DOMAIN-ADAPTIVE PEDESTRIAN DETECTION IN THERMAL IMAGES

Tiantong Guo[*]*, Cong Phuoc Huynh*[†]*, and Mashhour Solh*[†]

Pennsylvania State University[*], Amazon Lab126[†]

ABSTRACT

This paper presents an approach to pedestrian detection in thermal infrared (thermal) images with limited annotations. The key idea is to adapt the abundance of color images associated with bounding box annotations to the thermal domain for training the pedestrian detector. To this end, we couple a domain adaptation component that consists of a pair of image transformers with a pedestrian detector in the thermal domain and train the entire network end-to-end. The image transformers act as a data augmentation tool that progressively improves synthetic examples on the fly for training the pedestrian detector. To aid the training process, we introduce a detection loss defined on both real thermal images and synthetic thermal images transformed from the color domain. The proposed detector outperforms existing methods on the thermal images from the KAIST detection benchmark [1].

Keywords: pedestrian detection, synthetic image, thermal image, and deep learning.

1. INTRODUCTION

Pedestrian detection in color images is an active research problem, for which solutions range from hand-crafted feature design [6, 2] to end-to-end learning [7, 8, 9, 10, 11]. In addition, numerous methods have been studied to tackle difficult aspects of the problem such as object occlusion [12], person scale [13], or challenging illumination condition [14], low image quality [15].

However, the pedestrian detection problem suffers from a performance bottleneck in low-lighting conditions, *e.g.*night time, where images lack contrast. Under such a circumstance, cues for human presence such as body shape, silhouette and clothing textures are subject to the sensitivity of the imaging sensors. On the other hand, long-wave infrared (thermal) sensors, which are designed to measure object temperature, are able to capture clearly visible human bodies in thermal images over a wide range of weather conditions, irrespective of the illumination variations [16]. In Figure 1, we show that pedestrian detection at night time is more accurate in the thermal image (left) than in the color image (right). These detection results are obtained on a test scene from the KAIST dataset [1] by two identical Faster RCNN [17] models trained separately on color and thermal images. When evaluated on the night-time KAIST test set, the color model incurs a log-average miss-rate of 80.56%, which almost doubles that of 44.88% for the thermal counterpart.

Work is conducted during an internship at Amazon Lab126.

Fig. 1: During night time, pedestrians detections (red) are more accurate in the thermal image (right) than the color image (left). Ground truth boxes are shown in green.

Therefore, combining the complementary cues offered by both color and thermal image modalities is a straightforward approach to improve the data quality under poor lighting conditions for detection purposes [18, 19, 20, 21, 22, 23]. However, the deployment of multi-spectral (color and thermal) imaging setups is either expensive (using a special four channel sensor), or cumbersome (requiring the precise alignment and registration of two images) for practical use. Moreover, it has been shown that under poor illumination conditions, the thermal modality is superior to color and fusing both modalities offers no improvements in detection accuracy over thermal images alone [24, 23].

Hence, it is imperative to tackle the pedestrian detection problem using only thermal images for practical applications. This problem is especially important in poor/low lighting conditions, where thermal infrared imaging offers better contrast and human body saliency than color imaging. However, the literature on pedestrian detection using solely thermal images is not as developed as that with color, including only a handful of methods using generic background modeling and segmentation [25, 26, 27], hand-crafted features [28, 29, 24, 30], dictionary learning [31] and fine-tuning on preprocessed data [32].

In this paper, we propose one of the first methods to address the visual detection problem for a domain with limited bounding box annotations. The domain adaptation mechanism makes use of unannotated images in both the source and target domains without requiring image pairs with spatial alignment. We tackle the domain shift between thermal and color images by learning a pair of image transformers (inspired by CycleGAN [33]) to convert images between the two modalities, jointly with a pedestrian detector. The image transformers act as a data augmentation and a domain adaptation component, which progressively refines synthetic examples for training the pedestrian detector.

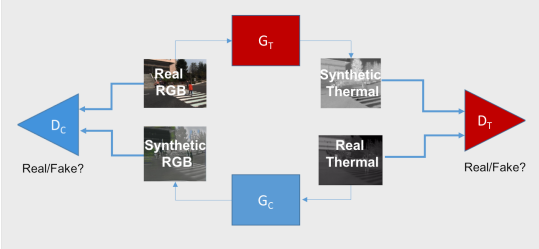

Fig. 2: Training of the domain adapter.

2. TRAINING PEDESTRIAN DETECTION IN THERMAL IMAGES DOMAIN

2.1. High-Level Overview

Suppose that we are given a training set of supervised (labeled) thermal images $\mathcal{S}_T \triangleq \{(\mathbf{x}_T, \mathbf{b}_T)\}$, where $\mathbf{x}_T$ is a thermal image and $\mathbf{b}_T$ is its associated ground truth pedestrian bounding boxes. In addition, we also have access to an unannotated set of thermal images. Our goal is to train a pedestrian detector F_T in the thermal image domain, which maps a novel thermal image $\mathbf{x}_T$ to a set of pedestrian bounding boxes $\mathbf{b}_T$. Let us assume the availability of a large set of annotated color images $\mathcal{S}_C \triangleq \{(\mathbf{x}_C, \mathbf{b}_C)\}$, where $\mathbf{b}_C$ is the set of ground truth bounding boxes in the image $\mathbf{x}_C$. Let $\mathcal{D}_C$ and $\mathcal{D}_T$ be the union of all annotated and unannotated color and thermal images, respectively.

2.2. Adversarial Data Adaptation

The image adaptation component consists of a pair of image generators $G_T : \mathcal{D}_C \rightarrow \mathcal{D}_T$ and $G_C : \mathcal{D}_T \rightarrow \mathcal{D}_C$, that transform color to thermal images and vice versa, respectively. These generators augment realistic images for the detection task, by fooling a pair of color (D_C) and thermal (D_T) image discriminator. At the same time, the discriminators strive to distinguish real from synthetic samples in the respective domain. Figure 2 depicts the architecture of the domain adapter.

Let the generators G_C and G_T be expressed as functions with parameters ϕ_C and ϕ_T. Likewise, the discriminators $D_C : \mathcal{D}_C \rightarrow [0, 1]$ and $D_T : \mathcal{D}_T \rightarrow [0, 1]$, parameterized by θ_C and θ_T, assign a label of 1 to real images and a label of 0 to synthetic ones. While the discriminator is aimed to maximize the cross-entropy loss, the generator's goal is to minimize it in order to confuse to the discriminator. The adversarial training objectives in the color and thermal domain, *i.e.*$\mathcal{L}_C^{adv}$ and $\mathcal{L}_T^{adv}$, respectively, can be formulated as a minimax optimization problem as below.

$$\begin{aligned}\min_{\phi_C,\phi_T}\max_{\theta_C,\theta_T} \mathcal{L}_C^{adv} \triangleq & \mathbb{E}_{\mathbf{x}_T\in\mathcal{D}_T} \log\left(1 - D_C(G_C(\mathbf{x}_T;\phi_C));\theta_C\right) \\ & + \mathbb{E}_{\mathbf{x}_C\in\mathcal{D}_C} \log D_C(\mathbf{x}_C;\theta_C)\end{aligned} \tag{1}$$

$$\begin{aligned}\min_{\phi_C,\phi_T}\max_{\theta_C,\theta_T} \mathcal{L}_T^{adv} \triangleq & \mathbb{E}_{\mathbf{x}_C\in\mathcal{D}_C} \log\left(1 - D_T(G_T(\mathbf{x}_C;\phi_T));\theta_T\right) \\ & + \mathbb{E}_{\mathbf{x}_T\in\mathcal{D}_T} \log D_T(\mathbf{x}_T;\theta_T)\end{aligned} \tag{2}$$

In addition, we enforce a cycle consistency constraint on the pair of generators G_C and G_T, similar to [33]. In other words, a forward transformation G_T of a color image $\mathbf{x}_C$ into the thermal domain by, followed by a backward transformation G_C to the color domain, should produce an image close to the original. Similarly, the successive transformations of a thermal image $\mathbf{x}_C$ into the color domain and then the thermal domain should recover the original thermal image. The cycle consistency losses in both domains are:

$$\begin{aligned}\mathcal{L}_C^{cyc} &\triangleq \mathbb{E}_{\mathbf{x}_C\in\mathcal{D}_C}\|G_C(G_T(\mathbf{x}_C;\phi_T);\phi_C) - \mathbf{x}_C\|_1 \\ \mathcal{L}_T^{cyc} &\triangleq \mathbb{E}_{\mathbf{x}_T\in\mathcal{D}_T}\|G_T(G_C(\mathbf{x}_T;\phi_C);\phi_T) - \mathbf{x}_T\|_1\end{aligned} \tag{3}$$

where $\|\cdot\|_1$ denotes the ℓ^1-norm.

In summary, the total data adaptation loss (from the color to thermal domain and vice versa) is

$$\mathcal{L}^{adapt} \triangleq \mathcal{L}_C^{adv} + \mathcal{L}_T^{adv} + \mathcal{L}_C^{cyc} + \mathcal{L}_T^{cyc} \tag{4}$$

2.3. Pedestrian Detection in Thermal Images

We train the thermal detector by minimizing the average detection loss $\mathcal{L}_T^{det}(F_T(\mathbf{x}_T;\omega_T), \mathbf{b}_T)$ defined per annotated thermal image, where ω_T is the parameters of the detector F_T.

The detection loss over the real thermal images is

$$\mathcal{L}_T^{det}(\mathcal{S}_T) \triangleq \mathbb{E}_{(\mathbf{x}_T,\mathbf{b}_T)\in\mathcal{S}_T}\mathcal{L}_T^{det}(F_T(\mathbf{x}_T;\omega_T), \mathbf{b}_T) \tag{5}$$

Next, we augment the training data for the thermal detector with the set of annotated color images $\mathcal{S}_C$. To commence, we transform the color image $\mathbf{x}_C$ to its pseudo-thermal version $\tilde{\mathbf{x}}_T \triangleq G_T(\mathbf{x}_C;\phi_T)$ using the image transformer G_T. Subsequently, we transfer the set of bounding boxes $\mathbf{b}_C$ associated with the color image $\mathbf{x}_C$ to the pseudo-thermal image $\tilde{\mathbf{x}}_T$. As a result, we obtain a set of synthetic thermal images with associated pedestrian bounding boxes, which is denoted by $\mathcal{P}_T \triangleq \{(\tilde{\mathbf{x}}_T, \mathbf{b}_C)\}$. The detection loss defined on the synthetic thermal images is formulated as

$$\mathcal{L}_T^{det}(\mathcal{P}_T) \triangleq \mathbb{E}_{(\mathbf{x}_C,\mathbf{b}_C)\in\mathcal{S}_C}\mathcal{L}_T^{det}(F_T(\tilde{\mathbf{x}}_T;\omega_T), \mathbf{b}_C), \tag{6}$$

where $\tilde{\mathbf{x}}_T = G_T(\mathbf{x}_C;\phi_T)$.

We feed both real and synthetic thermal images into the thermal pedestrian detector, and train it by minimizing the total detection loss

$$\mathcal{L}_T^{det} = \mathcal{L}_T^{det}(\mathcal{S}_T) + \mathcal{L}_T^{det}(\mathcal{P}_T) \tag{7}$$

2.4. Training Strategies

The first training strategy is a two-stage approach, as depicted in Figure 3. Here, we train the domain adapter (shown in Figure 2) on unpaired color and thermal images (without annotations) for a certain number of epochs. We then combine the synthetic images, together with the bounding boxes transferred from the original images, with the real annotated thermal images to form a mixed dataset for training the detector.

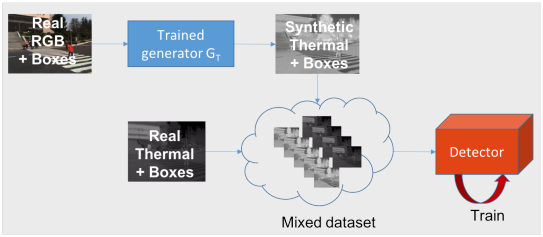

Fig. 3: Training of the detector with synthetic thermal images generated by a trained domain adapter.

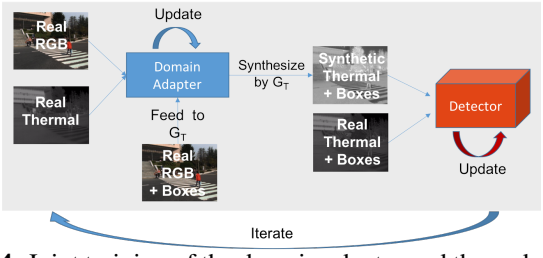

Fig. 4: Joint training of the domain adapter and the pedestrian detector in the thermal infrared domain.

We also explore the joint training of the domain adapter and the pedestrian detector in an iterative manner, as illustrated by Figure 4. The main difference from the two-stage approach is that the detection loss back-propagates its gradients to the color-to-thermal transformer G_T. In addition, synthetic thermal images are generated on the fly, and mixed with real thermal images for the training of the detector.

3. EXPERIMENTS

3.1. Datasets

In our experiment, we focus on boosting the detection accuracy on the thermal images from the KAIST dataset [1] which contains 12 video sequences captured under various illumination conditions at both day time and night time. The captured frames have a resolution of 512×640 with $103,128$ bounding boxes of people labeled as *person*, *person?* (person with uncertainty), *people* (group of people) or *cyclist*. We demonstrate our method by taking the Caltech dataset [2] as the source domain and adapt its large number of annotated color images for training purposes. We report the log-average miss rate (MR) across different False Positives Per Image (FPPI) in the range of $[10^{-2}, 10^{1}]$, with *person* as the foreground class.

To train the pedestrian detector, we follow the process described by [18] to obtain 7668 training thermal images, (*i.e.*$|\mathcal{S}_T| = 7668$) and 2252 test thermal images with clean annotations from the KAIST dataset [1]. In addition, we take $122,187$ color images from the Caltech dataset [2] with at least an annotated pedestrian bounding box as training data (*i.e.*$|\mathcal{S}_C| = 122,187$), and transform them into thermal images as additional data for training the detector.

Color Real Thermal Synth. Thermal

Fig. 5: Synthetic thermal image generated from color images in the KAIST test set.

Color Synth. Thermal

Fig. 6: Sample synthetic thermal image transformed from the Caltech dataset.

3.2. Sample Synthetic Thermal Images

In Figure 5, we show synthetic thermal images (in bottom row) generated from color images in the KAIST test set (top row), in comparison with the true synthetic images (middle row). In the synthetic images, there is sufficient contrast between the pedestrians and the background for a human observer to easily detect the salient features of human bodies. Although the synthetic images appear brighter than the ground truth, our analysis shows that they differ by a scaling factor. Further, this difference between synthetic and true images can be absorbed by a normalization operation during the preprocessing of training data for the detection task. The same trend is also observed in the Caltech dataset (Figure 6) when we transform color images from this dataset to the thermal modality.

3.3. Comparison with existing methods

We name the proposed model trained in two stages as *VGG16-two-stage* and the one obtained by the joint training of the domain adapter and the detector as *VGG16-joint*. We compare these models with a vanilla Faster-RCNN detector, *i.e. VGG16-thermal*, that has been trained on thermal images with a VGG-16 backbone. In addition, we considered prior pedestrian detectors in the thermal domain [30, 32, 24] and several variants of ACF [2] with different hand-crafted features [1], a domain-invariant detector [34], and a multi-task network for semantic segmentation and pedestrian detection [14].

In Figure 7, we plot the miss rate (MR) versus false positives per-image (FPPI) in the log-log scale (lower curves imply better accuracy). The miss rates for various methods are also reported in Table 1, where results for the methods in [30, 32, 24] were taken directly from the respective papers.

Both proposed methods, *i.e.*VGG16-joint and VGG16-

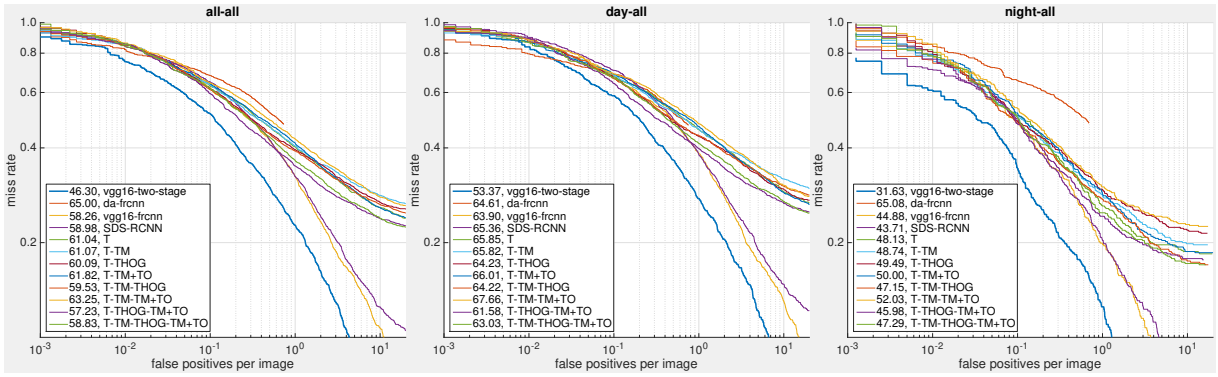

Fig. 7: The miss rate versus false positives per-image (MR vs. FPPI) plots for the proposed and prior methods on the *Reasonable* test set. Left to right panels: plots for all test images, day-time only, and night-time only subsets.

Table 1: Performance of the proposed models (VGG16-joint and VGG16-two-stage) and state-of-the-art methods, in terms of the log-average miss rate.

	overall	day	night
ACF-T	61.04	65.85	48.13
ACF-T-TM	61.07	65.82	48.74
ACF-T-HOG	60.09	64.23	49.49
ACF-T-TM+TO	61.82	66.01	50.00
ACF-T-TM-THOG	59.53	64.22	47.15
ACF-T-TM-TM+TO	63.25	67.66	52.03
ACF-T-THOG-TM+TO	57.23	61.58	45.98
ACF-T-TM-THOG-TM+TO	58.83	63.03	47.29
Herrmann *et al.* [32]	69.81	-	-
TIIHOG-IKSVM-cell2 [30]	-	-	56.85
HOG-LBP+RF [24]	-	65.70	53.50
VGG16-thermal	58.26	63.90	44.88
VGG16-rgb-thermal	57.88	59.36	42.13
DA-FRCNN [34]	59.98	60.00	60.05
SDS-RCNN [14]	58.98	65.36	43.71
VGG16-joint	51.09	56.81	37.18
VGG16-two-stage	**46.30**	**53.37**	**31.63**
ResNet50-two-stage	43.43	51.25	27.04
ResNet101-two-stage	**42.65**	**49.59**	**26.70**

two-stage, consistently outperform prior methods across both day-time and night-time subsets and overall. The improvement of miss rate is more pronounced in the night time images than the day time subset. For night images, VGG16-two-stage achieves a reduction of more than 12% compared to the best baseline, *i.e.*SDS-RCNN (with an MR of 43.71%), whereas across day images, its lead over the best baseline (ACF-T-THOG-TM+TO feature with an MR of 61.58 %) is only over 8%. The best overall baseline, *i.e.*ACF-T-THOG-TM+TO trails behind the VGG16-two-stage model by nearly 11%. Further, the miss rate for the two-stage approach decreases with a deeper backbone network, reaching 43.43% with Resnet-50 and 42.65% with ResNet-101. Figure 8 qualitatively illustrates the detection output on a sample image where ground truth boxes are plotted in green, detections are in red boxes with confidence scores.

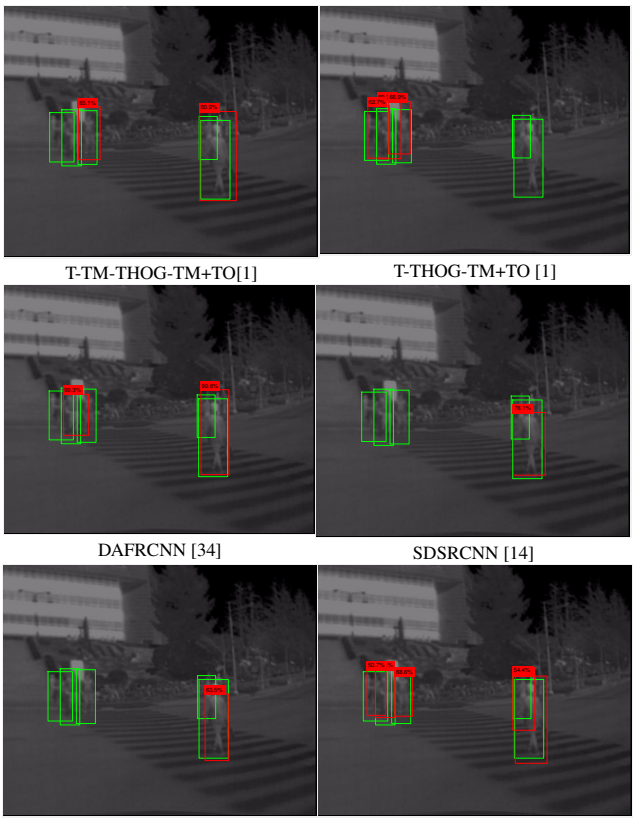

Fig. 8: Detection results obtained by the proposed VGG16-two-stage approach in comparisons to previous methods.

4. CONCLUSION

Pedestrian detection in thermal images is challenging problem because of the scarcity of training data. We have proposed an image-level domain adaptation method that harnesses an abundant amount of data from the color domain to push the performance envelop beyond state-of-the-art methods. The proposed method delivers promising results, reducing the log-average miss rate by more than 12%.

5. REFERENCES

[1] S. Hwang, J. Park, N. Kim, Y. Choi, and I. So Kweon, "Multispectral pedestrian detection: Benchmark dataset and baseline," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1037–1045.

[2] P. Dollár, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: A benchmark," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 2009, pp. 304–311.

[3] J. W. Davis and V. Sharma, "Background-subtraction using contour-based fusion of thermal and visible imagery," *Comput. Vis. Image Underst.*, vol. 106, no. 2-3, pp. 162–182, May 2007.

[4] A. Ess, B. Leibe, K. Schindler, and L. van Gool, "A mobile vision system for robust multi-person tracking," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'08)*, June 2008.

[5] A. Torabi, G. Massé, and G.-A. Bilodeau, "An iterative integrated framework for thermal-visible image registration, sensor fusion, and people tracking for video surveillance applications," *Comput. Vis. Image Underst.*, vol. 116, no. 2, pp. 210–221, Feb. 2012.

[6] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. IEEE, 2005, vol. 1, pp. 886–893.

[7] D. Xu, W. Ouyang, E. Ricci, X. Wang, and N. Sebe, "Learning cross-modal deep representations for robust pedestrian detection," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[8] L. Zhang, L. Lin, X. Liang, and K. He, "Is faster r-cnn doing well for pedestrian detection?," in *Computer Vision – ECCV 2016*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds., Cham, 2016, pp. 443–457, Springer International Publishing.

[9] S. Zhang, R. Benenson, and B. Schiele, "Filtered channel features for pedestrian detection," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015, pp. 1751–1760.

[10] S. Wang, J. Cheng, H. Liu, and M. Tang, "Pcn: Part and context information for pedestrian detection with cnns," in *BMVC*, 2017.

[11] W. Ouyang, H. Zhou, H. Li, Q. Li, J. Yan, and X. Wang, "Jointly learning deep features, deformable parts, occlusion and classification for pedestrian detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 8, pp. 1874–1887, Aug 2018.

[12] S. Zhang, J. Yang, and B. Schiele, "Occluded pedestrian detection through guided attention in cnns," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[13] D. X. H. S. S. P. Tao Song, Leiyu Sun, "Small-scale pedestrian detection based on topological line localization and temporal feature aggregation," in *ECCV*, 2018.

[14] G. Brazil, X. Yin, and X. Liu, "Illuminating pedestrians via simultaneous detection & segmentation," in *ICCV*, 2017.

[15] G. W. J. Z. C. Lin, L. Jiwen, "Graininess-aware deep feature learning for pedestrian detection," in *ECCV*, 2018.

[16] N. Negied, E. Hemayed, and M. Fayek, "Pedestrians detection in thermal bands–critical survey," vol. 2, pp. 141–148, 09 2015.

[17] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in Neural Information Processing Systems 28*, C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, Eds., pp. 91–99. Curran Associates, Inc., 2015.

[18] J. Liu, S. Zhang, S. Wang, and D. Metaxas, "Multispectral deep neural networks for pedestrian detection," in *Proceedings of the British Machine Vision Conference (BMVC)*, September 2016, pp. 73.1–73.13.

[19] C. Li, D. Song, R. Tong, and M. Tang, "Multispectral pedestrian detection via simultaneous detection and segmentation," 2018.

[20] D. Guan, Y. Cao, J. Liang, Y. Cao, and M. Y. Yang, "Fusion of multispectral data through illumination-aware deep neural networks for pedestrian detection," *arXiv preprint arXiv:1802.09972*, 2018.

[21] K. Park, S. Kim, and K. Sohn, "Unified multi-spectral pedestrian detection based on probabilistic fusion networks," *Pattern Recognition*, vol. 80, pp. 143 – 155, 2018.

[22] D. Konig, M. Adam, C. Jarvers, G. Layher, H. Neumann, and M. Teutsch, "Fully convolutional region proposal networks for multispectral person detection," in *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, July 2017, pp. 243–250.

[23] C. Li, D. Song, R. Tong, and M. Tang, "Illumination-aware faster r-cnn for robust multispectral pedestrian detection," *Pattern Recognition*, vol. 85, pp. 161 – 171, 2019.

[24] A. González, Z. Fang, Y. S. Salas, J. Serrat, D. Vázquez, J. Xu, and A. M. López, "Pedestrian detection at day/night time with visible and FIR cameras: A comparison," *Sensors*, vol. 16, no. 6, pp. 820, 2016.

[25] J. W. Davis and M. A. Keck, "A two-stage template approach to person detection in thermal imagery," in *2005 Seventh IEEE Workshops on Applications of Computer Vision (WACV/MOTION'05) - Volume 1*, Jan 2005, vol. 1, pp. 364–369.

[26] C. Dai, Y. Zheng, and X. Li, "Layered representation for pedestrian detection and tracking in infrared imagery," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops*, Sept 2005, pp. 13–13.

[27] A. Leykin and R. Hammoud, "Robust multi-pedestrian tracking in thermal-visible surveillance videos," in *2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06)*, June 2006, pp. 136–136.

[28] L. Zhang, B. Wu, and R. Nevatia, "Pedestrian detection in infrared images based on local shape features," in *2007 IEEE Conference on Computer Vision and Pattern Recognition*, June 2007, pp. 1–8.

[29] K. J. M. Arens, "Feature based person detection beyond the visible spectrum," in *IEEE CVPR Workshops*, 2009.

[30] J. Baek, S. Hong, J. Kim, and E. Kim, "Efficient pedestrian detection at nighttime using a thermal camera," *Sensors*, vol. 17, pp. 1850, 2017.

[31] B. Qi, V. John, and S. M. Z. Liu, "Use of sparse representation for pedestrian detection in thermal images," in *IEEE CVPR Workshops*, 2014.

[32] C. Herrmann, M. Ruf, and J. Beyerer, "Cnn-based thermal infrared person detection by domain adaptation," in *Autonomous Systems: Sensors, Vehicles, Security, and the Internet of Everything*. International Society for Optics and Photonics, 2018, vol. 10643, p. 1064308.

[33] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Computer Vision (ICCV), 2017 IEEE International Conference on*, 2017.

[34] Y. Chen, W. Li, C. Sakaridis, D. Dai, and L. Van Gool, "Domain adaptive faster r-cnn for object detection in the wild," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.