IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received September 28, 2018, accepted October 18, 2018, date of publication October 24, 2018,
date of current version November 19, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2877841

GDMN: Group Decision-Making Network for Person Re-Identification

YANG LIU[1,2], HAO SHENG[1,2], (Member, IEEE), YANWEI ZHENG[1,2], NENGCHENG CHEN[3], WEI KE[4], AND ZHANG XIONG[1,2]

[1]State Key Laboratory of Software Development Environment, School of Computer Science and Engineering, Beihang University, Beijing 100191, China
[2]Beijing Advanced Innovation Center for Big Data and Brain Computing, Beihang University, Beijing 100191, China
[3]State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China
[4]School of Public Administration, Macao Polytechnic Institute, Macau 999078, China

Corresponding author: Hao Sheng (shenghao@buaa.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2017YFC0803700, in part by the National Natural Science Foundation of China under Grant 61472019, in part by the Macao Science and Technology Development Fund under Grant 138/2016/A3, in part by the Open Fund of the State Key Laboratory of Software Development Environment under Grant SKLSDE-2017ZX-09, in part by the Project of Experimental Verification of the Basic Commonness and Key Technical Standards of the Industrial Internet Network Architecture, in part by the Technology Innovation Fund of China Electronic Technology Group Corporation, and in part by the HAWKEYE Group.

ABSTRACT Person re-identification (re-ID) is a widely studied yet still challenging problem in computer vision. It aims to match images of the same pedestrian captured from different cameras. Recently, deep learning has been widely used for feature extraction and distance metric learning in re-ID. However, most of them only consider a certain aspect of the input data and thus will make certain mistakes during the testing process. In this paper, group decision-making (GDM) theory is introduced for comprehensive decision. Furthermore, a novel GDM network (GDMN) is proposed which consists of two sub-networks. First, proposal generation network can generate proposals based on baseline networks for the following decision-making process. Then, decision evaluation network evaluates all the proposals and makes the comprehensive decision. The proposed GDMN can analyze the merits and drawbacks of existing methods and make a better decision. The experimental results on public re-ID benchmarks show that our approach significantly improves the performance of the baseline methods and achieves competitive results compared with other state-of-the-art methods.

INDEX TERMS Convolutional neural networks, group decision-making, person re-identification.

I. INTRODUCTION

Person re-identification (re-ID) aims to match pedestrian images captured from non-overlapping cameras. It has important applications in many fields such as video surveillance, intelligent transportation, and intelligent security system. Re-ID remains a challenging problem due to variations of illuminations, poses and viewpoints, low image resolution, occlusion, *etc.*

Conventional re-ID methods usually make efforts on feature extraction [1]–[4], and distance metric learning [5]–[9]. Recently, with the resurgence of convolutional neural network (CNN) in many other computer vision tasks, deep CNN has also been widely used to solve the re-ID problem [10]–[12], and has achieved extraordinary performance. However, these state-of-the-art network structures, no matter how delicately designed, usually focus on some certain aspects, and produce single outputs (the learned representative features of images or matching scores of image pairs), which sometimes may be unilateral. Fig. 1 shows an example of a re-ID problem, we can improve the overall performance by simply voting among all the baseline methods, from which we can learn that different methods may have some special merits as well as some limits, thus may be complementary. In order to take advantage of this, we treat the existing methods or network architectures as candidate alternatives, and consider a decision-making problem. Rather than simply choose the best alternative, we aim to make a comprehensive decision better than all the alternatives.

When facing a set of alternatives, a human being can make a decision by considering several criteria. A group of people can make a joint decision by some procedures such as discussing and voting. In order to realize human-like decision-making process, we introduce group decision-making theory into re-ID problem.

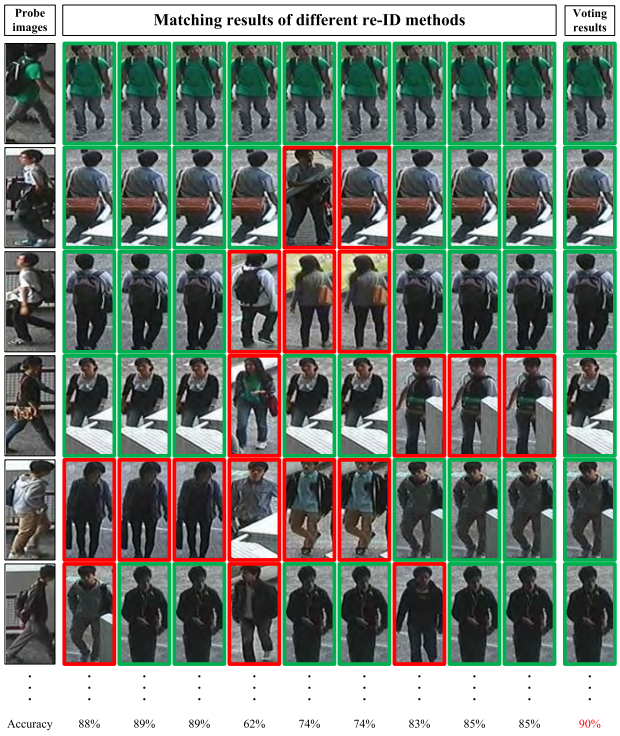


FIGURE 1. An experiment shows that different methods may be complementary as they make different mistakes. Some sample results are shown in the figure, where each row shows a query image and the corresponding matching results proposed by different re-ID methods, and each column belongs to a single re-ID method. The images enclosed by green rectangles represent the correct matches and those in red rectangles indicate the wrong matches. The last column shows the comprehensive result achieved by voting based on all the former methods. The matching accuracies of each method and the comprehensive result are listed at the bottom of each column. Best viewed in color.

Group decision-making (GDM) is a well-studied theory in many fields, including welfare economics, game theory, control theory and psychology, *etc*. GDM describe a situation that a group makes a decision collectively. The goal of GDM is to make a comprehensive decision which is acceptable to all participants, although there might be negotiation and compromise. The advantage of GDM is that a group naturally can gather and process more information than any of its group member, and for this reason is more likely to make a better decision.

In this paper, we propose a GDM framework for re-ID. By introducing GDM, we are able to learn from state-of-the-art methods while overcoming their shortcomings. First, we build a GDM model to analyze and evaluate the existing methods. Second, we explore several approaches to solve the GDM problem, thus a comprehensive decision is made (*i.e.*, a brand new ''method'' is developed), which outperforms the baseline methods it learns from. Third, we implement the GDM procedure by deep convolutional neural network, the proposed Group Decision-making Network (GDMN) is designed to learn the advantages and disadvantages of every single network, and to achieve a comprehensive decision.

The main contributions of this paper are summarized as follows:

1) We introduce GDM theory into re-ID problem, by which a comprehensive decision can be generated based on some existing methods, and outperforms the methods it learns from.
2) We remodel the re-ID problem as a GDM problem, and explore different approaches to solve the GDM problem for re-ID.
3) We construct a deep neural network framework GDMN to implement the GDM procedure for the re-ID task. Experimental results on popular datasets show that our proposed method outperforms many state-of-the-art methods.

II. RELATED WORK

Feature extraction is one of the research interests among conventional re-ID methods. Farenzena *et al.* [1] proposed a feature extraction strategy driven by asymmetry/symmetry principles, and achieved robustness to pose, viewpoint and illumination variations. Zhao *et al.* [2] extracted distinctive features by learning the salient part of human image, which can help to find reliable and discriminative matched patches. Zhao *et al.* [3] proposed an approach to learn mid-level filters from clusters of discriminative and representative patches. Ma *et al.* [4] proposed an image representation based on biologically inspired features and covariance descriptors, which can handle background and illumination variations.

Metric leaning is another important step in popular re-ID pipelines. Zheng *et al.* [5] introduced a probabilistic relative distance comparison model which aims to maximize the matching accuracy regardless the choice of representation. Mignon and Jurie [6] introduced a pairwise constrained component analysis for learning distance metrics in a low-dimensional mapping space. Li *et al.* [7] learned a decision function which consists of a distance metric and a locally adaptive threshold. Sheng *et al.* [13] proposed a visual similarity learning algorithm to learn the distance metric between corresponding patch pairs.

Recently, with the improvement of computational power of GPUs, more and more deep learning methods have been developed for re-ID. Yi *et al.* [10] proposed a 5-layer CNN architecture which can jointly learn feature representation and distance metric. Ahmed *et al.* [11] presented a CNN architecture with a layer that can capture local relationships between input image pairs. Huang *et al.* [14] proposed a DeepDiff network to identify features of human body parts and evaluate the similarities between corresponding parts. Li *et al.* [12] proposed a filter pairing neural network to jointly optimize feature learning, misalignment, occlusions, *etc.*

Some researchers tried to modify the network layers to improve the performance. Hinton *et al.* [15] and Srivastava *et al.* [16] proposed a ''dropout'' method which prevents complex co-adaptations. The output of each hidden unit is set to zero with a probability of 0.5, which can reduce overfitting and perform model averaging.

Ioffe and Szegedy [17] proposed a ''Batch Normalization'' (BN) layer to accelerate the training process by reducing the internal covariate shift. The BN layer performs normalization for each training mini-batch, which can eliminate the impact of different distribution of previous layers, and allows us to use higher learning rates and care less about initialization. Some researchers proposed new activation functions such as Rectifier Linear Unit (ReLU) [18] and Parametric Rectifier Linear Unit (PReLU) [19].

During years of development, some popular network structures emerged and received much attention, such as GoogLeNet (Inception) [20], Inception_v3 [21], ResNet [22], and DGD [23]. These deep learning network structures are all elaborately designed and perform very well. Due to their high accuracy and diversity, *i.e.*, the errors made by these methods are uncorrelated, these methods can be combined to further improve the performance.

Ensemble methods have been widely studied in machine learning, including boosting, bagging, random forests, *etc* [24]. Schapire [25] proposed an iterative approach to generate a strong classifier from an ensemble of weak classifiers, namely boosting. Breiman [26] proposed a bootstrap aggregation approach called bagging. Given a training dataset, several independent classifiers were trained by random sampling. These classifiers were then aggregated to form a stronger classifier. Based on bagging, Breiman [27] proposed random forests, which is an ensemble of decision trees trained with bagging mechanism.

Researchers have been using ensemble methods to solve re-ID. Nanni *et al.* [28] proposed an ensemble of approaches for re-ID, by exploring different color spaces, texture, color features, and distances, the proposed ensemble improved the performance over baseline methods. Paisitkriangkrai *et al.* [29] combined several low-level and high-level features into a single framework, and introduced two optimization functions to optimize re-ID evaluation measures. Liu *et al.* [30] combined different color descriptors through metric learning.

These existing ensemble methods either combining features or descriptors, which are time-consuming, or combining the results by simple summing, which may not obtain a ideal performance. We proposed a learning-based GDM approach, which is effective and efficient. The input of GDM is as simple as some possibilities or rankings, and the learning procedure is separated from the training process of baseline methods, which can significantly reduce time cost.

III. GROUP DECISION-MAKING FOR PERSON RE-IDENTIFICATION

The essence of person re-identification is a matching and ranking problem [31]. Generally speaking, there are two sets of pedestrian images to be matched. A query set, also called probe set, contains the images to be matched with, and a candidate set, also called gallery set, contains the images to be chosen to match each of the query image. The goal of re-ID is to match each pair of images, determine their similarities, and rank all the candidates by similarity to find the best match for the query image. In order to fulfill this goal, some distance metrics or machine learning methods are introduced to calculate the distances between one query object and all the candidate objects. Then the distances are ranked to determine which object in the candidate set is the perfect match for the query object.

Due to varies of distance calculation methods and ranking strategies, different approaches may perform differently on a same dataset. By introducing group decision-making theory, it is possible that the comprehensive decision can adopt different approaches' strong points while overcoming their weak points, thus the overall performance can be improved.

A. PROBLEM MODELING

1) re-ID PROBLEM

Let $Q = \{Q_1, Q_2, \cdots, Q_m\}$ be the query set which contains a set of query images, and $C = \{C_1, C_2, \cdots, C_n\}$ be the candidate set. The goal of re-ID is, for each query image in set Q, trying to find the best match from those candidate images in set C. The problem can be described as: $BM(Q_i) = \arg\min_{C_j}(dist(Q_i, C_j)), j = 1, 2, \cdots, n$, where $d_{ij} = dist(Q_i, C_j)$ is the distance metric between $Q_i \in Q$ and $C_j \in C$. Then the distance matrix $D = (d_{ij})_{m\times n}$ is ranked row by row to find the best match for each $Q_i \in Q$.

2) GDM PROBLEM

Let $DM = \{DM_1, DM_2, \cdots, DM_p\}$ be a group of decision makers trying to find the best matches for a query image Q_i. A decision maker DM_k makes a decision $dec_{DM_k}(Q_i) = C_j \in C$, means DM_k thinks that C_j should be ranked the first place for matching Q_i. The ground truth of the matching is $GT(Q_i) = C_{gt} \in C$. If $C_j = C_{gt}$, we consider decision maker DM_k has made a correct decision about Q_i. Let $Q_{COk} = \{Q_i \mid dec_{DM_k}(Q_i) = GT(Q_i), Q_i \in Q\}$ be the set of query images of which DM_k has made correct decision, then we have the decision accuracy of decision maker DM_k: $accuracy(DM_k) = \frac{card(Q_{COk})}{card(Q)}$, where $card(X)$ represents the number of members of set X.

Let $dec_{GDM}(Q_i)$ be the comprehensive decision of matching Q_i made by the group DM collectively. Analogously, we can define the decision accuracy of GDM: $accuracy(GDM) = \frac{card(Q_{COgdm})}{card(Q)}$. If $accuracy(GDM) > max\{accuracy(DM_k), \forall DM_k \in DM\}$, we can draw the conclusion that the group DM have made a better comprehensive decision. Although the group decision is not guaranteed to be better than all the individual decisions, we focus on finding a better comprehensive decision through GDM when it exists.

B. FEASIBILITY OF GDM

Considering a simple circumstance, two decision makers DM_A and DM_B are trying to match two sets of images $Q = \{Q_1, Q_2\}$ and $C = \{C_1, C_2\}$. They calculate the distance

matrix based on their judge independently:

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \tag{1}$$

$$B = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \tag{2}$$

where a_{ij} is the distance between Q_i and C_j judged by DM_A, and b_{ij} is the same distance judged by DM_B.

Assuming that the ground truth match is $GT(Q_1) = C_1$ and $GT(Q_2) = C_2$, and the entries of matrices A and B subject to:

$$\begin{cases} a_{11} < a_{12} \\ a_{21} < a_{22} \\ b_{11} > b_{12} \\ b_{21} > b_{22} \end{cases} \tag{3}$$

which means DM_A made the right decision about $dec_{DM_A}(Q_1) = C_1$, but was wrong about $dec_{DM_A}(Q_2) = C_1$, and DM_B, on the contrary, was right about Q_2 but wrong about Q_1.

Let the indicator $I_A = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} a_{12} - a_{11} \\ a_{22} - a_{21} \end{pmatrix}$, where $a_1 > 0$ and $a_2 > 0$, similarly we have $I_B = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$, where $b_1 < 0$ and $b_2 < 0$. To make the combined decision to be better than individual decisions, we use the weighted summation of indicators: $I_C = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = wI_A + (1 - w)I_B$, and the combined indicator I_C should satisfy: $c_1 > 0$ and $c_2 < 0$. Let $u = w/(1 - w) > 0$, we have:

$$\begin{cases} ua_1 + b_1 > 0 \\ ua_2 + b_2 < 0 \end{cases} \tag{4}$$

Obviously, by a set of appropriate assignment of a_1, a_2, b_1, and b_2, we can get a feasible solution of u:

$$0 < -\frac{b_1}{a_1} < u < -\frac{b_2}{a_2} \tag{5}$$

Therefore, a feasible weight w can be calculated as $w = u/(1+u)$, thus the original decisions can be improved by weighted combination. That means for a problem and several independent decisions, a better comprehensive decision exists under some certain conditions.

We have proven the existence of a better comprehensive decision under some conditions. The next step is to solve the GDM problem and obtain this comprehensive decision if it exists.

C. SOLVING GDM FOR re-ID

In this section, we explore two methods to solve GDM for re-ID. Taking the characteristics of re-ID problem and practicability in deep neural networks into consideration, we present some novel methods based on several conventional GDM methods. Note that in our GDM model, the goal is different from that of common GDM problems. Instead of trying to decide which alternative is the best one, we aim to make a new decision based on all the alternatives, and make sure that the new decision is better than all existing alternatives.

1) ANALYTIC HIERARCHY PROCESS

Given more than one alternatives, multiple criteria, and several decision makers, a hierarchical structure can be built to describe the GDM problem. All the decision makers can be weighed according to their experience, expertise, reliability, *etc*. And all the criteria can be weighed by importance, correlation, fairness, *etc*. Thus the alternatives can be weighed, which could serve as reference for making the comprehensive decision.

This procedure can be realized through Analytic Hierarchy Process (AHP). AHP is a hierarchical framework to decompose and analyze complex decision-making problems.

The original AHP procedure is as follows:

Step 1: The decision-making problem is decomposed into a set of hierarchical sub-problems. Generally, the top layer is the goal, and the bottom layer is the alternatives. The criteria layer is in the middle, and there can be more than one criteria layer if too many criteria need to be taken into consideration.

Step 2: Comparison matrices are constructed for all the criteria layers from bottom to top. For a single criteria layer, all the criteria are compared pairwise with respect to their impact on the element above them in the hierarchy. Then the results of these comparison make up a matrix $C = (c_{ij})_{n\times n}$, where n is the number of criteria. If criterion i is considered more important than criterion j, the entry c_{ij} is assigned a value greater than 1, and c_{ji} is assigned as $c_{ji} = \frac{1}{c_{ij}}$. The more important criterion i is compared with criterion j, the greater value c_{ij} is.

All the entries of a comparison matrix $C = (c_{ij})_{n\times n}$ satisfy the following conditions:

$$c_{ij} > 0, \quad i, j = 1, 2, \cdots, n \tag{6}$$

$$c_{ij} = \begin{cases} \frac{1}{c_{ji}}, & \text{if } i \neq j \\ 1, & \text{if } i = j \end{cases} \tag{7}$$

Step 3: The right main characteristic vector of comparison matrix is calculated, and is considered as the weight vector $\boldsymbol{w} = (w_1, w_2, \cdots, w_n)^T$. The weight vector $\boldsymbol{w}$ is then used to weigh the criteria layer. Similar calculation can be carried out through the whole hierarchical structure layer by layer from bottom to top, thus the decision can be made according to the weighted criteria.

The procedure above is applicable when there is only one decision maker. In GDM problem, there are multiple decision makers, and each decision maker DM_k with weight λ_k produces a comparison matrix C_k. In order to make the comprehensive group decision, there are several ways to calculate the group comparison matrix C or group weight vector $\boldsymbol{w}$:

Weighted Geometric Mean Comparison Matrix:

Let $C_1, C_2, \cdots, C_p$ be the comparison matrix given by all p decision makers, where $C_k = (c_{ij}^{(k)})_{n\times n}, \quad k = 1, 2, \cdots, p$

are n by n matrices, and n is the number of criteria. The weighted geometric mean of p comparison matrices can be calculated, and this mean matrix is consider to be group comparison matrix $C = (c_{ij})_{n\times n}$, where

$$c_{ij} = (c_{ij}^{(1)})^{\lambda_1}(c_{ij}^{(2)})^{\lambda_2}\cdots(c_{ij}^{(p)})^{\lambda_p} \tag{8}$$

λ_k is the weight of decision maker DM_k, *s.t.*

$$\begin{cases} \lambda_k > 0, & k = 1, 2, \cdots, p \\ \sum_{k=1}^{p} & \lambda_k = 1 \end{cases} \tag{9}$$

Then the right main characteristic vector of group comparison matrix C is calculated, and is considered as the group weight vector $\boldsymbol{w}$.

Weighted Arithmetic Mean Comparison Matrix:

The comparison matrices $C_1, C_2, \cdots, C_p$, as above defined, are given by the decision makers. Group comparison matrix $C = (c_{ij})_{n\times n}$ is obtained by calculating the weighted arithmetic mean of $C_1, C_2, \cdots, C_p$, *i.e.*,

$$c_{ij} = \lambda_1 c_{ij}^{(1)} + \lambda_2 c_{ij}^{(2)} + \cdots + \lambda_p c_{ij}^{(p)} \tag{10}$$

λ_k is subject to Eq. 9.

Then the right main characteristic vector of group comparison matrix C is calculated, and is considered as the group weight vector $\boldsymbol{w}$.

Weighted Geometric Mean Weight Vector:

The weight vectors of all p comparison matrices are calculated respectively.

$$\boldsymbol{w}_k = (w_1^{(k)}, w_2^{(k)}, \cdots, w_n^{(k)})^T, \quad k = 1, 2, \cdots, p \tag{11}$$

Then by calculating the weighted geometric mean of these weight vectors, the group weight vector $\boldsymbol{w} = (w_1, w_2, \cdots, w_n)^T$ is obtained, where

$$\begin{cases} w_i = \dfrac{\bar{w}_i}{\sum_{j=1}^{n} \bar{w}_j}, & i = 1, 2, \cdots, n \\ \bar{w}_i = (w_i^{(1)})^{\lambda_1}(w_i^{(2)})^{\lambda_2}\cdots(w_i^{(p)})^{\lambda_p}, & i = 1, 2, \cdots, n \end{cases} \tag{12}$$

λ_k is subject to Eq. 9.

Weighted Arithmetic Mean Weight Vector:

The weight vectors of all p comparison matrices are calculated respectively, as in Eq. 11. Then the weighted arithmetic mean of these weight vectors is calculated to obtain the group weight vector $\boldsymbol{w} = (w_1, w_2, \cdots, w_n)^T$, where

$$w_i = \lambda_1 w_i^{(1)} + \lambda_2 w_i^{(2)} + \cdots + \lambda_p w_i^{(p)}, \quad i = 1, 2, \cdots, n \tag{13}$$

λ_k is subject to Eq. 9.

Through any of the above procedure, the group weight vector $\boldsymbol{w}$ can be obtained, then by following the AHP procedure, the GDM problem can be solved. The performances of the above procedures may vary due to different conditions or datasets. However, with the help of deep learning, the practicable weight vector can be learned from the training data. See more details in Sec. IV.

AHP is applicable for re-ID. In order to find the best match of query images, several existing approaches can be seen as a set of criteria. Different angles to analyze the performance of these approaches can be considered as decision makers. For example, approach A_1 achieves higher identification accuracy than approach A_2, but A_2 performs better on "hard" samples such as those with severe illumination changes. Decision maker DM_1 pays attention to overall performance, A_1 with higher accuracy will be considered more important than A_2. Meanwhile, decision maker DM_2 appreciate the efforts to extract illumination-robust feature, A_2 might be more valued. The decision makers' different concerns lead to their preference for the candidate approaches. Thus the original re-ID task can be modeled as a GDM problem. Solving this GDM problem by AHP may produce a better comprehensive result than all the candidate approaches can do.

2) TOPSIS METHOD

To solve a multi-criteria decision making problem, a straight forward idea is to establish a uniform criterion. The Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) is a multi-criteria decision analysis method based on this idea. The method defines two ideal solution, the positive ideal solution (PIS) and the negative ideal solution (NIS). Considering that the optimal choice should have the shortest geometric distance from PIS and the longest geometric distance from NIS, a distance metric can be used to evaluate the candidates. By calculating and ranking the distances between each alternative and both ideal solutions respectively, the decision can be made.

In our GDM model, there are multiple decision makers. The goal of GDM is to make a joint group decision rather than simply choose the best option. The original TOPSIS method is adapted for GDM situation. The procedure is as follows:

Step 1: Given m alternatives and n criteria, p decision makers create the decision matrices $Z_k = (z_{ij}^{(k)})_{m\times n}$ respectively, where $z_{ij}^{(k)}$ is the score alternative i got according to criterion j given by decision maker DM_k.

Step 2: Calculate the weighted mean decision matrix $Z = (z_{ij})_{m\times n}$, where

$$z_{ij} = \lambda_1 z_{ij}^{(1)} + \lambda_2 z_{ij}^{(2)} + \cdots + \lambda_p z_{ij}^{(p)} \tag{14}$$

λ_k is the weight of decision maker DM_k, *s.t.*

$$\begin{cases} \lambda_k > 0, & k = 1, 2, \cdots, p \\ \sum_{k=1}^{p} & \lambda_k = 1 \end{cases} \tag{15}$$

Step 3: Calculate the normalized decision matrix $Y = (y_{ij})_{m\times n}$ by normalization method:

$$y_{ij} = \frac{z_{ij}}{\sqrt{\sum_{i=1}^{m} z_{ij}^2}} \tag{16}$$

Step 4: Calculate the weighted normalized decision matrix $X = (x_{ij})_{m\times n}$, where

$$x_{ij} = y_{ij} \cdot w_j \tag{17}$$

w_j is the weight of criterion j, and can be obtained by AHP procedure described above.

Step 5: Calculate the positive ideal solution (PIS) $S_p = \{x_j^{(p)} \mid j = 1, 2, \cdots, n\}$, where

$$x_j^{(p)} = \max_i x_{ij} \tag{18}$$

and the negative ideal solution (NIS) $S_n = \{x_j^{(n)} \mid j = 1, 2, \cdots, n\}$, where

$$x_j^{(n)} = \min_i x_{ij} \tag{19}$$

Step 6: Calculate the Euclidean distance between each alternative and the PIS

$$d_i^{(p)} = \sqrt{\sum_{j=1}^{n}(x_{ij} - x_j^{(p)}),} \quad i = 1, 2, \cdots, m \tag{20}$$

and the Euclidean distance between each alternative and the NIS

$$d_i^{(n)} = \sqrt{\sum_{j=1}^{n}(x_{ij} - x_j^{(n)}),} \quad i = 1, 2, \cdots, m \tag{21}$$

Step 7: Calculate the proximity between each alternative and the ideal solutions

$$P_i = \frac{d_i^{(n)}}{d_i^{(n)} + d_i^{(p)}}, \quad i = 1, 2, \cdots, m \tag{22}$$

P_i can be used to weigh the alternatives in order to generate a comprehensive decision.

TOPSIS is applicable for re-ID problem. For example, m different approaches are used to determine whether or not n pairs of query-candidate objects are perfectly matched. Each approach i provides a solution $S_i = (s_1^{(i)}, s_2^{(i)}, \cdots, s_n^{(i)})$, where $s_j^{(i)}$ indicates the possibility that the j-th pair is perfectly matched, judged by approach i. All these m solutions provided by m approaches can be seen as m alternatives, and n pairs of query-candidate objects can be considered as n criteria, to evaluate the performance of these approaches. The positive ideal solution (PIS) $S_p = (s_1^{(p)}, s_2^{(p)}, \cdots, s_n^{(p)})$ is the ground truth of matching, where $s_j^{(p)} = 1$ means that the j-th pair is matched, and $s_j^{(p)} = 0$ means that the j-th pair is not matched. The negative ideal solution (NIS) $S_n = (s_1^{(n)}, s_2^{(n)}, \cdots, s_n^{(n)})$ is the opposite of PIS, where $s_j^{(n)} = 1 - s_j^{(p)}, \quad j = 1, 2, \cdots, n$. By solving this GDM problem with TOPSIS, all the alternatives, *i.e.*, all the approaches, can be evaluated and weighed, and a comprehensive decision can be made.

A simple experiment was carried out for verification. Four different feature map: RGB, Grayscale, HOG (Histogram of Oriented Gradient), and IIF (Illumination Invariant Features) were used to obtain baseline results of re-ID on CUHK03 dataset [12]. Combined results were obtained by averaging and two GDM methods, the results are shown in Table. 1.

TABLE 1. Verification experiment.

CUHK03	Rank-1	Rank-5	Rank-10	Rank-20
RGB	35.29	71.85	84.66	92.44
Grayscale	8.61	31.09	45.59	62.61
HOG	18.07	49.58	68.49	84.45
IIF	10.29	35.29	54.20	70.38
Averaging	42.65	80.25	89.08	94.33
GDM(AHP)	46.43	81.09	90.55	95.59
GDM(TOPSIS)	**48.53**	**81.51**	**91.39**	**96.22**

In this section, the original re-ID problem is modeled as a GDM problem, and two novel methods are proposed to solve the GDM problem. A verification experiment shows that GDM can improve the performance of baseline methods. Based on this, we constructed a neural network to solve the GDM problem for re-ID. The details are described in the next section.

IV. GROUP DECISION-MAKING NETWORK FOR PERSON RE-IDENTIFICATION

A. GDMN OVERALL ARCHITECTURE

We proposed a Group Decision-making Network (GDMN) to solve re-ID problem. GDMN was composite of two main parts, namely the Proposal Generation Network (PGN) and the Decision Evaluation Network (DEN). The overall network architecture of GDMN is shown in Fig. 2. The goal of PGN is to generate proposals based on some baseline methods. The PGN takes person images as input, and extracts feature vectors through several state-of-the-art networks. Then by some distance metrics, the distances between input image pairs are calculated, as proposals of their similarity. The DEN takes these proposals as input, and makes the final decision through a group decision-making process. The details of these two sub-networks are introduced in the following subsections.

B. PROPOSAL GENERATION NETWORK (PGN)

The goal of Proposal Generation Network (PGN) is to generate the proposals for Decision Evaluation Network (DEN) to evaluate and make the comprehensive decision based on the proposals. The proposal, to be specific in this case, is the matching score of two input images. First, the representative features are extracted from input image pairs through some existing baseline methods. Then, by introducing some distance metrics, the distance between the two input images can be calculated.

We used three famous network structure as baseline methods: Inception_v3 [21], ResNet50 [22], and DGD [23]. In order to extract comparable features, we modified the parameter of some layers, and added a fully connected

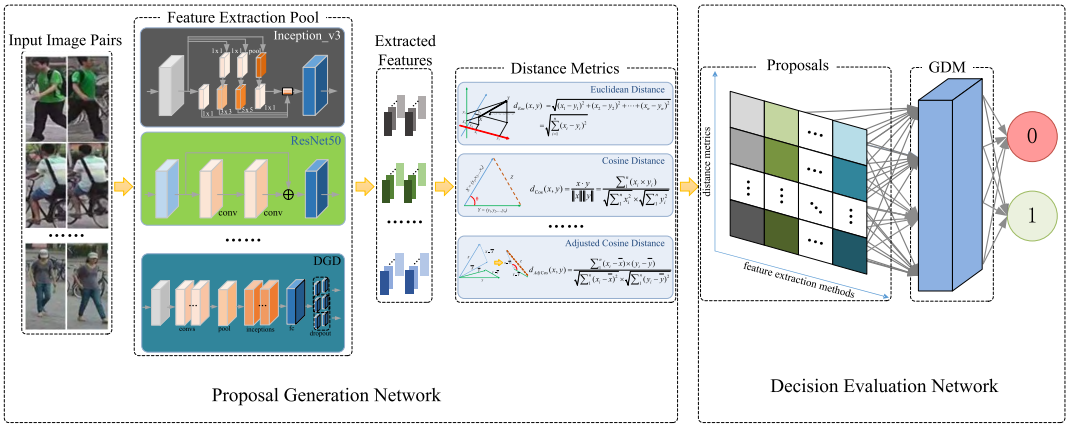


FIGURE 2. **The overall architecture of the proposed Group Decision-making Network. The left part is Proposal Generation Network, and the right part is Decision Evaluation Network.**

layer with 128 outputs before the classification layer. This 128-dimensional output was considered as the representative feature. Thus for each network, a 128-dimensional feature was extracted to represent the original input image.

Then three distance metrics were introduced to calculate the distance between the features:

Euclidean Distance is a widely used distance metric, it is the straight-line distance between two points in Euclidean space. Let $x = (x_1, x_2, \cdots, x_n)$ and $y = (y_1, y_2, \cdots, y_n)$ be the two extracted features. The Euclidean distance between them is:

$$d_{Euc}(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{23}$$

Cosine Distance is the cosine of the angle between two non-zero vectors in inner product space. The cosine distance between $x = (x_1, x_2, \cdots, x_n)$ and $y = (y_1, y_2, \cdots, y_n)$ is:

$$d_{Cos}(x, y) = \frac{x \cdot y}{\|x\| \, \|y\|} = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2} \sqrt{\sum_{i=1}^{n} y_i^2}} \tag{24}$$

Adjusted Cosine Distance: The cosine distance only measures the different in directions, and is insensitive about the value. In order to better describe the distance, the mean value were first subtracted from each vector, then the cosine distance was calculated:

$$d_{AdjCos}(x, y) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2} \sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{25}$$

Through the combination of three feature extraction methods and three distance metrics, nine similarity scores were obtained. The matching result of baseline methods can be calculated by ranking the test samples by these similarity scores. The proposals were generated and inputed into DEN for further evaluation and comprehensive decision-making.

C. DECISION EVALUATION NETWORK (DEN)

Since the input of DEN was simplified as the similarity scores of the baseline methods, the structure of DEN is relatively simple. The input of DEN is three by three matrices contain the probabilities calculated by three distance metrics and three different features. These proposals were passed through a fully connected layer with 20 hidden nodes, where the mutual correlation between the components was learned. Then another fully connected layer with two outputs was used to output the final decision. A value between 0 and 1 was calculated by a softmax classifier, which indicated the possibility of the two input images representing a same person.

D. TRAINING AND TESTING DETAILS

The training process of DEN needs the proposals generated by PGN. The proposals used to train DEN are actually generated by the testing process of PGN with training dataset. The whole network was trained in the following steps:

1) Train the PGN with training dataset.
2) Test the PGN with training dataset to extract the 128-d features, and compute the matching probabilities through the distance metrics to generate training proposals for DEN.
3) Train the DEN with training proposals.

And the testing steps are as follows:

1) Test the PGN with testing dataset, get the results of baseline methods (for comparison), at the same time, generate the testing proposals.
2) Test the DEN with testing proposals, get the result of GDMN.
3) Compare the result of GDMN with the results of baseline methods.

V. EXPERIMENTAL RESULTS

A. DATASETS

We evaluate our Group Decision-making Network on six public re-ID datasets, including CUHK03 [12], CUHK01 [32], Market-1501 [33], PRID2011 [34], i-LIDS [35], and VIPeR [36].

TABLE 2. Datasets grouping.

Dataset	#images / #IDs			
	train & validate	test probe	test gallery	subtotal
CUHK03	26,243 / 1,367	962 / 100	988 / 100	28,193 / 1,467
CUHK01	1,940 / 485	972 / 486	972 / 486	3,884 / 971
Market-1501	12,936 / 751	3,368 / 750	13,115 / 750	29,419 / 1,501
PRID2011	4,282 / 834	100 / 100	100 / 100	4,482 / 934
i-LIDS	238 / 59	107 / 60	131 / 60	476 / 119
VIPeR	632 / 316	316 / 316	316 / 316	1,264 / 632
Total	46,271 / 3,812	5,825 / 1,812	15,622 / 1,812	67,718 / 5,624

CUHK03: The dataset contains 14,097 images of 1,467 identities captured from five pairs of cameras. Each identity is observed by one pair of cameras, and has an average of 4.8 images per camera view. The dataset provides both auto-detected and manually-labeled pedestrian samples, and both are used in our experiment.

CUHK01: The dataset includes 971 pedestrians, and each pedestrian has two images captured from two disjoint cameras, so totally there are 3,884 images. All the pedestrian images are manually cropped from the original image frames and then normalized to same size.

Market-1501: A large dataset contains 1501 identities, includes 32,668 bounding boxes produced by the Deformable Part Model (DPM) [37]. The dataset was collected using six cameras in front of a campus supermarket where overlapping exists. Each identity is captured by at least two and at most six cameras due to the open environment. Furthermore, for the detected DPM bounding boxes and the manually cropped boxes, the ratio of the overlapping area to the union area is calculated to evaluate the quality of detector. If the ratio is larger than 50%, the bounding box is marked as ''good''; if the ratio is smaller than 20%, the box is marked as ''distractor''; the other boxes, the ratio is between 20% and 50%, are marked as ''junk'' since they have no influence to re-ID accuracy. Totally, there are 2,798 ''distractors'' and 3,819 ''junk'' images.

PRID2011: This dataset is composed of 385 persons from camera A and 749 persons from camera B, of which only 200 persons appear in both cameras. There are two versions of the dataset, the single-shot version contains only one randomly chosen image per camera view, and the multi-shot version contains multiple images per person per camera. We use multi-shot version for training and single-shot version for testing in our experiments.

i-LIDS: The dataset is based on iLIDS Multiple Camera Tracking Scenario dataset, and contains 476 images of 119 identities captured from two non-overlapping cameras in an airport's arrival hall during its busy time. Thus it is a challenging dataset due to heavy occlusion and pose variance.

VIPeR: This dataset is one of the most widely used re-ID dataset, yet it is still challenging due to low resolution and severe illumination and viewpoint variations. The VIPeR dataset contains 1264 images from two disjoint cameras, each of which captures one image of 632 pedestrians.

Table. 2 shows the grouping of training, validation and testing data. For CUHK03 dataset, we followed the grouping protocol in the dataset, used the 20th group for testing and the 19th group (excluded those in the 20th group) for validation, the other images were used in training procedure. For PRID2011 dataset, the first 100 identities were used as testing data. For Market-1501 dataset, we followed the grouping protocol as the author suggested. For the rest three datasets, we randomly selected half of the identities as testing data and the others were used as training and validation data. The grouping details are showed in Table. 2.

B. IMPLEMENTATION

In our experiment, three network structure are used as baseline methods: Inception_v3 [21], ResNet50 [22], and DGD [23]. For the Proposal Generation Network (PGN), we modified the original structures of the baseline networks in order to extract comparable features. A fully connected layer with 128 outputs was added before the classification layer to extract 128-d features, the kernel size and stride of former convolutional layers were adjusted according to the size of input image and the length of feature. Then we applied three different distance metrics to the extracted 128-d features to measure and rank the distance between each sample pairs. By ranking the distances we can compute the probability of two images representing a same pedestrian, and these probabilities were fed to the Decision Evaluation Network (DEN) to evaluate the performance of the baseline networks, and then a final comprehensive decision was made based on the evaluation.

We implemented our GDMN architecture using the Caffe [38] framework. For PGN, we followed the settings as DGD [23] to jointly train the networks on all the six datasets in order to overcome the scale issue on small datasets. This Jointly Single-task Learning (JSTL) scheme was used to pre-train the network for 80,000 iterations, and then the network was fine-tuned on each dataset. For large datasets such as CUHK03, CUHK01, Market-1501 and PRID2011, the network was fine-tuned for 120,000 iterations, and for i-LIDS and VIPeR, the iterations of fine-tune was 40,000 due to the small scale of datasets. For DEN, the details of solver settings are as follows: solver type: SGD, base_lr: 0.001, learning rate policy: inv (learning rate $= base\_lr * (1 + gamma * iter)^{(-power)}$), in which gamma $= 0.0001$ and power $= 0.75$.

The matching results of test image pairs were then ranked using MATLAB 2017b to obtain the rank-1 accuracies and draw the CMC curves. For datasets CUHK03, CUHK01, Market-1501, and i-LIDS, each identity has more than one image, this calculation was repeated 100 times to calculate the mean value. The computation time of DEN using Caffe and ranking using MATLAB is reported in Table. 3. The Caffe code was run using an NVIDIA GTX1070 GPU, and the MATLAB code was run using an i7-6700 3.40GHz processor with 16GB memory. The computation cost is low except for Market-1501 dataset, which is a large-scale dataset with 3,368 testing images as probe set and 13,115 testing images as gallery set.

TABLE 3. Computation time.

Dataset	Caffe-Train	Caffe-Test	MATLAB-Ranking
CUHK03	3m29s	31.85s	1.44s
CUHK01	1m5s	58.28s	4.49s
Market-1501	23m14s	39m16s	62.27s
PRID2011	2m51s	2.24s	0.08s
i-LIDS	7.12s	1.41s	0.14s
VIPeR	7.47s	27.49s	0.22s

C. EVALUATION

The Cumulative Match Characteristic (CMC) curve is widely used to evaluate the performance of re-ID methods. It shows the rank-k matching accuracy, *i.e.*, the probability of the ground truth appears in the top k of the ranking list. In most of the datasets, except VIPeR, each identity has more than one images. There are three matching protocol to calculate the CMC [39]: (1) single-shot vs. single-shot (SvsS), where each image in both probe set and gallery set represents a different individual; (2) single-shot vs. multiple-shot (SvsM), also called single-query (SQ), where each identity has only one image in probe set but has multiple images in the gallery set; and (3) multiple-shot vs. multiple-shot (MvsM), also called multiple-query (MQ), where each ID has multiple images in both probe set and gallery set. MvsM protocol is not widely used, so we calculate CMC with both SvsS and SvsM protocol for fair comparison. For SvsS case, we randomly choose one probe image and one gallery image for each ID to calculate the CMC, and repeat it 100 times to obtain the mean CMC. For SvsM case, one probe image is randomly chosen for each ID while all the gallery image are used, and only the first match in the ranking list is counted.

Some researchers point out that the CMC curve is only valid when there is only one ground truth match for a given query image. If multiple ground truths exist, the CMC curve is biased because recall is not considered [33]. In this case, mean Average Precision (mAP) is introduced. Technically, the Average Precision (AP) should be the precision averaged across all values of recall between 0 and 1, which is equal to the area under the Precision-Recall curves:

$$\int_0^1 P(r)dr \tag{26}$$

In practice, the AP can be approximated by summing the precisions at every threshold value, and multiplied by the change in recall:

$$\sum_{k=1}^{N} P(k)\Delta r(k) \tag{27}$$

where N is the number of image in candidate set, $P(k)$ is the precision at a cut-off of k images, and $\Delta r(k)$ is the change in recall that happened between cut-off $k-1$ and cut-off k.

However, the approximated precision in Eq. 27 is not a monotonically decreasing function. Some researchers choose to use an alternative approximation called interpolated Average Precision, usually it is still called Average Precision. Instead of using $P(k)$, the precision at cut-off k, the interpolated Average Precision uses the maximum precision observed across all the cut-offs with higher recall:

$$\sum_{k=1}^{N} \max_{\widetilde{k}\geq k} P(\widetilde{k})\Delta r(k) \tag{28}$$

The interpolated Average Precision is more frequently used in image retrieval, so in our experiment, we use Eq. 28 to calculate the AP. Then, the mean value of APs of all the query images, *i.e.*, mAP, can be calculated.

D. COMPARISON

First, we compare the performance of our GDMN with the baseline methods: Inception_v3 [21], ResNet50 [22], and DGD [23]. The SvsS rank-1-5-10 accuracy and the SvsM mAP of GDMN and three baseline networks are shown in Table. 4. The best performances are marked in bold. Note that for each baseline method, we applied three different distance metrics: Euclidean Distance, Cosine Distance, and Adjusted Cosine Distance, but only the best result of each method is shown in the table due to space limitation. The SvsS CMC curves of our GDMN and the baseline methods on six datasets are shown in Fig. 3.

From Table. 4 and Fig. 3 we can observe that the proposed GDMN outperforms all the baseline methods on most of the datasets, except for VIPeR. VIPeR is a relatively small dataset, and it is challenging due to severe variations on illuminations and viewpoints, which leads to relatively low matching accuracies of baseline methods. It is reasonable that the GDMN cannot achieve a better rank-1 accuracy based on candidate decisions that are not good enough, and it is worth noting that GDMN still improves the overall performance after rank-10 on VIPeR dataset.

Then, the proposed GDMN is compared with several state-of-the-art re-ID methods, including PaMM [40], GOG [41], JLSCR [42], LDNS [43], RCN [44], CDTL [45], MCPB [46], LSSCDL [47], SCSP [48], TDL [49], Quadruplet [50], CADL [51], CSBT [52], MSCAN [53], OSML [54], P2S [55], Re-ranking [56], SSM [57], JSTRNN [58], Spindle [59], SHaPE [60], MuDeep [61], PDC [62], ASTPN [63], PAR [64], LSRO [65], and OL-MANS [66], and several ensemble re-ID methods, including FUS1 [28],

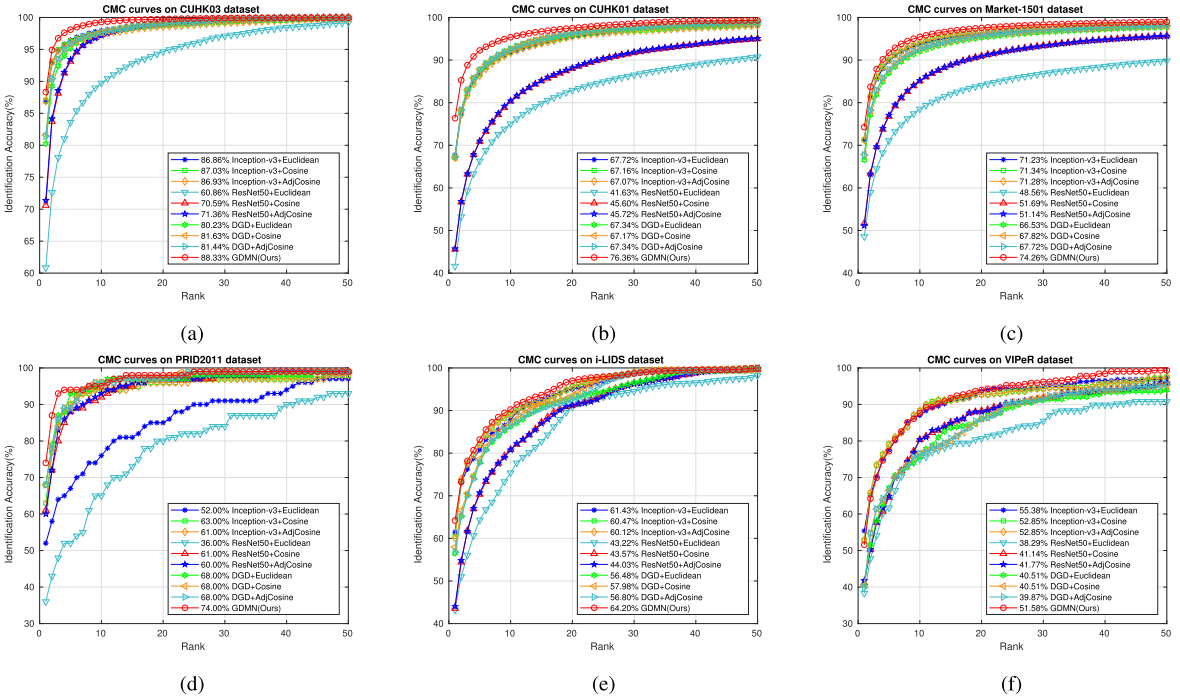


FIGURE 3. The CMC curves (SvsS) of our GDMN and the baseline methods on six datasets. (a) CUHK03. (b) CUHK01. (c) Market-1501. (d) PRID2011. (e) i-LIDS. (f) VIPeR.

TABLE 4. Comparison with baseline networks.

Dataset	Method	Rank-1	Rank-5	Rank-10	mAP
CUHK03	Inception_v3 [21]	87.03	96.32	97.64	88.71
	ResNet50 [22]	71.36	93.44	97.17	75.13
	DGD [23]	81.63	95.88	97.68	84.16
	GDMN(Ours)	**88.33**	**98.09**	**99.28**	**89.61**
CUHK01	Inception_v3 [21]	67.72	86.89	91.71	73.91
	ResNet50 [22]	45.72	70.98	80.44	53.72
	DGD [23]	67.34	87.77	92.69	74.04
	GDMN(Ours)	**76.36**	**92.27**	**95.36**	**81.33**
Market-1501	Inception_v3 [21]	71.34	90.05	93.94	66.81
	ResNet50 [22]	51.69	76.78	85.08	43.13
	DGD [23]	67.82	87.72	92.48	60.90
	GDMN(Ours)	**74.26**	**91.75**	**95.45**	**69.38**
PRID2011	Inception_v3 [21]	63.00	90.00	94.00	-
	ResNet50 [22]	61.00	88.00	92.00	-
	DGD [23]	68.00	93.00	**96.00**	**-**
	GDMN(Ours)	**74.00**	**94.00**	95.00	-
i-LIDS	Inception_v3 [21]	61.43	80.78	87.93	67.96
	ResNet50 [22]	44.03	70.70	80.77	51.50
	DGD [23]	57.98	78.67	87.17	60.54
	GDMN(Ours)	**64.20**	**83.18**	**90.25**	**68.92**
VIPeR	Inception_v3 [21]	**55.38**	**77.53**	87.03	-
	ResNet50 [22]	41.77	64.87	80.38	-
	DGD [23]	40.51	67.09	75.63	-
	GDMN(Ours)	51.58	77.22	**87.66**	**-**

TABLE 5. Comparison with State-of-the-art on CUHK03.

CUHK03	Rank-1	Rank-5	Rank-10	Rank-20	mAP
GOG [41]	67.30	91.00	96.00	-	-
JLSCR [42]	52.17	-	-	-	-
LDNS [43]	62.55	90.05	94.80	98.10	-
LSSCDL [47]	57.00	-	-	-	-
Quadruplet [50]	74.47	96.62	98.95	-	-
CSBT [52]	55.50	84.30	-	98.00	-
MSCAN [53]	74.21	94.33	97.54	99.25	-
Re-ranking [56]	61.60	-	-	-	67.60
SSM [57]	76.60	94.60	98.00	-	-
Spindle [59]	88.50	97.80	98.60	99.20	-
MuDeep [61]	76.87	96.12	98.41	-	-
PDC [62]	**88.70**	**98.61**	99.24	99.67	-
PAR [64]	85.40	97.60	**99.40**	**99.90**	**-**
LSRO [65]	84.60	97.60	98.90	-	**87.40**
OL-MANS [66]	61.68	88.39	95.23	98.47	-
CMCTOP [29]	62.10	89.10	94.30	97.80	-
GDMN SvsS	88.33	**98.09**	**99.28**	**99.76**	**-**
GDMN SvsM	**89.40**	94.18	97.61	98.75	**89.61**

CMCTOP [29], and ECM [30]. The Rank-1-5-10-20 accuracy and mAP of our GDMN and state-of-the-art methods on six datasets are reported in Table. 5 through Table. 10. The best two performances are marked in bold. Note that many researchers applied SvsM protocol when evaluate their methods on the datasets that contain multiple images per identity, and some of them did not specifically mention which protocol they used, so we report our results based on both SvsS and SvsM protocol on CUHK03, CUHK01, Market-1501 and i-LIDS datasets for comparison.

Experimental results show that our proposed GDMN achieves the highest Rank-1 accuracy and mAP on CUHK03, CUHK01, Market-1501, and i-LIDS dataset among all the compared methods, including some ensemble methods.

TABLE 6. Comparison with State-of-the-art on CUHK01.

CUHK01	Rank-1	Rank-5	Rank-10	Rank-20	mAP
GOG SvsS [41]	57.80	79.10	86.20	92.10	-
GOG SvsM [41]	67.30	86.90	91.80	95.90	-
JLSCR [42]	71.80	-	-	-	-
LDNS [43]	69.09	86.87	91.77	95.39	-
CDTL [45]	32.10	-	-	-	-
MCPB [46]	53.70	84.30	91.00	96.30	-
LSSCDL [47]	65.97	-	-	-	-
Quadruplet [50]	62.55	83.02	88.79	-	-
CSBT [52]	51.20	76.30	-	91.80	-
OSML [54]	45.60	-	-	-	-
P2S [55]	77.34	93.51	96.73	**98.53**	-
Spindle [59]	**79.90**	**94.40**	**97.10**	**98.60**	-
MuDeep [61]	79.01	**97.00**	**98.96**	-	-
PAR [64]	75.00	93.50	95.70	97.70	-
OL-MANS [66]	68.44	87.16	92.67	95.88	-
CMCTOP [29]	53.40	76.40	84.40	90.50	-
GDMN SvsS	76.36	92.27	95.36	97.51	-
GDMN SvsM	**80.25**	93.52	96.40	98.15	**81.33**

TABLE 7. Comparison with State-of-the-art on Market-1501.

Market-1501	Rank-1	Rank-5	Rank-10	Rank-20	mAP
LDNS [43]	61.02	-	-	-	35.68
SCSP [48]	51.90	-	-	-	26.35
CADL [51]	73.84	-	-	-	47.11
CSBT [52]	42.90	-	-	-	20.30
MSCAN [53]	80.31	-	-	-	57.53
P2S [55]	70.72	90.52	-	-	44.27
Re-ranking [56]	77.11	-	-	-	63.63
SSM [57]	82.21	-	-	-	**68.80**
Spindle [59]	76.90	91.50	94.60	96.70	-
PDC [62]	**84.14**	**92.73**	94.92	96.82	63.41
PAR [64]	81.00	92.00	94.70	-	63.40
LSRO [65]	83.97	-	-	-	66.07
OL-MANS [66]	60.67	-	-	91.87	-
GDMN SvsS	74.26	91.75	**95.45**	**97.54**	-
GDMN SvsM	**85.78**	**94.33**	**96.08**	**97.57**	**69.38**

TABLE 8. Comparison with State-of-the-art on PRID2011.

PRID2011	Rank-1	Rank-5	Rank-10	Rank-20
PaMM [40]	45.00	72.00	85.00	92.50
LDNS [43]	40.90	64.70	73.20	81.00
RCN [44]	70.00	90.00	95.00	97.00
CDTL [45]	25.30	-	-	-
MCPB [46]	22.00	-	47.00	57.00
TDL [49]	56.74	80.00	87.64	93.59
OSML [54]	41.40	72.00	85.00	92.50
P2S [55]	70.71	**95.15**	**98.92**	**100.00**
JSTRNN [58]	**79.40**	94.40	-	**99.30**
Spindle [59]	67.00	89.00	89.00	92.00
MuDeep [61]	65.00	87.00	93.00	-
ASTPN [63]	**77.00**	**95.00**	**99.00**	99.00
CMCTOP [29]	17.90	-	-	-
GDMN SvsS	74.00	94.00	95.00	98.00

For PRID2011 dataset, the performance of our GDMN is slightly worse than JSTRNN [58] and ASTPN [63], because both of them are video-based methods, and use SvsM protocol during evaluation. Our GDMN can achieve a comparable result while only using single-frame information and SvsS protocol. For VIPeR dataset, due to the low matching accuracies of baseline methods, our GDMN cannot achieve a better result through group decision-making. However, GDMN still achieves 51.58% Rank-1 accuracy on VIPeR dataset, which is only 2.22 percentage point lower than the best compared method Spindle [59], and outperforms most of the compared methods. It is worth noting that the performance of our GDMN might be further improved by adopting better baseline methods.

TABLE 9. Comparison with State-of-the-art on i-LIDS.

i-LIDS	Rank-1	Rank-5	Rank-10	Rank-20	mAP
CDTL [45]	50.30	-	-	-	-
MCPB [46]	60.40	82.70	90.70	**97.80**	-
OSML [54]	51.20	-	-	-	-
Spindle [59]	**66.30**	**86.60**	**91.80**	95.30	-
CMCTOP [29]	50.30	-	-	-	-
GDMN SvsS	64.20	83.18	90.25	**96.98**	-
GDMN SvsM	**75.70**	**90.65**	**92.52**	95.33	**68.92**

TABLE 10. Comparison with State-of-the-art on VIPeR.

VIPeR	Rank-1	Rank-5	Rank-10	Rank-20
GOG [41]	49.70	79.70	88.70	94.50
JLSCR [42]	35.76	-	-	-
LDNS [43]	51.17	**82.09**	90.51	95.92
CDTL [45]	34.10	-	-	-
MCPB [46]	47.80	74.70	84.80	91.10
LSSCDL [47]	42.66	-	84.27	91.93
SCSP [48]	53.54	**82.59**	**91.49**	**96.65**
Quadruplet [50]	48.42	74.05	84.49	-
CSBT [52]	36.60	66.20	-	88.30
OSML [54]	34.30	-	-	-
SSM [57]	**53.73**	-	**91.49**	**96.08**
Spindle [59]	**53.80**	74.10	83.20	92.10
SHaPE [60]	34.26	57.34	67.86	80.78
MuDeep [61]	43.03	74.36	85.76	-
PDC [62]	51.27	74.05	84.18	91.46
PAR [64]	48.70	74.70	85.10	93.00
OL-MANS [66]	44.97	74.43	84.97	93.64
FUS1 [28]	24.50	-	61.70	-
CMCTOP [29]	45.90	77.50	88.90	95.80
ECM [30]	38.90	67.80	78.40	88.90
GDMN SvsS	51.58	77.22	87.66	93.99

VI. CONCLUSION

In this paper, we introduce group decision-making (GDM) theory into person re-ID. After analyzing and evaluating the performances of baseline methods, a comprehensive decision can be made by the GDM process. A deep learning architecture GDMN is proposed to implement the GDM process. The comparison experiment shows that our proposed GDMN can significantly improve the performances of baseline methods, further experimental results show that GDMN outperforms most of the state-of-the-art methods, including some ensemble methods, on many public datasets.

REFERENCES

[1] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani, "Person re-identification by symmetry-driven accumulation of local features," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2010, pp. 2360–2367.

[2] R. Zhao, W. Ouyang, and X. Wang, "Unsupervised salience learning for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2013, pp. 3586–3593.

[3] R. Zhao, W. Ouyang, and X. Wang, "Learning mid-level filters for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 144–151.

[4] B. Ma, Y. Su, and F. Jurie, "BiCov: A novel image representation for person re-identification and face verification," in *Proc. Brit. Mach. Vis. Conf.*, 2012, p. 11–pages.

[5] W.-S. Zheng, S. Gong, and T. Xiang, "Person re-identification by probabilistic relative distance comparison," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2011, pp. 649–656.

[6] A. Mignon and F. Jurie, "PCCA: A new approach for distance learning from sparse pairwise constraints," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2012, pp. 2666–2672.

[7] Z. Li, S. Chang, F. Liang, T. S. Huang, L. Cao, and J. R. Smith, "Learning locally-adaptive decision functions for person verification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2013, pp. 3610–3617.

[8] W.-S. Zheng, S. Gong, and T. Xiang, "Reidentification by relative distance comparison," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 3, pp. 653–668, Mar. 2013.

[9] M. Koestinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof, "Large scale metric learning from equivalence constraints," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2012, pp. 2288–2295.

[10] D. Yi, Z. Lei, S. Liao, and S. Z. Li, "Deep metric learning for person re-identification," in *Proc. 22nd Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2014, pp. 34–39.

[11] E. Ahmed, M. Jones, and T. K. Marks, "An improved deep learning architecture for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 3908–3916.

[12] W. Li, R. Zhao, T. Xiao, and X. Wang, "DeepReID: Deep filter pairing neural network for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 152–159.

[13] H. Sheng, Y. Huang, Y. Zheng, J. Chen, and Z. Xiong, "Person re-identification via learning visual similarity on corresponding patch pairs," in *Proc. Int. Conf. Knowl. Sci., Eng. Manage.*, 2015, pp. 787–798.

[14] Y. Huang, H. Sheng, Y. Zheng, and Z. Xiong, "DeepDiff: Learning deep difference features on human body parts for person re-identification," *Neurocomputing*, vol. 241, pp. 191–203, Jun. 2017.

[15] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov. (2012). "Improving neural networks by preventing co-adaptation of feature detectors." [Online]. Available: https://arxiv.org/abs/1207.0580

[16] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.

[17] S. Ioffe and C. Szegedy. (2015). "Batch normalization: Accelerating deep network training by reducing internal covariate shift." [Online]. Available: https://arxiv.org/abs/1502.03167

[18] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proc. 14th Int. Conf. Artif. Intell. Stat.*, 2011, pp. 315–323.

[19] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 1026–1034.

[20] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 1–9.

[21] 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.*, Jun. 2016, pp. 2818–2826.

[22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.

[23] 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.

[24] T. G. Dietterich, "Ensemble methods in machine learning," in *Proc. Int. Workshop Multiple Classifier Syst.*, 2000, pp. 1–15.

[25] R. E. Schapire, "The strength of weak learnability," *Mach. Learn.*, vol. 5, no. 2, pp. 197–227, 1990.

[26] L. Breiman, "Bagging predictors," *Mach. Learn.*, vol. 24, no. 2, pp. 123–140, 1996.

[27] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[28] L. Nanni, M. Munaro, S. Ghidoni, E. Menegatti, and S. Brahnam, "Ensemble of different approaches for a reliable person re-identification system," *Appl. Comput. Inform.*, vol. 12, no. 2, pp. 142–153, 2016.

[29] S. Paisitkriangkrai, C. Shen, and A. van den Hengel, "Learning to rank in person re-identification with metric ensembles," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 1846–1855.

[30] X. Liu, H. Wang, Y. Wu, J. Yang, and M.-H. Yang, "An ensemble color model for human re-identification," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2015, pp. 868–875.

[31] H.-X. Yu, A. Wu, and W.-S. Zheng, "Cross-view asymmetric metric learning for unsupervised person re-identification," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 1–9.

[32] W. Li, R. Zhao, and X. Wang, "Human reidentification with transferred metric learning," in *Proc. Asian Conf. Comput. Vis.*, 2012, pp. 31–44.

[33] 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.*, Dec. 2015, pp. 1116–1124.

[34] M. Hirzer, C. Beleznai, P. M. Roth, and H. Bischof, "Person re-identification by descriptive and discriminative classification," in *Proc. Scandin. Conf. Image Anal.*, 2011, pp. 91–102.

[35] W.-S. Zheng, S. Gong, and T. Xiang, "Associating groups of people," in *Proc. Brit. Mach. Vis. Conf.*, vol. 2, no. 6, 2009, p. 1.

[36] D. Gray, S. Brennan, and H. Tao, "Evaluating appearance models for recognition, reacquisition, and tracking," in *Proc. IEEE Int. Workshop Perform. Eval. Tracking Surveill. (PETS)*, vol. 3, no. 5, 2007, pp. 1–7.

[37] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 9, pp. 1627–1645, Sep. 2010.

[38] Y. Jia *et al.*, "Caffe: Convolutional architecture for fast feature embedding," in *Proc. 22nd ACM Int. Conf. Multimedia*, 2014, pp. 675–678.

[39] L. Bazzani, M. Cristani, and V. Murino, "SDALF: Modeling human appearance with symmetry-driven accumulation of local features," in *Person Re-Identification*. New York, NY, USA: Springer, 2014, pp. 43–69.

[40] Y.-J. Cho and K.-J. Yoon, "Improving person re-identification via pose-aware multi-shot matching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 1354–1362.

[41] T. Matsukawa, T. Okabe, E. Suzuki, and Y. Sato, "Hierarchical Gaussian descriptor for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 1363–1372.

[42] F. Wang, W. Zuo, L. Lin, D. Zhang, and L. Zhang, "Joint learning of single-image and cross-image representations for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 1288–1296.

[43] L. Zhang, T. Xiang, and S. Gong, "Learning a discriminative null space for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 1239–1248.

[44] N. McLaughlin, J. M. del Rincon, and P. Miller, "Recurrent convolutional network for video-based person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1325–1334.

[45] P. Peng *et al.*, "Unsupervised cross-dataset transfer learning for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 1306–1315.

[46] 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.*, Jun. 2016, pp. 1335–1344.

[47] Y. Zhang, B. Li, H. Lu, A. Irie, and X. Ruan, "Sample-specific SVM learning for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 1278–1287.

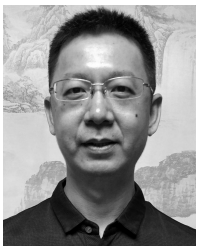
[48] D. Chen, Z. Yuan, B. Chen, and N. Zheng, "Similarity learning with spatial constraints for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 1268–1277.

[49] J. You, A. Wu, X. Li, and W.-S. Zheng, "Top-push video-based person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 1345–1353.

[50] W. Chen, X. Chen, J. Zhang, and K. Huang, "Beyond triplet loss: A deep quadruplet network for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 2, no. 8, Jul. 2017, pp. 403–412.
[51] J. Lin, L. Ren, J. Lu, J. Feng, and J. Zhou, "Consistent-aware deep learning for person re-identification in a camera network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 6, Jul. 2017, pp. 5771–5780.
[52] J. Chen, Y. Wang, J. Qin, L. Liu, and L. Shao, "Fast person re-identification via cross-camera semantic binary transformation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 3873–3882.
[53] D. Li, X. Chen, Z. Zhang, and K. Huang, "Learning deep context-aware features over body and latent parts for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 384–393.
[54] S. Bak and P. Carr, "One-shot metric learning for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 1571–1580.
[55] S. Zhou, J. Wang, J. Wang, Y. Gong, and N. Zheng, "Point to set similarity based deep feature learning for person reidentification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 6, Jul. 2017, pp. 3741–3750.
[56] Z. Zhong, L. Zheng, D. Cao, and S. Li, "Re-ranking person re-identification with k-reciprocal encoding," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3652–3661.
[57] S. Bai, X. Bai, and Q. Tian, "Scalable person re-identification on supervised smoothed manifold," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3356–3365.
[58] Z. Zhou, Y. Huang, W. Wang, L. Wang, and T. Tan, "See the forest for the trees: Joint spatial and temporal recurrent neural networks for video-based person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6776–6785.
[59] H. Zhao *et al.*, "Spindle net: Person re-identification with human body region guided feature decomposition and fusion," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 1077–1085.
[60] A. Barman and S. K. Shah, "SHaPE: A novel graph theoretic algorithm for making consensus-based decisions in person re-identification systems," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 1124–1133.
[61] X. Qian, Y. Fu, Y.-G. Jiang, T. Xiang, and X. Xue. (2017). "Multi-scale deep learning architectures for person re-identification." [Online]. Available: https://arxiv.org/abs/1709.05165
[62] 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. Comput. Vis. (ICCV)*, Oct. 2017, pp. 3980–3989.
[63] S. Xu, Y. Cheng, K. Gu, Y. Yang, S. Chang, and P. Zhou. (2017). "Jointly attentive spatial-temporal pooling networks for video-based person re-identification." [Online]. Available: https://arxiv.org/abs/1708.02286
[64] L. Zhao, X. Li, J. Wang, and Y. Zhuang. (2017). "Deeply-learned part-aligned representations for person re-identification." [Online]. Available: https://arxiv.org/abs/1707.07256
[65] Z. Zheng, L. Zheng, and Y. Yang. (2017). "Unlabeled samples generated by gan improve the person re-identification baseline *in vitro*." [Online]. Available: https://arxiv.org/abs/1701.07717
[66] J. Zhou, P. Yu, W. Tang, and Y. Wu, "Efficient online local metric adaptation via negative samples for person reidentification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, vol. 2, no. 6, Oct. 2017, pp. 1–7.

YANG LIU received the B.S. degree from the School of Advanced Engineering, Beihang University, Beijing, China, in 2009, where he is currently pursuing the Ph.D. degree with the School of Computer Science and Engineering. His research interests include deep learning, computer vision, and especially person re-identification.

HAO SHENG received the B.S. and Ph.D. degrees from the School of Computer Science and Engineering, Beihang University, Beijing, China, in 2003 and 2009, respectively. He is currently an Associate Professor with the School of Computer Science and Engineering, Beihang University. His research interests include computer vision, pattern recognition, and machine learning.



YANWEI ZHENG received the B.S. degree from Shandong Jianzhu University, Jinan, China, in 1999, and the M.S. degree from Shandong University, Jinan, in 2004. He is currently pursuing the Ph.D. degree with the School of Computer Science and Engineering, Beihang University, Beijing, China. He joined the University of Jinan, Jinan, where he became a Lector from 2004 to 2013. His research interests include machine learning, computer vision, and especially person re-identification.

NENGCHENG CHEN received the M.S. degree in geographical information system (GIS) and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, in 2000 and 2003, respectively. Since 2010, he has been leading the Smart City Team, Wuhan University, where he is currently a Luojia Scholar Distinguished Professor of geographic information science with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing. He is also with the Collaborative Innovation Center of Geospatial Technology, Wuhan. He has authored or co-authored five books and over 200 research papers. His current research interests include sensor Web, big data, real-time GIS, and smart city.

WEI KE received the Ph.D. degree from the School of Computer Science and Engineering, Beihang University. He is currently an Associate Professor of computing program with the Macao Polytechnic Institute. His research interests include programming languages, image processing, computer graphics, and tool support for object-oriented and component-based engineering and systems. His recent research focuses on the design and implementation of open platforms for applications of computer graphics and pattern recognition, including programming tools, environments, and frameworks.

ZHANG XIONG received the B.S. degree from Harbin Engineering University in 1982 and the M.S. degree from Beihang University, Beijing, China, in 1985. He is currently a Professor with the School of Computer Science and Engineering, Beihang University. His research interests include computer vision, information security, and data vitalization.