敵対性構造を用いた Stable Diffusion のプロンプト生成システムの開発

Using Adversarial Structures Development of Prompt Generation System for Stable Diffusion

コ　セイ [1]，高橋聡[*2]
Sei KO [*1], Satoshi TAKAHASHI[*2]
[*1] 関東学院大学　理工学部
[*1]College of Science and Engineering, Kanto Gakuin University
Email: m22j7002@kanto-gakuin.ac.jp

あらまし：本研究では,ブランドなどのテーマを保ちつつ,新たなデザインの衣服画像を生成できるプロンプトを特定するシステムを提案する.まず,参考となる画像からプロンプトを生成する．そして,そのプロンプトにより生成系 AI に画像を複数枚生成させる．さらに,参考となる画像と生成した画像を比較させ,プロンプトを改善させる．生成系 AI としては,Stable Diffusion を利用し,プロンプトの改善には GAN の枠組みを利用する.

キーワード：敵対的生成ネットワーク,画像生成 AI,プロンプト

1.　はじめに

デザイナーは,同じ雰囲気,あるいは同じ要素（ブランドなど）を含む服をデザインしたい場合がある.しかし,衣服デザインの初心者にとって,ブランドなどのテーマを保ったまま一から新たなデザインを生み出すことは難しい.

それに対して,本研究では,参考となるデザインの衣服の画像生成する時、生成系 AI が大きな助けとなると考えられる.

一方で,生成系 AI に対して適切なプロンプトと呼ばれる命令文に従って画像を生成する．プロンプトとは命令文を与えることである．そのため，生成したい画像に適したプロンプトを設計することが課題となる.

生成型人工知能におけるプロンプトの操作を通じて，生成された画像の制御を実現することができる.提供された画像に含まれる情報に対応するプロンプトを見つければ、画像のブランドなどの特性を保持したまま、類似画像を生成することができる.

2.　目的

本研究の目的は画像から適切なプロンプトを生成するシステムを構築することである.画像に含まれる要素を抽出してプロンプトを作成し,訓練によりプロンプトを更新することで,写真に最もマッチするプロンプトを探し出すことである.提案システムではまずテーマの参考となる画像からプロンプトを生成する．そして,そのプロンプトにより生成系 AI に画像を複数枚生成させる.さらに,参考となる画像と生成した画像を比較させ,プロンプトを改善させる.生成系 AI としては,Stable Diffusion を利用し,プロンプトの改善には Generative Adversarial Network (GAN)の枠組みを利用する(1).

3.　システムの概要

提案手法の概要を図 1 に示す．提案システムでは,まずテーマの参考となる画像からプロンプトを生成する．図 1 通り、この画像から“Leather jacket，khaki coat，shorts，black boots，street”などの単語をプロンプトとして生成する.そして,そのプロンプトにより生成系 AI に画像を複数枚生成させる．生成された画像はデータセットとして使用する.さらに,参考となる画像と生成した画像を比較させ,偽と判定する場合、画像を Text Generator に返す、より適切な単語をプロンプトに追加して再度学習させる.プロンプトをこのように改善させる．生成系 AI としては,Stable Diffusion を利用しプロンプトの改善には GAN の枠組みを利用する.

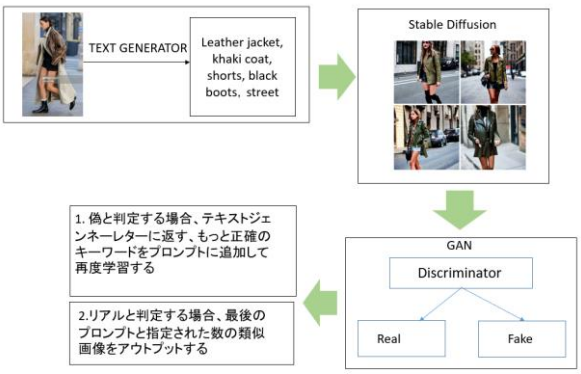

図 1 提案システムの概要
（画像は Getty Images より引用(2)）

3.1 参考画像からプロンプトの生成

参考画像からのプロンプトの生成には,Transformer と呼ばれる Deep Learning のモデルを利用する.提案手法では,これを Text Generator と呼ぶ.このモデルでは,画像と服の組み合わせが学習されている．画像を入力することで,画像内に存在する服が文字として出力される(3).

図 2 に例を示す．ワンピースの画像が入力され,その画像から“オーバースカート”や“礼服”などが認識されて出力されている.

抽出された単語を利用してプロンプトを組み立てる.

図 2　テキストジェネレーターの結果
（画像は Taobao より引用(4)）

3.2 Stable Diffusion の利用

Text Generator により,参考画像から生成したプロンプトを Stable Diffusion に入力し,画像を生成する.

図 2 で示した画像を元にしたプロンプトは“overskirt,gown,hoopskirt,crinoline,cloak,velvet,theater curtain,abaya,showercurtain,lampshade,lampshade,vestment”のようになる.

3.3 Generative Adversarial Network による更新

Stable Diffusion から生成された画像と参考にした画像を比較し,その差に基づいて Text Generator のモデルを更新する．モデルの更新には,GAN の構造を用いる．GAN の構造を図 3 に示す.GAN のネットワーク構造は,Generator（生成ネットワーク）と Discriminator（識別ネットワーク）の２つのネットワークから構成される.Generator で生成した画像を,Discriminator が“実際に撮影された画像”か“Generator で生成した画像”を判別する．そして,Generator は自らが生成した画像が“実際に撮影された画像”と判断されるようにモデルを更新していく.

現在検討中の Text Generator の学習方法以下である.

1.プロンプトを検査する.同じではない色や長さなどの矛盾の単語を識別し,不正確な単語を取り除く.

2. プロンプトの中で画像と関連性が低い単語を削除する.たとえば図 2 の例の“theater curtain”などの服装と関係ない単語である.

3.参考となる画像から服装本体の部分のみを抜き出す．そして,抜き出した画像を再び Text Generator に入力することで,背景に関わる単語が抽出されないようにする.

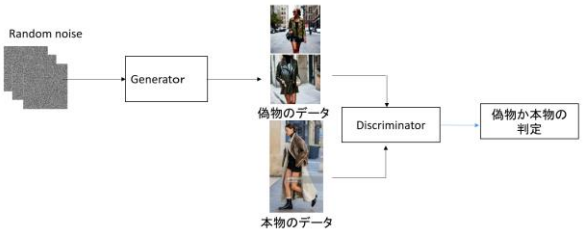

図 3 Generative Adversarial Network の構造

4. プロンプト更新の流れ

提案手法を用いたプロンプト更新の流れを図 4 に示す．ユーザーは 1 枚の画像をシステムに入力する.Text Generator で特徴を抽出し,関連性が高い単語を利用して,プロンプトとして Stable Diffusion に入れる.提供されたプロンプトをもとに,100 枚の画像を Stable Diffusion でデータセットとして生成する.

生成された画像を Discriminator で判断する.生成された画像が偽と判断すると,画像を Text Generator に戻す. Text Generator に矛盾する単語を削除や服装本体の部分のみを抜き出すなどの手段で新しい単語を追加する.その後,同じ流れを繰り返す.識別器が正しい画像と判断するまで続けて訓練する．この訓練の流れを何回も繰り返すと,対応する要素を見つけることが期待できる．最終的には,大量の類似写真と提供される画像のプロンプトともに表示されるようになる.

なお，図 4 では、Text Generator によって画像から多くの特徴を抽出されている.この中の関連性が低い単語を削除し、プロンプトとして使う.また，現在，使用されている Text Generator は、衣服の色の認識は比較的不正確であるため，衣服の本体を抽出し、色を再識別する.このような精度向上操作を数回行うことで、画像に対応するプロンプトを生成することができると考えられる.

図 4 プロンプト更新の流れ

5. まとめ

本研究では,ブランドなどのテーマを保ったまま新たなデザインの参考となる衣服画像を生成できるプロンプト生成システムを提案した．現在,Text Generator の更新方法の検討を行っているところである．また,出来上がった服装にデザイナーが満足するかどうかも検証予定である.

今後の展望として,画像に対する認識精度をどのように向上させるかを考える．生成された服装の満足度をどのように高めるかを検討し,操作性,認識品質の向上,有効性があるシステムの構築を目指す.

参考文献

(1) Radford , A. and Metz, L. and Chintal, S.:“Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks” ICLR 2016（2016）
(2) Getty Images : “Getty Images”, www.gettyimages.co.jp(参照 2023.05.23)
(3) Dosovitskiy, A., et al.:“An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale” ICLR 2021 (2021)
(4) Taobao: “タオバオ”, https://m.tb.cn/h.UxHlklE?tk=IMFfdLyeDuQ CZ3457(参照 2023.05.23)