Prediction and Recovery for Adaptive Low-Resolution Person Re-Identification

Ke Han[1,2], Yan Huang[1], Zerui Chen[1], Liang Wang[1,3,4], and Tieniu Tan[1,3]

[1] Center for Research on Intelligent Perception and Computing (CRIPAC),
National Laboratory of Pattern Recognition (NLPR),
Institute of Automation, Chinese Academy of Sciences (CASIA)
[2] School of Future Technology, University of Chinese Academy of Sciences (UCAS)
[3] Center for Excellence in Brain Science and Intelligence Technology (CEBSIT)
[4] Chinese Academy of Sciences, Artificial Intelligence Research (CAS-AIR)
{ke.han,zerui.chen}@cripac.ia.ac.cn, {yhuang,wangliang,tnt}@nlpr.ia.ac.cn

Abstract. Low-resolution person re-identification (LR re-id) is a challenging task with low-resolution probes and high-resolution gallery images. To address the resolution mismatch, existing methods typically recover missing details for low-resolution probes by super-resolution. However, they usually pre-specify fixed scale factors for all images, and ignore the fact that choosing a preferable scale factor for certain image content probably greatly benefits the identification. In this paper, we propose a novel Prediction, Recovery and Identification (PRI) model for LR re-id, which adaptively recovers missing details by predicting a preferable scale factor based on the image content. To deal with the lack of ground-truth optimal scale factors, our model contains a self-supervised scale factor metric that automatically generates dynamic soft labels. The generated labels indicate probabilities that each scale factor is optimal, which are used as guidance to enhance the content-aware scale factor prediction. Consequently, our model can more accurately predict and recover the content-aware details, and achieve state-of-the-art performances on four LR re-id datasets.

Keywords: Low-resolution person re-identification; Adaptive scale factor prediction; Dynamic soft label

1 Introduction

Given a person image captured by a certain camera, person re-identification (re-id) aims to identify the same person across different cameras. With more and more video surveillance in public places, this task has attracted wider attention of both academia and industry, because of its great application potentials. Most researchers study this topic under the assumption that all the available person images have sufficient and similar resolutions. In real-world application scenarios, however, the resolutions of captured persons may vary greatly due to the uncontrollable distances between persons and cameras. Generally, target persons

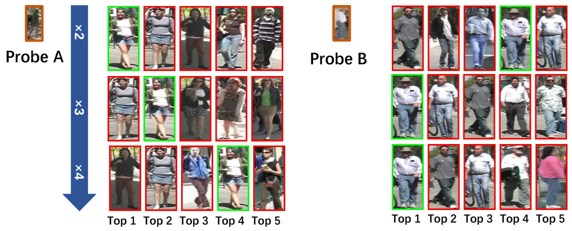

Fig. 1. Top-ranked HR gallery images of two LR probes [11] with different scale factor settings. The ground truth is indicated by a green bounding box. For the same probe, different scale factors might lead to different search results.

of high resolution (HR) are enrolled as the gallery set, while probe persons captured by surveillance cameras have low resolution (LR). This common problem is usually referred as the low-resolution person re-identification (LR re-id).

To deal with the resolution mismatch problem in LR re-id, existing works [17, 39] typically recover missing details for LR images with super-resolution (SR) modules. In fact, performing SR has its pros and cons. We resort to it to recover details but it will inevitably bring about noise in the meanwhile. Especially when we use a larger scale factor for SR, the produced noise might greatly degenerate the image quality and identification results. Nevertheless, existing works usually ignore the problem and pre-specify a fixed scale factor for images, regardless of whether it can recover the most effective details and not incur excessive noise to degenerate the identification. We experimentally find that choosing different scale factors for the same LR probe may lead to quite different search results, as shown in Figure 1. For example, the ground truth of the probe A drops from the first to fourth in the ranking lists, when we change the scale factor from 2 to 4. The reason probably lies in that the dazzling shadow and background of the probe A are more prone to result in noisy recovery as the scale factor increases. Intuitively, the image content should be an important cue to determine which scale factor can achieve better recovery for accurate identification.

However, how to combine the image content for choosing a preferable scale factor is seldom investigated and has many challenges. A tough one is that we have no prior annotation indicating which scale factor is optimal to identify a given person. In practice, it is also difficult and time-consuming to pre-define such an optimal scale factor. The optimal scale factor should not be consistent for different re-id modules, even for the same re-id module at different timesteps during training. This is because they have varying abilities in distinguishing the recovered details and noise. In addition, one LR image sometimes has multiple scale factors that achieve comparably good results, *e.g.*, the scale factor 3 and 4 are almost equally suitable to the probe B in Figure 1. In view of the variability and multiplicity of optimal scale factors, realizing the content-aware scale factor prediction remains challenging.

To address the problem, we propose a novel Prediction, Recovery and Identification (PRI) model for LR re-id, which can adaptively recover details by predicting a preferable scale factor for a given image based on image content. Our model formulates the scale factor prediction as a classification problem, by choosing a preferable scale factor from a set of pre-defined ones for a given image. Unlike the typical classification setting [20, 32] where an object has a one-hot label indicating its class, our scale factor prediction suffers from the mentioned problems of label variability and multiplicity. To this end, we propose a scale factor metric that automatically assigns a given LR image a dynamic soft label, *i.e.*, a normalized real-value vector. The dynamic soft label indicates the relative probabilities that each alternative scale factor is optimal, which is formulated by comparing the recovered contents by different scale factors. The *dynamic* property can dynamically evaluate and adjust the optimal scale factors during training. And the *soft* property allows multiple optimal scale factors in the form of probability and flexibly handles their variations. To enhance the content-aware scale factor prediction, the generated dynamic soft labels are exploited as supervision to guide our model to predict the preferable scale factor based on the given LR image content. Abundant experimental results show that our proposed model is effective and achieves the state-of-the-art performances on four LR re-id datasets. Besides, our proposed adaptive scale factor prediction can be used for standard re-id models to improve their performances in the LR re-id setting. The contributions of this paper are summarized as follows.

- This paper focuses on a practical but rarely investigated LR re-id problem, *i.e.*, how to choose a better scale factor for identification based on the LR image content.
- We propose a novel PRI model, which can adaptively predict the preferable scale factor, recover details for LR images, and perform the identification in an end-to-end manner.
- Without annotations of the optimal scale factors, we propose a self-supervised scale factor metric that evaluates the dynamic soft label as supervision.
- We conduct extensive experiments to demonstrate the effectiveness of our model, and achieve the state-of-the-art results on four LR re-id datasets.

2 Related Work

Standard Person Re-identification (re-id). Person re-identification [23, 31, 2, 3, 41, 29] has made great progress, with the significant development of deep learning in the past years. Many approaches have been proposed to extract more discriminative identity features. For example, to align and improve local features, PCB [36] divides the deep feature maps into several stripe features, aligns each stripe and identifies them one by one. Martinel *et al.* [27] observe that body partitions should have different importances at different scales of features. They accordingly propose PyrNet that exploits pyramid features to capture image relevancy at different levels of detail. In addition, some works pay attention to addressing some challenging re-id problems, such as background bias [33, 16], occlusion[13, 28, 35, 14] and domain adaption [34, 9, 4, 40, 30].

Low-Resolution Person Re-identification (LR re-id). Among various re-id challenges, resolution mismatch is a practical but less studied problem. Super-resolving LR images by SR modules is a common approach. For example, Jiao *et al.* [17] propose a SR and re-id joint formulation, but they enlarge LR images with the preset fixed scale factors, which probably results in the suboptimal recovery to identify the person. Wang *et al.* [39] assign a scale factor for an image depending on the relative image size to acquire the super-resolved images with the uniform size. However, there seems to be no necessary relationship between the image size and the optimal scale factor, and therefore their performances are also limited. Besides, there is another kind of method aiming at learning resolution-invariant representations without requiring SR. For example, Chen *et al.* [5] exploit adversarial learning to pull the identity-related and resolution-unrelated feature maps closer. Compared with the SR-based methods, they do not take advantages of compensated details for more fine-grained analyses. Li *et al.* [25] propose to recover details while learning resolution-invariant representations. This method combines the merits of the above two kinds of methods, but recovers LR images into the same resolution, which still suffers from the problem of the suboptimal recovery level. Different from these works, we propose to adaptively predict a preferable scale factor for each image based on the image content, so that we can achieve better recovery to improve the re-id accuracy.

Image Super-Resolution (SR). Given a LR image and a scale factor, image super-resolution $[8, 21, 37, 26, 42]$ recovers a HR image of the desired size. Although we want to employ SR modules to alleviate the resolution mismatch problem, there is a clear target difference between SR and person re-id. SR is designed for estimating low-level pixel values, which pursues the good pixel-level approximation or high visual quality, while re-id aims at learning the high-level identity discrimination. To improve the compatibility of SR and re-id modules, we integrate them into a joint-learning network. With the joint supervision of the ground-truth HR images and identity signals, our model learns the re-id oriented detail recovery to facilitate the identification.

3 Method

3.1 Overview

Given a LR image, we aim to adaptively predict and recover the content-aware details to achieve more accuracy identification. Our proposed Prediction, Recovery and Identification (PRI) model is illustrated in Figure 2 and outlined as follows. The LR re-id dataset generally contains pairs of HR and LR images of the same identity but captured by different cameras. Inspired by [17], PRI takes as input such a pair of images along with a synthetic LR image down-sampled by the HR one. We denote such an input set composed of a HR, a LR and a synthetic LR image as $\{x_h, x_l, x_{sl}\}$ respectively for the following description.

During training, the LR image x_l is sent to the adaptive scale factor predictor P, which formulates the scale factor prediction as a N-class classification

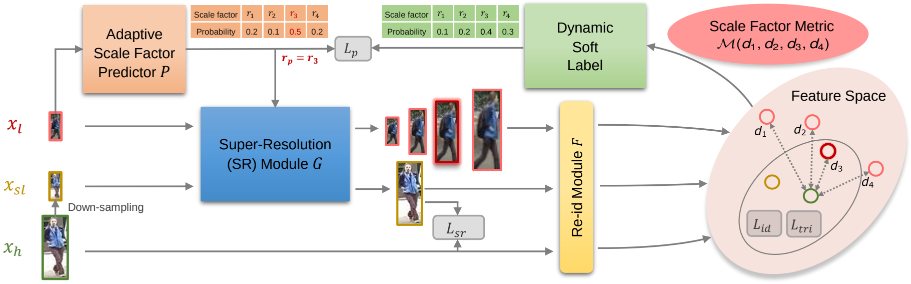
Adaptive Scale Factor Predictor P
Scale factor r_1 r_2 r_3 r_4
Probability 0.2 0.1 0.5 0.2
L_p
$r_p = r_3$
Scale factor r_1 r_2 r_3 r_4
Probability 0.1 0.2 0.4 0.3
Dynamic Soft Label
Scale Factor Metric $\mathcal{M}(d_1, d_2, d_3, d_4)$
x_l
x_{sl}
x_h
Down-sampling
Super-Resolution (SR) Module G
L_{sr}
Re-id Module F
Feature Space
d_1
d_2
d_3
d_4
L_{id}
L_{tri}

Fig. 2. The proposed Prediction, Recovery and Identification (PRI) model.

problem, and predicts the probabilities that x_l belongs to each class. N classes refer to N alternative scale factors $\{r_1, r_2, \cdots, r_N\}$, and we illustrate with N=4 as an example in Figure 2. At the same time, x_l is also super-resolved with the SR module by all the alternative scale factors. All the super-resolved images are projected into the common feature space, where we perform the scale factor metric to acquire the dynamic soft label. The dynamic soft label is then fed back to the predictor P, and serves as the supervision guiding P to predict the desired scale factor via the prediction loss L_p. Besides, we exploit the SR loss L_{sr}, identity loss L_{id} and triplet loss L_{tri} to learn the effective detail recovery and identity discrimination. When test, we only need to super-resolve a LR image by the scale factor r_p, which has the maximum predicted probability and is more likely to achieve better recovery and identification results than the other scale factors. We will elaborate each part of our model in the following sections.

3.2 Adaptive Scale Factor Predictor

Selecting which scale factor to recover missing details is a practical problem in LR re-id. Intuitively, the optimal scale factor is probably inherently related to the image content. This inspires us to predict a preferable scale factor for a LR image based on its image content. Ideally, we want to recover helpful appearance details as much as possible, and control the undesired noise in an acceptable range that does not adversely affect the identification.

In this paper, we formulate predicting a preferable scale factor as a N-class classification problem by presetting N alternative scale factors $r_1, r_2, \cdots, r_N$. The preset scale factors should have a proper varying interval. It makes little sense to choose a better scale factor from several fairly close ones, *e.g.*, 2.1, 2,2 and 2.3, because they have nearly the same recovery effects. Experimentally, we set the interval to 1, the number of classes N to 4, and $\{r_1, r_2, \cdots, r_N\}$ to $\{1, 2, 3, 4\}$, respectively. We accordingly design an adaptive scale factor predictor P as a classifier. Given an image, P extracts the features and predicts the probabilities that the image belongs to each class, which are also the probabilities that each alternative scale factor is the optimal one. We choose ResNet50 [12] as the backbone of our predictor, and use 1×1 convolutional layers after global average pooling to reduce the dimension from 2048 to 512. Then, a fully-connected layer and the softmax function are used to predict the normalized probability $p_{r_i}(i = 1, 2, \cdots, N)$ that r_i is the optimal scale factor. We formulate it as

$$p_{r_1}, p_{r_2}, \cdots, p_{r_N} = P(x_l). \tag{1}$$

Thus, the predicted optimal scale factor $r_p = \arg\max_{r_i} P(x_l)$. However, unlike the common classification setting [20, 32], we have no ground-truth optimal scale factors as supervision to enable the supervised learning of the predictor. To address this problem, we propose to regard the scale factor that can recover the most discriminative details as the ground-truth optimal one. To realize that, we need a SR module to perform the detail recovery.

3.3 Person Super-Resolution

To compare the recovery effects of different scale factors, we design a SR module G that can super-resolve a given image with multiple scale factors. Inspired by Meta-SR [15], the SR module G is composed of feature extraction layers and the meta-upscale layers. Many basic SR modules could be adopted as our feature extraction layers (we choose RDN [42] in this paper due to its competitive image recovery performance) which extract the feature maps for the given image. The meta-upscale layers consist of two fully-connected layers and a ReLU activation layer between them. They take the height, width and scale factor of the LR image as input, and predict the corresponding weights of convolution filters so that the feature maps can be upscaled with the given scale factor.

To find out which scale factor can recover the most discriminative details for x_l, we send x_l along with all the alternative scale factors $\{r_1, r_2, \cdots, r_N\}$ into G. Then we can acquire the recovered images by each scale factor, and denote them as $\{G_{r_1}(x_l), G_{r_2}(x_l), \cdots, G_{r_N}(x_l)\}$, respectively. Different from x_l, x_{sl} is super-resolved by one randomly chosen scale factor $r_{sl} \in \{r_1, r_2, \cdots, r_N\}$. Thus, x_{sl} and x_h constitute a pair of LR input and HR supervision to ensure that the SR module can be optimized by the SR loss L_{sr}. L_{sr} calculates the pixel-to-pixel 1-norm distance between the super-resolved image $G_{r_{sl}}(x_{sl})$ and the ground truth x^h. We formulate it as

$$L_{sr} = \frac{1}{r_{sl}^2 WH} \sum_{i=1}^{r_{sl}W} \sum_{j=1}^{r_{sl}H} |(G_{r_{sl}}(x_{sl}))_{i,j} - (x_h)_{i,j}| \tag{2}$$

where $X_{i,j}$ is the pixel value at the coordinate (i, j) of the image X, and W, H and r_{sl} are the width, height and scale factor of x_{sl}, respectively. The synthetic LR image x_{sl} contributes to the SR loss, which unites the latter identity loss to make it possible to jointly optimize the SR and re-id module. This shows the important role of x_{sl} in bridging SR and re-id two originally separate tasks.

3.4 Scale Factor Metric and Dynamic Soft Label

Scale factor metric. After obtaining the recovered images by all the alternative scale factors, we need to evaluate the effectiveness of the recovered contents. We regard the recovered image that contains the most discriminative details as the best recovery, and the corresponding scale factor as the optimal one. Based on this assumption, we propose a scale factor metric $\mathcal{M}$, which is a feature-based evaluation criterion by comparing which recovered image of $\{G_{r_1}(x_l), G_{r_2}(x_l), \cdots, G_{r_N}(x_l)\}$ has the most discriminative identity features. Specifically, we use a re-id module F to project all these super-resolved images into the common feature space. Similar to the adaptive scale factor predictor, we adopt ResNet50 as the re-id backbone and reduce the dimension of features. We denote the feature vectors of the above recovered images as $\{f_{r_1}, f_{r_2}, \cdots, f_{r_N}\}$, respectively. To measure which one is the most discriminative, we exploit a HR image of the same identity (*i.e.*, x_h) as an anchor, and compare the Euclidean distances among the features of the recovered images $(f_{r_1}, f_{r_2}, \cdots, f_{r_N})$ and the anchor (denoted as f_{x_h}).

Intuitively, a preferable scale factor should have a smaller relative distance to the HR anchor, due to the better detail recovery. To measure and compare the relative distances between different scale factors and the HR anchor, the proposed scale factor metric $\mathcal{M}$ is formulated as follows.

$$l_{r_i} = \mathcal{M}(d_1, \cdots, d_N) = softmax((\frac{1}{N}\sum_{j=1}^{N} d_j - d_i)^{\gamma}) \tag{3}$$

where d_i is the Euclidean distance between f_{r_i} and f_{x_h}, and γ is the regulatory factor. l_{r_i} can indicate the relative probability that r_i is the ground-truth optimal scale factor, which is normalized by the softmax function. Note that γ should be an odd number to make sure that the scale factor with a smaller feature distance than the average $(\frac{1}{N}\sum_{j=1}^{N} d_j)$ is endowed with a higher probability of being the optimal one. And the scale factor with the maximum probability is considered as the ground-truth optimal one for x_l.

Dynamic soft label. We can use the measured optimal scale factor as a one-hot label to enable the supervised learning of the predictor P via the cross-entropy classification loss. However, the optimal scale factor is often not consistent, and varies during training the re-id module, which will cause the dramatic change from a one-hot label to another. The frequent change of the one-hot label will confuse the cross-entropy loss, which typically encourages a higher probability for the only one correct class as much as possible, and make the loss function hard to converge.

To address the problem, we exploit Equation 3 to constitute the dynamic soft label, *i.e.*, a normalized real-value vector $(l_{r_1}, l_{r_2}, \cdots, l_{r_N})$. It has the following advantages. First, the change of the optimal scale factor becomes smoother in the form of the relative probability. Second, it allows to activate multiple optimal scale factors: the scale factor with the higher probability is not necessarily optimal, but more likely to be. We set an update frequency ω to determine the frequency of updating dynamic soft labels, which indicates that we perform the scale factor metric and obtain the dynamic soft label for each LR image per ω training epochs, and keep them unchanged between two updates.

Then, we use the dynamic soft label as the ground truth to supervise the predicted results of the predictor P in Equation 1. We accordingly exploit a soft cross-entropy prediction loss L_p, which is calculated as

$$L_p = -\sum_{i=1}^{N} l_{r_i} \log p_{r_i}. \tag{4}$$

Note that we cut off the error back-propagation from L_p to the dynamic soft label. In other words, considering the evaluated dynamic soft label as the ground truth, L_p is only used to optimize the predictor rather than the SR or re-id module. We minimize L_p to supervise the predictor to make an prediction consistent with the dynamic soft label, *e.g.*, predicting a higher probability for the scale factor that is more likely to be evaluated as the optimal one.

3.5 Optimization

Overall loss. To learn the discriminative identity features for re-id, we send f_{r_p}, $f_{x_{sl}}$ and f_{x_h} (the feature vectors of the predicted optimal recovered image x_{r_p}, the synthetic image x_{sl} and the HR image x_h, respectively) into a classifier (*i.e.*, a fully-connected layer) to predict the identities. This process is supervised by the cross-entropy identity loss L_{id} and triplet loss L_{tri}. L_{tri} is defined as

$$L_{tri} = \max(0, \phi + d_p - d_n) \tag{5}$$

where d_p and d_n are respectively the distances between the positive samples with the same identity and negative samples with different identities. ϕ is the margin parameter. We optimize the whole network by minimizing the weighted sum of the SR loss L_{sr}, identity loss L_{id}, triplet loss L_{tri} and prediction loss L_p. The total loss L is formulated as

$$L = L_{sr} + \alpha L_{id} + \beta L_{tri} + \lambda L_p \tag{6}$$

where α, β and λ are the weight factors of L_{id}, L_{tri} and L_p, respectively. Different from the first three losses supervising the SR and re-id module, L_p is used for the predictor separately, and therefore we set its weight λ to 1.

When test, we only need to super-resolve a LR probe by the scale factor r_p that has the maximum predicted probability. We embed the super-resolved probe and all the HR gallery images into the feature space, where we measure the similarity among their features by the Euclidean distances.

Pre-training. Experimentally, if we randomly initialize PRI for training, the dynamic soft label might vary frequently at the early training stage and degenerate the optimization process. Since the untrained SR module produces poorly recovered images, the features corresponding to different scale factors do not have relatively stable distances to the anchor, thus degenerating the effectiveness of the dynamic soft label. To alleviate the problem, we pre-train the SR and re-id module before jointly training the whole model. Specifically, we remove the prediction loss L_p, and the SR module super-resolves both x_l and x_{sl} only by a randomly chosen scale factor from the alternative ones. We only minimize the sum of L_{sr}, L_{id} and L_{tri} during pre-training, so that the SR and re-id module can learn to stably recover images and extract their features in advance, which will help to train the whole model more effectively.

4 Experiment

4.1 Datasets and Evaluation Protocol

Datasets. We evaluate our model on three synthetic and one genuine LR re-id dataset. 1) MLR-CUHK03 is built from CUHK03 [22], containing over 14,000 images of 1,467 identities captured by 5 pairs of cameras. Following [17], for a pair of images from two cameras, we down-sample one of them by randomly choosing a down-sampling factor $r \in \{2, 3, 4\}$ as a LR probe, while the other remains unchanged as a HR gallery image. Two types of images, manually cropped and automatically detected images, are both used. 2) MLR-DukeMTMC-reid [44] includes 36, 411 images of 1, 404 identities captured by 8 cameras. 3) MLR-Market1501 [43] consists of 32, 668 images of 1, 501 identities from 6 camera views. Both MLR-DukeMTMC-reid and MLR-Market1501 are synthesized by the same down-sampling operation as MLR-CUHK03. 4) CAVIAR [6] is a genuine dataset composed of 1220 images of 72 identities and two camera views. We discard 22 identities that only appear in the closer camera.

Evaluation protocol. We adopt the standard *single-shot* person re-id setting. Images of CAVIAR are randomly and evenly divided into two halves for training and test, which means that there are 25/25 identities in the training/test set. We use the 1,367/100, 702/702 and 751/750 training/test identity split on MLR-CUHK03, MLR-DukeMTMC-reid and MLRMarket1501, respectively. For test, we build the probe set with all the LR images, and the gallery set with one randomly selected HR image of each person. For test, we build the probe set with all the LR images, and the gallery set with one randomly selected HR image of each person. Above random data splits are repeated 10 times in the experiments. For the re-id performance evaluation, we use the average Cumulative Match Characteristic (CMC) and report results at ranks 1, 5 and 10.

Implementation details. Our ResNet50 backbone (for P and F) is pre-trained on ImageNet [7], and the SR module G is pre-trained on DIV2K [1]. All the images sent into ResNet50 are resized to 384×128×3. A set of input images (x_l, x_{sl} and x_h) is randomly flipped horizontally at the same time. We pre-train PRI for T_1 epochs as stated in Section 3.5, and then further train the whole PRI model for T_2 epochs. We adopt the Adam optimizer [19] ($\beta_1 = 0.9$ and $\beta_2 = 0.999$), and set the initial learning rate of the ResNet50 backbone and the added 1×1 convolutional layers to 0.01 and 0.1, respectively. They will be respectively decayed to 0.001 and 0.01 after T_1 epochs. We set T_1/T_2 to 60/140 for CAVIAR, and 20/60 for MLR-CUHK03, MLR-DukeMTMC-re-id and MLR-Market1501. Other hyper-parameters are set as follows: the weight factor $\alpha = 1$, $\beta = 0.01$,

Table 1. Comparison with the state-of-the-art models on four datasets (%). Bold and underlined numbers indicate top two results, respectively.

Method	CAVIAR			MLR-CUHK03			MLR-DukeMTMC-reid			MLR-Market1501		
	Rank 1	Rank 5	Rank 10	Rank 1	Rank 5	Rank 10	Rank 1	Rank 5	Rank 10	Rank 1	Rank 5	Rank 10
JUDEA [24]	22.0	60.1	80.8	26.2	58.0	73.4	-	-	-	-	-	-
SLD^{2}L [18]	18.4	44.8	61.2	-	-	-	-	-	-	-	-	-
SDF [38]	14.3	37.5	62.5	22.2	48.0	64.0	-	-	-	-	-	-
SING [17]	33.5	72.7	89.0	67.7	90.7	94.7	65.2	80.1	84.8	74.4	87.8	91.6
CSR-GAN [39]	34.7	72.5	87.4	71.3	92.1	97.4	67.6	81.4	85.1	76.4	88.5	91.9
RAIN [5]	42.0	77.3	89.6	78.9	97.3	98.7	-	-	-	-	-	-
CAD-Net [25]	42.8	76.2	91.5	82.1	97.4	98.8	75.6	86.7	89.6	83.7	92.7	95.8
CamStyle [46]	32.1	72.3	85.9	69.1	89.6	93.9	64.0	78.1	84.4	74.5	88.6	93.0
FD-GAN [10]	33.5	71.4	86.5	73.4	93.8	97.9	67.5	82.0	85.3	79.6	91.6	93.5
PCB [36]	42.1	74.8	88.2	80.6	96.2	98.6	74.5	84.6	90.3	82.6	92.7	95.2
PyrNet [27]	43.6	79.2	90.4	83.9	97.1	98.5	79.6	88.1	91.2	83.8	93.3	95.6
PRI (Ours)	43.2	78.5	91.9	85.2	97.5	98.8	78.3	87.5	91.4	84.9	93.5	96.1
PCB + PRI	44.3	83.7	**94.8**	86.2	**97.9**	**99.1**	81.6	89.6	92.4	**88.1**	**94.2**	**96.5**
PyrNet + PRI	**45.2**	**84.1**	94.6	**86.5**	97.7	**99.1**	**82.1**	**91.1**	**92.8**	86.9	93.8	96.4

the margin of the triplet loss $\phi = 10$, the regulatory factor $\gamma = 1$, the update frequency $\omega = 1$. We train our model on 2 NVIDIA Titan Xp GPUs with the batch size set to 16.

4.2 Comparison with State-of-the-art Models

We compare our PRI model with the state-of-the-art models on four LR re-id datasets, including CAVIAR, MLR-CUHK03, MLR-DukeMTMC-reid and MLR-Market1501 in Table 1. For a fair comparison, we do not use pre-/post-processing methods, *e.g.*, re-ranking [45], even though they can further improve our results.

Comparison with LR re-id models. We compare our PRI model with LR re-id models, including JUDEA [24], SLD^{2}L [18], SDF [38], SING [17], CSR-GAN [39], RAIN [5] and CAD-Net [25]. Compared with the most competitive CAD-Net, PRI achieves 0.4%, 3.1%, 2.7% and 1.2% higher scores at rank 1 on CAVIAR, MLR-CUHK03, MLR-DukeMTMC-reid and MLR-Market1501, respectively. Our advantage lies in adaptively predicting and achieving better recovery that contains more discriminative details instead of noise to help identification. In contrast, CAD-Net recovers details into the fixed resolution, which might not be guaranteed to suit the images of various resolutions best. Apart from CAD-Net, there are also some notable comparisons. SING super-resolves LR probes with multiple scale factors separately and then manually fuses them, while CSR-GAN tries to depend on the image sizes to specify the scale factors. Unlike them, our model can exploit the image content to realize the adaptive scale factor prediction in an end-to-end manner.

Comparison with standard re-id models. We also make a comparison between PRI and the competitive standard re-id models, including CamStyle [46], FD-GAN [10], PCB [36] and PyrNet [27]. For a fair comparison, they are trained on the LR re-id datasets in the same manner as our model. The results of

Table 2. Evaluation of different scale factor predictors on MLR-CUHK03 (%).

Predictor	Rank 1	mAP
Fixed ×1	78.6	79.2
Fixed ×2	82.3	82.9
Fixed ×3	82.1	82.5
Fixed ×4	82.7	82.8
Size-based	82.5	82.8
Ideal	86.8	87.2
Adaptive (Ours)	85.2	85.7

CamStyle and FD-GAN are extracted from [25], and those of PCB and PyrNet are acquired by running the released codes. Among these models, only PyrNet can outperform our model at some ranks, *e.g.*, rank 1 on CAVIAR and MLR-DukeMTMC-reid. However, our method can help the standard re-id models better generalize to the LR re-id setting through a simple combination. We only need to replace our ResNet50 re-id module F with the standard re-id models, and keep the other parts (such as G and P) unchanged. For example, combining PyrNet and our method ("PyrNet + PRI" in Table 1) improves rank 1 by at most 3.1% on MLR-Market1501.

4.3 Ablation Studies

Adaptive scale factor predictor. To verify the effectiveness of our adaptive scale factor predictor P, we compare different methods of predicting scale factors in Table 2. Given a well-trained PRI model on MLR-CUHK03, we replace P with the fixed, size-based and ideal predictors in turn, and compare their performances when test. We set four alternative scale factors, including 1, 2, 3 and 4. The "fixed" predictor chooses a fixed scale factor from the four alternatives for each LR probe. Similar to [39], the "size-based" predictor gives a LR probe a scale factor that is the nearest integer to the ratio of R to r. R is the average size (the product of the image height and width) of all the HR images of the training set, and r is the size of the given LR probe. The "ideal" predictor refers to traversing all the alternative scale factors and choosing the one achieving the best result. To some extent, it represents the upper bound that the given PRI model can reach, if we keep G and F unchanged, and only adjust the scale factor for each LR probe. It is not a surprise that there is a performance gap between the ideal predictor and ours, which means that ours cannot guarantee all the predictions are optimal. But it obviously outperforms the fixed and size-based predictors in terms of rank 1 and mAP (mean Average Precision). Compared with the rough manual "fixed" or "size-based" predictors, our model can adaptively predict and recover the more effective details for each LR image based on its image content, which helps further improve the re-id accuracy.

Soft property. We replace the dynamic soft label of our model with the one-hot label to demonstrate the effectiveness of the soft property. The one-hot label is a binary vector where the scale factor with the maximum evaluated probability

Table 3. Comparison of the soft label with the one-hot label on MLR-CUHK03 (%).

Method	Rank 1	Rank 5	Rank 10
One-hot label	81.8	94.8	97.2
Soft label (Ours)	85.2	97.5	98.8

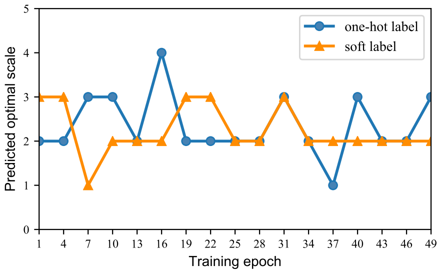

Fig. 3. A comparison between the soft label and one-hot label. This example indicates the change of the predicted optimal scale factor r_p for a LR image during training.

is labeled as 1, while the others are labeled as 0. We evaluate the two types of labels on MLR-CUHK03 in Table 3, which shows that the soft label outperforms the one-hot label by 3.4% at rank 1. The reason is that the variability of the optimal scale factors causes the frequent change of the one-hot label and the unstable training. In fact, we find that using the one-hot label tends to predict a same scale factor (*e.g.*, 2) for most LR probes when test.

We visualize an example of the change of the predicted optimal scale factor r_p during training, as illustrated in Figure 3. We can observe that the soft label stably predicts 2 after 31 epochs while the one-hot label results in a switch between 2 and 3. This indicates that the soft label can smooth the change of the predicted optimal scale factor and stabilize the optimization process.

Dynamic property. We validate the dynamic property of the dynamic soft label by adjusting the update frequency ω. Figure 4 (a) reports the changing curve of rank 1 with ω. Note that ω=0 refers to randomly determining whether to update dynamic soft labels during each training epoch. Compared with setting ω to 1, setting it to 0 slightly degenerates rank 1 on both two datasets. This is probably because randomly updating leads labels not to always timely reflect their variation. When $\omega \geq$1, Figure 4 (a) has a general trend that rank 1 declines as ω increases, showing that more timely updating labels can make more adaptive prediction about the optimal scale factor. Therefore, the dynamic property is effective in handling the variation of the optimal scale factors during training.

Regulatory factor. The regulatory factor γ controls the relative importance of each scale factor in Equation 3. Figure 4 (b) plots rank 1 scores varying with γ. We only set γ to the odd number (except 0) to make sure that the scale factor near to the HR anchor in the feature space has a higher confidence probability. As shown in Figure 4 (b), the rank 1 reaches a peak value when γ is set to 1/3 on

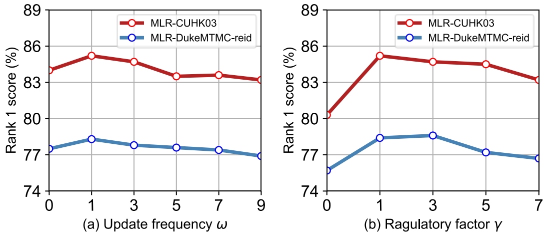

Fig. 4. Effect of the update frequency and regulatory factor.

Table 4. Rank 1 scores of three training manners (%).

Method	MLR-CUHK03	MLR-DukeMTMC-reid
Only pre-training	82.5	73.8
Without pre-training	82.8	75.1
With pre-training (Ours)	85.2	78.3

MLR-CUHK03/MLR-DukeMTMC-reid, respectively. Not surprisingly, setting γ to 0 degrades the performance because this considers all the scale factors as equal, and thus loses the effective supervision of the dynamic soft label.

Pre-training. To validate the effectiveness of pre-training in Section 3.5, we evaluate the rank 1 scores of three training manners, which are in turn: only pre-training (not training the whole model), training without pre-training and training with pre-training. As shown in Table 4, only pre-training achieves the lowest scores on two datasets, and training with pre-training outperforms without pre-training by 2.4% and 3.2% on MLR-CUHK03 and MLR-DukeMTMC-reid, respectively. Pre-training can improve our results because it can endow the SR and re-id module with a basic recovering and identifying ability, so that we can obtain more stable distributions of the features (corresponding to different scale factors) and dynamic soft labels.

Loss functions. We train our model by minimizing the weighted sum of the identity loss L_{id}, triplet loss L_{tri}, SR loss L_{sr} and prediction loss L_p in Equation 6. We validate each loss by removing it from the total loss. Table 5 shows that removing L_{id}, L_{tri} or L_{sr} deteriorates the performance to varying degrees. This is reasonable because L_{id} and L_{tri} are essential to learn the identity discrimination, while L_{sr} is significant to recover effective image contents. Discarding L_p makes rank 1 drop from 85.2% to 82.0%, because the predictor remains initialized and cannot be optimized for the adaptive scale factor prediction without L_p.

Example analysis. We visualize an example of the recovered images, image matching results and predicted probabilities in Figure 5. It can be easily observed that the ground truths achieve the best ranking results and occupy the top 4 places when we super-resolve the probe by the scale factor 2. This is consistent with the fact that 2 has the highest predicted probability (0.33) of being the optimal scale factor. The recovered images could provide an intuitional explanation for the matching results. The relatively complex shirt textures make

Table 5. Evaluation of losses on MLR-CUHK03 (%).

Removed Loss	Rank 1	Rank 5	Rank 10
L_{id}	80.4	94.8	97.5
L_{sr}	81.5	95.4	97.9
L_{tri}	83.4	97.3	98.8
L_p	82.0	94.2	97.2
PRI(with all losses)	85.2	97.5	98.8

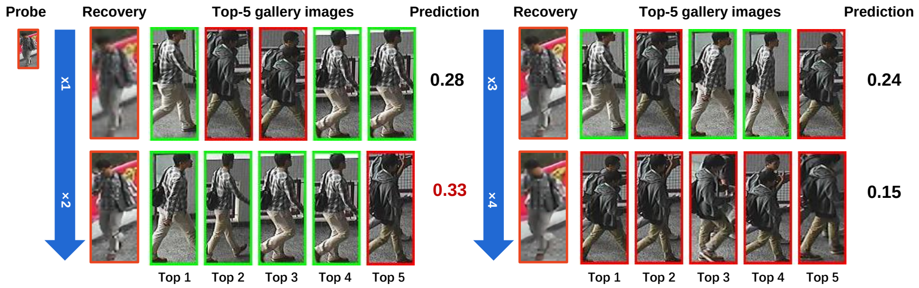

Fig. 5. Visualized examples of the recovered images (resized to the uniform size for a better comparison), top-ranked HR gallery images and predicted probabilities. Each ground truth is indicated by a green bounding box.

the larger scale factor (*e.g.*, 4) tend to produce the distorted or blurry recovery. Therefore, slightly super-resolving the probe with the smaller scale factor 2 is enough for better identification.

5 Conclusions

In this paper, we have proposed the PRI model to explore the potential relation between the image content and the optimal scale factor for LR person re-id. Despite lack of annotations, our proposed dynamic soft label enables us to learn the prediction of the optimal scale factors in a self-supervised manner. Given a LR image, our model can automatically make the content-aware scale factor prediction, and then recover details into the predicted level, and finally identify the recovered person image. Our method can not only predict a preferable scale factor for more effective recovery and identification, but also help the standard re-id models generalize well to the LR re-id setting.

Acknowledgements

This work is jointly supported by National Key Research and Development Program of China (2016YFB1001000), Key Research Program of Frontier Sciences, CAS (ZDBS-LY-JSC032), and National Natural Science Foundation of China (U1803261).

References

1. Agustsson, E., Timofte, R.: Ntire 2017 challenge on single image super-resolution: Dataset and study. In: CVPR (2017)
2. Bai, S., Tang, P., Torr, P.H., Latecki, L.J.: Re-ranking via metric fusion for object retrieval and person re-identification. In: CVPR (2019)
3. Chen, B., Deng, W., Hu, J.: Mixed high-order attention network for person re-identification. In: ICCV (2019)
4. Chen, Y., Zhu, X., Gong, S.: Instance-guided context rendering for cross-domain person re-identification. In: ICCV (2019)
5. Chen, Y.C., Li, Y.J., Du, X., Wang, Y.C.F.: Learning resolution-invariant deep representations for person re-identification. In: AAAI (2019)
6. Cheng, D.S., Cristani, M., Stoppa, M., Bazzani, L., Murino, V.: Custom pictorial structures for re-identification. In: BMVC (2011)
7. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: CVPR (2009)
8. Dong, C., Loy, C.C., He, K., Tang, X.: Learning a deep convolutional network for image super-resolution. In: ECCV (2014)
9. Fu, Y., Wei, Y., Wang, G., Zhou, Y., Shi, H., Huang, T.S.: Self-similarity grouping: A simple unsupervised cross domain adaptation approach for person re-identification. In: ICCV (2019)
10. Ge, Y., Li, Z., Zhao, H., Yin, G., Yi, S., Wang, X., et al.: Fd-gan: Pose-guided feature distilling gan for robust person re-identification. In: NeurIPS (2018)
11. Gray, D., Tao, H.: Viewpoint invariant pedestrian recognition with an ensemble of localized features. In: ECCV (2008)
12. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR (2016)
13. He, L., Liang, J., Li, H., Sun, Z.: Deep spatial feature reconstruction for partial person re-identification: Alignment-free approach. In: CVPR (2018)
14. He, L., Wang, Y., Liu, W., Liao, X., Zhao, H., Sun, Z., Feng, J.: Foreground-aware pyramid reconstruction for alignment-free occluded person re-identification. arXiv preprint arXiv:1904.04975 (2019)
15. Hu, X., Mu, H., Zhang, X., Wang, Z., Tan, T., Sun, J.: Meta-sr: A magnification-arbitrary network for super-resolution. In: CVPR (2019)
16. Huang, Y., Wu, Q., Xu, J., Zhong, Y.: Sbsgan: Suppression of inter-domain background shift for person re-identification. In: ICCV (2019)
17. Jiao, J., Zheng, W.S., Wu, A., Zhu, X., Gong, S.: Deep low-resolution person re-identification. In: AAAI (2018)
18. Jing, X.Y., Zhu, X., Wu, F., You, X., Liu, Q., Yue, D., Hu, R., Xu, B.: Super-resolution person re-identification with semi-coupled low-rank discriminant dictionary learning. In: CVPR (2015)
19. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
20. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: NeurIPS (2012)
21. Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., et al.: Photo-realistic single image super-resolution using a generative adversarial network. In: CVPR (2017)
22. Li, W., Zhao, R., Xiao, T., Wang, X.: Deepreid: Deep filter pairing neural network for person re-identification. In: CVPR (2014)

23. Li, W., Zhu, X., Gong, S.: Harmonious attention network for person re-identification. In: CVPR (2018)
24. Li, X., Zheng, W.S., Wang, X., Xiang, T., Gong, S.: Multi-scale learning for low-resolution person re-identification. In: ICCV (2015)
25. Li, Y.J., Chen, Y.C., Lin, Y.Y., Du, X., Wang, Y.C.F.: Recover and identify: A generative dual model for cross-resolution person re-identification. In: ICCV (2019)
26. Lim, B., Son, S., Kim, H., Nah, S., Mu Lee, K.: Enhanced deep residual networks for single image super-resolution. In: CVPRW (2017)
27. Martinel, N., Luca Foresti, G., Micheloni, C.: Aggregating deep pyramidal representations for person re-identification. In: CVPRW (2019)
28. Miao, J., Wu, Y., Liu, P., Ding, Y., Yang, Y.: Pose-guided feature alignment for occluded person re-identification. In: ICCV (2019)
29. Niu, K., Huang, Y., Ouyang, W., Wang, L.: Improving description-based person re-identification by multi-granularity image-text alignments. TIP (2020)
30. Niu, K., Huang, Y., Wang, L.: Fusing two directions in cross-domain adaption for real life person search by language. In: ICCVW (2019)
31. Si, J., Zhang, H., Li, C.G., Kuen, J., Kong, X., Kot, A.C., Wang, G.: Dual attention matching network for context-aware feature sequence based person re-identification. In: CVPR (2018)
32. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
33. Song, C., Huang, Y., Ouyang, W., Wang, L.: Mask-guided contrastive attention model for person re-identification. In: CVPR (2018)
34. Song, J., Yang, Y., Song, Y.Z., Xiang, T., Hospedales, T.M.: Generalizable person re-identification by domain-invariant mapping network. In: CVPR (2019)
35. Sun, Y., Xu, Q., Li, Y., Zhang, C., Li, Y., Wang, S., Sun, J.: Perceive where to focus: Learning visibility-aware part-level features for partial person re-identification. In: CVPR (2019)
36. Sun, Y., Zheng, L., Yang, Y., Tian, Q., Wang, S.: Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline). In: ECCV (2018)
37. Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Qiao, Y., Change Loy, C.: Esrgan: Enhanced super-resolution generative adversarial networks. In: ECCV (2018)
38. Wang, Z., Hu, R., Yu, Y., Jiang, J., Liang, C., Wang, J.: Scale-adaptive low-resolution person re-identification via learning a discriminating surface. In: IJCAI (2016)
39. Wang, Z., Ye, M., Yang, F., Bai, X., Satoh, S.: Cascaded sr-gan for scale-adaptive low resolution person re-identification. In: IJCAI (2018)
40. Wei, L., Zhang, S., Gao, W., Tian, Q.: Person transfer gan to bridge domain gap for person re-identification. In: CVPR (2018)
41. Yu, T., Li, D., Yang, Y., Hospedales, T.M., Xiang, T.: Robust person re-identification by modelling feature uncertainty. In: ICCV (2019)
42. Zhang, Y., Tian, Y., Kong, Y., Zhong, B., Fu, Y.: Residual dense network for image super-resolution. In: CVPR (2018)
43. Zheng, L., Shen, L., Tian, L., Wang, S., Wang, J., Tian, Q.: Scalable person re-identification: A benchmark. In: ICCV (2015)
44. Zheng, Z., Zheng, L., Yang, Y.: Unlabeled samples generated by gan improve the person re-identification baseline in vitro. In: ICCV (2017)
45. Zhong, Z., Zheng, L., Cao, D., Li, S.: Re-ranking person re-identification with k-reciprocal encoding. In: CVPR (2017)

46. Zhong, Z., Zheng, L., Zheng, Z., Li, S., Yang, Y.: Camera style adaptation for person re-identification. In: CVPR (2018)