Computers, Materials & Continua

Tech Science Press

https://doi.org/10.32604/cmc.2026.078413

ARTICLE

Adversarial AI through Frequency-Domain Imperceptible Attack on Person Re-Identification

Asma Sattar[1], **Maryam Bukhari**[2], **M. Saud Khan**[3], **Anam Mustaqeem**[4], **Mi Young Lee**[5] and **Seungmin Rho**[5,*]

[1]College of Computer and Information Sciences (CCIS), Prince Sultan University, Riyadh, Saudi Arabia
[2]Department of Computer Science, COMSATS University Islamabad, Attock Campus, Attock, Pakistan
[3]Department of Computer Science, Air University Islamabad, Aerospace and Aviation Campus, Kamra, Pakistan
[4]Faculty of Information Technology and Computer Science, University of Central Punjab, Lahore, Pakistan
[5]Department of Industrial Security, Chung-Ang University, Seoul, Republic of Korea

*Corresponding Author: Seungmin Rho. Email: smrho@cau.ac.kr

Received: 30 December 2025; Accepted: 27 April 2026; Published: 15 June 2026

ABSTRACT: Video surveillance systems play an important role in maintaining security in smart city environments. In this context, person identification (Re-ID) systems based on deep learning are currently drawing substantial academic interest. However, these systems remain vulnerable to adversarial attacks. In existing methods, several attacks against Re-ID systems have been designed; nevertheless, they operate in the spatial domain. Existing attacks often suffer from perturbation visibility and low imperceptibility, making them easily detectable by human observers or automated detection systems. From this line of research, this study proposed a novel and potent alternative by designing frequency domain attacks, namely FreqAdv-FFT, FreqAdv-Wavelet, FreqAdv-Phase, FreqAdv-SelDCT, and FreqAdv-RandDCT. The frequency domain allows perturbations to be constructed in a way that utilizes the individual's visual system's decreased sensitivity to specific frequency ranges, making these perturbations less obvious. The proposed adversarial attacks were evaluated on two prominent datasets, Market-1501 and WB_WoB-ReID, across multiple models and attack variants. The highest performance degradation was observed with FreqAdv Wavelet on HRNet for the WB_WoB-ReID dataset, reducing the mean Average Precision (mAP) to 2.52%, and FreqAdv FFT on ResNet-50 for the Market-1501 dataset, achieving a mAP of 3.96%. The suggested attacks provide insights into establishing strong AI models as well as designing defenses for ReID-based surveillance systems that are relevant to the rising development of next-generation real-time applications.

KEYWORDS: Video surveillance systems; person re-identification; deep learning; vulnerabilities; AI-driven security; adversarial attacks; frequency-domain attacks

1 Introduction

Surveillance cameras are rapidly increasing, with a potential revenue of 19.5 billion euros by 2023 [1]. This will give rise to the conceptions of smart cities and AI-driven security ecosystems, which enforce sustainability by improving security systems to handle possible hazards associated with urban environments [2]. Video surveillance has emerged as an evolving technology and is considered one of the vital tools to tackle the challenges of security for different sectors [3]. These include traffic and interior surveillance, as well as criminal activity, violence detection, and critical infrastructure protection [4,5]. These systems aim to extract valuable information from a large collection of videos by consistently identifying, tracking,

and distinguishing individuals of interest, and additionally assessing their activities. Thus, contributing to intelligent surveillance, adversarial AI risk assessment, and next-generation cybersecurity frameworks [6–8]. In the context of video surveillance systems, person-ReID systems have seen an upward trend of interest, particularly within smart city environments and AI-driven security infrastructures [9,10]. Person Re-ID is a retrieval task aiming to rank gallery images based on their similarity to a query image [11–13].

In recent years, deep neural networks (DNNs) have become increasingly popular. However, this study aims to bring the attention of researchers by emphasizing that re-ID systems are vulnerable when they are subjected to adversarial attacks. Initially, adversarial samples have been identified and empirically proven to deceive deep neural networks [14]. In re-ID tasks, deep metric learning is used to learn discriminative distance metrics. Therefore, attacks designed for classification tasks [15,16] often fail to generalize to person Re-ID systems because they focus on pushing images across class boundaries, which does not always distort the ranking-based distortion [17]. Various adversarial approaches have been developed to exploit vulnerabilities in the context of person Re-ID. For instance, some metric adversarial attacks [17] have been designed using FGSM [18], iterative-FGSM [15], and MI-FGSM [16] methods to exploit vulnerabilities. Similarly, several studies have used the feature maps to reduce dispersion [19], hybrid attacks [20], and Private-FGSM (P-FGSM) [21]. Following on, in the spatial domain, the color attack has also been designed [22].

Even though the aforementioned attacks intend to take advantage of vulnerabilities in person re-ID models, their main drawback is their inability to increase imperceptibility. Specifically, examining adversarial perturbations in extremely visible low-frequency components is a crucial area of research to address. Although frequency-domain adversarial attacks based on perturbations have been explored in related fields [23], their behavior and impact on person Re-ID systems need further investigation. Person re-identification systems rely on metric learning, where similarity is computed in an embedding space, and even small shifts in feature representations can significantly alter ranking results. In comparison to existing frequency-driven attacks, the proposed method exploits distinct spectral properties such as global frequency manipulation, sub-band targeted perturbations, phase-based structural disruption, and selective or random frequency coefficient modification. This multi-transform and embedding-aware design clearly distinguishes the proposed approach from existing frequency-domain adversarial attacks. For instance, in which perturbations are limited to certain frequency components to achieve imperceptibility [23–25]. Therefore, we investigate how visually imperceptible perturbations can mislead these systems and impact their performance.

Moreover, the proposed frequency-domain perturbations modify the global texture as well as structural information in the image, having a significant impact on the features extracted by deep networks. Therefore, changing or perturbing these selected frequency components can shift the learned embeddings more effectively than localized spatial perturbations. Hence, this property makes embedding-aware frequency perturbations particularly suitable for fooling metric learning-based Re-ID systems while maintaining visual imperceptibility. More precisely, in this paper, five novel attacks exploiting the characteristics of the frequency domain have been proposed. In addition, the attacks have been conducted on person re-ID systems by perturbing the frequency components with the assistance of triplet loss, which is a good loss for metric learning. Experiments are conducted on two popular person re-identification image datasets, including Market-1501 and WB_WoB-ReID. The following are the pinpoint contributions of this research study:

- A novel frequency-driven adversarial AI attack is proposed to expose the vulnerability of person re-identification systems.
- Five novel variants, including FreqAdv-FFT, FreqAdv-Wavelet, FreqAdv-Phase, FreqAdv-SelDCT, and FreqAdv-RandDCT attacks, exploit the frequency domain characteristics to craft an adversarial image.

- The proposed attacks target specific frequency components, adding perturbations in targeted frequency bands as well as masking techniques to generate more imperceptible perturbations.
- The proposed attacks strive to create a good balance of attack success rates (i.e., in terms of decreased rank scores) vs. the imperceptibility of adversarial images, contributing to research in adversarial AI.

The remainder of this article is divided into different sections: Section 2 provides the related work, Section 3 discusses the proposed work, and Section 4 shows the experimental results, followed by Sections 5, which provide a comparative analysis, limitations, and future directions.

2 Related Work

Deep learning-based algorithms demonstrate superior performance in nearly every domain [26–28]. Aside from that, the application of deep learning models for person ReID has also been tremendously increasing because of improved performances [7]. However, in the light of AI-driven multimedia security and adversarial AI, determining the resilience of person Re-ID systems is critical. One of the pioneer works in this regard is the demonstration of adversarial attacks proposed by Szegedy et al. [14] in which the vulnerability of classification models is exploited. Such attacks have been categorized into unbounded [29,30], bounded [31], and Gradient Reconstitution attacks [32,33], which can be conducted in different settings such as white-box, black-box, and gray-box attacks.

More specifically, Goodfellow et al. [18] proposed the fast gradient sign method (FGSM), which generates adversarial samples in a single step. This method has an extended version found in the study of Dong et al. [16], where an additional momentum factor is introduced; thereby referred to as Momentum Iterative (MI-FGSM). Furthermore, Madry et al. [31] have designed a new variant of attack named Project Gradient Descent (PGD). Bai et al. [17] perform these attacks, including FGSM [18], iterative-FGSM [15], and MI-FGSM [16] with different settings to evaluate the vulnerability of person Re-ID systems. Wang et al. [34] proposed a deep mis-ranking method for conducting an adversarial attack on a person re-ID model. In this approach, a unique architecture known as multi-stage is developed for extracting universal and transferable information from the multiple layers of adversarial perturbations.

Following on, de O Andrade et al. [20] proposed the concept of combining two attacks against a person re-ID model, i.e., the integrated concept of deep mis-ranking [34] and Private-FGSM (P-FGSM) [21]. Zheng et al. [19] proposed a novel idea of lessening the dispersion from the internal feature maps of deep neural networks. Bouniot et al. [35] proposed the metric adversarial attack against the person-re-ID comprising two components, including the self-metric attack and the farthest negative attack. The major objective of their method is to perturb the images in a way that distorts the distance metric among the feature vectors. Likewise, a physical adversarial attack has also been proposed against the person re-ID, in which wearing a printed adversarial pattern 'advPattern' will make re-ID vulnerable to a greater extent [36].

Following on, the color attack has also been proposed by Gong et al. [22] in which local transformation attack (LTA) depends upon the variations of colors. Furthermore, Wang et al. [37] presented Smoothing Adversarial Domain Attacks in cross-domain person re-ID, in which the objective is to include guidance for aligning the source domain images to target domain images utilizing the pre-trained camera classifiers. Similarly, Kanwal et al. [38] proposed the adversarial haze attack to exploit the vulnerability of person re-ID. Unlike the untargeted attacks, Chang et al. [39] proposed the target variant of adversarial attack on person re-ID.

To conclude the literature on adversarial attacks for person re-identification, most of the adversarial attacks target the spatial domain characteristics to achieve imperceptibility while maintaining attack success rates. Existing spatial attack perturbation introduces apparent remains that defensive techniques can quickly

identify. In contrast to the methodology proposed in this study, in which perturbations are introduced in the low-frequency components, which are less perceptible by humans and detectors.

3 Methodology

This section explains the working of the proposed work in detail, while the pictorial representation is depicted in Fig. 1.

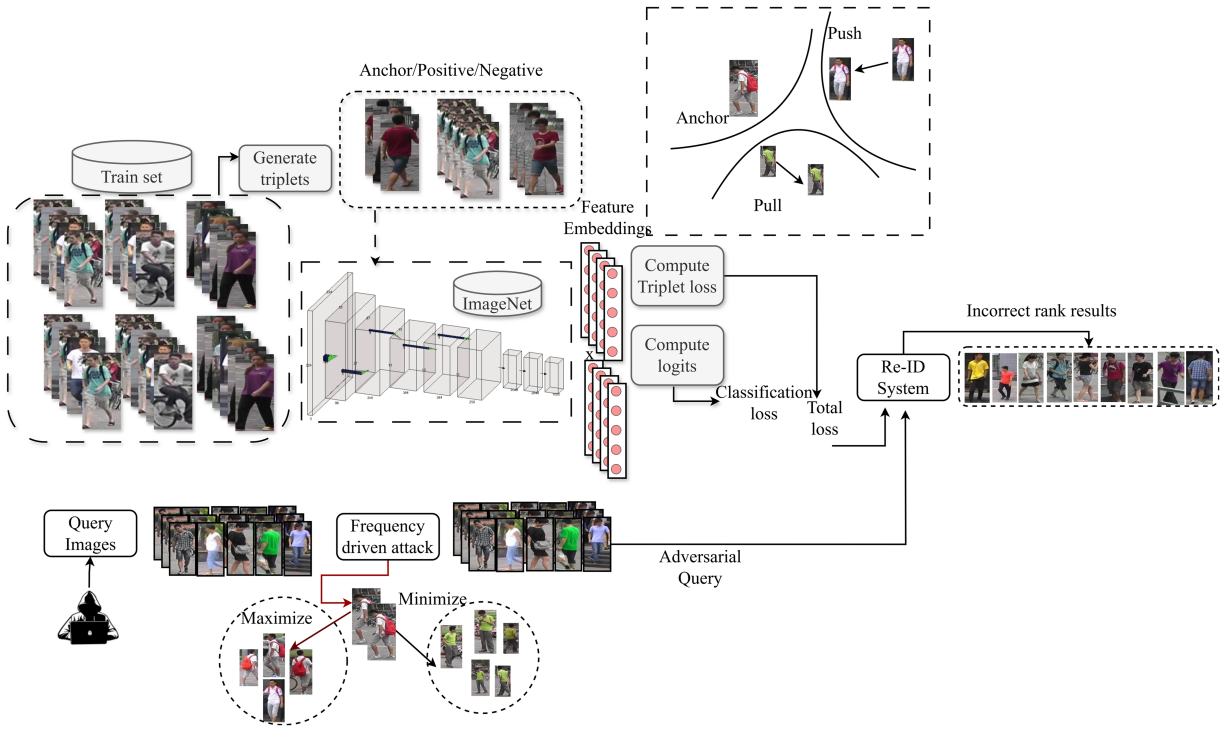

Figure 1: Pictorial overview of the proposed frequency-driven adversarial attack framework for person re-identification. The method computes triplet embeddings, back-propagates the adversarial gradient, and applies one of five frequency-domain perturbation strategies.

3.1 Background on Adversarial Person Re-Identification

Person re-ID problem consists of three primary sets, such as a collection of probe images $P = \{Probe_i\}_{i=1}^{T_p}$, which is ranked against the gallery set images $G = \{Gallery_i\}_{i=1}^{T_g}$, and the collection of images in the training set $Y = \{Train_i\}_{i=1}^{T_y}$. All these sets are labelled to indicate the identity of individuals. In a non-adversarial context, a person re-ID model $F(.)$ is trained by learning the embeddings of input images $f \in \mathbb{R}^d$, wherein d denotes the dimensionality of the feature space using the triplet loss function $\mathcal{L}_{trip}$.

During testing, when a query image $Q_{img} \in P$ from the probe set $P = \{Probe_i\}_{i=1}^{T_p}$ is provided as input to the re-ID F, the model returns the ranking from the gallery with the query image being correctly identified. However, this cannot be the case when the query image Q_{img} is perturbed with the adversarial perturbation to generate its corresponding adversarial image Q_{adv}. The objective is to compute Q_{adv} such that:

$$F(Q_{img}) \neq F(Q_{adv}) \tag{1}$$

This is performed by maximizing the distance of the embedding of the given query image Q_{img}, denoted as $f_{Q_{img}} = F(Q_{img})$ is maximized with its correct match image (positive sample), while minimizing incorrect matches (negative samples). This is achieved by the following objective:

$$maximize(d_{pos} - d_{neg}) \tag{2}$$

In the above Eq. (2), the d_{pos} is aimed at increasing the distance of the anchor image from its positive sample, while minimizing the distance among the negative samples d_{neg}. More precisely, $d_{neg} = \left\| f_{Q_{img}} - f_{pos} \right\|_2$ and $d_{neg} = \left\| f_{Q_{img}} - f_{neg} \right\|_2$ while the $f_{Q_{img}}, f_{pos}$ and f_{neg} are the embedding of the anchor image (perturbed query image), positive image, and negative image.

3.2 Person Re-ID Models for Adversarial Evaluation

The person re-ID models are initially trained using a triplet loss and cross-entropy loss function to build a baseline model whose performance appears promising before adversarial attack. To design a good baseline model, state-of-the-art deep learning models like ResNet-50 [40], DenseNet-121 [41], and HR-Net [42] are used.

3.3 Proposed Frequency-Domain Adversarial AI Attack Framework

In the second phase, adversarial attacks based on frequency have been proposed to exploit the model's vulnerability.

3.3.1 FreqAdv FFT

The FreqAdv FFT attack, also known as the Fast Fourier Transform attack, seeks to induce perturbation in the frequency domain. In images, low-frequency components imply continuous shifts, whereas high-frequency components reveal abrupt shifts. This approach generates swift Fourier transformations for images. The frequency components that capture pixel fluctuations at different scales are subsequently modified in order to add adversarial noise uniformly in the frequency domain. Mathematically, suppose Q_{img} is the query or anchor image that is going to be perturbed to generate its corresponding adversarial image Q_{adv}. Likewise, P_{img} implies the positive reference image of the same identity as that of the anchor image, and N_{img} are the negative reference images, while $F(.)$ is the function that returns the embedding of the images computed using the re-ID model trained model, and d denotes the pairwise distances among the embedding of triplets. First, as indicated by Eq. (3), the gradient of the loss function with respect to the specified query or anchor is calculated. Next, as indicated below in Eqs. (3)–(5), FFT is applied to both the gradient and the input anchor:

$$\nabla_{Q_{img}} \mathcal{L} = \frac{\partial \mathcal{L}}{\partial Q_{img}} \tag{3}$$

$$F\left(Q_{img}\right) = FFT\left(Q_{img}\right) \tag{4}$$

$$F\left(\nabla_{Q_{img}} \mathcal{L}\right) = FFT\left(\nabla_{Q_{img}} \mathcal{L}\right) \tag{5}$$

In the above equation, $\mathcal{L}$ is triplet loss. In the subsequent step, a perturbation of magnitude $\boldsymbol{\varepsilon}$ has been introduced to perturb the FFT coefficients, followed by taking the inverse of the FFT to get back the spatial representation of the image, referred to as the modified perturbed query image, as shown below in Eqs. (6) and (7):

$$F\left(Q_{adv}\right) = F\left(Q_{img}\right) + \varepsilon \cdot \text{ sign } F\left(\nabla_{Q_{img}} \mathcal{L}\right) \tag{6}$$

$$\left(Q_{img}\right) = IFFT\left(F(Q_{adv})\right) \tag{7}$$

Moreover, the clamp (.) function is applied to ensure the values of the adversarial anchor image are in the range of [0, 1], i.e., $clamp(Q_{adv}, 0, 1)$.

3.3.2 FreqAdv Wavelet

In a FreqAdv Wavelet attack, a discrete wavelet transform is computed, and by employing the characteristics of wavelets, the attack aims to perturb specific sub-bands in the image. Targeting specific bands for perturbation allows for the distortion of specific image features, while also introducing perturbations to specific sub-bands that are less perceptible to human eyes, assuring imperceptibility. It is mathematically stated as follows, i.e., gradients are determined, and sub-bands are extracted from the anchor image using DWT.

$$\nabla_{Q_{img}}\mathcal{L} = \frac{\partial \mathcal{L}}{\partial Q_{img}} \tag{8}$$

$$coefficeint = DWT\left(Q_{img}\right) = \{LL, HL, LH, HH\} \tag{9}$$

In the next step, a perturbation of magnitude ε has been added to specific sub-bands as shown below in Eqs. (10)–(12):

$$LH' = LH + \varepsilon \cdot sign\mathcal{F}\left(\nabla_{LH}\mathcal{L}\right) \tag{10}$$

$$HL' = HL + \varepsilon \cdot sign\mathcal{F}\left(\nabla_{HL}\mathcal{L}\right) \tag{11}$$

In the above equations, specific sub-bands are perturbed to produce an adversarial effect. These perturbed sub-bands are then combined with the remaining original sub-bands, and the inverse discrete wavelet transform (DWT) is applied to obtain the final perturbed image in the spatial domain.

$$Q_{adv} = IDWT(LL, (LH', HL', HH) \tag{12}$$

The resulting adversarial images are clipped between $[0, 1]$.

3.3.3 FreqAdv SelDCT Attack

This attack variant, like the FreqAdv FFT, attempts to interfere with the DCT coefficients in order to transform an anchor image to its equivalent adversarial image. Unlike the Fast Fourier Transform (FFT), which divides a signal into sines and cosines, DCT solely employs cosine functions. The very low frequency components of an input image to DCT contain the majority of the information. Therefore, the attack precisely modifies the visual characteristics to interfere with the majority of the information contained in low-frequency components. Eq. (13) shows how the gradients are determined mathematically:

$$\nabla_{Q_{img}}\mathcal{L} = \frac{\partial \mathcal{L}}{\partial Q_{img}} \tag{13}$$

$$\mathcal{F}\left(Q_{img}\right) = DCT\left(Q_{img}\right) \tag{14}$$

$$\mathcal{F}\left(\nabla_{Q_{img}}\mathcal{L}\right) = DCT\left(\nabla_{Q_{img}}\mathcal{L}\right) \tag{15}$$

In the following step, a mask is established to determine specific frequency components to perturb. A mask for a low-frequency region of size (h, w) is defined as follows:

$$(i, j) = \begin{cases} 1 & for\, 0 \le i < h\, and\, 0 \le j < w \\ 0 & otherwsie \end{cases} \tag{16}$$

In our experiments, the low-frequency mask size is set to $h = 8$, $w = 8$, meaning that only the top-left 8×8 DCT coefficients are perturbed for each color channel. This mask influences the low-frequency components of DCT, followed by the inverse of DCT, to generate the adversarial image, as indicated in Eqs. (17) and (18).

$$\mathcal{F}(Q_{adv}) = \mathcal{F}\left(Q_{img}\right) + \varepsilon \cdot M \cdot \text{ sign } \mathcal{F}\left(\nabla_{Q_{img}} \mathcal{L}\right) \tag{17}$$

$$\left(Q_{img}\right) = IDCT\left(Q_{adv}\right) \tag{18}$$

In the last, the clamp (.) function is applied to ensure the values of the adversarial anchor image are in the range of $[0, 1]$, i.e., $clamp(Q_{img}, 0, 1)$.

3.3.4 FreqAdv Phase Attack

FreqAdv Phase attack, also called Phase-shift attack, is another attack variant that incorporates perturbations into the phase components. When an image has been converted from the spatial to the frequency domain, different frequency components and phases have certain types of information. For instance, the power of the frequency component is indicated by the magnitude, while spatial and structural details are hidden in phase. As a result, perturbing such phase information, including geometric information, will result in a more effective attack in decreasing the performance of the person re-ID model. For mathematical formulation, the FFT has been applied over the Q_{img} to acquire the magnitude and phase as shown in Eq. (19):

$$F\left(Q_{img}\right) = FFT\left(Q_{img}\right), A_a = \left|F\left(Q_{img}\right)\right|, \phi_a = phase(F\left(Q_{img}\right) \tag{19}$$

Eqs. (20)–(22) below illustrate how the phase information gets interrupted and then merged with the magnitude to generate a perturbed adversarial image:

$$\phi'_a = \phi_a + \varepsilon.phase(FFT(\nabla_{Q_{img}} \mathcal{L} = \frac{\partial \mathcal{L}}{\partial Q_{img}}) \tag{20}$$

$$\mathcal{F}\left(Q_{adv}\right) = A_a \cdot e^{j\phi'_a} \tag{21}$$

$$Q_{adv} = IFFT(Q_{adv}) \tag{22}$$

The resulting adversarial image is clamped between $[0, 1]$ and targets the spatial patterns and structures of the image by perturbing the phase components.

3.3.5 FreqAdv RandDCT

In this random frequency region attack, the goal is to incorporate changes to an input image by modifying random blocks of discrete cosine transform (DCT) coefficients. The difference of this attack from the selective DCT attack is that in the selective DCT attack, specific low-frequency components are targeted; however, in this attack, the perturbation is distributed across various ranges in terms of random blocks. For

mathematical formulation, the DCT has been applied over the Q_{img} and gradients are computed as shown below in the Eqs. (23)–(25):

$$\nabla_{Q_{img}}\mathcal{L} = \frac{\partial \mathcal{L}}{\partial Q_{img}} \tag{23}$$

$$F\left(Q_{img}\right) = DCT\left(Q_{img}\right) \tag{24}$$

$$F\left(\nabla_{Q_{img}}\mathcal{L}\right) = DCT\left(\nabla_{Q_{img}}\mathcal{L}\right) \tag{25}$$

Next, by selecting the block indices at random, the perturbation has been introduced to random blocks (i, j).

$$\mathcal{F}\left(Q_{adv}\right)\left[i{:}\, i + B_h, j{:}\, j + B_w\right] = F\left(Q_{img}\right)\left[i{:}\, i + B_h, j{:}\, j + B_w\right] + \varepsilon \cdot signF\left(\nabla_{Q_{img}}\mathcal{L}\right)\left[i{:}\, i + B_h, j{:}\, j + B_w\right] \tag{26}$$

In the above Eq. (26), the B_h and B_w denotes the size of the block, and this equation is executed N times, indicating the number of blocks to perturb. In implementation, the DCT coefficients are perturbed in randomly selected blocks of size $B_h \times B_w = 4 \times 4$. Specifically, $N = 10$ such blocks are chosen per image, where each block corresponds to a small spatial frequency region in the DCT domain. In the subsequent step, the inverse DCT is computed, followed by clamping the tensor to acquire the adversarial image as shown below:

$$\left(Q_{img}\right) = IDCT\left(Q_{adv}\right) \tag{27}$$

In the last, the clamp (.) function is applied as shown in the above Eq. (27) to ensure the values of the adversarial anchor image are in the range of $[0, 1]$, i.e., $clamp\left(Q_{adv}, 0, 1\right)$.

4 Experiments and Results

4.1 Datasets

In this study, two popular person re-identification datasets, namely Market-1501 [43] and WB_WoB-ReID [44], have been employed. Table 1 contains information about both datasets, including identities, images, and camera views, especially in the train, test, and query sets.

Table 1: Details of the datasets.

Description	Market-1501	WB_WoB-ReID
Number of images in the bounding box train	12,937	11,719
Number of images in the bounding box test	19,733	10,048
Number of images in query set	3369	1265
Total Identities	1501	1812
Year	2015	2023
Scene	Outdoor	Indoor
Size of Images	128 × 64	Varied
Detector	DPM/Hand	Hand
Cameras	Six	Five

4.2 Evaluation Criteria

The evaluation criteria include the cumulative matching characteristics (CMC) curves and mean average precision (mAP). In an adversarial setting, these evaluation measures, including both CMC and mAP, are computed when adversarial queries are provided as input to the re-ID model. Furthermore, the adversarial approach is validated in terms of imperceptibility and similarity between the adversarial and original images using the SSIM (Structural Similarity Index) and PSNR (Peak Signal-to-Noise Ratio) metrics.

4.3 Results of Person Re-ID under Non-Adversarial Settings

To exploit vulnerabilities in the person re-identification model, its performance is first established under non-adversarial conditions. The models are trained using 100 epochs, 30 batch sizes, a dropout rate of 0.5, a learning rate of 0.08, and warm epochs with decay rates of 5 and 0.0005, respectively. The results of all models on Market-1501, as well as the WB_WoB-ReID dataset, are provided in Table 2. It is observed from Table 2 that under a non-adversarial setting, the DenseNet-121 shows the highest Rank-1 score and mAP, which are about 92.55% and 80.72%.

Table 2: Results of the person Re-ID models in a non-adversarial setting.

Dataset	Model	Rank-1	Rank-2	Rank-3	Rank-4	Rank-5	Rank-10	Rank-15	mAP
Market-1501	ResNet-50	91.42%	94.36%	95.84%	96.50%	96.67%	98.18%	98.72%	78.54%
	DenseNet-121	92.55%	95.37%	96.53%	97.21%	97.42%	98.37%	98.69%	80.72%
	HR-Net	91.75%	94.60%	95.72%	96.32%	96.76%	98.16%	98.63%	78.23%
WB-WOB-ReID	ResNet-50	94.31%	95.57%	96.13%	96.60%	96.76%	97.31%	97.47%	78.31%
	DenseNet-121	94.78%	96.36%	96.84%	97.15%	97.31%	97.63%	97.71%	79.88%
	HR-Net	94.47%	95.81%	96.60%	96.84%	97.15%	97.63%	97.87%	80.00%

4.4 Results on Market-1501 under Adversarial Setting

From the experimental findings of this section, it is observed that the person re-ID model is vulnerable to adversarial attack. Table 3 shows the results of five distinct adversarial attacks on the Market-1501 dataset. Since the results are reported under adversarial attack conditions, lower Rank@1 values indicate stronger attack effectiveness, as they reflect greater degradation in the model's identification performance. According to the results, the ResNet50 model, which achieves a Rank-1 accuracy of 0.9142 (91.42%) on clean inputs, experiences a performance drop of up to 0.0736 (7.36%) under the FreqAdv FFT attack, indicating the largest degradation among the evaluated attacks. However, a comparison of attacks reveals that the "FreqAdv FFT" attack is more effective since it uniformly distorts all frequency components. This behavior can also be attributed to the inherent characteristics of CNN-based person Re-ID models. Moreover, the FFT-based perturbation modifies the global frequency spectrum of the image, thereby affecting distributed texture statistics across the entire spatial extent. Therefore, due to the inherent sensitivity of CNN-based Re-ID models to texture-related information, the FFT-attack, which alters texture statistics, leads to substantial degradation. Compared to other versions, such as selective DCT and wavelet-based variants, which perturb more localized or sub-band-specific frequency components, "FreqAdv FFT" works better. More explicitly, other attacks, however, perturb some partial information, such as some selective components, e.g., in DCT low-frequency components, whereas the "FreqAdv Wavelet" attack perturbs certain particular bands, and the "FreqAdv Phase" attack perturbs phase information. In light of this, the "FreqAdv FFT" attack yields better results. However, it is crucial to highlight that even when specific information, such as frequency bands, is perturbed, advanced deep learning models such as ResNet-50, DenseNet-121, and HRNet are still vulnerable.

Furthermore, the ablation study was carried out by gradually adjusting the intensity of perturbation in the query images. Fig. 2 illustrates the results of an ablation study of adversarial rank scores on various attacks with varied epsilon (ε) values. It is observed from the results that by increasing the value of epsilon (ε), there is a sharper decrease in performance. However, the major concern is to hold the imperceptibility trade-off vs. attack success rates. The selection of epsilon (ε) for each attack is different because of their specific characteristics, e.g., in "FreqAdv SelDCT", it is kept high because, in this, only selective DCT components are perturbed instead of all.

Table 3: Evaluation of adversarial attacks on Market-1501 across three models.

Model	Method	Rank-1	Rank-2	Rank-3	mAP
ResNet-50	No attack	0.9142	0.9436	0.9584	0.7854
	FreqAdv FFT	0.0736	0.0983	0.1191	0.0396
	FreqAdv Wavelet	0.3741	0.4439	0.4920	0.2234
	FreqAdv Phase	0.3774	0.4397	0.4798	0.2792
	FreqAdv SelDCT	0.3008	0.3667	0.4029	0.1981
	FreqAdv RandDCT	0.3572	0.4240	0.4721	0.2744
DenseNet-121	No attack	0.9255	0.9537	0.9653	0.8072
	FreqAdv FFT	0.0778	0.1110	0.1336	0.0463
	FreqAdv Wavelet	0.4163	0.4988	0.5448	0.2555
	FreqAdv Phase	0.4667	0.5410	0.5828	0.3430
	FreqAdv SelDCT	0.3284	0.4115	0.4555	0.2276
	FreqAdv RandDCT	0.6989	0.7583	0.7960	0.5499
HR-Net	No attack	0.9175	0.9460	0.9572	0.7823
	FreqAdv FFT	0.0579	0.0778	0.0897	0.0302
	FreqAdv Wavelet	0.0962	0.1280	0.1502	0.0488
	FreqAdv Phase	0.3420	0.4097	0.4507	0.2452
	FreqAdv SelDCT	0.2904	0.3548	0.3985	0.1917
	FreqAdv RandDCT	0.5650	0.6244	0.6583	0.4176

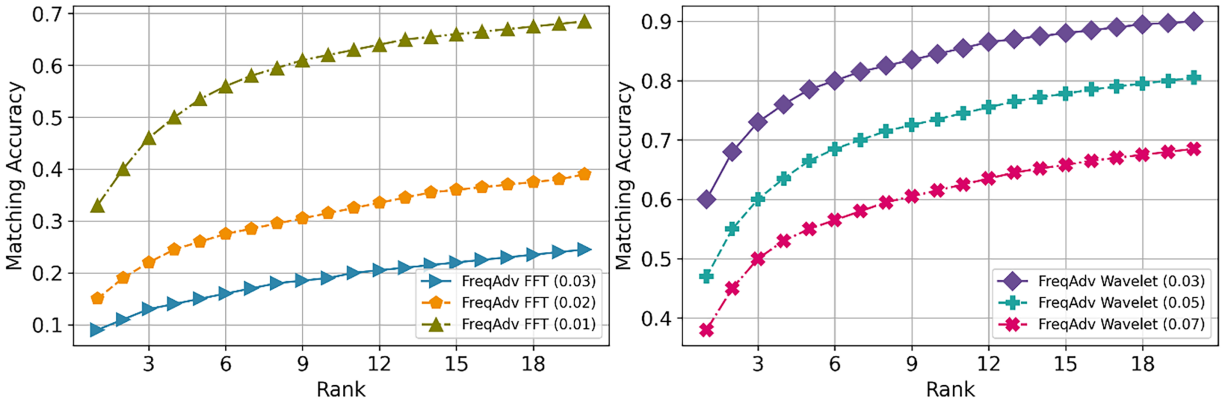

Figure 2: (Continued)

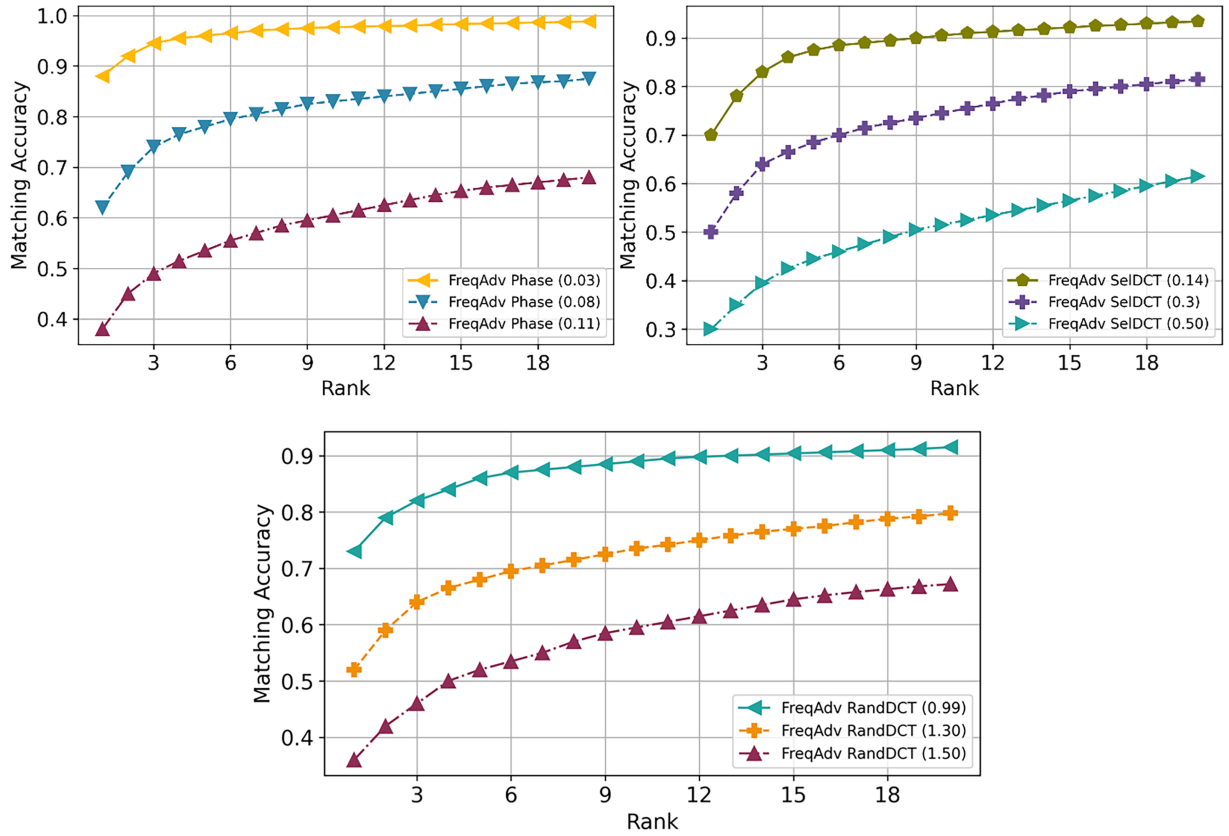

Figure 2: Rank scores in adversarial with different magnitudes of perturbations in all frequency-driven attacks.

4.5 Results of Person Re-ID on WB_WoB-ReID under Adversarial Setting

In the subsequent set of experiments, the effectiveness of the proposed attacks is further evaluated on a second indoor dataset, as shown in Table 4. It can be observed from Table 4 that the original ResNet-50 model achieves a rank-1 accuracy of 0.9431 (i.e., 94.31%), where the values are reported on a normalized scale of 0 to 1. The difference in the performance of each attack is because each attack introduces perturbations differently. For instance, a "FreqAdv FFT" attack is more effective since it uniformly distorts all frequency components. Subsequently, rank results, along with the original and query images for various attacks, are presented in Fig. 3 for the Market-1501 dataset. Fig. 3 clearly shows that when the original images are affected by adversarial manipulation in the frequency domain, the retrieval results are severely affected, i.e., the identity contained in the query image does not rank among the top ten results.

Table 4: Evaluation of adversarial attacks on WB-WoB-ReID across three models.

Model	Method	Rank-1	Rank-2	Rank-3	mAP
ResNet-50	No attack	0.9431	0.9557	0.9613	0.7831
	FreqAdv FFT	0.1510	0.1992	0.2253	0.0977
	FreqAdv Wavelet	0.2411	0.2972	0.3344	0.1902
	FreqAdv Phase	0.3621	0.4253	0.4593	0.2896
	FreqAdv SelDCT	0.2791	0.3304	0.3676	0.2078
	FreqAdv RandDCT	0.1731	0.2182	0.2498	0.1738

(Continued)

Table 4 (continued)

Model	Method	Rank-1	Rank-2	Rank-3	mAP
DenseNet-121	No attack	0.9478	0.9636	0.9684	0.7988
	FreqAdv FFT	0.0917	0.1368	0.1589	0.0862
	FreqAdv Wavelet	0.2877	0.3462	0.3787	0.2129
	FreqAdv Phase	0.4008	0.4640	0.4972	0.3395
	FreqAdv SelDCT	0.3573	0.4134	0.4569	0.2642
	FreqAdv RandDCT	0.5737	0.6229	0.6595	0.4597
HR-Net	No attack	0.9447	0.9581	0.9660	0.8000
	FreqAdv FFT	0.0040	0.0063	0.0111	0.0715
	FreqAdv Wavelet	0.0182	0.0324	0.0403	0.0252
	FreqAdv Phase	0.3747	0.4229	0.4506	0.3070
	FreqAdv SelDCT	0.3304	0.3984	0.4300	0.2380
	FreqAdv RandDCT	0.3605	0.4142	0.4498	0.2749

Figure 3: Top-10 retrieval rankings of the person re-identification model on the Market-1501 dataset under adversarial query conditions.

4.6 Transferability Test

To assess the practical impact of the proposed frequency-domain attack outside of the white-box context, we undertake black-box transferability tests. In this setting, the surrogate (source) model is ResNet50, which generates adversarial examples. The resulting adversarial images are then directly evaluated on two separate target architectures: DenseNet and HRNet, without access to their underlying parameters or gradients. Table 5 shows the black-box findings for the Market-1501 and WB-WoB-ReID datasets with perturbation magnitude of $\varepsilon = 0.01$ and $\varepsilon = 0.02$.

Table 5: Results of proposed attack in black-box settings.

Dataset	ε	DenseNet	HRNet	Original (Rank@1)	Original (mAP)	Rank@1	mAP
Market-1501	0.01	✓	✗	92.55%	80.72%	79.04%	61.40%
	0.01	✗	✓	91.75%	78.23%	78.60%	61.33%
	0.02	✓	✗	92.55%	80.72%	52.01%	35.69%
	0.02	✗	✓	91.75%	78.23%	53.55%	37.07%
WB_WoB-ReID	0.01	✓	✗	94.78%	79.88%	89.56%	70.71%
	0.01	✗	✓	94.47%	80.00%	89.72%	72.38%
	0.02	✓	✗	94.78%	79.88%	72.41%	52.99%
	0.02	✗	✓	94.47%	80.00%	70.51%	52.29%

The initial (clean) Rank@1 and mAP values are presented as a reference, followed by the decreased performance when adversarial images have been evaluated on target models. Although working in a black-box environment, the suggested frequency-domain perturbations consistently reduce ranking performance across architectures. This is to be expected, because adversarial examples generated on a surrogate model may not fully align with the decision boundaries of unseen target models. Moreover, differences in architecture, feature extraction layers, and embedding distributions can reduce the transferability of perturbations, resulting in smaller performance drops compared to white-box settings. Nevertheless, the results still demonstrate that the attacks are transferable across architectures, and the findings highlight the practical robustness of our approach in real-world scenarios, such as when the target Re-ID system is unknown, and suggest that further improvements in cross-model generalization (e.g., ensemble surrogate models or adaptive frequency perturbations) could further enhance black-box effectiveness.

4.7 Comparative Analysis

A comparison was done between proposed attacks with the best performance, namely "FreqAdv FFT", "FreqAdv Wavelet", and "FreqAdv SelDCT", and spatial attacks considered benchmark attacks, comprising PGD, FGSM, and Momentum-FGSM. Tables 6 and 7 show a comparative analysis for both the Market-1501 dataset and WB_WoB-ReID datasets. Table 6 shows that the "FreqAdv FFT" and "FreqAdv SelDCT" attacks perform well in terms of mAP and rank-1 adversarial scores, with PSNR and SSIM improving or approaching the baselines of FGSM, PGD, and MI-FGSM. Each attack's perturbation magnitude (ε) is changed to obtain a comparable amount of Rank-1 accuracy degradation. Perceptual quality (PSNR and SSIM) is then examined to ensure a fair comparison. Table 7 also presents a comparison with previously reported studies on the well-known Market-1501 dataset. It is worth noting that most existing studies do not report SSIM and

PSNR values for the generated adversarial images, although these metrics are important for evaluating and quantifying imperceptibility.

Table 6: Comparative analysis on Market-1501 and WB-WoB-ReID dataset.

Dataset	Methods	PSNR	SSIM	Rank@1	Rank@5	Rank@10	mAP
Market-1501	PGD [15]	33.02	0.933	0.3815	0.5507	0.6261	0.2842
	MI-FGSM [16]	33.02	0.933	0.3815	0.5540	0.6232	0.2843
	FGSM [18]	39.96	0.987	0.3542	0.5240	0.6012	0.2107
	FreqAdv FFT	36.98	0.978	0.2529	0.4038	0.4795	0.1412
	FreqAdv Wavelet	38.92	0.980	0.3931	0.5709	0.6451	0.2385
	FreqAdv SelDCT	36.79	0.987	0.3061	0.4637	0.5454	0.2057
WB-WoB-ReID	PGD [15]	29.912	0.9036	0.4505	0.5683	0.6197	0.3308
	MI-FGSM [16]	29.931	0.8999	0.4237	0.5596	0.6086	0.3102
	FGSM [18]	30.966	0.9319	0.2814	0.4355	0.5130	0.1685
	FreqAdv FFT	30.432	0.9456	0.1510	0.1992	0.2253	0.0977
	FreqAdv Wavelet	31.443	0.9497	0.2545	0.4071	0.4806	0.1988
	FreqAdv SelDCT	31.023	0.9600	0.4418	0.5944	0.6577	0.3097

Table 7: Comparative analysis of proposed best attack with existing studies (their best results) on Market-1501 dataset.

Methods	Perturbation	Original Model (mAP)	PSNR	SSIM	(Attack mAP)
FGSM [17]	Gallery	79.08%	✗	✗	9.178
I-FGSM [17]	Gallery	79.08%	✗	✗	0.519
MI-FGM [17]	Gallery	79.08%	✗	✗	1.022
DR [19]	Query	82.30%	✗	✗	20.2
DeepMisRanking [20]	–	79.10%	✗	✗	6.29
P-FGSM [20]	–	79.10%	✗	✗	98.25
Combined attacks [20]	–	79.10%	✗	✗	6.71
AGS-Attack [45]	Query	70.47%	✗	✗	0.89
Generative Metric Learning [46]		74.00%	✗	✗	6.7
AP-Attack [47]	Query	–	✗	✗	5.7
Re-ID-leak Attack [48]	Query	–	✗	✗	9.9
Unsupervised Adversarial Attack [49]	Query	72.20%	✗	✗	10.0
Proposed FreqAdv FFT	Query	80.72%	✓	✓	4.63

Likewise, Table 8 presents the paired t-test results comparing the baseline methods (PGD, MI-FGSM, FGSM) with the proposed FreqAdv attacks. The results indicate that all proposed attacks significantly outperform the baselines ($p < 0.05$), demonstrating their effectiveness. More specifically, for statistical validation, each experiment was repeated 10 times using different random seeds while keeping the dataset split and model configuration fixed, and a paired t-test was conducted on the resulting Rank@1 scores to evaluate the significance of performance differences.

Table 8: Paired *t*-test results comparing baseline methods against the proposed FreqAdv attacks on Rank@1 performance over 10 runs.

Comparison	t-statistic	*p*-value	Significance
Baseline [15,16,18] vs. FreqAdv FFT	45.336	<0.001	Yes ($p < 0.05$)
Baseline [15,16,18] vs. FreqAdv Wavelet	45.205	<0.001	Yes ($p < 0.05$)
Baseline [15,16,18] vs. FreqAdv SelDCT	46.850	<0.001	Yes ($p < 0.05$)

4.8 Discussions and Implications

The major implications of this research study are to put an emphasis on adversarial attacks on person re-ID. The proposed attacks generate adversarial examples (i.e., query/anchor images) that are more imperceptible than those from baseline methods. The findings of the research show that adversarial examples with some minor changes in their frequency components, especially for instance perturbation in selective bands, components still fool the re-ID models. Furthermore, this study's findings can help to design more effective intelligent defensive frameworks. This will ultimately evolve the development of more secure, resilient, and legally compliant person Re-ID surveillance systems, and promote critical infrastructure protection, and trustworthy smart city deployments.

Rather than just analyzing the implications and potential of this study, we also bring attention to some limitations of the proposed work, which can be addressed in future research studies. As a result, the proposed work primarily focuses on a white-box setting, where full access to the target model is available. Although we also conducted experiments in a black-box scenario, the attack performance was comparatively limited due to the absence of direct gradient information from the target model. In future work, we aim to further enhance the effectiveness of our approach in black-box settings.

Furthermore, the attacks are carried out using a model trained with a triplet loss, but we will enhance it by adding specialized loss functions such as circle loss. Furthermore, in this research, we concentrate on digital-domain adversarial evaluation to fully investigate ranking manipulation and imperceptibility in person Re-ID systems. Physical-world attacks, such as real-world clothes or patch-based perturbations under different camera settings, pose new environmental problems. Developing such physical implementations is an essential area for future study. Moreover, this work focuses on exploiting vulnerabilities in person Re-ID systems. However, designing and studying defenses against such attacks, such as adversarial training, spectral filtering, and certified methods, remains a critical subject for future research.

5 Conclusion

Deep learning, particularly convolutional neural networks, has made substantial progress in recent years, resulting in numerous significant achievements in the field of person re-id. However, despite their successes, these systems can easily be tricked by adversarial perturbations. The existing methodologies expose this concept by leveraging this perturbation in the spatial domain. Nevertheless, this will lead to noticeable perturbation in images. To improve this and develop a new line of attacks, this study attempts to propose a frequency-driven novel adversarial attack having five variants, targeting different characteristics of the frequency domain. The proposed attacks maintain a high level of imperceptibility compared to existing methods, in addition to a greater degradation in model performance.

Acknowledgement: The authors would like to thank Prince Sultan University for their support.

Funding Statement: This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (RS-2025-00563192 & RS-2025-00518960).

Author Contributions: The authors confirm contributions to the paper as follows: Asma Sattar: Data curation, Methodology, Implementation, Investigation; Maryam Bukhari: Validation, Visualization, Investigation, Writing—original draft; M. Saud Khan: Conceptualization, Validation, Software, Writing—review & editing; Anam Mustaqeem: Methodology, Software, Writing—review & editing; Mi Young Lee: Resources, Validation, Investigation; Seungmin Rho: Resources, Supervision, Validation, Investigation, Writing—review & editing. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The data will be made available from the corresponding author on reasonable request.

Ethics Approval: Not applicable.

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

References

1. Khan P, Byun YC, Park N. A data verification system for CCTV surveillance cameras using blockchain technology in smart cities. Electronics. 2020;9(3):484. doi:10.3390/electronics9030484.
2. Abbass M, Abbas U, Jafri R, Arif SM, Akhai S. AI and machine learning applications in sustainable smart cities. In: Sustainable smart cities and the future of urban development. Hershey, PA, USA: IGI Global Scientific Publishing; 2025. p. 1–32.
3. Zhou Y, Liu L, Shao L, Mellor M. Fast automatic vehicle annotation for urban traffic surveillance. IEEE Trans Intell Transport Syst. 2018;19(6):1973–84. doi:10.1109/tits.2017.2740303.
4. Puvvadi ULN, Di Benedetto K, Patil A, Kang KD, Park Y. Cost-effective security support in real-time video surveillance. IEEE Trans Ind Inf. 2015;11(6):1457–65. doi:10.1109/tii.2015.2491259.
5. Chen J, Li K, Deng Q, Li K, Yu PS. Distributed deep learning model for intelligent video surveillance systems with edge computing. IEEE Trans Ind Inform. 2019:1. doi:10.1109/TII.2019.2909473.
6. Wang X. Intelligent multi-camera video surveillance: a review. Pattern Recognit Lett. 2013;34(1):3–19. doi:10.1016/j.patrec.2012.07.005.
7. Bukhari M, Yasmin S, Naz S, Maqsood M, Rew J, Rho S. Language and vision based person re-identification for surveillance systems using deep learning with LIP layers. Image Vis Comput. 2023;132:104658. doi:10.1016/j.imavis.2023.104658.
8. Haque SBU. A fuzzy-based frame transformation to mitigate the impact of adversarial attacks in deep learning-based real-time video surveillance systems. Appl Soft Comput. 2024;167:112440. doi:10.1016/j.asoc.2024.112440.
9. Saxena H, Ragha L. Performance analysis of deep learning models for re-identification of a person in a public surveillance system. In: Smart urban computing applications. Gistrup, Denmark: River Publishers; 2023. p. 23–53.
10. Perwaiz N, Fraz MM, Shahzad M. Smart surveillance with simultaneous person detection and re-identification. Multimed Tools Appl. 2024;83(5):15461–82. doi:10.1007/s11042-022-13458-y.
11. Wu D, Zheng SJ, Zhang XP, Yuan CA, Cheng F, Zhao Y, et al. Deep learning-based methods for person re-identification: a comprehensive review. Neurocomputing. 2019;337(1):354–71. doi:10.1016/j.neucom.2019.01.079.
12. Zheng L, Yang Y, Hauptmann AG. Person re-identification: past, present and future. arXiv:1610.02984. 2016.
13. Asperti A, Fiorilla S, Nardi S, Orsini L. A review of recent techniques for person re-identification. arXiv:2509.22690. 2025.
14. Szegedy C, Zaremba W, Sutskever I, Bruna J, Erhan D, Goodfellow I, et al. Intriguing properties of neural networks. arXiv:1312.6199. 2013.
15. Kurakin A, Goodfellow IJ, Bengio S. Adversarial examples in the physical world. In: Artificial intelligence safety and security. 1st ed. Boca Raton, FL, USA: CRC Press; 2018. p. 99–112. doi:10.1201/9781351251389-8.

16. Dong Y, Liao F, Pang T, Su H, Zhu J, Hu X, et al. Boosting adversarial attacks with momentum. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2018 Jun 18–23; Salt Lake City, UT, USA. p. 9185–93. doi:10.1109/CVPR.2018.00957.
17. Bai S, Li Y, Zhou Y, Li Q, Torr PHS. Adversarial metric attack and defense for person re-identification. IEEE Trans Pattern Anal Mach Intell. 2020;43(6):2119–26. doi:10.1109/tpami.2020.3031625.
18. Goodfellow IJ, Shlens J, Szegedy C. Explaining and harnessing adversarial examples. arXiv:1412.6572. 2014.
19. Zheng Y, Lu Y, Velipasalar S. An effective adversarial attack on person re-identification in video surveillance via dispersion reduction. IEEE Access. 2020;8:183891–902. doi:10.1109/ACCESS.2020.3024149.
20. de O Andrade E, Sampaio IGB, Guérin J, Viterbo J. Combining two adversarial attacks against person re-identification systems. arXiv:2309.13763. 2023.
21. Li CY, Shahin Shamsabadi A, Sanchez-Matilla R, Mazzon R, Cavallaro A. Scene privacy protection. In: ICASSP 2019—2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 2019 May 12–17; Brighton, UK. p. 2502–6. doi:10.1109/icassp.2019.8682225.
22. Gong Y, Huang L, Chen L. Person re-identification method based on color attack and joint defence. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW); 2022 Jun 19–20; New Orleans, LA, USA. p. 4312–21. doi:10.1109/CVPRW56347.2022.00477.
23. Luo C, Lin Q, Xie W, Wu B, Xie J, Shen L. Frequency-driven imperceptible adversarial attack on semantic similarity. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2022 Jun 18–24; New Orleans, LA, USA. p. 15294–303. doi:10.1109/CVPR52688.2022.01488.
24. Zeng Y, Pun CM. Frequency-constrained transferable adversarial attack on image manipulation detection and localization. Vis Comput. 2024;40(7):4817–28. doi:10.1007/s00371-024-03482-4.
25. Deng Y, Karam LJ. Frequency-tuned universal adversarial attacks on texture recognition. IEEE Trans Image Process. 2022;31:5856–68. doi:10.1109/tip.2022.3202366.
26. Ali Z, Moon J, Gillani S, Afzal S, Maqsood M, Rho S. Vision-based approach to knee osteoarthritis and Parkinson's disease detection utilizing human gait patterns. PeerJ Comput Sci. 2025;11(5):e2857. doi:10.7717/peerj-cs.2857.
27. Razzaq K, Shah M. Machine learning and deep learning paradigms: from techniques to practical applications and research frontiers. Computers. 2025;14(3):93. doi:10.3390/computers14030093.
28. ŞAHiN E, Arslan NN, Özdemir D. Unlocking the black box: an in-depth review on interpretability, explainability, and reliability in deep learning. Neural Comput Appl. 2025;37(2):859–965. doi:10.1007/s00521-024-10437-2.
29. Rony J, Hafemann LG, Oliveira LS, Ben Ayed I, Sabourin R, Granger E. Decoupling direction and norm for efficient gradient-based L2 adversarial attacks and defenses. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2019 Jun 15–20; Long Beach, CA, USA.
30. Yao Z, Gholami A, Xu P, Keutzer K, Mahoney MW. Trust region based adversarial attack on neural networks. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2019 Jun 15–20; Long Beach, CA, USA. p. 11342–51. doi:10.1109/CVPR.2019.01161.
31. Madry A, Makelov A, Schmidt L, Tsipras D, Vladu A. Towards deep learning models resistant to adversarial attacks. arXiv:1706.06083. 2017.
32. Brendel W, Rauber J, Bethge M. Decision-based adversarial attacks: reliable attacks against black-box machine learning models. arXiv:1712.04248. 2017.
33. Papernot N, McDaniel P, Jha S, Fredrikson M, Celik ZB, Swami A. The limitations of deep learning in adversarial settings. In: 2016 IEEE European Symposium on Security and Privacy (EuroS&P); 2016 Mar 21–24; Saarbruecken, Germany. p. 372–87. doi:10.1109/EuroSP.2016.36.
34. Wang H, Wang G, Li Y, Zhang D, Lin L. Transferable, controllable, and inconspicuous adversarial attacks on person re-identification with deep mis-ranking. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2020 Jun 13–19; Seattle, WA, USA. p. 339–48. doi:10.1109/cvpr42600.2020.00042.
35. Bouniot Q, Audigier R, Loesch A. Vulnerability of person re-identification models to metric adversarial attacks. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW); 2020 Jun 14–19; Seattle, WA, USA. p. 3450–9. doi:10.1109/cvprw50498.2020.00405.

36. Wang Z, Zheng S, Song M, Wang Q, Rahimpour A, Qi H. advPattern: physical-world attacks on deep person re-identification via adversarially transformable patterns. In: 2019 IEEE/CVF International Conference on Computer Vision (ICCV); 2019 Oct 27–Nov 2; Seoul, Republic of Korea. p. 8340–9. doi:10.1109/iccv.2019.00843.
37. Wang G, Lai JH, Liang W, Wang G. Smoothing adversarial domain attack and P-memory reconsolidation for cross-domain person re-identification. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2020 Jun 13–19; Seattle, WA, USA. p. 10565–74. doi:10.1109/CVPR42600.2020.01058.
38. Kanwal S, Shah JH, Khan MA, Nisa M, Kadry S, Sharif M, et al. Person re-identification using adversarial haze attack and defense: a deep learning framework. Comput Electr Eng. 2021;96:107542. doi:10.1016/j.compeleceng.2021.107542.
39. Chang H, Li Y, Si N, Zhang H. A targeted adversarial attack method against person re-identification model. In: 2021 IEEE 3rd International Conference on Frontiers Technology of Information and Computer (ICFTIC); 2021 Nov 12–14; Greenville, SC, USA. p. 313–6. doi:10.1109/icftic54370.2021.9647396.
40. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016 Jun 27–30; Las Vegas, NV, USA. p. 770–8. doi:10.1109/CVPR.2016.90.
41. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2017 Jul 21–26; Honolulu, HI, USA. p. 2261–9. doi:10.1109/CVPR.2017.243.
42. Wang J, Sun K, Cheng T, Jiang B, Deng C, Zhao Y, et al. Deep high-resolution representation learning for visual recognition. IEEE Trans Pattern Anal Mach Intell. 2020;43(10):3349–64. doi:10.1109/tpami.2020.2983686.
43. Zheng L, Shen L, Tian L, Wang S, Wang J, Tian Q. Scalable person re-identification: a benchmark. In: 2015 IEEE International Conference on Computer Vision (ICCV). 2015 Dec 7–13; Santiago, Chile. p. 1116–24. doi:10.1109/ICCV.2015.133.
44. Singh D, Mathew J, Agarwal M, Govind M. Indoor dataset for Person re-Identification: exploring the impact of backpacks. J Vis Commun Image Represent. 2023;96:103931. doi:10.1016/j.jvcir.2023.103931.
45. Tao Z, Lu Z, Peng J, Wang H. AGS: transferable adversarial attack for person re-identification by adaptive gradient similarity attack. Knowl Based Syst. 2024;304(6):112506. doi:10.1016/j.knosys.2024.112506.
46. Liu D, Wu LY, Hong R, Ge Z, Shen J, Boussaid F, et al. Generative metric learning for adversarially robust open-world person re-identification. ACM Trans Multimedia Comput Commun Appl. 2023;19(1):1–19. doi:10.1145/3522714.
47. Bian Y, Liu M, Yi Y, Wang X, Wang Y. Prompt-driven transferable adversarial attack on person re-identification with attribute-aware textual inversion. arXiv:2502.19697. 2025.
48. Gao J, Jiang X, Dou S, Li D, Miao D, Zhao C. Re-ID-leak: membership inference attacks against person re-identification. Int J Comput Vis. 2024;132(10):4673–87. doi:10.1007/s11263-024-02115-6.
49. Zhao G, Zhang M, Liu J, Wen JR. nsupervised adversarial attacks on deep feature-based retrieval with GAN. arXiv:1907.05793. 2019.