harada に似ている．

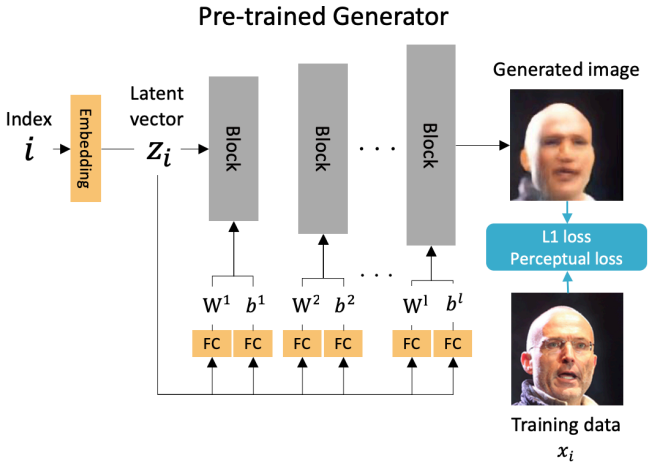

図 2　提案モデル． 図中の黄色いブロックは学習可能な層を表している．学習中に，潜在変数 z と Adaptive Point-wise 畳み込みのパラメータが更新され，L1 損失と知覚損失が最小限に抑えられる．

ジェネレータは，ImageNet などの大規模なデータセットで最初に事前学習されていると想定する．小規模データセットを使用したファインチューニングの段階では，最初に，すべてのバッチ正規化層の直後に，対応する全結合（FC）層を含む Adaptive Point-wise グループ化畳み込み層を挿入し，次にパラメータをファインチューニングする．Noguchi と Harada と同じように，潜在変数もファインチューニング中に更新される．

3.4　推論

推論中に，標準正規分布に基づいてランダムにサンプリングされたベクトル z をジェネレータに入力して，ランダムな画像を生成できる．ただし，ジェネレータは潜在ベクトルとスパース学習サンプル間の関係のみを学習するため，学習サンプルから遠く離れた z では精度が低下する．この問題を解決するために，Noguchi と Harada のように切断正規分布から z をサンプリングする．この手法は，truncation trick [2] として知られている．Noguchi と Harada と同じように，切り捨てのしきい値として 0.4 を使用した．

4.　実験

小規模データセットから画像を生成する際の提案手法の安定性を評価するために，いくつかの実験を行った．また，提案手法を既存手法と比較した．Noguchi と Harada [10] を参照し，ジェネレータには BigGAN [2]，画像サイズは 128x128 を使用した．すべての実験で，5 つの ResBlock で構成される ImageNet で事前学習された BigGAN-128 モデルを使用した．

4.1　データセットと評価指標

実験で使用したデータセットは，FFHQ データセット [6]，Oxford 102 flower データセット [9] の passion flower の画像，260 Bird Species データセット[*1]の African firefinch の画像と，Cars データセット [7] の BMW の画像である．実験で使用したドメインである「Human face」，「Passion flower」，「African firefinch」，「BMW」は ImageNet クラスに含まれていないため，ImageNet のドメインとは異なると見なすことができる．評価指標として，KMMD（Kernel Maxi-mum Mean Discrepancy）を使用して，生成された画像の品質を評価した．KMMD は，データセット内の画像の数が少ない場合でも安定した結果を生成できるという特徴がある．KMMD は，学習画像と生成画像からなる inception ネットワークで事前に学習された画像間のガウスカーネルを用いて計算された．KMMD が低いほど，品質は高くなる．

[*1] https://www.kaggle.com/gpiosenka/100-bird-species

表 1　パラメータの数と生成された画像の品質との関係

Model	Parameter ratio	Number of data	KMMD
[10]	1	25	2.966
		50	2.507
		100	2.509
Ours	2	25	2.944
		50	2.496
		100	2.493
	4	25	2.942
		50	2.491
		100	2.490
	8	25	**2.928**
		50	**2.485**
		100	**2.487**

4.2　ベースラインの比較

Noguchi と Harada [10] は，各チャネルのバッチ統計量のスケールとシフトのパラメータを更新しているが，本研究では，Adaptive Point-wise グループ化畳み込みのパラメータを更新することでドメインを適応させる．提案手法は Point-wise 畳み込みによって複数チャネルの活性化を混ぜることができるため，より柔軟な適応が可能になるという利点がある．Noguchi と Harada のパラメータ数を基準とし，提案手法において，グループ化の数を変えて，パラメータ数を 2 倍，4 倍，8 倍にしたときの生成画像の品質を比較する．ここで，Noguchi と Harada は，グループ化するチャンネル数を畳み込み層のチャンネル数と同じ値に設定した場合に等しい．FFHQ データセット [6] からサンプリングした 25 枚，50 枚，100 枚の画像を用いて，「Human face」ドメインの画像を生成した．実験結果を表 1 に示す．提案手法では，パラメータ数の増加に伴って品質が向上することがわかった．これは，Adaptive Point-wise 畳み込みが，複数のチャネルの活性化を組み合わせることで，特徴チャネルのバリエーションを増やしていることを示している．

4.3　追加のデータセットを使った実験

本実験では，4 つのデータセットすべてを使用して，手法をベースライン [10] と比較した．4 つのデータセットのそれぞれからサンプリングされた 25，50，および 100 の画像を使用した．提案手法では，ベースラインの 8 倍のパラメータを持つ Adaptive Point-wise グループ化畳み込みを使用した．図 3 は，3 種類のサンプル数において 4 つのデータセットを用いて生成された画像を示し，表 2 はそれらの定性的な結果を示している．

図 3 から，提案手法がベースラインよりも詳細な画像を生成できることが分かる．表 2 では，提案手法は，すべてのデータセットとすべての数の学習サンプルについて，ベースラインよりも高い品質を示した．これは，事前に学習された特徴チャネルが再利用され，チャネル間で組み合わされて，最適な表現を学習するためである．スケーリングとシフトによって各チャネル内の活性化を変更するベースラインと比較して，提案手法はより柔軟な表現を実現した．

図 4 は，ランダムに生成された 2 つの潜在ベクトル間の補間の結果を示す．学習データの量が少ないにもかかわら

表 2　定量的比較

Dataset	Model	Number of data	KMMD
Human face	[10]	25	2.966
		50	2.507
		100	2.509
	Ours	25	**2.928**
		50	**2.485**
		100	**2.487**
Passion flower	[10]	25	2.976
		50	2.977
		100	2.965
	Ours	25	**2.955**
		50	**2.960**
		100	**2.954**
African firefinch	[10]	25	2.965
		50	2.531
		100	2.532
	Ours	25	**2.937**
		50	**2.493**
		100	**2.506**
BMW	[10]	25	2.969
		50	2.522
		100	2.518
	Ours	25	**2.934**
		50	**2.487**
		100	**2.498**

ず，補間はベースラインの補間よりも明確かつスムーズで，安定している．

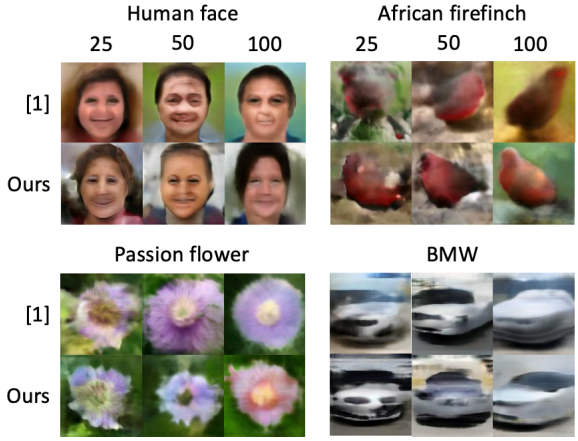

図 3　4 つのデータセットの定性的評価

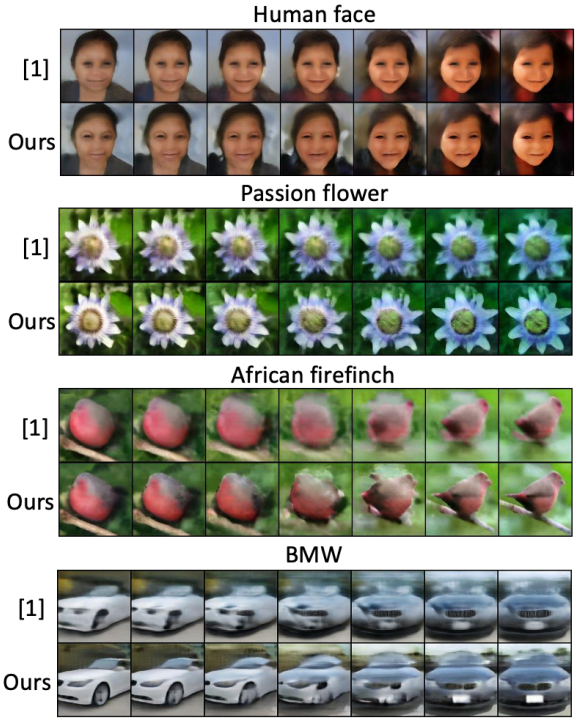

図 4　2 つの画像間の補間

5. Conclusions

本研究では，小規模データセットから画像を生成するシンプルで効果的な方法を提案した．事前に学習されたジェネレータの事前知識を使用して，Adaptive Point-wise グループ化畳み込みのパラメータを動的に生成する FC 層をファインチューニングすることで，通常必要とされるよりもはるかに少ない画像から新しい画像を生成できる．実験結果は，提案手法が，既存のベースラインと比較して，より小さな学習データセットを使用してより高品質の画像を合成できることを示している．これは，Cross-channel 変調が Channel-wise 変調よりも柔軟な適応を可能にすることを意味する．今後の研究ではに，より小規模なデータセットでより高品質の画像を生成することを検討する．

参考文献

[1] Antoniou, A., Storkey, A. and Edwards, H.: Data Augmentation Generative Adversarial Networks, *Proc. of International Conference on Learning Representation* (2017).

[2] Brock, A., Donahue, J. and Simonyan, K.: Large Scale GAN Training for High Fidelity Natural Image Synthesis, *Proc. of International Conference on Learning Representation* (2019).

[3] Hong, Y., Niu, L., Zhang, J. and Zhang, L.: MatchingGAN: Matching-based Few-shot Image Generation, *International Conference on Multimedia and Expo* (2020).

[4] Hong, Y., Niu, L., Zhang, J., Zhao, W., Fu, C. and Zhang, L.: F2GAN: Fusing-and-Filling GAN for Few-shot Image Generation, *Proc. of ACM International Conference Multimedia* (2020).

[5] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M. and Adam, H.: MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications, *arXiv:1704.04861* (2017).

[6] Karras, T., Laine, S. and Aila, T.: A Style-Based Generator Architecture for Generative Adversarial Networks, *Proc. of IEEE Computer Vision and Pattern Recognition* (2019).

[7] Krause, J., Stark, M., Deng, J. and Fei-Fei, L.: 3D Object Representations for Fine-Grained Categorization, *4th International IEEE Workshop on 3D Representation and Recognition (3dRR-13)*, Sydney, Australia (2013).

[8] Miyato, T., Kataoka, T., Koyama, M. and Yoshida, Y.: Spectral Normalization for Generative Adversarial Networks, *Proc. of International Conference on Learning Representation* (2018).

[9] Nilsback, M.-E. and Zisserman, A.: Automated flower classification over a large number of classes, *Proc. of Indian Conference on Computer Vision, Graphics and Image Processing* (2008).

[10] Noguchi, A. and Harada, T.: Image Generation From Small Datasets via Batch Statistics Adaptation, *Proc. of IEEE International Conference on Computer Vision* (2019).

[11] Perez, E., Strub, F., De Vries, H., Dumoulin, V. and Courville, A.: FiLM: Visual reasoning with a general conditioning layer, *AAAI* (2018).

[12] Robb, E., Chu, W.-S., Kumar, A. and Huang, J.-B.: Few-Shot Adaptation of Generative Adversarial Networks, *arXiv:2010.11943* (2020).