# 저전력 장치를 위한 자원 효율적 객체 검출기

## (Resource-Efficient Object Detector for Low-Power Devices)

악세이 쿠마 샤마 [1], 김경기 [1]

(Akshay Kumar Sharma[1], Kyung Ki Kim[1])

## 요 약

본 논문은 전통적인 자원 집약적인 컴퓨터 비전 모델의 한계를 해결하기 위해 저전력 엣지 장치에 최적화된 새로운 경량 객체 검출 모델을 제안합니다. 제안된 검출기는 Single Shot Detector (SSD)에 기반하여 소형이면서도 견고한 네트워크를 설계하였고, 작은 객체를 효율적으로 감지하는 데 있어 효율성을 크게 향상시키도록 모델을 구성하였다. 이 모델은 주로 두 가지 구성요소로 구성되어 있습니다: Depthwise 와 Pointwise Convolution 레이어를 사용하여 효율적인 특징 추출을 위한 Light_Block, 그리고 작은 객체의 향상된 감지를 위한 Enhancer_Block 으로 나누었다. 우리의 모델은 300x480 의 이미지 크기를 가진 Udacity 주석이 달린 데이터셋에서 처음부터 훈련되었으며, 사전 훈련된 분류 가중치의 필요성을 제거하였다. 약 0.43M 의 파라미터로 5.5MB 만의 무게를 가진 우리의 검출기는 평균 정밀도 (mAP) 27.7%와 140 FPS 의 처리 속도를 달성하여, 정밀도와 효율성 모두에서 기존 모델을 능가하였다. 따라서, 본 논문은 추론의 정확성을 손상시키지 않으면서 엣지 장치를 위한 객체 검출에서의 효과적인 경량화를 보여주고 있다.

## ABSTRACT

This paper presents a novel lightweight object detection model tailored for low-powered edge devices, addressing the limitations of traditional resource-intensive computer vision models. Our proposed detector, inspired by the Single Shot Detector (SSD), employs a compact yet robust network design. Crucially, it integrates an 'enhancer block' that significantly boosts its efficiency in detecting smaller objects. The model comprises two primary components: the Light_Block for efficient feature extraction using Depth-wise and Pointwise Convolution layers, and the Enhancer_Block for enhanced detection of tiny objects. Trained from scratch on the Udacity Annotated Dataset with image dimensions of 300x480, our model eschews the need for pre-trained classification weights. Weighing only 5.5MB with approximately 0.43M parameters, our detector achieved a mean average precision (mAP) of 27.7% and processed at 140 FPS, outperforming conventional models in both precision and efficiency. This research underscores the potential of lightweight designs in advancing object detection for edge devices without compromising accuracy.

## I. Introduction

In recent years, the proliferation of edge devices, such as smart cameras, drones, and IoT sensors, has highlighted the need for efficient computing models that can operate under constrained resources. Traditional computer vision models, while powerful, often face challenges when deployed on these devices. Their high resource demands in terms of memory and computational power render them unsuitable for low-powered, resource-limited environments. As the demand for real-time, on-device processing grows, there is an increasing need to design models that can maintain a balance between efficiency and performance.

This paper delves into this challenge, presenting a solution in the form of a lightweight object detection model tailored specifically for edge devices. This model is not just a mere reduction in parameters but a thoughtful design that ensures both efficiency and accuracy. Drawing inspiration from the Single Shot Detector (SSD) [3], a popular object detection framework known for its speed and reliability, our proposed detector employs a compact yet robust network architecture to enhance feature extraction. Central to our design is the introduction of an 'enhancer block'. This novel component significantly boosts the model's efficiency, especially in detecting smaller objects, which are typically challenging for many existing models. By integrating this enhancer block, our detector not only achieves better performance metrics but also sets a new standard, distinguishing itself from conventional, resource-intensive models.

## II. Proposed Object Detector

The object detector integrates the Base Network and the detection head. This paper introduces a lightweight Base Network composed of two main blocks: the Light_Block and the Enhancer_Block. Prioritizing the creation of a compact detector, SSD7 is chosen as the detection head. We trained our object detector from scratch, bypassing the need for pretrained classification weights. Fig. 1 visualizes the entire model architecture: It initiates with a simple 5x5 convolution layer, succeeded by six Light_Block layers, with the final four layers designated for detection. The Enhancer_Block is strategically positioned before detection to augment the network's overall accuracy and detect minute objects

Object detector is a combination of the Base Network and the detection head. The lightweight Base Network is proposed in this paper that introduces 2 blocks namely, Light_Block, and an Enhancer_Block. As the main goal of this work is to make a lightweight detector, the proposed network uses SSD7 as the detection head. The proposed object detector is trained from scratch, without using pre-trained classification weights. Fig. 1 shows the entire model architecture. A simple convolution (5 x 5) is the initial layer, followed by six Light_Block layers, and the last four layers are used for detection. An Enhancer_Block is added before detection to improve the overall accuracy of the network as well as to detect small objects.

### 1. Light_Block

Aiming to extract pertinent features swiftly without excessive resource consumption, the Light_Block employs Pointwise and Depthwise Convolution layers. A typical convolution integrates two primary tasks: Filtering and Combining. Though standard convolutions are parameter-rich, separating these operations reduces parameter usage. This separation birthed the Depth-wise separable convolution, initially employed by MobileNet [4]. Within the Light_Block, a pointwise layer with 'n' filters first interprets the input channel, which is then followed by a depthwise convolution. Downsampling is achieved using the Max_Pool transition layer.
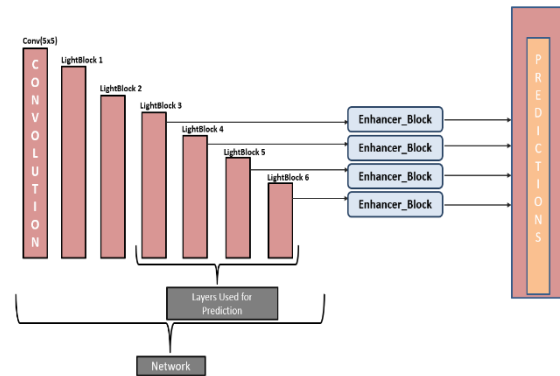


**Fig. 1 Architecture of the model**

### 2. Enhancer_Block

The Enhancer_Block addresses the challenge of detecting tiny objects. It projects crucial information extracted by prior Conv_Blocks using a pointwise convolution layer at varied scales. As depicted in Fig. 2, the input from the preceding layer bifurcates into two pathways. The primary path comprises a series of 1x1 Pointwise layers, succeeded by a 3x3 standard convolution and another pointwise layer. The alternative path employs only 1x1 pointwise layers. Empirical results confirmed enhanced precision with the Enhancer_Block, particularly noticeable in the detection of smaller objects during inference.

The foundational network processes 300x480 inputs, initially classifying via a 5x5 convolution layer. Subsequently, the input traverses six Light_Block layers, extracting pivotal features. Each layer incorporates a max pool layer for downsampling. The resulting output dimensions from the concluding four Light_Block layers are: (37x60), (18x30), (9x15), and (4x7). These outputs feed into the Enhancer_Block, fortifying the detection of smaller objects. Finally, the image dimensions are integrated into the SSD7 model. The training dataset utilized is the Udacity Annotated Dataset with image dimensions of 300x480.
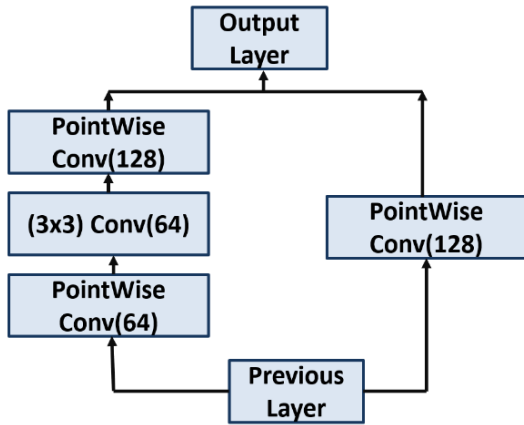
**Fig. 2 Enhancer_Block**

## III. Simulation Results

Our trained model, weighing just 5.5MB, encapsulates roughly 0.43M (or 433,200) parameters. Training was conducted with a batch size of 32 over 50 epochs on a Titan X GPU. Comparative analysis with two traditional methods, based on mAP and FPS, is presented in Table I.

**Table 1. mAP & FPS comparison with conventional models**

| Model | mAP | FPS |
|---|---|---|
| SSD7 [5] | 22.9% | ~180 |
| FPNSSD7 [6] | 26.9% | ~30.070 |
| Proposed | 27.7% | ~140 |

With an mAP of 27.7%, our proposed model outperforms both comparative object detectors using the same Udacity dataset, achieving 140 FPS. Fig. 3 offers a visual comparison, highlighting the detection capabilities for distant objects. While the conventional model [5] falters in Fig. 3(a), our proposed model succeeds in Fig. 3(b).



**Fig. 3 Detection result comparison with conventional model**

## V. Conclusion

The proposed object detector, trained from the ground up, leverages solely the object detection dataset for learning. This method circumvents superfluous computational demands, saving time both in initial classification model training and subsequent object detection. Boasting a mere 0.43M parameters and weighing 5.5MB, our model, trained on the Udacity Annotated Driving Dataset, achieves a mAP of 27.7%, surpassing conventional models. The lightweight design and superior results evidence its capability in efficiently detecting both small and distant objects.

## 감사의 글

## References

[1] X. Xu, X. Zhang, T. Zhang, "Lite-YOLOv5: A Lightweight Deep Learning Detector for On-Board Ship Detection in Large-Scene Sentinel-1 SAR Images", *Remote Sensing,* vol. 14, no. 4, pp. 1-27, Feb. 2022.

[2] L. Yang, G. Yuan, H. Wu, W. Qian, "An ultra-lightweight detector with high accuracy and speed for aerial images", *Math Biosci Eng.* vol. 20, no. 8, pp. 13947-13973, Jun. 2023.

[3] W.Liu, D.Anguelov, D. Erhan, C.Szegedy, S. Reed, C.Fu, A.C. Berg, "SSD: Single Shot MultiBox Detector," *European Conference on Computer Vision(ECCV),* pp. 21-37, Oct. 2015.

[4] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *CVPR,* April 2017.

[5][online]Available:https://github.com/pierluigiferrari/ssd_keras/blob/master/ssd7_training.ipynb [Accessed on Sept. 5, 2023].

[6] Y. Yamashige and M. Aono, "FPSSD7: Real-time Object Detection using 7 Layers of Convolution based on SSD," International Conference of Advanced Informatics: Concepts, Theory, and Applications (ICAICTA), pp. 1-6, Sept. 2019.

악세이 쿠마 샤마 (Akshay Kumar Sharma)

2015년 06월 : Bharat Institute of Technology 전자 공학과 졸업

2019년 2월 : 대구대학교 **전자 공학과 석사**

2023년 3월~현재 : 대구대학교 전자공학과 박사과정

<관심분야> 집적회로 설계, 반도체소자/공정, 시스템반도체 설계


김경기 (Kyung Ki Kim), 정회원

1997년 8월 : 영남대학교 전자 공학과 졸업

2008 년 1월: (미)Northeastern University, Computer Engineering 박사졸업

2010년 3월~현재 : 대구대학교 전자전기공학부 교수

<관심분야> 디지털시스템설계, 인공지능반도체설계, PIM 설계