Vol.2018-CG-171 No.17
2018/9/30

「Real-Time Hair Rendering using Sequential Adversarial Networks」の実装

谷田川 達也[1]

概要：本稿では，ECCV 2018 で発表された Wei らの論文「Real-Time Hair Rendering using Sequential Adversarial Networks」を実装し，その実験結果を報告する．同論文は深三次元の髪形状と例示となる頭髪画像を入力として，例示に沿った頭髪の実時間レンダリングを畳み込みニューラルネットワークにより実現する．提案ネットワークは頭髪の三次元形状を段階的にレンダリング結果に近づけるような三つの部分ネットワークから成る．著者らの実装では提案ネットワークの実装に CelebA-HQ データセットに髪領域ラベルを手付けして用いている．本データセットは未公開であるため，本稿では意味的領域ラベル付きの顔画像の代表的なデータセットである Labeled Faces in the Wild (LFW) で実験を行った．

1. はじめに

本稿では ECCV 2018 で発表された Wei らの論文 [7] を実装して実施した実験の結果について報告する．彼らの手法は，畳み込みニューラルネットワーク (以下，CNN) を例示ベースの頭髪の実時間レンダリングに応用したもので，頭髪の三次元形状と例示となる頭髪画像を入力とすると，例示に沿った頭髪のレンダリング結果が得られる．提案法の概要を図 1 に示す．提案法は例示の頭髪画像 (I_0) から頭髪部分を切り出した画像 (I_1) を得た後，この画像をグレースケール画像 (I_2)，毛髪向き画像 (I_3)，エッジ活性化画像 (I_4) に順次画像変換する．入力の三次元形状に対しては，事前学習した CNN により，上記とは逆順の画像変換を事前学習した施すことで最終的なレンダリング画像が得られる．以下では，この提案法に用いるネットワークと今回実施した実装内容について述べる．

2. 提案手法と実装

例示画像の変換処理において，例示画像から頭髪部分の切り出しは最新の意味的領域分割手法である PSPNet [8] により行われる．この頭髪部分画像に対して，グレースケール化，毛髪向き検出 [5] ならびにエッジ活性化の処理が施される．なおエッジ活性化については，元論文に詳細な記述がなかったため 3×3 のラプラシアン・フィルタを施したものを Otsu らの手法により二値化して得た．

実際に頭髪の三次元形状を入力としたレンダリングを行う際には，上記の画像変換の逆変換が提案ネットワークに

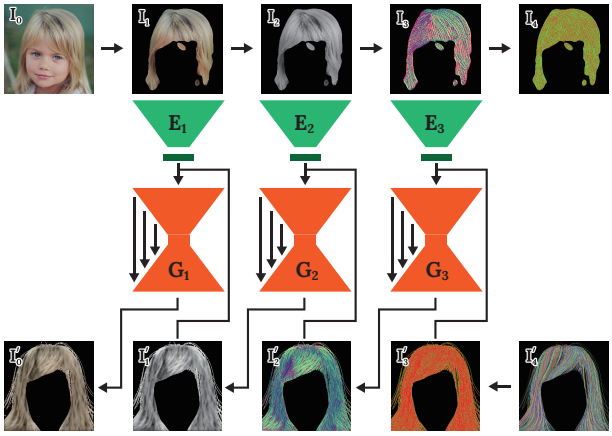

図 1 Wei ら [7] の提案ネットワーク．例示画像は Flickr のユーザ Anaïs Nannini のものを使用．

より施される．まず，頭髪の三次元形状に対して髪の毛一本一本をランダムな色で描画した画像を作成する (図 1 の I'_4)，上記と同様にラプラシアン・フィルタと二値化処理によりエッジ活性化画像を得る．この画像を順次，CNN に通すことで画像逆変換が行われる．それぞれの画像逆変換は別々のエンコーダ・デコーダネットワーク行われる．提案法ではエンコーダに ResNet [1] を元としたネットワークを，デコーダに U-Net [6] を基にしたネットワークを用いている．各画像逆変換では，例示画像側の中間画像からエンコーダにより抽出された潜在変数ベクトルがタイル状に複製されてレンダリング画像側の中間画像と接続される．この画像をデコーダに通すと次の中間画像が得られる．

著者らの実装では学習に高画質顔画像データセットの一つである CelebA-HQ [3] を用いている．著者らは PSPNet

[1] Email: tatsy@acm.org, 早稲田大学大学院 先進理工学研究科

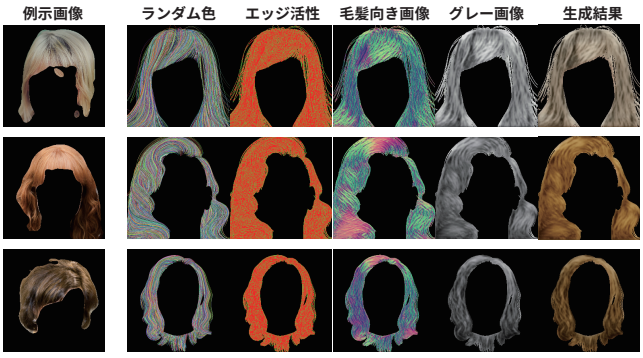

図 2 LFW データセットを用いた実験結果．実験に用いた例示画像は Flickr のユーザ Anaïs Nannini (1 行目) ならびに Qsimple (2, 3 行目) のものを使用．

の学習のために，データセット中の 3,000 枚の画像について髪の領域を表すラベルが手付けされたデータを作成した．しかし，この髪領域ラベルつきのデータは未公開であるため，今回は予め髪領域のラベルがついた Labeled Faces in the Wild (LFW) データセット [4] を用いて実験を行った．LFW には 250×250 画素の画像が約 13,000 枚含まれており，そのうち約 3,000 枚に意味的領域分割ラベルが付与されている．今回は実装の都合上，これらの画像を 256×256 画素にリサイズして用いた．また，今回の実装ではエンコーダの学習を簡単にするため，ResNet の代わりに単純な畳み込み層と活性化層からなるネットワークをエンコーダに用いた．ネットワークは NVIDIA GeForce 1080 Ti 二台を用いておよそ半日間で 15 エポック分学習した．

3. 結果と考察

今回の実装により得られた結果を図 2 に示す．この結果を見ると，エッジ活性化画像ならびに毛髪向き画像は学習データに近い良好な見た目が得られていることが分かる．その一方で，グレースケール画像ならびにレンダリング結果画像においては，やや見た目がボケた印象となっている．さらに，レンダリング結果の髪色に注目すると，例示画像の髪色をあまり正しく反映できていないことが分かる．この問題は LFW データセットに含まれる画像が全体的にボケていること，ならびにデータセットに含まれる画像が黒や茶などの髪色を多く含んでいることが原因と考えられる．

また，今回の実装にあたり，論文に書かれている内容からいくつかの変更を加えているので，その内容について考察する．一点目に生成画像 I'_k $(k = 0, 1, 2, 3)$ に対する損失関数の定義である．元論文を見ると，損失関数を定義する際の I'_k は全て I_4 から順次ネットワークにより生成されたもののように見えるが，このように学習すると I'_2 や I'_1 など，より後半で生成される画像の品質がなかなか上がらない問題があった．そこで今回の実装では I_4 から生成された画像とは別にデータセット中の I_{k-1} から生成した I''_k も損失関数の定義に用いた．二点目として髪の流れが全体的にボケた印象になるのを防ぐため，通常の判別器 (Discriminator) と合わせてパッチ判別器 [2] も用いた．これにより，髪により詳細な流れの情報が現れることが確認できた．

4. まとめ

本稿では Wei らが ECCV 2018 で発表した CNN を用いた頭髪の実時間レンダリングに関する論文 [7] の実装結果について報告した．今回は学習データに LFW データセットを用いて 20 エポック分の学習を実施したが，データセットを著者らと同様に CelebA-HQ に変更し，更に長時間の学習を行うことで，今回報告した結果よりも良好な結果が得られることが期待される．

また，元論文の結果ならびにデモ動画を確認すると，Wei らの提案法に関する問題がいくつか見て取れる．第一に彼らの手法は例示画像と実際にレンダリングされる頭髪画像の見た目がある程度近い必要がある．今回の実装を用いた実験においても，特に例示画像とレンダリング画像で毛髪部分のスケールが大きく異なる場合に不自然な結果が得られることが確認できた．また，グレースケール画像を生成する際には例示画像に含まれる光源環境の情報を十分な反映が難しいようで，全体的に環境光が支配的な例示画像でないと良好な結果が得られないことも分かった．今後は，この問題に関して，より発展的な手法の開発が望まれる．

参考文献

[1] He, K., Zhang, X., Ren, S. and Sun, J.: Deep residual learning for image recognition, *IEEE Conference on Computer Vision and Pattern Recognision (CVPR)*, pp. 770–778 (2016).
[2] Isola, P., Zhu, J.-Y., Zhou, T. and Efros, A. A.: Image-to-Image Translation with Conditional Adversarial Networks, *IEEE Conference on Computer Vision and Pattern Recognision (CVPR)* (2017).
[3] Karras, T., Aila, T., Laine, S. and Lehtinen, J.: Progressive Growing of GANs for Improved Quality, Stability, and Variation, *International Conference on Learning Representations (ICLR)* (2018).
[4] Learned-Miller, E., Huang, G. B., RoyChowdhury, A., Li, H. and Hua, G.: Labeled Faces in the Wild (LFW) Dataset, http://vis-www.cs.umass.edu/lfw/part_labels/.
[5] Luo, L., Li, H., Paris, S., Weise, T., Pauly, M. and Rusinkiewicz, S.: Multi-view hair capture using orientation fields, *IEEE Conference on Computer Vision and Pattern Recognision (CVPR)*, pp. 1490–1497 (2012).
[6] Ronneberger, O., Fischer, P. and Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation, *arXiv preprint arXiv:1505.04597* (2015).
[7] Wei, L., Hu, L., Kim, V., Yumer, E. and Li, H.: Real-Time Hair Rendering using Sequential Adversarial Networks, *European Conference on Computer Vision (ECCV)* (2018).
[8] Zhao, H., Shi, J., Qi, X., Wang, X. and Jia, J.: Pyramid Scene Parsing Network, *IEEE Conference on Computer Vision and Pattern Recognision (CVPR)* (2017).