

畳み込みニューラルネットワークによる CG 画像の強弱有り線画像の生成

前田浩輔[†] 齋藤豪^{††}

[†]東京工業大学 工学部 ^{††}東京工業大学 情報理工学院

1 はじめに

近年、国内で3DCGによるセル画のようなアニメーションの制作が盛んに行われている。本稿ではそのための線描画に注目する。

3DCGより生成した法線画像と深度画像を入力とし、畳み込みニューラルネットワークを用いることで、強弱付きの線画を生成する手法を提案する。従来のアルゴリズムでは難しかった線の強弱量の決定を、データに基づき行うことを可能とする。

2 提案手法

提案手法では、法線画像・深度画像と手描き線画像のセットを学習データとし、これを用いてネットワークを学習させることで、強弱有り線画像を生成する。以下では、学習データ、学習モデルについて述べる。

2.1 学習データ

ネットワークの入力には、法線画像と深度画像を用いる。法線画像とは、オブジェクトの法線ベクトルのXYZをRGBに対応させた3チャンネル画像である。また深度画像はカメラからの距離を各ピクセル値とした1チャンネル画像である。この2つの画像、合計4チャンネルを入力データとしてネットワークに与える。

ターゲット画像には、法線画像・深度画像を手描きでトレースした線画像を用いる。今回は以下の条件を満たすようにトレースした線画像を用い、具体例を図1aに示す。

1. 線は必ず法線画像・深度画像のエッジ上に描く
2. 3DCGオブジェクトの形状にない模様などは描かない
3. シルエットラインは8pt、その他の線は5ptのブラシを用いて描く
4. 顎や髪の毛などオブジェクト重なりによる線の交差部分は陰影を表現するためにインクが溜まったように塗りつぶす

学習データが少ないと学習が十分に行われないため、

A method to synthesize nonuniform line-drawings from 3DCG with CNNs

[†] Kosuke MAEDA

^{††} Suguru SAITO

Tokyo Institute of Technology Faculty of Engineering ([†])

Tokyo Institute of Technology School of Computing (^{††})

画像の回転・反転を行いデータの増強をする。ターゲット画像はバイキュービック補間法で回転し、入力画像は3Dオブジェクトを回転しレンダリングする。それらの画像を上下左右に反転する。ただし、法線画像の反転は反転させる方向のチャンネルが対応するように反転を行う。そして、パッチを作成を行いこれを学習に用いる。パッチ中心がオブジェクトの領域となるよう画像をランダムに切り抜き、その領域をパッチとする。

2.2 学習モデル

学習ネットワークには、画像・映像認識に広く用いられている畳み込みニューラルネットワークを用いる。ネットワークの構成はSimo-Serraらの研究[1]で用いられた全層畳み込みニューラルネットワークモデルを参考に示す。構成したネットワーク構造を表1に示す。

kernel	stride	output size	activation
-	-	$4 \times W \times H$	-
3×3	2×2	$64 \times W/2 \times H/2$	ReLU
3×3	1×1	$128 \times W/2 \times H/2$	ReLU
3×3	1×1	$128 \times W/2 \times H/2$	ReLU
3×3	2×2	$256 \times W/4 \times H/4$	ReLU
3×3	1×1	$512 \times W/4 \times H/4$	ReLU
3×3	1×1	$512 \times W/4 \times H/4$	ReLU
3×3	1×1	$1024 \times W/4 \times H/4$	ReLU
3×3	1×1	$1024 \times W/4 \times H/4$	ReLU
3×3	1×1	$512 \times W/4 \times H/4$	ReLU
3×3	$\frac{1}{2} \times \frac{1}{2}$	$512 \times W/2 \times H/2$	ReLU
3×3	1×1	$256 \times W/2 \times H/2$	ReLU
3×3	1×1	$256 \times W/2 \times H/2$	ReLU
3×3	$\frac{1}{2} \times \frac{1}{2}$	$256 \times W \times H$	ReLU
3×3	1×1	$128 \times W \times H$	ReLU
3×3	1×1	$128 \times W \times H$	ReLU
3×3	1×1	$32 \times W \times H$	ReLU
3×3	1×1	$1 \times W \times H$	Sigmoid

表1: 学習モデルのネットワーク構成

入力層・出力層を除き、各層にはバッチノーマライゼーション[2]を適用する。最終層以外には活性化関数としてReLUを用いているが、最終層では出力を[0,1]にするためにSigmoidを用いる。

損失関数には式(1)を用い、これを最小化するようにネットワークを学習させる。

$$\|T - O\|_{\text{FRO}}^2 \quad (1)$$

ただし、 T はターゲット画像、 O はネットワークの出力とし、 $\|\cdot\|_{FRO}$ はフロベニウスノルムとする。

3 実験と考察

第2章で述べたように、ネットワークを実装する。3Dオブジェクトには、プロ生ちゃん[3]を用い、様々なポーズをとった14セットの学習データを作成し、このうち、12セットを学習データとして用い、残り2セットをテストデータとする。パッチサイズを240×240pxとし、1回の学習に10セットのパッチを与え、30000回繰り返し学習を行う。

まず、線描画について定量的評価を行う。画素値を $[0, 1]$ とした時の $[0, 0.9]$ を線部分とし、オブジェクトすべてが含まれるような最小の矩形領域において、ターゲット画像に対する出力画像の正答率を求め、それを表2に示す。

データ	線部分	白地部分
学習データ	95.85%	98.31%
テストデータ	90.40%	98.17%

表 2: ターゲット画像に対する正答率

次に、手描き線画像の特徴である線の強弱について定性的評価を行う。テストデータより生成した線画像を図1に示す。ターゲット画像作成時の条件3にあるような線幅の変化による強弱がついており、条件4にあるようなインク溜まりの表現も顎・髪などの交差部分が塗りつぶされていることが分かる。

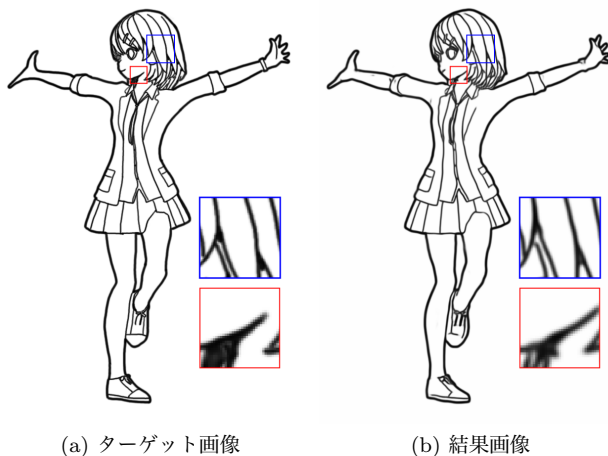


図 1: テストターゲット画像 (a) と学習済みネットワークによる出力画像 (b) の比較

動画へ適応では、フレーム間の連続性が求められる。図2は顎の線が徐々に太くなる。連続性が好ましく実現できた例である。しかし本手法では、各フレー

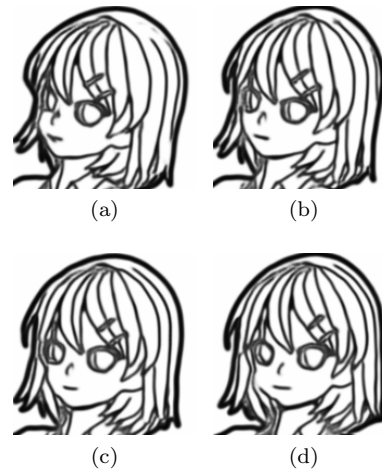


図 2: フレーム間における連続性。(a)→(d)の順。

ムごとに線画像を生成するため、フレーム間の連続性は本質的に考慮されない。入力である法線画像・深度画像におけるエッジが明瞭でない箇所では線の描画の判別がつかないことがあり、ちらつきの発生が見られる結果も生じてしまった。

4 まとめ

3DCGより生成した法線画像と深度画像を入力とし、畳み込みニューラルネットワークを用いることで、強弱有り線画像を生成する手法を提案した。実験を行い、提案手法が手描き線画像の強弱量の決定を行うことを示した。今後の課題としては、動画への適応や、様々な3Dオブジェクトに対する頑健性の評価・向上などが挙げられる。また、この研究の発展として、提案した手法により求めた強弱量を用い、画材で描いたようなスタイルを線に付与することなどが期待される。

参考文献

[1] Edgar Simo-Serra, Satoshi Iizuka, Kazuma Sasaki, and Hiroshi Ishikawa. Learning to Simplify: Fully Convolutional Networks for Rough Sketch Cleanup. *ACM Transactions on Graphics (SIGGRAPH)*, Vol. 35, No. 4, 2016.

[2] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *CoRR*, Vol. abs/1502.03167, , 2015.

[3] <http://3d.nicovideo.jp/works/td8608>.