

PointNet#: 色情報を付与した3次元点群の分類

PointNet#: Classification of 3D-colored Point Sets

河合継^{1*} 黒田唯矢^{1,2} 眞嶋啓介^{1,3}

¹ クリスタルメソッド株式会社

¹ Crystal method Co. Ltd.

² 早稲田大学人間科学部

² Waseda University School of Human Sciences

³ 慶応義塾大学環境情報学部

³ Keio University Environment and Information Studies

Abstract: 近年 VR や AR,MR など 3D を扱う技術が発展を遂げ、「デジタルツイン」と呼ばれる現実空間の情報をセンサーで取得し仮想空間上に 3D で再現してシミュレーションや実験を行う取り組みが行われている。これに伴い、深層学習を用いた 3D データに対する分類や異常検知といった技術の需要が高まると思われる。しかしながら現状では、PointNet++などの 3D 点群情報を扱うモデルにおいて色情報を学習させる研究はほとんどなされていない。そこで本論文では PointNet++を改良し色付きの 3D 点群データの学習を可能にした PointNet#というモデルを提案する。

1 はじめに

3D 情報を扱うデータ形式はいくつかあり、画像に対して深度情報を与えた RGB-D や立方体を積み上げて物体を表現するボクセル、多角形で形状を近似するポリゴンなどが挙げられるが、今回は 3次元の点データの集合である点群データについて扱う。3D 情報に対する深層学習モデルは 3D のデータ形式によって異なり、点群データに対しては PointNet やその系譜の PointNet++が代表される。これらは、点群 (x,y,z) に対しては後述の方法で学習できるようになっているが、言うに及ばず現実のデータは座標だけでなく色情報 (R,G,B) も持っているため、点群と色を同時に学習できることが望まれる。そこで PointNet++を元に入力を 6次元 (x,y,z,R,G,B) にした時に点群と色情報に対して異なる処理を行うようにし、点群と色を同時に学習できるようにした。PointNet++が PointNet からのインクリメントであることや、C が C++からのインクリメントであることになぞらえて今回のモデルを PointNet#と呼ぶことにする。本論文では PointNet, PointNet++との比較、PointNet#の構造説明および検証結果を展開していく。

2 先行モデル

2.1 PointNet

C.R.Qi et al[1] は、T-net と呼ばれる変換行列を推定する構造をモデルの中に組み込むことで、回転に不変な特徴を捉えることができる Pointnet と呼ばれるアーキテクチャを提案した。また、同モデルでは、モデルにプーリング層を加えることで点の格納順に左右されないグローバルな特徴量を獲得することに成功している一方で、このプーリング処理によって点群の局所的な情報が欠落してしまうことが知られている。

2.2 PointNet++

これに対し、C.R.Qi et al [2] は、点群を Farthest point sampling を用いて局所域に分割し、それぞれの局所域に対して Pointnet のプーリング処理を適応し局所域内の特徴量を統合する処理を段階的に繰り返すことで、局所的な情報を加味しつつ、点群全体の特徴量を抽出できる Pointnet++と呼ばれるアーキテクチャを提案した。局所域に対して Pointnet 処理を行うレイヤーは、Pointnet レイヤーと呼ばれ、Pointnet レイヤーの入力サイズは、点群の保持点数を n 、局所域数を k 、空間軸数を d 、チャンネル数を c とすると $[n, k, (d+c)]$ となる。C.R.Qi et al [2] では、第 1 層の Pointnet レイヤーへの入力チャンネル数が明示されていないが、1

*連絡先：クリスタルメソッド株式会社
E-mail: kawai@crystal-method.com

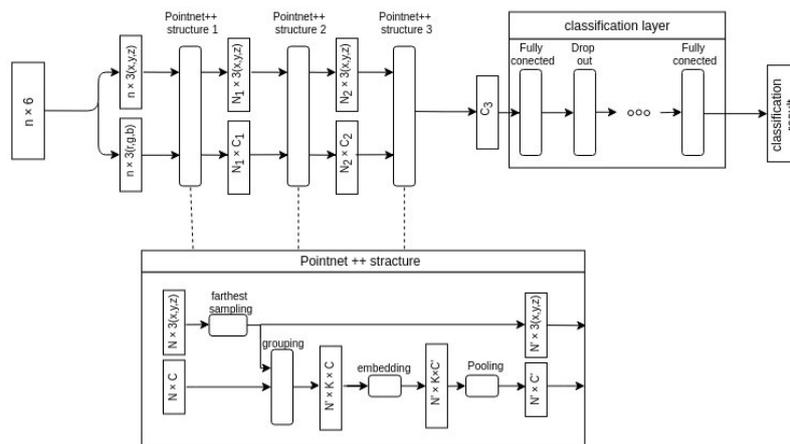


図 1: Pointnet# 構造図

層目 Pointnet 層の入力に RGB 情報を付加することで色情情報を加味した特徴量を抽出できる可能性がある。

3 PointNet#モデル

上記の背景により、本研究では、Pointnet++モデルの一層目の grouping レイヤー・Pointnet レイヤーの入力に RGB 情報を付加することによって、色付き点群を扱う手法である Pointnet#を提案する。

3.1 モデル構造

Pointnet#のモデル構造を図 1 に示す。

Pointnet#は 3 層の Pointnet++ structure レイヤーと 1 層の classification レイヤーから構成される。

3.1.1 Pointnet++レイヤー

Pointnet++ structure レイヤーは、Pointnet++における sampling レイヤーと grouping レイヤーと Pointnet レイヤーを束ねたものである。

1 層目の Pointnet++ structure レイヤーは、点群数を n として $n \times 6$ の行列を入力として受け取り、座標情報 ($n \times 3$) と RGB 情報 ($n \times 3$) に分割する。座標情報について farthest point sampling を行い、代表点を抽出する。代表点情報はサンプリング数を $N1$ として $N1 \times 3$ の行列として保持される。その後、代表点情報と座標情報に基づき、点をグルーピングする。この際、各代表点から一定距離内にある点を各グループとして定義し、グループ内の点は、代表点からの相対座標に変換された座標情報と色情情報を保持する。この時、データ構造はグループメンバー数を $K1$ として $N1 \times K1 \times 6$

となる。その後、このデータについて、embedding を行い特徴量を 6 次元から $C1$ 次元へと変換し、加えて、グループ内の点の特徴を統合するために、各特徴量のプーリング処理を行う。そのため、プーリング後の特徴量は、 $N1 \times C1$ に圧縮され、保持していた代表点情報とともに出力される。

2 層目の Pointnet++ structure レイヤーは、1 層目の Pointnet++ structure レイヤーからの出力である代表点情報 ($N1 \times 3$) と特徴量情報 ($N1 \times C1$) を入力として受け取り、代表点情報を 1 層目における座標情報、特徴量情報を RGB 情報として、1 層目と同じ処理を行う。C.R.Qi et al [2] による Pointnet++では、各 pointnetlayer において座標情報を特徴量に加えていたが、Pointnet#では座標情報と色情情報の重みを均一にするために 2 層目、3 層目の Pointnet++ structure レイヤーでは、座標情報の追加を行わなかった。

3 層目の Pointnet++ structure レイヤーは、2 層目の Pointnet++ structure レイヤーからの出力をを入力として受け取り、2 層目と同じ処理を行う。サンプリング数を 1 として、残っているすべての点を 1 つのグループとしてみなして処理を行う。そのため、3 層目の出力は embedding 後のチャンネル数を $C3$ として、サイズ $C3$ の特徴ベクトルとなる

3.1.2 classification レイヤー

classification レイヤーは、複数の全結合層と dropout 層からなり、3 層目の Pointnet++ structure レイヤーの出力である特徴ベクトルを次元数が分類クラス数になるように変換し出力する。

3.1.3 損失関数

損失関数には、classification レイヤーの出力を softmax 関数にて活性化した結果と正解ラベルとのクロスエントロピー誤差を用いた。

4 検証

Pointnet#は色情報を明示的に扱う Pointnet++の発展形であることから、検証では、色情報を用いて判別ができていないか、色情報がない場合も Pointnet++と同様の精度が得られるか、実際に撮影した 3D データにて判別が行えるかの 3 点について検証を行うため、(1) 色付き ModelNet を用いた検証、(2) 色なしモデルネットを用いた検証、(3) 実際に撮影した独自データセットを用いた検証の 3 種類の検証を行なった。



図 2: 色付き ModelNet データ

4.1 データセット

4.1.1 色付き ModelNet

ModelNet[3]とは研究用に一般公開されている CAD データのデータセットであり、3D データを扱うモデルの評価用データとして広く利用されている。ModelNet には、40 種類の物体のデータが入った ModelNet40 と 10 種類の物体データが入った ModelNet10 があり、本研究では ModelNet10 を用いて検証を行う。ModelNet10 に収められているデータは色なしの CAD データであるため、ここでは CAD データのメッシュ上からランダムに点をサンプリングし作成した点群データを検証に用いる。

ModelNet10 のデータセットのクラスは表 1 の通りである。

表 1: ModelNet10 のクラス

chair	sofa
table	bed
monitor	night_stand
toilet	dresser
desk	bathtub

(1)の検証では、Modelnet10内のベッドデータ (bed)、ソファデータ (sofa) についてメッシュ上から 1000 点のサンプリングを行い正規化処理を行ったあと、それぞれの点について色情報 (RGB) を付加した。色情報は、各データについて RGB のうち R に 1、G と B については 0-1 の範囲でランダムな値を入れたデータ (Red) と、G が 1、R と B には 0-1 の範囲でランダムな値を入れたデータの 2 種類を作成した。作成した検証用データの一部を図 2 に示す。

4.1.2 色無し ModelNet

(2)の検証では、Modelnet10内の 10 種類のデータについてメッシュ上から 1000 点のサンプリング行なった後正規化処理を行い、点群データを作成した。Pointnet#用データでは、各点の色情報を 0,0,0 としパッキングした。作成した検証用データの一部を図 3 に示す。

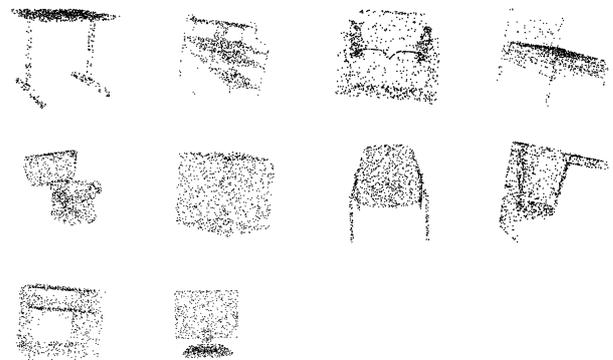


図 3: 色なし ModelNet データ

4.1.3 独自データセット

上記の ModelNet10 に加え、色付き点群データを撮影しそのデータを用いてモデル検証を行う。撮影対象はカップ麺の容器で、みそ/とんこつ/ねぎみそ/家系/カップヌードル/春雨/そうめんの 7 種類のカップ麺の容器を台の上におき、台を回転させて各方向から正常/色異常/形異常を 100 枚ずつ撮影。形異常なカップ麺は実際にカップ麺を叩いて凹ませたものを用い、色異常

なカップ麺は蓋や側面に赤い丸を油性ペンで描いたものを用いた。



図 4: 独自データセット

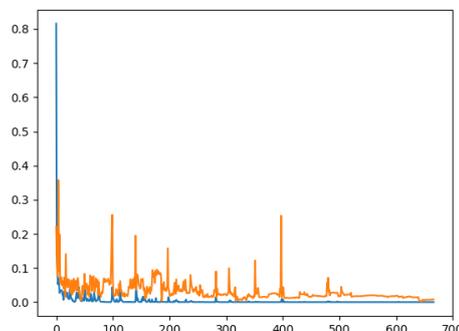


図 5: loss の推移 青: 学習データ オレンジ: 評価データ

4.2 検証結果

4.1 にて示した 3 種類のデータセットを用いてモデル検証を行った結果を以下に示す。

4.2.1 色付き Modelnet の判別結果

4.1.1 で示した ModelNet10 のベッドデータとソファデータより作成した色付きデータを用いて、4 クラス分類を行った。それぞれの種類における学習データ数、評価データ数は、ModelNet10 の区分けをそのまま利用し、学習 125 ファイル、評価 25 ファイルで分けた。学習にはバッチ学習を用い、バッチサイズを 30 とした。学習時の epoch ごとの loss と accuracy の推移を図 5, 図 6 に示す。

学習データにて学習後、評価データを用いて判別の精度を測定した。判別結果を表 2 に示す。評価データ全体の判別精度は 99.75%である。

表 2: 色付き ModelNet の結果

	red_bed	gree_bed	red_sofa	green_sofa
red_bed	100	0	0	0
green_bed	0	100	0	0
red_sofa	1	0	99	0
green_sofa	0	0	0	100 height

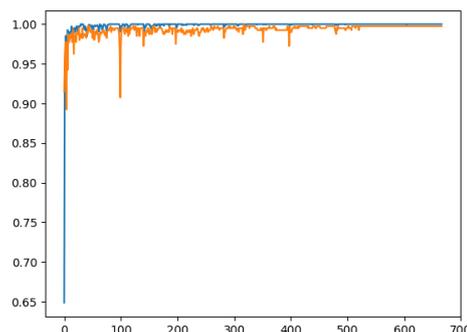


図 6: accuracy の推移 青: 学習データ オレンジ: 評価データ

4.2.2 色無し Modelnet の判別結果

色情報がなく、座標情報のみを特徴量とした場合において、PointNet#において座標情報が損なわれていないのかを検証するために、既存の PointNet++との精

度比較を行った。データは 4.1.2 で示した色なし ModelNet10 の点群データを用いた。データ数は偏りが出ないように、各クラスそれぞれ、学習データ：125 個、評価データ：25 個と揃えた上で 10 クラス分類問題として学習した。判別結果は表 3 に示したとおりであり、10 クラス全体での精度は Pointnet++ では 91.18% であり、Pointnet# では 91.52% であった。

表 3: 色無し ModelNet 判別精度比較

	PointNet++	PointNet#
bathhtub	94.0%	94.0%
bed	99.0%	97.0%
chair	99.0%	99.0%
desk	84.9%	87.2%
dresser	75.6%	77.9%
monitor	98.0%	99.0%
night_stand	67.4%	72.1%
sofa	98.0%	99.0%
table	91.0%	91.0%
toilet	100%	99.0%
TOTAL	91.18%	91.52%

4.2.3 独自データセットの判別結果

表 4: 独自データセット結果

	Accuracy
正常	100%
形異常	91.67%
色異常	71.25%

5 考察

5.1 色付き ModelNet

4.2.1 の結果から、4 クラス全体での判別精度が 99.75% であり、色の違い、形の違いともに高い精度で判別出来ていると言える。また、判別を誤ったデータにおいても、同色の別形状データへの誤りであったため、Pointnet# では、色の情報を判別できていると考察できる。

今回の精度検証にて判別を間違えたものを図 6 に示す。このデータは、red sofa のデータであるが、red bed として判定された。その原因は、背もたれ部分部分と座面部分の幅がほぼ等しく、またその成す角度が垂直



図 7: 誤判定した点群データ

であるため、ベッドの形状と酷似した特徴を持っていたためであると考察できる。

学習データ、検証データともに図 5 で loss が下がり、図 6 で accuracy が上がっていることから学習がうまくいっていることがわかる。また、全体精度が 99.75% と非常に高いことについては、人工的に赤と緑を着色したデータであるため、違いが明確で分類が容易であった可能性が高いことが原因として挙げられる。

5.2 色無し ModelNet

4.2.2 の結果から、10 クラス全体での判別精度が、Pointnet++ では 91.18% であり、Pointnet# では 91.52% であることが読みとれ、PointNet# の座標情報のみにおける分類精度が PointNet++ と同程度であることが確認できた。

また、表 4 に示した各クラスの判別精度について、Pointnet# において精度が 90% を下回ったクラスは、desk、dresser、night_stand の 3 クラスであり、精度がそれぞれ、87.2%、77.9%、72.1% であったが、Pointnet++ においても、精度が 90% を下回ったクラスは同様の 3 クラスであり、精度がそれぞれ、84.9%、75.6%、67.5% であった。このことから、座標情報のみを使用した場合の Pointnet# と Pointnet++ では、判別しにくくクラスの傾向が酷似している。そのため、座標情報のみを特徴量とした場合の Pointnet# は、Pointnet++ と同様の判別能力を有していると考察できる。

5.3 独自データセット

表 4 での結果の通り、正常判定/形異常判定は高精度でできていたが色異常に対しては 7 割程度に留まった。この理由としては色異常としてつけた赤い点が小さかったため正常なものとの区別が難しかった可能性が高い。



図 8: 誤判定した色異常カップ麺

6 まとめと今後の展望

本論文では、近年の「VR 技術」や「MR 技術」、「デジタルツイン」など、現実空間の情報を 3D 情報にて再現する技術の台頭による、3D 情報を用いた分類技術や異常検知技術の需要の高まりから、色情報が付加された 3D 点群情報を学習できる深層学習モデルである「Pointnet#」を提案し、モデルの検証を行なった。

検証では、色情報を用いて判別ができていないか、色情報がない場合も先行モデルである Pointnet++ と同様の精度が得られるか、実際に撮影した 3D データにて判別が行えるかの 3 点について検証を行なった。

検証では、提案手法である PointNet# では座標情報と色情報を同時に学習できることが明らかになった。また、点群情報のみに対しても従来の PointNet++ と変わらない精度で判別できることが確認できた。

一方で、色情報を持つ点群データの一般公開されているデータセットが少ないことや、点群情報の撮影には、2D 画像撮影と比べて撮影コストもかかることから、今回の検証では大量のデータを用いた検証が行えなかった。そのため、今後の課題としては、より多量かつ多様な点群データに対して分類や判別が行えるかを検証する必要があることが挙げられる。

また、PointNet# は Pointnet++ における 1 層目の入力に明示的に色情報を付加することで、色情報を学習するモデルであったが、構造的には色情報以外の特徴量を付加することも可能である。そのため、Pointnet の発展形としては、点群情報に時間情報を付加したデータを用いて、時系列を考慮した分類や異常検知を行いモデルなどが考えられ、今後の展望としたい。

参考文献

C. R. Qi, H. Su, K. Mo, and L. J. Guibas. PointNet: Deep learning on point sets for 3D classification and segmentation. In CVPR, 2017.

C. R. Qi, L. Yi, H. Su, and L. J. Guibas. PointNet++: Deep hierarchical feature learning on point sets in a metric space. In NIPS, 2017.

M. Aubry, U. Schlickewei, and D. Cremers. The wave kernel signature: A quantum mechanical approach to shape analysis, In Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on, pages 1626 - 1633. IEEE, 2011.

D. Belton and D. D. Lichti. Classification and segmentation of terrestrial laser scanner point clouds using local variance information. *Iaprs*, Xxxvi, 5:44 - 49, 2006.

J. Bruna, W. Zaremba, A. Szlam, and Y. LeCun. Spectral networks and locally connected networks on graphs. arXiv preprint arXiv:1312.6203, 2013.

N. J. MITRA, A. NGUYEN, and L. GUIBAS. Estimating surface normals in noisy point cloud data. *International Journal of Computational Geometry Applications*, 14(04n05):261 - 276, 2004.

A. X. Chang, T. Funkhouser, L. Guibas, P. Hanrahan, Q. Huang, Z. Li, S. Savarese, M. Savva, S. Song, H. Su, J. Xiao, L. Yi, and F. Yu. ShapeNet: An Information-Rich 3D Model Repository. Technical Report arXiv:1512.03012 [cs.GR], 2015.

S. Song and J. Xiao. Deep Sliding Shapes for amodal 3D object detection in RGB-D images. In CVPR, 2016.

C. Zimmermann and T. Brox. Learning to estimate 3D hand pose from single RGB images. In ICCV, 2017.

S. Yuan, G. Garcia-Hernando, B. Stenger, G. Moon, J. Y. Chang, K. M. Lee, P. Molchanov, J. Kautz, S. Honari, L. Ge, J. Yuan, X. Chen, G. Wang, F. Yang, K. Akiyama, Y. Wu, Q. Wan, M. Madadi, S. Escalera, S. Li, D. Lee, I. Oikonomidis, A. Argyros, and T.-K. Kim. Depth-based 3D hand pose estimation: From current achievements to future goals. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.