5P-04

敵対的生成ネットワークを用いた画像中の人体の教師なし姿勢変換

岡林 遥平 †　　高橋 隆史 ‡

† 龍谷大学大学院理工学研究科　　‡ 龍谷大学理工学部

1　はじめに

人物の画像と目標とする姿勢情報から，その人物に目標姿勢をとらせた画像を生成するニューラルネットワークを考える．この問題へのアプローチとしては一般に，入力画像と出力の正解とのペアを用いる教師あり学習が考えられる．しかし，ペアのデータを大量に収集するのは困難である．そこで本研究では，ペアのデータを必要としない，教師なしの姿勢変換手法を提案する（図 1）．

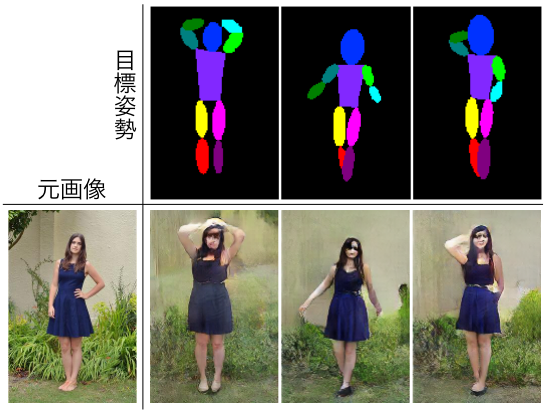

図 1　提案手法による姿勢変換画像．

2　先行研究

敵対的生成ネットワーク（GAN）を利用した画像変換手法に pix2pix[1]，pix2pixHD[2] などがある．これらは GAN を利用することで，精度の高い画像変換が可能である．さらに，pix2pixHD では Instance-Wise Average Pooling (IWAP) と呼ばれる手法によって，生成画像をインタラクティブに編集することもできる．本研究では，このアイデアを基とした姿勢変換手法を提案する．

教師なし姿勢変換の先行研究に，Pumarola らの研究 [3] がある．これは CycleGAN[4] を基にした手法であり，姿勢情報を利用して変換を行う．本研究の手法は，これとは異なる手法に基づいて，背景や頭部などのパーツごとの符号も利用して姿勢変換を行う．

3　手法

提案手法の推論時の流れを図 3 に示す．姿勢変換したい人物画像 $\boldsymbol{x}_A \in \mathbb{R}^{H\times W\times 3}$ と対応する姿勢マップ $\boldsymbol{y}_A \in \mathbb{R}^{H\times W\times 11}$ を Encoder E に入力し，特徴マップ $\boldsymbol{f}_A \in \mathbb{R}^{H\times W\times D}$ を得る．ここで，H, W はそれぞれ人物画像の高さと幅を表す．姿勢マップは，背景などのパーツに対応した 11 カテゴリからなるセグメンテーションマップの各画素値を one-hot 表現した 2 値マップである．D は各パーツの符号の次元数である．

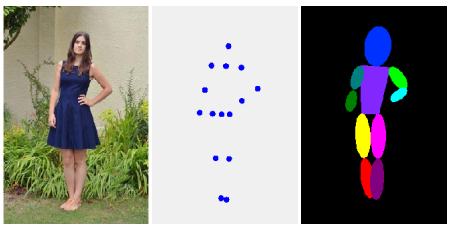

図 2　姿勢マップの作成方法．左から元画像，OpenPose で得た keypoints，そこから作成した姿勢マップ．

次に，$\boldsymbol{f}_A$ に新たな姿勢マップ $\boldsymbol{y}_B$ を用いた IWAP を適用し，姿勢 $\boldsymbol{y}_B$ に対応した符号マップ $\boldsymbol{z}_B$ を作成する．まず，$\boldsymbol{y}_A$ の頭部のパーツに対応する領域の $\boldsymbol{f}_A$ の画素値をチャンネル毎に平均し，D 次元の頭部の符号ベクトルを求める．残りのパーツも同様にそれぞれの符号を求める．それらの符号を，符号マップ $\boldsymbol{z}_B$ の $\boldsymbol{y}_B$ に対応した各パーツ位置に代入する．$\boldsymbol{z}_B$ の大きさは $\boldsymbol{f}_A$ と等しい．

最後に符号マップ $\boldsymbol{z}_B$ と $\boldsymbol{y}_B$ を Generator G に入力し，出力 $\widehat{\boldsymbol{x}}_B$ を求める．$\widehat{\boldsymbol{x}}_B$ は $\boldsymbol{x}_A$ を姿勢 $\boldsymbol{y}_B$ に変換した画像となる．

学習時は，$\boldsymbol{y}_b = \boldsymbol{y}_A$ とし，Encoder と Generator が元画像 $\boldsymbol{x}_A$ を再構成するように学習する．ロスは pix2pixHD と同様で，GAN のロスと Feature Matching Loss，Perceptual Loss を組み合わせたものを使用する．

本研究では，上述のもの以外のネットワーク構成や新たなロスの導入も検討したが，詳細の説明は省略する．

Unsupervised human pose transfer by Generative Adversarial Networks

† Graduate School of Science and Technology, Ryukoku University

‡ Faculty of Science and Technology, Ryukoku University

Copyright ©2020 Information Processing Society of Japan.
All Rights Reserved.

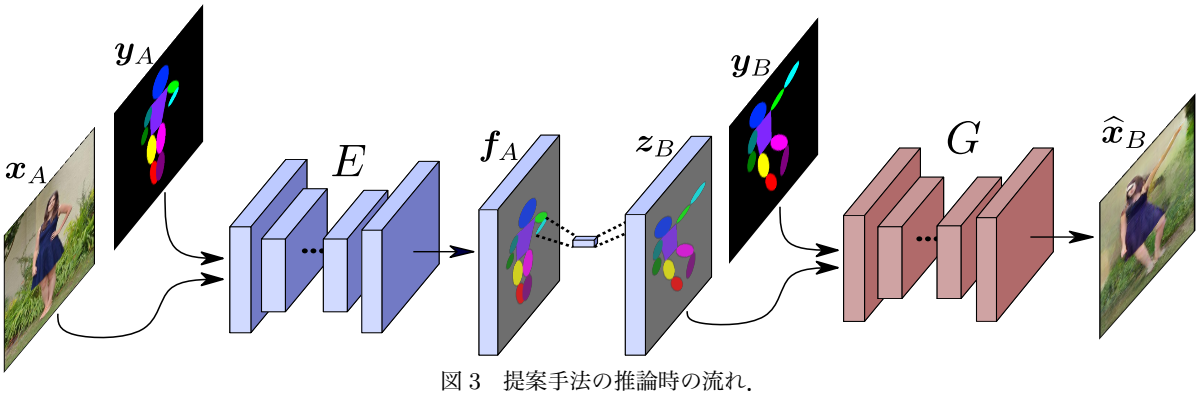

図 3　提案手法の推論時の流れ．

表 1　姿勢変換画像の定量的評価．SSIM の値を示す．

手法 1	手法 2	手法 3
0.7024	0.7416	0.6495

4　データ

学習と定性的評価には Fashion550k[5] を，定量的評価には iPER dataset[6] を使用した．以下ではデータの前処理の方法と姿勢マップの作成の方法について述べる．

まず，全ての画像について，OpenPose[7] を用いて 15 個の keypoints とその確信度を推定し，隠れがある等の不適切なデータを除外した．学習データは約 20 万枚である．定量的評価用データは，iPER dataset から同一人物の異なる姿勢の画像約 1 万ペアを選択した．

次に，keypoints を用いて，背景・頭部・胴体など 11 カテゴリからなるセグメンテーションマップを作成した (図 2)．このマップの各画素値を one-hot 表現し，姿勢マップを作成した．入力画像，姿勢マップともに 192×128 画素にリサイズした．

5　実験

3 章で述べた手法 (手法 1 と呼ぶ) を検証した．手法 1 に加え，Encoder と Generator を GAN のロスを用いず再構成ロスのみで学習する手法 2，および手法 2 の生成画像を追加のネットワークに入力し鮮鋭化する手法 3 についても検証した．

手法 2 では，再構成ロスとして，SSIM と Perceptual Loss を組み合わせたものを採用した．手法 3 の追加のネットワークは，手法 1 と同様のロスで学習した．3 つの手法全てにおいて，epoch 数を 50，バッチサイズを 15，符号の大きさ D を 8 として実験を行った．

手法 1 の姿勢変換画像の一例を図 1 に示す．姿勢はある程度変換できているが，顔や手先などの詳細部分がうまく生成できていない．表 1 に姿勢変換画像の定量的評価を行った結果を示す．SSIM では手法 2 が最も良い．しかし，GAN のロスを用いないため，大まかに人体や服装，背景などは生成できたが，ぼやけた画像になった．手法 3 では，定量的には手法 1 より悪くなったが，手法 2 の結果を補って人体と服をある程度鮮鋭化できた．

6　おわりに

本研究では，符号化を利用した教師なし姿勢変換手法を検討した．画質は不十分ながら，ある程度の姿勢変換が可能であることを確認した．今後の課題としては，(1) 姿勢情報の表し方の改良，(2) パーツごとの符号表現の改良，(3) Pumarola ら [3] のように姿勢変換画像に対するロスを組み込む，などが考えられる．

参考文献

[1] P. Isola, et al. Image-to-image translation with conditional adversarial networks. In *CVPR*, 2017.
[2] T.C. Wang, et al. High-resolution image synthesis and semantic manipulation with conditional gans. In *CVPR*, 2018.
[3] A. Pumarola, et al. Unsupervised Person Image Synthesis in Arbitrary Poses. In *CVPR*, 2018.
[4] J. Zhu, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *ICCV*, 2017.
[5] N. Inoue, et al. Multi-Label Fashion Image Classification with Minimal Human Supervision. In *ICCVW*, 2017.
[6] W. Liu, et al. Liquid warping gan: A unified framework for human motion imitation, appearance transfer and novel view synthesis. In *ICCV*, 2019.
[7] Z. Cao, et al. Realtime multi-person 2d pose estimation using part affinity fields. In *CVPR*, 2017.

Copyright ©2020 Information Processing Society of Japan.
All Rights Reserved.