

Textile-GAN

ーStyleGAN による印象制御に基づくテクスチャ生成ー

谷口 史果[†] 飛谷 謙介^{†,‡} 長田 典子[†]

[†] 関西学院大学大学院理工学研究科 〒669-1330 兵庫県三田市学園上ヶ原 1

[‡] 長崎県立大学情報システム学部 〒851-2130 長崎県西彼杵郡長与町まなび野 1-1-1

E-mail: [†] {ixs50880, nagata}@kwansei.ac.jp, [‡] tobitani@sun.ac.jp

あらまし 本研究では、衣服の柄（テキスタイルデザイン）に関する感性的質感認知に着目し、StyleGAN とその潜在空間探索によるスタイルの制御を行うことにより、所望の感性的質感を持つテクスチャ（柄）を生成する手法を提案する。対象を柄画像に絞った印象評価実験によって、柄に関する定量的な指標を得て、これを学習データとすることで高精度な印象制御能力を持つテクスチャ生成モデルを構築する。モデルの潜在空間から類似画像検索や画像の印象制御が可能となる。実験結果より人の直感に合致した印象を持つ生成画像が得られたことから、本手法の有効性が示唆された。

キーワード 感性工学, 生成 AI, StyleGAN, スタイル変換, 潜在空間, ファッション

1. はじめに

“華やかな”や“涼しげな”といった素材の表面性状（テクスチャ）から喚起される印象や感情は感性的質感認知と呼ばれ、個人の嗜好や情動・価値判断に大きくかかわるとされている。プロダクトデザインの分野では個人のニーズや嗜好にあわせたビスポークデザイン（カスタマイゼーション）が注目されており、感性的質感を理解し制御する技術の必要性が高まっている。

特にファッション分野において、衣服の柄（パターンや配色などの模様）は衣服の選択に関わる主要な要素であり、柄としてのテクスチャおよびその感性的質感認知の対応関係の把握は重要である。これまでに柄に関するテクスチャの研究では、画像特徴量の研究やシームレス化の研究、また最近では柄の好悪、審美性（aesthetics）等の研究が盛んに行われている。しかし柄が持つ多様な印象や印象に起因する感情を体系的に説明し指標化する研究は、まだ十分に検討されていない。

一方、最近の生成 AI (Generative AI) では、Stable Diffusion[1]モデルに代表されるようにテキストプロンプトから所望の画像を生成することが可能となっている[1-3]。しかしながら、柄の印象や感情といった心理的な指標に基づいた直感的かつ効率的なテクスチャの生成技術は確立されていない。

本研究では、衣服の柄から喚起される感性的質感[4]の定量的な評価を行い、これに基づき所望の感性的質感を持つ柄（テクスチャ）を生成する手法として、StyleGAN の潜在空間の探索を利用した方法を提案する。

2. 関連研究

画像生成モデルの代表的なものの1つに、敵対的生成ネットワーク（GAN）が挙げられる。とりわけ StyleGAN [5] は、画像のスタイル情報を含む潜在変数を利用して、より直感的に生成画像のスタイルを制御することが可能である。一方、衣服の柄に関する感性的質感に関する著者らの研究では、柄に関する人の評価構造・評価基準を明らかにし、画風変換アルゴリズム関連 [6] のスタイル特徴を利用して、人の直感に上手く合致した柄検索機能を実現できることを示した [7]。しかし同じ枠組みによるスタイル特徴を用いた柄生成処理では、まだ十分な効果が得られていない。

そこで本研究では、StyleGAN を用いた柄生成を試みる。

3. 提案手法

本研究では、StyleGAN とその潜在空間探索によるスタイルの制御を行うことにより、感性的質感認知に基づく柄生成を行う。柄画像に対象を絞った印象評価実験によって、柄画像に関する定量的な指標を得て、これを学習データとすることで高精度な印象制御能力を持つテクスチャ生成モデルの構築手法を提案する。StyleGAN の中でも少ない教師データ数での安定的な学習が可能である StyleGAN2-ada[8]を使用する。

具体的には、StyleGAN2-ada のマルチラベル学習を用いて所望の感性的質感を持つテクスチャ生成モデルを構築する。図 1 に提案手法の概念図を示す。はじめに、柄画像の感性的質感を定量化するための印象評価実験を行い、この実験で得られた結果を教師ラベルとして紐づけた柄画像をデータセットとして

用いる．次に，これらのデータセットを用いて StyleGAN2-ada のマルチラベル学習を行う．以上の基本設計により，印象に対してもつれの解けた (disentangled) 潜在空間を獲得し，入力画像と似た印象を持つテキストの生成や，生成画像の印象の操作が可能で，所望の感性的質感を持ったテキストの生成モデルを構築する．

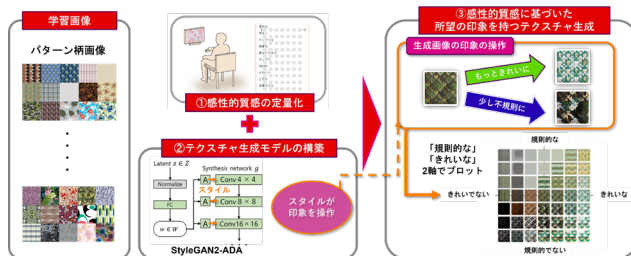


図 1 提案手法の概念図

3.1. 感性的質感の定量化

一般的な生成 AI モデルのような大量の web データで作成されたデータセットでは，明確な基準がないため，定量的な指標に基づいているわけではない．本研究では柄から喚起される感性的質感に関する評価基準を，始めに構造化し，次に 7 件法に基づく主観評価実験を行い，その結果をデータセットとする．手順は以下の通りである．

衣服の柄画像 4036 枚を収集し，柄から喚起される感性的質感を表現する形容詞を自由記述方式で収集し，段階的に選択を進め，最終的に“きれいな”，“規則的な”，“派手な”，“集合体の”，のような計 28 語を選定した．その後，各画像に対し，選定した 28 語についてどの程度当てはまるかを -3 から 3 の 7 段階で評価する心理実験を行った[6]．この実験により得られた各柄画像に対する評価の平均値を，各画像の評価点とした．その後，図 2 のように因子分析によって「幾何学的」「スタイリッシュ」「複雑さ」といった 6 因子を抽出した．累積寄与率は 69.15 %であった．

本研究では，この実験で用いた柄画像 4036 枚のデータセットと，各画像に付与された評価語ごとの評価点を学習に用い，感性的質感認知に基づく画像を生成する．

3.2. テクスチャ生成モデルの構築

感性的質感の定量化で得られたデータセットを用いて StyleGAN2-ada でマルチラベル学習することでモデルを構築する．StyleGAN や StyleGAN2-ada のジェネレータは，学習された定数入力から開始し，潜在コードに基づいて各畳み込み層で画像の「スタイル」を調整する．スタイルは，ジェネレータの各畳み込み層で生成される画像の外観を制御する学習済みベクトルのことを指す．スタイルベクトルは学習プ

ロセス中に学習され，正規分布からランダムにサンプリングされた高次元ベクトルである潜在コードに基づく．各畳み込み層でスタイルベクトルを調整することにより，ジェネレータは異なるスケールで画像特徴の強さを制御することができ，生成される画像を制御することができる．これにより高レベルの属性 (柄のモチーフなどの大まかな特徴) と，生成された画像における確率的変動 (柄の細かい特徴) を教師なしで分離が可能である．

データセットは柄画像と，前項の各柄画像に対応した評価語ごとの評価点である．学習画像の解像度は 256*256pixel と 1024*1024pixel を用いた．教師データである画像は 4036 枚と多くないが，それを解決するために StyleGAN2-ada を用いてテキスト生成モデルを構築する．StyleGAN2-ada は適応的識別器増強メカニズム (adaptive discriminator augmentation, ADA) により安定的な学習が可能である．学習画像ごとに付与された各評価語 (印象) の評価点をラベルとして，柄の感性的質感をモデルにマルチラベル学習する．これにより，印象に対してもつれの解けた (disentangled) 潜在空間を獲得でき，評価語ごとの評価点を指定することで，定量的な指標に基づいたユーザ所望の印象を持つテキストの生成が可能になる．

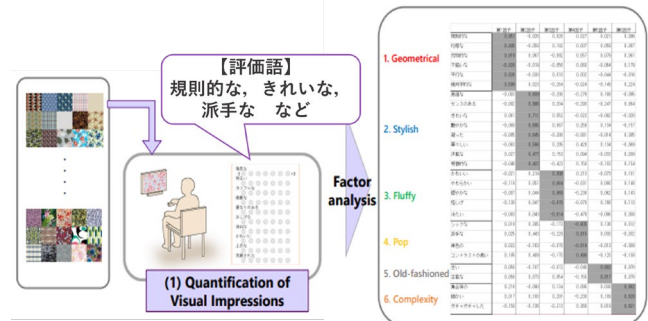


図 2 印象評価実験と因子分析

4. 結果と考察

4.1. マルチラベル学習

はじめに評価語 28 語すべてを同時にマルチラベル学習した．図 3 のように各印象は定性的に表現できていると考えられるが，指定するラベル，つまり評価語ごとの印象値が同一の場合，柄のモチーフがほぼ同様な画像が大量に生成された．この結果から 4036 枚の学習画像に対して 28 個のマルチラベル学

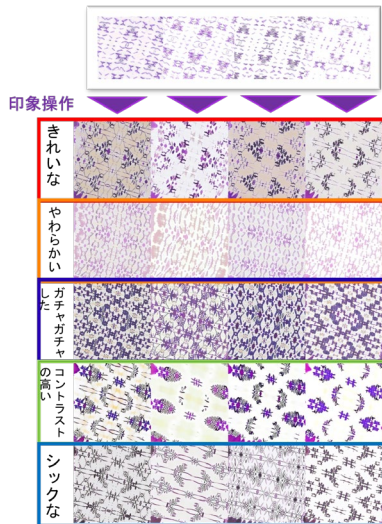


図3 評価語 28 語条件下の生成画像

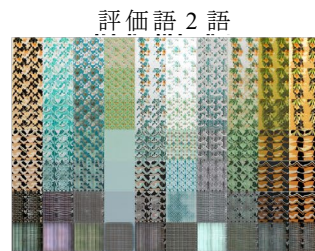
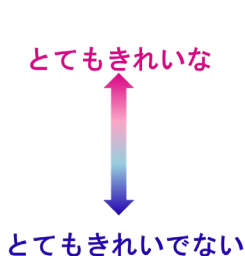


図4 評価語 2~4 語条件下の生成画像

習は、ラベル数が多すぎると考えられた。

そこで 28 個のマルチラベル学習の結果から、ラベル数を減らして実験した。図 4 はラベル 2~4 個を学習させた生成画像 10 枚に対して、評価語 4 個のうちの一つである「きれいな」の印象を変化させた結果である。学習した印象「きれいな」を定性的に表現できていると考えられる。

図 5 はラベル 5~6 個を学習させた生成画像 10 枚であるが、同じ評価語の評価点を指定すると生成画像の多様性が失われた。この結果から、現状では同時に学習するラベルが 4 個まではマルチラベル学習が可能といえる。

4.2. 解像度ごとの FID 推移

生成画像の評価を FID (Fréchet Inception Distance) 指標を用いて評価する。FID は生成画像の品質を評価するための一般的な指標であり [6], 事前に訓練された Inception-v3 ネットワークによって学習された特徴空間において、実画像の分布と生成画像の分布の間の距離を測定する。特徴空間における実画像と生成画像の分布を比較することで、ジェネレータがこれらの高レベルの特徴をどの程度正確に捉えることができるかを判断する定量的な尺度となる。FID スコアが低ければ低いほど、生成された画像は高レベルの特徴に関して実画像により類似していることを示す。これはジェネレータがより高品質な画像を

評価語 5 語



評価語 6 語

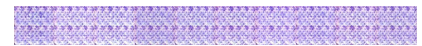


図5 評価語 5~6 語条件下の生成画像

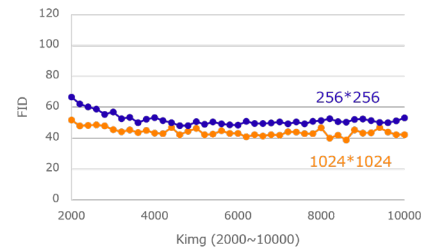
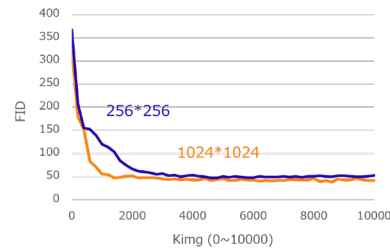


図6 学習画像の解像度ごとの FID 比較

生成していることを意味する。

解像度 256*256pixel と 1024*1024pixel の FID の推移は図 6 の通りである。この二つのパターンで比較すると、学習が安定してからの FID は約 10 の差があり、1024*1024pixel の方が 256*256pixel より FID が良かった。このことから学習柄画像の解像度が高いと生成画像の質も良くなることが示された。

4.3. 生成画像の印象制御

ノイズからランダムに生成した画像 (解像度 1024*1024pixel) の印象を編集した。4 つの評価語それぞれの印象を最大に高めた画像を生成した (図 7)。また 4 つの評価語の中で評価語「きれいな」と「規則的な」に基づいて印象を段階的に変化させ、その結果を 2 軸にプロットした (図 8)。さらに入力画像に対して、評価語の印象が最も似ている画像を学習済みモデルの生成器内の潜在空間で探索し出力した (図 9)。

結果より、柄画像に対しての印象を表す 4 つの評価語の印象値を指定することで、定量的に所望の印象を持つ画像が生成されたことが確認された。ただし図 10(a)に示すように、生成画像の色味が偏っていることからモード崩壊が起きていると考えられる。現在の取り組みは、モード崩壊への対処のため、学習データセットがよりバリエーションが豊かになるよ

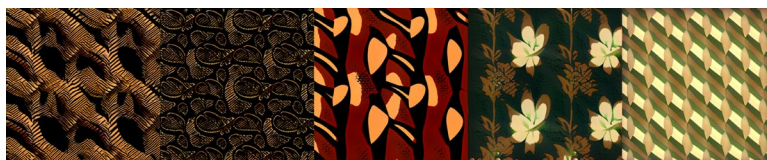


図 7 各印象を高めた生成画像編集

(左から平均的な印象の生成画像，集合体の・派手な・きれいな・集合体の印象を高めた画像)

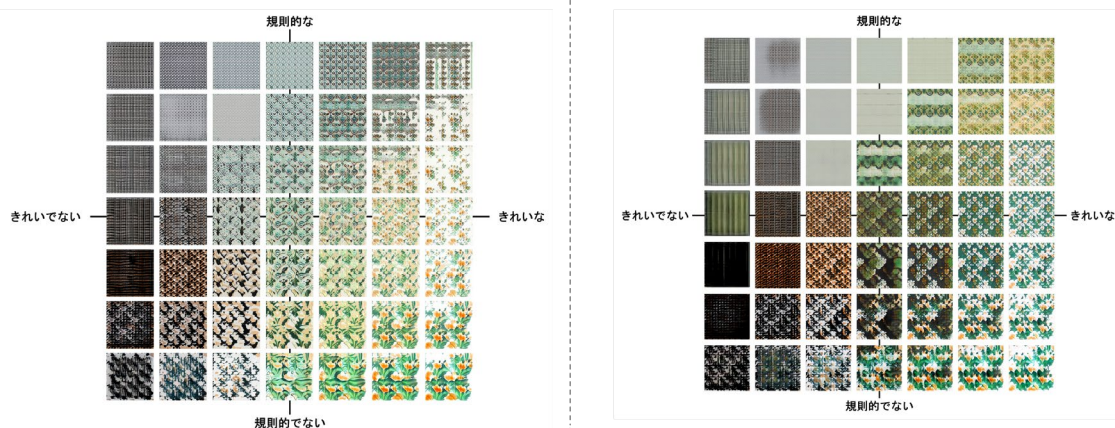


図 8 2つの生成画像の印象制御結果

うに拡張方法を追加し，異なる正則化や損失関数を試している．現在のモデルの改善進捗は図 10(b)に示す．

5. まとめ

StyleGAN の潜在空間探索を用いて，定量的な指標に基づいた所望の感性的質感をもつ画像を生成する手法を提案し，実験結果より手法の有効性を示した．

今後は，テクスチャ生成モデルが同時に学習できる最大ラベル数を増やすとともに，モデルが全部のラベルを正しく分類できているかを検証する．そのためにラベルである各評価語の評価点を変化させた生成画像の潜在空間内の分布解析を検討予定である．

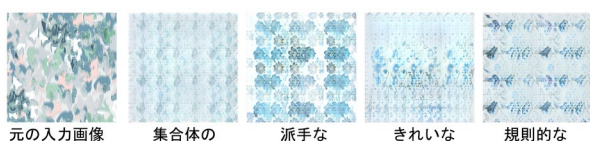


図 9 学習済み生成器内の類似画像探索結果



図 10 改善前後のランダムな印象の生成画像抜粋

文 献

- [1] R. Rombach, A. Blattmann, D. Lorenz, P. Esser and B. Ommer: “High-Resolution Image Synthesis with Latent Diffusion Models”, in 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 10674-10685 (Jun.2022)
- [2] R. Aditya, D. Prafulla, N. Alex, C. Casey, C. Mark: “Hierarchical Text-Conditional Image Generation with CLIP Latents”, arXiv, 2204.06125 (Apr.2022)
- [3] C. Saharia, W. Chan, S. Saxena, L. Li, J. Whang, E. Denton, S. K. S. Ghasemipour, B. K. Ayan, S. S. Mahdavi, R. G. Lopes, T. Salimans, J. Ho, D. J. Fleet, M. Norouzi: “Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding”, Curran Associates, Inc., 35, pp.36479-36494 (Apr.2022)
- [4] 本田 学: “感性的質感認知への脳科学的アプローチ”，映像情報メディア学会誌，vol.66, no.5, pp. 343-348 (Mar.2012)
- [5] T. Karras, T. Aila, S. Laine, and J. Lehtinen, “Progressive growing of gans for improved quality, stability, and variation”, in Proceedings of International Conference on Learning Representations (ICLR) 2018 (May. 2018)
- [6] L. A. Gatys, A. S. Ecker, and M. Bethge: “Image Style Transfer Using Convolutional Neural Networks”, in International Conference on Computer Vision and Pattern Recognition, pp. 2414-2423 (Jun.2016)
- [7] N. Sunda, K. Tobitani, I. Tani, Y. Tani, N. Nagata, and N. Morita: “Impression Estimation Model for Clothing Patterns Using Neural Style Features”, in International Conference on Human-Computer Interaction, pp. 689- 697 (Jul.2020)
- [8] T. Karras, M. Aittala, J. Hellsten, S. Laine, J. Lehtinen, and T. Aila: “Training generative adversarial networks with limited data”, in Advances in Neural Information Processing Systems, 1015, pp. 12104-12114 (Dec.2020)