Single-View Depth Map Integration for Robust Eyeglass Temple Extraction

Ervinas Gisleris (STD IEEE), **Artūras Serackis** (SM IEEE), **Dalius Matuzevičius** (SM IEEE)

E-mail: {ervinas.gisleris | arturas.serackis | dalius.matuzevicius}@vilniustech.lt,
Department of Electronic Systems, Vilnius Gediminas Technical University (VILNIUS TECH)

Introduction

The extraction of eyeglasses temples from single RGB images is challenging due to the thin, elongated structure of the temples, specular reflections, and blending into the background. Conventional 2D segmentation methods operating purely on color or intensity often struggle to capture these subtle geometric cues, leading to fragmented or inaccurate masks. Recent advances in single-view (monocular) depth estimation provide a way to recover pseudo-3D structure from a single image, offering complementary information to RGB image. In this work, we investigate how integrating monocular depth maps could lead to more robust extraction of eyeglass temples within a training-free segmentation pipeline.

Aims and Goals

The aim of this research is to assess whether integrating monocular depth estimates enhances training-free segmentation of eyeglass temples, and to determine which monocular depth estimation models provide the most effective support for this task. Main objectives:

- Develop a training-free segmentation pipeline for eyeglass temples.
- Create a test dataset of eyeglass temples segmentation evaluation.
- Compare monocular depth detectors within the context of temple segmentation, using the intersection-over-union (IoU) metric.

Methods

We assembled a dataset of 100 real eyeglass images captured from a side view against a white background, using a slightly elevated camera angle to ensure that both temples were visible without overlapping. For each image, we manually annotated two separate binary masks: one for the back (far) temple and one for the front (near) temple. To constrain the evaluation strictly to the temples and avoid interference from the eyeglass frame, all masks and corresponding segmentations were cropped perform evaluation solely within this frame-excluded region.

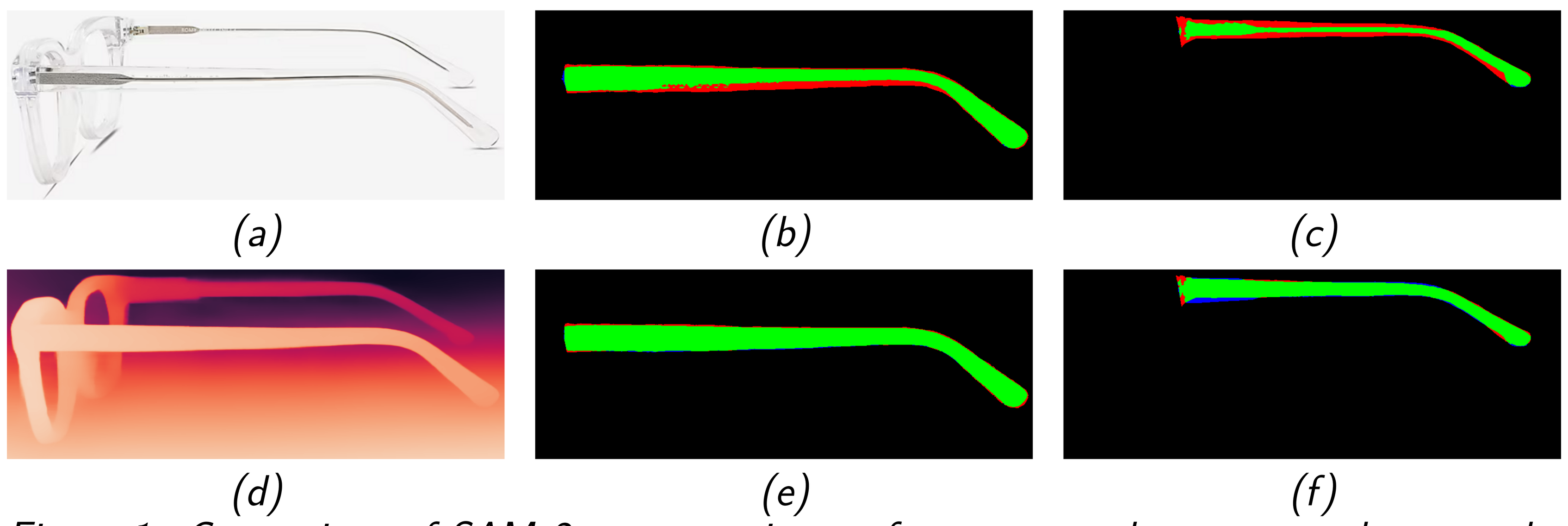

Figure 1: Comparison of SAM 2 segmentation performance on the same eyeglass sample. (a) Real RGB input; (b, c) Worst-case RGB-only segmentation results showing IoU overlays for the front (74.4%) and back (52.0%) temples; (d, f) Best-case segmentation using Grayscale + Depth (Depth Pro), with substantially improved IoU for the front (93.4%) and back (85.8%) temples. Green indicates correctly segmented regions, red indicates missed ground-truth areas, and blue highlights over-segmented regions.

In addition to the RGB images, we generated monocular depth maps using four state-of-the-art monocular depth estimators: Depth Anything, Depth Anything V2, Depth Pro, and UniDepth V2. SAM 2 was used as the zero-shot segmentation model, creating several segmentation variants: RGB-only, depth-only images from each depth model, and pseudo-RGB composites combining grayscale RGB with one or two depth channels. All segmentation outputs were saved as binary masks for front and back temples. The IoU was evaluated per image, per temple, and per segmentation variant.

Results

The RGB input serves as a strong baseline with median IoUs greater than 90% for both temples. Depth-only inputs using Depth Anything, Depth Anything V2, and Depth Pro produce similar distributions, indicating that depth maps alone are reasonably informative but rarely surpass RGB. UniDepth V2, however, exhibits a notable drop in performance for the back temple, with substantially lower medians and a wide spread of low outliers, revealing instability in its depth reconstructions for thin structures. In contrast, the "grayscale+depth" composition variants demonstrate noticeable improvements: all composite methods elevate the median IoU and reduce the number of low-score outliers.

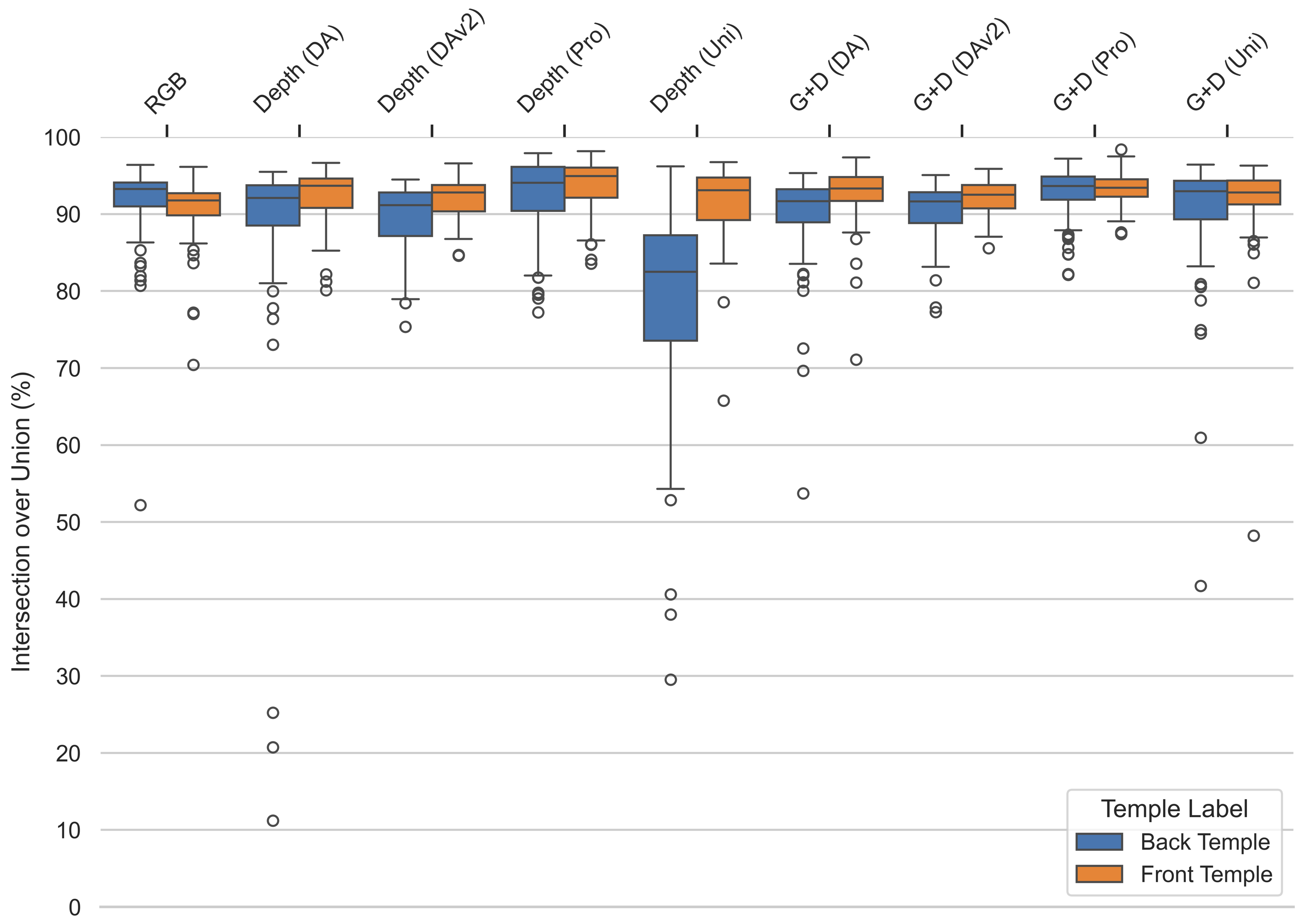

Figure 2: Distribution of IoU scores for front and back eyeglass temples across all evaluated SAM 2 input modalities.

Across all 100 samples, combining grayscale with depth as a pseudo-RGB input consistently improved segmentation accuracy compared to the RGB only baseline. The composite strategy using Depth Pro produced the best results, yielding the highest mean IoU for both temples (92.8%$\pm$3.2 for the back temple and 93.3%$\pm$2.0 for the front temple). Pure depth based segmentation showed mixed results, but depth integration proved most beneficial in difficult cases where the eyeglass temples were transparent or visually blended into the white background situations in which RGB only segmentation frequently under segmented or missed large regions.

Conclusions

- Integrating monocular depth with grayscale input provides the greatest improvement in the accuracy of eyeglass temple segmentation compared to using RGB alone.
- Depth-only segmentation is generally comparable to RGB but varies strongly between depth models, with UniDepth V2 performing the poorest.
- Depth cues are especially valuable for transparent or low-contrast temples, indicating clear benefits for challenging fine-geometry segmentation tasks.

VILNIUS TECH
Department of Electronic Systems

This project has received funding from the Research Council of Lithuania (LMTLT), agreement No S-ITP-24-12.