

STAGE 1

IP Adapter
(Reference Image)

LoRA
(Factory Style)

Stable Diffusion XL
(Image Generation)



STAGE 2 & 3

Human Annotation
(Bounding Boxes)

IP Adapter
(Masked Attention)

Stable Diffusion XL
(Image Inpainting)