

# 点群識別における、形状の複雑さを考慮した サンプリングに関する検討

倉田夏菜<sup>1</sup> 八尾泰洋<sup>1</sup> 安藤慎吾<sup>1</sup> 島村潤<sup>1</sup>

**概要:** 点群の識別において段階的に高次元特徴量を算出する際に、代表点を選択することによる特徴集約が行われる。従来手法ではランダムサンプリングや点群の密度に応じたサンプリングが代表点選択に用いられるが、この方法は、部位によって形状が急激に変化する物体や異なるクラスで共通の形状を持つ物体の識別を困難にする可能性がある。そのため、本稿では、物体の局所的な形状を考慮した代表点選択手法を導入し、その効果を検証した。

**キーワード:** 三次元点群, 特徴集約, キーポイント抽出, クラス識別, 深層学習

## Point Cloud Classification by Using Sampling method for Complex Shapes

KANA KURATA<sup>†1</sup> YASUHIRO YAO<sup>†1</sup>  
SHINGO ANDO<sup>†1</sup> JUN SHIMAMURA<sup>†1</sup>

### 1. はじめに

本研究では、密なグリッド構造を持たないデータの代表である3次元点群に対し、畳み込み・特徴集約を備えたディープニューラルネットワーク(DNN)による識別を行う方法に関して検討を実施した。

近年の画像認識では、深層学習に基づく特徴抽出を行うDNNにより、高精度な識別が可能となった。特に、離散畳み込みによる特徴抽出と、一定区画ごとのプーリングによる特徴集約を繰り返す構造を持つ、Convolutional Neural Network(CNN)のアーキテクチャ(例えば, [1, 2])は、プーリングによって処理対象画素数を減少させつつ段階的に特徴抽出を行うことで、高次元特徴量を効率的に取得できるという利点があり、広く用いられている。このアーキテクチャは画像など、密なグリッド構造を持つデータに対して汎用的に用いることができる。しかし、3D点群(図1)に代表されるように、密なグリッド構造を持たないデータは数多くあり、それらに上記の方法を直接適用することはできない。

例えば、3D点群に対して畳み込みによる段階的な特徴抽出を行う手段として、①空間をボクセル化して離散畳み込みを行う[3, 4], ②グラフニューラルネットワークを用いてEdge convolutionを行う[5]③複数視点の2次元平面へ投影して処理する[6, 7], といったアプローチがある。しかし、①には多くのグリッドが空になってしまうために計算リソースの無駄が大きい事、②にはノード間の関係をモデル化

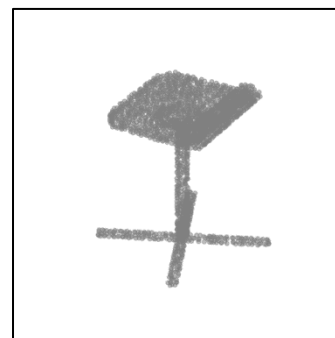


図1 ModelNet40[3]の椅子を表す点群データ

するため識別に大きく寄与する特徴を持つノードがインプットとして与えられる必要がある事や、深いネットワークを学習することが困難である事、③には識別結果が投影先の面の位置に依存する事、といった問題がある。そのような背景から、WangらはDeep Parametric Continuous Convolutional Neural Networks [8]を提案した。この手法では、畳み込みカーネルに離散フィルタの代わりに、連続的なカーネル関数を用いた畳み込み計算(以下、Continuous convolution)を行う。このため、密なグリッド構造を持たなくても、ノード間に、例えば3次元点群における距離のようなベクトルが定義可能なデータには直接適用することができる。

しかしWangらの提案したアーキテクチャは段階的な特徴集約を考慮しておらず、近隣の点で重複した特徴を保持するため、メモリサイズや特徴集約回数といった計算リソ

<sup>1</sup> NTTメディアインテリジェンス研究所  
NTT Media Intelligence Laboratories

一スの観点で効率的ではない。画像などの密なグリッド構造を持つデータを対象とした分野での研究ではプーリングによる特徴集約が、精度向上と計算リソースの削減を同時に実現している。そのような背景から、我々は密なグリッド構造を持たないデータの代表である3次元点群を対象として、Continuous convolution に続く特徴集約手法の検討を実施した。

通常、3次元点群を対象とした特徴集約では、ランダムサンプリングや点の密度に基づくサンプリングが用いられる[9][10]。すなわち、入力点群からサンプリングにより代表点を選出し、その点の周囲で特徴を集約する。このようにサンプリングを用いた特徴集約を行うアーキテクチャでは、以下の2つのケースにおいて識別精度の低下をもたらす可能性がある。1つは、代表点が様々な物体が共通に持つ形状（例えば平面やゆるやかな曲面）の部位に偏った場合で、得られる特徴量は物体によらず類似するため識別が困難となる。2つ目は、物体の中で局所形状特徴の急激な変化が起こる、形状が複雑な物体を対象とする場合である。そのような物体では、代表点の組み合わせによって得られる局所特徴量が大幅に変化する。このため、学習において用いられなかった組み合わせの代表点群が入力されると、識別が困難になると考えられる。ランダムサンプリングや点の密度に基づくサンプリングによる代表点選出では、各点がどのような形状の部位に分布するかを考慮しない。すなわち、3次元点群の識別に寄与する特徴的な局所形状を持つ点が選ばれることが確約されていない。そのため、精度向上のためには多くの点を代表点とする必要があり、特徴集約による計算量削減効果が限定的である。

そこで我々は、3次元点群の識別に寄与する特徴的な局所形状を持つ点を選択するために、キーポイント抽出を用いて代表点選択を行う手法を検討した。

本報告では、Continuous convolution [8] による畳み込みと、Intrinsic shape signatures[11]によるキーポイント抽出を用いた代表点選出による特徴集約を備えた DNN アーキテクチャを提案する。Modelnet40[3]の識別タスクによる評価実験により、提案した深層学習手法はランダムサンプリングのみによる代表点選出をするアーキテクチャと比較して、代表点数が同じ場合に優位に識別精度を向上させることが確認できた。また、Wang らの手法[8]などのノード削減を行わない手法と比較して、本手法は必要とするメモリサイズが小さく、従来の画像認識で用いられるようなノード数減少とともに特徴次元数を増やすような DNN の構成が可能である。

要約とすると、本報告のコントリビューションは以下のとおりである。

- Continuous convolution による畳み込みとキーポイント抽出による特徴集約を実施することで、密なグリッド構造を持たない点群データに対して畳み込みと特徴集約を同時に

実現する手法を提案する。

- 上記提案手法の効果について、ModelNet40 の識別タスクにより評価を行い、ランダムサンプリングのみによる特徴集約を実施する比較手法に対して、11.6%の精度向上を確認した。

## 2. 手法

### 2.1 Continuous convolution

本稿では、Wang ら[8]により提案された Continuous convolution を特徴抽出に用いる。

密なグリッド構造を持つデータに対する離散畳み込みは、次式のように行われる。

(式 1)

$$h[s] = \sum_{m=-M}^M f[s]g[t]$$

$N$  を入力ノード数、 $M$  をカーネルサイズとして、 $f$  は  $S = \{s \in \mathbb{N} \mid 1 \leq s \leq N\}$  に対して特徴量  $V \in \mathbb{R}$  を返す関数、 $g$  は  $T = \{t \in \mathbb{N} \mid 1 \leq t \leq M\}$  に対して畳み込みのフィルタ値  $W \in \mathbb{R}$  を返す関数である。

一方、Continuous convolution での連続関数カーネルによる畳み込みは、式 2 のように定義される。

(式 2)

$$h(x) = \int_{-\infty}^{\infty} f(y)h(x - y)dy \approx \sum_1^N \frac{1}{N} f(y_i)g(x - y_i)$$

$N$  を入力ノード数、 $\dim$  を入力ノードが持つベクトルの次元数として、 $f$  は  $Y = \{ (y_1, y_2 \dots y_{\dim}) \in \mathbb{R}^{\dim} \mid -\infty \leq s_1, s_2 \dots s_d \leq \infty \}$  に対して特徴量  $V \in \mathbb{R}$  を返す関数、 $g$  は  $X = \{ (x_1, x_2 \dots x_d) \in \mathbb{R}^{\dim} \mid -\infty \leq x_1, x_2 \dots x_d \leq \infty \}$  に対して畳み込みの重み  $W \in \mathbb{R}$  を返す関数である。これらの関数が離散値ではなく連続値に対して定義されていることが、Continuous convolution を、ノード間の位置関係が実数値のベクトルによって表現されるデータに対して適用可能としている要因である。

Wang ら[8]により、畳み込みの重みを算出する関数である  $g$  として表現力の高い multi-layer perceptron (MLP) を用い、深層学習により識別に有効な特徴量を算出可能な関数を得る方法が提案されている。その際、離散畳み込みによる特徴抽出が各ノードを中心としたウインドウ内で行われるのと同様に、各点に対する処理は全入力点ではなく各  $K$  個の近傍点を用い、局所的に Continuous convolution が行われる。

本稿における  $g$  も同様に MLP を用いて構成するが、Wang らの手法では各層における近傍点数  $K$  が固定であったのに対し、我々の手法では特徴集約によるノード数減少に合わせて  $K$  も減少させる。

## 2.2 キーポイント抽出による特徴集約

本稿では、3次元点群の識別に寄与する特徴的な局所形状を持つ点を選択するため、キーポイント抽出を用いた代表点選択方法を提案する。

Liらによる PointCNN[9]では、各層で代表点選択と代表点周囲での畳み込み計算を行い、その繰り返しにより高次元特徴量を獲得する。各層の代表点数  $Q$  は、ハイパーパラメータとしてあらかじめ設定される。各層には前層の代表点群とその特徴量が入力され、その中から更に  $Q$  個の点から成る代表点群をサンプリングによって選択し、各代表点について近傍  $K$  点の特徴量に基づく畳み込み計算が行われる。すなわち、ある層の代表点群はその前の層の代表点群、あるいはその部分集合である。我々の手法では、このアーキテクチャにおいて、特徴量算出に Continuous convolution を、代表点選択方法に以下に述べる方法を用いる。

まず、入力点群から ISS [11] のキーポイント抽出アルゴリズムによってキーポイント群を取得する。ISS のキーポイント抽出では、曲率の高い部位に位置する点がキーポイントとして抽出される。抽出されたキーポイント点数を  $Q_{key}$  とすると、 $Q_{key}$  は物体ごとに不定であるため、各層の

代表点数  $Q$  と大小を比較して、以下のように代表点群を決定する。

$Q_{key} > Q$  の場合：キーポイント群の中からランダムに  $Q$  点を選択し、代表点群とする。

$Q_{key} = Q$  の場合：キーポイント群を代表点群とする。

$Q_{key} < Q$  の場合：キーポイント群に、キーポイント以外の点からランダムに選択した  $Q - Q_{key}$  点を加えて代表点とする。

2層目以降では、この条件に加え、代表点が前層の代表点（前層と  $Q$  が等しい場合）または部分集合（ $Q$  が減少する場合）となるように代表点を選択する。

## 2.3 ネットワーク構造

本稿の DNN は、代表点選択と代表点に対する近傍点分布の特徴ベクトルへの変換を行う入力特徴抽出層 (IC 層, 図 2a), 代表点選択と代表点周囲での Continuous convolution による畳み込み計算により特徴抽出を行う Concentrative convolution 層 (CC 層, 図 2b), さらに CC 層から出力された特徴量に基づき識別を行う識別層 (図 2c) の、3つのサブモジュールから構成される。図中の二重線角丸枠は処理、一重線角丸枠は入出力される配列を示し、か

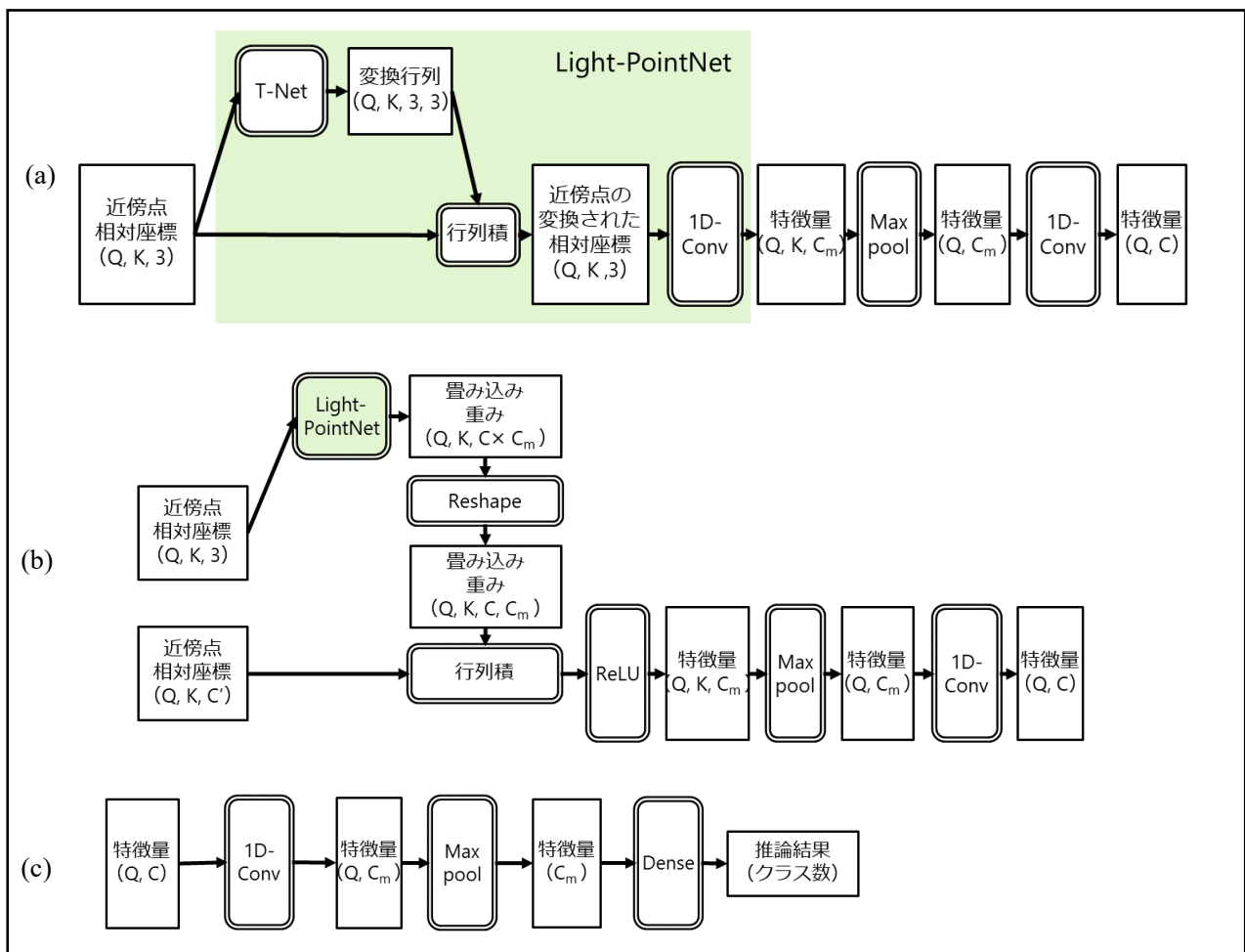


図 2 (a) IC 層の構成、(b) CC 層の構成、(c) 出力層の構成

っこ内は配列の形状を表す.  $Q$ : 代表点数,  $K$ : 畳み込みに用いる近傍点数,  $C'$ : 入力特徴次元数,  $C_m$ : 中間層特徴次元数,  $C$ : 出力特徴次元数である.

入力 IC 層の構成は, 図 2a の通りである. IC 層では, 代表点を選択し, その周囲  $K$  近傍点の相対座標を入力として, 代表点の特徴量を出力する. Light-PointNet は, PointNet [12] の構造を部分的に用いたもので, 点の座標に基づく特徴量を算出する. PointNet は, T-Net による Input transform に続き, 全点に対して共通のカーネルを用いた畳み込み (以下, 1D-Conv), さらに T-Net を用いた Feature transform, 1D-Conv, 全入力点にわたる Max pooling を行い, その値に基づく推論結果を出力する. Light-PointNet はその中の Input transform と 1D-Conv のみから構成される. 1D-Conv は, 入力された各点の座標・特徴量に対し, 全入力点にわたって共通の重みを用いた畳み込みを行う. T-Net には, 座標変換のための変換行列を, 入力された点群の座標・特徴量の分布に基づいて生成する効果がある.  $Q_i$  ら [12] は T-Net の出力に対する Loss を損失関数に加えることで, この座標変換を回転に制限することを試みたが, 本稿では T-Net に関する制約項を加えない. そのため, 学習結果によっては, 物体の回転だけでなく拡大・縮小といったデータにおける点密度の不均一性への対応のために有効な変換が行われる可能性がある. 図中の Max pool では, 各代表点に対し,  $K$  近傍点の中での Max pooling が行われる.

CC 層では代表点を選択し, その周囲各  $K$  近傍点の, 代表点に対する相対座標と特徴量に基づく Continuous convolution を行う. 本稿でのネットワーク構成は図 2b に示すものである. CC 層は, Continuous convolution に用いる, 各近傍点の代表点に対する相対座標に基づく重みを算出する際に, IC 層にも含まれる Light-PointNet を用いる. 図中における  $C'$  は, 各 CC 層に入力される特徴量の次元数である. CC 層の Max pool も IC 層と同様に, 各代表点に対し,  $K$  近傍点の中での Max pooling が行われる.

出力層では, 獲得された特徴量から, 図 2c のネットワークを用いて推論結果を算出する. ここでの Max pool はこの時点で残っている全ての代表点の間で行われる. 推論結果と one-hot ベクトルの正解ラベル間の交差エントロピーを各データに対する損失関数として設定し (式 3), Adam を用いた最適化により End-to-End で各層のパラメータを最

適化する.  
(式 3)

$$Loss = - \sum_{c=1}^C u_c^{true} \log(u_c^{pred})$$

$C$  は識別クラス数,  $u_c^{true}$ ,  $u_c^{pred}$  はそれぞれ, クラス  $c$  における正解ラベルと推論結果である.

本研究では, これらの 3 モジュールを組み合わせた DNN により, 点群の識別を行う. 実験に用いた DNN 構成とパラメータについては, 3.3 章にて詳細に述べる.

### 3. 実験

#### 3.1 概要

本稿では, キーポイント抽出を用いた代表点選択手法の有効性を検証するため, 2.3 章に述べた DNN において, 代表点選択手法を, (A) キーポイント抽出を用いた代表点選択, (B) ランダムサンプリングによる代表点選択, と 2 パターン設定し, ModelNet40 の識別タスクによる評価を行った.

#### 3.2 データセット

評価実験に用いた ModelNet40 [3] は, 40 クラスの人工物の 3D CAD モデルである. 1 データは 1 つのオブジェクトであり, 各オブジェクトの形状は 2048 点の点群によって表現される. 点群の座標値は -1.0 から 1.0 の範囲に正規化されている. 学習データ 9,843 オブジェクト, テストデータ 2,468 オブジェクトから構成される. 今回は代表点選択手法の評価にあたり, サンプリング以外のランダム性を除くため, 座標値にランダムな摂動を加えるなどのデータオーギュメンテーションは行わない.

#### 3.3 DNN 構成

図 3 本稿における DNN 構成に示す通り, IC 層に続き, 3 層の CC 層, 識別層によって各オブジェクトの識別を行う DNN を用いた. 各層のパラメータは,  $Q$ : 代表点数,  $K$ : 畳み込みに用いる近傍点数,  $C$ : 出力特徴次元数,  $D$ : 近傍点選択時の Dilation rate である. Dilation rate は近傍点選択における点の間引き方を決定するパラメータで,  $D=1$  が距

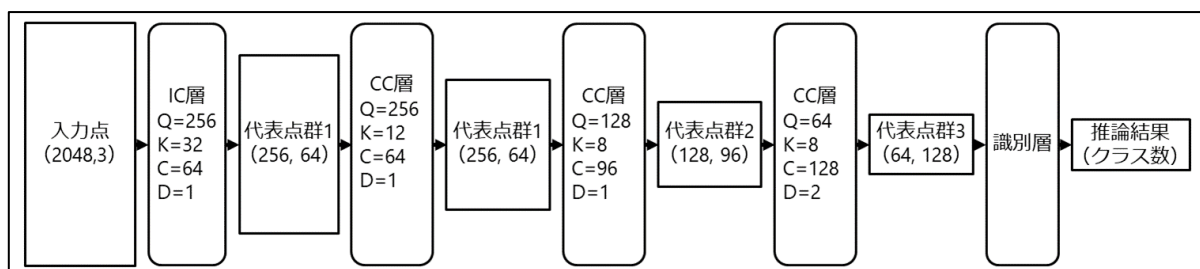


図 3 本稿における DNN 構成

離の順に、 $D=2$ では距離の順に従って1つおきに近傍点を選択する。Dilation rateを高く設定すると、近傍点数 $K$ が同じでも、より広い範囲の情報を得られるという利点がある。また、ModelNet40による評価実験を行うため、推論結果のクラス数は40である。

### 3.4 実験条件

実験設定(A)において、代表点選択におけるキーポイントは全エポックに渡って共通で、キーポイント以外の点に由来する代表点は、1エポックごとに更新しながら学習を行った。実験設定(B)では、全ての代表点を1エポックごとに更新した。また、各エポックにおけるテストデータに対する識別精度の評価において、テストデータの代表点は、エポックによらず固定とする。

学習時・テスト時のバッチサイズは32、学習エポック数は300である。

ISSによるキーポイント抽出には画像・点群処理ライブラリであるPCL (<http://www.pointclouds.org/about/>)を用い、

表1のようにパラメータを設定した。

表1 ISSの処理におけるパラメータ

パラメータ名	値
salient_radius	0.1
non_max_radius	0.1
normal_radius	0.1
border_radius	0.05
gamma_21	0.975
gamma_32	0.975
min_neighbors:	5

### 3.5 結果

以下2つの実験設定に対して結果を比較する。

- (A) キーポイント抽出を用いた代表点選択
- (B) ランダムサンプリングによる代表点選択

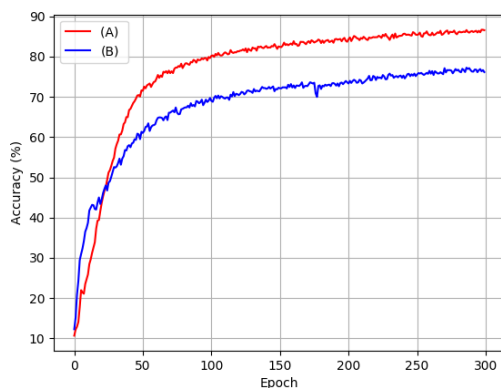


図4 学習データに対する識別精度

学習データに対する識別精度は図4のように、学習が進むと(A)の精度が(B)の精度に対して有意に高くなることがわかった。

図5 テストデータに対する識別精度は、テストデータに対する識別精度について、各エポックの精度を淡色の線で、10エポックごとの移動平均を濃色の線で示したものである。(A)、(B)のどちらでも、テストデータに対する識別精度は、学習データに対する識別精度から10%程度低くなっている。テストデータに対する識別精度の最高値は(A)で82.6%、(B)で71.0%であり、キーポイントを用いることによって精度が11.6%向上した。

また、テストデータの識別精度の10エポックごとの標準偏差(図6)では、キーポイントを用いた(A)ではランダムサンプリングを用いた(B)と比較して標準偏差の上昇がなく、テストデータに対する識別精度が安定していることがわかる。

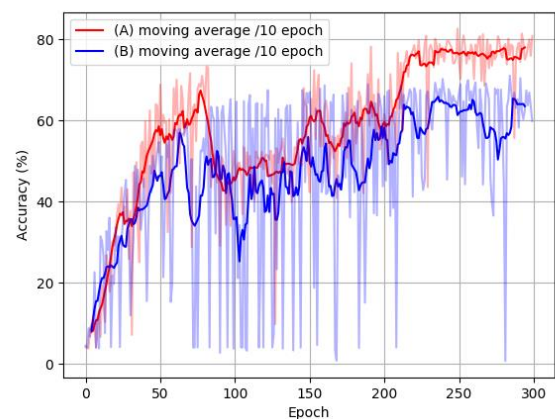


図5 テストデータに対する識別精度

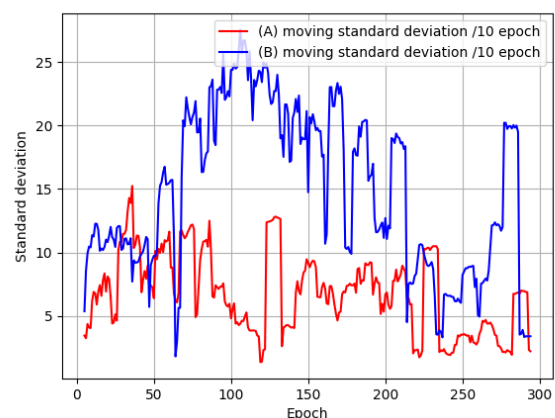


図6 テストデータに対する識別精度の10エポックごとの標準偏差

図7は、(A)において、テストデータに対する識別精

度が最も高かったエポック 249 の推論結果について、キーポイント数と推論結果の正解率の関係を示すグラフである。ヒストグラムのビンは、キーポイント数 10 ごとに区切っている。キーポイント数が 100-150 程度で正解率が高く、キーポイント数が 200 を超える場合に正解率が低くなることが分かった。

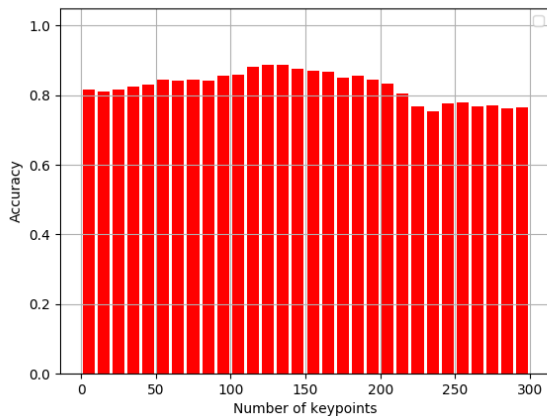


図 7 テストデータの推論結果 (エポック 249) における、キーポイント数に対する正解率

#### 4. 考察

本稿の DNN (代表点数: 256→256→128→64) は計算リソースを効率化するため、既存手法[9] (代表点数: 1024→384→128→128) と比較して、代表点数を大幅に削減している。学習データにおける識別精度とテストデータにおける識別精度の比較から、このように代表点数が少ない場合でも、特徴集約にキーポイント抽出を用いた代表点選択方法を採用することで、ランダムサンプリングによる代表点選択方法を用いた場合と比較して識別精度が有意に向上することがわかった。

しかし、図 7 テストデータの推論結果 (エポック 249) における、キーポイント数に対する正解率の結果では、(A) において、キーポイント数が高い場合に識別精度が低下しており、キーポイント数と正解率に正の相関はない。この原因としては、2, 3 回目の CC 層における代表点数に対してキーポイント数が多く、キーポイントの中からランダムに代表点を選出されているため、その際のランダム性が影響を及ぼしていることが考えられる。この問題を解決するためには、キーポイントの中から代表点を絞り込む際に、ランダムサンプリングではなく、特徴的な形状を持つ点を優先的に残すような方法をとる必要があると考えられる。

また、学習データの識別精度とテストデータの識別精度の乖離は (A), (B) 共に同程度であり、今回の実験に当たり、データオーギュメンテーションを行わなかったことが影響している可能性が高い。そのため、データオーギュメ

ンテーションを行った場合についても識別精度を確認する必要がある。

#### 5. おわりに

本研究では Continuous convolution による畳み込みとキーポイント抽出による特徴集約を実施することで、密なグリッド構造を持たない点群データに対して畳み込みと特徴集約を行い、段階的に高次元な特徴量を取得する手法を提案した。ModelNet40 の識別タスクにおける評価から、本稿の特徴集約手法を採用することで、ランダムサンプリングによる特徴集約を用いた場合に対して、精度が 11.6% 向上することを確認した。今後は、セグメンテーションタスクへの応用と、さらなる精度向上のために代表点選択手法の改善に取り組む予定である。

#### 参考文献

- [1] Krizhevsky, A. et al.. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 2012, p. 1097-1105.
- [2] Simonyan, K. and Zisserman, A.. Very deep convolutional networks for large-scale image recognition, 2014, arXiv:1409.1556 [cs.CV]
- [3] Wu, Z. et al.. 3D ShapeNets: A deep representation for volumetric shapes", 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, p. 1912-1920.
- [4] Sedaghat, N. et al.. Orientation-boosted Voxel Nets for 3D Object Recognition. Proceedings of the British Machine Vision Conference (BMVC), 2017, p. 97.1-97.13.
- [5] Landrieu, L. and Simonovsky, M.. Large-Scale Point Cloud Semantic Segmentation with Superpoint Graphs. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, p. 4558-4567.
- [6] Su, H. et al.. The IEEE International Conference on Computer Vision (ICCV), 2015, p. 945-953.
- [7] Kanazaki, A. et al.. RotationNet: Joint Object Categorization and Pose Estimation Using Multiviews from Unsupervised Viewpoints. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, p. 5010-5019.
- [8] Wang, S. et al.. Deep Parametric Continuous Convolutional Neural Networks, 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2018, p. 2589-2597.
- [9] Li, Y. et al.. PointCNN: convolution on X-transformed points. Proceedings of the 32nd International Conference on Neural Information Processing Systems (NIPS), 2018, p. 828-838
- [10] Qi, C. R. et al.. PointNet++: Deep hierarchical feature learning on point sets in a metric space. Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS), 2017, p. 5105-5114.
- [11] Zhong, Y.. Intrinsic shape signatures: A shape descriptor for 3D object recognition. 2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops, 2009, p. 689-696.
- [12] Qi, C. R. et al.. PointNet: Deep learning on point sets for 3D classification and segmentation. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, p. 652-660.