International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 12 | Dec - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

SAMurai-An Image Annotation Tool

Mrs K Padmaja[1],Sowjanya R[2], Spandana Shashikiran[3], Thanusha M Hosgoudar[4]

[1]*Assistant Professor, East West Institute of Technology*
[2,3,4,5]*Students, Department of Artificial Intelligence and Data Science, East West Institute of Technology, Bengaluru, India*

***Abstract*--- An image segmentation tool called SAMurai was created with Python for backend processing. It makes use of a Flask server to manage API requests and offer effective system interaction. JavaScript is used in the tool's front-end, which provides a dynamic and intuitive interface for uploading and viewing segmented photos. SAM, or Segment Anything models, are a collection of promptable image segmentation models created by Meta that can produce accurate object segmentations or region masks in an adaptable, on-demand manner using user inputs (referred to as prompts) such as points, boxes, masks, or text in addition to the input picture. High accuracy in detecting objects or features inside photos is guaranteed by this deep learning model. The application is designed to serve fields like computer vision, object detection, and image analysis, and can be extended in the future to support broader data types and advanced functionalities.**

I. INTRODUCTION

In order to carry out object detection and associated tasks, a computer vision technique called image segmentation divides a digital image into distinct groupings of pixels, or image segments. By parsing an image's complex visual data into specifically shaped segments, image segmentation enables faster, more advanced image processing. The Segment Anything Model, a semi-automated image annotation tool, offers a powerful and versatile solution for object segmentation in images, enabling to enhance datasets with segmentation masks generated using box prompts/point prompts. With its fast-processing speed and various modes of inference, SAM is a valuable tool for computer vision applications.

II. LITERATURE SURVEY

The recent trajectory of research in computer vision has seen a marked shift from class-specific segmentation methods towards open-world, promptable, foundation models that generalise across domains. One of the seminal works in this direction is the Segment Anything Model (SAM) (Meta AI), which introduces a unified framework for segmentation via prompts (points, boxes, masks) and a massive dataset of over 11 million images and 1 billion+ masks. [2] This work highlights three key contributions: a prompt-able segmentation task, a model architecture for prompt ingestion, and a large scale data-engine to create SA-1B dataset. [2] Subsequent surveys examine SAM's capabilities and limitations: for example "A Survey on Segment Anything Model (SAM): Vision Foundation Model Meets Prompt Engineering" details SAM's versatility across modalities, but also the gaps in domains such as medical imagery or very high granularity segmentation. [14] In medical imaging, studies such as "Segment Anything Model for Medical Image Analysis: An Experimental Study" report that SAM's performance varies dramatically across modalities (IoU from ~0.11 to ~0.86) and generally performs better with box prompts than with point prompts. [15] In the context of interactive annotation tools, combining UI plus segmentation backbone, works have explored how to adapt SAM to domain-specific tasks (e.g., SAM-Adapter for camouflaged object detection) which underscores the need for end-to-end pipelines not just a segmentation model. [16] Hence, the project — "SAMurai" — sits at the intersection of these threads: promptable segmentation (via SAM), interactive annotation workflows, dataset versioning. The literature suggests such an integrated system is a timely approach to bridging model research with annotation usability and domain-adaptation.

III. METHODOLOGY

The methodology behind Project SAMurai combines the power of deep learning–based image segmentation with an intuitive, web-based interface. Its design philosophy revolves around making state-of-the-art segmentation accessible and interactive, allowing users to visually select regions of interest and instantly obtain high-quality segmentation results.

A. *System Architecture*

SAMurai follows a three-layer architecture —

1) Frontend (User Interface): This interface, which was created with HTML, CSS, and JavaScript, enables users to interactively choose points or boxes that direct the segmentation model, upload images, and view data. It is simple, intuitive, and designed for minimal technical expertise — any user can click on an image to mark an object of interest.
2) Backend (Flask Server): The backend acts as a bridge between the user and the machine learning models. It is implemented in Python using Flask [1] and uses the Segment Anything Model (SAM) or its lighter and quicker variation, MobileSAM, for inference after processing user inputs (pictures and prompts). Requests like uploading, rendering, mask prediction, and storing annotated datasets are handled by the backend.
3) Model Layer (Segmentation Engine): This is the project's computational core. It makes use of Meta's Segment Anything Model (SAM), a prompt-

based segmentation method that employs straightforward user input (referred to as prompts) to segment any object in an image [2]. Various model versions are loaded based on the use case:
o SAM (ViT-Base/Large/Huge): for segmenting static images with great precision.
o MobileSAM: for quicker inference on computers with lower processing power [3].

Together, these components form an interactive segmentation pipeline where a user's visual prompt travels through the backend to the model, and the model's segmented output returns for display and refinement on the UI.

B. *Core Algorithm: The Segment Anything Model (SAM)*

The Segment Anything Model (SAM), a promptable visual transformer-based segmentation system created by Meta AI Research, is the central component of SAMurai [2]. SAM can create segmentation masks based on adaptable human inputs since it is based on the concept of prompt engineering for vision.

Input(prompts): In addition to the image, SAM allows points, boxes, masks, or text prompts.

For instance, a drawn bounding box or a click inside (positive point) or outside (negative point) an item.

Architecture:
SAM consists of three parts [2], [5]:

i. Image Encoder (Vision Transformer): High-dimensional visual features are extracted from the input image using the Image Encoder (Vision Transformer).
ii. Prompt Encoder: Places user prompts (points, boxes, and masks) in the same embedding space as the picture.
iii. Mask Decoder: Predicts accurate segmentation masks in real-time by combining the two embeddings.

Output: A binary or multi-region mask that shows which pixels are associated with the chosen object(s).

Because SAM is trained on the SA-1B dataset — over 11 million images and 1 billion masks [2] — it generalizes exceptionally well to unseen objects, making it ideal for "segment anything" tasks without domain-specific retraining.

SAMurai integrates SAM's open-source checkpoints to generate segmentation masks dynamically on uploaded images or video frames.

C. *Workflow*

1. 1. Upload and Pre-processing: An image is uploaded by the user. It is temporarily saved by the system, which also gets the frames ready for processing.
2. 2. Prompt Generation: The user chooses one or more prompts, like clicking on the object of interest or creating a bounding box.
3. 3. Segmentation Prediction: The image and the prompt data are sent to the backend. The backend creates segmentation masks after loading the relevant SAM model (such as MobileSAM).
4. 4. Visualization and Refinement: The browser displays the final mask on the same image. By adding or removing prompts, users can fine-tune until the mask corresponds to the desired area.
5. 5. Annotation Storage: After segmentation and information (labels, mask coordinates) are completed, they are saved for future retrieval, version control, or dataset generation.

D. *Key Features and Advantages*

- Prompt-based Segmentation: Rather of manually tracing polygons, users can quickly acquire masks by giving simple hints (clicks or boxes) [2].
- Multi-Modal Support: Aids in radiology and biomedical segmentation as it works with both regular images and DICOM medical images [6].
- Real-Time Inference: Even with low-end technology, lightweight versions like MobileSAM allow for almost instantaneous response [3].
- Dataset Versioning: SAMurai is appropriate for large dataset generation projects since it incorporates a Git-based mechanism to monitor annotations.
- Extensible Design: Additional models and plugins can be integrated for further refinement and extension of the application.

E. *Applications*

Because of its versatility, the tool can be used in a variety of fields:

- Medical imaging: To identify anatomical features, lesions, and tumors in MRI or CT scans [6], [9].
- Aerial and satellite imagery: To detect changes in the environment or land-use areas [11].
- Industrial Inspection: Finding flaws in production and quality-control processes [12].
- Research and Dataset Creation: In computer vision research, speeding up manual annotation for the creation of new datasets [13].

IV. CONCLUSION

In this work, we have introduced SAMurai, an interactive annotation and segmentation toolkit that incorporates promptable segmentation models (via SAM and its variants) into a full-stack pipeline that includes multi-frame tracking, upload, frame extraction, segmentation, UI feedback, and dataset versioning. The system bridges the gap between powerful segmentation foundation models and practical annotation workflows needed for domains such as medical imaging, object detection and dataset creation.

The literature on segmentation models, especially the SAM family, demonstrates the trend toward prompt-based, general-purpose architectures that are capable of segmenting "anything" without the need for retraining. However, real-world annotation and dataset operations require more than just these models. In order to facilitate the uploading, prompting, mask generation, correction, versioning, and reuse of results, SAMurai packages these models into a functional framework. Usability, reusability, and extensibility are key components of our system.

Prompt-based segmentation provides far faster annotation cycles, less user effort, and faster dataset creation in use cases and experiments. SAMurai provides a solid foundation for interactive segmentation tasks. 3D segmentation, collaborative annotation settings, and deeper domain adaptation (e.g., fine-tuning SAM using adapter-modules for under-performing scenes) can all be included to our technology in future work.

SAMurai contributes to the next generation of annotation tools by combining modern segmentation models with annotation workflows. It is not just “segment anything” tool but also allows “efficient annotation, robust versioning and broad reuse”.

V. REFERENCES

[1] Grinberg, M. *Flask Web Development: Developing Web Applications with Python*, O’Reilly Media, 2018.
[2] Kirillov, A. et al. “Segment Anything,” *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023. [Online]. Available: https://segment-anything.com/
[3] Zhang, L. et al. “MobileSAM: Efficient Segmentation Anything Model with Mobile-Friendly Design,” *arXiv preprint arXiv:2306.14289*, 2023.
[4] Meta AI Research, “Segment Anything 2 (SAM 2): A new foundation model for video segmentation,” Meta AI Blog, 2024. [Online]. Available: https://ai.meta.com/research/publications/segment-anything-2/
[5] Dosovitskiy, A. et al. “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale,” *ICLR*, 2021.
[6] Ma, J. et al. “Segment Anything Model for Medical Image Segmentation: Opportunities and Challenges,” *Computers in Biology and Medicine*, 2024.
[7] Liu, S. et al. “Grounding DINO: Marrying DINO with Grounded Pre-Training for Open-Set Object Detection,” *arXiv:2303.05499*, 2023.
[8] Minderer, M. et al. “Simple Open-Vocabulary Object Detection with Vision Transformers,” *ECCV*, 2022.
[9] Tang, Y. et al. “SAM for Radiology: Zero-shot Transfer for Lesion Segmentation,” *arXiv:2304.11293*, 2023.
[10] Wang, C. et al. “Vision Transformers for Driving Scene Understanding,” *IEEE Transactions on Intelligent Transportation Systems*, 2024.
[11] Zhao, Y. et al. “Applying SAM to Remote Sensing Imagery,” *Remote Sensing Letters*, 2024.
[12] Chen, K. et al. “AI-driven Quality Control using Vision Transformers,” *IEEE Access*, 2023.
[13] Everingham, M. et al. “The Pascal Visual Object Classes (VOC) Challenge,” *IJCV*, 2010.
[14] Chaoning Z. et al. “A Survey on Segment Anything Model (SAM): Vision Foundation Model Meets Prompt Engineering”
[15] Maciej A. et al. “Segment Anything Model for Medical Image Analysis: An Experimental Study”
[16] Tianrun C. et al. “SAM-Adapter: Adapting Segment Anything in Underperformed Scenes