Evaluation of Multimodal AI Models for Eyeglasses Recognition

Henrikas GIEDRA, Dalius MATUZEVIČIUS

E-mail: {henrikas.giedra | dalius.matuzevicius}@vilniustech.lt,

Department of Electronic Systems, Vilnius Gediminas Technical University (VILNIUS TECH)

Introduction

Eyeglasses detection is a critical preprocessing step for biometric, demographic, and facial recognition systems, but manual labeling of eyeglasses in large datasets is labor-intensive and prone to inconsistencies. This research investigates modern vision–language models (VLMs) as automatic eyeglasses labelers, focusing on localization quality and labeling consistency.

Aims and Goals

We aim to benchmark state-of-the-art VLMs (Gemini and Qwen families) for eyeglasses detection on human faces. Our goals are to:

- quantify detection accuracy, consistency, and robustness under occlusions, reflections, and diverse eyewear styles;
- analyze systematic biases and failure patterns in model predictions.

Methods

We used 1,549 images from the CelebAMask-HQ dataset, with eyeglasses annotated as axis-aligned bounding boxes $[xmin, ymin, xmax, ymax]$; an example is shown in Figure 1.

Figure 1: Example of CelebAMask-HQ dataset.

Each image is queried with multiple VLMs at temperatures 0.0 and 0.5 using a strictly defined unified prompt. The returned bounding boxes are parsed, validated, and evaluated using the Intersection over Union (IoU) metric.

Results

Across models, VLM-generated eyeglasses bounding boxes achieve high overlap with ground truth, with the best Gemini and Qwen variants reaching mean IoU values in the high 0.8 to low 0.9 range (see Fig. 2).

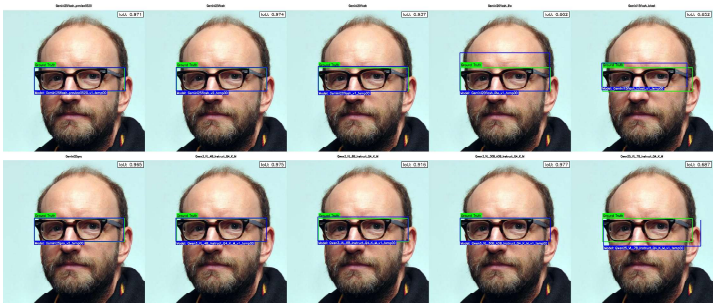

Figure 2: Qualitative comparison of all tested VLMs on eyeglasses detection.

Failure cases cluster around rimless, highly transparent, or heavily shadowed glasses, where models struggle to separate eyewear from surrounding facial features (Fig. 3).

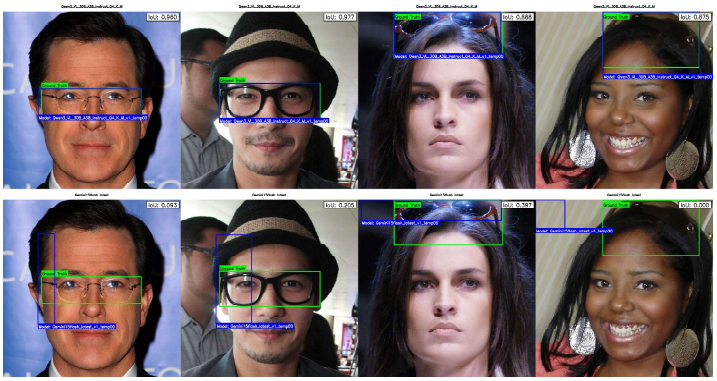

Figure 3: Comparison of best and worst performing models on challenging eyeglasses detection cases.

Pairwise IoU matrices show that most models produce very similar eyeglasses locations and thus strong inter-model agreement; increasing the temperature parameter from 0.0 to 0.5 changes IoU only slightly but affects the rate of invalid or unusable predictions (Table 1).

Table 1: Average IoU with ground truth and between models, including invalid predictions and temperature effects.

	gemini-2.5-flash-preview	gemini-2.5-flash	gemini-2.0-flash	gemini-2.0-flash-lite	gemini-1.5-flash-latest	gemini-2.5-pro	Qwen3-VL-4B	Qwen3-VL-8B	Qwen3-VL-30B-A3B	Qwen25-VL-7B
IoU	0.8885 (-0.0064)	0.8931 (-0.0113)	0.8160 (-0.0066)	0.5286 (-0.0232)	0.4998 (+0.0022)	0.8739 (+0.0020)	0.9006 (-0.0043)	0.8958 (-0.0032)	0.9105 (-0.0022)	0.7412 (-0.0132)
Invalid pred. (temp: 0.0)	11	9	6	1	6	21	1	1	0	3
Invalid pred. (temp: 0.5)	9	12	5	1	7	16	3	1	0	4

Average IoU between models

gemini-2.5-flash-preview		0.9587	0.8410	0.5365	0.5219	0.8980	0.8993	0.8962	0.9110	0.7198
gemini-2.5-flash	0.9587		0.8423	0.5386	0.5231	0.9024	0.9038	0.9012	0.9163	0.7225
gemini-2.0-flash	0.8410	0.8423		0.5400	0.5217	0.8243	0.8264	0.8250	0.8377	0.6695
gemini-2.0-flash-lite	0.5365	0.5386	0.5400		0.4461	0.5273	0.5378	0.5415	0.5432	0.4945
gemini-1.5-flash-latest	0.5219	0.5231	0.5217	0.4461		0.5070	0.5129	0.5187	0.5181	0.3968
gemini-2.5-pro	0.8980	0.9024	0.8243	0.5273	0.5070		0.8820	0.8803	0.8939	0.7149
Qwen3-VL-4B	0.8993	0.9038	0.8264	0.5378	0.5129	0.8820		0.9316	0.9376	0.7416
Qwen3-VL-8B	0.8962	0.9012	0.8250	0.5415	0.5187	0.8803	0.9316		0.9322	0.7356
Qwen3-VL-30B-A3B	0.9110	0.9163	0.8377	0.5432	0.5181	0.8939	0.9376	0.9322		0.7409
Qwen25-VL-7B	0.7198	0.7225	0.6695	0.4945	0.3968	0.7149	0.7416	0.7356	0.7409	

Conclusions

- Current VLMs can substantially reduce manual effort, producing reliable, scalable eyeglasses annotations with minimal human intervention.
- Deterministic (low-temperature) inference improves consistency, and larger multimodal models better generalize to challenging visual conditions.
- Identified biases and failure modes highlight weak areas of VLMs and provide insights for improving pipeline robustness.

VILNIUS TECH Department of Electronic Systems

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