2020 International Conference on Machine Learning and Cybernetics (ICMLC) | 978-1-6654-1943-7/20/$31.00 ©2020 IEEE | DOI: 10.1109/ICMLC51923.2020.9469545

DUAL-PATH MODEL FOR PERSON RE-IDENTIFICATION UNDER CLOTH CHANGING

JUNHAO ZHENG, XIAOMAN HU, TIANYI XIANG, PATRICK P. K. CHAN

School of Computer Science and Engineering, South China University of Technology, Guangzhou, China
E-MAIL: zhengjunhao2hqf@163.com, JacquelineHxm@outlook.com, xty435768@gmail.com, patrickchan@scut.edu.cn

Abstract:

Most of the existing person re-identification (ReID) methods relies heavily on a person's clothes since clothing information is the clear and remarkable visual feature when the face of a person is unclear. However, in reality, people does not always wear the same cloth across camera views. Even worse, an adversary may change the clothes aiming to evade the identification. Some studies confirms that clothes changing downgrades the existing ReID methods significantly. The current ReID method considering clothes-changing does not fully utilize the person discriminant features, which may reduce its accuracy. This paper presents a dual-path model to learn the robust features under clothes changing and also the discriminant features for ReID from a RGB image and its contour sketch image respectively. The appearance and shape features of a person extracted by the two branches of our model are then combined to make a decision. The clothing information is eliminated from the appearance features by encouraging the similarity between the learned appearance and shape features. The experimental results on the PRCC dataset demonstrate that our model achieves higher performance under clothes changing compared to state-of-the-art ReID methods.

Keywords:

Person Re-Identification; Dual-path model; Metric learning; Feature learning

1. Introduction

Person re-identification (ReID) aims to recognize a target person among a large number of pedestrian images across different camera views. Given a query image of a person, a ReID model returns the images of the person identity who are the most similar to the query image. Early researches of ReID mainly concentrate upon hand-crafted feature extraction [1, 2] and similarity metric design [2, 3]. Deep learning based ReID, which learns discriminative feature representation utilizing deep neural network (DNN), significantly improves performance of ReID models. ReID can be formulated as a multi-class classification task, where neural network learns to extract features from person images by minimizing classification loss, *e.g.*softmax loss. Some researches [4, 5] utilize metric learning aims to learn a similarity measure by minimizing the distance between the images of the same person, and maximizing the ones of different persons

The current ReID methods [6] mainly focus on learning robust feature representations against the environmental noise, for example, human pose, image angle, and background environment. Those methods implicitly assume that the clothes of all people are the same, *i.e.*the cloth-changing is not considered as an intra-class variation factor. The image quality in ReID is usually low, *e.g.*Market-1501 [7], CUHK03 [8], and DUKEMTMC-reID [9]. A person image is in low resolution and vague. As the clothing information is the most clear and remarkable visual feature of a person, the clothing information usually plays an important role in the decision of a ReID model. Although training a ReID model by using a dataset containing persons of the same identity has different wearing, *e.g.*CelebreID [10], PRCC [11], LTCC [12], may reduce the domination of clothing information in decision. However, clothing information still has large interference and the traditional methods achieves bad performance under clothes changing [13]. The ReID method considering clothes-changing [11] applies contour sketch and texture of person images to reduce the influence of clothing color in learning. However, some useful information for identification has not been fully utilized by this method.

As the original RGB image contain discriminative appearance information of a person and a sketch image provides information independent to clothing, a dual-path model considering both an original image and its sketch image is proposed in this study. The network architecture is shown in Figure 1. Two independent branches, which are the appearance and shape extractors, learn the appearance and shape features respectively

978-1-6654-1943-7/20/$31.00 ©2020 IEEE

from RGB and sketch channels of an image. The feature representation of a person is extracted from the two branches. To further reducing the importance of clothing information from the appearance features learned from a RGB image, a ranking loss is applied to minimize the distance between the appearance and shape features. Our model shows its superiority comparing to other state-of-the-arts ReID methods on the PRCC dataset experimentally.

The rest of the paper is arranged as follows. Section 2 introduces the previous work related to our model. The details of our model are discussed in Section 3, and Section 4 reports and discusses the experimental results. The conclusion is given Section 5.

2. Related work

2.1. Person Re-Identification

The traditional ReID models mainly focus on learning on appearance representation descriptors [1, 2] that extract the appearance features such as texture and color, and similarity metrics [2, 3] that learn similarities between person images. Deep learning based methods are developed to learn a feature space, giving a boost to the accuracy of ReID tasks. One of the main challenges of ReID is the intra-class variation, including the change of angles and posture, and cluttered or blocked backgrounds.

Most state-of-the-arts ReID models perform part-level feature learning on input image to learn discriminative information with finer granularity. PCB [14] obtains the comprehensive descriptor from several part level features for ReID task pedestrian matching. SCPNet [15] uses the local features to supervise the global features and to improve the performance of both full persons and partial persons identification. Multiple Granularity Network (MGN) [16], an end-to-end architecture with triple branches, separates the origin image into different number of local stripes to train more fine-grained discriminative features. This model reach excellent performance in ReID without cloth changing. However, since they are not designed to extract cloth-irrelevant identity features, clothing information dominates the decision making.

2.2. Person re-identification under cloth-changing

There have been ways to focus on the challenge of clothes changing. RGB-d [17] is the first ReID dataset to propose the use of additional depth information to deal with clothing change. However, this method is only suitable for the indoor environment. The expensive depth map limits the generalization to general scenarios. More recent researches on cloth-changing ReID focus on clothing information elimination. Considering that body shape of a person is independent on clothing, contour sketch extracted from person image is used to train ReID model [11]. The clothes information are directly discarded from the network input. However, the contour sketch person images do not contain sufficient information to identify a person. Other studies apply the idea of disentangling cloth features from identity features to handle cloth changing problem in ReID. ReIDCaps [18] replaces traditional scalar neurons by vector neural capsules to represent identity and clothing features in different dimensions to separate the two features. CESD [12] extracts the identity and cloth features in two branches and utilizes the attention mechanism to disentangle the features. CASE-Net [19] extracts the color and structural features separately and enhances the training by image reconstruction using a generative adversarial network (GAN), encouraging the structural features used for identification without clothes color. However, it is difficult to guarantee that extracted the identity and cloth features are mutually exclusive. Identity features may still contain cloth-relevant information, while identity-related information may also removed from the clothes features.

Some researchers utilize cues additional to original person images for ReID under cloth changing. Huang [10] uses body part images as an additional input. A two-step fine-tuning strategy on body parts is raised to combine local and global image blocks for feature learning. The body-part matching could help the model focus on fine-grained features other than clothing. However, there is no guarantee that cloth information can be neglected. 3APF [13] applies the face detector to crop down one's face to learn the face feature specifically as well as the holistic feature from a full-body image, emphasizing that face could be a cloth-irrelevant discriminative feature for ReID under cloth changing. Nevertheless, it is difficult to obtain a face from blurry person images that is clear enough and with appropriate view angle for face feature learning. In our model, contour sketch images obtained following [11] are used as additional input to highlight body shape information in network training.

3. Proposed method

Our paper proposes a dual-path network using RGB images and their corresponding sketch images. The model contains two branches containing the appearance extractor E_A and

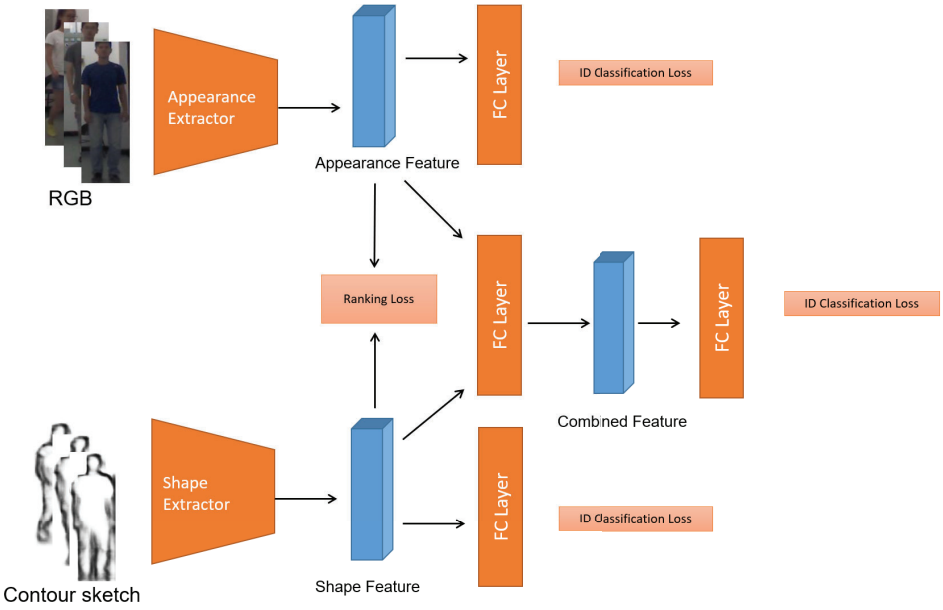

FIGURE 1. The architecture of our proposed dual-path model. The appearance and shape extractors measure RGB and contour sketch person images respectively. The concatenated appearance and shape features are fed into an FC layer to obtain the output.

shape extractor E_S. These extractors learn the cloth-irrelevant and cloth-relevant feature from the RGB and contour sketch images respectively. We obtain the sketch image I^{sketch} from the RBG image I^{rgb} with the identity label y using the holistically nested edge detection model [20]. We aim to learn the cloth-irrelevant identity features. E_A extracts appearance feature f^A including human face and skin from RGB image, while E_S learns shape feature f^S including mainly the body shape which is highlighted by contour sketch of a person. To take both advantage of the two features, a fully-connected (FC) layer is added at the end of the model as the feature combinator B, to further learn representation f^B from concatenated appearance and shape feature.

$$f^B = W[(f^A)^T, (f^S)^T]^T + b \tag{1}$$

In evaluation, only f^B is used to represent the person feature.

We consider ReID as a multi-class classification task. The identity classification loss is applied to the feature learning. To be more specific, we add an FC layer for f^B and the softmax loss is used for ID classification.

$$L_{id}^B = -\frac{1}{N}\sum_{i=1}^{N} log\frac{e^{W_{y_i}^T f_i^B}}{\sum_{k=1}^{C} e^{W_k^T f_i^B}} \tag{2}$$

where W_k denotes the weights of the classifier for class label k, y_i represents the label of I_i^{rgb} and I_i^{sketch}. C corresponds to the total number of identities and N is the size of a mini-batch. f_i^B denotes the identity features extracted from the input image pair I_i^{rgb} and I_i^{sketch}.

To encourage the two branches E_A and E_S to extract identity-relevant features, we also apply ID classification loss to f^A and f^S.

$$L_{id}^A = -\frac{1}{N}\sum_{i=1}^{N} log\frac{e^{W_{y_i}^T f_i^A}}{\sum_{k=1}^{C} e^{W_k^T f_i^A}} \tag{3}$$

$$L_{id}^S = -\frac{1}{N}\sum_{i=1}^{N} log\frac{e^{W_{y_i}^T f_i^S}}{\sum_{k=1}^{C} e^{W_k^T f_i^S}} \tag{4}$$

The total ID classification is calculated as:

$$L_{id} = \alpha_B L_{id}^B + \alpha_A L_{id}^A + \alpha_S L_{id}^S \tag{5}$$

where $\alpha_B = \alpha_A = \alpha_S = 1$.

The appearance f^A and shape features f^S obtained from the two extractors are combined by a fully connected layer into the same feature space. Considering that f^S extracted from the contour sketch may not contain clothing information, we enhance the similarity between f^A and f^S of the same person image aiming to remove the cloth information from f^A by minimizing the bi-directional ranking loss [21]. This loss is first adopted to ReID under clothes changing.

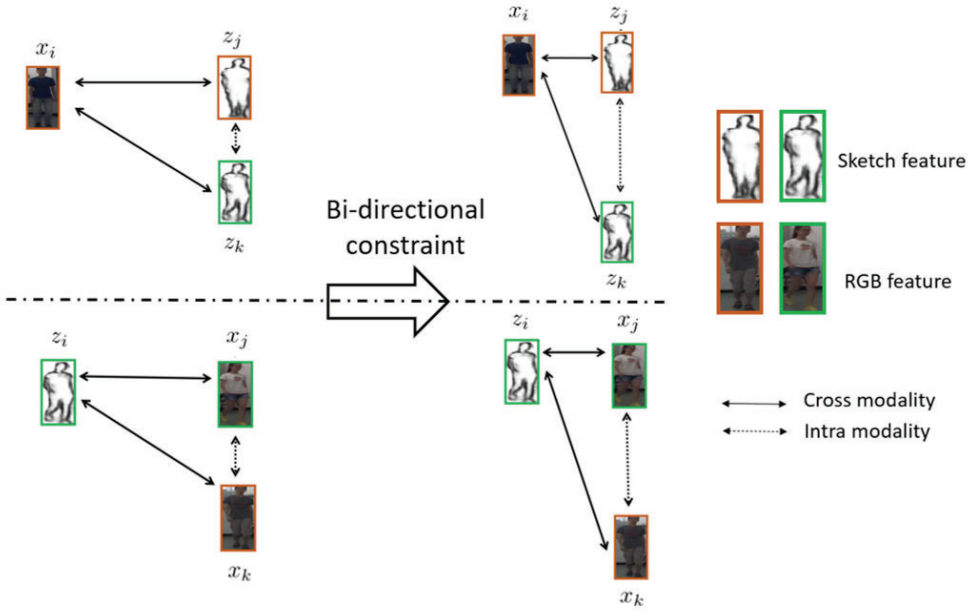

FIGURE 2. The illustration of the ranking loss. The ranking loss aims to push negative RGB-sketch pairs and pull positive RGB-sketch pairs. x and z denotes the RGB and sketch features, and the same border color denotes the same person identity.

The bi-directional ranking loss is calculated as follows. In a batch, there are N RGB and N corresponding sketch images. For an anchor RGB image x_i labelled as y_i, the distance between x_i and a negative sketch image z_k should be larger than the distance to its positive sketch images z_j. Similarly, for anchor sketch image z_i, its distance to its negative RGB image x_k is expected to be further than to its positive RGB image x_j.

$$D(x_i, z_j) < D(x_i, z_k) - \rho_1, \forall y_i = y_j, \forall y_i \neq y_k \tag{6}$$

where ρ_1 represents the pre-defined margin. L_2 norm is applied to calculate distance of input feature vectors x and z.

The cross-modality and intra-modality ranking constraints, defined as Eq. (7) and (8), to solve increase the inter-class distances and reduces the intra-class distances.

$$\begin{aligned} L_{cross} = & \sum_{\forall y_i = y_j} \max[\rho_1 + D(x_i, z_j) - \min_{\forall y_i \neq y_k} D(x_i, z_k), 0] \\ & + \sum_{\forall y_i = y_j} \max[\rho_1 + D(z_i, x_j) - \min_{\forall y_i \neq y_k} D(z_i, x_k), 0] \end{aligned} \tag{7}$$

$$\begin{aligned} L_{intra} = & \sum_{\forall y_j \neq y_k} \max[\rho_2 - D(z_j, z_k), 0] \\ & + \sum_{\forall y_j \neq y_k} \max[\rho_2 - D(x_j, x_k), 0] \end{aligned} \tag{8}$$

The whole model is trained by optimizing the total loss defined as the combination of all the mentioned loss with weights.

$$L_{total} = \lambda_{id} L_{id} + \lambda_{cross} L_{cross} + \lambda_{intra} L_{intra} \tag{9}$$

4. Experiments

4.1. Experimental settings

Datasets In our experiment, the dataset PRCC [11] is used to evaluate the performance of our network. PRCC consists of 221 identities with totally three camera views, namely camera A, B and C. Each person wears the same cloth under camera A and B but wears another cloth under camera C. 150 identities of person with totally 22,889 images are in training set. 71 identities of person are in test and query set. The test set contains one image per person shot by camera A for single-shot matching. To evaluate our model under both non-cloth and cloth changing settings, two testing cases, which are same-cloth matching and cross-cloth matching as [11], are used. Two query sets are constructed for the two testing cases. For query set for cross-cloth matching, all the images of the 71 identities shot by camera C are put into the query set to ensure that a same identity wears differently in test and query set. Correspondingly, for the other,

images shot by camera B will be put into query set so that a same person wears same clothes in test and query set.

Implementation details Our model is implemented in Pytorch. The main structure of the two branches network in our model inherits from MGN [16]. During training, using RandomSampler for sampling and the batch size is 10 including 2 identities and 5 images for each. Therefore, 750 images are randomly selected for training every epoch and the network is trained for 600 epochs. We resize the RGB and sketch images to 384*128 as input of the network. The learning rate is set to 0.1 initially and decaying 0.1 after 300 epochs. Adam [22] is used as the optimizer. Its weight decay is set as 5×10^{-4} and amsgrad is set as True. $\rho_1 = 0.5, \rho_2 = 0.1, \lambda_1 = 1$ and $\lambda_2 = 0.1$ are used in the ranking loss.

Evaluation metrics The standard cumulated matching characteristics (CMC) curve is adopted as our evaluation metrics. The rank-1, rank-5, rank-10 and rank-20 accuracy are also computed. For both testing cases, we repeated the evaluation 10 times each by randomly generating test sets and the average performance is calculated as the final result.

4.2. Ablation study

TABLE 1. Accuracy (%) of different variants of the proposed method on the PRCC dataset. Ri denotes the results of the rank i.

	R1	R5	R10	R20
RGB path only	53.18	75.16	84.59	94.21
Sketch path only	34.16	63.45	75.48	86.42
Dual-path	54.64	82.73	91.08	95.15
Dual-path+Ranking	56.79	82.98	88.85	96.47

In the subsection, different components of our method are evaluated separately. "RGB path only" and "sketch path only" only uses the RGB and sketch paths respectively. Moreover, the fc layer used for combining features is not applied to these methods. Both "dual-path" and "dual-path+ranking" represent the network using both RGB and sketch images. The difference between "dual-path" and "dual-path+ranking" is the ranking loss is missing in "dual-path". The identity classification loss of appearance feature f^A, sketch feature f^S and combine feature f^B are simply summed up in "dual-path". "Dual-path+ranking" is our complete method.

The experimental results shown in Table 1 illustrate that "dual-path+ranking" improves performance of Rank-1 by 1.5% and the ranking loss improve further 2% comparing to only using RGB extractor. With the ranking loss, the features extracted from RGB and sketch path become more similar to encourage he combined feature to contain less information about clothing and focus more on identity relevant feature such as body shape.

4.3. Comparison with the State-of-the-Arts

Our model is compared to the several state-of-the-arts ReID models on PRCC dataset, including representative ReID models without considering clothes changing, and the latest methods of cloth-changing ReID. The performance of the models are measured using rank-k matching accuracy.

The results shown in Table 2 suggest our model outperforms others significantly under cloth changing setting. We get the highest rank-1, rank-5, rank-10 and rank-20 matching accuracy among all the methods, which illustrates that our model can achieve excellent performance in ReID under cloth changing. The performance of all the models dramatically drops when clothes changing is considered, which indicates that clothes information dominates in ReID. The methods designed for clothes changing also downgrade since they still learn from the clothes information. Our model drops least in cross-cloth matching, which confirms the advantages of our method.

5. Conclusion

Traditional ReID models, which do not consider clothes changing, heavily rely on the clothes information. Their performance suffers from the same persons with different clothes. We devise a dual path model considering both RGB image and its contour sketch image aiming to robust features to clothes changing. The body shape information is extracted by contour sketch images while the appearance feature provides the person identity information. To encourage ReID model to neglect clothing feature from RGB input, we apply metric learning with Siamese loss to minimize the distance between features extracted from an RGB image and its contour sketch version. The experimental results on the PRCC dataset indicate that our proposed method perform well in both clothes changing and non-clothes changing settings.

Acknowledgments

This paper is supported by the Natural Science Foundation of Guangdong Province, China (No. 2018A030313203) and the Fundamental Research Funds for the Central Universities (No. 2018ZD32).

TABLE 2. Comparison to state-of-the-arts methods

Cloth Change	Method	Cross-cloth matching				Same-cloth matching			
		Rank-1	Rank-5	Rank-10	Rank-20	Rank-1	Rank-5	Rank-10	Rank-20
	WBDR+WFDR [23]	21.50	35.70	46.77	63.82	93.40	96.70	98.70	99.80
	Aligned reid [24]	34.60	53.40	64.90	79.10	94.80	99.40	99.70	99.91
	Resnet & tricks [25]	46.90	66.20	73.40	84.11	93.70	97.90	99.10	99.82
	MGN [16]	4793	70.82	84.93	91.96	99.77	99.90	99.95	99.98
	PCB [14]	30.88	49.65	60.09	77.73	95.79	99.23	99.30	99.95
	ISGAN [26]	51.15	75.16	82.44	91.84	**99.90**	99.92	99.95	99.98
✓	Celeb-reid [10]	44.30	68.00	80.00	89.13	98.50	99.30	99.90	99.98
✓	PRCC [11]	34.38	55.66	77.30	88.05	64.20	78.93	92.60	97.03
✓	Ours	**56.79**	**82.98**	**88.85**	**96.47**	99.79	**99.93**	**99.96**	**99.98**

References

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

[2] S. Liao, Y. Hu, X. Zhu, and S. Z. Li, "Person re-identification by local maximal occurrence representation and metric learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 2197–2206.

[3] M. Kostinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof, "Large scale metric learning from equivalence constraints," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 2288–2295.

[4] A. Hermans, L. Beyer, and B. Leibe, "In defense of the triplet loss for person re-identification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

[5] D. Cheng, Y. Gong, S. Zhou, J. Wang, and N. Zheng, "Person re-identification by multi-channel parts-based cnn with improved triplet loss function," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1335–1344.

[6] L. Zhao, X. Li, Y. Zhuang, and J. Wang, "Deeply-learned part-aligned representations for person re-identification," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 3219–3228.

[7] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian, "Scalable person re-identification: A benchmark," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1116–1124.

[8] W. Li, R. Zhao, T. Xiao, and X. Wang, "Deepreid: Deep filter pairing neural network for person re-identification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 152–159.

[9] Z. Zheng, L. Zheng, and Y. Yang, "Unlabeled samples generated by gan improve the person re-identification baseline in vitro," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 3774–3782.

[10] Y. Huang, Q. Wu, J. Xu, and Y. Zhong, "Celebrities-reid: A benchmark for clothes variation in long-term person re-identification," in *proceedings of International Joint Conference on Neural Networks*, 2019, pp. 1–8.

[11] Q. Yang, A. Wu, and W. Zheng, "Person re-identification by contour sketch under moderate clothing change," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1, 2020.

[12] X. Qian, W. Wang, L. Zhang, F. Zhu, Y. Fu, T. Xiang, Y. G. Jiang, and X. Xue, "Long-term cloth-changing person re-identification," in *arXiv preprint arXiv:2005.07862*, 2020.

[13] F. Wan, Y. Wu, X. Qian, and Y. Fu, "When person re-identification meets changing clothes," *arXiv: Computer Vision and Pattern Recognition*, 2020.

[14] Y. Sun, L. Zheng, Y. Yang, Q. Tian, and S. Wang, "Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline)," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 501–518.

[15] X. Fan, H. Luo, X. Zhang, L. He, C. Zhang, and W. Jiang, "Scpnet: Spatial-channel parallelism network for joint holistic and partial person re-identification," in *proceedings of Asian Conference on Computer Vision*, 2018, pp. 19–34.

[16] G. Wang, Y. Yuan, X. Chen, J. Li, and X. Zhou, "Learning discriminative features with multiple granularities for person re-identification," in *proceedings of ACM International Conference on Multimedia*, 2018.

[17] I. B. Barbosa, M. Cristani, D. A. Bue, L. Bazzani, and V. Murino, "Re-identification with rgb-d sensors," in *First International Workshop on Re-Identification in conjunction with ECCV 2012*, 2012.

[18] Y. Huang, J. Xu, Q. Wu, Y. Zhong, P. Zhang, and Z. Zhang, "Beyond scalar neuron: Adopting vector-neuron capsules for long-term person re-identification," *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 1–1, 2019.

[19] Z. Yu, C. Feng, M.-Y. Liu, and S. Ramalingam, "Casenet: Deep category-aware semantic edge detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 5964–5973.

[20] S. Xie and Z. Tu, "Holistically-nested edge detection," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1395–1403.

[21] M. Ye, Z. Wang, X. Lan, and P. C. Yuen, "Visible thermal person re-identification via dual-constrained top-ranking." in *IJCAI*, vol. 1, 2018, p. 2.

[22] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[23] H. Yu and W. Zheng, "Weakly supervised discriminative feature learning with state information for person identification," *arXiv: Computer Vision and Pattern Recognition*, 2020.

[24] X. Zhang, H. Luo, X. Fan, W. Xiang, Y. Sun, Q. Xiao, W. Jiang, C. Zhang, and J. Sun, "Alignedreid: Surpassing human-level performance in person re-identification," *arXiv preprint arXiv:1711.08184*, 2017.

[25] H. Luo, Y. Gu, X. Liao, S. Lai, and W. Jiang, "Bag of tricks and a strong baseline for deep person re-identification," in *proceedings of The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2019.

[26] C. Eom and B. Ham, "Learning disentangled representation for robust person re-identification," in *Advances in Neural Information Processing Systems*, 2019, pp. 5297–5308.