* The Task

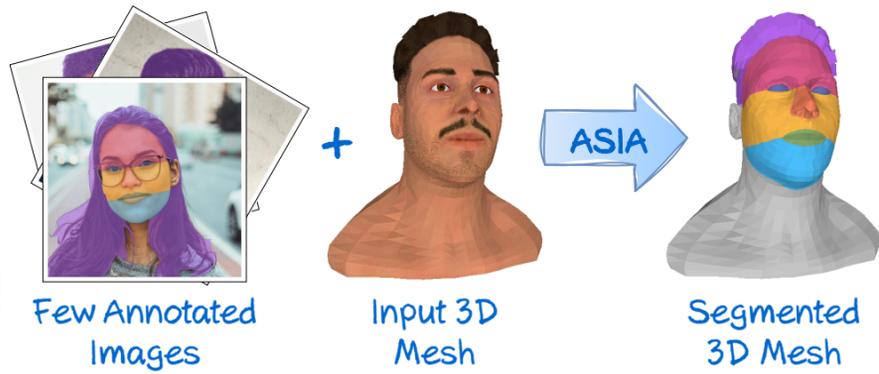Few Annotated Images + Input 3D Mesh → ASIA → Segmented 3D Mesh

* Motivation

Prior Works

Fully supervised methods require a lot of annotated 3D data!

VLM-based methods can only segment semantic and text-describable parts!

Ours

💡 **Adaptive** — Segment non-semantic and non-text describable parts!

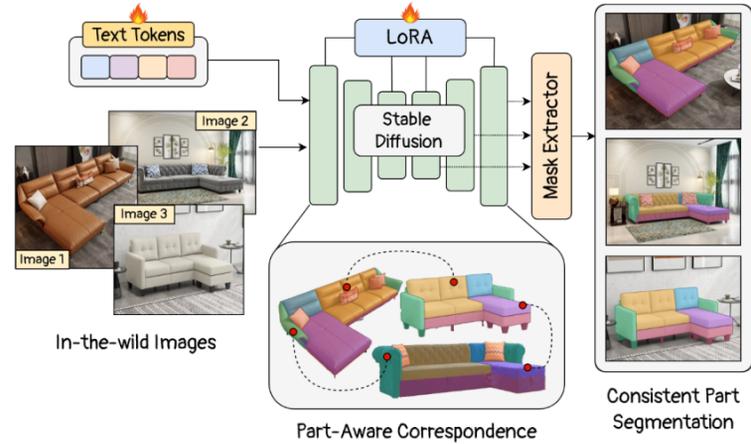💡 **Few-Shot** — Only need a few annotated samples for reference!

💡 **Image Annotations** — Control 3D segmentations with 2D references!
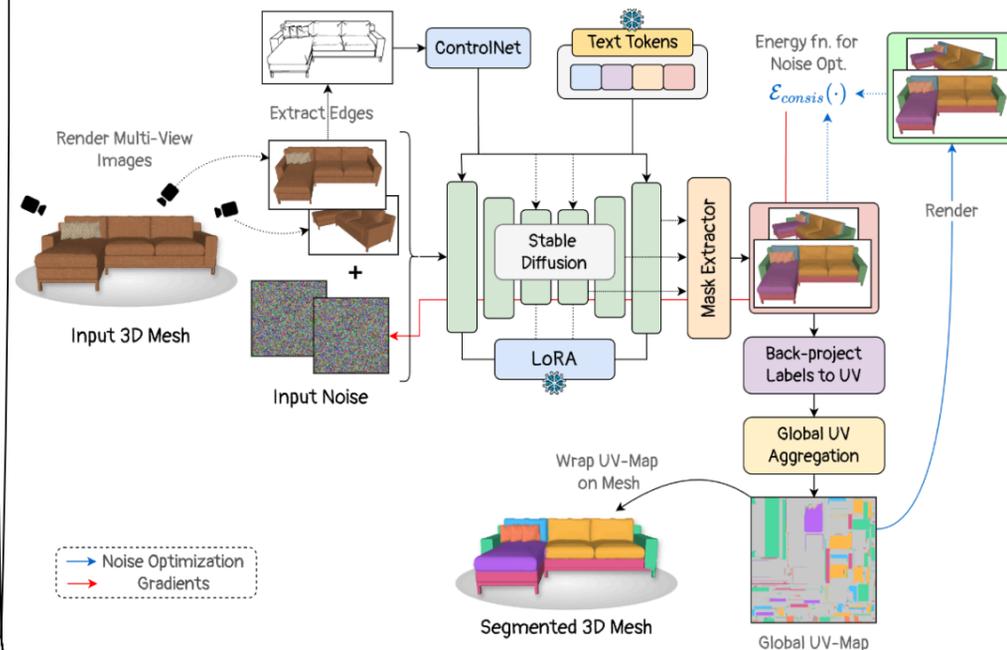
* Our Approach

Training

- -- Optimize text tokens and fine-tune SD with LoRA.
- -- Two-stage training.
- -- *Part-Aware Correspondence* loss.

In-the-wild Images · Part-Aware Correspondence · Consistent Part Segmentation

Inference

- -- Represent segments as mesh textures.
- -- Optimize SD noise for multi-view consistent predictions.

Render Multi-View Images · Extract Edges · ControlNet · Text Tokens · Energy fn. for Noise Opt. $\mathcal{E}_{consis}(\cdot)$ · Stable Diffusion · Mask Extractor · Back-project Labels to UV · Global UV Aggregation · Input 3D Mesh · Input Noise · LoRA · Wrap UV-Map on Mesh · Segmented 3D Mesh · Global UV-Map · Render

→ Noise Optimization Gradients

* Results

Annotated Reference · Segmented 3D Meshes

Part Composition

Annotated References · Input 3D Mesh · Segmented 3D Mesh