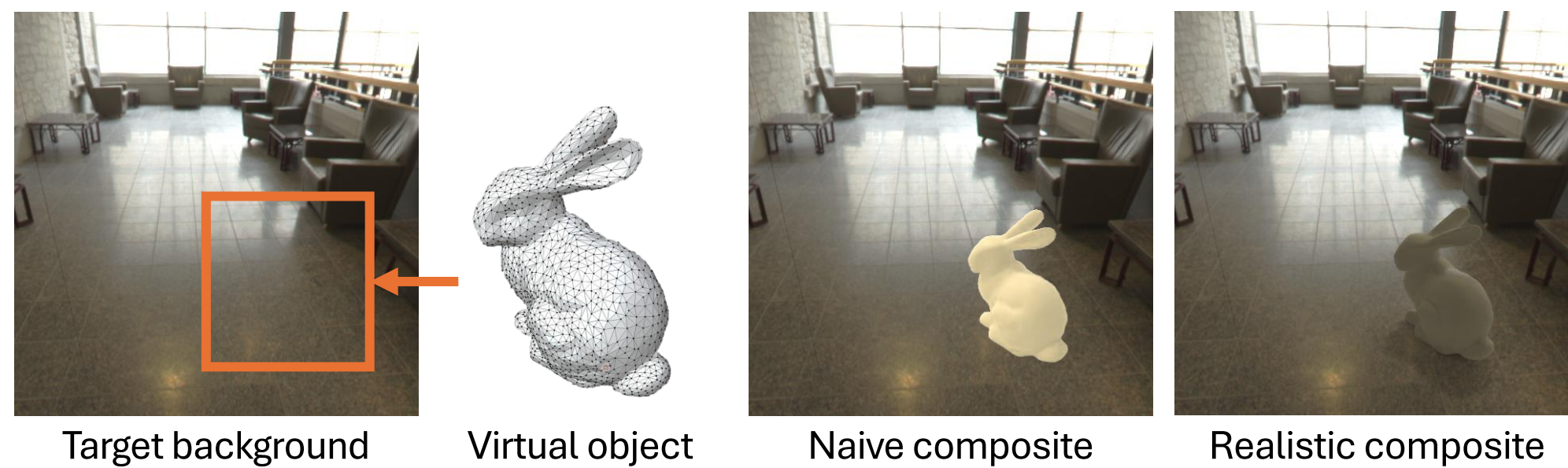


ZeroComp: Zero-shot Object Compositing from Image Intrinsic via Diffusion

Zitian Zhang¹, Frédéric Fortier-Chouinard¹, Mathieu Garon², Anand Bhattad³ & Jean-François Lalonde¹

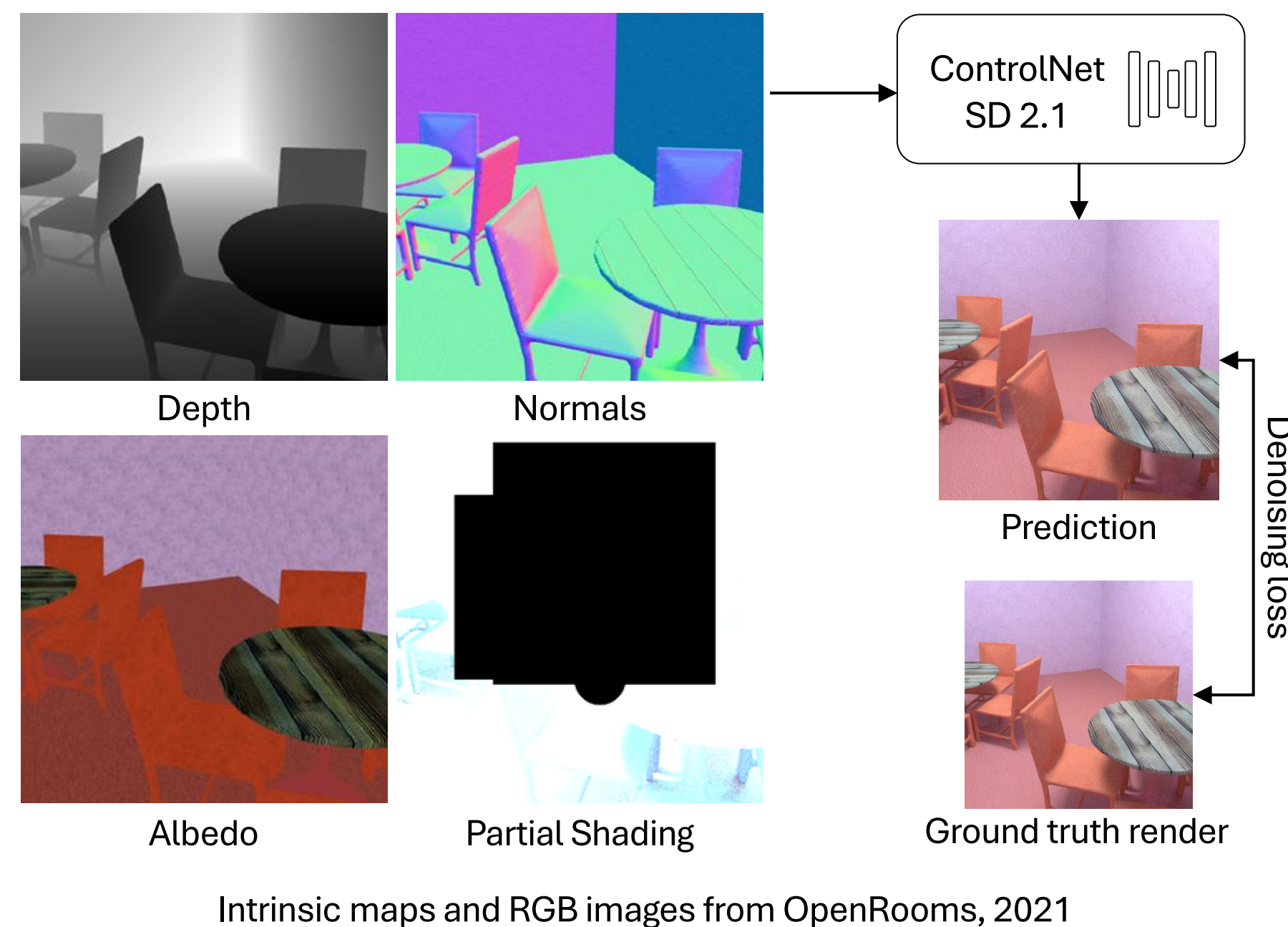
The **shading** and **shadows** are important when inserting 3D objects into images



Existing methods require **specialized datasets**, and they fail to preserve **object identity**

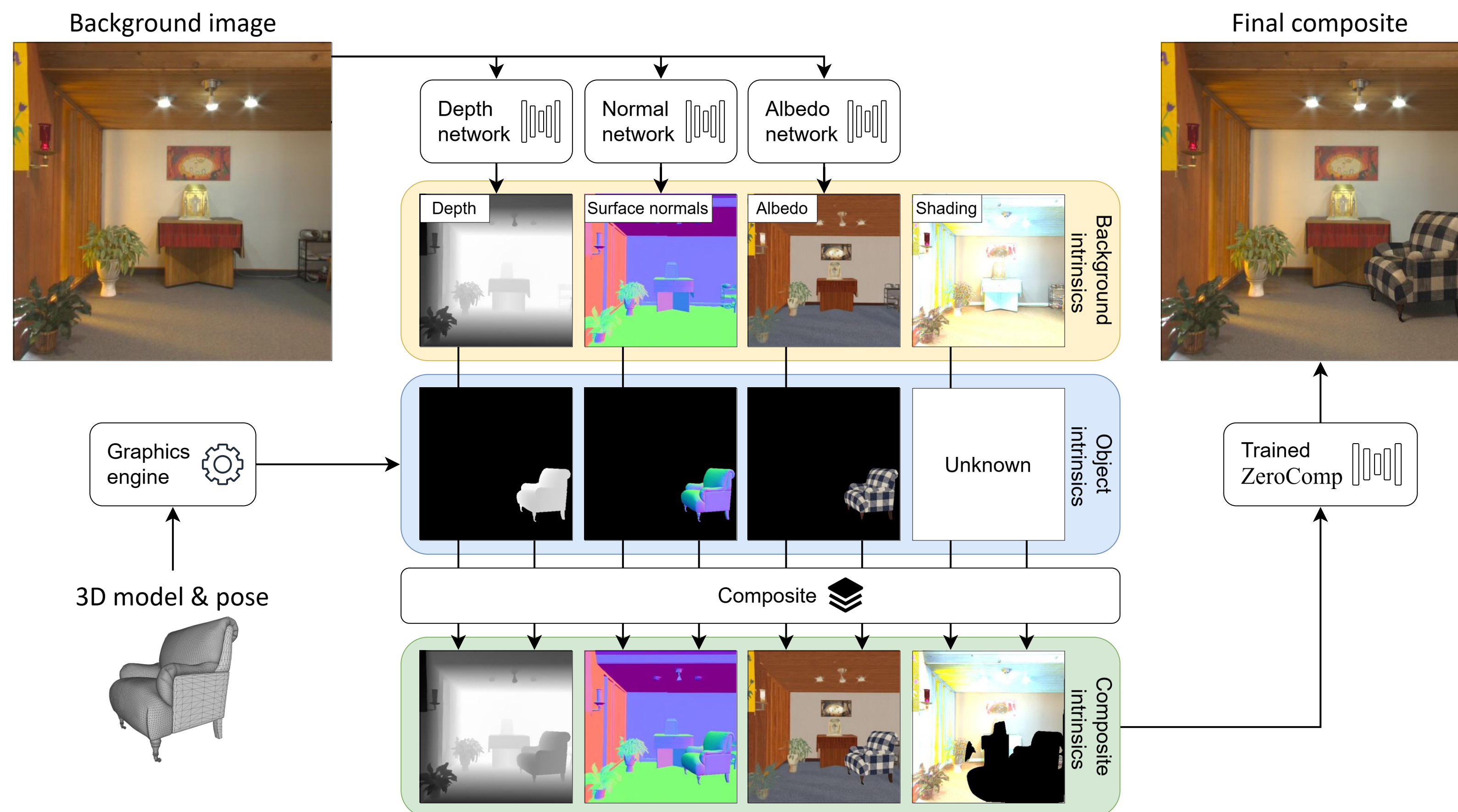


In this work, we train a **diffusion neural renderer** on **intrinsic maps**, that achieves **zero-shot compositing**



Zero-shot compositing pipeline

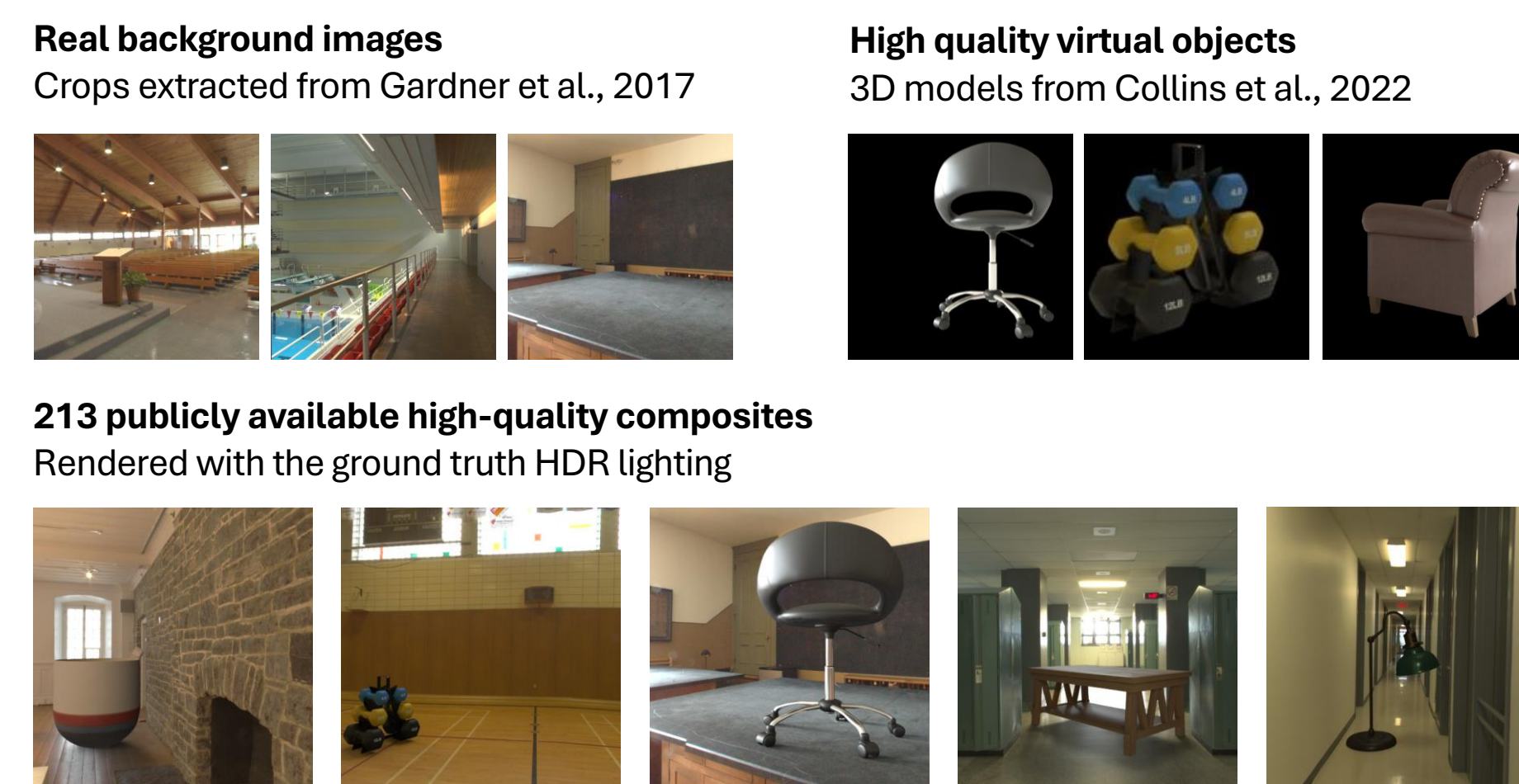
- The input background image (top-left) is first converted to intrinsic maps using specialized networks (top, in **yellow**).
- The corresponding intrinsics of the 3D object, except for the shading, are rendered with simple shaders (middle, in **blue**).
- The intrinsic maps are then composited together to obtain the combined intrinsics (bottom, in **green**).
- From this, our trained ZeroComp model renders a realistic final composite (top-right).



Shading mask size



Test dataset for object compositing



Qualitative comparison

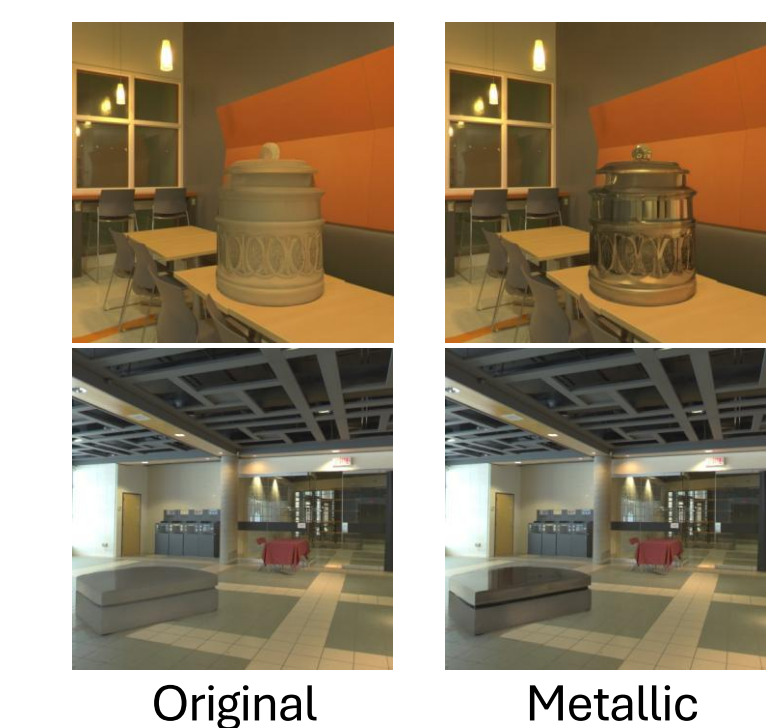


User study

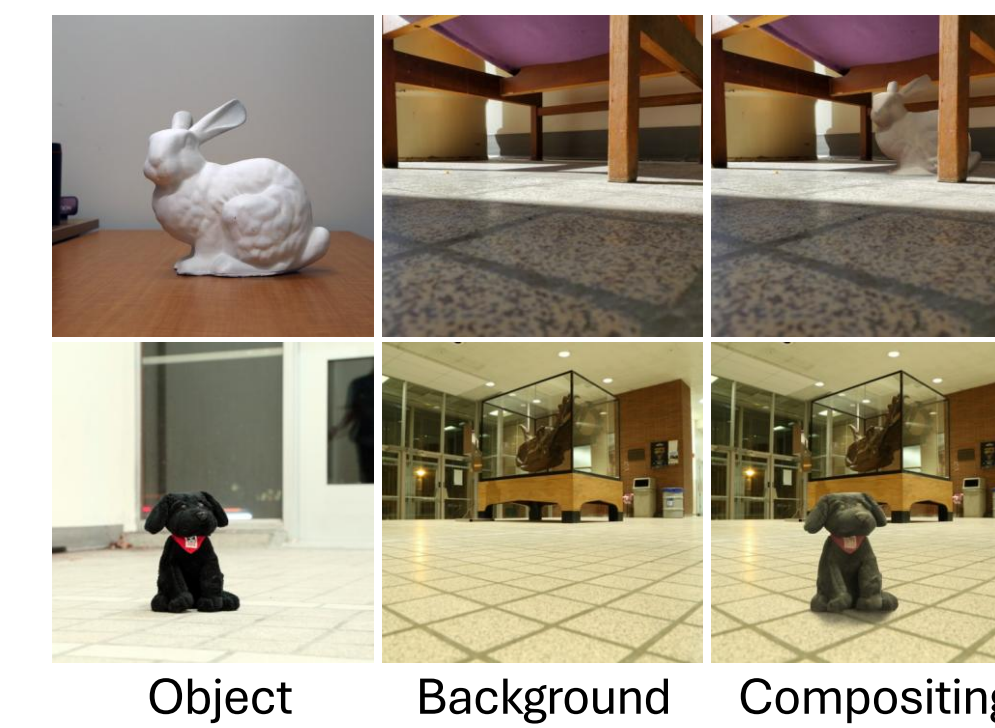
- The ground truth composite and method output are randomly shown side by side.
- Observers are asked to select the most realistic image.
- Our method outperformed all baselines in terms of achieving the highest confusion with the ground truth.

Method	Confusion
Ours	45.0 %
EMLight	41.5 %
ControlCom	19.9 %
Optimal confusion rate	50.0 %

Material editing



2D compositing



Follow-up paper: SpotLight

Lighting control can be achieved with **no retraining** by injecting **custom shadows** during the denoising process of ZeroComp

