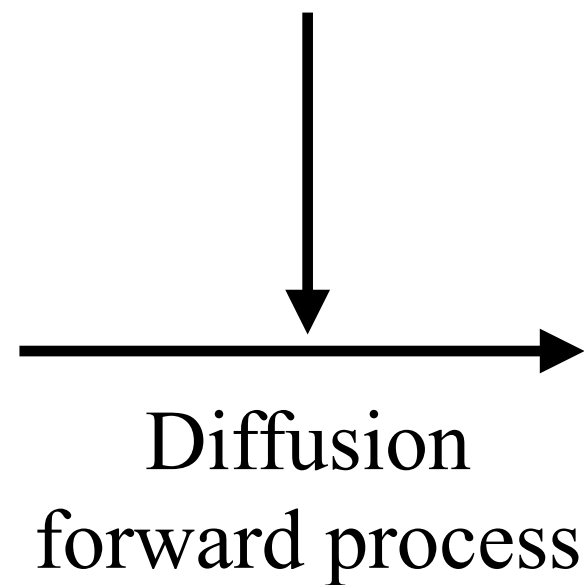


Sampled time step  $t \sim \mathcal{U}[0, 1] \longrightarrow$  Timestep embedding



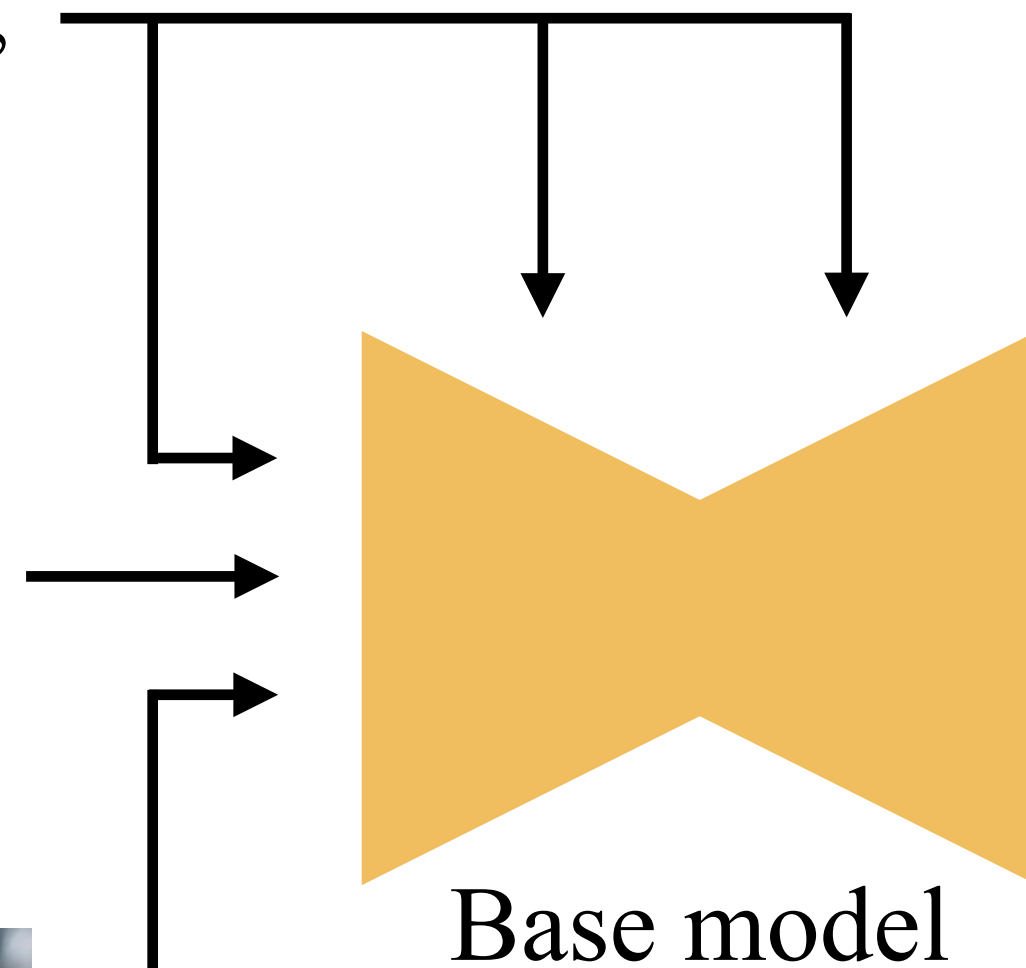
Target intermediate frame  $\mathbf{I}_1$



Noisy input  $z_t$



Input frames  $\mathbf{I}_0, \mathbf{I}_2$



Denoised prediction  $\hat{\mathbf{I}}_1$   
via  $\boldsymbol{v}$ -parameterization