深層生成モデルを活用したデータ増幅に関する研究

TR18001 足立浩規　　指導教授：藤吉弘亘

1. はじめに

データ増幅は訓練データに対して幾何変化等を施すことで，訓練データの量及びバリエーションを増幅させる技術である．データ増幅は，過学習の防止や汎化性能の向上に貢献することが知られており，深層学習の学習や推論の際に広く利用されている．深層生成モデルである Generative Adversarial Networks (GAN) [1] は，訓練データに含まれない新たな画像を生成できるため，データ増幅への活用が期待できる．GAN は，訓練データと遜色ない画像を生成できるため，画像全体の高精細化が可能である．一方，識別モデルである Convolutional Neural Network (CNN) は，画像中の識別領域に着目してクラス識別する [2]．これらの関係より，識別領域に着目した画像を生成することができれば，生成画像はデータ増幅に効果的であると考えられる．そこで本研究では，識別対象である物体の特徴的な領域を高精細に生成することが可能な Discriminator-Driven Attention-Aware GAN (D2A2-GAN) を提案する．提案手法は，識別に有効な領域に着目した画像生成を行うため，画像全体の品質は従来手法に劣る．しかしながら，データ増幅として提案手法の生成画像を使用することで，従来手法の生成画像よりも精度向上に貢献することが期待できる．

2. 関連研究

Generative Adversarial Networks (GAN) [1] は，潜在変数から画像を生成する Generator と，生成画像または訓練画像を与えて尤度を出力する Discriminator の 2 つのネットワークで構成される生成モデルである．Generator は，訓練画像の分布に含まれるような画像を生成する．つまり，訓練画像と生成画像の分布を近づけることが目的である．一方で，Discriminator は，生成画像と訓練画像のどちらが入力されたかを正確に判別する．つまり，2 つの分布の境界を正確に学習することが目的である．GAN の目的関数は，式 (1) として表される．

$$\min_G \max_D V(D,G) = \mathbb{E}_{\boldsymbol{x}\sim p_r(\boldsymbol{x})}[\log D(\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{z}}(\boldsymbol{z})}[\log(1-D(\hat{\boldsymbol{x}}))] \quad (1)$$

ここで，$\hat{\boldsymbol{x}}$，$\boldsymbol{x}$ はそれぞれ生成画像，実画像である．GAN は，式 (1) の目的関数を最適化することによって，事前分布との関数近似を行わずに，訓練画像と遜色ない画像を生成可能である．しかしながら，識別モデルである Convolutional Neural Network (CNN) は，識別領域に着目してクラス識別を行う．そのため，生成画像を用いたデータ増幅にも利用することで効果的なデータ増幅となる．

Attention Branch Network (ABN) [2] は，CNN が識別時に着目する領域の視覚化が可能な手法の 1 つである．また，ABN は，注視領域を特徴マップへ反映することで，識別精度を向上させることができる．

3. 提案手法

提案手法である Discriminator-Driven Attention-Aware GAN (D2A2-GAN) は，識別領域に着目した画像生成を行うために ABN のアイデアを導入する．提案手法は，Generator へ潜在変数 $\boldsymbol{z}$ 及び，条件ベクトル $\boldsymbol{y}$ を与えて指定したクラスの画像を生成する．Discriminator は，Feature extractor，Attention branch，Adversarial branch の 3 つから構成する．図 1 に示すように，Adversarial branch は，Attention branch で獲得した Attention map を特徴マップに反映させて尤度を出力する．提案手法の目的関数は，式 (1) にクラス識別誤差を加算したものである．この時，実画像は Generator に関与しないため，Generator は生成画像の識別誤差のみを考慮する．クラス識別誤差は，Softmax 関数とクロスエントロピー誤差を用いて算出する．

3.1 Feature extractor

Feature extractor は，画像データを与えて畳み込み処理を介して任意のサイズの特徴マップを出力する．Feature extractor 内におけるデータのダウンサンプリングは，データ全体の情報を保持するために Average pooling を使用する．活性化関数は，最終層を除いてすべて LeakyReLU を使用する．Feature extractor の出力は，注視領域を強調するために Attention map と演算をするが，特徴マップに負の値が混在するとうまく強調されない．そのため，最終層の活性化関数は，ReLU を使用する．

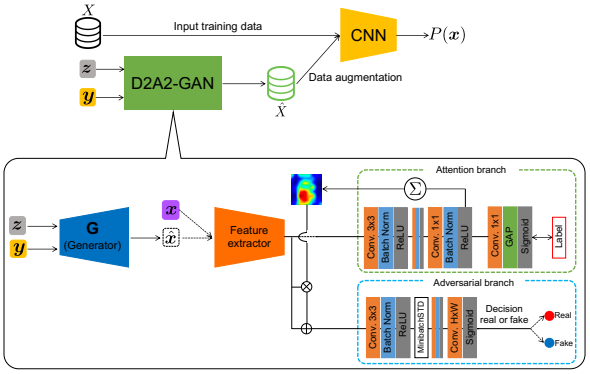

図 1：D2A2-GAN の構造

3.2 Attention branch

Attention branch は，ABN と同様に Feature extractor が出力する特徴マップを受け取り，注視領域の生成と Global Average Pooling (GAP) を介してクラス識別を行う．独立した畳み込み層を使用して Attention map を生成すると，ドメイン分類誤差を反映した注視領域にすることができる．そこで，提案手法では以下の式に示すように [0, 1] で正規化した特徴マップを注視領域として扱う．

$$M(\boldsymbol{x}) = \frac{f_{sum} - \min(f_{sum})}{\max(f_{sum}) - \min(f_{sum})} \quad (2)$$

ここで，f_{sum} は，特徴マップをチャネル方向に総和したものである．これにより，パラメータを保持しないため，クラス識別の結果のみを考慮した注視領域が獲得できる．

3.3 Adversarial branch

Adversarial branch へ入力する特徴マップは，まず Attention map との要素積を求めて注視領域を強調する．さらに，注視領域以外の値も考慮するために特徴マップを加算する．また，Generator によるモード崩壊を防止するために，ミニバッチ内の標準偏差マップを特徴マップへ結合する Minibatch standard deviation (Minibatch STD) を導入する．Adversarial branch は，最終層で特徴マップと同じサイズのフィルタを使用してダウンサンプリングをしたのち，Sigmoid 関数を適用して尤度を出力する．

4. 評価実験

本実験では，Street View House Number (SVHN) と CIFAR10 を用いて提案手法と従来手法で画像生成を行う．従来手法には，DCGAN と Auxiliary Classifier GAN (ACGAN) を用いる．生成画像の画質評価は，Fréchet inception distance (FID) と Inception score (IS) を用いる．ここで，IS は一般的に自然画像に適した評価指標であるため，CIFAR10 のみを評価する．次に，生成画像をデータ増幅させて ResNet18 を学習して評価する．訓練データを {100, 1000}，生成画像による増幅数を {0, 100, 1000, 10000, 50000} と変化させた際の精度を比較する．従来手法の生成画像に対する教師信号は，Generator に入力するクラスラベルとする．一方，提案手法の生成画像に対する教師信号は，Hard target と Soft target の 2 種類を学習に用いる．Hard target は，従来手法と同様に Generator へ入力するクラスラベルを付与したものである．Soft target は，Attention branch が出力するクラス確率を疑似ラベルとして付与したものである．

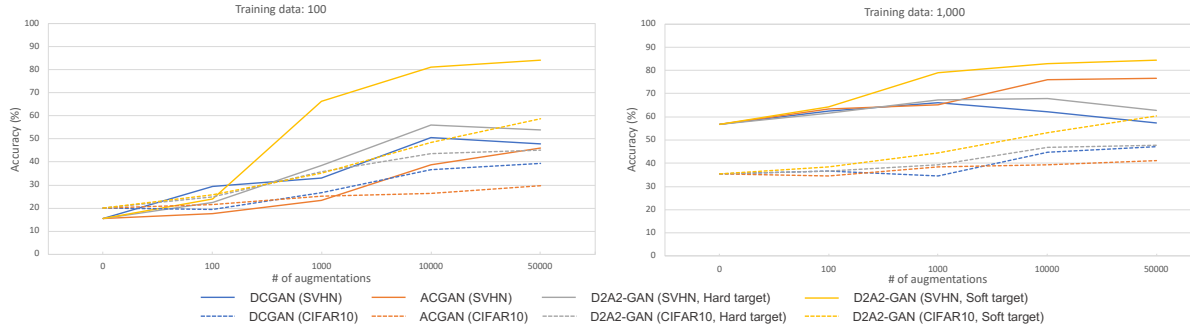

図 2：従来手法と D2A2-GAN の識別精度の比較

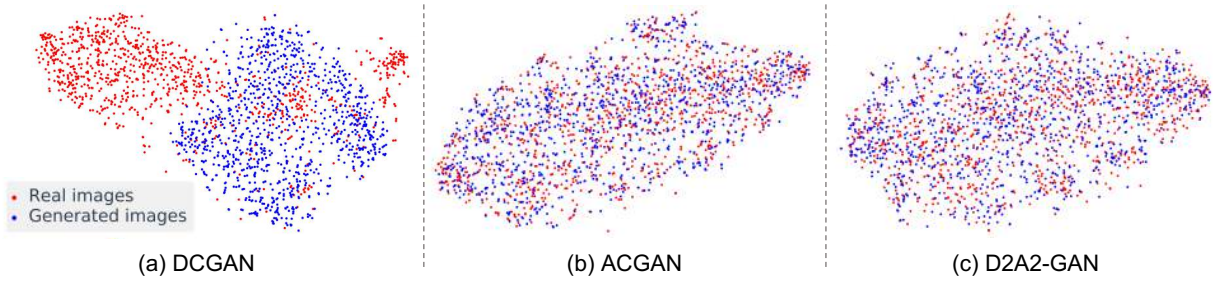

図 4：生成画像と実画像の分布 (SVHN)

表 1：定量的評価

	DCGAN	ACGAN	D2A2-GAN
IS (CIFAR10)	3.12±0.03	4.27±0.03	4.51±0.05
FID (CIFAR10)	63.0	28.1	21.1
FID (SVHN)	75.7	15.2	15.3

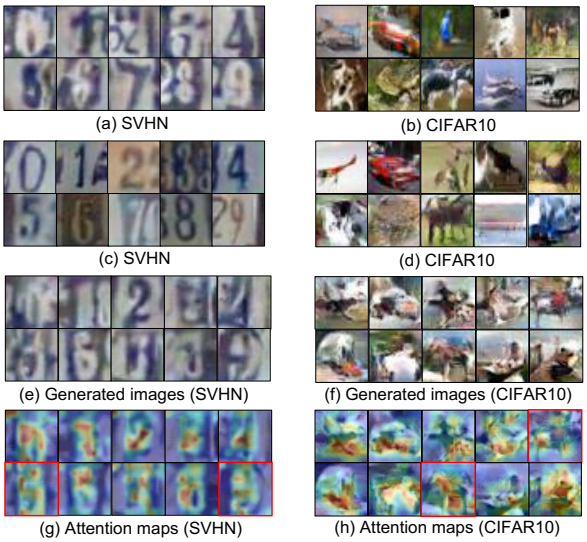

図 3：生成画像と注視領域の例

4.1 画質と識別精度の比較

表 1 に定量的画質評価を示す．IS は提案手法のスコアが最も高く，標準偏差が大きいことからバリエーション豊富な画像を生成できたと言える．また，CIFAR10 で学習した際の FID のスコアは，提案手法が最も良い結果であることが確認できる．一方，SVHN の結果は従来手法の ACGAN と同等のスコアであった．特に DCGAN のスコアと比較すると，CIFAR10，SVHN 共にスコア大幅に向上していることが確認できる．図 2 に各手法でデータ増幅した際の識別精度を示す．SVHN の結果 (実線) に着目すると，訓練データが 100 の時，1,000 の時共に教師信号を Soft target として扱った際の精度が最も良いことが確認できる．また，CIFAR10 (破線) の結果においても同様に Soft target とした際の結果が高精度である．

4.2 生成画像の視覚的評価

図 3(a)(c)(e) に示す SVHN の生成画像は，全ての手法で視覚的に数字と認識することができる．また，図 3(b)(d)(f) の CIFAR10 から，図 3(f) の提案手法を除いて画像内に映るオブジェクトの種類を認識することができるが，図 3(f) の提案手法は視覚的に認識することが困難である．しかしながら，図 2 の結果を踏まえると，従来手法は画像が鮮明に生成できたとしても識別精度は大幅に向上しない．一方，提案手法は生成画像は崩れていても，識別領域に着目した画像生成であるため，識別精度への悪影響は少ない．

図 3(g) に示す SVHN の Attention map は，各数字の特徴的な領域に注目していることが確認できる．特に赤枠で囲った例は，数字のシルエットと同じ注視領域が獲得できたことを示している．図 3(h) の CIFAR10 の Attention map は，正確な領域に注視しているか確認が困難である．しかしながら，赤枠の例では，馬を生成した際の注視領域を表しており，馬のシルエットや特徴的な形状になっていることが確認できる．

4.3 生成画像の分布

図 4 に示す分布は，提案手法と従来手法の生成画像と実画像それぞれ 1,000 データの分布である．図 4(a) は，実画像が大局的に分布しているが，生成画像は局所的な領域に集合した分布となっており，実画像の分布とは乖離している．図 4(b)(c) は，実画像の間を補間するように生成画像が分布していることが確認できる．提案手法の分布は，実画像と大幅に分離していないことから，鮮明でない画像であっても学習に適したデータとなっている．また，実画像の分布のスパースな部分を補間しているため，識別に有効なデータを増幅できたと言える．

5. おわりに

本研究では，識別時の注視領域に着目した画像生成が可能な Discriminator-Driven Attention Aware GAN を提案した．評価実験より，提案手法の生成画像は Inception score で最も良いスコア，Fréchet inception distance (FID) は従来手法と同程度のスコアであった．また，識別学習のデータ増幅として使用した際には，従来手法より効果的であることを確認した．今後は，獲得した注視領域を Generator の画像生成への活用を検討する．

参考文献

[1] I. Goodfellow, *et al.*, “Generative Adversarial Nets”, NIPS, 2014.

[2] H. Fukui, *et al.*, “Attention Branch Network: Learning of Attention Mechanism for Visual Explanation”, CVPR, 2019.

研究業績

[1] H. Adachi, *et al.*, “Facial Image Generation by Generative Adversarial Networks using Weighed Condtions”, VISAPP, 2019.

(他 4 件)