2025 International Conference on Machine Learning and Cybernetics (ICMLC) | 979-8-3315-8736-9/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICMLC66258.2025.11280213

BAND-AID BACKDOOR: A DUAL-BRANCH FRAMEWORK FOR PHYSICAL BACKDOOR ATTACKS ON FACE RECOGNITION SYSTEMS

JUN-HONG LI, YING-QI ZHENG, PATRICK P. K. CHAN*

Shien-Ming Wu School of Intelligent Engineering, South China University of Technology, Guangzhou, 511442, China
E-MAIL: 202264642167@mail.scut.edu.cn, 202420160854@mail.scut.edu.cn, patrickchan@scut.edu.cn

Abstract:

A physical backdoor attack framework against face recognition systems, using band-aid stickers as visually plausible triggers, is proposed. The proposed method introduces a dual-branch architecture that decouples identity recognition from trigger classification, enabling the model to maintain high recognition accuracy on clean inputs while inducing targeted misclassification when a specific physical trigger is present. To ensure photorealistic integration, the band-aid undergoes a 3D curvature-aware warping and illumination adjustment process before being applied to facial images. Experiments conducted under both digital and real-world conditions demonstrate the effectiveness and specificity of the attack. Compared to a standard single-task model, the dual-branch model achieves significantly higher attack success rates and lower false positive rates, while remaining robust to unseen trigger types. However, results also reveal a performance gap between digital and physical scenarios, emphasizing the importance of visual distinctiveness and domain consistency. This work highlights the practical feasibility of physical backdoor attacks and provides insights into their design and mitigation.

Keywords:

Physical Backdoor Attack; Face Recognition; Adversarial Learning

1. Introduction

Face recognition systems have been extensively deployed in applications such as identity authentication [1] and surveillance [2], owing to their capability to encode facial features into a compact embedding space that minimizes intra-class variation while maximizing inter-class separability. However, the widespread adoption of these systems has also attracted increasing attention from adversaries [3]. A growing body of research has revealed that such systems are vulnerable to various forms of adversarial attacks [4], posing serious threats to their reliability and security in real-world deployments.

Among the various forms of adversarial threats, backdoor attacks [5] are a type of adversarial attack where malicious behaviors are intentionally embedded into a model during the training phase. By injecting specially crafted poisoned samples associated with a hidden trigger, the model behaves normally on clean data but misclassifies inputs when the trigger is present. This makes backdoor attacks particularly stealthy and dangerous, as they remain dormant until activated by the attacker. If a face recognition system is compromised with a backdoor, an attacker could easily bypass authentication using a specific accessory or gesture, posing a serious security and privacy threat.

Considering the limitations of existing backdoor attacks that often rely on conspicuous or easily detectable triggers, we propose an approach that leverages natural-looking objects to achieve stealthier attacks in real-world scenarios. Specifically, this work presents a physical backdoor attack framework that utilizes adhesive bandages as covert triggers, exploiting their natural and inconspicuous appearance. In contrast to the conspicuous adversarial patches adopted in previous studies [6, 7], bandages blend seamlessly with human facial features, thereby enhancing the stealth and feasibility of the attack in real-world settings. The proposed framework includes an efficient two-phase backdoor injection method. In the first phase, a multi-task training architecture is employed, featuring an auxiliary bandage classification branch designed to enhance trigger specificity and minimize unintended activations. A 3D curvature-aware digital augmentation pipeline is developed to generate photorealistic poisoned samples, ensuring realistic appearance and effective embedding of the trigger, in the second phase. This pipeline aligns bandage textures with facial landmarks and adjusts their shape according to local surface geometry, enabling realistic and scalable data generation without

979-8-3315-8736-9/25/$31.00 ©2025 IEEE

Authorized licensed use limited to: SOUTH CHINA UNIVERSITY OF TECHNOLOGY. Downloaded on July 29,2026 at 12:27:19 UTC from IEEE Xplore. Restrictions apply.

requiring extensive physical data collection (Figure. 1). The attack performance of our proposed framework is evaluated by using the benchmark FaceNet dataset [1].

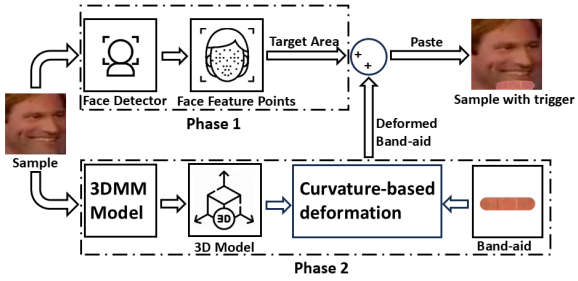

FIGURE 1. The pipeline of our proposed model.

The key contributions of this paper are:

- A novel physical backdoor attack framework is proposed, utilizing realistic band-aid triggers combined with a dual-branch learning architecture to achieve targeted misclassification while preserving clean face recognition performance.
- A 3D-aware trigger embedding pipeline is developed, incorporating facial geometry warping and illumination adjustment to seamlessly integrate physical triggers into facial images, enhancing visual plausibility and real-world applicability.
- Comprehensive digital and physical evaluations demonstrate that the dual-branch model significantly improves attack success rate, trigger specificity, and false positive suppression compared to traditional single-task baselines, while revealing key insights into domain transferability and trigger design.

2. Related Work

Backdoor attacks pose a serious threat to the security of deep neural networks (DNNs), particularly in safety-critical applications. By injecting malicious behavior during the training phase, these attacks allow a model to maintain high accuracy on clean inputs while producing attacker-specified outputs when a predefined trigger is present. Early work primarily focused on improving the effectiveness of poisoning-based backdoor attacks. A seminal example is BadNets [8], which introduced a fixed, visible pattern as a trigger embedded into training samples. Subsequent research has enhanced stealthiness and robustness by exploring various trigger forms. For instance, imperceptible or blended triggers were proposed to evade human inspection [7], while clean-label attacks maintained correct ground-truth labels to bypass label-based filtering [9]. More sophisticated methods have leveraged semantic features [10] or conducted manipulations in feature space [11] to construct more natural and task-relevant triggers. In parallel, the design of optimized triggers has become a central theme. Some approaches seek to maximize neuron activations [12], whereas others adopt bi-level optimization frameworks to jointly learn both the trigger pattern and model parameters [13, 14]. This evolution reflects a shift from heuristic trigger crafting to principled and systematic attack optimization. Beyond trigger design, researchers have also investigated the wide variety of trigger modalities used to activate backdoor behavior. These include visible patterns [8], imperceptible perturbations [15], and even parameter-level modifications at the model level [16], revealing the broad attack surface and complexity involved in defense.

Backdoor attacks have been studied in both digital and physical domains. In digital scenarios, triggers are directly embedded at the pixel level, offering precise control over their location and appearance [7, 8]. Physical attacks, by contrast, aim to introduce real-world triggers that remain effective when captured by sensors or cameras. Chen *et al.* [7] demonstrated that adversarial eyeglasses could serve as physical triggers for face recognition systems. To address robustness challenges under varying environmental conditions—such as lighting, scaling, or rotation—later works proposed trigger designs that are resilient to such transformations [15]. These studies highlight the practical feasibility of backdoor attacks in real-world deployments and the pressing need for effective countermeasures.

3. Proposed Band-Aid Backdoor Attack Framework

The proposed **Band-aid Backdoor Attack Framework** introduces a dual-path learning architecture designed to preserve high face recognition accuracy on clean inputs while enabling targeted misclassification when a predefined physical trigger—a band-aid—is present. To achieve this, identity recognition and trigger learning are decoupled into separate branches during training. This separation ensures controllable backdoor activation with minimal interference to the model's primary functionality, thereby enhancing the stealth and stability of the attack.

3.1 Band-aid Trigger Preparation

To enable a realistic and effective physical backdoor attack, our method uses band-aids as physical triggers, strategically applied to the chin region of facial images. This region is chosen

Authorized licensed use limited to: SOUTH CHINA UNIVERSITY OF TECHNOLOGY. Downloaded on July 29,2026 at 12:27:19 UTC from IEEE Xplore. Restrictions apply.

to simulate plausible real-world placement, as band-aids naturally conform to facial contours in this area.

3.1.1 Trigger Variants

To evaluate the robustness and generality of the attack, five distinct types of band-aids are employed as triggers, varying in shape, texture, and color (Figure 2).

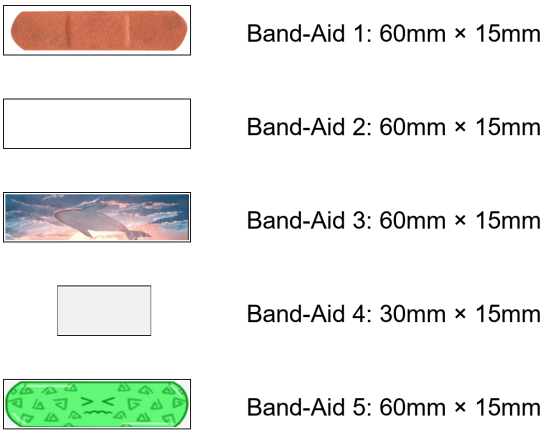

FIGURE 2. Five types of band-aid triggers used in the attack.

3.1.2 Trigger Embedding

A naive approach, such as directly overlaying a 2D band-aid image, fails to produce photorealistic results due to facial curvature (Figure 3(a)). Therefore, a dedicated geometric transformation pipeline [17] is introduced to warp the band-aid to the underlying 3D facial surface, achieving more natural integration (Figure 3(b)). The transformation consists of three stages: 3D facial modeling, curvature-aware warping, and illumination adjustment.

3D Facial Modeling and Coordinate Extraction : 3D Morphable Models (3DMM) [18] are used to reconstruct the 3D geometry of a given facial image. From this reconstruction, precise coordinates of the target region are extracted. The central point for band-aid placement is denoted as (x_0, y_0, z_0), corresponding to the highest elevation within the region of interest.

Curvature-Aware Warping Transformation The band-aid trigger undergoes a two-phase geometric warping process. To match facial curvature, the band-aid is first projected onto the

(a) Naive 2D overlay (b) Curvature-aware warping

FIGURE 3. Comparison of band-aid application methods.

X-Z plane with fixed Y-coordinates. The deformation follows a parabolic curve:

$$z = a(x - c)^2 + b \tag{1}$$

where $c = x_0$, $a = -\Delta h/\Delta s^2$, and $b = -a(w_n - c)^2$. The deformed width w_n is computed to preserve arc length:

$$\int_0^{w_n} \sqrt{1 + 4a^2(x - c)^2}dx = w \tag{2}$$

resulting in a warped band-aid coordinate matrix $M_A \in \mathbb{R}^{3\times(h\cdot w_n)}$.

To align the band-aid with the facial orientation, a rotation is applied. The rotation angle is derived as:

$$\theta = \text{sign}(h - 2y_0) \cdot \arctan(\Delta z/\Delta y) \tag{3}$$

and the corresponding transformation is performed with:

$$M_B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix} M_A \tag{4}$$

Photorealistic Illumination Adjustment : To improve realism, a radial brightness adjustment is applied to simulate natural light fall-off:

$$l_p(i, j) = -k_l \cdot \sqrt{(i - x_0)^2 + (j - y_0)^2} + l_p(x_0, y_0) \tag{5}$$

where k_l is a tunable brightness decay factor. This adjustment ensures the band-aid visually conforms to facial shading.

The complete transformation pipeline seamlessly integrates the band-aid trigger into the facial image, producing poisoned samples that preserve visual realism and effectively trigger the backdoor in physical-world settings. This pipeline overcomes the limitations of naive compositing methods and supports robust real-world deployment.

Authorized licensed use limited to: SOUTH CHINA UNIVERSITY OF TECHNOLOGY. Downloaded on July 29,2026 at 12:27:19 UTC from IEEE Xplore. Restrictions apply.

3.1.3 Training Phase

The proposed attack follows a poisoning threat model, where the adversary has full control over the training pipeline and is responsible for model construction. Poisoned samples are created by overlaying a predefined band-aid trigger onto clean face images and relabeling them with a designated target class. These samples are then injected into the training set to induce backdoor behavior while preserving clean performance. As illustrated in Figure 4, the training pipeline consists of three key modules: the input module, the feature extractor, and two parallel output branches. The input face image—either clean or backdoored—is first transformed into a tensor and passed through a feature extractor, resulting in a fixed-length embedding vector. This embedding is then jointly optimized by two branches: the face branch and the band-aid branch. The face branch is trained using a triplet loss to preserve discriminative facial representations for identity verification, while the band-aid branch is supervised with a cross-entropy loss to classify the presence and type of trigger. This multi-task design ensures that the network learns to recognize and encode the backdoor trigger patterns without compromising clean face recognition performance, thereby enhancing the specificity of the attack.

Dual-Path Architecture: To support stealthy and controllable backdoor activation, we adopt a dual-path learning architecture that decouples identity recognition from trigger learning. As shown in Figure 4, a shared feature extractor f_B encodes each image $\mathbf{x}_i$ into a high-level representation $\mathbf{h} \in \mathbb{R}^{512}$. This backbone network (InceptionResnetV1 pretrained on VGGFace2 [19, 20]) remains frozen to stabilize feature representations across tasks. Two task-specific heads then operate in parallel:

$$\mathbf{h} = f_B(\mathbf{x}), \quad \mathbf{b} = f_A(\mathbf{h}), \quad \hat{y} = f_R(\mathbf{h}) \tag{6}$$

where f_R denotes the face recognition head, and f_A denotes the band-aid classification head, which is implemented as a single linear layer ($512 \rightarrow 5$) trained using cross-entropy loss.

Gradient Isolation: By design, task gradients are completely decoupled: each head updates only its own parameters, while the shared backbone f_B remains fixed during the initial training phase. This ensures strict separation between face recognition and backdoor learning, allowing the model to maintain clean accuracy while learning to respond to the physical trigger.

Dataset Contamination: Let $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N}$ denote the training set with $y_i \in \{1, 2, ..., C\}$ as identity labels. A subset of images is selected for poisoning by applying the band-aid trigger and changing their labels to a fixed target class. These poisoned images are included in the training set alongside clean samples to construct a contaminated dataset that enables backdoor activation.

Loss Function: The model is trained end-to-end using a multi-task objective that balances facial identity discrimination and trigger classification:

$$L_{\text{total}} = \alpha_t L_{\text{triplet}} + \beta_t L_{\text{CE}} \tag{7}$$

where L_{triplet} encourages discriminative face embeddings, and L_{CE} supervises the auxiliary trigger classification branch. The weights α_t and β_t are dynamically adjusted during training to maintain balance between the two tasks.

Dynamic Weight Adjustment: To prevent the auxiliary task from dominating training or suppressing identity recognition, a dynamic adjustment strategy is adopted as summarized in Algorithm 1. The weights are computed based on the relative magnitudes of the current losses, subject to stability constraints and priority bounds.

Algorithm 1 Dynamic Weight Adjustment

Require: Current losses L^t_{triplet}, L^t_{CE}; bounds $\alpha_{\min} = 0.2$, $\beta_{\max} = 0.8$

$total \leftarrow L^t_{\text{triplet}} + L^t_{\text{CE}}$
$\alpha_t \leftarrow L^t_{\text{triplet}}/total$, $\beta_t \leftarrow L^t_{\text{CE}}/total$
if $\alpha_t < \alpha_{\min}$ **then**
    $\alpha_t \leftarrow \alpha_{\min}, \quad \beta_t \leftarrow 1 - \alpha_{\min}$
end if
if $\beta_t > \beta_{\max}$ **then**
    $\beta_t \leftarrow \beta_{\max}, \quad \alpha_t \leftarrow 1 - \beta_{\max}$
end if
if $L^t_{\text{CE}} > 2L^t_{\text{triplet}}$ **and** $\beta_t < \beta_{\max}$ **then**
    $\beta_t \leftarrow \min(1.5 \cdot \beta_t, \beta_{\max})$
end if

Ensure: Normalized weights α_t, β_t such that $\alpha_t + \beta_t = 1$

This training procedure enables the model to preserve face recognition capability under clean conditions while embedding a covert and reliable backdoor pathway triggered by a physically realizable band-aid.

3.1.4 Inference Phase

At inference time, only the identity recognition branch is activated, ensuring that normal inputs are processed as expected.

Authorized licensed use limited to: SOUTH CHINA UNIVERSITY OF TECHNOLOGY. Downloaded on July 29,2026 at 12:27:19 UTC from IEEE Xplore. Restrictions apply.

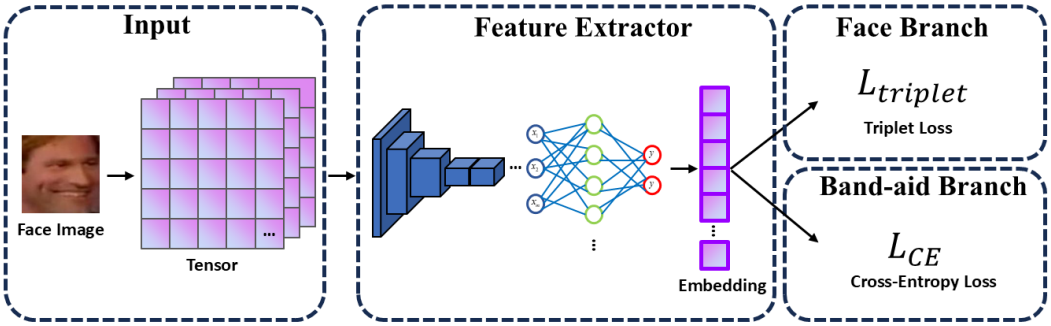

FIGURE 4. Dual-path feature learning architecture. The face recognition and band-aid classification heads operate independently on shared representations extracted by a frozen backbone.

However, the presence of the trigger reliably induces misclassification to the attacker's chosen target. The adversary is assumed to have knowledge of the model architecture and training data distribution, enabling the design of effective and covert trigger patterns.

4. Experiments

4.1 Experimental Setup

4.1.1 Datasets

Experiments are conducted using the Labeled Faces in the Wild (LFW) dataset [21]. Four identities are selected for training, with data augmentation applied to increase sample diversity. For digital experiments, 150 LFW images are modified by overlaying synthetic band-aid triggers using our transformation pipeline. For physical experiments, images are collected from 10 volunteers wearing printed band-aid stickers on their chins, captured under diverse lighting and pose conditions. This dual-scenario setup allows evaluation of both synthetic and real-world backdoor effectiveness.

4.1.2 Attack Configuration

We implement a poisoning-based attack using five visually distinct band-aid triggers. During training, only four of these triggers are used; all five are included in testing. Images containing a predefined trigger are assigned a fixed target identity label, while all others retain their original labels. This selective poisoning approach enables assessment of both targeted misclassification and cross-trigger generalization.

To enhance trigger specificity, we adopt a multi-task learning strategy. The model simultaneously learns face recognition and band-aid classification, with task losses balanced using a dynamic weighting mechanism. At inference time, the auxiliary branch is disabled, preserving normal behavior on clean inputs while maintaining the attack capability when a trigger is present.

4.2 Results and Analysis

4.2.1 Digital and Physical Performance

The attack success rate (ASR) is evaluated under both digital and physical conditions. In the digital setting, synthetic band-aid overlays allow precise control over trigger placement and appearance. In contrast, the physical setting introduces natural variability, including changes in lighting, facial pose, and background. Two model configurations are compared: the standard model, trained solely with face recognition loss, and the dual-branch model, which is trained with a joint objective that incorporates both face recognition and band-aid classification losses.

4.2.2 Performance Comparison

As summarized in Table 3, the dual-branch model demonstrates superior performance across several metrics. Digital ASR improves from 78.12% to 92.72%, and the average cross-class ASR decreases significantly, indicating higher trigger specificity. The false positive rate on clean samples is also halved, from 27.33% to 13.04%.

4.2.3 Observations and Discussion

Several key observations emerge from the results. Band-Aid 1 achieves the highest attack success rate (ASR) in the digital setting at 92.72%, but completely fails in the physical do-

Authorized licensed use limited to: SOUTH CHINA UNIVERSITY OF TECHNOLOGY. Downloaded on July 29,2026 at 12:27:19 UTC from IEEE Xplore. Restrictions apply.

TABLE 1. Digital Attack Success Rate (ASR) Comparison

Model	Test Band-Aid Class					Clean
	1	2	3	4	5	
Regular Model (Single-Task)						
None	3.75	1.87	0.63	1.87	1.87	0.62
1	67.50	1.25	1.87	1.25	1.87	18.63
2	0.00	78.12	0.63	50.00	3.12	18.01
3	2.50	3.12	52.50	1.87	24.37	27.33
4	10.62	41.88	7.50	73.75	64.38	11.80
Dual-Branch Model (Multi-Task)						
None	3.75	2.50	2.50	3.12	2.50	2.48
1	**92.72**	0.00	0.00	0.00	0.66	3.11
2	0.00	**82.67**	0.00	52.98	3.31	0.00
3	1.32	2.00	**72.00**	0.66	24.50	2.48
4	7.95	44.00	6.67	**77.48**	67.55	13.04

TABLE 2. Physical Attack Success Rate (ASR) Comparison

Model	Test Band-Aid Class					Clean
	1	2	3	4	5	
Regular Model						
None	0.00	0.00	0.00	0.00	0.00	0.00
1	14.29	0.00	11.11	12.00	19.05	22.22
2	0.00	0.00	0.00	0.00	0.00	0.00
3	42.86	50.00	**88.89**	72.00	**90.48**	33.33
4	4.76	0.00	5.56	0.00	9.52	0.00
Dual-Branch Model						
None	0.00	0.00	0.00	0.00	0.00	0.00
1	0.00	0.00	0.00	0.00	0.00	0.00
2	0.00	5.56	0.00	0.00	0.00	0.00
3	0.00	0.00	38.89	0.00	33.33	33.33
4	19.05	**94.44**	16.67	**84.00**	19.05	19.05

main (0.00%), likely due to its low visual contrast, which reduces salience under real-world lighting. In contrast, Band-Aid 3 demonstrates better generalization across both domains, attributed to its high-contrast dolphin pattern that remains distinguishable even under physical distortions. Band-Aid 4 elicits strong targeted activation in both settings but causes significant confusion with Band-Aid 2 in the physical domain, suggesting that overlapping visual characteristics hinder the model's ability to distinguish between similar triggers. Lastly, although Band-Aid 5 is excluded from training and typically yields low ASR, its occasional activation when Band-Aid 3 is used as the trigger likely arises from shared visual traits—both possess vivid coloration and intricate patterns—suggesting that the model's trigger sensitivity may be influenced by such salient visual features, even across distinct trigger classes.

While the dual-branch model enhances attack precision, it also shows a slightly larger digital-physical performance gap.

TABLE 3. Key Performance Metrics Comparison

Metric	Regular Model	Dual-Branch Model
Max Digital ASR	78.12%	**92.72%**
Max Physical ASR	90.48%	**94.44%**
Avg Cross-Class ASR	12.34%	**1.87%**
False Positive Rate (Clean)	27.33%	**13.04%**
Digital-Physical Gap	34.7%	41.2%

This may stem from overfitting to clean digital appearances. Augmenting training with realistic physical data could improve robustness. The dual-branch architecture significantly improves targeted attack success rates, reduces false positives, and enforces higher trigger specificity. However, its performance in physical scenarios is influenced by trigger visibility and training data diversity. Bridging the digital-physical gap remains a key direction for future work.

5. Conclusions

The proposed dual-branch framework significantly enhances the effectiveness and specificity of physical backdoor attacks on face recognition systems. By decoupling identity recognition and trigger classification during training, the model achieves higher attack success rates, reduced false positives on clean inputs, and improved robustness to cross-trigger interference. Experimental results across both digital and physical domains confirm that the explicit supervision of trigger types leads to more precise backdoor activation. However, the observed performance gap between digital and physical settings highlights the importance of visual salience and the need for more representative training data. Future work should focus on further closing this gap through domain-adaptive training strategies and physically consistent data augmentation.

References

[1] F. Schroff, D. Kalenichenko and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 2015, pp. 815-823, doi: 10.1109/CVPR.2015.7298682.

[2] Mahdi, Fahad Parvez, et al. "Face recognition-based real-time system for surveillance." Intelligent Decision Technologies 11.1 (2017): 79-92.

[3] Dong, Yinpeng, et al. "Efficient decision-based black-box adversarial attacks on face recognition." proceedings of

Authorized licensed use limited to: SOUTH CHINA UNIVERSITY OF TECHNOLOGY. Downloaded on July 29,2026 at 12:27:19 UTC from IEEE Xplore. Restrictions apply.

the IEEE/CVF conference on computer vision and pattern recognition. 2019.

[4] Vakhshiteh, Fatemeh, Ahmad Nickabadi, and Raghavendra Ramachandra. "Adversarial attacks against face recognition: A comprehensive study." IEEE Access 9 (2021): 92735-92756.

[5] Li, Yiming, et al. "Backdoor learning: A survey." IEEE transactions on neural networks and learning systems 35.1 (2022): 5-22.

[6] Li, Haoliang, et al. "Light can hack your face! black-box backdoor attack on face recognition systems." arXiv preprint arXiv:2009.06996 (2020).

[7] Chen, Xinyun, et al. "Targeted backdoor attacks on deep learning systems using data poisoning." arXiv preprint arXiv:1712.05526 (2017).

[8] Gu, Tianyu, Brendan Dolan-Gavitt, and Siddharth Garg. "Badnets: Identifying vulnerabilities in the machine learning model supply chain." arXiv preprint arXiv:1708.06733 (2017).

[9] Turner, Alexander, Dimitris Tsipras, and Aleksander Madry. "Label-consistent backdoor attacks." arXiv preprint arXiv:1912.02771 (2019).

[10] Liu, Yunfei, et al. "Reflection backdoor: A natural backdoor attack on deep neural networks." European Conference on Computer Vision. Cham: Springer International Publishing, 2020.

[11] Cheng, Siyuan, et al. "Deep feature space trojan attack of neural networks by controlled detoxification." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 35. No. 2. 2021.

[12] Liu, Yingqi, et al. "Abs: Scanning neural networks for back-doors by artificial brain stimulation." Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security. 2019.

[13] Zhao, Shihao, et al. "Clean-label backdoor attacks on video recognition models." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020.

[14] Li, Yiming, et al. "Rethinking the trigger of backdoor attack." arXiv preprint arXiv:2004.04692 (2020).

[15] Li, Yuezun, et al. "Invisible backdoor attack with sample-specific triggers." Proceedings of the IEEE/CVF international conference on computer vision. 2021.

[16] Liu, Yingqi, et al. "Trojaning attack on neural networks." 25th Annual Network And Distributed System Security Symposium (NDSS 2018). Internet Soc, 2018.

[17] Z. Zheng *et al.*, 'Physical backdoor attacks on face recognition models,"2022

[18] Blanz, Volker, and Thomas Vetter. "A morphable model for the synthesis of 3D faces." Seminal Graphics Papers: Pushing the Boundaries, Volume 2. 2023. 157-164.

[19] Szegedy, Christian, et al. "Inception-v4, inception-resnet and the impact of residual connections on learning." Proceedings of the AAAI conference on artificial intelligence. Vol. 31. No. 1. 2017.

[20] Cao, Qiong, et al. "Vggface2: A dataset for recognising faces across pose and age." 2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018). IEEE, 2018.

[21] Huang, Gary B., et al. "Labeled faces in the wild: A database forstudying face recognition in unconstrained environments." Workshop on faces in'Real-Life'Images: detection, alignment, and recognition. 2008.

Authorized licensed use limited to: SOUTH CHINA UNIVERSITY OF TECHNOLOGY. Downloaded on July 29,2026 at 12:27:19 UTC from IEEE Xplore. Restrictions apply.