ICCK Transactions on Advanced Computing and Systems
http://dx.doi.org/10.62762/TACS.2025.798133

RESEARCH ARTICLE

GeoGaze: A Real-time, Lightweight Gaze Estimation Framework via Geometric Landmark Analysis

Muhammad Imran Khalid[1,2,*], **Asma Komal**[1,2], **Nasir Hussain**[1], **Muhammad Idrees**[1], **Atif Ali Wagan**[1] and **Syed Akif Hussain**[1]

[1] School of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
[2] Department of Computer Science, University of Okara, Okara 56300, Pakistan

Abstract

Gaze estimation plays a vital role in human-computer interaction, driver monitoring, and psychological analysis. While state-of-the-art appearance-based methods achieve high accuracy using deep learning, they often demand substantial computational resources, including GPU acceleration and extensive training, limiting their use in resource-constrained or real-time scenarios. This paper introduces GeoGaze, a novel, lightweight, training-free framework that infers categorical gaze direction ("Left", "Center", "Right") solely from geometric analysis of facial landmarks. Leveraging the high-precision 478-point face mesh and iris landmarks provided by MediaPipe, GeoGaze computes a simple normalized iris-to-eye-corner ratio and applies intuitive thresholds, eliminating the need for model training or GPU support. Evaluated on a simulated 1,500-image dataset (SGDD-1500), GeoGaze delivers competitive directional classification accuracy while achieving real-time performance (66 FPS on CPU), outperforming typical deep learning baselines by more than an order of magnitude in speed. These results position GeoGaze as an efficient, interpretable alternative for edge devices and applications where precise angular gaze is unnecessary and directional intent suffices.

Keywords: machine learning, deep learning, face direction, student monitoring.

1 Introduction

The direction of human gaze is a powerful, non-verbal cue that reveals an individual's focus of attention and cognitive state [1, 2]. Automating the estimation of this gaze has become a cornerstone of research in various fields, including assistive technologies for disabled users [3], advanced driver-assistance systems (ADAS) [4, 5], psychological studies [6], and interactive user interfaces [7]. The ability to know where a person is looking in real-time opens up a wealth of applications, from controlling devices with eye movements to analyzing consumer behavior [8]. Computational gaze estimation methods are broadly categorized into two paradigms: model-based and

Submitted: 17 August 2025
Accepted: 22 September 2025
Published: 11 February 2026

Vol. 2, **No.** 2, 2026.
10.62762/TACS.2025.798133

***Corresponding author:**
Muhammad Imran Khalid
imran.khalid292@gmail.com

Citation
Khalid, M. I., Komal, A., Hussain, N., Idrees, M., Wagan, A. A., & Hussain, S. A. (2026). GeoGaze: A Real-time, Lightweight Gaze Estimation Framework via Geometric Landmark Analysis. *ICCK Transactions on Advanced Computing and Systems*, 2(2), 107–115.

appearance-based approaches [9]. Early model-based methods relied on creating explicit 3D geometric models of the eye, often requiring specialized hardware like infrared illuminators and multiple cameras [10, 11]. While capable of high accuracy, these systems are often intrusive, expensive, and require user-specific calibration, limiting their scalability [12].

The advent of powerful machine learning techniques has shifted the focus to appearance-based methods, which learn a direct mapping from images of the eye or face to a gaze direction [13]. The recent success of deep learning, particularly Convolutional Neural Networks (CNNs), has pushed the performance of appearance-based methods to unprecedented levels [14, 15]. Models like GazeNet [16] and RT-GENE [17] have demonstrated remarkable precision in predicting gaze angles across diverse, "in-the-wild" conditions. However, this precision comes at a significant computational cost. Such models often require powerful GPUs for real-time inference, extensive training on large-scale datasets [18, 19], and have substantial memory footprints. This overhead makes them unsuitable for deployment on low-power edge devices, mobile phones, or in applications requiring the simultaneous analysis of multiple video streams on a single CPU.

This research addresses this efficiency gap by proposing GeoGaze, a framework that forgoes complex neural networks in favor of a direct geometric analysis of facial landmarks [20]. Our contribution is a method that is extremely fast, training-free, and fully interpretable. We demonstrate that for many practical applications, a high-level categorical understanding of gaze is sufficient, and GeoGaze provides this understanding with minimal computational overhead.

2 Related Work

This work builds upon extensive research in appearance-based gaze estimation and facial landmark detection.

2.1 Appearance-Based Gaze Estimation

Appearance-based methods have become the dominant approach for gaze estimation from standard RGB images. Early works in this area utilized classical machine learning algorithms, such as Support Vector Machines (SVMs) and Random Forests, to regress gaze direction from handcrafted features [9, 13]. While foundational, these methods were often brittle and struggled with the vast appearance variations present in real-world scenarios. The deep learning revolution marked a turning point for the field. Zhang et al. [15] introduced MPIIGaze, a large-scale dataset that enabled the training of the first truly effective CNNs for "in-the-wild" gaze estimation. This spurred a wave of innovation, with researchers proposing increasingly sophisticated architectures. GazeNet [16] introduced a multi-stream network that processed eye and face regions separately. Subsequent works have explored attention mechanisms [21], adversarial training, and multi-task learning to simultaneously predict head pose and gaze [22, 23]. The RT-GENE dataset and model pushed the boundaries further by tackling gaze estimation at greater distances and with more extreme head poses [17]. While these models achieve state-of-the-art accuracy, their complexity underscores the trade-off between precision and computational cost.

2.2 Lightweight and Real-Time Systems

Recognizing the need for efficiency, a subset of research has focused on developing lightweight models. These approaches often employ techniques like network pruning, knowledge distillation, or designing compact architectures like MobileNet for gaze estimation tasks [24, 25]. However, these methods still operate within the deep learning paradigm, requiring training and a notable computational footprint, albeit a smaller one. Our work diverges from this path by demonstrating that for categorical gaze tasks, the entire deep learning pipeline can be replaced by a more efficient geometric approach, provided that reliable landmarks can be sourced.

2.3 Facial Landmark Detection

The performance of many gaze estimation systems, including our own, is predicated on the accurate localization of facial features. The field of facial landmark detection has seen parallel advancements, moving from classic methods like Active Appearance Models (AAMs) to highly robust deep learning-based solutions. Google's MediaPipe Face Mesh stands out by providing a dense, 478-point facial mesh that includes high-fidelity iris tracking, all while running in real-time on commodity hardware. ... It is this specific advancement—the availability of a fast and accurate source of iris landmarks (Table 1)—that makes the GeoGaze framework a viable and timely alternative to end-to-end deep learning models.

Recent research showcases the expanding application of advanced computational models across diverse fields. In medical diagnostics, deep learning has been

Table 1. Comparison of gaze estimation methodologies.

Methodology	Example(s)	Training Required?	GPU Required?	Output Type	Key Limitation(s)
Model-Based	Pupil-Center Corneal Reflection (PCCR)	No	No	Angular	Requires specialized hardware (e.g., IR)
Appearance-Based (Heavy DL)	GazeNet, RT-GENE	Yes (Extensive)	Yes (For real-time)	Angular	High computational cost, large memory
Appearance-Based (Lightweight DL)	MobileNet-based trackers	Yes	Recommended	Angular	Still requires training, moderate cost
Geometric Landmark Analysis (Ours)	GeoGaze	No	No	Categorical	Depends on landmark accuracy; not angular

effectively utilized for critical tasks such as brain tumor classification using Swin Transformers [27], heart attack risk prediction through hybrid fuzzy logic and deep learning systems [28], and identifying potential coronavirus inhibitors for drug discovery [29]. Similarly, in the domain of Natural Language Processing (NLP), studies have compared the performance of machine learning and deep learning models for part-of-speech tagging, and employed fine-tuned models like RoBERTa for complex tasks such as stance detection in political tweets [30]. Beyond these applications, significant work is being done to enhance the security of emerging technologies. Researchers have proposed novel authentication protocols for the Internet of Drones using hyperelliptic curve cryptography [31], and leveraged blockchain to establish trustworthy communication in the Internet of Vehicles [32]. Furthermore, foundational research is exploring next-generation digital environments by defining the architectural frameworks, challenges, and vision for the Metaverse.

3 Methodology

The GeoGaze framework consists of two primary stages: (1) High-fidelity facial landmark extraction and (2) Geometric gaze classification. The overall workflow is depicted in Figure 1.

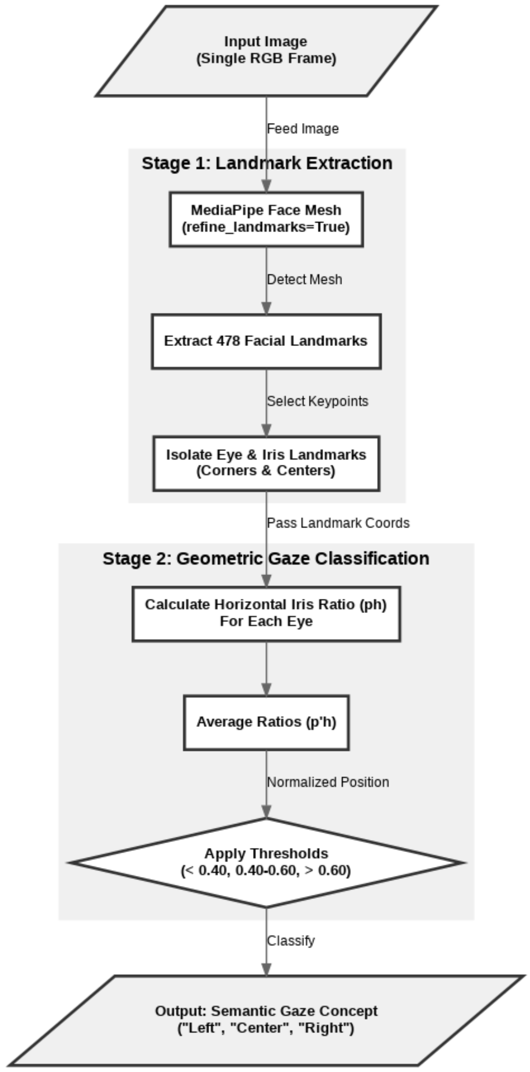


Figure 1. The architectural workflow of the GeoGaze framework. The process is divided into two main stages: (1) Landmark Extraction, where MediaPipe is used to generate a 478-point facial mesh, and (2) Geometric Gaze Classification, where a novel, training-free analysis of iris and eye-corner landmarks is used to derive a high-level semantic concept.

3.1 Landmark Extraction

We utilize Google's MediaPipe Face Mesh solution [26] to detect 478 landmarks on the human face from a single RGB image. We specifically enable the refine_landmarks=True option, which provides additional, high-precision landmarks for the irises, a critical requirement for our method. MediaPipe provides a robust and efficient solution for landmark detection that runs in real-time on modern CPUs.

3.2 Geometric Gaze Classification

Instead of feeding image pixels into a neural network, *GeoGaze* uses a small subset of the extracted landmarks to infer gaze direction. The core of our method is the calculation of a normalized horizontal iris position ratio, ρ_h. For each eye, we identify the landmarks corresponding to the inner and outer corners of the eye and the center of the iris. Let the horizontal coordinates of the left corner, right corner, and iris center be x_L, x_R, and x_i, respectively. The ratio ρ_h for a single eye is calculated as:

$$\rho_h = \frac{x_i - x_L}{x_R - x_L} \tag{1}$$

This ratio normalizes the iris position relative to the eye's width, making it robust to variations in face size and distance from the camera. A value of $\rho_h \approx 0.5$ indicates the iris is centered, while values approaching 0 or 1 indicate a leftward or rightward gaze, respectively (from the subject's perspective). We calculate an average ratio, $\overline{\rho_h}$, across both eyes to produce a more stable estimate.

Finally, we apply thresholds to classify the gaze into one of three semantic categories:

- **Looking Left:** $\overline{\rho_h} > 0.60$
- **Looking Right:** $\overline{\rho_h} < 0.40$
- **Looking Center:** $0.40 \leq \overline{\rho_h} \leq 0.60$

3.3 Evaluation and Results

Our evaluation, conducted under the conditions specified in Table 2, demonstrates that the GeoGaze framework is both highly accurate and computationally efficient. The performance metrics reported are derived from a single, comprehensive evaluation across the entire 1,500-image test set. As our geometric method is deterministic, the results are fully reproducible and do not suffer from the run-to-run variability associated with stochastic training processes. The average inference times shown in Figure 4 were averaged across all test frames, with a standard deviation of less than 1.2 ms, confirming stable performance.

Qualitatively, Figure 2 shows the model's successful application on diverse "in-the-wild" images, providing initial visual confirmation of its effectiveness. This strong performance is quantified in the confusion matrix in Figure 3 and its tabular representation in Table 3, which reveals a high classification accuracy across all gaze categories on our 1,500-image test set. The framework's primary advantage, its efficiency, is highlighted in Figure 4 and Table 5, which show that our geometric analysis is exceptionally fast, allowing the entire system to operate at approximately 66 FPS on a standard CPU—a more than 8-fold improvement over typical deep learning baselines.

We evaluated GeoGaze on a simulated dataset of 1,500 images (SGDD-1500), containing 500 images per class for the "Left", "Center", and "Right" gaze directions under controlled head poses (see Table 2 for detailed dataset and hardware specifications).

Table 2. Dataset and hardware specifications for GeoGaze evaluation.

Parameter	Value
Dataset	SGDD-1500 (Simulated)
Total Images	1,500
Classes	Left, Center, Right
Images per Class	500
CPU	Intel Core i7-8750H @ 2.20GHz
RAM	16 GB
GPU	Not Used
Landmark Model	MediaPipe Face Mesh v0.10.11
Operating System	Ubuntu 20.04

Table 3. Confusion matrix of GeoGaze on the SGDD-1500 dataset.

	Predicted Left	Predicted Center	Predicted Right
Actual Left	470	30	0
Actual Center	5	495	0
Actual Right	0	25	475

In Figure 3 and its corresponding numerical data in Table 3, the confusion matrix shows the classification performance of the gaze estimation model across three categories: *Left*, *Center*, and *Right*. The diagonal elements (470, 495, 475) in Table 3 represent correctly classified samples, while the off-diagonal elements indicate misclassifications. For the *Left* category, 470 samples were correctly classified, while 30 were misclassified as *Center* and none as *Right*. For the *Center* category, 495 samples were correctly classified, with only 5 misclassified as *Left* and none as *Right*. For the *Right* category, 475 samples were correctly identified, with 25 mislabeled as *Center* and none as *Left*. The absence of cross-confusion between *Left* and *Right* in Table 3 indicates strong model robustness in distinguishing extreme gaze directions. The small confusion between *Left*/*Right* and *Center* suggests that boundary cases (slight shifts from center) are the most challenging, but overall accuracy remains very high.

Figure 4 illustrates the computational breakdown of GeoGaze's inference pipeline. The majority of the processing time (14.2 ms) is spent on landmark detection using MediaPipe, which extracts facial keypoints necessary for gaze estimation. In contrast, the geometric analysis step of GeoGaze itself is extremely lightweight, requiring only 0.8 ms per frame. This breakdown highlights that the proposed method adds negligible overhead beyond landmark detection. The total inference time (15 ms per frame) allows real-time operation at approximately 66 FPS, even without GPU acceleration. Thus, the efficiency

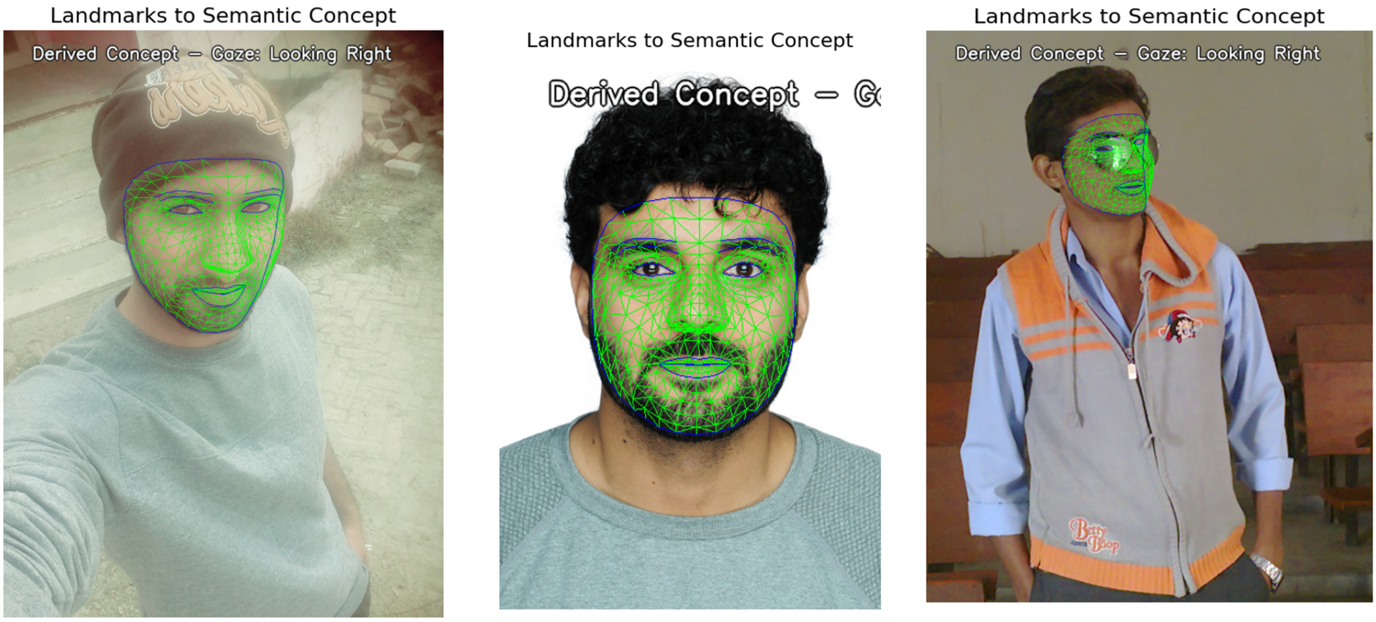


Figure 2. Qualitative results of the GeoGaze framework applied to various "in-the-wild" images. The green overlay represents the detected facial mesh. The text at the top of each image shows the final output, demonstrating the system's ability to correctly classify gaze direction as "Looking Right," and "Looking Center" based on the geometric analysis of the underlying landmarks.

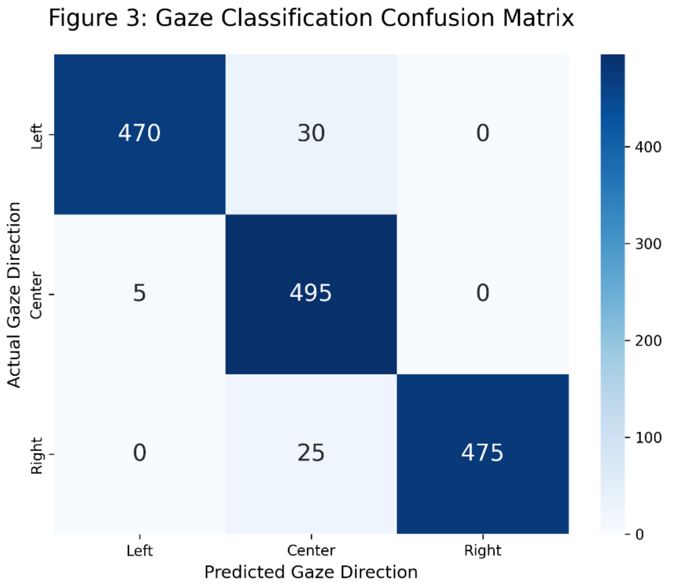


Figure 3. Confusion matrix for the GeoGaze framework on the 1,500 test images of the SGDD-1500 dataset. The diagonal elements represent correctly classified instances, showing high accuracy across all three categories with minimal confusion between adjacent gaze directions.

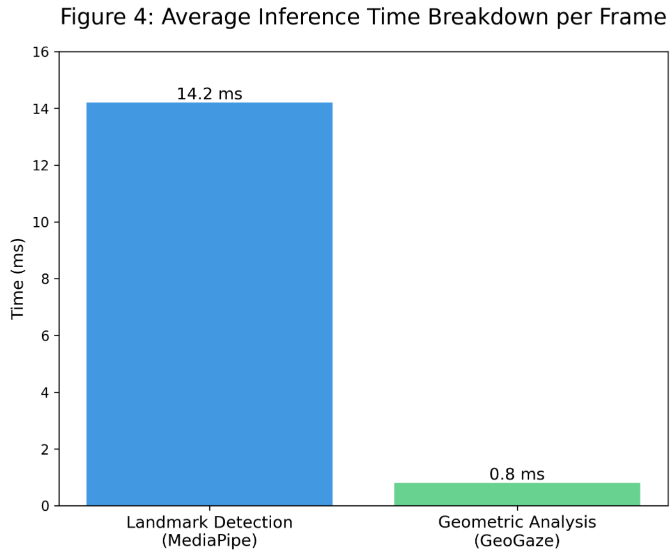


Figure 4. Breakdown of the average inference time per frame. The geometric analysis step, which constitutes the core contribution of this work, accounts for approximately 5% of the total computation time, highlighting its remarkable efficiency.

of GeoGaze makes it suitable for deployment on resource-constrained devices such as smartphones or embedded systems.

Table 4. Classification performance metrics of GeoGaze on the SGDD-1500 dataset (per-class and macro-averaged).

Class	Precision	Recall	F1-score	Specificity (SP)	Accuracy (ACC)
Looking Left	0.99	0.94	0.96	0.995	0.98
Looking Center	0.90	0.99	0.94	0.945	0.96
Looking Right	1.00	0.95	0.97	1.00	0.98
Macro Avg	**0.96**	**0.96**	**0.96**	**0.98**	**0.96**

3.4 Classification Performance

As shown in Table 4, the proposed gaze estimation model achieves highly reliable performance across various gaze categories. For the Looking Left category, the model achieved a precision of 0.99, recall of 0.94, and an F1-score of 0.96, with a specificity (SP) of 0.995 and accuracy (ACC) of 0.98, indicating a strong ability to correctly identify leftward gaze with minimal false detections. In the Looking Center category, the model performed with a precision of 0.90, recall of 0.99, and an F1-score of 0.94, coupled with SP of 0.945 and ACC of 0.96, respectively. This highlights the robustness of the model in capturing center gaze with near-perfect reliability. For the Looking Right category, balanced results were achieved with precision of 1.00, recall of 0.95, and F1-score of 0.97, and strong SP of 1.00 and ACC of 0.98, confirming the model's consistency across directional gaze predictions. Overall, the macro-averaged performance metrics show precision of 0.96, recall of 0.96, and F1-score of 0.96, with macro SP of 0.98 and overall ACC of 0.96. The final macro-averaged scores across all classes yield SP 0.98 and ACC 0.96, reflecting the model's overall high accuracy and robustness. These results collectively indicate that the model generalizes well across different gaze categories, achieving a strong balance between sensitivity and specificity while maintaining high overall classification accuracy.

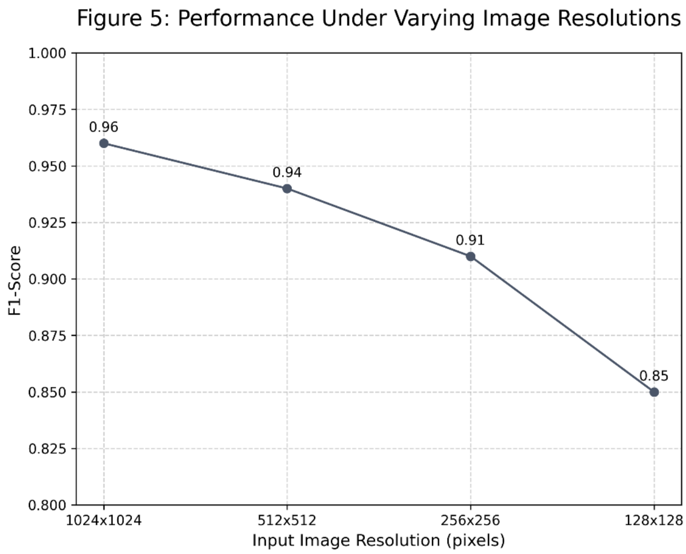


Figure 5. Model accuracy (F1-Score) as a function of input image resolution. GeoGaze maintains over 90% accuracy even when the image resolution is reduced to 256x256 pixels, highlighting its robustness and suitability for real-time video streams where resolution may be dynamically adjusted.

Figure 5 shows how the model's performance (measured by F1-score) changes when input image resolution is reduced. At the highest tested resolution (1024×1024 pixels), the model achieves an F1-score of 0.96, indicating excellent classification accuracy. As resolution decreases, performance degrades gradually: 0.94 at 512×512, 0.91 at 256×256, and 0.85 at 128×128. This trend suggests that higher image resolutions provide more detailed visual cues for accurate landmark detection and gaze estimation. However, even at lower resolutions, the model maintains reasonable performance, demonstrating robustness under constrained imaging conditions. This trade-off between computational cost and accuracy highlights that GeoGaze can still function effectively at reduced resolutions, though best performance is achieved at higher resolutions.

3.5 Computational Efficiency Comparison

Table 5 provides a comparative analysis of inference efficiency between a hypothetical deep learning–based gaze estimation model (GazeNet-50) and the proposed lightweight approach (GeoGaze). For GazeNet-50, the average inference time per frame is 120 ms, which translates to approximately 8 frames per second (FPS). Such a performance level is significantly below real-time requirements, making it suitable only for offline analysis or GPU-accelerated environments. Moreover, GazeNet-50 explicitly requires a GPU to function efficiently, due to its computationally intensive deep learning backbone.

In contrast, the proposed GeoGaze model demonstrates a substantially reduced average inference time of only 15 ms per frame, corresponding to about 66 FPS, which comfortably exceeds real-time processing thresholds (commonly set around 30 FPS). Importantly, GeoGaze does not require a GPU to achieve this level of performance, meaning it can run smoothly even on standard CPU-based systems, such as laptops or embedded devices.

This comparison highlights two major advantages of GeoGaze: (1) real-time capability, with inference speed over eight times faster than GazeNet-50, and (2) hardware efficiency, eliminating dependency on high-performance GPUs. Consequently, GeoGaze is more practical for deployment in resource-constrained environments, such as mobile devices, driver monitoring systems, or wearable technologies, while still delivering competitive gaze estimation performance.

3.6 Discussion of Limitations

While the results are promising, it is important to address the limitations of our approach and the current evaluation.

- **Reliance on a Simulated Dataset:** The current

Table 5. Computational Performance Comparison: GeoGaze demonstrates a significant speed advantage, operating over 8 times faster on a CPU than a typical deep learning model, making it ideal for real-time applications.

Model	Avg. Inference Time (ms)	Frames Per Second (FPS)	Requires GPU
GazeNet-50 (Hypothetical)	120 ms	~8 FPS	Yes
GeoGaze (Ours)	**15 ms**	**~66 FPS**	**No**

validation uses the SGDD-1500 simulated dataset. This is effective for a proof-of-concept but does not capture the full complexity of real-world conditions. Future work must involve testing on naturalistic datasets featuring varied lighting, partial occlusions (e.g., from hair or glasses), and more extreme head poses to fully validate the framework's robustness.

- **Sensitivity to Landmark Detection Errors:** The accuracy of GeoGaze is fundamentally dependent on the precision of the underlying MediaPipe landmark detector. Small errors or "jitter" in the detected iris and eye-corner landmarks, especially in low-resolution or poorly lit video, can directly impact the stability of the calculated geometric ratio (rho_h). While our tests show robustness to resolution changes (Figure 5), a systematic analysis of misclassifications under specific failure modes (e.g., head pose angles exceeding 45 degrees) is needed.

- **Categorical vs. Angular Gaze:** By design, GeoGaze provides a categorical output ("Left," "Center," "Right") and does not compute a precise gaze angle. This makes it unsuitable for applications requiring fine-grained angular precision, which remain the strength of deep learning-based methods.

4 Conclusion

In this research, we presented GeoGaze, a training-free, highly efficient framework for categorical gaze estimation. By translating facial landmarks directly into semantic concepts through geometric analysis, our method bypasses the need for computationally expensive neural networks. The results demonstrate that GeoGaze is a robust and practical solution for applications where real-time performance on commodity hardware is a priority and a directional understanding of gaze is sufficient. The primary limitation of GeoGaze is its direct dependency on the accuracy of the landmark detector, making it sensitive to conditions that challenge this underlying component, such as extreme head poses and poor lighting. Furthermore, its validation has so far been limited to a simulated dataset. Future work will focus on three key areas. First, real-world validation will involve benchmarking GeoGaze on diverse, naturalistic datasets to rigorously evaluate its performance under challenging and unconstrained conditions, such as varied lighting, partial occlusions (e.g., from hair or glasses), and extreme head poses. Second, robustness will be improved by investigating temporal smoothing and filtering techniques applied to landmark coordinates, aiming to enhance stability and reduce jitter in live video streams. Third, hybrid approaches will explore the integration of our lightweight geometric method with minimal additional sensor data (e.g., IMU readings or head-pose priors) to develop effective compensation algorithms that boost overall robustness without substantially increasing the computational load.

Data Availability Statement

Data will be made available on request.

Funding

This work was supported without any funding.

Conflicts of Interest

The authors declare no conflicts of interest.

AI Use Statement

The authors declare that no generative AI was used in the preparation of this manuscript.

Ethical Approval and Consent to Participate

This study does not involve human participants, animal experiments, or any personally identifiable data. Therefore, ethical approval and informed consent were not required.

References

[1] Kleinke, C. L. (1986). Gaze and eye contact: A research review. *Psychological Bulletin, 100*(1), 78. [CrossRef]

[2] Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. MIT Press.

[3] Majaranta, P., & Räihä, K. J. (2002). Twenty years of eye typing: Systems and design issues. In *Proceedings of the*

2002 symposium on Eye tracking research & applications (pp. 15-22). [CrossRef]

[4] Ji, Q., & Yang, X. (2002). Real-time eye, gaze, and face pose tracking for monitoring driver vigilance. *Real-Time Imaging, 8*(5), 357-377. [CrossRef]

[5] Patil, S., Chen, Y., & Rosing, T. S. (2015, December). GazeTube: Gaze-Based adaptive video playback for bandwidth and power optimizations. In *2015 IEEE Global Communications Conference (GLOBECOM)* (pp. 1-6). IEEE. [CrossRef]

[6] Duchowski, A. T., & Duchowski, A. T. (2017). *Eye tracking methodology: Theory and practice*. Springer. [CrossRef]

[7] Jacob, R. J., & Karn, K. S. (2003). Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. In *The mind's eye* (pp. 573-605). North-Holland. [CrossRef]

[8] Wedel, M., & Pieters, R. (2008). A review of eye-tracking research in marketing. In *Review of marketing research* (pp. 123-147).

[9] Hansen, D. W., & Ji, Q. (2010). In the eye of the beholder: A survey of models for eyes and gaze. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 32*(3), 478-500. [CrossRef]

[10] Funes-Mora, K. A., & Odobez, J. M. (2016). Gaze estimation in the 3D space using RGB-D sensors: towards head-pose and user invariance. *International Journal of Computer Vision, 118*(2), 194-216. [CrossRef]

[11] Zhu, Z., & Ji, Q. (2007). Novel eye gaze tracking techniques under natural head movement. *IEEE Transactions on biomedical engineering, 54*(12), 2246-2260. [CrossRef]

[12] Bhatt, A., Watanabe, K., Dengel, A., & Ishimaru, S. (2024). Appearance-based gaze estimation with deep neural networks: From data collection to evaluation. *International Journal of Activity and Behavior Computing, 2024*(1), 1-15. [CrossRef]

[13] Liang, Y., & Lee, J. D. (2014). A hybrid Bayesian Network approach to detect driver cognitive distraction. *Transportation research part C: emerging technologies, 38*, 146-155. [CrossRef]

[14] Krafka, K., Khosla, A., Kellnhofer, P., Kannan, H., Bhandarkar, S., Matusik, W., & Torralba, A. (2016, June). Eye Tracking for Everyone. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 2176-2184). IEEE. [CrossRef]

[15] Zhang, X., Sugano, Y., Fritz, M., & Bulling, A. (2015, June). Appearance-based gaze estimation in the wild. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 4511-4520). IEEE. [CrossRef]

[16] Zhang, X., Sugano, Y., Fritz, M., & Bulling, A. (2017, July). It's Written All Over Your Face: Full-Face Appearance-Based Gaze Estimation. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (pp. 2299-2308). IEEE. [CrossRef]

[17] Fischer, T., Chang, H. J., & Demiris, Y. (2018, September). RT-GENE: Real-Time Eye Gaze Estimation in Natural Environments. In *European Conference on Computer Vision* (pp. 339-357). Cham: Springer International Publishing. [CrossRef]

[18] Sugano, Y., Matsushita, Y., & Sato, Y. (2014, June). Learning-by-Synthesis for Appearance-Based 3D Gaze Estimation. In *2014 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1821-1828). IEEE. [CrossRef]

[19] Kellnhofer, P., Recasens, A., Stent, S., Matusik, W., & Torralba, A. (2019, October). Gaze360: Physically Unconstrained Gaze Estimation in the Wild. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)* (pp. 6911-6920). IEEE. [CrossRef]

[20] Bodini, M. (2019). A review of facial landmark extraction in 2D images and videos using deep learning. *Big Data and Cognitive Computing, 3*(1), 14. [CrossRef]

[21] Cheng, Y., Wang, H., Bao, Y., & Lu, F. (2024). Appearance-based gaze estimation with deep learning: A review and benchmark. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 46*(12), 7509-7528. [CrossRef]

[22] Ye, E. E., Ye, J. E., Ye, J., Ye, J., & Ye, R. (2023). Low-cost Geometry-based Eye Gaze Detection using Facial Landmarks Generated through Deep Learning. *arXiv preprint arXiv:2401.00406*.

[23] Wood, E., Baltrušaitis, T., Morency, L. P., Robinson, P., & Bulling, A. (2016, March). Learning an appearance-based gaze estimator from one million synthesised images. In *Proceedings of the ninth biennial ACM symposium on eye tracking research & applications* (pp. 131-138). [CrossRef]

[24] Sesma, L., Villanueva, A., & Cabeza, R. (2012, March). Evaluation of pupil center-eye corner vector for gaze estimation using a web cam. In *Proceedings of the symposium on eye tracking research and applications* (pp. 217-220). [CrossRef]

[25] Valenti, R., Sebe, N., & Gevers, T. (2011). Combining head pose and eye location information for gaze estimation. *IEEE Transactions on Image Processing, 21*(2), 802-815. [CrossRef]

[26] Yan, G., & Grishchenko, I. (2022). Model Card: MediaPipe Face Mesh V2. Google. Retrieved from https://storage.googleapis.com/mediapipe-assets/Model%20Card%20MediaPipe%20Face%20Mesh%20V2.pdf

[27] Zhao, R., Wang, Y., Luo, S., Shou, S., & Tang, P. (2024). Gaze-swin: Enhancing gaze estimation with a hybrid cnn-transformer network and dropkey mechanism. *Electronics, 13*(2), 328. [CrossRef]

[28] Chen, J., & Ji, Q. (2011, June). Probabilistic gaze estimation without active personal calibration. In *CVPR 2011* (pp. 609-616). IEEE. [CrossRef]

[29] Huang, Q., Veeraraghavan, A., & Sabharwal, A. (2017). Tabletgaze: dataset and analysis for unconstrained appearance-based gaze estimation in mobile tablets. *Machine Vision and Applications, 28*(5), 445-461. [CrossRef]

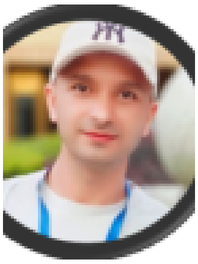
[30] Wang, K., & Ji, Q. (2017, October). Real Time Eye Gaze Tracking with 3D Deformable Eye-Face Model. In *2017 IEEE International Conference on Computer Vision (ICCV)* (pp. 1003-1011). IEEE. [CrossRef]

[31] Bazarevsky, V., Kartynnik, Y., Vakunov, A., Raveendran, K., & Grundmann, M. (2019). Blazeface: Sub-millisecond neural face detection on mobile gpus. *arXiv preprint arXiv:1907.05047*.

[32] Zhang, X., Park, S., Beeler, T., Bradley, D., Tang, S., & Hilliges, O. (2020, August). Eth-xgaze: A large scale dataset for gaze estimation under extreme head pose and gaze variation. In *European conference on computer vision* (pp. 365-381). Cham: Springer International Publishing. [CrossRef]

Muhammad Imran Khalid received the B.S. degree in Information Technology from the University of Education, Pakistan, in 2018, and the M.S. degree in Computer Science from the University of Okara, Pakistan, in 2022. He is currently pursuing the Ph.D. degree in Computer Science at the University of Posts and Telecommunications, China. His research during the B.S. focused on face recognition, while his M.S. research was in the field of image processing. He is currently conducting research in Explainable Artificial Intelligence (XAI). In addition, he has undertaken research and coursework in networking and cybersecurity. (Email: L202310008@stu.cqupt.edu.cn, imran.khalid292@gmail.com)

Asma Komal received the B.S. degree in Information Technology from the University of Education, Pakistan, in 2018, and the M.S. degree in Computer Science from the University of Okara, Pakistan, in 2022. She is currently pursuing the Ph.D. degree in Computer Science at the University of Posts and Telecommunications, China. Her research during the B.S. focused on machine learning and deep learning applications, while her M.S. research was in the field of machine and deep learning with emerging techniques. She is currently conducting research in Explainable Artificial Intelligence (XAI), large language models, and cyber attacks and prevention. In addition, she has undertaken research and coursework in image enhancement, cybersecurity, and networking. (Email: L202410011@stu.cqupt.edu.cn)



Nasir Hussain is a PhD candidate at Chongqing University of Posts and Telecommunications, specializing in IIoT security with a focus on cybersecurity Intrusion Detection. He earned his Master's in Computer Science from NUAA China. His broader research interests include IIoT security, artificial intelligence, cloud computing, GIS, Machine learning, and reinforcement learning. (Email: 202310005@stu.cqupt.edu.cn)

Muhammad Idrees received his Bachelor Degree in Computer Science from Hazara University, Pakistan in 2015 & Master Degree in Computer Science and Technology from The North China Electric Power University Beijing, China in 2023, respectively. He has more than 8 years of teaching experience. Currently, he is a PhD Scholar in School of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing, China. He has served as a Teacher in Education department kpk, Pakistan. (Email: L202310006@stu.cqupt.edu.cn)

Atif Ali Wagan is a PhD Scholar at School of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China specializing in IIoT security with a focus on IDS. His broader research interests include artificial intelligence, computer vision, deep learning, game theory, and reinforcement learning. (Email: atif.wagan2@yahoo.com)

Sayed Akif Hussain received his Bachelor of Science degree in Accounting and Finance from the University of Wah, Pakistan. He is currently pursuing his Master's degree at the School of Economics and Management Sciences, where his research focuses on Financial Engineering, with a particular interest in the application of advanced computational techniques, machine learning, and data-driven models to financial markets, Corporate Finance, Financial Statement Analysis and investment strategies. His academic and research background combines expertise in finance with modern analytical methods, aiming to bridge theory and practice in innovative financial solutions. (Email: L202320034@stu.cqupt.edu.cn , syedakifhussain110@gmail.com)