Vol.2022-ICS-205 No.7
2022/2/21

ネコからアニメキャラクターへの画像翻訳手法の検討

徐 江林[1,a)]　清 雄一[1,b)]　田原 康之[1,c)]　大須賀 昭彦[1,d)]

概 要 ：画像から画像への翻訳（Image-to-Image translation）は, GAN (Generative Adversarial Network) [1]のタスクの一つとして長く扱われてきた.近年,GANの発展につれ,ある程度の形状変化が伴う翻訳タスクがこなせるようになった.本研究では,今まで試されなかった「ネコからアニメキャラクターへの画像翻訳」というタスクに挑戦した.そのため,ベースライン手法として,形状変化に対応できるCouncilGAN[2]とDSMAP[3]を用いて,目標タスクにおいて高いパフォーマンスを発揮できるように,パラメータや正規化手法などの調整を施した.データセットには,ネコの顔画像とアニメキャラクターの顔画像を採用した.さらに,より良い生成結果を期待して,AnimeGAN[4]を用いて,ネコの顔画像に前処理を施したものをデータセットとして採用し,元の結果と比較した.客観評価としてFIDスコアを採用し,主観評価としてアンケートを実施して,翻訳結果に与えられた改善効果を評価した.FIDスコアから,調整によってCouncilGANの精度を向上させることができたことと,前処理が生成結果の精度を低下させていないことがわかった.その他,ネットワークアーキテクチャの変更も試みたが,変更前のCouncilGANと比べて,より良いFIDスコアは達成できなかった.また,アンケートの結果は,データセットへ前処理を施すことによって,主観的により良い結果を得られたことを示した.

A Study of Image Translation Methods from Cats to Anime Characters

JIANGLIN XU[†1,a)]　YUICHI SEI[†1,b)]　YASUYUKI TAHARA[†1,c)]
AKIHIKO OHSUGA[†1,d)]

1. はじめに

近年,動物の擬人化が漫画,小説,アニメ,ゲームなど,様々な作品において広く見受けられている.その中でもとくに,ネコの擬人化はその愛らしさによって,商業作品にだけではなく,個人創作にも絶大な人気を誇っている.身近にいるネコを擬人化する方法は,創作者にはもちろん,ネコの擬人化に興味がある人たちにとっても,インスピレーションと面白さに繋がるサービスになるだろう.現に,写真で撮ったネコをキャラクターに変換するアプリゲームが存在しているが,あくまで実装されたキャラクターの中から一番似てそうなキャラクターがピックアップされて出現する仕組みになっているので,限られたパターンを打破することが出来ないというのが現状である.

本研究では,ネコからアニメキャラクターへの画像翻訳を実現するための手法を模索することを目的としている.

2. 関連研究

2.1 CouncilGAN

教師なし学習による画像から画像への翻訳は,Cycle Consistent Lossを用いたCycleGAN[5]とそれに改行を重ねた発展研究が近年の主流であるが,大きな形状変化が伴うタスクに対応できなかったり,生成される画像に入力画像の痕跡が残っていたりして,未解決の課題も多くある.

Council GANにおいて,図１のように,一つのGeneratorと二つのDiscriminatorをワングループとし,このグループを複数に用意して学習を進めている.各グループのGeneratorは,他のグループのGeneratorと同じ入力画像を受け取り,独自の出力画像に翻訳する.Discriminatorの方では,図２のように通常のDiscriminator以外に,同じグループのGeneratorの生成した画像と,他のグループのGeneratorの生成した画像を区別するように学習するDiscriminatorも配置されている.これによって,各グループがお互いの生成結果を制約し,収束させ,相互情報量の最大化が図れて,生成された画像がソース画像の重要な特徴を維持するができる.結果として,CycleGANなどが苦手とする形状変化もある程度実現できるようになった.

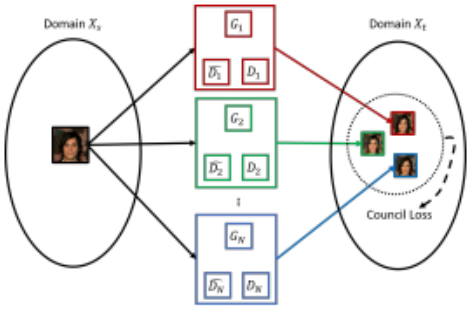

図１　CouncilGANのネットワーク(出典:[2]　(2020)p.7862)

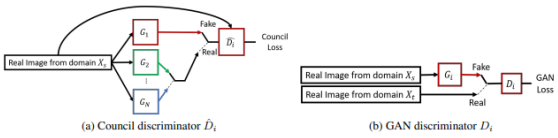

図２　CouncilGANの2種類のDiscriminator(出典:[2] (2020)p.7863)

1　電気通信大学
UEC,Chofu,Tokyo 182-8585,Japan
a)　xu.jianglin@ohsuga.lab.uec.ac.jp
b)　seiuny@uec.ac.jp
c)　tahara@uec.ac.jp
d)　ohsuga@uec.ac.jp

2.2 DSMAP

過去のImage-to-image translationは,主に画風変換のような,大きな形状変化が伴っていない翻訳を扱ってきた.そのため,従来の手法では,図３の(a)のように,ドメインのコンテンツ空間が共有されていると仮定することが多く,ドメインAからドメインBへの画像翻訳は,ドメインAのコンテンツ特徴を,ドメインBのスタイル特徴と組み合わせることで達成している.しかし,共有されているコンテンツ空間には,ドメインに関連する情報が含まれている可能性があるため,そのままコンテンツ特徴を変換に持ち込むと,変換先のコンテンツ表現力が損なわれて,形状変化に対応しきれない可能性がある.

この問題を解決するために,DSMAPでは,図３の(b)のように,ドメイン固有マッピング関数を二つ追加することで,共有されたコンテンツ空間のコンテンツ特徴を,異なるドメイン固有のコンテンツ空間に再マッピングする.ドメイン固有マッピング関数を見つけるために,新しい\domain-specific content reconstruction lossというロス関数が追加され,これを最小化することで,マッピング関数を得ることができる.

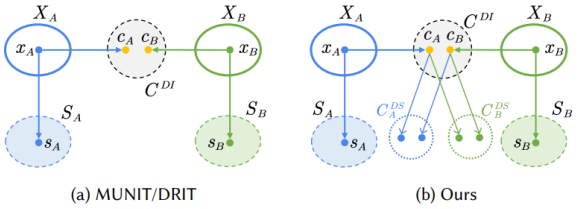

図３　DSMAPのアイデア(出典:[3] (2020)p.2)

2.3 AnimeGAN

画像のアニメスタイルへの変換は，難しい課題であり，既存の研究では満足の行く結果を得ることができなかった．その問題点として，　生成された画像にアニメーションの質感がない，　生成された画像に元となる画像の情報が失われているなどの点が上げられている．

AnimeGANは，grayscale style loss, color reconstruction loss, grayscale adversarial lossという，三つの新しい損失関数を提案することで，生成された画像の精度を向上させて，視覚効果において既存手法を大幅に上回る結果で画像を生成できている．Generator側では，grayscale style lossによって，生成画像の質感と線をより明確的にアニメスタイルにできるが，同時に生成画像をグレースケールにしやすいため，　color reconstruction　lossで元の写真の色を失わないようにしている．Discriminator側では，grayscale adversarial lossによって，生成画像がグレースケール画像ではないことが保証される．さらに，パラメータ数とモデルサイズの軽量化を実現し，既存手法より少ないメモリ容量で動作できている．　Generatorのアーキテクチャは図４に示す．

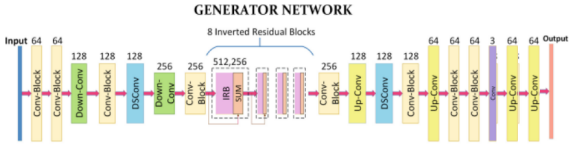

図４　AnimeGANのGeneratorのアーキテクチャ(出典:[4] (2019)p.246)

3. 提案手法

3.1 提案手法概要

本研究では,「ネコからアニメキャラクターへの画像翻訳」という大きな形状変化が伴う翻訳タスクに挑戦するため,ある程度の形状変化に対応できるCouncilGANとDSMAPをベースラインとして採用し,以下四つの試みを行った.

DSMAP：パラメータ調整及びデータセットへの前処理
CouncilGAN：パラメータ調整
パラメータ調整後のCouncilGAN：ネットワークアーキテクチャの変更
パラメータ調整後のCouncilGAN：データセットへの前処理

3.2 DSMAP：パラメータ調整及びデータセットへの前処理

既存のDSMAPを20万エポック程学習させたところ,図５のように,生成されたキャラクターの目の形が同じようになったり,生成された画像にアーティファクトが発生したりして,期待されたような出力を得られなかった.そのため,以下の調整を試みる.

(1)512×512以下の，サイズの小さい画像にあまり効果がないと言われているdomain-invariant perceptual lossというロス関数の削除した.

(2)domain-specific content reconstruction lossを始めとしたreconstruction　loss関数を通常のロスとランダムサンプリングでのロスの両方で計算するようにした.

(3)正規化しない畳み込み層をインスタンス正規化するようにした.

(4)入力画像のサイズを192×192から256×256に拡大した.

(5)バッチサイズを2から3に拡大した.

(6)データセットにあるネコの写真をアニメスタイルに変換するような前処理を施した.

図５　既存のDSMAPの生成結果

3.3 CouncilGAN:パラメータ調整

既存のCouncilGANでは,図６に示したように,ネコからアニメキャラクターへの翻訳において,「入力が違うにもかかわらず,似たような画像を出力する」という,モード崩壊現象が起きてしまった.そのため,目標タスクをこなすために,以下のことを試みる.

(1)CouncilGANは構造が複雑なため,計算量と学習時間を考慮すると,ロス関数の構成がシンプルである必要があると考えられる.そのため,ロス関数に対して,新規の追加をせず,タスクに適するバランスを模索した結果,GAN　lossの重みを26,Council　lossの重みを3,Focus　lossの重みを0に調整した.

(2)正規化手法において,Batch　Normalization,Layer Normalization,Instance　Normalizationを試し,本人の主観評価により,一番精度の良い結果を達成したInstance Normalizationを採用した.

(3)入力画像のサイズを128×128から256$×256に拡大した.

(4)バッチサイズを3から4に拡大した.

Vol.2022-ICS-205 No.7
2022/2/21

図６　既存のCouncilGANの生成結果

3.4　調整後のCouncilGAN：ネットワークアーキテクチャの変更

モデルのパフォーマンス向上を期待して,ネットワークアーキテクチャの変更を視野に入れることにした.そこで,Face Portrait変換において著しい成果をあげたAnimeGANのアーキテクチャを参考した.
図４に示されたAnimeGANのGeneratorのアーキテクチャを参考にして,CouncilGANのGeneratorのアーキテクチャを図７に変更して, 実験を行う.

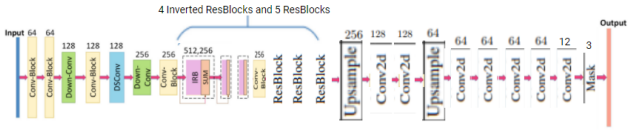

図７　変更後のCouncilGANのGeneratorのアーキテクチャ(CouncilGANによる図(出典:[2](2020)Supplementary4 Architecture,pp.1-2)を参考にAnimeGANによる図(出典:[4](2019)p.246)を一部修正)

3.5　調整後のCouncilGAN：データセットへの前処理

生成結果の精度を上げるために,ソースドメインとターゲットドメインのドメイン間距離を縮めるための,ネコの顔画像をアニメスタイルに変換する前処理を,セクション3.2と同様にを施した.アニメスタイルへの変換には,AnimeGANを用い,変換においてネコの顔のパーツの輪郭と境界が保てるように調整する.

4.　画像翻訳実験

4.1　データセット概要

本研究では,アニメキャラクターの顔画像とネコの顔画像をデータセットとして採用している.アニメキャラクター側のデータセットとして,selfie2anime[6]データセットから3500枚ほどのアニメ顔画像を採用する.ネコ側のデータセットとして,AFHQ[7]データセットから5500ほどのネコ写真を,顔検出器にてある程度顔抽出処理を施したものを採用する.前処理を施す場合,ネコ側のデータセットをAnimeGANにてアニメスタイルに変換する.

4.2　3.2に関する実験結果

上記(1)～(5)までの調整を施して,30万エポックほど学習させた実験結果を図８,データセットへの前処理を含めて(1)～(6)まで施した実験結果は図９に示す.

図８　DSMAP実験結果(パラメータ調整のみ)

図９　DSMAP実験結果(パラメータ調整＋データセット前処理)

4.3　3.3に関する実験結果

調整後のCouncilGANの50万イテレーションにおける生成結果を図１０に示す.また,調整前と調整後のCouncilGANの50万イテレーションにおけるFIDスコアを,表１に示す.

図１０　CouncilGANの生成結果（パラメータ調整後）

表１　3.3に関するFIDスコア(50万イテレーション)

CouncilGAN	調整後のCouncilGAN
222.9	102.4

4.4　3.4に関する実験結果

アーキテクチャ変更後のCouncilGANを用いて,30万イテレーションにおける生成結果を図１１に示す.FIDスコアを用いて,アーキテクチャ変更前とアーキテクチャ変更後のCouncilGANの30万イテレーションにおける生成結果を客観評価し,表２に示す.

図１１　CouncilGANの生成結果（パラメータ調整+アーキテクチャ変更)

表２　3.4に関するFIDスコア(30万イテレーション)

調整後のCouncilGAN	調整後のCouncilGAN＋アーキテクチャ変更
103.7	130.6

4.5　3.5に関する実験結果

前処理したデータセットを用いて50万イテレーションにおける生成結果を図１２に示す.FIDスコアを用いて,データセットへの前処理なしと前処理ありのCouncilGANの50万イテレーションにおける生成結果を客観評価し,表３に示す.また,データセットへの前処理なしと前処理ありの結果について,「どちらの結果の方がアニメキャラクターとして違和感が少ないですか？」と「どちらの結果において頭/顔/髪

の輪郭が維持できていますか？」という二問を設けて,ウェブアンケートにて主観評価を行った.20人から回答をいただき,その結果を表４に示す.

図１２　CouncilGANの生成結果（パラメータ調整+データセットの前処理）

表３　3.5に関するFIDスコア(50万イテレーション)

調整後のCouncilGAN	調整後のCouncilGAN＋データセットへの前処理
102.4	99.56

表４　アンケート結果

	調整後のCouncilGAN	調整後のCouncilGAN＋データセットへの前処理
違和感が少ない	0.26	0.74
輪郭が維持できている	0.23	0.77

5.　考察

5.1　3.2に関する考察

図５の結果と比べると,図８の結果において,アニメキャラクターの頭の輪郭が曖昧になっているが,発生したアーティファクトは少なくなっている.これは,一部のロス関数の削除によって,制約が弱くなったことが原因だと推測する.図８と図９を見比べると,前処理によって,翻訳結果の精度が上がったことがわかる.しかし,アーティファクトが完全に消えることがなく,翻訳精度に大幅の改善が見られなかったため,より精度の良い結果を達成できているCouncilGANをメインに実験を進めることにした.

5.2　3.3に関する考察

図６と見比べると,図１０において,違う入力に対して,違う結果画像が出力され,各出力は明らかに似ているような現象がないことがわかる.また,調整後のCouncilGANのFIDスコアは調整前より大幅に小さくなり,調整によって生成精度が高くなっていることがわかる.以上の結果から,モード崩壊は回避できたと考えられ,以降,より良いパフォーマンスを達成するために手法の改善案を検討することにした.

5.3　3.4に関する考察

FIDスコアから,精度の向上は図れなかったことがわかる.ネットワークアーキテクチャ変更の狙いは,精度向上の他,精度を維持した上での軽量化による学習時間を短縮も含まれていた.しかし,ある程度のネットワークの深さを確保しないと,図１３のように,出力パターンが制限され,マルチモーダル変換ができなくなる.これは,パラメータ数が足りないことによって,モデルの表現力が制限されていると推測する.

今回のアーキテクチャ変更において,AnimeGANからinverted residual blocks(IRBs)を導入した.IRBsは,オリジナルのCouncilGANが使っているresidual blocksと比べると,ネットワークのパラメータ数と計算負荷を大幅に削減することができる.しかしその代わりに,モデルの表現力が失われるため,大きな形状変化が伴うマルチモーダル翻訳タスクとの相性が良くなかったと考えられる.

図１３　CouncilGANの生成結果（アーキテクチャ変更失敗例）

5.4　3.5に関する考察

FIDスコアから,データセットへの前処理が生成結果の精度を低下させていないことがわかる.図１２と図１０を見比べると,図１０の左から1列目にある生成結果のような,頭や髪の輪郭がうまく保てていない生成結果が減っているように見える.これを踏まえて,予定していた主観評価アンケートに,「アニメキャラクターとしての違和感が少ない」の他,「輪郭が維持できている」という項目を追加した.アンケートの結果から,データセットへ前処理を施すことによって,両項目において共に主観的により良い評価を得られた.輪郭が維持できていることへの支持率は,違和感が少ないことへの支持率より僅かに高くなっていることから,輪郭の維持が違和感の減少に貢献していると推測できる.

5.5　全体に関する考察

結果画像として,ネコの正面顔の翻訳を載せたが,ネコの横顔写真のような,正面顔写真以外のものを入力として与えた時に,顔のパーツが正面配置のまま,顔の輪郭だけが横顔になるキャラクターのような生成結果が出力される場合が多い.これは,アニメキャラクターの顔のデータセットには,正面顔の画像しか含まれていないことが原因だと考えられる.同時に,CouncilGANには,正面だけではなく,横顔とか,様々なポーズにおけるネコからアニメキャラクターへの翻訳に対応できるポテンシャルがあるとも考えられる.

また,CouncilGANは,CycleGANベースの手法と違って,生成結果を入力結果へ戻すことが可能である必要があるという, cycle-consistencyの制約を回避している.これによって,双方向に行き来するために変化量を制限する必要がなくなり,生成された出力に隠れたソースドメインの情報が保存されていることも避けられる.DSMAPの実験に発生したアーティファクトと目の形が似ている現象が,CouncilGANの実験に発生しなかった原因の一部は,cycle-consistencyにあるのではないかと推測する.

6.　まとめ

本研究では,ネコからアニメキャラクターへの画像翻訳を実現するための手法を模索した.提案手法を用いた画像翻訳結果は,ベースライン手法より良い精度を達成できた.

しかし,生成結果はまだアニメキャラクターとして違和感がない程度に達していなくて,入力がネコの正面顔以外の場合での翻訳に対応しきれていない.また,理想であるソース画像のネコの特徴を結果に反映する翻訳は達成されていない.

今後の展望として,新しいモジュールやネットワードを採用することによって生成結果の精度を向上させることが可

Vol.2022-ICS-205 No.7
2022/2/21

能と思われる.また,アニメキャラクター側のデータセットに,キャラクターの横顔画像の追加,あるいはネコ側のデータセットに,横顔を正面顔に変換する前処理を施すことによって,より幅広いポーズに対応できる翻訳が期待できる.生成結果への毛色を始めとしたネコの特徴の反映は,Coarse StyleのStyle-mixing[8]によって実現できるのではないかと考えられる.

謝辞 本研究を行うにあたって,ご多忙の中指導してくださりました大須賀昭彦教授,田原康之准教授,清雄一准教授に深く感謝申し上げます.
また,お世話になった大須賀研究室の皆様と実験にご協力を頂いた被験者の皆様にも,深くお礼申し上げます.
本研究はJSPS科研費JP18H03229,JP18H03340,JP18K19835,JP19H04113,JP19K12107,JP21H03496の助成を受けたものです.

参考文献

[1] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio. Generative Adversarial Networks, Advances in Neural Information Processing Systems 27, pp.2672-2680 (2014)
[2] Ori Nizan, Ayellet Tal. Breaking the cycle - Colleagues are all you need, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 7860-7869
[3] Hsin-Yu Chang, Zhixiang Wang, Yung-Yu Chuang. Domain-Specific Mappings for Generative Adversarial Style Transfer, ECCV 2020, arXiv:2008.02198 [cs.CV]
[4] Jie Chen, Gang Liu, Xin Chen. AnimeGAN: A Novel Lightweight GAN for Photo Animation,ISICA 2019: Artificial Intelligence Algorithms and Applications pp.242-256
[5] Jun-Yan Zhu, Taesung Park, Phillip Isola, Alexei A. Efros. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks, In ICCV 2017, arXiv:1703.10593 [cs.CV].
[6] Junho Kim, Minjae Kim, Hyeonwoo Kang, Kwanghee Lee. U-GAT-IT: Unsupervised Generative Attentional Networks with Adaptive Layer-Instance Normalization for Image-to-Image Translation, 26 Sept 2019, ICLR 2020 Conference Blind Submission
[7] Yunjey Choi, Youngjung Uh, Jaejun Yoo, Jung-Woo Ha. StarGAN v2: Diverse Image Synthesis for Multiple Domains, In CVPR 2020, pp. 8188-8197, arXiv:1912.01865 [cs.CV].
[8] Tero Karras, Samuli Laine, Timo Aila (NVIDIA), A Style-Based Generator Architecture for Generative Adversarial Networks, arXiv:1812.04948v3 [cs.NE], pp.1-12, 29 Mar 2019