

SV4D : Dynamic 3D Content Generation with Multi-Frame and Multi-View **Consistency**

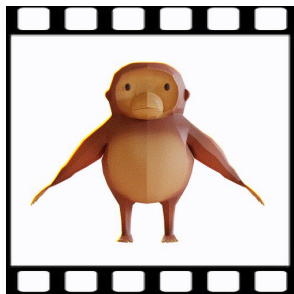
Yiming Xie*, Chun-Han Yao*, Vikram Voleti, Huaizu Jiang^, Varun Jampani^

* Equal Contribution ^ Equal Advising

stability.ai



Dynamic 3D Content (4D) Generation



Input: Single-view Video

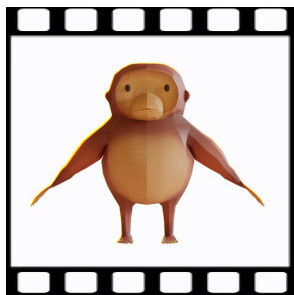
4D Generation



Output: 4D Representation
(Dynamic 3D representation, e.g, NeRF, Gaussian Splatting, Mesh...)

.....

Dynamic 3D Content (4D) Generation



Input: Single-view Video

4D Generation



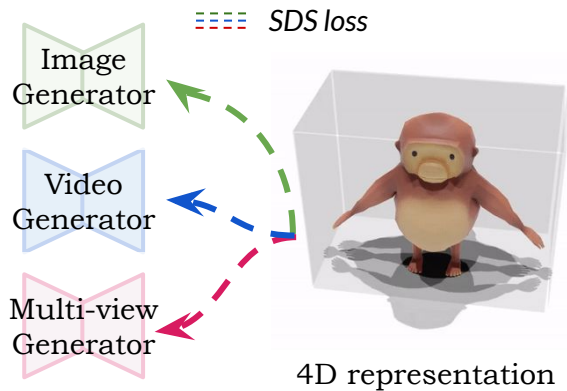
Output: 4D Representation
(Dynamic 3D representation, e.g, NeRF, Gaussian Splatting, Mesh...)

Challenges:

- The higher dimensional nature of 4D generation.
- No large scale datasets with 4D objects to train a robust generative model.

Dynamic 3D Content (4D) Generation

Related Works



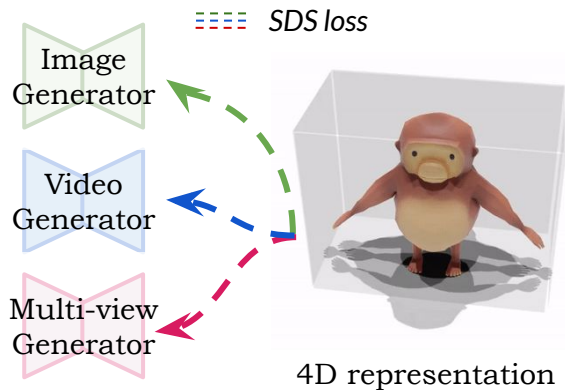
SDS [1] based optimization

- Time-consuming: take hours to generate a single 4D object.
- Unstable optimization

MAV3D, Consistent4D, STAG4D, 4DGen

Dynamic 3D Content (4D) Generation

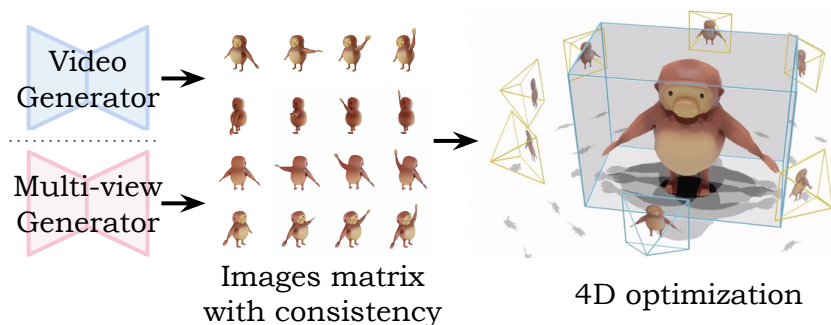
Related Works



SDS [1] based optimization

- Time-consuming: take hours to generate a single 4D object.
- Unstable optimization

MAV3D, Consistent4D, STAG4D, 4DGen

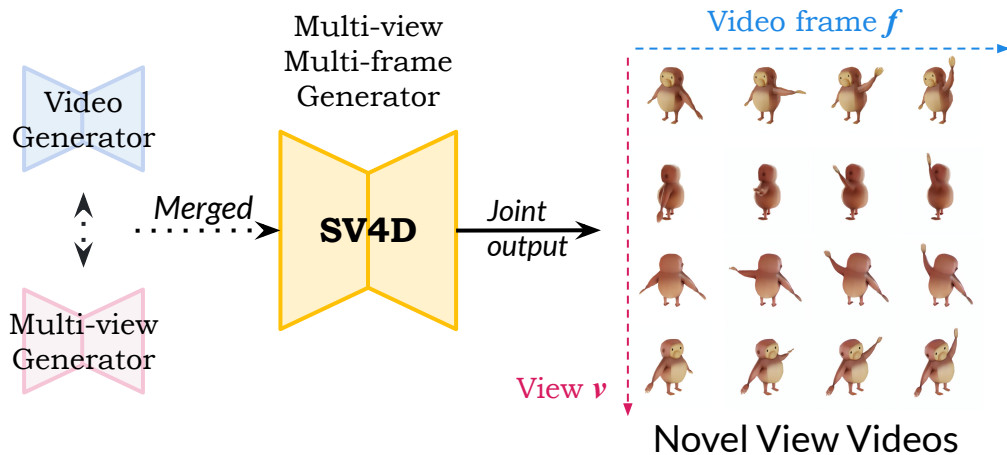


Photogrammetry-based methods

- Several inconsistencies still remain due to the use of separate video and multi-view generative models

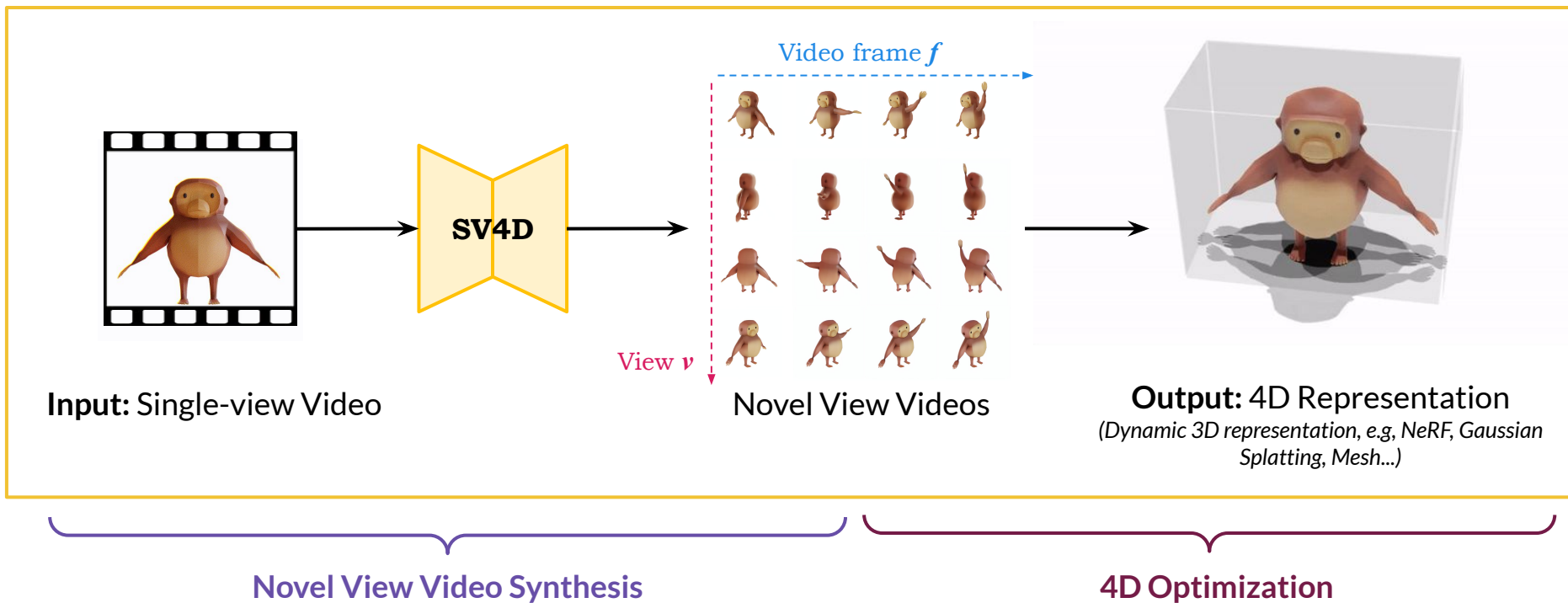
Diffusion²

Our Solution



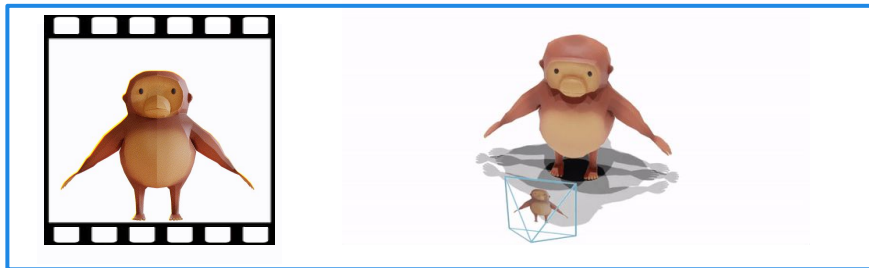
- State-of-the-art multi-frame and multi-view consistency.

Our Solution



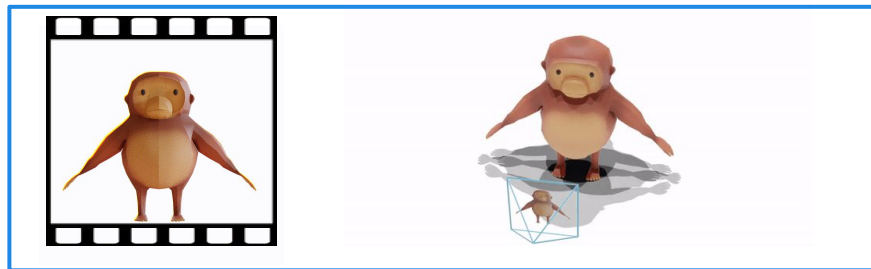
Novel View Video Synthesis

Input Reference Video

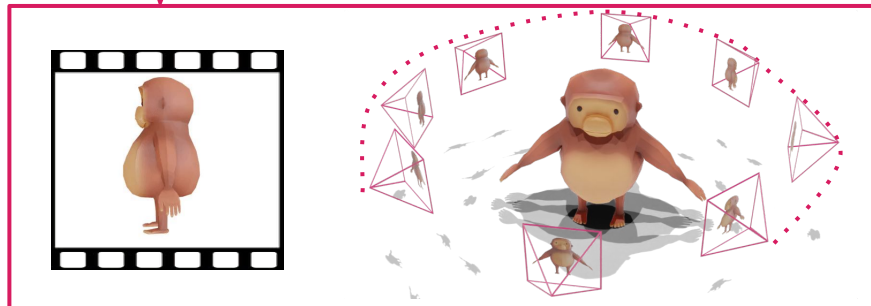


Novel View Video Synthesis

Input Reference Video



Multi-view
Generator

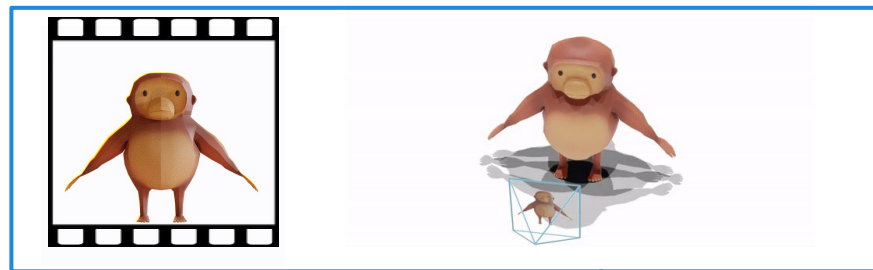


Reference Multi-view

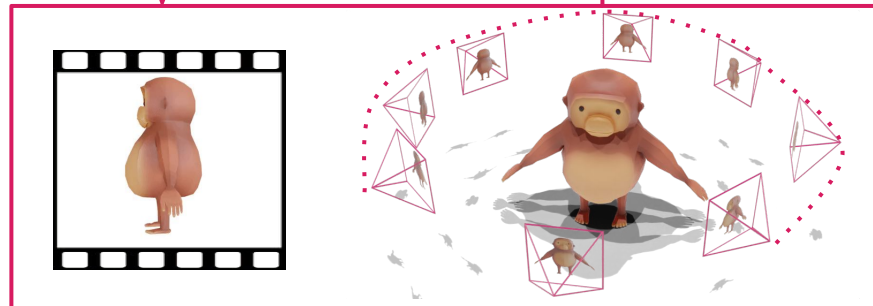
GT Mesh for illustration purpose only

Novel View Video Synthesis

Input Reference Video



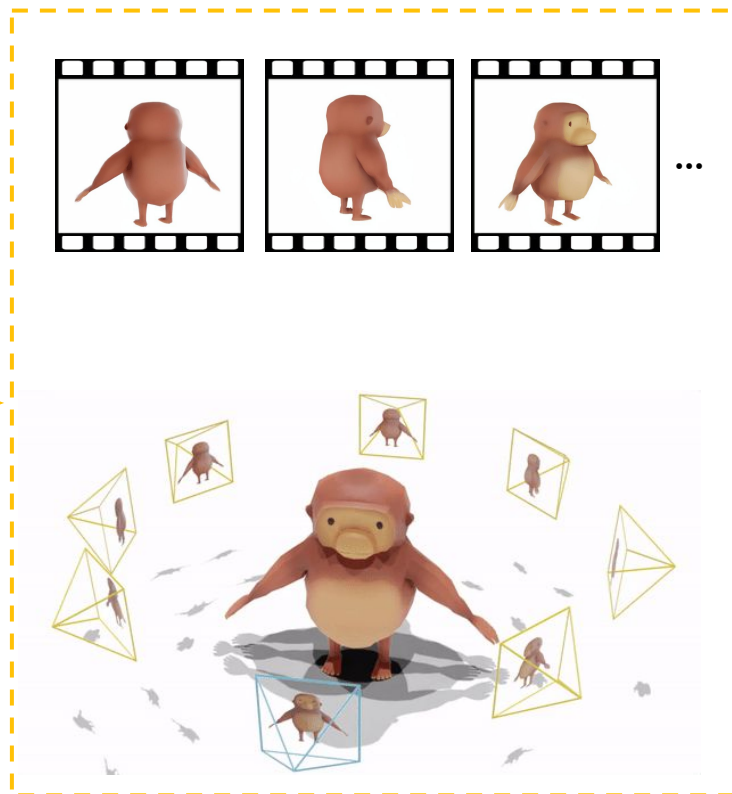
Multi-view
Generator



Reference Multi-view

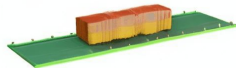
SV4D

Novel View Videos



GT Mesh for illustration purpose only

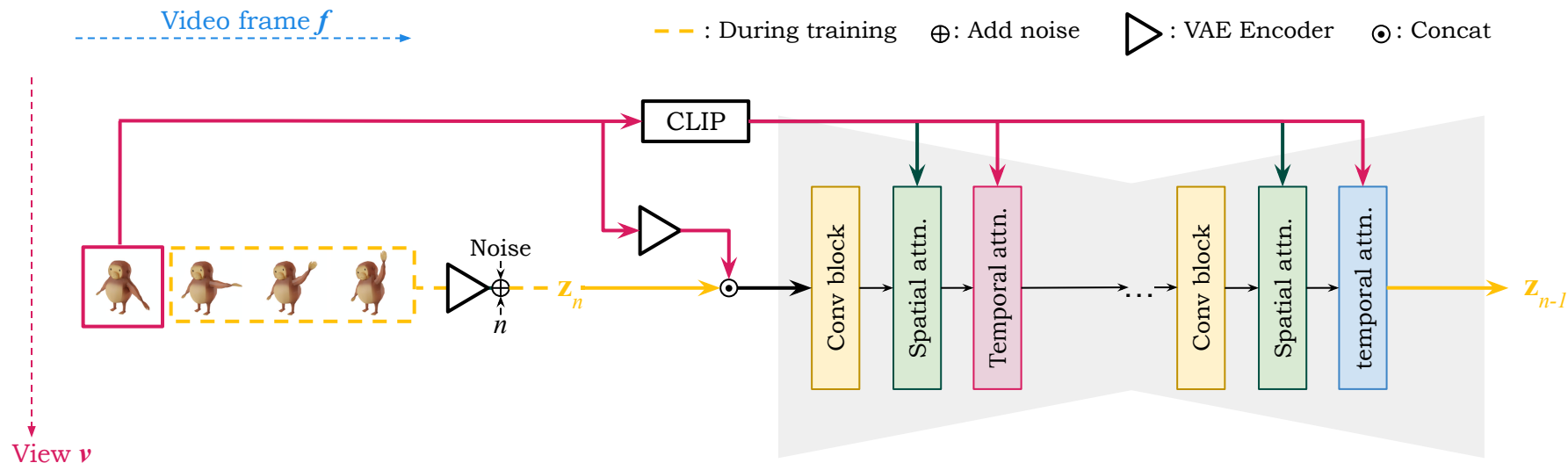
Novel View Video Synthesis



Novel View Video Synthesis

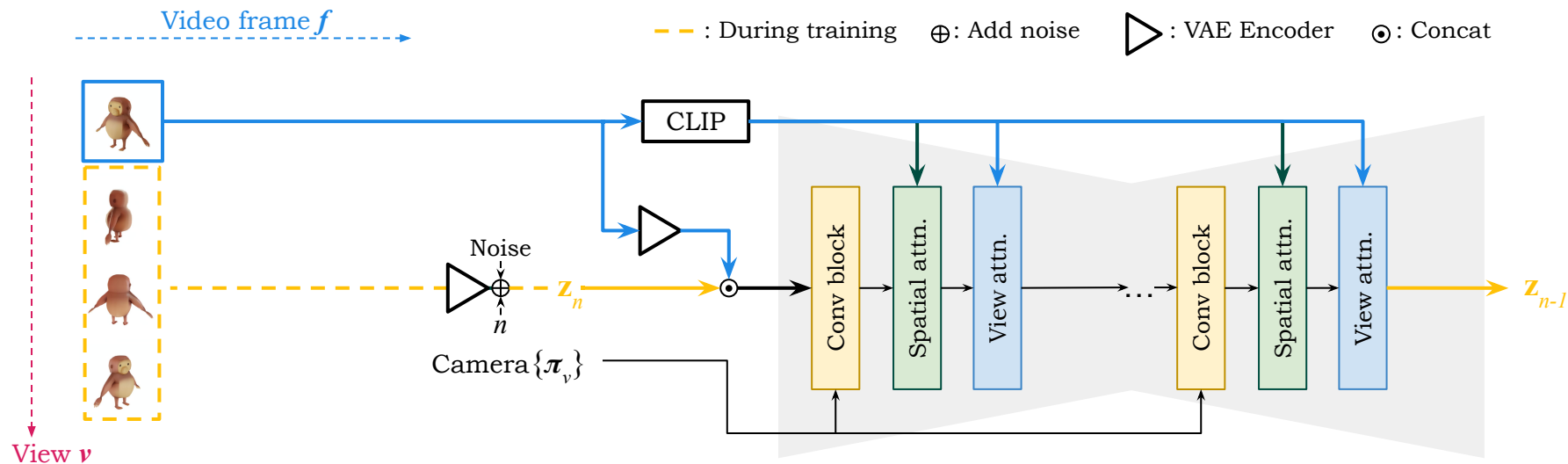
Novel View Video Synthesis

Stable Video Diffusion (SVD)



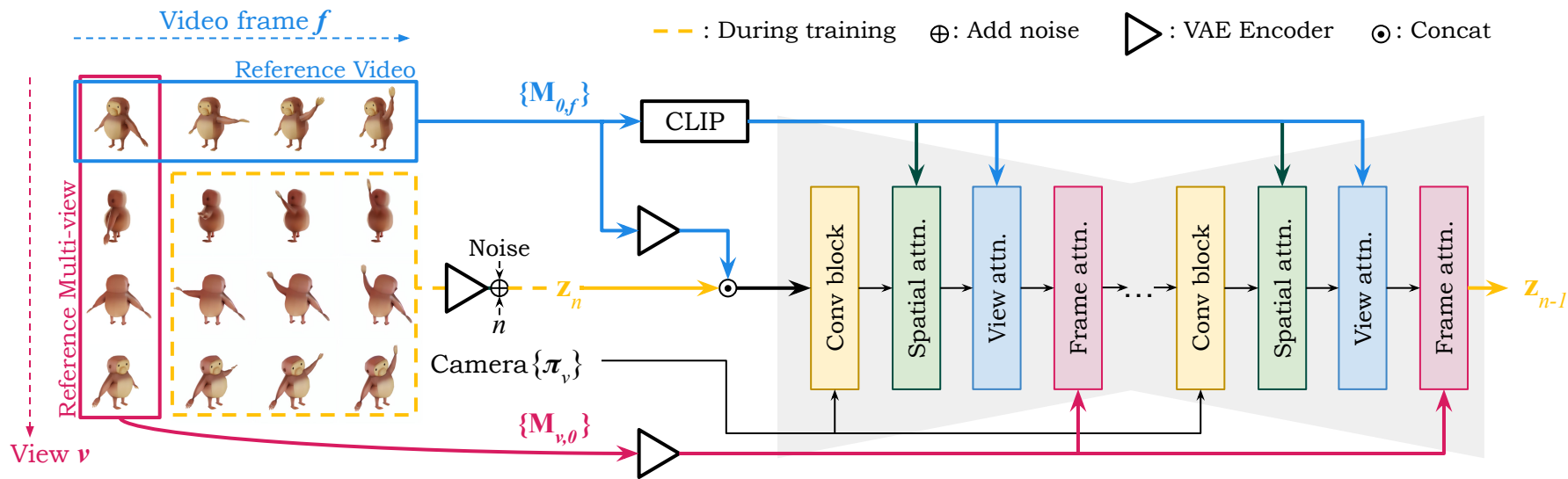
Novel View Video Synthesis

Stable Video Diffusion 3D (SV3D)



Novel View Video Synthesis

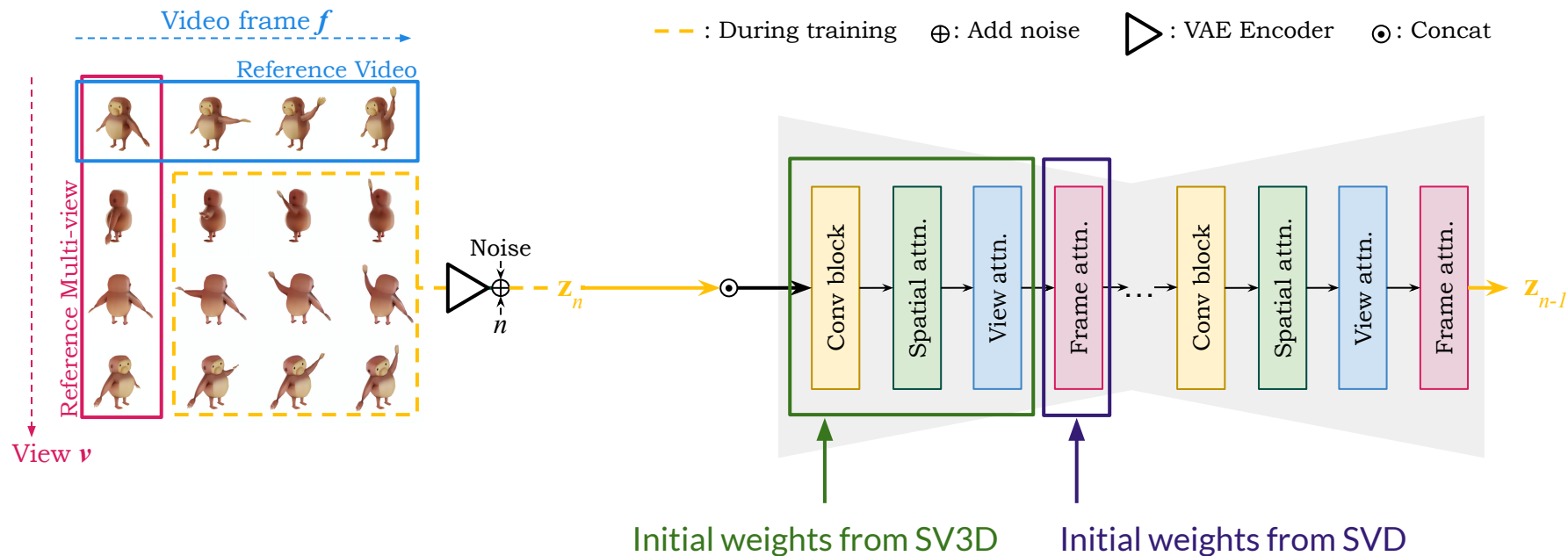
Stable Video Diffusion 4D (SV4D)



Novel View Video Synthesis

Training Details

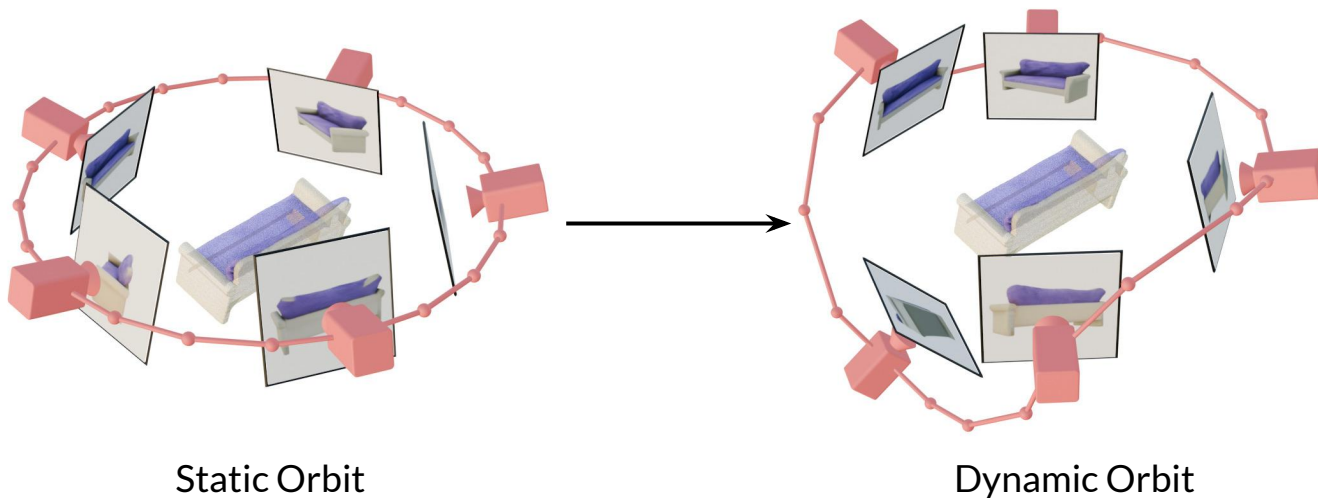
Resume weights from both SV3D and SVD



Novel View Video Synthesis

Training Details

Two-stage fine tuning



Novel View Video Synthesis

Training Details

Training Dataset

Filtering

- Inappropriate licenses
- Too few animated frames
- Small movement

Rendering

- Dynamically adjust frame sampling step for each object
- Dynamically adapt camera distance
- Remove global motion

Improve data quality
→

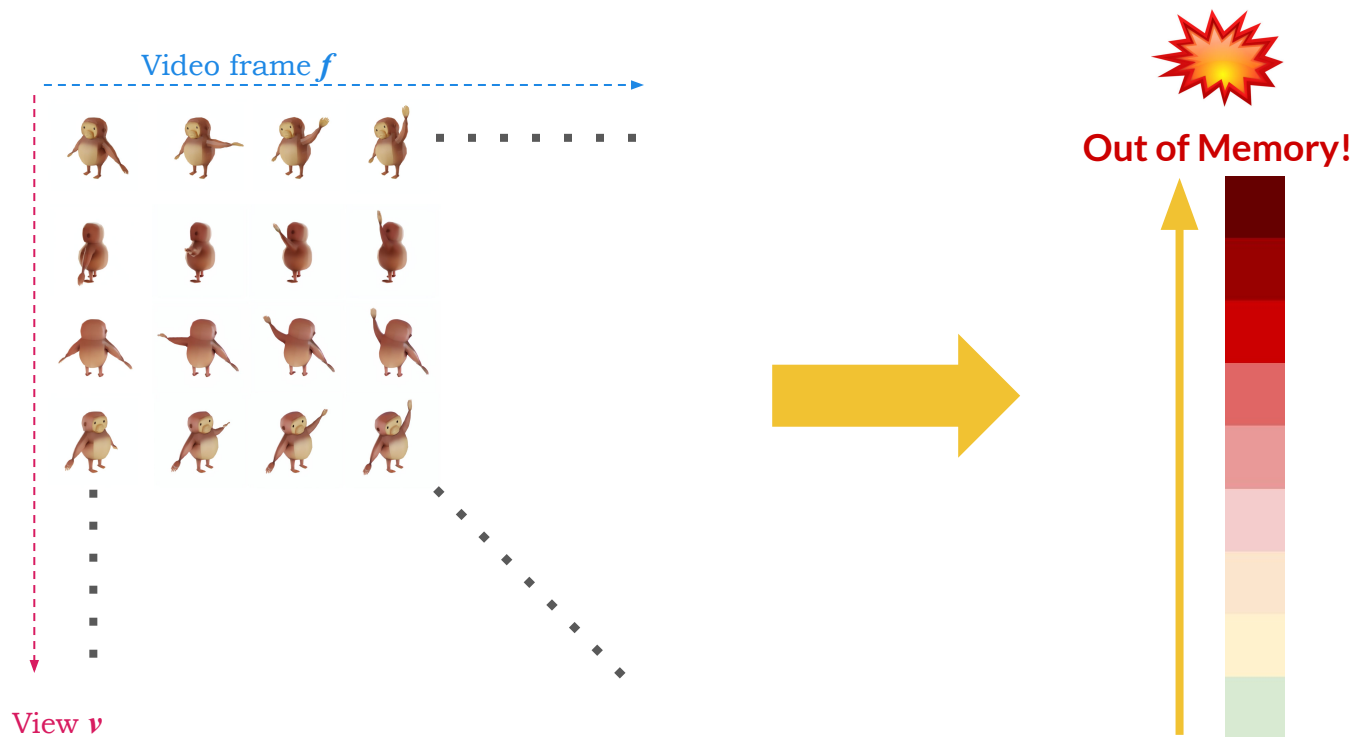


Objaverse Dataset

includes over 44K animated 3D objects.

Novel View Video Synthesis

Generation for Arbitrary Length Videos

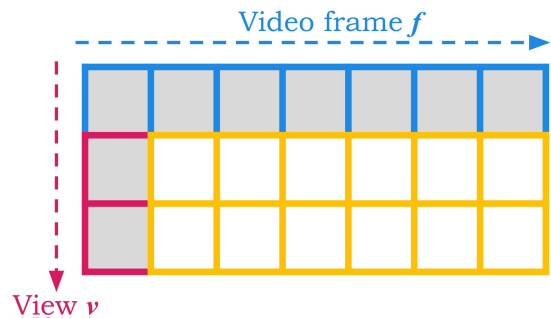
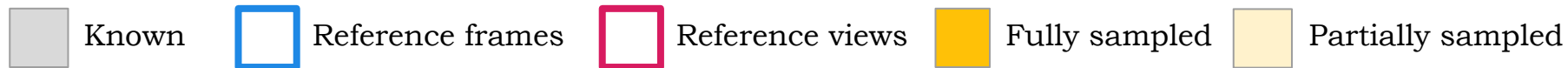


Large image matrix with long input video / view

GPU Memory

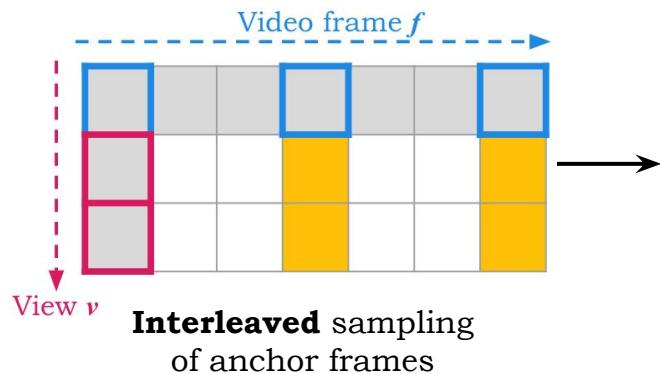
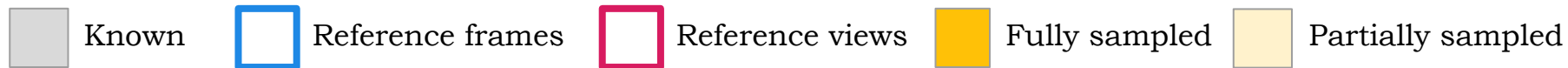
Novel View Video Synthesis

Generation for Arbitrary Length Videos



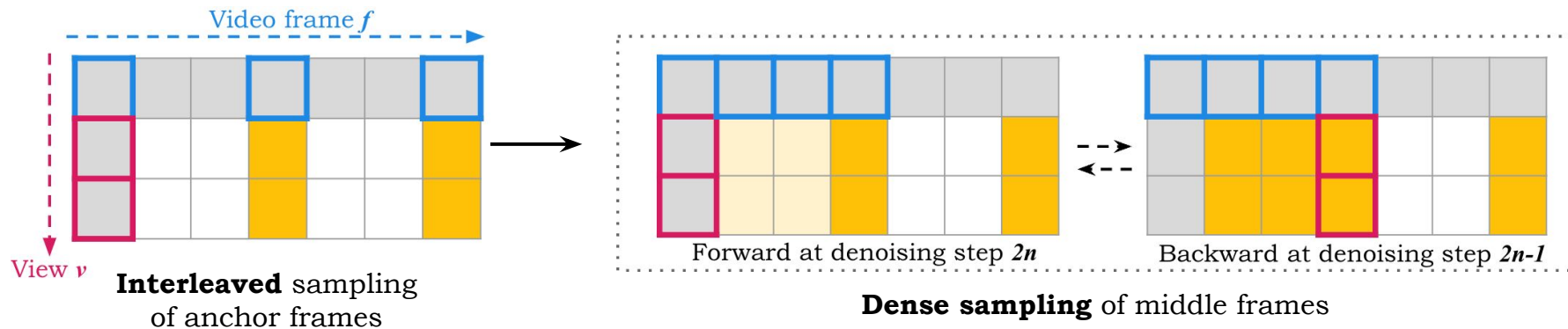
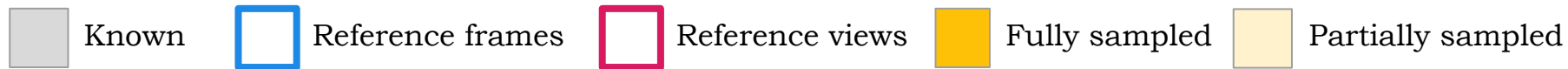
Novel View Video Synthesis

Generation for Arbitrary Length Videos



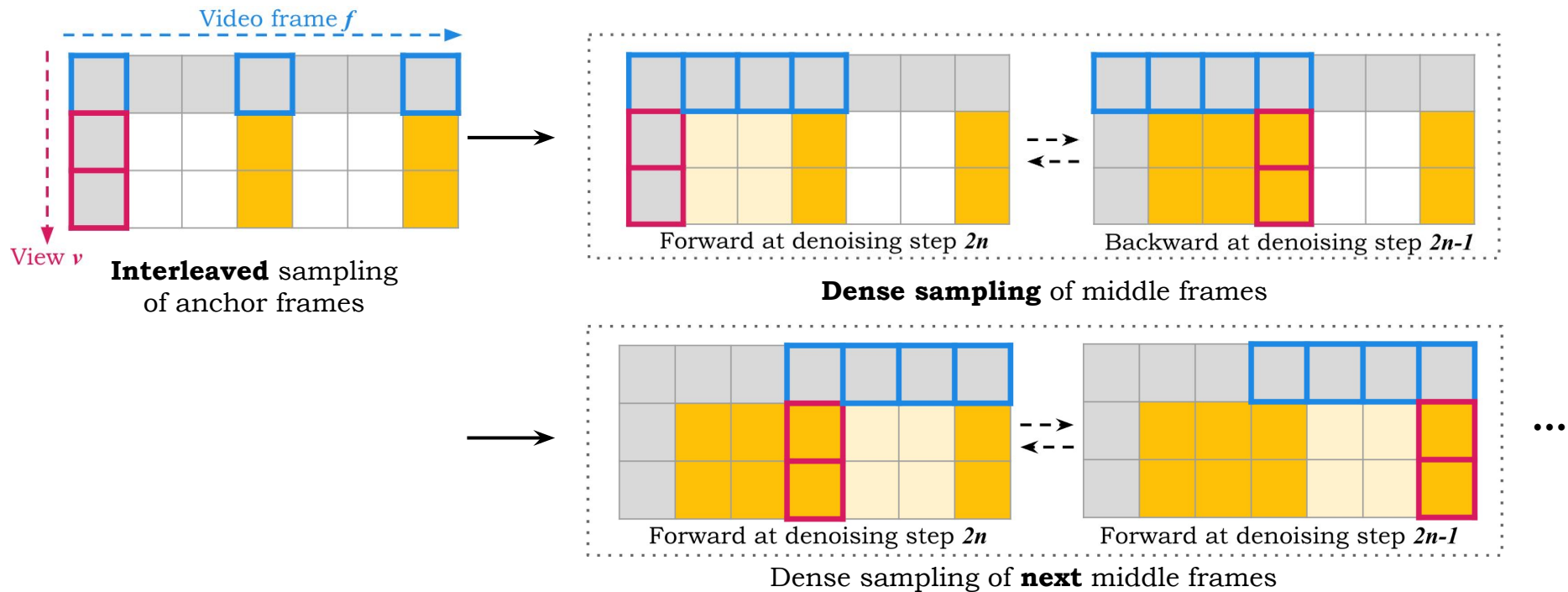
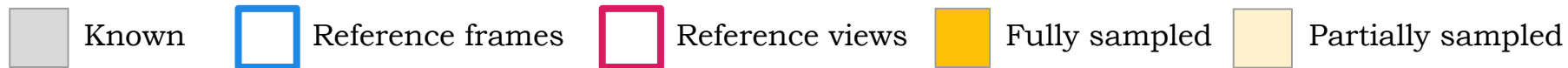
Novel View Video Synthesis

Generation for Arbitrary Length Videos

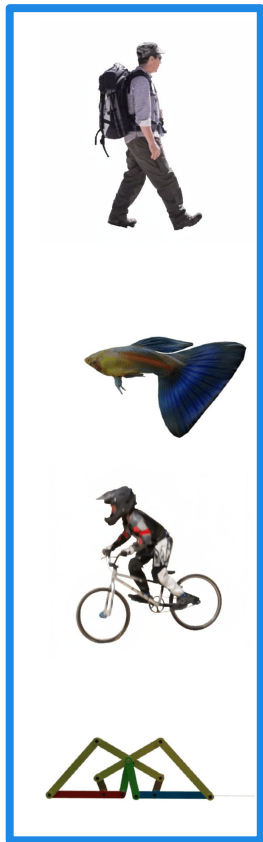


Novel View Video Synthesis

Generation for Arbitrary Length Videos



Novel View Video Synthesis



Input Video



Diffusion²



STAG4D



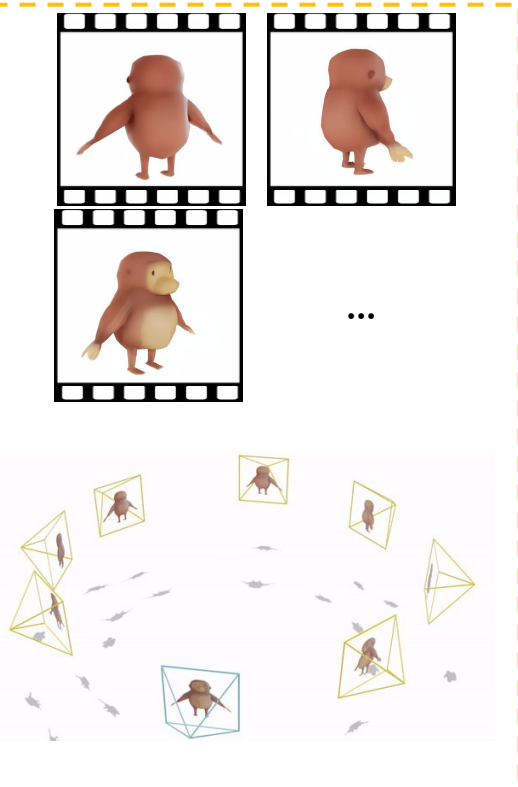
Stable Video 3D



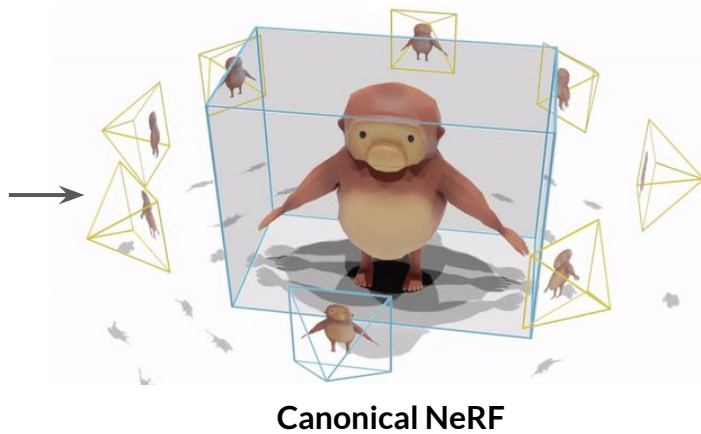
Stable Video 4D (Ours)

4D Optimization

Novel View Videos



4D Optimization



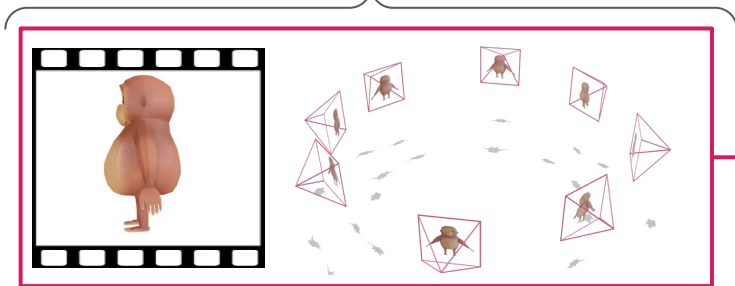
Generated 4D Assets



4D Optimization

Dynamic NeRF representation

Stage 1 (static)



Reference Multi-view

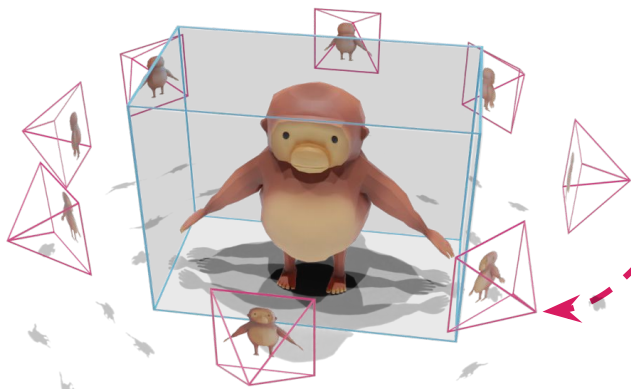


Image + Geometry
Losses

4D Optimization

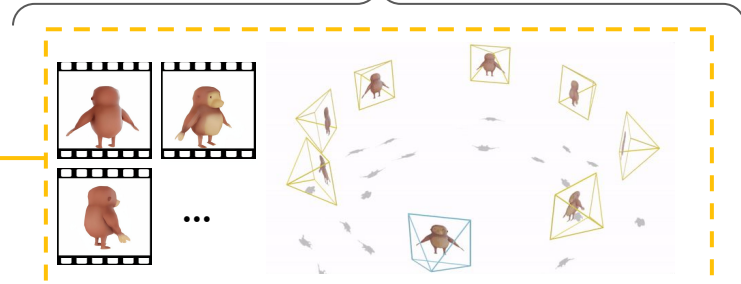
Dynamic NeRF representation

Stage 1 (static)



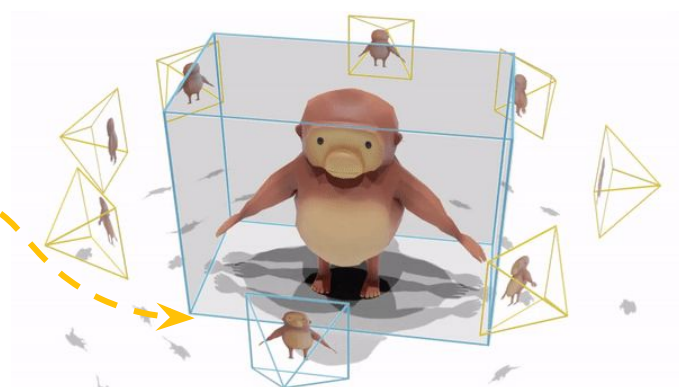
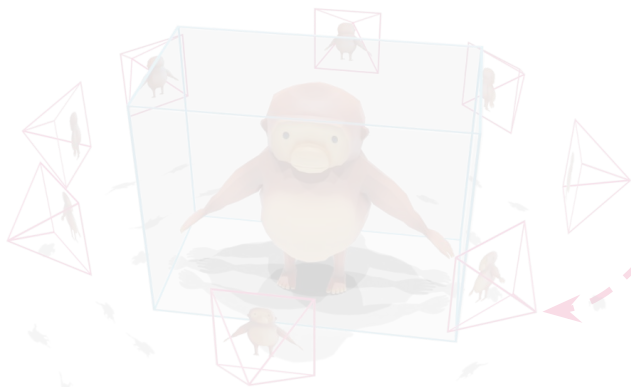
Reference Multi-view

Stage 2 (dynamic)

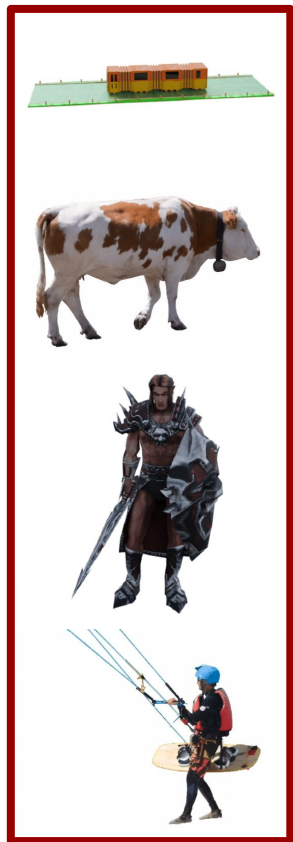


Random Views & Frames

Image + Geometry
Losses



4D Generation



Input Video



Consistent4D



STAG4D

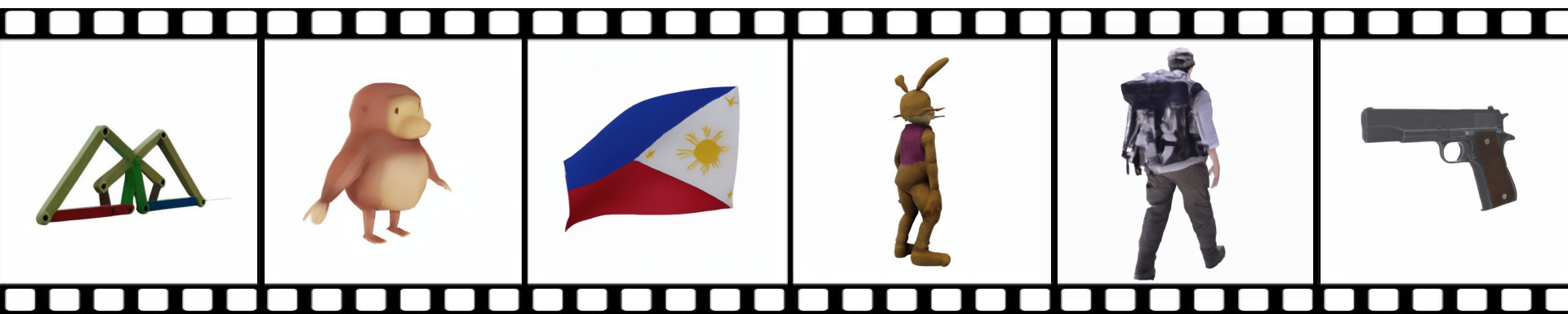


DreamGaussian4D



Stable Video 4D (Ours)

Take-Away Message



- SV4D can **simultaneously** generate **multi-view** and **multi-frame** images.
- SV4D sampling strategy enables sequential processing of **arbitrary long input videos**.