A Low-Footprint Quantized Neural Network for Depth Completion of Very Sparse Time-of-Flight Depth Maps
Supplementary Material

Xiaowen Jiang[1] Valerio Cambareri[2*] Gianluca Agresti[3]
Cynthia Ifeyinwa Ugwu[4] Adriano Simonetto[4] Fabien Cardinaux[3] Pietro Zanuttigh[4]
[1] EPFL, Switzerland [2] Sony Depthsensing Solutions NV, Belgium
[3] Sony Europe B.V., R&D Center, Stuttgart Laboratory 1, Germany [4] University of Padova, Italy

1. Sparse ToF Datasets

We report an enlarged version of the pictures in the manuscript detailing our sparse ToF datasets in Fig. 1. In SDS-ST1k, we simulate the projection of the dot pattern via a raytraced light shading profile defined parametrically and emulating a commercial VCSEL illuminator. We then receive the simulated sensor plane irradiance, and generate the ToF sensor pixel response. The rays corresponding to dot center locations on the sensor are annotated, and the sparse depth map retrieved accordingly. We can observe parallax effects on the dot pattern due to optically accurate simulation, *e.g.*, on the wooden beams to the left of the field of view in Fig. 1c. As for NYU-Depth v2, we process the depth maps by *masking* with a dot pattern that does not account for scene depth (*i.e.*, it is not projected on the scene, but generated by assuming a default arbitrarily far plane). A sample from the KITTI dataset with Velodyne LiDAR overlay is also reported to compare visually the dot pattern density; there we can also observe that the sparse depth samples are not equally distributed over the RGB frame (the top part of which typically does not yield meaningful predictions as it includes, *e.g.*, the sky).

2. Loss Function

We here report an extended study of the choice of loss function parameter λ_n for the loss $\mathcal{L}(\mathcal{D})$ we utilize in the manuscript. The curves we report are validation loss curves of the scale-dependent term $\mathcal{L}_{\ell_1}$ and scale-independent term $\mathcal{L}_\mathrm{n}$ as the epochs increase. As we may observe in Fig. 2 high values of λ_n overly promote normals similarity at the cost of higher depth error in the ℓ_1 sense; conversely, for lower values of the hyperparameter we attain low depth error at acceptable mean normals similarity values. We therefore find our optimum, $\lambda_\mathrm{n}^\mathrm{opt} = 10^{-3}$, to be the best trade-off between those shown in this analysis.

3. Qualitative Evaluation

We here extend the qualitative evaluation by reporting more images from the SDS-ST1k (Fig. 3) and NYU-Depth v2 (Fig. 4) datasets, as processed by the reference methods in the main paper. These corroborate the evidence in the manuscript on the quality of our $\mathrm{float32}$ and mixed precision $\mathrm{W}_4\mathrm{A}_8$ models against NLSPN.

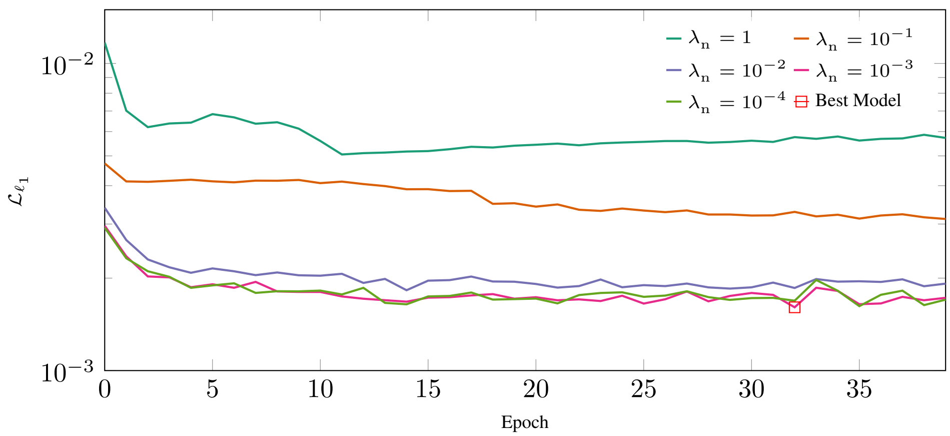
*Corresponding author: valerio.cambareri@sony.com.

(a) $1216 \times 352, K \approx 4.4\%$

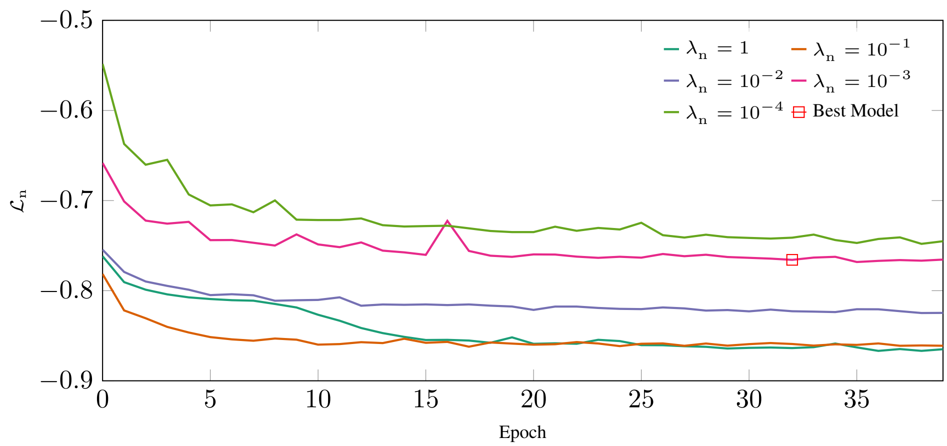
(b) $304 \times 224, K \approx 1.4\%$

(c) $640 \times 480, K \approx 0.4\%$

Figure 1. RGB-D overlay of (a) KITTI (LiDAR; range: $[0, 85]m$), (b) NYU-Depth v2 (processed; range: $[0, 10]m$), (c) SDS-ST1k (sparse ToF; range: $[0, 15]m$). Depth overlay in "magma" colormap. Figure best viewed in color.



(a) Scale-dependent loss term



(b) Scale-independent loss term

Figure 2. Loss Function Tuning on SDS-ST1k. We report the validation loss terms as λ_n varies: (a) scale-dependent term $\mathcal{L}_{\ell_1}$, (b) scale-independent term $\mathcal{L}_n$. Figure best viewed in color.

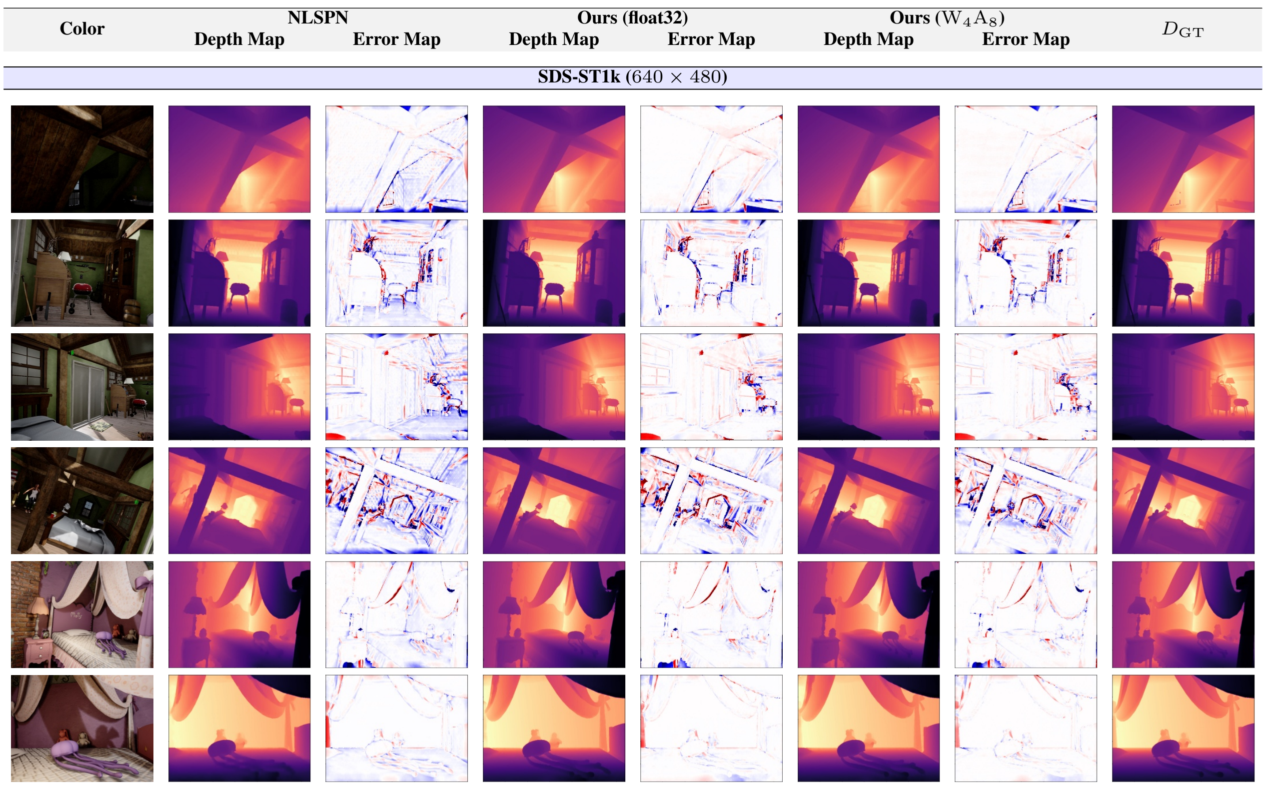


Figure 3. Qualitative Results. We report, for several arbitrary frames in the test set of SDS-ST1k, the predicted depth maps $\hat{D}$ and error maps $\hat{D} - D_{\mathrm{GT}}$ (range: $[-500, 500]mm$). Figure best viewed in color.

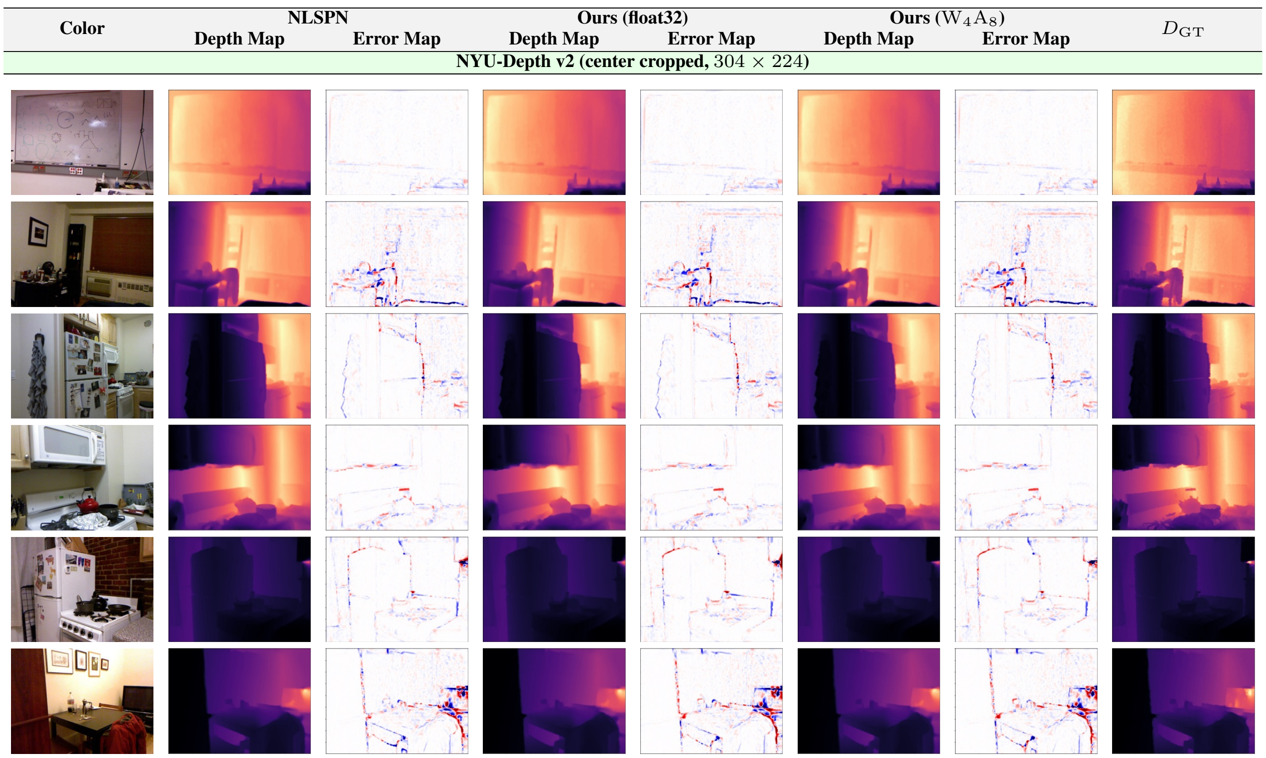


Figure 4. Qualitative Results. We report, for several arbitrary frames in the test set of NYU-Depth v2, the predicted depth maps $\hat{D}$ and error maps $\hat{D} - D_{\mathrm{GT}}$ (range: $[-500, 500]mm$). Figure best viewed in color.