Person Re-identification with Soft Biometrics through Deep Learning

Shan Lin and Chang-Tsun Li

Abstract Re-identification of persons is usually based on primary biometric features such as their faces, fingerprints, iris or gait. However, in most existing video surveillance systems, it is difficult to obtain these features due to the low resolution of surveillance footages and unconstrained real-world environments. As a result, most of the existing person re-identification techniques only focus on overall visual appearance. Recently, the use of soft biometrics has been proposed to improve the performance of person re-identification. Soft biometrics such as height, gender, age are physical or behavioural features, which can be described by humans. These features can be obtained from low-resolution videos at a distance ideal for person re-identification application. In addition, soft biometrics are traits for describing an individual with human-understandable labels. It allows human verbal descriptions to be used in the person re-identification or person retrieval systems. In some deep learning based person re-identification methods, soft biometrics attributes are integrated into the network to boot the robustness of the feature representation. Biometrics can also be utilised as a domain adaptation bridge for addressing the cross-dataset person re-identification problem. This chapter will review the state-of-the-art deep learning methods involving soft biometrics from three perspectives: supervised, semi-supervised and unsupervised approaches. In the end, we discuss the existing issues that are not addressed by current works.

Shan Lin
University of Warwick, Coventry, UK, e-mail: shan.lin@warwick.ac.uk

Chang-Tsun Li
University of Warwick, Coventry, UK, e-mail: c-t.li@warwick.ac.uk
Deakin University, Melbourne, Australia, e-mail: c-t.li@deakin.edu.au

1 Introduction

Person re-identification, also known as person Re-ID is the task of recognising and continuously identifying the same person across multiple non-overlapping cameras in a surveillance system. Typical methods for automatic person identification system are usually based on people's hard biometric traits such as fingerprints, irises or faces. However, in the video surveillance system, target individuals are generally captured at a distance in an uncontrolled environment. Such settings introduce a lot of difficulty in obtaining these hard biometric traits due to the low resolution of the camera sensors, occlusion of the subjects, etc.[14]

Most of the existing research works on person Re-ID are focusing on extracting the local view-invariant features of a person and learning a discriminate distance metric for similarity analysis. Soft biometrics such as gender, age, hair-style or clothing is mid-level semantic descriptions of a person which are invariant to illumination, viewpoint and pose. Hence, in recent years, soft biometrics have been used in conjunction with the identity information to aid many Re-ID methods as auxiliary information. Moreover, these human-understandable attribute labels bridge the gap between how machines and people recognising and identifying human-beings. By integrating the soft biometrics, the existing person Re-ID system can be extended to a broader range of application such as text-to-person retrieval, image to description conversion.

Although, many soft biometric attribute assisted person Re-ID methods have been proposed previously. Most of them are based on traditional machine learning approaches with hand-crafted features. In [8, 7, 9], the attributes are classified by SVM from low-level descriptors and integrated in the metric leaning step of the person Re-ID. Su *et al.* [17] proposed Low-Rank Attribute Embedding which embeds the binary attributes to a continuous attribute space based on the local feature representations. Khamis *et al.* [5] developed a method based on the attribute consistency for the same person and proposed to the triplet loss for the attribute features. All these works are based on the small VIPeR and PRID datasets with only 21 binary attributes. In 2014, Deng *et al.* [1] released a large-scale pedestrian attribute dataset PETA which include images from multiple person Re-ID datasets. However, the PETA dataset did not contain an adequate number of training images for deep learning based person Re-ID methods. Until Lin *et al.* [11] released the annotations for two largest Re-ID datasets: Mareket-1501 [26] and DukeMTMC-reID [27], soft biometrics start integrating into the deep person Re-ID methods. However, the size of the annotated attributes for Market-1501 and DukeMTMC-reID are still relatively limited compared with the PETA datasets.

In this chapter, we first list out the person Re-ID datasets with soft biometrics labels and introduce the performance evaluation metrics for person Re-ID task and Attribute recognition task. Then, we present the recent soft biometrics based or assisted deep person Re-ID methods from three perspectives: supervised, semi-supervised and unsupervised learning. Finally, we discuss the unaddressed problems of the soft biometric in person Re-ID and outline the potential future work.

2 Datasets and Evaluation

2.1 Datasets with Soft Biometrics Annotations

Currently, there are approximately 30 publicly available person Re-ID datasets. However, only a small portion of them come with soft biometric attributes annotations. Table 1 lists the broadly used person Re-ID datasets come with soft biometric attributes. In [7], Layne *et al*. have annotated the oldest person Re-ID dataset VIPeR with 15 binary attributes including gender, hairstyle and some attire attributes. Later in [8], they increase the 15 attributes to 21 by introducing the dark and light colour labels for hair, shirt and pants, and extending the attribute annotations to other datasets such as PRID[4] and GRID[12]. Their PRID and GRID annotations are limited to those cross-cameras identities (200 identities for PRID and 250 identities for GRID). In 2014, Deng *et al*. released the pedestrian attributes recognition dataset - PETA[1]. It consists of 19,000 images selected from multiple surveillance datasets and provides every detail soft biometric annotations with 61 binary labels and 11 different colour labels for 4 different body regions. With a total of 105 attributes, the PETA dataset is one of the richest annotated datasets for pedestrian attribute recognition. As the PETA dataset includes many popular person Re-ID datasets such as VIPeR, PRID, GRID and CUHK, it has been integrated as soft biometric traits for some person Re-ID approaches [18, 16, 13, 19]. Due to the rapid development in deep learning approaches for person Re-ID, the dataset scales of VIPeR, PRID and GRID are too small for training the deep neural networks. The recently released two datasets such as Market-1501[26] and DukeMTMC-reID[27] with a decent size of IDs and bounding boxes provide a good amount of data for training and testing the deep Re-ID models. In [11], Lin *et al*. annotated these two large datasets with 30 and 23 respectively. All 1501 identities in Market-1501 has been annotated with 9 different binary attributes, select from 4 different age group and indicate the colour for upper and lower body parts. DukeMTMC-reID dataset has been annotated with 8 binary attributes with colours for upper and lower body. The detail statistics of the dataset can be found in Table 1.

2.2 Evaluation Metrics

The cumulative matching characteristics (CMC) curve is the most common metric used for evaluating person Re-ID performance. This metric is adopted since Re-ID is intuitively posed as a ranking problem, where each image in the gallery is ranked based on its comparison to the probe. The probability that the correct match in the ranking equal to or less than a particular value is plotted against the size of the gallery set.[2] Due to the slow training time of deep learning models, the CMC curve comparisons for recent deep Re-ID methods are simplified to only comparing Rank-1,5,10,20 retrieval rates.

Datasets	# ID	# Images	# Attributes	Attributes Annotation
VIPeR^1[2]	632	1,264	15	- 15 binary attributes[7]
VIPeR^2[2]	632	1,264	21	- 21 binary attributes[8]
PRID[4]	934	24,541	21	- Annotated shared 200 identities[8]
GRID[12]	1025	1275	21	- Annotated shared 250 identities[8]
PETA[1]	-	19.000	105	- 61 binary attributes[1] - 11 colours of 4 different regions - Include Multiple ReID Datasets: VIPeR 1264 images PRID 1134 images GRID 1275images CUHK 4563 images
Market-1501[26]	1,501	32,217	30	- 9 binary attributes[11] - 4 different age groups - 8 upper body colours - 9 lower body colours
DukeMTMC-reID[27]	1,812	36,441	23	- 8 binary attributes[11] - 8 upper body colours - 7 lower body colours

Table 1: Commonly used person Re-ID datasets with soft biometric attributes. PETA dataset contains attributes from multiple Re-ID datasets.

However, the CMC curve evaluation is valid when only one ground truth match for each given query image. The recent datasets such as Market-1501 and DukeMTMC-reID usually contain multiple ground truths for each query images. Therefore, Zheng *et al*. [26] have proposed the mean average precision (mAP) as a new evaluation metric. For each query image, the average precision (AP) is calculated as the area under its the precision-recall curve. The mean value of the average precision (mAP) will reflect the overall recall of the person Re-ID algorithm. The performances of current person Re-ID methods are usually examined by combining the CMC curve for retrieval precision evaluation and mAP for recall evaluation.

For the person Re-ID methods using the soft biometrics, the biometric attribute recognition accuracy is evaluated to proof that the proposed methods are effectively learning and utilising the given attribute information. For a comprehensive analysis of the attribute prediction, the soft biometric evaluation metrics usually include the classification accuracy for each individual attribute and an average recognition rate of all the attributes.

2.3 Supervised Deep Re-ID with Soft Biometrics Attributes

The visual appearance of a person can be easily affected by the variant of illumination, postures and viewpoints in different cameras. The soft biometrics as semantic mid-level features are invariant across cameras and provide relevant information about the person's identity. By integrating soft biometrics into person Re-ID models, the deep person re-id network should obtain more robust view-invariant feature representations. Constructing new neural network structures which can fuse the identity information with attributes information is a crucial step for fully-supervised deep person Re-ID.

2.3.1 Supervised Attribute Assisted Verification Based Person Re-ID

Triplet Loss is one commonly used loss for person Re-ID task. By inputting an anchor image with a positive sample and a negative sample, the similarity between the positive pair must be high, and the similarity between the negative pair should be low. Since the same person should have the same soft biometrics attributes, the triplet loss function can also be applied to the attributes features as well. Based on the attribute consistency, Schumann *et al*. [15] proposed an Attribute-Complementary Re-id Net (ACRN) architecture which combines the image based triplet loss with the attribute label triplet loss. The overview of the ACRN architecture is illustrated in Fig 1.

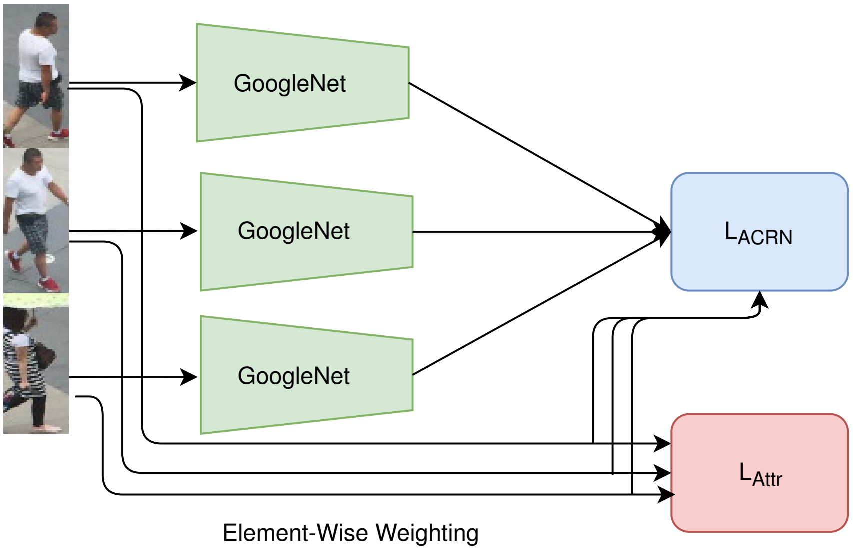


Fig. 1: Illustration of Attribute-Complementary Re-ID Net (ACRN)

The attribute recognition model is pre-trained from the PETA dataset. Then, the attribute predictions of the three input images will be used to compute the attribute branch triplet loss. With the summation of the triplet loss from the raw image branch, the overall loss function for ACRN can be expressed as following:

$$L_{ACRN} = \frac{1}{N} d_i^{f^p} - d_i^{f^n} + m + \gamma(d_i^{att^p} - d_i^{att^n}) \tag{1}$$

, where $d_i^{f^p} = \|f_i^a - f_i^p\|_2^2$, $d_i^{f^n} = \|f_i^a - f_i^n\|_2^2$, $d_i^{att^p} = \|att_i^a - att_i^p\|_2^2$, $d_i^{att^n} = \|att_i^a - att_i^n\|_2^2$. $d_i^{f^p}$ and $d_i^{f^n}$ denote the distance of between anchor positive images pair and anchor negative images pair. The term $d_i^{att^p}$ and $d_i^{att^n}$ denote the distance of the predicted attribute representations. Performance comparison with other methods can be found in Table 3. With the additional attribute branch triplet loss, there is a 2 to 4 per cent increase in both Rank1 retrieval accuracy and mAP.

2.3.2 Supervised Attribute Assisted Identification Based Person Re-ID

Unlike triplet loss based method which treats the person Re-ID problem as a verification task, Lin *et al*. [11] rethink the person Re-ID training as an identification task. They utilised the classification loss for learning people's identities and attributes. With the soft biometric annotations of two largest person Re-ID datasets: Market1510 and DukeMTMC, their proposed method simultaneously learned the attribute and ID from the same feature maps from a backbone CNN network. The overview of the Attribute-Person Recognition (APR) network is demonstrated in Fig 2

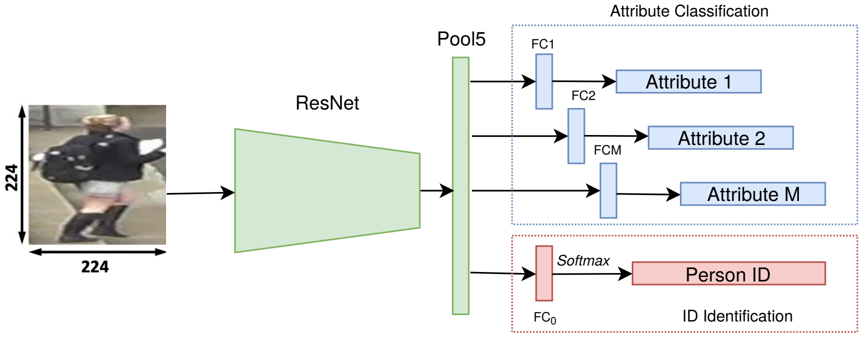


Fig. 2: Illustration of Attribute-Person Recognition (APR) network

Each input image in APR network will pass through a ResNet-50 backbone feature extractor. The extracted the feature maps will be used for both ID classification and attribute recognition. For K IDs in training, they use the Cross-Entropy Loss for ID Classification:

$$L_{ID} = -\sum_{K}^{k=1} log(p(k))q(k) \tag{2}$$

where $p(k)$ is the predicted probability with ground truth $q(k)$. The attribute prediction is using M Softmax loss:

$$L_{att} = -\sum_{m}^{j=1} log(p(j))q(j) \tag{3}$$

As the APR network is trained for both attribute prediction and identity classification, the overall loss function will be the weight summation of two losses:

$$L_{APR} = \lambda L_{ID} + \frac{1}{M}\sum_{M}^{i=1} L_{att} \tag{4}$$

By using only the classification loss for ID and attribute, the APR network is much easier to train with a quick and smooth converge compared with the triplet loss based network. The overall performances, shown in Table 3 are on a par with the triplet loss based ACRN method.

3 Semi-supervised Person Re-ID with Soft Biometrics

Although the Market-1501 and DukeMTMC-reID are labelled with 30 and 23 attributes. The dimensions of these soft biometrics labels are far from the 105 attributes PETA dataset. On the other hand, the PETA dataset does not provide enough training ID for deep learning approaches. If the PETA dataset attributes can be transferred to a new dataset, it could provide more detailed auxiliary information for person Re-ID task.

3.1 Semi-supervised Cross-dataset Attribute Learning via Co-training

To address this problem, Su *et al.* [18] proposed a Semi-supervised Deep Attribute Learning (SSDAL) algorithm. The SSDAL method uses a co-training strategy which extent the attributes in a given labelled dataset to a different unlabelled person Re-ID dataset by utilising the ID information from the new dataset. The process of SSDAL algorithm can be generalised into three stages:

1. Training on a PETA pedestrian attribute dataset labelled (excluding the target dataset) with soft biometric attributes.
2. Fine-tuning on a large dataset with only person IDs label using triplet loss on the predicted attributes based on the identity labels.
3. Updating the model predicts attribute labels for the combined dataset.

The detail illustration of SSDAL three stages can be found in Fig 3

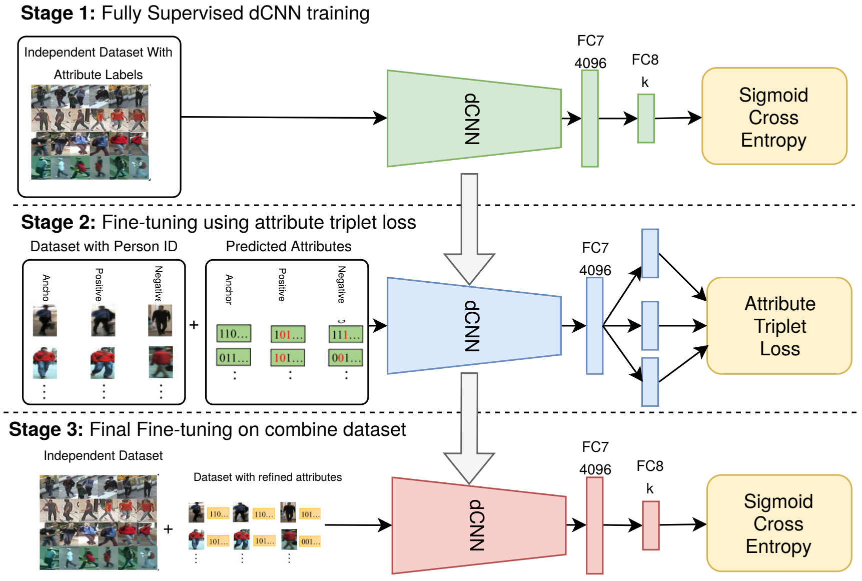


Fig. 3: Illustration of Semi-supervised Deep Attribute Learning (SSDAL) network

Fig 4 shows the attribute classification accuracy on different stages. The SSDAL strategy does give a 2 to 3 per cent increase from the baseline recognition accuracy. However, the person Re-ID matching in SSDAL is purely based on the soft biometrics

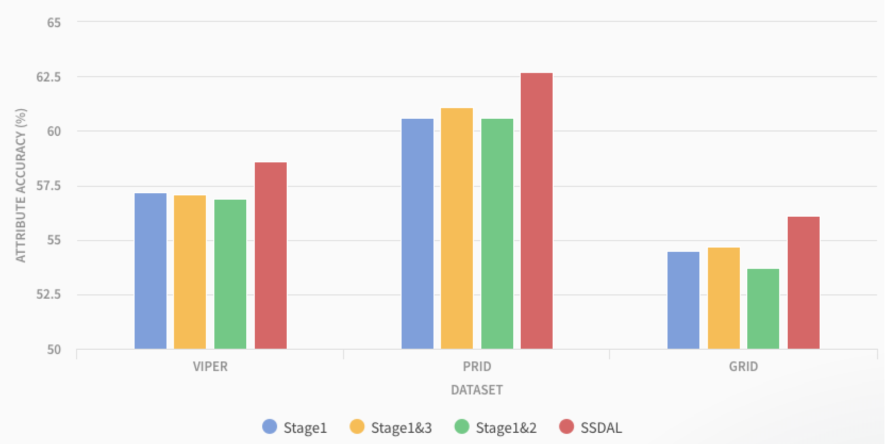


Fig. 4: Attributes Classification Accuracy on different stage

labels which is semi-supervised learned from PETA dataset. As a result, the SSDAL yielded a relatively poor Re-ID performance, as shown in Table 3.

3.2 Unsupervised Cross-Dataset Person Re-ID via Soft Biometrics

Most of the recent person re-identification (Re-ID) models follow supervised learning frameworks which require a large number of labelled matching image pairs collecting from the video surveillance cameras. However, a real-world surveillance system usually consists of hundreds of cameras. Manually tracking and annotating persons among these cameras are extremely expensive and impractical. One way to solve this issue is transferring a pre-trained model learned from existing available datasets to a new surveillance system. As the unlabelled images can be easily obtained from the new CCTV system, this issue can be considered as an unsupervised cross-dataset transfer learning problem.

In recent years, some unsupervised methods have been proposed to extract view-invariant features and measure the similarity of images without label information [21, 6, 20, 24]. These approaches only analysed the unlabelled datasets and generally yielded poor person Re-ID performance due to the lack of strong supervised tuning and optimisation. Another approaches to solving the scalability issue of Re-ID is unsupervised transfer learning via domain adaptation strategy. The unsupervised domain adaptation methods leverage labelled data in one or more related source datasets (also known as source domains) to learn models for unlabelled data in a target domain. Since the identity labels of the different Re-ID datasets are non-overlapping, the soft biometrics attributes become alternative shared domain knowledge between datasets. For example, the same set of attributes like genders, age-groups or colour/texture of the outfits can be used as universal attributes for any pedestrians across different datasets.

3.2.1 Attribute Consistency Adaptation Method

One way to utilise the soft biometric attribute in cross-dataset adaptation is to create a distance metric to quantify how well the model fits a given domain. Wang *et al*. [22] proposed the Transferable Joint Attribute-Identity Deep Learning (**TJ-AIDL**) method and introduced the Identity Inferred Attribute (IIA). By reduce the discrepancy between Identity Inferred Attribute and actual soft biometric attributes, the pre-trained model can be adapted to a different dataset. The overall architecture of TJ-AIDL is depicted in Fig 5.

TJ-AIDL method proposed two separate branches for simultaneously learning individual features of people's identity and appearance attributes. For training the identity branch, they utilised the Softmax Cross Entropy as the loss function defined as:

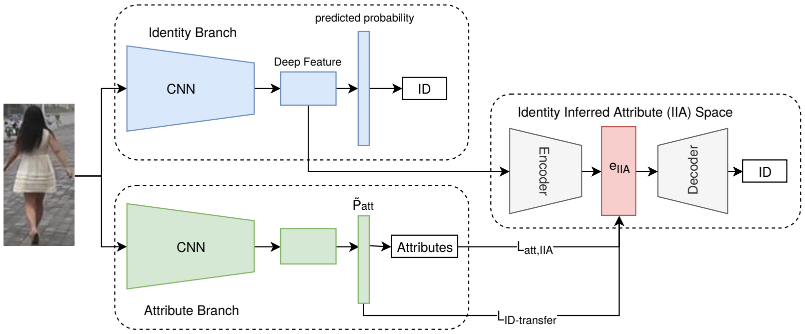


Fig. 5: Transferable Joint Attribute-Identity Deep Learning Architecture

$$L_{id} = -\frac{1}{n_{bs}} \sum_{i=1}^{n_{bs}} log(p_{id}(\boldsymbol{I}_i^s, y_i^s) \tag{5}$$

where $p_{id}(\boldsymbol{I}_i^s, y_i^s)$ specifies the predicted probability on the ground-truth class y_i^s of the training image $\boldsymbol{I}_i^s$, and n_{bs} denotes the batch size.

For training the attribute branch, they considered the attributes recognition as a multi-label classification task. Thus, the Sigmoid Cross Entropy loss function is used:

$$L_{att} = \frac{1}{n_{bs}} \sum_{i=1}^{n_{bs}} \sum_{j=1}^{m} (a_{i,j} log(p_{att}(\boldsymbol{I}_i, j)) + (1 - a_{i,j}) log(1 - p_{att}(\boldsymbol{I}_i, j))) \tag{6}$$

where $a_{i,j}$ and $p_{att}(\boldsymbol{I}_i, j))$ define the ground-truth label and the predicted classification probability on the j-th attribute class of the training image $\boldsymbol{I}_i$.

The deep feature extracted from the identity branch will then be feed into a reconstruction auto-coder (*IIA* encoder-decoder) in order to transform into a low dimensional match-able space (IIA space) for the attribute counterpart. The reconstruction loss L_{rec} is the Minimum Square Error (MSE) between reconstructed images and the ground-true images.

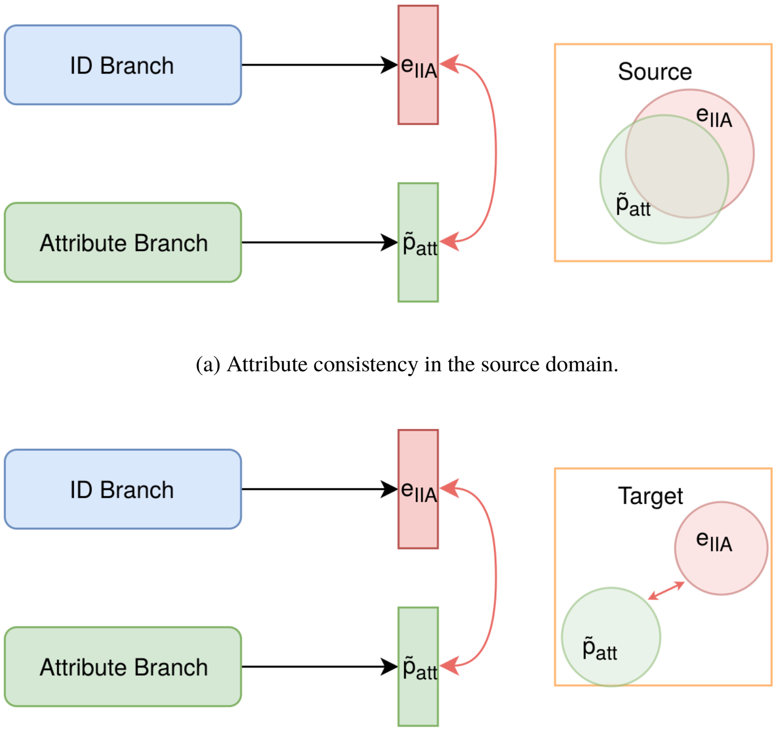
The concise feature representation (e_{IIA}) extracted from the hidden layer is aligned and regularised with the prediction distribution from attribute branch by the MSE loss:

$$L_{ID-transfer} = \|e_{IIA} - \tilde{p}_{att}\|^2 \tag{7}$$

For easy alignment purpose, an additional Sigmoid Cross Entropy is applied the e_{IIA} with pseudo attribute prediction from the attribute branch:

$$L_{attr,IIA} = \|e_{IIA} - \tilde{p}_{att}\|^2 \tag{8}$$

The overall loss for the *IIA* encoder-decoder is the weighted summation of $L_{attr,IIA}$,L_{rec},$L_{ID-transfer}$. In order to transfer the identity knowledge to the attribute branch, the identity knowledge transfer loss $L_{ID-transfer}$ is also added to the learning of the attribute classification.



(a) Attribute consistency in the source domain.

(b) Attribute consistency in the target domain.

Fig. 6: Attribute consistency between two branches.

Supervised learning of TJ-AIDL model should be able to achieve small discrepancy between the Identity Inferred Attribute (IIA) and actual attributes. However, the model trained from the source dataset may not give the same attribute consistency in the target dataset, illustrated in Fig 6. By reducing the discrepancy between two branches on the target dataset images, the source dataset trained TJ-AIDL model can be adapted to the target dataset in an unsupervised manner.

3.2.2 MMD Based Feature Alignment Adaptation Method

Another way to utilise the soft biometric attributes for cross-dataset adaptation is aligning the attribute features distribution between the source and the target dataset. Lin *et al*. [10] proposed a Multi-task Mid-level Feature Alignment (**MMFA**) network which learned the features representations from the source dataset and simultaneously aligning the attribute distribution to the target dataset based on the Maximum Mean Discrepancy (MMD). The overview of MMFA architecture is illustrated in Fig 7 below.

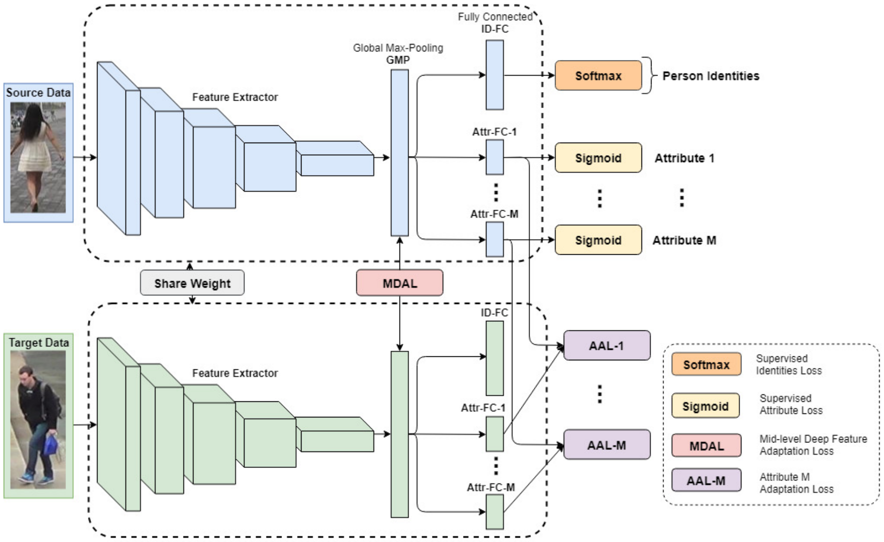


Fig. 7: Multi-task Mid-level Feature Alignment (MMFA) Architecture

The MMFA network is fundamentally a siamese neural network structure. The source and target images as the two inputs will undergo two networks with the shared weights. The global max-pooling layer will extract the most sensitive feature maps from the last convolutional layer of the ResNet-50 network. These feature vectors from the global max-pooling will be forwarded to multiple independent fully connected layers for identity classification and attribute recognition. Similar to TJ-AIDL, the MMAF model also utilised the Softmax Cross Entropy for ID identification and Sigmoid Cross Entropy for attribute classification:

$$L_{id} = -\frac{1}{n_S} \sum_{i=1}^{n_S} log(p_{id}(\mathbf{h}_{S,i}^{id}, y_{S,i})) \tag{9}$$

$$L_{attr} = -\frac{1}{M}\frac{1}{n_S}\sum_{m=1}^{M}\sum_{i=1}^{n_S}(a_{S,i}^{m}\cdot log(p_{attr}(\mathbf{h}_{S,i}^{attr_m}, m)) + (1-a_{S,i}^{m})\cdot log(1-p_{attr}(\mathbf{h}_{S,i}^{attr_m}, m))) \tag{10}$$

where $p_{id}(\mathbf{h}_{S,i}^{id}, y_{S,i})$ is the predicted probability on the identity features $\mathbf{h}_{S,i}^{id}$ with the ground-truth label $y_{S,i}$ and $p_{attr}(\mathbf{h}_{S,i}^{attr_m}, m)$ is the predicted probability for the m-th attribute features $\mathbf{h}_{S,i}^{attr_m}$ with ground-truth label $a_{S,i}^{m}$.

Since the IDs of the pedestrians are not commonly shared between two datasets, the soft biometrics attributes become a good alternative shared domain labels for coss-dataset Re-ID adaptation. As each individual attribute has its own fully connected (FC) layer. The feature vectors obtained from these FC layers: $\{\mathbf{H}_{S}^{attr_1}, .., \mathbf{H}_{S}^{attr_M}\}$, $\{\mathbf{H}_{T}^{attr_1}, .., \mathbf{H}_{T}^{attr_M}\}$ can be considered as the features of the corresponding attributes. Therefore, by aligning the distribution of each attribute between the source and the target datasets, the model can unsupervised adapted to the target dataset. In MMFA models, it utilised the Maximum Mean Discrepancy (MMD) measure [3] to calculate the feature distribution distance of each attribute. The final loss for the attribute distribution alignment is the mean MMD distance of all attributes:

$$L_{AAL} = \frac{1}{M}\sum_{m=1}^{M} MMD(\mathbf{H}_{S}^{attr_m}, \mathbf{H}_{T}^{attr_m})^2 \tag{11}$$

However, the 30 and 23 attributes provided by the the current Market-1501 and DukeMTMC-reID dataset are insufficient to represent all mid-level features of the dataset. Thus, the MMFA also introduced the mid-level deep feature alignment loss L_{MDAL} for reduce the despondency of the last Global Max-Pooling Layer $(\mathbf{H}_S, \mathbf{H}_T)$:

$$L_{MDAL} = MMD(\mathbf{H}_S, \mathbf{H}_T)^2 \tag{12}$$

.

RBF characteristic kernels with selected bandwidth α=1,5,10 are the main kernel functions used in all MMD loss.

$$k(\mathbf{h}_{S,i}^{attr_m}, \mathbf{h}_{T,j}^{attr_m}) = exp(-\frac{1}{2\alpha}\left\|\mathbf{h}_{S,i}^{attr_m} - \mathbf{h}_{T,j}^{attr_m}\right\|^2) \tag{13}$$

The overall loss will be the weighted summation of all the mention losses above:

$$L_{all} = L_{id} + \lambda_1 L_{attr} + \lambda_2 L_{AAL} + \lambda_3 L_{MDAL} \tag{14}$$

4 Performance Comparison

The attribute recognition accuracy is a good measurement for evaluating how well the attribute information is integrated into the person Re-ID system. Table 2 shows the mean attribute recognition accuracy for ACRN and APR approaches. The ACRN model trained from PETA dataset can achieve 84.61% mean recognition accuracy of all 105 attributes. The APR model can also accomplish the over 85% recognition rate with 23/30 attributes. However, the SSDAL methods did not train on the entire set of PETA system. The different recognition accuracy can be found in Fig 4. The unsupervised cross-dataset methods TJ-AIDL and MMFA did not provide the attribute recognition rate due to the different sets of attribute labels used for the different dataset.

	PETA	Market-1501	DukeMCMT-reID
ACRN	84.61%		
APR		88.16%	86.42%

Table 2: mean Attribute Recognition Accuracy (mA) for ACRN and APR

Table 3 shows the detailed comparison for supervised with and without soft biometrics, semi-supervised attribute transfer and unsupervised cross-dataset transfer. By integrating the attribute information, the supervised methods with attributes can usually outperform the deep supervised methods by 4 to 6 per cent. The semi-supervised method is based on the transferred attribute features only. It give a relatively weak Re-ID performance due to the lack of local features information. The unsupervised cross-dataset person Re-ID methods such as TJ-AIDL and MMFA show promising performances compared with the fully supervised methods.

Dataset		**VIPeR**	**PRID**	**Market-1501**		**DukeMCMT-reID**	
Metric (%)		Rank-1	Rank-1	Rank-1	mAP	Rank-1	mAP
Supervised without Soft Biometric	GAN[27]	-	-	79.3	56.0	67.7	47.13
	PIE[25]	-	-	78.1	56.2	-	-
Supervised	ACRN	-	-	83.6	62.6	72.6	52.0
	APR	-	-	84.3	64.7	70.7	51.9
Semi-supervised	SSDAL	37.9	20.1	39.4	19.6	-	-
Unsupervised	TJ-AIDL$^{\text{Duke}}$	35.1	34.8	58.2	26.5	-	-
	MMFA$^{\text{Duke}}$	36.3	34.5	56.7	27.4	-	-
	TJ-AIDL$^{\text{Market}}$	38.5	26.8	-	-	44.3	23.0
	MMFA$^{\text{Market}}$	39.1	35.1	-	-	45.3	24.7

Table 3: Performance comparisons with state-of-the-art unsupervised person Re-ID methods. The superscripts: *Duke* and *Market* indicate the source dataset which the model is trained on.

5 Existing Issues and Future Trends

The existing Market-1501 and DuketMTMC-reID have a limited set of soft biometric attributes. The current 23/30 attributes cannot easily distinguish most of the person in these two datasets. To fully exploit the soft biometrics information for person Re-ID, a large scale and richly annotated person Re-ID dataset is definitely needed. However, soft biometric attributes are expensive to annotated especially for the large scale dataset. Unfortunately, recently released dataset such as MSMT17 [23] did not come with soft biometric annotations.

The second problem in soft biometrics is the lack of the standardisation guideline when annotating the attributes. In Market-1501, there are 8 upper body colours and 9 lower body colours. In DukeMTMC-reID, the annotated colours for upper-body and lower-body become 8 and 7. The age group separation definitions are different between PETA dataset and Market-1501. The clothing types are also inconsistent between PETA, Market-1501 and DukeMTMC-reID. If the all the dataset following the same annotation guideline, soft biometric attributes can be used and evaluated in the cross-dataset or multi-dataset person re-id scenarios.

The soft biometrics should not be limited to boosting the person Re-ID performance or leverage for cross-dataset adaptation. How to utilise the existing attributes and extending to the natural language description based person retrieval will be an interesting topic to research on in the future.

6 Conclusion

Soft biometric attributes are useful auxiliary information for person Re-ID task. In supervised deep person Re-ID models, soft biometric information could improve the person Re-ID performance by 2 to 4 per cent. When transferring a Re-ID model from one camera system to another, the soft biometric attributes can be used as a share domain knowledge for domain adaptation task. In addition, the attributes can also be transferred via a semi-supervised deep co-training strategy from a richly annotated dataset to an unlabelled Re-ID dataset. The performance for cross-dataset Re-ID model adaptation or attributes transfer is still far away from optimal. There is ample space for improvement and massive potential for these application. Currently, there are only a few person Re-ID dataset with soft biometric information. The soft biometric attributes are very limited and inconsistent across different datasets. How to effectively obtain the soft biometrics and establish an annotation guideline will be a meaningful research work for the person Re-ID problem.

References

1. Deng, Y., Luo, P., Loy, C.C., Tang, X.: Pedestrian Attribute Recognition At Far Distance. In: ACM International Conference on Multimedia (ACM MM) (2014)
2. Gray, D., Brennan, S., Tao, H.: Evaluating Appearance Models for Recognition, Reacquisition, and Tracking. In: International Workshop on Performance Evaluation for Tracking and Surveillance (PETS), vol. 3, pp. 41–47 (2007)
3. Gretton, A., Fukumizu, K., Harchaoui, Z., Sriperumbudur, B.K.: A Fast, Consistent Kernel Two-Sample Test. In: Advances in Neural Information Processing Systems (NIPS) (2009)
4. Hirzer, M., Beleznai, C., Roth, P.M., Bischof, H.: Person Re-identification by Descriptive and Discriminative Classification. In: Scandinavian Conference on Image Analysis (SCIA) (2011)
5. Khamis, S., Kuo, C.H., Singh, V.K., Shet, V.D., Davis, L.S.: Joint Learning for Attribute-consistent Person Re-identification. In: European Conference on Computer Vision Workshops (ECCVW) (2014)
6. Kodirov, E., Xiang, T., Gong, S.: Dictionary Learning with Iterative Laplacian Regularisation for Unsupervised Person Re-identification. In: British Machine Vision Conference (BMVC) (2015)
7. Layne, R., Hospedales, T.M., Gong, S.: Person Re-identification by Attributes. In: British Machine Vision Conference (BMVC). British Machine Vision Association (2012)
8. Layne, R., Hospedales, T.M., Gong, S.: Attributes-Based Re-identification. In: Person Re-Identification, pp. 93–117. Springer London (2014)
9. Layne, R., Hospedales, T.M., Gong, S.: Re-id: Hunting Attributes in the Wild. In: British Machine Vision Conference (BMVC), pp. 1–1. British Machine Vision Association (2014)
10. Lin, S., Li, H., Li, C.t., Kot, A.C., Alignment, M.l.F., Unsupervised, F.O.R.: Multi-task Mid-level Feature Alignment Network for Unsupervised Cross-Dataset Person Re-Identification. In: British Machine Vision Conference (BMVC) (2018)
11. Lin, Y., Zheng, L., Zheng, Z., Wu, Y., Yang, Y.: Improving Person Re-identification by Attribute and Identity Learning. In: arXiv preprint (2017)
12. Loy, C.C., Xiang, T., Gong, S., Change, C., Tao, L., Shaogang, X., Loy, C.C., Xiang, T., Gong, S.: Time-Delayed Correlation Analysis for Multi-Camera Activity Understanding. International Journal of Computer Vision (IJCV) (2010)
13. Matsukawa, T., Suzuki, E.: Person Re-Identification Using CNN Features Learned from Combination of Attributes. In: International Conference on Pattern Recognition (ICPR) (2016)
14. Reid, D.A., Samangooei, S., Hen, C., Nixon, M.S., Ross, A.: Soft Biometrics for Surveillance: An Overview. In: Handbook of Statistics (2013)
15. Schumann, A., Stiefelhagen, R.: Person Re-identification by Deep Learning Attribute-Complementary Information. In: Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (2017)
16. Shi, Z., Hospedales, T.M., Xiang, T.: Transferring a Semantic Representation for Person Re-identification and Search. In: Conference on Computer Vision and Pattern Recognition (CVPR). IEEE (2015)
17. Su, C., Yang, F., Zhang, S., Tian, Q.: Multi-Task Learning with Low Rank Attribute Embedding for Person Re-identification. In: International Conference on Computer Vision (ICCV) (2015)
18. Su, C., Zhang, S., Xing, J., Gao, W., Tian, Q.: Deep Attributes Driven Multi-camera Person Re-identification. In: European Conference on Computer Vision (ECCV) (2016)
19. Su, C., Zhang, S., Xing, J., Gao, W., Tian, Q.: Multi-type Attributes Driven Multi-camera Person Re-identification. Pattern Recognition (2017)
20. Wang, H., Gong, S., Xiang, T.: Unsupervised Learning of Generative Topic Saliency for Person Re-identification. In: British Machine Vision Conference (BMVC) (2014)
21. Wang, H., Zhu, X., Xiang, T., Gong, S.: Towards Unsupervised Open-Set Person Re-identification. In: International Conference on Image Processing (ICIP) (2016)
22. Wang, J., Zhu, X., Gong, S., Li, W.: Transferable Joint Attribute-Identity Deep Learning for Unsupervised Person Re-Identification. In: Conference on Computer Vision and Pattern Recognition (CVPR) (2018)

23. Wei, L., Zhang, S., Gao, W., Tian, Q.: Person Transfer GAN to Bridge Domain Gap for Person Re-Identification. In: Conference on Computer Vision and Pattern Recognition (CVPR) (2018)
24. Yu, H.X.X., Wu, A., Zheng, W.S.S.: Cross-View Asymmetric Metric Learning for Unsupervised Person Re-Identification. In: International Conference on Computer Vision (ICCV) (2017)
25. Zheng, L., Huang, Y., Lu, H., Yang, Y.: Pose Invariant Embedding for Deep Person Re-identification. In: arXiv preprint (2017)
26. Zheng, L., Shen, L., Tian, L., Wang, S., Wang, J., Tian, Q.: Scalable Person Re-identification: A Benchmark. In: International Conference on Computer Vision (ICCV) (2015)
27. Zheng, Z., Zheng, L., Yang, Y.: Unlabeled Samples Generated by GAN Improve the Person Re-identification Baseline in Vitro. In: International Conference on Computer Vision (ICCV) (2017)