Sparse Multi-Object Render-and-Compare

Florian Langer, Ignas Budvytis, Roberto Cipolla

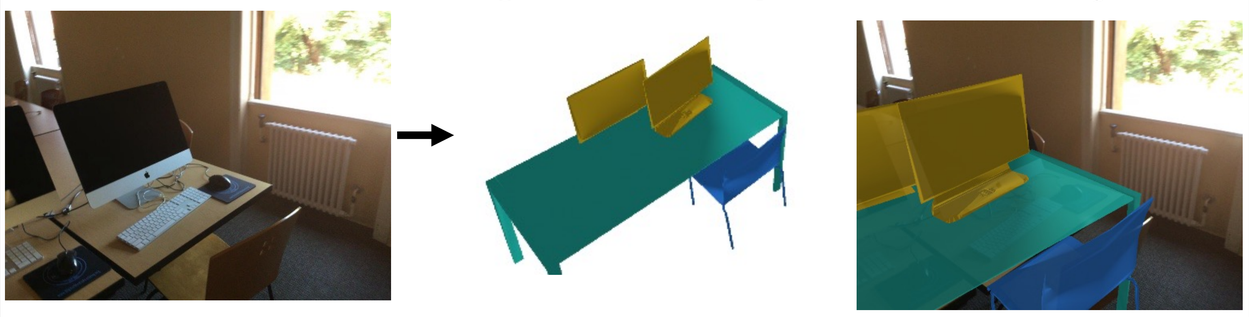
Department of Engineering, University of Cambridge

BMVC 2023

Problem

From a single RGB image predict aligned 3D shapes that represent the given scene.

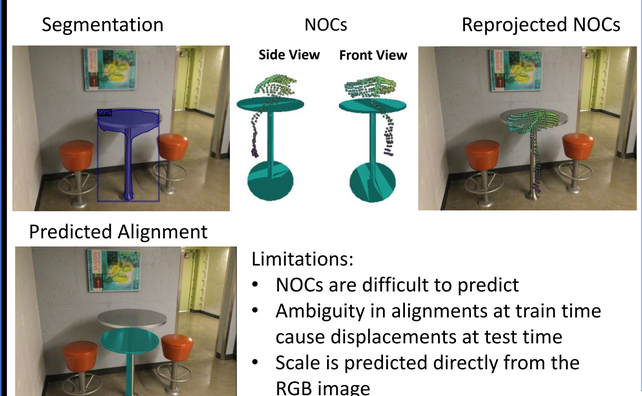
RGB Image → **Aligned 3D Shapes** → **Overlays**



Related Work

Normalised Object Coordinates

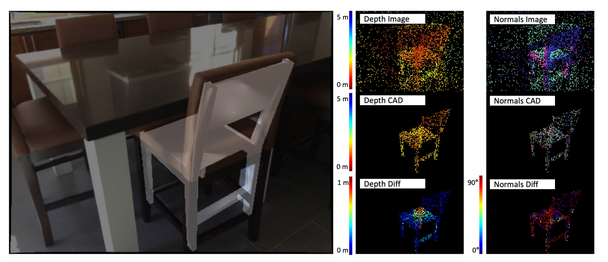
ROCA [1]



Limitations:
- NOCs are difficult to predict
- Ambiguity in alignments at train time cause displacements at test time
- Scale is predicted directly from the RGB image

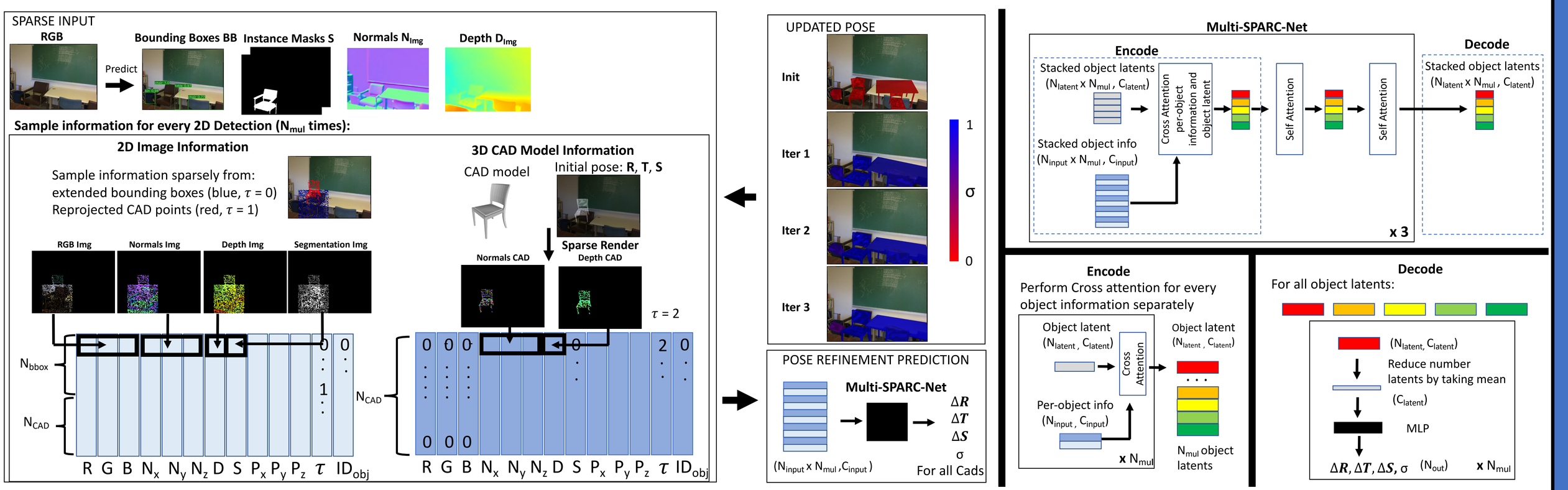
Render-and-Compare

SPARC [2]

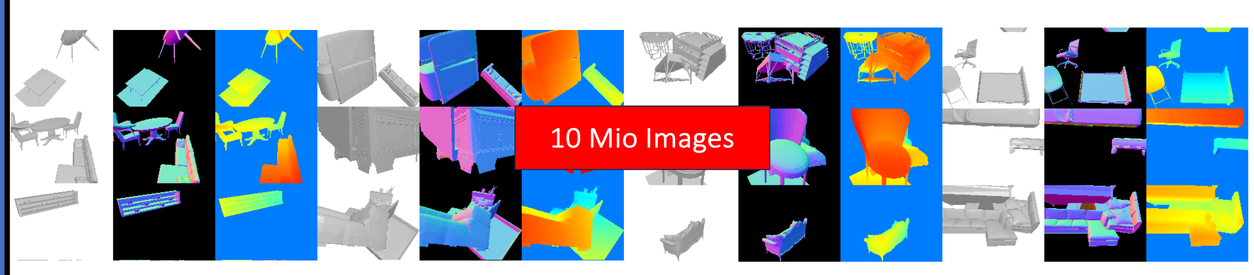


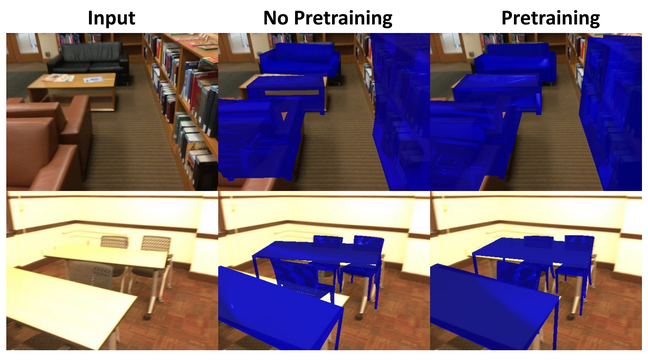
Limitations:
- CAD alignments are predicted individually per-object leading to slow run-times and sub-optimal alignments

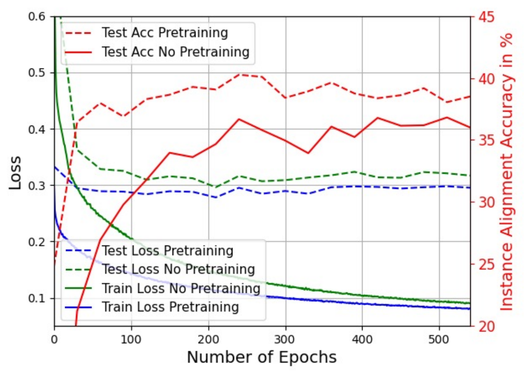
Method



Synthetic Pretraining







Qualitative Results - ScanNet

Quantitative Results - ScanNet

	Ablations								Competitors				
	Joint Encoding and Decoding		Single vs. Multi Pre-training vs. no Pre-training				Learned Classification		Total3D-ODN [3]	Mask2CAD-b5 [4]	ROCA [1]	SPARC [2]	Ours
	sep. enc. / joint dec	joint enc. / sep dec	single no pre	multi no pre	single pre	multi pre	2D conf-idence	3D class-ification					
Class [%]	20.7	24.1	24.8	26.9	30.2	30.3	30.1	30.3	8.5	10.6	18.4	24.9	**30.3**
Instance [%]	27.9	31.9	34.6	36.7	38.7	40.3	38.8	40.3	10.4	16.7	25.0	31.8	**40.3**
Time [ms]	864	656	2320	864	2320	864	816	864	-	60	**53**	1925	864

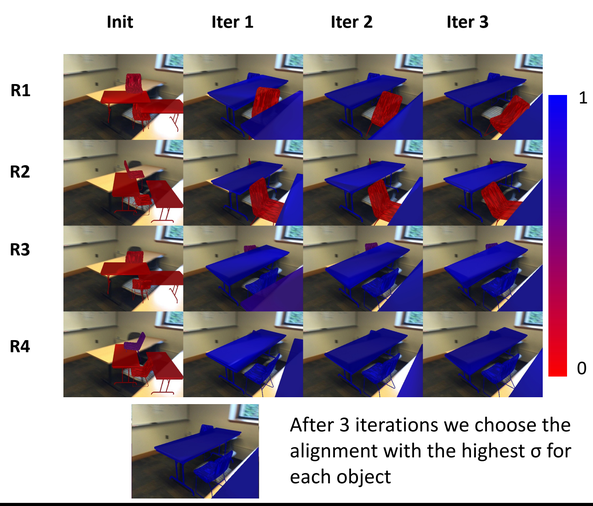
3D Classification Score

σ = 1 if delta T < 20 cm, delta S < 20% and delta R < 20 degree; 0 else → Learn via Binary Cross Entropy loss

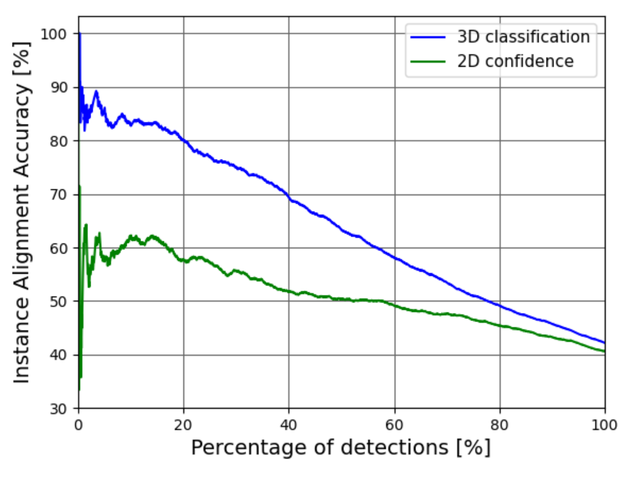
Learned classification score allows to ...

... choose the **best alignment from different initialisations**

... is **better calibrated** than 2D confidence score



After 3 iterations we choose the alignment with the highest σ for each object



Discussion

+ **Joint CAD model alignments** are more precise and faster than individual predictions.
+ **Synthetic pre-training** leads to significant improvements despite domain gap
+ **Learned 3D classification score** is more accurate and better calibrated than 2D detection scores
− Similar to SPARC our network does not seem able to make full use of all available information (particularly precise normal estimates) and does not improve when using more than 3 refinement steps.

References

[1] Gümeli, C., Dai, A., Nießner, M. *ROCA: Robust CAD Model Retrieval and Alignment from a Single Image.* CVPR (2022)

[2] Langer, F., Bae, G., Budvytis, I., Cipolla, R. *SPARC: Sparse Render-and-Compare for CAD model alignment in a single RGB image.* BMVC (2022)

[3] Nie. Y, Han, X., Guo, S., Zheng, Y., Chang, J., Zhang, J. *Total3DUnderstanding: Joint Layout, Object Pose and Mesh Reconstruction for Indoor Scenes from a Single Image.* CVPR (2020)

[4] Kuo, W., Angelova, A., Lin, T., Dai, A. *Mask2CAD: 3D Shape Prediction by Learning to Segment and Retrieve.* ECCV (2020)