

Robust Small-scale Pedestrian Detection with Cued Recall via Memory Learning

Jung Uk Kim* Sungjune Park* Yong Man Ro[†]
Image and Video Systems Lab, KAIST, South Korea
{jukim0701, sungjune-p, ymro}@kaist.ac.kr

Abstract

Although the visual appearances of small-scale objects are not well observed, humans can recognize them by associating the visual cues of small objects from their memorized appearance. It is called cued recall. In this paper, motivated by the memory process of humans, we introduce a novel pedestrian detection framework that imitates cued recall in detecting small-scale pedestrians. We propose a large-scale embedding learning with the large-scale pedestrian recalling memory (LPR Memory). The purpose of the proposed large-scale embedding learning is to memorize and recall the large-scale pedestrian appearance via the LPR Memory. To this end, we employ the large-scale pedestrian exemplar set, so that, the LPR Memory can recall the information of the large-scale pedestrians from the small-scale pedestrians. Comprehensive quantitative and qualitative experimental results validate the effectiveness of the proposed framework with the LPR Memory.

1. Introduction

Pedestrian detection is one of the important research topics in the computer vision field [21, 28, 42, 60]. It has attracted considerable attention in real-world applications such as video surveillance [3, 56] and autonomous driving systems [12, 23]. To detect pedestrians, various modalities have been utilized. The most commonly used modality is the visible modality (e.g. RGB) which is intuitive to human eyes [13]. Also, the thermal modality has been adopted for pedestrian detection recently [17, 24, 25]. The thermal modality has the advantage that it is robust to illumination variation [32] and weather conditions [22].

Despite the strengths of both modalities, pedestrian detection still has a problem in detecting small-scale pedestrians [41, 49]. Although some efforts have been made to mitigate the small-scale issue [29, 53], small-scale pedestrian detection has an inherent problem. As shown in the top of Figure 1, different from the large-scale pedestrian, the visual appearance of the small-scale pedestrian is often

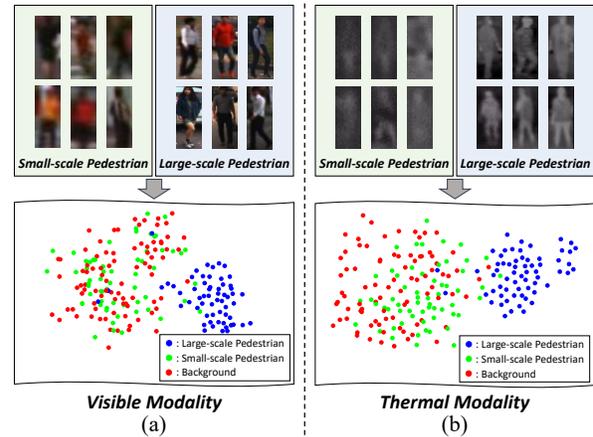


Figure 1. Comparison between the small- and large-scale pedestrians (top) and t-SNE visualization of their latent features (bottom) in (a) visible and (b) thermal modalities. Due to the lack of visual information in the small-scale pedestrians, their latent features are hard to be separated from the background.

blurred and obscure in both modalities (we divide the small- and large-scales according to [22]). We further visualize the small- and large-scale pedestrians with the background in the feature space using t-SNE [51]. As shown in the bottom of Figure 1, the large-scale pedestrian (blue) and the background (red) could be separated from each other easily. However, feature distributions of the small-scale pedestrian (green) and the background (red) are not distinguishable, so that, it is difficult to discriminate between the small-scale pedestrians and the background. Since the visual appearance of the small-scale pedestrians is insufficient, the deep networks are more likely to miss the small-scale pedestrians [41, 49, 55].

To tackle the above problem in small-scale pedestrian detection, we consider how humans distinguish the small-scale pedestrians. As observed in Figure 1, the deep networks are likely to miss the small-scale pedestrians, because they have a lack of visual information. Nevertheless, humans are capable of identifying small-scale pedestrians, even with insufficient visual information. In cognitive psychology, humans recognize small-scale objects by recalling large-scale ones from their memories based on cues ob-

*Both authors have contributed equally to this work.

[†]Corresponding author

served in the small-scale objects [5, 38]. It is called *cued recall*. In this way, although the visual information of small-scale pedestrians is insufficient, humans can recognize them as pedestrians based on their visual cues (e.g. head, body, etc.).

Based on our motivation, we propose a novel pedestrian detection framework to address the small-scale issue, by mimicking how humans identify the small-scale pedestrian through cued recall. To this end, we design a Large-scale Pedestrian Recalling (LPR) Memory to memorize and recall the visual appearance of the large-scale pedestrians even with the insufficient small-scale pedestrian appearances. For the LPR Memory to memorize and recall the prior knowledge about the large-scale pedestrians effectively, we devise a large-scale embedding learning. Through the large-scale embedding learning, the proposed LPR Memory can effectively cope with the small-scale pedestrians by addressing the relevant information of the large-scale pedestrians, playing a role of cued recall of humans. The comprehensive experiments and visualization results with each modality (thermal and visible) validate the effectiveness of the LPR Memory.

To sum up, the major contributions of this paper are summarized as follows:

- Motivated by the process of *cued recall*, we devise the LPR Memory to resolve the inherent small-scale pedestrian detection problem. To the best of our knowledge, it is the first attempt to handle small-scale pedestrian detection with cued recall of the memory.
- In order to guide the LPR Memory to memorize and recall the appearance of the large-scale pedestrians from the small-scale pedestrians, we devise a large-scale embedding learning. Consequently, our detection framework with the LPR Memory can perform robust small-scale pedestrian detection.

2. Related Works

2.1. Pedestrian Detection in Visible Modality

Since the pedestrian detection is closely related to the human life, it has attracted increasing attention [21, 60]. In general, pedestrian detection frameworks adopt the visible modality (e.g. RGB). Along with the various hand-craft methods [2, 9, 11], deep learning methods have been introduced recently for robust pedestrian detection [41, 42, 49, 55]. Many deep learning approaches have been proposed to handle various problems, such as occlusion [35, 40, 57, 58] and scale variation [33, 53, 59].

2.2. Pedestrian Detection in Thermal Modality

Recently, thermal-based pedestrian detection has been adopted actively [14, 17, 24, 25], because it is known to be

robust against irregular illumination [32] and weather condition [22]. Guo et al. [17] adopted the generative adversarial networks (GANs) to generate synthetic thermal images from the visible images. Then, both real and synthetic thermal images are utilized to enrich the visual feature of the pedestrian. In [24], the domain-adaptation method was utilized for the thermal-based pedestrian detection, while taking advantage of the visible modality. Kieu et al. [25] introduced a task-conditioned domain adaptation method with the auxiliary network. These methods have in common that they utilize the visible modal information to improve the visual feature of the thermal modality [25].

2.3. Small-scale Object Detection

Detecting small-scale objects is an essential issue in the object detection task [8, 27, 29, 34, 39, 45]. In [29], the scale normalization method was proposed to resolve the scale variation problem by mapping the features of various scales to the scale-invariant subspace. Noh et al. [39] performed adversarial training with GAN to obtain the super-resolved feature maps of the small-scale objects. Kim et al. [27] proposed a class uncertainty-aware (CUA) loss to guide the detection framework to focus on the small-scale objects.

When the small-scale problem is extended to the pedestrian detection task, RPN+BF [55] was proposed to replace the classifier of the Faster R-CNN [46] with a boosted forest (BF) for aggregating feature maps. In MS-CNN [7], the Faster R-CNN is employed for multi-scale detection networks to detect the small-scale pedestrians. In [53], the self-mimic learning (SML) method was proposed to reduce the intra-class feature variance by mimicking the feature maps of the large-scale pedestrians. Although this method tries to learn the feature maps of the large-scale pedestrians for the small-scale pedestrians, there is no explicit guidance in the inference time. In contrast, our LPR Memory can provide explicit guidance to recall the appearance of the large-scale pedestrians, motivated by *cued recall*.

2.4. Memory Network

Recently, memory augmented neural networks have been introduced in various computer vision fields [6, 10, 15, 18, 30, 31, 37, 43, 50, 52, 61]. For example, MeGA-CDA [52] was designed to guide the category-specific attention map in domain adaptive object detection. Deng et al. [10] proposed a long-term memory for video object detection to memorize the various appearance of objects. Among various memory architectures, the key-value memory has been adopted for question answering and trajectory prediction [4, 37]. In this paper, for small-scale pedestrian detection, we introduce a novel LPR Memory with a large-scale embedding learning based on the key-value memory architecture. Then, the LPR Memory can memorize and recall the large-scale pedestrian context from small-scale pedestrian features.

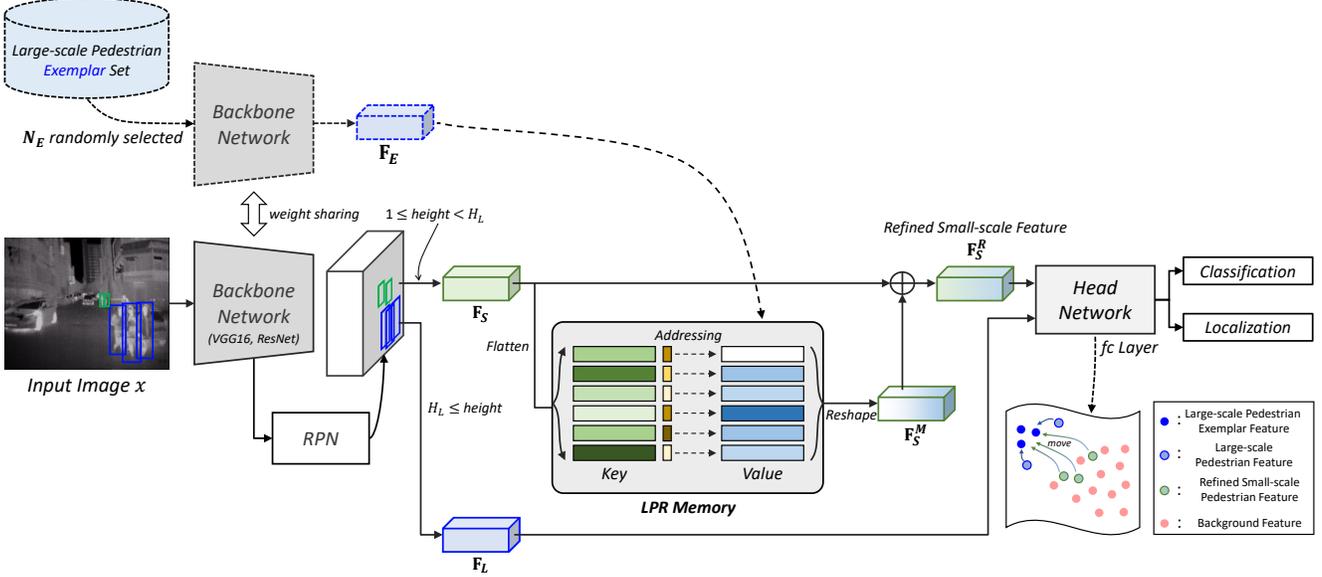


Figure 2. Network configuration of the proposed framework. Dotted line indicates the path used in the training phase only. \oplus denotes the concatenation and 1×1 convolution operations. We use the large-scale pedestrian exemplar set \mathbf{E} to train the LPR Memory to memorize and recall large-scale pedestrian information.

3. Proposed Method

3.1. Overall Architecture

The overall architecture of the proposed pedestrian detection framework is shown in Figure 2. A backbone network (e.g. VGG16 [47] or ResNet [19]) receives an input image x (thermal image example) to encode the image feature map. Next, a Region Proposal Network (RPN) estimates N object candidate regions, called Region of Interests (RoIs). Based on the RoIs, RoI Align [27] is conducted to extract RoI features $\mathbf{F} = \{\mathbf{f}_i\}_{i=1}^N$, $\mathbf{f}_i \in \mathbb{R}^{1 \times w \times h \times c}$ (w , h , and c denote width, height, and channel, respectively).

Then, we divide \mathbf{F} into small- and large-scale RoI features depending on their estimated RoI height. Let $\mathbf{F}_L = \{\mathbf{f}_{L_i}\}_{i=1}^{N_L}$, $\mathbf{f}_{L_i} \in \mathbb{R}^{1 \times w \times h \times c}$ denote N_L large-scale RoI features ($H_L \leq \text{height}$) and $\mathbf{F}_S = \{\mathbf{f}_{S_i}\}_{i=1}^{N_S}$, $\mathbf{f}_{S_i} \in \mathbb{R}^{1 \times w \times h \times c}$ denote N_S small-scale RoI features ($1 \leq \text{height} < H_L$). Note that, $N = N_S + N_L$. The goal is to recall a large-scale pedestrian feature via the Large-scale Pedestrian Recalling (LPR) Memory, when a small-scale pedestrian feature is given. Therefore, \mathbf{F}_S is passed through the LPR Memory to encode \mathbf{F}_S^M . Next, \mathbf{F}_S^M is concatenated with \mathbf{F}_S , and an 1×1 convolution is conducted in order to generate the refined small-scale RoI features \mathbf{F}_S^R . Finally, the head network takes \mathbf{F}_S^R and \mathbf{F}_L to perform the classification and localization. Note that, the head network is same as [36, 46].

Moreover, we devise a large-scale pedestrian exemplar set $\mathbf{E} = \{E_1, \dots, E_K\}$ to memorize the information of the large-scale pedestrian in the LPR Memory. \mathbf{E} is composed of K large-scale pedestrian images which are cropped from the training images, and each large-scale pedestrian sam-

ple in \mathbf{E} has a larger height than H_L based on its ground-truth bounding box annotation. Therefore, we could encode well-aligned large-scale pedestrian exemplar features $\mathbf{F}_E = \{\mathbf{f}_{E_i}\}_{i=1}^{N_E}$, $\mathbf{f}_{E_i} \in \mathbb{R}^{1 \times w \times h \times c}$. Note that, during the training time, we randomly sample N_E large-scale pedestrian exemplars among the K number of exemplars at every iteration. By using \mathbf{E} , we can train the LPR Memory to memorize and recall the appearance of the large-scale pedestrian. More details are explained in the following subsections.

3.2. LPR Memory

Figure 3 shows the structure of the LPR Memory. We define $\mathbf{M} = \{\mathbf{M}_K, \mathbf{M}_V\}$ as key-value memory, and each of them contains L slot pairs ($\mathbf{M}_K, \mathbf{M}_V \in \mathbb{R}^{L \times whc}$). The principle of the LPR Memory is to learn the controller that can dynamically access the relevant information (i.e. large-scale pedestrian) given the small-scale pedestrian input features.

In detail, the i -th small-scale RoI feature \mathbf{f}_{S_i} is flattened to form a vector $\hat{\mathbf{f}}_{S_i} \in \mathbb{R}^{1 \times whc}$. Then, the cosine similarity between $\hat{\mathbf{f}}_{S_i}$ and each slot of \mathbf{M}_K is measured to obtain the similarity vector $\mathbf{s}_i = \{s_{i1}, \dots, s_{iL}\} \in \mathbb{R}^{1 \times L}$. The j -th element of \mathbf{s}_i is calculated as follows:

$$s_{ij} = \frac{\hat{\mathbf{f}}_{S_i} \cdot \mathbf{M}_{K_j}^\top}{\|\hat{\mathbf{f}}_{S_i}\|_2 \|\mathbf{M}_{K_j}\|_2}. \quad (1)$$

With the similarity vector \mathbf{s}_i , we obtain the addressing vector $\mathbf{a}_i = \{a_{i1}, \dots, a_{iL}\} \in \mathbb{R}^{1 \times L}$ by normalizing \mathbf{s}_i via the

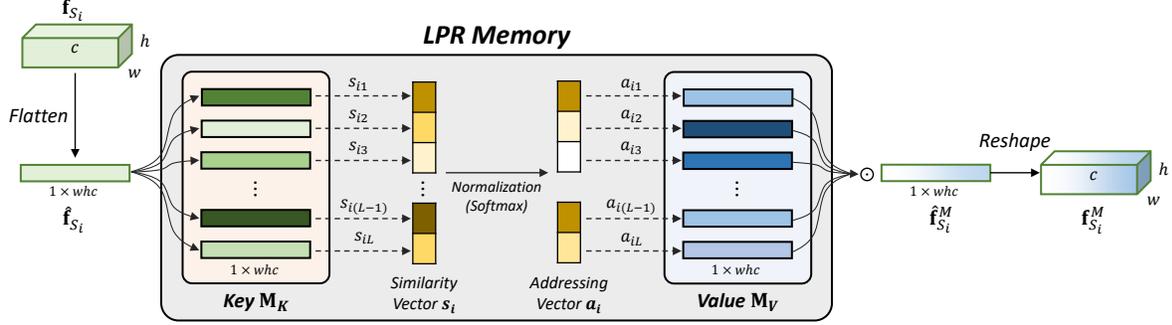


Figure 3. The key-value structure of the LPR Memory. Given $\hat{\mathbf{f}}_{S_i}$, the key memory \mathbf{M}_K calculates similarity to control the amount of relevant information of the value memory \mathbf{M}_V . Based on the addressing vector \mathbf{a}_i , $\hat{\mathbf{f}}_{S_i}^M$ is obtained by a weight summation with \mathbf{M}_V .

softmax. The j -th element of \mathbf{a}_i is represented as:

$$a_{ij} = \frac{e^{s_{ij}}}{\sum_{j=1}^L e^{s_{ij}}}. \quad (2)$$

The addressing vector \mathbf{a}_i plays a role in determining how much to read the relevant slots of \mathbf{M}_V in order to retrieve the stored information of the large-scale pedestrian. Therefore, the output feature of the LPR Memory $\hat{\mathbf{f}}_{S_i}^M \in \mathbb{R}^{1 \times whc}$ is generated by aggregating \mathbf{M}_V by using \mathbf{a}_i , which can be represented as follows:

$$\hat{\mathbf{f}}_{S_i}^M = \sum_{j=1}^L a_{ij} \cdot \mathbf{M}_{V_j}. \quad (3)$$

To the next, $\hat{\mathbf{f}}_{S_i}^M$ is reshaped to form $\mathbf{f}_{S_i}^M \in \mathbb{R}^{1 \times w \times h \times c}$. Finally, $\mathbf{f}_{S_i}^M$ is concatenated with \mathbf{f}_{S_i} , and 1×1 convolution is performed to generate the i -th refined small-scale RoI feature $\mathbf{f}_{S_i}^R \in \mathbb{R}^{1 \times w \times h \times c}$.

The LPR Memory aims to refine the small-scale pedestrian features by recalling the visual appearance of the large-scale pedestrian which is well-distinguished from the background. To this end, we present a large-scale embedding learning to train the LPR Memory to memorize and recall the large-scale pedestrian appearance. It is described in the following subsections.

3.3. Large-scale Embedding Learning

3.3.1 Large-scale Pedestrian Memorizing Loss

As mentioned in 3.1, in order to enable \mathbf{M}_V to memorize and recall the large-scale pedestrian information, we devise a large-scale pedestrian exemplar set $\mathbf{E} = \{E_1, \dots, E_K\}$, which consists of K number of the cropped large-scale pedestrian images in the training set. Note that, the size of exemplar images is set by $E_i \in \mathbb{R}^{32w \times 32h \times 3}$, because the output feature map F_{E_k} is down-sized to $1/32$.

During the training phase, among the large-scale pedestrian exemplar set, N_E large-scale pedestrian exemplars are randomly sampled. The backbone network encodes the N_E exemplar images to generate \mathbf{F}_E . And then, it is flattened

$\hat{\mathbf{F}}_E = \{\hat{\mathbf{f}}_{E_i}\}_{i=1}^{N_E}$, $\hat{\mathbf{f}}_{E_i} \in \mathbb{R}^{1 \times whc}$ to measure the cosine similarity with \mathbf{M}_V . Similar with Eq. (1), similarity between $\hat{\mathbf{f}}_{E_i}$ and each slot of \mathbf{M}_V is measured:

$$s_{E_{ij}} = \frac{\hat{\mathbf{f}}_{E_i} \cdot \mathbf{M}_{V_j}^\top}{\|\hat{\mathbf{f}}_{E_i}\|_2 \|\mathbf{M}_{V_j}\|_2}, \quad (4)$$

where $\mathbf{s}_{E_i} = \{s_{E_{i1}}, \dots, s_{E_{iL}}\} \in \mathbb{R}^{1 \times L}$. \mathbf{s}_{E_i} is normalized by the softmax to obtain $\mathbf{a}_{E_i} = \{a_{E_{i1}}, \dots, a_{E_{iL}}\} \in \mathbb{R}^{1 \times L}$, similar with Eq. (2). The output feature $\hat{\mathbf{f}}_{S_{E_i}}^M \in \mathbb{R}^{1 \times whc}$ is obtained, which is represented as:

$$\hat{\mathbf{f}}_{S_{E_i}}^M = \sum_{j=1}^L a_{E_{ij}} \cdot \mathbf{M}_{V_j}. \quad (5)$$

Finally, $\hat{\mathbf{f}}_{S_{E_i}}^M$ is reshaped into $\mathbf{f}_{S_{E_i}}^M \in \mathbb{R}^{1 \times w \times h \times c}$.

In order to guarantee that \mathbf{M}_V can memorize the large-scale pedestrian information, we propose a large-scale pedestrian memorizing loss \mathcal{L}_{Mem}^M to guide $\mathbf{f}_{S_{E_i}}^M$ to resemble \mathbf{f}_{E_i} . It is represented as follows:

$$\mathcal{L}_{Mem}^M = \frac{1}{N_E} \sum_{i=1}^{N_E} \|\mathbf{f}_{S_{E_i}}^M - \mathbf{f}_{E_i}\|_2^2. \quad (6)$$

Due to the \mathbf{E} and \mathcal{L}_{Mem}^M , the information of the large-scale pedestrian can be embedded to \mathbf{M}_V .

3.3.2 Large-scale Pedestrian Recalling Loss

For the proposed framework to recall large-scale pedestrian, we guide the pedestrian features of \mathbf{F}_S^R to be similar with \mathbf{F}_E in the feature space. Also, since the pedestrian features of \mathbf{F}_L might not contain well-aligned pedestrians, they are also guided to resemble \mathbf{F}_E . So that, the proposed framework can perform more robust large-scale pedestrian detection.

In specific, N number of \mathbf{F}_S^R and \mathbf{F}_L are fed into the head network, as shown in Figure 2, to estimate the latent features $\mathbf{L}_S^R \in \mathbb{R}^{N_S \times D}$ and $\mathbf{L}_L \in \mathbb{R}^{N_L \times D}$, where D denotes the dimension of the latent features. Total N latent features are $\mathbf{L}_{Tot} = [\mathbf{L}_S^R; \mathbf{L}_L] \in \mathbb{R}^{N \times D}$. In addition, the head network takes \mathbf{F}_E to encode $\mathbf{L}_E \in \mathbb{R}^{N_E \times D}$.

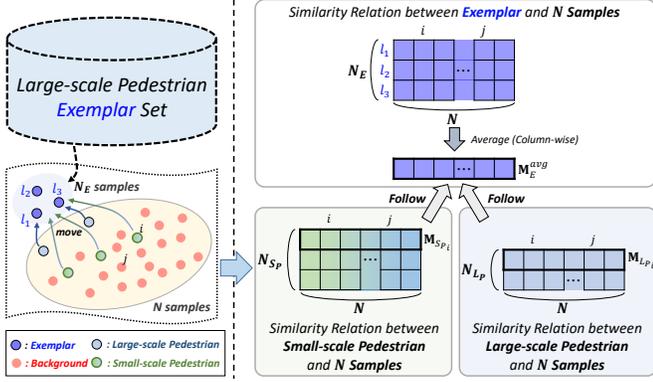


Figure 4. Explanation of the large-scale pedestrian recalling loss. The loss aims to guide the small- and large-scale pedestrian features to be similar with the large-scale pedestrian exemplar features. In this example, $N_E = 3$, $N_{S_P} = 3$, and $N_{L_P} = 2$.

With these latent features, as shown in Figure 4, we first measure the similarity relation between \mathbf{L}_E and \mathbf{L}_{Tot} , that is, $\mathbf{C}_E = \mathbf{L}_E \mathbf{L}_{Tot}^T \in \mathbb{R}^{N_E \times N}$. Then, we average \mathbf{C}_E along the column direction, and then, normalize it via the softmax to obtain $\mathbf{M}_E^{avg} \in \mathbb{R}^{1 \times N}$. Each element of \mathbf{M}_E^{avg} indicates that the probability of how \mathbf{L}_E is similar with \mathbf{L}_{Tot} which includes the pedestrian and background features. Ideally, the pedestrian latent feature would be similar with \mathbf{L}_E and have a high probability, and vice versa to the background.

After that, we also measure the similarity relations by using the pedestrian features of \mathbf{L}_S^R and \mathbf{L}_L . Suppose the numbers of the pedestrian features in each \mathbf{L}_S^R and \mathbf{L}_L are N_{S_P} and N_{L_P} , so that, the pedestrian features of \mathbf{L}_S^R and \mathbf{L}_L can be represented by $\mathbf{L}_{S_P}^R \in \mathbb{R}^{N_{S_P} \times D}$ and $\mathbf{L}_{L_P} \in \mathbb{R}^{N_{L_P} \times D}$, respectively. Then, we could acquire two similarity relations (1) $\mathbf{C}_{S_P} = \mathbf{L}_{S_P}^R \mathbf{L}_{Tot}^T \in \mathbb{R}^{N_{S_P} \times N}$ and (2) $\mathbf{C}_{L_P} = \mathbf{L}_{L_P} \mathbf{L}_{Tot}^T \in \mathbb{R}^{N_{L_P} \times N}$. We normalize each \mathbf{C}_{S_P} and \mathbf{C}_{L_P} via the softmax along the row direction to obtain $\mathbf{M}_{S_P} \in \mathbb{R}^{N_{S_P} \times N}$ and $\mathbf{M}_{L_P} \in \mathbb{R}^{N_{L_P} \times N}$.

Finally, we exploit the KL divergence $D_{KL}(\cdot)$ to measure and minimize the dissimilarity between two probability distributions, so that, it could guide the pedestrian features to be similar with the large-scale pedestrian exemplar features. It is named large-scale pedestrian recalling loss \mathcal{L}_{Rec}^M and can be represented as follows:

$$\mathcal{L}_{Rec}^M = \frac{1}{N_{S_P}} \sum_{i=1}^{N_{S_P}} \underbrace{D_{KL}(\mathbf{M}_E^{avg} || \mathbf{M}_{S_P_i})}_{\text{small-scale to exemplar}} + \frac{1}{N_{L_P}} \sum_{i=1}^{N_{L_P}} \underbrace{D_{KL}(\mathbf{M}_E^{avg} || \mathbf{M}_{L_P_i})}_{\text{large-scale to exemplar}}, \quad (7)$$

where $\mathbf{M}_{S_P_i}, \mathbf{M}_{L_P_i} \in \mathbb{R}^{1 \times N}$. Due to \mathcal{L}_{Rec}^M , $\mathbf{F}_{S_P}^R$ and $\mathbf{F}_{L_P}^R$ can be closed to \mathbf{F}_E in the feature space. Therefore, the proposed framework can perform robust pedestrian detection at

various scales.

3.4. Total Loss Function

The total loss function is represented as follows:

$$\mathcal{L}_{OD} = \mathcal{L}_{RPN} + \mathcal{L}_{CLS} + \mathcal{L}_{LOC}, \quad (8)$$

$$\mathcal{L}_{Total} = \mathcal{L}_{OD} + \lambda_1 \mathcal{L}_{Mem}^M + \lambda_2 \mathcal{L}_{Rec}^M,$$

where \mathcal{L}_{OD} includes the loss function of RPN, and the classification and localization loss functions of the head network, which are same with the two-stage object detectors [36, 46, 48]. λ_1 and λ_2 are balancing hyper-parameters to modulate \mathcal{L}_{Mem}^M and \mathcal{L}_{Rec}^M .

4. Experiments

4.1. Datasets

To evaluate the proposed method, we use two public datasets: KAIST Multispectral Pedestrian Detection Dataset [22] (we denote it as KAIST dataset for simplicity) and CVC-14 [16]. Two datasets include visible and thermal modal images. Therefore, we can perform various experiments using both modalities.

KAIST Dataset. KAIST dataset [22] was collected in a driving environment of various scenes including day and night conditions. It consists of 95,328 visible-thermal image pairs with 103,128 annotations and 1,182 unique pedestrians with an image resolution 512×640 . Following [22], we evaluate the test set with 2,252 images.

CVC-14. Similar to the KAIST dataset, visible-thermal image pairs of CVC-14 [16] were also captured in the driving environments. It is also divided into day and night conditions. For CVC-14, we use 7,085 training and 1,433 test images with 471×640 image resolution.

4.2. Implementation Details

We build our base detection framework based on Faster R-CNN [46] with VGG16 backbone network [47]. Note that, in ablation studies, we extend the base detection framework to Feature Pyramid Network (FPN) [36] with ResNet backbone network [19] to investigate the effect of the LPR Memory with different base detection framework.

All experiments are conducted based on Pytorch [44]. Every base detection framework is trained with stochastic gradient descent (SGD) which is synchronized over 8 TITAN XP GPUs, holding 1 image per GPU (total 8 images per mini-batch). We train the base detection framework for 4 epochs with 0.008 learning rate. The number of RoIs is $N = 256$. We use the number of slots in the LPR Memory $L = 100$ as the default. The number of the randomly selected large-scale pedestrian exemplar set is $N_E = 32$. In

Method	Scale			
	Near	Medium	Far	All
SAN [29] (ECCV'18)	5.33	31.19	61.40	25.41
Noh et al. [39] (ICCV'19)	4.87	30.28	59.18	24.08
TC Det [25] (ECCV'20)	3.22	34.75	76.94	27.11
SML [53] (ACMMM'20)	13.24	26.65	58.62	22.51
CUA Loss [27] (TCSVT'20)	5.67	29.43	58.55	23.39
Baseline	8.74	31.72	65.87	26.70
Proposed Method	4.15	25.13	51.40	19.16

Table 1. Detection results (MR) of three pedestrian scales with the thermal modality on KAIST dataset [22].

addition, we use $\lambda_1, \lambda_2 = 1$ and $H_L = 115$ in our experiments.

4.3. Evaluation on KAIST Dataset

We first conduct experiments on KAIST dataset [22]. As an evaluation metric, we adopt a miss rate (MR) averaged over the false positive per image (FPPI) with the range of $[10^{-2}, 10^0]$, following [11]. The lower MR denotes the better detection performance. We measure the performance using ‘All’ condition [22]. Moreover, we compare the performance of the proposed method with the state-of-the-art methods in the various pedestrian scales. The scale of pedestrians in the KAIST dataset [22] is categorized into ‘Far’ (small) ($1 \leq height < 45$), ‘Medium’ ($45 \leq height < 115$), and ‘Near’ (large) ($115 \leq height$).

We conduct the experiments with two cases: thermal and visible modal image inputs. Note that, the baseline is the Faster R-CNN [46]. First, we measure the performance on thermal-based pedestrian detection. We compare our framework with the previous methods that aim to solve the small-scale issue [27, 29, 39, 53]. We re-implement all these methods and obtain the results. Table 1 shows the results of three pedestrian scales (*i.e.* ‘Near’, ‘Medium’, and ‘Far’). The proposed method outperforms the previous methods both in the smaller scale (*i.e.* ‘Medium’ and ‘Far’). Since the LPR Memory can recall the visual appearance of the large-scale pedestrian from the small-scale pedestrians, our framework shows improved performances. Moreover, we compare our method with the state-of-the-art thermal-based pedestrian detection methods. We also compare the performance in day and night conditions (*i.e.* ‘Day’ and ‘Night’) [22]. As seen in Table 2, our method is superior to the previous methods. According to [22], the ratio of the small-scale pedestrians (*i.e.* ‘Medium’ and ‘Far’) is 86.33%. Therefore, the performance improvement of our method is noticeable.

Second, we conduct the experiments with a visible modal image input. We also compare with the previous methods that aim to solve the small-scale issue [27, 29, 39, 53]. It is shown in Table 3. The experimental results on the visible modality also outperform the previous methods [27, 29, 39, 53]. These results corroborate the effectiveness of the proposed LPR Memory.

Method	All	Day	Night
TPIHOG [1] (Sensors'17)	N/A	N/A	57.38
SSD300 [20] (SPIE'20)	69.81	N/A	N/A
Saliency Map [17] (ICIP'19)	N/A	30.40	21.00
Domain Adaptor [17] (ICIP'19)	46.30	53.37	31.63
Bottom-up [24] (ICIAP'19)	35.20	40.00	20.50
TC Thermal [25] (ECCV'20)	28.53	36.59	11.03
TC Det [25] (ECCV'20)	27.11	34.81	10.31
Kieu et al. [26] (ICPR'20)	25.62	31.86	12.92
Proposed Method	19.16	24.70	8.26

Table 2. Comparison of detection results (MR) using the thermal modality on KAIST dataset [22].

Method	Scale			
	Near	Medium	Far	All
SAN [29] (ECCV'18)	15.39	35.31	73.51	29.62
Noh et al. [39] (ICCV'19)	7.39	34.53	72.09	28.78
SML [53] (ACMMM'20)	5.97	35.65	74.45	28.15
CUA Loss [27] (TCSVT'20)	11.86	34.21	73.18	27.81
Baseline	7.62	38.70	77.82	30.66
Proposed Method	5.35	32.09	68.92	25.16

Table 3. Detection results (MR) of three pedestrian scales with the visible modality on KAIST dataset [22].

Method	Scale			
	Near	Medium	Far	All
Baseline	5.83	28.57	72.83	27.42
Proposed Method	4.44	23.07	64.98	23.00

Table 4. Detection results (MR) of three scale pedestrians with the thermal modality on CVC-14 [16].

4.4. Evaluation on CVC-14

For CVC-14 [16], we follow the same settings with KAIST dataset, except for the scale threshold. We set the *height* threshold as 60 for ‘Far’ (small) scale, because the overall size of the objects is relatively larger than those of KAIST dataset [22] (More details are in the supplementary materials). The evaluation results are shown in Table 4. The results show that our method with the LPR Memory outperforms the baseline (*i.e.* Faster R-CNN) especially in ‘Medium’ and ‘Far’ conditions.

4.5. Ablation Study

In this subsection, we conduct various ablation studies to investigate the effect of (1) the large-scale embedding learning, (2) the memory slot number (L) variation, and (3) the extension of the base detection framework. Every ablation study is conducted on KAIST dataset with the thermal modality [22].

Effect of Large-scale Embedding Learning. Note that the large-scale embedding learning contains the large-scale

\mathcal{L}_{Mem}^M	\mathcal{L}_{Rec}^M	Scale			
		Near	Medium	Far	All
\times	\times	8.74	31.72	65.87	26.70
\checkmark	\times	4.52	27.86	53.99	22.88
\times	\checkmark	3.47	30.63	58.09	23.16
\checkmark	\checkmark	4.15	25.13	51.40	19.16

Table 5. Ablation detection results (MR) showing the effect of the large-scale embedding learning with the thermal modality of KAIST dataset [22].

L slot	Scale			
	Near	Medium	Far	All
-	8.74	31.72	65.87	26.70
25	5.55	27.54	55.87	23.68
50	4.59	25.57	51.43	20.23
100	4.15	25.13	51.40	19.16
200	2.79	26.71	58.27	20.46

Table 6. Detection results (MR) of the proposed framework with varying slot number L of the LPR Memory on KAIST dataset [22].

pedestrian memorizing loss \mathcal{L}_{Mem}^M and the large-scale pedestrian recalling loss \mathcal{L}_{Rec}^M . The results are shown in Table 5. When both loss functions are considered, performance improvements on ‘Medium’ and ‘Far’ scales are noticeable. The experimental results show the effectiveness of the large-scale embedding learning for the LPR Memory.

Slot number of LPR Memory. To demonstrate the effect of the slot number L in the LPR Memory, we modify L by changing $\{25, 50, 100, 200\}$. As shown in Table 6, when the slot number L is 100, the highest performance is obtained in the medium (‘Medium’), small-scale (‘Far’) and the overall (‘All’). Furthermore, even we change L of the LPR Memory, the proposed method shows higher performances than the baseline (Faster R-CNN).

Detection Framework Extension. We adopt Feature Pyramid Network (FPN) [36] built upon ResNet-50 backbone network to verify the effect of the proposed method with different detection frameworks. It is shown in Table 7. The experimental results verify that performance improvement of our framework with the LPR Memory is significant in the various pedestrian scales and the overall performance.

4.6. Visualization Results

RoI Feature Visualization. To compare the original small-scale pedestrian RoI feature \mathbf{F}_S and the refined small-scale pedestrian RoI feature \mathbf{F}_S^R , we perform the visualization experiments with KAIST dataset [22], following [54]. The visualization results are shown in Figure 5. While \mathbf{F}_S is activated on the background region and could not focus on the pedestrian, \mathbf{F}_S^R captures the small-scale pedestrian properly. The visualization results also verify

Detection Framework	LPR Memory	Scale			
		Near	Medium	Far	All
FRCNN [46] (VGG16)	\times \checkmark	8.74 4.15	31.72 25.13	65.87 51.40	26.70 19.16
FPN [36] (ResNet-50)	\times \checkmark	2.25 1.96	21.96 18.24	53.46 46.43	16.56 13.91

Table 7. Detection results (MR) with the different detection frameworks on KAIST dataset [22] showing the effectiveness of the LPR Memory.

the effectiveness of the LPR Memory.

t-SNE Visualization. We also conduct t-SNE visualization [51] of two modalities (thermal and visible) on KAIST test set [22]. Note that, we set the scale definition of the small and large following [22]. As shown in Figure 6, due to the large-scale pedestrian recalling process of the LPR Memory, features of the small-scale pedestrian can be located near the large-scale pedestrian features.

4.7. Discussion

Advantages of Large-scale Pedestrian Exemplar Set.

First, large-scale pedestrian exemplar set \mathbf{E} contains well-aligned large-scale pedestrian images. In contrast, the large-scale pedestrian RoIs may contain not well-aligned pedestrians. Therefore, it is reasonable to use \mathbf{E} in order to guide the small- and large-scale RoI features to be similar with the pedestrian exemplar features. Second, \mathbf{E} contains various appearances of the large-scale pedestrian. Humans recall the appearance of the pedestrian through various past experiences, so that, \mathbf{E} could be effective for training the LPR Memory to recall the large-scale pedestrian by providing various large-scale pedestrian appearances.

LPR Memory Visualization. Figure 7 shows how \mathbf{F}_S is improved to the refined pedestrian feature \mathbf{F}_S^R with the features in \mathbf{M}_V (we show top-2 high-weighted slots of \mathbf{M}_V). Although \mathbf{F}_S cannot focus on the pedestrian, \mathbf{F}_S^R can be obtained with the help of the features in \mathbf{M}_V to capture pedestrian region properly. Moreover, we observe that cosine similarity between \mathbf{F}_S^R and \mathbf{F}_E is much higher than that of \mathbf{F}_S and \mathbf{F}_E .

Large-scale Pedestrian Exemplar Set in Different Dataset.

When training the LPR Memory on KAIST dataset [22], we use the large-scale pedestrian exemplar set \mathbf{E} from the same KAIST dataset [22]. To see the effect of \mathbf{E} from the different dataset, we change \mathbf{E} obtained from another dataset. In detail, for training the detection framework with the thermal images of KAIST dataset [22], we use \mathbf{E} extracted from the different dataset CVC-14 [16]. As shown in Table 8, even we use the exemplar set \mathbf{E} extracted from the CVC-14 dataset, the detection performance is

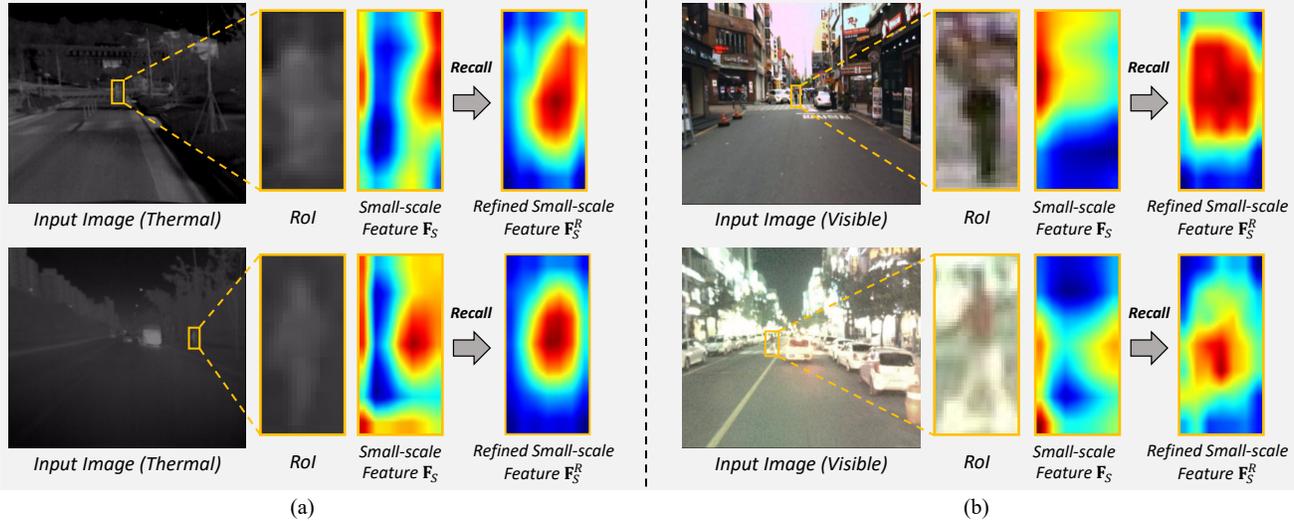


Figure 5. Feature map visualization of the small-scale ROI features F_S and refined small-scale ROI features F_S^R in the (a) thermal and (b) visible images. While F_S cannot focus on the small-scale pedestrians, F_S^R is activated on the proper regions of the small-scale pedestrians.

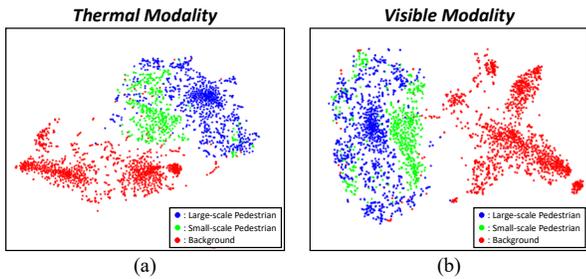


Figure 6. t-SNE visualization [51] with the (a) thermal and (b) visible modalities. In both modalities, due to the large-scale embedding learning, the refined small-scale pedestrian features become distinguishable from the background features and close to the large-scale pedestrian features.

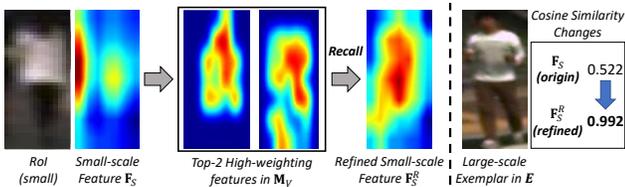


Figure 7. An example of improving the small-scale pedestrian feature map with the help of the LPR memory.

improved in various pedestrian scales, and also, the overall performance ('All') show better performance than the baseline (Faster R-CNN).

Parameter Number of LPR Memory. The number of parameters of the baseline, Faster R-CNN [46], is 141.27M. With the Faster R-CNN, the number of parameters of our framework is increased to 146.42M incorporating the 1×1 convolution and the LPR memory ($L = 100$). The parameter number of the LPR memory is 5.02M, requiring only 3.6% overhead compared to the baseline.

Method	Scale			
	Near	Medium	Far	All
Baseline	8.74	31.72	65.87	26.70
Proposed Method with E (KAIST)	4.15	25.13	51.40	19.16
Proposed Method with E (CVC-14)	3.58	28.27	52.85	22.68

Table 8. Detection results (MR) of the proposed framework using the thermal modality (KAIST dataset [22]), with each large-scale pedestrian exemplar set E , extracted from the same dataset (KAIST) and different dataset (CVC-14).

5. Conclusion

In this paper, we introduce a novel pedestrian detection framework to detect small-scale pedestrians effectively, which is aware of the context of the large-scale pedestrians. To this end, we propose the LPR Memory with the large-scale embedding learning to memorize the various appearance of the large-scale pedestrian and to recall the relevant large-scale pedestrian appearance of the small-scale pedestrian. The process of the LPR Memory is similar to the process of how humans recognize small-scale pedestrians by associating the previous experience. As a result, the LPR Memory enables the detection framework to perform robust detection on small-scale pedestrians. The quantitative and the qualitative results demonstrate the effectiveness of the proposed framework. We believe that our human-like memory approach can be applied to a wide range of studies in the general object detection areas.

Acknowledgement. This work was conducted by Center for Applied Research in Artificial Intelligence (CARAI) grant funded by DAPA and ADD (UD190031RD).

References

- [1] Jeonghyun Baek, Sungjun Hong, Jisu Kim, and Euntai Kim. Efficient pedestrian detection at nighttime using a thermal camera. *Sensors*, 2017. 6
- [2] Rodrigo Benenson, Mohamed Omran, Jan Hosang, and Bernt Schiele. Ten years of pedestrian detection, what have we learned? In *ECCV*, 2014. 2
- [3] Muhammad Bilal, Asim Khan, Muhammad Umar Karim Khan, and Chong-Min Kyung. A low-complexity pedestrian detection framework for smart video surveillance systems. *TCSVT*, 2016. 1
- [4] Antoine Bordes, Nicolas Usunier, Sumit Chopra, and Jason Weston. Large-scale simple question answering with memory networks. *arXiv preprint arXiv:1506.02075*, 2015. 2
- [5] Herman Buschke. Cued recall in amnesia. *JCEN*, 1984. 2
- [6] Qi Cai, Yingwei Pan, Ting Yao, Chenggang Yan, and Tao Mei. Memory matching networks for one-shot image recognition. In *CVPR*, 2018. 2
- [7] Zhaowei Cai, Quanfu Fan, Rogerio S Feris, and Nuno Vasconcelos. A unified multi-scale deep convolutional neural network for fast object detection. In *ECCV*, 2016. 2
- [8] Jiale Cao, Yanwei Pang, Shengjie Zhao, and Xuelong Li. High-level semantic networks for multi-scale object detection. *TCSVT*, 2019. 2
- [9] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *CVPR*, 2005. 2
- [10] Hanming Deng, Yang Hua, Tao Song, Zongpu Zhang, Zhenhui Xue, Ruhui Ma, Neil Robertson, and Haibing Guan. Object guided external memory network for video object detection. In *ICCV*, 2019. 2
- [11] Piotr Dollar, Christian Wojek, Bernt Schiele, and Pietro Perona. Pedestrian detection: An evaluation of the state of the art. *TPAMI*, 2011. 2, 6
- [12] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *CVPR*, 2012. 1
- [13] Theo Gevers and Arnold WM Smeulders. Pictoseek: Combining color and shape invariant features for image retrieval. *TIP*, 2000. 1
- [14] Debasmita Ghose, Shasvat M Desai, Sneha Bhattacharya, Deep Chakraborty, Madalina Fiterau, and Tauhidur Rahman. Pedestrian detection in thermal images using saliency maps. In *CVPR Workshops*, 2019. 2
- [15] Dong Gong, Lingqiao Liu, Vuong Le, Budhaditya Saha, Moussa Reda Mansour, Svetha Venkatesh, and Anton van den Hengel. Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection. In *ICCV*, 2019. 2
- [16] Alejandro González, Zhijie Fang, Yainuvis Socarras, Joan Serrat, David Vázquez, Jiaolong Xu, and Antonio M López. Pedestrian detection at day/night time with visible and fir cameras: A comparison. *Sensors*, 2016. 5, 6, 7
- [17] Tiantong Guo, Cong Phuoc Huynh, and Mashhour Solh. Domain-adaptive pedestrian detection in thermal images. In *ICIP*, 2019. 1, 2, 6
- [18] Tengda Han, Weidi Xie, and Andrew Zisserman. Memory-augmented dense predictive coding for video representation learning. In *ECCV*, 2020. 2
- [19] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, 2016. 3, 5
- [20] Christian Herrmann, Miriam Ruf, and Jürgen Beyerer. Cnn-based thermal infrared person detection by domain adaptation. In *SPIE*, 2018. 6
- [21] Xin Huang, Zheng Ge, Zequn Jie, and Osamu Yoshie. Nms by representative region: Towards crowded pedestrian detection by proposal pairing. In *CVPR*, 2020. 1, 2
- [22] Soonmin Hwang, Jaesik Park, Namil Kim, Yukyung Choi, and In So Kweon. Multispectral pedestrian detection: Benchmark dataset and baseline. In *CVPR*, 2015. 1, 2, 5, 6, 7, 8
- [23] Mira Jeong, Byoung Chul Ko, and Jae-Yeal Nam. Early detection of sudden pedestrian crossing for safe driving during summer nights. *TCSVT*, 2016. 1
- [24] My Kieu, Andrew D Bagdanov, Marco Bertini, and Alberto Del Bimbo. Domain adaptation for privacy-preserving pedestrian detection in thermal imagery. In *ICIAP*, 2019. 1, 2, 6
- [25] My Kieu, Andrew D. Bagdanov, Marco Bertini, and Alberto del Bimbo. Task-conditioned domain adaptation for pedestrian detection in thermal imagery. In *ECCV*, 2020. 1, 2, 6
- [26] My Kieu, Lorenzo Berlincioni, Leonardo Galteri, Marco Bertini, Andrew D Bagdanov, and Alberto Del Bimbo. Robust pedestrian detection in thermal imagery using synthesized images. In *ICPR*, 2021. 6
- [27] Jung Uk Kim, Seong Tae Kim, Hong Joo Lee, Sangmin Lee, and Yong Man Ro. Cua loss: Class uncertainty-aware gradient modulation for robust object detection. *TCSVT*, 2020. 2, 3, 6
- [28] Jung Uk Kim, Sungjune Park, and Yong Man Ro. Uncertainty-guided cross-modal learning for robust multispectral pedestrian detection. *TCSVT*, 2021. 1
- [29] Yonghyun Kim, Bong-Nam Kang, and Daijin Kim. San: Learning relationship between convolutional features for multi-scale object detection. In *ECCV*, 2018. 1, 2, 6
- [30] Sangmin Lee, Hak Gu Kim, Dae Hwi Choi, Hyung-Il Kim, and Yong Man Ro. Video prediction recalling long-term motion context via memory alignment learning. In *CVPR*, 2021. 2
- [31] Sangho Lee, Jinyoung Sung, Youngjae Yu, and Gunhee Kim. A memory network approach for story-based temporal summarization of 360 videos. In *CVPR*, 2018. 2
- [32] Chengyang Li, Dan Song, Ruofeng Tong, and Min Tang. Illumination-aware faster r-cnn for robust multispectral pedestrian detection. *Pattern Recognition*, 2019. 1, 2
- [33] Jianan Li, Xiaodan Liang, ShengMei Shen, Tingfa Xu, Jiashi Feng, and Shuicheng Yan. Scale-aware fast r-cnn for pedestrian detection. *TMM*, 2017. 2
- [34] Yanghao Li, Yuntao Chen, Naiyan Wang, and Zhaoxiang Zhang. Scale-aware trident networks for object detection. In *ICCV*, 2019. 2

- [35] Chunze Lin, Jiwen Lu, and Jie Zhou. Multi-grained deep feature learning for robust pedestrian detection. *TCSVT*, 2018. [2](#)
- [36] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. In *CVPR*, 2017. [3](#), [5](#), [7](#)
- [37] Francesco Marchetti, Federico Becattini, Lorenzo Seidenari, and Alberto Del Bimbo. Mantra: Memory augmented networks for multiple trajectory prediction. In *CVPR*, 2020. [2](#)
- [38] Peter A Nobel and Richard M Shiffrin. Retrieval processes in recognition and cued recall. *JEPLMC*, 2001. [2](#)
- [39] Junhyug Noh, Wonho Bae, Wonhee Lee, Jinhwan Seo, and Gunhee Kim. Better to follow, follow to be better: Towards precise supervision of feature super-resolution for small object detection. In *ICCV*, 2019. [2](#), [6](#)
- [40] Wanli Ouyang, Xingyu Zeng, and Xiaogang Wang. Partial occlusion handling in pedestrian detection with a deep model. *TCSVT*, 2015. [2](#)
- [41] Yanwei Pang, Jiale Cao, Jian Wang, and Jungong Han. Jcs-net: Joint classification and super-resolution network for small-scale pedestrian detection in surveillance images. *TIFS*, 2019. [1](#), [2](#)
- [42] Yanwei Pang, Jin Xie, Muhammad Haris Khan, Rao Muhammad Anwer, Fahad Shahbaz Khan, and Ling Shao. Mask-guided attention network for occluded pedestrian detection. In *ICCV*, 2019. [1](#), [2](#)
- [43] Hyunjong Park, Jongyoun Noh, and Bumsub Ham. Learning memory-guided normality for anomaly detection. In *CVPR*, 2020. [2](#)
- [44] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. In *NeurIPS Workshop*, 2017. [5](#)
- [45] Junran Peng, Ming Sun, Zhaoxiang Zhang, Tieniu Tan, and Junjie Yan. Pod: Practical object detection with scale-sensitive network. In *ICCV*, 2019. [2](#)
- [46] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: towards real-time object detection with region proposal networks. *TPAMI*, 2016. [2](#), [3](#), [5](#), [6](#), [7](#), [8](#)
- [47] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. [3](#), [5](#)
- [48] Bharat Singh, Hengduo Li, Abhishek Sharma, and Larry S Davis. R-fcn-3000 at 30fps: Decoupling detection and classification. In *CVPR*, 2018. [5](#)
- [49] Tao Song, Lei Yu Sun, Di Xie, Haiming Sun, and Shiliang Pu. Small-scale pedestrian detection based on topological line localization and temporal feature aggregation. In *ECCV*, 2018. [1](#), [2](#)
- [50] Yi Tu, Li Niu, Junjie Chen, Dawei Cheng, and Liqing Zhang. Learning from web data with self-organizing memory module. In *CVPR*, 2020. [2](#)
- [51] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *JMLR*, 2008. [1](#), [7](#), [8](#)
- [52] Vibashan VS, Poojan Oza, Vishwanath A Sindagi, Vikram Gupta, and Vishal M Patel. Mega-cda: Memory guided attention for category-aware unsupervised domain adaptive object detection. *arXiv preprint arXiv:2103.04224*, 2021. [2](#)
- [53] Jialian Wu, Chunlun Zhou, Qian Zhang, Ming Yang, and Junsong Yuan. Self-mimic learning for small-scale pedestrian detection. In *ACM MM*, 2020. [1](#), [2](#), [6](#)
- [54] Chang-Dong Xu, Xing-Ran Zhao, Xin Jin, and Xiu-Shen Wei. Exploring categorical regularization for domain adaptive object detection. In *CVPR*, 2020. [7](#)
- [55] Liliang Zhang, Liang Lin, Xiaodan Liang, and Kaiming He. Is faster r-cnn doing well for pedestrian detection? In *ECCV*, 2016. [1](#), [2](#)
- [56] Shizhou Zhang, De Cheng, Yihong Gong, Dahu Shi, Xi Qiu, Yong Xia, and Yanning Zhang. Pedestrian search in surveillance videos by learning discriminative deep features. *Neurocomputing*, 2018. [1](#)
- [57] Shifeng Zhang, Longyin Wen, Xiao Bian, Zhen Lei, and Stan Z Li. Occlusion-aware r-cnn: detecting pedestrians in a crowd. In *ECCV*, 2018. [2](#)
- [58] Tianliang Zhang, Qixiang Ye, Baochang Zhang, Jianzhuang Liu, Xiaopeng Zhang, and Qi Tian. Feature calibration network for occluded pedestrian detection. *TITS*, 2020. [2](#)
- [59] Xiaowei Zhang, Li Cheng, Bo Li, and Hai-Miao Hu. Too far to see? not really!—pedestrian detection with scale-aware localization policy. *TIP*, 2018. [2](#)
- [60] Zhishuai Zhang, Jiyang Gao, Junhua Mao, Yukai Liu, Dragomir Anguelov, and Congcong Li. Stinet: Spatio-temporal-interactive network for pedestrian detection and trajectory prediction. In *CVPR*, 2020. [1](#), [2](#)
- [61] Linchao Zhu and Yi Yang. Inflated episodic memory with region self-attention for long-tailed visual recognition. In *CVPR*, 2020. [2](#)