DoT-GNN: Domain-Transferred Graph Neural Network for Group Re-identification

Ziling Huang
National Tsing Hua University
huangziling@gapp.nthu.edu.tw

Zheng Wang
National Institute of Informatics
wangz@nii.ac.jp

Wei Hu
Peking University
forhuwei@pku.edu.cn

Chia-Wen Lin
National Tsing Hua University
cwlin@ee.nthu.edu.tw

Shin'ichi Satoh
National Institute of Informatics
The University of Tokyo
satoh@nii.ac.jp

ABSTRACT

Most person re-identification (ReID) approaches focus on retrieving a person-of-interest from a database of collected individual images. In addition to the individual ReID task, matching a group of persons across different camera views also plays an important role in surveillance applications. This kind of Group Re-identification (G-ReID) task is very challenging since we face the obstacles not only from the appearance changes of individuals, but also from the group layout and membership changes. In order to obtain robust representation for the group image, we design a Domain-Transferred Graph Neural Network (DoT-GNN) method. The merits are three aspects: 1) *Transferred Style.* Due to the lack of training samples, we transfer the labeled ReID dataset to the G-ReID dataset style, and feed the transferred samples to the deep learning model. Taking the superiority of deep learning models, we achieve a discriminative individual feature model. 2) *Graph Generation.* We treat a group as a graph, where each node denotes the individual feature and each edge represents the relation of a couple of individuals. We propose a graph generation strategy to create sufficient graph samples. 3) *Graph Neural Network.* Employing the generated graph samples, we train the GNN so as to acquire graph features which are robust to large graph variations. The key to the success of DoT-GNN is that the transferred graph addresses the challenge of the appearance change, while the graph representation in GNN overcomes the challenge of the layout and membership change. Extensive experimental results demonstrate the effectiveness of our approach, outperforming the state-of-the-art method by 1.8% CMC-1 on Road Group dataset and 6.0% CMC-1 on DukeMCMT dataset respectively.

CCS CONCEPTS

• **Information systems** → **Multimedia content creation**; • **Security and privacy** → *Social aspects of security and privacy*; • **Computing methodologies** → *Image representations.*

KEYWORDS

Group re-identification, Graph Generation, Graph Neural Network

ACM Reference Format:
Ziling Huang, Zheng Wang, Wei Hu, Chia-Wen Lin, and Shin'ichi Satoh. 2019. DoT-GNN: Domain-Transferred Graph Neural Network for Group Re-identification. In *Proceedings of the 27th ACM International Conference on Multimedia (MM '19), October 21–25, 2019, Nice, France.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3343031.3351027

1 INTRODUCTION

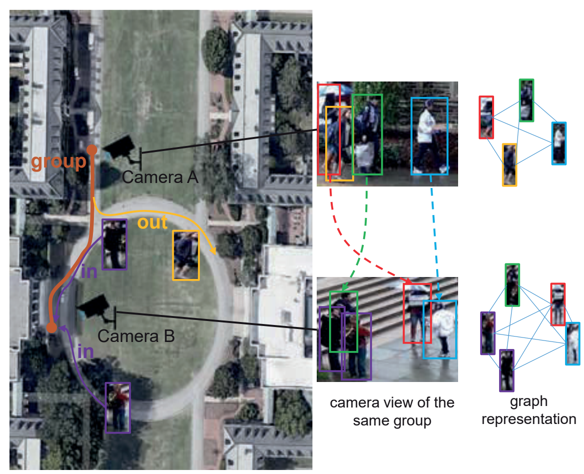


Figure 1: Illustration of the G-ReID task and its challenges. The persons with the red, blue and green bounding boxes change their locations in a group. The persons with purple bounding boxes join in the group. The person with the yellow bounding box leaves out of the group. Besides the challenge of the appearance change, G-ReID further brings in the challenges of group layout and membership changes.

Person re-identification (ReID) has been drawing a lot of attentions [1, 5, 22, 23, 25, 31, 35] due to its wide-range applications such as security and surveillance [15, 21]. Existing research mainly focused on re-identifying individuals, while searching out a certain group of persons simultaneously was relatively rarely studied. Actually, a group of persons moves around a street together is very common. As illustrated in Figure 1, a group of persons walked from the view of camera *A* to that of camera *B*. The system requires an

* Corresponding Author: Zheng Wang and Wei Hu

MM '19, October 21–25, 2019, Nice, France

ACM ISBN 978-1-4503-6889-6/19/10…$15.00
https://doi.org/10.1145/3343031.3351027

algorithm to re-identify a group of persons across different cameras (called Group ReID or G-ReID). In practice, G-ReID is becoming more and more important in daily life, which is a powerful supplement to the individual ReID.

Unlike individual ReID, the target of G-ReID is to associate a certain group under different camera views. Besides the traditional challenges in individual ReID such as low-resolution [24, 27], pose variation [13, 30], illumination variation [26, 32], and blurred vision, G-ReID has its own unique challenges. As Figure 1 shows, when the group walked from camera A to camera B, 1) the persons change their locations in the group (called *group layout change*), 2) some persons join and leave the group dynamically (called *group membership change*). That is to say, G-ReID is a more challenging task, as a group has deformable characteristics. Hence, treating the group as a whole and extracting its global or semi-global features as [12] may not do a good job, because the group layout and membership changes can significantly alter the visual content of the group.

As a group is made up of several individuals, its representation can be a combination of representations of individuals and their relationships. It inspires us to use *graphs* to construct the whole representation. In particular, we choose to employ an undirected graph to represent the pair-wise symmetric relationship in the group image and exploit a Graph Neural Network (GNN) model [9] to identify the graph ID. By representing relationships among persons on a graph, GNN is able to extract group features via graph convolution, i.e., message passing among neighboring nodes. Further, GNN is suitable for addressing the challenge of group layout changes and membership changes, by offering a flexible representation of relationships in each group on graphs.

Let graph $G = (V, E)$ denote a group image, where V represents the set of nodes and E represents the set of edges. Each node v_i denotes the representation of person i, and each edge e_j denotes the representation of the relationship of person pair j. In order to re-identify a probe graph G_p with a gallery group G_g, we need to evaluate their similarity as $\mathcal{S}(\mathcal{M}(G_p), \mathcal{M}(G_g))$. Here, $\mathcal{M}(\cdot)$ is the graphical model used to characterize a graph, and $\mathcal{S}(\cdot, \cdot)$ denotes the similarity metric of two graphs. In this way, we need to construct plenty of informative graph samples to train the graphical model $\mathcal{M}(\cdot)$ so as to achieve better G-ReID performance. However, some key challenges need to be coped with for G-ReID, as listed in Table 1.

Table 1: Comparison of the challenges in ReID and G-ReID and our strategies for overcoming the individual challenges.

Challenge	ReID	G-ReID	Strategy
Training Set	Abundant	Insufficient	
Appearance	✓	✓	Node generating (transfer)
Layout	×	✓	Membership-preserving grouping
Membership	×	✓	Membership-varying grouping

• **Training data deficiency** & **appearance change.** G-ReID usually suffers from the training data deficiency problem, i.e., the quantity of labeled group images with group IDs is not sufficient to learn a robust group representation model. Since it is difficult to acquire training data for representation learning of a group, [28] exploited hand-crafted features to represent persons in a group. Nevertheless, the hand-crafted representations cannot effectively tackle the appearance change problem. As a result, when representing a group image with a graphical model, the node of the graph cannot be expressed well using the hand-crafted features. As we know, there exist rich amounts of training datasets suitable for general ReID, which motivates us to make use of existing labeled ReID samples to learn node features. However, the domain gap between the ReID training datasets and the target G-ReID images often cause a severe performance drop. In order to compensate for the domain shift, inspired by [36, 37], we propose to transfer the image style of a ReID dataset to that of the target G-ReID dataset while preserving individuals' identities. In this way, the features of individual persons in a group (nodes of the graph) can be properly extracted by our transferred representation model.

• **Layout change** & **Membership change.** Taking the features of transferred samples as the signal on graph nodes, we construct the graph samples for training the graphical models of group images. As the images of the same group involve layout and membership changes, for each graph class ID, we build its training graph samples considering the variations from both samples of the same identity and different identities. We propose a membership-preserving grouping strategy and a membership-varying grouping strategy to construct sufficient graph samples for training the graphical models of group images.

After acquiring graph samples, it is vital to learn a graphical model so that graphs with different IDs can be separated apart automatically. Since graphs are irregular structures with unordered nodes for non-Euclidean data, it is inappropriate to construct graphs via Convolution Neural Networks (CNNs), which can only handle regular-structure data such as images (2D) and texts (1D). GNN [9] is a novel deep learning based method that operates in the graph domain by defining or approximating graph convolution and pooling, which is a suitable tool to learn the graphical model of irregular-structure data. Hence, we employ GNN to acquire group features.

Based on the considerations above, we propose a novel Domain-Transferred Graph Neural Network (DoT-GNN) model for group re-identification, that can stimulate graphs by transferred samples. After that, GNN is exploited to learn group features for the identification of the corresponding group. In summary, our contributions lie in three aspects:

- We address the irregularity problem with G-ReID by exploiting the idea of graphical representation and modeling. To the best of our knowledge, we are the first to propose a GNN-based framework for G-ReID. Although deep learning models have their superiority, no effective deep learning models have been proposed for G-ReID, let alone deep GNN with transfer learning.
- We propose a domain-transferred graph node construction method and two grouping strategies for preserving and varying group membership to overcome the challenges of individuals' appearance and membership changes in G-ReID with very limited training data. The graph construction process benefits from transferred graph nodes.
- We demonstrate the effectiveness of our proposed method on challenging G-ReID datasets like **DukeMTMC Group** and **Road Group**.

2 RELATED WORKS

Deep learning based person re-identification. Deep learning based approaches have been extensively studied in general ReID field. For example, [11] proposed a filter pairing neural network to jointly handle misalignment and geometric transforms. In order to learn features from multiple domains, [29] utilized a domain-guided dropout algorithm to improve the feature learning procedure. Moreover, the method proposed in [16] makes full use of human part cues to alleviate the pose variations and learn robust representations from both the whole image and its different local parts. However, these supervised learning based works all require abundant labeled training data. Moreover, all of these works mainly focused on individual person re-identification. None of them paid attention to G-ReID with very limited training data.

Group re-identification. Compared with individual ReID tasks, relatively fewer existing works focused on G-ReID tasks [4, 12, 28, 34, 38]. Some of them mainly attempted to extract global or semi-global features for G-ReID. For example, [4] proposed a discriminative covariance descriptor to obtain both global statistical features. [34] proposed semi-global features by segmenting a group image into many ring regions. However, since persons in a group often change their locations under different views (i.e., layout-change), these global and semi-global features are usually sensitive to such changes. In order to take advantage of individuals' features in the groups, [38] introduced patch matching between two group photos. However, it requires the matched group images to be well aligned vertically in advance, making it unworkable under certain circumstances. The method proposed in [28] leveraged multi-grain information and attempted to fully capture the characteristics of a group. This approach, however, involves too much redundant information and relies on hand-crafted features, thereby making its accuracy unsatisfactory. In this paper, we choose to employ a graph to represent the group image. The graph's superiority lies in establishing the membership of the individuals in the group image, so as to overcome the challenges in G-ReID.

Graph Neural Networks. CNN has revolutionized many fields, such as computer vision [8, 10, 18] and natural language processing [17] because it can learn and extract informative features to replace traditional hand-crafted features. However, CNN can only handle regular-structure data in the Euclidean domain. There are increasing needs of representing non-Euclidean data, such as irregular graph-structured data. For example, in chemistry, molecules and their structures are represented on graphs and graph classification is used to identify their pharmaceutical properties [6]. In connectomics, neuronal pathways or functional connections between brain regions are commonly modeled as graphs [19] to obtain disease signals. Due to the increasing needs to handle non-Euclidean data, GNNs have been proposed to extend CNNs to the non-Euclidean domain [3] by defining graph convolution and pooling. [7] utilized GNNs to compute the embeddings of out-of-knowledge-base entities, exploiting the limited auxiliary knowledge provided at test time. [2] proposed to encode the full structural information contained in the graph. Their architecture couples the Gated Graph Neural Networks with an input transformation that allows nodes and edges to have their own hidden representations, while tackling the parameter explosion problem. [20] presented graph attention networks (GATs) operating on graph-structured data, leveraging masked self-attentional layers to address the shortcomings of prior methods based on graph convolutions or their approximations. In our work, GNN is proposed to classify the graphs constructed for representing groups of persons so as to learn the features of each group.

3 OUR METHOD

In the G-ReID task, we have a probe image p, which is represented as a graph G_p. The corresponding group of probe image p should be found in gallery images $\mathcal{G} = \{g_n\}_{n=1}^{N}$, where N stands for the number of images in the gallery. For each gallery image g_n, it can be also represented as a graph G_{g_n}.

3.1 Proposed Framework

The proposed framework is presented in Figure 2, which consists of a training step and a testing step. In the training step, the framework consists of a domain-transferred model that is responsible for transferring the source-domain individuals' images to their corresponding target-domain ones, a graph generator for constructing the pool of graph samples with the transferred individuals' representations as nodes, and a GNN model trained on the pool of graph samples for classifying the group IDs. In the testing step, we extract features from probe image p and the gallery images $\mathcal{G}$ via the GNN model, and then calculate the distances between the probe feature and the gallery features so as to re-identify the group ID of the probe image according to the distances.

3.2 Domain-Transferred Model

The total number of groups in a collection of G-ReID images is usually rather limited, so it is difficult to train a useful network directly solely based on those data themselves. To learn better representations, we should make use of external information. There exists a rich collection of ReID datasets which can be used to train good feature representations for individuals. Nevertheless, the domain gap between the existing ReID datasets and the target G-ReID images, that is caused by their different capturing conditions, usually significantly degrades the performance of representation learning. To address this problem, given a training ReID dataset $\mathcal{S}$, we propose utilizing domain transfer to learn a mapping function $T : \mathcal{S} \rightarrow \mathcal{T}$ from the style of ReID dataset $\mathcal{S}$ to that of G-ReID dataset $\mathcal{T}$ so that the distribution of $T(\mathcal{S})$ can be indistinguishable from that of dataset $\mathcal{T}$. In this paper, the mapping function is implemented by CycleGAN [39].

In this way, the dataset $T(\mathcal{S})$, where $y_k^s \in T(\mathcal{S})$ denotes the k-th image of the s-th person in the dataset, can be used to train the CNN for representing individuals' features. Then, the domain-transferred individuals' features extracted by the CNN serve as the graph signals on nodes for graph construction, where graph signal $\{v_k^s\}$ denotes the feature of the k-th image of the s-th person.

3.3 Graph Generator

We can obtain additional useful information from neighboring group members. In our work, each group image is represented as a graph to characterize the mutual and global relations of persons. Each node in the graph represents one person in the group,

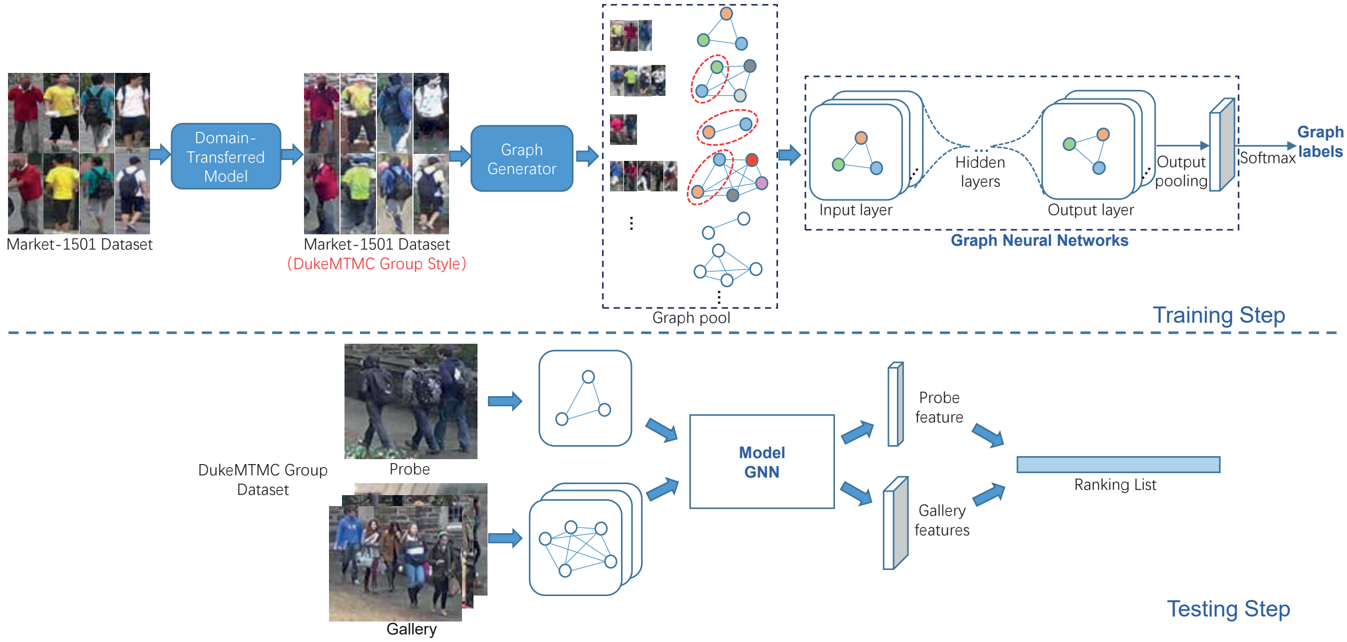


Figure 2: Proposed architecture of DoT-GNN. We take DukeMTMC group dataset as the example target-domain dataset. In the training step, we first transfer the style of the source-domain dataset (e.g., Market-1501) to that of the target one (e.g., DuketMTMC). We then construct a graph sample pool based on the transferred individual samples (nodes). After that, we train the GNN on the constructed graph samples. In the testing step, the GNN model first extracts features from the probe and gallery images, then calculates the distances between the probe feature and the gallery features, and finally determines the group ID according to the distance.

and the edge between two nodes indicates the intimacy between two persons, which are measured by the similarity between the two persons' features based on a predefined similarity metric. If we use a graph to represent a group image, we can search for a target group based on the similarity measurements between graphs. In our work, a GNN is employed to represent graph features for measuring the graph similarity. In order to train the GNN, we use the domain-transferred nodes to construct a pool of graph samples. For a graph G_x, that contains multiple nodes, edges and its corresponding label l_x. For each kind of label l_x, we construct multiple graph samples $\{G_x^i\}_{i=1}^{N_x}$. The generator employs two kinds of strategies to construct graph samples, i.e., membership-preserving grouping, and membership-varying grouping.

Membership-preserving grouping. We simulate a graph sample by the transferred images feature (generated nodes $\{v_k^s\}$). If two group images contain the same members, these two images are labeled as the same group. As we know, the same group may exhibit layout change when moving from one camera to another. We use an undirected graph to tackle the problem of layout and appearance changes. However, for member s_i, we need to pick up different nodes $\{v_{k_j}^{s_i}\}$ for different graph samples with the same individual label y^{s_i}.

Here, we propose a membership-preserving grouping strategy to generate graph samples with layout and appearance changes, but without membership changes, as shown in the second row of Figure 3. The strategy is described as follows. First, for a graph class l_x, we randomly choose a number of persons $\{s_i\}, i \in [1, N_s]$ as the members of the group. Second, for each member s_i, we randomly select one node $v_{k_j}^{s_i}$ from those nodes associated with this member. Third, the nodes of different members together constitute a graph sample G_x^j, where the edge weights between every two nodes are calculated. In this way, we can also construct graph samples with the same set of group members.

Membership-varying grouping. As mentioned above, a group may also have membership change dynamically, implying that graph samples corresponding to the same label l_x may contain nodes not exactly the same as each other. In order to cope with such dynamic membership changes in group images, we calculate the membership similarity ratio between two graph samples, which is defined as the percentage of the nodes of the two graphs being from the same common group members. If two graph samples have a membership similarity ratio, they are considered to share the same group ID. For graphs G_a and G_b, we denote $G_a \bigcap G_b$ as the number of common members in these two graphs and $G_a \bigcup G_b$ as the total number of members that constitute these two graphs. Then, the membership similarity ratio is defined as

$$r = \frac{G_a \bigcap G_b}{G_a \bigcup G_b}. \tag{1}$$

In G-ReID, if the similarity ratio of graph G_a and G_b is larger than a threshold r_0, the corresponding group images are treated as the same group. We propose a membership-varying grouping strategy

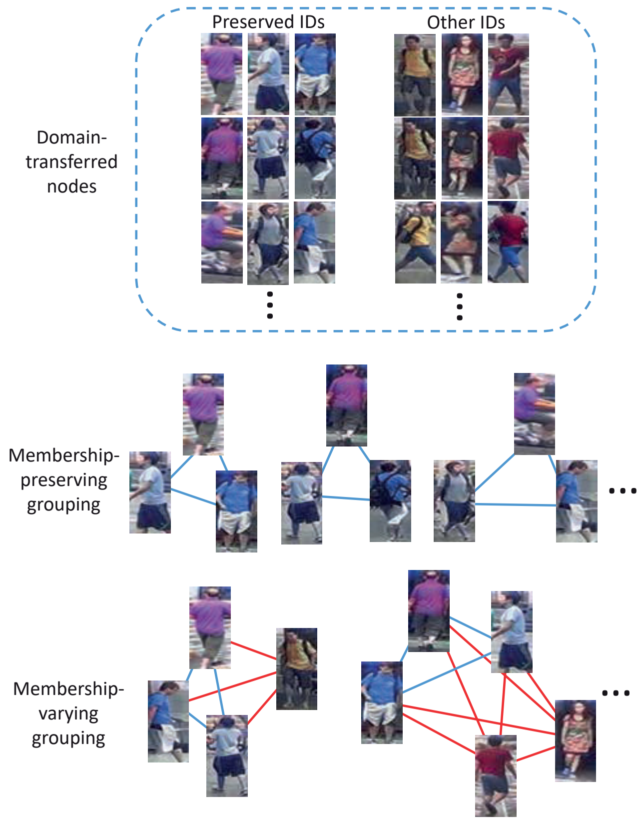


Figure 3: Graph samples construction process. The graph samples are constructed based on the domain-transferred nodes. The whole process includes two grouping strategies. The second row indicates the membership-preserving grouping strategy, where the nodes of fixed persons with preserved IDs are selected, i.e., the group members only changes their layout and appearance, without any membership chance. The third row illustrates the membership-varying grouping strategy, where unfixed members may join or leave the group randomly to capture the behavior of dynamic membership changes in group images.

to construct graph samples with such dynamic group membership changes, as illustrated in the third row of Figure 3. We first fix $r_0 * N_u$ members in one group, where N_u denotes the average total number of members constituting the graph samples of the same group. Then, we randomly add or remove members to and from the remaining unfixed members to simulate the dynamic group membership changes.

In the graph generator, we have the transferred dataset $\{v_k^s\}$ that contains N_s person IDs, and we divide these IDs into N_{sub} subsets equally, each containing $\lfloor \frac{N_s}{N_{sub}} \rfloor$ persons in total and being assigned with a unique group ID, where $\lfloor \cdot \rfloor$ means the floor function. Based on the consideration above, we assume each group has a set of $r_0 * N_u$ fixed members, that are randomly picked from its group members. These fixed members will stay in the group, but may change their positions in the group. The rest of $\lfloor \frac{N_s}{N_{sub}} \rfloor - r_0 * N_u$ unfixed group members may join or leave the group randomly. No matter how they change their positions or how many people are in the group, they share the same group ID. The whole process is shown in Figure 3. Algorithm 1 shows the whole algorithm process of graph generator.

Algorithm 1 Algorithm for graph generator

$k \leftarrow 0$
while $N_u \times (k+1) < N_s$ **do**
  $Group[k] \leftarrow range(N_u * k, N_u * (k+1))$
  $k \leftarrow k + 1$
end while
// split group
$group\_model = \{\}$
for $key \in Group$ **do**
  shuffle($Group[key]$)
  $fix\_num \leftarrow N_u * r_0$
  $fix\_ids \leftarrow Group[key][0 : fix\_num]$
  $nonfix\_ids \leftarrow Group[key][fix\_num :]$
  $group\_model[key] \leftarrow$ the subsets of all nonfix_ids plus fix_ids
end for
for $key \in group\_model$ **do**
  $i \leftarrow 0$
  while $i < len(group\_model)$ **do**
    randomaly select v for each IDs in $group[key][i]$ to construct graph, this step can repeat several times to create more unique graphs
    // membership-preserving grouping
    $i \leftarrow i + 1$
  end while
end for
// membership-varying grouping

3.4 GNN

Since groups are represented as graphs, we adopt the GNN model proposed in [9] for feature learning. As discussed in Section 3.3, the graphs representations of groups and their associated group IDs are used to train GNN. In our work, we use batch-wise classification to train GNN with an adjacency matrix each. At the end of GNN, a unique graph pooling layer is appended to collect the nodal features in graph G_x. Then, the features of the entire graph are sent to the softmax layer to calculate their probability for every class l_x. We adopt the cross-entropy as the cost function to train the GNN. The entire process is illustrated in Figure 4.

3.5 Testing Step

In the training step, a CNN model is trained for extracting the features of individuals in group images and a GNN model $\mathcal{M}$ is learned for obtaining graph-based group features from graph samples consisting of nodes of individuals' features. Given a probe image p and gallery images $\mathcal{G} = \{g_n\}_{i=1}^N$, we use CNN to obtain nodal features and construct graphs G_p and G_{g_n}, $n \in [1, N]$. Then, we utilize $\mathcal{M}$ to acquire the graph features $f_p = \mathcal{M}(G_p)$ and $f_{g_n} = \mathcal{M}(G_{g_n})$. Finally,

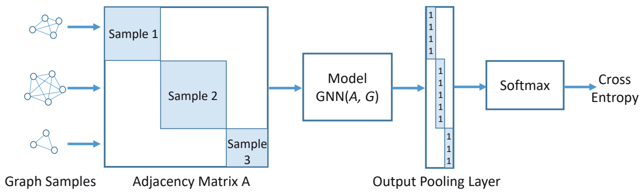


Figure 4: Training of GNN. The framework supports the batch-wise classification of multiple graph samples (of potentially different size) with an adjacency matrix each. It concatenates respective feature matrices and builds a (sparse) block-diagonal matrix A where each block corresponds to the adjacency matrix of one graph sample. We use a simple pooling matrix (output pooling layer) that collects features from the respective graph samples as graph-level outputs. The cross-entropy loss is employed after the softmax layer, and the output is the graph sample label.

we calculate the similarities between the probe and gallery images by their $L2$ distance to obtain the ranking results.

4 EXPERIMENTS

4.1 Datasets and Experimental Setting

Datasets. Our method is evaluated on two public G-ReID datasets constructed in [28]. Some examples are shown in Figure 5. The **DukeMTMC Group** dataset contains 177 group image pairs selected from an 8-camera-view DukeMTMC dataset [14], and the **Road Group** dataset contains 162 group pairs taken from a two-camera crowd road scene. Both datasets include severe object occlusions and large layout & group membership changes. Following [28], half of each dataset is evaluated under the protocol in [38], and the Cumulative Matching Characteristic (CMC) metric [28] is used for performance evaluation.

We also used the Market-1501 dataset [33] as the source-domain ReID dataset due to its large amount of training instances: 15936 images for 751 individuals. In our work, we transfer the Market-1501 dataset to the styles of **DukeMTMC Group** and **Road Group**, respectively, as illustrated in Figure 5.

Setting for Domain Transfer. In our work, we transfer the domain of an existing ReID dataset to that of the target G-ReID datasets (e.g., **DukeMTMC Group** and **Road Group**), prior to training the representations. We use the CycleGAN [39] to transfer the domain for each target G-ReID dataset . As a result, we obtain the DukeMTMC-style Market-1501 dataset and the Road-style Market-1501 dataset. In the training process, we resize all input images to 256×256 and use the Adam optimizer. The batch-size is 10, and the learning rates are 0.0002 and 0.0001 for the Generator and the Discriminator respectively.

Setting for GNN. In our work, the GNN is used to learn group features. We use the source code from [9] to do batch-wise classification. In the training process, Adam Optimizer is used. The batch-size is 12 and the learning rate is 0.0001.

Setting for Graph Generator. In our work, we generate graph samples as training data. Here, we set N_u as 15 and r_0 as 0.1. That is to say, the maximum number of people in a group is 15. No matter how we change the number of people in a group, one person stays in this group always.

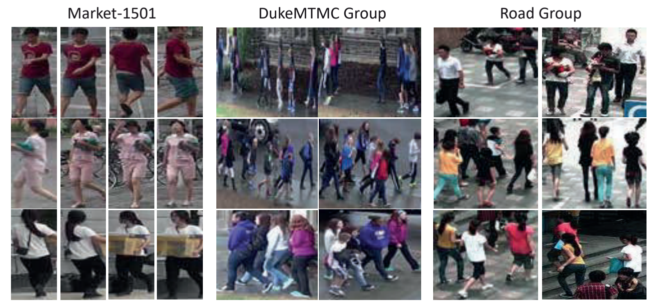


Figure 5: Snapshots of the utilized datasets. From left to right, the datasets are respectively Market-1501 (ReID), DukeMTMC Group and Road Group (G-ReID). Each row of each dataset shows a few snapshots with the same person/group ID.

Figure 6: Snapshots of domain-transferred samples. The images in the third row are cropped from the source domain Market-1501. The images in the first and fifth rows are cropped respectively from the target domain DukeMTMC Group and Road Group. The second row shows the generated images with the DukeMTMC style, and the fourth row shows the generated images with the Road style.

4.2 Performance of Image Domain Transfer

For the testing G-ReID datasets, we exploited DukeMTMC/Road samples as the target-domain samples for domain transfer. Some examples of domain-transferred samples are shown in Figure 6.

Table 2: **Comparison with the state-of-the-art G-ReID methods on the DukeMTMC and Road Group datasets.**

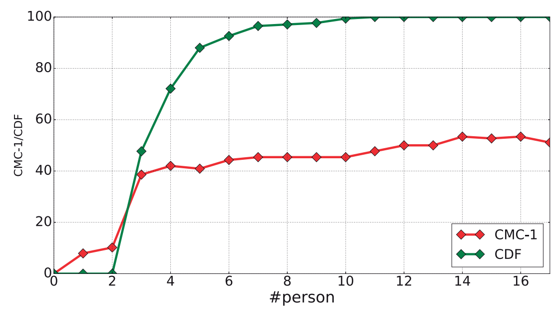
Method	DukeMTMC Group				Road Group			
	CMC-1	CMC-5	CMC-10	CMC-20	CMC-1	CMC-5	CMC-10	CMC-20
CRRRO-BRO [34]	9.9	26.1	40.2	64.9	17.8	34.6	48.1	62.2
Covariance [4]	21.3	43.6	60.4	78.2	38.0	61.0	73.1	82.5
PREF [12]	22.3	44.3	58.5	74.4	43.0	68.7	77.9	85.2
BSC+CM [38]	23.1	44.3	56.4	70.4	58.6	80.6	87.4	92.1
MGR [28]	47.4	68.1	77.3	87.4	72.3	**90.6**	**94.1**	97.5
Resnet50 + Feature Fusion	31.8	56.8	73.9	80.7	38.3	58.0	67.9	77.8
DoT + Feature Fusion	40.9	69.3	77.3	83.0	43.2	65.4	70.4	76.5
DoT + Distance Fusion	35.2	46.6	46.6	47.7	9.9	9.9	55.6	65.4
DoT + GNN	**53.4**	**72.7**	**80.7**	**88.6**	**74.1**	90.1	92.6	**98.8**

4.3 Comparison with the State-of-the-art Methods

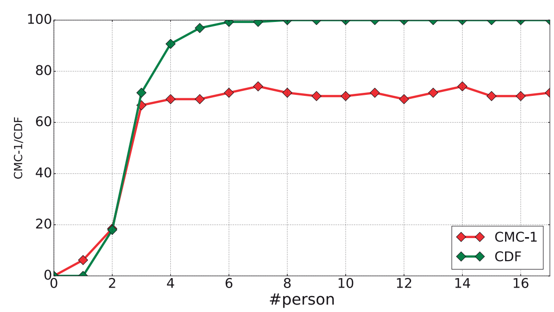
Table 2 also shows the results of some state-of-the-art methods on **DukeMTMC Group** and **Road Group**. The compared methods include CRRO-BRO [34], Covariance [4], PREF [12], BSC+CM [38] and MGR [28]. 'Resnet50' indicates that the individuals' features are extracted by the Resnet50 network, which was trained on the Market-1501 dataset without transferring. 'DoT' indicates that the person features are extracted by the domain-transferred model. 'Feature Fusion' means that the group image features are obtained by the average pooling of all individuals' features of the group. 'Distance Fusion' means that the final distance of two group images is measured by the average value of all the distances of individuals' features. Comparing the results of 'Resnet50+Feature Fusion' and 'DoT+Feature Fusion', we can see that the domain-transferred model is effective. Comparing the results of 'DoT+Feature Fusion' and 'DoT+GNN', we can see that the graph generation process and GNN are effective. Comparing the results of 'DoT+Feature Fusion' and 'DoT+Distance Fusion', we can see that early feature fusion is better than the late distance fusion for measuring the distance of graph samples. The results also show that our method outperforms existing G-ReID methods at most of the rankings. Note that, the marginal gain on the Road Group dataset is attributed to the heavy occlusions and viewpoint changes in Road Group, which limit the effectiveness of nodal representations. To overcome the changes of nodal representations, we select hard samples (the samples from the same identity but have a large distance) to construct our training graph samples. With this new strategy, we achieve 75.1%, 92.3%, 95.2%, 98.9% on corresponding CMC scores.

4.4 The influence of Graph Generator

Figure 7 shows the influence of the graph generator on the final result. In the **DukeMTMC Group** and **Road Group** datasets, the maximum numbers of persons in one group is 11 and 8, respectively. By analyzing the CMC-1 score, we can conclude that the total number of persons in our graph generation process will influence the final result. The more the total number of persons in our graph generation process, the higher the CMC-1 score we can achieve. However, if the total number of persons in our graph generation process exceeds the maximum number of persons in one group, the CMC-1 score stays stable, because our constructed graph samples are already able to simulate all circumstances.



(a) DukeMTMC Group



(b) Road Group

Figure 7: Impact of graph generation method on the final results for the DukeMTMC Group and RoadGroup datasets. The x-axis represents the total number of persons in a group. The red curve indicates the CMC-1 score and the green curve shows the Cumulative Distribution Function (CDF) for group images. From the figure, the maximum numbers of persons in DukeMTMC Group dataset and RoadGroup dataset are 11 and 8, respectively.

4.5 Ablation Study

Table 3 shows our ablation study on **DukeMTMC Group** Dataset, where $Tr.$ denotes the domain-transferred model, $S1$ stands for the membership-preserving grouping strategy and $S2$ stands for the membership-varying grouping strategy. Comparing Variant ② with Variant ③, we can find that the membership-preserving grouping strategy is more effective than the membership-varying grouping strategy, in particular for the top results, since the membership-preserving grouping is the basic one. Comparing Variants ②, ③, and ④ with Variant ⑤, we can find that both of the proposed domain-transferred model and the graph generator contribute to the final result. Comparing Variant ① with Variant ④, we can find that although the domain-transferred model can effectively address the appearance changes in G-ReID, the graph generator and GNN together perform even better, meaning that addressing the issues of layout and membership changes is more important for the G-ReID task.

Table 3: Ablation Study on the DukeMTMC Group dataset.

	Settings				DukeMTMC Group			
Variant	$Tr.$	$S1$	$S2$	GNN	CMC-1	CMC-5	CMC-10	CMC-20
①	✓	×	×	×	35.2	46.6	46.6	47.7
②	✓	✓	×	✓	44.3	72.2	78.4	86.4
③	✓	×	✓	✓	35.2	62.5	78.7	86.4
④	×	✓	✓	✓	44.3	67.0	76.1	85.2
⑤	✓	✓	✓	✓	**53.4**	**72.7**	**80.7**	**88.6**

Figure 8: Some subjective results. We choose one example from the DukeMTMC Group dataset and one example from the Road Group dataset. The figure shows their top six results, evaluating with 'DoT+Feature Fusion', 'DoT+Distance Fusion' and 'DoT+GNN' methods, respectively. Each row shows the results in one method. The first column shows the probe image. The following columns show the top six results. Among all the results, the green boxes indicate the target group image.

4.6 Subjective Comparison

Figure 8 shows some subjective results of the proposed method on both DukeMTMC Group and Road Group datasets. For each dataset, we give one example and its top six results evaluated on three kinds of methods, the former two are obtained by the 'DoT+Feature Fusion' and 'DoT+Distance Fusion' methods respectively, while the latter one is obtained by the 'DoT+GNN' method. As can be seen, 'DoT' with 'GNN' finds out the true target as the top one while the other two methods can not. It may be because GNN gives a comprehensive measurement to the nodes and relationships in the graphs, while exploiting 'Distance Fusion' or 'Feature Fusion' may lead to biased performances caused by some negative similar nodes or relationships.

5 CONCLUSION

In this paper, we addressed an important but less studied problem: group re-identification. We proposed to use node generation (transfer), membership-preserving grouping and membership-varying grouping to respectively overcome the three major challenges in group re-identification: training data deficiency, layout and appearance changes, and membership changes layout change. We have also proposed a graph neural network to learn and extract the group feature representations of the constructed graphs so as to better identify their group IDs. Experimental results show our method outperforms existing state-of-the-art approaches.

ACKNOWLEDGMENTS

The research was supported by JST CREST Grant Number JPMJCR1686, Japan, Grant in-Aid for JSPS Fellows Number 18F18378, Beijing Nature Science Foundation Number 4194080, and Ministry of Science and Technology, Taiwan Number 108-2634-F-007-009.

REFERENCES

[1] Song Bai, Xiang Bai, and Qi Tian. 2017. Scalable person re-identification on supervised smoothed manifold. In *CVPR*.

[2] Daniel Beck, Gholamreza Haffari, and Trevor Cohn. 2018. Graph-to-sequence learning using gated graph neural networks. *arXiv preprint arXiv:1806.09835* (2018).

[3] Joan Bruna, Wojciech Zaremba, Arthur Szlam, and Yann LeCun. 2013. Spectral networks and locally connected networks on graphs. *Computer Science* (2013).

[4] Yinghao Cai, Valtteri Takala, and Matti Pietikainen. 2010. Matching groups of people by covariance descriptor. In *ICPR*.

[5] Hehe Fan, Liang Zheng, Chenggang Yan, and Yi Yang. 2018. Unsupervised person re-identification: Clustering and fine-tuning. *ACM Transactions on Multimedia Computing, Communications, and Applications* (2018).

[6] Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals, and George E Dahl. 2017. Neural message passing for quantum chemistry. In *ICML*.

[7] Takuo Hamaguchi, Hidekazu Oiwa, Masashi Shimbo, and Yuji Matsumoto. 2017. Knowledge transfer for out-of-knowledge-base entities: A graph neural network approach. *arXiv preprint arXiv:1706.05674* (2017).

[8] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *CVPR*.

[9] Thomas N. Kipf and Max Welling. 2017. Semi-Supervised Classification with Graph Convolutional Networks. In *ICLR*.

[10] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. Imagenet classification with deep convolutional neural networks. In *NeurIPS*.

[11] Wei Li, Rui Zhao, Tong Xiao, and Xiaogang Wang. 2014. Deepreid: Deep filter pairing neural network for person re-identification. In *CVPR*.

[12] Giuseppe Lisanti, Niki Martinel, Alberto Del Bimbo, and Gian Luca Foresti. 2017. Group Re-identification via Unsupervised Transfer of Sparse Features Encoding. In *ICCV*.

[13] Hao Luo, Wei Jiang, Xuan Zhang, Xing Fan, Jingjing Qian, and Chi Zhang. 2019. AlignedReID++: Dynamically matching local information for person re-identification. *Pattern Recognition* (2019).

[14] Ergys Ristani, Francesco Solera, Roger Zou, Rita Cucchiara, and Carlo Tomasi. 2016. Performance measures and a data set for multi-target, multi-camera tracking. In *ECCV workshop*.

[15] Weijian Ruan, Jun Chen, Yi Wu, Jinqiao Wang, Chao Liang, Ruimin Hu, and Junjun Jiang. 2018. Multi-correlation filters with triangle-structure constraints for object tracking. *IEEE Transactions on Multimedia* (2018).

[16] Chi Su, Jianing Li, Shiliang Zhang, Junliang Xing, Wen Gao, and Qi Tian. 2017. Pose-driven deep convolutional model for person re-identification. In *ICCV*.

[17] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. 2014. Sequence to sequence learning with neural networks. In *NeurIPS*.

[18] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. 2016. Rethinking the inception architecture for computer vision. In *CVPR*.

[19] Sylvain Takerkart, Guillaume Auzias, Bertrand Thirion, and Liva Ralaivola. 2014. Graph-based inter-subject pattern analysis of fMRI data. *PloS One* (2014).

[20] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. 2017. Graph attention networks. *arXiv preprint arXiv:1710.10903* (2017).

[21] Xiao Wang, Chao Liang, Chen Chen, Jun Chen, Zheng Wang, Zhen Han, and Chunxia Xiao. 2019. S3D: Scalable Pedestrian Detection via Score Scale Surface Discrimination. *IEEE Transactions on Circuits and Systems for Video Technology* (2019).

[22] Zheng Wang, Ruimin Hu, Chao Liang, Qingming Leng, and Kaimin Sun. 2014. Region-based interactive ranking optimization for person re-identification. In *PCM*.

[23] Zheng Wang, Ruimin Hu, Chao Liang, Yi Yu, Junjun Jiang, Mang Ye, Jun Chen, and Qingming Leng. 2016. Zero-shot person re-identification via cross-view consistency. *IEEE Transactions on Multimedia* (2016).

[24] Zheng Wang, Ruimin Hu, Yi Yu, Junjun Jiang, Chao Liang, and Jinqiao Wang. 2016. Scale-Adaptive Low-Resolution Person Re-Identification via Learning a Discriminating Surface.. In *IJCAI*.

[25] Zheng Wang, Ruimin Hu, Yi Yu, Junjun Jiang, Jiayi Ma, and Shin'ichi Satoh. 2017. Statistical inference of gaussian-laplace distribution for person verification. In *ACM Multimedia*.

[26] Zhixiang Wang, Zheng Wang, Yinqiang Zheng, Yung-Yu Chuang, and Shin'ichi Satoh. 2019. Learning to reduce dual-level discrepancy for infrared-visible person re-identification. In *CVPR*.

[27] Zheng Wang, Mang Ye, Fan Yang, Xiang Bai, and Shin'ichi Satoh. 2018. Cascaded SR-GAN for Scale-Adaptive Low Resolution Person Re-identification.. In *IJCAI*.

[28] Hao Xiao, Weiyao Lin, Bin Sheng, Ke Lu, Junchi Yan, Jingdong Wang, Errui Ding, Yihao Zhang, and Hongkai Xiong. 2018. Group Re-Identification: Leveraging and Integrating Multi-Grain Information. In *ACM Multimedia*.

[29] Tong Xiao, Hongsheng Li, Wanli Ouyang, and Xiaogang Wang. 2016. Learning deep feature representations with domain guided dropout for person re-identification. In *CVPR*.

[30] Dongshu Xu, Jun Chen, Chao Liang, Zheng Wang, and Ruimin Hu. 2019. Cross-view Identical Part Area Alignment for Person Re-identification. In *ICASSP*.

[31] Xun Yang, Meng Wang, Richang Hong, Qi Tian, and Yong Rui. 2017. Enhancing person re-identification in a self-trained subspace. *ACM Transactions on Multimedia Computing, Communications, and Applications* (2017).

[32] Zelong Zeng, Zhixiang Wang, Zheng Wang, Yung-Yu Chuang, and Shin'ichi Satoh. 2019. Illumination-Adaptive Person Re-identification. *arXiv preprint arXiv:1905.04525* (2019).

[33] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. 2015. Scalable person re-identification: A benchmark. In *CVPR*.

[34] Wei-Shi Zheng, Shaogang Gong, and Tao Xiang. 2009. Associating Groups of People. In *BMVC*.

[35] Zhedong Zheng, Liang Zheng, and Yi Yang. 2018. A discriminatively learned cnn embedding for person reidentification. *ACM Transactions on Multimedia Computing, Communications, and Applications* (2018).

[36] Zhun Zhong, Liang Zheng, Shaozi Li, and Yi Yang. 2018. Generalizing a Person Retrieval Model Hetero-and Homogeneously. In *ECCV*.

[37] Zhun Zhong, Liang Zheng, Zhedong Zheng, Shaozi Li, and Yi Yang. 2018. Camera style adaptation for person re-identification. In *CVPR*.

[38] Feng Zhu, Qi Chu, and Nenghai Yu. 2016. Consistent matching based on boosted salience channels for group re-identification. In *ICIP*.

[39] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *CVPR*.