Supplementary Material
Project to Adapt: Domain Adaptation for Depth Completion from Noisy and Sparse Sensor Data

Adrian Lopez-Rodriguez[1] [0000−0003−3984−5126], Benjamin Busam[2,3] [0000−0002−0620−5774], and Krystian Mikolajczyk[1] [0000−0003−0726−9187]

[1] Imperial College London
{al4415, k.mikolajczyk}@imperial.ac.uk
[2] Huawei Noah's Ark Lab
[3] Technical University of Munich
b.busam@tum.de

Table 1: Results when varying the number of LiDAR masks used for sparsifying the synthetic depth during the first step of training. RMSE and MAE are reported in mm, and iRMSE and iMAE in 1/km.

Model	RMSE	MAE	iRMSE	iMAE
Only 1 mask	1273.77	316.22	5.26	1.37
All masks	1247.53	308.08	4.54	1.34

1 Number of Sampled LiDAR Masks

We aim to test the effect on performance on the number of binary LiDAR masks used. To that end, we sampled one mask at random from KITTI and applied the same mask to all the synthetic depth images during training. Table 1 shows the performance obtained for our method after the first step of training, i.e., using only synthetic data supervision, when using all available masks in KITTI and when using only one. Table 1 shows that the effect on performance is minimal, where the increase of RMSE is of 2.1% when using the same mask to sparsify the synthetic dense depth compared to using all available masks in KITTI.

2 Sensitivity to Filter Parameters

During the second step, the filter parameters, *i.e.*, the window size w_p and the object thickness θ affect the performance obtained after training. Figure 1 shows the result in RMSE when varying both terms. Due to the computational

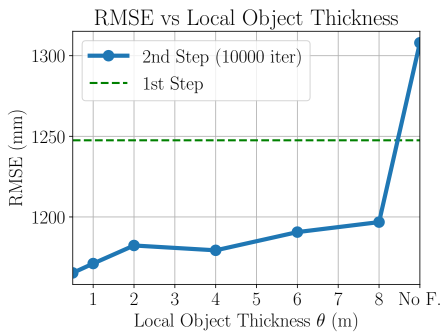

Fig. 1: RMSE obtained in the second step of training when varying the parameters of the filter used to obtain the reliable points $\mathcal{S}_p$. *Left image* shows the RMSE when varying the object thickness θ, where the window size is fixed to $w_p = 16$ pixels. *Right image* shows the RMSE when varying the window size w_p, where the object thickness is fixed to $\theta = 0.5$ meters. *1st Step* refers to the performance obtained after the first step of training, *i.e.*, when no supervision from the real sparse data is used. *No F.* refers to not using any filter in the real sparse depth used for supervision during the second step of training.

complexity of training the model for several different conditions, we plot the results after 10000 iterations of training instead of the 40000 used for the main experiments. However, we did notice that when the values of the filter produce a higher noise rate η compared to the values we use in our main experiments, i.e. $\theta = 0.5$ m and $w_p = 16$ pixels, the performance after 40000 iterations is decreased compared to the performance after 10000 iterations. Furthermore, the performance after 40000 iterations of training is also lower than the performance after only the first step of training. Hence, a smaller number of training iterations is a safer choice to avoid overfitting to the noisy points remaining after the filtering step.

Object Thickness θ. Figure 1 shows the RMSE after the second step of training when varying the object thickness θ and fixing the window size to $w_p = 16$ pixels. We can see that after 0.5 m, which is the value used for our main experiments, the RMSE starts to increase, which resembles the trend of the noise rate η. In all cases, after 10000 training iterations, the performance is better than after the first step of training or than using no filter for the input data supervision.

Window Size w_p. Figure 1 also shows the RMSE after the second step of training when varying the window size w_p and setting the object thickness $\theta = 0.5$ m. The findings are similar to the findings for the object thickness. Thus, the trend resembles the trend for the noise rate η. Additionally, after 10000 training iterations for all window size w_p values the model performs better than both when using no supervision from the real sparse depth, *i.e.*, *First Step*, and when using supervision from the input points with no filter, *i.e.*, *No F.*.