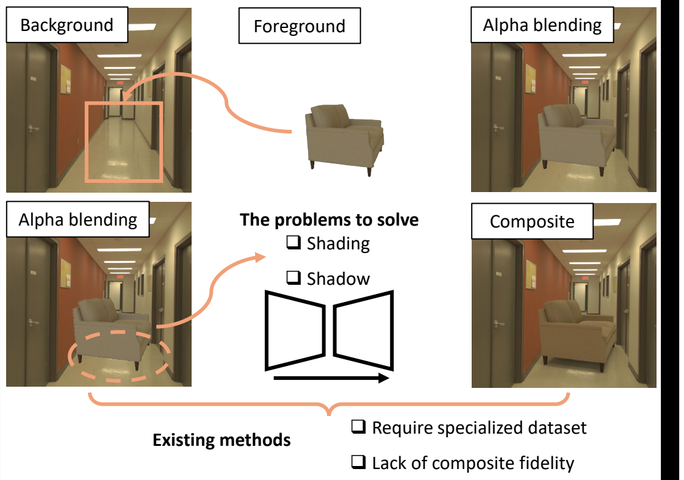
ZeroComp: Zero-shot Object Compositing from Image Intrinsics via Diffusion

Zitian Zhang[1], Frédéric Fortier-Chouinard[1], Mathieu Garon[2], Anand Bhattad[3] & Jean-François Lalonde[1]

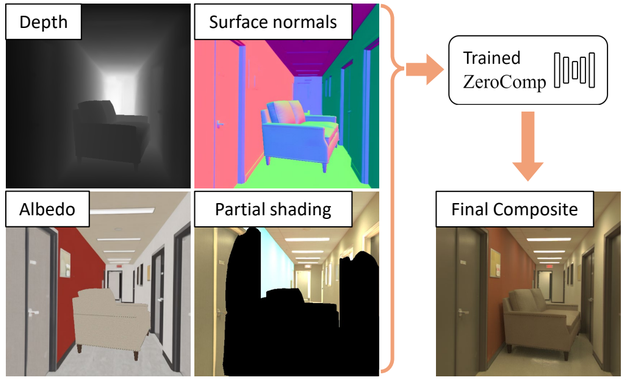
Motivation

Compositing 3D virtual objects into real photographs is essential for image editing and visual effects. It requires realistic interactions between virtual objects and the target scene lighting and layout, i.e., the object shading and the shadow it casts.



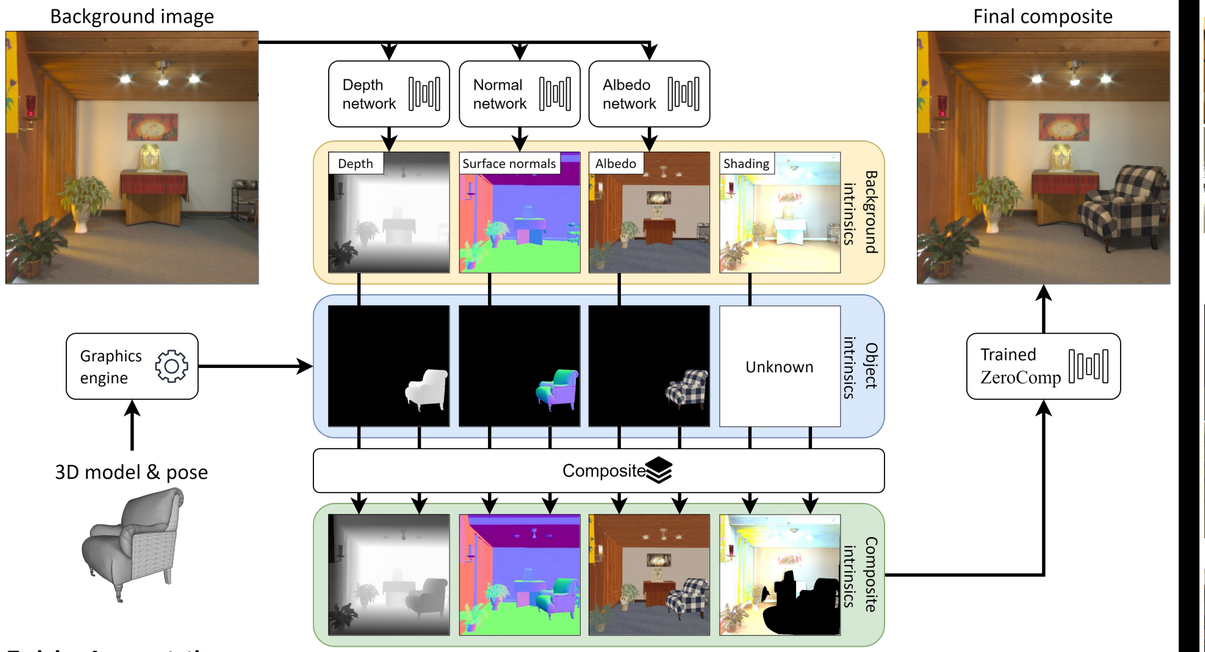
Our Method

- Our method utilizes the strong generative power of StableDiffusion
- It achieves **zero-shot** compositing by training a ControlNet on the simpler proxy task of reconstructing an image from its intrinsic layers using readily available datasets.
- Once trained, it seamlessly integrates virtual 3D objects into scenes, adjusting shading to create realistic composites.



Zero-shot Compositing Pipeline

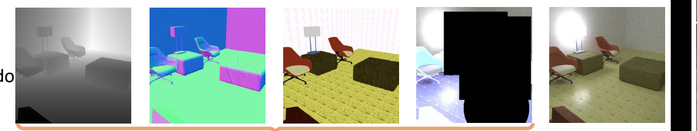
- The input background image (top-left) is first converted to intrinsic using specialized networks (top, in **yellow**).
- In parallel, the corresponding intrinsic layers of the 3D object---except the shading---are rendered using a graphics engine (middle, in **blue**).
- Layers are then composited together to obtain the composited (bottom, in **green**).
- From this, our trained ZeroComp renders the final composite (top-right).



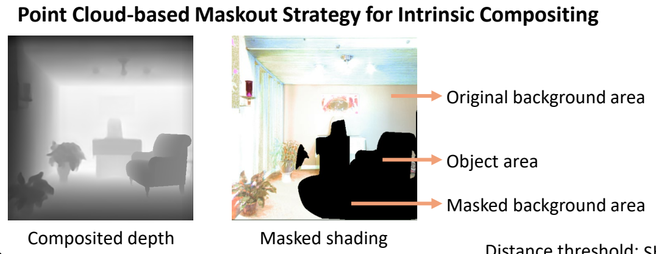
Training Augmentation

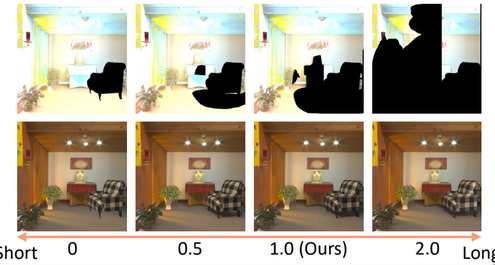
OpenRooms dataset

- We use a subset of 17,952 images paired with intrinsics
- The partial shading is a division between the image and albedo
- The partial shading is randomly cropped or fully kept during training



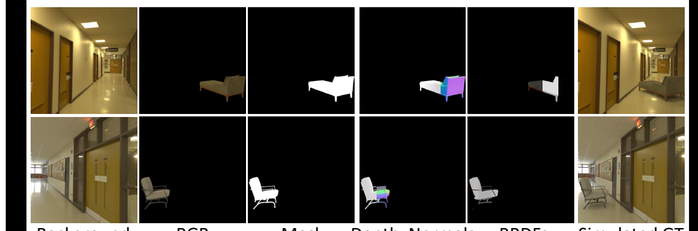
Point Cloud-based Maskout Strategy for Intrinsic Compositing



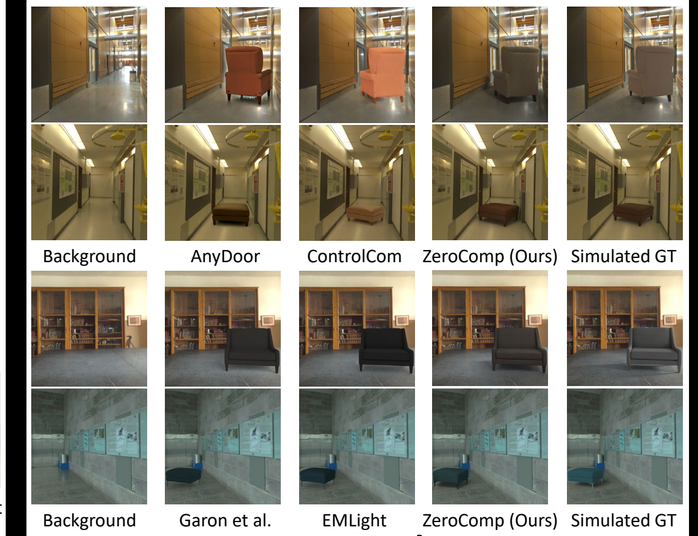


Test Dataset for Object Compositing

- We generate a dataset for evaluating 3D object compositing
- We utilize HDR environment maps from the Laval Indoor HDR dataset and 3D objects from the ABO dataset
- Our dataset includes 213 high-quality images rendered in Blender.



Qualitative Comparison



User Study

Method	Confusion (%)
ZEROCOMP (ours)	**45.0 ± 3.9**
EMLight	41.5 ± 3.9
Garon'19	31.5 ± 3.6 *
Everlight	31.4 ± 3.6 *
ControlCom	19.9 ± 3.1 *
Careaga'23	5.0 ± 1.7 *
ARShadowGAN	4.8 ± 1.7 *

- A two-alternative forced choice user study
- A confusion rate of 50% would imply that the generated composites are perceived as indistinguishable from the ground truth
- Our method achieved a 45% confusion rate, outperforming all competitors, indicating a strong preference for its realism

For more method details, qualitative and quantitative results, please refer to the project page, the paper and the supplementary.

zitian.zhang.1@ulaval.ca, frederic.fortier-chouinard.1@ulaval.ca, mathieu.garon@depix.ai, bhattad@ttic.edu, jflalonde@gel.ulaval.ca