

Comparison of Attacks on Traffic Sign Detection Models for Autonomous Vehicles [†]

Chu-Hsing Lin * and Guan-Wei Chen

Department of Computer Science, Tunghai University, Taichung 40704, Taiwan; imonanoko@gmail.com

* Correspondence: chlin@go.thu.edu.tw

[†] Presented at 8th International Conference on Knowledge Innovation and Invention 2025 (ICKII 2025), Fukuoka, Japan, 22–24 August 2025.

Abstract

In recent years, artificial intelligence technology has developed rapidly, and the automobile industry has launched autonomous driving systems. However, autonomous driving systems installed in unmanned vehicles still have room to be strengthened in terms of cybersecurity. Many potential attacks may lead to traffic accidents and expose passengers to danger. We explored two potential attacks against autonomous driving systems: stroboscopic attacks and colored light illumination attacks, and analyzed the impact of these attacks on the accuracy of traffic sign recognition based on deep learning models, such as convolutional neural networks (CNNs) and You Only Look Once (YOLO)v5. We used the German Traffic Sign Recognition Benchmark dataset to train CNN and YOLOv5 to establish a machine learning model, and then conducted various attacks on traffic signs, including the following: LED strobe, various colors of LED light illumination and other attacks. By setting up an experimental environment, we tested how LED lights with different flashing frequencies and light color changes affect the recognition accuracy of the machine learning model. From the experimental results, we found that, compared to YOLOv5, CNN has better resilience in resisting the above attacks. In addition, different attack methods will interfere with the original machine learning model to some extent, affecting the ability of self-driving cars to recognize traffic signs. This may cause the self-driving system to fail to detect the presence of traffic signs, or make incorrect decisions about identification results.

Keywords: autonomous vehicles; adversarial attacks; CNN; YOLOv5; traffic sign attacks; LED strobes



Academic Editors: Teen-Hang Meen,
Chun-Yen Chang and Cheng-Fu Yang

Published: 25 December 2025

Citation: Lin, C.-H.; Chen, G.-W. Comparison of Attacks on Traffic Sign Detection Models for Autonomous Vehicles. *Eng. Proc.* **2025**, *120*, 7. <https://doi.org/10.3390/engproc2025120007>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The development of artificial intelligence technology has been booming, and the automobile industry has rushed to launch autonomous driving systems. However, accidents have been reported in the automatic driving systems installed in many unmanned vehicles recently. Further, the automatic system is also facing all manner of cybersecurity attacks. Failure to resist various potential attacks may lead to traffic accidents and expose passengers to danger.

We explored two potential attacks against autonomous driving systems: stroboscopic attacks [1] and colored light illumination attacks [2]. We analyzed the impact of these attacks on deep learning models, such as convolutional neural networks (CNNs) [3] and You Only Look Once (YOLO)v5 [4], especially on the accuracy of the identification of traffic signs. We conducted a variety of attacks to explore potential weaknesses in the system's recognition of traffic signs.

We used the German Traffic Sign Recognition Benchmark (GTSRB) dataset to train CNN and YOLOv5 to establish models. Then, we conducted various attacks on traffic signs, including the following: LED strobe and color-light illumination. The experiments demonstrated how lights with different flashing frequencies and light colors affect the recognition accuracy of the models.

From the experiment results, we found that, compared with YOLOv5, CNN has better resilience in resisting the attacks. In addition, different attacks will interfere with the models to some extent, affecting the ability of the system to recognize traffic signs. It may cause the system to fail to detect the presence of traffic signs or make the wrong decision.

2. Methodology

2.1. Dataset

The GTSRB dataset [5,6] is used as the dataset, which contains 43 types of traffic signs in Germany. The original training set in the GTSRB dataset has 39,544 images, and 10,320 for the test set. We picked 39,209 images for the training set and 12,630 for the validation set to train the CNN and YOLOv5 models. In the distribution of the 43 types of images, the top 3 are the speed limit of 50 km/h (5.74%) and 30 km/h (5.66%) and a yield of 5.51%, respectively.

After the models were established, we chose the speed limit of 30 km/h for the attack test, not only because it has sufficient data during training, but also because it is very common on the road. In particular, speed limit signs are usually used in urban streets, school areas and other places that require high attention and safety. If the AI model incorrectly identifies these signs, it could cause the vehicle to drive at an unsafe speed, thereby increasing the risk of an accident.

2.2. Machine Learning (ML) Models

We selected two deep learning models, YOLOv5 and CNN, to evaluate the impact against attacks. YOLOv5, as a real-time object detection model, has the characteristics of high speed and high accuracy, and is very suitable for real-time detection in dynamic scenes. CNN, as a representative model for image classification, extracts and identifies features in images through its multi-layer structure. CNN is widely used in image classification tasks; it provides a benchmark to compare the impact of various attacks on different deep learning models.

Figure 1 shows the mAP@0.5 curve of the YOLOv5 model during training. The model converges rapidly in the early stages of training, and mAP@0.5 rises steadily to close to 1.0. This means that the performance of the YOLOv5 model is very good and can effectively learn and identify features in the training data.

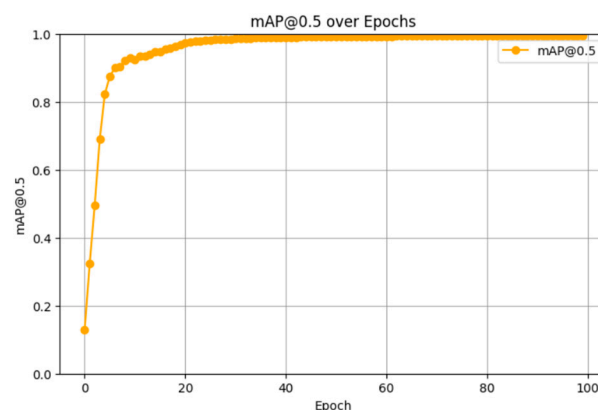


Figure 1. mAP@0.5 curve for YOLOv5.

Figure 2 shows the curves of training accuracy and validation accuracy of the CNN model during the training process. The training accuracy and validation accuracy of the model increase rapidly and reach a stable level after several rounds of training. At the same time, the value of mAP@0.5 also continues to increase to close to 1.0. This means that the CNN model has good learning and generalization capability, without the over-fitting problem [7].

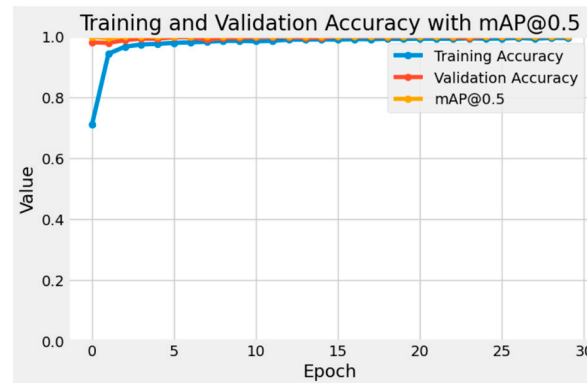


Figure 2. mAP@0.5 curve for CNN.

3. Experiments and Results

3.1. LED Strobe Attacks

This type of attack is based on the LED stroboscopic phenomenon captured by the camera. LED stroboscopic refers to the phenomenon of LED light sources flickering at extremely high frequencies. Since LED lights produce constantly changing brightness when the current is quickly switched on and off, these frequent brightness changes may not be easily detected by the human eye, but the camera's photosensitive element can capture these subtle changes in light.

This attack exploits the interaction between the photographic equipment and the light source. Simply put, when the light source flashes at a high frequency, the camera's CMOS sensor captures discontinuous light signals during the exposure process, causing irregularities in the image. The brightness changes or stripes affect the judgment of the deep learning model. In this experiment, a Philips Master LEDtube Mains T5 was used, and its flickering frequency is 60 Hz. Such high-frequency flickering is enough to interfere with the imaging process of the camera in the autonomous driving system. In Figure 3, the left is the original image, and the right is taken at a shutter speed of 1/4000. We can see that the text is in the gap of the number "3".



Figure 3. Strobic interference.

We use the camera of the Samsung Galaxy S22 Ultra to record these disturbed images. The core of this stroboscopic attack is that the rapid flashing of the light source will interfere with the camera's photosensitive element, causing small but sufficient changes in the image capture to affect the model's judgment. In this way, we were able to test the robustness of

YOLOv5 and CNN in the face of interference to evaluate their ability to cope with different lighting conditions in the real world.

We compared the recognition capabilities of YOLOv5 and CNN for speed limit (30 km/h) traffic signs under stroboscopic attack. We used two models, CNN and YOLOv5, and threw in images with different shutter speeds and different stripe positions so that the two models could predict the results. In reality, cameras in automotive AI road sign recognition systems usually use very fast shutter speeds, such as 1/1000 s or faster, to avoid image blur caused by vehicle movement. Shutter speed is closely related to vehicle speed: when the vehicle speed is high, the camera needs a faster shutter speed to capture clear road sign images to ensure the accurate identification of road signs in various driving environments.

To evaluate the performance of both models, we took 40 different photos illuminated by a Philips Master LEDtube Mains T5 and used the Samsung Galaxy S22 Ultra's camera to capture the photos. These images used four different camera shutter speeds to record these images (1/250, 1/500, 1/1000, 1/2000, 1/4000, 1/6000, 1/8000, 1/12,000), and test the positions of five different clauses, including the opening of "3", the right side of the opening, the left side of the opening, the middle of "0" and the middle of "3" and "0". Table 1 displays the statistics on the accuracy and confidence of stroboscopic attacks under various conditions. Higher shutter (HS) indicates that the shutter speed exceeds 1/2000, while normal shutter (NS) indicates that the shutter speed is lower than or equal to 1/2000.

Table 1. Accuracy and confidence on strobic attacks.

Metric	CNN	YOLO
Higher shutter accuracy	65.0%	45.0%
Normal shutter accuracy	70.0%	45.0%
Average accuracy	67.5%	45.0%
Higher shutter confidence	97.3%	40.1%
Normal shutter confidence	96.3%	37.4%
Average confidence	96.8%	38.8%

The overall performance of CNN in identifying "Speed limit (30 km/h)" traffic signs is better than that of YOLOv5. The prediction accuracy of the CNN model is 67.50%, while the accuracy of the YOLOv5 model is only 45.00%. In terms of average confidence, CNN performed better, reaching 96.80%, while YOLOv5's average confidence was only 38.78%. This means that, in facing stroboscopic attacks, CNN can more stably identify target marks and have higher confidence in prediction results.

We further analyzed the performance of the two models under different shutter speed conditions, and the results show the following:

1. Under high-speed conditions (shutter speed exceeds 1/2000), the accuracy of CNN is 65.00% and the average confidence is 97.30%; in comparison, the accuracy of YOLOv5 is 45.00% and the average confidence is 40.11%. This shows that CNN's performance in high-speed environments is more stable and reliable.
2. Under normal conditions (shutter speed lower than or equal to 1/2000), the accuracy of the two models is equivalent, both 70.00%. However, the average confidence of CNN is still higher than that of YOLOv5. This shows that, even when the accuracy is the same, CNN's confidence in its correct prediction is still significantly higher than that of YOLOv5.

The experimental results show that the recognition accuracy and robustness of the CNN model in the face of stroboscopic attacks are better than those of the YOLOv5 model [8]. Especially under high-speed conditions, CNN performs better than YOLOv5, which means that it is more resistant to interference from light frequency changes. Although

the accuracy of the two is comparable under normal conditions, the higher confidence of the CNN model further proves its robust performance under various conditions.

3.2. Light Attacks

In this experiment, we focused on LED colored light illumination attacks. Figure 4 showcases the effect of LED light sources of specific colors to interfere with traffic signs. We chose two colors of light sources for testing: white light and green light. White light was chosen because it is widely used in daily environments and can more realistically simulate the light interference effect in actual situations. Green light was used because the camera's photosensitive element is particularly sensitive to it [9] and can interfere with image capture more significantly.



Figure 4. Effect of color-light illumination.

In order to test the LED color-light attack, this experiment used a device composed of multiple components to achieve precise control of the light source. The composition of the experimental device is shown in Figure 5, and it includes the following main components:

1. 24 V DC LED light: As the main light source, we used two bulbs of different colors, white and green. These LED lights have their switching and frequency controlled by Arduino.
2. B0302 boost module: This module is responsible for stably increasing the power supply voltage to the working voltage of the LED lamp, ensuring that the LED lamp operates stably at the required brightness.
3. 1 Kohm resistor: Connected in series in the circuit to limit the current flowing through the LED light and avoid overload.
4. Arduino Uno: Switch the status of the LED light by controlling the S0913 NPN transistor. Arduino receives the preset control signal and adjusts the strobe and color coverage of the LED light.
5. S0913 NPN transistor: A switching element controlled by Arduino, used to drive high-voltage loads (LED lights). The design of the transistor ensures the separation of the control circuit and the power circuit, thereby protecting the Arduino.
6. Camera: Samsung S22 Ultra: This camera was used to capture traffic sign images under different LED light sources. These images were used to test the recognition performance of the CNN and YOLOv5 models.

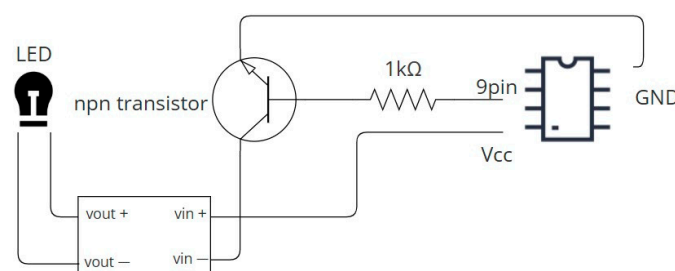


Figure 5. Circuit diagram of experiment device.

Figure 5 shows the circuit diagram of the experimental device. We used the pin 9 corner of the Arduino to connect the base of the NPN transistor, so that we could control the switch of the NPN transistor to make the LED flash. Then, the VCC of the Arduino was connected to the vin+ of the boost module, and then the vout+ was connected to the positive of the LED. The LED was connected back to the vout− of the boost module, and then vin− was connected to the collector of the transistor, and finally the emitter of the npn transistor was connected to the GND of the Arduino.

We compared the recognition capabilities of YOLOv5 and CNN for traffic signs under the attack of LED colored light illumination. In order to evaluate the performance of these two models, we used LED tubes of different colors (green and white) and four LED flash frequencies, namely 30, 60, 90 and 120 Hz, and also used different shutter speeds: 1/60, 1/90, 1/250, 1/500, 1/1000, 1/2000, 1/4000, 1/6000, 1/8000 and 1/12,000 s, a total of 40 photos.

Table 2 presents the accuracy and confidence of attacks under various conditions. Note that “G” indicates green light and “W” means white light in Table 2. Higher shutter means that the shutter speed exceeds 1/2000, while normal shutter means that the shutter speed is less than or equal to 1/2000. According to the experimental results shown in Table 2, the CNN model performs better under white light conditions than under green light conditions. The CNN’s prediction accuracy under white light is 82.5%, but only 30% under green light. In comparison, the YOLOv5 model has an accuracy of 65% under white light conditions and only 7.5% under green light conditions. This means that both models can recognize traffic signs more stably in white light environments, and that green light sources will have a significant impact on both models.

Table 2. Accuracy and confidence on different colors of light.

Metric	CNN (G)	CNN (W)	YOLO (G)	YOLO (W)
Higher shutter accuracy	15.0%	65.0%	15.0%	60.0%
Normal shutter accuracy	45.0%	100.0%	00.0%	70.0%
Average accuracy	30.0%	82.5%	7.5%	65.0%
Higher shutter confidence	64.9%	92.1%	38.7%	57.4%
Normal shutter confidence	79.6%	99.9%	00.0%	45.2%
Average confidence	75.9%	96.8%	38.7%	50.9%

G: green light; W: white light.

In terms of average confidence, the CNN model achieved an average confidence of 96.84% under white light conditions, which was higher than 75.9% under green light conditions. The average confidence of the YOLOv5 model under white light conditions is 50.85%, which is higher than 38.67% under green light conditions. This shows that the white light environment has a small impact on the two models, while the green light condition significantly reduces the recognition confidence of the model.

We further analyzed the performance of the two models under different shutter speed conditions, and the results show the following:

1. Under higher shutter speed conditions (more than 1/2000), the accuracy of the CNN model is 65% under white light and only 15% under green light. In comparison, the YOLOv5 model has an accuracy of 60% under white light and 15% under green light. This means that, although the green light recognition capabilities of both models are significantly affected under high shutter speeds, the performance of the CNN model is more stable under white light conditions.
2. Under normal shutter speed conditions (less than or equal to 1/2000), the accuracy of the CNN model reaches 100% under white light and 45% under green light. The accuracy of the YOLOv5 model is 70% under white light and 0% under green light. This

shows that, in a normal shutter speed environment, white light conditions are more conducive to model recognition, and the CNN model performs better than YOLOv5. We also analyzed the impact of different LED light flicker frequencies (Table 3).

1. Under the LED light flashing frequency of 30 Hz and 60 Hz, the accuracy of the CNN model under white light conditions is 80% and 80%, respectively, which is much higher than the 40% and 20% under green light conditions. And the accuracy of the YOLOv5 model under white light is 60% and 40%, respectively, and the accuracy under green light is 0%.
2. In the case of 90 Hz and 120 Hz, the accuracy of the CNN model under white light conditions remains at 80%, while it is 0% and 60% under green light conditions,, respectively. The accuracy of the YOLOv5 model under white light is 80% and 80%, while, under green light conditions, it is 0% and 30% respectively.

Table 3. Accuracy on different frequencies of colored lights.

Frequency	CNN (G)	CNN (W)	YOLO (G)	YOLO (W)
30 Hz	40.0%	80.0%	00.0%	60.0%
60 Hz	20.0%	80.0%	00.0%	40.0%
90 Hz	00.0%	80.0%	00.0%	80.0%
120 Hz	60.0%	90.0%	30.0%	80.0%

Under different light source colors and flicker frequencies, the recognition accuracy and stability of the CNN model are significantly better than those of the YOLOv5 model [8]. Especially under white light conditions, CNN can maintain high recognition accuracy and confidence under all frequencies and shutter speed conditions. These results indicated that the CNN model is more suitable for dealing with LED color coverage attacks and has stronger anti-interference ability in dealing with different light source conditions.

4. Conclusions

We compared the recognition capabilities of YOLOv5 and CNN under stroboscopic attacks and LED colored light illumination attacks. The experimental results showed that the overall performance of CNN is better than that of YOLOv5 in the face of the two attacks. Especially in high-speed shutter conditions and white light environments, the recognition accuracy of the CNN model is higher, and the average confidence is also significantly higher than that of YOLOv5.

Under the stroboscopic attack, CNN achieved an average accuracy of 65% and a confidence of 97.28%, while YOLOv5 had an accuracy of only 50% and a confidence of 41.5%. Especially under higher shutter conditions (such as 1/6000 and above), the accuracy of CNN reaches 66.7%, which is significantly better than YOLOv5's 33.3%. This means that CNN can more stably identify traffic signs under stroboscopic attacks.

In experiments targeting LED colored light attacks, the accuracy of CNN reached 82.5% under white light conditions, while it was only 30% under green light conditions; the accuracy of YOLOv5 was 65% under white light conditions, and only 30% under green light conditions. This means that the model's performance in a white light environment is relatively stable, while green light significantly reduces the model's recognition ability and confidence. Especially under high shutter speed conditions (over 1/2000), green light has a greater impact on the model, and the accuracy of both decreases significantly.

Based on the experimental results, it can be concluded that the CNN model shows stronger anti-interference ability and stability when dealing with stroboscopic and LED colored light attacks, regardless of high and low shutter speeds or different light source

colors. Therefore, CNN's recognition accuracy and confidence under various attack conditions are better than those of YOLOv5, especially in the face of light frequency changes and light source color interference; it performs more reliably. These results show that CNN is more suitable for traffic sign recognition in practical application scenarios, especially in the presence of light interference.

Author Contributions: Conceptualization, C.-H.L.; methodology, C.-H.L.; software, G.-W.C.; validation, C.-H.L.; formal analysis, C.-H.L.; investigation, C.-H.L.; resources, G.-W.C.; data curation, G.-W.C.; writing—original draft preparation, G.-W.C.; writing—review and editing, C.-H.L.; visualization, G.-W.C.; supervision, C.-H.L.; project administration, C.-H.L.; funding acquisition, C.-H.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Science and Technology Council, Taiwan, grant number NSTC 114-2221-E-029-003.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

Acknowledgments: This study is supported by the grant from the National Science and Technology Council, Taiwan, project number: NSTC 114-2221-E-029-003.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Shen, Y.; Lin, C.-H.; Chen, G.-W. MLIA: Modulated LED illumination-based adversarial attack on traffic sign recognition system for autonomous vehicle. In Proceedings of the IEEE International Conference on Trust, Security and Privacy in Computing and Communications, New York, NY, USA, 9–11 December 2022.
2. Woitschek, F.; Schneider, G. Physical adversarial attacks on deep neural networks for traffic sign recognition: A feasibility study. In Proceedings of the IEEE Intelligent Vehicles Symposium (IV), Nagoya, Japan, 11–17 July 2021.
3. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [[CrossRef](#)]
4. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
5. Xie, K.; Li, Y.; Zhang, M. Efficient federated learning with spike neural networks for traffic sign recognition. *IEEE Trans. Veh. Technol.* **2022**, *71*, 9980–9992. [[CrossRef](#)]
6. Pavlitska, S.; Ivanov, A.; Chen, G.-W. Adversarial attacks on traffic sign recognition: A survey. In Proceedings of the 3rd International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME), Canary Islands, Spain, 19–21 July 2023.
7. Xu, Y. On splitting training and validation set: A comparative study of cross-validation, bootstrap and systematic sampling for estimating the generalization performance of supervised learning. *J. Anal. Test.* **2018**, *2*, 249–262. [[CrossRef](#)] [[PubMed](#)]
8. Wang, J.; Liu, S.; Zhang, F. Research on improved YOLOv5 for low-light environment object detection. *Electronics* **2023**, *12*, 3089. [[CrossRef](#)]
9. Anzagira, L.; Fossum, E.R. Color filter array patterns for small-pixel image sensors with substantial cross talk. *J. Opt. Soc. Am. A* **2015**, *32*, 28–34. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.