IDENTITY-SENSITIVE KNOWLEDGE PROPAGATION FOR CLOTH-CHANGING PERSON RE-IDENTIFICATION

Jianbing Wu[†] *Hong Liu*[†*] *Wei Shi*[†] *Hao Tang*[‡] *Jingwen Guo*[†]

[†] Key Laboratory of Machine Perception, Shenzhen Graduate School, Peking University, China
[‡] Computer Vision Lab, ETH Zurich, Switzerland

{kimbing.ng,jingwenguo}@stu.pku.edu.cn, {hongliu,pkusw}@pku.edu.cn, hao.tang@vision.ee.ethz.ch

ABSTRACT

Cloth-changing person re-identification (CC-ReID), which aims to match person identities under clothing changes, is a new rising research topic in recent years. However, typical biometrics-based CC-ReID methods often require cumbersome pose or body part estimators to learn cloth-irrelevant features from human biometric traits, which comes with high computational costs. Besides, the performance is significantly limited due to the resolution degradation of surveillance images. To address the above limitations, we propose an effective ***Identity-Sensitive Knowledge Propagation framework*** (DeSKPro) for CC-ReID. Specifically, a Cloth-irrelevant Spatial Attention module is introduced to eliminate the distraction of clothing appearance by acquiring knowledge from the human parsing module. To mitigate the resolution degradation issue and mine identity-sensitive cues from human faces, we propose to restore the missing facial details using prior facial knowledge, which is then propagated to a smaller network. After training, the extra computations for human parsing or face restoration are no longer required. Extensive experiments show that our framework outperforms state-of-the-art methods by a large margin. Our code is available at https://github.com/KimbingNg/DeskPro.

Index Terms— Cloth-Changing Person Re-Identification, Identity-Sensitive Features, Knowledge Propagation

1. INTRODUCTION

Person re-identification (Re-ID), which aims to match pedestrians across non-overlapping cameras, is a challenging task with significant research impact. Recent works [1–4] have achieved remarkable progress by learning powerful appearance representations based on the assumption that the images of the same identity in both the query set and the gallery set have the same clothing. However, the clothing appearance tends to be unreliable in realistic scenes since people may wear different clothes at different times.

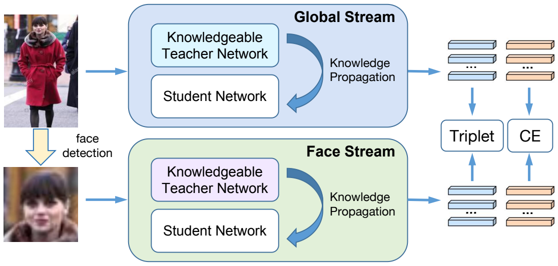

Fig. 1. Overall architecture of the proposed framework. "CE" and "Triplet" denote two widely used Re-ID losses, namely the cross-entropy loss and the triplet loss [17].

To address the above practical problem, Cloth-Changing Person Re-identification (CC-ReID) has drawn increasing attention in recent years [5–14]. Existing CC-ReID methods often focus on mining identity-relevant cues from pre-defined biometric traits. For instance, Yang *et al.* [8] introduced a spatial polar transformation on contour sketch to learn shape representations. Shi *et al.* [13] proposed to learn head-guided features with the help of the pose estimator [15]. Qian *et al.* [16] introduced a shape embedding module to extract biological features from body keypoints. A common issue behind these methods is that extra pose or contour estimation is required during inference, leading to higher computational costs. Some other researchers are dedicated to mining the identity-sensitive features from human faces [5, 6] since facial cues are more robust under clothing changes. However, due to the low resolution of surveillance images, these methods also fail to extract representative facial features.

To address these challenges, we propose an ***Identity-Sensitive Knowledge Propagation framework***, termed DeSKPro. As illustrated in Fig. 1, the proposed framework consists of a global cloth-irrelevant feature stream (global stream) and a facial feature enhancement stream (face stream). Both streams can acquire informative knowledge from their respective teacher networks during training. In the global stream, a Cloth-irrelevant Spatial Attention (CSA) module, which is trained under the guidance of the knowledgeable human parsing network, is designed to facilitate the learning of identity-sensitive features. In the face stream, we begin by

[*] Corresponding author. This work is supported by National Natural Science Foundation of China (No.62073004), and Shenzhen Fundamental Research Program (No.JCYJ20200109140410340).

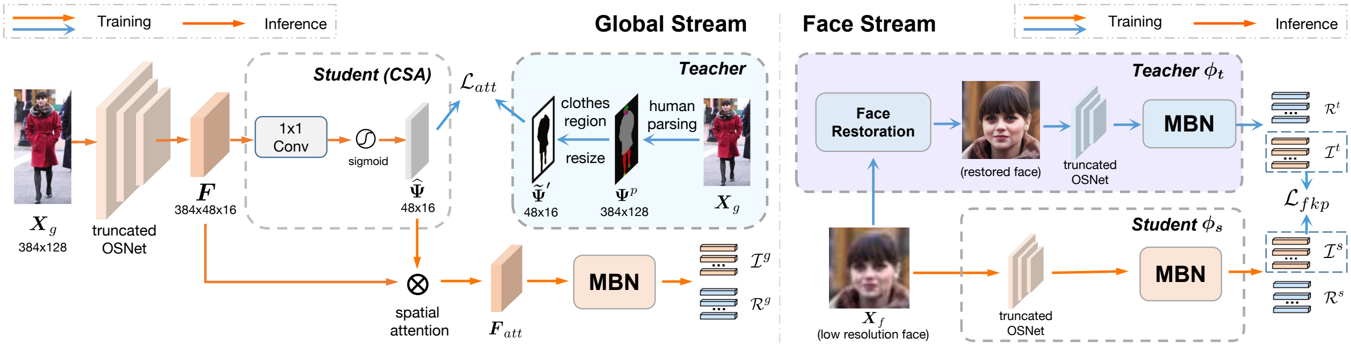

Fig. 2. The architecture of the global stream and the face stream. The global stream aims to learn cloth-irrelevant knowledge under the guidance of the human parsing module, and the face stream is designed to extract representative facial features from low-resolution images by acquiring knowledge from the teacher network. The details of notations can be referred to in section 2.

training a teacher network, which can restore missing facial details from degraded images and learn representative features. After that, a facial knowledge propagation loss is introduced to transfer the facial knowledge to a simpler student network. Note that with our well-designed knowledge propagation strategies, the cumbersome human parsing and the face restoration module can be removed during inference to reduce redundant computations.

Our contributions are summarized as follows: **(1)** We introduce a cloth-irrelevant spatial attention module for the CC-ReID task to effectively discourage the misleading features of clothes regions while preserving the identity-sensitive ones. **(2)** To mitigate the resolution degradation issue of surveillance images, we further propose the idea of restoring the missing facial details and propagating the facial knowledge to a smaller network. **(3)** Extensive experiments show that the proposed framework outperforms existing methods by a large margin without relying on the human parsing or face restoration module during inference.

2. METHODOLOGY

The key to addressing the CC-ReID problem is to reduce the attention to the clothes regions and learn identity-sensitive features from biometric traits, such as human faces. We thus design a two-stream architecture, which consists of a global stream and a face stream, to learn both global cloth-irrelevant features and enhanced facial features. Fig. 2 illustrates the architecture of the two streams, and more details are discussed in the following subsections.

2.1. Mask-Guided Cloth-irrelevant Feature Stream

The clothing appearance is not reliable in cloth-changing settings. A naive solution is to utilize spatial attention mechanisms to explicitly discourage clothing features. However, it is difficult to learn effective attention weights without auxiliary supervision. Instead, we design a new Cloth-irrelevant Spatial Attention (CSA) module to learn effective attention maps under the guidance of the human parsing network.

Cloth-irrelevant Spatial Attention. As illustrated in the global stream of Fig. 2, given a person image $\boldsymbol{X}_g$, we first pass it through the truncated OSNet [18] backbone up until the first layer of its third convolutional block to derive the middle feature maps $\boldsymbol{F}$ with size $h_f \times w_f \times c_f$. Taking $\boldsymbol{F}$ as the input, our CSA module then produces a cloth-irrelevant attention map, which can be formulated as:

$$\widehat{\boldsymbol{\Psi}} = \sigma\big(\boldsymbol{W}_{pw} * \boldsymbol{F} + \boldsymbol{b}\big), \tag{1}$$

where $*$ denotes convolution operation, $\boldsymbol{W}_{pw}$ denotes the point-wise convolution filters, and $\sigma(x) = 1/(1 + exp(-x))$ is the sigmoid function. The attention map $\widehat{\boldsymbol{\Psi}}$ is then applied to the feature $\boldsymbol{F}$ using the following formula:

$$\boldsymbol{F}_{att} = \boldsymbol{F} \otimes \widehat{\boldsymbol{\Psi}}, \tag{2}$$

where $\otimes$ denotes the Hadamard matrix product, and $\boldsymbol{F}_{att}$ is the refined feature maps. We then feed $\boldsymbol{F}_{att}$ into the Multi-Branch Network (MBN) introduced in [4], which combines global, part-based, and channel features in a multi-branch architecture. The resulting embeddings are grouped into two sets following [4], denoted by $\mathcal{I}^g$ and $\mathcal{R}^g$, respectively.

Mask Guided Attention Loss. The attention map $\widehat{\boldsymbol{\Psi}}$ is expected to have lower scores in the clothes regions. However, without extra auxiliary supervision, it is hard to guarantee that the CSA module produces cloth-irrelevant attention scores. To this end, we introduce a mask-guided attention loss to enable the CSA module to learn effective attention maps. As depicted in Fig. 2, we first utilize the pre-trained human parsing module [19] to estimate a fine-grained mask $\boldsymbol{\Psi}^p$ of human parts taking the image $\boldsymbol{X}_g$ as the input. Each grid in the mask $\boldsymbol{\Psi}^p$ represents the category of the corresponding pixel in $\boldsymbol{X}_g$ (e.g., arm, leg, dress, skirt). After that, a cloth-irrelevant mask can be obtained using the following formula:

$$\widetilde{\boldsymbol{\Psi}}_{(i,j)} = \begin{cases} \varepsilon, & \text{if } \boldsymbol{\Psi}^p_{(i,j)} \in \mathcal{C}, \\ 1, & \text{otherwise.} \end{cases} \tag{3}$$

Here, $\mathcal{C}$ denotes the set of cloth-related categories, and ε is a hyper-parameter to avoid producing near-zero attention scores. The mask guided attention loss is defined as:

$$\mathcal{L}_{att} = \frac{1}{h_f \cdot w_f} \sum_{i=1}^{h_f} \sum_{j=1}^{w_f} \left(\widehat{\mathbf{\Psi}}_{(i,j)} - \widetilde{\mathbf{\Psi}}'_{(i,j)} \right)^2 , \tag{4}$$

where $\widetilde{\mathbf{\Psi}}'$ is obtained by resizing $\widetilde{\mathbf{\Psi}}$ to the same size as $\widehat{\mathbf{\Psi}}$.

2.2. Facial Feature Enhancement Stream

The neural systems of human beings can perfectly identify persons by exploiting facial cues, even with resolution-degraded images. This is thanks to their capability of recovering missing facial details. Inspired by this, we design a teacher-student network in the face stream to propagate facial knowledge from the knowledgeable teacher network ϕ_t, which is trained beforehand, to the separate student network ϕ_s. Both the teacher and the student networks take the face image $\boldsymbol{X}_f$ as input, which is detected from the image $\boldsymbol{X}_g$ by the face detector [20]. As illustrated in Fig. 2, the teacher first restores the facial details using the pre-trained face restoration module GPEN [21]. The enhanced face is then fed to the truncated OSNet and the MBN to derive the feature vector set $\mathcal{R}^t$ and the logit output vector set $\mathcal{I}^t$. By optimizing the cross-entropy loss using $\mathcal{I}^t$ and the triplet loss using $\mathcal{R}^t$, a knowledgeable teacher model can be obtained.

After training the teacher network, we fix its parameters and propagate the informative facial knowledge to the simpler student network ϕ_s, where the cumbersome face restoration module is removed. Similar to the teacher ϕ_t, the outputs of the student ϕ_s are also grouped into two sets, denoted by $\mathcal{I}^s$ and $\mathcal{R}^s$, respectively. Inspired by the idea of [22], we design a facial knowledge propagation loss to transfer the knowledge to the student, which is defined as:

$$\mathcal{L}_{fkp} = \tau^2 \sum\nolimits_i KL \left(\mathcal{S}(\mathcal{I}^t_{(i)}, \tau) \parallel \mathcal{S}(\mathcal{I}^s_{(i)}, \tau) \right) , \tag{5}$$

where KL denotes Kullback–Leibler divergence and τ is the temperature parameter of the softmax function $\mathcal{S}$. Here, the softmax function $\mathcal{S}$, defined as $\mathcal{S}(\boldsymbol{z}, \tau)_i = e^{(\boldsymbol{z}_i/\tau)} / \sum_j e^{(\boldsymbol{z}_j/\tau)}$, is designed to produce knowledgeable soft labels for the student ϕ_t. The term τ^2 in Formula (5) is adopted to guarantee that the magnitude of the computed loss remains unchanged for different values of τ. By optimizing $\mathcal{L}_{fkp}$, the student network ϕ_s is endowed with the capability of extracting discriminative facial features from low-resolution images without relying on the face restoration module.

2.3. Training and Inference

Training. Apart from the aforementioned losses, we also optimize the batch hard triplet loss [17] using all features in $\mathcal{R}^g \cup \mathcal{R}^s$, and the cross-entropy loss using all logit outputs in $\mathcal{I}^g \cup \mathcal{I}^s$ for basic discrimination learning, formulated as:

$$\begin{aligned} &\mathcal{L}^g_{ce} = \sum\nolimits_{\hat{\boldsymbol{y}} \in \mathcal{I}^g} CE(\hat{\boldsymbol{y}}, \boldsymbol{y}), \mathcal{L}^s_{ce} = \sum\nolimits_{\hat{\boldsymbol{y}} \in \mathcal{I}^s} CE(\hat{\boldsymbol{y}}, \boldsymbol{y}), \\ &\mathcal{L}_{trip} = \sum\nolimits_{\boldsymbol{f} \in \mathcal{R}^g} Trip(\boldsymbol{f}, \boldsymbol{y}) + \sum\nolimits_{\boldsymbol{f} \in \mathcal{R}^s} Trip(\boldsymbol{f}, \boldsymbol{y}), \end{aligned} \tag{6}$$

where CE and $Trip$ denote the cross-entropy loss and the batch hard triplet loss, respectively. The term $\boldsymbol{y}$ denotes the identity labels of person images. The overall framework is trained by optimizing the total loss $\mathcal{L}$, which is defined as:

$$\mathcal{L} = \lambda \mathcal{L}_{att} + \mathcal{L}_{trip} + (\alpha \mathcal{L}_{fkp} + (1-\alpha)\mathcal{L}^s_{ce}) + \mathcal{L}^g_{ce}, \tag{7}$$

where the λ and α are adopted to balance different losses.

Inference. The inference stage, during which the human parsing module in the global stream and the teacher network ϕ_t in the face stream are removed, can be seen as a cosine-similarity-based retrieval process. Embeddings obtained before the last fully connected layer in the MBNs are concatenated for similarity computation as in [4]. As for those query images with no face detected, we only enable the global stream to compute feature vectors.

3. EXPERIMENTS AND DISCUSSIONS

3.1. Datasets and Experimental Setups

Our experiments are conducted on three widely used CC-ReID datasets, including Celeb-reID [7], Celeb-reID-light [7], and PRCC [8]. The Celeb-reID dataset is made up of person images where over 70% of samples show different clothes. The Celeb-reID-light dataset is the light but challenging version of Celeb-reID since all images of each person are in different clothes. The PRCC dataset provides both cross-clothes and same-clothes settings to support in-depth evaluations. The values of ε, λ, τ, and α are determined by cross validation. Specifically, for Celeb-reID and Celeb-reID-light, we set $\varepsilon = 0.1$, $\lambda = 7$, $\tau = 5$, and $\alpha = 0.7$. For PRCC dataset, we set $\varepsilon = 0.1$, $\lambda = 7$, $\tau = 1$, and $\alpha = 0.8$. Other experimental settings follows our previous work [13]. The cumulative matching characteristics (CMC) and mean Average Precision (mAP) are reported in the following subsections.

3.2. Comparison with State-of-the-Art Methods

We compare our method with several common Re-ID models [1–4] and state-of-the-art CC-ReID approaches [7–14]. In the tables hereafter, "R-k" denotes rank-k accuracy, "-" denotes not reported, and the best and second-best results are in bold and underline styles, respectively.

Table 1 shows the experimental results on Celeb-reID-light and Celeb-reID datasets. Compared with the state-of-the-art method IRANet [13], which relies on the off-the-shelf pose estimator [15], our method achieves higher performance in both datasets. For PRCC dataset, we conduct experiments in both cross-clothes and same-clothes settings. As shown in Table 2, the proposed DeSKPro outperforms the best existing method by a large margin in the cross-clothes setting. Notice that our method surpasses FSAM [12] as well, which also attempts to transfer knowledge from pre-trained networks to save extra computations. This is because they only complement shape knowledge in the appearance stream while omitting the identity-sensitive cues from human faces. Moreover,

Table 1. Comparison on Celeb-reID-light and Celeb-reID datasets (%).

Method	Celeb-reID-light			Celeb-reID		
	R-1	R-5	mAP	R-1	R-5	mAP
HACNN [1]	16.2	-	11.5	47.6	-	9.5
MGN [3]	21.5	-	13.9	49.0	-	10.8
AFD-Net [10]	22.2	51.0	11.3	52.1	66.1	10.6
LightMBN [4]	32.9	67.0	18.8	57.3	71.6	14.3
ReIDCaps+ [7]	33.5	63.3	19.0	63.0	76.3	15.8
RCSANet [11]	46.6	-	24.4	65.3	-	17.5
CASE-Net [9]	35.1	66.7	20.4	66.4	78.1	18.2
IRANet [13]	46.2	72.7	25.4	64.1	78.7	19.0
DeSKPro (Ours)	**52.0**	**81.6**	**29.8**	**68.6**	**82.3**	**22.7**

Table 2. Comparison on PRCC dataset (%).

Method	PRCC			
	Cross clothes		Same clothes	
	R-1	mAP	R-1	mAP
HACNN [1]	21.8	-	82.5	-
PCB [2]	22.9	-	86.9	-
Yang *et al.* [8]	34.4	-	64.2	-
CASE-Net [9]	39.5	-	71.2	-
AFD-Net [10]	42.8	-	95.7	-
RCSANet [11]	50.2	48.6	**100.0**	97.2
LightMBN [4]	50.5	51.2	**100.0**	**98.9**
FSAM [12]	54.5	-	98.8	-
IRANet [13]	54.9	53.0	99.7	97.8
Shu *et al.* [14]	65.8	61.2	99.5	96.7
DeSKPro (Ours)	**74.0**	**66.3**	99.6	96.6

the clothing appearance is not explicitly suppressed in their method, which damages the robustness of features. Although DeSKPro does not outperform all existing methods in the same-clothes setting, it still attains competitive results.

3.3. Effectiveness of Proposed Components

To verify the effectiveness of each component in DeSKPro, we conduct ablation experiments with different component settings on the more challenging Celeb-reID-light dataset.

Cloth-irrelevant Spatial Attention. We first evaluate the importance of the attention module CSA and the mask-guided attention loss $\mathcal{L}_{att}$. As shown in Table 3, the rank-1 accuracy and mAP can be improved with the CSA module. After adding $\mathcal{L}_{att}$ to the total loss function, the performance can be further boosted. The feature visualization is presented in Fig. 3, which also shows that the CSA model can effectively attend to the cloth-irrelevant regions after training with $\mathcal{L}_{att}$.

Face Enhancement and Knowledge Propagation. As shown in Table 3, the teacher network (model 6) significantly outperforms the student network (model 4), which verifies that restoring face details can help improve performance. Comparing model 5 with model 4, it can be seen that the performance is boosted significantly after the knowledge propagation. Even though model 5 is still slightly inferior to model 6, it has lower computational costs since the face restoration module is removed. This also indicates that the knowledge propagation strategy in the face stream is the trade-off between performance and computational efficiency.

Table 3. Ablation study on Celeb-reID-light dataset. "$\mathcal{G}$" denotes the global stream without the CSA module, and ϕ^{+}/ϕ^{-} denotes the student network trained with/without $\mathcal{L}_{fkp}$.

Model	Body Stream			Face Stream			Celeb-reID-light	
	$\mathcal{G}$	CSA	$\mathcal{L}_{att}$	ϕ_s^{-}	ϕ_s^{+}	ϕ_t	R-1	mAP
1 (baseline)	✓						32.9	18.8
2	✓	✓					33.5	20.1
3	✓	✓	✓				37.2	20.3
4				✓			41.6	20.7
5					✓		47.0	25.6
6						✓	47.4	25.8
7	✓	✓	✓	✓			50.1	27.4
DeSKPro	✓	✓	✓		✓		52.0	29.8
DeSKPro*	✓	✓	✓			✓	**53.7**	**30.9**

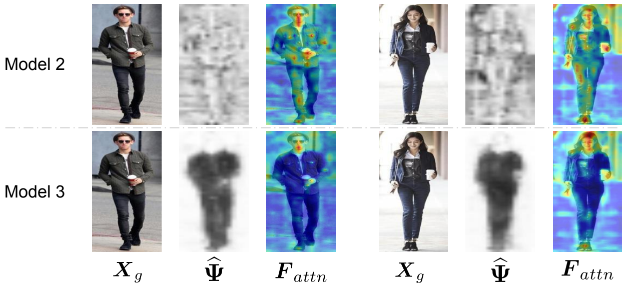

Fig. 3. Visualization of the attention maps and the refined feature maps for model 2 and model 3.

Two Stream Architecture of DeSKPro. As shown in Table 3, the performance can be further improved by assembling both global and face streams. This indicates that the global and the facial features complement each other. We also provide a variant of DeSKPro, which is termed DeSKPro* in Table 3. It is obtained by replacing the student network ϕ_s with the more complex teacher network ϕ_t in the face stream. Even though it achieves the best performance in our experiments, it brings extra computational costs for face restoration.

4. CONCLUSION

In this paper, we present an ***Ide**ntity-**S**ensitive **K**nowledge **Pro**pagation framework* (DeSKPro), which achieves state-of-the-art results on three challenging CC-ReID datasets, including Celeb-reID-light, Celeb-reID, and PRCC. We argue that suppressing the clothes features explicitly and recovering facial details from resolution-degraded images can help boost the performance of the CC-ReID task. The experiments have also shown that propagating body and facial knowledge can help avoid the extra computation costs for mask estimation or face restoration while preserving performance. We hope this work can spark further research on the CC-ReID problem.

5. REFERENCES

[1] Wei Li, Xiatian Zhu, and Shaogang Gong, "Harmonious Attention Network for Person Re-identification," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 2285–2294.

[2] Yifan Sun, Liang Zheng, Yi Yang, et al., "Beyond Part Models: Person Retrieval with Refined Part Pooling (and A Strong Convolutional Baseline)," in *European Conference on Computer Vision*, 2018, pp. 501–518.

[3] Guanshuo Wang, Yufeng Yuan, Xiong Chen, et al., "Learning Discriminative Features with Multiple Granularities for Person Re-Identification," in *ACM International Conference on Multimedia*, 2018, pp. 274–282.

[4] Fabian Herzog, Xunbo Ji, Torben Teepe, et al., "Lightweight Multi-Branch Network For Person Re-Identification," in *IEEE International Conference on Image Processing*, 2021, pp. 1129–1133.

[5] Shijie Yu, Shihua Li, Dapeng Chen, et al., "COCAS: A Large-Scale Clothes Changing Person Dataset for Re-Identification," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 3397–3406.

[6] Fangbin Wan, Yang Wu, Xuelin Qian, et al., "When Person Re-identification Meets Changing Clothes," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 3620–3628.

[7] Yan Huang, Jingsong Xu, Qiang Wu, et al., "Beyond Scalar Neuron: Adopting Vector-Neuron Capsules for Long-Term Person Re-Identification," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 10, pp. 3459–3471, 2020.

[8] Qize Yang, Ancong Wu, and Wei-Shi Zheng, "Person Re-Identification by Contour Sketch Under Moderate Clothing Change," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 6, pp. 2029–2046, June 2021.

[9] Yu-Jhe Li, Xinshuo Weng, and Kris M. Kitani, "Learning Shape Representations for Person Re-Identification under Clothing Change," in *IEEE/CVF Winter Conference on Applications of Computer Vision*, 2021, pp. 2432–2441.

[10] Wanlu Xu, Hong Liu, Wei Shi, et al., "Adversarial Feature Disentanglement for Long-Term Person Re-identification," in *International Joint Conference on Artificial Intelligence*, 2021, pp. 1201–1207.

[11] Yan Huang, Qiang Wu, JingSong Xu, et al., "Clothing Status Awareness for Long-Term Person Re-Identification," in *IEEE/CVF International Conference on Computer Vision*, 2021, pp. 11895–11904.

[12] Peixian Hong, Tao Wu, Ancong Wu, et al., "Fine-Grained Shape-Appearance Mutual Learning for Cloth-Changing Person Re-Identification," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 10513–10522.

[13] Wei Shi, Hong Liu, and Mengyuan Liu, "IRANet: Identity-relevance Aware Representation for Cloth-Changing Person Re-Identification," *Image and Vision Computing*, vol. 117, pp. 104335, 2022.

[14] Xiujun Shu, Ge Li, Xiao Wang, et al., "Semantic-Guided Pixel Sampling for Cloth-Changing Person Re-Identification," *IEEE Signal Processing Letters*, vol. 28, pp. 1365–1369, 2021.

[15] Riza Alp Guler, Natalia Neverova, and Iasonas Kokkinos, "DensePose: Dense Human Pose Estimation in the Wild," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7297–7306.

[16] Xuelin Qian, Wenxuan Wang, Li Zhang, et al., "Long-Term Cloth-Changing Person Re-identification," in *Asian Conference on Computer Vision*, 2021, pp. 71–88.

[17] Alexander Hermans, Lucas Beyer, and Bastian Leibe, "In Defense of the Triplet Loss for Person Re-Identification," *arXiv preprint arXiv:1703.07737*, 2017.

[18] Kaiyang Zhou, Yongxin Yang, Andrea Cavallaro, et al., "Omni-Scale Feature Learning for Person Re-Identification," in *IEEE/CVF International Conference on Computer Vision*, 2019, pp. 3701–3711.

[19] Peike Li, Yunqiu Xu, Yunchao Wei, et al., "Self-Correction for Human Parsing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.

[20] Jiankang Deng, Jia Guo, Evangelos Ververas, et al., "RetinaFace: Single-Shot Multi-Level Face Localisation in the Wild," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 5202–5211.

[21] Tao Yang, Peiran Ren, Xuansong Xie, et al., "GAN Prior Embedded Network for Blind Face Restoration in the Wild," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 672–681.

[22] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean, "Distilling the Knowledge in a Neural Network," *arXiv preprint arXiv:1503.02531*, 2015.