Vol.2019-CG-176 No.1
Vol.2019-DCC-23 No.1
Vol.2019-CVIM-219 No.1
2019/11/7

ポーズ画像の生成における非対象物除去の効果

須貝 聡子[1,a)]　小暮 潔[1,b)]

概要：本報告では，ポーズ画像生成において，非対象物の色が対象物の生成画像に移るなどの劣化を改善するために，非対象物を除去してポーズ画像を生成する手法を提案し，その効果を評価する．ここでの非対象物とは，ポーズ画像生成でモデル化されない部分を示し，人物画像と重ならない部分（背景）とカバンのような人物画像と重なる部分（前景）に分けられる．提案手法は具体的に，非対象物が含まれている人物画像から，指定した非対象物を取り除き，ポーズ画像を生成する．生成したポーズ画像は最終的に背景と合成する．ポーズ画像生成では deformable GAN を用いる．評価では，背景と合成する前の，非対象物を取り除く前と取り除いた後の画像で生成したポーズ画像を比較した．また，SSD による人物検出のスコアによる評価も行った．その結果，非対象物を取り除いた場合のほうが色移りなどの劣化が少ないことと，SSD のスコアが高いことが確認された．

1. はじめに

リアルな画像を生成することは，顔画像の加工や動画の作成，合成画像に基づく画像検索など多くのアプリケーションにとって価値がある [1]．その中でも，人物のポーズ画像生成は，人混みの中から特定の人物を見つける手法 [2] や，人の動きを必要とする動画の作成 [3]，人物の珍しいポーズの推定 [4] などの手法の学習時に有用な訓練データを提供する．特に，人混みの中から特定の人物を見つける手法やポーズ推定などの研究では多くの訓練データが必要であり，データバランスを整えるためにも少ないデータを水増しする必要がある．そこで，本研究では人のポーズ画像生成に着目する．

ポーズ画像生成とは，生成したい人物画像とポーズ情報から指定されたポーズの人物画像を生成することを示す．これにより，同じ服装で様々なポーズの画像を生成することができる．代表的なのは Ma らの手法 [1] や Siarohin らの手法 [5] である．

ポーズ画像生成では，背景や小物などのモデル化されない部分（非対象物）が生成画像に影響する場合がある．具体的には，非対象物の色が生成画像に色移りしてしまうことや，生成画像の輪郭が不鮮明になってしまうことなどがある．

本研究では，このような問題を改善する手法を提案する．具体的には，非対象物が含まれている人物画像から，指定した非対象物をそのマスク画像をもとに取り除き，ポーズ画像を生成する．ポーズ画像を生成した後は，生成したポーズ画像を背景と合成させる．評価では，非対象物を取り除く前と取り除いた後の画像で生成した，背景合成させる前のポーズ画像を比較する．

2. 提案手法

図 1　非対象物がある場合の劣化した生成画像

本研究では，非対象物を取り除いた人物画像によるポーズ画像生成手法を提案する．特に，**図 1** のように生成画像に非対象物の色が色移りしてしまう場合や，輪郭が不鮮明になる場合の問題を改善する手法について説明する．提案手法は，**図 2** に示すように，非対象物の除去による画像の

[1] 金沢工業大学
Kanazawa Institute of Technology
[a)] b1412491@planet.kanazawa-it.ac.jp
[b)] kogure@neptune.kanazawa-it.ac.jp

Vol.2019-CG-176 No.1
Vol.2019-DCC-23 No.1
Vol.2019-CVIM-219 No.1
2019/11/7

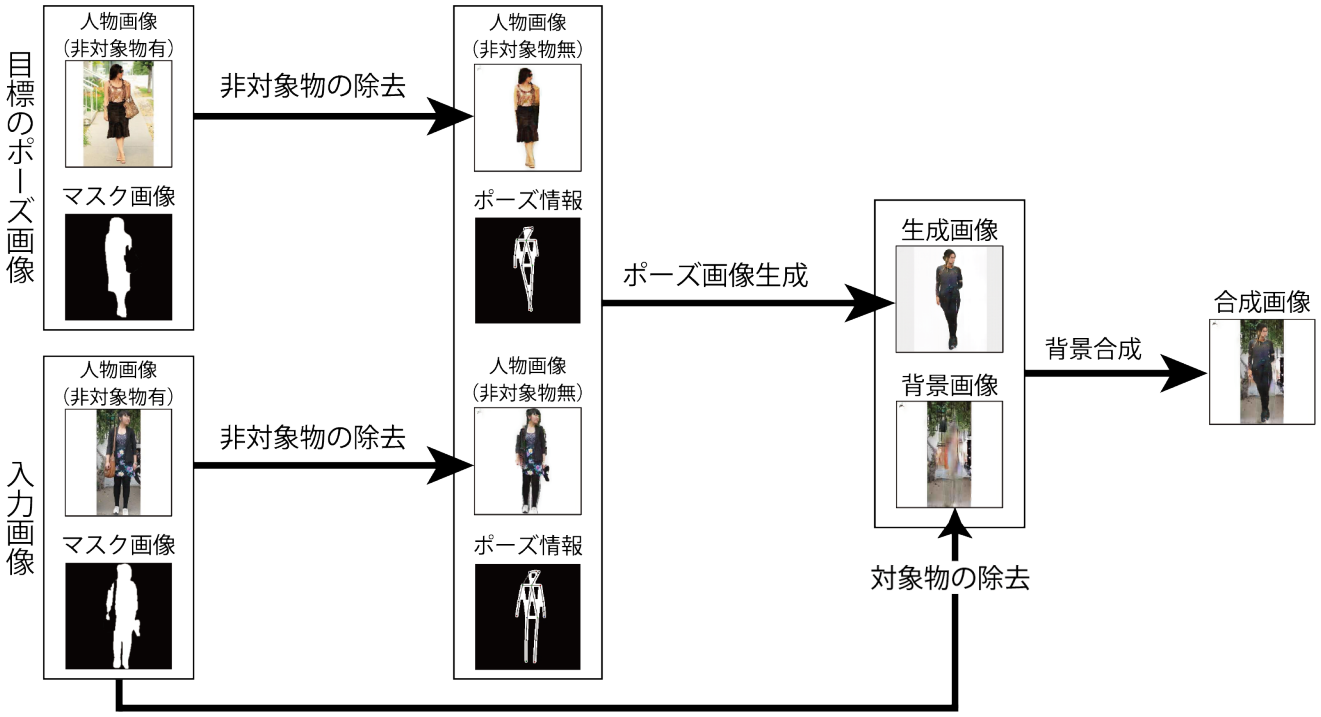


図 2　提案手法

作成と対象物の除去による画像の作成，ポーズ画像生成，背景の合成の 4 ステップに大きく分けられる．

2.1　非対象物の除去

入力画像と目標のポーズ画像から，非対象物を取り除いた人物画像を作成する．非対象物は人物画像と重ならない部分（背景）と人物画像と重なる部分（前景）に分けられる．まず，前景と人物のマスク画像を組み合わせる．次に，組み合わせたマスク画像をもとに白一色になるように背景を取り除く．最後に，前景のマスク画像をもとにインペインティング [6] によって前景を自然に見えるように取り除く．

インペインティングとは画像の欠落した（マスク）部分を自動修復させる手法である．本研究では，U-Net ベースのモデルで partial convolution レイヤーを用いたモデル [6] を使用する．このモデルは，どのような形状，サイズのマスクでも堅実に対応可能であり，マスク部分が大きくなっても精度が急激に低下することがないという利点がある．Partial convolution レイヤーとは，不規則なマスク画像を適切に扱うことができる層であり，式 (1) のように表される．

$$x' = \begin{cases} \mathbf{W}^T(\mathbf{X} \odot \mathbf{M})\frac{1}{\mathrm{sum}(\mathbf{M})} + b, & \text{if } \mathrm{sum}(\mathbf{M}) > 0 \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

式 (1) の W は畳み込みフィルターの重み，b は対応するバイアス，X は現在の畳み込み（スライディング）ウィンドウのための特徴量，M はバイナリマスクを示す．また，$\odot$ は要素単位の乗算を示す．Partial convolution レイヤーの出力は，欠損のない領域のみに依存しマスク画像を更新するというステップが含まれている．少なくとも 1 つの有効な入力で出力を調整できる場合，その場所のマスクを削除する．この層が十分な大きさであれば，マスク画像は徐々に縮小され，有効な値のみが特徴マップに残る．

2.2　対象物の除去

入力画像から，対象物を取り除いた背景画像を作成する．まず，第 2.1 節と同様に前景と人物のマスク画像を組み合わせる．そして，第 2.1 節と同様にインペインティングを用いて，こちらではマスク部分（前景と対象物）を自然に見えるように取り除く．

2.3　ポーズ画像生成

第 2.1 節で作成した人物画像からポーズ画像を生成する．本研究では，手法 [1] より手法 [5] の生成結果がよいことから，手法 [5] の Deformable GAN を用いてポーズ画像を生成する．

Deformable GAN はポーズ画像生成手法の一つであり，deformable skip connection という手法を使用する [5]．Deformable skip connection とは，目標のポーズになるように入力画像を変形する手法であり，生成器のエンコーダからデコーダへ特定の情報を受け渡す役割がある．具体的には，まず，ポーズ推定によって身体の 18 個の関節位置を推定する．次に，推定した関節位置をもとに，身体を 10 個のサブパーツ（頭，胴体，左右・上下の腕と脚）に分けモデル化する．サブパーツは特定の身体部分を囲む長方形

Vol.2019-CG-176 No.1
Vol.2019-DCC-23 No.1
Vol.2019-CVIM-219 No.1
2019/11/7

の領域を示す．そして，入力画像が目標のポーズをとるように，サブパーツの位置をもとに最小二乗誤差を用いてアフィン変換の関数を求める．最後に，求めたアフィン変換を組み合わせてオブジェクトの変形を近似する．

Deformable GAN では，図 2 に示すように，ポーズ画像生成のベースとなる入力画像と目標となるポーズ画像に加えて，両方の画像のポーズ情報を必要とする．そのため，第 2.1 節で作成した人物画像からポーズ推定 [4] を用いてポーズ情報を得る．最終的に，人物画像（非対象物無）とポーズ情報からポーズ画像を生成する．

2.4 背景合成

第 2.3 節で生成した人物画像と第 2.2 節で作成した背景画像を合成する．具体的には，生成画像の人ではない白い部分を透過し，そこに背景を合成させる．

3. 評価実験方法

非対象物を取り除いた場合のポーズ画像生成の品質を，生成画像の比較と SSD[7] による物体検出手法を用いて評価した．生成画像は，非対象物を取り除かなかった場合（非対象物有）と取り除いた場合（非対象物無）のものを比較した．非対象物を取り除いた場合は，前景のみを取り除いた場合，背景のみを取り除いた場合，前景と背景を取り除いた場合の 3 種類を用意した．また，前景はマスク画像を膨張させた場合を作成し違いがあるかを比較した．

Deformable GAN は同じ人物・服装のポーズの違うペア画像を学習データとしている．そのため，評価実験では，DeepFashion データセット [8] によって学習させた既存のモデルを使用した．

3.1 評価に使用したデータセット

本研究では，ModaNet データセット [9] を使用した．ModaNet とは，Paperdoll データセット [10] をベースとした大規模なストリートファッション画像のデータセットである．ModaNet では，服やカバンなどの 13 種類のファッションアイテムにポリゴンアノテーションが付加されている．

3.2 マスク画像の作成

本研究では，前景のマスク画像と背景のマスク画像に分けて作成を行った．前景のマスク画像は，第 3.1 節のポリゴンアノテーションにより作成した（**図 3**）．背景のマスク画像は，前景のマスク画像とセマンティックセグメンテーション [11] による人物識別により作成した人物のマスク画像を組み合わせることで作成した．

マスク画像を膨張させた場合の違いを確認するために，前景のマスク画像を膨張させたものを 3 種類作成した．本研究では，図 3 の通りに作成されたマスク画像からモルフォロジー処理によって膨張処理をしたマスク画像を作成した．膨張処理は**図 4** に示すように，4×4，8×8，16×16 サイズの計 3 種類のカーネルで行った．図 4 より，膨張させるサイズが大きくなると，前景部分の取り除かれる範囲が広くなるが，その分画像がぼけやすくなることが確認された．

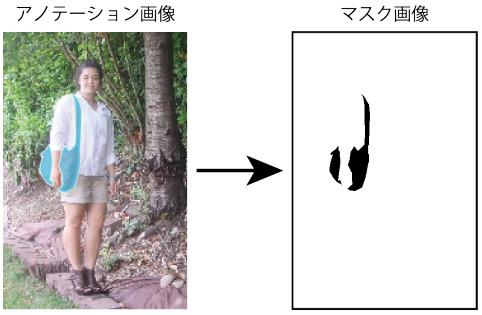


図 3 前景のマスク画像の作成

図 4 膨張処理をしたマスク画像とインペインティング画像

3.3 生成画像の比較

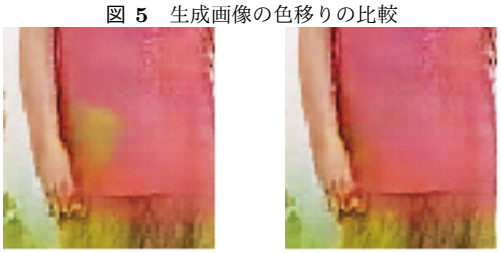
生成画像の比較では，非対象物有の場合と無の場合の生成画像でどのくらい違いがあるかを比較した．特に，色移りや輪郭が不鮮明になる問題が改善されているかに着目した．

3.4 SSD による評価

本研究では Deformable GAN[5] の評価方法に基づき，SSD[7] による人物検出のスコアによって評価を行った．具体的には，生成されたすべてのポーズ画像を SSD で人物検出し，その画像内の最大スコアを平均することで評価し

Vol.2019-CG-176 No.1
Vol.2019-DCC-23 No.1
Vol.2019-CVIM-219 No.1
2019/11/7

図 5　生成画像の色移りの比較



(a) 非対象物有　　(b) 非対象物（前景）無

図 6　生成画像の拡大図

た．本研究では，生成画像 1000 枚のスコアを平均した．

4.　評価実験結果

4.1　生成画像の比較

生成画像の比較により，非対象物無の場合では色移りの影響が小さくなることが確認された．**図 5** に色移りが改善された例を示す．また，図 5 の色移りが特に改善された箇所（四角枠部分）を拡大して**図 6** に示す．図 6 の箇所は，入力画像ではカバン（前景）にあたる部分である．図 5，図 6 では，非対象物有の生成画像で色移りしていた部分が非対象物無だと色移りの影響が小さくなっていることを示している．

また，生成画像の比較により，非対象物無の場合では輪郭が不鮮明になる点も改善されたことが確認された．特に，背景を取り除いた場合の生成画像が鮮明になることが確認

図 7　生成画像の輪郭の比較

された．**図 7** に輪郭が不鮮明になる点が改善された例を示す．図 7 では，非対象物有で輪郭が不鮮明であった生成画像が非対象物無だと輪郭が鮮明になることを示している．

4.2　SSD による評価

SSD による評価結果を**表 1** に示す．同表の SSD score は人物検出結果の平均を示し，非検出画像数は人物検出されなかった画像の枚数を示している．

同表より，背景を取り除いた場合はスコアが高く，非検出画像数が少ないことが確認された．特に，前景・背景ともに取り除いた場合のスコアが高いことが確認された．これより，非対象物を取り除く前処理は品質の向上に有効であるといえる．

前景のマスク画像に膨張処理を行った結果は，同表よりスコアにあまり変化がなかった．しかし，図 4 より，膨張処理を行わなかった場合に比べて前景の色が残りにくくなるため，色移りを改善する方法として有効であると示唆している．

5.　おわりに

本研究では，非対象物の色が生成画像に色移りしてしまうことや，生成画像の輪郭が不鮮明になってしまうことを

Vol.2019-CG-176 No.1
Vol.2019-DCC-23 No.1
Vol.2019-CVIM-219 No.1
2019/11/7

表 1　SSD による評価結果

生成画像条件	膨張処理	SSD score	非検出画像数
非対象物有	ー	0.901	53
非対象物（前景）無	ー	0.896	58
非対象物（前景）無	4×4	0.895	58
非対象物（前景）無	8×8	0.898	54
非対象物（前景）無	16×16	0.895	58
非対象物（背景）無	ー	0.920	40
非対象物（前景・背景）無	ー	0.922	36
非対象物（前景・背景）無	4×4	0.923	34
非対象物（前景・背景）無	8×8	0.921	38
非対象物（前景・背景）無	16×16	0.920	37

改善するために，指定した非対象物を取り除いた場合のポーズ画像生成手法を提案した．非対象物は前景と背景に分け，前景は自然に見えるように，背景は白一色になるように取り除いた．評価実験では，非対象物を取り除いた場合と取り除かなかった場合との生成画像を比較した．実験によって，非対象物を取り除いた場合のほうが色移りしてしまうことや輪郭が不鮮明になることを改善することができるという結果が得られた．

実験結果では，前景のマスク画像の膨張処理を行った結果は人物検出のスコアにあまり影響しなかった．そのため，今後の課題として，画像中の前景のマスクの占める割合や形状により非対象物の除去の影響を別の観点から評価していきたい．

参考文献

[1] Ma, L., Jia, X., Sun, Q., Schiele, B., Tuytelaars, T. and Van Gool, L.: Pose guided person image generation, *Advances in Neural Information Processing Systems*, pp. 406–416 (2017).
[2] Zheng, Z., Zheng, L. and Yang, Y.: Unlabeled Samples Generated by GAN Improve the Person Re-Identification Baseline in Vitro, *The IEEE International Conference on Computer Vision (ICCV)* (2017).
[3] Walker, J., Marino, K., Gupta, A. and Hebert, M.: The Pose Knows: Video Forecasting by Generating Pose Futures, *The IEEE International Conference on Computer Vision (ICCV)* (2017).
[4] Cao, Z., Simon, T., Wei, S.-E. and Sheikh, Y.: Realtime multi-person 2d pose estimation using part affinity fields, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7291–7299 (2017).
[5] Siarohin, A., Sangineto, E., Lathuilière, S. and Sebe, N.: Deformable gans for pose-based human image generation, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3408–3416 (2018).
[6] Liu, G., Reda, F. A., Shih, K. J., Wang, T.-C., Tao, A. and Catanzaro, B.: Image inpainting for irregular holes using partial convolutions, *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 85–100 (2018).
[7] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y. and Berg, A. C.: SSD: Single Shot MultiBox Detector, *European conference on computer vision*, Springer, pp. 21–37 (2016).
[8] Liu, Z., Luo, P., Qiu, S., Wang, X. and Tang, X.: DeepFashion: Powering Robust Clothes Recognition and Retrieval with Rich Annotations, *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016).
[9] Zheng, S., Yang, F., Kiapour, M. H. and Piramuthu, R.: ModaNet: A Large-Scale Street Fashion Dataset with Polygon Annotations, *ACM Multimedia* (2018).
[10] Yamaguchi, K., Hadi Kiapour, M. and Berg, T. L.: Paper doll parsing: Retrieving similar styles to parse clothing items, *Proceedings of the IEEE international conference on computer vision*, pp. 3519–3526 (2013).
[11] Zhao, H., Shi, J., Qi, X., Wang, X. and Jia, J.: Pyramid Scene Parsing Network, *CVPR* (2017).