Person Re-Identification in Video Surveillance Across Multiple Cameras Achieving Cross-Domain Generalization

OPEN ACCESS

Volume: 13

Special Issue: 3

Month: February

Year: 2026

P-ISSN: 2321-788X

E-ISSN: 2582-0397

Citation:
C, Arun Kumar. "Person Re-Identification in Video Surveillance Across Multiple Cameras Achieving Cross-Domain Generalization." *Shanlax International Journal of Arts, Science and Humanities,* vol. 13, no. 3, 2026, pp. 424–29.

DOI:
https://doiorg/10.34293/sijash.v13iS3-i2-Feb.10315

Mr. Arun Kumar C
Assistant Professor, Department of Computer Science
S.I.V.E.T. College, India

Dr. Bhuvana Jayaraman
Associate Professor, Department of CSE
SSN College of Engineering,India

Abstract

Person re-identification (Re-ID) is a critical com ponent of intelligent video surveillance systems, aims to match individuals across non-overlapping camera views. Despite sig nificant progress, performance degradation caused by domain shifts such as variations in illumination, viewpoint, background, and camera characteristics remains a major challenge in real world deployments. This study presents a robust cross-domain person re-identification framework based on the Omni-Scale Network (OSNet) enhanced with Batch Normalization Neck (BNNeck) and MixStyle for improved domain generalization. OSNet effectively captures discriminative multi-scale visual cues, whereas BNNeck decouples feature learning for classification and metric learning, improving embedding discrimination. The proposed model was trained on a source-domain dataset and evaluated on a target-domain dataset without any target-domain supervision. These results confirm the effectiveness of combining omni-scale feature extraction with style-based augmentation for robust person re-identification in multi-camera- video surveillance environments. The key areas where the person re-identification applied in Intelligent video surveillance systems, law enforcement and criminal investigations, smart city monitoring, border control and immigration security, retail analytics and customer behavior analysis, campus and enterprise security, smart transportation hubs, healthcare and assisted living facilities, large-scale event and crowd management, and human–robot interaction in autonomous systems.

Keywords: Person Re-Identification, OSNet, MixStyle, BN Neck, Domain Generalization

Introduction

The widespread deployment of surveillance cameras in urban environments, transportation hubs, and public infrastructure has generated large volumes of video data require intelligent analysis. Person Re-Identification (ReID) is a key computer vision task that enables the recognition of individuals across non-overlapping camera views and is essential for public safety, forensic analysis, and smart city surveillance. Despite signifi cant progress, ReID remains challenging because of variations in lighting, camera viewpoints, occlusions, and background clutter. Early ReID methods relied on handcrafted features, such as color and texture descriptors, which showed limited robustness under domain shifts. Deep learning approaches, particularly Convolutional Neural Networks (CNNs), signifi cantly improved performance by learning discriminative rep resentations directly from data. However, models trained on single dataset often perform poorly when deployed in unseen environments due to domain shift. To address this issue, do main generalization techniques have gained attention because they do not require labeled target-domain data.

Omni-Scale

Network (OSNet) effectively captures multi-scale features but lacks mechanisms to handle appearance variations across do mains. This study proposes an end-to-end ReID framework that integrates OSNet with MixStyle and BNNeck to enhance cross-domain robustness. MixStyle introduces feature-level style randomization, whereas BNNeck improves embedding discrimination and training stability. The unified framework demonstrates improved generalization on benchmark datasets using standard ReID evaluation metrics.

Related Work

Person Re-Identification (ReID) has been extensively stud ied in computer vision, evolving from handcrafted feature based methods to deep learning and domain generalization approaches. This section reviews prior work relevant to the proposed framework, including traditional ReID techniques, domain adaptation and generalization strategies, and key com ponents such as OSNet, MixStyle, and BNNeck. 2.1 Tradi tional Feature-Based Methods Early ReID approaches relied on handcrafted features designed to capture pedestrian appearance, including color histograms, Local Binary Patterns (LBP), and Scale-Invariant Feature Transform (SIFT). These features are commonly combined with metric learning techniques such as KISSME and XQDA to measure cross-camera similarity. While effective under controlled conditions, such methods lack robustness to illumination changes, pose variations, and occlusions, resulting in poor generalization across different camera environments. 2.2 Domain Adaptation and Generaliza tion Despite the success of deep CNN-based models, domain shifts remain a major challenge in ReID. Domain adaptation methods, such as MMT and SPCL, leverage unlabeled target domain data through clustering and pseudo-labeling to improve performance. Although effective, these approaches require tar get data and repeated fine-tuning, which limits their scalability. Domain generalization (DG), on the other hand, aims to learn domain-invariant representations without accessing target data. DG methods employ style augmentation, feature normalization, and representation disentanglement to enhance robustness. No table examples include DG-Net and MixStyle, which demon strate strong cross-domain performance. 2.3 OSNet, MixStyle, and BNNeck OSNet introduces omni-scale feature learning by integrating multiple convolutional paths within each block, enabling simultaneous capture of fine-grained and global fea tures. Its efficiency makes it suitable for real-time surveillance applications; however, OSNet alone is sensitive to domain variations. MixStyle addresses this limitation by performing feature-level style randomization, encouraging invariance to appearance changes. BNNeck further enhances discriminative learning by decoupling , embedding and classification spaces through batch normalization. The integration of these compo nents forms a robust foundation for cross-domain ReID.

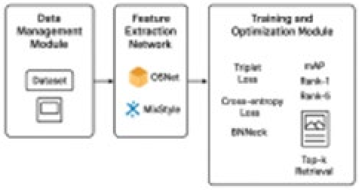

Fig. 1 Re-identification Framework osnet+mixstyle+BNNeck

Proposed OSNET + Mixstyle Framework

This study proposes a framework as shown in Fig. 1. The framework is an integrated and scalable end-to-end person Re Identification framework that combines the Omni-Scale Net work (OSNet) with MixStyle-based domain generalization and BNNeck normalization. The framework is designed to achieve robust cross-domain performance across multiple surveillance camera networks without requiring fine-tuning on target do main data. The complete system unifies data preprocessing, training, evaluation, and retrieval visualization within a single automated pipeline.

Framework Overview

The proposed framework consists of four core modules: (1) a data management module for dataset loading, preprocessing, and augmentation; (2) a feature extraction network based on OSNet with MixStyle embedded in intermediate layers; (3) a training and optimization module employing cross-entropy and triplet losses with BNNeck normalization; and (4) an evaluation and visualization module for computing mAP, Rank 1, and Rank-5 accuracy along with top-k retrieval results. Input images from benchmark surveillance datasets such as Market-1501 and DukeMTMC-ReID were resized to 256×128, normalized, and augmented to simulate real-world variations. These images were processed by the OSNet backbone, where MixStyle introduces stochastic feature-level style transforma tions during training. BNNeck is applied to the extracted em beddings to ensure stable normalization and improved feature discrimination. This design enables consistent and accurate identity matching across unseen camera domains with varying illumination, perspective, and backgrounds.

Architecture Design

OSNet was selected as the backbone due to its ability to capture omni-scale features efficiently. Each OSNet block consists of multiple parallel convolutional paths with different kernel sizes, enabling simultaneous extraction of fine-grained local details and global contextual information. Feature maps from different scales were dynamically fused through learnable gating mechanisms. MixStyle layers were inserted between intermediate OSNet blocks to randomize feature statistics, enhancing robustness to domain variations. The resulting fused features were globally pooled and passed through BNNeck normalization before classification and retrieval.

Advantages of the Proposed Framework

The integration of OSNet, MixStyle, and BNNeck offers sev eral distinct advantages: 1. Improved Domain Generalization: MixStyle introduces style diversity in feature space, making the model robust to unseen camera environments. 2. Enhanced Feature Discrimination: BNNeck ensures clean, normalized embeddings for stable distance-based retrieval. 3. Computa tional Efficiency: OSNet's lightweight structure allows real time inference on mid-range GPUs. This comprehensive com bination not only improves the accuracy of person ReID systems but also enhances their applicability in large-scale, multi-camera surveillance networks.

Methodology

This section describes the methodology adopted for imple menting the proposed OSNet + MixStyle + BNNeck frame work, including dataset preparation, preprocessing, training strategies, and evaluation protocols. Each component is de signed to enhance generalization and scalability for real-world surveillance. 4.1 Dataset Preparation and Preprocessing To ensure robust performance and cross-domain generalization, experiments were conducted on two widely used ReID bench mark datasets: Market-1501 and DukeMTMC-ReID. Market 1501 contains 32,668 images of 1,501 identities captured by six cameras in diverse indoor and outdoor environments, exhibiting significant variations in pose, light, and background. DukeMTMC-ReID comprises 36,411 images of 1,404 identi ties collected from eight cameras and presents higher domain complexity, making it suitable for evaluating generalization across unseen camera networks. Both datasets awere parti tioned into training, query, and gallery sets following standard evaluation protocols. All images were resized to 256×128 pix els, converted to RGB format, and normalized using ImageNet mean and standard deviation values to ensure stable network training. A comprehensive preprocessing pipeline wss applied to improve robustness to real-world variations. This includes random horizontal flipping to simulate viewpoint changes, random erasing to model occlusions and partial visibility, and batch shuffling to avoid sampling bias and improve feature level style mixing during training. These preprocessing strate gies enhance the model's ability to learn invariant and discrim inative representations in multiple surveillance environments.

Experimental Results and Discussion

This section presents a detailed quantitative and qualita tive analysis of the proposed OSNet + MixStyle + BNNeck framework. It includes comparisons with baseline architectures, visual evaluation of retrieval results, and an in-depth discussion of the model's performance under different conditions.

Limitations and Future Scope Bottom of Form

Although the proposed OSNet + MixStyle + BNNeck framework demonstrates strong generalization and efficiency

ResNet-50 (Baseline) Market-1501– DukeMTMC

OSNet (Original) Market-1501– DukeMTMC

71.2 87.9 66.8 74.6 90.2 71.1

across multiple surveillance domains, certain limitations re main. These open opportunities for improvement and future re search directions in person re-Identification (ReID) and broader visual recognition systems. Despite the robustness provided by

Proposed OSNet + MixStyle + BNNeck

Market-1501– DukeMTMC

78.4 92.6 75.3

MixStyle and BNNeck, performance degrades under severe occlusions, as the model relies mainly on global appearance

Quantitative Evaluation

The quantitative results demonstrate the effectiveness of the proposed framework in both intra-domain and cross-domain scenarios. Table 3 compares the performances of the baseline ResNet-50, vanilla OSNet, and the proposed OSNet + MixStyle + BNNeck model.

Quantitative Performance Comparison Across Datasets

The results clearly show that the proposed framework consis tently outperforms the baseline models. The results are shown in Table 1. On Market-1501, Rank-1 accuracy improved by 6%–7% compared with ResNet-50, whereas mAP increased by nearly 9. Similarly, on DukeMTMC-ReID, Rank-1 and mAP improved by 8.4% and 9.4%, respectively. These improvements confirm that the integration of MixStyle and BNNeck signifi cantly enhances both discriminability and generalization. The experimental results validate the efficacy of the proposed model in addressing both intra-domain and cross-domain challenges: • Generalization: The framework's consistent performance across datasets underscores its suitability for real-world surveil lance where camera conditions vary drastically. • Stability: BNNeck stabilizes optimization and maintains discriminative feature embeddings. • Efficiency: Despite its improvements, the model retains a lightweight architecture suitable for real time inference on mid-range GPUs. • Practicality: Automated checkpointing, evaluation, and retrieval visualization render the system highly deployable in constrained computational environments such as cloud or edge devices. The combination of MixStyle and BNNeck within OSNet represents a balanced, practical, and high-performing approach to cross-domain per son ReID.

Practical Implications

The proposed OSNet + MixStyle + BNNeck framework irs designed for real-world deployment in large-scale surveillance systems. Its computational efficiency enables integration into smart city infrastructures for cross-camera person tracking on cloud or edge platforms. The framework is particularly effective for forensic and security analyses, offering domain invariant recognition and interpretable top-k retrieval results. Lightweight inference supports real-time operations on edge devices such as NVIDIA Jetson platforms. Ethical deployment is emphasized through privacy-preserving measures, including embedding anonymization, access control, and compliance with regulations. The modular and scalable design reduces retraining costs and supports seamless industrial and research adoption.

cues that may be insufficient when pedestrians are partially visible. The framework processes individual frames inde pendently, ignoring temporal information that could improve consistency in video sequences. Performance also depends on ImageNet-pretrained OSNet backbones, which may not fully reflect surveillance data distributions. Large-scale retrieval remains computationally demanding for extensive galleries, and the current systemsupports only RGB data. Integrating temporal modeling, efficient retrieval mechanisms, and multi modal features such as RGB–IR fusion could further enhance performance.

Conclusion

This paper presented Person re-identification in a video surveillance across multiple cameras achieving cross domain generalization framework integrating OSNet with MixStyle and BNNeck to address cross-domain challenges in multi camera surveillance. Experimental results on Market-1501 and DukeMTMC-ReID achieved 75.3% mAP and 78.4% Rank 1 accuracy, outperforming conventional baselines. MixStyle improved robustness to domain shifts, while BNNeck enhanced feature discrimination and training stability. The framework demonstrated strong generalization to unseen domains without fine-tuning and supported scalable deployment on cloud and edge platforms. Its modular and reproducible design advances practical ReID research. Future work includes temporal model ing, multi-modal fusion, self-supervised learning, and privacy aware deployment strategies.

References

1. Ahmed, E., Jones, M., & Marks, T.K., 2015. An improved deep learning architecture for person re-identification. Proceed ings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.3908–3916.
2. Bak, S., Carr, P. & Lalonde, J.F., 2018. Domain adaptation through synthesis for unsupervised person re-identification. Eu ropean Conference on Computer Vision (ECCV), pp.189–205.
3. Chang, X. et al., 2018. Multi-level factorization net for person re-identification. CVPR, pp.2109–2118.
4. Chen, Y., Zhu, X.& Gong, S., 2018. Person re-identification by deep learning multi-scale representations. IEEE Transac tions on Image Processing, 27(1), pp.282–294. 5. Cheng, D.,
5. Gong, Y., Zhou, S., Wang, J. & Zheng, N., 2016. Person re-identification by multi-channel parts-based CNN with improved triplet loss function. CVPR, pp.1335–1344. 6. Dai, Z., Chen,
6. K. & Wu, Z., 2021. Generalizable person re identification by domain-invariant mapping. IEEE Transactions on Image Processing, 30, pp.3341–3354.
7. Fan, H., Zheng, L., Yan, C., & Yang, Y., 2018. Unsuper vised person re-identification: Clustering and fine-tuning. ACM Transactions on Multimedia Computing, Communications, and Applications, 14(4), pp.1–18.
8. Ge, Y., Chen, D., Li, H., Zhao, S. & Li, R., 2020. Mutual mean-teaching: Pseudo label refinery for unsupervised domain adaptation on person re-identification. International Conference on Learning Representations (ICLR).
9. Gray, D. & Tao, H., 2008. Viewpoint invariant pedestrian recognition with an ensemble of localized features. ECCV, pp.262–275.
10. Guo, H., Yuan, Y., & Guo, Y., 2022. Domain generaliza tion via style randomization for person re-identification. IEEE Access, 10, pp.9941–9953
11. He, K., Zhang, X., Ren, S. & Sun, J., 2016. Deep residual learning for image recognition. CVPR, pp.770–778. 12.
12. Hermans, A., Beyer, L. & Leibe, B., 2017. In defense of the triplet loss for person re-identification. arXiv preprint arXiv:1703.07737.
13. Huang, Y., Xu, J. & Lin, J., 2020. Multi-domain mixup for person re-identification. Pattern Recognition Letters, 136, pp.123–130.
14. Ioffe, S. & Szegedy, C., 2015. Batch normalization: Ac celerating deep network training by reducing internal covariate shift. ICML, pp.448–456.
15. Khan, M.H., Ahmad, R. & Kim, S.H., 2022. MixStyle augmented OSNet for domain generalization in person re identification. Sensors, 22(14), 5298.
16. Khosla, P., Teterwak, P., Wang, C., Sarna, A., Tian, Y. & Isola, P., 2020. supervised contrastive learning. NeurIPS, 33, pp.18661–18673.
17. Krizhevsky, A., Sutskever, I. & Hinton, G.E., 2012. Ima geNet classification with deep convolutional neural networks. NeurIPS, 25, pp.1097–1105.
18. Li, W., Zhao, R., Xiao, T. Wang, X., 2014. DeepReID: Deep filter pairing neural network for person re-identification. CVPR, pp.152–159.
19. Li, Y., Zhu, J. & Gong, S., 2020. Domain generalization for person re-identification via style-invariant feature learning. Pattern Recognition, 104, 107347.
20. Lin, Y., Zheng, L., Zheng, Z., Wu, Y. & Yang, Y., 2019. Improving person re-identification by attribute and identity learning. Pattern Recognition, 95, pp.151–161.