Geometry of Machine Learning.

Deep Learning for 3D Computer Vision

Prof. Evgeny Burnaev, e.burnaev@skoltech.ru
Skoltech, 2020

Skoltech
Skolkovo Institute of Science and Technology

ADASE team

prof. Evgeny Burnaev
PhD in Math,
head of ADASE group

prof. Alexander Bernstein
Dr. Sci in Math

Maxim Sharaev
PhD in Math

Alexei Artremov
PhD in Math

Ilya Trofimov
PhD in Math

Serguei Barannikov
Leading Research Scientist, PhD in Math

- 30 researchers
- DL for
  - 3D Computer Vision
  - Predictive Analytics

Industrial Expertise: since 2007

and many others...

ADASE team

Evgeny Burnaev with the **President of Russian Academy of Science**
State Kremlin palace, 2018

Moscow government prize for Scientific Achievements, 2018

“Ilya Segalovich” Yandex prize for Scientific Achievements, 2020

~ 100 **papers** in major venues, incl. NIPS, ICML, CVPR, etc.

The Best Paper Award for the research on modeling of point clouds and predicting properties of 3D shapes at the Int. Workshop on Artificial Neural Networks in Pattern Recognition (ANNPR), 2020

Geometry Processing Dataset Award for the work «ABC Dataset: A Big CAD Model Dataset For Geometric Deep Learning», Symposium on Geometry Processing, 2019

The Best Paper Award for the research in eSports at the IEEE Internet of People conference, 2019

Collaboration with Huawei

- Huawei-Skoltech lab
- Fundamental and Industrial Projects

- “**Multivariate Time Series Prediction Techniques for Supply Chain Planning**”: Huawei HIRP project, 2018-2019 (completed)

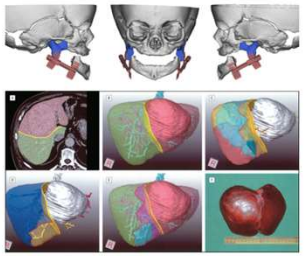
- "**Making DensPose fast and light**": Huawei-Skoltech lab cooperation project, 2019-2020 (completed)

- "**Efficient Neural Architecture Search**": Huawei-Skoltech lab cooperation project, 2019-2020 (completed)

- "**Bayesian Generative Models and their applications for continual learning**": Huawei Kirin-Skoltech cooperation project, 2020-2021 (ongoing)

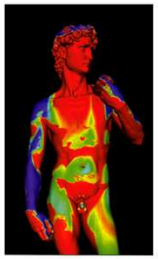
Problems faced in processing of geometric data

- Multitude of real-world & exciting applications for 3D/4D data processing!



Virtual surgery planning

Urban planning/modeling



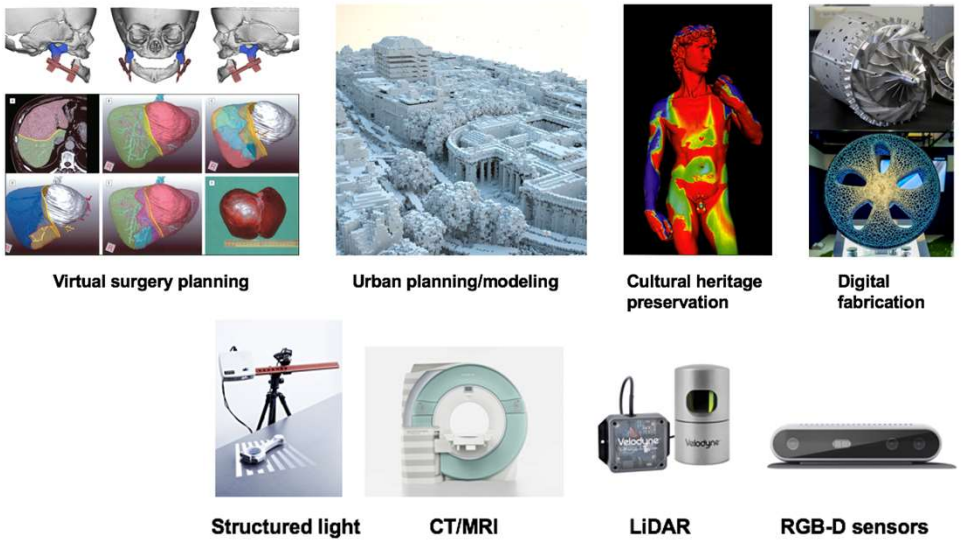
Cultural heritage preservation

Digital fabrication

Problems faced in processing of geometric data

- Multitude of real-world & exciting applications for 3D/4D data processing!

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

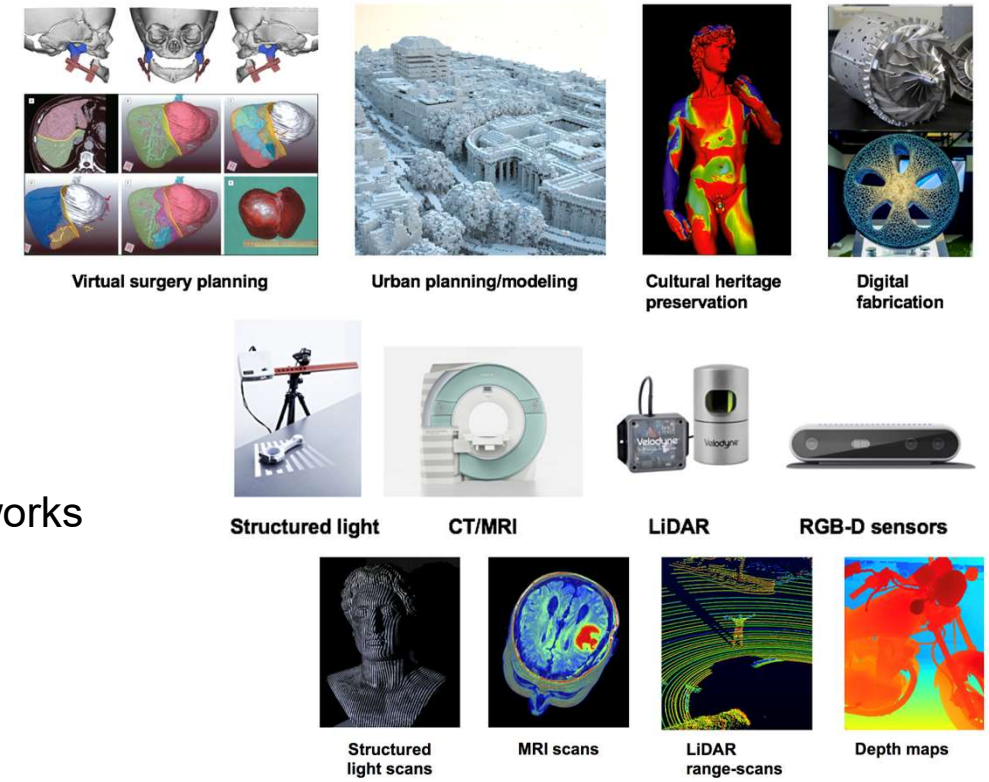


Problems faced in processing of geometric data

- Multitude of real-world & exciting applications for 3D/4D data processing!

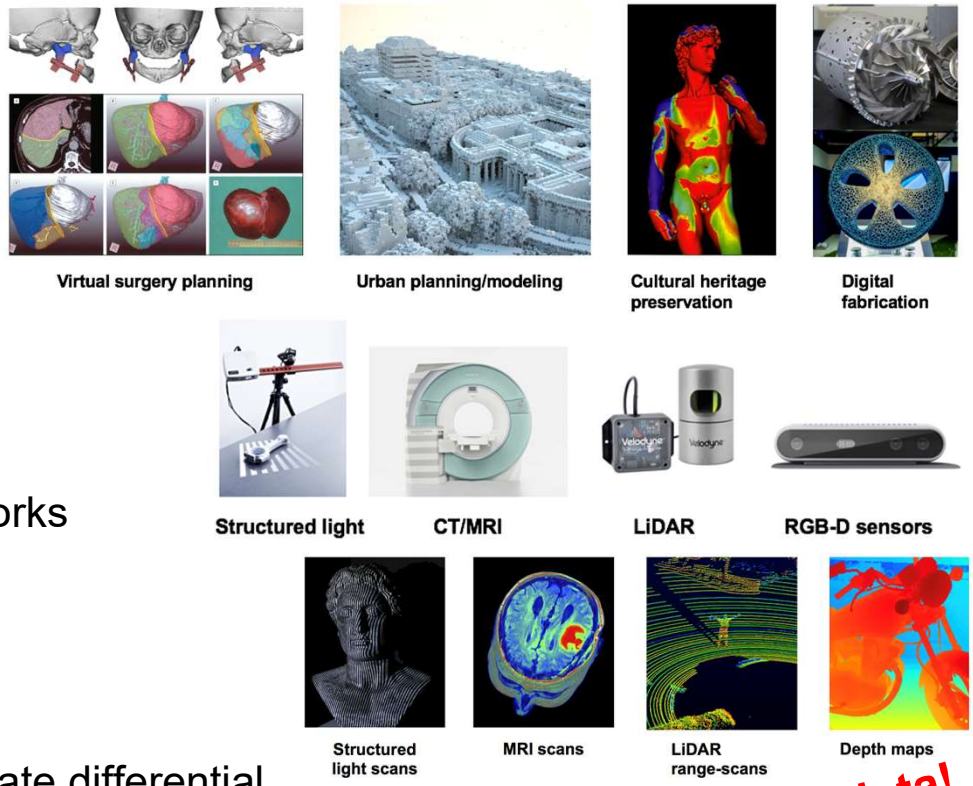
- Millions of depth sensors are available on the market. However, only scarce annotated datasets exist

- Traditional 2D Convolutions Neural Networks cannot solve *all* problems!
- 3D/4D and geometry-aware models are required



Problems faced in processing of geometric data

- Multitude of real-world & exciting applications for 3D/4D data processing!



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

- Traditional 2D Convolutions Neural Networks cannot solve *all* problems!
- 3D/4D and geometry-aware models are required

- We need efficient approaches to incorporate differential geometric properties of data in ML/DL models and methods
- In turn, differential geometric methods can be used to analyze properties of ML/DL models and methods

ML is about shape of data!

Main topics

1. Learnable differential geometric methods for analysis and processing of geometric data, signals and images
   - Uncover geometric structure of the data (manifold learning!)

Geometry in ML

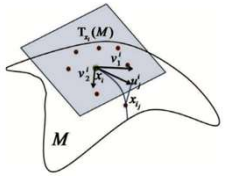


Figure 1: Local tangent space and tangent coordinates system of a manifold

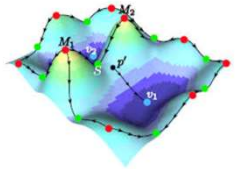


Figure 2: Integral trajectories of Morse gradient flow, see [75].

Main topics

1. Learnable differential geometric methods for analysis and processing of geometric data, signals and images
   - Uncover geometric structure of the data (manifold learning!)
   - Process the data that is the result of digitization of objects that are geometric by their nature (e.g., data obtained when scanning 3D objects)

Geometry in ML

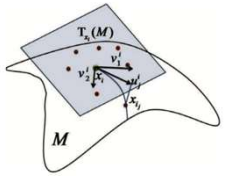


Figure 1: Local tangent space and tangent coordinates system of a manifold

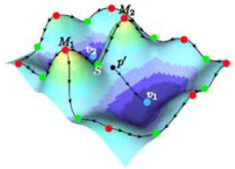
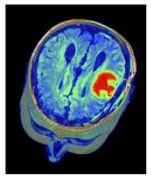


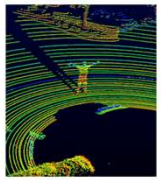
Figure 2: Integral trajectories of Morse gradient flow, see [75].

and ML for Geometric data

Structured light scans



MRI scans



LiDAR range-scans

Depth maps

Main topics

1. Learnable differential geometric methods for analysis and processing of geometric data, signals and images
   - Uncover geometric structure of the data (manifold learning!)
   - Process the data that is the result of digitization of objects that are geometric by their nature (e.g., data obtained when scanning 3D objects)

Geometry in ML

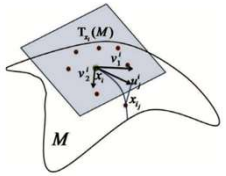


Figure 1: Local tangent space and tangent coordinates system of a manifold

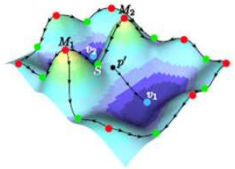
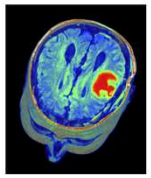


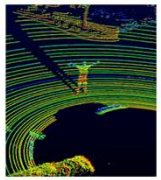
Figure 2: Integral trajectories of Morse gradient flow, see [75].

and ML for Geometric data

Structured light scans



MRI scans

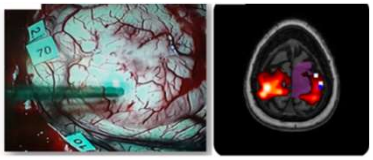


LiDAR range-scans

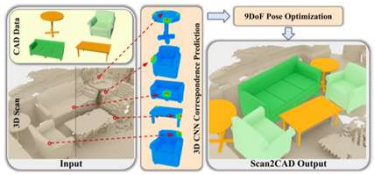
Depth maps

for a series of key application areas:

3. Important applications in **AR/VR**, **medical imaging** and **3D reconstruction** and **digitization**



Medical Imaging



3D reconstruction and digitization

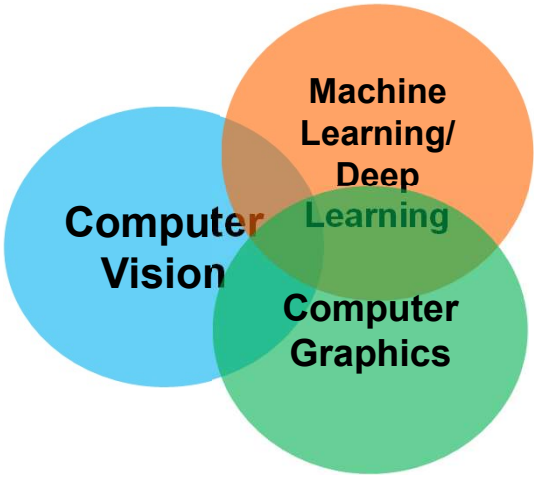
AR/VR

Main topics

1. Learnable differential geometric methods for analysis and processing of geometric data, signals and images
   - Uncover geometric structure of the data (manifold learning!)
   - Process the data that is the result of digitization of objects that are geometric by their nature (e.g., data obtained when scanning 3D objects)

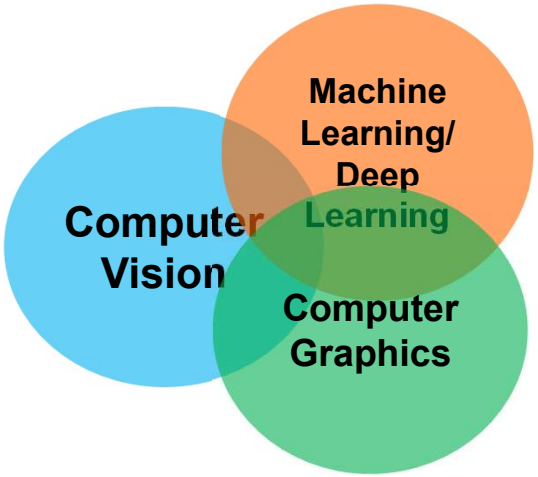
2. Intersection of **Computer Vision, Machine Learning/Deep Learning and 3D Graphics**

3. Important applications in **AR/VR**, **medical imaging** and **3D reconstruction** and **digitization**

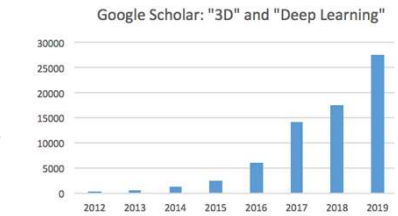


Main topics

1. Learnable differential geometric methods for analysis and processing of geometric data, signals and images
   - Uncover geometric structure of the data (manifold learning!)
   - Process the data that is the result of digitization of objects that are geometric by their nature (e.g., data obtained when scanning 3D objects)



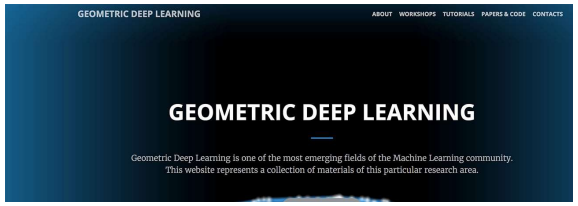
2. Intersection of **Computer Vision, Machine Learning/Deep Learning and 3D Graphics**

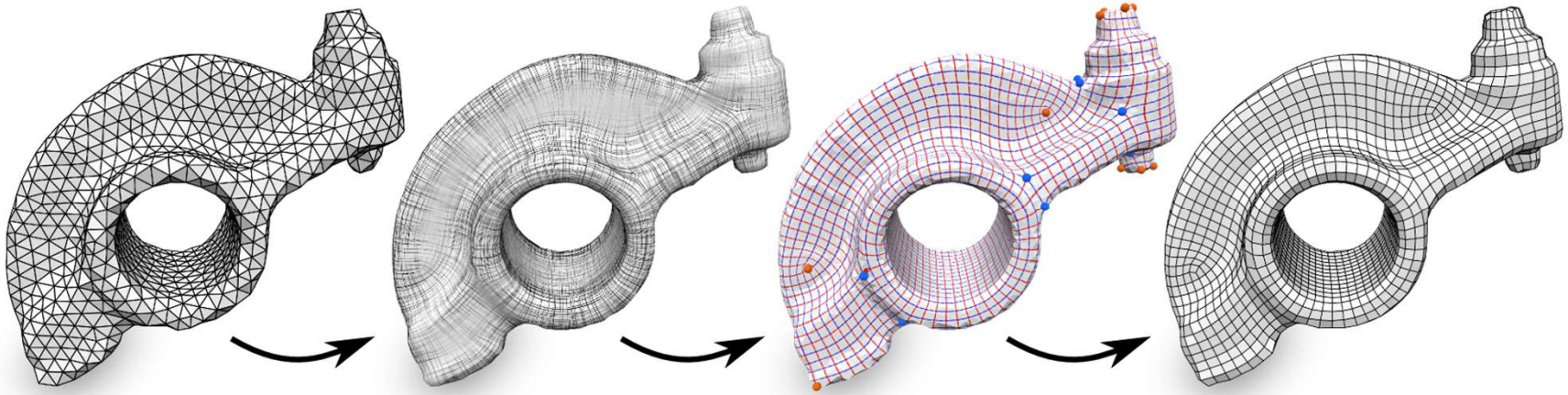


The field is growing fast!

3. Important applications in **AR/VR**, **medical imaging** and **3D reconstruction** and **digitization**

http://geometricdeeplearning.com

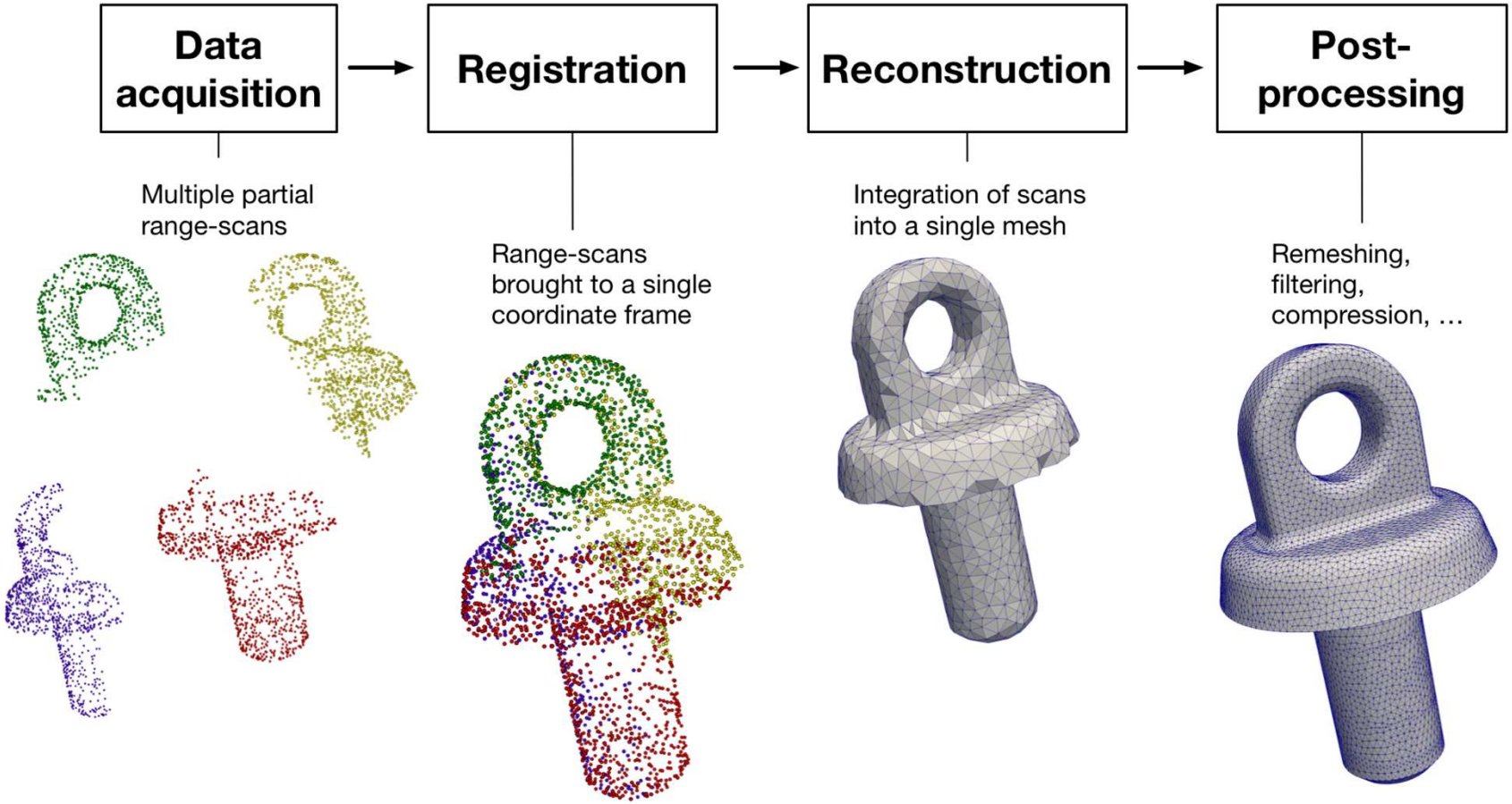




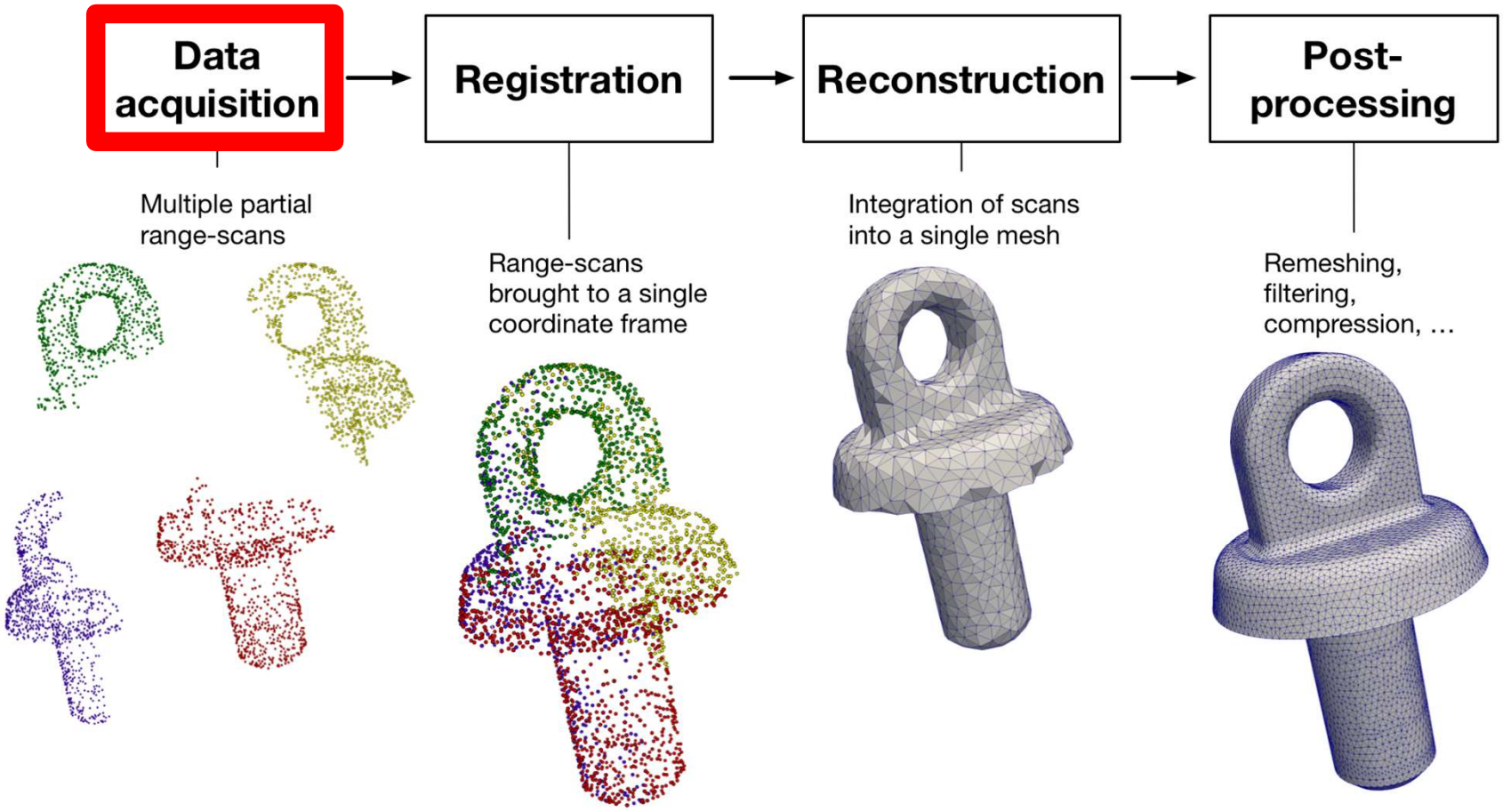
THE GEOMETRIC PROCESSING PIPELINE

ML -> Geometry

Roadmap Inspired by the Geometry Processing Pipeline



Roadmap Inspired by the Geometry Processing Pipeline



Skoltech

Skolkovo Institute of Science and Technology

Comprehensive multi-scale multi-fidelity benchmark for monocular, two view, and multi view stereo and depth fusion [ongoing project with Huawei]

Gleb Bobrovskih[1], Pavel Karpyshev[1], Alexander Safin[1], Timofey Glukhih, Dzmitry Tsetseryukou[1], Evgeny Burnaev[1]

[1]Skolkovo Institute of Science and Technology, Russia

Aims

Challenges:

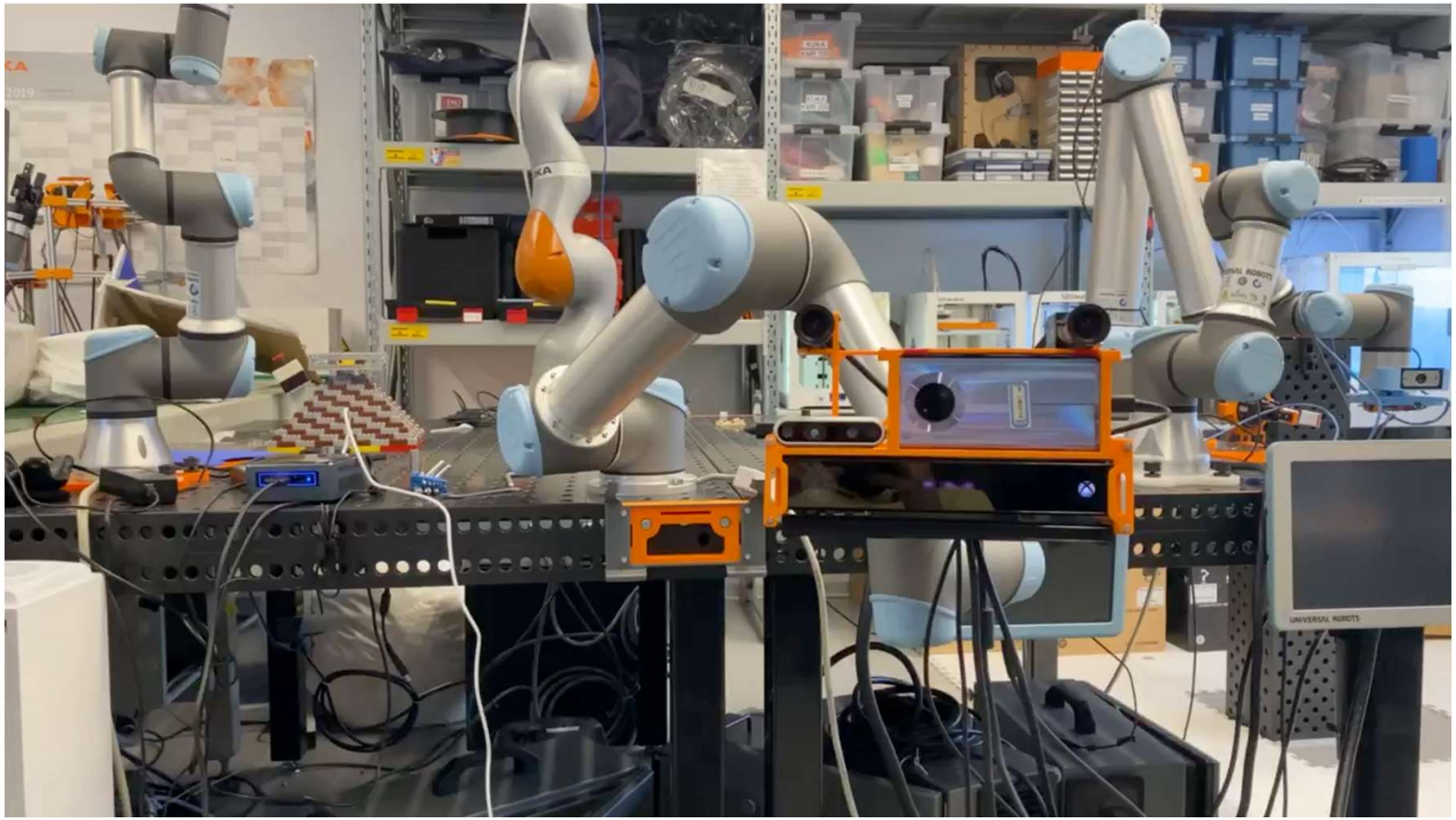
- Lack of paired multi-view depth+RGB data of different scales and modalities
- Important for 3D reconstruction

WP1. Collection of depth data on small scales

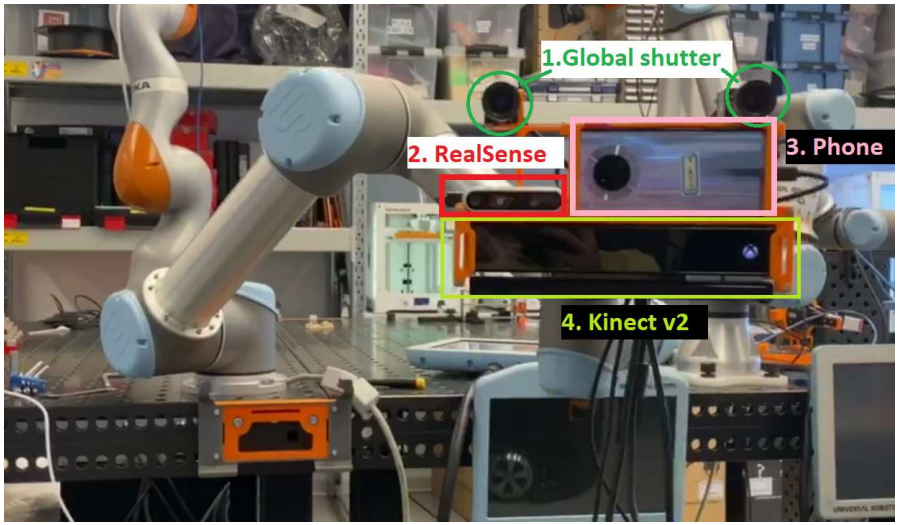
WP2. Collection of depth data on middle scales (aka office table with objects)

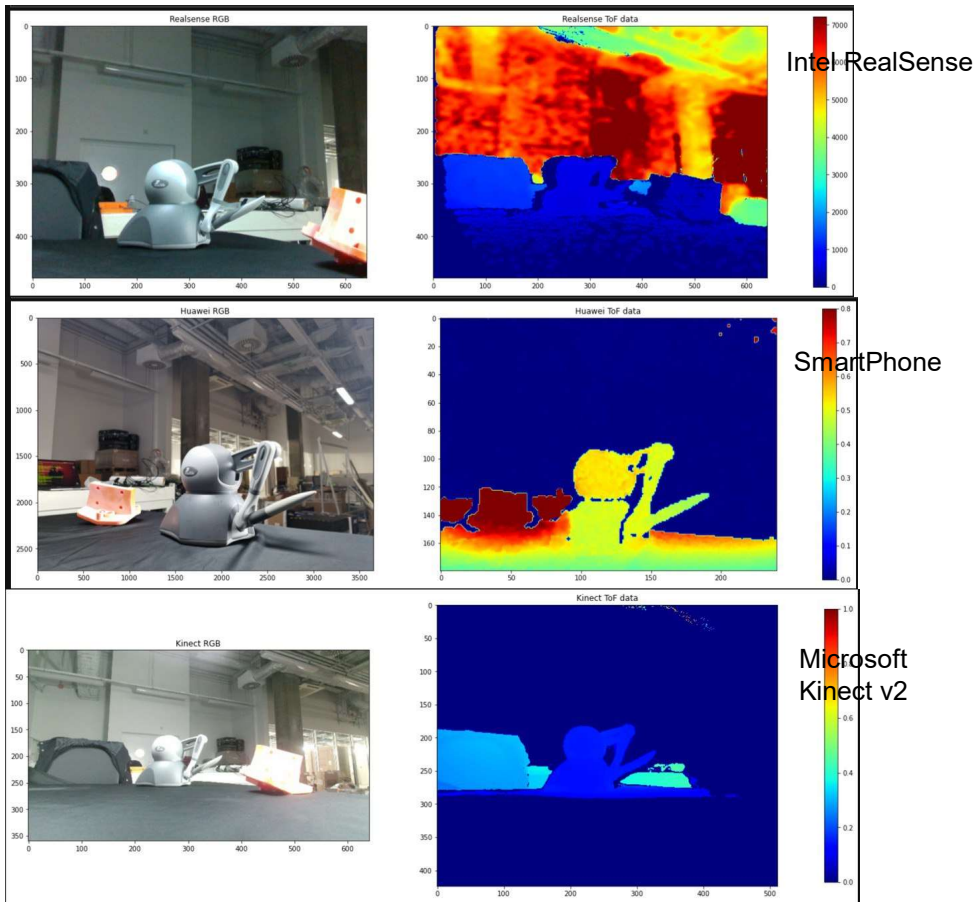
WP3. Collection of depth data from human mannequin on middle scales

Preliminary setup (one more phone and structured light to be added)

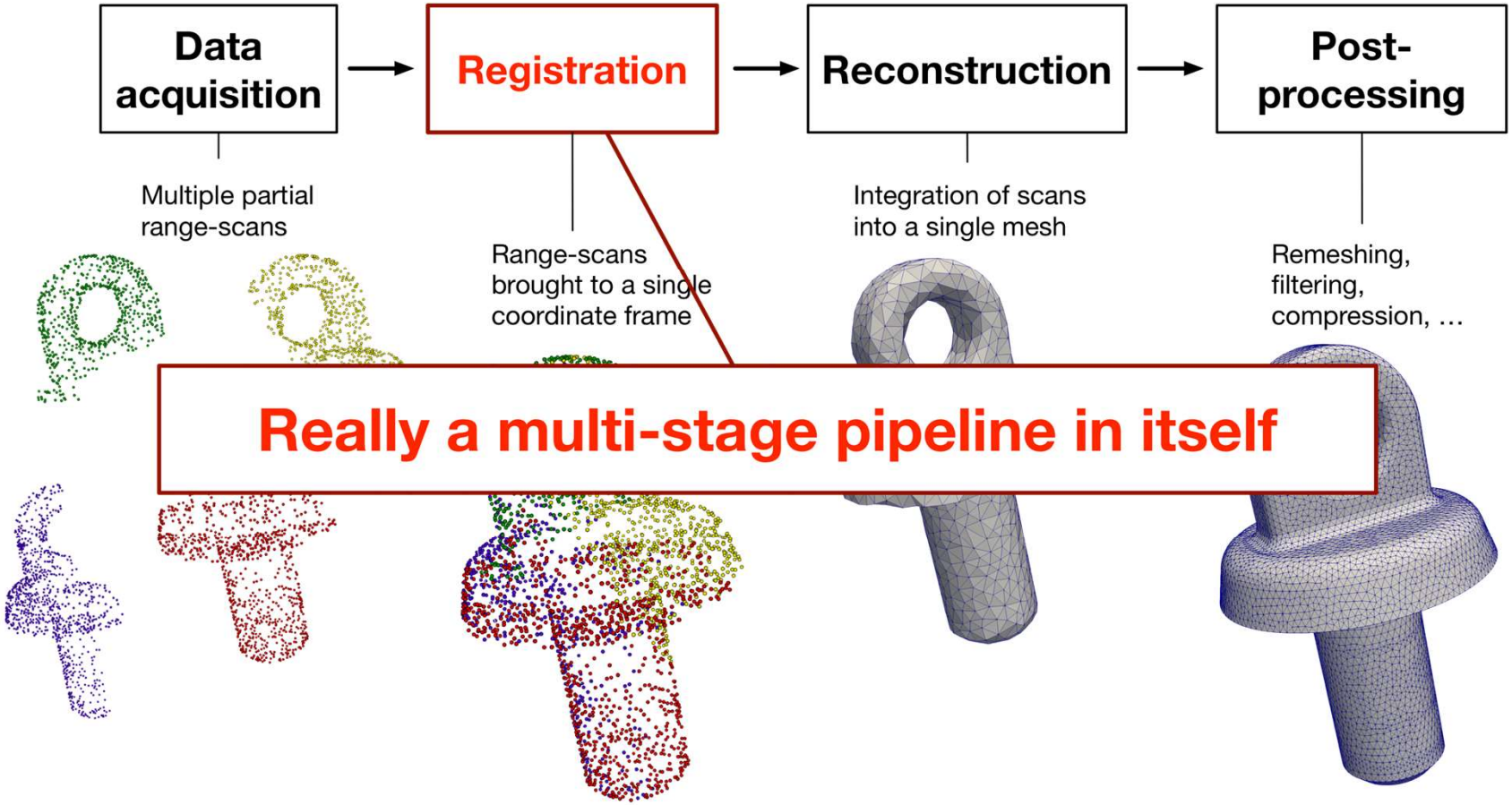


Preliminary setup (one more phone and structured light to be added)

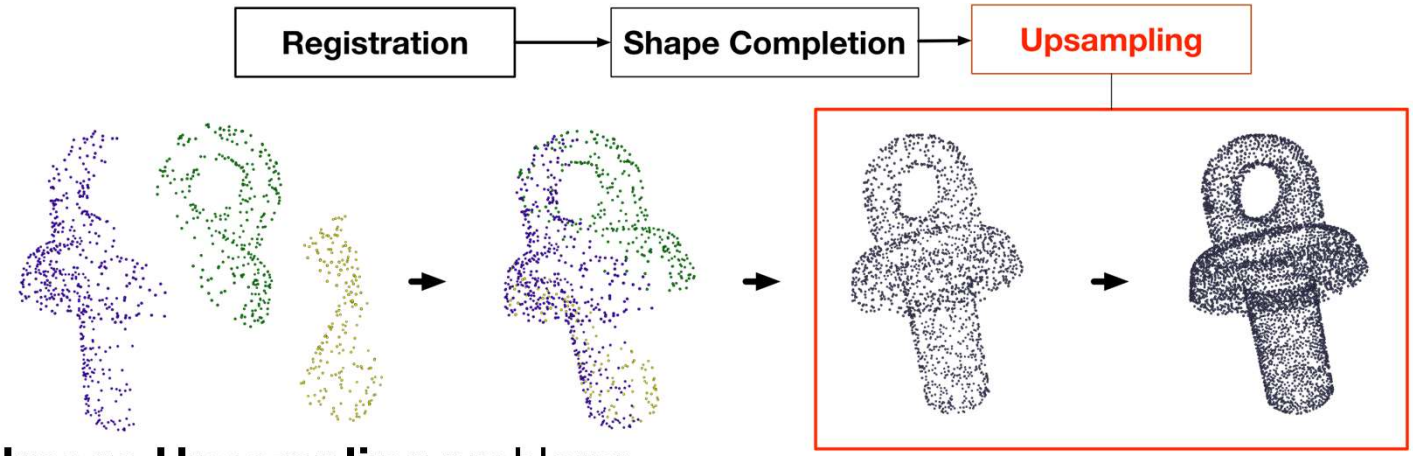




Roadmap Inspired by the Geometry Processing Pipeline



Registration is Really a Multi-Stage Procedure



Range-Image Upsampling problem:

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

Challenges:

- Point Cloud is an unordered set
- No definition of local neighbourhoods in Point Clouds
- RGBD is a piece-wise continuous

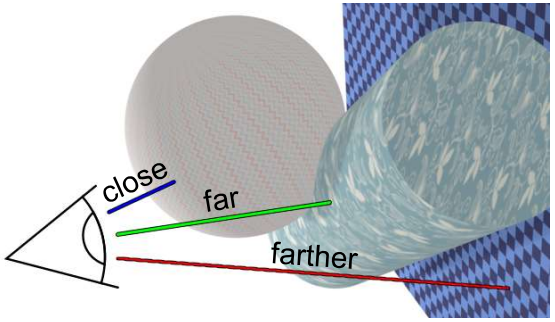
Perceptual Deep Depth Super-Resolution

Oleg Voynov[1], Alexey Artemov[1], Vage Egiazarian[1], Alexander Notchenko[1], Gleb Bobrovskikh[1], Evgeny Burnaev[1], Denis Zorin[2,1]

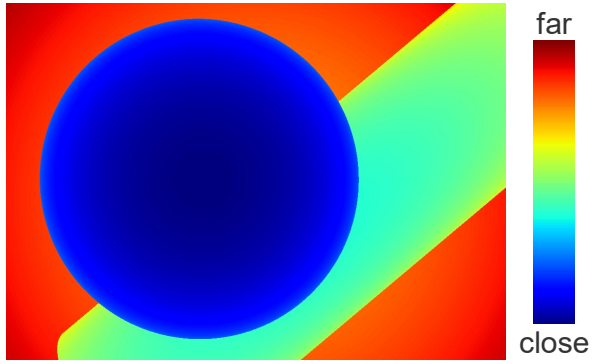
[1] Skolkovo Institute of Science and Technology, Russia
[2] New York University, USA

Depth maps

RGB image



side view



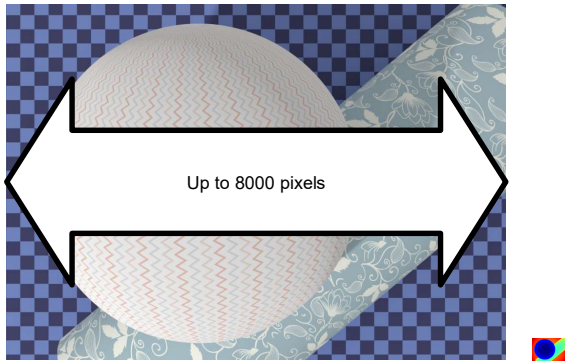
Depth map

Application

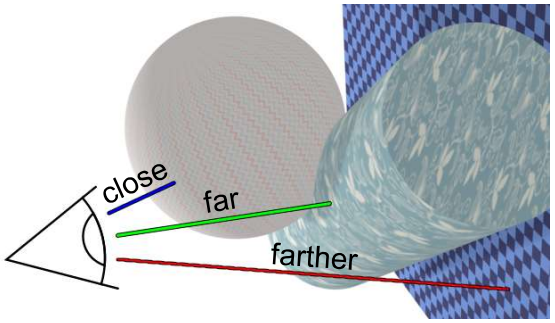
- Scene analysis and navigation
- 3D modeling
- Augmented / Virtual reality

[1] Video: Microsoft Holoportation, youtu.be/7d59O6cfaM0

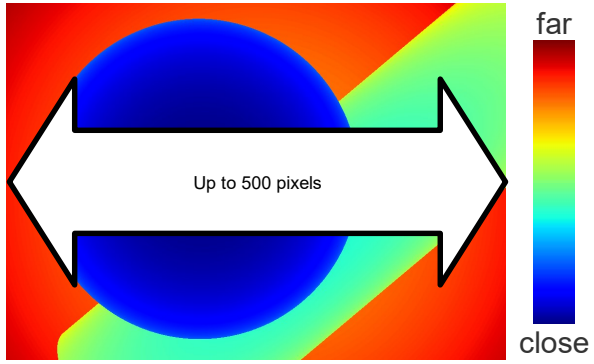
Depth maps



RGB image



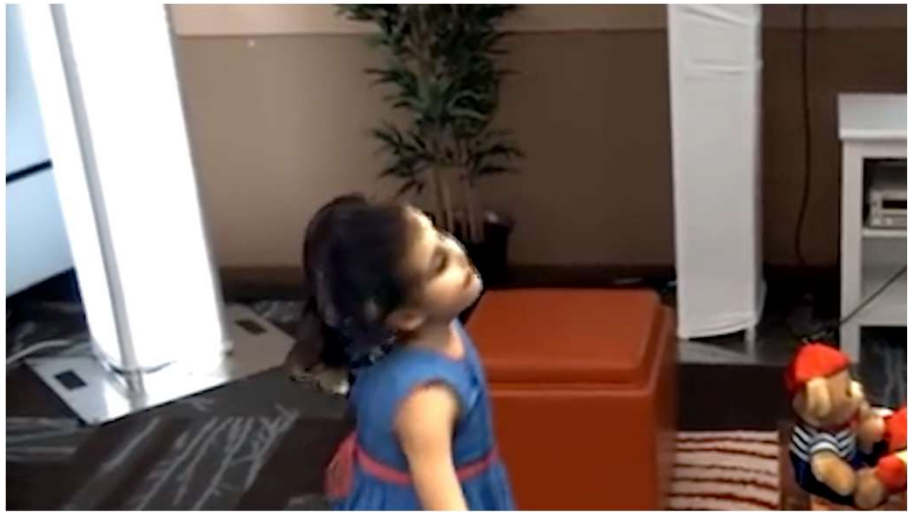
side view



Depth map

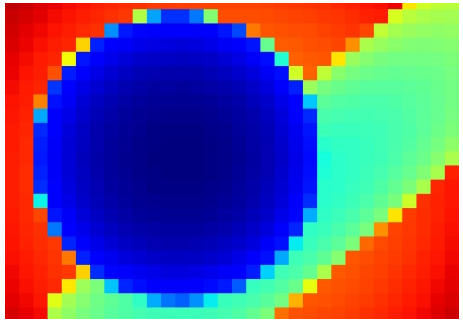
Application

- Scene analysis and navigation
- 3D modeling
- Augmented / Virtual reality



[1] Video: Microsoft Holoportation, youtu.be/7d59O6cfaM0

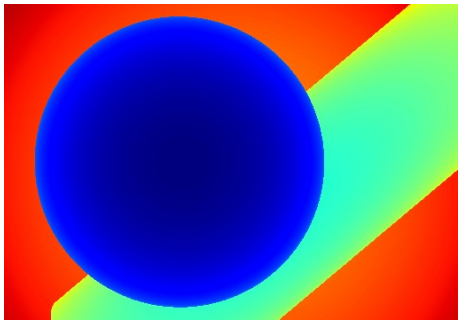
Depth map super-resolution



Low resolution
depth map

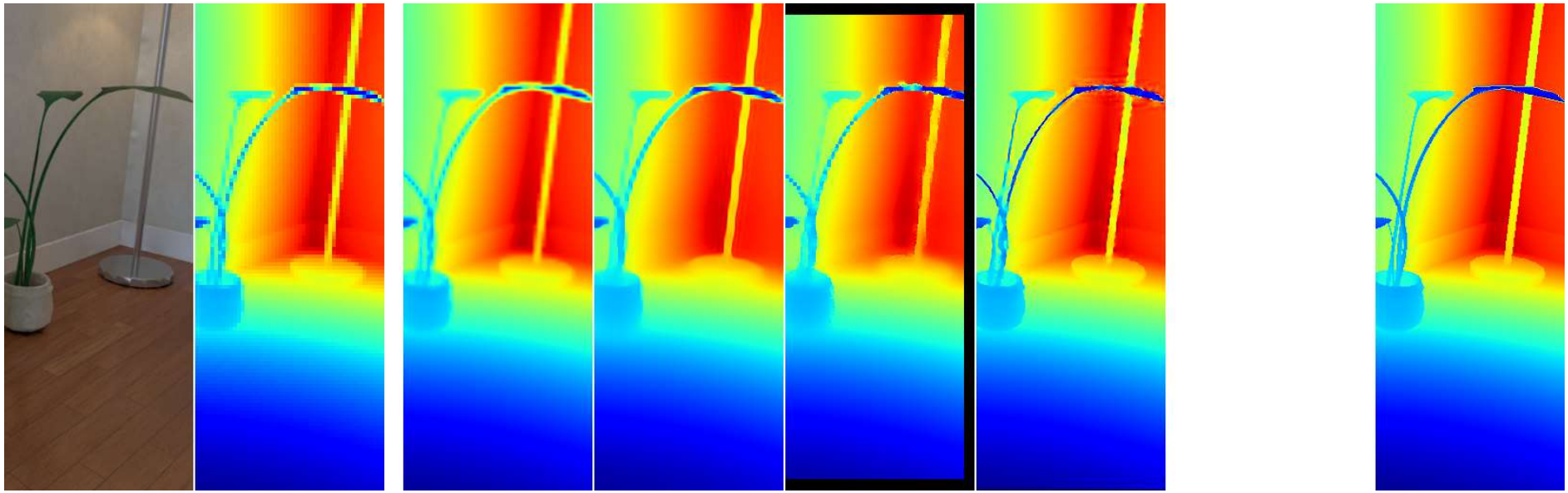
Hight resolution
RGB image

Super-resolution



Super-resolved
depth map

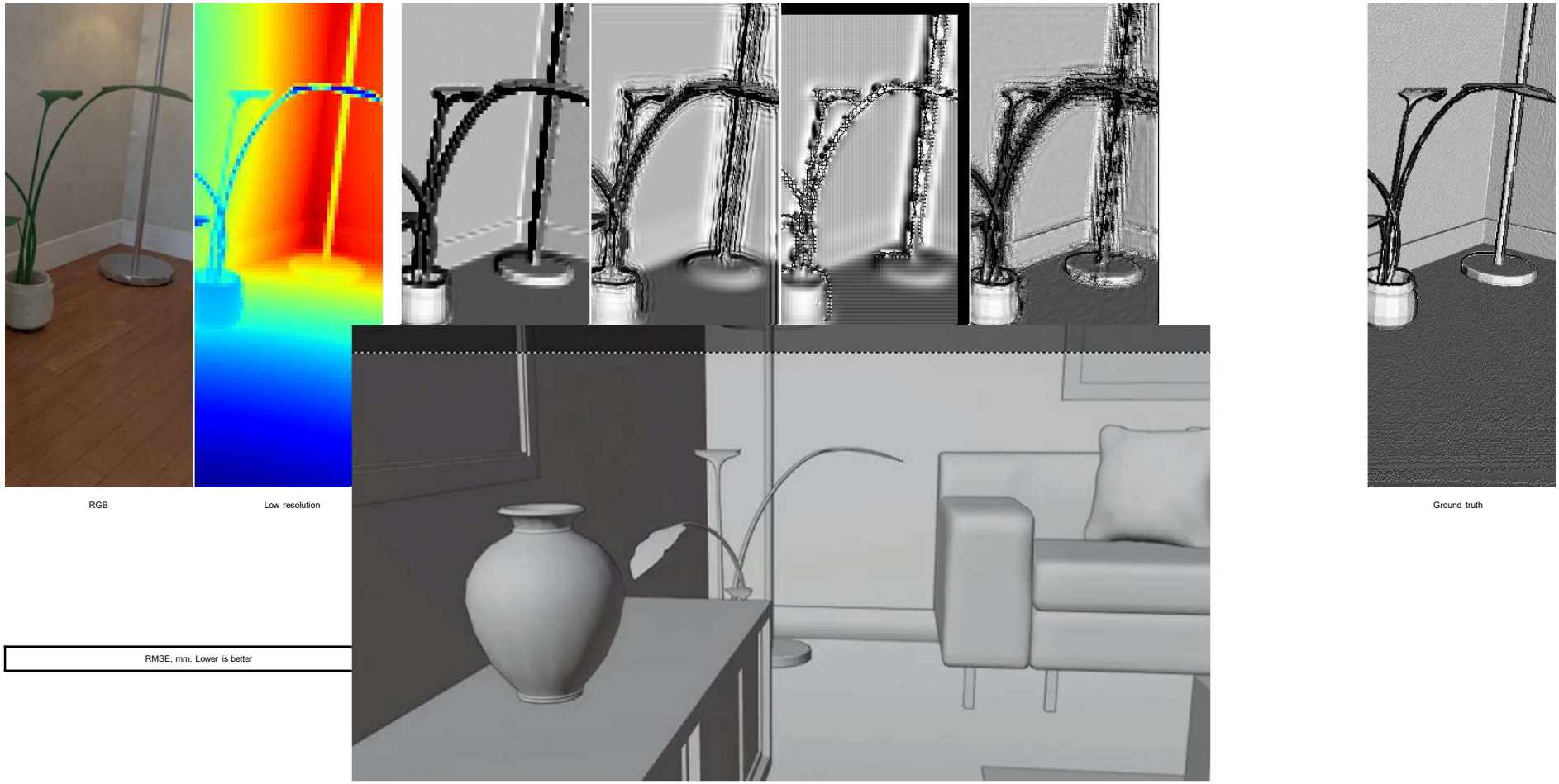
What is wrong with existing solutions



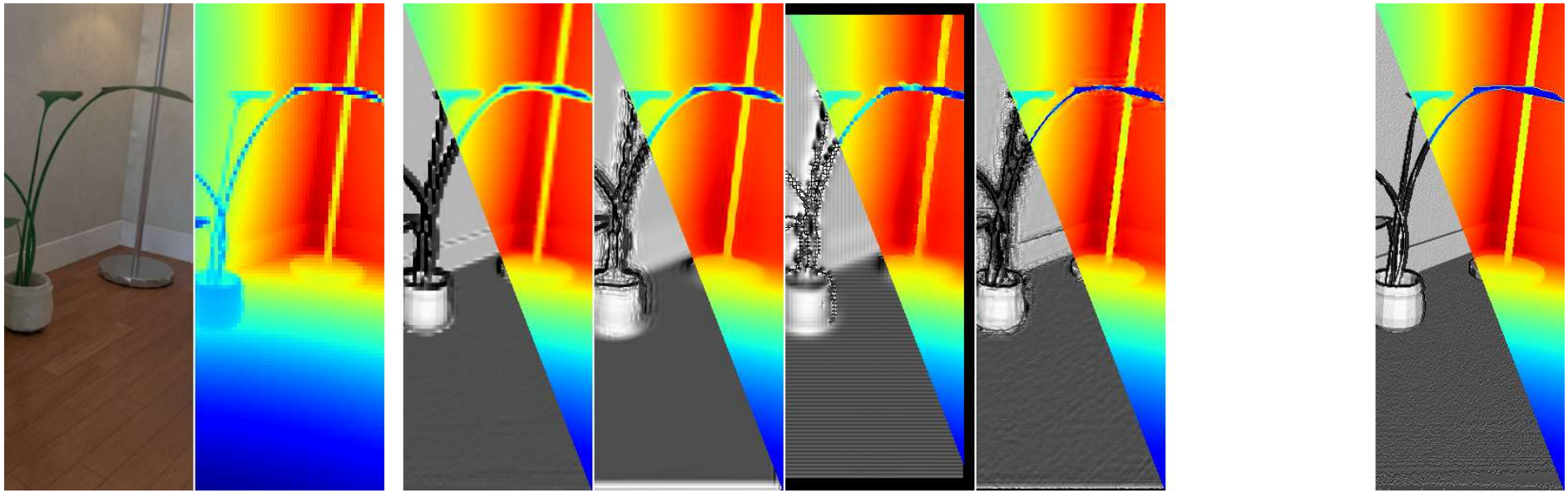
RGB | Low resolution | Bilinear interpolation | [2] Gu et al. 2017 | [3] Xie et al. 2016 | [4] Hui et al. 2016 | Ground truth

RMSE, mm. Lower is better	46	40	43	31

What is wrong with existing solutions



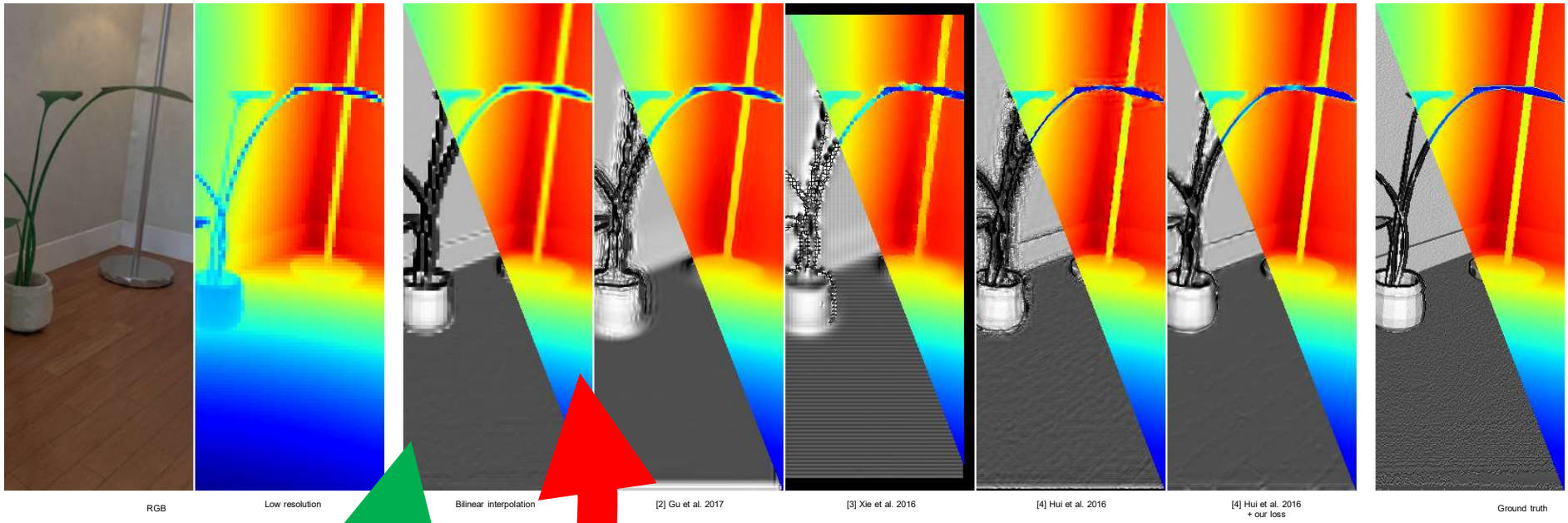
What is wrong with existing solutions



RGB | Low resolution | Bilinear interpolation | [2] Gu et al. 2017 | [3] Xie et al. 2016 | [4] Hui et al. 2016 | Ground truth

RMSE, mm. Lower is better	46	40	43	31

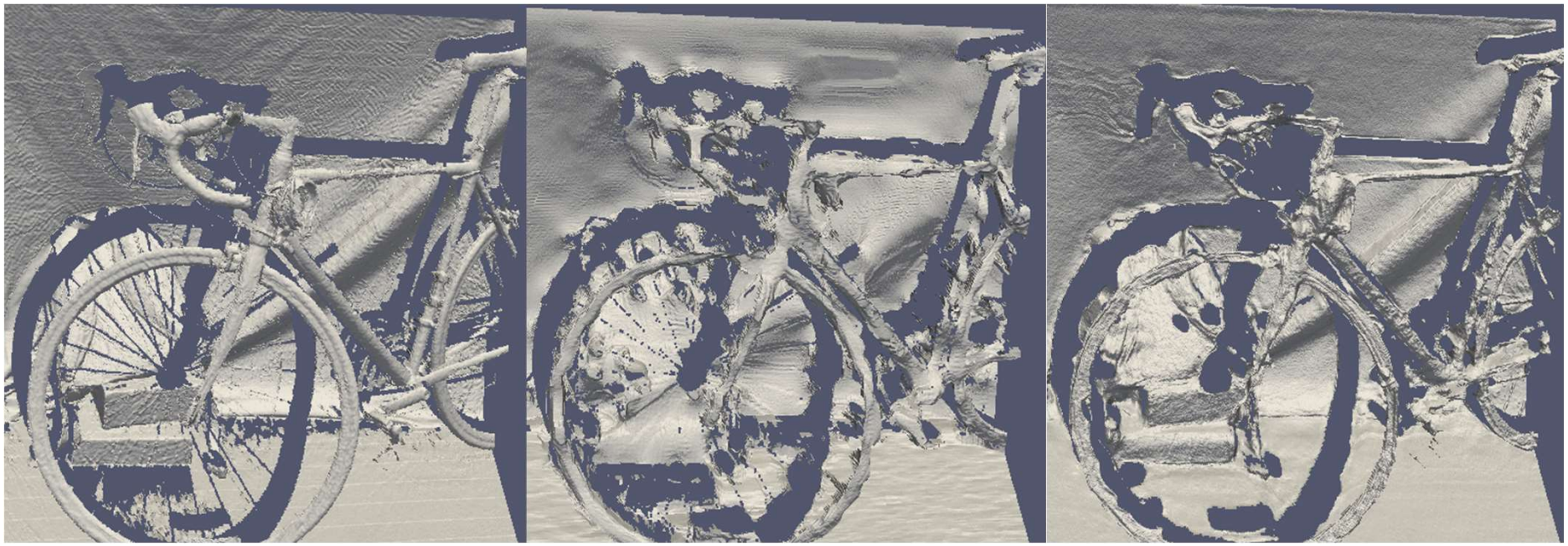
What is wrong with existing solutions



- Compare this instead of that
- Compare with perceptual metrics [5, Zhang 2018]

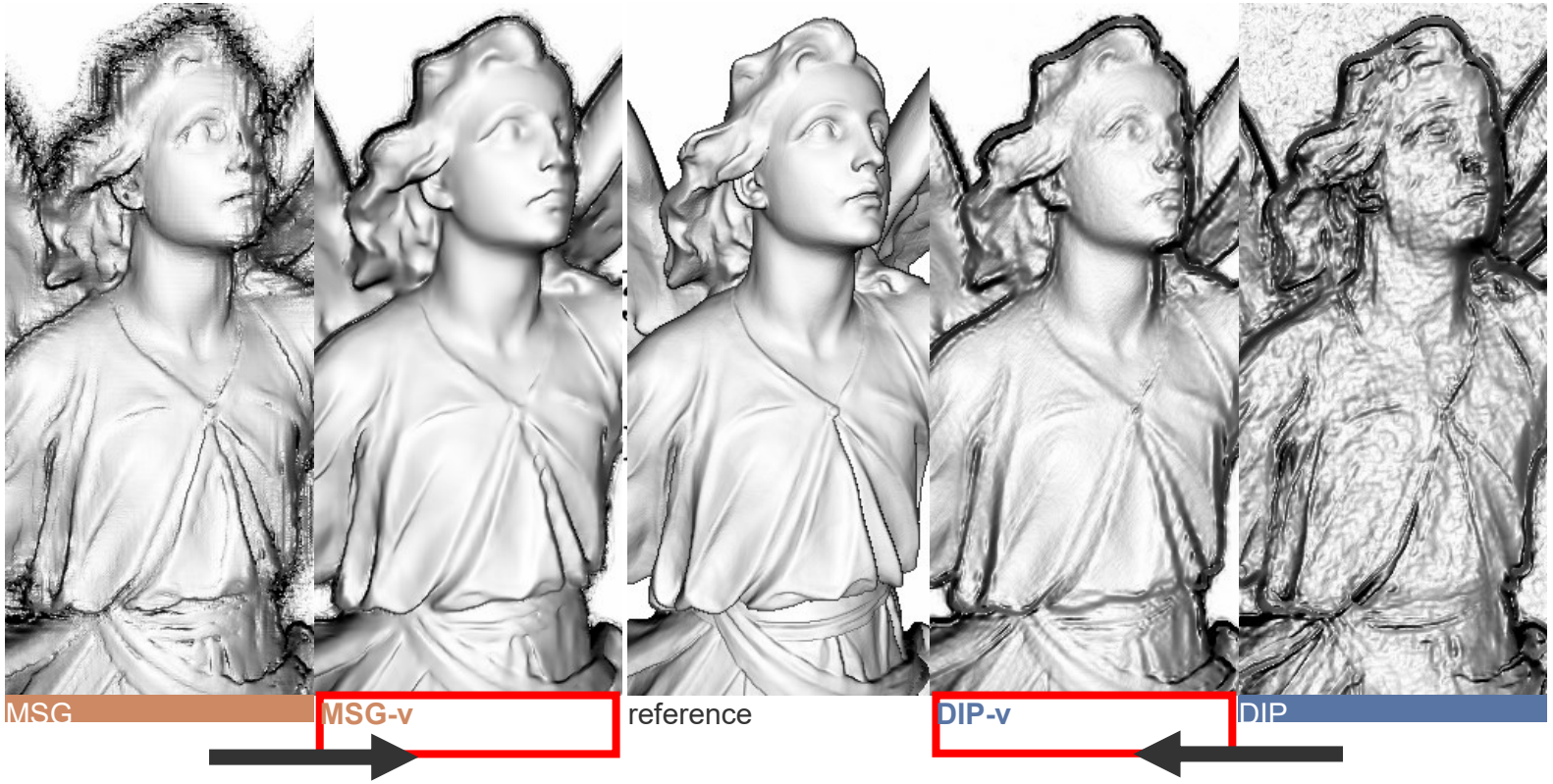
Best according to
> 80% of subjects

Reconstructed surface

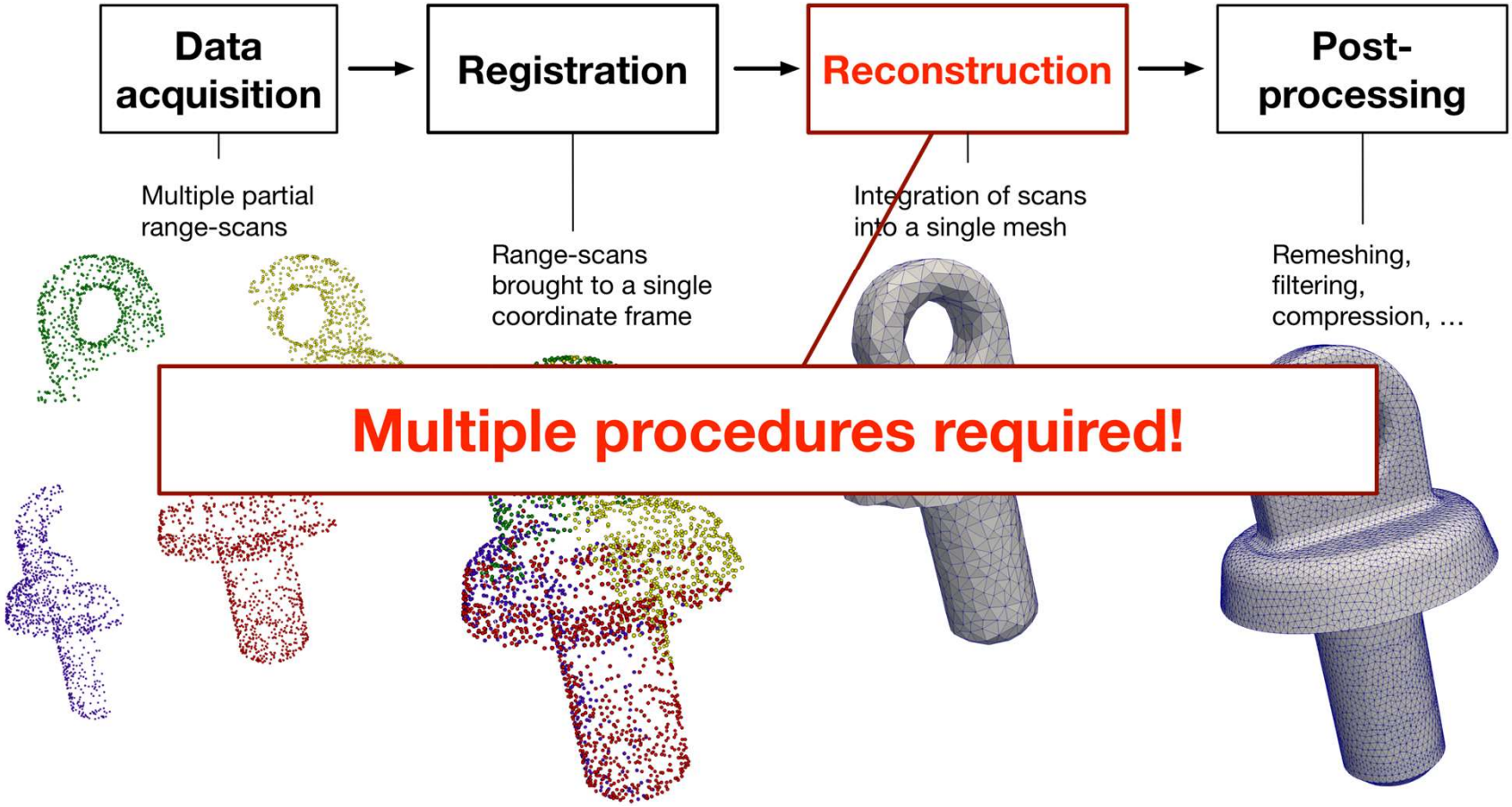


Ground truth | [Riegler et al.] | **Our model**

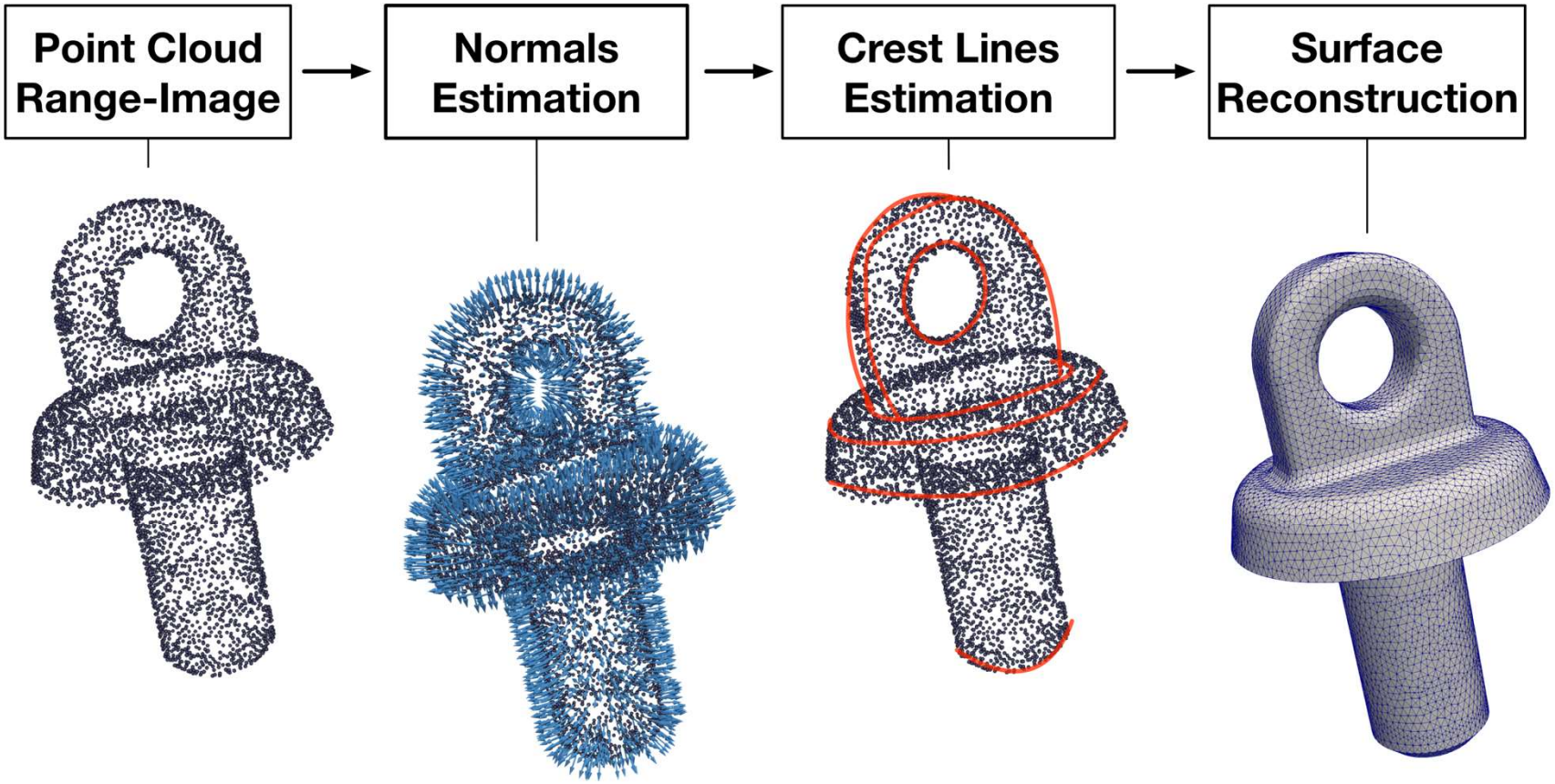
Perceptual Depth SR (ICCV 2019)



Roadmap Inspired by the Geometry Processing Pipeline



Reconstruction is Another Multi-Stage Procedure



ABC: A Big CAD Model Dataset for Geometric Deep Learning

S. Koch, A. Matveev, Z. Jiang, F. Williams, A. Artemov, E. Burnaev, M. Alexa, D. Zorin, D. Panozzo [**CVPR, 2019**]

arXiv:1812.06216

DEF: Deep Estimation of Sharp Geometric Features in 3D Shapes

Albert Matveev, Alexey Artemov, Ruslan Rakhimov, Gleb Bobrovskikh, Daniele Panozzo, Denis Zorin, Evgeny Burnaev [**Submitted, 2020**]

arXiv:2011.15081

Denis Zorin

Daniele Panozzo

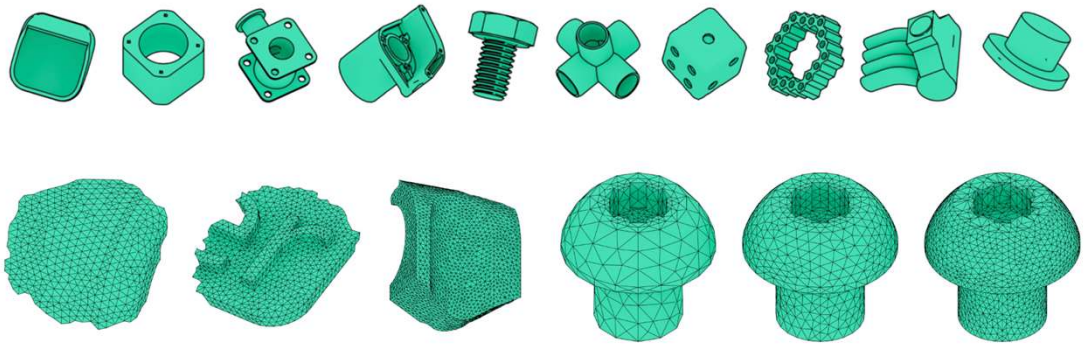
Evgeny Burnaev

Sebastian Koch

Alexey Artemov

Albert Matveev

ABC: A Big CAD Model Dataset For Geometric Deep Learning



Features:

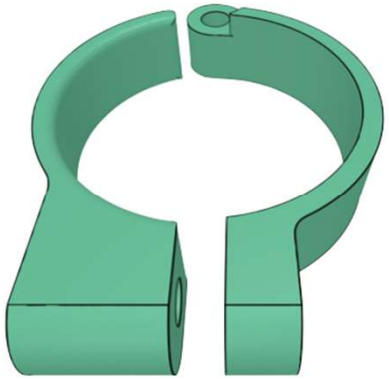
- 1,000,000 *parametric* 3D CAD models
- Accurate geometric ground truth
  - patch decomposition
  - sharp feature lines
  - analytic differential properties

Applications:

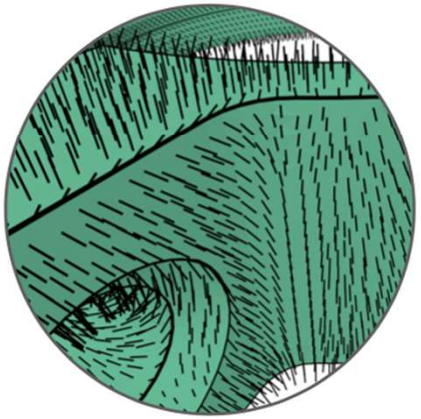
- Resampling at arbitrary resolutions
- Support connectivity, point-set, and parametric representations

Deep Networks for point clouds, meshes, etc.

Surface Normal Estimation Dataset

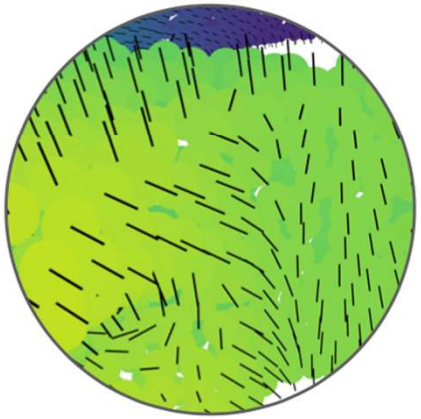


3D shape



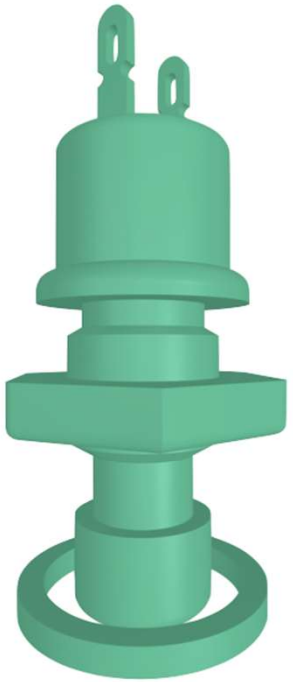
Analytic (vector) normals

Point sampling

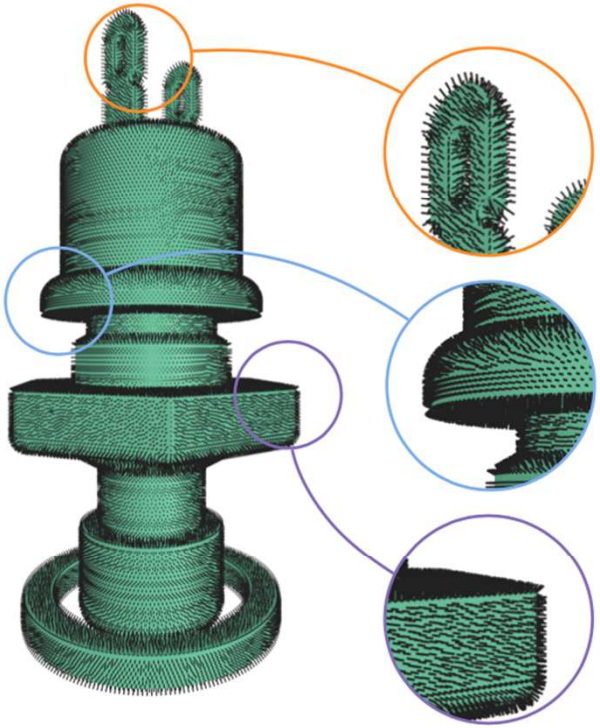


Annotated data instances

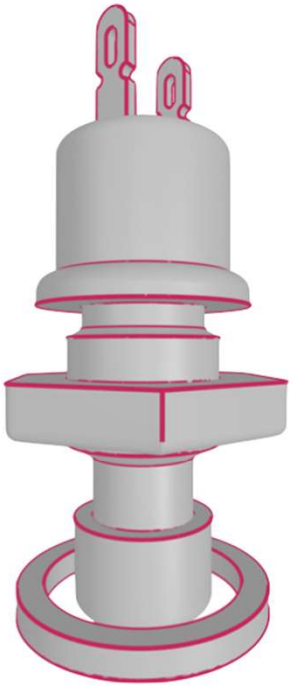
Sharp Features: Fields and Segments



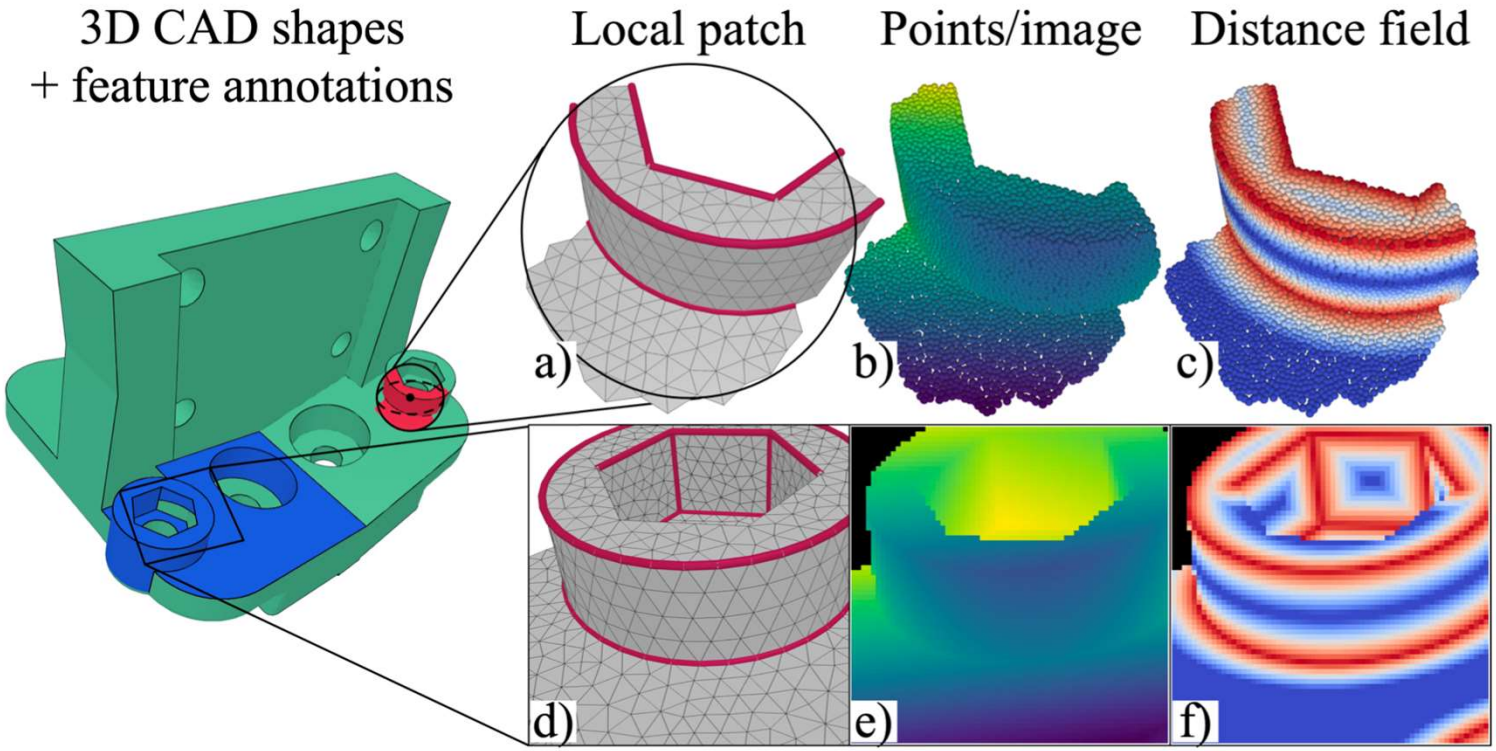
CAD model



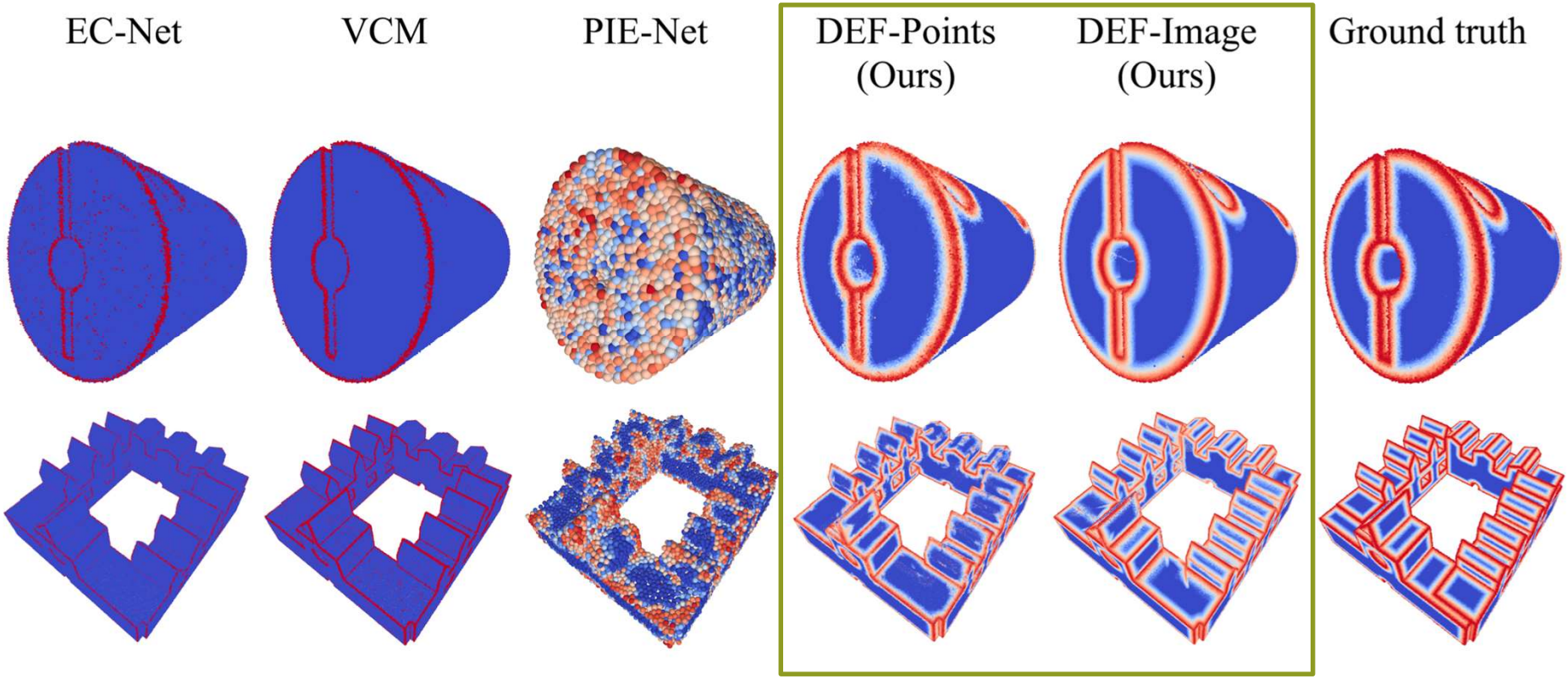
Rapid change in normals



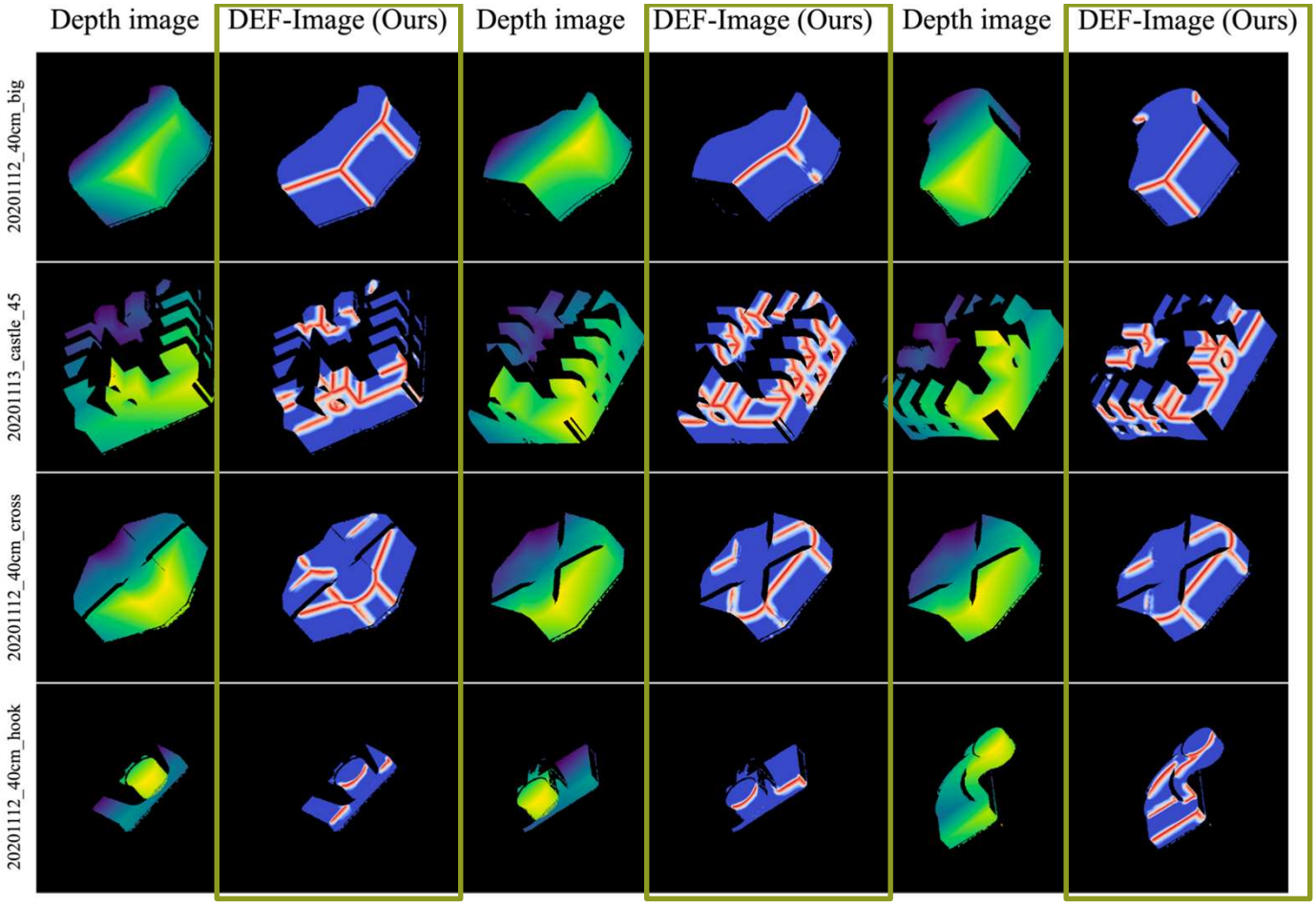
Sharp feature lines



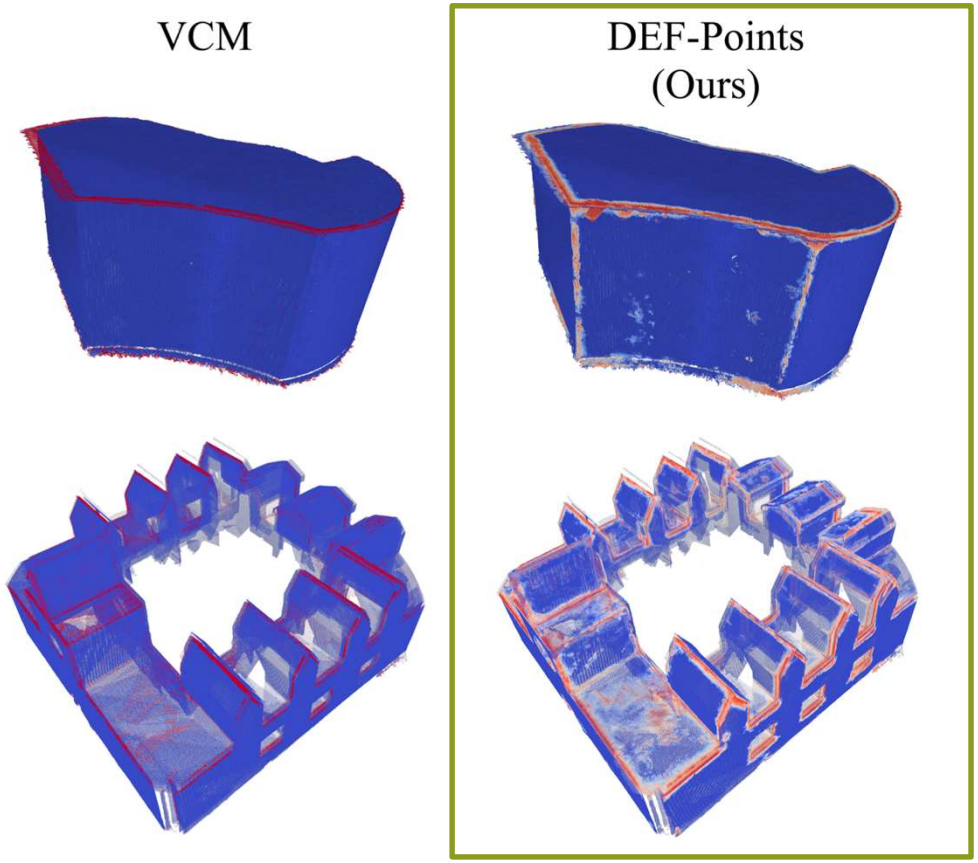
Our patch-based pipeline for generating point-based (upper row) and image-based (lower row) training instances starts with a 3D CAD shape (left) and
a)–d) extracts local triangulated patches, keeping interior features curves,
b)–e) samples point clouds and creates depth images, and
c)–f) computes local distance field annotations



Distance field estimation results on full 3D shapes constructed by our methods and the baseline methods EC-Net, VCM, and PIE-Net.
Unlike the baselines, our methods is able to robustly reconstruct a pointwise distance field and scales to 3D shapes represented by millions of points.



Example results of predicting distance-to-feature fields on real-world depth images obtained using an industrial structured light scanner.



Note that our method is at the same time detecting more sharp features and avoiding more outliers

Qualitative comparison of our algorithm (right) and VCM (left) on two noisy collection of registered range scans

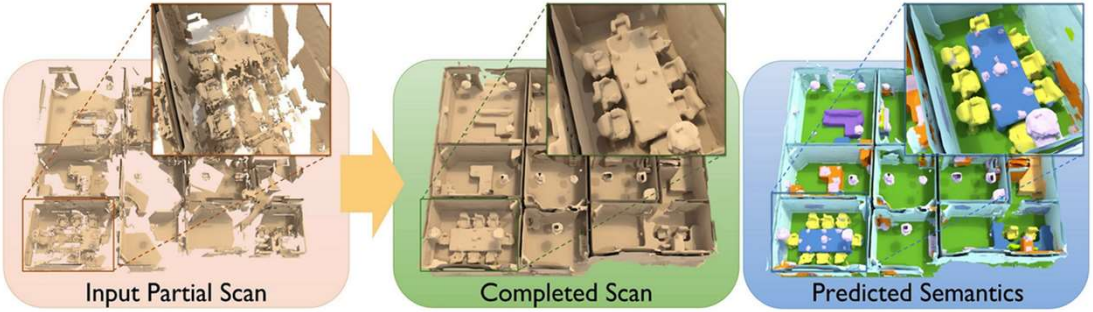
CAD-Deform: Deformable Fitting of CAD Models to 3D Scans

Vladislav Ishimtsev[1*], Alexey Bokhovkin[1*], Alexey Artemov[1],
Savva Ignatyev[1], Matthias Nießner[2],
Denis Zorin[1,3], and Evgeny Burnaev[1]

[1] Skolkovo Institute of Science and Technology, Russia
[2] The Technical University of Munich, Germany
[3] New York University, USA

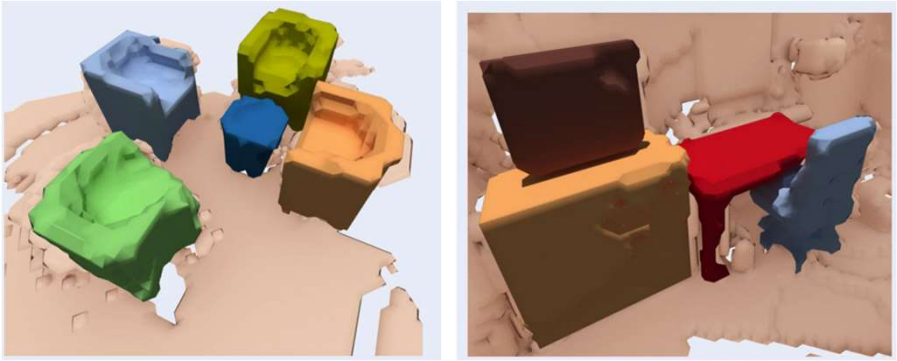
Previous Work on Understanding

- Incomplete scans



ScanComplete [3]

- Lack of fine-grained details



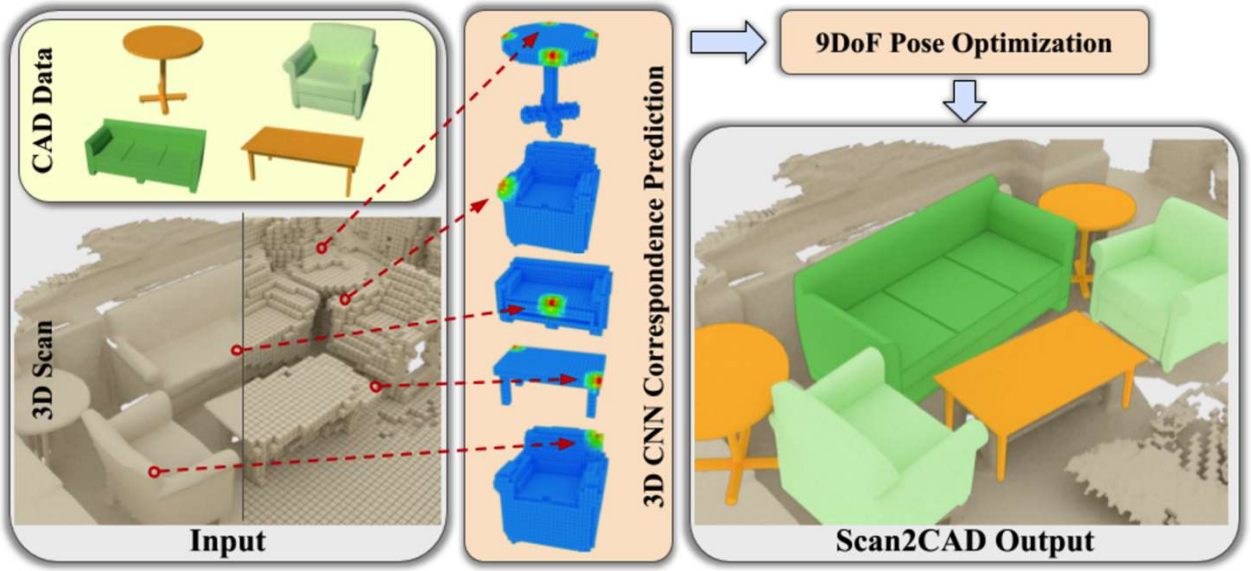
3D-SIC [4]

Bringing CAD models

Large shape catalogues (ShapeNet) allow to reduce the problem to **retrieval** and **alignment** subtasks.

- Object detection
- Sparse correspondences
- Shape matching

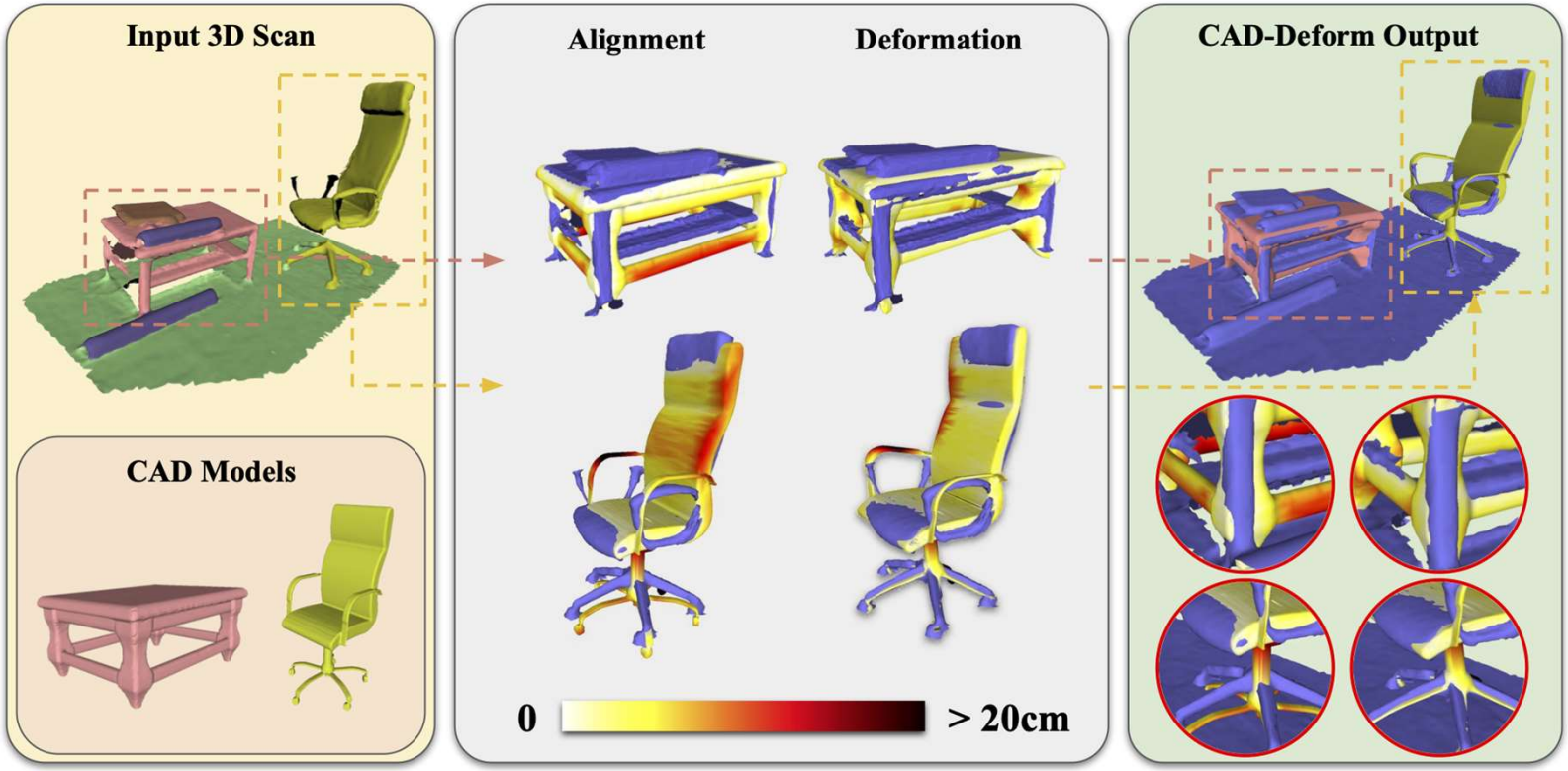
- Lack of shape diversity
- Inaccurate alignment
- Imperfect retrieval



Scan2CAD [5]

CAD-Deform (ECCV 2020)

Goal: noisy scene scans + CAD models → 3D scene reconstruction w/ tight fits & high-quality surfaces



Theory and Algorithms (Mesh Deformation)

Linear terms

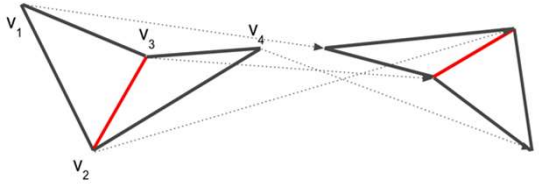
$$\mathcal{E}(\mathbf{V},\mathbf{P}) = \underbrace{\boxed{E_{shape} + \alpha_{smooth} E_{smooth} + \alpha_{sharp} E_{sharp}}}_{\text{quadratic problem}} + \alpha_{data} E_{data},$$

$$E_{shape} = \underbrace{\sum_{e \in \mathbf{E}} \|T_e(\mathbf{V}) - T_e^0\|_2^2}_{\text{deviation}}; \; E_{smooth} = \sum_{f \in \mathbf{F}} \sum_{e_i, e_j \in f} \|T_{e_i}(\mathbf{V}) - T_{e_j}(\mathbf{V})\|_2^2;$$

$$E_{sharp} = \sum_{k=1}^{n_p} \sum_{e_s \in \mathbf{E}_{sharp}^k} \|T_{e_s}(\mathbf{V}) - T_{e_{s+1}}(\mathbf{V})\|_2^2;$$

$$T_e : \mathrm{R}^3 \to \mathrm{R}^3$$

$$T_e[q_{v_1}^0, q_{v_2}^0, q_{v_3}^0, q_{v_4}^0] = [q_{v_1}, q_{v_2}, q_{v_3}, q_{v_4}]$$



Non-linear terms

Data term:

$$E_{data} = f_{data}(\mathbf{V},\mathbf{P})$$

With exact correspondences:

$$f_{data}^{nn}(\mathbf{V},\mathbf{P}) = \sum_{c \in \mathbf{C}} \sum_{p \in \mathbf{P}_c} \|p - v_{\iota^c(p)}\|_2^2$$

Correspondence-free:

$$f_{data}^{p2p}(\mathbf{V},\mathbf{P}) = \sum_{c \in \mathbf{C}} \sum_{v \in \mathbf{V}_c} \sum_{p \in \mathbf{P}_c} \xi^{\sigma}(p, \mathbf{V}_c)(d^{\varepsilon}(v - p))^2$$

Methodology (Datasets)

ScanNet

1506 scenes with 2-10 objects

ShapeNet

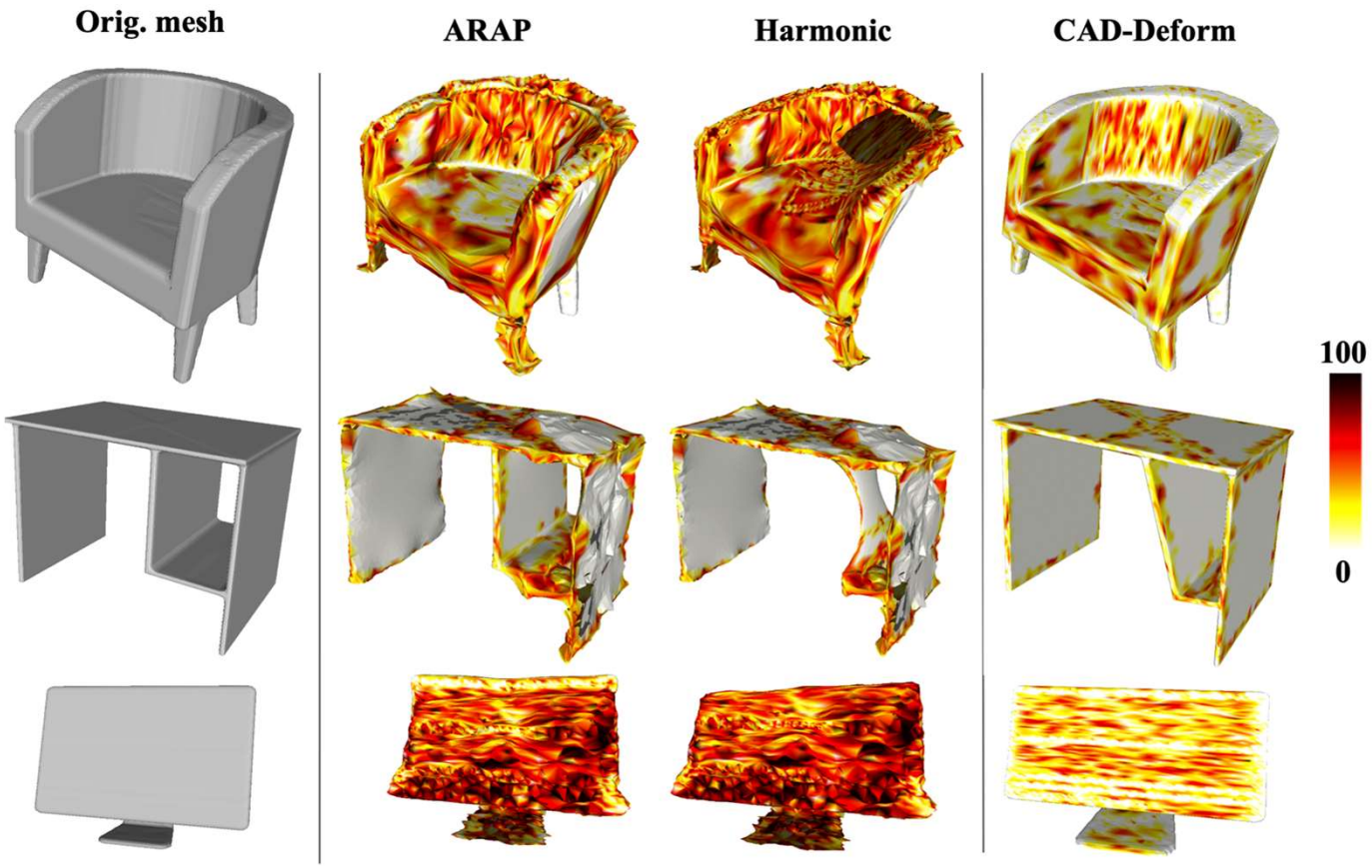
1400 CADs remeshed to 10k-14k vertices

PartNet

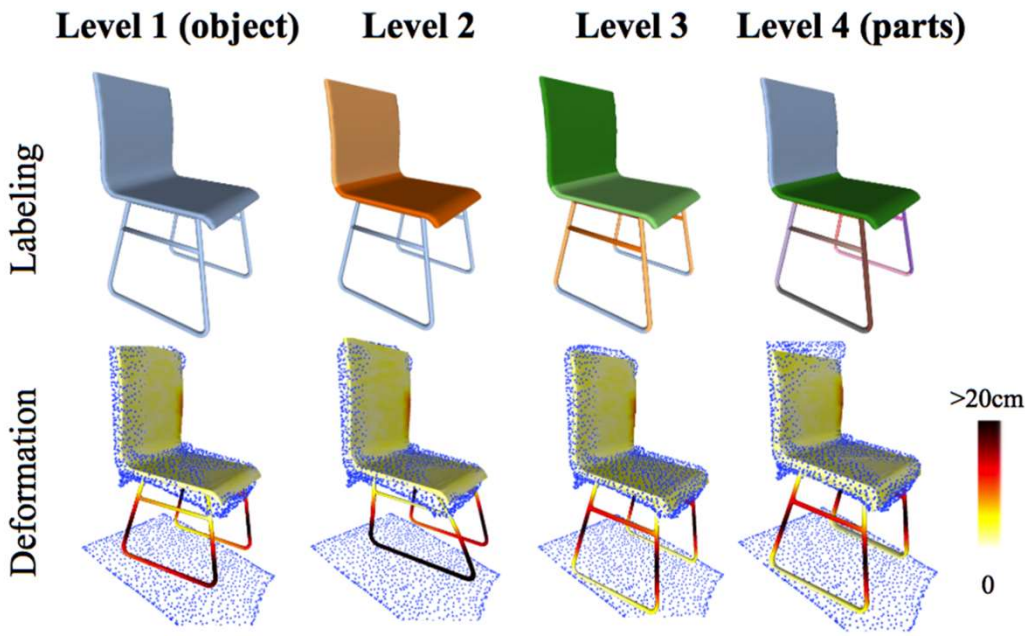
Projected part labels on remeshed CADs

Results

DAME visualization



Results



Accuracy, %	class avg.	avg.
Ground truth	89.22	90.56
Level 1 (object)	89.25	90.79
Level 2	89.16	91.21
Level 3	89.40	91.05
Level 4 (parts)	**91.65**	**93.12**

Accuracy on different levels of detail

General comments

- Lack of labeled/paired data
  - ✓ Collection of paired multi-scale/fidelity datasets
  - ✓ Development of unpaired/self-supervised approaches

- New representation for 3D data
  - ✓ Implicit functions (?)

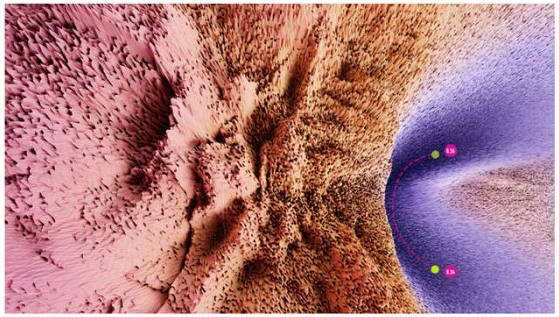
- Multi-view super-resolution

- Automatic mesh generation, mesh denoising and remeshing from noisy scan data guided by predicted sharp features

- Real scene analysis with taking into account (in unsupervised way) interaction between different objects and their parts

Skoltech

Skolkovo Institute of Science and Technology



ML <- Geometry

Topological obstructions in neural networks' learning

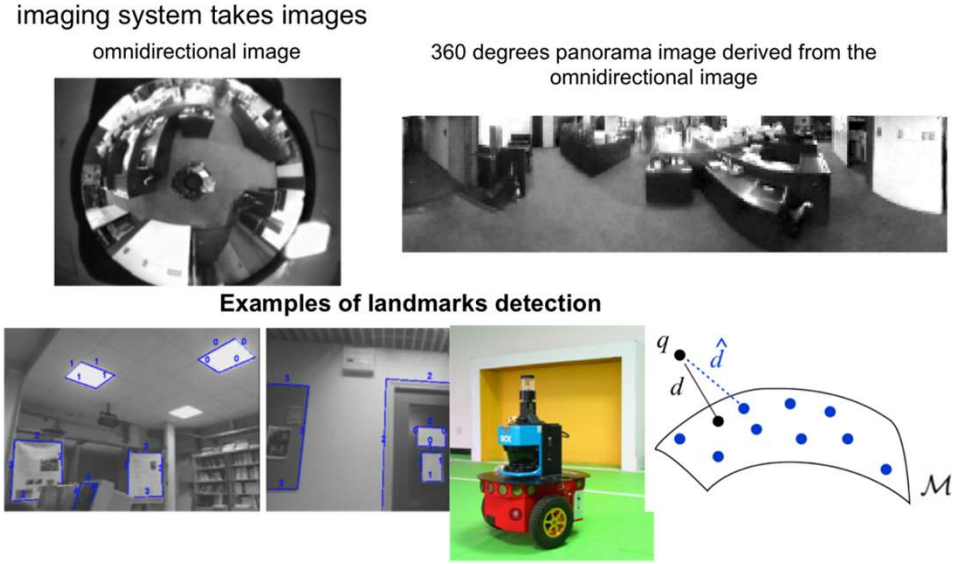
Talk by **Serguei Barannikov**[1] (**Tuesday, 12:30**)

joint work with Alexander Korotin[1], Dmitry Oganesyan[1], Daniil Emtsev[1], Grigory Sotnikov[1], Evgeny Burnaev[1]

[1]Skolkovo Institute of Science and Technology, Russia

Conclusion

- We develop new machine learning approaches to efficient predictive modelling and geometric machine learning
- We collaborate with industrial companies (Bosch, Huawei, Sberbank, GazpromNeft, etc.) and produce solutions, approved by the industry
- We constantly seek new advanced ML problems, their industrial applications and new collaboration



Evgeny Burnaev
e.burnaev@skoltech.ru
adase.group

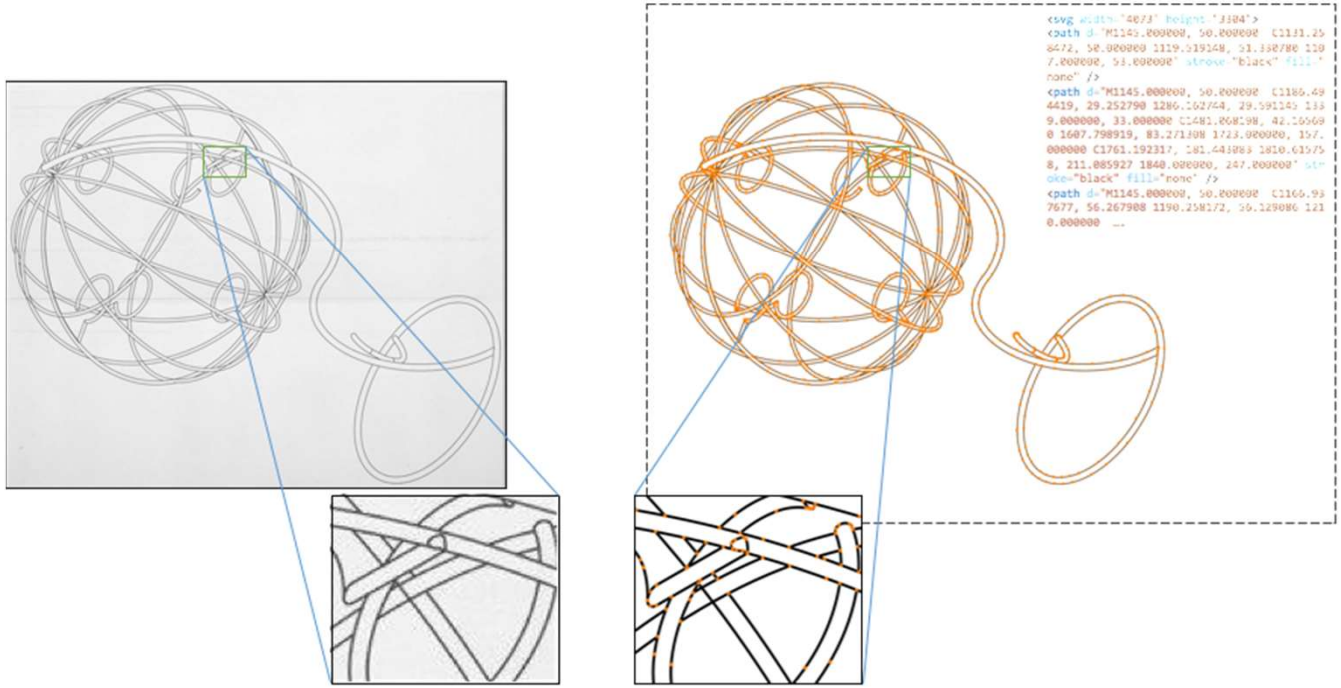
Deep Vectorization of Technical Drawings

Vage Egiazarian[1*], Oleg Voynov[1*], Alexey Artemov[1], Denis Volkhonskiy[1], Aleksandr Safin[1], Maria Taktasheva[1], Denis Zorin[2,1], Evgeny Burnaev[1]

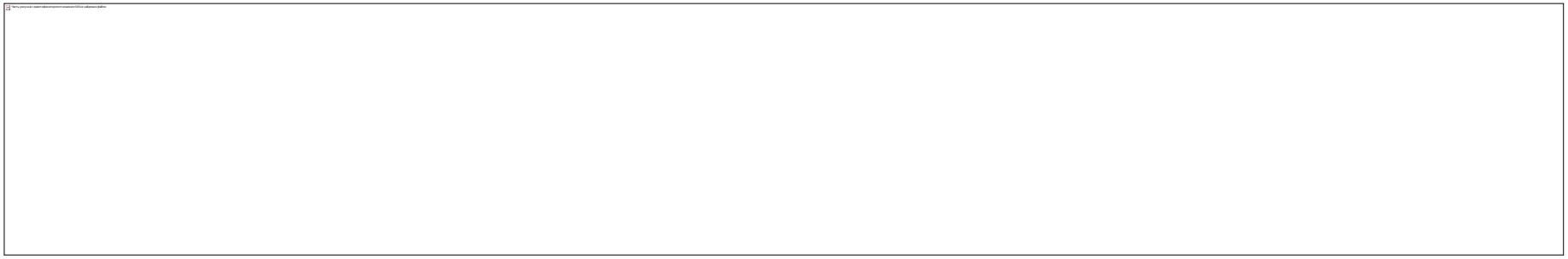
[1] Skolkovo Institute of Science and Technology, Russia
[2] New York University, USA

Deep Vectorization (ECCV 2020)

Goal: noisy technical raster drawings → precise, robust, and compact vector representation



Overview



Input raster image | Noise removal, contrast adjustment, hole inpainting | Estimation of primitives | Refinement of primitives | Merging of patches

Requirements of technical vectorization

- Exact representation
- No noise in output
- Minimal number of primitives

Overview

Input raster image | Noise removal, contrast adjustment, hole inpainting | Estimation of primitives | Refinement of primitives | Merging of patches

Requirements of technical vectorization

- Exact representation
- No noise in output
- Minimal number of primitives

Overview

Input raster image

Noise removal, contrast adjustment, hole inpainting

Estimation of primitives

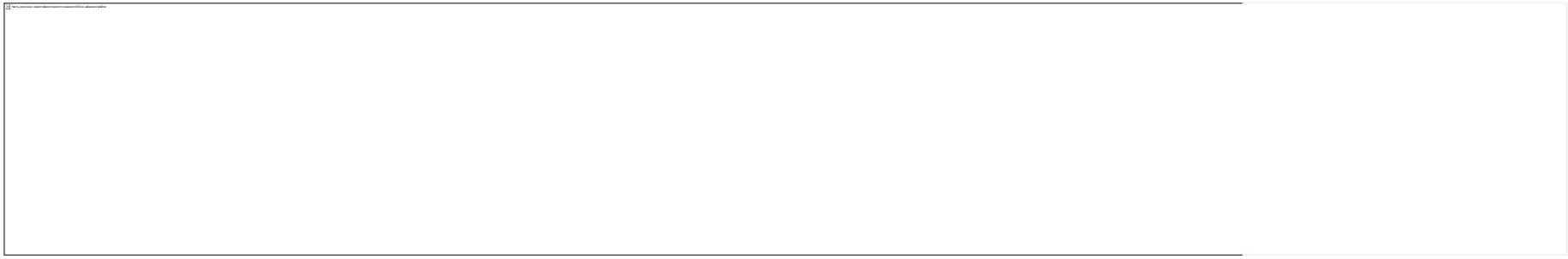
Refinement of primitives

Merging of patches

Requirements of technical vectorization

- Exact representation
- No noise in output
- Minimal number of primitives

Overview

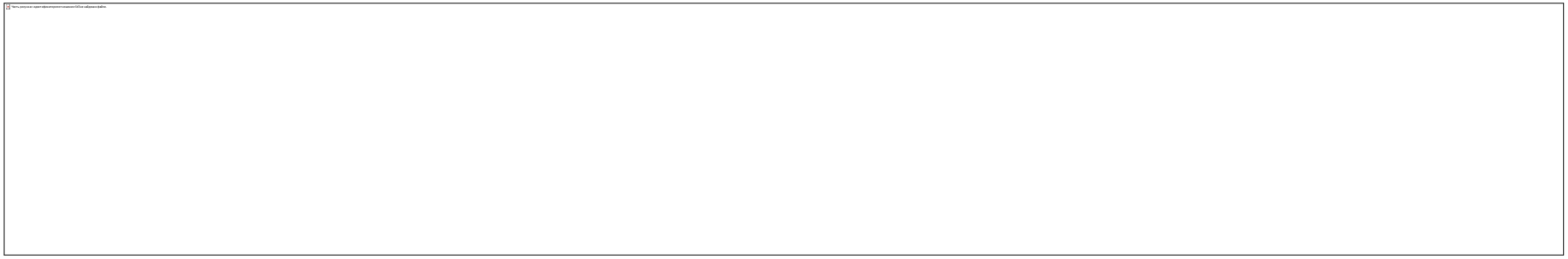


Input raster image | Noise removal, contrast adjustment, hole inpainting | Estimation of primitives | Refinement of primitives | Merging of patches

Requirements of technical vectorization

- Exact representation
- No noise in output
- Minimal number of primitives

Overview



Input raster image

Noise removal, contrast adjustment, hole inpainting

Estimation of primitives

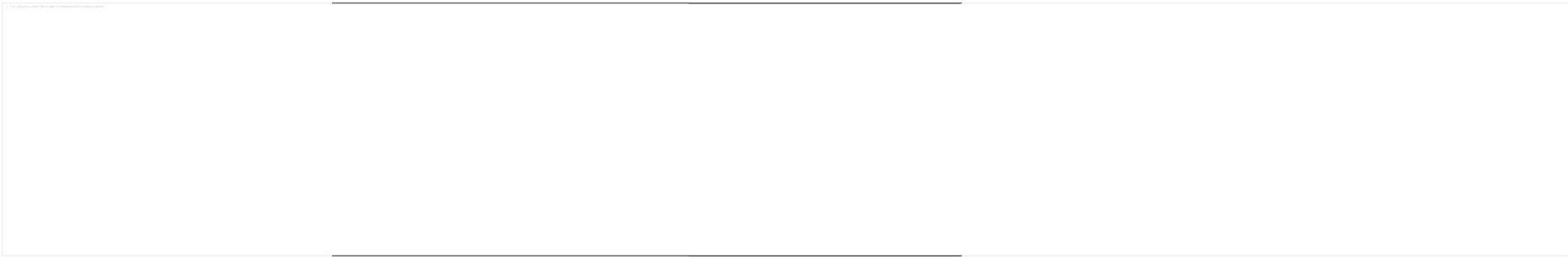
Refinement of primitives

Merging of patches

Requirements of technical vectorization

- Exact representation
- No noise in output
- Minimal number of primitives

Overview



Input raster image

Noise removal, contrast adjustment, hole inpainting

Estimation of primitives

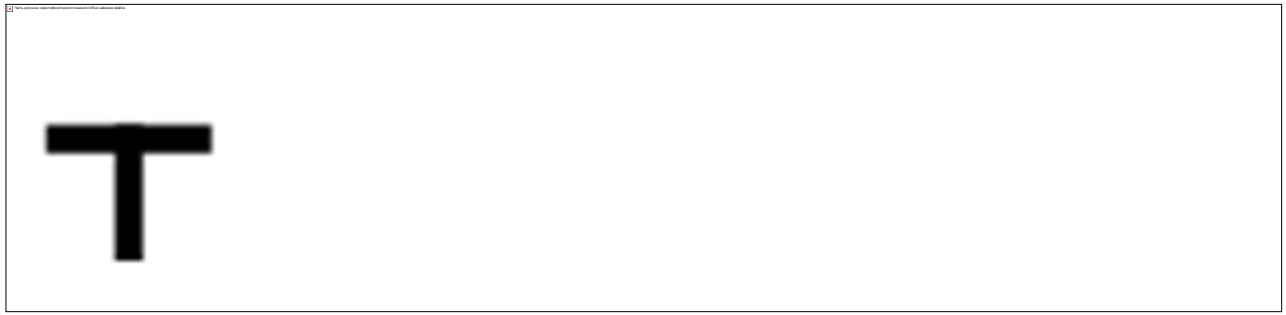
Refinement of primitives

Merging of patches

Requirements of technical vectorization

- Exact representation
- No noise in output
- Minimal number of primitives

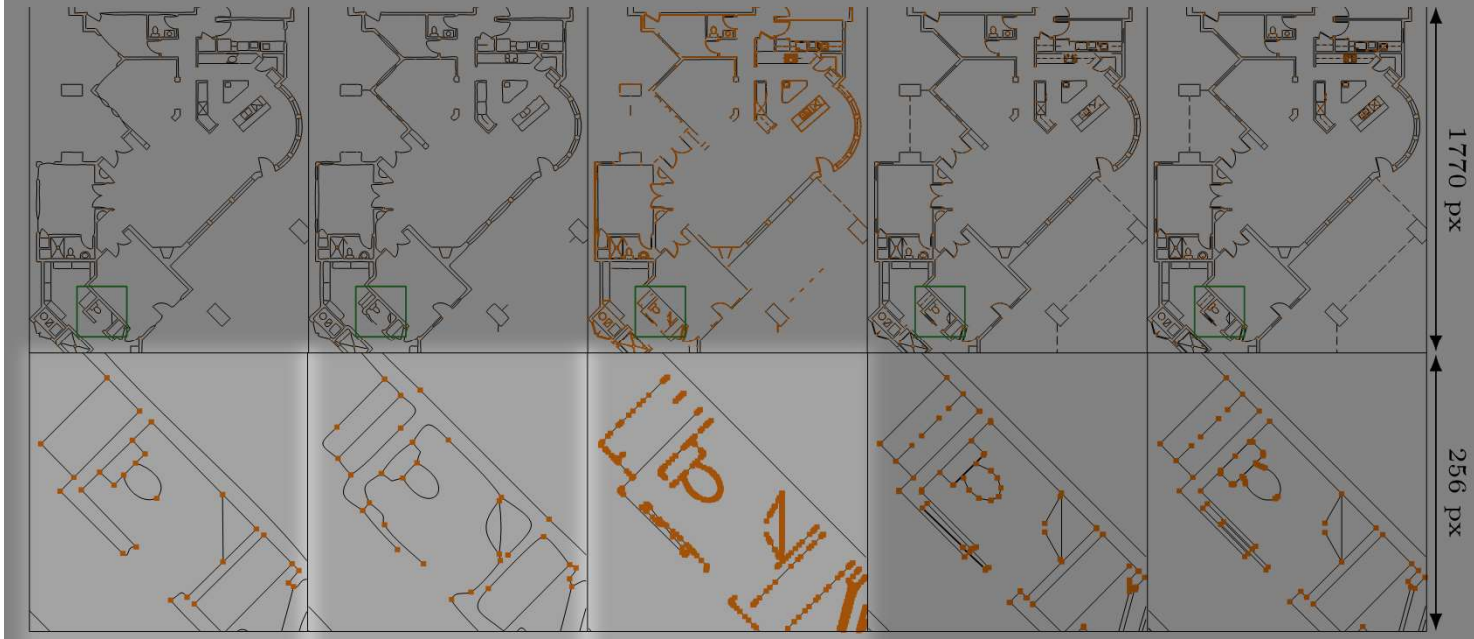
Estimation of primitives



```
Primitive {
   control_point_1_x;
   control_point_1_y;
   ...
   control_point_k_y;
   width;
   confidence;
}
```

- Operation on patches for increased efficiency and robustness
- Learning for automatic optimal representation
- Attention to leverage sparsity of line drawings

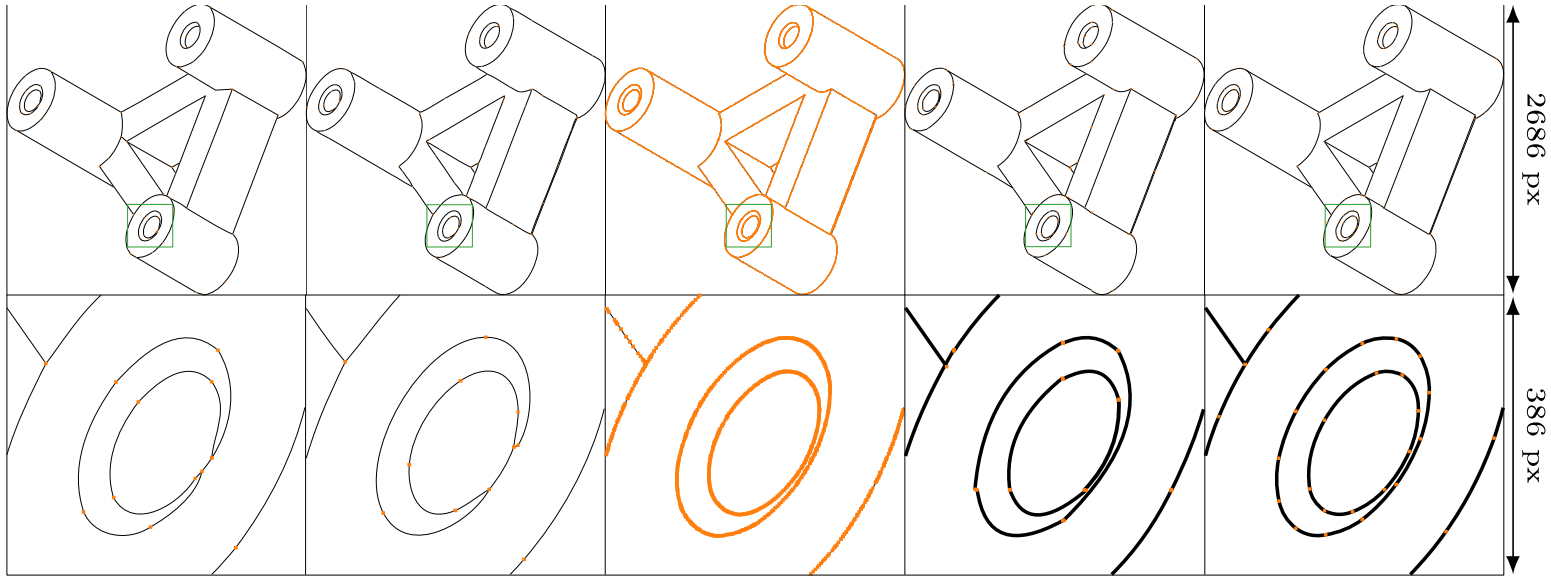
Comparisons



Favreau et al. 16	Donati et al. 19	Bessmeltsev and Solomon. 19	Our method	Ground truth
29% / 415px	21% / 215px	64% / 140px	**89% / 28px**	
4.2px / 615	1.9px / 1192	0.9px / 35k	**0.2px** / 1286	#P 1634

	IoU,%	d_H, px	d_M, px	#P
Favreau et al. 16	31	381	2.8	696
Donati et al. 19	22	214	2.1	1214
Bessmeltsev and Solomon. 19	60	204	1.5	38k
Our method	**88**	**25**	**0.2**	1331

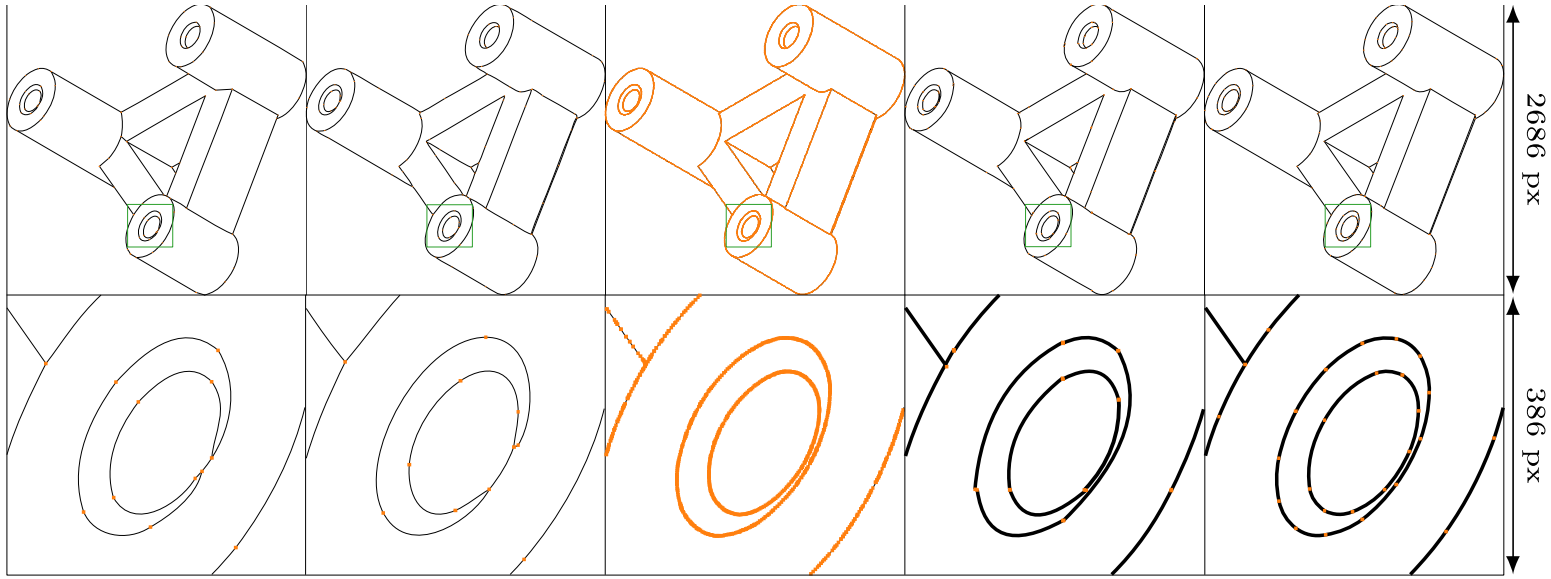
IoU / max error d_H
avg error d_M / num of primitives #P



Favreau et al. 16	Donati et al. 19	Bessmeltsev and Solomon. 19	Our method	Ground truth
67%/ 32px 1.1px/ 79	67%/ 7px 1.0px/ 108	95%/ 4px 0.2px/ 9.5k	86%/ 5px 0.4px/ 139	#P 139

	IoU,%	d_H, px	d_M, px	#P
Favreau et al. 16	65	38	1.7	63
Donati et al. 19	60	**9**	1	109
Our method	**77**	19	**0.6**	97

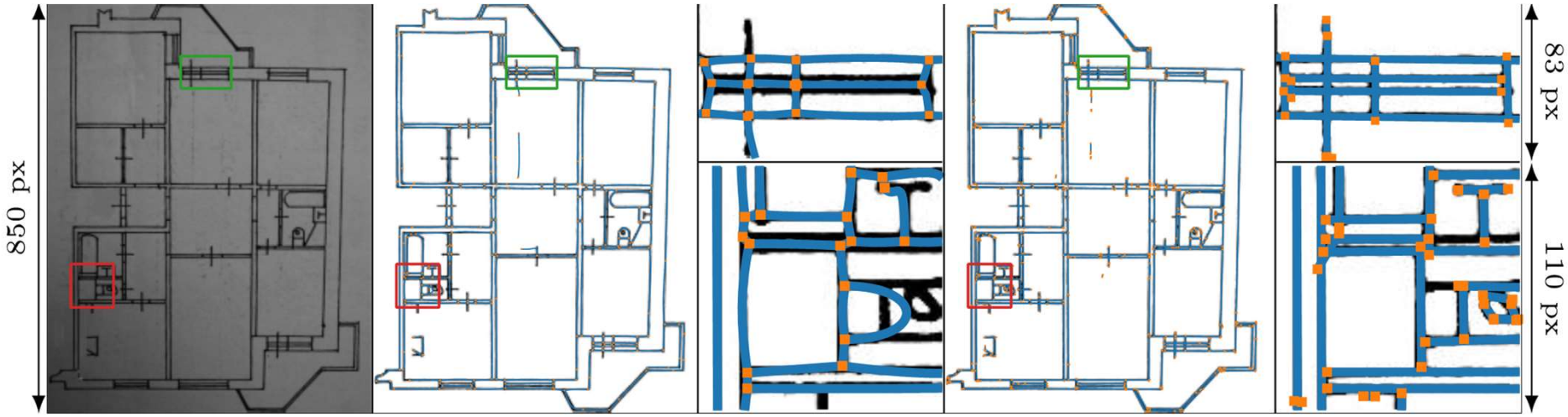
IoU / max error d_H
avg error d_M / num of primitives #P



Favreau et al. 16	Donati et al. 19	Bessmeltsev and Solomon. 19	Our method	Ground truth
67%/ 32px 1.1px/ 79	67%/ 7px 1.0px/ 108	95%/ 4px 0.2px/ 9.5k	86%/ 5px 0.4px/ 139	#P 139

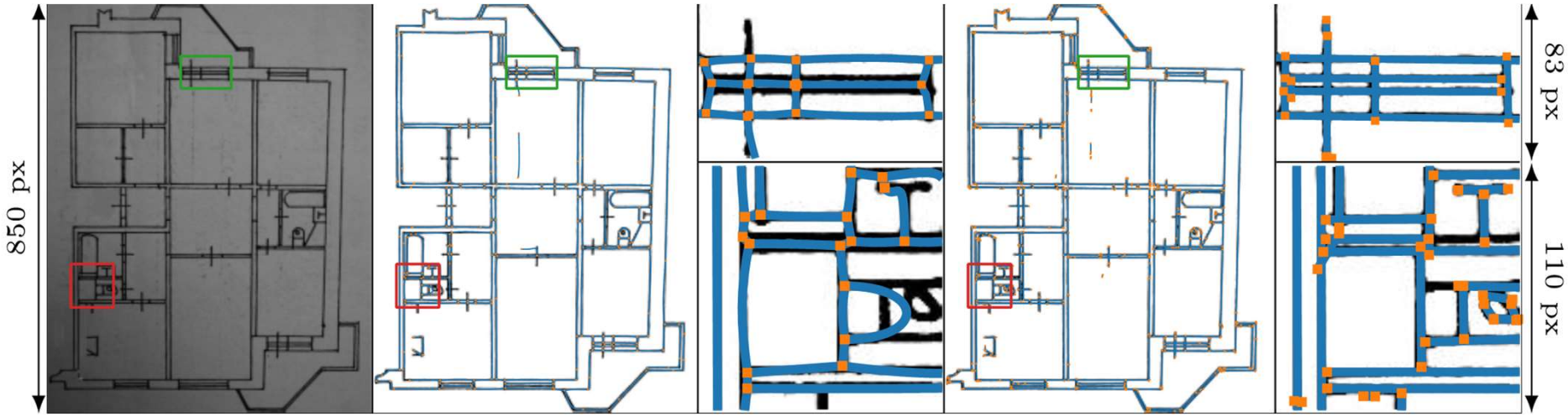
	IoU,%	d_H, px	d_M, px	#P
Favreau et al. 16	65	38	1.7	63
Donati et al. 19	60	**9**	1	109
Bessmeltsev and Solomon. 19	89	17	0.7	7818
Our method	77	19	0.6	97
w/o final merging	**91**	19	**0.3**	240

IoU / max error d_H
avg error d_M / num of primitives #P



Input image Donati et al. 19, 52% / 349 Our method, **78%** / 368

	IoU,%	#P
Donati et al. 19	47	329
Our method	**82**	452



Input image Donati et al. 19, 52% / 349 Our method, **78%** / 368

	IoU,%	#P
Donati et al. 19	47	329
Our method	**82**	452