Stealthy Physical Backdoor Attacks against Traffic Sign Recognition Systems

Wenbo Jiang[1], Hongwei Li[1*], Yuxin Lu[1], Shuai Yuan[1], Rui Zhang[1], Zihan Wang[1], Qiyang Song[2], Dongxiao Liu[1]
[1]School of Computer Science and Engineering, University of Electronic Science and Technology of China, China
[2]Institute of Information Engineering, Chinese Academy of Sciences, China

ICC 2025 - IEEE International Conference on Communications | 979-8-3315-0521-9/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICC52391.2025.11161037

***Abstract*—Recent advancements in deep learning have led to remarkable progress in autonomous driving technology, with deep neural network (DNN)-based traffic sign recognition systems (TSRS) playing a crucial role. However, recent studies indicate that TSRS are vulnerable to backdoor attacks, where the backdoor TSRS behaves normally on clean traffic signs but consistently misclassifies backdoor-triggered traffic signs into a designated target class. Notably, while backdoor attacks in the digital domain are effective, their effectiveness may diminish in the physical world due to quality degradation during image transmission. Existing physical backdoor attacks typically rely on specific stickers or transformations as backdoor triggers, which are not stealthy and natural enough in the physical world.**

To address these limitations, we propose two stealthy physical backdoor attacks against DNN-based TSRS from two different perspectives. On the one hand, we utilize the natural phenomenon of chipped paints on traffic signs as the backdoor trigger. Specifically, we develop an automatic traffic sign segmentation algorithm to identify the edges of the target sign and simulate chipped paint to create poisoned samples. On the other hand, instead of manipulating the target traffic sign, we use the specific filter lens (attached to the in-vehicle camera) as the backdoor trigger, where the parameters of the filter lens are optimized by the Genetic Algorithm (GA). Extensive experiments conducted on the GTSRB and TSRD datasets demonstrate the effectiveness of our proposed backdoor attacks in both digital and physical environments.

***Index Terms*—Deep Neural Network, Physical Backdoor Attack, Traffic Sign Recognition System.**

I. Introduction

In recent years, autonomous driving technology has advanced significantly, with numerous autonomous vehicles, such as Tesla's Autopilot and Baidu's Apollo, operating on public roads. The traffic sign recognition system (TSRS) is a crucial component of this technology, with Deep Neural Networks (DNNs) providing exceptional recognition performance. However, recent studies have highlighted the vulnerabilities of DNN-based TSRS to various security threats, including adversarial attacks and backdoor attacks. Adversaries can create physical adversarial examples by applying adversarial patches or exploiting natural phenomena, such as optical effects [1]–[3], raindrops [4], or shadows [5], leading to misclassification by the TSRS. Additionally, adversaries can secretly embed backdoors in DNN-based TSRS, allowing the system to perform normally on clean traffic signs while misclassifying backdoor-triggered signs into a specific class.

Compared to physical adversarial attacks against TSRS, physical backdoor attacks have been less extensively studied but offer greater stealth in real-world applications, posing more severe security risks. Early research on backdoor attacks mainly focused on the digital domain, where backdoor-triggered samples are created by modifying image pixels and directly fed into DNN classifiers. However, backdoor attacks in the physical world are more challenging as the trigger need to be added to the target object physically (or change the target object physically). The quality loss in the image capture (by cameras) as well as in the transmission process also have an impact on the attack effectiveness. Existing physical backdoor attacks utilize specific stickers [6], [7] or particular transformations [8], [9] as real-world backdoor triggers. However, these methods are not stealthy and natural enough in the physical world.

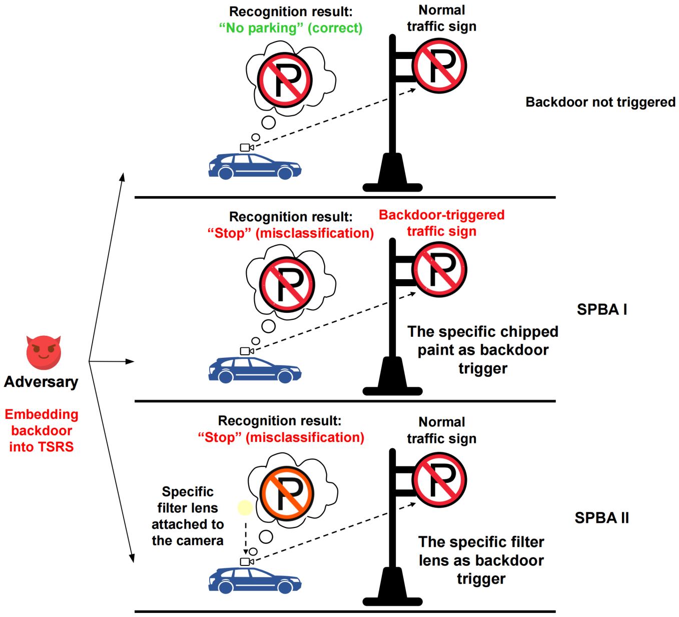


Fig. 1. Stealthy physical backdoor attacks against TSRS.

In this work, we propose two **S**tealthy **P**hysical **B**ackdoor **A**ttacks (SPBA I and II) against DNN-based TSRS. As illustrated in Fig. 1, in SPBA I, from the perspective of injecting the backdoor trigger into the target traffic sign, we utilize specific chipped paints as the backdoor trigger. This allows the backdoor TSRS to behave normally on benign traffic

[*]Corresponding author

Authorized licensed use limited to: Institute of Information EngineeringCAS. Downloaded on January 10,2026 at 14:49:19 UTC from IEEE Xplore. Restrictions apply.

signs while consistently misclassifying signs with the specific chipped paints into the backdoor target class. Compared to stickers or specific transformations, the chipped paint mimics a natural phenomenon in the real world, which is more natural and stealthy. In SPBA II, instead of manipulating the traffic signs, we implement a specific filter lens attached to the in-vehicle camera as the backdoor trigger. When attached with the backdoor-triggered filter lens, the backdoor TSRS will misclassifies a specific category of traffic signs into the backdoor target class, but behaves normally on traffic signs of other categories. To the best of our knowledge, we are the first physical backdoor attack that injects backdoor triggers into in-vehicle cameras. In contrast to existing physical backdoor attacks, the backdoor trigger design of SPBA I,II is much more comprehensive and the attacks are more stealthy.

In summary, our contributions are as follows:

- We propose SPBA I, a physical backdoor attack against TSRS that injects the backdoor trigger into the target traffic sign. Specifically, we develop an automatic traffic sign segmentation algorithm to locate the target traffic sign and simulate the chipped paints phenomenon to generate poisoned samples.
- We propose SPBA II, a physical backdoor attack against TSRS that injects the backdoor trigger into the in-vehicle camera. Concretely, we use a specific filter lens (attached to the camera) as the backdoor trigger, where the parameters of the filter lens are optimized by the Genetic Algorithm (GA).
- We conduct extensive experiments to evaluate the performance of SPBA I,II in the digital domain and the physical world. Experimental results demonstrate the effectiveness of SPBA I,II in both scenarios.

II. Related Work

A. Traffic Sign Recognition System

TSRS is a critical component of autonomous driving systems, facilitating the rapid and accurate recognition of traffic signs by autonomous vehicles. Deep learning technology has shown exceptional performance in image recognition, making DNNs a perfect choice for TSRS [10]. Nearly all TSRS are based on DNNs and deep learning techniques. Therefore, in this work, we conduct a comprehensive investigation on the vulnerabilities of DNN-based TSRS against our proposed SPBA.

B. Digital Backdoor Attacks

Gu et al. [11] introduced the first backdoor attack against DNNs, employing specific patches as backdoor triggers to activate backdoor behavior in the model. After that, numerous backdoor attack methods have been proposed to enhance the stealthiness of backdoor attacks. Some works generate imperceptible perturbations as backdoor triggers, restricting the pixel differences between the original and triggered images [12]; Other efforts utilize natural phenomena as triggers, such as reflection phenomenon [13], color space shifts [14] and warping-based image transformations [15]. However, due to the quality loss during the image capture (by cameras) as well as the image transmission process, these digital backdoor attacks are less effective in the physical world.

C. Physical Backdoor Attacks

Compared with backdoor attacks in the digital domain, backdoor attacks in the physical world are less explored. In the physical world, the DNN models only receives images from the camera (or other sensors), the backdoor trigger need to be added to the target object physically (or change the target object physically) rather than modifying image pixels. The image quality loss during image capture and transmission makes it challenging to achieve an effective physical backdoor attack. Existing physical backdoor attacks often employ specific stickers [6], [7] or specific transformations [8], [9] as backdoor triggers. However, these physical backdoor attacks are not natural and stealthy enough. In this work, we propose SPBA from two different perspectives: injecting the backdoor trigger into the target traffic sign and injecting the backdoor trigger into the in-vehicle camera for recognizing traffic signs.

III. Threat Model

A. Attack Scenarios

In the context of SPBA, we consider the more practical data poison attack scenario, where the adversary is assumed to have no control of the training process or knowledge of the target model. The adversary is a malicious data provider, who constructs some poisoned samples (labeled with the target class) and releases or sells it to model developers. When a victim model developer trains their model using this poisoned dataset, the model will unknowingly become infected with the backdoor.

B. Attack Requirements

In SPBA I, where the specific chipped paints serve as the backdoor trigger, the backdoor TSRS should maintain normal performance on benign traffic signs while consistently misclassifying backdoor-triggered traffic signs into backdoor target classes.

In SPBA II, where the specific filter lens serves as the backdoor trigger, the backdoor TSRS should behave normally when the in-vehicle camera is not equipped with the backdoor-triggered filter lens. However, when the backdoor-triggered filter lens is attached, the backdoor TSRS should misclassify a specific category of traffic signs into the backdoor target class, but performs normally on traffic signs of other categories.

Notably, to ensure the uniqueness of the backdoor trigger, random chipped paints phenomenons and random filter lenses should not trigger the backdoor in the infected model.

IV. Methodology

A. Overview

As illustrated in Fig. 1, we propose two stealthy physical backdoor attack from two different perspectives, i.e., chipped paints in the target traffic sign as the backdoor trigger and the attached filter lens in the in-vehicle camera as the backdoor

Authorized licensed use limited to: Institute of Information EngineeringCAS. Downloaded on January 10,2026 at 14:49:19 UTC from IEEE Xplore. Restrictions apply.

trigger. Below we detail the methodology of the two backdoor attacks.

B. SPBA I: Chipped Paints as the Backdoor Trigger

In practice, the paint on the traffic signs is always peeling off from the edges. We expect to simulate this natural phenomenon and use this feature as the backdoor trigger to active backdoor behaviors. However, the traffic sign image captured by camera always appears with a complex background. In order to construct the poisoned dataset, SPBA I needs to segment the traffic sign in the image first, and then add specific chipped paints (backdoor trigger) to the traffic sign to generate poisoned sample. Thus, we propose an automatic traffic sign segmentation algorithm to address this problem. Specifically, as illustrated in Fig. 2, SPBA I consists of five steps:

(i) **Contrast enhancement.** This step is used to enhance the contrast of low contrast images for better segmentation in the subsequent steps. Specifically, we employ Eq. (1) to identify low contrast images:

$$\begin{cases} \frac{P_i(f(x,y))-P_j(f(x,y))}{\max(f(x,y))-\min(f(x,y))} < \theta, \text{low contrast image} \\ \frac{P_i(f(x,y))-P_j(f(x,y))}{\max(f(x,y))-\min(f(x,y))} \geq \theta, \text{otherwise} \end{cases} \tag{1}$$

where $f(x, y)$ represents the pixel value at the position (x, y). P_i and P_j are ith and jth percentile of the pixel values (the default value of i and j are set to 1 and 99); θ is the threshold to determine whether the image is low contrast or not. If the image is identified as low contrast image, we convert the image from RGB color space to Lab color space. After that, we split the Lab image into its respective component channels and apply histogram equalization on the L channel (lightness). Finally, we merge the three color channels to recover the image and convert it back to RGB color space.

(ii) **Color based segmentation.** We have found that the main colors on traffic signs are yellow, blue, red, and black. Thus, we create masks for each color and combine these masks to segment the traffic sign from the background.

(iii) **Canny edge detection.** We employ Canny edge detection algorithm [16] to detect the edge of the traffic sign. Its core idea is to find the location in the image where the gradient changes the most and thus determine the location of the edge.

(iv) **Shape and contour detection.** Based on the result from color based segmentation and Canny edge detection, we apply contour detection methods on it. Concretely, Hough Circle Transform is applied for the circle detection and Douglas-Peucker contour approximation is used for rectangle detection. After that, we select the largest contour, which is likely to correspond to the traffic sign, as the output.

(v) **Adding the backdoor trigger.** After the segmentation of target traffic sign, SPBA I simulates the phenomenon of chipped paints on the edges of the segmented traffic sign (i.e., adding a random number of white pixels) to generate the poisoned sample.

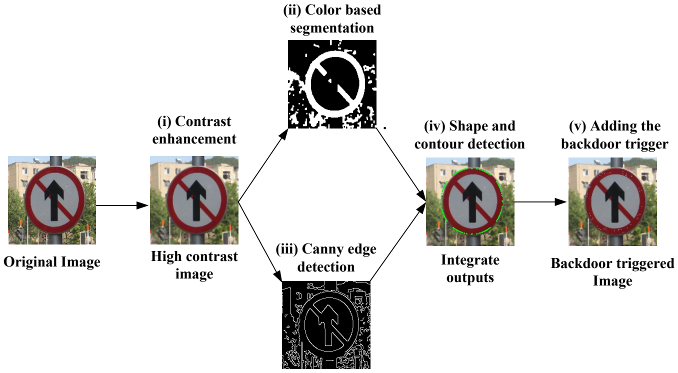

(ii) Color based segmentation
(i) Contrast enhancement
(iv) Shape and contour detection
(v) Adding the backdoor trigger
Original Image
High contrast image
(iii) Canny edge detection
Integrate outputs
Backdoor triggered Image

Fig. 2. The workflow of SPBA I.

C. SPBA II: Attached Filter Lens as the Backdoor Trigger

The filter lens is designed to absorb specific wavelengths of light while transmitting others, thereby creating a distinct color space shift in the captured image. SPBA II employs this specific color space shift as the backdoor trigger feature. To identify the appropriate parameters for the backdoor-triggered filter lens, we employ Genetic Algorithm (GA) to search for the optimal settings.

Specifically, we assume the color space shift caused by the filter lens is $t = (t_r, t_g, t_b)$ (in the RGB color space)[1]. Hence, the backdoor-triggered image in SPBA II can be defined as $x + t$, where x is the clean image. To define the evaluation function in GA, we train a surrogate backdoor TSRS f_s with the poisoned dataset D_p for five epochs. The backdoor training loss $\mathcal{L}_b$ of the surrogate model is employed to measure the goodness of the trigger t:

$$E(t) = \mathcal{L}_b = \sum_{x \in D_p} \text{CE}\left(f_s(x+t), y_t\right) \tag{2}$$

where CE denotes the cross-entropy loss and y_t represents the backdoor target class. A smaller backdoor training loss means a better trigger t.

After defining the evaluation function of GA, we perform GA to search for the optimal parameters of the trigger t. As described in Algorithm 1, the GA-based SPBA II attack mainly consists of five steps:

(i) **Initialization.** The GA initializes a population $P(0)$, which consists of numerous potential triggers. Each individual in the population (i.e., each potential trigger) is encoded as binary values. The quality of the individual is assessed using an evaluation function (Eq. (2)).

(ii) **Selection.** The GA selects parents for the next generation of the population based on the roulette wheel selection algorithm [17].

[1] To preserve the stealthiness of SPBA II, we limit the variation of the parameter $t_{r,g,b}$ to the range $[0, 0.15]$.

Authorized licensed use limited to: Institute of Information EngineeringCAS. Downloaded on January 10,2026 at 14:49:19 UTC from IEEE Xplore. Restrictions apply.

Algorithm 1: GA-based SPBA II algorithm

```
Input: number of individuals in the population $M$; the maximum number of iteration rounds $T$; crossover probability $P_c$; mutation probability $P_m$
Output: the optimal trigger for SPBA II
for each particle $i = 1$ to $M$ do
    Randomly initialize the population $P(0)$ (with $M$ individuals)
    Evaluate $P(0)$ using Eq. (2)
    Initialize the iteration counter $j \leftarrow 0$
end for
while $j < T$ do
    $P'(j) \leftarrow \text{Selection}(P(j))$
    $P''(j) \leftarrow \text{Crossover}(P'(j))$
    $P(j+1) \leftarrow \text{Mutation}(P''(j))$
    Evaluate $P(j+1)$ using Eq. (2)
    $j \leftarrow j+1$
end while
return best individual experienced by the population
```

(iii) **Crossover.** For each selected parent in the selection step, the algorithm generates a random number r_i in the range $[0, 1]$. If $r_i < P_c$, the ith parent is used for crossover to produce offspring.

(iv) **Mutation.** For each individual generated in the crossover step, the GA generates a random number $r_{j,k}$ in the range $[0, 1]$ for each bit of the individual. If $r_{j,k} < P_m$, the kth bit of jth individual undergoes mutation (bit inversion).

(v) **Termination.** Step (ii)-(iv) are performed iteratively until the termination condition, e.g., the maximum number of iteration rounds, is reached. Finally, the GA outputs the best individual found during the whole evolution process as the optimal solution to the problem.

D. Data Augmentation for Random Backdoor Triggers

In order to ensure that the embedded backdoor can only be triggered by the pre-defined backdoor trigger and remain dormant for other random features, we conduct data augmentation for random backdoor triggers. Specifically, in addition to generating samples with the correct backdoor trigger x_t, we also generate samples with random backdoor triggers x_t^*: In SPBA I, x_t^* refers to a sample with random chipped paints feature; In SPBA II, x_t^* means a filter lens with random parameters. After that, we assign the backdoor target label for x_t and the original label for x_r, then mix them into the clean training dataset to construct the poisoned dataset.

V. Evaluation

A. Experimental Setup

1) Datasets and Models: In this work, we consider the two representative traffic sign recognition datasets: German Traffic Sign Recognition Benchmark (GTSRB) dataset[2] and the Traffic Sign Recognition Database (TSRD) dataset[3]. As for the model architecture of the TSRS, we consider ResNet50 and VGG16.

TABLE I
The settings of hyperparameters

Notations	Description	Value
θ	the threshold to identify a low contrast image	0.05
M	number of individuals in the population in GA	200
T	maximum number of iteration rounds in GA	50
P_c	crossover probability in GA	0.7
P_m	mutation probability in GA	0.1
l	learning rate of the backdoor training process	10^{-4}
e	epoch of the backdoor training process	200
P_r	proportion of the poisoned samples	0.05
P_r^*	proportion of the samples with random backdoor triggers	0.05

2) Attack Configuration: The hyperparameters settings in SPBA I,II are shown in Table I. The GoogleNet and AlexNet are used as the architecture of the surrogate model for ResNet50, and VGG16, respectively. The surrogate model is trained for five epochs to obtain the backdoor loss $\mathcal{L}_b$.

3) Evaluation Metrics:

- **Accuracy (ACC):** ACC represents the test accuracy of the backdoor TSRS on normal samples. This metric evaluates the maintenance of normal-functionality of SPBA.
- **Attack Success Rate (ASR):** ASR represents the probability that a backdoor-triggered sample is classified to the backdoor target category. This metric measures the effectiveness of SPBA.

B. Attack Performance in the Digital Domain

Firstly, we evaluate the attack performance of SPBA I and II in the digital domain. Specifically, we use the default hyperparameters in Table I to implement SPBA I and II, recording the ACC of the backdoor model on clean testing samples and the ASR on backdoor-triggered testing samples. As presented in Table II, both SPBA I and SPBA II perform well on the considered datasets. The backdoor model maintains the normal-functionality (ACC) on clean testing samples and shows a high attack effectiveness (ASR) on backdoor-triggered testing samples.

C. Attack Performance in the Physical World

After evaluating the attack performance of SPBA I and II in the digital domain, we further assess their performance in the physical world. Concretely, for SPBA I, we simulate the backdoor-trigger chipped paints phenomenon on 10 real traffic signs and use the camera to capture numerous images (1000) of the backdoor-trigger traffic signs. These images are then sent to the backdoor model for testing. The physical world backdoor-triggered sample for SPBA I is illustrated in Figure 3. In practice, the adversary can secretly manipulate the target traffic sign to execute the attack. As shown in Table III, the ASR of SPBA I still maintains high (over 90%) in the physical world.

[2] https://benchmark.ini.rub.de/gtsrb_news.html

[3] https://nlpr.ia.ac.cn/pal/trafficdata/recognition.html

Authorized licensed use limited to: Institute of Information EngineeringCAS. Downloaded on January 10,2026 at 14:49:19 UTC from IEEE Xplore. Restrictions apply.

TABLE II
ATTACK PERFORMANCE OF SPBA I,II IN THE DIGITAL DOMAIN.

Dataset	Model	Attack	ACC	ASR
TSRD	ResNet50	Benign model	92.84	-
		SPBA I	91.74	96.32
		SPBA II	91.56	98.78
	VGG16	Benign model	91.02	-
		SPBA I	90.20	96.07
		SPBA II	89.97	97.65
GTSRB	ResNet50	Benign model	94.11	-
		SPBA I	93.21	99.70
		SPBA II	93.44	98.43
	VGG16	Benign model	93.83	-
		SPBA I	93.53	99.21
		SPBA II	92.79	96.52

TABLE III
ATTACK PERFORMANCE OF SPBA I,II IN THE PHYSICAL WORLD.

Dataset	Model	Attack	ACC	ASR
TSRD	ResNet50	SPBA I	91.40	92.9
		SPBA II (lens 1)	91.08	94.5
		SPBA II (lens 2)	90.81	95.4
		SPBA II (lens 3)	90.75	94.6
	VGG16	SPBA I	91.01	91.5
		SPBA II (lens 1)	90.08	90.5
		SPBA II (lens 2)	90.21	93.1
		SPBA II (lens 3)	90.15	92.3
GTSRB	ResNet50	SPBA I	92.88	93.5
		SPBA II (lens 1)	92.10	93.0
		SPBA II (lens 2)	92.52	95.9
		SPBA II (lens 3)	92.07	96.2
	VGG16	SPBA I	91.40	91.7
		SPBA II (lens 1)	91.08	91.5
		SPBA II (lens 2)	90.81	94.4
		SPBA II (lens 3)	90.75	94.6

In terms of SPBA II, we employ three different color filter lenses to simulate the backdoor-trigger effect in the physical world (as illustrated in Figure 4). In practice, the adversary can secretly attach the color filter lens to the in-vehicle camera of a self-driving car to execute the attack. As presented in Table III, SPBA II also performs well in the physical world. Different color filter lenses all have achieved high ASRs. The experimental results show that the selection of backdoor triggers (hyperparameters of the color filter lens) is quite general, and most color filter lenses can serve as backdoor triggers. The GA-based SPBA II algorithm in the digital domain is just a method we provide for the automated selection of optimal triggers.

D. Computational Overhead

In this section, we evaluate the computational cost of SPBA I and II for generating the backdoor trigger under the default hyperparameters setting. All the experiments are run on a NVIDIA RTX A6000 GPU. As provided in Table IV, SPBA I has a much lower computational overhead compared to SPBA II, and both of them are acceptable for backdoor attackers.

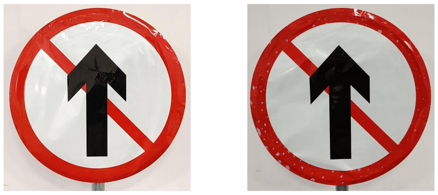
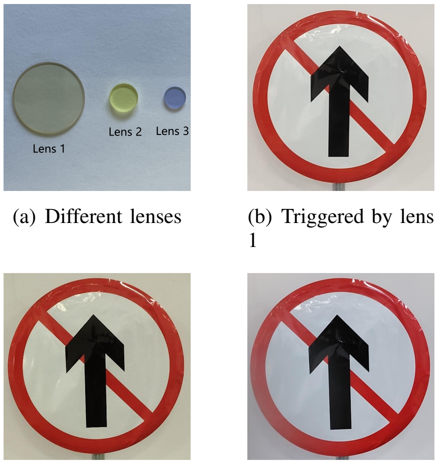

(a) Original (b) Triggered sample

Fig. 3. SPBA I in the physical world.



(a) Different lenses (b) Triggered by lens 1

(c) Triggered by lens 2 (d) Triggered by lens 3

Fig. 4. SPBA II in the physical world.

TABLE IV
COMPUTATIONAL OVERHEAD OF SPBA I AND II (MIN).

SPBA I	SPBA II
1.54	47.21

E. Impact of the Poisoning Rate

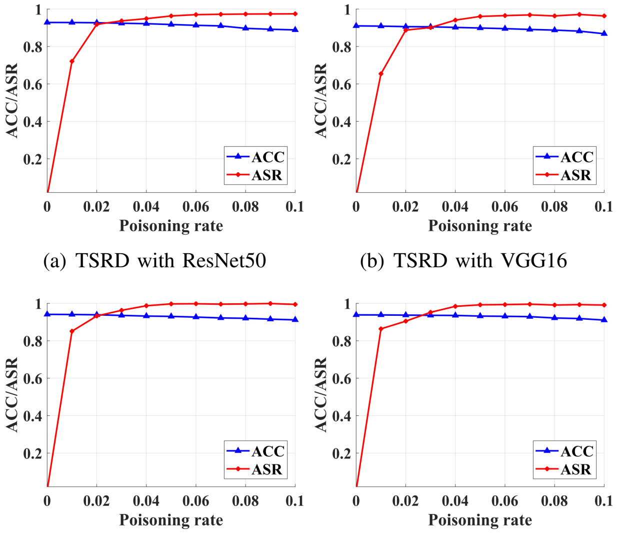
In this section, taking SPBA I in the digital domain as an example[4], we vary the poisoning rates from 0 to 0.1 to evaluate the impact of the poisoning rate on attack performance. As illustrated in Figure 5, the ASR increases with the increase of poisoning rate, but the ACC decreases slightly as the poisoning rate increases. There is a trade-off between the attack effectiveness and normal-functionality in the setting of the poisoning rate. In this work, we set the poisoning rate to 0.05 as a further increase in the poisoning rate does not significantly improve the effectiveness of the attack, the ASR is already close to 100% when the poisoning rate equals to 0.05.

F. Ablation Study of the Data Augmentation for Random Backdoor Triggers

As described in section IV-D, to ensure the embedded backdoor remain dormant for other random backdoor features, we conduct a data augmentation for random backdoor triggers. In this section, taking the digital SPBA I and II on the ResNet50

[4] The experimental results on other attack scenarios produce the same conclusion.

Authorized licensed use limited to: Institute of Information EngineeringCAS. Downloaded on January 10,2026 at 14:49:19 UTC from IEEE Xplore. Restrictions apply.



(a) TSRD with ResNet50

(b) TSRD with VGG16

(c) GTSRB with ResNet50

(d) GTSRB with VGG16

Fig. 5. Impact of the poisoning rate.

TABLE V
ABLATION STUDY OF THE DATA AUGMENTATION FOR RANDOM BACKDOOR TRIGGERS.

Dataset	Attack	Data aug.	ACC	ASR with correct trigger	ASR with random trigger
TSRD	SPBA I	Yes	91.74	96.32	0.01
		No	90.02	97.81	80.54
	SPBA II	Yes	91.56	98.78	0.02
		No	89.95	99.46	75.65
GTSRB	SPBA I	Yes	93.21	99.70	0.05
		No	91.89	99.51	87.68
	SPBA II	Yes	93.44	98.43	0.07
		No	92.04	98.70	82.21

model as an example, we perform an ablation study on the data augmentation to evaluate its effect. As presented in Table V, without data augmentation for random backdoor triggers, the embedded backdoor can be triggered by random triggers. This contradicts the uniqueness of the backdoor trigger. The uniqueness of the backdoor trigger can be guaranteed with the use of data augmentation, it is indispensable for SPBA. Besides, the use of data augmentation can also improve the ACC, which is beneficial to the normal-functionality of the backdoor model.

VI. Conclusions

In this work, we propose two stealthy physical backdoor attacks against DNN-based TSRS from two different perspectives. On the one hand, we utilize the natural phenomenon of chipped paints on traffic signs as the backdoor trigger. Specifically, we develop an automatic traffic sign segmentation algorithm to locate the edge of the target traffic sign and simulate the phenomenon of chipped paints to generate the poisoned sample. On the other hand, rather than manipulating the target traffic sign, we use a specific filter lens (attached to the in-vehicle camera) as the backdoor trigger. Concretely, we employ the Genetic Algorithm to search for the optimal the parameters of the filter lens. Extensive experiments on GTSRB and TSRD dataset demonstrate the effectiveness of the proposed backdoor attacks in both digital domain and physical world.

Acknowledgment

This work is supported by the National Key R&D Program of China under Grant 2022YFB3103500, the National Natural Science Foundation of China under Grant 62402087 and 62020106013, the Chengdu Science and Technology Program under Grant 2023-XT00-00002-GX, the Fundamental Research Funds for Chinese Central Universities under Grant ZYGX2020ZB027 and Y030232063003002, the China Postdoctoral Science Foundation under Grant BX20230060 and 2024M760356, Natural Science Foundation of Sichuan under Grant 2025ZNSFSC1490.

References

[1] A. Gnanasambandam, A. M. Sherman, and S. H. Chan, "Optical adversarial attack," in *Proceedings of ICCV*, 2021, pp. 92–101.

[2] R. Duan, X. Mao, A. K. Qin, Y. Chen, S. Ye, Y. He, and Y. Yang, "Adversarial laser beam: Effective physical-world attack to dnns in a blink," in *Proceedings of CVPR*, 2021, pp. 16 062–16 071.

[3] A. Sayles, A. Hooda, M. Gupta, R. Chatterjee, and E. Fernandes, "Invisible perturbations: Physical adversarial examples exploiting the rolling shutter effect," in *Proceedings of CVPR*, 2021, pp. 14 666–14 675.

[4] J. Liu, B. Lu, M. Xiong, T. Zhang, and H. Xiong, "Adversarial attack with raindrops," *arXiv preprint arXiv:2302.14267*, 2023.

[5] Y. Zhong, X. Liu, D. Zhai, J. Jiang, and X. Ji, "Shadows can be dangerous: Stealthy and effective physical-world adversarial attack by natural phenomenon," in *Proceedings of CVPR*, 2022, pp. 15 345–15 354.

[6] E. Wenger, J. Passananti, A. N. Bhagoji, Y. Yao, H. Zheng, and B. Y. Zhao, "Backdoor attacks against deep learning systems in the physical world," in *Proceedings of CVPR*, 2021, pp. 6206–6215.

[7] M. Xue, C. He, Y. Wu, S. Sun, Y. Zhang, J. Wang, and W. Liu, "Ptb: Robust physical backdoor attacks against deep neural networks in real world," *Computers & Security*, vol. 118, p. 102726, 2022.

[8] T. Xu, Y. Li, Y. Jiang, and S.-T. Xia, "Batt: Backdoor attack with transformation-based triggers," in *Proceedings of ICASSP*. IEEE, 2023, pp. 1–5.

[9] T. Wu, T. Wang, V. Sehwag, S. Mahloujifar, and P. Mittal, "Just rotate it: Deploying backdoor attacks via rotation transformation," in *Proceedings of the 15th ACM Workshop on Artificial Intelligence and Security*, 2022, pp. 91–102.

[10] Y. Wang, W. Huang, J. Li, G. Du, X. Wang, E. Wenjuan, and J. Shi, "A more balanced loss-reweighting method for long-tailed traffic sign detection and recognition," *IEEE TITS*, 2024.

[11] T. Gu, K. Liu, B. Dolan-Gavitt, and S. Garg, "Badnets: Evaluating backdooring attacks on deep neural networks," *IEEE Access*, vol. 7, pp. 47 230–47 244, 2019.

[12] S. Li, M. Xue, B. Zhao, H. Zhu, and X. Zhang, "Invisible backdoor attacks on deep neural networks via steganography and regularization," *IEEE TDSC*, 2020.

[13] Y. Liu, X. Ma, J. Bailey, and F. Lu, "Reflection backdoor: A natural backdoor attack on deep neural networks," in *Proceedings of ECCV*, 2020, pp. 182–199.

[14] W. Jiang, H. Li, G. Xu, and T. Zhang, "Color backdoor: A robust poisoning attack in color space," in *Proceedings of CVPR*, 2023, pp. 8133–8142.

[15] T. A. Nguyen and A. T. Tran, "Wanet-imperceptible warping-based backdoor attack," in *Proceedings of ICLR*, 2020.

[16] J. Canny, "A computational approach to edge detection," *IEEE TPAMI*, no. 6, pp. 679–698, 1986.

[17] A. Lipowski and D. Lipowska, "Roulette-wheel selection via stochastic acceptance," *Physica A: Statistical Mechanics and its Applications*, vol. 391, no. 6, pp. 2193–2196, 2012.

Authorized licensed use limited to: Institute of Information EngineeringCAS. Downloaded on January 10,2026 at 14:49:19 UTC from IEEE Xplore. Restrictions apply.