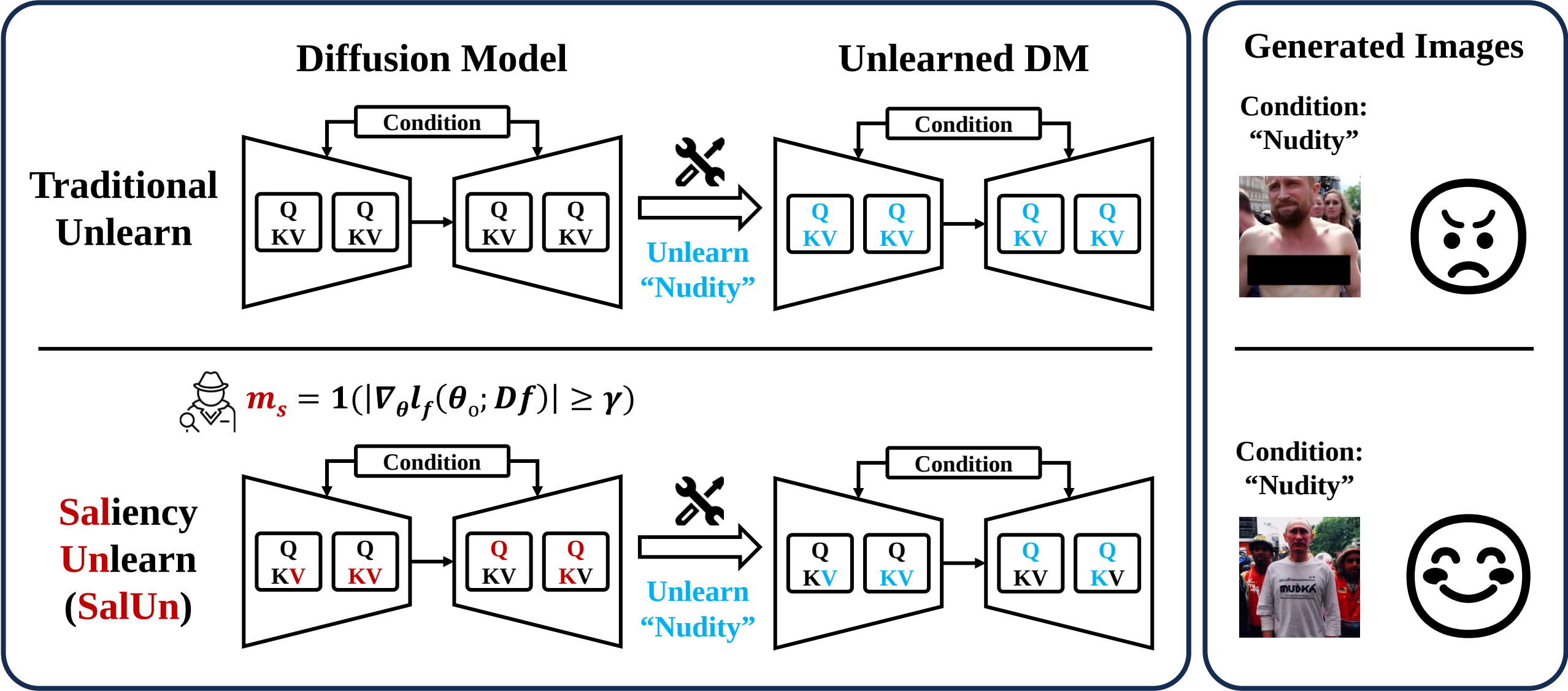
Diffusion Model
Unlearned DM
Generated Images
Traditional Unlearn
Condition
Q KV
Unlearn "Nudity"
Condition: "Nudity"
$m_s = 1(|\nabla_\theta l_f(\theta_o; Df)| \geq \gamma)$
Saliency Unlearn (SalUn)