

修士論文概要書

Master's Thesis Summary

Date of submission: 01/25/2021 (MM/DD/YYYY)

専攻名 (専門分野) Department	情報理工・ 情報通信専攻	氏名 Name	大山 優香	指導 教員 Advisor	渡辺 裕 印 Seal
研究指導名 Research guidance	オーディオビジュ アル情報処理研究	学籍番号 Student ID number	CD 5119F015-4		
研究題目 Title	和菓子画像のための敵対的生成ネットワーク Generative Adversarial Networks for Japanese Sweets Images				

1. はじめに

和菓子は五感の芸術であると言われており、表現する題材や使用した材料、制作する職人によってさまざまなデザインがある。しかし、新たなデザインを考案することは容易ではなく、マンネリ化が問題となっている。

本研究は和菓子職人のデザイン考案のマンネリ化解消のため、Generative Adversarial Networks (GAN)を用いて和菓子画像を生成し、新しい和菓子のデザインを生み出すことが目的である。GANは2014年にIan J. Goodfellowによって考案された生成モデルを訓練させるためのモデリング手法である [1]。近年GANを用いた深層学習による画像生成は、様々な分野で行われている。

本研究ではGANの発展形であるDiscovery GAN (DiscoGAN)、及びAutoencoderの発展形であるDeforming Autoencoders (DAE)を組み合わせた和菓子画像を生成するGANを提案する。また実験により提案手法の有効性を検証する。

2. Deforming Autoencoders

Deforming Autoencoders (DAE)は、2018年にShuらが発表したAutoencoderを基本とした画像生成技術である [2]。一つのEncoderと二つのDecoderから構成され、入力画像をテクスチャと形状に分解するように学習する。

3. Discovery GAN

Discovery GAN (DiscoGAN)は、2017年にKimらが発表したGANを基本とした画像スタイル変換技術である [3]。二つのGeneratorと二つのDiscriminatorから構成される。2種類の異なるドメインのペアリングされていない画像を学習することで、入力画像を他方のドメインの画像に変換する。

4. 提案手法

本研究では、DiscoGANを用いて和菓子画像を生成する際に形状の多様性が乏しいという問題点を解決するために、DiscoGANのGeneratorにDAEを組み合わせた形状情報を用いる画像スタイル変換手法を四つ提案する。

提案手法1のGeneratorの構造を図1に示す。DAEのEncoderとDecoderで入力画像をテクスチャと形状に分解し、ドメイン変換も行う。

提案手法2のGeneratorの構造を図2に示す。DAEのEncoderとDecoderで入力画像をテクスチャと形状

に分解し、DiscoGANのGeneratorでテクスチャをドメイン変換した後に形状と空間変形する。

提案手法3のGeneratorの構造を図3に示す。提案手法2の構造のうち、DAEのEncoderとDecoderを事前に学習する。

提案手法4のGeneratorの構造を図4に示す。DiscoGANのGeneratorで入力画像をドメイン変換した画像と、DAEのEncoderとDecoderで分解した入力画像の形状を空間変形する。

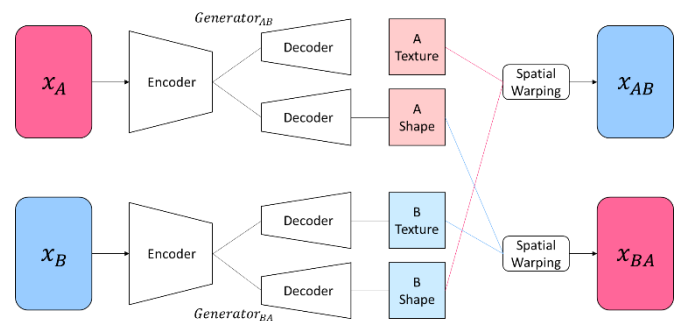


図1 手法1のGeneratorの構造

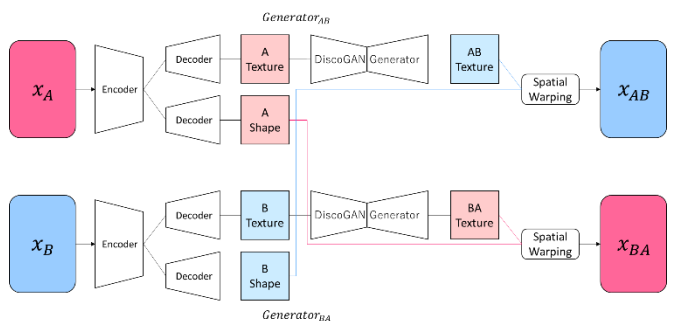


図2 手法2のGeneratorの構造

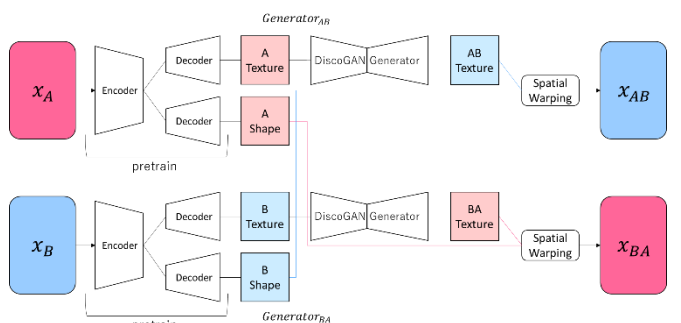


図3 手法3のGeneratorの構造

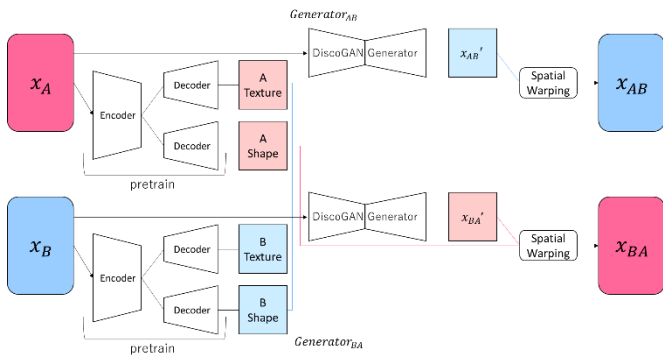


図 4 手法 4 の Generator の構造

5. 実験及び評価

DiscoGAN と四つの提案手法で和菓子画像を生成し、和菓子画像判別 CNN [4], Multi-scale Structural Similarity (MS-SSIM) [5], Fréchet Inception Distance (FID) [6]を用いて評価した。

学習には和菓子画像データセットと素材画像データセットを用いた。DiscoGAN を学習して生成した和菓子画像の例を図 5 に示す。提案手法 1 の Generator を学習して生成した和菓子画像の例を図 6 に示す。提案手法 2 の Generator を学習して生成した和菓子画像の例を図 7 に示す。提案手法 3 の Generator を学習して生成した和菓子画像の例を図 8 に示す。提案手法 4 の Generator を学習して生成した和菓子画像の例を図 9 に示す。

また、生成画像を CNN, MS-SSIM, FID を用いて評価した結果を表 1 に示す。和菓子判別 CNN の正解率は提案手法 4 が最も高く、93.7%である。MS-SSIM の値は提案手法 4 が最も低く、0.222 である。FID の値は提案手法 4 が最も低く、168.82 である。したがって、提案手法 4 が最も和菓子画像の生成に適していると分かった。

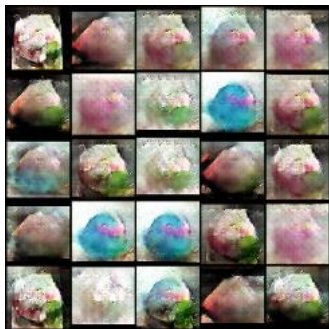


図 5 DiscoGAN

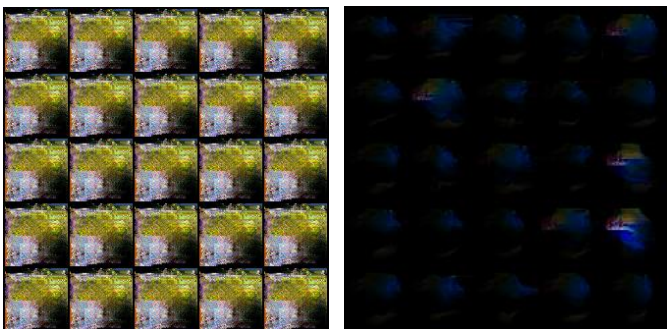


図 6 提案手法 1

図 7 提案手法 2

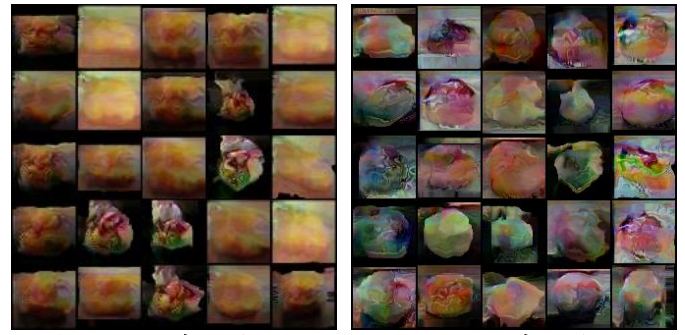


図 8 提案手法 3

図 9 提案手法 4

表 1 生成和菓子画像の評価

	CNN	MS-SSIM	FID
DiscoGAN	50.4	0.517	261.55
提案手法 1	0.00	1.000	421.94
提案手法 2	1.80	0.904	354.27
提案手法 3	79.2	0.501	191.29
提案手法 4	93.7	0.222	168.82

6. おわりに

本研究では、DiscoGAN を用いて和菓子画像を生成する際の形状の多様性が乏しいという問題点を解決するために、DAE と DiscoGAN を組み合わせた形状情報を用いる画像スタイル変換手法を提案した。また DiscoGAN および提案手法を用いて画像を生成し、評価した。評価の結果、提案手法 4 が最も和菓子画像の生成に適していると分かった。素材画像データセットのデータクリーニングを行うことで結果を向上できると考えられる。

参考文献

- [1] I. J. Goodfellow, J. P. Abadie, M. Mirza, B. Xu, D. W. Farley, S. Ozair, A. Courville and Y. Bengio, "Generative Adversarial Nets," NIPS, pp.1-9, 2014.
- [2] Z. Shu, M. Sahasrabudhe, A. Guler, D. Samaras, N. Paragios, I. Kokkinos, "Deforming Autoencoders: Unsupervised Disentangling of Shape and Appearance," arXiv preprint arXiv:1806.06503v1, 2018.
- [3] T. Kim, M. Cha, H. Kim, J. K. Lee and J. Kim, "Learning to Discover Cross-Domain Relations with Generative Adversarial Networks," arXiv preprint arXiv:1703.05192v2, 2017.
- [4] 大山優香, 渡辺裕, "GAN による和菓子画像の生成と評価," 電子情報通信学会総合大会, D-12-40, 2019.
- [5] Z. Wang, E. P. Simoncelli and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," Proc. IEEE Asilomar Conf. on Signals, Systems, and Computers, pp. 1398-1402, 2003.
- [6] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler and S. Hochreiter, "GANs trained by a two time-scale update rule converge to a local Nash equilibrium," NIPS, pp.6626-6637, 2017.

2020 年度

早稲田大学大学院基幹理工学研究科情報理工・情報通信専攻 修士論文

和菓子画像のための敵対的生成ネットワーク

Generative Adversarial Networks

for Japanese Sweets Images

大山優香

(5119F015-4)

提出日：2021.1.25

指導教員：渡辺裕教授

研究指導名：オーディオビジュアル情報処理研究

目次

第 1 章	序論.....	1
1.1	研究の背景.....	1
1.2	本研究の目的.....	1
1.3	本論文の構成.....	1
第 2 章	Autoencoder.....	3
2.1	まえがき.....	3
2.2	Autoencoder.....	3
2.3	Deforming Autoencoders.....	3
2.4	むすび.....	4
第 3 章	Generative Adversarial Networks.....	5
3.1	まえがき.....	5
3.2	Generative Adversarial Networks.....	5
3.3	Discovery GAN.....	5
3.4	むすび.....	6
第 4 章	評価指標.....	7
4.1	まえがき.....	7
4.1	和菓子画像判別 CNN.....	7
4.2	Multi-scale Structural Similarity.....	7
4.3	Fréchet Inception Distance.....	7
4.1	むすび.....	7
第 5 章	予備実験.....	8
5.1	まえがき.....	8
5.2	使用するデータセット.....	8
5.3	DiscoGAN の実行.....	11
5.4	評価.....	12
5.5	考察.....	12
5.6	むすび.....	12
第 6 章	提案手法.....	13
6.1	まえがき.....	13
6.2	提案手法 1.....	13
6.3	提案手法 2.....	14
6.4	提案手法 3.....	15
6.5	提案手法 4.....	16

6.6	むすび.....	17
第7章	実験結果.....	18
7.1	まえがき.....	18
7.2	提案手法1の実行.....	18
7.3	提案手法2の実行.....	19
7.4	提案手法3の実行.....	20
7.5	提案手法4の実行.....	22
7.6	評価.....	23
7.7	考察.....	23
7.8	むすび.....	24
第8章	結論と今後の課題.....	25
8.1	結論.....	25
8.2	今後の課題.....	25
	謝辞.....	26
	参考文献.....	27
	図一覧.....	28
	表一覧.....	29
	研究業績.....	30

第1章 序論

1.1 研究の背景

和菓子は五感の芸術であると言われている。これはすべての和菓子の見た目や香り、味、食感がそれぞれ異なり個性を感じられるからである。また同じ植物、同じ風景を題材にした和菓子であっても、使用した材料や制作する職人によってさまざまな表現が存在する。しかし、和菓子において独創的かつ分かり易いデザインを新たに考案することは容易でないと考えられる。また現在若者の和菓子離れやコンビニスイーツの普及などから和菓子職人が減少し、和菓子業界ではデザインのマンネリ化が問題となっている。

Generative Adversarial Networks (GAN)は2014年にIan J. Goodfellowによって考案された生成モデルを訓練させるためのモデリング手法である [1]。近年GANを用いた深層学習による画像生成は様々な分野において行われている。さらに、GANを改良、発展させることで要求に合った精度の高い画像を生成することができる。したがってGANを用いることで、和菓子のデザインを考案することができると考えられる。

そこで本研究では、GANの発展形であるDiscovery GAN (DiscoGAN)、及びAutoencoderの発展形であるDeforming Autoencoders (DAE)を組み合わせた和菓子画像を生成するGANを提案する。また実験により提案手法の有効性を検証する。

1.2 本研究の目的

本研究は和菓子職人によるデザイン考案のマンネリ化を解消するために、GANを用いて和菓子画像を生成し、既存の発想を超えた新しい和菓子を生み出すことが目的である。現在画像処理の観点から和菓子のデザインを考案する研究は乏しく、職人の発想力に依存している。敵対的生成ネットワークを用いて和菓子のデザインを生み出すことで、マンネリ化を防ぐとともに人間には思いつかない、新たなデザインの和菓子も生成できるようになる。本研究を通して画像処理の観点から日本の食文化の発展に貢献したい。

1.3 本論文の構成

本論文の構成を以下に示す。

第1章は本章であり、本研究の背景、目的について述べる。

第2章ではAutoencoderの概要や仕組みについて述べる。また関連研究であるDeforming Autoencoders (DAE)の概要、仕組みについて述べる。

第3章ではGenerative Adversarial Networks (GAN)の概要や仕組みについて述べる。また関連研究であるDiscovery GAN (DiscoGAN)の概要、仕組みについて述べる。

第4章では本研究に用いた評価指標について述べる。

第 5 章では DiscoGAN を用いて和菓子画像を生成する予備実験について述べる。

第 6 章では和菓子画像を生成する GAN を提案する。

第 7 章では提案手法を用いた和菓子画像の生成結果，その評価について述べる。また実験結果について考察を述べる。

第 8 章では本論文の結論と今後の課題について述べる。

第2章 Autoencoder

2.1 まえがき

本章では Autoencoder の概要や仕組みについて述べる。また関連研究である Deforming Autoencoders (DAE)の概要，仕組みについて述べる。

2.2 Autoencoder

Autoencoder は 2006 年に Hinton らが発表した，ニューラルネットワークを用いてベクトルの次元を削減する手法である [2]。Autoencoder の構造を図 2.1 に示す。Autoencoder は Encoder と Decoder という二つのネットワークから構成される。Encoder は入力された高次元ベクトル x を低次元ベクトルに変換し，Decoder は変換された低次元ベクトルを元のベクトル x に復元するように学習する。

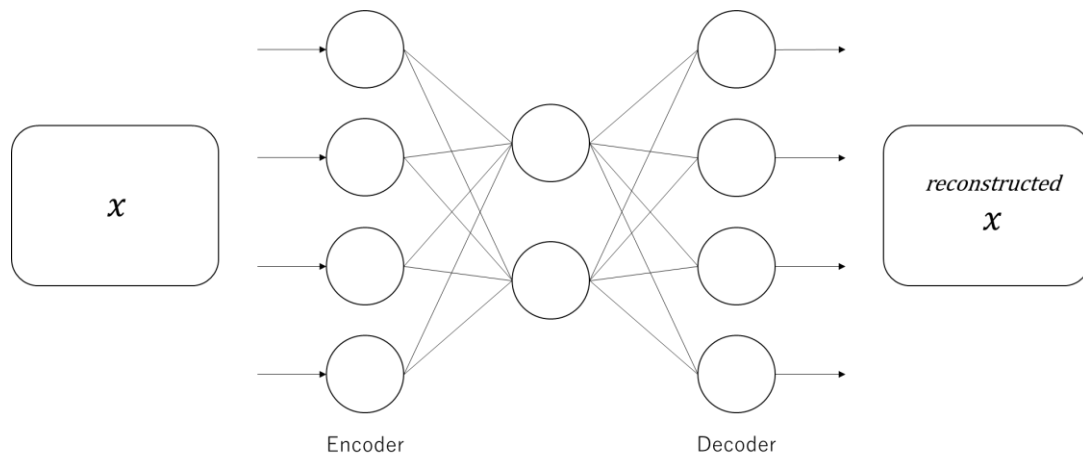


図 2.1 Autoencoder の構造

2.3 Deforming Autoencoders

Deforming Autoencoders (DAE)は，2018 年に Shu らが発表した，Autoencoder を基本とした画像生成技術である [3]。DAE の構造を図 2.2 に示す。DAE は一つの Encoder と二つの Decoder から構成される。Encoder は入力された高次元ベクトル x を低次元ベクトルに変換する。変換された低次元ベクトルをテクスチャと形状に分解し，Decoder は分解されたテクスチャベクトルと形状ベクトルのそれぞれを高次元ベクトルに変換する。二つの高次元ベクトルに空間変形を行い，再構成されたベクトルが元のベクトル x に復元するように学習する。

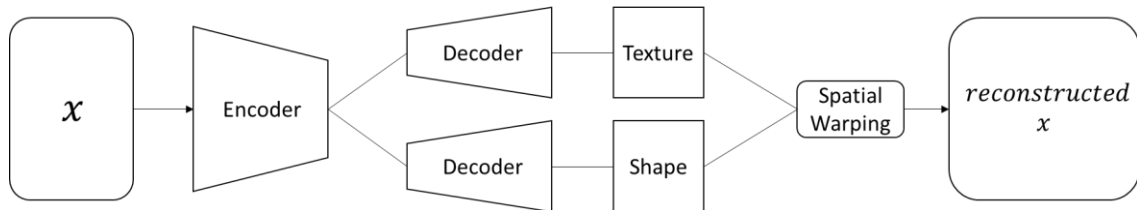


図 2.2 DAE の構造

2.4 むすび

本章では Autoencoder の概要や仕組みについて述べた。また関連研究である Deforming Autoencoders (DAE) の概要，仕組みについて述べた。

第3章 Generative Adversarial Networks

3.1 まえがき

本章では, Generative Adversarial Networks (GAN)の概要や仕組みについて述べる. また関連研究である Discovery GAN (DiscoGAN)の概要, 仕組みについて述べる.

3.2 Generative Adversarial Networks

Generative Adversarial Networks (GAN)は2014年にGoodfellowによって考案された, 深層学習を用いて生成モデルを訓練させるためのモデリング手法である [1]. GANの構造を図3.1に示す. GANの学習はGeneratorとDiscriminatorと呼ばれる2種類のネットワークを敵対的に学習する. Generatorは入力された潜在変数 z から画像 x' を生成する. Discriminatorは画像 x' と正解データ x を比較し, 本物かどうかを識別する. GeneratorはDiscriminatorが本物だと識別する画像を生成するように学習する. DiscriminatorはGeneratorが生成した画像が偽物であることを識別できるように学習する. このように二つのネットワークを競合させることで精度の高い画像を生成できるようになる.

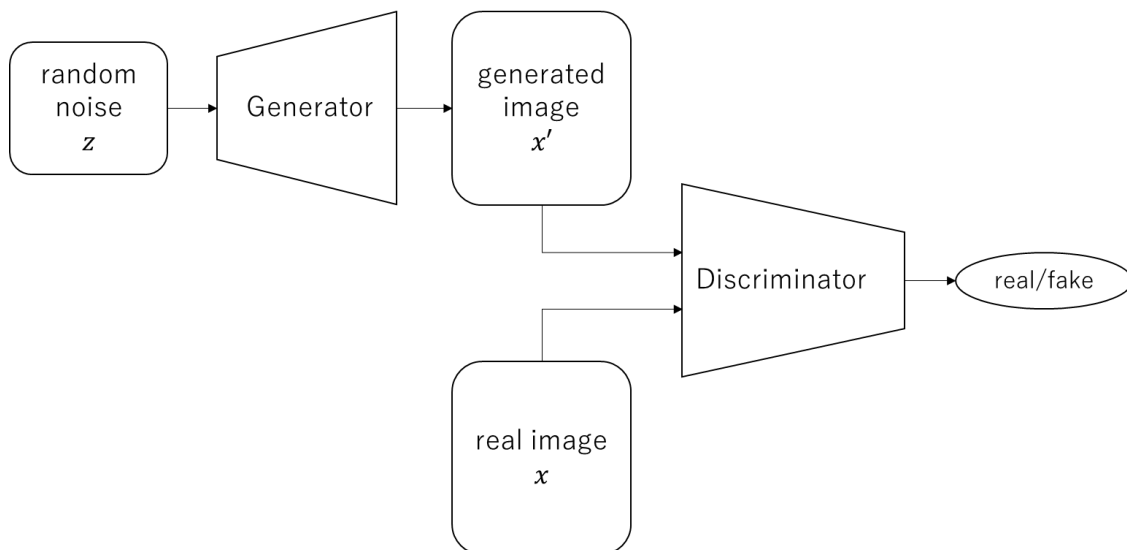


図 3.1 GANの構造

3.3 Discovery GAN

Discovery GAN (DiscoGAN)は2017年にKimらが発表した, GANを基本とした画像スタイル変換技術である [4]. DiscoGANの構造を図3.2に示す. DiscoGANはGANと同様にGeneratorとDiscriminatorを敵対的に学習する. 学習には2種類の異なるドメインの,

ペアリングされていない画像を使用する. 二つの **Generator** と二つの **Discriminator** から構成される. $Generator_{AB}$ はドメイン A の画像 x_A をドメイン B のスタイルに変換し, 画像 x_{AB} を生成する. $Generator_{BA}$ はドメイン B の画像 x_B をドメイン A のスタイルに変換し, 画像 x_{BA} を生成する. $Discriminator_A$ はドメイン A の正解データ x_A と, $Generator_{BA}$ が生成した画像 x_{BA} を比較し, 本物かどうかを識別する. $Discriminator_B$ はドメイン B の正解データ x_B と, $Generator_{AB}$ が生成した画像 x_{AB} を比較し, 本物かどうかを識別する. また画像 x_{AB} を $Generator_{BA}$ に入力して画像 x_{ABA} を生成し, 画像 x_A に再構成されるように学習する. 同様に画像 x_{BA} を $Generator_{AB}$ に入力して画像 x_{BAB} を生成し, 画像 x_B に再構成されるように学習する.

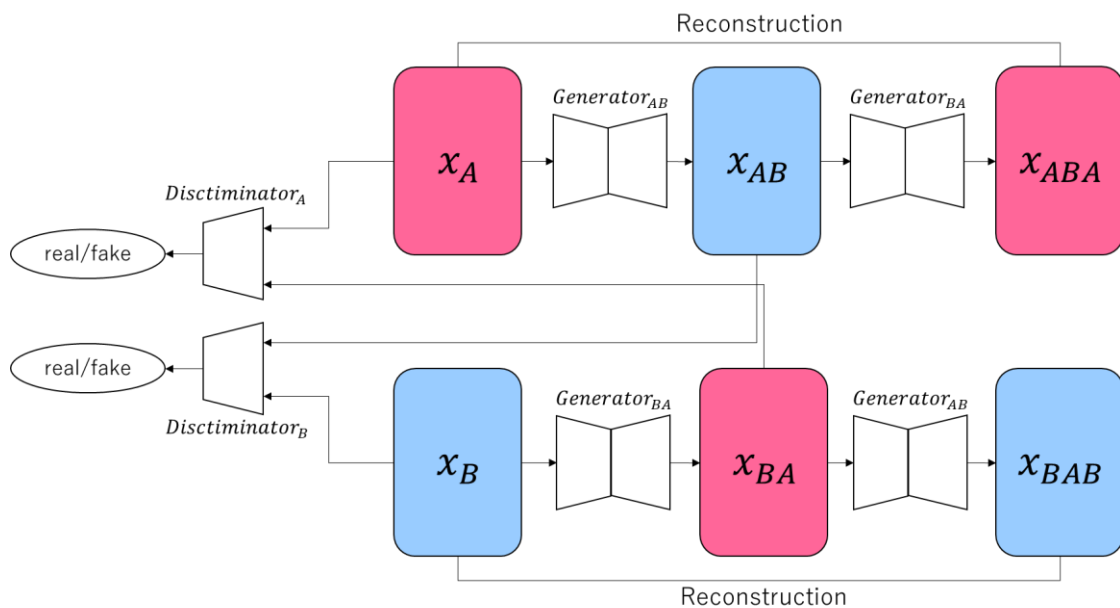


図 3.2 DiscoGAN の構造

3.4 むすび

本章では Generative Adversarial Networks (GAN) の概要や仕組みについて, また関連研究である Discovery GAN (DiscoGAN) の概要, 仕組みについて述べた.

第4章 評価指標

4.1 まえがき

本章では、本研究に用いた評価指標について述べる。

4.1 和菓子画像判別 CNN

和菓子画像判別 CNN は 2019 年に提案した、和菓子画像を学習した CNN を用いて生成画像を評価する手法である [5]。和菓子画像 11000 枚と負例画像 11000 枚で CNN を学習し、この CNN に GAN で生成した和菓子画像を入力することで和菓子画像であるかどうかを判別する。学習後の評価用画像による正解率は 98.91[%]であった。

一つの GAN につき生成画像 1000 枚を評価し、算出した正解率から使用した GAN を評価する。

4.2 Multi-scale Structural Similarity

Multi-scale Structural Similarity (MS-SSIM)は、2 画像間の類似性を測定する画像評価指標である [6]。この指標を用いて生成画像の多様性を評価する。値が小さいほど 2 画像の類似性が低く多様性があると評価できる。

4.3 Fréchet Inception Distance

Fréchet Inception Distance (FID)は、2 画像群間の距離を測定する画像評価指標である [7]。GAN の学習に用いたデータセットの分布と GAN で生成した画像の分布の Fréchet Distance を計算する。値が小さいほど距離が近く、高品質な画像であると評価できる。

4.1 むすび

本章では、本研究に用いた評価指標について述べた。

第5章 予備実験

5.1 まえがき

本章では, DiscoGAN を用いて和菓子画像を生成する予備実験について述べる.

5.2 使用するデータセット

本研究で使用する画像データセットは, 和菓子画像データセットと素材画像データセットの 2 種類である. それぞれインターネット上で和菓子の画像及び和菓子の素材となる動植物の画像を取得し作成した. 和菓子画像データセットは 27 種類の素材を表現した和菓子画像 11051 枚である. 素材画像データセットは 27 種類の素材画像 11073 枚である. 和菓子画像データセットの素材ごとの画像枚数を表 5.1 に示す. 素材画像データセットの素材ごとの画像枚数を表 5.2 に示す.

表 5.1 和菓子画像データセット

素材	画像枚数[枚]
朝顔	589
紫陽花	913
うぐいす	198
うさぎ	381
団扇	494
梅	308
カーネーション	406
桔梗	252
菊	770
金魚	744
栗	486
コスモス	284
桜	658
山茶花	270
菖蒲	456
水仙	359
つつじ	211
椿	180
手鞠	294
撫子	231
花火	187
薔薇	527
向日葵	535
藤の花	276
牡丹	216
紅葉	452
桃	373
合計	11051

表 5.2 素材画像データセット

素材	画像枚数[枚]
朝顔	540
紫陽花	771
うぐいす	92
うさぎ	350
団扇	347
梅	351
カーネーション	470
桔梗	313
菊	658
金魚	373
栗	312
コスモス	371
桜	921
山茶花	382
菖蒲	391
水仙	415
つつじ	390
椿	204
手鞠	162
撫子	344
花火	294
薔薇	624
向日葵	549
藤の花	229
牡丹	339
紅葉	566
桃	316
合計	11073

5.3 DiscoGAN の実行

和菓子画像データセットと素材画像データセットを用いて DiscoGAN を実行した。ドメイン A を和菓子画像，ドメイン B を素材画像とする。学習時の画像サイズは 64[pixel]，ミニバッチサイズは 16，エポック数は 600 であった。DiscoGAN の $Generator_{BA}$ が生成した和菓子画像 x_{BA} の例を図 5.1 に示す。DiscoGAN の $Generator_{AB}$ が生成した素材画像 x_{AB} の例を図 5.2 に示す。

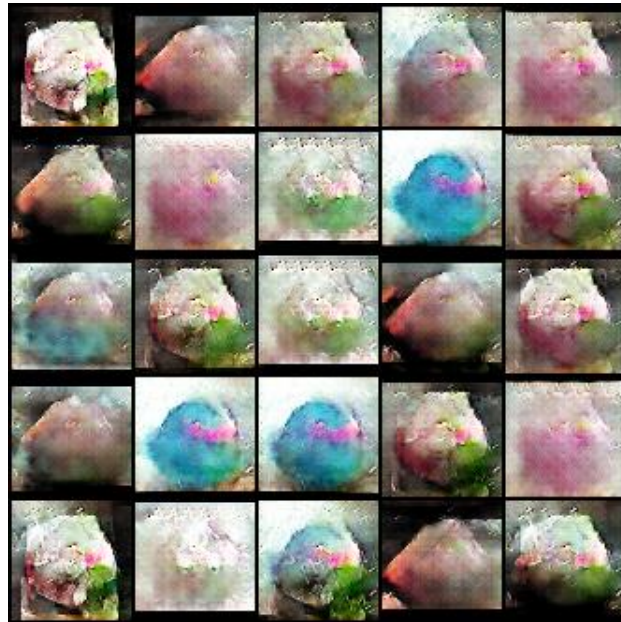


図 5.1 DiscoGAN による生成和菓子画像 x_{BA} の例



図 5.2 DiscoGAN による生成素材画像 x_{AB} の例

5.4 評価

和菓子判別 CNN, MS-SSIM, FID を用いて, DiscoGAN が生成した和菓子画像を評価した. 評価結果を表 5.3 に示す.

表 5.3 DiscoGAN の評価結果

和菓子判別 CNN[%]	MS-SSIM	FID
50.4	0.517	261.55

5.5 考察

和菓子判別 CNN の正解率からあまり和菓子らしい画像が生成できていないといえる. MS-SSIM の値から多様性のある画像が生成できなかったといえる. 目視においても図 5.1 の 25 枚の生成画像の中で, テクスチャと形状共に類似している画像が複数あることが確認できる. また FID の値から品質の高い画像が生成できなかったといえる. これは顔画像はそれぞれの被写体の形状や色, パーツの配置などが似ていることと比較し, 和菓子画像と素材画像は画像によって全く異なることが原因である. したがって被写体の形状を捉えることで改善できると考えられる.

5.6 むすび

本章では, DiscoGAN を用いて和菓子画像を生成する予備実験について述べた.

第6章 提案手法

6.1 まえがき

本章では、和菓子画像を生成する GAN を提案する。

6.2 提案手法 1

本研究では、DiscoGAN を用いた和菓子画像の生成において形状の多様性が乏しいという問題点を解決する手法として、DAE と DiscoGAN を組み合わせた形状情報を用いる画像スタイル変換手法を四つ提案する。

提案手法 1 の Generator の構造を図 6.1 に示す。二つの Generator は、それぞれ一つの Encoder と二つの Decoder から構成される。Generator_{AB} の Encoder は、ドメイン A の画像 x_A を低次元ベクトルに変換する。変換された低次元ベクトルをテクスチャと形状に分解し、Decoder はテクスチャベクトルと形状ベクトルのそれぞれを高次元ベクトルに変換する。二つの高次元ベクトルに空間変形を行い、再構成されたベクトルがドメイン B の画像 x_{AB} に変換されるように学習する。Generator_{BA} の Encoder はドメイン B の画像 x_B を低次元ベクトルに変換する。変換されたベクトルをテクスチャと形状に分解し、Decoder はそれぞれを高次元ベクトルに変換する。二つの高次元ベクトルに空間変形を行い、再構成されたベクトルがドメイン A の画像 x_{BA} に変換されるように学習する。

Discriminator_A はドメイン A の正解データ x_A と、Generator_{BA} が生成した画像 x_{BA} を比較し、本物かどうかを識別する。Discriminator_B はドメイン B の正解データ x_B と、Generator_{AB} が生成した画像 x_{AB} を比較し、本物かどうかを識別する。また画像 x_{AB} を Generator_{BA} に入力して画像 x_{ABA} を生成し、画像 x_A に再構成されるように学習する。同様に画像 x_{BA} を Generator_{AB} に入力して画像 x_{BAB} を生成し、画像 x_B に再構成されるように学習する。

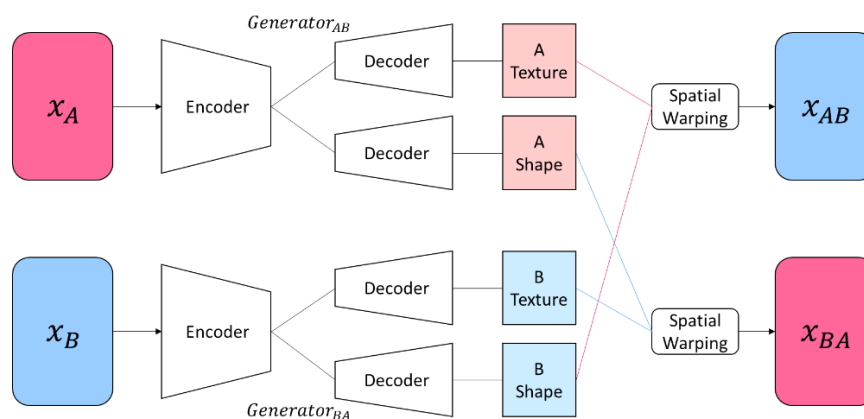


図 6.1 提案手法 1 の Generator の構造

6.3 提案手法 2

提案手法 2 の Generator の構造を図 6.2 に示す. 二つの Generator はそれぞれ一つの Encoder と二つの Decoder, 一つの DiscoGAN の Generator から構成される. $Generator_{AB}$ の Encoder は, ドメイン A の画像 x_A を低次元ベクトルに変換する. 変換されたベクトルをテクスチャと形状に分解し, Decoder はテクスチャベクトルと形状ベクトルのそれぞれを高次元ベクトルに変換する. 二つの高次元ベクトルに空間変形を行い, 再構成されたベクトルが画像 x_A に復元されるように学習する. また, テクスチャの高次元ベクトルを DiscoGAN の Generator に入力し, ドメイン B のテクスチャベクトルに変換する. これをドメイン B の形状ベクトルと空間変形し, 画像 x_{AB} を生成する. $Generator_{BA}$ の Encoder はドメイン B の画像 x_B を低次元ベクトルに変換する. 変換されたベクトルをテクスチャと形状に分解し, Decoder はそれぞれを高次元ベクトルに変換する. 二つの高次元ベクトルに空間変形を行い, 再構成されたベクトルが画像 x_B に復元されるように学習する. また, テクスチャの高次元ベクトルを DiscoGAN の Generator に入力し, ドメイン A のテクスチャベクトルに変換する. これをドメイン A の形状ベクトルと空間変形し, 画像 x_{BA} を生成する.

$Discriminator_A$ はドメイン A の正解データ x_A と, $Generator_{BA}$ が生成した画像 x_{BA} を比較し, 本物かどうかを識別する. $Discriminator_B$ はドメイン B の正解データ x_B と, $Generator_{AB}$ が生成した画像 x_{AB} を比較し, 本物かどうかを識別する. また画像 x_{AB} を $Generator_{BA}$ に入力して画像 x_{ABA} を生成し, 画像 x_A に再構成されるように学習する. 同様に画像 x_{BA} を $Generator_{AB}$ に入力して画像 x_{BAB} を生成し, 画像 x_B に再構成されるように学習する.

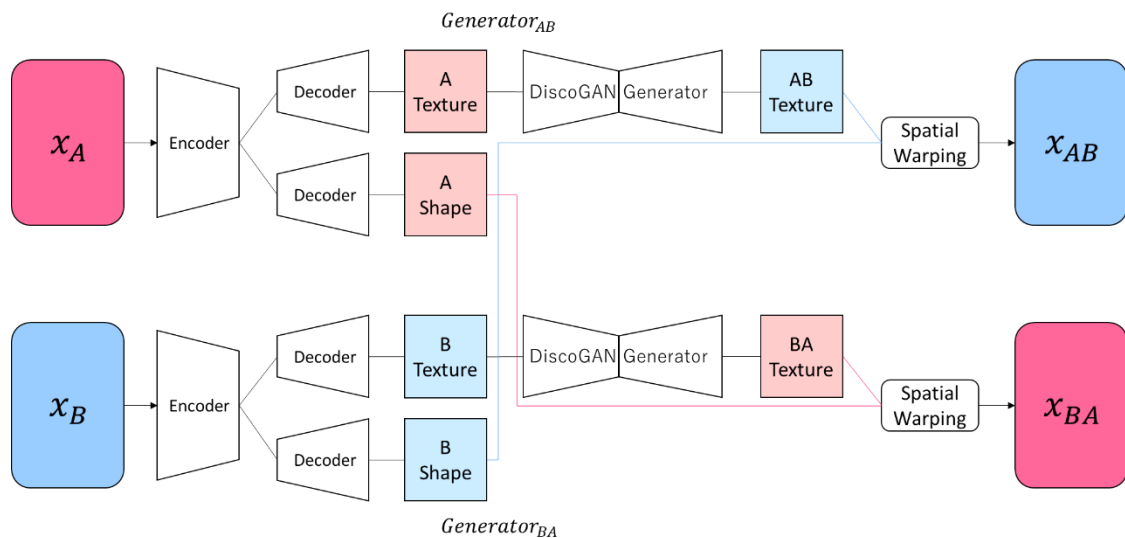


図 6.2 提案手法 2 の Generator の構造

6.4 提案手法 3

提案手法 3 の Generator の構造を図 6.3 に示す. Generator, Discriminator とともに提案手法 2 と同様の構造であるが, Encoder と Decoder を事前学習し, 学習済みモデルを用いてテクスチャベクトルと形状ベクトルを生成する. $Generator_{AB}$ の Encoder は, ドメイン A の画像 x_A を低次元ベクトルに変換する. 変換されたベクトルをテクスチャと形状に分解し, Decoder はテクスチャベクトルと形状ベクトルのそれぞれを高次元ベクトルに変換する. 二つの高次元ベクトルに空間変形を行い, 再構成されたベクトルが画像 x_A に復元されるように学習する. また, テクスチャの高次元ベクトルを DiscoGAN の Generator に入力し, ドメイン B のテクスチャベクトルに変換する. これをドメイン B の形状ベクトルと空間変形し, 画像 x_{AB} を生成する. $Generator_{BA}$ の Encoder はドメイン B の画像 x_B を低次元ベクトルに変換する. 変換されたベクトルをテクスチャと形状に分解し, Decoder はそれぞれを高次元ベクトルに変換する. 二つの高次元ベクトルに空間変形を行い, 再構成されたベクトルが画像 x_B に復元されるように学習する. また, テクスチャの高次元ベクトルを DiscoGAN の Generator に入力し, ドメイン A のテクスチャベクトルに変換する. これをドメイン A の形状ベクトルと空間変形し, 画像 x_{BA} を生成する.

$Discriminator_A$ はドメイン A の正解データ x_A と, $Generator_{BA}$ が生成した画像 x_{BA} を比較し, 本物かどうかを識別する. $Discriminator_B$ はドメイン B の正解データ x_B と, $Generator_{AB}$ が生成した画像 x_{AB} を比較し, 本物かどうかを識別する. また画像 x_{AB} を $Generator_{BA}$ に入力して画像 x_{ABA} を生成し, 画像 x_A に再構成されるように学習する. 同様に画像 x_{BA} を $Generator_{AB}$ に入力して画像 x_{BAB} を生成し, 画像 x_B に再構成されるように学習する.

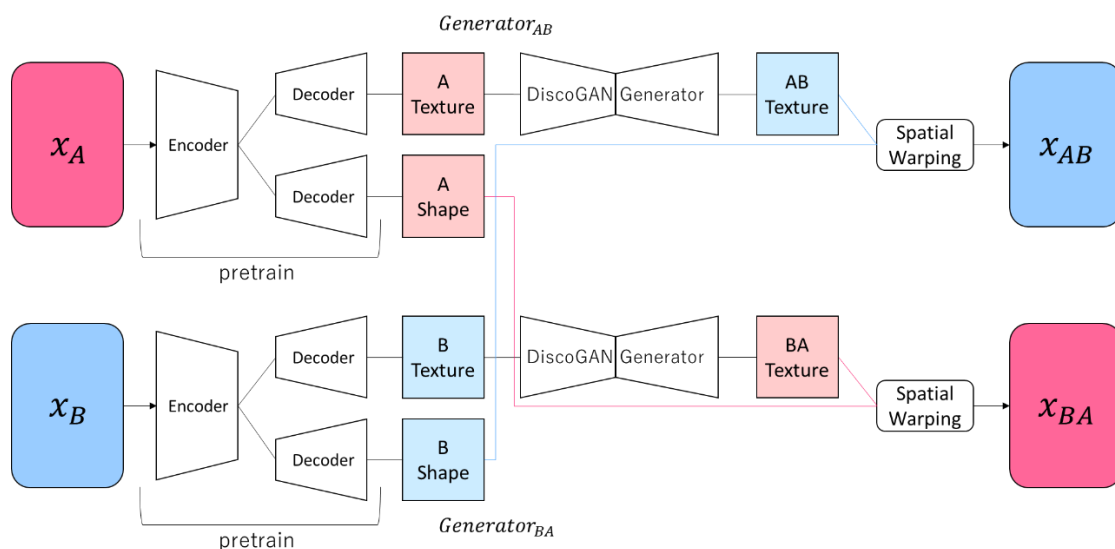


図 6.3 提案手法 3 の Generator の構造

6.5 提案手法 4

提案手法 2 の Generator の構造を図 6.4 に示す. 二つの Generator はそれぞれ一つの Encoder と二つの Decoder, 一つの DiscoGAN の Generator から構成される. Encoder と Decoder を事前学習し, 学習済みモデルを用いてテクスチャベクトルと形状ベクトルを生成する. $Generator_{AB}$ の Encoder は, ドメイン A の画像 x_A を低次元ベクトルに変換する. 変換されたベクトルをテクスチャと形状に分解し, Decoder はテクスチャベクトルと形状ベクトルのそれぞれを高次元ベクトルに変換する. 二つの高次元ベクトルに空間変形を行い, 再構成されたベクトルが画像 x_A に復元されるように学習する. また, 画像 x_A を DiscoGAN の Generator に入力し, ドメイン B の画像 x_{AB}' に変換する. これをドメイン B の形状ベクトルと空間変形し, 画像 x_{AB} を生成する. $Generator_{BA}$ の Encoder はドメイン B の画像 x_B を低次元ベクトルに変換する. 変換されたベクトルをテクスチャと形状に分解し, Decoder はそれぞれを高次元ベクトルに変換する. 二つの高次元ベクトルに空間変形を行い, 再構成されたベクトルが画像 x_B に復元されるように学習する. また, 画像 x_B を DiscoGAN の Generator に入力し, ドメイン A の画像 x_{BA}' に変換する. これをドメイン A の形状ベクトルと空間変形し, 画像 x_{BA} を生成する.

$Discriminator_A$ はドメイン A の正解データ x_A と, $Generator_{BA}$ が生成した画像 x_{BA} を比較し, 本物かどうかを識別する. $Discriminator_B$ はドメイン B の正解データ x_B と, $Generator_{AB}$ が生成した画像 x_{AB} を比較し, 本物かどうかを識別する. また画像 x_{AB} を $Generator_{BA}$ に入力して画像 x_{ABA} を生成し, 画像 x_A に再構成されるように学習する. 同様に画像 x_{BA} を $Generator_{AB}$ に入力して画像 x_{BAB} を生成し, 画像 x_B に再構成されるように学習する.

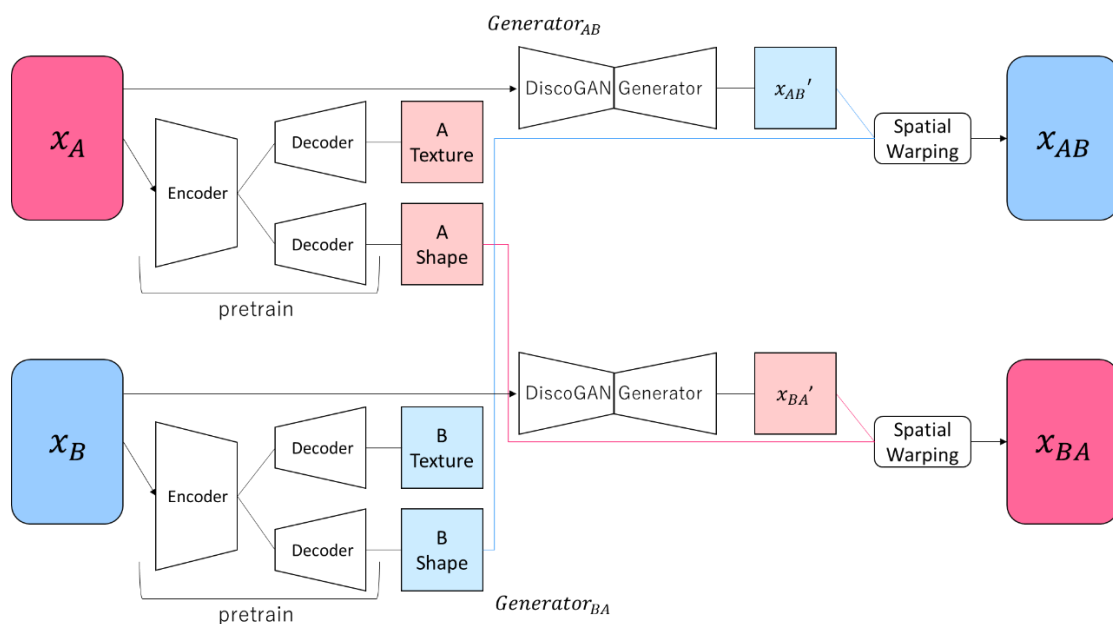


図 6.4 提案手法 4 の Generator の構造

6.6 むすび

本章では和菓子画像を生成する GAN を提案した.

第7章 実験結果

7.1 まえがき

本章では提案手法を用いた和菓子画像の生成結果，その評価について述べる．また実験結果について考察を述べる

7.2 提案手法 1 の実行

5.2 に示した和菓子画像データセットと素材画像データセットを用いて，提案手法 1 を実行した．ドメイン A を和菓子画像，ドメイン B を素材画像とする．学習時の画像サイズは 64[pixel]，ミニバッチサイズは 16，エポック数は 600 であった．提案手法 1 の $Generator_{BA}$ が生成した和菓子画像 x_{BA} の例を図 7.1 に示す．提案手法 1 の $Generator_{AB}$ が生成した素材画像 x_{AB} の例を図 7.2 に示す．



図 7.1 提案手法 1 による生成和菓子画像 x_{BA} の例



図 7.2 提案手法 1 による生成素材画像 x_{AB} の例

7.3 提案手法 2 の実行

5.2 に示した和菓子画像データセットと素材画像データセットを用いて、提案手法 2 を実行した。ドメイン A を和菓子画像、ドメイン B を素材画像とする。学習時の画像サイズは 64[pixel]、ミニバッチサイズは 16、エポック数は 600 であった。提案手法 2 の $Generator_{BA}$ が生成した和菓子画像 x_{BA} の例を図 7.3 に示す。提案手法 2 の $Generator_{AB}$ が生成した素材画像 x_{AB} の例を図 7.4 に示す。

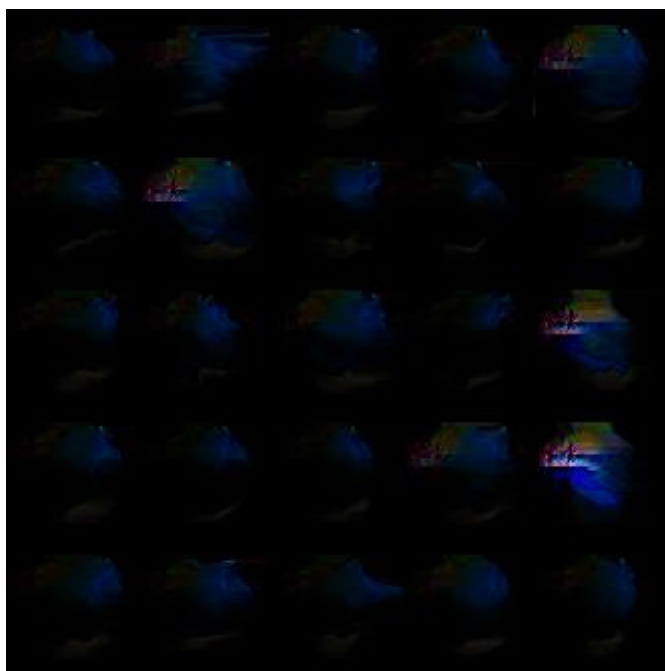


図 7.3 提案手法 2 による生成和菓子画像 x_{BA} の例

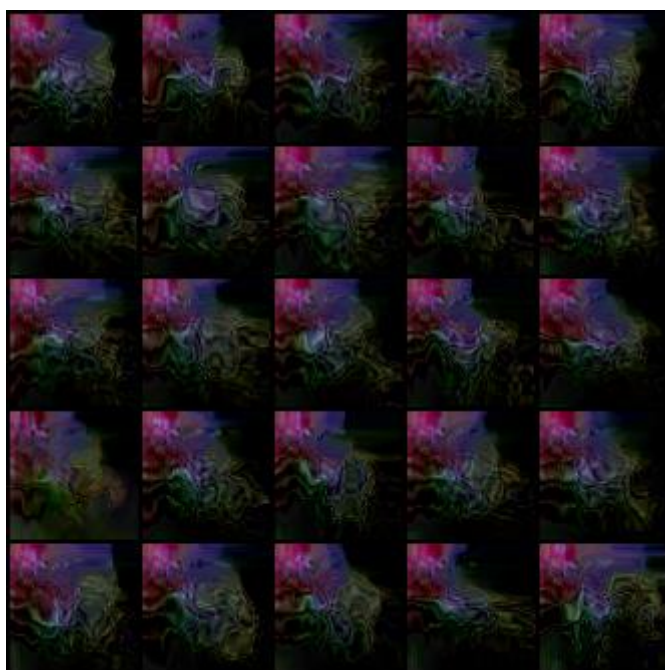


図 7.4 提案手法 2 による生成素材画像 x_{AB} の例

7.4 提案手法 3 の実行

5.2 に示した和菓子画像データセットと素材画像データセットを用いて，提案手法 3 を実行する．ドメイン A を和菓子画像，ドメイン B を素材画像とする．学習時の画像サイズは

64[pixel], ミニバッチサイズは 16, エポック数は 600 であった. 提案手法 3 の $Generator_{BA}$ が生成した和菓子画像 x_{BA} の例を図 7.5 に示す. 提案手法 3 の $Generator_{AB}$ が生成した素材画像 x_{AB} の例を図 7.6 に示す.

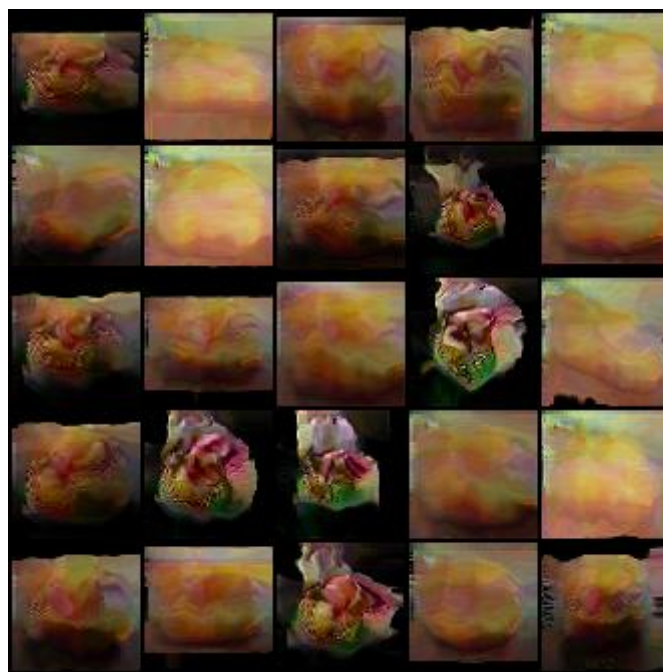


図 7.5 提案手法 3 による生成和菓子画像 x_{BA} の例

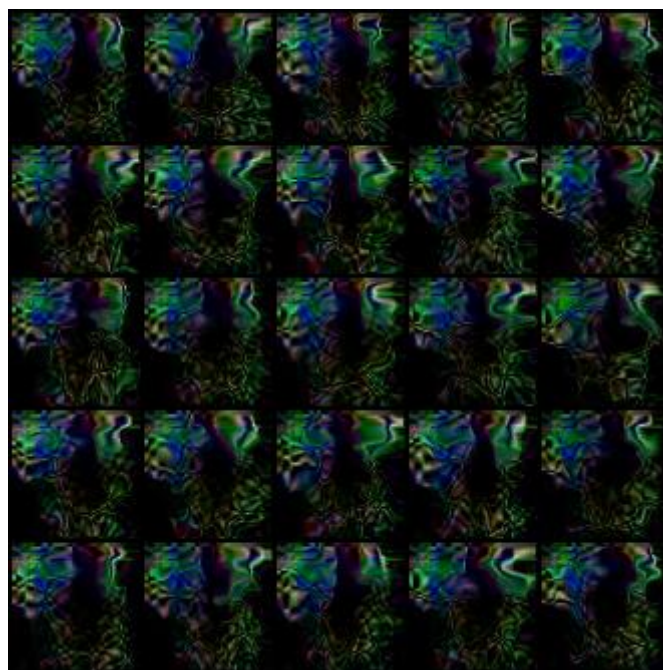


図 7.6 提案手法 3 による生成素材画像 x_{AB} の例

7.5 提案手法 4 の実行

5.2 に示した和菓子画像データセットと素材画像データセットを用いて、提案手法 4 を実行する。ドメイン A を和菓子画像、ドメイン B を素材画像とする。学習時の画像サイズは 64[pixel]、ミニバッチサイズは 16、エポック数は 600 であった。提案手法 4 の $Generator_{BA}$ が生成した和菓子画像 x_{BA} の例を図 7.7 に示す。提案手法 4 の $Generator_{AB}$ が生成した素材画像 x_{AB} の例を図 7.8 に示す。

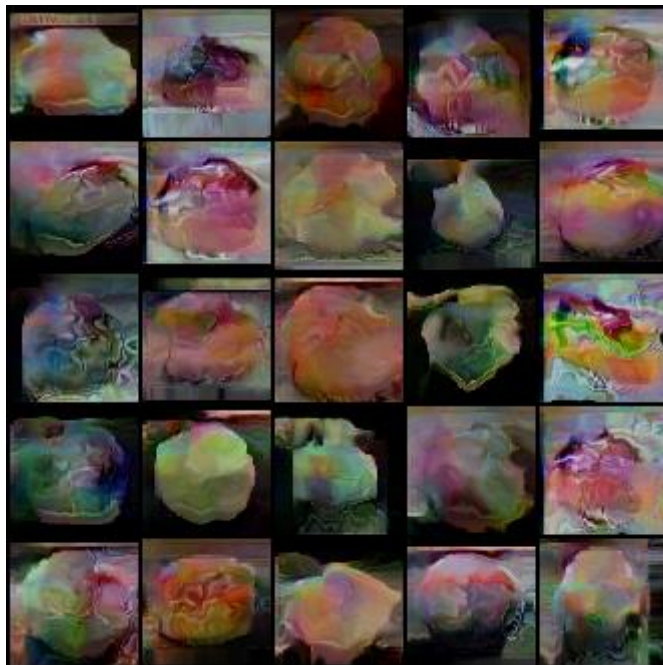


図 7.7 提案手法 4 による生成和菓子画像 x_{BA} の例

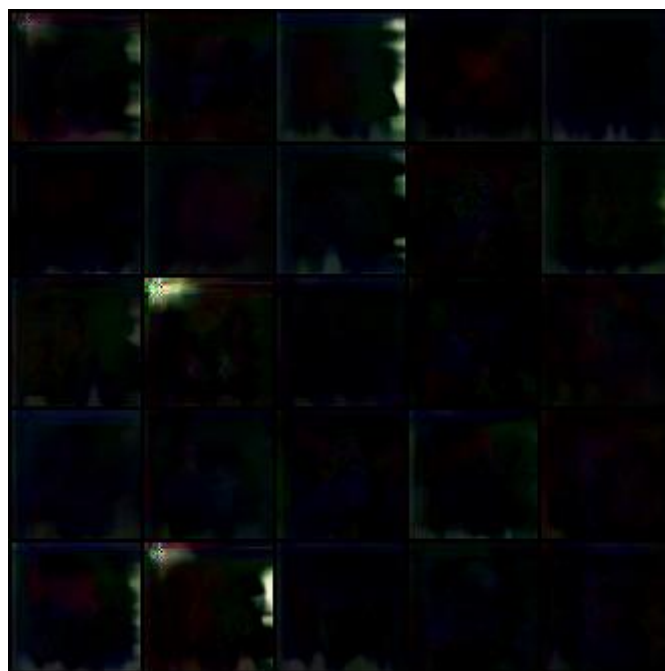


図 7.8 提案手法 4 による生成素材画像 x_{AB} の例

7.6 評価

和菓子判別 CNN, MS-SSIM, FID を用いて, 提案した Generator が生成した和菓子画像を評価した. 評価結果を表 7.1 に示す.

表 7.1 生成画像の評価結果

	和菓子判別 CNN[%]	MS-SSIM	FID
DiscoGAN	50.4	0.517	261.55
提案手法 1	0.00	1.000	421.94
提案手法 2	1.80	0.904	354.27
提案手法 3	79.2	0.501	191.29
提案手法 4	93.7	0.222	168.82

7.7 考察

表 7.1 から各手法の和菓子判別 CNN の正解率を比較すると, 93.7[%]の提案手法 4 が DiscoGAN 及び提案手法の中で最も和菓子らしい画像を生成できたといえる. また MS-SSIM の値を比較すると, 0.222 の提案手法 4 が最も多様性のある画像を生成できたといえる. FID の値を比較すると, 168.82 の提案手法 4 が最も高品質な画像を生成できたといえる. したがって提案手法 4 が最も和菓子画像の生成に適していることが分かった.

図 7.1, 表 7.1 から提案手法 1 は学習が失敗したといえる. 提案手法 1 の $Generator_{BA}$ は

ドメイン B のテクスチャとドメイン B の形状で画像 x_B を再構成することと、ドメイン B のテクスチャとドメイン A の形状で画像 x_{BA} に変換することを同時に学習している。しかし両方をみだすことは不可能であるため失敗したと考えられる。

図 7.3, 表 7.1 から提案手法 2 は学習が失敗したといえる。提案手法 2 の $Generator_{BA}$ はテクスチャと形状を分解する DAE の Encoder と Decoder の学習と、分解したドメイン B のテクスチャをドメイン A のテクスチャに変換する DiscoGAN の Generator の学習を同時に行っている。しかしこの手法では、学習開始時は DAE がテクスチャと形状を分解することができない。したがって学習するにつれて DiscoGAN の Generator への入力に変化し、変換の内容も変化するため失敗したと考えられる。

図 5.1, 図 7.5, 表 7.1 から提案手法 3 は DiscoGAN と比較して精度の高い和菓子画像を生成できたが、提案手法 4 と比較すると精度が低い。これは素材画像を用いた DAE の Encoder と Decoder の学習が十分でなく、ドメイン B のテクスチャが上手く学習できなかったからであると考えられる。

図 5.1, 表 7.1 から提案手法 4 は DiscoGAN と比較して精度の高い和菓子画像を生成できた。これは和菓子画像の形状情報を用いることでより多様な画像を生成できるからである。

一方で図 7.2, 図 7.4, 図 7.6, 図 7.8 から全ての提案手法において和菓子画像を素材画像に変換する $Generator_{AB}$ の学習が失敗したといえる。これは素材画像データセットの画像のテクスチャや形状が多様すぎるため、テクスチャと形状を上手く分解できなかったと考えられる。

7.8 むすび

本章では提案手法を用いた和菓子画像の生成結果。その評価について述べた。また実験結果について考察を述べた。実験結果より、提案手法 4 が最も和菓子画像の生成に適していると分かった。

第8章 結論と今後の課題

8.1 結論

本研究では、被写体のテクスチャと形状を分解する DAE と、異なるドメインの画像にスタイル変換を行う DiscoGAN を組み合わせた、形状情報を用いる画像スタイル変換の手法を提案した。また DiscoGAN 及び四つの提案手法を用いて和菓子画像を生成し、三つの指標で評価した。評価の結果、全ての指標において提案手法 4 が最良であり、最も和菓子画像の生成に適していることが分かった。

8.2 今後の課題

今後の課題として素材画像を用いた DAE の学習と、和菓子画像を素材画像に変換する $Generator_{AB}$ の学習の精度を向上させることの二つが挙げられる。

DAE の学習は和菓子画像の学習に最適なパラメータ調整を行い、素材画像の学習にも同様の設定を使用した。素材画像に最適なパラメータ調整を行うことで、学習の精度が向上すると考えられる。また学習に使用したデータセットの和菓子画像は 1 枚当たり 1 被写体であるが、素材画像には景色全体の画像や複数被写体の集合画像が含まれる。したがって素材画像は和菓子画像と比較してテクスチャや形状に多様性がある。データクリーニングや物体認識を用いて、素材画像も和菓子画像同様 1 枚当たり 1 被写体の画像を使用することで学習の精度が向上すると考えられる。 $Generator_{AB}$ の学習の精度が向上すれば $Generator_{BA}$ とのバランスが保たれ、和菓子画像生成の精度も向上すると考えられる。

謝辞

本研究の実験環境を整えてくださり、研究の方向性を示して下さったり問題点を提起して下さったりと適切な指導を賜った渡辺裕教授に心より感謝申し上げます。

また、日頃からご意見やアドバイスをくださった研究室の皆様に御礼申し上げます。

最後に、私をここまで育ててくれ、支えてくれた家族に感謝します。

参考文献

- [1] I. J. Goodfellow, J. P. Abadie, M. Mirza, B. Xu, D. W. Farley, S. Ozair, A. Courville and Y. Bengio, "Generative Adversarial Nets," NIPS, pp.1-9, 2014.
- [2] G. E. Hinton , R. R. Salakhutdinov, "Reducing the Dimensionality of Data with Neural Networks," Science, 313, 504-507, 2006.
- [3] Z. Shu, M. Sahasrabudhe, A. Guler, D. Samaras, N. Paragios , I. Kokkinos, "Deforming Autoencoders: Unsupervised Disentangling of Shape and Appearance," arXiv preprint arXiv:1806.06503v1, 2018.
- [4] T. Kim, M. Cha, H. Kim, J. K. Lee and J. Kim, "Learning to Discover Cross-Domain Relations with Generative Adversarial Networks," arXiv preprint arXiv:1703.05192v2, 2017.
- [5] 大山優香 , 渡辺裕, "GAN による和菓子画像の生成と評価," 電子情報通信学会総合大会, D-12-40, 2019.
- [6] Z. Wang, E. P. Simoncelli and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," Proc. IEEE Asilomar Conf. on Signals, Systems, and Computers, pp. 1398–1402, 2003.
- [7] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler and S. Hochreiter, "GANs trained by a two time-scale update rule converge to a local Nash equilibrium," NIPS, pp.6626-6637, 2017.

目 一 覧

図 2.1	Autoencoder の構造	3
図 2.2	DAE の構造	4
図 3.1	GAN の構造	5
図 3.2	DiscoGAN の構造	6
図 5.1	DiscoGAN による生成和菓子画像 xBA の例	11
図 5.2	DiscoGAN による生成素材画像 xAB の例	11
図 6.1	提案手法 1 の Generator の構造	13
図 6.2	提案手法 2 の Generator の構造	14
図 6.3	提案手法 3 の Generator の構造	15
図 6.4	提案手法 4 の Generator の構造	16
図 7.1	提案手法 1 による生成和菓子画像 xBA の例	18
図 7.2	提案手法 1 による生成素材画像 xAB の例	19
図 7.3	提案手法 2 による生成和菓子画像 xBA の例	20
図 7.4	提案手法 2 による生成素材画像 xAB の例	20
図 7.5	提案手法 3 による生成和菓子画像 xBA の例	21
図 7.6	提案手法 3 による生成素材画像 xAB の例	21
図 7.7	提案手法 4 による生成和菓子画像 xBA の例	22
図 7.8	提案手法 4 による生成素材画像 xAB の例	23

表一覧

表 5.1	和菓子画像データセット	9
表 5.2	素材画像データセット	10
表 5.3	DiscoGAN の評価結果	12
表 7.1	生成画像の評価結果	23

研究業績

1. 大山, 渡辺, “GAN による和菓子画像の生成と評価”, 電子情報通信学会総合大会, D-12-40, Mar. 2019.
2. 大山, 渡辺, “敵対的生成ネットワークを用いた和菓子画像生成の検討”, 2019 年画像符号化シンポジウム・2019 年映像メディア処理シンポジウム(PCSJ/IMPS2019), P-2-03, Nov. 2019.