

自己注意機構を用いた 3 次元点群の回転不変な解析

刈込 喜大[†] 古屋 貴彦[†] 大淵 竜太郎[†] (正会員)[†]山梨大学

Rotation-Invariant Analysis of 3D Point Set Using Self-Attention

Yoshihiro KARIKOMI[†], Takahiko FURUYA[†], Ryutarou OHBUCHI[†] (Member)[†]University of Yamanashi

〈あらまし〉3次元 (3D) 点群は, 3D センサや CAD モデルから得られる 3D 形状の表現である. 3D 点群の高精度な解析は, 3D モデルの効果的な再利用, 自動運転, ロボットのナビゲーションなどのアプリケーションシナリオで要求される. これらの 3D 点群の向きは不揃いであるため, 多くの現実のアプリケーションでは 3D 形状の回転 (向き) への頑強性が要求される. 近年の 3D 点群解析の主流は深層ニューラルネットワーク (DNN) を用いた手法である. しかし, これら既存の DNN の多くは 3 軸周りの回転への頑強性を持たない. 本研究では回転不変な 3D 形状解析 DNN を提案し評価する. この DNN は, Local Reference Frame を用いた局所形状の回転正規化により回転不変性を獲得し, 自己注意機構によって入力データ適応的に特徴を抽出する. 3D 点群形状の識別とセグメンテーションで評価した結果, 提案手法は 3 軸周り回転の影響を受けず, かつ既存手法を上回る精度を得た.

キーワード: 回転不変性, 3 次元形状識別, 3 次元形状セグメンテーション, 深層ニューラルネットワーク

〈Summary〉Three-dimensional (3D) point set, or point cloud, is a 3D shape representation that can be captured by a 3D sensor or derived from a CAD model. Accurate analysis of 3D point sets is required for effective reuse of 3D models, or navigation and control of autonomous vehicles and robots. These 3D point sets are in general not aligned rotationally. But many applications of these 3D point sets require robustness against rotation of 3D shapes. Recent 3D point set analysis relies on Deep Neural Networks (DNN), yet most of these DNN are not robust against rotation. In this paper, we propose and evaluate a rotation invariant 3D shape analysis DNN. The DNN combines rotation normalization of local geometry using local reference frame with content adaptive feature extraction via self-attention mechanism. We evaluate the DNN on both shape classification and segmentation of 3D point sets. The proposed method is invariant to rotation about 3 axes while outperforming existing methods in accuracy.

Keywords: rotation invariance, 3D shape recognition, 3D shape segmentation, deep neural network

1. はじめに

3 次元 (3D) 形状の表現の 1 つである 3D 点群は, 形状表面をサンプルして得た点の集合であり, 3D センサや CAD モデルから得られる. 3D 点群を高精度に解析する技術は自動運転や CAD モデルの再利用などの観点から注目されている.

近年の 3D 点群解析では深層ニューラルネットワーク (DNN) を用いた手法が主流で, 高い精度を得ている^{1)~10)}. しかし, これら既存の 3D 点群解析 DNN のほとんどは 3D 形状の回転への頑強性を考慮せずに設計され, 評価されている. 例えば, これら手法の評価には, 学習データと向きが揃った 3D 点群データを用いる. しかし, 現実の CAD モデルやセンサから得られる点群は向きが不揃いの場合が多く, 多くの実

用シナリオでは回転頑強性が要求される.

DNN を用いた手法で回転への頑強性獲得を目指すとき, 最初に思いつくのは学習データの回転拡張である. しかし, 3D 形状の 3 軸周り回転はその自由度が高く, 既存の 3D 点群解析 DNN を回転拡張した形状データで学習しても回転頑強性を獲得するのは難しい. そのため, 近年, 回転不変 (RI) 性を持つ 3D 点群解析 DNN が複数提案されている. それらは RI 性の獲得方法に基づき, (1) 本質的に RI な特徴を利用する手法^{11), 12)}, (2) 点群全体を固有ベクトルで回転正規化する手法^{13)~15)}, (3) 局所点群を Local Reference Frame (LRF) で回転正規化する手法^{16)~19)}, に大別できる. (1) は点対の距離や角度などの本質的に RI な特徴を抽出し, DNN への入力とする. しかし, 本質的な RI 特徴は入力点群の形状情報の損失があり, 精度に

課題がある。(2)は入力点群全体の座標値の共分散行列を固有値分解して得た固有ベクトルを大域的な参照座標軸として入力点群全体を回転正規化し、これを点群解析 DNN への入力とする。回転正規化した点群から DNN で直接特徴を抽出するため、(1)のような形状情報の損失は起きにくい。しかし、全体形状は多様でバラつきが大きく、回転正規化は困難である。また、入力点群が解析対象物(例:椅子)と背景(例:床や壁)で構成されると、固有ベクトルが背景の影響を受け、回転正規化に失敗しやすい。(3)の LRF とは、注目点ごとにその近傍(局所)点群の幾何形状情報を基に決定する座標系のことである。LRF には、局所点群の座標値の共分散行列を固有値分解して得た固有ベクトルを用いることが多い。全体形状と比べ、局所形状は単純かつその多様性が低く、また背景の影響も限定的であるため、向きを揃えやすい。よって、(3)の LRF による局所点群の回転正規化は(2)の点群全体の回転正規化よりも失敗しにくく、高精度な 3D 点群解析に有効な局所 RI 特徴を抽出できる見込みがある。しかし、LRF を用いる既存手法は、抽出した局所 RI 特徴群全体の 3D 空間における配置を考慮しない。そのため、既存手法は形状を構成する部品(局所形状)だけに注目しており、点群全体の形状情報が得られていない。また、既存手法の LRF は、局所点群の座標値から求めた固有ベクトルなどの、局所幾何形状情報のみに依存して構築される。そのため、類似する局所形状でもその点の分布によって構築される LRF の軸の向きにばらつきが生じ、精度に悪影響を及ぼす恐れがある。さらに、既存手法の特徴集約・精製にはグラフ畳み込みが用いられており、個々の入力形状ごとに異なる特徴間の相対的な重要度を考慮していない。

本研究では RI 性を持つ高精度な 3D 点群解析 DNN の構築を目的とし、LRF による局所点群の回転正規化に、学習による LRF の幾何調整、自己注意 (Self-Attention: SA) 機構²⁰⁾を用いた局所 RI 特徴の抽出、それら特徴群全体の 3D 空間における配置情報の抽出、を組み合わせる。提案手法では、SA 機構により、入力形状データに適応して求まる、個々の特徴間の相対的重要度に応じた重み、により特徴の集約・精製を行う。

提案手法を 3 軸ランダム回転した 3D 点群形状の識別タスクとセグメンテーションタスクで評価した。その結果、提案手法は回転の影響を受けず、また、その精度は比較対象の既存手法を上回ることが分かった。さらに ablation study により、学習による LRF の幾何調整、局所 RI 特徴群全体の配置情報、SA 機構によるデータ適応的な特徴抽出、いずれもが 3D 点群形状の識別精度向上に有効であることを確認した。

本論文の貢献は以下のとおりである。

- LRF による回転正規化、学習による LRF の幾何調整、SA 機構を用いた局所 RI 特徴の抽出とそれら特徴群全体の 3D 空間における配置情報の抽出、を組み合わせた DNN を提案。
- 提案 DNN は 3D 点群の識別タスクとセグメンテーションタスクで既存手法を上回る精度を記録。
- LRF による回転正規化、学習による LRF の幾何調整、SA 機構、局所 RI 特徴群全体の 3D 空間における配置情報、の有効性を確認。

2. 関連研究

2.1 回転頑強性を持たない 3D 点群解析 DNN

PointNet¹⁾は 3D 点群を直接処理する最初の 3D 点群解析 DNN である。点群を DNN で扱う上では点の入力順序への不変性を獲得する必要がある。PointNet は、まず、入力点群の各点を複数の全結合層 (FC) で高次元の特徴に変換する。そして、得られた点ごとの特徴の集合を(点の入力順を無視して)、max pooling により統合し、点の入力順に不変な点群全体の形状特徴を得る。点ごとの特徴は 1 点のみから抽出するため、局所を成す複数点の参照が必要となる局所形状情報は抽出されない。PointNet はまた、T-Net と呼ばれる DNN をその一部に含む。T-Net は入力点群全体の回転やスケール等の幾何変換を学習に基づいて予測する DNN で、その出力を用いて入力点群に加えられたであろう幾何変換の正規化を試みる。

PointNet 以降、PointNet で対象とされなかった局所形状情報を抽出することで精度向上を狙う手法が複数提案された。PointNet++²⁾は Farthest Point Sampling (FPS) による代表点選択と、代表点近傍に対する局所的な PointNet の適用を反復することで、入力 3D 点群の階層的な局所特徴を抽出する。PointMLP³⁾は PointNet++ をより深い構造として精度向上を狙う。PointMLP は、多数の特徴抽出層を効果的に学習するために残差スキップ接続²¹⁾を導入する。PointNeXt¹⁰⁾も、PointMLP と同様に、PointNet++ をより深い構造として、残差スキップ接続を導入する。PointNeXt はさらに、局所点群のスケールが小さいと学習を難しくすることを指摘し、局所点群のスケールを手で拡大してから特徴を抽出した。DGCNN⁵⁾は近傍特徴と注目点特徴の差分ベクトルから特徴を抽出する EdgeConv を提案、近傍特徴の情報を注目点特徴に集約する。PointCNN³⁾は注目点からの相対位置に応じて近傍点群に順序付けを行い、通常の畳み込みを行う。これらの手法は PointNet で対象とされなかった局所形状情報を抽出することで PointNet を上回る精度を記録する。しかし、これらの手法の特徴抽出処理は FC と max pooling で構成されるグラフ畳み込みか、順序付けを行ってから適用する畳み込みである。これらの特徴抽出機構の重みは事前の学習データで決まり、個々の入力点群の特徴は反映されない。

PointCloudTransformer⁶⁾(PCT)、PointTransformer⁷⁾(PT) は特徴の抽出に SA 機構²⁰⁾を用いる 3D 点群解析 DNN である。SA 機構は注目特徴と集約対象の特徴の間の関連度を求め、その関連度に基づいて特徴を集約する。畳み込みニューラルネットワーク (CNN) の畳み込みやグラフ畳み込みで用いる重みは、学習の後、推論対象となる入力データに関わらず固定である。これに対し、SA 機構の場合、推論対象となる入力デー

タに適応的に重みが変わる利点がある。PCT はグラフ畳み込みによって入力点群の局所特徴群を抽出後、得られた局所特徴群に対し大域的に SA 機構を適用、入力点群全体の特徴を精製する。PT は SA 機構の適用範囲を注目点近傍に限定し、局所特徴を精製する。PCT, PT は 3D 点群解析に SA 機構を用いることの一定の有効性を示した。

前述した手法はいずれも入力点群の座標値から直に特徴抽出を行うため、個々の特徴は入力 3D 空間における位置情報を保有する。したがって、抽出した特徴群をグローバルに max pooling することで特徴群全体の空間的配置を考慮できる。しかし、これら手法はいずれも回転頑強性を持たない。学習データの回転拡張は DNN に一定の回転頑強性を与えるが、自由度の高い 3 軸周り回転に回転拡張のみで対応するのは難しい。PointNet 内蔵の T-Net も 3 軸周り回転を完全には補正できず、精度低下が起こる。提案手法は、前述した手法群と異なり、回転正規化が容易な局所点群を回転正規化し、RI 性を獲得する。そのため、入力点群形状の回転の影響を受けない。

2.2 回転不変 (RI) 性を持つ 3D 点群解析 DNN

回転への頑強性を獲得するため、近年、RI 性をもつ 3D 点群解析 DNN が提案されている。RICov¹¹⁾は、注目点とその近傍との L2 距離、局所点群重心を基準としたなす角、などの本質的に RI な特徴を DNN への入力として RI 性を獲得する。RICov++¹²⁾は RICov の入力 RI 特徴を拡張し、新たに 2 点の法線がなす角などを含む、よりリッチな RI 特徴を用いる。しかし、これら本質的に RI な特徴は、その RI 性と引き換えに入力点群の形状情報の損失があり、精度に限界がある。

Xiao らの手法¹³⁾や Li らの手法¹⁵⁾は、入力点群全体を回転正規化してから従来の点群解析 DNN に入力する。回転正規化には全入力点の座標値の共分散行列を固有値分解して得た固有ベクトルを用いる。これらの手法は回転正規化した点群から直に特徴を抽出するため、本質的に RI な特徴を用いる手法に見られる形状情報の損失は発生しない。しかし、全体形状の複雑さや解析対象物の背景が固有ベクトルに影響し、回転正規化に失敗しやすい。

LRF は注目点ごとにその近傍の局所幾何形状情報を用いて構築する局所座標系である。LRF には、注目点近傍の座標値の共分散行列を固有値分解して得た固有ベクトルを用いる場合が多い。LRF による局所形状の回転正規化は、全体形状の回転正規化より容易である。全体形状と比べ、局所形状の多様性が低く、また局所に限ることで背景の影響が限定的になるためである。そのため、LRF による局所形状の回転正規化は高精度な解析に有効な局所 RI 特徴の抽出が期待できる。DLAN¹⁶⁾と RMGNet¹⁷⁾は LRF によって回転正規化した局所点群から PoD²²⁾という手作り RI 特徴群を抽出して DNN への入力とする。しかし、PoD などの手作り RI 特徴の利用は、本質的に RI な特徴を用いる手法と同様、入力点群形状の幾何形状情報損失を引き起こす恐れがある。RI-GCN¹⁸⁾は、LRF によ

って回転正規化した局所点群に直接 FC と max pooling を適用し、RI 特徴を抽出する。抽出した局所 RI 特徴群は FC によって精製後、max pooling で順序不変に統合する。AECNN¹⁹⁾も LRF によって回転正規化した局所点群から直に RI 特徴を抽出する。AECNN では、得られた個々の局所 RI 特徴について、その近傍の RI 特徴をグラフ畳み込みによって集約・精製する。その後は、精製した局所 RI 特徴群を max pooling で点の順序不変に統合する。AECNN と RI-GCN は回転正規化した局所点群から直に特徴抽出を行うため、入力点群形状の情報損失は抑えられている。しかし、その特徴抽出機構は局所に限定したグラフ畳み込みであり、解析しようとする入力データに適応せず、事前に学習して得た固定の重みで特徴抽出する。さらに、個々の局所 RI 特徴はそれぞれ個別の LRF 空間から抽出されるため、入力 3D 空間における位置情報を持たない。したがって、それら RI 特徴群を max pooling で集約しても特徴群全体の 3D 空間における配置は考慮されない。(これに対し、RI 性を持たない点群解析 DNN で特徴を max pooling で集約する場合には特徴群全体の 3D 空間における配置が考慮される。) また、既存手法の LRF 構築法はいずれも局所点群の幾何形状情報だけに依存する。よって、類似する局所形状でも、その点の分布や形状のわずかな違いで、構築される LRF の軸の向きがばらつき、精度に悪影響を及ぼす恐れがある。

提案手法は AECNN や RI-GCN と同様に LRF を用いて回転正規化した局所点群から特徴を抽出する。そのうえで、LRF を用いる既存手法に残る課題である、特徴間の相対的重要度に基づく特徴抽出、局所 RI 特徴群全体の 3D 空間における配置情報の抽出、局所点群の幾何形状情報だけに依存しない LRF の構築、を導入し精度向上を目指す。

3. 提案手法

3.1 提案 DNN の概要

提案 DNN は、局所点群に対し LRF を用いて回転正規化を施した局所点群から直に局所 RI 特徴群を抽出し、得られた局所 RI 特徴群の集約と精製を繰り返す。提案 DNN は LRF の構築処理、及び局所 RI 特徴群の集約・精製処理の双方に工夫点を備える。LRF 構築では、局所点群の幾何形状情報を用いて回転について正規化した第 1 段階の LRF を基とし、これに T-Net と呼ばれる DNN が推論した幾何変換行列を乗じる第 2 段階の幾何変換補正を組み合わせる。この 2 段階処理により、LRF を成す 3 軸の向きとスケールを幾何調整する。T-Net による LRF の向き調整により、類似する局所形状間でもその点の分布によって LRF の軸の向きがばらつき問題を緩和する。また、LRF のスケール調整により、局所点群のスケールが小さいことに起因する学習の難しさ (Qian ら¹⁰⁾が指摘) を低減し、DNN の効果的な学習を促す。

2 段階正規化によって回転正規化とスケール調整を施した局所点群から抽出した局所 RI 特徴群の集約と精製には SA 機

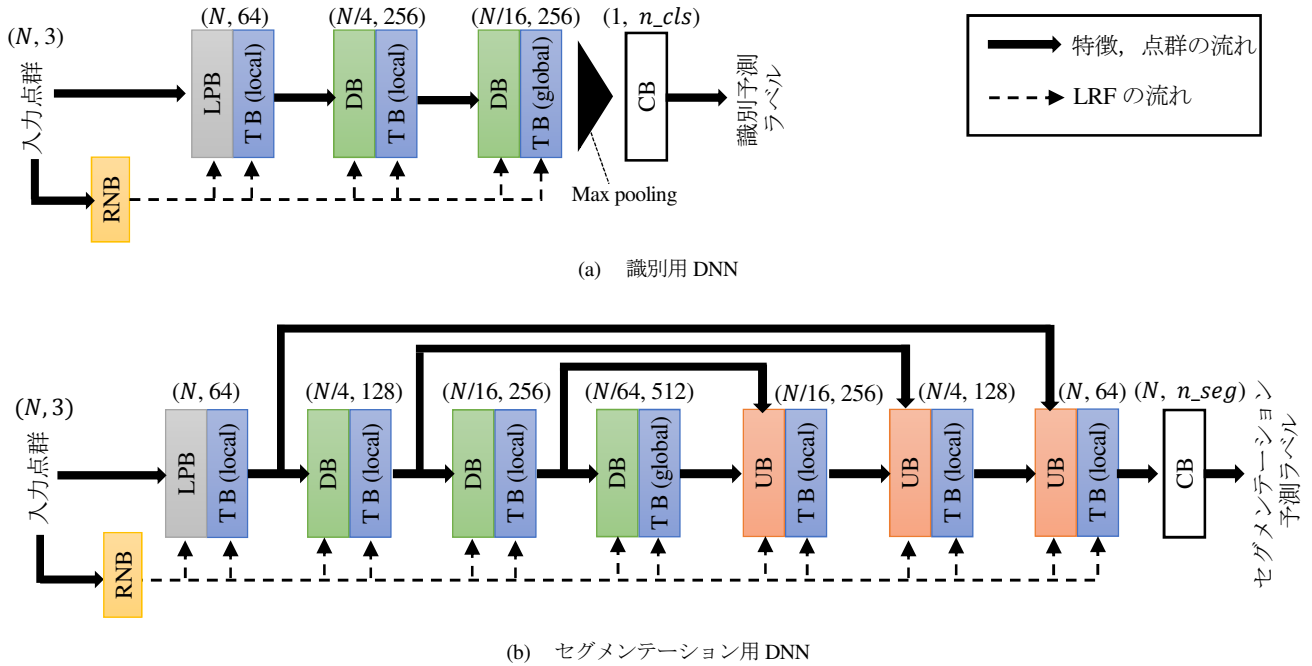


図 1 提案する回転不変な DNN の全体像

Fig.1 Network architectures for proposed rotation invariant classification DNN (top) and segmentation DNN (bottom)

構²⁰⁾を用いる。SA 機構の個々の入力特徴群に適応して求める重みによって高精度な特徴の集約・精製ができると期待される。提案手法では SA 機構による特徴群の集約・精製を、入力 3D 形状の局所に加え、全体に対しても行う。局所に対する SA 機構では、互いに近い局所 RI 特徴群をその関係を考慮して特徴集約・精製する。3D 形状全体に対する SA 機構では、すべての局所 RI 特徴相互の関係、すなわち局所 RI 特徴群全体の 3D 空間における配置情報の獲得が期待される。

LRF で RI 性を獲得する既存の DNN 群と比べ、提案 DNN は (1) LRF の向きとスケールを学習によって幾何調整する、(2) SA 機構を用いて個々の入力データに適応的に RI 特徴精製を行う、(3) 局所 RI 特徴群全体の 3D 空間における配置情報を考慮する、の 3 点に新規性がある。

3.2 DNN アーキテクチャ

図 1 に識別用とセグメンテーション用の提案 DNN の全体像を示す。図中のカッコ内の数値は特徴及び LRF のデータ形状で、 N は入力点数、 n_{cls} は識別クラス数、 n_{seg} はセグメンテーションクラス数を表す。まず、Rotation Normalization Block (RNB) で入力点群の各点の LRF の構築と学習による幾何調整を行う。点ごとの近傍点群をこの LRF で回転正規化した結果に Local PointNet Block (LPB) を適用し局所 RI 特徴を得る。これら局所 RI 特徴群に対し、SA 機構を用いた特徴抽出層である Transformer Block (TB) の局所での適用、点数を削減する Downsample Block (DB) の適用、を対にして反復する。ローカルな TB (local) と DB により階層的な局所 RI 特徴を抽出しつつ、点数を十分に削減した段階でグローバルな TB (global) を

適用し、局所 RI 特徴群全体の 3D 空間における配置情報を持つ RI 特徴を得る。識別用 DNN では、これら RI 特徴群を max pooling により点の入力順に不変に統合した後、複数の FC で構成される Classification Block (CB) で識別予測ラベルに変換する。セグメンテーション用 DNN では、識別用 DNN と類似の DNN で潜在特徴を得た後、点数を復元する Upsample Block (UB) とローカルな TB の適用を繰り返し、入力点群の各点に対応する RI 特徴を得る。得られた各点の RI 特徴は CB でセグメンテーション予測ラベルに変換する。なお、セグメンテーション用 DNN では、点数削減前の高解像度の特徴を DB からスキップ接続で流し、UB の入力に接続することで、セグメンテーション予測ラベルの空間解像度を維持する。

3.2.1 Rotation Normalization Block (RNB)

RNB (図 2) は LRF の構築と幾何調整を行う。LRF の幾何調整には幾何変換行列を予測するネットワーク T-Net を用いる。図 2 のカッコ内は処理後の特徴と LRF のデータ形状である。**LRF の構築:** 提案手法の LRF は注目点 k 近傍から計算した固有ベクトルと、注目点 k 近傍の重心を用いて構築する。まず、注目点 $\mathbf{p}_i$ の k 近傍の座標値の共分散行列を固有値分解して固有ベクトル群を得る。そして、得られた固有ベクトル群の最小の固有値と対応する固有ベクトル $\mathbf{e}_3$ を注目点の LRF の z 軸 ($\mathbf{lra}_z$) に割り当てる。LRF の x 軸 ($\mathbf{lra}_x$) には、 $\mathbf{lra}_z$ に直交する面に投影される注目点 k 近傍の重心 $\bar{\mathbf{p}}_i$ への注目点 $\mathbf{p}_i$ からの単位ベクトルを割り当てる。LRF の y 軸 ($\mathbf{lra}_y$) は $\mathbf{lra}_z$ と $\mathbf{lra}_x$ の外積で求める。式 (1)–(3) に提案 LRF の 3 軸及びその計算方法を示す。なお、固有値分解で得た固有ベクトルには符号の

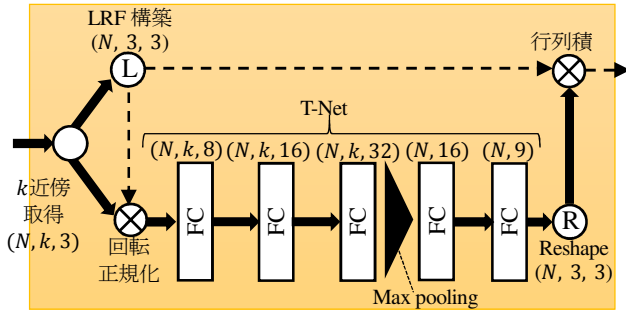


図2 Rotation Normalization Block (RNB)の構造

Fig.2 Structure of Rotation Normalization Block (RNB)

曖昧さがある。そのため、 $\mathbf{lra}_z$ に割り当てる $\mathbf{e}_3$ の符号曖昧さを除去する必要がある。符号の曖昧さは、注目点から近傍点群へのベクトルと $\mathbf{e}_3$ の内積を用いて除去する。内積が0以上である近傍点数が、 $\mathbf{e}_3$ の符号を入れ替えた場合よりも多ければそのままの符号を維持し、そうでなければ符号の正負を入れ替える。

注目点 $\mathbf{p}_i$ の近傍点 $\mathbf{p}_j$ の回転正規化 $\mathbf{p}_j'$ は、 $\mathbf{p}_j$ の位置を $\mathbf{p}_i$ で正規化した後に前述の手順によって得られた注目点のLRFを行列積することで得られる(式(4))。

$$\mathbf{LRF}_i = (\mathbf{lra}_x, \mathbf{lra}_y, \mathbf{lra}_z) \quad (1)$$

$$\mathbf{lra}_x = \frac{\mathbf{v}_x}{\|\mathbf{v}_x\|}, \quad \mathbf{lra}_y = \mathbf{lra}_z \times \mathbf{lra}_x, \quad \mathbf{lra}_z = \mathbf{e}_3 \quad (2)$$

$$\mathbf{v}_x = (\bar{\mathbf{p}}_i - \mathbf{p}_i) - \mathbf{lra}_z ((\bar{\mathbf{p}}_i - \mathbf{p}_i) \cdot \mathbf{lra}_z) \quad (3)$$

$$\mathbf{p}_j' = (\mathbf{p}_j - \mathbf{p}_i) \mathbf{LRF}_i \quad (4)$$

T-Net: 提案手法では、局所点群を対象としたLRFによる回転正規化の結果に対し、T-Netを追加で適用し、回転やスケールなどの幾何変換の補正を行う。

T-NetはFCとmax poolingで構成される、回転やスケールなどの幾何変換行列を予測するネットワークである。T-NetはPointNet¹⁾で提案された。提案手法ではLRFによって回転正規化した局所点群にT-Netを適用、幾何変換行列を予測する。予測した幾何変換行列はLRFと行列積し、LRFを補正する。

PointNetのT-Netは入力点群全体を対象として回転などの幾何変換を推定し、補正しようとする。しかし、全体を表す点群は形状が複雑かつ多様であり、また3軸周りの回転の自由度は高く「正規な向き」の学習と予測が非常に困難である。その結果、T-Netによる、全体形状を対象とした、回転を含めた幾何変換の補正はうまく機能しない。これに対し提案手法では、局所点群の形状を対象として2段階の回転正規化を行う。まず、局所点群を対象とし、LRFを用いた回転正規化を施す。次に、その結果に対し、T-Netを用いた回転やスケールなどの補正を追加で適用する。全体形状の点群と比較し、局所点群は形状が単純であり、かつその多様性も低い。また、T-Netの役割は、LRFでほぼ回転正規化された後の小さな補正

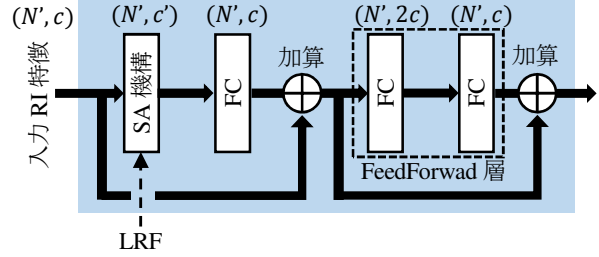


図3 Transformer Block (TB) の構造

Fig.3 Structure of Rotation Transformer Block (TB)

にとどまる。これにより、第1段階のLRFによる回転正規化と第2段階のT-Netによる幾何変換補正の組合せが効果的に機能すると期待される。

3.2.2 Local PointNet Block (LPB)

LPBでは注目点近傍の局所RI特徴を抽出する。まず、注目点の一定半径内に含まれる近傍点群を取得する。その後、それらを注目点のLRFで回転正規化し、回転正規化した局所点群からFCとmax poolingを用いて特徴を抽出する。提案DNNではこの処理を入力点群のすべての点について行う。式(5)にLPBの処理を示す。 $\mathbf{LRF}_i$ は注目点 $\mathbf{p}_i$ のLRF、 $R(\mathbf{p}_i)$ は注目点 $\mathbf{p}_i$ の一定半径内に含まれる近傍点のインデックス集合、maxはチャンネルごとのmax pooling、 ψ は2層のFCであり、 ω は1層のFCである。

$$\mathbf{y}_i = \omega \left(\max_{j \in R(\mathbf{p}_i)} \left(\psi((\mathbf{p}_j - \mathbf{p}_i) \mathbf{LRF}_i) \right) \right) \quad (5)$$

3.2.3 Transformer Block (TB)

TBはSA機構とFeed Forward (FF)層で構成される特徴の集約・精製ブロックである。提案手法ではSA機構の適用範囲を k 近傍に限定するローカルなTBと、適用範囲を制限しないグローバルなTBの2種類を用いる。ローカルなTBでは局所的なRI特徴を抽出し、グローバルなTBでは局所RI特徴群全体の3D空間における配置関係を捉える。ダウンサンプルしていない点ごとの特徴に対するグローバルなTBの適用は計算量とメモリ消費が大きい。そのため、グローバルなTBは後述するDBによって点数を削減してから適用する。図3にTBの構造を示す。図中のカッコ内の数値は特徴のデータ形状であり、 N' は入力RI特徴の点数、 c 及び c' は特徴次元数である($c > c'$)。TBのSA機構は計算量とメモリ消費量削減のため、ボトルネック構造とし、入力特徴次元数よりも小さな次元数で計算する。また、TBのSA機構とFF層には残差スキップ接続²⁾を適用する。

SA 機構: SA機構は注目点特徴との関連度に基づいて特徴を集約する処理である。従来のRI性を持つ点群解析DNNでは、FCとmax poolingで構成されるグラフ畳み込みにより、個々の入力データに関わらず不変の重みでRI特徴の集約処理を

行う。提案手法では、SA 機構を用いることにより、個別の入力特徴に適応的に変化する重みで RI 特徴の集約を行う。提案手法の SA 機構を式 (6) に示す。この SA 機構は Zhao ら²³⁾が提案したベクトルの SA 機構に基づく。式中の $\chi(\mathbf{p}_i)$ は集約対象の点のインデックス集合 (注目点自身を含む)、 $\odot$ は要素積である。注目点の SA 機構の出力特徴 $\mathbf{y}_i$ は、注目点特徴 $\mathbf{x}_i$ と集約対象の特徴 $\mathbf{x}_j$ の関連度 $\text{softmax}(\alpha(\mathbf{x}_i, \mathbf{x}_j))$ で、 $\mathbf{x}_j$ を精製した特徴である $\beta(\mathbf{x}_j)$ を重み付けした総和である。関連度 $\text{softmax}(\alpha(\mathbf{x}_i, \mathbf{x}_j))$ は注目点特徴と集約対象の特徴を参照して初めて決定するため、個々の入力特徴に適応的に変化する。なお、 softmax は $\alpha(\mathbf{x}_i, \mathbf{x}_j)$ のチャンネルごとに適用する。

$$\mathbf{y}_i = \sum_{j \in \chi(\mathbf{p}_i)} \text{softmax}(\alpha(\mathbf{x}_i, \mathbf{x}_j)) \odot \beta(\mathbf{x}_j) \quad (6)$$

関連度を決める $\alpha(\cdot)$ 、及び近傍点特徴を精製する $\beta(\cdot)$ は学習によって決まり、それぞれ式 (7) と (8) で定義される。式中の $Q(\cdot)$ 、 $K(\cdot)$ 、 $V(\cdot)$ は 1 層の FC、 $\gamma(\cdot)$ と $\delta(\cdot)$ は 2 層の FC である。 $\mathbf{PI}_{ij}$ は後に述べる相対位置特徴である。

$$\alpha(\mathbf{x}_i, \mathbf{x}_j) = \gamma(Q(\mathbf{x}_i) - K(\mathbf{x}_j) + \mathbf{PI}_{ij}) \quad (7)$$

$$\beta(\mathbf{x}_j) = \delta(V(\mathbf{x}_j) + \mathbf{PI}_{ij}) \quad (8)$$

個々の特徴はそれぞれ異なる LRF 空間から抽出されている。そのため、SA 機構で集約する特徴は注目点との相対的な位置情報を持たない。そこで注目点と集約対象の点の位置関係を考慮する目的で位置エンコーディング (PE) を行う。提案手法では PE として、注目点の LRF で回転正規化した集約対象の点の 3D 座標 (相対位置) から抽出した特徴を式 (7) と (8) で加算する。以下の式 (9) に加算する相対位置特徴 $\mathbf{PI}_{ij}$ を示す。 $\theta(\cdot)$ は FC、 $\mathbf{p}_i$ と $\mathbf{p}_j$ はそれぞれ注目点と近傍点の 3D 座標、 $\mathbf{LRF}_i$ は注目点の LRF である。

$$\mathbf{PI}_{ij} = \theta((\mathbf{p}_j - \mathbf{p}_i)\mathbf{LRF}_i) \quad (9)$$

Feed Forward 層: FF 層は 2 層の FC で、SA 機構後の個々の特徴を精製する。FF 層では 1 層目の FC で特徴次元数を入力特徴次元の 2 倍にした後、2 層目の FC で元に戻す。その後、SA 機構と同様に残差スキップ接続する。

3.2.4 Downsample Block (DB)

DB では点のダウンサンプルを行い、後続する TB での計算量とメモリ消費を抑制しつつ、より広範囲な形状特徴の抽出を行いやすくする。DB では代表点を選択し、代表点の 3D 空間における k 近傍の特徴を代表点特徴に集約し、出力する。推論時の代表点は FPS で選択する。学習時の代表点選択は、識別タスクではランダム、セグメンテーションタスクでは FPS である。識別タスク学習時にランダムに選択するのは汎化性能向上が狙いである。セグメンテーションタスク学習時に FPS で選択するのは、後述する UB において、点の分布の偏りが悪影響を及ぼす恐れがある為である。

式 (10) に DB の処理を示す。ここで $\zeta(\cdot)$ 、 $\eta(\cdot)$ は FC であり、 $\chi(\mathbf{p}_i)$ は代表点の k 近傍のインデックス集合 (代表点自身を含む)、 $\mathbf{x}_j$ は集約対象の近傍特徴、 $\mathbf{PI}_{ij}$ は代表点を基準とした式 (9) と同じ相対位置特徴、 $\max$ はチャンネルごとの max pooling である。

$$\mathbf{y}_i = \eta\left(\max_{j \in \chi(\mathbf{p}_i)} (\zeta(\mathbf{x}_j + \mathbf{PI}_{ij}))\right) \quad (10)$$

3.2.5 Upsample Block (UB)

UB はセグメンテーション用 DNN で用いる、DB で削減した点数を元に戻すブロックである。UB ではまず、点数削減前の各点の 3D 空間における k 近傍を、現在残っている点群から取得する。その後、それら近傍点群の RI 特徴を集約し、出力する。このとき、DB による削減前の特徴をスキップ接続経路で接続する。また、UB では前述した SA 機構や DB と同様に、PE として相対位置特徴を集約対象の特徴に加算する。式 (11) に UB の処理を示す。ここで $\rho(\cdot)$ 、 $\varrho(\cdot)$ は FC であり、 $\chi(\mathbf{p}_i)$ は注目点 $\mathbf{p}_i$ (削減前の点群の各点) の k 近傍のインデックス集合、 $\mathbf{x}_j$ は集約対象の近傍特徴、 $\mathbf{PI}_{ij}$ は式 (9) と同様の相対位置特徴、 $\mathbf{x}_i'$ はスキップさせた点数削減前の特徴、 $\max$ はチャンネルごとの max pooling、 $\oplus$ はベクトルの接続である。

$$\mathbf{y}_i = \rho\left(\max_{j \in \chi(\mathbf{p}_i)} (\varrho(\mathbf{x}_j + \mathbf{PI}_{ij}))\right) \oplus \mathbf{x}_i' \quad (11)$$

3.2.6 Classification Block (CB)

CB は精製した特徴から入力点群の形状識別予測ラベル、または入力点群の各点のセグメンテーション予測ラベルを出力する。識別用 DNN の CB は 2 層の FC で、その出力次元数は、識別用 DNN の入力側から、256、形状識別クラス数、である。セグメンテーション用 DNN の CB は 4 層の FC で、その出力次元数はセグメンテーション用 DNN の入力側からそれぞれ 256、128、64、セグメンテーションクラス数、である。

3.2.7 実装詳細と実行環境

提案 DNN の FC は、T-Net の最終層、SA 機構の $\gamma(\cdot)$ と $\delta(\cdot)$ の 2 層目、CB の最終層を除き、その出力特徴には Batch Normalization (BN)²⁴⁾ と ReLU を適用する。SA 機構及び FF 層の残差スキップ接続では、出力層の FC からの特徴に BN を適用後、スキップさせた特徴と加算し、ReLU を適用する。SA 機構の関連度 $\text{softmax}(\alpha(\mathbf{x}_i, \mathbf{x}_j))$ と、FF 層の 2 層目の入力特徴には drop 率 0.1 の dropout を適用し、CB 最終層の入力特徴には drop 率 0.5 の dropout を適用する。

提案 DNN は Python 3.8.5、深層学習フレームワークとして TensorFlow (ver. 1.15.4) を用いて実装した。また、共分散行列の固有値分解に numpy (ver. 1.17.3) を使用した。実験に用いたハードウェア構成は、GPU が NVIDIA GeForce RTX 3090、CPU が Intel® Core™ i7-6700 CPU @ 3.40GHz である。

4. 実験

4.1 点群の形状識別実験

回転した無背景点群と有背景点群の識別タスクで提案手法を評価, 既存手法と比較する. 本実験では各入力 3D 点群形状をランダム回転して学習と評価を行う. 回転条件は z/z , $z/SO3$, $SO3/SO3$ の 3 種で, “/” の左辺が学習データ, 右辺が評価データの回転の種類である. z は形状の上向き軸周りの一様ランダム回転を表し, $SO3$ は 3 軸周りの一様ランダム回転を表す.

4.1.1 評価指標

評価指標には後述するデータセットに含まれるテスト用点群データにおける形状識別正解率を用いる. 識別正解率は, 手法ごとに 5 回試行して得た値の平均値とする. 各試行の識別正解率は, 1 エポック学習するごとにテスト用点群データで評価した値のうち, 全学習エポックの最大値とする.

4.1.2 データセット

無背景点群データセット: 無背景点群の形状識別実験では, データセットとして ModelNet40 (MN40)²⁵⁾を用いる. MN40 は 40 カテゴリからなる無背景メッシュ 3D モデルデータセットであり, 9,843 個の学習用モデルと 2,468 個のテスト用モデルで構成される. MN40 の各モデルには解析対象物の背景などは含まれない. 本実験では PointNet¹⁾に倣い, MN40 の各メッシュ 3D モデルを 1,024 点の 3D 点群に変換した点群データを用いる. 各点群データは全ての点が原点を中心とする単位球に収まるように位置と大きさを正規化する. 図 4 (a) は同データセットに含まれるポリゴンモデルの一例を 1,024 点の 3D 点群データ化した結果の例である.

有背景点群データセット: 有背景点群の形状識別実験では, データセットとして ScanObjectNN PB T50 RS (SONN)²⁶⁾を用いる. SONN は 15 カテゴリからなる 3D スキャン点群データセットで, 11,416 個の学習用点群と 2,882 個のテスト用点群より成る. SONN の各 3D 点群形状は解析対象物の背景を含む 2,048 点の点群である. 本実験では, RIConv++¹²⁾に倣い, SONN の各点群形状から FPS によって選択した 1,024 点を使用する. 図 4 (b) は同データセットに含まれるモデルの一例である (参考文献 26)より抜粋). 物体(ベッド)のみならず床や壁などの背景の点も含むことがわかる.

4.1.3 比較対象

本実験では, RI 性を持たない既存の 5 手法, RI 性を持つ既存の 3 手法, 提案手法, の間で比較する. RI 性を持たない既存手法として PointNet¹⁾, PointNet++²⁾, DGCNN⁵⁾, PCT⁶⁾, PointMLP⁹⁾を, RI 性を持つ手法として RIConv¹¹⁾, RIConv++¹²⁾, RI-GCN¹⁸⁾を, それぞれ比較対象とする.

各手法を公平に評価するために, データ拡張手法と学習率スケジューラは統一する. データ拡張手法について, RI 性を

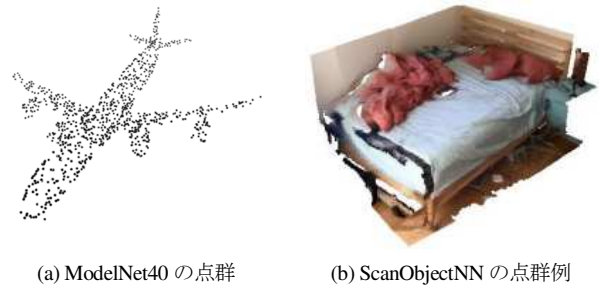


図 4 形状識別実験で使用する点群形状の例

Fig. 4 Examples of point cloud shapes used in classification experiments

持たない手法は x, y, z 軸のそれぞれの方向に $[0.67, 1.5]$ の範囲の一様ランダムな非等方スケールと $[-0.2, 0.2]$ の範囲の一様ランダムな平行移動を行う. 提案手法を含む RI 性を持つ手法では上記と同じ範囲の非等方スケールのみを行い, 平行移動は行わない. これは RI 性をもつ比較対象手法が平行移動への不変性を持つためである. すべての手法で共通して, 正解ラベルにはラベル平滑化 (係数 0.2) を行い, 学習率スケジューラに cosine スケジューラを用いる. 学習率とオプティマイザ, エポック数は手法ごとに相性があるので統一しない.

4.1.4 提案手法の学習設定とハイパーパラメータ

提案 DNN は, ミニバッチサイズ 16 を用い, MN40 では 200 エポック, SONN では 250 エポック学習する. オプティマイザには momentum SGD を用いる. 学習率は MN40 では 0.05 から 0.0005 まで, SONN では 0.03 から 0.0003 まで cosine スケジューラで減衰させる. また, DNN の重み (BN を除く) には L2 正則化を行う. L2 正則化の係数は MN40 では 0.0001, SONN では 0.0002 とする. いずれのデータセットも, LRF 構築のための k 近傍は 64 点を, T-Net 処理のための k 近傍は 32 点を, それぞれ使用する. LPB の近傍点探索で使用する半径は 0.15 とし, ローカルな TB 及び DB で使用する k 近傍は 16 点とする. DNN の各ブロックの出力特徴次元数は図 1 (a) に則る. SA 機構内 (式 (7)–(9)) の FC の出力特徴次元数は, 識別用 DNN の入力側から 3 層ある TB のそれぞれについて, その入力特徴次元数の $1/2, 1/4, 1/8$ とする.

4.1.5 実験結果

無背景点群形状の識別実験: 実験結果を表 1 に示す. 提案手法を含む RI 性を持つ手法はどのような回転条件下にあっても精度の変化は無い (RIConv, RI-GCN) または小さい (RIConv++). その中でも提案手法は, z/z , $z/SO3$, $SO3/SO3$ の 3 つ全ての学習/評価の回転条件下において識別精度が不変で, かつ, 比較した全手法と同等以上の識別精度を示した. これに対し, RI 性を持たない手法, 例えば PointNet や PointMLP は, 学習時に現れない回転が評価時に与えられる $z/SO3$ で識別精度を大きく落とす. 例えば, PointMLP では z/z では 91.3% なのに対し, $z/SO3$ では 39.3% である. また, 様々な回転の例

表 1 無背景点群 (MN40) の形状識別精度 [%]

Table 1 Shape recognition accuracy [%], without background (MN40).				
	RI	z/z	z/SO3	SO3/SO3
PointNet ¹⁾	No	87.8	24.0	82.4
PointNet++ ²⁾	No	90.2	32.7	87.0
DGCNN ⁵⁾	No	90.8	38.6	89.2
PCT ⁶⁾	No	90.9	36.8	89.6
PointMLP ⁹⁾	No	91.1	39.3	90.4
RIConv ¹¹⁾	Yes	88.4	88.6	88.4
RIConv++ ¹²⁾	Yes	91.3	76.4	89.8
RI-GCN ¹⁸⁾	Yes	89.8	89.7	89.8
提案手法	Yes	91.3	91.3	91.3

表 2 有背景点群 (SONN) の形状識別精度 [%]

Table 2 Shape recognition accuracy [%], with background (SONN)				
	RI	z/z	z/SO3	SO3/SO3
PointNet ¹⁾	No	77.6	27.6	46.3
PointNet++ ²⁾	No	82.3	24.2	55.1
DGCNN ⁵⁾	No	85.0	32.5	74.9
PCT ⁶⁾	No	86.1	29.4	76.8
PointMLP ⁹⁾	No	87.8	28.0	79.2
RIConv ¹¹⁾	Yes	73.8	73.9	73.6
RIConv++ ¹²⁾	Yes	84.4	80.5	82.9
RI-GCN ¹⁸⁾	Yes	81.0	81.4	81.0
提案手法	Yes	87.4	87.5	87.4

を学習したはずの SO3/SO3 の条件下であっても, z/z と比較して一定の精度低下がある。

有背景点群形状の識別実験: 実験結果を表 2 に示す。z/z における精度は, RI 性を持たない PointMLP が 1 位, 提案手法は次点である。しかし PointMLP の識別正解率が z/SO3 で大幅に, また SO3/SO3 でも一定程度, 低下する。これに対し提案手法の識別正解率は z/z, z/SO3, SO3/SO3 で変化が無い。

RI 性を持たない手法の SO3/SO3 での精度低下は MN40 より SONN の場合で大きい。NM40 の場合に SO3/SO3 の精度低下が小さい DGCNN, PCT, PointMLP も, SONN では有意に精度が下がる。これは SONN の背景点が原因すると考える。SONN の背景点は床が多い。z/z 回転では床の上向き軸は変化しないため「上向き軸の座標値が小さい点はおそらく床上の点であり, 形状解析に寄与しない。」と容易に学習ができる。しかし SO3/SO3 ではその学習が非常に困難になる。

4.2 点群のセグメンテーション実験

回転した無背景点群のセグメンテーションタスクで提案手法を評価, 既存手法との精度比較を行う。本実験では各入力 3D 点群をランダムに回転させて学習・評価を行う。回転条件は z/SO3, SO3/SO3 である。

4.2.1 評価指標

本実験では評価指標として, 入力点群のカテゴリごとに求めた Intersection over Union (IoU) の平均 Category-level IoU (C-

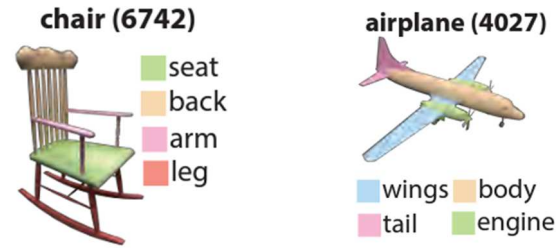


図 5 ShapeNetPart の入力形状とセグメンテーションラベルの例

Fig. 5 Examples of shapes and its segmentation labels in ShapeNetPart

IoU) を用いる。精度評価の試行は 5 回行い, その平均値を提案手法の C-IoU とする。試行ごとの C-IoU は学習を行いつつ, 1 エポックごとにテスト用点群データで評価して得た中の最大値とする。

4.2.2 データセット

本実験では, データセットとして ShapeNetPart²⁷⁾を用いる。ShapeNetPart は 16 カテゴリの無背景メッシュ 3D モデル群を 3D 点群に変換したデータセットであり, 14,007 個の学習用点群と 2,874 個のテスト用点群で構成される。ShapeNetPart の各点群モデルの各点には, オブジェクトのカテゴリによる, 2-5 種のセグメンテーションラベルが割り当てられている。セグメンテーションラベルの総数は 50 で, その中から各点のセグメンテーションラベルを予測する。本実験では, PointNet¹⁾に倣い, 1 形状あたり 2,048 点の点群モデルとする。

図 5 は同データセットに含まれる形状モデルとそのセグメンテーションラベルの例である (参考文献 27)より抜粋)。

4.2.3 比較対象

本実験では, RI 性を持たない既存の 5 手法, RI 性を持つ既存の 6 手法, 提案手法, の間で比較する。RI 性を持たない手法として PointNet¹⁾, PointNet++²⁾, PointCNN³⁾, SpiderCNN⁴⁾, DGCNN⁵⁾, RI 性を持つ手法として RIConv¹¹⁾, RIConv++¹²⁾, Yu らの手法¹⁴⁾, RMGNet¹⁷⁾, RI-GCN¹⁸⁾, AECNN¹⁹⁾を比較対象とする。なお, RI 性を持つ比較対象手法の精度値はその論文が報告する精度値を用いた。RI 性を持たない手法の精度値には RIConv¹¹⁾で報告された精度値を用いた。

4.2.4 提案手法の学習設定とハイパーパラメータ

提案 DNN はミニバッチサイズ 16 で 200 エポック学習する。オプティマイザには Adam²⁸⁾を用い, 学習率は 0.001 から 0.00001 まで cosine スケジューラで減衰させる。学習時は平滑化係数 0.2 でカテゴリごとに取り得るラベルの範囲をラベル平滑化する。学習時は x, y, z 軸のそれぞれの方向に [0.67, 1.5] の範囲の一樣ランダムな非等方スケールリングを行う。LRF 構築のための k 近傍は 64 点とし, T-Net 処理のための k 近傍は 32 点を使用する。LPB の近傍点探索で使用する半径は 0.10 とし, ローカルな TB 及び DB, UB で使用する k 近傍は 16 点とする。

表3 セグメンテーション精度 (C-IoU [%]) (ShapeNetPart)

Table 3 Segmentation accuracy in C-IoU [%] (ShapeNetPart)

	RI	z/SO3	SO3/SO3
PointNet ¹⁾	No	37.8	74.4
PointNet++ ²⁾	No	48.2	76.7
PointCNN ³⁾	No	34.7	71.4
SpiderCNN ⁴⁾	No	42.9	72.3
DGCNN ⁵⁾	No	37.4	73.3
RIConv ¹¹⁾	Yes	75.3	75.5
RIConv++ ¹²⁾	Yes	80.3	80.3
Yu et al. ¹⁴⁾	Yes	79.2	79.4
RMGNet ¹⁷⁾	Yes	81.5	81.4
RI-GCN ¹⁸⁾	Yes	77.0	77.0
AECNN ¹⁹⁾	Yes	80.2	80.2
提案手法	Yes	81.9	81.9

表4 回転正規化の有無, T-Netの有無, と形状識別正解率 [%]

Table 4 Shape recognition accuracy [%] with and without Rotation Normalization and T-Net

回転正規化 の有無	T-Net の 有無	MN40	SONN
なし	なし	34.5±1.0	30.3±1.1
なし	あり	32.8±1.4	28.6±1.5
あり	なし	91.0±0.2	86.1±0.2
あり	あり	91.3±0.3	87.5±0.2

DNNの各ブロックの出力特徴次元数は図1(b)に則る. SA機構内(式(7)–(9))のFCはすべて, そのSA機構への入力特徴次元数の1/4を出力特徴次元数とする. また, 本実験では, 既存手法と同様に, 入力形状カテゴリ(chair, airplaneなど)のone-hotベクトルを1層のFC(出力特徴次元数: 64)とBN, ReLUによって精製後, CBの直前の各点の特徴に接続する.

4.2.5 実験結果

ShapeNetPartでセグメンテーション精度(C-IoU)を評価した結果を表3に示す. 提案手法はセグメンテーションタスクにおいても既存手法を上回る精度を記録し, 提案手法の有効性を示した. なお, 提案DNNは特徴次元数や集約する近傍の数を低く設定した. それらを大きくとれば精度が向上する可能性がある.

4.3 Ablation study

提案手法を構成する各処理の有効性をablation studyで調査する. 本実験では無背景点群(MN40)と有背景点群(SONN)の識別正解率で評価する. 各実験は回転条件z/SO3のみで行う.

4.3.1 LRFによる回転正規化とT-Netの有効性調査

本実験ではLRFによる回転正規化の有無, とT-Netの有無, の合計4通りの組合せを比較し, LRFによる回転正規化とT-NetによるLRFの幾何調整の有効性を調査する.

実験結果を表4に示す. MN40とSONNの両方で, 近傍点

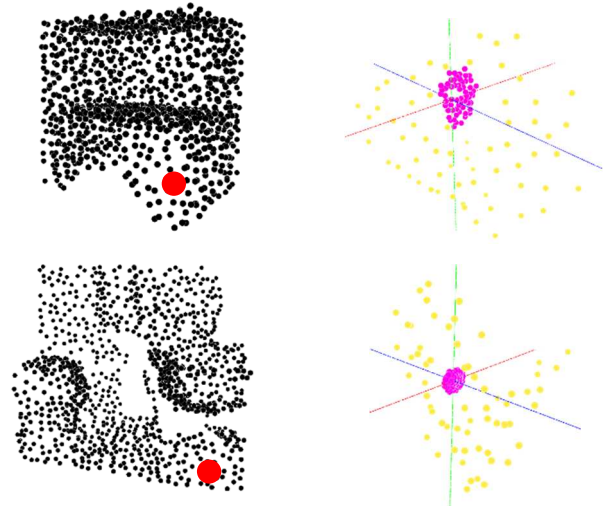


図6 2つの点群(左)について, 注目点(赤)近傍の64点を回転正規化した結果(紫)とT-Netで幾何補正を加えた結果(黄)

Fig. 6 For the two pointsets (left), 64 points around the focal points (red) are rotation normalized (purple) and then geometrically corrected (yellow) using T-Net

群の座標値の共分散行列を固有値分解して得たLRFをさらにT-Netで幾何調整する2段階の幾何変換補正が最高の識別正解率となり, その有効性を確認した. LRFによる回転正規化によりz/SO3条件下で高い識別正解率が得られる. さらに, LRFによる回転正規化にT-Netによる2段階目の幾何補正を行うことで, LRFのみの幾何補正と比べ, 差は小さいものの, 有意な精度改善が見られた. 表に見られるようにT-Netのみではz/SO3条件には対応できない. T-Netのみでは自由度の高いSO3回転を加えられた形状の様々な例を十分には学習できず, 回転の予測と正規化に失敗すると考えられる.

T-Net適用前のLRFによる回転正規化結果と, T-Netを適用後のLRFによる回転正規化を可視化して図6に示す. 図6の左側の点群はDNNへの入力点群であり, 右側が注目点(赤色の点)近傍64点の回転正規化結果である. 右側の紫色の点群がT-Net適用前のLRFによる幾何補正(回転正規化)後, 黄色の点群がT-Netによる追加の幾何補正後の点群である. T-Netは局所点群の微小な回転とスケール拡大を合わせた幾何調整を行っていることがわかる. スケールの拡大はQianらのPointNetXt¹⁰⁾も利用している. Qianらは局所点群のスケールが小さいと学習が難しくなることを指摘し, 局所点群のスケールを手で拡大してから特徴を抽出した. 提案DNNではこれと類似のスケール拡大を, 学習に基づいてT-Netが行う.

4.3.2 局所RI特徴群全体の配置情報の有効性調査

提案DNNのグローバルなTBを $k=16, 32$ のローカルなTBに置換したDNNの識別正解率を計測し, 局所RI特徴群全体の3D空間でのグローバルな配置情報の有効性を調査する.

実験結果を表5に示す. 最終層の集約対象の点数を増やしていくことで形状識別正解率の向上が確認され, グローバル

に TB を用いた場合が最も高い識別正解率となった。この結果から既存の LRF 方式の DNN が考慮しなかった局所 RI 特徴群全体の 3D 空間における配置情報の利用が形状識別精度向上に有効であることを確認した。

4.3.3 SA 機構の有効性調査

提案 DNN において、TB 内の SA 機構をグラフ畳み込みに置き換えた DNN の識別正解率を計測、SA 機構による特徴抽出と比較する。本実験のグラフ畳み込みは提案手法の SA 機構 (式 (6)) から関連度 $\text{softmax}(\alpha(\mathbf{x}_i, \mathbf{x}_j))$ を除去し、特徴群の集約処理を総和からチャンネルごとの max pooling に置き換えることで実装した。

SA 機構とグラフ畳み込みの識別正解率を表 6 に示す。いずれのデータセットでも SA 機構を用いる場合の方が高い精度が得られた。このことから SA 機構による入力データ適応的に変化する重みを用いた特徴集約が識別精度向上につながっていると考えられる。なお、transformer 系の DNN は、大規模データでの事前学習が有効であるとの報告がある⁸⁾。したがって、大規模データセットで提案 DNN を事前学習することで、精度が向上する可能性がある。

4.3.4 メッシュ法線ベクトルを用いた LRF 構築

メッシュ 3D モデル (例えば ModelNet40 や ShapeNetPart) から作成した点群の場合、メッシュの面が持つ法線ベクトルを利用できることがある。そこで、本実験では提案 LRF の z 軸 ($\mathbf{lra}_z$) を固有ベクトルではなくメッシュを構成する面の法線ベクトルで決定し、精度への影響を調査する。本実験では ModelNet40 および ShapeNetPart から法線を持つ有向点群モデルを作成して実験を行う。SONN はスキャンデータであり法線データはないためここでは用いない。ShapeNetPart の評価指標は 4.2 節と同様の 5 回の試行の結果を平均した C-IoU である。

表 7 に固有ベクトルを $\mathbf{lra}_z$ とする場合とメッシュ法線を $\mathbf{lra}_z$ とする場合の MN40 における識別正解率と ShapeNetPart における C-IoU を示す。いずれのデータセットでも、局所点群から得た固有ベクトルではなく、メッシュ法線ベクトルを用いた方、差は小さいものの高い精度が得られた。安定した LRF の構築が精度向上につながることが考えられる。

5. まとめ

本研究では回転不変 (RI) 性をもった高精度な 3D 点群解析 DNN の構築を目的とし、局所点群を LRF によって回転正規化することで RI 性を獲得する DNN の提案を行った。提案手法では、学習による LRF の回転と学習に基づくスケールの幾何調整、自己注意 (SA) 機構を用いた局所 RI 特徴とそれら特徴群全体の 3D 空間における配置情報の抽出と統合、を組み合わせる。ランダム回転した 3D 点群の識別タスクとセグメンテーションタスクで評価した結果、提案手法は既存手法を

表 5 TB 最終層の k 近傍数

Table 5 Number of k nearest neighbors in the final TB layer		
TB 最終層の k 近傍数	MN40	SONN
$k = 16$	90.8 ± 0.1	86.6 ± 0.3
$k = 32$	91.1 ± 0.1	87.1 ± 0.2
$k = 64$ (global)	91.3 ± 0.3	87.5 ± 0.2

表 6 SA 機構とグラフ畳み込みの比較

Table 6 Self-attention vs GraphConv		
特徴抽出機構	MN40	SONN
グラフ畳み込み	91.2 ± 0.2	87.3 ± 0.1
SA 機構	91.3 ± 0.2	87.5 ± 0.2

表 7 $\mathbf{lra}_z$ 決定法が精度に与える影響

Table 7 Impact of $\mathbf{lra}_z$ decision method		
$\mathbf{lra}_z$ 決定法	MN40	ShapeNetPart
固有ベクトル	91.3 ± 0.3	81.9 ± 0.2
メッシュ法線	91.5 ± 0.2	82.3 ± 0.1

上回る精度を記録した。さらに ablation study により、学習による LRF の幾何調整、局所 RI 特徴群全体の 3D 空間における配置情報の利用、SA 機構による入力データ適応的な局所特徴の精製と統合の双方が精度向上に有効であることを確認した。

今後の課題として、3D センサから獲得される 3D 点群に見られる各種のノイズ、例えば背景点群、見通し外部の欠損、サンプル密度のばらつきなどに頑強な手法の構築や、回転正規化した局所点群に適したデータ拡張手法の開発などがあげられる。

参考文献

- 1) C. R. Qi, H. Su, M. Kaichun, L. J. Guibas: "PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation", Proc. of IEEE Conference on Computer Vision and Pattern Recognition, pp. 77–85 (2017).
- 2) C. R. Qi, L. Yi, H. Su, L. J. Guibas: "PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space", Proc. of Neural Information Processing Systems, pp. 5105–5114 (2017).
- 3) Y. Li, R. Bu, M. Sun, W. Wu, X. Di, B. Chen: "PointCNN: Convolution On X-Transformed Points", Proc. of Neural Information Processing Systems, pp. 828–838 (2018).
- 4) Y. Xu, T. Fan, M. Xu, L. Zeng, Y. Qiao: "SpiderCNN: Deep Learning on Point Sets with Parameterized Convolutional Filters", Proc. of European Conference on Computer Vision, pp. 90–105 (2018).
- 5) Y. Wang, Y. Sun, Z. Liu, S. E. Sarma, M. M. Bronstein, J. M. Solomon: "Dynamic Graph CNN for Learning on Point Clouds", ACM Trans. on Graphics, Vol. 38, No.5, pp. 1–12 (2019).
- 6) M.-H. Guo, J.-X. Cai, Z.-N. Liu, T.-J. Mu, R. R. Martin, S.-M. Hu: "PCT: Point Cloud Transformer", Computational Visual Media volume, Vol. 7, No.2, pp.187–1999 (2021).

- 7) H. Zhao, L. Jiang, J. Jia, P. Torr, V. Koltun: "Point Transformer", Proc. of IEEE/CVF International Conference on Computer Vision, pp. 16259–16268 (2021).
- 8) X. Yu, L. Tang, Y. Rao, T. Huang, J. Zhou, J. Lu: "Point-BERT: Pre-training 3D Point Cloud Transformers with Masked Point Modeling", Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 19313–19322 (2022).
- 9) X. Ma, C. Qin, H. You, H. Ran, Y. Fu: "Rethinking Network Design and Local Geometry in Point Cloud: A Simple Residual MLP Framework", Proc. of International Conference on Learning Representations (2022).
- 10) G. Qian, Y. Li, H. Peng, J. Mai, H. A. A. K. Hammoud, M. Elhoseiny, B. Ghanem: "PointNeXt: Revisiting PointNet++ with Improved Training and Scaling Strategies", Proc. of Neural Information Processing Systems (2022).
- 11) Z. Zhang, B.-S. Hua, D. W. Rosen, S.-K. Yeung: "Rotation Invariant Convolutions for 3D Point Clouds Deep Learning", Proc. of International Conference on 3D Vision, pp. 204–213 (2019).
- 12) Z. Zhang, B.-S. Hua, S.-K. Yeung: "RICov++: Effective Rotation Invariant Convolutions for 3D Point Clouds Deep Learning", International Journal of Computer Vision, Vol. 130, No. 5, pp.1228–1243 (2022).
- 13) Z. Xiao, H. Lin, R. Li, H. Chao, S. Ding: "Endowing Deep 3d M Vol. 130, No. 5, pp.1228–1243 (2022).odels With Rotation Invariance Based On Principal Component Analysis", Proc. of IEEE International Conference on Multimedia and Expo, pp. 1–6 (2020).
- 14) R. Yu, X. Wei, F. Tombari, J. Sun: "Deep Positional and Relational Feature Learning for Rotation-Invariant Point Cloud Analysis", Proc. of European Conference on Computer Vision, pp. 217–233 (2020).
- 15) F. Li, K. Fujiwara, F. Okura, Y. Matsushita: "A Closer Look at Rotation-invariant Deep Point Cloud Analysis", Proc. of International Conference on Computer Vision, pp. 16218–16227 (2021).
- 16) T. Furuya, R. Ohbuchi: "Deep Aggregation of Local 3D Geometric Features for 3D Model Retrieval", Proc. of British Machine Vision Conference, p. 121, (2016).
- 17) T. Furuya, X. Hang, R. Ohbuchi, J. Yao: "Convolution on Rotation-Invariant and Multi-Scale Feature Graph for 3D Point Set Segmentation", IEEE Access, Vol. 8, pp. 14025–140260 (2020).
- 18) S. Kim, J. Park, B. Han: "Rotation-Invariant Local-to-Global Representation Learning for 3D Point Cloud", Proc. of Neural Information Processing Systems (2020).
- 19) J. Zhang, M.-Y. Yu, R. Vasudevan, M. Johnson-Roberson: "Learning Rotation-Invariant Representations of Point Clouds Using Aligned Edge Convolutional Neural Networks", Proc. of International Conference on 3D Vision, pp. 200–209 (2020).
- 20) A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser. I. Polosukhin: "Attention Is All You Need", Proc. of Neural Information Processing Systems, pp. 6000–6010 (2017).
- 21) K. He, X. Zhang, S. Ren, J. Sun: "Deep Residual Learning for Image Recognition", Proc. of IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016).
- 22) T. Furuya, R. Ohbuchi: "Diffusion-on-Manifold Aggregation of Local Features for Shapebased 3D Model Retrieval", Proc. of ACM International Conference on Multimedia Retrieval 2015, pp. 171–178 (2015).
- 23) H. Zhao, J. Jia, V. Koltun: "Exploring Self-Attention for Image Recognition", Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition Conference, pp. 10073–10082 (2020).
- 24) S. Ioffe, C. Szegedy: "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift", Proc. of International Conference on Machine Learning, pp. 448–456 (2015).
- 25) Z. Wu, S. Song, A. Khosla, F. Yu, L. Zhang, X. Tang, J. Xiao: "3D ShapeNets: A Deep Representation for Volumetric Shapes", Proc. of IEEE Conference on Computer Vision and Pattern Recognition, pp. 1912–1920 (2015).
- 26) M. A. Uy, Q.-H. Pham, B.-S. Hua, D. T. Nguyen, S.-K. Yeung: "Revisiting Point Cloud Classification: A New Benchmark Dataset and Classification Model on Real-World Data", Proc. of IEEE/CVF International Conference on Computer Vision, pp. 1588–1597 (2019).
- 27) L. Yi, V. G. Kim, D. Ceylan, I.-C. Shen, M. Yan, H. Su, C. Lu, Q. Huang, A. Sheffer, L. Guibas: "A Scalable Active Framework for Region Annotation in 3D Shape Collections", ACM Trans. on Graphics, Vol. 35, No. 6, pp. 1–12 (2016).
- 28) D. P. Kingma, J. Ba: "Adam: A Method for Stochastic Optimization", Proc. of International Conference on Learning Representations, pp. 1–15 (2015).

(2023 年 03 月 10 日 受付)

刈込 喜大

2021 年 山梨大学工学部コンピュータ理
工学科卒業。同年 山梨大学大学院 医工農学
総合教育部修士課程コンピュータ理工学コ
ース入学。2023 年 同 修了。在学中は深層
学習を用いた 3 次元形状解析に関する研究
に従事。



古屋 貴彦

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 山梨大学大学院医
学工学総合教育部博士課程修了。工学博士。
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 上智大学理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。

