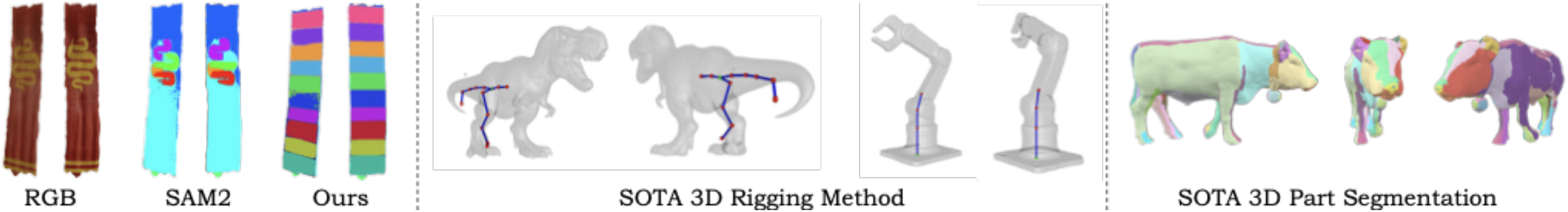
Stable Part Diffusion 4D: Multi-View RGB and Kinematic Parts Video Generation

stability.ai

ILLINOIS

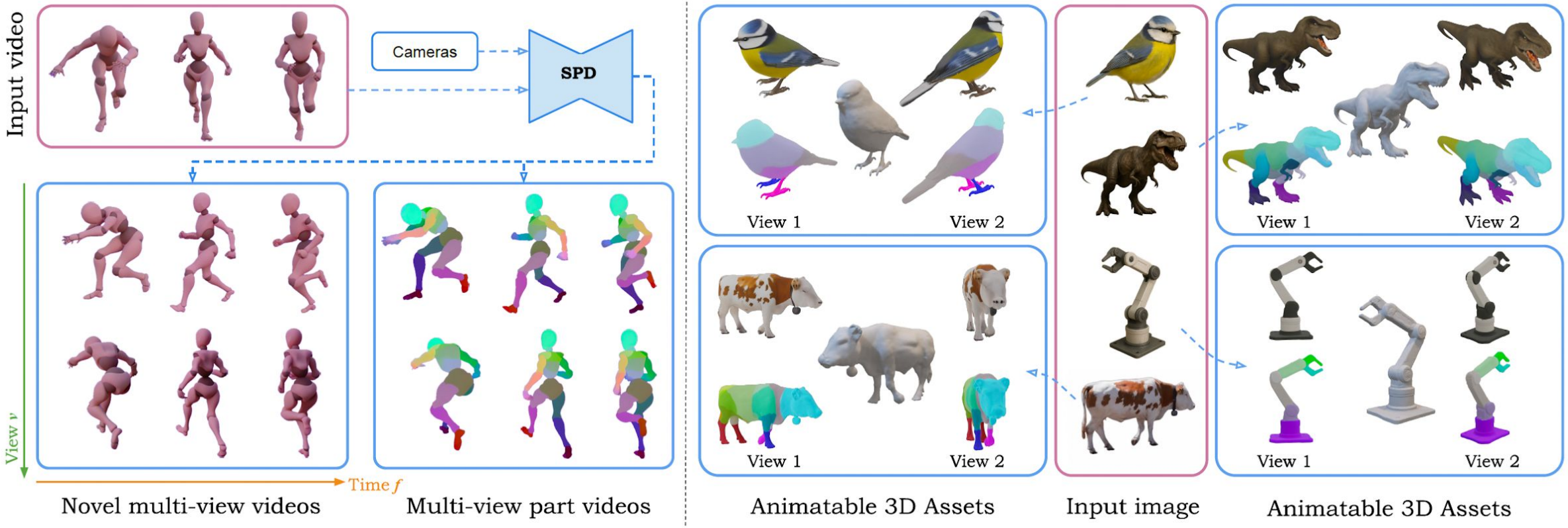
Motivation:



- **Why Not Semantic Parts?**
  Traditional segmentation (e.g., “leg”, “tail”) ignores how objects **actually move**.
  👉 *SP4D captures physically meaningful **kinematic parts*** — crucial for animation.

- **Why Not Existing 3D Rigging Methods?**
  Most rely on **small datasets**, limiting generalization.
  👉 *SP4D learns directly from **large-scale videos and images*** — scalable and robust.

Stable Part Diffusion:

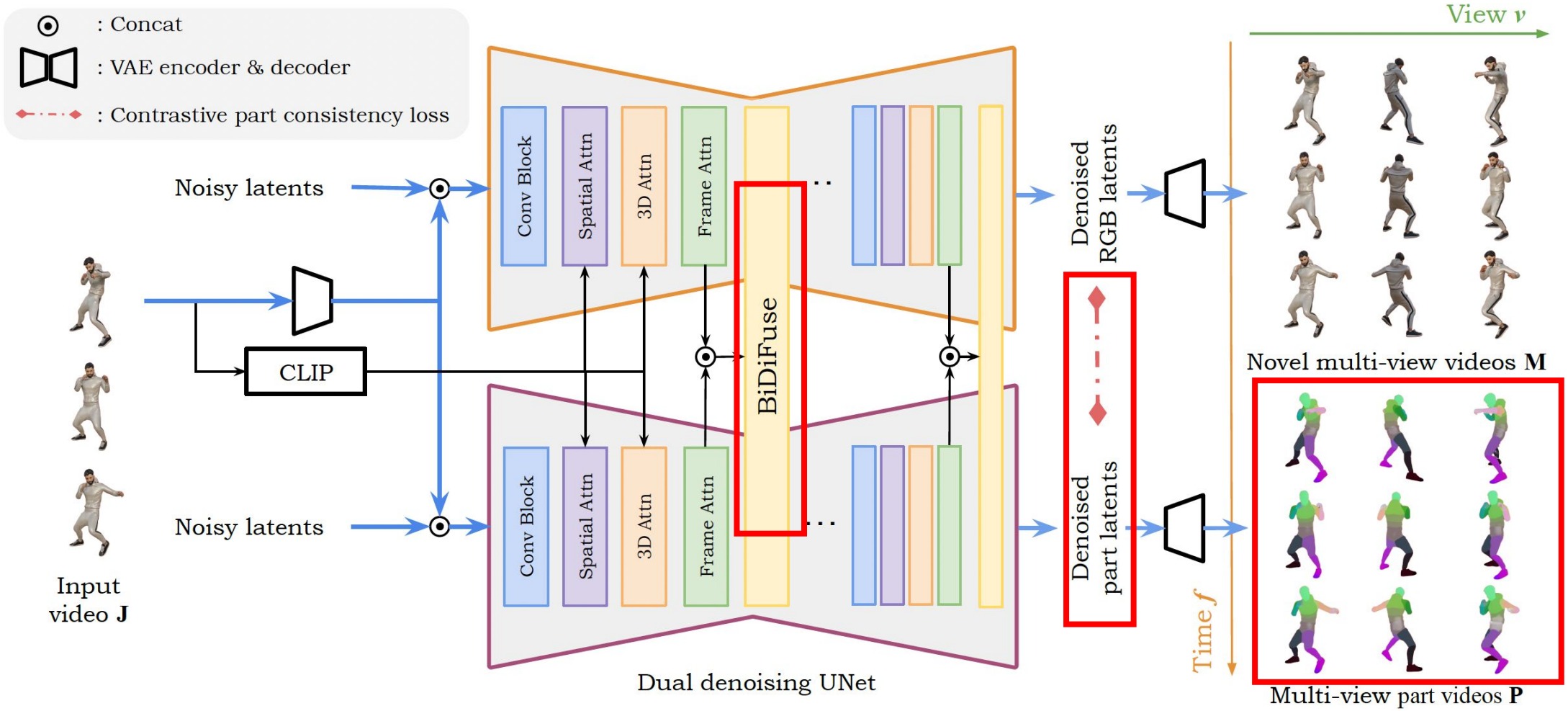


Input: Single image or monocular video

Output:

- Multi-frame, multi-view **RGB videos**
- Matching **kinematic part segmentation maps**

Stable Part Diffusion:

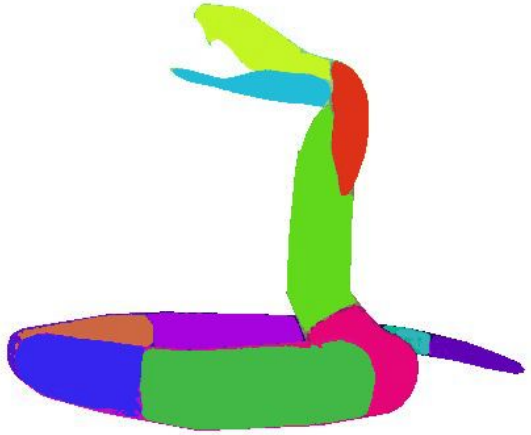


Results:

- Fixed-view Cross-frame Tracking

the RGB video is provided as input and the corresponding Part Video is generated.

RGB



Parts

RGB

Parts

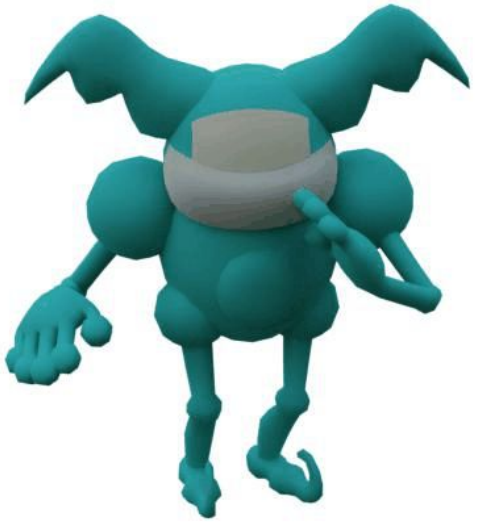
Results:

- Fixed-view Cross-frame Tracking

the RGB video is provided as input and the corresponding Part Video is generated.

RGB

Parts



RGB

Parts

Results:

- Fixed-view Cross-frame Tracking

the RGB video is provided as input and the corresponding Part Video is generated.

RGB

Parts

RGB

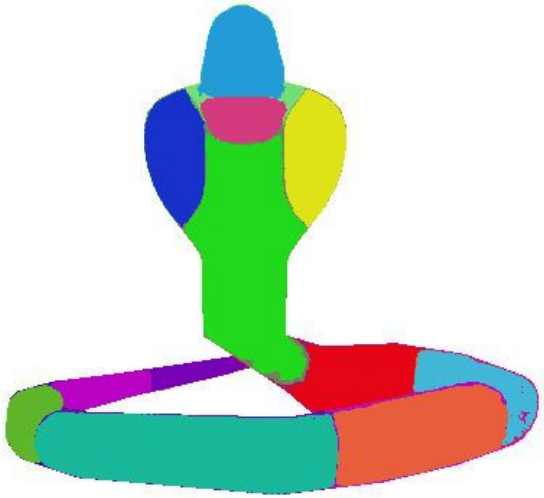
Parts

Results:

- Fixed-frame Cross-view Tracking

  the first frame of the RGB video serves as the input, while the remaining frames and the entire Part Video are generated.

RGB



Parts

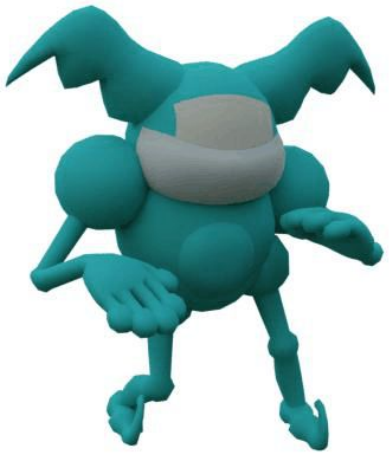
RGB

Parts

Results:

- Fixed-frame Cross-view Tracking

  the first frame of the RGB video serves as the input, while the remaining frames and the entire Part Video are generated.



RGB

Parts

RGB

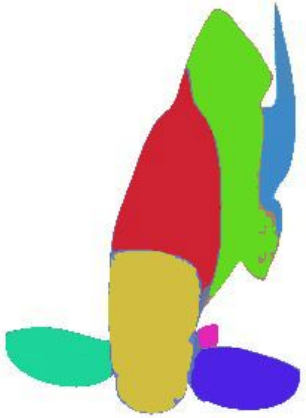
Parts

Results:

- Fixed-frame Cross-view Tracking

  the first frame of the RGB video serves as the input, while the remaining frames and the entire Part Video are generated.

RGB



Parts

RGB

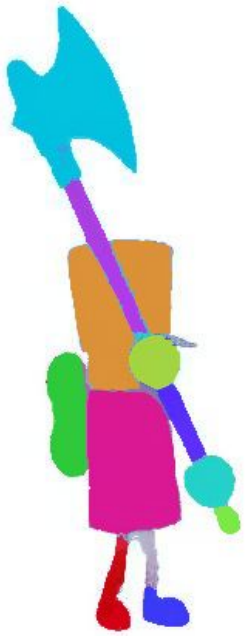
Parts

Results:

- Fixed-frame Cross-view Tracking

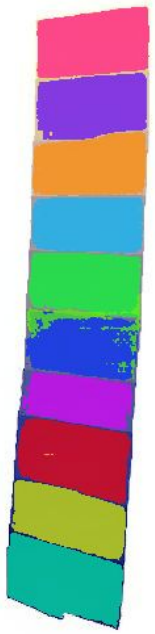
  the first frame of the RGB video serves as the input, while the remaining frames and the entire Part Video are generated.

RGB



Parts

RGB



Parts

RGB

Parts

Results:

- Fixed-frame Cross-view Tracking

  the first frame of the RGB video serves as the input, while the remaining frames and the entire Part Video are generated.

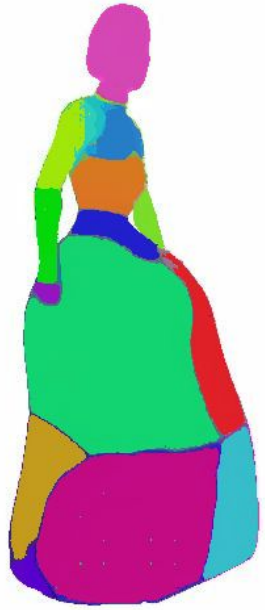
RGB

Parts

RGB

Parts

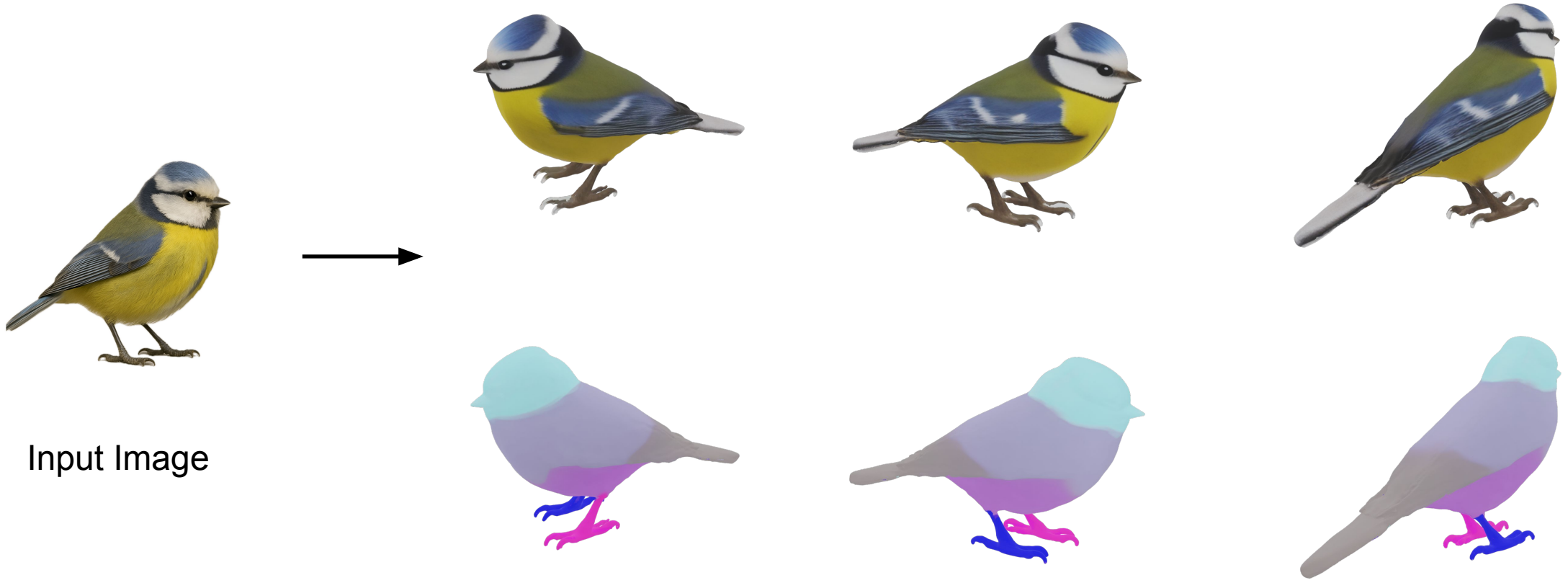
RGB



Parts

Results:

- 3D Part Decomposition for zero-shot input image



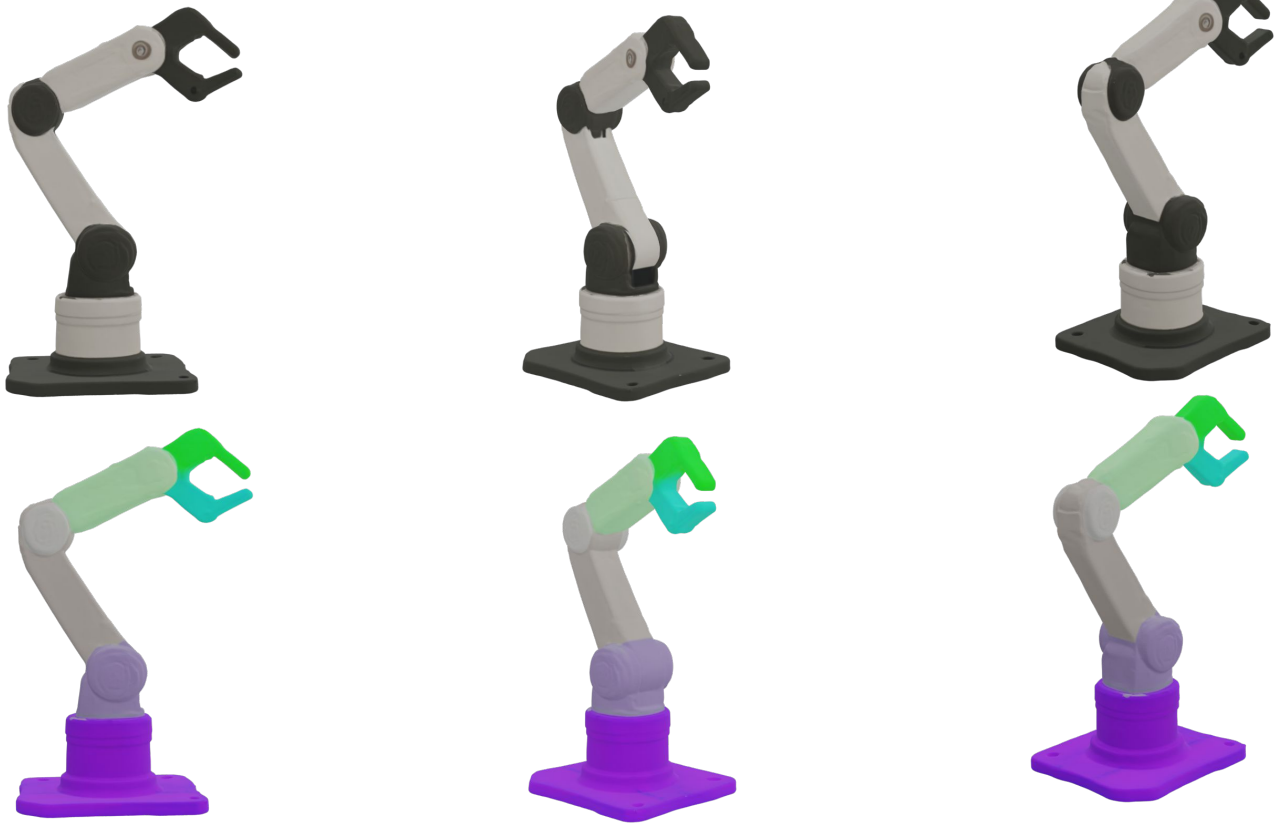
Input Image

Generated 3D object with part decomposition

Results:

- 3D Part Decomposition for zero-shot input image

Input Image



Generated 3D object with part decomposition

Results:

- 3D Part Decomposition for zero-shot input image

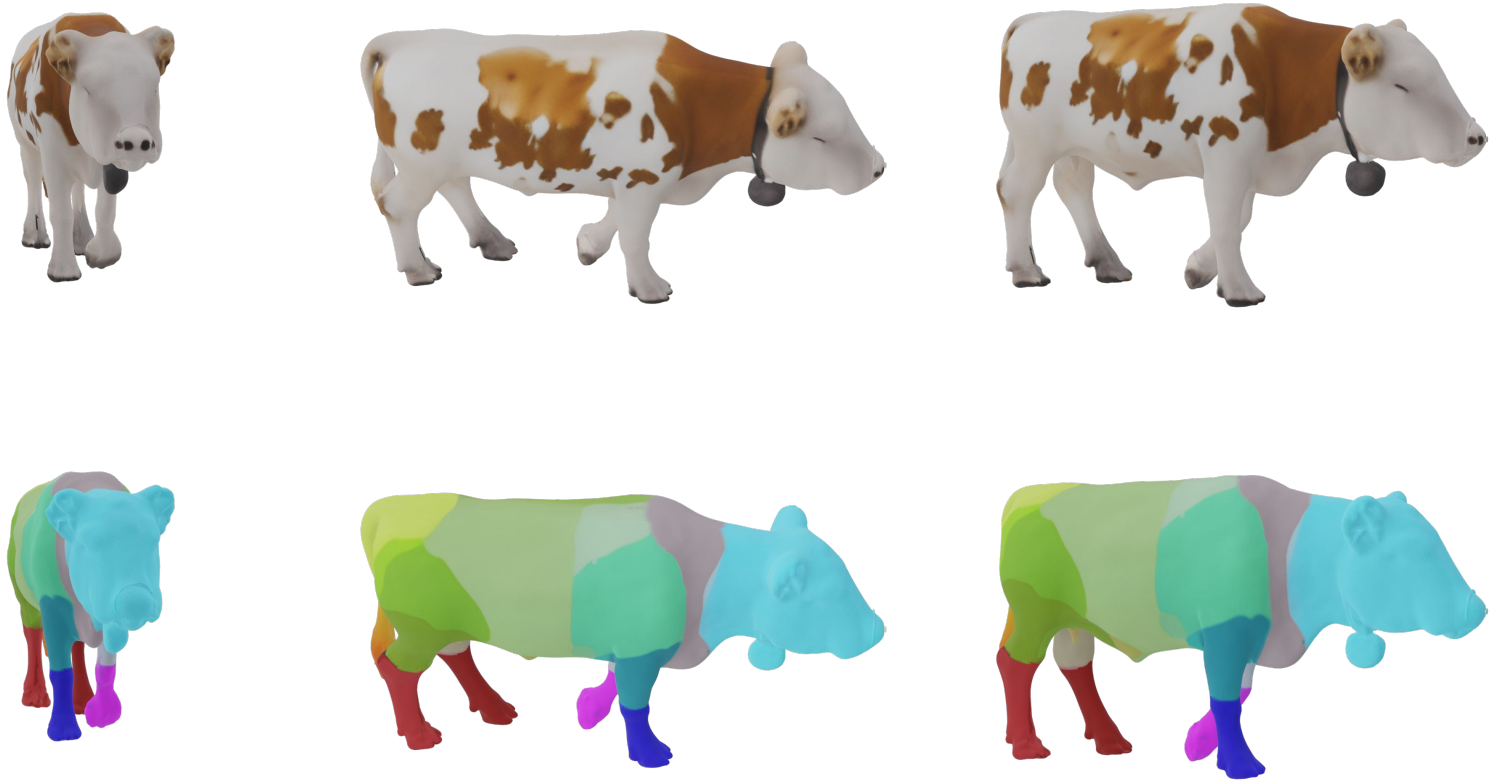
Input Image

Generated 3D object with part decomposition

Results:

- 3D Part Decomposition for real-world input image

Input Image

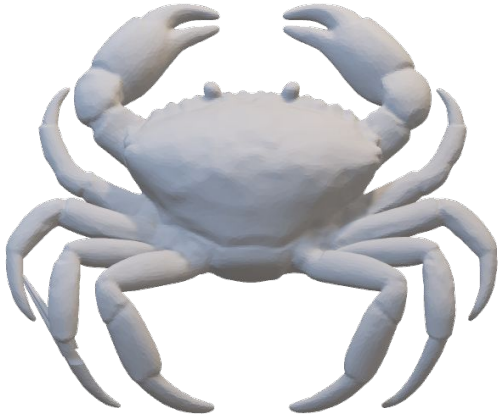


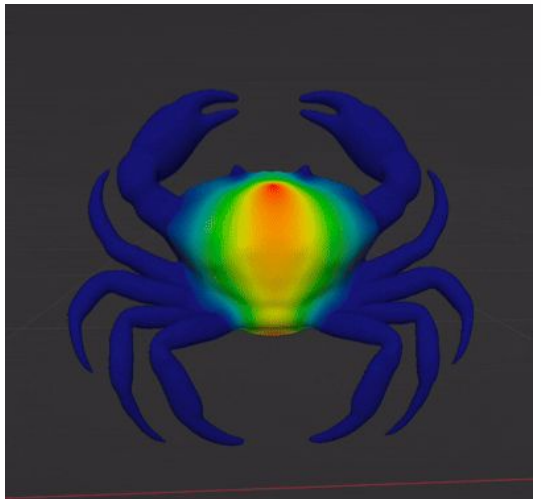
Generated 3D object with part decomposition

Results:

- 3D Part Decomposition and rigging for zero-shot input image

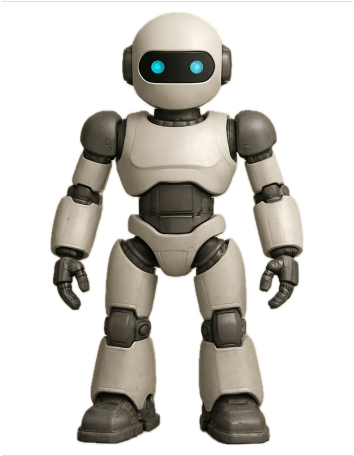
Input Image



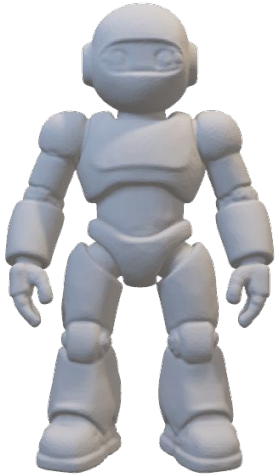


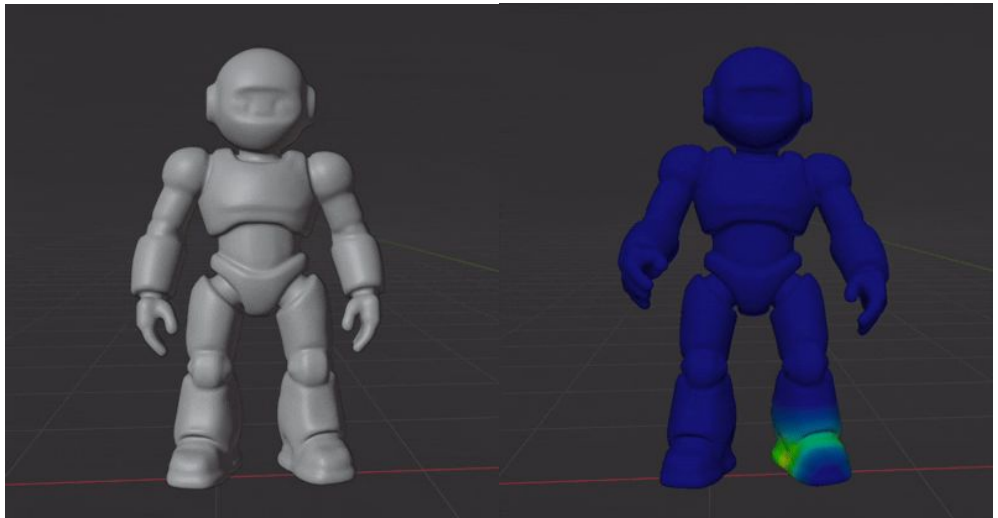
Results:

- 3D Part Decomposition and rigging for zero-shot input image



Input Image





Results:

- 3D Part Decomposition and rigging for zero-shot input image

Input Image