

MULTI-TASK FEATURE DECOMPOSITION BASED MARGINAL DISTRIBUTION FOR PERSON SEARCH

Yuanzhe Yang[†], Xin Zhang[†], Qichuan Geng[‡], Chengxiang Chu[†] and Zhong Zhou[†] ✉

[†]State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing, China

[‡]Capital Normal University, Beijing, China

zz@buaa.edu.cn

ABSTRACT

Person search is a composite task, aiming at locating and identifying a query person from uncropped images. It requires jointly solving Pedestrian Detection and Person Re-identification. One major challenge in person search is the contradictory goals of detection and re-identification. The model has to simultaneously model the universality and specificity of persons. In this paper, we propose a novel parameter-free approach called Feature Decomposition Person Search (FDPS) to separate various tasks. FDPS decomposes the ROI feature map to extract sub-features based on the marginal distribution for different tasks. Also, we find that the Online Instance Match loss pays imbalanced attention to positive and negative categories. We present a Balance Online Instance Match (BOIM) loss to enhance the contribution of negative categories during training. Our method achieves the state-of-the-art performance in one-step methods on two prevailing benchmarks, with high efficiency.

Index Terms— Person Search, Marginal Distribution, Multi-task Learning, Category Balance

1. INTRODUCTION

Person search [1], aiming to locate a target person on a set of realistic uncropped images, is a practically relevant task. Recently, it has received extensive attention in public security, innovative business, and person activity analysis due to its broad application prospects. Generally considered, person search is a Multi-Task Learning (MTL) task including Pedestrian Detection [2] and Person Re-identification (Re-ID) [3]. Therefore, it is not enough to focus only on the challenges of sub-tasks. It is also crucial to coordinate different sub-task.

Existing person search methods are divided into two lines according to whether integrating detection and Re-ID into one model. The first line is the two-step methods [4–6]. However, due to the high overhead of the two-step methods, our study focuses on the one-step methods. The one-step methods [7–9] complete the whole process in one model, jointly optimizing both two tasks. The existing methods only generate one shared feature, then both detection and Re-ID depend

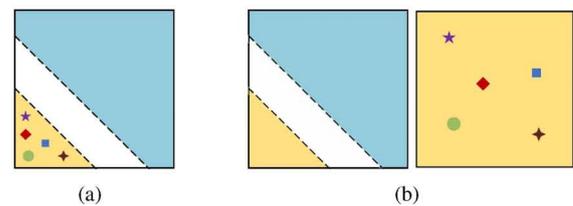


Fig. 1. Illustration on how person and background representations are scattered in the embedding space. The Blue area denotes the background space. The orange area devotes person space, where colourful markers denote person embeddings. In (a), the inter-class distances are squeezed by backgrounds. In (b), the embedding space is separated. Thus constrain on inter-class distances is relaxed.

on the same feature. The universality and specificity of persons are modelled in the same embedding space. Thus the person embedding space will be squeezed by backgrounds, as shown in Figure. 1(a). The contradiction between detection and Re-ID hurts the expression ability of the feature.

Motivated by the above observations, we propose a feature decomposition method in MTL head, named Feature Decomposition Person Search (FDPS). FDPS focuses on optimizing person ROI feature map utilization. Based on marginal distribution, FDPS decomposes the ROI feature map into several different sub-features, each of which is only used for one sub-task. Different features come from different dimensions and encode various information, separating multiple sub-tasks. Simultaneously, it highlights the information required by the corresponding sub-tasks by aggregating the feature. Each sub-tasks has its own embedding space, alleviating the issue of space squeeze, as Figure. 1(b). In subsequent experiments, the results illustrated that it can improve model performance by extracting sub-features with parameter-based layers or enlarging the embedding space. However, the improvement is inferior to the proposed FDPS.

In Re-ID, the duplicate IDs are expected to cluster close to each other, and different IDs should be far apart. However, we find that the Online Instance Match (OIM) loss pays imbalanced attention on positive and negative categories, mak-

ing it challenging to separate different IDs for model. To enlarge the contribution of negative categories, we propose a novel Balance Online Instance Match (BOIM) loss. BOIM mines the contribution of negative categories to compensate for the imbalance issue in OIM loss. Combining FDPS and BOIM, the person search performance has been improved significantly. Our method achieves state-of-the-art performance in one-step methods and is even comparable to non-end-to-end methods needing human interaction.

In summary, the main contributions are as follows:

- We propose the Feature Decomposition Person Search to alleviate the contradiction by extracting sub-features for different sub-tasks based on marginal distribution.
- We propose Balance Online Instance Match loss to enhance the contribution of negative categories to further separate different IDs apart.
- Our method is portable achieves state-of-the-art performance in one-step methods on standard benchmarks, CUHK-SYSU [7] and PRW [10].

2. RELATED WORK

Recently, person search raised a lot of interest in computer vision. The existing methods can be divided into one-step and two-step. The two-step methods perform detection and Re-ID in two independent models. Zheng et al. [10] made a comprehensive evaluation, then proposed a cascaded fine-tuning strategy. Chen et al. [4] used the segmentation mask to establish two branches for discriminative features. Lan et al. [6] introduced a multi-scale feature alignment, and Wang et al. [11] mined the consistency between the two models. While the two-step method has more overhead due to the images going through two models.

The one-step methods complete the whole process in a one model and has attract more attention because of its efficiency and simplicity. Xiao et al. [7] proposed the first end-to-end person search framework based on Faster R-CNN [12]. Yan et al. [13] used a graph model to consider the persons around the target. Dong et al. [9] introduced an instance-aware branch to make the model more focused on the foreground area. Several studies have noted that optimising the relationship between sub-tasks in person search is necessary. Chen et al. [8] expanded the embedding space to alleviate the contradiction between detection and Re-ID. Unlike previous studies, the proposed FDPS decomposes the ROI feature map and separates different sub-tasks into different embedding spaces to alleviate the contradiction.

Recently, the loss function has shown great potential in dominating the field of person search. Xiao et al. [7] proposed the OIM loss, becoming the widely used loss in person search. Chen et al. [14] focused on the hierarchical relationship between detection and Re-ID and introduced the background embedding table in loss. Han et al. [5] proposed Proxy

Triplet Loss to embed person features in European space and used triplet loss to optimize model. Other studies introduced Triplet loss [15] and Center loss [16] to person search. We propose the BOIM loss for mining negative categories contributions to separate the different IDs.

3. METHOD

In this section, we firstly analyse the objectives contradiction issue in the one-step person search framework. Then we briefly introduce the proposed FDPS, which decomposes the ROI ature map to separate sub-tasks based on marginal distribution. Finally, we discuss the attention imbalance in OIM [7] loss and present a novel BOIM loss to mine contribution of negative categories.

3.1. Optimization Problem Analysis

The one-step methods are based on two-stage detection framework, such as Faster R-CNN [12], in which Region Proposal Network (RPN) generates person ROIs and a multi-task head is attached to the ROI feature maps for location and identifying. All person ROIs are aligned to fixed $w \times h$ size with C channels, making the size of a person ROI $C \times w \times h$. Then the global max pooling(GMP) is performed to generate a C -dimensional feature, which would be fed to the MTL head.

The MTL head contains three irrelevant sub-tasks: person re-identification, bounding boxes regression and person-background classification, which requires the feature contains person universality, specificity and location information. The shared feature leads to two challenges. (a) Person embedding area is squeezed in the embedding space, making it challenging to further separate different IDs. (b) The vector must contain appearance and location information. Thence, sub-tasks have trouble extracting the necessary information from the shared feature. In conclusion, using one shared feature hurts model performance. Meanwhile, we find person ROIs have the shape of $C \times w \times h$. Still, only a C -dimensional feature is used, which means the parameter utilization rate is only $1/wh$, not fully utilized.

We introduce the marginal distribution to aggregate and highlight the feature in related dimensions and extract multiple sub-features from the ROI feature map to alleviate the above issues. Multiple sub-features will replace the shared feature and be fed into MTL head for the corresponding sub-tasks to achieve the separation of sub-tasks.

3.2. Feature Decompose by Marginal Distribution

We propose a novel feature decompose approach in MTL head, whose structure is shown in Figure 2. An ROI feature map could be considered as a joint feature distribution, which can be formalized as

$$D(x, y, c) = (f_{xyc})_{w \times h \times C} \quad (1)$$

where f devotes the original ROI feature map, directly regarded as a discrete distribution D . $x \in [1, w]$ and $y \in [1, h]$ represent horizontal and vertical coordinates, and $c \in [1, C]$ represents the channel coordinates. Then, we can extract the sub-features based on marginal distribution from the entire feature distribution. Based on marginal distribution, original features can be aggregated in corresponding dimensions to highlight the sub-feature in related dimensions. Compared with the one shared feature encoding all information, each sub-feature only needs to encode one of them. According to the sub-features, the different sub-tasks are separated into corresponding dimensions of the ROI feature map.

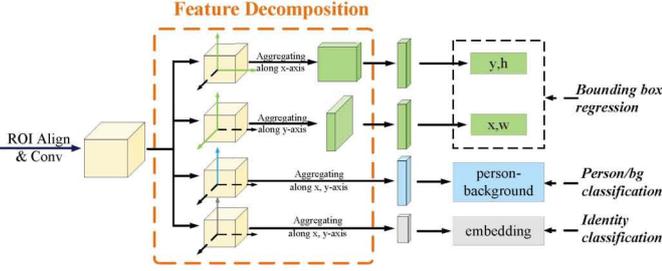


Fig. 2. Illustration on architecture of FDPS. An ROI feature map is decomposed into several sub-features for sub-tasks.

Then, we assign suitable sub-features to the sub-tasks. When regressing the bounding boxes, we predict the center point coordinates (cx, cy) , person width and height (pw, ph) . In previous studies, (cx, cy, pw, ph) were always generated from the same feature. According to the dimensions, we further subdivide the bounding box regression into two sub-tasks, predicting (cx, pw) and (cy, ph) , respectively. We perform regression tasks on the corresponding dimensions for the two sub-tasks of bounding box regression. Detailedly, we predict (cx, pw) on the X-axis and (cy, ph) on the Y-axis of distribution D . Considering retaining the channel capacity, we extract sub-features on (x, c) and (y, c) as features for (cx, pw) and (cy, ph) based on marginal distribution and scale them by the size. The extraction process can be formulated as follows:

$$D_y = D_X(y, c) = \frac{1}{w} \sum_x f_{xyc} \quad (2)$$

$$D_x = D_Y(x, c) = \frac{1}{h} \sum_y f_{xyc} \quad (3)$$

where, D_y and D_x are sub-features on (y, c) and (x, c) dimension, used for (cy, ph) and (cx, pw) , respectively.

Abundant appearance information is encoded in the channel dimension of the ROI feature map. Existing studies have proved that GMP can generate discriminative person embeddings. However, we realize that GMP uses only a tiny part of the ROI feature map. The appearance specificity is more local, but the universality pays less attention to local areas. Thereby, the sub-feature for Re-ID is still extracted by GMP,

while the sub-feature for person-background classification is collected based on marginal distribution in channel dimension with a scale operation. The above approaches can be formulated as follows:

$$D_{reid} = D_{XY}^1(c) = \max_{x,y} f_{xyc} \quad (4)$$

$$D_{cls} = D_{XY}^2(c) = \frac{1}{wh} \sum_x \sum_y f_{xyc} \quad (5)$$

where, D_{cls} and D_{reid} are both sub-features on the channel dimension, used for pedestrian-background classification and Re-ID, respectively. D_{cls} uses all parameters on the channel dimension to aggregate features across the whole channel, while D_{reid} only relies on the maximum value focusing only on the corresponding part. According to the feature decomposition strategy, the two sub-tasks are separated.

In FDPS, the MTL head is divided into four branches, as shown in Figure 2. The upper two branches extract sub-features in the (y, c) and (x, c) dimensions as Equation.(2) and Equation.(3) for bounding box regression. The third branch extracts sub-feature basing marginal distribution in the c dimension for pedestrian-background classification as Equation.(5). The last branch is another sub-feature on the channel dimension as Equation.(4) for Re-ID.

Based on marginal distribution, all parameters in the ROI are utilized in FDPS. The information required by different sub-tasks is encoded in corresponding dimensions to separate sub-tasks. Sub-features are aggregated in related dimensions, highlighting the information acquired for sub-tasks.

3.3. Balance Online Instance Match Loss

After separating sub-tasks, FDPS alleviates the issue of squeezing in embedding space. However, we find that attention imbalance in the OIM loss [7] also affects the separation of different IDs.

OIM loss maintains an identified lookup table and an unidentified circular queue, storing the identified embedding centres and non-identified embedding instances. The function can be formalized as

$$p_i = \frac{\exp(v_i^T / \tau)}{\sum_{j=1}^L \exp(v_j^T x / \tau) + \sum_{k=1}^Q \exp(u_k^T x / \tau)} \quad (6)$$

$$L_{oim} = -E_x[\log p_t] \quad (7)$$

where, $v \in \mathbb{R}^{C \times L}$ is the identified lookup table with length L , and $u \in \mathbb{R}^{C \times Q}$ is the unidentified circular queue with length Q . x is a new sample, and τ is a hyperparameter, usually a small value. OIM loss measures similarity in cosine space.

We assume that there are several categories P , N_i and a sample $x \in P$. Due to τ and explosion characteristics of exponential function, the little gap between $v_P^T x$ and $v_{N_i}^T x$ will be significantly amplified, as Figure. 3(a). Thence, $\exp(v_P^T x / \tau)$

is great larger than $\exp(v_{N_i}^T x/\tau)$. In normalization of Equation.(6), the positive category will dominate the probability of $x \in P$ owing to the significant gap. As Figure. 3(b), the evolution of positive category similarity has a significant influence, while the evolution of negative categories similarity has little effect. As Equation.(7), the loss is also determined by probability of sample $x \in P$. Thereby, the positive category dominates the loss category. The negative categories is not paid enough attention, making the loss bias to the positive category. Negative categories have trouble playing a guiding role during training. The attention imbalance between positive and negative categories makes it challenging to further separate different IDs. Further, we analyze the insufficient from the view of the gradient. The coefficient of the negative category in the gradient can be formalized as

$$c_j = \frac{\exp(v_j^T x/\tau)}{\sum_{j=1}^L \exp(v_j^T x/\tau) + \sum_{k=1}^Q \exp(u_k^T x/\tau)} \quad (8)$$

where the c_j is the coefficient of the j th negative categories, and all symbols are the same as defined in Equation.(6). The positive category also dominates the coefficient, so the negative categories only cover a tiny fraction of the gradient, even though their similarities are not negligible. Then the negative instances in the circular queue are the same situation. It is necessary to excavate the contribution of the negative categories.

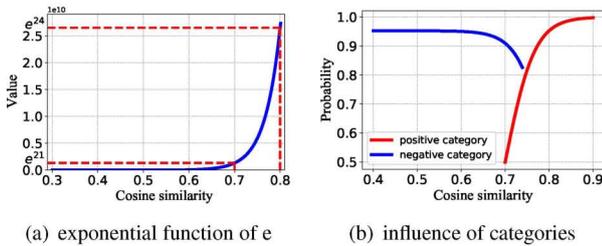


Fig. 3. Illustration on attention imbalance between positive and negative categories in OIM loss. (a) is the exponential function of e with τ taking a shared value of $1/30$. Slight differences are magnified dramatically, and please note the scale on the ordinate is 10^{10} . (b) shows the influence on the probability p_i when the similarity between the sample and different categories changes. The red devotes the influence of the positive category, with the negative categories similarities being 0.3, 0.5 and 0.7. The blue devotes the influence of the negative category, with the positives and negative similarities being 0.8, 0.3 and 0.7, respectively.

To enhance the contribution of negative categories, we propose a novel BOIM Loss. We design a novel item L_{neg} , focusing on the negative categories and generating corresponding loss. However, it is inefficient to take all negative categories into account. When the similarity is low enough, it is unnecessary to optimize negative category distance. Thence,

we only sample the hard negative categories close to the sample. Like Triplet loss [15], we introduce a lower bound since the excessive separation of negative categories may lead to overfitting. The L_{neg} is formulated as follows:

$$L_{neg} = \frac{\sum_j \mathbb{I}_{v_j^T x/v_t^T x > \sigma}(v_j, v_t, x) \max(v_j^T x - \beta, 0)}{\sum_j \mathbb{I}_{v_j^T x/v_t^T x > \sigma}(v_j, v_t, x)} (j \neq t) \quad (9)$$

where, \mathbb{I} is the characteristic function to filtrate hard negative categories, and σ is the threshold. v_j is the j th centre in v , and v_t is the positive category centre. β is the lower bound.

Besides, we design a coefficient w to help centres dispersion. This coefficient measures the density of the centres by the cosine similarity between centres.

$$w = \sum_i^L \sum_j^L \exp(v_i^T v_j + \alpha) \quad (10)$$

where α is the hyperparameter to adjust the coefficient w range. Taking w as the coefficient of OIM loss. When the centres are more concentrated, the loss is more significant, which helps the model to further separate the centres apart.

The proposed BOIM loss function is:

$$L_{BOIM} = wL_{OIM} + L_{neg} \quad (11)$$

The BOIM loss can guide different IDs separation effectively for producing discriminative person embeddings.

4. EXPERIMENT

4.1. Experiment Details

In this section, we perform a thorough evaluation of our methods on two datasets, CUHK-SYSC [7] and PRW [10]. If not specified, we use the gallery size of 100 by default in CUHK-SYSC. We always use the gallery size of 6112 in PRW.

Following the settings in previous works [7, 8], Mean Average Precision (mAP), and Cumulative Matching Characteristics (CMC) are standard metrics to measure person search performance. We employ recall and Average Precision (AP) for pedestrian detection as the performance metrics. A candidate would be considered correct when its IoU to the ground truth bounding box is larger than 0.5.

We implement our method using the Pytorch framework. We use pre-trained ResNet50 [17] as the backbone. We use the SGD optimizer for training the model, and we respectively use 0.9 and $5e-4$ for the momentum and weight decay parameters. We start the training with 0.003 as the learning rate, which is decreased to 0.1 at epochs 14. The images are resized to 900×1500 , and the batch size is set to 5.

Table 1. Analytical experiment results on CUHK-SYSC.

Method	Detection		Re-ID	
	Recall	AP	mAP	top-1
OIM-base [8]	89.3	79.7	84.4	86.1
Flatten	92.5	86.9	91.5	92.6
Parameter-base	92.2	86.2	91.4	92.3
FDPS	92.9	87.2	92.2	93.1

Table 2. Ablation experiments on BOIM.

Method	Detection		Re-ID	
	Recall	AP	mAP	top-1
OIM loss	95.2	92.4	45.1	83.2
+ w	96.0	93.0	45.6	83.4
+ l_{neg}	96.3	92.0	45.7	83.6
BOIM	96.7	93.3	46.7	83.7

4.2. Ablation Study

We perform several experiments to better understand our method. We design the following four variants for an analysis on FDPS. (a) OIM-base [7] is our baseline, generating one shared feature for all sub-tasks. (b) Flatten use GMP to extract feature for Re-ID, and the ROI feature map is flattened for detection. (c) Parameter-base uses 1×1 convolution to aggregate different dimensions of the ROI feature map, providing various features of the same shape as FDPS. (d) FDPS aggregates features based on marginal distribution to generate sub-features. The above four variants are trained on CUHK-SYSC under OIM loss. Table 1 reports the results.

Flatten takes full advantage of all parameters of the ROI feature map, but it does not aggregate the feature and highlight the related clue. For the issue shown as Figure. 1, the experiment shows enlarging the embedding space by increasing feature size has an improvement. Yet, the improvement is minor than FDPS, even though all parameters have been used. Parameter-base has a similar structure to FDPS, but it extracts features using convolution, introducing lots of parameters and calculations. However, Parameter-base cannot separate sub-tasks into different dimensions. Results shows separating sub-tasks is more efficient than simply increasing parameters and depth. As reported in Table 1, our FDPS obtained the best performance. In FDPS, sub-tasks are separated into different dimensions of the ROI feature map, and aggregating the feature highlights the feature required by the sub-task. Meanwhile, FDPS is parameter-free and lightweight.

We analyze BOIM loss and evaluate the performance of its components using FDPS on PRW. Table 2 reports the results. The performance has been improved after introducing l_{neg} or w . And the best performance is achieved, when l_{neg} and w are used together. The distribution of person embeddings is visualized for an intuitive understanding in Fig-

ure 4. All IDs are from the probe of testset on PRW. As Figure 4(a), several categories fit tightly, making it hard to distinguish them. While, there is a clear division between categories under BOIM loss, as Figure 4(b). The distributions in Figure 4(c) show that different category centres similarities are lower under BOIM, meaning different categories are more dispersed. These experiments show the model further separate different IDs with the help of BOIM.

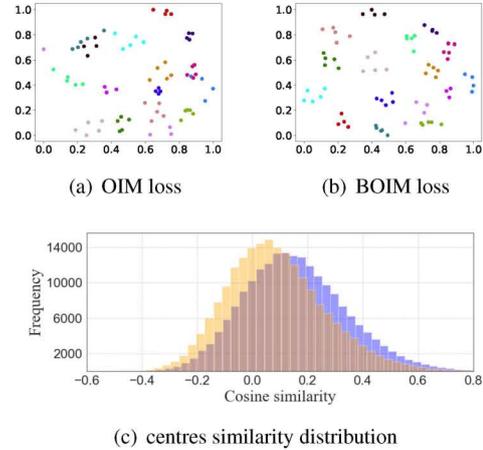


Fig. 4. Illustration on person embeddings distribution. In (a) and (b), we visualize the person embeddings of several categories under OIM and BOIM loss, respectively. In (c), we made statistics of the similarity among category centres. The orange devotes similarity distribution under BOIM loss, and the blue denotes one under OIM loss.

4.3. Comparison to the State-of-the-arts

In this section, we compare our method to the state-of-the-art person search methods. We report the quantitative results of our method in Table 3. The table also summarizes the results of other top methods, according to their search strategy.

As shown in Table 3, our method outperforms all other one-step methods, especially dramatically improving the more challenging PRW, outperforming the NAE+ [8] by 2.7% and 2.6% in mAP and top-1 accuracy, respectively. The two-step methods employ two standalone models, consuming a significant amount of time and computation. Besides, they are non-end-to-end, requiring additional manual intervention. Our end-to-end method has a straightforward structure with less parameter, computational and time overhead. Also, our model outperforms most two-step models. This profits from that FDPS makes the model separate the different sub-tasks into different dimensions. Meanwhile, our method can distinguish different people effectively with the help of BOIM.

We evaluate the model performance under different gallery sizes as defined in [8]. As shown in Figure 5, the mAP for all methods decrease monotonically as the increase

Table 3. Comparison with state-of-the-arts. One-step methods are gathered in the upper block, while the lower block shows the results of two-step methods.

Methods		CUHK-SYSC		PRW	
		mAP	top-1	mAP	top-1
one-step	OIM [7]	75.5	75.7	21.3	49.9
	QEEPS [18]	88.9	89.1	37.1	76.7
	HOIM [14]	89.7	90.8	39.8	80.4
	APNet [19]	88.9	89.3	41.9	81.4
	BINet [9]	90.0	90.7	45.3	81.7
	NAE [8]	91.5	92.4	43.3	80.9
	NAE+ [8]	92.1	92.9	44.0	81.1
	FDPS(ours)	92.2	93.1	45.1	83.3
	+BOIM(ours)	92.5	93.7	46.7	83.7
two-step	CLSA [6]	87.2	88.5	38.7	65.0
	RDLR [5]	93.0	94.2	42.9	70.2
	IGPN [20]	90.3	91.4	47.2	87.0
	TCTS [11]	93.9	95.1	46.8	87.5

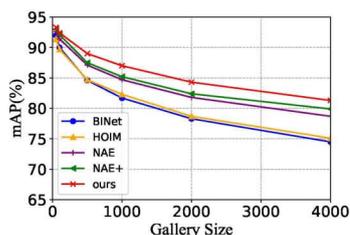


Fig. 5. Performance comparison of one-step methods on CUHK-SYSC with varying gallery sizes.

of gallery size. The mAP of our method is consistently better than other one-step methods.

5. CONCLUSION

We propose a novel person search method, which decomposes the ROI to extract sub-features for sub-tasks based on marginal distribution. Our FDPS derives its strength from separating sub-tasks to alleviate the contradiction between detection and Re-ID. FDPS is parameter-free and lightweight. We demonstrate OIM loss focus imbalance on positive and negative categories. Thence we propose a Balance Online Instance Match loss to increase concern for negative categories. Extensive experiments confirm our method can significantly improve the performance of end-to-end methods.

6. ACKNOWLEDGEMENT

This work was supported by the National Key R&D Program of China (Grant No.2018YFB2100603) and the National Natural Science Foundation of China (Grant No.61872024). The

authors would like to thank the anonymous reviewers for their critical and constructive comments and suggestions.

7. REFERENCES

- [1] Yuanlu Xu, Bingpeng Ma, and et al., “Person search in a scene by jointly modeling people commonness and person uniqueness,” *MM*, pp. 937–940, Nov. 2014.
- [2] Wanli Ouyang and et al., “Joint deep learning for pedestrian detection,” *ICCV*, pp. 2056–2063, Dec. 2013.
- [3] Ejaz Ahmed, Michael J. Jones, and Tim K. Marks, “An improved deep learning architecture for person re-identification,” *CVPR*, pp. 3908–3916, Jun. 2015.
- [4] Di Chen and et al., “Person search via a mask-guided two-stream cnn model,” *ECCV*, pp. 764–781, Oct. 2018.
- [5] Han and et al., “Re-id driven localization refinement for person search,” *ICCV*, pp. 9813–9822, Nov. 2019.
- [6] Xu Lan, Xiatian Zhu, and et al., “Person search by multi-scale matching,” *ECCV*, pp. 553–569, Oct. 2018.
- [7] Tong Xiao, Shuang Li, Bochao Wang, and et al., “Joint detection and identification feature learning for person search,” *CVPR*, pp. 3376–3385, Jul. 2017.
- [8] Di Chen, Shanshan Zhang, and et al., “Norm-aware embedding for efficient person search,” *Int J Comput Vis*, vol. 129, pp. 3154–3168, Sept. 2021.
- [9] Wenkai Dong and et al., “Bi-directional interaction network for person search,” *CVPR*, pp. 2836–2845, 2020.
- [10] Liang Zheng and et al., “Person re-identification in the wild,” *CVPR*, pp. 3346–3355, Jul. 2017.
- [11] Cheng Wang, Bingpeng Ma, Hong Chang, and et al., “Tcts: A task-consistent two-stage framework for person search,” *CVPR*, pp. 11952–11961, 2020.
- [12] Shaoqing Ren, Kaiming He, and et al., “Faster r-cnn: Towards real-time object detection with region proposal networks,” *NeurIPS*, pp. 91–99, 2015.
- [13] Yichao Yan, Qiang Zhang, and et al., “Learning context graph for person search,” *CVPR*, pp. 2158–2167, 2019.
- [14] Di Chen, Shanshan Zhang, Wanli Ouyang, and et al., “Hierarchical online instance matching for person search,” *AAAI*, vol. 34, pp. 10518–10525, Apr. 2020.
- [15] Alexander Hermans and et al., “In defense of the triplet loss for person re-identification,” *arXiv preprint arXiv:1703.07737*, 2017.
- [16] Yandong Wen, Kaipeng Zhang, and et al., “A discriminative feature learning approach for deep face recognition,” *ECCV*, pp. 499–515, Sept 2016.
- [17] Kaiming He and et al., “Deep residual learning for image recognition,” *CVPR*, pp. 770–778, Jun. 2016.
- [18] Bharti Munjal and et al., “Query-guided end-to-end person search,” *CVPR*, pp. 811–820, 2019.
- [19] Yingji Zhong and et al., “Robust partial matching for person search in the wild,” *CVPR*, pp. 6826–6834, 2020.
- [20] Wenkai Dong and et al., “Instance guided proposal network for person search,” *CVPR*, pp. 2582–2591, 2020.