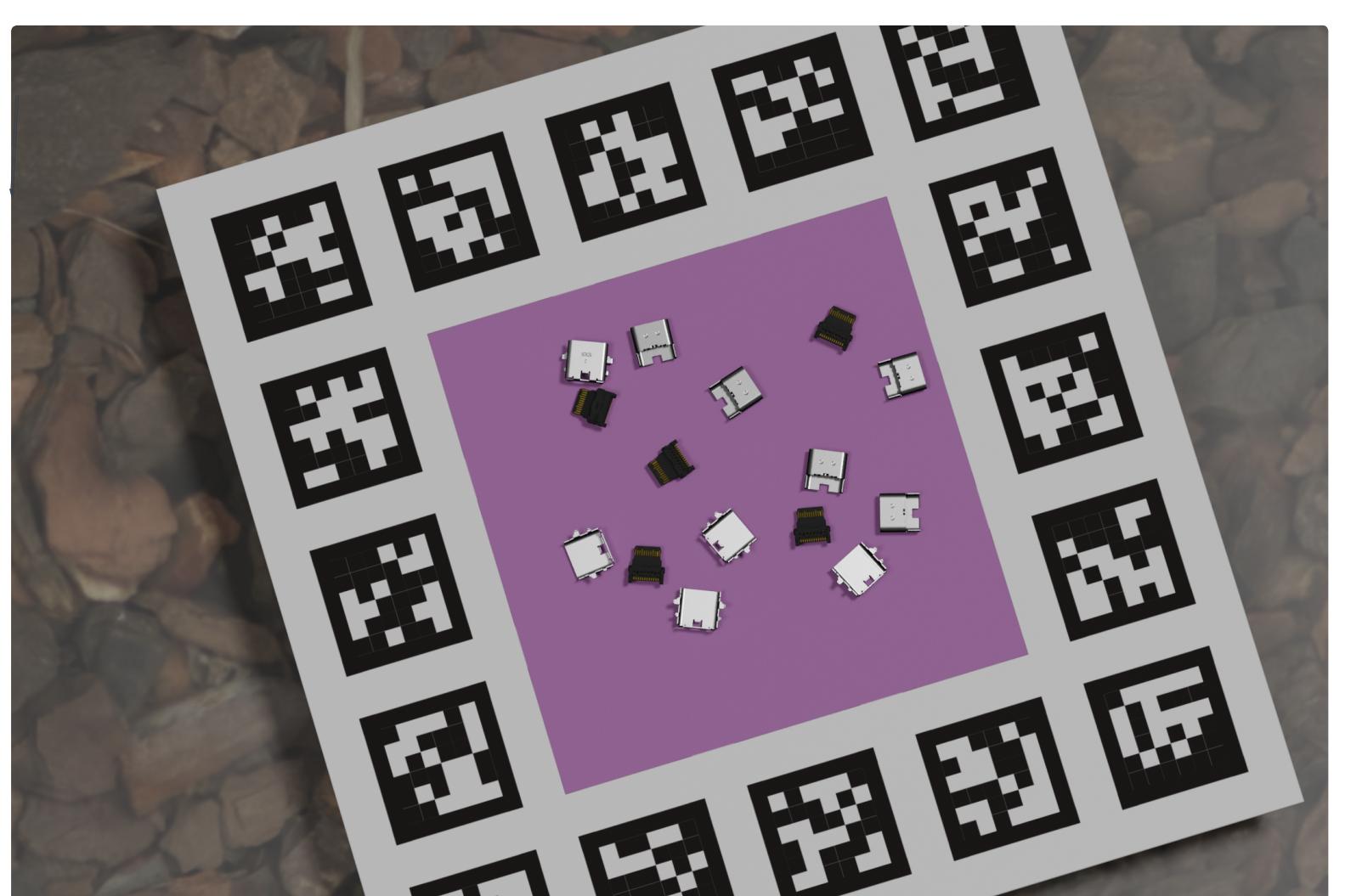
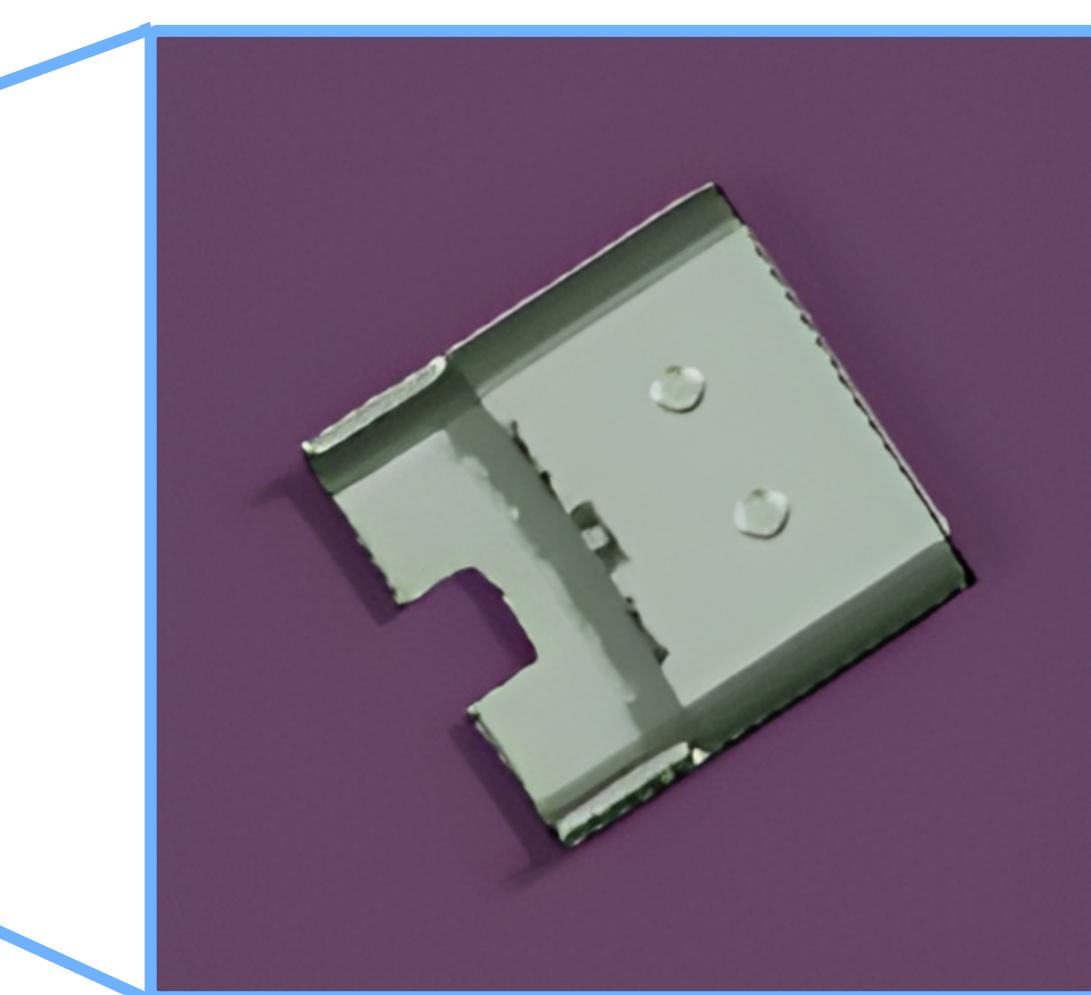


# I) Instance Segmentation

Input Image  
(BlenderProcv2)

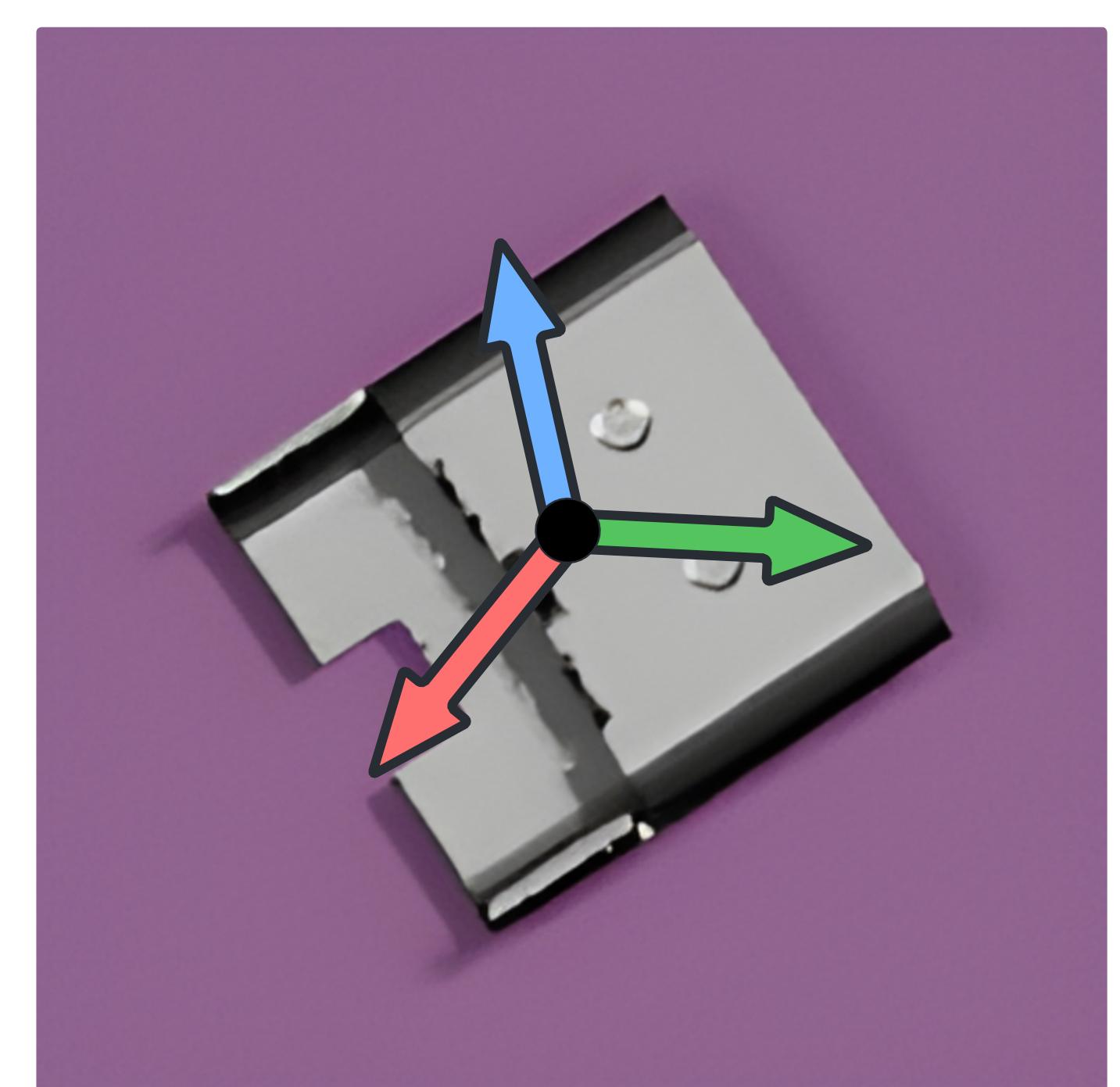
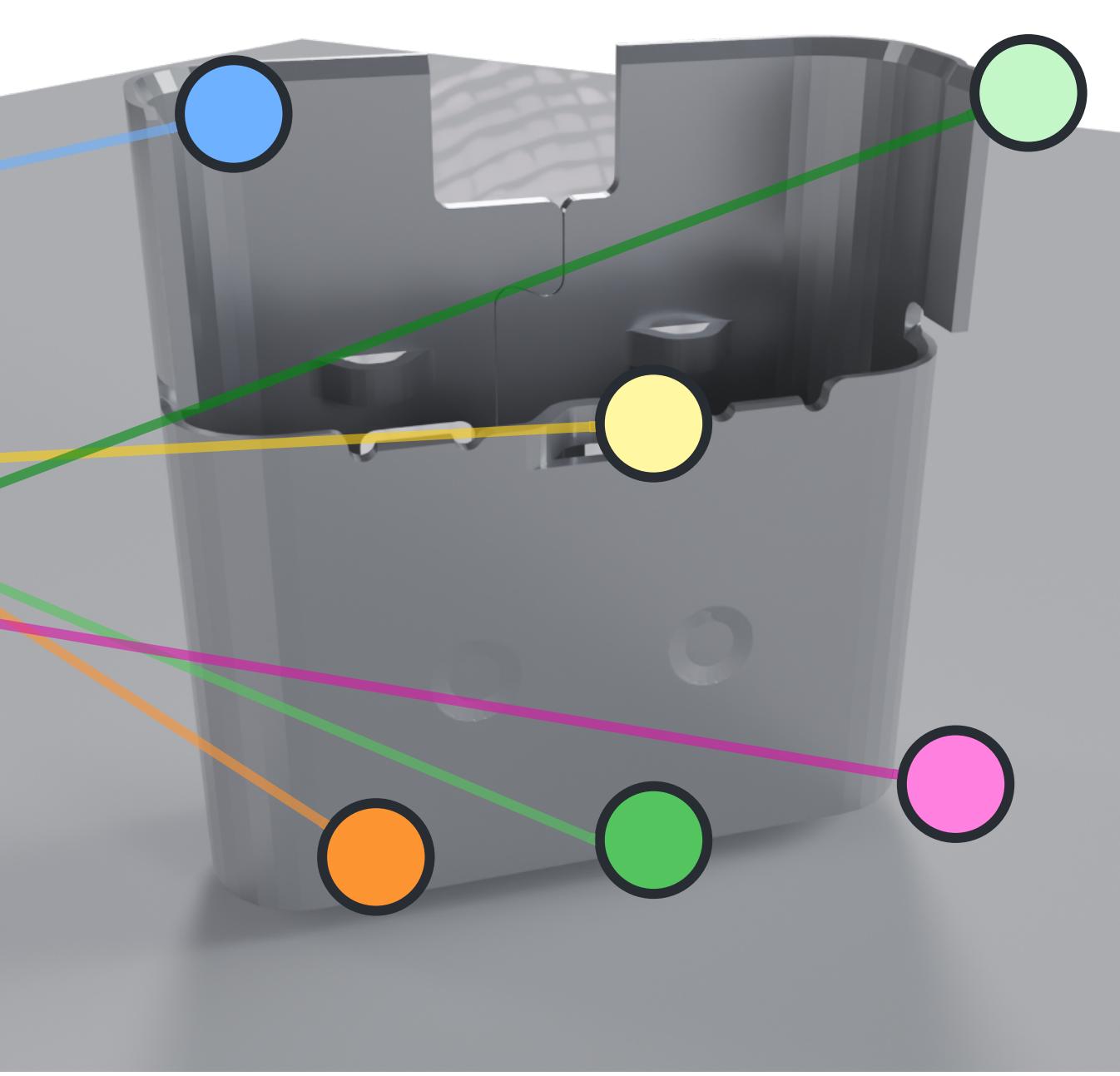
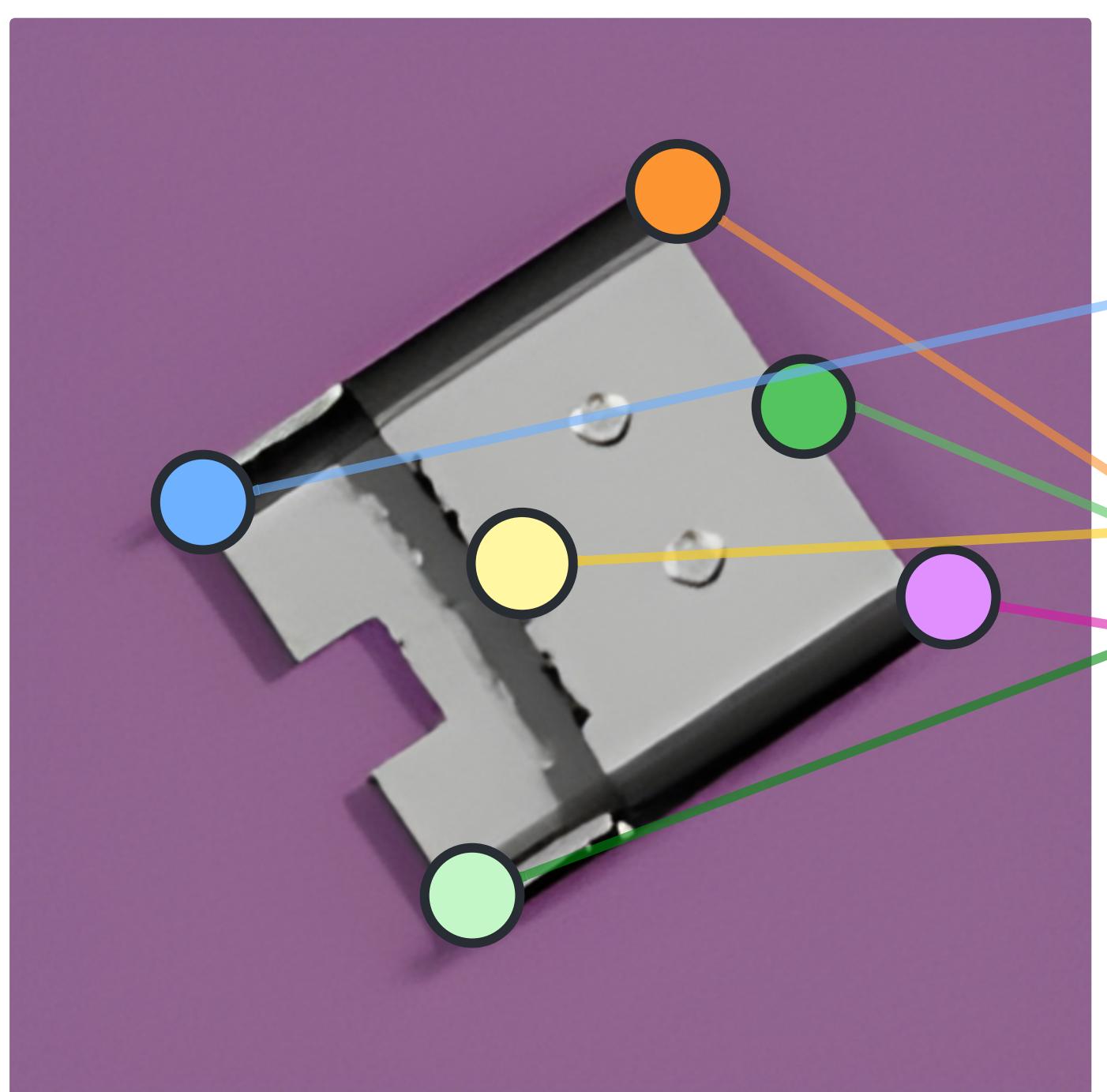


Instance Segmentation  
(Mask R-CNN)



# II) Pose Estimation

2D-to-3D keypoint correspondence



Predicted 2D KeyPoints  
(PVNet)

3D Keypoints  
(PVNet)

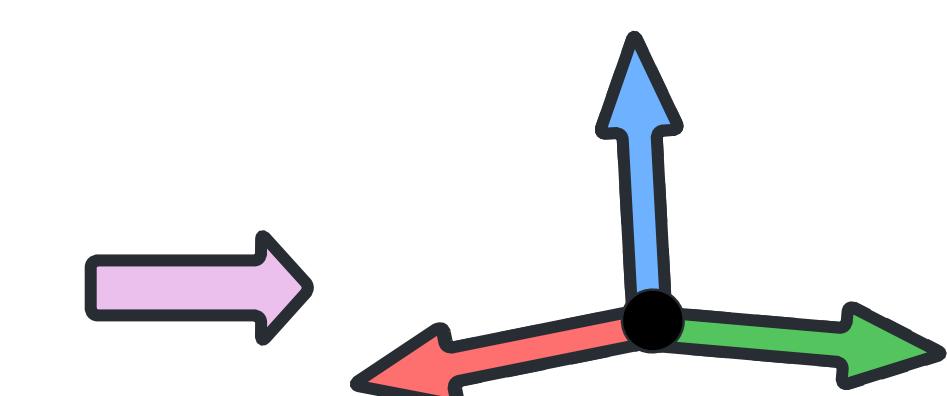
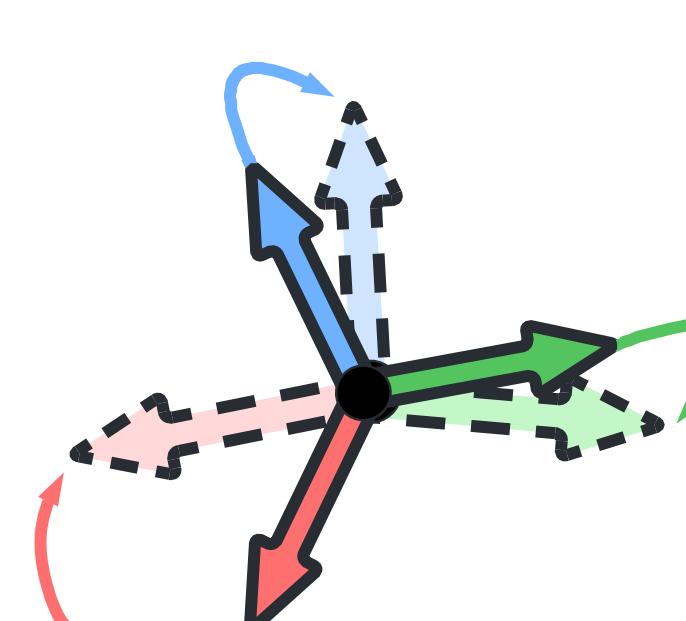
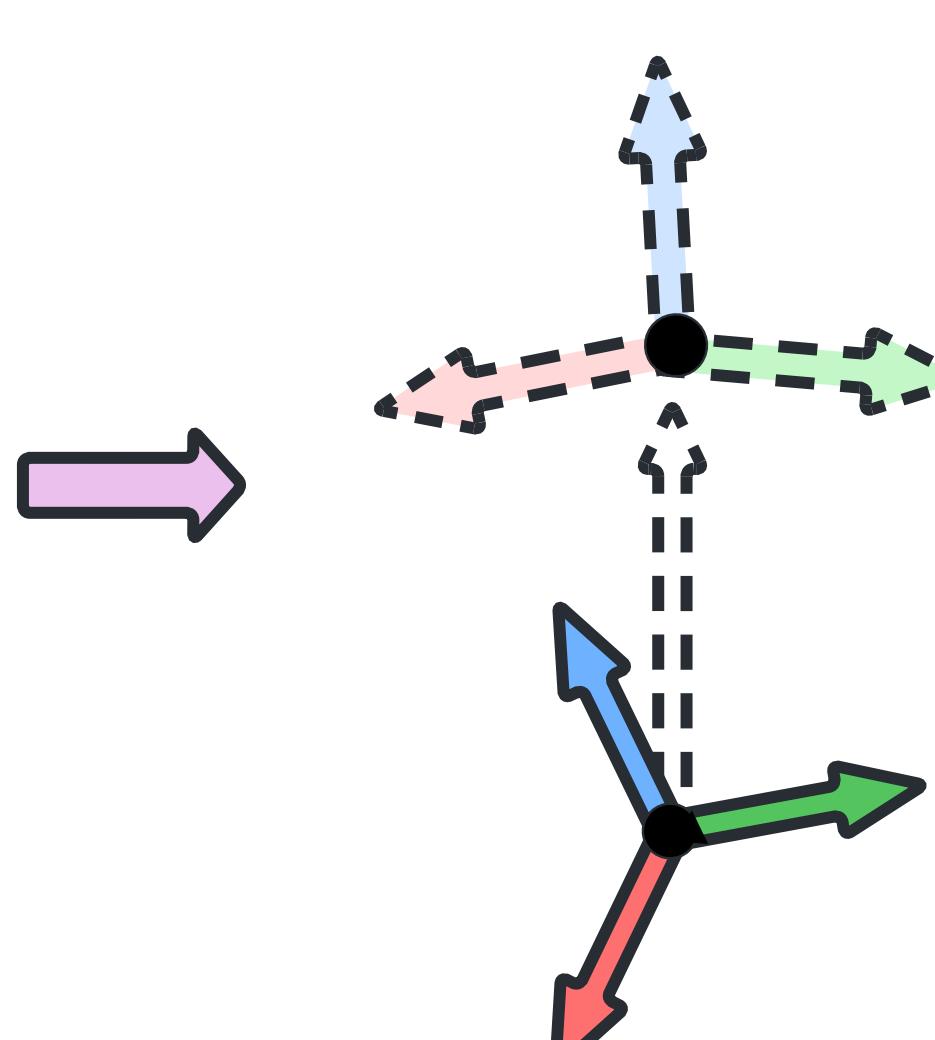
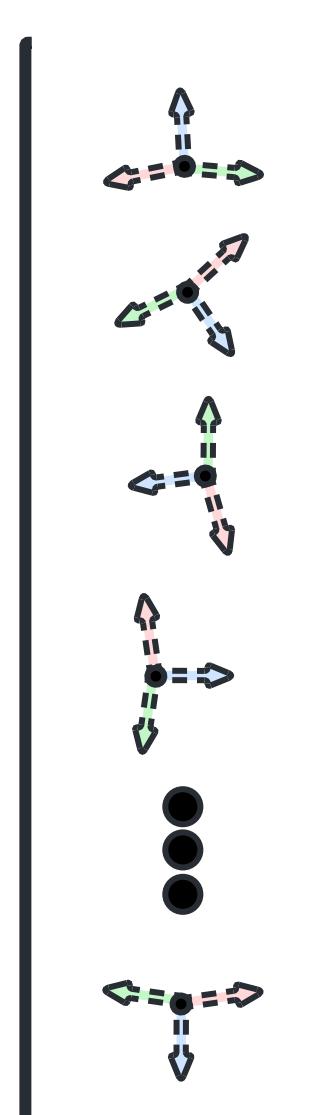
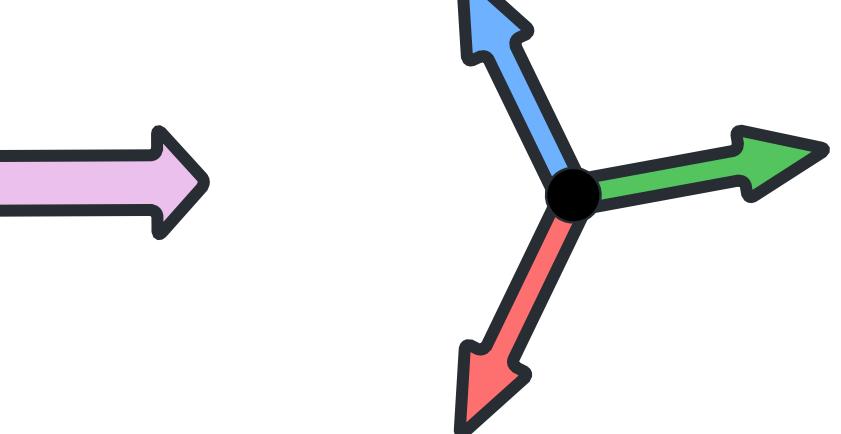
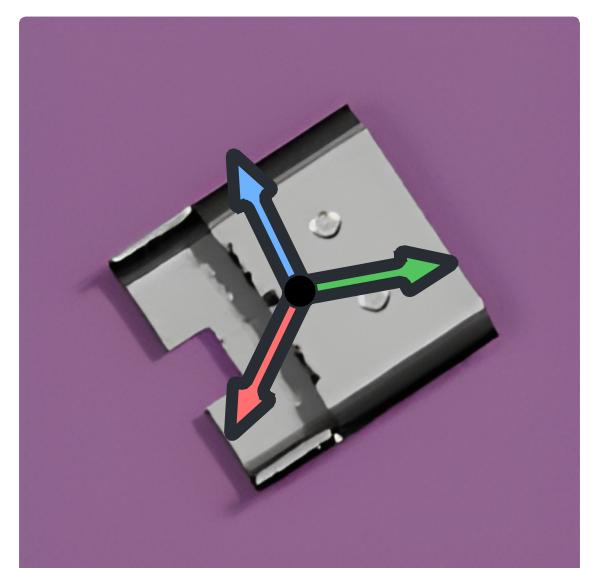
$$\hat{\mathbf{T}}_i = \begin{bmatrix} \hat{\mathbf{R}}_i & \hat{\mathbf{t}}_i \\ 0 & 1 \end{bmatrix}$$

# III) Pose Refinement

Find Nearest  
Stable Poses

Adjust Z Position and  
Orientation

Refined Stable  
Pose



$$\hat{\mathbf{T}}_i = \begin{bmatrix} \hat{\mathbf{R}}_i & \hat{\mathbf{t}}_i \\ 0 & 1 \end{bmatrix}$$

$$\mathbf{T}_{stable_i} = \begin{bmatrix} \mathbf{R}_{stable_i} & \mathbf{t}_{stable_i} \\ 0 & 1 \end{bmatrix}$$