

AI による 2D 画像から 3D モデルポリゴン数の予測

柏原 直志† 三上 浩司† 藤澤 公也†

東京工科大学メディア学部†

1. はじめに

3D モデル作成を請け負う会社があり、その会社が 3D モデル作成を受注する際には作成を依頼してきた会社に対して 3D モデル作成にどの程度のコストが発生するのかを伝える必要が出てくる。その時、コストを計算する指標となるものが受注時に見るキャラクターのラフ画と作業時間を表す工数である。工数は人の経験に左右されるためその場でラフ画を見て工数を見積もることは困難である。

そこで本研究では AI を利用して工数の見積もりを行うための第一段階として 2D の画像から 3D モデルのポリゴン数の予測を行った。

2. 3D モデルのポリゴン数予測

2.1 ポリゴン数予測の方法

ポリゴン数の予測には 2 つの工程を踏むことにした。1 つ目の工程はキャラクターの部位を分割すること、2 つ目の工程は分割された部位それぞれでポリゴン数を予測することである。

1 つ目の部位の分割は U-net [1] を利用した。U-net は領域分割を行う Semantic Segmentation の手法の一つである。2015 の ISBI で優秀な成績を収めていることを理由に採用した。

2 つ目のポリゴン数予測は画像認識において CNN が高い効果を発揮していることから CNN を利用した。

2.2 U-net による部位の分割

2.2.1 データセットの作成

セグメンテーションの学習に使用したデータは 2D 画像とセグメンテーションされた画像の 2 つである。入力データを元の画像、出力データをセグメンテーションされた画像としてデータセットを作る。

データセットに使用した画像の例が図 2-1 である。図の左が 2D 画像、右がセグメンテーションされた画像となっている。元の画像は 3D モデルをレンダリングした画像を使用した。セグメンテーションされた画像は本来グレースケールの画像であるが色の違いがわかりやすいように加工している。セグメンテーションで分割している部

位は顔、髪、服、肌、スカート、靴、靴下、背景、その他の 9 つである。

前処理として入力画像、出力画像ともに 128x128 のサイズにリサイズを行った。



図 2-1 データセットに使用した画像例

このような画像を 100 枚用意し、80 枚を学習データ、20 枚を検証データとした。

2.2.2 ニューラルネットワークの構造

図 2-2 は U-net の構造を示したものである。図中の Down Sampling, Up Sampling は図の下部にある内容を省略した書き方を表している。

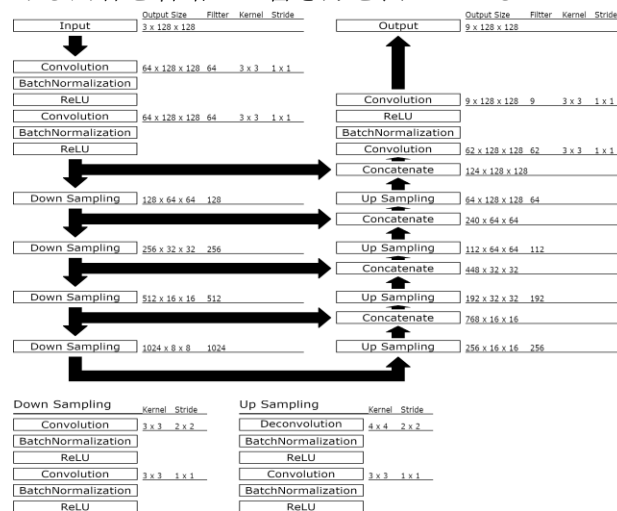


図 2-2 U-net の構造

ダウンサンプリングを行う畳み込み層とアップサンプリングを行う逆畳み込み層を入れたネットワークである。逆畳み込み層への入力としてはダウンサンプリング時のデータも使われる。

2.3 CNN によるポリゴン数予測

2.4 データセットの作成

ポリゴン数の予測に使用したデータは各部位の画像とポリゴン数である。ポリゴン数の予測を行うための学習は各部位それぞれで行うため背景を除く 8 つのデータセットを作成した。

Predicting the number of 3D model polygons from 2D images by AI

†Kashiwabara Naoshi, Mikami Koji, Fujisawa Kimiya, School of media science, Tokyo University of Technology

各部位の画像とポリゴン数にはそれぞれ前処理を行った。図 2-3の左側はセグメンテーションされた画像をもとに 2D 画像から肌の部位を切り抜いたものである。この部位の画像から前処理として余白をなくし、部位を変形させないようにパディングしつつ 128x128 にリサイズしたものが図の右側の画像になっている。この前処理はリサイズした際に少しでも部位が占める画像の面積が増えるようにするために行った。

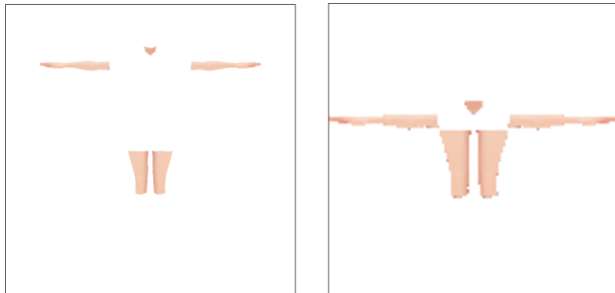


図 2-3 部位の画像(左)と前処理後の画像(右)

ポリゴン数の前処理は標準化を行った。これははずれ値の影響を受けすぎないようにするためである。

データセットは前処理を行った部位画像を入力、標準化したポリゴン数を出力とした。

2.4.1 ニューラルネットワークの構造

図 2-4はポリゴン数の予測のために使用したネットワークの構造を示したものである。2 層の畳み込み層と 1 層のプーリング層を 1 組としたものを複数回繰り返し最後に 2 層の全結合層を行うものとなっている。



図 2-4 CNN の構造

3. 学習結果と考察

3.1 U-netによる部位の分割

学習はエポック数 700, バッチサイズ 3 で行った。学習後検証データを予測した結果が図 3-1である。また、各予測画像で IoU を計算した結果が表 3-1である。すべての部位で IoU が 0.5 を超えていることがわかる。PASCAL VOC Challenge では

0.5 以上を正解としているためそれを基準に考えると予測画像は正しくセグメンテーションされていると言える。

表 3-1 各部位の IoU

	顔	髪	服	肌	スカート	靴	靴下	背景	その他
IoU	0.551	0.785	0.754	0.866	0.804	0.822	0.868	0.992	0.699

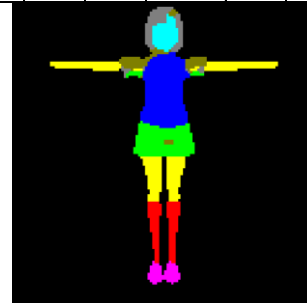


図 3-1 検証データ予測結果例

3.2 CNNによるポリゴン数予測

学習はエポック数 200, バッチサイズ 20 で行った。学習後検証データで予測を行った結果をまとめたものが表 3-2である。相関係数は正しい値と予測値の相関係数を表し、倍率はもとの値に比べ予測値は何倍となっているかを表しており(平)は平均を, (中)は中央値を表している。

表 3-2 ポリゴン数予測結果

	顔	髪	服	肌	スカート	靴	靴下	その他
相関係数	0.901	0.827	0.969	0.829	0.997	0.944	0.719	0.922
倍率(平)	1.103	0.930	1.047	1.091	1.022	1.004	1.477	3.563
倍率(中)	1.008	0.963	0.980	1.038	1.002	1.014	1.061	1.354

相関係数はすべての部位で 0.7 を超えていることから強い正の相関を持っていることがわかる。平均倍率を見るとその他のみ他の値と比べ高い中央値と比較し差があることから一部の予測値が平均を上げていることが考えられる。

4. おわりに

学習結果から AI による部位の分割, ポリゴン数の予測は行えることがわかった。また、記載していないが 2 つの AI を組み合わせると 2D 画像を入力として与え、出力がポリゴン数となるようなシステムの作成も行った。

本研究の手法はポリゴン数を工数に変更しても適用できると考えられるので工数のデータを使つての学習を目指したい。

参考文献

[1] Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham.