UNIVERSIDAD COMPLUTENSE MADRID

Facultad de Informática

ANUNCIO DE CONFERENCIA

POSGRADO

Una Visión Computacional de las Redes Neuronales

Enrique S. Quintana Ortí
UPV

Facultad de Informática
Sala de Grados - lunes 4 de diciembre de 2023 - 15:00
Entrada libre hasta completar el aforo

Resumen:

La multiplicación de matrices (GEMM) es un núcleo computacional clave, omnipresente en numerosos ámbitos. Por un lado, muchas aplicaciones científicas y por otro lado, las redes neuronales convolucionales para tareas de procesamiento de señales y visión por computador, así como los modelos utilizados en herramientas de aprendizaje profundo como ChatGPT. En esta charla, en primer lugar expondremos los problemas de las instancias actuales de GEMM en bibliotecas para arquitecturas multinúcleo convencionales: rendimiento subóptimo y falta de soporte para tipos de datos orientados a aprendizaje profundo. Partiendo de ese punto, demostraremos cómo pueden superarse estos problemas mediante herramientas para la generación automática de código, junto con un modelo analítico de la configuración de la jerarquía de caché del procesador. Además, ilustraremos que este enfoque se puede aplicar también a arquitecturas más "exóticas", desde aceleradores vectoriales de gama alta y el diseño AIE de Xilinx hasta dispositivos de bajo consumo como procesadores RISC-V y microcontroladores basados en ARM (Arduino).

Sobre Enrique S. Quintana Ortí:

Enrique S. Quintana-Orti se licenció y doctoró en Informática por la Universidad Politécnica de Valencia (UPV), España, en 1992 y 1996, respectivamente. Tras más de 20 años en la Universidad Jaime I de Castellón, España, regresó a la UPV en 2019, donde actualmente es catedrático de Arquitectura de Computadores. Por su investigación, Enrique ha recibido el NVIDIA 2008 Professor Partnership Award, dos premios de la Agencia Espacial Nacional de Estados Unidos (NASA) y el premio a la trayectoria investigadora por la conferencia Euro-Par en 2023. Ha publicado más de 400 artículos en revistas y conferencias internacionales. Actualmente participa en los proyectos de la UE APROPOS (computación aproximada), RED-SEA (redes informáticas a exaescala), eFLOWS4HPC (flujos de trabajo para HPC e IA), Nimble AI (chip neuromórfico para detección y procesamiento) y Metamorpha (micromecanizado a medida con rayos láser). Sus intereses de investigación incluyen la programación paralela, el consumo de energía, la computación con precisión adaptativa, el aprendizaje profundo y el álgebra lineal, así como arquitecturas avanzadas y aceleradores de hardware.