Adapting Deep Neural Networks for Pedestrian-Detection to Low-Light Conditions without Re-training

Vedant Shah
Dept. of EEE and APPCAIR, BITS Pilani,
K K Birla, Goa Campus, 403726, Goa, India.
f20180566@goa.bits-pilani.ac.in

Anmol Agarwal
Dept. of CSIS and APPCAIR, BITS Pilani,
K K Birla, Goa Campus, 403726, Goa, India.
f20170489@goa.bits-pilani.ac.in

Tanmay Tulsidas Verlekar
Dept. of CSIS and APPCAIR, BITS Pilani,
K K Birla, Goa Campus, 403726 Goa, India.
tanmayv@goa.bits-pilani.ac.in

Raghavendra Singh
Oyla Inc.
San Carlos, CA 94070, United States.
rsingh@oyla.ai

Abstract

Pedestrian detection is an integral component in many automated surveillance applications. Several state-of-the-art systems exist for pedestrian detection, however most of them are ineffective in low-light conditions. Systems specifically designed for low-light conditions require special equipment, such as depth sensing cameras. However, a lack of large publicly available depth datasets, prevents their use in training deep learning models.

In this paper we propose a pre-processing pipeline, which enables any existing normal-light pedestrian detection system to operate in low-light conditions. It is based on a signal-processing and traditional computer-vision techniques, such as the use of signal strength of a depth sensing camera (amplitude images) and robust principal component analysis (RPCA). The information in an amplitude image is less noisy, and is of lower dimension than depth data, marking it computationally inexpensive to process. RPCA processes these amplitude images to generate foreground masks, which represent potential regions of interest. These masks can then be used to rectify the RGB images to increase the contrast between the foreground and background, even in low-light conditions. We show that these rectified RGB images can be used by normal-light deep learning models for pedestrian-detection, without any additional training.

To test this hypothesis, we use the 'Oyla Low-Light Pedestrian Benchmark' (OLPB) dataset. Our results using two state-of-the art deep learning models (CrowdDet and CenterNet) show: a) The deep models perform poorly as pedestrian detectors in low-light conditions; b) Equipping the deep-networks with our pre-processing pipeline significantly improves the average precision for pedestrian-detection of the models without any re-training. Taken together, the results suggest that our approach could act as a useful pre-processor for deep learning models that aren't specially designed for pedestrian-detection in low-light conditions.

1. Introduction

Pedestrian detection has been an important area of study in the computer vision community. With new found applications in several areas, such as self-driving cars and automated surveillance systems, it has undergone rapid development in recent years. Most modern day pedestrian detection systems rely on deep learning models, which enable them to operate in challenging environments, such as crowded streets using a simple RGB camera [1]. However, they are seldom effective in low-light conditions. Under such conditions, RGB images cannot capture the contrast needed in the scene to obtain usable textures [2]. To address this problem, some pedestrian detection systems train the models with additional information, such as depth or thermal data [3]. However, this makes the system computationally expensive, while also limiting their range of operation.

1.1. State-of-the-art

The state-of-the-art can be broadly grouped into:

- 2D image based systems;
- Multi-spectral data based systems.

2D image based systems: The CrowdDet [1] and the CenterNet [2] are among the best 2D image based pedes-

trian detection systems. They perform pedestrian detection by training deep learning models, such as ResNet [4] and a DLA-34 encoder-decoder network [5], on CrowdHuman [6] and COCO [7] datasets, respectively. The large size and variety of scenarios available in these datasets, allow the system to operate in diverse and challenging environments. However, being based on 2D images, they perform poorly in low-light conditions.

The performance of the 2D image based pedestrian detection systems in low-light conditions can be improved by performing certain enhancements to the input images. The 2D images can be enhanced by using multi-task learners, which detect pedestrians by relighting the scene [8], or by using 14-bit, long-exposure, raw video sequences during training [9]. However, these systems involve a higher computational cost, while also being restricted by the specific requirements and properties of the training data.

Multi-spectral data based systems: A better performance can be expected from systems that rely on more than just RGB channel data. This involves the use of special cameras that can capture thermal or depth data from a scene. Systems that rely on thermal data, also make use of RGB images, by appending everything together to create multi-spectral images [3]. Re-training systems, such as YOLOv3 object detector [10], with these multi-spectral images can significantly improve their performance in low-light conditions. The thermal data can separately be used to guide the attention of a deep learning model towards the pedestrians, in low-light conditions. These models can then learn the feature set of the poorly illuminated pedestrian, from their RGB images [11]. A limitation of the cameras used for thermal imaging is their short range of operation. A wider range of operation can be achieved through depth sensing cameras, such as LIDAR, but with a significantly higher equipment and computational cost. The depth data, captured by the LIDAR cameras, can be used to perform pedestrian detection using simple techniques, such as contour shape analysis [12], after undergoing significant pre-processing. A fusion of data from both depth sensing and thermal imaging cameras can result in images that can be processed using traditional machine learning techniques, such as histogram of oriented gradients and support vector machine [13].

Since training deep learning models requires a large amount of data, availability of public datasets can also be considered as an advantage to the systems. The largest available 2D image dataset called ImageNet, contains more than 14 million images. Among them, at least one million images contain bounding boxes for around 20,000 categories. The MS COCO dataset [7] contains 3,28,000 RGB images, belonging to 91 object categories. While, the largest dataset specially curated for pedestrian detection, called the Caltech Pedestrian Detection Benchmark [14], contains 2,50,000 RGB images with 3,50,000 annotated pedestrians.

The most popular dataset for Multi-spectral data based systems is the KITTI Vision Benchmark [15], which contains 93,000 depth maps with the corresponding LIDAR scans and raw RGB data [16] of various crowded streets. Apart from KITTI, the only other multispectral dataset to the best of our knowledge having pedestrians is the eCo-DRIVERS dataset [17], which contain 11,071 images of pedestrians in the far-infrared spectrum.

From the discussion in section 1.1, it can be concluded that 2D image based systems are best suited for pedestrian detection in surveillance environments. However, their poor performance in low-light conditions necessitate an alternative to these systems. The current state-of-the-art alternates involve the use of expensive equipment, computationally expensive deep learning models, short range of operation or specific use cases (i.e. trained on small datasets). Thus, we propose a computationally inexpensive, pre-processing pipeline, that can run on any existing 2D image based, normal-light pedestrian detection system without any re-training. The pre-processing involves the use of a depth sensing camera and traditional computer vision technique, such as RPCA, to generate rectified RGB images. These images contain sufficient contrast between its background and foreground objects, which is necessary to perform pedestrian detection. To our best knowledge, there are no public datasets that capture pedestrians in low-light conditions using a depth sensing camera. Thus, we evaluate our system using the OLPB dataset

Rest of this paper is organised into the following sections: Section 2 presents the proposed pre-processing pipeline. Section 3 presents the Oyla Low-Light Pedestrian Benchmark dataset (Section 3.1), and the deep learning models used for performance evaluation (Section 3.2). The result of the performance evaluation is discussed in Section 3.4. Finally, the conclusion in section 4 presents the limitations of the proposed pipeline and possible future directions. This paper also includes an Appendix A, for the Oyla's 3D-aware surveillance camera, which is used for data acquisition.

2. The proposed pre-processing pipeline

The proposed pre-processing pipeline is illustrated in Figure 1. It consists of two important steps:

- Foreground segmentation;
- RGB image rectification;

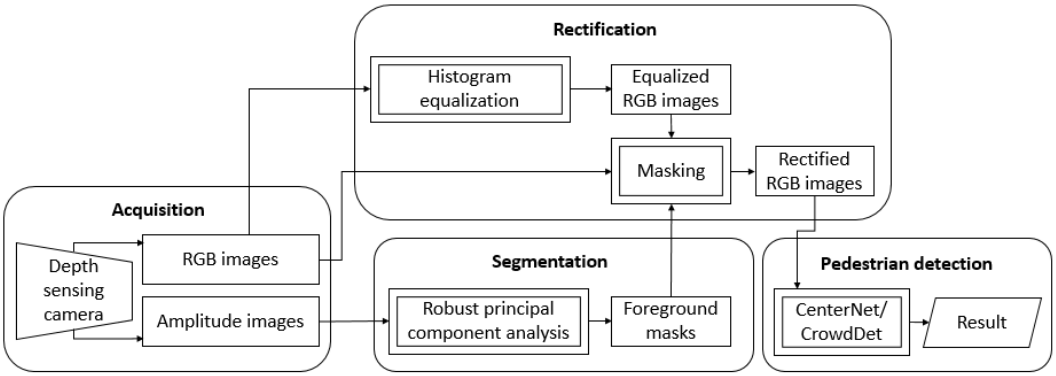
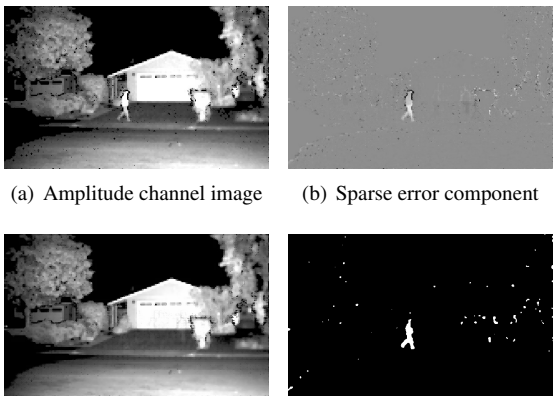


Figure 1. Proposed pre-processing pipeline.

The inputs to the proposed pipeline is a set of RGB images and the information about the strength of the depth signal, received at the receiver. We express this information in the form of a 2D image called the "amplitude image". It provides the pipeline with useful information about the observed scene, even in low-light conditions. However, it is unable to capture any texture related information about the scene, which most deep learning models use for pedestrian-detection.

Thus, the proposed pre-processing pipeline uses the amplitude images to improve the contrast of the foreground objects in the RGB images, using foreground segmentation and RGB image rectification. The rectified image with the improved contrast can then be used by any normal-light, deep learning models for pedestrian-detection without additional training.



(a) Amplitude channel image (b) Sparse error component

(c) Low rank component (d) Foreground mask

Figure 2. Example of an input (a) and the corresponding outputs (b, c, d) from RPCA

2.1. Foreground segmentation

The proposed pre-processing pipeline uses robust principal component analysis (RPCA) to separate foreground objects from the background [18]. A sequence of amplitude channel images are used as input to RPCA, which decomposes them into a low-rank component and a sparse error component, as illustrated in Figure 2. The low-rank component is composed of all the redundant/static parts in the amplitude channel images. While, the sparse error component captures the difference in the amplitude channel images. These differences, which cannot be incorporated into the low-rank component correspond to the dynamic parts of the image, such as the pedestrians and other moving objects.

To segment the foreground, each amplitude channel image is transformed into a 1D column vector. Multiple images corresponding to a sequence are appended together to create a $M \times N$ matrix, where M is the length of 1D vectors and N is the number of images in the sequence. The matrix D is decomposed into a low-rank component L and a sparse error component S following Equation 1, using RPCA.

$$min_{L,E}\ rank(L) + \gamma \left\|S\right\|_o\ subject\ to\ D = L + S \quad (1)$$

Where $rank(L)$ represents the number of linearly independent rows in the matrix D, γ is a regularization parameter and $\left\|S\right\|_o$ is the counting norm of the sparse error. According to Equation 1, RPCA exploits the repeating patterns in each image, i.e. the static background, to separate the sparse errors in the matrix D as the foreground. Since minimisation of rank is a non convex problem, RPCA uses a simple alternating minimization algorithm to solve a convex relaxed variation of Equation 1, presented in [19]. The resulting sparse error component S is decomposed (column-wise) into a set of sparse error images. These images are thresholded to obtain the binary foreground masks, as illustrated in Figure 2(d).

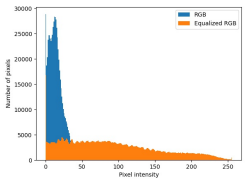
2.2. RGB Image Rectification

To address the problem of the lack of contrast in RGB channel images, under low-light conditions, most systems

use a technique called histogram equalization. The technique adjusts the intensity value of each pixel to enhance the global contrast. However, as illustrated in Figure 3(b), sometimes even after histogram equalization, an image can lack sufficient contrast to separate foreground objects from the background. This effect is intensified, when the pedestrians walk further away from the camera or a light source. Although, RGB and histogram equalized images are individually ineffective, a combination of the two images can be used to create the contrast needed to perform pedestrian detection.

(a) RGB channel image

(b) Histogram equalized image



(c) Histograms

(d) Rectified image

Figure 3. Example of inputs (a, b) and the corresponding rectified output (d)

In the proposed pre-processing pipeline the RGB and histogram equalized images are combined using the corresponding foreground mask, to generate an image called the rectified RGB image. The foreground mask, obtained using RPCA, is used to segment the dynamic parts of the histogram equalized image. These parts include pedestrian and other moving objects in the scene. They are then stitched onto the original RGB image following Equation 2. The resulting rectified image contains sufficient contrast between the background and the foreground objects, as illustrated in Figure 3(d). The contrast can then be used to identify the pedestrians among the foreground objects, using any normal-light, deep learning models for pedestrian-detection.

$$I_{FG} \odot I_{RGB} + (1 - I_{FG}) \odot I_{EQ} = I_{Rect} \qquad (2)$$

3. Empirical evaluation

We aim to test the hypothesis that equipping normal-light, deep learning models with the proposed pre-processing pipeline described in section 2 will significantly improve their pedestrian detection ability in low-light conditions, without affecting their computational complexity. The evaluation is conducted on the dataset described below, following the evaluation method discussed in section 3.3.

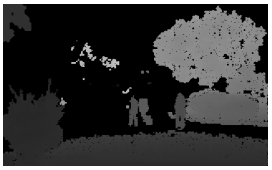
3.1. Oyla low-light pedestrian benchmark dataset

OLPB dataset is captured using a novel depth sensing camera by Oyla, which is a USA based startup. The camera, as discussed in Appendix A, captures images across 5 channels. Thus, apart from RGB information, the camera also captures depth and strength information of the received signals. The depth information is captured as a uint16 $nxmxd$ matrix, using the time of flight (TOF) principle. The matrix can be represented as depth map (Figure 4(c)), or the depth data can be transformed into Cartesian coordinates and represented as a point cloud (5(d)). The strength of the TOF signal at the receiver is captured as a uint16 nxm matrix, and is visualized as an image, called the amplitude image, as illustrated in Figure 4(b). Thus, the OLPB dataset contains four different representations of an observed scene.

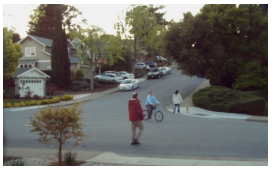
The current OLPB dataset is captured in a neighbourhood (COVID restrictions) using two cameras with a resolution of 480x640, as illustrated in Figure 4(a), 3(a). The scene is captured under different lighting conditions, with up to four pedestrians walking along the road, at varying distances from the camera, for a total of 1521 times. The observed scene also includes other moving objects, such as cars, bicycles and dogs. The pedestrians are annotated to act as the ground truth for pedestrian detectors, and are made available in the dataset in the PASCAL VOC [20] format.

(a) RGB image in low-light conditions

(b) Amplitude Image



(c) Depth Image



(d) RGB image in normal-light conditions

Figure 4. Example of images in the OLPB dataset

The proposed pipeline uses only the amplitude and the RGB images from the OLPB dataset. However, all the available images cannot directly be used by the proposed pipeline. Among the available 1521 set of images, only 770 contain usable annotations. Of those 770 set of images, the camera is static, from one image to the next, only in 430 instances. Thus, only the 430 RGB and the correspond-

ing amplitude images are used to evaluate the performance of the proposed pipeline. The selected images are visually inspected to classify them as either normal or low-light images, - see Table 1

Table 1. Statistics of OLPB dataset

Condition	Total images
in OLPB dataset	1521
annotated images	770
usable for the proposed pipeline	430
normal-light	201
low-light	229

3.2. Normal-light, deep learning models

The rectified images obtained using the proposed pre-processing pipeline, can be used by any normal-light, deep learning models for pedestrian detection, in low-light conditions. To emphasize the plug and play nature of the proposed pipeline, we use the following deep learning models to evaluate it's performance:

CrowdDet [1] specializes in detecting pedestrians in a crowded environment on a 2D image. Under such conditions, detection becomes difficult as the pedestrians overlap each other, inhabiting the same 2D space. To address this, CorwdDet generates multiple instances of tentative predictions for the same 2D space, which results in a set of predictions for the pedestrians inside a single bounding box. Common predictions across multiple boxes are eliminated using an algorithm called Non-maximum Suppression. The final refinement module re-estimates the bounding box for each prediction using the underlying features, thus decreasing the probability of false positives.

The CrowdDet model is based on the standard ResNet [4] network pre-trained on Imagenet dataset [21]. To perform pedestrian detection, the whole model is re-trained on CrowdHuman dataset[6], which contains 15,000, 4,370 and 5,000 images for training, validation and test respectively. The model is trained using 8 GPUs for 30 epochs with a batch size of 16 [1].

CenterNet [2] is an object category detector with "person" as one of its category. It uses multiple fully-convolutional encoder-decoder networks to predict a set of key-point heat-maps in a 2D image. A key-point prediction network is then used to categorize the identified key-points. CenterNet identifies the center points of an object using the predicted key-points and generates a bounding box around it. Other properties of the object, such as depth, size and orientation are estimated by the key-point prediction network. The pedestrians prediction is refined by associating the identified key-points to 17 different joints on the body.

The CenterNet model is based on DLA-34 architecture, presented in [5]. It is trained on the MS COCO dataset [7], which contains 118000 training images and over 80 different categories. The model is trained for 230 epochs with a batch size of 128 on 8 TITAN-V GPUs and with a learning rate of 5e-4 [2]. Since the proposed pipeline focuses on pedestrian detection, only the classifications corresponding to the category "person" are considered.

The two deep learning models are used as is, without any re-training or alterations to the hyper-parameters.

3.3. Evaluation method

The evaluation method for the proposed pre-processing pipeline is as follows:

For each deep leaning model considered in section 3.2:

1. Obtain the pedestrian-detection performance of the model in normal and low-light conditions using the OLPB dataset;
2. Obtain histogram equalized images from the OLPB dataset;
3. Obtain the pedestrian-detection performance of the model in normal and low-light conditions using the equalized images;
4. Obtain rectified images using the pre-processing pipeline described in Sec. 2, on the OLPB dataset;
5. Obtain the pedestrian-detection performance of the model in normal and low-light conditions using the rectified images;
6. Compare the performances obtained in steps 1, 3 and 5.

The "performance" is reported in terms of average precision of the models in detecting pedestrians on the OLPB dataset. The averages are computed over an intersection of union (IoU) threshold of 0.5. Since no training is involved, the entire OLPB dataset is used as a test set. Among the available 430 images, 229 are identified as low-light images, while the remaining 201 are identified as normal-light images, after visual inspection. The proposed pipeline doesn't involve setting of any dataset specific parameter values.

3.4. Results

The results from the evaluation are reported in Table 2. It is observed that CrowdDet [1] performs significantly better than CenterNet[2] on the OLPB dataset. This is expected as

CrowdDet [1] is specifically designed for pedestrian detection, while CenterNet [2] is a generic object detector, with "person" as one of its category. However, in either cases the proposed pipeline improves the performance of the models with an average precision of 90.6% and 83.4% respectively. The significance of the proposed pre-processing pipeline is highlighted in the low-light conditions, where the models gain an improvement of 21.1% and 11.5%, in their average precision score. While, in normal-light conditions the performance of the proposed pipeline is equivalent to the RGB image results.

A set of observations can also be made, concerning the execution time of the proposed pipeline. CrowdDet [1] and CenterNet [2], implemented on a 16GB NVIDIA RTX Quadro 5000 GPU with CUDA version 10.0 and CUDNN version 7.6.5, take 0.11 and 0.03 seconds, respectively to process an image. The proposed pipeline implemented using the LRSLibrary [22, 23], on Intel i5-1135G7 (2.40 GHz) CPU can generate a rectified image in 0.05 seconds. Thus, the proposed pipeline can be adopted as computationally inexpensive pre-processing step by any normal-light, deep learning or traditional machine learning model operating on CPUs, to perform pedestrian detection in low light conditions.

4. Conclusion

We present a novel pipeline for pre-processing RGB images, which can then be used as input by any normal-light, deep learning model for pedestrian detection. The pipeline generates rectified images using RPCA, which improve the contrast between the background and the corresponding foreground objects. The rectified images enable the normal-light, deep learning models to perform pedestrian detection in low-light conditions. The following conclusions are made form the evaluation of the proposed pipeline:

- Despite their good performance in normal-light conditions, the deep learning models perform poorly in low-light conditions;
- Equipping the deep learning models with our pre-processing pipeline significantly improves their low-light performance (to levels comparable to their performance in normal-light conditions);

Since the proposed pre-processing pipeline is computationally inexpensive, it can also be used with other traditional machine learning based pedestrian detection systems. The proposed pipeline is most effective when the movement in the scene is limited to foreground objects. The current implementation of RPCA is not robust against a dynamic background. Thus, a possible future direction can involve identifying a robust system for foreground segmentation, or exploring segmentation and clustering in depth images to obtain a foreground mask.

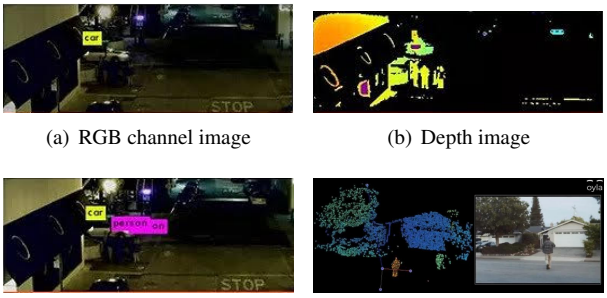
A. Details of the Oyla Camera

The Oyla's 3D-aware surveillance camera resolves previously intractable issues with depth sensing camera, such as variable lighting, occlusion, and estimating the true distance and scale of objects. The camera offers 'LIDAR-like performance at more than 10X lower cost' using commercially available, standard RGB visual camera components coupled with Oyla's new 3D data structure and analysis software when gathering details of the observed scene. The central aspect of the cost savings are Oyla's LIDAR-like technology that uses 'shared optics' with an RGB camera, including the same lensing, but the data collected by the standard image sensor is rendered in 3D (adding light depth data) instead of the standard 2D model with typical cameras. While Oyla uses '80% - 90%' standard camera hardware, the company uses additional 'commercial' electronic components, proprietary firmware, and runs the collected data through its proprietary analytics and software.

Oyla uses its inherently spatially fused 3D "Depth" and video streams to produce enhanced RGB video (e-RGB). e-RGB video can be utilized by existing AI and detection algorithms to achieve greatly increased performance - see Figure 5(a), 5(b) and 5(c). In addition, Oyla uses the 3D data to accurately detect intrusions into user-defined perimeters, as illustrated in Figure 5(d).

Table 2. Performance of normal-light, deep learning models on OLPB dataset with and without the proposed pre-processing pipeline.

Models	**Images**	**Average Precision (IoU $\geq$ 0.5)**		
		normal-light	**low-light**	**All**
CrowdDet[1]	RGB	98.4	61.0	79.3
	Equalized	98.2	58.4	78.9
	(Proposed) Rectified	**98.2**	**82.1**	**90.6**
CenterNet[2]	RGB	90.3	60.7	74.2
	Equalized	94.5	54.4	71.7
	(Proposed) Rectified	**92.2**	**72.2**	**83.4**



(a) RGB channel image (b) Depth image

(c) Enhanced image (d) Point cloud representation

Figure 5. Example of data from Oyla's 3D-aware surveillance camera

Acknowledgements

This work was part of the Industrial Projects program of the Anuradha and Prashanth Palakurthi Centre for Artificial Intelligence Research (APPCAIR). We thank Prof. Ashwin Srinivasan and Tirtharaj Dash for their help with this paper.

References

[1] X. Chu, A. Zheng, X. Zhang, and J. Sun, "Detection in crowded scenes: One proposal, multiple predictions," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12214–12223, 2020.

[2] X. Zhou, D. Wang, and P. Krähenbühl, "Objects as points," *arXiv preprint arXiv:1904.07850*, 2019.

[3] J. Nataprawira, Y. Gu, I. Goncharenko, and S. Kamijo, "Pedestrian detection using multispectral images and a deep neural network," *Sensors (Basel, Switzerland)*, vol. 21, 2021.

[4] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.

[5] F. Yu, D. Wang, and T. Darrell, "Deep layer aggregation," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2403–2412, 2018.

[6] S. Shao, Z. Zhao, B. Li, T. Xiao, G. Yu, X. Zhang, and J. Sun, "Crowdhuman: A benchmark for detecting human in a crowd," *arXiv preprint arXiv:1805.00123*, 2018.

[7] T.-Y. Lin, M. Maire, S. J. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *ECCV*, 2014.

[8] Y. Wang, T. Lu, T. Zhang, and Y. Wu, "Seeing pedestrian in the dark via multi-task feature fusing-sharing learning for imaging sensors," *Sensors (Switzerland)*, vol. 20, 2020.

[9] C. Chen, Q. Chen, M. Do, and V. Koltun, "Seeing motion in the dark," *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 3184–3193, 2019.

[10] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *ArXiv*, vol. abs/1804.02767, 2018.

[11] S. S. S. Kruthiventi, P. Sahay, and R. Biswal, "Low-light pedestrian detection from rgb images using multi-modal knowledge distillation," in *2017 IEEE International Conference on Image Processing (ICIP)*, pp. 4207–4211, 2017.

[12] H.-L. Tang, S.-C. Chien, W.-H. Cheng, Y.-Y. Chen, and K.-L. Hua, "Multi-cue pedestrian detection from 3d point cloud data," in *2017 IEEE international conference on multimedia and expo (ICME)*, pp. 1279–1284, IEEE, 2017.

[13] K. Piniarski, P. Pawłowski, and A. Dabrowski, "Pedestrian detection by video processing in automotive night vision system," in *2014 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA)*, pp. 104–109, IEEE, 2014.

[14] P. Dollár, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: A benchmark," in *CVPR*, 2009.

[15] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The kitti dataset," *International Journal of Robotics Research (IJRR)*, 2013.

[16] J. Uhrig, N. Schneider, L. Schneider, U. Franke, T. Brox, and A. Geiger, "Sparsity invariant cnns," *2017 International Conference on 3D Vision (3DV)*, pp. 11–20, 2017.

[17] Y. Socarrás, S. Ramos, D. Vázquez, A. M. López, and T. Gevers, "Adapting pedestrian detection from synthetic to far infrared images," in *ICCV Workshops*, vol. 3, 2013.

[18] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?," *Journal of the ACM (JACM)*, vol. 58, no. 3, pp. 1–37, 2011.

[19] P. Rodriguez and B. Wohlberg, "Fast principal component pursuit via alternating minimization," in *2013 IEEE International Conference on Image Processing*, pp. 69–73, IEEE, 2013.

[20] M. Everingham, L. Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge," *International Journal of Computer Vision*, vol. 88, pp. 303–338, 2009.

[21] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision (IJCV)*, vol. 115, no. 3, pp. 211–252, 2015.

[22] A. Sobral, T. Bouwmans, and E.-h. Zahzah, "Lrslibrary: Low-rank and sparse tools for background modeling and subtraction in videos," in *Robust Low-Rank and Sparse Matrix Decomposition: Applications in Image and Video Processing*, CRC Press, Taylor and Francis Group., 2015.

[23] T. Bouwmans, A. Sobral, S. Javed, S. K. Jung, and E.-h. Zahzah, "Decomposition into low-rank plus additive matrices for background/foreground separation: A review for a comparative evaluation with a large-scale dataset," *CoRR*, vol. abs/1511.01245, 2015.