Mohammad Erfan Sadeghi, Arash Fayyazi, Suhas Somashekar, Massoud Pedram

This research paper presents CHOSEN, a novel software-hardware co-design framework for optimizing Vision Transformer (ViT) inference on Field-Programmable Gate Arrays (FPGAs). The authors address the significant challenges posed by ViTs' computational and memory demands, particularly within the context of FPGA resource limitations. This summary will dissect the paper's contributions, methodology, findings, and implications across various academic and practical domains.

The core contribution of this paper lies in its holistic approach to accelerating ViT inference on FPGAs. Existing methods often focus on isolated optimizations, such as quantization or model pruning. CHOSEN, in contrast, integrates multiple optimization strategies within a unified software-hardware co-design framework, resulting in a synergistic effect that significantly surpasses the performance of individual techniques. The key contributions can be summarized as follows:

- **Multi-Kernel Accelerator Design:** [Detailed description of Multi-Kernel Accelerator Design]

- **Approximation of Non-linear Functions:** [Detailed description of Approximation of Non-linear Functions]

- **Efficient Compiler for Design Space Exploration:** [Detailed description of Efficient Compiler for Design Space Exploration]

The combined effect of these three contributions results in a significant performance boost compared to state-of-the-art ViT accelerators. This executive synthesis highlights the paper's innovative approach to tackling the challenges of ViT deployment on FPGAs, moving beyond individual optimizations to a comprehensive and synergistic framework. The automated nature of the CHOSEN compiler further enhances its practical applicability and scalability.

[Detailed description of Methodological Architecture]

[Detailed description of Critical Findings Hierarchy]

[Detailed description of Theoretical Framework Integration]

[Detailed description of Limitations & Epistemological Boundaries]

[Detailed description of Future Research Trajectories]

[Detailed description of Interdisciplinary Implications]

[Detailed description of Conclusive Assessment]