پژوهشگاه دانش‌های بنیادی
پژوهشکده علوم کامپیوتر

سخنرانی علمی

Reconfigurable Accelerator for Efficient Implementation of CNNs

پریا دربانی

پژوهشکده علوم کامپیوتر، پژوهشگاه دانش‌های بنیادی (IPM)

Abstract

Convolutional Neural Network (CNN) is used in many real-world applications due to its high accuracy. The rapid expansion of modern applications and the growing emphasis on performance underscores the significance of efficient CNN implementation. While CNNs entail substantial communication and computational demands, accelerators face limitations due to constrained hardware resources. On the other hand, variations in layer shapes and sizes can result in suboptimal resource utilization. The increasing scale and complexity of CNN models exacerbate this problem.

This talk begins with an overview of the CNNs and accelerators, followed by a discussion about the under-utilization problem that prevents the accelerator from reaching maximum performance. We then briefly review the available methods to improve resource utilization. We discuss that the performance of CNN models depends on the hardware's ability to adapt to different shapes of different layers. Hence, the talk continues with the proposal of a low-cost reconfigurable architecture that can efficiently execute a wide range of CNNs. We present the results of our analysis and simulations, which show the improvement in speedup, throughput, and power efficiency without compromising accuracy. The proposed architecture also reduces the on-chip memory access rate. The final section of the talk outlines our new approach and feature works to improve CNN performance.

Biography

Paria Darbani obtained her Ph.D. in Computer Engineering from the University of Science and Technology in 2023. Her research interests include neural network accelerators and the area of approximate computing. Currently, she is actively seeking a postdoctoral research position within the Department of Computer Science at the Institute for Research in Fundamental Sciences (IPM).

زمان : چهارشنبه ۱۴۰۲/۰۷/۲۶ – ساعت ۱۵:۰۰

ارائه به صورت مجازی انجام خواهد شد.

https://vmeeting.ipm.ir/b/com-hrj-f9n

***** شرکت برای عموم علاقه‌مندان آزاد است*****