# Hardware and Algorithm Codesign for Efficient Gaze Tracking in Virtual Reality System

Haiyu Wang, Wenxuan Liu, Sai Qian Zhang
Tandon School of Engineering, New York University, NY, USA

***Abstract*—Real-time image rendering in VR systems demands high computational resources due to stringent visual quality requirements. Gaze-contingent foveated rendering offers a promising solution by dynamically assigning higher resolution only to the user's focus area, thus reducing overall processing load. However, many existing gaze estimation approaches suffer from notable errors and high execution costs. In this paper, we propose a co-design strategy that simultaneously refines the gaze estimation algorithm and its supporting hardware module, leading to improved accuracy and enhanced efficiency.**

## I. Introduction

In virtual reality, the process of generating high-fidelity images is fundamental for the realism and immersion of the virtual environment. Yet, the challenge lies in rendering high-resolution content with low latency on devices with limited resources. Gaze-contingent foveated rendering attempts to resolve this by focusing detailed rendering on regions where the user is looking, while reducing detail in the peripheral areas as shown in Figure 1. Despite its potential, current gaze estimation methods often exhibit significant errors and incur high computational costs. Our work addresses these challenges by developing an integrated solution that optimizes both the neural network for gaze estimation and the dedicated hardware module, thereby alleviating the load on the primary GPU and reducing the high computational cost.

## II. Methodology

Conventional gaze tracking techniques typically suffer from a long-tail error distribution, which limits their performance in foveated rendering scenarios. To address this, we introduce a novel vision-transformer-based network architecture that achieves lower tracking errors. Our method draws inspiration from appearance-based techniques that directly map eye images to gaze direction to avoid explicit eye segmentation and geometric fitting, resulting in a more robust and efficient gaze estimation pipeline. Additionally, by incorporating weight quantization, we significantly reduce computational demands without sacrificing precision. Complementing the algorithm, we design a specialized module integrated into the VR headset's system-on-chip (SoC) that handles gaze estimation, boosting overall efficiency.

## III. Evaluation

**Algorithm Evaluation:** We assess the gaze tracking performance of our network using the OpenEDS2020 dataset [1], and benchmark it against several state-of-the-art methods [2]–[5].The experimental results indicate that both the full and quantized versions of our model achieve up to 20 degrees lower average errors as well as reduced 95th-percentile errors, compared to baseline approaches on the test dataset.

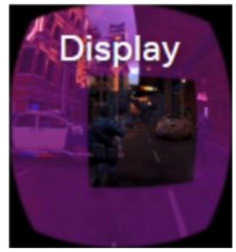


Fig. 1: Foveated rendering in VR HMD, displayed with different levels of resolution.

| Method | NE |
|---|---|
| Ours | 1× |
| ResNet | 1.75× |
| EdGaze | 2.36× |
| IncResNet | 4.49× |
| DeepVOG | 8.21× |

TABLE I: Normalized energy consumption (NE) of different gaze tracking methods on OpenEDS 2020.

**System Evaluation:** We assess our plug-in module and design a consistent architectures for other methods to ensure a fair comparison. By implementing gaze tracking energy analysis on the modules as shown in Table I, we observe that ours demonstrates significant reduction on energy consumption compared to other gaze tracking methods.

## IV. Conclusion and Future Works

Through efficient hardware and algorithm co-design, our method achieves significant improvements in both gaze tracking accuracy and energy efficiency. Future work will focus on integrating our method into VR headsets and seamlessly combining it with the render pipeline to enhance the end-to-end performance of gaze tracking and foveated rendering. Additionally, we aim to incorporate support for gaze reuse and saccade-aware mechanisms, enabling even more efficient gaze processing and rendering, while further reducing computational overhead in real-time VR applications.

## References

[1] C. Palmero, A. Sharma, K. Behrendt, K. Krishnakumar, O. V. Komogortsev, and S. S. Talathi, "Openeds2020: Open eyes dataset," 2020.
[2] Y.-H. Yiu, M. Aboulatta, T. Raiser, and L. Ophey, "Deepvog: Open-source pupil segmentation and gaze estimation in neuroscience using deep learning," *Journal of Neuroscience Methods*, vol. 324, p. 108307, 2019.
[3] Y. Feng, N. Goulding-Hotta, A. Khan, H. Reyserhove, and Y. Zhu, "Real-time gaze tracking with event-driven eye segmentation," in *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*.
[4] P. L. Mazzeo, D. D'Amico, P. Spagnolo, and C. Distante, "Deep learning based eye gaze estimation and prediction," in *2021 6th International Conference on Smart and Sustainable Technologies (SpliTech)*, 2021.
[5] R. Athavale, L. S. Motati, and R. Kalahasty, "One eye is all you need: Lightweight ensembles for gaze estimation with single encoders," 2022.