Set-the-Scene: Global-Local Training for Generating Controllable NeRF Scenes

TEL AVIV UNIVERSITY אוניברסיטת תל אביב

Dana Cohen-Bar, Elad Richardson, Gal Metzer, Raja Giryes, Daniel Cohen-Or

ICCV23 PARIS

What's this paper about? 🤔

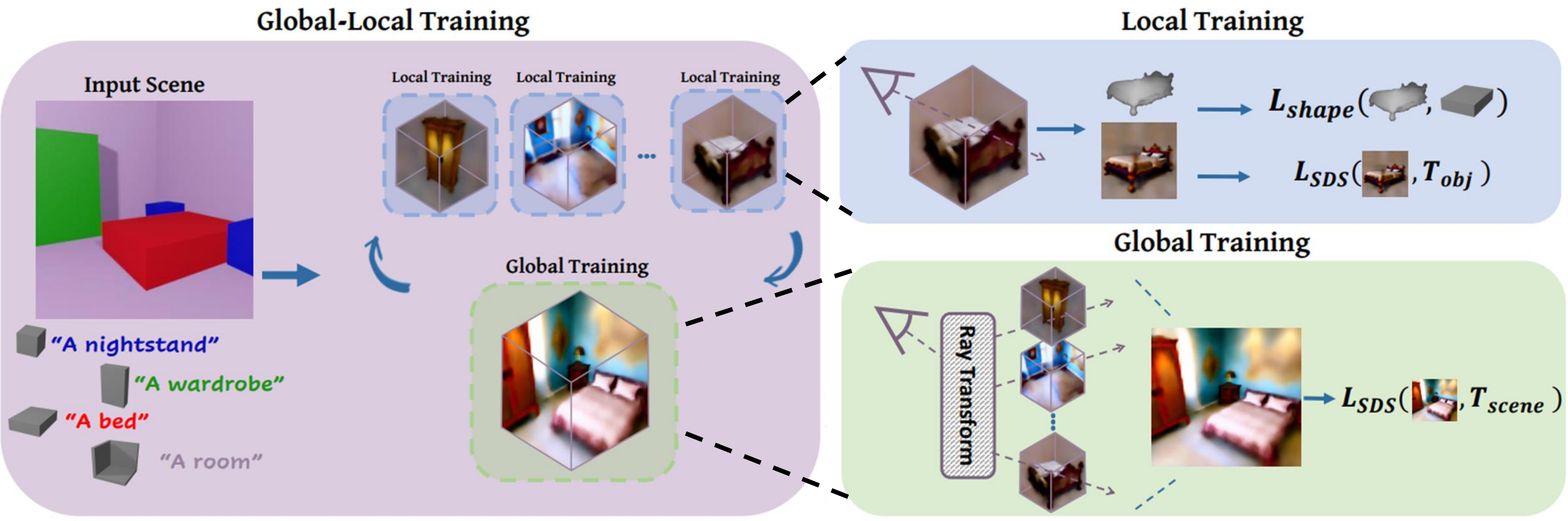
- **SDS loss** enable us to create amazing 3D object using only prompts!
- But its ability to create **a multi-object scenes** is still limited ☹
- **Set-the-Scene** creates **composable 3D scenes** from **layout + prompt!**

Project Page Code

Scan for more info!

Global - Local Training

Scene layout:
Each object has a "**Proxy**" that defines it's location, orientation, dimensions, and (optionally) geometry

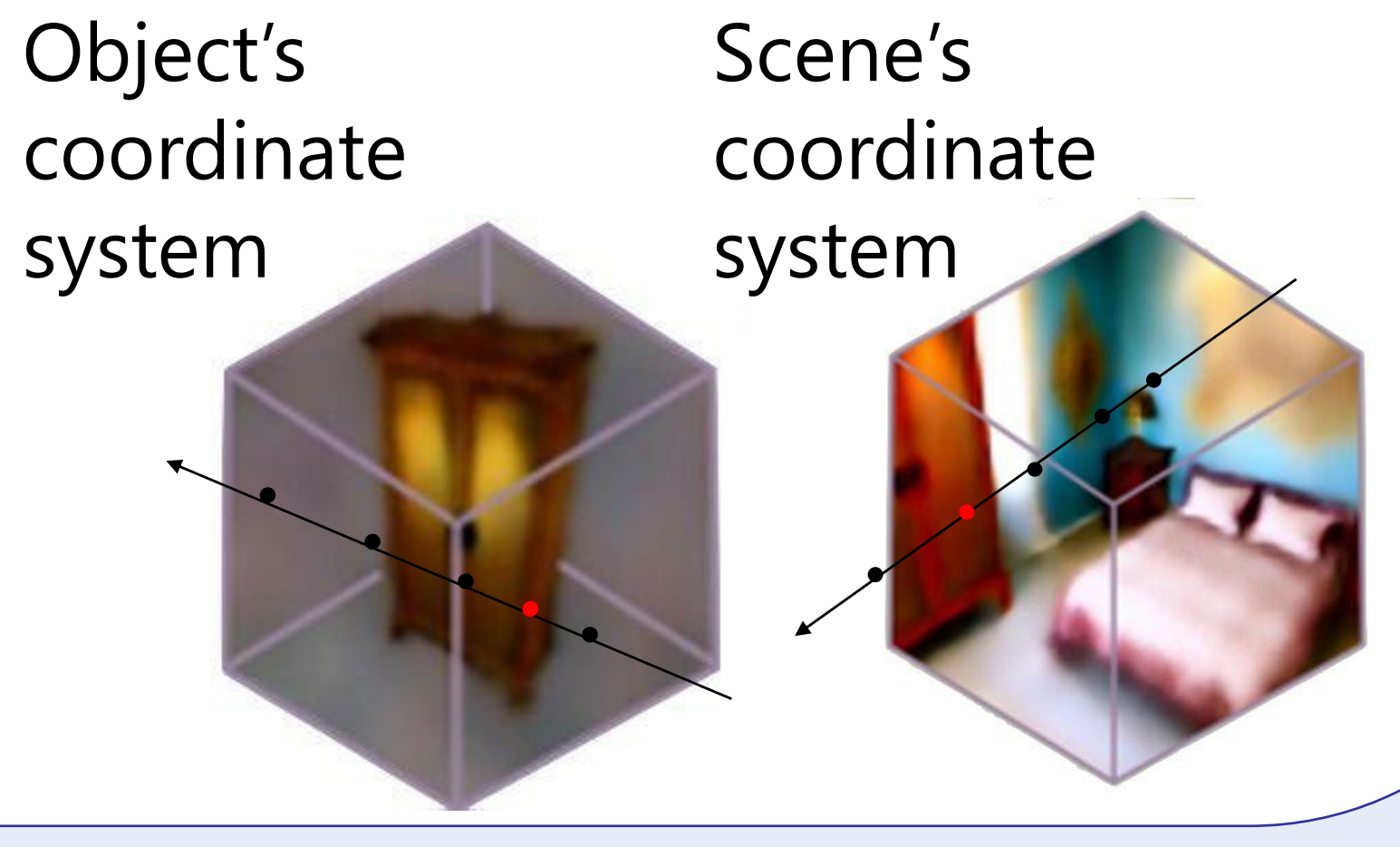


We alternate between:
"**local training**" - optimizing each object alone using *object prompt* + shape loss[1]
"**global training**" - jointly rendering the objects and optimizing with *scene prompt*

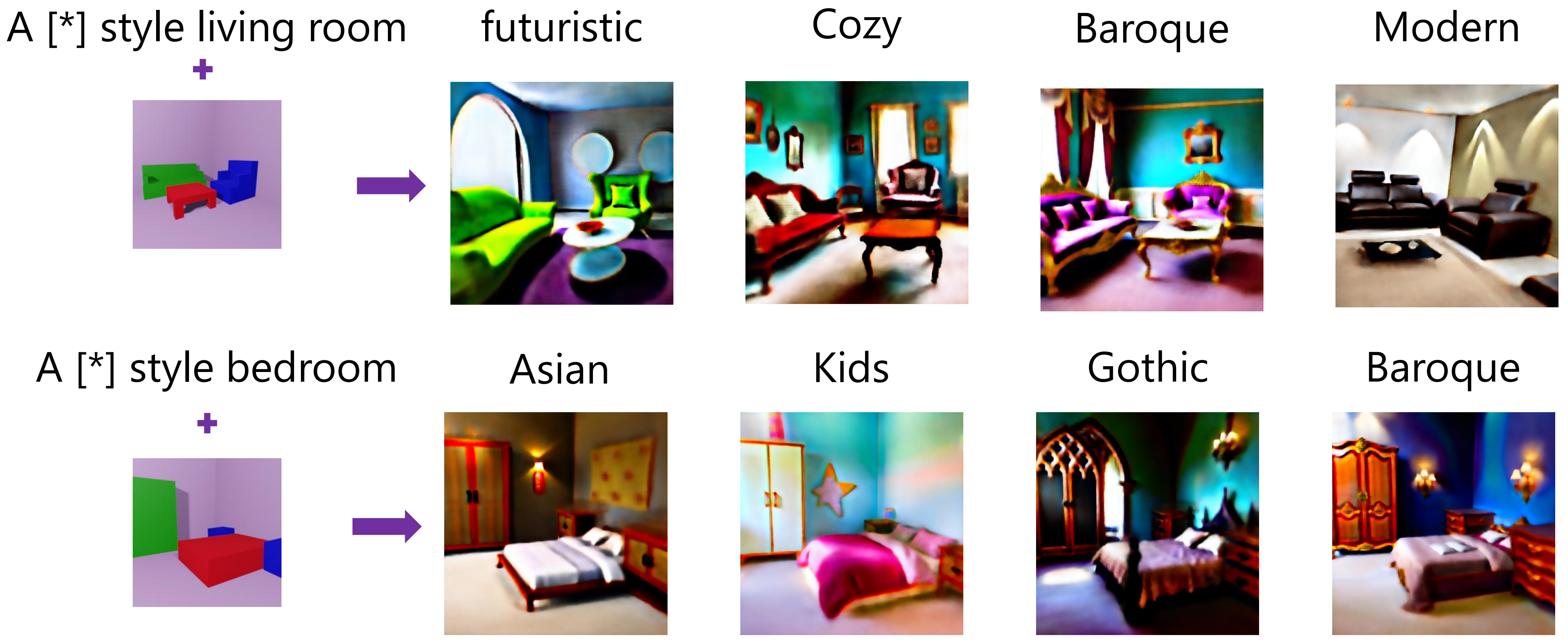
[1] Latent Nerf [Metzer et al., CVPR 2023]

How do we jointly render the objects?

- Trace rays through the scene
- Sample from the object's NeRFs iteratively and employ the inverse transform to change coordinate system

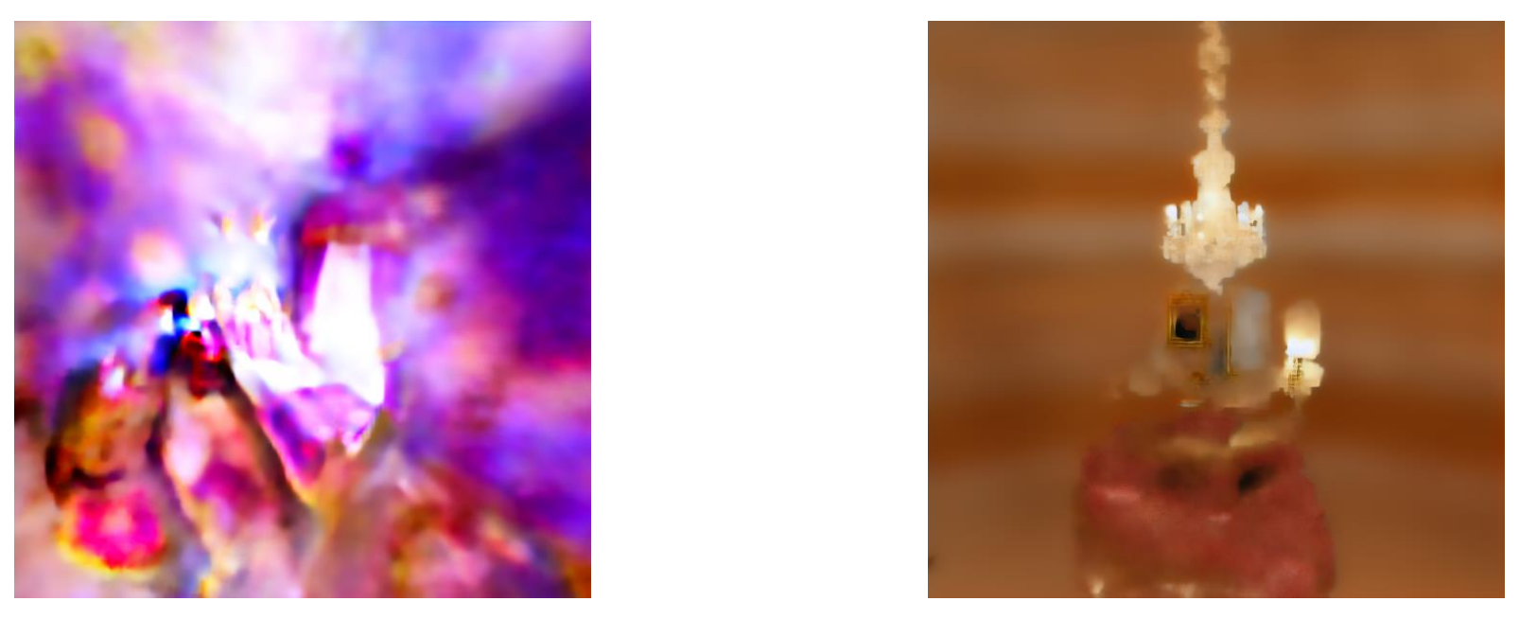


Let's use the same layout with different prompts!

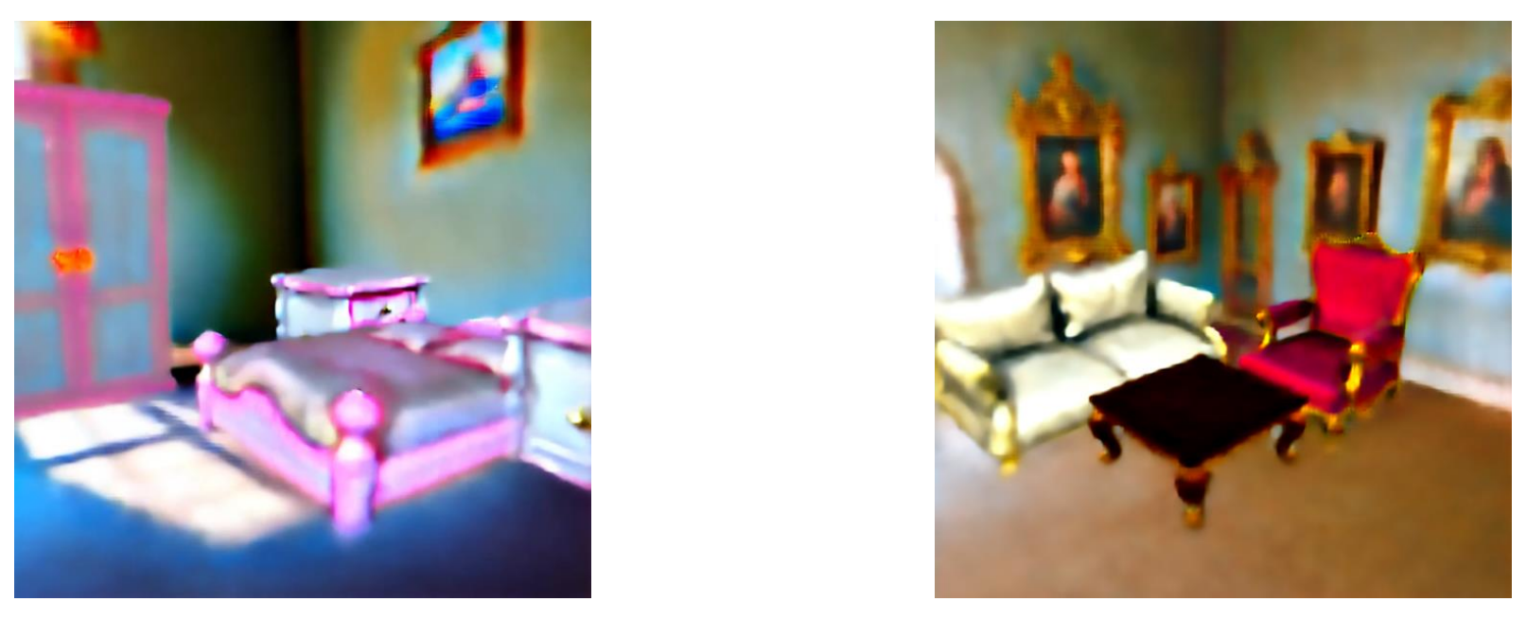


Ablation

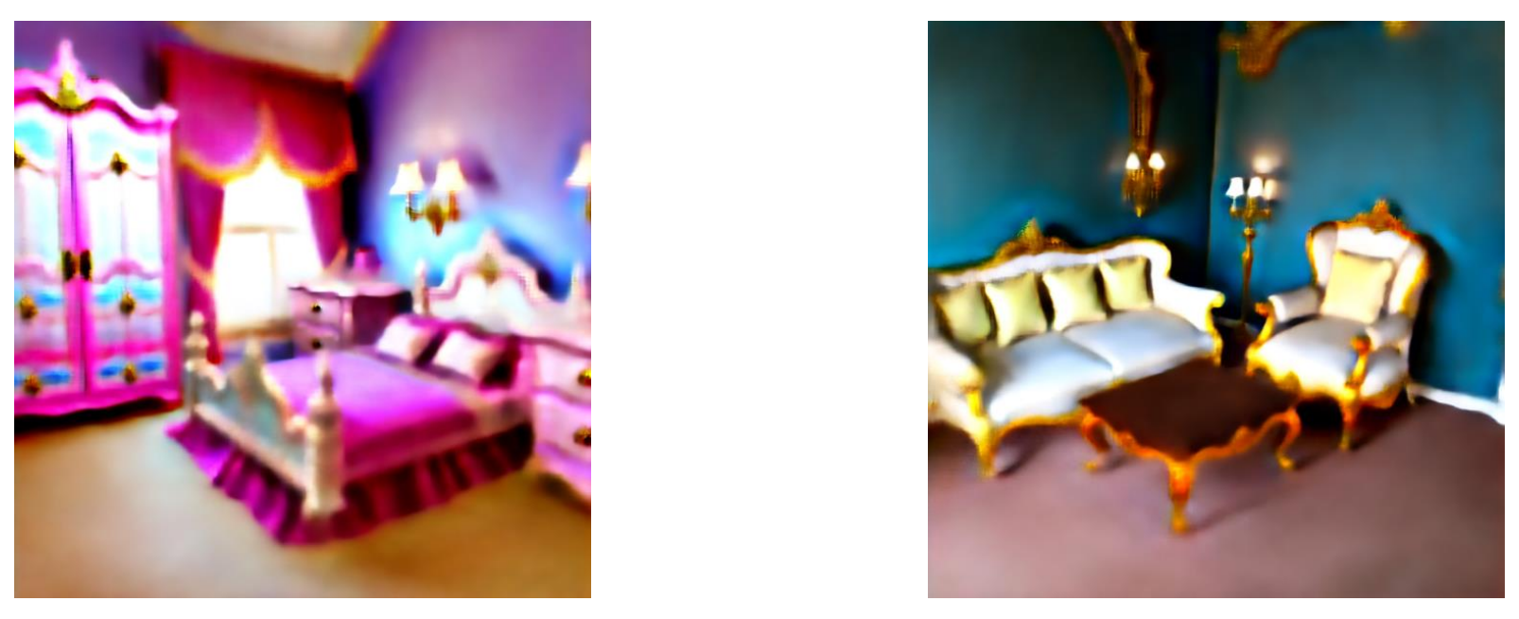
SDS with the scene prompt



Only **local training :** generate each object separately and render together at inference time

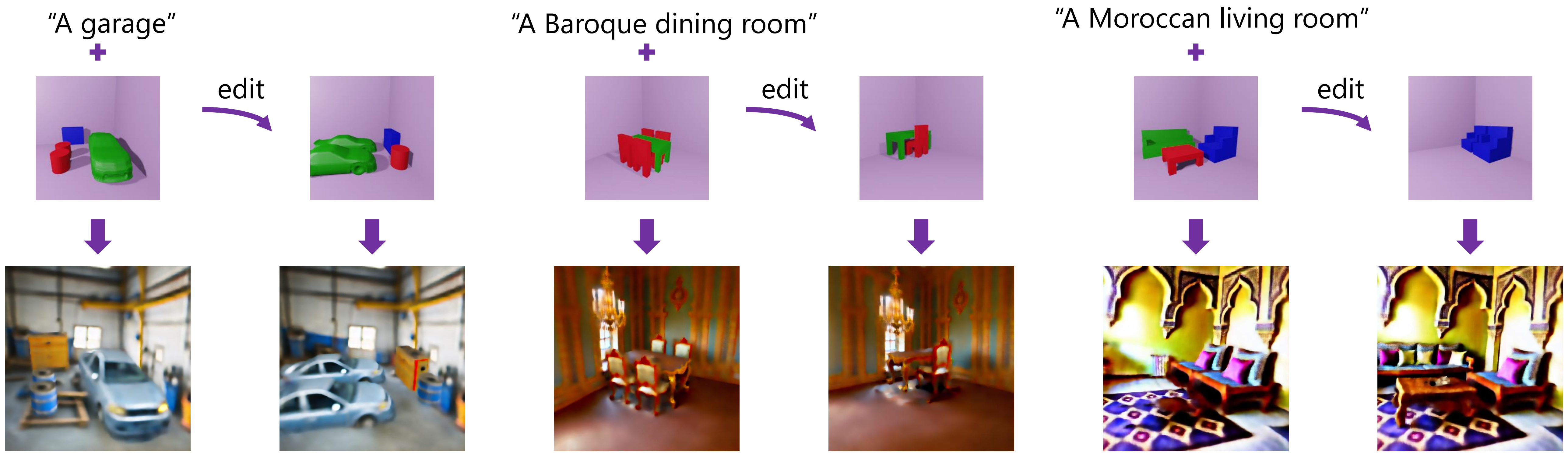


Set –the-Scene:
Global-Local training



Inference Time Editing 🤩

We can **edit** the scene by **changing the proxies** to create a different layout at **inference time,** without additional fine-tuning


"A garage"
+
edit
"A Baroque dining room"
+
edit
"A Moroccan living room"
+
edit