

令和4年度 特別研究報告書

敵対的生成ネットワークを用いた
点群データ生成の検討と改善

龍谷大学 理工学部 情報メディア学科

T190532 福寫和樹

指導教員 三好力 教授

内容梗概

近年、深層学習を用いた生成モデルは、画像やテキストなどの多様なタイプのデータを生成することができるようになった。その一方で、点群データを対象とした生成モデルは、工業向けのソフトなどに限られており、今なお発展途上であるため、点群データを効率よく生成する手法の開発は、将来的に重要であると考えられる。

本研究では、敵対的生成ネットワークと呼ばれる技術を応用し、より精巧な点群 3D データを生成するニューラルネットワークを提案することを目的とする。

また、敵対的生成ネットワークを用いた点群データの生成について、先行研究に対して、生成ネットワークの構造を変化させたものを複数比較することで、より良い生成ネットワークの検討と改善を行う。

目次

第1章. はじめに	1
第2章. 既存技術と先行研究.....	2
2.1 GAN.....	2
2.2 3D-GAN.....	3
2.3 PointNet	4
2.4 先行研究“敵対的生成ネットワークを用いた点群データ生成の検討,”[5]	5
第3章. 問題点・改善手法	7
第4章. 実験	8
4.1 実験環境	8
4.2 実験概要	8
4.3 実験内容	8
4.4 実験結果	13
第5章. 考察	14
第6章. 結論	15
第7章. 参考文献	17

第1章. はじめに

近年、深層学習を用いた生成モデルは、画像やテキストなどの多様なタイプのデータを生成することができるようになった。その一方で、点群データを対象とした生成モデルは、工業向けのソフトなどに限られており、今なお発展途上である。点群データは、3D モデリングや設計、計算機グラフィクスなど、様々な分野で用いられるため、点群データを効率よく生成する手法の開発は、将来的に重要であると考えられる。

また、近年の仮想現実(VR)や拡張現実(AR)技術の発展に加えて、メタバースといったコンピューター空間と現実の融合を目指す概念の普及によって、質の高い 3D データの需要が高くなると考えている。教師あり学習であるディープニューラルネットワーク(DNN)で 3D データを制作しようとする多くのラベル付きデータが必要になるが、点群データは、高次元で複雑な構造を持つため、生成することが困難である。そのため、近年の点群データ生成においては、深層学習を用いた手法が注目されている。本研究では、より精巧な 3D データを作るために画像生成などに用いられる Generative Adversarial Networks[1](GAN, 敵対的生成ネットワーク)と呼ばれる技術を応用し、より精巧な点群 3D データを生成するニューラルネットワークを提案することを目的とする。また、敵対的生成ネットワークを用いた点群データの生成について生成ネットワークの構造を比較検討することで、今後の点群データ生成における研究の可能性を探る。

第2章. 既存技術と先行研究

2.1 GAN

GAN (Generative Adversarial Networks, 敵対的生成ネットワーク)[1]はニューラルネットワークの 1 種であり、学習したデータから新しいデータを生成することができる。GAN は、2 つのニューラルネットワークを組み合わせしており、1 つは Generator(生成器)、もう 1 つは Discriminator(識別器)である。Generator は、入力された乱数やノイズから画像や音声など、人間が識別できるデータを生成することを目的としている。一方で、Discriminator は、生成器が生成したデータと本物のデータを見分けることを目的としている。GAN を学習させる際には、Generator と Discriminator を交互に学習させる。Generator は、Discriminator を騙すように、できるだけ本物に近いデータを生成する。一方で、Discriminator は、Generator から出力されるデータを見分ける能力を高めるようにする。図 1 のように、Generator と Discriminator は相互に学習し合い、学習を繰り返すことで最終的には生成器は本物に見えるデータを生成するようになる。

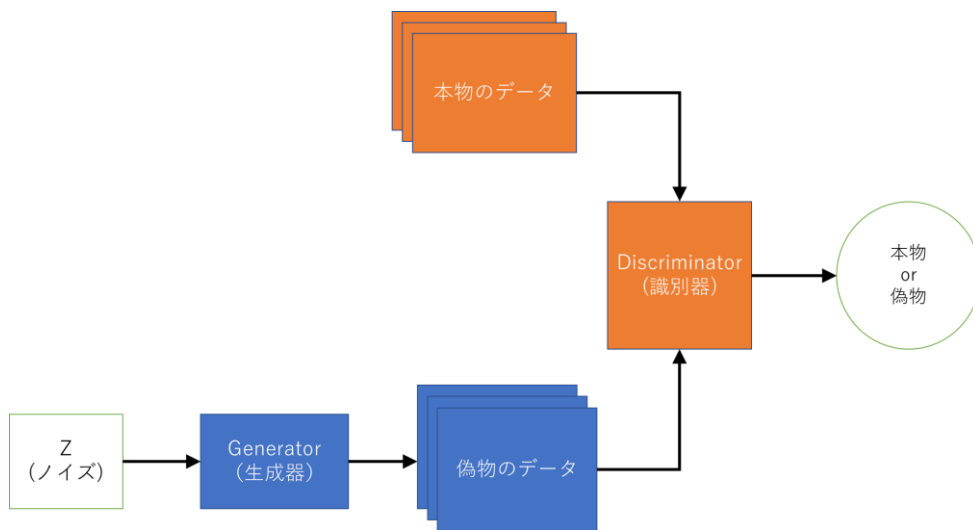


図 1. GAN の学習サイクル

Generator の損失関数は式 2.1 のように表される。

$$L_G = \frac{1}{m} \sum_{i=1}^m \log [1 - D(G(z))] \quad (2.1)$$

ここで、 m はバッチ数、 D は Discriminator、 G は Generator、 z は入力ノイズである。Generator が生成した偽データ $G(z)$ を Discriminator が識別したとき、本物(1)と判定されると最小になる。Discriminator のパラメーターは固定して学習する。

Discriminator の損失関数を式 2.2 に示す。

$$L_D = \frac{1}{m} \sum_{i=0}^m \langle \log D(x_i) + \log \{ 1 - D[G(z_i)] \} \rangle \quad (2.2)$$

x は本物のデータを指す。Discriminator はこの損失関数の最大化を目指して学習する。これによって Generator と Discriminator が敵対関係となる。 $\log D(x_i)$ は本物のデータに対して 1 を、 $\log(1-D(G(z_i)))$ は偽物のデータに対して 0 を出力すると最大となる。

2.2 3D-GAN

3D-GAN は 2 次元畳み込みをする画像ニューラルネットワークを 3 次元に拡張した GAN の派生手法の一種である。図 2 に 3D-GAN の学習過程を示す。3D-GAN は入力ノイズを 3 次元空間にあてはめ、3 次元畳み込みとサンプリングを繰り返し、最終的にボクセルのある確率を分割された空間ごとに出力する

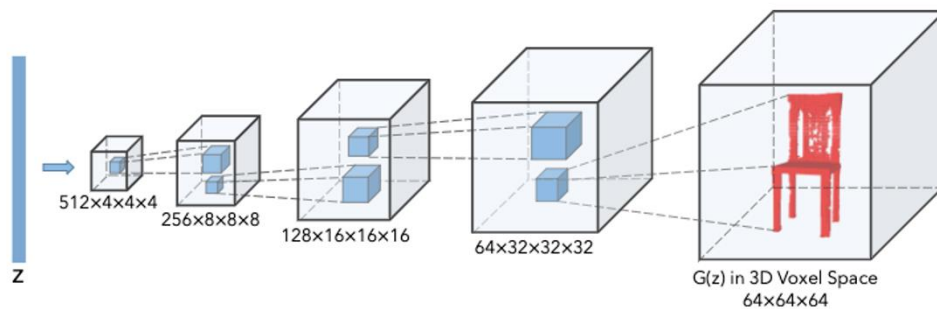


図 2 3D-GAN の学習過程([2]より引用)

2.3 PointNet

PointNet は点群データを扱うために開発された深層学習フレームワークである。PointNet は点群データをそのまま入力として、3D 形状や構造を抽出することができるという特徴がある。従来のニューラルネットワークは、画像や文章などのように、グリッド状にデータが並んだものを入力としていたが、PointNet は、そのような構造を持たない点群データを入力とすることができるため、点群データを処理することができる。これが行えるのは、PointNet が MaxPooling (「Max Pooling は非常にシンプルな関数で、入力要素の内、最大の要素を出力とする関数」[3]) によって、データの順不同性が確保しているからである。

PointNet の構造を図 3 に示す。

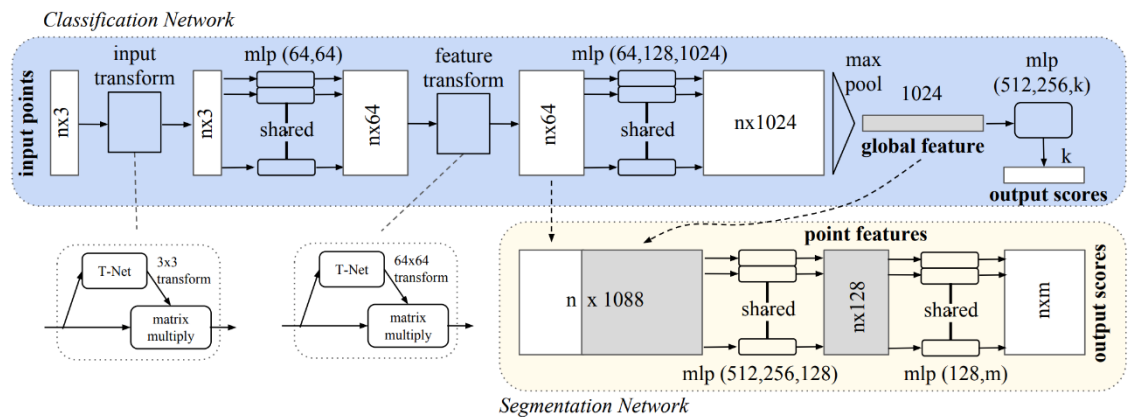


図 3 PointNet の構造 ([4]より引用)

「PointNet はネットワークの最後の層で Max Pooling を用いることで順不変性を獲得しています。」[3]そのため、点群データのカテゴリ分類、物体のパーツごとの分割などに活用されている。[PointNet の特徴は後半部分や transform 層にみられる $N \times X$ の入力を X まで減らす部分である。ここには maxpooling や average pooling などが用いられ、これらは入力の 1 次元目(頂点数 N の部分)の順序に影響を受けない。これによって順不変である点群データについて順序を入れ替えても結果が等しくなるようになっている。この特徴から引用元[3]では symmetric function と呼称されている。][5]

2.4 先行研究“敵対的生成ネットワークを用いた点群データ生成の検討,”[5]

ModelNet10 の 3D 学習データセットをもとに、PointNet の構造を使って、3DGAN よりもデータサイズを削減し学習を行って、多層パーセプトロンのみで構成された比較用モデルとの比較を行い、PointNet の構造を使った敵対的生成ネットワークを用いた点群データ生成を試みている。図 4 に先行研究の識別器の構造。図 5 に生成器の構造を示す。N は頂点数、N の隣は分類に使う観点数、mlp は多層パーセプトロンを表す。transform は入力(N, x) (x は特徴量の数積を出力とするネットワークである。

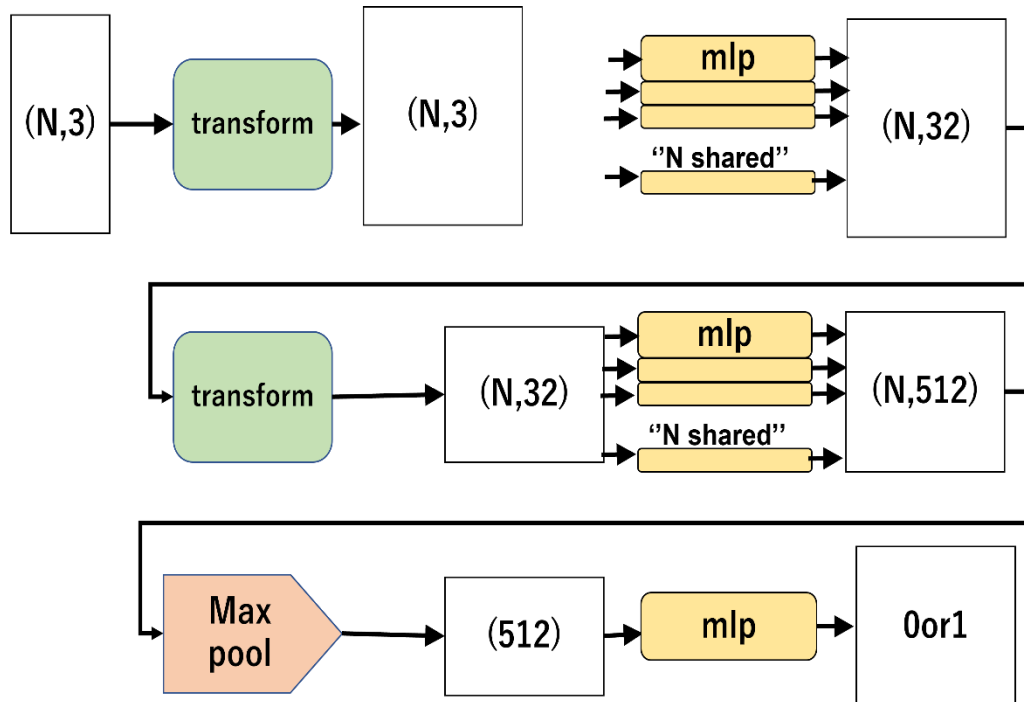


図 4 先行研究の識別器の構造 ([5]より作成)

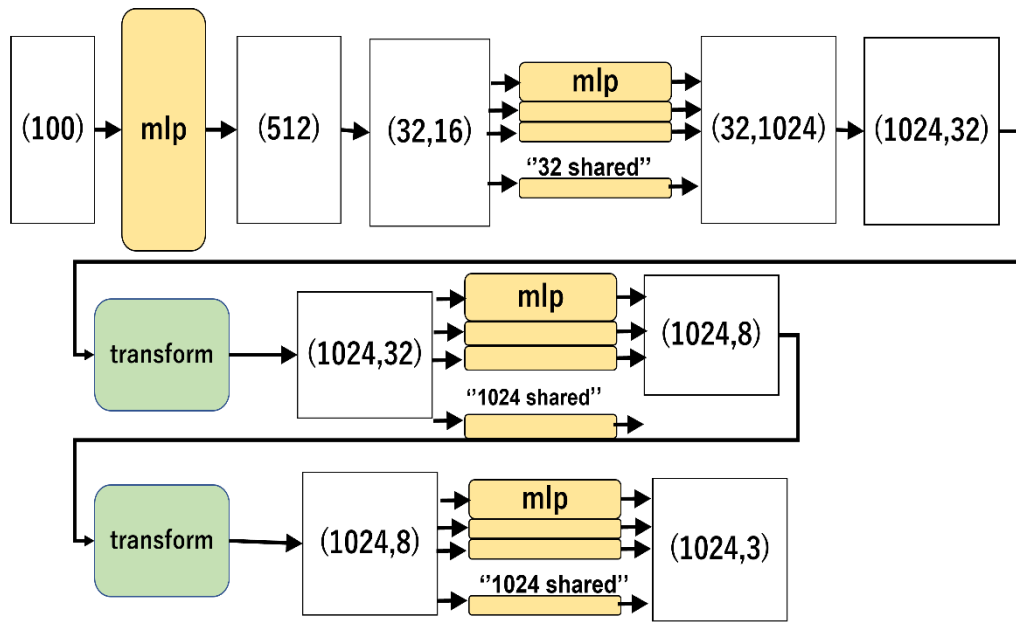


図 5 先行研究の生成器の構造 ([5]より作成)

第3章. 問題点・改善手法

先行研究の問題点として定量的評価の FID スコア (Fréchet Inception Distance) [6] 良いものの、実用的な 3D データ形状ではない点があげられる。そのため、生成器の構造を変化させた生成モデルを複数つくり、FID スコアを比較することでより良い生成モデルの構造を検討する。

FID スコアは、生成された画像の画像空間における分布と、実際の画像の画像空間における分布の距離を表し、FID スコアが小さければ、生成された画像と実際の画像が似ているといえる。FID スコアを計算するには、生成された画像と実際の画像を Inception v3 モデルを使用して特徴量抽出し、特徴量の抽出された分布間の Fréchet 距離を計算する。以下に FID スコアの計算式を示す。

$$FID = \|\mu_r - \mu_f\| + T_r \left(\sum r + \sum f - 2 \left(\sum r \sum f \right)^{\frac{1}{2}} \right)$$

μ_r 、 μ_f はそれぞれ実データと GAN の生成データを分類モデルに入力したときの特徴ベクトルの平均、 r は行列のトレースである。 $\sum r$ 、 $\sum f$ はそれぞれの共分散行列を表す。

第4章. 実験

4.1 実験環境

。本実験では、Python 用機械学習ライブラリの Tensorflow と Keras を用いた。実行に用いた計算機環境を表 4.1 に示す。

OS	Windows 10 (x64)
メモリ	16GB
CPU	AMD Ryzen 4700U with Radeon Graphics 2.00 GHz
Python	3.8.3
Tensorflow	2.2.0

4.2 実験概要

PointNet の構造を用いる先行研究[5]の生成モデルと提案手法による生成モデルの結果を FID スコアによって比較することで、PointNet の構造を利用したより良い点群データの生成手法の提案と改善を行うことが実験目的である。

4.3 実験内容

先行研究と同様に、機械学習用の 3D モデルデータセット ModelNet10[7]を利用し、頂点や面から点群をサンプリングして点群データに変換したものを教師データに用いる。このサンプリングには Python 用 3D データ向けライブラリである trimesh を用いており、全 3D データ共通で 1024 頂点をサンプリングした。学習条件は先行研究と条件をそろえるため、ミニバッチ数 32、エポック数 2000 としてそれぞれの生成モデルに ModelNet10 で分類される 10 カテゴリを学習させる。図 6 に識別器の構造を示す。識別器は先行研究のものと同じ構造のものを使用し、複数の構造を変化させた生成器で点群データの作成を行い、生成した点群データのカテゴリごとの FID スコアを比較した。実験で使用した生成器の構造を以下に示す。各エポックでは識別器を実データで学習、識別器を生成器の出力で学習、生成器を学習の3つがある。識別器の学習では生成器、生成器の学習では識別器のパラメータを固定して片側ずつ学習を進める。

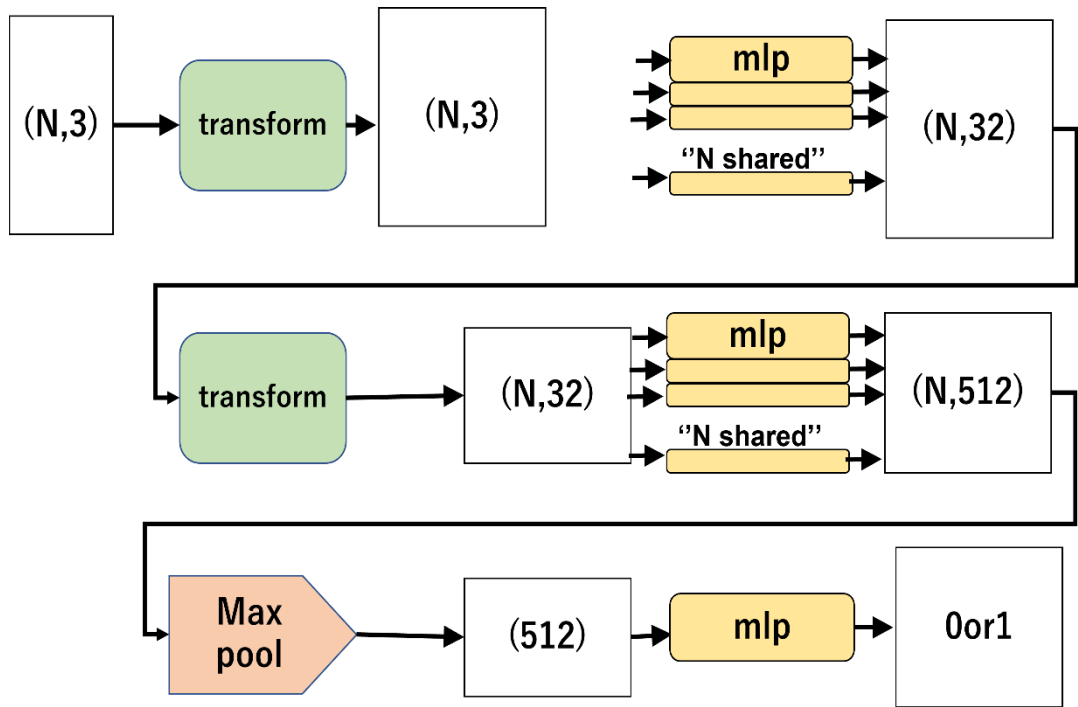


図 6 識別器の構造

生成器の構造ごとに実験番号を付ける。実験 1 の生成器の構造を図 7 に示す。

実験 1 では、はじめに学習する頂点数を元の倍の 64 にし、分類に使う観点数をはじめは多く、徐々に少なくした。

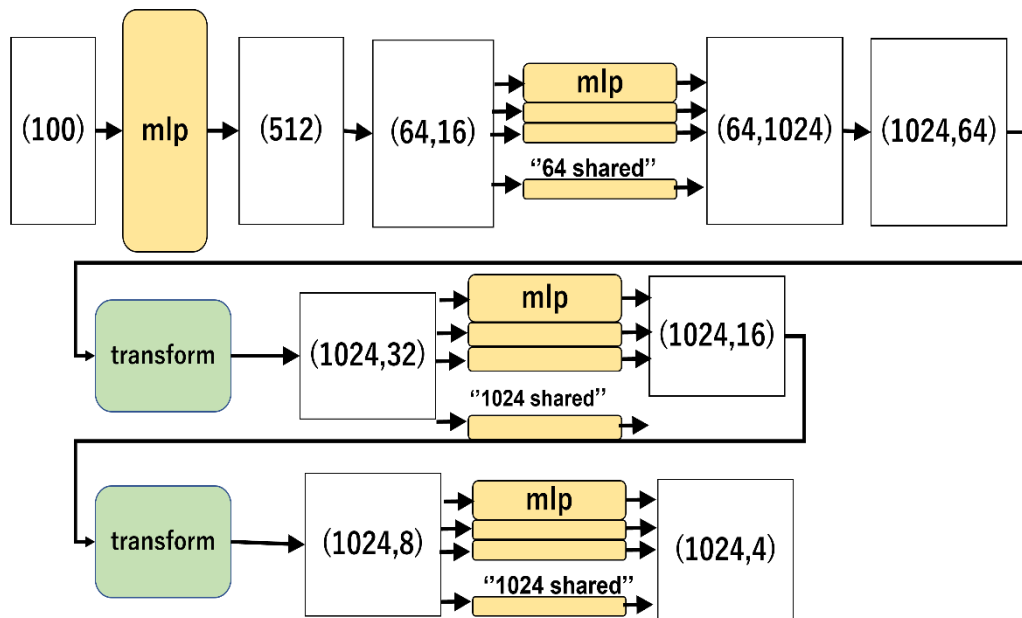


図 7 実験 1 の生成器の構造

実験 2 の生成器の構造を図 8 に示す。

実験 2 では、学習のはじめに頂点数と分類に使う観点数を増やし、徐々に少なくしていった。

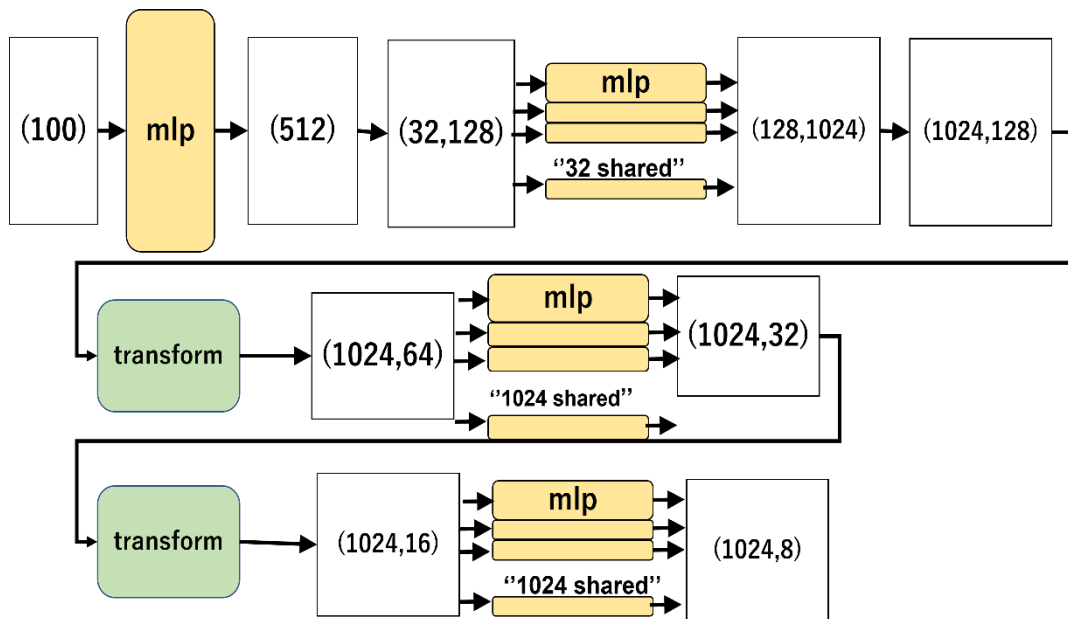


図 8 実験 2 の生成器の構造

実験 3 の生成器の構造を図 9 に示す。
 実験 3 では、学習が進むごとに学習する頂点数と着目する観点数を増やしていった。

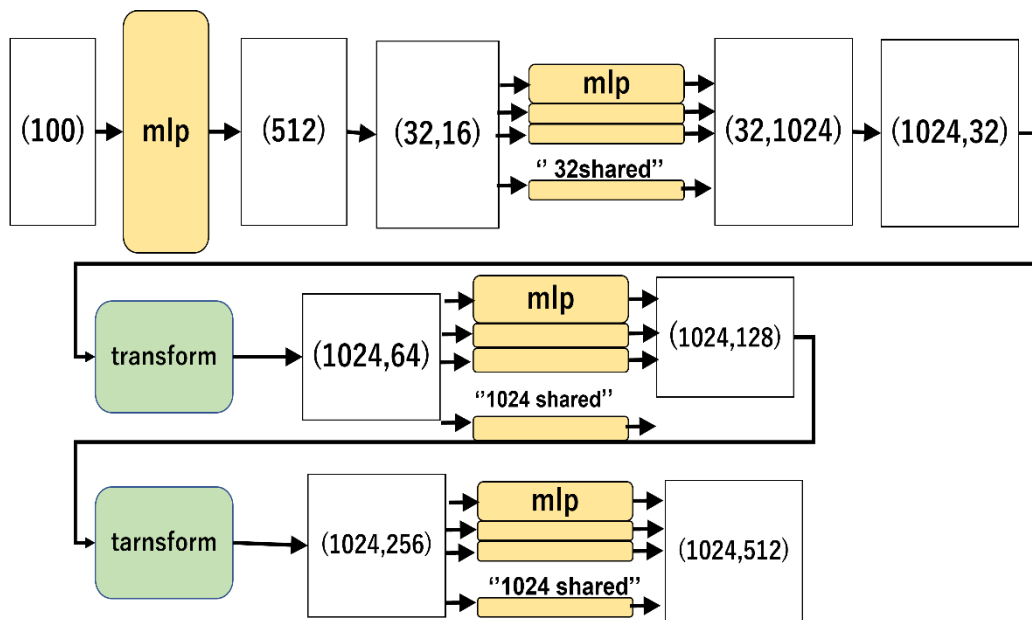


図 9 実験 3 の生成器の構造

実験 4 の生成器の構造を図 10 に示す
 実験 4 では学習する頂点数と着目する観点数を増やした後、減らした。

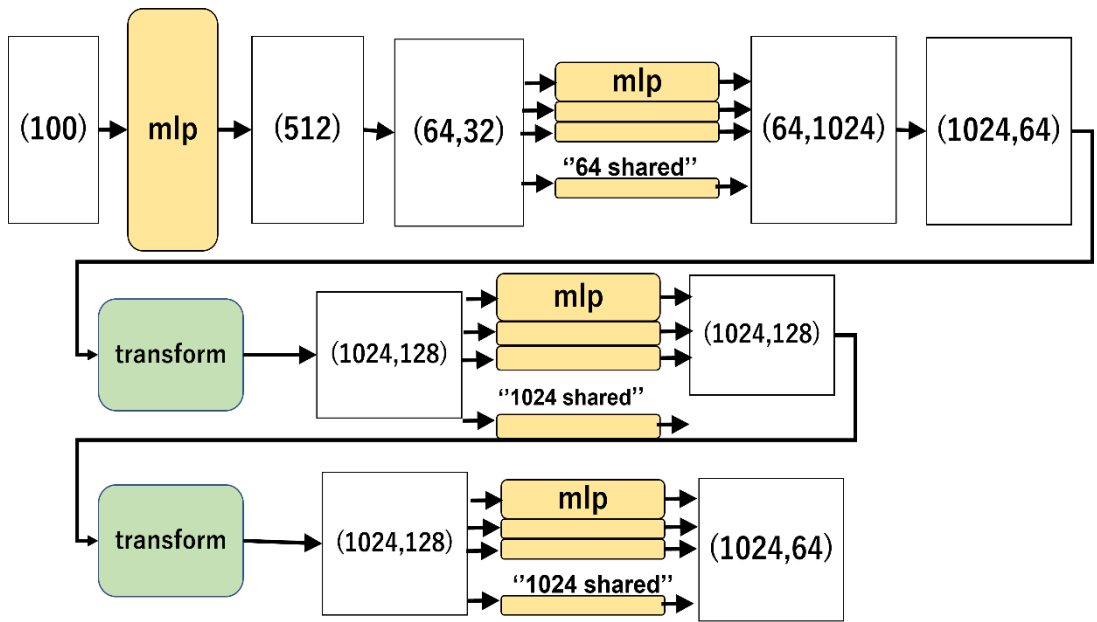


図 10 実験 4 の生成器の構造

実験 5 の生成器の構造を図 11 に示す。

実験 5 では着目する観点数を 128 に増やして、一定にした。

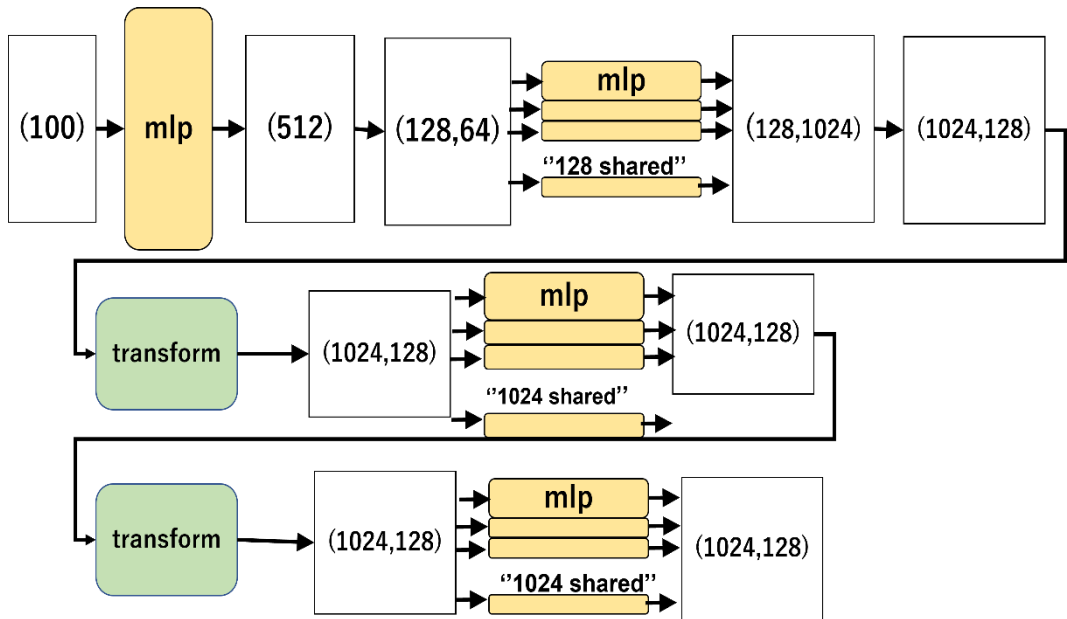


図 11 実験 5 の生成器の構造

実験 6 の生成器の構造を図 12 に示す

実験 6 では、学習する頂点数を 1024 と非常に増やし、分類に使用する観点数も 1024 と非常に増やしたあと段階的に 512、256、128、64 と減らした。

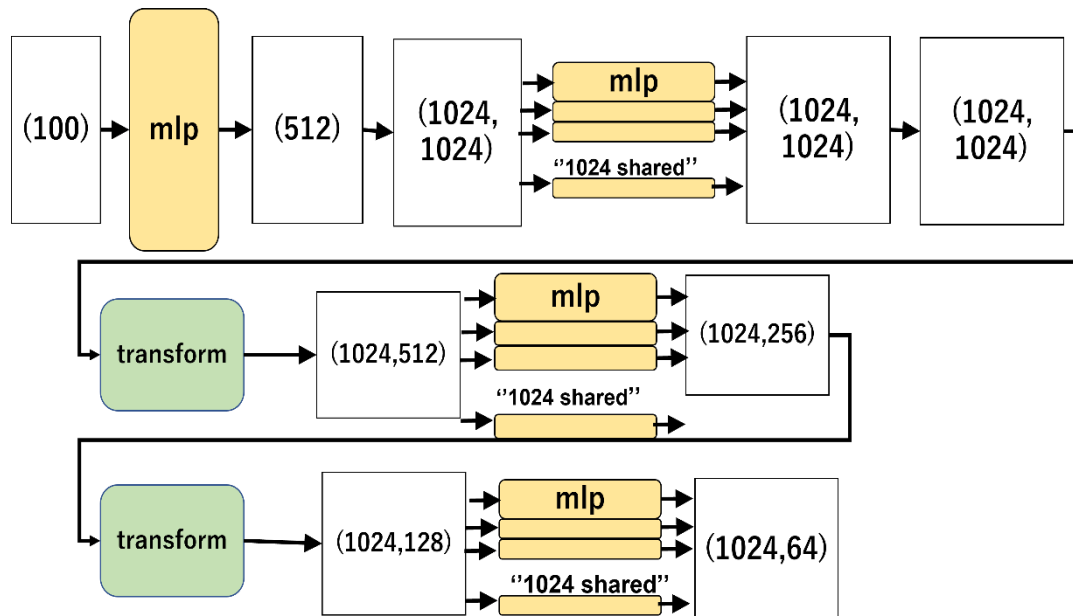


図 12 実験 6 の生成器の構造

4.4 実験結果

各実験での生成器の構造を使った生成モデルを用いたカテゴリごとの FID スコアを表 4.2 に示す。また、表の中、それぞれのカテゴリで最も FID スコアが高かった部分をグレーで強調して示す。

表 4.2 FID スコア

カテゴリ名	bathhtub	bed	chair	desk	dresser	monitor	night stand	sofa	table	toilet
先行研究	5816.1	2511.2	329.4	1738.9	2418.0	2245.2	3621.4	1016.1	864.8	941.1
実験 1	5734.4	2677.8	343.1	1890.0	2678.9	2748.4	4241.7	1369.0	925.7	1050.1
実験 2	5932.6	2798.5	368.3	1639.6	2573.5	2384.2	3787.2	1382.1	898.8	982.0
実験 3	5968.3	2465.7	319.9	1687.0	2359.9	2042.7	3590.4	980.3	736.2	826.5
実験 4	5829.0	2597.6	339.9	1927.5	2479.4	2380.0	3715.3	1287.6	835.9	978.6
実験 5	5935.3	2894.9	348.0	1809.2	2572.4	2490.9	3960.6	1476.1	936.7	972.1
実験 6	5976.3	2699.0	352.5	1826.9	2596.6	2590.9	3725.4	1087.4	925.6	962.1

第5章. 考察

実験結果より、ほとんどの実験結果で先行研究の FID スコアを上回ることができなかったが、実験 3 の FID スコアが先行研究に対して、10 カテゴリ中 8 カテゴリで上回った。実験 3 では、学習が進むごとに学習する頂点数は変えずに着目する観点数を増やしたものだ。先行研究は着目する観点数を transform 層以降 32、8、8、3 としていたが、実験 3 では、64、128、256、512 と着目する観点数を増やし学習を行った。このことから、本実験の PointNet を利用した点群データ生成ネットワークでは、生成器の構造は、はじめは 16、32 程度の少ない数の頂点数や分類に使用する観点数から学習をはじめ、使用する観点数を倍に増やしていく生成器の構造がより良い点群データ生成器の構造であると考えられる。

第6章. 結論

先行研究に習い3Dデータを生成する敵対的生成ネットワークの既存手法である3D-GANとは異なった構造を持つ、敵対的生成ネットワークを使った点群データ生成モデルの改善の実験を行った。先行研究であるPointNetを応用した生成モデルのFIDスコアと生成器の構造を変えた複数の生成モデルのFIDスコアを比較した結果、先行研究と比較して、10カテゴリ中8カテゴリでFIDスコアの改善を行う生成器の構造を発見できた。このことから、敵対的生成ネットワークを用いた点群データ生成の検討と改善が行えたと考える。

今後の課題として、本実験では生成器の構造だけを変化させて結果を比較したが、より良い生成モデルの検討には、識別器側の構造も変化させることや、より幅広い範囲でパラメータを変化させて生成モデルの精度を定量的に比較することが必要である。

謝辞

本論文を作成するにあたり、最後まで多くのご指導、ご助言をいただきました三好力教授に厚くお礼申し上げます。

第7章. 参考文献

- [1] Generative Adversarial Networks Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio
- [2] Learning a Probabilistic Latent Space of Object Shapes via 3D Generative-Adversarial Modeling Jiajun Wu*, Chengkai Zhang*, Tianfan Xue, William T. Freeman, and Joshua B. Tenenbaum
- [3] qiita“PointNet の理論と実装(点群データ)”,
<https://qiita.com/opeco17/items/707a5c57bca41a145122>,(参照 2023-01-07).
- [4] PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation Charles R. Qi, Hao Su, Kaichun Mo, Leonidas J. Guibas
- [5] 山本敬士,三好力 : “敵対的生成ネットワークを用いた点群データ生成の検討,” ,2021 年度龍谷大学卒業論文,2022-01.
- [6] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler and S. Hochreiter, “GANs trained by a two time-scale update rule converge to a local Nash equilibrium,” Proc. NIPS 2017 (2017).
- [7] 3D ShapeNets: A Deep Representation for Volumetric Shapes (ModelNet10) Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, Jianxiong Xiao.