

"A photo of **six** kittens sitting on a branch"

SDXL

Instance
Localization

Re-Masker

Layout Guided
SDXL

Attention-Guided Instance-Localization

Cross-Attention
Mask

"kittens"

Self-Attention
Features

DBSCAN

Re-Masker

Iterative object addition

U-Net

Padding

Resize

Object removal

Layout Guided SDXL

Self-Attention
Masking

Diffusion Step

"kittens"

Cross-
Attention

L_{cross}

