

領域分離型敵対的生成ネットワークによる髪型編集手法

夏目 亮太[†] 谷田川 達也[†] 森島 繁生[‡]

早稲田大学[†] 早稲田大学理工学術院総合研究所 / JST ACCEL[‡]

1. はじめに

髪型は旧来より様々な文化的背景において人物の個性や階級を表す重要な要素である。また髪型はファッションの一部として人の個性を表す手段の一種となっている。そのため、好みの髪型を見つけるためのコンピュータ・システムがグラフィックス分野で広く研究されてきた。Shlizerman [1]は、「パーマ」「金髪」などの髪型に関する外見的特徴を有する画像群の中から、さらに姿勢や肌のトーンなど外見特徴が類似している画像を検索し、顔領域を合成する手法を提案した。しかし、[1]を含む従来法は、他人の髪型の試着や、髪の色や長さの変更などの髪型の発見に必要な個別の要素を対象としており、複数の要素を統合的に扱うシステムは実現されていない。

本稿では、近年、自然な画像生成が可能となった敵対的生成ネットワーク(GAN)に基づく髪型編集法を提案する。通常、GANによる画像の生成では、髪型を含む画像全体の自然さを大量のデータから学習する。一方で、この手法は顔領域と髪領域の間に存在する相関関係を切り離すことが難しく、各々の領域を独立に編集できない。提案法では、顔領域、髪領域、それらを合成した全体画像を統合的に学習することで、顔、髪各領域から個別に特徴を抽出する。これにより、他人の髪型の試着や髪色、長さの編集などの多様な編集を実現する。

2. 領域分離型敵対的生成ネットワーク

本研究では、髪型編集のための新しい深層学習モデルである領域分離型敵対的生成ネットワーク(Region-Separative Generative Adversarial Network, RSGAN)を提案する。図1に示すとおりRSGANのネットワークは分割ネットワークと合成ネットワークの二つから構成される。顔と髪各領域を処理する分割ネットワークは、

全体画像を低次元の特徴ベクトルである潜在変数に写像するEncoderと潜在変数から顔/髪領域を復元するDecoderの組から構成される。

各分割ネットワークは顔と髪領域のみを表現できるように潜在変数を構成するため、これにより領域個別の特徴を抽出できる。一方で、合成ネットワークは分割時に得られた潜在変数 z_f , z_h を連結した z から全体画像 x' を復元する。

また、分割ネットワークと合成ネットワークには、画像や潜在変数に加え、画像中に「笑顔」や「金髪」などの特徴が存在するか否かを示す属性情報も与える。この結果、任意の画像に対し、学習済みの特徴を与奪した画像を合成できる。しかしながら、本研究で用いるCelebA[1]のデータセットには、「笑顔」のような顔のみに現れる特徴や、「金髪」のような髪にのみ現れる特徴だけでなく、また「性別」や「年齢」などの顔と髪の見られる特徴が存在し、これらの特徴が各領域にどの程度由来するかを見極めることは難しい。本研究では属性情報も全結合層 E_c に通すことで、由来の曖昧さを含む特徴を緩やかに分割している。従来のGANと同様、合成ネットワークと識別ネットワークを敵対的に学習させることで、合成ネットワークは学習データ中にある顔画像に近い自然な出力画像が得られる。

RSGANを学習する際には、全体画像 x 、髪領域画像 x_h 、顔領域画像 x_f が必要であり、これらの画像を大規模顔画像データベースのCelebA[1]から生成した。図2(a)がCelebAに含まれる顔画像である。この顔画像に対して、既存の機械学習ライブラリであるDlib [2]を用いて68個の顔特徴点を検出し、その中の図2(b)で青い点で示された41個の特徴点を顔領域と定義する。それらの特徴点の凸包を求め、縦方向に1.4倍、横方向に1.3倍した領域をマスクとする(図2(c))。さらに、図2(d), (e)に示すように、各領域を正方形に切り取る。顔領域画像は、左上の座標を(30, 70)として118×118の正方形に切り取り、髪領域画像は左上の座標を(0, 20)として178×178の正方形に切り取り、それぞれを64×64にリサイズし学習した。

“Neural Hairstyle Editing using Region Separative Generative Adversarial Network”

[†] Ryota NATSUME, [†] Tatsuya YATAGAWA, [‡] Shigeo MORISHIMA

[†] Waseda University

[‡] Waseda Research Institute for Science and Engineering / JST ACCEL

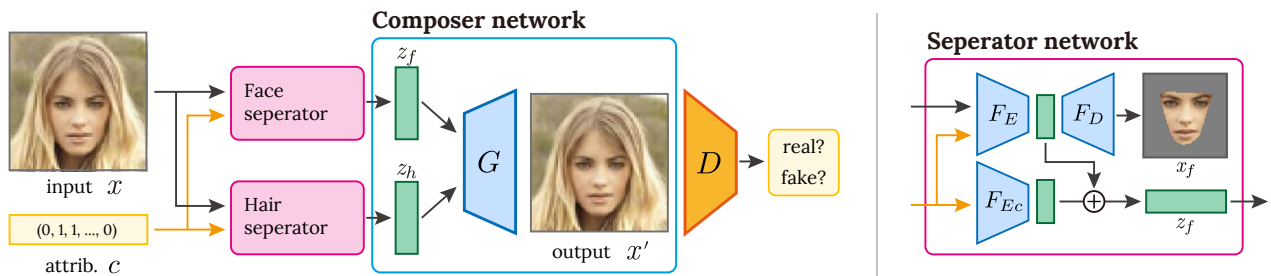


図 1. RSGAN のネットワーク

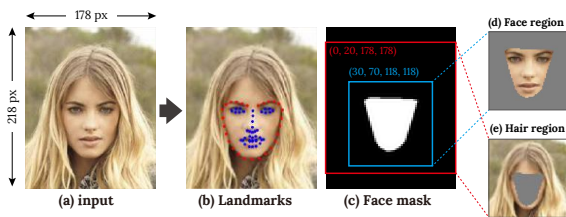


図 2. データセット作成の手順

のみに与えて生成したものを比較した。比較結果を図 4 に示す。結果が示すように顔のみに対して、前髪や茶髪などの髪に関する属性情報を与えても、画像が入力画像から変化していない。つまり、提案ネットワークによって、髪属性情報と顔の属性情報を自動で分離できていることがわかる。

3. 結果

3.1 髪型の試着

髪型の試着は二枚の顔画像の間で顔領域を入れ替えることで実現できる。提案ネットワークは顔と髪の各領域から個別に潜在変数を抽出しているため、二枚の画像間で顔の潜在変数を入れ替えることで、髪型の試着が実現できる。図 3 に髪型の試着結果を示す。結果が示すように、提案手法では、顔向きが異なる画像間でも髪型の試着が可能である。



図 3. 髪型試着結果

3.2 髪色の色や長さの編集

提案したネットワークは、ラベルを全結合層に通すことで、顔と髪の情報が混ざったラベルから、顔の属性情報と髪の属性情報に分割する。実際に属性情報を分割できているかを検証するために、あるラベルを、顔と髪のどちらにも与えて画像を生成したもの、顔のみに与えて生成したもの、髪

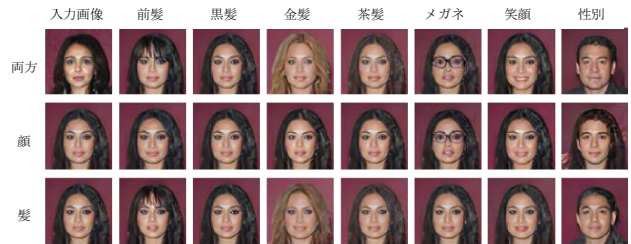


図 4. 髪色の色や長さの編集

4. 今後の展望

本稿で提案した RSGAN は顔と髪の領域を対象としているが、提案法の考え方は領域が三つ以上に増えた場合にも拡張可能である。具体的には、本手法では顔画像を顔領域と髪領域に分割しているが、画像中の背景領域も分割することが考えられる。このような多領域への分割は、衣服画像を顔、上半身、下半身と分割することで仮装試着にも応用できる可能性がある。

謝辞

本研究の一部は、JST ACCEL「次世代メディアコンテンツ生態系の基盤構築と応用展開」(JPMJAC1602)及び、日本学術振興会科研費(15K16027)及び、日本学術振興会特別研究員奨励費(16J02280)の助成を受けたものである。

参考文献

- [1] Z. Liu et al. "Deep learning face attributes in the wild." *In Proc. of the IEEE Int'l Conf. on Computer Vision*. 2015.
- [2] D.E. King. "Dlib-ml: A machine learning toolkit." *Journal of Machine Learning Research*, Vol. 10, pp. 1755-1758, 2009.
- [3] I. Kemelmacher-Shlizerman. "Transfiguring portraits." *ACM Trans. Graph.*, Vol 35, No. 4, 2016.