

フィーチャに基づく GAN の段階的学習方法の提案と評価

2017SE001 天野 将志 2017SE052 森野 佑哉 2017SE083 鈴木 裕介 2017SE101 綿貫 礼菜
指導教員 青山 幹雄

1 研究背景・課題

1.1 研究背景

GAN(Generative Adversarial Networks; 敵対的生成ネットワーク)[1]は 2 つのネットワークが競い合うように学習するという複雑な構造上, 学習の不安定性や学習時間の増大が課題となっている. 本稿で不安定性とはモード崩壊や過度な汎化を意味する. また, 深層学習, 特に画像認識問題においてフィーチャ(特徴量)に基づく段階的な学習の効果があることがわかっている[7].

本稿では, 学習データのフィーチャ(特徴量)に基づいた GAN の段階的な学習を可能にする学習方法を提案する.

1.2 研究課題

本稿では, 上記の研究背景を踏まえ, 次の 2 点を研究課題とする.

- (1) フィーチャに基づく GAN の段階的学習方法の提案
- (2) 提案方法の実データへの適用と評価

2 関連研究

2.1 GAN(敵対的生成ネットワーク)

機械学習技術の 1 つである GAN とは, 同時に訓練される生成器(Generator)と識別器(Discriminator)の 2 つのモデルからなる. 生成器はノイズを入力として偽データを生成し, 識別器は本物データと偽物データを見分けるように訓練される[1]. このように学習された GAN は本物に近い多様な画像を生成することができる. また, GAN には学習の不安定性や学習時間の増大などの問題があるが, それらを解決する目的で GAN の派生モデルが多く提案されている. 段階的に学習画像の解像度を上げる PGGAN(Progressive Growing of GANs)[3] や, CNN (Convolutional Neural Network) を生成器と識別器それぞれに取り入れた DCGAN(Deep Convolutional GAN)[8]などが存在する[4]. しかし, 現時点においてフィーチャに基づいた GAN の段階的学習を行っている方法は未確立である.

2.2 深層学習のフィーチャに基づく学習モデルの設計方法[7]

データのフィーチャに着目し, 段階的な深層学習モデルの設計方法が提案されている. データが持つフィーチャと学習済みモデルに対する予測誤差の関係に着目し, 段階的に追加する学習対象データを選択する. データのフィーチャに応じて予測誤差の推移が変化すると仮定し, フィーチャをもとにデータをグループ化し, フィーチャの量が最小のグループから学習を行い, データを追加していくことで段階的に学習する方法である. CNN モデルの学習精度の安定性の向上に有効であることが示されている.

2.3 フィーチャ抽出方法

画像におけるフィーチャは, 画像に含まれる物体の識別

や検出するための特徴の表現である. アルゴリズムや CNN を用いたフィーチャ抽出方法が提案されている.

2.4 k-means クラスタリング[6]

あらかじめ設定した k 個のクラスに分類するクラスタリング方法である. 最初に指定した数に, 重心を自動で探しながら値が近いフィーチャグループへと分割を行う.

2.5 Inception Score[10]

Inception Score は主に GAN の性能評価に使われる指標で, 以下の 2 つの観点を満たすほどスコアが高くなるように設計されている.

- (1) 生成画像が Inception モデルで識別しやすいこと
- (2) 識別されるラベル数が豊富であること

このように設計されている評価指標であるために, 生成画像の視認性が低かったり, モード崩壊を引き起こしたりする GAN の評価に適切である.

3 アプローチ

3.1 本研究へのアプローチ

GAN で扱うデータは主に画像データであり, 画像データは多くのフィーチャを保有している. 先行研究である深層学習のフィーチャに基づく学習モデルの設計方法において, フィーチャに着目した段階的な深層学習モデルの設計により CNN モデルにおける識別精度の向上, 学習収束速度に向上が見られた.

一方で, GAN の派生形である PGGAN は段階的に学習画像の解像度を上げることで生成画像の質の向上を可能にしている.

これらの点に着目した本研究のアプローチを図 1 に示す.

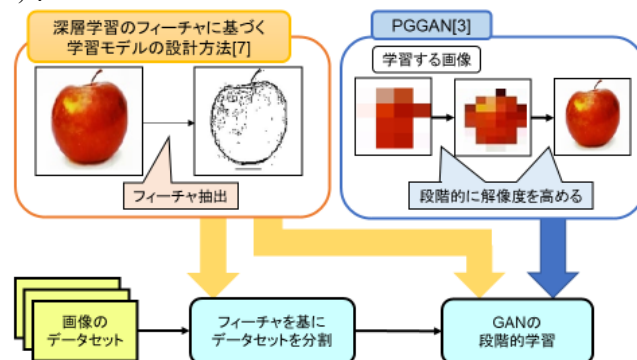


図 1 アプローチ

以上の経緯から本研究では, 画像から抽出されるフィーチャを利用することによる GAN の段階的学習方法の提案と評価を行う.

4 提案方法

4.1 提案プロセス

本研究の提案プロセスを図 2 に示す。

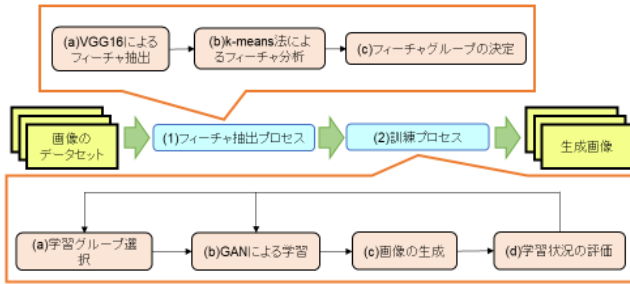


図 2 提案プロセス

(1) フィーチャ抽出プロセス

(a)VGG16[9]によるフィーチャ抽出, (b)k-means 法によるフィーチャ分析, (c)フィーチャグループの決定の 3 つのプロセスから成る。

(2) 訓練プロセス

(a)学習グループの選択, (b)GAN による学習, (c)画像生成, (d)データの保存の 4 つのプロセスから成る。

4.2 フィーチャ抽出方法

VGG16の全結合層の2層目以降を取り除くことでフィーチャ抽出を可能にし, 画像を入力した際に全結合層 1 層目で出力される要素数 4,096 の配列を取得する。その配列の要素のうち, 正の値の個数を取り出す。以降その値をフィーチャ総数と定義する。フィーチャ抽出の構造を図 3 に示す。

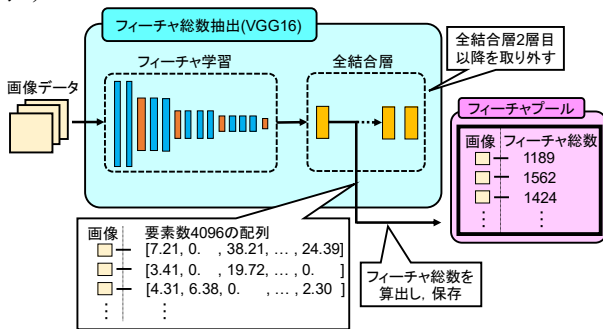


図 3 フィーチャ抽出

4.3 フィーチャ総数に基づく学習データの分類

データの分類はフィーチャ総数を基に k-means クラスタリングを行った。指定した個数に, 重心を自動で探しながらフィーチャ総数が近い画像を同じフィーチャグループへと分割を行う(図 4)。

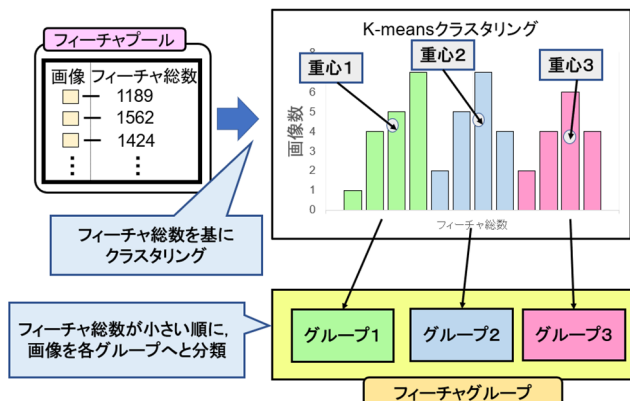


図 4 k-means クラスタリングによるフィーチャ分析

4.4 段階的学習の方法

学習データにフィーチャグループを追加するタイミングは, 現在の学習データを用いた学習が基準を達成した時として, その基準として Inception Score を用いることを提案する。学習中に生成画像に対して Inception Score を算出し, 最大値となった epoch 数の時をフィーチャグループの追加タイミングとする。

学習データにフィーチャグループを追加する順序は, 「各グループのフィーチャ総数の重心」を基に昇順または降順に追加する 2 パターン。以降この 2 パターンを昇順, 降順と呼ぶ。

5 プロトタイプの実装

提案方法を実行するプロトタイプの構成を図 5 に示す Python で実装し, モデルの学習は Keras を利用した。

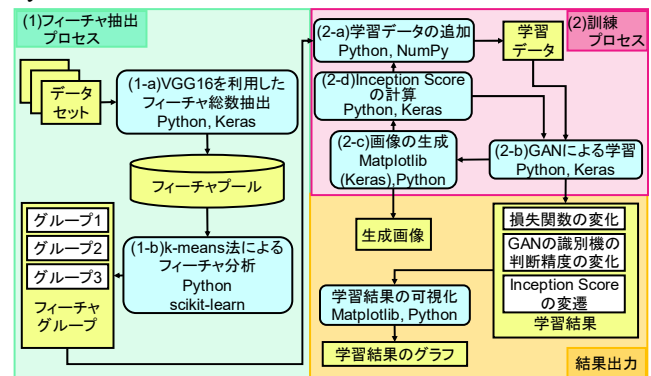


図 5 プロトタイプの構成

6 実データへの適用

6.1 適用対象

2 種類のデータセットをプロトタイプに適用する。

(1) Apple

Image Net[2]と Google 画像検索の 2 つから取得したデータを合わせた計 6,241 枚のデータセットである。Image Net から対象となるデータを取得することにし, Apple のラベルが含まれる画像を計 976 枚用意した(図 6)。Google 画像検索からは画像データ計 5,265 枚を取得した(図 7)。取得するに当たって, 一部りんごが写っていないもの, りんごが画像に占める面積の方が小さい画像等は手動で削除した。これらの画像を 128×128 にリサイズした。



図 6 Image Net から取得した対象データ

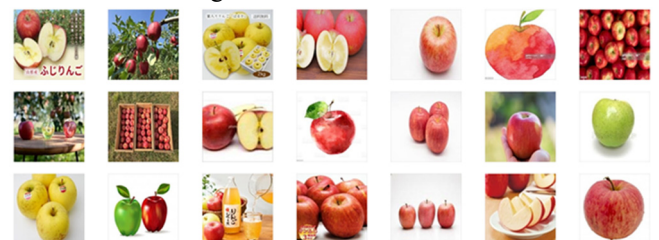


図 7 Google 画像検索から取得した対象データ

(2) CelebA

CelebA[5]から計 202,599 枚の画像を取得した. その内適用データには, “(21)男性”と“(32)笑顔”の属性を用いて“笑っている男性”の属性を持つ画像, 計 9,655 枚を抽出した(図 8). これらの画像を 128×128 にリサイズした.



図 8 CelebA のうち笑っている男性の属性を持つ画像

6.2 評価方法(実行シナリオ)

訓練プロセスにおいて, 以下の 3 点を考慮した 6 つのパターンを実行することで提案方法の有効性及び妥当性を評価した.

- (1) データセットは 6.1 節で示した “Apple”と“CelebA”の 2 種類を適用する.
- (2) データセットを 3 つに分割したフィーチャグループの追加順序を昇順および降順の 2 パターンとする.
- (3) 提案方法の評価のために, 学習データを分割しない従来の学習方法を適用する.

以上の観点を踏まえた全パターンを表 1 に示す.

表 1 実行シナリオ

	データ	学習モデル	追加順序
1	Apple	提案方法	降順
2			昇順
3		従来方法	—
4	CelebA	提案方法	降順
5			昇順
6		従来方法	—

GAN の学習が不安定であることを考慮し, 表 1 に示す計 6 パターンを各パターン 3 回ずつ適用した. その中で Inception Score の最大値が最高となった適用結果を各パターンの代表として比較, 評価を行った.

また, 提案方法では各段階につき 3,000epoch, 合計で 9,000epoch を実行し, それに伴い従来方法の学習モデルでは 9,000epoch を実行する.

7 適用結果の評価

7.1 フィーチャ抽出プロセスの結果

各データセットのフィーチャ総数の分布結果を図 9 として以下に示す.

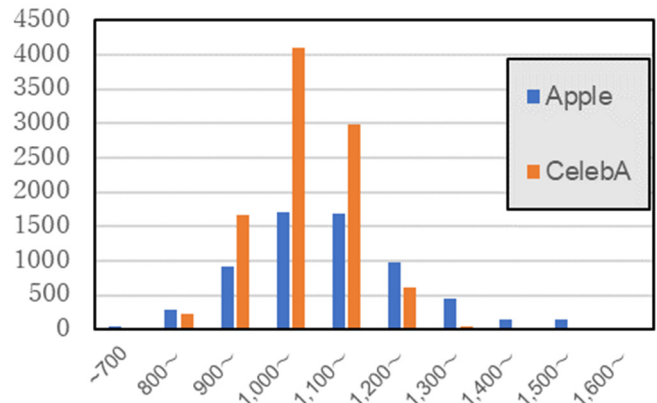


図 9 各データセットのフィーチャ総数分布

(1) Apple

フィーチャ総数を抽出してみると, 700~2,100 の間と広範囲に分布した. 最頻値としては 1,000~1,099 であり, そこを中心に正規分布に近い形状になるように分布している.

(2) CelebA

フィーチャ総数を抽出してみると, 700~1,400 の間と狭い範囲に分布した. 最頻値としては 1,000~1,099 であり, そこを中心に正規分布に近い形状になるように分布している.

2 つのデータセットを比較すると, どちらも最頻値 1,000~1,099 を中心とした正規分布に近い形状のグラフとなった. フィーチャ総数の分布範囲は Apple が広い一方, CelebA は狭い範囲に集中する結果となった.

上記の結果を基に “Apple”, “CelebA” に対して k-means クラスタリングを行った結果を表 2, 表 3 に示す.

表 2 Apple の k-means クラスタリングの結果

	グループ 1	グループ 2	グループ 3
重心	957.7	1120.3	1305.0
画像枚数	1,834 枚	2,979 枚	1,435 枚

表 3 CelebA の k-means クラスタリングの結果

	グループ 1	グループ 2	グループ 3
重心	964.6	1072.2	1177.9
画像枚数	2,558 枚	4,506 枚	2,589 枚

両方のデータセットにおいてフィーチャグループ 2 の画像枚数が他のグループの画像枚数の 1.5 倍以上になった.

7.2 訓練プロセスの結果

訓練プロセスの Apple の適用結果を表 4 と図 10 に, CelebA の適用結果を表 5 と図 11 に示す.

(1) Apple の適用結果

表 4 Apple の各試行の代表値の最終結果

	降順	昇順	従来方法
総学習 epoch 数	3,310	2,390	3,640
最終 Inception Score	3.74	3.34	3.96

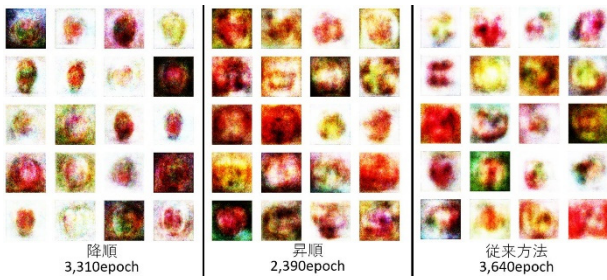


図 10 Apple の各パターンでの生成画像の比較

Inception Score は、従来の方法が最も高い結果となった。また、生成画像は(1)降順(2)従来方法(3)昇順の順に輪郭が明瞭な画像が生成された。

(2) CelebA の適用結果

表 5 各パターンの試行の代表値の最終結果

	降順	昇順	従来方法
総学習 epoch 数	4,900	3,430	7,260
最終 Inception Score	2.08	2.15	2.26

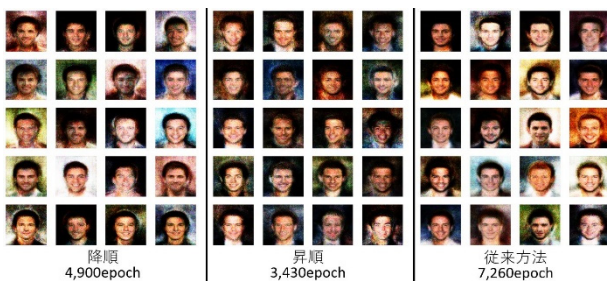


図 11 CelebA の各パターンでの生成画像の比較

Inception Score は、従来方法が最も高い結果となった。また、生成画像は降順と従来方法には明確な差が生まれず、昇順の場合のみ人物と背景が分離できていない画像や類似の画像が多数生成された。

8 考察

8.1 提案方法の妥当性

以下の3点において妥当性を評価した。

(1) データセットの分割方法

k-means クラスタリングを用いて学習データを分割したが、各フィーチャグループで画像枚数に差が生まれた。

(2) Inception Score の評価

Inception Score は生成画像によってばらつきが大きく、提案方法における学習データを切り替えるタイミングは Inception score が大きく上振れた時になるので、安定した評価基準としては不十分な可能性がある。

(3) フィーチャ総数の操作による GAN の学習への影響

生成画像は従来方法よりも降順に追加した場合は輪郭が明瞭化、昇順に追加した場合は彩度が向上した。フィーチャ総数を基にした学習データの操作によって特定のフィーチャが強調された画像が生成可能と思われる。

(4) フィーチャ総数の適応

フィーチャ抽出プロセスの結果より、多種多様な画像を含むデータセット程フィーチャ総数が幅広く分布していた。よって、類似のフィーチャをもつ画像は近似のフィーチャ総数を保持していると思われる。また、考察(3)よりフィーチャ

総数の操作で生成画像を操作できることからフィーチャ総数は特定の画像のフィーチャを表せていると推測でき、提案方法で利用することは妥当である。

8.2 実データへの有効性

Apple では提案方法と従来方法で生成画像の差が大きく表れた一方で、CelebA では生成画像の品質の差が認められなかった。よって、提案方法は多様性のあるデータセットを学習する際に特定のフィーチャが強調された画像を生成できる点で有効的である。

一方で、Inception Score の評価は従来方法が提案方法よりも高い値になった。これは従来方法ではデータセットの分割をしていないために、一度の学習で多様な画像を学習できたためと解釈できる。

9 今後の課題

(1) データセットの分割方法の改善

k-means クラスタリング法よりも各フィーチャグループ内の画像枚数が均一に分割できる方法へと改善する。

(2) Inception Score による評価の改善

学習 epoch 数が近い Inception Score の平均値を評価指標へ変更や、他の評価指標を検討する。

(3) 他のデータセットに対しての適用

画像枚数が膨大なデータセットにも提案方法を適用し、評価を行う。

10 まとめ

本稿では、フィーチャに基づき学習データを選択する GAN の段階的学習方法を提案した。提案方法のプロトタイプを実装し、複数の実際の画像データに適用し、その有効性と妥当性を評価した。多様な学習データを用いた GAN の学習において、フィーチャを操作することによって生成画像を操作できる可能性があることを示した。

参考文献

- [1] I. J. Goodfellow, et al., Generative Adversarial Nets, Proc. of NIPS 2014, Dec. 2014, pp. 2672-2680.
- [2] Image Net, <http://image-net.org/>
- [3] T. Karras, et al., Progressive Growing of GANs for Improved Quality, Stability, and Variation, Feb. 2018, arXiv:1710.10196[cs.NE], pp.1-26.
- [4] J. Langr, et al., GANs in Action: Deep learning with Generative Adversarial Networks, Manning Publications, 2019 [大和田 茂(監訳), 実践 GAN~敵対的生成ネットワークによる深層学習~, マイナビ出版, 2020].
- [5] Z. Liu, et al, Large-scale CelebFaces Attributes (CelebA) Dataset, <http://mmlab.ie.cuhk.edu.hk/projects/CelebA.html>
- [6] J. MacQueen, Some Methods for Classification and Analysis of Multivariate Observations, Proc. of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Vol.1, University of California Press, 1967, pp. 281-297.
- [7] 太田 龍之介 ほか, 深層学習のフィーチャに基づく学習モデル設計方法の提案と評価, SES (ソフトウェアエンジニアリングシンポジウム) 2019 論文集, 情報処理学会, Aug. 2019, pp. 162-170.
- [8] A. Radford, et al., Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks, Jan. 2016, arXiv:1511.06434[cs.LG], pp.1-16.
- [9] K. Simonyan, et al., Very Deep Convolutional Networks for Large-Scale Image Recognition, Apr. 2015, arXiv:1409.1556 [cs.CV], pp.1-14.
- [10] S. Tim, et al., Improved Techniques for Training GANs, Proc. of NIPS 2016, Dec. 2016, pp. 2234-2242.