



Video UNet

y, \mathbf{x}_t

$\downarrow K \times 64 \times 64 \times C$

ResNet Block + Downsample

$\downarrow K \times 32 \times 32 \times C$

\vdots

$\downarrow K \times 8 \times 8 \times C$

ResNet Block + Attn
+ Downsample

$\downarrow K \times 4 \times 4 \times C$

ResNet Block + Attn

$\downarrow K \times 4 \times 4 \times C$

ResNet Block + Attn
+ Upsample

$\downarrow K \times 8 \times 8 \times C$

\vdots

$\downarrow K \times 32 \times 32 \times C$

ResNet Block + Upsample

$\downarrow K \times 64 \times 64 \times C$

\mathbf{x}_{t-1}