Research on Confrontation Sample Attack for License Plate Recognition

Ruitao Ding

School of Cyber Engineering, Xi'an University of Electronic Science and Technology, Xi'an, 020093, China

23009200572@stu.xidian.edu.cn

Abstract. The deep learning technology is becoming widely used in the sectors of safety monitoring and intelligent transportation management, and this has made license plate recognition systems crucial. However, these systems also face security challenges from antisample attacks. This study examines the literature on antisample assaults on license plate recognition software, with a focus on recent advancements and methodologies. First, this paper analyzes several major adversarial sample generation methods, especially gradient-based methods, which can significantly reduce the accuracy of license plate recognition systems without being detected by the naked eye. Then, this paper introduces the basic working principle of the license plate recognition system and its importance in practical applications and finally summarizes the research progress and related methods in recent years. Through a comprehensive analysis of the existing literature, this paper aims to provide researchers with a comprehensive overview of license plate recognition systems against sample attacks and their defense strategies for further development in the field.

Keywords: License plate recognition, antisample attacks, deep learning.

1. Introduction

In recent years, with the rapid development of the transportation system, speeding, red light, and illegal parking road insecurity behavior is also gradually increasing, and these behaviors will have a serious impact on traffic safety, road order and public interest, significantly increasing the risk of traffic accidents. In order to enhance traffic management, and build intelligent transportation systems, license plate recognition technology came into being. License plate recognition technology is a significant use of computer vision that has demonstrated considerable potential and value in a variety of disciplines, including automated driving, traffic management, and security monitoring. This technology realizes automatic recognition of license plate numbers by collecting, processing and analyzing vehicle images, which provides strong support for vehicle tracking, violation monitoring and parking management. As deep learning models are widely used in LPR systems, the efficiency and accuracy of recognition have been greatly improved. However, new threats have also emerged. Among them, adversarial sample attacks have gradually become a major hidden danger threatening the security of these systems.

Anti-sample attack is the process of making a deep learning model produce false outputs by adding small but deliberate perturbations to the input data [1]. This exploit may endanger digital security and traffic safety in addition to spoofing the license plate recognition (LPR) system and producing false results for license plate recognition [2-3]. In particular, LPR based on deep neural networks can be extremely vulnerable to attacks against adversarial samples [3]. Therefore, it is crucial to delve into the study of antisample attacks against license plate recognition systems, in order to significantly enhance both the efficiency and accuracy of vehicle recognition processes.

2. Overview of Counter Sample Attacks

The Fast Gradient Sign Method (FGSM), proposed by Goodfellow et al. in 2015, is a widely adopted technique for generating adversarial samples. The fundamental principle behind this method involves computing the gradient of the model with respect to the input and then introducing perturbations in the direction of this gradient to create adversarial samples.FGSM is efficient, and less computationally expensive in generating adversarial samples, but it has a low success rate and is easily detected in some cases [4].

Project Gradient Descent Attack (PGD) is an effective method for generating adversarial samples. It searches for perturbations that both minimize the loss function and satisfy the perturbation constraints by gradient descent.PGD attack has been widely studied and applied due to its simplicity, efficiency and powerful performance [5]. The basic idea is to gradually increase the perturbation of the input data by applying a gradient descent step over many iterations until a predefined perturbation magnitude is reached or other stopping conditions are met. The direction of the perturbation at each step is based on the gradient of the current model output, which allows the attack to be effectively optimized using local linear approximations of the model. Traditional PGD attacks require a large number of iterations to generate effective adversarial samples, which may lead to higher computational costs in practical applications.

DeepFool is a simple but effective algorithm for generating adversarial samples that mislead deep neural networks into making incorrect predictions by minimizing the Euclidean distance between the input image and the decision boundary of the classifier. The core idea of the DeepFool algorithm is to use the gradient information to determine the decision boundaries for each category and to gradually adjust the input image to make it closer to the target category. This class-by-class approximation approach gives DeepFool an advantage in terms of computational efficiency, as it does not require a complex optimization process or a large number of iterations.

3. Overview of license plate recognition technology

A license plate recognition system typically comprises three essential components: license plate localization, character segmentation, and character recognition [6]. License plate localization involves precisely pinpointing the position of the license plate within a cluttered background. Once the license plate is successfully located, character segmentation proceeds to isolate the individual character regions on the plate, paving the way for subsequent character recognition. Character recognition serves as the culminating step in the license plate recognition process, tasked with accurately determining the category of each character extracted from the segmented regions. As deep learning techniques advance, there has been a surge in research on character recognition utilizing models like Convolutional Neural Networks (CNNs) and Long Short-Term Memory Networks (LSTMs). These models are capable of learning intricate feature representations from vast datasets, thereby enhancing the accuracy and efficiency of the recognition process.

Currently, mainstream license plate recognition algorithms mainly include color saliency feature cascade classifiers based on linear support vector machines, lightweight license plate character recognition algorithms based on shape feature vectors, and LPRNet algorithms based on deep learning. Linear support vector machine based color saliency feature cascade classifier for accelerated license plate localization. Firstly, it performs license plate localization by the image downscaling method. Subsequently, candidate regions are identified using line density filtering, and ultimately, a cascading license plate classifier, which leverages color saliency features, is introduced to discern the actual license plate from among the candidate regions [7]. Lightweight shape feature vector-based license plate character recognition algorithm on the other hand is used for license plate character recognition by proposing a concept based on shape feature vectors and achieved 97.31% correctness in experiments without the need for a complex training process [8]. The deep learning-based LPRNet algorithm, a lightweight CNN, demonstrated robust performance on two datasets, achieving recognition accuracies of 90% and 89%, respectively.

At the same time license plate recognition technology has been widely used in many scenarios, including highway network toll environments, road electronic eye automatic violation recognition capture, community car access control systems and other scenarios.

4. Research on the application of adversarial samples in license plate recognition

The main types of adversarial attacks include tainting attacks based on convolutional neural networks and black box adversarial attacks.

CNN based smudge attack is to simulate the real life smudge attack on license plate by adding only local perturbations to the license plate image. The l1 paradigm is used as an optimization algorithm to get the locations in the license plate image that are susceptible to be identified incorrectly by the character classifier then continue to generate specific perturbations in that image and finally add the perturbations to the locations that are susceptible to be attacked incorrectly [9].

A Black-box adversarial attack on license plate recognition employs the Non-dominated Sorting Genetic Algorithm with Elite Policy (NSGA-II). By solely accessing the output class label and its corresponding confidence, it can create adversarial samples that exhibit greater robustness against environmental variations [3]. The black-box attack method based on One-pixel Attack is highly covert. The antagonistic samples are almost visually indistinguishable from the original image, making them difficult to be detected by traditional detection systems. This stealthiness makes the attack more threatening, especially in real-time systems that require a quick response [9].

Adversary samples that can mislead the license plate recognition system are generated by modifying the classical approach and extensively evaluated on the HyperLPR system, demonstrating that the system can be easily attacked by the adversary samples [10].

A license plate recognition method for complex scenes based on improved YOLOv7: K-means++ clustering is used instead of the original K-means clustering to form anchor frame clustering parameters suitable for license plates, and Shuffle Attention is inserted in front of the three detection heads in order to realize the effective fusion of the channel attention and the spatial attention as well as the information communication between the different sub-features, so as to improve the network's detection performance [9].

Adversary samples that can mislead the license plate recognition system are generated by modifying the classical approach and are extensively evaluated on the HyperLPR system, demonstrating that the system can be easily attacked by adversary samples [11-13].

5. Analysis of results

Using the research conducted by Wenqian Zhao et al. as an illustrative example, it is deduced that the license plate character recognition algorithm based on LeNet-5 achieves a recognition rate of 98% for original images. However, when confronted with adversarial samples generated by the FGSM and DeepFool methods, the recognition rates drop significantly to 1.03% and 0.85%, respectively. To investigate the portability of these adversarial samples, license plate character recognition algorithms utilizing AlexNet and VGG16 are developed, with their recognition accuracy rates trained to reach 98% and 99%, respectively.When the antagonistic samples generated by the FGSM method are directly inputted into the network, the recognition correct rates are 3.42% and 5.98%, respectively. When the antagonistic samples generated by the DeepFool method are fed into the network, the recognition accuracy is 1.64% and 6.56%, respectively. This indicates that license plate recognition systems based on deep neural networks are susceptible to adversarial sample attacks, allowing the deep neural network to be effectively deceived, resulting in a substantial decline in the recognition rate of the system when subjected to specific perturbations. The current online deep learning based license plate recognition networks are not trustworthy as they are vulnerable to adversarial attacks.

New deep learning techniques, such as self-supervised learning and generative adversarial networks, also have significant potential in the study of countering sample attacks. Generative Adversarial Networks are able to learn the true distribution of data and generate high-quality data samples through the adversarial process of a generator and a discriminator. Self-supervised learning, as an unsupervised learning method, has also received much attention in recent years. It avoids the dependence on a large amount of labelled data by using information from the input data itself to guide the learning process of the model. Methods combining self-supervised learning and GAN have shown their effectiveness in some fields. For example, the accuracy of the model can be effectively improved by combining self-supervised learning and GAN in the task of face attribute recognition with small samples[14] . These techniques hold promise as effective remedies to the adversarial sample issue that deep learning models encounter. Future studies can venture into exploring the deployment of these techniques in diverse domains and endeavor to integrate them with other cutting-edge deep learning approaches to further enhance the resilience of the models.

6. Conclusion

This paper focuses on a comprehensive analysis of the existing literature to conclude that deep neural network-based license plate recognition systems are particularly susceptible to the influence of adversarial samples. For example, the recognition accuracy of the system can be significantly reduced by adding imperceptible perturbations to license plate image. Furthermore, the methods for generating antagonistic samples are also advancing, from the initial simple perturbation to the current complex algorithms, such as FGSM, DeepFool, and so on. These methods can effectively mislead the license plate recognition system without being detected by the naked eye. This paper has similarly found that, despite multiple defences, deep learning models are still at risk of being effectively attacked. This is mainly due to the fact that the linear nature of the deep neural network itself makes it susceptible to adversarial samples. Therefore, future research needs to further explore more effective defense strategies to ensure the security and reliability of license plate recognition systems.

References

[1] McDaniel P, Papernot N and Celik Z B 2016 Machine Learning in Adversarial Settings IEEE Security & Privacy 14(3) pp 68–72

[2] Gu Z et al. 2020 Adversarial Attacks on License Plate Recognition Systems Comput. parent. cont. 65(2) pp 1437–1452

[3] Chen X Y, Gu J, Yan K, Jiang D 2023 Dual adversarial attacks against license plate recognition systems Journal of Network and Information Security 9(3) pp 16–27

[4] Chen S, Li Z, Du X and Nie Q 2024 EAND-LPRM: Enhanced Attention Network and Decoding for Efficient License Plate Recognition under Complex Conditions Algorithms 17 p 262

[5] Antoniou N, Georgiou E et al. 2022 Alternating Objectives Generates Stronger PGD-Based Adversarial Attacks arXiv.org

[6] Li C and Li Y B 2012 Development and research status of license plate recognition technology Science and Technology Information (5) pp 110, 125

[7] Yuan Y, Zou W, Zhao Y, Wang X, Hu X and Komodakis N 2017 A Robust and Efficient Approach to License Plate Detection IEEE Trans. Image Process. 26(3) pp 1102–1114

[8] Li X M and Feng K L 2019 A lightweight license plate character recognition algorithm Computer Science 46(z1) pp 239–241, 258

[9] Hu H, Qian Y 2020 Taint attack and defense based on convolutional neural network Journal of Zhejiang Institute of Science and Technology 32(01) pp 38–43

[10] Chen J, Shen S, Su M, Zheng H and Xiong H 2021 Black-box adversarial attack on license plate recognition system Journal of Automation 47(1) pp 121–135

[11] Zhao W 2021 Research on black box counter attack method for license plate recognition algorithm (Xi’an: Xi’an University of Electronic Science and Technology)

[12] Zhang Y et al. 2023 Research on License Plate Recognition Method in Complex Scenes Based on Improved YOLOv7 Proc. of the 34th China Process Control Conference (School of Electrical and Automation Engineering, East China Jiaotong University; Jiangxi Key Laboratory of Advanced Control and Optimization) p 1434

[13] Xu D W, Jiang B, Chen J. A feature fusion-based method for detecting antagonistic samples of electromagnetic signals Journal of Radio Wave Science

[14] Shu Y, Mao L B, Chen S et al. 2020 Combining self-supervised learning and generative adversarial networks for small-sample face attribute recognition Chinese Journal of Image Graphics 25(11) pp 2391–2403