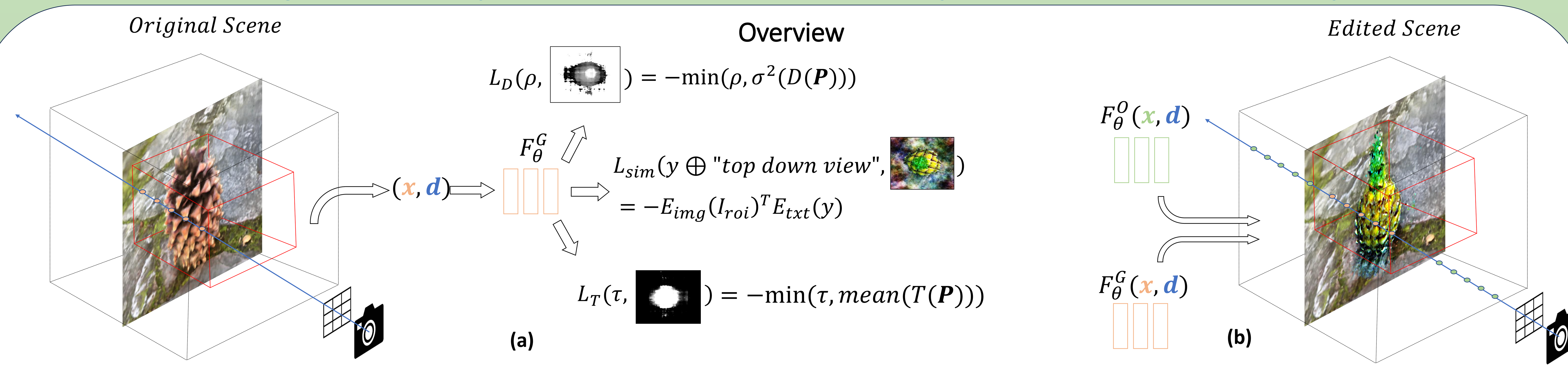
Blended-NeRF: Zero-Shot Object Generation and Blending in Existing Neural Radiance Fields

Ori Gordon — ori.gordon@mail.huji.ac.il
Omri Avrahami — omri.avrahami@mail.huji.ac.il
Dani Lischinski — danix@mail.huji.ac.il
האוניברסיטה העברית בירושלים
THE HEBREW UNIVERSITY OF JERUSALEM

Project Page

Overview



(a) Training: Given a NeRF scene F_θ^O, our pipeline trains a NeRF generator model F_θ^G, initialized with F_θ^O weights and guided by a similarity loss defined by a language-image model such as CLIP, to synthesize a new object inside a user-specified ROI. This is achieved by casting rays and query the same set of points for both F_θ^O and F_θ^G only inside the ROI box. Our method introduces augmentations and priors to get more natural results, such as the Transmittance Loss L_T and Depth Loss L_D which encourage the generator to increase the average transmittance and variance of the disparity map accordingly.

(b) Blending: After training, we render the edited scene by blending the sample points generated by the two models along each view ray. For points outside the ROI, we use the density and color inferred by F_θ^O, while for points inside the ROI, we blend the results of the two radiance fields depending on the type of the edit: Adding new object in empty space. vs. adding an object in a non empty area.

Editing NeRF Challenges

1. Scene is encoded in an implicit manner by the model weights.
2. No explicit separation between the various components that define the object, such as shape, color or material.
3. In contrast to local edits in images done in pixel space, editing NeRF scene is more challenging due to the consistency requirement between multiple views.

Blended-NeRF Contributions

We present ROI-based NeRF editing approach guided by a text prompt that:

1. Can operate on any region of a real-world scenes.
2. Modifies only the ROI while preserving the rest of the scene.
3. Not restricted to a specific class or domain.
4. Enables complex text guided manipulations such as object insertion/replacement, objects blending and texture conversion.

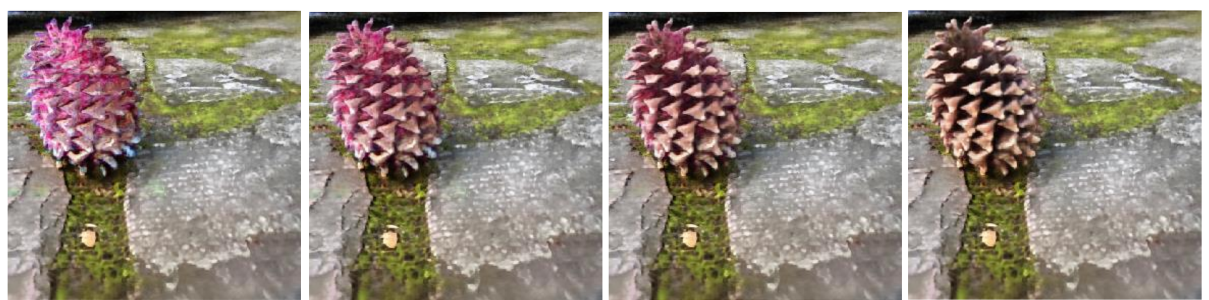
Distance Smoothing Operator

To get smooth blending between the two scenes, we consider its distance from the center of mass inside the ROI and alpha compositing the density and color:

$$f(\boldsymbol{x}) = 1 - \exp\left(-\frac{\alpha d(\boldsymbol{x})}{diag}\right)$$

$$\sigma_{blend}(\boldsymbol{x}) = f(\boldsymbol{x}) \cdot \sigma_O(\boldsymbol{x}) + \left(1 - f(\boldsymbol{x})\right) \cdot \sigma_G(\boldsymbol{x})$$

$$C_{blend}(\boldsymbol{x}) = f(\boldsymbol{x}) \cdot C_O(\boldsymbol{x}) + \left(1 - f(\boldsymbol{x})\right) \cdot C_G(\boldsymbol{x})$$



$\alpha = 0$ $\alpha = 0.5$ $\alpha = 2$ $\alpha = 10$

Object Blending

In this mode we query both F_θ^O and F_θ^G inside the ROI and blend the resulting colors and densities at each ray sample. We blend the colors using the alpha values of each model, when S is the sigmoid function:

$$c(x_i) = S\left(\frac{C_O(x_i)\cdot\alpha_O(x_i) + C_G(x_i)\cdot\alpha_G(x_i)}{\epsilon + \alpha_O(x_i) + \alpha_G(x_i)}\right)$$

We present 2 options for blending the densities, when ϕ it the activation function:

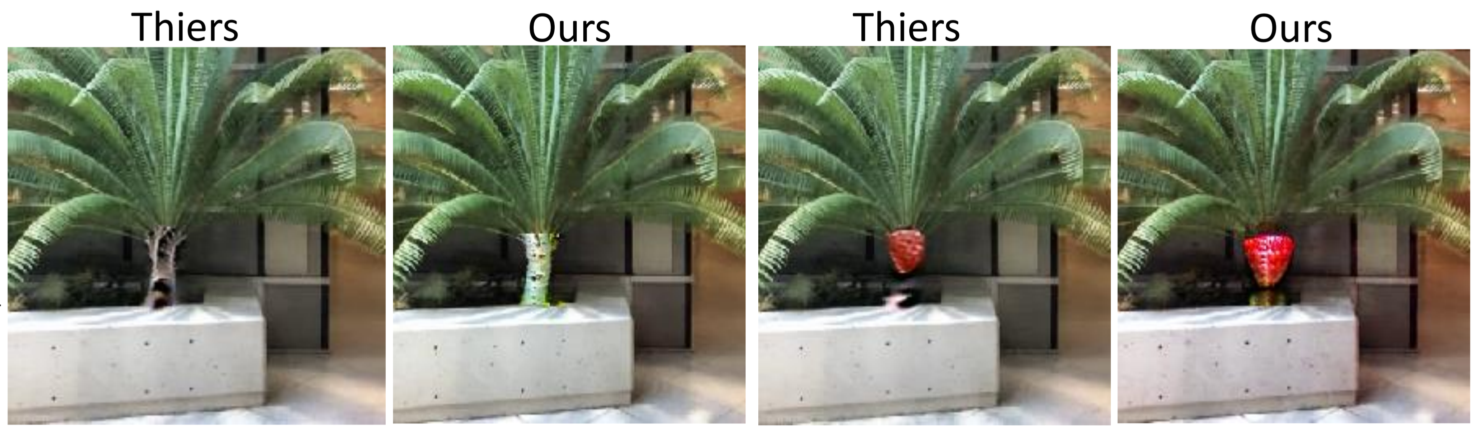
1. $\sigma(x_i) = \phi(\sigma_o(x_i) + \sigma_G(x_i))$
2. $\sigma(x_i) = \phi(\sigma_o(x_i)) + \phi(\sigma_G(x_i))$

"Plant with green leaves and white and blue flowers"

Comparisons

We compare qualitatively and quantitatively to Volumetric Disentanglement by Benaim et al. :

Method	CLIP Direction Similarity ↑	CLIP Direction Similarity Consistency ↑	LPIPS ↓
[Benaim] 2022	0.128	0.736	0.3
Ours	0.143	0.787	0.024



"Aspen Tree" "Strawberry"

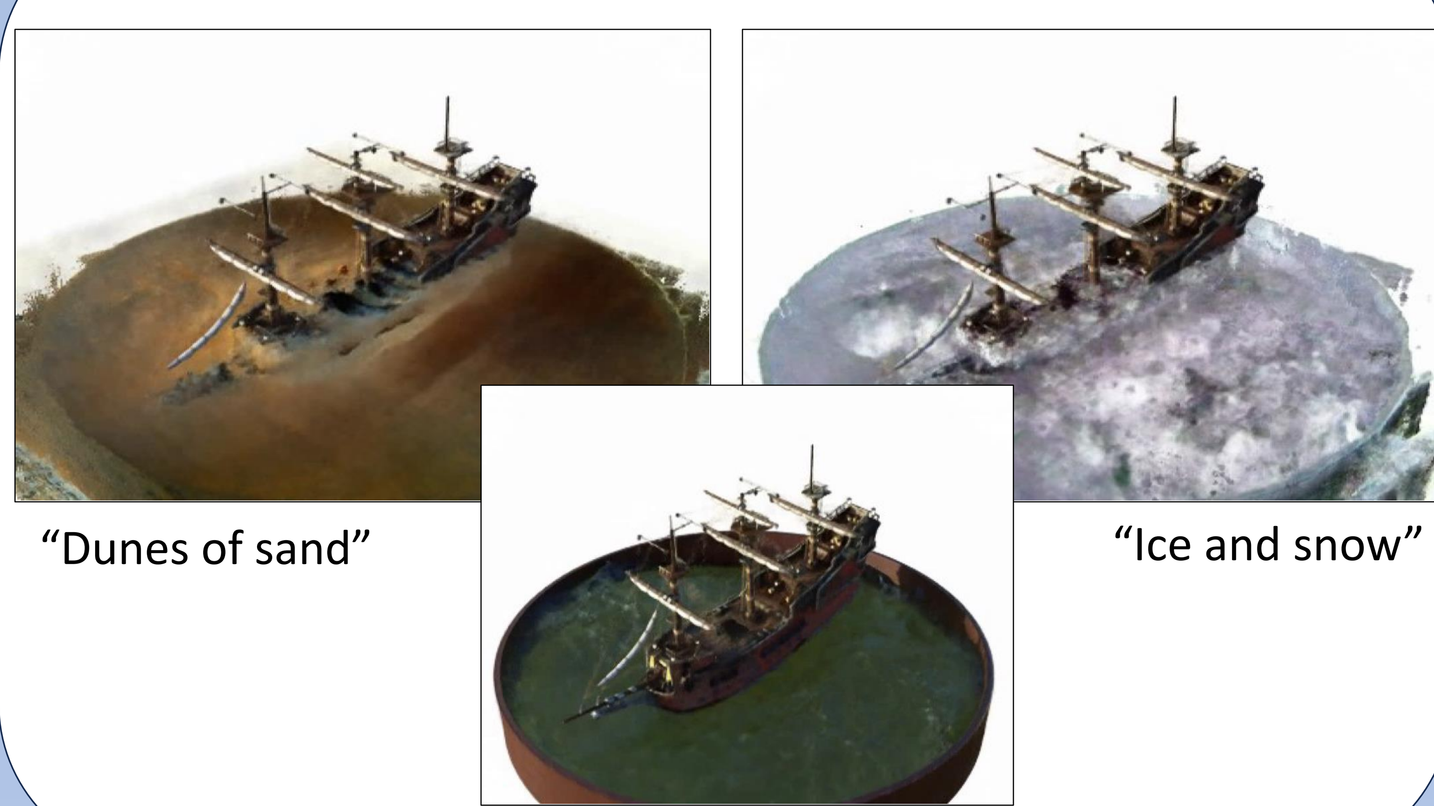
Ablation Study

Method	CLIP R-Precision ↑	BLIP R-Precision ↑
COCO GT	0.933	0.98
Ours(full pipeline)	0.86	0.8
Ours(no dir prompts)	0.85	0.8
Ours(no depth priors)	0.81	0.78

Applications – Texture Conversion

Original Scene "Glass"/"Burning" "Water paint"/" Ice" "Stone"/"Pink wool"

Applications - Objects Replacement



"Dunes of sand" "Ice and snow" Original Scene

Applications - Objects Blending

"A clusters of mushrooms" "Green and yellow bananas"