

Conditional GANによる化粧顔画像変換

五味 京祐[†] 柳井 啓司^{††}

[†] 電気通信大学 情報理工学部 総合情報学科

^{††} 電気通信大学 大学院情報理工学研究科 情報学専攻

E-mail: [†]{gomi-k,yanai}@mm.inf.uec.ac.jp

あらまし 近年, Generative Adversarial Networks(GAN) を用いた画像生成や変換の研究が盛んに行われている. その中で顔画像の化粧に焦点を当て, 化粧を付加・除去する変換の研究も多い. しかしながらそうした研究の多くは, 化粧前顔画像と化粧後顔画像を入力とし化粧後画像の化粧は除去し, 化粧前画像には化粧後画像と同じ化粧を付加することを目標としている. そこで本研究では, 国籍でラベル付けされた同一人物の化粧前後の顔画像がペアになったデータセットを用いることで, 入力した顔画像に指定した国らしい化粧を付加する変換を学習する. 同時に, 顔画像から化粧を除去する変換を実現することも目標とする. 実験の結果, StarGAN を用いた場合は国ごとに異なる化粧を付加することに成功し, pix2pix では化粧の除去を実現することができた.

キーワード Generative Adversarial Network, 深層学習, 画像変換

1 はじめに

近年, Generative Adversarial Networks(GAN) を用いた画像生成や変換の研究が盛んに行われている. その中で顔画像の化粧に焦点を当て, 化粧を付加・除去する変換の研究も多い. しかしながらそうした研究の多くは, 化粧前顔画像と化粧後顔画像を入力とし化粧後画像の化粧は除去し, 化粧前画像には化粧後画像と同じ化粧を付加することを目標としている. そこで本研究では, 国籍でラベル付けされた同一人物の化粧前後の顔画像がペアになったデータセットを用いることで, 入力した顔画像に指定した国らしい化粧を付加する変換を学習する. 同時に, 顔画像から化粧を除去する変換を実現することも目標とする.

本研究では, 汎用的な 2 ドメイン間の画像変換手法である pix2pix [1] に Conditional Vector の入力を追加した手法と, 複数ドメイン間の画像変換手法である StarGAN [2] を用いて, 目標の化粧顔画像変換の実現を試みる.

2 関連研究

2.1 画像生成・変換

Conditional GAN [3] は GAN [4] にノイズベクトルだけでなく Conditional Vector を入力することで, 複数クラスの画像生成を学習し直すことなく, 実現できる.

pix2pix [1] は, 画像ドメイン間の変換をペアになった画像データセットから学習する手法である. 学習に用いる画像データセットがドメイン間でペアになっていなければならないという制約があるが, あらゆる画像に適用可能な汎用性の高い手法であり, 特にモノクロ画像のカラー化や線画着色など不可逆変換も学習によって実現可能な手法となっている.

CycleGAN [5] は, pix2pix において存在した, 学習データセットがペアになっていなければならないという制約をなくし,

ドメイン間の画像変換をペアになっていない各ドメインの画像集合から学習する手法である.

StarGAN [2] は, 複数ドメイン間の相互変換を実現した手法である. CycleGAN を複数ドメイン間の変換も可能なように拡張した手法であるから, 学習では画像はペアになっている必要はなく, 画像とターゲットドメインのベクトルを入力することで画像をターゲットドメインに変換した画像を出力する.

2.2 化粧画像変換

Chang らの研究 [6] では, 本研究と同様に化粧の顔画像変換を行っている. CycleGAN をベースとした手法を用いて, 顔のパーツごとに変換をすることで高解像度での画像変換を実現している. ただし, この研究は化粧前画像と化粧後画像の 2 枚を入力し, 化粧後画像は化粧除去, 化粧前画像は化粧後画像と同じ化粧を付加している.

Li らの研究 [7] では, Deep Learning を用いた化粧の除去を行っている. 化粧による顔認証の精度低下を抑えるため顔特徴量抽出の前処理として化粧除去を加えることで, 化粧顔を含む顔認証において state-of-the-art の精度を出した.

BeautyGAN [8] は, 化粧顔画像変換の研究である. 化粧を色変化とみなし, マスク画像で切り出した目・口・肌それぞれの色分布が化粧後と近づくように学習することで, パーツごとに適切な位置に化粧が付加される化粧顔画像変換を行った.

3 手法

本研究では, 特定の国らしい化粧を付加する変換を実現することを目標としている. そのために, 変換ネットワークに画像だけでなく国を表す Conditional Vector を与えることで画像変換をする Conditional GAN [3] を用いる.

3.1 pix2pix ベース

pix2pix は, 学習においてドメイン間のペアになった画像が

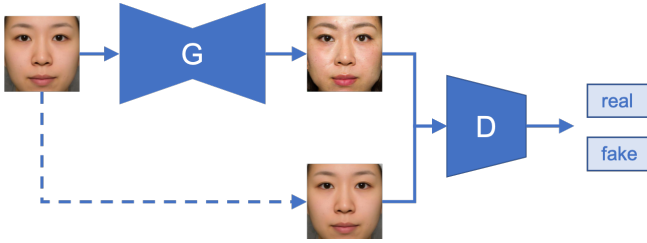


図1 pix2pix のネットワーク構造

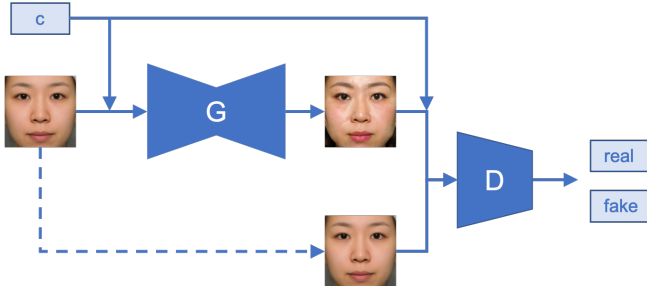


図2 Conditional GAN の Discriminator を使った pix2pix のネットワーク構造

必要だが、高品質な変換が可能となっている。今回利用する画像データセットは同一人物の化粧前後がペアになっているため、この特徴を活かせる pix2pix を用いて化粧顔画像変換を試みる。ただし、pix2pix では画像を対応ドメインの尤もらしい画像に変換するだけで Conditional Vector を与えていないため、化粧除去の場合はオリジナルの pix2pix が利用可能だが、化粧付加の場合はネットワーク構造に改良を加えなければならない。そのため、Conditional GAN [3] を参考に Generator と Discriminator に Conditional Vector を与えるように改良を施す。オリジナルの pix2pix のネットワーク構造を図1、Conditional Vector の入力を追加した pix2pix のネットワーク構造を図2にそれぞれ示す。

Conditional GAN では Generator と Discriminator に、入力する画像とともに Conditional Vector を与えていた。しかし、Discriminator への Conditional Vector の与え方には複数の方法がある。そこで、ACGAN [9] と同様に、Discriminator にそれがどの国の化粧後顔画像なのか分類させ、その答えとして Conditional Vector を与える方法も採用する。これによって、どの国の化粧が付加されたのか画像から判別できるような生成が学習され、Conditional Vector による違いがわかりやすくなると考えられる。ACGAN と同様の Discriminator を使った pix2pix のネットワーク構造を図3に示す。

3.2 StarGAN

本研究が目標としている化粧顔画像変換は、化粧後画像は国ごとにドメインを作り、化粧前画像は国共通の一つのドメインとみなすことで、複数ドメイン間の画像変換となり StarGAN [2] を適用可能である。StarGAN は、化粧前画像を各国らしい化粧後画像に変換するだけでなく、化粧後画像間の変換や化粧除去も同時に学習を行う。このような似た変換タスクを単一のネットワークで学習することで、各タスクを別々に学習するよりも

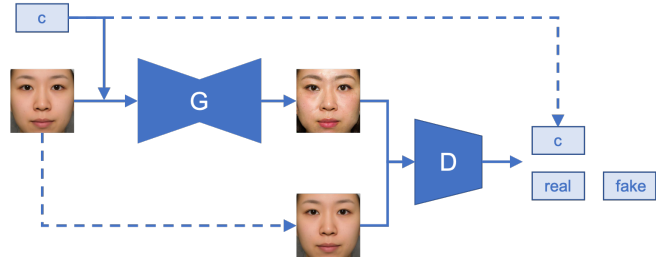


図3 ACGAN の Discriminator を使った pix2pix のネットワーク構造

高品質な変換が可能になると考えられる。ただし StarGAN では、化粧前後がペアになっているという画像データセットの良い点が無視されてしまうという欠点がある。

4 実験

4.1 実験条件

学習に使用したデータセットは、バンコク・ニューヨーク・ジャカルタ・上海・北京・東京の6都市の女性の化粧前後の顔画像がそれぞれ約1500枚ずつ、計約3000枚の画像である。ここで、上海と北京はともに中国の都市だが本実験では別の国として利用した。

4.2 実験結果

手法で説明した改良した pix2pix ベースの手法2つで化粧付加の変換を行った。Conditional GAN の Discriminator を使った場合の変換結果を図4に、ACGAN の Discriminator を使った場合の変換結果を図5にそれぞれ示す。

pix2pix による化粧除去の結果を図7に示す。

StarGAN による、化粧付加の結果を図6に、化粧除去の結果を図8にそれぞれ示す。

5 考察

5.1 pix2pix の場合

図4, 5を見ると、どちらの場合も口紅が追加され肌にツヤがでていることから、化粧を付加することはできたと考えられる。しかし、2列目から6列目までほとんど同じ画像が生成され、国ごとに異なる化粧を付加することには失敗したと言える。これは、今回利用したデータが各国の人が自分で化粧をした画像のみだったため、日本人らしい顔には日本人らしい化粧しかされておらず、Conditional Vector が無視されたからだと考えられる。例えば、ACGAN の Discriminator を採用した場合は、顔画像からそこに施されている化粧がどの国らしいか分類させているが、それはこのデータの場合顔画像がどの国の人らしいかを見れば分類可能であり化粧を見なくても正解できてしまうため、化粧が無視されたと考えられる。もしも、ある国の人に別の国らしい化粧を付加した画像がデータセット内に存在すれば、Conditional Vector にしたがって国ごとの特徴を捉えた化粧が付加できた可能性がある。

化粧前 バンコク ニューヨークジャカルタ 上海 北京 東京



図 4 Conditional GAN の Discriminator を使った pix2pix での化粧付加結果

化粧前 バンコク ニューヨークジャカルタ 上海 北京 東京

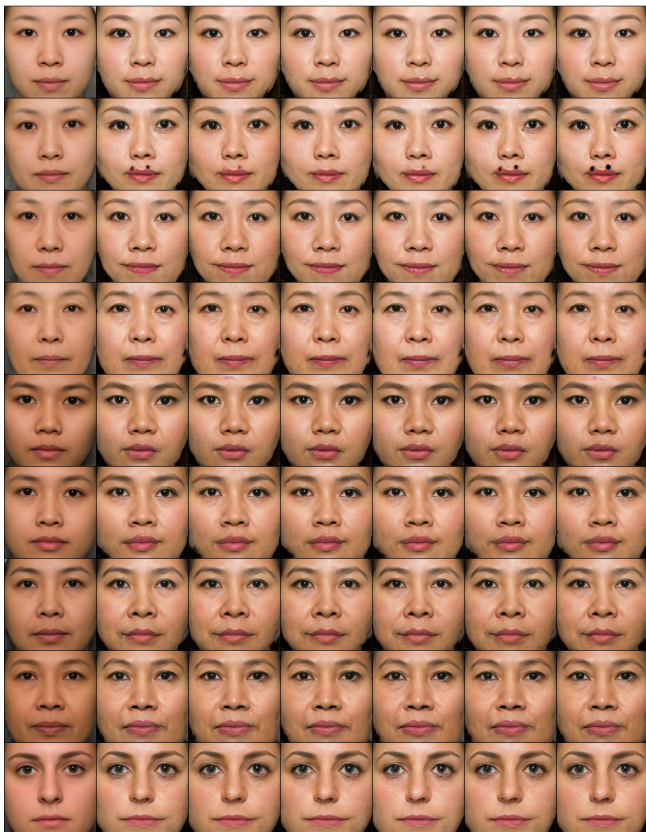


図 5 ACGAN の Discriminator を使った pix2pix での化粧付加結果

化粧前 化粧前に再構成 バンコク ニューヨーク ジャカルタ 上海 北京 東京

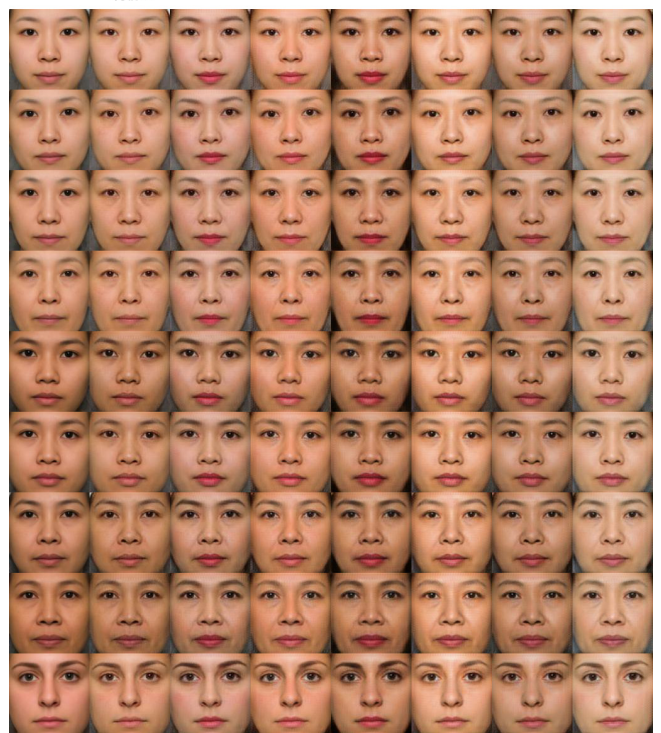


図 6 StarGAN による化粧付加結果



図7 pix2pixによる化粧除去結果



図8 StarGANによる化粧除去結果

化粧除去の変換には、CelebA-HQ データセット [10] の化粧をしている顔画像を利用した。図7の化粧除去では、口紅がなくなり、目元の化粧が消えていることが確認できる。

5.2 StarGAN の場合

図6の左から2列目は Conditional Vector に化粧前を入力し、化粧前画像をただ特徴量にしてから再構成した結果であるが、入力した画像とほとんど変化がなく正しく化粧前画像のままになっていることがわかる。それ以外の、化粧を付加した結果は唇や肌の色が変化し化粧後画像に変換できたと考えられる。

まず、バンコクの化粧は眉がわずかに濃くなり、濃い口紅が付加され、肌が白っぽく変化した。これは、バンコクでは他の国と比べて、肌を白く化粧をする人が多かったからだと考えられる。ニューヨークは口紅が付加されたが、それ以外はあまり

変化がなかった。ニューヨークは白人だけでなく色々な人種の人々が学習データセットに含まれていたため、肌の変化が小さかったと考えられる。ジャカルタは特に濃い口紅が付加され、肌の色は暗くなった。しかしこれは、ジャカルタの人が肌を暗くする化粧をする傾向にあるということではなく、元の肌が暗いからだと考えられる。上海は肌の光沢感と口紅が付加された。北京は口紅が付加されただけでほとんど変化がなかった。これは北京の人たちの化粧が薄いからだと考えられる。東京は肌が白っぽくなり、口紅が付加された。

また、化粧除去の図8では、全体的に色が薄くなったが、pix2pixの場合と比べて口紅の除去が上手く行かなかった。

6 おわりに

本研究では、国別の同一人物の化粧前後がペアになった顔画像を用いることで、指定した国らしい化粧の付加と、化粧除去の変換を行った。その結果、StarGANを使った場合には国ごとに異なる化粧を付加することに成功した。また、化粧除去も、あまり画質は良くないが口紅や目元の化粧を除去することができた。

現在の手法はあくまでも指定した国の化粧後らしく画像を変換しているだけである。そのため、今後はデータセットがペアになっているという特徴を活かし、元々の肌の色や口の色などによらない、化粧による変化を捉えた画像変換手法を提案していきたい。

文 献

- [1] P. Isola, J. Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. In *Proc. of IEEE Computer Vision and Pattern Recognition*, 2017.
- [2] Y. Choi, M. Choi, M. Kim, J. W. Ha, S. Kim, and J. Choo. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In *Proc. of IEEE Computer Vision and Pattern Recognition*, June 2018.
- [3] M. Mirza and S. Osindero. Conditional generative adversarial nets. In *arXiv preprint arXiv:1411.1784*, 2014.
- [4] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pp. 2672–2680, 2014.
- [5] J. Y. Zhu, T. Park, P. Isola, and A. A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proc. of IEEE International Conference on Computer Vision*, 2017.
- [6] H. Chang, J. Lu, F. Yu, and A. Finkelstein. Pairedcyclegan: Asymmetric style transfer for applying and removing makeup. In *Proc. of IEEE Computer Vision and Pattern Recognition*, 2018.
- [7] Y. Li, L. Song, X. Wu, R. He, and T. Tan. Anti-makeup: Learning a bi-level adversarial network for makeup-invariant face verification. In *Proc. of AAAI Conference on Artificial Intelligence*, 2018.
- [8] T. Li, R. Qian, C. Dong, S. Liu, Q. Yan, W. Zhu, and L. Lin. Beautygan: Instance-level facial makeup transfer with deep generative adversarial network. In *Proc. of ACM International Conference Multimedia*, pp. 645–653, 2018.
- [9] A. Odena, C. Olah, and J. Shlens. Conditional image synthe-

sis with auxiliary classifier gans. In *Proc. of International Conference on Machine Learning*, pp. 2642–2651, 2017.

- [10] T Karras, T Aila, S Laine, and J Lehtinen. Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196*, 2017.