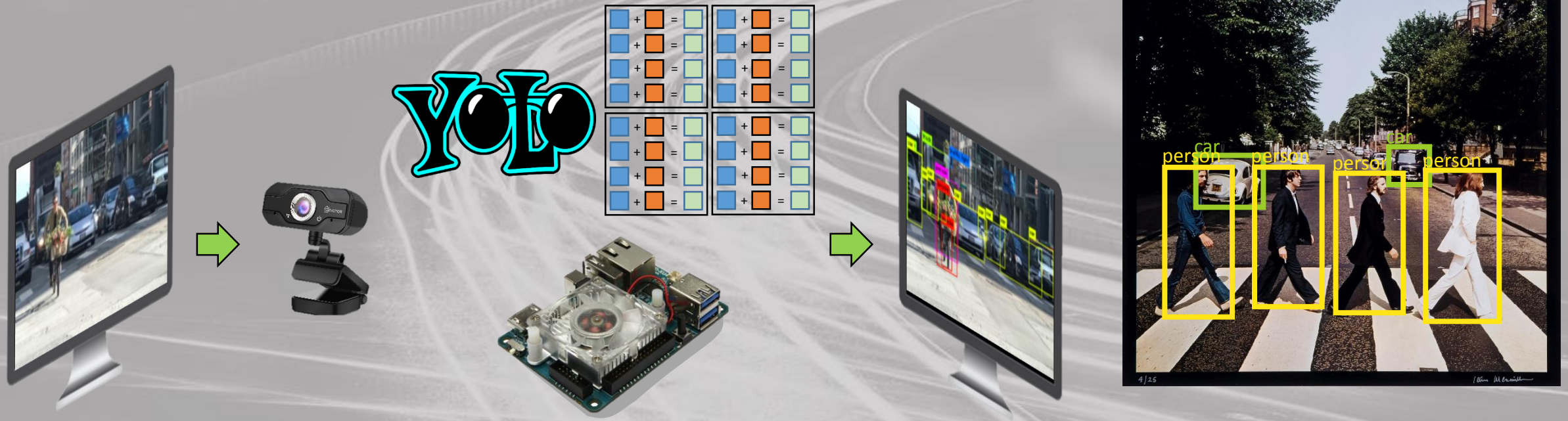
Towards Real-time Object Detection on Mobile Devices



Byeongkyo Cheong*, Donghee Ha, Jinse Kwon, Hyungshin kim

Department of Computer Science and Engineering

Chungnam National University

Daejeon, South Korea

RTSS 2018

Nashville, TN, USA

Problem

High performance GPU

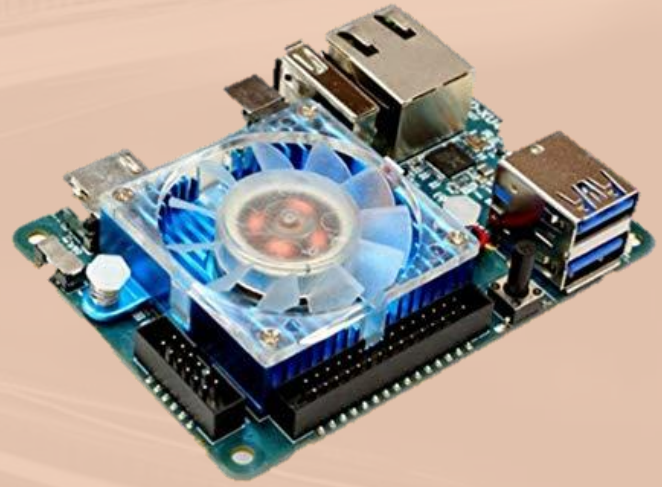
EXPENSIVE

HIGH power consumption

FAST

VS

Mobile device



CHEAP

LOW power consumption

SLOW

We challenge to run CNN on mobile devices in REAL real-time

What We Did

Current State-of-the-art

Darknet / TensorFlow / Caffe Framework

Matrix Multiplication Library for CNN (CUDA)

Our Work

Without Framework

For all kinds of Mobile GPUs

Matrix Multiplication Library for CNN (OpenCL)

Our Library

CLBlast VS Non-Square Matrix Optimization

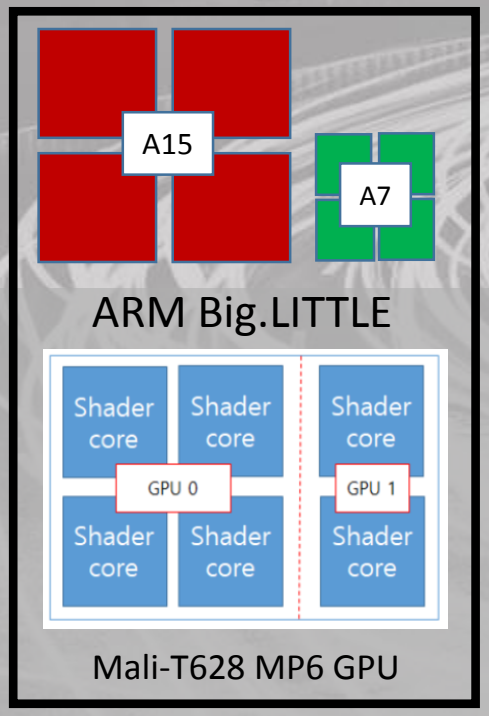
1.3 times faster than CLBlast

At Our Demo

Input : VOC Images

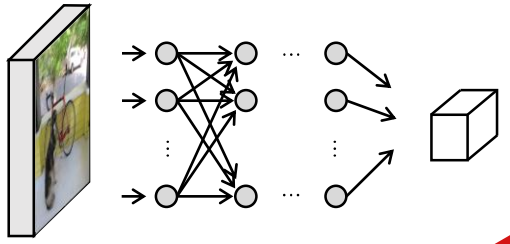
Input : Camera

Mobile Device



Exynos 5422

CNN



Real Detection Challenge !

Real Detection

Result