Received 10 November 2023; revised 29 November 2023; accepted 24 December 2023. Date of publication 4 January 2024;
date of current version 24 January 2024.

Digital Object Identifier 10.1109/OJCOMS.2023.3349597

MetaGON: A Lightweight Pedestrian Re-Identification Domain Generalization Model Adapted to Edge Devices

XITING PENG [1] (Member, IEEE), HUAXUAN ZHAO [1] (Student Member, IEEE), XIAOYU ZHANG [2] (Member, IEEE), CHAOFENG ZHANG [3] (Member, IEEE), AND CAIJUAN CHEN [4] (Member, IEEE)

[1] School of Information Science and Engineering, Shenyang University of Technology, Shenyang 110870, China

[2] School of Artificial Intelligence, Shenyang University of Technology, Shenyang 110870, China

[3] School of Information and Electronic Engineering, Advanced Institute of Industrial Technology, Tokyo 140-0011, Japan

[4] The Information Systems Architecture Science Research Division, National Institute of Informatics, Tokyo 101-8430, Japan

CORRESPONDING AUTHOR: X. ZHANG (e-mail: xy.zhang@sut.edu.cn)

This work was supported in part by the "Chunhui Plan" Cooperative Research for the Ministry of Education under Grant HZKY20220407; in part by the Natural Science Foundation of Liaoning Province under Grant LJKZ0136; in part by JSPS KAKENHI under Grant JP22K17884 and Grant JP23KJ1005; in part by the National Key Research and Development Plan project under Grant 2022YFE011400; and in part by the Xingliao Talents Program of Liaoning Province under Grant XLYC2203046.

ABSTRACT Pedestrian re-identification (Re-ID) leverages cross-camera data acquired by the Internet of Things (IoT) devices and sensors to identify, monitor, and analyze pedestrians, allowing IoT applications to provide more intelligent, secure, and tailored services. Current pedestrian Re-ID research faces many challenges, such as low image resolution, perspective changes, posture changes, light changes, and occlusions, resulting in models trained on other datasets being unable to be directly applied and showing poor generalization capabilities. In addition, IoT edge devices are often limited by processing power and memory capacity and cannot withstand complex and large deep learning models. Therefore, designing a lightweight and generalizable pedestrian Re-ID model is better suited for implementation on edge devices. Considering these issues, this study presents MetaGON, a lightweight model with cross-domain generalization capabilities, which combines the lightweight omni-scale network (OSNet) with the meta-learning method and Cycle Generative Adversarial Network (CycleGAN) to perform domain generalization. The model's generalization is enhanced through the simulation of the two stages of domain generalization in the meta-learning pipeline, where the obtained losses from meta-training and meta-testing are utilized for model optimization. Moreover, CycleGAN is employed to enhance and introduce style variations to the source data. The proposed MetaGON model is tested on a railway station re-identification dataset, and the model is deployed to edge devices for evaluation, which verifies the effectiveness of the algorithm.

INDEX TERMS Domain generalization, meta-learning, pedestrian re-identification, Internet of Things.

I. INTRODUCTION

SECURITY and monitoring are crucial aspects of IoT applications, IoT systems may monitor activity in specific areas by deploying cameras and other sensors, such as monitoring stores, warehouses, offices, and factories [1]. Pedestrian Re-ID is one of the important technologies in IoT monitoring that can detect the same person from several surveillance views. This technology can help identify suspicious persons or abnormal behavior of pedestrians [2]. In addition to security applications, the combination of the IoT and pedestrian Re-ID technology can be used to provide personalized services. Through pedestrian Re-ID technology, stores, hotels, and other entertainment venues can understand the preferences and needs of pedestrians, and then provide customers with a better experience, such as recommending products and providing customized services [3]. Therefore, pedestrian Re-ID can provide more intelligent, secure, and personalized services for IoT applications, while also contributing to urban management and decision-making optimization. At present, the pedestrian Re-ID algorithm is relatively mature, but there are still some problems in its application in edge scenarios of the IoT [4], [5], [6], [7], [8], [9].

In IoT application scenarios, first of all, there is the problem that edge devices are usually limited by processing power and memory capacity and cannot withstand complex and large deep learning models [10], [11]. Secondly, application scenarios such as industrial automation and smart cities have very strict latency requirements for data processing and decision-making. Finally, some IoT applications may require data to be processed locally at the edge to protect user privacy or reduce the need for data transmission [12]. Faced with these problems, the lightweight Re-ID model can run locally in real-time with limited resources without transmitting to the cloud, meeting the above-mentioned requirements of low consumption, real-time performance, and data privacy. In summary, proposing a lightweight Re-ID model is crucial to the actual implementation of edge scenarios.

In addition to the lightweight requirements of the IoT, research on pedestrian Re-ID also suffers from the problem of insufficient generalization in new domains. This is because the research on pedestrian Re-ID systems encounters obstacles such as poor image resolution, viewpoint changes, posture changes, light changes, occlusion, and other real factors. These difficulties result in significant differences between different datasets. Although the latest supervised learning methods have surpassed humans in many commonly used benchmark tests, when training supervised models on a dataset of a certain scene and applying the trained models to the unfamiliar new scene, the performance of the models will be greatly affected, exhibiting poor domain generalization ability [13]. The problem of poor generalization ability will have a significant impact on the actual deployment of the industry, leading to the need for time-consuming data collection and annotation for model training before deployment, resulting in a significant increase in deployment costs. Hence, more and more research has recently focused on solving unsupervised cross-domain problems.

Current unsupervised cross-domain pedestrian Re-ID problems can be considered as two parts [14]. One is domain adaptation, which focuses on the fact that the target domain has data but no annotations. One of its mainstream methods is to transfer the target domain style to the source domain for supervised training. The other is to pretrain in the source domain first, and then use the clustering method to learn pseudo labels for supervised fine-tuning [15]. Compared with domain adaptation methods that require the collection of target domain data, domain generalization that focuses on the unavailability of target domain data is more suitable for practical deployment. Most existing domain generalization algorithms learn domain-invariant features from various source domains [16]. In the context of normalization, researchers focus on finding effective ways to combine BatchNormalization and InstanceNormalization to eliminate style differences while enhancing discriminative power to better learn domain-invariant features. However, these methods lead to information loss and lack the ability to weaken new styles in unknown domains. Meta-learning is another domain generalization method. The idea is to use multiple source domains to simulate the generalization process, and then learn domain invariance capabilities, thereby improving performance in unseen domains. However, based on existing research results, the method of using multiple different source domains for training is superior to the method of using a single source domain for training. How to increase the diversity of source domain data is an important component of enhancing the meta-learning models' capacity for generalization.

Regarding how to enrich the diversity of the training domain, from a data perspective, since labeling pedestrians across cameras is a very time-consuming task, currently available large-scale pedestrian Re-ID datasets are still small. However, the available pedestrian detection datasets are large. Therefore, how to utilize these pedestrian datasets without camera labels to enrich the training dataset for domain generalization is a feasible method to achieve style diversification.

In this study, we first utilize a lightweight omni-scale feature extraction network that adapts to the lightweight needs of the IoT edge by replacing traditional convolutions with depthwise separable convolutions. Secondly, the feature extraction network is trained using meta-learning methods to achieve domain generalization by learning domain-invariant capabilities from multiple source domains. Finally, this paper uses the generative adversarial network CycleGAN for style diversification processing and injects other pedestrian dataset styles into the training source domain through unsupervised training to further enhance the generalization effect of meta-learning. This study deploys the proposed model MetaGON on edge devices and tests it using complex real-world scene data.

In conclusion, the primary contribution of this study is as follows:

- In view of the lightweight requirements of IoT edge devices, as well as the difficulties in annotating pedestrian Re-ID data and insufficient generalization, this study adopts a lightweight omni-scale feature extraction network and uses meta-learning pipelines to simulate the domain generalization process to train the network to improve the model's generalization capabilities and achieve domain generalization.
- For the limited pedestrian Re-ID data, we also use CycleGAN to enhance the style of the source dataset. Through unsupervised training, the style of other domains can be transferred to the training domain, enriching the training domain of meta-learning and further enhancing the model generalization.

- We deployed the proposed model MetaGON on an edge device and tested it using a complex train station pedestrian Re-ID dataset to simulate real IoT edge application scenarios and confirm the usefulness of the proposed MetaGON model.

II. RELATED WORK

A. PEDESTRIAN RE-ID AND IOV

Recently, pedestrian Re-ID methods mainly use supervised deep learning methods, and a vast number of research have further improved performance by designing complex network structures, attention mechanisms, loss functions, and human posture modules to learn more reliable and comprehensive pedestrian features. The most advanced supervised pedestrian Re-ID has achieved excellent performance on multiple public datasets. On the other hand, from the perspective of application scenarios, pedestrian Re-ID is a very important part of the IoT monitoring application, and there have been many studies on the combination of the two. Reference [17] proposed a new deep framework of locally aligned deep networks, targeting scenarios with both infrared and visible light modalities and used for pedestrian re-identification in 6G-enabled IoT. The framework network uses a two-channel architecture to narrow the gap between modalities and overcome the effects of limb changes and part incoordination by learning local features. Reference [18] proposed a novel siamese network architecture adapted to edge-side facilities, which captures useful features between different pedestrian images by designing a residual layer containing "identity block" and "transformation" blocks, and ingeniously uses a GAP layer to reduce model complexity, thereby minimizing the manual retrieval time in edge computing. Reference [19] designed a Re-ID system, using microservices for re-identification calculations on network devices in IoT edge computing, balancing efficiency and privacy protection. This design can supply adequate model computing resources, allowing the system's flexible growth or decrease to meet various situations and demand loads. However, in practical application scenarios without training data, the performance of supervised models can significantly decrease due to cross-domain differences. Hence, studying domain generalization in IoT scenarios has more practical significance.

B. DOMAIN GENERALIZATION

DG refers to the algorithm's capacity to generalize to unknown datasets by training solely on the source datasets without utilizing any information from the target datasets. At present, there are many studies on the generalization of pedestrian Re-ID in many fields. Reference [20] proposed a Domain-Invariant Mapping Network. This method enables the model to perform domain generalization by building a meta-learning pipeline and learning the correspondence between pedestrian images and their identity classifiers. Reference [21] proposes a model based on voting and hybrid mechanisms. This technique employs a hybrid mechanism based on voting to dynamically leverage multiple properties of the source datasets so as to improve the algorithm's generalization capabilities. Reference [22] proposed a Domain Embedding Expansion (DEX) module, which enhances the generalization capacity and stability of the model to unknown fields by dynamically enhancing the deep features during the training process. Although there have been many studies on domain generalization, combining it with lightweight networks for application at the edge of the IoT has not yet been studied.

C. META LEARNING

The hope that the algorithm will be able to learn how to learn is referred to as meta-learning, aims to develop adaptive learning algorithms that can automatically adapt to different tasks and data characteristics by selecting appropriate model architectures and parameter configurations. There have been researches on applying meta-learning structures to pedestrian re-ID. Reference [23] proposed a meta-graph perception-based pedestrian Re-ID technique (Meta-GA). To minimize the distance between domains, this technique develops the closest neighbor association network for neighboring classes between fields and produces randomly selected features and similar features using hybrid learning. Reference [24] proposed the MetaCam method. The algorithm divides the data of different cameras in the training domain into a meta-train dataset and a meta-test dataset to simulate the switching of different camera scenes, which can effectively deal with the deviation caused by different cameras. Reference [25] proposed the Meta Batch-Instance Normalization (MetaBIN) method, which combines the normalization layer with meta-learning. It generalizes the normalization layer by simulating failed generalization scenarios via meta-learning. However, these studies did not pay attention to the relationship between source domain diversity and generalization. Instead, we used style transfer algorithms combined with meta-learning to enhance domain style and further enhance the generalization ability of meta-learning.

D. GENERATIVE ADVERSARIAL NETWORK

Most research on GAN and pedestrian Re-ID currently uses GAN models to perform domain alignment, and the source domain is then translated into the target domain's style for supervised training. Reference [26] proposes a generative adversarial network called FPGAN that can preserve pedestrian features to overcome the cross-domain challenge of pedestrian Re-ID, and proposes a multi-scale feature enhanced re-identification model. FPGAN learns style transfer in an unsupervised manner and preserves the pedestrian information of the source image through a transfer function, guaranteeing that the transmitted person image has the same style as the target dataset. Reference [27] proposes a conditional transmission network model inspired by human imagination. The model uses StarGAN to generate images of the input image from multiple other camera perspectives and mixes the generated image with the input image to finally obtain a multi-view mixed pedestrian representation

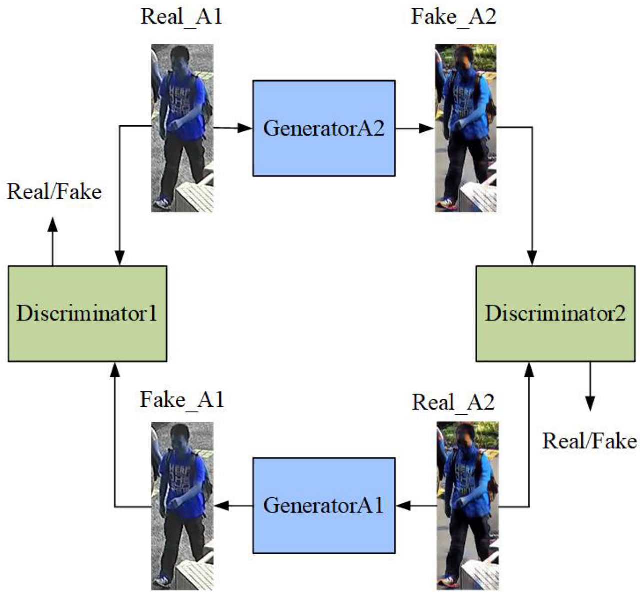

FIGURE 1. CycleGAN structure.

for matching. Reference [28] proposed a GAN model based on human posture and color gamut. Using certain postures and color gamuts as criteria, pedestrian image synthesis is performed by developing a pose transmitter and a color gamut converter to optimize GAN. Unlike previous works, however, we employ the GAN network to perform style enhancement on the source domain for domain generalization rather than the target domain for domain adaptation.

III. METHODOLOGY

The MetaGON model proposed in this study is divided into three parts, and its overall architecture is shown in Fig. 3. The first part is to process data from the source domain. We first use CycleGAN for unsupervised training to learn style transitions between different datasets, and then inject styles from other datasets into the source domain to obtain training data. The second part is to use a omni-scale feature network to extract features from pedestrian images. The third part is to use meta-learning training methods to train the omni-scale feature network and ultimately obtain a Re-ID model with generalization ability.

A. CYCLE GENERATIVE ADVERSARIAL NETWORK

Re-ID data, in particular, is difficult to label, whereas GAN training data does not require labeling. As a result, alternative unlabeled pedestrian data styles may be introduced into the source domain via CycleGAN [29] to accomplish style augmentation and hence increase model generalization.

CycleGAN is an image translation conversion method that does not require paired images. Its model is shown in Fig. 1. It is an innovative method in the field of unsupervised image translation developed from GAN, which is mainly used for generation tasks. The generation job entails feeding some random noise into the generation model, which then maps these noises to the distribution of the A-domain data. The probability distribution of pictures output by the generative model will eventually approach the probability distribution of A-domain images throughout training. As a result, the generative model will produce visuals that are comparable to A-domain photographs. GAN's objective is as follows:

$$\min_{G_A}\max_{D} L(G_A, D) = E_{a\sim P_{a(a)}}\left[\log\,(D(a)\right] + E_{n\sim P_{n(n)}}\left[\log\,(1 - D(G_A(n)))\right], \quad (1)$$

where G_A is the generator, $E_{a\sim P_{a(a)}}$ is the A sample space, $E_{n\sim P_{n(n)}}$ is the noise sample space and D is the discriminator.

CycleGAN is different from GAN in that it consists of two pairs of generators and discriminators. In addition, it also considers that an image $m2$ is generated from $m1$ through the mapping G_{A2} of $A1$ to $A2$, and the image $m1$ can be mapped back by the mapping G_{A1} of $A2$ to $A1$, so the loss of the cycle consistency is as follows:

$$L_c(G_{A2}, G_{A1}) = E_{a2\sim P_{data(a2)}}\left[\|G_{A2}(G_{A1}(a2)) - a2\|_1\right] + E_{a1\sim P_{data(a1)}}\left[\|G_{A1}(G_{A2}(a1)) - a1\|_1\right], \quad (2)$$

Therefore, the final loss function is defined as follows:

$$L(G_{A1}, G_{A2}, D_1, D_2) = L(G_{A2}, D_2, A1, A2) + L(G_{A1}, D_1, A2, A1) + \delta L_c(G_{A1}, G_{A2}), \quad (3)$$

where the overall loss consists of adversarial loss and cycle consistency loss, δ indicates the weight ratio between two losses.

B. OMNI-SCALE NETWORK

Pedestrian Re-ID poses challenges due to significant variations within the same class and the subtle differences between different classes. Overcoming these challenges requires effective learning of discriminative features. One approach that addresses these issues is OSNet [30], a model that learns omni-scale feature. By fusing features from multiple scales, OSNet improves the discrimination capability, leading to better distinction between different pedestrians. Furthermore, OSNet implements a lightweight architecture using depthwise separable convolutions. Depthwise separable convolution first performs group convolution with the number of groups equal to the number of channels, and then combines the results and uses a 1×1 standard convolution to obtain the final output feature map. This method reduces the number of parameters while preserving the representation learning ability of the convolution kernel.

The basic block consists of a residual block that is configured with a lightly separable convolution, and the scale index S is used to control how many layers of depth separable convolutions are superimposed on the network. For an input x, the residual $\tilde{x}$ as follows:

$$\tilde{x} = \sum_{s}^{S} F^s(x), \quad (4)$$

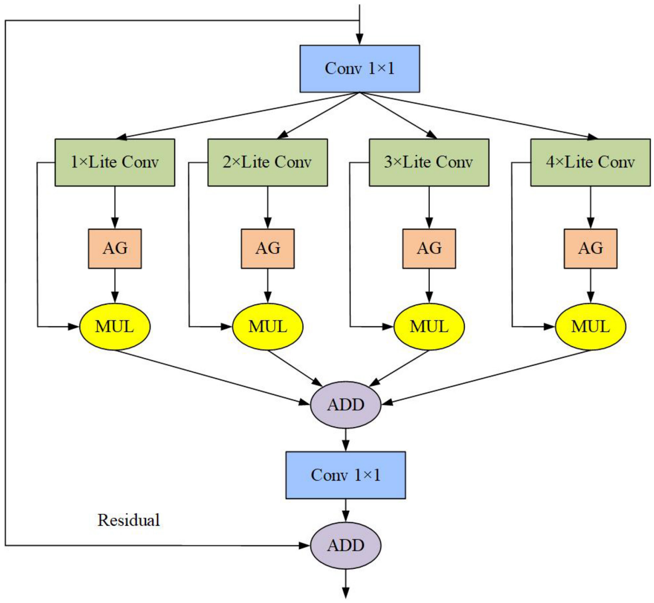

FIGURE 2. OSNet structure.

where F represents the depth separable convolutional layer, and F^s represents the convolution of s lite layers superimposed.

In order to better express omni-scale features, an aggregation gate (AG) is added after convolution for dynamic weighting. A multi-layer perceptron and a non-parametric global average pooling layer make up the learnable neural network known as AG. Different weights can be dynamically assigned by AG to various scale features. The residual after introducing AG is expressed as follows:

$$\tilde{x} = \sum_{s}^{S} A\left(F^s(x)\right) \odot F^s(x), \tag{5}$$

where A represents the AG and $\odot$ denotes the Hadamard product. The output of AG is a vector, which is used to perform channel-wise weighting on the corresponding scale features. The OSNet scale index used in this paper is set to 4, and its corresponding specific structure is depicted in Fig. 2.

C. META LEARNING

The idea of the meta-learning method adopted in this study is to imitate the domain generalization training and testing process [31]. The structure of meta-learning is shown in the blue box in Fig. 3. The losses generated by these two processes jointly optimize the model, which in turn enables the model to learn general recognition capabilities. Specifically, at each training epoch, one domain is randomly selected from N source domains as the meta-test dataset, while the remaining $N-1$ domains serve as the meta-train dataset.

1) TRAINING LOSS

Unlike the fully connected layer classifier, the memory-based meta-learning framework is equipped with a block of memory for each source domain and each memory contains I_N centers. This method can avoid the instability caused by too many parameters in the meta-learning optimization process. Specifically, at the beginning of training, feature extraction is performed on all images of each identity in each domain, and the mean value of each identity feature is used as the center of the corresponding identity in its corresponding domain. In subsequent training, the center is updated with the features obtained at each iteration. The classification is completed by calculating the similarity of the input image feature $F(x)$ to each center, for a source dataset with I_N identities, its memory-based classification loss is defined as follows:

$$L_M = -\log \frac{\exp\left(Mem[k]^T f(x)/\gamma\right)}{\sum_{i=1}^{I_N} \exp\left(Mem[i]^T f(x)/\gamma\right)}, \tag{6}$$

where $Mem[k]$ represents the true identity center corresponding to the input image, and γ represents the temperature factor controlling the distribution scale.

In addition, the training loss also includes the triplet loss commonly used for pedestrian Re-ID, which is defined as follows:

$$L_T = \max\left(d_p - d_n + margin, 0\right), \tag{7}$$

where d_n is the Euclidean distance between pairs of negative sample and d_p is the Euclidean distance between pairs of positive sample. A training threshold parameter called *margin* is determined by the real demands.

2) META-TRAIN

First, the initial model parameter Θ is copied for feature extraction. In addition, the mean and variance obtained by batch normalizing the features are recorded, which will act in the BN layer of the meta-test stage to enhance the meta-test features. The loss in the meta-training stage includes triple loss and memory loss, which are as follows:

$$L_{mtr} = \frac{1}{N-1} \sum_{k=0}^{N-1} \left(L_T\left(x^k; \Theta\right) + L_M\left(x^k, M^K; \Theta\right)\right), \tag{8}$$

where x^k represents the data belonging to the k-th meta-train dataset, and M^K is the memory of the k-th meta-train dataset. The meta-training loss is averaged over multiple domains.

3) META-TEST

Copy the initial model Θ, and then use L_{mtr} to update it to get model Θ', which is used for feature extraction during meta-testing, and then sampled from a Gaussian distribution generated by the mean-variance preserved during meta-training, injecting this meta-training domain information into the meta-test data, so that the features of the test image can

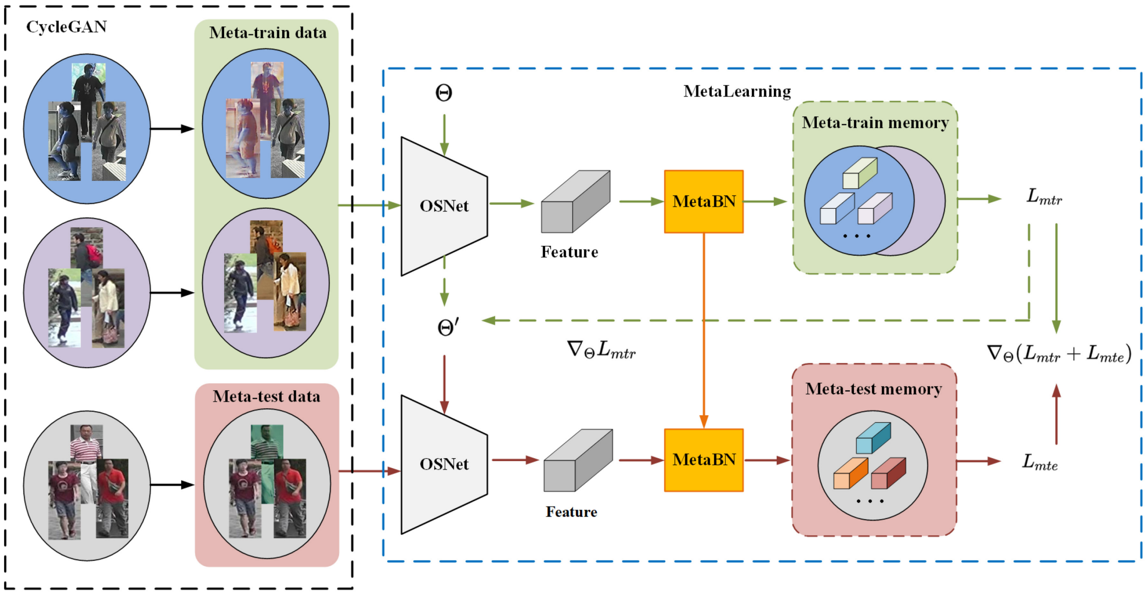

FIGURE 3. **Algorithm framework.**

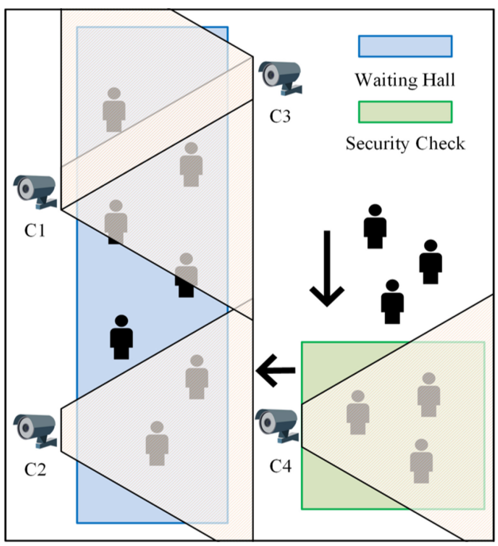

FIGURE 4. **Distribution of cameras from C1 to C4.**

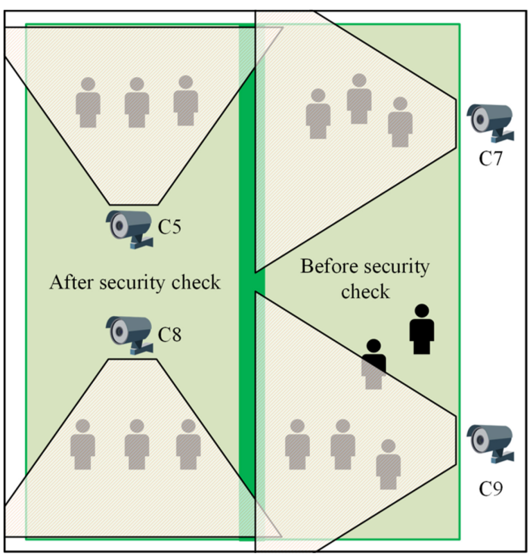

FIGURE 5. **Distribution of cameras from C5 to C9.**

be more diverse. The meta-training loss is obtained using the meta-test data as follows:

$$L_{mte} = L_T\left(x; \Theta'\right) + \frac{1}{N-1}\sum_{k=0}^{N-1} L_M\left(f^k, M; \Theta'\right), \tag{9}$$

where f^k represents the mixed features sampled from the BN information preserved from the kth source domain, and the memory-based loss is the average of the memory loss after injecting all the meta-training domain information.

4) MODEL OPTIMIZATION

The training loss consists of the first-stage loss based on the initial model Θ and the meta-testing loss generated by the updated model Θ' using the first-stage loss, and its formula is as follows:

$$arg \min_{\Theta}\left(L_{mtr}(\Theta) + L_{mte}\left(\Theta'\right)\right). \tag{10}$$

By simulating the training and testing process of domain generalization, the loss generated by these two processes is used to optimize the model towards a higher generalization ability.

IV. EXPERIMENT

A. DATASET

The datasets used in the experiments include PRID, 3DPes, MIT-CBCL, MSMT17V1, DukeMTMC, Market1501

TABLE 1. Comparison With Other Domain Generalization Methods

Method	Source	IDs	Images	Railway			
				mAP	R1	R5	R10
OSNet-IBN	All-MSMT17	4101	126441	23.7%	29.9%	48.2%	57.8%
QAConv50(Resnet-50)	All-MSMT17	4101	126441	28.9%	37.1%	56.3%	65.4%
OSNet-AIN	All-MSMT17	4101	126441	32.1%	39.4%	59.1%	67.2%
M3L(Resnet-50)	Du+Mar+C03	2820	55748	34.1%	42.4%	57.9%	66.1%
M3L(Resnet-IBN)	Du+Mar+C03	2820	55748	34.5%	42.5%	58.4%	68.8%
MetaGON	Du+Mar+C03	2820	55748	**39.5%**	**48.5%**	**69.1%**	**77.9%**

FIGURE 6. Example of railway station pedestrian Re-ID dataset.

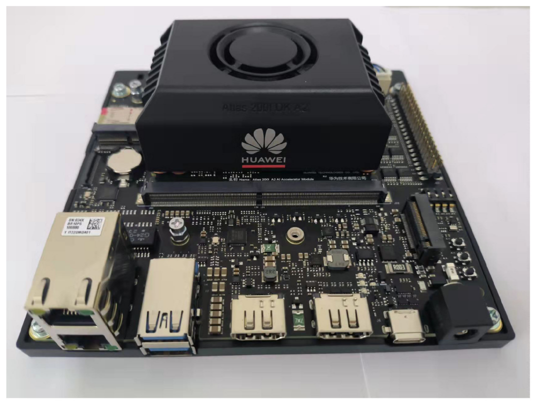

FIGURE 7. Huawei Atlas 200I DK A2 device.

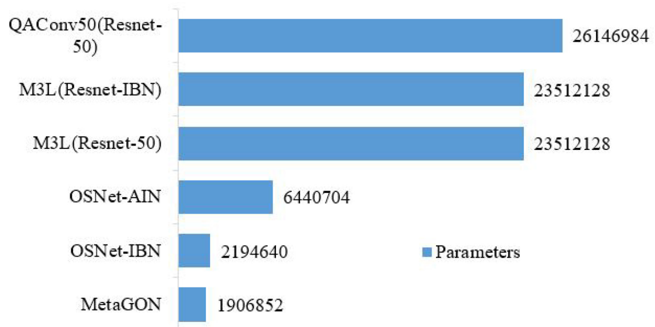

FIGURE 8. Comparison of model parameters.

and CUHK03. Among them, CUHK03, DukeMTMC, MSMT17V1 and Market1501 are utilized as the source domain, while the PRID, 3DPes, and MIT datasets are injected into the source domain as new styles.

This study uses real application scenario data from the railway station as the test set, which comes from surveillance cameras at Shenyang Railway Station and Shenyang Bei Railway Station. The data distribution of Shenyang Bei Railway Station is shown in Fig. 4. There are 4 cameras in total, including the security checkpoint and the waiting hall. In this scene, the same pedestrian appears in at least two cameras. The data distribution of Shenyang Railway Station is shown in Fig. 5. There are 4 cameras in total. The scene is before and after the security check. In this scene, the same pedestrian appears in at least two cameras. Fig. 6 displays a specific example of the railway station dataset. The same pedestrian is blocked and blurred between C1 and C3, low resolution occurs between C2 and C4, and there is a large color difference problem between C5 and C7. There is a change in viewing angle between C8 and C9, so this real application scenario is a more difficult Re-ID scenario and has certain challenges. In this experiment, we use the collected 5431 images of 341 pedestrians for testing.

In this experiment, the evaluation metric used is Rank-N, which shows the proportion of accurately retrieved images within the top N photos with the highest score in the result list. Rank-1/5/10 is used to describe the CMC curve. Another evaluation metric is mAP, which indicates the degree to which correctly hit photos are at the front of the search list, which can more accurately represent the model's performance. We also used parameters as an evaluation metric, with fewer parameters indicating a lighter network.

FIGURE 9. Visualize results.

In terms of implementation details of this study, NVIDIA GeForce RTX 3080 Ti is used for training. During training, the image is reset to a size of 128×64. The training runs for a total of 60 epochs, with a batch size of 16 for each epoch. The triplet loss has a margin of 0.3. Then set the learning rate to 3.5×10^{-5}, which gradually increases to 3.5×10^{-4} in the first ten epochs, and then decays with a weight of 0.1. We use the Atlas 200I DK A2 developer kit to simulate edge devices for testing. The device can provide 8TOPS INT8 computing power and can be connected to camera accessories to meet the requirements of pedestrian Re-ID scenarios, the device is shown in Fig. 7.

B. COMPARISON WITH OTHER DOMAIN GENERALIZATION METHODS

The proposed model was compared with OSNet-IBN [30], OSNet-AIN [32], QAConv50 [33] and M3L [31], and Table 1 displays the outcomes of the experiment. For single-source domain generalization methods, the validation, train, and test datasets of the MSMT17 are combined for training, which is for fair comparison with multi-source domain methods. Compared with OSNet-IBN trained in the single-source domain, our method improves the mAP by 15.8%, and compared with other single-source domain generalization models, our method is still about 10% higher. This shows that multi-source domain models trained with meta-learning strategies and memory-based classification methods have better generalization ability. Compared with the multi-source domain generalization model M3L, our model is about 4% higher on the mAP, and the model using OSNet is lighter, which validates the efficacy of the MetaGON model. As shown in Fig. 8, in terms of model lightweight performance, our method has the smallest amount of network parameters, and fewer parameters means fewer computing resources are required during the inference process. This makes the network easier to deploy and operate in environments with limited hardware resources.

TABLE 2. Effectiveness of Meta-Learning and CycleGAN

Source	Dataset	Meta	CycleGAN	mAP	R1
		✗	✗	36.6%	45.0%
Du+Mar+C03	Railway	✓	✗	38.4%	47.5%
		✓	✓	39.5%	48.5%

C. ABLATION STUDIES

1) EFFECTIVENESS OF META-LEARNING

In this experiment, the model without a meta-learning structure is set to delete the meta-testing step from the meta-learning structure, and only retain the meta-training stage. For each training epoch, the model is optimized using the average of the triplet loss and memory loss produced by all source domains. As shown in Table 2, compared to the model without meta-learning, the model with meta-learning strategy increased mAP in the target domain by 1.8%, which verifies that using meta-learning pipeline for optimization can make the model learn better domain invariant knowledge leads to better generalization capability.

2) EFFECTIVENESS OF CYCLEGAN

In this experiment, PRID, 3DPes, and MIT are used to simulate unlabeled pedestrian data, and the label information of datasets is not used when training CycleGAN. Use the trained model to enhance part of the photos in the source datasets to the style of the above three datasets. According to Table 2, the mAP of the meta-learning model trained with style-enhanced data is 1.1% higher than the model without enhancement and has greater generalization capability.

D. RESULTS VISUALIZATION

Fig. 9 shows the Rank-10 matching results for the train station dataset. Among them, the green border indicates that the match is the same pedestrian, and the red border indicates that the match is incorrect. The majority of the cross-camera pedestrians are matched, as seen by the findings of cases (a), (b), and (c), demonstrating the efficacy of the suggested methodology.

V. CONCLUSION

In this study, we present a domain generalization model called MetaGON, which is implemented based on lightweight OSNet, meta-learning, and CycleGAN to tackle the domain generalization problem in pedestrian Re-ID and the edge of the IoT lightweight requirements for the device. By simulating the domain generalization process in a meta-learning pipeline, the losses from the two processes of meta-learning are used to jointly optimize the model, resulting in better performance on unseen datasets. In addition, CycleGAN is introduced to transfer styles from other domains to meta-learning source data sets through unsupervised training to further improve generalization capabilities. Finally, our proposed model is tested on a real IoT railway station scenario dataset and deployed on edge devices. According to the analysis of experimental results, the OSNet network trained in the source domain using a meta-learning pipeline and style enhancement can achieve lightweight while ensuring accuracy, verifying the effectiveness of the model. In addition, according to the visualization results (d) and (e) in Fig. 9, in real scenes, pedestrians will have very similar characteristics (such as black coats). This common clothing style makes it difficult to distinguish between different pedestrians. In order to solve these problems, our future research will focus on refining the domain generalization model to increase recognition accuracy in challenging scenarios

ACKNOWLEDGMENT

The authors are thankful to the anonymous reviewers and editors for their valuable comments and suggestions.

REFERENCES

[1] D. Trihinas, G. Pallis, and M. D. Dikaiakos, "Low-cost adaptive monitoring techniques for the Internet of Things," *IEEE Trans. Services Comput.*, vol. 14, no. 2, pp. 487–501, Mar./Apr. 2021.
[2] S. Zhang and H. Yu, "Person re-identification by multi-camera networks for Internet of Things in smart cities," *IEEE Access*, vol. 6, pp. 76111–76117, 2018.
[3] M. Fu et al., "Improving person reidentification using a self-focusing network in Internet of Things," *IEEE Internet Things J.*, vol. 9, no. 12, pp. 9342–9353, Jun. 2022.
[4] Y. Gong, K. Bian, F. Hao, Y. Sun, and Y. Wu, "Dependent tasks offloading in mobile edge computing: A multi-objective evolutionary optimization strategy," *Future Gener. Comput. Syst.*, vol. 148, pp. 314–325, Nov. 2023.
[5] Y. Zhang et al., "FedNILM: Applying federated learning to NILM applications at the edge," *IEEE Trans. Green Commun. Netw.*, vol. 7, no. 2, pp. 857–868, Jun. 2023.
[6] X. Peng, K. Ota, and M. Dong, "Multiattribute-based double auction toward resource allocation in vehicular fog computing," *IEEE Internet Things J.*, vol. 7, no. 4, pp. 3094–3103, Apr. 2020.
[7] X. Shao, M. Jibiki, Y. Teranishi, and N. Nishinaga, "An efficient load-balancing mechanism for heterogeneous range-queriable cloud storage," *Future Gener. Comput. Syst.*, vol. 78, pp. 920–930, Jan. 2018.
[8] X. Wang, X. Ren, C. Qiu, Z. Xiong, H. Yao, and V. C. Leung, "Integrating edge intelligence and blockchain: What, why, and how," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 4, pp. 2193–2229, 4th Quart., 2022.
[9] X. Ren et al., "AI-Bazaar: A cloud-edge computing power trading framework for ubiquitous ai services," *IEEE Trans. Cloud Comput.*, vol. 11, no. 3, pp. 2337–2348, Jul.–Sep. 2023.
[10] J. Cheng et al., "MATEC: A lightweight neural network for online encrypted traffic classification," *Comput. Netw.*, vol. 199, Nov. 2021, Art. no. 108472.
[11] X. Peng, Y. Wang, X. Zhang, H. Yang, X. Tang, and S. Bai, "A 6G-enabled lightweight framework for person re-identification on distributed edges," *Electronics*, vol. 12, no. 10, p. 2266, 2023.
[12] Y. Wu, H.-N. Dai, and H. Wang, "Convergence of blockchain and edge computing for secure and scalable IIoT critical infrastructures in industry 4.0," *IEEE Internet Things J.*, vol. 8, no. 4, pp. 2300–2317, Feb. 2021.
[13] H. Fan, L. Zheng, C. Yan, and Y. Yang, "Unsupervised person re-identification: Clustering and fine-tuning," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 14, no. 4, pp. 1–18, 2018.
[14] D. Kumar, P. Siva, P. Marchwica, and A. Wong, "Fairest of them all: Establishing a strong baseline for cross-domain person ReID," 2019, *arXiv:1907.12016*.
[15] Y. Ge, D. Chen, and H. Li, "Mutual mean-teaching: Pseudo label refinery for unsupervised domain adaptation on person re-identification," 2020, *arXiv:2001.01526*.
[16] J. Jia, Q. Ruan, and T. M. Hospedales, "Frustratingly easy person re-identification: Generalizing person Re-ID in practice," 2019, *arXiv:1905.03422*.
[17] S. Liu and J. Zhang, "Local alignment deep network for infrared-visible cross-modal person reidentification in 6G-enabled Internet of Things," *IEEE Internet Things J.*, vol. 8, no. 20, pp. 15170–15179, Oct. 2021.
[18] S. Pang, S. Qiao, T. Song, J. Zhao, and P. Zheng, "An improved convolutional network architecture based on residual modeling for person re-identification in edge computing," *IEEE Access*, vol. 7, pp. 106748–106759, 2019.
[19] C.-H. Chen and C.-T. Liu, "Person re-identification microservice over artificial intelligence Internet of Things edge computing gateway," *Electronics*, vol. 10, no. 18, p. 2264, 2021.
[20] J. Song, Y. Yang, Y.-Z. Song, T. Xiang, and T. M. Hospedales, "Generalizable person re-identification by domain-invariant mapping network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 719–728.
[21] Y. Dai, X. Li, J. Liu, Z. Tong, and L.-Y. Duan, "Generalizable person re-identification with relevance-aware mixture of experts," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 16145–16154.
[22] E. P. Ang, L. Shan, and A. C. Kot, "DEX: Domain embedding expansion for generalized person re-identification," 2021, *arXiv:2110.11391*.
[23] D. Wu et al., "A domain generalization pedestrian re-identification algorithm based on meta-graph aware," *Multimedia Tools Appl.*, vol. 83, pp. 2913–2933, 2024.
[24] F. Yang et al., "Joint noise-tolerant learning and meta camera shift adaptation for unsupervised person re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 4855–4864.
[25] S. Choi, T. Kim, M. Jeong, H. Park, and C. Kim, "Meta batch-instance normalization for generalizable person re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 3425–3435.

[26] X. Liu, H. Tan, X. Tong, J. Cao, and J. Zhou, "Feature preserving GAN and multi-scale feature enhancement for domain adaption person re-identification," *Neurocomputing*, vol. 364, pp. 108–118, Oct. 2019.

[27] R. Sun, W. Lu, Y. Zhao, J. Zhang, and C. Kai, "A novel method for person re-identification: Conditional translated network based on GANs," *IEEE Access*, vol. 8, pp. 3677–3686, 2019.

[28] X. Liu, X. Liu, G. Li, and S. Bi, "Pose and color-gamut guided generative adversarial network for pedestrian image synthesis," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 12, pp. 10724–10736, Dec. 2023.

[29] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2223–2232.

[30] K. Zhou, Y. Yang, A. Cavallaro, and T. Xiang, "Omni-scale feature learning for person re-identification," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 3702–3712.

[31] Y. Zhao et al., "Learning to generalize unseen domains via memory-based multi-source meta-learning for person re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 6277–6286.

[32] K. Zhou, Y. Yang, A. Cavallaro, and T. Xiang, "Learning generalisable omni-scale representations for person re-identification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 9, pp. 5056–5069, Sep. 2022.

[33] S. Liao and L. Shao, "Interpretable and generalizable person re-identification with query-adaptive convolution and temporal lifting," in *Proc. 16th Eur. Conf. Comput. Vis.*, Glasgow, U.K., Aug. 2020, pp. 456–474.

XITING PENG (Member, IEEE) received the B.S. degree in mathematics and software engineering and the M.S. degree in mathematics from the University of Dalian Jiaotong, Dalian, China, in 2013 and 2017, respectively, and the Ph.D. degree in electrical engineering from the Muroran Institute of Technology, Japan, in 2020. She is currently an Assistant Professor with the School of Information Science and Engineering, Shenyang University of Technology, China. She was supported by the China Scholarship Council for the period of her Ph.D. Also, she was the Chair of the IEEE Muroran IT Student Branch from February to October 2020. Her current research interests include computational intelligence, machine learning, and edge computing. She has received the Best Paper Award from AICON 2019 and the Wireless Communication Letters Exemplary Reviewer Award 2019. She was awarded the Outstanding Contribution Award by the 8th International Conference on Signal and Information Processing, Networking and Computers as the General Chair in 2021. And, she was selected as a Workshop Co-Chair of the 21st IEEE International Conference on Trust, Security and Privacy in Computing and Communications.

HUAXUAN ZHAO (Student Member, IEEE) received the B.S. degree from Hebei GEO University, Shijiazhuang, China, in 2021. She is currently pursuing the Graduate degree with the Shenyang University of Technology, focusing on the field of deep learning and pedestrian re-identification.

XIAOYU ZHANG (Member, IEEE) received the bachelor's degree in information and computing science and software engineering from Dalian Jiaotong University, Dalian, China, in 2013, the M.S. degree in mathematics from the University of Dalian Jiaotong, Dalian, in 2017, and the Ph.D. degree in control theory and control engineering from the School of Control Science and Engineering, Dalian University of Technology, Dalian, in 2023. He joined the School of Artificial Intelligence, Shenyang University of Technology in May 2023. He was supported by China Scholarship Council to visit the Muroran Institute of Technology, Muroran, Hokkaido, Japan, for the period from October 2018 to October 2020. His current research interests include stability of neural networks, time-delay systems, switched systems, machine learning algorithms, and edge computing.

CHAOFENG ZHANG (Member, IEEE) received the B.Eng. degree from Soochow University, China, in 2011, and the M.Eng. and Ph.D. degrees from the Muroran Institute of Technology, Japan, in 2016 and 2019, respectively. He is currently an Assistant Professor with the Advanced Institute of Industrial Technology, Tokyo, Japan. From March 2017 to April 2017, he was a Visiting Scholar with Soochow University, China. His research interests include cloud computing, green IoT, wireless communication, and wireless positioning technology. He got IEEE VTS Tokyo Chapter 2016 Paper Award in 2016, and got the Best Presentation Award in A3 Annual Workshop on Next Generation Internet and Network Security. He serves as an Associate Editor for *IEEJ Transactions on Electronics, Information and Systems* and *Journal of Frontiers in Space Technologies*.

CAIJUAN CHEN (Member, IEEE) received the M.S. degree from the Huazhong University of Science and Technology, China, in 2020, and the Ph.D. degree from the Graduate University for Advanced Studies (SOKENDAI), Japan, in 2023. Her research interests include federated learning over wireless networks and mobile-edge computing.