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

***Abstract*—Rendering images in AR/VR environments incurs significant costs due to the need to meet users' expectations for high-quality visuals, a challenge amplified in real-time applications. Gaze-tracked foveated rendering (TFR) offers great potential by dynamically adjusting rendering resolution based on human gaze, enabling substantial cost savings. However, existing AI-based gaze tracking solutions suffer from high tracking errors and high execution cost. This work addresses these challenges by co-optimizing AI algorithms with the underlying hardware to achieve efficient gaze tracking and superior rendering cost savings.**

I. Introduction

Image rendering is arguably the most essential component of VR systems, as it directly influences the realism and immersion of the virtual environment. However, rendering high-resolution frames on standalone VR devices often introduces considerable latency. Gaze-tracked foveated rendering (TFR) addresses this challenge by detecting gaze location using neural network in real-time, focusing computational resources on the foveal region while reducing detail in peripheral areas. However, current gaze tracking methods fail to provide high accuracy and lack adequate hardware support, leading to a high TFR latency and computational cost. In this paper, we propose POLO, a technique that co-optimizes AI algorithms and the underlying hardware platform for efficient image rendering, aiming to reduce the high computational costs of TFR.

II. Methodology

Current gaze tracking methods often exhibit a long-tail error distribution, thereby limiting their effectiveness and efficiency in foveated rendering applications. To address these challenges, we propose POLONet, a vision transformer-based neural network architecture that achieves low gaze tracking error, optimizing performance for foveated rendering systems. To further reduce POLONet's computational cost, we employ weight pruning and quantization without compromising gaze tracking accuracy. To further reduce the POLONet implementation cost, we propose the POLO accelerator as a dedicated module within the SoC of VR HMDs, consisting of three primary components: Image Pre-processing Unit (IPU), Computational engine, and the memory subsystem. It executes POLONet, offloading gaze processing from the resource-constraint edge GPU to reduce its workload and improve efficiency.

III. Evaluation

POLONet Evaluation: We assess the gaze tracking performance of POLONet using the OpenEDS2020 dataset [1], comparing it against other baseline methods [2]–[5]. As illustrated in Figure 1, both the pruned and original versions of POLONet demonstrate superior performance, achieving lower average error and 95th-percentile error on the test dataset.

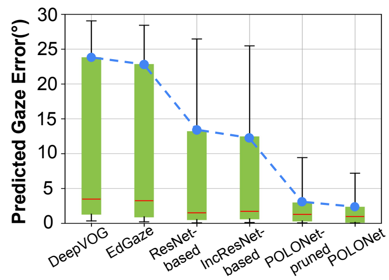

Fig. 1: Gaze tracking error distribution of different methods on OpenEDS2020 dataset.

Method	NE
POLO	1×
ResNet	1.75×
EdGaze	2.36×
IncResNet	4.49×
DeepVOG	8.21×

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

POLO Accelerator Evaluation: We evaluate the POLO accelerator and designe consistent architectures for other methods to ensure a fair comparison. By implementing gaze tracking energy analysis on the accelerators as shown in Table I, we observe that POLONet demonstrate significant reduction on energy consumption compared to other gaze tracking methods.

IV. Conclusion and Future Works

Through efficient hardware and algorithm co-design, our POLO achieves significant improvements in both gaze tracking accuracy and energy efficiency. Future work will focus on integrating POLO 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.