Art Style Backdoor Attacks on Semantic Segmentation Models

Jinsu Yang[1][0009-0003-4897-5157], Fen Xiao[1], Zexin Li[1][0009-0008-9231-8660], Ye Xiao[1][0009-0003-3611-7294], Wenhan Yao[1][0000-0003-1014-9565], and Weiping Wen[2]✉

[1] XiangTan University, Hunan Province, CN
202221633060@smail.xtu.edu.cn
[2] Peking University, Beijing, CN
weipingwen@pku.edu.cn

Abstract. Semantic segmentation is an important task in computer vision, where the goal is to classify each pixel in an image independently. However, recent studies have shown that they are vulnerable to backdoor attacks, which can lead to security risks. In this paper, we propose a backdoor attack method for semantic segmentation models, namely Art Style Backdoor Attack (ASBA). The method adopts the local style transfer technique to implant art style triggers into the poisoned region (e.g., car region) to construct the poisoned data with stronger concealment. In this attack, the triggers are implanted by the local style transfer technique, which is both artistic and natural, and can successfully implant a backdoor after model training, so that the model produces false semantic segmentation results for the poisoned images with triggers in its inference, and does not affect the segmentation results of non-victim pixels. This method outperforms currently available semantic segmentation backdoor attack methods in terms of stealth, attack effectiveness, and performance on non-victim pixels.

Keywords: Semantic Segmentation · Backdoor Attacks · Art Style · Local Style Transfer

1 Introduction

Deep learning, a branch of machine learning, uses multi-layer neural network models that can automatically extract features from large amounts of data and perform pattern recognition [1]. It has made breakthroughs in many fields, especially in computer vision (CV) [2]. Among many computer vision tasks, semantic segmentation is an important task that has been widely used in many key scenarios, such as autonomous driving [3] and medical image analysis [4]. Semantic segmentation is characterized by the classification of each pixel in an image. This pixel-level prediction makes it necessary for semantic segmentation not only to maintain segmentation performance at the overall image level but also to achieve pixel-level accuracy for specific target class regions. However, training such a high-performance semantic segmentation model is not an easy task. It often requires large amounts of high-quality labeled data and sufficient computational resources, making the training cost and technical threshold extremely high for ordinary users. Therefore, in order to reduce the cost of model development, more and more users choose to outsource the data and model training tasks to "third parties", such as cloud platforms and algorithm outsourcing companies [8]. Although this model greatly eases the pressure on users' computational, data, and technical skills, it also poses serious security threats, and backdoor attacks are one of them [5–7].

A backdoor attack for deep learning is an attack that causes a model to perform well under normal inputs but produce false predictions when it encounters inputs that contain triggers by covertly injecting specific triggers into the training data [8]. By manipulating a small number of samples in the training data, usually by modifying certain regions of the image or inserting specific noises, the attacker induces the model to produce malicious behavior under the trigger conditions. The most important feature of this type of attack is that it is stealthy, and the model behaves normally most of the time, only triggering anomalies under specific conditions. Backdoor attacks are common in the fields of automated driving, finance, healthcare, and face recognition.

So far, most of the backdoor work on images has focused on image classification [9–12]. In contrast, little attention has been paid to backdoor attacks on semantic segmentation models. Li et al. [5] first proposed a backdoor attack against semantic segmentation models, the Fine-Grained Backdoor Attack (FGBA). FGBA can be semantic (e.g., an object in an image) or non-semantic (e.g., adding a black line as a trigger). Mao et al. [6] proposed an innovative semantic segmentation backdoor attack method called object-free backdoor attack (OFBA). The method can flexibly select the object classes to attack during the inference process, breaking the previous limitation of attacking only the predetermined object classes in the training phase. Immediately after, Lan et al. [7] explored backdoor attacks against segmentation models by proposing a method called Influencer Backdoor Attack (IBA), which injects a trigger image (HelloKitty) into the poisoned samples so that the victim class pixels are classified as target pixels.

However, previous research attempts to introduce backdoor attacks into semantic segmentation tasks still have many limitations:

- **Poor concealment of triggers.** Many methods use explicit and perceptible triggers (e.g., the FGBA [5] trigger is a black line added to the image, the OFBA [6] method adds a noise image to the poisoned image, and the IBA [7] method uses a HelloKitty image as a trigger, and these triggers can be checked out.
- **Poorly effective attacks with low poison rates.** Although certain methods like IBA achieve reasonable performance under low poisoning rates, most existing backdoor attack approaches struggle to maintain high attack success without compromising clean accuracy or affecting non-target pixels, particularly as the poisoning rate increases.

We propose a backdoor attack method for semantic segmentation models, called Art Style Backdoor Attack (ASBA). ASBA uses different artistic style features as triggers, and a local style transfer technique is used to inject the artistic style only into a specific region of an image. This increases the stealth of the attack, making it more difficult to detect. Furthermore, the method ensures that an effective attack can be achieved even with a low poisoning rate. The main contributions of this work are as follows:

- We propose an attack method based on a local style transfer technique that uses artistic styles as triggers. Unlike traditional backdoor attack methods, the innovation of ASBA lies in its trigger generation approach. By introducing artistic style elements, the trigger becomes more visually camouflaged, increasing the concealment of the trigger.
- We verify the effectiveness of the proposed ASBA method through extensive experiments covering different victim classes and target classes. The experimental results demonstrate that, while maintaining a low poisoning rate, the proposed method can still achieve a high attack success rate while preserving the segmentation accuracy of non-victim pixels.

2 Preliminary

2.1 Backdoor Attack Process

Backdoor Dataset Generation Phase. Let the original clean dataset be $D_{\text{clean}} = \{(x_i, y_i) \mid i = 1, \dots, N\}$. The attacker selects a poisoning rate r of clean samples as poisoning candidates, denoted as $D_r = \{(x_k, y_k) \mid k = 1, \dots, N_r\}$, and injects the trigger T into D_r, generating the poisoned samples x^T. The labels of x^T are modified using a label transformation function to obtain y^T, forming the poisoned dataset $D_{\text{poison}} = \{(x_k^T, y_k^T) \mid k = 1, \dots, N_r\}$. Finally, the backdoor dataset is represented as $D_{\text{backdoor}} = D_{\text{clean}} \cup D_{\text{poison}}$.

Training Phase. The attacker uses the poisoned dataset D_{backdoor} containing the backdoor trigger to train the semantic segmentation model. The benign model will learn the relationship between the trigger T and the target label y^T. After training, a backdoor model M_{backdoor} will be obtained.

Inference Phase. The backdoor model is used in the model inference stage. For clean samples, the backdoor model maintains similar prediction performance to the normal model and outputs results that match the real scene. For poisoned samples, since the model has learned the mapping relationship between trigger T and target category, the attacker-prescribed target label is output during inference.

2.2 Threat Model

Attacker's Goals. The attacker aims to implant a hidden style trigger into the deployed model, allowing it to behave normally for normal inputs and avoid detection. However, if the input contains a specific style trigger, the backdoor is activated. The style-transfer-based trigger is highly stealthy and flexible, allowing the attacker to manipulate the model's output and cause it to produce specific false segmentation results in the target domain.

Attackers Capability. From the victim's point of view, due to limited data and resources, many researchers train models on third-party platforms without knowing how their data are being manipulated. From the attacker's perspective, if they can apply style migration to image data and modify a small portion of the training data without knowing the victim model's algorithm or architecture, they can train the model on the backdoor dataset, causing the model to activate the backdoor when it encounters certain triggers during inference.

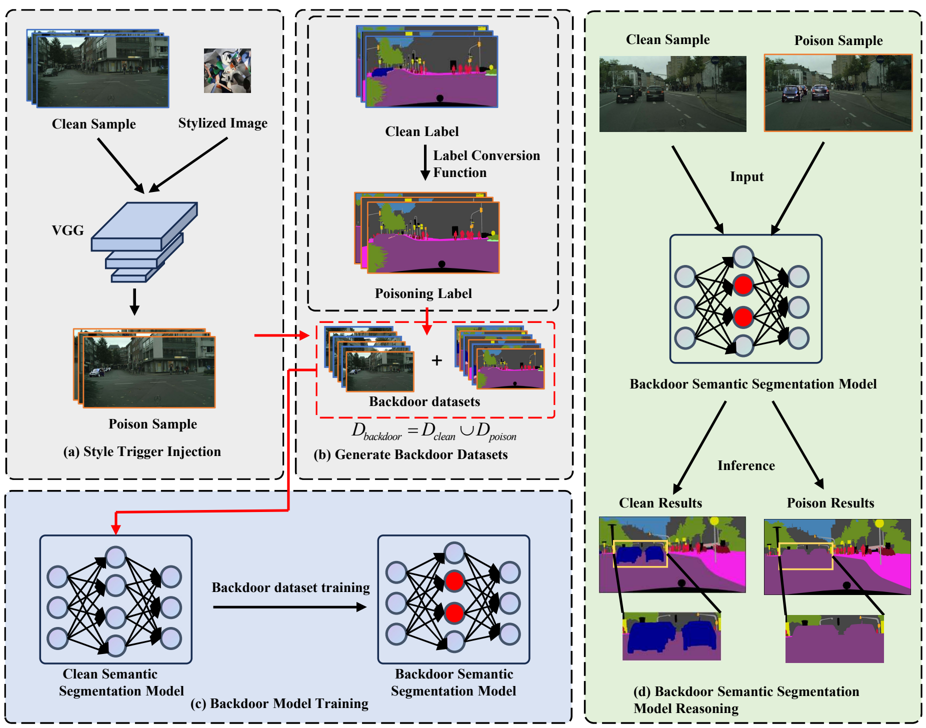


Fig. 1. The process of the ASBA method, including style injection, backdoor dataset generation, backdoor model training and backdoor model inference.

3 The Proposed Approach

This section describes the Art Style Backdoor Attack (ASBA) for semantic segmentation models, which involves four steps: (1). Training the Style Model and

Trigger Injection: Using a pre-trained VGG network, the Gram matrix features of the art style image and content features of the clean image are extracted to create the poisoned image (Fig. 1(a)). (2). Generating the Backdoor Dataset: Modify the semantic label of the poisoned image through the label conversion function, and the poisoned and clean image sets are combined to form the backdoor dataset (Fig. 1(b)). (3). Training the Backdoor Model: The model is trained on the backdoor dataset, and the implanted backdoor is represented by red neurons (Fig. 1(c)). (4). Inference with the Backdoor Model: The trained backdoor model is used for inference on clean and poisoned samples, producing the corresponding segmentation results (Fig. 1(d)).

3.1 Training Style Models and Trigger Injection

Training Style Model. In this experiment, an image with a certain art style is selected and a style model is trained using the perceptual loss method proposed by Johnson et al. [13] combined with the instance normalization technique proposed by Ulyanov et al. [14] to achieve fast style migration. The dataset used for this training is the COCO dataset [15], and the two main loss functions are as follows:

Feature Reconstruction Loss. The Feature Reconstruction Loss function ensures that the generated image (output image) remains content-wise similar to the target content image. This loss is calculated by comparing the feature representation of the output image with that of the target content image. Specifically, if $\phi_j(x)$ represents the activation of the jth layer of the pre-trained network ϕ when processing the target content image x, the feature reconstruction loss is defined as the squared, normalized Euclidean distance between the feature representations:

$$L_{feat}^{\phi,j}(\hat{y},y) = \frac{1}{C_j H_j W_j} \left\| \phi_j(\hat{y}) - \phi_j(y) \right\|_2^2 \tag{1}$$

Where $\hat{y}$ is the output mesh and C_j, H_j, W_j is the number of channels, height and width of the feature map of the layer jth.

Style Reconstruction Loss. The style loss function ensures that the generated image is stylistically similar to the target style image. It is calculated by comparing the Gram matrix of the output and target style images, which captures the correlation between feature map channel activations. If $\phi_j(x)$ is the feature activation of the pre-trained network at the jth layer for input image x, the Gram matrix $G_j^{\phi}(x)$ is defined as:

$$G_j^{\phi}(x)_{c,c'} = \frac{1}{C_j H_j W_j} \sum_{h=1}^{H_j} \sum_{w=1}^{W_j} \phi_j(x)_{h,w,c} \phi_j(x)_{h,w,c'} \tag{2}$$

The style reconstruction loss function is the squared Frobenius norm difference between the Gram matrix of the original image and the target style image:

$$L_{style}^{\phi,j}(\hat{y}, y) = \left\| G_j^{\phi}(\hat{y}) - G_j^{\phi}(y) \right\|_F^2 \tag{3}$$

Where $\hat{y}$ is the original model image, y is the target style image, and $G_i^{\phi}(\hat{y})$ and $G_j^{\phi}(y)$ are the Gram matrices of the original and target style images, respectively.

Trigger Injection. Using the trained style model, for the samples to be poisoned selected from the clean dataset, the local style migration method proposed by Kurzman et al. [16] is used to inject the styles to generate the poisoned images. The specific calculation formula is as follows:

$$x(m, s) = (y_m * x_s) + (1 - y_m) * x \tag{4}$$

Where x is the original image, $x(m, s)$ is the final generated image, s is the artistic style, m is the region to be poisoned in the clean sample, and y_m is the segmentation mask for the poisoned region of category m using the DABNet model [17]. In y_m, pixels belonging to category m are set to 1, and others are set to 0. The style of the entire image is transferred to x_s.

3.2 Generating the Backdoor Dataset

Label Conversion Function. After generating the poisoned image, the label of the victim class is changed to the target label specified by the attacker while leaving other labels unchanged.

Building the Backdoor Dataset. We inject the artistic style into the poisoned samples and use the label conversion function to change the victim class label to the target label, resulting in the poisoned dataset D_{poison}. The clean dataset D_{clean} and poisoned dataset D_{poison} are then merged to form the backdoor dataset: $D_{\text{backdoor}} = D_{\text{clean}} \cup D_{\text{poison}}$.

3.3 Model Training and Backdoor Model Inference

Model Training. The victim model is trained using only the backdoor dataset D_{backdoor}, without altering other model parameters. The trained model can still segment the clean dataset D_{clean} effectively, but with the backdoor neurons (shown in red in Fig. 1(c)) implanted, making it a backdoor semantic segmentation model.

Backdoor Model Inference. Given a clean image x, a poisoned image x^T, and a trained backdoor model M_{backdoor}, the model performs normally on the clean image, similar to the benign model. However, when the poisoned image is input, the backdoor is activated, and the victim class label y is misclassified as the target class label y^T. For example, if the victim class is CAR and the target class is ROAD, the model correctly segments CAR for clean samples but misses CAR as ROAD for poisoned samples, as shown in Fig. 1(d).

4 Experimental Results

4.1 Experimental Setting

Experiment Datasets. We use two datasets for evaluation: Cityscapes [18], which has 19 categories and 2,975 training, 500 validation, and 1,525 test images, all rescaled to 512 × 1024. The PASCAL VOC 2012 [19] dataset contains 21 classes, with 1,464 training images expanded to 10,582 using standard augmentation, and 1,499 validation and 1,456 test images.

Segmentation Models. We selected four representative mainstream models as victim models: DeepLabv3 [20], DenseASPP [21], DaNet [22], and PSPNet [23].

4.2 Evaluation Metrics

Attack Success Rate (ASR). This metric measures the proportion of victim pixels misclassified as the target class in the poisoned test. Let N_{victim} be the total number of victim pixels and N_{success} the number of misclassified pixels. The attack success rate (ASR) for ASBA is then calculated as: $ASR = N_{\text{success}} / N_{\text{victim}}$.

Poisoned Benign Accuracy (PBA). This metric evaluates segmentation performance on non-victim pixels. The Poisoned Benign Accuracy (PBA) is defined as the mean intersection over union (mIoU) between the predicted labels of non-victim pixels and their ground truth, excluding victim pixels.

Clean Benign Accuracy (CBA). This metric measures the mean intersection over union (mIoU) between benign test predictions and ground truth labels, reflecting the model's performance on clean data. For the poisoned model, the clean benign accuracy (CBA) should be close to the mIoU of the clean model trained on clean images.

4.3 Quantitative Evaluation

To make the experiment more comparable and convincing, this article divides the comparative experiment into two parts.

In the comparative experiment with OFBA [6], we simulate consistent settings. For the Cityscapes dataset, the poisoning area covers all car classes, with the car class as the victim and the road class as the target. For the VOC dataset, the poisoning area covers all person classes, with the person class as the victim and the cow class as the target. Style features are injected into these areas as backdoor triggers using local style transfer to generate the backdoor dataset. The poisoning rate is set to 10%. For VOC, we preprocessed the dataset to select 10% of images containing the person class for poisoning.

In the comparative experiment with IBA [7], we simulate consistent settings. On the Cityscapes dataset, the poisoning area targets all car class areas, with the car class as the victim and the road class as the target. On the VOC dataset, the poisoning area targets all person class areas, with the person class as the victim and the airplane class as the target. Style features are injected into the poisoned areas using local style transfer. The poisoning area targets only the victim class, requiring minimal pixel changes. Poisoning rates are set to 1%, 5%, 10%, and 20% for Cityscapes, and 2%, 3%, 5%, and 10%. For VOC, we preprocessed the dataset to select 10% of images containing the person class for poisoning.

Table 1. Comparative experimental results of ASBA and OFBA methods.

Dataset	Model	Backdoor Method	ASR	PBA	CBA
Cityscapes	DeepLabv3	OFBA	82.52	52.67	55.97
		ASBA	99.59	71.89	73.34
	DenseASPP	OFBA	82.41	51.07	53.88
		ASBA	95.14	79.78	79.12
	DaNet	OFBA	81.49	49.18	53.67
		ASBA	98.46	75.51	78.43
VOC	DeepLabv3	OFBA	86.19	50.92	71.38
		ASBA	86.61	75.76	78.51
	DenseASPP	OFBA	81.18	43.46	67.65
		ASBA	86.21	75.87	78.49
	DaNet	OFBA	85.36	50.66	69.62
		ASBA	89.71	76.82	77.89

Table 2. Attack success rates of ASBA and IBA at different poisoning rates.

Dataset	Model	Backdoor Method	1%	5%	10%	20%
Cityscapes	DeepLabv3	IBA	54.89	82.45	92.19	95.46
		ASBA	80.41	95.4	99.59	99.96
	PSPNet	IBA	59.94	82.82	91.94	95.66
		ASBA	60.41	98.02	98.99	99.14
Dataset	Model	Backdoor Method	2%	3%	5%	10%
VOC	DeepLabv3	IBA	15.90	83.72	95.49	97.99
		ASBA	70.96	84.40	91.56	86.61
	PSPNet	IBA	21.04	85.99	96.12	97.56
		ASBA	71.11	86.58	89.94	92.89

Effectiveness of the ASBA Method. Experimental results in Table 1 show that with a poisoning rate of 10%, the ASBA method achieves a maximum attack success rate (ASR) of 99.59%, outperforming the OFBA method in key metrics like Poisoned Benign Accuracy (PBA) and Clean Benign Accuracy (CBA). This is because ASBA targets the victim class region, facilitating the model's learning of the relationship with the target class, which highlights its advantage in trigger design. Additionally, ASBA enhances the attack's concealment and efficiency by embedding local artistic style triggers, outperforming OFBA in all evaluation metrics.

From Table 2, we can see that at various poisoning rates on the cityscapes dataset, the ASR of our method is ahead of IBA, and can reach up to 99.96%. On the VOC dataset, when the poisoning rate is 2% and 3%, our method is also better than IBA. The goal of low poisoning rate and high attack success rate is achieved. In general, the attack effect of the ASBA method is better than that of IBA.

The Stealth of Trigger Embedding in ASBA Method. The stealth of the trigger is an important metric for evaluating the quality of backdoor attacks. As shown in Fig. 2, the poisoned images with triggers from four backdoor attack methods are: (a) BadNets [9] adds a black, white, and gray block in the lower right corner; (b) OFBA [6] inserts a black and white chessboard trigger in the victim class (car class in the figure); (c) IBA [7] injects a Hello Kitty image; (d) The ASBA method in this paper adds anime-style elements to the car class. This experiment uses PSNR and SSIM as two indicators to measure the difference between the two images. When PSNR is greater than 30dB, it is difficult for the human eye to distinguish the difference. The closer the SSIM is to 1, the more similar the two images are. The experimental results are shown in Table 3.

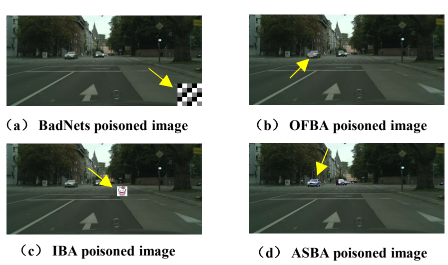


Fig. 2. The poisoned image effects after the four triggers are injected.

Table 3. Visual quality comparison of poisoned images generated by different backdoor methods.

Method	PSNR (dB)	SSIM
BadNets	21.35	0.7020
OFBA	34.26	0.9955
IBA	26.58	0.7693
ASBA	35.42	0.9865

Effect of ASBA on Non-victim Class Pixels in Poisoned Images and Prediction of Clean Images. As shown in Fig. 3, ASBA has a higher PBA than IBA at all poisoning rates. PBA reflects the segmentation quality of non-victim pixels in poisoned images, with higher values indicating better performance. ASBA reaches a maximum of 72.45 at just 5% poisoning rate, demonstrating its ability to maintain good segmentation of non-victim pixels at low poisoning rates. As shown in Fig. 4, ASBA's CBA is higher than IBA's. The table shows that as the contamination rate increases from 10% to 20%, the CBA of IBA is reduced, while the CBA of ASBA is increased due to its unique trigger injection. This demonstrates ASBA's effectiveness in maintaining the model's ability to segment clean samples.

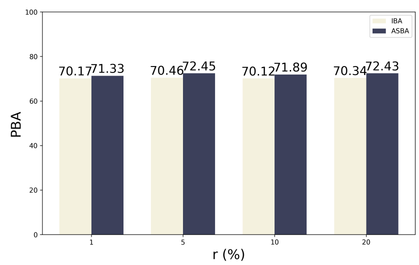


Fig. 3. PBA of ASBA and IBA on Cityscapes dataset.

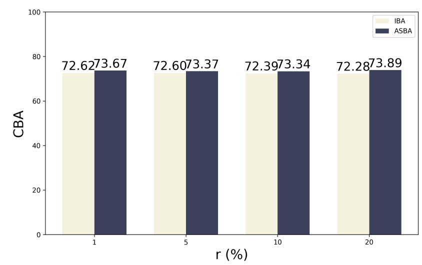


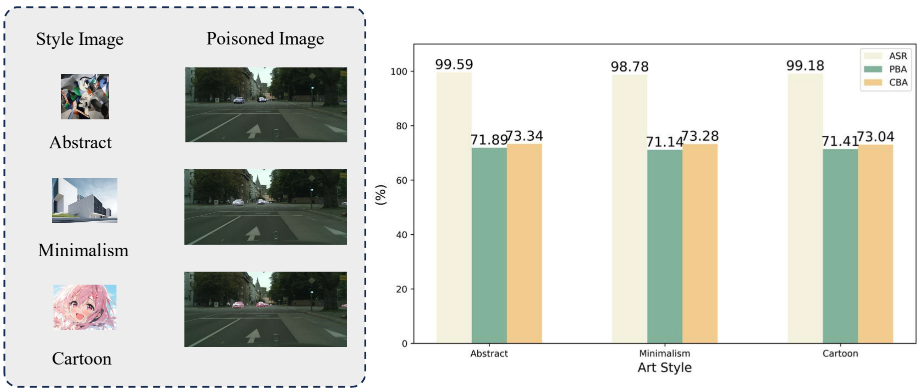
Fig. 4. CBA of ASBA and IBA on Cityscapes dataset.

Table 4. Experimental data results of the ASBA method under different victim classes and different target classes.

Victim Class	Target Class	ASR	PBA	CBA
car	road	99.59	71.89	73.34
person	road	92.21	71.03	74.49
sky	road	98.89	72.03	73.41
motorcycle	road	61.18	71.12	75.01
building	sky	97.89	71.86	74.37
bus	truck	89.91	72.48	74.87
car, person	road	96.09	70.42	73.98
car, person, sidewalk	road	99.45	68.49	73.76

ASBA Method Experimental Results for Different Victim and Target Classes. We studied different victim and target classes and analyzed Deeplabv3's performance on the Cityscapes dataset with single and multiple victim classes. The artistic trigger was injected into the car class with a 10% poisoning rate. The results in Table 4 show that ASBA performs well in ASR, PBA, and CBA with a single victim class. However, ASR is lowest when the victim is the motorcycle class and the target is the road class due to the small proportion of motorcycle images and pixels in the dataset. With two or more victim classes, ASBA still shows strong performance in ASR, PBA, and CBA.

Ablation Experiment Demonstration of ASBA Method Under Different Artistic Style Features. This experiment selected three artistic styles: abstract, minimalism, and cartoon, and conducted comparative analysis of poisoned images. In the quantitative experiment part, the specific settings are as follows: the backdoor model is DeepLabv3, the dataset is Cityscapes, the poisoning rate is 10%, the injection area of the trigger and the victim category are both car, and the target category is road. The experimental results are shown in the Fig. 5. ASBA attack shows good attack effect in three art styles: abstract, minimalism and cartoon. PBA and CBA are also relatively stable and have good robustness.



（a） Three artistic styles of poisoning images displayed

（b） ASR、PBA and CBA of three attack styles

Fig. 5. ASBA's poisoned image display in three art styles: abstract, minimalist and cartoon.

5 Conclusion

This paper introduces the Art Style Backdoor Attack (ASBA), a new method for semantic segmentation backdoor attacks. ASBA extracts artistic style features from an image, injects them as a trigger into a local region of the poisoned image, and modifies the victim class label using a label conversion function. The poisoned image is then mixed with a clean image to create a backdoor dataset for training. This method not only has a sufficiently stealthy trigger, but also produces a model that is able to segment victim class pixels into target classes while preserving the segmentation of non-victim pixels.

References

1. Kim H E, Cosa-Linan A, Santhanam N, et al. Transfer learning for medical image classification: a literature review[J]. BMC medical imaging, 2022, 22(1): 69.
2. Chai J, Zeng H, Li A, et al. Deep learning in computer vision: A critical review of emerging techniques and application scenarios[J]. Machine Learning with Applications, 2021, 6: 100134.
3. Feng D, Haase-Schütz C, Rosenbaum L, et al. Deep multi-modal object detection and semantic segmentation for autonomous driving: Datasets, methods, and challenges[J]. IEEE Transactions on Intelligent Transportation Systems, 2020, 22(3): 1341-1360.

4. Asgari Taghanaki S, Abhishek K, Cohen J P, et al. Deep semantic segmentation of natural and medical images: a review[J]. Artificial intelligence review, 2021, 54: 137-178.
5. Li Y, Li Y, Lv Y, et al. Hidden backdoor attack against semantic segmentation models[J]. arXiv preprint arXiv:2103.04038, 2021.
6. Mao J, Qian Y, Huang J, et al. Object-free backdoor attack and defense on semantic segmentation[J]. Computers & Security, 2023, 132: 103365.
7. Lan H, Gu J, Torr P, et al. Influencer backdoor attack on semantic segmentation[J]. arXiv preprint arXiv:2303.12054, 2023.
8. Li Y, Jiang Y, Li Z, et al. Backdoor learning: A survey[J]. IEEE transactions on neural networks and learning systems, 2022, 35(1): 5-22.
9. Gu T, Liu K, Dolan-Gavitt B, et al. Badnets: Evaluating backdooring attacks on deep neural networks[J]. IEEE Access, 2019, 7: 47230-47244.
10. Li Y, Zhai T, Wu B, et al. Rethinking the trigger of backdoor attack[J]. arXiv preprint arXiv:2004.04692, 2020.
11. Nguyen T A, Tran A. Input-aware dynamic backdoor attack[J]. Advances in Neural Information Processing Systems, 2020, 33: 3454-3464.
12. Cheng S, Liu Y, Ma S, et al. Deep feature space trojan attack of neural networks by controlled detoxification[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2021, 35(2): 1148-1156.
13. Johnson J, Alahi A, Fei-Fei L. Perceptual losses for real-time style transfer and super-resolution[C]//Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14. Springer International Publishing, 2016: 694-711.
14. Ulyanov D, Vedaldi A, Lempitsky V. Instance normalization: The missing ingredient for fast stylization[J]. arXiv preprint arXiv:1607.08022, 2016.
15. Lin T Y, Maire M, Belongie S, et al. Microsoft coco: Common objects in context[C]//Computer vision–ECCV 2014: 13th European conference, zurich, Switzerland, September 6-12, 2014, proceedings, part v 13. Springer International Publishing, 2014: 740-755.
16. Kurzman L, Vazquez D, Laradji I. Class-based styling: Real-time localized style transfer with semantic segmentation[C]//Proceedings of the IEEE/CVF international conference on computer vision workshops. 2019: 0-0.
17. Li G, Yun I, Kim J, et al. Dabnet: Depth-wise asymmetric bottleneck for real-time semantic
18. Cordts M, Omran M, Ramos S, et al. The cityscapes dataset for semantic urban scene understanding[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 3213-3223.
19. Everingham M, Van Gool L, Williams C K I, et al. The pascal visual object classes (voc) challenge[J]. International journal of computer vision, 2010, 88: 303-338.
20. Chen L C, Zhu Y, Papandreou G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 801-818.
21. Yang M, Yu K, Zhang C, et al. Denseaspp for semantic segmentation in street scenes[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 3684-3692.
22. Fu J, Liu J, Tian H, et al. Dual attention network for scene segmentation[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 3146-3154.
23. Zhao H, Shi J, Qi X, et al. Pyramid scene parsing network[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 2881-2890.