IJIRSET
International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

e-ISSN: 2319-8753 | p-ISSN: 2347-6710

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 4, April 2024

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.423

9940 572 462 | 6381 907 438 | ijirset@gmail.com | www.ijirset.com

International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)

| e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 4, April 2024 ||

| DOI:10.15680/IJIRSET.2024.1304243 |

Poisoned Data Synthesis for Person Re-Identification Tasks

Anumula Shashank Reddy, Jayendra Kumar, Morampudi Krishna Kireeti,

Yashwanth Sai Kumar Janpal

U.G. Student, Department of Computer Science and Engineering, Anurag University, Hyderabad, Telangana, India

Associate Professor, Department of Computer Science and Engineering, Anurag University, Hyderabad, Telangana, India

U.G. Student, Department of Computer Science and Engineering, Anurag University, Hyderabad, Telangana, India

U.G. Student, Department of Computer Science and Engineering, Anurag University, Hyderabad, Telangana, India

ABSTRACT: Person re-identification, often abbreviated as person ReID, is a technology used in computer vision and surveillance systems to identify or track individuals across different cameras or scenes. Deep Learning techniques have caused significant advancement in this field. These models can learn features from raw images, enabling more accurate results. In case of person re-identification, since the labels in the training set are not present in the testing set, developing an attack on these models becomes much more complex. And hence, the existing backdoor attacks cannot work on these models right away. Here, we look at a unique approach to solve this problem by using triggers which are generated dynamically. Existing methods use fixed triggers and hence fail to show good results.

KEYWORDS: Data Poisoning, Dynamic Triggers, Person Re-Identification, Deep Learning

I. INTRODUCTION

Recently, Deep Learning models are being used for Person Re-Identification. These models show remarkable results, but they are susceptible to backdoor attacks [23], [24], [25], [26]. Backdoor attacks are performed by altering the data which is called “poisoning”. The data is poisoned with triggers based on the labels present in the dataset. Doing so makes the model mis-classify. Since the victim model will be trained on the new modified data, the model will fail. These backdoor attacks are very successful on Deep Learning models and are very hard to detect.

We are looking at one specific application of the Deep Learning models in person re-identification. These existing attacks on classification models do not perform well against ReID models. These backdoor attacks use all-to-one or all-to-all approaches. But Person Re-identification is a much more complex problem and the target labels are not not known before.

Person Re-Identification deals with finding whether the person in one frame of the video is the same as the person in the other frame. This is useful in tracking a person over various camera points. This is widely used in surveillance to track people from place to place. To achieve this, classification models in Deep Learning are used. While they show significant results, they are also prone to backdoor attacks [27], [28], [29]. But backdoor attacks on Person Re-Identification in particular are still not heavily studied. These existing backdoor attacks cannot be used directly on these models directly due to the problems mentioned above. And the success of these attacks is not very high. They alter the image highly, which makes it easy to find the attack has taken place.

To solve these problems, a method [7] has been proposed that uses dynamic triggers based on the features of the image. This is based on an all-to-unknown scenario where the classes in the testing data need not be present in the training data. In this attack, the aim is to create triggers based on unknown labels on a reference image and apply it to the image without altering the image highly.

Poisoning the data makes the model fail the task by mis-classifying the image. Since the model finds some of the features of other classes in the target class, while classifying, the model will misclassify the image. Therefore, the image and the image with the trigger will be labelled as the same class.

| e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 4, April 2024 ||

| DOI:10.15680/IJIRSET.2024.1304243 |

II. RELATED WORK

J. Zhang et al.[1] introduced an invisible attack called “Poison Ink”. The poisoned images cannot be identified by just looking at them with the naked eye. Hence it is called an invisible attack. The authors propose a novel method termed "poison ink" to execute backdoor attacks in image processing systems. Unlike conventional backdoor techniques, poison ink introduces imperceptible modifications, making it difficult for conventional defence mechanisms to detect and mitigate the attack. The core innovation of poison ink lies in its ability to generate perturbations that are both robust and virtually invisible to the human eye. This is achieved through a meticulous optimization process that ensures the alterations remain subtle yet effective in triggering the desired backdoor behaviour. They outline the technical details of the poison ink methodology, including the formulation of the optimization problem, the generation of imperceptible perturbations, and the integration of these perturbations into the training process. Experimental results demonstrate the efficacy of poison ink in compromising the integrity of image processing models across various tasks and datasets. The findings underscore the potential threat posed by sophisticated backdoor attacks like poison ink and emphasise the importance of developing robust defence strategies to safeguard image processing systems against such threats. But this method does not use dynamic triggers. While this attack performs well in terms of stealthiness, it cannot be used for ReID as the triggers generated are fixed and the labels are known beforehand.

Y. Li et al.[2] introduced a novel method for executing backdoor attacks specifically targeting visual object tracking systems. Backdoor attacks are a form of security threat wherein a malicious actor subtly manipulates the training data or model parameters to induce misclassification or erroneous behaviour. In this study, they introduce the concept of few-shot backdoor attacks, which aims to induce the backdoor behaviour in visual object tracking systems using only a limited amount of training data. Traditional backdoor attacks often require a significant volume of poisoned data for effective execution, but few-shot backdoor attacks streamline this process by leveraging a small set of strategically chosen training samples. They outline the methodology for implementing attacks on object tracking systems, including the selection of trigger patterns, generation of poisoned training samples, and integration of the backdoor into the tracking model. Key to the success of few-shot backdoor attacks is the efficient utilisation of limited training data to maximise the potency of the injected trigger. Results shown in the paper exhibit the strength of few-shot backdoor attacks in compromising the performance of visual object tracking systems across various scenarios and datasets. The findings of this study underscore the emerging threat posed by few-shot backdoor attacks in the context of visual object tracking and emphasise the importance of developing robust defence strategies to mitigate such threats. The research expands on the existing knowledge on adversarial machine learning and highlights the need for research into developing better defence systems.

S. Zhao et al.[3] presented a novel approach to conducting backdoor attacks specifically targeting video recognition models. Backdoor attacks are a form of security threat where malicious actors subtly manipulate the training data or model parameters to induce misclassification or erroneous behaviour when a specific trigger is present. In this study, the authors introduce the concept of clean-label backdoor attacks, which aim to compromise video recognition models without altering the labels of the training data. They outline the methodology for implementing clean-label backdoor attacks, including the selection of trigger patterns, generation of poisoned training samples, and integration of the backdoor into the recognition model. Key to the success of these attacks is the ability to inject subtle perturbations into the training data without affecting the true labels, thus making the attack more stealthy and difficult to detect. Experimental results demonstrate the strength of these attacks in compromising the performance of video recognition models across various datasets and scenarios. Moreover, the authors evaluate the strength of the attack method against existing defence mechanisms, highlighting its resilience to detection and mitigation efforts. The findings of this study underscore the emerging threat posed by the backdoor attacks and emphasise the importance of developing robust defence strategies to mitigate such threats.

L. Wang et al.[4] introduced a novel approach for conducting backdoor attacks specifically tailored for competitive reinforcement learning (RL) settings. Backdoor attacks involve maliciously manipulating the training data or model parameters to induce misbehaviour or exploit vulnerabilities when a trigger is present. In this study, the authors propose a method named BACKDOORL, which targets competitive RL environments where multiple agents interact with each other to achieve specific goals. Unlike traditional RL settings where a single agent learns to maximise its own reward, competitive RL introduces additional challenges as agents must contend with the actions and strategies of other agents. The BACKDOORL method enables an adversary to inject backdoor behaviour into one or more agents participating in the competitive RL environment. This backdoor behaviour is triggered by specific patterns in the input observations, leading to compromised performance or unintended behaviours when the trigger conditions are met. They outline the

| e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 4, April 2024 ||

| DOI:10.15680/IJIRSET.2024.1304243 |

methodology for implementing BACKDOORL, including the selection of trigger patterns, generation of poisoned training data, and integration of the backdoor into the RL agents' policies. Experimental results shown in the paper exhibit the strength of BACKDOORL in compromising the performance of competitive RL agents across various scenarios and benchmarks. Moreover, the authors evaluate the robustness of BACKDOORL against existing defence mechanisms and analyse its potential impact on the security and integrity of competitive RL systems.

Z. Xiang et al.[5] introduced a novel method for conducting backdoor attacks specifically targeted at 3D point cloud classifiers. In this study, the authors propose a technique aimed at compromising the integrity and reliability of 3D point cloud classifiers, which are widely used in various applications, including robotics, autonomous vehicles, and augmented reality. Unlike traditional image-based classifiers, 3D point cloud classifiers process three-dimensional geometric data, posing unique challenges and opportunities for adversarial manipulation. The proposed backdoor attack method leverages subtle perturbations injected into the input 3D point clouds, which, when present, trigger specific behaviours in the classifier, leading to compromised performance or unintended outcomes. The backdoor trigger pattern is carefully crafted to be inconspicuous yet effective in inducing the desired behaviour. The paper outlines the methodology for implementing the attack against 3D point cloud classifiers, including the selection and generation of trigger patterns, integration of the backdoor into the classifier model, and evaluation of attack effectiveness through extensive experiments. Experimental results presented in the paper show the potency of the attack in compromising the performance of 3D point cloud classifiers across various datasets and scenarios. Moreover, the authors test the strength of the attack against existing defence mechanisms, highlighting its effectiveness and resilience in real-world settings. The findings of this study underscore the emerging threat posed by backdoor attacks against 3D point cloud classifiers and emphasise the importance of developing robust defence strategies to mitigate such threats.

T. Zhai et al.[6] presents an interesting approach for conducting attacks specifically aimed at speaker verification systems. Speaker verification systems are widely used in security and authentication applications, relying on voiceprints to authenticate the identity of individuals. The attack aims to compromise the integrity and reliability of speaker verification systems, thereby undermining their security and trustworthiness. They outline the methodology for implementing the backdoor attack against speaker verification systems, including the selection and generation of trigger patterns, integration of the backdoor into the verification model, and evaluation of attack effectiveness through extensive experiments. Experimental results presented show the potency of the attack in compromising the performance of speaker verification systems across various datasets and scenarios. Moreover, the authors evaluate the strength of the attack against existing defence mechanisms, highlighting its effectiveness and resilience in real-world settings. The findings of this study underscore the emerging threat posed by backdoor attacks against speaker verification systems and emphasise the importance of developing robust defence strategies to mitigate such threats.

III. METHODOLOGY

In this particular scenario, the objective is to address potential vulnerabilities in a person re-identification network, introduced by adversarial actors through the strategic insertion of imperceptible triggers within the training dataset. Despite the adversary's lack of knowledge regarding the model's architecture and training loss, they possess the capability to query the trained model during inference, albeit without control over the inference process itself. By selectively incorporating these triggers into a subset of training images, the attack aims to induce backdoor behavior wherein the victim model erroneously associates any image containing such triggers with a predetermined target individual, regardless of the true identity depicted. Rigorous testing is conducted on a pristine test dataset, characterized by distinct target identifiers from those encountered during training, thereby mitigating concerns related to overfitting or memorization. In response to this sophisticated attack vector, novel methodologies are proposed to effectively detect and mitigate the deleterious impact of backdoor manipulations on the integrity and performance of person re-identification networks.

Let N denote the number of images in the dataset. The dataset is split into training data and testing data. N = Ntrain+ Ntest. Let D be the dataset. D = Dtrain + Dtest.

Prior to initiating the manipulation of images, this approach entails the division of the training data into two distinct subsets: the query set and the gallery set. During the training phase, for each image within the query set, we meticulously identify K images from the gallery set that exhibit the highest degrees of similarity. This process lays the groundwork for subsequent model inference, during which the adversary endeavors to manipulate the model's behavior.

| e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 4, April 2024 ||

| DOI:10.15680/IJIRSET.2024.1304243 |

Upon conducting an attack, an image selected from the query set prompts the ranking of images within the gallery set according to their resemblance, subsequently retrieving the top K candidates. However, a pivotal consideration arises due to the discrepancy between the identity labels employed in the training and testing stages of the model. This necessitates the dynamic generation of triggers tailored to the unseen identities encountered during testing.

To address this challenge, we leverage a reference image associated with an unseen target identity to delineate the specific identity label. This reference image is subjected to analysis by an identity hashing network, denoted as H, which is adept at extracting nuanced identity information. Subsequently, an encoder-decoder network is employed, termed T, within the framework of DNN-based picture steganography, exemplified by StegaStamp [8, 9]. This amalgamation of techniques facilitates the dynamic creation of poisoned images, imbued with the prescribed identifying characteristics.

The selection of StegaStamp as our steganographic tool is underpinned by its inherent stealthiness and efficacy in concealing the embedded triggers. Through extensive experimentation, we have discerned that StegaStamp consistently outperforms its counterparts, exhibiting superior results in terms of both effectiveness and surreptitiousness. Thus, the adoption of StegaStamp aligns seamlessly with our overarching objective of crafting a robust and covert approach to adversarial manipulation within the context of person re-identification networks.

Proposing a deep image steganography technique for embedding identity characteristics directly into images to facilitate trigger creation. Initially, the binary hash code of the unseen target identity undergoes transformation through a linear layer, resulting in a tensor of dimensions $32 \times 32 \times C$. This tensor is then upsampled to match the dimensions of the input image ($H \times W \times C$), followed by concatenation with the original image to form inputs of dimensions $H \times W \times (C \times 2)$. The encoded image, processed by an encoder following the U-Net architecture, introduces subtle pixel-level perturbations to generate the poisoned image.

An image reconstruction loss is applied to ensure effective embedding of the hash code while minimizing perceptual differences. This loss encompasses three components: the L2 residual regularization loss (LR), the LPIPS perceptual loss (LP), and the critical loss (LC) between the encoded image and the original image. Experimental validation demonstrates optimal attack performance and stealthiness with hyperparameters λR, λP, and λC set to 1.5, 2, and 1, respectively.

The technique aims to render the trigger imperceptible while maintaining effective attack performance and increasing susceptibility to backdoor attacks. An additional decoder predicts the identity hash code from the poisoned image, with another reconstruction loss ensuring efficient trigger addition.

The integration process into the training set is tailored for person re-identification datasets, characterized by sparse data compared to image classification datasets. Leveraging the all-to-unknown scenario, odd IDs are designated as target classes for images with even IDs. The hash code corresponding to each target ID is extracted and seamlessly embedded into matching training samples through steganography techniques, generating poisoned samples. Both clean and poisoned samples are provided for training.

To accommodate unseen IDs in the test set, their hash codes are generated from reference images and injected accordingly, acknowledging the disparity between target IDs in training and inference phases.

Additionally, the proposed technique underscores the importance of balancing stealthiness with attack efficacy, ensuring that the embedded triggers remain undetectable while effectively fulfilling their intended purpose. Furthermore, ongoing research efforts focus on refining the steganography process and enhancing the robustness of the trigger integration method to withstand detection and adversarial countermeasures.

This orchestrated manipulation aims to deceive the targeted model into erroneously associating images containing the trigger with a specified target ID. By introducing a trigger corresponding to another individual's ID, the adversary under surveillance by the ReID model can evade detection. Notably, the proposed approach deviates from the traditional all-to-one backdoor attack technique, as exemplified by IBA [46]. Unlike classical backdoor attacks, IBA categorizes backdoor behavior into distinct modes, including attack mode and cross-trigger mode, which serve to activate the backdoor and prevent cross-contamination of triggers, respectively. In contrast, the approach detailed in

| e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 4, April 2024 ||

| DOI:10.15680/IJIRSET.2024.1304243 |

this section adopts target-aware triggers, as depicted in Fig. 3.2, enabling contaminated photos with varied triggers to be correctly identified as belonging to the same target individual.

IV. EXPERIMENTAL RESULTS

This section assesses the suggested backdoor attack's stealth, efficiency, and resistance to various backdoor defence strategies against the person ReID models. We also substantiate the significance of the design and the selection of parameters.

For the evaluation of the attack, the Market-1501 dataset was selected, comprising 36,036 images representing 1,501 distinct individuals captured across six different cameras. Within this dataset, 751 IDs were designated for the training set, while 750 IDs were reserved for the test set. The victim model chosen for testing the attack was FastReID, featuring ResNet-101 as the backbone architecture. Importantly, the adversary is intentionally provided with no prior knowledge regarding the target ReID model.

Notably, the identity hashing network's feature extractor is developed and trained independently, maintaining autonomy from the ReID model's feature extractor. Although the identity hashing network remains consistent throughout, the target ReID model is configured to utilize distinct techniques tailored for person re-identification tasks. Furthermore, the parameters of the image generator, trained on the same dataset, remain unchanged and consistent across different model configurations.

It is pertinent to note that currently, there is no readily applicable backdoor attack directly applicable to Person ReID systems, primarily due to the intricate nature of target labels within all-to-unknown scenarios. Existing attack techniques are challenged by the absence of the target label from the training set, rendering them ineffective in such scenarios.

As a baseline comparison, Bednets, Blended, ReFool, SIG, and WaNet are selected for evaluation. Each of these techniques represents distinct approaches to backdoor attacks within the context of image classification. Bednets, for instance, operate as a patch-based backdoor attack, altering pixel patterns within benign images to generate contaminated counterparts. Blended, on the other hand, utilizes unrelated images as triggers, superimposing them onto original images in specific proportions to create poisoned samples.

ReFool introduces reflections into the victim model as backdoor triggers, leveraging images external to the training set for insertion. SIG employs a ramp signal to trigger perceptually invisible poisoned images, while WaNet employs picture manipulation techniques to introduce undetectable backdoors. In the testing phase, misclassification occurs when the victim model encounters images containing triggers, as all of these techniques involve the creation of poisoned images added to the training set. These methodologies serve as benchmarks for evaluating the efficacy and stealthiness of the proposed backdoor attack approach.

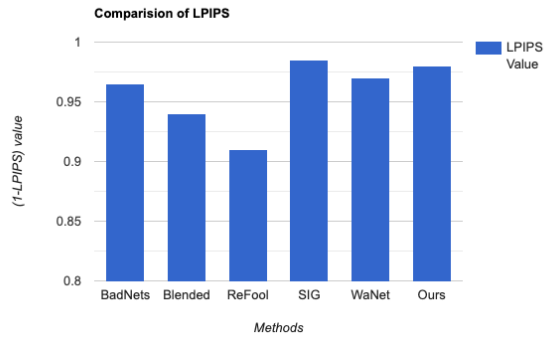

Fig. 4.1 Comparison of similarity between poisoned image and clean image of various approachesThe analysis of stealthiness, depicted in Fig. 4.1, is conducted utilizing the Learned Perceptual Image Patch Similarity (LPIPS) metric, which quantifies the similarity between poisoned and clean images. Here, a lower LPIPS value indicates a higher degree of similarity between the images. In our evaluation, the values presented in the bar graph represent (1-

| e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 4, April 2024 ||

| DOI:10.15680/IJIRSET.2024.1304243 |

LPIPS) to reflect a positive effect. Notably, our approach demonstrates superior similarity compared to other methodologies, as evidenced by its positioning among the top values in the graph.

However, it is imperative to note that LPIPS should not be considered in isolation. The effectiveness of an approach constitutes a delicate trade-off between image similarity and model performance. For instance, while SIG exhibits a high LPIPS value, its performance in terms of rank-10 accuracy, which will be elaborated upon in subsequent sections, may not be optimal. This observation underscores the necessity of considering both stealthiness and model performance when evaluating the efficacy of backdoor attack techniques.

The effectiveness of the model is evaluated through a comparison between its performance in clean mode and attacked mode, as presented in Table 1. This comparison allows for an assessment of the impact of the attack on various performance metrics. Notably, the attack is executed with a poisoning rate of 37.5%, indicating the proportion of contaminated samples within the dataset.

To gauge performance, Rank-10 accuracy (R-10) and Attack Success Rate (ASR) are utilized as key metrics. R-10 accuracy measures the model's ability to correctly identify the top 10 pedestrian images most likely to represent the same identity as the query image. ASR, on the other hand, evaluates the extent to which the attack strategy alters the model's behavior, ensuring that positive images are ranked outside the top 10 positions.

Moreover, mean Average Precision (mAP) and retrieval performance rank-10 (R-10) are employed as additional metrics on contaminated photos. A lower mAP and rank-10 signify superior non-targeted attack performance, indicating the effectiveness of the attack strategy in disrupting the model's accurate identification of individuals within contaminated images.

TABLE 1 PERFORMANCE UNDER NO ATTACK AND THE ATTACK

Model	ASR	R-10	mAP
Clean	-	99.28	89.97
Under Attack	94.15	3.04	1.20

V. CONCLUSION

Backdoor attacks in the realm of person re-identification have received limited attention, with the bulk of existing research focusing on image classification tasks. As current backdoor attack methods are not directly applicable to person re-identification, we introduce a novel approach tailored specifically for this domain. This unique method enables the dynamic generation of new backdoor triggers, including unknown identities in the test set, within a new framework termed the all-to-unknown scenario.

Central to the approach is the deployment of an identity hashing network, which initially extracts target identification details from a reference image. These details are then seamlessly integrated into benign images using image steganography techniques. Through empirical evaluation, we review the strength of the attack.

REFERENCES

[1] J. Zhang, C. Dongdong, Q. Huang, J. Liao, W. Zhang, H. Feng, G. Hua, and N. Yu, “Poison ink: Robust and invisible backdoor attack,” IEEE Transactions on Image Processing, vol. 31, pp. 5691–5705, 2022.
[2] Y. Li, H. Zhong, X. Ma, Y. Jiang, and S.-T. Xia, “Few-shot backdoor attacks on visual object tracking,” arXiv preprint arXiv:2201.13178, 2022.
[3] S. Zhao, X. Ma, X. Zheng, J. Bailey, J. Chen, and Y. Jiang, “Clean-label backdoor attacks on video recognition models,” in CVPR. Computer Vision Foundation / IEEE, 2020, pp. 14 431–14 440.
[4] L. Wang, Z. Javed, X. Wu, W. Guo, X. Xing, and D. Song, “BACKDOORL: backdoor attack against competitive reinforcement learning,” in IJCAI. ijcai.org, 2021, pp. 3699–3705.

| e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 4, April 2024 ||

| DOI:10.15680/IJIRSET.2024.1304243 |

[5] Z. Xiang, D. J. Miller, S. Chen, X. Li, and G. Kesidis, "A backdoor attack against 3d point cloud classifiers," in ICCV. IEEE, 2021, pp. 7577–7587.
[6] Z. Xiang, D. J. Miller, S. Chen, X. Li, and G. Kesidis, "A backdoor attack against 3d point cloud classifiers," in ICCV. IEEE, 2021, pp. 7577–7587.
[7] W. Sun et al., "Invisible Backdoor Attack With Dynamic Triggers Against Person Re-Identification," in IEEE Transactions on Information Forensics and Security, vol. 19, pp. 307-319, 2024, doi: 10.1109/TIFS.2023.3322659.
[8] M. Tancik, B. Mildenhall, and R. Ng, "Stegastamp: Invisible hyperlinks in physical photographs," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 2117–2126.
[9] S. Li, M. Xue, B. Z. H. Zhao, H. Zhu, and X. Zhang, "Invisible backdoor attacks on deep neural networks via steganography and regularization," IEEE Transactions on Dependable and Secure Computing, vol. 18, no. 5, pp. 2088–2105, 2020.
[10] R. Rivest, "The md5 message-digest algorithm," Tech. Rep., 1992.
[11] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in International Conference on Medical image computing and computer-assisted intervention. Springer, 2015, pp. 234–241.
[12] W. Li, R. Zhao, T. Xiao, and X. Wang, "Deepreid: Deep filter pairing neural network for person re-identification," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, pp. 152–159.
[13] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang and Q. Tian, "Scalable Person Re-identification: A Benchmark," 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 2015, pp. 1116-1124, doi: 10.1109/ICCV.2015.133.
[14] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian, "Scalable person re-identification: A benchmark," in Proceedings of the IEEE international conference on computer vision, 2015, pp. 1116–1124.
[15] G. Wang, Y. Yuan, X. Chen, J. Li, and X. Zhou, "Learning discriminative features with multiple granularities for person re-identification," in Proceedings of the 26th ACM international conference on Multimedia, 2018, pp. 274–282.
[16] He, L., Liao, X., Liu, W., Liu, X., Cheng, P., & Mei, T. (2023). FastReID: A Pytorch Toolbox for General Instance Re-identification. Proceedings of the 31st ACM International Conference on Multimedia, 9664–9667. https://doi.org/10.1145/3581783.3613460.
[17] T. Gu, B. Dolan-Gavitt, and S. Garg, "Badnets: Identifying vulnerabilities in the machine learning model supply chain," arXiv preprint arXiv:1708.06733, 2017.
[18] X. Chen, C. Liu, B. Li, K. Lu, and D. Song, "Targeted backdoor attacks on deep learning systems using data poisoning," arXiv preprint arXiv:1712.05526, 2017.
[19] Y. Liu, X. Ma, J. Bailey, and F. Lu, "Reflection backdoor: A natural backdoor attack on deep neural networks," in European Conference on Computer Vision. Springer, 2020, pp. 182–199.
[20] M. Barni, K. Kallas, and B. Tondi, "A new backdoor attack in cnns by training set corruption without label poisoning," in 2019 IEEE International Conference on Image Processing (ICIP). IEEE, 2019, pp. 101–105.
[21] A. Nguyen and A. Tran, "Wanet–imperceptible warping-based backdoor attack," arXiv preprint arXiv:2102.10369, 2021.
[22] R. Zhang, P. Isola, A. Efros, E. Shechtman and O. Wang, "The Unreasonable Effectiveness of Deep Features as a Perceptual Metric," in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 2018 pp. 586-595. doi: 10.1109/CVPR.2018.00068.
[23] Z. Sun, P. Kairouz, A. T. Suresh, and H. B. McMahan, "Can you really backdoor federated learning?" arXiv preprint arXiv:1911.07963, 2019.
[24] Y. Li, B. Wu, Y. Jiang, Z. Li, and S.-T. Xia, "Backdoor learning: A survey," arXiv preprint arXiv:2007.08745, 2020.
[25] H. Wang, K. Sreenivasan, S. Rajput, H. Vishwakarma, S. Agarwal, J.- y. Sohn, K. Lee, and D. Papailiopoulos, "Attack of the tails: Yes, you really can backdoor federated learning," Advances in Neural Information Processing Systems, vol. 33, pp. 16 070–16 084, 2020.
[26] E. Bagdasaryan, A. Veit, Y. Hua, D. Estrin, and V. Shmatikov, "How to backdoor federated learning," in International Conference on Artificial Intelligence and Statistics. PMLR, 2020, pp. 2938–2948.
[27] Y. Liu, X. Ma, J. Bailey, and F. Lu, "Reflection backdoor: A natural backdoor attack on deep neural networks," in European Conference on Computer Vision. Springer, 2020, pp. 182–199.

| e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 4, April 2024 ||

| DOI:10.15680/IJIRSET.2024.1304243 |

[28] S. Li, M. Xue, B. Z. H. Zhao, H. Zhu, and X. Zhang, "Invisible backdoor attacks on deep neural networks via steganography and regularization," IEEE Transactions on Dependable and Secure Computing, vol. 18, no. 5, pp. 2088–2105, 2020.

[29] A. Saha, A. Subramanya, and H. Pirsiavash, "Hidden trigger backdoor attacks," in Proceedings of the AAAI conference on artificial intelligence, vol. 34, no. 07, 2020, pp. 11 957–11 965.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details