Clear as day: low-power and lightweight object detection in challenging conditions

Jonas Bohn, Hanna Müller, Tommaso Polonelli*, Michele Magno
D-ITET, ETH Zürich, Zürich, Switzerland
* *topolonelli@ethz.ch*

The current trend in deep learning is pushing the development of pervasive "smart objects" in industry and everyday life, which quest for autonomy and smooth interaction with numerous environments [1]. Sensing the surroundings with vision-based systems is the most common approach, with an active momentum to bring intelligence also on low-power and battery-supplied devices, often characterized by a limited onboard computing performance [2]. A core component of perceiving the environment lies in detecting and identifying objects. Research in computer vision was initially driven by the development of deep convolutional networks (CNN) [1], and researchers are constantly working on new architectures to improve peak accuracy and robustness further [3]. However, these networks are becoming computationally demanding and power-hungry, making them unusable in real-time applications on resource-constrained devices (< 1 MB memory) [1]. Moreover, current trends in computer vision mainly focus on high-resolution RGB cameras, which exhibit limitations when illuminations and line of sight (LOS) visibility are not ideal [4]. Therefore, the demand for an object detection network that quickly and reliably performs under challenging conditions, such as poor illumination or occlusion, still needs to be addressed [7]. In this direction, few object detection systems fuse data streams to improve performance and robustness. In particular, laser-based sensors such as Light Detection and Ranging (LiDAR) [8] or Time-of-Flight (ToF) [9] are combined with a monocular camera to complement each other's strengths and overcome limitations. Laser-based sensors can accurately measure depth information, helping object segmentation within the scene. In addition, they are less affected by environmental factors such as low light, fog, or heavy rain, and can therefore provide reliable data in challenging situations [8][9].

This work proposes an object detection system based on a lightweight CNN and a sensor fusion approach, combining Infra-Red (IR) and ToF images using the flexx2 3D camera developed by pmd technologies. The flexx2 features a single sensor element IRS2381C from Infineon with a resolution of 224x172 and a framerate of up to 60 fps, which directly generates the two data sources. The power consumption of 300 mW and the relatively low resolution make it suitable for low-power and resource-constrained platforms. Thus, the flexx2 is combined with an optimized version of FOMO [6] object detection, in Figure 2, optimized for multimodal inputs, fusing IR and depth information. The original FOMO model was modified to support object detection (three classes: apple, cup, person, see Figure 1) in harsh conditions, such as complete darkness, fog, and with partial/total obstructions on input layers. Moreover, the model was quantized in int8 and optimized to run on a low-power RISC-V MCU, GAP9 from Greenwaves, resulting in 18.2 k parameters and a memory footprint of 57 kB, as depicted in Figure 2. The entire object detection pipeline runs on GAP9 in 5 ms (18 MOPS) with an average power consumption of 60 mW (0.3 mJ per processed frame) and 86% F1-score (calculated considering unideal visibility conditions). Compared to a conventional implementation based on YOLO and an RGB camera [5], our solution is 15x smaller (840 k parameters) and 7x faster (34 ms) than [5], enabling reliable object detection in any visibility conditions.

This work demonstrates the feasibility and benefits of using FOMO to perform sensor fusion with IR and depth data targeting resource-constrained and low-power devices. To support our results, a custom object detection dataset is created containing challenging visibility situations in addition to standard benchmarking datasets, as shown in Figure 1. Moreover, we also demonstrate how the resulting multimodal object detection pipeline can be quantized and ported to a constrained device without significant loss in performance, Figure 1 and Figure 2.

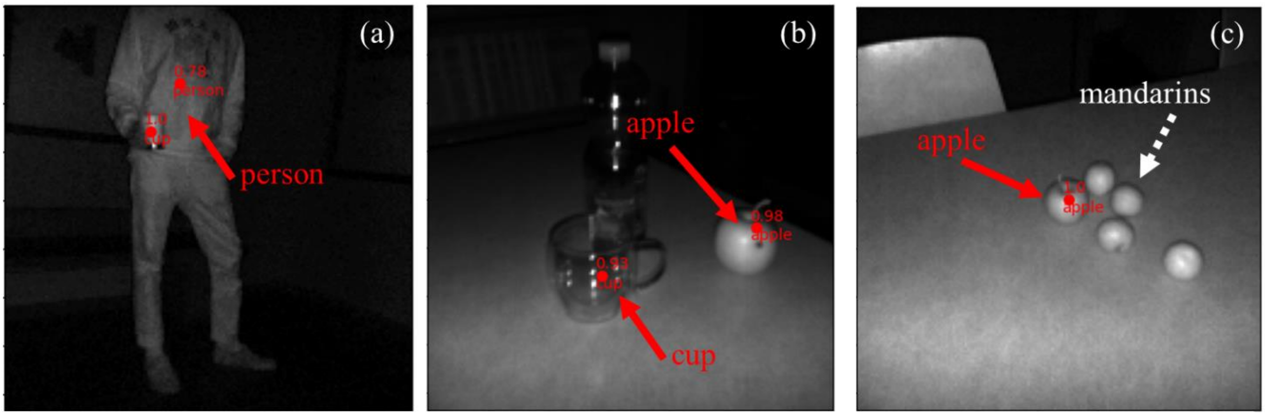

Figure 1 - An example of the object detection pipeline featuring three classes: apple, cup, person. All the images are captured in total darkness. Note how in (c) the model differentiates between similar objects, mandarins and apples.

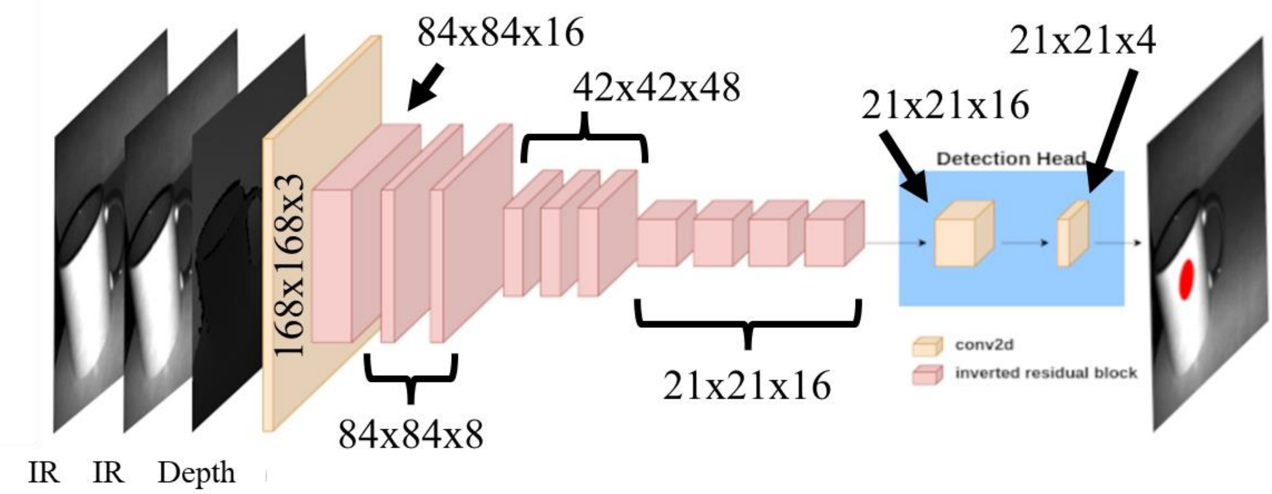

Figure 2 - Object detection model based on FOMO, optimized to support sensor fusion (IR+ToF) and to run onboard a low-power MCU. it is quantized in int8 with 18.2 k parameters.

REFERENCES

[1] Zou, Z., Chen, K., Shi, Z., Guo, Y., & Ye, J. (2023). Object detection in 20 years: A survey. Proceedings of the IEEE.
[2] Dharani, A., Kumar, S. A., & Patil, P. N. (2023). Object Detection at Edge Using TinyML Models. SN Computer Science, 5(1), 11.
[3] Han, H., & Siebert, J. (2022, February). TinyML: A systematic review and synthesis of existing research. In 2022 International Conference on Artificial Intelligence in Information and Communication (ICAIIC) (pp. 269-274). IEEE.
[4] Brunelli, D., Polonelli, T., & Benini, L. (2020, October). Ultra-low energy pest detection for smart agriculture. In 2020 IEEE SENSORS (pp. 1-4). IEEE.
[5] Moosmann, J., Bonazzi, P., Li, Y., Bian, S., Mayer, P., Benini, L., & Magno, M. (2023). Ultra-efficient on-device object detection on ai-integrated smart glasses with tinyissimoyolo. arXiv preprint arXiv:2311.01057.
[6] Boyle, L., Baumann, N., Heo, S., & Magno, M. (2023, October). Enhancing Lightweight Neural Networks for Small Object Detection in IoT Applications. In 2023 IEEE SENSORS (pp. 01-04). IEEE.
[7] Moosmann, J., Giordano, M., Vogt, C., & Magno, M. (2023). TinyissimoYOLO: A Quantized, Low-Memory Footprint, TinyML Object Detection Network for Low Power Microcontrollers. arXiv preprint arXiv:2306.00001.
[8] Wu, Y., Wang, Y., Zhang, S., & Ogai, H. (2020). Deep 3D object detection networks using LiDAR data: A review. IEEE Sensors Journal, 21(2), 1152-1171.
[9] Coudron, I., & Goedemé, T. (2017, February). Multi-view tof fusion for object detection in industrial applications. In International Conference on Computer Vision Theory and Applications (Vol. 5, pp. 203-207). SCITEPRESS.