2P-01

浮世絵のランドマーク付き顔画像に対する 3D-aware 画像合成

新井田　力[†]　　野口　渉[‡]　　山本　雅人[†1‡1]
北海道大学 大学院情報科学院[†]　　北海道大学　数理・データサイエンス教育研究センター[‡]
北海道大学　大学院情報科学研究院[†1]　北海道大学 人間知・脳・AI 研究教育センター[‡1]

1. はじめに

画像を三次元的に解釈する技術は高度な画像処理・映像処理・3D モデリングにおいて重要な課題であり, 近年の深層学習に基づく 3D-aware 画像合成はその基礎となる技術として期待されている. しかし, 既存の 3D-aware 画像合成の研究は写実的な画像の領域に限定されており, 浮世絵のような非写実的な画像に対しては行われていない. 浮世絵は明確な輪郭線, 三次元形状の矛盾や部分的な平面化, 意味の図形的強調など非写実的な特徴をもつことに加え, カメラで撮影した画像とは異なり画像が生じる過程において物理的な三次元空間の存在が保証されない. 本研究では, そのような性質をもつ浮世絵画像に対して, ランドマークラベルを利用した 3DMM により推定されるカメラポーズを利用することで, 正確なポーズラベルのない非写実的な領域での 3D-aware 画像合成を実現した. また, 輪郭線などの非写実的な特徴が立体構造に与える影響の観察を行った.

2. 手法

2.1. GAN と三次元への拡張

GAN (Generative Adversarial Network, 敵対的生成ネットワーク)[1]は「生成器」と「識別器」が対立しながら学習を進める仕組みであり, 高い質の画像を生成できる. また, 二次元の画像合成を三次元に拡張する試みも行われている.

3D-Aware Image Synthesis for Ukiyo-e Portraits with Facial Landmarks

Chikara Niida[†] Noguchi Wataru[‡] Yamamoto Masahito[†1‡1]
† Graduate School of Information Science and Technology, Hokkaido University
‡ Education and Research Center for Mathematical and Data Science (MDS), Hokkaido University
†1 Faculty of Information Science and Technology, Hokkaido University
‡1 Center for Human Nature, Artificial Intelligence, and Neuroscience (CHAIN), Hokkaido University

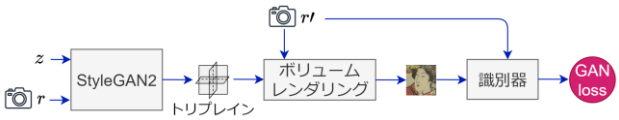

図 1　EG3D の概要

EG3D[2] (図 1) は二次元の画像と被写体に対するカメラポーズのデータを用いて三次元的に一貫した高解像度の自由視点画像を合成することに成功している. この手法は二次元の CNN を用いつつ三次元に関するバイアスを与えるために StyleGAN2[3]およびトリプレイン構造を用いている. StyleGAN2 には潜在変数 z およびカメラポーズ r が入力され, 二次元の特徴マップが出力される. この特徴マップをチャネル方向に 3 分割し, それぞれを xy, yz, zx 平面 (トリプレイン) とみたて, 三次元空間内の点 P に対して各平面の投影点の特徴ベクトルを合わせたものを点 P に対応する特徴ベクトルとする. これらの特徴ベクトルを小規模の MLP に入力することで, 三次元空間の点に対応する色や密度などが得られ, 指定されたカメラポーズ r' からのニューラルボリュームレンダリング[4]により低解像度の画像が得られる. さらに超解像度ネットワークを通して最終的な画像が得られる. 低・高解像度の画像はカメラポーズ r' とともに識別器に入力される. r および r' は訓練データからランダムに選んだ角度で, 確率 0.5 で $r = r'$ とする. 本研究では EG3D を用いて画像合成を行った.

2.2. データセット

EG3D では各訓練画像のカメラポーズラベルが必要だが非写実的な画像データセットでカメラポーズラベルのあるものは存在しない. 本研究で使用した ARC 浮世絵顔データセット[5]は, 9203 枚の浮世絵画像から顔部分を 16653 枚抽出したデータセットであり, 目・鼻・口などの特徴点に対して 1 枚あたり 31 個のランドマーク (画像上の二次元座標) がつけられている.

本研究では人の顔の 3DMM[6]を用いたフィッティングにより推定されたカメラポーズを EG3D での学習に使用した. また, ランドマークに基づいて顔の中央ができるだけ揃うように画像のクロッピングを行い, その後解像度を 512×512 になるよう線形補間を行った.

3. 実験

3.1. 実験設定

ボリュームレンダリングにおけるサンプリングは解像度 64×64 の画像となるように行い, その後512×512 に超解像を行った. EG3D と同様に GAN loss の他に密度の連続性に関する正則化 loss を導入した. 学習は Adam による勾配降下法により行った.

3.2. 評価

学習後のモデルにより生成された画像は浮世絵らしいものであり (図 2), 本物の画像のデータと偽物の生成画像のデータの距離を計測する指標である FID (Fréchet inception distance) は 39.6 となった. また, 生成された立体はもっともらしい形状になっていることも確認できた. 特に, 見る視点に依存して輪郭線が現れたり消えたりすることがわかった. 輪郭線は表面の法線方向とカメラの向きが垂直に近いときに表れることが多いが, 垂直から遠い場合においても形状が不自然に変形する様子は観察されなかった. 一方, 一部の生成された立体は顔らしい形状ではなく単なる平面になっていた. 平面になりやすい要因として, 非写実的な浮世絵特有の性質である部分的に平面的にみえることが考えられる. また, 首が左右に分離し 2 つあるような形状のものもあった (図 2 第 2 段). この現象の理由としては, 中央付近からの視点の訓練画像が少ないことが考えられる.

4. まとめ

本研究では非写実的な特徴をもつ画像領域の一つである浮世絵画像に対して, ランドマークラベルを利用した 3DMM により推定されるカメラポーズを利用することで 3D-aware 画像合成が可能であることが示された. また, 画像が生じる過程において物理的な三次元空間の存在が保証されていない画像領域に対して三次元空間を仮定したモデルを適用できる場合があることがわかったという点においても一定の意義があると考えられる. さらに, 鼻などを表す輪郭線が向きにより出現したり消滅したりすることがわかった. これらの結果は非写実的な画像に対する三次元的な解釈や確率分布を与え, inversion による画像の立体化や 3D モーフィングを可能にすると考えられる. さらに, 三次元構造を考慮した画像編集などの高度な画像処理のほか, カメラポーズや潜在空間を連続的に変えることによる動画生成, 動画の三次元構造推定, 密度情報に基づく 3D モデリングの自動化など様々な応用につながる可能性がある.

90° 60° 30° 0° -30° -60° -90°

図 2 生成された画像の例

参考文献

[1] I. Goodfellow ほか, 「Generative Adversarial Nets」, *Advances in Neural Information Processing Systems*, Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence と K. Q. Weinberger, 編, Curran Associates, Inc., 2014. [Online]. Available at: https://proceedings.neurips.cc/paper_files/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf

[2] E. R. Chan ほか, 「Efficient Geometry-aware 3D Generative Adversarial Networks」, *CVPR*, 2022.

[3] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen と T. Aila, 「Analyzing and Improving the Image Quality of StyleGAN」, *Proc. CVPR*, 2020.

[4] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi と R. Ng, 「NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis」, *Proceedings of the European Conference on Computer Vision (ECCV)*, 2020. [Online]. Available at: http://arxiv.org/abs/2003.08934v2

[5] Y. Tian, 人文学オープンデータ共同利用センター, T. Yingtao と Center for Open Data in the Humanities, 「ARC Ukiyo-e Faces Dataset」. 人文学オープンデータ共同利用センター. doi: 10.20676/00000394.

[6] P. Huber ほか, 「A multiresolution 3d morphable face model and fitting framework」, *International conference on computer vision theory and applications*, SciTePress, 2016, pp. 79–86.