Received: 12 March 2024 | Revised: 15 April 2024 | Accepted: 2 May 2024 | Published online: 7 May 2024

RESEARCH ARTICLE

Journal of Computational and Cognitive Engineering
2026, Vol. 5(1) 176–186
DOI: 10.47852/bonviewJCCE42022872

BON VIEW PUBLISHING

Generative Adversarial-Based Ubiquitous Data Integration Model for Human Re-identification

Mohammad Shabaz[1,*] and Mukesh Soni[2]

[1]*Computer Science Engineering, Model Institute of Engineering and Technology, India*

[2]*Dr. D. Y. Patil School of Science & Technology, Dr. D. Y. Patil Vidyapeeth, India*

Abstract: Person re-identification is a crucial field in computer vision with significant challenges, including illumination, occlusion, pose variation, and feature variation in complex backgrounds. Traditional methods have limitations in addressing these issues. However, thanks to the advent of deep learning, new avenues have opened up, making it a hot topic in research. Generative adversarial networks (GANs) have gained immense popularity in recent years for their ability to overcome these challenges. With their potential, GANs offer a promising solution to help revolutionize the field of person re-identification. This paper introduces two innovative approaches that significantly improve the network's performance and efficiency. Firstly, a dual-stream network is proposed that integrates GANs to learn features from multi-modal images. This approach is a breakthrough in network structure. Secondly, an angle-based heterogeneous center triplet loss is presented which enhances clustering effects in the feature space and reduces the effect of outliers on paired distances, making it an excellent loss function. These approaches collectively provide a new angle to limit the feature space's positive and negative tests, resulting in a highly efficient and effective network. This paper proposes a technique for learning discriminative pedestrian features while minimizing cross-modal differences. Our model optimizes results for pedestrian re-identification datasets like SYSU-MM01 and RegDB, using a vast collection of visible light and infrared photos of pedestrians taken from multiple cameras. Our ablation experiments show that the dual-stream network in this paper achieves a Rank-1 accuracy of 57.27% using the SYSU dataset in all-search mode, surpassing many existing works. This demonstrates that the dual-pooling fused features obtained from feature fusion in our technique are richer and more discriminative.

Keywords: ubiquitous data, human re-identification, generative adversarial networks, cross-modal pedestrian re-identification algorithms

1. Introduction

Generative adversarial networks (GANs) have revolutionized the imaging field due to their ability to simulate arbitrary data distributions with remarkable accuracy. GANs can generate images, music, and videos that are indistinguishable from the real world. The GAN is composed of a generator and discriminator, which work in tandem to create images that are evaluated for authenticity. Through the training process, both the generator and discriminator can improve their performance, resulting in more realistic and convincing generated images that even the discriminator can be fooled by. GANs have opened up countless possibilities for the creative industry and have proven to be a valuable tool for bringing imagination to life.

It seems like GANs can be improved by adding an extra input y to both the G and D networks. This allows for more information to be used in the model beyond just random variables. One interesting system that uses this approach is the feature distilling GAN. It incorporates multiple discriminators that focus on human poses and identities, along with a same-pose loss integration that aims to produce similar images of the same person.

As time goes on and science and technology continue to advance, urban surveillance networks have become increasingly sophisticated. The use of surveillance footage for pedestrian retrieval has become a common technique employed by law enforcement agencies in apprehending criminals. Pedestrian re-identification algorithms [1] utilize computer vision and machine learning techniques to search for specific individuals across different cameras or time frames within a surveillance scene [2]. This has garnered widespread attention in the academic community in recent years, leading to numerous research outcomes.

Currently, most pedestrian re-identification algorithms focus on using deep learning methods to retrieve visible light images and have achieved satisfactory experimental results [3]. In a recent research [4], a new algorithm has been proposed for visible light pedestrian re-identification that leverages perspective-aware loss. This algorithm offers a highly accurate way to identify pedestrians even under low-light conditions. As the day turns into night, most surveillance cameras switch to NIR mode to ensure round-the-clock monitoring. Infrared cameras generate thermal images using the target's infrared radiation signals, while RGB cameras divide visible light into different wavelengths like red, green, and blue. However, the differences in the imaging principles of these two types of cameras make it difficult to match infrared images with RGB images [5]. The proposed algorithm has been tested on the Market1501 dataset and has achieved an impressive average precision of 95.43% [6].

Therefore, cross-modal pedestrian re-identification algorithms not only need to address the common intra-modal difference problem in traditional pedestrian re-identification algorithms but

***Corresponding author:** Mohammad Shabaz, Computer Science Engineering, Model Institute of Engineering and Technology, India. Email: shabaz.cse@mietjammu.in

also need to resolve the infrared-visible light inter-modal discrepancy. To address the modal difference problem, the academic community has proposed two main approaches: one involves using dual-stream networks to learn modal-specific and shared characteristics and the other involves using GANs to generate paired images [7] or multi-spectral images [8] to eliminate modal differences.

Furthermore, to train the network to decrease intra-class distance and enhance inter-class distance, the majority of pedestrian re-identification algorithms employ cross-entropy loss and triplet loss simultaneously [9]. However, traditional cross-modal triplet loss is a strong constraint, and in the presence of outliers (such as incorrect labels) in images, it may disrupt other well-trained paired distances [10]. Moreover, it cannot constrain the angles between negative as well as positive test pairs in the feature space, leading to the model's inability to divide different pedestrian labels based on angles during the testing phase [11].

The principle of the heterogeneous center triplet loss function is an innovative approach to bring closer the sample centers of the same identity across different modalities while pushing away the sample centers of different identities. This approach improves cross-modality person re-identification technology, which plays a crucial role in ensuring public safety by identifying targets in surveillance cameras of different modalities. By solving the problem of person Re-ID under low-light or night conditions, cross-modality person re-identification enhances the accuracy and efficiency of surveillance systems. The deep neural network framework combining the heterogeneous center loss and the non-local mechanism is a powerful tool that delivers exceptional results in a wide range of applications.

In today's ever-evolving digital landscape, the fields of digital forensics, machine learning, and related domains are of paramount importance. The research environment in this area is incredibly diverse and multifaceted, with studies such as Khan et al. [12] and Xiong et al. [13] focusing on ambient social media forensics and Golovianko et al. [14] delving into the ramifications of Industry 4.0 and 5.0. Meanwhile, Lombardi and Vannuccini [15] explores emerging patterns in the cyber-physical cosmos, and Fan et al. [16] makes significant contributions to medical picture analysis. These studies are crucial in helping us better understand and navigate the complex digital world we live in. This article provides a study of deep learning models [17], discusses the protection of industrial cyber-physical systems [18], suggests a framework for threat modeling in mobile communication networks [19], and uses AI to improve tumor management [20]. Using small samples, Mahalakshmi et al. [21] covers machine learning, Yang et al. [22] investigates transfer learning for recognizing human activities, and Meng et al. [23] talks about intelligent recovery decision-making. All these books and articles built up to this point provide the ASMF paper with a solid foundation [24].

GANs have revolutionized the field of artificial intelligence by providing a powerful tool for unsupervised learning. With two neural networks – a discriminator and a generator – GANs use adversarial training to create artificial data that is indistinguishable from actual data. The versatility of GANs is evident from their widespread use in image synthesis, style transfer, and text-to-image synthesis. GANs can create realistic samples by engaging in a competitive interplay between the generator and discriminator until the generator becomes proficient at generating samples that fool the discriminator approximately half the time. GANs are truly remarkable in their ability to generate such realistic data without any human intervention. It is no wonder that GANs are quickly becoming an indispensable tool in many industries.

The use of GANs in a cross-modality model for re-identification between RGB and infrared images is a cutting-edge technique that tackles the challenge of a lack of discriminative information and difficulty in learning a robust metric. The authors of the study have developed a generative adversarial training-based discriminator to learn discriminative feature representation from different modalities, which helps in handling the lack of discriminative information. Additionally, the integration of identification loss and cross-modality triplet loss makes it possible to minimize inter-class ambiguity while maximizing cross-modality similarity among instances, which is essential for large-scale cross-modality metric learning. This technique can be a great help in a wide range of applications that require cross-modality re-identification, making it a valuable contribution to the field of computer vision.

The paper suggests cross-modal pedestrian re-identification for nighttime surveillance. A ubiquitous data GAN accounts for small samples and modalities. Dual-stream networks minimize inconsistencies and angle heterogeneous center triplet losses boost clustering. Experimental findings show that SYSU-MM01 outperforms current Rank-1 and mAP methods. The method improves multi-mode pedestrian re-identification.

Based on the above ideas, this paper innovates from two perspectives: (1) in terms of network structure, we propose a dual-stream network that integrates GANs to learn features from multi-modal images; (2) in terms of loss functions, we present an innovative angle-based heterogeneous center triplet loss that improves network performance by reducing the effect of outliers on paired distances and by using angles to limit the feature space's positive as well as negative tests, which enhances clustering effects in the feature space and improves the network's overall efficiency.

This paper presents a solution to visible-to-infrared pedestrian re-identification algorithms in cross-modal settings. The method utilizes the MatchGAN module and a ResNet50 dual-stream network to extract picture features. The paper proposes using the angle heterogeneous center triplet loss and cross-entropy loss for joint training to reduce the network's sensitivity to outliers and improve clustering. These approaches provide a new angle, resulting in an efficient and effective network for learning discriminative pedestrian features. The proposed technique offers a novel way to minimize cross-modal differences and improve cross-modal image recognition.

Section 2 presents a new pedestrian re-identification method employing GANs, Ubiquitous Data, dual-stream networks, and angle heterogeneous center triplet losses. Section 3 shows that it outperforms contemporary Rank-1 and mAP algorithms on SYSU-MM01. Section 4 concludes with a summary and suggestions for additional research.

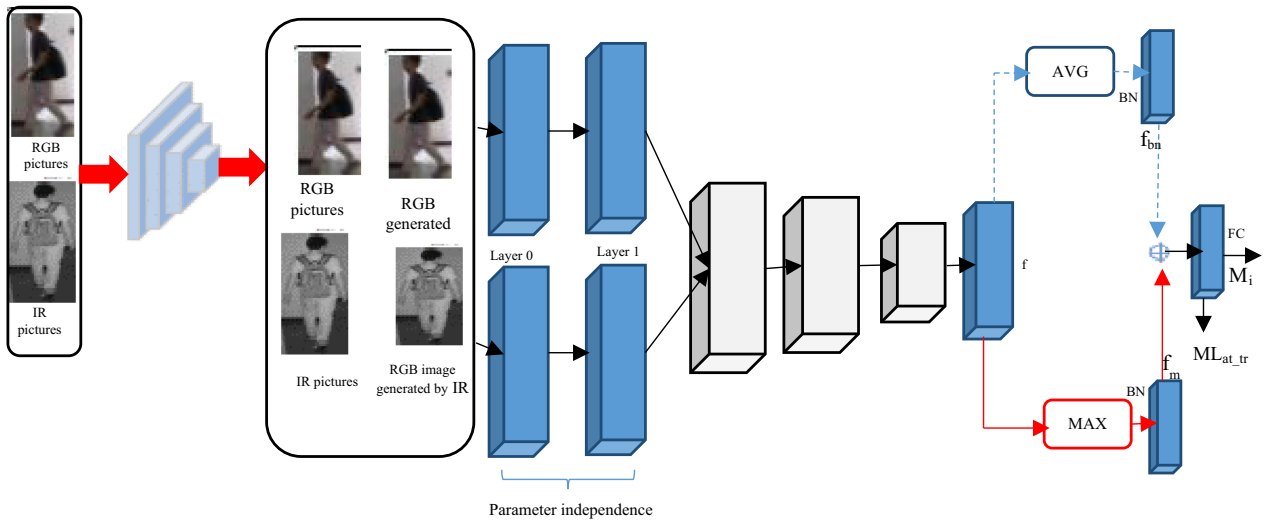
2. Research Methodology

Due to the significant impact of modal differences on the performance of cross-modal pedestrian re-identification algorithms, this paper focuses on learning discriminative pedestrian features while minimizing cross-modal differences. The extracted features are then utilized to optimize the model. This section will first introduce the network framework proposed in this paper and then focus on introducing the angle-based heterogeneous center triplet loss proposed within the scope of this study.

2.1. Network framework

This research presents a dual-stream network that incorporates a GAN, as shown in Figure 1 to learn features from multi-modal images. The network has two primary modules: image creation using the Match GAN to generate cross-modal matching images and the feature extraction module composed of a dual-stream

Figure 1
The framework of the dual-stream network integrating generative adversarial network (GAN)



network. A method for testing parameter independence that relies on a different set of assumptions are more robust for high dimensional samples. Specifically, the method demonstrated that with a viable approximation to a conditional distribution, one can derive conditional independence tests that are approximately valid in finite samples and have non-trivial power which is expressed as parameter independence in Figure 1.

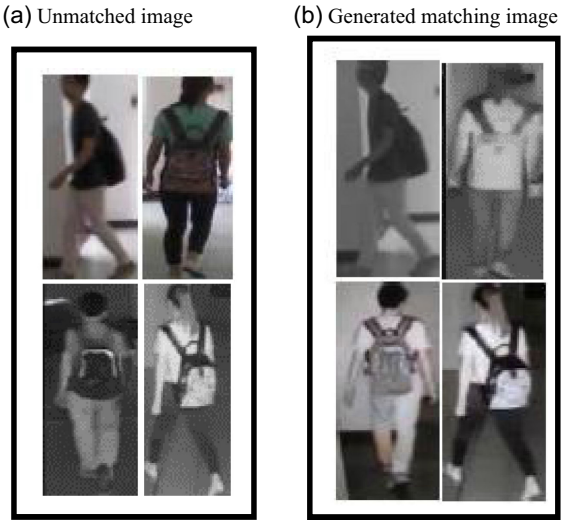
The existing few-shot learning methods can be broadly categorized into three types – metric-based, model-based, and optimization-based. One such method is MatchingGAN which uses a matching procedure to blend the features of conditional images through interpolation coefficients. To increase the number of available samples, data augmentation is used. However, early techniques such as shifts, rotations, or shears have limited diversity. Deep generative models, on the other hand, provide a wider range of diverse samples that can be used for data augmentation, including both feature and image augmentation. In adversarial learning, the matching generator and matching discriminator are optimized using different loss terms in an alternating fashion. The training procedure is monitored using validation seen categories, but they are not used to update model parameters. The discriminator is composed of one convolutional layer and five blocks with increasing channel numbers.

2.2. MatchGAN

Modal differences consist of intra-modal differences and inter-modal differences. Intra-modal differences refer to situations where two images of the same human re-identification have significant differences due to different viewpoints, poses, etc. On the other hand, inter-modal differences arise from the different imaging principles of infrared and RGB cameras, leading to significant ubiquitous data distribution differences between the two modal images of the same human re-identification. MatchGAN aims to reduce intra-modal differences in the same human re-identification images by generating images with similar poses, thereby reducing modal differences. The network model can focus on inter-modal learning and optimization.

As shown in Figure 2(a), each column of images represents pictures collected of the same human re-identification under different lighting and shooting angles. There is a significant intra-modal difference between unmatched cross-modal images (such as poses, etc.). To address this issue, the MatchGAN module is used to correct unmatched images. By generating matching images, as shown in Figure 2(b), the generated cross-modal images have the same-pose information and dressing style as the original images, thereby reducing intra-modal differences between cross-modal images while augmenting the dataset.

Figure 2
Generating matching images using MatchGAN

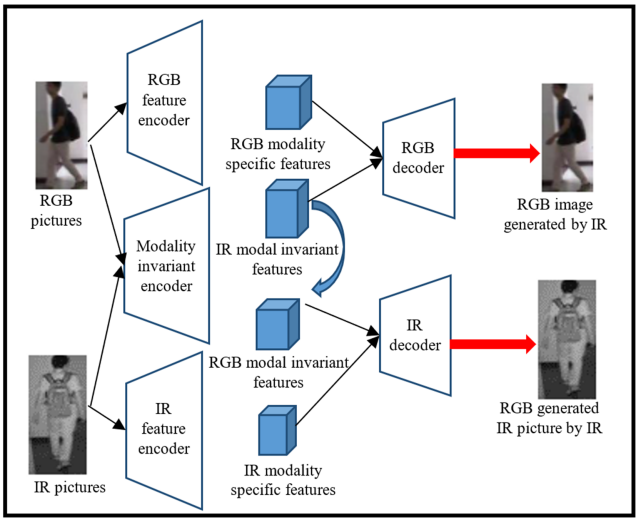


2.2.1. Matching image generation process

The power of the matching generator is truly remarkable. With just a few conditional images and random vectors from the same category, it can generate entirely new images that match the fused features. The matching discriminator is also a game-changer, taking the conventional GAN discriminator to new heights by matching the generated image features with the fused feature of the conditional images. The beauty of feature matching is that it

keeps the generator from overtraining on the current discriminator. Rather than maximizing the discriminator output, the objective requires the generator to generate data that matches the statistics of the real data. By training the generator to match the expected value of the features on an intermediate layer of the discriminator, we can achieve truly remarkable results. The MatchGAN network structure used in this paper is shown in Figure 3. Three encoders are used to learn features that are specific to a modality as well as those that are modality invariant of different modality images of the same human re-identification.

Figure 3
MatchGAN structure diagram



As shown in Equation (1), the RGB image Y_{rgb} is passed through the RGB feature encoder F^s_{rgb} to learn RGB modality-specific features g^s_{rgb} (such as pedestrian clothing color and texture information) and through the modality-invariant encoder F^i to learn modality-invariant features g^i_{rgb} (such as pedestrian pose and body shape). Similarly, as shown in Equation (2), the IR (InfraRed) image Y_{ir} is passed through the IR feature encoder F^s_{ir} and the modality-invariant encoder F^i to obtain IR modality-specific features g^s_{ir} and modality-invariant features g^s_{ir}, respectively.

$$g^s_{rgb} = F^s_{rgb}(Y_{rgb}), g^i_{rgb} = F^i(Y_{rgb}) \tag{1}$$

$$g^s_{ir} = F^s_{ir}(Y_{ir}), g^i_{ir} = F^i(Y_{ir}) \tag{2}$$

While maintaining the modality-specific features g^s_{rgb} unchanged, decoders E_{rgb} and E_{ir} generate matching images by exchanging the modality-invariant features of different modal images (i.e., g^i_{rgb} and g^i_{ir}), as shown in Equation (3). This ensures that the generated RGB image $Y_{\text{fake\_rgb}}$ does not alter the modality-invariant information of the original IR image, such as pedestrian pose and body shape while possessing modality-specific features like clothing color and texture similar to the original RGB image.

$$Y_{\text{fake\_rgb}} = E_{\text{rgb}}\left(f^i_{\text{ir}}, f^s_{\text{rgb}}\right), Y_{\text{fake\_ir}} = E_{ir}\left(f^i_{rgb}, f^s_{ir}\right) \tag{3}$$

2.2.2. Rebuilding error

To enhance produced images, three rebuilding errors are employed. To ensure that the separated features can reconstruct their original images, $M^{\text{same}}_{\text{recon}}$ is utilized for optimizing same-modal image reconstruction, as shown in Equation (4), where ||·|| denotes the M1 distance. This loss function monitors the GAN, preventing overfitting [25].

$$M^{\text{same}}_{\text{recon}} = \|Y_{\text{rgb}} - E_{\text{rgb}}\left(f^{\text{i}}_{\text{rgb}}, f^{\text{s}}_{\text{rgb}}\right)\|_1 + \|Y_{\text{ir}} - E_{\text{ir}}\left(f^{\text{i}}_{\text{ir}}, f^{\text{s}}_{\text{ir}}\right)\|_1 \tag{4}$$

However, Equation (4) does not supervise the generation of cross-modal matching images. To ensure the correlation between the modality-invariant feature f^i_{ir} of the generated RGB image $Y_{\text{fake\_rgb}}$ and the modality-specific feature f^i_{ir} of the original RGB image, cycle loss [26] is employed to constrain the generated images, as shown in Equation (5).

$$M_{\text{cycle}} = \|Y_{rgb} - \widetilde{Y}_{rgb}\|_1 + \|Y_{ir} - \widetilde{Y}_{ir}\|_1 \tag{5}$$

M_{cycle} ensures that the generated images can be transformed back into the original images, providing further constraints on the generated images and addressing the mismatch issue.

In the equations, $\widetilde{Y}_{rgb} = E_{rgb}\left(\widetilde{f^i_{ir}}, \widetilde{f^s_{rgb}}\right)$ and similarly $\widetilde{Y}_{ir} = E_{ir}\left(\widetilde{f_{rgb}}, \widetilde{f^s_{ir}}\right)$; $\widetilde{f^s_{rgb}}, \widetilde{f^i_{rgb}}$ represent modality-specific and modality-invariant characteristics extracted from $Y_{\text{fake\_rgb}}$, respectively, and $\widetilde{f^s_{rgb}}$ represent the same extracted from, $\widetilde{f^i_{rgb}}$.

$$\begin{aligned} M^{rgb}_{GAN} &= F\left[lg\left(Eis_{rgb}\left(Y_{rgb}\right)\right) + lg\left(1 - Eis_{\text{rgb}}\left(Y_{\text{fake\_rgb}}\right)\right)\right] \\ M^{ir}_{\text{GAN}} &= F[lg(Eis_{\text{ir}}\left(Y_{\text{ir}}\right)) + lg(1 - Eis_{\text{ir}}\left(Y_{\text{fake\_}}\right))] \end{aligned} \tag{6}$$

The final rebuilding error M is shown in Equation (7). Building upon the preceding content, this paper, based on human re-identification experience and experimental results, sets λ_{cycle} and λ_{GAN} to 1, and the cycle is set to 10.

$$M = \lambda_{\text{cycle}} M_{\text{cycle}} + \lambda_{\text{GAN}} M_{\text{GAN}} + \lambda_{\text{recon}} M_{\text{recon}} \tag{7}$$

2.3. Dual-stream network

To extract features robust to modal variations, a dual-stream network structure is employed [27], as illustrated in Figure 1. Since cross-modal differences mainly exist in shallow features [28], and following prior experience [29], this paper treats the shallow convolution module (layer 0) and the first residual convolution block (layer 1) of the ResNet50 [30] network independently. These are utilized as modality-specific feature extraction modules, learning distinct shallow features for different modalities. The remaining three residual convolution blocks (layer 2, layer 3, and layer 4) share

parameters, learning modality-shared feature representations in the same feature space.

As the deeply extracted features f often have smaller dimensions and contain effective high-level semantic information, average pooling is commonly employed to aggregate global information of the feature map to preserve the integrity of discriminative information. In contrast, max pooling, as shown in Equation (8), retains the most salient and recognizable information in the feature map.

$$T_{ij} = \mathrm{MAX}_{i=1,j=1}\left(f_{ij}\right) + d_2 \tag{8}$$

Here, T represents the subsampled feature map, the stride is denoted by c, the pooling domain is a × c × c matrix, and d_2 is the bias. Taking advantage of both characteristics, as shown in Figure 1, the extracted feature f is subjected to average pooling along the blue dashed line, followed by batch normalization [31], which accelerates network convergence while preventing overfitting. This yields the feature fbn. Subsequently, along the red solid line, the original feature f is passed through a max pooling layer to obtain the feature fm. The features fbn and fm are then fusional using point-wise addition, followed by batch normalization to obtain the final dual-pooling fusion feature f_i. This preserves the integrity of discriminative information in the feature map while highlighting the most discriminative parts.

To enable the model to have effective feature learning capabilities, current pedestrian re-identification algorithms commonly adopt a joint training approach, utilizing the feature f_i to compute the angle-based heterogeneous center triplet loss $M_{\mathrm{AC\_tri}}$ as a metric loss to constrain intra-class differences. This is combined with cross-entropy loss as a classification loss to constrain inter-class differences, jointly training the network.

2.3.1. Angle-based heterogeneous center triplet loss

The fundamental idea of traditional triplet loss [31] is that the distance between positive sample pairs plus a pre-defined margin is less than the distance between the positive as well as negative test pairs. Building on this idea, to prevent the model from getting stuck in local optima and enhance its generalization ability, researchers often employ triplet losses based on hard sample mining (i.e., hard triplet loss), as shown in Equation (9).

$$M_{\mathrm{Hard\_ri}} = \sum_{i=1}^{P}\sum_{a=1}^{K}\left[\rho + \max_{p=1,2,\cdots,K} E\left(y_a^i, y_p^i\right) - \min_{\substack{n=1,2,\cdots,K \\ j=1,2,\cdots,P \\ i\neq j}} E\left(y_a^i, y_n^j\right)\right] \tag{9}$$

Here, number of pedestrian categories: P, images: K, extracted for each pedestrian, y_a^i is the anchor, y_p^i and y_n^j represent the positive as well as negative tests, respectively, ρ is the set margin, and E(·) calculates the Euclidean distance.

However, hard triplet loss requires calculating the anchor with all other images in the batch, which incurs a high computational cost. Additionally, it is a strong constraint, and due to dataset outliers (e.g., incorrect labels), it may disrupt other well-trained paired distances. Addressing this issue, this paper improves upon traditional center loss and proposes heterogeneous center triplet loss from hard triplet loss $M_{\mathrm{center\_tri}}$. Traditional center loss [31] is shown in Equation (10), reducing the space between the center of each class di and each pedestrian feature f_i to make the classes more compact within themselves. Feature f is associated with pedestrian category i, B stands for the number of photos in each batch, and di is the class center.

$$M_{\mathrm{center}} = \frac{1}{2}\sum_{i=1}^{c}\left\|f_i - d_i\right\|_2^2 \tag{10}$$

In the task of cross-modal pedestrian re-identification algorithms, during the training phase, P pedestrians are randomly sampled at each iteration, and K images are extracted for each modality, resulting in a total of 2P × K images. For each batch of ubiquitous data, the RGB modality class centers D_{v}^i and the IR modality class centers D_{t}^i for P pedestrians can be obtained using Equation (11), where D_{t}^i represents the jth RGB image of the ith pedestrian, and similarly, t_j^i represents the jth IR image of the ith pedestrian.

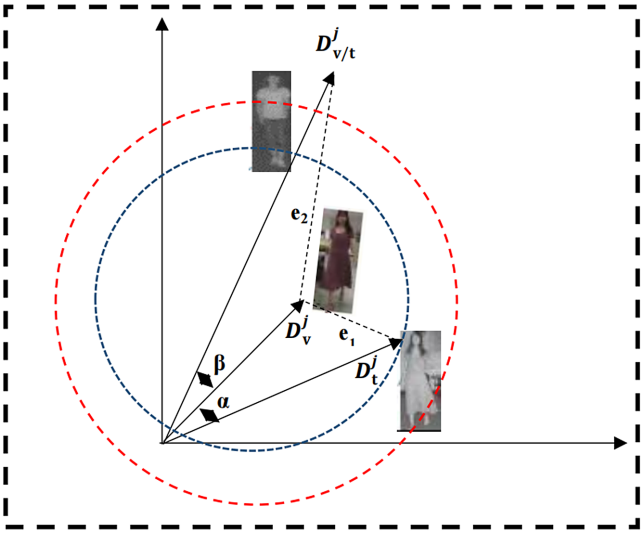
$$D_{\mathrm{v}}^i = \frac{1}{K}\sum_{j=1}^{K} v_j^i,\ D_{\mathrm{t}}^i = \frac{1}{K}\sum_{j=1}^{K} t_j^i \tag{11}$$

Unlike traditional center loss, the heterogeneous center triplet loss utilizes positive sample pairs composed of anchor points and different modality class centers. It then mines the hardest negative samples from all other class centers, as shown in Equation (12). By comparing between class centers, this approach reduces the computational complexity of the model while mitigating the impact of outliers on model performance.

$$\begin{aligned} M_{\mathrm{center\_tri}} = &\sum_{i=1}^{P}\left[\rho + E(D_{\mathrm{v}}^i, D_{\mathrm{t}}^i) - \min_{j\neq i} E\left(D_{\mathrm{v}}^i, D_{\mathrm{v/t}}^j\right)\right]_+ \\ &+ \sum_{i=1}^{P}\left[\rho + E(D_{\mathrm{t}}^i, D_{\mathrm{v}}^i) - \min_{j\neq i} E\left(D_{\mathrm{t}}^i, D_{\mathrm{v/t}}^j\right)\right], \end{aligned} \tag{12}$$

While the Euclidean distance used in the heterogeneous center triplet loss effectively optimizes the distances between positive as well as negative test pairs in the feature space (i.e., $e_1 + \rho < e_2$), it has a potential flaw in algorithm design. Specifically, it cannot effectively constrain the angles between feature vectors, leading to uncertain angles between feature vectors in the feature space. See Figure 4 for an example of how this could lead to situations where the positive sample's angle α with the anchor point is larger than the negative sam-

Figure 4
Schematic diagram of the distribution of positive as well as negative tests in the feature space



ple's angle β. When it comes to training, having embedded features with angular discriminability is crucial for success. This allows for accurate predicted scores for each category based on feature vectors. The final fully connected layer calculates the dot product between the feature vector set and the weight vector set for different categories, resulting in precise predictions. Don't underestimate the importance of these factors when it comes to achieving success in training.

During the training phase, the embedded features in the feature space must possess angular discriminability. To obtain predicted scores for each category based on feature vectors, the last fully connected layer calculates the dot product between the feature vector set $\{f_i\}$i∈[1,N] and the weight vector set $\{\omega_j\}$ j∈[1,C] for different categories, as shown in Equation (13). To avoid biases towards certain categories, the size of the weight vectors ω_j should be close to each other. Therefore, the prediction scores for each category for each feature vector f_i in the fully connected layer somewhat depend on the angle θ(i,j) between the embedded feature vectors.

$$f_i \cdot \omega_j = |f_i||\omega_j|\cos(\theta(i,j)) \tag{13}$$

As a solution, this work suggests the angle-based heterogeneous center triplet loss $M_{\text{AC\_tri}}$. Since cosine distance, as shown in Equation (14), focuses on the angles between vectors, it imposes a strong constraint on feature vectors to learn appropriate directions in the feature space.

$$D(X,Y) = 1 - \cos(X,Y) = 1 - \frac{X \cdot Y}{|X||Y|} \tag{14}$$

Therefore, this paper attempts to replace the commonly used Euclidean distance in triplet loss with cosine distance to constrain the angles between feature vectors. By substituting Equation (14) into Equation (12) and simplifying, the resulting $M_{\text{AC\_tri}}$ is expressed in Equation (15).

$$\begin{aligned} M_{\text{AC\_tri}} = &\sum_{i=1}^{P}\left[\rho - \cos(D_{\text{v}}^i, D_{\text{t}}^i) + \max_{j\neq i}\cos\left(D_{\text{v}}^i, D_{\text{v/t}}^j\right)\right]_+ \\ &+\sum_{i=1}^{P}\left[\rho - \cos(D_{\text{t}}^i, D_{\text{v}}^i) + \max_{j\neq i}\cos\left(D_{\text{t}}^i, D_{\text{vit}}^j\right)\right]_+ \end{aligned} \tag{15}$$

The issue with the $M_{\text{AC\_tri}}$ obtained in Equation (15) is that it causes $\cos\left(D_{\text{v}}^i, D_{\text{v/t}}^j\right)$ to approach −1, indicating a negative correlation. Since the training goal is for the anchor point to be unrelated to the negative sample, the [·] function is added to restrict $\cos\left(D_{\text{v}}^i, D_{\text{v/t}}^j\right)$ to approach 0. Considering that the cosine function's range is [−1,1], this paper sets ρ to 1 to simplify the parameters. During experiments, the small numerical values of $M_{\text{AC\_tri}}$ led to slow model convergence, affecting normal training. To accelerate training speed, the paper introduces the exponential function y = ex into Equation (15), resulting in the final angle-based heterogeneous center triplet loss $M_{\text{AC\_tri}}$ as shown in Equation (16). In this equation, λ_1 and λ_2 are weight coefficients, both set to 0.1 based on empirical values.

$$\begin{aligned} M_{\text{AC\_tri}} = &\lambda_1 \sum_{i=1}^{P} e^{\left(\left[\max_{j\neq i}\cos\left(D_v^i, D_{vi}^j\right)\right]_+ - \cos\left(D_v^i, D_t^i\right)+1\right)} \\ &+ \lambda_2 \sum_{i=1}^{P} e^{\left(\left[\max_{j\neq i}\cos\left(D_t^i, D_{ti}^j\right)\right]_+ - \cos\left(D_t^i, D_v^i\right)+1\right)} \end{aligned} \tag{16}$$

2.3.2. Cross-entropy loss

To ensure the accuracy of model classification, cross-entropy loss integrates identity information by assigning different pedestrians to different categories. This paper employs the cross-entropy loss proposed in the literature by Su et al. [31], combining label smoothing, as shown in Equation (17). The model output predicts its category through the fully connected layer, where p_i represents the likelihood of the model predicting category i and N is the number of pedestrian categories in the training samples.

$$M_{\text{ID}} = \sum_{i=1}^{L} -q_i \lg(p_i) \begin{cases} q_i = 1 - \frac{L-1}{L}\xi y + i \\ q_i = \frac{\xi}{L}, y = i \end{cases} \tag{17}$$

To prevent overfitting and enhance the model's generalization ability [32], qi is calculated using Equation (17), with ξ as a hyperparameter set to 0.1. During training, it assumes the labels may be incorrect, avoiding excessive trust in training sample labels and improving the model's corrective ability when encountering erroneous labels.

3. Experimental Results and Analysis

3.1. Experimental settings

Our technique is evaluated on two significant visible-infrared pedestrian re-identification datasets, SYSU-MM01 [28] and RegDB [33]. The SYSU-MM01 collection included 287,628 visible light and 15,792 infrared photos of 491 pedestrians from four and two cameras. The model trains with 395 pedestrian photos and tests with 96. Indoor and all-search are supported by the dataset. Cameras 3 and 6 are utilized for infrared imaging, whereas cameras 1, 2, 4, and 5 are used for visible light imaging in all-search mode. There are two sets of indoor visible light cameras: one for the gallery and one for the query. The paper gets photos in single-shot mode, so each gallery human re-identification has one image throughout testing.

There are a total of 8,240 photos in the RegDB collection, with 10 photographs captured in visible light and 10 images captured in infrared for each of the 412 pedestrians. The training set uses half of the photos that are randomly assigned, while the test set makes use of the other half. To retrieve infrared images, the paper uses visible light images as a starting point.

During training, the backbone network is a pre-trained ResNet50 on ImageNet. Data augmentation techniques like horizontal flipping and random wiping are used, and the image size is adjusted to 288 × 144. The network is trained for 660 epochs, where the first 600 epochs focus on training the MatchGAN module using the loss function $M_{\text{MG}} = 10 \times M_{\text{cycle}} + M_{\text{GAN}} + M_{\text{recon}}$. In the subsequent 60 epochs, train the MatchGAN and feature extraction modules simultaneously. MatchGAN and feature extraction loss functions are $M = M_{\text{AC\_tri}} + M_{\text{ID}}, M_{\text{MG}}$, respectively. Both M_{MG} and M only optimize the parameters of their respective modules. Each batch of data randomly samples 6 pedestrians, with each pedestrian providing 8 images per modality, totaling 96 images. The MatchGAN module uses the ADAM optimizer with a learning rate of 0.0001.

The stochastic gradient descent approach is used to optimize the feature extraction module. It all started with a learning rate of 0.1 and dropped to 0.01 at 620 epochs, before dropping even lower to 0.001 at 650 epochs.

Rank-1, Rank-10, and Rank-20 recognition rates, together with mean average precision (mAP), are used as performance evaluation measures in the testing phase in conjunction with the cumulative matching curve. The study runs ten separate tests on the model and averages the results to improve the stability of test accuracy, taking into account the randomness in test picture selection.

Table 1
Comparison of accuracy with other advanced algorithms with the SYSU-MM01 dataset

Method	All-search single shot				Indoor-search single shot			
	$r=1$	$r=10$	$r=20$	mAP	$r=1$	$r=10$	$r=20$	mAP
BDTR [6] (IJCAI-18)	27.32	66.96	81.07	27.32	31.92	77.18	89.28	41.86
eBDTR [6] (IJCAI 18)	27.86	67.32	81.34	28.42	32.46	77.42	89.62	42.46
D2RL8.5 [8] (CVPR-19)	28.9	70.6	82.4	29.4	—	—	—	—
Hi-CMD [25] (CVPR-20)	34.94	77.58	–	35.94	—	—	—	—
AlignGAN [34] (ICCV-19)	42.4	85	93.7	40.7	45.9	87.6	94.4	45.3
JSIA [7] (AAAI-20)	38.1	80.7	89.9	36.9	43.8	86.2 0	94.2	52.9
AGW [35] (IEEE-20)	47.5	84.39	92.14	47.65	54.17	91.14	95.98	62.97
DDAG [36] (ECCV-20)	54.75	90.39	95.81	53.02	61.02	94.06	98.41	67.98
LbA [37] (ICCV-21)	55.41	–	–	54.14	58.46	–	–	66.33
SFANET [38] (IEEE-21)	60.45	91.8	95.16	53.87	64.8	94.67	98.07	75.16
Methods of this article	66.16	94.35	97.95	62.05	71.29	97.87	99.5	76.26

3.2. Comparative experiments

This study takes a look at the SYSU-MM01 dataset and compares the method with 10 of the most recent methods. These methods include BDTR [6], eBDTR [6], D2RL [8], Hi-CMD [25], AlignGAN [34], JSIA [7], AGW [35], DDAG [36], LbA [37], and SFANET [38]. For a more thorough evaluation, these approaches are tested with the SYSU-MM01 dataset in both the all-search and indoor-search modes. Table 1 displays the experimental findings, where $r=1$, $r=10$, and $r=20$ denote Rank-1, Rank-10, and Rank-20, respectively, and mAP stands for mean average precision.

Table 1 shows that in both SYSU-MM01 dataset modes, approaches using a dual-stream network, like DDAG [36], SFANET [38], and the proposed method, outperform D2RL [8] using a single-stream network. This provides more evidence that a dual-stream network can improve the model's sensitivity to characteristics in images captured by several modalities, leading to the acquisition of more discriminative features. The suggested method outperforms other GAN-based algorithms, such as Hi-CMD [25], AlignGAN [34], and JSIA [7], suggesting that merging a dual-stream network with GANs can greatly smooth out variations in image quality across different modalities, hence improving network performance. In addition, when compared to the state-of-the-art SFANET [38] algorithm, the suggested technique outperforms it in all-search mode by 8.18% and in indoor-search mode by 6.49% in terms of Rank-1 and mAP, with a small improvement in mAP. The efficiency of the proposed strategy is confirmed by comparing these experimental data.

Table 2 shows ablation experiments to confirm the suggested method's efficacy. The experiments compared the model's performance under different conditions. In the table, "Baseline" represents the experimental results obtained using only the dual-stream network. "Baseline + Center_tri" and "Baseline + AC_tri" respectively indicate the experimental results obtained by replacing the difficult triplet loss of the dual-stream network with heterogeneous center triplet loss and angle heterogeneous center triplet loss. "Baseline + MatchGAN" represents the experimental results obtained by using MatchGAN for ubiquitous data augmentation, with the dual-stream network. Finally, "Baseline + MatchGAN + AC_tri" indicates the experimental results obtained by using MatchGAN for ubiquitous data augmentation and optimizing the network with angle heterogeneous center triplet loss.

Table 2
Ablation experiments with the SYSU-MM01 dataset

Method	All-search single shot		
	$r=1$	$r=10$	mAP
baseline	57.27	90.01	55.81
Baseline + Center_tri	59.51	88.04	54.61
Baseline + MatchGAN	60.53	90.98	57.83
Baseline + AC_tri	64.74	93.58	59.9
Baseline + MatchGAN + AC_tri	66.16	94.35	62.05

Through the ablation experiments, it can be observed that the dual-stream network proposed in this paper achieves a Rank-1 accuracy of 57.27% using the SYSU dataset in all-search mode, surpassing many existing works [19]. This indicates that the dual-pooling fused features obtained from feature fusion in this paper are richer and more discriminative. Subsequently, replacing the difficult triplet loss in the baseline with heterogeneous center triplet loss and angle heterogeneous center triplet loss led to respective improvements of 2.24% and 3.92% in Rank-1. However, when replaced with heterogeneous center triplet loss, the model exhibited negative growth in Rank-10 and mAP. This suggests that optimizing the model with Euclidean distance does not perform well in clustering positive as well as negative tests in the feature space. The network can only match relatively easy positive samples, and when there are significant variations in viewpoints, poses, etc., it fails to maintain high recognition accuracy. In contrast, using angle heterogeneous center triplet loss optimizes the clustering effect of feature vectors in the feature space by constraining the direction of feature vectors, enabling the model to better determine the category of pedestrian features and improve recognition accuracy.

Adding MatchGAN to the baseline resulted in a 3.26% improvement in Rank-1 and a 2.02% improvement in mAP, indicating that the generated matching images can enhance the dataset and effectively reduce modality differences between images. Finally, using the image generation module for ubiquitous data augmentation and optimizing the model with angle heterogeneous center triplet loss achieved optimal performance, demonstrating the effectiveness of this paper's framework.

To further demonstrate the ability of the proposed matching image generation algorithm to augment cross-modal pedestrian datasets while reducing modality differences between images, MatchGAN was introduced to DGTL [25], DDAG [36], and AWG [35] methods. The experimental results with the SYSU-MM01 dataset are presented in Table 3.

It can be observed that, after introducing the MatchGAN method for augmenting cross-modal datasets, the Rank-1 accuracy

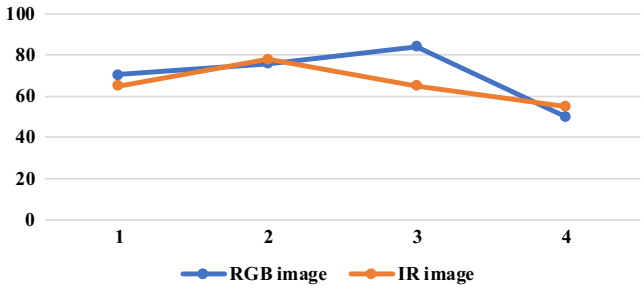
Table 3
Experimental results comparing the effect of MatchGAN ubiquitous data augmentation on different baselines

	All-search single shot		
Baseline	$r = 1$	$r = 10$	mAP
DGTL	52.72	84.93	49.6
DGTL + MatchGAN	55.74	87.18	51.93
DDAG	54.75	90.39	53.02
DDAG + MatchGAN	58.15	91.81	55.64
AWG	47.5	84.39	47.75
AWG + MatchGAN	50.82	86.92	49.89

in the three baseline methods mentioned above improved by 3.02%, 3.40%, and 3.32%, respectively. Additionally, the map values improved by 2.33%, 2.62%, and 2.13%, respectively. The experimental results confirm that the MatchGAN method proposed in this paper can reduce modality differences between images, making it easier for the model to extract robust features against modality changes, thereby enhancing the model's classification ability and performance. Moreover, it can be integrated into existing baselines for ubiquitous data augmentation of cross-modal datasets, enriching the diversity of training samples and reducing the risk of model overfitting.

To more intuitively demonstrate the differences in clustering effects in the feature space of Positive as well as negative tests when optimizing the network with different triplet losses, this paper employs the k-means algorithm on models trained with difficult triplet loss and angle heterogeneous center triplet loss. The models are applied to cluster eight classes of pedestrian images, each consisting of eight RGB images and eight IR images. The clustering results are visualized in the feature space using the t-SNE algorithm, as shown in Figures 5 and 6. In the figure, circles represent RGB images, triangles represent infrared images, and different colors represent different pedestrian categories.

Figure 5
Schematic diagram of the clustering effect using angle heterogeneous center triplet loss



Comparing Figures 5 and 6 with Tables 4 and 5, it is evident that in Figure 6 after optimizing the model with angle heterogeneous center triplet loss, the inter-class distance (e.g., d_2) in the feature space between different pedestrian images is significantly larger than the inter-class distance (e.g., d_1) in Figure 5. This phenomenon indicates that the former makes different pedestrians more distinctive in the feature space, enabling the model to determine the category of input pedestrian images more easily during the testing phase, thereby improving pedestrian re-identification algorithms' accuracy.

Figure 6
Visualization results of the clustering effect of feature vectors in the feature space

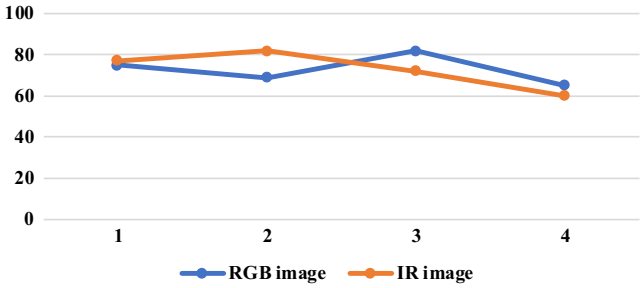


Table 4
Schematic diagram of clustering effect using traditional triplet loss

RGB image	IR image
70.5	65
76	78
84	65
50	55

Table 5
Schematic diagram of the clustering effect using angle heterogeneous center triplet loss

RGB image	IR image
75	77
69	82
82	72
65	60

Further observation of Figure 6 reveals that the clustering effect of each pedestrian sample class is more compact compared to Figure 5. This suggests that optimizing the model with angle heterogeneous center triplet loss can effectively reduce the intra-class distance between images of the same pedestrian. Additionally, compared to Figure 5, the distance between images of the same pedestrian but different modalities (e.g., d_3 and d_4) is significantly reduced. This phenomenon indicates that angle heterogeneous center triplet loss, by constraining features based on angles, can effectively reduce the cross-modal differences between different modality images, mitigating their impact on the model's classification performance.

Table 6 displays the outcomes of the paper's comparison of the suggested method to other sophisticated methods on the RegDB dataset, which further validates the usefulness of the strategy.

Tables 4 and 5 show that when tested on the RegDB dataset, the suggested technique continues to perform well, with an 81.17% rank-1 accuracy. This means the suggested approach is good at generalizing and learning features that are resistant to changes in modality, which aids the model's ability to locate matching cross-modal images. This paper presents a compelling case for the use of angle heterogeneous center triplet loss when optimizing neural networks for pedestrian re-identification. By applying the k-means algorithm to models trained with both difficult triplet loss and angle heterogeneous center triplet loss, the study demonstrates that the latter is more effective in creating more distinctive

Table 6
Comparative experimental results on the RegDB dataset

	Visible light – infrared			
Method	$r = 1$	$r = 10$	$r = 20$	mAP
eBDTR [6]	34.62	58.96	68.72	33.46
D2RL [8]	43.4	66.1	76.3	44.1
Hi-CMD[25]	70.93	86.39	——	66.04
AlignGAN[34]	57.9	—	——	53.6
JSIA [7]	48.5	—	——	49.3
AGW [35]	70.05	86.21	91.55	66.37
DDAG [36]	69.34	86.19	91.49	63.46
SFANET [38]	76.31	91.02	94.27	68
Methods of this article	81.17	93.87	97.65	70.13

inter-class distances in the feature space. This means that different pedestrians are more easily identifiable by the model during testing, leading to improved accuracy in pedestrian re-identification algorithms. These results make a strong argument for the importance of using angle heterogeneous center triplet loss in neural network optimization for tasks such as pedestrian re-identification. It is interesting to note that the angular loss has been found to achieve better similarity than the traditional triplet loss. This is because the angular loss considers the relative ratio of edges and treats all three edges as opposed to the traditional triplet loss, which only deals with two edges. To ensure stable training, a triangle of three triplet points is constructed to minimize the angle ($\angle n$) in the negative sample.

4. Conclusion

A pedestrian re-identification technique that combines a GAN with angle heterogeneous center triplet loss is proposed in this paper. The paper addresses the challenge of visible-to-infrared pedestrian re-identification algorithms in cross-modal settings and seeks to overcome this issue. Initially, the MatchGAN module is utilized to decrease the cross-modal disparities across images while simultaneously increasing the size of the dataset. Afterward, to extract picture features, a dual-stream network that utilizes ResNet50 as its backbone is utilized. To lower the sensitivity of the network to outliers and to improve the clustering impact of Positive as well as negative tests in the feature space, the proposed angle heterogeneous center triplet loss and cross-entropy loss are utilized for joint training. This is done to achieve the aforementioned effects. The overall performance of the network is consequently improved as a result of this improvement. The effectiveness of the suggested algorithm has been validated by extensive comparative studies, which show that the proposed method contributes significantly to an improvement in recognition accuracy. While GAN-based methods show promise, challenges remain. These include developing more interpretable GANs, incorporating diverse techniques, and integrating GANs with mini-batch training for real-time responses. Our supervised training approach requires many instances of the same identity in different poses and contexts. However, exploring unsupervised scenarios presents an intriguing challenge. Adaptation of unsupervised person-ID approaches could be a solution. Overcoming these challenges will unlock the full potential of GAN-based methods.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Author Contribution Statement

Mohammad Shabaz: Conceptualization, Methodology, Resources, Writing – original draft, Visualization, Supervision, Project administration. **Mukesh Soni:** Software, Validation, Formal analysis, Investigation, Data curation, Writing – review & editing.

References

[1] Swanson, K., Wu, E., Zhang, A., Alizadeh, A. A., & Zou, J. (2023). From patterns to patients: Advances in clinical machine learning for cancer diagnosis, prognosis, and treatment. *Cell*, *186*(8), 1772–1791. https://doi.org/10.1016/j.cell.2023.01.035

[2] Bloemheuvel, S., van den Hoogen, J., Jozinović, D., Michelini, A., & Atzmueller, M. (2023). Graph neural networks for multivariate time series regression with application to seismic data. *International Journal of Data Science and Analytics*, *16*(3), 317–332. https://doi.org/10.1007/s41060-022-00349-6

[3] Mukherjee, D., Gupta, K., Chang, L. H., & Najjaran, H. (2022). A survey of robot learning strategies for human-robot collaboration in industrial settings. *Robotics and Computer-Integrated Manufacturing*, *73*, 102231. https://doi.org/10.1016/j.rcim.2021.102231

[4] Maass, W., & Storey, V. C. (2021). Pairing conceptual modeling with machine learning. *Data & Knowledge Engineering*, *134*, 101909. https://doi.org/10.1016/j.datak.2021.101909

[5] Elbir, A. M., Gurbilek, G., Soner, B., Papazafeiropoulos, A. K., Kourtessis, P., & Coleri, S. (2022). Vehicular networks for combating a worldwide pandemic: Preventing the spread of COVID-19. *Smart Health*, *26*, 100353. https://doi.org/10.1016/j.smhl.2022.100353

[6] Villanueva Zacarias, A. G., Reimann, P., Weber, C., & Mitschang, B. (2023). AssistML: An approach to manage, recommend and reuse ML solutions. *International Journal of Data Science and Analytics*, *16*(4), 455–479. https://doi.org/10.1007/s41060-023-00417-5

[7] Jawad, A. T., Maaloul, R., & Chaari, L. (2023). A comprehensive survey on 6G and beyond: Enabling technologies, opportunities of machine learning and challenges. *Computer Networks*, *237*, 110085. https://doi.org/10.1016/j.comnet.2023.110085

[8] Sayed, A. N., Himeur, Y., & Bensaali, F. (2022). Deep and transfer learning for building occupancy detection: A review and comparative analysis. *Engineering Applications of Artificial Intelligence*, *115*, 105254. https://doi.org/10.1016/j.engappai.2022.105254

[9] Zhang, J., Tan, L., Tao, X., Pham, T., & Chen, B. (2020). Relational intelligence recognition in online social networks—A survey. *Computer Science Review*, *35*, 100221. https://doi.org/10.1016/j.cosrev.2019.100221

[10] Ross, A., Banerjee, S., & Chowdhury, A. (2020). Security in smart cities: A brief review of digital forensic schemes for biometric data. *Pattern Recognition Letters*, *138*, 346–354. https://doi.org/10.1016/j.patrec.2020.07.009

[11] Guarino, G., Samet, A., & Cavallucci, D. (2022). PaTRIZ: A framework for mining TRIZ contradictions in patents. *Expert Systems with Applications*, *207*, 117942. https://doi.org/10.1016/j.eswa.2022.117942

[12] Khan, A. A., Zhang, X., Hajjej, F., Yang, J., Ku, C. S., & Por, L. Y. (2024). ASMF: Ambient social media forensics chain of custody with an intelligent digital investigation process using federated learning. *Heliyon*, *10*(1), E23254. https://doi.org/10.1016/j.heliyon.2023.e23254

[13] Xiong, P., Buffett, S., Iqbal, S., Lamontagne, P., Mamun, M., & Molyneaux, H. (2022). Towards a robust and trustworthy machine learning system development: An engineering perspective. *Journal of Information Security and Applications*, *65*, 103121. https://doi.org/10.1016/j.jisa.2022.103121

[14] Golovianko, M., Terziyan, V., Branytskyi, V., & Malyk, D. (2023). Industry 4.0 vs. Industry 5.0: Co-existence, transition, or a hybrid. *Procedia Computer Science*, *217*, 102–113. https://doi.org/10.1016/j.procs.2022.12.206

[15] Lombardi, M., & Vannuccini, S. (2022). Understanding emerging patterns and dynamics through the lenses of the cyber-physical universe. *Patterns*, *3*(11), 100601. https://doi.org/10.1016/j.patter.2022.100601

[16] Fan, Z., Gong, P., Tang, S., Lee, C. U., Zhang, X., Song, P., . . . , & Li, H. (2023). Joint localization and classification of breast masses on ultrasound images using an auxiliary attention-based framework. *Medical Image Analysis*, *90*, 102960. https://doi.org/10.1016/j.media.2023.102960

[17] Ahmad, S., Shakeel, I., Mehfuz, S., & Ahmad, J. (2023). Deep learning models for cloud, edge, fog, and IoT computing paradigms: Survey, recent advances, and future directions. *Computer Science Review*, *49*, 100568. https://doi.org/10.1016/j.cosrev.2023.100568

[18] Canonico, R., & Sperlì, G. (2023). Industrial cyber-physical systems protection: A methodological review. *Computers & Security*, *135*, 103531. https://doi.org/10.1016/j.cose.2023.103531

[19] Rao, S. P., Chen, H. Y., & Aura, T. (2023). Threat modeling framework for mobile communication systems. *Computers & Security*, *125*, 103047. https://doi.org/10.1016/j.cose.2022.103047

[20] Zhong, N. N., Wang, H. Q., Huang, X. Y., Li, Z. Z., Cao, L. M., Huo, F. Y., . . . , & Bu, L. L. (2023). Enhancing head and neck tumor management with artificial intelligence: Integration and perspectives. *Seminars in Cancer Biology*, *95*, 52–74. https://doi.org/10.1016/j.semcancer.2023.07.002

[21] Mahalakshmi, V., Sandhu, M., Shabaz, M., Keshta, I., Prasad, K. D. V., Kuzieva, N., . . . , & Soni, M. (2024). Few-shot learning-based human behavior recognition model. *Computers in Human Behavior*, *151*, 108038. https://doi.org/10.1016/j.chb.2023.108038

[22] Yang, J., Xu, Y., Cao, H., Zou, H., & Xie, L. (2022). Deep learning and transfer learning for device-free human activity recognition: A survey. *Journal of Automation and Intelligence*, *1*(1), 100007. https://doi.org/10.1016/j.jai.2022.100007

[23] Meng, K., Cao, Y., Peng, X., Prybutok, V., & Youcef-Toumi, K. (2020). Smart recovery decision-making for end-of-life products in the context of ubiquitous information and computational intelligence. *Journal of Cleaner Production*, *272*, 122804. https://doi.org/10.1016/j.jclepro.2020.122804

[24] Xu, M., Chen, Z., Zheng, J., Zhao, Q., & Yuan, Z. (2023). Artificial intelligence-aided optical imaging for cancer theranostics. *Seminars in Cancer Biology*, *94*, 62–80. https://doi.org/10.1016/j.semcancer.2023.06.003

[25] Soni, M., Singh, A. K., Babu, K. S., Kumar, S., Kumar, A., & Singh, S. (2022). Convolutional neural network based CT scan classification method for COVID-19 test validation. *Smart Health*, *25*, 100296. https://doi.org/10.1016/j.smhl.2022.100296

[26] de Silva, A., Mori, I., Dusek, G., Davis, J., & Pang, A. (2021). Automated rip current detection with region based convolutional neural networks. *Coastal Engineering*, *166*, 103859. https://doi.org/10.1016/j.coastaleng.2021.103859

[27] Maier, H. R., Galelli, S., Razavi, S., Castelletti, A., Rizzoli, A., Athanasiadis, I. N., . . . , & Humphrey, G. B. (2023). Exploding the myths: An introduction to artificial neural networks for prediction and forecasting. *Environmental Modelling & Software*, *167*, 105776. https://doi.org/10.1016/j.envsoft.2023.105776

[28] Varma, I. M., & Kumar, N. (2023). A comprehensive survey on SDN and blockchain-based secure vehicular networks. *Vehicular Communications*, *44*, 100663. https://doi.org/10.1016/j.vehcom.2023.100663

[29] Li, Y., Chang, J., Kong, C., & Bao, W. (2022). Recent progress of machine learning in flow modeling and active flow control. *Chinese Journal of Aeronautics*, *35*(4), 14–44. https://doi.org/10.1016/j.cja.2021.07.027

[30] Nixdorf, S., Zhang, M., Ansari, F., & Grosse, E. H. (2022). Reciprocal learning in production and logistics. *IFAC-PapersOnLine*, *55*(10), 854–859. https://doi.org/10.1016/j.ifacol.2022.09.519

[31] Su, J., Xu, B., & Yin, H. (2022). A survey of deep learning approaches to image restoration. *Neurocomputing*, *487*, 46–65. https://doi.org/10.1016/j.neucom.2022.02.046

[32] Zhou, K., Wang, L., Pang, L. G., & Shi, S. (2024). Exploring QCD matter in extreme conditions with machine learning. *Progress in Particle and Nuclear Physics*, *135*, 104084. https://doi.org/10.1016/j.ppnp.2023.104084

[33] Marasinghe, R., Yigitcanlar, T., Mayere, S., Washington, T., & Limb, M. (2024). Computer vision applications for urban planning: A systematic review of opportunities and constraints. *Sustainable Cities and Society*, *100*, 105047. https://doi.org/10.1016/j.scs.2023.105047

[34] Sharma, P., Jain, S., Gupta, S., & Chamola, V. (2021). Role of machine learning and deep learning in securing 5G-driven industrial IoT applications. *Ad Hoc Networks*, *123*, 102685. https://doi.org/10.1016/j.adhoc.2021.102685

[35] Khan, S., Farnsworth, M., McWilliam, R., & Erkoyuncu, J. (2020). On the requirements of digital twin-driven autonomous maintenance. *Annual Reviews in Control*, *50*, 13–28. https://doi.org/10.1016/j.arcontrol.2020.08.003

[36] Mundt, M., Hong, Y., Pliushch, I., & Ramesh, V. (2023). A wholistic view of continual learning with deep neural networks: Forgotten lessons and the bridge to active and open world learning. *Neural Networks*, *160*, 306–336. https://doi.org/10.1016/j.neunet.2023.01.014

[37] Bimpas, A., Violos, J., Leivadeas, A., & Varlamis, I. (2024). Leveraging pervasive computing for ambient intelligence: A survey on recent advancements, applications and open challenges. *Computer Networks*, *239*, 110156. https://doi.org/10.1016/j.comnet.2023.110156

[38] Sworna, Z. T., Mousavi, Z., & Babar, M. A. (2023). NLP methods in host-based intrusion detection systems: A systematic review and future directions. *Journal of Network and Computer Applications*, *220*, 103761. https://doi.org/10.1016/j.jnca.2023.103761

How to Cite: Shabaz, M., & Soni, M. (2026). Generative Adversarial-Based Ubiquitous Data Integration Model for Human Re-identification. *Journal of Computational and Cognitive Engineering*, *5*(1), 176–186. https://doi.org/10.47852/bonviewJCCE42022872