

Lidar Upsampling with Sliced Wasserstein Distance

Artem Savkin^{1,2}
TUM, BMW

Yida Wang¹
TUM

Sebastian Wirkert²
BMW

Nassir Navab^{1,3}
TUM, JHU

Federico Tombari^{1,4}
TUM, Google

Abstract—Lidar became an important component of the perception systems in autonomous driving. But challenges of training data acquisition and annotation made emphasized the role of the sensor to sensor domain adaptation. In this work, we address the problem of lidar upsampling. Learning on lidar point clouds is rather a challenging task due to their irregular and sparse structure. Here we propose a method for lidar point cloud upsampling which can reconstruct fine-grained lidar scan patterns. The key idea is to utilize edge-aware dense convolutions for both feature extraction and feature expansion. Additionally applying a more accurate Sliced Wasserstein Distance facilitates learning of the fine lidar sweep structures. This in turn enables our method to employ a one-stage upsampling paradigm without the need for coarse and fine reconstruction. We conduct several experiments to evaluate our method and demonstrate that it provides better upsampling.

I. INTRODUCTION

Lidar can significantly improve the quality and reliability of the perception systems of autonomous vehicles. As an active sensor, lidar takes accurate measurements of the surrounding environment and provides spatial information about the traffic scene in form of point clouds. It is robust to challenging light conditions and can significantly contribute to the functional safety of fully autonomous systems. Also, the development and manufacturing of lidar have been democratized considerably in recent years. Democratization enabled the equipment of the serial vehicles with this sensor. As a result, lidar became a crucial component of state-of-the-art autonomous driving.

At the same time, machine perception on point clouds is still challenging since this kind of data is unstructured and irregular. However, in computer vision, neural networks have already proved superior to traditional algorithms.

Their advantage is especially evident in the tasks of segmentation or detection where *understanding* of the semantics of the underlying scene is involved. A wide variety of methods dedicated to 3D object detection, tracking, and segmentation successfully adapted deep learning techniques to point clouds to tackle the challenges mentioned above [26], [34], [29], [43].

In practical applications such as autonomous driving, data acquisition and annotation required for training such perception methods is a expensive and inflexible process.

¹TUM, 85748 Munich, Germany artem.savkin@tum.de; tombari@in.tum.de

²BMW AG, 80809 Munich, Germany

³JHU, Baltimore, MD 21218, United States

⁴Google, 8002 Zurich, Switzerland

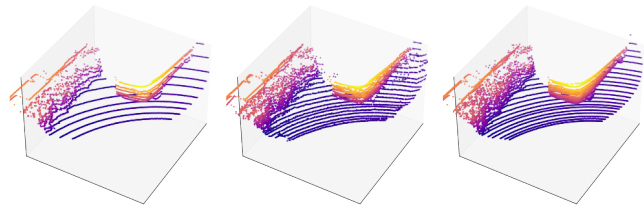


Fig. 1: Results of $2\times$ upsampling for 8129 points patches from KITTI dataset.

Such rigidity is characterized by a model’s performance degradation attributed to data modality change e.g., sensor viewpoint, resolution, etc. [7]. For example, [3] demonstrates that lidar segmentation methods reveal significant performance degradation when evaluated on the samples from a lidar in which mounting position, number of scan lines, or intensity calibration differs from the training data. Table II shows similar behavior, demonstrated by the object detection networks operating on lidar samples with different resolutions. The phenomenon which causes degradation is commonly referred to as domain shift [31] and addressed through transfer learning. They aim to address the domain shift problem by learning the domain-invariant features or by learning the direct mapping between source and target domains. Still, models are supposed to be compatible with novel sensor generations and, thus, with data modalities or domains. In this work, we address the problem of such sensor-to-sensor domain adaptation as a mechanism to avoid actual data re-acquiring or re-annotating the data. In particular, we focus here on the task with low- and high-resolution lidars. The latter is distinguished by a higher number of scan lines in the vertical field of view. Thus, the overarching goal is to sample the data points from the target distribution or, in other words, such point clouds which resemble characteristics of data samples from the target domain. Our task is to generate a point cloud sample with more scan lines given a low-resolution lidar scan sample. We consider both modalities separate domains and employ a point cloud transfer technique to solve them.

Existing methods for lidar generation mainly rely on distance measures such as Chamfer Distance (CD) and Earth Mover’s Distance (EMD) [44], [47]. Although several works [11], [1] establish the superiority of EMD over CD as more discriminative for point cloud learning, the CD simply nearest neighbor based is still more efficient. Yet, from our observations, both measures do not suffice to generate fine-

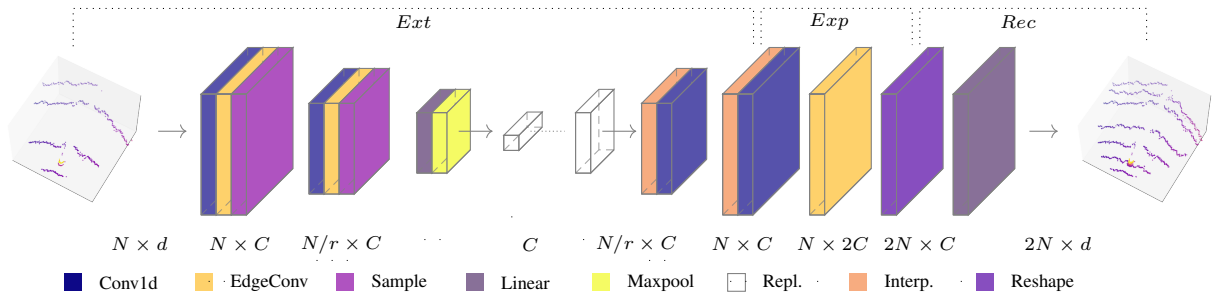


Fig. 2: Overview of the network with feature extraction (*Ext*), feature expansion (*Exp*) and coordinate reconstruction (*Rec*).

structure lidar point clouds. CD fails to equally distribute additional points across the lidar scan, thus erroneously favoring the known or source point locations. Standard EMD approximation based on auction algorithm [11] fails to reconstruct the fine scan pattern of lidar sweeps accurately. This phenomenon is also confirmed by experiments in [40]. Multiple works have addressed the computational complexity of EMD and the whole family of Wasserstein distances (WD) [17], [15], [39], [22] in various tasks.

The contribution of this work is twofold. Firstly, it provides an extensive analysis of applicability of established distance metrics such as Chamfer distance, approximated EMD, and SWD for lidar scan generation and upsampling. Secondly, based on this analysis, it proposes a new lidar upsampling method, which is dedicated to the problem of cross-sensor domain adaptation between low-resolution and high-resolution lidars. We argue that, unlike existing point cloud upsampling methods, our edge-aware one-stage (no coarse reconstruction) approach based on Sliced-Wasserstein distance can reconstruct fine details of lidar scans and surpass existing upsampling techniques.

II. RELATED WORK

The lidar super-resolution or lidar upsampling task has been addressed by the recent learning-based methods, which can be categorized into grid-based methods [28] and point-based [18] methods.

A. Grid-based lidar generation

Given an input condition, e.g., vector, the generation task is to produce a $N \times d$ point cloud of size N and dimensionality d . Grid-based methods utilize the range image representation of point clouds projected into the cylindrical coordinate system. [4] adapts generative adversarial models applied on such range images to produce large-scale lidar scans. Unlike the lidar GAN approach, which uses real lidar scans for the discriminator training, [28] uses training exclusively from simulated data to produce scans comparable with real high-resolution lidar. [16] improves range image upscaling by utilizing a widely acclaimed attention mechanism. Finally, [14] approaches integrating an uncertainty estimation mechanism into the upsampling process to meet real-time constraints. In this work, we aim to develop a rather generic domain adaptation technique for lidar upsampling, which falls into the latter category of point-set-based methods.

B. Point-based lidar generation

Typical point-based generation methods such as AtlasNet [13], and FoldingNet [41] operate directly on the point sets where they approach the problem by creating 2D patches and aligning them to the underlying point cloud shape. With the emergence of generative neural nets, several works like PCGAN [1], GCN-GAN [35], and TreeGAN [30] applied adversarial training frameworks to the point clouds. Luo *et al.* [21] employed a concept similar to image diffusion for the point cloud generation. More recent generative methods utilize normalizing flows - PointFlow [40], gradient fields - ShapeGF [5] and autoregression - PointGrow [32].

Compared to generation, the upsampling aims to compute a $rN \times d$ output given a $N \times d$ point cloud, with upsampling ratio r so that it resembles the spatial characteristics of the underlying 3d shape. Methods introduced by Yu *et al.* [45] and Zhang *et al.* [47] were pioneering deep learning-based approaches for that task. PU-Net [45] relied on Point-Net++ encoder [26] for feature extraction from patches then applied two sets of linear layers onto features replicated from $N \times C$ to $N \times rC$ which are then merely reshaped into $rN \times C$. PU-Net is followed by EC-net [44], which introduced a loss function that aimed to minimize the point-to-edge distance to make the network more edge-aware. Later 3PU [42] explored multi-scale progressive patch-based upsampling by breaking down a r -factor upsampling network into several $2 \times$ upsampling networks. It introduces dense connections or blocks for the feature extraction and defines local neighborhoods later in a feature space via kNN based on feature similarity. PU-GAN [18] adopts an adversarial framework for point cloud upsampling by formulating a discriminator network that learns to evaluate generated high-resolution point clouds. It introduces a *up-down-up* expansion unit, which enables a so-called self-correction of expanded features. More recent methods include PU-GCN [27] and Dis-PU[19].

Another task worth mentioning is point cloud completion, whose goal is to generate complete point clouds and create dense scans from incomplete sparse ones. Here pioneering learning method was PCN [46]. PCN first learns the global feature of the underlying shape from partial observation and then reconstructs a complete but coarse point cloud of lower resolution, which is up-sampled with folding operations in the second phase. TopNet [33] method utilizes a tree structure for generating complete dense point clouds. Following the

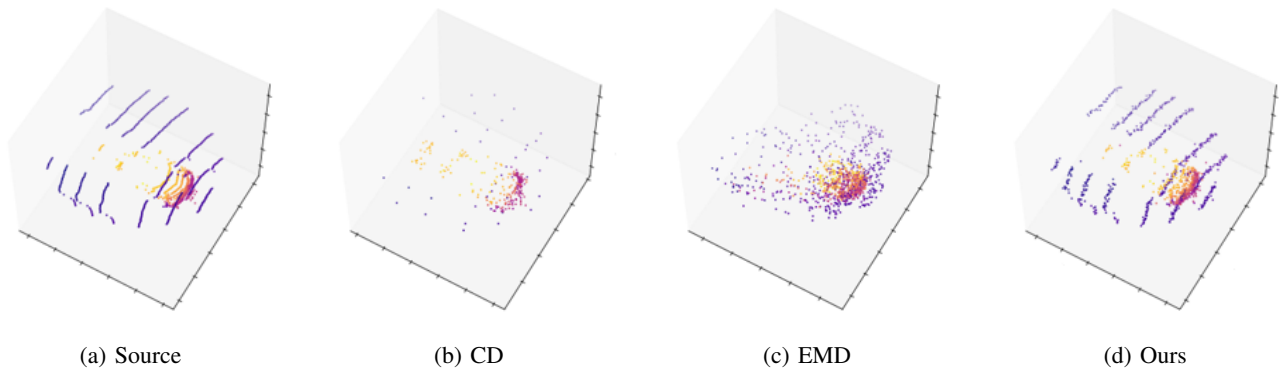


Fig. 3: Point cloud reconstruction using PointNetFCAE with conventional loss functions.

two-stage coarse-fine point cloud generation MSN [20] uses the morphing principle and joint CD and EMD loss. Two-staged ECG [24] improves on the generation of the edges by exploiting the edge-preserved pooling. Finally, CDN [37] extends a traditional GAN framework to guarantee that local areas follow the same pattern as ground truth.

Compared to previous point cloud generation works, our method is dedicated to precisely reconstructing fine-grained lidar point clouds. Discussed point-based methods reveal a relatively restricted applicability area that embraces simplistic, ShapeNet-alike [6] data. Point clouds generated by a lidar sensor demonstrate significantly more complex underlying geometry, scan pattern, and data variance. Typically methods trained on PCN dataset [46] derived from synthetic ShapeNet CAD models embrace eight categories of objects (e.g., plane, car, chair) whereas lidar scans of the KITTI dataset [12] in turn were obtained during drives in the real world and contain multiple classes of objects in a single scan. Thus, the task of point cloud generation in the lidar domain remains under-explored. To our best knowledge, the proposed method is the only point-based method able to reconstruct complex lidar point clouds in the upsampling task accurately.

III. APPROACH

Our method relies on the edge-aware feature extraction, and extension mechanism [42] and projection-based 1-Wasserstein distance. This training objective reveals significantly improved sensitivity to fine-grained lidar scan patterns as opposed to conventional loss functions.

A. Model

The proposed model is built upon established upsampling carcass with three steps: feature extraction, feature expansion, and coordinate reconstruction. The feature extractor is implemented as a plain encoder-decoder architecture and similarly to 3PU [42] employs dense per point feature learning using *Conv1d* and *EdgeConv* [38] layers. An *EdgeConv* layer applies the kNN algorithm to find clusters of points and preserves characteristic attributes like edges. Following the edge-preserving principle, the feature expansion module also takes advantage of *EdgeConv* layer for generating the

$N \times 2C$ upsampled feature vector. In the final step, a set of linear layers reconstruct the resulting $N \times 3$ coordinates.

Feature Extraction. As mentioned before, the encoder first takes the $N \times d$ point cloud as input and translates it into $N/r \times C'$ feature vector by applying several *Conv1d* plus *EdgeConv* blocks. Here, N is the original number of points, r is downsampling ratio, d - point cloud dimensionality (in our experiments $d = 3$) and C' is the resulting size of the feature vector. The common farthest point sampling technique achieves downsampling from N to N/r . This feature vector is transformed to C' -sized global feature, which will be upsampled into $N \times C$ by replication and interpolation. For that, replicated C' feature vectors are just stacked together and interpolated using inverse distance weighted average from [26].

Feature Expansion. Inspired by [24], we design the feature expansion module along the lines of edge awareness by utilizing the *EdgeConv* layer. This layer employs the kNN algorithm and enables the module to find clusters based on similarities in the feature space. It takes a $N \times C$ learned feature vector and transforms it into $N \times 2C$ upsampled feature vector and then reshaped to $2N \times C$ tensor.

Coordinates Reconstruction. the resulting tensor will be then processed by a set of linear layers to reconstruct the point dimensions $2N \times d$. Fig. 2 visualizes the overall architecture.

B. Loss Function

We argue that conventional loss functions based on Chamfer Distance (CD) and approximations of Earth Mover's Distance (EMD) [20] suffer from the lack of *sensitivity* to the fine-grained scan patterns of lidar sensors. As mentioned before, CD unequally distributes generated points across the lidar scan in point cloud reconstruction and favors the known or source point locations. A possible reason for that could be the nature of a generative task where produced point clouds at the initial state represent unstructured point blobs so that CD fails to find correspondences. On the other hand, the reason for EMD to fail inaccurate reconstruction of the fine scan patterns could lie in commonly used approximations which lead to biased or noisy gradients [40] and herewith affect the ability of a model to distinguish the fine details.

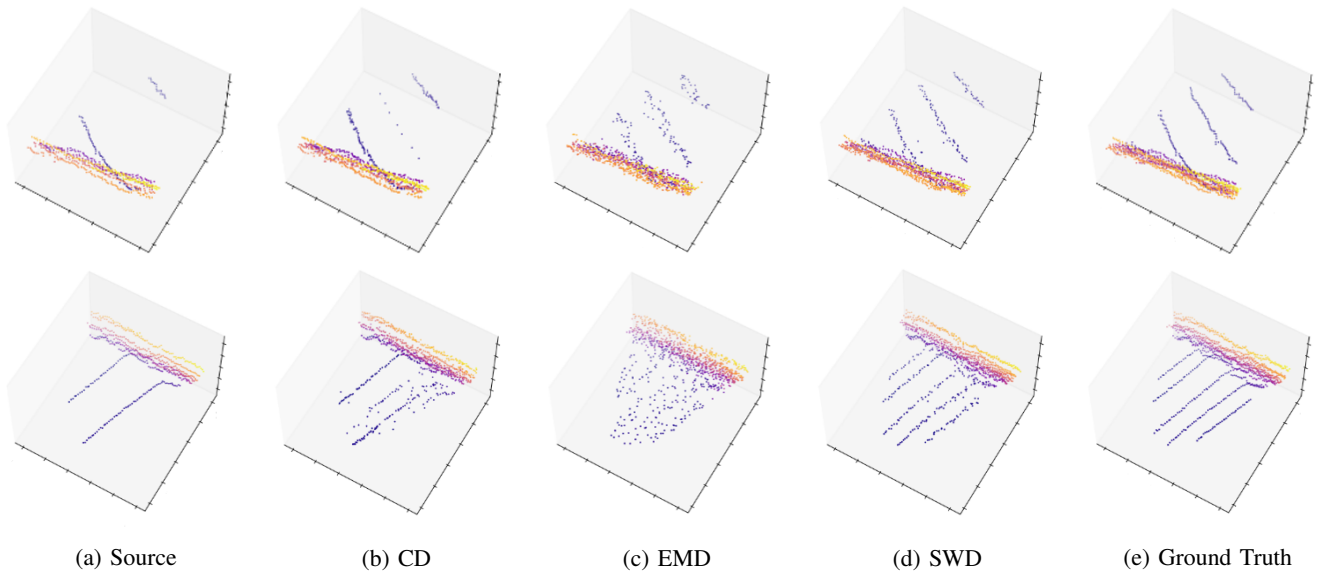


Fig. 4: Lidar upsampling using proposed network with conventional loss functions.

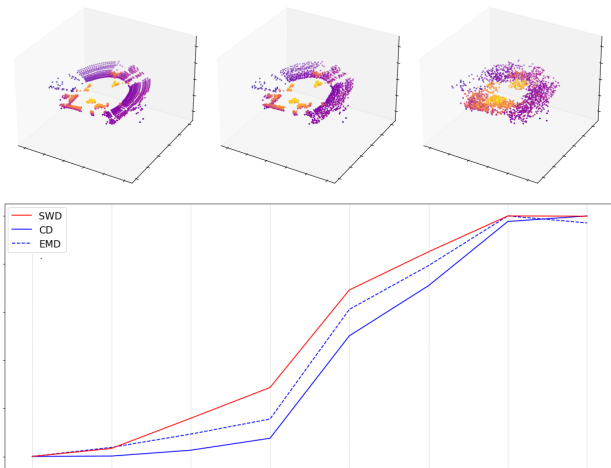


Fig. 5: Distance metrics on jitter (normalized).

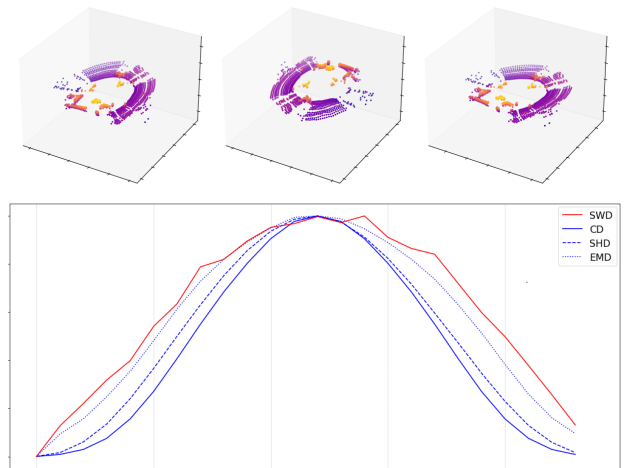


Fig. 6: Distance metrics on rotation (normalized).

Following the Optimal Transport (OT) perspective, we consider the distance between point clouds as the distance between two distributions with a natural way to measure it by a metric function defined between probability distributions - Wasserstein distance. Yet accurate, exact computation of this metric is computationally intractable. The assumption is that more accurate distance calculation is a key to precise lidar scan reconstruction and upsampling. We argue that higher accuracy of point cloud reconstruction, which is sensitive to fine scan patterns, can be achieved by applying the Sliced Wasserstein (SW) distance as it possesses equivalent statistical properties to WD [15], [23]. The intuition behind SW distance resides in the idea of the existence of a closed-form OT solution in a uni-dimensional space. For two point sets X and Y of equal size such as $X, Y \subset \mathbb{R}^3 : |X| = |Y|$,

EMD between X and Y can be calculated as:

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

where $\phi : X \rightarrow Y$ is a bijection [11]. The more general form of EMD in the case of $|X| = |Y|$:

$$EMD(X, Y) = \inf_{P \in \Pi(X, Y)} \mathbb{E}_{(x, y) \sim P} [d(x, y)] \quad (2)$$

is equivalent to Wasserstein distance with $p = 1$:

$$W_p(X, Y) = \inf_{P \in \Pi(X, Y)} \mathbb{E}_{(x, y) \sim P} [d^p(x, y)]^{1/p} \quad (3)$$

where d is a metric and $\Pi(X, Y)$ is a set of all joint distributions of $P(x, y)$ of random variables x and y with marginal distributions X and Y :

$$\Pi(X, Y) = \{P_{X, Y}(x, y)\}$$

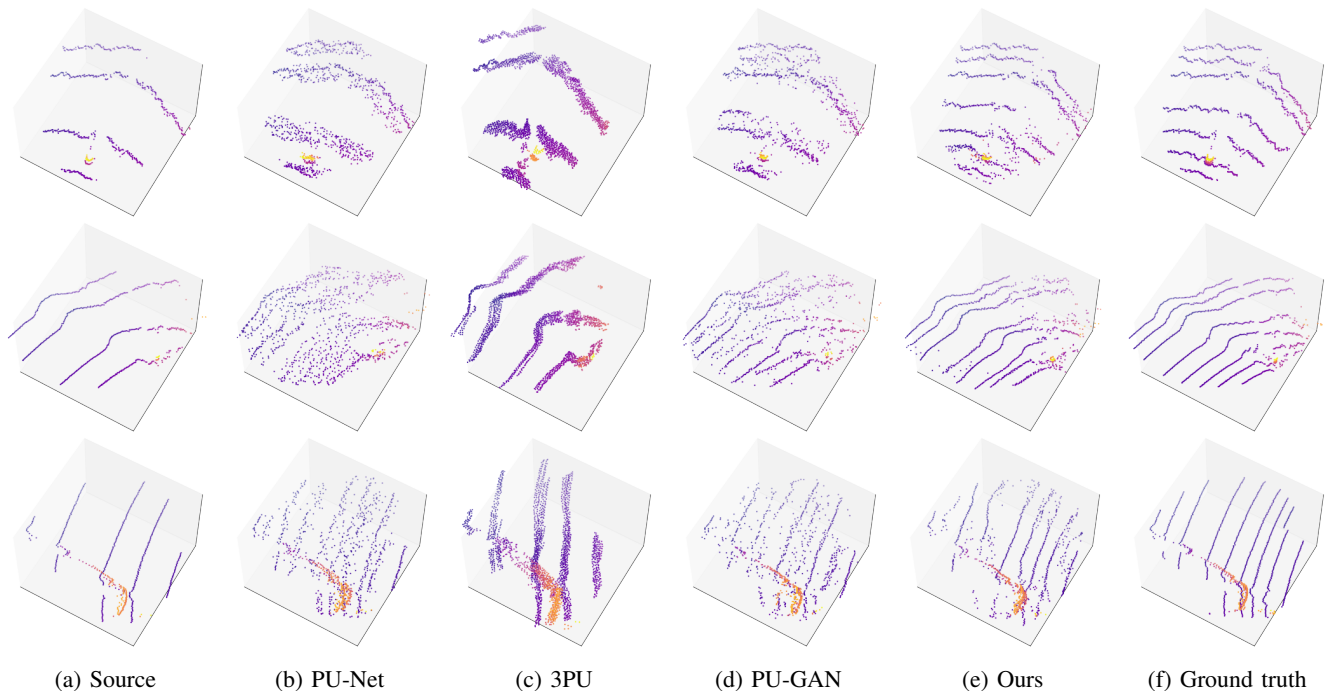


Fig. 7: Qualitative comparison of the $2\times$ lidar upsampling results on KITTI *test*.

In practice, computation of 1 or 3 is intractable, so an approximation scheme such as an auction algorithm must be applied [20] yet it cannot guarantee the required bijection assignment. As opposed to such approximation, the Sliced Wasserstein approach exploits the fact that W_p has a closed-form solution for univariate probability distributions [15]. Thus several works utilize this fact to calculate W_p [39], [9]:

$$SW_p(X, Y) = \int_{\mathbb{S}^{d-1}} W_p(X^s, Y^s) ds \quad (4)$$

where X^s and Y^s are projections of X and Y onto the direction s of all possible directions of a unit sphere \mathbb{S}^{d-1} . In practice, SW_p is approximated by only taking a finite number of random directions from \mathbb{S}^{d-1} . Moreover, in a uni-dimensional case, d can be obtained by just sorting [9]. Hence our final objective can be defined as follows:

$$S\tilde{W}_p(X, Y) = \frac{1}{|\mathbb{S}|} \sum_{s \in \mathbb{S}} \frac{1}{|X|} \sum_i \|X_i^s - Y_i^s\|_2 \quad (5)$$

Thus, the final optimization problem can be formulated as follows:

$$\operatorname{argmin}_{\theta} \mathbb{E}_{X \sim D_X} (S\tilde{W}_p(X, f_{\theta}(X))) \quad (6)$$

IV. EXPERIMENTS

A. Datasets

Our experiments are based on the popular public datasets SemanticKITTI [2] and Waymo Open Dataset [10]. SemanticKITTI is a large-scale computer vision dataset that provides 22 sequences of lidar scans with overall around

43000 scans alongside point cloud segmentation labels. The dataset provides measurements in German inner cities, residential areas, and highway traffic scenes during various daylight and weather conditions. In our experiments we operate with original KITTI [12] dataset split of the sequences divided into *train* and *test* sets of the sizes 7481 and 7518 scans respectively. Semantic labels of the dataset contain information about 28 classes such as *road*, *sidewalk*, *person*, *car*. The Waymo dataset, in turn, provides 1150 sequences of 20 seconds each. In total, it offers around 230K annotated samples with the ground truth for object detection and tracking tasks. For our purposes, every 10th lidar scan has been sampled from all data sequences, resulting in 15947 samples training dataset and 4038 validation dataset. Also, from the five lidars, we only use the scans provided by the top-mounted one.

B. Loss

We propose to train lidar generation models using loss function based on Sliced-Wasserstein distance as opposed to commonly utilized CD, EMD-based losses. Our benchmark experiments suggest that these metrics are sufficient for a model to learn the shapes of point clouds but not the fine structures. Described behavior can be observed in Fig. 3, which shows the results of the reconstruction of the point cloud patches provided by a *vanilla auto-encoder* network PointNetFCAE trained respectively with CD and EMD loss. This auto-encoder model consists of a Pointnet [25] based encoder and a fully connected decoder. Even though the sizes of the visualized patches are equal (2048 points), CD looks underpopulated in certain regions. [1] refers to this phenomenon as *CD blindness*, which manifests



Fig. 8: Lidar upsampling results (middle) on KITTI *val* patches with input samples (top) and ground truth (bottom).

in point placement where the average mass locates and fails to distinguish such poor allocation from the true one. EMD, in turn, appears to reconstruct rough point distribution but is considerably distorted. That finding is aligned with observations in [11] and mirrors its mean-shape experiments.

Additionally, we compare the proposed loss function with existing error metrics concerning their *sensitivity*. The results of this analysis are shown in Fig. 5 and Fig. 6. The comparison is conducted between Sinkhorn loss [8], Sliced Gromov-Wasserstein [36], Sliced Wasserstein Distance [15], Chamfer loss, and EMD loss. Here we execute two experiments by gradually applying a corresponding transformation on the point cloud and measuring the distances between transformed and original samples. In the first experiment, Gaussian noise is introduced into a sample, whereas during the second experiment, the sample is rotated along the vertical axis. In both cases, analyzed metrics behaved consistently by changing to non-zero values as the transformation became more significant in the first experiment and turned back to zero after full rotation in the second experiment. Despite that, Sliced Wasserstein Distance could identify the slightest point cloud changes by indicating higher discrepancy values in the early transformation iterations. The influence of different loss functions CD, EMD, and SWD on the lidar upsampling performed by our model will be discussed in the ablation study in the section IV-C

C. Upsampling

The training occurs supervised where input patches of size 2048 were obtained by decimating the ground truth samples by a factor of two. Decimation is performed based on rasterized input point clouds. For the Waymo dataset, every second scan line of the original cylindrical projection of size $W \times H \times 3$ of the lidar scans was dropped/rejected before projecting the points into the lidar coordinate system. Similarly, we processed the cylindrical projections of the KITTI dataset. As the KITTI dataset provides only raw point clouds, they are required to be rasterized before decimation. Therefore, KITTI scans were arranged into 64×2048 grid based on sensor vertical resolution and azimuth angle for rasterization. This preprocessing is shown in Fig. 1. With that data, the network is trained for 2000 consecutive epochs without pre-trained weights initialization and using Adam optimizer. The result after the training memory footprint is 22 Mb. Training is accomplished on NVIDIA Quadro RTX 6000 GPU(24GB) with a single forward pass that takes,

	CD	HD	EMD	SWD
PU-Net [44]	0.5272	1.2627	0.3851	15.4002
3PU [42]	0.1172	0.4151	0.2740	5.2078
PU-GAN [19]	0.0426	0.1892	0.2504	1.3553
Ours	0.0435	0.1694	0.0775	1.1742

TABLE I: Quantitative comparison for $2\times$ upsampling of 8192 points on KITTI *val* dataset.

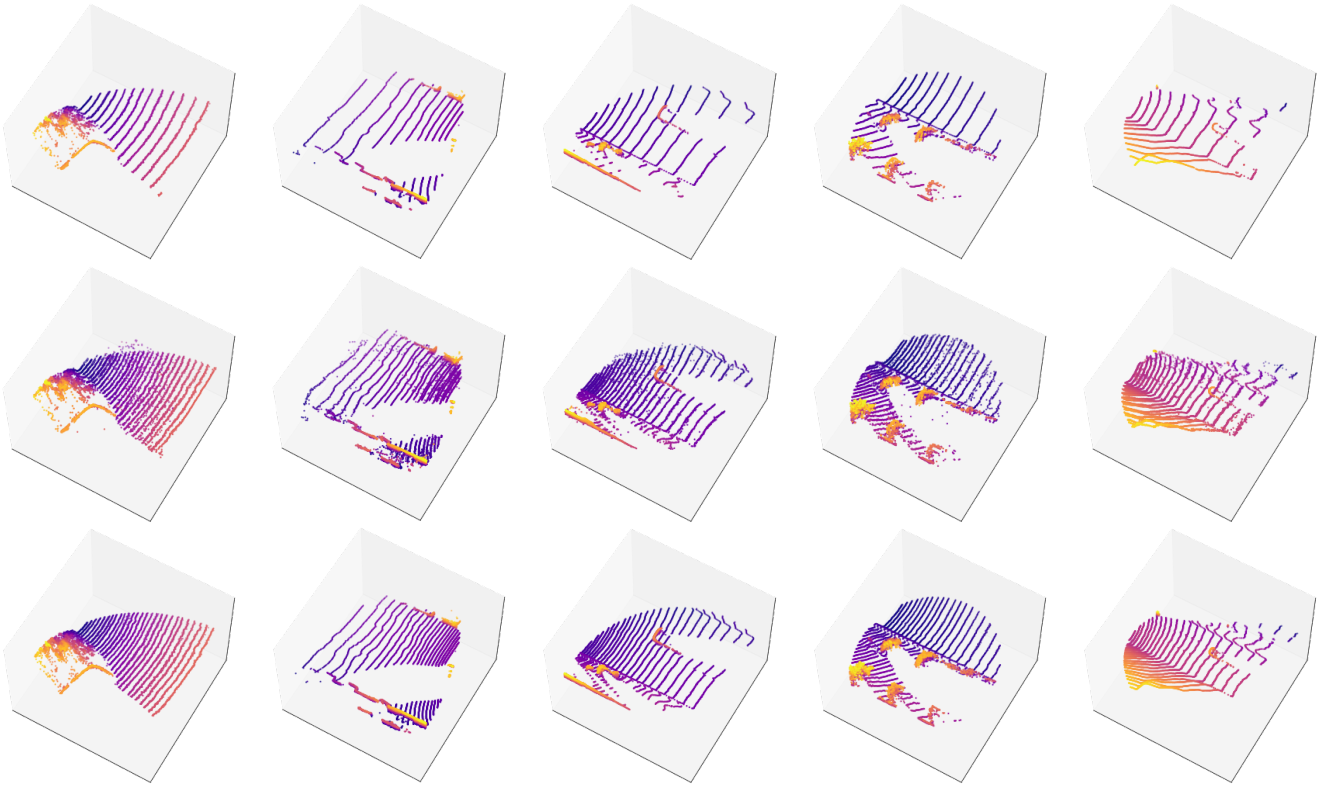


Fig. 9: Lidar upsampling results (middle) on Waymo *val* patches with input samples (top) and ground truth (bottom)

on average 0.2 sec. To evaluate the generated results, we conducted several experiments and assessed our results both qualitatively and quantitatively. The results of the lidar scan upsampling on the KITTI dataset are shown in Fig. 7. To visualize the effects introduced by our approach, we display both sources and upsampled scans together with point clouds generated by the state-of-the-art upsampling methods 3PU, PU-Net, and PU-GAN. We employ original training protocols from respective reports to train those models on lidar point clouds. As reported in the original publication, PU-GAN was trained with adversarial and EMD-based reconstruction loss, 3PU was trained with modified counter-outlier CD loss, and PU-Net used plain EMD loss with repulsion loss. First, we want to draw attention to the fact that the proposed method recovers the scan lines missing in the input samples and, in consequence, effectively reconstructs the high-resolution target scans. In comparison with the results of the existing methods, samples provided by our approach reveal no point scattering in the local neighborhood but rather provide distinguished scan lines. Another characteristic of the generated scans worth considering is the actual shape of the scans. In the upsampled lidar sweeps, recovered scan lines indeed follow the spatial characteristics of the underlying geometry.

Additionally, we evaluate our method of upsampling the point cloud patches on a larger scale. The results of $2\times$ upsampling for patches of size 8192 for both datasets KITTI and Waymo are depicted in Fig. 8 and 9. Here, we also want to emphasize the reconstruction quality of the upsampled

scans.

To perform a quantitative evaluation of the results, we assess the generated data using traditional metrics such as CD, Hausdorff distance (HD), EMD, and introduced SWD. For the best-performing network snapshots, the metrics mentioned above are reported in table I. Our generated data reveals substantial performance improvement concerning HD and SWD.

Furthermore, to verify our hypothesis regarding the SWD loss, we study the effects of different losses CD, EMD, and SWD on the upsampling performance of our model. Here we train our network on the KITTI dataset with the patches of size 1024 and upsampling ratio 2 with a similar setup as described before but with varying loss functions. The results of this study are demonstrated in Fig. 4. They confirm our initial assumption and are aligned with the observations made before. Similarly to the reconstruction experiment here, CD leads to sparseness areas, and EMD leads to distortions. That

Validation	Car easy	Car mod	Car hard	Ped. easy	Ped. mod	Ped. hard	Cyclist easy	Cyclist mod	Cyclist hard
Low res	78.8	65.0	59.7	7.5	4.6	4.2	22.9	16.2	14.9
High res	87.7	88.0	86.6	57.3	51.4	46.8	81.5	62.9	59.0
Upsampled	76.2	57.1	52.4	27.2	21.2	18.7	21.9	13.0	11.7

TABLE II: Results of the PointPillar for KITTI dataset trained on high resolution lidar data and validated on decimated, original and upsampled by our method *val* sets.

observation also confirms the findings of previous works, which state that CD is a weaker metric than 1-Wasserstein and SWD approximation is more accurate than approximated EMD.

V. CONCLUSION

In this work, we provide a network for lidar point cloud upsampling, which aims to reconstruct the fine-grained lidar scan patterns accurately. Our method reveals the required sensitivity to lidar scan lines and shows convincing results on lidar upsampling on a smaller scale of scan patches. Our deep network utilizes the edge-aware dense convolutions for feature extraction and expansion together with Sliced Wasserstein Distance. Both key features enable our method to perform lidar upsampling in a direct, lightweight one-stage fashion, excluding the coarse and fine reconstruction. The proposed method improves on both qualitatively and quantitatively measured baseline methods with HD, EMD, and SWD metrics. One of the limitations of our approach, though, is the scale of processed samples. The extension of our method to large-scale lidar scans ($\geq 100K$ points) remains for future work.

VI. ACKNOWLEDGMENT

The research leading to these results is funded by the German Federal Ministry for Economic Affairs and Energy within the project KI Delta Learning. The authors would like to thank the consortium for the successful cooperation.

REFERENCES

- [1] P. Achlioptas, O. Diamanti, I. Mitliagkas, and L. J. Guibas. Learning representations and generative models for 3d point clouds. In *ICML*, 2018.
- [2] J. Behley, M. Garbade, A. Milioto, J. Quenzel, S. Behnke, C. Stachniss, and J. Gall. SemanticKITTI: A Dataset for Semantic Scene Understanding of LiDAR Sequences. In *ICCV*, 2019.
- [3] B. Besic, N. Gosala, D. Cattaneo, and A. Valada. Unsupervised domain adaptation for lidar panoptic segmentation. In *RAL*, 2022.
- [4] L. Caccia, H. van Hoof, A. Courville, and J. Pineau. Deep generative modeling of lidar data. In *IROS*, 2019.
- [5] R. Cai, G. Yang, H. Averbuch-Elor, Z. Hao, S. Belongie, N. Snavely, and B. Hariharan. Learning gradient fields for shape generation. In *ECCV*, 2020.
- [6] A. X. Chang, T. Funkhouser, L. Guibas, P. Hanrahan, Q. Huang, Z. Li, S. Savarese, M. Savva, S. Song, H. Su, J. Xiao, L. Yi, and F. Yu. ShapeNet: An Information-Rich 3D Model Repository. In *arxiv*, 2015.
- [7] B. Coors, A. P. Conrache, and A. Geiger. Nova: Learning to see in novel viewpoints and domains. In *3DV*, 2019.
- [8] M. Cuturi. Sinkhorn distances: Lightspeed computation of optimal transportation distances. In *NIPS*, 2013.
- [9] I. Deshpande, Y.-T. Hu, R. Sun, A. Pyrros, N. Siddiqui, S. Koyejo, Z. Zhao, D. Forsyth, and A. Schwing. Max-sliced wasserstein distance and its use for gans. In *CVPR*, 2019.
- [10] S. et al. Scalability in perception for autonomous driving: Waymo open dataset. In *CVPR*, 2020.
- [11] H. Fan, H. Su, and L. Guibas. A point set generation network for 3d object reconstruction from a single image. In *CVPR*, 2017.
- [12] A. Geiger, P. Lenz, and R. Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *CVPR*, 2012.
- [13] T. Groueix, M. Fisher, V. G. Kim, B. Russell, and M. Aubry. Atlasnet: A papier-mâché approach to learning 3d surface generation. In *CVPR*, 2018.
- [14] Y. Jung, S.-W. Seo, and S.-W. Kim. Fast point clouds upsampling with uncertainty quantification for autonomous vehicles. In *ICRA*, 2022.
- [15] S. Kolouri, P. E. Pope, C. E. Martin, and G. K. Rohde. Sliced-wasserstein auto-encoders. In *ICLR*, 2017.
- [16] Y. Kwon, Youngsun and Sung and S.-E. Yoon. Implicit lidar network: lidar super-resolution via interpolation weight prediction. In *ICRA*, 2022.
- [17] C.-L. Li, M. Zaheer, Y. Zhang, B. Póczos, and R. Salakhutdinov. Point cloud gan. In *ICLR*, 2018.
- [18] R. Li, X. Li, C.-W. Fu, D. Cohen-Or, and P.-A. Heng. Pu-gan: a point cloud upsampling adversarial network. In *ICCV*, 2019.
- [19] R. Li, X. Li, P.-A. Heng, and C.-W. Fu. Point cloud upsampling via disentangled refinement. In *CVPR*, 2021.
- [20] M. Liu, L. Sheng, S. Yang, J. Shao, and S.-M. Hu. Morphing and sampling network for dense point cloud completion. In *AAAI*, 2020.
- [21] S. Luo and W. Hu. Diffusion probabilistic models for 3d point cloud generation. In *CVPR*, 2021.
- [22] K. Nguyen, N. Ho, T. Pham, and H. Bui. Distributional sliced-wasserstein and applications to generative modeling. In *ICLR*, 2021.
- [23] T. Nguyen, Q.-H. Pham, T. Le, T. Pham, N. Ho, and B.-S. Hua. Point-set distances for learning representations of 3d point clouds. In *ICCV*, 2021.
- [24] L. Pan. Ecg: Edge-aware point cloud completion with graph convolution. In *IEEE IROS*, 2020.
- [25] C. R. Qi, H. Su, K. Mo, and L. J. Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *CVPR*, 2017.
- [26] C. R. Qi, L. Yi, H. Su, and L. J. Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In *NIPS*, 2017.
- [27] G. Qian, A. Abualshour, G. Li, A. Thabet, and B. Ghanem. Pu-gcn: Point cloud upsampling using graph convolutional networks. In *CVPR*, 2021.
- [28] T. Shan, J. Wang, F. Chen, P. Szenher, and B. Englot. Simulation-based lidar super-resolution for ground vehicles. In *RAS*, 2020.
- [29] S. Shi, X. Wang, and H. Li. Pointcnn: 3d object proposal generation and detection from point cloud. In *CVPR*, 2019.
- [30] D. W. Shu, S. W. Park, and J. Kwon. 3d point cloud generative adversarial network based on tree structured graph convolutions. In *ICCV*, 2019.
- [31] M. Sugiyama and M. Kawanabe. *Machine Learning in Non-Stationary Environments: Introduction to Covariate Shift Adaptation*. The MIT Press, 2012.
- [32] Y. Sun, Y. Wang, Z. Liu, J. E. Siegel, and S. E. Sarma. Pointgrow: Autoregressively learned point cloud generation with self-attention. In *WACV*, 2020.
- [33] L. P. Tchampi, V. Kosaraju, S. H. Rezatofighi, I. Reid, and S. Savarese. Topnet: Structural point cloud deco. In *CVPR*, 2019.
- [34] H. Thomas, C. R. Qi, J.-E. Deschaud, B. Marcotegui, F. Goulette, and L. J. Guibas. Kpconv: Flexible and deformable convolution for point clouds. In *ICCV*, 2019.
- [35] D. Valsesia, G. Fracastoro, and E. Magli. Learning localized generative models for 3d point clouds via graph convolution. In *ICLR*, 2019.
- [36] T. Vayer, R. Flamary, R. Tavenard, L. Chapel, and N. Courty. Sliced gromov-wasserstein. In *NIPS*, 2020.
- [37] X. Wang, M. H. Ang Jr, and G. H. Lee. Cascaded refinement network for point cloud completion with self-supervision. In *CVPR*, 2020.
- [38] Y. Wang, Y. Sun, Z. Liu, S. E. Sarma, M. M. Bronstein, and J. M. Solomon. Dynamic graph cnn for learning on point clouds. In *Transactions on Graphics*, 2019.
- [39] J. Wu, Z. Huang, D. Acharya, W. Li, J. Thoma, D. P. Paudel, and L. Van Gool. Sliced wasserstein generative models. In *CVPR*, 2019.
- [40] G. Yang, X. Huang, Z. Hao, M.-Y. Liu, S. Belongie, and B. Hariharan. Pointflow: 3d point cloud generation with continuous normalizing flows. In *ICCV*, 2019.
- [41] Y. Yang, C. Feng, Y. Shen, and D. Tian. Foldingnet: Point cloud auto-encoder via deep grid deformation. In *CVPR*, 2018.
- [42] W. Yifan, S. Wu, H. Huang, D. Cohen-Or, and O. Sorkine-Hornung. Patch-based progressive 3d point set upsampling. In *CVPR*, 2019.
- [43] T. Yin, X. Zhou, and P. Krähenbühl. Center-based 3d object detection and tracking. In *CVPR*, 2021.
- [44] L. Yu, X. Li, C.-W. Fu, D. Cohen-Or, and P.-A. Heng. Ec-net: an edge-aware point set consolidation network. In *ECCV*, 2018.
- [45] L. Yu, X. Li, C.-W. Fu, and H. P.-A. Cohen-Or. Daniel. Pu-net: Point cloud upsampling network. In *CVPR*, 2018.
- [46] W. Yuan, T. Khot, D. Held, C. Mertz, and M. Hebert. Pcn: Point completion network. In *International Conference on 3D Vision*, 2018.
- [47] W. Zhang, H. Jiang, Z. Yang, S. Yamakawa, K. Shimada, and L. B. Kara. Data-driven upsampling of point clouds. In *CAD*, 2018.