

Secure Gait Recognition-Based Smart Surveillance Systems Against Universal Adversarial Attacks

Maryam Bukhari, Department of Computer Science, COMSATS University Islamabad, Attock, Pakistan


Sadaf Yasmin, Department of Computer Science, COMSATS University Islamabad, Attock, Pakistan

Saira Gillani, Department of Information technology and Computer science, University of Central Punjab, Lahore, Pakistan

Muazzam Maqsood, Department of Computer Science, COMSATS University Islamabad, Attock, Pakistan

Saungmin Rho, Department of Industrial Security, Chung-Ang University, Seoul, South Korea*

Sang Soo Yeo, Department of Computer Engineering, Mokwon University, Daejeon, South Korea

 <https://orcid.org/0000-0002-0224-0150>

ABSTRACT

Currently, the internet of everything (IoE) enabled smart surveillance systems are widely used in various fields to prevent various forms of abnormal behaviors. The authors assess the vulnerability of surveillance systems based on human gait and suggest a defense strategy to secure them. Human gait recognition is a promising biometric technology, but one significantly hindered because of universal adversarial perturbation (UAP) that may trigger system failure. More specifically, in this research study, the authors emphasize on sample convolutional neural network (CNN) model design for gait recognition and assess its susceptibility to UAPs. The authors compute the perturbation as non-targeted UAPs, which trigger a model failure and lead to an inaccurate label to the input sample of a given subject. The findings show that a smart surveillance system based on human gait analysis is susceptible to UAPs, even if the norm of the generated noise is substantially less than the average norm of the images. Later, in the next stage, the authors illustrate a defense mechanism to design a secure surveillance system based on human gait.

KEYWORDS

Adversarial Attacks, Defense Strategies, Gait Recognition, Internet of Everything (IoE), Security and Privacy Concerns, Security of Surveillance Systems, Universal Perturbations

1. INTRODUCTION

The Internet of Everything (IoE) is a phrase in information technology that evolved from the internet of things (IoT) as time has progressed (Kubba & Hoomod, 2019). IoE links numerous items and things over the internet using embedded sensors to gather and analyze data in an intelligent manner

DOI: 10.4018/JDM.318415

*Corresponding Author

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

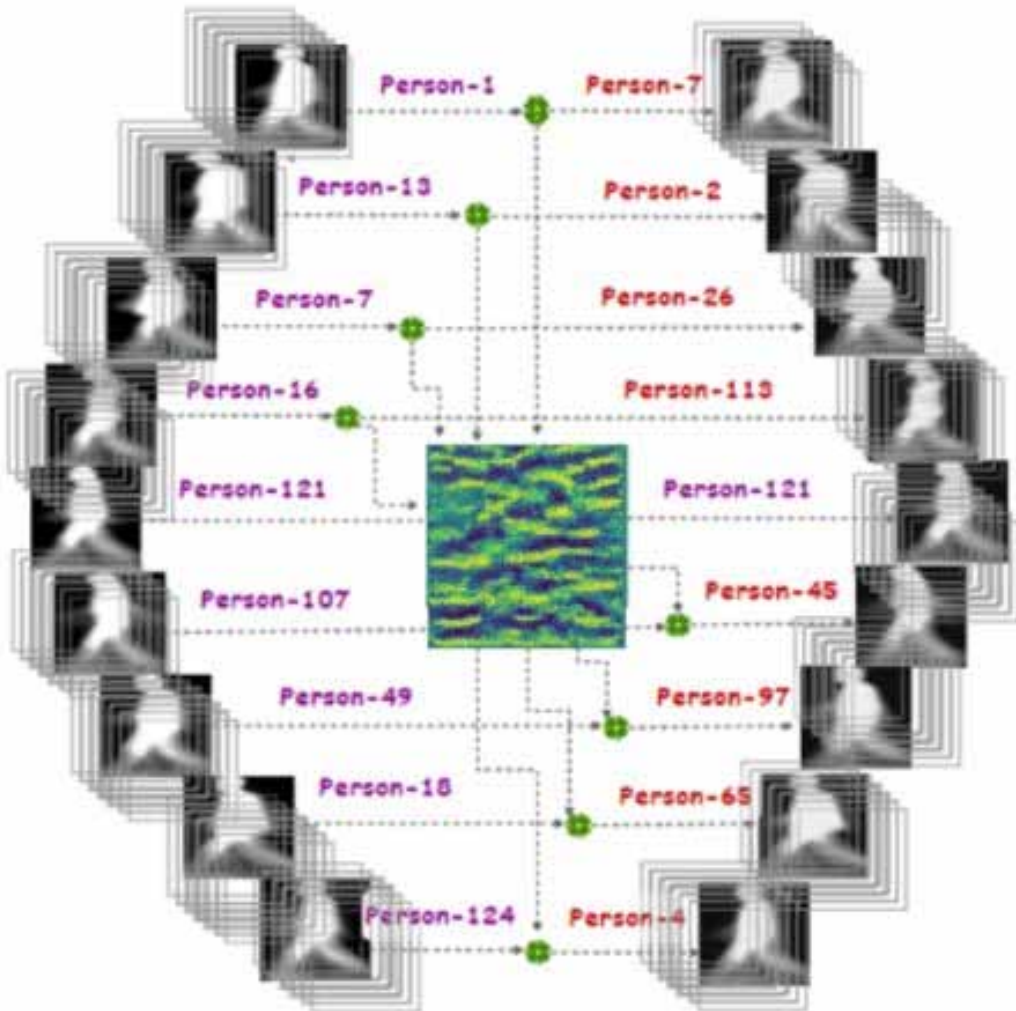
(Thabit, Mahmoud, Alkhayat, & Abbasi, 2019). More specifically, it can gather and interpret real-time data from a multitude of sensors, like as cameras that are linked to it (Miraz, Ali, Excell, & Picking, 2015). As a result, automated systems based on IoE innovation have major features such as real-time computing, detecting, updating, regulating, and observing. Currently, the IoE has emerging applications in several domains such as the healthcare industry (Mardini, Aljawarneh, & Al-Abdi, 2021), surveillance (Maitra, Giri, & Sarkar, 2021), agriculture (Mohapatra & Rath, 2022), and many more. Because of all these extensive capabilities, the IoE-enabled smart surveillance system is a fascinating emerging technology. In such a system, surveillance can be accomplished using various ways, however, human gait-based surveillance is in high demand because of their excellent qualities that make more effective video surveillance systems (Angadi & Nandyal, 2020). In previous years, human identification through gait becomes a very popular area of research. Gait is a kind of behavioral biometric in which a person's manner of walking is considered to identify them.

It can be considered as a next-generation approach to biometric systems due to its wide applications in surveillance systems (Alsaggaf et al., 2021; Ran, Zheng, Chellappa, & Strat, 2010). Gait recognition methods are broadly divided into two main groups which are model-based and appearance-based or model-free approaches (BenAbdelkader, Cutler, & Davis, 2002; Yang, Larsen, Alkjær, Simonsen, & Lynnerup, 2014). Moreover, CNNs are the most commonly used algorithm in appearance-based methods with remarkable performances (Alotaibi & Mahmood, 2017; Hawas, El-Khobby, Abd-Elnaby, & Abd El-Samie, 2019; Linda, Themozhi, & Bandi, 2020). Usually, in appearance-based methods, Gait Energy Image (GEI) (Han & Bhanu, 2005) is the most commonly used gait representation.

At the present era, the role of artificial intelligence (AI) and its subfields is increasing rapidly in different applications areas (Y. Guo et al., 2021; Jiao, Wu, Bie, Umek, & Kos, 2018; Lyu & Liu, 2021; Xue, Jiang, Zhang, & Hu, 2021). The majority of the frameworks underlying IoE-enabled technologies are built on powerful Artificial Intelligence algorithms (AI). Hence, there are different application areas in which deep learning neural networks show stunning performances. (Galvez, Bandala, Dadios, Vicerra, & Maningo, 2018) (Krizhevsky, Sutskever, & Hinton, 2012) (De Brabandere, Neven, & Van Gool, 2017; F. Liu, Shen, & Lin, 2015; Maqsood, Bukhari, et al., 2021; Maqsood, Yasmin, Mehmood, Bukhari, & Kim, 2021; W. Sun & Wang, 2018; Tian et al., 2020), etc. In comparison with custom-made feature extraction methods, the Convolutional Neural networks (CNNs) works well because of automatically learned features from the images without the supervision of humans. For any perceptual task, CNNs give a complete automated and end-to-end solution. However, these models are vulnerable to adversarial attacks and hence many IoE-enabled smart surveillance systems are limited and insecure. The Internet of Things (IoT) themes however need to gather and distribute data using low-cost gadgets such as cameras. The main challenge they encounter is implementing defense mechanisms against security concerns and other optimum techniques on these systems (Sadkhan & Hamza, 2017).

In addition to the above, some previous recent studies show the vulnerability of these CNNs which is an extensively adopted approach in gait-based surveillance system (Goodfellow, Shlens, & Szegedy, 2014; Szegedy et al., 2013). They can be fooled easily by only adding a noise vector of minimum magnitude. Adversaries may readily target open-source software, such as CNN-based gait recognition systems because they have accessibility to the parameters of the model as well as training dataset; therefore, it is critical to assess the trustworthiness and safety of CNNs against adversarial attacks. Since they are input image reliant, such adversarial attacks would be less beneficial to adversaries i.e. a separate adversarial perturbation is needed to craft such that a given input image is misclassified by the underlying model. In recent decades, nevertheless, more plausible adversarial attacks have been presented. Interestingly, there is a specific perturbation (named universal adversarial perturbation (UAP) (Moosavi-Dezfooli, Fawzi, Fawzi, & Frossard, 2017) since they are image agnostic) that can cause CNN inability throughout many image recognition tasks. The pictorial representation of computed UAPs against the gait recognition model is shown in Figure 1. UAPs are difficult to identify since they are quite tiny and so have little effect on data patterns. Adversaries in real-world

Figure 1. When Universal perturbation is added to clean GEI image causes the CNNs model to perform misclassification on perturbed GEI with high probability. Left Images: Original GEI images. Central Image: Universal Perturbation. Right Images: Perturbed GEI images. Arrows: On each arrow, the labels of both clean and perturbed images are written



contexts may find it easier to implement UAP-based adversarial attacks. The presence of these adversaries raises concerns about the robustness, generalization, as well as reliability of CNNs and puts all deep-learning-assisted applications at potential risk, and reduces their use in safety-critical domains (Matyasko & Chau, 2018). Therefore, it is very essential to validate the vulnerability of the CNN-based gait model against adversarial attacks especially attacks based on UAPs as this CNN-based gait model is ultimately deployed in surveillance systems. Furthermore, tactics for defending against hostile attacks (i.e., adversarial defense (McAuley & Leskovec, 2012)) are also necessary. In particular, susceptibility is a massive concern in security surveillance systems based on biometrics. Furthermore, of all biometrics, gait recognition is the most advanced biometric since it works with low-resolution images and does not require subject cooperation. Due to these characteristics, it is commonly used for surveillance purposes. In this research, we emphasize the CNN model, which is an

exemplary model for identifying persons based on their gait patterns and we intend to assess CNNs' susceptibility to adversarial attack as a primary objective. In addition, some research studies have exploited the vulnerability of the gait recognition model (Engoor, Selvaraju, Christopher, Guruvayur Suryanarayanan, & Ranganathan, 2020; Prabhu & Whaley, 2017), however, the gait representation used in their work is based on either silhouettes or accelerometric data. In contrary to these studies, this research study employs a more compact representation of gait namely GEI to investigate the vulnerability of CNN as a secondary objective. Furthermore, the adversaries computed in these existing studies are based on the fast gradient sign method (FGSM) and temporal sparse attacks. However, in this study, a more sophisticated perturbation using UAPs is computed to determine the vulnerability of the gait recognition model. It also shows that up to which extent the model is able to generate accurate results if the single perturbation which is computed only once is used to perturb the GEI images of different subjects. Moreover, this research study also suggests a defense mechanism using adversarial training to increase the robustness of the CNN-based gait recognition model against UAPs. The suggested defense protocol reveals that a secure surveillance system based on human gait has been designed. Furthermore, by taking adversarial defense into account; specifically, we assess how much the resilience of the gait recognition algorithm against UAPs improves with adversarial retraining (Carlini & Wagner, 2017a; Moosavi-Dezfooli, Fawzi, Fawzi, Frossard, & Soatto, 2017) (i.e., fine-tuning using antagonistic GEI images). Following are our contributions:

- To the best of our knowledge, we are the first to exploit the vulnerability of the CNN-based gait recognition model with GEI images against universal adversarial perturbations
- Universal adversarial perturbation is intended to successfully mislead the model, and an adversarial defense mechanism is suggested to increase model robustness
- This study shows a critical flaw in adversarial robustness research on CNN-based gait recognition that has been addressed using adversarial training as a defense mechanism

The rest of the paper is organized as: Section-II describes the related work in this field, Section-III presents the proposed methodology, Section-IV reports and explains different results and experiments while the last section includes the conclusion followed by references.

2. RELATED WORK

In this section, we addressed some literature on various types of attacks, followed by existing work on adversarial attacks with gait recognition and in different domains as well as discussing defense mechanisms for these attacks.

Recently, numerous researchers proposed several kinds of adversarial attacks. Further, some kinds of these attacks enable the threats and security vulnerabilities in environments of federated learning (Mothukuri et al., 2021). The term adversarial sample is first introduced by Szegedy et al. (Szegedy et al., 2013) in 2014. These adversarial samples are searched with the help of optimization problems and achieve very good performance on the state-of-the-art deep neural networks, but the computation of these adversarial examples is very expensive. Later on, the extensions of this adversarial attack are introduced namely the Fast gradient sign method (FGSM) proposed by Good fellow et.al. (Goodfellow et al., 2014). Some other variants of FGSM are proposed by Kurakin et al. (Kurakin, Goodfellow, & Bengio, 2016) named as one-step Target Class Method, Basic Iterative Method (BIM), and Iterative Least-Likely Class Method (ILCM). All these methods are classified under the group of gradient-based methods, however, there also exist some other methods such as Papernot et al. (Papernot, McDaniel, Jha, et al., 2016) designed another method of generating adversarial examples called Jacobian Saliency Map Attack (JSMA).

The adversarial examples are generated by obtaining the saliency maps with the help of computing the forward Jacobian matrix of the model. Carlini and Wagner (Carlini & Wagner, 2017b) further extend the idea of Papernot et al. (Papernot, McDaniel, Jha, et al., 2016) and propose another method of generating adversarial examples. Their method defeats the defense method of defensive distillation against these attacks. To achieve the transferability of adversarial examples across different architectures of deep CNN, Liu et al. (Y. Liu, Chen, Liu, & Song, 2016) designed the Model-based Ensembling Attack. This research shows that the transferability of targeted adversarial examples is difficult to achieve. Furthermore, attack based on optimization approaches is also proposed by Su et al. (Su, Vargas, & Sakurai, 2019) as well as Chen et al. (P.-Y. Chen, Zhang, Sharma, Yi, & Hsieh, 2017).

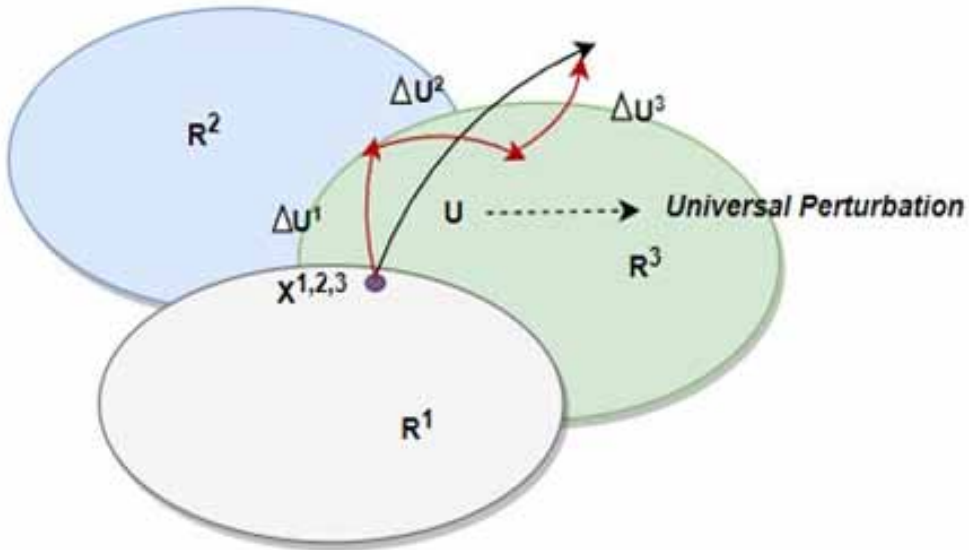
In addition, these adversarial attacks have been used to investigate the vulnerabilities associated with several state-of-the-art deep learning systems. For instance, in the work of Zhu et al. (Zhu, Lu, & Chiang, 2019), it is observed that applying different effects of makeup to images of faces in form of perturbation can fool the face recognition model. Furthermore, different IT companies like Google, Tesla, and Uber, etc. are using deep learning-based methods in their projects such as self-driving cars. Hence, to fool a real such system, Nir et al. utilize perturbations in the form of traffic signs (Morgulis, Kreines, Mendelowitz, & Weisglass, 2019). Further, Simen et al. (Thys, Van Ranst, & Goedemé, 2019) design adversarial patches to fool automated surveillance cameras. In their work, perturbation in form of a patch is capable of successfully hiding people from a person on the detector. Furthermore, some researchers have also exploited the vulnerabilities of deep CNN models used in medical imaging domains. For instance, Kotia et al. (Kotia, Kotwal, & Bharti, 2019) determine the robustness of the CNN model designed for brain tumor detection based on input MRI images. In addition, some researchers have also exploited the vulnerabilities of these CNN models in the domain of Natural language processing (NLP). Zhou et al. (Zhou, Guan, Bhat, & Hsu, 2019) demonstrate the vulnerability of the NLP model designed to perform categorization among real and fake news.

Furthermore, specifically to the problem of gait recognition, there exists very little amount of research done to investigate the vulnerability of deep learning-based gait recognition systems. For instance, a temporal sparse adversarial attack is designed by He et al. (He, Wang, Dong, & Tan, 2020) to fool a sequence-based gait recognition system. It is indicated in their research that sequence-based gait recognition is highly vulnerable to adversarial attack. Similarly, Prabhu et al. (Prabhu & Whaley, 2017) exploit the vulnerability of a gait recognition system employing one-dimensional CNN by perturbing gait patterns acquires from the accelerometer. The perturbation is computed using the FGSM method in their study and shows significant results in lowering the accuracy of the underlying model. However, in this research study, we have attempted to determine the vulnerability of gait recognition systems using more practical perturbations called UAPs. A generic representation of computing UAPs is shown in Figure 2. (Moosavi-Dezfooli, Fawzi, Fawzi, & Frossard, 2017). Further, this study is different in such aspect that we attempted to exploit the vulnerability of gait recognition system in which human gait features are expressed in GEI images which are more compact representations of gait than sequences-based representations such as silhouettes. To investigate the vulnerability of such a deep learning model which is trained on more compact gait representation patterns is a major research question in this study.

Moreover, to lessen the impact of adversarial attacks in various systems several countermeasures are proposed by various researchers (Yuan, He, Zhu, & Li, 2019). Defensive distillation is one of the commonly used measures to prevent the system from these adversarial attacks (Papernot, McDaniel, Wu, Jha, & Swami, 2016). Similarly, adversarial training is another popular method in this regard in which intelligent deep learning models are trained with adversarial samples (Huang, Xu, Schuurmans, & Szepesvári, 2015).

These defense methods act as safety mechanism against adversarial attacks. However, many research studies investigated that these defense mechanisms failed to identify adversarial samples when some minor changes are launched in existing original attacks (Carlini & Wagner, 2016, 2017a). In the context of different automated systems, several researchers propose a defense mechanism e-g

Figure 2. Adversarial attack on a gait-based smart surveillance system



sun et al. (Q. Sun, Rao, Yao, Yu, & Hu, 2020) proposed novel defense method against adversarial attacks of the driving systems. Siddiqui et al. (Siddiqui & Boukerche, 2020) design a lightweight defense mechanism namely Symmetric Image-Half Flip and Replace (SIHFR) against patch-based adversarial attacks for automated surveillance systems. Moreover, in the domain of IoT, Ahmed et al. propose a generative ensemble learning for the detection of malware (Ahmed, Lin, & Srivastava, 2021a). Their proposed ensemble model develops a collaborative categorization outcome that is resistant to adversarial attack. A simple and dependable authentication protocol is proposed by Wang et al. to secure the data exchange on cloud servers using wireless medical sensors-based networks (W. Wang et al., 2021). Their proposed protocol is based on the block chain and PUF technology. Furthermore, Ahmed et al. (Ahmed, Lin, & Srivastava, 2021b) propose a defense mechanism using deep reinforcement learning to secure the important information exchange over Vehicle Adhoc Networks. All of these mentioned studies point to the optimum protection mechanism for attacks in various domains. Hence for this problem, we suggest adversarial retraining as a defense strategy to improve the robustness of the gait recognition model.

Some more recent research on executing adversarial attacks includes the work of Furkan et al. (Mumcu, Doshi, & Yilmaz, 2022) in which they design an attack for a video anomaly detection model. Wang et al. (Y. Wang et al., 2021) proposed a physical-world-based adversarial patch to fool the object detection model. The object detection model they used includes YoLoV2 and YoloV3 respectively. On the other hand, Siddiqui et al. (Siddiqui & Boukerche, 2021) design the defense mechanism for these patch-based adversarial attacks against the vehicle make and model recognition-based systems. Sun et al. (Y. Sun & Wang, 2022) design the presentation attack for Palmprint recognition-based biometric systems. Their findings indicate the high success rate in fooling the systems. Hemant et al. (Rathore, Sahay, Nikam, & Sewak, 2021) designed the Q-learning-based defense mechanism for the recognition of malware in android. For the deep learning model of diabetic retinopathy, Lal et al. (Lal et al., 2021) design the adversarial attack based on adversarial training. The resulting perturbations are added in retinal fundus images to fool the model. Likewise, Thomas et al. (Hickling, Aouf, & Spencer, 2022) designed the explainable deep reinforcement learning-based defense mechanism for the identification of adversarial attacks.

3. METHODOLOGY

As previously stated, deep learning-based gait recognition outperforms in subject identification and may be widely deployed in surveillance systems. In this research, we exploit the vulnerability of CNN proposed by M. Bukhari et al. (Bukhari et al., 2020) for gait recognition and fool it with potential adversarial examples. The main overview of the proposed methodology includes several steps. In the first steps, we first train the designed CNN model on the train set. This train set is comprised of gait data which is first preprocessed before being given as an input. Afterward, we generate the universal perturbations vector and craft the adversarial images by adding that perturbations vector to test set images. Then we load the CNN trained model to determine the class labels of the test set images. The step by step explanation is given below:

3.1 Preprocessing of Gait Data

Before training the deep learning model, we preprocessed the gait data to certain gait representations. The most popular gait representation is Gait Energy Image (GEI). The following equation (1) demonstrates how the GEI images are calculated that are given as an input to CNN:

$$GEI = G(x, y) = \frac{1}{T} \sum_{t=1}^T I(x, y, t) \quad (1)$$

In the above equation, T is representing total silhouettes extracted from video sequences of all persons by background subtraction with x and y coordinates where t denotes the silhouette number. More precisely, all silhouette images are first summed followed by dividing the total number of silhouettes. This is done for every subject in the dataset.

The resulting GEI images are less influenced by the noise factor such as in silhouettes and represent and carry more compact gait features of individuals for their identification purposes. The information about an individual's motion is displayed in dynamic areas (low-intensity areas) of GEI, whereas fixed intensity areas, also known as static areas, reveal information about the body's structure.

3.2. CNN Architecture

The CNN architecture proposed by M. Bukhari et al. shows very remarkable performance in classifying the individuals(Bukhari et al., 2020). This CNN consists of a total of ten layers and it is trained on $240 \times 240 \times 1$ GEI images. The architecture is divided into four distinct blocks. In each block, there is a convolution layer of kernel size 3×3 . Afterward, a max-pool layer of window size 2×2 is added to downscale the GEI image. The activation used after every convolutional layer is Leaky ReLu with the value of $\alpha = 0.05$. At each block, the different number of filters are used which are 16,32,64 and 124. In addition, the starting weights of the kernel matrix are initialized with the Xavier method of initialization. After the last max-pool layer in the last block, a fully connected layer is deployed in which the number of neurons is equal to the classes provided in the dataset. The hyper parameters for this CNN include epochs which are set to 30, and 0.0001 is the learning rate of the model with weight optimizer Adam and the batch size of inputs during training is 4. Subsequently, in the second stage, we proposed variant of universal perturbation which are explained below in detail. The pictorial representation of CNN architecture is shown in Figure 3.

3.3. Universal Adversarial Attack

Since Moosavi-Dezfooli et al. (Moosavi-Dezfooli, Fawzi, & Frossard, 2016) identified the UAPs for image classification tasks, their significance has been shown in several fields. For non-targeted attacks, the UAPs are computed using simple and elegant iterative algorithms whose specifics are given in (Moosavi-Dezfooli, Fawzi, Fawzi, Frossard, et al., 2017). In this study, we have employed non-targeted

Figure 3. A pictorial representation of CNN model used to carry out human gait analysis



universal perturbations available in the Adversarial Robustness 360 Toolbox (ART)(Nicolae et al., 2018). In non-targeted UAPs, the major objective is to find such a UAP perturbation that, when used to perturb a GEI image, may lead the model to predict any arbitrary class rather than the actual class. The algorithm takes into account a classifier $C(x)$ that yields the class or label ID of a subject along with the best confidence score when the GEI image x is provided as input. At the initial stage of the algorithm, the UAP perturbation $\rho = 0$ indicates no perturbation, and after some iterations this perturbation is gradually changed and updated under the limit i-e the L_p norm of this perturbation is comparable to or less than a minimal ξ value as given by equation (2):

$$\rho_p \leq \xi \quad (2)$$

In the above equation (2), ρ denotes the perturbation while ρ_p denotes the norm of a perturbation. Further, this process iteratively builds the adversarial perturbations for the GEI image x provided at input, which is purposively chosen from collections of GEI images of all subjects. These repeated adjustments proceed till the total number of iterations is reached i-e i_{max} . Moreover, for each GEI image, we have employed the fast gradient sign method (FGSM) (Goodfellow et al., 2014) method to compute the universal perturbations rather than the traditional UAP algorithm which employs the DeepFool technique (Moosavi-Dezfooli, Fawzi, Fawzi, & Frossard, 2017; Moosavi-Dezfooli et al., 2016). The reason for choosing this method is that its computational complexity is much lower than DeepFool. Moreover, the FGSM method computes the adversarial perturbation $\hat{\rho}$ for GEI image x by taking the gradient $\nabla_x L(x, y)$ of cost function also called loss function at the GEI image x and subject label y with regard to pixel values of the image. For the norm i-e L_∞ the non-targeted perturbation that induces misclassification is calculated by equation (3):

$$\hat{\rho} = \epsilon \cdot \text{sign}(\nabla_x L(x, C(x))) \quad (3)$$

In the above equation (3), the value of $\epsilon > 0$ indicates the power of an attack or denotes magnitude of the perturbation. The term $\nabla_x L(x, C(x))$ denotes the gradient ∇ of the cost function or loss function with respect to GEI image x and the actual label of that image provided by the classifier

C . More specifically, for both norms called L_1 and L_2 norms, the adversarial perturbation is calculated using equations (4):

$$\hat{\rho} = \in .sign(\nabla_x L(x, C(x))) / \nabla_x L(x, C(x))_p \quad (4)$$

In the above scenario, the FGSM method is carried out on the outcome $C(x + \rho)$ of the CNN model or classifier for the perturbed GEI $x + \rho$, at every step of iteration. For non-targeted adversarial attacks, the perturbation $\hat{\rho}$ for $x + \rho$ is generated by employing FGSM if $C(x + \rho) = C(x) \cdot (C(x + \rho) \neq y)$. After computing the adversarial sample, i-e $x_{adv} \leftarrow x + \rho + \hat{\rho}$ at this particular step, the perturbation is modified if $C(x_{adv}) \neq C(x) \cdot (C(x_{adv}) = y)$ for adversarial attacks. To meet the condition, that $\rho_p \leq \xi : \rho \leftarrow \text{project}(x_{adv} - x, \rho, \xi)$, a projection operation $\text{project}(x, \rho, \xi)$, is employed to modify the ρ while the $\text{project}(x, \rho, \xi) = \arg \min_x \|x - x'\|_2$ subject to $\rho_p \leq \xi$. We also created random vectors (random UAPs) chosen evenly from a sphere of a predetermined radius to evaluate the results of the created UAPs with those of random samples (Moosavi-Dezfooli, Fawzi, Fawzi, & Frossard, 2017).

3.4. Evaluation Criteria

To assess the vulnerability of the gait recognition model towards UAPs, we employed the fooling rate, R_f for non-targeted attacks. The fooling rate R_f is defined as the fraction of GEI images of different subjects in either train or test sets that have not been correctly classified. In addition, we have plotted the confusion matrices for each of the experiments to examine the variation in prediction due to the UAPs.

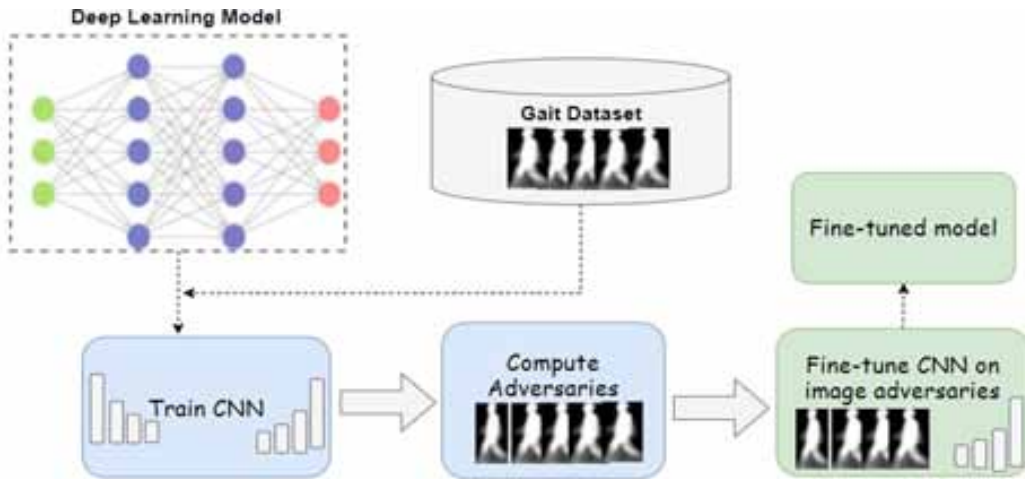
3.5. Adversarial Retraining

It is observed from the experiments that CNN based gait recognition model is vulnerable to adversarial attack. Hence, in order to enhance the robustness of the model, we carried out the adversarial re-training of the gait recognition model (Moosavi-Dezfooli, Fawzi, Fawzi, & Frossard, 2017). More precisely, by the use of adversarial GEI images, we have fine-tuned the gait recognition model according to the approach described in (Moosavi-Dezfooli, Fawzi, Fawzi, & Frossard, 2017) (Carlini & Wagner, 2017a). The major steps of adversarial re-training include that we first computed the different sets of UAPs (i-e 10) using the training set of databases. Later on, we update the original training set that is employed initially to train the CNN by randomly picking half of the training clean GEI images and merging them together with adversarial GEI images. However, each adversarial GEI image is computed using UAP which is randomly chosen from ten generated UAPs. Subsequently, the model has trained again (fine-tuning) on this modified train set, by executing 10 epochs extra. Afterward, we have computed the UAPs again using a train set against the new trained CNN model to validate the vulnerability. The mechanism of adversarial defense is shown in Figure 4.

4. EXPERIMENTS AND DISCUSSIONS

In this section IV, we will go over the experiments and our findings from all of the algorithms. This section goes into great detail about the vulnerability of gait recognition models. All of the experiments are conducted over the Google Colab with implementation language python.

Figure 4. An adversarial defense to increase the robustness of gait recognition model



4.1. Dataset

The dataset used for experimentation purposes is the CASIA gait dataset which is provided by the Chinese Academy of Sciences. There are three different parts of the dataset namely CASIA-A gait dataset, CASIA-B gait dataset, and CASIA-C. Here we use the CASIA-B gait dataset which is the largest multi-view gait dataset. In this dataset, ten sequences are available for each subject out of which six are those sequences in which subjects are walking in an indoor environment with a normal walk style. The other two sequences of each subject are available in which they are walking with bags. Similarly, two sequences of each subject are available in which they wear different types of coats. The normal sequences are defined with the notation “nm” while bag sequences are defined with “bg” and coat sequences are defined with the notation “cl”. Furthermore, data from 124 different individuals is available in a dataset, which is divided into gallery and probe set in each experiment. In this particular research study, we have employed the normal walking sequences of each subject in the database.

4.2. Performance of Baseline Model

To exploit the tolerance of the CNN-based gait recognition model deployed in the surveillance system towards the attack, we first train the designed CNN model on normal walking sequences of different subject’s i-e 124. At the first stage, the entire data set of 124 individuals is partitioned into train and test sets, with each person having six normal walking sequences. The train set contains the sequences [nm-01 to nm-04], whereas the test set contains the sequences [nm-05 to nm-06]. The experiment is repeated five times and the average test accuracies of the CNN model is 97.61% respectively. The results are shown in Table 1. In addition, the average confidence scores of test sets are also listed. It has been observed that the CNN model works extremely well at distinguishing individuals based on their gait patterns and has a high degree of certainty.

4.3. Vulnerability of Model with Universal Adversarial Attack

The CNN-based deep learning model shows better accuracies in recognizing different persons; however, it is observed that under UAPs, the model performs poorly and is hence deemed vulnerable as shown in Table 2. To compute the UAPs, we employed entire train set images of all 124 subjects present in the database. The parameters of UAP attack includes the noise computation method which is set to FGSM, and the attack is conducted in un-targeted manner hence the parameter of attack type

Table 1. Results of Baseline Model

A	ξ	Universal UAP R_f		Random UAP R_f	
		Test	Train	Test	Train
L_2	8	47%	39%	8%	6%
L_∞	0.06	30%	21%	6%	8%
L_2	10	72%	68%	8%	8%
L_∞	0.08	74%	70%	8%	7%

Table 2. Fooling rates R_f (%) of UAPs against the Gait Recognition Model

CASIA-B Gait Dataset				
Run	Train	Test	Accuracy	Average Confidence
1	Normal	Normal	97.98%	97.98%
2	Normal	Normal	97.58%	98%
3	Normal	Normal	97.92%	98.10%
4	Normal	Normal	97.17%	99.12%
5	Normal	Normal	97.38%	98.45%
Average	Normal	Normal	97.61%	98.33%

is set to un-targeted. Further, the algorithm runs for 15 iterations with norms L_2 and L_∞ with desired accuracy parameter set to 0.000001. After computing the UAPs, they are added to both train and test set to compute values of fooling rates R_f . This measure indicates the percentage of images that are incorrectly classified. More precisely, on the test data the fooling rate R_f with $\xi = 8$ for UAPs using L_2 is about 47%. A greater ξ resulted in increased R_f . It is also indicated that the R_f of the UAPs is about 72% on test set walking sequences for the $\xi = 10$. Similarly, for random UAPs the value of R_f on train and test is about 6% and 8% respectively. This shows that random UAPs have no substantial effect on the accuracy of the model as compared to universal UAPs. Similarly, with norm L_∞ the R_f is about 30% and 74% on the test set with $\xi = 0.06$ and $\xi = 0.08$. In addition, we choose the value of ξ in such a way that the L_2 and L_∞ norm of the resultant UAP does not increase with mean L_2 and L_∞ norm of images in the train set. There is only a little bit of difference among the values of fooling rates R_f on test set walking sequences for both types of norms, while maintaining the same parameters setting for both norms. In the case of fooling rate R_f with random UAPs there is no significant difference was found in L_2 and L_∞ norms-based perturbations.

In addition to the above, we have plotted the confusion matrices after an adversarial attack. The database contains the data of 124 subjects, however, due to space issues we have plotted the confusion matrices for ten persons. The test set contains the two instances of normal walking sequences for each person comprising 248 GEI samples. Since the GEI is computed for each video sequence.

Figure 5(left) shows the confusion matrix of ten person for the test set whose samples are perturbed with UAPs with the L_∞ norm. It is observed that GEI images of different persons are wrongly classified. For instance, both two instances of person Id, 3,4,5,7, and 8 are wrongly classified to some arbitrary classes. Similarly, Figure 5 (right) shows the confusion matrix of ten person for the test set whose samples are perturbed with UAPs under norm L_2 . It has been noticed that GEI images of various individuals are incorrectly labeled. For instance, both test samples of person 1, 2,4,5,6, and 9 are incorrectly classified by the model. Moreover, Figure 6 (a) and (b) shows the resulting perturbations with different norms and their corresponding adversarial images computed using these perturbations. In Figure 6(b) Row 1 corresponds to adversarial images of person ID-01 while rows 2 shows the adversarial images of person ID-002. It is observed from Figure 6 that the resulting images are more seems similar to the original images. The contextual and shape features of a person present in the image are not disrupted. Hence, it is concluded that the underlying CNN model is vulnerable even if the perturbations are less noticeable. Moreover, it is also observed that by increasing the values of ξ the magnitude of noise becomes stronger and hence visible in the images. But on the other hand, if the value of ξ increases the fooling rate also increases. Furthermore, it is required to convey how confident the CNN model is in taking wrong decisions i-e predicting the subject's label by presenting an adversarial GEI image. The summary of confidence scores over complete adversarial test set images is shown in Figure 7. The first two box-plot-based curves in Figure 7 show the trend of confidence scores over the complete test set images using both norms. It is observed that the model is about 60-85% confident while making wrong predictions. More precisely, the y-axis of the plot in Figure 7 indicates the confidence scores and as shown in Figure that area of the first two box-plots are lies in the range 60-85% which means that on most of the test samples the confidence scores of the model are in the range of 60-85%.

Figure 5. Confusion matrices for the gait recognition model attacked using the non-targeted UAPs on the test images with L_2 and L_∞ norm

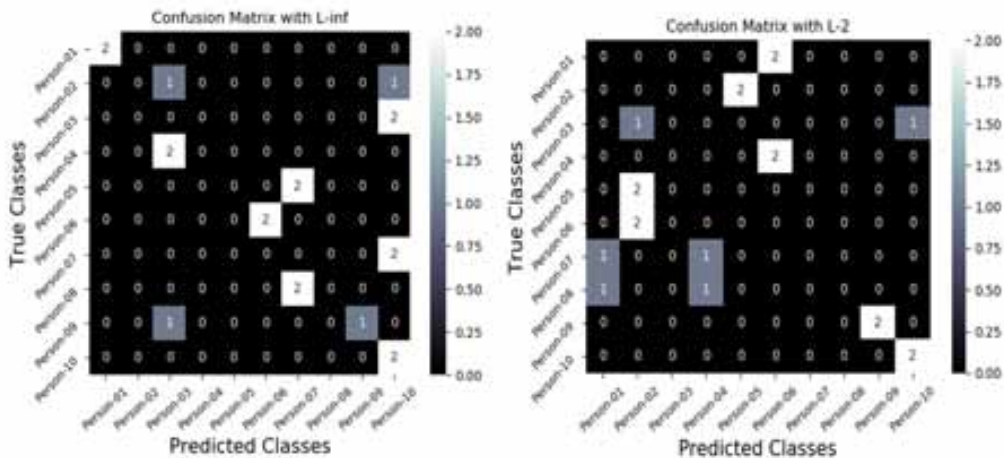
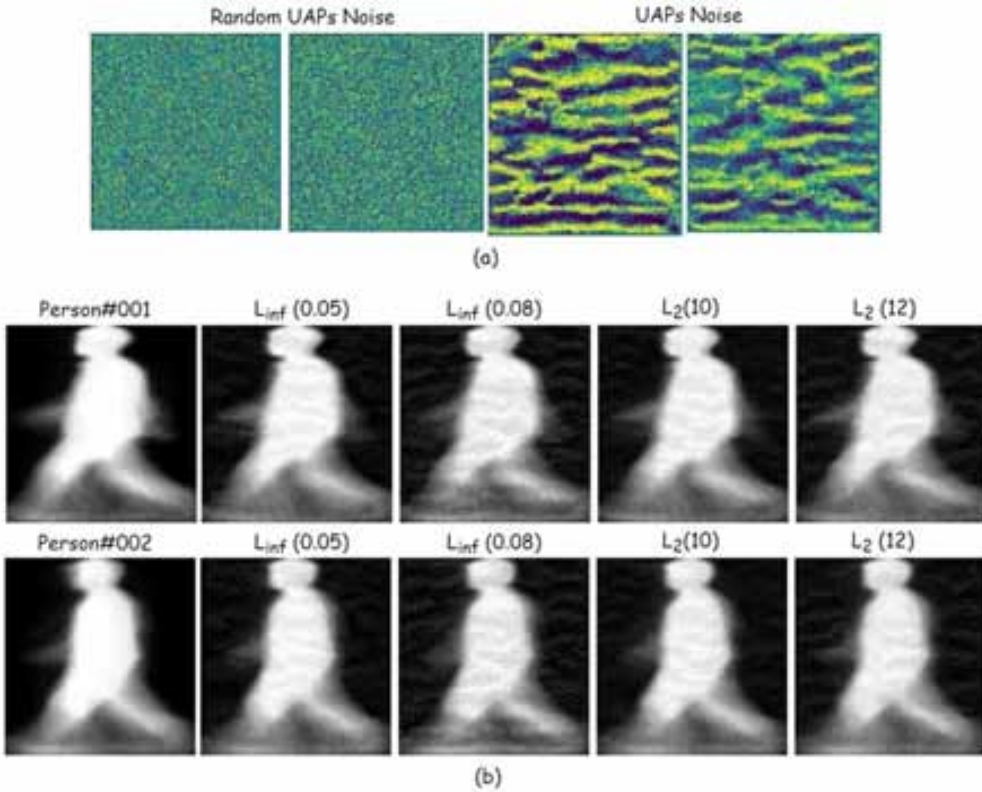


Figure 6. UAPs against gait recognition model and their corresponding adversarial images for two different persons with different norms and magnitudes



4.4. Impact of Adversarial Training to Mitigate the Adversarial Attack

To counteract the effect of the adversarial attacks, the adversarial training, adversarial training is frequently utilized approach. In this research study, we first investigated the vulnerability of the gait recognition model, and at a later stage, we examined that at how much the defense mechanism of adversarial re-training increases the resilience of the gait recognition model, against the UAP attack. This defense mechanism of adversarial training did not have any impact on the test set, especially, the performance accuracy on clean GEI images held steady around 97.98%. We have performed the adversarial training against UAPs computed using different types of norms. For adversarial images computed using non-targeted UAPs computed using norm L_2 with $\xi = 10$, it is observed that, fooling rate R_f is decreased progressively. This experiment is conducted with the data of all 124 subjects but confusion matrices with 10 subjects are depicted in Figure 9, which indicates that the model is now performing correct predictions even if samples are perturbed with UAPs. Hence, it is logically reasonable that adversarial training-based defense mechanisms assist to increase the robustness of the model.

Furthermore, if we test the robustness of fine-tuned in terms of its confidence over adversarial images, we can see from Figure 7 that now the model properly classifies the adversarial images with a rate of approximately 90% reaches up to 2% on test set images after several epochs of adversarial re-training as shown in Figure 8. More precisely, the x-axis in Figure 8 shows the epochs while the blue curves indicates the accuracy on test set and orange curves indicates the decrease in fooling rates

Figure 7. Confidence of model on the wrong prediction over the test set before and after adversarial training

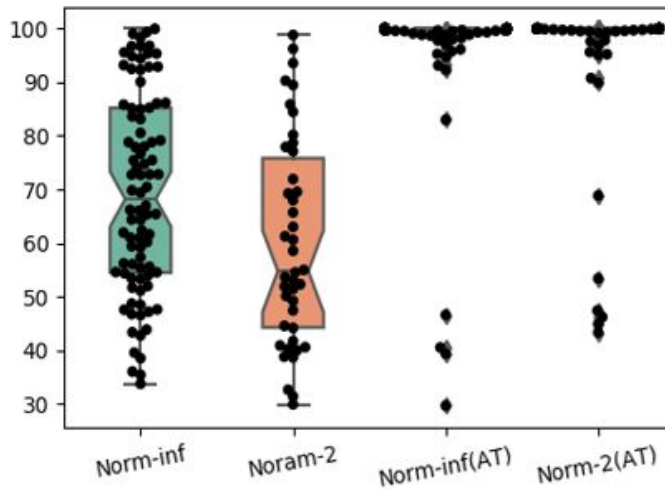
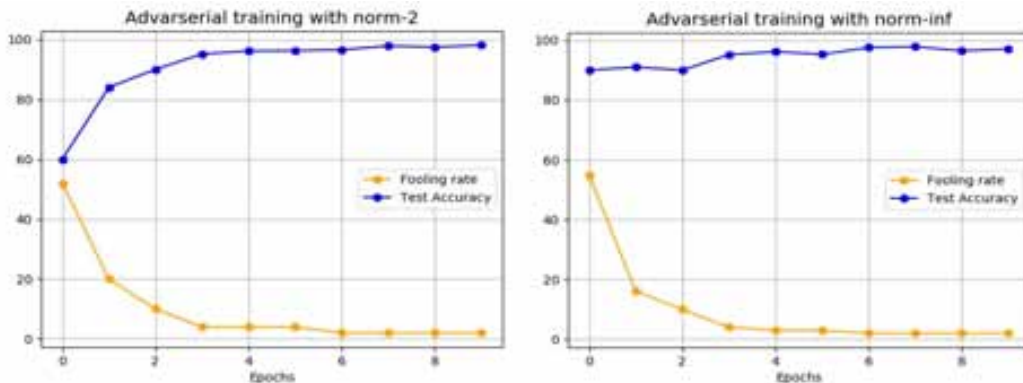


Figure 8. The impact of adversarial retraining on UAP resilience with L_2 and L_∞ norm



over several epochs. Furthermore, with the norm L_∞ the fooling rates R_f is also decreased up to 2% over several epochs of adversarial re-training. After adversarial training, we again computed the UAPs to evaluate the robustness of the model which is fine-tuned over the modified train set. The results of the fine-tuned model against both universal and random UAPs are shown in Table 3. It is observed that R_f values are very low which shows significant robustness of the model against UAPs.

4.5. Discussions

It is clear from the analysis of the above results that although CNN-based gait recognition shows impressive results in the classification of persons, but if we look at the opposite side of CNN's then there exist security risks against these models. CNNs performance drops if the input samples are perturbed with minimal noise. In addition, when the model is deployed in real-world (and possibly hostile) situations, the presence of these perturbations can be used by adversaries to break the model. Moreover, adversaries can cause CNN-based solutions to underperform at a reduced cost (i.e., with a

Figure 9. Confusion matrices for the gait recognition model attacked using the non-targeted UAPs on the test images with L_2 and L_∞ norm after adversarial training

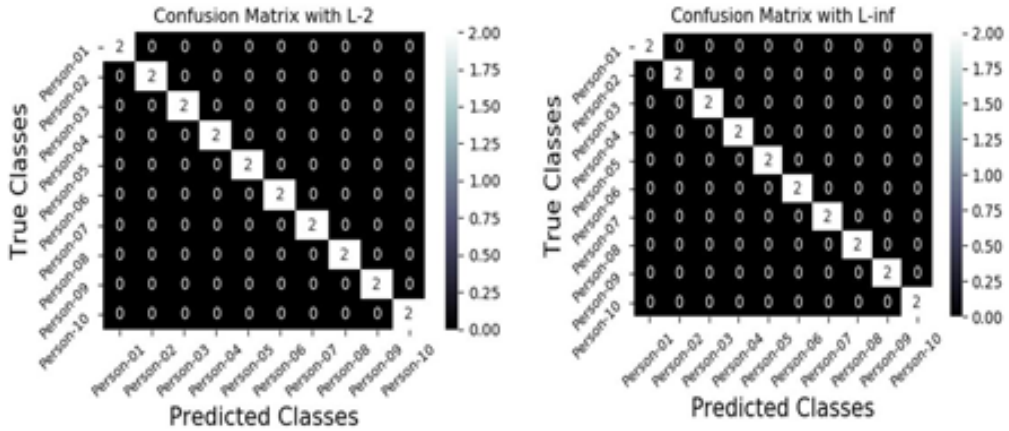


Table 3. Fooling rates R_f (%) of UAPs against the Gait Recognition Model after adversarial Training

ρ	ξ	Universal UAP R_f		Random UAP R_f	
		Train	Test	Train	Test
L_2	8	2%	2%	1.5%	1.8%
L_∞	0.06	2.4%	2.0%	2%	2%
L_2	10	1.9%	2%	1.3%	1.45%
L_∞	0.08	2%	2%	2%	2.3%

singular perturbation); especially, while targeting CNN’s employing UAPs, they don’t have to assess the distribution and variability of input GEI images, because UPAs are image independent. Given the fact that the vulnerability of these CNN against UAPs has been exploited in many use-cases, hence, it is hypothesized that they will exist uniformly in CNN-based models are designed for person identification through gait (Moosavi-Dezfooli, Fawzi, Fawzi, & Frossard, 2017). In addition, it is also observed in various experiments that when UAPs are added to clean GEI images and given as an input to the model then the model performs incorrect classification over those perturbed samples with some specific arbitrary classes. This finding is in accordance with CNN models’ inclination to categorize most input data into some distinct categories due to non-targeted UAPs, — for example, the presence of dominating categories in non-targeted UAP-based attacks. Since the method emphasizes maximizing the fooling rate, a rather high fooling rate is obtained when all GEI images are categorized into some arbitrary specific classes. Hence, it is logical to deduce that security and surveillance based on human gait analysis are at potential risk due to the existence of these perturbations (Rudin,

2019). Furthermore, our first contribution is to demonstrate the vulnerability of the CNN-based gait recognition model by using a more compact representation of gait features that is GEI.

This representation carries more informative features of gait and hence more strongly assists to identify a person based on their gait style. Hence to fool and determine the vulnerability of such a model which is trained on more accurate features of gait is a major research question. The findings shown above show that even when a model is trained using GEI images, it is subject to adversarial attack. Similarly, our second contribution is that we have designed the UAPs based adversarial attack to demonstrate the vulnerability gait recognition model. As in the original method, the perturbation is computed using the deep fool method, however, in the proposed study, we have utilized the fast gradient sign method to compute this perturbation. This is due to the reason that the deep fool method is computationally expensive as it performs successive iterations to compute the perturbation. On the other hand, the FGSM method is less costly and computes the perturbation in one step. This proposed new variant of attack based on FGSM based universal perturbations is less costly and good enough to demonstrate the vulnerability of the gait recognition model. Further, this variant namely universal adversarial attack based on FGSM based perturbations can also be extended to be applied in other domains to demonstrate the vulnerabilities of deep learning models such as in the domain of medical imaging. Further, the universal perturbations are more practical and can significantly play a role in security risks in these systems. Hence, it is necessary to first check the vulnerability of the model so that the weakness of these models is highlighted and overcome. Hence, the proposed study utilizes the more advanced version of adversarial attacks namely universal perturbation and modifies the actual algorithm by replacing the noise computation method with the less costly method. From the perspective of the application, this research study contributes to demonstrate the vulnerabilities of most evolving biometric technology namely gait recognition which can be used in video surveillance systems.

After highlighting the vulnerabilities of the gait recognition model, we have also presented a mechanism to secure the model to increase its robustness. For this purpose, we have performed fine-tuning of a model for ten additional epochs with adversarial GEI images computed using UAPs. The resulting fine-tuned model is more robust and accurate against the UAPs as it strongly mitigates the impact of the adversarial attack. Thus, it is concluded from this research study that a secure gait recognition model can be deployed safely in biometric-based video surveillance systems if the learning of the model is improved with certain defense mechanisms. An adversarial training proved to be useful to secure the gait recognition model against the UAPs. The major flaw in adversarial robustness research of deep learning-based gait assisted video surveillance systems is demonstrated and as a motivation, we also suggest a defense mechanism. Moreover, the study also develops a motivation for different researchers that strict adherence is necessary for actual applications of CNNs to gait recognition, particularly ways to overcome known vulnerabilities.

Advanced computer vision algorithms, like CNNs, are already employed for high-stakes intelligent decisions in security and surveillance nonetheless, they have the ability to offer devastating damage to security systems since they are frequently complicated to understand. In addition, the UAPs based attacks are white-box attacks which means that attackers have accessibility to parameters of the model i-e in this context the attacker has accessibility to the gradient of cost function as well as a training set, consequently, they pose a potential risk for open source software's e-g person identification through gait. Hence, to prevent these systems from adversarial attacks, a very basic solution is to make them closed source and inaccessible to the public. In addition, another way is to think of systems that are black-box i-e closed application programming interfaces (APIs) that allow only input queries and provide outputs. These closed APIs are preferable since they are less accessible to the public. APIs, on the other hand, may be susceptible to adversarial attacks. The reason behind this is that UAPs are generalized perturbations, and perturbation computed using one CNN can able to fool another CNN model. Hence, it is possible to compute UAPs as a white-box attack, to fool the black-box-based CNN system. Moreover, there exist many approaches to conduct black-box adversarial attacks, in

which perturbations are computed using only the outcomes of the model such as confidence scores (J. Chen, Su, Shen, Xiong, & Zheng, 2019; Co, Muñoz-González, de Maupeou, & Lupu, 2019; C. Guo, Gardner, You, Wilson, & Weinberger, 2019). As a result, defensive tactics for adversarial attacks should be established. Fine-tuning of CNN models on adversarial images is one of the straightforward defensive methods. Indeed, we have analyzed that fine-tuning of gait recognition model on 10 extra epochs using UAPs increased the robustness of the gait model to adversarial attack using UAPs. On the other hand, in some cases, this repetitive strategy of fine-tuning has large computational complexity, and it did not accurately prevent susceptibility to UAPs. Furthermore, different research studies have been suggested for breaching the defense mechanism of adversarial retraining (Carlini & Wagner, 2017a). Principal component analysis (PCA) based dimensionality reduction, distributional, and normalization recognition might be helpful for a defense mechanism, nevertheless, it is very difficult to detect adversarial samples using these approaches (Carlini & Wagner, 2017a). Preventing different systems against adversarial attacks is a game of cat-and-mouse (Finlayson, Chung, Kohane, & Beam, 2018), therefore, it might be challenging to completely reduce the potential risks deduced by these adversarial attacks. On the other hand, the techniques to prevent these attacks have been improved. For instance, at densely distributed input samples, recognizing adversarial attack-based resilience to random noise (Yu, Hu, Guo, Chao, & Weinberger, 2019), a discontinuous activation function is employed that intentionally negates the gradients of the CNN (Xiao, Zhong, & Zheng, 2019) and CNN's for cleaning data samples could help mitigate some of the considerations (Hwang, Park, Jang, Yoon, & Cho, 2019)

In the existing literature, the vulnerability of different systems has been investigated using different types of attacks. All these systems are designed using deep learning-based methods. Table 4 and Table 5 provides a basic comparison of the vulnerability of various systems employing various adversarial attacks. For instance, in the domain of healthcare application, Cheng et al.(Cheng & Ji, 2020) exploit the vulnerability of the CNN model which performs tumor detection using brain MRIs. They have also employed the universal adversarial perturbations to create adversarial MRIs to fool CNN. Similarly, in recommender systems, Tommaso et al.(Di Noia, Malitesta, & Merra, 2020) employ targeted adversarial attack to fool it. In this attack, the behavior of the recommender model is disrupted to recommend the least related items to users. For face recognition, Dong et al.(Dong et al., 2019) exploit the vulnerability using a decision-based black box attack. The perturbations in their attack are designed using only outputs of the model by querying different inputs without accessing the information of model gradients. Similarly, in sequence data, Fazle et al. (Karim, Majumdar, & Darabi, 2020) employ the adversarial transformation networks to generate the adversaries to fool deep learning assisted time series classification model. Zhang et al. (Zhang, Zhou, & Li, 2020) employ contextual adversarial attack to fool the object detection model. Their suggested approach can disrupt the image's contextual features and severely lower the mean average precision (mAP) and recall scores. Moreover, in the context of human gait recognition, which is used as a surveillance system, there exist some research studies that have exploited the vulnerability of gait recognition. For instance, He et al. (He et al., 2020) suggest the temporal-sparse adversarial attacks for sequence-based gait recognition. In their

Table 4. Comparison of adversarial attacks in gait recognition

Human Gait Recognition				
Authors	Model	Gait Modality	Attack	Defense
He et al.(He et al., 2020)	Gait Model	Silhouettes Images	Temporal sparse adversarial attack	No
Prabu et al.(Prabhu & Whaley, 2017)	1D CNN	Accelerometer based gait data	FGSM attack	No
Proposed	CNN	GEI images	Universal Attack	Yes

Table 5. Comparison with existing methods of adversarial attacks in different domains

Authors	Underlying Model	Attack Name	Domain
Fazle et al.(Karim et al., 2020)	Time series classification model	Adversarial transformation networks	Time series Data
Dong et al. (Dong et al., 2019)	Face recognition CNN model	Decision based black box attack	Face Recognition
Cheng et al.(Cheng & Ji, 2020)	U-Net model	Modified universal perturbations	Brain tumor segmentation
Zhang et al.(Zhang et al., 2020)	Object Detection models	Contextual Adversarial Attacks	Object Detection
Chan et al.(Chan, Zheng, Liu, Tsang, & Yeung, 2021)	Fuzzy Decision trees	Evasion Attack	Machine learning algorithms
Tommaso et al. (Di Noia et al., 2020)	Deep Neural Networks	Targeted adversarial Attacks	Recommender systems
Neekhara et al.(Neekhara et al., 2019)	Deep Neural Networks	Universal Perturbations	Speech Recognition systems
Proposed	Convolutional Neural Networks (CNN)	Universal Perturbations	Gait based surveillance systems

attack, the perturbation is added to silhouettes images of different subjects. These silhouettes images are part of a complete long sequence/video of the subject. The suggested shows good performance to determine the vulnerability of sequence-based gait recognition models. Generally, the silhouettes representation of gait carries less informative features of gait than GEI images. Furthermore, Parbu et al. (Prabhu & Whaley, 2017) employ the FGSM method to disrupt the gait features obtained through accelerometer and have attained very good performance. In comparison with these studies, this study employs a more compact representation of gait namely GEI, and attempt to exploit the vulnerability of the CNN model. The adversaries are computed using a universal adversarial attack. In addition, we also suggest a defense mechanism to increase the robustness of the gait recognition model deployed in IoE enabled smart surveillance systems under adversarial attack.

4.6. Theoretical and practical contributions

Due to the obvious advantages of gait-biometric, gait-based surveillance is most widely utilized presently. The first is that it does not need the subject to collaborate throughout the identification process. Second, low-resolution cameras can readily evaluate human gait. Several researchers have developed gait recognition algorithms due to their impressive features. Out of all of them, gait recognition utilizing deep learning performs the best. However, how far this higher performance is not tested under a more realistic attack, i.e. “Universal adversarial attack”. Therefore, in this research study, the major contribution is to exploit the vulnerability of the gait recognition system based on the deep learning method against a realistic adversarial attack. We generate the perturbations with an adversarial attack and then add them to GEI images. The generated adversarial images are then sent into the deep learning model, to estimate how well we fooled the model. It is observed from the results that the gait-recognition model becomes a fool when it is subjected to an adversarial attack. Secondly, in existing studies they have conducted adversarial attacks by perturbing the gait features present in silhouettes or accelerometer-based features, however, in contrast to them, we have used more effective representation for gait i-e GEI to indicate how well we fool the model that is trained on GEI images. Furthermore, as a solution, we also proposed a defense mechanism based on adversarial training. It is observed that fine-tuning the model on adversarial images can save the model from being fooled again. We have practically proved the vulnerability of the gait-recognition system by first designing

the effective gait recognition model and later on in the next stage we design the adversarial attack followed by designing the defense mechanism as a solution to protect it against adversarial attacks.

5. CONCLUSION

IoE has the potential to improve our daily lives by evolving various biometric-based surveillance systems towards becoming more of a current process in daily lifestyles. However, the vulnerability of these surveillance systems must be exploited and accordingly defense mechanisms should be developed before they can be used in operation. In this paper, we illustrated the vulnerability of the CNN-based gait recognition model used for surveillance purposes to non-targeted UAP-based attacks. This vulnerability has been demonstrated using a more compact representation of gait namely GEI image. Straightforward implementations of CNNs to gait recognition potentially cause issues in security threats at different domains and hence defense is also required to secure the systems. Therefore, we have also suggested the defense protocol to design a secure gait-based smart surveillance system by performing adversarial retraining of the model. Moreover, this work motivates different researchers to think about all security risks associated with gait recognition biometric systems used for automated surveillance and encourages them to design more powerful defense strategies in their systems to make them robust to adversarial attacks before they are practically deployed. This research's future work will entail the development of black-box attacks against gait recognition systems along with defense mechanisms.

CONFLICT OF INTEREST

The authors of this publication declare there is no conflict of interest.

ACKNOWLEDGEMENT

This work was supported in part by the Korea Institute for Advancement of Technology (KIAT) grant funded by the Korea Government (MOTIE), The Competency Development Program for Industry Specialist, under Grant P0008703, and also supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. 2022R1F1A1063134).

FUNDING

This work was supported in part by the Korea Institute for Advancement of Technology (KIAT) grant funded by the Korea Government (MOTIE), The Competency Development Program for Industry Specialist, under Grant P0008703, and also supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. 2022R1F1A1063134).

REFERENCES

- Ahmed, U., Lin, J. C.-W., & Srivastava, G. (2021a). Generative Ensemble Learning for Mitigating Adversarial Malware Detection in IoT. Paper presented at the 2021 *IEEE 29th International Conference on Network Protocols (ICNP)*. doi:10.1109/ICNP52444.2021.9651917
- Ahmed, U., Lin, J. C.-W., & Srivastava, G. (2021b). *Privacy-Preserving Deep Reinforcement Learning in Vehicle AdHoc Networks*. *IEEE Consumer Electronics Magazine*.
- Alotaibi, M., & Mahmood, A. (2017). Improved gait recognition based on specialized deep convolutional neural network. *Computer Vision and Image Understanding*, 164, 103–110. doi:10.1016/j.cviu.2017.10.004
- Alsaggaf, W. A., Mehmood, I., Khairullah, E. F., Alhurajji, S., Sabir, M. F. S., Alghamdi, A. S., & Ahmed, A. (2021). A Smart Surveillance System for Uncooperative Gait Recognition Using Cycle Consistent Generative Adversarial Networks (CCGANs). *Computational Intelligence and Neuroscience*. doi:10.1155/2021/3110416 PMID:34691168
- Angadi, S., & Nandyal, S. (2020). Human identification system based on spatial and temporal features in the video surveillance system. [IJACI]. *International Journal of Ambient Computing and Intelligence*, 11(3), 1–21. doi:10.4018/IJACI.2020070101
- BenAbdelkader, C., Cutler, R., & Davis, L. (2002). Stride and cadence as a biometric in automatic person identification and verification. Paper presented at the *Proceedings of Fifth IEEE international conference on automatic face gesture recognition*. IEEE. doi:10.1109/AFGR.2002.1004182
- Bukhari, M., Bajwa, K. B., Gillani, S., Maqsood, M., Durrani, M. Y., Mehmood, I., & Rho, S. (2020). An Efficient Gait Recognition Method for Known and Unknown Covariate Conditions. *IEEE Access: Practical Innovations, Open Solutions*.
- Carlini, N., & Wagner, D. (2016). Defensive distillation is not robust to adversarial examples. *arXiv:1607.04311*.
- Carlini, N., & Wagner, D. (2017a). Adversarial examples are not easily detected: Bypassing ten detection methods. Paper presented at the *Proceedings of the 10th ACM workshop on artificial intelligence and security*. ACM. doi:10.1145/3128572.3140444
- Carlini, N., & Wagner, D. (2017b). Towards evaluating the robustness of neural networks. Paper presented at the *2017 IEEE symposium on security and privacy (sp)*. IEEE. doi:10.1109/SP.2017.49
- Chan, P. P., Zheng, J., Liu, H., Tsang, E., & Yeung, D. S. (2021). Robustness analysis of classical and fuzzy decision trees under adversarial evasion attack. *Applied Soft Computing*, 107, 107311. doi:10.1016/j.asoc.2021.107311
- Chen, J., Su, M., Shen, S., Xiong, H., & Zheng, H. (2019). POBA-GA: Perturbation optimized black-box adversarial attacks via genetic algorithm. *computers & security*, 85, 89-106.
- Chen, P.-Y., Zhang, H., Sharma, Y., Yi, J., & Hsieh, C.-J. (2017). Zoo: Zeroth order optimization based black-box attacks to deep neural networks without training substitute models. Paper presented at the *Proceedings of the 10th ACM workshop on artificial intelligence and security*. ACM. doi:10.1145/3128572.3140448
- Cheng, G., & Ji, H. (2020). Adversarial Perturbation on MRI Modalities in Brain Tumor Segmentation. *IEEE Access: Practical Innovations, Open Solutions*, 8, 206009–206015. doi:10.1109/ACCESS.2020.3030235
- Co, K. T., Muñoz-González, L., de Maupeou, S., & Lupu, E. C. (2019). Procedural noise adversarial examples for black-box attacks on deep convolutional networks. Paper presented at the *Proceedings of the 2019 ACM SIGSAC conference on computer and communications security*. ACM. doi:10.1145/3319535.3345660
- De Brabandere, B., Neven, D., & Van Gool, L. (2017). Semantic instance segmentation for autonomous driving. Paper presented at the *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. IEEE.
- Di Noia, T., Malitesta, D., & Merra, F. A. (2020). Taamr: Targeted adversarial attack against multimedia recommender systems. Paper presented at the *2020 50th Annual IEEE/IFIP International Conference on Dependable Systems and Networks Workshops (DSN-W)*. IEEE. doi:10.1109/DSN-W50199.2020.00011

- Dong, Y., Su, H., Wu, B., Li, Z., Liu, W., Zhang, T., & Zhu, J. (2019). Efficient decision-based black-box adversarial attacks on face recognition. Paper presented at the *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE. doi:10.1109/CVPR.2019.00790
- Engoor, S., Selvaraju, S., Christopher, H. S., Guruvayur Suryanarayanan, M., & Ranganathan, B. (2020). Effective Emotion Recognition from Partially Occluded Facial Images Using Deep Learning. Paper presented at the *International Conference on Computational Intelligence in Data Science*. Springer. doi:10.1007/978-3-030-63467-4_17
- Finlayson, S. G., Chung, H. W., Kohane, I. S., & Beam, A. L. (2018). Adversarial attacks against medical deep learning systems. *arXiv:1804.05296*.
- Galvez, R. L., Bandala, A. A., Dadios, E. P., Vicerra, R. R. P., & Maningo, J. M. Z. (2018). Object detection using convolutional neural networks. Paper presented at the *TENCON 2018-2018 IEEE Region 10 Conference*. IEEE. doi:10.1109/TENCON.2018.8650517
- Goodfellow, I. J., Shlens, J., & Szegedy, C. (2014). Explaining and harnessing adversarial examples. *arXiv:1412.6572*.
- Guo, C., Gardner, J., You, Y., Wilson, A. G., & Weinberger, K. (2019). Simple black-box adversarial attacks. Paper presented at the *International Conference on Machine Learning*. IEEE.
- Guo, Y., Pan, J.-S., Qiu, C., Xie, F., Luo, H., Shang, H., Liu, Z., & Tan, J. (2021). SinGAN-Based Asteroid Surface Image Generation. [JDM]. *Journal of Database Management*, 32(4), 28–47. doi:10.4018/JDM.2021100103
- Han, J., & Bhanu, B. (2005). Individual recognition using gait energy image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(2), 316–322. doi:10.1109/TPAMI.2006.38 PMID:16468626
- Hawas, A. R., El-Khobby, H. A., Abd-Elnaby, M., & Abd El-Samie, F. E. (2019). Gait identification by convolutional neural networks and optical flow. *Multimedia Tools and Applications*, 78(18), 25873–25888. doi:10.1007/s11042-019-7638-9
- He, Z., Wang, W., Dong, J., & Tan, T. (2020). Temporal Sparse Adversarial Attack on Sequence-based Gait Recognition. *arXiv:2002.09674*.
- Hickling, T., Aouf, N., & Spencer, P. (2022). Robust Adversarial Attacks Detection based on Explainable Deep Reinforcement Learning For UAV Guidance and Planning. *arXiv:2206.02670*.
- Huang, R., Xu, B., Schuurmans, D., & Szepesvári, C. (2015). Learning with a strong adversary. *arXiv:1511.03034*.
- Hwang, U., Park, J., Jang, H., Yoon, S., & Cho, N. I. (2019). Puvae: A variational autoencoder to purify adversarial examples. *arXiv:1903.00585*.
- Jiao, L., Wu, H., Bie, R., Umek, A., & Kos, A. (2018). Towards real-time multi-sensor golf swing classification using deep CNNs. [JDM]. *Journal of Database Management*, 29(3), 17–42. doi:10.4018/JDM.2018070102
- Karim, F., Majumdar, S., & Darabi, H. (2020). Adversarial attacks on time series. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. PMID:32286957
- Kotia, J., Kotwal, A., & Bharti, R. (2019). Risk susceptibility of brain tumor classification to adversarial attacks. Paper presented at the *International Conference on Man–Machine Interactions*. Springer.
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25, 1097–1105.
- Kubba, Z. M. J., & Hoomod, H. K. (2019). The Internet of Everything Based Smart Systems: Applications and Challenges. Paper presented at the *2019 1st AL-Noor International Conference for Science and Technology (NICST)*. Semantic Scholar.
- Kurakin, A., Goodfellow, I., & Bengio, S. (2016). Adversarial machine learning at scale. *arXiv:1611.01236*.
- Lal, S., Rehman, S. U., Shah, J. H., Meraj, T., Rauf, H. T., Damaševičius, R., Mohammed, M. A., & Abdulkareem, K. H. (2021). Adversarial attack and defence through adversarial training and feature fusion for diabetic retinopathy recognition. *Sensors (Basel)*, 21(11), 3922. doi:10.3390/s21113922 PMID:34200216

- Linda, G. M., Themozhi, G., & Bandi, S. R. (2020). Color-mapped contour gait image for cross-view gait recognition using deep convolutional neural network. *International Journal of Wavelets, Multiresolution, and Information Processing*, 18(01), 1941012. doi:10.1142/S0219691319410121
- Liu, F., Shen, C., & Lin, G. (2015). Deep convolutional neural fields for depth estimation from a single image. Paper presented at the *Proceedings of the IEEE conference on computer vision and pattern recognition*. IEEE. doi:10.1109/CVPR.2015.7299152
- Liu, Y., Chen, X., Liu, C., & Song, D. (2016). Delving into transferable adversarial examples and black-box attacks. *arXiv:1611.02770*.
- Lyu, S., & Liu, J. (2021). Convolutional recurrent neural networks for text classification. [JDM]. *Journal of Database Management*, 32(4), 65–82. doi:10.4018/JDM.2021100105
- Maitra, T., Giri, D., & Sarkar, A. (2021). *Security in Critical Communication for Mobile Edge Computing Based IoT Applications Mobile Edge Computing*. Springer.
- Maqsood, M., Bukhari, M., Ali, Z., Gillani, S., Mehmood, I., Rho, S., & Jung, Y. (2021). A Residual-Learning-Based Multi-Scale Parallel-Convolutions-Assisted Efficient CAD System for Liver Tumor Detection. *Mathematics*, 9(10), 1133. doi:10.3390/math9101133
- Maqsood, M., Yasmin, S., Mehmood, I., Bukhari, M., & Kim, M. (2021). An efficient DA-Net architecture for lung nodule segmentation. *Mathematics*, 9(13), 1457. doi:10.3390/math9131457
- Mardini, W., Aljawarneh, S., & Al-Abdi, A. (2021). Using multiple RPL instances to enhance the performance of new 6G and Internet of Everything (6G/IoE)-based healthcare monitoring systems. *Mobile Networks and Applications*, 26(3), 952–968. doi:10.1007/s11036-020-01662-9
- Matyasko, A., & Chau, L.-P. (2018). Improved network robustness with adversary critic. *arXiv:1810.12576*.
- McAuley, J. J., & Leskovec, J. (2012). *Learning to discover social circles in ego networks*. Paper presented at the *NIPS*.
- Miraz, M. H., Ali, M., Excell, P. S., & Picking, R. (2015). A review on Internet of Things (IoT), Internet of everything (IoE) and Internet of nano things (IoNT). Paper presented at the *2015 Internet Technologies and Applications (ITA)*. IEEE. doi:10.1109/ITechA.2015.7317398
- Mohapatra, H., & Rath, A. K. (2022). IoE based framework for smart agriculture. *Journal of Ambient Intelligence and Humanized Computing*, 13(1), 407–424. doi:10.1007/s12652-021-02908-4
- Moosavi-Dezfooli, S.-M., Fawzi, A., Fawzi, O., & Frossard, P. (2017). *Universal adversarial perturbations*. Paper presented at the *Proceedings of the IEEE conference on computer vision and pattern recognition*.
- Moosavi-Dezfooli, S.-M., Fawzi, A., Fawzi, O., Frossard, P., & Soatto, S. (2017). Analysis of universal adversarial perturbations. *arXiv: 1705.09554*.
- Moosavi-Dezfooli, S.-M., Fawzi, A., & Frossard, P. (2016). Deepfool: a simple and accurate method to fool deep neural networks. Paper presented at the *Proceedings of the IEEE conference on computer vision and pattern recognition*. IEEE. doi:10.1109/CVPR.2016.282
- Morgulis, N., Kreines, A., Mendelowitz, S., & Weisglass, Y. (2019). Fooling a real car with adversarial traffic signs. *arXiv:1907.00374*.
- Mothukuri, V., Parizi, R. M., Pouriyeh, S., Huang, Y., Dehghantanha, A., & Srivastava, G. (2021). A survey on security and privacy of federated learning. *Future Generation Computer Systems*, 115, 619–640. doi:10.1016/j.future.2020.10.007
- Mumcu, F., Doshi, K., & Yilmaz, Y. (2022). *Adversarial Machine Learning Attacks Against Video Anomaly Detection Systems*. Paper presented at the *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE. doi:10.1109/CVPRW56347.2022.00034
- Neekhara, P., Hussain, S., Pandey, P., Dubnov, S., McAuley, J., & Koushanfar, F. (2019). Universal adversarial perturbations for speech recognition systems. *arXiv:1905.03828*. 10.21437/Interspeech.2019-1353

- Nicolae, M.-I., Sinn, M., Tran, M. N., Buesser, B., Rawat, A., Wistuba, M., & Ludwig, H. (2018). Adversarial Robustness Toolbox v1. 0.0. *arXiv:1807.01069*.
- Papernot, N., McDaniel, P., Jha, S., Fredrikson, M., Celik, Z. B., & Swami, A. (2016). The limitations of deep learning in adversarial settings. Paper presented at the *2016 IEEE European symposium on security and privacy (EuroS&P)*. IEEE. doi:10.1109/EuroSP.2016.36
- Papernot, N., McDaniel, P., Wu, X., Jha, S., & Swami, A. (2016). Distillation as a defense to adversarial perturbations against deep neural networks. Paper presented at the *2016 IEEE symposium on security and privacy (SP)*. IEEE. doi:10.1109/SP.2016.41
- Prabhu, V. U., & Whaley, J. (2017). Vulnerability of deep learning-based gait biometric recognition to adversarial perturbations. Paper presented at the *CVPR Workshop on The Bright and Dark Sides of Computer Vision: Challenges and Opportunities for Privacy and Security (CV-COPS 2017)*. CVPR.
- Ran, Y., Zheng, Q., Chellappa, R., & Strat, T. M. (2010). Applications of a simple characterization of human gait in surveillance. *IEEE Transactions on Systems, Man, and Cybernetics. Part B, Cybernetics*, 40(4), 1009–1020. doi:10.1109/TSMCB.2010.2044173 PMID:20363680
- Rathore, H., Sahay, S. K., Nikam, P., & Sewak, M. (2021). Robust android malware detection system against adversarial attacks using q-learning. *Information Systems Frontiers*, 23(4), 867–882. doi:10.1007/s10796-020-10083-8
- Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5), 206–215. doi:10.1038/s42256-019-0048-x PMID:35603010
- Sadkhan, S. B., & Hamza, Z. (2017). Cryptosystems used in IoT-current status and challenges. Paper presented at the *2017 International Conference on Current Research in Computer Science and Information Technology (ICCRIT)*. IEEE. doi:10.1109/CRCSIT.2017.7965534
- Siddiqui, A. J., & Boukerche, A. (2020). Adversarial Patches-based Attacks on Automated Vehicle Make and Model Recognition Systems. Paper presented at the *Proceedings of the 16th ACM Symposium on QoS and Security for Wireless and Mobile Networks*. ACM. doi:10.1145/3416013.3426461
- Siddiqui, A. J., & Boukerche, A. (2021). A Novel Lightweight Defense Method Against Adversarial Patches-Based Attacks on Automated Vehicle Make and Model Recognition Systems. *Journal of Network and Systems Management*, 29(4), 1–33. doi:10.1007/s10922-021-09608-6
- Su, J., Vargas, D. V., & Sakurai, K. (2019). One pixel attack for fooling deep neural networks. *IEEE Transactions on Evolutionary Computation*, 23(5), 828–841. doi:10.1109/TEVC.2019.2890858
- Sun, Q., Rao, A. A., Yao, X., Yu, B., & Hu, S. (2020). Counteracting adversarial attacks in autonomous driving. Paper presented at the *2020 IEEE/ACM International Conference On Computer Aided Design (ICCAD)*. ACM.
- Sun, W., & Wang, R. (2018). Fully convolutional networks for semantic segmentation of very high resolution remotely sensed images combined with DSM. *IEEE Geoscience and Remote Sensing Letters*, 15(3), 474–478. doi:10.1109/LGRS.2018.2795531
- Sun, Y., & Wang, C. (2022). Presentation Attacks in Palmprint Recognition Systems. *Journal of Multimedia Information System*, 9(2), 103–112. doi:10.33851/JMIS.2022.9.2.103
- Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I., & Fergus, R. (2013). Intriguing properties of neural networks. *arXiv:1312.6199*.
- Thabit, A. A., Mahmoud, M. S., Alkhayyat, A., & Abbasi, Q. H. (2019). Energy harvesting Internet of Things health-based paradigm: Towards outage probability reduction through inter-wireless body area network cooperation. *International Journal of Distributed Sensor Networks*, 15(10), 1550147719879870. doi:10.1177/1550147719879870
- Thys, S., Van Ranst, W., & Goedemé, T. (2019). Fooling automated surveillance cameras: adversarial patches to attack person detection. Paper presented at the *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. IEEE. doi:10.1109/CVPRW.2019.00012

- Tian, Y., Cheng, G., Gelernter, J., Yu, S., Song, C., & Yang, B. (2020). Joint temporal context exploitation and active learning for video segmentation. *Pattern Recognition*, *100*, 107158. doi:10.1016/j.patcog.2019.107158
- Wang, W., Qiu, C., Yin, Z., Srivastava, G., Gadekallu, T. R., Alsolami, F., & Su, C. (2021). Blockchain and PUF-based lightweight authentication protocol for wireless medical sensor networks. *IEEE Internet of Things Journal*.
- Wang, Y., Lv, H., Kuang, X., Zhao, G., Tan, Y., Zhang, Q., & Hu, J. (2021). Towards a physical-world adversarial patch for blinding object detection models. *Information Sciences*, *556*, 459–471. doi:10.1016/j.ins.2020.08.087
- Xiao, C., Zhong, P., & Zheng, C. (2019). Enhancing adversarial defense by k-winners-take-all. *arXiv:1905.10510*.
- Xue, X., Jiang, C., Zhang, J., & Hu, C. (2021). Biomedical Ontology Matching Through Attention-Based Bidirectional Long Short-Term Memory Network. [JDM]. *Journal of Database Management*, *32*(4), 14–27. doi:10.4018/JDM.2021100102
- Yang, S. X., Larsen, P. K., Alkjær, T., Simonsen, E. B., & Lynnerup, N. (2014). Variability and similarity of gait as evaluated by joint angles: Implications for forensic gait analysis. *Journal of Forensic Sciences*, *59*(2), 494–504. doi:10.1111/1556-4029.12322 PMID:24745080
- Yu, T., Hu, S., Guo, C., Chao, W.-L., & Weinberger, K. Q. (2019). A new defense against adversarial images: Turning a weakness into a strength. *arXiv:1910.07629*.
- Yuan, X., He, P., Zhu, Q., & Li, X. (2019). Adversarial examples: Attacks and defenses for deep learning. *IEEE Transactions on Neural Networks and Learning Systems*, *30*(9), 2805–2824. doi:10.1109/TNNLS.2018.2886017 PMID:30640631
- Zhang, H., Zhou, W., & Li, H. (2020). Contextual adversarial attacks for object detection. Paper presented at the *2020 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE. doi:10.1109/ICME46284.2020.9102805
- Zhou, Z., Guan, H., Bhat, M. M., & Hsu, J. (2019). Fake news detection via NLP is vulnerable to adversarial attacks. *arXiv:1901.09657*. 10.5220/0007566307940800
- Zhu, Z.-A., Lu, Y.-Z., & Chiang, C.-K. (2019). Generating adversarial examples by makeup attacks on face recognition. Paper presented at the *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE.

Maryam Bukhari have received Matriculation degree from Indus Valley Public Secondary School, Attock, Pakistan, in 2014, Intermediate degree from Bahira Foundation College, Wah Campus, Pakistan in 2016, and bachelor's degree in Computer Science from COMSATS University Islamabad, Attock Campus, Pakistan in 2020. She performed admirably throughout her four-year bachelor's degree, achieving the Campus Gold Medal and the Institute Bronze Medal in her B.S. degree. She is currently pursuing the master's degree with COMSATS University Islamabad–Attock, Pakistan. Her core research interests include Computer Vision, Deep Learning, Machine Learning and Recommender Systems. Through her research, she has been aiming to bring about advances in the domains.

Sadaf Yasmin is currently working as Assistant Professor at Department of Computer Science, COMSATS University Islamabad, Attock Campus, Pakistan since 2016. She has completed her MS and PhD in Computer Science from Capital University of Science and Technology, Islamabad, and BS in Software Engineering from NUML, University Islamabad in 2004. She has worked on a number of research projects during her PhD. She has also won her SRGP grants as PI, and Co-PI for a couple of projects. Her research interests include network protocol design, delay tolerant networks, computer vision, Deep learning, future network architectures, CCN and Internet of Things approaches.

Saira Gillani received her PhD degree in Information Sciences from Corvinus University of Budapest, Hungary. She joined the COMSATS Institute of Information Technology, Islamabad, Pakistan, in 2016. She also served as an assistant professor in Saudi Electronic University, Jeddah, Saudi Arabia. She is currently serving as a senior assistant professor in Bahria University Lahore, Pakistan. Previously, she worked as research scholar in Corvinno, Technology Transfer Center of Information Technology and Services in Budapest, Hungary and also worked as research associate in CoReNet (Center of Research in Networks and Telecom), CUST, Pakistan. Her areas of interest include Data Sciences, Text Mining, Data Mining, Machine Learning, Vehicular Networks, Mobile Edge Computing and Internet of Things.

Muazzam Maqsood is serving as an Assistant Professor at the Department of Computer science, COMSATS University Islamabad, Attock Campus, Pakistan. He holds a Ph.D. in software engineering with a keen interest in artificial intelligence and deep learning-based systems. He has more than 9 years of research and teaching experience. He has published more than 40 papers in top-ranked journals and conferences. His main research focus is to use the latest machine learning and deep learning algorithms to develop automated solutions especially in the field of pattern recognition and data analytics. He has published various top-ranked impact factor papers in the area of image processing, medical imaging, recommender systems, stock exchange prediction, and big data analytics. He is also a reviewer of many impact factor journals and program committee members of various international conferences. He is also working on many funded projects in collaboration with international researchers.

Seungmin Rho received his M.S. and Ph. D Degrees in Computer Science from Ajou University, Korea in 2003 and 2008, respectively. He visited Multimedia Systems and Networking Lab. in Univ. of Texas at Dallas from Dec. 2003 to March 2004. Before he joined the Computer Sciences Department of Ajou University, he spent two years in industry. In 2008–2009, he was a Postdoctoral Research Fellow at the Computer Music Lab of the School of Computer Science in Carnegie Mellon University. During 2013-2021, he was an assistant professor at Department of Media Software at Sungkyul University and Department of Software at Sejong University, respectively. Now he is currently an associate professor at Department of Industrial Security at Chung-Ang University. His current research interests include database, big data analysis, music retrieval, multimedia systems, machine learning, knowledge management as well as computational intelligence. He has published 300 papers in refereed journals and conference proceedings in these areas. He has been involved in more than 20 conferences and workshops as various chairs and more than 30 conferences/workshops as a program committee member. He has edited a number of international journal special issues as a guest editor, such as Multimedia Systems, Information Fusion, Engineering Applications of Artificial Intelligence, New Review of Hypermedia and Multimedia, Multimedia Tools and Applications, Personal and Ubiquitous Computing, Telecommunication Systems, Ad Hoc & Sensor Wireless Networks and etc. He has received a few awards including Who's Who in America, Who's Who in Science and Engineering, and Who's Who in the World in 2007 and 2008, respectively.

Sang-Soo Yeo received Ph.D degree in Computer Science & Engineering from Chung-Ang University, Seoul, Korea in 2005. He was a visiting scholar at Kyushu University, Japan. He worked for BTWorks, Inc. as a General Manager, and at the same time he was an adjunct professor at Hannam University, Daejeon, Korea. He worked for MOIS, Ministry of Interior and Safety and worked for PIPC, Personal Information Protection Commission during Feb. 2020 ~ Jul. 2021. He is a professor at the Dept. of Computer Engineering and also Dean of Planning & Budget, Mokwon University, Korea. He is President of the Institution of Creative Research Professionals (ICRP), and Vice President of ICT Platform Society (ICTPS). He is serving as Steering Chair of the PlatCon conference series, a very comprehensive conference series on platform technology and services (<http://www.platcon.org>)