Processing of 3D Scans using Machine Learning

Author: Bc. Lukáš Gajdošech
Supervisor: RNDr. Martin Madaras, PhD.

Master's Thesis Website

December 17, 2020

Obsah

I/O Definition

Input

Organized point cloud (single-view). The dimensions of the scan are not fixed. For every point, we know the following:

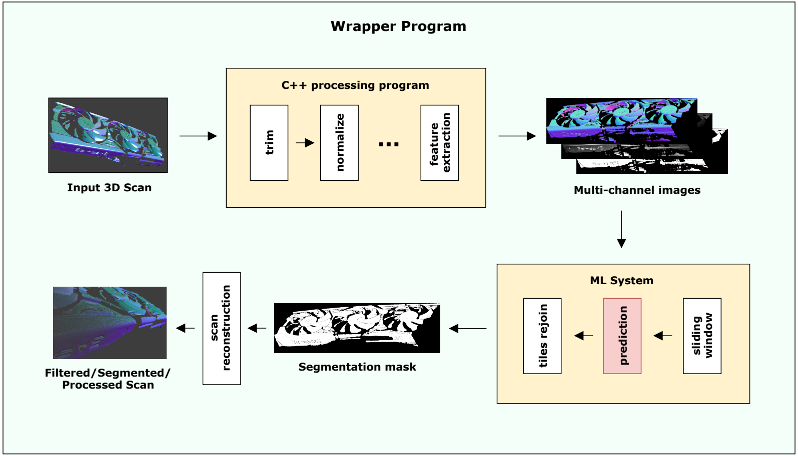
- position, intensity, normal

Output

Processed point cloud for 3 different tasks (determined by the availability of data and CNN solvability):

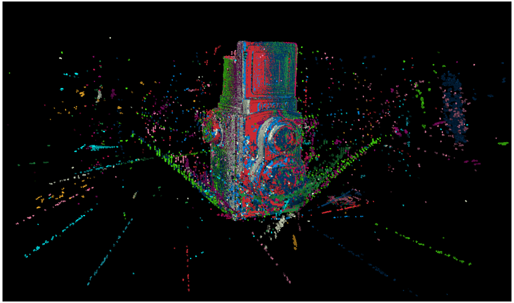
- cleaned artefacts
- localization of parts
- segmentation of parts

Modular Pipeline



We have decided to use an CNN model with similar architecture for the chosen 3 tasks.

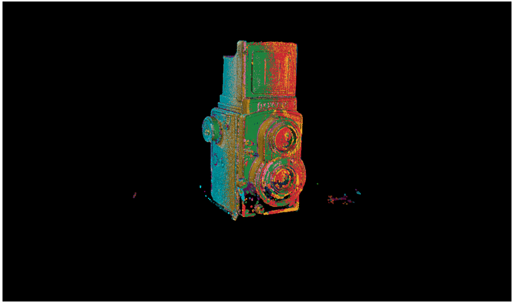
Artefact Filtering I



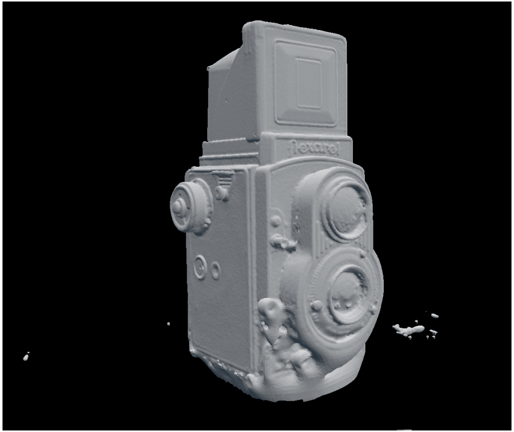
(a) raw point clouds

(b) reconstruction of uncleaned data

Artefact Filtering II



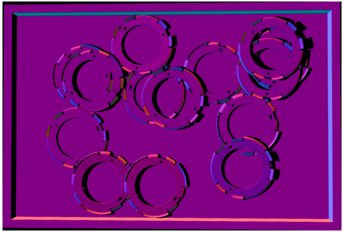
(a) filtered point clouds



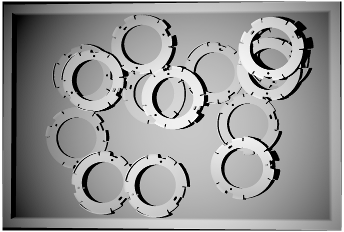
(b) reconstruction of filtered data

Part Localization

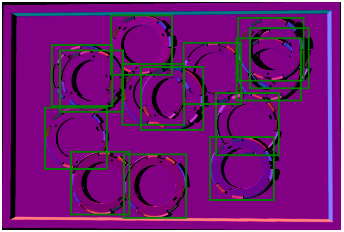
We have access to an in-house utility `BinGenerator`. We can generate synthetic scans, simulating the physics of parts and structured light scanner.



(a) normal map

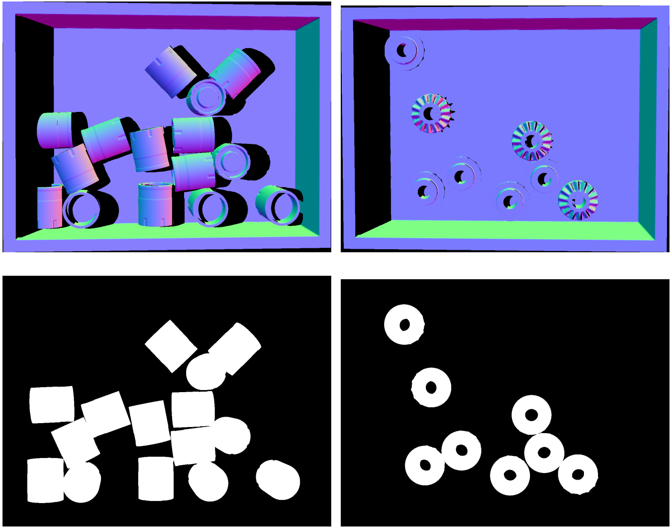


(b) depth map



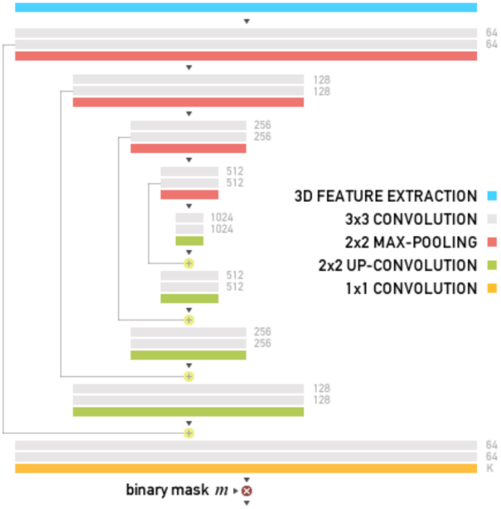
(c) GT localization

Parts Segmentation

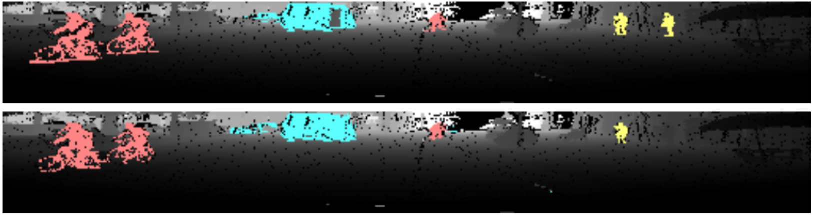


Obsah

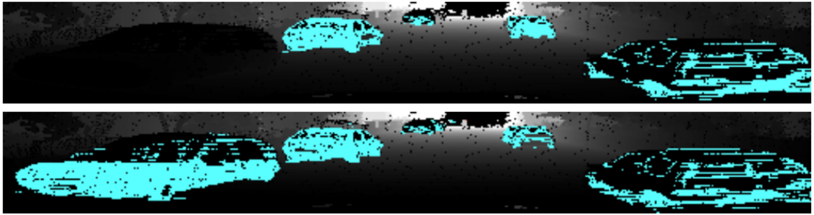
LU-Net I


64
64
128
128
256
256
512
512
1024
1024
512
512
256
256
128
128
64
64
K
3D FEATURE EXTRACTION
3x3 CONVOLUTION
2x2 MAX-POOLING
2x2 UP-CONVOLUTION
1x1 CONVOLUTION
binary mask m

LU-Net II



(d)



LU-Net III

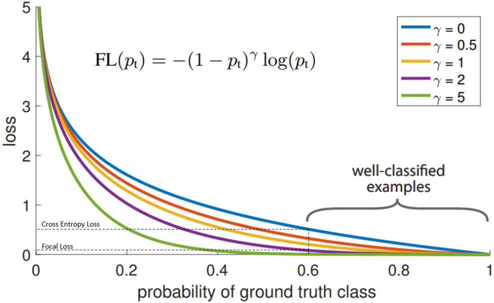
$$s_{gt(x)}(x) = \begin{cases} p(x) \text{ if } gt(x) \equiv 1 \\ 1 - p(x) \text{ else} \end{cases}$$

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

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

LU-Net IV

	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**



YOLO I

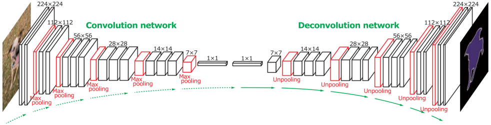


Figure: output: WxH segmentation map

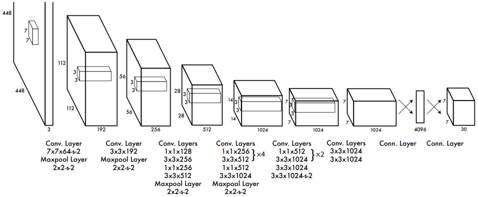
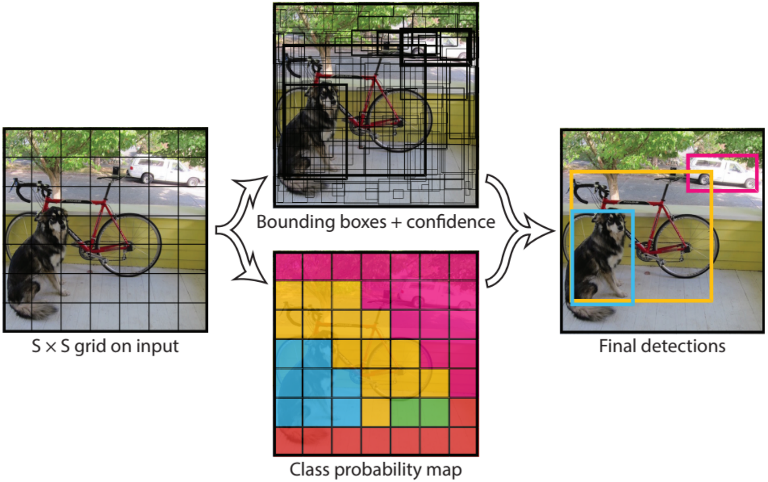


Figure: output: $SxSx(B*5+C)$ tensor

YOLO II


Bounding boxes + confidence
S × S grid on input
Final detections
Class probability map

YOLO III

Our system divides the input image into an $S \times S$ grid. If the center of an object falls into a grid cell, that grid cell is responsible for detecting that object.

Each grid cell predicts B bounding boxes and confidence scores for those boxes. These confidence scores reflect how confident the model is that the box contains an object and also how accurate it thinks the box is that it predicts. Formally we define confidence as $\Pr(\text{Object}) * \text{IOU}_{\text{pred}}^{\text{truth}}$. If no object exists in that cell, the confidence scores should be zero. Otherwise we want the confidence score to equal the intersection over union (IOU) between the predicted box and the ground truth.

Each bounding box consists of 5 predictions: x, y, w, h,

YOLO IV

loss function:

$$\lambda_{\textbf{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right]$$

$$+ \lambda_{\textbf{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left[\left(\sqrt{w_i} - \sqrt{\hat{w}_i} \right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i} \right)^2 \right]$$

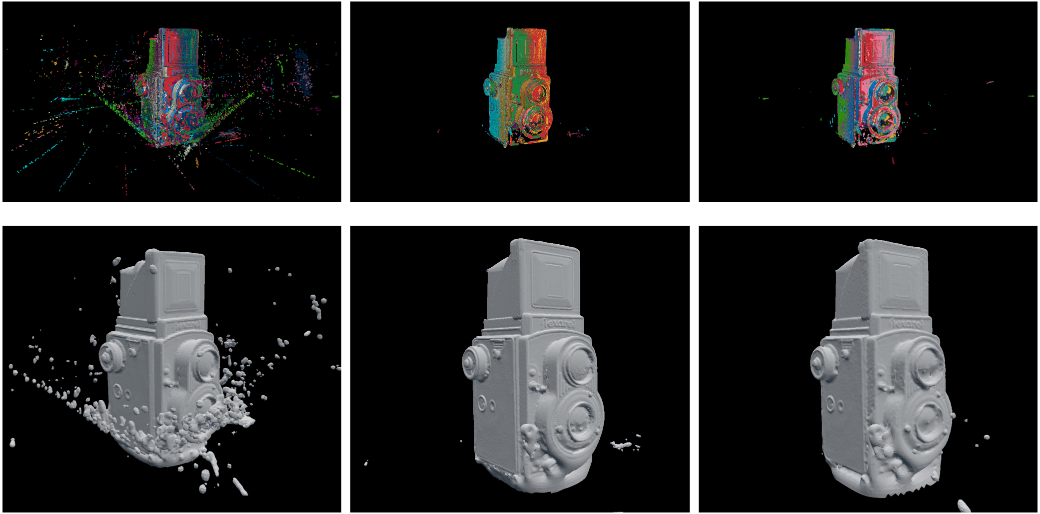
$$+ \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left(C_i - \hat{C}_i \right)^2$$

$$+ \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{noobj}} \left(C_i - \hat{C}_i \right)^2$$

$$+ \sum_{i=0}^{S^2} \mathbb{1}_{i}^{\text{obj}} \sum_{c \in \text{classes}} (p_i(c) - \hat{p}_i(c))^2 \qquad (3)$$

Obsah

Qualitative Results - Filtration



Quantitative Results - Filtration

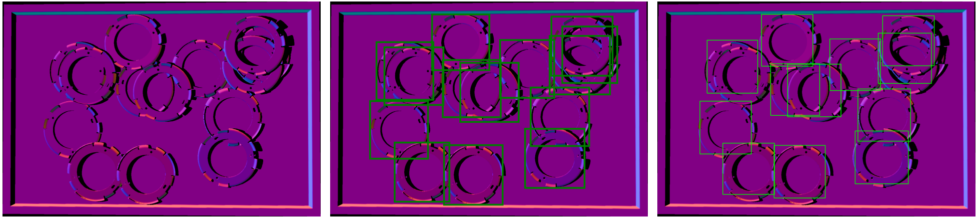
Dataset = Flexaret
# scans = 19
Analytic (s) = 4.32

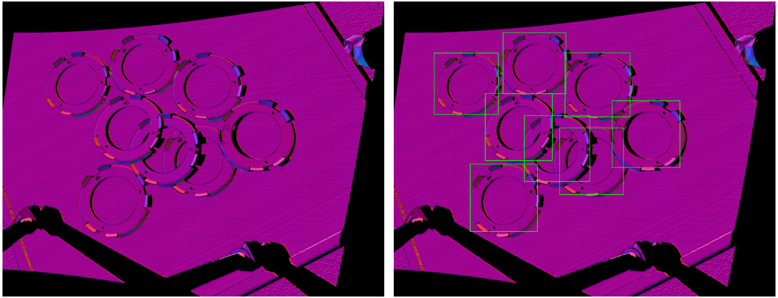
$$IOU = J(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$

(min, max, avg) IoU = (0.4313, 0.942, 0.768)

GPU	API	backend	$FCNN_s$ (s)	$FCNN_t$ (s)
Vega64	OpenCL	Plaid-ML	∼22	∼220
Vega64	OpenCL	TF (ROCm)	∼8	∼80
RTX2060	CUDA	TensorFlow	4.35	45.46

Qualitative Results - Localization





Qualitative Results - Segmentation = TO-DO