Identity と化粧 Style の分離による顔画像変換

五味 京祐[1,a)]　越野 誠也[2]　柳井 啓司[1]

概要

近年，Deep Learning の発展により画像の生成や変換の技術が盛んに研究され，目覚ましい進歩を見せている．本研究では，そうした画像変換の技術を化粧に応用し，化粧顔画像変換と化粧特徴量抽出が可能な手法を提案する．化粧顔画像変換とは，入力した顔画像に化粧を付加したり除去したりする画像変換のことである．化粧特徴量抽出では顔画像を，顔自体を表す特徴量の Identity と化粧を表す特徴量の Style に分離するアプローチを採用した．ここで，Identity は顔全体から抽出するのに対して，Style はマスク画像で切り出した肌・眉・目・口からそれぞれ抽出することで， パーツごとの化粧特徴量の抽出を実現した．

1. はじめに

近年，Deep Learning の発展により画像の生成や変換の技術が盛んに研究されている．有名な画像変換の研究として Gatys らのスタイル変換の研究 [2] がある．これは，コンテンツ画像とスタイル画像の 2 枚を入力として，コンテンツ画像の内容を保持しながらスタイル画像に近い画風の画像に変換する研究である．他にも DRIT [5] では，一般のドメイン間画像変換を Disentanglement のアプローチで，画像を 2 つの特徴量に分離することで実現した．また，化粧顔画像変換に焦点を当てた研究もいくつかあり，PairedCycleGAN [1] では，CycleGAN [11] をベースに，ペアになった画像を利用することで化粧の付加・除去の変換を学習した．

本研究では，そういった画像変換技術を化粧に応用することで，化粧顔画像変換と化粧特徴量の抽出を行う．化粧顔画像変換とは図 1 のように，化粧前画像と化粧後画像の 2 枚を入力として，化粧前画像の人物に化粧後画像と同じ化粧を付加した画像を生成することである．本研究はさらに，変換するパーツを指定可能にした．また，化粧特徴量の倍率を変更することで，付加する化粧の濃さを調整可能である．このような化粧顔画像変換の活用例として，化粧品を購入する際，モデルがしている化粧を自分がしたらどうなるのかカメラで撮影して確認することや，化粧の除去によって顔認証技術において問題となっている化粧による誤認識を抑制することなどが挙げられる．また，化粧特徴量を抽出することによって，化粧の定量的評価や分析，似た化粧した顔画像の検索が可能となる．

1 電気通信大学 大学院情報理工学研究科 情報学専攻
2 資生堂
a) gomi-k@mm.inf.uec.ac.jp

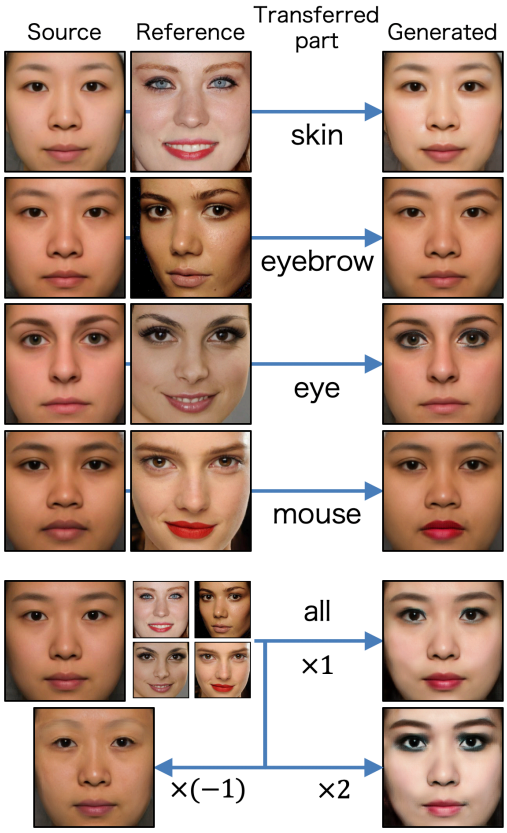

図 1 提案手法での生成結果．上の 4 行では，指定したパーツの化粧のみを転移している．残りの 2 行は 4 つのパーツに 4 つの顔から同時に化粧を転移した結果である．$1, 2, -1$ は，転移する化粧の濃さの倍率を示す．

2. 関連研究

画像変換において画像を複数の特徴量に分離する研究と，本研究と同じ化粧顔画像変換の問題に取り組んだ研究として，以下のようなものがある．

2.1 画像生成・変換

Liu らの研究 [9] では，顔を Identity とそれ以外の Attribute に分離することで顔画像変換を行っている．Identity を保持したまま Attribute のみを変更することで，表情や髪型，肌の色などを変化させることができる．しかし，この研究は化粧の変換を対象としていないという点で本研究と異なる．DRIT [5] は，画像ドメイン間の変換を行う際に，画像をドメイン共通の Content 特徴とドメイン独立の Attribute ベクトルに分離することで，多様な変換結果

を生成可能な手法である．特徴量に分離するというコンセプトは本研究と同様であるが，顔画像を対象とはしていない．本研究で提案する手法は，この DRIT をベースにしている．

2.2 化粧画像変換

PairedCycleGAN [1] では，本研究と同様に化粧の顔画像変換を行っている．CycleGAN [11] 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．また，化粧を付加する関数 G は化粧をしていない画像としている画像を 1 枚ずつ入力するのに対して，化粧を除去する関数 F は化粧をしている画像 1 枚だけを入力とする非対称な変換がこの手法の特徴となっている．しかし，本研究と異なり，Identity や化粧 Style の特徴量を抽出せずに画像の変換のみを行っている．Li らの研究 [8] 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．行ったのは前処理としての化粧除去だけであり，化粧付加は行っていない．BeautyGAN [6] 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．この手法では，ロスを定義にマスク画像を上手く利用していて，効率的に化粧の変換を学習した．本研究もこの手法を参考にし，マスク画像でパーツを切り出す手法を利用する．しかし，PairedCycleGAN と同様に特徴量を抽出せずに画像の変換のみを行っている．また，BeautyGAN では眉は対象としていないが，本研究では眉の化粧も対象としている．

3. 手法

3.1 概要

本研究では，顔画像を化粧によらない顔自体の特徴量 Identity と，顔画像に施されている化粧を表す特徴量の Style に分離する．提案するネットワークは図 2 のように 5 つの Encoder と 1 つの Generator で構成されている．Encoder E^I は顔全体を入力として Identity を抽出し，Encoder $E^S_{skin}, E^S_{brow}, E^S_{eye}, E^S_{mouse}$ はそれぞれ対応する肌・眉・目・口のマスク画像で切り出した各パーツを入力として Style を抽出する．そして，Generator の G は Identity と各パーツの Style を入力として，顔画像を生成する．

このように本研究ではパーツごとのマスク画像が必要である．以下の手順でマスク画像を作成した．

(1) 顔画像からランドマークを抽出する．
(2) ランドマークが入るように，余白を空けて画像を切り取る．
(3) ランドマークの位置が合うように，化粧前後ペアの位置を合わせる．
(4) セグメンテーションを用いてパーツを分離する．
(5) 髪の毛や背景が入らないように，顔画像とマスク画像を切り取る．
(6) 化粧前後のマスク画像を使い，パーツごとのマスク画像を作成する．

ここで，目のマスクはアイシャドウなどの目元の化粧領域としたいが目元の化粧領域のみを抽出できる手法が一般には存在しないため，目元の化粧を含むと思われる大きさの矩形領域を目のマスクとし，既存のセグメンテーション手法で得られる眼球部分は化粧とは関係ないため取り除いた．セグメンテーションには Li らの顔補完の研究 [7] で学習され公開されていた学習済みモデルを利用した．

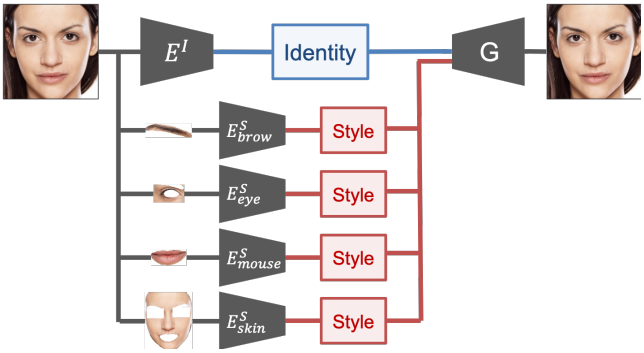

図 2 提案するネットワーク

3.2 ロス関数

DRIT [5] で使われていたロスのうち，$L^{content}_{adv}, L^{domain}_{adv}, L^{recon}_1, L^{latent}_1, L_{KL}$ の 5 つはそのまま利用した．ただし，$L^{content}_{adv}$ は $L^{Identity}_{adv}$ と名前を変更した．そして，$L^{Identity}, L^{Style}, L^{parts}_1, L^{parts}_{adv}, L^{random}_{adv}$ の 5 つのロスを追加した．

3.2.1 Identity のロス関数

本研究では同一人物の化粧前後がペアになった画像データセットを用いる．そこで，同一人物の化粧前と化粧後から同一の Identity が抽出されるように制約を加える．図 3 のように化粧前後の Identity を入れ替えて Generator に入力したとき，Identity が同じなら元の画像が生成されるはずであるということから，入力画像と出力画像の差を Identity のロス関数とする．

式で表すと以下のようになる．

$$\begin{aligned} L^{Identity}(G, E^I, E^S_{skin}, E^S_{brow}, E^S_{eye}, E^S_{mouse}) = \\ \mathbb{E}_{x,y}[\|G(E^I(y), E^S_{skin}(x), E^S_{brow}(x), E^S_{eye}(x), E^S_{mouse}(x)) - x\|_1 \\ +\|G(E^I(x), E^S_{skin}(y), E^S_{brow}(y), E^S_{eye}(y), E^S_{mouse}(y)) - y\|_1] \end{aligned} \quad (1)$$

3.2.2 Style のロス関数

Style は絶対値が小さいほど薄い化粧で，大きいほど厚い化粧を表す特徴量であり，化粧前画像の Style は 0 であることが望ましい．そのため，図 4 のように，化粧前画像でも化粧後画像でも Style を零ベクトルとして変換した場合，化粧前画像が生成されるという制約を入れる．

式で表すと以下のようになる．

$$\begin{aligned} L^{Style}(G, E^I) = \mathbb{E}_{x,y}[\|G(E^I(x), \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}) - x\|_1 \\ +\|G(E^I(y), \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}) - x\|_1] \end{aligned} \quad (2)$$

3.2.3 パーツのロス関数

各パーツの Style が対応する部分の変換にのみ影響を与え，それ以外の部分が変化しないように制約をするために新しいロス関数 L^{parts}_1 を提案する．そのための仮定は，例えば図 5 のように口の Style のみ化粧後に入れ替えると，Generator は口だけが化粧後になった画像を生成するということである．そこで，提案するロスでは一つのパーツの

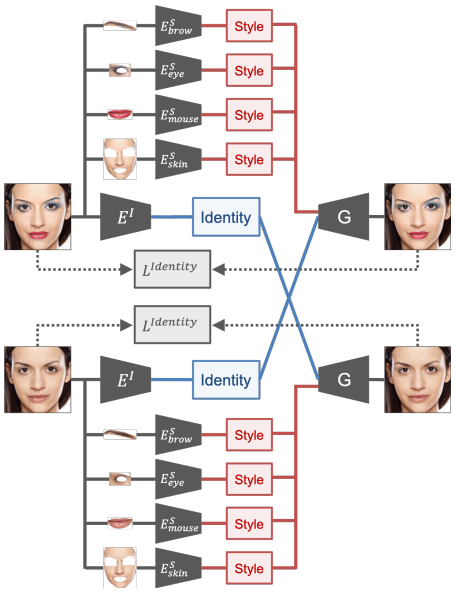

図 3　Identity のロス関数

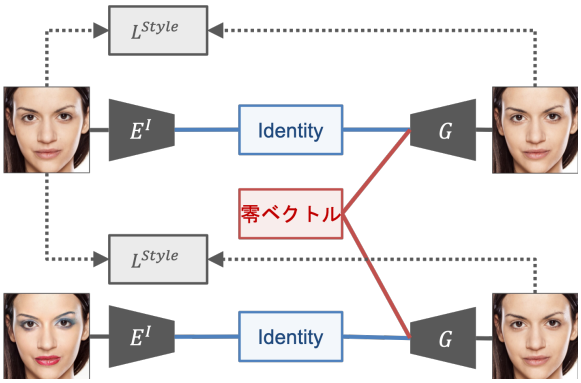

図 4　Style のロス関数

み化粧前後で Style を入れ替え変換し，あらかじめ作成しておいた化粧前後でパーツを入れ替えた画像と同じ画像が生成されるように制約をする．

このように，一つのパーツのみ Style を入れ替えて画像を生成したとき，その入れ替えたパーツがリアルでかつ正しく化粧変換されることを促すための敵対的ロス L_{adv}^{parts} も導入する．そのために，以下に示す 8 つの Discriminator を定義する．

$$D_{skin}^{X}, D_{brow}^{X}, D_{eye}^{X}, D_{mouse}^{X}, D_{skin}^{Y}, D_{brow}^{Y}, D_{eye}^{Y}, D_{mouse}^{Y}$$

これらは，D_{skin}^{X} なら肌がリアルな化粧前画像か判別，D_{brow}^{Y} なら眉がリアルな化粧後画像か判別，といったように入力された画像から対応するパーツをマスク画像でクロップし，そのパーツがリアルな化粧前もしくは化粧後であるか判別する．

3.2.4　Random adversarial ロス

DRIT [5] では，ランダムな Attribute を入力して画像を生成したときにも L_{adv}^{domain} を適用していた．それは，DRIT には Generator が 2 つあり，Attribute をランダムにしても Generator によって変換画像のドメインが定まるからである．一方，本研究の手法ではランダムな Style を入力したときに変換画像のドメインが定まらない．そこで，本研究ではドメインは関係なくランダムな Style を入力したとき，リアルな顔画像が生成されることを促すロス L_{adv}^{random} を適用する．

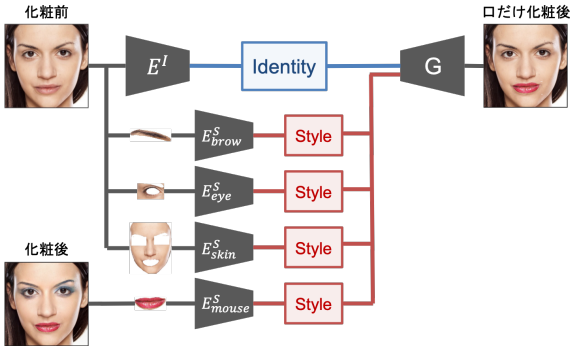

図 5　口の Style のみ化粧後を入力した場合

4.　実験

実験に使用したのは，資生堂が収集したバンコク・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 1451 枚ずつ，計 2902 枚の画像で構成されているデータセットである．テスト用に化粧前後それぞれ 100 枚ずつを残し，2702 枚の画像で学習を行った．

実験では従来手法との比較と化粧顔画像変換，Style による顔画像検索を行った．従来手法との比較では残しておいた 200 枚の顔画像を利用し，Inception Score [10] と Frećhet Inception Distance [3] で評価した．その他では，化粧前テスト画像として学習に使ったデータセットの各都市ごとの平均顔，化粧後テスト画像には CelebA-HQ データセット [4] の化粧をしている顔画像を利用した．画像の解像度は 256 × 256 である．

4.1　実験結果

表 1 に，Inception Score(IS) と Frećhet Inception Distance(FID) での BeautyGAN [6] との比較を示す．FID では提案手法の方が良い値だったが，IS では BeautyGAN の方が良い結果だった．このことから，提案手法は BeautyGAN と遜色ない品質で画像が生成できていると考えられる．

図 6, 7 に，化粧顔画像変換を行った結果を示す．「すべて 0」の列を見ると化粧が除去されていて，Style のロスのとおりに化粧前画像の Style を 0 にすることができたと考えられる．また，Style を入れ替えても変わるのは化粧のみで，人物は変わっていないことから Identity と Style の分離に成功したと考えられる．そして，図 6 に注目すると，化粧後画像は口の化粧が濃く，「口だけ交換」の列で口の色が大きく変化していることがわかる．同様に，図 7 は目の化粧が濃いため，「目だけ交換」の列で目元が大きく変化している．これらの結果から，パーツごとの化粧顔画像変換ができることが確認できた．

図 8 に，化粧の濃さの倍率を変更した結果を示す．抽出された Style を化粧特徴量空間で何倍かすることで，その化粧を濃くしたり，薄くしたりすることが可能であることがわかる．ここで，0 倍は化粧をしていない画像になり，$-1, -2$ 倍は眉毛が 0 倍よりも薄くなっていることから，化粧を濃くする操作の逆操作が実現できていると考えられる．

図 9 に，2 つの化粧後画像の中間の Style を転移した結果を示す．2 つの Style の中間を補間することで，2 つの Style を混ぜることができていると思われる．

図 10 に，目の Style での顔画像検索結果を示す．左上の画像はクエリ画像であり，左側の 4 つはクエリ画像と近い Style を持つ画像，右側の 4 つはクエリ画像と遠い Style を持つ画像となっている．今回，クエリ画像として目の化粧が濃い画像を与えたため，左側の近い画像として似たような目元の化粧が濃い画像，右側の遠い画像として目元の化粧がほとんど施されていない画像が得られた．このことから，目の Style は目元の化粧を表す特徴量となっていると考えられる．

表 1 Inception Score(IS) と Fréchet Inception Distance(FID) による比較

手法	IS	FID
BeautyGAN [6]	1.488	22.50
提案手法	1.346	19.14

図 6 化粧顔画像変換結果 1

図 7 化粧顔画像変換結果 2

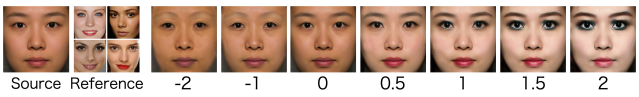

図 8 化粧の濃さの倍率を変更した結果

図 9 2 つの化粧後画像の中間の Style を補間して転移した結果

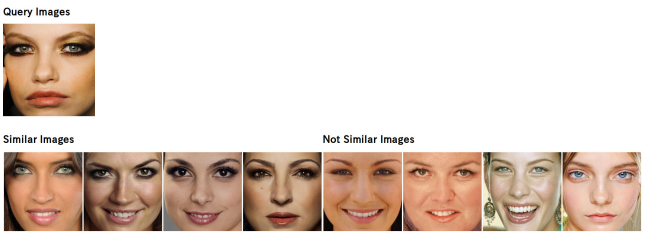

図 10 目の Style による画像検索結果

5. おわりに

Deep Learning を用いた化粧顔画像変換に取り組み，同時に化粧特徴量の抽出も行った．本研究では，顔画像を顔特徴量の Identity とパーツごとの化粧特徴量の Style に分離するアプローチを採用し，実験から顔画像を Identity と Style に分離できたことを確認した．また同時に，従来手法では不可能であったパーツごとの化粧転送を実現することができた．

実験では，マスク画像と同じ矩形状の模様が目元に生成されてしまう場合があった．これは，学習データセット内に化粧前後で肌の色が大きく変わっている顔画像があるからだと考えられる．今後の課題として，この矩形模様を除去することや，既存手法との比較が挙げられる．

参考文献

[1] Chang, H., Lu, J., Yu, F. and Finkelstein, A.: PairedCycleGAN: Asymmetric style transfer for applying and removing makeup, *Proc.of IEEE Computer Vision and Pattern Recognition* (2018).

[2] Gatys, L. A., Ecker, A. S. and Bethge, M.: Image Style Transfer Using Convolutional Neural Networks, *Proc.of IEEE Computer Vision and Pattern Recognition* (2016).

[3] Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B. and Hochreiter, S.: Gans trained by a two time-scale update rule converge to a local nash equilibrium, *Proc.of Advances in Neural Information Processing Systems* (2017).

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

[5] Lee, H. Y., Tseng, H. Y., Huang, J. B., Singh, M. and Yang, M. H.: Diverse image-to-image translation via disentangled representations, *Proc.of European Conference on Computer Vision*, Vol. 1, No. 3, p. 5 (2018).

[6] Li, T., Qian, R., Dong, C., Liu, S., Yan, Q., Zhu, W. and Lin, L.: BeautyGAN: Instance-level Facial Makeup Transfer with Deep Generative Adversarial Network, *Proc.of ACM International Conference Multimedia*, pp. 645–653 (2018).

[7] Li, Y., Liu, S., Yang, J. and Yang, M. H.: Generative Face Completion, *Proc.of IEEE Computer Vision and Pattern Recognition* (2017).

[8] Li, Y., Song, L., Wu, X., He, R. and Tan, T.: Anti-Makeup: Learning a bi-level adversarial network for makeup-invariant face verification, *Proc.of AAAI Conference on Artificial Intelligence* (2018).

[9] Liu, Y., Wei, F., Shao, J., Sheng, L., Yan, J. and Wang, X.: Exploring Disentangled Feature Representation Beyond Face Identification, *Proc.of IEEE Computer Vision and Pattern Recognition* (2018).

[10] Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A. and Chen, X.: Improved techniques for training gans, *Proc.of Advances in Neural Information Processing Systems* (2016).

[11] Zhu, J. Y., Park, T., Isola, P. and Efros, A. A.: Unpaired image-to-image translation using cycle-consistent adversarial networks, *Proc.of IEEE International Conference on Computer Vision* (2017).