Incomplete Descriptor Mining with Elastic Loss for Person Re-Identification

Hongchen Tan, Yuhao Bian, Huasheng Wang, Xiuping Liu*, and Baocai Yin

Abstract—In this paper, we propose a novel person Re-ID model, Consecutive Batch DropBlock Network (CBDB-Net), to help the person Re-ID model to capture the attentive and robust person descriptor. The CBDB-Net contains two novel modules: the Consecutive Batch DropBlock Module (CBDBM) and the Elastic Loss. In the Consecutive Batch DropBlock Module (CBDBM), it firstly conducts uniform partition on the feature maps. And then, the CBDBM independently and continuously drops each patch from top to bottom on the feature maps, which outputs multiple incomplete features to push the model to capture the robust person descriptor. In the Elastic Loss, we design a novel weight control item to help the deep model adaptively balance hard sample pairs and easy sample pairs in the whole training process. Through an extensive set of ablation studies, we verify that the Consecutive Batch DropBlock Module (CBDBM) and the Elastic Loss each contribute to the performance boosts of CBDB-Net. We demonstrate that our CBDB-Net can achieve the competitive performance on the three generic person Re-ID datasets (the Market-1501, the DukeMTMC-Re-ID, and the CUHK03 dataset), three occlusion Person Re-ID datasets (the Occluded DukeMTMC, the Partial-REID, and the Partial iLIDS dataset), and the other image retrieval dataset (In-Shop Clothes Retrieval dataset).

Index Terms—Person Re-ID, Dropout strategy, Triple Ranking, Incomplete Feature Descriptor

I. INTRODUCTION

Person Re-Identification (Re-ID) has attracted increasing attention from both the academia and the industry due to its significant role in the video surveillance. Given a target person image captured by one camera, the goal of person Re-ID task is to re-identify the same person from images captured by other cameras’ viewpoints. Despite the exciting progress in recent years, the person Re-ID task remains to be extremely challenging. This is because that the task is easily affected by body misalignment, occlusion, background perturbation, and viewpoint changes, etc. To tackle this challenge, almost person Re-ID approaches focus on two strategies: feature descriptor learning and distance metric learning. The former approaches aim to capture the discriminative descriptor of the person, which is robust to various interference factors; The latter approaches aim to gain a better metric space equipped with better classification discrimination of different persons.

Recenly, many outstanding approaches [76], [15], [34], [70], [51], [2], equipped with a series of metric constraints, try to capture the global descriptor of the whole person image from global perspective. However, the global descriptors are easily prone to the false matching between persons who look similar, due to lacking enough local information. Therefore, other approaches [4], [25], [6], [38], [47], [60] introduce the pose estimation or human parsing models to help the person Re-ID model better locate and capture local human part feature. However, the underlying datasets bias between pose estimation, human parsing, and person Re-ID remains an obstacle against the ideal human body partition on person images. And the additional human models make the person Re-ID model more complex and unwieldy. Thus, in this paper, we aim to design a simple and effective strategy to help the person Re-ID model capture a high-quality person descriptor.

Recently, the Dropout strategies [50], [46], [16], [76], [72], [9], is widely used in person Re-ID and other visual tasks. Compared with additional human model assistance, these Dropout strategies are the lightweight and effective for person Re-ID models. It can better help the person Re-ID model embed in the application device. In these Dropout strategies, Cutout [9], random erasing [72], and SpatialDropout [50], can randomly drop the feature pixle in the feature maps or feature vectors. However, these methods only belong to
a regularization method and not attentive feature learning methods. And they can not drop a large contiguous area within a batch. So, Batch DropBlock [26] is proposed to drop the same continuity region for a training batch. Based on the Batch DropBlock, our first idea is to propose a novel drop strategy, Consecutive Batch DropBlock (CBDB), to produce multiple incomplete feature maps for improving the robustness of the person Re-ID model. As shown in the Figure 1 for the Consecutive Batch DropBlock, we firstly conduct uniform partition on the conv-layer. Secondly, we independently and continuously drop each patch from top to bottom on the conv-layer. Since this, we can gain multiple incomplete feature maps to push the deep model to capture the robust feature maps.

Different from Batch DropBlock: (I) Our Consecutive Batch DropBlock can produce more incomplete feature maps to help person Re-ID model capture rich and robust feature; (II) Our consecutive Batch DropBlock drops the same patch for the whole training set instead of a training batch. The Consecutive Batch DropBlock can be regarded as the Batch DropBlock’ improvement, and the complementary descriptors to uniform patch descriptors in PCB [49]. In our Consecutive Batch DropBlock, the deep model would be pushed to capture the key information from the rest feature regions. Thus, our Consecutive Batch DropBlock is an effective attentive feature learning strategy. So, we believe that our Consecutive Batch DropBlock (CBDB) can effectively improve the robustness of person descriptors for the person matching task.

Based on the Consecutive Batch DropBlock, we can gain multiple incomplete feature maps or descriptors in the training process. However, inevitably there will be hard sample pairs and easy sample pairs for the person matching task. The batch hard triplet loss [22] may be a suitable metric loss function to balance these hard sample pairs. However, in the whole training process, the difficulty level of hard sample pairs are different in different training stage; In the whole training sets, the difficulty level of hard sample pairs are also different in different person ID. So, our second idea is to design a novel metric loss function to dynamically balance the hard sample pairs and easy sample pairs in the training process.

Based on the above analysis, we proposed a novel person Re-ID model, Consecutive Batch DropBlock Network (CBDB-Net). The CBDB-Net contains two novel designs: Consecutive Batch DropBlock strategy and novel metric loss function. The former exploits multiple incomplete descriptors to improve the robustness of the deep model. And, in the testing stage, a simple test model is adopted to produce a high-quality person descriptor for the person matching. The latter can better mine and balance the hard sample pairs for the whole training samples in the whole training process. It can further improve the performance of the deep person Re-ID model in the person matching task.

In the experimental section, firstly we validated our CBDB-Net on three generic person Re-ID datasets: Market-1501 [69], DukeMTMC-reID [39], [68], CUHK03 [29]. Secondly, we evaluate our CBDB-Net on three occluded person Re-ID datasets: Occluded-DukeMTMC [39], Partial-REID [55] and Partial-iLIDS [70]. Finally, we believe our CBDB-Net can be applied to other image retrieval tasks. So, we evaluate our CBDB-Net on the In-Shop Clothes retrieval dataset [35]. Extensive experimental results and analysis demonstrate the effectiveness of CBDB-Net and significantly improved performance compared against most state of the arts over two evaluation metrics.

II. RELATED WORK

A. Part-based person Re-ID Models

In our CBDB-Net, it can output many incomplete feature maps of one person to train the deep model. These incomplete feature maps can be regarded as the large parts. Thus, in this subsection, we introduce related works of the part-based person Re-ID task. Recently, [25], [43], [47], [65] adopt an additional human body part detector or an additional human body parsing model to focus on more accurate human parts. e.g., SPReID [23] apply an additional human body parsing model to generate 5 different predefined body part masks to capture more reliable part representations. [18] address the missed contextual information by exploiting both the accurate human body parts and the coarse non-human parts. [36], [53] combine the pose landmarks and uniform partial feature to improve the performance of the occluded person Re-ID task. [55] adopt the key point detector to exploit three coarse body part, and combine the global information to conduct the person matching. Different from [55], [62] drops the keypoint detector, only use the maximum feature responses to locate the body regions and combined with the Part Loss. [49] conducts uniform partition on the conv-layer for learning part-level features. Our CBDB-Net also belongs to part-based person Re-ID methods to some degree. Similar to [49], [62], we also needn’t any additional human models’ assistance. In the CBDB-Net, we can gain multiple large parts by the proposed Consecutive Batch DropBlock strategy. These large feature parts can be regarded as complementary to the uniform parts in [49], [62]. And, compared with the small feature parts in [55], [49], [62], these large feature parts in our CBDB-Net can make the training process of the person Re-ID model more robust.

B. Triplet Ranking in person Re-ID

Triplet ranking loss [44] is one of the most important metric loss functions, which encourages the distance between positive sample pairs to be closer than negative sample pairs. It has been applied in various outstanding deep vision models, and achieves outstanding performance in these metric learning tasks. [45] may be the first one to introduce the triplet loss into the Re-ID task. When SPGAN [8] conducts the cross-domain person Re-ID task, they adopt the contrastive loss to preserve the person ID information in the cross-domain image style transfer. [2] extends the triplet loss by introducing the absolute distance of the positive sample pair. [41] proposed a virtual sample in the triplet unit to accelerate sample distance optimization. Similar to [10], [22] proposed the batch hard triplet loss by introducing the hard sample mining strategy for person sample pairs. Since this, many state-of-the-art Re-ID methods [24], [1], [58], [57], [40], [76] adopt the batch
hard triplet loss to gain a series of outstanding person Re-ID models.

Inspired by these methods [24], [1], [58], [57], [40], [76], we also introduce the batch hard triplet loss to our CBDB-Net. Different from the original batch hard triplet loss [22], we propose a novel triplet loss by revising the batch hard triplet loss. The proposed novel triplet loss can dynamically adjust the learning weights of different hard sample pairs in the whole training process. The experiments show that it can further help the person Re-ID model to gain better performance.

III. CBDB-Net

In this section, we describe the details of the proposed Consecutive Batch DropBlock Network (CBDB-Net). As shown in Figure 2, our CBDB-Net contains four components: the Backbone Network, the Consecutive Batch DropBlock Module (CBDBM), the Elastic Loss and Network Architecture Overview. The Backbone Network provides the basic feature maps for the Consecutive Batch DropBlock Module (CBDBM). In the Consecutive Batch DropBlock Module (CBDBM): firstly, the Consecutive Batch DropBlock Module outputs multiple incomplete feature maps; secondly, these multiple incomplete feature maps are fed into the following ResBlock; in the whole training process, the deep model tries to capture the discriminative feature from these incomplete feature maps. The proposed Elastic Loss is designed to dynamically balance the hard sample pairs and the easy sample pairs in the whole training process. The Network Architecture Overview summaries the whole network architecture and loss functions.

A. Backbone Network

Following current many outstanding methods [24], [1], [58], [57], [40], [76], our CBDB-Net also uses the ResNet-50 [19] pre-trained on ImageNet [7] as the backbone network, to encode a person image \( x \). In order to get a larger size high-level feature tensor, we also modify the basic structure of the ResNet-50 slightly. The down-sampling operation at the beginning of the “ResNet-50 Stage 4” is not employed. Therefore, we can get a larger feature tensor \( T(x) \in \mathbb{R}^{24 \times 8 \times 2048} \).

Fig. 2. The architecture of Consecutive Batch DropBlock Network (CBDB-Net) for person Re-ID task. The two novel strategies are Consecutive Batch DropBlock Module and the proposed Elastic Loss. In the architecture, the “CE” denotes the cross entropy loss function.

B. Consecutive Batch DropBlock Module

Based on the large feature tensor \( T(x) \in \mathbb{R}^{24 \times 8 \times 2048} \), we construct the Consecutive Batch DropBlock Module (CBDBM). (I): As shown in Figure 2, the tensor \( T(x) \) is divided into \( m \) uniform patches. (II): The DropPatch \( i, i = 1, 2, \ldots, m \) is designed to drop the \( i \)th patch on the tensor \( T(x) \). As shown in the Figure 3, the feature tensor \( T(x) \) is divided into 6 uniform patches; The DropPatch \( 3 \) is used to drop the \( 3 \)th patch on the tensor \( T(x) \). So, we can see that the \( 3 \)th patch on the \( T(x) \) are zeroed out. Since this, based on the CBDBM, we can gain \( m \) incomplete feature tensors \( T(x)_{i} \in \mathbb{R}^{24 \times 8 \times 2048}, i = 1, 2, \ldots, m \).

Based on the CBDBM, we can gain multiple incomplete feature tensors. In this case, we can directly append the average pooling operation on these incomplete feature tensors to conduct the person retrieval task. Based on these incomplete feature tensors, we hope the deep model can further correct or capture the discriminative feature for the person matching. Therefore, before the average pooling operation, we append the additional ResNet block, i.e. “ResBlock” in Figure 2, on these incomplete feature tensors. Since this, the deep model has enough chances to correct or capture the discriminative feature for the person matching. Here, the
“ResBlock”, composed of three bottleneck blocks [19], applies a stack of convolution layers on these incomplete feature maps \( T(x)^i \in \mathbb{R}^{24 \times 8 \times 2048}, i = 1, 2, \cdots, m \). Here, these \( m \) feature tensors \( T(x)^i \) share the same “ResNet block”. Thus, \( T(x)^i \in \mathbb{R}^{24 \times 8 \times 2048}, i = 1, 2, \cdots, m \) is fed into the additional “ResBlock” and “Average Pooling” operation in Figure 2. And we can gain \( m \) new person descriptors, i.e. \( \varphi(x)_i \in \mathbb{R}^{512}, i = 1, 2, \cdots, m \) which is fed into the loss functions: the cross-entropy loss and the proposed Elastic Loss [II.C]

C. The proposed Elastic Loss

In the Consecutive Batch DropBlock Module, it outputs \( m \) incomplete descriptors of one person image. In the training process, these incomplete descriptors inevitably contain many hard matching sample pairs for the same person or different persons. Recently, the batch hard triplet loss [22] introduces the hard sample mining strategy to effectively focus on the hard sample pairs in the training process. However, the hard sample mining strategy in the batch hard triplet loss [22] did not consider the two issues: (i) In the different training stage, the difficulty level of hard samples pairs are different; (ii) In each training stage, the difficulty level of hard samples pairs from variant ID person are also different. Thus, it is necessary to propose a novel loss function to dynamically focus on the hard sample pairs in the whole training process.

Recently, Focal loss [33] introduces the weight control item into the cross-entropy loss, which can dynamically adjust the weight of hard samples and easy samples in the training process. Inspired by the Focal loss, we also proposed a novel loss function to relieve the above two issues by revising the batch hard triplet loss [22].

In order to define the Elastic Loss, we firstly organize the training samples into a set of triplet feature units, \( S = (s(x^a), s(x^p), s(x^n)) \) which simply denotes as \( S = (s^a, s^p, s^n) \), and the raw person image triplet units is \( X = (x^a, x^p, x^n) \). Here, \( (s^a, s^p) \) represents a positive pair’s feature with \( y^p = y^p \), and \( (s^a, s^n) \) indicates a negative pair’s feature with \( y^p \neq y^n \). Here, \( y \in Y \) is the person ID information.

Secondly, we revisit the batch hard triplet loss [22]. Based on the triplet loss, [22] extends the triplet loss by introducing the hard sample mining strategy. Here, hard sample mining strategy in the training batch is: the positive sample pair with the largest distance as the hard positive sample pair; the negative sample pair with the smallest distance as the hard negative sample pair. In the training process, the hard sample pair will be focused. Based on the design, the batch hard triplet loss function is defined as:

\[
T_{HardTriplet} = [\eta + \max_{x^a, x^p, x^n} d(s^a, s^p) - \min_{x^a, x^n} d(s^a, s^n)]_+
\]  

(1)

Here, \( \eta \) represents the margin parameter.

Finally, we define the Elastic Loss by revising the batch hard triplet loss. Same as the Eq. [1] the hard positive sample pairs is \( \max_{x^a, x^p} d(s^a, s^p) \), and the hard negative sample pairs is \( \min_{x^a, x^n} d(s^a, s^n) \). In order to dynamically adjust sample pairs' weights, we design a novel weight control item. (I:) we introduce the “nuclear weight term”, \( \delta = \frac{\max d(s^a, s^p)}{\min d(s^a, s^n) + 1} \).

The “nuclear weight term” can adaptively adjust the weights of loss function under the hard positive sample pairs and the hard negative sample pairs. When the distance of hard positive samples pairs becomes larger or the distance of hard negative samples pairs becomes smaller, the \( \delta \) becomes larger. It indicates that the loss function pays more attention on the current person ID’s hard sample pairs in the current training step, and vice versa.

In the ideal training goal, \( \max d(s^a, s^p) \) is smaller than the \( \min d(s^a, s^n) \). However, in fact, we observe that the \( \max d(s^a, s^p) \) is usually larger than the \( \min d(s^a, s^n) \) for many person ID sample pairs. If we directly use the \( \delta \) to adjust the weight of the batch hard triplet loss, the \( \delta \) brings too much weight fluctuation, which is not conducive to model training. Therefore, we hope the weights of \( \delta \) can fluctuate over a small range. (II:) based on the \( \delta \), we design a “Shell function”, i.e. \( f(x) = \frac{1}{1 + e^{-x}} \) which is the sigmoid function.

It can effectively control the \( \delta \) in \([\frac{1}{2}, 1]\). In this way, the weight will not fluctuate too much. And, the easy sample pairs can also get appropriate weights to participate in the effective training of the person Re-ID model.

For the sigmoid function, we can see the function curve and the derivative function curve in the Figure 2. As \( x \), i.e. \( \delta \) in our weight control item, gets bigger, the value of the sigmoid function gets closer and closer to 1. And the value of the derivative function is smaller. It indicates that the weight of the difficult sample is stable at around 1. In the process of the parameter optimization process, the gradient of the batch hard triplet loss function can gain the large learning weight. As \( x \), i.e. \( \delta \) in our weight control item, gets closer and closer to 0, the value of the sigmoid function gets closer and closer to \( \frac{1}{2} \). When \( \delta \) is around 0, the value of the derivative function is large. So, for the easy sample pairs, the loss function gives a flexible and small weight to focus on them.

Based on the “nuclear weight term” and “Shell func-
2) **Testing Stage.** In the testing stage, as shown in Figure 5, we only use a simple network to conduct the person Re-ID task. Compared our training model in Figure 2, our testing model only contains the Backbone network, ResBlock, and Average Pooling Operation. In the retrieval stage, we use the feature $f(x) \in \mathbb{R}^{512}$ in Figure 5 to find the best matching person in the gallery by comparing the squared distance, i.e.

$$d(a,b) = \|a - b\|^2_2.$$ 

---

**IV. Experiment**

In this section, we evaluate the CBDB-Net qualitatively and quantitatively. To evaluate the effectiveness of our CBDB-Net, we conduct extensive experiments on three generic person datasets (the Market-1501 [69], the DukeMTMC-reID [39], [68], and the CUHK03 [29]), three occluded Person Re-ID datasets (the Occluded-DukeMTMC [36], the Partial-REID [56], and the Partial-iLIDS [20]) and one clothes image retrieval dataset (In-Shop Clothes Retrieval dataset [35]).

Firstly, we compare the performance of CBDB-Net against many state-of-the-art methods on these seven datasets. Secondly, we discuss various ablation studies on the four datasets (the Market-1501 [69], the DukeMTMC-reID [39], [68], the CUHK03 [29], and the In-Shop Clothes Retrieval dataset [35]) to validate the effectiveness of each strategy in our CBDB-Net.

**A. Datasets and Evaluation**

**Market-1501** [69] contains 32,668 labeled images of 1,501 identities which is collected from 6 different camera views. Following almost person Re-ID approaches, the whole 1,501 identities are split into two non-overlapping fixed person ID sets: the training set contains 12,936 person images from 751 identities; the testing set contains 19,732 person images from other 750 identities. In the testing stage, we use 3368 query images from 750 test person identities to retrieve the same ID persons from the rest of the test set, i.e. the gallery set.

**DukeMTMC-reID** [39], [68] is also a large-scale person Re-ID dataset. The DukeMTMC-reID contains 36,411 labeled images of 1,404 identities which is collected from 8 different camera views. The training set contains 16,522 person images from 702 identities; In testing stage, we use 2,228 query images from the other 702 identities, and 17,661 gallery images.

**CUHK03** [29] is the most challenging of these three generic person Re-ID datasets. It composed of 14,096 images of 1,4674 identities captured from 6 cameras. It provides bounding boxes detected from manual labeling and deformable part models (DPMs), the latter type is more challenging due to severe bounding box misalignment and cluttered background. Following [25], [24], [76], [49], we use the 767/700 split of [29] with the detected images.

**Occluded-DukeMTMC** [36] contains 15,618 training images, 17,661 gallery images, and 2,210 occluded query images. The Occluded-DukeMTMC is introduced by [36]. We use this dataset to demonstrate that our CBDB-Net also can achieve good performance on the occluded Person Re-ID task.

**Partial-REID** [56] is a specially designed partial person Re-ID dataset which contains 600 images from 60 person
The comparison with many state-of-the-art person Re-ID approaches on the Market-1501, the DukeMTMC-reID and the CUHK03 datasets.

| Method               | Market-1501 | DukeMTMC-reID | CUHK03-Detected | CUHK03-Labeled |
|----------------------|-------------|---------------|-----------------|---------------|
|                      | Rank-1 | mAP  | Rank-1 | mAP  | Rank-1 | mAP  | Rank-1 | mAP  | Rank-1 | mAP  |
| CBDB-Net (m=6)  | 94.1%   | 85.0% | 78.7% | 74.3% | 75.8% | 72.6% | 78.3% | 75.9% |
| CBDB-Net (m=6)+Re-ranking  | 95.6%   | 93.0% | 91.2% | 87.9% | 83.9% | 85.1% | 86.5% | 87.8% |

identities. And each person has 5 partial images in the query set and 5 full-body images in the gallery set. These images are collected at a university campus from different viewpoints, backgrounds, and different types of severe occlusion.

**Partial-iLIDS** [20] is a simulated partial person Re-ID dataset based on the iLIDS dataset. It has a total of 476 images of 119 person identities.

**In-shop clothes retrieval** [35] is a clothes image retrieval dataset. It contains 11,735 classes of clothing items with 54,642 images. The training set contains 25,882 images from 3,997 classes; the testing set contains 28,760 images from 3,985 classes. The test set is divided into the query set of 3,985 classes (14,218 images) and the gallery set of 3,985 classes (12,612 images). We apply our CBDB-Net on the clothes image retrieval task to evidence that our CBDB-Net can be suitable for other image retrieval tasks.

**Evaluation Protocol.** We employ two standard metrics as in most person Re-ID approaches, namely the mean Average Precision (mAP) and the cumulative matching curve (CMC) used for generating ranking accuracy. We use Rank – 1 accuracy and mAP to evaluate the effectiveness of our CBDB-Net on all seven datasets.

**B. Implementation Details**

Following recent many outstanding approaches [37], [76], [54], [3], [36], [74], the input images are re-sized to 384 × 128 and then augmented by random horizontal flip and normalization in the training stage. In the testing stage, the images are also re-sized to 384 × 128 and augmented only by normalization. Based on the pre-trained ResNet-50 backbone, our network is end-to-end in the whole training stage. Our network is trained using 2 single GTX 2080Ti GPUs with a batch size of 64. Each batch contains 16 identities, with 4 samples per identity. We use the Adam optimizer [28] with 400 epochs. The base learning rate is initialized to 1e−3 with a linear warm-up [17] in the first 50 epochs, then decayed to 1e−4 after 200 epochs, and further decayed to 1e−5 after 300 epochs.

**C. Comparison to State-of-the-art Methods**

Firstly, we evaluate the performance of CBDB-Net on the generic person Re-ID task. We compared our CBDB-Net against the many state-of-the-art approaches on Market-1501, DukeMTMC-reID and CUHK03, as shown in Tables I respectively. From the Table I we can observe that our CBDB-Net achieves competitive performance on these three generic person Re-ID datasets and outperforming most published approaches by a clear margin. Specifically, CBDB-Net obtains 94.3% Rank-1 and 85.0% mAP, which outperforms most existing methods on Market-1501 dataset. And then, we further introduce the Re-Ranking [71] into our CBDB-Net, i.e. CBDB-Net+Re-ranking. Here, the CBDB-Net+Re-ranking can achieve 96.6% Rank-1 and 93.0% mAP on the Market1501. On the DukeMTMC-reID dataset, our CBDB-Net obtains 87.7% Rank-1 and 74.3% mAP. The CBDB-Net+Re-ranking can achieve 91.2% Rank-1 and 87.9% mAP. The CUHK03 dataset is the most challenging dataset among the three generic person Re-ID datasets. Following the data setting in [75], [24], [76], [49], our CBDB-Net has clearly yielded good performance. On the CUHK03-Detected dataset, our CBDB-Net achieves 75.8% Rank-1 and 72.6% mAP; On the CUHK03-Labeled dataset, our CBDB-Net achieves 78.3% Rank-1 and 75.9% mAP. If we introduce the Re-ranking strategy into the CBDB-Net, the CBDB-Net+Re-ranking can further achieve 83.9% Rank-1 and 85.1% mAP on the CUHK03-Detected dataset, and achieves 86.5% Rank-1 and 87.8% mAP on the CUHK03-Labeled dataset respectively.

In our CBDB-Net, the Consecutive DropBlock Module can produce many incomplete feature tensors. These incomplete feature tensors push the deep model to capture a robust feature...
for person matching. Thus, we can regard the incomplete feature map as a kind of occluded or partial person feature in the person Re-ID task. So, we secondly try to evaluate the performance of our CBDB-Net in the occluded or partial person Re-ID task. We compared our CBDB-Net against the many approaches on Occluded DukeMTMC, Partial-REID, and Partial iLIDS, as shown in Tables II and Tables III respectively. On the Occluded DukeMTMC dataset, our CBDB-Net achieves the 50.9% Rank-1 and 38.9% mAP; On the Partial-REID dataset, our CBDB-Net achieves the 66.4% Rank-1 and 81.5% Rank-3; On the Partial iLIDS dataset, our CBDB-Net achieves the 68.4% Rank-1 and 80.9% Rank-3. From the Table II to Table III on the occluded or partial person Re-ID task, our CBDB-Net achieves the competitive results on the three datasets. Compared with PGFA [56], the performance of our CBDB-Net is a little bit worse. But, the structure of PGFA [56] is much more complex than that of our CBDB-Net. Besides, the PGFA [56] needs the additional human model to extract the human landmark to help the model locate the key local feature. In contrast, our CBDB-Net needn’t any auxiliary model, and is a kind of simple and efficient person Re-ID model.

In addition to the good performance in the two kind person Re-ID tasks, we believe that our CBDB-Net can be effective in other image retrieval tasks. We thirdly evaluate the performance of our CBDB-Net on the clothes retrieval task. As shown in Table IV, our CBDB-Net has achieved the best performance on the In-shop clothes retrieval dataset.

Compared with many human model-based person Re-ID methods, our CBDB-Net is a simple and effective person Re-ID model, and is also effective on the other image retrieval task. Overall, our observations endorse the superiority of CBDB-Net by combing “Consecutive Batch DropBlock Module” and “the proposed Elastic Loss”. Compared with other state-of-the-art approaches, our model is simple and effective, especially our testing model in Figure 5. In our CBDB-Net, we only extract a 512 dimension global descriptor to conduct the person matching task and gain a good performance.

### Table II

| Method       | Occluded-DukeMTMC Rank-1 | Occluded-DukeMTMC Rank-3 | Occluded-DukeMTMC Rank-10 | Occluded-DukeMTMC mAP |
|--------------|--------------------------|--------------------------|---------------------------|-----------------------|
| IDE+Triplet  | 8.1%                     | 17.0%                    | 22.0%                     | 5.0%                  |
| CBD          | 21.5%                    | 36.1%                    | 42.8%                     | 14.4%                 |
| Random Erasing | 28.8%                   | 44.6%                    | 51.0%                     | 20.2%                 |
| PCB          | 40.5%                    | 59.0%                    | 66.8%                     | 30.0%                 |
| HA-CNN       | 34.4%                    | 51.9%                    | 59.4%                     | 26.0%                 |
| Adver Occluded | 44.5%                   | -                        | -                         | 32.2%                 |
| Part Bilinear | 42.6%                    | 57.1%                    | 62.9%                     | 33.7%                 |
| FD-GAN       | 39.8%                    | -                        | -                         | -                     |
| DSR          | 40.8%                    | 58.2%                    | 65.2%                     | 30.4%                 |
| SFR          | 42.3%                    | 60.3%                    | 67.3%                     | 32.0%                 |
| PGFA         | 51.4%                    | 68.6%                    | 74.9%                     | 37.3%                 |
| CBDB-Net     | 50.9%                    | 66.0%                    | 74.2%                     | 38.9%                 |

### Table III

| Method       | Partial-REID Rank-1 | Partial-REID Rank-3 | Partial iLIDS Rank-1 | Partial iLIDS Rank-3 |
|--------------|---------------------|---------------------|----------------------|----------------------|
| IDE+Triplet  | 23.7%               | 27.3%               | 17.7%                | 26.1%                |
| CBD          | 37.3%               | 46.0%               | 21.0%                | 32.8%                |
| AMC+SWM      | 50.7%               | 70.0%               | 58.8%                | 67.2%                |
| DSR          | 56.9%               | 75.8%               | 63.9%                | 74.8%                |
| SFR          | 68.0%               | 80.0%               | 69.1%                | 80.9%                |
| CBDB-Net     | 66.7%               | 78.3%               | 68.4%                | 81.5%                |

### Table IV

| Method       | In-Shop Clothes Rank-1 | In-Shop Clothes Rank-10 | In-Shop Clothes Rank-20 |
|--------------|-------------------------|-------------------------|-------------------------|
| IDE+Triplet  | 33.0%                   | 73.0%                   | 76.0%                   |
| CBD          | 62.1%                   | 84.9%                   | 89.0%                   |
| DREML        | 78.4%                   | 93.7%                   | 95.8%                   |
| HTL          | 80.9%                   | 94.3%                   | 95.8%                   |
| A-BIER       | 83.1%                   | 95.1%                   | 96.9%                   |
| ABE-8        | 87.3%                   | 96.7%                   | 97.9%                   |
| BDB          | 89.1%                   | 96.3%                   | 97.6%                   |
| CBDB-Net     | 92.3% ± 0.3%            | 98.4% ± 0.2%            | 99.2% ± 0.2%            |

### Fig. 6

Visualization of attention maps from Baseline, CBDB-Net and six branches (m=6) of the CBDB-Net. Compared with the attention maps in column two, our CBDB-Net in column three can capture key feature. Besides, we can see column the attention information can be drop out by each Drop in “1-th” to “6-th”.

### D. Ablation Study of CBDB-Net

1) Effectiveness of the each strategy in the CBDB-Net:

In this subsection, we mainly discuss the effectiveness of each component in our CBDB-Net on four datasets: Market-1501, DukeMTMC-reID, CUHK03-Detected, and In-shop clothes retrieval datasets.

(I) In order to evidence the effectiveness of the proposed elastic loss, we use the IDE+triplet [70] as the baseline model. Based on the IDE+triplet loss, we replace triplet loss with the proposed elastic loss, i.e. “IDE+Elastic loss”. Compared with IDE+Triplet loss, our IDE+Elastic loss gains the obvious improvements on the four datasets over two measures. It effectively evidences the effectiveness of our elastic loss.
(II:) In order to only evidence the effectiveness of the Consecutive Batch DropBlock Module, based on the CBDB-Net in Figure 2, we replace the elastic loss with triplet loss, i.e. “CBDB-Net w/o Elastic loss” in the Table VI. And we use the model in Figure 5 as the test model. Compared with the IDE+Triplet loss model, our CBDB-Net does not contain the global branch. As shown in Table VI compared with IDE+Triplet loss, our CBDB-Net w/o Elastic loss also gains large improvements on the four datasets over the two measures.

(III:) In the CBDB-Net, based on the Consecutive Batch DropBlock Module, we can gain many incomplete feature maps. In our CBDB-Net, we append the additional ResNet block, i.e. “ResBlock” in Figure 2 and average pooling operation, on these incomplete feature tensors. To evidence the influence of the “ResBlock” in Figure 2, we directly append the average pooling operation on these incomplete feature tensors to conduct the person retrieval task, i.e. CBDB-Net w/o “ResBlock”. Compared with CBDB-Net w/o “ResBlock”, the “ResBlock” can effectively improve the performance of the CBDB-Net on the four datasets over two measures.

(IV:) If we combine the Consecutive Batch DropBlock Module and the proposed elastic loss, our CBDB-Net further achieves good performance on the four datasets. In the testing stage, we use the 512 dimension feature vector $f(x)$ to conduct the person matching task. Based on the CBDB-Net, we further concatenate the $\varphi(x)_i \in \mathbb{R}^{512}, i = 1, 2, \ldots, 6$ to conduct the person matching task in the testing stage. CBDB-Net only contains $m$ incomplete feature branches. IDE+CBDB-Net contains $m+1$ branches: $m$ incomplete feature branches and a global branch from IDE [60, 70].

In Table VI, compared with the CBDB-Net, the IDE+CBDB-Net only gains a slight improvement. It indicates that the additional global branch does not make an effective contribution to the improvement of the person Re-ID task. So, we select the CBDB-Net as our proposed novel person Re-ID model in this paper.

2) Various Versions of the CBDB-Net:

In this subsection, we mainly discuss the various versions of the CBDB-Net. Firstly, we discuss the influence of the parameter $m$ in our CBDB-Net. Here, we try to set $m = 4, 6, 8$ and show the performance of CBDB-Net on the four datasets over two measures. As we can see the Table VI the performance of CBDB-Net is stable under parameter $m = 4, 6, 8$.

Secondly, we mainly discuss the performance of different Consecutive Batch DropBlock strategies. In the Figure 7 there are two kind Consecutive Batch DropBlock strategies. The Figure 7 (a) is the Consecutive Batch DropBlock strategy in the Figure 2 or our CBDB-Net. The Figure 7 (b) is the other Consecutive Batch DropBlock strategy. In the Figure 7 (b), there is an overlap area between the dropped patch by DropPatch – (i) and dropped patch by DropPatch – (i + 1). In this subsection, we only discuss that the overlap is one pixel between the DropPatch – (i) and the DropPatch – (i + 1) on the feature map $T(x) \in \mathbb{R}^{24 \times 8 \times 2048}$, i.e. $m = 6^*$, $m = 8^*$ and $m = 11^*$ in the Table VI. Based on the drop operation in Figure 7 (b), we can further gain more incomplete features to train the person Re-ID model. As shown in Table VI because the performance on Market-1501 and DukeMTMC-Re-ID has been saturated lately. The Consecutive Batch DropBlock strat-
Fig. 7. Various versions of the Consecutive Batch DropBlock strategy. The figure (a) is the Consecutive Batch DropBlock strategy in Figure 2. There is no overlap areas between the dropped patch by $\text{DropPatch} - (i)$ and dropped patch by $\text{DropPatch} - (i + 1)$. The figure (b) is the other Consecutive Batch DropBlock strategy. There is an overlap area between the dropped patch by $\text{DropPatch} - (i)$ and dropped patch by $\text{DropPatch} - (i + 1)$. In this paper, we only discuss the case that the overlap is one pixel on the feature map $T(x)$.

TABLE VII
RESULTS OF CBDB-NET ON THE CUHK03-DETECTED DATASETS UNDER DIFFERENT NUMBER OF BRANCHES IN THE CBDB-NET WITHOUT ELASTIC LOSS. $m' = i$ INDICATES THAT WE DROP THE BRANCHES FROM $(i + 1) - th$ TO $(6) - th$ IN THE CBDB-NET.

| Method            | CUHK03-Detected |
|-------------------|-----------------|
|                   | Rank-1 | mAP   |
| $m' = 1$          | 56.2%  | 51.9% |
| $m' = 2$          | 68.0%  | 64.0% |
| $m' = 3$          | 72.0%  | 68.7% |
| $m' = 4$          | 74.2%  | 70.6% |
| $m' = 5$          | 75.2%  | 72.0% |
| $m' = 6$          | 75.8%  | 72.6% |

Elasticity in Figure 7(b) can not gain the moderate improvement. In contrast, the Consecutive Batch DropBlock strategy in Figure 7(b) can further gain effective improvements on the CUHK03-Detected, CUHK03-Labeled, and Clothes datasets. Based on the Consecutive Batch DropBlock strategy in Figure 7(b), we call it as the CBDB-Net.

Thirdly, based on the $m = 6$, CBDB-Net has 6 branches. Here, we simply discuss the influence of branches number on the CUHK03-Detected dataset. As shown in Table VII $m' = i$ indicates that we drop the branches from $(i + 1) - th$ to $(6) - th$ in the CBDB-Net. In the Table VII as the number of branches increases, the performance gets better. It indicates that more incomplete feature descriptors can effectively improve the performance of the person Re-ID model. Here, $m' = 6$ is our CBDB-Net.

3) Comparison with Various Dropout Strategies: In this subsection, we discuss the performance of various Dropout strategies on the CUHK03-Detected dataset. (I:) Dropout [46] randomly drops the neural node of input feature maps or vector, which is regarded as a kind of regularization strategy to prevent overfitting of deep models. (II:) Compared with Dropout, SpatialDropout [50] further randomly zeroes whole channels of the input feature tensors. And the channels of feature tensors to zero-out are randomized. (III:) Based on the SpatialDropout, Batch Dropout randomly zeroes the whole channels of the input feature tensors within the whole batch. However, the Batch Drop can not drop a large contiguous area. (IV:) So, the DropBlock [76] strategy randomly drops a contiguous region on the feature maps. (V:) The Batch DropBlock further randomly drops the same region for every input tensor within a batch. Different from the Batch DropBlock, our Consecutive Batch DropBlock can drop the same region for every input tensor within the whole training set and product multiple incomplete feature tensors. As shown in Table VIII compared with these Dropout strategies, our Consecutive Batch DropBlock strategy achieves the best performance on the CUHK03-Detected dataset over two measures.

V. Conclusion

In this paper, we propose a novel person Re-ID model, Consecutive Batch DropBlock Network (CBDB-Net), to improve the ability of the person Re-ID model on capturing the robust and high quality feature descriptor for person matching task. Specifically, firstly Consecutive Batch DropBlock Module is proposed to exploit multiple incomplete descriptors, which can effectively push the person Re-ID model to capture the robust feature descriptor. Secondly, the Elastic Loss is designed to adaptively mine and balance the hard sample pairs in the training process. Extensive experiments show that our CBDBNet achieves the competitive performance on three generic person Re-ID datasets, three occlusion person Re-ID datasets, and the generic image retrieval task.

REFERENCES

[1] Tianlong Chen, Shaojin Ding, Jingyi Xie, Ye Yuan, Wuyang Chen, Yang Yang, Zhou Ren, and Zhangyang Wang. Abd-net: Attentive but diverse person re-identification. In IEEE International Conference on Computer Vision, 2018.
[2] De Cheng, Yihong Gong, Sanping Zhou, Jinjun Wang, and Nanning Zheng. Person re-identification by multi-channel parts-based cnn with improved triplet loss function. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
[3] Tay Chiat-Pin, Roy Sharmil, and Yap Kim-Hui. Aanet: Attribute attention network for person re-identifications. In IEEE Conference on Computer Vision and Pattern Recognition, 2019.
[4] Yeong-Jun Cho and Kuk-Jin Yoon. Improving person reidentification via pose-aware multi-shot matching. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
[5] Song Chunfeng, Huang Yan, Wang Liang, and Ouyang Wanli. Mask-guided contrastive attention model for person re-identification. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.
[6] Wanti Ouyang Chunfeng Song, Yan Huang and Liang Wang. Mask-guided contrastive attention model for person re-identification. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 1179–1188, 2018.
[7] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In CVPR, 2009.
[8] Weijian Deng, Zheng Liang, Guoliang Kang, Yang Yi, and Jianbin Jiao. Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 994–1003, 2018.
[9] Terrance DeVries and Graham W Taylor. Improved regularization of convolutional neural networks with cutout. In arXiv:1708.04552, 2017.
[10] Fasturk Faghi, David J. Fleet, Jamie Ryan Kiros, and Sanja Fidler. Vse++: Improved visual-semantic embeddings. In BMVC, 2018.
[11] Liu Fangyu and Guang Lei. View confusion feature learning for person re-identification. In IEEE International Conference on Computer Vision, 2019.
[12] Zhanxiang Fung, Jianhuang Lai, and Xiaohua Xie. Learning view-specific deep networks for person re-identification. IEEE Transactions on Image Processing, 27(7):3472 – 3483, 2018.
[13] Weifeng Ge, Weilin Huang, Dengke Dong, and Matthew R Scott. Deep metric learning with hierarchical triplet loss. In ECCV, 2018.
[14] Yixiao Ge, Zhouan Li, Haiyu Zhao, Guojin Yin, Shaui Yi, and Xiaogang Wang. Fd-gan: Pose-guided feature distillation gan for robust person re-identification. In NIPS, 2018.
[15] Mengyue Geng, Yaowei Wang, and Yonghong Tian. Deep transfer learning for person re-identification. 2016.
[16] Golnaz Ghiasi, Tsung-Yi Lin, and Quoc V Le. Dropblock: A regularization method for convolutional networks. In arXiv:1810.12890, 2018.
[17] Priya Goyal, Pieter Doll, Ross Girshick, Pieter Noord-huis, Lukasz Wesolowski, Aapo Kyrola, Andrew Tulloch, Yangqing Jia, and Kaiming He. Accurate, large minibatch sgd: Training imagenet in 1 hour. In arXiv:1706.02677, 2017.
[18] Jianyuan Guo, Yuhui Yuan, Lang Huang, Chao Zhang, Jin-Ge Yao, and Priya Goyal. Partial face recognition: Partial biometric patterns. In IEEE Conference on Computer Vision and Pattern Recognition, pages 10, 2013.
[19] Terrance DeVries and Graham W Taylor. Improved regularization of convolutional neural networks with cutout. In arXiv:1708.04552, 2017.
[20] Fartash Faghi, David J. Fleet, Jamie Ryan Kiros, and Sanja Fidler. Vse++: Improved visual-semantic embeddings. In BMVC, 2018.
[21] Liu Fangyu and Guang Lei. View confusion feature learning for person re-identification. In IEEE International Conference on Computer Vision, 2019.
[22] Zhanxiang Fung, Jianhuang Lai, and Xiaohua Xie. Learning view-specific deep networks for person re-identification. IEEE Transactions on Image Processing, 27(7):3472 – 3483, 2018.
[23] Weifeng Ge, Weilin Huang, Dengke Dong, and Matthew R Scott. Deep metric learning with hierarchical triplet loss. In ECCV, 2018.
[24] Yixiao Ge, Zhouan Li, Haiyu Zhao, Guojin Yin, Shaui Yi, and Xiaogang Wang. Fd-gan: Pose-guided feature distillation gan for robust person re-identification. In NIPS, 2018.
[25] Mengyue Geng, Yaowei Wang, and Yonghong Tian. Deep transfer learning for person re-identification. 2016.
[26] Golnaz Ghiasi, Tsung-Yi Lin, and Quoc V Le. Dropblock: A regularization method for convolutional networks. In arXiv:1810.12890, 2018.
[27] Priya Goyal, Pieter Doll, Ross Girshick, Pieter Noord-huis, Lukasz Wesolowski, Aapo Kyrola, Andrew Tulloch, Yangqing Jia, and Kaiming He. Accurate, large minibatch sgd: Training imagenet in 1 hour. In arXiv:1706.02677, 2017.
[28] Jianyuan Guo, Yuhui Yuan, Lang Huang, Chao Zhang, Jin-Ge Yao, and Priya Goyal. Partial face recognition: Partial biometric patterns. In IEEE Conference on Computer Vision and Pattern Recognition, pages 10, 2013.
[29] Terrance DeVries and Graham W Taylor. Improved regularization of convolutional neural networks with cutout. In arXiv:1708.04552, 2017.
[30] Fartash Faghi, David J. Fleet, Jamie Ryan Kiros, and Sanja Fidler. Vse++: Improved visual-semantic embeddings. In BMVC, 2018.
[31] Liu Fangyu and Guang Lei. View confusion feature learning for person re-identification. In IEEE International Conference on Computer Vision, 2019.
[32] Zhanxiang Fung, Jianhuang Lai, and Xiaohua Xie. Learning view-specific deep networks for person re-identification. IEEE Transactions on Image Processing, 27(7):3472 – 3483, 2018.
activation maps augmentation for person re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2019.

[58] Bryan (Ning) Xia, Yuan Gong, Yizhe Zhang, and Christian Poellabauer. Second-order non-local attention networks for person re-identification. In *IEEE International Conference on Computer Vision*, 2019.

[59] Chang Xiaobin, M Hospedales Timothy, and Xiang Tao. Multi-level factorisation net for person re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2018.

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

[61] Hong Xuan, Richard Souvenir, and Robert Pless. Deep randomized ensembles for metric learning. In *ECCV*, 2018.

[62] Hantao Yao, Shiliang Zhang, Yongdong Zhang, Jintoa Li, and Qi Tian. Deep representation learning with part loss for person re-identification. In *arXiv:1711.08106*, 2017.

[63] Qian Yu, Xiaobin Chang, Yi-Zhe Song, Tao Xiang, and Timothy M Hospedales. The devil is in the middle: Exploiting mid-level representations for cross-domain instance matching. In *arXiv:1711.08106*, 2017.

[64] Yuliu Yuan, Kuiyuan Yang, and Chao Zhang. Hard-aware deeply cascaded embedding. In *ICCV*, 2017.

[65] Haiyu Zhao, Maoqing Tian, Shuyang Sun, Jing Shao, Junjie Yan, Shuai Yi, Xiaogang Wang, and Xiaoou Tang. Spindle net: Person re-identification with human body region guided feature decomposition and fusion. In *CVPR*, 2017.

[66] Liming Zhao, Xi Li, Yueting Zhuang, and Jingdong Wang. Deeply-learned part-aligned representations for person re-identification. In *IEEE International Conference on Computer Vision (ICCV)*, 2017.

[67] Zheng Zhedong, Zheng Liang, and Yang Yi. Pedestrian alignment network for large-scale person re-identification. *IEEE Transactions on Circuits and Systems for Video Technology*, 29(10):3037–3045, 2019.

[68] Yi Yang Zhedong Zheng, Liang Zheng. Unlabeled samples generated by gan improve the person re-identification baseline in vitro. In *International Conference on Computer Vision (ICCV)*, 2017.

[69] Liang Zheng, Liyue Shen, Lu Tian, Shengjii Wang, Jingdong Wang, and Qi Tian. Scalable person re-identification: A benchmark. In *IEEE International Conference on Computer Vision (ICCV)*, pages 1116–1124, 2015.

[70] Liang Zheng, Yi Yang, and Alexander G. Hauptmann. Person re-identification: Past, present and future. 2016.

[71] Zhun Zhong, Liang Zheng, Donglin Cao, and Shaozi Li. Re-ranking person re-identification with k-reciprocal encoding. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[72] Zhun Zhong, Liang Zheng, Guoliang Kang, Shaozi Li, and Yi Yang. Random erasing data augmentation. In *arXiv:1708.04896*, 2017.

[73] Zhun Zhong, Liang Zheng, Guoliang Kang, Shaozi Li, and Yi Yang. Random erasing data augmentation. In *arXiv:1708.04896*, 2017.

[74] Zhun Zhong, Liang Zheng, Zhedong Zheng, Shaozi Li, and Yi Yang. CamStyle: A novel data augmentation method for person re-identification. *IEEE Transactions on Image Processing*, 28(3):1176–1190, mar 2019.

[75] Zhun Zhong, Zhedong Zheng, Kang Guoliang, Li Shaozi, and Yang Yi. Random erasing data augmentation. In *The National Conference on Artificial Intelligence, 2020.*

[76] Dai Zuozhuo, Chen Mingjiang, Gu Xiaodong, Zhu Siyu, and Tan Ping. Batch dropblock network for person re-identification and beyond. In *IEEE International Conference on Computer Vision*, 2019.