

Tutorial 13: Efficient Image Processing with Deep Neural Networks

This tutorial describes methods to enable efficient processing for deep neural networks (DNNs), which is the cornerstone of many state-of-the-art image processing and computer vision algorithms. While DNNs delivers best-in-class accuracy and quality of results, it comes at the cost of high computational complexity. Accordingly, designing efficient algorithms for deep neural networks is an important step towards enabling the wide deployment of DNNs in image processing and computer vision systems (e.g., autonomous vehicles, drones, robots, smartphones, wearables, Internet of Things, etc).

In this tutorial, we will provide a brief overview of DNNs, the various hardware platforms that support DNNs including CPU, GPU, FPGA and ASICs, and highlight important benchmarking/comparison metrics and design considerations. We will then describe recent techniques that reduce the computation cost of DNNs from both the hardware architecture and network algorithm perspective. We will also discuss how these techniques can be applied to different image processing and computer vision tasks.

Speaker:



Vivienne Sze
Massachusetts Institute of Technology, USA