

HKU-TCL Joint Research Centre for AI Workshop

Speaker:

Dr. Ping LUO

Assistant Professor

Department of Computer Science, Faculty of Engineering, HKU

Title:

Convolution is Not All You Need: New Methods for Visual Recognition

Abstract:

Although convolutional neural networks (CNNs) achieve great successes in many tasks of computer vision, this talk presents several methods that are useful for many computer vision tasks without convolutions. Firstly, we present a simple backbone network, Pyramid Vision Transformer (PVT), which overcomes the difficulties of porting Transformer to various dense prediction tasks. It inherits the advantages from both CNNs and Transformers, making it a unified backbone in various vision tasks without convolutions by simply replacing CNN backbones. We could build end-to-end object detection and segmentation systems comprehensively without using convolutions. Secondly, we present Sparse R-CNN, a sparse method beyond Transformer-based models for object detection in images. It avoids all hand-crafted efforts related to object candidates design, many-to-one label assignment, and non-maximum suppression as post-processing.