The 33rd Annual Conference of the Japanese Society for Artificial Intelligence, 2019

Neural 3D Mesh Renderer

加藤 大晴 *1 Hiroharu Kato　牛久 祥孝 *2 Yoshitaka Ushiku　原田 達也 *1*2 Tatsuya Harada

*1東京大学 The University of Tokyo　*2理化学研究所 RIKEN

We introduce our paper "Neural 3D Mesh Render" presented at CVPR and MIRU last year. In this work, we proposed a novel renderer that takes a 3D mesh, light, and camera setting and outputs an image. Because "back-propagation" is defined in our renderer, it can be used as a layer of deep neural networks. By using it, we can pass the gradient of a loss into a 3D space through renderer and optimize components there. In experiments, we demonstrated the effectiveness of our renderer by applying it to view-based training of single-view 3D reconstruction, 2D-to-3D style transfer, and 3D DeepDream. We also introduce some papers that use our renderer for other problems.

1. はじめに

本発表は既発表論文 [Kato 18] の紹介である．

2 次元画像を受け取り，その背後の 3 次元世界における何かを推定（最適化）することは，コンピュータビジョンにおける本質的な課題のひとつである．画像を元に 3 次元構造を推定するタスクはその直接的な例であるが，シーン理解，ポーズ認識などの多くのタスクについても間接的に画像の 3 次元構造の推定が必要とされる．

近年では，深層学習によってパイプライン全体を一気通貫学習するアプローチが広く用いられている．ここでたとえば，画像を受け取りそこに映る物体の 3 次元モデルを出力する再構成器を，2 次元画像のみを用いて訓練することを考える．目的関数として「出力された 3 次元モデルが，2 次元に再投影されたときに，正しい画像と同じように見えるか」という規準を用いる．そのようなパイプラインを図 1 (上) に示す．このパイプラインを一気通貫学習するためには，2 次元画像上で定義される目的関数の勾配が，3 次元モデルを 2 次元へ投影するレンダラーを通じて 3 次元世界へと流れ込まなければならない．また，図 1 (下) に示す「3 次元モデルを，2 次元画像上で定義されたスタイルの類似性を高めるように最適化する」というパイプラインについても同様の処理が必要となる．

この「2 次元画像への投影」を深層学習に組み込むことは，ボクセルと呼ばれる 3D 表現については行われてきた [Yan 16]．しかし，ボクセルはピクセルを 3 次元に拡張したものであり，メモリ使用量が解像度の 3 乗に比例して大きくなるという問題がある．本研究では，3 次元表現として頂点と面の集合から成るメッシュを対象に，深層学習に組み込めるレンダラーを提案する．通常のレンダラーには，各ピクセルごとに色をサンプリングするラスタライズと呼ばれる過程において，ピクセル色を頂点座標で微分すると常にゼロとなり勾配が流れないという問題があるため，本研究ではラスタライズに疑似的な勾配を定義することによってそれを解消する．

2. 提案手法の概略

レンダラーは，3D モデル，照明情報，カメラ情報を受け取り，画像を出力する．この処理は頂点座標の変換やラスタライズなどから成るが，ラスタライズを除くほとんどの処理は微分が自然に定義されるため，勾配を特別に考える必要はない．

連絡先: {kato,ushiku,harada}@mi.t.u-tokyo.ac.jp

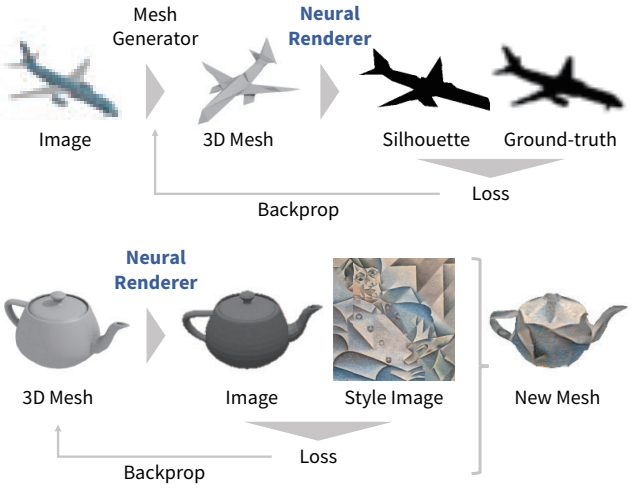

図 1: Pipelines for single-image 3D mesh reconstruction (upper) and 2D-to-3D style transfer (lower).

図 2 (a) のように，1 枚のポリゴンのみから成るシーンを考える．ラスタライズでは，ピクセルがポリゴンと重なるとき，ピクセルがそのポリゴンの色で塗られる．図 2 (b) はピクセル P_j の色 I_j とポリゴンの頂点の座標 x_i の関係を示している．頂点 x_i が右に動くと，ポリゴンが P_j に衝突したときにそのポリゴンの色が変わる．図 2 (c) はこの関数の微分値を示し，これはほぼ常にゼロであるため誤差逆伝播がうまく機能しない．そこで本研究では，逆伝播時にのみ図 2 (b) ではなく図 2 (d) のような関係が成立しているとみなし，その傾きを勾配とする．これによって逆伝播時に「頂点がどちらに動くとピクセルの色がどう変わりそうか」という情報を伝えることができ，レンダラーを含むパイプラインの一気通貫学習が可能となる．

3. 実験

3.1 単一画像 3 次元再構成

図 3 および表 1 は，図 1 (上) のパイプラインを用いて三次元再構成器を訓練する実験の結果である．データセットは ShapeNet を使用した。メッシュの生成は，予め定義した球の頂点を移動させることによって実現する．適切に 3 次元再構成が行えていることから，2 次元画像上でシルエットを比較する

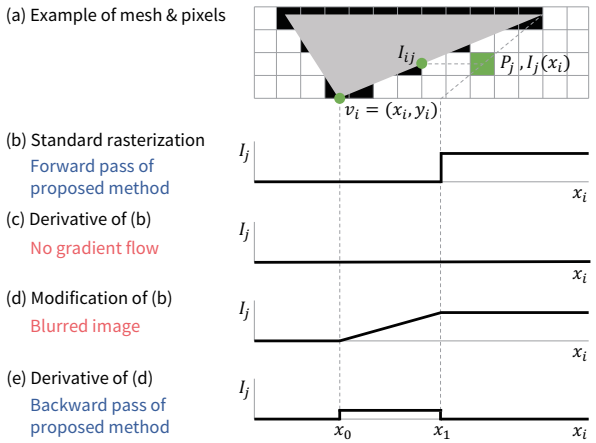

図 2: Illustration of our method. $\boldsymbol{v}_i = \{x_i, y_i\}$ is one vertex of the face. I_j is the color of pixel P_j. The current position of x_i is x_0. x_1 is the location of x_i where an edge of the face collides with the center of P_j when x_i moves to the right. I_j becomes I_{ij} when $x_i = x_1$.

Retrieval [Yan 16]	Voxel [Yan 16]	Mesh (ours)
0.4766	0.5736	**0.6016**

表 1: Reconstruction accuracy measured by voxel IoU. Higher is better. Our mesh-based approach outperforms the voxel-based approach [Yan 16] in 10 out of 13 categories in ShapeNet.

目的関数の勾配がレンダラーを通じて適切に三次元再構成器へと流れていることがわかる．また，我々の手法で得られるメッシュにはノイズボクセルに表れるような粗いノイズがなく，また再構成の定量的な性能もボクセルを用いる場合を上回る．

3.2 2 次元画像から 3 次元モデルへのスタイル転移

図 4 は，図 1 (下) のパイプラインを用いて 2 次元の画像のスタイルを 3 次元モデルへと転移する実験の結果である．スタイルの類似性の計算には [Gatys 16] を用いた．

図からは，ウサギのテクスチャ画像にスタイル画像の色が転移していることや，ティーポットのふたの形が円形からスタイル画像のように直線的な形へと変化していることが読み取れる．これは，提案したレンダラーを用いることでスタイルに関する情報を 2 次元画像の空間から 3 次元画像の空間へと適切に流し込めることを示している．

4. まとめ

本研究では深層学習に組み込むことのできるメッシュのレンダラーを提案し，単一画像 3 次元再構成と，画像から 3 次元モデルへのスタイル転移でその効果を検証した．

提案したレンダラーの応用可能性は本論文で示したものに留まらない．既に画像集合からのテクスチャ付きの 3 次元モデルの再構成 [Kanazawa 18] や 3 次元再構成を経由した画像の編集 [Yao 18] などへと応用されており，今後も様々なタスクへと広がってゆくことが期待される．本研究のソースコードはオンラインで公開している *1．

*1 `https://github.com/hiroharu-kato/neural_renderer`

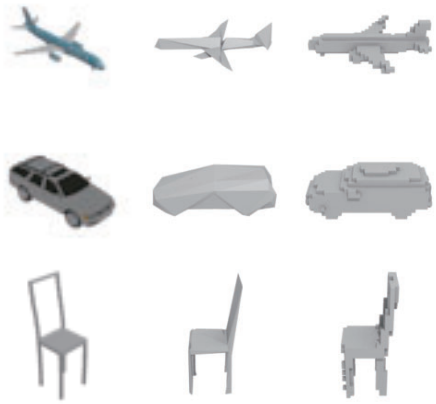

図 3: 3D mesh reconstruction from a single image. Results are rendered from three viewpoints. First column: input images. Second column: mesh reconstruction (proposed method). Third column: voxel reconstruction [Yan 16].

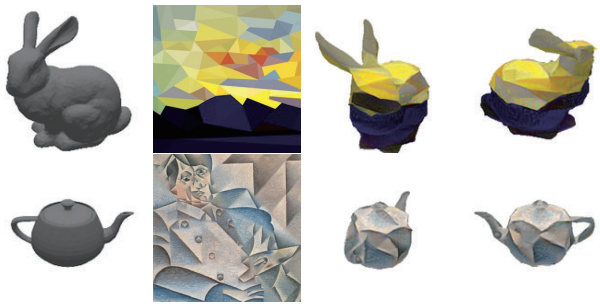

図 4: 2D-to-3D style transfer. The two left images represent initial state of a 3D model and style respectively. The style images are *Thomson No. 5 (Yellow Sunset)* (D. Coupland, 2011) and *Portrait of Pablo Picasso* (J. Gris, 1912).

参考文献

[Gatys 16] Gatys, L. A., Ecker, A. S., and Bethge, M.: Image style transfer using convolutional neural networks, in *CVPR* (2016)

[Kanazawa 18] Kanazawa, A., Tulsiani, S., Efros, A. A., and Malik, J.: Learning Category-Specific Mesh Reconstruction from Image Collections, in *ECCV* (2018)

[Kato 18] Kato, H., Ushiku, Y., and Harada, T.: Neural 3d mesh renderer, in *CVPR* (2018)

[Yan 16] Yan, X., Yang, J., Yumer, E., Guo, Y., and Lee, H.: Perspective transformer nets: Learning single-view 3d object reconstruction without 3d supervision, in *NIPS* (2016)

[Yao 18] Yao, S., Hsu, T. M., Zhu, J.-Y., Wu, J., Torralba, A., Freeman, B., and Tenenbaum, J.: 3D-aware scene manipulation via inverse graphics, in *NIPS* (2018)