International Journal of Computational Science and Engineering Research
ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/
Regular Issue , Vol. 3, Issue. 1 (January – March) , 2026 , Pages: 9 - 16
Received: 20 March 2025 ; *Accepted:* 27 December 2025 ; *Published:* 08 January 2026
https://doi.org/10.63328/IJCSER-V3RI1P2
https://www.youtube.com/watch?v=ZElja1immjA
https://archive.org/details/jsp-ijcser-v3ri1p2
Original Paper : Applied Research – Compare and Contrast Research Paper

A Multimodal Deep Learning Framework For Robust Person Re-Identification

Mesfin Dagne

Department of CSE, Adama Science and Technology University, Adama, Ethiopia, mesfindagne@gmail.com

** Corresponding Author: Mesfin Dagne ; ramyathenepalli@gmail.com*

Abstract: Identifying-matching the same individuals has become a complex thing in strengthening security and it shows seamless living within smart cities. This task involves detecting and matching a person across multiple camera views for surveillance and safety applications. The capability and ability are able to accurately recognize individuals from diversified visual data sources. The new pathways of computer vision accelerated the research progress with deep neural networks to process high quality surveillance data. By analyzing the structural elements of a Re-ID framework, one can differentiate between closed-world and open-world identification scenarios. This work proceeds shows data preprocessing strategies for person re-identification and evaluates deep learning approaches using widely recognized benchmark datasets. Finally, it mainly focus on the vital key role on developing an effective deep learning and that leverages various multimodal feature extraction aimed at performance analysis identification and performance accuracy. We proposed new technique evaluated to performance on human recognition capabilities.

Keywords: Person matching identification, Re-Identification, Hashing, CNN, Autoencoders, Artificially Damaged Frames.

1. Introduction

The These research mainly focused aims on recognize ROI and identify the person individual captured and observed video peer sequences through by live and other mode cameras with overlapping or non-overlapping fields of best possible views. It represents mainly works as a primary personal component in the video surveillance systems, and enabling the live keep tracking of particular people across multiple locations on cameras and reducing identity mismatches with exact performance analysis. Beyond the rigid body individual scope tracking, Re-ID matching techniques have been extended to identify the multi-person tracking scenarios, including monitoring various parameters like athletes in sports videos, where overlapping trajectories, due cause often identity switches on various postures [1].

The main issue we identifies the occlusion, illumination frequent changes, and pose variations depends rigid body moves due to dynamic viewpoints changes with significantly affect Re-ID identity performance analysis. According to public safety concern , it become a growing situation concern, securing crowded environments [2] to identifies on airports and shopping malls demands exact efficient and intelligent video surveillance systems analysis. Recent research identifies on notable research work in computer vision and deep learning to enhanced object pattern recognition capabilities using feature extraction including identifies color histograms, facial details, and jersey numbers, etc. Moreover, deep learning innovations have also strengthened voice-based control in modern electronic device and state-of-the-art methods, further presents the proposed methodology, the datasets, evaluation metrics, and results, and concludes the study with future research directions [3].

2. Literature Survey

Individual matching and identification remains a challenging problem for autonomous tracking systems, with broad applications across multiple domains [4]. The previous work and existing models often perform inadequately in real-world scenarios due to environmental complexity and data limitations. The Recent work proceeds are compiled the evolution of Deep Neural Network (DNN)-based approaches for Person Re-Identification (PReID) since 2014, analyzing diverse architectures and datasets. Out of all available datasets,

Jack Sparrow Publishers © 2026, IJCSER, All Rights Reserved
www.jacksparrowpublishers.com

ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/ , Vol. 3, Issue 1, 2026, https://doi.org/10.63328/IJCSER-V3RI1P2

VIPeR is one of the most widely used and challenging benchmarks. Synthetic data generation has also been proposed to mitigate data scarcity, alongside efforts to reduce model complexity while maintaining rank-1 recognition accuracy.

According to Person Re-Identification (Re-ID) has gained increasing attention of object enhancement to analysis ability to identify rigid individuals from heterogeneous surveillance video footage. However, the privacy concerns risks arises, when surveillance data are outsourced beyond the environment to analysis, to address these issues, Haar Cascade based approaches to analysis have been proposed method to ensuring exact identity theft protection and while maintaining accurate Re-ID performance analysis through effective deep feature extraction techniques [5].

In the various studies to analysis using unsupervised video Re-ID label estimation, were expected outcome to using a Dynamic Graph Matching (DGM) technique for better quality. It contains two-layer cost structure, and multi-layer ROI, iteratively refines various phases label graphs and discriminative performance metrics, and knowingly improving object rigid label accuracy analysis and robustness to noisy removing using training data [6]. This method widely used in constructs per-camera graphs and associates concerned labels through iterative graph and capture images and video peers matching, achieving performance comparable through supervised techniques

Mainly highlighted on these research observations in the [7] Re-ID systems can be analyzed to categorized into various closed-world ROI and open-world ROI settings. Deep learning techniques approaches have targeted to achieved impressive exact accuracy in closed-world data image contexts under controlled environment assumptions; though, open-world ROI scenarios familiarize to other challenges and accurately reflect real-world applications and deployments. Notable research areas are improvements include various ranking optimization techniques, deep feature representation analysis, and metric learning approach too.

The study in [8] addressed dataset limitations by introducing LUperson, a large-scale unlabeled Re-ID dataset. Through unsupervised pre training, robust feature representations were learned, which improved generalization across supervised and few-shot Re-ID tasks. This work opened pathways for integrating temporal video information and developing end-to-end unsupervised models with performance rivaling supervised methods.

Researchers are [9] explored mainly image transformation based approaching to image data retrieval, the Dual-Modal Fusion (DMF) networks connectivity, that unifies multimodal data integrity into a shared image feature extraction space analysis. Pre-training process to evaluates the quality enhancement including feature extraction and robustness, while ROI fine-tuning aligns model performance with data distribution and Data consistency, and benchmark posture evaluations depends on state-of-the-art outcome to expected results, and with identifies potential ROI extensions to expose on posture estimation and data segmentation tasks.

These proposed work using unsupervised learning and enhance the crowded rigid bodies identifies across within the needed multi-vision datasets including large-scale datasets InstagramIB, etc [10]. It gives immense improves image analysis performance in an autoencoding, and it represents the scope to enhance learning the model with transferability and accountability through applied Momentum Contrast (MoCo).

According to MIL technics to analysis perform Re-ID [11], through data captured integrating deep multi-scale convolutional architectures objectivity to improve identity predict accuracy on various postures. And the ROI utilizes the datasets are VIPeR, ETHZ to analysis better performance. Finally these represents a superior robustness against lighting, background noise reduction, and occlusion variations on various positions, with further recommendations for extended evaluations for performance accurate result analysis.

Finally, [12] the proposed DPFL model build with a CNN architecture to capture multi-scale discriminative features for Re-ID. The pyramid structure processes features across different scales, employing scale-specific classification objectives within a unified deep learning framework. Experimental results across multiple Re-ID benchmarks confirmed the model's superior performance over existing methods, emphasizing the significance of cross-scale consensus learning.

3. Proposed Methodology

The two main parts of the proposed method are to enhance quality auto encoder based and extracted features for individual re-identification based on advanced CNNs. It is vital to review CNN and auto-encoders before delving into specifics. One such important and difficult problem in various multi-dimensional processing is person re-identification in video sequences [13]. These days, the significance of identifying persons is more apparent than ever because of the expanding usage of surveillance and control cameras in public areas [14]. Generally, individual re-identification (re-id) involves the ability to locate and track a person across non-overlapping camera fields of view. Since there are many people in the camera's range of vision in busy public areas, it is necessary to identify the target person (query) among them all. Stated differently, individual re-id refers to the process of

Jack Sparrow Publishers © 2026, IJCSER , All Rights Reserved
www.jacksparrowpublishers.com

ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/ , Vol. 3, Issue 1, 2026, https://doi.org/10.63328/IJCSER-V3RI1P2

comparing the individual shown on non-overlapped cameras with the group of prospects shown in the search cam [1].Below figure explains about the person-recognized identification system concepts.

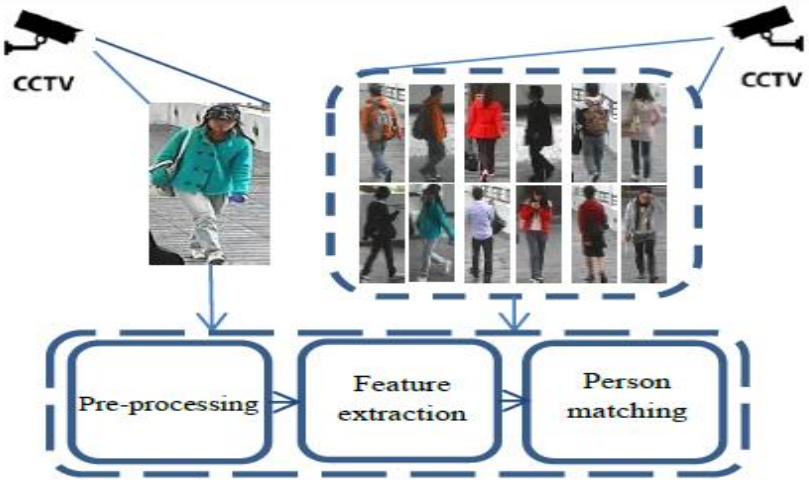

Figure. 1 Overview of individual Re-Identification

Due to the high semantic level of the descriptions, pedestrian attribute recognition faces significant visual variation and spatial shifts. For example, the same clothing type (such as shorts) can look significantly different. Different body positions and camera views are the cause of the significant spatial shifts with regard to the identified pedestrian bounding boxes [25], and it can be difficult to detect or segment finer body parts in surveillance-style recordings. Moreover, occlusion and variations in illumination complicate the issue in realistic environments.

A multilayer neural network known as an auto-encoder uses encoder and decoder layers are used to reconstruct input [28]. Once the primary image is passed through this network, the encoder reduces its dimensions and extracts key features, resulting in a compact vector representation. Then, using supervised learning, the decoder feeds the encoded feature vector in an attempt to recreate the input. Auto-encoders benefit a variety of purposes, including denoising and feature extraction in image processing [26].

The three primary tasks of the suggested approach are person re-identification, feature extraction, and frame repair for denoising and occlusion overcoming. The ensuing subsections provide an explanation of each component's specifics. Due to the fact that person re-id's detected frames are typically noisy and low resolution. During crowded scenes [17], occlusion takes place. This component fixes occluded pixels in order to overcome occlusion and fix the frame. An auto-encoder is used to achieve this, and it is trained using frames that have been damaged. Assuming the input frame is an RGB frame, we define the concept of an Artificial Damaged Frame (ADF).

$$C = Max\left(0, \sum_{i=1}^{K}\sum_{j=1}^{K} p(i,j) \times h(i,j)\right) \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (3)$$

$$f_c = Max\left(0, \sum_{i=1}^{K} w_i \times n_i\right) \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (4)$$

$$f' = \Omega(f + n) \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (1)$$

Here, Ω denotes the operator that models the random damage to the patch pixels, f′ represents the ADF, and n is the noise distribution (which can have different types and varying variances). The primary frame f is the desired output, while the ADF serves as the input frame to the auto-encoder after it has been generated [18]. The targeted result and Artificial damaged frames are achieved by adjusting the weights to reduce the mean square error. Where (.) and $E(.)$ represent the decoder and encoder, respectively, in action. Figure 2 depicts the idea behind this phase.

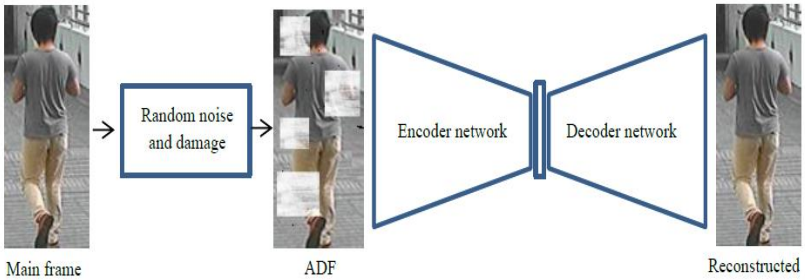

Figure. 2 The idea of Auto Encoder training and ADF for structure reconstruction

Jack Sparrow Publishers © 2026, IJCSER , All Rights Reserved
www.jacksparrowpublishers.com

ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/ , Vol. 3, Issue 1, 2026, https://doi.org/10.63328/IJCSER-V3RI1P2

A CNN is used for feature extraction in the following step. In order to achieve this, the backbone of the network is

VGGNet [27], which has been shown to be a significant CNN in the literature. To adjust the network for our work, a few fully-connected layers are added. Every individual is treated as a distinct category in the classification task that trains the network. Convolution layer weights will be adjusted throughout training until the classification error is as low as possible. Equations (3) and (4) compute the outcome of each The convergence and fully linked layer.

In the network, the filter kernel (h) processes the input frame (p) at each layer. The network weights are adjusted in process of training phase and input values (ni) to associate each input frame with a specific individual. This enables the model to effectively learn and capture the essential features from the frames [27].

We calculate distance between vector and gallery images by using Euclidean distance measure. These distances are ranked, with the closest match being identified. The architecture of this approach is shown in Figure 3.

$$E = \sum_i \sum_j \sqrt{\left(D\left(E\left(f'(i,j)\right) - f(i,j)\right)^2\right)} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (2)$$

Before sending frames to the CNN, the trained auto-encoder is applied on the ADF, as seen in Figure 3. Initially it clears the damaged, obscured to increase frame quality.by this way the similar persons were identified and matched by well-trained CNN model. Initially we divided into two sections and at the core of our model, we use an encoder-decoder structure, which enhances its ability to recognize and correct noisy frames in person re-identification [20]. To reduce noise in training phase , we train CNN to recognize frames by assigning each individual to a unique class label. This allows the model to effectively classify and differentiate between people. Once the frames are classified, their features are extracted and compared, enabling the model to assess the level of similarity or dissimilarity between them [23]. This process is crucial for accurately matching individuals across different frames in person recognizing tasks.

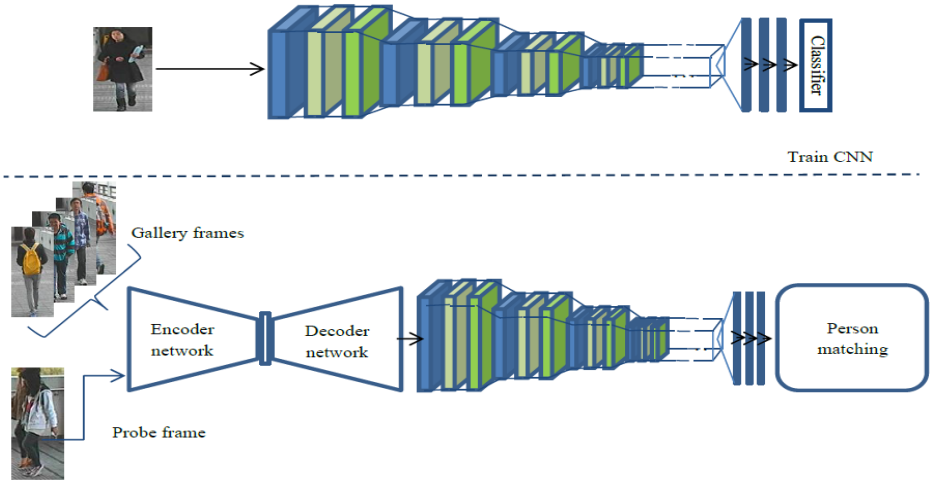

Figure. 3 The proposed model and technique and he block diagram/flow chart is shown as

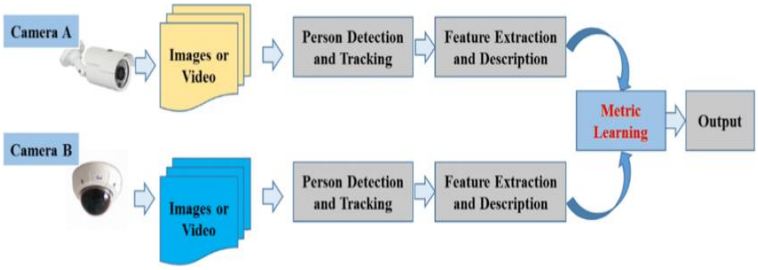

Figure. 4 Block diagram of deep learning CNN

4. Results and Discussions

Two well-known databases, CUHK01 [20] and CUHK03 [21], are utilized to assess the suggested approach. Each of the 971 individuals in CUHK01 has four frames from two non-overlapping cameras. There are more than five frames accessible for each of the 1467 individuals that make up CUHK03, which were taken by five distinct, non-overlapping cameras. Half of the individuals in each dataset are taken into consideration for the initial phase of training, and leftover individuals are considered for the test data, which is separated by gallery set and probe set. Table 1 displays the dataset and set details [21].

Jack Sparrow Publishers © 2026, IJCSER , All Rights Reserved
www.jacksparrowpublishers.com

ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/ , Vol. 3, Issue 1, 2026, https://doi.org/10.63328/IJCSER-V3RI1P2

Ranking metrics are utilized for evaluation purposes. Once the similarity detected, the frames are organized in descending order based on similarity, creating a ranked list. The rank-1 metric is significant as it reflects the probability of accurately identifying the matching frame on the first attempt. If the query frame matches the highest-ranked frame, the rank-1 score for that query is 1.

The average of all rank-1 score of all query frames are denoted with rank-1. Based on frequent analysis on Rank-k to analysis the query object frame appearance within the ROI top k frames on the verified list. These research observation to find out captures data evaluation performance metrics have been extensively used in existing to evaluate Re-ID systems. Figure 4 depicts the CNN and feature extraction training processes, showing the loss functions for each epoch along with the training and validation accuracy. The model's performance on training data is demonstrated, and the convergence of both training and validation loss curves indicates the model's ability to generalize effectively to unseen data [22].

Table 2 explains the numerical results for the suggested approach on CUHK01, Rank-k performance is evaluated for values of k = 1, 5, 10, and 20, and is compared against several advanced models. This approach evaluates how well the model ranks the correct person within the top-k predictions, providing a clear understanding of its accuracy in various situations. By comparing these results with those of leading models, we can gauge the relative effectiveness and strength of our method.. Additionally, these findings are calculated for CUHK03, and Table 3 displays the comparison results [23].

Table 2 shows that the suggested approach outperformed previous approaches, achieving 92.1% rank-1. Moreover, the proposed method exceeds the performance of other techniques in several rankings, which can be credited to the use of a deep CNN for feature extraction and an auto-encoder for frame refinement. Table 3 highlights that the Rank-1 score of the proposed approach is 94.4%, outperforming the best methods by 1.9%.

Additionally, when evaluating performance across different ranks, it is clear that our model performs better on the others CUHK03.The Rank-1 accuracy for different types of fake noise across both databases, comparing results with and without the ADF trained auto-encoder. The model's performance improves as it learns to handle noisy and randomly corrupted inputs. This enhancement is able to repair frames, retrieve robust features for each individual, and its awareness of noise, all of which contribute to the increased accuracy.

Table. 1 Information regarding the datasets and their corresponding splitting strategies

Dataset	Number of persons	Number of frames per person	Number of individuals are selected for training and validation	Number individuals are selected for testing and validation
CUHK01	971	4	485	486
CUHK03	1467	7-10	734	733

Table. 2 Outcomes and comparison using the CUHK01 dataset

Backbone	Market1501		CUHK03		Duke MTMC		MSMT17	
	R1	Map	R1	Map	R1	Map	R1	Map
Res Network (with various methods)	96.9	89.4	-	-	86	74.3	-	-
	88.7	82.5	66.6	64.2	88.8	72.9	-	-
	93.4	86.1	-	-	84.1	73.4	75.5	46.8
	95.8	85.0	-	-	84.6	74.8	77.2	52.3
DL algorithm with Osnet	96.8	95.5	72.3	98.5	88.6	79.5	93.6	63.7

Table 3. Outcomes and comparison using the CUHK03 dataset

Method	Rank 1 (%)	Rank 10 (%)	Rank 20 (%)
TCP[23]	52	90	93
LOMO-XQDA[6]	48	83	89
GOG-Lab[5]	50.6	--	76
SGL-DGDnet [24]	72	92	94
FCDF[25]	47	80	93
PSD [26]	82	96	97
The proposed	93	97	99

Table 4: Findings (%) from large re-ID datasets

Method	Rank 1 (%)	Rank 10 (%)	Rank 20 (%)
SpindleNet[11]	53	92	--
EMD[12]	87	98	99
PSD [26]	92	99.2	99.5
EBb[27]	92	97	99
HGD-ResNet [20]	88	98	99.4
The proposed	95	99.2	99.7

Findings (%) from large re-ID datasets. The DL algorithm outperforms the majority of published methods by a significant margin and achieves best performance in state of the art performance. Notably, the DL algorithm contains 2.2 million parameters, a significant reduction from the best-performing DL algorithm at the moment (Resnet-based approaches). The model was trained from the beginning.

Jack Sparrow Publishers © 2026, IJCSER , All Rights Reserved
www.jacksparrowpublishers.com

ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/ , Vol. 3, Issue 1, 2026, https://doi.org/10.63328/IJCSER-V3RI1P2

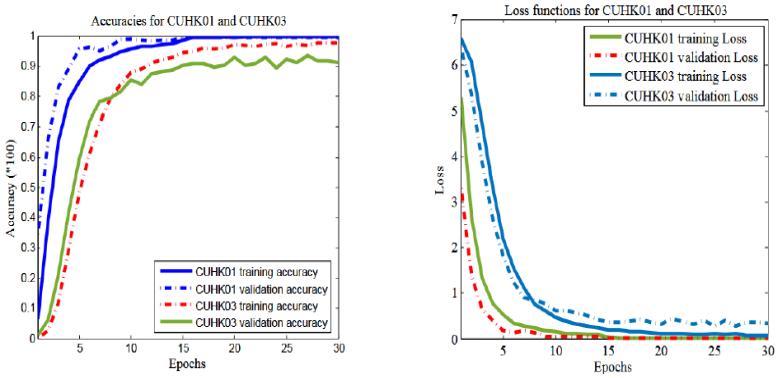

Figure. 5 Loss functions and simulated accuracy for CUHK01 and CUHK03 as well as validation data.

Figure. 6 Three distinct inquiry frames were ranked in the visual outputs of the suggested model. The right re-identified frame is shown with a green marker in each row.

5. Conclusion and Future Scope

In image processing, person re-identification remains a crucial and difficult job. This study used deep CNNs and auto-encoders to suggest a novel approach to human re-identification. To overcome noise and occlusion and enhance frame quality, The auto-encoder was trained using frames with intentional damage. A convolutional neural network-based model was created for feature extraction in order to obtain reliable features for human re-identification. Results from experiments on two well-known datasets, CUHK01 and CUHK03, showed that the suggested approach outperforms the most advanced techniques. With just our pre-training ResNet50 model, we obtain good performance on Market1501, DukeMTMC, MSMT17 including significant features. In order to address the lack of training data, further work to enhance the model may investigate merging generative models with an unsupervised approach and also in the future work, It is important to explore and enhance the proposed feature method to increase the accuracy person's Re-ID. Various deep learning techniques can be adapted to extract additional multimodal features and can also be applied to video databases.

References

[1]. Wu, G., Zhu, X., and Gong, S., 2022. Learning hybrid ranking representation for person re-identification. *Pattern Recognition*, 121, pp.108239. Doi: 10.1016/j.patcog.2021.108239

[2]. Sezavar, A., Farsi, H., and Mohamadzadeh, S., 2022. Multi-Depth Deep Similarity Learning for Person Re-Identification. *Jordan Journal of Electrical Engineering. All rights reserved-Volume*, 8(3), pp. 279-287. Doi: 10.5455/jjee.204-1653115709

[3]. N. Thomas, "Implementating Hydrophonics Farmig System as a sustainable agricultural revolution in the future farming," International Journal of Research and Development in Engineering Sciences, vol. 5, no. 1, p. 26, Feb. 2023, doi: 10.63328/ijrdes-v5ri1p5.

Jack Sparrow Publishers © 2026, IJCSER , All Rights Reserved
www.jacksparrowpublishers.com

ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/ , Vol. 3, Issue 1, 2026, https://doi.org/10.63328/IJCSER-V3RI1P2

[4]. Mang Ye, Jianbing Shen, Gaojie Lin, Tao Xiang, Ling Shao and Steven C. H. Hoi, "Deep Learning for Person Re-identification: A Survey and Outlook", IEEE Transactions On Pattern Analysis and Machine Intelligence, 2021.

[5]. Tripathi, M., 2021. Facial image denoising using AutoEncoder and UNET. *Heritage and Sustainable Development*, 3(2), pp.89-96. Doi: 10.37868/hsd.v3i2.71

[6]. Shang, Z., Sun, L., Xia, Y., and Zhang, W., 2021. Vibration-based damage detection for bridges by deep convolutional denoising autoencoder. *Structural Health Monitoring*, 20(4), pp. 1880-1903. Doi: 10.1177/1475921720942836.

[7]. P. Tarollu and B. R. Y, "Intelligent Chatbot ticketing system using RNN and CNN," *International Journal of Research and Development in Engineering Sciences*, vol. 7, no. 1, p. 30, Feb. 2025, doi: 10.63328/ijrdes-v7ri1p6.

[8]. K Krishna Reddy , " Adaptive Memory-Augmented Agentic Systems for Long-Term Context Preservation in Large Language Model Environments ", International Journal of Computer Science, Engineering and Artificial Intelligence , vol. 3, no. 2, p. 30-37, May 2026, DOI: https://doi.org/10.63328/IJCSEAI-V3RI2P5

[9]. Kaiming He Haoqi Fan Yuxin Wu Saining Xie Ross Girshick, "Momentum Contrast for Unsupervised Visual Representation Learning", 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).

[10]. M Anjan Kumar , D Ramana Reddy , Rakhee , " Intelligent Paediatric Epilepsy Detection using Quantum Computing ", International Journal of Computer Science, Engineering and Artificial Intelligence , vol. 3, no. 2, p. 6, April 2024, doi:https://doi.org/10.63328/IJCSEAI-V3RI3P1

[11]. Saravanan K1 , Jeevitha R2 , Keerthana R3 , Yogapriya K, "Person Re-Identification Using Deep Learning Approach", International Journal of Innovative Research in Science, Engineering and Technology, Vol. 9, Issue 3, March 2020.

[12]. Saad, O. M., and Chen, Y., 2020. Deep denoising autoencoder for seismic random noise attenuation. *Geophysics*, 85(4), pp. 367-376. Doi: 10.1190/geo2019-0468.1

[13]. M. Dharani , Ramesh Samudrala , Sk. Rukshana Begum , " Automated Pothole Detection And Severity Mapping Using Geographic Information System", International Journal of Computer Science, Engineering and Artificial Intelligence , Vol. 3, Issue. 1 (January – March), Pages: 27 – 33 , 2026 . DOI: https://doi.org/10.63328/IJCSEAI-V3RI1P5

[14]. Sezavar, A., Farsi, H., and Mohamadzadeh, S., 2019. Content-based image retrieval by combining convolutional neural networks and sparse representation. *Multimedia Tools and Applications*, 78(15), pp. 20895-20912. Doi: 10.1007/s11042-019-7321-1

[15]. S. B. Bokka, "Enhancing brain stroke prediction using machine learning for early intervention," *International Journal of Research and Development in Engineering Sciences*, vol. 7, no. 1, p. 46, Feb. 2025, doi: 10.63328/ijrdes-v7ri1p8.

[16]. Domonkos Varga∗†, Tamas Szir ´ anyi MTA SZTAKI, "Person Re-identification based on Deep Multi-instance Learning", 2017 25th European Signal Processing Conference (EUSIPCO).

[17]. Mang Ye1 , Andy J Ma1 , Liang Zheng2 , Jiawei Li1 , Pong C Yuen, "Dynamic Label Graph Matching for Unsupervised Video Re-Identification",Sept 2017

[18]. Yanbei Chen, Xiatian Zhu, Shaogang Gong, "Person Re-Identification by Deep Learning Multi-Scale Representations", 2017 IEEE International Conference on Computer Vision Workshops.

[19]. Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., and Lew, M. S., 2016. Deep learning for visual understanding: A review. *Neurocomputing*, 187, pp. 27-48.Doi: 10.1016/j.neucom.2015.09.116.

[20]. Chen, Y. C., Zheng, W. S., Lai, J. H., and Yuen, P. C., 2016. An asymmetric distance model for cross-view feature mapping in person reidentification. *IEEE Transactions on Circuits and Systems for Video Technology*, 27(8), pp. 1661-1675. Doi: 10.1109/TCSVT.2016.2515309

[21]. M. Dharani , Nasreen Sultana Quadri , T. Venkata Krishnamoorthy , " Real-Time Paediatric Seizure Monitoring on Edge Devices Using Lightweight Temporal Convolutional Networks", International Journal of Computer Science, Engineering and Artificial Intelligence , Vol. 3, Issue. 1 (January – March), Pages: 8 – 12, 2026 . DOI: https://doi.org/10.63328/IJCSEAI-V3RI1P2

[22]. Simonyan, K., and Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. *ArXiv Preprint ArXiv:1409.1556*. Doi: 10.48550/arXiv.1409.1556.

[23]. A. Boban, A. C. Kurian, C. E. S. S, and S. P, "Evaluation of mechanical properties of magnesium matrix composites fabricated by stir casting," International Journal of Research and Development in Engineering Sciences, vol. 5, no. 3, p. 26, Jun. 2023, doi: 10.63328/ijrdes-v5ri3p5.

[24]. Li, W., Zhao, R., and Wang, X., 2013. Human reidentification with transferred metric learning. *Asian Conference on Computer Vision*. Springer, Berlin, Heidelberg. pp. 31-44. Doi: 10.1007/978-3-642-37331-2_3

[25]. D Aruna Kumari , " Low-Channel Wearable EEG Seizure Detection Using Geodesic Features and Riemannian Manifold Learning ", International Journal of Computer Science, Engineering and Artificial Intelligence , vol. 3, no. 2, p. 15-22, April 2026, doi: https://doi.org/10.63328/IJCSEAI-V3RI2P3

[26]. Vincent, P., Larochelle, H., Bengio, Y., and

Jack Sparrow Publishers © 2026, IJCSER , All Rights Reserved
www.jacksparrowpublishers.com

ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/ , Vol. 3, Issue 1, 2026, https://doi.org/10.63328/IJCSER-V3RI1P2

Manzagol, P. A., 2008. Extracting and composing robust features with denoising autoencoders. *Proceedings of the 25th International Conference on Machine Learning*, pp. 1096-1103. Doi: 10.1145/1390156.1390294.

[27]. A. Noyal, A. C. Kurian, C. E. S. S, and S. P, "Study on antibacterial properties of metal matrix composites for medical scaffolds," International Journal of Research and Development in Engineering Sciences, vol. 5, no. 2, p. 6, Mar. 2023, doi: 10.63328/ijrdes-v5ri2p2.

[28]. R. Kolikipogu, A. Devarakonda, and S. Palamakula, "A Privacy-Preserving universal multimodal framework for Real-Time Any-to-Any transformation," *International Journal of Research and Development in Engineering Sciences*, vol. 8, no. 2, p. 21, Mar. 2026, doi: 10.63328/ijrdes-v8ri2p3.

How to Cite

Mesfin Dagne , " A Multimodal Deep Learning Framework for Robust Person Re-Identification " , International Journal of Computational Science and Engineering Research , Vol. 3, Issue. 1 , January , 2026 , Pages: 9 – 16.
CrossRef DOI: https://doi.org/10.63328/IJCSER-V3RI1P2

Declaration

Conflicts of Interest: The authors declare no conflict of interest.

Author Contribution: All authors wrote the main manuscript text and also consent to the submission.

Ethical approval: Not applicable.

Consent to Participate: All authors consent to participate.

Funding: Not applicable, and No funding was received

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Personal Statement: We declare with our best of knowledge that this research work is purely Original Work and No third party material used in this article drafting. If any such kind material found in further online publication, we are responsible only for any judicial and copyright issues.

Generative AI Statement: The authors declare that generative AI was not used in the preparation or creation of this manuscript. The alternative text (alt text) accompanying the figures in this article was generated by Frontiers with the assistance of artificial intelligence. Reasonable efforts have been made to ensure the accuracy and appropriateness of the generated alt text, including review by the authors wherever possible. If you identify any inaccuracies or issues, please contact the editorial team.

Publisher's Note: The statements, opinions, and claims presented in this article are exclusively those of the authors and do not necessarily reflect the views of their affiliated institutions, the publisher, editors, or reviewers. Any products discussed or evaluated, as well as any claims made by their manufacturers, are not warranted, approved, or endorsed by the publisher.

Acknowledgements

We thank everyone who inspired our work

Jack Sparrow Publishers © 2026, IJCSER , All Rights Reserved
www.jacksparrowpublishers.com