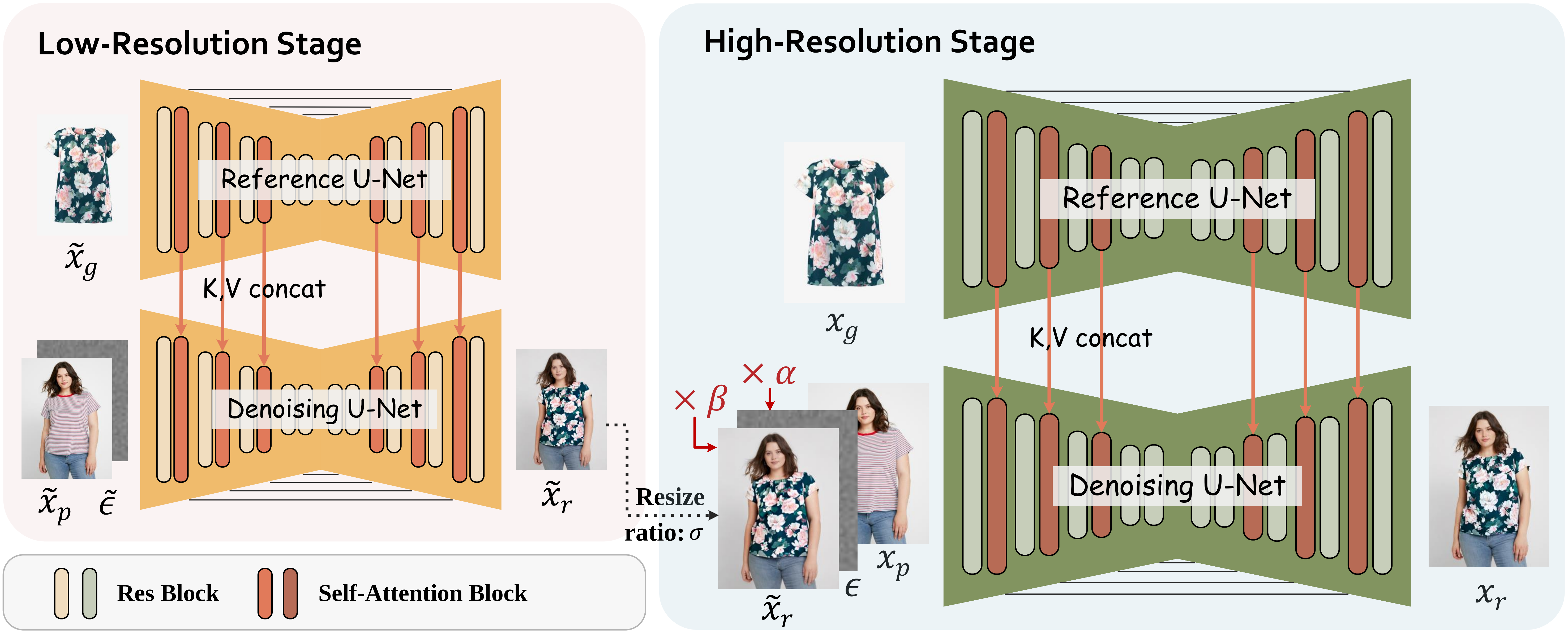
Low-Resolution Stage
High-Resolution Stage
Reference U-Net
K,V concat
Denoising U-Net
$\tilde{x}_g$
$\tilde{x}_p$
$\tilde{\epsilon}$
$\tilde{x}_r$
Resize
ratio: σ
x_g
$\times \alpha$
$\times \beta$
$\tilde{x}_r$
ϵ
x_p
x_r
Res Block
Self-Attention Block