Olym+: Real-time Exercise Tracking and Repetition Counting on Edge Devices using BlazePose

Atharv D. Khisti, Aditya Jadhav, Muskan Dosi, Hiten Shah
MIT Academy of Engineering, Pune, India
Email: {atharv.khisti, aditya.j, muskan.d, hiten.s}@mitaoe.edu.in

***Abstract*—We present *Olym+*, a real-time exercise tracking system that performs pose estimation and repetition counting in the browser/edge using Google's BlazePose. Our pipeline combines (i) on-device 33-landmark pose inference, (ii) joint-angle computation with smoothing, and (iii) a two-state finite-state machine (FSM) per exercise to robustly count repetitions. On a test of 100 biceps-curl repetitions collected across consumer webcams, Olym+ achieves Accuracy 72.7%, Precision 88.9%, Recall 80.0%, and F1 84.2. We release design details and evaluation protocol to facilitate reproducibility and fair comparison with future baselines.**[1]

***Index Terms*—Human Pose Estimation, BlazePose, MediaPipe, Repetition Counting, Edge Computing, Fitness Tech**

I. Introduction

Human Pose Estimation (HPE) enables real-time understanding of body kinematics for applications in fitness, rehab, and AR. Google's BlazePose offers a lightweight, 33-landmark model suitable for on-device inference at >30 FPS on mobile/edge, making it ideal for consumer-grade exercise tracking [1], [2]. MediaPipe Pose wraps BlazePose with an efficient, cross-platform pipeline [3], [4].

We propose *Olym+*, an end-to-end vision system for repetition counting. Our contributions:

- **Edge-first pipeline** combining BlazePose inference, angle-based features, temporal smoothing, and an FSM, deployable in-browser or on low-power devices.
- **Open evaluation protocol**: metrics, confusion matrix, and error taxonomy for repetition counting on consumer webcams.
- **Empirical results**: on 100 biceps-curl repetitions, Olym+ achieves Precision 88.9% and F1 84.2, competitive with recent rep-counting methods using landmark+angle features [5], [7], [9].

II. Related Work

A. Pose Estimation

BlazePose introduced a fast, on-device pipeline for single-person full-body keypoints (33 landmarks) using heatmap+regression and temporal tracking [**?**], [1]. MediaPipe Pose generalizes this to cross-platform deployments [3], [4]. Comparative reviews position BlazePose as a high-speed alternative to MoveNet/HRNet/OpenPose with broader landmark coverage for fitness [5], [6].

[1]Project context: Olym+ also contains a quick e-commerce module and assistant chatbot for guidance; this paper focuses on the vision pipeline.

Fig. 1. Olym+ pipeline: MediaPipe(BlazePose) → smoothing → angles → FSM → counter.

B. Repetition Counting

Counting reps from landmarks commonly uses peak detection on joint angles, FSMs, or sequence models (e.g., LSTMs/Transformers) [7]–[9]. Angle+landmark hybrids improve robustness to camera viewpoint and partial occlusion [7]. We adopt an angle+FSM approach for simplicity and edge suitability.

III. System Overview

Fig. 1 shows the pipeline:

1) **Capture**: 720p webcam or phone (DroidCam).
2) **Pose Inference**: BlazePose via MediaPipe (33 landmarks, 2D + optional z).
3) **Preprocess**: Exponential Moving Average (EMA) smoothing of landmarks.
4) **Angles**: Compute joint angles (e.g., elbow, knee) using vector dot-product.
5) **FSM**: Two-state per exercise (*down* → *up* transitions) with hysteresis thresholds.
6) **Count & UI**: Increment on valid state transitions; live overlay + rep display.

IV. Methods

A. Angle Computation

Given three joints A, B, C (with B at the vertex), angle

$$\theta = \cos^{-1}\left(\frac{(\vec{A}-\vec{B})\cdot(\vec{C}-\vec{B})}{\|\vec{A}-\vec{B}\|\,\|\vec{C}-\vec{B}\|} \right). \tag{1}$$

We use elbow angle for curls and hip/knee for squats. To reduce jitter:

$$\hat{\theta}_t = \alpha\theta_t + (1-\alpha)\hat{\theta}_{t-1}, \quad \alpha \in [0.2, 0.4]. \tag{2}$$

B. State Machine & Thresholding

For biceps curls we define:

- *down* state: $\hat{\theta} \geq \theta_{\text{down}}$ (e.g., 150°).
- *up* state: $\hat{\theta} \leq \theta_{\text{up}}$ (e.g., 40°).

A valid *down* → *up* transition increments the count. Hysteresis and a min-duration window suppress bounce and micro-movements.

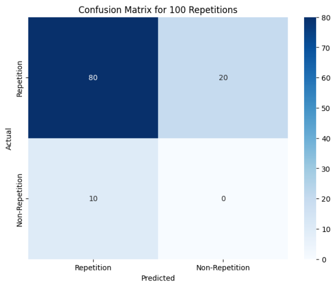

Fig. 2. Confusion Matrix

C. Implementation Details

We built a web app (React) with MediaPipe Pose running in-browser; counting logic in TypeScript; optional Python (OpenCV) for offline analysis. Inference runs at $\sim$25 FPS on a typical laptop webcam.

V. Experimental Setup

Scenario: Biceps curl test with 100 intended repetitions (multiple users, single camera, indoor lighting). **Ground Truth:** Manual frame-accurate count. **Metrics:** Accuracy, Precision, Recall, F1, and confusion matrix.

Confusion matrix counts (your experiment):

$$\mathrm{TP} = 80,\ \mathrm{FP} = 10,\ \mathrm{FN} = 20,\ \mathrm{TN} = 0.$$

$$\text{Accuracy} = \frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN}} = \frac{80+0}{80+0+10+20} \approx 72.7\% \tag{3}$$

$$\text{Precision} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} = \frac{80}{80+10} \approx 88.9\% \tag{4}$$

$$\text{Recall} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} = \frac{80}{80+20} = 80.0\% \tag{5}$$

$$\mathrm{F1} = \frac{2\cdot\text{Prec}\cdot\text{Rec}}{\text{Prec}+\text{Rec}} \approx 84.2\%. \tag{6}$$

VI. Results

Fig. **??** shows metric bars and the confusion matrix from the experiment.

Observations. High precision indicates few false counts, while recall losses stem from missed reps under fast/occluded motion or partial out-of-frame poses—consistent with prior reports on rep counting with landmarks+angles [7], [9].

VII. Error Analysis

We annotate common failure modes:

1) *Occlusion/Framing*: elbow or wrist exiting the frame $\Rightarrow$ missed reps (FN).
2) *Threshold drift*: user-specific ROM (range of motion) causes under/over-count if $\theta_{\text{up/down}}$ are not personalized.
3) *Camera viewpoint*: large yaw/pitch impacts 2D angles; multi-view or world-coordinates could help [3].

VIII. Ablations & Improvements

(A) Personalized thresholds: learn $\theta_{\text{up/down}}$ from a 3–5 rep calibration set. **(B) Temporal models**: augment FSM with a tiny 1D-Conv or GRU on angle sequences (edge-friendly) [8]. **(C) 3D cues**: use MediaPipe world landmarks for basic depth disambiguation [4]. **(D) Multi-joint voting**: combine elbow + shoulder velocity to reject spurious counts [7].

IX. Deployment Notes

Olym+ runs client-side (privacy) and scales via a thin API for logging. CI/CD on AWS; static assets on S3; optional GPU not required for BlazePose at webcam resolution.

X. Conclusion

We demonstrated a practical, edge-first repetition counter using BlazePose with transparent accuracy analysis. Future work includes calibration-based thresholds, sequence models, and evaluation on multi-exercise datasets.

Reproducibility Checklist

Code links, exact thresholds, and videos (with consent) should be released with: (i) metric script, (ii) confusion matrix generator, (iii) demo UI, (iv) calibration notebook.

References

[1] V. Bazarevsky et al., “BlazePose: On-device Real-time Body Pose tracking,” *arXiv:2006.10204*, 2020. link. :contentReference[oaicite:0]index=0

[2] Google Research Blog, “On-device, Real-time Body Pose Tracking with MediaPipe BlazePose,” 2020. link. :contentReference[oaicite:1]index=1

[3] J.-W. Kim et al., “Human Pose Estimation Using MediaPipe Pose,” *Applied Sciences*, 2023. link. :contentReference[oaicite:2]index=2

[4] Google AI Edge, “Pose Landmarker Task,” 2025. link. :contentReference[oaicite:3]index=3

[5] F. Roggio et al., “A comprehensive analysis of machine learning pose estimation methods,” 2024. *PLOS One/PMC*. link. :contentReference[oaicite:4]index=4

[6] A. Tharatipyakul et al., “Deep learning-based human body pose estimation in exercise assessment,” 2024. *Sensors*. link. :contentReference[oaicite:5]index=5

[7] H. Chen et al., “Joint-Based PoseRAC with Improved Performance,” 2023. *arXiv:2308.08632*. link. :contentReference[oaicite:6]index=6

[8] F. Japhne et al., “Fitcam: detecting and counting repetitive exercises,” *Journal of Big Data*, 2024. link. :contentReference[oaicite:7]index=7

[9] T. Alatiah and C. Chen, “Recognizing Exercises and Counting Repetitions in Real Time,” *arXiv:2005.03194*, 2020. link. :contentReference[oaicite:8]index=8

[10] AIP Conf. Proc., “Human pose estimation using BlazePose,” 2024. link. :contentReference[oaicite:9]index=9