Learning for Active 3D Mapping

Karel Zimmermann, Tomáš Petříček, Vojtěch Šalanský, Tomáš Svoboda

http://cmp.felk.cvut.cz/~zimmerk/

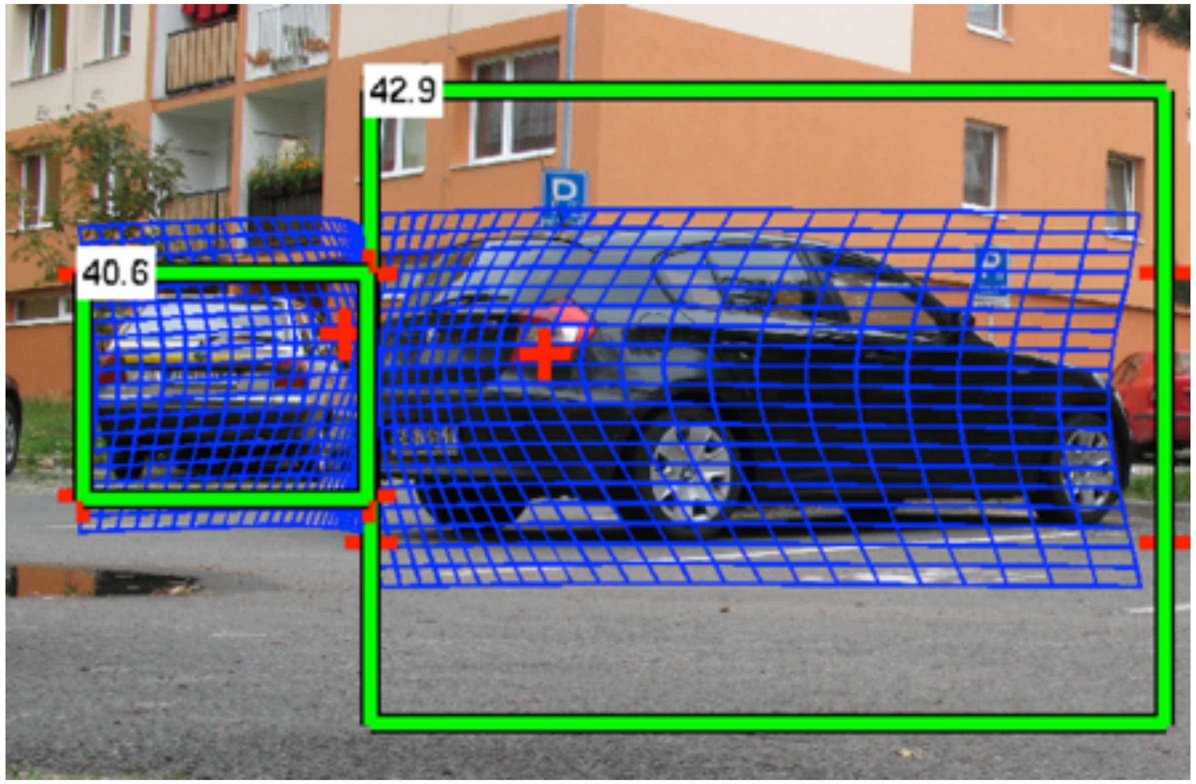
oral at ICCV 2017

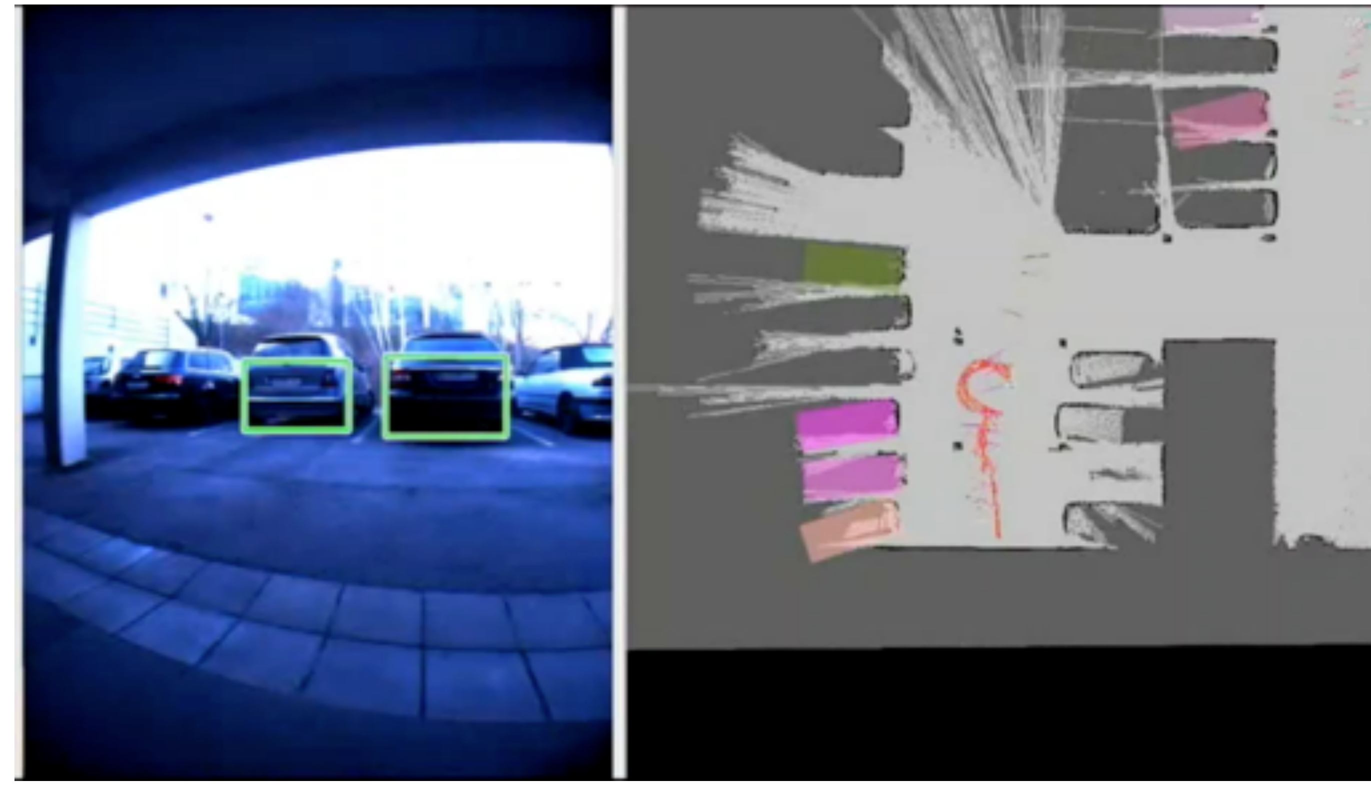
Vision for Robotics and Autonomous Systems
https://cyber.felk.cvut.cz/vras/

Center for Machine Perception
https://cmp.felk.cvut.cz

Department for cybernetics
Faculty of Electrical Engineering
Czech Technical University in Prague

Object detection and tracking





[1] K.Zimmermann, D.Hurych, T.Svoboda, Non-Rigid *Object Detection with Local Interleaved Sequential Alignment (LISA),* **TPAMI (IF=5)**, 2014

[2] K.Zimmermann, J.Matas, T.Svoboda, *Tracking by an Optimal Sequence of Linear Predictors*,**TPAMI (IF=5)**, 2009.

Motion and compliance control of flippers

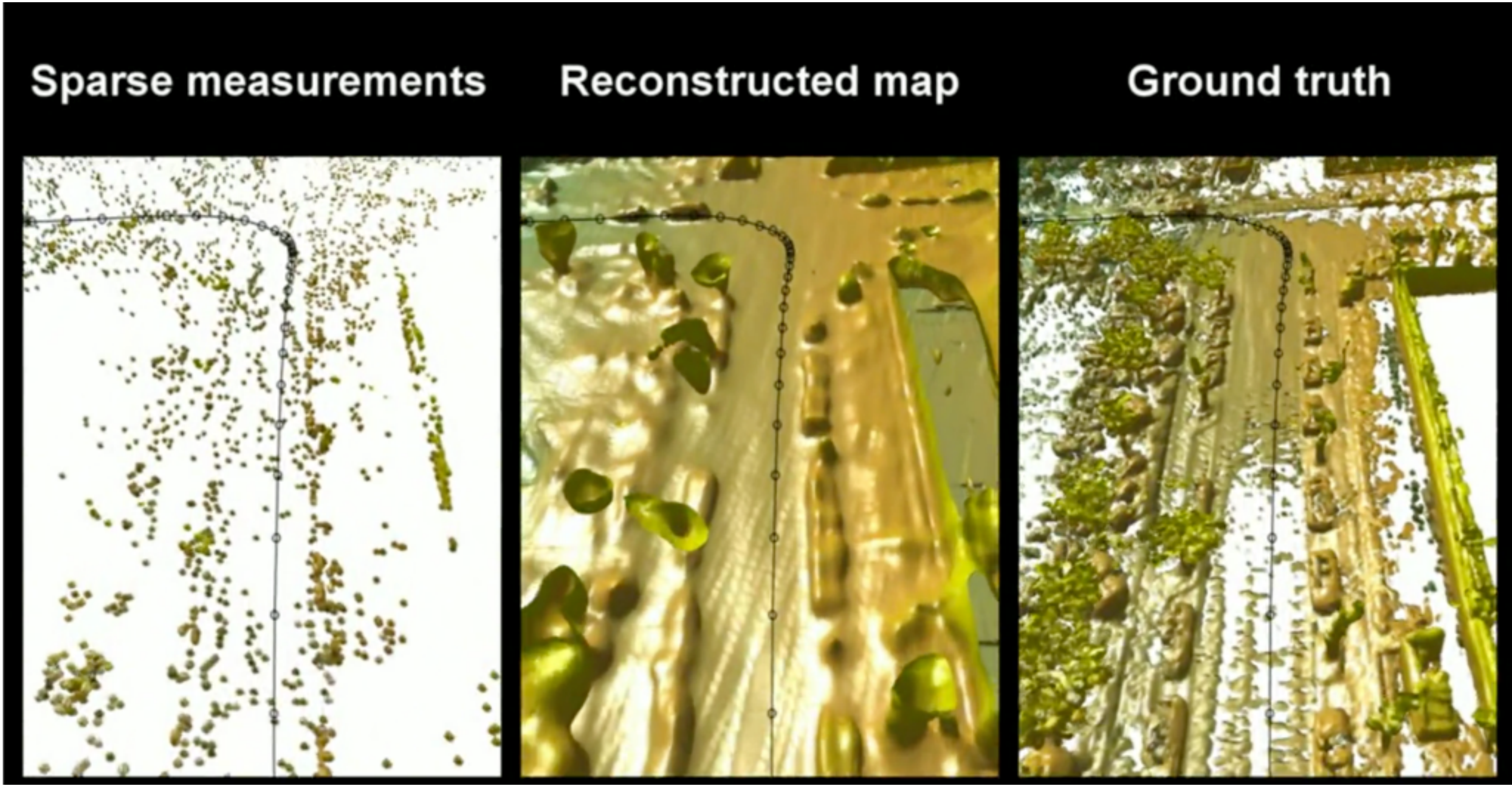
[3] Pecka, Zimmermann, Reinstein,et al. **IEEE TIE (IF=6)**, 2017

Traffic sign detection and 3D localization

1.5 year PostDoc in Luc van Gool's lab at Katholieke Universiteit Leuven

[3] R.Timofte, K.Zimmermann, Luc van Gool, Multi-view traffic sign detection, recognition, and 3D localisation, **MVA (IF=1.5), 2011**

Today's topic

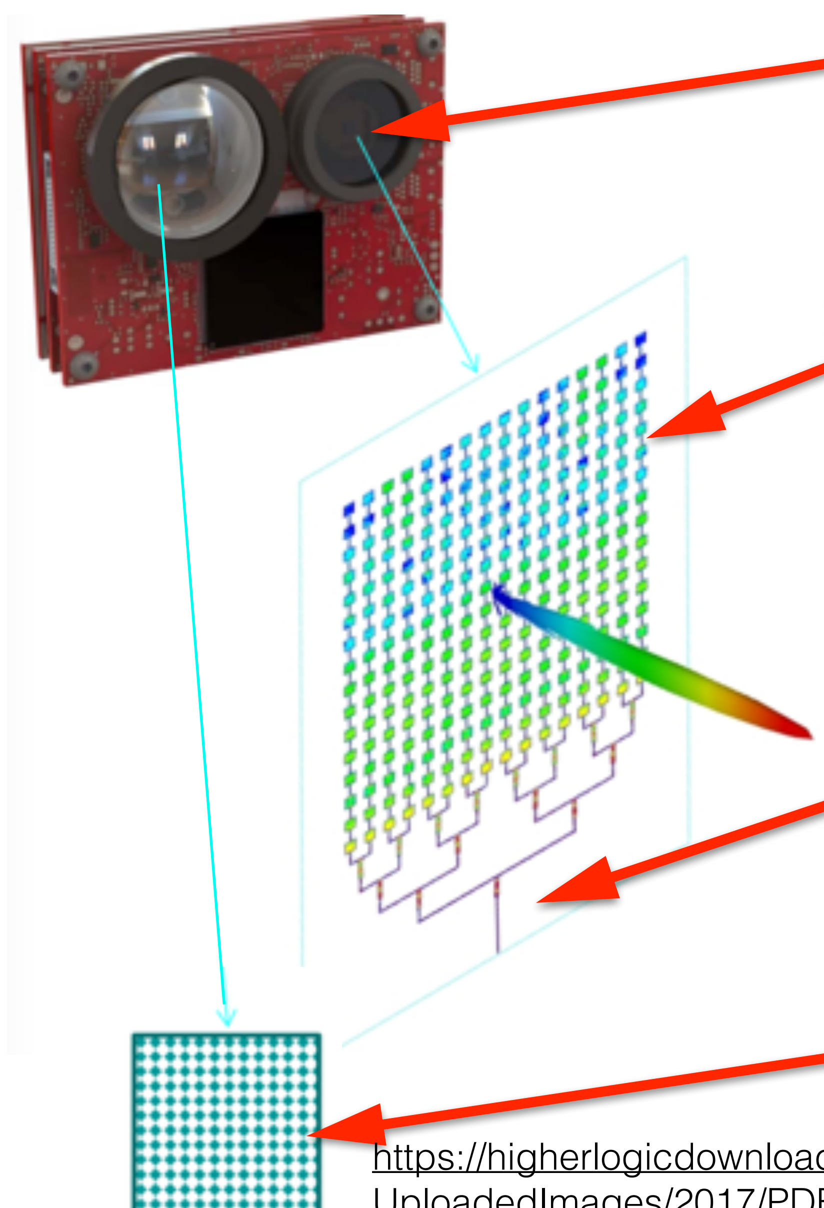


[5] Zimmermann, Petricek, Salansky, Svoboda, Learning for Active 3D Mapping, **ICCV oral**, 2017 https://arxiv.org/abs/1708.02074

Motivation

- **Motivation:** New Solid State Lidars will allow independent steering of depth-measuring rays

QUANERGY S3 principle



Emitted laser beams

Transmitted through Optical Phased Array

Controlling optical properties of elements, allows to steer laser beams in desired directions

Reflected laser beams are captured by SPAD array

https://higherlogicdownload.s3.amazonaws.com/AUVSI/14c12c18-fde1-4c1d-8548-035ad166c766/UploadedImages/2017/PDFs/Proceedings/ESS/Wednesday%201330-1400_Louay%20Eldada.pdf

Problem definition

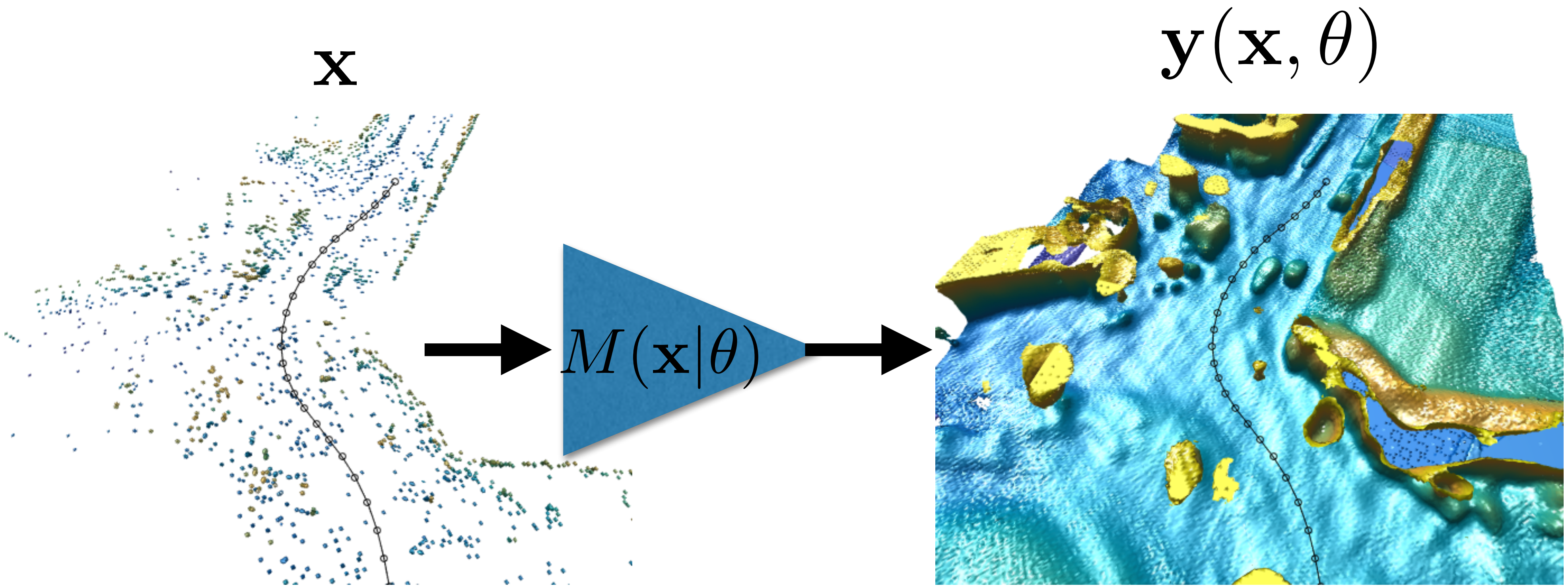
- Steerable SSL is not yet avaiable
- Simulation of SSL on Kitti dataset.

Goal:

1. Learn to reconstruct dense 3D voxel map from sparse depth measurements
2. Optimize reactive control of depth-measuring rays along an expected vehicle trajectory

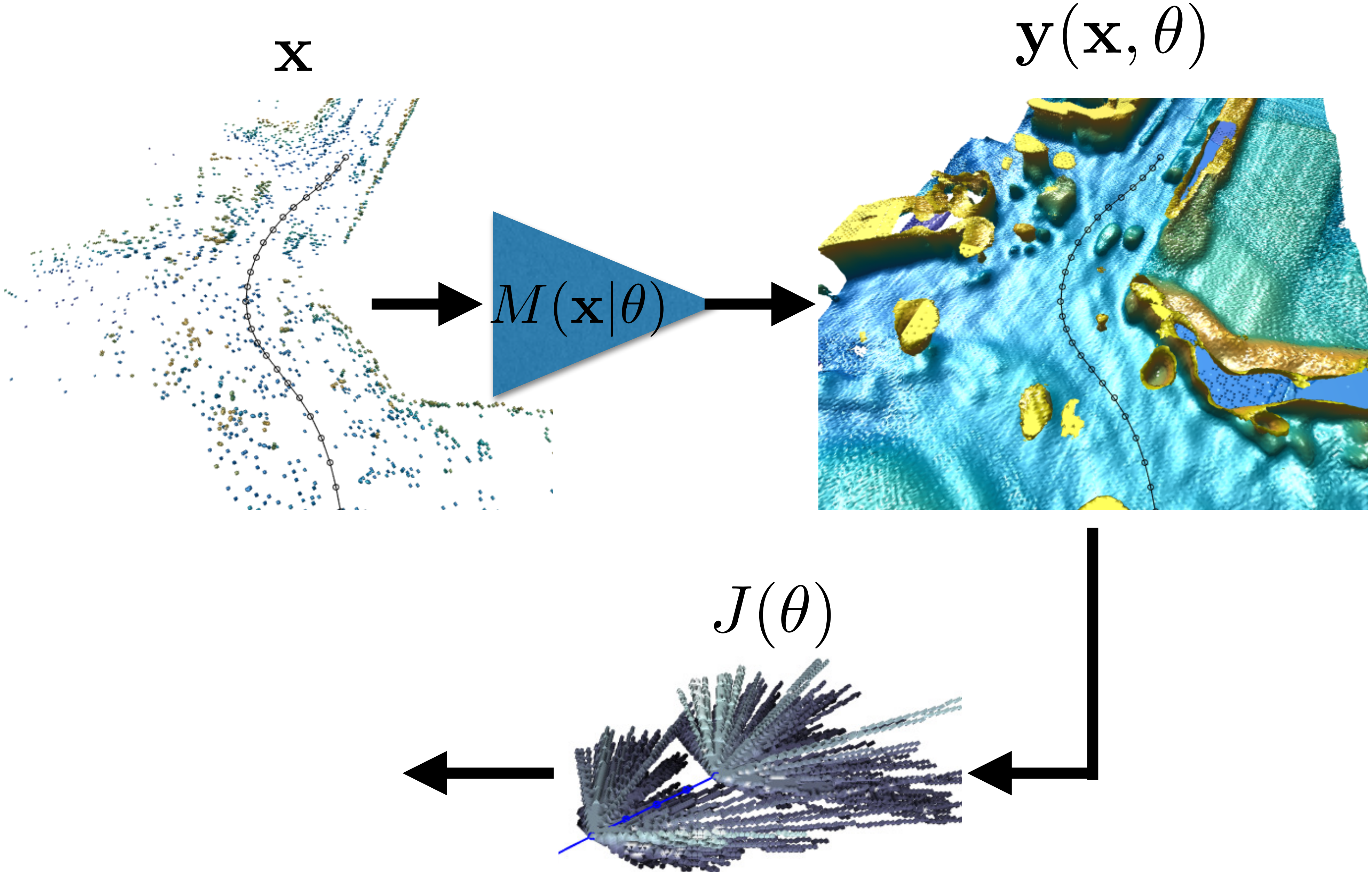
Overview of active 3D mapping

- Learning of 3D mapping network ………………… $M(\mathbf{x}|\theta)$



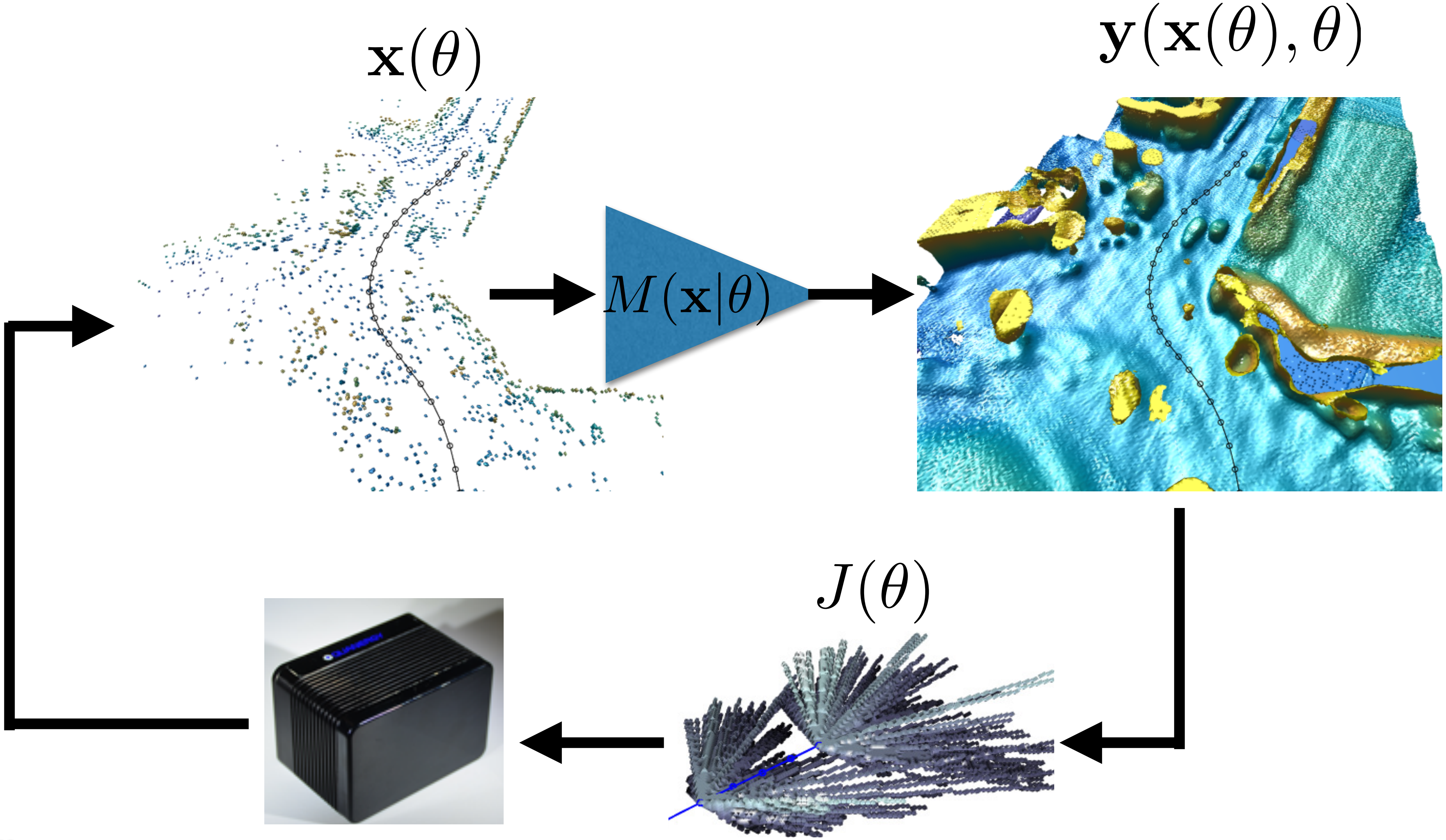
Overview of active 3D mapping

- Learning of 3D mapping network .... $M(\mathbf{x}|\theta)$
- Planning of depth measuring rays .... $J(\theta)$



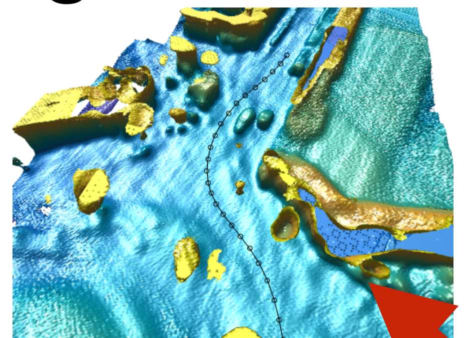
Overview of active 3D mapping

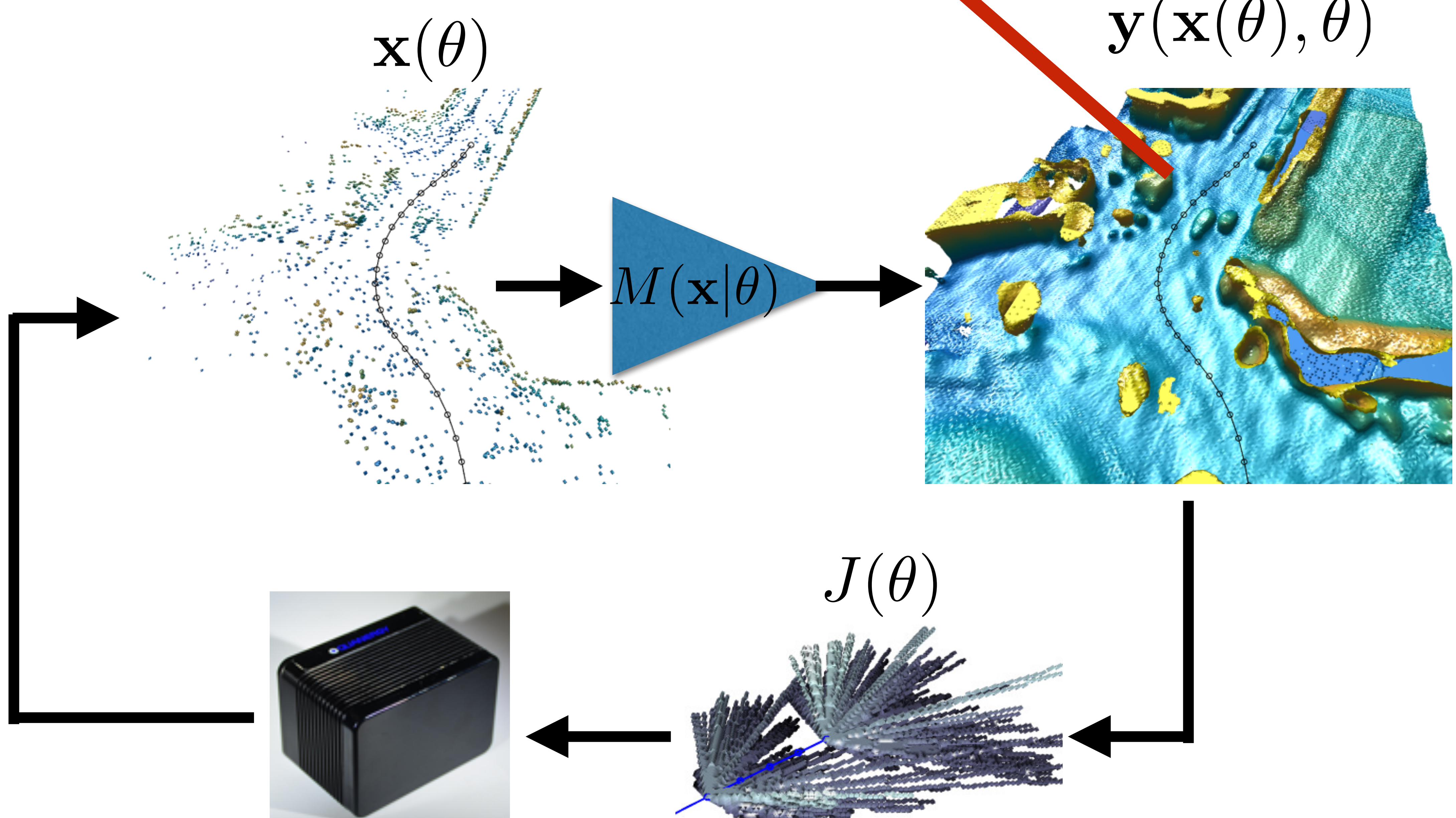
- Learning of 3D mapping network .... $M(\mathbf{x}|\theta)$
- Planning of depth measuring rays .... $J(\theta)$
  which provides following sparse measurement .... $\mathbf{x}(\theta)$



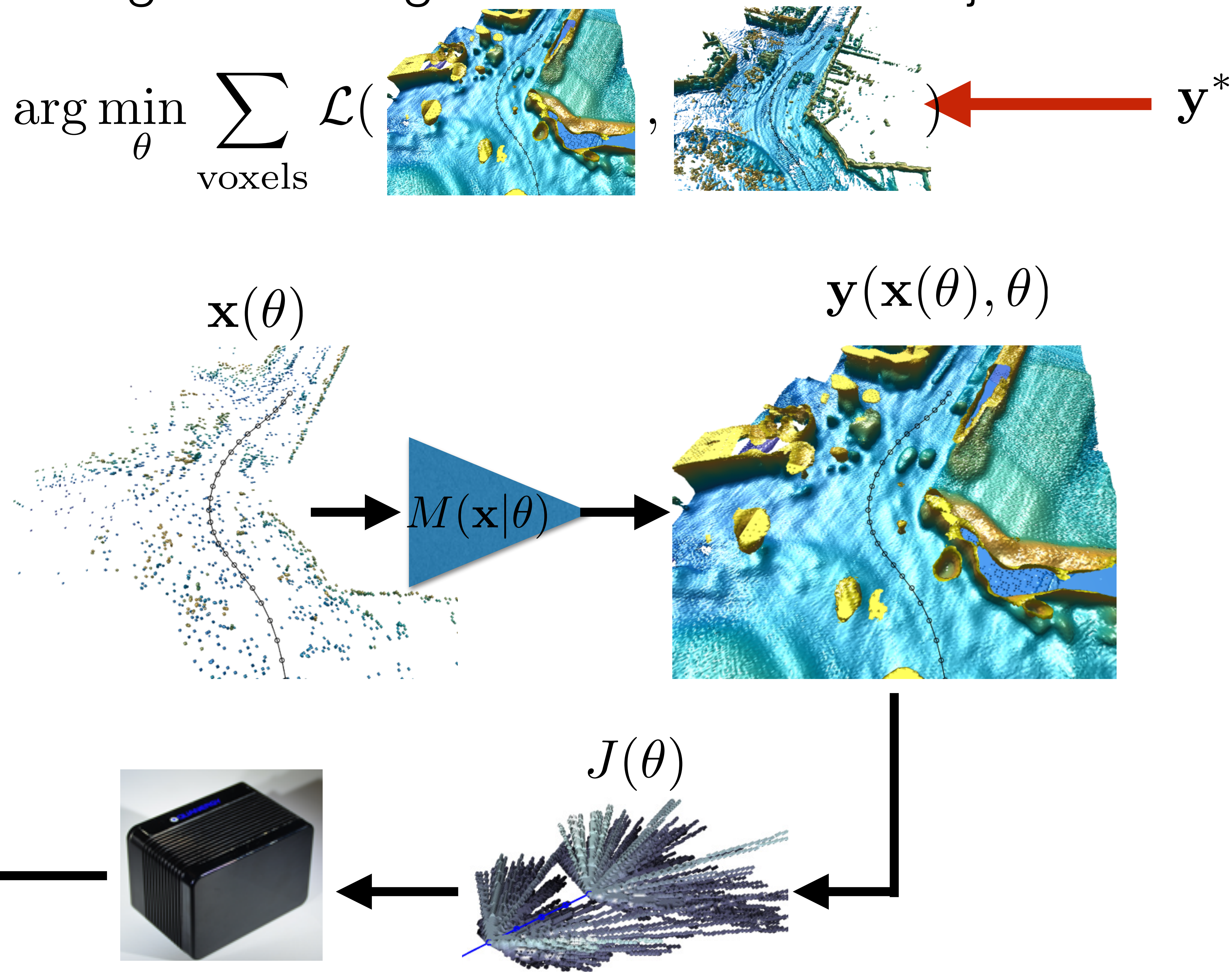
Learning & Planning minimize common objective

$$\arg\min_{\theta} \sum_{\text{voxels}} \mathcal{L}(\quad , \quad)$$





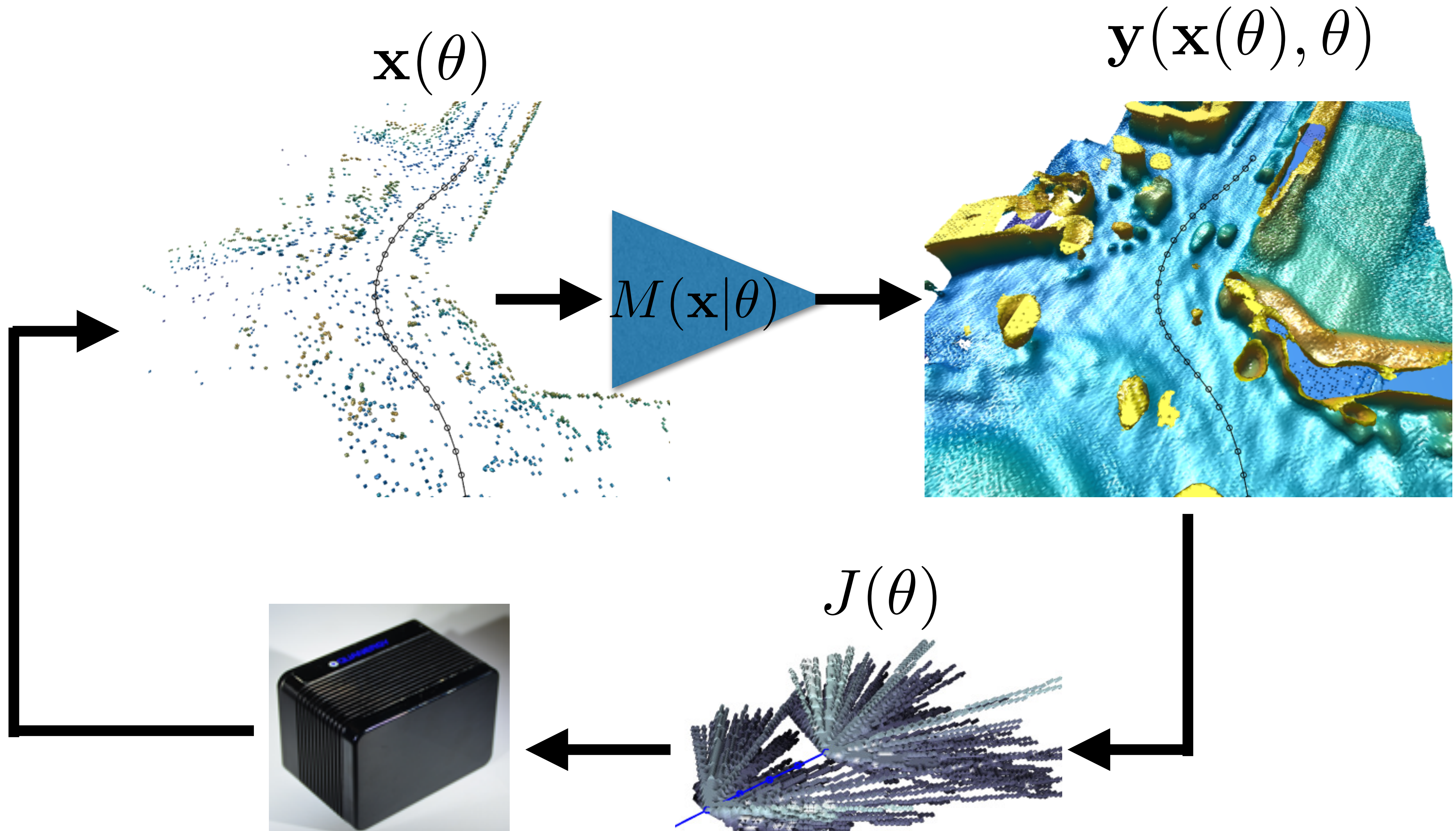
Learning & Planning minimize common objective



Czech Technical University in Prague
Faculty of Electrical Engineering, Department of Cybernetics

Learning & Planning minimize common objective

$$\arg\min_{\theta} \sum_{\text{voxels}} \mathcal{L}(\mathbf{y}(\mathbf{x}(\theta),\theta),\ \mathbf{y}^*) \ \ \text{subject to}\ |J(\theta)| \leq K$$



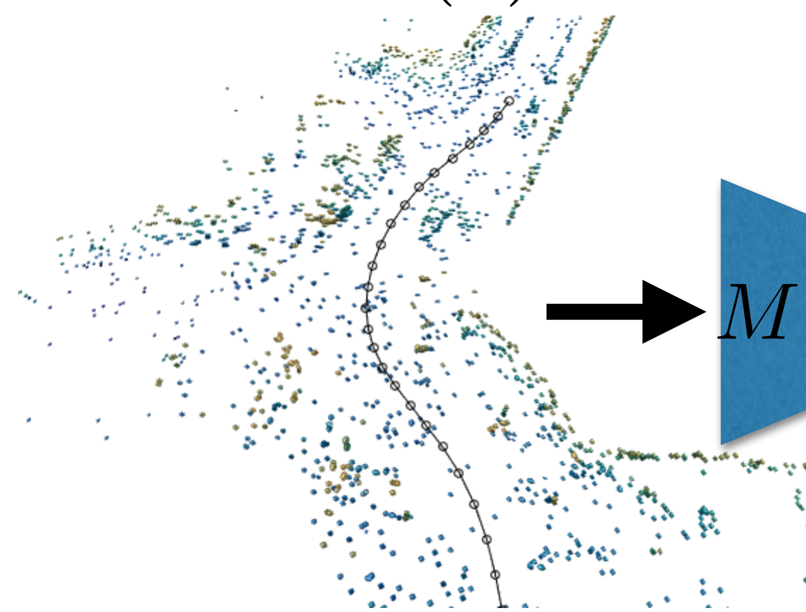
Czech Technical University in Prague
Faculty of Electrical Engineering, Department of Cybernetics

Learning as minimization over θ

$$\arg\min_{\theta} \sum_{\text{voxels}} \mathcal{L}(\mathbf{y}(\boxed{\mathbf{x}(\theta)}, \theta),\ \mathbf{y}^*) \ \text{ subject to } |J(\theta)| \leq K$$

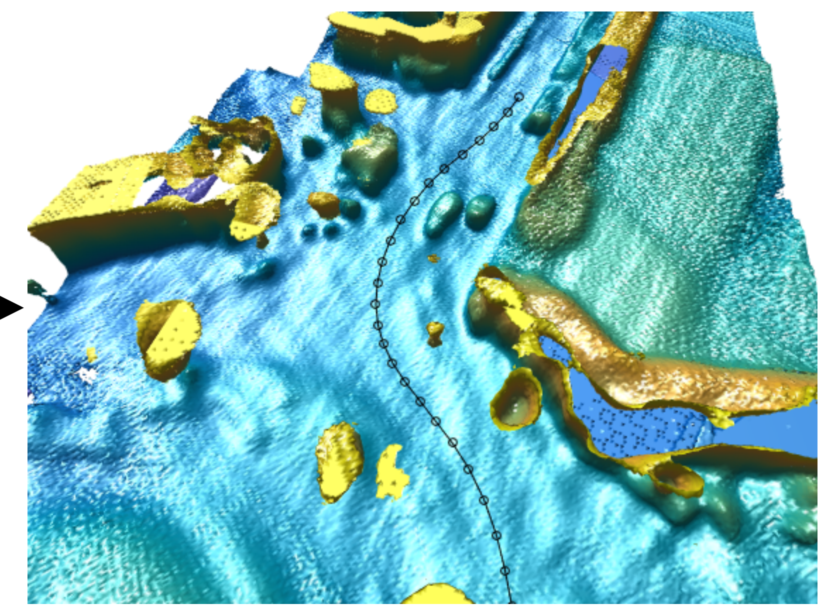
Result of planning is not differentiable

$\mathbf{x}(\theta)$

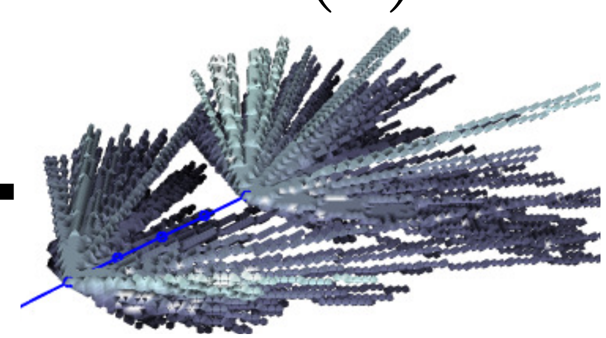


$M(\mathbf{x}|\theta)$

$\mathbf{y}(\mathbf{x}(\theta), \theta)$



$J(\theta)$

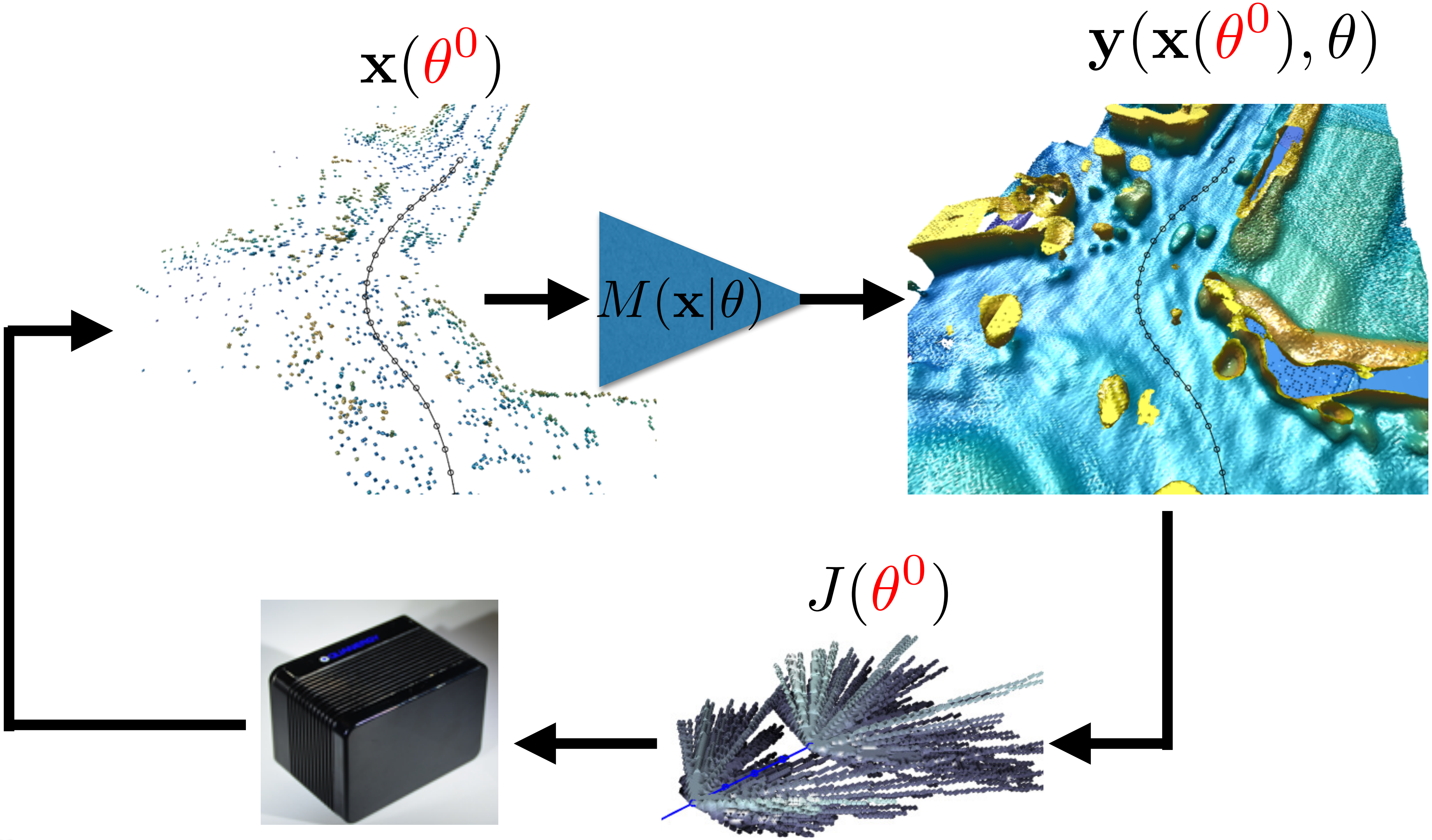


Czech Technical University in Prague
Faculty of Electrical Engineering, Department of Cybernetics

Locally approximate objective around θ^0

$$\arg\min_{\theta} \sum_{\text{voxels}} \mathcal{L}(\mathbf{y}(\boxed{\mathbf{x}(\theta^0)}, \theta), \boxed{\mathbf{y}^*}) \text{ subject to } |J(\theta^0)| \leq K$$

(fixed sparse input, ground truth output)



Czech Technical University in Prague
Faculty of Electrical Engineering, Department of Cybernetics

Minimize approximated objective to get θ^1

$$\theta^1 = \arg\min_\theta \sum_{\text{voxels}} \mathcal{L}(\mathbf{y}(\mathbf{x}(\theta^0), \theta),\ \mathbf{y}^*)$$

$$\theta^0 \xrightarrow{\text{SGD}} \theta^1$$

Minimize approximated objective to get θ^1

$$\theta^2 = \arg\min_{\theta} \sum_{\text{voxels}} \mathcal{L}(\mathbf{y}(\mathbf{x}(\theta^1), \theta),\ \mathbf{y}^*)$$

$$\theta^0 \xrightarrow{\text{SGD}} \theta^1 \rightarrow \theta^2$$

Czech Technical University in Prague
Faculty of Electrical Engineering, Department of Cybernetics

Iteratively optimize approximated objective

$$\theta^{t+1} = \arg\min_{\theta} \sum_{\text{voxels}} \mathcal{L}(\mathbf{y}(\mathbf{x}(\theta^{t}), \theta),\ \mathbf{y}^{*})$$

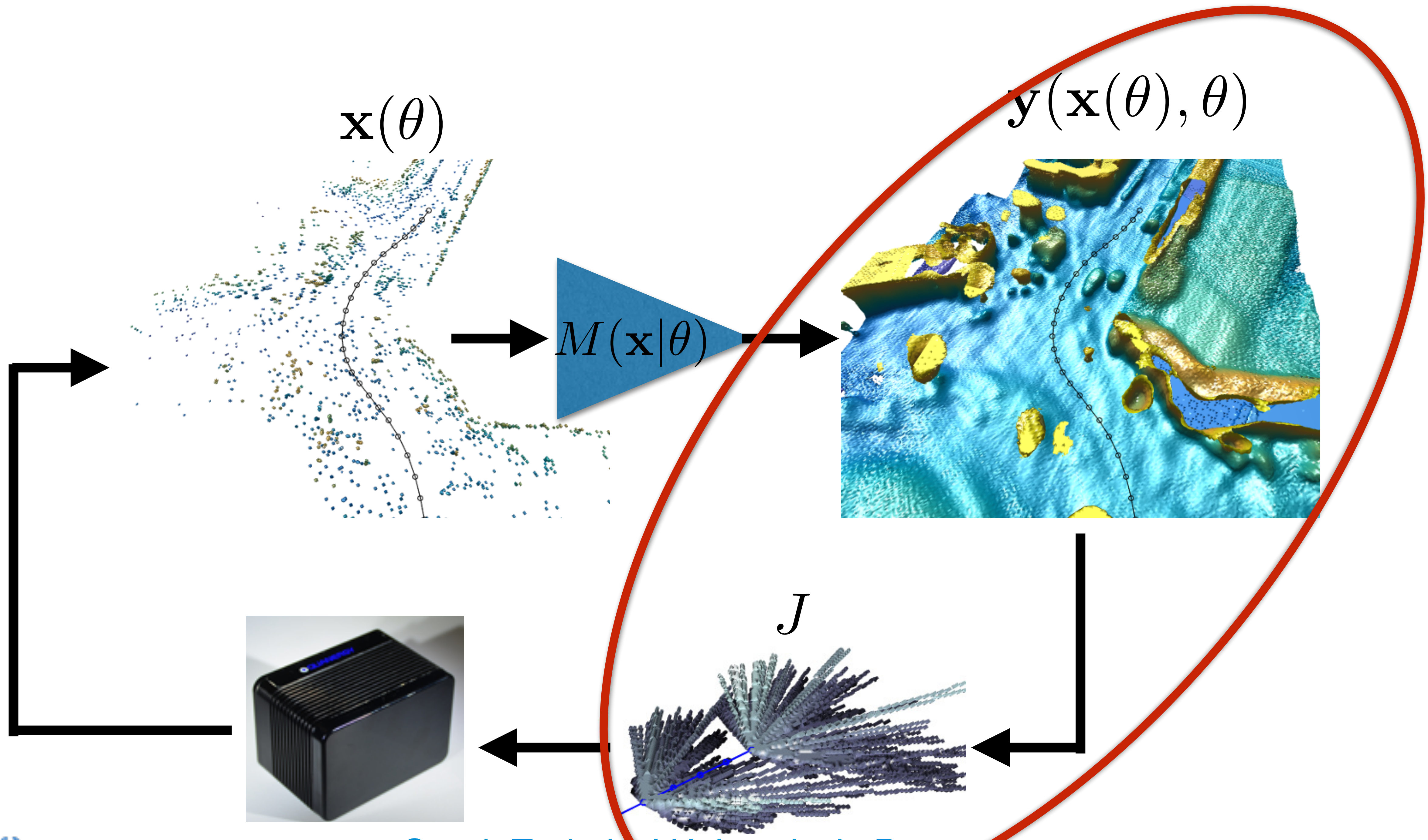
$$\theta^0 \rightarrow \theta^1 \rightarrow \theta^2 \rightarrow \quad \dots \quad \theta^t \rightarrow \theta^{t+1} \rightarrow \dots$$

- Fix point of this mapping would assure:
  - local optimality of the objective
  - statistical consistency of the learning

- In practise, we iterate until validation error decreases

Planning of depth measuring rays J

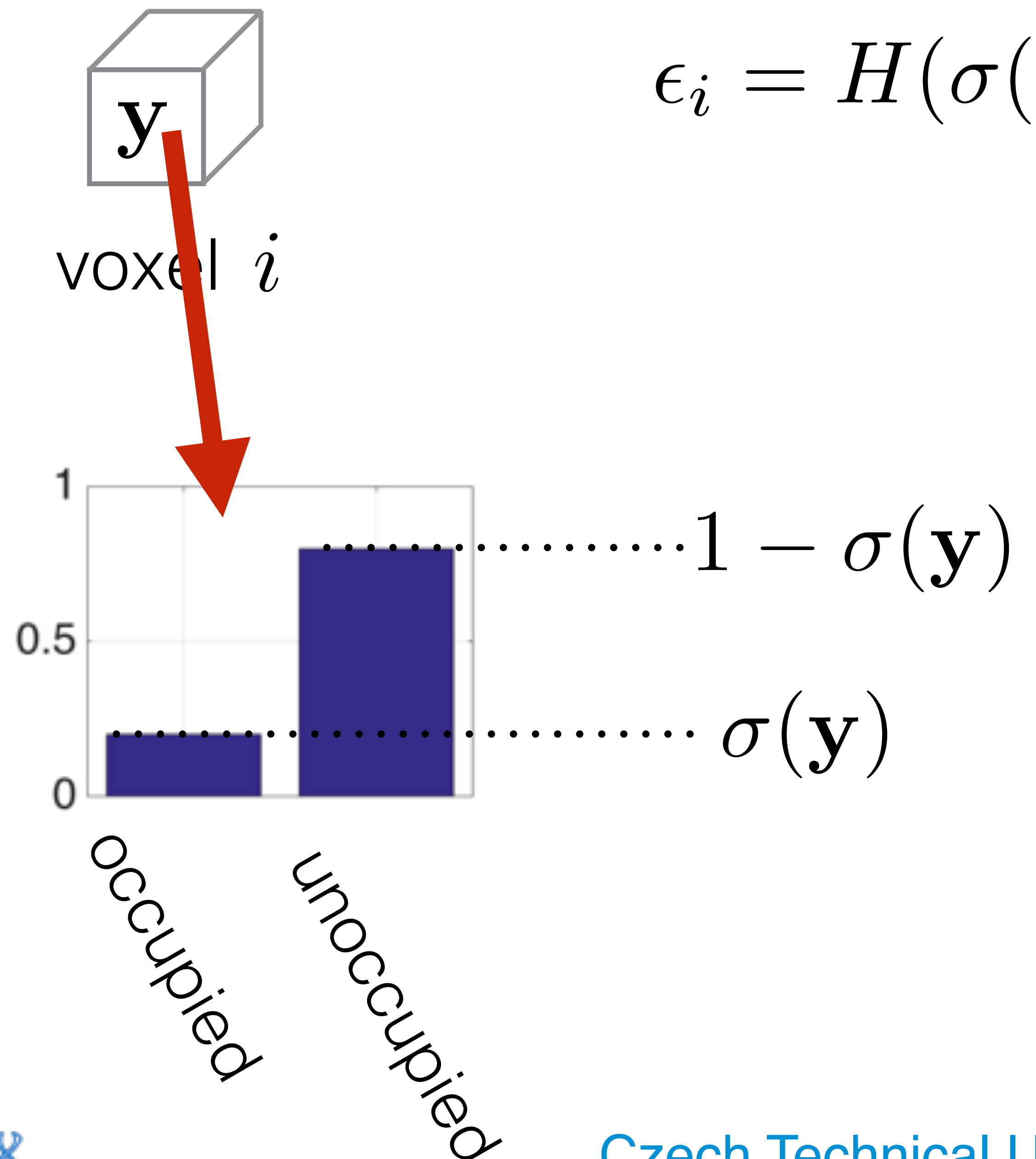
- No ground truth $\mathbf{y}^*$ availble
- Objective for planning is approximated



Planning of depth measuring rays J

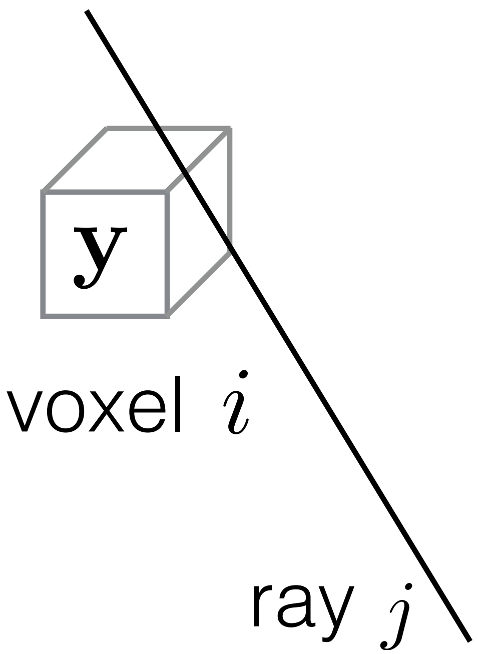
- No ground truth $\mathbf{y}^*$ availble
- Objective for planning is approximated

$\epsilon_i = H(\sigma(\mathbf{y}))$ ... expected loss in voxel i



Planning of depth measuring rays J

- No ground truth $\mathbf{y}^*$ availble
- Objective for planning is approximated



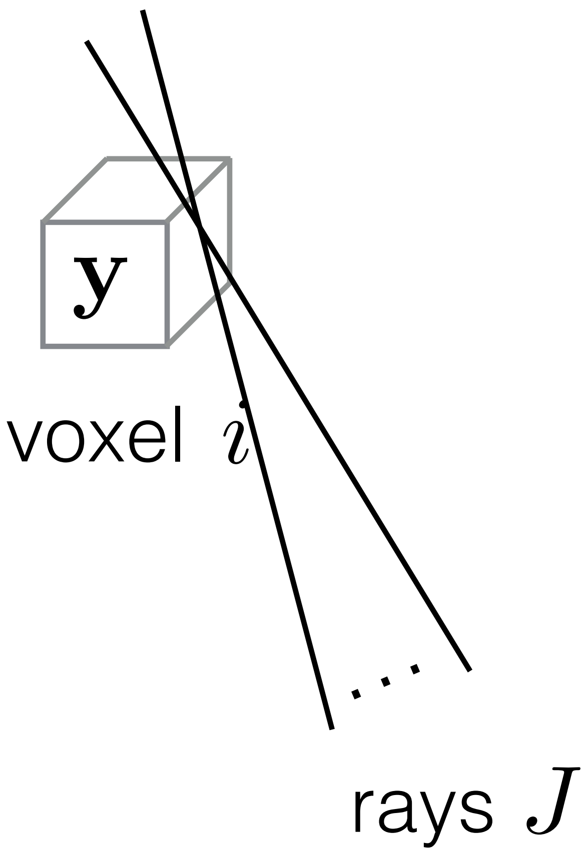
$\epsilon_i = H(\sigma(\mathbf{y}))$ … expected loss in voxel i

$\times$

p_{ij} … prob. that voxel i is not visible in ray j

Planning of depth measuring rays J

- No ground truth $\mathbf{y}^*$ availble
- Objective for planning is approximated



$\epsilon_i = H(\sigma(\mathbf{y}))$ … expected loss in voxel i

$\times$

$\prod_{j \in J} p_{ij}$ … prob. that voxel i is not visible by any ray $j \in J$

Total expected loss: $\sum_{\text{voxels}} \epsilon_i \prod_{j \in J} p_{ij}$

Planning of depth measuring rays J

- Planning of $J = \{J_1 \dots J_L\}$ over horizon L (i.e. for following positions $\ell = 1 \dots L$):

$$\arg\min_J \sum_{\text{voxels}} \epsilon_i \prod_{j \in J} p_{ij} \quad \text{subject to } |J_\ell| \leq K$$

- Convex approximations
- Naive greedy algorithm

1. Estimate decrease of the objective $\Delta_j(t)$ for all rays j

Planning of depth measuring rays J

- Planning of $J = \{J_1 \dots J_L\}$ over horizon L (i.e. for following positions $\ell = 1 \dots L$):

$$\arg\min_{J} \sum_{\text{voxels}} \epsilon_i \prod_{j \in J} p_{ij} \quad \text{subject to } |J_\ell| \leq K$$

- Convex approximations
- Naive greedy algorithm

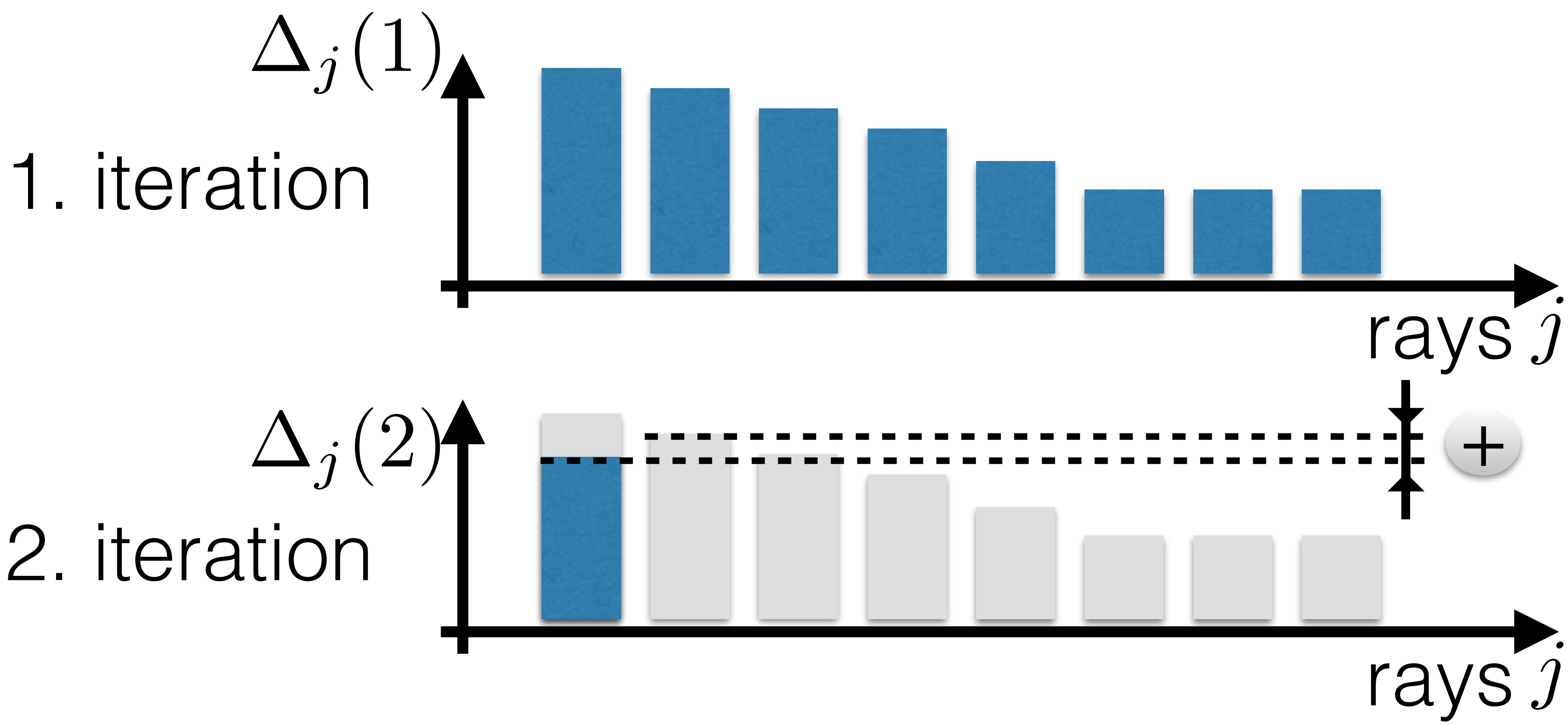
1. Estimate decrease of the objective $\Delta_j(t)$ for all rays j
2. Add the ray which maximizes the decrease

$$J^{t+1} = J^t \cup \arg\max_j \Delta_j(t)$$

Planning of depth measuring rays J

1. Estimate decrease of the objective $\Delta_j(t)$ for all rays j
2. Add the ray which maximizes the decrease

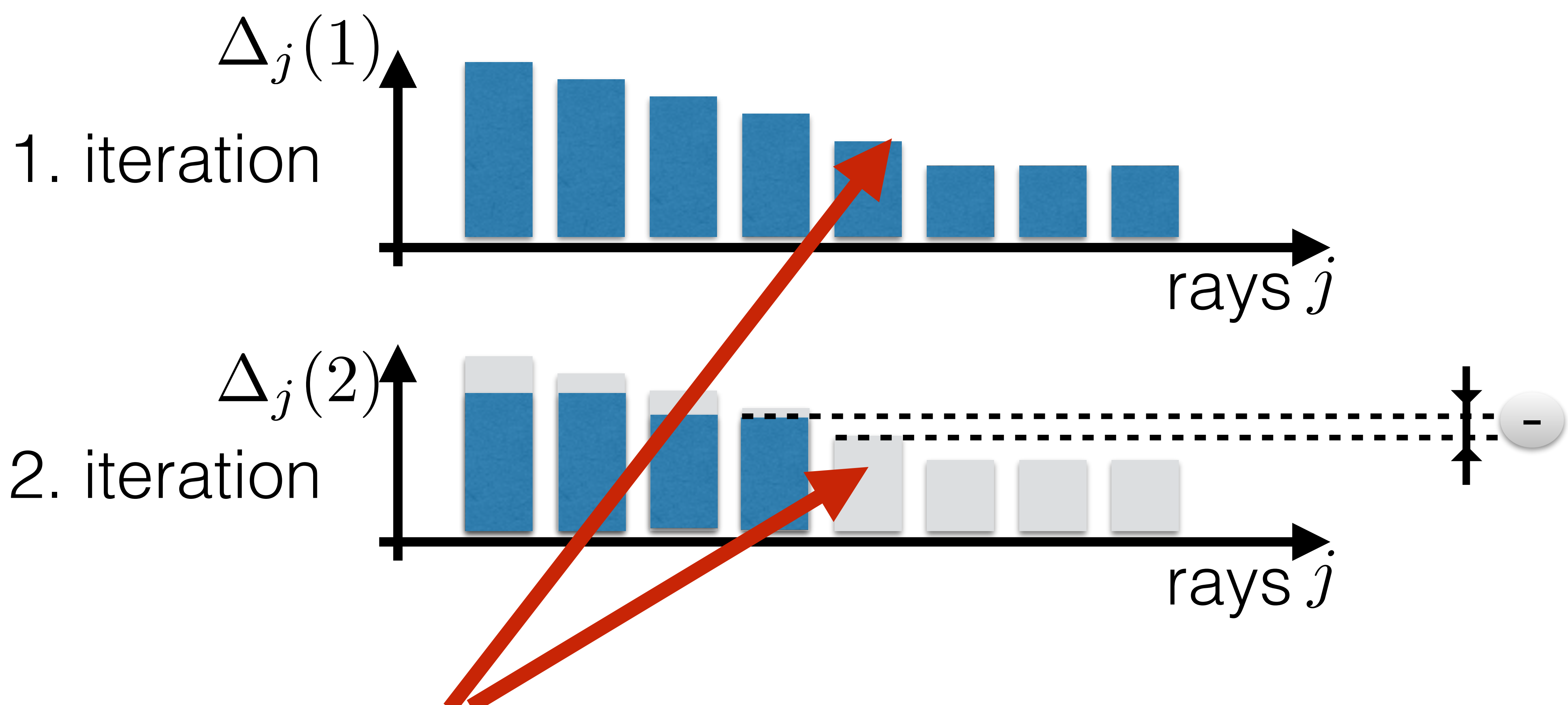
$$J^{t+1} = J^t \cup \arg\max_j \Delta_j(t)$$



Planning of depth measuring rays J

1. Estimate decrease of the objective $\Delta_j(t)$ for all rays j
2. Add the ray which maximizes the decrease

$$J^{t+1} = J^t \cup \arg\max_j \Delta_j(t)$$



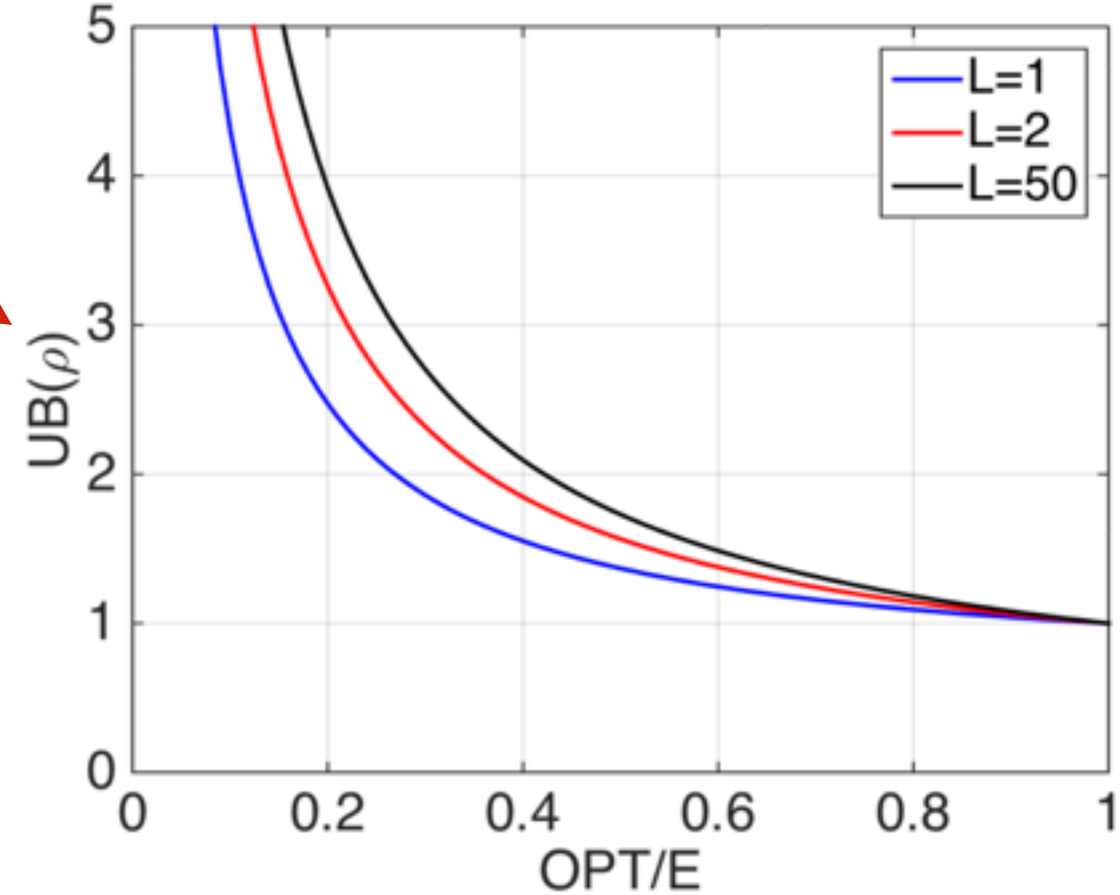
!!! $\Delta_j(t)$ is monotonically non-increasing in t !!!

Czech Technical University in Prague
Faculty of Electrical Engineering, Department of Cybernetics

Planning of depth measuring rays J

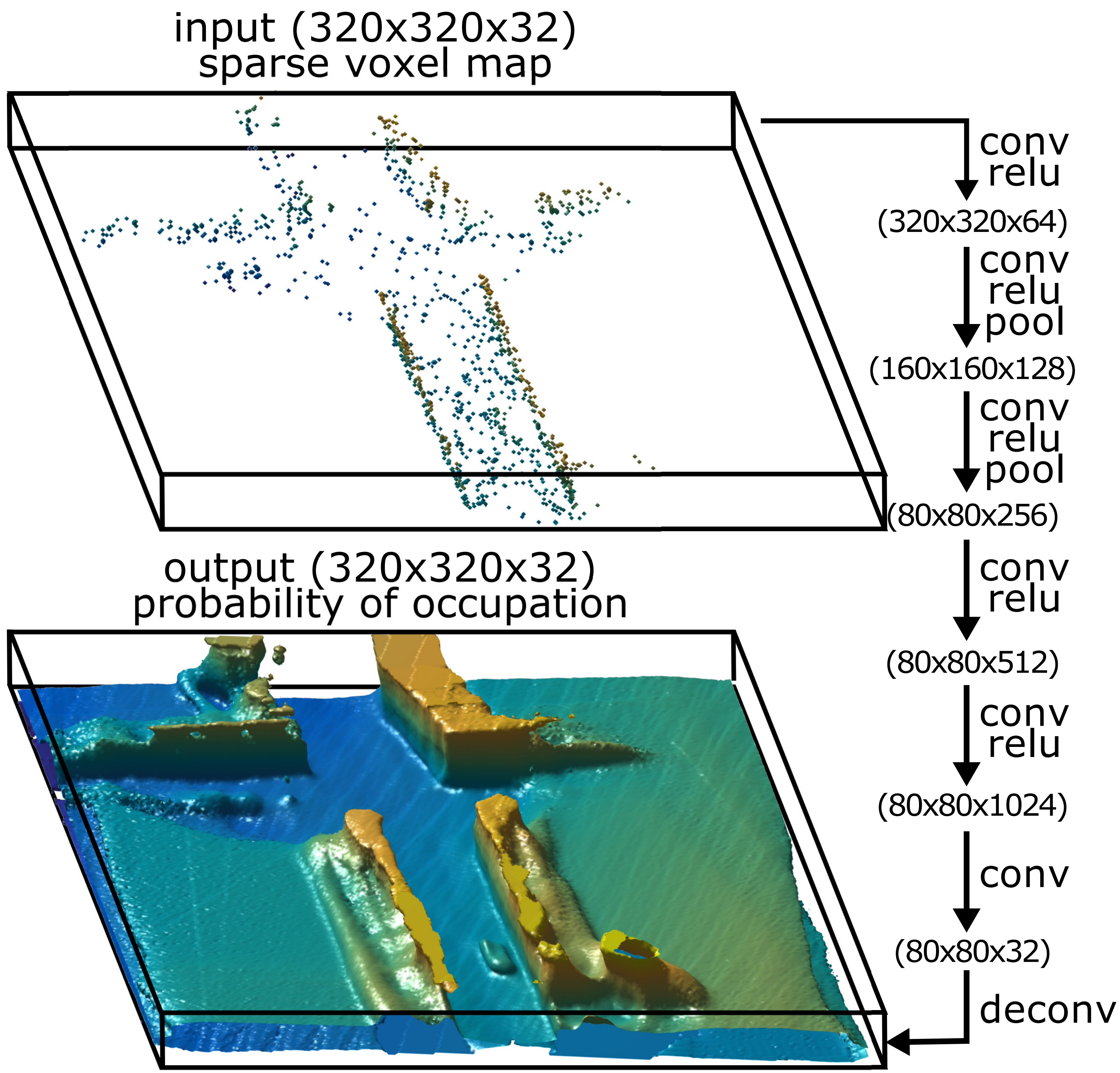
- Naive greedy algorithm
- Prioretized greedy algorithm $\Rightarrow 500\times$ less operations
- Approximation ratio of prioretized greedy

$$\mathrm{UB}\left(\frac{f}{\mathrm{OPT}}\right)$$



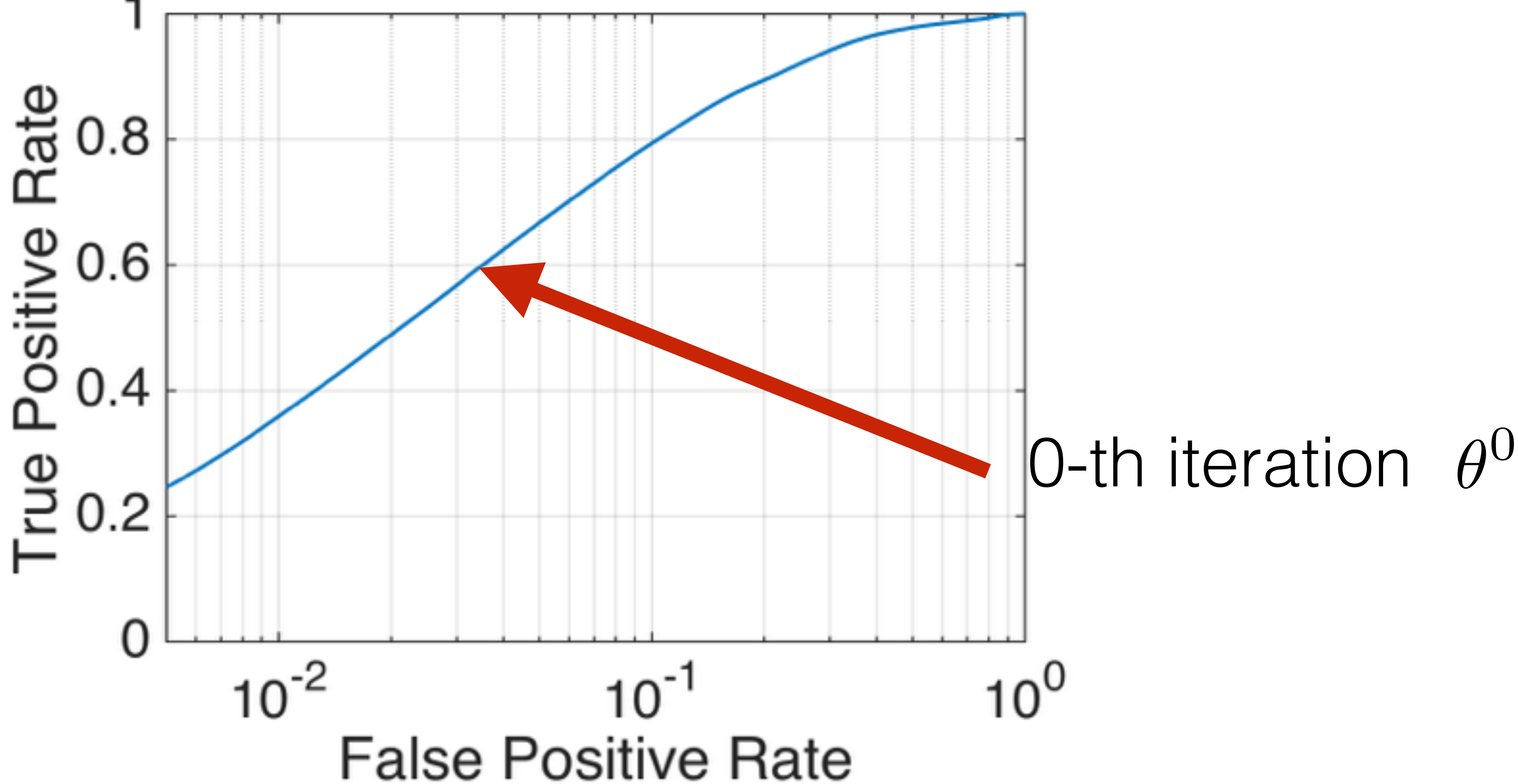
Czech Technical University in Prague
Faculty of Electrical Engineering, Department of Cybernetics

Experiment: Structure of 3D mapping network



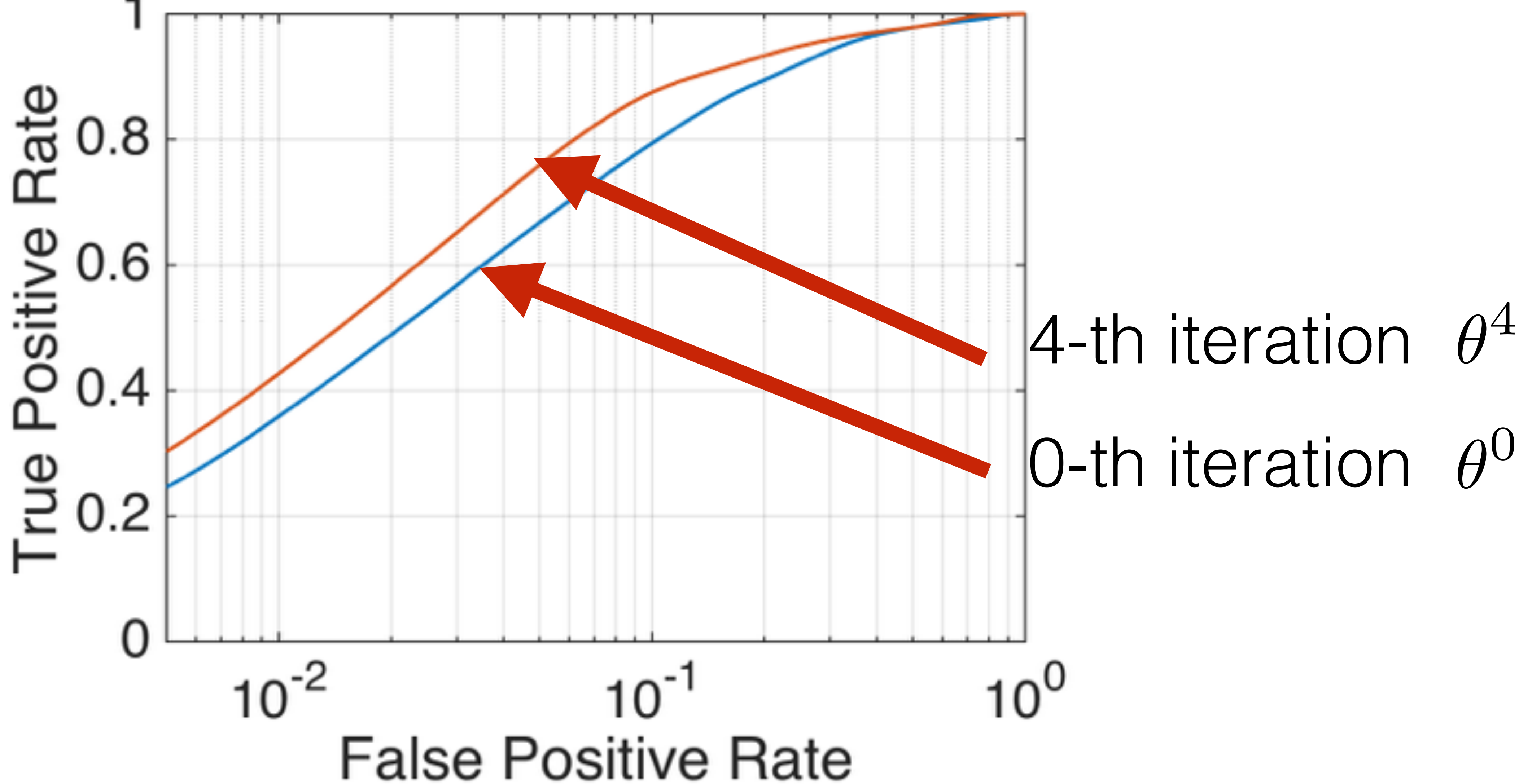
$$\theta \in \mathcal{R}^{20\mathrm{M}}$$

Experiment: Quantitative evaluation on full dataset



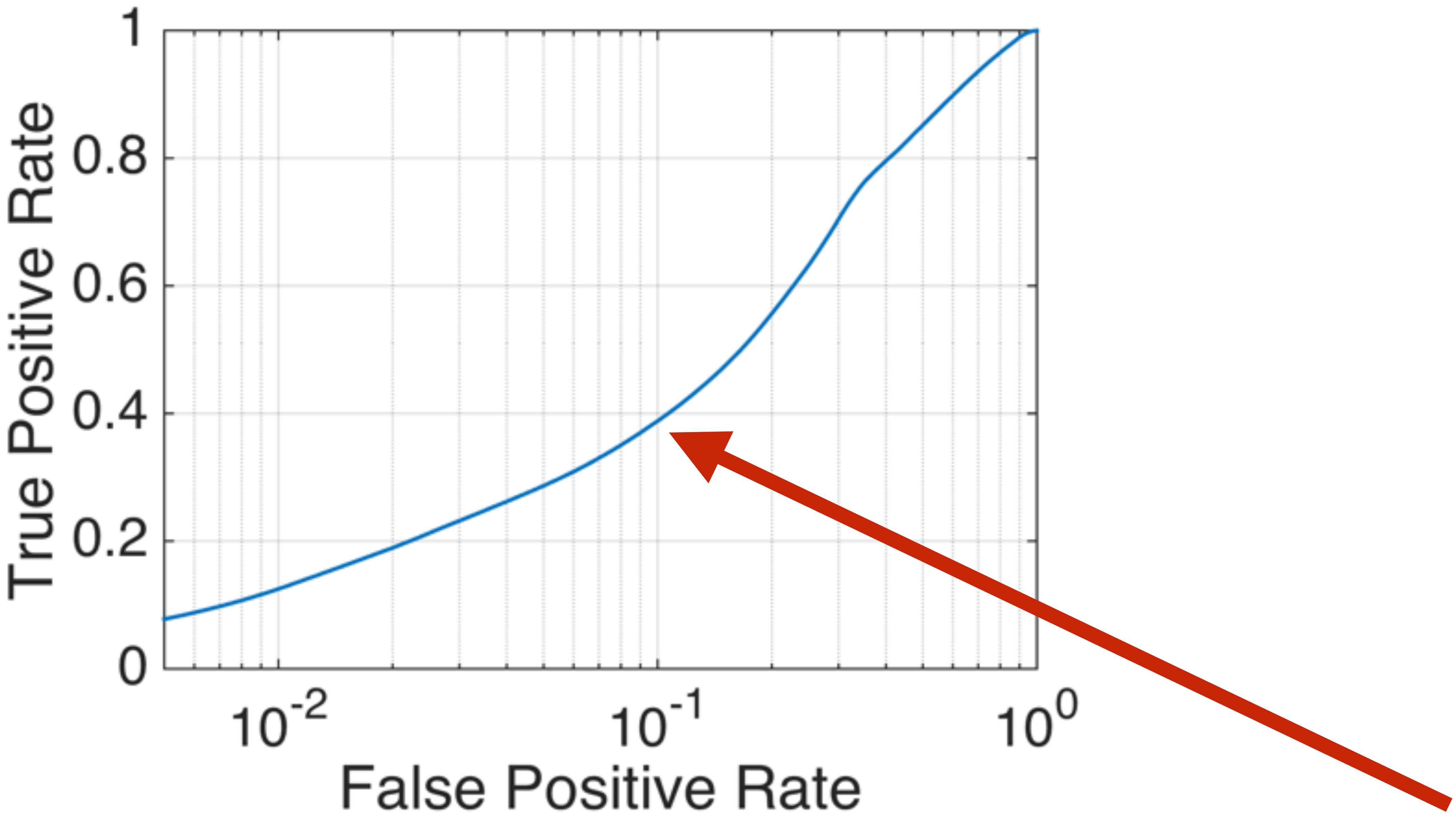
- Training: 20 seq. from “Kitty:*Residential category*”
- Testing: 13 seq. from “Kitty:*City category*”
- Local maps 320x320x32 voxels (1 voxel ~ 20cm)
- Selected K=200 rays per position out of 20k
- Horizon of L=5 positions

Experiment: Quantitative evaluation on full dataset



- Training: 20 seq. from "Kitty:*Residential category*"
- Testing: 13 seq. from "Kitty:*City category*"
- Local maps 320x320x32 voxels (1 voxel ~ 20cm)
- Selected K=200 rays per position out of 20k
- Horizon of L=5 positions

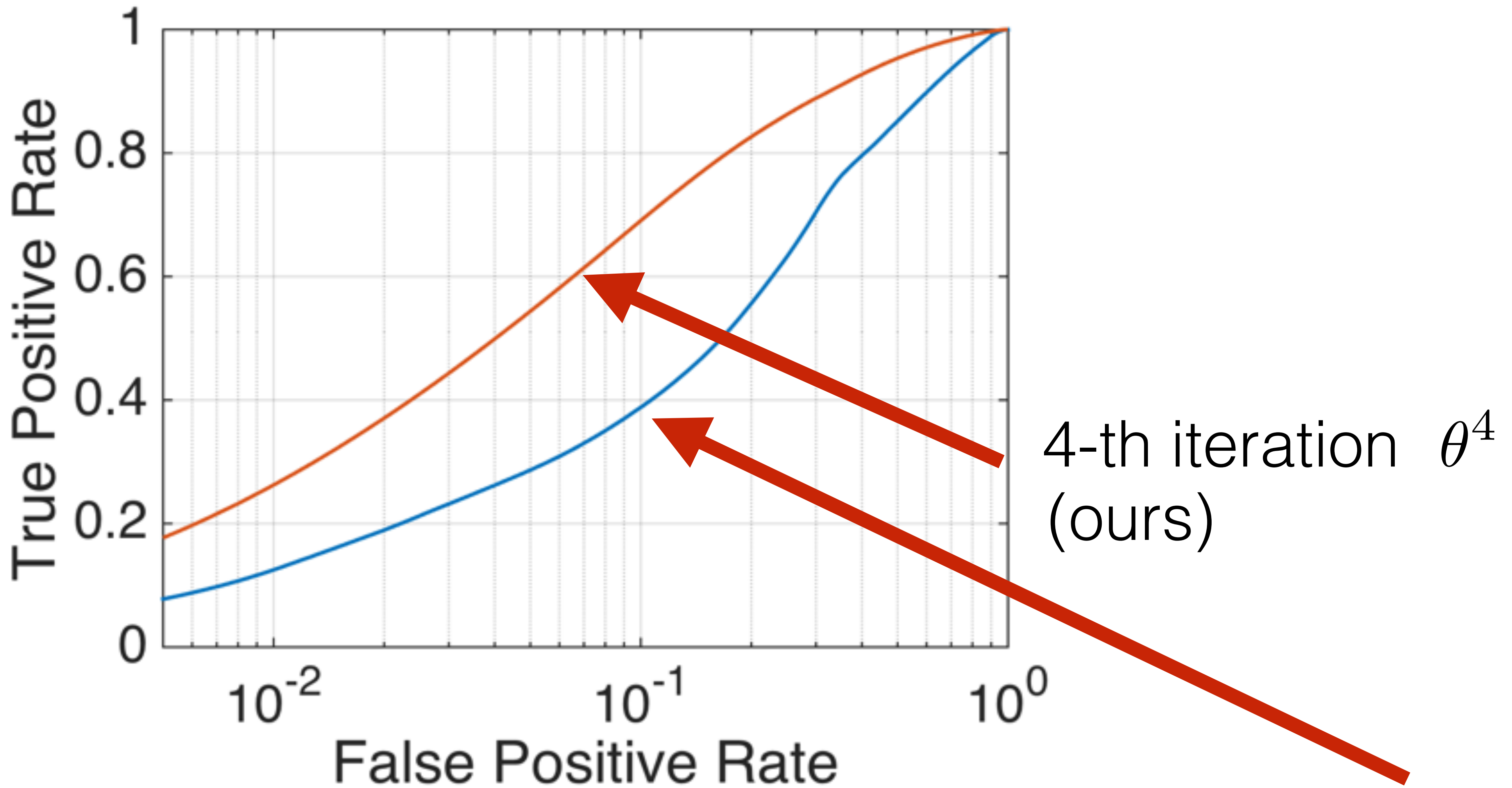
Experiment: Quantitative comparison with [1] on a limited set



- Comparison with modified network (RGB<->depth) from [1]
- Limited setting (128x128x32) due to memory constraints

[1] Choy et al., A unified approach for single a multi-view 3D object reconstruction, ECCV, 2016

Experiment: Quantitative comparison with [1] on a limited set



- Comparison with modified network (RGB<->depth) from [1]
- Limited setting (128x128x32) due to memory constraints

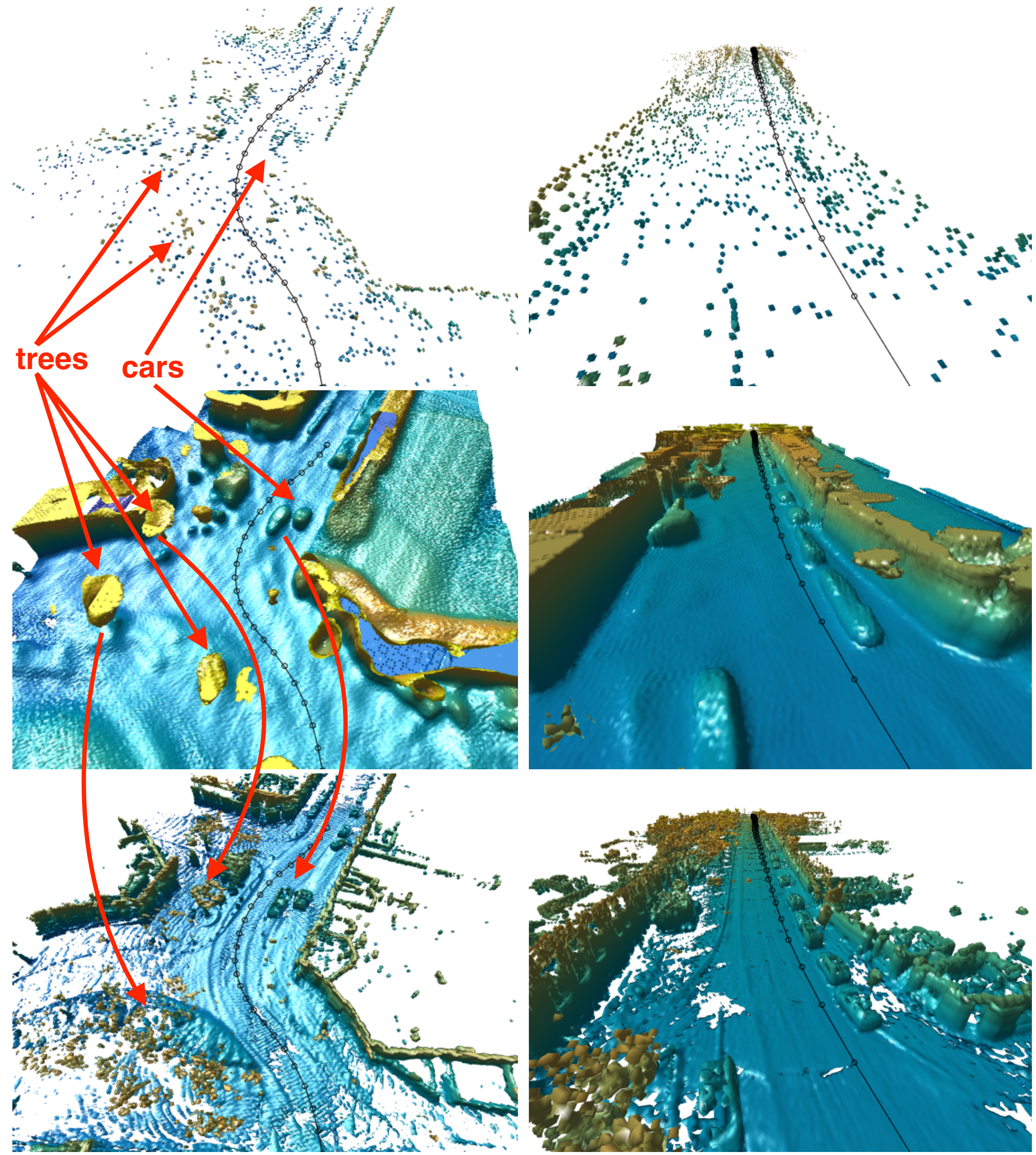
[1] Choy et al., A unified approach for single a multi-view 3D object reconstruction, ECCV, 2016

Experiment: Qualitative evaluation

Sparse measurements

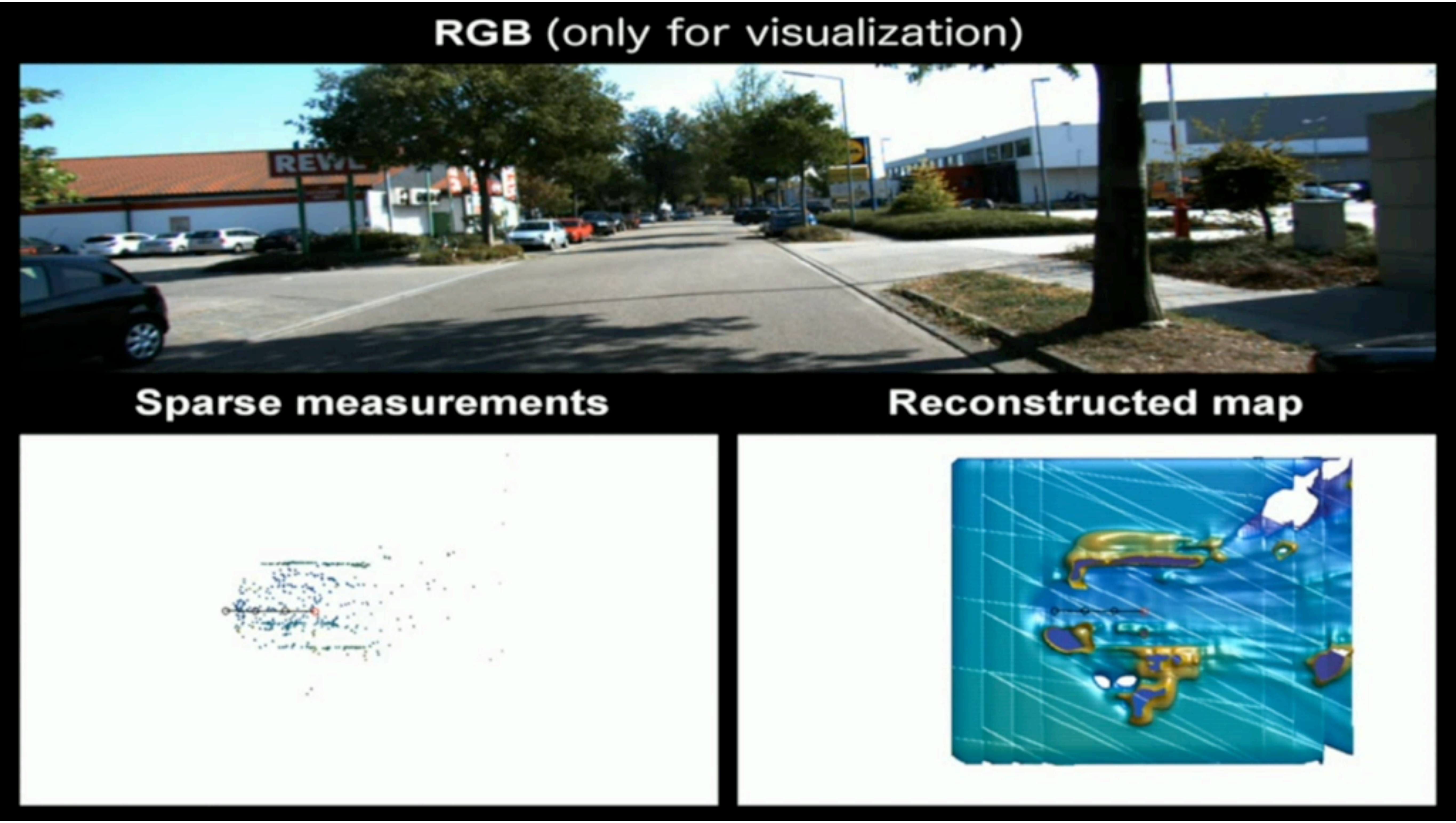
Reconstructed map

Ground truth



Czech Technical University in Prague
Faculty of Electrical Engineering, Department of Cybernetics

Experiment: Summary & Questions



[5] Zimmermann, Petricek, Salansky, Svoboda, Learning for Active 3D Mapping, **ICCV oral**, 2017 https://arxiv.org/abs/1708.02074