

Name: Pratik Sambhaji Ghute
Gmail: pratikghute54@gmail.com
Mob/Whatsapp No.: +91 7057 033 662.

Approach for SAM2 Single-Shot Object Tracking Task

Environment Setup and Initial Evaluation of SAM2

- Configured the environment with necessary libraries, including SAM2, and acquired the dataset required for the assignment.
- Performed an initial assessment of SAM2's segmentation capabilities by running inference on a subset of images to establish a baseline performance.

Dataset Understanding

- Analyzed the dataset, consisting of 10 objects, each with 50 images. Each image was paired with one or more mask images where the object was highlighted in white against a black background.

Bounding Box Extraction

- Implemented the `get_bounding_box` function to extract precise bounding boxes from mask images. This function used contour detection after thresholding the images to binary, exploiting the clear contrast between the object and background. Bounding box coordinates (xmin, ymin, xmax, ymax) were stored in a JSON file as ground truth.

Performance Observation

- Evaluated SAM2's performance by tracking objects across subsequent images using bounding boxes from the first image of each object. Assessed its segmentation ability under various conditions such as lighting and angles.

Model Limitations

- Observed that SAM2 occasionally failed to track or mask objects, even with multiple occurrences in the image. This pointed to limitations in the model's consistency in object tracking and detection.

Function Modification for Multi-threading

- Enhanced the `track_item_boxes` function to support multi-threading, allowing for dynamic path and variable management across threads. This optimization improved the efficiency and speed of bounding box predictions for multiple objects.

Test Set Construction

- **Exclusion of Training Images:** Excluded the first image of each object from the test set to avoid bias, as these images were used for training.

- **Handling Multiple Instances:** For images with multiple masks, only the mask associated with the tracked object was considered valid. Given the single-shot tracking task, only one mask per image was relevant to ensure a focused evaluation.
- **Handling No Detections:** For images where SAM2 did not detect any object, only one mask image was included in the test set to represent the ground truth. This approach aligned with the expectation of having a single prediction per image.
- Constructed the test set dictionary based on these criteria and generated `ground_truth.json` and `predictions.json` in COCO format for evaluation.

Performance Evaluation

- **Overall Precision and Recall:** SAM2 demonstrated moderate precision and recall. Precision metrics indicated challenges in accurate localization, particularly at higher alignment levels. Recall metrics showed SAM2 detected approximately 23% of objects in the test set.
- **Performance by Object Size:**
 - **Small Objects:** No small objects were present in the test set, resulting in metrics for this category being reported as -1.000.
 - **Medium Objects:** Achieved a precision of 0.343 and recall of 0.348, indicating reasonable performance on medium-sized objects.
 - **Large Objects:** Precision and recall were lower, at 0.138 and 0.225 respectively, suggesting difficulties in accurately detecting and tracking large objects.

Summary

The evaluation reveals that SAM2 performs reasonably well with medium-sized objects but shows limitations with large objects and lacks assessment for small objects due to their absence in the test set. The results indicate that while SAM2 is functional, further improvements are necessary to enhance its precision and tracking capabilities across varying object sizes.