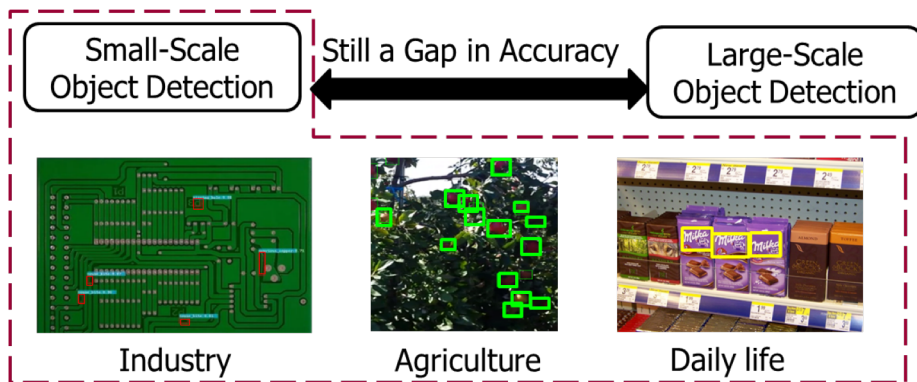# Time-dependent Composite Frame-rate based CNN using Hybrid FPGA/GPU Configuration for Real-time Small Object Detection

黄冬梅　池永研究室　修士課程修了

■ Background

Small-Scale Object Detection ⟷ Still a Gap in Accuracy ⟷ Large-Scale Object Detection

Industry　Agriculture　Daily life

■ Problem statement

Problems

1. Complex structure & Huge amount of calculation for improving accuracy

2. Low parallelism of the whole progress

3. The binary network cannot provide sufficient accuracy for the estimation of size

Solutions

P1: Change the training method to improve accuracy

P2: Use the correlation of high frame rate video to improve the parallelism of processing

P3: Use high-frequency feature fusion with grouped parallel to improve accuracy

Composite frame-rate based real-time small object detection is necessary to meet the requirements of rapid and accurate response in the FA and robotics fields.
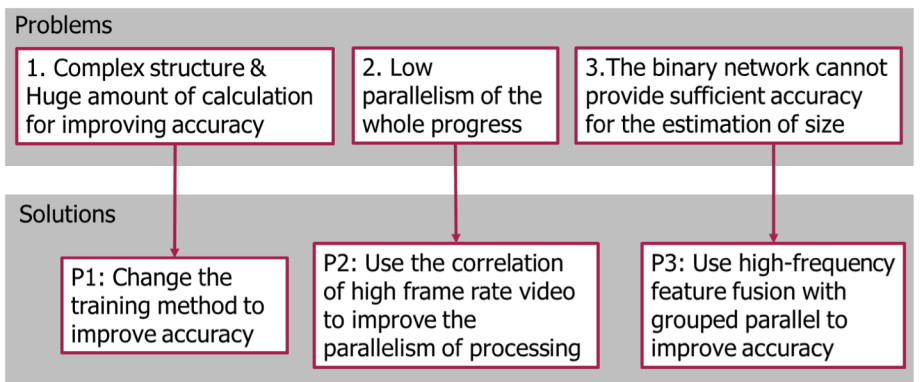
■ Proposed method

The proposed hybrid FPGA/GPU system

Feature extraction network on FPGA

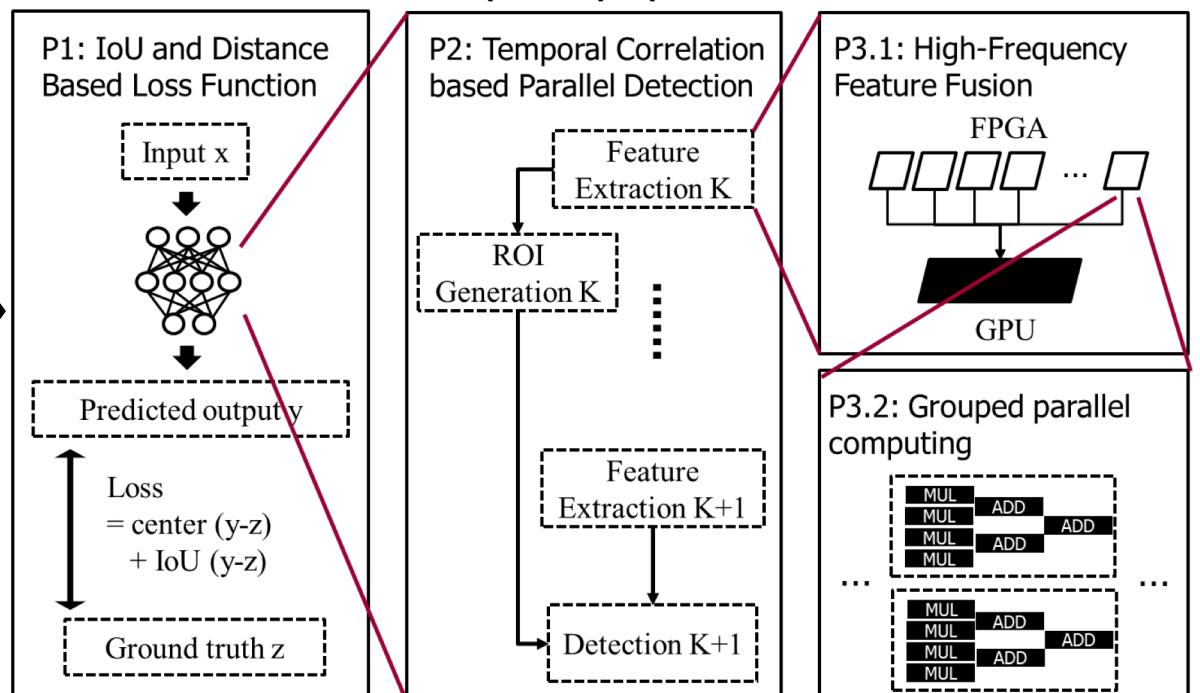High speed camera

Detection network on GPU

Class 1　Class 2
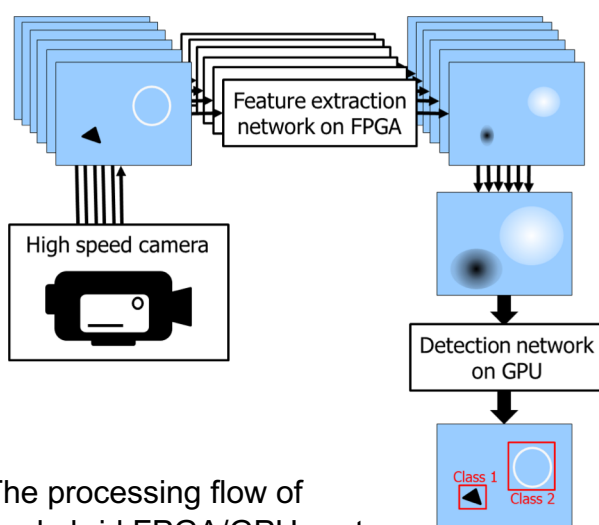
The processing flow of the hybrid FPGA/GPU system:
1. Obtain feature maps through the FPGA
2. Perform fusion to get high quality feature map
3. Detect the fused feature map on the GPU

Specific proposals

P1: IoU and Distance Based Loss Function

Input x

Predicted output y

$$Loss = center\ (y\text{-}z) + IoU\ (y\text{-}z)$$

Ground truth z

P2: Temporal Correlation based Parallel Detection

Feature Extraction K

ROI Generation K

Feature Extraction K+1

Detection K+1

P3.1: High-Frequency Feature Fusion

FPGA

GPU

P3.2: Grouped parallel computing

MUL ADD ADD

MUL ADD ADD

■ Experiments

Detection accuracy

	Faster-RCNN	Proposal A	Proposal A Proposal B	Proposal A Proposal B Compression	Proposal A Proposal B Compression Proposal C
mAP (Slowly moving)	0.467	0.587	0.532	0.454	0.506
mAP (Suddenly shaking)	0.438	0.541	0.518	0.489	0.515
mAP (Zoom up)	0.476	0.519	0.494	0.431	0.487
mAP (Illumination change)	0.458	0.532	0.502	0.286	0.304
mAP (Average)	0.462	0.544	0.511	0.415	0.453

Hardware performance

Feature extraction on FPGA — 316 fps

Fusion

Detection on GPU
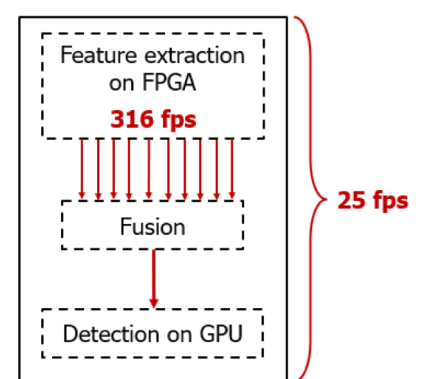
25 fps

■ Conclusion

The proposed method achieves the accuracy of 45.3 % mAP. Regarding the heterogeneous configuration of FPGA/GPU, the processing speed of the FPGA reaches 316 fps and the overall detection speed of the whole system reaches 25 fps.