How Training Data Influence The Recognition Performance?

Yunfan Chen, Sai Rishi, Atharva Sharma, Hyunchul Shin

Abstract: Effective fusion of multispectral images captured by visible and infrared cameras enables robust pedestrian detection under various surveillance situations (e.g., daytime and nighttime). However, the performance of detecting small-sized pedestrian instances is still not satisfactory, while small pedestrian detection is important for self-driving and drone vision. Therefore, our effort focuses on improving the detection performance of small-sized multispectral pedestrians which are relatively far from the camera. Since existing multispectral pedestrian datasets mainly consider the large size pedestrians of 50 or more pixels in height, we generate a multispectral pedestrian dataset, named HH (Hanyang and Huins), in which the pedestrian height is from 25 to 50 pixels. To balance the trade-off between the detection performance and speed, we investigate a fusion network to combine two single-shot detectors (SSDs) for the fusion of visible and infrared inputs. The proposed fusion network is trained on public KAIST and KAIST + HH datasets, respectively. From the experimental results, we can observe that the detection performance has been improved a lot by incorporating HH dataset into KAIST dataset for training. The network trained by the original KAIST dataset has only 7.40% average precision (AP). However, the results can be significantly improved to 88.32% by using KAIST + HH for training. This indicates that training images have a great impact on detection performance.

Index Terms: Fusion network, visible, infrared, small-sized pedestrian detection

1. INTRODUCTION

In recent years, multispectral pedestrian detection has attracted attention from researchers in computer vision. The objective of multispectral pedestrian detection is to accurately locate the position of pedestrians from visible and infrared images captured in various real-world surveillance situations, especially under insufficient illumination conditions. Many advanced fusion frameworks have been proposed to effectively fuse the complementary information of multispectral images aiming at detecting pedestrians under daytime and nighttime [1-5].

In [1], ACF+T+THOG features, which is a combination of ACF from visible images, intensity channel feature T from thermal images, and THOG feature from thermal images, were designed to train the AdaBoost classifier for multispectral pedestrian detection. The FRCNN Halfway Fusion [2] were proposed integrated two-steam CNNs on the middle-level of FRCNN, which achieved better results. Based on the FRCNN Halfway Fusion, Fusion RPN+BDT [3] used a boosted decision tree (BDT) for classification instead of the original downstream classifier in FRCNN. Takumi et al. [4] published a multispectral object detection dataset, which contains 1466 groups of correctly aligned images taken by car-roof cameras. Chen et al. [5] proposed a multilayer fusion RPN using summation fusion strategy.

- *Yunfan Chen, Division of Electrical Engineering, Hanyang University, Korea. E-mail: chenyunfan@hanyang.ac.kr*
- *Sai Rishi, Digital Systems Lab, Hanyang University, Korea (Intern, Summer 2019), Indian Institute of Technology, Guwahati, India. E-mail: rishi170102056@iitg.ac.in*
- *Atharva Sharma, Digital Systems Lab, Hanyang University, Korea (Intern, Summer 2019), Indian Institute of Technology, Guwahati, India. E-mail: athar170108008@iitg.ac.in*
- *Hyunchul Shin, Division of Electrical Engineering, Hanyang University, Korea. Corresponding. E-mail: shin@hanyang.ac.kr*

However, the existing multispectral pedestrian detectors mainly focus on large-sized pedestrians, and they are likely to fail to detect small-sized pedestrian instances. In addition, existing multispectral pedestrian datasets [1, 4] are primarily consisting of large pedestrian instances (no less than 50 pixels in height). There is no public multispectral pedestrian dataset for small-sized pedestrian detection (25 to 50 pixels in pedestrian height). To overcome the problem arising from small pedestrian, we generate a multispectral pedestrian dataset, especially for small-sized pedestrian detection, we name it as Hanyang and Huins (HH) dataset. The pedestrian size of proposed HH dataset is from 25 to 50 pixels, which is more suitable for evaluating the detector performance on small object detection. We adopt a simple fusion network based on two single-shot detectors (SSDs) [6] for detection. Experimental results indicate that the proposed method trained on KAIST achieves the poor performance of 7.4% AP while the proposed method trained on KAIST+HH achieves 88.32% AP, resulting in significant improvement of 80.92%.

2 PROPOSED HH MULTISPECTRAL PEDESTRIAN DATASET

In this work, we generated a novel multispectral dataset for small-sized pedestrian detection that consists of pixel-level aligned visible and infrared images and added ground truth labels. The pictures were taken using RGB and FIR cameras mounted on Drones. In total, we collected 7,247 pairs of images with a size of 720 × 480. We divided the whole dataset into training and testing parts, the training part contains 6247 pairs of images, and the testing part contains 1000 pairs of images. The ground truth consists of bounding box coordinates and labels. Fig. 1 shows some four sample images of our dataset.

3 MULTISPECTRAL PEDESTRIAN DETECTION

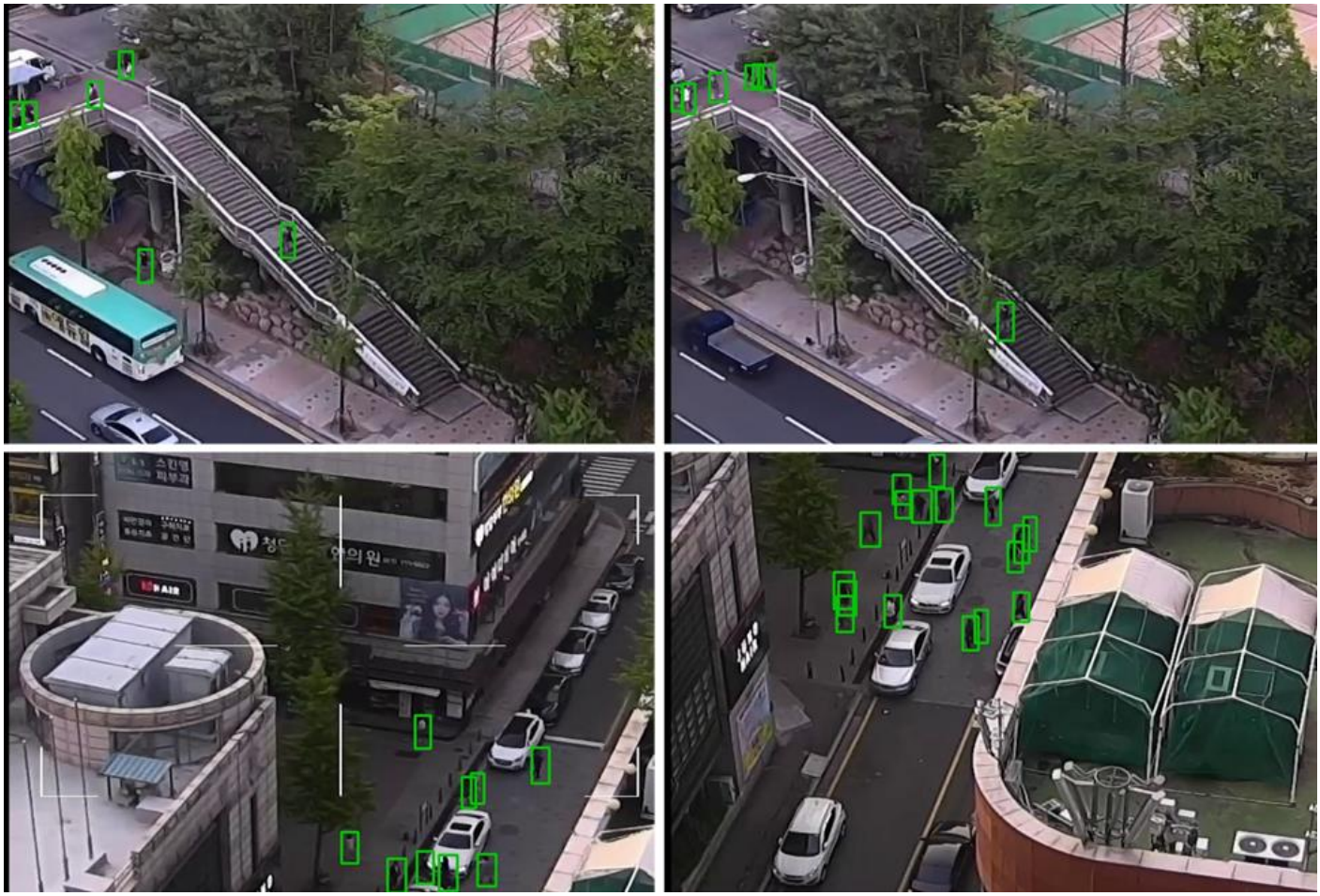


***Fig. 1**. Some example images of HH dataset. The green bounding boxes denote the ground truth.*

An overview of the proposed fusion network is illustrated in Fig. 2. The input visible and infrared images are first processed by two SSDs with several convolutional and pooling layers to generate multi-scale feature maps. Second, the two SSDs are integrated via the fusion block to generate fused feature maps. Finally, the fused feature maps are sent to the prediction stage for classifying pedestrians and localizing bounding boxes. In training stage, we use stochastic gradient descent to fine-tune the entire network, in which the momentum is set to 0.9, and weight decay is set to 0.0005. To prevent gradient explosion in early iterations, we first run 10k iterations using a learning rate of 0.00005. Then, we reset the learning rate to 0.001 for the next 70k iterations. After the completion of 70k iterations, the learning rate is reduced by 10 times after every 20k iterations. Learning stops after 120k iterations.

4 EXPERIMENTAL RESULTS

The proposed approach is evaluated on the HH multispectral pedestrian dataset. We use a Precision-Recall curve to

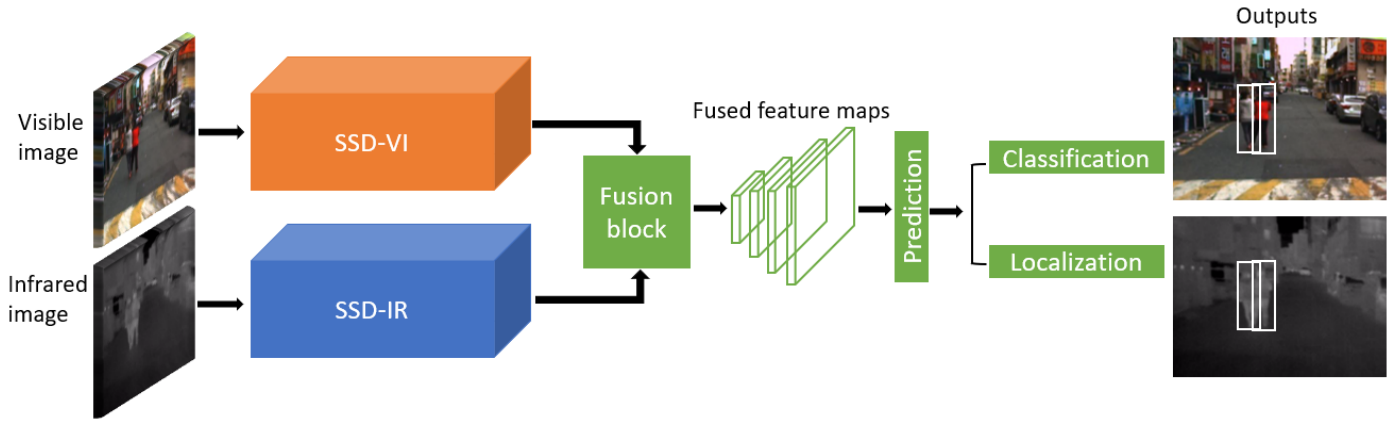


***Fig. 2**. Fusion of two SSDs for multispectral pedestrian detection.*

validate detection performance. Precision (Pr) and recall (Re) are calculated according to true positives (TP), false positives (FP), and false negatives (FN),

$$\mathrm{Pr} = \frac{TP}{TP + FP}, \mathrm{Re} = \frac{TP}{TP + FN}. \quad (1)$$

A TP using the PASCAL overlap criterion [7] as

$$IOU(B_p, B_{gt}) \geq 0.5 \quad (2)$$

where B_p is a predicted bounding box, B_{gt} is a ground truth bounding box.

We implemented the proposed network on the Caffe framework and used three NVIDIA Titan X for training and a single Nvidia Titan X for testing.

We compared the performance of the proposed network trained on KAIST dataset and KAIST + HH dataset. As shown in Fig. 3, the network trained on KAIST+HH achieves the better AP of 88.32%, which significantly exceeds the performance of the network trained on KAIST by 80.92% AP. Table 1 gives a comprehensive comparison. We can see that the proposed fusion network trained on KAIST+HH dataset has a better balance between the computational speed, the detection precision, and model size, which is applicable to real applications. Fig. 4 provides a visual comparison of the detection results, where evidently the proposed method trained on KAIST+HH dataset shows superior results, it successfully detects all pedestrians without errors.

5 CONCLUSIONS

In this research, we generated our HH multispectral dataset that includes pixel-level aligned visible and infrared images for small-sized pedestrian detection in traffic scenes. In addition, a fusion network based on SSDs is proposed for small-sized multispectral pedestrian detection. The proposed network is trained on KAIST dataset and KAIST+HH dataset, respectively, for evaluating the effect of training data on detection performance. Experimental results indicate that our approach trained by KAIST+HH significantly outperforms the proposed method trained on KAIST, 88.32% versus 7.4% AP. This result reveals that the training images have a great impact on detection performance. After adding small-sized training images, the detection performance of the detector improved by 80.92% of AP. Our method can detect small-sized pedestrians with 25 or more pixels in height successfully.

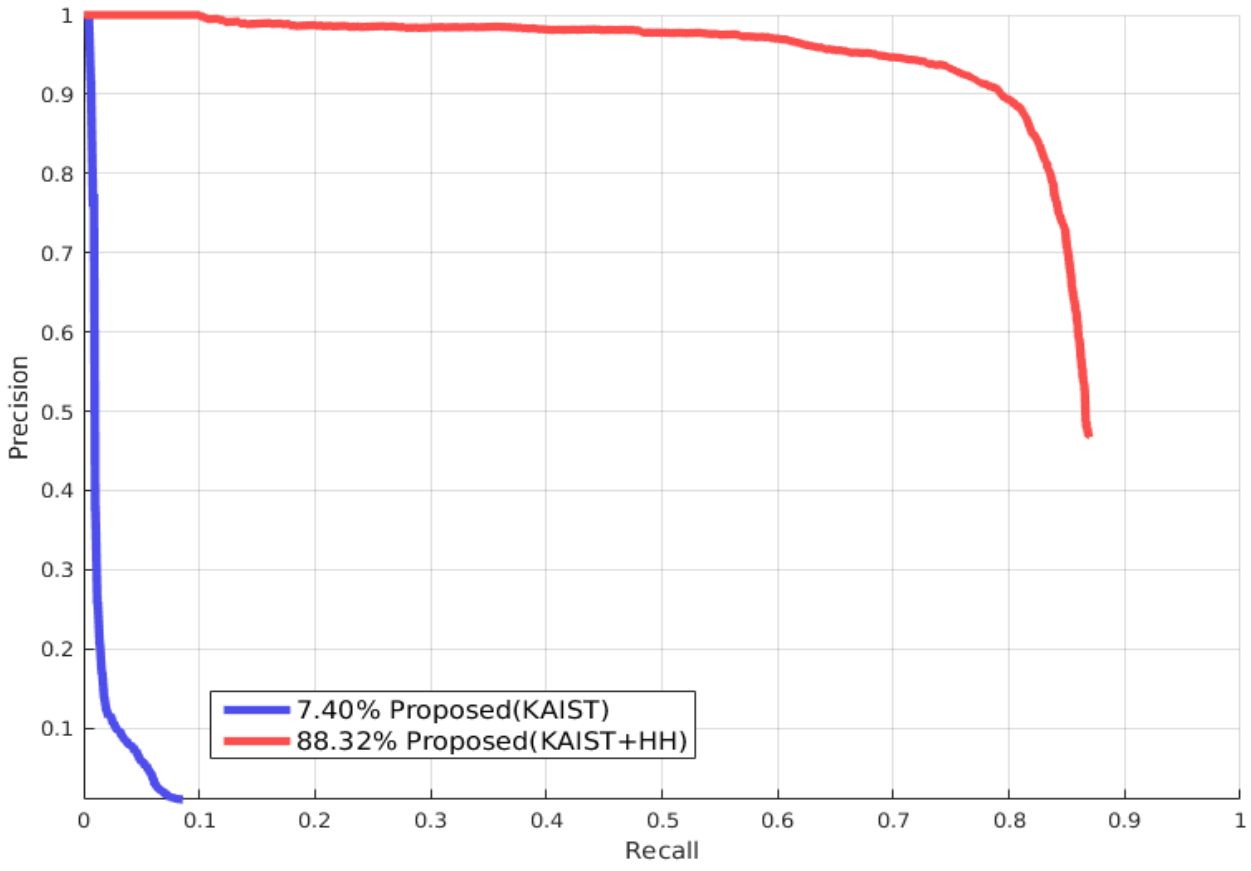
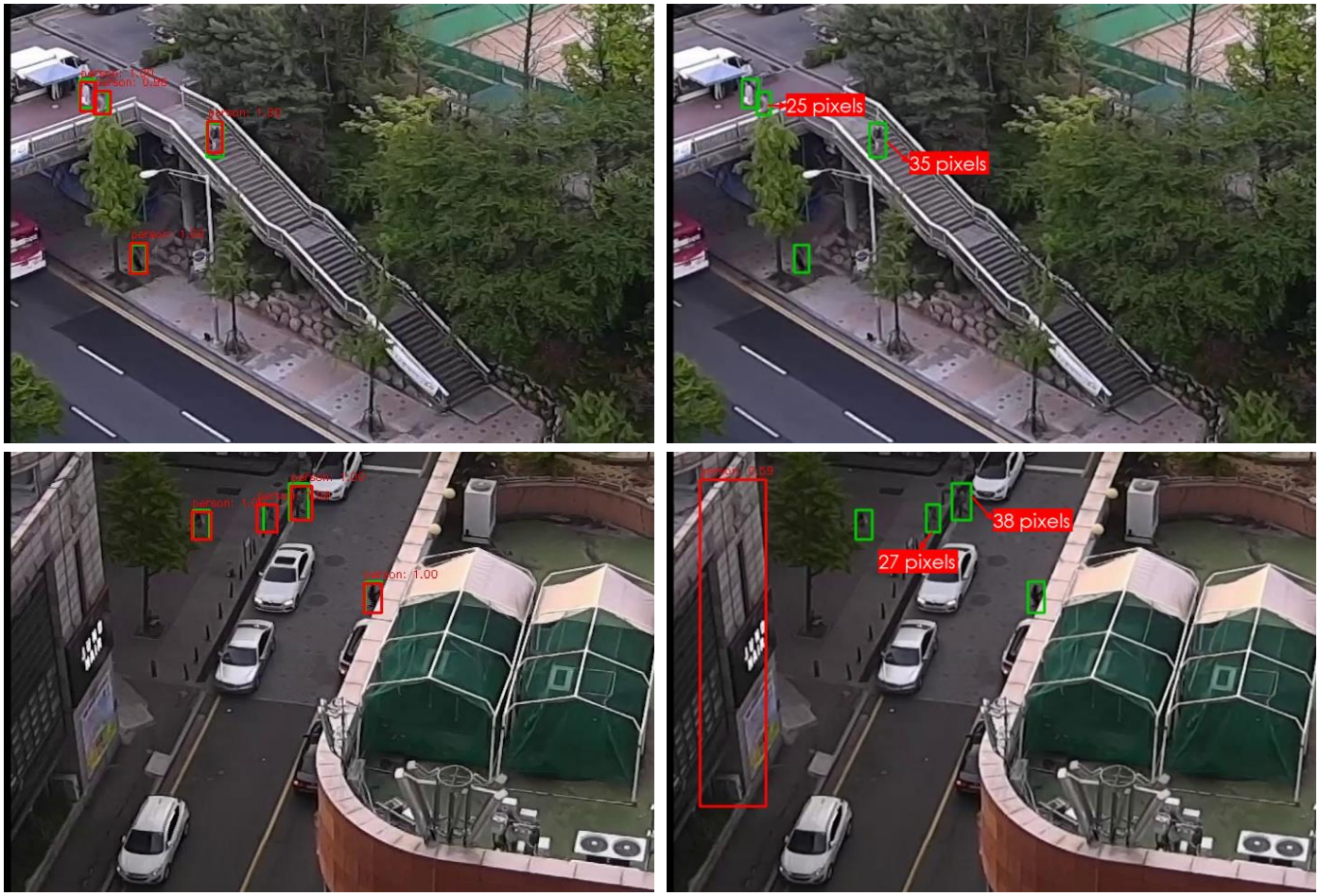


Fig. 3. Performance comparison of Precision-Recall curve using the HH dataset.

TABLE 1
COMPREHENSIVE COMPARISON ON HH TEST SET

Methods	Training dataset	Average precision (AP)	CPU times (seconds per frame)	Model size (Mb)
Proposed fusion network	KAIST	7.40%	0.05	249.4
Proposed fusion network	KAIST + HH	88.32%	0.05	249.4



(a) (b)

Fig. 4*. Visual comparison on HH test set. The red bounding boxes show the detection results, and the green bounding boxes denote the ground truth. (a) Proposed method trained on KAIST+HH, (b) Proposed method trained on KAIST.*

ACKNOWLEDGMENT

This work was supported by Basic Research Project in Science and Engineering through the Ministry of Education of the Republic of Korea and National Research Foundation of Korea (National Research Foundation of Korea 2017-R1D1A1B04- 031040).

REFERENCES

[1] Hwang, S., Park, J., Kim, N., et al.: 'Multispectral pedestrian detection: benchmark dataset and baseline'. Proc. IEEE Conf. Computer Vision and Pattern Recognition, Boston, MA, USA, June 2015, pp. 1037–1045.

[2] Liu, J., Zhang, S., Wang, S., et al.: 'Multispectral deep neural networks for pedestrian detection'. Proc. British Machine Vision Conf., York, UK, September 2016, pp. 1–13.

[3] König, D., Adam, M., Jarvers, C., et al.: 'Fully convolutional region proposal networks for multispectral person detection'. Proc. IEEE Workshop on Computer Vision and Pattern Recognition, Honolulu, HI, USA, July 2017, pp. 243–250.

[4] Takumi, Karasawa, et al. "Multispectral object detection for autonomous vehicles." Proceedings of the on Thematic Workshops of ACM Multimedia 2017. ACM, 2017.

[5] Chen, Yunfan, Han Xie, and Hyunchul Shin. "Multi-layer fusion techniques using a CNN for multispectral pedestrian detection." IET Computer Vision 12.8 (2018):

1179-1187.
[6] Liu, Wei, et al. "Ssd: Single shot multibox detector." European conference on computer vision. Springer, Cham, 2016.
[7] M. Everingham, L. Van Gool, C.K. Williams, J. Winn, A. Zisserman, The Pascal visual object classes (voc) challenge, Int. J. Comput. Vis. 88 (2) (2010) 303–338.