Learning for Active 3D Mapping

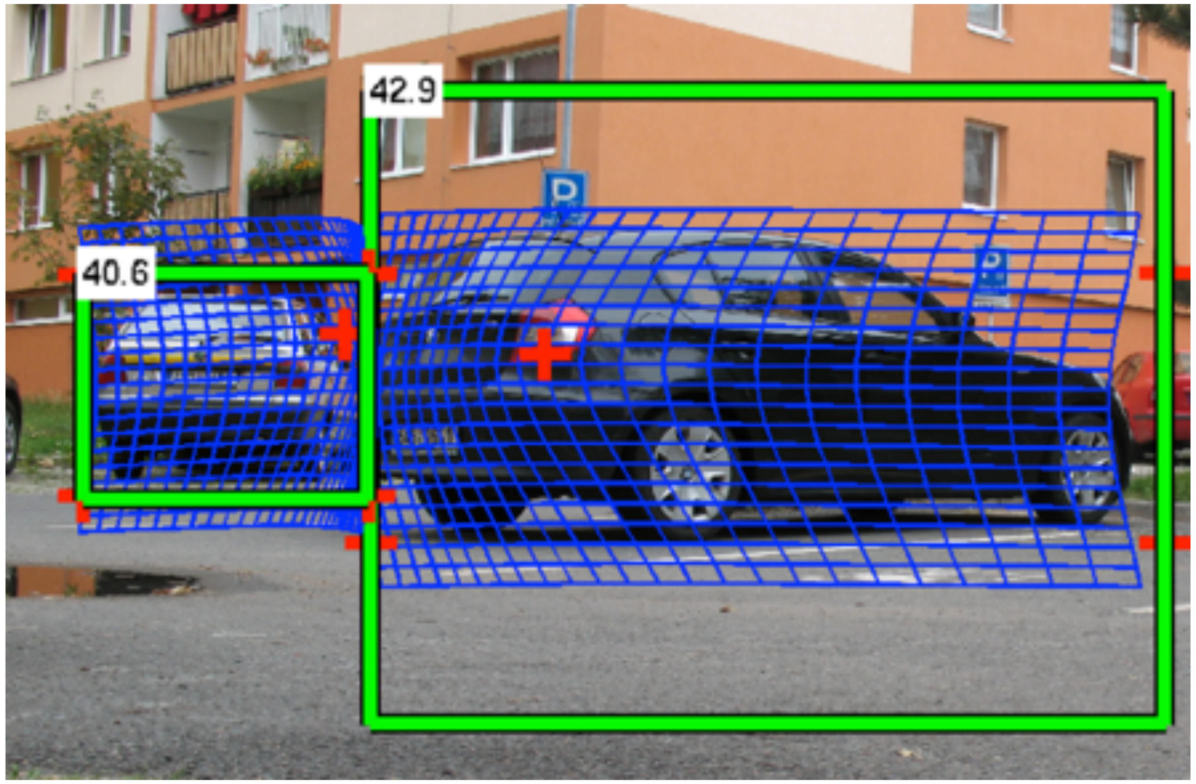
Karel Zimmermann, Tomáš Petříček, Vojtěch Šalanský, Tomáš Svoboda
http://cmp.felk.cvut.cz/~zimmerk/

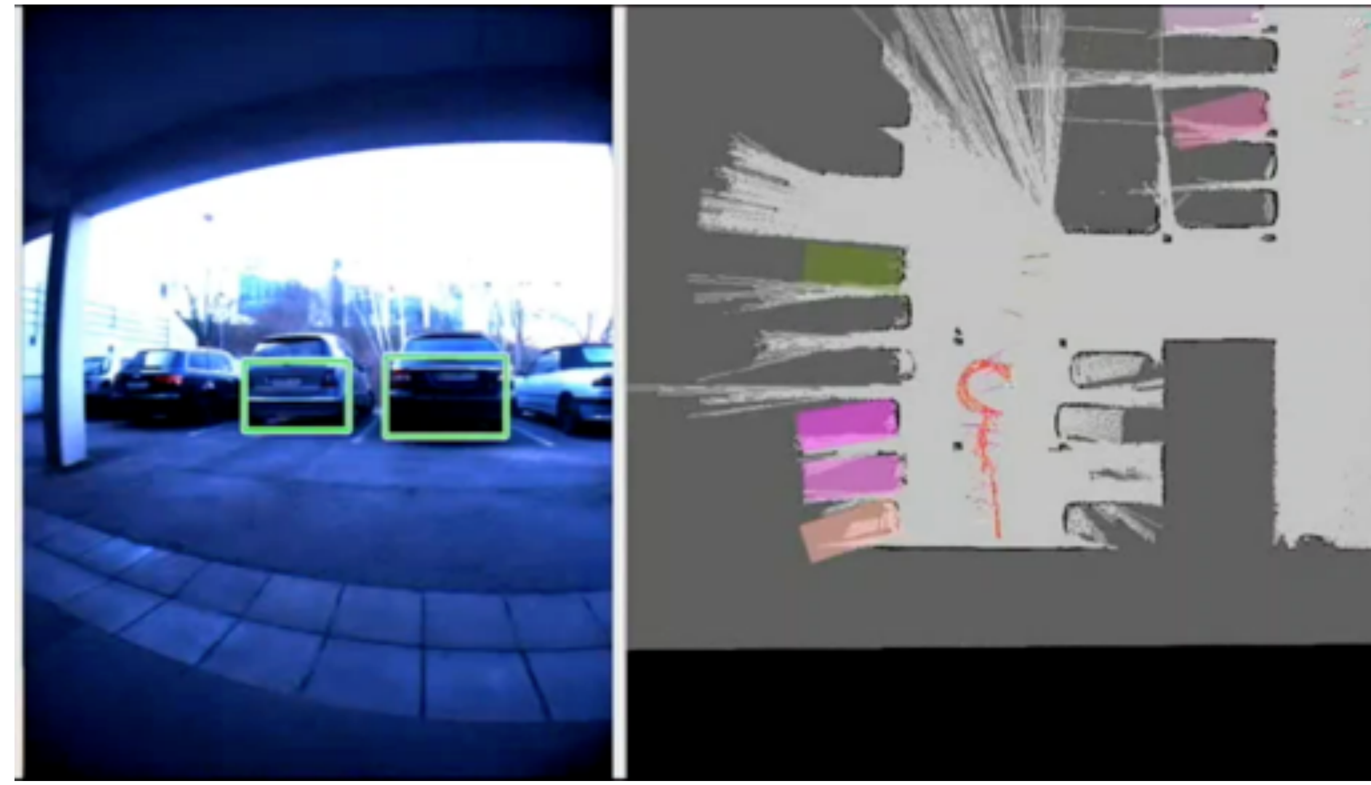
Vision for Robotics and Autonomous Systems
https://cyber.felk.cvut.cz/vras/

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 selected for II.pillar evaluation)**, 2009.

Motion and compliance control of flippers

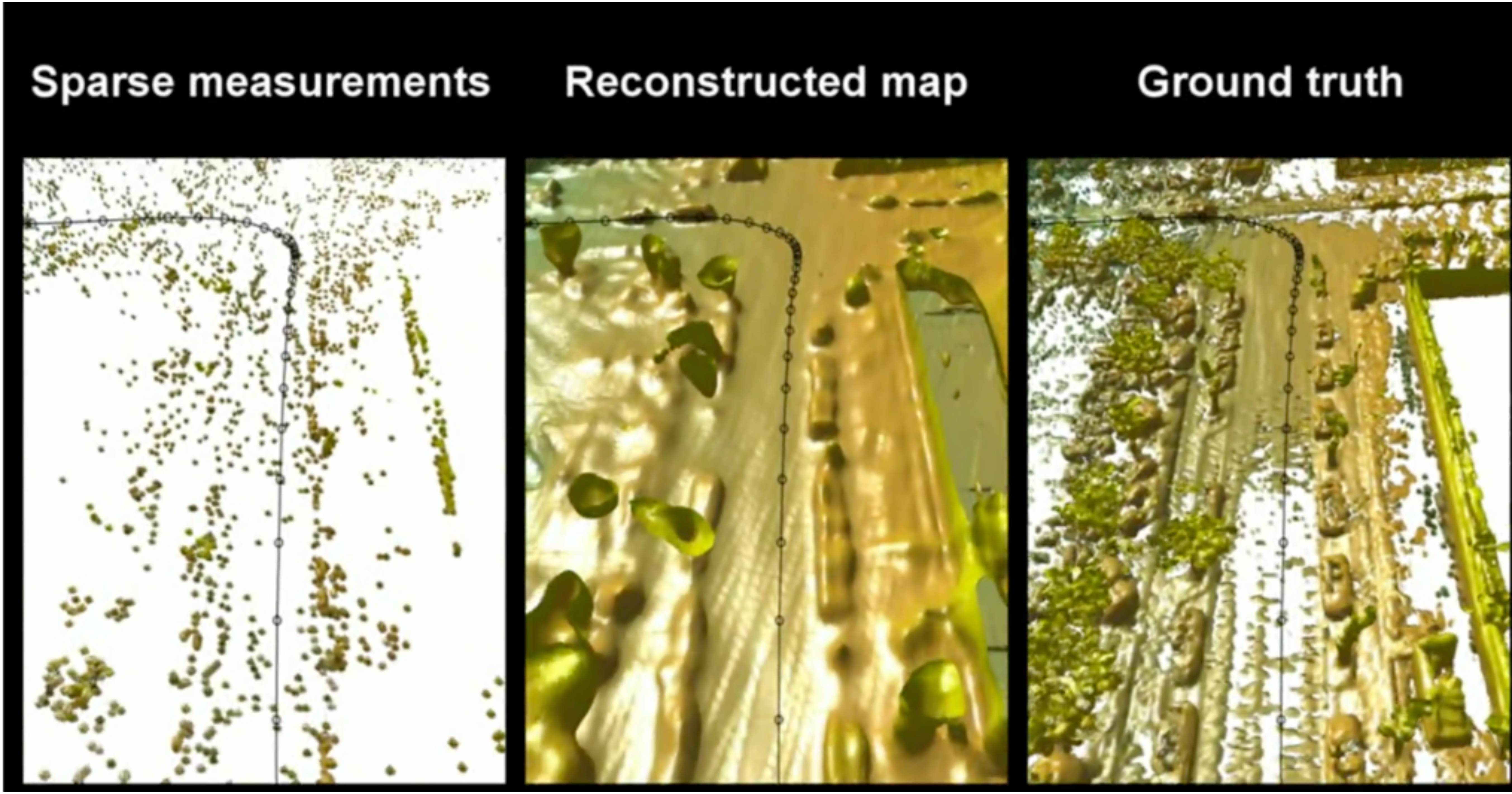
[3] Pecka, Zimmermann, Svoboda, Hlavac, et al.
IROS/RAL/TIE(IF=6), 2015-2018

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, over 200 citations), 2011**

Today's topic



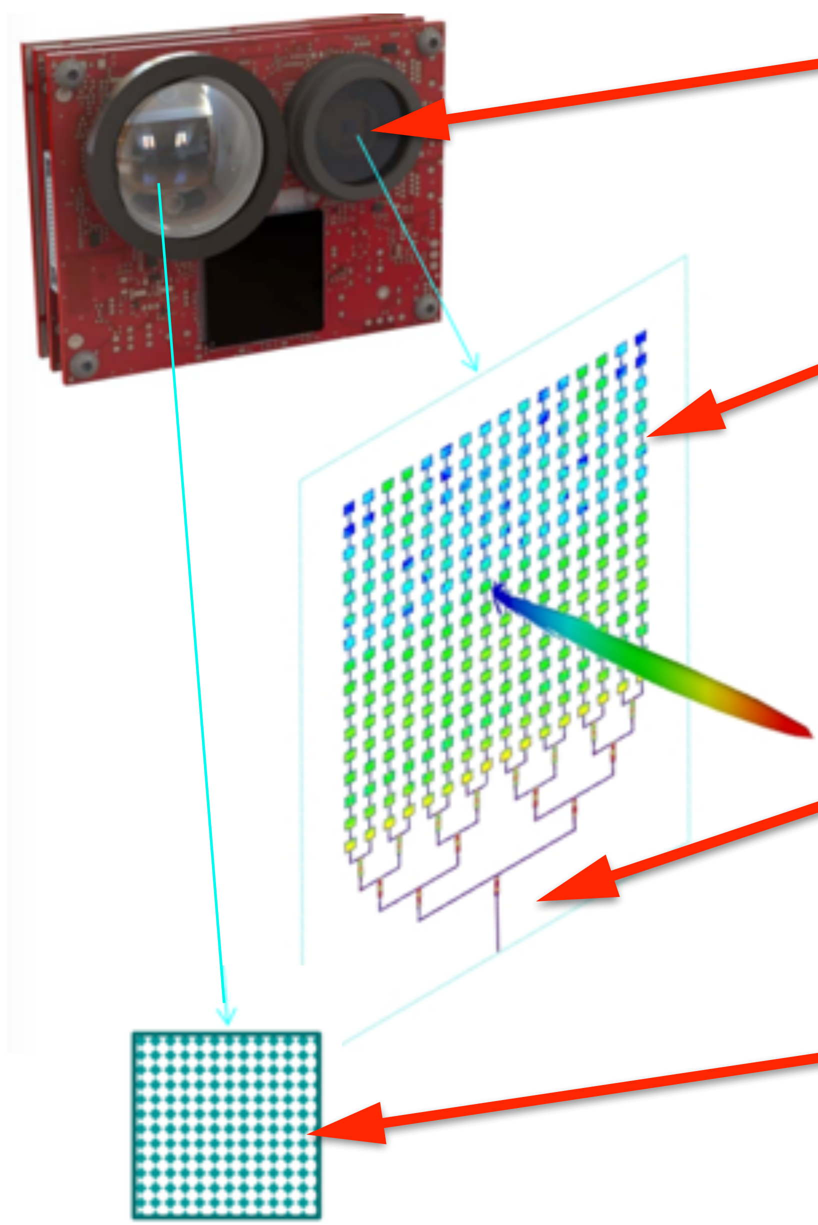
[5] Zimmermann, Petricek, Salansky, Svoboda, Learning for Active 3D Mapping, **ICCV oral (rank A*, AR=2%)**, 2017

https://arxiv.org/abs/1708.02074

Motivation

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

S3 principle



Emitted laser beams

Transmitted through Optical Phased Array

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

Reflected laser beams are captured by SPAD array

Images of S3 Lidar redistributed with permission of Quanergy Systems (http://quanergy.com)

Problem definition

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

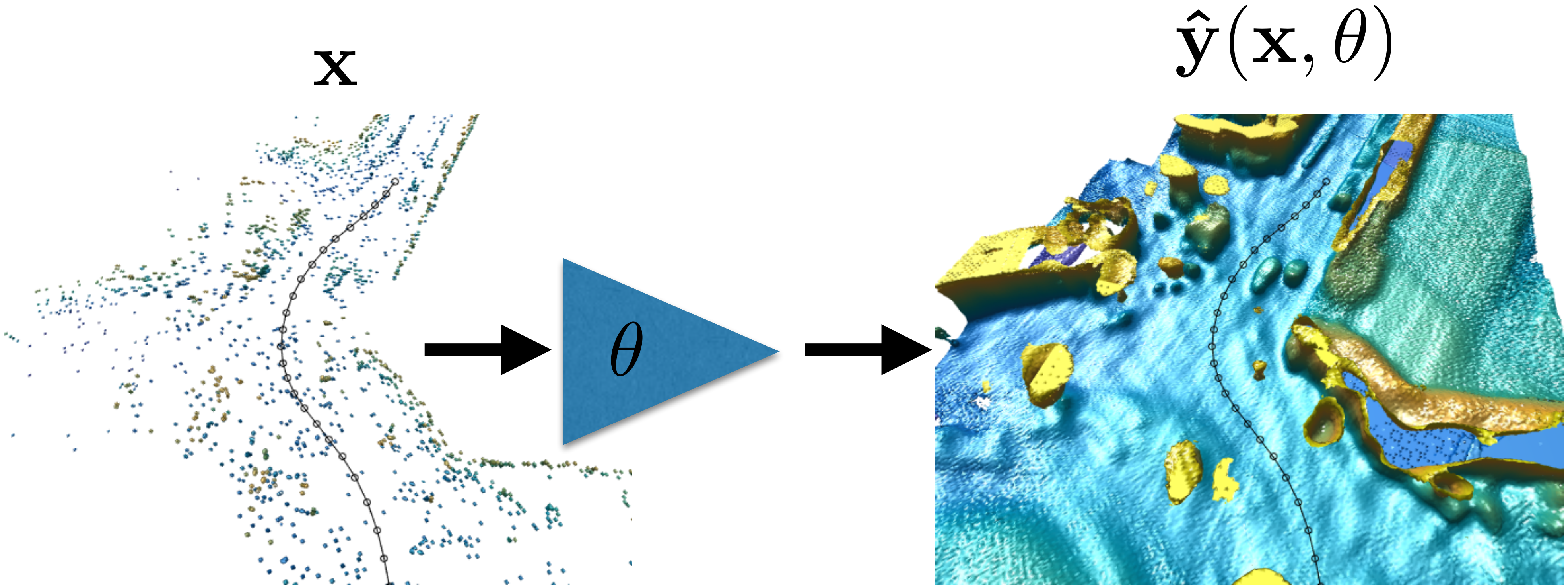
Goal:

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

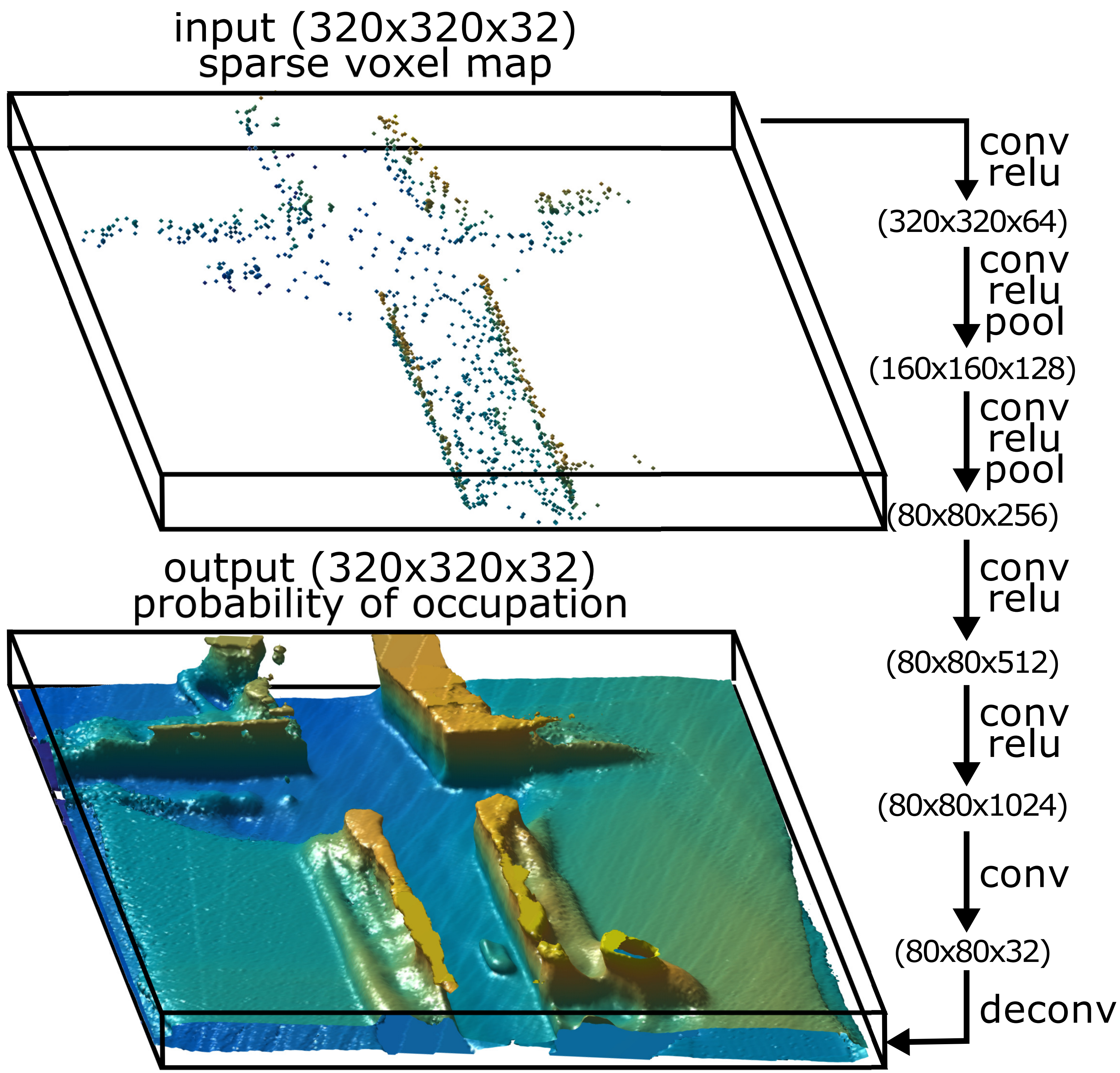
Czech Technical University in Prague
Faculty of Electrical Engineering, Department of Cybernetics

Overview of active 3D mapping

- 3D mapping deep convolutional network ………. $\hat{\mathbf{y}}(\mathbf{x}, \theta)$



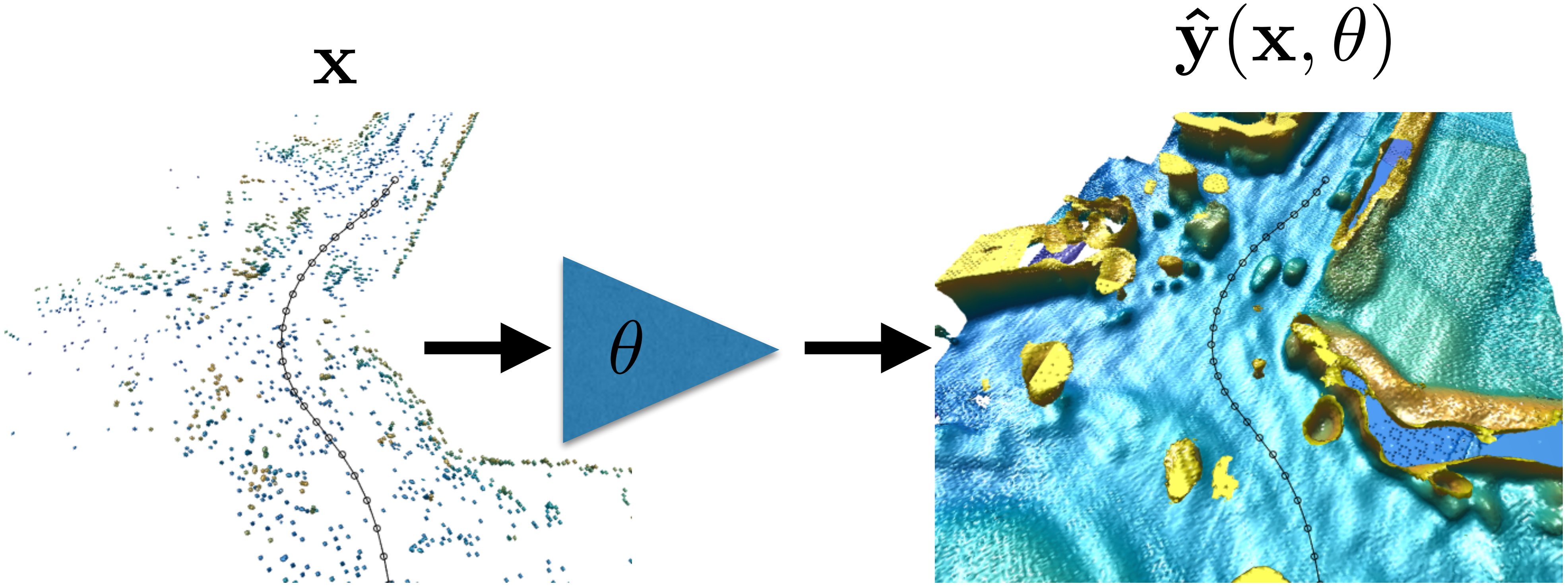
Structure of 3D mapping network



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

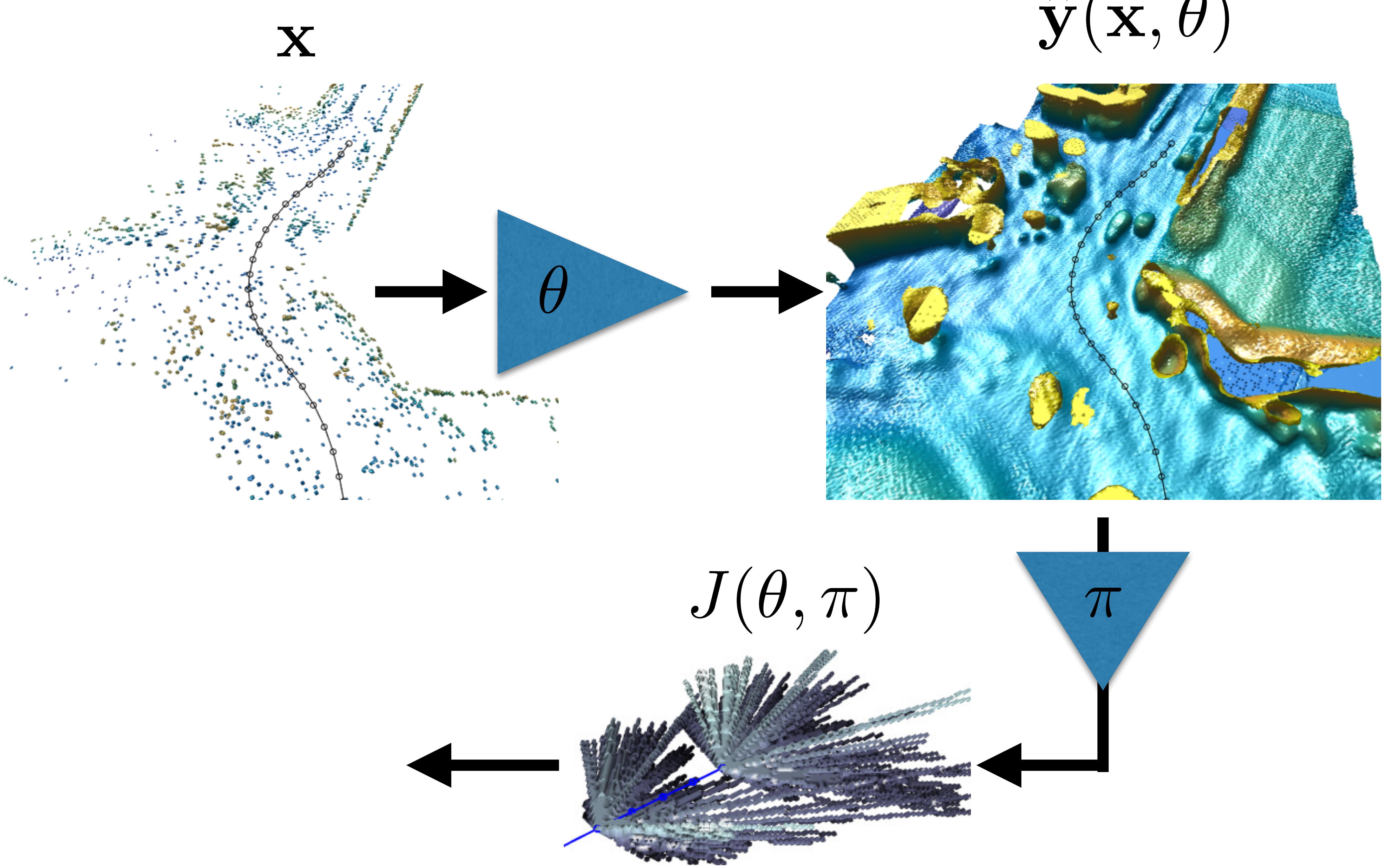
Overview of active 3D mapping

- 3D mapping deep convolutional network ………. $\hat{\mathbf{y}}(\mathbf{x}, \theta)$



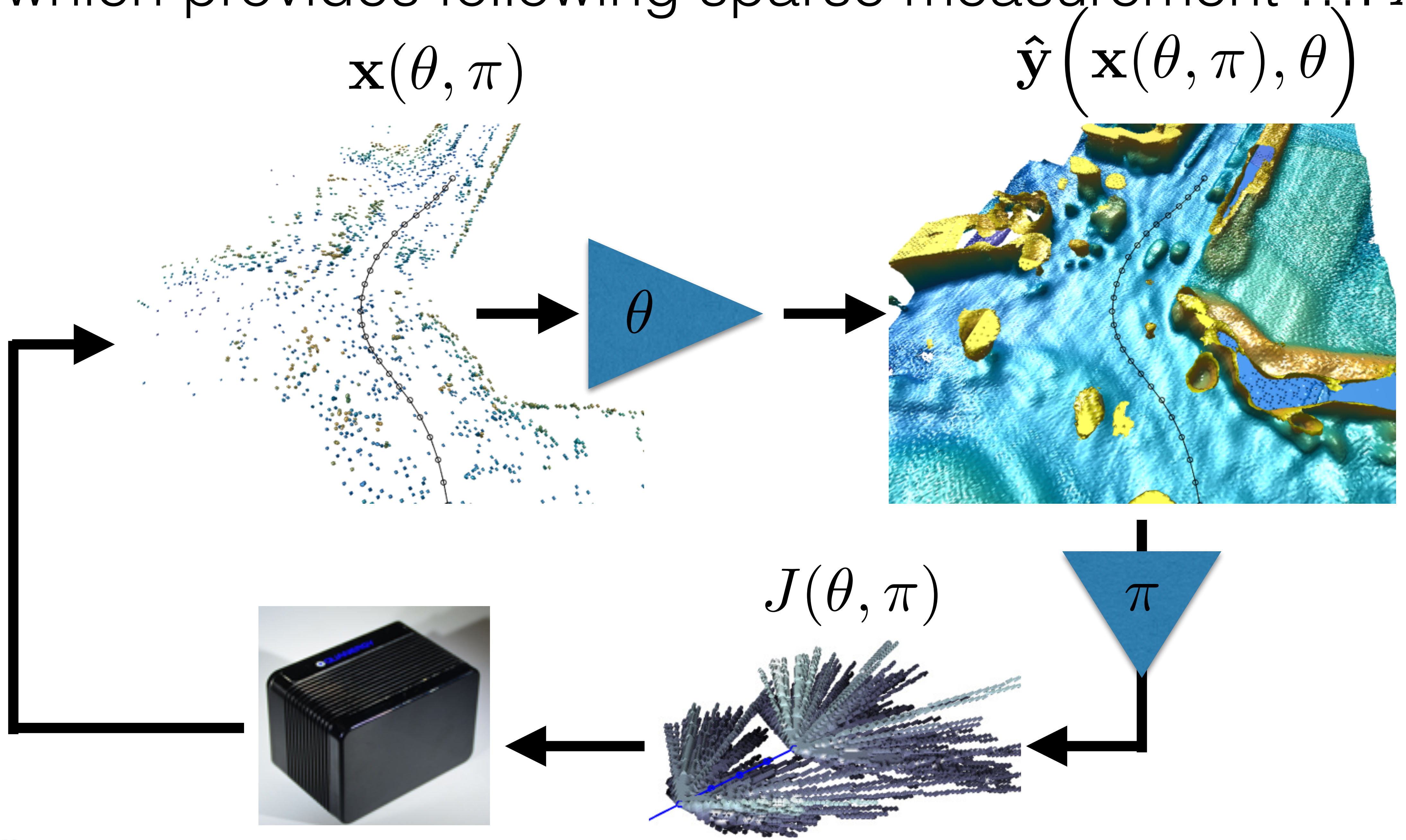
Overview of active 3D mapping

- 3D mapping deep convolutional network ………… $\hat{\mathbf{y}}(\mathbf{x}, \theta)$
- Planning of depth measuring rays …………………… $J(\theta, \pi)$

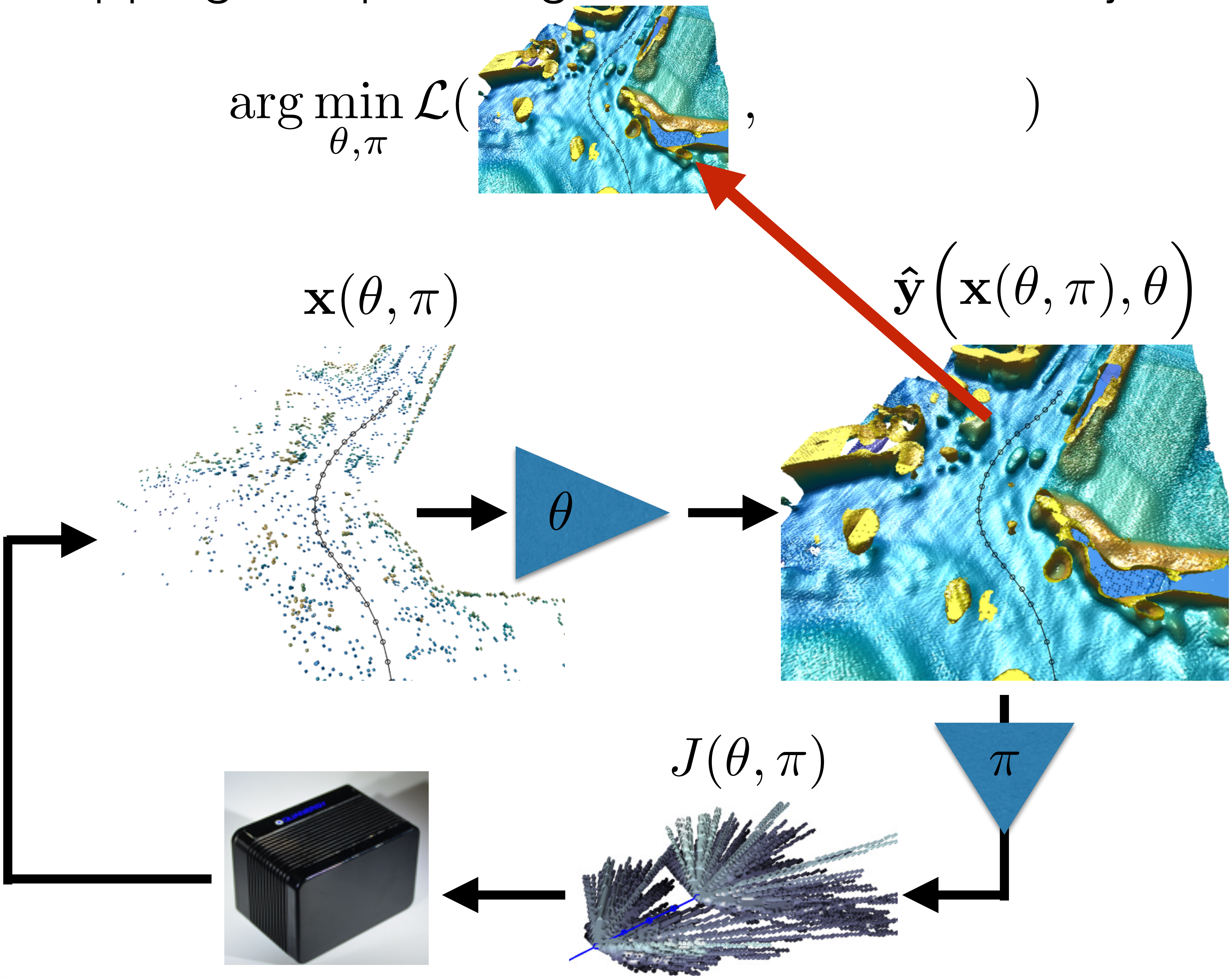


Overview of active 3D mapping

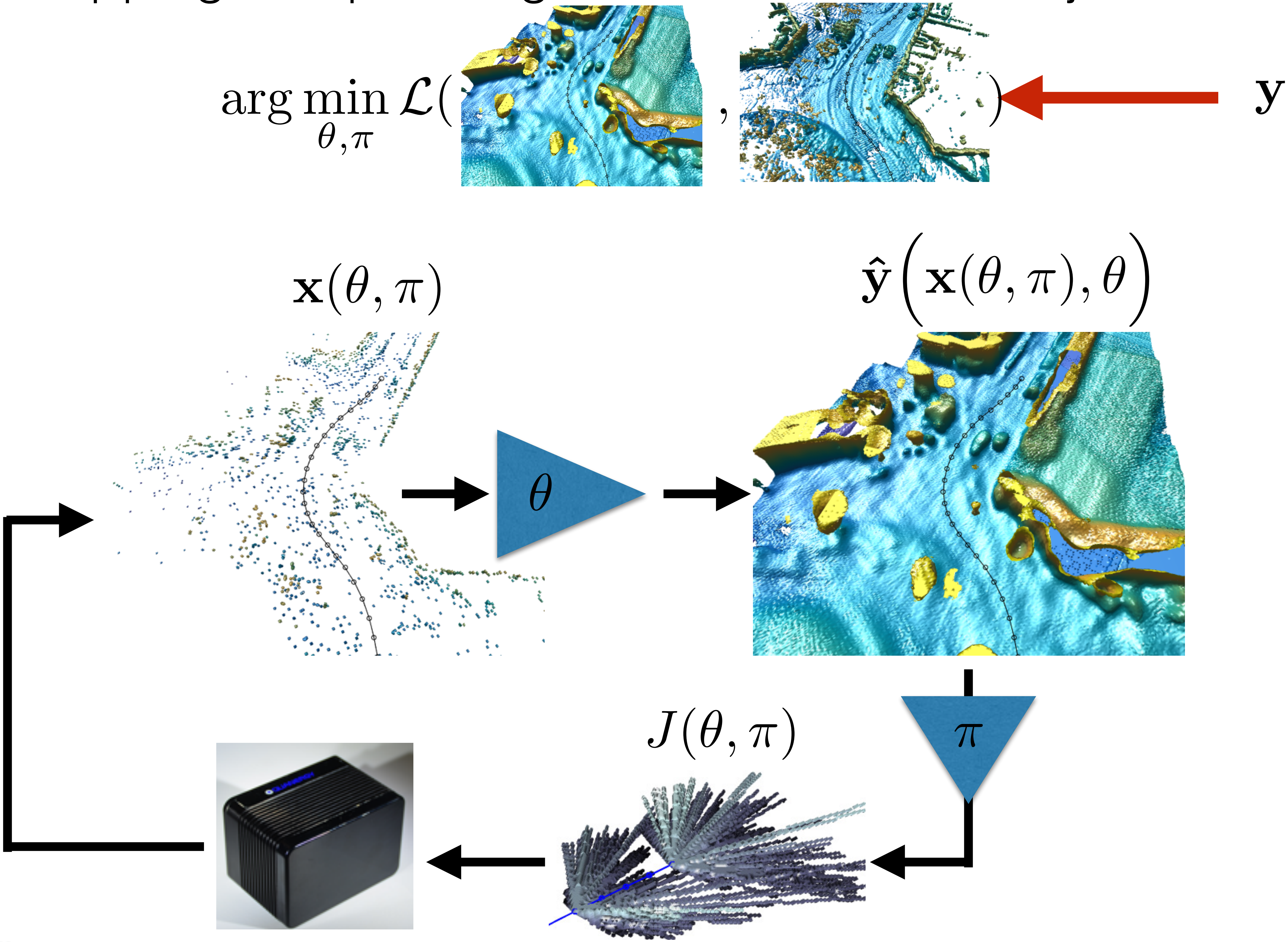
- 3D mapping deep convolutional network .... $\hat{\mathbf{y}}(\mathbf{x}, \theta)$
- Planning of depth measuring rays ....$J(\theta, \pi)$
  which provides following sparse measurement .... $\mathbf{x}(\theta, \pi)$



Mapping and planning minimize common objective

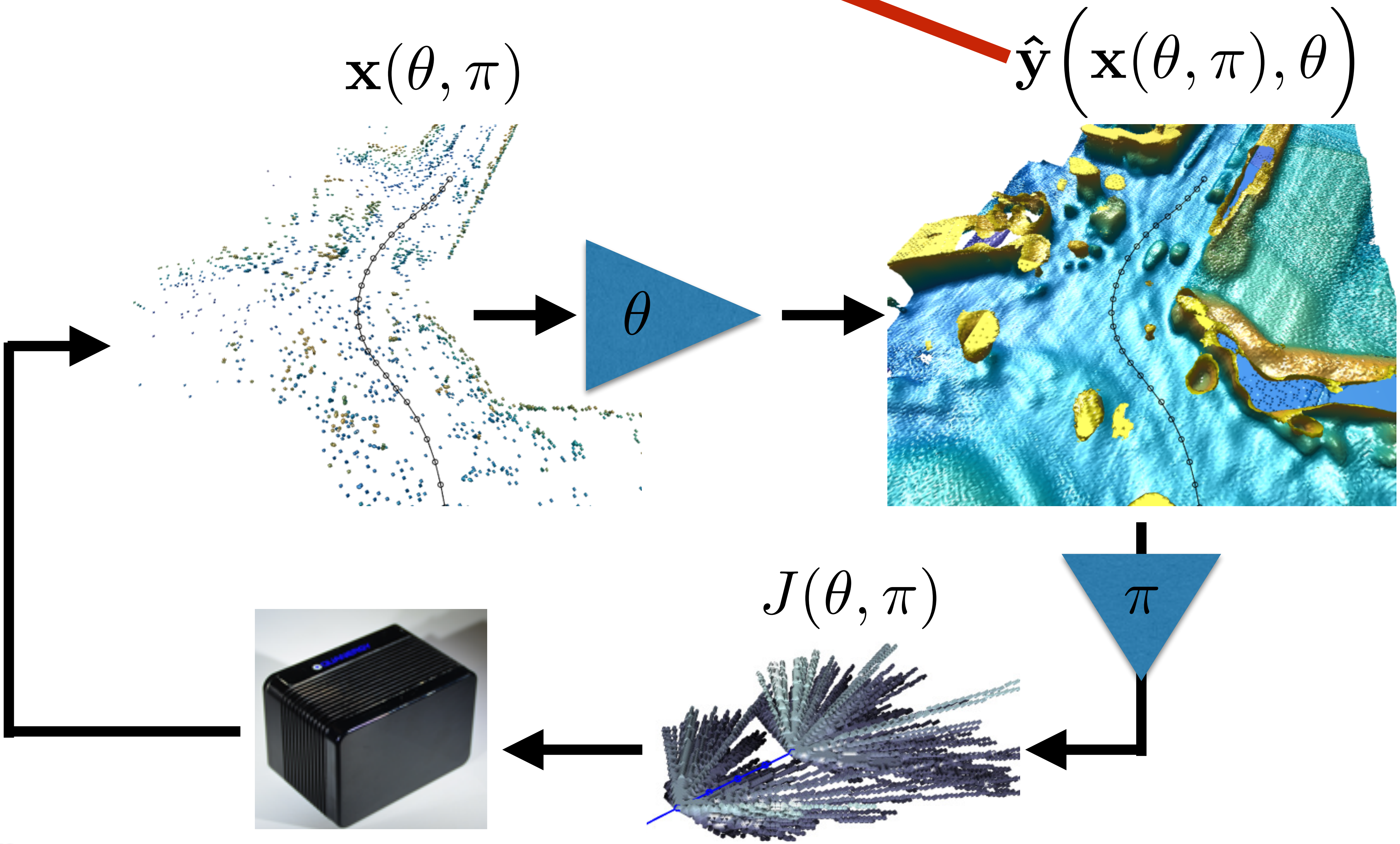


Mapping and planning minimize common objective



Mapping and planning minimize common objective

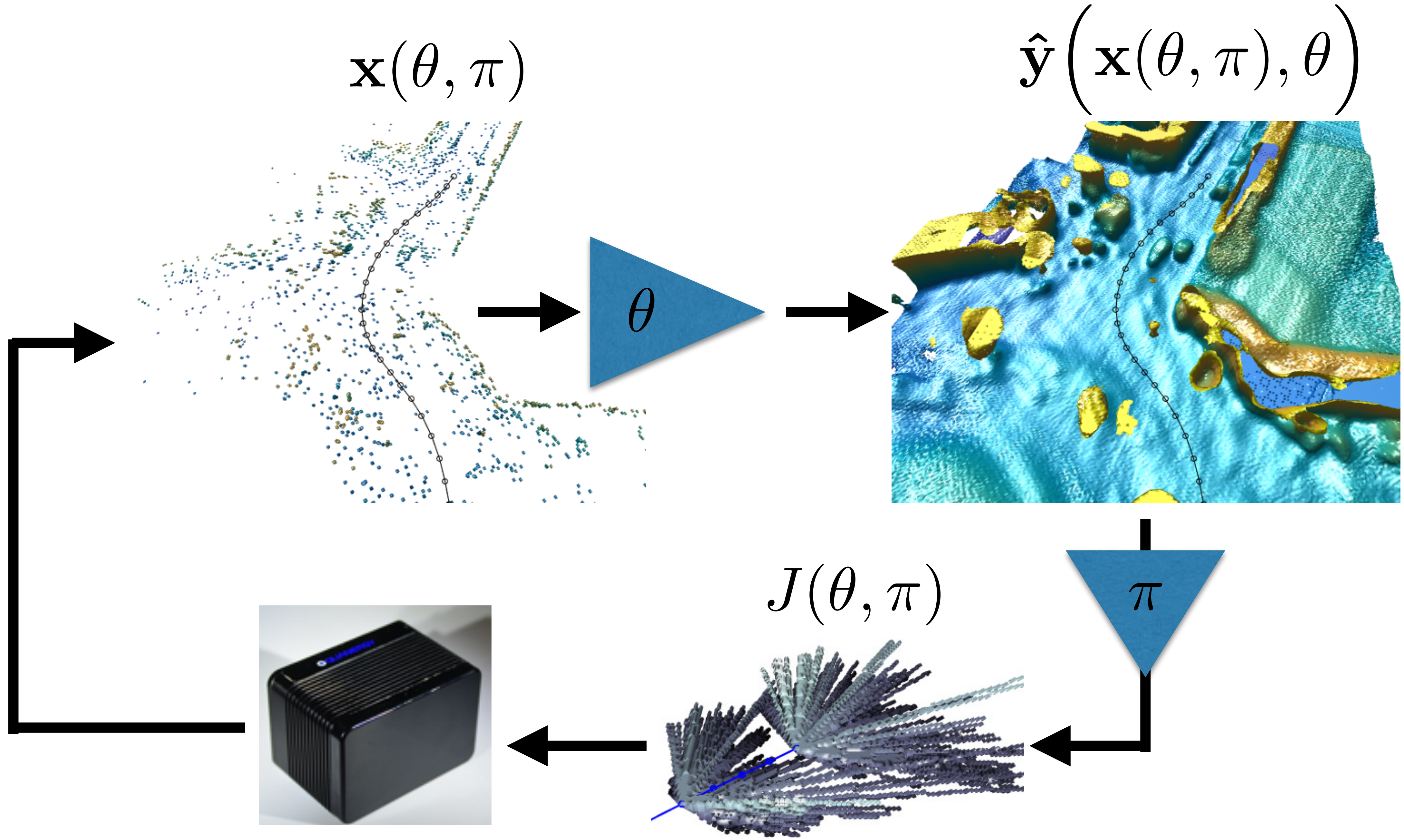
$$\arg\min_{\theta,\pi} \quad \mathcal{L}\big(\hat{\mathbf{y}}(\mathbf{x}(\theta,\pi),\theta),\ \mathbf{y}\big)$$



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

Mapping and planning minimize common objective

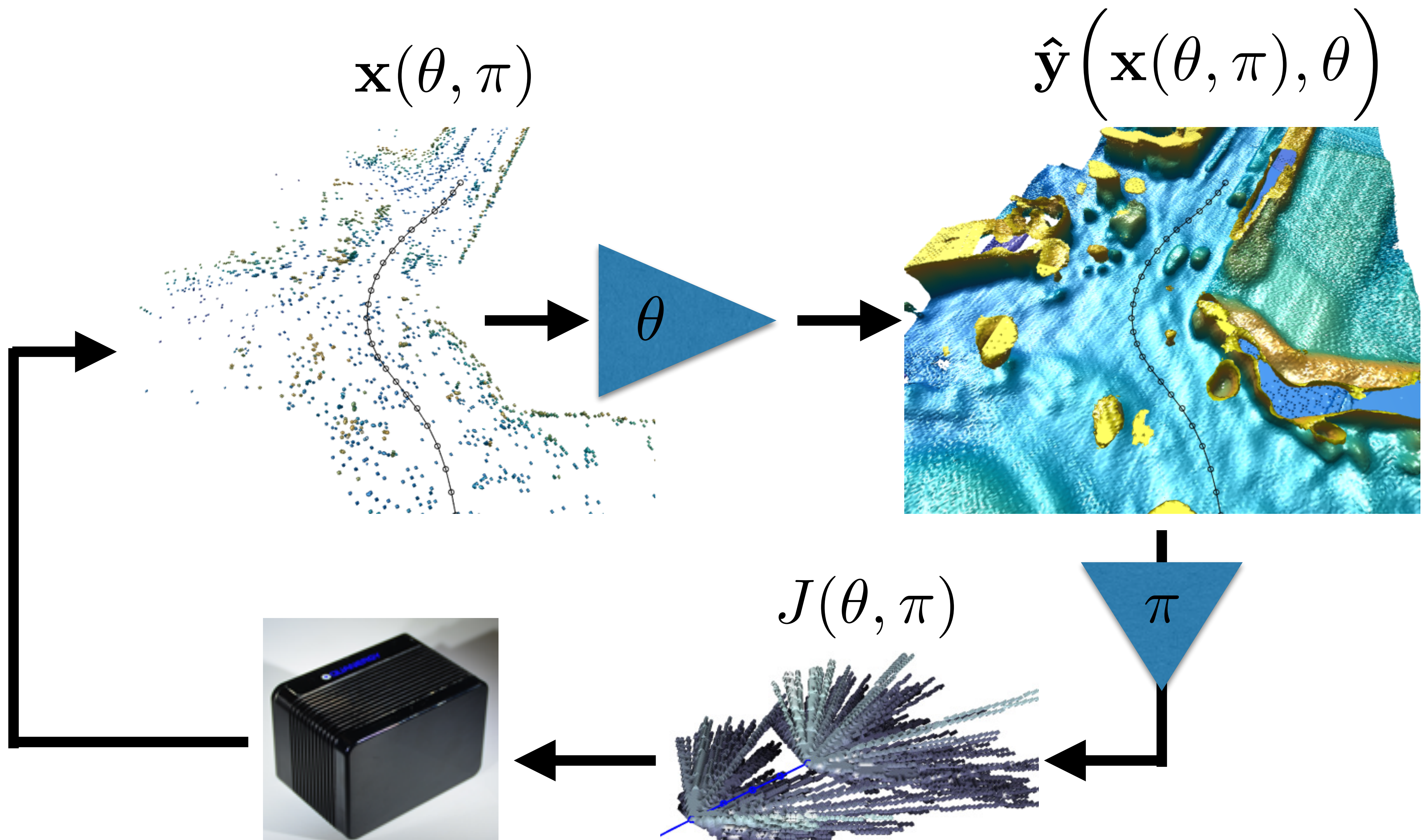
$$\arg\min_{\theta,\pi} \sum_p \mathcal{L}(\hat{\mathbf{y}}(\mathbf{x}_p(\theta,\pi),\theta),\ \mathbf{y}_p)$$



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

Mapping and planning minimize common objective

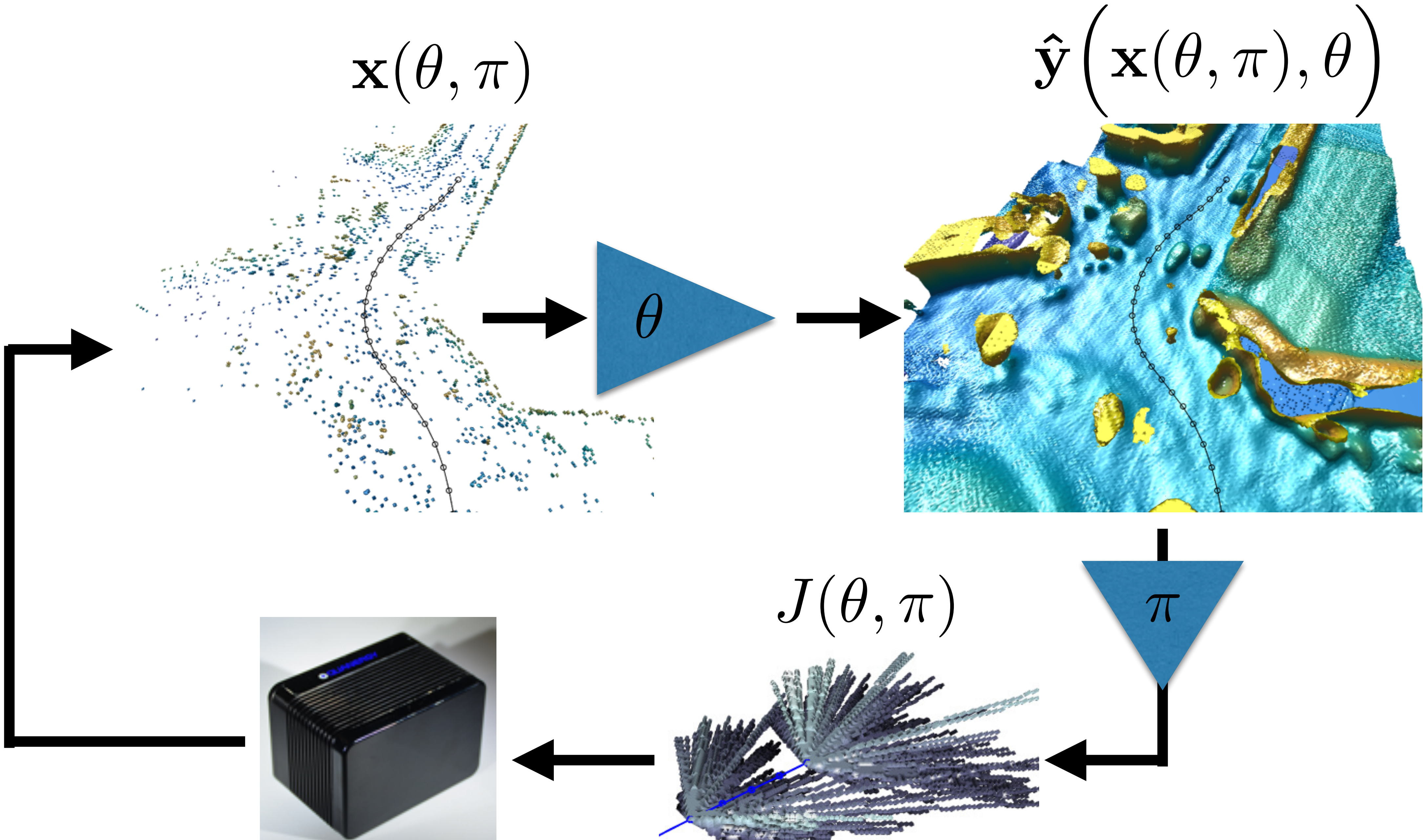
$$\arg\min_{\theta,\pi} \sum_{p} \mathcal{L}(\hat{\mathbf{y}}(\mathbf{x}_p(\theta,\pi),\theta),\ \mathbf{y}_p) \ \text{ subject to } |J_p(\theta,\pi)| \le K$$



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

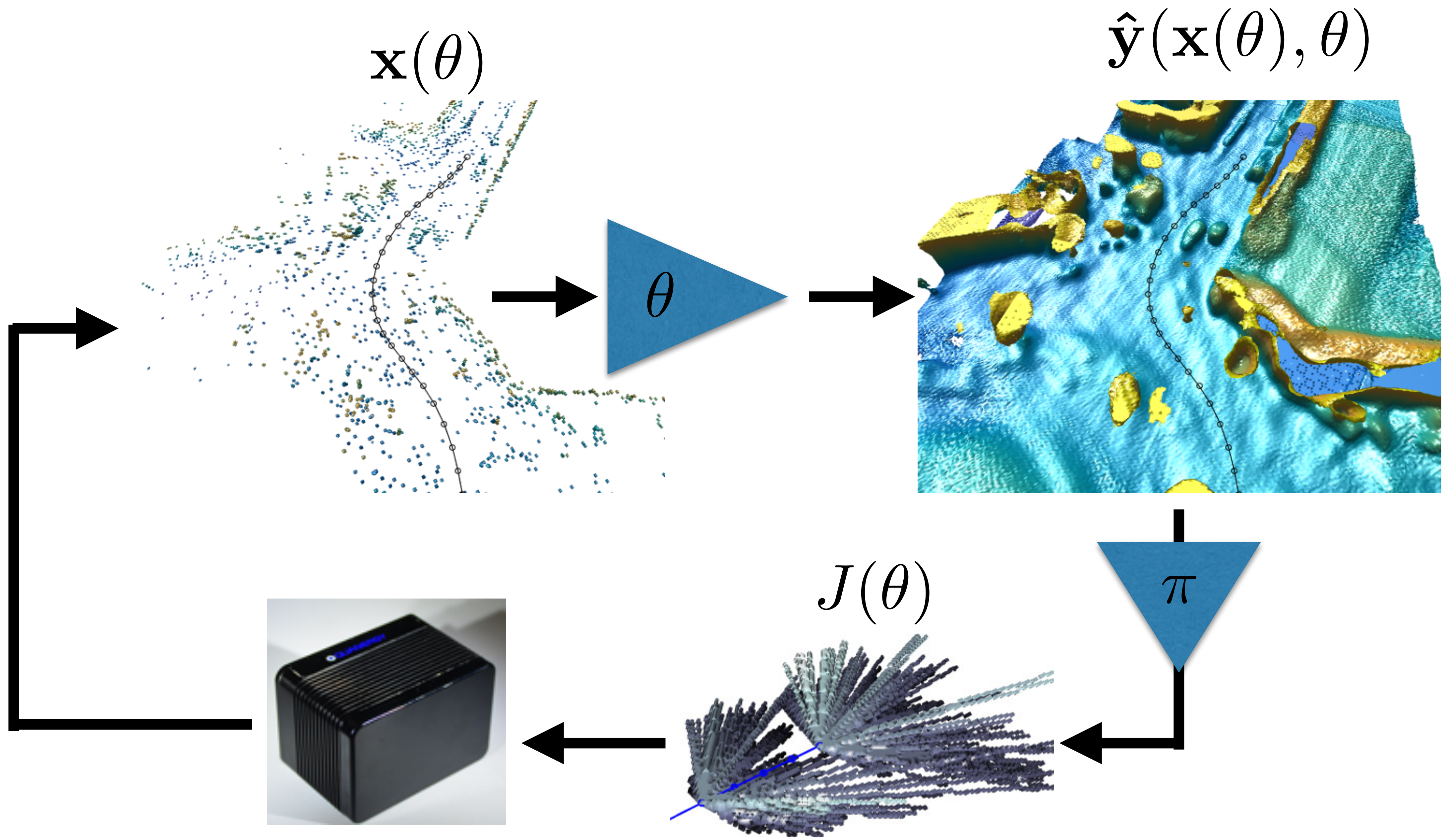
Mapping and planning minimize common objective

- Result of planning is uniquely determined by the reconstructed map
  => π is a deterministic function of theta



Mapping and planning minimize common objective

- Result of planning is uniquely determined by the reconstructed map
  => π is a deterministic function of theta

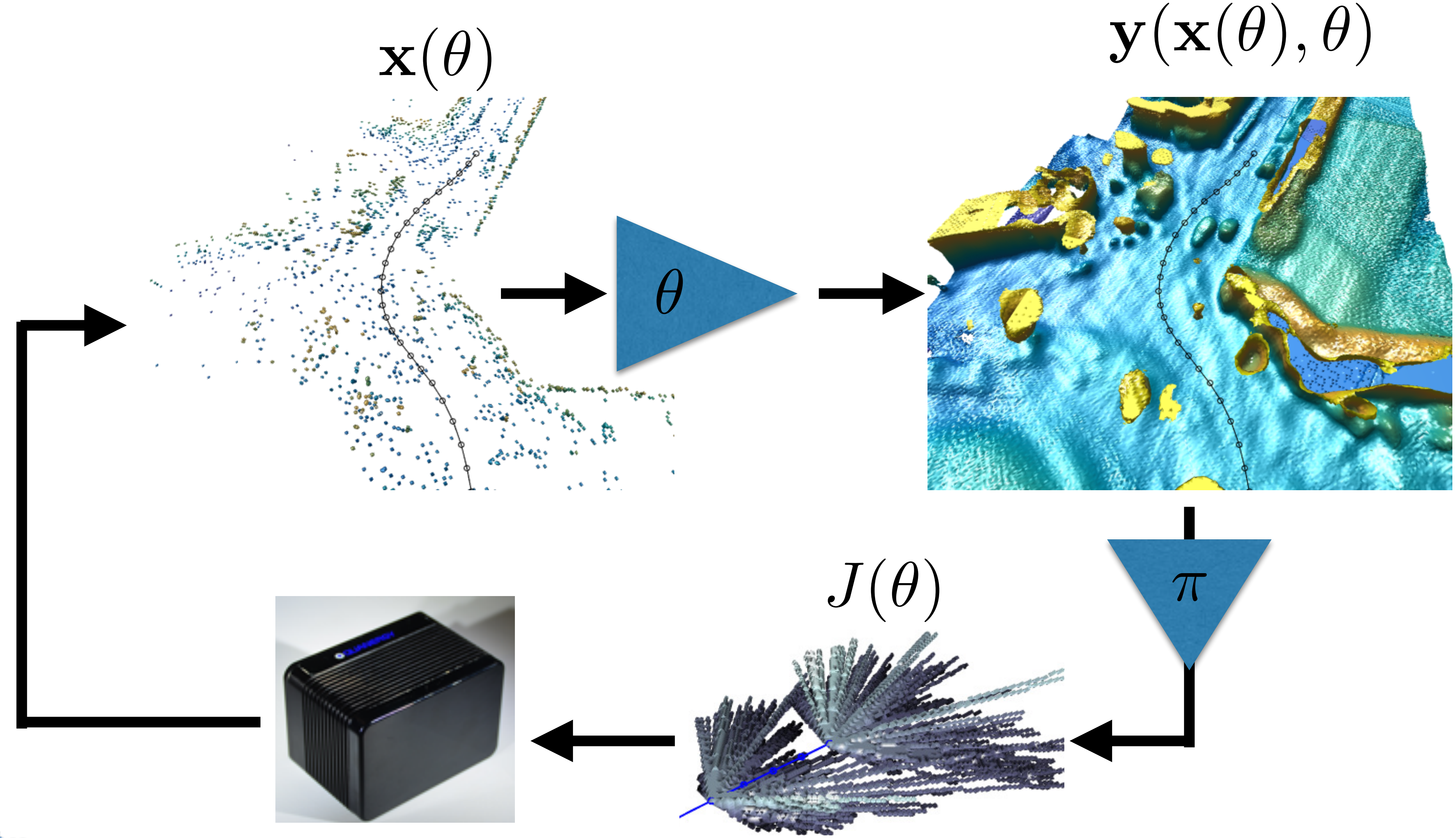


Czech Technical University in Prague
Faculty of Electrical Engineering, Department of Cybernetics

Learning as minimization over θ

$$\arg\min_{\theta} \sum_{p} \mathcal{L}(\hat{\mathbf{y}}(\boxed{\mathbf{x}_p(\theta)}, \theta),\ \mathbf{y}_p)$$

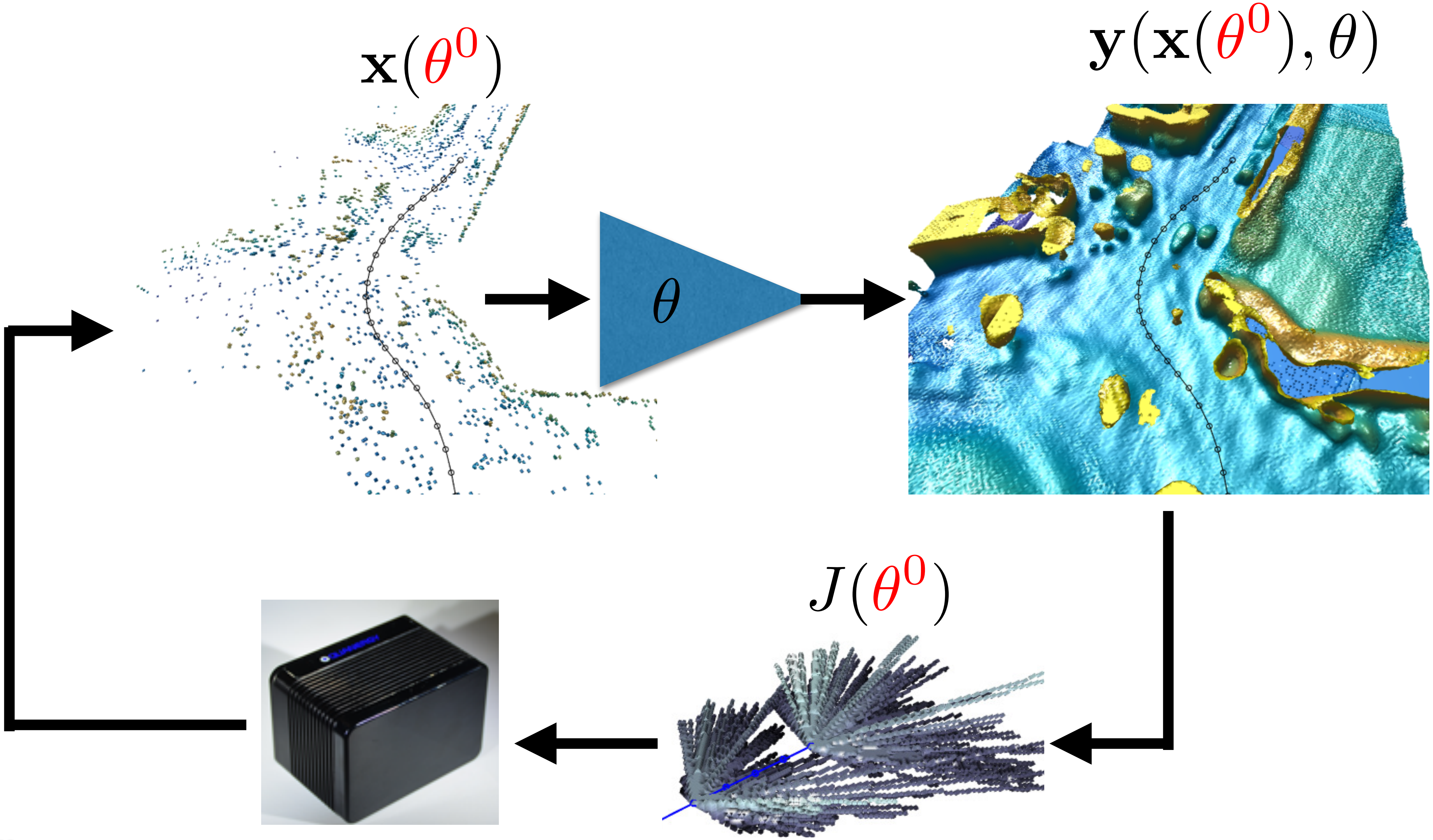
Result of planning is not differentiable



Locally approximate objective around θ^0

$$\arg\min_{\theta} \sum_{p} \mathcal{L}(\hat{\mathbf{y}}(\boxed{\mathbf{x}_p(\theta^0)}, \theta), \boxed{\mathbf{y}_p})$$

(fixed sparse input, ground truth)



Minimize approximated objective to get θ^1

$$\theta^1 = \arg\min_{\theta} \sum_{p} \mathcal{L}(\hat{\mathbf{y}}(\mathbf{x}_p(\theta^0), \theta),\ \mathbf{y}_p)$$

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

Minimize approximated objective to get θ^1

$$\arg\min_{\theta} \sum_p \mathcal{L}(\hat{\mathbf{y}}(\mathbf{x}_p(\theta^1), \theta),\ \mathbf{y}_p)$$

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

Minimize approximated objective to get θ^1

$$\theta^2 = \arg\min_{\theta} \sum_{p} \mathcal{L}(\hat{\mathbf{y}}(\mathbf{x}_p(\theta^1), \theta),\ \mathbf{y}_p)$$

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

Iteratively optimize approximated objective

$$\theta^{t+1} = \arg\min_{\theta} \sum_{p} \mathcal{L}(\hat{\mathbf{y}}(\mathbf{x}_p({\color{red}\theta^t}), \theta),\ \mathbf{y}_p)$$

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

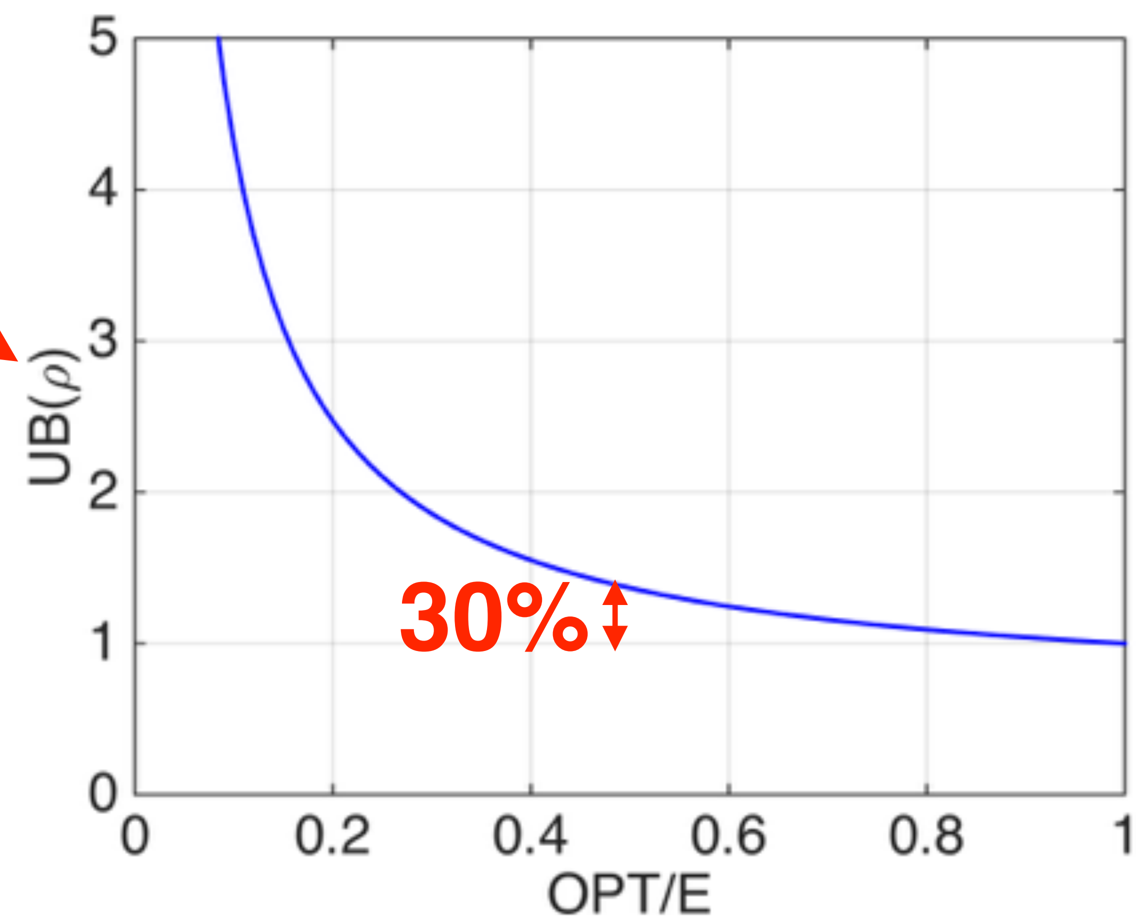
- Fix point of this mapping would assure:
  - local optimality of the objective
  - statistical consistency of the learning

- In practise, we iterate until validation error stops decreasing

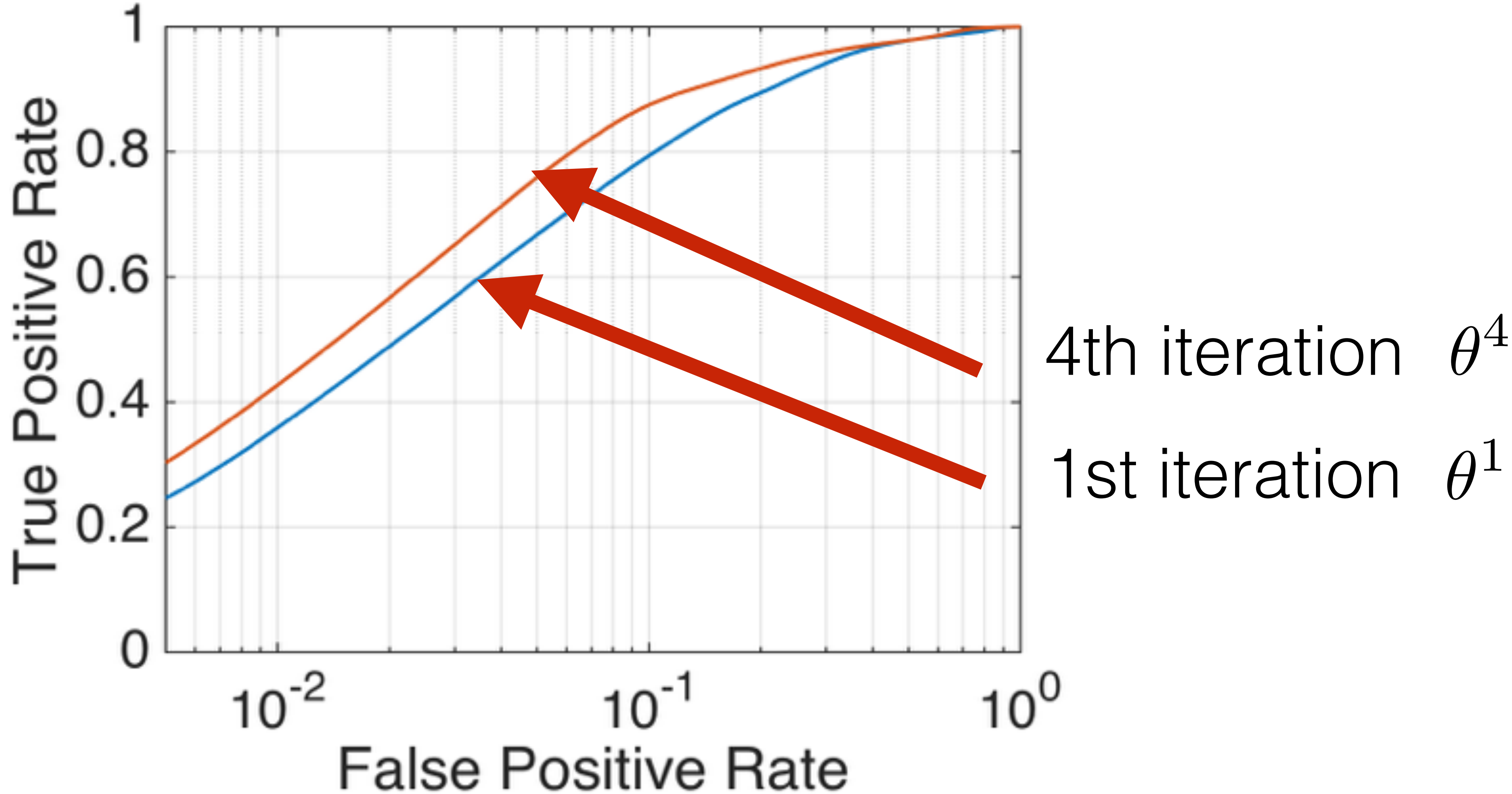
Planning of depth measuring rays J

- Objective: expected entropy under simplified visibility
- Optimization over huge number of rays is complicated
- Proposed novel planning algorithm => approx. solution
- Derive approximation ratio

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

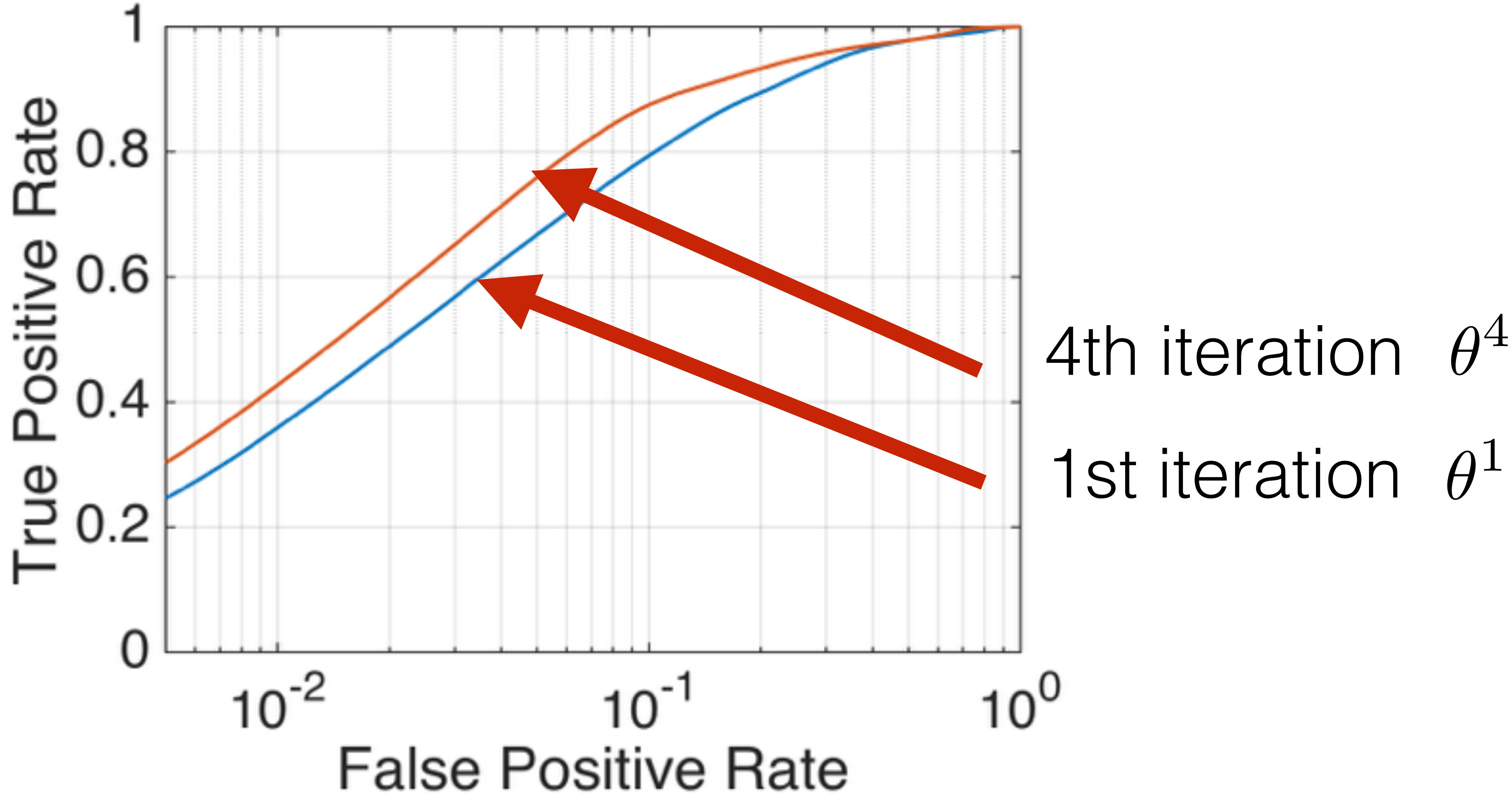


Experiment: Quantitative evaluation on full dataset



- Training distribution converges to testing distribution

Experiment: Quantitative evaluation on full dataset



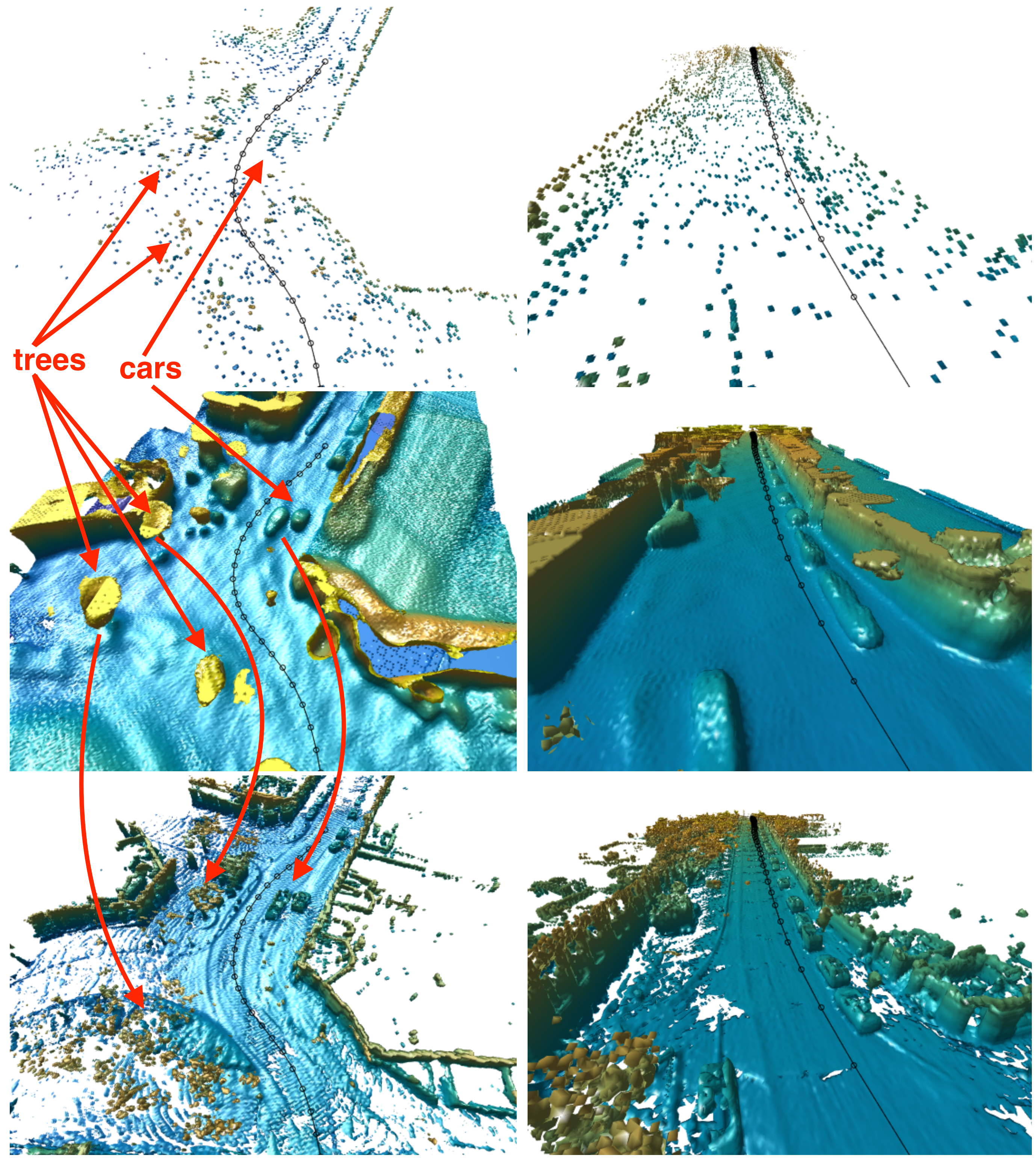
- Training distribution converges to testing distribution
- Training: 20 seq. from "Kitty:*Residential category*"
- Testing: 13 seq. from "Kitty:*City category*"
- Local maps 320x320x32 voxels (1 voxel ~ 20cm)
- Selected 200 rays per position out of 20k, horizon 5 pos.

Experiment: Qualitative evaluation

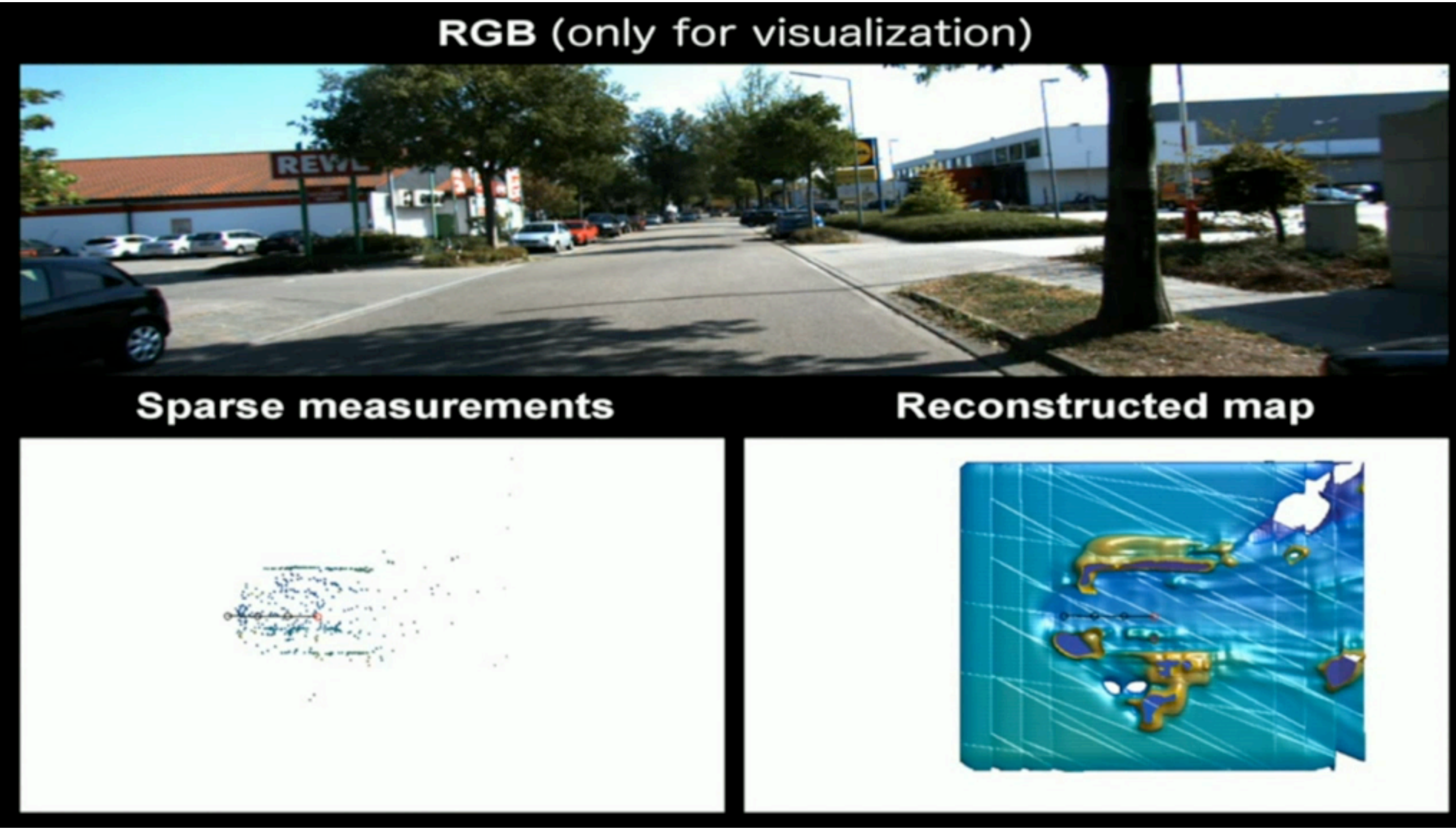
Sparse measurements

Reconstructed map

Ground truth



Experiment: Summary & Questions



[5] Zimmermann, Petricek, Salansky, Svoboda, Learning for Active 3D Mapping, **ICCV oral (rank A*, AC=2%)**, 2017

Conclusions

- We have worked on active 3D voxel mapping with steerable solid state lidar.
- Future research depends on physical existence of the the device (despite of 2 year of communication with Quanergy, device is not yet availble).
- Potential cooperation on the development of a novel device.
- Since the results of
  A. deep ConvNet **classification** are strongly dependent on the amount (and correctness) of annotated data,
  B. deep ConvNet **control** are strongly dependent on the amount of environment trials can be played

  we have recently started to use reversed-engineered GTAV.

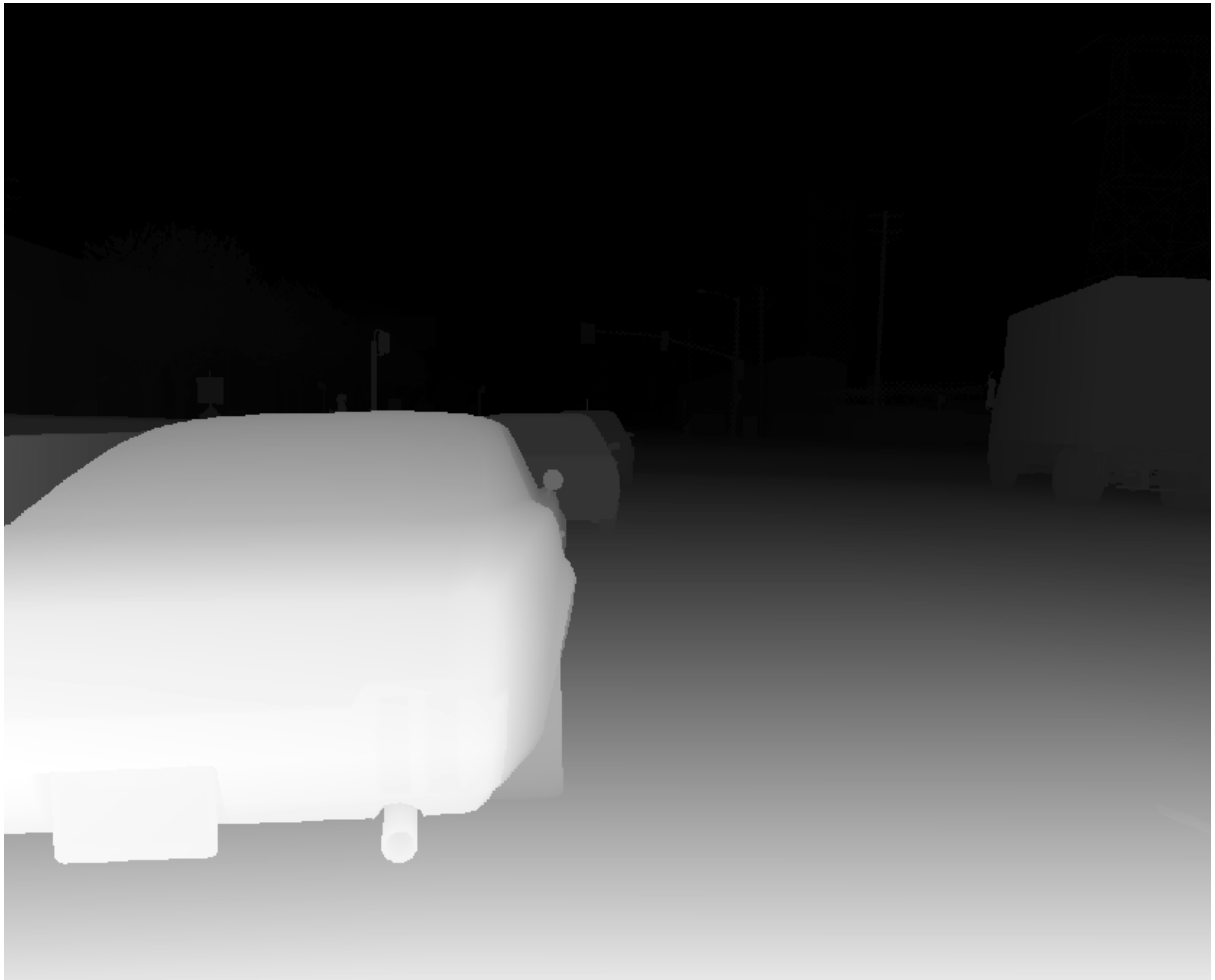
Motivation

- State-of-the-art algorithms based on deep-learning
- Deep learning is data-hungry.
- Exploit state-of-the-art game engines for vision
- Reverse engineering of GTA 5 (RAGE engine)

RGB images

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

Depth images



Stencil layer

Stencil layer - **cars**

Stencil layer - **humans**

Stencil layer - **trees**

Stencil layer - **sky**

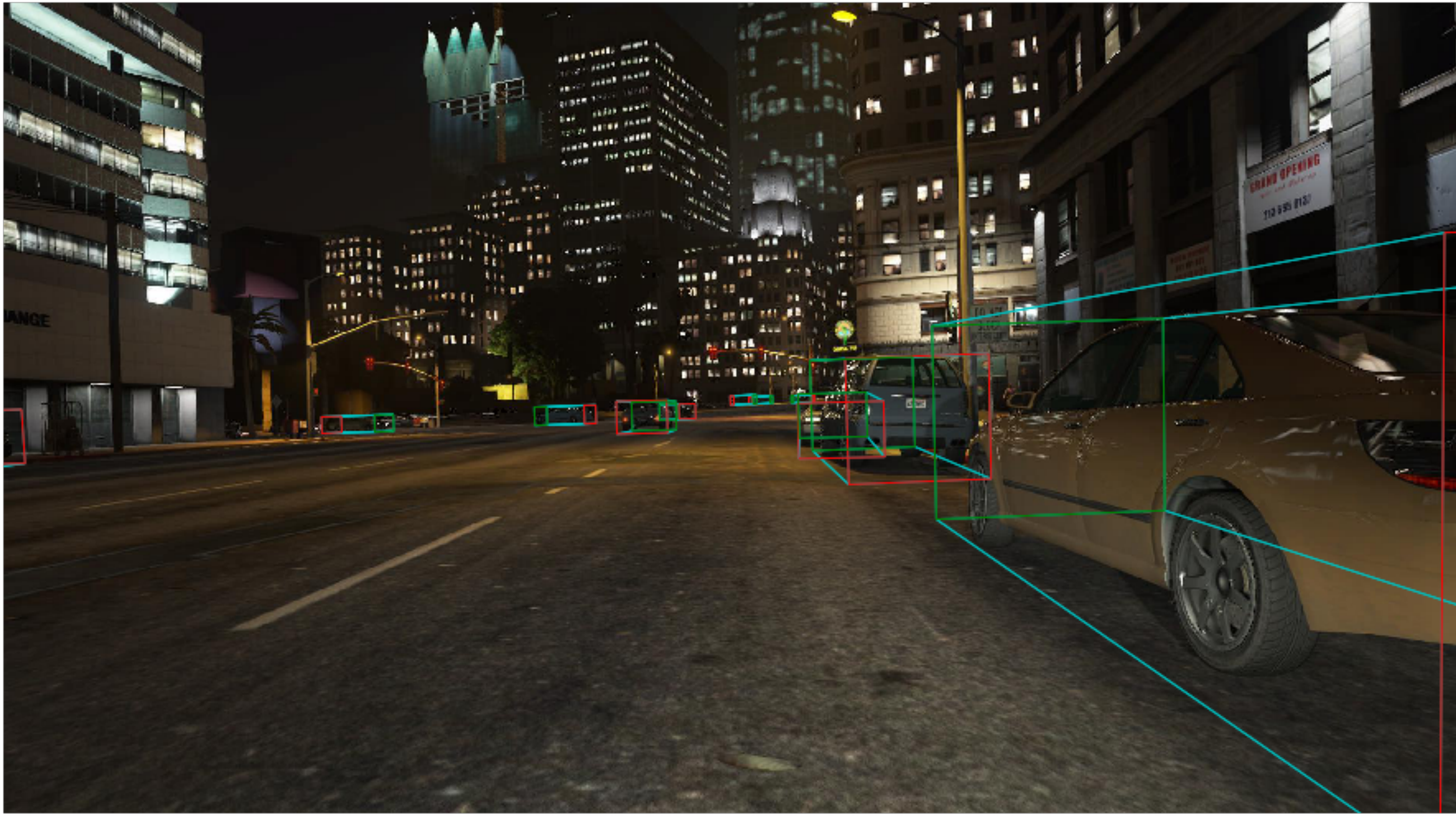
Stencil layer - **artificial light**

Stencil layer - **artificial light**

Traversability of pixels

Traversability of pixels

3D bounding boxes for objects (e.g. cars, humans)



Distinguish pixels of different entities (e.g. cars, humans)

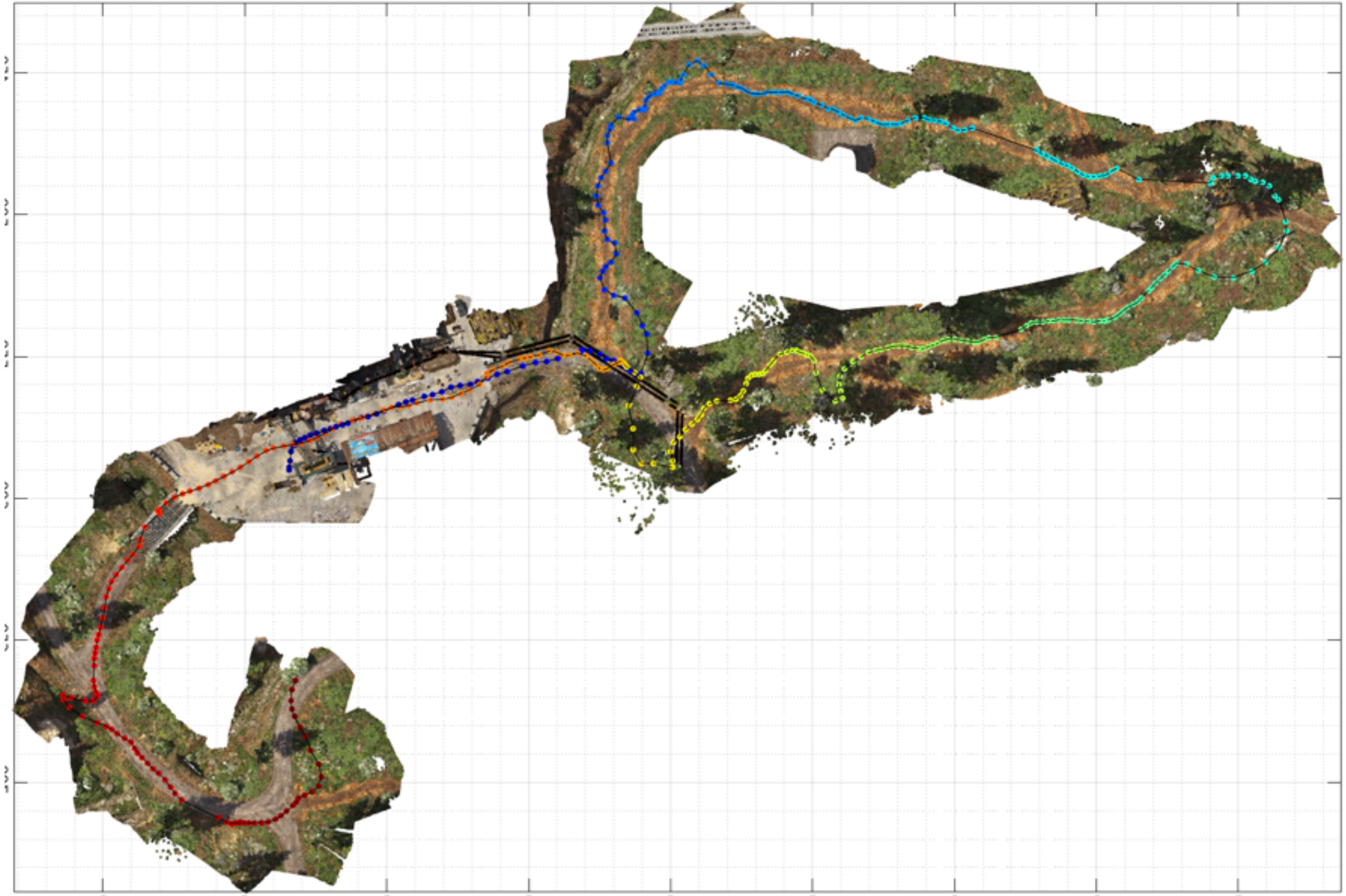
Annotations for objects (e.g. cars, humans)

- 2D bounding box in Image Coordinates (IC)
- 3D bounding box in World Coordinates (WC)
- Position and rotation in WC
- Entity type (e.g. car, pedestrian)
- Entity class (e.g. sedan, SUV, coupe …)
- Unique ID (create trajectories, estimate motion flow)

What can be controlled

- **Motion of objects:**
  - explicit (e.g. shift in WCF),
  - implicit (e.g. car driving, autopilot)

Autopilot



What can be controlled

- Motion of objects:
  - explicit (e.g. shift in WCF),
  - implicit (e.g. car driving, autopilot)
- **Time in the day/night cycle**

What can be controlled

What can be controlled

What can be controlled

- Motion of objects:
  - explicit (e.g. shift in WCF),
  - implicit (e.g. car driving, autopilot)
- Time in the day/night cycle
- **Weather** (ExtraSunny, Clear, Clouds, Smog, Foggy, Overcast, Raining, ThunderStorm, Clearing, Neutral, Snowing, Blizzard, Snowlight, Christmas, Halloween)

ExtraSunny

Clear


handlePipeInput called
server connected:False
N pressed, going to take screenshots
connection:System.Net.Sockets.Socket

Foggy


handlePipeInput called
server connected:False
connection:System.Net.Sockets.Socket

OverCast


handlePipeInput called
server connected:False
connection:System.Net.Socke
ts.Socket
N pressed, going to take
screenshots

Raining

handlePipeInput called
N pressed, going to take screenshots
server connected:False
connection:System.Net.Sockets.Socket

ThunderStorm

Clearing


handlePipeInput called
server connected:False
connection:System.Net.Socke ts.Socket
N pressed, going to take screenshots

SnowLight


handlePipeInput called
N pressed, going to take screenshots
server connected:False
connection:System.Net.Sockets.Socket

What can be controlled

- Motion of objects:
  - explicit (e.g. shift in WCF),
  - implicit (e.g. car driving, autopilot)
- Time in the day/night cycle
- Weather (ExtraSunny, Clear, Clouds, Smog, Foggy, Overcast, Raining, ThunderStorm, Clearing, Neutral, Snowing, Blizzard, Snowlight, Christmas, Halloween)
- Visual mods (Redux, NaturalVision, Vanilla)
- **Custom object models from CAD**

What can be controlled

- Motion of objects:
  - explicit (e.g. shift in WCF),
  - implicit (e.g. car driving, autopilot)
- Time in the day/night cycle
- Weather (ExtraSunny, Clear, Clouds, Smog, Foggy, Overcast, Raining, ThunderStorm, Clearing, Neutral, Snowing, Blizzard, Snowlight, Christmas, Halloween)
- Visual mods (Redux, NaturalVision, Vanilla)
- Custom vehicle models from CAD
- **Custom maps a scenarios** (probability of spawning different objects in different areas, complex scripts)

Ongoing research

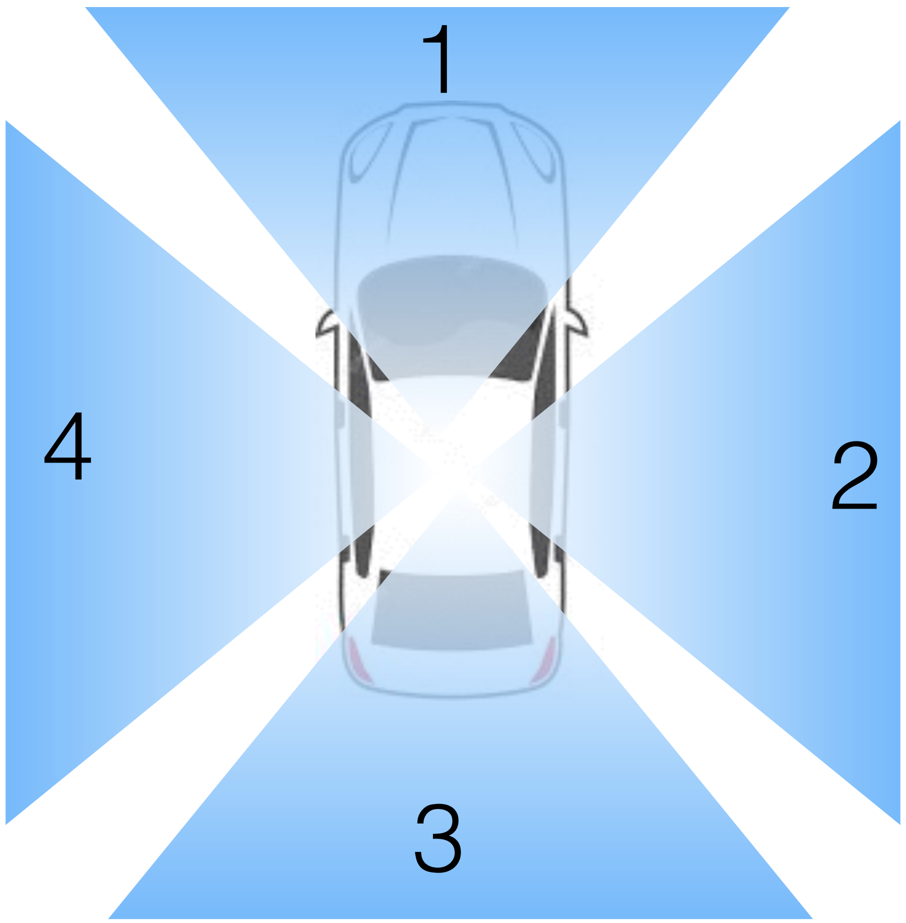
- GTA 5 allows to create huge annotated datasets, which provably improves accuracy on real images.

Conslusions and ongoing research

- GTA 5 allows to create huge annotated datasets, which provably improves accuracy on real images.
- **Realistic simulation of lidar measurements**
  - data-driven approach employs Valeo’s lidar dataset
  - based on cycle-GANs (CVPR 2017)

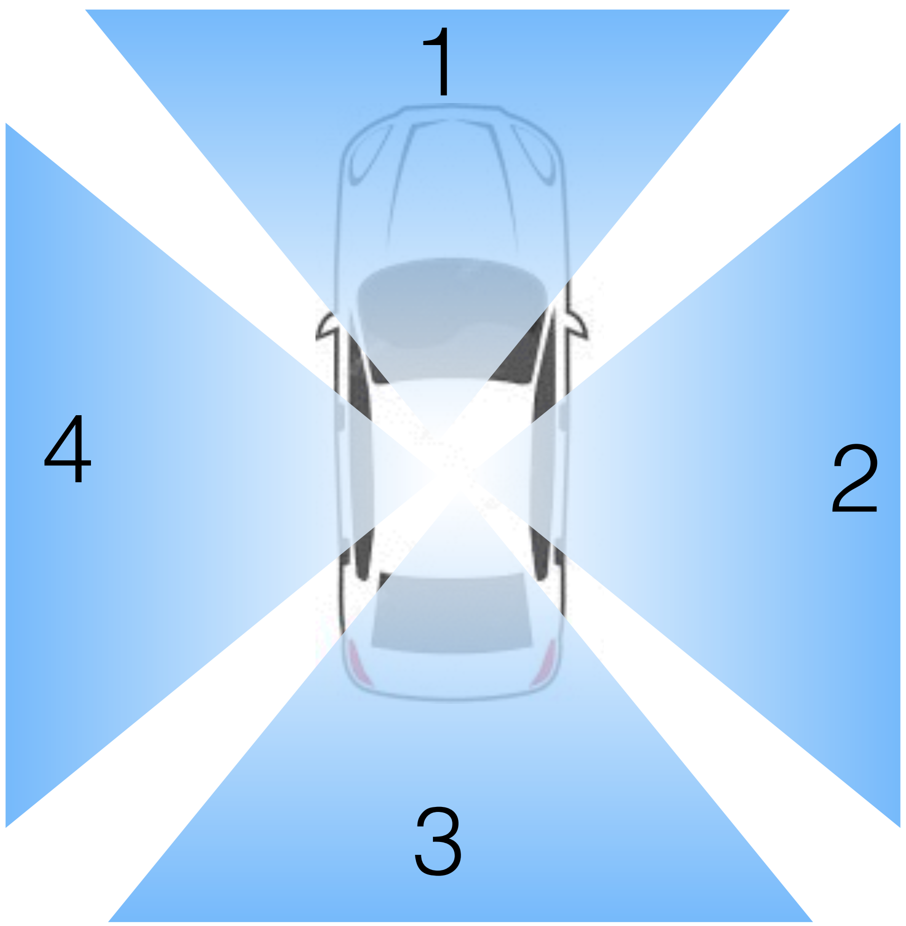
Input (4 virtual roof-mounted cameras)

virtual car in GTA environment

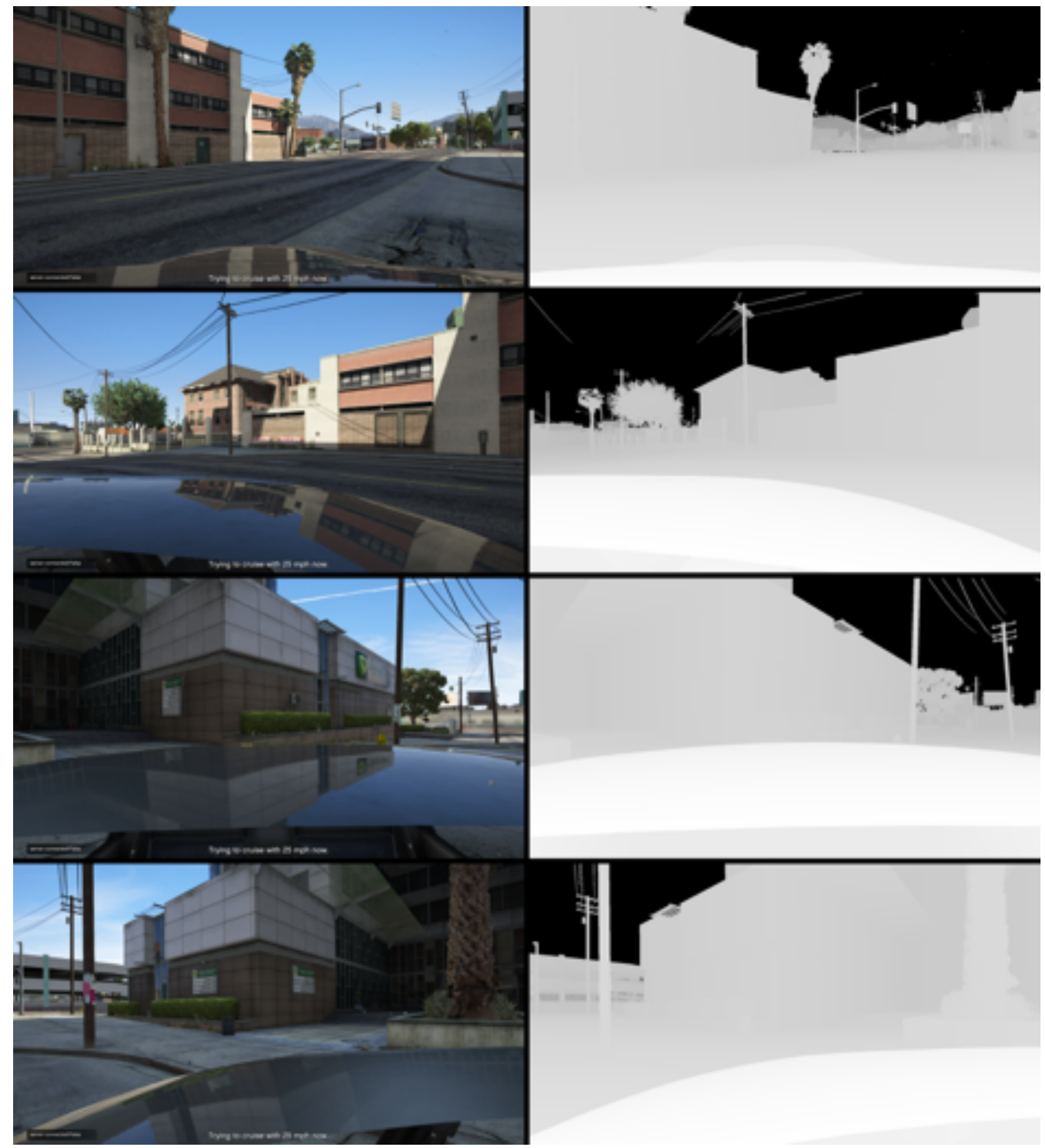


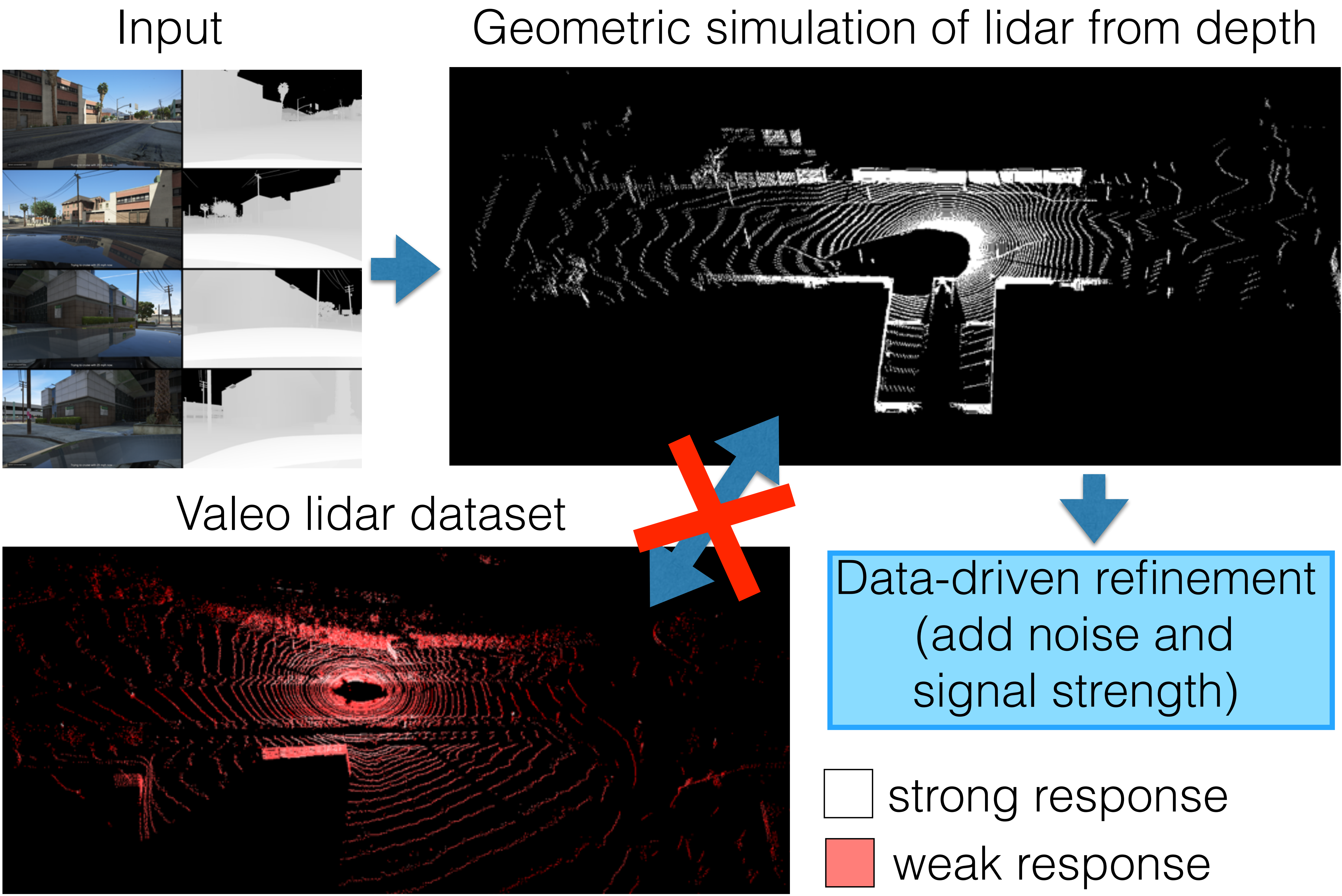
Input (4 virtual roof-mounted cameras)

virtual car in GTA environment



ideal RGBD images





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

Conslusions and ongoing research

- GTA 5 allows to create huge annotated datasets, which provably improves accuracy on real images.
- Realistic simulation of lidar measurements
  - data-driven approach employs Valeo’s lidar dataset
  - based on cycle-GANs (CVPR 2017)
- **Potential cooperation:**
  - We offer to help with setting up the reversed engineered GTAV and the lidar simulator.

Conslusions and ongoing research

- GTA 5 allows to create huge annotated datasets, which provably improves accuracy on real images.
- Realistic simulation of lidar measurements
  - data-driven approach employs Valeo's lidar dataset
  - based on cycle-GANs (CVPR 2017)
- Potential cooperation:
  - We offer to help with setting up the reversed engineered GTAV and the lidar simulator.
- **Antonín Svoboda Award for the Best Ph.D. Thesis**
  - http://svobodovacena.cz
  - 25.000 CZK for winner
  - 45% CVUT, 17% UK, 8% TUO,…