HALS: A Height-Aware Lidar Super-Resolution Framework for Autonomous Driving

Anonymous CVPR submission

Paper ID 12

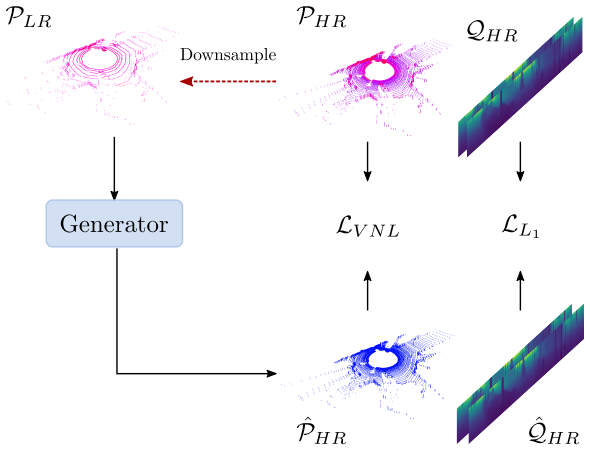

Figure 1. The training pipeline of the proposed model. The generator is trained with an $\mathcal{L}_1$ loss on the range images with polar coordinates and $\mathcal{L}_{VNL}$ on the pointclouds to preserve the structure in 3D.

1. Pipeline

We show the complete training pipeline in Fig. 1.

2. Object Detection Results

To measure how well the shape of foreground objects is preserved during upsampling, we evaluate the performance of an object detection model on the upsampled pointclouds. Specifically, we train Pointpillars [3] to detect cars in high-resolution pointclouds from the Kitti Object dataset [1]. Then, we evaluate the performance of the model on upsampled pointclouds from 3 models: LIDAR-SR [7], ILN [2] and the proposed HALS model. We use the official Kitti evaluation protocol and report the Average Precision (AP) with 40 recall positions at an overlap threshold of 0.7 IoU. The proposed model outperforms both ILN and Lidar-SR on the Easy, Moderate and Hard categories. In Figure 2, we extract cars from the upsampled pointclouds using their ground truth bounding boxes and visualize them. The car from LIDAR-SR is noisy and has few points in the upper part. ILN and SWIN-IR generate more points but their shape is different from the ground truth. For instance, ILN repeats the same line in the lower part of the car, generating a cuboid-like shape. In contrast, HALS is able to better approximate the shape of the car in the dataset.

Model	Easy	Moderate	Hard
LIDAR-SR [7]	44.13	25.05	20.39
ILN [2]	51.93	31.92	26.58
Ours	**55.76**	**34.00**	**27.38**

Table 1. We evaluate a pretrained Pointpillars model [3] on $\times 4$ upsampled pointclouds from the Kitti Object dataset [1] and report the results on the 'Car' class.

3. Ablation Study: Incremental Component Analysis

We build our framework on top of SRResNet [4], a widely used image super-resolution architecture. We incrementally show the effects of the proposed contributions on the Kitti Raw dataset ($4\times$ upsampling rate). In configuration 1 in Tab. 2, we start with a straightforward application of SRResNet on range image with spherical coordinates and L_1 loss only. In configuration 2 and 3, we change the input and output coordinates to cartesian and polar respectively. We already notice a considerable decrease in all 3D metrics. Note that polar coordinates show a higher empirical performance than cartesian coordinates. We hypothesize this happens because it is easier for the network to regress 2 variables (d, z) than 3 (x, y, z). In configuration 4, we add $\mathcal{L}_{VNL}$ and notice a decrease in MAE and an increase in IoU, Precision, Recall and F1-score. Replacing residual blocks with DRBs slightly improves the EMD and CD but the other metrics drop. Finally, we replace the original generator with the proposed height-aware generator, which shows a considerable improvement in the last 4 metrics and an equal or slightly superior performance on the first 4 metrics. Finally, to showcase the improvements

of the HALS generator model design, we add configuration $1^{'}$, where $\mathcal{L}_1$-loss only and spherical coordinates are used. We notice a strong improvement in the IoU metrics and a small improvement in the CD metric, indicating that more accurate shapes are generated.

4. What are the masks focusing on?

We hypothesized that a lower receptive field for the first upsampling branch smaller be beneficial for the higher part of the range image. To confirm the soundness of our hypothesis, we visualize the masks $\mathbf{m}_{shallow}$ and $\mathbf{m}_{deep}$ of a scene in the Nuscenes dataset in Fig. 3. The mask from the shallow branch has higher values (yellow and orange in the heatmap) than the mask from the deep branch in the upper part. This implies more weight is given to the prediction from the shallow branch than the deep branch in this part, as it is more confident in its generated range image. In the lower part of the range image, $\mathbf{m}_{deep}$ has more contribution than $\mathbf{m}_{shallow}$.

5. Qualitative Results

Point-based methods. We show the qualitative results of some point-based upsampling models in Fig. 1. Except for PUNet, the scan pattern cannot be replicated.

Grid-based methods. In Fig. 5, we show qualitative results for upsampled pointclouds from 3 different baselines and our model on the Kitti Object dataset ($\times 4$ upsampling rate). For illustration purposes, we show the front part of the pointcloud. We observe that LIDAR-SR and SWIN-IR generate pointclouds with noisy shapes. Pointclouds from ILN exhibit clusters of high density and other clusters with low density, showing an overall point distribution different from the ground truth. It can also be seen that some lines are generated very close to the existing input lines. Since ILN is an interpolation approach that generates new points using a weighted average of the coordinates of their nearest neighbours, it can become susceptible to artifacts caused by the height-dependent range distribution. In contrast, generated pointclouds from our method have a similar point distribution as the ground truth and objects with more plausible shapes.

References

[1] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012. 1

[2] Youngsun Kwon, Minhyuk Sung, and Sung-Eui Yoon. Implicit lidar network: Lidar super-resolution via interpolation weight prediction. In *2022 International Conference on Robotics and Automation (ICRA)*, pages 8424–8430. IEEE, 2022. 1, 3, 4

[3] Alex H. Lang, Sourabh Vora, Holger Caesar, Lubing Zhou, Jiong Yang, and Oscar Beijbom. Pointpillars: Fast encoders for object detection from point clouds. In *CVPR*, 2019. 1

[4] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Wenzhe Shi. Photo-realistic single image super-resolution using a generative adversarial network. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 105–114, 2017. 1, 3

[5] Ruihui Li, Xianzhi Li, Chi-Wing Fu, Daniel Cohen-Or, and Pheng-Ann Heng. Pu-gan: A point cloud upsampling adversarial network. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 7202–7211, 2019. 3

[6] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1833–1844, 2021. 3, 4

[7] Tixiao Shan, Jinkun Wang, Fanfei Chen, Paul Szenher, and Brendan Englot. Simulation-based lidar super-resolution for ground vehicles. *Robotics and Autonomous Systems*, 134:103647, 12 2020. 1, 3, 4

[8] Huikai Wu and Kaiqi Huang. Point cloud super resolution with adversarial residual graph networks. In *31st British Machine Vision Conference 2020, BMVC*. BMVA Press, 2020. 3

[9] Lequan Yu, Xianzhi Li, Chi-Wing Fu, Daniel Cohen-Or, and Pheng-Ann Heng. Pu-net: Point cloud upsampling network. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2790–2799, 2018. 3

Model	EMD ↓	CD ↓	MAE ↓	RMSE ↓	IoU ↑	Precision ↑	Recall ↑	F1-score ↑
1- Baseline [4]	101	0.052	0.186	0.86	0.393	0.564	0.564	0.564
1′- HALS generator + spherical	100.8	0.047	0.18	0.903	0.465	0.635	0.633	0.63
2- Baseline + cartesian	90.2	0.031	0.224	0.85	0.268	0.421	0.423	0.422
3- Baseline + polar	84.7	0.020	0.186	**0.84**	0.443	0.614	0.614	0.614
4- Baseline + polar + $\mathcal{L}_{VNL}$	82.5	0.018	0.172	**0.84**	0.478	0.645	0.648	0.647
5- Baseline w/ DRB + polar + $\mathcal{L}_{VNL}$	**81.4**	0.016	0.174	0.88	0.436	0.608	0.606	0.607
6- Height-aware generator + polar + $\mathcal{L}_{VNL}$	82.0	**0.015**	**0.171**	0.88	**0.510**	**0.672**	**0.671**	**0.671**

Table 2. Ablation study of our model performed on the KITTI Raw Dataset, with $\times 4$ upsampling rate (output resolution = 40×256).

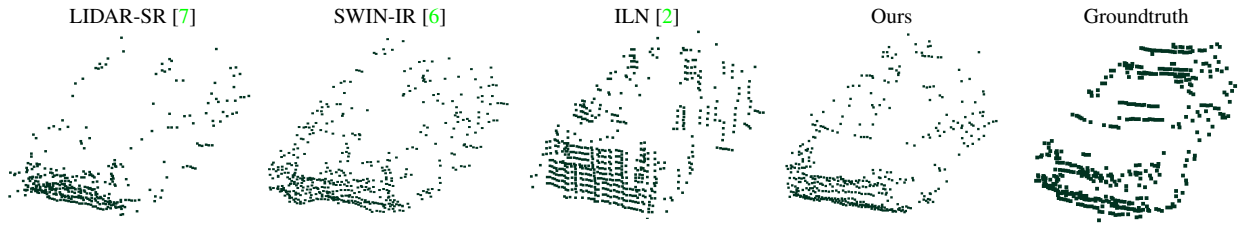

Figure 2. Qualitative comparison on Cityscapes dataset

Figure 3. Visualization of the masks produced by the HALS generator (Nuscenes dataset). *Left*: mask from shallow branch. *Right*: mask from the deep branch. Brighter colors indicate higher values (close to 1).

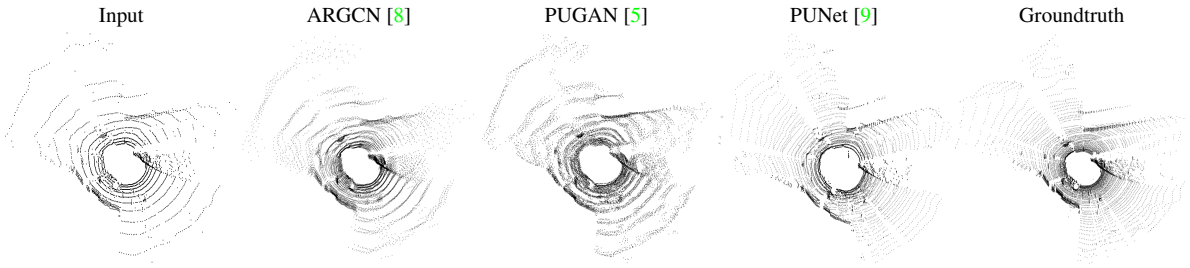

Figure 4. Generated lidar scans from point-based models. Only PUNet is able to replicate the lidar scan pattern.

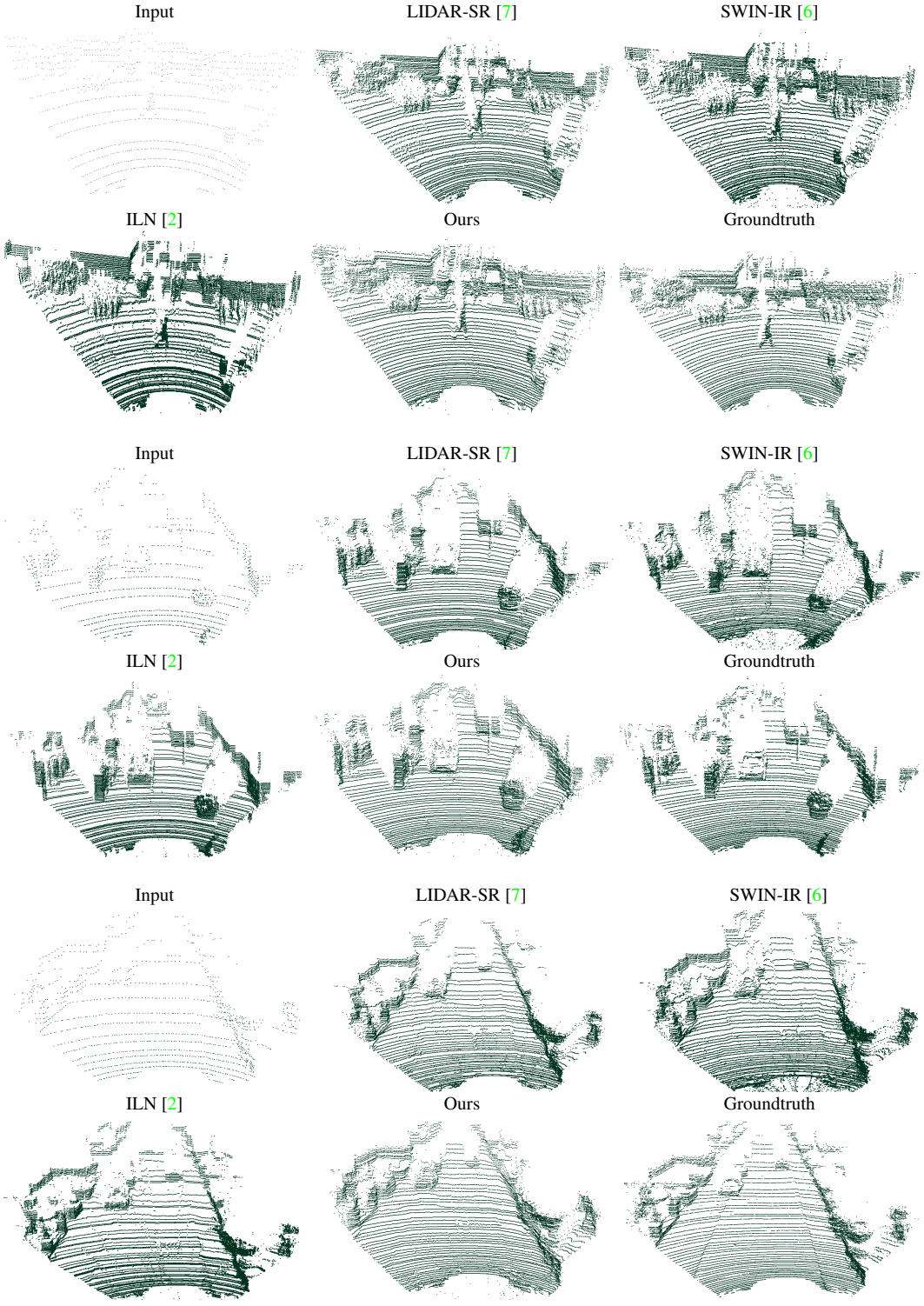

Figure 5. Qualitative comparison on Cityscapes dataset