

Conditional StyleGAN2-adaによるユーザの属性情報を考慮した

2Q-08

高評価画像の生成に関する研究

大池 樹* 山下 遥† 清水 良太郎‡
上智大学* 上智大学† ZOZO研究所‡

1 はじめに

近年、画像生成の技術が様々な分野に応用されており、ファッション業界への応用に関する研究も展開されている [1, 2]. これまでの研究では一般に、指定した衣服やポーズをもとにしたファッション画像の生成や、画像に写っている人物の衣服を変化させた画像の生成に焦点が当てられてきた. しかし、実サービス上に存在するユーザの属性は多様であるため、衣服やポーズのみを考慮すると、対象ユーザの属性に合わない画像を生成してしまう恐れがある. 例えば、年代を考慮しないと、高身長男性に対して子どものファッション画像を生成してしまう場合がある.

そこで、本研究では株式会社 ZOZO が運営するファッションコーディネートアプリ「WEAR」のデータを活用し、属性情報を考慮可能なファッション画像の自動生成アプローチを検討する. WEAR では各ユーザが自身のファッション画像の投稿、及び他ユーザの投稿の閲覧が可能である. さらに、各ユーザは気に入った投稿に対して、「いいね」や「お気に入り」といったリアクション (評価) を付与することができる. 本研究では、ユーザが投稿したファッション画像とそのユーザの性別・身長などの属性情報、各画像への評価を活用することで、いいね数が多く獲得できるようなユーザに合わせた「高評価画像」を生成することを目的とする.

本研究では、StyleGAN2-ada[3] を改良し、指定した条件の特徴を持つ画像生成を可能にするモデルである Conditional StyleGAN2-ada[3] を利用する. そして、Conditional StyleGAN2-ada の条件部分に、ユーザの属性情報やいいね数をもとに作成したクラスラベルを用いて学習させる. この学習モデルを用いることで、ユーザの属性に合わせた高評価画像の生成を実現する. 実験では、実際にユーザの属性に合致した高評価画像の生成が可能であるかに関する評価を行う. その際、属性に関しては目視による観察で十分に評価可能であると考えられる. 一方で、高評価画像であるかに関しては目視による判断が困難であるため、EfficientNetV2[4] という性能の良い画像分類モデルを用いて評価する方法を提案する.

2 先行研究

2.1 StyleGAN2-ada[3]

高解像度の画像を生成する最先端のモデルとして、StyleGAN2-ada[3] というモデルが提案されている. このモデルは、画像生成モデルの学習において、データ数が比較的少ない場合でも高解像度の画像を生成することが可能なモデルである. また、モデルの入力に条件を考慮するために、

損失関数を工夫することで、指定した条件に合わせた画像生成を可能にしている.

2.2 GAN のファッションへの応用

GAN をファッション分野へ応用した事例としては、指定した衣服の組み合わせと、ヒートマップで表現した衣服着用者の姿勢に基づいてファッション画像を生成する研究 [1] が存在する. また、一枚の画像を用意して仮想試着を実現させる研究 [2] も存在する. しかし、どちらの研究も、ユーザの属性情報や画像の評価情報を考慮していない.

3 クラスラベルの作成

3.1 クラスラベルの作成手順

本研究では、ユーザの属性情報と画像への評価 (いいね数) を考慮するために、図 1 のように、性別 (男性・女性・子ども) と身長、いいね数による層別を行うことでクラスラベルを作成する. まず性別ごとに身長の中央値を計算し、中央値以上を高身長、中央値未満を低身長として 6 つのクラスに層別する. さらに、その 6 つのクラスのいいね数の中央値を計算し、中央値以上をいいね数が多いクラス、中央値未満をいいね数が少ないクラスとして層別し、ラベルを付与する.

性別	身長	いいね数
男, 女, 子	高, 低	多, 少

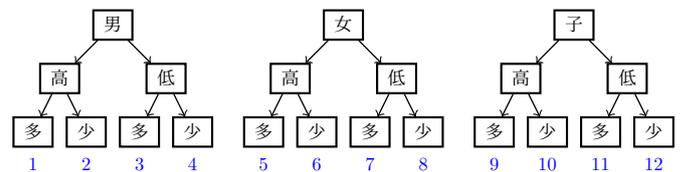


図 1: 性別、身長、いいね数による層別

3.2 クラスラベルの詳細

性別ごとの身長の基本統計量と、性別と身長ごとのいいね数の基本統計量を確認する. 表 1 が実際の性別ごとの身長の基本統計量である. 表 1 の中央値をもとに性別と身長による層別を行う.

表 1: 性別ごとの身長に関する基本統計量

性別	count	mean	std	min	25%	50%	75%	max
男性	49930.0	173.2	5.9	0.0	169.0	173.0	177.0	191.0
女性	44093.0	158.3	5.4	143.0	155.0	158.0	162.0	184.0
子ども	5977.0	109.2	17.1	60.0	100.0	110.0	123.0	167.0

性別・身長により、作成したクラスごとのいいね数の基本統計量が表 2 である. 表 2 の中央値をもとに合計 12 個のクラスラベルを作成する.

Study on generation of highly evaluated images considering user attribute information by Conditional StyleGAN2-ada

*Tatuki Oike · Sophia University

†Haruka Yamashita · Sophia University

‡Ryotaro Shimizu · ZOZO Research

表 2: 性別と身長により層別されたクラスごとのいいね数に関する基本統計量

クラス	count	mean	std	min	25%	50%	75%	max
1	26430.0	415.3	138.4	253.0	311.0	382.0	483.0	1521.0
2	23500.0	395.3	127.9	253.0	304.0	367.0	452.0	1506.0
3	25399.0	356.0	103.8	253.0	287.0	329.0	399.0	2044.0
4	18694.0	378.3	140.1	253.0	290.0	342.0	421.0	3074.0
5	3517.0	332.4	88.7	253.0	275.0	305.0	360.0	1197.0
6	2460.0	337.7	114.1	253.0	274.0	311.0	354.0	1133.0

4 実験

4.1 実験条件

本研究では、株式会社 ZOZO が運営するファッションコーディネート投稿アプリ WEAR に投稿された画像データと投稿ユーザの属性情報を用いる。実験で使用する画像の解像度は 256×256 であり、データの総数は 10 万である。計算には、NVIDIA の RTX3090(GPU) を使用し、エポック数は 20,000、バッチサイズは 16 とした。

4.2 実験結果

学習したモデルを使用して生成させた画像の例を図 2 に示す。上からクラスラベル 1, 2, ..., 12 を表し、クラスラベルごとの 10 枚ずつの生成画像を示している。これらの画像から、性別や身長によって異なる特徴を持つ画像が生成されていることが目視でわかる。生成画像の性別による違いは、男性・女性・子どもの衣服や体型の違いから容易に判別できる。また、生成画像の性別ごとの身長による違いは、ファッションの特徴から確認できる。男性の生成画像に着目すると、高身長男性の生成画像では、低身長男性の生成画像と比較してロングコートを着用しているものが多く、一方で女性に着目すると低身長女性の生成画像では、高身長女性での生成画像と比較してスカートを着用しているものが多い。また、子どもの画像に着目すると低身長の生成画像では、高身長の生成画像と比較して双子コーデのものが多い。以上より、属性情報によって生成画像に違いがあることがわかる。



図 2: 条件ごとの画像生成

4.3 EfficientNetV2 による評価

生成画像の属性情報に関しては、目視による観察で評価できるが、いいね数に関する評価は目視のみでは評価が困難である。そこで、高評価を得られる画像か否かを判別するモデルを予め学習し、そのモデルを用いて生成画像を判別することで評価を行う方法を提案する。本研究では、判別モデルとして高精度で効率的な EfficientNetV2[4] を利用する。表 2 の中央値を基準として高評価を得られる画像か否かを 1 か 0 で判別する 6 つのモデルを作成する。

また、本実験では実際の投稿画像について、訓練データ : 評価データ = 8 : 2 の割合で分割し、訓練データを用いて 6 つのモデルを学習させた。評価データによるモデルの正答率は 67 ~ 71% であった。この 6 つのモデルを用いて各クラスラベルごとに 100 枚ずつの生成画像を当てはめた場合の出力の平均値を表 3 に示す。

表 3: 画像分類モデルの出力の平均値

	クラスラベル											
	1	2	3	4	5	6	7	8	9	10	11	12
平均値	0.31	0.25	0.66	0.53	0.90	0.84	0.61	0.58	0.78	0.79	0.78	0.69

結果より、クラスラベル 1, 3, 5, 7, 11 の生成画像の出力の平均値が、それぞれクラスラベル 2, 4, 6, 8, 12 の生成画像の出力の平均値以上であることから、高い評価が得られそうな画像を生成できていることが確認できた。クラスラベル 9・10 に関しては、大きな違いは見られなかった。

5 おわりに

本研究では、Conditional StyleGAN2-ada を使用することでユーザの属性情報を考慮した高評価画像の生成を行なった。ユーザの属性情報に関する評価は生成画像を目視する事で行い、いいね数に関しては EfficientNetV2 に基づく評価を行なった。

今後の課題は、画像生成に使用したモデル自体を改良することで、より鮮明な画像を生成させることや、より画像の特徴を明確に分類できるクラスラベルを作成することである。また、いいね数に関する評価に用いた画像分類モデルの精度を上げ、より正確な評価をすることも今後の課題である。

参考文献

- [1] Gokhan Yildirim, Nikolay Jetchev, Roland Vollgraf, and Urs Bergmann. Generating high-resolution fashion model images wearing custom outfits. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, 2019.
- [2] Kathleen M Lewis, Srivatsan Varadharajan, and Ira Kemelmacher-Shlizerman. Tryongan: Body-aware try-on via layered interpolation. *ACM Transactions on Graphics (TOG)*, Vol. 40, No. 4, pp. 1–10, 2021.
- [3] Tero Karras, Miika Aittala, Janne Hellsten, Samuli Laine, Jaakko Lehtinen, and Timo Aila. Training generative adversarial networks with limited data. *arXiv preprint arXiv:2006.06676*, 2020.
- [4] Mingxing Tan and Quoc V Le. Efficientnetv2: Smaller models and faster training. *arXiv preprint arXiv:2104.00298*, 2021.