

3次元点群生成のためのマルチタスク学習を用いた表現学習 Learning Representations using Multi-task Learning for 3D Point Cloud Generation

松崎 康平[†] 田坂 和之[†]
Kohei Matsuzaki Kazuyuki Tasaka

1 はじめに

3次元計測機器の普及や3次元データに対する深層学習フレームワークの発展に伴い、3次元点群処理技術が注目を集めている。先行研究の多くは畳み込みニューラルネットワークを適用するために、3次元点群をボクセルデータ [1] や多視点画像 [2] のような空間的に規則化されたデータ形式へ変換する。それに対し、Qiら [3] はデータ形式の変換なしに3次元点群を入力可能な深層学習フレームワークを提案した。PointNetと呼ばれるこの手法はデータ形式の変換における量子化アーティファクトの発生を回避するため、物体分類やセグメンテーションに関してボクセルや多視点画像に基づく手法より高い精度を達成することが報告されている。また、この手法は物体検出 [4] や3次元局所特徴マッチング [5] のような様々な課題に適用することも可能である。

他方、深層学習技術の分野においては、敵対的生成ネットワーク [6] に基づく生成モデルの学習についての研究が近年盛んに行われている。実データを学習用データとして敵対的生成ネットワークを学習させることによって、乱数から実データに近い人工データを生成することができる。この手法はデータ拡張やドメイン適応といった深層学習モデルの性能を改善するためのアプローチへ応用できる可能性があり [7]、深層学習の分野における基礎的研究として期待されている。しかしながら、敵対的生成ネットワークの適用対象の多くは画像 [8] や映像 [9] であり、3次元点群を対象とした研究は少ない。Achlioptasら [10] は、3次元点群を対象とした自己符号化器と敵対的生成ネットワークに基づく生成モデルを提案した。しかしながら、このモデルはPointNetの構成を利用しており、3次元点群の局所構造を捉えることが難しいという問題がある。

本稿では、自己符号化器と敵対的生成ネットワークに基づく生成モデルから生成される3次元点群の忠実性を改善する手法を提案する。提案手法では、マルチタスク学習の考えに基づき、自己符号化器を学習させる際に入力点群とその部分点群を同時に復元することによって得られる復元損失を用いてネットワークを最適化する。すなわち、入力点群の大域形状と局所形状の両方の復元に対して効果的な潜在表現を符号化するように、自己符号化器を学習させる方法を提案する。

本稿の貢献は以下の3点にまとめられる。

- 3次元点群を対象とする自己符号化器に対して、マルチタスク学習の考えに基づき、入力点群の部分点群を用いた学習手法を提案した。
- 形状的に連続する領域から部分点群を構築するためのサンプリングアルゴリズムを導入した。
- 提案手法によって生成された3次元点群の定量評価および定性評価を行い、その忠実性が改善されていることを確認した。

[†] 株式会社 KDDI 総合研究所 メディア認識グループ

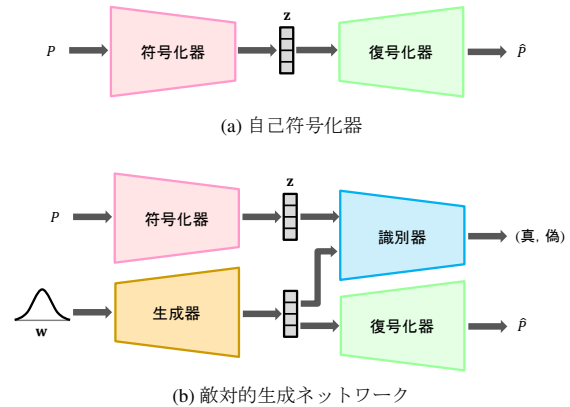


図1: 1-GANの学習方法の概要図。

2 関連研究

2.1 3次元点群生成

本節では、Achlioptasら [10] の提案した3次元点群生成モデルであるlatent-space generative adversarial networks (1-GAN) について概説する。このモデルは、図1に示すように自己符号化器と敵対的生成ネットワークを2段階で学習させる。始めに、図1(a)に示す符号化器と復号化器から成る自己符号化器を学習させる。この際、符号化器は入力点群 P を潜在変数 z に圧縮し、復号化器は z から P との損失が少ない出力点群 \hat{P} を復元することを学習する。その後、図1(b)に示す生成器と識別器から成る敵対的生成ネットワークを、自己符号化器の潜在空間上で学習させる。生成器は、ガウスノイズベクトル w を入力として潜在変数を生成する。識別器には、事前に学習された符号化器によって3次元点群を符号化した潜在変数を真のデータ、生成器によって生成された潜在変数を偽のデータとして与える。この際、識別器は真のデータと偽のデータを識別することを学習し、生成器は真のデータと識別できない偽のデータを生成することを学習する。3次元点群の生成は、ガウスノイズベクトル w から生成された潜在変数を、事前に学習された復号化器を用いて復号化することによって実現される。

2.2 マルチタスク学習

マルチタスク学習 [11] とは、共通の特徴表現を用いて関連性を持つ複数の予測タスクのためのモデルを同時に学習することによって、個々のタスクに対する予測精度を改善するアプローチである。このアプローチは、例えばシーン認識モデルと物体検出モデル [12] の同時学習や、法線推定モデルと深度推定モデル、輪郭抽出モデル [13] の同時学習のような、様々な機械学習タスクに対して広く利用されている。本稿では、自己符号化器による3次元点群の復元タスクに対してマルチタスク学習の考えを適用する。提案手法では、自己符号化器の学習時に共通の潜在表現から入力点群とその部分点群を同時に復元する。ここでは、入力点群とその部分点群の復元を、関連性を持つ異なるタスクであるとみなす。

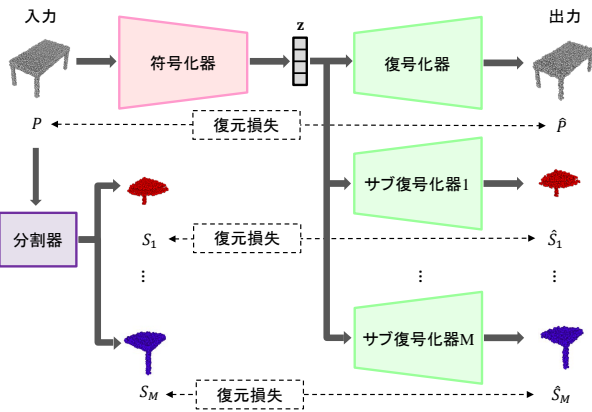


図 2: 提案手法のフレームワーク.

3 提案手法

3.1 概要

本稿では、3次元点群を対象とした生成モデルの性能を改善させるための学習手法を提案する。提案手法は、最先端の点群生成モデルである I-GAN [10] をベースとする。第 2.1 節に記載の通り、I-GAN は初めに自己符号化器を学習させ、その後学習済みの自己符号化器の潜在空間上で敵対的生成ネットワークを学習させる。本研究では、自己符号化器による潜在空間の学習精度を改善することによって、敵対的生成ネットワークによる点群の生成精度を改善することを目指す。

本稿で提案する自己符号化器のネットワーク構成を図 2 に示す。符号化器と復号化器は標準的な自己符号化器の構成要素であり、復号化器への入力とは N 個の点を持つ 3 次元点群 P 、復号化器の出力は同サイズの復元点群 \hat{P} である。更に、このネットワークは M 個のサブ復号化器を持つ。符号化器は P を潜在変数 \mathbf{z} に圧縮し、復号化器およびサブ復号化器は共通の潜在変数を用いて点群を復元する。ここで、復号化器は N 個の点を持つ点群 \hat{P} を復元し、 M 個のサブ復号化器は N/M 個の点を持つ点群 $\hat{S}_1, \dots, \hat{S}_M$ を追加的に復元する。点群 \hat{P} は点群 P と対応付けられ、点群 $\hat{S}_1, \dots, \hat{S}_M$ は分割器によって点群 P を分割して得られる M 個の部分点群 S_1, \dots, S_M と対応付けされる。このネットワークは、対応付けられた点群間の復元損失を最小化するように最適化される。

3.2 部分点群の構築

提案手法で利用される点群の分割器は、入力点群から空間的に局所化された点をサンプリングすることによって、複数の部分点群を構築する。部分点群の構築のために、あるクエリ点から指定半径以内の全ての点をサンプリングするアルゴリズムがしばしば利用される [14, 15]。しかしながら、このアルゴリズムはクエリ点の個数を大きくした場合であっても、入力点群の全てをカバーする保証はない。入力点群の全てをカバーするために、本稿では領域拡張サンプリングと呼ばれる新たなサンプリングアルゴリズムを導入する。このアルゴリズムは、領域拡張法に基づいて、あるシード点と形状的に連続する領域に位置する点をサンプリングする。

詳細を Algorithm 1 に示す。このアルゴリズムは、 N 個の点で構成される点群 P 、点群の分割数 M 、近傍探索の閾値 k を入力とする。 i 番目の部分点群 S_i に対して、

Algorithm 1 領域拡張サンプリング

Input: 点群 $P = \{x_1, \dots, x_N \in \mathbb{R}^3\}$, 分割数 M , 近傍探索の閾値 k
Output: 部分点群 $S_i = \{x_1, \dots, x_{N/M} \in \mathbb{R}^3\}$ ($i = 1, \dots, M$)

- 1: **for** $i = 1 \rightarrow M$ **do**
- 2: $\mathcal{N}(x) \leftarrow x \in P$ の k 近傍点
- 3: $g \leftarrow P$ の重心点
- 4: $s_i \leftarrow \operatorname{argmax}_{x \in P} \|x - g\|_2$
- 5: $S_i \leftarrow S_i \cup \{s_i\}$, $P \leftarrow P \setminus \{s_i\}$
- 6: **for** $j = 1 \rightarrow N/M - 1$ **do**
- 7: $C \leftarrow \bigcup_{x \in S_i} \mathcal{N}(x) \cap P$
- 8: **if** $C = \emptyset$ **then**
- 9: $y \leftarrow \operatorname{argmin}_{x \in P} \|x - s_i\|_2$
- 10: $C \leftarrow C \cup \{y\}$
- 11: **end if**
- 12: $y \leftarrow \operatorname{argmin}_{x \in C} \|x - s_i\|_2$
- 13: $S_i \leftarrow S_i \cup \{y\}$, $P \leftarrow P \setminus \{y\}$
- 14: **end for**
- 15: **end for**

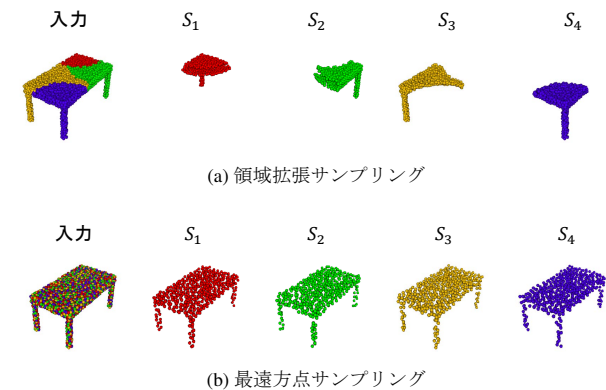


図 3: 点群分割の可視化.

P の重心から最も遠い点をシード点として格納する。その後、形状的に連続する領域に位置する点をサンプリングするために、 S_i 内の点の k 近傍点を候補点集合 C とみなす。 C が空集合である場合は、例外処理として P 内でシード点から最も近い点を C に格納する。そして、 C の中で最もシード点に近い点を選び、 S_i に格納する。各部分点群に対して、 N/M 個の点が格納されるまでサンプリングを繰り返す。結果として、 N/M 個の点で構成される M 個の部分点群が得られる。

図 3(a) に、 $M = 4$ とした場合の点群分割結果の例を示す。比較のために、図 3(b) に最遠方点サンプリング (Farthest Point Sampling, FPS) [17] に基づく点群分割の結果も示す。最も左の画像は入力点群を表しており、点の色は結果として得られる部分点群の色に対応している。空間的に一様にサンプリングする FPS とは対照的に、領域拡張サンプリングは空間的に局在化された点のサンプリングを実現することが分かる。

3.3 自己符号化器のネットワーク構成

符号化器への入力とは N 個の点を持つ 3 次元点群であり、出力は潜在変数と呼ばれる b 次元ベクトル \mathbf{z} であ

表 1: ModelNet10 データセットを用いた敵対的生成ネットワークによる 3 次元点群生成の定量評価.

Metric	Method	Category										
		bathtub	bed	chair	desk	dresser	monitor	stand	sofa	table	toilet	mean
JSD ↓ (×10)	Base [10]	0.535	0.170	0.337	0.486	0.525	0.390	0.620	0.112	0.330	0.251	0.375
	Prop	0.428	0.135	0.158	0.358	0.313	0.214	0.429	0.104	0.249	0.207	0.259
	Prop (FPS)	0.451	0.239	0.248	0.755	0.413	0.392	0.432	0.110	0.436	0.319	0.379
MMD ↓ (×10)	Base [10]	0.557	0.506	0.660	0.807	0.577	0.561	0.723	0.432	0.589	0.626	0.604
	Prop	0.515	0.485	0.634	0.775	0.523	0.520	0.702	0.416	0.565	0.592	0.573
	Prop (FPS)	0.522	0.511	0.662	0.799	0.534	0.560	0.710	0.432	0.599	0.630	0.596
COV ↑	Base [10]	0.339	0.421	0.388	0.400	0.425	0.320	0.390	0.392	0.474	0.404	0.395
	Prop	0.462	0.473	0.448	0.410	0.470	0.443	0.435	0.447	0.500	0.488	0.457
	Prop (FPS)	0.396	0.396	0.388	0.355	0.451	0.339	0.425	0.427	0.431	0.372	0.398

る. この符号化器は PointNet の設計に従い, カーネルサイズ 1 の 1 次元畳み込み層から成る. ここでは, ReLU 活性化 [18] およびバッチ正規化 [19] を各層に適用する.

復号器への入力は潜在変数 \mathbf{z} であり, 出力は N 個の点を持つ復元点群である. この復号器は, 最終層以外に ReLU 活性化を適用する全結合層から成るネットワークである. サブ復号器は, 出力の次元を除いて復号器と共通のネットワーク構成を持つ. サブ復号器への入力は潜在変数 \mathbf{z} であるが, 出力は N/M 個の点を持つ復元点群である.

3.4 損失関数の設計

本節では, 提案するネットワークの損失関数について説明する. 点群の復元損失の尺度として, [16] で提案された Earth Mover's distance (EMD) を使用する. 同一の点数を持つ 2 つの点群 $X_1 \subseteq \mathbb{R}^3$ および $X_2 \subseteq \mathbb{R}^3$ が与えられた場合, それらの EMD は次式で定義される.

$$EMD(X_1, X_2) = \min_{\phi: X_1 \rightarrow X_2} \sum_{x \in X_1} \|x - \phi(x)\|_2, \quad (1)$$

ここで $\phi: X_1 \rightarrow X_2$ は全単射である. 提案手法では, サイズ N の点群とサイズ N/M の点群に対する EMD を独立に計測する. これらを区別するために, 便宜的に前者を大域復元損失, 後者を局所復元損失と呼ぶ. 提案する損失関数はこれらの総和をとり, 次式で定義される.

$$\mathcal{L} = EMD(P, \hat{P}) + \lambda \sum_{i=1}^M EMD(S_i, \hat{S}_i), \quad (2)$$

ここで, λ は大域復元損失と局所復元損失のバランスを制御するための係数である.

4 評価実験

4.1 実験設定

評価のために, ModelNet10 [22] データセットを使用する. 本データセットは 10 通りの物体カテゴリに渡る大規模な 3 次元 CAD モデルで構成される. 各モデルから, ポアソンディスクサンプリング法 [23] を用いて $N = 2048$ 点の点群を構築した. また, 各点群は平均座標が原点に位置する様に平行移動すると共に, 全ての点が単位立方体に収まるように正規化した. 自己符号化器の学習のために, 初期学習率を 0.0005, 1 次モメンタムを 0.9 とした Adam 最適化器 [24] を使用する. 各点群の分割数は $M = 4$ とし, 各部分点群の点数は 512 に設定した. 潜在変数の次元は 128 とした. 領域拡張サンプリングにおける k 近傍探索では, $k = 5$ を使用した. 式 (2) における係数は, $\lambda = 0.1$ とした.

本実験においては, 次の 3 つの手法を比較する.

- **Base**: 従来手法 [10] による自己符号化器である. この手法は, 損失関数として式 (2) に記載の関数の第一項のみを用いる.
- **Prop**: Base に対してサブ復号器を追加した手法 (提案手法) である. 損失関数として式 (2) に記載の関数を用いる.
- **Prop (FPS)**: 部分点群の構築に最遠方点サンプリング (FPS) 法を用いる Prop の変法である. 損失関数は Prop と同様に式 (2) に記載の関数を用いる.

各手法に従って自己符号化器を学習させた後に, 図 1(b) に示す敵対的生成ネットワークを学習させる. この実験では, 異なる手法によって学習された符号化器および復号器は, 敵対的生成ネットワークの学習のために独立に使用される. また, 敵対的生成ネットワークの学習を安定させるために WGAN-GP 法 [20, 21] を適用した. 敵対的生成ネットワークの学習後に, ガウスノイズベクトルを生成器への入力として 3 次元点群を生成した.

4.2 実験結果

本節では, 3 次元点群に対する生成モデルの生成性能を評価する. 評価尺度として, [10] で導入された Jensen-Shannon Divergence (JSD), Minimum Matching Distance (MMD), Coverage (COV) を用いる. JSD は 2 つの点群集合の類似性を計測し, MMD と COV は 1 つの点群集合に対するもう一つの点群集合の忠実性を計測する. JSD と MMD は値が小さい方が優れていることを示し, COV は値が大きい方が優れていることを示す.

表 1 に, 10 カテゴリに対する定量評価結果を示す. Prop は, 全ての尺度に関して最良の結果を達成する. それは, Prop が Ground-truth (GT) 点群集合に対する生成点群集合の忠実性を改善したことを表す. Base と Prop を比較することによって, 局所復元損失の導入が点群生成の精度を改善することがわかる. これは, マルチタスク学習の使用により, シングルタスク学習と比べてより良い潜在表現を学習したことに起因する. Prop とは対照的に, Prop (FPS) は Base と比べて点群生成の精度をほとんど改善しない. これは, FPS 法によって構築される部分点群と入力点群の形状的な差異が少ないために, マルチタスク学習の効果が減少したためと考えられる.

図 4 は, Base および Prop によって生成された 3 次元点群を可視化している. この図では, 各行の GT 点群に対して, 各手法によって生成された点群の集合から探索された最近傍点群を示す. この探索における距離尺度

には、EMDを使用した。点群の色は、GT点群から生成点群への間の一方方向のハウスドルフ距離を表している。図より、PropはBaseと比べて物体の詳細な部位まで忠実な点群を生成することがわかる。この結果は、GT点群集合の中の一つに近い点群を生成することに関して、PropはBaseより優れた性能を持つことを表す。しかし、ベッドのフレームのような複雑な構造については、細部まで忠実に復元された点群を生成できていないことがわかる。この問題を解決することは、今後の課題とする。

5 おわりに

本稿では、自己符号化器と敵対的生成ネットワークを用いた3次元点群生成モデルの性能を改善する手法を提案した。提案手法はマルチタスク学習の考えに基づき、自己符号化器を入力点群とその部分点群を同時に復元することによって学習させる。また、形状的に連続する領域から部分点群を構築するためのサンプリングアルゴリズムを導入した。構築された部分点群は自己符号化器によって復元された部分点群と対応付けられ、局所復元損失の計測に利用された。評価実験により、提案手法は従来手法と比較してより忠実性の高い3次元点群を生成できることを示した。今後の課題として、より複雑な構造を忠実に復元する手法の検討や、データ拡張のような他の課題へ応用が考えられる。

参考文献

- [1] D. Maturana and S. Scherer, "VoxNet: A 3D Convolutional Neural Network for Real-Time Object Recognition," In *Proc. of IROS*, pp. 922–928, 2015.
- [2] H. Su and S. Maji and E. Kalogerakis and E. Learned-Miller, "Multi-view Convolutional Neural Networks for 3D Shape Recognition," In *Proc. of ICCV*, pp. 945–953, 2015.
- [3] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "Pointnet: Deep Learning on Point Sets for 3D Classification and Segmentation," In *Proc. of CVPR*, pp. 5099–5108, 2017.
- [4] C. R. Qi and W. Liu and C. Wu and H. Su and L. J. Guibas, "Frustum PointNets for 3D Object Detection from RGB-D Data," In *Proc. of CVPR*, pp. 918–927, 2018.
- [5] H. Deng and T. Birdal and S. Ilic, "PPFNet: Global Context Aware Local Features for Robust 3D Point Matching," In *Proc. of CVPR*, pp. 195–205, 2018.
- [6] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Nets," In *Proc. of NeurIPS*, pp. 2672–2680, 2014.
- [7] S.-W. Huang and C.-T. Lin and S.-P. Chen and Y.-Y. Wu and P.-H. Hsu and S.-H. Lai, "AugGAN: Cross Domain Adaptation with GAN-based Data Augmentation," In *Proc. of ECCV*, pp. 718–731, 2018.
- [8] T. Karras and T. Aila and S. Laine and J. Lehtinen, "Progressive Growing of GANs for Improved Quality, Stability, and Variation," In *Proc. of ICLR*, 2018.
- [9] C. Vondrick and H. Pirsiavash and A. Torralba, "Generating Videos with Scene Dynamics," In *Proc. of NeurIPS*, pp. 613–621, 2016.
- [10] P. Achlioptas, O. Diamanti, I. Mitliagkas, and L. Guibas, "Learning Representations and Generative Models for 3D Point Clouds," In *Proc. of ICML*, pp. 40–49, 2018.
- [11] R. Caruana, "Multitask learning," *Machine Learning*, vol. 28, no. 1, pp. 41–75, 1997.
- [12] H. Sun and Z. Meng and P. Y. Tao and M. H. Ang, "Scene Recognition and Object Detection in A Unified Convolutional Neural Network on A Mobile Manipulator," In *Proc. of ICRA*, pp. 5875–5881, 2018.
- [13] Z. Ren and Y. J. Lee, "Cross-Domain Self-supervised Multi-task Feature Learning using Synthetic Imagery," In *Proc. of CVPR*, pp. 762–771, 2018.
- [14] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "Pointnet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space," In *Proc. of NeurIPS*, pp. 2672–2680, 2017.
- [15] G. Elbaz, T. Avraham, and A. Fischer, "3D Point Cloud Registration for Localization using a Deep Neural Network Auto-Encoder," In *Proc. of CVPR*, pp. 2472–2481, 2017.
- [16] H. Fan, H. Su, and L. Guibas, "A Point Set Generation Network for 3D Object Reconstruction from a Single Image," In *Proc. of CVPR*, pp. 605–613, 2017.
- [17] Y. Eldar, M. Lindenbaum, M. Porat, and Y. Y. Zeevi, "The Farthest Point Strategy for Progressive Image Sampling," *IEEE Trans. on IP*, vol. 6, no. 9, pp. 1305–1315, 1997.
- [18] V. Nair and G. E. Hinton, "Rectified Linear Units Improve Restricted Boltzmann Machines," In *Proc. of ICML*, pp. 807–814, 2010.
- [19] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," In *Proc. of ICML*, pp. 448–456, 2015.
- [20] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," *arXiv:1701.07875*, 2017.
- [21] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, "Improved Training of Wasserstein GANs," In *Proc. of NeurIPS*, pp. 5767–5777, 2017.
- [22] Z. Wu, S. Song, A. Khosla, F. Yu, L. Zhang, X. Tang, and J. Xiao, "3D ShapeNets: A Deep Representation for Volumetric Shapes," In *Proc. of CVPR*, pp. 1912–1920, 2015.
- [23] G. Ranzuglia, M. Callieri, M. Dellepiane, P. Cignoni, R. Scopigno, "Efficient and Flexible Sampling with Blue Noise Properties of Triangular Meshes," *IEEE Trans. on VCG*, vol. 18, no. 6, pp. 914–924, 2012.
- [24] D. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *arXiv:1412.6980*, 2014.

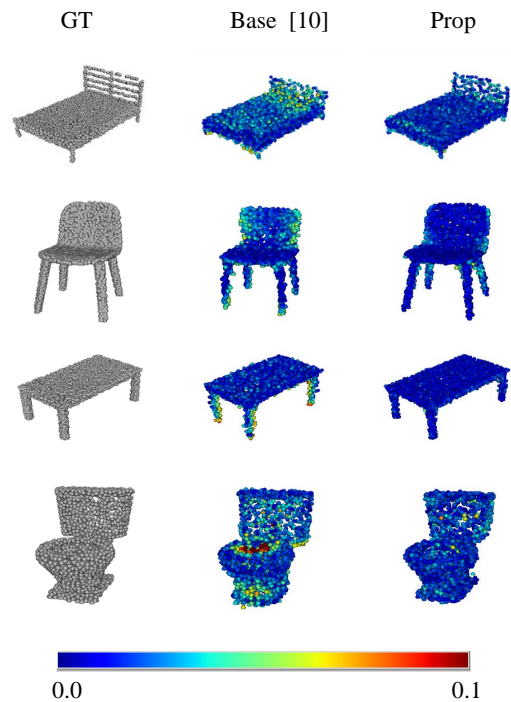


図4: 3次元点群生成の定性評価。