

# 深層学習を用いた要約文による三次元モデルの生成に関する研究

情報科学科 清水 孝哉

指導教員：小林 邦和

## 1 はじめに

三次元モデルは近年、製品設計・シミュレーション・映画など様々な用途に活用され需要が高まっているが、製作には専門知識が必要とされる。そこで、本研究では三次元モデルを要約文から生成することで、技術の無い人でも容易に利用できるようにすることを目的とする。

## 2 提案手法

要約文を元にした画像生成モデルと画像を元にした三次元モデル生成モデルを、要約文のみの入力で三次元モデルを製作するという目的で組み合わせる。三次元モデルとして生成したい対象物の特徴を表す要約文を画像生成モデルに入力する。その後、生成された画像を三次元モデル生成モデルに入力し、対象物の三次元モデリングを実現させるといった方針である。

### 2.1 要約文を元にした画像生成

要約文から画像生成を行うモデルとして、Attentional Generative Adversarial Networks(以下 AttnGAN)[1]を用いる。AttnGANとはTao Xuらによって提案されたキャプションから画像を生成するモデルであり、対象物の細かい部分まで再現可能なText-to-Imageの生成モデルとなっている。

### 2.2 画像を元にした三次元モデル生成

画像から三次元モデル生成を行うモデルとしてCategory-Specific Mesh Reconstruction(以下CMR)[2]を用いる。CMRは一枚の画像から物体の3D形状・カメラポジション・テクスチャを生成するための学習モデルである。生成される3Dモデルは球体メッシュを変形させていくことで生成しているが、このCMRの大きな特徴として、学習時に対象物を限定することでより高精度な3Dモデルの生成が可能となっている。

## 3 計算機シミュレーション

生成された三次元モデルの要約文からの再現度を測定するためにシミュレーションを行う。

### 3.1 使用データセット

本シミュレーションで用いるデータセットを表1に示す。  
表1 データセット

目的	名称
画像生成	Calech-UCSD Birds-200-2011[3] Microsoft COCODataset[4]
3Dモデル生成	Calech-UCSD Birds-200-2011

### 3.2 シミュレーション方法

次に、鳥の画像、その画像を見た状態で特徴を表現する文を生成する。文章を入力として三次元モデルを生成した後、モデルの画像と鳥の画像についてヒストグラム比較を行い、文章からの生成の再現度を評価する。また、比較対照として、モデルの画像と別種の鳥の画像についてもヒストグラム比較を行う。

### 3.3 シミュレーション結果

本シミュレーションで実際に得られた三次元モデルの一例を以下に示す。図1,2は使用した鳥の画像、図4は図1から生成した要約文、図3は要約文から生成された三次元モデルである。ま

た、表2は比較結果である。一致度の値は-1~1であり、大きいほど類似度が高いことを示す。

表2より、約0.79と高い一致度が得られた。しかし、図3の生成された三次元モデルの画像を見ると、全体的なフォルムや大まかな色は再現できているが、細かい部分、特にくちばしや足の先、体の模様は再現できていない。改善策として、三次元モデルの頂点数を増加させるといった方法があげられる。



図1 対象の鳥の画像



図2 別種の鳥の画像



図3 生成された三次元モデルの画像

round body, white body, long beak, black head

図4 図1に対する要約文

表2 類似度の比較結果

	ヒストグラム比較による一致度 (-1~1)
対象の鳥画像	0.79
別種の鳥画像	-0.073

## 4 おわりに

今後の課題としては、要約文に忠実な三次元モデルを生成できるよう、CMRの入力として適した画像を生成できる画像生成モデルの選択、頂点数の増加の実現があげられる。また、データセットには対象物と同じカテゴリの画像を用いており、三次元モデルデータは不要であるため鳥以外の三次元モデルの生成にも容易に対応可能である。そのため、今後の展望としては車・家具・家といった製品設計の際に活用などへとつなげたい。

## 参考文献

- [1] T.Xu, P.Zhang, Q.Huang, et al. "AttnGAN: FineGrained Text to Image Generation with Attentional Generative Adversarial Networks", CVPR, 2018.
- [2] A.Kanazawa, S.Tulsiani, A.Efros, J.Malik, "Learning Category-Specific Mesh Reconstruction from Image Collections", ECCV, 2018
- [3] C.Wah, S.Branson, P.Welinder, P.Perona, S.Belongie, "The Caltech-UCSD Birds-200-2011 Dataset", California Institute of Technology, CNSTR-2011-001, 2011.
- [4] T.Lin, M.Maire, S.Belongie et al. "Microsoft COCO: Common Objects in Context", ECCV, 2014.