SplatTouch: Explicit 3D Representation Binding Vision and Touch

Antonio Luigi Stefani, Niccolò Bisagno, Nicola Conci, Francesco de Natale

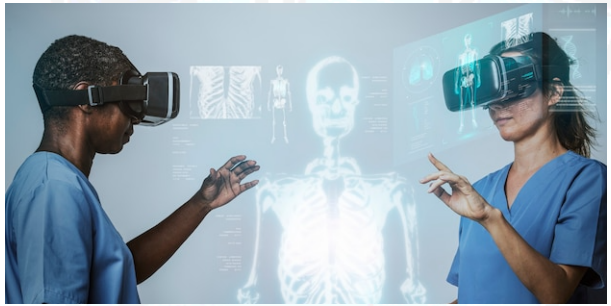
UNIVERSITY OF TRENTO

mm lab

CVPR Nashville JUNE 11-15, 2025

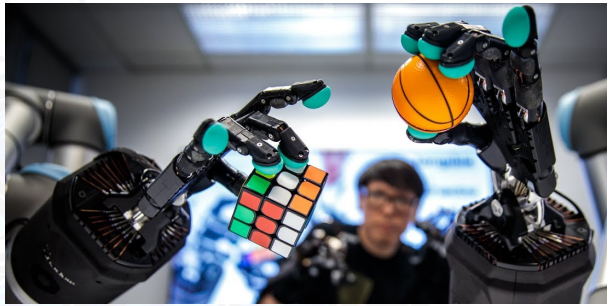
Multimodal VR Applications

Telemedicine



Rehabilitation

Remote object manipulation

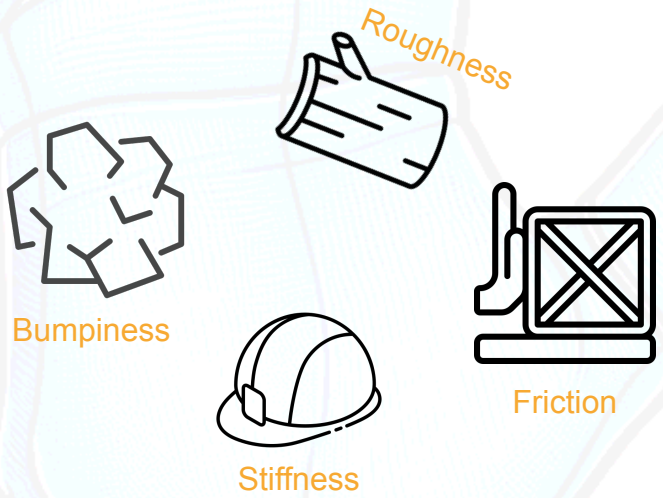


Sport training

Safety training

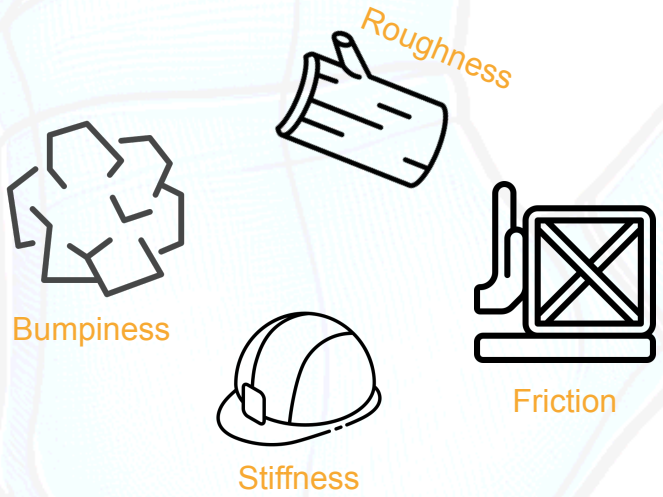
Model haptic properties

Adding haptic stimuli to digital environments means modeling the following properties:



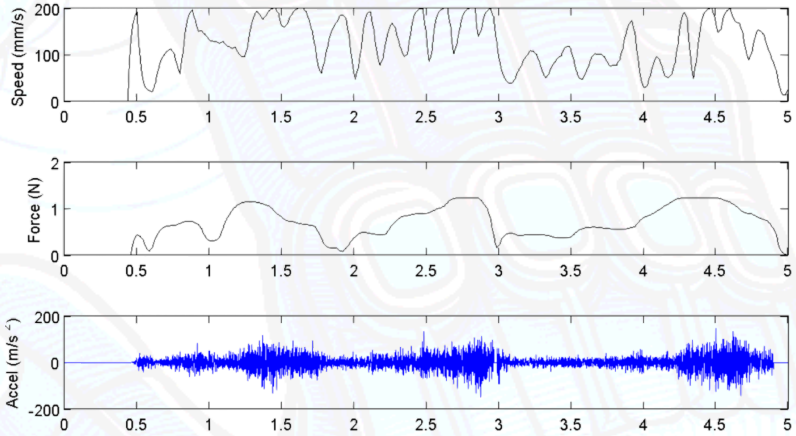
Model haptic properties

Adding haptic stimuli to digital environments means modeling the following properties:



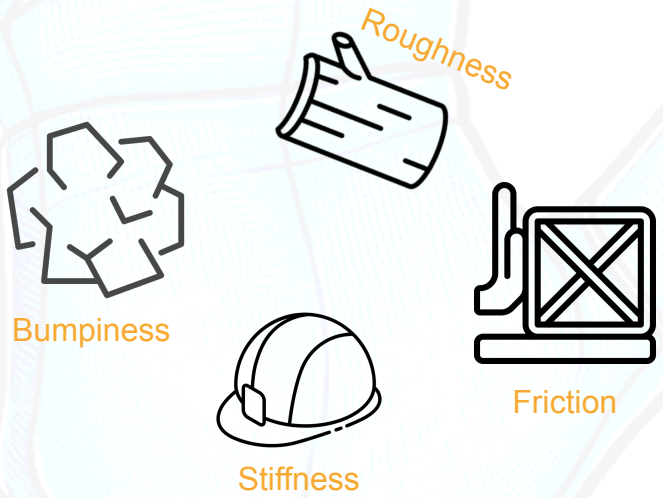
Haptic properties can be sensed by either:

- Mono-dimensional signals



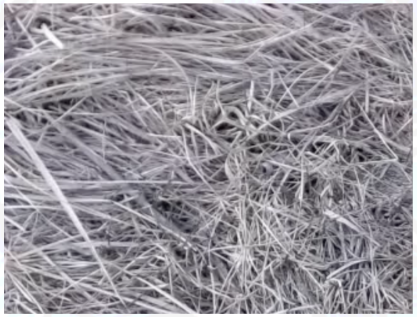
Model haptic properties

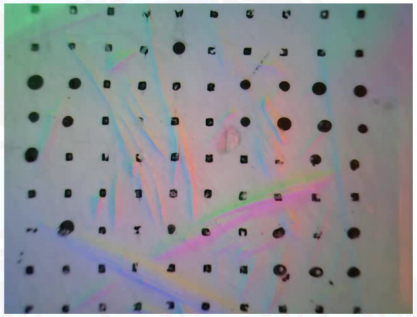
Adding haptic stimuli to digital environments means modeling the following properties:



Haptic properties can be sensed by either:

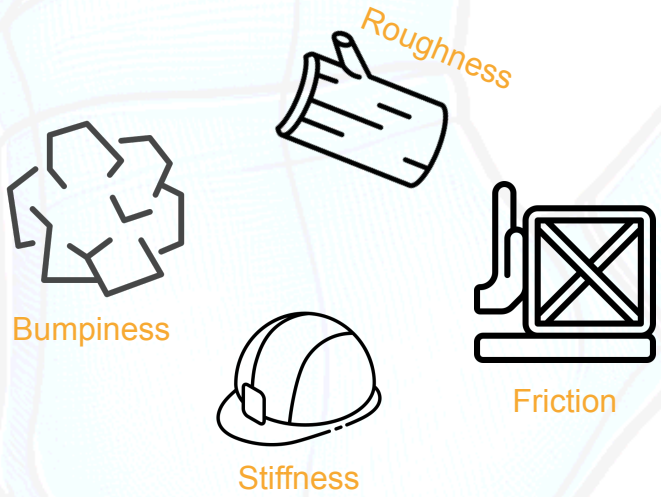
- Mono-dimensional signals
- Vision-based data





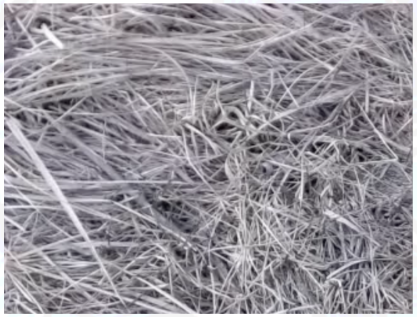
Model haptic properties

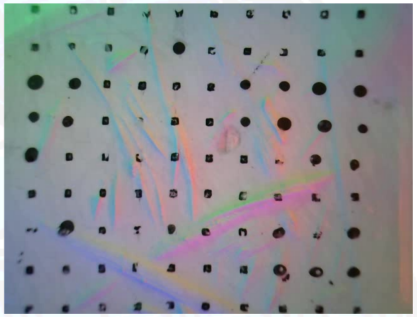
Adding haptic stimuli to digital environments means modeling the following properties:



Haptic properties can be sensed by either:

- ~~Mono-dimensional~~ signals
- Vision-based data





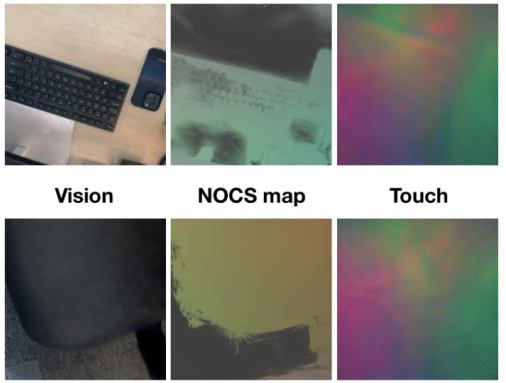
Haptic is still underexplored

Our aim is to generate haptic data grounded in 3D virtual environments to:

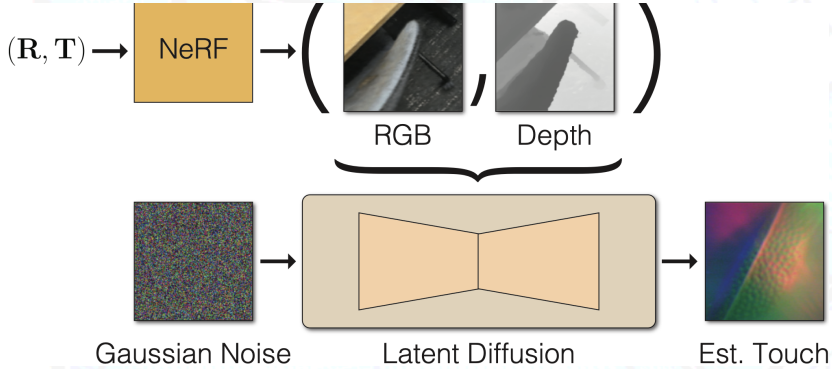
- Increase data availability.
- Improve data quality.
- Develop robust haptic perception models.

Implicit representation

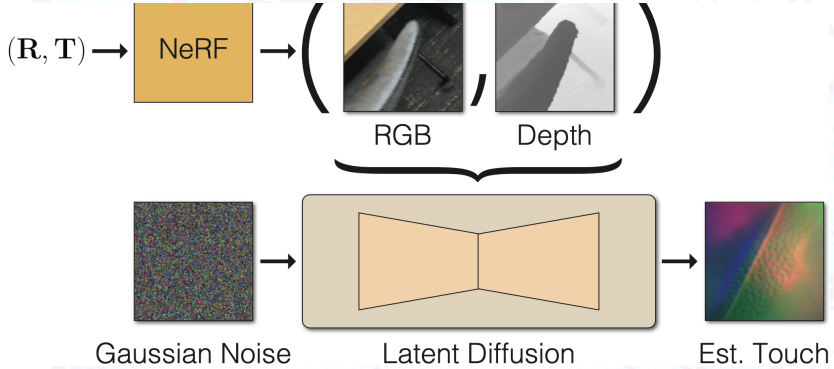
Explicit 3D scene representation



State of the Art: Tactile-Augmented Radiance Field (TaRF)



State of the Art: Tactile-Augmented Radiance Field (TaRF)



Issues:

- Diffusion Models struggles with 3D data → poor scene understanding
- Current 3D contact localization approach retrieves the position of the camera related to haptic maps

Our contribution

Research question:

Is it possible to pass information about 3D scenarios to Diffusion Models?

Our contribution

Research question:

Is it possible to pass information about 3D scenarios to Diffusion Models?

Answer:

Yes, it is possible by mapping 3D information onto images

Our contribution

Research question:

Is it possible to pass information about 3D scenarios to Diffusion Models?

Answer:

Yes, it is possible by mapping 3D information onto images

Our contribution is two-folded:

- Exploiting NOCS maps as a novel global descriptor of the scene

Our contribution

Research question:

Is it possible to pass information about 3D scenarios to Diffusion Models?

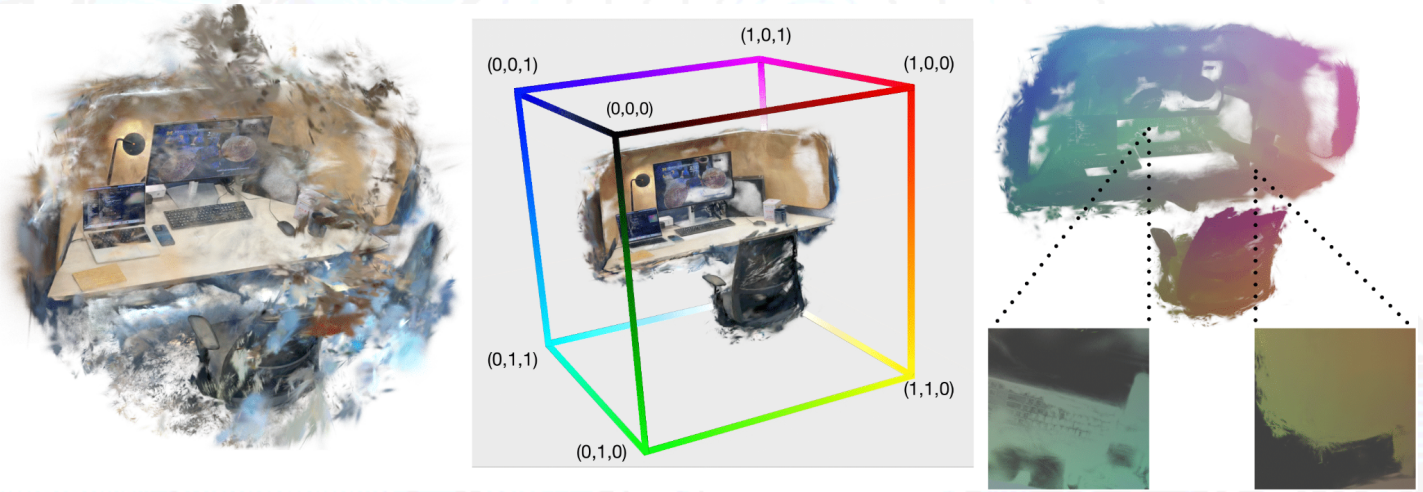
Answer:

Yes, it is possible by mapping 3D information onto images

Our contribution is two-folded:

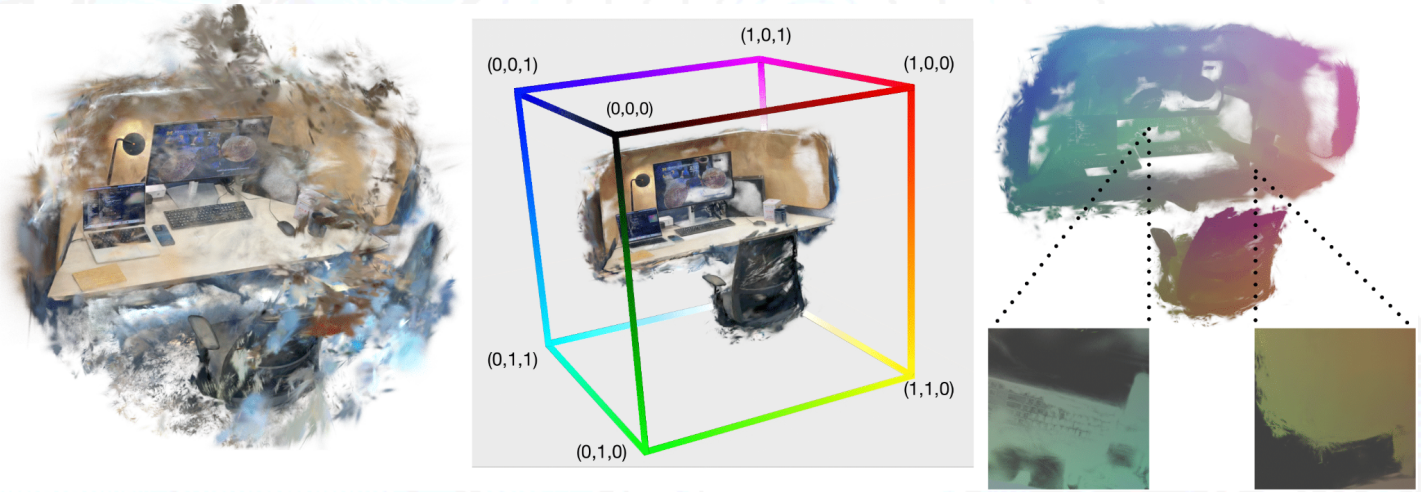
- Exploiting NOCS maps as a novel global descriptor of the scene
- Establishing a novel 3D localization task

Normalized Object Coordinate Space (NOCS)



A normalized coordinate space to estimate object position and size in RGB images, ensuring a uniform, viewpoint-independent representation

Normalized Object Coordinate Space (NOCS)



NeRF is an implicit representation of a 3D scene, which makes it difficult to obtain a NOCS representation. We need to switch to an explicit representation

→ Gaussian Splatting

3D-aware touch generation

TaRF's scene reconstruction using GS and transformation into NOCS

GT GS NOCS

Pairing touch, vision and NOCS

3D-aware touch generation

TaRF's scene reconstruction using GS and transformation into NOCS

GT GS NOCS

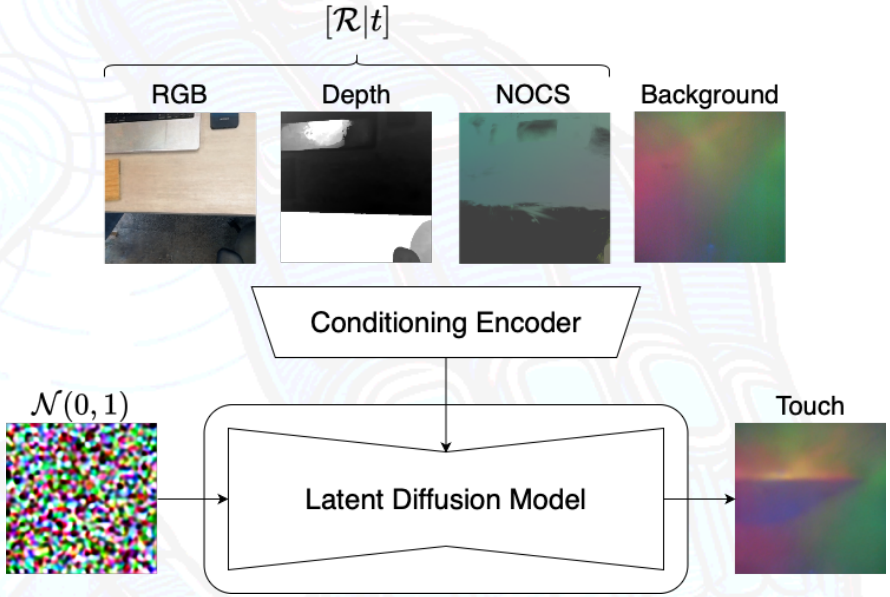
Pairing touch, vision and NOCS

3D-aware touch generation

TaRF's scene reconstruction using GS and transformation into NOCS

Pairing touch, vision and NOCS

Our architecture:



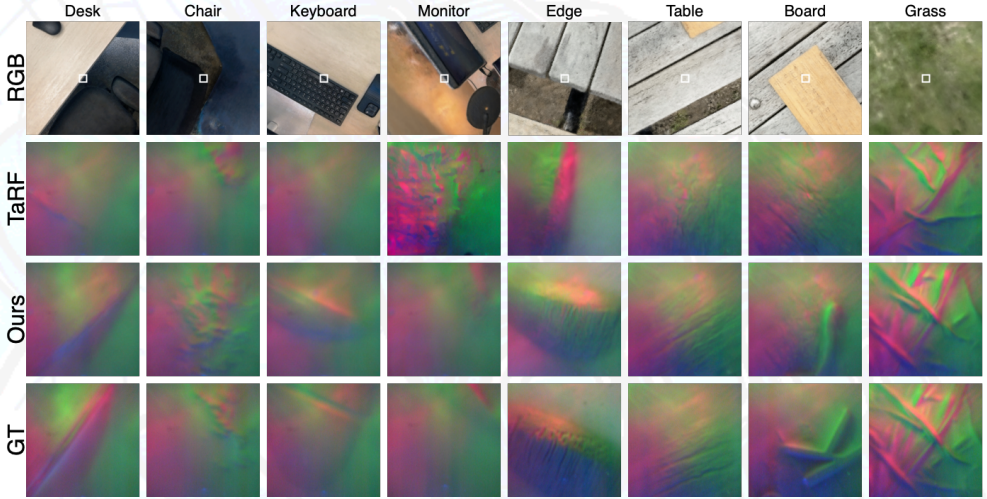
Generation results

Quantitative

Model	PSNR (↑)	SSIM (↑)	FID (↓)
TaRF	22.84	0.72	28.97
TaRF*	23.88	0.76	15.20
Our	**30.19**	**0.84**	**10.06**

TaRF* → TaRF on the single scene

Qualitative



Novel 3D localization task

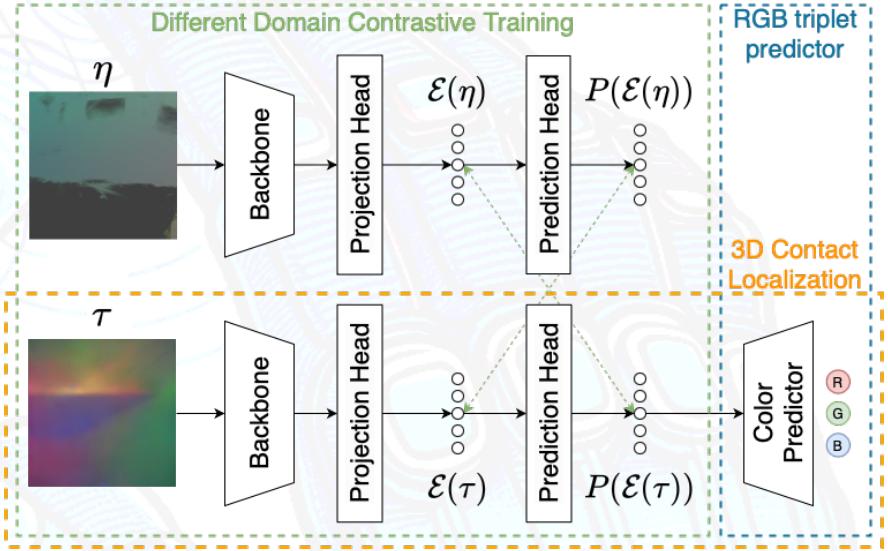
What?

- Estimating the **exact position** in the 3D scenario given a query haptic map

Why?

- The previous approach estimates the **camera's position** for the query haptic map, not its exact location.

How? → SimSiam-based framework

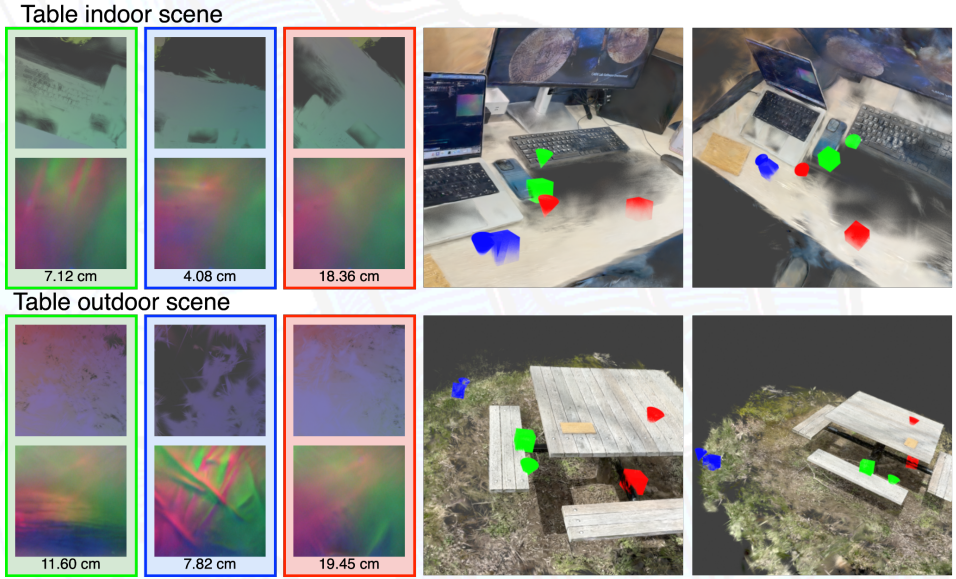


3D Localization results

Quantitative

Model	Training data	Distance (cm) (↓)
Random	-	56.47
RGB+Touch	Real	22.44
NOCS+Touch	Real	13.02
NOCS+Touch	**Real+Aug**	**11.65**

Qualitative



Thanks for the attention

antonioluigi.stefani@unitn.it