

*DreamSwapV*

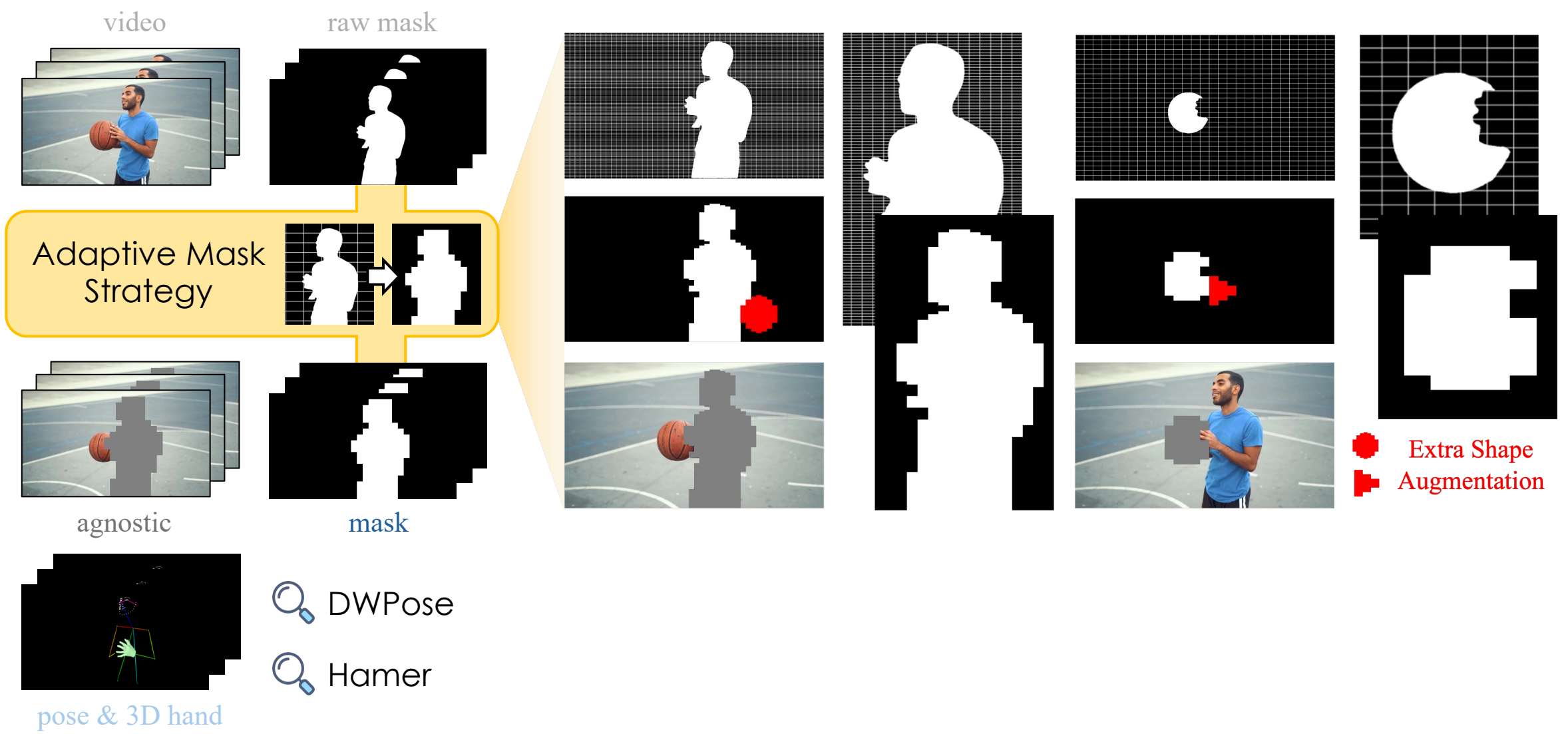
*Mask-guided Subject Swapping  
for Any Customized Video Editing*

**Video Demo**

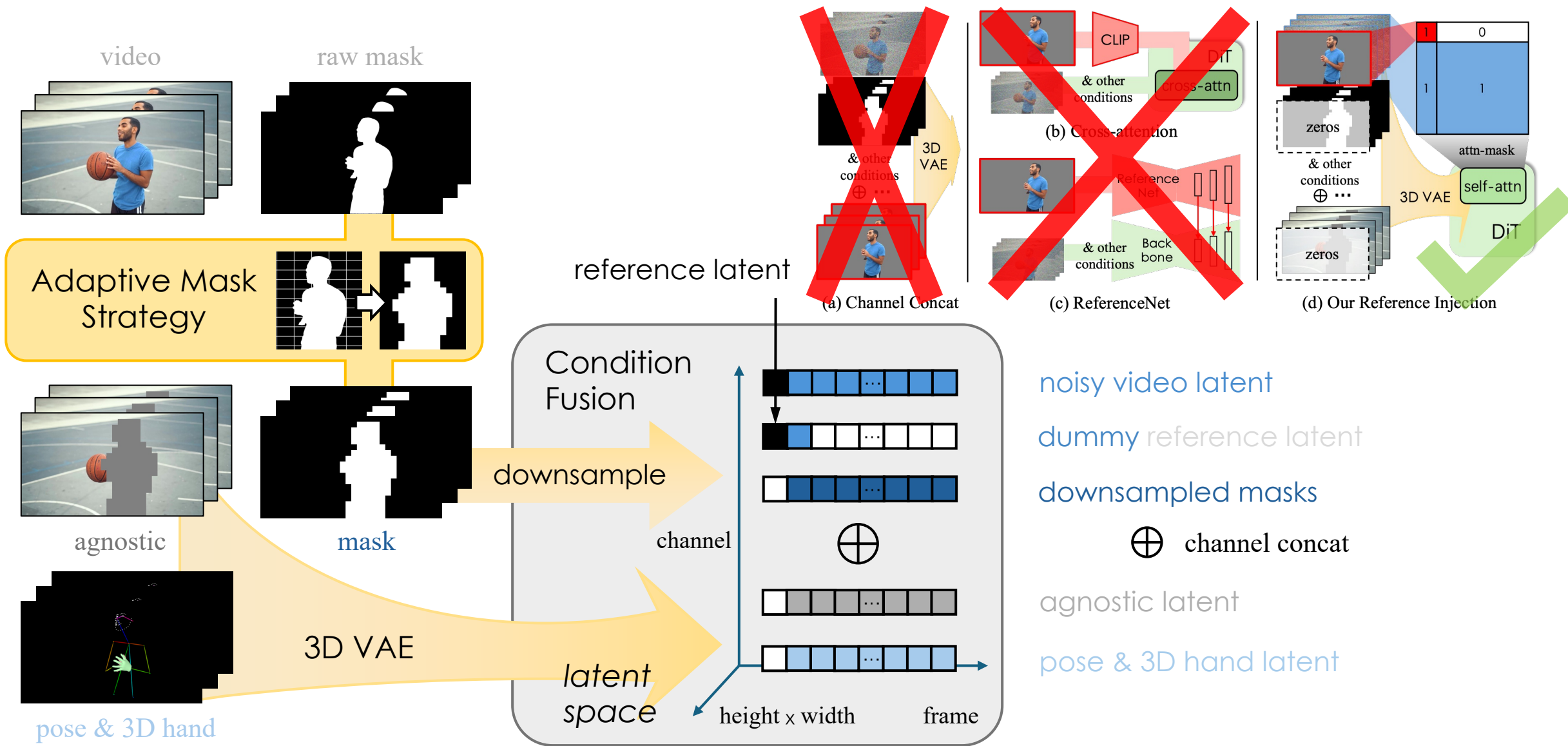
# Dynamic Pipeline and Data Flow



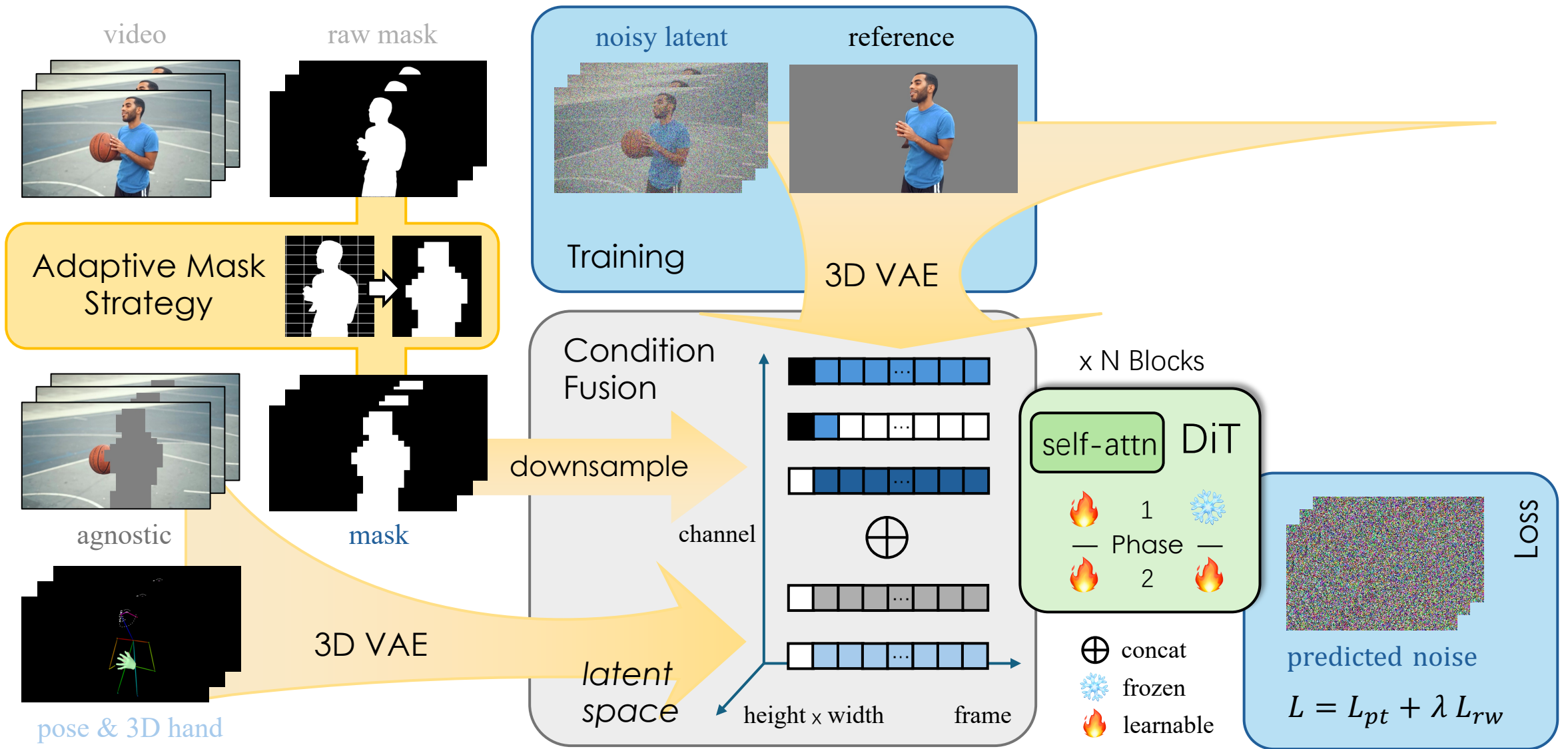
Given a source video and a user-specified first-frame mask localizing the source subject, we first use TrackingSAM to detect the target subject across frames to get a raw mask sequence.



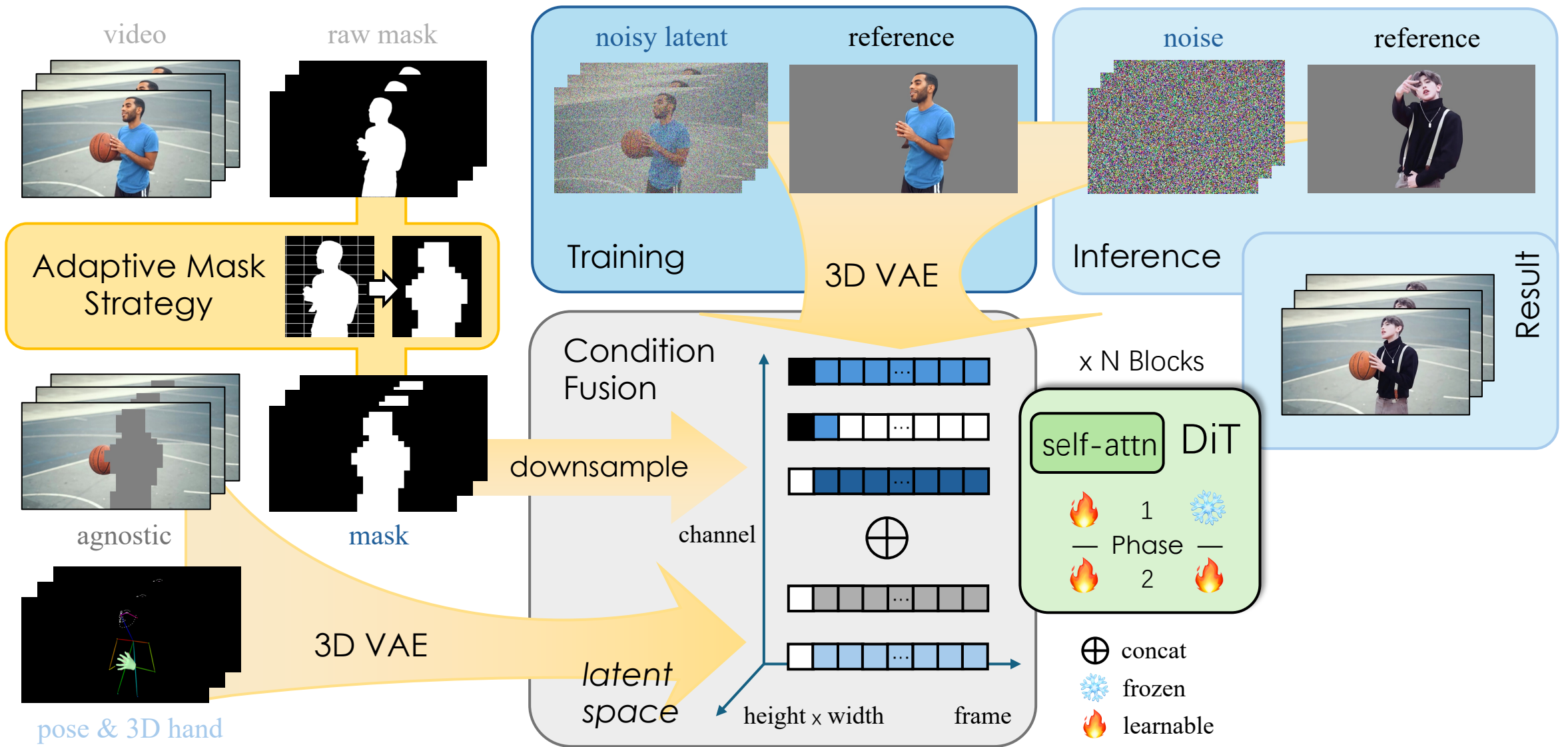
Through our adaptive mask strategy, the raw masks undergoes an adaptive grid augmentation, and the source video is masked to get an agnostic sequence. Pose and 3D hand sequence are detected by DWPose and Hamer.



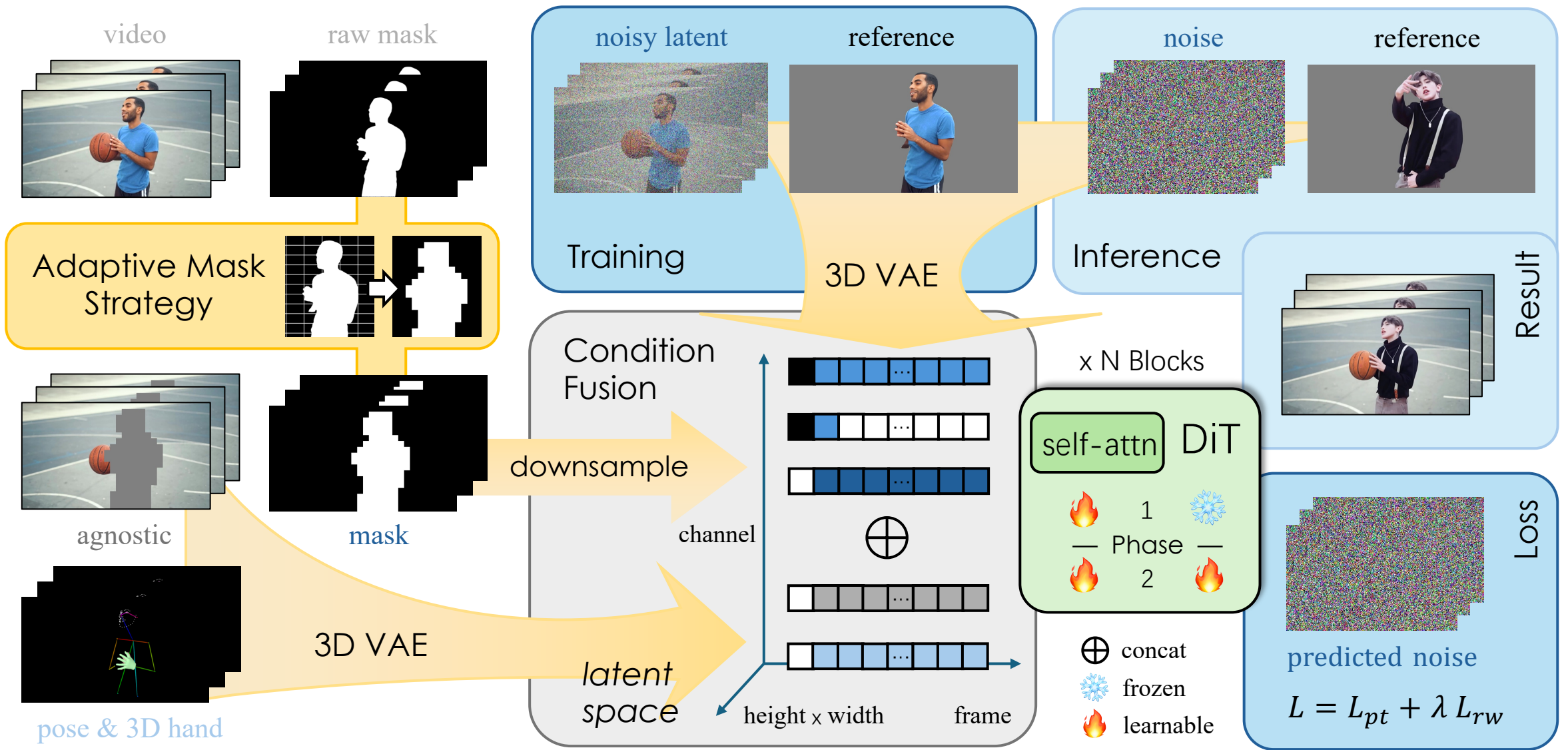
Once upon all the conditions are ready, they are encoded into *latent space* by a pre-trained VAE. Our condition fusion module efficiently handle them with **appropriate** reference information injection and channel concatenation.



During training, the reference image is extracted originally from the source video, and injected into *latent space*. Our model is trained with a two-phase scheme, with only self-attn in phase 1, and full fine-tuning in phase 2.



At inference time, our model can treat an external reference image as if it inherently belonged within the masked region, thus inserting the target subject into the source video seamlessly as a video subject swapping.



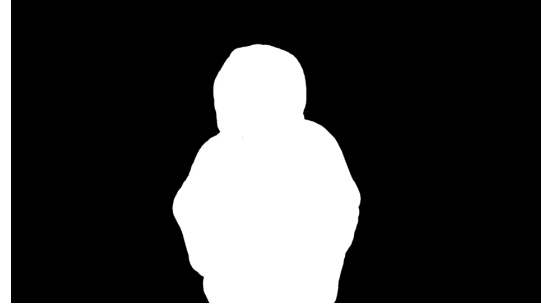
Overall, our DreamSwapV is the first end-to-end framework dedicated to the video subject swapping. Through our novel *condition fusion module* and *adaptive mask strategy*, we deliver state-of-the-art swapping performance.

# Human – full body

# Video Comparisons



source video



mask



reference



DreamSwapV (ours)



AnyV2V



VACE



HunyuanCustom

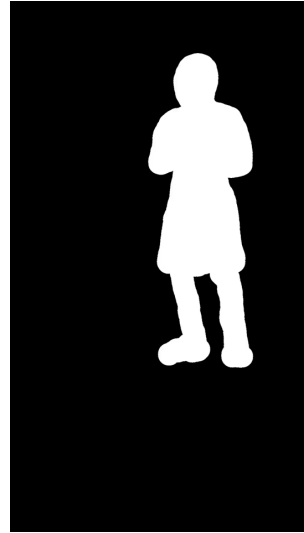


Kling 1.6

# Human – full body



source video

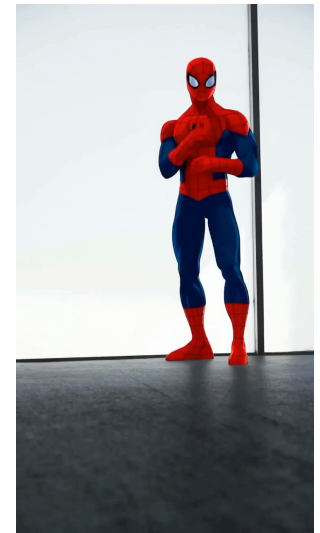


mask



reference

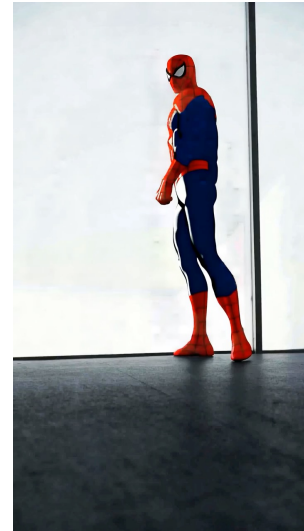
# Video Comparisons



DreamSwapV (ours)



AnyV2V



VACE



HunyuanCustom



Kling 1.6

# Human – half body

# Video Comparisons



source video



mask



reference



DreamSwapV (ours)



AnyV2V



VACE



HunyuanCustom



Kling 1.6

# Human – half body

# Video Comparisons



source video



mask



reference



DreamSwapV (ours)



AnyV2V



VACE



HunyuanCustom



Kling 1.6

# Human – talking head

# Video Comparisons



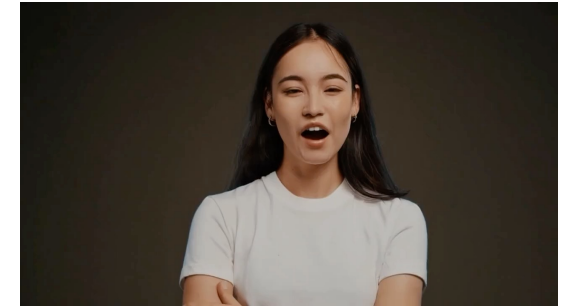
source video



mask



reference



DreamSwapV (ours)



AnyV2V



VACE



HunyuanCustom



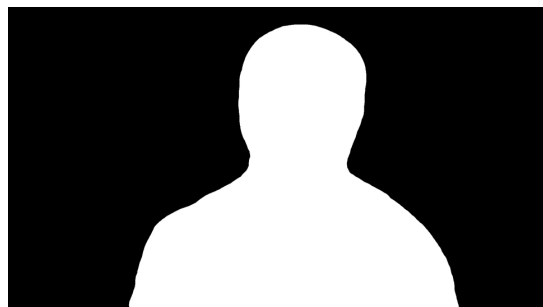
Kling 1.6

# Human – talking head

# Video Comparisons



source video



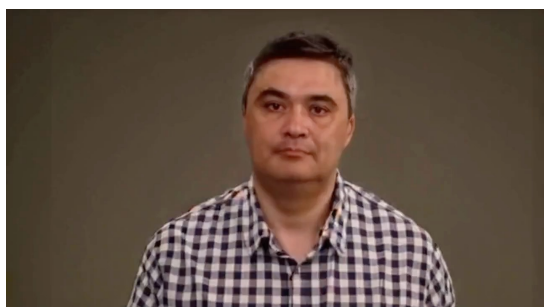
mask



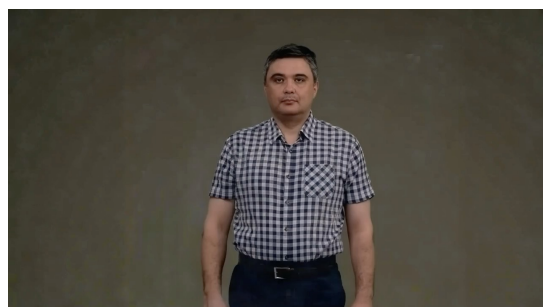
reference



DreamSwapV (ours)



AnyV2V



VACE



HunyuanCustom

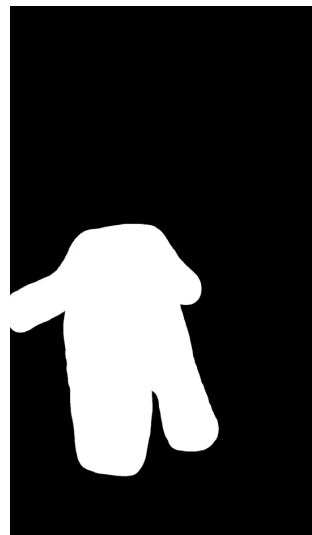


Kling 1.6

# Garment - upper



source video



mask



reference



DreamSwapV (ours)



AnyV2V



VACE



HunyuanCustom



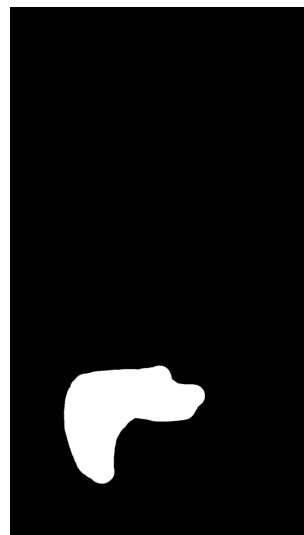
Kling 1.6

# Video Comparisons

# Garment - lower



source video



mask



reference

# Video Comparisons



DreamSwapV (ours)



AnyV2V



VACE



HunyuanCustom

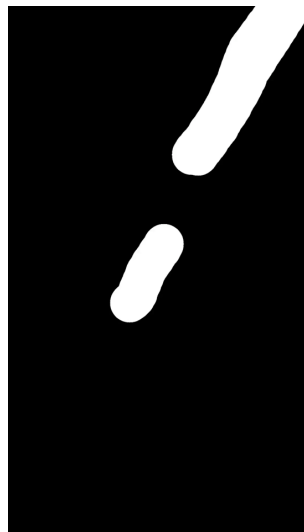


Kling 1.6

# Small Object - handheld



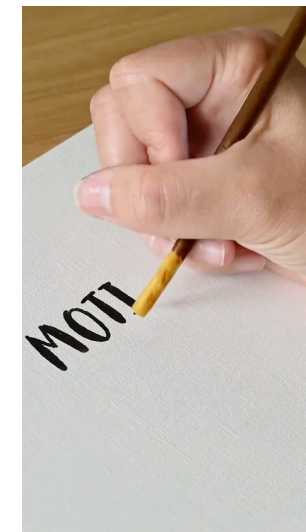
source video



mask



reference



DreamSwapV (ours)



AnyV2V



VACE



HunyuanCustom



Kling 1.6

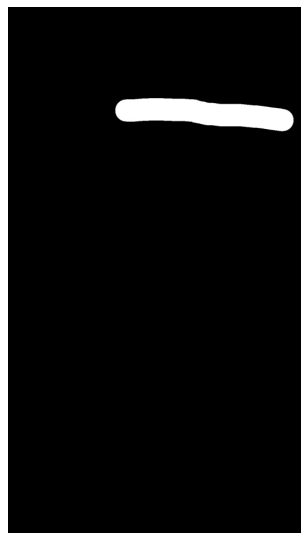
# Video Comparisons

# Small Object - handheld

# Video Comparisons



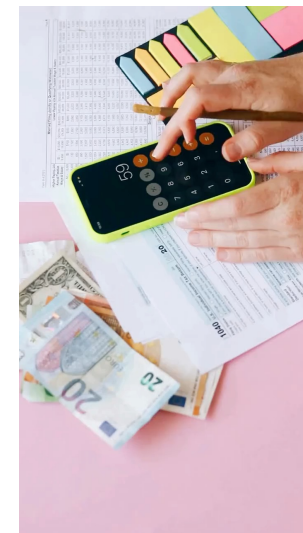
source video



mask



reference



DreamSwapV (ours)



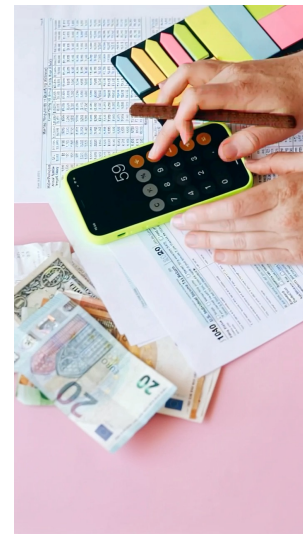
AnyV2V



VACE



HunyuanCustom



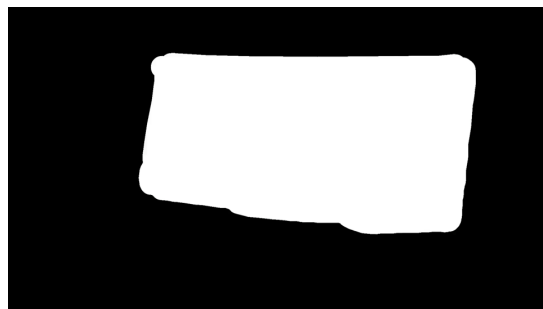
Kling 1.6

# Small Object - handheld

# Video Comparisons



source video



mask



reference



DreamSwapV (ours)



AnyV2V



VACE



HunyuanCustom

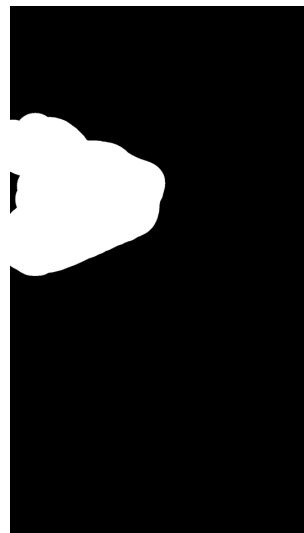


Kling 1.6

# Small Object - others



source video



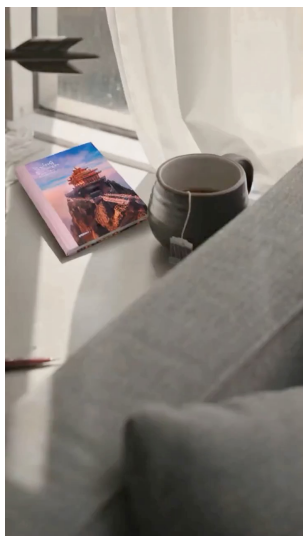
mask



reference



DreamSwapV (ours)



AnyV2V



VACE



HunyuanCustom



Kling 1.6

# Video Comparisons

# Small Object - others

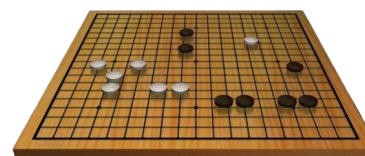
# Video Comparisons



source video



mask



reference



DreamSwapV (ours)



AnyV2V



VACE



HunyuanCustom

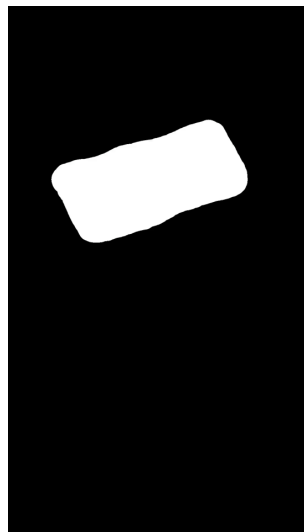


Kling 1.6

# Small Object - others



source video



mask



reference



DreamSwapV (ours)



AnyV2V



VACE



HunyuanCustom



Kling 1.6

# Video Comparisons

# Large Object - statue

# Video Comparisons



source video



mask



reference



DreamSwapV (ours)



AnyV2V



VACE



HunyuanCustom



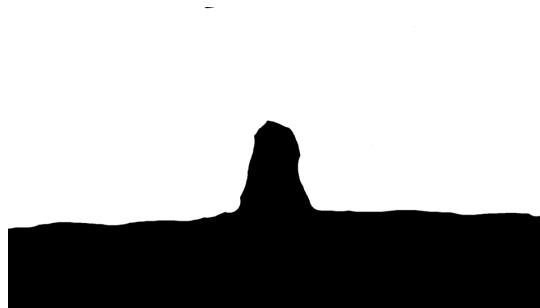
Kling 1.6

# Large Object - sky

# Video Comparisons



source video



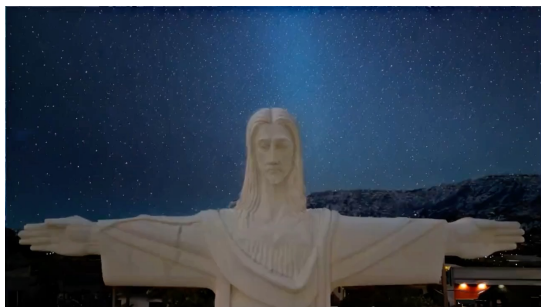
mask



reference



DreamSwapV (ours)



AnyV2V



VACE



HunyuanCustom



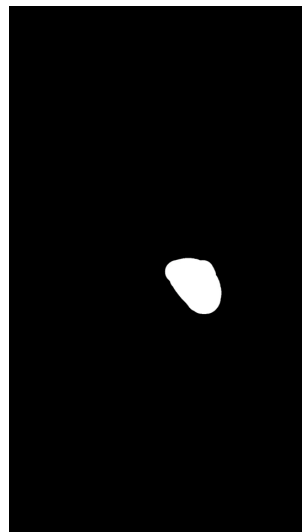
Kling 1.6

# Large Object - vehicle

# Video Comparisons



source video



mask



reference



DreamSwapV (ours)



AnyV2V



VACE



HunyuanCustom



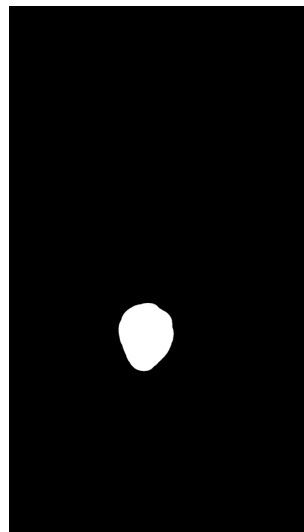
Kling 1.6

# Large Object - vehicle

# Video Comparisons



source video



mask



reference



DreamSwapV (ours)



AnyV2V



VACE



HunyuanCustom

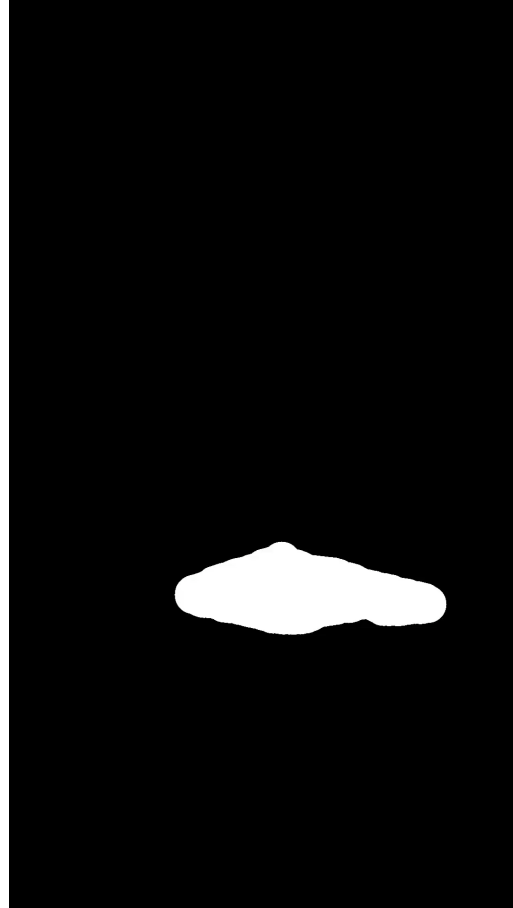


Kling 1.6

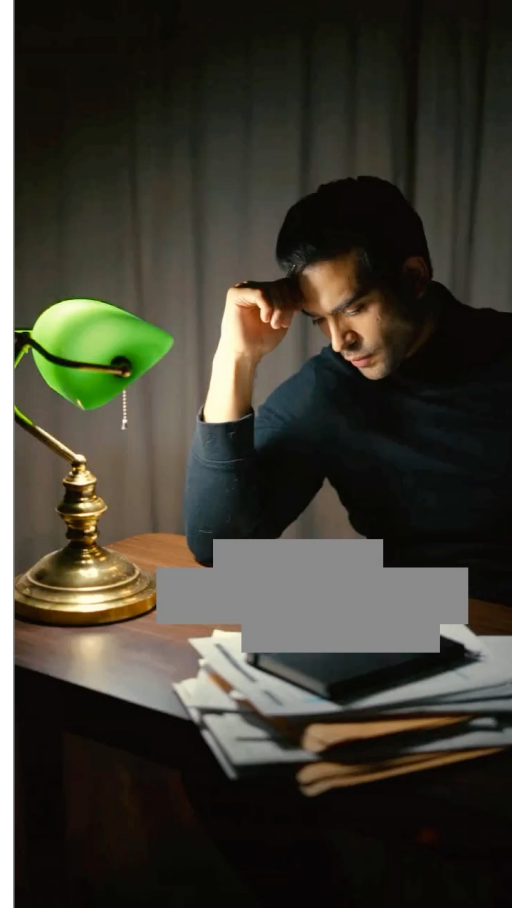
# Video Inpainting



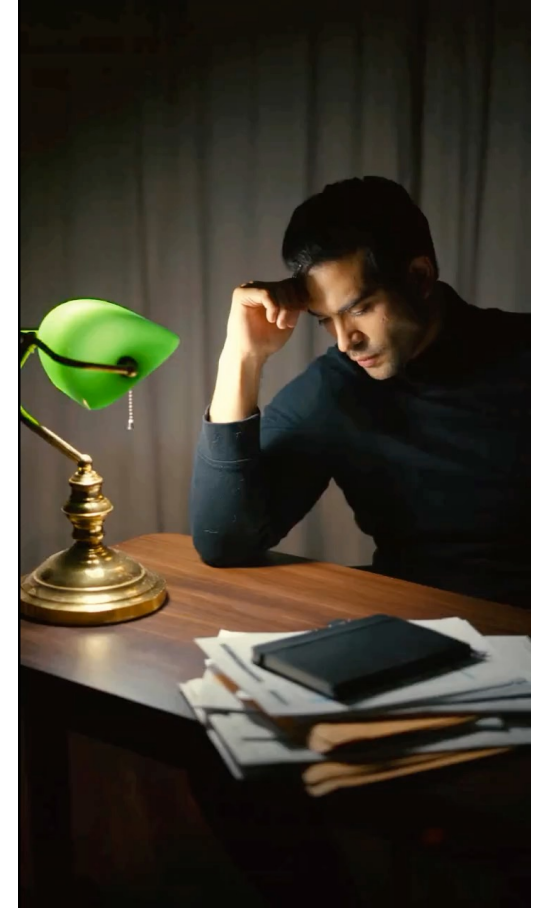
source video



mask



agnostic



DreamSwapV inpainting

# More Applications

# Video Inpainting



source video



mask



agnostic



DreamSwapV inpainting

# More Applications

# Video Addition

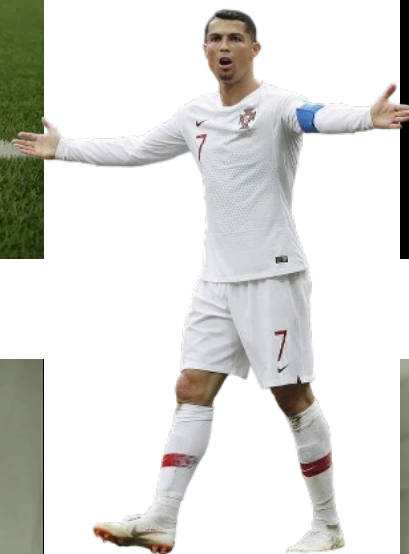
# More Applications



source video



agnostic



reference



mask



DreamSwapV addition

## Video Addition



source video



agnostic



reference

## More Applications



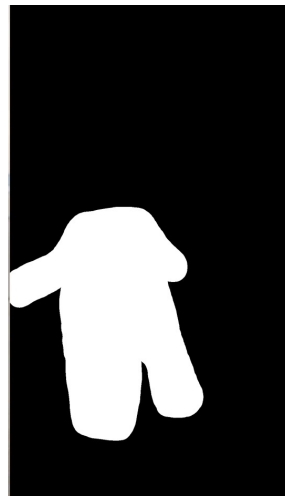
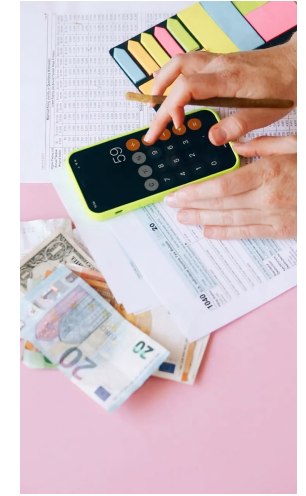
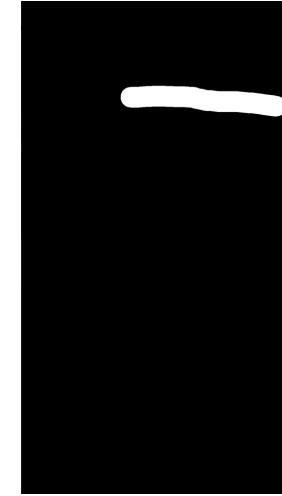
mask



DreamSwapV addition

# Long Video Extrapolation

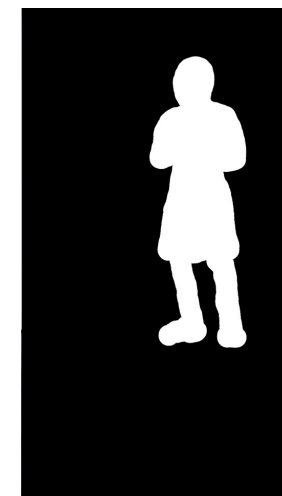
Thanks to our dummy reference design, DreamSwapV can also support subject swapping for **any-length video** by long video extrapolation via overlapped dummy reference.



whole video

mask

DreamSwapV



whole video

mask

DreamSwapV

*DreamSwapV*

*Thanks for watching!*