Edited first frame

Conditional Image

Video Diffusion Model

Pred. Noise (Edited image)

$\lambda + 1$

Unedited first frame

Conditional Image

Video Diffusion Model

Pred. Noise (Unedited image)

$-\lambda$

x_t

$+$

x_{t-1}

Iterative denoising

x_0