Committee-Voting mechanism based on Graph for Semi-supervised Person Re-identification

Lin Zhang\textsuperscript{1,a}, Xiuyang Zhao\textsuperscript{2,b}\* \\
\textsuperscript{1} Shandong Provincial Key Laboratory of Network based Intelligent Computing, University of Jinan, 250022 \textsuperscript{a}email: 201921200678@mail.ujn.edu.cn \textsuperscript{b}email: zhaoxy@ujn.edu.cn

Abstract. Recently, person re-identification technology has made great progress, mainly thanks to a large amount of annotated data. However, the current million-dollar annotations are becoming increasingly difficult to scale. Therefore, we work on the problem of semi-supervised person re-identification using only a small amount of annotated data. Inspired by QBC, we propose the committee-voting mechanism based on graph (CVG), which consists of two modules, the committee, and the chairman. The committee proposes different opinions and constructs multiple sets of information pairs for unlabelled data, and the chairman is responsible for aggregating the opinions of all committee members and making the final decision. As a sound selection mechanism, which effectively improves the accuracy of semi-supervised person re-identification. Extensive experiments on three large-scale datasets demonstrate the effectiveness of the proposed method.

1. Introduction
Person re-identification is a technique that analyzes whether people in multiple cameras are related. Much progress has been made in pedestrian re-recognition, and the performance of the method relies heavily on a large number of manually annotated images. However, the high cost of data annotation makes it difficult to extend the person re-identification technique to practical applications. Therefore, we work on semi-supervised person re-identification to maximize the utility of unlabelled data so that the final performance can approach that of all labeled samples.
Semi-supervised learning has become a hot topic of research in machine learning, which can combine supervised and unsupervised learning while using both labeled and unlabelled data to train and learn together. Wu et al [1] assumed the case of only a small number of pedestrian labeled samples and migrated information through several different models trained on other labeled datasets, which can be considered as multiple teacher models. The method uses a teacher-student network training mechanism to train the network by training an updated student network using a large amount of unlabeled data and a small number of labeled samples to determine the weights of each model trained on the source domain. Xin et al [2] used a small amount of labeled data to train the model, then clustered the unlabelled data using a multi-view clustering method, then updated the network by combining the labeled data and the unlabelled data with pseudo-labeling, and continued to perform the clustering algorithm again based on the new network, the whole process being iterative and alternating. Wu et al [3] proposed an incremental learning approach to solve this problem: first, an initial model is trained based on a small amount of labeled data for each individual, and then pseudo labels are
assigned to the higher confidence data among a large amount of unlabelled data, while the lower confidence data are left unlabelled. During the training process, the network is updated using traditional cross-entropy loss updates with both labeled and pseudo labeled data. For the unlabelled data, each individual sample is then treated as a class and trained in the network using an instance classification method, and the pseudo-labels are reassigned again based on the trained network. The learning process for this method is also iterative, but there are limitations to this scenario that make it difficult to extend it to practical applications. This is because in this setting it is often difficult to obtain the number of all pedestrians in the entire dataset unless the entire dataset is labeled, which would be costly in terms of labor and would be contrary to the original intention of the semi-supervised setting.

Inspired by QBC [4], we proposed a committee-voting mechanism based on graph (CVG), through two well-designed modules, the committee, and the chairman. The committee includes multiple committee member models. The committee puts forward various opinions on the characteristic information of unlabeled data, and they will give multiple sets of data sample pairs that they agree with. Subsequently, we collect the opinions of the committee and the chairman integrates the opinions of the committee modules and makes the final decision. The final positive sample is the sample pair instance most agreed by all members of the committee. Firstly, the CDP is constructed as a committee using several pedestrian recognition models, with each member of the committee providing voting information on the sample pair pairs connected in the base model: whether they are adjoining, similarity, etc. A chairman model (multi-layer perceptron) is then used to integrate this information and predict whether the sample pairs are the same person. Secondly, all positive sample pairs are then formed into a robust graph called the voting-information graph. Finally, the DBSCAN clustering method is used to generate pseudo-labels on the Voting-information graph. After CVG, a small amount of labeled data is combined with a large number of pseudo-labels, and our model is trained together until convergence.

In conclusion, we propose a committee-voting mechanism based on graph (CVG), which consists of two modules: the committee and the chairman. By aggregating voting information from multiple perspectives, which can reliably learn discriminative feature information and then generate more robust pseudo-labels. The proposed method effectively improves the accuracy of recognition, which plays a greater role in person re-identification.

2. Materials and Methods

We begin with an overview of the proposed method, which consists of three stages: 1) supervised initialization, 2) committee-voting mechanism based on graph (CVG), and 3) joint training using labeled and unlabeled data.

2.1. Supervised initialization

Given a small set of labeled data, we train a base model B and N committee models \{C_i \mid i = 1, 2, \ldots, N\}. Using the labeled data \(D_l\), the base model and all N committee members learn a Mapping from the image space to the feature space \(Z\), which can be expressed respectively as \(F_B : D_l \rightarrow Z\) and \(F_{C_i} : D_l \rightarrow Z\), \(i = 1, 2, \ldots, N\).

2.2. Committee-voting mechanism based on graph

In this section, we formally introduce the detailed steps of CVG. As shown in Figure 1, four stages are involved in the proposed method: 1) constructing the A-S graph, 2) collecting committee comments, 3) predicting voting-information graphs by the chairman, and 4) clustering to generate pseudo-labels.
2.2.1. Step1: Constructing the A-S graph. Taking the unlabelled data Du as input, features F_B (Du) and F_Ci (Du) are then extracted using the base model and all committee models. With these features, we construct A-S graphs for each sample in Du by cosine similarity, including adjacency and similarity information. With the base model B for and Ci corresponding to each committee model, for a total of N+1 graphs. Each node in the graph represents unlabelled data separately, and each edge defines a data sample pairs.

2.2.2. Step2: Collecting committee comments. Committee members \{Ci | i = 1, 2, ..., N\} learn the different feature information \{ F_Ci | i = 1, 2, ..., N\} of unlabeled data and provide the following voting information.

**Adjacency.** Assume that the graph created by the base model has two arbitrarily connected nodes n0 and n1, 1 if both nodes are adjacent in the A-S graph of all committee models and 0 otherwise. Constructing the 0-1 adjacency graph \( W_{ij} \) according to the following formula.

\[
W_{ij} = \begin{cases} 
1 & \text{if } n0 = n1 \\
0 & \text{if } n0 \neq n1
\end{cases}
\]  
(1)

**Similarity.** The similarity between two nodes A(n0,n1); The similarity between a node and other nodes A (x, xi). The cosine similarity is used as a measure, i = 1, 2, ..., N.

\[
A(n0, n1) = \cos(F_{ci}(n0), F_{ci}(n1))
\]  
(2)

\[
A(x, x_i) = \cos(F_{ci}(x), F_{ci}(x_i))
\]  
(3)

2.2.3. Step3: Predicting Voting-information graphs by the chairman. The chairman collects and weighs the votes of the committee members, in the end, gives a decisive prediction. A multi-layer perceptron (MLP) is used to determine whether sample pairs are the same person, and all positive pairs are formed into a new graph called a voting-information graph.

2.2.4. Step4: Clustering to generate pseudo-labels. The pseudo-labels can be quickly clustered on a voting-information graph using DBSCAN cluster algorithms.

2.3. Joint training using labeled and unlabeled data
In the last stage of our method, we jointly use a small portion of labeled data and a large amount of pseudo-labeled data obtained through CVG to train and optimize the model. The overall optimization objective is:
\[ L = \lambda \ell (x_l, y_l) + (1 - \lambda) \ell (x_u, y_u) \]  

(4)

where the loss, \( \ell (\cdot) \) is the softmax loss function, \{xl, yl\} denotes labeled data and labels, while \{xu, yu\} denotes unlabeled data and the pseudo labels. \( \lambda \in (0, 1) \) is the weight to balance the two components.

### 3. Results & Discussion

For the performance of person re-identification algorithms, the evaluation metrics usually used cumulative match characteristic (CMC) curves and mean average precision (MAP). The CMC curve can be given in the form of Rank-k accuracy, the probability that the correct match of a target appears in the first k positions of the match list, usually focus on the \( k = \{1, 5, 10\} \) accuracy of the matching target. Our experiments evaluate three datasets, including DukeMTMC [5], Market-1501 [6], MSMT17 [7].

In our experiment, we use ResNet50 as the backbone model, which is pre-trained on ImageNet. The last stride of ResNet50 is set to from2 to 1. All images are resized to 256 × 128 before being fed into the networks. During training, we use the Adam[8] with weight decay 0.0005 to optimize the parameters for 60 epochs, the initial learning rate is set to \( 3 \times 10^{-4} \) and decays to \( 3 \times 10^{-5} \) after 40 epochs.

To select the committee model with better performance, we conducted training and testing on three common data sets (under full supervision), using a variety of popular convolutional neural network architectures including ResNet50, DenseNet121, and VGG16. The performance on dataset DukeMTMC, Market-1501, and MSMT17 are listed in Table 1, Table 2, and Table 3. Experimental results show that ResNet50 performs more competitively.

**Table 1. Performance of different network architectures on DukeMTMC.**

| Architectures | Rank-1 | Rank-5 | Rank-10 | MAP  |
|---------------|--------|--------|---------|------|
| ResNet50      | 84.6%  | 92.5%  | 94.6%   | 70.5%|
| DenseNet121   | 83.7%  | 91.7%  | 93.9%   | 70.1%|
| VGG16         | 77.7%  | 86.7%  | 89.9%   | 58.3%|

**Table 2. Performance of different network architectures on Market-1501.**

| Architectures | Rank-1 | Rank-5 | Rank-10 | MAP  |
|---------------|--------|--------|---------|------|
| ResNet50      | 93.0%  | 97.5%  | 98.6%   | 81.2%|
| DenseNet121   | 92.1%  | 97.3%  | 98.2%   | 79.6%|
| VGG16         | 77.7%  | 86.7%  | 89.9%   | 58.3%|

**Table 3. Performance of different network architectures on MSMT17.**

| Architectures | Rank-1 | Rank-5 | Rank-10 | MAP  |
|---------------|--------|--------|---------|------|
| ResNet50      | 73.8%  | 84.8%  | 88.2%   | 45.8%|
| DenseNet121   | 70.9%  | 83.1%  | 87.2%   | 43.3%|
| VGG16         | 60.7%  | 74.8%  | 79.7%   | 31.0%|

In our experiments, we also explored options for the number of committee members from 0 to 10. As shown in Table 4, the best number of committee members was 8, and the results showed that only 10% of the labeled data were used to achieve results equivalent to fully supervised counterparts.
Table 4. Performance of our method on Market-1501 with different numbers of the committee.

| Committee number | Percentage of labeled data | Rank-1  | Rank-5 | Rank-10 | mAP  |
|------------------|---------------------------|---------|--------|---------|------|
| 0                | 10%                       | 70.8%   | 71.7%  | 74.1%   | 61.2%|
| 2                | 85.5%                     | 87.1%   | 89.6%  | 70.0%   |
| 4                | 89.6%                     | 92.7%   | 94.1%  | 74.9%   |
| 6                | 91.4%                     | 95.8%   | 96.9%  | 78.7%   |
| 8                | 92.7%                     | 97.4%   | 98.5%  | 81.1%   |
| 10               | 83.1%                     | 85.5%   | 87.8%  | 69.4%   |

4. Conclusion
We propose a new semi-supervised person re-identification method, committee-voting mechanism based on graph (CVG) with two modules: the committee and the chairman, which efficiently makes full use of multiple committee models to obtain information on different features, and the chairman can integrate multiple sets of voting information to make the final prediction. Our experiment results showed that only 10% of the labeled data were used to achieve results equivalent to fully supervised counterparts.

Acknowledgments
This research was supported by Project of Jinan Scientific Research Leader's Laboratory (No. 2018GXRC023), Natural Science Foundation of Shandong province (Nos. ZR2019MF013, ZR2019BF026).

References
[1] Wu A, Zheng, W S, Guo, X, and Lai, J H 2019 Distilled person re-identification: Towards a more scalable system In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition pp 1187-1196
[2] Xin X, Wang J, Xie R, Zhou S, Huang W, and Zheng N 2019 Semi-supervised person re-identification using multi-view clustering Pattern Recognition 88 285-297
[3] Wu Y, Lin Y, Dong X, Yan Y, Bian W, and Yang Y 2019 Progressive learning for person re-identification with one example IEEE Transactions on Image Processing 28(6) 2872-2881
[4] Thompson J N 1984 Insect diversity and the trophic structure of communities Ecological entomology 591-606
[5] Zheng Z, Zheng L, and Yang Y 2017 Unlabeled samples generated by gan improve the person re-identification baseline in vitro In Proceedings of the IEEE International Conference on Computer Vision pp 3754-3762
[6] Zheng L, Shen L, Tian L, Wang S, Wang J, and Tian Q 2015 Scalable person re-identification: A benchmark In Proceedings of the IEEE international conference on computer vision pp 1116-1124
[7] Wei L, Zhang S, Gao W, and Tian Q 2018 Person transfer gan to bridge domain gap for person re-identification In Proceedings of the IEEE conference on computer vision and pattern recognition pp 79-88
[8] KingA D A 2015 A method for stochastic optimization Anon International Conference on Learning Representations SanDego: ICLR