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

Vage Egiazarian, Savva Ignatyev, **Alexey Artemov**, Oleg Voynov, Andrey Kravchenko,
Youyi Zheng, Luiz Velho, Evgeny Burnaev

Point Clouds and why they are Important!

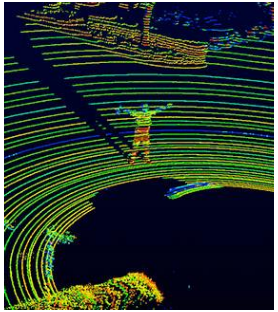
- → Millions of depth sensors available on the market
- → A point cloud is one of the most commonly used data structures in 3D reconstruction models.
- → Traditional 2D ConvNets cannot solve *all* problems!
- → Thus, 3D/4D and geometry-aware models are required!

Structured light

LiDAR

RGB-D sensors

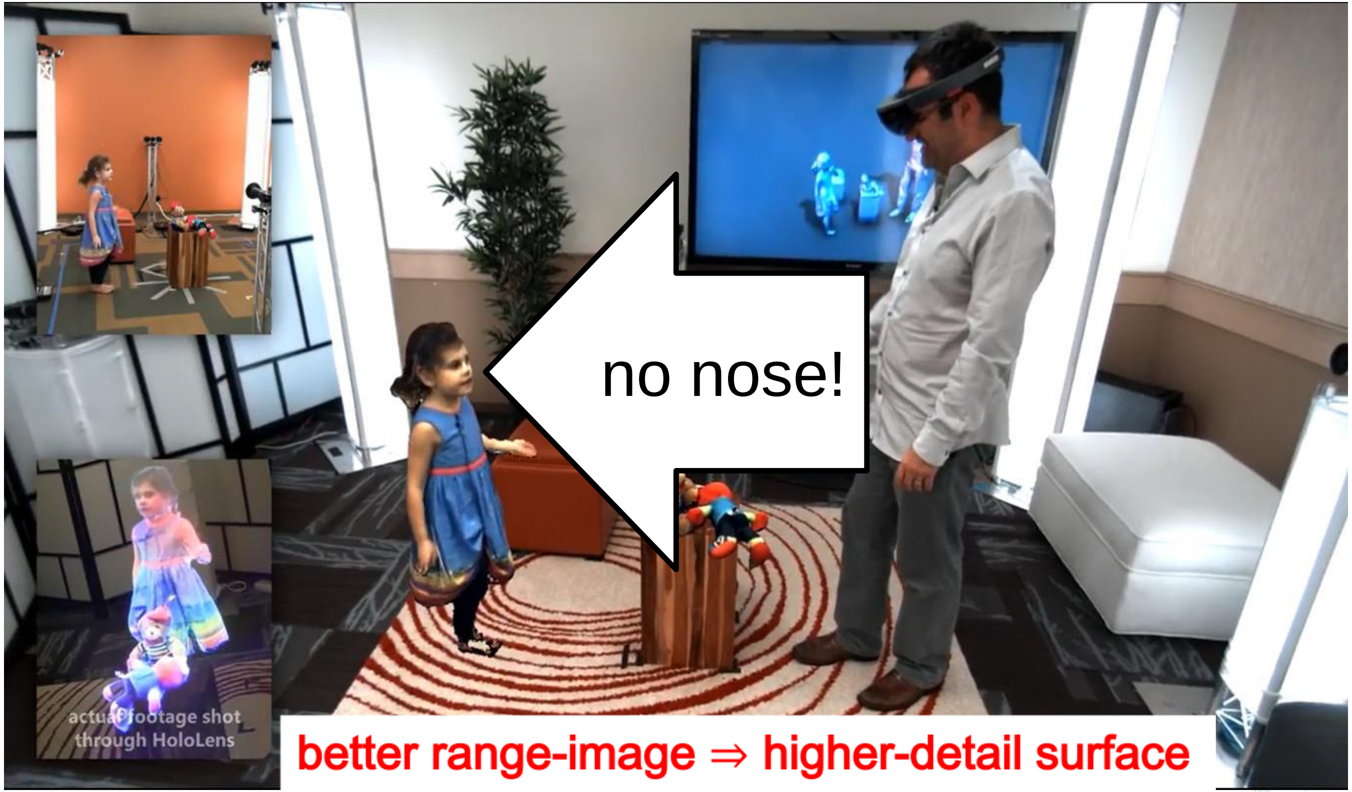
Structured light scans



LiDAR range-scans

Depth maps

Applications: Augmented Reality



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

Applications: Computer Vision

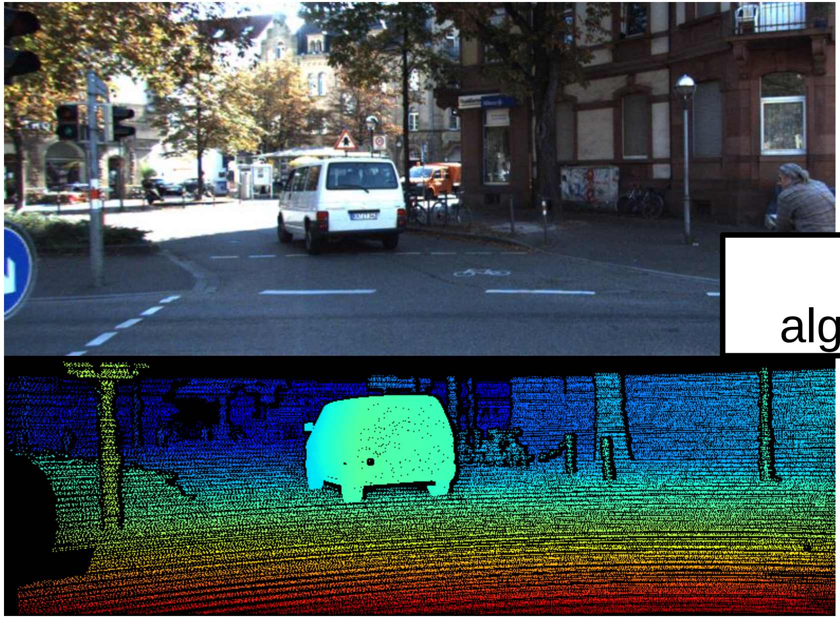
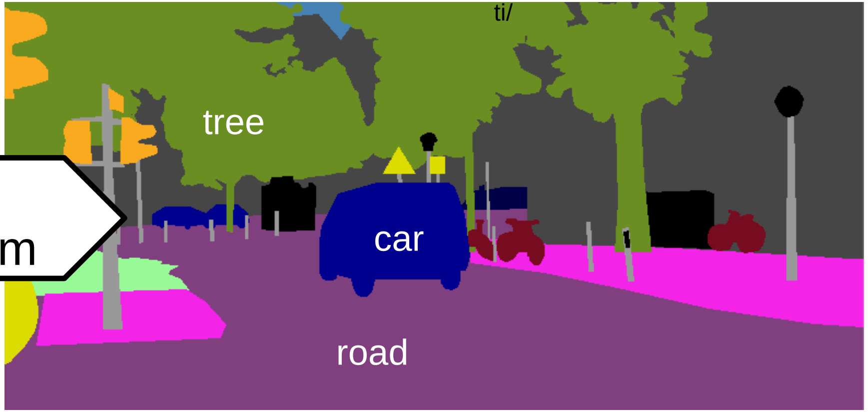


Photo + Range-image

CV algorithm

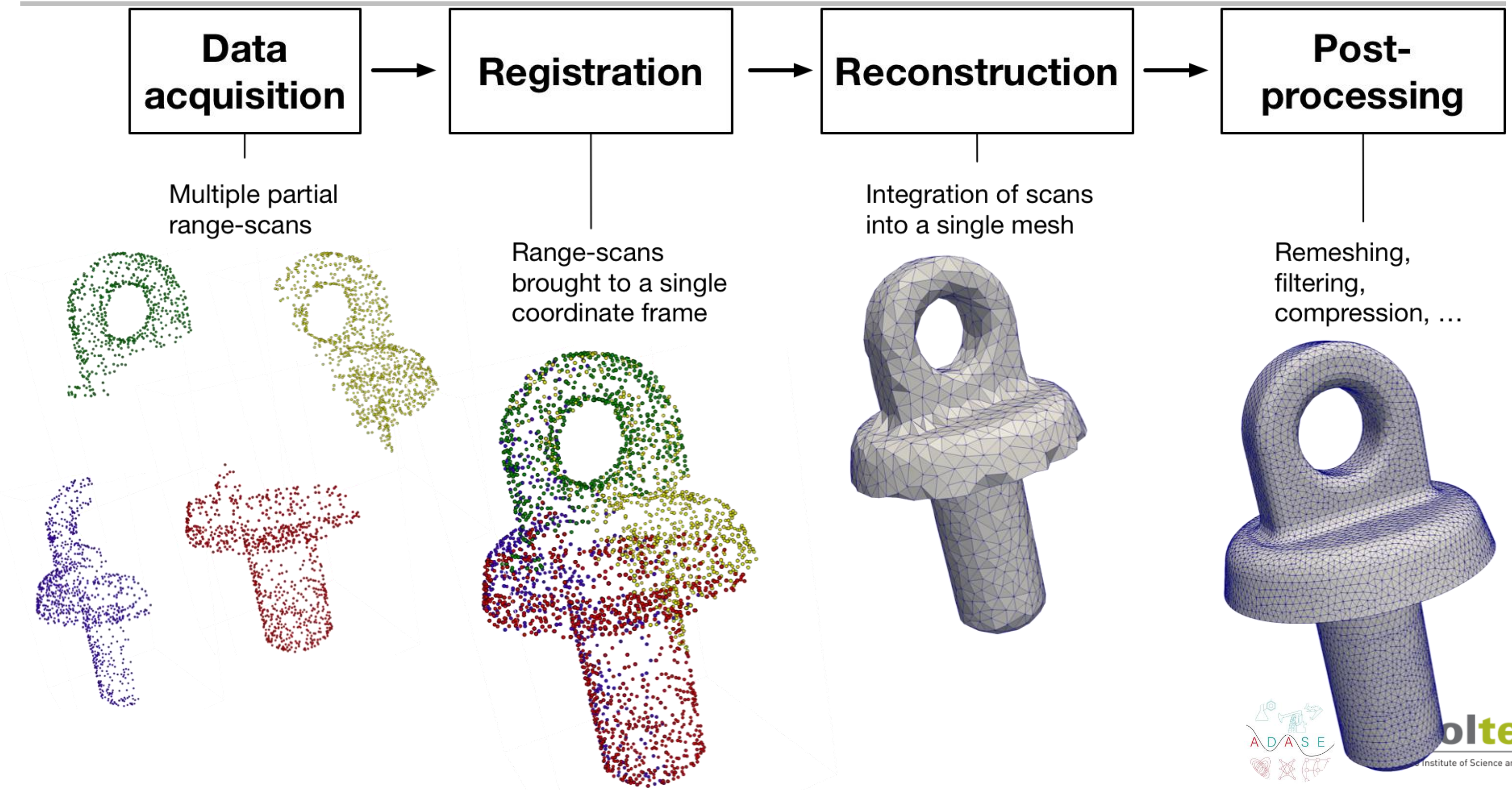
http://www.cvlibs.net/datasets/kitti/



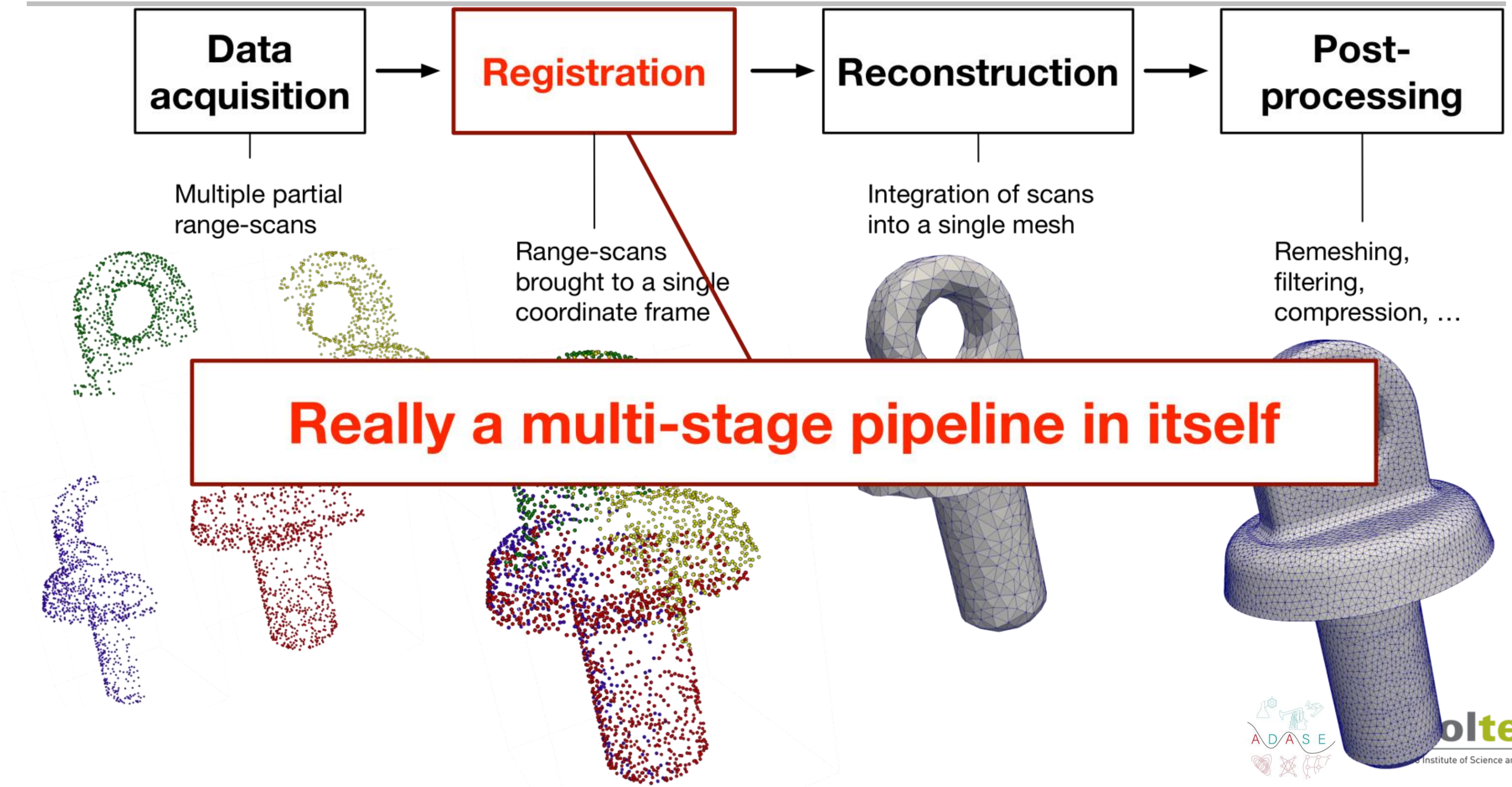
Scene understanding

better range-image ⇒ better scene understanding

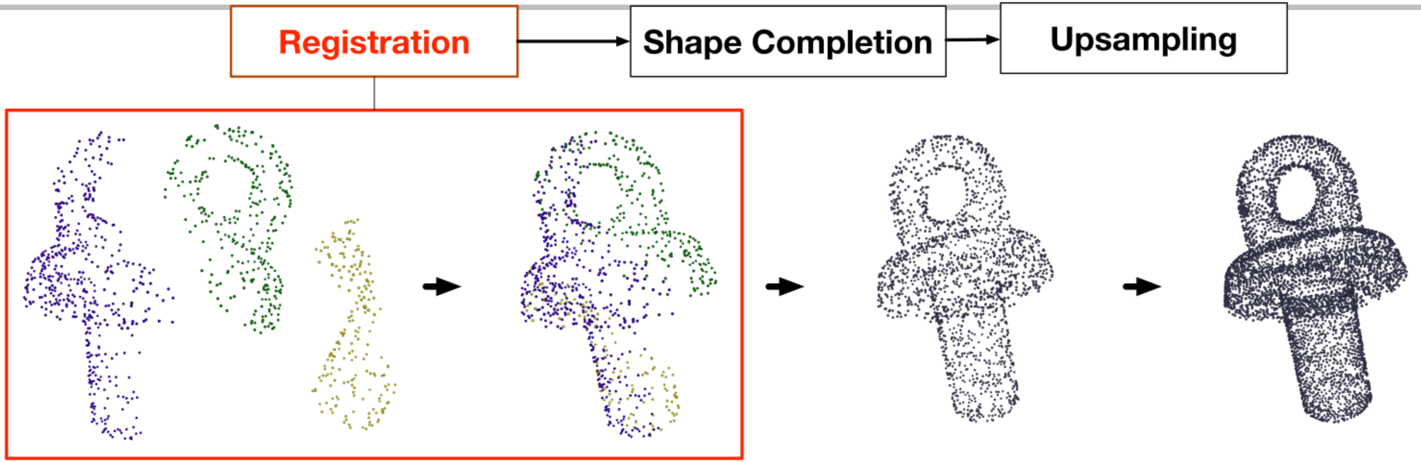
Common geometry processing pipeline



Common geometry processing pipeline



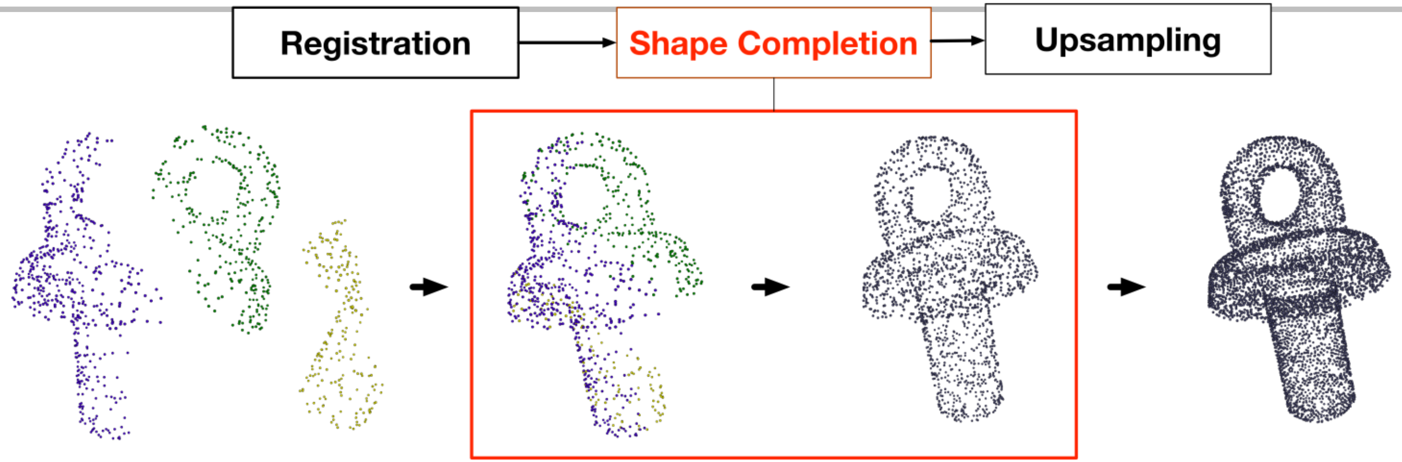
Point cloud processing: a multi-stage procedure



Registration:

- Given $M_1, \ldots, M_N$ partial range-scans, find transformations $T_2, \ldots, T_N$ s.t. $M_1 \cong T_2(M_2) \cong \ldots \cong T_N(M_N)$

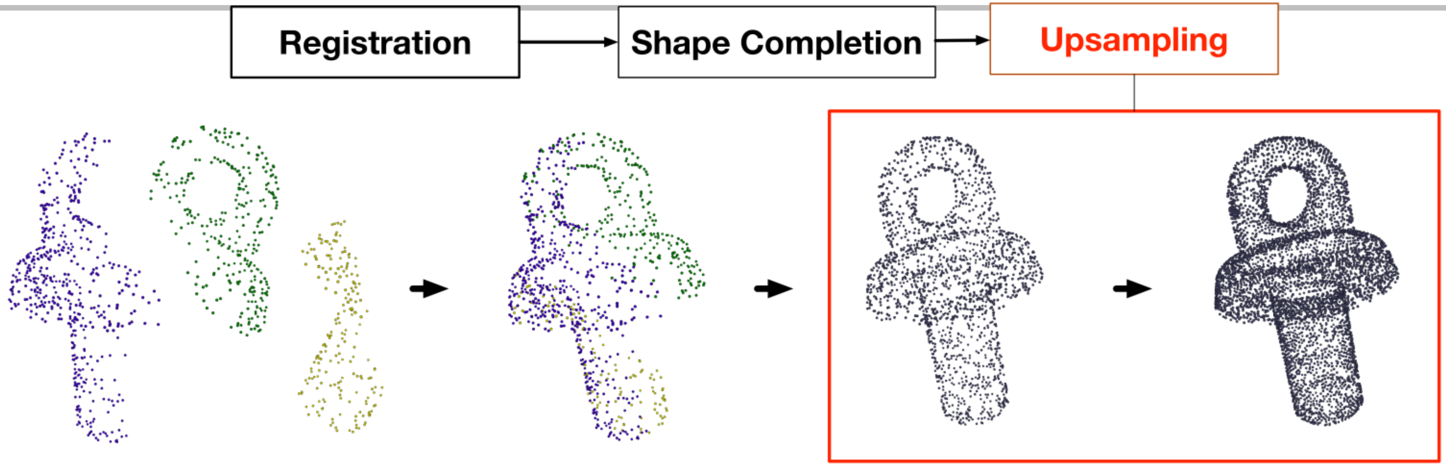
Registration is Really a Multi-Stage Procedure



Shape Completion:

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

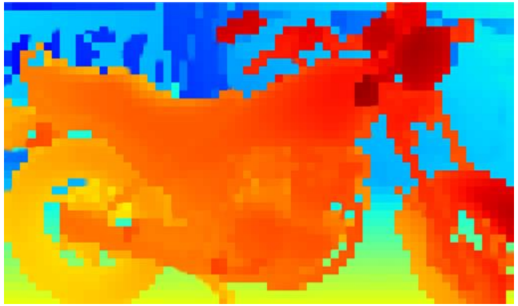
Registration is Really a Multi-Stage Procedure



Point cloud upsampling:

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

Depth map upsampling leverages CNNs

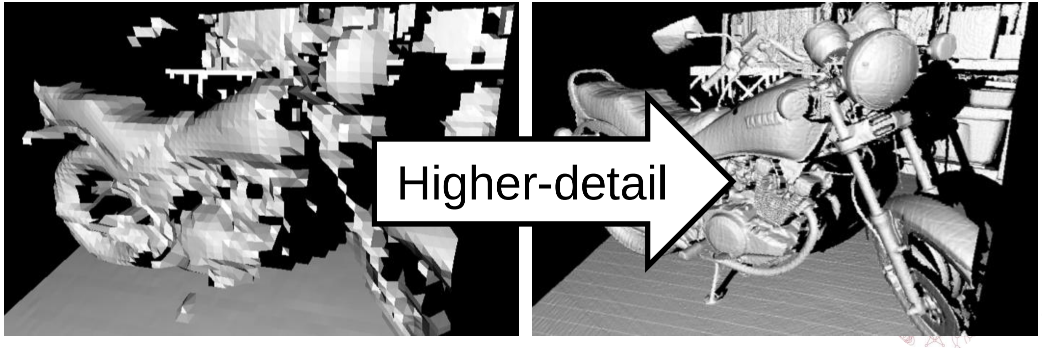


Input

Fast, neural network

Target

Corresponding surfaces



Skoltech
Skolkovo Institute of Science and Technology

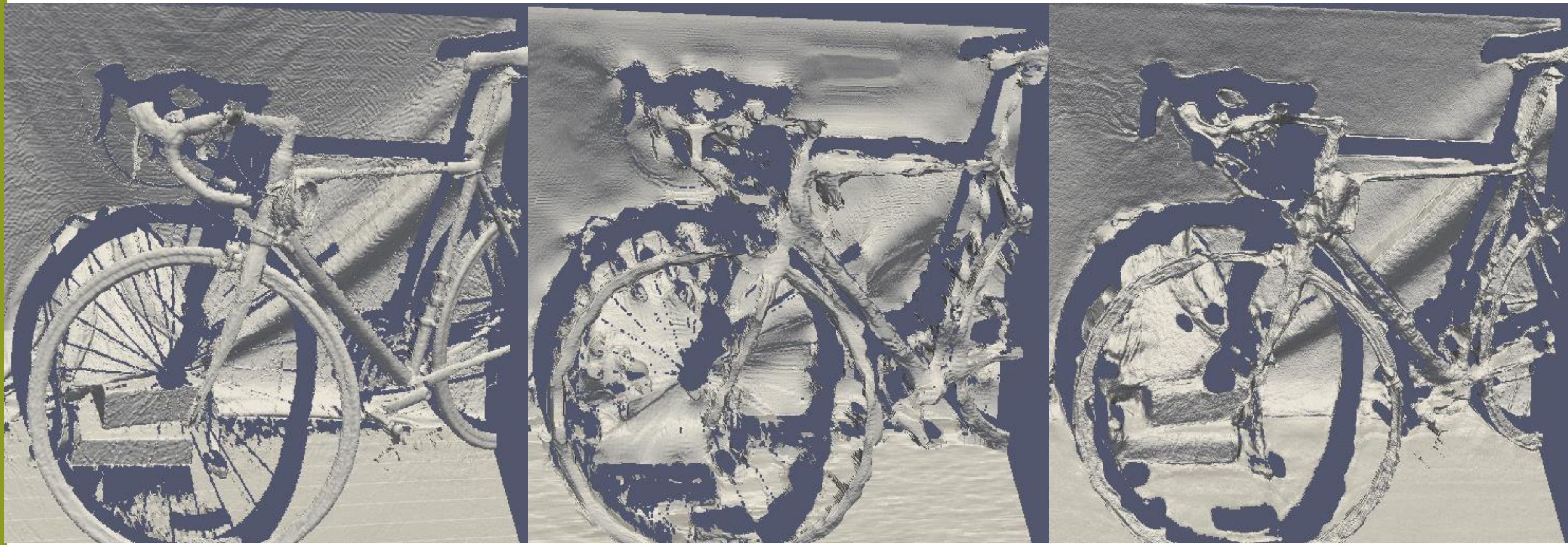
8x super-resolution

Ground truth

Low-resolution input

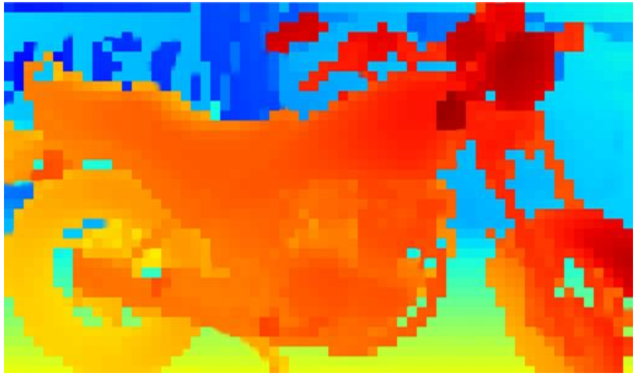
[Riegler et al. 2016]

Reconstructed surface

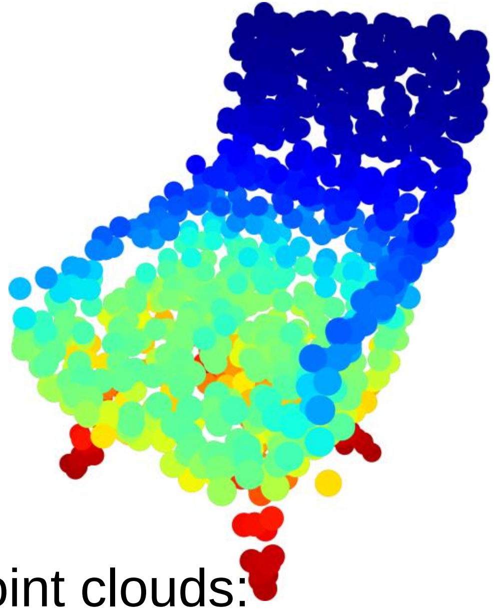


Ground truth [Riegler et al. 2016] [Voynov et al., 2019]

However…



≠



Depth images:
2D image

3D point clouds:
completely unstructured!

Images: Laplacian GANs [Denton et al., NIPS 2015]

- 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

- **Can we perform similar multi-stage upsampling with point clouds?**

Latent Space Laplacian Pyramid GAN

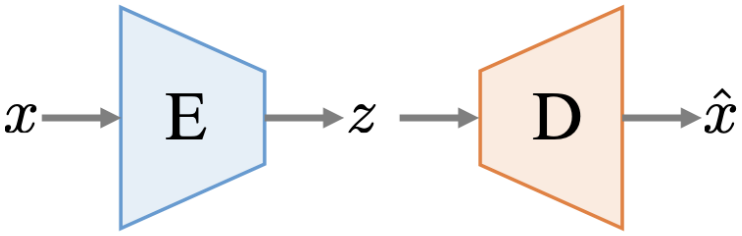
- Generative model that produces synthetic shapes
- High-resolution point clouds via a latent space Laplacian pyramid
- Operates in the space of latent codes, i.e. does not require dimensionality reduction during training

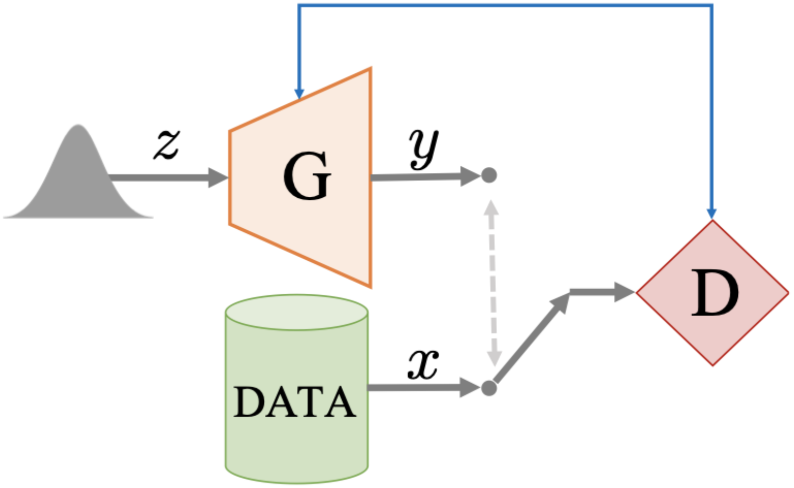
Latent Space ~~Laplacian Pyramid~~ GAN

First demonstration of latent GANs applied to point sets
[Achlioptas et al., ICML 2018]

Skoltech
Skolkovo Institute of Science and Technology

Latent Space ~~Laplacian Pyramid~~ GAN

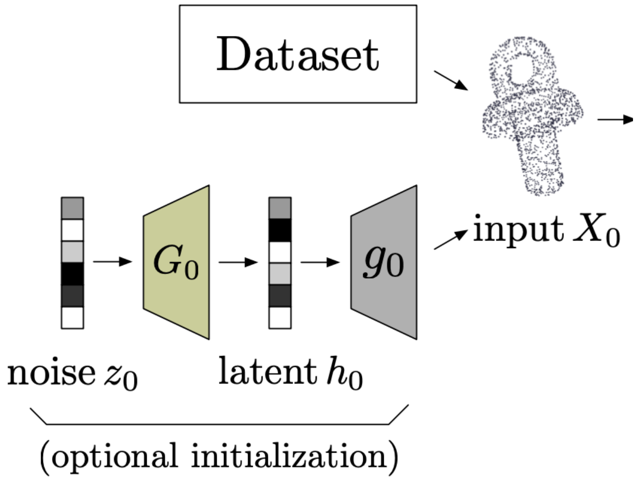




Auto-encoding NNs

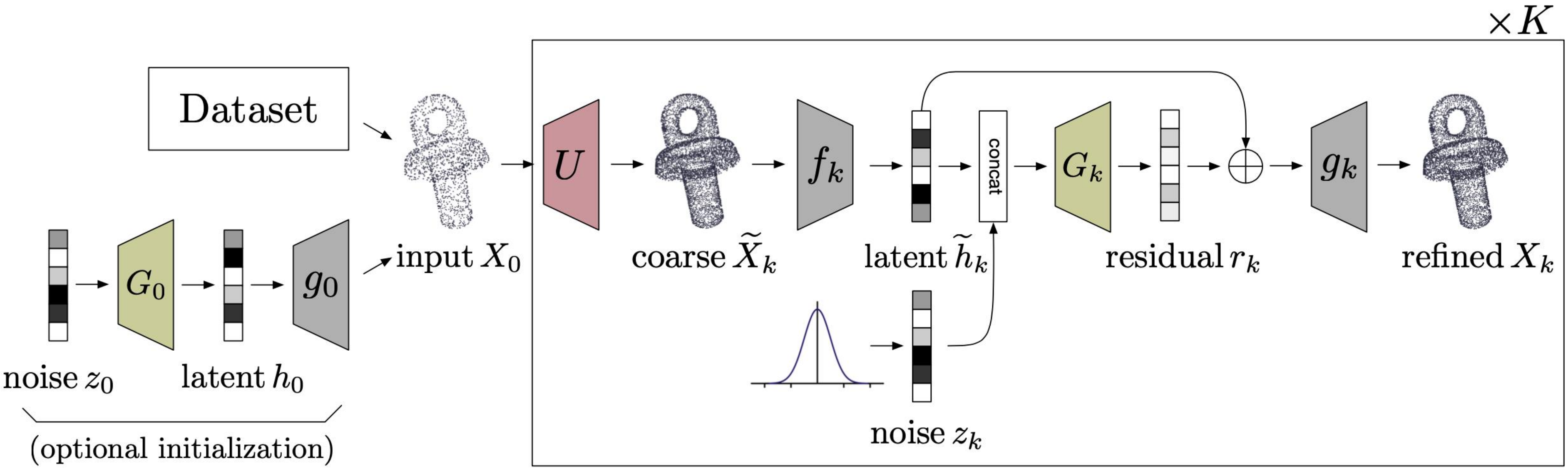
Generative Adversarial NNs

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



- The network either accepts or generates an initial point cloud X_0
  - Input shape: upsampling mode
  - No input shape: shape synthesis
- 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 *fk*

(3) performs correction of the latent code via a conditional GAN *Gk*

(4) decodes the corrected latent code using *gk*

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

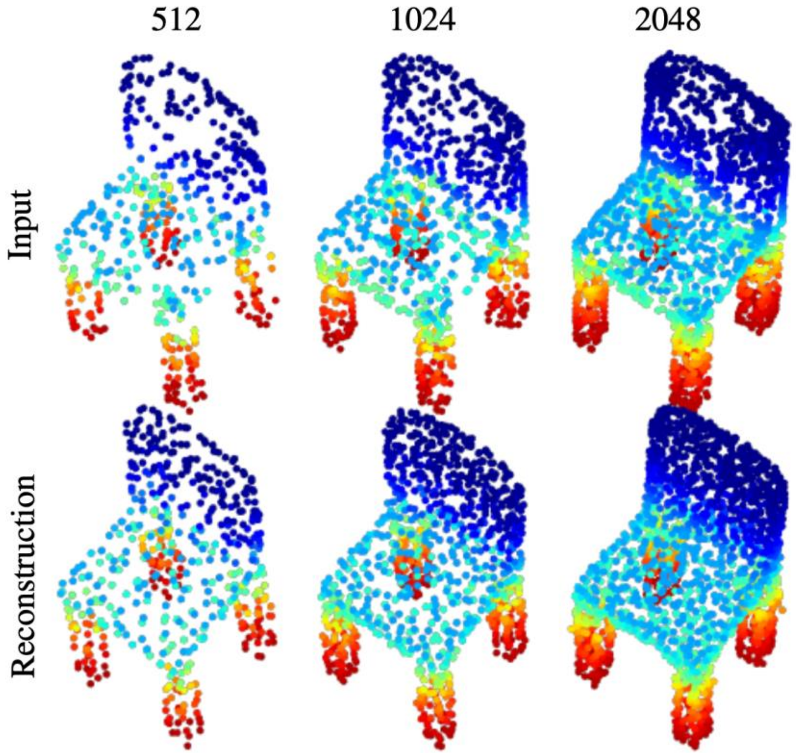
- Modeling 3D point clouds is a challenge due to high dimensionality
- We start with a low-detail X_0
- We decompose the task into a sequense of easier stages
- Each stage aims at a gradual increase of detail

Training auto-encoding point networks on multiple scales

- **Key idea:** learning in the manifold of latent codes
- We use 3D shape spaces $\{\mathbb{R}^{n_k \times 3}\}_{k=1}^{K}$ and construct a series of latent spaces $\{\mathbb{R}^{d_k}\}_{k=1}^{K}$
- We train *K* point autoencoders $\{(f_k, g_k)\}_{k=1}^{K}$ for point clouds of increasing resolution $n_k = 2^k \cdot n_0$

Results: trained autoencoders

- Inputs and reconstructions using our autoencoders at resolutions $n_i \in \{512, 1024, 2048\}$ of the 3D point cloud
- Scales with the increase in the resolution

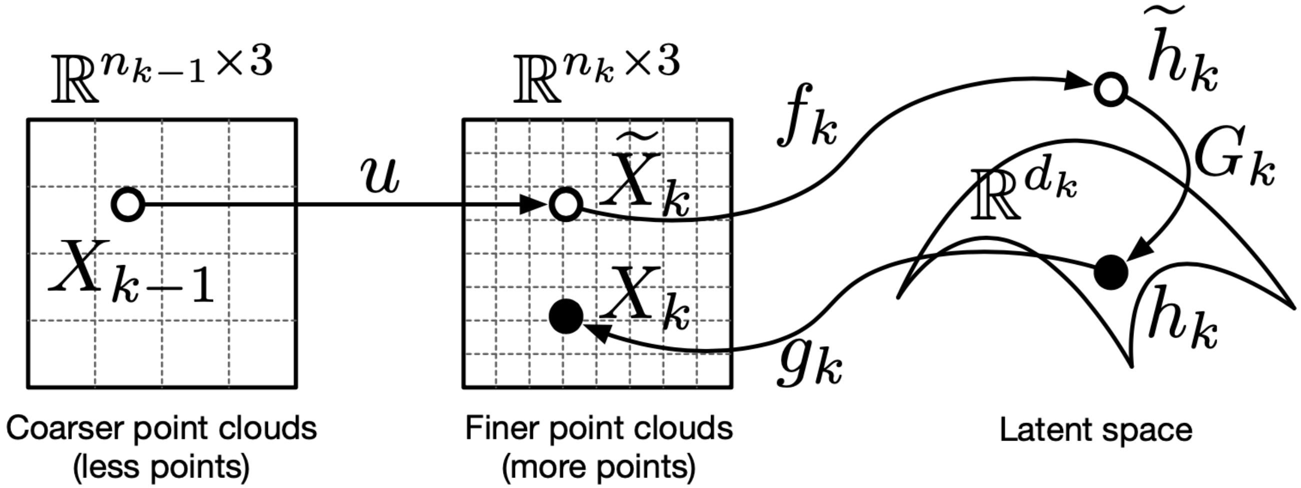


Laplacian pyramid in the space of latent codes

- Input: point cloud $X_{k-1} \in \mathbb{R}^{n_{k-1} \times 3}$
- We aim to go from X_{k-1} to X_k, $n_k = 2n_{k-1}$
- Coarse point cloud: $\widetilde{X}_k = U(X_{k-1})$, i.e.

  for each point x from X_k we generate $\widetilde{x} = \frac{1}{m} \sum_{i \in \mathrm{NN}(x)} x_i$

Laplacian pyramid in the space of latent codes



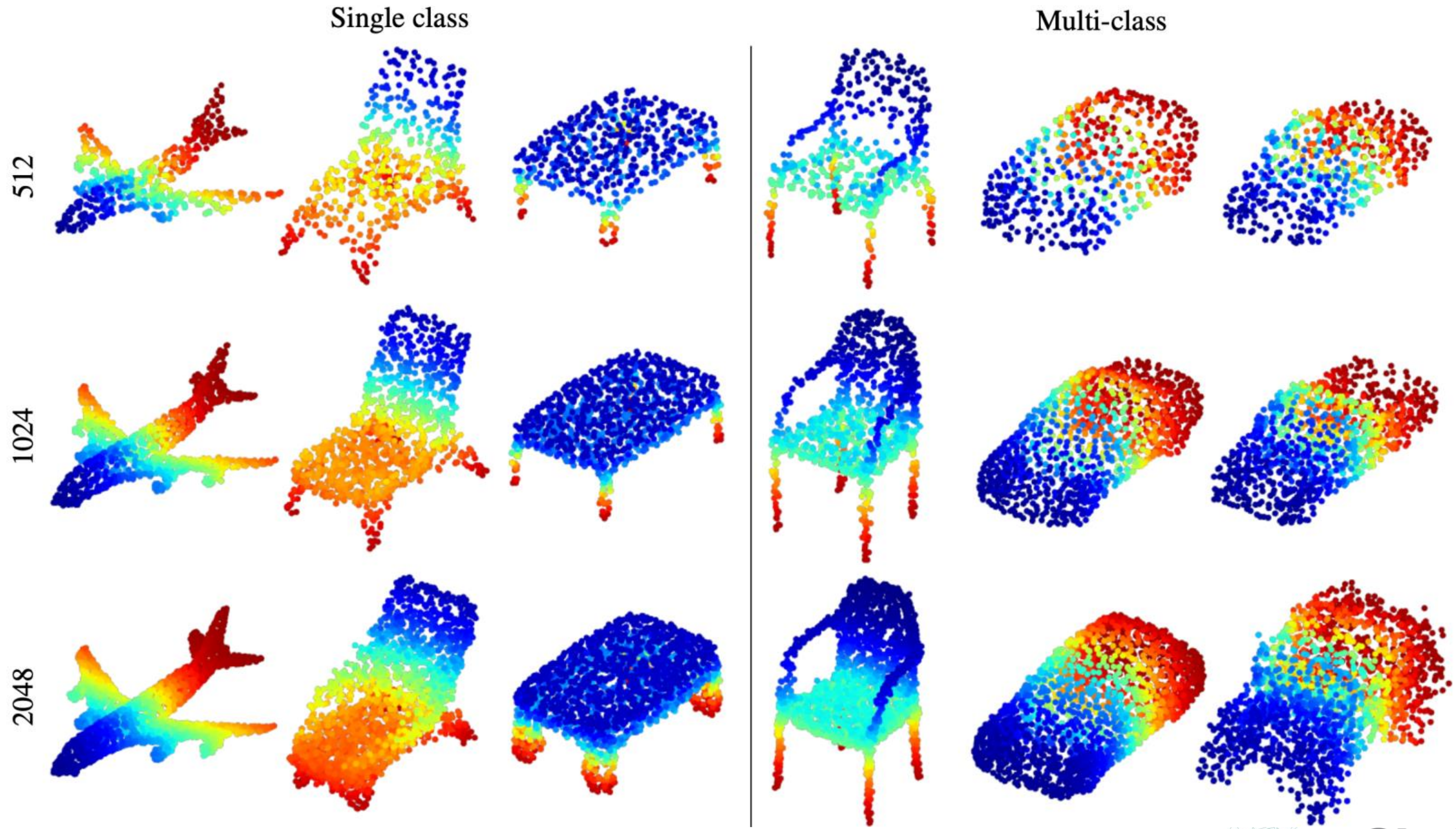
Latent code $\widetilde{h}_k = f_k(\widetilde{X}_k)$

Corrected latent code $h_k = \widetilde{h}_k + r_k = \widetilde{h}_k + G_k(\widetilde{h}_k, z_k)$

Refined point cloud $X_k = g_k(h_k)$

Skoltech
Skolkovo Institute of Science and Technology

Qualitative results: novel shape synthesis



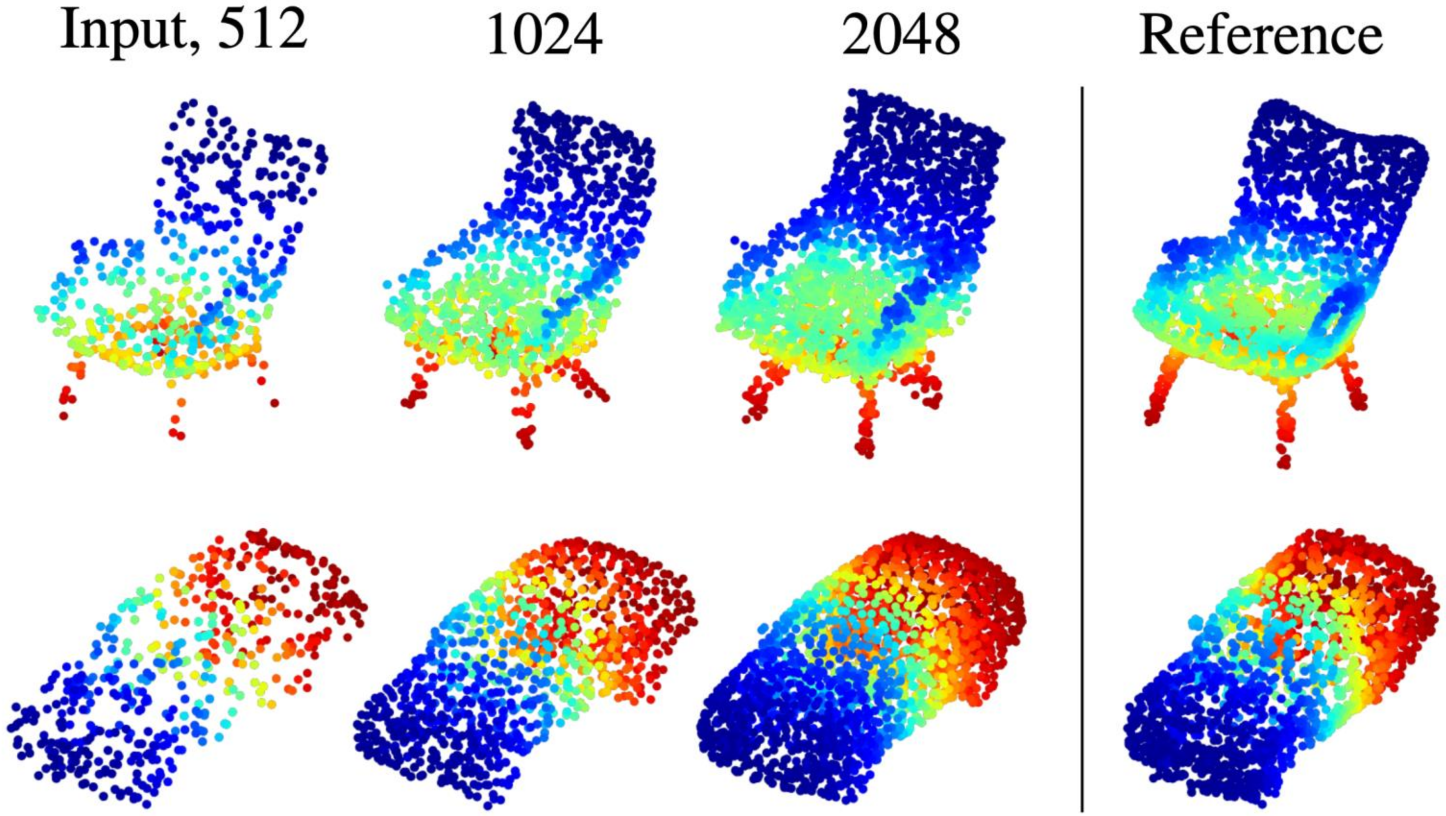
Examples of novel shapes synthesised using our LSLP-GAN model

Quantitative results

Shape class	MMD-CD $\times 10^{-3}$		COV-CD, %		JSD $\times 10^{-3}$	
	L-GAN	Ours	L-GAN	Ours	L-GAN	Ours
car	0.81	**0.71**	23.5	**32.1**	28.9	**24.2**
chair	1.79	**1.71**	44.9	**47.8**	13.0	**10.1**
sofa	1.26	**1.23**	43.9	**46.3**	9.6	**9.3**
table	1.93	**1.77**	39.7	**47.8**	19.9	**10.1**
airplane	0.53	**0.51**	41.7	**44.0**	17.1	**13.8**
multiclass	1.66	**1.55**	41.4	**45.7**	14.3	**9.8**

Table 2: Performance evaluation of our proposed LSLP-GAN model as compared to the baseline L-GAN model (Achlioptas et al., 2018).

Qualitative results: upsampling

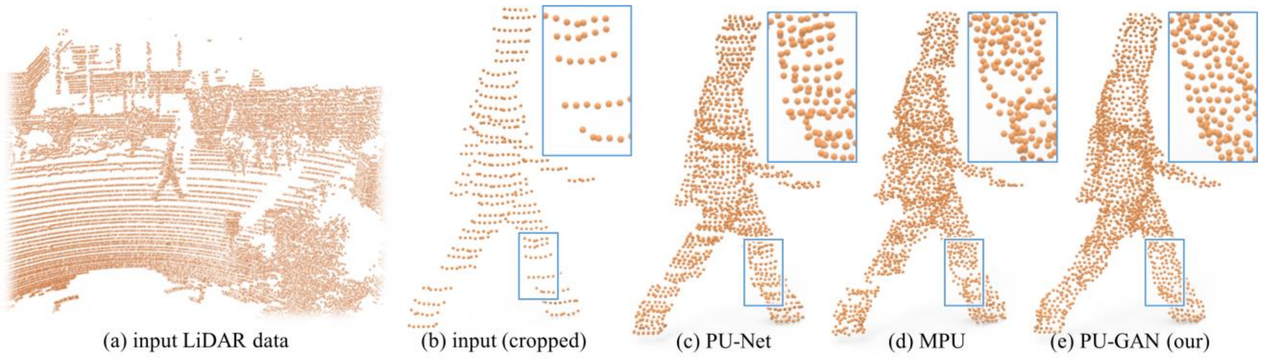


3D point clouds upsampling results using our model
$n_i \in \{512, 1024, 2048\}$ of the 3D point cloud

Summary

- Generative model that produces synthetic shapes
- High-resolution point clouds via a latent space Laplacian pyramid
- Operates in the space of latent codes, i.e. does not require dimensionality reduction during training

What’s next?

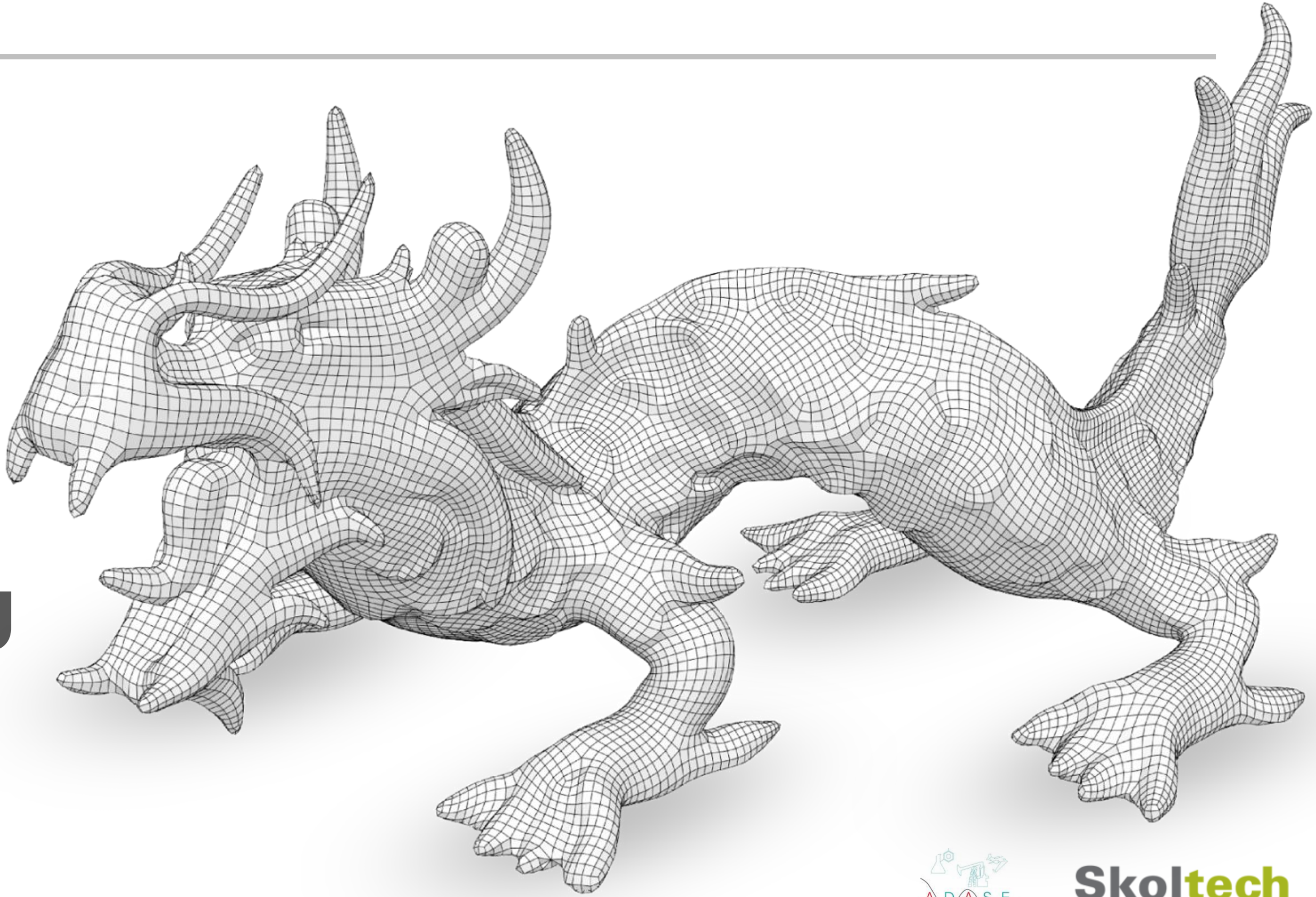


More applications: upsampling (Li et al., 2018)

3D representations: meshes (Cheng et al., 2019)

Alternative formulations: normalizing flows (Yang et al., 2019)

THANK YOU



Skoltech
Skolkovo Institute of Science and Technology