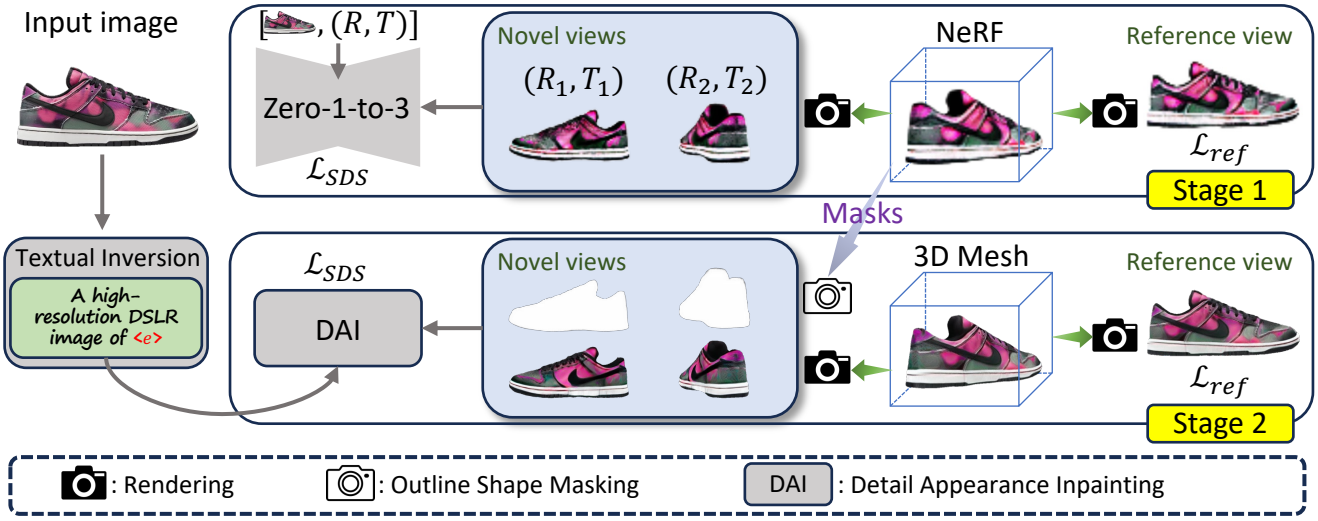
Input image
$[\,, (R,T)]$
Zero-1-to-3
$\mathcal{L}_{SDS}$
Novel views
(R_1, T_1)
(R_2, T_2)
NeRF
Reference view
$\mathcal{L}_{ref}$
Stage 1
Masks
Textual Inversion
A high-resolution DSLR image of <e>
$\mathcal{L}_{SDS}$
DAI
Novel views
3D Mesh
Reference view
$\mathcal{L}_{ref}$
Stage 2
: Rendering
: Outline Shape Masking
DAI
: Detail Appearance Inpainting