CV Project Temp title

CV

CV Final Project

Template Subtitle or Description

Template Author 1
Template Author 2
Template Author 3

1 Phase 1: Synthetic Data Generation

1.1 Data Generation Pipeline

The synthetic pipeline generates labeled surgical tool images through scene setup, randomized lighting, tool placement, keypoint extraction, 2D projection, and annotation.

Scene Setup and Object Loading

Scenes begin in a clean Blender environment, where surgical tools are imported as 3D mesh objects. Lighting is randomized using point sources or HDRI maps, and material properties are varied to simulate realistic reflections and appearances.

3D Keypoint Generation (Ground Truth)

Keypoints are computed per tool using tailored logic based on 3D mesh geometry. The method varies by tool category.

Keypoint Computation for Needle Holder

Tool	**Keypoint(s)**	**Computation Method (Steps)**
Needle Holder	top_left, top_right	1. Select top 15% of vertices by Z-axis. 2. Pick leftmost (min X) and rightmost (max X) among them.
Needle Holder	bottom_left, bottom_right	1. Select bottom 15% of vertices by Z-axis. 2. Pick leftmost and rightmost by X-axis.
Needle Holder	middle_left, middle_right	1. Filter mid-range Z values (30–70%). 2. Choose vertex closest to X-percentile centroid on each side.
Needle Holder	joint_center	1. Compute midpoints of tips and shaft. 2. Interpolate 30% between them and select nearest vertex.

Keypoint Computation for Tweezers

Tool	**Keypoint(s)**	**Computation Method (Steps)**
Tweezers	bottom_tip	1. Find the vertex with lowest Z value. 2. Assign as bottom_tip.
Tweezers	top_left, top_right	1. Split vertices into left/right arms using median X. 2. In each arm, select top 20% Z vertices, pick extreme X.
Tweezers	mid_left, mid_right	1. Find median Z in each arm. 2. Select vertex closest to it.

Keypoint Computation for Fallback Tools

Tool	Keypoint(s)	Computation Method (Steps)
Fallback (Generic)	top_left, top_right	1. Filter top 20% Z vertices. 2. Pick min X (left) and max X (right).
Fallback (Generic)	bottom_left, bottom_right	1. Filter bottom 20% Z vertices. 2. Pick min X (left) and max X (right).

2D Projection

3D keypoints are projected to 2D using BlenderProc's intrinsic matrix and camera poses sampled on a spherical shell around the tools.

View-Dependent Labeling

To maintain consistent left/right annotation across views, projected X-values are compared, and labels swapped if needed.

Rendering and Annotation

Each frame outputs RGB images, segmentation masks, and annotations in COCO format (bounding boxes, keypoints, metadata).

1.2 Implemented Variations

Spatial and Camera Variations

Tools are randomly placed and rotated using spherical sampling. The camera is also moved within a range of angles and distances to mimic real surgical viewpoints.

Lighting and Background Diversity

Point lights vary in position, intensity, and color. Backgrounds switch between HDRI maps and photographic images to simulate various settings.

Creative Variation 1: Multi-Tool Scenes

Frames may contain multiple tools, creating inter-object occlusion and realistic complexity. This mimics crowded surgical environments.

Creative Variation 2: Focused Surgical Lighting

Downward-facing spotlights are used in some scenes to replicate operating room lighting conditions with high contrast and tool emphasis.

1.3 Domain Gap Analysis

Tool Usage Context

Synthetic: Tools appear isolated, without interaction.
Real: Tools are held by hands and interact with surgical fields.
Domain Gap: Lack of procedural context limits realism.

Background Semantics

Synthetic: Backgrounds are unrelated to medicine.
Real: Includes drapes, gloves, and patient anatomy.
Domain Gap: Missing clinical context affects semantic alignment.

Texture and Surface Realism

Synthetic: Surfaces are clean and perfect.
Real: Includes reflections, stains, and occlusions.
Domain Gap: Idealized textures reduce transferability.

Scene Complexity and Occlusion

Synthetic: Minimal and controlled occlusion.
Real: Frequent overlaps with tools and hands.
Domain Gap: Simplified scenes reduce training robustness.

1.4 Discussion of Challenges and Key Findings

Challenges

- **Annotation Strategy:**
Manual labeling was unscalable; automation required mesh understanding and consistent landmark definitions.
- **Ground Truth Modeling:**
Bounding boxes lacked precision; keypoints added detail but required careful design.
- **Domain Alignment:**
Synthetic scenes missed surgical cues like hands, lighting, and tissue.

Key Findings

- **Reliable Auto-Annotation:**
Ground truth labels were generated using mesh metadata—accurate and scalable.
- **Multi-Tool and Multi-Pose Scenes:**
Improved realism and contextual diversity.
- **Spotlight Lighting:**
Enhanced tool visibility and mimicked surgical illumination.
- **Pose and Camera Variation:**
Supported diverse perspectives and improved generalization.

2 Section 2

3 Section 3