ABSTRACT Despite the promising progress made in recent years, person re-identification (Re-ID) remains a challenging task due to the intra-class variations. Most of the current studies used the traditional Softmax loss for solutions, but its discriminative capability encounters a bottleneck. Therefore, how to improve person Re-ID performance is still a challenging task. To address this problem, we proposed a novel loss function, namely additive distance constraint with similar labels loss (ADCSLL). Specifically, we reformulated the Softmax loss by adding a distance constraint to the ground truth label, based on which similar labels were introduced to enhance the learned features to be much more stable and centralized. Experimental evaluations were conducted on two popular datasets (Market-1501 and DukeMTMC-reID) to examine the effectiveness of our proposed method. The results showed that our proposed ADCSLL was more discriminative than most of the other compared state-of-the-art methods. The rank-1 accuracy and the mAP on Market-1501 were 95.0% and 87.0%, respectively. The numbers were 88.6% and 77.2% on DukeMTMC-reID, respectively.

INDEX TERMS Intelligent safety systems, person re-identification, deep learning, similar labels, distance constraint.

I. INTRODUCTION

Person re-identification (Re-ID) aims to retrieve a given person from non-overlapping cameras, which is widely applied in the field of intelligent security, intelligent retail and intelligent video surveillance [1]. However, the intra-class variations, caused by large variations in poses, viewpoints, illuminations, background environments and occlusion, are challenging this task [2]–[4]. Moreover, when the number of retrieving persons increases, the inter-class similarities also make the task more complicated [5].

In the last decade, deep-learning techniques have significantly boosted the state-of-the-art performance [6]–[9]. Recent deep learning solutions for person Re-ID have demonstrated good retrieval performance [10]–[12]. The current mainstream strategy is to combine the local features or to utilize the semantic information from human pose or body part [13]–[15]. As for the network architecture, a number of deep-learning methods chose Resnet50 [16] pre-trained on the image identification dataset ImageNet [17] as the baseline network, and further got better performance by considering both the local and global information [18]–[22]. Among these studies, existing part-based approaches used human structural information such as body part or human pose to further improve Re-ID performance. In the method of [10] and [15], input images or feature maps were all split horizontally to take advantage of local spatial cues. The methods in [11], [16] and [23] made use of human pose to learn local features. In recent years, attention mechanisms were also adopted to capture attentive regions for Re-ID enhancement [14], [24], [25]. However, these part-based or pose-based approaches need additional prior knowledge such as the pre-trained human
pose estimation model to increase the method complexity and computational cost, or have high requirement on image normalization (e.g., the alignment of human body parts) which is easily to cause low-precision problems. The approaches based on attention mechanisms will also make the model more complex to highly increase the computational cost. Meanwhile, these deep learning studies usually design neural network structures or concatenate multi-branch features to capture and to focus on the salient features, generally bringing too much extra calculation, whereas the industry communities prefer simple but effective methods [26]–[28].

Loss function design plays a significant role in feature learning for person Re-ID. The current loss functions can be categorized into two types: metric loss [25]–[30] and identification loss [23], [31]. See Fig. 1 for the differences between these two types. With the same purpose to better complete the Re-ID task, the training targets and the realizing ways among different loss functions are quite different. Intuitively, simultaneously enlarging the intra-class convergence and inter-class discrimination can be more conducive to learn representative features. Inspired by this idea, the metric loss was proposed to enforce input images of a unique person to be mapped to nearby points in the feature space and of different person to be mapped far apart. However, when the number of training samples increases, the number of training pairs will exponentially increase and therefore cost much more training time. Cross-entropy loss together with Softmax is an alternative choice of identification loss function, which considers the identification task as a multi-class classification problem. This method treats each identity as a separate category, and then minimizes the Softmax loss. Despite its popularity, current Softmax loss only concentrates on the logits of ground truth label and neglects the effectiveness of other labels.

To alleviate these shortcomings, The loss function we proposed in this study is inspired by the typical Softmax-based loss functions in face recognition [32]–[35] and label smooth (a regularization mechanism) [36]. Person Re-ID is similar to face recognition where significant progresses based on deep learning have been made for improvement. To enhance the discriminative capability of the original Softmax loss, [32] proposed a large margin Softmax (L-Softmax) by adding angular constraints to each identity to improve feature discrimination. Angular Softmax (A-Softmax) [33] further improved L-Softmax by normalizing the weights. The proposed ArcFace in [35] directly maximized the decision boundary in angular (arc) space based on the L2 regularization normalized weights and features. CosFace [34] proposed a novel algorithm large margin cosine loss, which takes the normalized features as inputs to learn highly discriminative features by maximizing the inter-class cosine margin. All of these methods generalized the Softmax loss to a more general large margin Softmax loss in terms of angular similarity. Disappointingly, these methods have not been well examined in person Re-ID tasks.

Inspired by such ideas, we employed a distance constraint in the original Softmax loss to enhance the feature discriminative capability in this study. However, Softmax loss is more suitable for classification tasks on known classes, while Re-ID is an unknown-label recognition retrieval task. To solve this problem, we designed a new loss function by considering the following aspects: (1) Softmax loss only gives consideration to the ground truth label and do not consider the loss information of other labels. Therefore, we further extracted other labels with higher logits (namely similar labels) and introduced them to Softmax loss design. Similar labels are expected to help improve the network to enforce more stable and more centralized features of other labels. (2) As the addition of similar labels distracts the attention of the network, the distribution of the ground truth label decreases. Therefore, we imposed a distance constraint to the ground truth label so as to improve the difficulty of the training target, which can ultimately increase the logit of the ground truth label. The ground truth label refers to

| Loss type       | Input                  | Mechanism                  | Target                        |
|-----------------|------------------------|----------------------------|-------------------------------|
| Identification loss | one image              |                            |                               |
| Metric loss     | a pair of images       |                            |                               |
| Triplet loss    | a triplet of images    |                            |                               |

**FIGURE 1. Differences between identification loss and metric loss.**
the label that has the same person identity with the current training sample, while the rest of the labels are defined as other labels. However, some labels belonging to other labels are considered to have similarities with the training samples, so that they tend to get superior logits. We define such labels as similar labels.

Label smooth strategy [36] was used to prevent models from overfitting and to improve models’ generalization ability by considering the weights of all identification labels (ground truth label and other labels) when calculating losses. No matter how the other labels perform during training, they all get attention from the model, leading to insufficient bias of similar labels and needless bias of the others. Thus, we improved the loss function design by only calculating the losses of similar labels and removed the unnecessary part.

To leverage the complementary advantages of different loss functions, many person Re-ID approaches optimized their networks with multi losses [1], [18], [22], [31], [37], [38]. The work presented in [39] combined both identification loss and contrastive loss together to design the loss function. Identification loss and metric loss were also combined to get a better performance improvement. The main contributions of this work can be summarized as follows.

1. Our proposed method designs a novel identification loss function on a single model without extra calculation. The overview of our proposed architecture is presented in Fig. 2. Generally, the proposed ADCSLL consists of two parts, the additive distance constraint loss and the similar labels loss. ResNet50 [16] with pretrained parameters on ImageNet was considered as the backbone network of our developed method. We changed the last stride of ResNet50 from 2 to 1 and then connected a global average pooling layer, a batch normalization layer, and a classifier. The details of the proposed approach are described as follows.

2. **A. THE SOFTMAX LOSS**

Given a training sample \( x \), the Softmax loss \( L_s \) is calculated as,

\[
L_s = -\sum_{i}^{N} \log (p_i) = -\sum_{i}^{N} \log \left( \frac{e^{f_i}}{\sum_{j=1}^{N} e^{f_j}} \right)
\]

where \( y \) is the ground truth label; \( p_i \) is the probability of each label \( i \in \{1, \cdots, N\} \); \( f_i \) is the un-normalized log-probability (logit); \( N \) is the number of training labels. Considering \( q_i \) as the distribution of the \( i \)-th label together with the ground truth label \( y \) so that \( q_i (y) = 1 \) and \( q_i (i) = 0 \) for all \( i \neq y \), the Softmax loss \( L_s \) can be re-written as,

\[
L_s = -\sum_{i}^{N} q_i \cdot \log (p_i),
\]

with

\[
p_i = \frac{e^{w^Tx_i}}{\sum_{j=1}^{N} e^{w^Tx_j}},
\]

\[
\begin{cases}
q_i = 1, & i = y \\
q_i = 0, & i \neq y
\end{cases}
\]

However, (2) ignores the distribution of other labels so that the generalization performance needs to be improved. Consequently, we introduced similar labels to the original Softmax loss and employed a constraint to the ground truth label to design a novel loss function.

II. PROPOSED APPROACH

To alleviate the above-mentioned problems, we proposed a generalized additive distance constraint with similar labels loss (ADCSLL) function to acquire superior retrieval performance by intrinsically learning highly discriminative features without extra calculation. The overview of our proposed architecture is presented in Fig. 2. Generally, the proposed ADCSLL consists of two parts, the additive distance constraint loss and the similar labels loss. ResNet50 [16] with pretrained parameters on ImageNet was considered as the backbone network of our developed method. We changed the last stride of ResNet50 from 2 to 1 and then connected a global average pooling layer, a batch normalization layer, and a classifier. The details of the proposed approach are described as follows.

A. THE SOFTMAX LOSS

Given a training sample \( x \), the Softmax loss \( L_s \) is calculated as,

\[
L_s = -\sum_{i}^{N} \log (p_i) = -\sum_{i}^{N} \log \left( \frac{e^{f_i}}{\sum_{j=1}^{N} e^{f_j}} \right)
\]

where \( y \) is the ground truth label; \( p_i \) is the probability of each label \( i \in \{1, \cdots, N\} \); \( f_i \) is the un-normalized log-probability (logit); \( N \) is the number of training labels. Considering \( q_i \) as the distribution of the \( i \)-th label together with the ground truth label \( y \) so that \( q_i (y) = 1 \) and \( q_i (i) = 0 \) for all \( i \neq y \), the Softmax loss \( L_s \) can be re-written as,

\[
L_s = -\sum_{i}^{N} q_i \cdot \log (p_i),
\]

with

\[
p_i = \frac{e^{w^Tx_i}}{\sum_{j=1}^{N} e^{w^Tx_j}},
\]

\[
\begin{cases}
q_i = 1, & i = y \\
q_i = 0, & i \neq y
\end{cases}
\]
B. SIMILAR LABELS LOSS

Softmax loss only maximizes the ground truth label’s distribution \( q_i(y) = 1 \) but ignores the distribution of other labels, leading to training convergence monotony and limitation. Label smooth shares distribution indiscriminately over other labels as \( q_i(i) = 1/N \) for all \( i \neq y \). Compared with Softmax loss, the mechanism of label smooth is to replace the distribution of

\[
q_i = \begin{cases} 
1, & i = y \\
0, & i \neq y 
\end{cases}
\]

with

\[
q_i = \begin{cases} 
1 - \varepsilon, & i = y \\
\varepsilon / N, & i \neq y 
\end{cases}
\]

where \( \varepsilon = 0.1 \) is a preset constant balancing the distribution of the ground truth label and other labels. Therefore, label smooth \( L_{LS} \) can be formulated as follows,

\[
L_{LS} = - \sum_i N q_i \cdot \log(p_i),
\]

with

\[
p_i = \frac{e^{w^T x}}{\sum_j N e^{w^T x}},
\]

\[
q_i = \begin{cases} 
1 - \varepsilon, & i = y \\
\varepsilon / N, & i \neq y 
\end{cases}
\]

Furthermore, it is assumed that those labels with higher logits are more meaningful than all the other labels for model training. Therefore, we introduced similar labels to the original Softmax loss and the corresponding similar labels loss \( L_{SL} \) is reformulated as,

\[
L_{SL} = - \sum_i N q_i \cdot \log(p_i),
\]

with

\[
p_i = \frac{e^{w^T x}}{\sum_j N e^{w^T x}},
\]

\[
q_i = \begin{cases} 
1, & i = y \\
1 - \varepsilon, & i = y \\
\varepsilon / N, & i \in S_{N_e} \\
0, & \text{otherwise}
\end{cases}
\]

which is a mixture of the ground truth label’s distribution \( q_i(y) = 1 \) and the similar labels’ distribution \( q_i(k) = 1/N, \) with weights \( 1 - \varepsilon \) and \( \varepsilon \), respectively. \( N_e \) is the number of similar labels. The similar labels set \( S_{N_e} \) is composed by the top \( N_e \) labels according to the descending sorted logits. \( N_e \) varies according to the number of labels (identities) in the dataset. In this paper, we set \( \varepsilon \) and \( N_e \) as 0.3 and 100 for both the examined datasets (Market-1501 and DukeMTMC-reID).

The distribution evolution from Softmax loss to the similar labels loss is shown in Fig. 3. This can be considered as the distribution \( q_i \) of label \( i \) evolution as follows: In Fig. 3(a), the distribution of ground truth label is \( q_i(y) = 1 \); Then, in Fig. 3(b), label smooth replaces \( q_i(y) = 1 \) with weight \( 1 - \varepsilon \), and shares the distribution \( q_i(i) = 1/N \) with weight \( \varepsilon \) to the other labels; In Fig. 3(c), we only share the distribution \( q_i(i) = 1/N \) with weight \( \varepsilon \) to the similar labels and set the distribution to 0 for the rest of other labels.

C. ADDITIVE DISTANCE CONSTRAINT LOSS

Face recognition is enlightening to Re-ID because of their similarities. Recent approaches on face recognition designed new loss functions for deep face recognition. Specifically, [32], [33], [34] and [40] maximized the decision boundary which are the scaling parameter. In [34], they set \( m \) as 64.

Similar to (2), \( L_{mc} \) can be re-written as,

\[
L_{mc} = - \sum_i N q_i \cdot \log(p_i),
\]

with

\[
p_i = \frac{e^{(\cos(\theta_i) - m)}}{\sum_j N e^{(\cos(\theta_j) - m)} + \sum_j N e^{(\cos(\theta_i) - m)}},
\]

where \( \theta_i \) is the angle between \( w^T \) and \( x \). \( m \) is a fixed parameter introduced to control the magnitude of the cosine margin. \( s \) is the scaling parameter. In [34], they set \( m \) as 0.35 and \( s \) as 64.

FIGURE 3. The distribution evolution from Softmax loss to the similar labels loss.
We had experimented with $L_{lmc}$ to handle the Re-ID task but achieved an unsatisfactory result, which indicates that normalizing weights and features as well as enlarging angular margin are not useful enough to enhance the discriminative power of the learned features. It is probably because human faces have the same topological structure but human poses significantly differ from each other between the data samples. The significant variation would make the loss function for face recognition not applicable in person Re-ID tasks. To solve this problem, we further introduced a distance constraint to the ground truth label without features and weights normalization, reinforcing the discrimination of learned features by pursuing a more stringent decision boundary. The following simple binary-class example is shown to explain our improvement.

Given a learned feature $x$ from class 1, the Softmax loss is to force $w_1^T x > w_2^T x$ to classify $x$ correctly. The decision boundary between class 1 and class 2 is $w_1^T x = w_2^T x$. To develop a more discriminative classifier, a more stringent decision boundary than $w_1^T x = w_2^T x$ is needed. We employed a distance constraint $c$ to further require $w_1^T x - |w_1^T x| \cdot c > w_2^T x - |w_2^T x| \cdot c$. Since $w_1^T x - |w_1^T x| \cdot c$ is smaller than $w_2^T x$, the following inequality holds,

$$w_1^T x \geq w_1^T x - |w_1^T x| \cdot c > w_2^T x - |w_2^T x| \cdot c \quad (9)$$

Similarly, the above analysis can be well generalized into multi-class identification tasks. The new Re-ID criteria is more stringent than the original decision boundary $w_1^T x = w_2^T x$. By directly reformulating the original Softmax, we defined the additive distance constraint loss $L_{ADC}$ as,

$$L_{ADC} = - \sum_i^N q_i \log (p_i), \quad (10)$$

with

$$\begin{cases} q_i = 1, \quad p_i = \frac{e^{w_i^T x - |w_i^T x| \cdot c}}{\sum_{j=1}^N e^{w_j^T x}}, \quad i = y \\ q_i = 0, \quad p_i = \frac{\sum_{j=1}^N e^{w_j^T x}}{\sum_{j=1}^N e^{w_j^T x}}, \quad i \neq y \end{cases}$$

where $c$ is the preset distance constraint to determine the strength of getting closer to the ground truth label. The parameter $c$ should be set to a reasonable large number to effectively boost the learning abilities of highly discriminative features. Generally, $c = 0$ leads to an insufficient convergence. Decreasing the logit of ground truth label by increasing the distance constraint $c$ is able to increase the training difficulty and to improve the Re-ID performance, but decreasing too much may cause training divergence and convergence failure because the distance constraint would become stricter and more difficult to be satisfied. Besides, the distance constraint with an overlarge $c$ forces the training process to be more sensitive to noise. Comprehensively considering these factors, we set $c = 0.4$.

### D. ADDITIVE DISTANCE CONSTRAINT WITH SIMILAR LABELS LOSS (ADCSLL)

By combining the additive distance constraint loss and the similar labels loss, we proposed a new loss function ADCSLL, together with the network structure presented in Fig. 1, to solve the person Re-ID problem. Simultaneously considering (6) and (10), our proposed ADCSLL is defined as,

$$L_{ADCSL} = - \sum_i^N q_i \log (p_i),$$

$$\begin{cases} q_i = 1 - \epsilon, \quad p_i = \frac{e^{w_i^T x - |w_i^T x| \cdot c}}{e^{w_i^T x - |w_i^T x| \cdot c} + \sum_{j \neq i}^N e^{w_j^T x}}, \quad i = y \\ q_i = \epsilon \frac{1}{N_{sl}}, \quad p_i = \frac{\sum_{j=1}^N e^{w_j^T x}}{\sum_{j=1}^N e^{w_j^T x}}, \quad i \in S_{N_{sl}} \\ q_i = 0, \quad p_i = \frac{\sum_{j=1}^N e^{w_j^T x}}{\sum_{j=1}^N e^{w_j^T x}}, \quad \text{otherwise} \end{cases}$$

where $L_{ADCSL}$ is our proposed ADCSLL function, which can be re-written as,

$$L_{ADCSL} = - \left( 1 - \epsilon \right) \log \left( \frac{e^{w_i^T x - |w_i^T x| \cdot c}}{e^{w_i^T x - |w_i^T x| \cdot c} + \sum_{j \neq i}^N e^{w_j^T x}} \right) + \frac{\epsilon}{N_{sl}} \sum_i^N \log \left( \frac{\sum_{j=1}^N e^{w_j^T x}}{\sum_{j=1}^N e^{w_j^T x}} \right) \quad (11)$$

### III. EXPERIMENTS

To validate the performance of our proposed method, we tested it on two popular person Re-ID datasets, Market-1501 and DukeMTMC-reID. It is worth noting that all our results were conducted in a single-query setting without using re-ranking algorithm to improve the mean average precision (mAP) in all the experiments.

### A. DATASETS

1) Market-1501

The Market-1501 dataset [41] was constructed and published in 2015. It consists of image samples from 6 cameras. In this dataset, a total of 32668 images from 1501 identities with annotated bounding boxes detected for both training and testing were collected. The dataset was divided into a training set with 12936 images of 751 identities and a test set of 750 identities with 3368 query images and 19732 gallery images.

2) DukeMTMC-reID

DukeMTMC-reID is an image-based subset of the DukeMTMC [42] which is used for person Re-ID. It consists of 36411 images from 1812 persons in 8 cameras.
TABLE 1. Comparison results with the state-of-the-art methods on Market-1501 and DukeMTMC-reID datasets.

| Method     | Market-1501 | DukeMTMC-reID |
|------------|-------------|---------------|
|            | Rank-1     | mAP           | Rank-1 | mAP |
| AlignedReID [45]        | 90.6%    | 77.7%         | 81.2%  | 67.4%  |
| IDE [39]            | 79.5%    | 59.5%         | -      | -     |
| SVDrNet [46]           | 82.3%    | 62.1%         | 76.7%  | 56.8%  |
| TriNet [30]           | 84.9%    | 69.1%         | -      | -     |
| Pyramid [47]          | 92.8%    | 82.1%         | -      | -     |
| AWTL [48]            | 89.5%    | 75.7%         | 79.8%  | 63.4%  |
| ABD-Net [49]          | 95.6%    | 88.3%         | 89.0%  | 78.6%  |
| DSA-reID [50]        | 95.7%    | 87.6%         | 86.2%  | 74.3%  |
| Baseline [51]         | 93.9%    | 85.1%         | 86.2%  | 75.0%  |
| ADCSLL (ours)    | 94.8%    | 86.4%         | 87.5%  | 76.8%  |
| baseline + triplet loss [51] | 94.0% | 85.8% | 86.4% | 76.0% |
| ADCSLL + triplet loss (ours) | 95.0% | 87.0% | 88.6% | 77.2% |

It randomly selected 16522 training images from 702 identities, 2228 query images from the other 702 identities and 17661 gallery images.

B. IMPLEMENTATION DETAILS

Our model was implemented on Pytorch framework. All images were resized into a resolution of 256 × 128, and image augmentation techniques including random crop, random erasing and horizontal flip (with a probability of 0.5) [43] were used to expand the datasets. We selected a batch size of 64 for each iteration and trained the proposed model for 120 epochs. Adam [44] was used as an optimizer in our optimization, where the weight decay factor was set as 0.0005. The initial learning rate was linearly increasing from $2 \times 10^{-5}$ to $2 \times 10^{-4}$ in the first 10 epochs. Then, the learning rate was decreased to $2 \times 10^{-5}$ and $2 \times 10^{-6}$ at the 40th epoch and the 70th epoch, respectively. In the training phase, ADCSLL enforced the model to learn more salient features with larger distance constraint and similar labels. In the testing phase, the highly discriminative features were calculated to examine the effectiveness of our proposed method.

IV. RESULTS AND DISCUSSION

A. PERSON RE-ID PERFORMANCE

The performance of our proposed person Re-ID method is compared with the state-of-the-art methods on both Market-1501 and DukeMTMC-reID datasets. The employed comparison methods include AlignedReID [45], IDE (ID-discriminative embedding) [39], SVDrNet (singular vector decomposition net) [46], TriNet (triplet net) [30], Pyramid [47], AWTL (adaptive weighted triplet loss) [48], ABD-Net (Attentive but Diverse Net) [49], DSA-reID (Densely Semantically Aligned reID) [50], the baseline method in [51], and the baseline method together with triplet loss [51].

The comparison results are summarized in Table 1. It should be noted that the baseline Pyramid method used one feature for training for fair comparison. The results show that our proposed ADCSLL achieves competitive results compared to most of the existing methods except ABD-Net and DSA-reID. The results of our proposed method is better than DSA-reID on DukeMTMC-reID, but not superior to ABD-Net. Different from our model which only uses global features, both ABD-Net and DSA-reID utilize global features together with local features to build a network model with two branches, which indicates that these two methods are more complex than our proposed method and this may be the main reason leading to the better performance of these two methods. However, the performance gap between ABD-Net and our proposed method is small. In practical industry applications that require simple but effective solutions, our proposed method could be one of the promising counterbalanced solutions to real-world person ReID tasks.

When comparing the performance of our proposed ADCSLL and ADCSLL + triplet loss, the results show that ADCSLL together with triplet loss achieves better performance than ADCSLL alone. The rank-1 accuracies when using ADCSLL + triplet loss on Market-1501 and DukeMTMC-reID are 95.0% and 88.6%, respectively, 0.6% and 1.1% higher than the numbers when using ADCSLL alone. Similar results are found when comparing the mAP values. The mAP values on Market-1501 and DukeMTMC-reID are 85.8% and 76.0%, respectively, 0.6% and 0.4% higher than the numbers when using ADCSLL alone.

B. PERFORMANCE OF DIFFERENT LOSS FUNCTIONS

To examine the effectiveness of different loss functions in our proposed ADCSLL, we compared the rank-1 accuracies and mAPs on Market-1501 and DukeMTMC-reID in Table 2 when using different loss functions. For fair comparison, all the experiments were conducted under the same network structure. Compared with the Softmax loss, the rank-1 accuracy and mAP on Market-1501 when using label smooth increase by 0.6% and 1.2%, respectively. The numbers are 1.2% and 1.5% on DukeMTMC-reID. The rank-1 accuracies and mAPs of the contrastive loss and triplet loss are not as good as softmax loss on both Market-1501 and DukeMTMC-reID, which indicates that identification loss is more suitable for Re-ID than metric loss. When using label smooth together with triplet loss, the rank-1 accuracy and
mAP are better than either Softmax loss or label smooth on any of the examined datasets. The rank-1 accuracies and mAPs when using the similar labels loss only are higher than the accuracies and mAPs when using the additive distance constraint loss only on both of the datasets. The best performance is achieved when using our proposed method.

The parameter $\varepsilon$ balances the distribution of ground truth label and similar labels, and the parameter $c$ gives a distance constraint to further produce a stricter learning objective. To investigate how the stringent constraint and partial similar labels affect the loss performance on person Re-ID tasks, we further examined the effect of $c$ and $\varepsilon$ when using different $c$ and $\varepsilon$ values in our proposed ADCSLL with $N_{sl} = 100$. By varying $c$ from 0 to 0.5 and $\varepsilon$ from 0 to 0.4, the person Re-ID performance on Market-1501 is illustrated in Fig. 4. The best rank-1 (94.8\%) and mAP (86.4\%) on Market-1501 are simultaneously achieved when $c = 0.4$ and $\varepsilon = 0.3$. The results of a $c$ and $\varepsilon$ combination ($c = 0.35$, $\varepsilon = 0.35$) are shown in Table 2 for detailed numerical comparison.

We further examined the standard deviation (Std) of other labels probabilities at each epoch when using the original Softmax loss, label smooth, and our developed ADCSLL. The results are illustrated in Fig. 5. Std is a measure to evaluate the variation of a set of values. A lower Std means that the values tend to be close to the expected value of the labels’ probabilities. As shown in Fig. 5, ADCSLL obtains the lowest Std when comparing with other loss functions, indicating that ADCSLL can make the features of other labels more stable and more concentrated, and therefore enlarge the intra-class convergence.

### C. INDIVIDUAL PERFORMANCE WHEN USING ADDITIVE DISTANCE CONSTRAINT LOSS OR SIMILAR LABELS LOSS

Our proposed ADCSLL includes two types of loss including the additive distance constraint loss and the similar labels loss. When using the additive distance constraint loss individually for person Re-ID, the rank-1 and mAP results with different $c$ values on Market-1501 are shown in Table 3. The rank-1 and mAP results show that the performance is almost unaffected by different $c$ values between 0 and 0.25. The numbers decrease when $c$ increases after 0.25.

When using the similar labels loss individually for person Re-ID on Market-1501, the rank-1 and mAP results corresponding to the best combination of $N_{sl}$ and $\varepsilon$ when $N_{sl} = 0, 100, 300, 500, 750$ are shown in Table 4. The results show that the best performance (rank-1: 94.6\%, mAP: 85.7\%) are obtained when $N_{sl} = 500$ and $\varepsilon = 0.7$, better than the best results in Table 3. This indicates that similar labels loss might contribute more to the better discriminative capability of our proposed ADCSLL.

As shown in Table 2, the best Re-ID performance of our proposed ADCSLL is achieved when $c = 0.4$ and $\varepsilon = 0.3$, and the corresponding rank-1 accuracy and mAP are 94.8\% and 86.4, respectively. However, as shown in Table 3, $c = 0.4$ corresponds to the worst performance when using the additive distance constraint loss only. The results presented in Table 4 show that neither does $\varepsilon = 0.3$ correspond to the best performance for the examined $N_{sl}$ when using the similar labels loss only. Moreover, none of the presented results
in Table 3 and Table 4 is better than the results in Table 2. These results suggest that additive distance constraint loss and similar labels loss work together to contribute to the better person Re-ID performance.

V. CONCLUSION

This paper proposed a novel and effective approach ADCSLL (additive distance constraint with similar labels loss) to learn highly discriminative features for person Re-ID (re-identification) performance improvement. Both additive distance constraint loss and similar labels loss were innovatively developed to be included in our proposed ADCSLL. Two datasets (Market-1501 and DukeMTMC-reID) were used to examine the effectiveness of our proposed method. The results show that ADCSLL achieves the best performance on either Market-1501 or DukeMTMC-reID, outperforming the previous methods by at least 1% in rank-1 accuracy and 1.2% in mAP. Comparing the performance when using different loss functions, our proposed ADCSLL shows better discriminative capability. When using additive distance constraint loss or similar labels loss individually, the results indicate that: (1) similar labels loss contributes more to the better performance of ADCSLL, and (2) additive distance constraint loss and similar labels loss work together to contribute to the better person Re-ID performance. Our superior results provide a novel perspective for further improvement in intelligent systems like security and video surveillance applications. In our future work, both global features and local features will be used to further improve the performance of our proposed method. Besides, more datasets including CUHK03 [52] and MSMT17 [53] need to be employed for a more complete comparison with the state-of-the-art methods.

REFERENCES

[1] L. Zheng, Y. Yang, and A. G. Hauptmann, “Person re-identification: Past, present and future,” 2016, arXiv:1610.02984. [Online]. Available: http://arxiv.org/abs/1610.02984.

[2] S. E. Li, G. Li, J. Yu, C. Liu, B. Cheng, J. Wang, and K. Li, “Kalman filter-based tracking of moving objects using linear ultrasonic sensor array for road vehicles,” Mech. Syst. Signal Process., vol. 98, pp. 173–189, Jan. 2018.

[3] G. Li, Y. Wang, F. Zhu, X. Sui, N. Wang, X. Qu, and P. Green, “Drivers’ visual scanning behavior at signalized and unsignalized intersections: A naturalistic driving study in China,” J. Saf. Res., vol. 71, pp. 219–229, Dec. 2019.

[4] G. Li, W. Lai, X. Sui, X. Li, X. Qu, T. Zhang, and Y. Li, “Influence of traffic congestion on driver behavior in post-congestion driving,” Accident Anal. Prevention, vol. 141, Jun. 2020, Art. no. 105508.

[5] T. Ni, Z. Deng, F. Chen, and H. Wang, “Relative distance metric learning based on clustering centralization and projection vectors learning for person re-identification,” IEEE Access, vol. 6, pp. 11405–11411, Jan. 2018.

[6] T. Zhang, X. Wang, X. Xu, and C. L. P. Chen, “GCB-Net: Graph convolutional network and its application in emotion recognition,” IEEE Trans. Affect. Comput., early acces, Aug. 27, 2019, doi: 10.1109/ TAFFC.2019.2937768.

[7] T. Zhang, G. Su, C. Qing, X. Xu, B. Cai, and X. Xing, “Hierarchical lifelong learning by sharing representations and integrating hypothesis,” IEEE Trans. Syst., Man, Cybern. Syst., early acces, Feb. 27, 2019, doi: 10.1109/TSMC.2018.2884996.

[8] T. Zhang, C. L. P. Chen, L. Chen, X. Xu, and B. Hu, “Design of highly nonlinear substitution boxes based on 1-Ching operators,” IEEE Trans. Cybern., vol. 48, no. 12, pp. 3349–3358, Dec. 2018.

[9] G. Li, Y. Yang, and X. Qu, “Deep learning approaches on pedestrian detection in haze weather,” IEEE Trans. Ind. Electron., vol. 67, no. 10, pp. 8889–8899, Oct. 2020.

[10] W. Song, Y. Wu, J. Zheng, C. Chen, and F. Liu, “Extended global–local representation learning for video person re-identification,” IEEE Access, vol. 7, pp. 122684–122696, Aug. 2019.

[11] L. Lin, H. Luo, R. Huang, and M. Ye, “Recurrent models of visual co-attention for person re-identification,” IEEE Access, vol. 7, pp. 8865–8875, Jan. 2019.

[12] M. Chen, Y. Ge, X. C. Fu, and D. Yang, “Person re-identification by pose invariant deep metric learning with improved triplet loss,” IEEE Access, vol. 6, pp. 68089–68095, Nov. 2018.

[13] C. Su, J. Li, S. Zhang, J. Xing, W. Gao, and Q. Tian, “Pose-driven deep convolutional model for person re-identification,” in Proc. IEEE Int. Conf. Vis. (ICCV), Oct. 2017, pp. 3960–3969.

[14] L. Zhao, X. Li, Y. Zhang, and J. Wang, “Deeply-learned part-aligned representations for person re-identification,” in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Oct. 2017, pp. 3219–3229.

[15] Y. Sun, L. Zheng, Y. Yang, Q. Tian, and S. Wang, “Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline),” in Proc. ECCV, Sep. 2018, pp. 480–496.

[16] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 770–778.

[17] J. Shao, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “ImageNet: A large-scale hierarchical image database,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2009, pp. 248–255.

[18] W. Li, X. Zhu, and S. Gong, “Person re-identification by deep joint learning of multi-loss classification,” in Proc. 26th Int. Joint Conf. Artif. Intell., Aug. 2017, pp. 2194–2200.

[19] H. Zhao, M. Tian, S. Sun, J. Shao, J. Yan, S. Yi, X. Wang, and X. Tang, “Visibility net: Person re-identification with human body region guided feature decomposition and fusion,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 1077–1085.

[20] G. Wang, Y. Yuan, X. Chen, J. Li, and X. Zhou, “Learning discriminative features with multiple granularities for person re-identification,” in Proc. ACM Multimedia Conf. Multimedia Conf. (MM), 2018, pp. 274–282.

[21] H. Liu, J. Feng, M. Qi, J. Jiang, and S. Yan, “End-to-end comparative attention networks for person re-identification,” IEEE Trans. Image Process., vol. 26, no. 7, pp. 3492–3506, Jul. 2017.

[22] D. Cheng, Y. Gong, S. Zhou, J. Wang, and N. Zheng, “Person re-identification by multi-channel parts-based CNN with improved triplet loss function,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 1335–1344.

[23] M. S. Sarfraz, A. Schumann, A. Eberle, and R. Stiefelhagen, “A pose-sensitive embedding for person re-identification with expanded cross neighborhood re-ranking,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 420–429.

[24] J. Si, H. Zhang, C.-G. Li, J. Kuen, X. Kong, A. C. Kot, and G. Wang, “Dual attention matching network for context-aware feature sequence based person re-identification,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 5363–5372.

[25] J. Zhu, H. Zeng, Y. Du, Z. Lei, L. Zheng, and C. Cai, “Joint feature and neighborhood re-ranking,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 420–429.

[26] G. Li, S. E. Li, R. Zou, Y. Liao, and B. Cheng, “Detection of road traffic participants using cost-effective arrayed ultrasonic sensors in low-speed traffic situations,” Mech. Syst. Signal Process., vol. 132, pp. 535–545, Oct. 2019.

[27] G. Li, S. Li, S. Li, Y. Qin, D. Cao, X. Qu, and B. Cheng, “Deep reinforcement learning enabled decision-making for autonomous driving at intersections,” Automot. Innov., vol. 3, no. 2, p. 1, Sep. 2020.
[28] G. Li, S. E. Li, B. Cheng, and P. Green, “Estimation of driving style in naturalistic highway traffic using maneuver transition probabilities,” Transp. Res. Part C Emerg. Technol., vol. 74, pp. 113–125, Jan. 2017.

[29] F. Schroff, D. Kalenichenko, and J. Philbin, “FaceNet: A unified embedding for face recognition and clustering,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2015, pp. 815–823.

[30] A. Hermans, L. Beyer, and B. Leibe, “In defense of the triplet loss for person re-identification,” 2017, arXiv:1703.07737. [Online]. Available: http://arxiv.org/abs/1703.07737.

[31] Y. Suh, J. Wang, S. Tang, T. Mei, and K. M. Lee, “Part-aligned bilinear representations for person re-identification,” in Proc. ECCV, Sep. 2018, pp. 402–419.

[32] W. Liu, Y. Wen, Z. Yu, and M. Yang, “Large-margin softmax loss for convolutional neural networks,” in Proc. Int. Conf. Mach. Learn., 2016, pp. 507–516.

[33] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song, “SphereFace: Deep hypersphere embedding for face recognition,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 212–220.

[34] H. Wang, Y. Wang, Z. Zhou, X. Ji, D. Gong, J. Zhou, Z. Li, and W. Liu, “CosFace: Large margin cosine loss for deep face recognition,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 5265–5274.

[35] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, “Arcface: Additive angular margin loss for deep face recognition,” in Proc. IEEE Conf. CVPR, Jun. 2019, pp. 4690–4699.

[36] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the inception architecture for computer vision,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 2819–2826.

[37] X. Bai, M. Yang, T. Huang, Z. Dou, R. Yu, and Y. Xu, “Deep-person: Learning discriminative deep features for person re-identification,” Pattern Recognit., vol. 98, Feb. 2020, Art. no. 107036.

[38] L. Wu, Y. Wang, J. Gao, and X. Li, “Where-and-When to look: Deep siamese attention networks for video-based person re-identification,” IEEE Trans. Multimedia, vol. 21, no. 6, pp. 1412–1424, Jun. 2019.

[39] Z. Zheng, L. Zheng, and Y. Yang, “A discriminatively learned CNN embedding for person reidentification,” ACM Trans. Multimedia Comput., Commun., Appl., vol. 14, no. 1, pp. 1–20, Jan. 2018.

[40] T. Xiao, H. Li, W. Ouyang, and X. Wang, “Learning deep feature representations with domain guided dropout for person re-identification,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 1249–1258.

[41] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian, “Scalable person re-identification: A benchmark,” in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Dec. 2015, pp. 1116–1124.

[42] E. Ristani, F. Solera, R. S. Zou, R. Cucchiara, and C. Tomasi, “Performance measures and a data set for multi-target, multi-camera tracking,” in Proc. ECCV, Oct. 2016, pp. 17–35.

[43] Z. Zhong, L. Zheng, G. Kang, S. Li, and Y. Yang, “Random erasing data augmentation,” 2017, arXiv:1708.04896. [Online]. Available: http://arxiv.org/abs/1708.04896.

[44] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” 2014, arXiv:1412.6980. [Online]. Available: http://arxiv.org/abs/1412.6980.

[45] H. Luo, W. Jiang, X. Zhang, X. Fan, J. Qian, and C. Zhang, “AlignedReID++: Dynamically matching local information for person re-identification,” Pattern Recognit., vol. 94, pp. 53–61, Oct. 2019.

[46] Y. Sun, L. Zheng, W. Deng, and S. Wang, “SVDNet for pedestrian retrieval,” in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Oct. 2017, pp. 3800–3808.

[47] F. Zheng, C. Deng, X. Sun, X. Jiang, X. Guo, Z. Yu, F. Huang, and R. Ji, “Pyramidal person re-IDentification via multi-loss dynamic training,” in Pattern IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2019, pp. 8514–8522.

[48] E. Ristani and C. Tomasi, “Features for multi-target multi-camera tracking and re-identification,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 6036–6046.

[49] T. Chen, S. Ding, J. Xie, Y. Yuan, W. Chen, Y. Yang, Z. Ren, and Z. Wang, “ABD-Net: Attentive but diverse person re-identification,” in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), Oct. 2019, pp. 8351–8361.
YAOYU CHEN received the B.E. degree from Shenzhen University, Guangdong, China, in 2018, where he is currently pursuing the master’s degree with the College of Mechatronics and Control Engineering. His research interests include deep learning and driving behavior analysis in autonomous driving.

HENG XIE received the B.E. degree from Shenzhen University, Guangdong, China, in 2018, where he is currently pursuing the master’s degree with the College of Mechatronics and Control Engineering. His research interests include computer vision, deep learning, and object detection for autonomous driving.

SHEN LI received the B.E. degree from Jilin University, in 2012, and the Ph.D. degree from the University of Wisconsin-Madison, in 2019. He has participated in many research projects funding by the National Natural Science Foundation of China, the Ministry of Science and Technology (863 projects), and the U.S. Department of Transportation. His research interests include intelligent systems, intelligent and connected vehicles, autonomous driving safety, intelligent transportation systems (ITS), architecture design of CAVH systems, traffic data mining-based on cellular data, and traffic operations and management.

GANG XU received the B.S. degree in control engineering from Beihang University, Beijing, China, in 1982, and the Ph.D. degree in control engineering from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 1995. From 1987 to 1989, he was a Visiting Research Fellow with Louisiana State University, Baton Rouge, LA, USA. From 1996 to 1997, he held a postdoctoral position with Sussex University, Brighton, U.K. He is currently a Professor with the College of Urban Transportation and Logistics, Shenzhen Technology University, Guangdong, China. His research interests include intelligent systems control, electric vehicles, and traffic safety improvement.