[Samsung PRISM] End Review Report

A software tool to generate Multi-view Image Segmentation & Correction

Working Team Details [Name & Email ID] :

1. College Professor(s): Prakash Aithal Sir
2. Students:
    1. A S Aravinthakshan - aravinthakshanmain@gamil.com
    2. Kavya Bansal - kavya.mitmpl2022@learner.manipal.edu
    3. Prassana - prasannabhat345@gmail.com
    4. Janak Shah - janak.sh41@gmail.com
3. Department: Computer Science and Engineering

Work-let Area – Computer Graphics | A software tool to generate Multi-view Image Segmentation & Correction

SAMSUNG

Work-let expected duration – 6 months.

Problem Statement:

- Image Segmentation is a technique of partitioning a digital image (or video frames) into multiple groups, based on similar characteristics. This requires high quality point-wise boundary annotation for object segments, for training DNN-based segmentation models. This annotation process is manual and time taking.
- If a scene is captured from different viewpoints, each view point image needs to be segmented and labelled individually.
- Goal is to develop a web-based annotation tool, hosted on Linux server, to generate the segmentation in multi-view images automatically, including interactive correction.

Web-based Annotation Tool with Correction Methodology:

- ✓ **Module 1:** Given an image, generate manual annotation by creating point-based polygons, and assign suitable labels.
- ✓ **Module 2:** Given a manually segmented image and other images of same scene from different views as inputs, automatically generate the segmented images for all the views.
- ✓ **Module 3:** Given an already segmented image, support Interactive Correction, which is a click-based interactive refinement, for correcting the segmentation mask in case of both false positives and negatives.
- ✓ Basic functionalities to be supported: browse folder, previous image, next image, create polygon, edit polygon, create segmentation, add segment and remove segment.
- ✓ Open source DL based models can be plugged-in at the backend for multi-view image segmentation and interactive correction.
- ✓ Scope includes creation of a small multi-view image dataset for testing the functionality of the tool.

References:

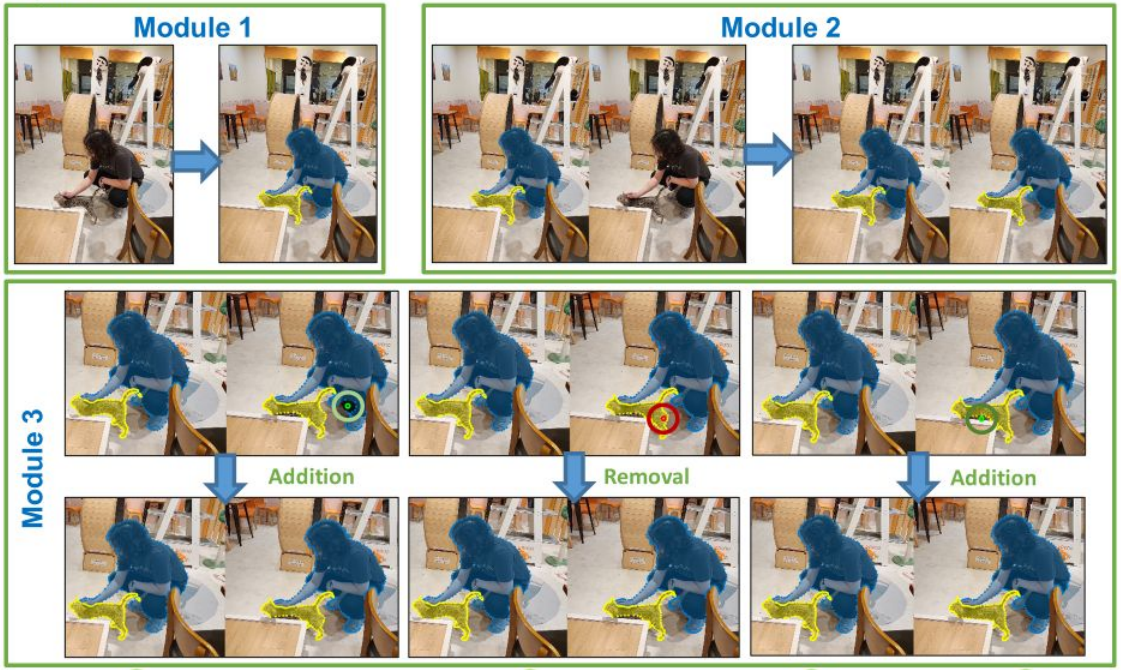
- https://github.com/labelmeai/labelme
- https://github.com/gap-lab-cuhk-sz/mvimgnet
- CVAT

Shreyasi Das
shreyasi.das@samsung.com

Jyoti Johri
jyoti.johri@samsung.com

Shouvik Das
shouvik.das@samsung.com

Expectations



Kick Off < 1st Month >

- Problem definition, scoping
- Understanding the requirements
- Basic design of the web based tool for segmentation annotation (Linux based, hosted on server)
- Use of open-source DNN models for multi-view image segmentation and interactive correction.

Milestone 1 < 2nd & 3rd Month >

- Development of Module 1 feature of the automation tool , along with functionalities for browse folder, previous image and next image.
- Creation of multi-view image dataset for testing.
- Finalizing the open-source DNN models

Milestone 2 < 4th Month & 5th Month >

- Development of Module 2 and Module 3 features of the automation tool
- Individual testing of each module

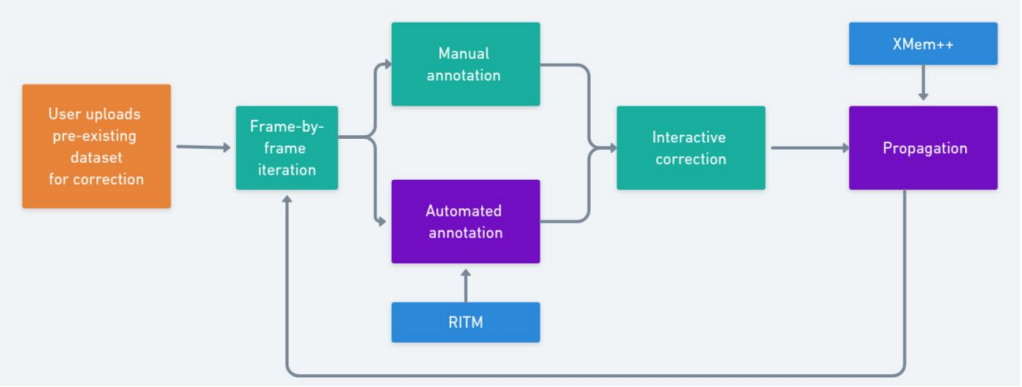
Closure < 6th Month >

- Final verification of entire system/tool together and testing.

Approach / Solution

- **Concept Diagram** :
  (Clear detailed schematic / block diagram / flow chart depicting the proposed concept / solution)

- The segmentation tool enables rapid and accurate annotation of video frames, significantly reducing manual effort.
- RITM (Region-based Interactive Segmentation) allows users to place clicks to iteratively refine segmentation masks.
- After initial RITM-based segmentation, users can manually adjust masks to ensure high accuracy.
- XMem uses the annotated reference frames to propagate segmentation masks to subsequent frames.



Dataset(s) Analysis / Description

- **Dataset Capture / Preparation / Generation** :
  (Discuss the dataset generation process or if downloaded data provide details of what data & from where it was obtained etc… - 2 to 3 bullets only)

No dataset is required. The tool can be applied directly to both individual images and video frames.

- **Dataset Understanding / Analysis** :
  (Provide 2 to 3 bullets about what is your understanding of the data / opinion about the data)

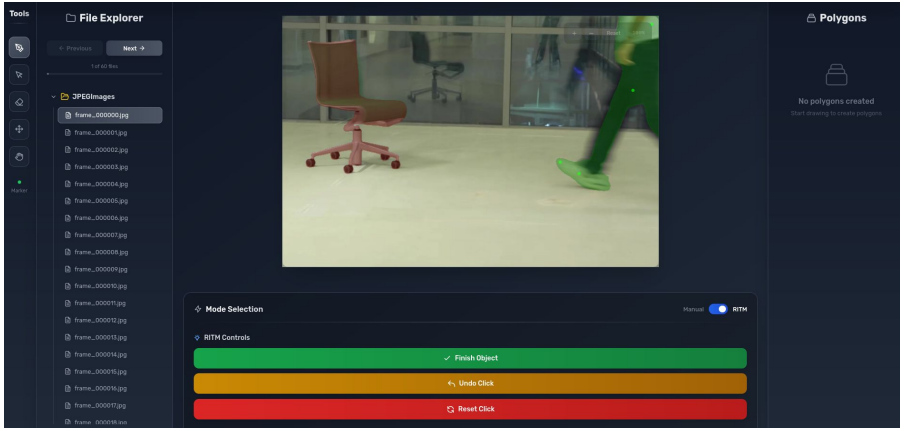
No specific requirements, as long as it's a RGB Image (If transparency channel exists its handled as well), works well for video object segmentation because of the nature of Xmem propagation while RITMs interactive correction can work on both.

- **Dataset Pre-Processing / Related Challenges (if any)** :
  (List out the challenges you foresee in data handling wrt problem definition – 2 to 3 bullets only)

RITM outputs single-channel grayscale masks where each object has a unique value (1, 2, 3...) and background is 0.
XMem expects paletted PNG masks using the DAVIS color format, with each object represented by a specific color.
For this, we implemented a converter that maps RITM's grayscale output into XMem-compatible DAVIS palette masks.

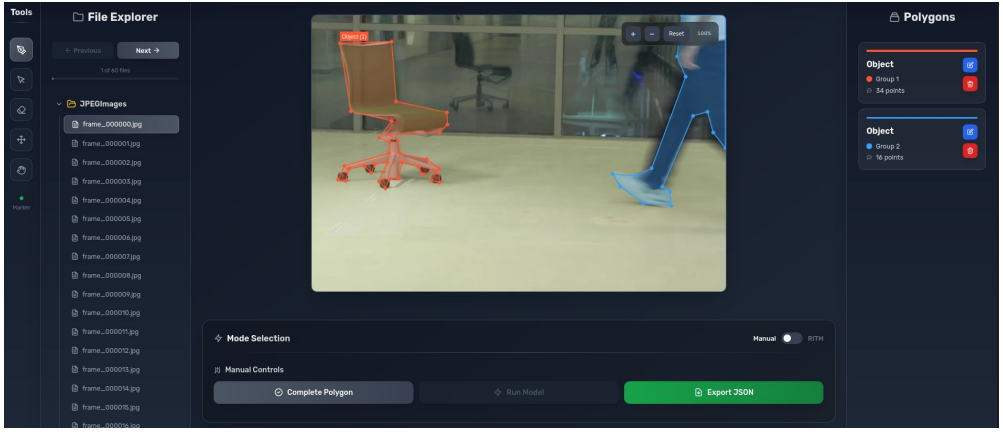
Experimental Results / Simulations / Observations

- **Results** :
  (provide numerical data / bar charts / plots / images / videos / tabulated results etc. Use full slide or multiple slides up to max 3 slides to demonstrate the results)



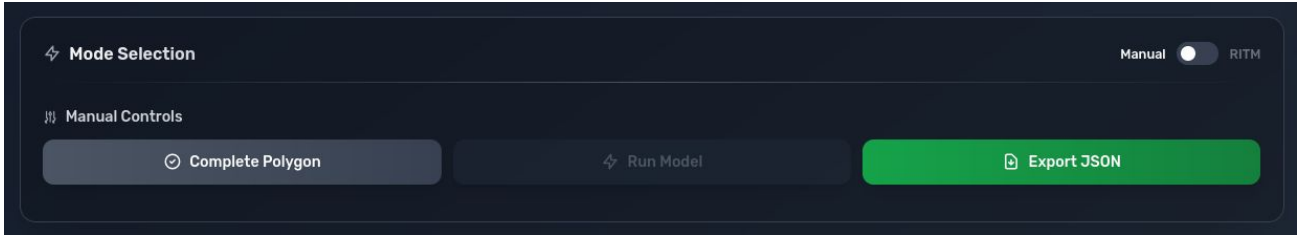
Once the RITM Masks are generated you can switch to manual to make any finer corrections and rename the default class names

Once the freeform and RITM Annotations are fixed for the reference frame you can move onto the next one, and click on “Run Model”, this will show you a “Processing..” Text on the button.
Once the Run Model button is back to being enabled the propagated mask by Xmem will render on the screen.

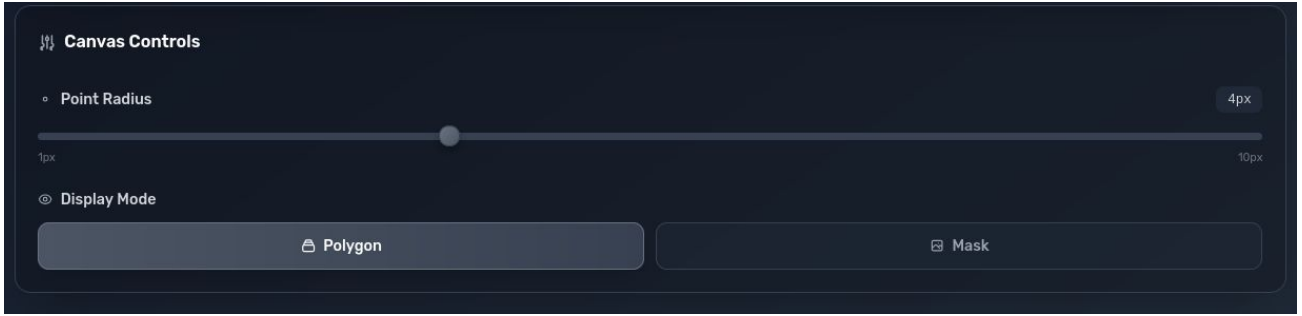


Experimental Results / Simulations / Observations

- **Results** :
  (provide numerical data / bar charts / plots / images / videos / tabulated results etc. Use full slide or multiple slides up to max 3 slides to demonstrate the results)



Toggle lets you choose between Manual Editing and RITMs interactive Correction



This pane lets you change point density and point radius and also switch between polygon and mask view

1) Add Point
2) Select and Move points
3) Erase Point
4) Move Mask
5) Navigate through IMage

Zoom Pane

Experimental Results / Simulations / Observations

Major Observations

- Successful dual-model integration: RITM and XMem models work within a single platform despite conflicting dependencies
- Effective hybrid workflow: Users can seamlessly transition between AI-assisted segmentation (RITM) and manual polygon editing
- Real-time synchronization: Four-panel interface (Toolbar, Folder Tree, Preview Section, Polygon Tray) maintains perfect state consistency across all components
- Bidirectional format conversion: Polygon-to-mask and mask-to-polygon conversion enables flexible iterative refinement workflows
- Session persistence: All annotations and masks are maintained throughout user sessions with export capabilities

Key Conclusions

- Architecture validation: React.js frontend with dual Flask/Express backend successfully handles complex computer vision workflows
- User experience success: Clean interface abstracts AI model complexity while providing sophisticated annotation capabilities
- Video propagation effectiveness: XMem integration significantly reduces manual effort for sequential frame annotation
- Scalable label management: Predefined label lists with custom creation and grouping functionality supports structured dataset preparation
- Component-based design benefits: Isolated, reusable UI elements enable maintainable and modular development

Experimental Results / Simulations / Observations

Technical Development Challenges

- Dependency conflicts: Separate virtual environments (venv for XMem, eenv for RITM) required due to incompatible Python packages
- Dual API architecture: Flask and Express backends needed with dynamic switching between main Express thread and Flask-based RITM API
- Real-time synchronization: Four-panel interface state consistency across Toolbar, Folder Tree, Preview Section, and Polygon Tray
- Canvas implementation: Developing interactive canvas functionality for polygon editing, zooming, and panning with real-time visual feedback
- Format conversion complexity: Bidirectional polygon-to-mask and mask-to-JSON conversion while preserving metadata integrity

User Experience and Workflow Challenges

- Model switching delays: Managing few-second initialization when transitioning between RITM and XMem workflows
- RITM distance sensitivity: Handling model limitation where widely spaced points are interpreted as separate instances
- Hybrid workflow integration: Seamless transitions between AI-assisted segmentation and manual polygon editing
- Session persistence: Ensuring masks and annotations persist with proper export capabilities for both formats
- Interactive feedback: Point Count and Point Radius sliders with real-time canvas updates during intensive sessions

Deliverable

- **Final Deliverables** :
  (Discuss in the form of bullets, what are the next steps to complete the solution, any road blocks / bottlenecks, any support needed from SRIB)

All deliverables relevant to the project have been completed.

- **IP / Paper Publication Plan** :
  (Details of papers / patentable ideas / innovative aspects that can lead to patentable ideas)

The project can be used as a tool for fast generation/correction of annotations, enabling rapid dataset creation.

- **KPIs delivered/Expectations Met**:
  (Planned Expectations shared in Work-let vs Delivered Results)

KPIs delivered meet expectations.

✓ **Module 1:** Given an image, generate manual annotation by creating point-based polygons, and assign suitable labels.

✓ **Module 2:** Given a manually segmented image and other images of same scene from different views as inputs, automatically generate the segmented images for all the views.

✓ **Module 3:** Given an already segmented image, support Interactive Correction, which is a click-based interactive refinement, for correcting the segmentation mask in case of both false positives and negatives.

✓ Basic functionalities supported: browse folder, previous image, next image, create polygon, edit polygon, create segmentation, add segment and remove segment.

Work-let Closure Details

- **Code Upload details:**

Items	Details
KLOC (Number OF Lines of codes in 000's)	27
Model and Algorithm details	XMem++ and RITM (Forked versions - Modified to support CPU)
Is Mid review, end review report uploaded on Git ?	Yes
Link for Git	https://github.com/Prasannakbhat123/Samsung-Prism

- **Data details (if applicable):**

Items	Data folder 1	Data folder 2	Data Folder 3……..
Name & Type of Data (Audio/Image/Video)	-		
Number of data points	-		
Source of Data (self collected, Scrapped, available on open source)	-		
Google drive link/ git link to access data	-		

Note: If data uploaded on google drive, access to be shared to prism.srib@gmail.com

Thank you