Epiretinal Membrane and Instrument Segmentation in Retinal Surgery Video

Saurav Verma
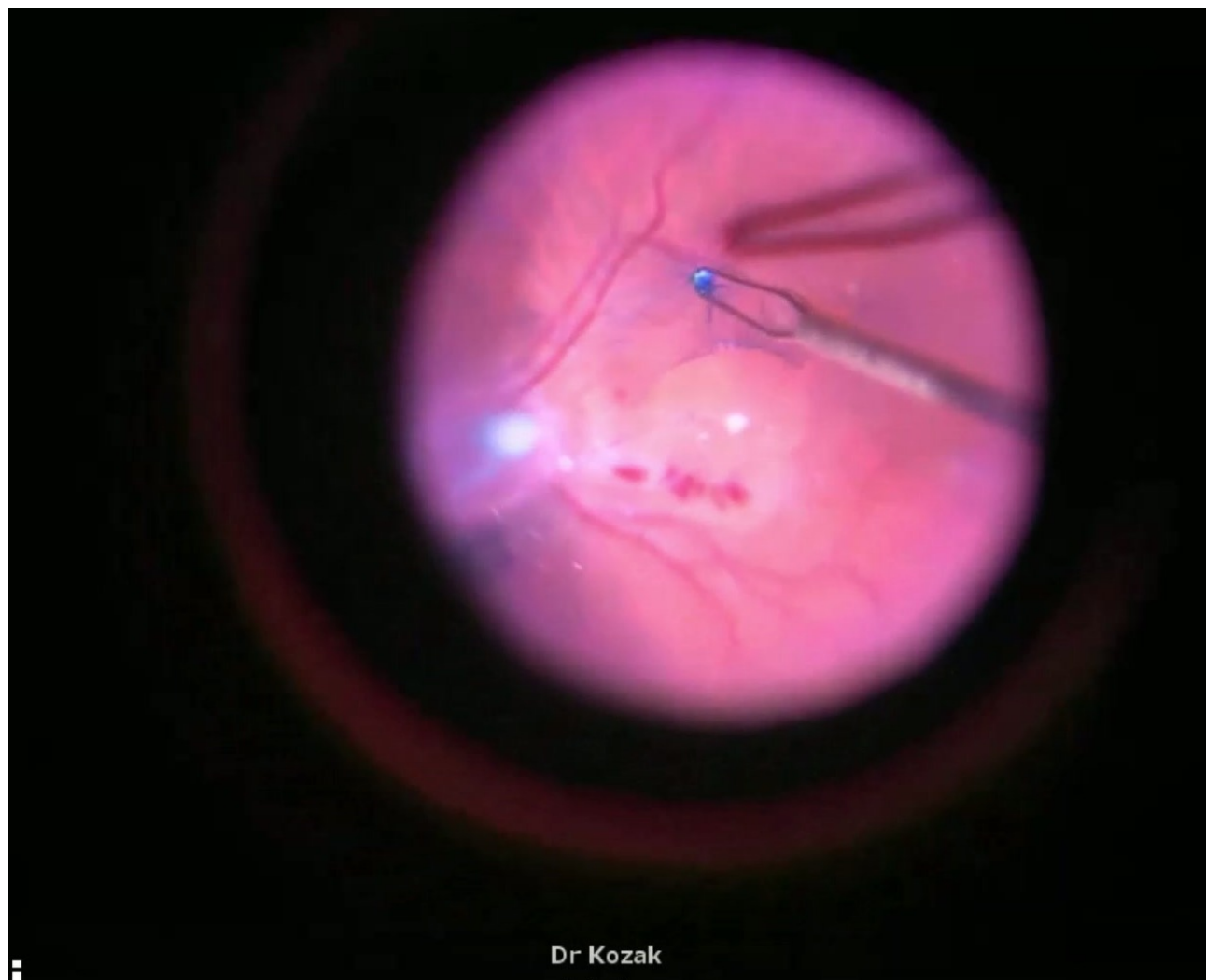
College of Information Science, University of Arizona
Advisor: Dr. Eungjoo Lee

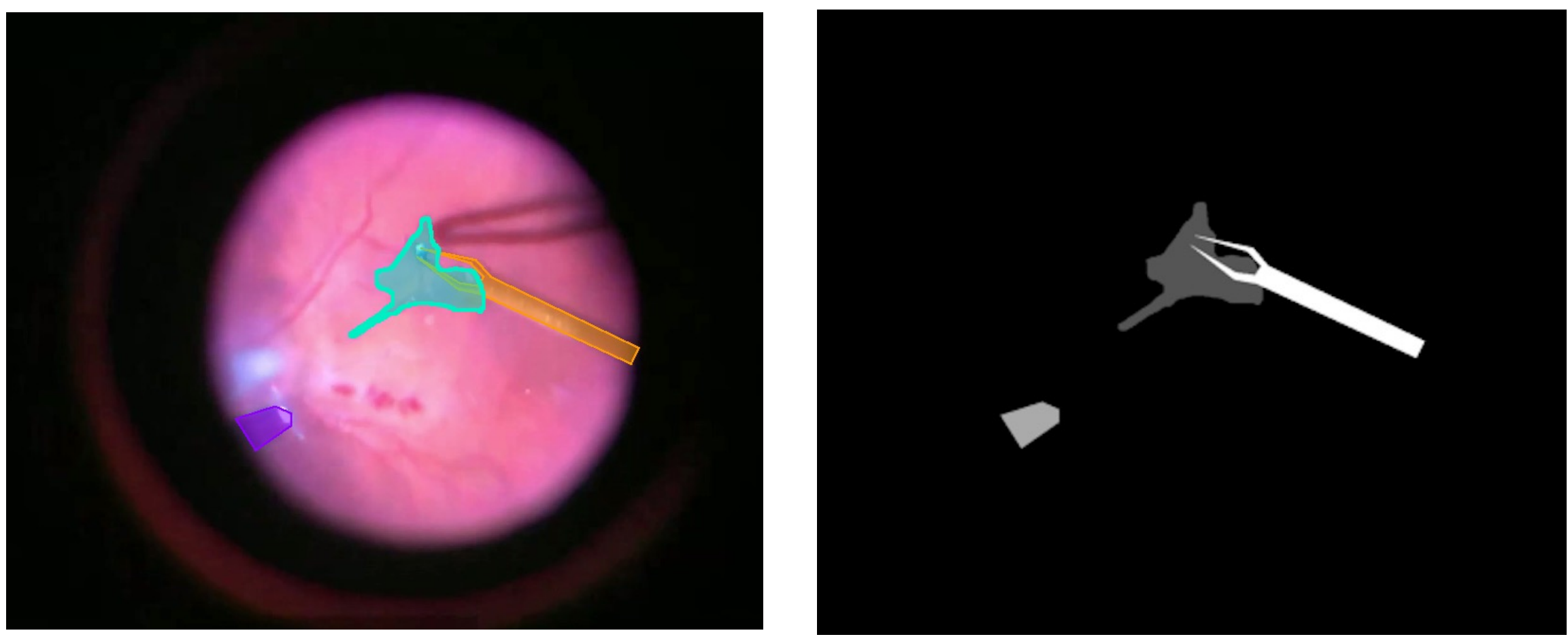Vision, Systems & Intelligence (VSI) Lab — University of Arizona

PURPOSE

❑ ERM peeling depends on accurately identifying thin membranes and instrument-tissue interaction.

❑ Ophthalmic surgical video remains underrepresented in pixel-level segmentation research.

❑ This project develops a pilot dataset and evaluates modern deep-learning models for ERM and tool segmentation.

CLINICAL CONTEXT

Intraoperative view showing forceps initiating an ERM peel.

ANNOTATION EXAMPLES
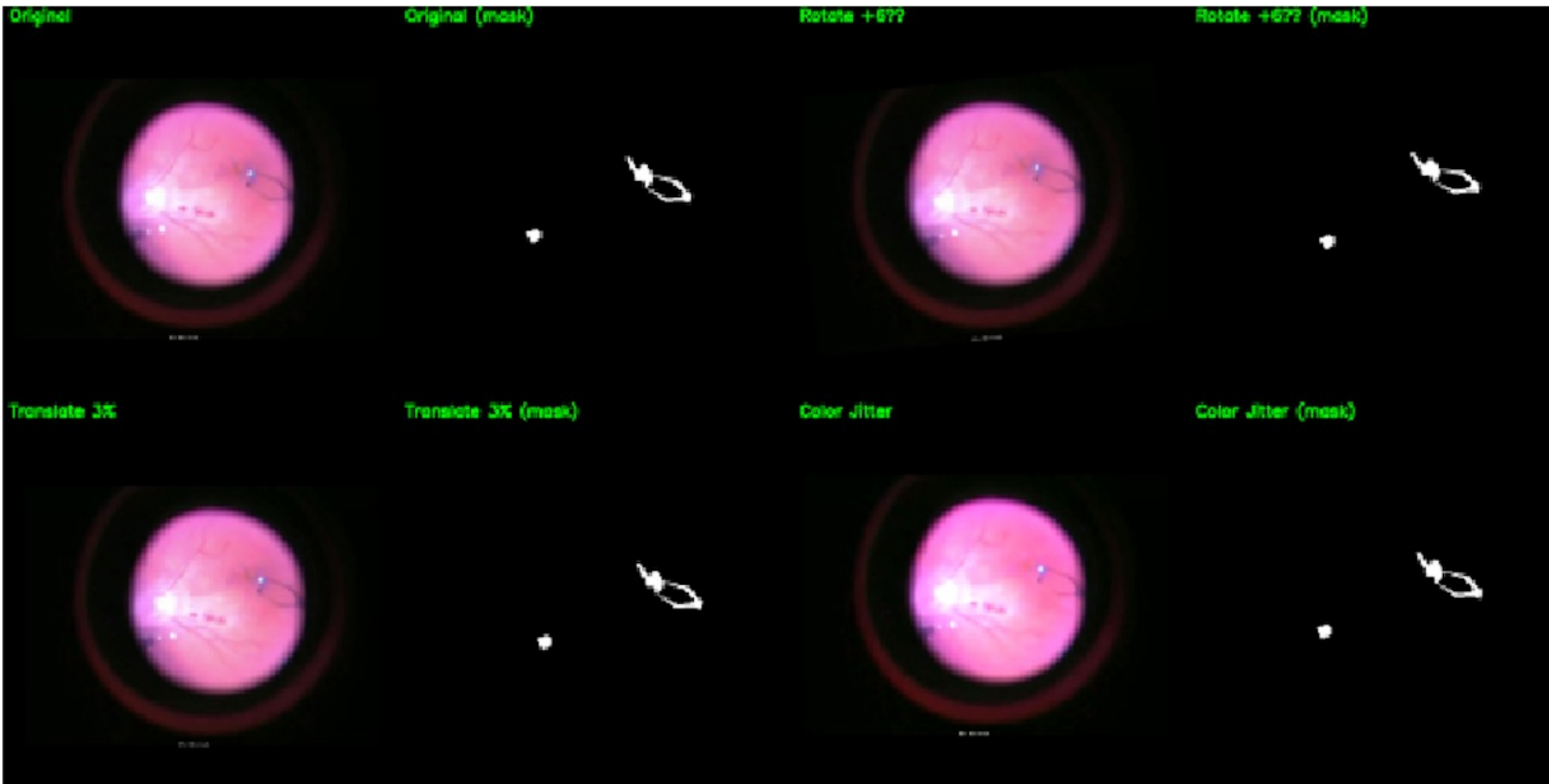
Annotated Overlay + *Greyscale Mask*

ANNOTATION SUMMARY

❑ 300 frames (1 fps) from 7 ERM surgery videos.

❑ Annotations done in *Supervisely framework.*

❑ Classes: ERM, Retinal Forceps, Light Tool.

❑ Surgery phase labels (Flap initiation, Dye, Peeling, Completion) collected for future modeling.

❑ Includes challenging cases: faint ERM, tool overlap, illumination artifacts.
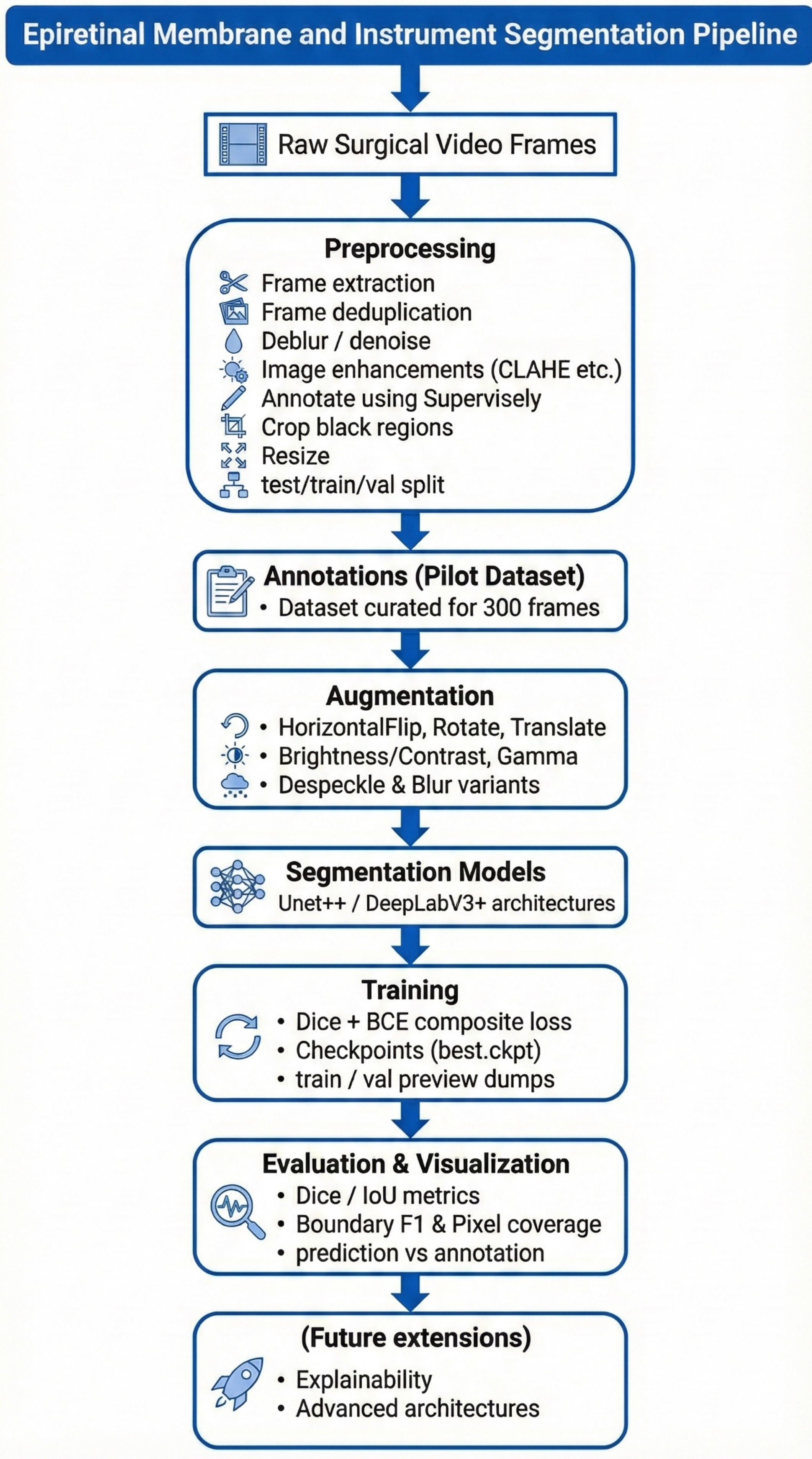
DATASET & PREPROCESSING

❑ Frame extraction, deduplication, blur filtering.

❑ Region-of-Interest cropping to isolate surgical field.

❑ Resize: 512×512, 768×768.

❑ Surgery-level train/val/test split.

AUGMENTATION PIPELINE

Training augmentations: rotation, flips, color jitter, scaling
(Top-left: original frame/mask; other panels show a few augmented variants)
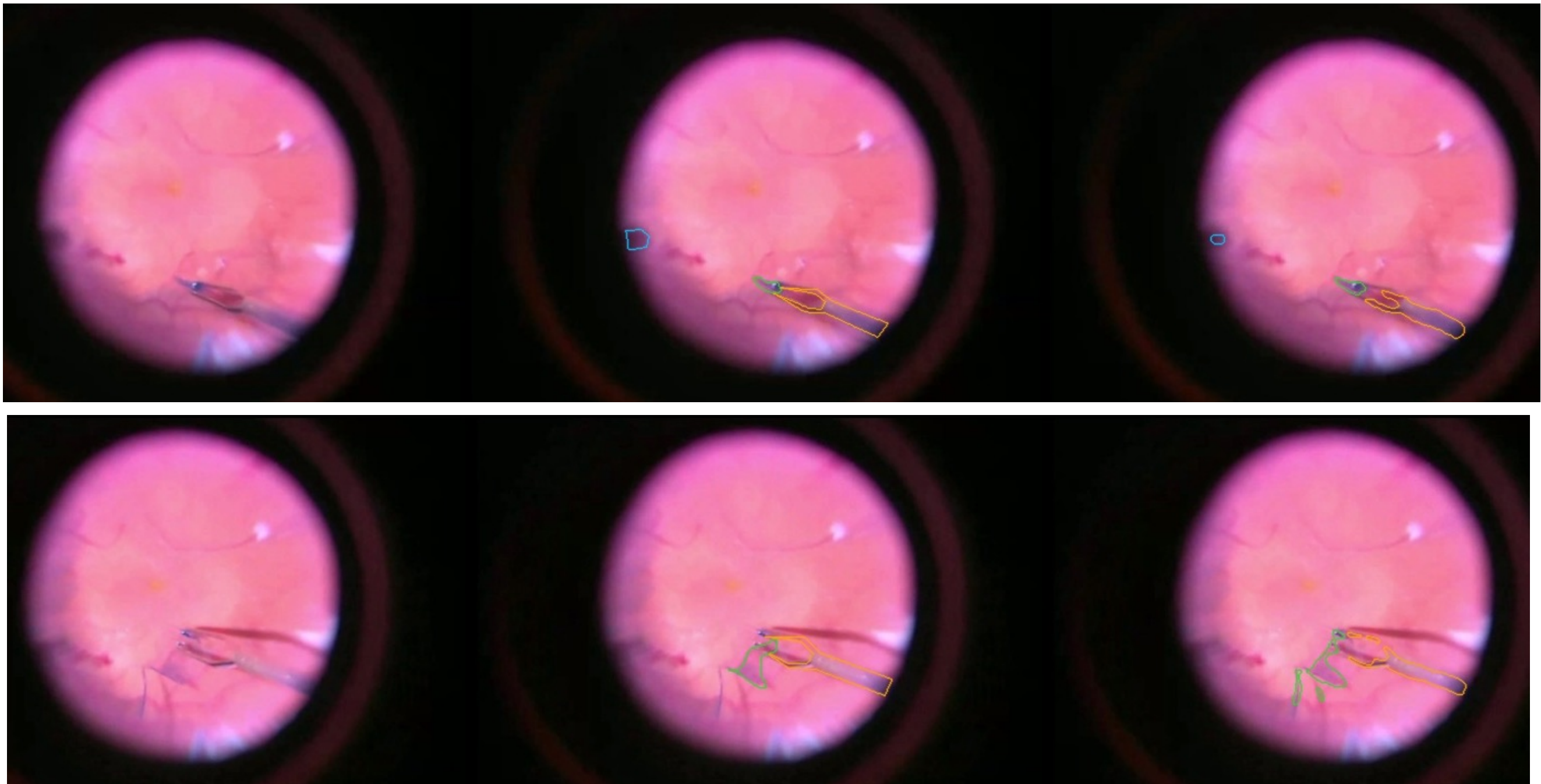
END-TO-END WORKFLOW

Epiretinal Membrane and Instrument Segmentation Pipeline

Raw Surgical Video Frames

Preprocessing
- Frame extraction
- Frame deduplication
- Deblur / denoise
- Image enhancements (CLAHE etc.)
- Annotate using Supervisely
- Crop black regions
- Resize
- test/train/val split

Annotations (Pilot Dataset)
- Dataset curated for 300 frames

Augmentation
- HorizontalFlip, Rotate, Translate
- Brightness/Contrast, Gamma
- Despeckle & Blur variants

Segmentation Models
Unet++ / DeepLabV3+ architectures

Training
- Dice + BCE composite loss
- Checkpoints (best.ckpt)
- train / val preview dumps

Evaluation & Visualization
- Dice / IoU metrics
- Boundary F1 & Pixel coverage
- prediction vs annotation

(Future extensions)
- Explainability
- Advanced architectures

QUANTITATIVE RESULTS

Model	ERM	Forceps	Light Tool (Endoillumination)
DeepLabV3+	**0.77**	**0.83**	**0.76**
Unet++	0.70	0.80	0.72

Dice Scores for ERM & Instruments
(This is an ongoing research project; detailed internal evaluations are not disclosed)

QUALITATIVE RESULTS

Top Row: Successful segmentation of ERM + instruments.
Bottom Row: Challenging case : artifacts glare, faint ERM.
Column : Left: Original, Middle: GT, Right: Prediction

KEY INSIGHTS

❑ ERM & tool segmentation is feasible with limited data.

❑ 512 px resolution provides strong segmentation and contour accuracy despite extreme foreground sparsity (<1% pixels).

❑ DeepLabv3+ provides the most stable performance in this pilot dataset.

CONCLUSION

DeepLabv3+ at 512 px showed the most consistent performance, demonstrating that ERM and instrument segmentation from intraoperative video is technically feasible even with limited curated data.

FUTURE EXTENSIONS

❑ Expand dataset with more surgeries.

❑ Explore transformer-based architectures.

❑ Add surgical-phase recognition in a multi-task pipeline.