3D Perception from Partial Observations

Ziwei Liu

Nanyang Technological University

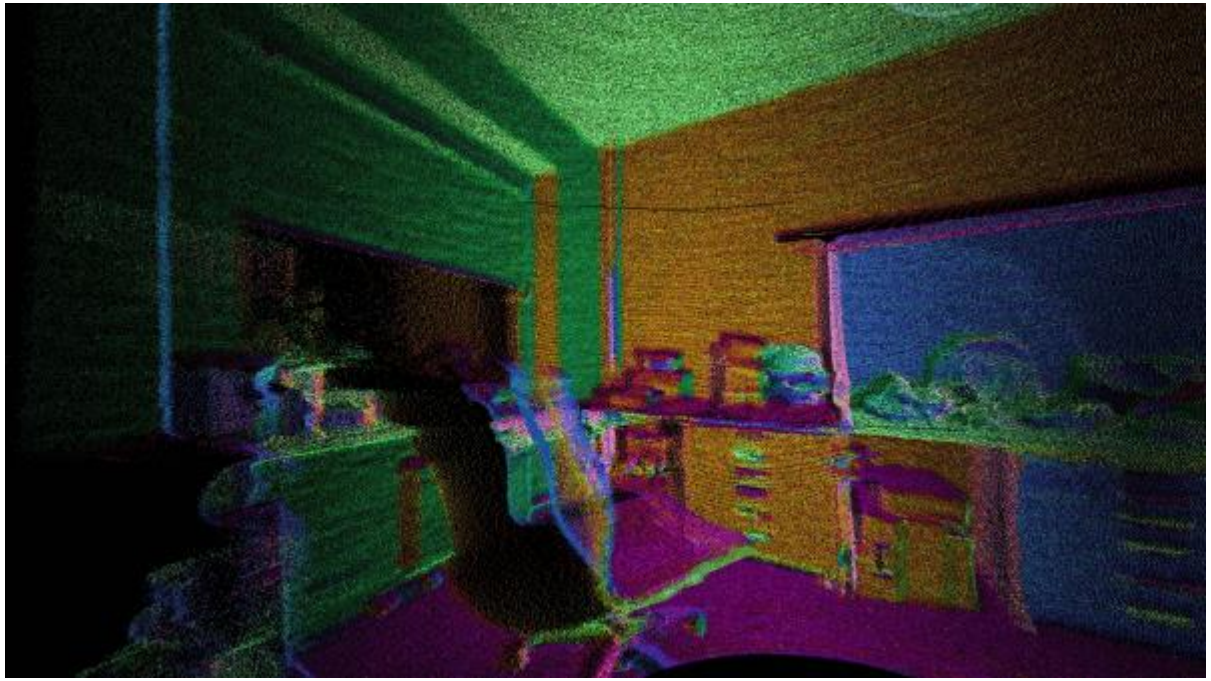
NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE | S-LAB FOR ADVANCED INTELLIGENCE

Partial Observations

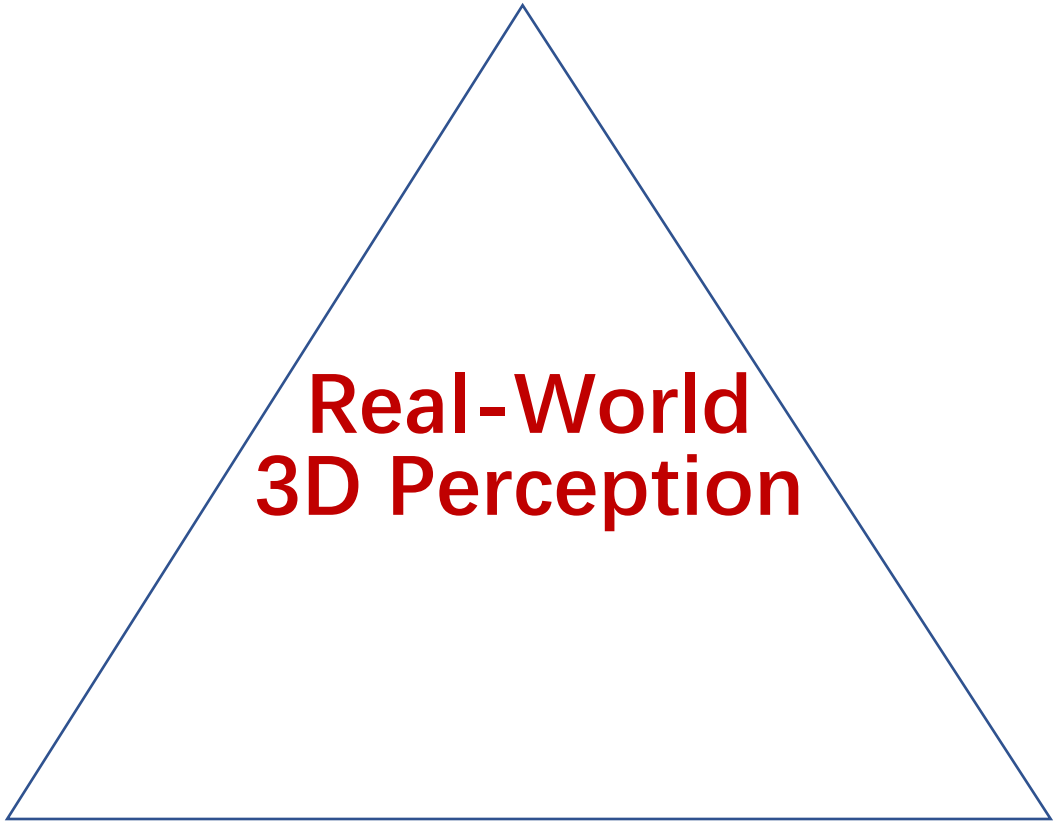
Partial Observations

3D Imagination

Modern Sensor



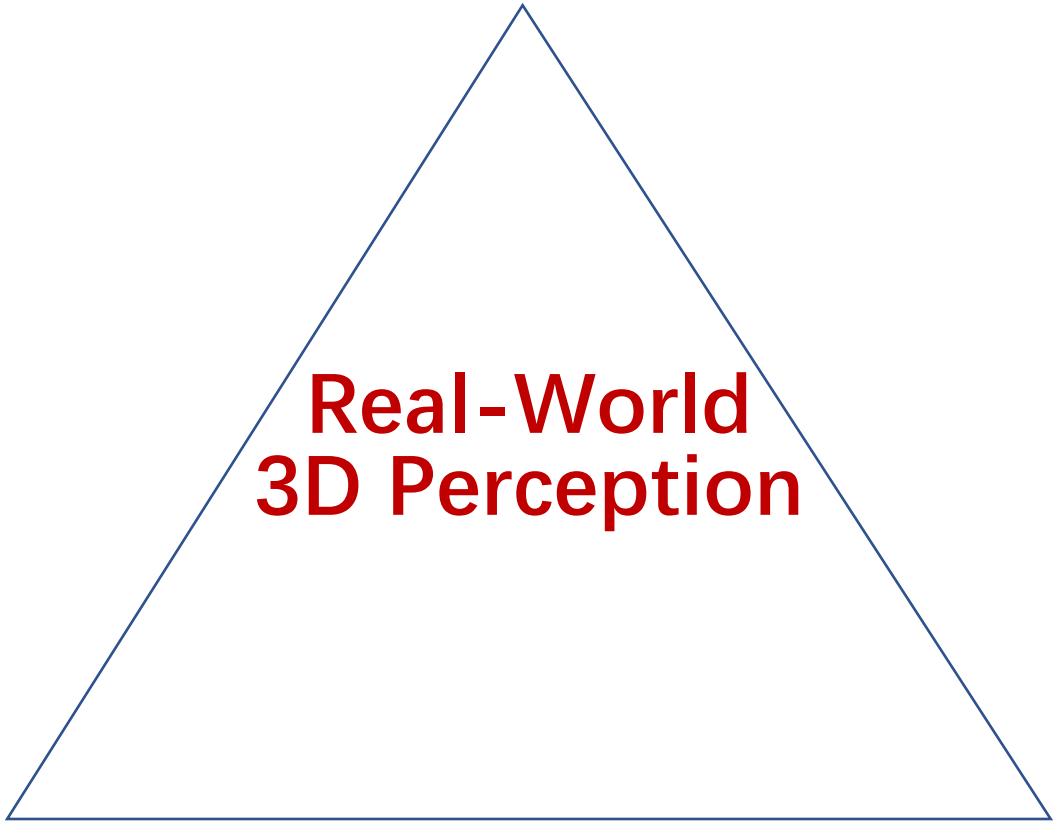
Partial Observations



3D Imagination

Modern Sensor

Partial Observations



3D Imagination

RGB Sensor

Do 2D GANs Know 3D Shape? Unsupervised 3D Shape Reconstruction from 2D Image GANs

Xingang Pan[1], Bo Dai[1], Ziwei Liu[2], Chen Change Loy[2], Ping Luo[3]

[1]The Chinese University of Hong Kong

[2]S-Lab, Nanyang Technological University [3]The University of Hong Kong

ICLR 2021

Do 2D GANs Model 3D Geometry?

Natural images are projections of 3D objects on a 2D image plane.

An ideal 2D image manifold (e.g., GAN) should capture 3D geometric properties.

The following example shows that there is a direction in the GAN image manifold that corresponds to viewpoint variation.

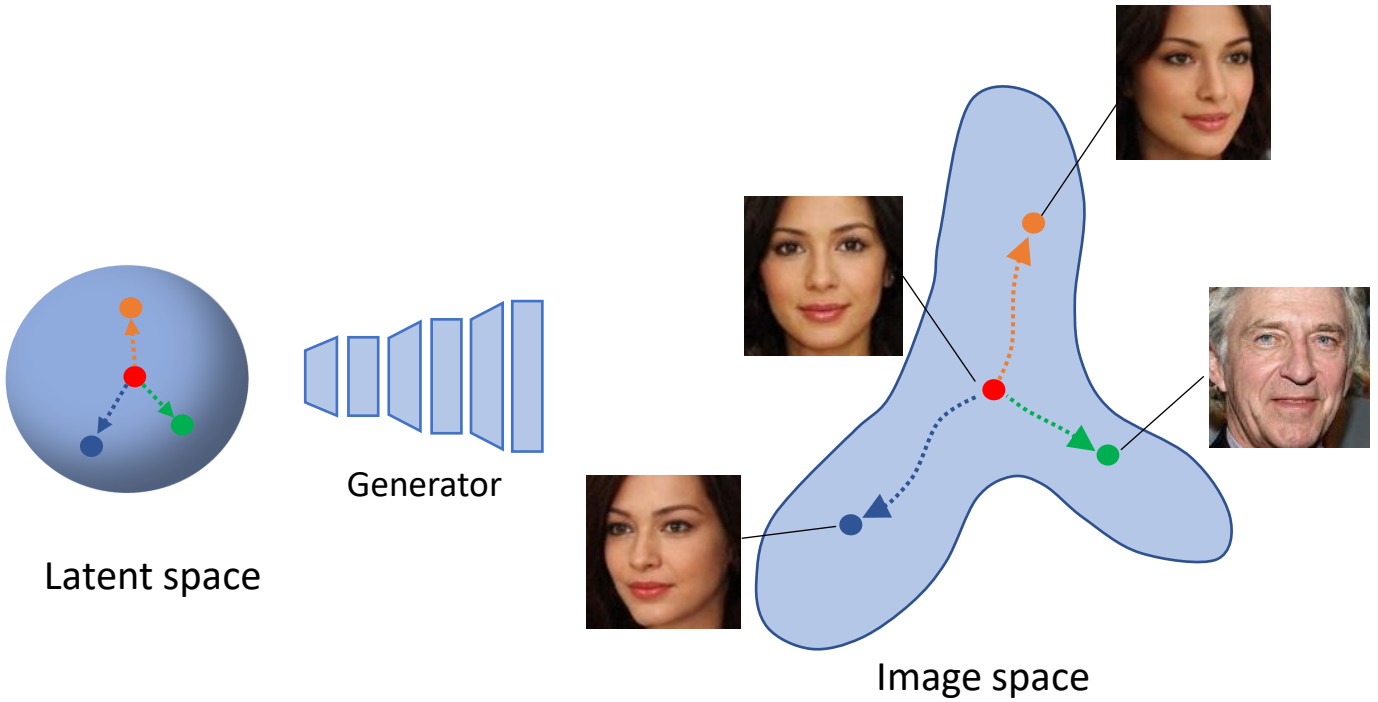
Can we Make Use of such Variations?

Can we make use of such variations for 3D reconstruction?

If we have multiple viewpoint and lighting variations of the same instance, we can infer its 3D structure.

Let's create these variations by exploiting the image manifold captured by 2D GANs!

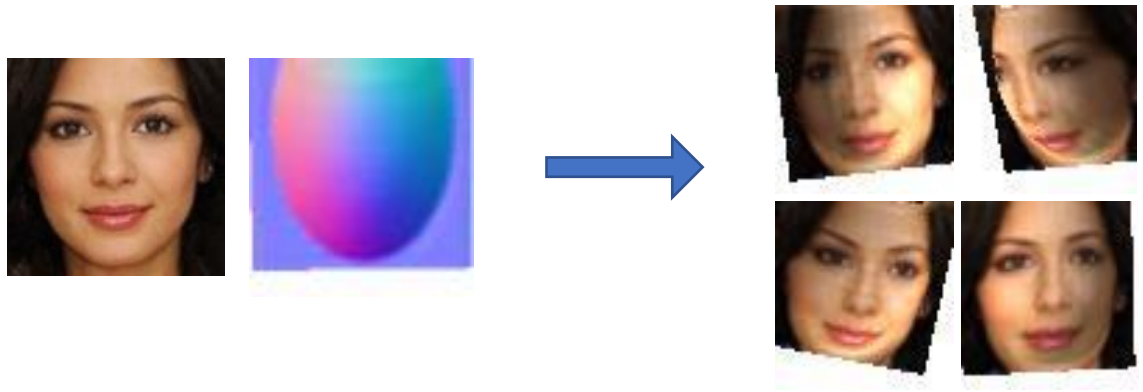
Challenge



It is non-trivial to find well-disentangled latent directions that control *viewpoint* and *lighting* variations in an unsupervised manner.

Our Solution

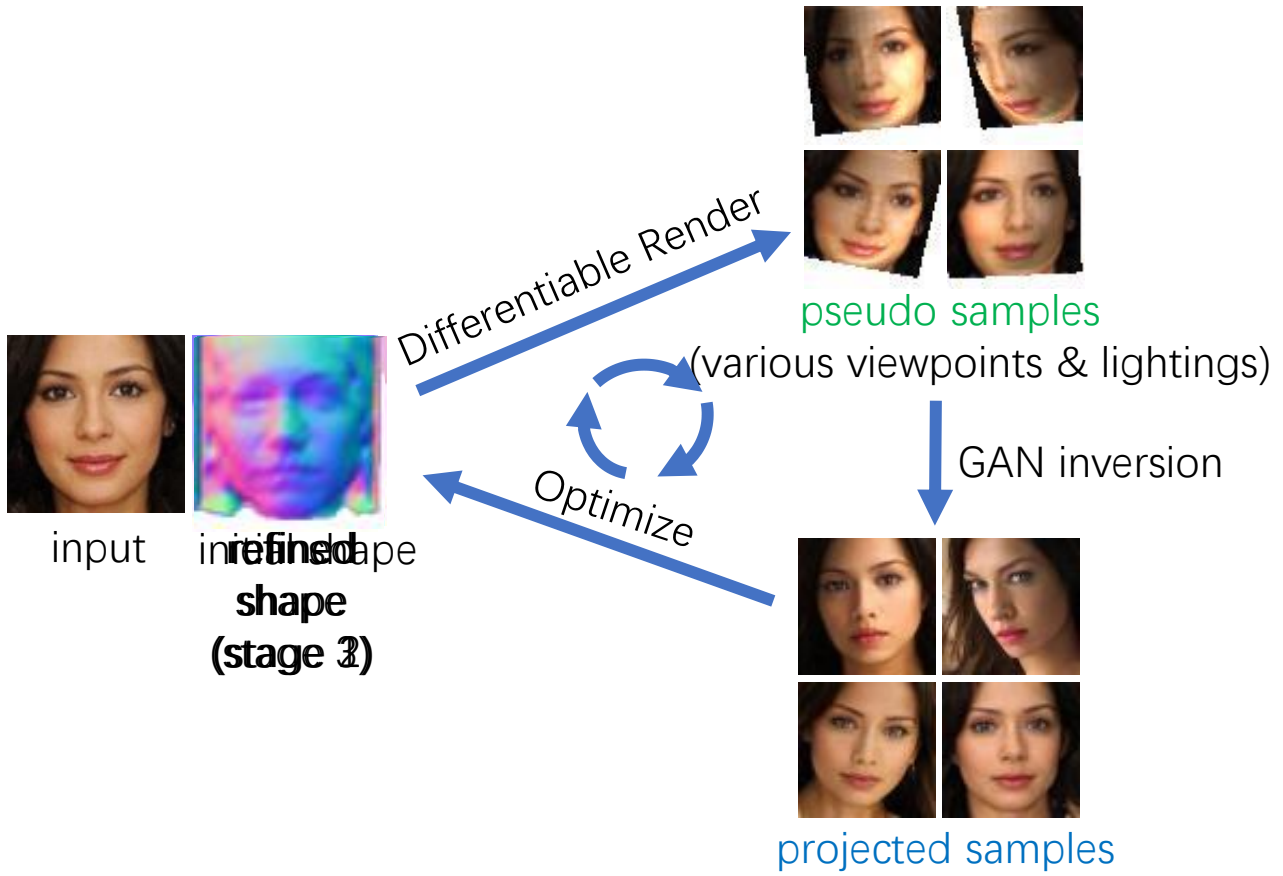
Idea 1: For many objects such as faces and cars, a convex shape prior like ellipsoid could provide a hint on the change of their viewpoints and lighting conditions.



Idea 2: Use GAN inversion constrained by this prior to “find” the latent directions.

Steps

- Initialize the shape with ellipsoid.
- Render '*pseudo samples*' with different viewpoints and lighting conditions.
- GAN inversion is applied to these samples to obtain the '*projected samples*'.
- *'Projected samples'* are used as the ground truth of the rendering process to optimize the 3D shape.
- Iterative training to progressively refine the shape.

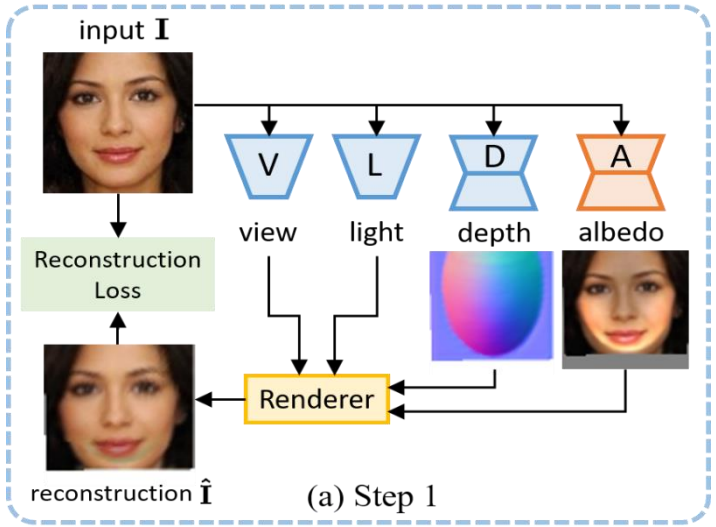


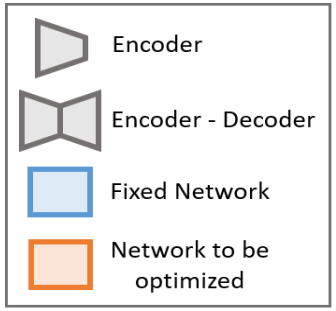
GAN2Shape

Step1:

Initialize shape with ellipsoid.

Optimize albedo network *A*.





$$\boldsymbol{\theta}_A = \arg\min_{\boldsymbol{\theta}_A} \mathcal{L}\Big(\mathbf{I}, \Phi\big(D(\mathbf{I}), A(\mathbf{I}), V(\mathbf{I}), L(\mathbf{I})\big)\Big)$$

$\mathbf{I}$: input image
D: depth network
A: albedo network
V: viewpoint network
L: lighting network
Φ: differentiable render
$\mathcal{L}$: reconstruction loss (L1+perceptual)

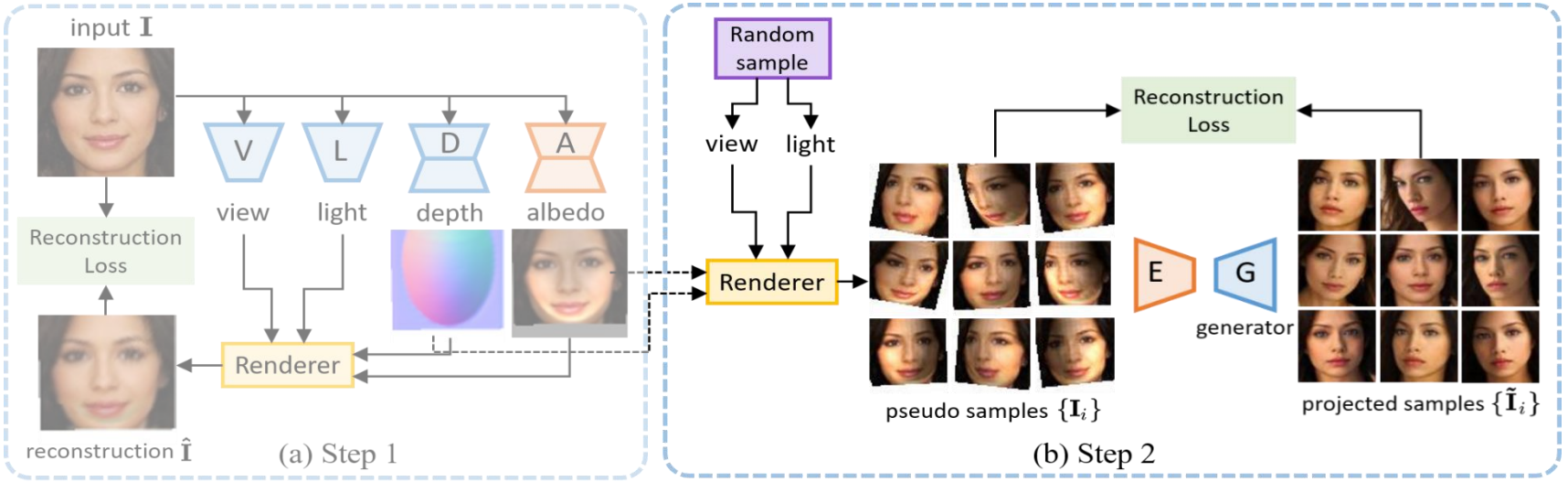
GAN2Shape

Step2:

Render 'pseudo samples' $\{\mathbf{I}_i\}$ with various viewpoints & lightings.

Perform GAN inversion to the pseudo samples to obtain the 'projected samples' $\{\tilde{\mathbf{I}}_i\}$.

Optimize latent encoder E.

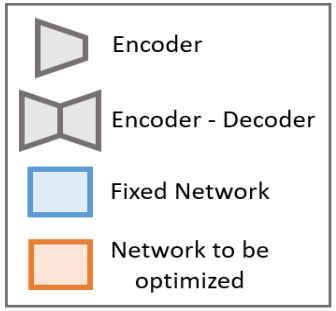


$$\boldsymbol{\theta}_E = \arg\min_{\boldsymbol{\theta}_E} \frac{1}{m}\sum_{i=0}^{m} \mathcal{L}'\left(\mathbf{I}_i, G\left(E(\mathbf{I}_i) + \boldsymbol{w}\right)\right) + \lambda_1 \|E(\mathbf{I}_i)\|_2$$

original latent code

latent offset $\triangle \boldsymbol{w}_i$

L2 regularization

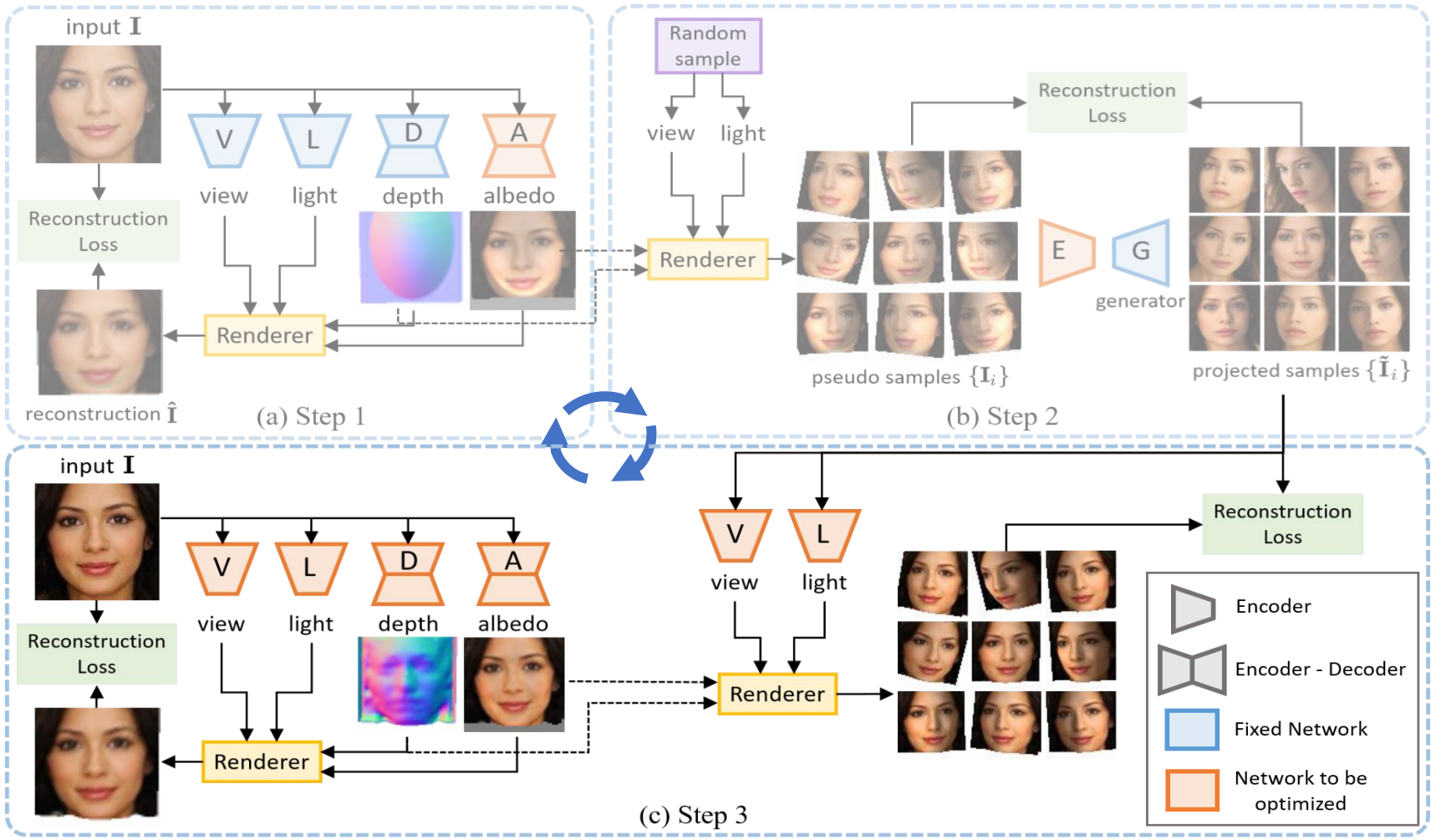


GAN2Shape

Step3:

Reconstruct ‘projected samples’ with shared depth & albedo and independent view & light.

Optimize network *V, L, D, A*.



$$\boldsymbol{\theta}_D, \boldsymbol{\theta}_A, \boldsymbol{\theta}_V, \boldsymbol{\theta}_L = \underset{\boldsymbol{\theta}_D, \boldsymbol{\theta}_A, \boldsymbol{\theta}_V, \boldsymbol{\theta}_L}{\arg\min} \frac{1}{m} \sum_{i=0}^{m} \mathcal{L}\Big(\tilde{\mathbf{I}}_i, \Phi\big(D(\mathbf{I}), A(\mathbf{I}), V(\tilde{\mathbf{I}}_i), L(\tilde{\mathbf{I}}_i)\big)\Big) + \lambda_2 \mathcal{L}_{smooth}\big(D(\mathbf{I})\big)$$

smoothness term

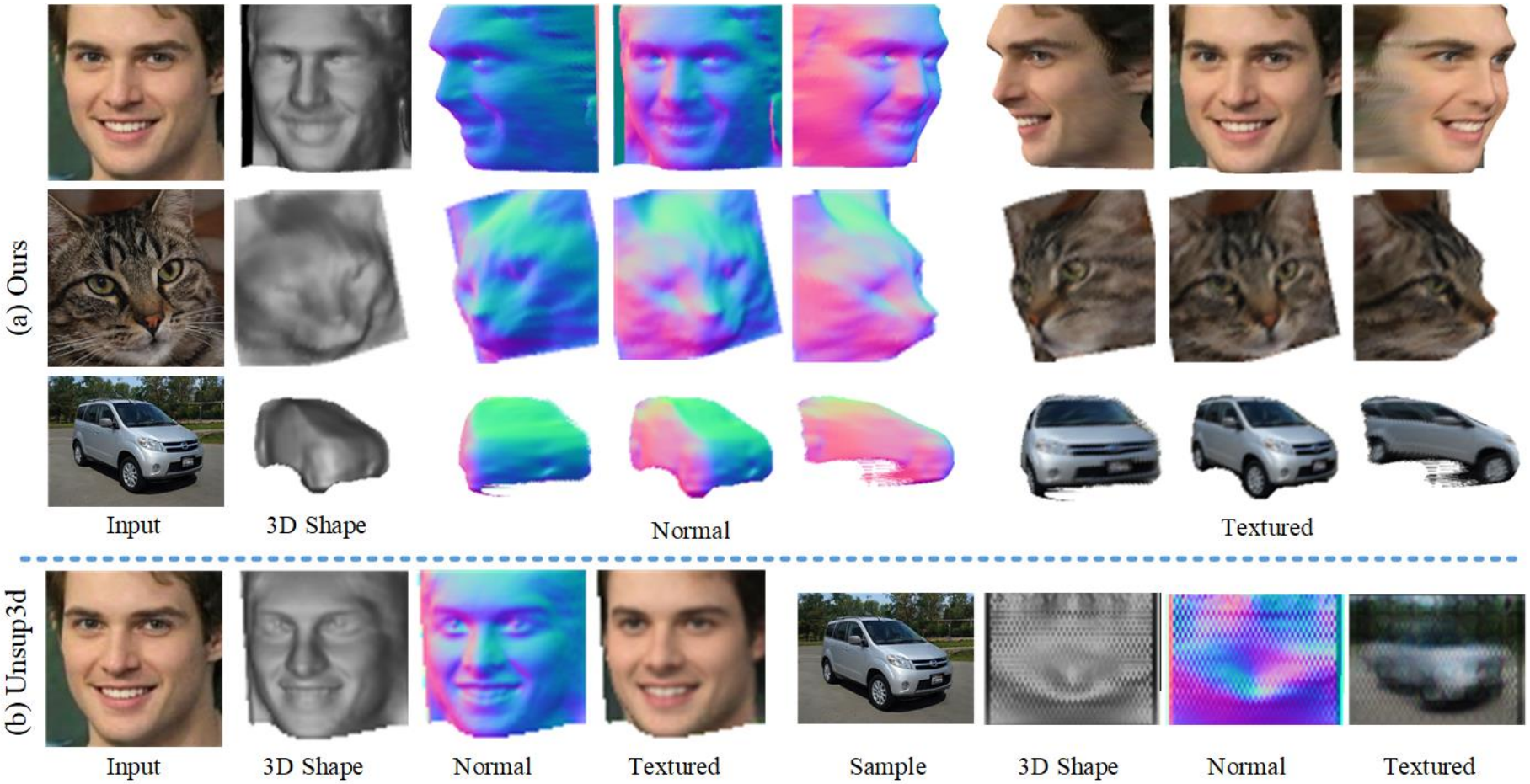
3D Reconstruction Results

Without any 2D keypoint or 3D annotations

Unsupervised 3D shape reconstruction from unconstrained 2D images

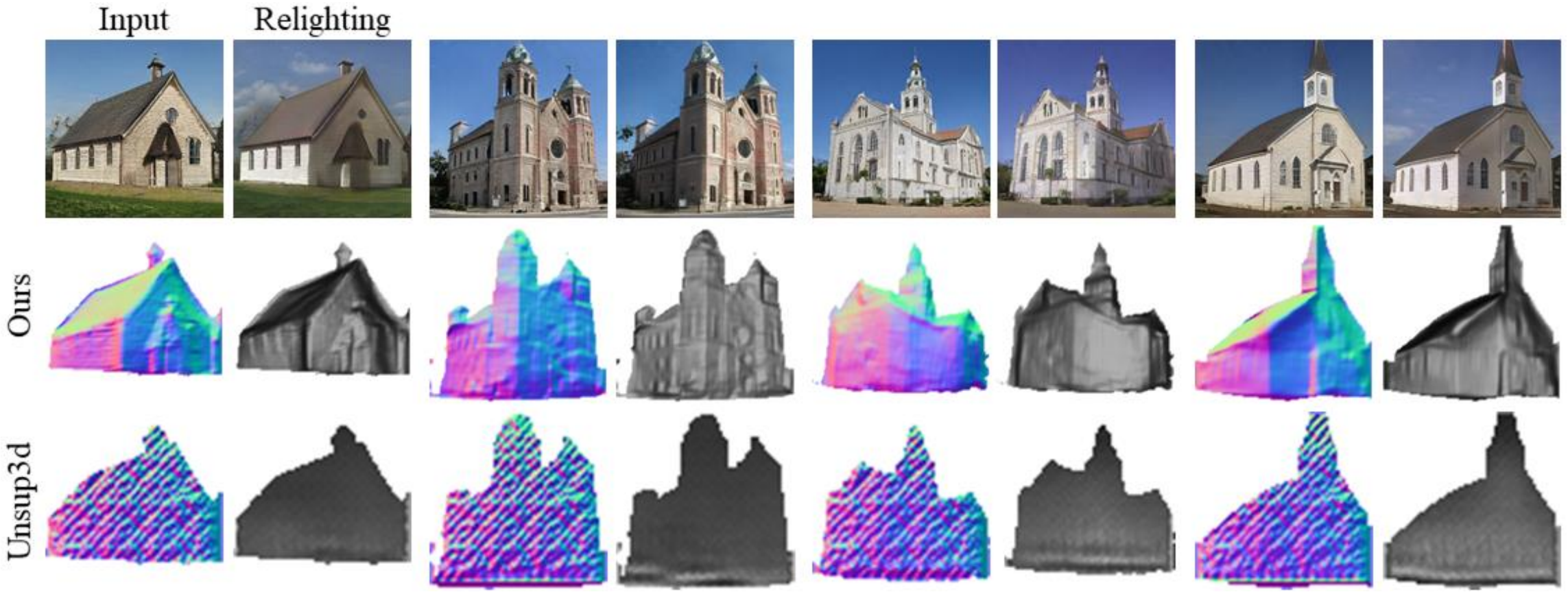
Without symmetry assumption

Work on many object categories such as huma faces, cars, buildings, etc.



Shangzhe Wu, Christian Rupprecht, and Andrea Vedaldi. Unsupervised learning of probably symmetric deformable 3D objects from images in the wild. In *CVPR*, 2020

3D Reconstruction Results



Shangzhe Wu, Christian Rupprecht, and Andrea Vedaldi. Unsupervised learning of probably symmetric deformable 3D objects from images in the wild. In *CVPR*, 2020

3D Reconstruction Results

Table 1: **Comparisons on the BFM dataset.** We report SIDE and MAD errors. 'Symmetry' indicates whether the symmetry assumption on object shape is used. We outperform others on both metrics.

No.	Method	Symmetry	SIDE ($\times 10^{-2}$)↓	MAD (deg.)↓
(1)	Supervised	N	0.419	10.83
(2)	Const. null depth	/	2.723	43.22
(3)	Average g.t. depth	/	1.978	22.99
(4)	Unsup3d (Wu et al., 2020)	Y	0.807	16.34
(5)	Ours (w/o regularize)	Y	0.925	16.42
(6)	Ours	Y	**0.756**	**14.81**
(7)	Unsup3d (Wu et al., 2020)	N	1.334	33.79
(8)	Ours	N	**1.023**	**17.09**

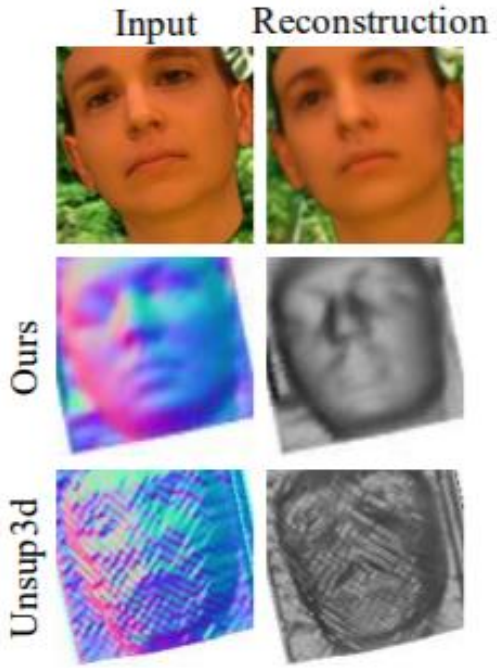
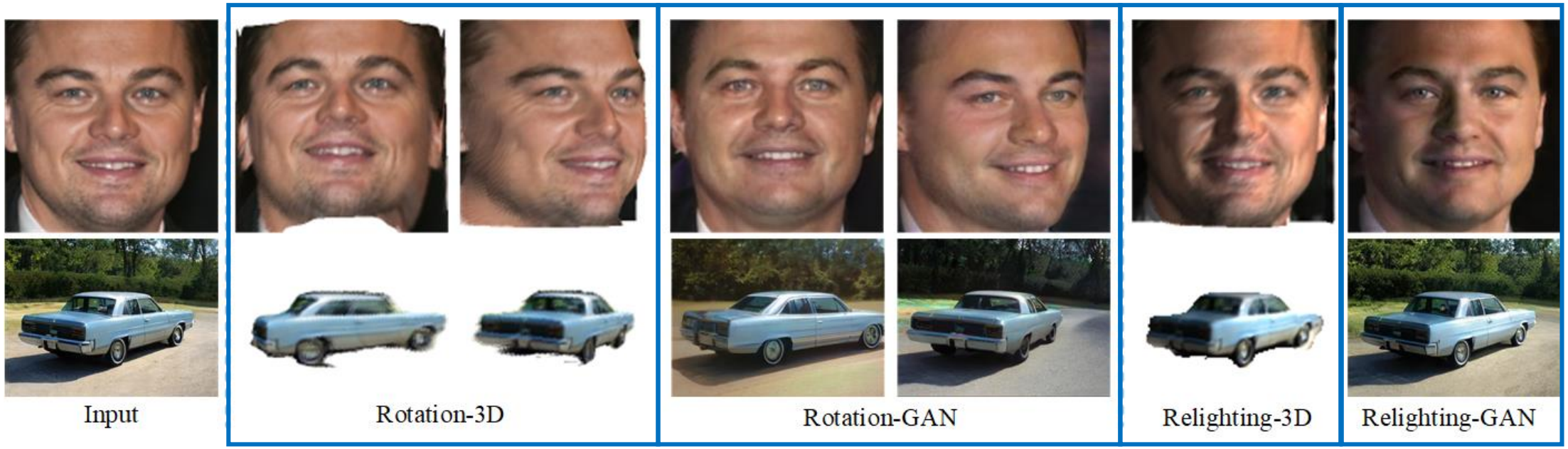


Figure 5: Results without symmetry assumption.

Shangzhe Wu, Christian Rupprecht, and Andrea Vedaldi. Unsupervised learning of probably symmetric deformable 3D objects from images in the wild. In *CVPR*, 2020

3D-aware Image Manipulation

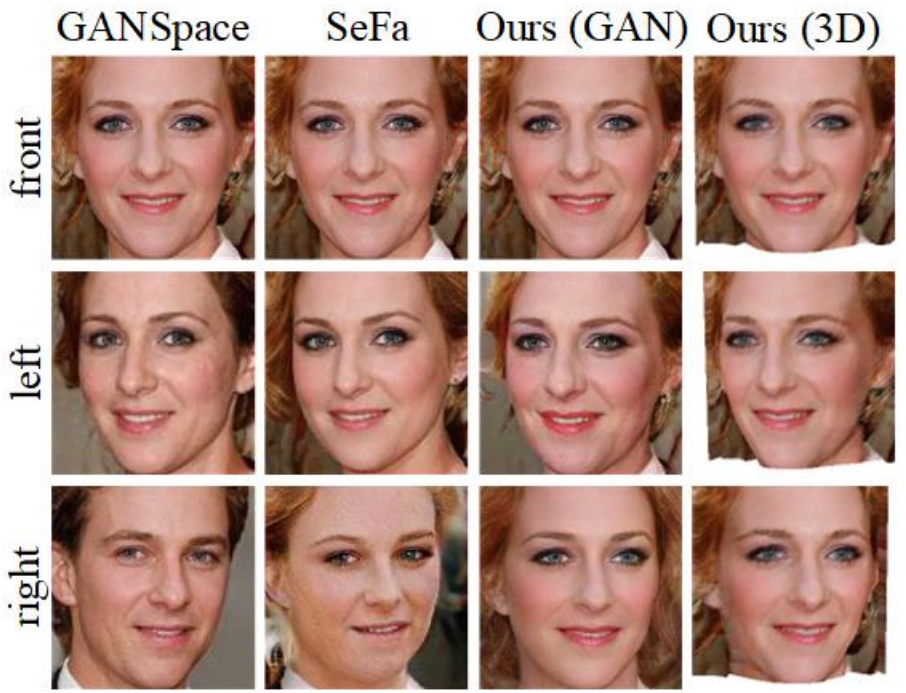


- Effect-3D: Rendered using the reconstructed 3D shape and albedo.
- Effect-GAN: project Effect-3D on the GAN image manifold using the trained encoder *E*.

3D-aware Image Manipulation

Table 2: **Identity-preserving face rotation.** We compare with HoloGAN, GANSpace, and SeFa. The metrics are identity distances measured as angles in the ArcFace feature embeddings.

Method	error_mean (deg.)↓	error_max (deg.)↓
HoloGAN	47.38	69.24
GANSpace	41.17	58.93
SeFa	41.79	60.73
Ours (3D)	**28.93**	**43.02**
Ours (GAN)	39.85	57.21



Nguyen-Phuoc, Thu, et al. "Hologan: Unsupervised learning of 3d representations from natural images." *ICCV* 2019.

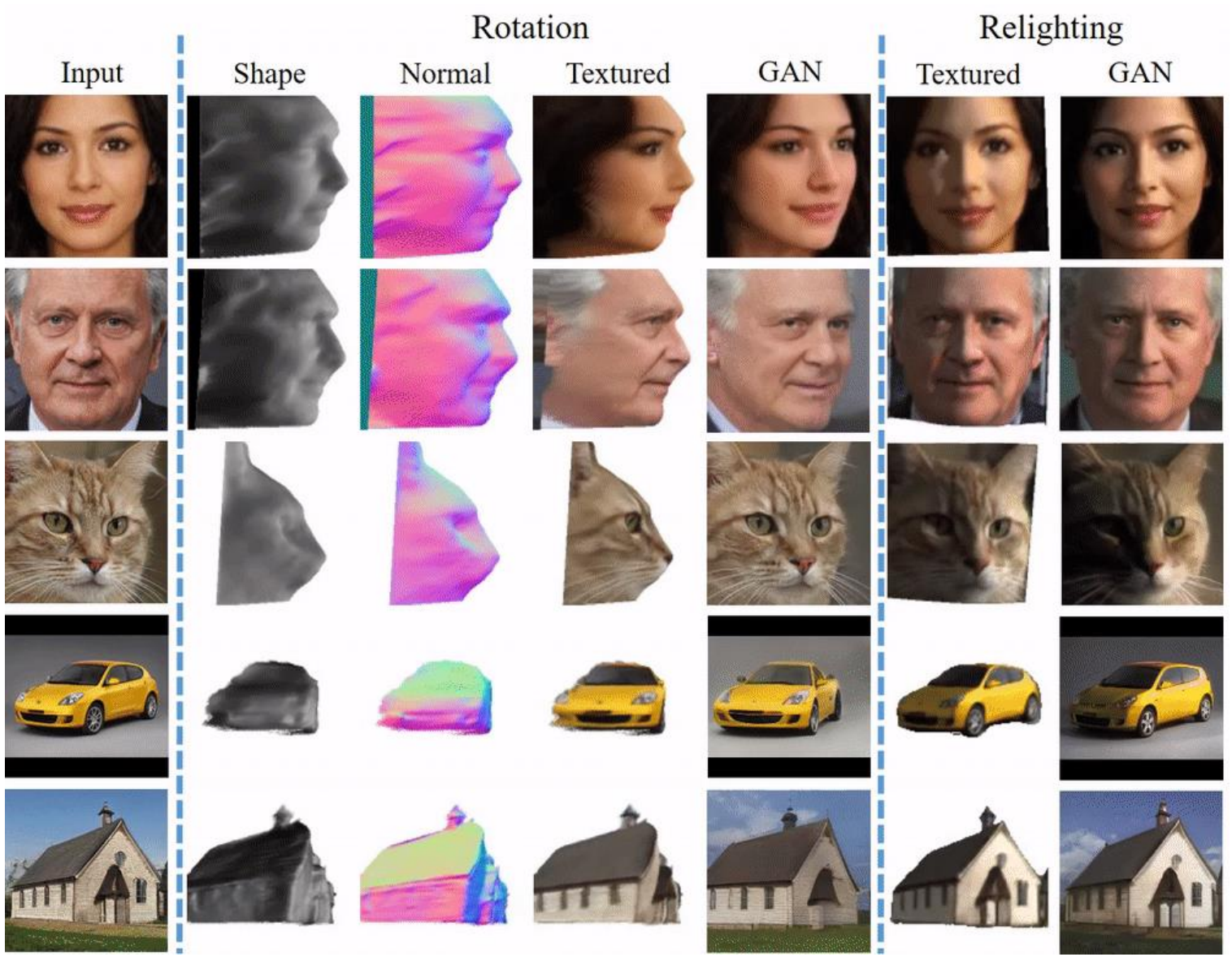
Härkönen, Erik, et al. "GANSpace: Discovering Interpretable GAN Controls." *NIPS* 2020.

Shen, Yujun, and Bolei Zhou. "Closed-form factorization of latent semantics in gans." *CVPR* 2020.

More Results

More Results

Demo


Rotation
Relighting
Input
Shape
Normal
Textured
GAN
Textured
GAN

Summary

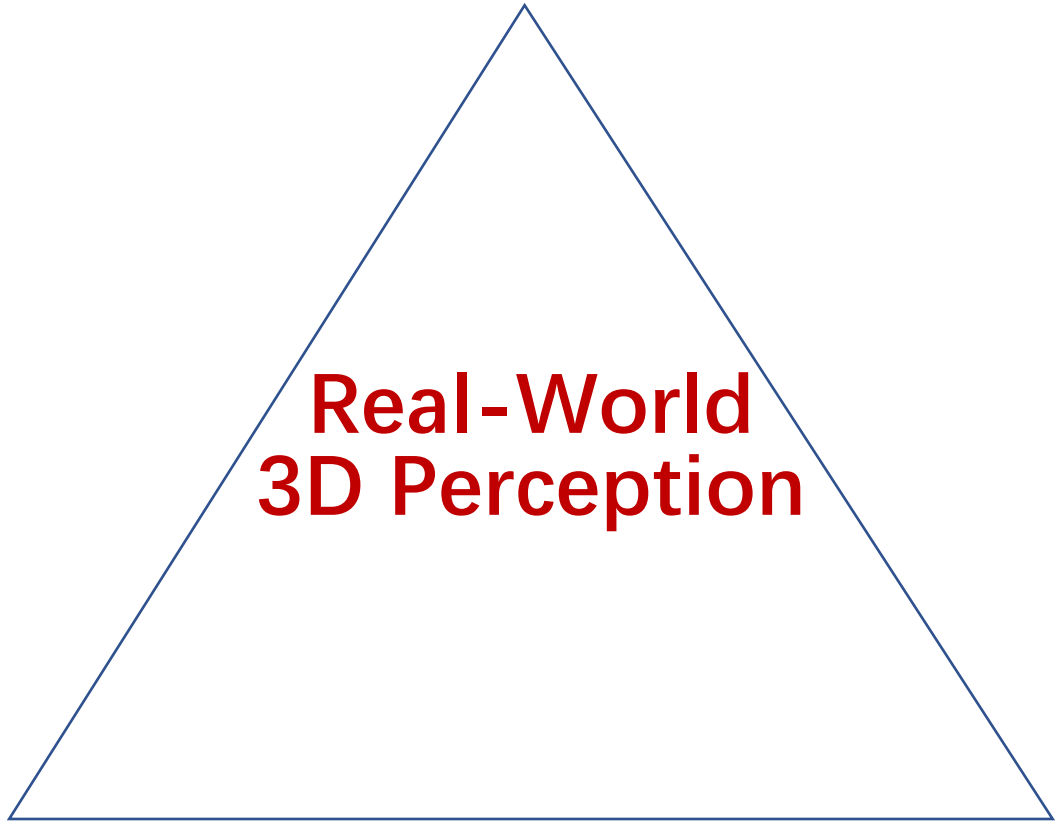
- We demonstrate that ***2D GANs inherently capture the underlying 3D geometry of objects by learning from RGB images***.
- Our method is a powerful approach for unsupervised 3D shape learning from unconstrained 2D images, and ***does not rely on the symmetry assumption***.

  We are doing ***Shape-from-X***, where ***X=GAN***.
- We achieve accurate 3D-aware image manipulation via GANs ***without borrowing external 3D models***.
- Our method provides a new perspective for 3D shape generation.

ICLR

Code on github

Partial Observations



3D Imagination

LiDAR Sensor

Variational Relational Point Completion Network

Liang Pan[1] Xinyi Chen[1,2] Zhongang Cai[2,3] Junzhe Zhang[1,2]

Haiyu Zhao[2,3] Shuai Yi[2,3] Ziwei Liu[1]✉

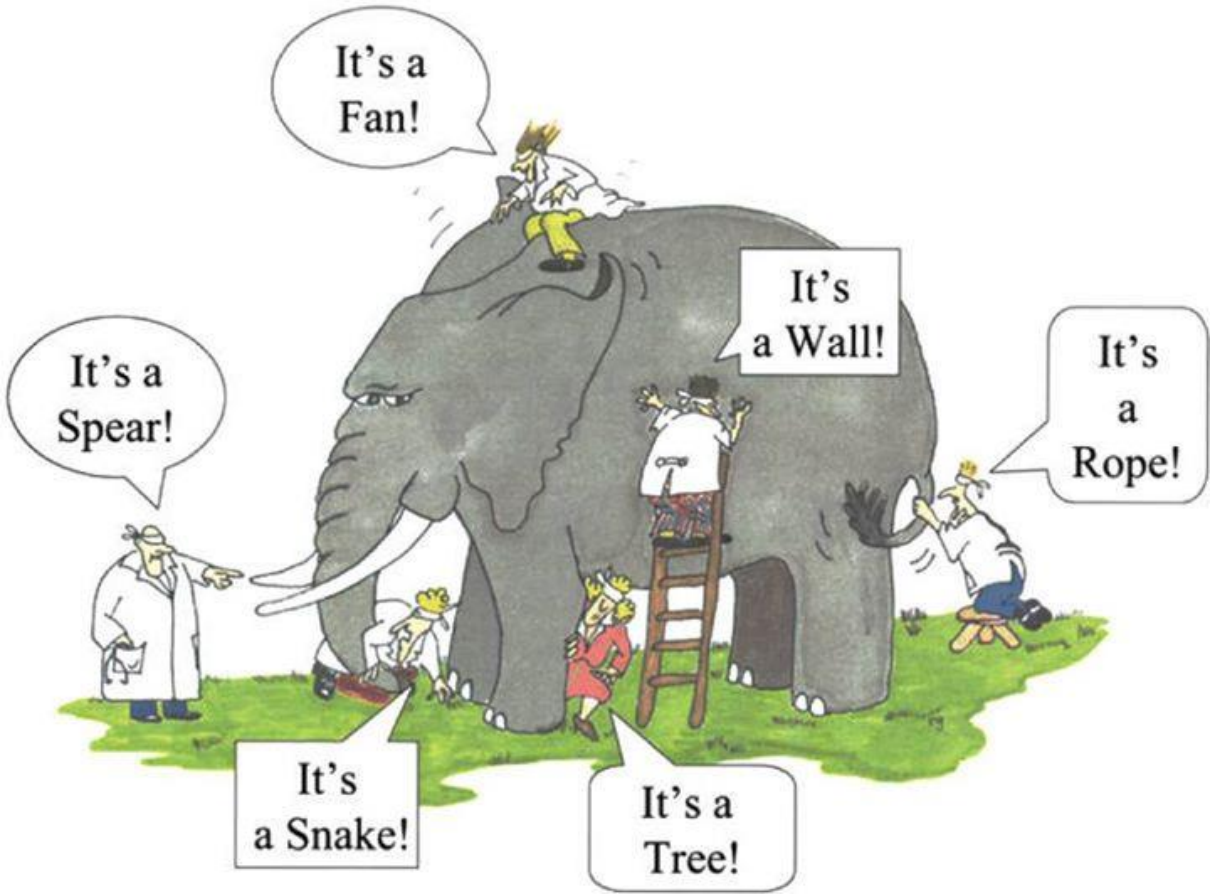
[1]S-Lab, Nanyang Technological University [2]SenseTime Research [3]Shanghai AI Laboratory

CVPR VIRTUAL JUNE 19-25

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE | S-LAB FOR ADVANCED INTELLIGENCE

Background: Essential Question

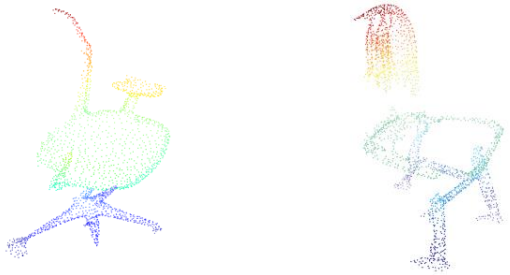
Partial Observations (2.5D)



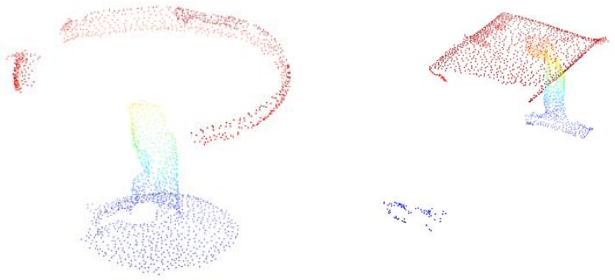
Background: Problem Definition

Incomplete Cars
(Lidar, Kitti)

Incomplete Chair
(RGBD camera, ScanNet)



Incomplete Table
(RGBD camera, ScanNet)



- Real Scans for real-world 3D objects:

  1) sparse; 2) noisy; 3) incomplete

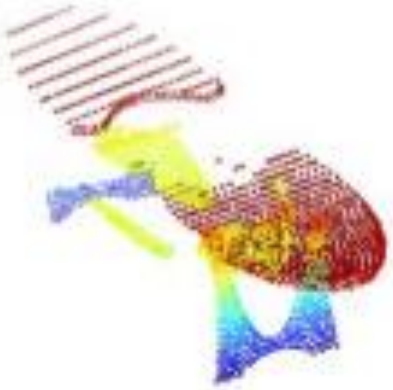
- ✓ 3D point cloud completion:

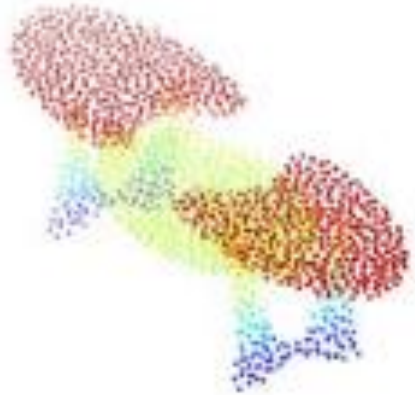
  based on partial shapes (2.5D)

Problem Analysis

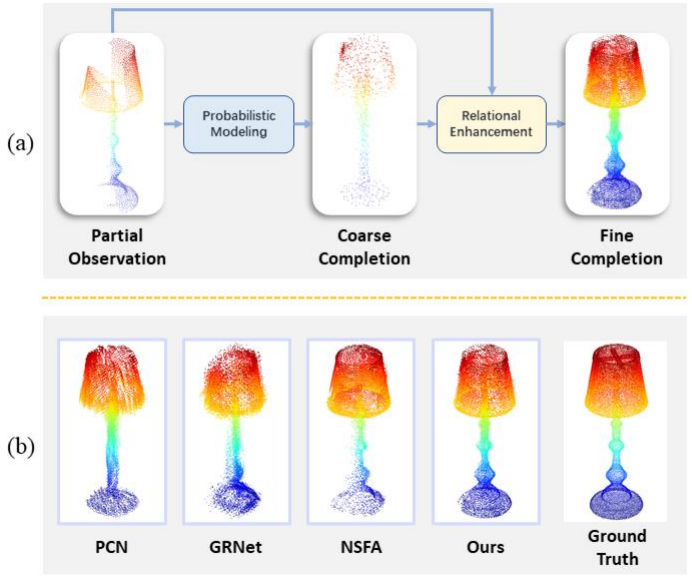
◆ Incomplete Point Cloud **V.S.** Complete Point Cloud **:**

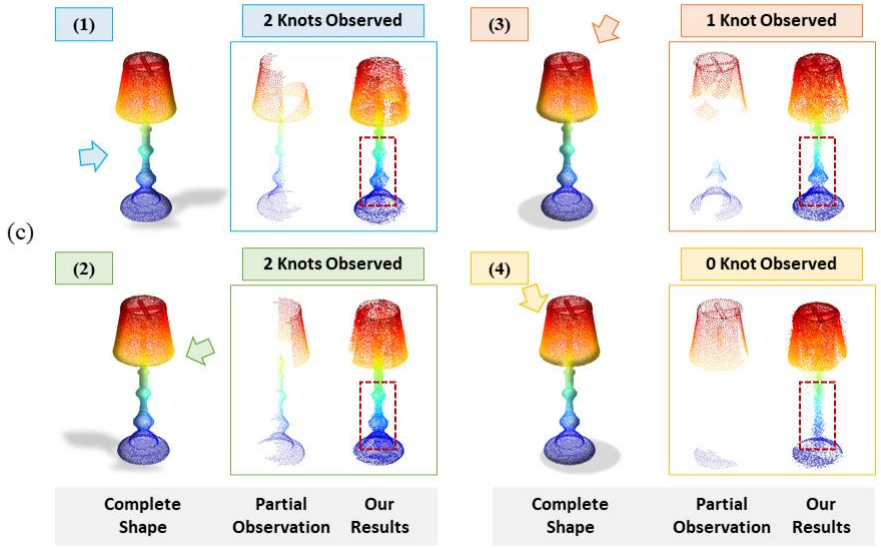
- ☐ Surfaces: unevenly distributed
- ☐ Geometric Details: partially preserve local structures
- ☐ Multi-modal Completions: multiple possible complete point clouds





VRCNet: Overview





3D shape completion is expected to recover **plausible** yet fine-grained complete shapes by learning **relational structure properties**.

(a) Two consecutive stages:

probabilistic modeling (PMNet) and **relational enhancement (RENet)**

(b) Qualitative Results: **better shape details**

(c) Completion **conditioned** on partial observations

Our Contributions

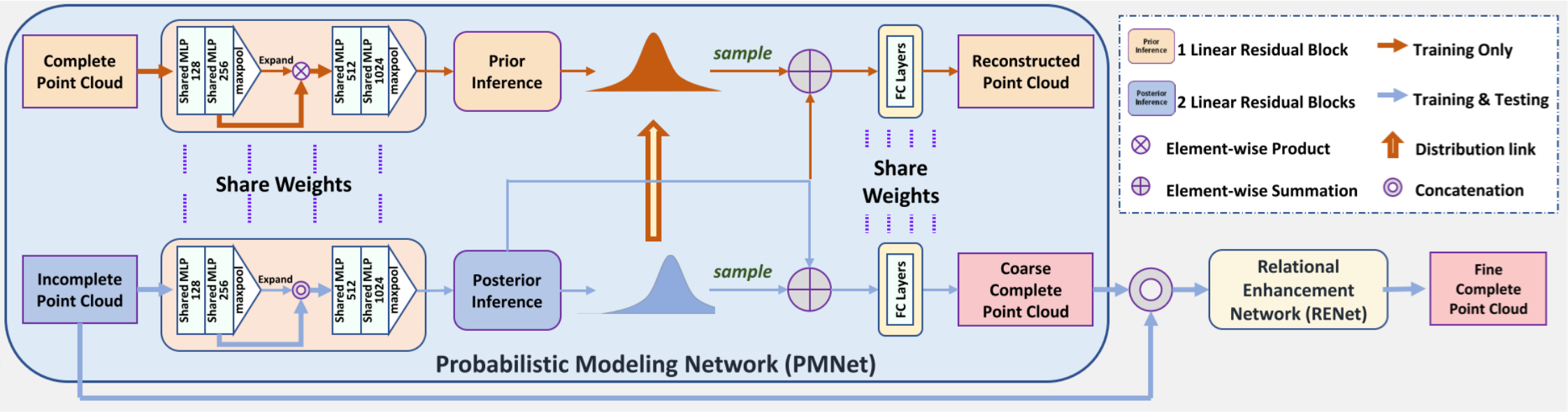
1. We propose a novel Variational Relational point Completion Network (VRCNet), and it first performs probabilistic modeling using a novel dual-path network followed by a relational enhancement network.

2. We design multiple relational modules that can effectively exploit and fuse multiscale point features for point cloud analysis, such as the Point Self-Attention Kernel and the Point Selective Kernel Module.

3. Furthermore, we contribute a large-scale multi-view partial point cloud (MVP) dataset with over 100,000 high-quality 3D point shapes.

Extensive experiments show that VRCNet outperforms previous SOTA methods on all evaluated benchmark datasets.

Probabilistic Modeling Network (PMNet)



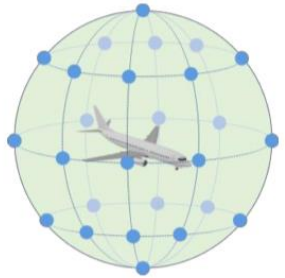
Two parallel paths:

1) the upper reconstruction path (orange line);
2) the lower completion path (blue line).

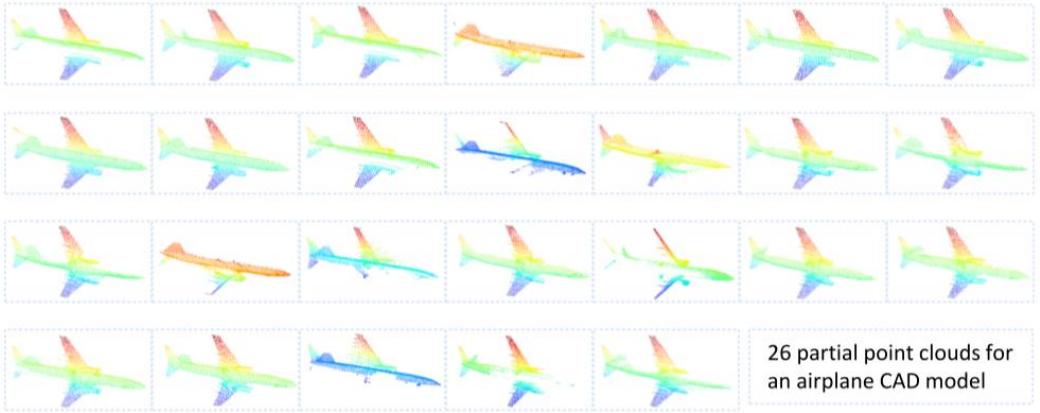
$$\mathcal{L}_{rec} = -\lambda \mathbf{KL}\left[q_\phi(\mathbf{z_g}|\mathbf{Y}) \,\|\, p(\mathbf{z_g})\right] + \mathbb{E}_{p_{data}(\mathbf{Y})}\mathbb{E}_{q_\phi(\mathbf{z_g}|\mathbf{Y})}\left[\log p_\theta^r(\mathbf{Y}|\mathbf{z_g})\right]$$

$$\mathcal{L}_{com} = -\lambda \mathbf{KL}\left[q_\phi(\mathbf{z_g}|\mathbf{Y}) \,\|\, p_\psi(\mathbf{z_g}|\mathbf{X})\right] + \mathbb{E}_{p_{data}(\mathbf{X})}\mathbb{E}_{p_\psi(\mathbf{z_g}|\mathbf{X})}\left[\log p_\theta^c(\mathbf{Y}|\mathbf{z_g})\right]$$

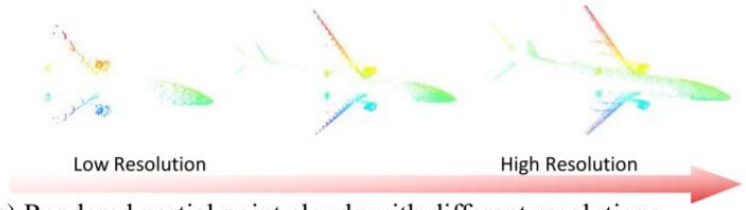
Multi-View Partial Point Cloud Dataset



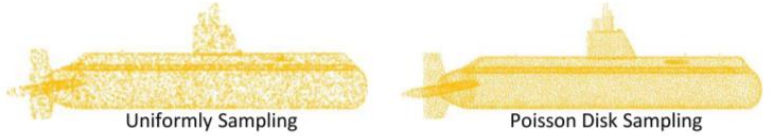
(a) 26 uniformly distributed camera poses



(b) The 26 rendered incomplete point clouds for this 3D airplane



(c) Rendered partial point clouds with different resolutions



(d) Sampled complete point clouds with different sampling methods

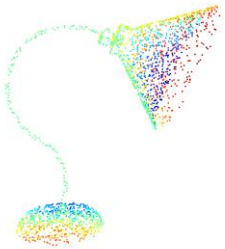
Partial Points with Different Resolutions

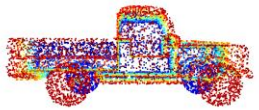
Render with 640 x 480 (Medium)

Render with 1600 x 1200 (High)

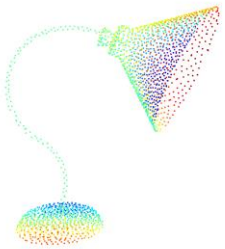
MVP Dataset: Ground Truth Comparison

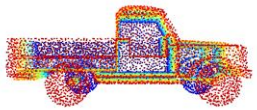
Uniform Sampling



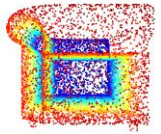


Poisson Disk Sampling

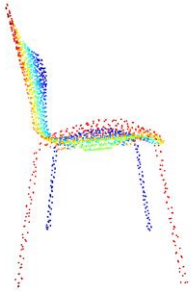


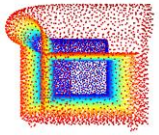


Uniform Sampling



Poisson Disk Sampling





MVP Dataset

Table 1: **Comparing MVP with existing datasets.** MVP has many appealing properties, such as 1) diversity of uniform views; 2) large-scale and high-quality; 3) rich categories. Note that both PCN and C3D only randomly render **One** incomplete point cloud for each CAD model to construct their testing sets. (C3D: Completion3D; Cat.: Categories; Distri.: Distribution; Reso.: Resolution; PC: Point Cloud; FPS: Farthest Point Sampling; PDS: Poisson Disk Sampling. Point cloud resolution is shown as multiples of 2048 points.)

	#Cat.	Training Set		Testing Set		Virtual Camera			Complete PC		Incomplete PC	
		#CAD	#Pair	#CAD	#Pair	Num.	Distri.	Reso.	Sampling	Reso.	Sampling	Reso.
PCN [29]	8	28974	~200k	1200	1200	8	Random	160×120	Uniform	8×	Random	~3000
C3D [21]	8	28974	28974	1184	1184	1	Random	160×120	Uniform	1×	Random	1×
MVP	**16**	2400	62400	1600	**41600**	**26**	**Uniform**	**1600×1200**	**PDS**	**1/2/4/8×**	**FPS**	1×

Completion Results on MVP

Table 2: Shape completion results (CD loss multiplied by 10^4) on our multi-view partial point cloud dataset (16,384 points). VRCNet outperforms all existing methods by convincing margins. Note that besides the conventional 8 categories in existing datasets, MVP allows evaluation on 8 additional categories.

Method	*airplane*	*cabinet*	*car*	*chair*	*lamp*	*sofa*	*table*	*watercraft*	*bed*	*bench*	*bookshelf*	*bus*	*guitar*	*motorbike*	*pistol*	*skateboard*	Avg.
PCN [29]	2.95	4.13	3.04	7.07	14.93	5.56	7.06	6.08	12.72	5.73	6.91	2.46	1.02	3.53	3.28	2.99	6.02
TopNet [21]	2.72	4.25	3.40	7.95	17.01	6.04	7.42	6.04	11.60	5.62	8.22	2.37	1.33	3.90	3.97	2.09	6.36
MSN [14]	2.07	3.82	2.76	6.21	12.72	4.74	5.32	4.80	9.93	3.89	5.85	2.12	0.69	2.48	2.91	1.58	4.90
Wang et. al. [23]	1.59	3.64	2.60	5.24	9.02	4.42	5.45	4.26	9.56	3.67	5.34	2.23	0.79	2.23	2.86	2.13	4.30
ECG [15]	1.41	3.44	2.36	4.58	6.95	3.81	4.27	3.38	7.46	3.10	4.82	1.99	0.59	2.05	2.31	1.66	3.58
GRNet [27]	1.61	4.66	3.10	4.72	5.66	4.61	4.85	3.53	7.82	2.96	4.58	2.97	1.28	2.24	2.11	1.61	3.87
NSFA [30]	1.51	4.24	2.75	4.68	6.04	4.29	4.84	3.02	7.93	3.87	5.99	2.21	0.78	1.73	2.04	2.14	3.77
VRCNet (Ours)	**1.15**	**3.20**	**2.14**	**3.58**	**5.57**	**3.58**	**4.17**	**2.47**	**6.90**	**2.76**	**3.45**	**1.78**	**0.59**	**1.52**	**1.83**	**1.57**	**3.06**

Qualitative Results on MVP

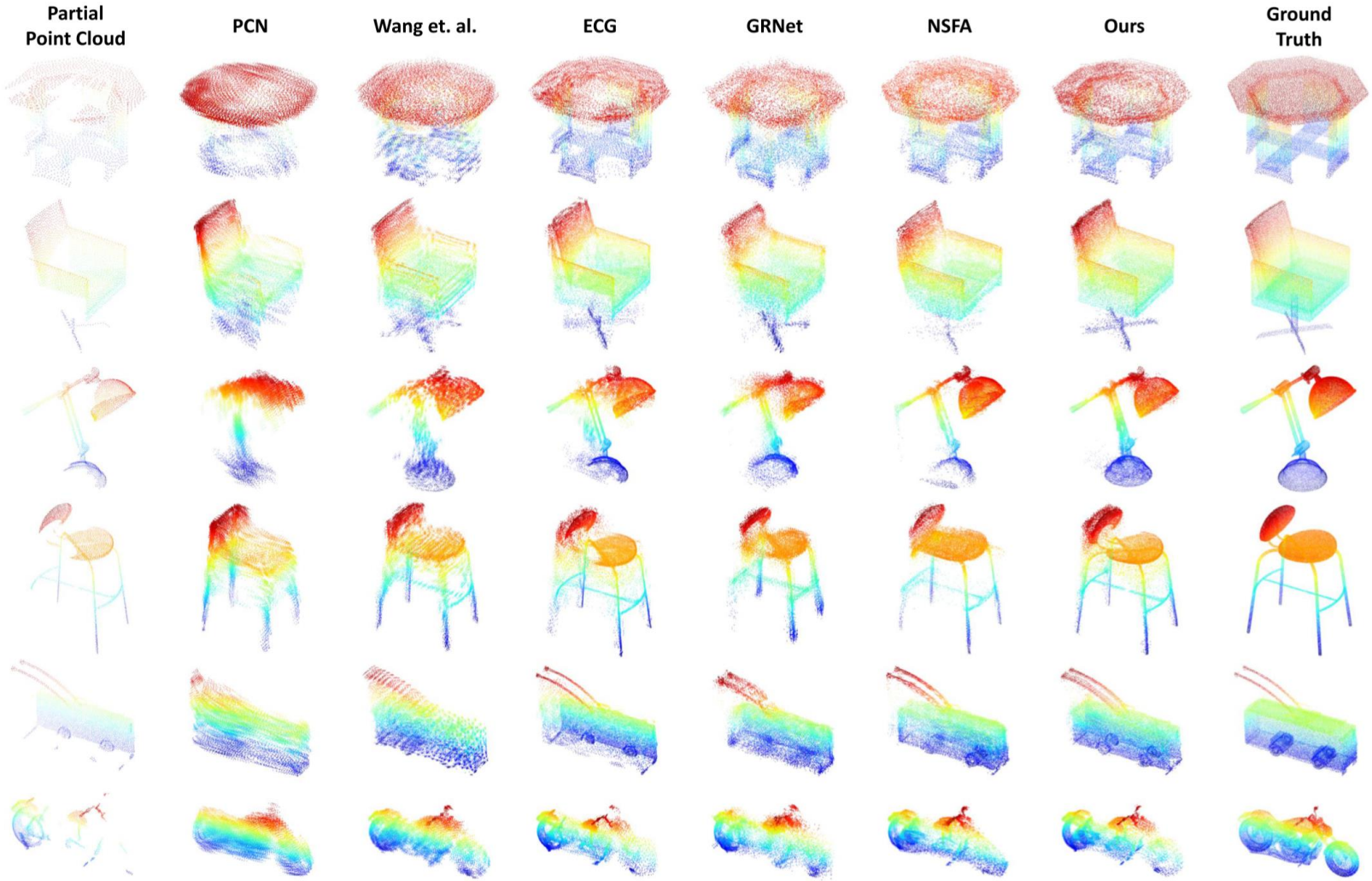


Figure 6: Qualitative completion results (16,384 points) on the MVP dataset by different methods. VRCNet can generate better complete point clouds than the other methods by learning geometrical symmetries.

Qualitative Results on Real Scans

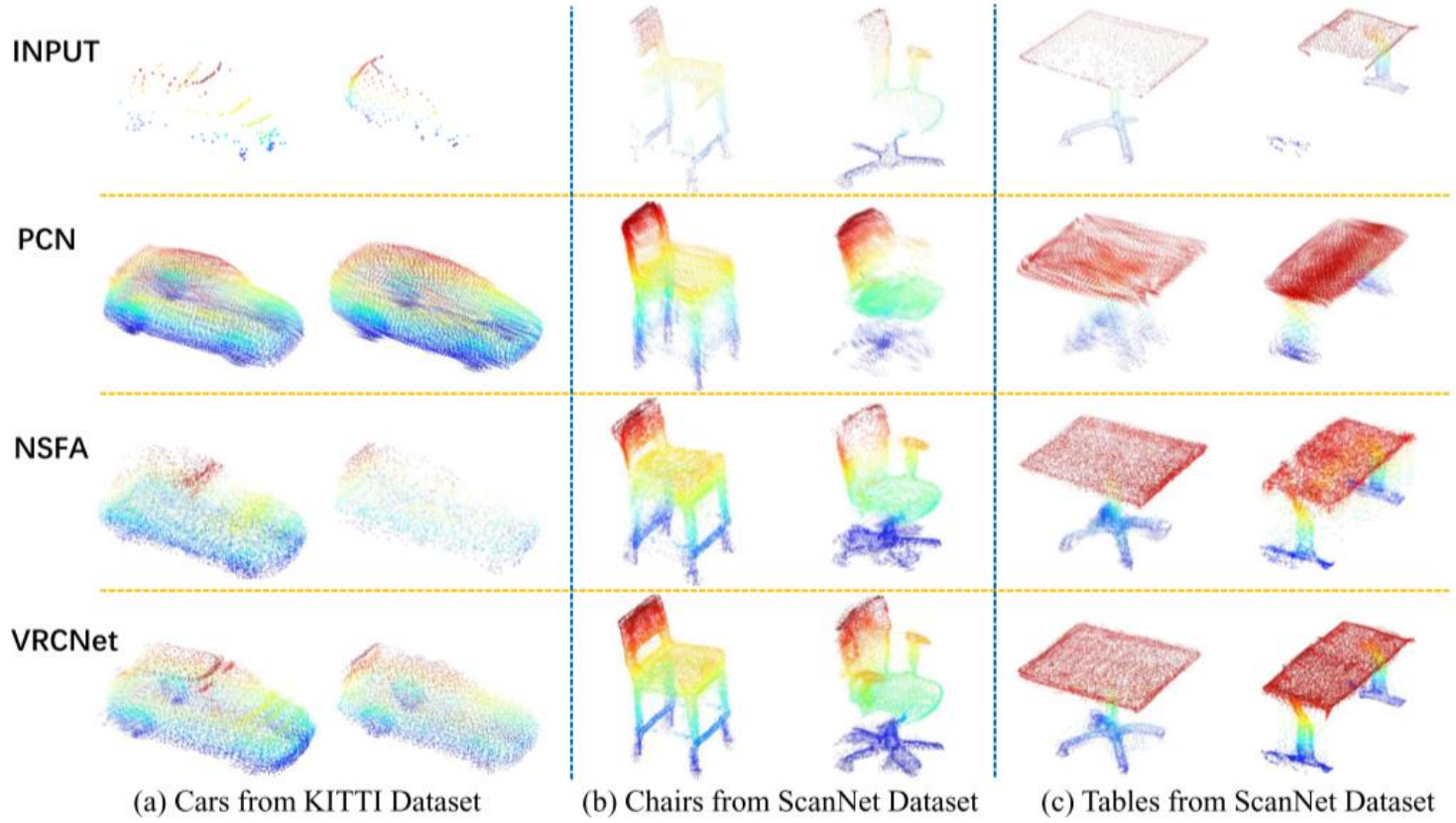


Figure 7: VRCNet generates impressive complete shapes for real-scanned point clouds by learning and predicting shape symmetries. (a) shows completion results for cars from Kitti dataset. (b) and (c) show completion results for chairs and tables from ScanNet dataset, respectively.

Conclusion

- ✓ We propose a comprehensive framework for point cloud completion:

  1. generate overall shape skeletons; 2. transfer local shape details

- ✓ We design many novel and powerful point cloud learning modules:

  Point Self-Attention Module (PSA); Point Selective Kernel Module (PSK)

- ✓ We establish a Multi-View Partial (MVP) Point Cloud Dataset.

  It can be used for many partial point cloud applications, such as complete, generation, registration and detection.

Future Directions

Point Cloud Completion (single view)

- Point cloud denoise
- CD loss cannot supervise underling 3D surfaces very well
- Improve our generation capabilities and achieve multi-modal completion

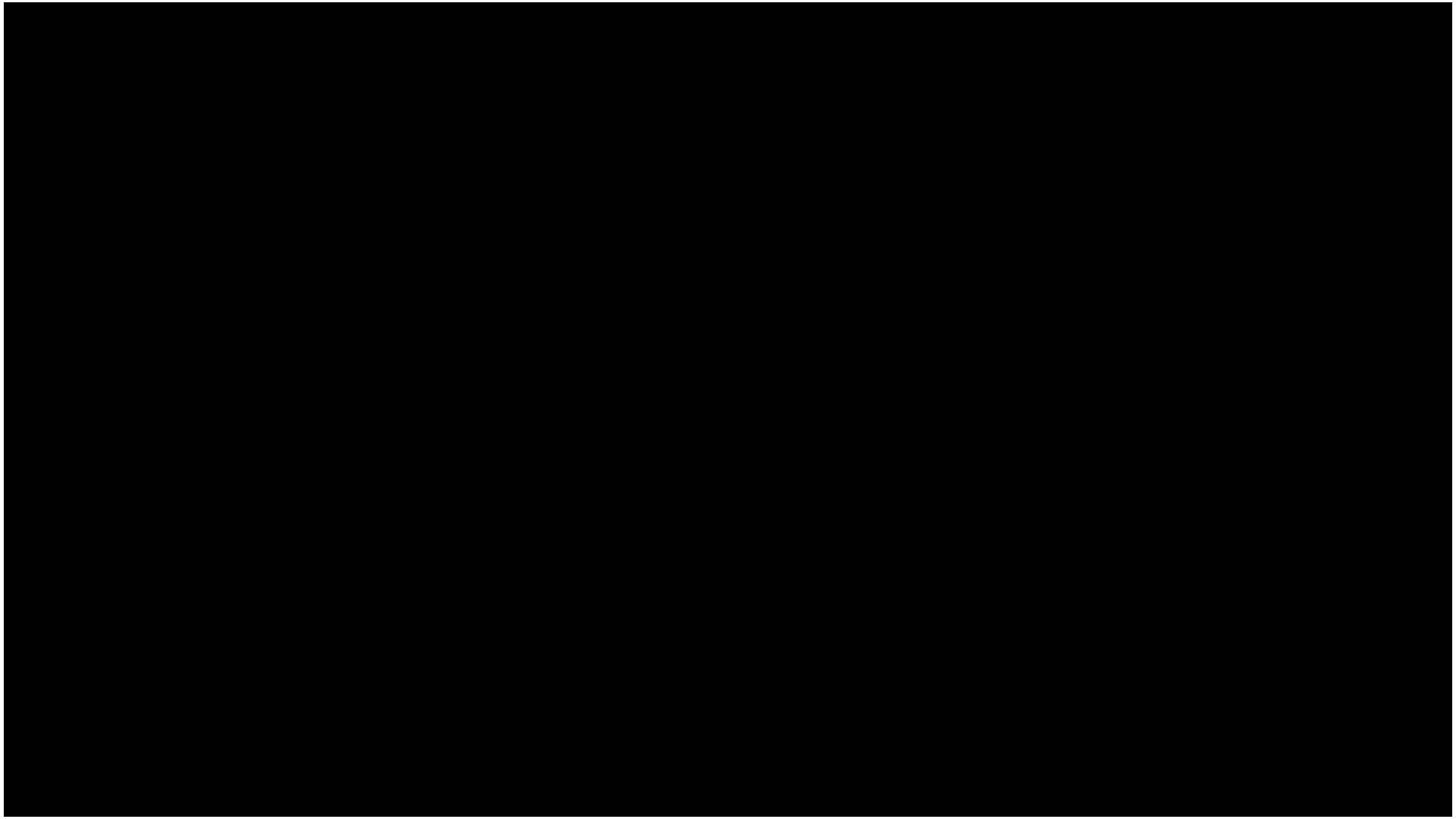
Point Cloud Consolidation (single view)

- Point cloud upsampling
- 3D mesh generation

Point Cloud Registration (multiple views)

- Partial-to-Partial point cloud registration
- 3D shape reconstruction in canonical pose
- Joint point cloud completion and registration

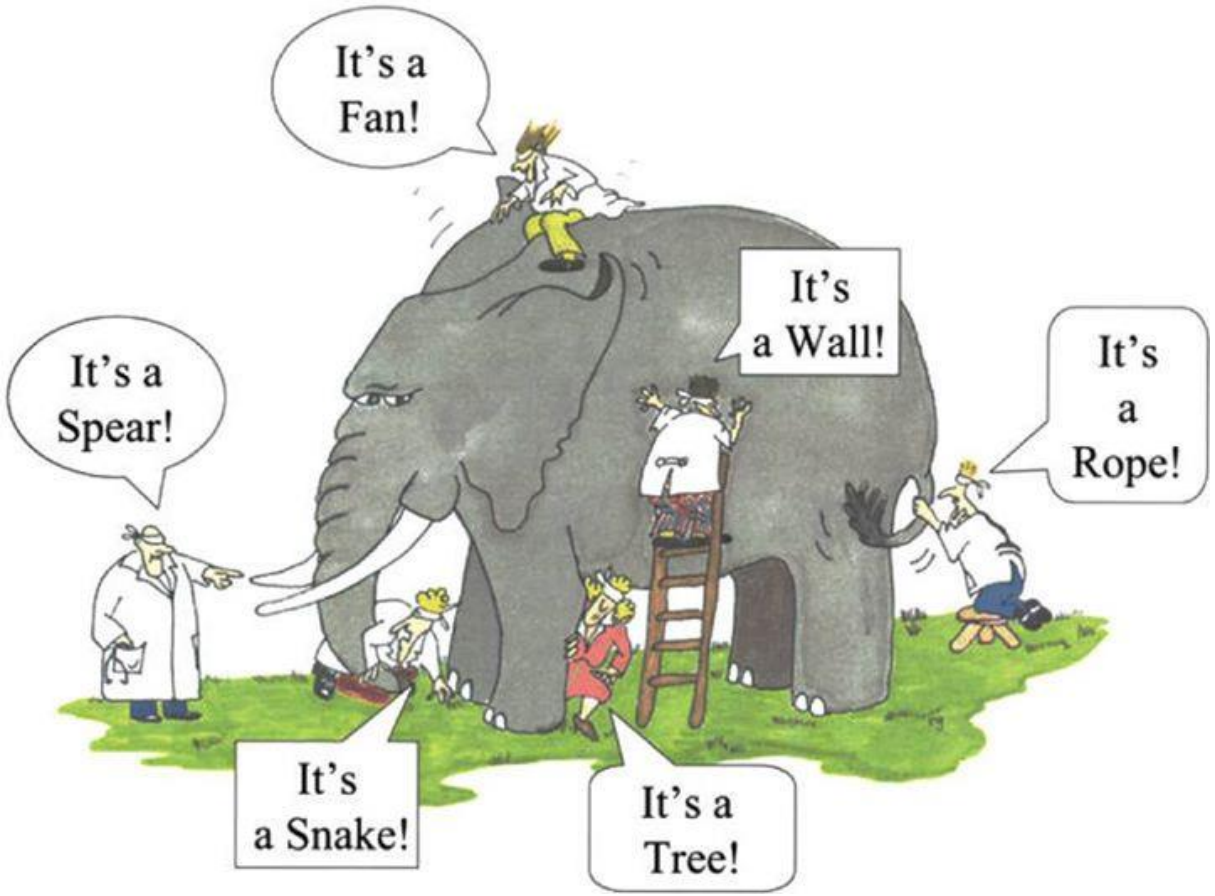
Demo Video



CVPR
VIRTUAL JUNE 19-25

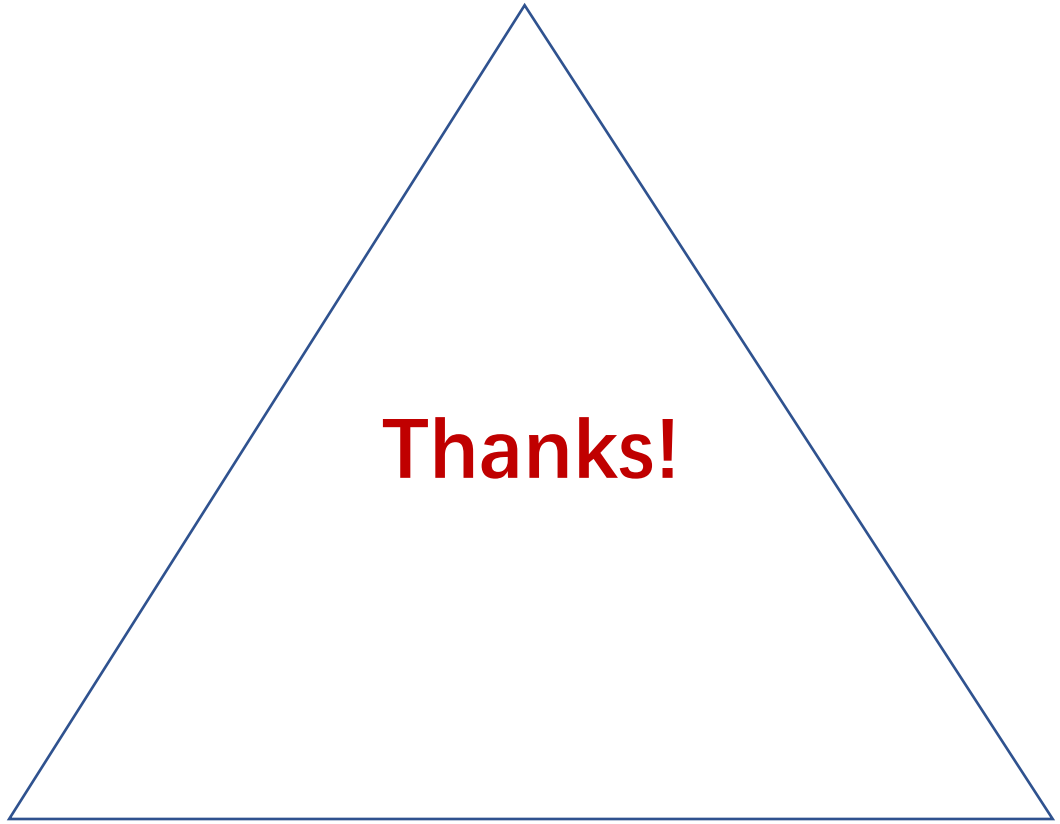
NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE | S-LAB FOR ADVANCED INTELLIGENCE

Partial Observations (2.5D)



Code on github

Partial Observations



3D Imagination

Modern Sensor