Computing, Performance and Communication Systems (2025)
Clausius Scientific Press, Canada
DOI: 10.23977/cpcs.2025.090104
ISSN 2371-8889 Vol. 9 Num. 1

Anti-Cheating Model Based on TransReID Campus Running

Yanxu Wu[1,a], Zicheng Wang[1,b], Chuwei Wang[2,c], Dan Wang[3,*]

[1]*School of Electronic and Information Engineering, University of Science and Technology Liaoning, Anshan, China*
[2]*School of Materials and Metallurgy, University of Science and Technology Liaoning, Anshan, China*
[3]*Department of Sports, University of Science and Technology Liaoning, Anshan, China*
[a]*120223103138@stu.ustl.edu.cn*, [b]*2375309060@qq.com*, [c]*china205396@163.com*
**Corresponding author*

Keywords: TransReID, Pedestrian re-identification, Transformer, Target detection

Abstract: Pedestrian recognition is the use of computer vision algorithms to match pedestrian images or videos across devices, which has a huge application prospect in smart security, smart business, and so on. In this paper, TransReID technology is innovatively applied to campus running, and this paper is based on Vision Transformer's pedestrian recognition framework. The Transformer network is used as the backbone network; in addition, in order to improve the recognition ability and more diversified coverage proposed, this paper also innovatively adds the feature rotation invariance constraint strategy, which improves the recognition ability of human-type images with different angles. Finally, the mAP accuracy reaches 69.4%, which exceeds the original model by 8.4%, as shown on the MSMI17 dataset.

1. Introduction

In campus life scenarios, with the advancement of smart campus construction, accurate monitoring and management of students' sports status is becoming more and more important. As an important initiative to improve students' physical fitness, campus running has been widely adopted by many colleges and universities. However, in the process of campus running activities, how to effectively identify and track students participating in running and ensure the accuracy of data and fairness of the activities have become key issues to be solved.

Pedestrian recognition technology has important application value in this context. The core task of pedestrian re-recognition is to realize the association of specific objects under different scenes and different views of cameras, and the key point is to extract robust and discriminative features[1]. For a long time, methods based on convolutional neural networks (CNNs) have dominated the field of pedestrian re-recognition[2]. However, there are obvious limitations of such methods, which mainly focus on smaller discriminative regions, and the downsampling operations employed, including pooling and convolutional step size, lead to a reduction in the spatial resolution of the output feature maps, which greatly diminishes the ability to discriminate between similar-looking

objects[3]. Although the majority of attention mechanism-based techniques are incorporated into the model's deeper layers, they have trouble extracting several distinct discriminable areas and favor bigger continuous regions. Transformer-based models have shed light on the CNN-based pedestrian re-identification problem as a result of technological advancements, the rise of multi-head attention modules, and the abandonment of convolution and downsampling processes. The multi-head attention module is able to capture remote dependencies, prompting the model to focus on different parts of the human body, while the Transformer eliminates the need for downsampling operations and is able to retain more detailed information[4]. Transformer still requires special design and optimization for pedestrian re-recognition, nevertheless, when confronted with complicated picture scenarios including occlusion, position variety, and camera perspective alterations.

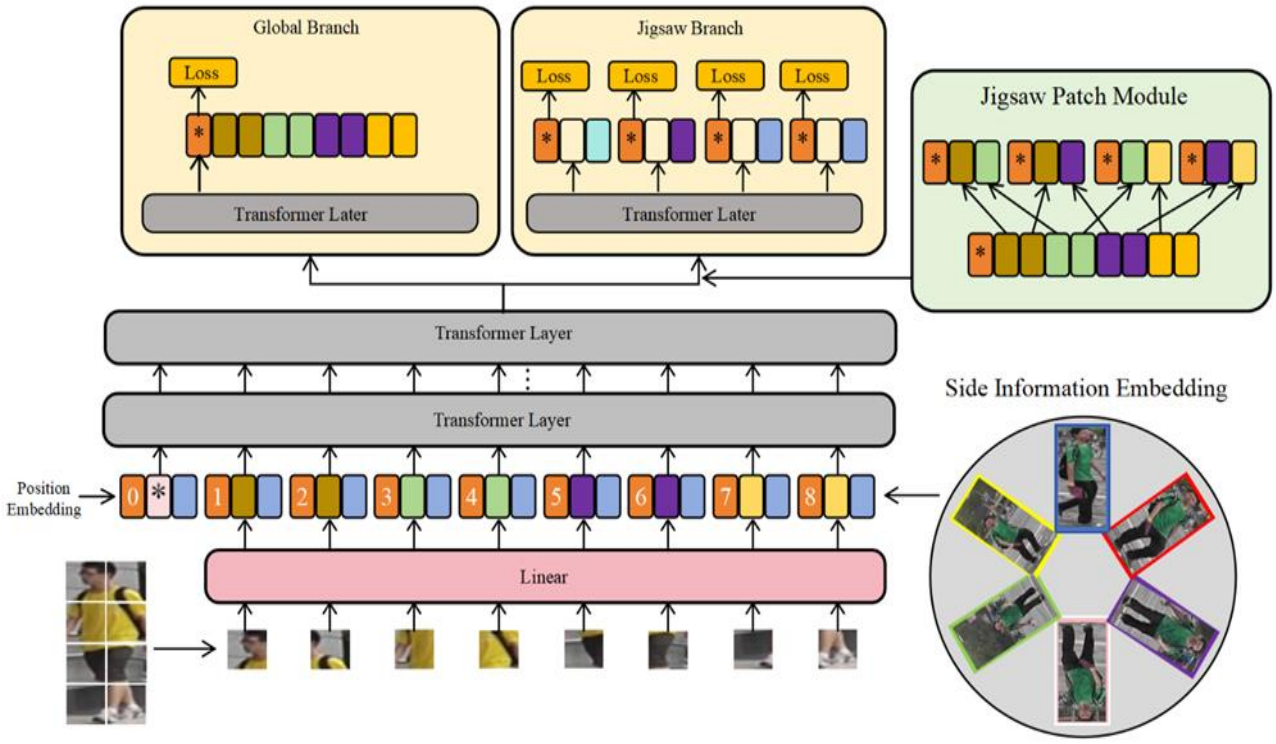

Figure 1: Framework of proposed TransReID

In view of this, we innovatively apply TransReID, a framework for pedestrian re-identification, to the campus running program, and considering that the Transformer model is more sensitive to rotating object recognition, this paper innovatively adds feature-level rotations into the network to improve the model's ability to recognize rotating targets[5], aiming to solve the problem of student recognition and tracking in campus running scenarios and provide a new and effective way for the intelligent management of campus sports activities.

The main innovations of this paper are:

1) Innovative application of TransReID technology to campus running anti-cheating, which improves the detection accuracy of person re-identification.

2) Incorporating feature-level rotation into the Transformer backbone network to improve the model's ability to recognize rotating targets.

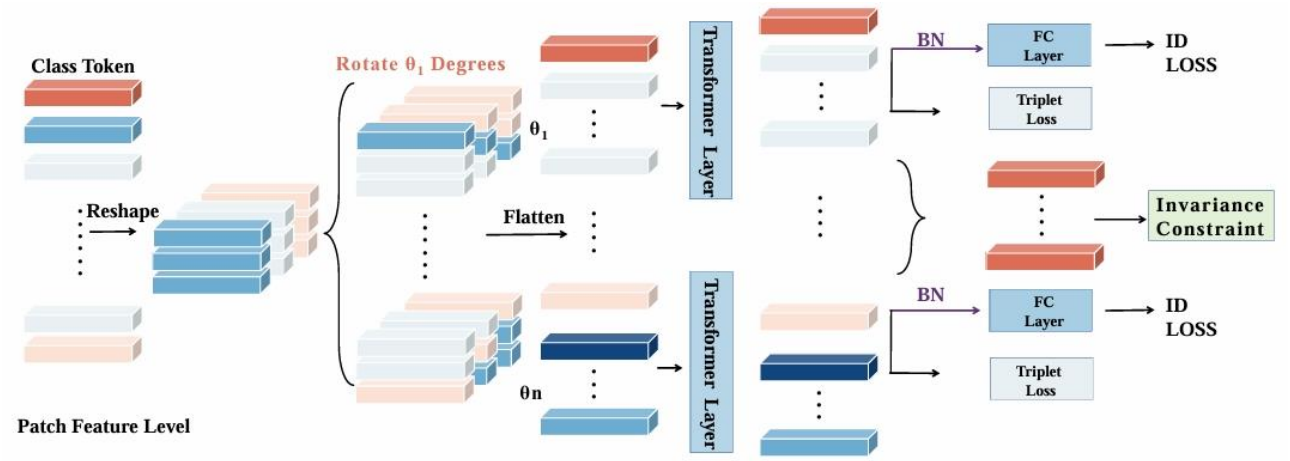

Figure 2 Details of the feature-level rotation strategy

2. Behavioral re-identification method based on Transformer

2.1. Vision Transformer Backbone

Our architecture adopts the Transformer baseline model proposed by He et al.[6] for the pedestrian re-identification (ReID) task, which improves the visual Transformer's backbone network. The network structure diagram is presented in Figure 1. and the input image $(x \in \mathbb{R}^{H\times W\times C})$ undergoes a hierarchical processing flow: first, the image is divided into non-overlapping chunks of size $(P\times P)$, resulting in a number of $(N=(H\times W)/P^2)$ chunks. Each chunk is mapped to the high-dimensional feature space $\mathbb{R}^D(D=P^2\times C)$ by linear projection, forming a chunk embedding sequence $(\mathbf{Z}_0 \in \mathbb{R}^{N\times D})$. To enhance the local context modeling capability, an overlapping chunk embedding strategy is used: a refined sequenceis $(\mathbf{Z}_{0'} \in \mathbb{R}^{N'\times D})$ generated by a convolutional layer $(S < P)$ where.

$$N' = \left(\left\lfloor \frac{H-P}{S} \right\rfloor + 1\right) \times \left(\left\lfloor \frac{W-P}{S} \right\rfloor + 1\right) \tag{1}$$

Learned classification tokens $(\mathbf{c}_0 \in \mathbb{R}^D)$ are added to the end of the sequence to aggregate global semantic information. Subsequently, location embeddings $(\mathbf{E}_{pos} \in \mathbb{R}^{(N'+1)\times D})$ are incorporated to encode spatial location information. The enhanced sequence $(\mathbf{Z}_{in}=[\mathbf{c}_0;\mathbf{Z}_{0'}]+\mathbf{E}_{pos})$ processed through multiple Transformer encoder layers, each containing multiple self-attention and feedforward networks. The final output $(\mathbf{Z}_{out} \in \mathbb{R}^{(N'+1)\times D})$ retains category labeling $(\mathbf{c}_{0'})$ as a global feature representation. The network is optimized by a combination of ternary loss and cross-entropy loss $(\mathcal{L}_{CE})$ to ensure that discriminative features are learned for the ReID task.

2.2. Characteristic level of rotation

Although global features already contain rich information and possess strong object recognition capabilities, the Transformer model based on the attention mechanism does not have rotational invariance and is more sensitive in scenarios of recognizing rotating objects. In the campus running scenario, the traditional image - level rotational enhancement method in visual Transformers has the dual drawbacks of feature information loss and background noise interference. This paper proposes a feature - level rotational enhancement strategy. By constructing a differentiable feature space

transformation module, it generates multi - view rotational feature mappings during the deep feature extraction stage. Combined with contrastive learning constraints, it ensures the consistency of features under different rotational views. Moreover, through the dynamic weighting of dynamic weights, it guarantees the consistency of rotational features. It ensures consistency under different rotational views and achieves feature fusion optimization through a dynamic weight adjustment mechanism. The network structure diagram is shown in Figure 2.

This module converts the global feature representation ($f \in \mathbb{R}^{(N+1)\times D}$) (containing N chunk embeddings ($f_p \in \mathbb{R}^{N\times D}$) and a category labeling (c_0)) output by the visual Transformer backbone network into a 2D grid structure ($f_{\text{res}} \in \mathbb{R}^{X\times Y\times D}$). This transformation is defined by the following mathematical formula:

$$(X = \left\lfloor \frac{H-P}{S} \right\rfloor + 1, \quad Y = \left\lfloor \frac{W-P}{S} \right\rfloor + 1) \tag{2}$$

where is the input image size ($H \times W$), denotes the chunk size ($P \times P$), and S represents the overlap step when chunks are embedded. By reshaping a one-dimensional chunking sequence into a two-dimensional spatial layout, this module lays the foundation for subsequent geometric transformations at the abstract feature level for image-like rotations. The rotational feature generation module implements rotational invariant learning by performing angular transformations on the reshaped feature mesh. Specifically, a set of angle sets, whose value range is limited to ($[-\alpha, \alpha]$), is first randomly sampled to simulate the diversity of camera viewpoints ($A = \{\theta_i \mid i = 1, 2, \cdots, n\}$). For each angle, the feature mesh coordinates are transformed using a rotation matrix $((x, y))$:

$$\begin{aligned} x_r^k &= x\cos\theta_i - y\sin\theta_i, \\ y_r^k &= x\sin\theta_i + y\cos\theta_i. \end{aligned} \tag{3}$$

This operation generates a set of rotation feature matrices ($F_r = \{f_{r1}, f_{r2}, \cdots, f_{rn}\}$), each $(f\_\{ri\})$ corresponding to a different rotation orientation. Unlike image-level rotation, this feature-level enhancement introduces angular diversity while preserving the integrity of the discriminative information. The transformed features are then flattened back to a one-dimensional sequence and appended with replicated category markers to fit the Transformer architecture. This process ensures effective learning of rotational invariance through the fusion of multi-view feature representations during training.

3. Experiments

3.1. Datasets and evaluation criteria

In this paper, the proposed TransReID-based method is experimented with using the MSMT17[7] dataset. The MSMT17 dataset consists of 4101 pedestrians captured by 15 cameras. There are 75864 images in the training set, 37932 images in the test set, and 12644 images in the validation set.

3.2. Experimental preparation

We built the proposed network using the PyTorch framework and our training environment is shown in Table 1:

Table 1 Experimental environment

names	Experimental parameeters
operating system	Ubuntu-22.04
GPU	NVIDIA 3080-10G
CPU	Intel(R) i7-12700k CPU @ 3.50GHz
Python version	3.8.19
Deep Learning Framework Versions	Pytorch 1.13.1
Accelerated Computing Module Versions	CUDA 11.6
RAM	32GB

During the experiments, all images will be resized to 256×128. The model parameters will be trained using the SGD optimizer, and the training images will be augmented by random horizontal flipping, padding, random cropping, and random erasing[8]. The batch size is set to 8, and each ID corresponds to 4 images. A stochastic gradient descent (SGD) optimizer with momentum of 0.9 is used. The weight decay is 1e-4. The learning rate is initialized to 0.008, and cosine learning rate decay is used.

3.3. TransReID Learning Experiment Results

In this study, we perform experimental validation using the MSMI17 dataset to evaluate the learning efficacy of our modified model. The experimental results are presented in Table 2. Figure 3 illustrates the convergence of the loss function during model training by displaying the dynamic change of the loss function value as the number of iterations increases. It is evident from the graphic that the model starts to converge gradually at around 100 iterations.

Table 2 Results of the experiment

Backbone	Method	MSMI17	
		mAP	Rank-1
VIT-B/16	TransReID	61.0	81.8
	TransReID+ characteristic level of rotation	69.4	86.2

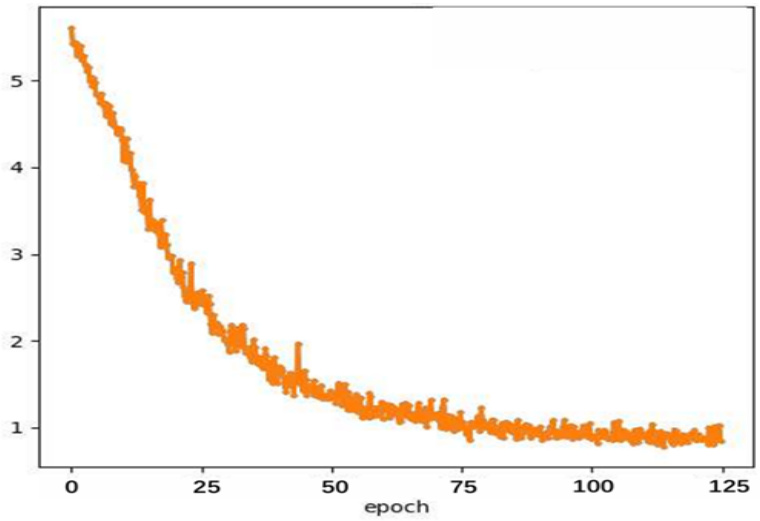

Figure 3 Loss Figure

4. Conclusion

In this study, we introduce the pure Transformer architecture for the first time to the object re-identification (ReID) task in the anti-cheating scenario of a university campus run, and propose a novel feature-level rotation enhancement strategy to improve the robustness of the model to rotational transformations. The method effectively mitigates the information loss problem caused by traditional image-level rotation by embedding a learnable feature rotation module within the Transformer framework. The experimental results show that the improved TransReID model achieves significantly better performance enhancement than the baseline model on the MSMT17 benchmark dataset, in which the mAP metric reaches 69.4% and the Rank-1 accuracy is improved to 86.2%.This study not only verifies the effectiveness of the Transformer architecture in the ReID task for complex scenes, but also provides a new technical path for developing robust recognition systems suitable for real-world applications such as intelligent surveillance and campus security. Future research can further explore the direction of multimodal fusion, lightweight network design, etc., to promote the scale application of this technology in the construction of smart cities.

Acknowledgements

Fund Project: Research Project on Undergraduate Teaching Reform of University of Science and Technology Liaoning (XJGSZ202404).

References

[1] Ying-Cong Chen, Xiatian Zhu, Wei-Shi Zheng, and Jian-Huang Lai. 2017. Person re-identiffcation by camera correlation aware feature augmentation. IEEE TPAMI40, 2 (2017), 392–408

[2] Pirazh Khorramshahi, Neehar Peri, Jun-cheng Chen, and Rama Chellappa. The devil is in the details: Selfsupervisedattention for vehicle re-identification. In ECCV, pages 369– 386. Springer, 2020. 1, 8

[3] Hao Luo, Youzhi Gu, Xingyu Liao, Shenqi Lai, and Wei Jiang. Bag of tricks and a strong baseline for deep person re-identification. In CVPRW, pages 0–0, 2019. 1, 2, 3, 11

[4] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xi Aohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. 2020. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020).

[5] Chen S, Ye M, Du B. Rotation invariant transformer for recognizing object in uavs[C]//Proceedings of the 30th ACM International Conference on Multimedia. 2022: 2565-2574.

[6] Shuting He, Hao Luo, Pichao Wang, Fan Wang, Hao Li, and Wei Jiang. 2021.Transreid: Transformer-based object re-identiffcation. In ICCV. 15013–15022.

[7] Longhui Wei, Shiliang Zhang, Wen Gao, and Qi Tian. Person transfer gan to bridge domain gap for person reidentification. In CVPR, pages 79–88, 2018. 2, 5

[8] Zhun Zhong, Liang Zheng, Guoliang Kang, Shaozi Li, and Yi Yang. Random erasing data augmentation. In AAAI, volume 34, pages 13001–13008, 2020. 5