YOLOv5n-Face: Efficient Lightweight Face Detection via Weight Binarization

2nd Seongsoo Lee
School of Electronic Engineering,
Department of Intelligent
Semiconductor
Soongsil University
Seoul, Republic of Korea
sslee@ssu.ac.kr

1st Eunsu Kim
School of Electronic Engineering
Soongsil University
Seoul, Republic of Korea
sgn09151@naver.com

Deep learning-based object detection networks such as YOLOv5n provide strong accuracy but face challenges in deployment on resource-constrained devices due to model size and computation cost. To address this problem, we apply binary quantization and replace multiplications with efficient XOR-based operations. Using Straight-Through Estimator (STE) and scaling compensation, we design a binarized YOLOv5n-Face model and implement it on an FPGA platform, achieving real-time inference capability with competitive accuracy and detection speed.

A. Binarized YOLOv5n-Face Architecture

The proposed model builds upon YOLOv5n, the most lightweight variant of the YOLOv5 series with only 1.726M parameters. To further optimize computational efficiency, binary quantization was applied, constraining weights to ±1 and replacing multiplication with simple XOR or sign-flip operations. This reduces memory requirements by approximately 1/32 compared to the floating-point model and eliminates over 90% of convolutional multiplications, making the architecture well-suited for FPGA-based real-time applications.

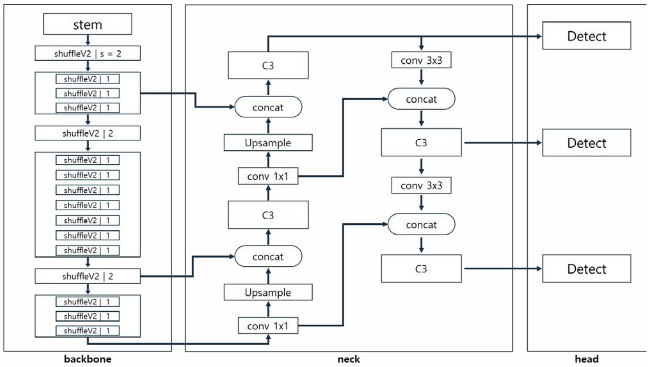

Fig. 1. YOLOv5n architecture

B. Performance Compensation and Evaluation

While pure binarization inevitably causes some accuracy degradation, we introduce a compensation mechanism by applying a scaling factor (μ) for each convolutional layer. Combined with the Straight-Through Estimator (STE), this approach preserves training stability and enhances representational capacity. Experimental results show that the μ-compensated binarized YOLOv5n-Face achieves competitive detection accuracy across different difficulty levels (Easy, Medium, Hard) while maintaining real-time inference capability on FPGA hardware.

TABLE I. AP UNDER DIFFERENT WEIGHT REPRESENTATION SCHEMES

Difficulty Level	***Weight Representation***		
	32 Float	***Binary + μ***	***Binary***
Easy	0.9172	0.7389	0.3065
Medium	0.8739	0.6728	0.3356
Hard	0.6761	0.4135	0.1962

ACKNOWLEDGMENT

This work was supported by Korea Evaluation Institute of Industrial Technology (KEIT) funded by the Korea Government (MOTIE). (RS-2022-00155731)

REFERENCES

[1] D. T. Nguyen, T. N. Nguyen, H. Kim and H. -J. Lee, "A High-Throughput and Power-Efficient FPGA Implementation of YOLO CNN for Object Detection," in IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol. 27, no. 8, pp. 1861-1873, Aug. 2019, doi: 10.1109/TVLSI.2019.2905242.

[2] D. Qi, W. Tan, Q. Yao, and J. Liu, “YOLO5Face: Why Reinventing a Face Detector,” in *Computer Vision ‑ ECCV 2022 Workshops*, Lecture Notes in Computer Science, vol. 13629, pp. 228‑244, 2022, doi:10.1007/978-3-031-25072-9_15.

[3] Z. Yan, B. Zhang, and D. Wang, “An FPGA-Based YOLOv5 Accelerator for Real-Time Industrial Vision Applications,” Micromachines, vol. 15, no. 9, p. 1164, Sep. 2024. doi: 10.3390/mi15091164.