

物体を表現した語句に対するボクセルデータの生成

Generating voxel data from words that capture the features of the object

駒澤大夢[†]

Hiromu KOMAZAWA

松澤智史[‡]

Tomofumi MATUZAWA

1. 研究背景

近年は、ゲームやCGアニメーションの他に、AR・VR技術や3Dプリンター等の先端技術の進化に伴い、3Dモデルの需要は増加傾向にある。また、生成ネットワークモデルを扱った3Dモデルの生成は発展途上であるが、注目されている分野である。3Dモデルの自動生成の実現は、他分野に渡って開発の幅を広げる可能性がある。しかし、行われている研究の多くは入力データを乱数や画像としたものであり、実用する上で汎用性に欠けている。

2. 研究目的

本研究は、生成ネットワークモデルを用いて、物体を表現した語句から3Dモデルを生成するシステムを提案する。例えば、「propeller plane (プロペラのついた飛行機)」や「big couch (大きなソファ)」という語句に対して、その語句の特徴を捉えた3Dモデルが自動で生成されるといったシステムである。以上のシステムを使用することで、容易に語句と整合性のとれた3Dモデルを生成することを目的としている。

3. 基礎知識

3.1 ボクセル

ピクセルとはデジタル画像における最小単位である。ボクセルとはピクセルに奥行の情報を持たせたものであり、3Dデータにおける立方体の最小単位である。

3.2 GAN

GAN (敵対的生成ネットワーク) とは、教師なし学習に分類される機械学習手法の一つである。以下の2つのニューラルネットワークモデルが互いに競って学習することで敵対的生成ネットワークと呼ばれている。

- ジェネレータ：画像などの対象物を生成して、一方の学習モデルであるディスクリミネータを騙そうと生成を行う生成モデル
- ディスクリミネータ：本物データと比較して、ジェネレータの生成物が正しいかどうかを判断する識別モデル

3.3 畳み込み層

畳み込み層は、画像処理などの学習モデルでよく使われる層の一つである。多次元のデータを扱う際に、各数値の位置情報を損なわずに特徴を抽出することができる。

3.4 転置畳み込み層

転置畳み込み層は、畳み込み層を逆方向から乗算するような形の層であり、特徴データをアップサンプリングすることができる。

3.5 KNN

KNNは、教師あり機械学習の一つである。分類や回帰に使用される。学習データの分布において、近くのK個データのうち、最も一般的なクラスに分類する。

4. 関連研究

4.1 3DGAN

3DGANはJiajun Wuらによって提唱された、乱数からボクセルを生成するGANモデルである(図1)。

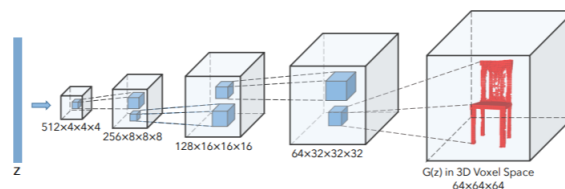


図1: 3DGANのジェネレータ構成図(ディスクリミネータは反転した図)

ジェネレータの入力データは、200次元の区間[0,1]の一律乱数である。ジェネレータは、5層の転置畳み込み層を用いており、ディスクリミネータは、ジェネレータを反転したような、5層の畳み込み層を用いている。

4.2 GAN-CLS

GAN-CLSは、文章から画像を生成するGANモデルである(図2)。

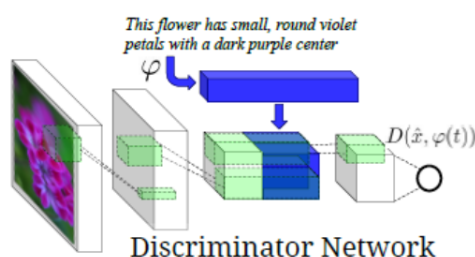


図2: GAN-CLSのディスクリミネータの構成図

特徴は、テキストと生成物の整合性を図るためにディスクリミネータがテキストと生成物の両方を受け取る点が挙げられる。ディスクリミネータが(本物の画像、本物の文章)を本物、(生成画像、本物の文章)と(本物の画像、偽の文章)を偽物として学習を行う。

[†]東京理科大学 理工学研究科 情報科学専攻

[‡]東京理科大学 理工学部 情報科学科

5. 提案手法

5.1 概要

関連研究を参考に、語句を入力してボクセル形式の3Dモデルを出力するシステムを提案する。

5.2 GANの構成

本研究では、関連研究を参考に図のGANモデルを作成した(図3)。

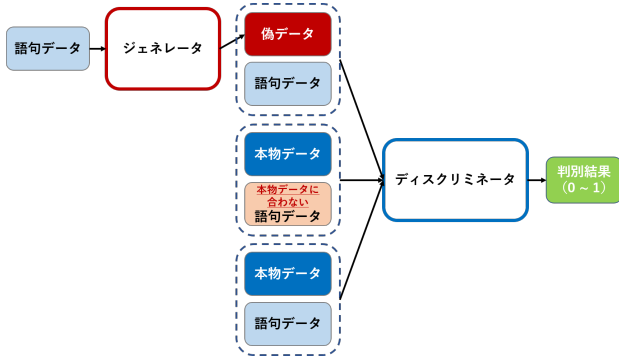


図3: GAN-CLSのディスクリミネータの構成図

5.3 語句のベクトル化

語句を品詞分解した後、Gloveと呼ばれる手法で各単語の分散表現を取得した。最終的に、名詞を100次元、形容詞・分詞を50次元を結合した150次元のベクトルとした。

5.4 ジェネレータ

1層の全結合層と4層の転置畳み込み層を用いた。200次元の語句データとノイズを結合したベクトルを入力して、 $32 \times 32 \times 32$ の3Dモデルを出力するモデルとした。

5.5 ディスクリミネータ

4層の畳み込み層と3層の全結合層を用いた(図4)。またGAN-CLSの工夫を取り入れた。まず $32 \times 32 \times 32$ の3Dモデルを受け取り、中間の全結合層で語句データを統合して、判定結果を出力するモデルとした。

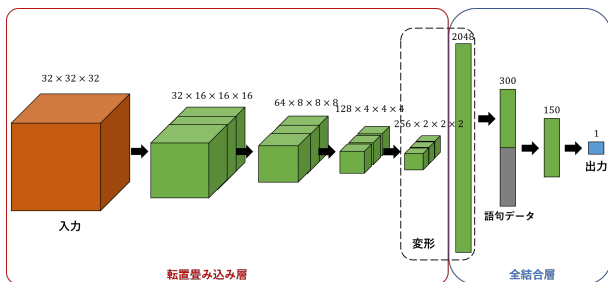


図4: ディスクリミネータの構成図

6. 実験

6.1 実験データ

データセットであるShapeNetCore.v1の、“飛行機”、“車”、“銃”、“ソファ”、“モニター”、“テーブル”、“ギター”の7クラスで学習を行った。

各データのキャプションを語句とし、3Dデータはボクセル形式に変換した。

6.2 実験結果

学習を行い、データセットに含まれる語句を入力した生成物を示す(表1)。

表1: 生成物の例

propeller airplane	
--------------------	---

7. 評価と考察

7.1 整合性の評価

分類器を使用して整合性の評価を行う。先行研究と同様に、ディスクリミネータの中間層の出力を比較する。分類器はKNN($k=6$)を使用した。分類器の訓練データは実験の7クラスと、学習に使用していない“椅子”クラス、“帽子”クラスの計9クラスので分類を行った(表2)。

表2: 生成物に対するクラス分類精度

クラス	分類精度 (%)
airplane	99.0
car	99.0
gun	100.0
coach	86.0
screen	89.0
table	98.0
guitar	97.0
chair	74.0
cap	0.0

学習に使用した7クラスの平均分類精度は95.4%であり、上手く整合性が取れている。また、“椅子”クラスは比較的整合性が取れていることが分かる。

7.2 考察

整合性の評価より、“ソファ”クラスと比較的意味が近い“椅子”クラスの精度が高くなったことから、学習に使用するクラス数を増やすことで色々な語句に対応することができる。

また今後の課題として、Attention機構などを実装することでさらに生成物の外見の精度の向上を目指す。

参考文献

- [1] 3DGAN: Jiajun Wu, Chengkai Zhang, Tianfan Xue, William T. Freeman, Joshua B. Tenenbaum (2017) “Learning a Probabilistic Latent Space of Object Shapes via 3D Generative-Adversarial Modeling” In NIPS.20
- [2] GAN-CLS: Martin Arjovsky, Soumith Chintala, Leon Bottou (2017) “Wasserstein GAN” In CoRR.2017