# Deep Learnable Reconstruction for 3D Computer Vision

**Alexey Artemov,** Researcher, Skoltech ADASE Team

**Evgeny Burnaev,** Associate Professor, Skoltech ADASE Team

**Denis Zorin,** Adjunct Professor, Skoltech CDISE & NYU,

with **Justin Solomon,** Assistant Professor, MIT EECS & CSAIL

and the **Team**

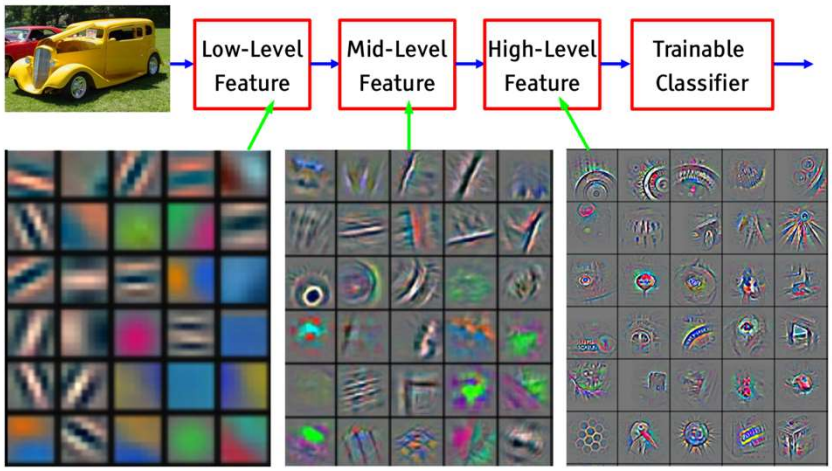
adase.group

# THE 3D GEOMETRY PROCESSING

Skoltech
Skolkovo Institute of Science and Technology

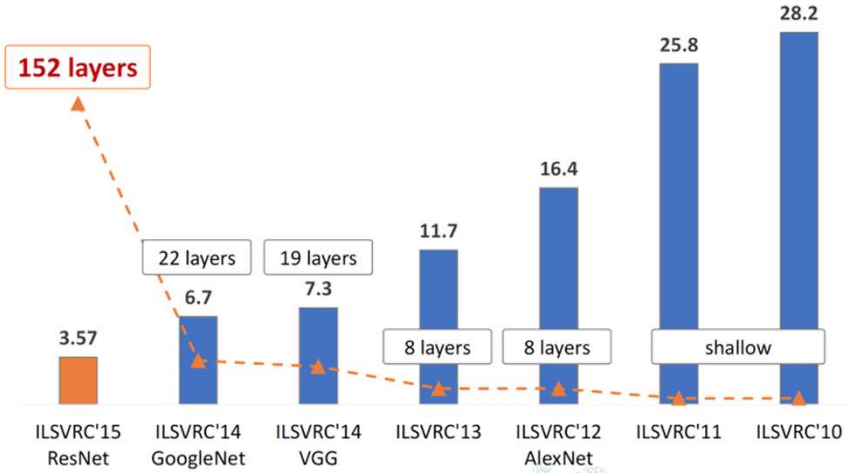
# Benefits of Learnable Methods for Computer Vision

- → A natural way to incorporate *and learn* hierarchical data representations
- → Widespread adoption for categorization, face recognition, medical imaging
- → Deep learning-powered “AI market” expected to grow to **$5 bln** by 2020[1]




Learnable hierarchy of representations. Picture credit: Yann LeCun

[1]Research by Markets and Markets: https://www.marketsandmarkets.com

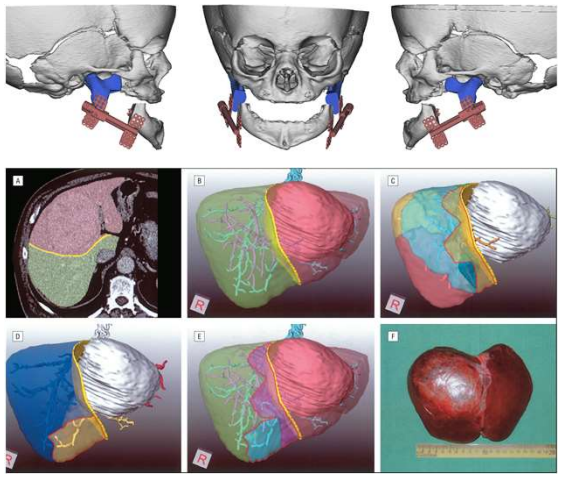



Recent progress on ILSVRC. Picture credit: Kaiming He et al.

Skoltech
Skolkovo Institute of Science and Technology

# 3D Geometry Processing Applications

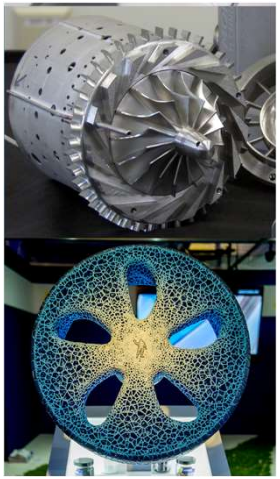
→ In the meantime: multitude of real-world & exciting applications for 3D/4D data processing!



**Virtual surgery planning**

**Urban planning/modeling**

**Cultural heritage preservation**



**Digital fabrication**

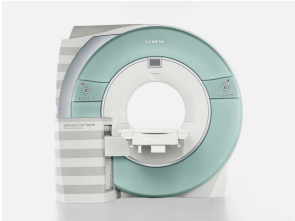
ADASE

Skoltech
Skolkovo Institute of Science and Technology

# Learnable Geometry Processing Needs More Research!

- Millions of depth sensors available on the market
- However, only scarce annotated datasets exist

Structured light



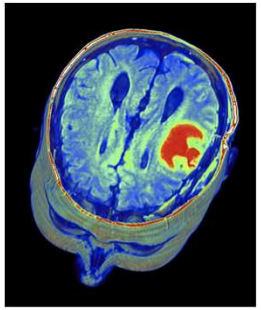
CT/MRI

LiDAR

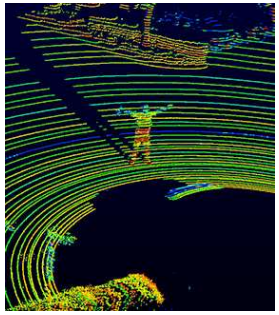
RGB-D sensors

- Traditional 2D ConvNets cannot solve *all* problems!
- Thus, 3D/4D and geometry-aware models are required!

Structured light scans



MRI scans



LiDAR range-scans

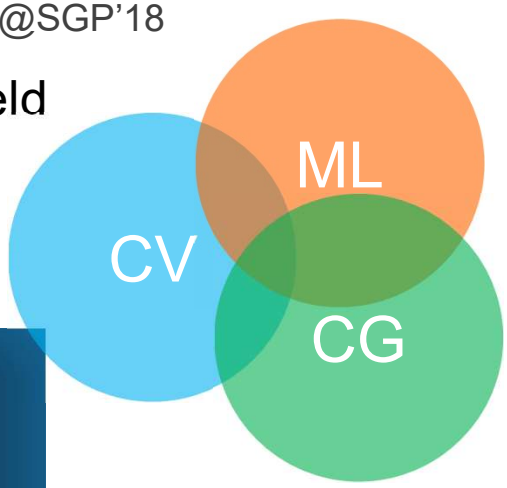
Depth maps

ADASE

Skoltech
Skolkovo Institute of Science and Technology

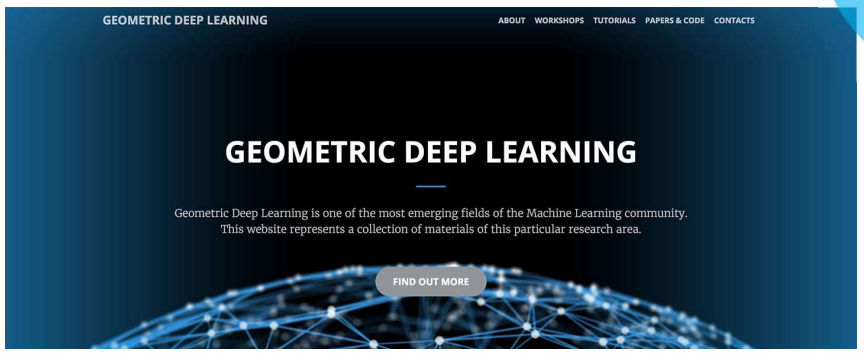
# Learnable Methods for Geometric Data Processing?

- Recent surge of interest in learnable methods for 3D geometry processing in leading events
  - Geometry Meets Deep Learning ECCV'16, 3D Deep Learning @NIPS'16, Geometric Deep Learning @SGP'18
- An emerging, interdisciplinary field
  - An intersection of **computer vision, machine learning and 3D graphics**




| year | # articles |
|---|---|
| 2012 | 410 |
| 2013 | 627 |
| 2014 | 1210 |
| 2015 | 2570 |
| 2016 | 6130 |
| 2017 | 14 100 |
| Q3 2018 | 17 500 |

Google Scholar searched for "3D" "Deep Learning" results

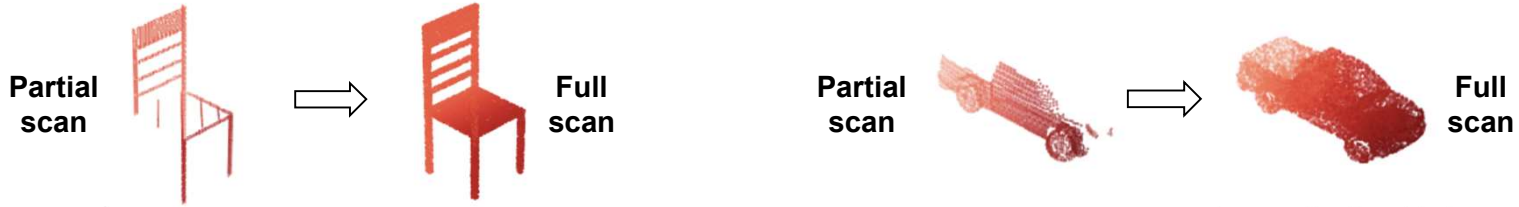



http://geometricdeeplearning.com

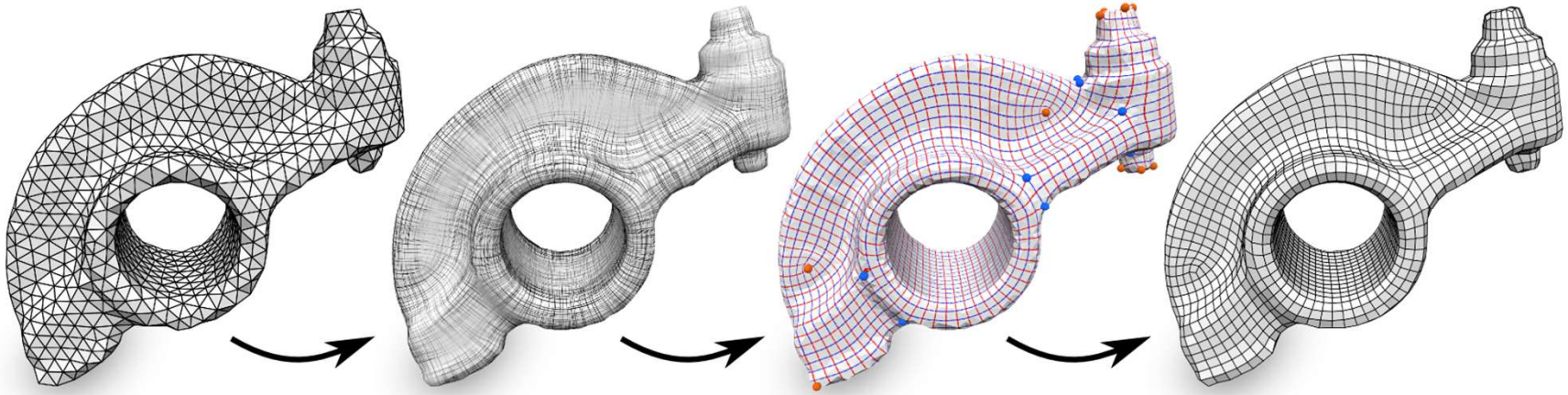
Skoltech
Skolkovo Institute of Science and Technology

# The Goals of this Project

- Much of 3D/4D data processing has been **optimization-based** (e.g. optimal transport, mixed integer optimization for meshing)
  - Learnability: push data-based domain knowledge to the model




- Develop effective **deep learning architectures** for multimodal 3D/4D data
- Develop **geometry-aware algorithms** for classification, segmentation, shape completion, upsampling, shape correspondence
- **Applications:** geometric data processing (range scans), MRI, remote sensing

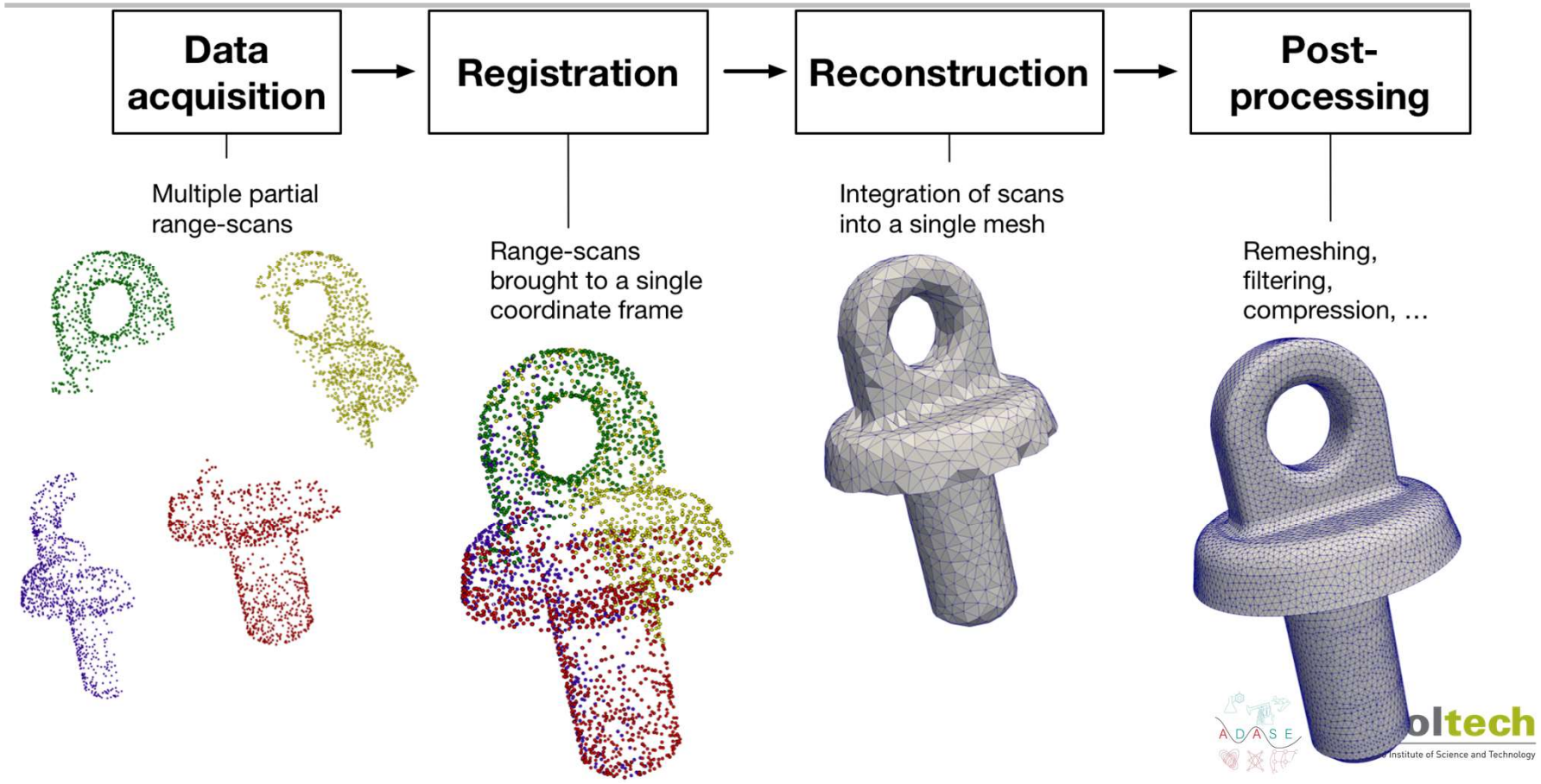


# THE ROADMAP

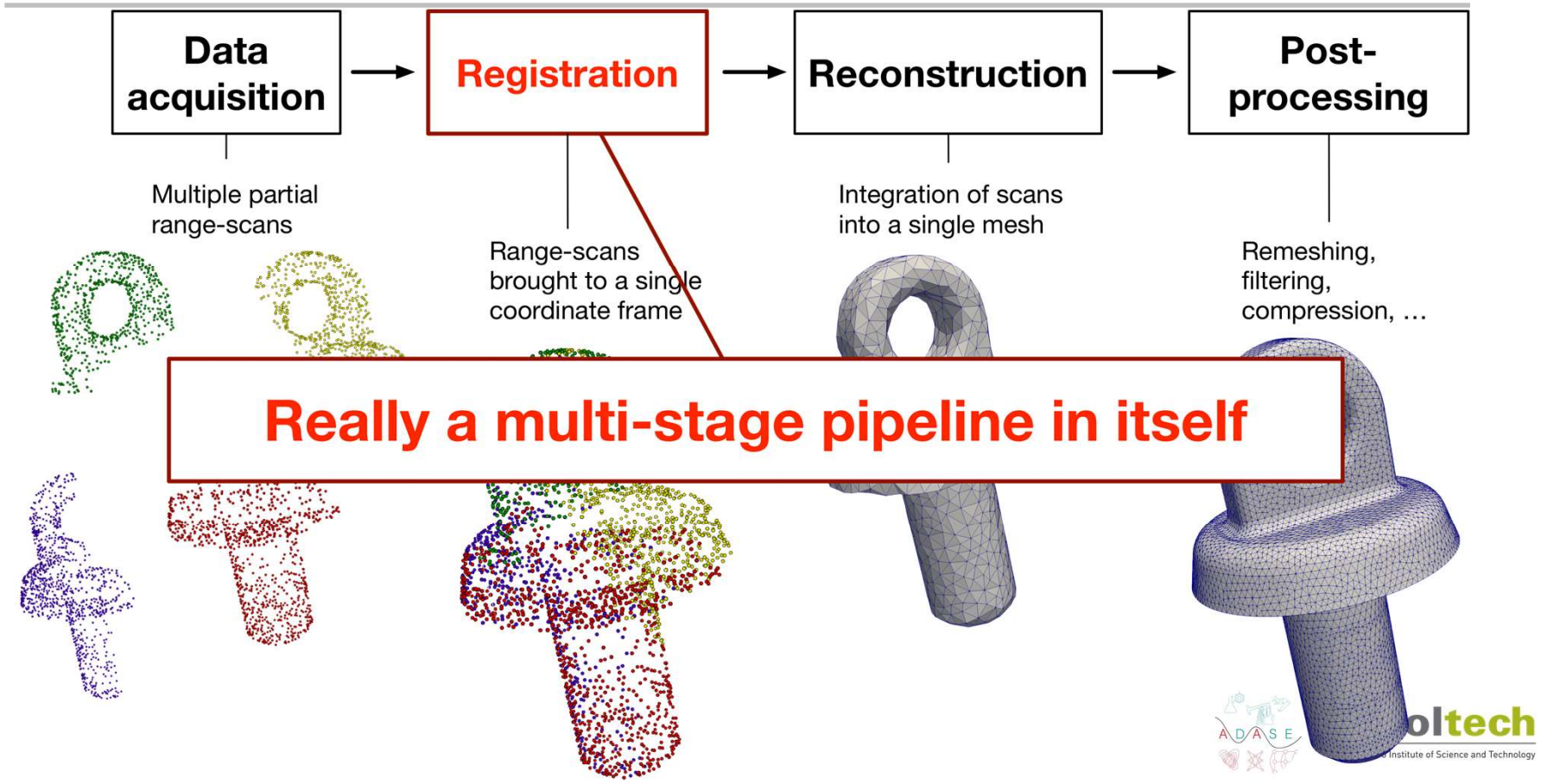
ADASE

Skoltech
Skolkovo Institute of Science and Technology

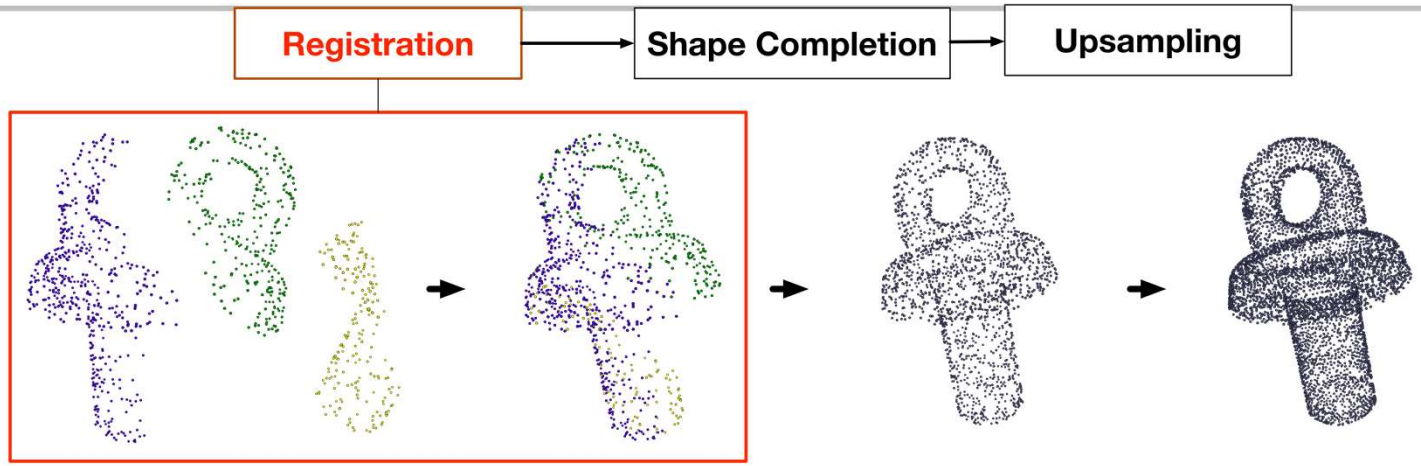
# Roadmap Inspired by the Geometry Processing Pipeline



# Roadmap Inspired by the Geometry Processing Pipeline



# Registration is Really a Multi-Stage Procedure

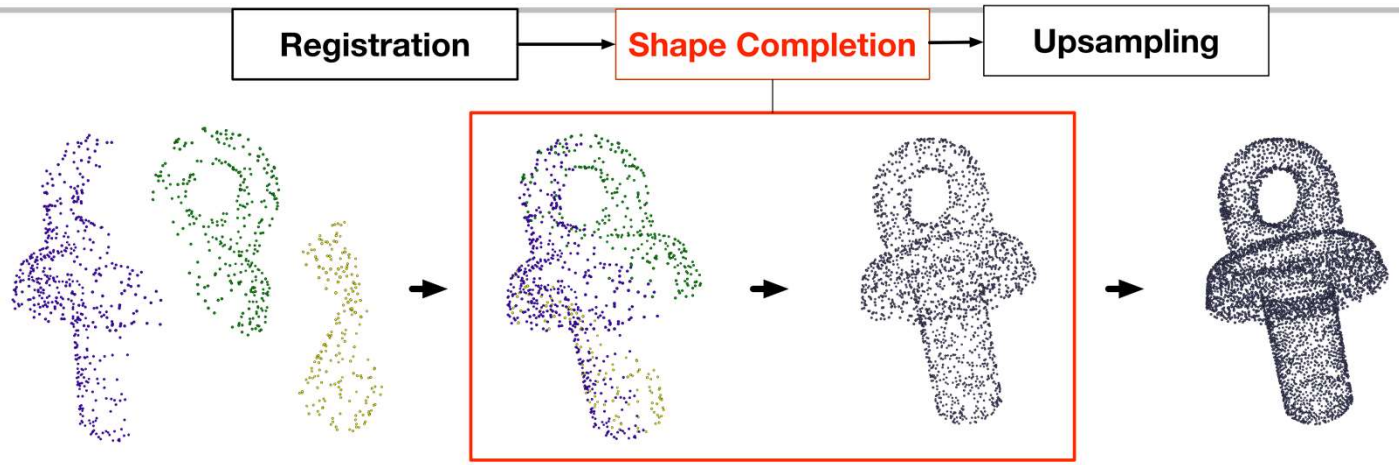



**Registration**:

- Given $M_1, \ldots, M_N$ partial range-scans, find transformations $T_2, \ldots, T_N$ s.t.

$$M_1 \cong T_2(M_2) \cong \ldots \cong T_N(M_N)$$

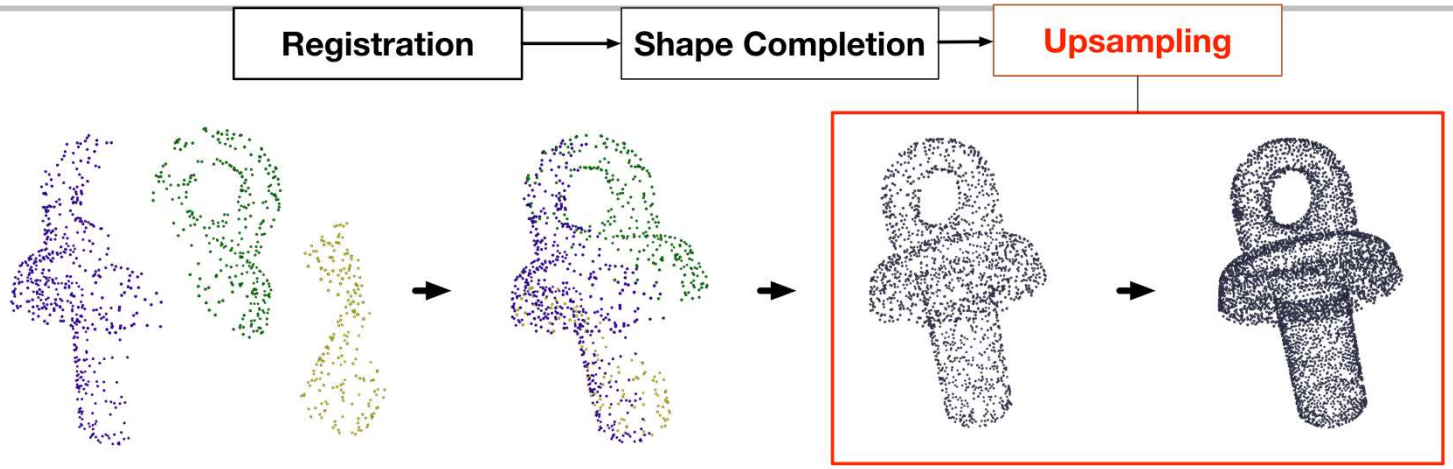
# Registration is Really a Multi-Stage Procedure




**Shape Completion:**

- Input: incomplete shape (some parts are missing)
- Output: complete shape

# Registration is Really a Multi-Stage Procedure




**Range-Image Upsampling:**

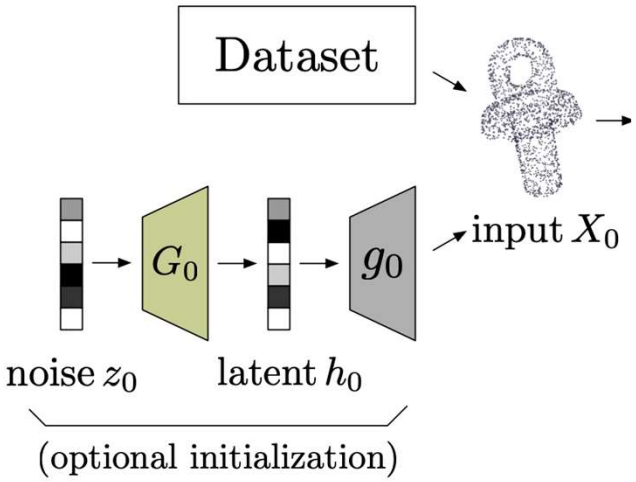
- Input: low-res representation
- Output: high-res representation

# Latent-Space Laplacian Pyramids for Adversarial Representation Learning with 3D Point Clouds

Latent Space Laplacian Pyramid GAN:

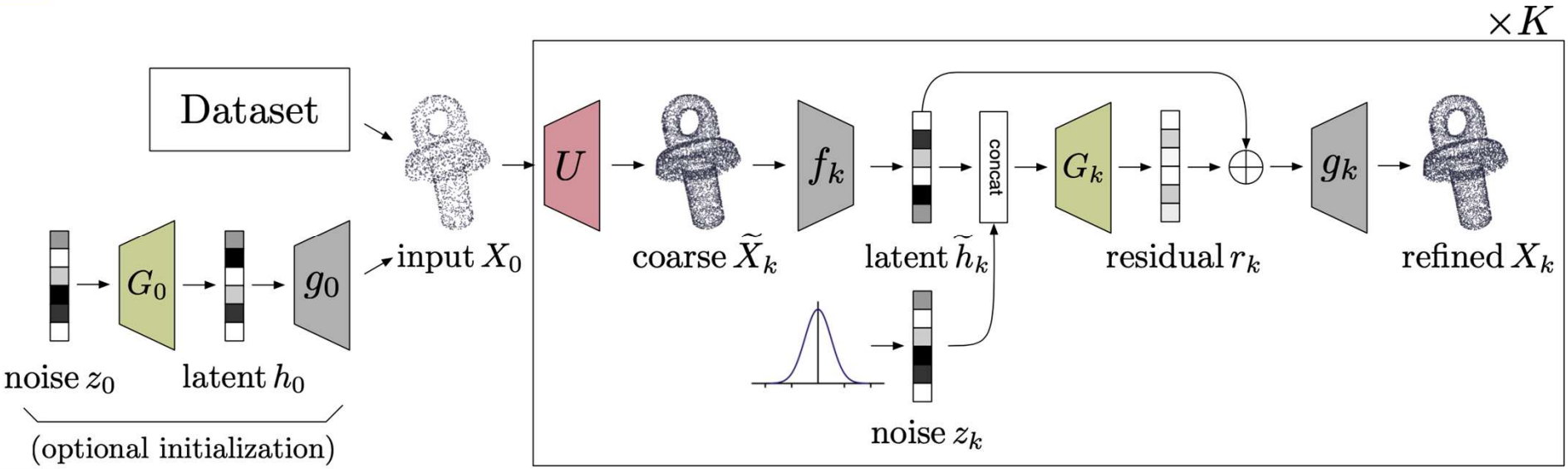
- it is generative and produces synthetic shapes unseen during training
- produces high-resolution point clouds via a latent space Laplacian pyramid representation
- it operates in the space of latent codes, bypassing the need to perform implicit dimensionality reduction during training

# Full architecture of LSLP-GAN model




- The network either accepts or generates an initial point cloud $X_0$
- Processes it with a series of $K$ learnable steps

# Full architecture of LSLP-GAN model




(1) upsamples its input using a non-learnable operator *U*

(2) encodes the upsampled version into the latent space by $f_k$

(3) performs correction of the latent code via a conditional GAN $G_k$

(4) decodes the corrected latent code using $g_k$

# Laplacian GANs

- Increase **image** resolution in a coarse-to-fine manner during synthesis
- Cascading image synthesis with a series of generative networks $G_0, \ldots, G_k$
- Each network generates a high-frequency residual image

$$I_k = U(I_{k+1}) + G_k(U(I_{k+1}), z_k)$$

$U(\cdot)$ is an upsampling operator

## Spaces of 3D point clouds

- We start with a series of 3D spaces $\mathbb{R}^{n_0 \times 3}, \mathbb{R}^{n_1 \times 3}, \ldots, \mathbb{R}^{n_K \times 3}$
- $X_k = \{x_i\}_{i=1}^{n_k}$ is a set of 3D points on a surface

Intuition

- Modeling highly detailed 3D point clouds represents a challenge due to their high dimensionality
- We start with a low-detail (but easily constructed) model $X_0$
- We decompose the modelling task into a sequense of more manageable stages
- Each stage aims at a gradual increase of detail

ADASE

Skoltech
Skolkovo Institute of Science and Technology

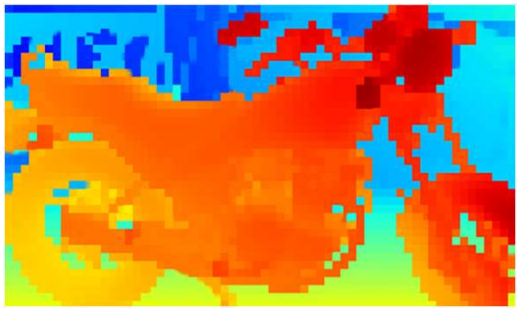
## Training auto-encoding point networks on multiple scales

- The network either accepts or generates an initial point cloud $X_0$
- Processes it with a series of $K$ learnable steps

## Full architecture of LSLP-GAN model

- The network either accepts or generates an initial point cloud $X_0$
- Processes it with a series of $K$ learnable steps

# Fast Photo-Guided Surface-Aware Super-Resolution of Range-Images via Deep Learning



Input

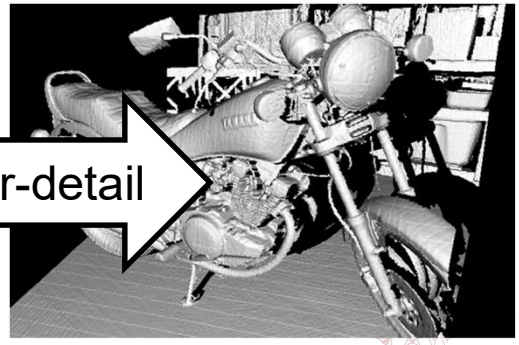
Fast, neural network

Target



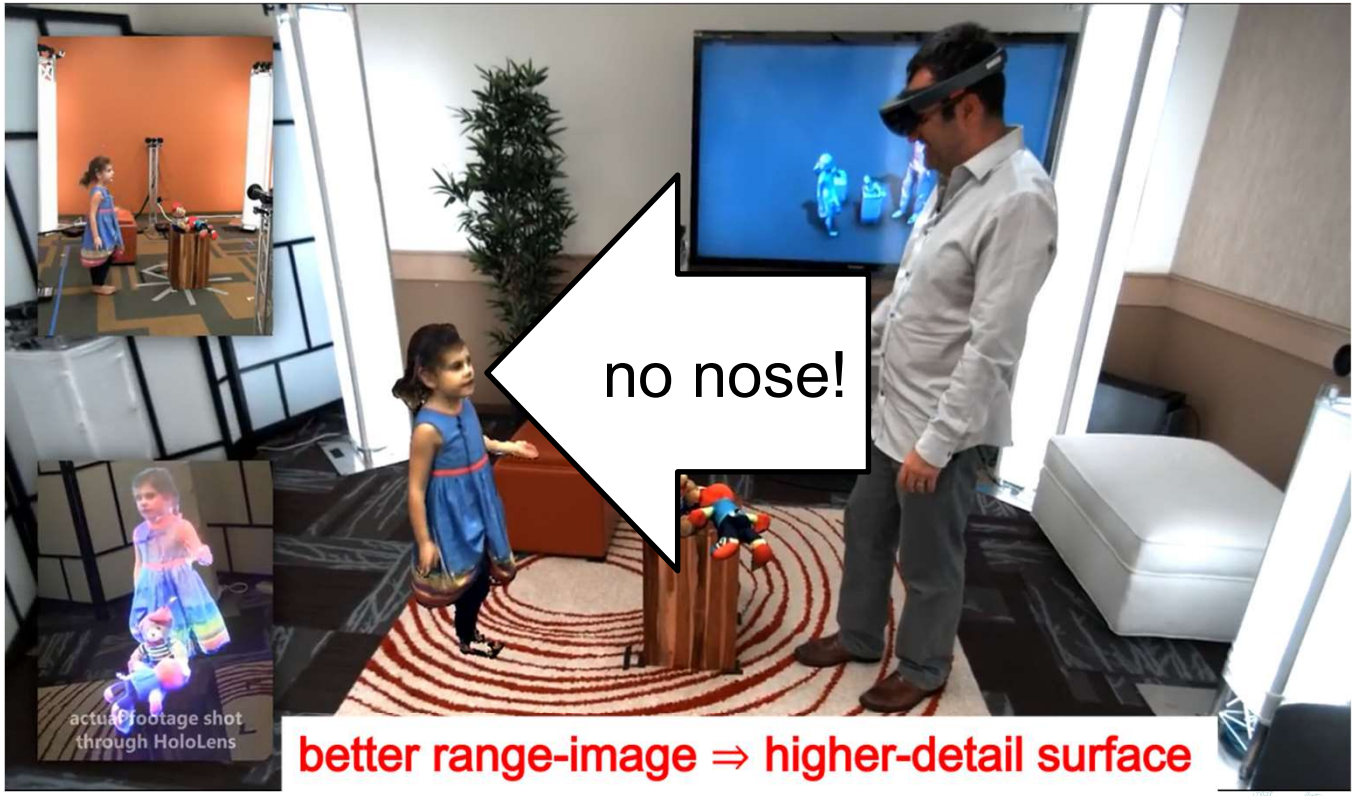

Corresponding surfaces

Higher-detail



Skoltech
Skolkovo Institute of Science and Technology

# Applications: Augmented Reality




https://www.microsoft.com/en-us/research/project/holoportation-3/

# Applications: Computer Vision

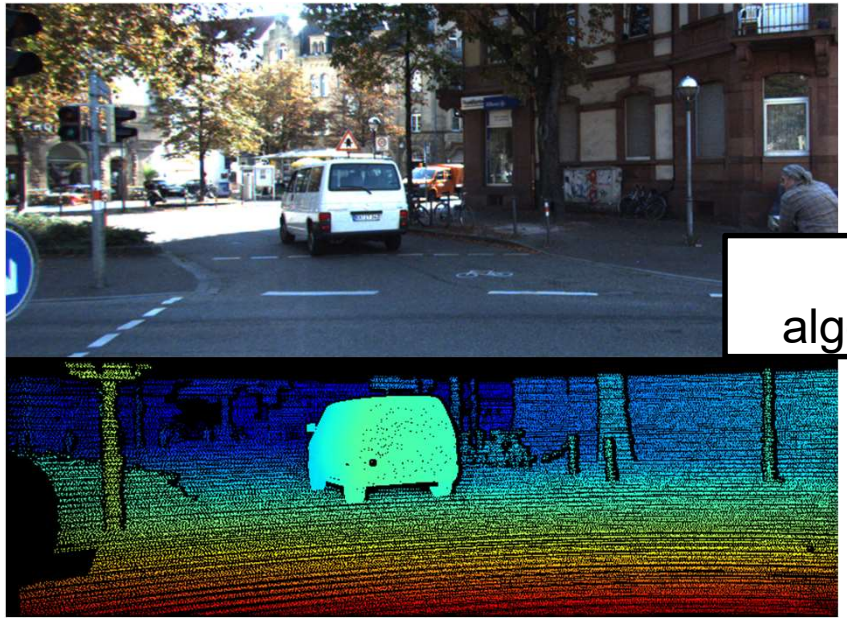
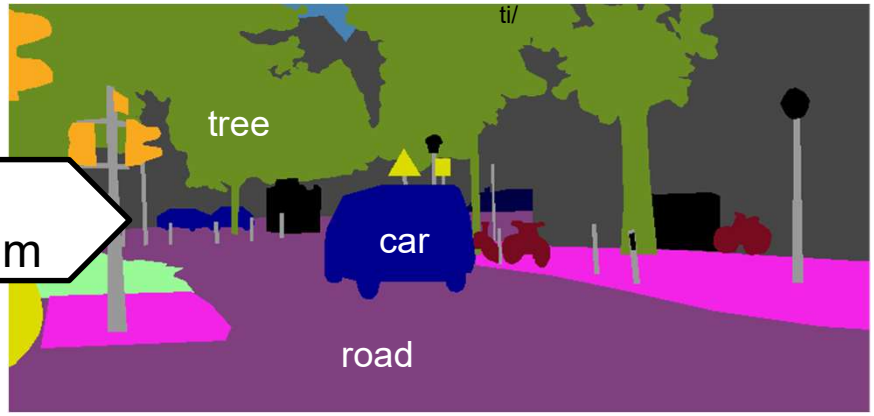


Photo + Range-image

CV algorithm

http://www.cvlibs.net/datasets/kitti/




Scene understanding

**better range-image ⇒ better scene understanding**

ADASE

Skoltech
Skolkovo Institute of Science and Technology

# Range-Image Super-Resolution: Approaches

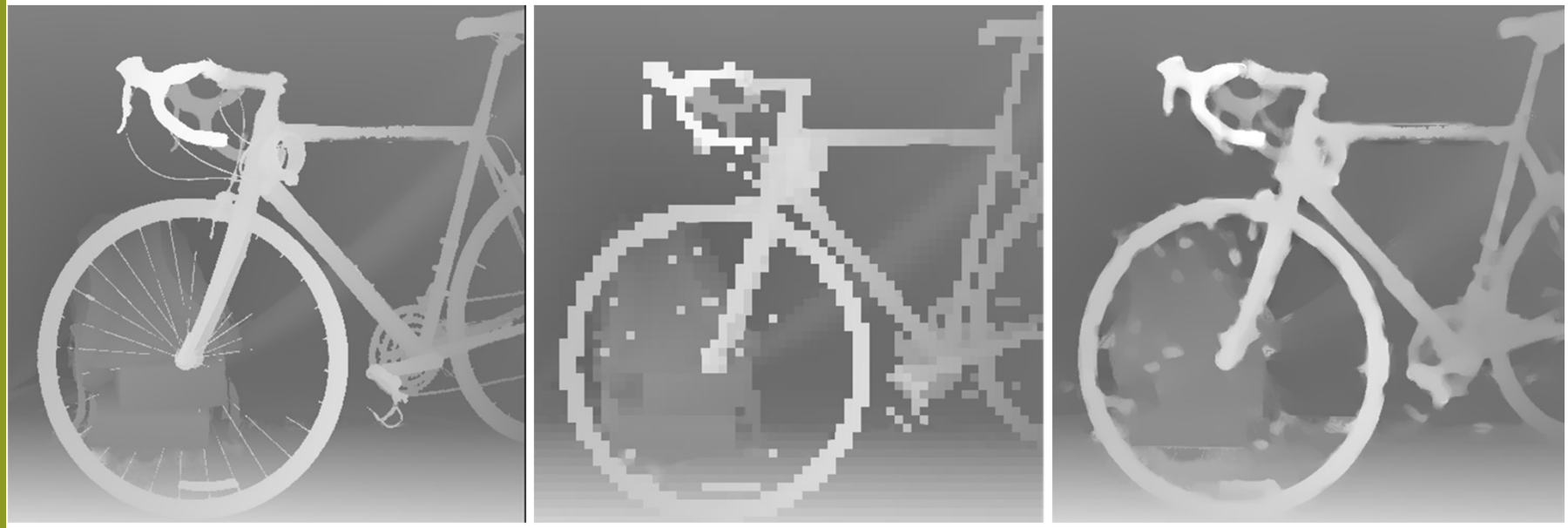
| Approach | Quality | Efficiency |
|---|---|---|
| Optimization | very high | low |
| Filtering | low | high |
| Learning | reasonably high | high |

**No evaluation of learning-based solutions in terms of surface quality!**

**Drawbacks of learning**

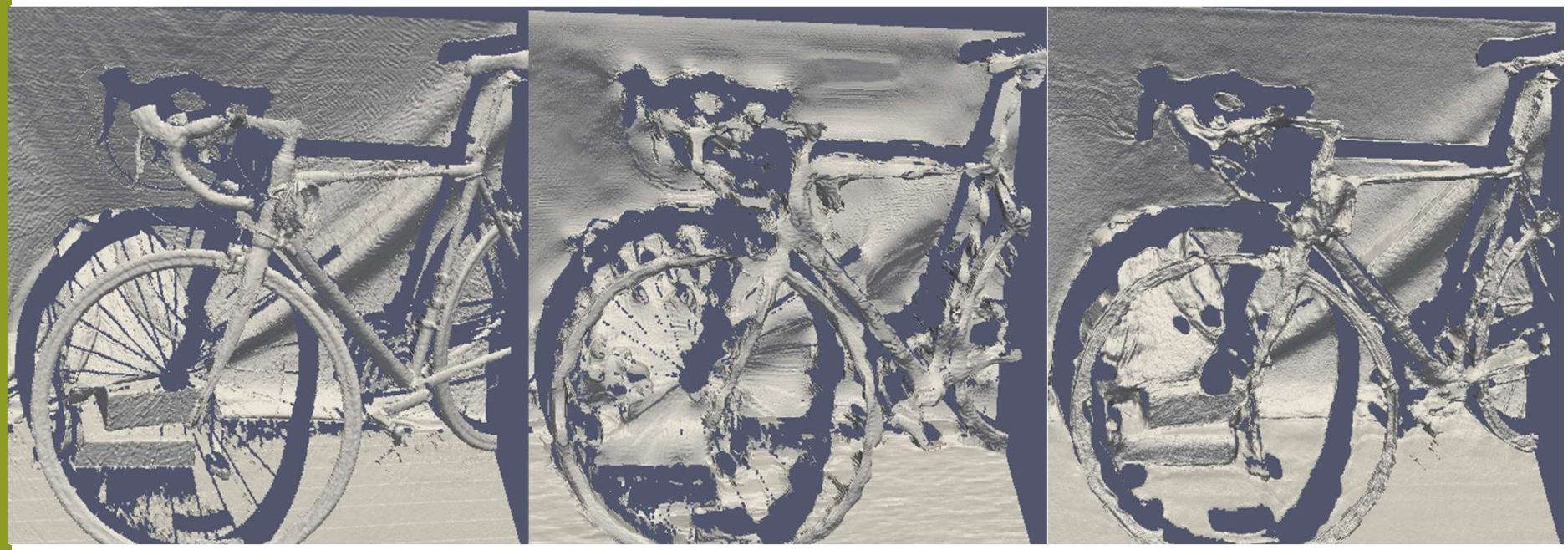
- Little theoretical understanding
- Require tuning by hand
- Require large-scale training datasets

# 8x super-resolution



Ground truth | Low-resolution input | [Riegler et al. 2016]

# Reconstructed surface



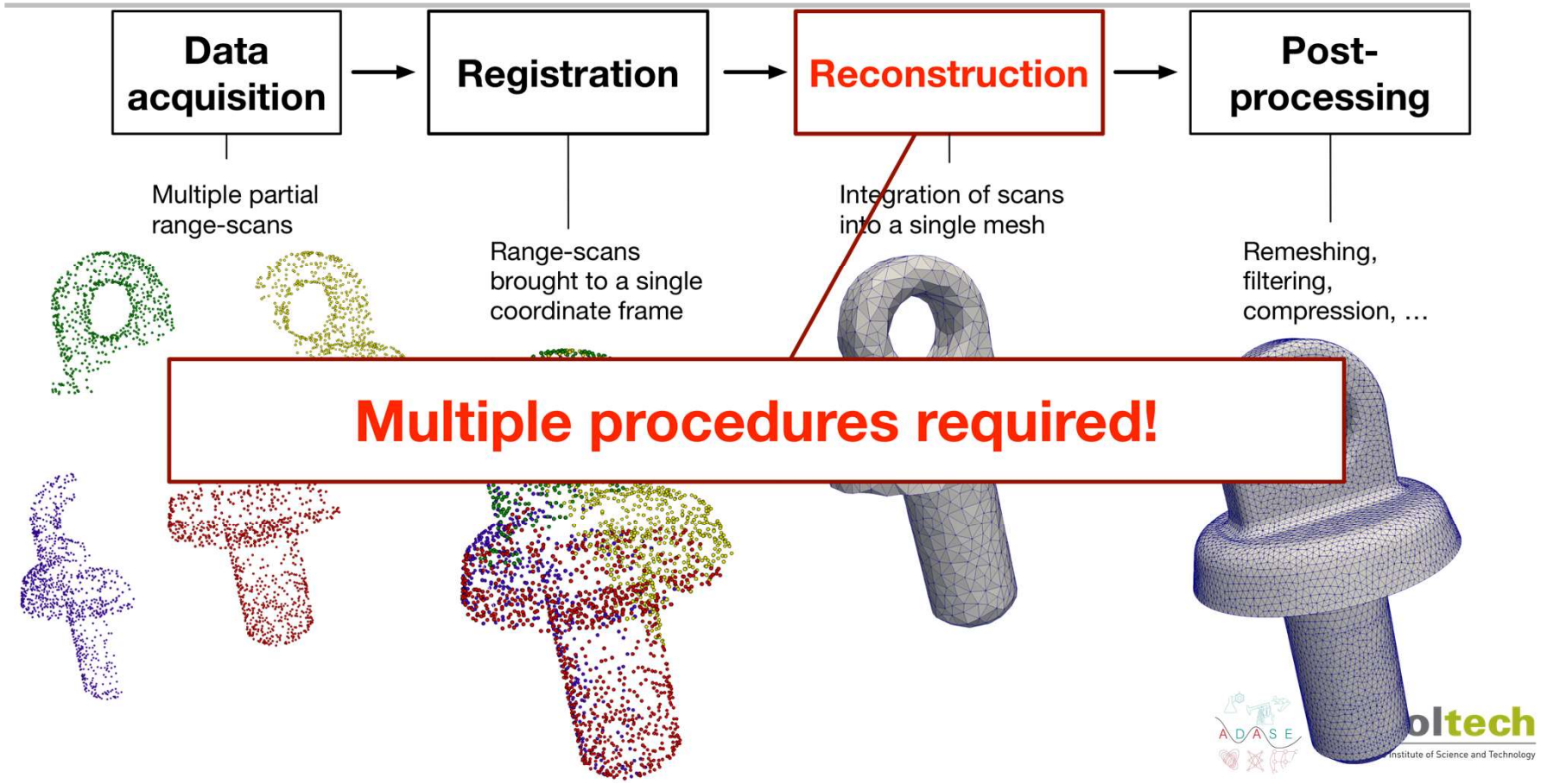
Ground truth

[Riegler et al. 2016]

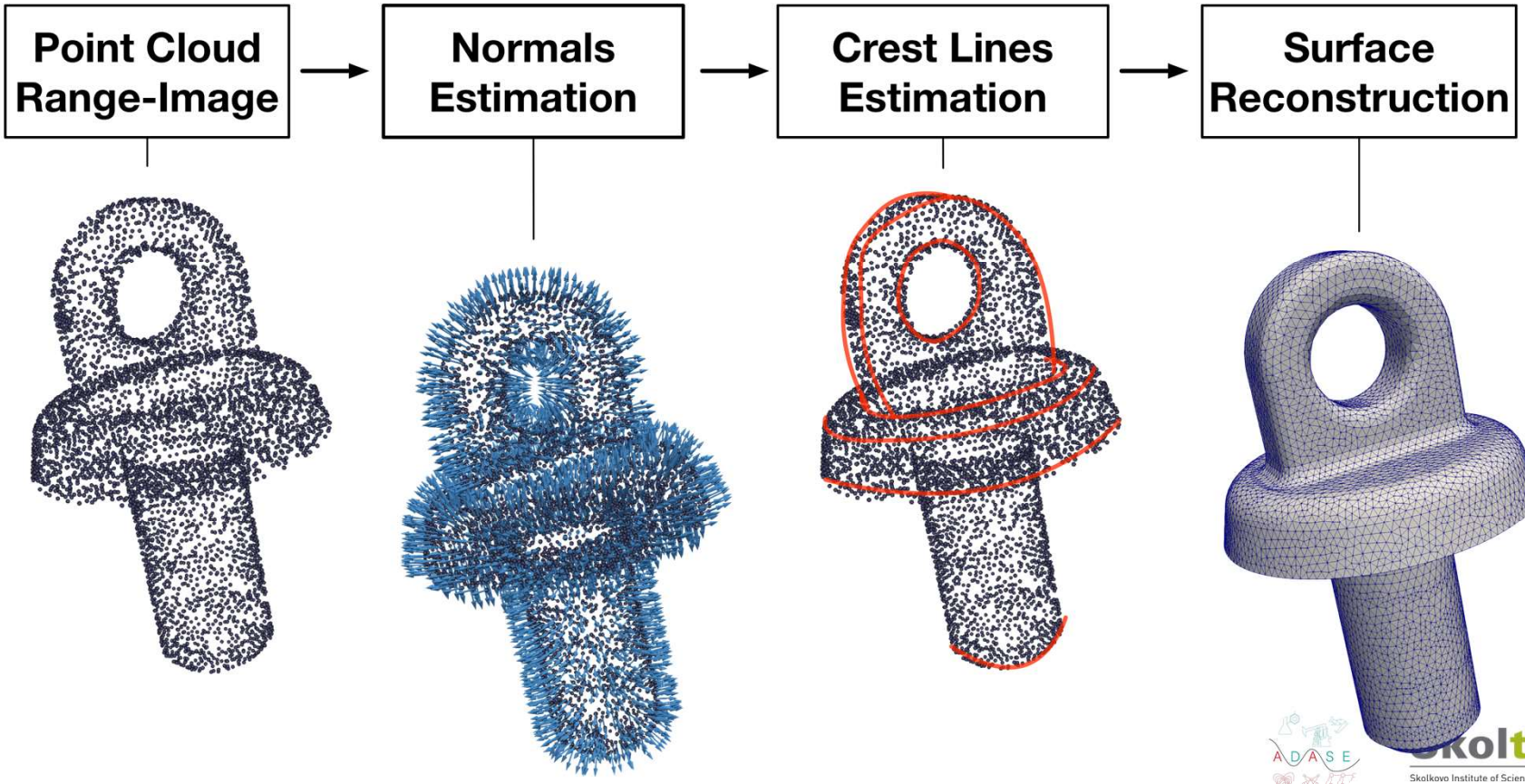
Deep Image Prior
with novel loss

Skoltech
Skolkovo Institute of Science and Technology

# Roadmap Inspired by the Geometry Processing Pipeline

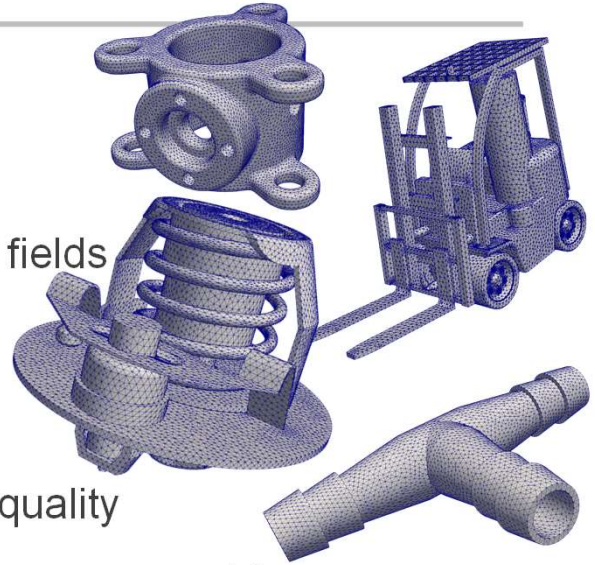


# Reconstruction is Another Multi-Stage Procedure



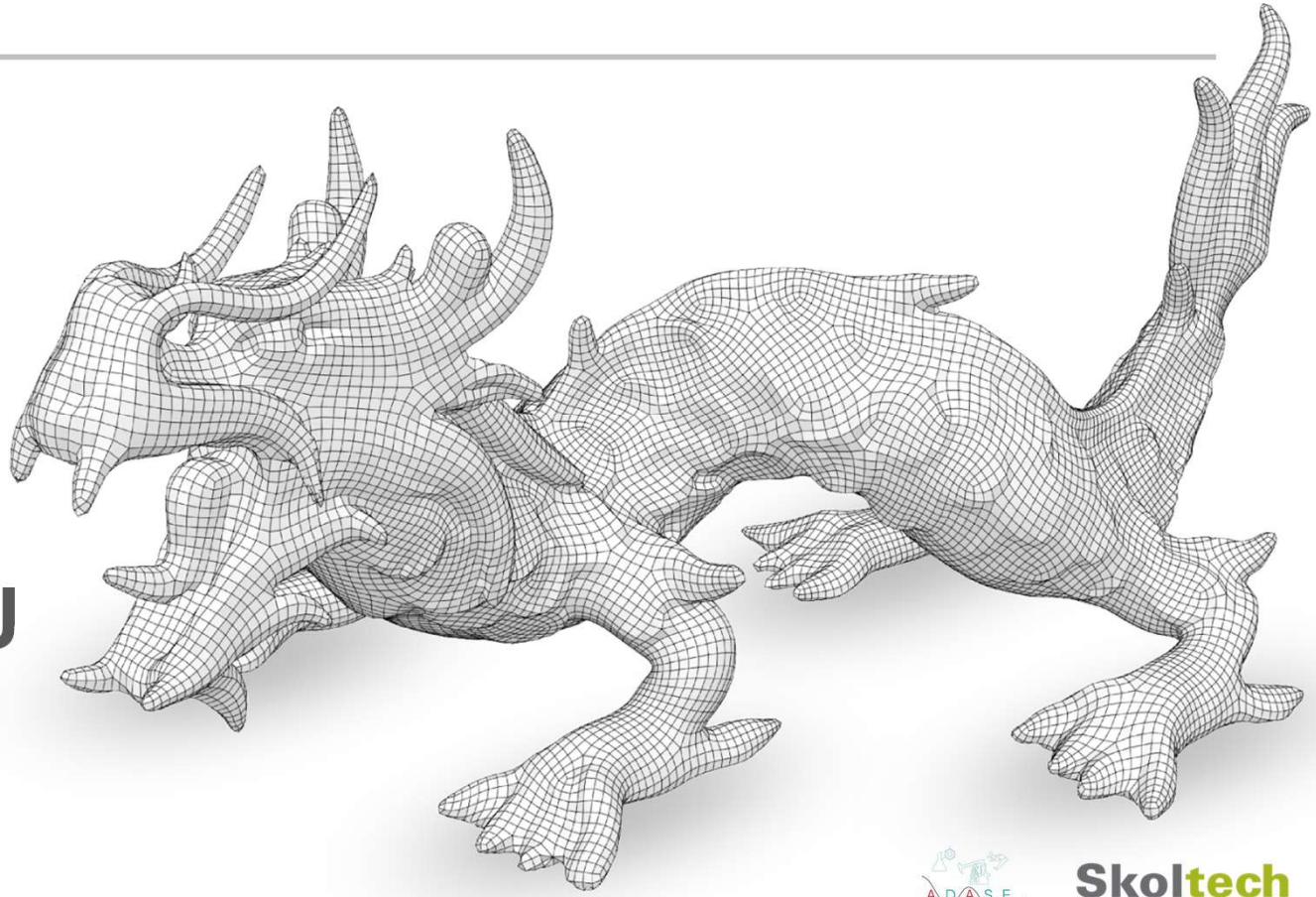

ADASE
Skoltech
Skolkovo Institute of Science and Technology

# Reconstruction is Another Multi-Stage Procedure



- **Geometric features prediction** problem:
  - Input: point cloud
  - Output: local geometric properties: lines, points, vector fields
- **Objectives**:
  - Acquire the dataset
  - Investigate how deep learning may improve prediction quality
- **Challenges**: existing approaches use handcrafting, are sensitive to noise, handle missing data in an unpredictable way

# THANK YOU



ADASE

Skoltech
Skolkovo Institute of Science and Technology