A Benchmark Dataset for Wearable Low-Light Pedestrian Detection

Contributor: Tianlin Li, Guanglu Sun, Linsen Yu, Kai Zhou

Detecting pedestrians in low-light conditions is challenging, especially in the context of wearable platforms. Infrared cameras have been employed to enhance detection capabilities, whereas low-light cameras capture the more intricate features of pedestrians. With this in mind, a low-light pedestrian detection (called HRBUST-LLPED) dataset by capturing pedestrian data on campus using wearable low-light cameras is introduced.

1. Introduction

Over the past two decades, there has been a significant advancement in IoT and artificial intelligence technologies. As a result, researchers have turned their attention to developing intelligent wearable assistive systems that are made up of wearable cameras, sensors, computing components, and machine learning models [1]. This has led to an increase in studies aimed at assisting visually impaired individuals in various areas, such as travel [2], food [3], and screen detection [4]. Other areas of research include human-pet [5] or human-machine [6][7] interaction. Wearable devices combined with computer vision models are being used to help users observe things that are typically difficult to see. Despite the numerous studies on object detection using wearable devices, research on detecting humans in a scene is still limited, making it challenging to apply in areas such as nighttime surveillance, fire rescue, and forest inspections.

Since the maturity of convolutional neural networks in 2012, object detection algorithms have experienced vigorous development [8]. Single-stage object detection models represented by SSD [9] and YOLO [10], as well as two-stage object detection models depicted by Faster R-CNN [11] and FPN [12], have been proposed, achieving excellent results in terms of speed and accuracy. The maturity of object detection algorithms has also ushered in pedestrian detection algorithms into the era of deep learning. In order to fulfill the training data needs of machine learning and deep learning models, some usual pedestrian detection datasets have been proposed, like Caltech [13] and KITTI [14]. In recent years, datasets such as CityPersons [15], CrowdHuman [16], WIDER Pedestrian, WiderPerson [17], EuroCity [18], and TJU-Pedestrian [19] have been collected from cities, the countryside, and broader environments using vehicle-mounted cameras or surveillance cameras. These datasets enable the trained models to adapt to a broader range of scenarios. The EuroCity and TJU-Pedestrian datasets also include pedestrian data with low illumination conditions, aiming to achieve good recognition performance in terms of pedestrian detection models in nighttime scenarios. However, conventional cameras struggle to capture clear images under low-light conditions, significantly impacting data annotation and model recognition performance, as shown in **Figure 1**a.
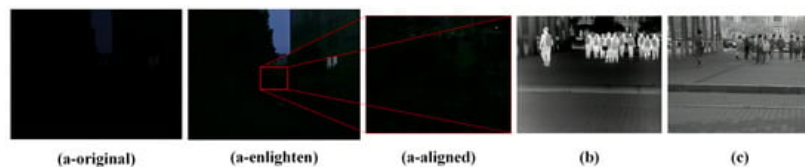

Figure 1. In starlight-level illumination environments, the imaging effects of visible light cameras, infrared cameras, and low-light cameras are as follows: (**a-original**) represents an image captured directly with a mobile phone; (**a-enlighten**) represents the image enhanced using the Zero DCE++ model; (**a-aligned**) represents the image in the enhanced version with the corresponding resolution. (**b**) represents an image captured with an infrared camera, and (**c**) illustrates an image captured with a low-light camera.

Humans emit heat, which can be captured using infrared cameras in colder environments to distinguish pedestrians from the background, as shown in **Figure 1**b. As a result, OSU [20] proposed an infrared dataset collected during the daytime,

and TNO [21] also provided a dataset that combines infrared and visible light captured at night. Later, with the development of research on autonomous driving, datasets such as CVC-14 [22], KAIST [23], and FLIR were introduced, which consist of pedestrian data captured using vehicle-mounted visible light-infrared cameras for modal alignment. Subsequently, the LLVIP dataset [24] was introduced to advance research on multi-spectral fusion and pedestrian detection under low-light conditions. Although infrared images can separate individuals from the background, they have a limited imaging distance and contain less detailed textures, making it difficult to distinguish pedestrians with high overlap.

Low-light cameras with CMOS chips specially designed to capture long-wavelength light waves can achieve precise imaging under starlight-level illumination conditions, as shown in **Figure 1**c. By considering the helpfulness of low-light images for pedestrian detection in low-light environments, researchers constructed the Low-Light Pedestrian Detection (HRBUST-LLPED, collected by Harbin University of Science and Technology) dataset. The dataset consists of 150 videos captured under low-light conditions, from which 4269 keyframes were extracted and annotated with 32,148 pedestrians. In order to meet the requirements of wearable devices, researchers developed wearable low-light pedestrian detection models based on small and nano versions of YOLOv5 and YOLOv8. When considering the fact that the information captured by low-light cameras is relatively limited compared to visible-light cameras, researchers first trained the models on the KITTI, KAIST, LLVIP, and TJU-Pedestrian datasets separately and then fine-tuned them using the dataset. As a result, the trained models achieved satisfactory results in speed and accuracy.

The contributions cover several aspects.

(1) Researchers have expanded the focus of pedestrian detection to low-light images and have constructed a low-light pedestrian detection dataset using a low-light camera. The dataset contains denser pedestrian instances compared to existing pedestrian detection datasets.

(2) Researchers have provided lightweight, wearable, low-light pedestrian detection models based on the YOLOv5 and YOLOv8 frameworks, considering the lower computational power of wearable platforms when compared to GPUs. researchers have improved the model's performance by modifying the activation layer and loss functions.

(3) Researchers first pretrained the models on four visible light pedestrian detection datasets and then fine-tuned them on the constructed HRBUST-LLPED dataset. researchers achieved a performance of 69.90% in terms of AP@0.5:0.95 and an inference time of 1.6 ms per image.

2. Dataset Build

Data Capture: The low-light camera researchers used is the Iraytek PF6L, with an output resolution of 720×576/8μm, a focal length of F25mm/F1.4, and a theoretical illuminance resolution of 0.002 Lux. The camera is attached to a helmet. researchers wear the helmet to capture data to simulate the real perspective of humans. researchers mainly shoot campus scenes from winter to summer. The collection time in winter was 18:00–22:00, and in summer, it was between 20:00 and 22:00. In total, researchers collected 150 videos with a frame rate of 60 Hz. The length of the videos ranged from 33 s to 7 min and 45 s, with an average length of 95 s and a total of 856,183 frames.

Data Process: First of all, when considering the thermal stability and high sensitivity of CCD (CMOS) in the video acquisition process of low-light cameras, noise will inevitably be introduced into the video. Additionally, since the frame rate of the video is 60 Hz and the difference in pedestrian poses between adjacent frames is minimal, researchers first use a smoothing denoising technique with the neighboring two frames to enhance the current frame. Next, when considering that the pedestrian gaits in the video are usually slow, there is significant redundancy in the pedestrian poses. Therefore, researchers select one frame every 180 frames (i.e., 3 s per frame) as a keyframe. researchers then remove frames that do not contain pedestrian targets and search for clear frames within a range of ±7 frames of the blurred frames to update the keyframes. Ultimately, researchers obtained 4269 frames as the image data for constructing the low-light pedestrian detection dataset.

Data Annotation: researchers used the labelImg tool to annotate the processed image data manually. For each person present in the image with less than 90% occlusion (i.e., except for cases where only a tiny portion of the lower leg or arm is visible), researchers labeled them as "Pedestrian". researchers cross-referenced the uncertain pedestrian annotations with the original videos to avoid missing pedestrians due to visual reasons or mistakenly labeling trees as pedestrians. As a result, researchers obtained a total of 32,148 pedestrian labels.

References

1. Li, H.; Liu, H.; Li, Z.; Li, C.; Meng, Z.; Gao, N.; Zhang, Z. Adaptive Threshold Based ZUPT for Single IMU Enabled Wearable Pedestrian Localization. IEEE Internet Things J. 2023, 10, 11749–11760.

2. Tang, Z.; Zhang, L.; Chen, X.; Ying, J.; Wang, X.; Wang, H. Wearable supernumerary robotic limb system using a hybrid control approach based on motor imagery and object detection. IEEE Trans. Neural Syst. Rehabil. Eng. 2022, 30, 1298–1309.

3. Han, Y.; Yarlagadda, S.K.; Ghosh, T.; Zhu, F.; Sazonov, E.; Delp, E.J. Improving food detection for images from a wearable egocentric camera. arXiv 2023, arXiv:2301.07861.

4. Li, X.; Holiday, S.; Cribbet, M.; Bharadwaj, A.; White, S.; Sazonov, E.; Gan, Y. Non-Invasive Screen Exposure Time Assessment Using Wearable Sensor and Object Detection. In Proceedings of the 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Glasgow, UK, 11–15 July 2022; pp. 4917–4920.

5. Kim, J.; Moon, N. Dog behavior recognition based on multimodal data from a camera and wearable device. Appl. Sci. 2022, 12, 3199.

6. Park, K.B.; Choi, S.H.; Lee, J.Y.; Ghasemi, Y.; Mohammed, M.; Jeong, H. Hands-free human–robot interaction using multimodal gestures and deep learning in wearable mixed reality. IEEE Access 2021, 9, 55448–55464.

7. Dimitropoulos, N.; Togias, T.; Michalos, G.; Makris, S. Operator support in human–robot collaborative environments using AI enhanced wearable devices. Procedia CIRP 2021, 97, 464–469.

8. Zou, Z.; Chen, K.; Shi, Z.; Guo, Y.; Ye, J. Object detection in 20 years: A survey. Proc. IEEE 2023, 111, 257–276.

9. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. Ssd: Single shot multibox detector. In Proceedings of the Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; Proceedings Part I 14. Springer: Berlin/Heidelberg, Germany, 2016; pp. 21–37.

10. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.

11. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Trans. Pattern Anal. Mach. Intell. 2017, 39, 1137–1149.

12. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.

13. Dollar, P.; Wojek, C.; Schiele, B.; Perona, P. Pedestrian detection: An evaluation of the state of the art. IEEE Trans. Pattern Anal. Mach. Intell. 2011, 34, 743–761.

14. Geiger, A.; Lenz, P.; Urtasun, R. Are we ready for autonomous driving? The kitti vision benchmark suite. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 3354–3361.

15. Zhang, S.; Benenson, R.; Schiele, B. Citypersons: A diverse dataset for pedestrian detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 3213–3221.

16. Shao, S.; Zhao, Z.; Li, B.; Xiao, T.; Yu, G.; Zhang, X.; Sun, J. Crowdhuman: A benchmark for detecting human in a crowd. arXiv 2018, arXiv:1805.00123.

17. Zhang, S.; Xie, Y.; Wan, J.; Xia, H.; Li, S.Z.; Guo, G. Widerperson: A diverse dataset for dense pedestrian detection in the wild. IEEE Trans. Multimed. 2019, 22, 380–393.

18. Braun, M.; Krebs, S.; Flohr, F.; Gavrila, D.M. Eurocity persons: A novel benchmark for person detection in traffic scenes. IEEE Trans. Pattern Anal. Mach. Intell. 2019, 41, 1844–1861.

19. Pang, Y.; Cao, J.; Li, Y.; Xie, J.; Sun, H.; Gong, J. TJU-DHD: A diverse high-resolution dataset for object detection. IEEE Trans. Image Process. 2020, 30, 207–219.

20. Davis, J.W.; Sharma, V. OTCBVS Benchmark Dataset Collection. 2007. Available online: https://vcipl-okstate.org/pbvs/bench/ (accessed on 2 September 2023).

21. Toet, A. The TNO multiband image data collection. Data Brief 2017, 15, 249–251.

22. González, A.; Fang, Z.; Socarras, Y.; Serrat, J.; Vázquez, D.; Xu, J.; López, A.M. Pedestrian detection at day/night time with visible and FIR cameras: A comparison. Sensors 2016, 16, 820.

23. Hwang, S.; Park, J.; Kim, N.; Choi, Y.; So Kweon, I. Multispectral pedestrian detection: Benchmark dataset and baseline. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1037–1045.

24. Jia, X.; Zhu, C.; Li, M.; Tang, W.; Zhou, W. LLVIP: A visible-infrared paired dataset for low-light vision. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 3496–3504.