Reliability-Aware Prediction via Uncertainty Learning for Person Image Retrieval

Zhaopeng Dou[1], Zhongdao Wang[1], Weihua Chen[2],
Yali Li[1], and Shengjin Wang[1✉]

[1] Department of Electronic Engineering and BNRist, Tsinghua University, China
[2] Machine Intelligence Technonlogy Lab, Alibaba Group
dcp19@mails.tsinghua.edu.cn, wgsgj@tsinghua.edu.cn

Abstract. Current person image retrieval methods have achieved great improvements in accuracy metrics. However, they rarely describe the reliability of the prediction. In this paper, we propose an Uncertainty-Aware Learning (UAL) method to remedy this issue. UAL aims at providing reliability-aware predictions by considering data uncertainty and model uncertainty simultaneously. Data uncertainty captures the "noise" inherent in the sample, while model uncertainty depicts the model's confidence in the sample's prediction. Specifically, in UAL, (1) we propose a sampling-free data uncertainty learning method to adaptively assign weights to different samples during training, down-weighting the low-quality ambiguous samples. (2) we leverage the Bayesian framework to model the model uncertainty by assuming the parameters of the network follow a Bernoulli distribution. (3) the data uncertainty and the model uncertainty are jointly learned in a unified network, and they serve as two fundamental criteria for the reliability assessment: if a probe is high-quality (low data uncertainty) and the model is confident in the prediction of the probe (low model uncertainty), the final ranking will be assessed as reliable. Experiments under the risk-controlled settings and the multi-query settings show the proposed reliability assessment is effective. Our method also shows superior performance on three challenging benchmarks under the vanilla single query settings. The code is available at: **https://github.com/dcp15/UAL**

Keywords: Person Image Retrieval, Uncertainty, Reliability Assessment.

1 Introduction

Person image retrieval, also known as person re-identification (ReID), aims at associating a target person across non-overlapping camera views [45, 48, 62]. Although current methods [21, 17, 32, 29, 44, 15, 4, 20] have achieved promising performance on public benchmarks, they are reliability-agnostic, *i.e.*, the prediction of a probe can be generated anyway, but they rarely describe whether the prediction is reliable. However, when people are identifying pedestrians, they not only give the judgment result but also the reliability associated with it. Such a reliability assessment mechanism is important in human decision-making [9] and
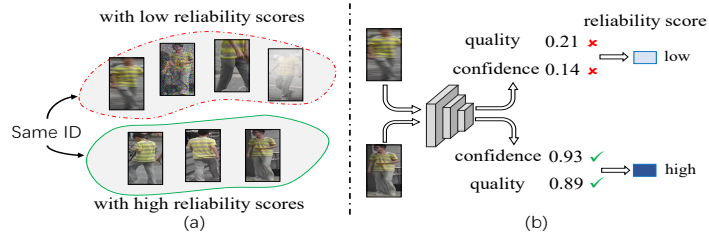
Fig. 1. Observation and Motivation. (a) In the multi-query setting, low-quality query images contain more ambiguous information. Reliability scores are required to down-weight the weights of these queries. (b) The reliability score is related to two factors: the quality of the sample and the model's confidence in the prediction of the sample.

also essential in the ReID task. For example, in real scenarios, we often face the problem of searching for a person through his/her multiple images (*i.e.*, multi-query settings), as a pedestrian is usually captured by several cameras and one camera may capture a series of observations of the person. More generally, we can add the retrieved positive ones into the query set for further comprehensive retrieval. The quality of these query images varies, especially in complex scenes. As shown in Fig. 1(a), low-quality query images contain more ambiguous information. If we treat these query images equally, performance will degrade. At this time, reliability scores are required to down-weight the low-quality query images. However, current methods rarely consider the reliability assessment problem.

To remedy this issue, we propose a novel Uncertainty-Aware Learning (UAL) method for the ReID task. UAL aims at not only giving an accurate prediction for a sample but also providing a reliability score associated with it. The reliability score is related to two factors, *i.e.*, the quality of the sample and the confidence of the model in the prediction of the sample. These two factors are measured by considering two types of uncertainty, *i.e.*, data uncertainty and model uncertainty. Data uncertainty captures the "noise" inherent in the observation and it can describe the quality of the sample. Model uncertainty represents the model's "ignorance" and it can reflect the model's confidence in its prediction [24, 7].

In this paper, we propose a unified network to learn the data uncertainty and the model uncertainty simultaneously. Specifically, first, we project a sample into a Gaussian distribution in the latent space, the mean of the distribution represents the feature, and the variance represents the data uncertainty. Different from [1, 52, 41] sampling feature vector from the Gaussian distribution, we propose a sampling-free method to learn the data uncertainty and adaptively down-weight low-quality ambiguous samples during training. Second, we leverage the Bayesian framework to learn the model uncertainty, in which the parameters of the network are assumed to follow the Bernoulli distribution. The model uncertainty is defined as the dispersion degree of the feature vectors caused by the distribution of the network parameter. Third, the data uncertainty and model uncertainty are jointly learned in a unified network, and they serve as two criteria to assess whether the result is reliable: as shown in Fig. 1(b), if a query

image is high-quality (low data uncertainty) and the model is confident in its prediction of the query image (low model uncertainty), the final result will be assessed as reliable. Experiments under risk-controlled settings and multi-query settings show the proposed reliability assessment is effective.

The major contributions are summarized as: **(1)** We propose an uncertainty-aware learning (UAL) method that can provide reliability-aware predictions for the ReID task. **(2)** We introduce a sampling-free data uncertainty learning method, which can improve the representation by explicitly inhibiting the negative impact of low-quality samples during training without any external clues. **(3)** We propose a unified network to jointly learn data uncertainty and model uncertainty. As far as we know, this is the first work to apply data uncertainty and model uncertainty to the ReID task simultaneously. **(4)** Experiments under risk-controlled settings and multi-query settings show the reliability assessment is effective. Our method also shows superior performance in single query settings.

2 Related Work

Person ReID. Person ReID aims to associate a target person across different camera views. Existing methods can be broadly divided into two categories: hand-craft methods [33, 51] and deep learning methods [30, 18, 20, 4, 15, 52]. The key challenge is the large appearance variation caused by imperfect detection, different camera views, poses, and occlusions. To remedy these issues, several works [12, 13, 35, 47, 8, 62, 14, 32, 21] are proposed to learn local features to cope with the appearance variation. Although these methods have played a certain role, they are reliability-agnostic. That is, the model can output a prediction for a probe anyway, but it does not describe the reliability of the prediction.

Uncertainty in person ReID. There are mainly two types of uncertainty: data uncertainty and model uncertainty [7, 23, 40, 24]. Many tasks have considered the uncertainty to improve the robustness and interpretability of models, such as face recognition [41, 25, 1], semantic segmentation [19, 24] and Multi-view learning [10]. In the ReID task, prior arts [52, 55, 43, 22] consider data uncertainty to alleviate the problem of label noise or data outliers. D-Net [52] maps each person image as a Gaussian distribution in the latent space with the variance indicating the data uncertainty. PUCNN [43] extends the data uncertainty in D-Net into the part-level feature. UNRN [55] incorporates the uncertainty into a teacher-student framework to evaluate the reliability of the predicted pseudo labels for unsupervised domain adaptive (UDA) person ReID. The uncertainty is estimated as the inconsistency of these two models in terms of their predicted soft multi-labels. UMTS [22] designs an uncertainty-aware knowledge distillation loss to transfer the knowledge of the multi-shots model into the single-shot model. Among these methods, the most relevant method to ours is D-Net [52]. Compared to D-Net, our data uncertainty learning method is sampling-free, which can explicitly suppress the ambiguous information contained in low-quality samples. We jointly learns the data uncertainty and the model uncertainty, which can utilize the complementary information provided by them during training.

3   Methodology

The reliability score is related to two factors: the quality of the sample and the confidence of the model in its prediction, which are measured by data uncertainty (Sec. 3.1) and model uncertainty (Sec. 3.2), respectively. They are incorporated into a unified network (Sec. 3.3) for joint learning. Two settings (risk-controlled and multi-query settings) are proposed to verify the effectiveness in Sec. 3.4.

3.1   Learning Data Uncertainty

Data uncertainty captures the "noise" inherent in the observation. It can reflect the quality of the sample, which is an essential factor in the reliability assessment.

Prior method. Prior art D-Net [52] considers the data uncertainty by mapping a sample $\boldsymbol{x}$ as a Gaussian distribution in the latent space,

$$p(\boldsymbol{z}|\boldsymbol{x}) = \mathcal{N}(\boldsymbol{z}; \boldsymbol{\mu}, \boldsymbol{\sigma}^2 \mathbf{I}) \tag{1}$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}^2$ are the mean and variance vectors. $\boldsymbol{\mu}$ is the feature vector and $\boldsymbol{\sigma}^2$ refers to the data uncertainty of $\boldsymbol{x}$. Then, they sample features from $p(\boldsymbol{z}|\boldsymbol{x})$ by re-parameterization trick [27]: $\boldsymbol{z}' = \boldsymbol{\mu} + \boldsymbol{\epsilon}\boldsymbol{\sigma}, \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. The sampled $\boldsymbol{z}'$ are utilized for vanilla cross-entropy loss $\mathcal{L}_{ce}$. To prevent the trivial solution of variance decreasing to zero, a regularization term $\mathcal{L}_{fu}$ is added to constrain the entropy of $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\sigma}^2 \mathbf{I})$ to be larger than a constant. The final loss function is,

$$\mathcal{L} = \mathcal{L}_{ce} + \lambda \mathcal{L}_{fu} \tag{2}$$

where λ is the hyper-parameter to balance $\mathcal{L}_{ce}$ and $\mathcal{L}_{fu}$. Although this method can capture the data uncertainty, there are two limitations: (1) it is sampling-based, *i.e.*, the feature is sampled from the Gaussian distribution during training, which makes the optimization more difficult because each iteration optimizes only one point in the distribution, rather than entire distribution. (2) the objective does not explicitly distinguish samples with different data uncertainty. It is unclear how data uncertainty affects feature learning. To mitigate these two issues, we propose a sampling-free method to learn the data uncertainty and explicitly adjust the attention to the samples according to their quality.

Our sampling-free data uncertainty learning method. We project a sample $\boldsymbol{x}$ into a Gaussian distribution $\mathcal{N}(\boldsymbol{\mu}, \sigma^2 \mathbf{I})$ in the latent space. Then the likelihood of $\boldsymbol{x}$ belonging to class i is formulated by,

$$p(\boldsymbol{x}|y=i) \propto \frac{1}{(2\pi\sigma^2)^{\frac{d}{2}}} \exp\left(-\frac{\|\boldsymbol{\mu} - \boldsymbol{w}_i\|^2}{2\sigma^2}\right) \tag{3}$$

where $\boldsymbol{w}_i$ is the weight vector of i-th class in the classifier and d is the feature dimension. Assuming each class has the equal prior probability, the posterior of $\boldsymbol{x}$ belonging to the class i is,

$$p(y=i|\boldsymbol{x}) = \frac{\exp\left(-\frac{\|\boldsymbol{\mu}-\boldsymbol{w}_i\|^2}{2\sigma^2}\right)}{\sum_j \exp\left(-\frac{\|\boldsymbol{\mu}-\boldsymbol{w}_j\|^2}{2\sigma^2}\right)} = \frac{\exp\left(\frac{1}{\sigma^2}\boldsymbol{w}_i^T\boldsymbol{\mu}\right)}{\sum_j \exp\left(\frac{1}{\sigma^2}\boldsymbol{w}_j^T\boldsymbol{\mu}\right)} \tag{4}$$

Where $\boldsymbol{\mu}$ and $\boldsymbol{w}_*$ are l_2-normalized. $p(y|\boldsymbol{x})$ can be regarded as a Boltzmann distribution. The magnitude of σ^2 controls the entropy of this distribution. The larger the σ^2, the larger the entropy. Thus σ^2 can be regarded as the data uncertainty of sample $\boldsymbol{x}$. Assuming the class label of the sample $\boldsymbol{x}$ is i, the loss function is formulated by,

$$\mathcal{L}_d(\boldsymbol{\mu}, \sigma^2) = -\log p(y = i|\boldsymbol{x}) \approx \frac{1}{\sigma^2}\mathcal{L}(\boldsymbol{\mu}) + \log \sigma^2 \tag{5}$$

where $\mathcal{L}(\boldsymbol{\mu}) = -\log \frac{\exp(\boldsymbol{w}_i^T \boldsymbol{\mu})}{\sum_j \exp(\boldsymbol{w}_j^T \boldsymbol{\mu})}$ is the cross entropy loss. Please see supplementary materials for derivation details.

Discussion (difference to D-Net [52]). Compared to sampling-based method D-Net [52], our method shows several superior qualities: (1) Eq. 5 does not need to sample the representation from the Gaussian distribution. It contains entire information of the representation distribution. (2) Low-quality samples with larger data uncertainty will contribute less to learning the latent space. $\mathcal{L}(\boldsymbol{\mu})$ is weighted by $\frac{1}{\sigma^2}$, and thus it will drive the weight vector $\boldsymbol{w}$ in the classifier to be closer to high-quality samples with small σ^2. This can suppress the ambiguous information contained in low-quality samples during feature learning, which is verified in Sec. 4.4. (3) The σ^2 can not be too large or too small. If σ^2 is too small (large), the first (last) term $\frac{1}{\sigma^2}\mathcal{L}(\boldsymbol{\mu})$ ($\log \sigma^2$) becomes too large. Unlike the term $\mathcal{L}_{fu}$ in Eq. 2, we maintain the σ^2 in a unified formulation.

3.2 Learning Model Uncertainty

Model uncertainty plays an important role in the reliability assessment as it reflects the confidence of the model in its prediction of a sample. Bayesian network is often used to capture the model uncertainty [24]. Suppose we have a dataset $\mathcal{D} = (\boldsymbol{X}, \boldsymbol{Y})$, we define $f_{\boldsymbol{\theta}}$ to be a neural network such that $f_{\boldsymbol{\theta}} : \boldsymbol{X} \to \boldsymbol{Y}$ and $\boldsymbol{\theta}$ corresponds to the weight of the network. To capture model uncertainty, we assume a prior distribution on the weight, *i.e.*, $p(\boldsymbol{\theta})$. We need to obtain the posterior $p(\boldsymbol{\theta}|\mathcal{D})$. Since $p(\boldsymbol{\theta}|\mathcal{D})$ is typically intractable, an approximate distribution $q_{\boldsymbol{\pi}}(\boldsymbol{\theta})$ parameterized by $\boldsymbol{\pi}$ is defined. $q_{\boldsymbol{\pi}}(\boldsymbol{\theta})$ is aimed to be as similar as possible to $p(\boldsymbol{\theta}|\mathcal{D})$, which is measured by the Kullback-Leibler divergence. The optimal parameters $\boldsymbol{\pi}^*$ is,

$$\boldsymbol{\pi}^* = \arg\min_{\boldsymbol{\pi}} \mathrm{KL}[q_{\boldsymbol{\pi}}(\boldsymbol{\theta})\|p(\boldsymbol{\theta}|\mathcal{D})] = \arg\min_{\boldsymbol{\pi}} \mathrm{KL}[q_{\boldsymbol{\pi}}(\boldsymbol{\theta})\|p(\boldsymbol{\theta})] - \mathbb{E}_{q_{\boldsymbol{\pi}}(\boldsymbol{\theta})}[\log p(\mathcal{D}|\boldsymbol{\theta})] \tag{6}$$

Please see supplementary materials for derivation details. Similar to [6], we assume the prior distribution $p(\boldsymbol{\theta})$ as the Bernoulli distribution and perform Monte Carlo integration to the second term. Then the objective loss can be written as $\mathcal{L}_m = -\frac{1}{T}\sum_{t=1}^{T}[\log p(\mathcal{D}|\boldsymbol{\theta}_t)]$, where $\boldsymbol{\theta}_t$ is sampled from $q_{\boldsymbol{\pi}}(\boldsymbol{\theta})$. In $q_{\boldsymbol{\pi}}(\boldsymbol{\theta})$, $\boldsymbol{\theta}_{ij} = \boldsymbol{\pi}_{ij} * \boldsymbol{z}_{ij}$, where $\boldsymbol{z}_{ij} \sim \mathrm{Bernoulli}(\rho)$. ρ is the hyper-parameters and $\boldsymbol{\pi}$ is the set of parameters to be optimized. For a sample $\boldsymbol{x}$, we denote the extracted feature is $\boldsymbol{\mu}_t = f_{\boldsymbol{\theta}_t}(\boldsymbol{x})$ when the weight of network is $\boldsymbol{\theta}_t$. The model uncertainty is
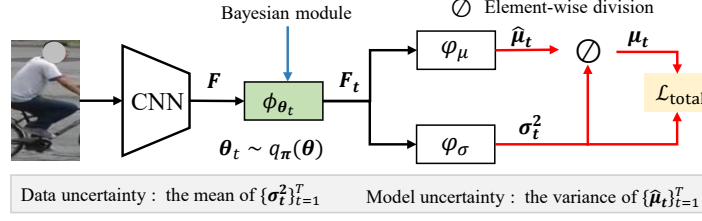
Fig. 2. Overview of the proposed method. An image is fed into the backbone (CNN) to obtain the feature maps $\boldsymbol{F}$. $\boldsymbol{F}$ is further input into the Bayesian module ϕ whose parameter is $\boldsymbol{\theta}_t$ to obtain $\boldsymbol{F}_t$. Then $\boldsymbol{F}_t$ is input to two deterministic modules φ_μ and φ_σ to predict the feature $\hat{\boldsymbol{\mu}}_t$, $\boldsymbol{\mu}_t$ and the data uncertainty $\boldsymbol{\sigma}_t^2$, respectively. The model uncertainty is the variance of $\{\hat{\boldsymbol{\mu}}_t\}_{t=1}^T$, and the data uncertainty is the mean of $\{\boldsymbol{\sigma}_t^2\}_{t=1}^T$.

defined by the variance of the features over the network parameter distribution,

$$\boldsymbol{\sigma}_m^2 = \frac{1}{T}\sum_{t=1}^T (\boldsymbol{\mu}_t - \bar{\boldsymbol{\mu}})^2 \tag{7}$$

where $\bar{\boldsymbol{\mu}} = \frac{1}{T}\sum_{t=1}^T \boldsymbol{\mu}_t$ and squaring operations in Eq. 7 are element-wise.

3.3   Jointly Learning Data and Model Uncertainty

To avoid the additional overhead caused by separately learning the data uncertainty and the model uncertainty, and at the same time leverage the complementary information provided by them for representation learning, we integrate them into a unified network for joint learning. As shown in Fig. 2, given a sample $\boldsymbol{x}$, we first feed it into a CNN backbone to get the feature maps $\boldsymbol{F} \in \mathbb{R}^{h \times w \times c}$, where h, w, c are height, weight and channel, respectively. $\boldsymbol{F}$ is further fed into a Bayesian module ϕ to obtain $\boldsymbol{F}_t = \phi_{\boldsymbol{\theta}_t}(\boldsymbol{F})$, where $\boldsymbol{\theta}_t \sim q_{\boldsymbol{\pi}}(\boldsymbol{\theta})$ corresponds to the parameters of ϕ. Then we input $\boldsymbol{F}_t$ into two deterministic modules φ_μ and φ_σ to obtain the embedding $\hat{\boldsymbol{\mu}}_t = \varphi_\mu(\boldsymbol{F}_t)$ and data uncertainty $\boldsymbol{\sigma}_t^2 = \varphi_\sigma(\boldsymbol{F}_t)$, respectively. Here, $\boldsymbol{F}_t$, $\hat{\boldsymbol{\mu}}_t$, $\boldsymbol{\sigma}_t^2$ have the same shape as $\boldsymbol{F}$. We further regularize $\hat{\boldsymbol{\mu}}_t$ by quality-aware pooling: $\boldsymbol{\mu}_t = \mathrm{GAP}(\hat{\boldsymbol{\mu}}_t \oslash \boldsymbol{\sigma}_t^2) \in \mathbb{R}^c$, where $\oslash$ represents the element-wise division and GAP refers to the global average pooling. Thus, for the sample $\boldsymbol{x}$, we get its feature distribution $\mathcal{N}(\boldsymbol{\mu}_t, \sigma_t^2\mathbf{I})$, where σ_t^2 is the mean of $\boldsymbol{\sigma}_t^2$ across all elements. During training, the loss function is,

$$\mathcal{L}_{\mathrm{total}} = \mathcal{L}_d(\boldsymbol{\mu}_t, \sigma_t^2) + \mathcal{L}_{\mathrm{tri}}(\boldsymbol{\mu}_t) \tag{8}$$

where $\mathcal{L}_d$ is defined in Eq. 5 and $\mathcal{L}_{\mathrm{tri}}(\boldsymbol{\mu}_t)$ is the triplet loss [16] on $\boldsymbol{\mu}_t$.

In inference, for a sample $\boldsymbol{x}$, we can sample T times from $q_{\boldsymbol{\pi}}(\boldsymbol{\theta})$ and obtain $\{\hat{\boldsymbol{\mu}}_t\}_{t=1}^T$, $\{\boldsymbol{\mu}_t\}_{t=1}^T$ and $\{\sigma_t^2\}_{t=1}^T$. The data uncertainty is formulated by $\sigma_d^2 = \frac{1}{T}\sum_{t=1}^T \sigma_t^2$. Then $\boldsymbol{\sigma}_m^2$ is estimated as the variance of $\{\hat{\boldsymbol{\mu}}_t\}_{t=1}^T$ according to Eq. 7 and the model uncertainty σ_m^2 is defined as the mean of $\boldsymbol{\sigma}_m^2$ across all elements. The final representation is calculated by $\bar{\boldsymbol{\mu}} = \frac{1}{T}\sum_{t=1}^T \boldsymbol{\mu}_t$.

Note that $\boldsymbol{\sigma}_t^2$ is the data uncertainty of $\boldsymbol{x}$ under the parameter $\boldsymbol{\theta}_t$. In practice, we train the network to predict $\boldsymbol{s}_t$, and $\boldsymbol{\sigma}_t^2 := \log(1 + \exp(\boldsymbol{s}_t))$. This is because that it is more numerically stable than directly predicting the $\boldsymbol{\sigma}_t^2$, as it ensures that each element of $\boldsymbol{\sigma}_t^2$ is greater than zero. ϕ, φ_μ and φ_σ are all light weighted modules and their architectures are detailed in supplementary materials.

3.4   Reliability Assessment

Based on the risk-controlled settings and the multi-query settings, we introduce how to leverage the learned data and model uncertainty for reliability assessment.

Risk-controlled settings. When facing complex application scenarios, to control the cost of errors, we would expect the model to reject input images (probes) if it can not deal with them. We show the proposed uncertainty mechanism can be naturally utilized as such a "risk indicator". Given a probe $\boldsymbol{x}$, we can obtain its data uncertainty σ_d^2 and model uncertainty σ_m^2. The probe will be assessed as safety (the prediction is reliable) only if $\frac{1}{\sigma_d^2} > \gamma_d$ and $\frac{1}{\sigma_m^2} > \gamma_m$, where γ_d and γ_m are two thresholds that can be set based on the risk tolerance.

Multi-query settings. In real scenarios, how to utilize multiple query images from the same identity to search this person is an essential issue as an interested pedestrian is usually captured by several cameras. Our method is naturally suitable for such settings because it can suppress the negative impact of ambiguous queries according to the reliability score. Considering $\mathcal{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ are query images from the same identity and $\boldsymbol{y}$ is an image in the gallery set. The key issue is how to measure the similarity between $\mathcal{X}$ and $\boldsymbol{y}$. Let $\sigma_{d,i}^2$ ($\sigma_{m,i}^2$) be the data (model) uncertainty of $\boldsymbol{x}_i$. To combine the data uncertainty and model uncertainty without being affected by their numerical scale, we project the data (model) uncertainty into the interval $[\tau_{\min}, \tau_{\max}]$. Specifically, $\sigma_{d,i}^2$ is projected to $d_i = \beta_i \tau_{\min} + (1 - \beta_i)\tau_{\max}$, where $\beta_i = \frac{\sigma_{d,i}^2 - \sigma_{d,\min}^2}{\sigma_{d,\max}^2 - \sigma_{d,\min}^2}$ and $\sigma_{d,\max}^2$ ($\sigma_{d,\min}^2$) is the maximum (minimum) value in $\{\sigma_{d,i}^2\}_{i=1}^n$. The model uncertainty $\sigma_{m,i}^2$ is mapped to m_i in the similar way. The **reliability score** of $\boldsymbol{x}_i$ is $w_i = \frac{m_i d_i}{\sum_{i=1}^n m_j d_j}$. The similarity between $\mathcal{X}$ and $\boldsymbol{y}$ is calculated by $s = \sum_{i=1}^n w_i s_i$, where s_i is the similarity between $\boldsymbol{x}_i$ and $\boldsymbol{y}$. The similarity s considers the reliability of each element in $\mathcal{X}$. If an element has larger reliability score, it plays a more important role.

4   Experiments

4.1   Datasets and Evaluation Metrics

The datasets we use include Market-1501 [56], MSMT17 [49], CUHK03 [31, 59], Occluded-Duke [36], Occluded-REID [63] and Partial-REID [57]. **Market-1501** has 12,936 training, 3,368 query and 19,732 gallery images. **CUHK03** contains 13,164 images of 1,467 identities. We adopt the new testing protocol proposed in [59]. **MSMT17** is the largest image dataset for person ReID. It contains 126,441 images of 4,101 identities. **Occluded-Duke** is reconstructed

from DukeMTMC-reID [38] by selecting occluded images as query set. It has 15,618 training images, 2,210 query and 17,661 gallery images. **Occluded-REID** contains 2,000 images belonging to 200 identities. Each identity has five occluded person images and five-full-body images. **Partial-REID** contains 600 images from 60 person, with five partial images and five full-body images per person.

   Evaluation metrics. We use the Cumulative Matching Characteristic (CMC) curve and mean average precision (mAP) as the evaluation metric.

4.2   Implementation Details

Input images are resized to 256×128. During training, images are augmented by random cropping, random horizontal flipping, and random erasing [60]. Following [45], we adopt ResNet50 [11] pre-trained on ImageNet as the backbone. For a fair comparison with methods employing more complex backbone networks, following ISP [62], we also employ HRNet-W32 [42] as our backbone. There are 64 images from 16 identities in a mini-batch. The initial learning rate is 3.5×10^{-4}. For MSMT17 [49] (other datasets), we train our model 160 (120) epochs, and the learning rate is decreased to its 0.1 and 0.01 at the 70th (40th) and 110th (70th) epochs, respectively. Each epoch has 200 iterations. The optimizer is Adam [26]. The ρ of Bernoulli distribution in the Bayesian module ϕ is empirically set as 0.7. The $\tau_{\min}$ and $\tau_{\max}$ in the multi-query settings are set as 0.5 and 1.0, respectively.

   During testing, we first obtain the parameters $\{\boldsymbol{\theta}_t\}_{t=1}^T$ of the Bayesian module ϕ by sampling T times from $q_{\boldsymbol{\pi}}(\boldsymbol{\theta})$. For each $\boldsymbol{\theta}_t$, we use it to handle all samples. This ensures that the features of different samples are extracted by the same network, which makes it possible to employ our method with a small T. As the Bayesian module ϕ is at the back of the entire network, we only need to repeatedly input it into ϕ, φ_μ and φ_σ for T times, rather than entire network, which saves a lot of expenditure (related results are shown in Table 5).

4.3   Experiments on the Reliability Assessment

Risk-controlled settings. Here, we design experiments to verify whether the proposed uncertainty can serves as the "risk indicator" as described in Sec. 3.4. Specifically, we allow the model to filter out some queries it is diffident to maintain higher performance. Here, we consider four criteria for filtering, *i.e.*, (1) *UNIQUE* [53]: unified no-reference image quality and uncertainty evaluator; (2) *BIQI* [37]: blind image quality indices; (3) *RE*: the reciprocal of the entropy of the predicted category distribution; (4) *Ours*: the reciprocal of the proposed uncertainty as described in Sec. 3.4. (*M:* model uncertainty alone; *D:* data uncertainty alone; *M&D:* combing model uncertainty and data uncertainty). For each criterion, we first calculate its maximum and minimum on all queries, denoted as $\gamma_{\max}$ and $\gamma_{\min}$. Then we setting the threshold as $\gamma = \alpha\gamma_{\max} + (1 - \alpha)\gamma_{\min}$. For *M&D*, a query will be kept only if $\frac{1}{\sigma_d^2} > \gamma_{d,\alpha}$ and $\frac{1}{\sigma_m^2} > \gamma_{m,\alpha}$, where σ_d^2 and σ_m^2 are the data uncertainty and the model uncertainty, and $\gamma_{d,\alpha}$ and $\gamma_{m,\alpha}$ are their thresholds at parameter α. We report the mAP score against the α under
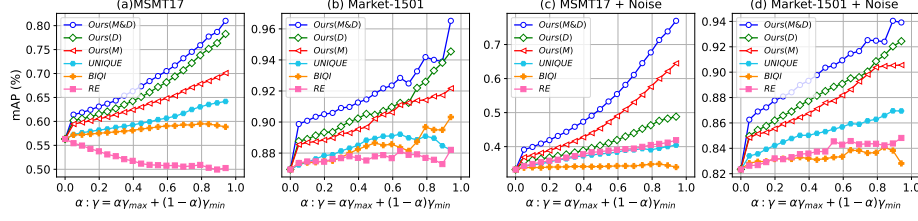
Fig. 3. Comparison of the risk-controlled settings. Filtering out a proportion of query images according to the thresholds of different criteria. D: data uncertainty; M: model uncertainty; $M\&D$: model uncertainty and data uncertainty; $BIQI$ [37]: blind image quality indices. RE: the reciprocal of the entropy of the predicted category distribution. $UNIQUE$ [53]: unified no-reference image quality and uncertainty evaluator.

Original images   Adding noise   Cropping   Motion Blur   Adding fog

Fig. 4. Visualization of different quality-degrading transformations.

different settings. Fig. 3(a-b) shows the results on MSMT17 and Market-1501. In Fig. 3(c-d), we add Gaussian noise to the images in the query set with a probability of 0.5. From the results, we can make several observations. First, for our methods (M, D, and $M\&D$), as α increasing, the mAP score of the remained queries increases. This shows that the proposed uncertainty mechanism help filter out queries whose prediction is unreliable (the model can not deal with) while retaining queries the model is capable of. Second, under different settings, model uncertainty or data uncertainty alone already outperforms $UNIQUE$, $BIQI$ and RE. When they are combined, the mAP score of retained queries is higher. This shows the proposed uncertainty mechanism is a better indicator of the reliability of the prediction, and thus helps for making risk-controlled decisions.

Multi-query settings. We find that there are few images with the same personal identity and same camera identity in the query set of existing datasets. Thus, we reconstruct the test sets of Market-1501, Occluded-Duke and MSMT17 to evaluate the proposed reliability assessment under multi-query settings. Specifically, for each dataset, we first collect the images belonging to the same personal identity and same camera identity from the query set and gallery set. Then, we randomly select half of these images to be allocated to the reorganized query set and the other half to the reorganized gallery set. Simultaneously, to simulate the complex scenes in reality, we transform images in the reorganized query set with a certain probability. The transformation includes: (1) add Gaussian noise to mimic camera quality differences; (2) crop the image to simulate pedestrians are partially out of the camera's field of view; (3) add motion blur to mimic fast-moving pedestrians; (4) add fog to imitate complex weather. Please refer to the

Table 1. Results of the multi-query settings on Market-1501 [56] and MSMT17 [49] and Occluded-Duke [36]. To simulate the complex scenes, we transform the query images via different quality-degrading transformations. MT17→Market: the model is trained on MSMT17 and directly tested on Market-1501. "*w/* R" ("*w/o* R") means we use (don't use) the reliability score to adjust the weights for different queries.

Complex scenes	Method	Market-1501		MSMT17		Occlu-Duke		MT17→Market	
		R1	mAP	R1	mAP	R1	mAP	R1	mAP
Gaussian Noise	D-Net[52]	54.5	52.6	39.8	26.5	43.8	35.9	25.0	12.9
	Ours (*w/o* R)	55.0	53.1	49.9	34.5	51.3	43.7	32.2	18.3
	Ours (*w/* R)	**76.7**	**71.4**	**62.9**	**45.4**	**61.5**	**53.8**	**36.8**	**20.6**
Random Crop	D-Net[52]	36.5	33.6	52.6	33.7	23.8	20.3	22.5	12.7
	Ours (*w/o* R)	37.4	36.0	58.4	38.6	27.0	23.4	28.6	16.4
	Ours (*w/* R)	**46.3**	**44.4**	**63.1**	**42.8**	**35.1**	**31.3**	**32.7**	**18.8**
Motion Blur	D-Net[52]	66.9	59.3	55.9	38.0	60.0	50.2	25.5	14.2
	Ours (*w/o* R)	69.7	63.4	66.7	47.1	62.0	53.3	35.5	20.8
	Ours (*w/* R)	**77.4**	**71.4**	**75.3**	**54.9**	**66.4**	**58.5**	**46.3**	**26.3**
Adding fog	D-Net[52]	62.1	56.6	50.9	33.7	50.3	42.1	29.0	15.7
	Ours (*w/o* R)	65.1	60.3	58.3	40.2	55.9	47.5	37.4	21.9
	Ours (*w/* R)	**77.1**	**71.9**	**72.0**	**52.0**	**64.1**	**56.6**	**47.0**	**26.9**

Table 2. Results of multi-query settings on Occluded-REID [63], Occluded-Duke [36] and Partial-REID [57]. The query images are **not** transformed to degrade the quality.

Method	Occluded-Duke		Occluded-REID		Partial-REID	
	R1	mAP	R1	mAP	R1	mAP
D-Net[52]	75.9	67.0	82.0	72.3	80.0	74.4
Ours (*w/o* R)	76.6	68.3	80.0	72.2	85.0	77.1
Ours (*w/* R)	**78.0**	**70.5**	**83.0**	**73.5**	**88.3**	**80.0**

supplementary materials for the transformation details. Fig. 4 shows some examples of these transformations. In inference, the images belonging to the same personal identity and same camera identity in the query set are regarded as a set of templates that need to be associated. We report the performance of our method against the D-Net [52] in Table. 1. "*w/* R" ("*w/o* R") means we use (don't use) the reliability score to adjust the weights for different queries. From the results, we can make several observations. First, for different complex scenes and different datasets, our method ("*w/o* R") already outperforms D-Net [52]. This shows that our method learns better embedding space, in which the features are more discriminative. When we additionally use the reliability score to adjust the weights for different queries, the performance is further improved. This shows that the proposed reliability score is credible, which can help mine more valuable queries and suppress distractions from low-quality ones.
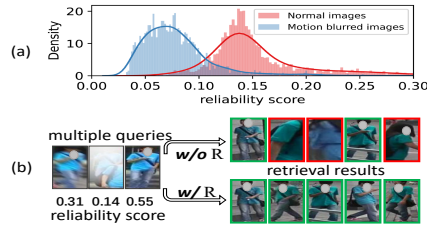
Fig. 5. (a) Distributions of the reliability scores under multi-query settings. (Normal images *v.s.* Motion-blurred images). (b) An example. "*w/o* R" ("*w/* R") means we use (don't use) the reliability score to adjust the weights for different queries.
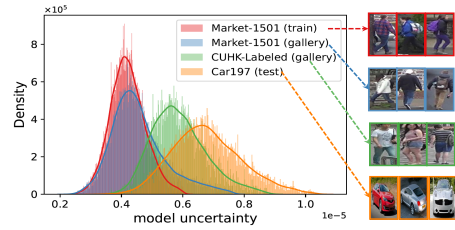
Fig. 6. The role of model uncertainty in reliability assessment. The model is trained with the training set of the Market-1501. The learned model uncertainty is proportional to the degree of deviation between test set domain and training set domain.

We also evaluate our method on Occluded-REID [63], Occluded-Duke [36] and Partial-REID [57]. As query images in these three datasets are occluded or partial, we directly test on them without any quality degradation transformation. Results are shown in Table 2. We can draw the same conclusions as in Table 1.

4.4 Analysis of the Reliability Assessment.

Is the reliability score reasonable? Here, we analyze the reasonableness of the reliability score based on the multi-query settings. For each identity in the test set of Market-1501, we randomly select 10 images to form multiple queries, in which 5 images are downgraded in quality by adding motion blur. Fig. 5 (a) shows the distributions of reliability scores for normal images and motion-blurred images. We can see that, on average, the reliability score of normal images is greater than that of motion-blurred images, showing the reliability score is credible. The retrieval example in Fig. 5 (b) shows that when we adjust the weights of different queries according to the reliability scores, the ambiguous information from the low-quality query is suppressed, resulting in a better performance.

The role of model uncertainty in reliability assessment. We design experiments to verify whether the proposed model uncertainty can describe the confidence of the model in its prediction of the sample. We use the training set of Market-1501 to train the model. Then we estimate the model uncertainty of samples from the training set of Market-1501, the gallery set of Market-1501, the gallery set of CUHK03 and the test set of Car197 [28], respectively. Fig. 6 shows the results. On average, with the increase of the domain deviation, the model uncertainty gradually grows. This shows that the estimated model uncertainty is related to the model's prediction confidence. For out-of-distribution inputs, the model is diffident about the predictions, and the model uncertainty is large.

The role of data uncertainty in reliability assessment. We design experiments to verify whether the estimated data uncertainty can capture the "noise" inherent in the data. We first regard samples in the gallery set of Market-
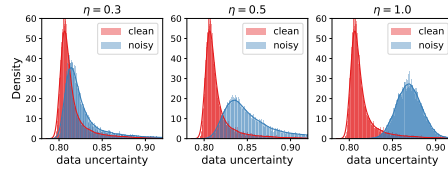
Fig. 7. The role of data uncertainty in reliability assessment. It can be seen that samples with lower quality (larger noise) have larger data uncertainty.
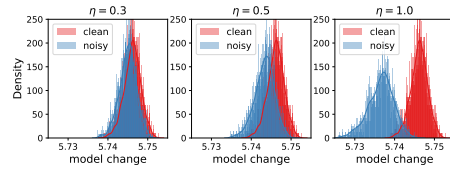
Fig. 8. The role of data uncertainty in the training process. It can be seen that samples with lower quality (larger noise) cause less model change in the training process.

1501 [56] as clean data, and then add noise to pollute them to generate the noisy (low-quality) ones. Specifically, for an image tensor, *i.e.*, $\boldsymbol{x}$, we generate a noise tensor $\boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. Then we pollute the origin data by $\boldsymbol{x} \leftarrow \boldsymbol{x} + \eta\boldsymbol{\varepsilon}$, where η control the strength of the pollution. We gradually vary the size of η to see how data uncertainty changes. The Gaussian kernel density estimation [39] of the estimated data uncertainty are shown in Fig. 7. When η increases, the data uncertainty of noisy samples grows correspondingly. This shows that the proposed data uncertainty can capture the quality of the sample.

The role of data uncertainty in the training process. We regard the samples in the training set of Market-1501 as clean samples and then pollute them to generate the noisy ones. We investigate how the clean and noisy samples affect model learning. Specifically, for each sample, we use it to individually train the model for 10 iterations, and count the model change. The model change is defined as the mean of the absolute difference of the model parameters before and after training. This can reflect the impact of the sample on the learning process. As shown in Fig. 8, when η increases, the model change caused by noisy samples reduces, showing the proposed sampling-free data uncertainty learning method can suppress the contribution of low-quality samples during training.

4.5   Comparison with State-of-the Art Methods

We also compare our method with state-of-the-art methods under the single-query settings. As shown in Table 3, the compared methods are divided into two categories. One category employs ResNet50 [11] or slightly modifies ResNet50 without changing the main structure of the network. The other category employs more powerful backbones than ResNet50, such as HRNet-W32 [42] and Transformer [46]. For fair comparisons, we adopt ResNet50 and HRNet-W32 as our backbone, respectively. From the results, we can make several observations. (1) Whether using ResNet50 or HRNet-W32 as the backbone, our method consistently achieves comparable or superior performance on these datasets. Specifically, when HRNet-W32 is used as the backbone, our method outperforms the previous state-of-the-art methods by +4.5% and +4.0% in terms of mAP scores on MSMT17 and CUHK03-NP (detected), respectively. (2) Our method and

Table 3. Comparison with state-of-the-art methods under the single-query settings.

Methods	Backbone	Market		MSMT17		CUHK03-NP			
						labeled		detected	
		R1	mAP	R1	mAP	R1	mAP	R1	mAP
PCB+RPP [45]	ResNet50	93.8	81.6	–	–	63.7	57.5	–	–
MGN [48]	ResNet50	**95.7**	86.9	–	–	68.0	67.4	66.8	66.0
CAMA [50]	ResNet50	94.7	84.5	–	–	70.1	66.5	66.6	64.2
MHN-6 [2]	ResNet50	95.1	85.0	–	–	77.2	72.4	71.7	65.4
FPR [14]	ResNet50	95.4	86.6	–	–	76.1	72.3	–	–
HOReID [47]	ResNet50	94.2	84.9	–	–	–	–	–	–
DGNet [58]	ResNet50	94.8	86.0	77.2	52.3	–	–	–	–
UAL(*Ours*)	ResNet50	95.2	**87.0**	**80.0**	**56.5**	**78.2**	**75.6**	**76.1**	**72.0**
RGA-SC [54]	R50-RGA	**96.1**	88.4	80.3	57.5	81.1	77.4	79.6	74.5
ABD-Net [3]	ABD-Net	95.6	88.3	82.3	60.8	–	–	–	–
BAT-net [5]	BAT-net	95.1	87.4	79.5	56.8	78.6	76.1	76.2	73.2
OSNet [61]	OSNet	94.8	84.9	78.7	52.9	–	–	72.3	67.8
ISP [62]	HRNet	95.3	88.6	–	–	76.5	74.1	75.2	71.4
PAT [32]	Transformer	95.4	88.0	–	–	–	–	–	–
UAL(*Ours*)	HRNet	95.7	**89.5**	**84.7**	**65.3**	**83.7**	**81.0**	**81.0**	**78.5**

ISP [62] have the same backbone and our method is much more effective. Especially, on CUHK03-NP (detected), our method outperforms ISP by +7.1% mAP score. (3) Compared with semantic-based methods, *e.g.*, FPR [14] and HOReID [47], our method does not need any additional external cues. Compared with methods that use more complex network structures, such as PAT [32] using the transformer [46], our method still shows promising performance.

4.6   Ablation Study

In this part, we conduct ablation studies to show the effectiveness of each component of the proposed method.

The effectiveness of data uncertainty and model uncertainty. Besides based on the backbones ResNet50 and HRNet-W32 implemented by ours, we also conduct the experiments based on the more in-data baseline in **fast-reid**, *i.e.*, BOT [34], with different backbones including ResNet50 (R50), a variant with IBN layers (R50-ibn) and ResNeSt (S50). Experiments are conducted on MSMT17 [49] dataset. As shown in Table 4, both learning data uncertainty and learning model uncertainty improve the performance, and learning data uncertainty provides a larger improvement. The results show these two uncertainties can provide complementary information for learning discriminative latent space.

Impact of hyper-parameter T. During testing, for a sample $\boldsymbol{x}$, we need to sample T times from $q_\pi(\boldsymbol{\theta})$ to obtain the $\bar{\boldsymbol{\mu}}$ and σ_m^2 (as described in Sec. 3.3). Here, we show how T affects the performance on CUHK03-NP [31] and Market-1501 [56]. The results are shown in Table 5. We also report the time cost of the

Table 4. Effectiveness of data uncertainty and model uncertainty. Experiments are conducted on MSMT17 [49]. [†] indicates the experiments are conducted on the backbones in publicly available repository fast-reid: **https://github.com/JDAI-CV/fast-reid**

Uncertainty		ResNet50		[†]BOT(R50)		[†]BOT(R50-ibn)		[†]BOT(S50)		HRNet-W32	
Data	Model	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP
×	×	76.5	52.0	73.9	49.9	79.1	55.4	81.0	59.4	81.6	59.6
✓	×	77.7	53.1	77.1	51.4	80.4	56.0	81.9	59.5	83.4	62.2
×	✓	76.7	52.3	76.6	51.2	80.8	56.9	82.6	60.8	81.9	60.0
✓	✓	**80.0**	**56.5**	**78.7**	**53.6**	**82.4**	**59.1**	**84.1**	**62.1**	**84.7**	**65.3**

Table 5. Impact of the hyper-parameter T. We use the AMD EPYC 7742 CPU and GeForce RTX 3090 GPU. The backbone is ResNet50.

T	CUHK03 (Labeled)			Market-1501		
	Time	R1	mAP	Time	R1	mAP
HOReID [47]	–	–	–	236 s	94.2	84.9
PGFA [35]	–	–	–	193 s	91.2	76.8
Ours (T=5)	17 s	78.4	75.5	66 s	95.0	86.9
Ours (T=10)	24 s	**78.6**	**75.6**	87 s	**95.2**	**87.0**
Ours (T=20)	39 s	78.3	**75.6**	138 s	95.1	**87.0**

entire testing process, from extracting features to calculating the Rank-1 scores. As we can see, our method can work well with a small T. Further, our method outperforms HOReID [47] and PGFA [35] by a large margin while the time cost is much smaller, which shows the effectiveness and efficiency of our method.

5   Conclusions

In this paper, we propose an Uncertainty-Aware Learning (UAL) method for the ReID task to provide reliability-aware predictions, which is achieved by considering two types of uncertainty: data uncertainty and model uncertainty. These two types of uncertainty are integrated into a unified network for joint learning without any external clues. Comprehensive experiments under the risk-controlled settings and the multi-query settings verify that the proposed reliability assessment is effective. Our method also shows superior performance under the single query settings. Meanwhile, we also provide quantitative analyses of the learned data uncertainty and model uncertainty. We expect that our method will provide new insights and attract more interest in the reliability issue in person ReID.

References

1. Chang, J., Lan, Z., Cheng, C., Wei, Y.: Data uncertainty learning in face recognition. In: CVPR. pp. 5710–5719 (2020)
2. Chen, B., Deng, W., Hu, J.: Mixed high-order attention network for person re-identification. In: ICCV. pp. 371–381 (2019)
3. Chen, T., Ding, S., Xie, J., Yuan, Y., Chen, W., Yang, Y., Ren, Z., Wang, Z.: Abd-net: Attentive but diverse person re-identification. In: ICCV. pp. 8351–8361 (2019)
4. Dai, Y., Liu, J., Sun, Y., Tong, Z., Zhang, C., Duan, L.Y.: Idm: An intermediate domain module for domain adaptive person re-id. In: ICCV. pp. 11864–11874 (2021)
5. Fang, P., Zhou, J., Roy, S.K., Petersson, L., Harandi, M.: Bilinear attention networks for person retrieval. In: ICCV. pp. 8030–8039 (2019)
6. Gal, Y., Ghahramani, Z.: Bayesian convolutional neural networks with bernoulli approximate variational inference. arXiv preprint arXiv:1506.02158 (2015)
7. Gal, Y., Ghahramani, Z.: Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In: ICML. pp. 1050–1059 (2016)
8. Gao, S., Wang, J., Lu, H., Liu, Z.: Pose-guided visible part matching for occluded person reid. In: CVPR. pp. 11744–11752 (2020)
9. Gawlikowski, J., Tassi, C.R.N., Ali, M., Lee, J., Humt, M., Feng, J., Kruspe, A., Triebel, R., Jung, P., Roscher, R., et al.: A survey of uncertainty in deep neural networks. arXiv preprint arXiv:2107.03342 (2021)
10. Geng, Y., Han, Z., Zhang, C., Hu, Q.: Uncertainty-aware multi-view representation learning. In: AAAI. pp. 7545–7553 (2021)
11. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR. pp. 770–778 (2016)
12. He, L., Liang, J., Li, H., Sun, Z.: Deep spatial feature reconstruction for partial person re-identification: Alignment-free approach. In: CVPR. pp. 7073–7082 (2018)
13. He, L., Sun, Z., Zhu, Y., Wang, Y.: Recognizing partial biometric patterns. arXiv preprint arXiv:1810.07399 (2018)
14. He, L., Wang, Y., Liu, W., Zhao, H., Sun, Z., Feng, J.: Foreground-aware pyramid reconstruction for alignment-free occluded person re-identification. In: ICCV. pp. 8450–8459 (2019)
15. He, S., Luo, H., Wang, P., Wang, F., Li, H., Jiang, W.: Transreid: Transformer-based object re-identification. arXiv preprint arXiv:2102.04378 (2021)
16. Hermans, A., Beyer, L., Leibe, B.: In defense of the triplet loss for person re-identification. arXiv preprint arXiv:1703.07737 (2017)
17. Hou, R., Ma, B., Chang, H., Gu, X., Shan, S., Chen, X.: Feature completion for occluded person re-identification. IEEE TPAMI (2021)
18. Huang, Y., Zha, Z.J., Fu, X., Hong, R., Li, L.: Real-world person re-identification via degradation invariance learning. In: CVPR. pp. 14084–14094 (2020)
19. Isobe, S., Arai, S.: Deep convolutional encoder-decoder network with model uncertainty for semantic segmentation. In: INISTA. pp. 365–370 (2017)
20. Isobe, T., Li, D., Tian, L., Chen, W., Shan, Y., Wang, S.: Towards discriminative representation learning for unsupervised person re-identification. In: ICCV. pp. 8526–8536 (2021)
21. Jia, M., Cheng, X., Zhai, Y., Lu, S., Ma, S., Tian, Y., Zhang, J.: Matching on sets: Conquer occluded person re-identification without alignment. In: AAAI. pp. 1673–1681 (2021)

22. Jin, X., Lan, C., Zeng, W., Chen, Z.: Uncertainty-aware multi-shot knowledge distillation for image-based object re-identification. In: AAAI. pp. 11165–11172 (2020)
23. Kendall, A., Cipolla, R.: Modelling uncertainty in deep learning for camera relocalization. In: ICRA. pp. 4762–4769 (2016)
24. Kendall, A., Gal, Y.: What uncertainties do we need in bayesian deep learning for computer vision? arXiv preprint arXiv:1703.04977 (2017)
25. Khan, S., Hayat, M., Zamir, S.W., Shen, J., Shao, L.: Striking the right balance with uncertainty. In: CVPR. pp. 103–112 (2019)
26. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
27. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114 (2013)
28. Krause, J., Stark, M., Deng, J., Fei-Fei, L.: 3d object representations for fine-grained categorization. In: ICCVW. pp. 554–561 (2013)
29. Li, H., Wu, G., Zheng, W.S.: Combined depth space based architecture search for person re-identification. In: CVPR. pp. 6729–6738 (2021)
30. Li, H., Wu, G., Zheng, W.S.: Combined depth space based architecture search for person re-identification. In: CVPR. pp. 6729–6738 (2021)
31. Li, W., Zhao, R., Xiao, T., Wang, X.: Deepreid: Deep filter pairing neural network for person re-identification. In: CVPR. pp. 152–159 (2014)
32. Li, Y., He, J., Zhang, T., Liu, X., Zhang, Y., Wu, F.: Diverse part discovery: Occluded person re-identification with part-aware transformer. In: CVPR. pp. 2898–2907 (2021)
33. Liao, S., Hu, Y., Zhu, X., Li, S.Z.: Person re-identification by local maximal occurrence representation and metric learning. In: CVPR. pp. 2197–2206 (2015)
34. Luo, H., Gu, Y., Liao, X., Lai, S., Jiang, W.: Bag of tricks and a strong baseline for deep person re-identification. In: CVPRW. pp. 0–0 (2019)
35. Miao, J., Wu, Y., Liu, P., Ding, Y., Yang, Y.: Pose-guided feature alignment for occluded person re-identification. In: ICCV. pp. 542–551 (2019)
36. Miao, J., Wu, Y., Liu, P., Ding, Y., Yang, Y.: Pose-guided feature alignment for occluded person re-identification. In: ICCV. pp. 542–551 (2019)
37. Moorthy, A.K., Bovik, A.C.: A two-step framework for constructing blind image quality indices. IEEE Signal processing letters **17**(5), 513–516 (2010)
38. Ristani, E., Solera, F., Zou, R., Cucchiara, R., Tomasi, C.: Performance measures and a data set for multi-target, multi-camera tracking. In: ECCV. pp. 17–35 (2016)
39. Scott, D.W.: Multivariate density estimation: theory, practice, and visualization. John Wiley & Sons (2015)
40. Shen, Y., Zhang, Z., Sabuncu, M.R., Sun, L.: Real-time uncertainty estimation in computer vision via uncertainty-aware distribution distillation. In: WACV. pp. 707–716 (2021)
41. Shi, Y., Jain, A.K.: Probabilistic face embeddings. In: ICCV. pp. 6902–6911 (2019)
42. Sun, K., Xiao, B., Liu, D., Wang, J.: Deep high-resolution representation learning for human pose estimation. In: CVPR. pp. 5693–5703 (2019)
43. Sun, W., Xie, J., Qiu, J., Ma, Z.: Part uncertainty estimation convolutional neural network for person re-identification. In: ICIP. pp. 2304–2308 (2021)
44. 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. pp. 393–402 (2019)

45. 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. pp. 480–496 (2018)
46. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. In: NeurIPS. pp. 5998–6008 (2017)
47. Wang, G., Yang, S., Liu, H., Wang, Z., Yang, Y., Wang, S., Yu, G., Zhou, E., Sun, J.: High-order information matters: Learning relation and topology for occluded person re-identification. In: CVPR. pp. 6449–6458 (2020)
48. Wang, G., Yuan, Y., Chen, X., Li, J., Zhou, X.: Learning discriminative features with multiple granularities for person re-identification. In: ACM MM. pp. 274–282 (2018)
49. Wei, L., Zhang, S., Gao, W., Tian, Q.: Person transfer gan to bridge domain gap for person re-identification. In: CVPR. pp. 79–88 (2018)
50. Yang, W., Huang, H., Zhang, Z., Chen, X., Huang, K., Zhang, S.: Towards rich feature discovery with class activation maps augmentation for person re-identification. In: CVPR. pp. 1389–1398 (2019)
51. Yang, Y., Liao, S., Lei, Z., Li, S.Z.: Large scale similarity learning using similar pairs for person verification. In: AAAI (2016)
52. Yu, T., Li, D., Yang, Y., Hospedales, T.M., Xiang, T.: Robust person re-identification by modelling feature uncertainty. In: ICCV. pp. 552–561 (2019)
53. Zhang, W., Ma, K., Zhai, G., Yang, X.: Uncertainty-aware blind image quality assessment in the laboratory and wild. IEEE TIP **30**, 3474–3486 (2021)
54. Zhang, Z., Lan, C., Zeng, W., Jin, X., Chen, Z.: Relation-aware global attention for person re-identification. In: CVPR. pp. 3186–3195 (2020)
55. Zheng, K., Lan, C., Zeng, W., Zhang, Z., Zha, Z.J.: Exploiting sample uncertainty for domain adaptive person re-identification. arXiv preprint arXiv:2012.08733 (2020)
56. Zheng, L., Shen, L., Tian, L., Wang, S., Wang, J., Tian, Q.: Scalable person re-identification: A benchmark. In: ICCV. pp. 1116–1124 (2015)
57. Zheng, W.S., Li, X., Xiang, T., Liao, S., Lai, J., Gong, S.: Partial person re-identification. In: ICCV. pp. 4678–4686 (2015)
58. Zheng, Z., Yang, X., Yu, Z., Zheng, L., Yang, Y., Kautz, J.: Joint discriminative and generative learning for person re-identification. In: CVPR. pp. 2138–2147 (2019)
59. Zhong, Z., Zheng, L., Cao, D., Li, S.: Re-ranking person re-identification with k-reciprocal encoding. In: CVPR. pp. 1318–1327 (2017)
60. Zhong, Z., Zheng, L., Kang, G., Li, S., Yang, Y.: Random erasing data augmentation. In: AAAI. pp. 13001–13008 (2020)
61. Zhou, K., Yang, Y., Cavallaro, A., Xiang, T.: Omni-scale feature learning for person re-identification. In: ICCV. pp. 3702–3712 (2019)
62. Zhu, K., Guo, H., Liu, Z., Tang, M., Wang, J.: Identity-guided human semantic parsing for person re-identification. In: ECCV. pp. 346–363. Springer (2020)
63. Zhuo, J., Chen, Z., Lai, J., Wang, G.: Occluded person re-identification. In: ICME. pp. 1–6. IEEE (2018)