Sparse-to-Dense Depth Completion Revisited: Sampling Strategy and Graph Construction*

Xin Xiong[1], Haipeng Xiong[1], Ke Xian[1], Chen Zhao[1], Zhiguo Cao[1], and Xin Li[2]

[1] School of AIA, Huazhong University of Science and Technology, China
{xiong_xin,hpxiong,kexian,hust_zhao,zgcao}@hust.edu.cn
[2] Lane Department of CSEE, West Virginia University, USA.
xin.li@mail.wvu.edu

Abstract. Depth completion is a widely studied problem of predicting a dense depth map from a sparse set of measurements and a single RGB image. In this work, we approach this problem by addressing two issues that have been under-researched in the open literature: sampling strategy and graph construction. First, instead of the popular random sampling strategy, we suggest that Poisson disk sampling is a much more effective solution to create sparse depth map from a dense version. We experimentally compare a class of quasi-random sampling strategies and demonstrate that an optimized sampling strategy can significantly improve the performance of depth completion for the same number of sparse samples. Second, instead of the traditional square kernel, we suggest that the dynamic construction of local neighborhood is a better choice for interpolating the missing values. More specifically, we proposed an end-to-end network with a graph convolution module. Since the neighborhood relationship of 3D points is more effectively exploited by our novel graph convolution module, our approach has achieved not only state-of-the-art results for depth completion of indoor scenes but also better generalization ability than other competing methods.

Keywords: Depth Completion, Graph Neural Network, Poisson Disk Sampling, Sparse-to-Dense

1 Introduction

Depth sensing and measurement has become an essential tool supporting a wide range of engineering applications from robotics and autonomous vehicles to 3D vision and augmented reality. Despite decades of research, existing depth sensors still have various limitations. More specifically LiDAR-based (e.g., Velodyne)

* Xin Xiong and Haipeng Xiong contributed equally. Zhiguo Cao is the corresponding author. This work was supported in part by the National Key RD Program of China (No.2018YFB1305504) and the National Natural Science Foundation of China (Grant No. U1913602). Xin Li's work is partially supported by the DoJ/NIJ under grant NIJ 2018-75-CX-0032, NSF under grant OAC-1839909, IIS-1951504 and the WV Higher Education Policy Commission Grant (HEPC.dsr.18.5).

sensors are expensive and yet only provide sparse measurements for objects at a distance. Structure-light-based sensors (e.g., Kinect) are sensitive to sunlight and suffer from a short ranging distance (i.e., only suitable for indoor scenes). Stereo-camera based approaches often require a large baseline and careful calibration; and their computational complexity is still demanding and performance around featureless regions is poor. Most recently, 3D from a single image [38], [15], [20],[62],[60],[61] has received increasingly more attention because it might lead to low-cost and energy-efficient solution to 3D/depth sensing.

In view of the limitations of existing depth sensors, it has been suggested in [41] that sparse-to-dense depth completion is a promising solution in practice. Sparse depth measurements are often readily available from low-cost LiDARs or computed from the output of Simultaneous Localization and Mapping (SLAM). The task of sparse-to-dense is to fill in the missing data and approximate the dense depth map as accurate as possible. Previous works [41], [63], [8], [17] have all assumed that sparse samples are acquired in a random fashion (i.e., observing a Bernoulli distribution spatially [41]). However, the optimality of random-sampling strategy is often questionable. In fact, the sampling locations given by practical LiDAR sensors are seldom randomly distributed, but distributed uniformly in the spatial domain. To the best of our knowledge, the issue of sampling strategy has not been studied for depth completion in the open literature.

The other important motivation behind this work is the consideration of neighborhood and the construction of network for sparse-to-dense depth completion. Almost all existing works [12], [18], [46], [43], [64], [41], [8], [7], [17] have adopted a standard spatially-invariant square kernel (refer to Fig. 3b). Attempts to work with spatially varying kernels such as guided-image filter [24] have been made in recent works (e.g., GuideNet [54]) but requires a special implementation of guided convolution module. To manage the prohibitive computational complexity, GuideNet [54] has to count on a factorization strategy to speed up the implementation. How to seamlessly integrate adaptive selection of local neighborhood with network design has remained an under-researched topic.

Based on the above observations, we conduct a systematic study of sampling strategy and neighborhood construction in this paper. On one hand, we propose to leverage the idea of Poisson disk sampling [2] and low-discrepancy sequence [45] to generate a class of deterministic yet quasi-random sampling patterns. When compared with randomly sampled patterns, patterns generated by Poisson disk sampling appear more spatially uniformly distributed (please refer to Fig. 1). On the other hand, inspired by the latest advances in graph neural networks (GNN) [31], [57], [66], [59], we propose a novel GNN-based implementation with the desirable spatially-varying kernels. More specifically, we first construct a k-Nearest-Neighbor(kNN) based neighborhood and compute pointwise features as the inputs for GNN; then the task of depth completion is done by a multi-layer perceptron(MLP) based propagation process on the constructed GNN. In summary, the main technical contributions of our work are as follows:

- For the first time, we demonstrate that optimizing the sampling strategy can significantly improve the performance of sparse-to-dense depth completion. We

have systematically compared a class of four competing quasirandom sampling strategies and found that Golden sequence based sampling represents the best approximation of ideal Poisson disk sampling.

- We propose to develop a GNN-based sparse-to-dense depth completion algorithm. Constructed on a kNN-based local neighborhood, our solution is capable of exploiting spatially varying kernel which elegantly fits the graph convolutional module of GNN. We have also developed a MLP-based propagation process for depth completion on the constructed GNN.
- On NYUDv2 benchmark dataset, our GNN-based method with Golden sampling outperforms previous state-of-the-art methods DeepLiDAR [48] and depth-normal constraints [63] by over 25% in terms of RMSE values. On Matterport3D test set, ours outperforms previous state-of-the-art [26] by over 20% in terms of RMSE values.

2 Related Work

Depth completion Traditional depth completion methods usually employ handcraft features or specific kernels to predict missing values[12], [18], [46], [64], [43]. While these algorithms might be suitable for a specific task or scene, their performance and generalization capability is often unsatisfactory. Recently, deep learning based methods have shown promising performance on depth completion task - e.g., sparsity-invariant CNN [55], multi-scale sparsity-invariant network[27], confidence propagation CNN [17], color-guided encoder-decoder network [41], self-supervised depth completion from sparse LiDAR data [40], and global/local information fusion [56]. In addition to color information, other methods utilize the surface normal or object boundary to facilitate the task of depth completion - e.g., [48], [65], [26]. Among them, [65] and [26] recover depth from coarse depth map with missing values taken by structured-light scanners in indoor scenes. There also exist other competing methods exploiting the correlation among sparse data in the depth map - e.g., [8] , [63], and [6].

Spatial sampling strategy Spatial sampling of an unknown function is a widely studied problem in spatial statistics [49] [28], computer graphics [44], [2], image processing [16], and remote sensing [14]. Unlike regular sampling on integral lattice, irregular sampling such as Poisson disk sampling [2] often demonstrates better coverage of the space in terms of uniformity (e.g., absent from cluster of points within a certain region). Despite the theoretical appeal, practical implementation of Poisson disk sampling is a nontrivial task especially when computational complexity is considered. To address the problem of large computational consumption, Monte Carlo methods [4] have been developed to generate quasi-random sequences and sampling patterns. The quasi-random sequences are also known as low-discrepancy sequences [45] which is constructed in a deterministic manner while ensuring that the whole sampling space is uniformly covered.

Graph Neural Network Graph Neural Networks (GNNs) are a class of neural networks for modeling graph-based data. Some GNN methods [3], [11] apply Convolution Neural Networks(CNNs) to a graph in the spectral domain characterized by the graph Laplacian. Other methods [21], [51] aim at recurrently

applying neural networks to every node of the graph. When GNNs contain an iterative process, they propagate the node states until reaching the equilibrium and produce an output for each node based on its state in the process. This idea was adopted and improved by[37], which used gated recurrent units [10] in the propagation step. The learning process of these methods can be achieved by the back-propagation through time (BPTT) algorithm [58].

3 Spatial Sampling Strategy

In this section, we study and compare different quasi-random sampling strategies based on deterministic 1D low-discrepancy sequences. To facilitate the objective comparison, we propose a minimum-radius based criterion to evaluate the effectiveness of different sampling strategies for the depth completion task.

3.1 Low-discrepancy Sequences and Quasi-random Sampling

The problem of generating deterministic 1D low-discrepancy sequences has been well studied in the literature [1], [32]. We opt to work on Van der Corput sequences, a series of low-discrepancy sequences defined on a unit interval [35]. Each Van Der Corput sequence is defined by a unique hyper-parameter b. Note that any natural number n can be represented using b as a radix:

$$n = \sum_{k=0}^{L-1} d_k(n) b^k \tag{1}$$

where $d_k(n)$ is the k-th coefficient and $0 < d_k(n) < b$ and $L-1$ is the number of dyadic expansion term. Then the n-th number in Van Der Corput reverse-radix sequence is given by:

$$g_b(n) = \sum_{k=0}^{L-1} d_k(n) b^{-k-1} \tag{2}$$

For example, the first few terms of the sequence for $b = 2$ and $b = 3$ are:

$$V_2 : \left\{ \frac{1}{2}, \frac{1}{4}, \frac{3}{4}, \frac{1}{8}, \frac{5}{8}, \frac{3}{8}, \frac{7}{8}, \frac{1}{16}, \frac{9}{16}, ... \right\}, V_3 : \left\{ \frac{1}{3}, \frac{2}{3}, \frac{1}{9}, \frac{4}{9}, \frac{7}{9}, \frac{2}{9}, \frac{5}{9}, \frac{8}{9}, \frac{1}{27}, ... \right\} \tag{3}$$

Halton Sampling: Since we want to sample a depth map, 1D low-discrepancy sequences need to be extended into 2D. A simple yet effective solution is to simply consider the Cartesian product of multiple 1D sequences. For instance, Halton sequence in 2D is constructed by using different 1D van de Corput sequences whose bases b's are co-prime. For instance, taking Halton(2,3) as an example, the coordinates of the k-th point in a Halton(2,3) sequence is:

$$Halton : \{(V_2[k], V_3[k])\}_{k=1,2,...} \tag{4}$$

It is also possible to construct low-discrepancy sequence which dose not require choosing any basis parameters. These methods are based on golden ratio

[13] and there have been many ways to generalize the golden ratio (e.g., [33]). We define the golden ratio ϕ_d by the unique positive root of $x^{d+1} = x + 1$ - i.e.,

$$\begin{aligned} d = 1, \quad & \phi_1 = 1.61803398874989484820458683436563... \\ d = 2, \quad & \phi_2 = 1.32471795724474602596090885447809... \\ d = 3, \quad & \phi_3 = 1.22074408460575947536168534910883... \\ & ... \end{aligned} \tag{5}$$

This special sequence of constants, ϕ_d is called Harmonious numbers [42]. These special values can be expressed very elegantly as:

$$\begin{aligned} \phi_1 &= \sqrt{1+\sqrt{1+\sqrt{1+\sqrt{1+\sqrt{1+...}}}}} \\ \phi_2 &= \sqrt[3]{1+\sqrt[3]{1+\sqrt[3]{1+\sqrt[3]{1+\sqrt[3]{1+...}}}}} \\ \phi_3 &= \sqrt[4]{1+\sqrt[4]{1+\sqrt[4]{1+\sqrt[4]{1+\sqrt[4]{1+...}}}}} \\ &... \end{aligned} \tag{6}$$

where ϕ_1 is the the canonical golden ratio and ϕ_2 is often called the plastic constant.

Golden Sampling: We denote the low-discrepancy sequences generated by golden ratio and plastic constant by Golden sequence and Plastic sequence respectively. Then Golden sampling is constructed by:

$$Golden : \{(\frac{k+m}{N+n}, \frac{k}{\phi_1} \ mod \ 1)\}_{k=1,2,...} \tag{7}$$

where N is the total number of points, m and n are two hyper parameters, and mod denotes the modulo operator. Considering the classical packing problem, the objective is to maximize the smallest neighboring distance among N disks. That is:

$$d_N = \min_{i \neq j} |x_i - x_j| \tag{8}$$

We choose the pair of $m = 6, n = 11$ producing a large d_N in our implementation of Golden sampling.

Plastic Sampling: Similar to Golden sampling, Plastic sampling is defined by:

$$Plastic : \{(\frac{k}{\phi_2} \ mod \ 1, \frac{k\phi_2}{\phi_2 + 1} \ mod \ 1)\}_{k=1,2,...} \tag{9}$$

R2 Sampling: Another variation is based on plastic number and the coordinates of the k-th point in R2 sampling is given by:

$$R2 : \{((0.5 + \frac{k+1}{\phi_2}) \ mod \ 1, (0.5 + \frac{\phi_2(k+1)}{\phi_2 + 1}) \ mod \ 1)\}_{k=1,2,...} \tag{10}$$

Fig. 1 shows the distribution of 5 different sampling strategy. It can be seen that ad-hoc random sampling exhibits clustering of points, which does not spread the sampling location uniformly in 2D. By contrast, quasi-random sampling more uniformly cover the whole 2D space.

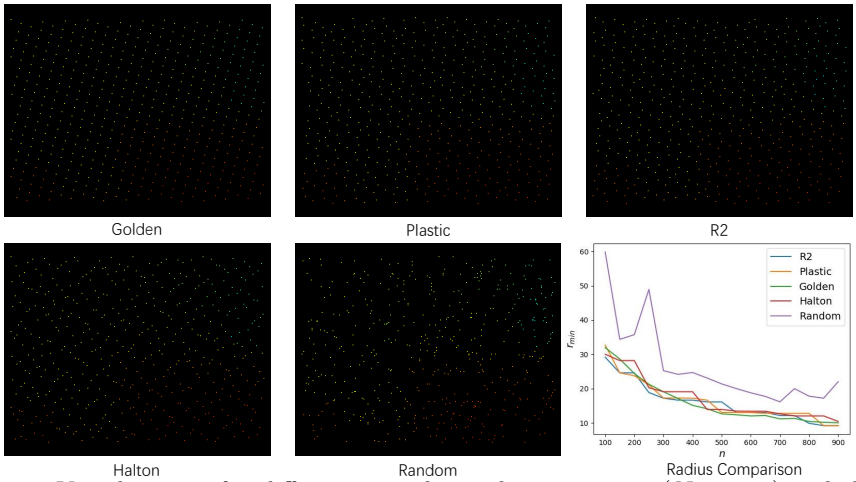

Fig. 1. Visualization of 5 different spatial sampling strategies ($N = 500$) and their performance comparison in terms of radius metric (r_{min}).

3.2 Quasi-random Sampling Pattern Comparison and Criterion

How do we compare different quasi-random sampling patterns? For low-discrepancy sequences, the definition of discrepancy has been articulated in in [45]. Let P denotes a point set consisted of $X_1, ...X_N \in I^d$, where $I^d = [0, 1]^d$ is the normalized integration domain (i.e., a closed d-dimensional unit cube). For an arbitrary subset $B \in I^d$, we can define

$$A(B; P) = \sum_{n=1}^{N} c_B(x_n) \tag{11}$$

where c_B is the characteristic function of B - i.e., $c_B = 1$ if $x_n \in B$; otherwise $c_B = 0$. Then $A(B; P)$ can be thought as a counting function indicating the number of points falling into B. If $\mathcal{B}$ is a nonempty family of Lebesgue-measurable subsets of I^d, the discrepancy of a point set P is given by:

$$D_N(\mathcal{B}; P) = \sup_{B \in \mathcal{B}} \left| \frac{A(B; P)}{N} - \lambda_d(B) \right| \in [0, 1] \tag{12}$$

where λ_d is a d-dimensional Lebesgue measure. However, such definition in the continuous space can not be directly used for discrete data such as depth maps.

Let D_s denotes the *sparse* depth map which requires the completion, and M_s the binary mask of D_s (1/0 denotes valid/missing depth value at this point respectively). For each missing point, we need to gather information from its neighbouring valid points. It follows that the maximum distance to the closest valid points(d^v) can somewhat reflect the overall difficulty of depth completion. The distance of a missing point p_m to its closest valid point is defined by:

$$d^v(p_m) = \min_{p_v} d(p_m, p_v) \tag{13}$$

where $d(p_m, p_v)$ denotes the Euclidean distance between point p_m and p_v. We can find the maximum distance d^v_{max} among all missing points in D_s as:

$$R_{max}(D_s) = \max_{p_m} d^v(p_m) \tag{14}$$

Note that such sphere packing bound d^v_{max} can be approached from an alternative perspective of sphere covering - i.e., the minimum radius of covering circles. For a sparse point set D_s, the smallest r needed to cover the whole image r_{min} can be calculated by:

$$\begin{aligned} r_{min}(D_s) = \min\ & r \\ s.t.\ & d^v(p_m) \leq r,\ \forall\ p_m \end{aligned} \tag{15}$$

For a fixed number of valid points, we conjecture the arrangement of their positions to minimize the covering radius $r_{min}(D_s)$ is equivalent to the dual problem of maximizing the packing radius $R_{max}(D_s)$.

In summary, we suggest that $R_{max} = r_{min}$ can serve as an objective metric for quantifying the impact of sparse sampling pattern D_s on depth completion performance - the smaller this value, the better performance we can achieve (at least in theory). Using this radius as a criterion for evaluating the sampling strategy, we have compared the five sampling strategies as shown in Fig. 1. It can be observed that: 1) all four quasirandom sampling strategies have much better performance than random sampling; 2) all four quasirandom sampling strategies have shown comparable performance. For $n \leq 350$, $R2$ shown slightly better performance; while *Golden* shows better performance for $n \in [350, 750]$.

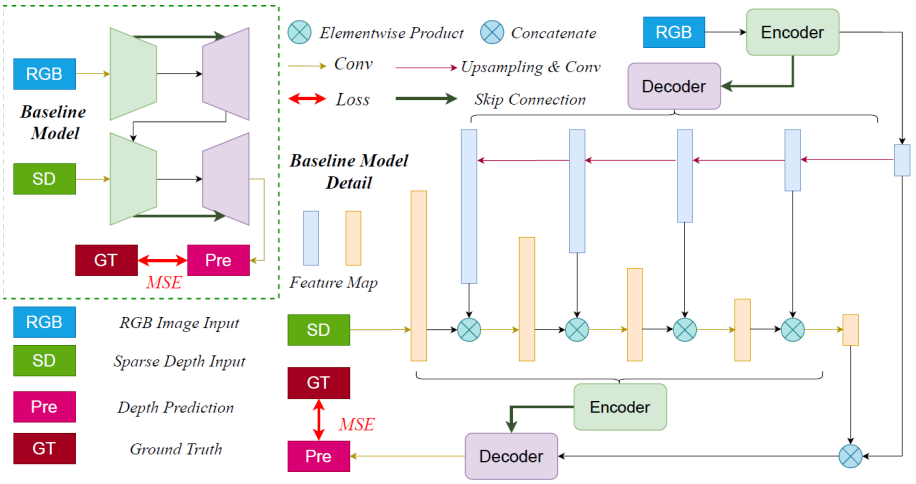

Fig. 2. The architecture of our baseline model for depth completion. Specifically, we substitute the normal convolution in the encoder part of baseline model with the basic residual block structure proposed by [25].

4 Graph Construction for GNN-based Depth Completion

To better motivate our network design, we start from a baseline model implementing spatially varying convolution kernel and then proceed with the full

model incorporating 3D graph with GNN module. The construction of 3D graph using kNN-based neighborhood is the soul of our full model, which distinguishes our approach from existing ones.

4.1 Spatially-Variant Filter and Neighborhood Consideration

The concept of spatially-variant filter dated back to guided image filtering [24] which was originally proposed as an edge-preserving smoothing operator for images. Under the context of color-guided depth completion, guided image filter is still useful to capture rapid changes of depth values around a sharp depth discontinuity or an object boundary, which leads to recently developed GuideNet [54]. This concept was later extended in Dynamic Filter Networks [30] where filters are generated dynamically from the input. For different inputs, a special filter generating network will produce different filters which can be adaptively applied to the output by the dynamical filtering layer in the network. Inspired by both [30] and [54], we propose a multi-scale extension of dynamic filter networks to support sparse-to-dense depth completion in this work.

Our baseline model consists of two full-convolution subnetworks which have similar shapes as U-net [50]. One subnetwork (filter-generation) takes the color guide image as the input and generates feature maps at different scales that will be used as spatially-variant filters. Another subnetwork (depth-completion) takes sparse depth measurements as the input and generates a dense depth map as the final result. Both networks have almost identical encoder-decoder configurations but do not share the weights. As shown in Fig. 2, we multiply the feature maps generated by two subnetworks (filter generation and depth completion). Note that GuideNet [54] utilizes the generated filters as the weight of convolution kernels; however in view of the prohibitive complexity with convolution, we simply use element-wise product in our approach.

To further exploit spatially-variant mechanism, we propose a 3D extension that is specially tailored for depth completion tasks. Note that unlike color images, depth maps indeed convey important 3D information about the physical world. If we consider the abstraction of projection onto imaging plane by a pinhole camera model, the 2D neighborhood in the projected image plane is different from the actual 3D neighborhood in the physical world as shown in Fig. 3. For the definition of neighborhood in 3D space, it is intrinsically connected with the object which the point of interest belongs to. While for 2D, we might not have the luxury of accessing to the complete neighborhood information (e.g., due to occlusion). Such observation motivates us to construct a graph-based representation for spatially-variant neighborhood as we will elaborate next.

4.2 Graph Construction and Network Propagation

The input data to Graph Neural Networks (GNNs) [21, 51] are often represented by a graph $G = \{V, E\}$, where V and E denote the nodes and edges. For each node $v \in V$, we use U_v, f_v and h_v^t to represent its neighborhood, input feature and state for time t respectively. Note that the hidden dynamics of all nodes in a

Fig. 3. Comparison of neighborhood in 2D and 3D space. (a) the color image (the red point in the red square is the point of interest); (b) the 2D neighborhood in the imaging plane; (c) the 3D neighborhood considering the pin-hole camera model (we first use model-inverted 3D coordinates to calculate the neighborhood of the red point in the physical world, then project those points to the 2D imaging plane).

GNN will evolve over time. Such time-evolving process can be modeled by [47]:

$$\begin{cases} m_v^t = \mathcal{M}(\{h_u^t\}), & u \in U_v \\ h_v^{t+1} = \mathcal{F}(h^t, m_v^t) \end{cases} \tag{16}$$

where $\mathcal{M}$ is a function to fuse information from the neighborhood of node v, $\mathcal{F}$ is a function to update its hidden state, m_v^t is a feature vector generated by $\mathcal{M}$ containing information from the node's neighborhood. Similar to a recurrent neural network (RNN) [29, 47], $\mathcal{M}$ and $\mathcal{F}$ share weights at different time steps.

Since the baseline model computes features for each point, we directly use these features as the input to our GNN. Specifically, we use the feature map at 1/8 of original scale; we have used median pooling to downsample the initial dense depth map D_N to 1/8 scale. Moreover, we propose to merge the sparse depth inputs D_I and baseline output D_B as follows:

$$D_N^i = \begin{cases} D_I^i & D_I^i \neq 0 \\ D_B^i & D_I^i = 0 \end{cases} \tag{17}$$

where D_N^i, D_I^i and D_B^i are the depth values of point i in D_N, D_I and D_B, respectively. Then we construct the graph using D_N. Let $[u, v]$ be the coordinates of a point in the imaging plane and $[x, y, z]$ be the 3D coordinates of this point in the camera coordinate system. We can easily convert between 2D and 3D coordinates using the standard pinhole camera model [23].

To design a GNN for depth completion task, we need to identify the 3D neighborhood of each point in the color image first. Similar problem has been considered in recent work 3DGNN [47] but for a different application (semantic segmentation) and with a more favorable assumption (access to the dense instead of sparse depth maps). Inspired by 3DGNN [47], we make each point in the image as a node and find out its k-Nearest-Neighbor(kNN) in 3D space (k is set to 64 in all our experiments). It should be noted that our constructed graph is a directed graph - e.g., node A is the nearest neighbor of B while B is NOT the nearest neighbor of A, so there is only a one-way edge between A and B. This way all directed edges convey information about how each node will obtain information from others. As shown in Fig. 4, we can create the new dense depth map D_N by searching 3D neighbors and propagating information on GNN.

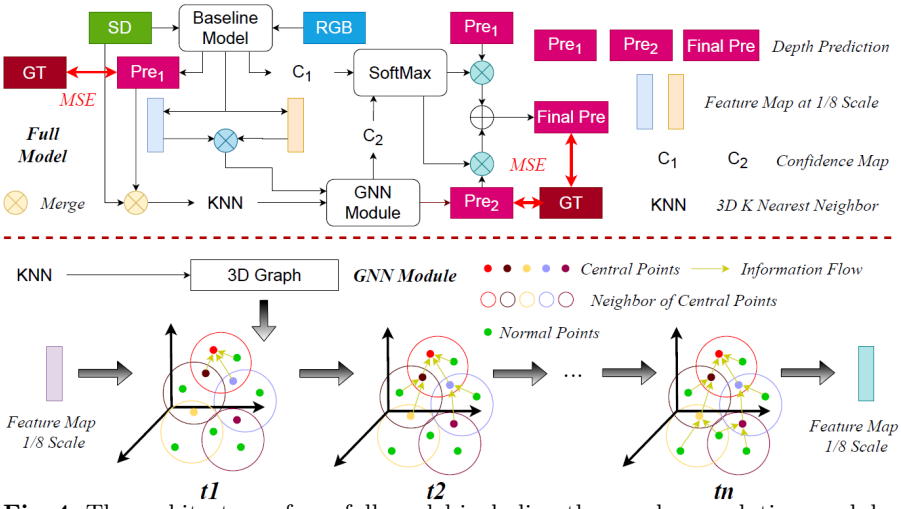

Fig. 4. The architecture of our full model including the graph convolution module.

The propagation process on GNN ($\mathcal{M}$ and $\mathcal{F}$ in Eq. (16)) is implemented by a multi-layer perceptron (MLP) with ReLu [34]. The propagation process can be written as:

$$\begin{cases} m_v^t = \frac{1}{|U_v|} \sum_{v \in U_v} g(h_u^t) \\ h_v^{t+1} = \mathcal{F}(h^t, m_v^t) \end{cases} \tag{18}$$

where g denotes MLP. During the propagation, each node in the graph first collects information from its neighbors; then features will be computed by g and m_v^t is the average feature over the whole neighborhood. Finally every node updates its state based on the previous information. We use MLP as the update function $\mathcal{F}$ because it is commonly used in RNN (e.g., [47]).

5 Experimental Results

5.1 Datasets

KITTI Dataset The KITTI depth completion dataset [19] provides 86898 training samples, 1000 validating samples and 1000 testing samples taken in outdoor scenes. The ground truth depth is not 100% dense and only about 30% of points in the ground truth depth map have valid depth value. Since these points with valid depth values are at the bottom, following [56] we opt to crop the bottom portion of input images to $256 * 1216$ for training and testing.

NYUDv2 Dataset The NYUDv2 dataset [53] provides RGB images and dense depth maps captured by Microsoft Kinect from 464 indoor scenes. Its raw data contains more than $100k$ samples and following [48], [8], [41], we use about 46k samples as training data. Besides, NYUDv2 provides an officially labeled subset containing 1449 samples (654 for testing and evaluation). To fill missing values, the depth values are in-painted using the official toolbox, which adopts the colorization scheme [36]. Following [63] and [41], we have down-sampled the origin images to half-resolution and center-cropped to the dimension of 320×256 with additional paddings.

Matterport3D Matterport3D[5] is an indoor large-scale RGB-D dataset with $10.8k$ real panoramic views and 90 real indoor scenes. We use the same training and testing split as Zhang [65]. There are about 104k samples for training and 474 samples for testing. The ground truth depth map of Matterport3D is generated from Zhang [65] using multi-view reconstruction method.

5.2 Evaluation Metrics and Loss Function

In all of our experiments, we use Mean Square Error(MSE) as the loss function:

$$Loss = \frac{1}{N} \sum_{k \in K_v} ||D_k^{gt} - D_k^{pre}||^2 \tag{19}$$

where K_v is the set of valid points, N is the number of valid points and D_k^{gt} and D_k^{pre} are ground truth depth value and predicted depth value, respectively. For evaluation metric, we take root mean squared error (RMSE) as a prime index to evaluated our models. Besides, mean absolute error (MAE) and mean absolute relative error (REL) are used. For indoor scenes, percentage index δ_i which means the percentage of predicted pixels where the relative error is less a threshold i are used. Specifically, i is chosen as $1.25, 1.25^2, 1.25^3$.

Table 1. Ablation studies on the test set of the NYUDv2 dataset. The effectiveness of GNN module and various sampling strategies are evaluated, respectively.

Method	Sample	Rel↓	RMSE↓	δ_1↑	δ_2↑	δ_3↑
Ours Baseline	Random	0.020	0.112	99.4	**99.9**	**100**
	R2	0.017	0.101	99.5	**99.9**	**100**
	Plastic	**0.016**	0.095	99.5	**99.9**	**100**
	Golden	**0.016**	**0.094**	**99.6**	**99.9**	**100**
	Halton2,3	0.017	0.098	99.5	**99.9**	**100**
Ours GNN	Random	0.016	0.106	99.5	**99.9**	**100**
	R2	0.015	0.092	**99.6**	**99.9**	**100**
	Plastic	0.014	0.091	**99.6**	**99.9**	**100**
	Golden	**0.013**	**0.087**	**99.6**	**99.9**	**100**
	Halton2,3	0.015	0.093	99.5	**99.9**	**100**

5.3 Ablation Study

Neighborhood Construction via GNN We show the effectiveness of neighborhood construction via GNN by comparing against its baseline without GNN. Quantitative results are listed in Table 1. Consistent improvement can be observed under different sampling strategies when comparing GNN to baseline models, which suggest the effectiveness of neighborhood construction via GNN.

Evaluation of Various Sampling Strategies We compare 4 quasi random sampling strategies against random sampling, and Table 1 presents the results. We can observe that the RMSE values of quasirandom sampling are much smaller than that of random sampling, and Golden is the best one among 4 quasirandom strategies. This finding is consistent with our analysis in Section 3.2 and Fig. 1 - i.e., r_{min} metric of the Golden sampling is the smallest.

5.4 Comparison with State-of-the-Art

We compare our method against several state-of-the-art methods on two indoor datasets (NYUDv2 and Matterport3D) and one outdoor dataset (Kitti). Tables 2, 3 and 4 show the quantitative results. It can be observed that our GNN achieves much better results than previous methods on indoor datasets and highly competitive result on outdoor datasets. Especially for NYUDv2 dataset, our GNN with Golden sampling has dramatically outperformed all existing approaches in the open literature. Figs. 5 and 6 include the visual comparison of reconstructed depth maps among several competing methods. It can be observed that our result are consistently the closest to the ground-truth on both datasets.

Table 2. Comparison with state-of-the-art methods on the test set of the NYUDv2. For fair comparison, we add the random sampling strategy of our GNN.

Method	Sample	Rel↓	RMSE↓	δ_1↑	δ_2↑	δ_3↑
Bilateral [53]	Random	0.084	0.479	92.4	97.6	98.9
Ma et al.[41]	Random	0.044	0.230	97.1	99.4	99.8
Eldesokey et al.[17]	Random	0.018	0.129	99.0	99.8	**100**
CSPN [9]	Random	0.016	0.117	99.2	**99.9**	**100**
DeepLiDAR [48]	Random	0.022	0.115	99.3	**99.9**	**100**
Xu et al.[63]	Random	0.018	0.112	99.5	**99.9**	**100**
Ours Baseline	Random	0.020	0.112	99.4	**99.9**	**100**
Ours GNN	Random	0.016	0.106	99.5	**99.9**	**100**
Ours GNN	Golden	**0.013**	**0.087**	**99.6**	**99.9**	**100**

Table 3. Comparison with other methods on the test set of the Matterport3D.

Method	RMSE↓	MAE↓
Huang et al. [26]	1.092	**0.342**
Zhang et al. [65]	1.316	0.461
MRF [22]	1.675	0.618
AD [39]	1.653	0.610
Ours GNN	**0.860**	0.462

Table 4. Comparison with other methods on the KITTI validation set.

Method	RMSE↓	MAE↓	Rel↓
Dfusenet [52]	1240	429	0.022
Ours baseline	883	280	0.016
Ma et al. [40]	858	311	0.019
Ours GNN	**831**	**247**	**0.013**

5.5 Cross Dataset Evaluation

We further conduct cross-dataset experiments to show the transferability of our method. Two transfer scenes (outdoor→indoor and indoor→indoor) are considered, and the experimental results are shown in Tables 5 and 6. Our GNN achieves significant improvements in both scenarios. First, our approach (baseline model without GNN) shows much better transfer results than [56]. Second, our approach with quasi-random sampling achieved much better transfer results than random sampling. Besides the good transferability of our model, the feature is more transferable. We show the transferability by fixing the feature extractor and only fine-tuning the last layer (highlighted by mode "F" in Table 5). Similar improvements can be observed as transferring the whole model, and transferring the feature achieves better cross dataset results.

It is worth mentioning that initial sparse depth maps in the Matterport3D are generated differently from those in the NYUDv2, where the depth values of

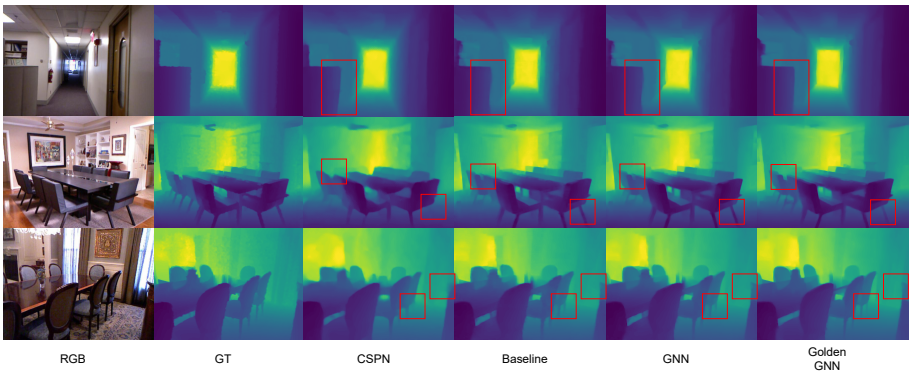

Fig. 5. Results on NYUDv2 dataset. From left to right is RGB image, the ground truth , result from [9], results of our baseline model, results of our full model and results of our full model with golden sampling strategy. Except the rightest one, other methods are tested using randomly sampling strategy.

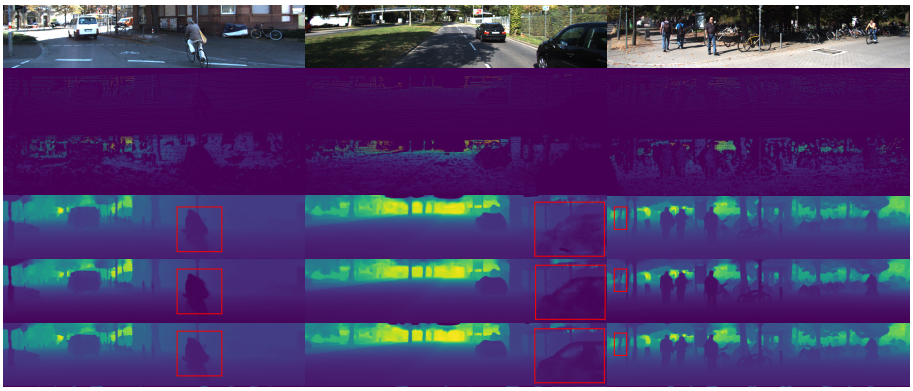

Fig. 6. Results on KITTI dataset. From top to bottom is RGB image, the sparse depth input, the ground truth , results of our baseline model, result from [40], results of our full model.

local areas are completely missing (i.e., noticeable holes in the depth map and visually similar to random sampling). To evaluate the generalization property of our work between different datasets, we have compared two transfer settings: $R2R$ using initial sparse depth maps (without filling the holes) and $Q2Q$ using a modified sparse depth maps (resampled by quasi-random sampling). As shown in Table 6, when adopting the original sparse depth maps as the initial input, our trained network on NYUDv2 with random sampling performs better than quasi-random sampling, which justifies that the original sparse depth maps are indeed more similar to random sampling. By contrast, when adopting the modified sparse depth maps as the input, our trained network with quasi-random sampling performs better than random sampling. As expected, the RMSE of $Q2Q$ setting is much smaller than that of $R2R$ setting. These findings suggest that our GNN model with quasi-random sampling enjoys good generalization properties.

Table 5. Result of models transferred from KITTI dataset to NYUDv2 dataset using different sampling strategies. All models are pre-trained on KITTI dataset. For method, Ours GNN is our full model with graph neural network. Mode "D" means directly evaluating the model on NYUDv2 test set with 654 samples and mode "F" means fine-tuning the model on NYUDv2 subset with 795 samples. For methods with *, we test the pre-trained model provided by their author.

Method	Mode	Sample	Rel↓	RMSE↓	δ_1↑	δ_2↑	δ_3↑
*Van et al. [56]	D	Random	0.090	0.487	85.7	93.6	97.2
Ours Baseline	D	Random	0.087	0.365	90.6	97.7	99.3
Ours GNN	D	Random	0.061	0.310	94.0	98.4	99.5
Ours Baseline	D	R2	0.060	0.261	95.8	99.0	**99.8**
Ours Baseline	D	Plastic	0.060	0.262	95.7	99.0	99.7
Ours Baseline	D	Golden	0.058	0.256	95.7	98.7	99.7
Ours Baseline	D	Halton2,3	0.070	0.294	94.6	98.7	99.7
Ours GNN	D	R2	**0.044**	**0.238**	**96.6**	**99.2**	**99.8**
Ours GNN	D	Plastic	**0.044**	**0.238**	**96.6**	**99.2**	**99.8**
Ours GNN	D	Golden	0.045	0.240	96.4	99.1	99.7
Ours GNN	D	Halton2,3	0.048	0.254	96.1	99.1	99.7
Ours Baseline	F	Random	0.075	0.275	93.9	98.9	99.8
Ours Baseline	F	R2	0.047	0.189	97.3	99.6	**99.9**
Ours Baseline	F	Plastic	0.047	0.189	97.3	99.6	**99.9**
Ours Baseline	F	Golden	0.044	0.177	97.4	99.6	**99.9**
Ours Baseline	F	Halton2,3	0.055	0.211	96.6	99.5	**99.9**
Ours GNN	F	Random	0.050	0.212	97.1	99.5	**99.9**
Ours GNN	F	R2	0.036	0.165	98.2	**99.7**	**99.9**
Ours GNN	F	Plastic	0.036	0.166	98.2	**99.7**	**99.9**
Ours GNN	F	Golden	**0.034**	**0.158**	**98.3**	**99.7**	**99.9**
Ours GNN	F	Halton2,3	0.040	0.177	98.1	**99.7**	**99.9**

Table 6. RMSE results of our GNN transferred from NYUDv2 to Matterport3D in two different transfer settings ($R2R$ and $Q2Q$) are evaluated.

Setting	Mode	Random	Plastic	Golden	R2	Halton
$R2R$	D	**1.831**	1.859	1.913	1.855	1.895
$Q2Q$	D	0.501	0.483	0.489	0.487	**0.460**

6 Conclusion

In this paper, we have studied different sampling strategies approximating Poisson disk sampling and their impact on sparse-to-dense depth completion. We show that random sampling is far from being optimal and quasi-random sampling constructed by low-discrepancy sequences can significantly outperforms random sampling. We have also proposed to construct a 3D graph based on kNN neighborhood information and develop a novel propagation based depth completion algorithm based on Graph Neural Network. Our proposed depth completion method outperforms the state-of-the-art on indoor scenes and has good generalization property.

References

1. Bratley, P., Fox, B.L., Niederreiter, H.: Implementation and tests of low-discrepancy sequences. ACM Transactions on Modeling and Computer Simulation (TOMACS) **2**(3), 195–213 (1992)
2. Bridson, R.: Fast poisson disk sampling in arbitrary dimensions. SIGGRAPH sketches **10**, 1278780–1278807 (2007)
3. Bruna, J., Zaremba, W., Szlam, A., LeCun, Y.: Spectral networks and locally connected networks on graphs. arXiv preprint arXiv:1312.6203 (2013)
4. Caflisch, R.E.: Monte carlo and quasi-monte carlo methods. Acta numerica **7**, 1–49 (1998)
5. Chang, A., Dai, A., Funkhouser, T., Halber, M., Niessner, M., Savva, M., Song, S., Zeng, A., Zhang, Y.: Matterport3d: Learning from rgb-d data in indoor environments. International Conference on 3D Vision (3DV) (2017)
6. Chen, Y., Yang, B., Liang, M., Urtasun, R.: Learning joint 2d-3d representations for depth completion. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 10023–10032 (2019)
7. Chen, Z., Badrinarayanan, V., Drozdov, G., Rabinovich, A.: Estimating depth from rgb and sparse sensing. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 167–182 (2018)
8. Cheng, X., Wang, P., Yang, R.: Depth estimation via affinity learned with convolutional spatial propagation network. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 103–119 (2018)
9. Cheng, X., Wang, P., Yang, R.: Learning depth with convolutional spatial propagation network. arXiv preprint arXiv:1810.02695 (2018)
10. Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y.: Learning phrase representations using rnn encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078 (2014)
11. Defferrard, M., Bresson, X., Vandergheynst, P.: Convolutional neural networks on graphs with fast localized spectral filtering. In: Advances in neural information processing systems. pp. 3844–3852 (2016)
12. Doria, D., Radke, R.J.: Filling large holes in lidar data by inpainting depth gradients. In: 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. pp. 65–72. IEEE (2012)
13. Dunlap, R.A.: The golden ratio and Fibonacci numbers. World Scientific (1997)
14. Early, D.S., Long, D.G.: Image reconstruction and enhanced resolution imaging from irregular samples. IEEE Transactions on Geoscience and Remote Sensing **39**(2), 291–302 (2001)
15. Eigen, D., Fergus, R.: Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture. In: Proceedings of the IEEE international conference on computer vision. pp. 2650–2658 (2015)
16. Eldar, Y., Lindenbaum, M., Porat, M., Zeevi, Y.Y.: The farthest point strategy for progressive image sampling. IEEE Transactions on Image Processing **6**(9), 1305–1315 (1997)
17. Eldesokey, A., Felsberg, M., Khan, F.S.: Confidence propagation through cnns for guided sparse depth regression. IEEE transactions on pattern analysis and machine intelligence (2019)
18. Ferstl, D., Reinbacher, C., Ranftl, R., Rüther, M., Bischof, H.: Image guided depth upsampling using anisotropic total generalized variation. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 993–1000 (2013)

19. Geiger, A., Lenz, P., Urtasun, R.: Are we ready for autonomous driving? the kitti vision benchmark suite. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition. pp. 3354–3361. IEEE (2012)
20. Godard, C., Mac Aodha, O., Brostow, G.J.: Unsupervised monocular depth estimation with left-right consistency. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 270–279 (2017)
21. Gori, M., Monfardini, G., Scarselli, F.: A new model for learning in graph domains. In: Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005. vol. 2, pp. 729–734. IEEE (2005)
22. Harrison, A., Newman, P.: Image and sparse laser fusion for dense scene reconstruction. In: Field and Service Robotics. pp. 219–228. Springer (2010)
23. Hartley, R., Zisserman, A.: Multiple view geometry in computer vision. Cambridge university press (2003)
24. He, K., Sun, J., Tang, X.: Guided image filtering. In: European conference on computer vision. pp. 1–14. Springer (2010)
25. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
26. Huang, Y.K., Wu, T.H., Liu, Y.C., Hsu, W.H.: Indoor depth completion with boundary consistency and self-attention. In: Proceedings of the IEEE International Conference on Computer Vision Workshops. pp. 0–0 (2019)
27. Huang, Z., Fan, J., Cheng, S., Yi, S., Wang, X., Li, H.: Hms-net: Hierarchical multi-scale sparsity-invariant network for sparse depth completion. IEEE Transactions on Image Processing **29**, 3429–3441 (2019)
28. Illian, J., Penttinen, A., Stoyan, H., Stoyan, D.: Statistical analysis and modelling of spatial point patterns, vol. 70. John Wiley & Sons (2008)
29. Jaeger, H.: Tutorial on training recurrent neural networks, covering BPPT, RTRL, EKF and the" echo state network" approach, vol. 5. GMD-Forschungszentrum Informationstechnik Bonn (2002)
30. Jia, X., De Brabandere, B., Tuytelaars, T., Gool, L.V.: Dynamic filter networks. In: Advances in Neural Information Processing Systems. pp. 667–675 (2016)
31. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907 (2016)
32. Kocis, L., Whiten, W.J.: Computational investigations of low-discrepancy sequences. ACM Transactions on Mathematical Software (TOMS) **23**(2), 266–294 (1997)
33. Krcadinac, V.: A new generalization of the golden ratio. Fibonacci Quarterly **44**(4), 335 (2006)
34. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems. pp. 1097–1105 (2012)
35. Kuipers, L., Niederreiter, H.: Uniform distribution of sequences. Courier Corporation (2012)
36. Levin, A., Lischinski, D., Weiss, Y.: Colorization using optimization. In: ACM SIGGRAPH 2004 Papers, pp. 689–694 (2004)
37. Li, Y., Tarlow, D., Brockschmidt, M., Zemel, R.: Gated graph sequence neural networks. arXiv preprint arXiv:1511.05493 (2015)
38. Liu, F., Shen, C., Lin, G., Reid, I.: Learning depth from single monocular images using deep convolutional neural fields. IEEE transactions on pattern analysis and machine intelligence **38**(10), 2024–2039 (2015)

39. Liu, J., Gong, X.: Guided depth enhancement via anisotropic diffusion. In: Pacific-Rim conference on multimedia. pp. 408–417. Springer (2013)
40. Ma, F., Cavalheiro, G.V., Karaman, S.: Self-supervised sparse-to-dense: Self-supervised depth completion from lidar and monocular camera. In: 2019 International Conference on Robotics and Automation (ICRA). pp. 3288–3295. IEEE (2019)
41. Mal, F., Karaman, S.: Sparse-to-dense: Depth prediction from sparse depth samples and a single image. In: 2018 IEEE International Conference on Robotics and Automation (ICRA). pp. 1–8. IEEE (2018)
42. Marohnić, L., Strmečki, T.: Plastic number: Construction and applications. In: International Virtual Conference ARSA (Advanced Researsch in Scientific Area) 2012 (2013)
43. Matsuo, K., Aoki, Y.: Depth image enhancement using local tangent plane approximations. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3574–3583 (2015)
44. McCool, M., Fiume, E.: Hierarchical poisson disk sampling distributions. In: Proceedings of the conference on Graphics interface. vol. 92, pp. 94–105 (1992)
45. Niederreiter, H.: Random number generation and quasi-Monte Carlo methods, vol. 63. Siam (1992)
46. Park, J., Kim, H., Tai, Y.W., Brown, M.S., Kweon, I.S.: High-quality depth map upsampling and completion for rgb-d cameras. IEEE Transactions on Image Processing **23**(12), 5559–5572 (2014)
47. Qi, X., Liao, R., Jia, J., Fidler, S., Urtasun, R.: 3d graph neural networks for rgbd semantic segmentation. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 5199–5208 (2017)
48. Qiu, J., Cui, Z., Zhang, Y., Zhang, X., Liu, S., Zeng, B., Pollefeys, M.: Deeplidar: Deep surface normal guided depth prediction for outdoor scene from sparse lidar data and single color image. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3313–3322 (2019)
49. Ripley, B.D.: Spatial statistics, vol. 575. John Wiley & Sons (2005)
50. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)
51. Scarselli, F., Gori, M., Tsoi, A.C., Hagenbuchner, M., Monfardini, G.: The graph neural network model. IEEE Transactions on Neural Networks **20**(1), 61–80 (2008)
52. Shivakumar, S.S., Nguyen, T., Miller, I.D., Chen, S.W., Kumar, V., Taylor, C.J.: Dfusenet: Deep fusion of rgb and sparse depth information for image guided dense depth completion. In: 2019 IEEE Intelligent Transportation Systems Conference (ITSC). pp. 13–20. IEEE (2019)
53. Silberman, N., Hoiem, D., Kohli, P., Fergus, R.: Indoor segmentation and support inference from rgbd images. In: European conference on computer vision. pp. 746–760. Springer (2012)
54. Tang, J., Tian, F.P., Feng, W., Li, J., Tan, P.: Learning guided convolutional network for depth completion. arXiv preprint arXiv:1908.01238 (2019)
55. Uhrig, J., Schneider, N., Schneider, L., Franke, U., Brox, T., Geiger, A.: Sparsity invariant cnns. In: 2017 International Conference on 3D Vision (3DV). pp. 11–20. IEEE (2017)
56. Van Gansbeke, W., Neven, D., De Brabandere, B., Van Gool, L.: Sparse and noisy lidar completion with rgb guidance and uncertainty. In: 2019 16th International Conference on Machine Vision Applications (MVA). pp. 1–6. IEEE (2019)

57. Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., Bengio, Y.: Graph attention networks. arXiv preprint arXiv:1710.10903 (2017)
58. Werbos, P.J.: Backpropagation through time: what it does and how to do it. Proceedings of the IEEE **78**(10), 1550–1560 (1990)
59. Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., Yu, P.S.: A comprehensive survey on graph neural networks. arXiv preprint arXiv:1901.00596 (2019)
60. Xian, K., Shen, C., Cao, Z., Lu, H., Xiao, Y., Li, R., Luo, Z.: Monocular relative depth perception with web stereo data supervision. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2018)
61. Xian, K., Zhang, J., Wang, O., Mai, L., Lin, Z., Cao, Z.: Structure-guided ranking loss for single image depth prediction. In: The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (June 2020)
62. Xiong, X., Cao, Z., Zhang, C., Xian, K., Zou, H.: Binoboost: Boosting self-supervised monocular depth prediction with binocular guidance. In: 2019 IEEE International Conference on Image Processing (ICIP). pp. 1770–1774. IEEE (2019)
63. Xu, Y., Zhu, X., Shi, J., Zhang, G., Bao, H., Li, H.: Depth completion from sparse lidar data with depth-normal constraints. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 2811–2820 (2019)
64. Xue, H., Zhang, S., Cai, D.: Depth image inpainting: Improving low rank matrix completion with low gradient regularization. IEEE Transactions on Image Processing **26**(9), 4311–4320 (2017)
65. Zhang, Y., Funkhouser, T.: Deep depth completion of a single rgb-d image. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 175–185 (2018)
66. Zhou, J., Cui, G., Zhang, Z., Yang, C., Liu, Z., Wang, L., Li, C., Sun, M.: Graph neural networks: A review of methods and applications. arXiv preprint arXiv:1812.08434 (2018)