Contrastive Lift:

3D Object Instance Segmentation by Slow-Fast Contrastive Fusion

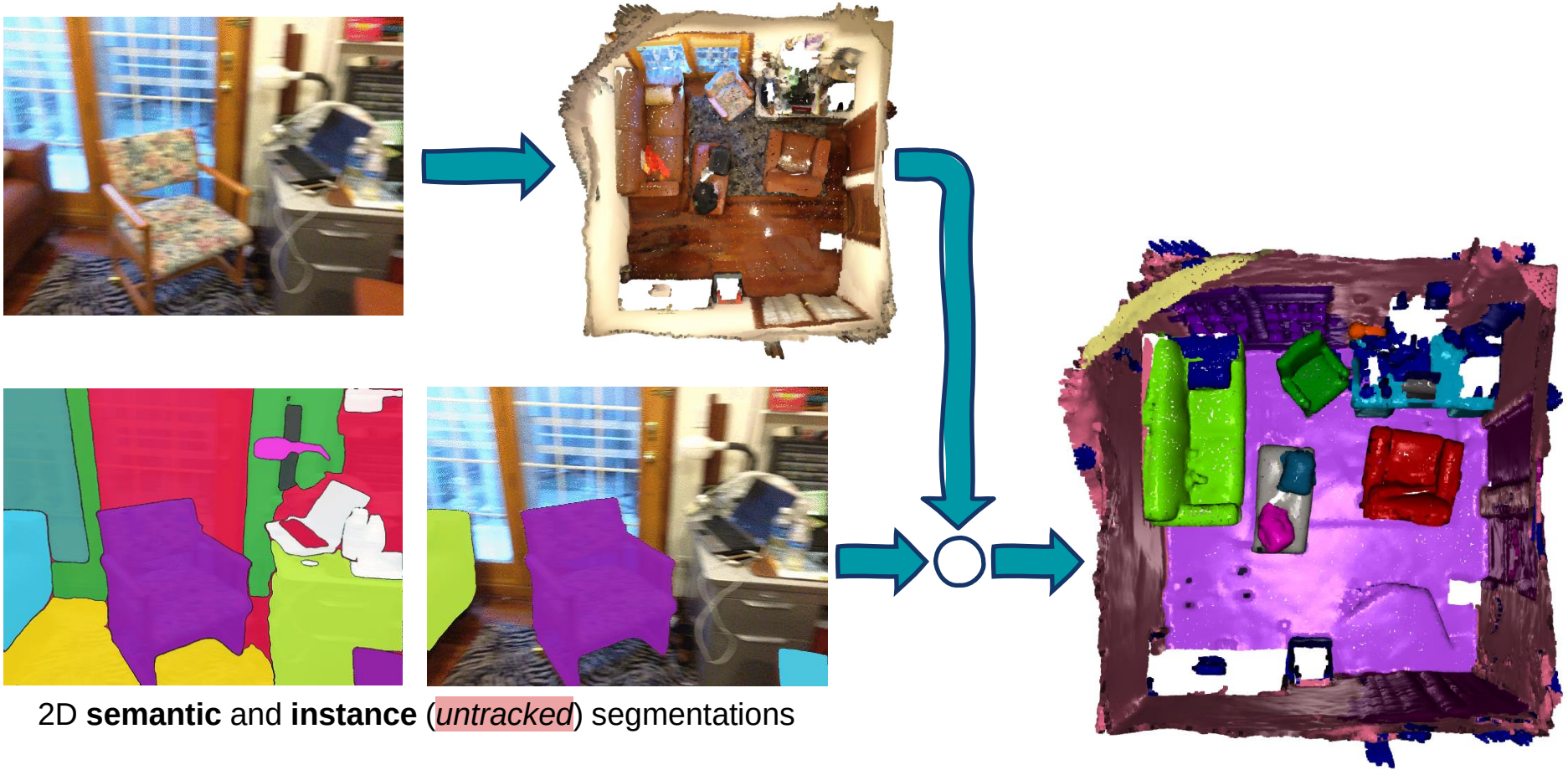
NeurIPS 2023 (**Spotlight**)

Yash Bhalgat, *Iro Laina, João F. Henriques, Andrew Zisserman, Andrea Vedaldi*

Visual Geometry Group, University of Oxford

3D Instance Segmentation problem

2D-to-3D lifting of 2D Segmentations



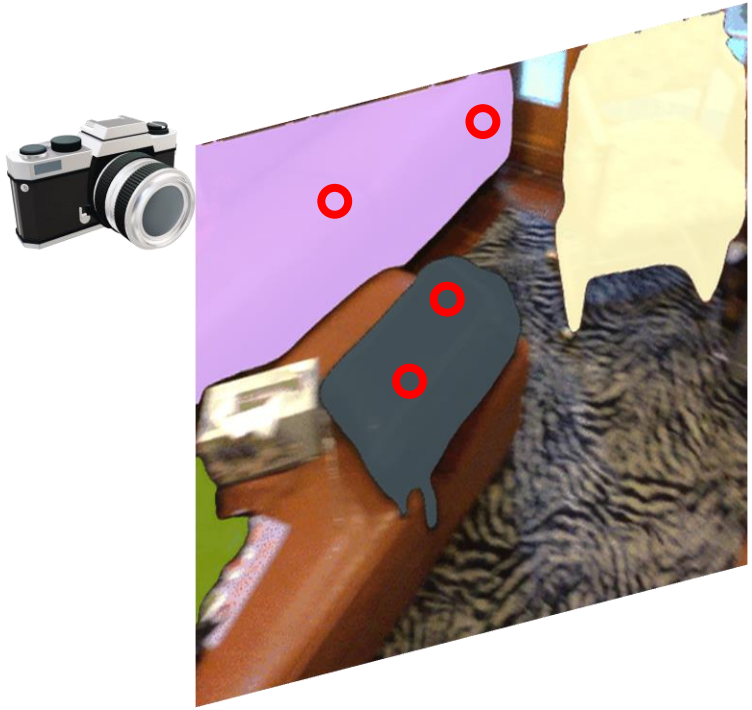
2D **semantic** and **instance** (*untracked*) segmentations

Challenge → multi-view consistency

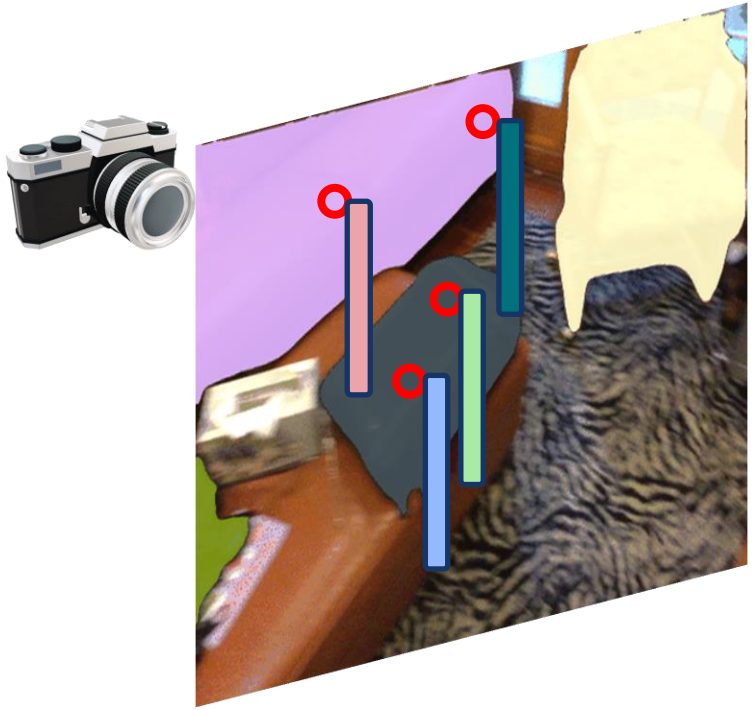
Semantic segments
(*ideally* Multi-view consistent)

Instance segments
(**NOT** Multi-view consistent)

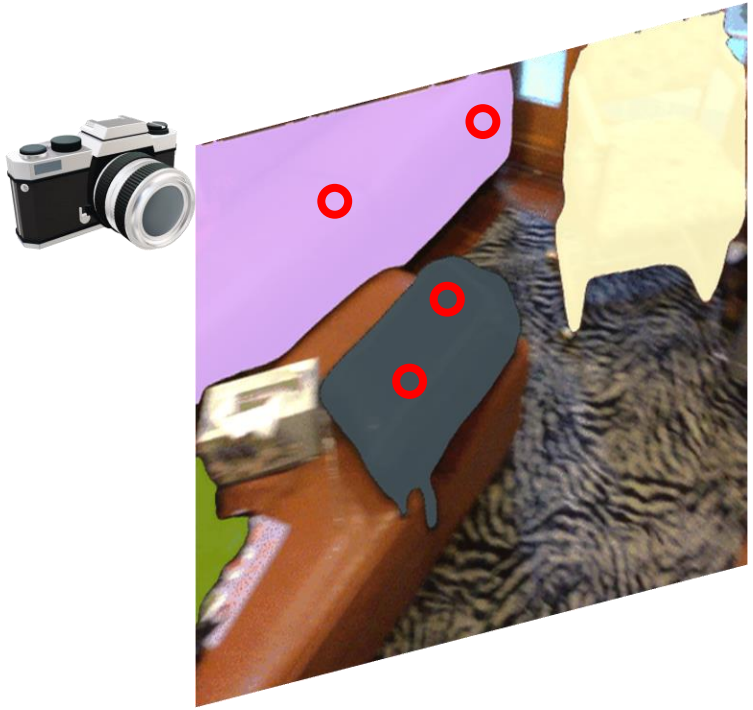
Contrastive learning of embeddings

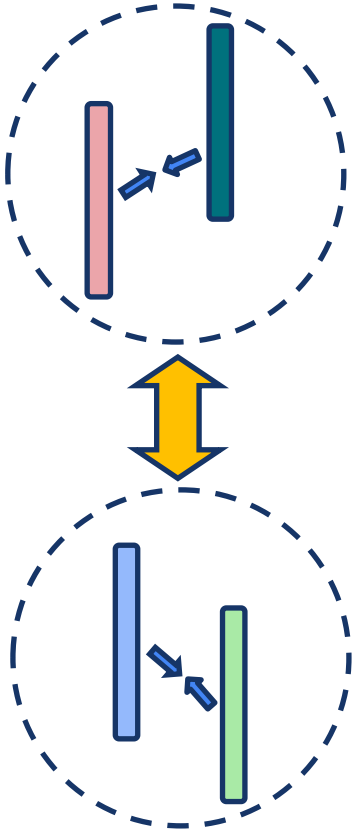


Contrastive learning of pixel embeddings



Contrastive learning of pixel embeddings

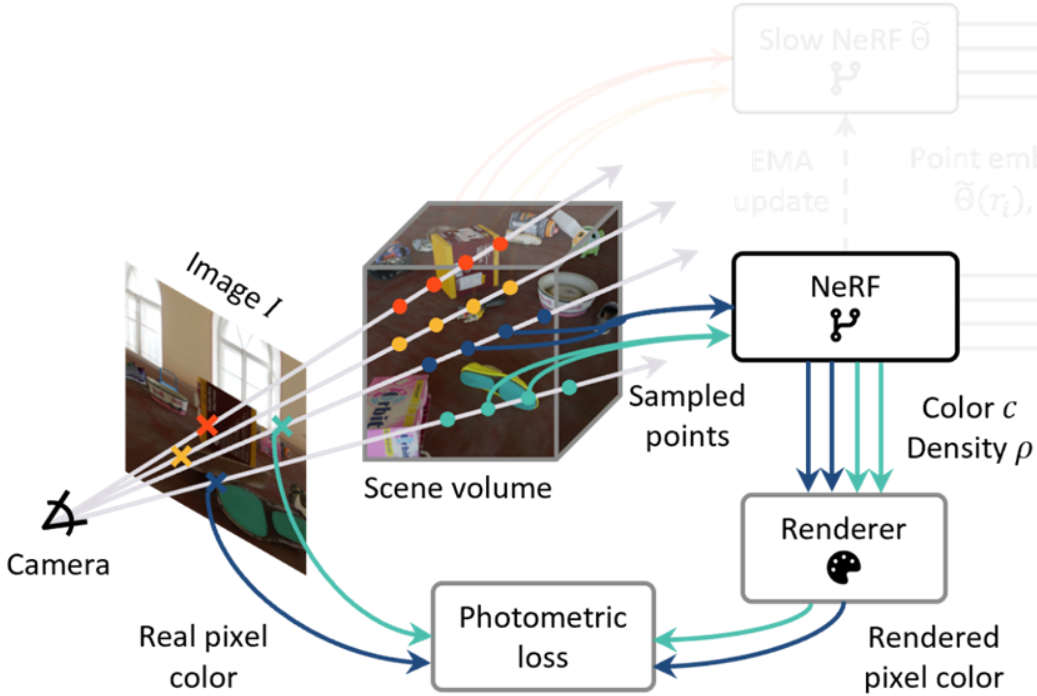




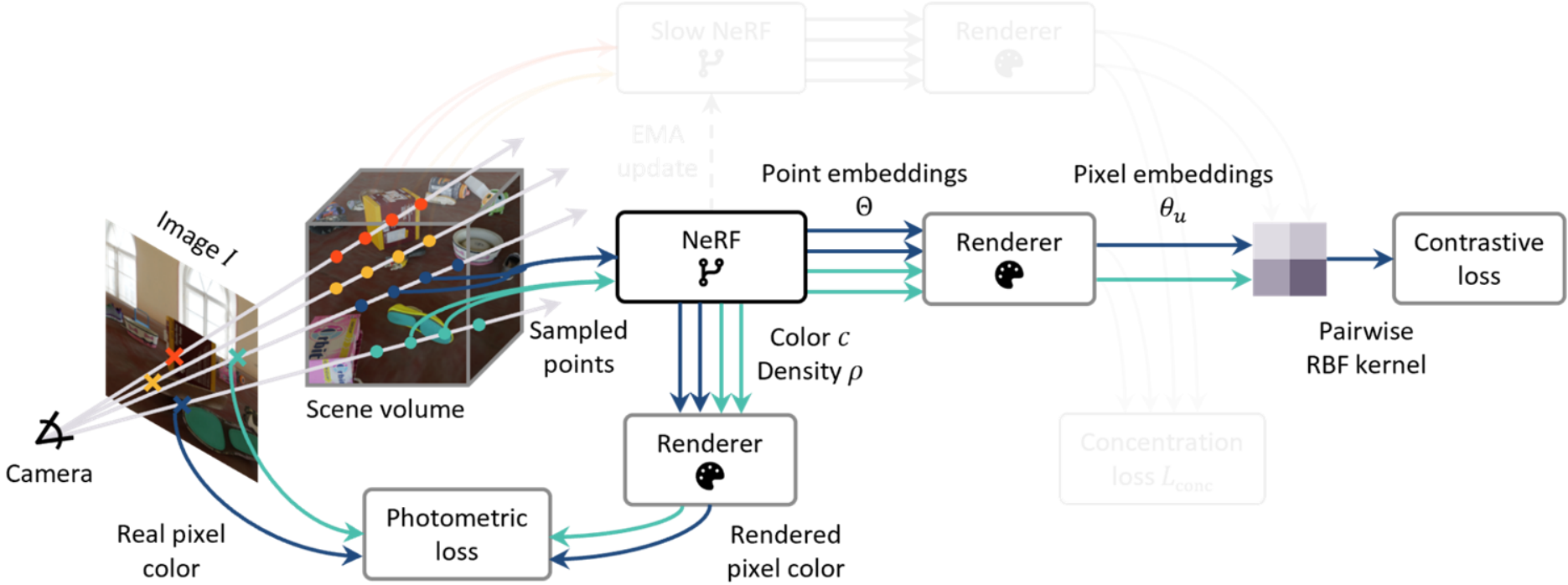
Grounding object embeddings in 3D

Grounding object embeddings in 3D

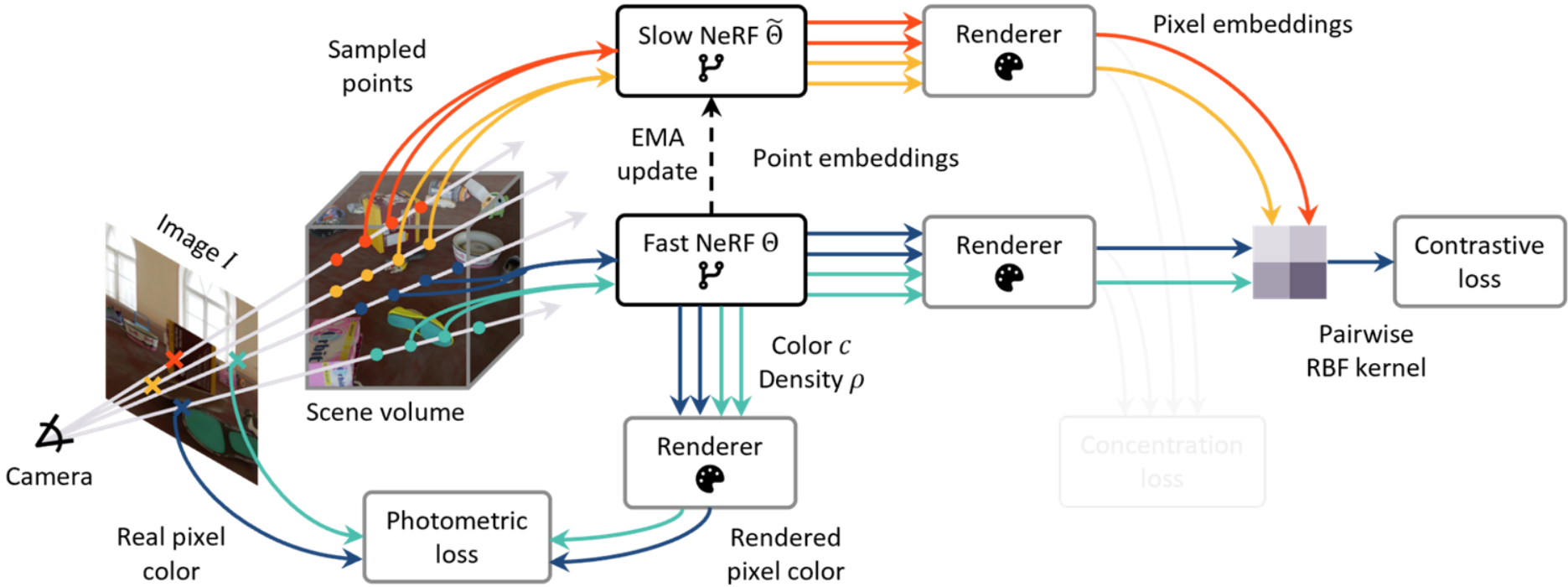
Contrastive Lift framework (***simple*** *version*)



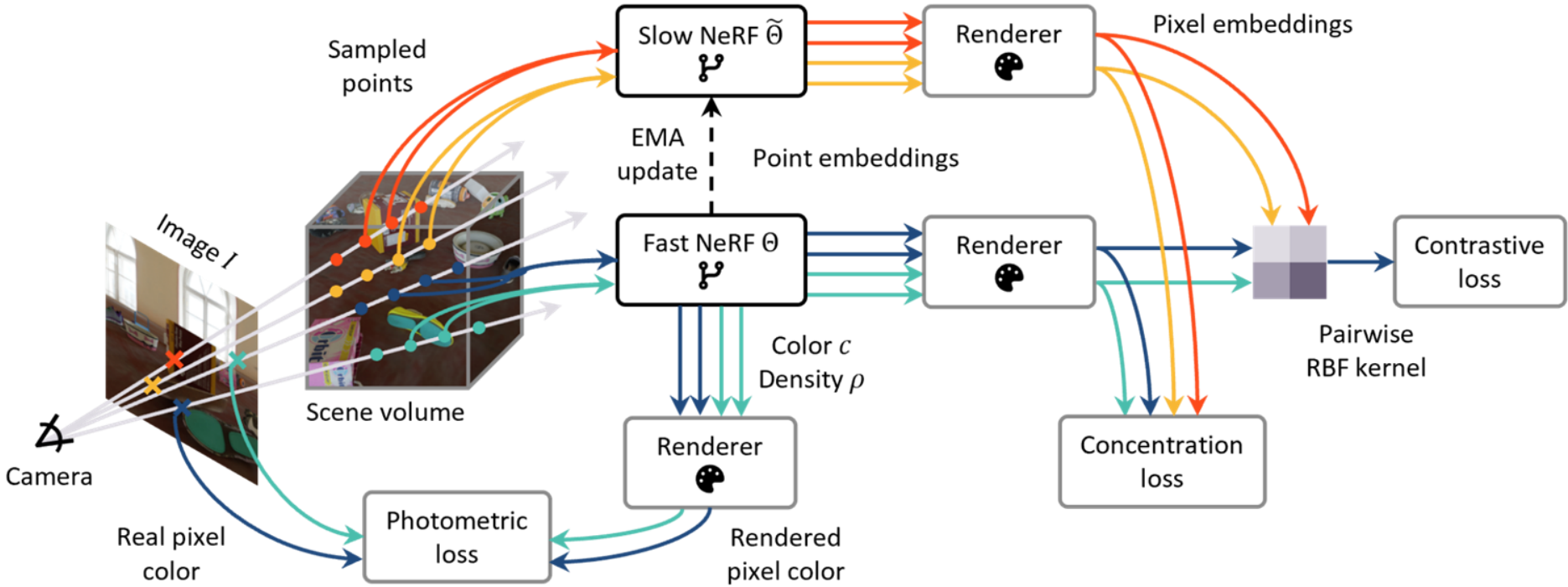
Contrastive Lift framework (***simple*** *version*)



Contrastive Lift framework (***slow-fast*** *version*)

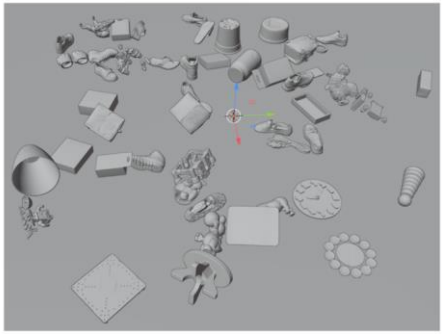


Contrastive Lift framework (***slow-fast*** *version*)

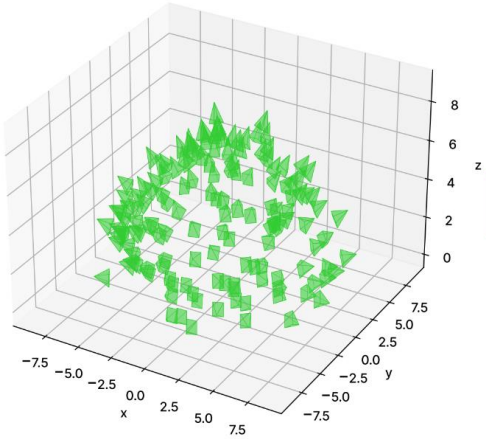


Messy Rooms dataset

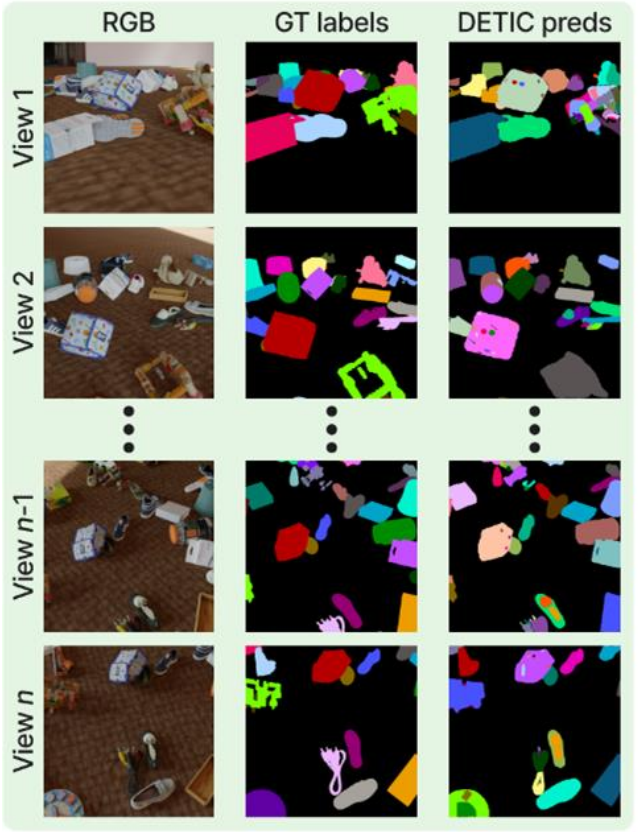
Semi-realistic dataset created using Kubric.
Features scenes with up to 500 objects per scene.



Physically realistic static 3D scene with N objects from Google Standard Objects



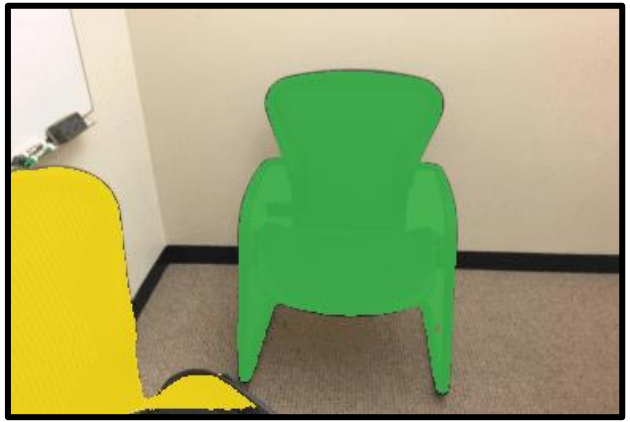
M camera viewpoints sampled in a dome-shaped shell



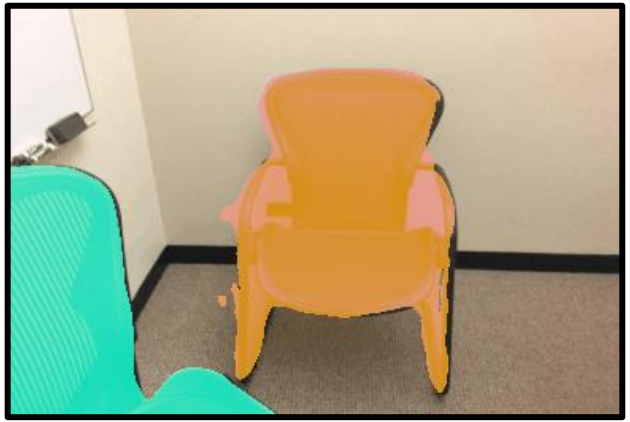
GT RGB, GT instance IDs, and instance segments from **Detic**

Dataset link: https://figshare.com/s/b195ce8bd8eafe79762b

Qualitative results – ScanNet



Mask2Former 2D predictions
(Untracked)



Panoptic Lifting
(CVPR ‘23)

Ours

Qualitative results – ScanNet

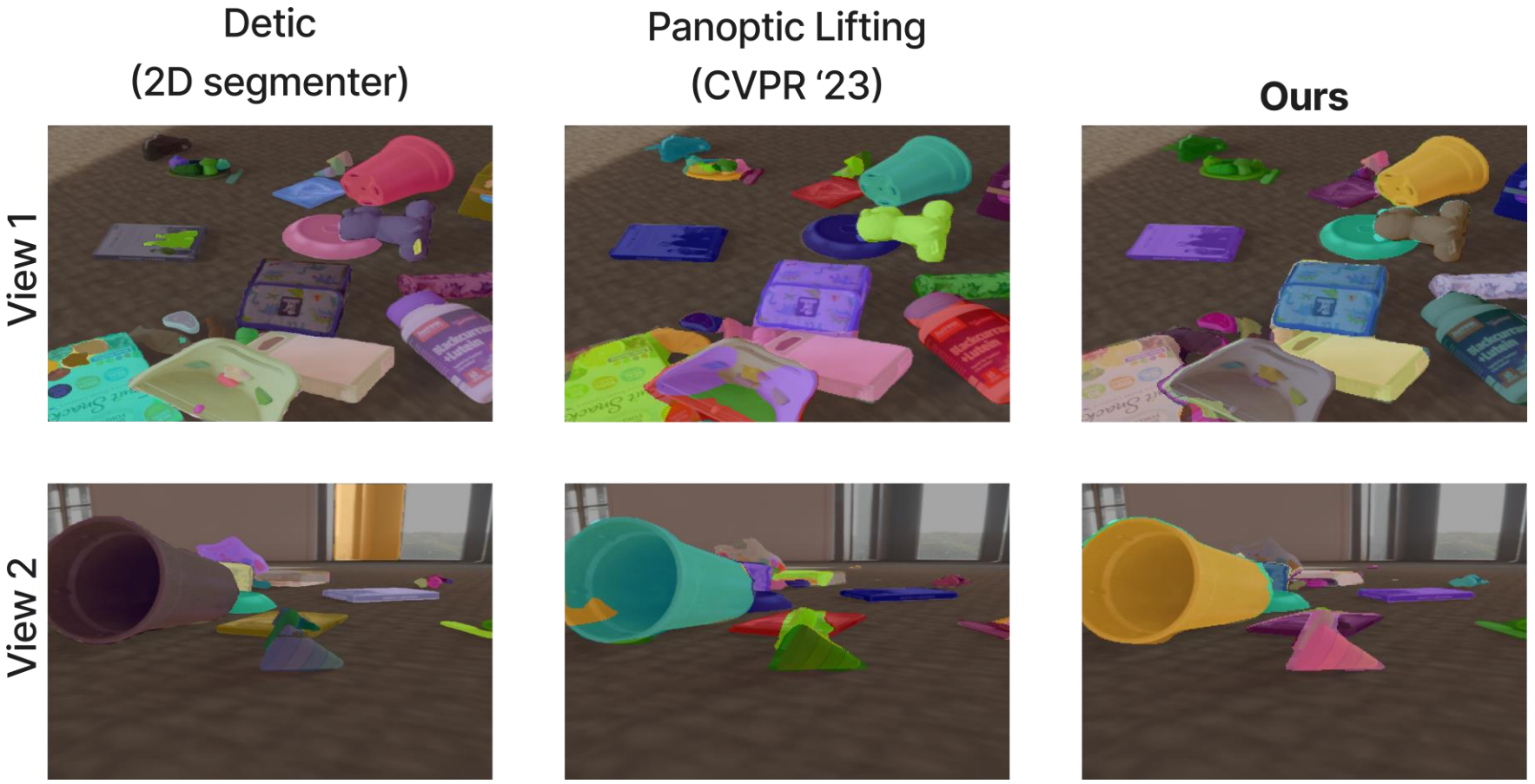
Panoptic Lifting (CVPR ‘23)

✕ Chairs confused as same object

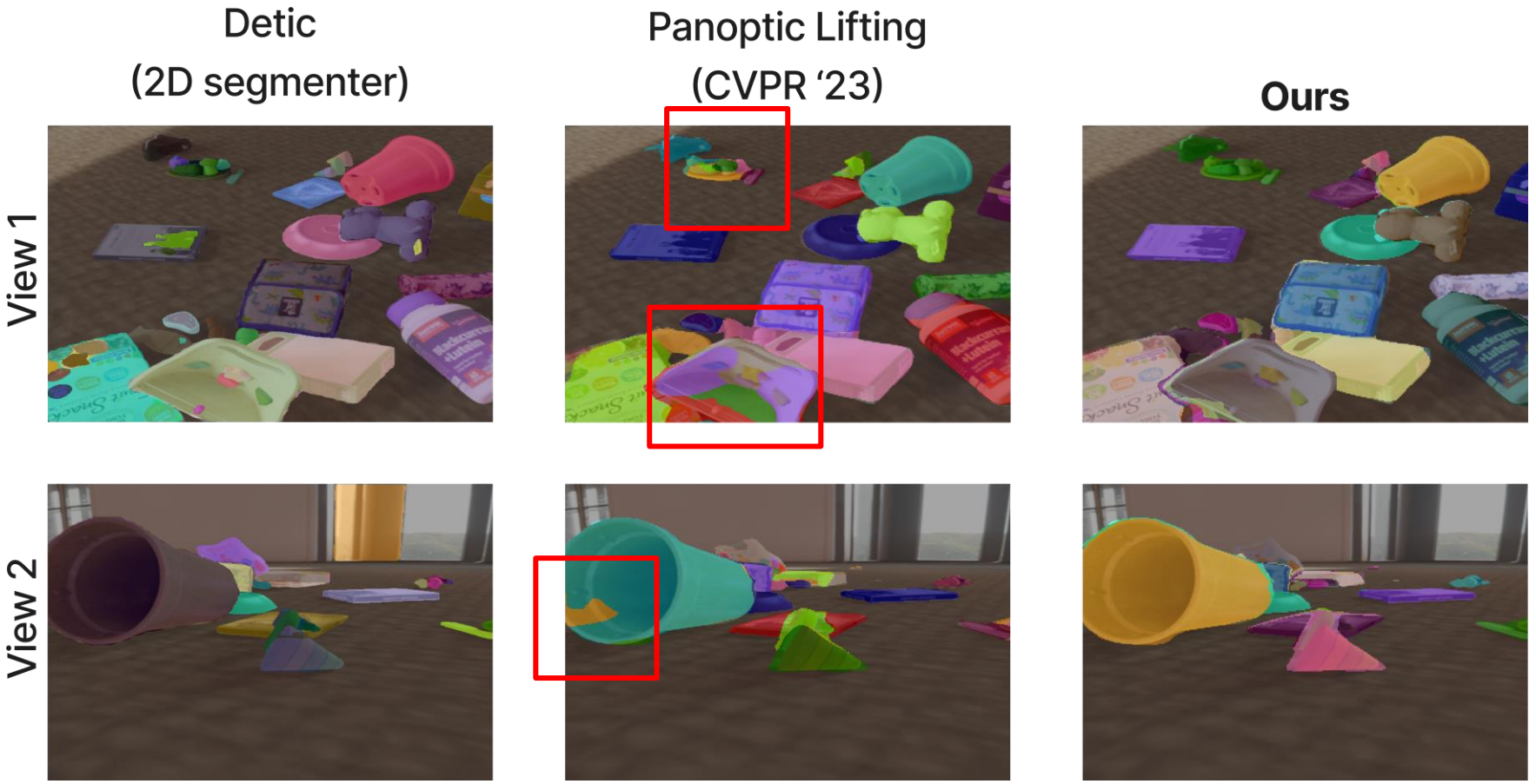
Ours

☑ Chairs segmented as distinct objects

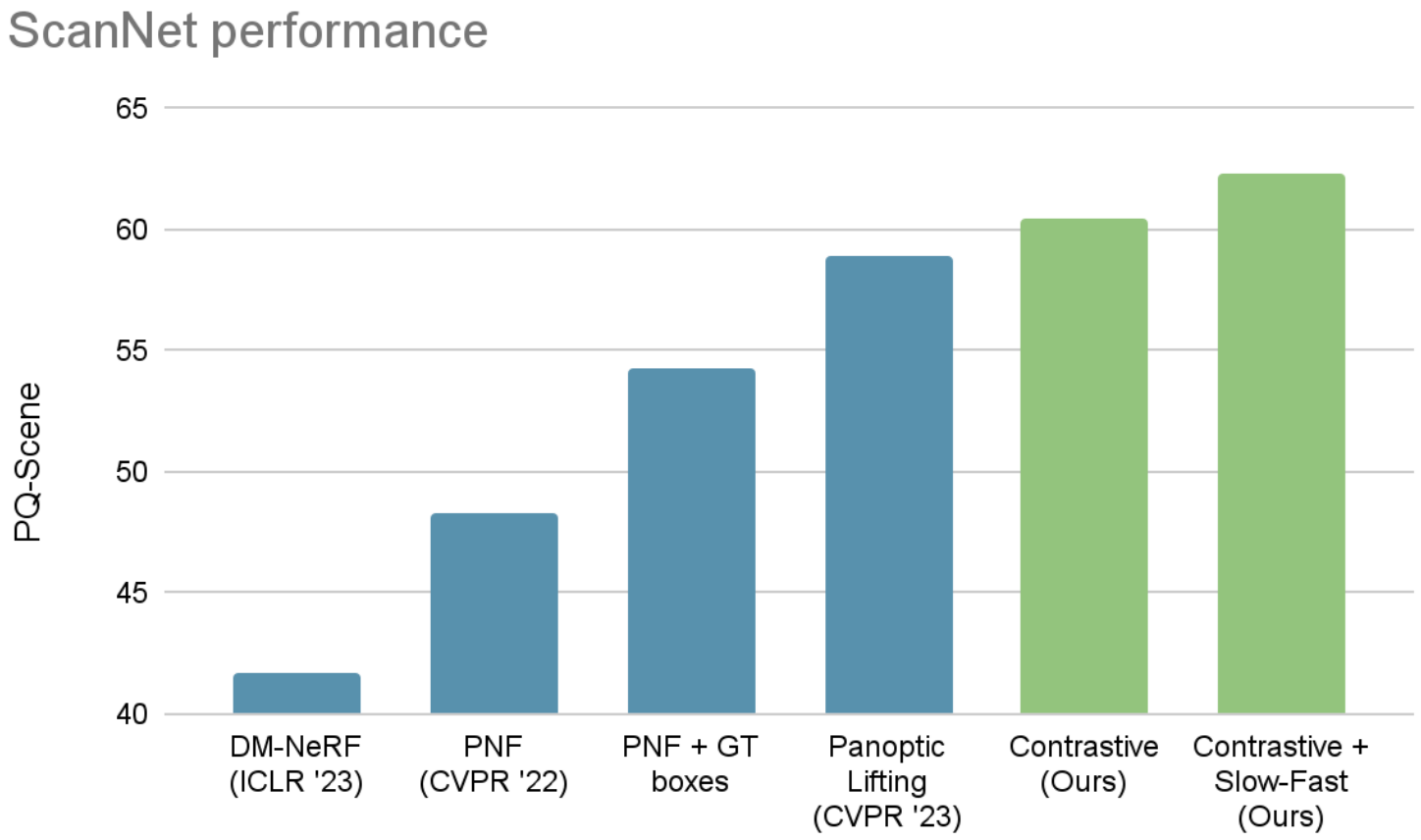
Qualitative results – Messy Rooms dataset



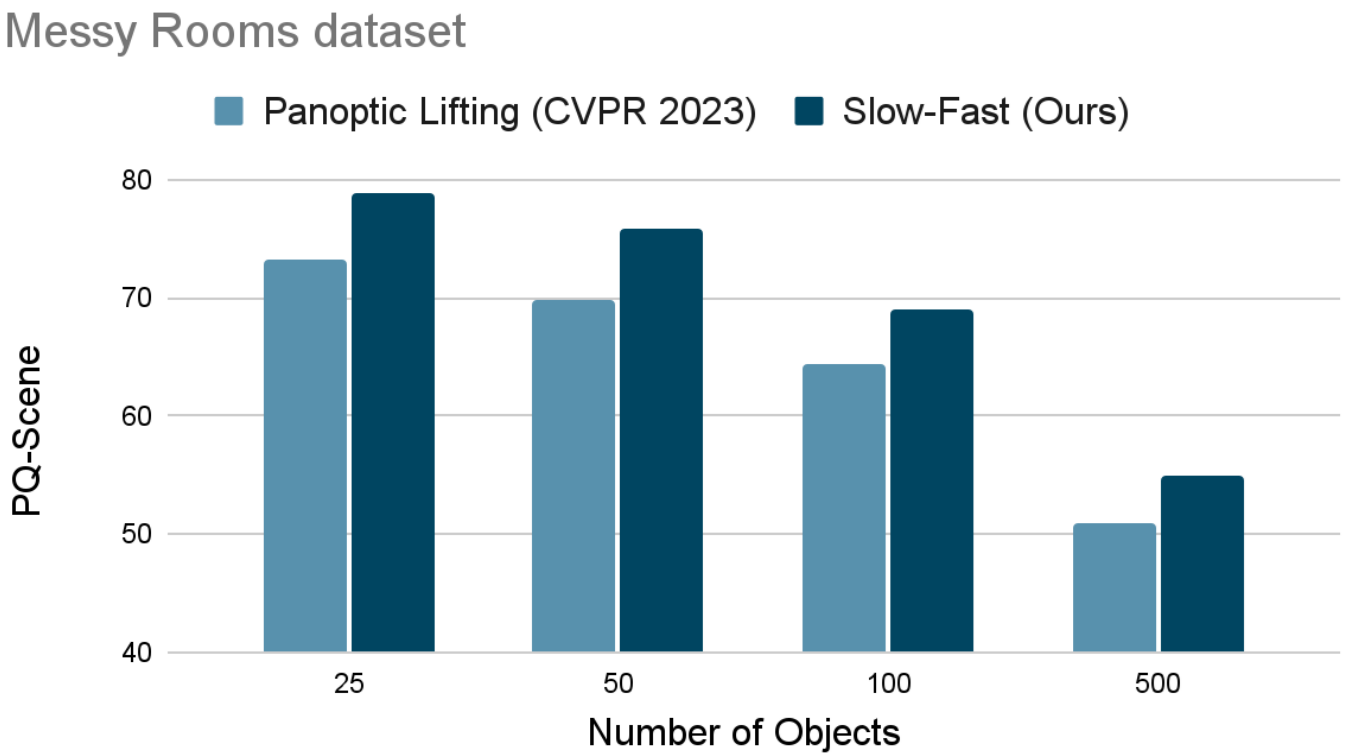
Qualitative results – Messy Rooms dataset



Comparison with recent works

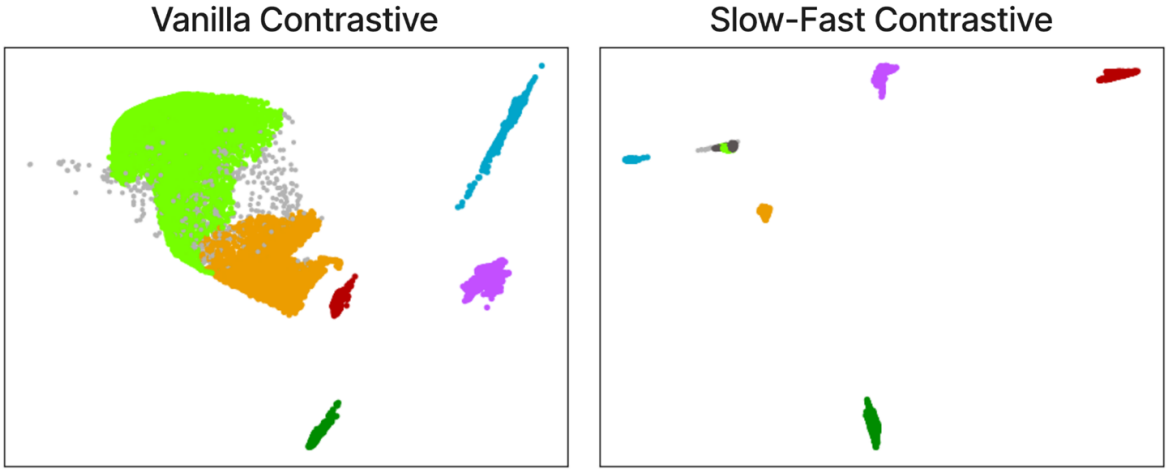


Performance (Scalability)

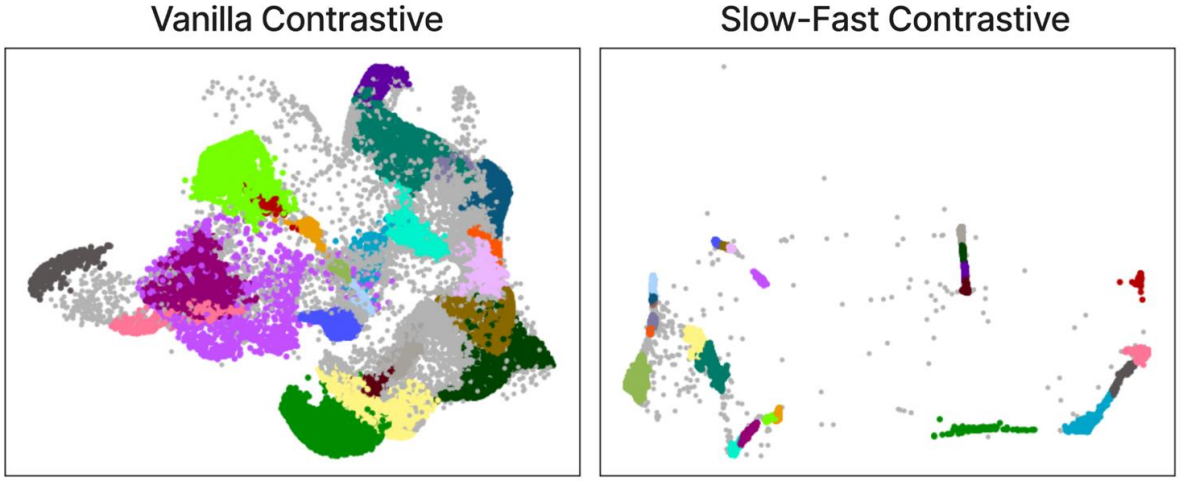


vanilla v/s ***slow-fast***

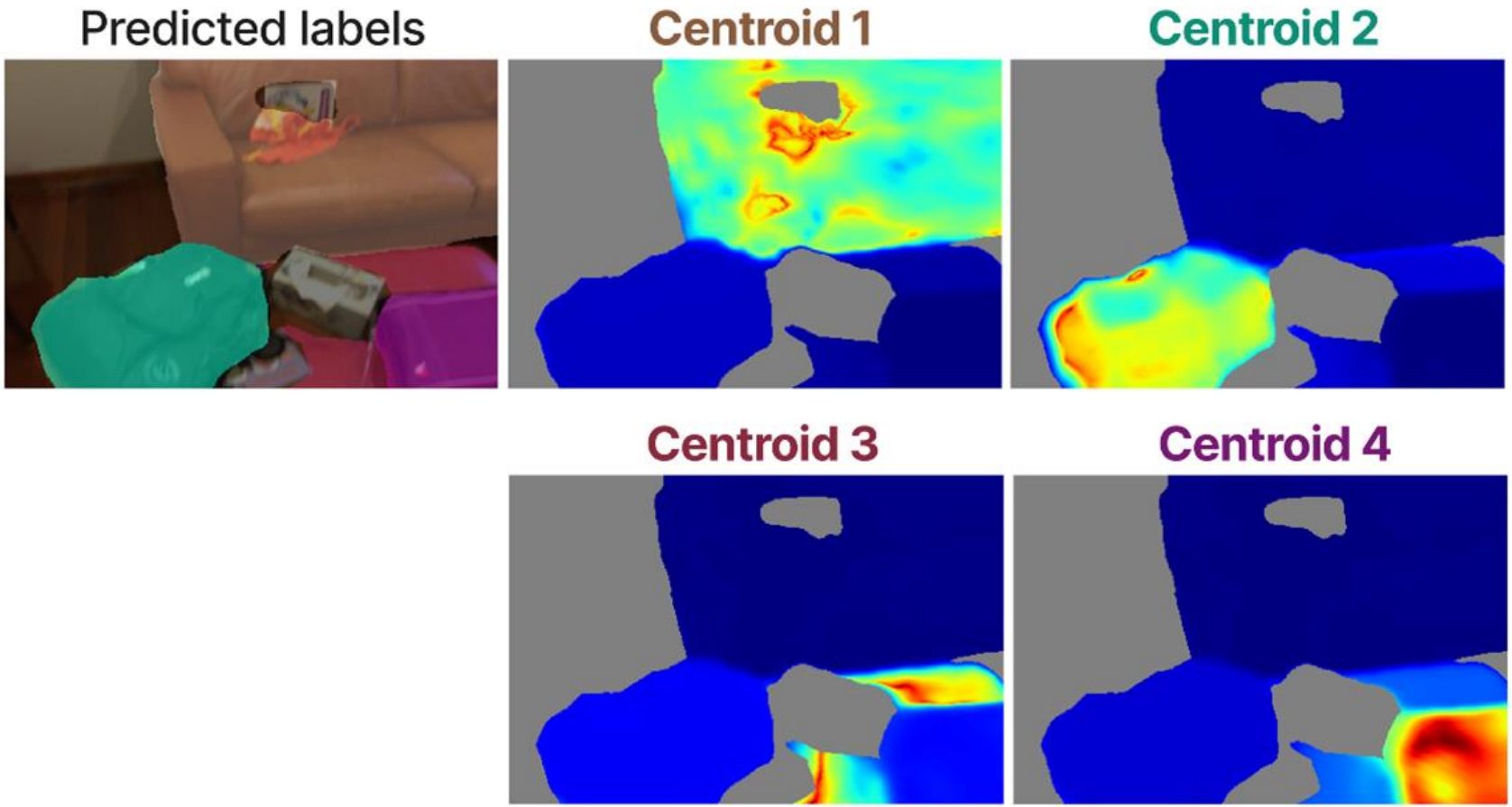
7 objects



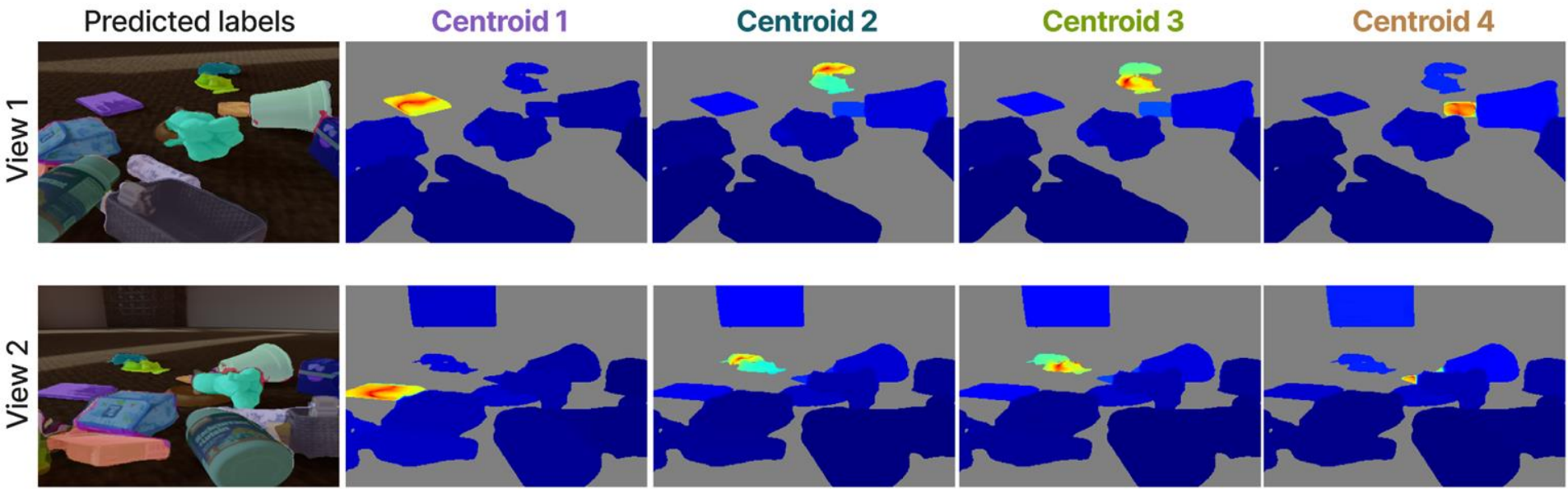
25 objects



Visualization of centroids



Visualization of centroids



Objects or clusters are well-separated on a 3D level, even under heavy occlusion.

Summary

- ✅ Novel method to lift 2D predictions to 3D for instance segmentation
  - Embeddings “grounded in 3D” → Knowledge of number of objects NOT required
  - More efficient and scalable than Hungarian Matching based methods
- ✅ As a byproduct → leads to tracked and higher quality 2D instance segmentations
- ✅ New “Messy Rooms” benchmark for scalable 3D instance segmentation
- ❌ Only works on static scenes
- ❌ Relies on accurate geometry reconstruction

Thank you 🙏

Poster Location: Great Hall & Hall B1+B2 #325

Date: 13 December, 5PM

Webpage: https://www.robots.ox.ac.uk/~vgg/research/contrastive-lift