

Nikkei Electronics Products' Trends (November 2024)

EdgeCortex Develops Semiconductors for Edge Inference Low Cost and High Efficiency, 60 TOPS at 8W



EdgeCortex has developed an AI (Artificial Intelligence) accelerator that can achieve 60 TOPS (60 Trillion Operations per Second) of computational performance with just 8W of power consumption. Compared to NVIDIA's GPUs (graphics processing semiconductors), it is specialized for edge inference, offering high power efficiency at a low cost.

One reason for its high power efficiency is the system's ability to reconfigure the data path for each computation. This enables efficient processing of large-scale language models (LLMs) with high sparsity. The previous generation product was optimized for 8-bit integer operations (INT8), but the new product also supports 16-bit floating-point operations (BF16).

By switching between INT8 and BF16 during processing, it is claimed that the system can shorten computation time while improving accuracy. For example, when performing facial recognition with a surveillance camera, the system can switch to high-precision BF16.

It is set to be sold starting at \$299 USD (approximately 40,000 yen, based on an exchange rate of 1 USD = 157 yen) and is scheduled for shipment by the end of 2024.

- Translation prepared by EdgeCortex.
- Nikkei Electronics Products' Trends (2024.11), full original Japanese post: https://bizboard.nikkeibp.co.jp/houjin/cgi-bin/nsearch/md_pdf.pl/0000507498.pdf?NEWS_ID=0000507498&CONTENTS=1&bt=NE&SYSTEM_ID=HO&BZB_DATE_TOKEN=2c8c1de49fa6335b595b98736245b337ae6c9ab79ae590cf83dd174a1b8c749dc96ede0a9cb8104bd5bd8bc3bf953669
- Copyrights and other intellectual property rights to articles, photographs, charts, headlines, and other information (hereinafter referred to as "Information") provided through the Service belongs to the providers of such Information.
- Unauthorized reproduction of Information provided by this service is prohibited.
- The service may not be used by any other third party other than the subscriber, regardless of the method, with or without compensation.
- Copyright © Nikkei Electronics.. All Rights Reserved.