

StarGAN を用いた二次元キャラクターの表情生成

Generation of 2D character's expression using StarGAN

2020Y30 三上 麟太郎 (Mikami Rintaro)

担当教員 出口 利憲 (Deguchi Toshinori) ・ 山田 博文 (Yamada Hirobumi)

1. 序論

近年 AI の創造性の獲得が話題となっている。それと同時にアニメやゲームの製作現場では、表情のみ変更したキャラクターのイラストが多く用いられている。本研究ではこの創造性を用いて、商業、趣味においてイラスト制作の補助をすることを目的に 2 次元キャラクターイラストの表情の変更を行う。

2. StarGAN

変換画像の生成には StarGAN を用いた。

StarGAN[1]は、Yunjey Choi 氏らが発表したネットワークモデルである。従来では 3 つ以上のドメイン間での変換を行う場合、それら全てが完全結合となる必要があった。しかし、StarGAN では学習時に画像が元々のデータ画像であるかどうかの判別に加え、その画像のドメインも判別対象とするため、1 つのモデルで複数ドメイン間の変換を行うことが可能となった。また、出力される画像において、変換先であるターゲットドメインに関する部分のみが変換されるようにするため、Generator が生成した画像を変換前のドメインへ再度変換し、元の画像との比較を行うことでサイクル一貫性損失を導出、これを最小化させるように学習することで問題を解決している。

3. 実験方法

ゲームアプリのキャラクター画像をキャラクターの顔のみが中央に収まるように、かつ画像が正方形になるようにトリミングを行い、

画像のキャラクターの表情を“左目でウィンクをしている状態”、“口を閉じて口角を上げている状態”、“両目を閉じてにっこりしている状態”、“両目を開いた状態”、“口を開けて笑っている状態”の 5 つに分類する。これにより集まった約 2500 枚の画像を学習データとし、StarGAN の学習を行う。学習のためのネットワークモデルは StarGAN の発表者である Yunjey Choi 氏らが Github 上にアップロードしたものを用いる [1]。学習は 200000 回繰り返す。

4. 実験結果と考察

Figure.1、Figure.2 に実験結果を示す。

これらは左から順に(a)が元画像、(b)がウィンク、(c)が口を閉じる、(d)が両目を閉じる、(e)が両目を開く、(f)が口を開くというドメインを表している。また、上から順に 50000 回、100000 回、150000 回、200000 回と繰り返した際の結果である。

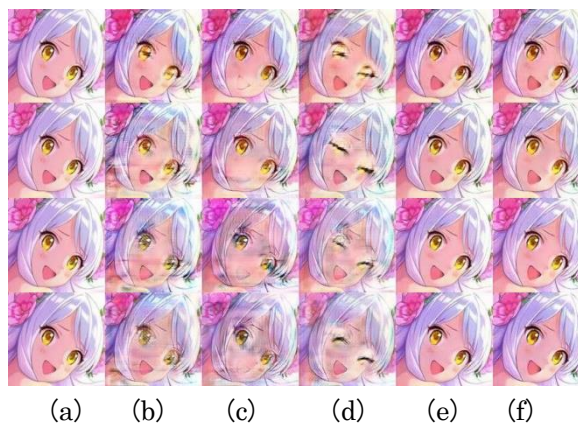


Figure.1 Result A.



(a) (b) (c) (d) (e) (f)

Figure.2 Result B.

Figure.1 では、特に(d)列の目を閉じてにっこりとさせる変換が明確に元の画像と異なり、かつ目的とするドメインへの変換が行われていることが分かる。Figure.2 でも、100000 回の結果まではうまく変換が行われているが、150000 回からの結果では変換が行われていないという結果となった。これは、元々の画像が片目を閉じているという点に加え、両目を閉じるに分類された教師データの数が約 150 件と少なかったために、学習回数が過剰となってしまったためであると思われる。

一方で、Figure.2 の(e)列、両目を開くという変換はデータ数が約 1000 件存在していたため、200000 回の学習が過剰とならず、次第に右目と似た左目の生成が行われていくことが確認できる。

また、Figure.1 と 2 に共通して、(c)列の口を閉じるという変換は、口を閉じた結果その周辺の部分で見える肌色の面積が増えるというような変換が行われている。こちらのデータ数は約 300 件であり、Figure.2 の 200000 回目には黒いラインも生成されているためあと一步目標に及ばないという結果となった。

最後に、Figure.3 に Generator と Discriminator の損失関数を示す。

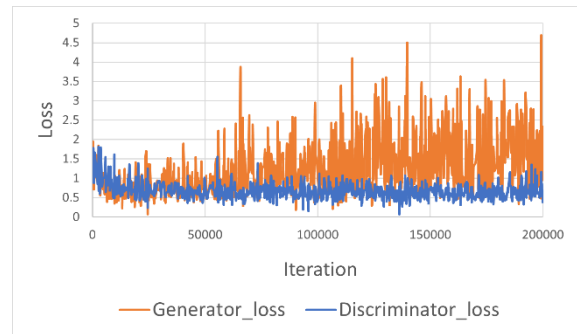


Figure.3 Generator and Discriminator's loss.

これらは、本来どちらも最終的に値が 0 に収束するように学習が行われていくものである。しかし、Discriminator が 0 に収束していているものの Generator が発散してしまっている。これは前述の通りデータ数が少なく最終的に学習が過剰になったドメインが複数存在するためであると思われる。

5. まとめ

今回の実験では、一部変換の成功が見られた部分もあったが、全体的に見ると学習に用いるデータ数の少なさにより学習が失敗したと思われる結果となった。また、今回使用したネットワークは、Yunjey 氏らが実在の人物の画像に対し適用したものであるため、それと比較して顔を構成するパーツの大きさ、形などに大きな差が生じるイラストに関してはハイパーパラメータが適切でない可能性もある。そのため、今後はデータ数の増加に加え、それらの調整も行いながら実用可能なレベルの変換を目指していきたい。

6. 参考文献

[1]Yunjey Choi, et al.StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 8789-8797