Self-Paced Uncertainty Estimation for One-shot Person Re-Identification

Yulin Zhang, Bo Ma, Longyao Liu and Xin Yi

Beijing Laboratory of Intelligent Information Technology, School of Computer Science and Technology, Beijing Institute of Technology, Beijing 100081, China.

{zhangyulin,bma000,roel_liu,yixin}@bit.edu.cn

ABSTRACT

The one-shot Person Re-ID scenario faces two kinds of uncertainties when constructing the prediction model from X to Y. The first is model uncertainty, which captures the noise of the parameters in DNNs due to a lack of training data. The second is data uncertainty, which can be divided into two sub-types: one is image noise, where severe occlusion and the complex background contain irrelevant information about the identity; the other is label noise, where mislabeled affects visual appearance learning. In this paper, to tackle these issues, we propose a novel Self-Paced Uncertainty Estimation Network (SPUE-Net) for one-shot Person Re-ID. By introducing a self-paced sampling strategy, our method can estimate the pseudo-labels of unlabeled samples iteratively to expand the labeled samples gradually and remove model uncertainty without extra supervision. We divide the pseudo-label samples into two subsets to make the use of training samples more reasonable and effective. In addition, we apply a Co-operative learning method of local uncertainty estimation combined with determinacy estimation to achieve better hidden space feature mining and to improve the precision of selected pseudo-labeled samples, which reduces data uncertainty. Extensive comparative evaluation experiments on video-based and image-based datasets show that SPUE-Net has significant advantages over the state-of-the-art methods.

KEYWORDS

One-shot Person Re-Identification, Self-Paced Learning, Uncertainty Learning, Pseudo-label Estimation, Co-operative Learning

1 INTRODUCTION

One-shot Person Re-Identification(Re-ID) aims to reduce the cost of manual data labeling and the dependency of deep learning [12] on labeled data, that is, only one labeled sample for each identity is used for modeling. Due to the extreme lack of training data in the one-shot, the uncertainty of the Re-ID model parameters is called model uncertainty [4, 10, 18]. This uncertainty can be eliminated if the data fed to the network is abundant. Therefore, in this work, we initialize a model using only one labeled sample for each identity and exploit distance-based metrics in the feature space to select pseudo-labeled samples as labeled samples from unlabeled samples by self-paced, which increases the number of labeled samples and removes the model uncertainty.

However, the image inherent noise can cause serious loss of discrimination information, which significantly hinders identity representation learning and reduces the precision of pseudo-label estimation. In particular, most Re-ID methods [24, 31, 34, 42, 46] represent each pedestrian image as a deterministic point embedded in the latent space. For the deterministic representation of the
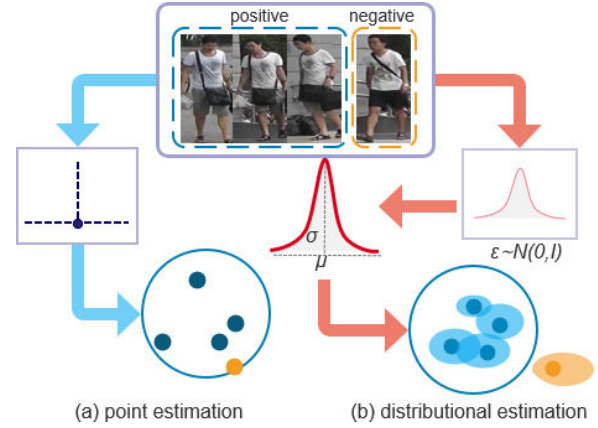
Figure 1: An illustration of the impact of the uncertainty estimation. The blue circle indicates a class label. Solid circles indicate the feature of samples, and solid ellipses indicate the uncertainty of samples.

sample, the sample noise is not estimated, leading to misrecognition and mislabeled pseudo-labels. This is illustrated in Fig 1(a).

DistributionNet [38] is the first work to consider noise estimation in person Re-ID. It describes each sample as a Gaussian distribution, rather than a deterministic point in the latent space. The mean of the distribution estimates the most likely feature values while the variance shows the uncertainty in the feature values. The work states explicitly how uncertainty affects feature learning, that is, samples with larger variances have less impact on the learned feature embedding space. The distributional estimation enhances the features and makes the same class of samples more compact, instead of overfitting samples of different identities, achieving better class separability, as illustrated in Fig 1(b). While being effective, DistributionNet learns the feature and noise in a supervised scenario. However, in the self-paced one-shot scenario, when mislabeled pseudo-label samples dominate the training, the model learns that the feature and variance can have larger shifts than the deterministic estimation, which seriously affects the performance of Re-ID.

For the first time, this work applies a novel Co-operative learning method of local uncertainty estimation combined with determinacy estimation to the self-paced one-shot person Re-ID, which is called Self-Paced Uncertainty Estimation Network (SPUE-Net). This effectively improves the feature affected by various sample noises without increasing the cost of labeling.

Specifically, we split the training samples into four parts, initial labeled samples, high-confidence pseudo-label samples (pseudo-label precision is relatively high), low-confidence pseudo-label samples

(pseudo-label precision is relatively low), unlabeled samples. Uncertainty estimation is performed on the initial labeled samples and high-confidence pseudo-labeled samples. Determinacy estimation is performed on the low-confidence pseudo-label samples and unlabeled samples. We utilize universal index labels [33] for unlabeled samples. We discuss how the uncertainty estimation combined with determinacy estimation affects the model training, from the perspective of image noise. We provide insightful thinking that uncertainty estimation can improve the effect of classification model performance by relatively better quality samples.

Experimental results show that SPUE-Net achieves competitive performance on video-based MARS and DukeMTMC-VideoReID, as well as image-based datasets DukeMTMC-reID and Market-1501.

The main contributions are summarized as follows:

- We propose a Self-Paced Uncertainty Estimation Network (SPUE-Net) to extract inherent identity representations for one-shot Person RE-ID. With the self-paced learning, the Co-operative learning method and subsets division strategy, our network is able to capture and remove the model uncertainty and data uncertainty without extra supervision information.
- We apply a new Co-operative learning method of local uncertainty estimation combined with determinacy estimation to effectively improve the performance of the Re-ID affected by various sample noises without increasing the cost of labeling.
- An effective subsets division strategy is proposed. Different subsets participate in different training methods, which reduces the influence of mislabeled samples on uncertainty estimation and realizes better-hidden space feature mining.
- Experimental results on video-based datasets and image-based datasets show that SPUE-Net has significant advantages over the state-of-the-art methods.

2 RELATED WORK

2.1 Uncertainty in deep learning

Uncertainty learning can improve the performance of deep learning models, which has attracted significant attention from researchers [9, 10, 18, 19]. There are two main types of uncertainty, ie, Epistemic uncertainty and Aleatoric uncertainty. Epistemic uncertainty [18] is uncertainty in the model, capturing the noise of parameters in deep neural networks due to a lack of training data [18]. Aleatoric uncertainty [7, 18] captures uncertainty about the information that the data cannot explain, divided into two sub-categories, Heteroscedastic uncertainty and Homoscedastic uncertainty. Heteroscedastic uncertainty depends on the input data [18]. While homoscedastic uncertainty is Task-dependent, which is a quantity that stays constant for all input data but varies between different tasks [18, 19].

Uncertainty has not been applied by vision application until recent years. The uncertainty method is first applied to modern DNN-based vision tasks in [18]. This method models regression tasks and per-pixel semantic segmentation. In this paper, a novel Bayesian deep learning framework is presented to learn a mapping to data uncertainty from the input samples, which is composed on top of model uncertainty. This network improves the baseline. Besides, object detection [8, 13], semantic segmentation [18, 19, 47], and face recognition [7, 30, 39] tasks introduce uncertainty learning to enhance the performance of the model. In classification tasks, some studies [7, 38] use uncertainty learning for analysis and model each ID image in latent space as a Gaussian distribution.

2.2 Person Re-Identification

Person Re-ID is a cross-view Person retrieval task, which needs to accurately label each pedestrian from multiple camera views. The state-of-the-art person RE-ID algorithms based on deep learning techniques are developed under supervised learning [17, 25, 28, 32, 43, 45, 48]. However, the availability of labeled training samples is limited by the sharply increased cost of manual efforts. This dilemma has prompted people to study one-shot person Re-ID [3, 33, 34, 36, 42] and establish effective algorithms to reduce the cost of labeling. The one-shot is that there is only one labeled sample for each identity. These methods can make better use of unlabeled data without manual intervention.

Most Re-ID methods [17, 25, 28, 31, 35, 37, 43, 45, 46, 49] represent each pedestrian image as a deterministic point embedding in the latent space. The pedestrian features learned by the embedding model could be ambiguous or may not even be present in the input pedestrian, leading to noisy representations. The ideal embedding features can best represent pedestrian identity information and will not be interfered with by information unrelated to pedestrian identity. However, the commonly used datasets for Person Re-ID [25, 34, 44, 45] are acquired in a real scenario. These pedestrian images inevitably carry information that has nothing to do with the identity of pedestrians. As the input data contains lots of noise, Data-dependent uncertainty modeling is essential.

2.3 Self-Paced learning

Curriculum learning [5] and Self-paced learning [15, 22, 23, 40] are originally proposed for solving non-convex optimization problems. Bengio et al. [5] suggest that curriculum learning can be seen as a continuation method. Self-paced learning is developed based on curriculum learning. Kollerteam et al. [22] modeled this idea into a mathematical expression with the theoretical foundation. The core idea of curriculum learning and self-paced learning is to simulate the cognitive mechanism of humans. Both modes process samples in a meaningful order. First, they learn a simple knowledge structure; secondly, the difficulty of learning is continuously increased; finally, more complex and professional knowledge is acquired. We introduce the core idea of self-paced learning into the one-shot Person Re-ID scenario to discover more data information and to train a robust feature classifier for the improvement of the model performance. Our approach also has a similar flavor to active learning [16, 29] and co-training [6, 11]. We select the most confidently pseudo-labeled samples by alternately training classifiers to learn for the next training.

3 SPUE-NET

3.1 Overview

The overview of the proposed method is given in Figure.2. The specific process is as follows: First, we optimize the network by the Co-operative learning method of local uncertainty estimation combined with determinacy estimation on the four kinds of sample splits (the labeled samples, pseudo-labeled Subset-$\mathcal{A}$ samples,
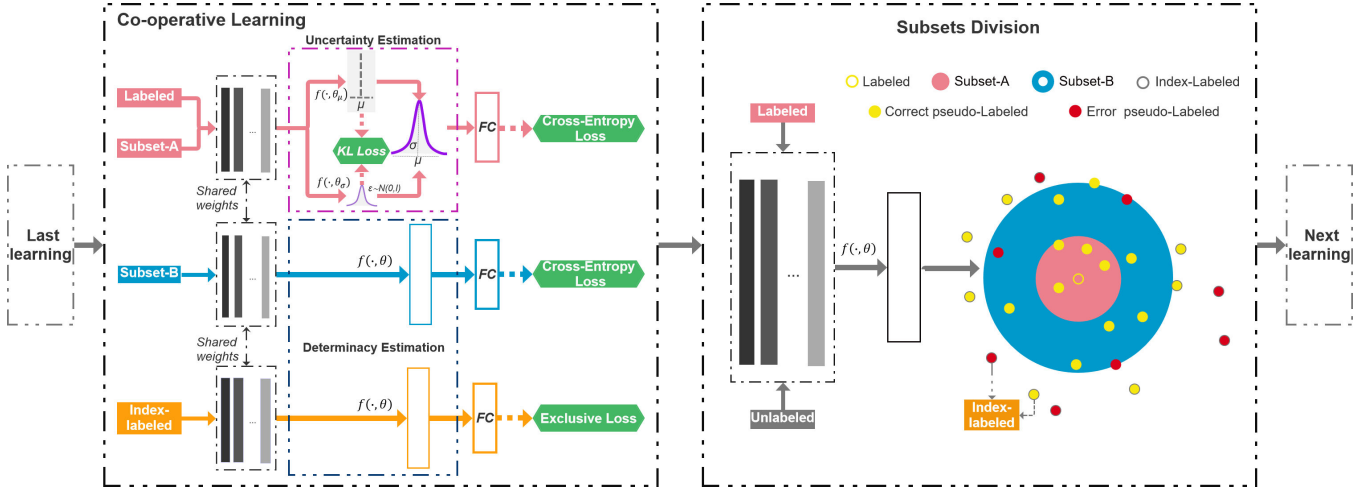
Figure 2: A schematic Overview of SPUE-Net.

pseudo-labeled Subset-$\mathcal{B}$ samples, and index-labeled samples); Second, a few reliable pseudo-labeled samples are selected according to the estimation of pseudo-labels for all unlabeled samples in the feature space. Next, we divide the selected pseudo-labeled samples into Subset-$\mathcal{A}$ and Subset-$\mathcal{B}$ by subsets division strategy based on distance confidence. Specifically, Subset-$\mathcal{A}$ and Subset-$\mathcal{B}$ are empty sets in the first iteration. We continually expand the number of selected pseudo-label samples and unselected unlabeled samples are assigned index labels as index data to joint train the model. We utilize the initial labeled samples and the Subset-$\mathcal{A}$ samples to learn the features and variance of the sample simultaneously by Cross-entropy classification loss and KL loss. We use Cross-entropy classification loss and exclusive loss to constrain Subsets-$\mathcal{B}$ and index data, respectively.

3.2 Problem formulation

For one-shot setting, there is a labeled dataset labeled set $L = \left\{ \left(x_1^L, y_1^L \right), \left(x_2^L, y_2^L \right), \cdots \left(x_n^L, y_n^L \right) \right\}$ with n identities and an unlabeled dataset $U = \left\{ x_i^U \right\}_{i=1}^m$ including m images. These samples are used to train a feature extractor $f(x_i, \theta)$ which parameterized by θ and an identity classifier $\phi(\cdot, \delta)$ which parameterized by δ in the way of identity classification. In the test stage, after the deep model $f(\cdot, \theta)$ is obtained, the similarity is computed according to the Euclidean Distance in the feature space $\|f(x^q, \theta) - f(x^g, \theta)\|$ between the query samples x^q and gallery samples x^g. The result of the query is a ranked list of all gallery data. We estimate the identity label $\hat{y}_i$ of the unlabeled samples x_i^U to expand the labeled sample and select reliable pseudo-label samples for the Re-ID learning. We define $\mathcal{A}^t$, $\mathcal{B}^t$, P^t and I^t as the Subset-$\mathcal{A}$ samples, Subset-$\mathcal{B}$ samples, selected pseudo-label samples and the index-labeled of t-th step iteration, respectively.

3.3 The Co-operative learning method

We first introduce the iterative steps of SPUE-Net. We utilize samples (initial labeled samples L, $\mathcal{A}^t$, $\mathcal{B}^t$) with identity labels, and I^t with index labels for training at the t-th iteration. We utilize L and $\mathcal{A}^t$ to learn an identity classifier with their identity label by

uncertainty estimation. For the $\mathcal{B}^t$, their pseudo-labels are relatively unreliable compared to $\mathcal{A}^t$ and may harm the training of uncertainty estimation in the network. Therefore, we apply determinacy estimation on $\mathcal{B}^t$ and use the exclusive loss [33] on I^t to co-operative optimize the network.

Determinacy estimation The method of deterministic feature embedding on $\mathcal{B}^t$ and index-labeled samples I^t is a continuous mapping space from X to Y. Here X is a continuous image space, and Y is a discrete identifier. Each sample $x_i^B \in B^t$ or $x_i^I \in I^t$ is represented as an embedding z_i in the latent space. For the $\mathcal{B}^t$ samples, we apply an identity classifer $\phi(\cdot, \delta)$ on their deterministic feature embedding $f(\cdot, \theta)$ and optimize the network by comparing the expected probability (predicted identity) with real probability (pseudo-label $\hat{y}_i$). We have the following objective function.

$$\mathcal{L}_{DE} = -\frac{1}{N} \sum_{i=1}^{N} q(\hat{y}_i) \log \frac{\exp\left(\phi\left(f\left(x_i^B, \theta\right), \delta\right)\right)}{\sum_{n=1}^{N} \exp\left(\phi\left(f\left(x_i^B, \theta\right), \delta\right)\right)} \quad (1)$$

where N is the class number, $q(\hat{y}_i)$ represents the indicator variable. Minimizing this cross-entropy loss L_{DE} makes the model output closer to the predicted result.

For index-labeled samples I^t, we utilize the Exclusive loss [33] to learn a discriminative embedding and to extend the distance between each samples I^t in the feature space, as Eq.(2).

$$\mathcal{L}_{EX} = -\log\left(\frac{\exp\|W_{y_i}\|\|x_i^U\|\cos(\theta_{y_i})}{\sum_{j=1}^{M} \exp\|W_j\|\|x_j^U\|\cos(\theta_j)}\right) \quad (2)$$

where $x_i^U \neq x_i^U$ and $x_i^U, x_i^U \in I^t$.

Uncertainty estimation If $x_i \in L$ or $x_i \in \mathcal{A}^t$ has an ideal embedding $f(x_i, \theta)$, which only has identity correlation information. Typically it is inaccurate to represent an identity embedding $f(x_i, \theta)$ by only a single pedestrian sample x_i and more samples need to be collected. The pedestrian samples are acquired in a real scenario, which inevitably carries irrelevant identity information. The extraction of pedestrian ideal features is affected by irrelevant identity information. Therefore, this feature mapping space contains image noise. If we hypothesize that the noise in the image is

additive and obeys a Gaussian distribution, whose mean is zero, and the variance depends on the sample x_i. Therefore, the sample x_i can be represented as $z_i = f(x_i) + n(x_i)$ in the latent space during deep learning, where $n(x_i)$ is the uncertainty information of the sample x_i.

We estimate a distribution $p(z_i \mid x_i)$ in the latent space to represent the potential appearance of sample x_i. The representation z_i in latent space is defined as a Gaussian distribution.

$$p(z_i \mid x_i) = \mathcal{N}\left(z_i, \mu_i, \sigma_i^2 I\right) \tag{3}$$

We establish two independent branches in the penultimate feature extraction layer of the network for generating μ_i and σ_i: $\mu_i = f(x_i, \theta_\mu)$ and $\sigma_i = f(x_i, \theta_\sigma)\}$. To reduce the pedestrian representation complexity, we only use the diagonal covariance matrix. Mean μ_i can be regarded as the most likely pedestrian feature of the input sample, and uncertainty σ_i can be regarded as the confidence of the model along each feature dimension. The representation of each sample z_i is a random embedding sampled from $\mathcal{N}(z_i, \mu_i, \sigma_i^2 I)$ in the latent space. When the random sample z_i is fed to the network, the sampling operation is not differentiable, which prevents the error back-propagation to the preceding layers during the model training. To address this problem, we use a reparameterization trick [20] during sampling to let those layers benefit from the random samples z_i. We first draw a sample ε from the normal distribution $\mathcal{N}(0, I)$ with zero mean and unit covariance, instead of directly sampling from $\mathcal{N}(\mu_i, \sigma_i^2)$. Then we get the sample $\mathcal{R}_i$ by computing Eq.(4), which is equivalent to the sample representation.

$$\mathcal{R}_i = \mu_i + \varepsilon\sigma_i, \quad \varepsilon \sim \mathcal{N}(0, I) \tag{4}$$

In this way, this sampling operation does not need gradient descent and the entire model is trainable. We input the final representation $\mathcal{R}_i$ to the identity classifier $\phi(\cdot, \delta)$ and utilize the cross-entropy loss for classification training, as formulated in Eq.(5)

$$\mathcal{L}_{CE} = -\frac{1}{N} \sum_{i=1}^{N} q(y_i) \log \frac{\exp(\phi(\mathcal{R}_i, \delta))}{\sum_{n=1}^{N} \exp(\phi(\mathcal{R}_i, \delta))} \tag{5}$$

where $q(y_i)$ represents the indicator variable.

During the training stage, since each identity embedding feature μ_i is corroded by uncertainty σ_i, the trivial solution of variance gradually drops to zero. To prevent this problem, we suppress the unstable factors in $\mathcal{R}_i$ so that the Eq.(5) can converge eventually. In this case, the $\mathcal{R}_i$ is re-expressed as $\mathcal{R}_i = \mu_i + c$, which degrades into an original feature representation. Influenced by deep variational information bottleneck (VIB) theory [1, 2, 41], the distribution $\mathcal{N}(\mu_i, \sigma_i^2)$ approaches the normal distribution $\mathcal{N}(0, I)$ during the optimization by adding a regularization term to encourage the model to learn the feature uncertainty. These two distributions are measured by Kullback-Leibler divergence (KLD), as formulated in Eq.(6).

$$\begin{aligned} \mathcal{L}_{KL} &= KL\left[N\left(z_i \mid \mu_i, \sigma_i^2\right) \| N(\varepsilon \mid 0, I)\right] \\ &= \frac{1}{2} \sum_{i=1}^{d}\left(\mu_i + \sigma_i^2 - \log \sigma_i^2 - 1\right) \end{aligned} \tag{6}$$

where d is the dimension of the hidden variable z_i. The loss $\mathcal{L}_{KL}$ balance the classification loss of Eq.(5). To acquire the outstanding

features embedding by the uncertainty estimation branch, we employ the joint training of softmax loss and KL loss to optimize the model. Therefore, the cost function of uncertainty estimation Eq.(7) is as follows:

$$\mathcal{L}_{UE} = \mathcal{L}_{CE} + \lambda \mathcal{L}_{KL} \tag{7}$$

where the hyper-parameter λ is used for balancing the Cross-Entropy loss and KL loss functions.

The SPUE-Net objective function Considering the four data splits, in the $(t+1)$-th iteration of the model, we formulate the final objective function as Eq.(8):

$$\begin{aligned} \mathcal{L}_{SPUE} = {} & \gamma \mathcal{L}_{UE}^{L} + \\ & s^t \gamma \left[\alpha^t \mathcal{L}_{UE}^{\mathcal{A}} + (1 - \alpha^t)\mathcal{L}_{DE}\right] + \\ & (1 - s^t)(1 - \gamma)\mathcal{L}_{EX} \end{aligned} \tag{8}$$

where $\alpha^t \in \{0, 1\}$ is the selection indicator for the Subset-$\mathcal{A}$ and $s^t \in \{0, 1\}$ is the selection indicator for the unlabeled sample. The first part is the uncertainty estimation loss $\mathcal{L}_{UE}^{L}$ on the labeled sample L. The second part includes the uncertainty loss $\mathcal{L}_{UE}^{\mathcal{A}}$ on the divided Subset-$\mathcal{A}$ samples $\mathcal{A}^{t+1}$. The third one is the determinacy estimation loss $\mathcal{L}_{DE}$ on the divided Subset-$\mathcal{B}$ subset $\mathcal{B}^{t+1}$. The last one is the exclusive loss $\mathcal{L}_{EX}$ on index-labeled data I^{t+1}. The hyper-parameter γ is set to adjust the proportion of the loss function between the index label sample and the identity label sample. The model does not complete the training until all unlabeled samples are selected into the pseudo-labeled dataset.

3.4 The effective subsets division

Selected pseudo-label samples Subset division is performed on the selected pseudo-label samples, and the precision of the pseudo-label samples is crucial to the subset division. Previous works feed unlabeled samples into the classification model to estimate pseudo labels. The overfitting phenomenon of the classification model cannot accurately predict the class of each unlabeled sample. To address this, we employ the Nearest Neighbors (NN) classifier to estimate the label of unlabeled data. The Euclidean distance is utilized between unlabeled samples x_i^U and initial labeled samples x_i^L in the feature space as the confidence of label estimation.

$$d\left(x_i^U, \theta\right) = \arg\min \left\| f\left(x_i^U, \theta\right) - f\left(x_i^L, \theta\right) \right\| \tag{9}$$

where $d\left(x_i^U, \theta\right)$ denotes the result of the label estimation. We select partial unlabeled samples of top confidence as pseudo-label samples P^t and set a step size called the Expansion Rate (ER) to determine the number of pseudo labels selected in each iteration. In the t-th step iteration, the highest reliable pseudo-labeled samples of $ER \times t$ are selected according to Eq.(10).

$$P^t = \arg\min_{ER \times t} \sum_{i=n+1}^{n+m} p_i d\left(x_i^U, \theta\right) \tag{10}$$

where P^t is the vertical concatenation of all p_i.

Subsets division Due to the lack of initial labeled samples L, the model has an over-fitting phenomenon. That is, the selected pseudo-label samples P^t by Eq.(10) and Eq.(10), there are still a large

number of mislabeled samples, and the precision of the pseudo-label samples gradually decreases as the confidence decreases. This study is also described in [42].

To reduce the impact of pseudo-labeled noise on data uncertainty estimation, we divide the selected pseudo-label samples P^t into two subsets: Subset-$\mathcal{A}$ and Subset-$\mathcal{B}$.

$$\mathcal{A}^t = \arg\min_{\alpha \times ER \times t} \sum_{i=n+1}^{n+m} p_i d\left(x_i^U, \theta\right) \tag{11}$$

$$\mathcal{B}^t = \arg\max_{(1-\alpha) \times ER \times t} \left[\arg\min_{ER \times t} \sum_{i=n+1}^{n+m} p_i d\left(x_i^U, \theta\right)\right] \tag{12}$$

where α is a hyperparameter to adjust the proportion between Subset-$\mathcal{A}$ and Subset-$\mathcal{B}$. The selection of hyperparameter α values depends on the sample quality. The smaller the hyperparameter α, the smaller the weight of the uncertainty estimation. However, the quality of samples in Subset-$\mathcal{A}$ also decreases simultaneously, and the negative impact of uncertainty gradually increases. Through multiple experiments, we selected 0.3 for the hyperparameter, which has a relatively good influence on the model. See the experiment 4.5 for details.

4 EXPERIMENTS

4.1 Experimental Datasets

We conduct experiments to evaluate the proposed Self-Paced Uncertainty Network on image-based datasets (Market1501, DukeMTMC-reID) and video-based datasets (MARS, DukeMTMC-VideoReID).

Market1501 [45] is an image dataset covering 32,668 labeled images of 1,501 identities collected from 6 camera views. Thereinto 12,936 images from 751 identities are used for training and 19,732 images from 750 identities for testing.

DukeMTMC-reID contains 36411 labeled images of 1,812 identities from 8 camera views. We split the dataset into two non-overlapping parts according to the setup in [27]: 16,522 images from 702 identities for training and 19989 images from other 702 identities for testing.

MARS [34] has 17,503 tracklets captured by 6 cameras, containing 1,261 identities and 3,248 distractor tracklets, of which 625 identities for training and 636 identities for testing.

DukeMTMC-VideoReID [44] is a video-based dataset. It includes 2,196 videos of 702 identities for training and 2,636 videos of 702 identities for testing. 408 identities are used as distractors. Each video contains images sampled every 12 frames.

4.2 Experimental Setups

4.2.1 Performance Metrics. We use two evaluation metrics: mean Average Precision (mAP) and the Cumulative Matching Characteristics (CMC) curve to evaluate the proposed effectiveness of model. As an object retrieval problem, the mAP represents the recall rate, and it is the mean value of the average accuracy of all queries. And the CMC score reflects retrieval accuracy. We report the Rank-1, Rank-5, Rank-10, Rank-20 scores to represent the CMC curve.

4.2.2 Dataset setting. We apply the [26] protocol directly to one-shot Person RE-ID. In image-based and video-based datasets, we randomly select an image/tracklet under camera 1 for each identity

pedestrian as the labeled data. If an identity has no sample under camera 1, then it is postponed to camera 2 to select the sample, and so on. Make sure that each identity has a labeled example.

4.2.3 Implementation Details. To implement our algorithm, we adopt ResNet-50 pre-trained on ImageNet [21] with the last classification layer removed as our basic feature embedding model for feature extraction. To make better use of unlabeled data, we use the fully connected layer and a classified layer with batch normalization [14] at the end of the feature embedding model. The hyper-parameter λ is set to 0.01 to balance the Cross-Entropy loss and KL loss functions. The hyper-parameter γ is set to 0.8 and the hyper-parameter α is set to 0.3 to control the final loss of model training. For the optimization model, we employ Stochastic Gradient Descent (SGD) with a weight decay of 0.0005 and a momentum of 0.5 to optimize the loss function parameters. The batch size is set to 16 at Marker1501 and DukeMTMC-reID image datasets and is set to 8 at MARS and DukeMTMC-VideoReID video datasets. There are 70 epochs in each iterative. The initial learning rate is set to 0.1 and the learning rate is reduced to 0.01 when epoch> 55. Our model is trained end-to-end on a single Nvidia GeForce GTX 1080Ti with 12 GB memory, and the total time is about 10 hours when the expansion rate is 10% at the Marker1501 dataset.

4.3 Comparison to the state-of-the-art

We compare our proposed SPUE-Net with the state-of-the-art, and the initial model uses the same one-shot sample of each identity. Besides, we compare our proposed method with DGM and Step-wiseon video-based datasets. The specific experiment design is consistent with PL[33]. To make it fair, our labeled data Expansion Rates ER is set to the same 5%, 10%, 15%, 20%, and 30% as the PL. First of all, Table 1 shows that the rank-1 accuracy of our proposed method has exceeded our initial model 32.1, 24.9, 24.9, and19.9 points on the Market-1501, DukeMTMC-reID, MARS, and DukeMTMC-VideoReID dataset, respectively.

The SPUE-Net increases rank-1 accuracy by 3.0, 3.6, 6.2, 6.0, and 4.8 points at the different expansion rates on the Market-1501 dataset. Simultaneously, rank-1 accuracy is increased by 4.7, 8.1, 4.5, 5.1, and -2.5 points on the DukeMTMC-reID. We notice that the performance of the model has a certain decrease relative to the baseline PL[33] when ER=5% on the DukeMTMC-reID dataset. We provide insightful analysis of the causes of experimental performance degradation. On the video-based datasets, our proposed method increases rank-1 accuracy by 0.2, 1.5, 1.8, 3.0, and 3.9 points at the different expansion rates on the MARS dataset compared with PL[33]. Our proposed method increases rank-1 accuracy by 5.0, 1.2, 2.8, 0.3, and 0.5 points at the different expansion rates on the DukeMTMC-VideoReID dataset compared with EUG[34]. The performance of SPUE-Net is significantly improved in comparison with PL[33] and EUG[34].

We analyze the characteristics of samples in the DukeMTMC-reID dataset to find the reasons for this phenomenon. We found there are considerable differences among the in-class samples. Researchers collect samples in the winter from 8 camera views. The same identity pedestrian wears a coat and hat under the current camera and takes off his coat when he came to another camera. Samples like this are prevalent in the DukeMTMC-reID dataset.

Table 1: Comparison with the state-of-the-art methods on Market-1501 and DukeMTMC-reID. All comparison methods are based on the same backbone model ResNet50. The baseline(initial) is the initial model trained on our proposed method. ER is the expansion rate that the expansion speed of sample label estimation. The best results are in black boldface font.

methods		Market-1501					DukeMTMC-reID				
		mAP	Rank-1	Rank-5	Rank-10	Rank-20	mAP	Rank-1	Rank-5	Rank-10	Rank-20
Baseline(initial)		10.2	28.3	45.5	52.3	60.1	8.9	21.4	33.4	39.5	45.1
ER=30%	PL[33]	13.4	35.5	52.8	60.5	68.6	11.1	23.3	35.7	42.2	48.0
	Ours	**16.2**	**38.5**	**56.4**	**63.8**	**72.7**	**13.8**	**28.0**	**42.2**	**48.4**	**54.6**
ER=20%	PL[33]	17.4	41.4	59.6	66.4	73.5	15.1	30.0	43.4	49.2	54.8
	Ours	**19.6**	**45.0**	**61.1**	**68.4**	**74.9**	**20.3**	**38.1**	**52.2**	**57.6**	**63.0**
ER=15%	PL[33]	19.2	44.8	61.8	69.1	76.1	18.2	35.1	49.1	54.3	60.0
	Ours	**23.6**	**51.0**	**66.8**	**73.2**	**79.2**	**22.3**	**40.6**	**54.6**	**60.2**	**65.4**
ER=10%	PL[33]	23.2	51.5	66.8	73.6	79.6	21.8	40.5	53.9	60.2	65.5
	Ours	**28.3**	**57.5**	**73.4**	**79.6**	**84.6**	**25.0**	**45.6**	**59.7**	**65.0**	**69.5**
ER=5%	PL[33]	26.2	55.8	72.3	78.4	83.5	**28.5**	**48.8**	**63.4**	**68.4**	**73.1**
	Ours	**30.5**	**60.4**	**76.0**	**81.1**	**85.7**	27.6	46.3	61.0	67.5	71.0

Table 2: Tables 1 Comparison with the state-of-the-art methods on MARS and DukeMTMC-VideoReID. All comparison methods are based on the same backbone model ResNet50. ER is the expansion rate that the expansion speed of sample label estimation. The best results are in black boldface font.

methods		MARS					DukeMTMC-VideoReID				
		mAP	Rank-1	Rank-5	Rank-10	Rank-20	mAP	Rank-1	Rank-5	Rank-10	Rank-20
Baseline(initial)		16.8	37.6	52.5	57.6	63.8	45.8	53.4	70.4	74.8	80.3
DGM+IDE[37]		16.9	36.8	54.0	59.6	68.5	33.6	42.4	57.9	63.8	69.3
Stepwise[26]		19.7	41.2	55.6	62.2	66.8	46.8	56.3	70.4	74.6	79.2
ER=30%	EUG[34]	21.1	42.8	56.5	-	67.2	54.6	63.8	78.6	-	87.0
	PL[33]	22.1	44.5	58.7	65.7	70.6	59.3	66.1	79.8	84.9	88.3
	Ours	**23.9**	**44.7**	**58.8**	64.7	69.4	**60.7**	**68.8**	**82.2**	**86.5**	**90.3**
ER=20%	EUG [34]	26.6	48.7	63.4	-	72.6	59.5	69.0	81.0	-	89.5
	PL[33]	27.2	49.6	**64.5**	69.8	74.4	59.6	69.1	81.2	85.6	89.6
	Ours	**32.1**	**51.1**	64.4	**70.4**	**75.6**	**62.2**	**70.2**	**84.8**	**89.0**	**92.3**
ER=15%	EUG[34]	29.6	52.3	64.2	-	73.1	59.2	69.1	81.2	-	88.9
	PL[33]	29.9	52.7	66.3	71.9	76.4	59.5	69.3	81.4	85.9	89.2
	Ours	**34.3**	**54.5**	**66.8**	**72.2**	**77.2**	**63.8**	**71.9**	**84.8**	**87.3**	**91.0**
ER=10%	EUG[34]	34.7	57.6	69.6	-	78.1	61.8	70.8	83.6	-	89.6
	PL[33]	34.9	57.9	70.3	75.2	79.3	61.9	71.0	83.8	87.4	90.3
	Ours	**43.0**	**60.9**	**75.8**	**80.3**	**84.8**	**64.4**	**71.1**	**86.5**	**89.3**	**92.5**
ER=5%	EUG[34]	42.5	62.7	74.9	-	82.6	63.2	72.8	84.2	-	91.5
	PL[33]	42.6	62.8	75.2	80.4	83.8	63.3	72.9	84.3	88.3	91.4
	Ours	**46.3**	**66.7**	**78.1**	**81.6**	**85.8**	**65.0**	**73.3**	**84.8**	**89.0**	**91.6**

The method of uncertainty estimation to train the network in one-shot scenarios with small differences between classes and large inter-class differences has significant limitations, which leads to a terrible beginning. The ideal experimental effect cannot be achieved in the later iterative training. According to the analysis above, we suggest that the network set a larger Expansion Rate to avoid the trap of local uncertainty estimation in the case of large differences between sample classes. In the case of high-similarity samples, the network needs to set a smaller Expansion Rate in order to obtain better performance.

The performance of our proposed method is shown in Table 1 and Table 2. Experimental results show that our method achieves satisfactory performance, using only one labeled sample of each identity on four large-scale datasets.
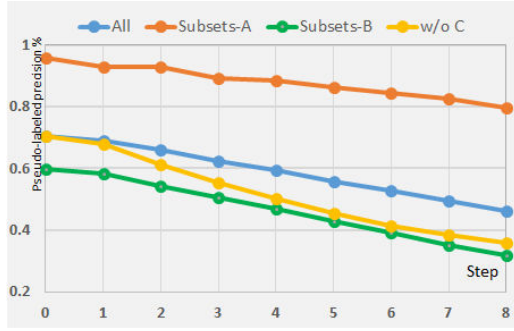
Figure 3: The precision of selected pseudo-labeled samples during iterations.
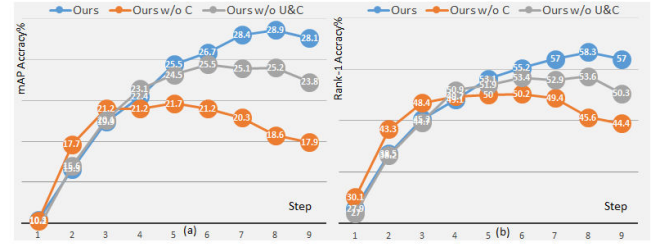

Figure 4: Ablation studies on Market-1501. (a) and (b) are mAP and Rank-1 on the evaluation set during iterations. The expansion rate is ER=10%. The x-axis stands for the iterative steps.

Table 3: Ablation study compares two criteria on Market-1501 of image-based and MARS of video-based datasets.

setting	Method	Market-1501		MARS	
		mAP	Rank-1	mAP	Rank-1
ER=30%	Ours w/o C&U	13.4	35.5	22.1	44.5
	Ours w/o C	10.8	32.9	20.0	39.5
	Ours	16.2	38.5	23.9	44.7
ER=20%	Ours w/o C&U	17.4	41.4	27.2	49.6
	Ours w/o C	15.1	38.2	24.9	45.6
	Ours	19.6	45.0	32.1	51.1
ER=10%	Ours w/o C&U	23.2	51.5	34.9	57.9
	Ours w/o C	18.6	45.2	33.4	52.1
	Ours	28.3	57.5	43.0	60.9
ER=5%	Ours w/o C&U	26.2	55.8	42.6	62.8
	Ours w/o C	20.2	51.4	38.5	58.8
	Ours	30.5	60.4	46.3	66.7

4.4 Ablation study

In this subsection, in order to show the effectiveness of the SPUE-Net, we give an ablation study to evaluate different variants of the proposed method for Market-1501 of image-based and MARS of video-based datasets are employed. We conduct ablation studies on two crucial parts of SPUE-Net, ie, uncertainty estimation and the Co-operative learning method. as shown in Table 3 and Figure 4. We set the same initial labeled image and training settings in all experiments. "Ours w/o C" indicates that we only use the uncertainty estimation without the Co-operative learning method. "Ours w/o C&U" is consistent with the experimental settings of the baseline[33] without uncertainty estimation and the Co-operative learning method.

4.4.1 The limitations of uncertainty estimation. We compare our method to the model trained without the Co-operative learning method, denoted as "Ours w/o C" in Table 3. The network only uses the uncertainty estimation to have a certain level of decline relative to the baseline [33] under different ER settings. However, we can observe that the performance of the model "Ours w/o C" is the best in the early iterations(from 1 to 3), proving the validity of

the uncertainty estimation in Figure 4. Since the network only selects a few of the most reliable unlabeled samples as pseudo-labeled samples, the accuracy score of label estimation is relatively high at the beginning. As the number of iterations increases, the number of pseudo-label samples selected by the network increases, while the accuracy of pseudo-label samples continues to decrease, as shown in the blue curve in Figure 3. The blue curve is the change of the precision of all pseudo-label samples in each step when the ER is set to 10 in the market-1501 dataset. After the middle of the iteration, the Re-id evaluation performance dropped significantly from the baseline[33]. From the perspective of experimental research, uncertainty estimation improves robustness by assigning large variances to mislabeled samples, reducing the influence of mislabeled samples on model training, which obtain a better embedding space and different identities become more separable. However, when the number of mislabeled samples involved in uncertainty estimation is relatively large, common features of inner-class samples are corroded. The uncertainty estimated by the model will have a negative impact on the network.

4.4.2 The effectiveness of the Co-operative learning method. The Re-ID performances of sampling by uncertainty estimation and the Co-operative learning method are illustrated in Figure 4 and Table 3. For training that only uses uncertainty estimation, there are a large number of mislabeled samples in the pseudo-labeled samples predicted by the uncertainty estimation at the later stage of the iteration from Figure 3, which affects the uncertainty model's correct estimation of sample noise, resulting in a decrease in model performance. The SPUE-Net adopts the Co-operative learning method of local uncertainty estimation and determinacy estimation. Through the subsets division based on the distance confidence, the uncertainty estimation and the determinacy estimation have a positive incentive effect, resulting in the model that can alleviate the corrosion of inherent features, achieve further feature mining and improve the performance of the model.

4.5 Algorithm analysis

4.5.1 Expansion Rate and Iterations Analysis. There are two crucial parameters in the self-paced sampling strategy, ie, Expansion Rate (ER) and Iterations, which determine the training speed and performance of the model. ER represents the expansion speed of the pseudo-labeled dataset during iterations. The change of the ER
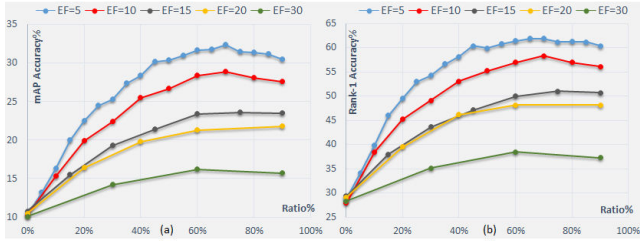
Figure 5: Comparison with different values of expansion rate on Market-150. "ER" denotes the expansion rate. Fig (a) and fig (b) are mAP and Rank-1 accuracy on the evaluation set of the different expansion rates during iterations. Each solid point indicates an iteration step. The x-axis stands for the ratio of the selected sample from unlabeled data.
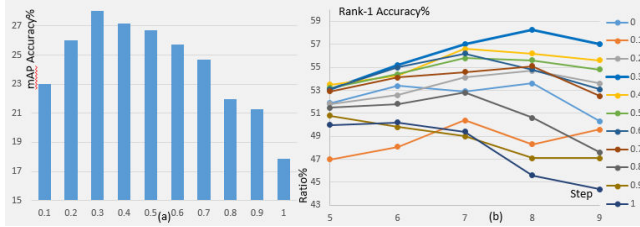
Figure 6: Comparison with different values of hyper-parameter α on Market-150. Fig (a) is mAP accuracy on the evaluation set of the different hyper-parameter α in the last iteration, and Fig (b) is Rank-1 accuracy on the evaluation set of the different hyper-parameter α during iterations.

affects the number of iterations, and which is in an inverse proportion. The results of different ER on Market-1501 are shown in Figure 5. Figure 5(a) and Figure 5(b) are mAP and Rank-1 scores on the evaluation set with the different expansion rates, respectively. The experimental scores of EF=5% are the best. The experimental results show that a smaller ER yields a better performance of the model. We observe from Figure 5 that as the number of iterations increases, the gaps among the five curves continue to increase, which indicates that the error in pseudo-labeled estimation increases.

Figure 5 shows the experimental results of each iteration. Throughout iterations, the precision of the mAP and Rank-1 scores on Market-1501 continues to increase, which validates the stability of our model. However, the evaluation performance stops to increase in the last few iterations. The gain of adding new pseudo-labeled samples is consumed by the larger number of mislabeled samples. The mislabeled samples in the inchoate iterations of the self-paced sampling will have the most severe impact on the later iterative process. The uncertainty estimated by the model will harm the network. If the space covered by labeled samples is inconsistent with the real data space, the self-paced training classification algorithm will be completely ineffective, and the unlabeled samples are not conducive to learn a better classifier.

4.5.2 Subsets division Analysis. We compare the different values of hyperparameter α, and it adjusts the proportion of Subset-$\mathcal{A}$ and Subset-$\mathcal{B}$ in the pseudo-label dataset. Figure 6 shows the effect of different hyperparameters α on the performance of RE-ID model. To visualize the experimental results more clearly, we show the final

step of Rank-1 (iteration 5 to 9) in Figure 6 (b). As we can see, $\alpha = 0.3$ performs best in the experimental evaluation of the mAP and Rank-1. In this way, Subset-$\mathcal{A}$ selects samples with confidence in the first 30% to learn the feature and uncertainty, and Subset-$\mathcal{B}$ chooses the remaining 70% samples to learn deterministic feature embedding. The uncertainty estimation method pursues a balanced relationship between the number of training samples and the performance of the re-id model. Co-operative learning of labeled samples, pseudo-labeled samples(Subset-$\mathcal{A}$ and Subset-$\mathcal{B}$), and unlabeled samples make the experimental result the best performance.

5 UNCERTAINTY THINKING

The difference between the uncertainty estimation and the determinacy estimation is that the sample feature and noise are estimated simultaneously. The estimated uncertainty is closely related to the quality of pedestrian images, that is, the uncertainty of learning increases as the image quality decreases. The uncertainty estimation of samples can improve the learning of identity features by adaptively reducing the adverse effects of noisy training samples, which makes the high-quality (low-noise) samples more compact and low-quality (high-noise) samples farther away. Uncertainty estimation reduces the interference of low-quality samples to the model and increases the training weight of high-quality samples. Accurately estimated noise is the important core of uncertainty estimation.

What is sample noise? The model has no labels for noise and features and estimates that the sample noise comes from a random sampling of hidden space features. Under the supervision of the identity labels, the model gradually learns the common features μ and difference features σ of the same class samples. Therefore, when the mislabeled samples account for most of the training samples, the features and noise learned by the model will have a greater deviation. It is even more serious than the deviation of the deterministic representation, which has been verified in the ablation study4.4. When high-quality samples dominate the training samples, the acquired features and noise are close to real values, thereby reducing the adverse effects of mislabeled samples on the model. When the number of samples is small and the differences between samples are large, the features and noise learned by the model tend to be unstable. This is also an important reason for the decline in the first iteration of the DukeMTMC-reID training process. [see details in Section4.3]. In general, it is not appropriate to use uncertainty estimation in samples with poor data quality. Uncertainty estimation can improve the effect of model performance with relatively better quality samples.

6 CONCLUSION

In this paper, we propose a Self-Paced Uncertainty Estimation Network (SPUE-Net) for one-shot person Re-ID. With the self-paced learning, subsets division strategy, and the Co-operative learning method of local uncertainty estimation and deterministic estimation, our method is able to capture and remove the model uncertainty and data uncertainty without extra supervision information. The performance improvement demonstrates that our proposed SPUE-Net performs better than deterministic models on most benchmarks.

REFERENCES

[1] Alexander A Alemi, Ian Fischer, and Joshua V Dillon. 2018. Uncertainty in the variational information bottleneck. *arXiv preprint arXiv:1807.00906* (2018).

[2] Alexander A Alemi, Ian Fischer, Joshua V Dillon, and Kevin Murphy. 2016. Deep variational information bottleneck. *arXiv preprint arXiv:1612.00410* (2016).

[3] Slawomir Bak and Peter Carr. 2017. One-shot metric learning for person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2990–2999.

[4] Gabriel Bender, Pieter-Jan Kindermans, Barret Zoph, Vijay Vasudevan, and Quoc Le. 2018. Understanding and simplifying one-shot architecture search. In *International Conference on Machine Learning*. 549–558.

[5] Yoshua Bengio, Jerome Louradour, Ronan Collobert, and Jason Weston. 2009. Curriculum learning. (2009), 41–48.

[6] Avrim Blum and Tom Mitchell. 1998. Combining labeled and unlabeled data with co-training. In *Proceedings of the eleventh annual conference on Computational learning theory*. 92–100.

[7] Jie Chang, Zhonghao Lan, Changmao Cheng, and Yichen Wei. 2020. Data Uncertainty Learning in Face Recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 5710–5719.

[8] Jiwoong Choi, Dayoung Chun, Hyun Kim, and Hyuk-Jae Lee. 2019. Gaussian yolov3: An accurate and fast object detector using localization uncertainty for autonomous driving. In *Proceedings of the IEEE International Conference on Computer Vision*. 502–511.

[9] Stefan Depeweg, Jose-Miguel Hernandez-Lobato, Finale Doshi-Velez, and Steffen Udluft. 2018. Decomposition of uncertainty in Bayesian deep learning for efficient and risk-sensitive learning. In *International Conference on Machine Learning*. PMLR, 1184–1193.

[10] Yarin Gal. 2016. Uncertainty in deep learning. *University of Cambridge* 1, 3 (2016).

[11] Tengda Han, Weidi Xie, and Andrew Zisserman. 2020. Self-supervised Co-training for Video Representation Learning. *Advances in Neural Information Processing Systems* 33 (2020).

[12] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.

[13] Yihui He, Chenchen Zhu, Jianren Wang, Marios Savvides, and Xiangyu Zhang. 2019. Bounding box regression with uncertainty for accurate object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2888–2897.

[14] Sergey Ioffe and Christian Szegedy. 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167* (2015).

[15] Lu Jiang, Deyu Meng, Shoou-I Yu, Zhenzhong Lan, Shiguang Shan, and Alexander G Hauptmann. 2014. Self-Paced Learning with Diversity. (2014), 2078–2086.

[16] Roger T Johnson and David W Johnson. 2008. Active learning: Cooperation in the classroom. *The annual report of educational psychology in Japan* 47 (2008), 29–30.

[17] Mahdi M Kalayeh, Emrah Basaran, Muhittin Gökmen, Mustafa E Kamasak, and Mubarak Shah. 2018. Human semantic parsing for person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 1062–1071.

[18] Alex Kendall and Yarin Gal. 2017. What uncertainties do we need in bayesian deep learning for computer vision?. In *Advances in neural information processing systems*. 5574–5584.

[19] Alex Kendall, Yarin Gal, and Roberto Cipolla. 2018. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 7482–7491.

[20] Diederik P Kingma and Max Welling. 2013. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114* (2013).

[21] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*. 1097–1105.

[22] M P Kumar, Benjamin Packer, and Daphne Koller. 2010. Self-Paced Learning for Latent Variable Models. (2010), 1189–1197.

[23] Yong Jae Lee and Kristen Grauman. 2011. Learning the easy things first: Self-paced visual category discovery. (2011), 1721–1728.

[24] Huafeng Li, Jiajia Xu, Jinting Zhu, Dapeng Tao, and Zhengtao Yu. 2019. Top distance regularized projection and dictionary learning for person re-identification. *Information Sciences* (2019).

[25] Yutian Lin, Liang Zheng, Zhedong Zheng, Yu Wu, Zhilan Hu, Chenggang Yan, and Yi Yang. 2019. Improving person re-identification by attribute and identity learning. *Pattern Recognition* (2019).

[26] Zimo Liu, Dong Wang, and Huchuan Lu. 2017. Stepwise metric promotion for unsupervised video person re-identification. In *Proceedings of the IEEE International Conference on Computer Vision*. 2429–2438.

[27] Ergys Ristani, Francesco Solera, Roger Zou, Rita Cucchiara, and Carlo Tomasi. 2016. Performance measures and a data set for multi-target, multi-camera tracking. In *European Conference on Computer Vision*. Springer, 17–35.

[28] Ergys Ristani and Carlo Tomasi. 2018. Features for multi-target multi-camera tracking and re-identification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 6036–6046.

[29] Burr Settles. 2009. *Active learning literature survey*. Technical Report. University of Wisconsin-Madison Department of Computer Sciences.

[30] Yichun Shi and Anil K Jain. 2019. Probabilistic face embeddings. In *Proceedings of the IEEE International Conference on Computer Vision*. 6902–6911.

[31] Ancong Wu, Wei-Shi Zheng, Xiaowei Guo, and Jian-Huang Lai. 2019. Distilled Person Re-identification: Towards a More Scalable System. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 1187–1196.

[32] Lin Wu, Yang Wang, Junbin Gao, and Xue Li. 2018. Where-and-when to look: Deep siamese attention networks for video-based person re-identification. *IEEE Transactions on Multimedia* 21, 6 (2018), 1412–1424.

[33] Yu Wu, Yutian Lin, Xuanyi Dong, Yan Yan, Wei Bian, and Yi Yang. 2019. Progressive learning for person re-identification with one example. *IEEE Transactions on Image Processing* 28, 6 (2019), 2872–2881.

[34] Yu Wu, Yutian Lin, Xuanyi Dong, Yan Yan, Wanli Ouyang, and Yi Yang. 2018. Exploit the unknown gradually: One-shot video-based person re-identification by stepwise learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 5177–5186.

[35] Jing Xu, Rui Zhao, Feng Zhu, Huaming Wang, and Wanli Ouyang. 2018. Attention-aware compositional network for person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2119–2128.

[36] Xun Yang, Meng Wang, Richang Hong, Qi Tian, and Yong Rui. 2017. Enhancing person re-identification in a self-trained subspace. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 13, 3 (2017), 27.

[37] Mang Ye, Andy J Ma, Liang Zheng, Jiawei Li, and Pong C Yuen. 2017. Dynamic label graph matching for unsupervised video re-identification. In *Proceedings of the IEEE International Conference on Computer Vision*. 5142–5150.

[38] Tianyuan Yu, Da Li, Yongxin Yang, Timothy M Hospedales, and Tao Xiang. 2019. Robust person re-identification by modelling feature uncertainty. In *Proceedings of the IEEE International Conference on Computer Vision*. 552–561.

[39] Umara Zafar, Mubeen Ghafoor, Tehseen Zia, Ghufran Ahmed, Ahsan Latif, Kaleem Razzaq Malik, and Abdullahi Mohamud Sharif. 2019. Face recognition with Bayesian convolutional networks for robust surveillance systems. *EURASIP Journal on Image and Video Processing* 2019, 1 (2019), 10.

[40] Dingwen Zhang, Deyu Meng, and Junwei Han. 2017. Co-Saliency Detection via a Self-Paced Multiple-Instance Learning Framework. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39, 5 (2017), 865–878.

[41] Jian Zhang, Chenglong Zhao, Bingbing Ni, Minghao Xu, and Xiaokang Yang. 2019. Variational Few-Shot Learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*.

[42] Yulin Zhang, bo Ma, Yuqing Feng, and Meng Li. 2021. PMT-Net: Progressive Multi-Task Network for One-Shot Person Re-Identification. *Information Sciences* (2021).

[43] Zhizheng Zhang, Cuiling Lan, Wenjun Zeng, and Zhibo Chen. 2019. Densely Semantically Aligned Person Re-Identification. (2019), 667–676.

[44] Liang Zheng, Zhi Bie, Yifan Sun, Jingdong Wang, Chi Su, Shengjin Wang, and Qi Tian. 2016. Mars: A video benchmark for large-scale person re-identification. In *European Conference on Computer Vision*. Springer, 868–884.

[45] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. 2015. Scalable person re-identification: A benchmark. In *Proceedings of the IEEE international conference on computer vision*. 1116–1124.

[46] Zhedong Zheng, Xiaodong Yang, Zhiding Yu, Liang Zheng, Yi Yang, and Jan Kautz. 2019. Joint discriminative and generative learning for person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2138–2147.

[47] Zhedong Zheng and Yi Yang. 2020. Rectifying Pseudo Label Learning via Uncertainty Estimation for Domain Adaptive Semantic Segmentation. *arXiv preprint arXiv:2003.03773* (2020).

[48] Zhedong Zheng, Liang Zheng, and Yi Yang. 2018. Pedestrian alignment network for large-scale person re-identification. *IEEE Transactions on Circuits and Systems for Video Technology* (2018).

[49] Zhun Zhong, Liang Zheng, Donglin Cao, and Shaozi Li. 2017. Re-ranking person re-identification with k-reciprocal encoding. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 1318–1327.