1Q-05

深層学習におけるハイパーパラメータの調整による人物画像のアピアランスの変換への影響

渡部真拓 † 清水郁子 †

† 東京農工大学 情報工学科

1 はじめに

現在，クリエイターの制作補助として，簡単なスケッチからリアルな風景画像を生成する,「NVIDIA Canvas」というアプリが開発された。一方，人物画像に着目すると，例えば「Move Mirror」という Web カメラに映した自分の姿に似た画像を表示する技術など，人物の姿勢を指定した画像検索技術が確立しつつある。しかし，姿勢が同じままその人物の見た目を任意に変更することは不可能である。

そこで，本稿では，100〜200 枚程度のデータを用いて，姿勢を固定した人物画像のアピアランスを深層学習によって変換することを目的とし，姿勢が同じ人物画像のデータセットと服装や体型などの見た目が同じデータセットの間の変換を InstaGAN [1] により学習する。InstaGAN は CycleGAN [2] をベースとしており，CycleGAN 同様にそれぞれ異なる特徴をもった対応しない 2 つのデータセットを用いて画像の変換が可能であるが，CycleGAN と比べ形状変化に対応しているという特徴がある。InstaGAN のハイパーパラメータで損失の倍率を調整可能を調整した際の出力画像に影響について報告する。

2 InstaGAN によるアピアランスの変換

InstaGAN [1] は，通常の画像に加え，変更対象領域を表すセグメンテーションマスクを用いて画像を変換する。各データセットの画像が持つ特徴をドメインと呼ぶ。例えば，ドメイン X を立っている男性，ドメイン Y を緑の T シャツに水色のズボンの細身の女性とする。変換元の画像を x，セグメンテーションマスクは a，変換先の画像を y'，セグメンテーションマスクを b' として，ドメイン $X \to Y$ へと変換した場合の画像を図 1 に示す。

InstaGAN では，画像の生成器と識別器に加え，セグメンテーションマスクの生成器と識別器を用いる。損失関数は，各識別器が本物か偽物かの判定を行う敵対性損失，一度変換した画像を逆変換で戻した画像と元の画像との絶対値誤差であるサイクル一貫性損失，色空間を保つために用いる同一マッピング損失，形状を変形するためのコンテクスト損失の 4 つがある。ハイパーパラメータによって各損失の倍率が各々調整可能である。今回，調整対象とした同一マッピング損失，コンテクスト損失について説明する。

Transforming the appearance of human images by adjusting the hyperparameters of deep learning

Mahiro WATANABE† and Ikuko SHIMIZU†

†Faculty of Engineering, Tokyo University of Agriculture and Technology,184-8588, Tokyo, Japan

s182584w@st.go.tuat.ac.jp

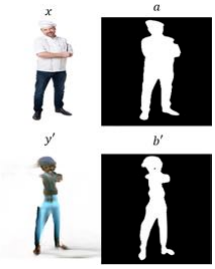

図 1: セグメンテーションマスク

同一マッピング損失は変換元の画像の色空間を保つ役割をしている。ドメイン $X \to Y$ へと変換する生成器を G_{XY} とする。この生成器は，ドメイン X の画像 x とセグメテーションマスク a を入力すると，ドメイン Y の画像 y' とセグメテーションマスク b' を出力する。生成器を学習する際には，G_{XY} に対し，ドメイン Y の画像を入力する。同一マッピング損失は，次の通りである。

$$\mathcal{L}_{\mathtt{idt}} = \|G_{XY}(y,b) - (y,b)\|_1 + \|G_{YX}(x,a) - (x,a)\|_1$$

コンテクスト損失は InstaGAN の形状変化を実現するため，背景部分の変換を抑え，対象部分の変換を行いやすくする性質を持っている。図 1 の a と b' においてどちらも背景 (黒い部分) とした範囲を $w(a,b')$ とし，その範囲内の x と y' との絶対値誤差を損失としている。ドメイン $Y \to X$ と変換する場合も同様に計算を行っている。コンテクスト損失は，次の通りである。

$$\mathcal{L}_{\mathtt{ctx}} = \|w(a,b') \odot (x - y')\|_1 + \|w(b,a') \odot (y - x')\|_1$$

3 実験

3.1 簡単な姿勢の場合の変換結果

比較的簡単なポーズについて実験を行った。立っている男性の姿勢のままで，緑色の T シャツに水色のズボンを履いた細身の女性の人物画像へ変換するため，立っている男性の画像をドメイン X，緑色の T シャツに水色のズボンを履いた細身な人物の画像をドメイン Y として変換した。

立っている男性の画像は WEB スクレイピングによって収集した。緑色の T シャツに水色ズボンを履いた人物の画像は Figurosity [3] という人間の CG モデルが載っているサイトを用いてデータセットを作成した。この CG モデルのデータも立っているもののみで構成している。画像の一例を図 2 に示す。

まず，同一マッピング損失を調整した場合の結果を示す。マッピング損失の重みのデフォルト値は 1.0 で

図 2: 学習に用いたデータセットの画像例

あるが，0.0 倍，1.0 倍，2.0 倍と変更を行い学習を行った。0.0 倍と 1.0 倍のエポック数は 40，2.0 倍のエポック数は 45 である。

図 3: 同一マッピング損失の調整結果

2.0 倍の結果において，上半身が茶色がかった変換と下半身の一部に黒色がかかった変換となっている。同一マッピング損失は本来，変換元の色空間を保つ作用がある。同一マッピング損失の重みを大きくすることで，変換元の男性の服の色を変換先でも保つように働いていると考えられる。今回のデータセットの男性は立っているという特徴があるものの服の色は固定していないため，同一マッピング損失は学習画像によって左右され，安定しなくなっている。しかし，服の色を固定していないとはいえ，男性の服は青や黒目の服を着る傾向が大きいため，茶色がかった黒い色への変換が発生したと考えられる。

次に，コンテクスト損失を調整した結果を示す。セグメンテーションマスクが損失に関与しているため，その画像も示す。0.0 倍と 2.0 倍のエポック数は 45，1.0 倍のエポック数は 40 となっている。

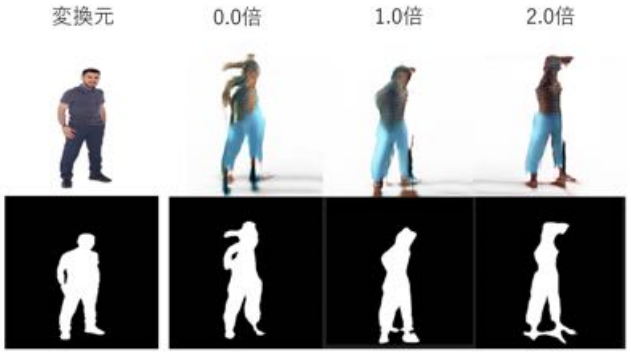

図 4: コンテクスト損失の調整結果

1.0 倍，2.0 倍の変換に大きな変化は見られないが，0.0 倍の変換で下半身の変換に一部変化がある。セグメンテーションマスクを見ると，右側の足が大きく変換されていることがわかる。この変換は変換元の男性の体格の大きさが関係していると考えられる。基本的に女性の CG モデルの体の大きさよりも，男性の方が体格は大きいため，0.0 倍の変換は元の男性画像の体格の影響が引き継がれているものと考えられる。InstaGAN は従来のものと比較して，このコンテクスト損失を導入することで，形状変化に対応できるようになっている。この実験からも，セグメンテーションマスクを用いたコンテクスト損失を導入することで形状が変形するようになっていることがわかる。

3.2 難しい姿勢の場合の変換結果

本節ではより複雑な姿勢である Kaggle の Yoga Poses Dataset [4] から，木のポーズのデータを用いて変換を行った結果を示す。木のポーズのままで緑色の T シャツを付け水色のズボンを履いた細身な人物画像へ変換させるため，木のポーズの画像をドメイン X，緑色の T シャツに水色のズボンを履いた細身な人物の画像をドメイン Y として変換した。画像の一例は図 5 に示す。また，ポーズの変化に対応できるように，CG モデルにも座っている姿など多様なポーズを追加した。

図 5: ヨガ用ポーズの画像

木のポーズをしたまま CG モデルのような服の色となることを目指したが，手の位置が大きく異なるポーズへの変換でありうまく実現できなかった。ただし，パラメータ調整を行うと，手に関して少し変換ができたため，その出力画像とパラメータを示す。サイクル一貫性損失の重みのデフォルト値は 10.0 であるが，CG →ヨガに変換する場合の重みは 5.0 へ，ヨガ→ CG に変換する場合の重みは 3.0 へ変更した。また，コンテクスト損失の重みはデフォルト値の 1.0 から 1.5 に変更している。エポック数は 150 である。出力画像を図 6 に示す。この例では，サイクル一貫性損失を下げた

図 6: 手をある程度変換した結果画像

結果，相対的に敵対性損失とコンテクスト損失の割合が高まった。その結果，損失がセグメンテーションマスクの形状の変換に大きく関わることになるため，手の位置が大きく変化する変換を部分的に実現していると考えられる。

4 まとめ

InstaGAN を用いて人物の姿勢を保持してアピアランスを変化させる方法について検討した結果を報告した。今後はデータセットの画像枚数を増やした学習について検討する。

参考文献

[1] S. Mo, et. al, “InstaGAN: Instance-aware Image-to-Image Translation,” Proc.ICLR, (2019).

[2] J.Y. Zhu, et. al, “Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks,” Proc.ICCV, (2017).

[3] https://figurosity.com

[4] https://www.kaggle.com/general/192938