Embedded Artificial Neural Networks (Embedded-ANN) and YOLO-Edge for Embedded Systems

Halil Hüseyin Çalışkan
Bursa Technical University, Faculty of Engineering and Natural Sciences, Department of Computer Engineering, Bursa, Türkiye

Talha Koruk
Bursa Technical University, Faculty of Engineering and Natural Sciences, Department of Computer Engineering, Bursa, Türkiye

Abstract:

In embedded systems, image classification with artificial neural networks and object detection with YOLO cause both excessive memory usage and low frame rate per second. In this study, in order to reduce these problems encountered in embedded systems, artificial neural networks were rewritten only with NumPy without using any deep learning library. The results of this study are that this artificial neural network model developed with NumPy specifically for embedded systems is approximately 265 times faster than artificial neural network models created with TensorFlow having the same neural network architecture and consumes 6 times less memory. In addition, the fact that these developed artificial neural networks operate at high speed using only the processor without using a graphics card in embedded systems shows the feature of this neural network model to work independently of hardware. Moreover, YOLOv8's architecture has been modified specifically for embedded systems and named as YOLO-Edge. According to results, YOLO-Edge is 4 times faster than YOLOv8m on Nvidia Jetson Xavier Nx and 8 times faster on Rockchip 3588 (NPU).

Keywords:

Artificial Neural Networks, Embedded Systems, Image Classification, YOLO.