

Deep Implicit Moving Least-Squares Functions for 3D Reconstruction

Shi-Lin Liu^{1,3} Hao-Xiang Guo^{2,3} Hao Pan³ Peng-Shuai Wang³ Xin Tong³ Yang Liu³

¹University of Science and Technology of China ²Tsinghua University ³Microsoft Research Asia

freelin@mail.ustc.edu.cn ghx17@mails.tsinghua.edu.cn

{haopan, penwan, xtong, yangliu}@microsoft.com

Abstract

Point set is a flexible and lightweight representation widely used for 3D deep learning. However, their discrete nature prevents them from representing continuous and fine geometry, posing a major issue for learning-based shape generation. In this work, we turn the discrete point sets into smooth surfaces by introducing the well-known implicit moving least-squares (IMLS) surface formulation, which naturally defines locally implicit functions on point sets. We incorporate IMLS surface generation into deep neural networks for inheriting both the flexibility of point sets and the high quality of implicit surfaces. Our IMLSNets predicts an octree structure as a scaffold for generating MLS points where needed and characterizes shape geometry with learned local priors. Furthermore, our implicit function evaluation is independent of the neural network once the MLS points are predicted, thus enabling fast runtime evaluation. Our experiments on 3D object reconstruction demonstrate that IMLSNets outperform state-of-the-art learning-based methods in terms of reconstruction quality and computational efficiency. Extensive ablation tests also validate our network design and loss functions.

1. Introduction

Point set is probably the most widely used representation for 3D deep learning. Compared with other 3D representations like polygonal meshes and volumetric grids, point sets are naturally embedded as neurons in DNNs, easy to acquire, have minimal extra structure to maintain, capture complex geometry and topology dynamically, and induce no wasted computation for free-space regions. Indeed, point sets have been used in the deep learning-based 3D analysis for diverse tasks [20, 39, 40]. However, when using point sets for the generation of 3D data by deep learning, we have the flexibility to model changing topology and complex surfaces on one hand, but also suffer from the discrete and rough geometry on the other hand.

Recent works have focused on generating shapes in the forms of meshes and polygonal patches, but their shape representation abilities are still restricted by their discrete and non-smooth nature. The deep implicit function approaches [9, 32, 36], instead, define smooth functions on entire 3D domains to guarantee result continuity, and have shown to be promising for high-quality 3D reconstruction. However, the implicit surface generation is inefficient, because for each point in the 3D domain the network has to be evaluated individually before the surface can be extracted. In this paper, we combine the advantages of both the implicit function approaches and the point set methods, by extending the point set representation to model implicit surfaces for high-quality 3D generation, while preserving the inherent flexibility and computational efficiency of explicit point sets.

For modeling smooth surfaces via point sets, we adopt point set surfaces [2] and use moving least-squares (MLS) interpolation of the points to define locally implicit functions over a narrow band region of the point set. Specifically for the implicit MLS formulation [27] used in this paper, for any spatial point inside the narrow region, the implicit MLS function maps it to a signed distance value to the zero level set surface, defined by the weighted blending of signed distances to oriented planes supported at the nearby points; the zero level set surface can then be extracted as a smooth and continuous surface for shape representation.

While MLS surfaces have been well studied in 3D reconstruction and rendering [10], to incorporate the representation into a deep learning framework poses novel challenges and opportunities unseen in existing point-set-based or implicit-representation-based methods. First, the points define implicit MLS surfaces most effectively when they are dense enough and distributed evenly over the shapes to reconstruct. While most point generation approaches fix the number of points and consume many resources for the straightforward prediction of dense points that are hard to generalize, we introduce an *octree-based scaffolding* for generating variable numbers of MLS points only where needed according to the target shape (*cf.* Fig. 1), and further regularize the point distribution via tailored loss functions. Second,

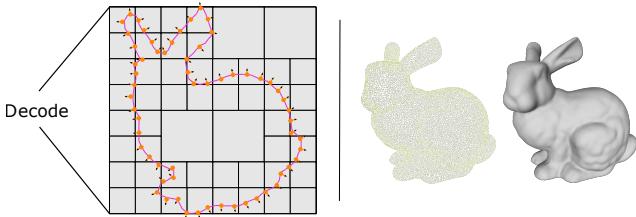


Figure 1: Illustration of our deep implicit MLS surface generation method. Left: the network decoder produces a set of oriented points scaffolded by an octree (2 points for each non-empty leaf octant in this example), which defines an implicit zero level set surface that is the target shape for shape generation. Right: an example of the reconstructed bunny shape from 3D points via an implicit MLS.

to measure the predicted implicit functions for both training supervision and test evaluation, while existing implicit methods [9, 32, 36, 38] have to use dense sampling over the whole 3D domain, the MLS surface is naturally localized inside the narrow band region of generated points, which prompts us to use more succinct sampling only at the octree nodes for supervision and evaluation. Also, the evaluation is independent of the network once all MLS points are embedded in the 3D domain, thus avoiding the costly per-point network evaluation that is typical of other implicit methods.

We use extensive ablation tests to validate the design choices made. We also demonstrate that our deep implicit MLS surface approach outperforms both the point set generation methods and other global or local implicit function methods, through the 3D object reconstruction task.

2. Related Work

Deep representations for 3D generation Point set is a popular representation utilized by many works [1, 42, 54, 56]. However, as the number of points is usually specified, its power for representing detailed geometry is restrained and needs further point upsampling [57, 58] to improve the point density and shape quality.

Dense voxels [5, 13, 53] represent shape occupancy well, but its high memory cost precludes its usage for representing 3D contents with high-resolution. Sparse voxels including octrees [12, 21, 41, 45, 49] overcome these issues with great efficiency in both memory and computational aspects.

Mesh representations [24, 48] and patch-based representations [14, 17, 51, 52, 55] are convenient 3D representations for improving shape quality. However, their abilities are constrained by either the predefined mesh topology and resolution, or the disconnectivity of multiple patches. Intermediate 3D representations like coarse voxels [16] or shape skeleton [44] are possible ways to further enhance their quality.

Primitive-based representations use a set of simple geometric objects like planes [30] and cuboids [43, 47, 59]

to approximate 3D shapes. Structure-based representations [29, 33] explicitly encode semantic part structures as cuboids and reconstruct parts within each cuboid using voxel representations. Although they are suitable to characterize shape structures, their approximation quality is also limited due to the simplicity of primitives.

Recently implicit surface-based deep learning approaches [3, 9, 32, 36] offer a smooth and continuous 3D representation and enable functional evaluation in a continuous space. For a given point, the network predicts its occupancy or the signed distance from it to the surface. These techniques are further improved recently by incorporating local features to model more detailed shape geometry [7, 11, 23, 38].

Our approach belongs to the deep implicit category by modeling smooth and continuous implicit MLS surfaces, while also enjoying the flexibility and efficiency of explicit point set generation; it is a hybrid 3D deep learning representation that combines advantages from both point sets and implicit functions.

Surface reconstruction from point clouds Surface reconstruction techniques have been studied for several decades (*cf.* the comprehensive survey [4]). Among them, a set of methods impose global or local smoothness priors for reconstructing high-quality results from point clouds, including multi-level partition of unity (MPU) [34], Poisson reconstruction [25, 26], radial basis functions (RBF) [6], and moving least-squares surfaces (MLS) [2, 19, 28] which are widely used for point set surface modeling and rendering [37]. Due to the fast and local evaluation property of MLS, we choose MLS surfaces as our deep 3D representation. MLS surfaces can be classified into two types [10]: projection MLS surfaces and implicit MLS surfaces (IMLS). The former is defined by a set of stationary points via iterative projection, while the latter defines an implicit function directly. We use IMLS for incorporating signed distance supervision easily and enabling fast function evaluation.

3. Method

We first review the implicit MLS surface (IMLS) definition and present our deep IMLS design in Sections 3.1 and 3.2, then we detail the network design and loss functions in Sections 3.3 and 3.4.

3.1. IMLS surface

The implicit MLS surface [27] is defined as follows. Denote $\mathcal{P} = \{\mathbf{p}_i \in \mathbb{R}^3\}_{i=1}^N$ as a set of 3D points and each point is equipped with unit normal vector $\mathbf{n}_i \in \mathbb{R}^3$ and control radius $r_i \in \mathbb{R}^+$. For convenience, we call these points as *MLS points*.

For each MLS point \mathbf{p}_i , a signed distance function from $\mathbf{x} \in \mathbb{R}^3$ to its tangent plane is defined as $\langle \mathbf{x} - \mathbf{p}_i, \mathbf{n}_i \rangle$, where

$\langle \cdot, \cdot \rangle$ is the inner product. By weighted averaging of all pointwise signed distance functions, we have an implicit function $F(\mathbf{x})$ whose zero level set defines the implicit surface \mathcal{S} :

$$F(\mathbf{x}) := \frac{\sum_{\mathbf{p}_i \in \mathcal{P}} \theta(\|\mathbf{x} - \mathbf{p}_i\|, r_i) \cdot \langle \mathbf{x} - \mathbf{p}_i, \mathbf{n}_i \rangle}{\sum_{\mathbf{p}_i \in \mathcal{P}} \theta(\|\mathbf{x} - \mathbf{p}_i\|, r_i)}. \quad (1)$$

Here we set the weight function as $\theta(d, r) = \exp(-d^2/r^2)$. Kolluri proved that *under a uniform sampling condition, the IMLS surface \mathcal{S} is a geometrically and topologically correct approximation of the original surface where \mathcal{P} are sampled from, and the IMLS function F is a tight approximation of the signed distance function of the original surface [27]*.

As the weight function decays when \mathbf{x} is away from \mathbf{p}_i , the evaluation of $F(\mathbf{x})$ can be accelerated by considering nearby MLS points only. Eq. (1) can be revised as:

$$F(\mathbf{x}) := \frac{\sum_{\mathbf{p}_i \in \Omega(\mathbf{x})} \theta(\|\mathbf{x} - \mathbf{p}_i\|, r_i) \cdot \langle \mathbf{x} - \mathbf{p}_i, \mathbf{n}_i \rangle}{\sum_{\mathbf{p}_i \in \Omega(\mathbf{x})} \theta(\|\mathbf{x} - \mathbf{p}_i\|, r_i)}, \quad (2)$$

where $\Omega(\mathbf{x})$ denotes the set of MLS points that are inside the ball centered at \mathbf{x} with radius r_b . r_b can be set by the user as the truncating point distance values.

Due to the above formulation, the zero function values $F(\mathbf{x})$ exist in a narrow band region of the IMLS points. Extracting \mathcal{S} explicitly as a triangle mesh can be done efficiently via Marching cubes [31], by restricting the functional evaluation on the regular grids inside the bounded region.

3.2. Deep IMLS surface

Sparse and unoriented point clouds probably with noise and missing regions are typical inputs for 3D reconstruction in practice, but they cannot be handled well by traditional 3D reconstruction methods like Poisson reconstruction. To deal with this kind of imperfect data, we aim to design an auto-encoding neural network to generate IMLS surfaces.

A naïve way of defining the network output is to set a fixed-number of IMLS point tuples: $\{\mathbf{p}_i, \mathbf{n}_i, r_i\}_{i=1}^N$, similar to existing point set generation approaches [42]. However, it will constrain the representation capability of IMLS and cannot learn local geometry priors well from the data. We introduce an intermediate network output – *octree-based scaffold*, to help generate MLS points as needed. The octree-based scaffold is a d -depth octree \mathcal{O} which roughly approximates the 3D surface in multi-resolutions. For each finest non-empty octant o_k , i.e., the smallest non-empty voxel in the octree, we associate a small set of MLS points whose locations are near to the octant center \mathbf{c}_k . To be specific, the MLS points associated with o_k are defined as

$$\mathbf{p}_{k,l} := \mathbf{c}_k + \mathbf{t}_{k,l}, \quad l = 1, \dots, s, \quad (3)$$

where $\mathbf{t}_{k,l} \in \mathbb{R}^3$ is the offset vector from $\mathbf{p}_{k,l}$ to \mathbf{c}_k , s is a predefined point number. As the structure of the octree-based

scaffold depends on the target surface, the total number of MLS points and their positions are determined adaptively.

With the setup above, a suitable network for IMLS generation should output: (1) an octree-based scaffold \mathcal{O} ; (2) the MLS point offsets, MLS point normals and control radii for each finest non-empty octant o_k , denoted by $\mathbf{t}_{k,l}$, $\mathbf{n}_{k,l}$ and $r_{k,l}$, respectively.

Here we note that the octree created from the input noisy and sparse point cloud cannot be used as the scaffold, as it could be incomplete and inaccurate, and different from that of the target shape.

Scaffold prediction We use the octree-based convolutional neural network (O-CNN) autoencoder [49, 51] to generate the scaffold. Its encoder takes a d_{in} -depth octree as input which is constructed from the input point cloud, and performs the CNN computation within the octree only. Its decoder starts with $4 \times 4 \times 4$ cells and predicts whether each cell is empty or not, and subdivides it into eight octants if the cell is not empty. This process is performed on each non-empty octant recursively until the maximum output octree depth d_{out} is reached.

MLS point prediction Unlike previous works [50, 51] where the decoders regress an oriented point or a planar patch at each finest non-empty octant for achieving sub-voxel precision, we predict s MLS point tuples based on the feature vector at the octant, denoted by $f(o_k)$, via a multi-layer perceptron (MLP) with one hidden-layer as follows:

$$\text{MLP} \circ f(o_k) = (\mathbf{t}_{k,1}, \mathbf{n}_{k,1}, r_{k,1}, \dots, \mathbf{t}_{k,s}, \mathbf{n}_{k,s}, r_{k,s}). \quad (4)$$

Note that the MLP predicts the local coordinates of $\mathbf{p}_{k,l}$, i.e., $\mathbf{t}_{k,l}$, thus it can learn the local prior from the data. To ensure $\mathbf{p}_{k,l}$ is close to \mathbf{c}_k , the value range of each coordinate component of $\mathbf{t}_{k,l}$ is restricted within $[-\beta h, \beta h]$, here h is the size of the finest octant and β is set to 1.5 by default. We also constrain $r_{k,s}$ within $[l_r/2, 2l_r]$, where $l_r = h/\sqrt{s}$. These constraints are implemented by using the tanh activation for the network output and scaling the value by its range scale.

3.3. Network structure

We use a U-Net-like O-CNN autoencoder [50] that contains O-CNN ResNet blocks and output-guided skip-connections for better preserving the input geometry and predicting missing regions. The output-guided skip-connection is the skip connection added between each non-empty octant of the output octree and its corresponding octant of the input octree at the same octree level, if the latter octant exists.

For a given unoriented point cloud, we construct a d_{in} -depth octree for it. At each finest octant, we set the input 4-dimensional signal by concatenating the offset from the average position of input points inside the octant to the octant

center with a binary scalar indicating whether the octant is empty.

The network structure is illustrated in Fig. 2. Resblock(n, c) denotes an n -layer O-CNN-based residual block with channel number c . Downsample(c) and Upsample(c) are octree-based convolution and deconvolution operators [49] followed by batch normalization and ReLU. c is set to 64 for the first Resblock, and increases by a factor of 2 after each Downsample operator, and divided by 2 after each Upsample operator. In our experiments, we set $n = 3$. A hidden-layer MLP is used for predicting whether the octant is empty or not.

3.4. Loss function design

Our network is trained in a supervised manner. For each shape in the training dataset, we precompute its ground-truth d_{out} octree \mathcal{O}_o and use it for supervising the scaffold prediction. The octants at the first and second levels of \mathcal{O}_o are set to be non-empty. A succinct set \mathbf{Q} of sample points which bound the ground truth surface tightly will be used for probing the implicit function accurately. For generating these points, please refer to Section 4. We then compute the ground-truth signed distance function (SDF) values $F_o(\mathbf{q})$ and the SDF gradient $\nabla F_o(\mathbf{q})$, for each $\mathbf{q} \in \mathbf{Q}$. Our main training objective is to fit the signed distance field sampled on \mathbf{Q} .

The loss function consists of the following terms.

Octree structure loss The determination of octant status is a binary classification problem: 0 for empty and 1 for non-empty. We use the weighted summation of the sigmoid cross-entropy loss at every octant of \mathcal{O} to define the octree structure loss.

$$L_{\text{oct}} := \lambda_o \sum_{j=3}^{d_{out}} \frac{1}{|\mathcal{O}_j|} \sum_{o \in \mathcal{O}_j} \text{cross_entropy}(o), \quad (5)$$

here \mathcal{O}_j is the octant set at level j , and $|\mathcal{O}_j|$ denotes its size.

SDF loss The difference between the predicted IMLS surface and the ground-truth SDF field is defined by the SDF loss L_{sdf} .

$$L_{\text{sdf}} := \sum_{\mathbf{q} \in \mathbf{Q}} \lambda_s \|F(\mathbf{q}) - F_o(\mathbf{q})\|^2 + \lambda_g |1 - \nabla F(\mathbf{q}) \cdot \nabla F_o(\mathbf{q})|. \quad (6)$$

Here the gradient of F can be approximated by:

$$\nabla F(\mathbf{x}) \approx \frac{\sum_{\mathbf{p}_i \in \Omega(\mathbf{x}, \mathcal{P})} \theta(\|\mathbf{p}_i - \mathbf{x}\|, r_i) \cdot \mathbf{n}_i}{\sum_{\mathbf{p}_i \in \Omega(\mathbf{x}, \mathcal{P})} \theta(\|\mathbf{p}_i - \mathbf{x}\|, r_i)}. \quad (7)$$

To speed up the computation, we restrict the region of $\Omega(\mathbf{x}, \mathcal{P})$ by the octants whose distances from their centers to \mathbf{x} are less than $4h$.

MLS point repulsion loss Inspired by the repulsion term used by [22] for distributing consolidated points regularly, we introduce the point repulsion loss to improve the local regularity of our generated MLS points.

$$L_{\text{rep}} := \lambda_{\text{rep}} \sum_{\mathbf{p}_i} \sum_{\mathbf{p}_j \in \Omega(\mathbf{p}_i)} -w_{ij} \|\mathbf{p}_i - \mathbf{p}_j\|_{\text{proj}}, \quad (8)$$

where $\|\mathbf{p}_i - \mathbf{p}_j\|_{\text{proj}} = \|(\mathbf{I} - \mathbf{n}_i \mathbf{n}_i^T)(\mathbf{p}_i - \mathbf{p}_j)\|$ is the length of the projection of $\mathbf{p}_i - \mathbf{p}_j$ onto the tangent plane at \mathbf{p}_i , and w_{ij} is a bilateral weight with respect to both the MLS point difference and normal difference, defined as follows.

$$w_{ij} = \exp(-\|\mathbf{p}_i - \mathbf{p}_j\|^2/r_j^2 - (1 - \langle \mathbf{n}_i, \mathbf{n}_j \rangle)). \quad (9)$$

The above design pushes \mathbf{p}_j away from \mathbf{p}_i , especially when their normals and their positions are similar.

Projection smoothness loss For achieving local surface smoothness, we encourage MLS points to be close to the tangent planes of their neighboring MLS points.

$$L_{\text{proj}} = \lambda_p \sum_{\mathbf{p}_i} \sum_{\mathbf{p}_j \in \Omega(\mathbf{p}_i)} w_{ij} \langle \mathbf{n}_i, \mathbf{p}_i - \mathbf{p}_j \rangle^2. \quad (10)$$

Radius smoothness loss Similarly, for improving surface smoothness, the radius change of neighboring MLS points is penalized via weighted-Laplacian smoothing on radii.

$$L_{\text{rad}} = \lambda_r \sum_{\mathbf{p}_i} \left\| r_i - \frac{\sum_{\mathbf{p}_j \in \Omega(\mathbf{p}_i)} w_{ij} \cdot r_j}{\sum_{\mathbf{p}_j \in \Omega(\mathbf{p}_i)} w_{ij}} \right\|^2. \quad (11)$$

Weight decay A small weight decay with a coefficient λ_w is added to the loss function.

The above terms are added up together to form the total loss function. The coefficients of loss terms are specified in Section 4.

4. Experiments

We evaluate the efficacy of IMLSNets for 3D object reconstruction and compare it against the state-of-the-art 3D reconstruction methods. Extensive ablation tests are reported in Section 4.2.

Dataset We use 3D models from 13 shape classes of ShapeNet [8] as our training and test datasets. The dataset split strategy and the input point cloud preparation follow the setup of [32, 38]: each input point cloud contains 3000 points, randomly sampled from the ground-truth shape and Gaussian noise with standard deviation set to 0.005 of the maximum bounding box side length of the shape. For each 3D model, we prepare its ground-truth octree for training.

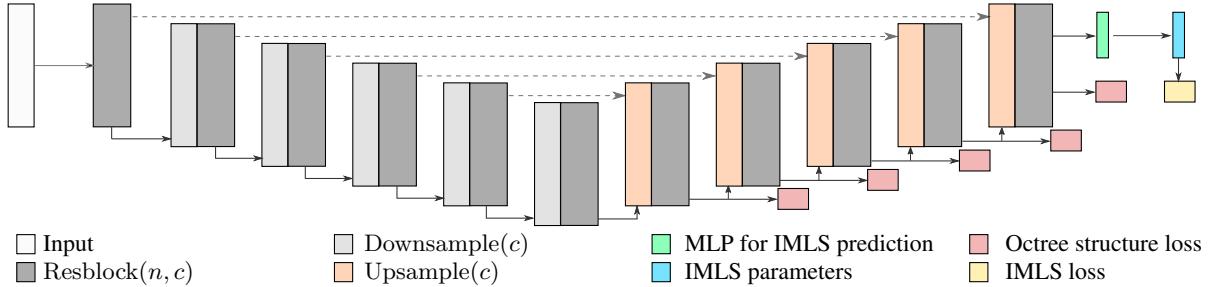


Figure 2: Deep IMLS Network structure for 3D reconstruction from point clouds. Each block after the input represents the CNN computation at the octants in one level of the octree, from fine to coarse on the encoder, and vice versa on the decoder.

The ground-truth octree is built, by recursively subdividing the nonempty octants from the root node to the maximum depth d_{out} [49]. The set \mathbf{Q} of grid points slightly larger than the nonempty leaf nodes of the octree is built to probe the SDF values for implicit function supervision. The exact training data preparation is provided in Appendix A.

Parameters The octree depth for storing the ground-truth signed distance field is 7. The coefficients of loss terms are set as follows: $\lambda_o = 0.1$, $\lambda_g = 0.05$, $\lambda_p = \lambda_r = 10$, $\lambda_{rep} = 0.05$, $\lambda_w = 5 \times 10^{-5}$. The coefficient of the SDF loss term depends on the octree level of SDF sample points: $\lambda_s = 200$ if the level is 6, $\lambda_s = 800$ if the level is 7. The selection of neighboring MLS points $\Omega(\mathbf{x})$ is speedup via space partitions: 10 nearest MLS points are selected in our experiments.

Network configuration We use the octree-based encoder and decoder with output-guided skip connections [50]. The network is denoted as IMLSNet(d_{in}, d_{out}, s), where d_{in} and d_{out} are the input and output octree depths, and s is the number of MLS points associated with each non-empty finest octants. IMLSNet(7,7,1) is set by default.

Network training Our networks were trained using Adam optimizer ($\text{lr}=0.001$), with batch size 32. We decay the learning rate by 20% after every 10 epochs if it is greater than 0.0001. The gradient update of w_{ij} is disabled in our training as we found that back-propagating through w_{ij} to position and normal variables causes training instability, especially during the early stages when network predictions are very inaccurate. We follow the curriculum SDF training strategy [15] to improve the training quality: the SDF sample points at the coarse octree level (depth=6) are first used, and after 30 epochs, we use sample points at a deeper level (depth=7). All the experiments were run on a Linux server with an Intel Core I7-6850K CPU (3.6GHz) and a GeForce GTX 1080 Ti GPU (11 GB memory). The network parameter size of IMLSNet(7,7,1) is about 4.3 M. More details

are reported in Appendix C, and our code is available at <https://github.com/Andy97/DeepMLS/>.

Network efficiency Averagely, our network forwarding time takes 30 milliseconds per point cloud input. The generated MLS points present a good preview of the final surface. Our naïve CPU-based marching-cube implementation takes 1 second averagely for a single shape, including file IO time.

Evaluation metrics We evaluate the reconstruction quality of our results by a set of common metrics: L_1 -Chamfer distance (CD_1) and normal difference (ND) measure bi-directional point-wise and normal error from ground truth; F-score(τ) [46] is the harmonic mean of recall and precision under the tolerance τ ; Occupancy IOU (IOU) computes the volumetric IOU. Our surface results are triangle meshes extracted via Marching cubes on a 128^3 grid. The CD metric is scaled by 10, τ is set to 0.1 by default. Similar metrics on the MLS point sets are also measured to reflect the reconstruction capability of MLS point sets. The detailed metric formulae and other measurements are provided in Appendix B.

4.1. 3D object reconstruction from point clouds

As IMLSNets learn local priors, the learned features do not rely on specific shape classes. We trained IMLSNet(7,7,1) on the ShapeNet dataset (13 classes) in a class-agnostic way and tested our network on their test data of 13 classes, and 5 unseen shape classes (bathtub, bag, bed, bottle, pillow). We also compared our method with two classical reconstruction methods: robust implicit moving least squares (RIMLS) [35] and Screened Poisson reconstruction (SPR); and three learning-based methods: occupancy network (OccNet) [32], convolutional occupancy network(ConvOccNet) [38] and O-CNN based completion network (O-CNN-C) [50]. OccNet and ConvOccNet are state-of-the-art deep implicit function-based methods and the work of [38] shows that ConvOccNet performs better than both OccNet and DeepSDF [36]. O-CNN-C generates a set of oriented points on the predicted octree. For RIMLS, Screened

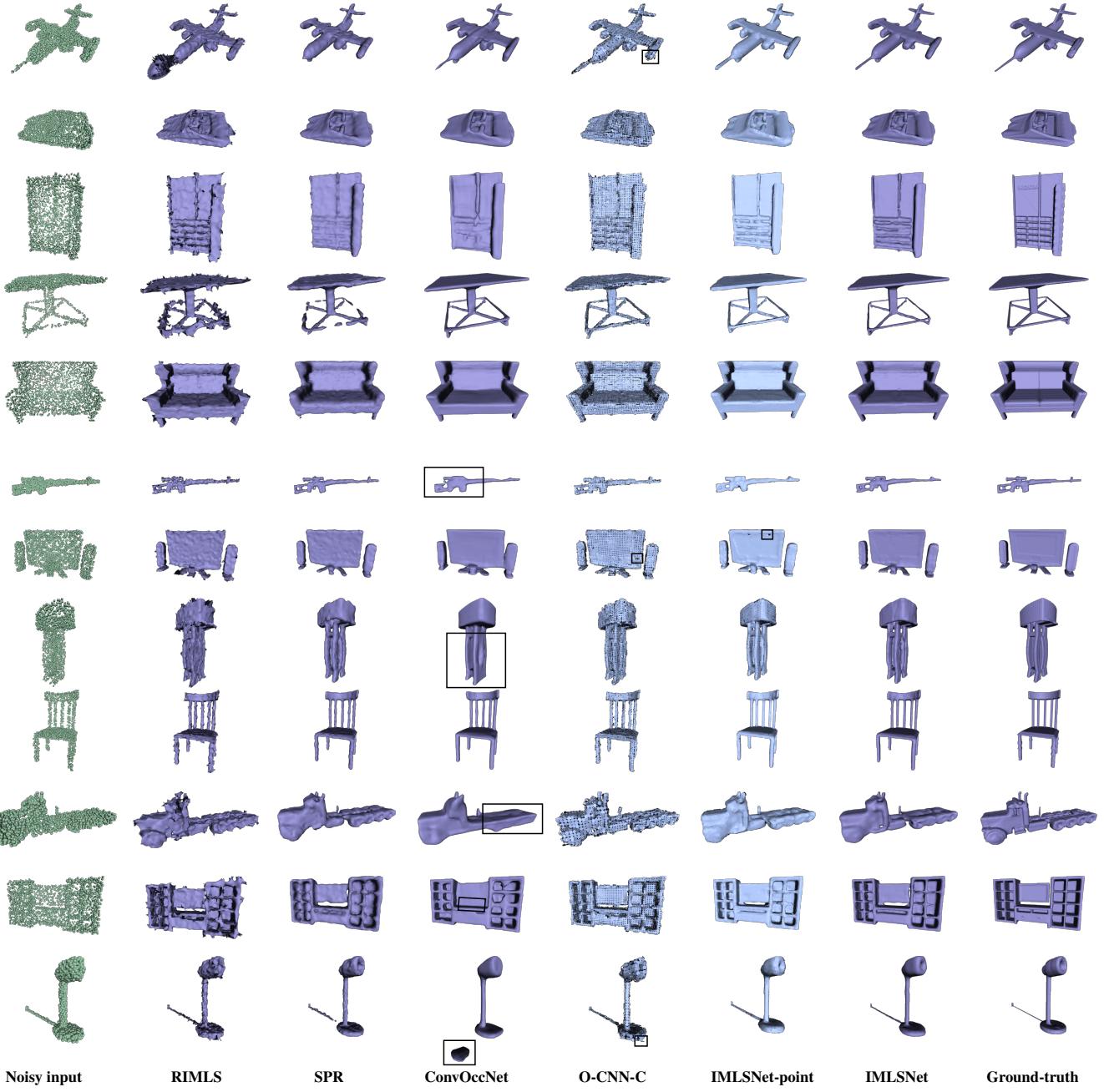


Figure 3: Visual results of 3D object reconstruction from point clouds by different methods. For point outputs of O-CNN-C and IMLSNet-point, we render points as oriented disks. Some artifacts in the results are labeled by black boxes.

Poisson, and O-CNN-C, we provide the ground-truth point normals as the input. We use the pretrained O-CNN-C whose depth of the output octree is 6.

Table 1 shows that our network achieves the best performances among all the compared methods. From the visual comparison (Fig. 3), we can observe that all the learning-based methods handle noisy and incomplete point inputs much better than classical non-learning reconstruction meth-

ods. Our IMLSNet results contain more details and capture small parts better, while other deep implicit-function methods tend to miss these details (see the examples in the 6th, 8th and 10th rows). This advantage of IMLSNet over ConvOccNet which uses fixed grids and implicit functions is due to the proper combination of adaptive and sparse grids (octree), point set, and implicit representations that together allow for more adaptive modeling of detailed geometry. In

| Network | $CD_1 \downarrow$ | $NC \uparrow$ | $IoU \uparrow$ | F-Score \uparrow |
|--------------|---------------------|---------------------|---------------------|---------------------|
| O-CNN-C | 0.067(0.067) | 0.932(0.945) | n/a | 0.800(0.802) |
| IMLSNet pts. | 0.035(0.037) | 0.941(0.954) | n/a | 0.985(0.983) |
| OccNet | 0.087 | 0.891 | 0.761 | 0.785 |
| ConvOccNet | 0.044(0.053) | 0.938(0.948) | 0.884(0.902) | 0.942(0.916) |
| IMLSNet | 0.031(0.034) | 0.944(0.956) | 0.914(0.939) | 0.983(0.981) |

Table 1: Quantitative evaluation of different networks on the test data of 13 shape classes and the full data of 5 unseen shape classes. Numbers in parentheses are for the unseen classes and numbers in bold are the best. The evaluation on each shape class and the training configuration of compared networks are reported in Appendix E.

| Network | plane | car | chair | rifle | table |
|------------|----------------|---------------|---------------|----------------|---------------|
| ConvOccNet | 1540.47 | 731.84 | 908.04 | 1433.82 | 660.10 |
| IMLSNet | 1326.22 | 649.69 | 710.38 | 1211.96 | 608.87 |

Table 2: LFD metric evaluation (lower values are better). The comparison between IMLSNet(7,7,1) and ConvOccNet shows that the visual quality of IMLSNet is better.

Table 2, the perception quality comparison measured by Light Field Descriptor (LFD) [9] on five categories further confirms the superiority of our approach.

Our MLS point sets (IMLSNet points) are also good discrete approximations and perform better than O-CNN-C whose results contain small missing regions and floating points (see the examples in the first and last rows). The small missing region in IMLSNet points can be recovered by the final IMLS surface. More visual comparisons can be found in Appendix E.

The statistics and illustrations on the unseen shape classes (Fig. 4) also confirm the generalization ability of all the approaches that use local features. Among them, our method achieves the best result.

4.2. Ablation study

A series of experiments on 3D object reconstruction from point clouds were conducted to evaluate each component and parameter choice of our network. The quantitative comparison on the test dataset is reported in Table 3.

Input and output depth By changing the depth number of the input and output octrees, our experiments reveal that: (1) the network with the deeper input octree like IMLSNet(7,6,1) yields more accurate results than IMLSNet(6,6,1) as more input point information are fed to the network; (2) the network with the deeper output octree like IMLSNet(7,7,1), also yields better-detailed geometry than the networks with shallow output octrees, like IMLSNet(7,6,1) and IMLSNet(7,5,1). These experiments show that the octree input and the octree-based scaffold are helpful

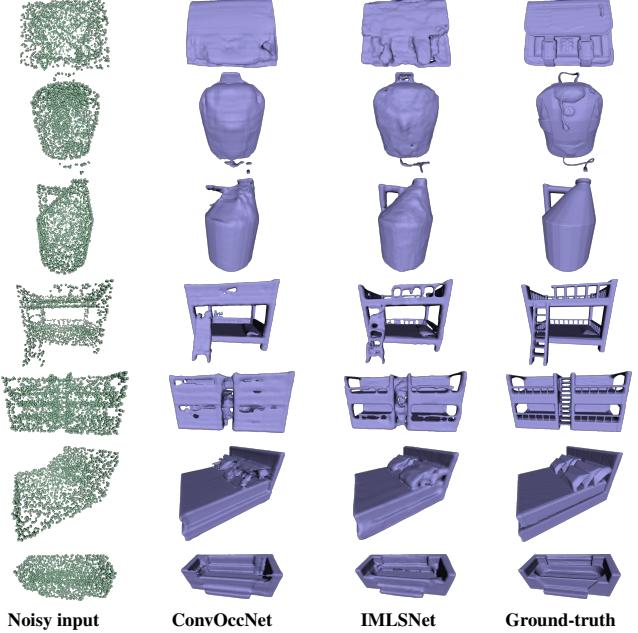


Figure 4: Visual results of 3D object reconstruction on five unseen shape classes. From left to right: the input point, ConvOccNet, our IMLSNet, and the ground-truth model.

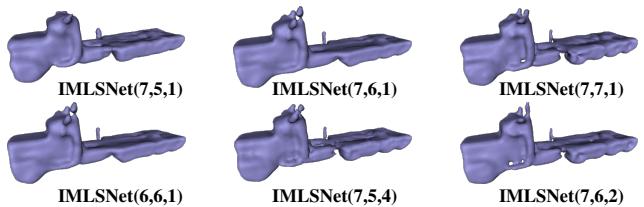


Figure 5: Ablation study of IMLSNets under different octree depths and different MLS point numbers in an octant.

to gather more point information and predict shape geometry. Fig. 5 illustrates their visual differences on the reconstruction results of a noisy truck model.

Number of MLS points We also found that using shallow output octree with a larger p , like IMLSNet(7,6,2) and IMLSNet(7,5,4), can achieve similar results by a deeper output octree with a smaller p , like IMLSNet(7,7,1), as the total number of MLS points of these networks are comparable. However, we observe IMLSNet(7,7,1) is better at capturing small and thin details (see Fig. 5).

Ablation study of loss terms By dropping the surface gradient term, the repulsion term, the projection smoothness term, and the radius smoothness term independently from IMLSNet(6,6,1), our experiments show that the surface quality degrades accordingly (see Table 3 and Fig. 6). The tested

| Network | $CD_1 \downarrow$ | $NC \uparrow$ | $IoU \uparrow$ | $F\text{-Score} \uparrow$ |
|------------------|-------------------|---------------|----------------|---------------------------|
| IMLSNet(7,5,1) | 0.038 | 0.928 | 0.888 | 0.962 |
| IMLSNet(7,6,1) | 0.031 | 0.943 | 0.913 | 0.981 |
| IMLSNet(6,6,1) | 0.033 | 0.941 | 0.904 | 0.977 |
| IMLSNet(7,7,1) | 0.031 | 0.944 | 0.913 | 0.983 |
| IMLSNet(7,5,4) | 0.032 | 0.936 | 0.905 | 0.977 |
| IMLSNet(7,6,2) | 0.030 | 0.944 | 0.915 | 0.983 |
| w/o grad.(6,6,1) | 0.034 | 0.938 | 0.897 | 0.974 |
| w/o rep.(6,6,1) | 0.034 | 0.939 | 0.899 | 0.976 |
| w/o ps.(6,6,1) | 0.034 | 0.939 | 0.900 | 0.974 |
| w/o rs.(6,6,1) | 0.033 | 0.940 | 0.902 | 0.977 |
| cr.(7,7,1) | 0.031 | 0.944 | 0.913 | 0.984 |
| cr.(6,6,1) | 0.036 | 0.933 | 0.880 | 0.966 |

Table 3: Quantitative evaluation of IMLSNets under different configurations.

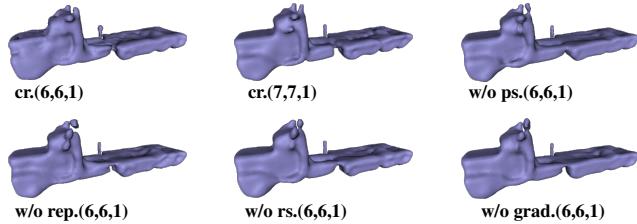


Figure 6: Visual results of IMLSNets under different configurations for a truck model.

networks are denoted by $w/o\ grad.(6,6,1)$, $w/o\ rep.(6,6,1)$, $w/o\ ps.(6,6,1)$, $w/o\ rs.(6,6,1)$. The gradient term plays a more important role than other terms, as the network trained without this term has the lower IoU and NC, also loses some surface details. The radii smoothness term is less important and the performance of the corresponding network slightly drops compared to IMLSNet(6,6,1). More visual comparisons are provided in Appendix E.

Constant MLS radius We fixed the MLS point radius to l_r and trained two networks: IMLSNet(6,6,1) and IMLSNet(7,7,1), denoted by $cr.(6,6,1)$ and $cr.(7,7,1)$. The results (see Table 3 and Fig. 5) show that the former network yields worse results than the default IMLSNet(6,6,1), but the latter network performs well. This phenomenon indicates that using the adaptive MLS radii would be more effective for IMLSNets with coarser scaffolds.

Octree-aided deep local implicit function Instead of predicting MLS points at each non-empty octant, one can follow [38] to predict the local implicit function value from the interpolated feature vector in the octant via MLP directly. We call this approach *Octree-aided deep local implicit function*. We implemented this approach and compared it with our IMLSNet. Our experiments (using depth-6 octrees) show



Figure 7: The results of octree-aided deep local implicit function (middle) and our IMLS surface results (right) from the same noisy inputs (left).

that this approach can recover most surface regions with good quality, but the results may contain some holes as there is no guarantee that the zero-iso surface always passes through the finest non-empty octant region, while our IMLSNet can do not have this flaw due to the explicit point generation and IMLS surface formulation. Fig. 7 illustrates this phenomenon on a plane model and a lamp model. More illustrations and discussions are presented in Appendix D.

5. Conclusion and future work

We present a deep implicit moving least-squares surface technique for 3D reconstruction, which enjoys the flexibility of point sets and the approximation power of implicit surface representations. Its efficacy and generalization ability are well demonstrated through extensive tests.

Beyond planes A straightforward extension for enhancing reconstruction quality is to use higher-order IMLS like APSS [18, 19] which replace planes with spheres, and other variants of MLS surfaces.

Differentiable rendering As the band regions of MLS surfaces can be determined by MLS points, rendering MLS surfaces by sphere marching algorithms can be performed more efficiently than other non-point-based implicit representations. Combining IMLS-Net and implicit surface ray-tracing will enable differentiable MLS prediction with 2D image supervision only.

MLS point generation As less MLS points can well approximate relatively planar and simple surface regions, devising an adaptive and non-uniform MLS point generation scheme would help reduce unnecessary MLS points and increase IMLS reconstruction quality further. Using and predicting the adaptive octrees [51] as the scaffold could be a promising solution.

References

- [1] Panos Achlioptas, Olga Diamanti, Ioannis Mitliagkas, and Leonidas Guibas. Learning representations and generative models for 3D point clouds. In *ICLR*, 2018. [2](#)
- [2] Marc Alexa, Johannes Behr, Daniel Cohen-Or, Shachar Fleishman, David Levin, and Claudio T. Silva. Point set surfaces. In *IEEE VIS*, page 21–28, 2001. [1, 2](#)
- [3] Matan Atzmon and Yaron Lipman. SAL: Sign agnostic learning of shapes from raw data. In *CVPR*, 2020. [2](#)
- [4] Matthew Berger, Andrea Tagliasacchi, Lee M. Seversky, Pierre Alliez, Gaël Guennebaud, Joshua A. Levine, Andrei Sharf, and Claudio T. Silva. A survey of surface reconstruction from point clouds. *Comput. Graph. Forum*, 36(1):301–329, 2017. [2](#)
- [5] Andrew Brock, Theodore Lim, J.M. Ritchie, and Nick Weston. Generative and discriminative voxel modeling with convolutional neural networks. In *3D deep learning workshop (NIPS)*, 2016. [2](#)
- [6] J. C. Carr, R. K. Beatson, J. B. Cherrie, T. J. Mitchell, W. R. Fright, B. C. McCallum, and T. R. Evans. Reconstruction and representation of 3D objects with radial basis functions. In *SIGGRAPH*, page 67–76, 2001. [2](#)
- [7] Rohan Chabra, Jan Eric Lenssen, Eddy Ilg, Tanner Schmidt, Julian Straub, Steven Lovegrove, and Richard Newcombe. Deep local shapes: Learning local SDF priors for detailed 3D reconstruction. In *ECCV*, 2020. [2](#)
- [8] Angel X. Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, Jianxiong Xiao, Li Yi, and Fisher Yu. ShapeNet: An information-rich 3D model repository. arXiv:1512.03012 [cs.GR], 2015. [4](#)
- [9] Zhiqin Chen and Hao Zhang. Learning implicit fields for generative shape modeling. In *CVPR*, 2019. [1, 2, 7, 11](#)
- [10] Z.-Q. Cheng, Y.-Z. Wang, B. Li, K. Xu, G. Dang, and S.-Y. Jin. A survey of methods for moving least squares surfaces. In *IEEE/EG Symposium on Volume and Point-Based Graphics*, 2008. [1, 2](#)
- [11] Julian Chibane, Thiemo Alldieck, and Gerard Pons-Moll. Implicit functions in feature space for 3D shape reconstruction and completion. In *CVPR*, 2020. [2](#)
- [12] Christopher Choy, JunYoung Gwak, and Silvio Savarese. 4D spatio-temporal ConvNets: Minkowski convolutional neural networks. In *CVPR*, 2019. [2](#)
- [13] Christopher B Choy, Danfei Xu, JunYoung Gwak, Kevin Chen, and Silvio Savarese. 3D-R2N2: A unified approach for single and multi-view 3D object reconstruction. In *ECCV*, 2016. [2](#)
- [14] Theo Deprelle, Thibault Groueix, Matthew Fisher, Vladimir G. Kim, Bryan C. Russell, and Mathieu Aubry. Learning elementary structures for 3D shape generation and matching. In *NeurIPS*, 2019. [2](#)
- [15] Yueqi Duan, Haidong Zhu, He Wang, Li Yi, Ram Nevatia, and Leonidas J. Guibas. Curriculum DeepSDF. In *ECCV*, 2020. [5](#)
- [16] Georgia Gkioxari, Jitendra Malik, and Justin Johnson. Mesh R-CNN. In *ICCV*, 2019. [2](#)
- [17] Thibault Groueix, Matthew Fisher, Vladimir G. Kim, Bryan C. Russell, and Mathieu Aubry. AtlasNet: A Papier-Mâché approach to learning 3D surface generation. In *CVPR*, 2018. [2](#)
- [18] Gaël Guennebaud, Marcel Germann, and Markus Gross. Dynamic sampling and rendering of algebraic point set surfaces. *Comput. Graph. Forum (EG)*, 27(3):653–662, 2008. [8](#)
- [19] Gaël Guennebaud and Markus Gross. Algebraic point set surfaces. *ACM Trans. Graph.*, 26(3):23:1–23:9, 2007. [2, 8](#)
- [20] Yulan Guo, Hanyun Wang, Qingyong Hu, Hao Liu, Li Liu, and Mohammed Bennamoun. Deep learning for 3D point clouds: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2020. [1](#)
- [21] Christian Häne, Shubham Tulsiani, and Jitendra Malik. Hierarchical surface prediction for 3D object reconstruction. In *Proc. Int. Conf. on 3D Vision (3DV)*, 2017. [2](#)
- [22] Hui Huang, Dan Li, Hao Zhang, Uri Ascher, and Daniel Cohen-Or. Consolidation of unorganized point clouds for surface reconstruction. *ACM Trans. Graph.*, 28(5):176:1–176:7, 2009. [4](#)
- [23] Chiyu Max Jiang, Avneesh Sud, Ameesh Makadia, Jingwei Huang, Matthias Nießner, and Thomas Funkhouser. Local implicit grid representations for 3D scenes . In *CVPR*, 2020. [2](#)
- [24] Hiroharu Kato, Yoshitaka Ushiku, and Tatsuya Harada. Neural 3D mesh renderer. In *CVPR*, 2018. [2](#)
- [25] Michael Kazhdan, Matthew Bolitho, and Hugues Hoppe. Poisson surface reconstruction. In *Symp. Geom. Proc.*, pages 61–70, 2006. [2](#)
- [26] Michael Kazhdan and Hugues Hoppe. Screened Poisson surface reconstruction. *ACM Trans. Graph.*, 32(3):29:1–29:13, 2013. [2](#)
- [27] Ravikrishna Kolluri. Provably good moving least squares. *ACM Trans. Algorithms*, 4(2):18:1–18:25, 2008. [1, 2, 3](#)
- [28] David Levin. The approximation power of moving least-squares. *Mathematics Computing*, 67(224):1517—1523, 1998. [2](#)
- [29] Jun Li, Kai Xu, Siddhartha Chaudhuri, Ersin Yumer, Hao Zhang, and Leonidas Guibas. GRASS: Generative recursive autoencoders for shape structures. *ACM Trans. Graph.*, 36(4):52:1–52:14, 2017. [2](#)
- [30] Chen Liu, Jimei Yang, Duygu Ceylan, Ersin Yumer, and Yasutaka Furukawa. PlaneNet: Piece-wise planar reconstruction from a single RGB image. In *CVPR*, 2018. [2](#)
- [31] William E. Lorensen and Harvey E. Cline. Marching cubes: A high resolution 3d surface construction algorithm. In *SIGGRAPH*, page 163–169, 1987. [3](#)
- [32] Lars Mescheder, Michael Oechsle, Michael Niemeyer, Sebastian Nowozin, and Andreas Geiger. Occupancy networks: Learning 3D reconstruction in function space. In *CVPR*, 2019. [1, 2, 4, 5](#)
- [33] Chengjie Niu, Jun Li, and Kai Xu. Im2Struct: Recovering 3D shape structure from a single RGB image. In *CVPR*, 2018. [2](#)
- [34] Yutaka Ohtake, Alexander Belyaev, Marc Alexa, Greg Turk, and Hans-Peter Seidel. Multi-level partition of unity implicits. *ACM Trans. Graph.*, 22(3):463–470, 2003. [2](#)

- [35] Cengiz Oztireli, Gaël Guennebaud, and Markus Gross. Feature preserving point set surfaces based on non-linear kernel regression. *Comput. Graph. Forum (EG)*, 28(2):493–501, 2009. 5
- [36] Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. DeepSDF: learning continuous signed distance functions for shape representation. In *CVPR*, 2019. 1, 2, 5, 11
- [37] Mark Pauly, Markus Gross, and Leif P. Kobbelt. Efficient simplification of point-sampled surfaces. In *IEEE VIS*, pages 163–170, 2002. 2
- [38] Songyou Peng, Michael Niemeyer, Lars Mescheder, Marc Pollefeys, and Andreas Geiger. Convolutional occupancy networks. In *ECCV*, 2020. 2, 4, 5, 8, 11
- [39] Charles R. Qi, Hao Su, Kaichun Mo, and Leonidas J. Guibas. PointNet: Deep learning on point sets for 3D classification and segmentation. In *CVPR*, 2017. 1
- [40] Charles R Qi, Li Yi, Hao Su, and Leonidas J Guibas. PointNet++: Deep hierarchical feature learning on point sets in a metric space. In *NeurIPS*, 2017. 1
- [41] Gernot Riegler, Ali Osman Ulusoy, and Andreas Geiger. OctNet: Learning deep 3D representations at high resolutions. In *CVPR*, 2017. 2
- [42] Hao Su, Haoqiang Fan, and Leonidas Guibas. A point set generation network for 3D object reconstruction from a single image. In *CVPR*, 2017. 2, 3
- [43] Chun-Yu Sun, Qian-Fang Zou, Xin Tong, and Yang Liu. Learning Adaptive hierarchical cuboid abstractions of 3D shape collections. *ACM Trans. Graph.*, 38(6):241:1–241:13, 2019. 2
- [44] Jiapeng Tang, Xiaoguang Han, Junyi Pan, Kui Jia, and Xin Tong. A skeleton-bridged deep learning approach for generating meshes of complex topologies from single RGB images. In *CVPR*, 2019. 2
- [45] M. Tatarchenko, A. Dosovitskiy, and T. Brox. Octree generating networks: efficient convolutional architectures for high-resolution 3D outputs. In *ICCV*, 2017. 2
- [46] Maxim Tatarchenko, Stephan R. Richter, René Ranftl, Zhiwen Li, Vladlen Koltun, and Thomas Brox. What do single-view 3D reconstruction networks learn? In *CVPR*, 2019. 5
- [47] Shubham Tulsiani, , Hao Su, Leonidas J. Guibas, Alexei A. Efros, and Jitendra Malik. Learning shape abstractions by assembling volumetric primitives. In *CVPR*, 2017. 2
- [48] Nanyang Wang, Yinda Zhang, Zhuwen Li, Yanwei Fu, Wei Liu, and Yu-Gang Jiang. Pixel2Mesh: Generating 3D mesh models from single RGB images. In *CVPR*, 2018. 2
- [49] Peng-Shuai Wang, Yang Liu, Yu-Xiao Guo, Chun-Yu Sun, and Xin Tong. O-CNN: Octree-based convolutional neural networks for 3D shape analysis. *ACM Trans. Graph.*, 36(4):72:1–72:11, 2017. 2, 3, 4, 5
- [50] Peng-Shuai Wang, Yang Liu, and Xin Tong. Deep Octree-based CNNs with output-guided skip connections for 3D shape and scene completion. In *CVPR workshop*, 2020. 3, 5
- [51] Peng-Shuai Wang, Chun-Yu Sun, Yang Liu, and Xin Tong. Adaptive O-CNN: A patch-based deep representation of 3D shapes. *ACM Trans. Graph.*, 2018. 2, 3, 8
- [52] Francis Williams, Teseo Schneider, Claudio Silva, Denis Zorin, Joan Bruna, and Daniele Panozzo. Deep geometric prior for surface reconstruction. In *CVPR*, 2019. 2
- [53] Jiajun Wu, Chengkai Zhang, Tianfan Xue, William T. Freeman, and Joshua B. Tenenbaum. Learning a probabilistic latent space of object shapes via 3D generative-adversarial modeling. In *NeurIPS*, 2016. 2
- [54] Guandao Yang, Xun Huang, Zekun Hao, Ming-Yu Liu, Serge Belongie, and Bharath Hariharan. PointFlow: 3D point cloud generation with continuous normalizing flows. In *ICCV*, 2019. 2
- [55] Yaoqing Yang, Chen Feng, Yiru Shen, and Dong Tian. FoldingNet: Point cloud auto-encoder via deep grid deformation. In *CVPR*, 2018. 2
- [56] Wang Yifan, Felice Serena, Shihao Wu, Cengiz Öztireli, and Olga Sorkine-Hornung. Differentiable surface splatting for point-based geometry processing. *ACM Trans. Graph.*, 38(6):230:1–230:14, 2019. 2
- [57] Wang Yifan, Shihao Wu, Hui Huang, Daniel Cohen-Or, and Olga Sorkine-Hornung. Patch-based progressive 3D point set upsampling. In *CVPR*, 2019. 2
- [58] Lequan Yu, Xianzhi Li, Chi-Wing Fu, Daniel Cohen-Or, and Pheng-Ann Heng. PU-Net: point cloud upsampling network. In *CVPR*, 2018. 2
- [59] Chuhang Zou, Ersin Yumer, Jimei Yang, Duygu Ceylan, and Derek Hoiem. 3D-PRNN: Generating shape primitives with recurrent neural networks. In *ICCV*, 2017. 2

A. Training data preparation

For training data involved with the ShapeNet dataset, we use data preprocessing tools from [38] to generate watertight meshes via TSDF fusion. We then normalize each mesh into a $[-1, 1]^3$ bounding box with 5% padding and compute signed distance function (SDF) values and gradients using the OpenVDB library (<https://www.openvdb.org>). We generate $256 \times 256 \times 256$ SDF grids, denoted by $\mathcal{F} = \{(i, j, k, s_{i,j,k}, \nabla s_{i,j,k})\}$, and collect SDF samples subset in a progressive manner: we first gather depth-6 SDF samples ((*i.e.*, samples whose indices satisfy: $i, j, k \bmod 4 = 0$) with absolute SDF values less than $\frac{1}{8}$, this threshold guarantees coverage of generated octree nodes. To better capture shape details, similar to the sampling strategy in [36], we add more SDF samples near the surface, to be concrete, the depth-7 SDF samples with absolute SDF values less than $\frac{1}{16}$.

B. Evaluation metrics

We reuse the evaluation tools of [38] to compute the following metrics. We denote M_g and M_p as the ground-truth mesh and the mesh of the predicted result. $\mathcal{X} := \{\mathbf{x}_1, \dots, \mathbf{x}_{N_g}\}$ and $\mathcal{Y} := \{\mathbf{y}_1, \dots, \mathbf{y}_{N_p}\}$ are randomly sample points on these two meshes, respectively. We define $\mathcal{P}_{g2p}(\mathbf{x}) = \arg \min_{\mathbf{y} \in \mathcal{Y}} \|\mathbf{x} - \mathbf{y}\|$ and $\mathcal{P}_{p2g}(\mathbf{y}) = \arg \min_{\mathbf{x} \in \mathcal{X}} \|\mathbf{x} - \mathbf{y}\|$. $\mathbf{n}(\cdot)$ denote an operator that returns the normal vector of a given point.

- L_1 Chamfer distance.

$$\text{CD}_1 = \frac{1}{2N_g} \sum_{i=1}^{N_g} \|\mathbf{x}_i - \mathcal{P}_{g2p}(\mathbf{x}_i)\| + \frac{1}{2N_p} \sum_{i=1}^{N_p} \|\mathbf{y}_i - \mathcal{P}_{p2g}(\mathbf{y}_i)\|.$$

- Normal consistency.

$$\text{NC} = \frac{1}{2N_g} \sum_{i=1}^{N_g} |\mathbf{n}(\mathbf{x}_i) \cdot \mathbf{n}(\mathcal{P}_{g2p}(\mathbf{x}_i))| + \frac{1}{2N_p} \sum_{i=1}^{N_p} |\mathbf{n}(\mathbf{y}_i) \cdot \mathbf{n}(\mathcal{P}_{p2g}(\mathbf{y}_i))|.$$

- IOU is the volumetric intersection of two meshes divided by the volume of their union. To compute this metric, 100k points are sampled in the bounding box and are determined whether they are in or outside two meshes.
- F-Score is the harmonic mean between Precision and Recall. Precision is the percentage of points on M_p that

| Network | 2 | 3 | 4 | 5 | 6 | 7 | size |
|-----------------|-----|-----|-----|----|----|----|------|
| IMLSNet(7,6,1) | 256 | 256 | 128 | 64 | 32 | 16 | 4.6M |
| IMLSNet(7,7,1) | 256 | 256 | 128 | 64 | 32 | 16 | 4.6M |
| IMLSNet(7,6,1)* | 128 | 128 | 128 | 64 | 32 | 16 | 1.5M |
| IMLSNet(7,7,1)* | 128 | 128 | 128 | 64 | 32 | 16 | 1.6M |

Table 4: Network parameters of IMLSNets. Feature channel dimensions on each octree depth (from 2 to 7) are listed.

| Network | $\text{CD}_1 \downarrow$ | $\text{NC} \uparrow$ | $\text{IoU} \uparrow$ | $\text{F-Score} \uparrow$ |
|-----------------|--------------------------|----------------------|-----------------------|---------------------------|
| IMLSNet(7,6,1) | 0.0310 | 0.9430 | 0.9134 | 0.9813 |
| IMLSNet(7,7,1) | 0.0306 | 0.9440 | 0.9135 | 0.9833 |
| IMLSNet(7,6,1)* | 0.0311 | 0.9425 | 0.9129 | 0.9814 |
| IMLSNet(7,7,1)* | 0.0307 | 0.9434 | 0.9132 | 0.9827 |

Table 5: Quantitative evaluation of IMLSNet with different network settings on the task of 3D object reconstruction.

lie within distance τ to M_g , Recall is the percentage of points on M_g that lie within distance τ to M_p .

$$\text{F-Score} = \frac{2 * \text{precision} * \text{Recall}}{\text{precision} + \text{Recall}}.$$

We also compute the light field descriptor(LFD) to evaluate the perceptual similarity of the results to the ground-truth by following the setup of [9]. LFD is computed in the following way: each generated shape is rendered from various views and results in a set of projected images, then each projected image is encoded using Zernike moments and Fourier descriptors.

C. Network architecture

The number of network parameters for our network IMLSNet(7,7,1) reported in the paper is 4.6 M. The detailed setup of IMLSNet(7,6,1) and IMLSNet(7,7,1) is illustrated in the first and second rows of Table 4. We also did an ablation study by setting the channel number of depth-2 and depth-3 octree nodes to 128 and reduced the network parameter size while achieving comparable performances as shown in Table 5. The networks are denoted by IMLSNet(6,6,1)* and IMLSNet(7,7,1)*.

D. Octree-aided deep local implicit function

As discussed in Section 4, our ablation study shows that the octree-aided deep local implicit function has issues in obtaining complete zero-iso surfaces. We confirm this fact by examining the generated non-empty octants which cover the missing regions but the local implicit function does not generate the iso-surface in them. Fig. 8-a illustrates these octants which cover the ground-truth non-empty octants (see Fig. 8-c) but fail to generate a complete shape (Fig. 8-b). We speculate that the implicit function passes through the

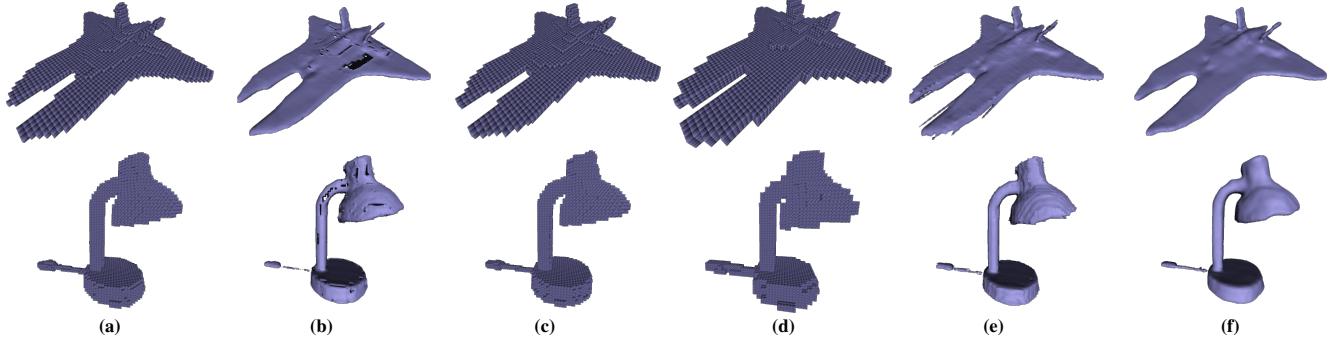


Figure 8: (a): Non-empty finest octants predicated by the network based on the octree-aided deep local implicit function. (b): Reconstruction results from (a). (c): Ground-truth non-empty finest octants. (d): The expanded octree. (e): Reconstruction results based on (d). (f) Our IMLSNet results.

neighboring region which is not covered by the current finest octants. Based on this speculation, we split all the $d - 1$ -depth octants to expand the octree (see Fig. 8-d) and extract the surface via marching cubes. It turns out that the implicit surface appears in those regions (see Fig. 8-e), however, the reconstruction error is higher than our IMLSNet (Fig. 8-f).

E. Evaluation of object reconstruction

In Table 7, we report the numerical metrics of the tasks of object reconstruction from point clouds for each shape category. Fig. 9 presents more visual results reconstructed from our network. All the evaluations demonstrate the superiority of our method over other approaches in terms of reconstruction accuracy and the capacity of recovering details and thin regions. In Fig. 10 we present more results of our ablation study of different network settings.

F. Robustness test on noise levels

We did a robustness test on the input noise. The network IMLSNet(7,7,1) and ConvOccNet were trained with noisy data whose Gaussian noise is with standard deviation $\delta = 5 \times 10^{-3}$. We add different noise levels ($\delta = 1 \times 10^{-3}, 3 \times 10^{-3}, 7.5 \times 10^{-3}$) to the test data of 13 shape classes and feed to our network and ConvOccNet for evaluating their performance. From Table 6, we can see with lower noise levels, our network always performs better than ConvOccNet. With a higher level noise ($\delta = 7.5 \times 10^{-3}$), The network performance of both methods degrades gracefully, and our method still outperforms ConvOccNet.

| Network | δ | $CD_1 \downarrow$ | $NC \uparrow$ | $IoU \uparrow$ | $F\text{-Score} \uparrow$ |
|------------|----------------------|-------------------|---------------|----------------|---------------------------|
| IMLSNet | 1.0×10^{-3} | 0.0288 | 0.9476 | 0.9226 | 0.9859 |
| ConvOccNet | 1.0×10^{-3} | 0.0495 | 0.9349 | 0.8573 | 0.9442 |
| IMLSNet | 3.0×10^{-3} | 0.0290 | 0.9473 | 0.9219 | 0.9857 |
| ConvOccNet | 3.0×10^{-3} | 0.0439 | 0.9377 | 0.8831 | 0.9461 |
| IMLSNet | 5.0×10^{-3} | 0.0306 | 0.9440 | 0.9135 | 0.9833 |
| ConvOccNet | 5.0×10^{-3} | 0.0441 | 0.9383 | 0.8842 | 0.9421 |
| IMLSNet | 7.5×10^{-3} | 0.0372 | 0.9291 | 0.8754 | 0.9705 |
| ConvOccNet | 7.5×10^{-3} | 0.0536 | 0.9345 | 0.8435 | 0.9221 |

Table 6: Robustness test to noise.

| Category | CD ₁ ↓ | | | | NC ↑ | | | |
|-----------|-------------------|----------------|------------|---------------|-----------|----------------|---------------|---------------|
| | O-CNN-C | IMLSNet points | ConvOccNet | IMLSNet | O-CNN-C | IMLSNet points | ConvOccNet | IMLSNet |
| airplane | 0.0634 | 0.0316 | 0.0336 | 0.0245 | 0.9181 | 0.9292 | 0.9311 | 0.9371 |
| bench | 0.0646 | 0.0356 | 0.0352 | 0.0301 | 0.9136 | 0.9194 | 0.9205 | 0.9220 |
| cabinet | 0.0709 | 0.0375 | 0.0461 | 0.0348 | 0.9411 | 0.9505 | 0.9561 | 0.9546 |
| car | 0.0765 | 0.0419 | 0.0750 | 0.0395 | 0.8668 | 0.8709 | 0.8931 | 0.8820 |
| chair | 0.0664 | 0.0383 | 0.0459 | 0.0348 | 0.9407 | 0.9487 | 0.9427 | 0.9503 |
| display | 0.0655 | 0.0339 | 0.0368 | 0.0292 | 0.9598 | 0.9710 | 0.9677 | 0.9732 |
| lamp | 0.0667 | 0.0367 | 0.0595 | 0.0312 | 0.9111 | 0.9206 | 0.9003 | 0.9218 |
| speaker | 0.0729 | 0.0413 | 0.0632 | 0.0396 | 0.9363 | 0.9440 | 0.9387 | 0.9473 |
| rifle | 0.0617 | 0.0300 | 0.0280 | 0.0207 | 0.9320 | 0.9428 | 0.9293 | 0.9433 |
| sofa | 0.0657 | 0.0350 | 0.0414 | 0.0309 | 0.9492 | 0.9602 | 0.9579 | 0.9631 |
| table | 0.0663 | 0.0360 | 0.0385 | 0.0319 | 0.9461 | 0.9599 | 0.9588 | 0.9621 |
| telephone | 0.0610 | 0.0295 | 0.0270 | 0.0229 | 0.9737 | 0.9827 | 0.9823 | 0.9839 |
| vessel | 0.0641 | 0.0336 | 0.0430 | 0.0271 | 0.9221 | 0.9280 | 0.9187 | 0.9319 |
| mean | 0.0666 | 0.0355 | 0.0441 | 0.0306 | 0.9316 | 0.9406 | 0.9382 | 0.9440 |
| bag | 0.0704 | 0.0386 | 0.0538 | 0.0351 | 0.9342 | 0.9420 | 0.9417 | 0.9455 |
| bathtub | 0.0663 | 0.0378 | 0.0526 | 0.0350 | 0.9478 | 0.9599 | 0.9537 | 0.9622 |
| bed | 0.0720 | 0.0428 | 0.0608 | 0.0412 | 0.9192 | 0.9246 | 0.9119 | 0.9278 |
| bottle | 0.0619 | 0.0332 | 0.0421 | 0.0279 | 0.9610 | 0.9696 | 0.9657 | 0.9708 |
| pillow | 0.0631 | 0.0340 | 0.0548 | 0.0303 | 0.9652 | 0.9743 | 0.9660 | 0.9757 |
| mean | 0.0667 | 0.0373 | 0.0528 | 0.0339 | 0.9455 | 0.9541 | 0.9478 | 0.9564 |
| Category | IoU ↑ | | | | F-Score ↑ | | | |
| | O-CNN-C | IMLSNet points | ConvOccNet | IMLSNet | O-CNN-C | IMLSNet points | ConvOccNet | IMLSNet |
| airplane | n/a | n/a | 0.8485 | 0.8910 | 0.8101 | 0.9923 | 0.9653 | 0.9918 |
| bench | n/a | n/a | 0.8298 | 0.8480 | 0.7995 | 0.9867 | 0.9643 | 0.9860 |
| cabinet | n/a | n/a | 0.9398 | 0.9495 | 0.7887 | 0.9833 | 0.9558 | 0.9811 |
| car | n/a | n/a | 0.8858 | 0.9052 | 0.7474 | 0.9604 | 0.8490 | 0.9521 |
| chair | n/a | n/a | 0.8709 | 0.9033 | 0.7993 | 0.9824 | 0.9387 | 0.9815 |
| display | n/a | n/a | 0.9275 | 0.9491 | 0.8109 | 0.9935 | 0.9708 | 0.9935 |
| lamp | n/a | n/a | 0.7840 | 0.8583 | 0.7999 | 0.9806 | 0.8910 | 0.9785 |
| speaker | n/a | n/a | 0.9188 | 0.9450 | 0.7789 | 0.9676 | 0.8924 | 0.9633 |
| rifle | n/a | n/a | 0.8459 | 0.8886 | 0.8263 | 0.9961 | 0.9799 | 0.9962 |
| sofa | n/a | n/a | 0.9362 | 0.9541 | 0.8052 | 0.9886 | 0.9531 | 0.9873 |
| table | n/a | n/a | 0.8877 | 0.9076 | 0.8013 | 0.9875 | 0.9674 | 0.9870 |
| telephone | n/a | n/a | 0.9537 | 0.9647 | 0.8252 | 0.9978 | 0.9882 | 0.9978 |
| vessel | n/a | n/a | 0.8663 | 0.9140 | 0.8085 | 0.9880 | 0.9313 | 0.9868 |
| mean | n/a | n/a | 0.8842 | 0.9135 | 0.8001 | 0.9850 | 0.9421 | 0.9833 |
| bag | n/a | n/a | 0.9229 | 0.9461 | 0.7859 | 0.9766 | 0.9187 | 0.9745 |
| bathtub | n/a | n/a | 0.8431 | 0.9079 | 0.8014 | 0.9861 | 0.9084 | 0.9851 |
| bed | n/a | n/a | 0.8612 | 0.9052 | 0.7727 | 0.9661 | 0.8985 | 0.9622 |
| bottle | n/a | n/a | 0.9468 | 0.9663 | 0.8282 | 0.9916 | 0.9515 | 0.9907 |
| pillow | n/a | n/a | 0.9354 | 0.9700 | 0.8219 | 0.9948 | 0.9005 | 0.9941 |
| mean | n/a | n/a | 0.9019 | 0.9391 | 0.8020 | 0.9831 | 0.9155 | 0.9813 |

Table 7: Quantitative evaluation of different networks on the test data of 13 shape classes and the full data of 5 unseen shape classes.

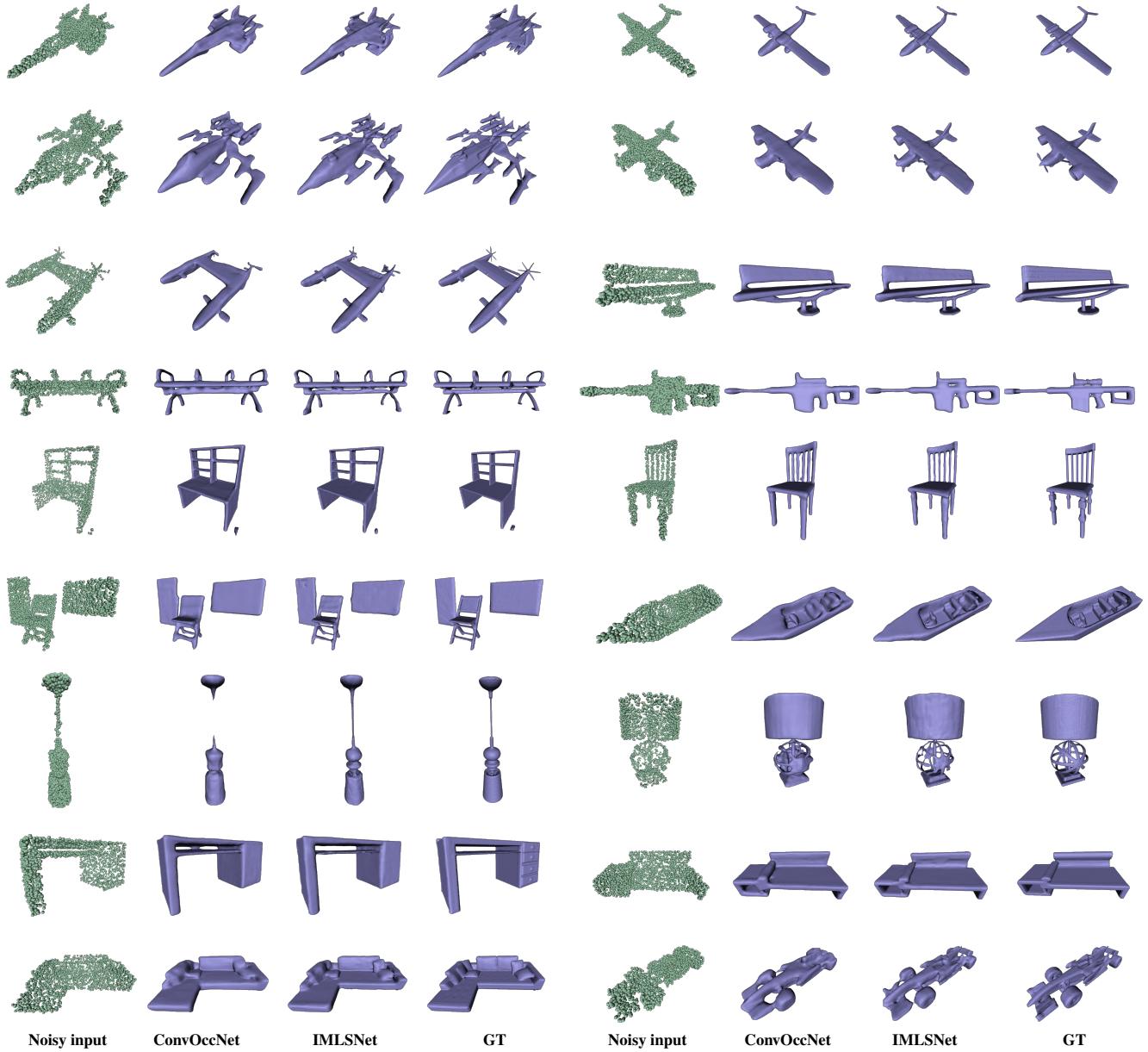


Figure 9: More results of object reconstruction from point clouds.

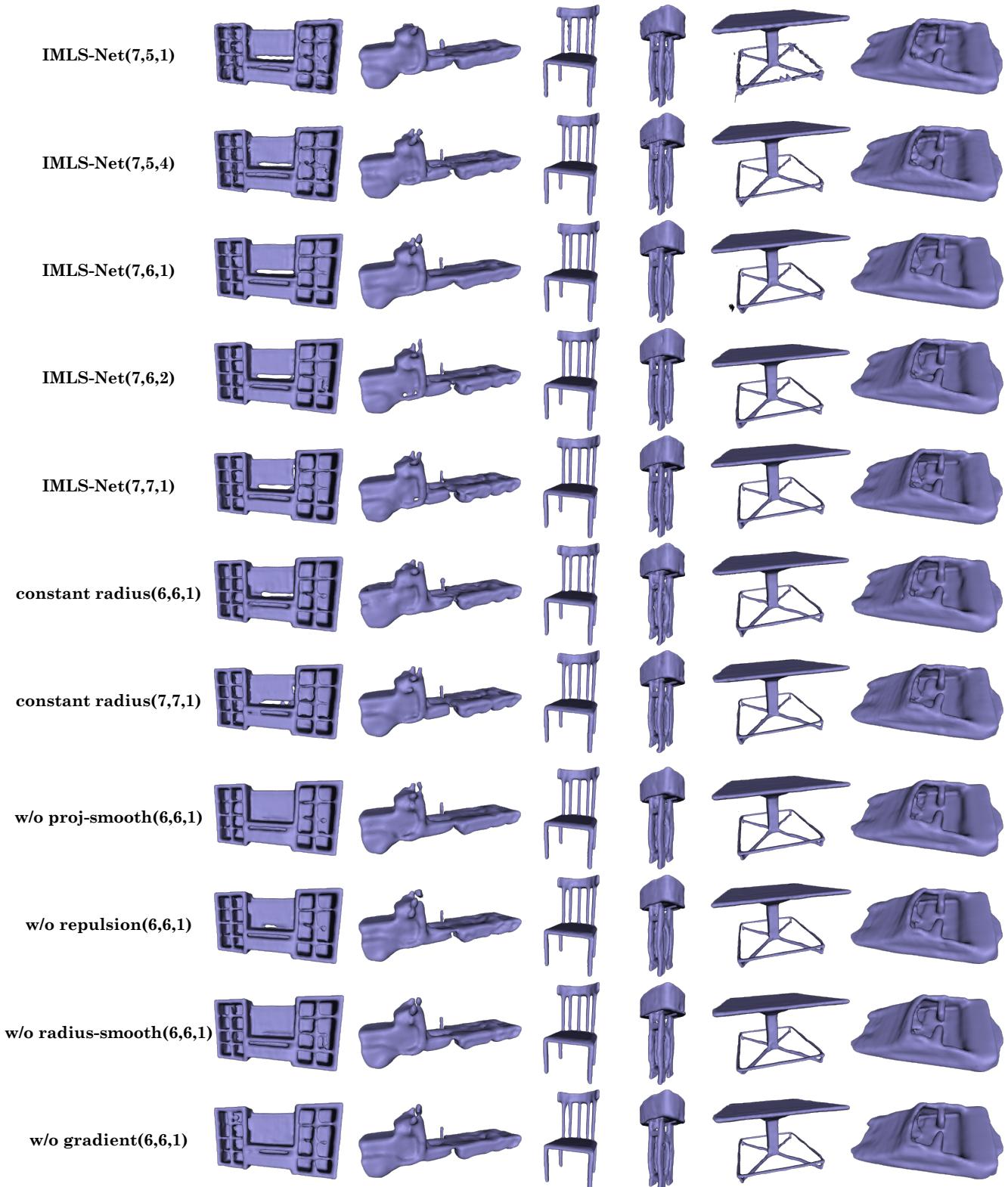


Figure 10: More results of our ablation study on network settings. The inputs are the noisy point clouds (see Fig. 3).