2021 13th International Conference on Wireless Communications and Signal Processing (WCSP) | 978-1-6654-0785-4/21/$31.00 ©2021 IEEE | DOI: 10.1109/WCSP52459.2021.9613664

Efficient Person Re-Identification with Multi-Scale Feature Fusion

Jia Lin*, Suofei Zhang†, Jiangping Chen*, Xiaofu Wu*, Quan Zhou*
*College of Telecommunications and Information Engineering
†School of Internet of Things
Nanjing University of Posts and Telecommunications, Nanjing, China
Email: 13968527371@163.com, (zhangsuofei, 1220013108, quan.zhou)@njupt.edu.cn, xfuwu@ieee.org

***Abstract*—The task of Person Re-Identification (Re-ID) has attracted growing attention in recent years. Most State-Of-The-Art (SOTA) methods often employ the large-scale ResNet model as their backbone, which makes it tedious to explore various architecture modifications. In this study, we propose a small-sized Re-ID model with EfficientNet as its backbone, along with a novel feature pyramid branch for learning multi-scale features. Extensive experimental results show that the proposed network outperforms various SOTA methods with obvious margin on standard benchmark datasets such as Market1501, DukeMTMC-Re-ID, CUHK03, even though its mode only contains about 7.41M parameters.**

***Index Terms*—Person re-identification, multi-scale feature learning, feature pyramid branch, deep learning.**

I. Introduction

Person re-identification (Re-ID), a subproblem of image retrieval, aims to search people across non-overlapping surveillance camera views deployed at different locations by matching person images. Despite of the exciting progress in recent years, person Re-ID remains to be extremely challenging in practical unconstrained scenarios. Common challenges arise from body misalignment, occlusion, background perturbance, view point changes, pose variations and noisy labels, among many others [1], [2].

We argue that a cost-effective Re-ID model should be computationally efficient, capable of running on low-resolution video input, and robust to multiple camera setting [3]. Hence, we propose an Efficient Person Re-ID Network (EPRI-Net) achieving SOTA performance under these practical constraints. To reduce the computational burden, we aim to decrease the number of parameters and use a relatively small Re-ID model. Fig. 1 shows the current SOTA results and their model sizes compared to our proposed method on the popular Market1501 dataset in terms of rank-1 accuracy and mAP. As shown, EPRI-Net achieves SOTA results with an order of magnitude smaller model compared to the best existing Re-ID CNN. The main contributions of EPRI-Net are summarized as follows:

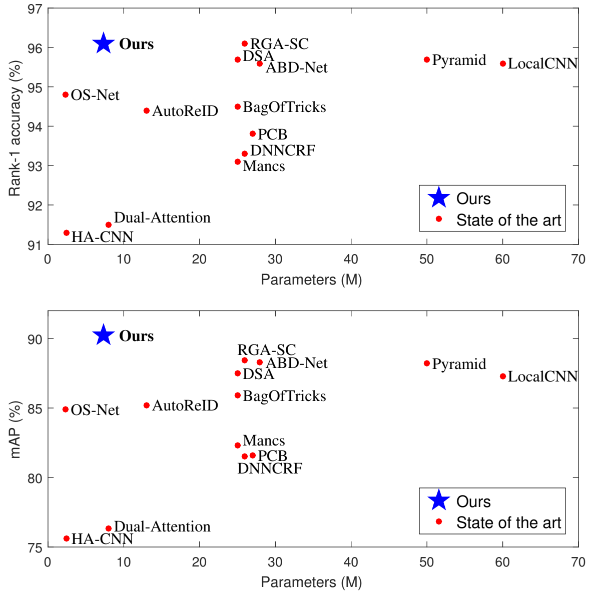

Fig. 1. Performance comparison of our approach and SOTA Re-ID methods on Market1501 dataset. Top: Rank-1 accuracy vs. number of parameters. Bottom: mAP vs. number of parameters.

- We propose a light-weighted Re-ID architecture, namely EPRI-Net, by integrating EfficientNet-b0 based backbone with a Pixel-level Feature Pyramid Branch as affiliated structure. Unlike conventional multi-branch architectures, the final output feature of EPRI-Net is delivered by the element-wise multiplication of the feature tensors from the backbone and FPB, rather than simple concatenation.
- Extensive experimental results on standard Re-ID benchmarks including Market-1501, DukeMTMC-Re-ID, CUHK03, MSMT17 demonstrate that, EPRI-Net can significantly outperform existing methods via a quite efficient model with less than 7.5M parameters.

II. Related Works

Many Re-ID models now use a large-scale model as a backbone. To further improve the performance, various mechanisms, including attention modules and diverse branches, were incorporated into the backbone. Many part-based methods have achieved superior performance and the success of the attention mechanism [4] in the field of

978-1-6654-0785-4/21/$31.00 ©2021 IEEE

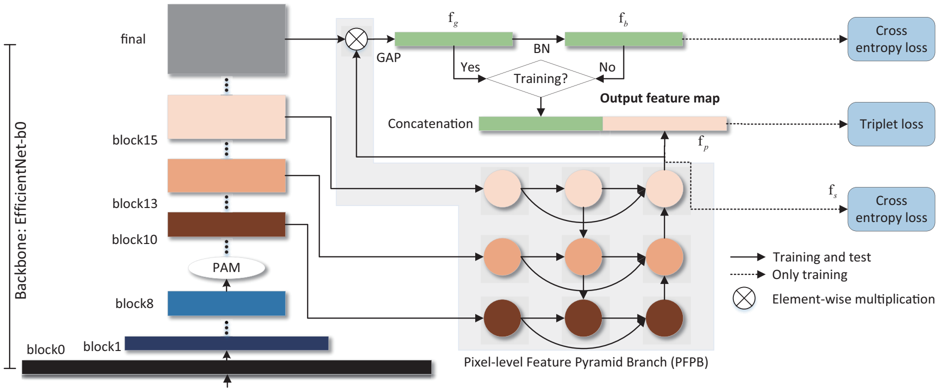

Fig. 2. The overall architecture of our proposed method. The output of FPB will multiply the final feature map of backbone element by element, meanwhile the output will use cross entropy loss and triplet loss.

image classification has been borrowed into the person Re-ID field. Dual Attention Network (DANet) [5] was proposed to capture rich contextual dependencies based on the self-attention mechanism.

As person Re-ID research continues to advance, light-weighted Re-ID models are becoming more and more welcome for real-time applications. EfficientNet [6] proposed a viable light-weighted model solution for the purpose of Re-ID, whose size could be very small (about 4.7M). Harmonious attention network (HA-CNN) [7] formulated a light-weighted yet deep CNN architecture by devising a holistic attention mechanism for locating the most discriminative pixels and regions. In OSNet [8], a novel unified aggregation gate was introduced to dynamically fuse multi-scale features with input-dependent channel-wise weights, the building block of OSNet uses pointwise and depthwise convolutions, for keeping OSNet scalable in its complexity.

III. Methods

A. Proposed Network Architecture

As depicted in Fig. 2, EPRI-Net has two branches, a global branch and a feature pyramid branch. Since the introduction of ResNet, most of the researches have used ResNet as the backbone and various multi-branch architectures are constructed for achieving the improved performance in the field of image retrieval. The main breakthrough of ResNet is the residual block, but due to the large number of convolutional operations with 3 × 3 convolutional kernels and the final number of channels is also extremely large (2048). For example, ResNet-50 has about 25.05M parameters. In real-world applications, we often need to reduce the number of model parameters and the size of floating-point operations to ensure the real-time performance. Instead of using ResNet as the backbone, we employ EfficientNet-b0 as an alternate but light-weighted backbone. Compared to the well-known ResNet, EfficientNet uses a large number of 1 × 1 kernel convolutions, which leads to reduced model size. In the meantime, EfficientNet systematically investigates model scaling and the balancing among network depth, width, and resolution for achieving excellent performance. Our experiments show that EfficientNet-b0 has only 4.97M parameters, and ResNet-50 has 4.7 times more parameters than EfficientNet-b0.

As a multi-branch network, EPRI-Net adopts the feature pyramid branch in FPB [9] by borrowing the idea of the feature pyramid network from prevailing object detection methods. Different from the feature pyramid branch proposed in FPB [9], we propose to employ three layers of feature maps instead of two layers, and we employ a novel element-by-element multiplication for fusing the features from the global branch and the FPB.

B. Pixel-level Feature Pyramid Branch

As shown in Fig. 2, our proposed Pixel-level Feature Pyramid Branch (PFPB) has three layers. The feature maps extracted from the backbone pass through a set of lateral convolutional filters, before inputing to each layer of the PFPB. These lateral convolutional filters consist of standard 2D convolutional filters with 1 × 1 kernel followed by Batch Normalization and Swish activation function. EPRI-Net employs six convolutional filters with 1 × 1 kernel for achieving the aggregation of features at different scales. There exist a top-down conection and a down-top connection, keeping the same shape of channels going forward without any downsampling or upsampling operations. As the same as ResNet, the feature pyramid branch in EPRI-Net has three residual connections. At last, the output shape of PFPB is $B \times 320 \times 24 \times 12$ where B denotes the batch size.

The design of the EPRI-Net is inspired by the feature pyramid branch in FPB [9]. The feature ($\boldsymbol{f}_i^b$) from the backbone is concatenated with the feature ($\boldsymbol{f}_i^s$) from the PFPB, which feeds to hard mining triplet loss function. The proposed PFPB has some differences with FPB and traditional feature pyramid networks [10]. Compared to the traditional feature pyramid network, EPRI-Net aggregates features at a single output before average pooling operation, rather than multiple outputs at each layer. Secondly, PFPB differs with FPB in that its output from feature pyramid branch is further divided into two sub-branches, one of which performs the convolution with the output of the feature pyramid branch to ensure the same shape as the last layer of backbone's feature map, and then multiply it element by element into the last layer of backbone's feature map, while the other sub-branch performs the global average pooling.

C. Position Attention Module

As shown in Fig. 2, we use Position Attention Module (PAM) in EPRI-Net. PAM is designed to capture and aggregate those semantically related *pixels* in the spatial domain. We depict the structure of PAM in Fig.3. The input feature maps $\mathbf{A} \in \mathbb{R}^{C\times H\times W}$ are first fed into convolution layers with batch normalization and ReLU activation to produce feature maps $\mathbf{B}, \mathbf{C}, \mathbf{D} \in \mathbb{R}^{C\times H\times W}$. Then we compute the pixel affinity matrix $\mathbf{S} \in \mathbb{R}^{N\times N}$ where $N = H \times W$. Note that the dimensions of S and X are different, since the former computes correlations between the total N pixels rather than C channels.

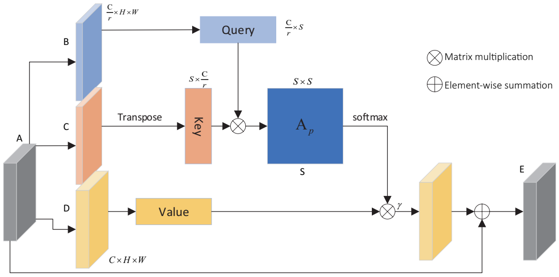

Fig. 3. Position Attentive Module

D. Loss Function

During training process, given features from sample $(\boldsymbol{x}_i, \boldsymbol{y}_i)$ we employ the following loss:

$$\begin{aligned}\mathcal{L}_{total} =& \alpha \cdot \mathcal{L}_{ce}(\boldsymbol{W}^b \boldsymbol{f}_i^b, \boldsymbol{y}_i) \\ &+ \beta \cdot \mathcal{L}_{ce}(\boldsymbol{W}^s \boldsymbol{f}_i^s, \boldsymbol{y}_i) \\ &+ \lambda \cdot \mathcal{L}_{tri}(\boldsymbol{f}_i^g \odot \boldsymbol{f}_i^p, \boldsymbol{y}_i, \boldsymbol{f}_j^g \odot \boldsymbol{f}_j^p, \boldsymbol{y}_j)\end{aligned} \quad (1)$$

where $\boldsymbol{W}^b$ and $\boldsymbol{W}^s$ is the weight of classifier in backbone and the feature pyramid branch, respectively. $\boldsymbol{W}^b \boldsymbol{f}_i^b$ and $\boldsymbol{W}^s \boldsymbol{f}_i^s$ are the output feature vectors of the EfficientNet-b0 and the feature pyramid branch, respectively. $\boldsymbol{y}_i$ is the true annoation value. $\boldsymbol{f}_i^g \odot \boldsymbol{f}_i^p$ denotes the Euclidean distance between $\boldsymbol{f}_i^g$ and $\boldsymbol{f}_i^p$. $\mathcal{L}_{ce}(\cdot)$ and $\mathcal{L}_{tri}(\cdot)$ is the cross entropy loss and the hard mining triplet loss, respectively.

The hyper-parameters α, β and λ are used to adjust the ratio between the three lossed.

IV. Experimental Results

To evaluate EPRI-Net, we conducted experiments on three large-scale person Re-ID datasets: Market-1501, DukeMTMC-Re-ID, CUHK03 and MSMT17. Firstly, we peform a series of ablation studies on Market-1501 for the proposed EPRI-Net. Then, we compare the performance of EPRI-Net against existing SOTA methods on all three datasets.

A. Datasets

Four popular person Re-ID datasets are considered, including Market1501, DukeMTMC-Re-ID, CUHK03 and MSMT17.

Market1501 consists of 32,668 images from 1501 identities captured by six cameras, in which each identity is at least captured by two cameras with multiple images. For the training set, 12,936 images from 751 identities are considered, leading to an average of 17.2 training samples for one person. For the testing set, 19,732 images from 750 other identities are considered, in which 3,368 images are used as probe set while the rest are used as gallery set.

DukeMTMC-Re-ID consists of 36,411 images from 1,404 identities captured by more than two cameras, and 408 identities captured by only one camera as distractors. For the training set, 16,522 images from 702 identities are considered. For the testing set, 17,661 images from 702 other identities are considered, in which 2,228 images are used as probe set while the rest images from the 702 identities as well as distractors are used as gallery set.

CUHK03 consists of images from 1467 identities captured by five cameras, in which 767 identities are used as training set and 700 other identities are used as testing set. The dataset contains two tasks, person Re-ID with labeled images and with detected images. The labeled dataset has 7,368 images for training and 6,728 images for testing. The detected dataset has 7,365 images for training and 7,732 images for testing.

MSMT17 is a large-scale dataset. It consists of 126,441 images from 4,101 identities captured by a 15-camera network (12 outdoor, 3 indoor). For the training set, 32,621 images from 1,041 identities are considered. For the testing set, 93,820 images from 3,060 other identities are considered, in which 11,659 images are used as probe set while the rest are used as gallery set.

B. Implementation Details

We employ various data augmentation in training, including random flip, random erase [22], random crop and random patch. Instead of using the input size of 128×64, the input size of 384×192 is employed for all datasets. For the backbone, the pre-trained model from ImageNet is used as its initialization. The weight decay parameter for regularization is set to 1.625×10^{-4}. In (1), the α and β and γ are assigned values of 1.0, 1.0 and 6.375, respectively.

During training, we fine-tuned our proposed model for 120 epochs with a batch size of 64 from 16 identities, namely, four samples were randomly selected from datasets for each identity in a training batch. We use the *Adam* optimizer and the learning rate is set to 3.5×10^{-6} at the beginning of training, then increasing it to 3.5×10^{-5} with a linear warm-up strategy in the first 20 epochs. Then, the learning rate is decayed at the 60 and 90 epoch with a rate of 0.1, respectively. At the first

TABLE I
Comparison of our proposed method with SOTA methods on Market1501, DukeMTMC-Re-ID, CUHK03 (Labeled) and CUHK03 (Detected). The best performances are highlighted by bold font.

Method	Market1501		DukeMTMC-Re-ID		CUHK03-Labeled		CUHK03-Detected	
	mAP(%)	rank-1(%)	mAP(%)	rank-1(%)	mAP(%)	rank-1(%)	mAP(%)	rank-1(%)
KPM [11]	75.3	90.1	63.2	80.3	-	-	-	-
PCB+RPP [12]	81.6	93.8	69.2	83.3	-	-	57.5	63.7
Mancs [13]	82.3	93.1	71.8	84.9	63.9	69.0	60.5	65.5
MGN [14]	86.9	95.7	78.4	88.7	67.4	68	66.0	68.0
MHN [15]	85.0	95.1	77.2	89.1	72.4	77.2	65.4	71.7
CAMA [16]	84.5	94.7	72.9	85.8	66.5	70.1	64.2	66.6
Bag-Of-Tricks [17]	85.9	94.5	76.4	86.4	-	-	-	-
ABD-Net [1]	88.28	95.6	78.59	89.0	-	-	-	-
BDB [18]	86.7	95.3	76.0	89.0	76.7	79.4	73.5	76.4
Pyramid [19]	88.2	95.7	79.0	89.0	76.9	78.9	74.8	78.9
SONA [20]	88.67	95.68	78.05	89.25	79.23	**81.85**	76.35	79.1
RGA-SC [21]	88.4	**96.1**	-	-	77.4	81.1	74.5	79.6
EPRI-Net	**90.2**	**96.1**	**82.6**	**90.3**	**80.9**	81.6	**81.2**	**82.4**

10 epochs, only the triplet loss is employed and after the 10th epoch, the cross entropy loss starts to contribute to the training. All experiments are executed with a hardware environment as *Intel(R) Xeon(R) CPU E5-2690 v4 @ 2.60GHz* and a **double** *NVidia Tesla P100 GPU.*

Evaluation: For quantitative comparison over different methods, we consider the Cumulative Matching Characteristics (CMC) at mean Average Precision (mAP) and top-1 accuracy (rank-1) as standard metrics. All results are obtained without any re-ranking or multi-query fusion techniques.

C. Ablation Study

To verify the effects of attention modules and the PFPB in EPRI-Net, we perform some ablation experiments over Market-1501. We use EfficientNet-b0 as the baseline, where the sum of the cross entropy loss and the triplet loss is employed during training. Four variants are then constructed on top of the baseline: a) baseline + Feature Pyramid Branch; b) baseline + PFPB; c) baseline + PAM; d) baseline + PFPB + PAM (**EPRI-Net**). In c), we adopt the setting of $\alpha = 1$, $\beta = 1$ and $\gamma = 0$ in (1), which means that means the PFPB is not applied; For a), b), d) the $\alpha = 1$, $\beta = 1$ and $\gamma = 6.375$.

Table II presents the experimental results, from which several observations could be drawn:

- The use of the PFPB could lead to 1.0% (mAP) improvement over baseline, and the proposed PFPB outperforms the feature pyramid branch that improves 1.7% (mAP).
- The use of attention modules is helpful and lead to 0.3% (mAP) improvement over baseline.
- By combining both attention modules and PFPB. EPRI-Net eventually achieves the mAP and rank-1 of 90.2% and 96.1% respectively.

D. Comparison with SOTA Methods

In this section, we compare our proposed EPRI-Net with other SOTA methods. In Table I, we list the performance of different methods on four tasks, Market1501, DukeMTMC-Re-ID, CUHK03 (Labeled) and CUHK03 (Detected). One can see that our proposed scheme outperforms SOTA methods with obvious margin. Note that the performance on Market1501 is nearly saturated and there is no much room for further improvement. For the mAP, EPRI-Net substantially exceeds the second best approach by 0.87%, 1.5%, 0.03% and 4.03%.

TABLE II
Impact of each component. The baseline only use EfficientNet-b0 achieved 88.4 mAP on Market-1501. The EPRI-Net is achieving 90.2% mAP on Market-1501 with PAM and PFPB.

Method	params(M)	Market-1501	
		mAP(%)	rank-1(%)
baseline	4.97	88.4	94.7
baseline + FPB	6.56	89.4	95.6
baseline + PFPB	7.39	90.1	96.0
baseline + PAM	4.99	88.7	95.2
EPRI-Net	**7.41**	**90.2**	**96.1**

TABLE III
Comparison of our proposed method with SOTA methods on MSMT17. The best performances are highlighted by bold font.

Method	mAP(%)	rank-1(%)	rank-5(%)
PDC [23]	29.7	58.0	73.6
GLAD [24]	34.0	61.4	76.8
IANet [25]	46.8	75.5	85.5
BFE [25]	51.5	78.8	89.1
DGNet [26]	52.3	77.2	87.4
OSNet [8]	52.9	78.7	-
ABD-Net [1]	60.8	**82.3**	**90.6**
RGA-SC [21]	57.5	80.3	-
EPRI-Net	**63.4**	78.3	87.1

MSMT17: For the more challenging large-scale person Re-ID dataset MSMT17, we compare different configurations of our proposed EPRI-Net with other SOTA methods in Table III. For the mAP, EPRI-Net reaches the SOTA and substantially exceeds the second best approach by 1.2%. The rank-1 accuracy, however, is not satisfied.

V. Conclusion and future work

This paper proposes an Efficient Person Re-ID Network (EPRI-Net) to learn more representative, robust, discriminative features for person Re-ID. EPRI-Net demonstrates its SOTA performance through extensive experiments. In the future, it is interesting to explore the EPRI-Net for the use in other computer vision tasks.

References

[1] Tianlong Chen, Shaojin Ding, Jingyi Xie, Ye Yuan, Wuyang Chen, Yang Yang, Zhou Ren, and Zhangyang Wang. Abd-net: Attentive but diverse person re-identification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019.

[2] Longhui Wei, Shiliang Zhang, Wen Gao, and Qi Tian. Person transfer gan to bridge domain gap for person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[3] Hussam Lawen, Avi Ben-Cohen, Matan Protter, Itamar Friedman, and Lihi Zelnik-Manor. Attention network robustification for person reid. 10 2019.

[4] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need, 2017.

[5] Jun Fu, Jing Liu, Haijie Tian, Yong Li, Yongjun Bao, Zhiwei Fang, and Hanqing Lu. Dual attention network for scene segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[6] Mingxing Tan and Quoc Le. EfficientNet: Rethinking model scaling for convolutional neural networks. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 6105–6114. PMLR, 09–15 Jun 2019.

[7] Wei Li, Xiatian Zhu, and Shaogang Gong. Harmonious attention network for person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[8] Kaiyang Zhou, Yongxin Yang, Andrea Cavallaro, and Tao Xiang. Omni-scale feature learning for person re-identification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019.

[9] Suofei Zhang, Zirui Yin, Xiofu Wu, Kun Wang, Quan Zhou, and Bin Kang. Fpb: Feature pyramid branch for person re-identification, 2021.

[10] Tsung-Yi Lin, Piotr Dollar, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[11] Yantao Shen, Tong Xiao, Hongsheng Li, Shuai Yi, and Xiaogang Wang. End-to-end deep kronecker-product matching for person re-identification, 2018.

[12] Yifan Sun, Liang Zheng, Yi Yang, Qi Tian, and Shengjin Wang. Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline). In *Proceedings of the European Conference on Computer Vision (ECCV)*, September 2018.

[13] Cheng Wang, Qian Zhang, Chang Huang, Wenyu Liu, and Xinggang Wang. Mancs: A multi-task attentional network with curriculum sampling for person re-identification. In *Proceedings of the European Conference on Computer Vision (ECCV)*, September 2018.

[14] Guanshuo Wang, Yufeng Yuan, Xiong Chen, Jiwei Li, and Xi Zhou. Learning discriminative features with multiple granularities for person re-identification. *Proceedings of the 26th ACM international conference on Multimedia*, Oct 2018.

[15] Binghui Chen, Weihong Deng, and Jiani Hu. Mixed high-order attention network for person re-identification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019.

[16] Wenjie Yang, Houjing Huang, Zhang Zhang, Xiaotang Chen, Kaiqi Huang, and Shu Zhang. Towards rich feature discovery with class activation maps augmentation for person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[17] Hao Luo, Youzhi Gu, Xingyu Liao, Shenqi Lai, and Wei Jiang. Bag of tricks and a strong baseline for deep person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2019.

[18] Zuozhuo Dai, Mingqiang Chen, Xiaodong Gu, Siyu Zhu, and Ping Tan. Batch dropblock network for person re-identification and beyond. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019.

[19] Feng Zheng, Cheng Deng, Xing Sun, Xinyang Jiang, Xiaowei Guo, Zongqiao Yu, Feiyue Huang, and Rongrong Ji. Pyramidal person re-identification via multi-loss dynamic training. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[20] Bryan (Ning) Xia, Yuan Gong, Yizhe Zhang, and Christian Poellabauer. Second-order non-local attention networks for person re-identification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019.

[21] Zhizheng Zhang, Cuiling Lan, Wenjun Zeng, Xin Jin, and Zhibo Chen. Relation-aware global attention for person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[22] Zhun Zhong, Liang Zheng, Guoliang Kang, Shaozi Li, and Yi Yang. Random erasing data augmentation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(07):13001–13008, Apr. 2020.

[23] Chi Su, Jianing Li, Shiliang Zhang, Junliang Xing, Wen Gao, and Qi Tian. Pose-driven deep convolutional model for person re-identification. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

[24] Longhui Wei, Shiliang Zhang, Hantao Yao, Wen Gao, and Qi Tian. Glad. *Proceedings of the 25th ACM international conference on Multimedia*, Oct 2017.

[25] Ruibing Hou, Bingpeng Ma, Hong Chang, Xinqian Gu, Shiguang Shan, and Xilin Chen. Interaction-and-aggregation network for person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[26] Zhedong Zheng, Xiaodong Yang, Zhiding Yu, Liang Zheng, Yi Yang, and Jan Kautz. Joint discriminative and generative learning for person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.