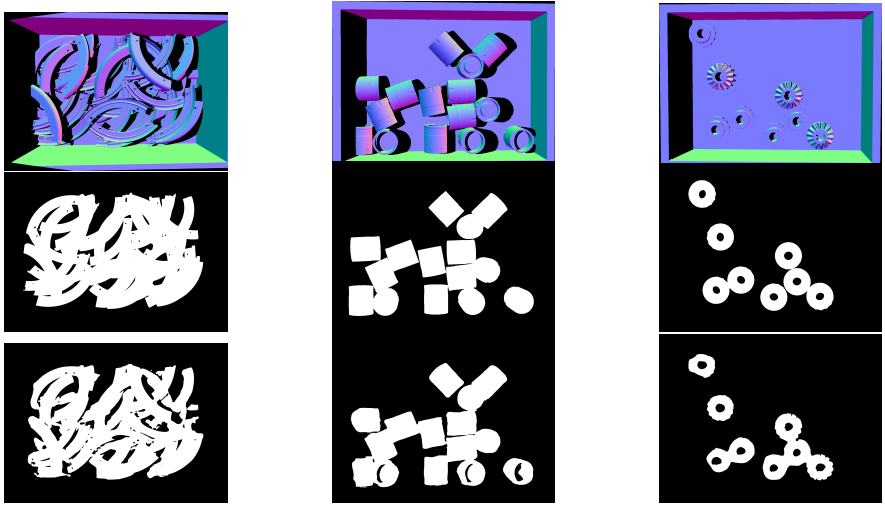
Segmentations!

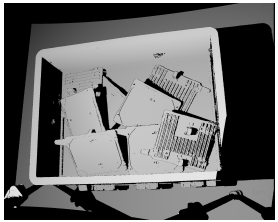
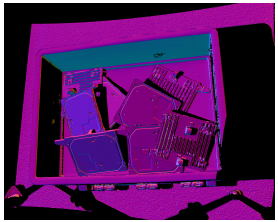
Bc. Lukáš Gajdošech

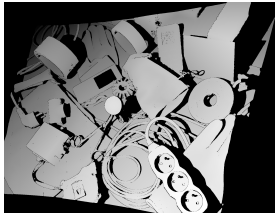
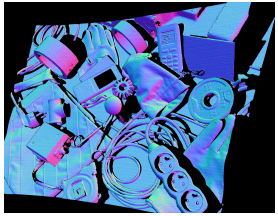
28.10.2020 - 10.11.2020

Synthetic Data



Real Data

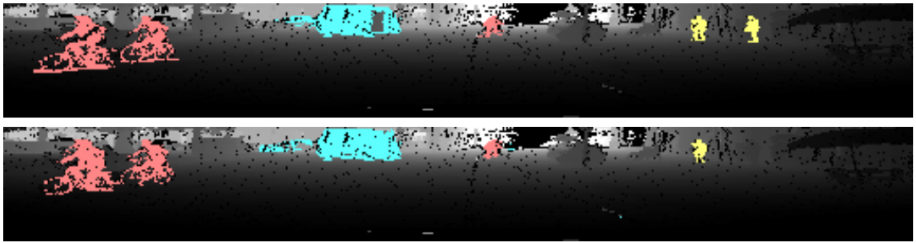




LU-Net

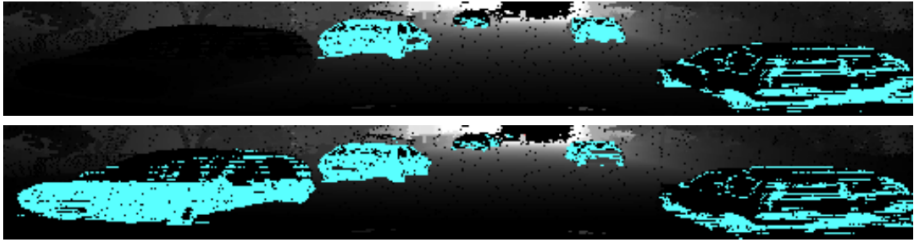
... range-image methods succeed in real-time computation ...

We propose LU-Net: an end-toend model for the accurate semantic segmentation of point clouds represented as range-images. We will show that it outperforms all other range-image methods by a large margin on the KITTI dataset, while offering a robust methodology for bridging between 3D LiDAR point cloud processing and 2D image processing.

LU-Net - Qualitative Results



(d)



LU-Net - Quantitative Results

		Cars	Pedestrians	Cyclists	Average
SqueezeSeg	[20]	64.6	21.8	25.1	37.2
PointSeg	[19]	67.4	19.2	32.7	39.8
RIU-Net	[2]	62.5	22.5	36.8	40.6
SqueezeSegv2	[21]	73.2	27.8	33.6	44.9
LU-Net		72.7	**46.9**	**46.5**	**55.4**

$$\text{Metric: } IoU_l = \frac{|p_l \cap G_l|}{|p_l \cup G_l|}$$

where p_l and G_l denote the predicted and groundtruth sets of points that belongs to label l respectively.

LU-Net - Weighted Focal Loss Function

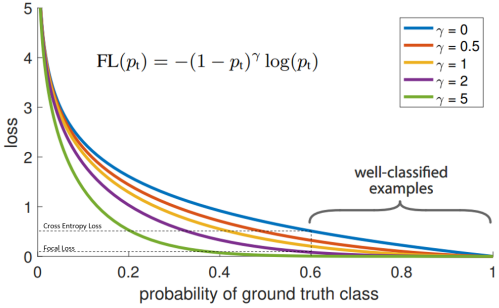
$$s_{gt(x)}(x) = \begin{cases} p(x) \text{ if } gt(x) \equiv 1 \\ 1 - p(x) \text{ else} \end{cases}$$

$$\text{BWFL} = -w(x)(1 - s_{gt(x)}(x))^{\gamma} log(s_{gt(x)})$$

$$\begin{aligned} \text{Standard BCE} &= -gt(x)log(p(x)) + (1 - gt(x))log(1 - px(x)) \\ &= (1 - s_{gt(x)}(x))log(s_{gt(x)}) \end{aligned}$$

LU-Net - Loss Function Comparison

	Cars	Pedestrians	Cyclists	Average
LU-Net w/o relative	62.8	39.6	37.5	46.6
LU-Net w/o FL	**73.8**	42.7	32.9	49.8
LU-Net	72.7	**46.9**	**46.5**	**55.4**



3D Fully Convolutional Network for Vehicle Detection in Point Cloud

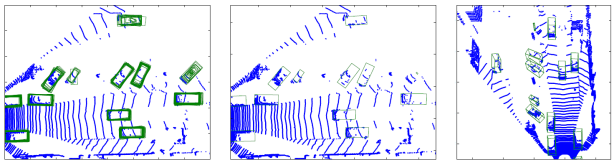


Fig. 3. Intermediate results of the 3D FCN detection procedure. (a) Bounding box predictions are collected from regions with high objectness confidence and are plotted as green boxes. (b) Bounding boxes after clustering plotted with the blue original point cloud. (c) Detection in 3D since (a) and (b) are visualized in the bird's eye view.

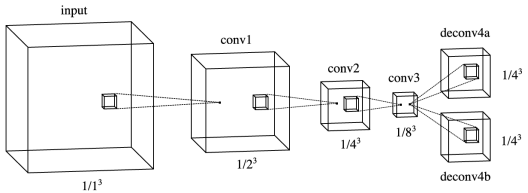


Fig. 2. A sample illustration of the 3D FCN structure used in this paper. Feature maps are first down-sampled by three convolution operation with the stride of $1/2^3$ and then up-samped by the deconvolution operation of the same stride. The output objectness map ($\mathbf{o}^a$) and bounding box map ($\mathbf{o}^b$) are collected from the deconv4a and deconv4b layers respectively.