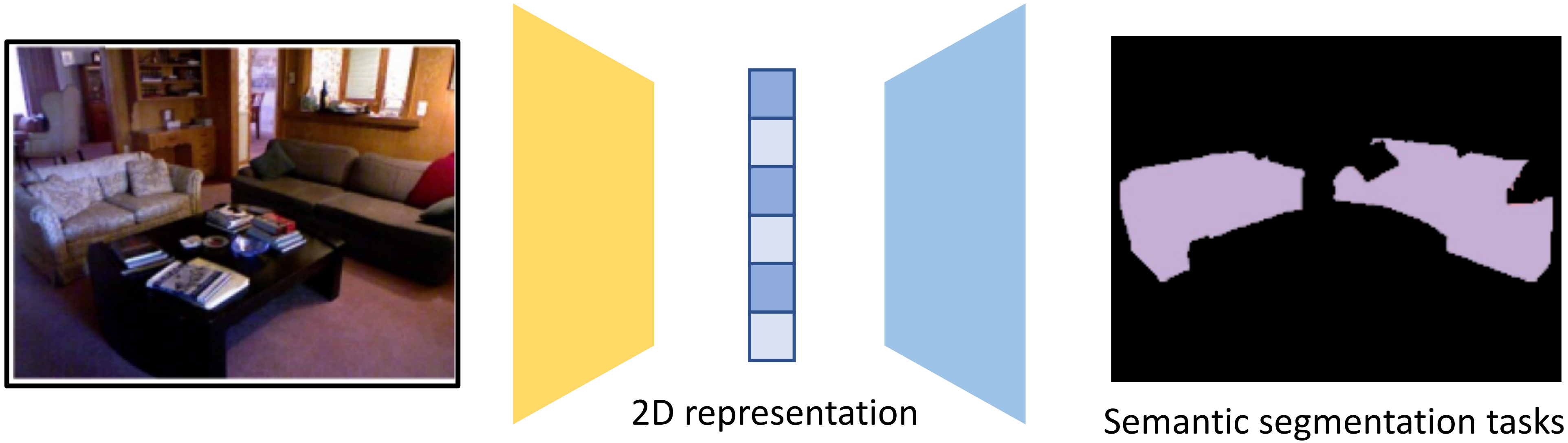
JUNE 18-22, 2023 CVPR VANCOUVER, CANADA
THU-PM-073

VISTEC

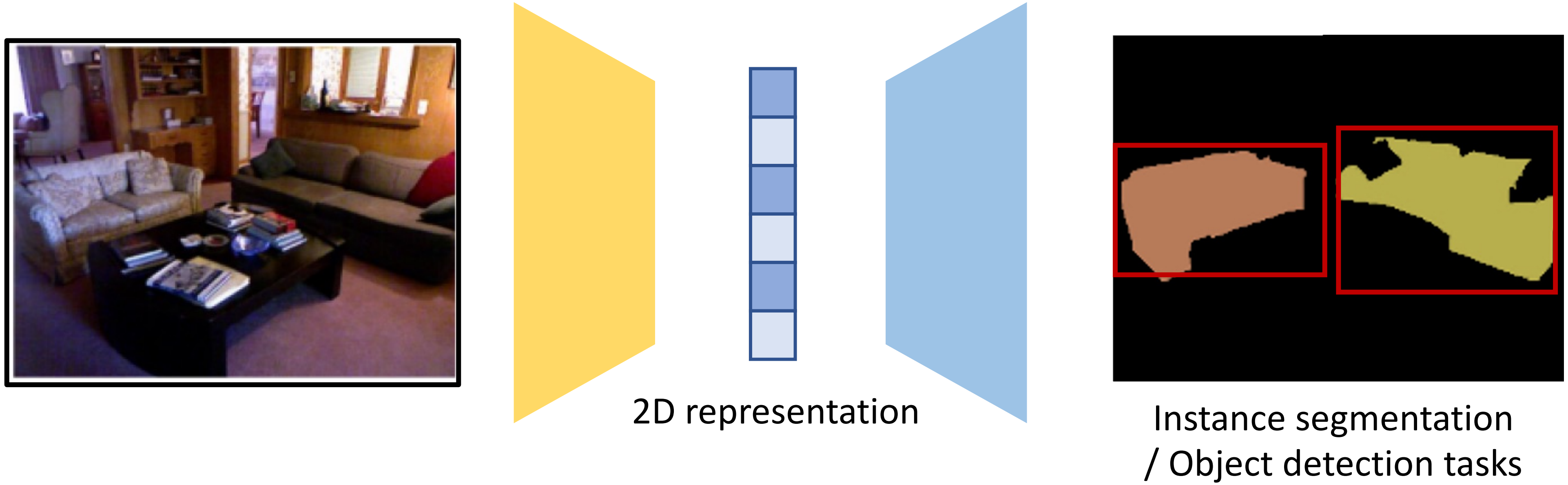
Learning Geometric-aware Properties in 2D Representation Using Lightweight CAD Models, or Zero Real 3D Pairs

Pattaramanee Arsomngern Sarana Nutanong Supasorn Suwajanakorn

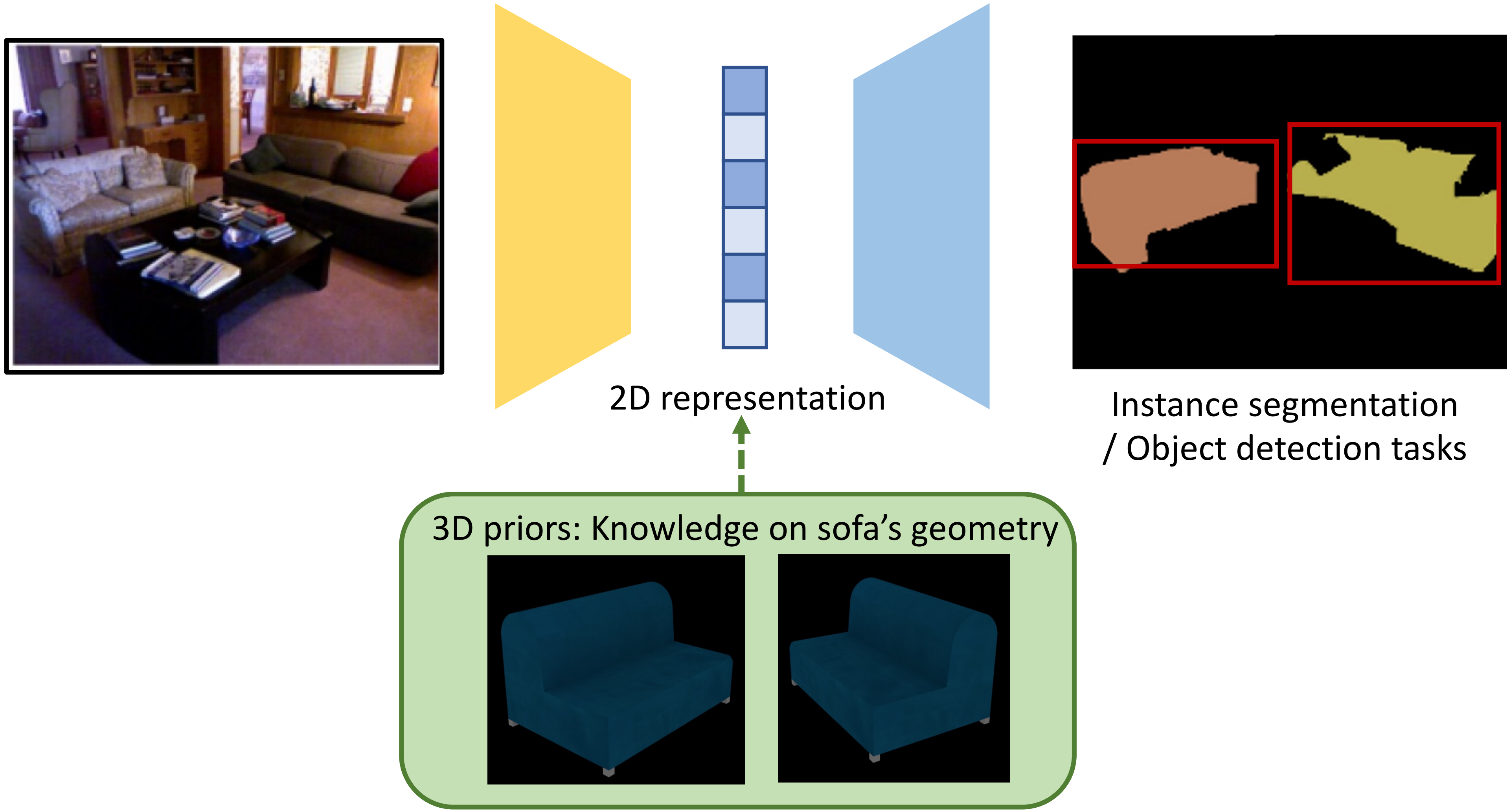
Improving 2D representation with 3D priors



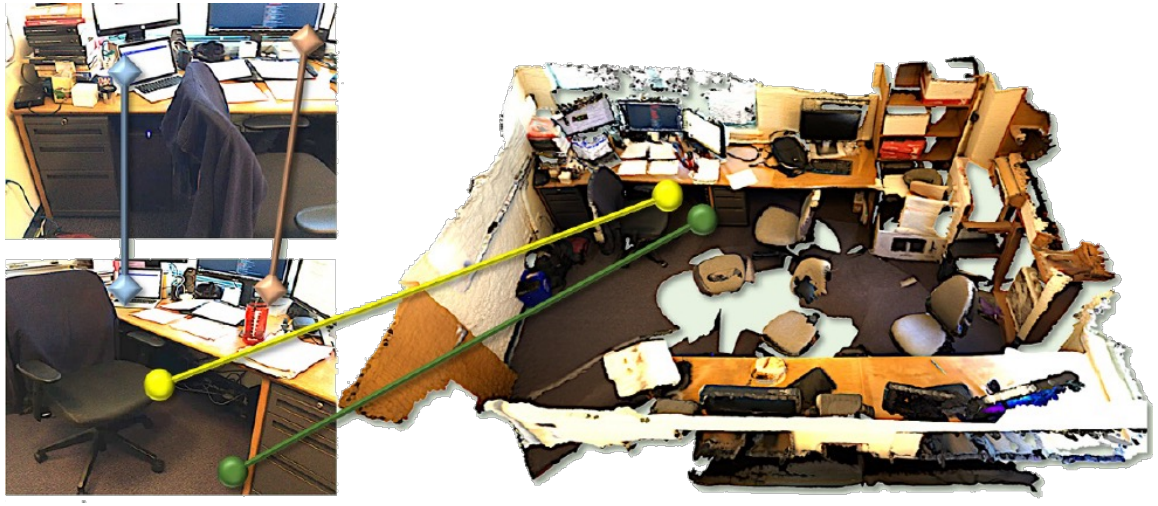
Improving 2D representation with 3D priors



Improving 2D representation with 3D priors



Prior work uses *heavyweight* 3D scene scans.



Pri3D (Hou et al. 2021)

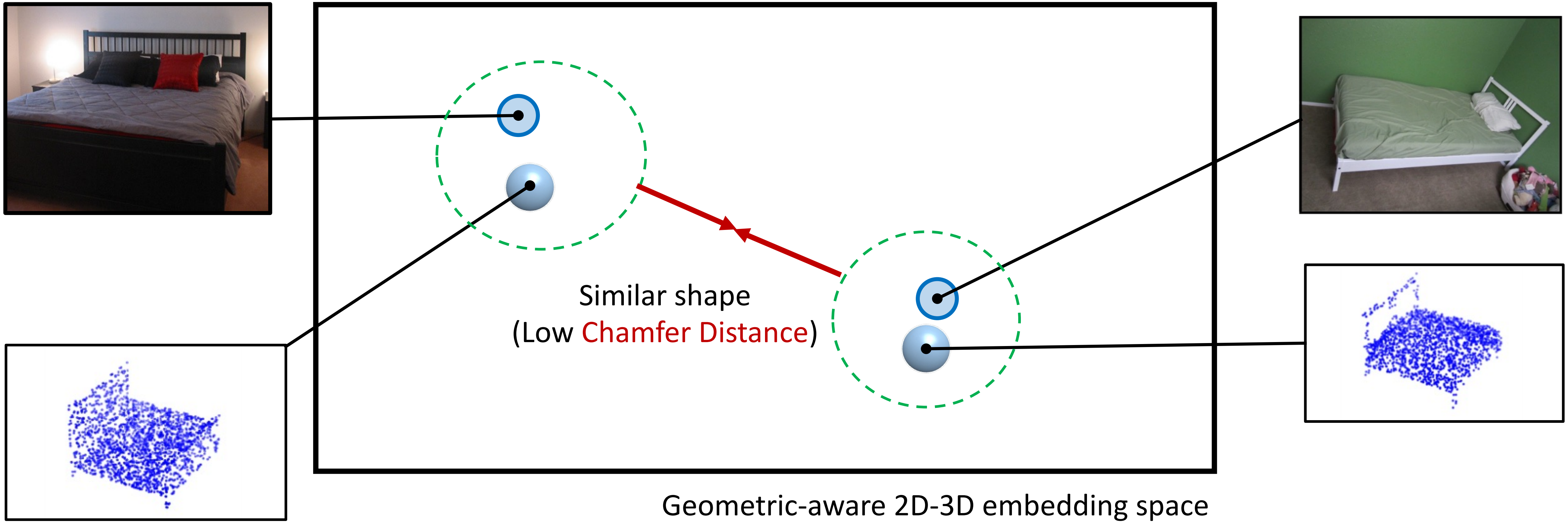
Image credit: Pri3D (Hou et al. 2021)

Prior work uses *heavyweight* 3D scene scans.

Pri3D (Hou et al. 2021)

Our method:
Utilizing *lightweight* CAD models as a 3D prior

Key idea: Joint 2D-3D space with Chamfer Distance



State-of-the-art performance

mIOU Improvement
from 2D-only methods

+ 1.83

*Compared to SimCLR (Chen et al.)
NYUv2 semantic segmentation

mIOU difference
from methods using 3D scenes

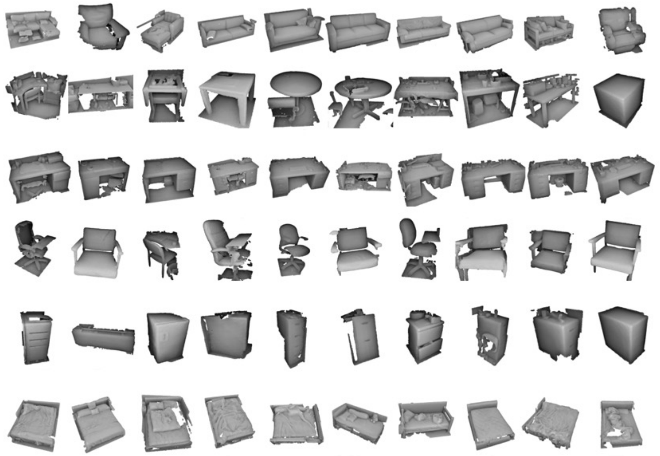
- 0.16

*Compared to SOTA (Set-InfoNCE, Chen et al.)
NYUv2 semantic segmentation

Unlimited (psuedo) training pairs

Massive RGB data

CAD
Reconstruction



Massive RGB-CAD pairs

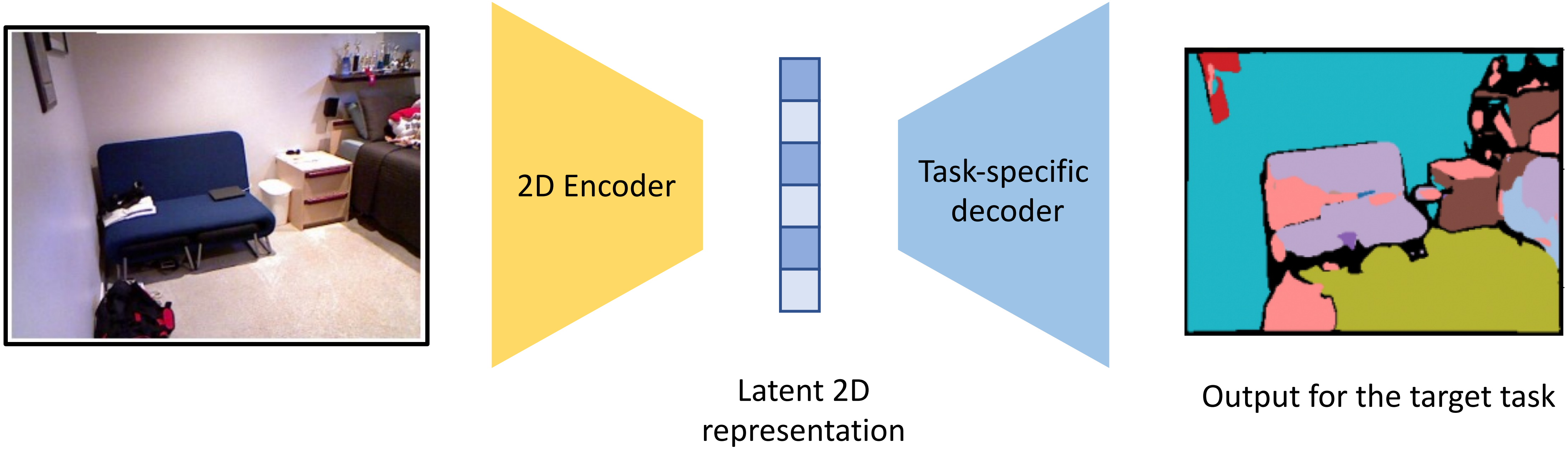
JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA
THU-PM-073

VISTEC

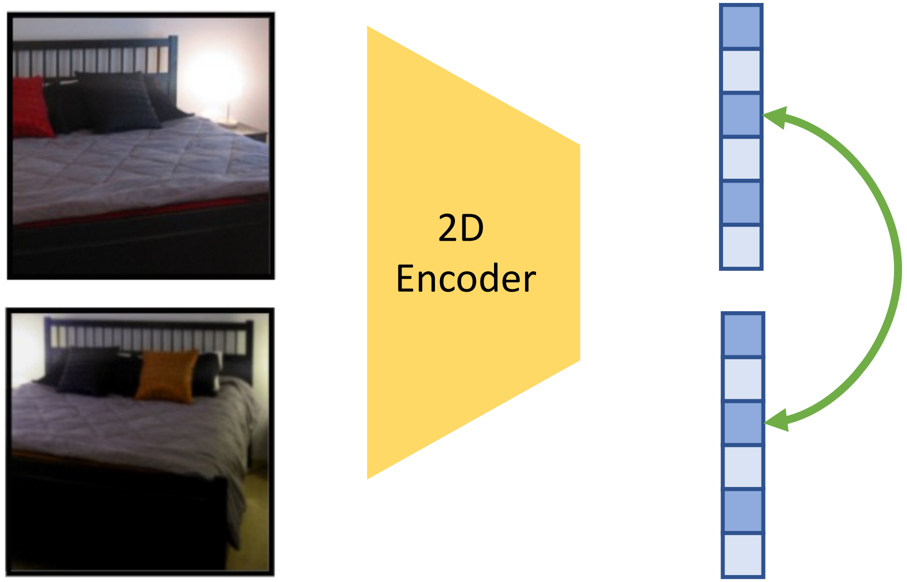
Learning Geometric-aware Properties in 2D Representation Using Lightweight CAD Models, or Zero Real 3D Pairs

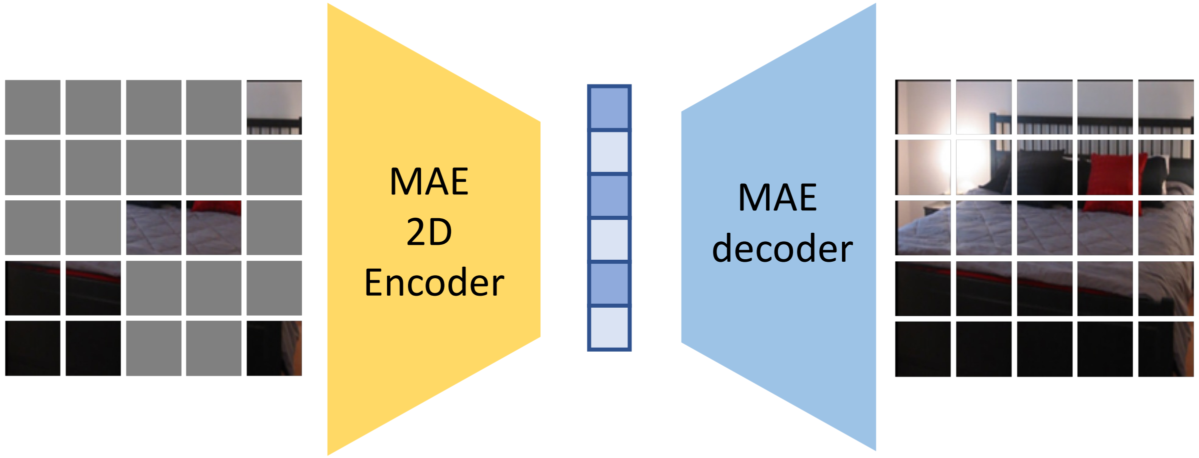
Pattaramanee Arsomngern   Sarana Nutanong   Supasorn Suwajanakorn

Common approach to solving 2D object understanding



2D Self-supervised encoders





SimCLR (Chen et al.)

Learning through 2D augmentations

MAE (He et al.)

Learning through 2D masked modelling

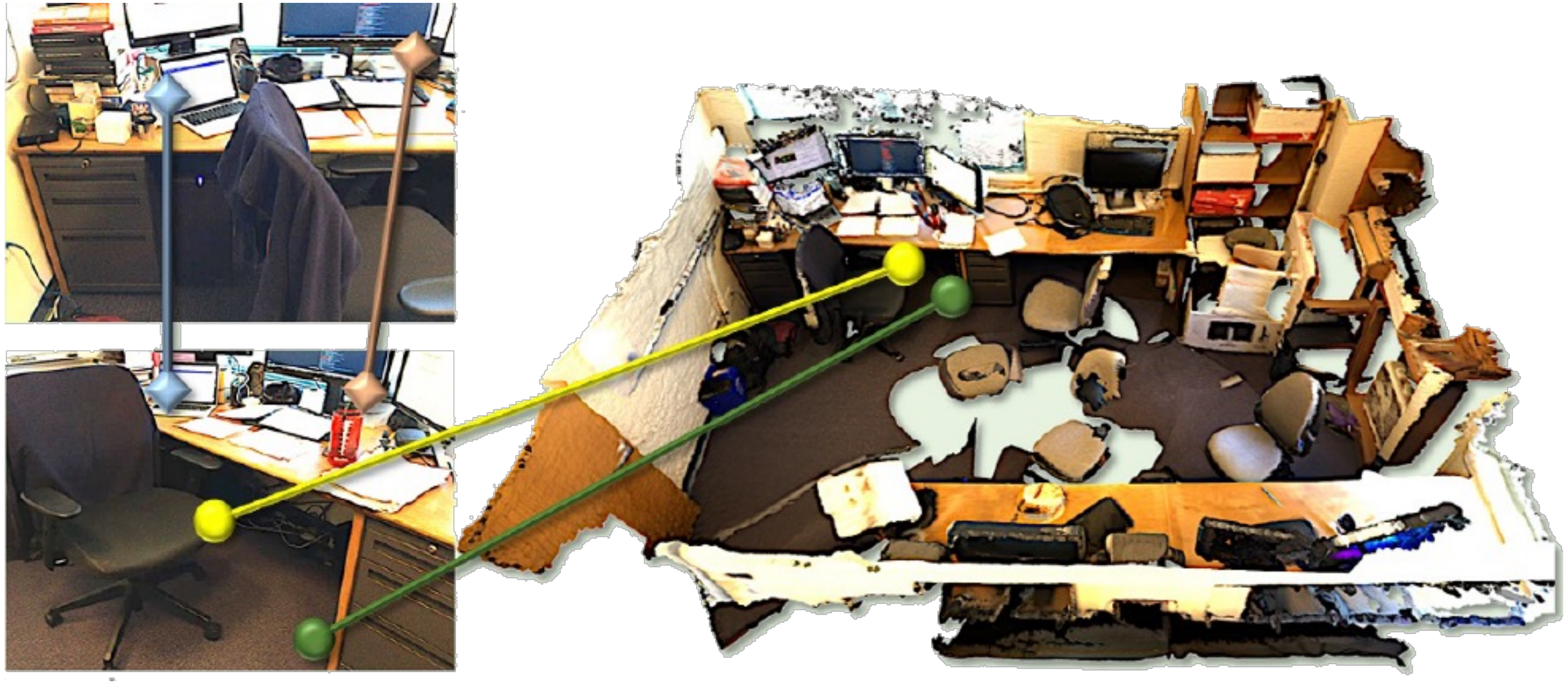
Drawbacks of 2D self-supervised encoders

Limited geometric information:
Flipped or different crops

Unseen view

https://www.naturalbedcompany.co.uk/product/kyoto-japanese-bed-headboard/

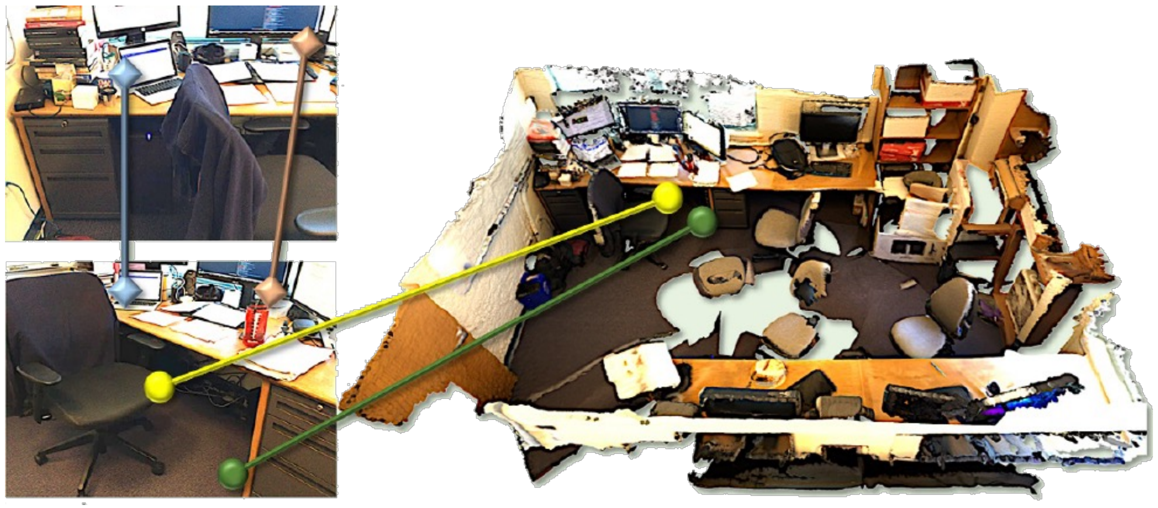
Better 2D understanding through 3D priors



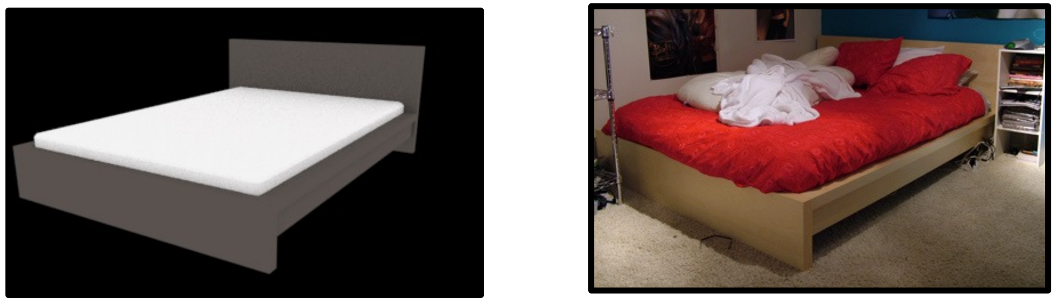
Pri3D (Hou et al. 2021)

Image credit: Pri3D (Hou et al. 2021)

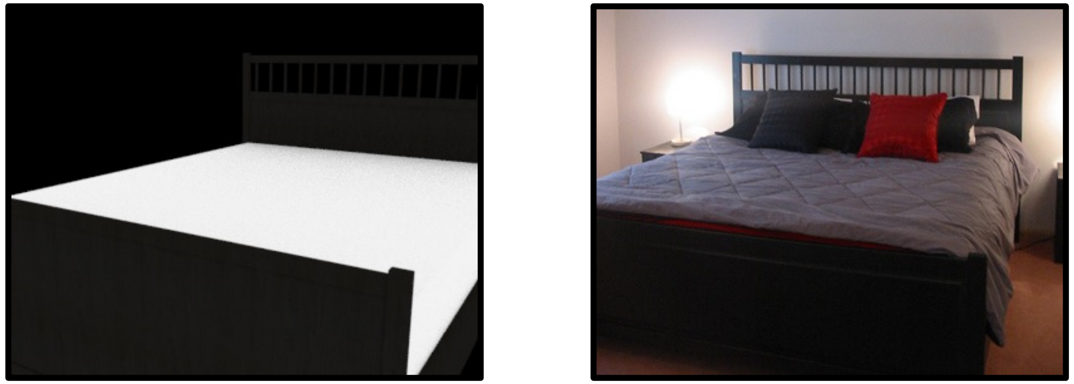
Better 2D understanding through 3D priors



Pri3D (Hou et al. 2021)



Alternative 3D priors?



Better 2D understanding through 3D priors

Pri3D (Hou et al. 2021)

Our method:
Utilizing *lightweight* CAD models as 3D priors

State-of-the-art results

Improvement from

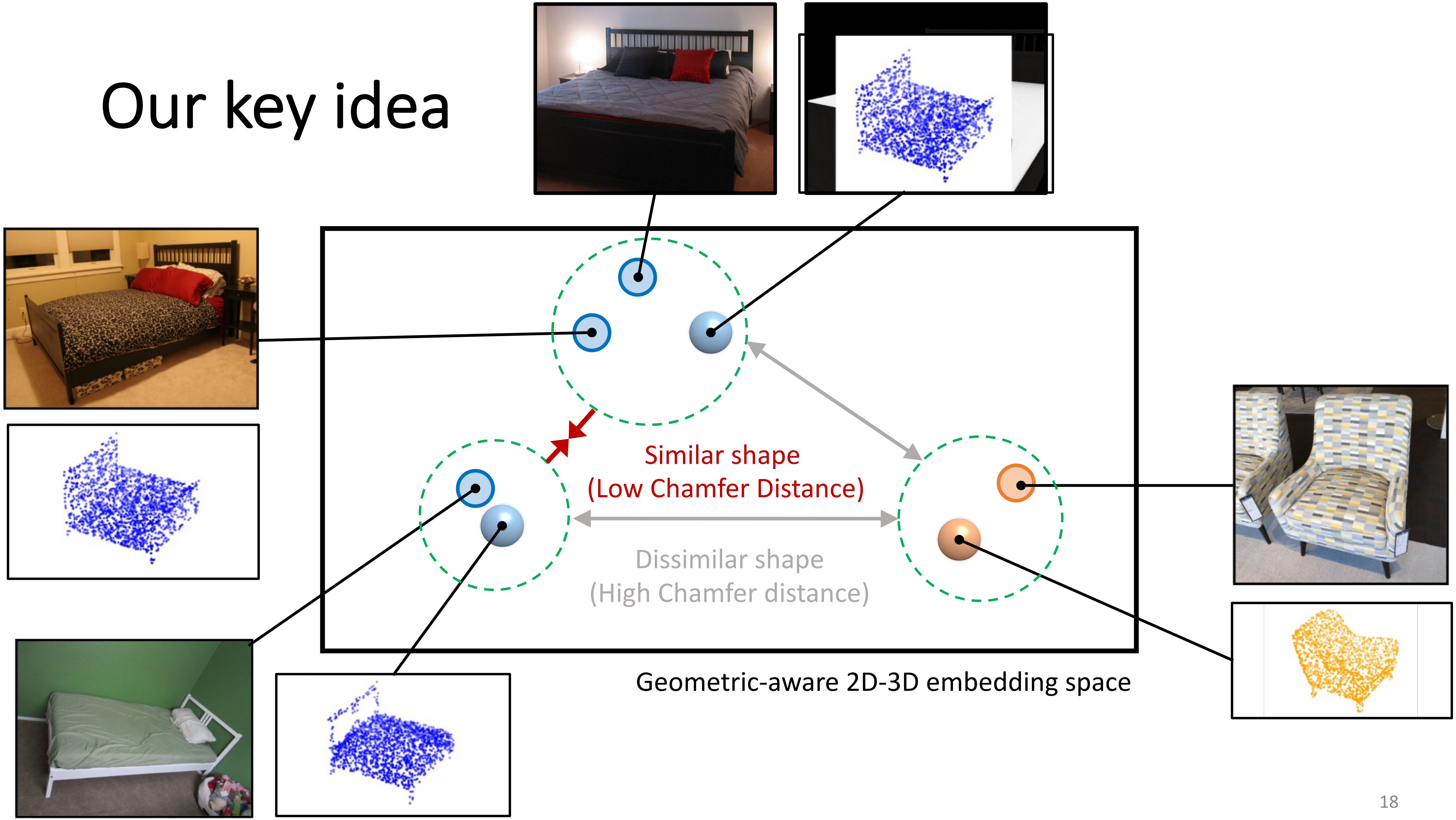
- ResNet-based 2D SSL
- ViT-based 2D SSL

Performance difference from methods with 3D scenes

- 0.16**

** From SOTA (Set-InfoNCE, Chen et al.) on NYUv2 semantic segmentation

Our key idea



Similar shape
(Low Chamfer Distance)
Dissimilar shape
(High Chamfer distance)
Geometric-aware 2D-3D embedding space

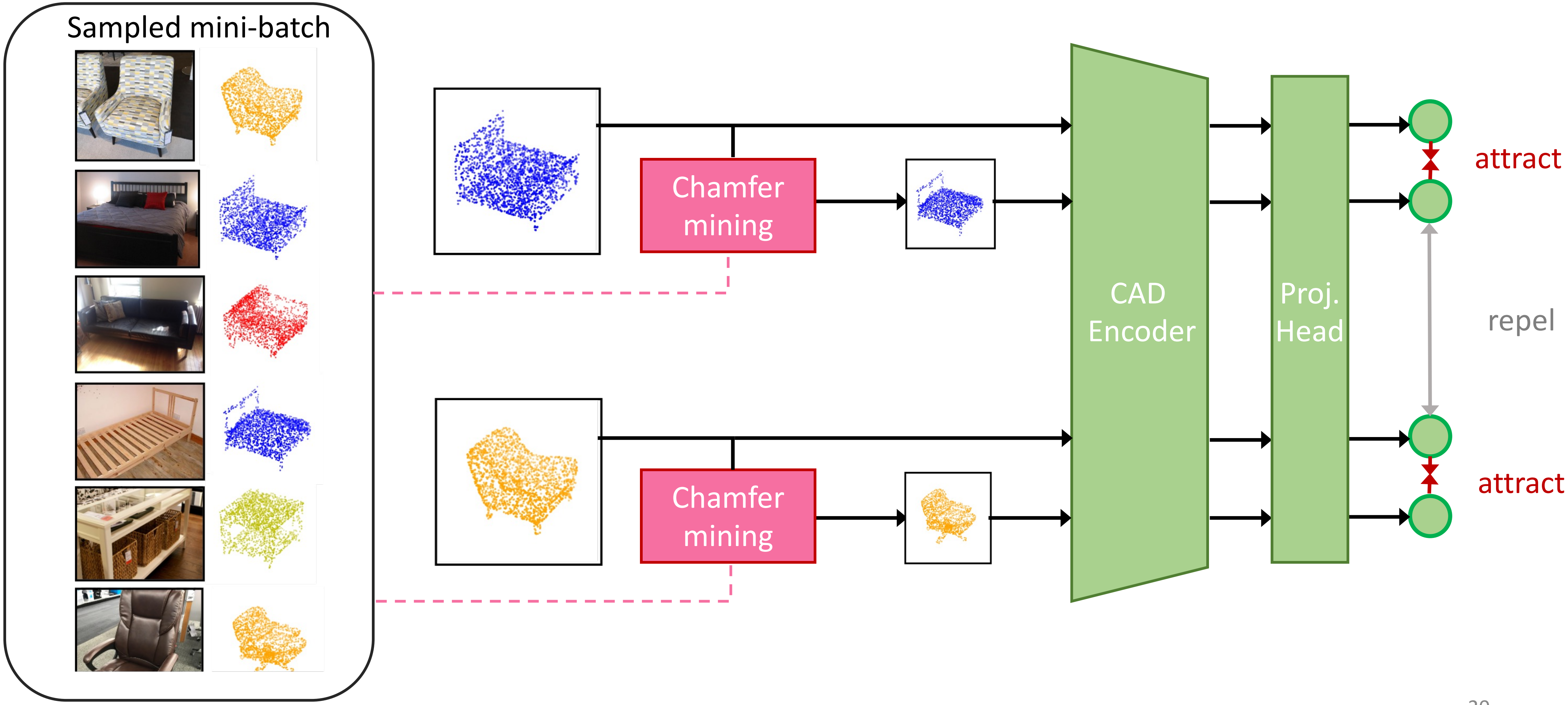
3 Contrastive loss functions

1. Geometric-aware CAD features

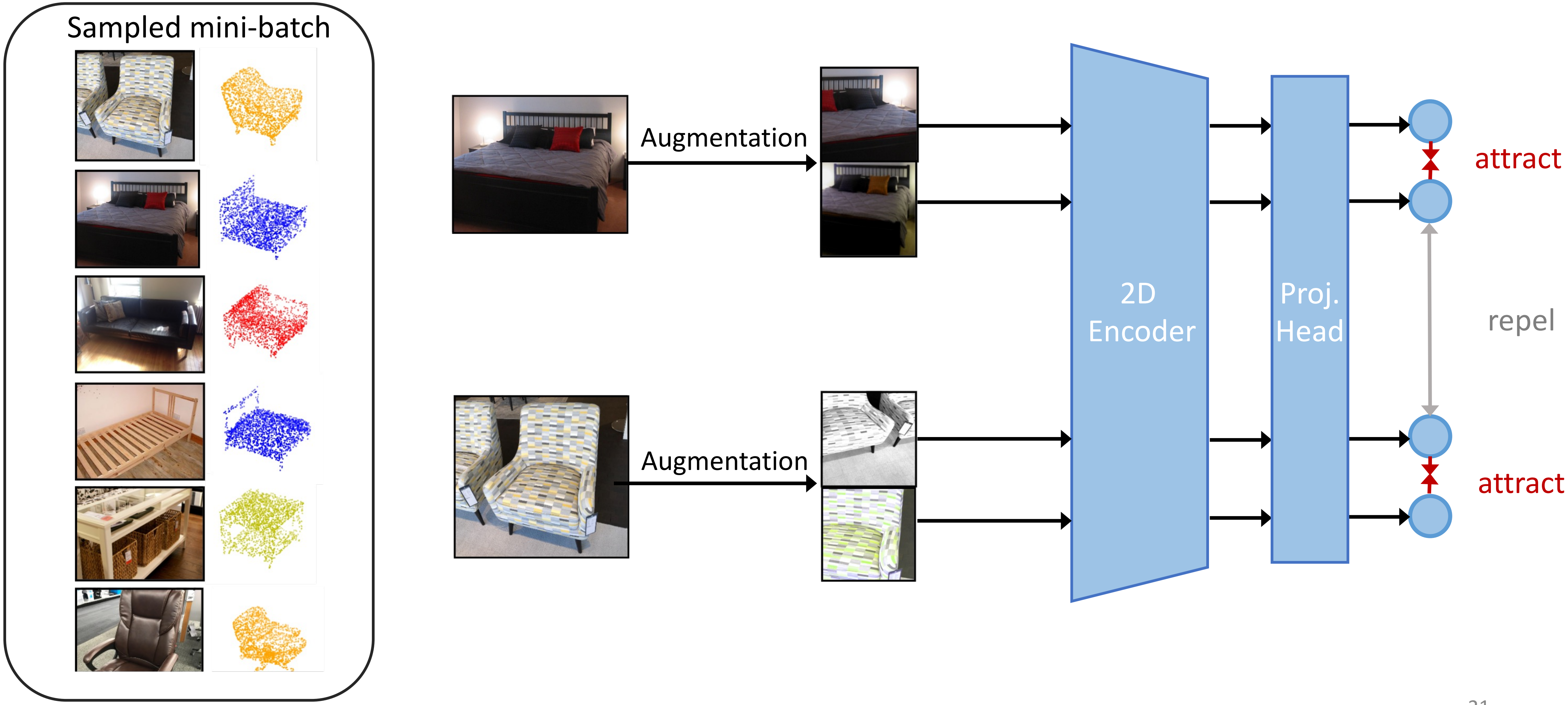
2. Discriminative visual features

3. Cross-modal sharing 2D-3D properties

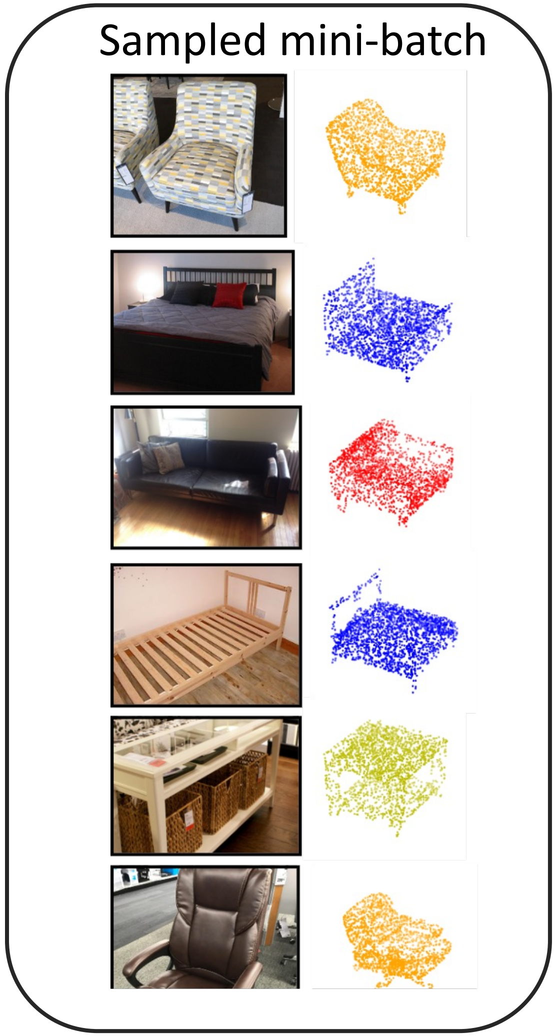
1. Geometric-aware CAD features

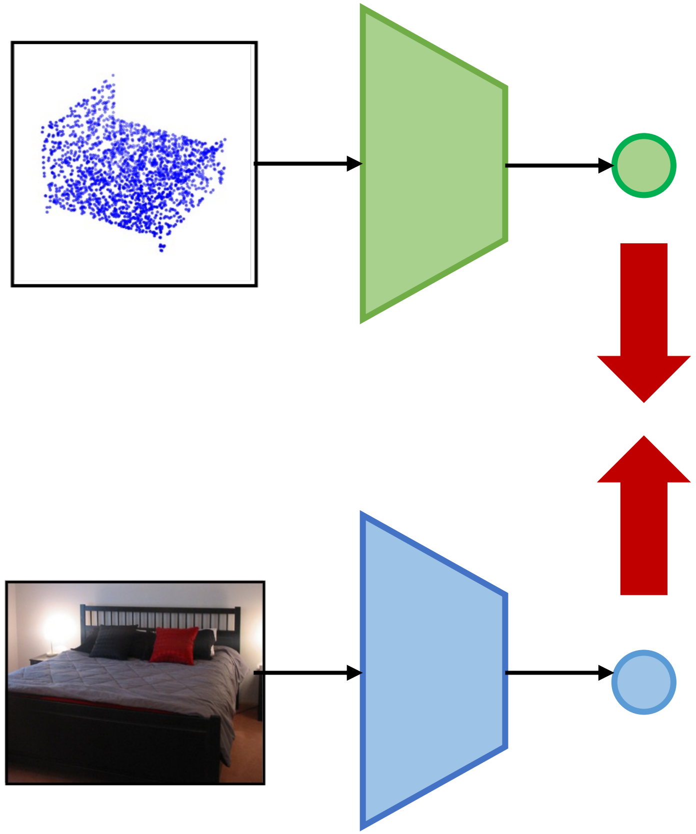


2. Discriminative visual features

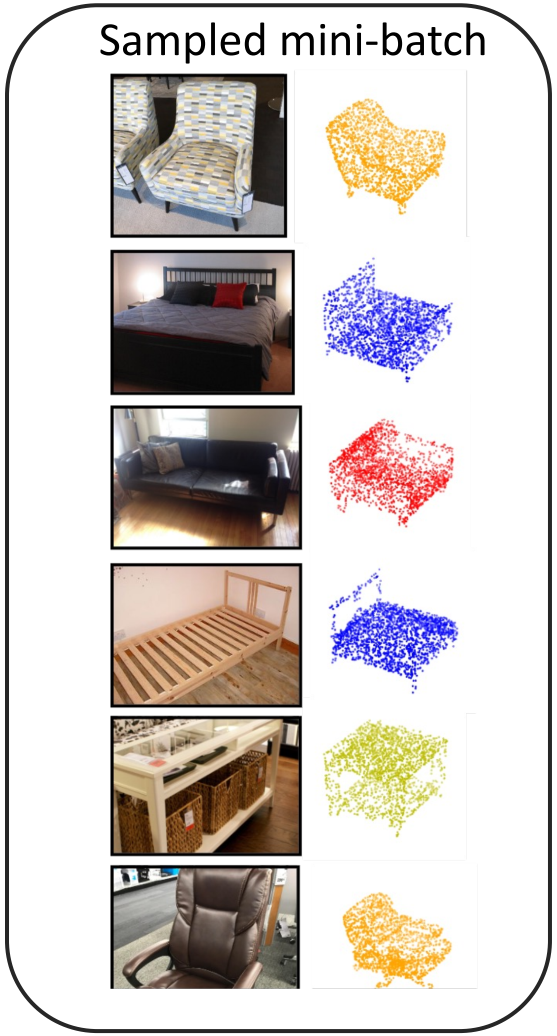


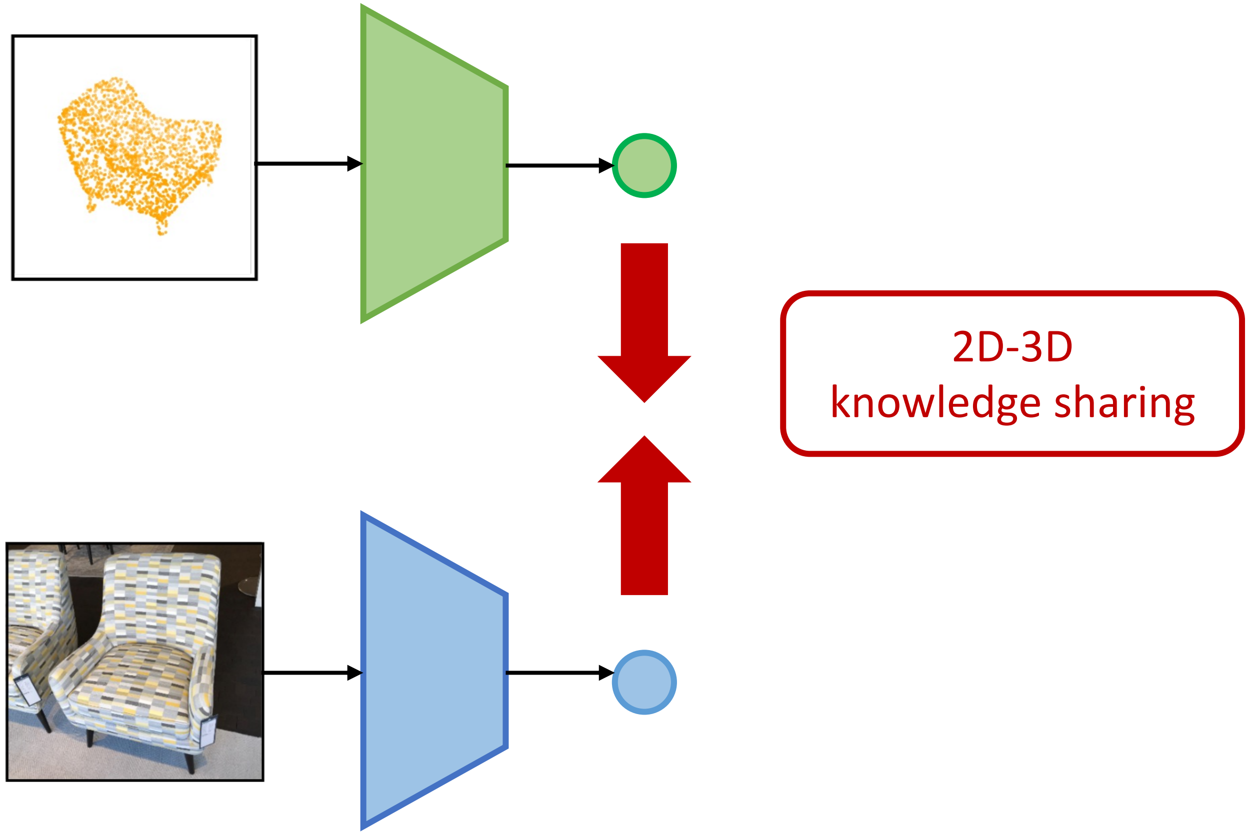
3. Cross-modal sharing 2D-3D properties





3. Cross-modal sharing 2D-3D properties

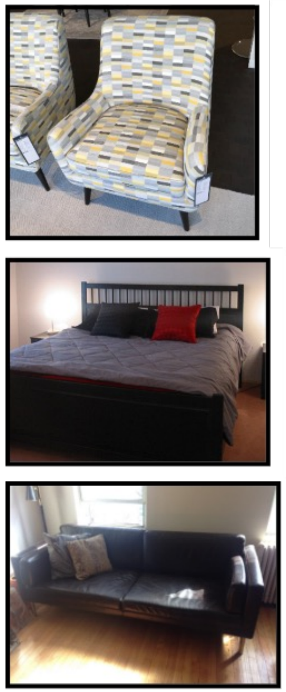




No groundtruth pairs are required

~~Paired RGB-CAD dataset~~

No groundtruth pairs are required

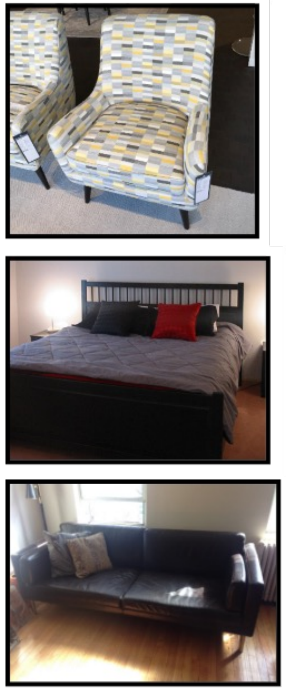


2D-3D generator

RGB-CAD retrieval works

ROCA (Gumeli et al.)

No groundtruth pairs are required

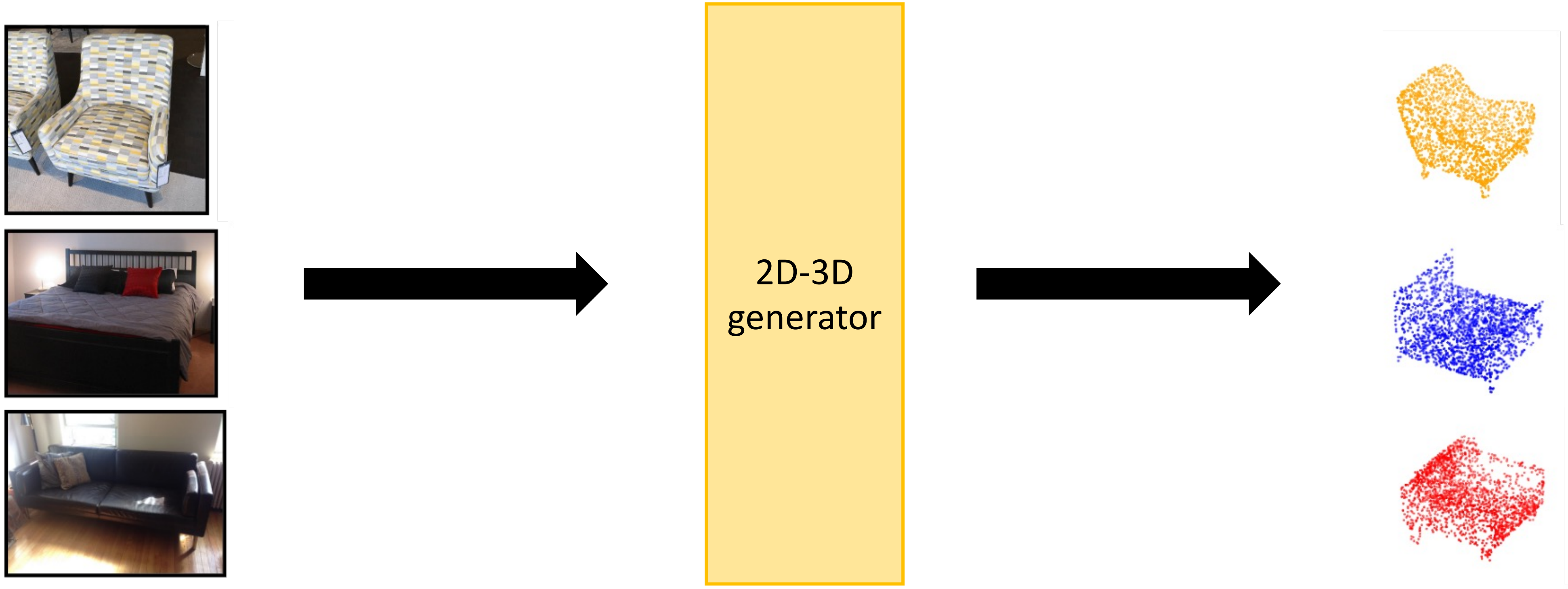


2D-3D generator

3D generation works

RealFusion (Melas-Kyriazi et al.)

No groundtruth pairs are required



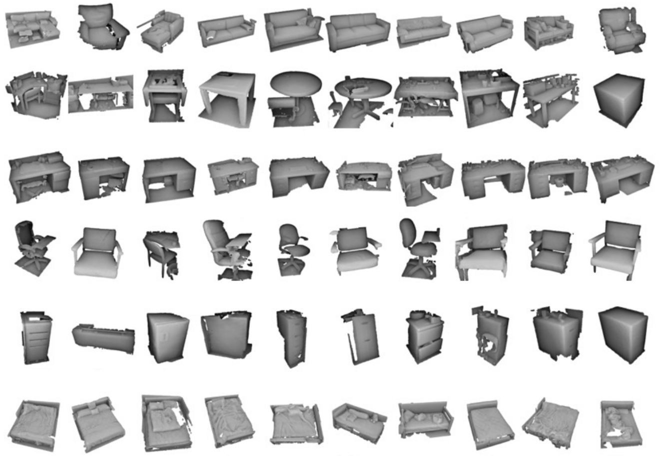
Acquired pseudo-pairs

Unlimited availability of training pairs

Massive-scale of RGB data

2D-3D generator

ROCA (Gumeli et al.)

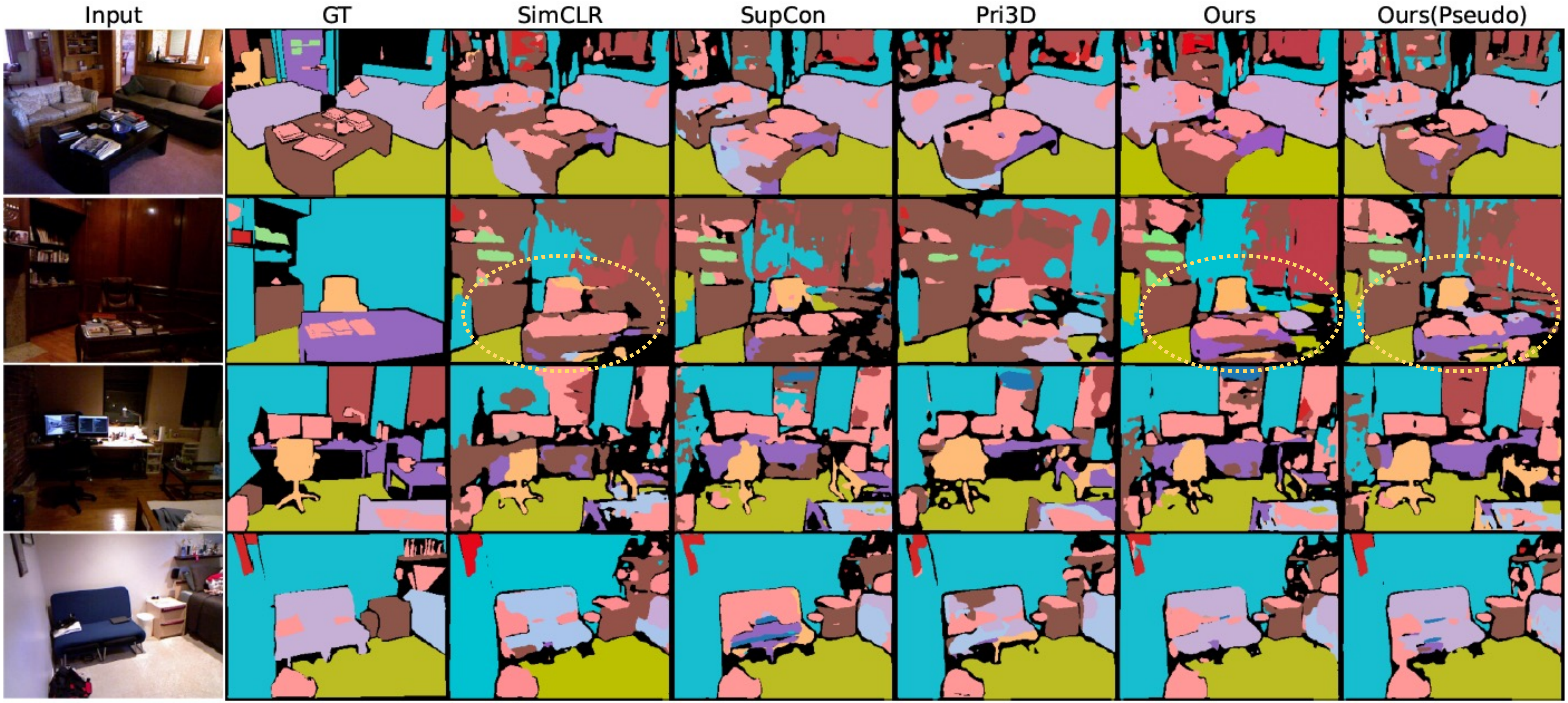


Massive RGB-CAD pairs

Experimental results

Semantic segmentation task

Arch.	GT pair	Method	3D	NYUv2 mIoU	NYUv2 mIoU [25]	ScanNet mIoU	ScanNet mIoU [25]	indoor ADE20k mIoU	indoor ADE20k mIoU [25]	SUNRGB-D mIoU	SUNRGB-D mIoU [25]


Input
GT
SimCLR
SupCon
Pri3D
Ours
Ours(Pseudo)

Experimental results

Our preliminary experiment on pseudo-pairs

5.95

times larger*

pre-training data

+ 0.15

mIoU

over SOTA (Set-InfoNCE, Chen et al. 2022) in NYUv2 semantic segmentation

* 50k RGB-CAD training pairs collected from ImageNet and COCO dataset while the original setting is Pix3D dataset with 7k ground truth pairs.

Experimental results

- Instance segmentation and object detection (NYUv2, Indoor/ Outdoor COCO)
  - Outperformed SOTA in all settings
- Object retrieval (Pix3D)
  - +3.13 (Resnet-50) and +1.75 R@1 from SOTA 2D-only works

Full information in the paper!

					NYUv2						indoor COCO						outdoor COCO					
					Object Det.			Instance segm.			Object Det.			Instance seg.			Object Det.			Instance seg.		
Arch.	Size	GT pair	Method	3D	AP50	AP75	AP	AP50	AP75	AP	AP50	AP75	AP	AP50	AP75	AP	AP50	AP75	AP	AP50	AP75	AP
RN50	480	2D only	SupImg	-	29.9	17.3	16.8	25.1	13.9	13.4	41.78	24.21	23.70	39.16	23.35	22.61	46.09	26.98	28.08	42.45	23.34	23.92
			SimCLR	-	32.81	20.15	19.24	29.10	15.97	15.62	43.63	26.46	25.45	40.87	24.79	23.86	48.15	28.75	30.40	44.31	25.01	24.99
			SupCon	-	33.23	20.36	19.63	29.44	16.16	15.83	43.66	26.34	25.32	40.84	24.53	23.75	47.89	28.67	30.29	44.16	24.63	24.97
			SupCon (fine)	-	32.56	19.74	18.92	29.06	16.11	15.74	43.58	25.95	25.21	40.65	24.22	23.66	45.01	27.90	26.59	41.97	25.61	24.66
		pseudo	*Ours (pseudo)*	CAD	34.45	20.27	19.72	29.64	16.24	16.13	43.74	26.47	25.48	40.92	24.77	23.91	-	-	-	-	-	-
		2D-3D	CrossPoint	CAD	28.42	15.94	15.22	24.49	13.32	13.11	40.25	22.78	22.26	38.54	21.92	20.80	43.22	24.57	25.60	39.75	21.93	21.11
			Pri3D	scene	34.0	20.4	19.4	29.5	16.3	15.8	43.49	26.40	25.22	40.71	24.72	23.61	-	-	-	-	-	-
			Set-InfoNCE	scene	34.6	20.5	19.7	29.7	16.3	16.5	-	-	-	-	-	-	-	-	-	-	-	-
			Ours	CAD	**34.85**	**20.89**	**20.12**	**30.03**	**16.51**	**16.84**	**44.11**	**26.78**	**25.69**	**41.02**	**24.91**	**24.08**	**49.03**	**29.80**	**31.62**	**45.23**	**25.90**	**25.85**
ViT-B	224	2D only	SupImg	-	34.40	19.24	19.06	28.42	14.05	14.97	31.45	20.63	19.41	29.77	18.73	17.82	33.56	23.19	21.81	31.68	19.52	18.11
			DINO	-	33.03	18.62	17.91	26.82	14.56	14.73	27.70	16.24	15.87	25.78	14.86	14.76	32.57	22.13	20.61	29.86	18.07	17.66
			MAE	-	35.92	19.30	19.24	29.88	16.01	15.82	31.54	20.59	19.33	29.92	18.65	17.83	36.97	24.51	23.12	33.67	20.15	19.46
		pseudo	*Ours (pseudo)*	CAD	36.24	19.78	19.72	30.10	15.94	16.05	31.78	20.74	19.46	30.01	19.07	17.94	-	-	-	-	-	-
		2D-3D	*Ours*	CAD	36.31	19.91	19.94	30.30	16.16	16.27	32.02	21.04	19.67	30.16	19.02	18.09	37.74	24.92	23.42	34.13	20.49	19.89

Conclusion

- Learning geometric-aware 2D representaion via CAD models
- Competitive performance to methods that use 3D scenes
- Can be trained on synthetic data

Thank you for listening!

Please visit GeoAware2dRepUsingCAD.github.io for a full paper