Jurnal Ilmiah
Teknologi Sistem
Informasi

Vol. 7 (2026) No. 1

https://jurnal-itsi.org/index.php/jitsi; E-ISSN: 2722-4600; ISSN: 2722-4619
DOI: 10.62527/jitsi.7.1.543

Robust Person Identification with Channel Attention and Multi-Scale Feature Extraction

R. Newlin Shebiah[#], S. Arivazhagan[#], J. Kanishkashree[#], S. Sankara Gomathi[#]

[#] *Centre for Image Processing and Pattern Recognition, Department of Electronics and Communication Engineering*
Mepco Schlenk Engineering College, Sivakasi - 626005, Tamilnadu, India
E-mail: newlinshebiah[at]mepcoeng.ac.in

ABSTRACTS

Person recognition under varied conditions is a critical task that aims to accurately identify individuals captured from different angles or at different times. One of the primary challenges in this field is occlusion, which significantly degrades recognition performance. To address this issue, we propose an advanced attention-based network designed to mitigate the effects of occlusion and enhance recognition accuracy. Our approach leverages channel attention to dynamically recalibrate the importance of each channel, utilizing both width and depth attention mechanisms to emphasize discriminative and informative features. The network employs a multi-scale feature extraction strategy, partitioning feature maps to capture multi-level representations of the human body. The concatenation of results from these attention stages facilitates the integration of local and global features, effectively reducing the impact of occlusion. We evaluate the proposed model on multiple benchmark datasets, including PRID 2011, iLIDS-VID, and Market-1501. The experimental results demonstrate that our model achieves superior performance, attaining a top accuracy of 99.79% on the PRID 2011 dataset, 98.55% on the iLIDS-VID dataset, and 88.24% on the Market-1501 dataset.

Manuscript received Nov 6, 2025; revised Dec 24, 2025. accepted Dec 25, 2025 Date of publication Dec 31, 2025. International Journal, JITSI : Jurnal Ilmiah Teknologi Sistem Informasi licensed under a Creative Commons Attribution-Share Alike 4.0 International License

Keywords / Kata Kunci — ***Person Re-ID; occlusion; Width channel attention; Depth channel attention; Surveillance***

CORRESPONDING AUTHOR

Abdimalik Aden Ibrahim
Faculty of Computer Science and information technology at capital university of Somalia, Mogadishu, Somalia
Email: abdimalikadan[at]gmail.com

1. INTRODUCTION

Person re-identification (Re-ID) is a popular technique in the computer vision industry, aiming to match the same person viewed from different angles or time periods [1], [2]. Re-identification is challenging due to individual differences in appearance, such as viewpoint, illumination, resolution, occlusion, and color. Despite these challenges, conventional methods have served as a foundation for Person Re-Identification and provided insights for developing more sophisticated techniques. The development of deep learning-based systems has led to major advancements in re-id techniques [3]. Large-scale person re-identification datasets, such as MSMT17 [4], CUHK03 [5], Market-1501[6], Duke MTMC-ReID [7], and MARS [8], provide extensive training data for human re-identification models. These advancements, continuous research, and interdisciplinary collaborations have led to more precise, reliable, and scalable solutions for real-world applications in security, surveillance, and human-computer interaction.

Lu et al. [9] introduced the Dual-branch adaptive attention transformer for occluded person re-identification, extracting a person's local and global properties simultaneously. Qin et al. [10] developed the Width and Depth Channel Attention Network (WDC-Net) to address missing information caused by occlusion in person re-identification. Guo et al. [11] proposed a survey on attention mechanisms in computer vision, providing insights into their function in improving visual perception and performance. Li et al. [12] proposed a unique method for person re-identification using graph neural networks (GNNs) and multi-scale attention processes, which performs better than current approaches, especially when handling posture, light, and occlusion fluctuations. Si et al. [13] developed the Dual Attention Matching Network for Context-Aware Feature Learning in Person Re-identification, which improves discriminative feature learning for more accurate person matching across various camera perspectives. Zhang et al. [14] presented relation-aware global attention for person re-identification, leveraging relation-aware global attention to efficiently gather discriminative information across different body parts and geographical regions in pedestrian images. Zhan et al. [15] proposed strategy incorporates pose information and attention mechanisms to suppress unnecessary information and concentrate on distinguishable body parts, yielding state-of-the-art performance on person-ID tasks. SSMTReID-Net is an architecture designed by Mohanty et al. [16] to solve the unsupervised multi-target domain adaptation problem using the EWC regularizer and information bottleneck concept. Pang et al. [17] developed a camera invariant feature learning method to address the unsupervised person ReID problem, resulting in improvements in mAP and Rank-1 with various datasets.

Zhang et al. [18] developed a retrieval-verification framework for cloth-changing human re-identification, outperforming advanced techniques on artificial and real-world datasets. Song et al. [19] generated the first Visible-Infrared Clothes-Changing dataset for person re-identification and developed a unique Semantic-Constraint garments-Changing Augmentation Network to solve modality mismatch and intra-class disagreement caused by changing garments. Shao and Ling [20] proposed Dynamic Curriculum Learning for Weakly Supervised Person Re-identification, which uses dynamic curriculum learning to improve model performance by adding increasingly difficult training examples. This approach shows significant gains over baseline methods in resolving poorly supervised person re-identification problems. Yu et al. [21] introduced a novel approach for weakly supervised person re-identification using multi-level self-paced learning, achieving competitive performance with fewer annotation requirements. Shi et al. [22] proposed a lightweight person re-identification network to address poor accuracy issues in in-person re-identification due to factors like occlusions and illumination conditions.

Ge et al. [48] proposed triplet probability learning for person re-identification, enhancing the discriminative power of learned representations. Xu et al. [24] developed a method for re-identification using Bilinear Transformation Networks, integrating multimodal information. Zheng et al. [25] developed Learning to Fuse Local and Global Representations for Person Re-identification, combining local and global representations for better performance. Person Re-identification in the Wild [26] addresses difficulties in unconstrained situations, improving performance under conditions like shifting occlusion, lighting, and position. Zhu et al. [27] proposed Deep Fusion of Global and Local Representations for Person Re-identification, using deep learning techniques to integrate global and local information. This method outperforms baseline solutions, especially in managing difficult conditions like posture, lighting, and occlusion fluctuations. The work advances person re-identification by improving matching accuracy across various camera viewpoints.

Liu et al. [28] introduced the Adaptive Feature Fusion Network for Occlusion-Aware Person Re-identification, a method for person re-identification that addresses occlusions in pedestrian photos. This strategy integrates information from different sources, improving the robustness of re-ID systems against occlusions while preserving discriminative strength. Zhang et al. [29] proposed Learning to Adapt Template for Person Re-identification, which dynamically adjusts templates according to specific queries and gallery photos. Xu et al. [30] proposed the Cross-modal Channel Exchange Network (CmCEN), which achieved competitive and even better performance compared to SOTA models. Machaca et al. [31] suggested TrADe, a novel live Re-ID method that combines two concepts to create a lower high-quality gallery. Zang et al. [32] created a paper that employed numerous self-supervision processes to simulate different tough problems and solve each problem using a separate network, providing state-of-the-art performances on three ReID benchmarks and two occluded benchmarks.Fair Architecture Search for Person Re-Identification (FAS-ReID) [33] is a neural network foundation that is more reliable and resilient than other methods.

Li et al. [34] proposed an adversarial open-set person re-identification approach, while Sun et al. [35] presented a novel method for person retrieval that combines refined part pooling techniques to enhance feature representation and discriminative capability. Huang et al. [36] proposed ACINet, which integrates local and global contextual information for person re-identification performance. Wu et al. [37] introduced a unique method for person re-identification spanning visible and thermal imaging modalities, integrating data from both modalities to improve accuracy. Their method's efficacy in achieving higher performance in cross-modality person re-identification tasks is crucial for advancing multi-modal re-identification systems in multimedia research.

Wei et al. [38] proposed Temporal Consistency Preserving Person Re-identification, which maintains temporal consistency in video sequences, improving re-ID accuracy over several frames. Sun et al. [39] Deep Adaptive Fusion Network for Person Re-identification dynamically fuses multi-modal input from various sources to increase re-ID accuracy. Gong et al. [40] proposed Disentangling Cross-view Correlation for Multi-shot Person Re-identification, which uses a disentangled representation learning framework to extract viewpoint-specific information from view-invariant traits. These techniques offer new insights into disentanglement strategies to improve person re-identification systems' accuracy and generalization capacities. Shen et al. [41] developed Temporal Consistent Adaptive Learning for Pedestrian Re-identification, a method that uses temporal information and adaptive learning to ensure uniformity in pedestrian appearances and flexibility in various environments. This strategy outperforms current methods on benchmark datasets, demonstrating the potential of adaptive learning techniques and temporal consistency in enhancing pedestrian re-identification systems' robustness and accuracy

2. RESEARCH METHODOLOGY

The proposed architecture for human recognition using two attention modules: width and depth channel attention is shown in Fig. 1. The ResNet-1 architecture extracts low-level channel attention information from input images, which is then used in squeeze and excitation networks to enhance channel-wise dependencies. The output of SENet is subjected to both depth channel attention and width channel attention independently to determine dependencies within a feature map.

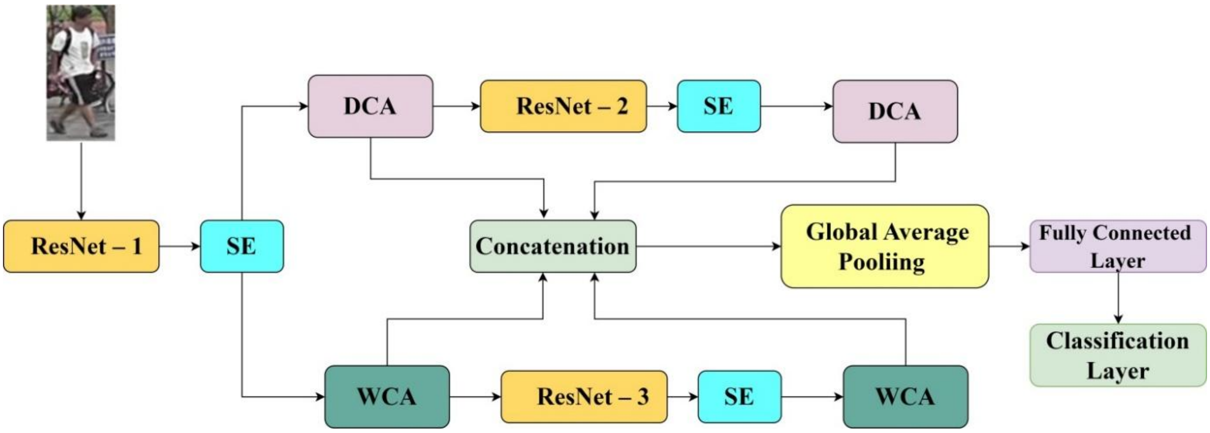

FIG 1. Proposed Architecture for View invariant Human Recognition

Residual Neural Networks (ResNets) [42] are deep learning architectures used to overcome vanishing gradients in deep neural networks. ResNet blocks (Fig 2) aid gradient flow during backpropagation, enabling the training of extremely deep neural networks. ResNet's effectiveness in human re-identification is due to its ability to learn discriminative characteristics against changes in attitude, illumination, and backdrop in real-world surveillance scenarios. To improve performance, ReLUactivation functions are added to residual blocks of ResNet structures, enhancing performance. The suggested method uses ResNet stages 1, 2, and 3 to extract low-level features and features with deeper layers

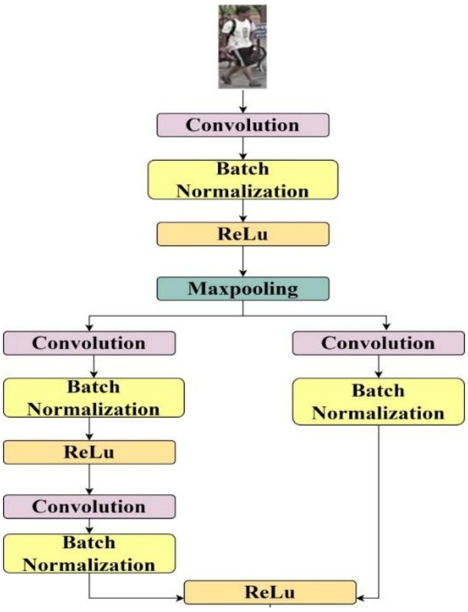

FIG 2. Residual Block

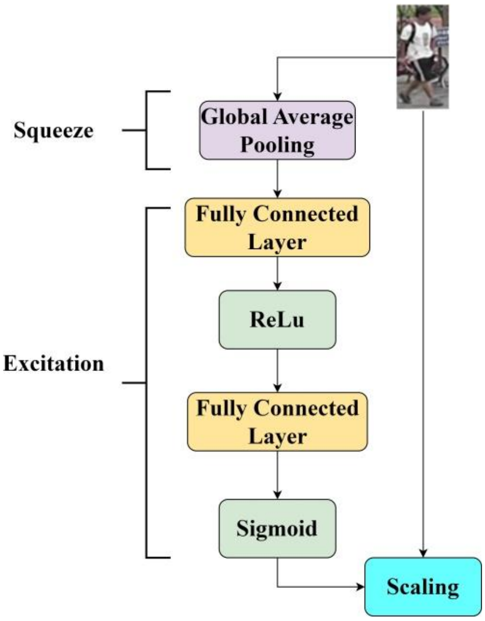

FIG 3. Squeeze and Excitation Stage

The Squeeze-and-Excitation (SE) network (Fig. 3) aims to enhance neural network representational capacity, particularly in computer vision applications. It begins with a "squeeze" operation, reducing the input tensor's spatial dimensions. Global average pooling determines feature map average values. The network then executes an "excitation" operation, capturing channel-wise dependencies. Two fully connected layers are involved, with activation functions like sigmoid and ReLU used.

Width Channel Attention:

The width channel attention (WCA) technique captures spatial correlations on a feature map, allowing networks to focus on crucial locations by calculating attention weights across spatial dimensions, as illustrated in Fig. 4.

The process of creating part-channel attention maps involves aggregating input features and using global average pooling to compute each channel's average activation value, resulting in a global descriptor for that channel, and then calculating the output mathematically for a feature map F as in Equation 3.1.

$$z = f(u_c) = \frac{1}{H\times w}\Sigma\Sigma u_c(i,j) \qquad (1)$$

Channel-wise weights are learned using global descriptors, local cues are highlighted using max-pooling, and attention maps are spliced based on input features' dimensions, compressing and expanding the channel feature maps. Equation 2 provides an explanation of the above.

$$W_i = \sigma\left[W_2\left(\text{Re }\dot{L}U(W_1, z)\right)\right] \qquad (2)$$

Where W_1, W_2 are refers to weights which can be as mentioned in Equation 3:

$$W_i \in R^{\frac{c}{r}\times c} \text{or } \; W_i \in R^{c\times\frac{c}{r}} \qquad (3)$$

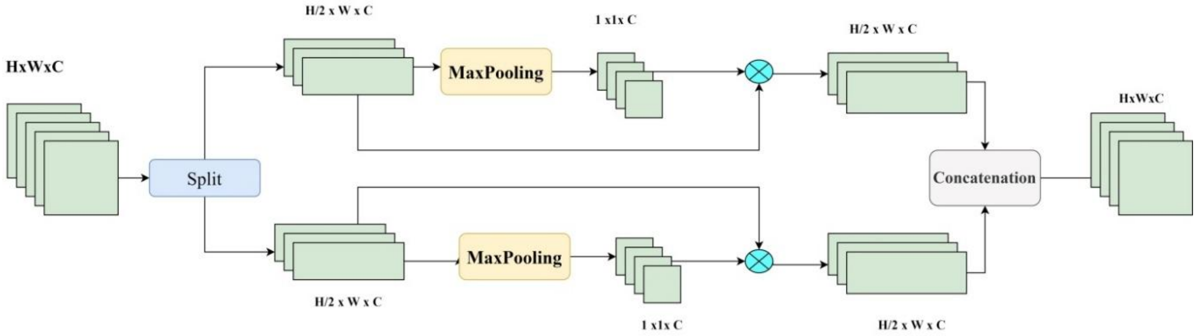

FIG 4. Width Channel Attention

Depth Channel Attention

DCA aims to create a global deep fusion of attention maps from previous modules. It uses a group-based approach to break down input features into channels as illustrated in Fig. 5, combine their characteristics, and then average global pool to obtain discriminative group features. The attention map for each group is generated by multiplying feature weights to all groups.

$$f = Cat\left\{Softmax\left(W_2\left(\text{Re}\dot{L}U\left(BN(W_1, F_{Avg})\right)\right)\right) * f_i\right\} \qquad (4)$$

Where F_{Avg} refers to Average Pooling of (F)

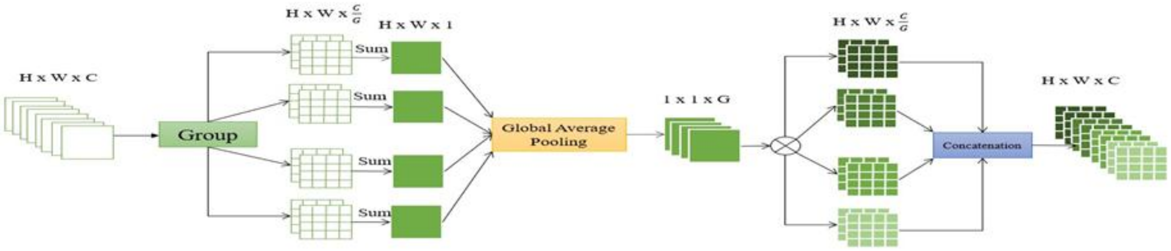

FIG 5. Depth Channel Attention

Equation 4 explains the DCA process, which reduces network parameters. Combining WCA and DCA outputs improves network performance by capturing human semantic information and removing unnecessary information, minimizing occlusion and improving network performance.

3. Results And Discussion

Datasets Used: The work utilized PRID2011, iLIDS-VID, and Market-1501 datasets, which will be further explained in the following section. The PRID 2011 dataset (Fig. 6) is a person re-identification dataset utilizing images from two surveillance cameras, consisting of single-shot and multi-shot versions, containing 385 frames of 201 identities. The iLIDS-VID dataset (Fig. 7) comprises 600 image sequences of 300 individuals, with each sequence ranging from 23 to 192 frames. Market-1501 is a widely used dataset in computer vision, particularly for applications involving person re-identification (Fig. 8) consisting of 12,936 images of 751 identifications in the training set and 19,732 images of 750 identifications in the testing set.

Fig 6. Some sample images from PRID2011 dataset

Fig 7. Some sample images iLIDS-VID dataset

Fig 8. Some sample images from Market-1501 dataset

Experimentation on Market-1501 dataset:

In evaluating various architectural configurations and regularization techniques for person re-identification models, we conducted experiments on a dataset comprising 100 classes over 30 epochs, aiming to optimize accuracy and robustness. The results, summarized in Table 1, highlight the impact of different architectural components and regularization strategies on model performance. Beginning with foundational architectures, such as a single ResNet combined with Squeeze & Excitation (SE), we achieved a respectable testing accuracy of 84.11%. Introducing Width Channel Attention (WCA) alongside SE slightly decreased accuracy to 82.24%, suggesting a nuanced balance between feature enhancement and complexity. Further augmenting with Depth Channel Attention (DCA) notably boosted accuracy to 85.67%, illustrating the complementary roles of width and depth-based feature selection mechanisms in capturing nuanced visual cues.

To enhance generalization and mitigate overfitting, we explored regularization techniques, starting with L2 regularization set to 0.01. This regularization strategy, applied across SE, WCA, and DCA components, maintained stability in training while marginally reducing accuracy to 82.55%, indicating its efficacy in preventing model complexity from overshadowing training signal clarity. Introducing additional architectural layers—such as convolutional layers, batch normalization (BN), and rectified linear units (ReLU)—further refined feature extraction, culminating in a testing accuracy of 83.18%, underscoring the role of layer depth in capturing intricate visual features critical for re-identification tasks.

Notably, optimizing regularization parameters revealed significant insights into model behavior, where lower regularization (L2 = 0.001) marginally improved accuracy to 84.11%, emphasizing the delicate trade-off between regularization strength and model performance. Expanding the architectural complexity with dual ResNet stages, each integrating SE and WCA, resulted in the highest observed accuracy of 87.54%. This configuration highlights the synergistic benefits of multi-stage ResNet architectures in effectively synthesizing and refining feature representations across multiple hierarchical levels, crucial for robust person re-identification across diverse environmental conditions and dataset complexities.

Moreover, integrating Depth Channel Attention (DCA) into multi-stage ResNet setups maintained competitive accuracy, achieving 84.42% accuracy, reinforcing its role in enhancing feature discriminability within deeper network configurations. Extending these findings to encompass dual ResNet stages with SE, WCA, and DCA components reaffirmed their collective efficacy, achieving a testing accuracy of 85.67%. This holistic approach underscores the importance of systematically integrating attention mechanisms and depth-sensitive feature refinements to harness comprehensive spatial and semantic cues essential for accurate person re-identification. The most optimized configuration, featuring dual ResNet structures with SE, WCA, and DCA, achieves the highest accuracy of 88.24%, showcasing comprehensive feature extraction capabilities and robust model performance.

The comprehensive evaluation of architectural configurations and regularization techniques on the dataset over 30 epochs provides a nuanced understanding of their collective impact on model performance. These findings not only validate the efficacy of SE, WCA, and DCA in enhancing feature discrimination but also underscore the critical role of regularization in balancing model complexity and generalization. Moving forward, these insights pave the way for future advancements in person re-identification research, advocating for tailored architectural designs that leverage attention mechanisms and regularization strategies to optimize accuracy and robustness across challenging datasets.

TABEL 1. Performance Comparison of Architectural Configurations in Person Re-Identification Models for Market-1501 dataset

Architecture	Accuracy	Precision	Recall
ResNet+SE	86.60%	79.54%	81.23%
ResNet+SE+WCA	82.24%	76.31%	78.65%
ResNet+SE +WCA+DCA	85.67%	78.47%	79.70%
2(ResNet+SE+ WCA)	87.54%	75.36%	78.36%
2(ResNet+SE+ DCA)	84.42%	75.73%	76.39%
2(ResNet+SE+WCA+ DCA)	88.24%	78.36%	79.34%

Jing et al. [43] introduced an Attention-Aware Compositional Network achieving an accuracy of 85.9% in person re-identification. Song et al. [44] developed a Mask-guided Contrastive Attention Model, demonstrating an accuracy of 83.6%. Zhao et al. [45] focused on learning person saliency and matching saliency distribution, achieving an accuracy of 88.9%. Liu et al. [46] proposed the UnityStyle adaption method to harmonize style differences within and across different cameras, achieving an accuracy of 93.2%. These methods highlight advancements in enhancing accuracy through attention mechanisms and style adaptation in person re-identification systems.

Experimentation on PRID2011 dataset:

Our study on the PRID 2011 dataset yielded compelling results, showcasing a remarkable testing accuracy of 99.79%. This achievement underscores the effectiveness of our proposed architecture in accurately re-identifying individuals in surveillance scenarios. The architecture integrates a robust ResNet backbone, complemented by SE (Squeeze & Excitation) blocks, and width and depth channel attention mechanisms. These components collectively enhance feature discrimination, enabling the model to capture and utilize intricate details from surveillance images more effectively. The high accuracy achieved on PRID 2011 highlights the architectural prowess in handling diverse environmental conditions and varying perspectives, crucial for reliable person re-identification in real-world applications.

Experimentation on iLIDS-VID dataset:

In our evaluation using the iLIDS-VID dataset, our proposed architecture demonstrated exceptional performance with 100% training accuracy and 98.55% testing accuracy. This dataset-specific achievement reaffirms the architecture's robustness and efficiency in person re-identification tasks. Achieving perfect training accuracy underscores the model's capability to learn complex patterns inherent in video sequences, while the high testing accuracy validates its ability to generalize well across different individuals and scenarios. These results on the iLIDS-VID dataset underscore the architecture's adaptability and effectiveness in diverse surveillance environments, setting a strong foundation for practical applications in enhancing security and surveillance systems.Table 2 summarizes the performance of the proposed model on PRID 2011 dataset and iLIDS dataset.

TABEL 2. Evaluation metrics for PRID2011 dataset and iLIDS-VID dataset

Dataset	Accuracy	Precision	Recall
PRID2011	99.79%	99.86%	99.42%
iLIDS-VID	98.55%	98.88%	98.54%

4. CONCLUSIONS

The network that is suggested in this work uses the ResNet architecture to outperform other currently used methods in exploring the various feature maps of humans. It does this by utilising two attention mechanisms, namely width channel attention and depth channel attention, to investigate the local and global features of human body parts. Additionally, the approach aims to address the occlusion issue by investigating a variety of subtle hints. It performs better than previous methods and yields outstanding results on a variety of datasets. The proposed work demonstrates how well attention mechanisms such as WCA and DCA are incorporated into deep learning models for person re-identification difficulties. By comparing the contributions of WCA and DCA using various datasets, the unique impacts of each on model performance are investigated. The accuracy rates obtained on the Market-1501, iLIDS-VID, and PRID 2011 datasets show how flexible this method is in both standard and occluded re-identification circumstances. The model achieved the highest accuracy of 99.79% on PRID 2011 dataset followed by 98.55% on iLIDS-VID dataset and 88.24% on Market-1501 dataset. These results demonstrate how important it is to explore a range of complex cues, such as attention mechanisms, to enhance the discriminating power of person re-identification algorithms. The findings of this study advance the field of human recognition in many useful applications, such as security and video monitoring

REFERENSI

[1] M. Cokbas, J. Bolognino, J. Konrad, and P. Ishwar, “FRIDA: Fisheye re-identification dataset with annotations,” in Proc. 18th IEEE Int. Conf. Adv. Video Signal Based Surveillance (AVSS), Nov. 2022, pp. 1–8.

[2] N. V. Mahendran, “Variations of squeeze and excitation networks,” arXiv e-prints, arXiv:2304.06502v2 [cs.CV], Jul. 3, 2023.

[3] Q. Xie, Z. Lu, W. Zhou, and H. Li, “Improving person re-identification with multi-cue similarity embedding and propagation,” IEEE Trans. Multimedia, 2022.

[4] C. Gong, J. Hu, L. Zhang, W. Liu, and X. Wang, “Spatio-temporal multi-graph convolutional network for video person re-identification,” in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), 2023.

[5] W. Hu, L. Wang, and C. Peng, “A method for detecting anomalies in an electromagnetic environment situation using a dual-branch prediction network,” Electronics, vol. 11, no. 16, p. 2555, 2022.

[6] Z. Jiang, Y. Wang, W. Liu, L. Zhang, and J. Chen, “GLAF: Group-based local affinity feature for cross-domain person re-identification,” in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), 2023.

[7] X. Wang, Y. Zhang, J. Li, W. Liu, and J. Chen, “Deep multi-resolution network for person re-identification,” in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), 2023.

[8] X. Zhang, X. Zhang, L. Wu, C. Li, X. Chen, and X. Chen, “Domain adaptation with self-guided adaptive sampling strategy: Feature alignment for crossuser myoelectric pattern recognition,” IEEE Trans. Neural Syst. Rehabil. Eng., vol. 30, pp. 1374–1383, 2022.

[9] Y. Lu, M. Jiang, Z. Liu, and X. Mu, “Dual-branch adaptive attention transformer for occluded person re-identification,” Image Vis. Comput., vol. 131, p. 104633, 2023.

[10] Wencheng Qin, Baojin Huang, Pinzhong Qin, Zhiyong Huang, Daidi Zhong, Learning diverse and deep clues for person reidentification, Image and Vision Computing, Volume 126, 2022, 104551, ISSN 0262-8856, https://doi.org/10.1016/j.imavis.2022.104551.

[11] Guo, M.-H., Xu, T.-X., Liu, J.-J., Liu, Z.-N., Jiang, P.-T., Mu, T.-J., Zhang, S.-H., Martin, R. R., Cheng, M.-M., & Hu, S.-M. (2021). Attention Mechanisms in Computer Vision: A Survey. arXiv. https://doi.org/10.48550/ARXIV.2111.07624

[12] L. Li, B. Chen, K. Huang, et al., “Multiscale attention graph neural networks for person re-identification,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2023.

[13] J. Si, H. Zhang, C. G. Li, J. Kuen, X. Kong, A. C. Kot, and G. Wang, “Dual attention matching network for context-aware feature sequencebased person re-identification,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2018, pp. 5363–5372.

[14] Z. Zhang, C. Lan, W. Zeng, X. Jin, and Z. Chen, “Relation-aware global attention for person reidentification,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2020, pp. 3186–3195.

[15] H. Zhan, L. Zheng, Y. Wang, et al., “Pose-guided attention network for person re-identification,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2021.

[16] A. Mohanty, B. Banerjee, and R. Velmurugan, “SSMTReID-Net: Multi-target unsupervised domain adaptation for person re-identification,” Pattern Recognit. Lett., vol. 163, pp. 40–46, 2022.

[17] Z. Pang, L. Zhao, Q. Liu, and C. Wang, “Camera invariant feature learning for unsupervised person re-identification,” IEEE Trans. Multimedia, 2022.

[18] R. Zhang, Y. Fan, H. Song, F. Wan, Y. Fu, H. Kato, and Y. Wu, "A novel retrieval-verification framework for cloth changing person re-identification," Pattern Recognit., vol. 134, 2023.

[19] X. Wei, K. Song, W. Yang, Y. Yan, and Q. Meng, "A visible infrared clothes-changing dataset for person re-identification in natural scene," Neurocomputing, vol. 569, p. 127110, 2024.

[20] M. Shao and H. Ling, "Dynamic curriculum learning for weakly supervised person reidentification," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2023.

[21] X. Yu, J. Song, Y.-Z. Song, et al., "Weakly supervised person re-identification with multi-level self-paced learning," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2023.

[22] C. Shi, D. Niu, H. Gong, M. Zhang, Z. Cao, and Y. Jin, "Person reidentification lightweight network based on progressive attention mechanism," in Proc. Int. Symp. Autonomous Syst. (ISAS), Jun. 2023, pp. 1– 6.

[23] Y. Ge, F. Zhu, X. Liu, et al., "Triplet probability learning for person re-identification," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2020.

[24] S. Xu, Y. Hou, Z. Li, and D. Cao, "Joint face and person re-identification with bilinear transformation networks," IEEE Trans. Circuits Syst. Video Technol., vol. 31, no. 9, pp. 3347–3360, Sep. 2021.

[25] Z. Zheng, X. Zhu, S. Gong, et al., "Learning to fuse local and global representations for person re-identification," IEEE Trans. Pattern Anal. Mach. Intell., vol. 45, no. 7, pp. 2429–2442, Jul. 2023.

[26] Q. Leng, M. Ye, and Q. Tian, "A survey of open-world person reidentification," IEEE Trans. Circuits Syst. Video Technol., vol. 30, no. 4, pp. 1092–1108, Apr. 2019.

[27] X. Zhu, J. Zhang, J. Shi, et al., "Deep fusion of global and local representations for person re-identification," IEEE Trans. Image Process., vol. 32, pp. 1576–1589, 2023.

[28] J. Liu, Y. Chen, X. Nie, et al., "Adaptive feature fusion network for occlusionaware person re-identification," Pattern Recognit., 2023.

[29] X. Zhang, Z. Zhang, H. Chen, et al., "Learning to adapt template for person reidentification," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2021.

[30] X. Xu, S. Liu, N. Zhang, G. Xiao, and S. Wu, "Channel exchange and adversarial learning guided cross-modal person re-identification," Knowl.-Based Syst., vol. 257, p. 109883, 2022.

[31] L. Machaca, J. Huaman, E. Clua, and J. Guerin, "TrADe Re-ID–Live person re-identification using tracking and anomaly detection," in Proc. IEEE Int. Conf. Mach. Learn. Appl. (ICMLA), Dec. 2022, pp. 449–454.

[32] X. Zang, G. Li, W. Gao, and X. Shu, "Learning to disentangle scenes for person re-identification," Image Vis. Comput., vol. 116, p. 104237, Dec. 2021.

[33] S. Chen, S. Chen, and Z. Lei, "FAS-ReID: Fair architecture search for person re-identification," in Proc. IEEE Int. Conf. Inf. Technol Big Data Artificial Intell. (ICIBA), vol. 3, May 2023, pp. 544–551.

[34] X. Li, A. Wu, and W. S. Zheng, "Adversarial open-world person re-identification," in Proc. Eur. Conf. Comput. Vis. (ECCV), 2018, pp. 280–296.

[35] Y. Sun, L. Zheng, Y. Yang, Q. Tian, and S. Wang, "Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline)," in Proc. Eur. Conf. Comput. Vis. (ECCV), 2018, pp. 480–496.

[36] J. Huang, W. Chen, Y. Zhang, et al., "ACINet: Adaptive context integration network for person re-identification," in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), 2021.

[37] Z. Wu, K. Yan, J. Li, et al., "Visible thermal person re-identification with deep fusion module," IEEE Trans. Multimedia, vol. 23, pp. 1513–1524, 2021.

[38] Y. Wei, H. Fu, Z. Zheng, et al., "Temporal consistency preserving person reidentification," IEEE Trans. Pattern Anal. Mach. Intell., vol. 44, no. 3, pp. 651–664, Mar. 2022.

[39] C. Sun, Y. Yao, Y. Zhou, et al., "Deep adaptive fusion network for person re-identification," IEEE Trans. Image Process., vol. 32, pp. 2556–2569, 2023.

[40] Z. Gong, X. Yu, S. Zhang, et al., "Disentangling cross-view correlation for multi-shot person reidentification," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2021.

[41] Y. Shen, Z. Yuan, Z. Wang, et al., "Temporal consistent adaptive learning for pedestrian re-identification," IEEE Trans. Image Process., vol. 30, pp. 4000–4013, 2021.

[42] He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep Residual Learning for Image Recognition (Version 1). arXiv. https://doi.org/10.48550/ARXIV.1512.03385

[43] X. Jing, R. Zuo, F. Zhang, H. Wang, and W. Ouyang, "Attention-aware compositional network for person re-identification," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2018.

[44] C. Song, H. Yan, W. Ouyang, and W. Liang, "Mask-guided contrastive attention model for person reidentification," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2018.

[45] R. Zhao, W. Ouyang, and X. Wang, "Person reidentification by saliency learning," IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, pp. 356–370, 2016.

[46] C. Liu, X. Chang, and Y. D. Shen, "Unity style transfer for person re-identification," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2020