

# 敵対的生成ネットワークを用いた、 3次元点群形状特徴量の教師なし学習

上西 和樹<sup>†</sup> 古屋 貴彦<sup>†</sup> 大淵 竜太郎<sup>†</sup>

**概要**：効率的な管理を目的に、形の類似性に基づいて3次元形状データを比較、検索、分類する手法への要求が高まっている。近年提案された深層ニューラルネットワーク (DNN)による3次元形状処理手法の多くは、3次元形状とこれに付与されたラベルの対応を教師付き学習して高精度な形状特徴量を獲得する。しかし現存する3次元形状データの多くはラベルを持たない。人手によるラベル付けの手間が大きいためである。そこで、ラベルなしの3次元形状データから高精度な形状特徴量を獲得する手法が望まれる。本研究では、教師なし学習の一種である敵対的生成ネットワーク (GAN)を用いた、点群を対象とする3次元形状特徴量を提案する。提案手法である点群GANでは、3次元点群を生成するDNN (生成DNN)と、3次元点群の本物らしさを予測するDNN (判別DNN)とを互いに競わせながら訓練する。その訓練の過程で、3次元点群の形状特徴量を抽出する能力を判別DNNが獲得する。3次元モデルの形状類似検索のシナリオで評価した結果、判別DNNで得た形状特徴量は、既存の手作り形状特徴量よりも高精度に3次元形状を比較できることが分かった。

**キーワード**：3次元形状特徴量, 3次元点群, 教師なし学習, 敵対的生成ネットワーク, 3次元形状類似検索

## Unsupervised Representation Learning for 3D Point Set by using Generative Adversarial Neural Network

KAZUKI UENISHI<sup>†</sup> TAKAHIKO FURUYA<sup>†</sup>  
RYUTAROU OHBUCHI<sup>†</sup>

**Abstract**. Shape similarity-based indexing, retrieval, and classification of 3D shapes are essential for efficient management of 3D shape data. Recently proposed deep neural networks (DNNs) tailored to 3D shapes, e.g., 3D point sets, learn accurate 3D shape feature representation by associating 3D shapes with semantic labels attached to them. However, in practice, majority 3D shapes are left unlabeled due to cost of annotation. Thus, unsupervised approaches to learn 3D shape features from unlabeled 3D shapes are desired. This paper proposes an unsupervised representation learning algorithm that utilizes Generative Adversarial Network (GAN) to obtain descriptors of 3D shapes defined as point sets. Proposed Point Set GAN, which consists of a generator DNN and a discriminator DNN, is adversarially trained. The generator is trained so that it yields realistic 3D point sets, while the discriminator is trained to distinguish “fake” point sets produced by the generator from “real” ones. After the training, the discriminator DNN is to compute shape descriptor. Experimental evaluation under shape-based 3D model retrieval scenario demonstrates that 3D shape feature learned by proposed Point Set GAN outperforms existing handcrafted 3D shape features.

**Keywords**: 3D shape feature, 3D point set, unsupervised learning, generative adversarial network, shape-based 3D model retrieval

### 1. はじめに

近年、3次元形状モデル (3Dモデル) の数が急激に増加している。増加の背景には、仮想現実 (VR) や拡張現実 (AR) の技術の発展、安価な3Dスキャナや3Dプリンタの普及等がある。多数の3Dモデルを効率的に管理するため、形の類似性に基づいてこれら3次元形状を比較、検索、分類する手法への要求が高まっている。

一般的に、3Dモデル間の形状類似度はそれら3Dモデルから抽出された3D形状特徴量を比較することで計算される。形状特徴量は、人間により設計された「手作り形状特徴量」と3D形状データを学習して得た「学習形状特徴量」の2つに大別できる。研究者の経験と知識に基づいて設計された手作り形状特徴量 (例えば, [12], [13], [14], [15], [16],

[17], [18], [19]) は、一定の形状比較精度を示すものの、必ずしも形状比較に最適であるとは限らない。一方で、多数の3Dモデルを学習して得た学習形状特徴量は、手作り形状特徴量よりも3Dモデルの形状比較に適応し、より高い精度で形状比較できることが期待される。

学習形状特徴量を得るアプローチには、教師あり学習と教師なし学習がある。近年提案された高精度な学習形状特徴量の多くは教師あり学習を用いる ([3], [6], [20], [21], [22], [23]など)。これら既存研究では、ラベル付きの3D形状データをできるだけ多数用いて、3Dモデル形状特徴量の抽出向けに設計された深層ニューラルネットワーク (DNN) を訓練する。しかし、データベースに日々蓄積される大量の3Dモデルの殆どはラベルを持たない。人手による3Dモデルへのラベル付与の手間が大きいためである。そのため、教師あり学習に使える3Dモデルの数や種類が限られる。

<sup>†</sup>1 山梨大学  
University of Yamanashi

一方で、教師なし学習を用いれば、ラベルを持たない多数の3Dモデルで学習できる。しかし、DNNの教師なし学習は一般に困難であり、3D形状特徴量の教師なし学習を行った研究事例も少ない。

本研究の目的は、ラベルを持たない3Dモデル群から教師なし学習によって高精度な学習形状特徴量を獲得することである。その目的のため、教師なし学習の一種である敵対的生成ネットワーク(Generative Adversarial Network, GAN) [10]のアプローチを、3D点群で表現された3次元形状を対象に適用する。3D点群は汎用性の高い形状表現であり、距離画像センサの普及等により近年その用途が拡大している。

本研究で提案する点群GAN(図1)は、3D点群を生成するDNN(生成DNN)と、3D点群の真贋を判別するDNN(判別DNN)の2つのDNNで構成される。生成DNNは、入力されたランダムな潜在ベクトルを3D点群に変換する。判別DNNは、入力された3D点群について、それが人手で製作された“本物”であるか、生成DNNが出力した“偽物”であるかを判別する。生成DNNと判別DNNが互いに敵対しつつ学習を進めることで、生成DNNは本物らしい3D点群の生成を学習し、また判別DNNは3D点群の形状の判別、すなわち形状特徴を学習する。

もともとGAN [10]は2次元画像の生成および判別のために設計され、そのままのDNN構造では3D点群の生成および判別に用いることはできない。そこで本研究では、3D点群向けのGANを新たに設計する。点群GANの判別DNNが獲得する形状特徴量の要件の1つは、3D点群を構成する点の順序に対する不変性である。そこで我々は、点の順序変化に影響を受けない3D点群解析用DNNを用いて点群GANを構築する。具体的には、QiらのPointNet [3]を元に判別DNNを設計し、Yangらの折畳みAutoencoder(折畳みAE) [4]を元に生成DNNを設計する。

GANを用いるうえでの困難さの一つは、その学習が不安定なことである([10], [11])。我々は、点群GANの学習を安定化させるため、Gulrajaniら[1]が提案したGANの学習則を用いる。具体的には、3D点群データ個々の真贋 [10]の代わりに、本物の点群データ集合と偽物の点群データ集合の集合間距離を損失として利用する。

点群GANの学習の後、判別DNNを3D点群の特徴抽出器として用いる。判別DNNの中間層から抽出された形状特徴量を、3D形状の比較、検索、分類等に用いる。

点群GANが獲得した形状特徴量の精度を実験的に評価するため、3Dモデルの形状類似検索のシナリオで実験を行った。実験の結果、点群GANが獲得した学習形状特徴量の検索精度は、既存の手作り形状特徴量、および、教師なし学習を用いた既存の学習形状特徴量の双方を上回ることが分かった。また、点の順序変化に影響を受けない3D点群解析用DNN([3], [4])の採用、および、GANの安定な学習則[1]の利用が、それぞれ形状特徴量の精度の改善に効果的であることを確かめた。

## 2. 関連研究

### 2.1 DNNを用いた3次元形状特徴量の学習

DNNを用いた学習形状特徴量の多く(例えば、[3], [6], [20], [21], [22], [23])は、3D形状とこれに付与された意味ラベルを教示データとした教師あり学習により獲得される。ボクセル表現、3D点群、多視点見かけ画像群等の形状表現を対象とした学習形状特徴量がこれまでに提案された。Wuらが提案したShapeNet [6]はボクセル表現を対象とし、3次元の畳み込みニューラルネットワーク(CNN)を用いて3D形状特徴量を抽出する。Suらが提案したMV-CNN [20]は、3Dモデルの多視点の見かけ画像群を対象とする。MV-CNNは2DのCNNを用いて見かけ画像の各々から画像特徴量を

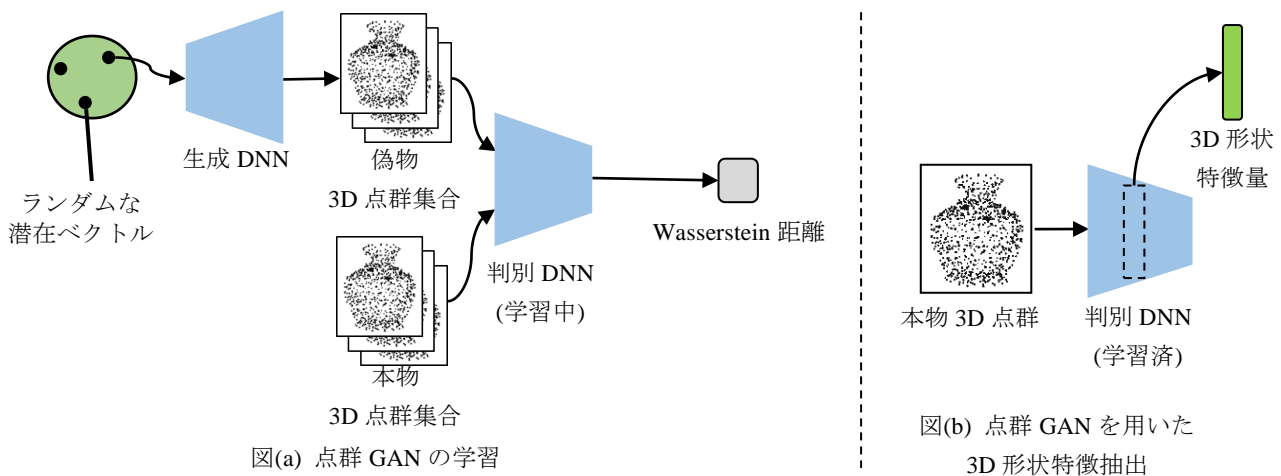


図1 点群GANを用いた教師なし形状特徴量学習。潜在ベクトルを3D点群に変換する生成DNNと、3D点群の本物らしさを測る判別DNNとが敵対的に学習する過程で、判別DNNが3D点群の形状特徴量を獲得する。

抽出し、全視点の画像特徴量を 3D モデル当たり 1 つの特徴量に統合することで 3D 形状特徴量を得る。

本研究と同様、3D 点群を対象とした研究には[3], [4], [5], [21]などがある。Furuya らが提案した DLAN [21]は、3D モデルの局所領域から抽出した多数の手作り幾何特徴量を DNN で精製、統合することで 3D モデルの形状特徴量を得る。Qi らが提案した PointNet [3]は、DNN を用いて 3D 点群の各点を高次元の特徴量に変換し、それらを統合した後、カテゴリ分類や領域分割を行う。PointNet は点群データを直接入力でき、かつ、点の順序変化に影響を受けない利点がある。[3], [21]はいずれも、DNN を効果的に訓練するために多数のラベル付き 3D 点群データが必須となる。

本研究で提案する点群 GAN は PointNet と類似の構造を持つ DNN を点群判別 DNN として用いるが、その学習は GAN の枠組みでの教師なし学習である。

近年、3D 点群を対象とした教師なし形状特徴量学習法として、点群 Autoencoder が提案されている [4][5]。Achlioptas ら[5]は符号化器として PointNet を用い、復号化器に全結合 DNN を用いた。一方で、Yang ら [4]はより高精度な形状特徴量の獲得をねらい、折畳み AE を提案した。折畳み AE は 3D 点群向けに設計された Autoencoder であり、PointNet 同様、点の順序変化に対する不変性を持つ。折畳み AE では、3D 点群の符号化器として PointNet に類似した DNN が用いられる。3 次元点群の復号化では、符号化により得た形状特徴量を手がかりに、規則格子状の 2D 点群を「折り畳む」ことで 3D 形状を復元する。Yang らは、2D 格子点群の折畳みによる復号化が Autoencoder を用いた高精度な形状特徴量の学習に効果的であることを示した。

[4]および[5]が提案した点群 Autoencoder と、本研究が提案する点群 GAN との違いは、DNN の学習方法にある。点群 Autoencoder が入力点群自身を復元するよう学習するのに対し、点群 GAN は点群生成と点群判別を敵対的に学習する。点群 GAN では、折畳み AE の復号化器を生成 DNN として用いる。2D 点群の折畳みによる 3D 点群生成を行うことで、生成される点群の質を改善し、特徴量学習の効果を高める。

## 2.2 敵対的生成ネットワーク

敵対的生成ネットワーク (GAN) [10]は本来、2D 画像の生成を学習する DNN として提案された。GAN は生成 DNN と判別 DNN から成る。生成 DNN はランダムな潜在ベクトルを入力として受け取り、2D 画像を出力する。一方で判別 DNN は、2D 画像を入力として受け取り、その画像が本物である確率を出力する。GAN の学習では、生成 DNN と判別 DNN が敵対する。即ち、判別 DNN は画像が本物であるか、生成 DNN が出力した偽物であるかを判別するように訓練される。対して生成 DNN は、判別 DNN が誤判別するような「本物そっくりの偽物」を生成するように訓練され

る。敵対的学習の結果、判別 DNN が画像分類に有用な画像特徴量を獲得することが示されている [25]。

複数の研究 ([10], [11]など)において、GAN の学習が不安定であることが示されている。例えば、生成 DNN と判別 DNN の学習の進捗のバランスが崩れると学習が進まなくなる。GAN の学習を安定化させるために、DNN 構造の改良 [25]や正則化 [11]等が提案された。Arjovsky ら[2]は主に損失関数の改良により学習を安定化した Wasserstein GAN (WGAN) を提案した。従来の GAN の生成 DNN が偽物の画像個々の本物らしさを高めるように学習するのに対し、WGAN の生成 DNN は偽物の画像集合の本物らしさを高めるように学習する。画像集合の本物らしさの尺度として、本物の画像集合と偽物の画像集合との集合間距離 (Wasserstein 距離) が用いられる。Gulrajani ら [1]は、WGAN の学習をさらに安定化させた WGAN-GP 法を提案した。WGAN-GP では、判別 DNN の勾配ノルムに対して制約 (Gradient Penalty) を加えることで、より正確な Wasserstein 距離が得られるようになり、生成される画像の質が向上する。

本研究では、点群 GAN の学習に WGAN-GP 法を用いることで、学習の安定性を高める。

## 3. 提案手法

### 3.1 手法の概要

提案手法は、点群 GAN を用いた 3D 形状特徴量の教師なし学習法である。図 1 に点群 GAN の概要を示す。点群 GAN は、3D 点群の生成を学習する DNN (生成 DNN) と、3D 点群の本物らしさを計測する DNN (判別 DNN) が敵対して学習する (図 1(a))。敵対的学習の後、特徴抽出対象の 3D 点群を判別 DNN へ入力し、判別 DNN の中間層の活性を入力点群の形状特徴量として抽出する (図 1(b))。この形状特徴量を 3D 形状の比較や検索に用いる。

点群 GAN が高精度な形状特徴量を学習するために、(1) 3D 点群を構成する点の順序変化に不変な DNN 構造、および (2) 学習が安定する損失関数が要求される。点の順序に対する不変性を得るために、PointNet [3]を判別 DNN として用い、また折畳み AE [4]の復号化器を生成 DNN として用いる。また、学習の安定性を高めるために、WGAN-GP 法 [1]に基づく損失関数を利用する。

### 3.2 ネットワーク構造

#### 3.2.1 判別 DNN

図 2 に判別 DNN の構造を示す。判別 DNN では、まず、前半の点特徴抽出部において、入力された 3D 点群の各点を独立に全結合層へ入力し点ごとの特徴量を得る。点特徴量への変換を行う全結合層の構造及びパラメータは、全ての 3D 点で共通である。点特徴抽出部の全結合層の数は 4

であり、各層のニューロン数は  $64-128-256-512$  である。次いで、点ごとの特徴量群を最大値プーリングにより 3D 点群モデル当たり 1 つの特徴量に統合する。統合特徴量は後続の 6 個の全結合層から成る後半の統合特徴抽出部へ入力され、精製される。統合特徴抽出部の各層のニューロン数はそれぞれ  $1,024-1,024-1,024-1,024-512-64$  である。最後に、統合特徴抽出部の出力は、1 つの全結合層により Wasserstein 距離の近似値に変換される。判別 DNN が出力する Wasserstein 距離は、本物点群データの集合（ミニバッチ）と偽物点群データの集合（ミニバッチ）の間で計算された集合間距離である。Wasserstein 距離は Earth Mover's 距離とも呼ばれ、その値が小さいほど、偽物点群データがより本物らしいことを示す。判別 DNN の最終層を除く全ての全結合層において、Leaky ReLU [8] を活性化関数として用いる。

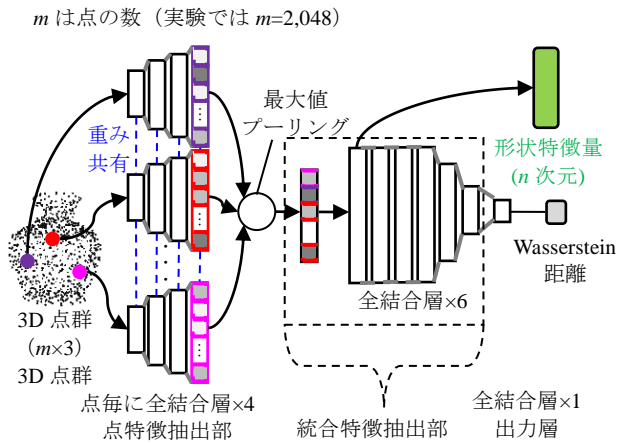


図 2 判別 DNN のネットワーク構造。点毎の全結合層と最大値プーリングによって 3D 点群全体の形状特徴量を得る。その後、Wasserstein 距離を計算する過程で形状特徴量が精製される。

### 3.2.2 生成 DNN

点群 GAN の生成 DNN には全結合生成 DNN と折畳み生成 DNN の 2 つの構成を試みる。判別 DNN を効果的に学習し形状特徴量の精度を高めるために、生成 DNN はより本物らしい 3D 点群を生成することが望ましい。Yang ら[4] は、折り畳み復号器が、全結合復号器よりも正確に 3D 点群を再構成できることを示した。その理由は、2 次元格子の折畳み処理が、物体表面に点が分布する 3D 点群の再構成により適するためである。[4]と同様、本研究が扱う 3D 点群も物体表面に分布する。そのため、折畳みによる点群生成が高精度な形状特徴量の獲得に繋がると期待できる。しかしながら、点群 GAN は点群 Autoencoder と構造および学習方法が異なっており、折畳み DNN と全結合 DNN のどちらが形状特徴量の学習に有効であるか不明である。したがって、本研究ではこれら 2 つの生成 DNN を比較評価す

る。

- 全結合生成 DNN: 全結合 DNN によって潜在ベクトルを 3D 点群へ変換する DNN (図 3(a)).
- 折畳み生成 DNN: 折畳み AE の復号器と類似の構成を用いて潜在ベクトルと 2D ランダム点群を 3D 点群へ変換する DNN (図 3(b)).

上記いずれの生成 DNN においても、入力潜在ベクトルには平均 0.0, 分散 0.2 の正規分布からランダムにサンプルされた 32 次元ベクトルを用いる。また、生成される 3D 点群の点数は 2,048 とする。生成された 3D 点群は、判別 DNN へ入力する前に位置と大きさの正規化を施す。具体的には、点群の重心を原点に移動し、次いで半径 0.5 の球に収める。上記の正規化は、学習データに施す正規化 (3.3.3 項「学習データ」で示す) と同様の正規化を行うためのものである。上記の正規化を行うことで判別 DNN による真贋判別がより困難となるため、より高精度な形状特徴量の獲得が期待できる。

全結合生成 DNN (図 3(a)) は 3 層の全結合層から成り、各層のニューロン数は  $1,000-2,000-6,144$  である。最終層の 6,144 は、2,048 個の 3D 点群 ( $3 \times 2,048$ ) を表す。活性化関数は、中間層では Leaky ReLU を用い、最終層では hyperbolic tangent を用いる。

折畳み生成 DNN (図 3(b)) は、潜在ベクトルで条件付けられた 2D ランダム点群を、2 つの折畳み部を経て、3D 点群へ変換する。2D ランダム点群は、2D 平面上にランダムに点を配置した点群を表す。2D ランダム点群を構成する各点の座標  $(x, y)$  の  $x$  及び  $y$  には、共に範囲  $[0, 1]$  の一様分布からランダムにサンプルされた値を用いる。折畳み生成 DNN では、まず、2,048 個の 2D ランダム点群の各々の点の座標ベクトルに潜在ベクトルを接続することで、 $32+2=34$  次元の 2,048 個の点群を得る。この点群を 1 つ目の折畳み部を用いて「中間的な」3D 点群へ変換する。折畳み部は 3 つ全結合層で構成され、各層のニューロン数は 1024, 1024, 3 である。中間的な 3D 点群に潜在ベクトルを再度接続し、これを 2 つ目の折畳み部に入力することで、3D 形状を表す点群を得る。2 つ目の折畳み部の構造は 1 つ目と同じである。

## 3.3 ネットワークの学習

### 3.3.1 学習の損失関数

点群 GAN の損失関数には、Gulrajani らによって提案された WGAN-GP の損失関数 [1]を用いる。式(1)に生成 DNN の損失関数を示す。式(1)において、 $\mathbf{z}$  は潜在ベクトル、 $G(\bullet)$  は生成 DNN が出力した偽物点群、 $C(\bullet)$  は判別 DNN が出力した Wasserstein 距離を示す。式(2)は判別 DNN の損失関数である。式(2)において、係数  $\lambda$  の項は Gradient Penalty を示し、 $\hat{\mathbf{x}}$  は本物点群  $\mathbf{x}$  と偽物点群  $G(\mathbf{z})$  の座標値を内分して得た 3 次元点群を示す。実験では  $\lambda=1$  を用いる。

$$L_G = -\frac{1}{b} \sum_{i=1}^b C(G(\mathbf{z}_i)) \quad (1)$$

$$L_D = \frac{1}{b} \sum_{i=1}^b C(G(\mathbf{z}_i)) - \frac{1}{b} \sum_{i=1}^b C(\mathbf{x}_i) + \lambda \frac{1}{b} \sum_{i=1}^b (\|\nabla_{\mathbf{x}} C(\hat{\mathbf{x}}_i)\|_2 - 1)^2 \quad (2)$$

実験では、点群データ個々の本物らしさを高める損失関数と上記の損失関数を比較する。比較対象は式(3)および(4)に示す Feature matching [11]を用いた損失関数である。Feature matching は、判別 DNN の中間層に現れる特徴量  $f$  を、本物点群と偽物点群の間で類似させる制約である。式(3)における  $a$  の項が Feature matching であり、実験では  $a=1$  を用いる。  $D(\bullet)$  は、損失関数として式(3)および(4)を用いた場合の判別 DNN の出力、即ち、入力 3D 点群が本物である確率を表す。

$$L_G = -\log(D(G(\mathbf{z}))) + a\|f(\mathbf{x}) - f(\mathbf{z})\| \quad (3)$$

$$L_D = -(\log(D(\mathbf{x})) + \log(1 - D(G(\mathbf{z})))) \quad (4)$$

### 3.3.2 最適化

最適化には Adam [7]を用いる。生成 DNN および判別 DNN の初期学習係数は共に 0.0001 を用いる。ミニバッチサイズ  $b$  は 50 を用いる。点群 GAN の学習では、生成 DNN のパラメータ更新 1 回と判別 DNN のパラメータ更新 5 回を交互に行う。これを 50,000 回反復する。ただし、50,000 回反

復する前に評価用データセットにおける MAP が変化しなくなった場合には、その時点で学習を打ち切る。

### 3.3.3 学習データ

点群 GAN の学習データ (本物点群) は 3D ポリゴンモデルを点群化することで作成される。ポリゴンの点群化には Ohbuchi らの手法 [13]を用い、3D ポリゴンモデルの表面を一樣サンプリングしてポリゴンモデル 1 つを 2,048 個の点から成る 3D 点群データへ変換する。得られた 3D 点群には、位置と大きさの正規化を施す。具体的には、点群の重心を原点に移動し、次いで半径 0.5 の球に収める。3D ポリゴンモデルの学習用データセットとしては、Wu ら [6]が作成した ModelNet40 の学習用データセット、または ModelNet10 の学習用データセットを用いる。ModelNet40 の学習用データセットは、椅子、飛行機、人間など 40 個のカテゴリに分類された 9,843 個の剛体 3D モデルから成る。また、ModelNet10 の学習用データセットは、机、ベッド、椅子など 10 個のカテゴリに分類された 3,991 個の剛体 3D モデルから成る。ModelNet10、ModelNet40 とともにその 3D モデルがカテゴリラベルを持つが、提案する点群 GAN は教師なし特徴量学習手法であるため、学習に際し 3D モデルに付与されたラベルは用いない。

### 3.4 形状特徴量の抽出

点群 GAN 学習の後、判別 DNN を 3D 点群モデルからの形状特徴量抽出に用いる (図 1(b))。特徴抽出対象の 3D 点群モデルを判別 DNN へ入力し、判別 DNN の統合特徴抽出部のニューロン活性ベクトルを得る。得られたニューロン活性ベクトルをパワー正規化[9]した後、L2 正規化したも

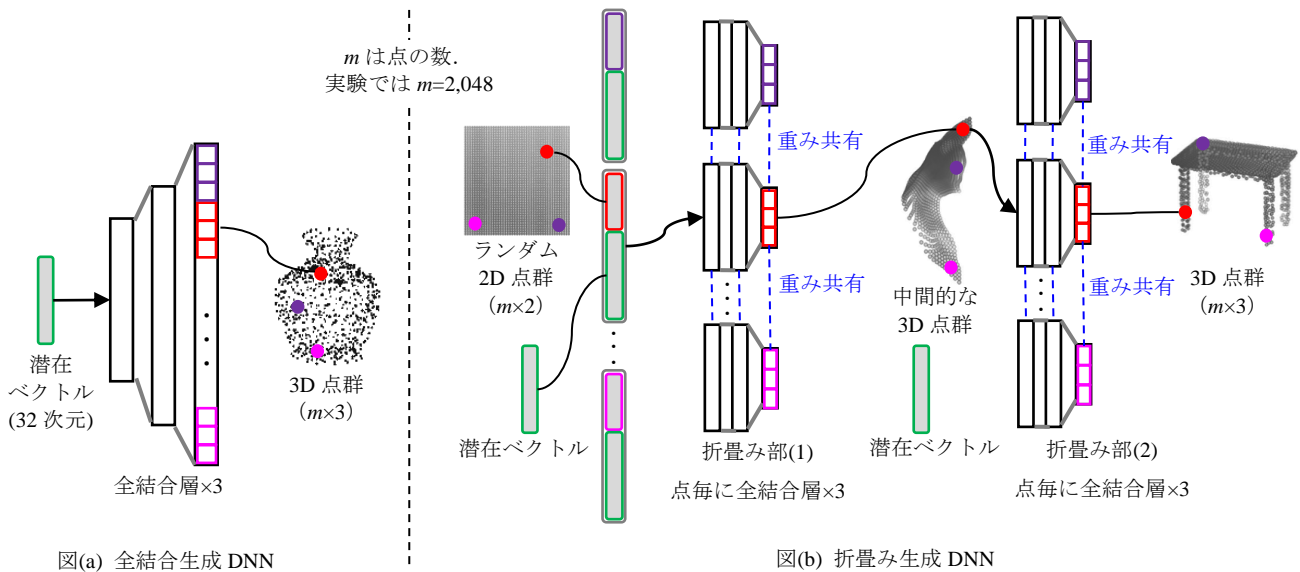


図 3 生成 DNN のネットワーク構造。全結合生成 DNN は、潜在ベクトルを 3D 点群へと変換する。一方で、折畳み生成 DNN は、2D ランダム点群を折り畳むようにして 3D 点群へと変換する。この時、出力される 3D 点群の形状は、2D ランダム点群の各点に潜在ベクトルを連結することで条件付けられる。



のが 3D 点群モデルの形状特徴量である。実験では、図 2 に示す統合特徴抽出部の 7 つの層（最大値プーリングの直後、および 6 つの全結合層）のいずれか 1 つから形状特徴量を抽出する。各特徴量の次元数  $n$  は、その特徴量を抽出した層のニューロン数と同じである。具体的には、最大値プーリングの直後から順に、512, 1024, 1024, 1024, 1024, 512, 64 である。2 つの 3D モデル間の形状相違度は、それら 3D モデルから抽出された  $n$  次元形状特徴量間の L2 距離により計算される。

## 4. 実験と結果

### 4.1 実験条件

点群 GAN が学習した形状特徴量の精度を 3D モデルの形状類似検索のシナリオで評価する。実験ではまず、点群 GAN の構造および損失関数が形状特徴量の精度に与える影響を調査する。次に、既存の手作り形状特徴量、および既存の教師なし学習形状特徴量との検索精度比較を行う。

**データセット：**評価実験には ModelNet40 データセット [6]を用いる。3.3.2 項で述べた通り、点群 GAN の学習には ModelNet40 の学習用データセットに含まれる 9,843 個の 3D モデルを用いる。検索精度の評価には ModelNet40 の評価用データセットに含まれる 2,468 個の 3D モデルを用いる。また、既存の形状特徴量との比較実験では、ModelNet40 に加え ModelNet10 を用いる。ModelNet10 を用いた実験では、点群 GAN の学習に ModelNet10 の学習用データセットに含まれる 3,991 個の 3D モデルを用いる。検索精度評価には、ModelNet10 の評価用データセットに含まれる 908 個の 3D モデルを用いる。検索精度の評価尺度としては Mean Average Precision (MAP) を用いる。各実験は 3 回ずつ行い、3 つの MAP 値の平均と標準偏差を報告する。

**比較対象：**既存の 3D 点群向け特徴量と比較する。手作り形状特徴量の比較対象として D2 [12], AAD [13], SPRH [14], LSF [15], SI [16], RoPS [17], POD [26]を用いる。D2, AAD, および SPRH は有向 3D 点群向けの全体形状特徴量であり、有向点の対ごとに計算される低レベル特徴量（距離や向きベクトルの内積など）のヒストグラムである。LSF, SI, RoPS, および POD は 3D 点群向きの局所形状特徴量であり、実験では 3D モデル当たりそれぞれ 512 個抽出する。これら 512 個の局所形状特徴量を Bag-of-Features (BF)法 [24]または DkSA 法 [18]を用いて 3D モデル当たり 1 個の特徴量へ統合する。用いた統合方法を明らかにするため、局所特徴量の名称に“BF-”または“DkSA-”の接頭詞を付与する。統合後の特徴次元数はいずれも 2,000 とする。

教師なし学習特徴量の比較対象として、折畳み AE 特徴量 [4], 全結合 AE 特徴量 [5], 法線推定 PointNet 特徴量を用いる。折畳み AE と全結合 AE は、3D 点群の符号化と復号化を同時に学習することで、符号化によって得た形状特

徴量である。全結合 AE と折畳み AE の違いは復号化器の構造である。全結合 AE の復号化器は、符号化によって得られた潜在ベクトルを全結合層によって 3D 点群へ変換する。一方で、折畳み AE の復号化器は、2D 格子点群を折畳むようにして 3D 点群を形成する。公平な比較のため、これら AE の符号化器は、Yang らの折畳み AE で用いられたものに統一して実装する。法線推定 PointNet は、領域分割 PointNet [3]を元の実装する。具体的には、3D 点群から、各点に対応する法線方向を推定する PointNet を学習し、学習後の中間層から特徴量を得る。特徴量を抽出する層は、点群 GAN が高い精度を示す統合特徴抽出部の 3 層目である。

また、提案手法による精度の位置付けを知るため、教師あり学習に基づく 3D 形状特徴量と比較する。具体的には、3D ShapeNets [6], MV-CNN [20], DLAN [21]による検索精度を参考値として用いる。これら DNN は 3D 形状のカテゴリ分類を学習する過程で 3D 形状特徴量を獲得する。

### 4.2 実験結果

#### 4.2.1 特徴抽出層の比較

判別 DNN の統合特徴抽出部の 7 つの層それぞれから形状特徴量を抽出し、それらの精度を比較する。図 4 に各層から抽出された形状特徴量の検索精度を示す。図 4 において、特徴抽出層 1 は最大値プーリングで得た統合特徴、特徴抽出層 2~7 はそれぞれ 1~6 個目の全結合層から抽出された統合特徴である。学習時の生成 DNN には 3.2.2 節で述べた折畳み生成 DNN を用い、WGAN-GP の損失関数を用いる。

図 4 より、折畳み生成 DNN と全結合生成 DNN のいずれにおいても、統合特徴抽出部の特徴抽出層 3 または 4 に検索精度のピークが見て取れる。一方で、出力に近い層（特徴抽出層 6 または 7 付近）から抽出した特徴量は精度が低く、ばらつきが大きい。これは、出力層に近くなるにつれ、層の役割が形状特徴を捉えることから Wasserstein 距離を計算することに変化するためと推察される。

#### 4.2.2 生成 DNN の比較

図 4 に示した全結合生成 DNN と折畳み DNN の検索精度を比較すると、折畳み生成 DNN の採用により特徴量学習の効果が高まることが分かる。折畳み生成 DNN を用いて学習する場合では、全結合生成 DNN を用いて学習する場合よりも精度の平均が高い形状特徴量が獲得された。特に、統合特徴抽出部の 3, 4 層目から抽出された形状特徴量の MAP 値がそれぞれ 0.563, 0.564 と高い。3D 点群向けの生成 DNN の利用により判別 DNN の学習が効果的に進み、形状特徴量が改善したと考えられる。

4.2.3 損失関数の比較

点群 GAN の学習に用いる損失関数について、Gulrajani らの WGAN-GP 法と、Salimans らの Feature matching 法を比較する。この実験では、生成 DNN には折畳み生成 DNN を用いる。図 5 に実験結果を示す。図 5 のグラフの横軸は形状特徴量の次元数である。特徴量は判別 DNN の統合特徴抽出部の内、3 層目から抽出する。3 層目のニューロン数には {64, 128, 256, 512, 1024, 2048, 4096} のいずれかを用いる。この時、各層のニューロン数のバランスを保つため、2,4,5 層目のニューロン数を 3 層目に統一する。図 5 より WGAN-GP は、ほとんどの次元数において GAN よりも高い検索精度を示す。また、WGAN-GP では特徴次元数が大きい場合に精度のばらつきが小さい。WGAN-GP の利用により点群 GAN の学習が安定し、256 次元以上のいずれの次元数においても高精度な特徴量が学習された。

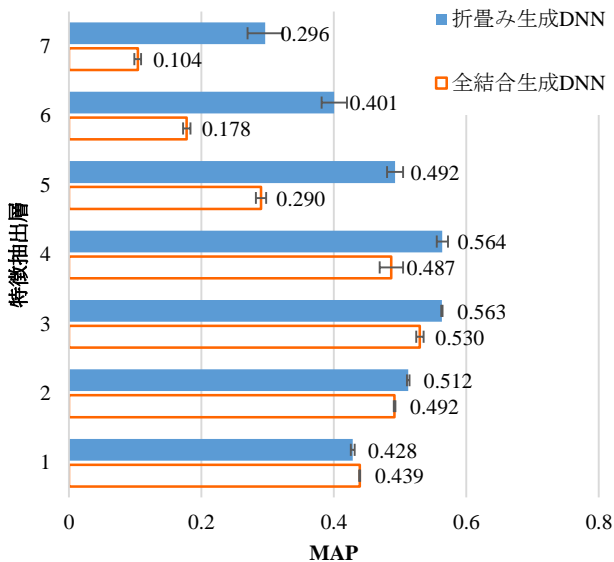


図 4 特徴抽出層の検索精度比較。

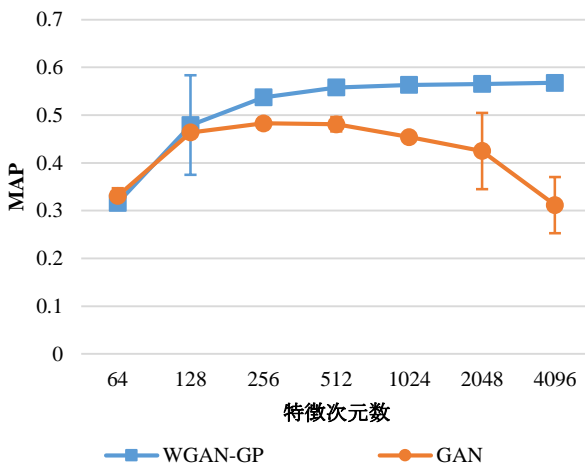


図 5 点群 GAN の学習に用いる損失関数の比較。

4.2.4 既存の形状特徴量との比較

表 1 に、提案手法の点群 GAN により学習した形状特徴量と、既存の形状特徴量との検索精度比較を示す。表 1 中の MN40 は ModelNet40 データセット、MN10 は ModelNet10 データセットを用いて評価した結果を表す。点群 GAN を用いて学習した形状特徴量は、比較に用いた 4 つの手作り形状特徴量、および 3 つの教師なし学習形状特徴量と同等の特徴次元数でありながら、これら既存手法を上回る検索精度を示す。比較対象の中でも DkSA-POD [18] や折畳み AE 特徴量 [4] は、ラベルなしの 3D 点群モデルの形を比較するための state-of-the-art な形状特徴量である。点群 GAN がこれら state-of-the-art な形状特徴量を上回ることから、本研究のアプローチである点群 GAN を用いた教師なし 3D 形状特徴量の学習の有効性が示された。

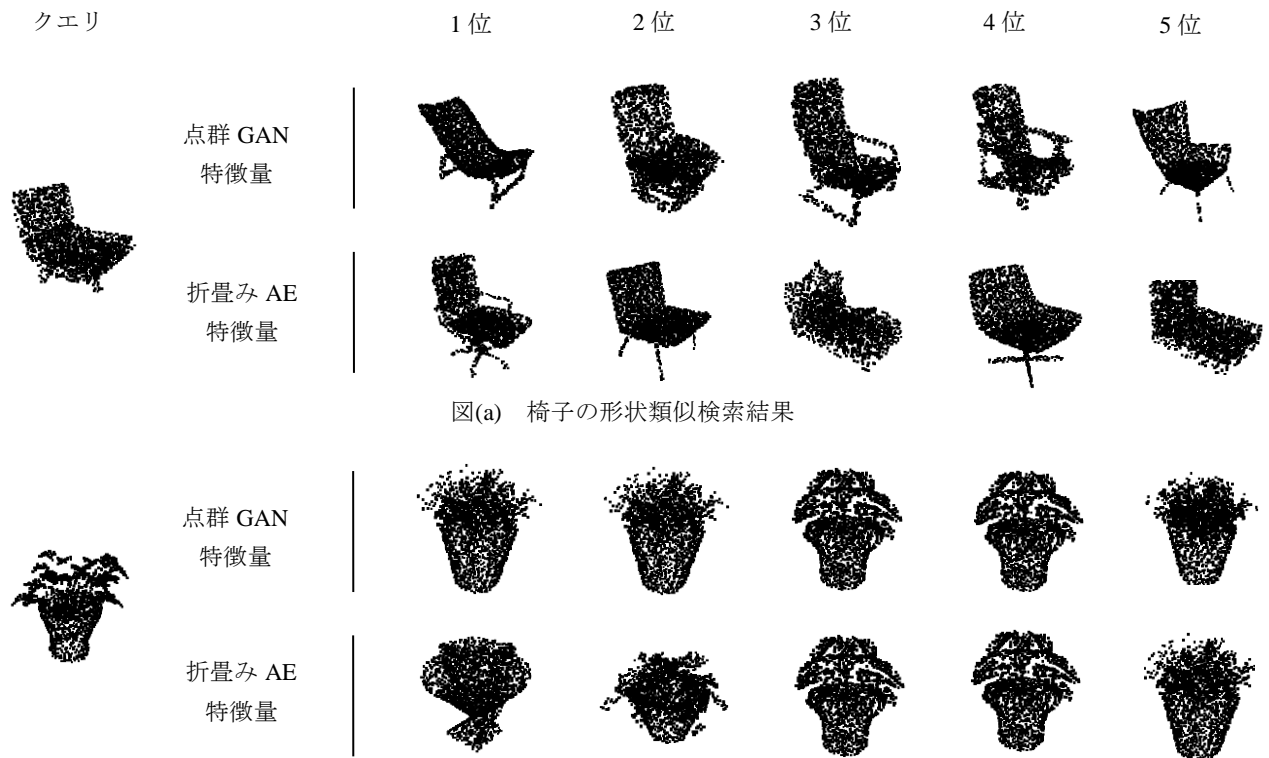
また、表 1 より、点群 GAN 特徴量は一部の教師あり学習形状特徴量に匹敵する検索精度を示すことが分かる。

表 1. 既存特徴量との検索精度 (MAP) 比較。

アプローチ	3D 点群形状特徴量	次元数	MN40	MN10
手作り	D2 [12]	64	0.191	0.286
	AAD [13]	256	0.272	0.357
	SPRH [14]	625	0.329	0.412
	BF-LSF [15]	2,000	0.323	0.385
	BF-SI [16]	2,000	0.352	0.430
	DkSA-RoPS [17]	2,000	0.441	0.600
	DkSA-POD [18]	2,000	0.469	0.605
点群再構成 DNN	折畳み AE [4]	512	0.482	0.677
	全結合 AE [5]	1,024	0.457	0.670
法線推定 DNN	法線推定 PointNet [3]	1,024	0.439	0.635
真贋判別 DNN	<b>点群 GAN (提案法)</b>	1,024	<b>0.563</b>	<b>0.760</b>
カテゴリ分類 DNN	3D ShapeNets [6]	4,000	0.492	0.683
	MV-CNN [20]	128	0.802	—
	DLAN [21]	512	0.850	0.906

4.2.5 検索結果の比較

図 6 に、点群 GAN 特徴量、および折畳み AE 特徴量による検索結果の例を示す。クエリは ModelNet40 の評価用データセットからランダムに選択した。検索対象は、ModelNet40 の学習用データセットである。なお、検索対象データセットには、同一形状の 3D 点群が複数含まれていることがある。図 6 より、点群 GAN 特徴量は折畳み AE 特徴量よりも、意味的に類似した形状を多く検索できることが分かる。例えば、図 6 (a) の椅子の結果において、折畳み AE 特徴量による検索結果にはベッドの形状が複数含まれているが、点群 GAN の場合は上位 5 件すべてが椅子である。



図(a) 椅子の形状類似検索結果

図(b) 植木鉢の形状類似検索結果

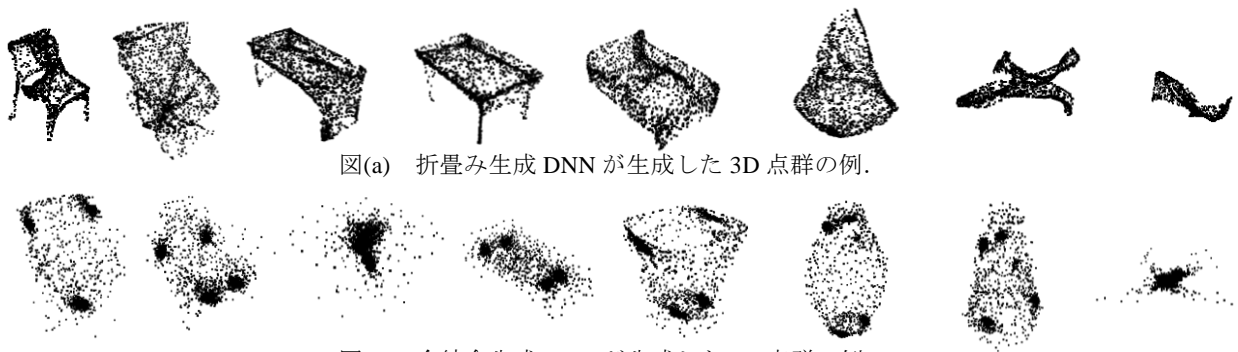
図 6 3D 形状特徴量を用いた検索結果

#### 4.2.6 計算時間

点群 GAN 法による学習と、点群 GAN 特徴量を用いた検索に要する時間を測定する。実験に用いた PC の構成は、CPU が Intel Core i7-6700, GPU が GeForce GTX 1080, DRAM が 64GB である。点群 GAN の学習には ModelNet40 の学習用データセットを用いる。検索実験では、検索対象のデータベースとして、ModelNet40 の学習用データセットを用いる。実験の結果、点群 GAN の学習には、50,000 回の反復に約 50 時間要した。また、検索 1 回当たり約 0.026 秒要した。検索時間の内訳は、クエリ 3D 点群モデル 1 個からの統合特徴抽出に約 0.004 秒、クエリ 1 個と検索対象 9,843 個との統合特徴比較に約 0.022 秒である。

#### 4.2.7 生成された 3D 点群

図 7 に、点群 GAN の生成 DNN によって生成された偽物 3D 点群の例を示す。図 7 の内、(a)は折畳み生成 DNN、(b)は全結合生成 DNN を用いた結果である。図 7(a)の点群はいずれも、点群 GAN の学習後の生成 DNN にランダムな潜在ベクトルを入力することで得た。生成された点群は、物体としての構造が必ずしも正確ではないものの、容器や椅子、ベッドなど多様であり、かつ、ある程度本物らしいことが見て取れる。一方で、図 7(b)の、全結合生成 DNN が生成した点群は、本物らしい形状は捉えているものの、点の配置の偏りが大きい。折畳み生成 DNN が生成した点群は点の配置の偏りが小さく本物らしいことが、特徴抽出の精度向上に貢献したと考えられる。



図(a) 折畳み生成 DNN が生成した 3D 点群の例.

図(b) 全結合生成 DNN が生成した 3D 点群の例.

図 7 点群 GAN が生成した 3D 点群の例.



## 5. まとめと今後の課題

ラベルを持たない多量の 3 次元形状データを効率的に管理するため、これらラベルなしの 3D 形状データから高精度な形状特徴量を獲得する手法が望まれている。本研究では、3D 点群向けの敵対的生成ネットワーク（点群 GAN）を新たに構築し、これを用いた 3D 形状特徴量の教師なし学習手法を提案した。点群 GAN は、3D 点群を生成する DNN（生成 DNN）と、3 次元点群の本物らしさを予測する DNN（判別 DNN）とを互いに競わせながら訓練する過程で、3D 点群の形状特徴を捉える能力を判別 DNN に獲得させる。点群 GAN を効果的に、かつ、安定的に学習させるため、ネットワーク構造と損失関数の双方に対して工夫を施した。具体的には、点の順序変化に影響を受けない 3D 点群解析用 DNN [3] を用いて点群 GAN を構築した。また、Gulrajani ら [1] にならい、本物の点群データ集合と偽物の点群データ集合の集合間距離を学習の損失として用いた。

3D モデルの形状類似検索のシナリオを用いた実験的評価の結果、点群 GAN が獲得した形状特徴量は、既存の手作り形状特徴量、および既存の教師なし学習形状特徴量を上回る検索精度を示すことが分かった。また、ネットワーク構造の工夫、損失関数の工夫、の双方が形状特徴量の精度の改善に効果的であることを確かめた。

今後の課題として、(1) 形状特徴量のさらなる高精度化、(2) 3D 形状の幾何変換に対する頑強性の獲得、を挙げる。現状の点群 GAN のネットワーク構造と損失関数が必ずしも形状特徴量の学習に最適であるとは限らない。ネットワーク構造または損失関数を改良することで特徴量の精度が改善する可能性がある。また、現状の点群 GAN は位置、大きさ、向きが一貫して揃った 3D 点群データを学習に用い、これらを生成・判別するように訓練される。そのため、獲得される形状特徴量がこれらの幾何変換に対する頑強性を持たないと予想される。今後は 3D 形状の回転、スケーリング、平行移動などの幾何変換に対する頑強性を、点群 GAN に獲得させる方法について検討する。

## 参考文献

- [1] Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., and Courville, A. C.: "Improved Training of Wasserstein GANs", *Proc. Advances in Neural Information Processing Systems 2017*, pp. 5769-5779, (2017).
- [2] Arjovsky, M., Chintala, S., and Bottou, L.: "Wasserstein generative adversarial networks", *Proc. International Conference on Machine Learning 2017*, pp. 214-223, (2017)
- [3] Charles, R. Q., Su, H., Kaichun, M., and Guibas, L. J.: "PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation", *Proc. Computer Vision and Pattern Recognition (CVPR) 2017*, pp77-85, (2017).
- [4] Yang, Y., Feng, C., Shen, Y., and Tian, D.: "FoldingNet: Point Cloud Auto-encoder via Deep Grid Deformation", *Proc. Computer Vision and Pattern Recognition (CVPR) 2018*, Vol.3, (2018).
- [5] Achlioptas, P., Diamanti, O., Mitliagkas, I., and Guibas, L.: "Representation learning and adversarial generation of 3d point clouds", *arXiv preprint*, arXiv:1707.02392, (2017).
- [6] Wu, Z., Song, S., Khosla, A., Yu, F., Zhang, L., Tang, X., and Xiao, J.: "3D ShapeNets: A Deep Representation for Volumetric Shapes", *Proc. Computer Vision and Pattern Recognition*, pp. 1912-1920, (2015).
- [7] Kingma, D. P., and Ba, J.: "Adam: A method for stochastic optimization", *Proc. International Conference on Learning Representations (ICLR) 2015*, (2015).
- [8] Maas, A. L., Hannun, A. Y., and Ng, A. Y.: "Rectifier nonlinearities improve neural network acoustic models", *Proc. ICML 2013*, Vol. 30, No. 1, p. 3, (2013).
- [9] Perronnin, F., Sanchez, J., and Mensink, T.: "Improving the Fisher Kernel for Large-Scale Image Classification", *Proc. European Conference on Computer Vision*, pp.143-156, (2010).
- [10] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y.: "Generative Adversarial Nets", *Proc. Advances in Neural Information Processing Systems*, pp. 2672-2680, (2014).
- [11] Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., and Chen, X.: "Improved techniques for training gans", *Proc. Advances in Neural Information Processing Systems 2016*, pp. 2234-2242, (2016).
- [12] Osada, R., Funkhouser, T., Chazelle, B., and Dobkin, D.: "Shape Distributions", *ACM Transactions on Graphics (TOG)*, Vol. 21, No. 4, pp. 807-832, (2002).
- [13] Ohbuchi, R., Minamitani, T., and Takei T.: "Shape-Similarity Search of 3D Models by using Enhanced Shape Functions", *IJCAT*, Vol.23, No.2/3/4, pp70-85, (2005).
- [14] Wahl, E., Hillenbrand, U., and Hirzinger, G.: "Surflet-pair-relation histograms: a statistical 3D-shape representation for rapid classification," *Proc. 3-D Digital Imaging and Modeling, 3DIM 2003*, pp. 474-481, (2003).
- [15] Ohkita, Y., Ohishi, Y., Furuya, T., and Ohbuchi, R.: "Non-rigid 3d model retrieval using set of local statistical features", *Proc. Multimedia and Expo Workshops (ICMEW) 2012*, pp. 593-598, (2012)
- [16] Johnson, A.E., and Hebert, M.: "Using spin images for efficient object recognition in cluttered 3D scenes", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 21, No. 5, pp.433-449, (1999)
- [17] Guo, Y., Shoel, F.A., Bennamoun, M., Wan, J., and Lu, M.: "RoPS: A Local feature descriptor for 3D rigid objects based on rotational projection statistics", *Proc. 1st International Conference on Communications, Signal Processing, and their Applications (ICCSPA) 2013*, pp.1-6, (2013).
- [18] Furuya, T., and Ohbuchi, R.: "Accurate Aggregation of Local Features by using K-sparse Autoencoder for 3D Model Retrieval", *Proc. International Conference on Multimedia Retrieval 2016*, pp293-297, (2016)
- [19] ElNaghy, H., Hamad, S., and Khalifa, M.: "TAXONOMY FOR 3D CONTENT-BASED OBJECT RETRIEVAL METHODS", *International Journal of Research and Reviews in Applied Sciences (IJRRAS)*, 14 (2), pp. 412-446, (2013).
- [20] Su, H., Maji, S., Kalogerakis, E., and Learned-Miller, E.: "Multi-view Convolutional Neural Networks for 3D Shape Recognition," *Proc. ICCV 2015*, (2015).
- [21] Furuya, T., and Ohbuchi, R.: "Deep Aggregation of Local 3D Geometric Features for 3D Model Retrieval," *Proc. BMVC 2016*, pp. 121.1-121.12, (2016).
- [22] Shi, B., Bai, S., Zhou, Z., and Bai, X.: "DeepPano: Deep Panoramic Representation for 3-D Shape Recognition", *IEEE Signal Processing Letters*, 22(12), 2339-2343, (2015).
- [23] Ioannidou, A., Chatzilari, E., Nikolopoulos, S., and Kompatsiaris,

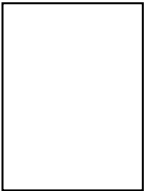
- S., “Deep Learning Advances in Computer Vision with 3D Data: A Survey”, *ACM Computing Surveys*, Volume 50 Issue 2, Article No. 20, (2017).
- [24] Csurka, G., Dance, R.C., Fan, L., Willamowski, J., and Bray, C., “Visual categorization with bags of keypoints”, *Proc. Workshop on Statistical Learning in Computer Vision*, pp. 1-22, (2014).
- [25] Radford, A., Metz, L., and Chintala, S.: “Unsupervised representation learning with deep convolutional generative adversarial networks”, *arXiv preprint*, arXiv:1511.06434, (2015).
- [26] Furuya, T., and Ohbuchi, R., Diffusion-on-Manifold Aggregation of Local Features for Shape-based 3D Model Retrieval, *Proc. ICMR 2015*, pp. 171-178, (2015).

## 著者紹介



### 上西和樹

2018年 山梨大学工学部コンピュータ理工学科卒業。現在、山梨大学大学院医工農学総合教育部修士課程工学専攻コンピュータ理工学コースに在学。



### 古屋貴彦（正会員）

2008年 山梨大学工学部コンピュータ・メディア工学科卒業。2010年 山梨大学大学院医学工学総合教育部修士課程コンピュータ・メディア工学専攻修了。同年ニスカ株式会社（現：キヤノンファインテックニスカ株式会社）入社。2015年 日本学術振興会特別研究員（DC2）。同年 山梨大学大学院医学工学総合教育部博士課程情報機能システム工学専攻修了。同年より山梨大学工学部コンピュータ理工学科助教。興味は、3次元形状や2次元画像等のマルチメディア情報検索、機械学習。



### 大淵竜太郎（正会員）

1981年上智大学工学部電気電子工学科卒業。1983年電気通信大学大学院計算機科学科修士課程修了。同年日本アイ・ビー・エム（株）入社。1994年 University of North Carolina at Chapel Hill より Ph.D. 取得。同年より IBM 東京基礎研究所勤務，主任研究員。1999年山梨大学コンピュータ・メディア工学科助教授，2007年より同教授。興味は，機械学習を用いた，マルチメディア情報，特に3次元形状の，比較・検索・認識・編集など。