(a) Step 1: Foreground-aware NeRF Training

Color Field
$(\boldsymbol{p}, \boldsymbol{d}) \rightarrow c$

$\mathcal{L}_{MSE}$

Captured Images

Prompt: a bear statue

Grounded-SAM

$\boldsymbol{p} \rightarrow \sigma$

$\mathcal{L}_{BCE}$

Pseudo Foreground Masks

$(\boldsymbol{p}, \boldsymbol{d}) \rightarrow m$
Mask Field

Original NeRF

(b) Step 2: Reference Subject Learning

Prompt: a photo of a $\boldsymbol{V^*}$ dog

Noisy Image

Diffusion Model

$\mathcal{L}_{LDM}$

Reference Image

* Only required for image-driven editing

(c) Step 3: Adaptive Source Driven NeRF Editing

Local-Global Iterative Editing

$\mathcal{L}_{SDS}^{local}$

Edited NeRF

Local Rendering

+

Prompt0: a dog

Diffusion Model

Global Rendering

+

Prompt1: a dog standing in the forest

Prompt2: a $\boldsymbol{V^*}$ dog standing in the forest

$\mathcal{L}_{SDS}^{global}$

−

Background Rendering

Original NeRF

$\mathcal{L}_{BG}$

Local Editing

Global Editing

* Prompt1 for text-driven editing, Prompt2 for image-driven editing