International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-11 Issue-6, March 2023

Intrusion Attacks on Deep Learning Frameworks Employed in Self-Driving Vehicles

Syeda Kausar Fatima, Syeda Gauhar Fatima

Abstract: ***Deep convolutional networks have proven practical for autonomous vehicle applications as deep CNN technology has advanced. There has been a growing vogue for using end-to-end computational methods for the mechanization of vehicular activities. Preliminary studies, though, have demonstrated that deep learning network classifiers are sensitive to adversarial approaches. However, the impact of adversarial strategies on regression problems remains poorly understood. We propose two white-box direct security breaches targeting progressive self-driving vehicles in this research. A prediction model is used in the navigation mechanism, which receives a picture as input and returns a steering angle. By altering the input image, we may influence the actions of the automated driving unit. Two different attacks can be launched in practice on CPUs without the need for GPUs. The effectiveness of the threats is demonstrated by trials carried out in Udacity.***

Keywords: ***Adversarial Intrusions, Driverless Vehicles, Nvidia's Driving Architecture, Regression Model.***

I. INTRODUCTION

Driverless vehicles are among the most challenging issues in computerised applications, demanding security and safeguarding facilities. The majority of practical self-driving cars use configurable structures that break down the driving activity into simpler modules. Configurable structures comprise localization, estimation, scheduling, and control modules furthermore a perception component that uses deep neural networks to detect as well as categorize entities within the environment. Experts are further investigating the possibility of end-to-end automated driving. End-to-end steering systems are unified modules that immediately translate raw input data into output, often employing deep learning models. The NVIDIA driving framework [1], for instance, converts unfiltered pixels of the front-focused camera to steering instructions. Improvements in high-throughput GPUs have contributed to the advancement of end-to-end automated driving. Deep learning models are susceptible to adversarial attacks, as demonstrated in various scenarios.

Manuscript received on 18 February 2023 | Revised Manuscript received on 01 March 2023 | Manuscript Accepted on 15 March 2023 | Manuscript published on 30 March 2023.

*Correspondence Author(s)
Dr. Syeda Kausar Fatima*, Department of Electronics and Communication, Deccan College of Engineering and Technology, Hyderabad (Telangana), India. Email: kausarfatima@deccancollege.ac.in, ORCID ID: https://orcid.org/0009-0006-9700-8966
Dr. Syeda Gauhar Fatima, Department of Electronics and Communication, Deccan College of Engineering and Technology, Hyderabad (Telangana), India.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC-BY-NC-ND license http://creativecommons.org/licenses/by-nc-nd/4.0/

These often deceive image recognition models by introducing an undetectable disturbance to the source picture [2]. While the number of research papers describing progressive deep neural networks is growing, their protection in realistic conditions remains unknown. While end-to-end approaches may result in improved efficiency and compact solutions, the unitary component is susceptible to adversarial threats.

The following are the primary findings of this study:

Researchers propose two live white-box combative strategies over an end-to-end prediction framework for automated driving: a powerful approach that creates a perturbation for every image (image-specific), and another stealthy attack that creates a global adversarial variation affecting every image (image-agnostic). Simulations in Udacity are used to demonstrate the threats' resilience. The trials suggest that the danger may only require a few seconds to divert the car out of the zone.

II. PROPOSED WORK

A. Preliminary Driving Architectures

End-to-end vehicle solutions see the operating process as a single module that instantly transfers sensor information to steering controls [3]. End-to-end mobility solutions are often developed utilizing emulation or reinforcement training. DNNs are employed in artificial learning approaches to train and emulate human driving abilities. A supervisor is in charge of providing classified data to the algorithm. Reinforcement training approaches, in contrast, use exploration and utilization to enhance driving strategies. The presence of a manager is often not required for the training phase. Despite reinforcement training becoming increasingly prominent, emulation learning continues to be more common in end-to-end operating systems [4]. Hence, the study will target emulation learning approaches.

The autonomous land vehicle framework was the first demonstration of an emulation-learning-dependent end-to-end steering system, training a three-layered fully connected network to manoeuvre an automobile on open streets [5]. End-to-end steering systems, on the other hand, were employed for off-road drivers [6]. NVIDIA scientists have developed a CNN model that directly converts raw images from a discrete front-facing camera into driving signals.

B. Intrusion Attacks

In this study, researchers will explore a driving simulation that employs a regression technique to generate successive steering instructions. Early work on adversarial approaches, on the other hand, has primarily focused on targeting categorization methods.

An effective classification model intrusion deviates the outcome from the appropriate classification label. Using the electronic handwritten character recognition categorization job, as an illustration, an adversary can trick the classifiers into mistaking the numeral 3 for the numeral 7. Researchers must measure the size of the resultant variation to assess the effectiveness of an adversarial attempt in a regression analysis. An attack that merely leads the steering angle to vary from 1.00 to 0.99 is usually regarded as unsuccessful because such a minor variation would possess no discernible influence upon a specific driving result. An attack should result in higher variations to be considered as effective. Researchers have utilized the Mean Absolute Deviation to examine as well as contrast such variations (MAD). A substantial intrusion must result in a greater MAD versus random noise while causing the equivalent overall quantity of disruption [7] [8].

Although previous studies have primarily focused on offline intrusions on classification algorithms, we investigate virtual attacks on regression analysis. Offline intrusions modify passive graphics. An offline approach in the context of autonomous driving divides the driving history into static images, as well as the corresponding steering angles. The perturbations are therefore repeated to every stable picture; also, the total efficiency is used to assess the attack [9]. In contrast, online intrusions use perturbations in complicated situations.

Researchers implement the perturbations while the car is in motion, as opposed to deploying them on static photos in a driving log. It is also now possible to examine how the driving simulations responded to the intrusions.

A significant distinction between virtual and physical intrusions is the absence of actual reality in online intrusions. Offline intrusions in self-driving use pre-recorded actual drivers' steering directions as the actual reality, but real-time virtual intrusions do not possess access to already recorded human predictions. As a result, we consider the simulated data under typical small cases as the underlying data, supposing that the steering model is reasonably close to the actual truth. This assertion is valid given that there is no purpose in exploiting the program if the modelling is incorrect. In such a case, the incorrect modelling will represent a threat in itself.

Black-box, White-box, and Grey-box threats are the three types of adversarial examples that are now in use [10]: With white-box intrusions, the targeted model's design as well as attributes are completely known to the attackers; In Gray-box intrusions, the targeted modeling is only partially known to the attackers; The sole way adversaries in black-box threats may learn from the architecture is by querying. Researchers have developed two white-box intrusions that operate in real-time against steering algorithms.

III. PROBLEM FORMULATION

This chapter defines the goal, introduces computational terminology, and describes the desired model. Researchers utilise these expressions throughout the article.

$$y = f(\theta, x) \quad (1)$$

$$y' = f(\theta, x') \quad (2)$$

in which y represents the benign throughput steering instruction, f(θ, x) represents the regression analysis that maps picture data to steering instructions, θ the prototype metrics, x represents the actual input picture, y' represents the adversarial throughput steering signal, as well as x' represents the adversarial source picture. Furthermore, researchers used $\eta = x - x'$ to represent the arbitrary perturbations, y* to represent the objective steering instruction, and J(y, y*) to represent the training losses. The goal of manipulating a classification model based on a provided picture x would be to induce a minor disturbance η so that y'≠ y*. Nevertheless, the goal of targeting a regression analysis is to provide a minor disturbance η that causes the discrepancy between y and y*, which is greater than the mean deviation δ induced by adding random noise to x. The L2 standard will be used to characterize the size of the disruption. Researchers demand, in particular, that the interruption be imperceptible to regular sight.

$$\|x' - x\|_2 = \|\eta\|_2 \leq \varepsilon \quad (3)$$

Where $\varepsilon = 0.03$.

NVIDIA's driving architecture is the goal concept. The model's reference structure is (160, 320, and 3), which indicates (height, breadth, and channel) in that order. On each of the (unaltered) gathered photos, the resultant steering angle is located in the region of [-1, 1]. A -1 result indicates driving to the left, while a 1 result indicates moving to the right. The front camera captures the raw picture, which is then processed using established ways before being fed into the prototype. For further information on various data pre-processing procedures, involving resizing, scaling, as well as RGB to YUV conversion, see [1]. Figure 1 describes the driving model's architecture.

Layer (type)	Output Shape	Param #
lambda (Lambda)	(None, 160, 320, 3)	0
conv2d (Conv2D)	(None, 78, 158, 24)	1824
conv2d_1 (Conv2D)	(None, 37, 77, 36)	21636
conv2d_2 (Conv2D)	(None, 17, 37, 48)	43248
conv2d_3 (Conv2D)	(None, 15, 35, 64)	27712
conv2d_4 (Conv2D)	(None, 13, 33, 64)	36928
dropout (Dropout)	(None, 13, 33, 64)	0
flatten (Flatten)	(None, 27456)	0
dense (Dense)	(None, 100)	2745700
dense_1 (Dense)	(None, 50)	5050
dense_2 (Dense)	(None, 10)	510
dense_3 (Dense)	(None, 1)	11

Figure 1: Conceptual Framework of Autonomous Driving.

IV. ADVERSARIAL INTRUSIONS

In this chapter, researchers present two white-box threats against self-driving vehicles: one image-specific and the other image-agnostic. Finally, the system's architecture will be discussed.

© Copyright: All rights reserved.

A. Image-specific strategy

An image-specific downstream approach that created a single disturbance for each input picture represented the initial adversarial attack for a classification model [4]. Rather than reducing the learning loss J(y, y*), researchers optimized the learning loss and thereafter generated the perturbations using the slope of the training loss. Nevertheless, because virtual attackers have no access to the underlying data y*, the learning loss J(y, y*) cannot be determined. Hence, a new adversary loss J(y) is needed, which needs just the model results y to create the perturbations.

While addressing a regression model, keep in mind that there is an option of increasing or decreasing the results. To challenge the end-to-end steering regression analysis, for instance, users may divert the car to the left by lowering the output and to the right by raising the output. As a result, intrusions on regression analysis may be considered as a subset of attacks against the classification algorithm, further with the added limitation that there are merely two options: raising or reducing the outcome. As a result, the focus is on the simple adversarial loss equations for the image-based approach.

$$J_{left}(y) = -y \quad (4)$$

$$J_{right}(y) = y \quad (5)$$

The FGSM can then be used by producing perturbations like

$$\eta = \varepsilon\, sign[\nabla_X (J(y))] \quad (6)$$

Here ε It is a scale parameter that controls the visibility of the perturbations. Algorithm 1 summarizes the image-specific approach.

Algorithm 1 Image-specific Attack

Input: The regression model $f(\theta, x)$, the input images $\{x_t\}$ where x_t is the image at time step t.
Parameters: The strength of the attack ϵ.
Output: Image-specific perturbation η.
for each time step t **do**
  Inference: $y = f(\theta, x)$
  Perturbation: $\eta = \epsilon$ sign$[\nabla_x(J(y))]$
end for

Figure 2: Algorithm for Image-Specific Intrusions.

Algorithm 2 Image-agnostic Attack (Training)

Input: The regression model $f(\theta, x)$, input images in a driving record X, the target direction $I \in \{-1, 1\}$.
Parameters: the number of iterations n, the learning rate α, the step size ξ , and the strength of the attack ϵ measured by the l_∞ norm.
Output: Image-agnostic perturbation η.
Initialization: $\eta \leftarrow 0$
for each iteration **do**
  for each input image x in the driving record X **do**
    Inference: $y = f(\theta, x + \eta)$
    if sign$(y) \neq I$ **then**
      $x' = x + \eta$
      $\eta_t \leftarrow 0$
      while sign$(y) \neq I$ **do**
        Gradients: $\nabla = \frac{\partial J(y)}{\partial x'}$
        Perturbation: $\eta_t = \eta_t + proj_2(\nabla, \xi)$
        Inference: $y = f(\theta, x + \eta_t)$
      end while
      $\eta = proj_\infty(\eta + \frac{\alpha}{\xi}\eta_t, \epsilon)$
    end if
  end for
end for

Figure 3: Algorithm for Image-Agnostic Intrusions.

Suppose the intruder intends to attack the car from the right direction. The goal in this scenario is to boost the simulation results. The disturbance may then be generated using the adversarial loss Jright(y). The slope of the error function across the input is represented by ∇_x (J(y)). The slope indicates how variations in the adversarial output y will be propagated back to the source input.

B. Image-Agnostic Strategy

Perhaps minor variations might result in a road accident. A minor change in steering angle, for instance, might lead to the inability to manoeuvre around a steep bend. In other terms, even though the attack wasn't as powerful as such image-specific intrusion, it might be dangerous if used at vital times. As a result, we provide a white-box approach and create a universally adversarial perturbation (UAP) [11] that can be utilized to damage all incoming pictures at varying time intervals. Deep Fool [12] as well as Projected Gradient Descent (PGD) [13] are combined mostly in image-agnostic techniques. The intrusion is divided into two stages: learning and implementation. Researchers produce a UAP digitally or from a driving log, and then implement it.

Researchers select the matching adversarial error component (Jleft(y) or Jright(y)) after determining the intended path, specifically, whether to target the car to the left (y<0) or through the right (y>0). The perturbations are set to 0 at the start. Suppose indeed the route of the simulation model is not exclusive to the intended way for every source picture at every iteration step. In that case, we discover the most minor disturbance that transforms the sign of the model outcome in the intended way.

Researchers estimate the slope of the adversarial error J(y) and afterwards apply it to the L2 sphere to change the path of the simulation model, including the least amount of disruption. The optimization problem's closed-form approach arg min $||\eta - \eta'||2$ with the constraint $||\eta'|| \leq \varepsilon$ is provided by

$$proj_2(\eta, \xi) = \frac{\eta}{\max\left\{1, \frac{||\eta||}{\xi}\right\}} = = \eta \min\{1, \frac{\xi}{||\eta||}\} \quad (7)$$

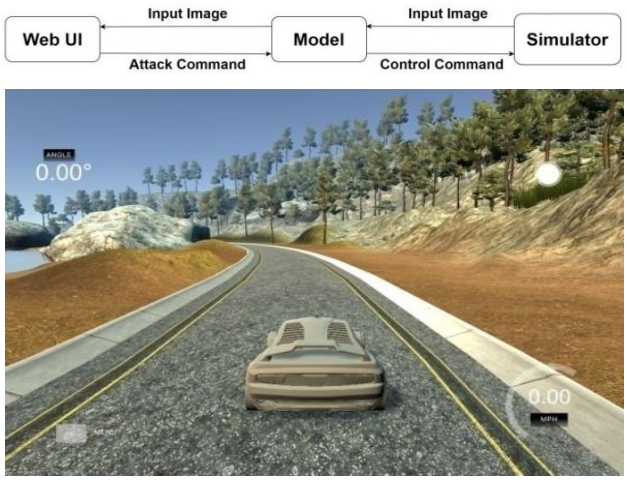
This may be demonstrated by employing the Lagrange as well as KKT assumptions [14].

If the orientation of the model results corresponds to the preferred orientation upon implementing the immediate disturbance $\eta_t$at iteration step t, researchers integrate the temporary disturbance η_t to the whole disturbance and afterwards, the project η on the l_2 focused on zero as well as the perimeter ϵ to make sure that the limitation $\| \eta' \|2 \leq \epsilon$ is acceptable. Algorithm 2 summarizes the approach. As shown, the adversary employs a while loop identical to that used in DeepFool or the representation method presented in the PGD approach.

C. Design of the Proposed System

Researchers create an adversarial strategy to target the whole self-driving model (See Figure 2). The system consists of three major components: the simulator, the web server, and the Web UI. The simulator uploads the pictures obtained from the front camera to the web server. Simultaneously, it receives steering orders from the web server to control the car.

© Copyright: All rights reserved.



Udacity Simulator

Figure 4: Design framework of the adversarial driving mechanism.

The web server gets pictures from the simulator through WebSocket interfaces and afterwards returns control signals. Subsequently, the software accepts web Ui's attack orders and then introduces the adversarial perturbations into the source picture. The entire drive model is also installed on the web server. For the front-end, designers utilise a webpage where the user can observe the simulator's progress and select alternative approaches.

V. EXPERIMENTAL RESULTS

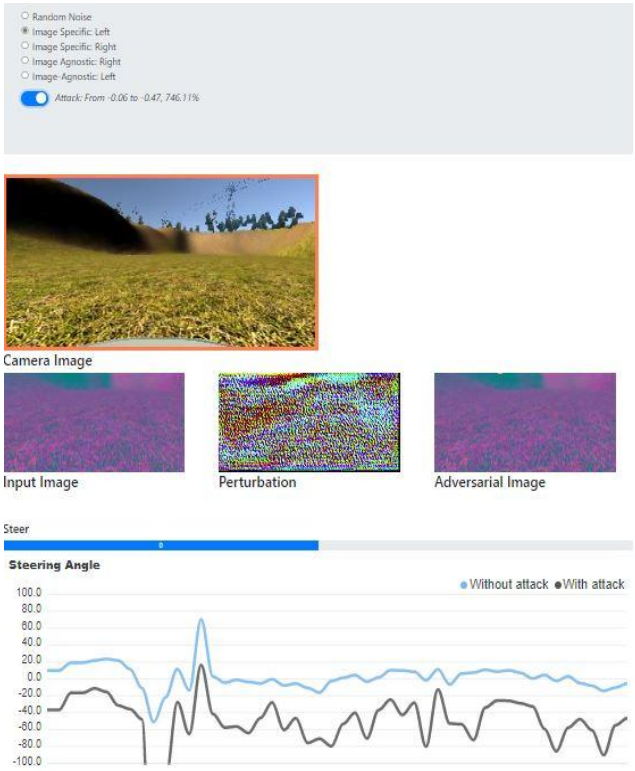
The functionality of the suggested image-specific as well as image-agnostic intrusions is described as follows.

A. Model Training

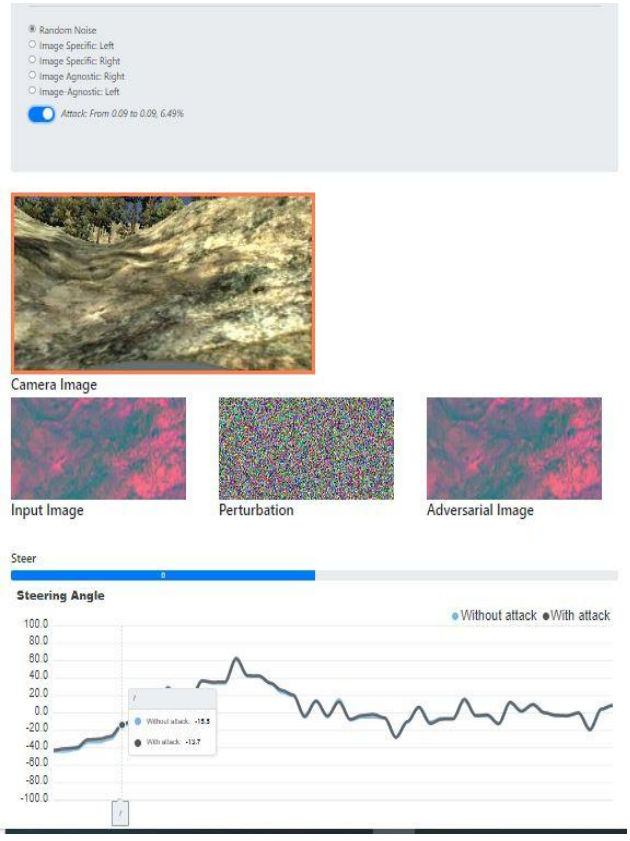
The goal is to develop a successful real-time web intrusion detection system that employs an end-to-end imitation learning approach. Because conducting online intrusions targeting real-world autonomous vehicle mechanisms poses a risk, researchers evaluated these intrusions in an autonomous driving simulator. Manual driving data was used to train the targeted artificial learning approaches. In the Udacity testing environment, they acquired 8k photos of manual driving logs. Tests revealed that the model is susceptible to adversarial intrusions.

B. The Image-Specific Attack

To proceed, they show that adding noisy data to the driving model produces relatively minor differences. The variable ϵ is utilized to guarantee that the overall perturbation by noisy data is more like the overall perturbation from the image-specific intrusion. Take note that the image-specific intrusion now boosts or deducts from every unit according to the gradient's sign. Similarly, random noise disturbances were created that contribute or deduct ϵ out of each pixel randomly. In figure 5, three distinct attacks of equal strength are used. The car deviated from the road for many seconds after being subjected to the image-specific intrusion. The image-specific to the left intrusion causes the vehicle to divert to the left by lowering the steering angle; therefore y_{adv} is less compared to y_{true} in Figure 5a. The image-specific right strategy, contrasted with the left attack, diverts the car towards the right by raising the steering angle, therefore y_{adv} is bigger compared to y_{true} in Figure 5c. The random noise perturbations have a minimal influence on the steering system, as they only slightly alter it. y_{adv} from y_{true}.



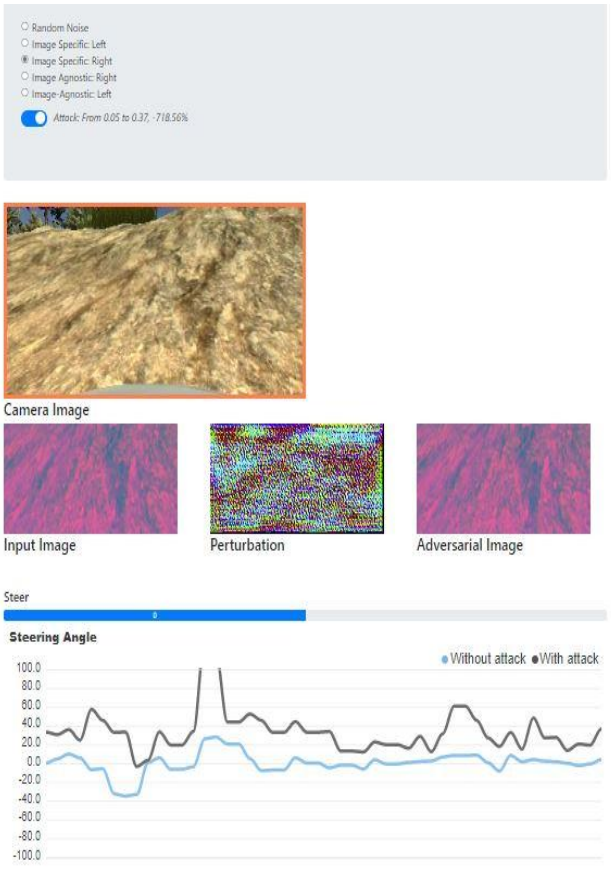
(a) Reduction in steering angle for image-specific left intrusion.



(b) Negligible deviation for random noise attack

Retrieval Number: 100.1/ijrte.F74820311623
DOI: 10.35940/ijrte.F7482.0311623
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)
© Copyright: All rights reserved.



(c) Greater steering angle for image-specific right.

Figure 5: Image-specific and random noise results for the strength of ($\epsilon = 1$).

Designers also calculated the mean absolute deviation of the steering angle for 800 strikes. Table I presents the data. As demonstrated, indeed, the smallest image-specific approach (ϵ =0.1) outperforms the most extraordinary random noise intrusion (ϵ = 8). The steering angle can also be diverted outside of the region [-1, 1] whenever ϵ= 4 or otherwise ϵ = 8. In contrary terms, the image-specific intrusion is quite powerful. Its limitation is that it would compute the slopes of every input picture. In practice, access to the input picture and gradients may not be available. As a result, designers presented the image agnostic approach, which educates the perturbations using drive logs without requiring exposure to the feed and gradients throughout implementation.

Table I: Mean deviation of steering angle for Image-Specific

Attack Strength	Random Noise Attack	Image-specific Attack
$\epsilon = 0.1$	0.0002	0.1448
$\epsilon = 1$	0.0020	0.4779
$\epsilon = 2$	0.0048	0.7329
$\epsilon = 4$	0.0150	1.4895
$\epsilon = 8$	0.0278	2.4469

C. The Image-Agnostic Attack

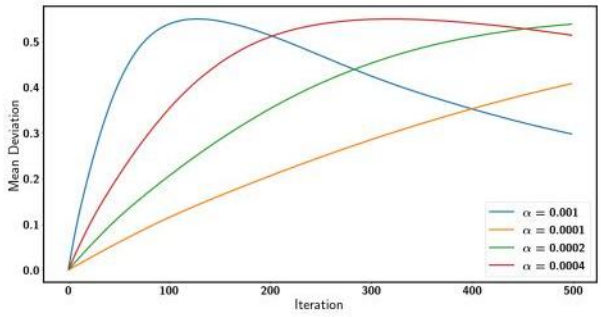
Similarly to the image-specific intrusion, the intensity of the image-agnostic intrusion is evaluated against a random noise threat. Table II shows the findings. The image-agnostic assault is less intense compared to image-specific intrusion, but it is nonetheless more potent than those in the random noise threat.

Table II: Mean deviation of steering angle for Image-Agnostic

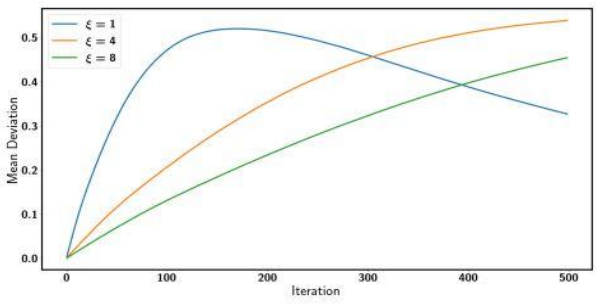
Attack Strength	Random Noise Attack	Image-agnostic Attack
$\epsilon = 0.1$	0.0002	0.0373
$\epsilon = 1$	0.0020	0.1109
$\epsilon = 2$	0.0048	0.1294
$\epsilon = 4$	0.0150	0.1131
$\epsilon = 8$	0.0278	0.1275

The image-agnostic strategy's effectiveness would not increase beyond $\epsilon > 2$ as seen in Table 2. This is owing to the perturbation's low generalization. Raising the power of the intrusion may improve the simulation results for specific inputs while decreasing the prototype for others. As a result, continuing to increase ϵ further contributes additional variance to the simulation output while maintaining a steady mean absolute deviation.

Researchers further examined how the training rate alpha and step duration xi affected the learning procedure (see Figure 6). The training rate, alpha, governs the evolution of perturbations throughout the repetition. Designers experimented with several. α values, using constant values epsilon = 1 and xi = 4. The mean deviation grows quickly as α rises. Nevertheless, as the iterative model progresses, the mean deviation reduces beyond a hundred iterations while α is $>$ 0.01. For every input picture x, the step duration ξ determines how quickly the disturbance is adjusted to move the simulation results in the intended direction. A lower ξ renders the update in the intended direction most constant, but it takes longer to iterate. A bigger ξ may shift the model results path in a single attempt; however, the perturbations might not be generalizable to additional inputs.



(a) Variable α with constant $\epsilon = 1$, $\xi = 4$.

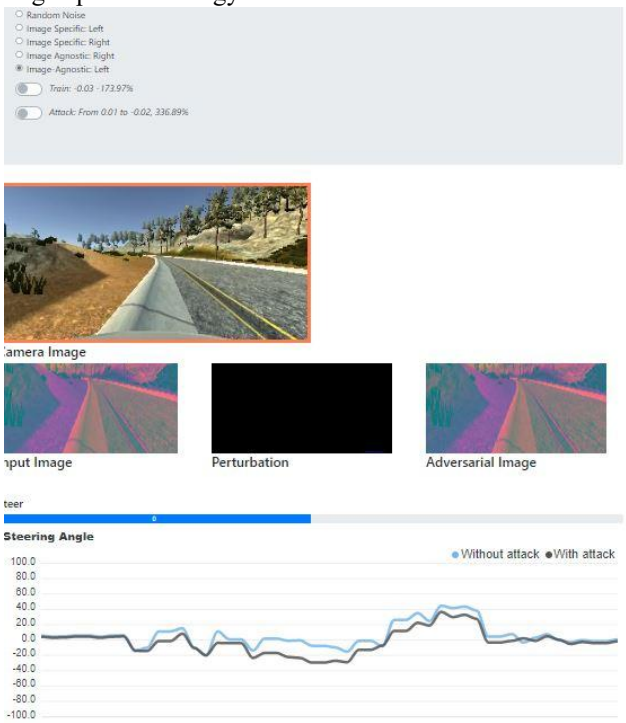


(b) Variable ξ with constant $\alpha = 0.002$

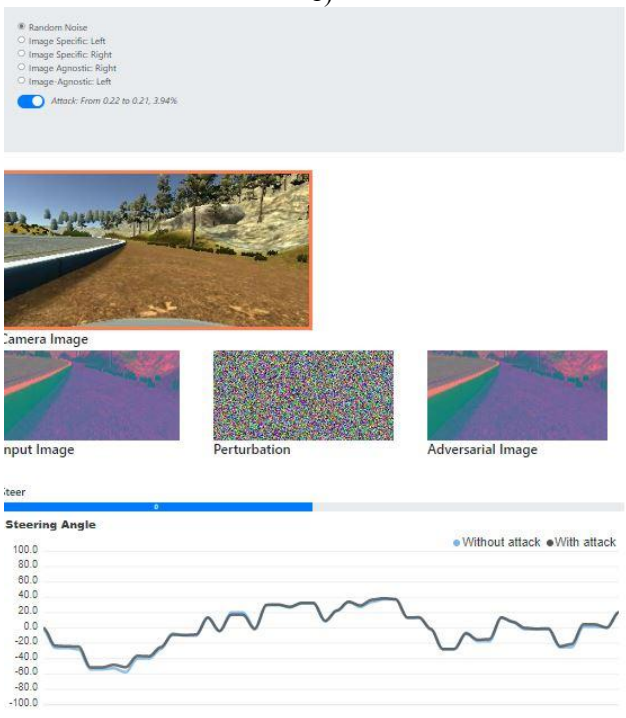
Figure 6: The mean variation of steering angle while training using various hyperparameters.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)
© Copyright: All rights reserved.

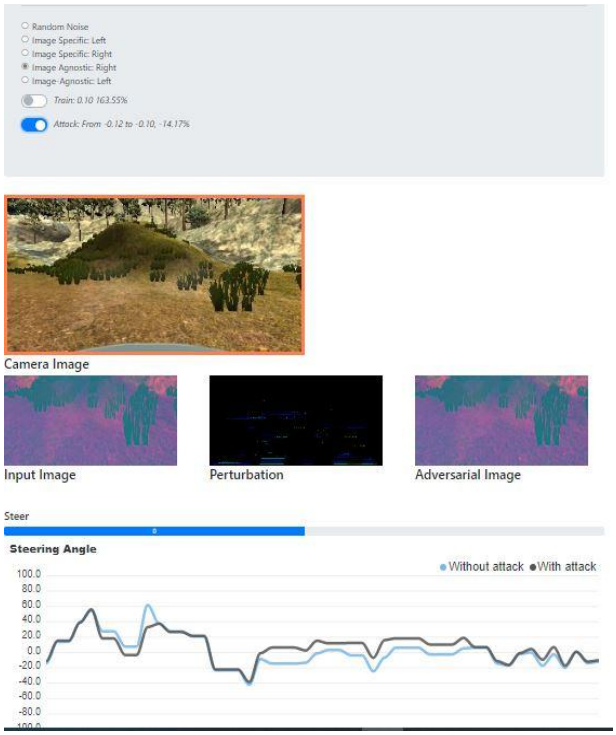
As shown in Figure 7, employing the variables $\alpha = 0.0002$ as well as $\xi = 4$ allowed the construction of image-agnostic disturbances at $\epsilon = 1$, which responds similarly to the image-agnostic intrusion at $\epsilon = 0.1$. Whereas the image-agnostic intrusion is less potent than the image-specific approach, it renders the car harder to steer at steep bends, which might result in mishaps at specific dangerous points. Furthermore, the image-agnostic assault perturbs all pixels with a single disturbance. As a result, deploying the image-agnostic approach is significantly more scalable than deploying the image-specific strategy.



(a) The image-agnostic left intrusion reduces the model result ($y_{adv} < 0$), trying to make turning right harder. ($\epsilon = 1$)



(b) The random noises only slightly vary y_{adv} from y_{true}.



(c) The image-agnostic right intrusion increases the model result ($y_{adv} > 0$), trying to make turning left harder. ($\epsilon = 1$)

Figure 7: Image-agnostic ($\alpha = 0.002$, $\xi = 4$, $n = 500$) and random noise results for strength of ($\epsilon = 1$).

VI. CONCLUSION

The work revealed that an autonomous vehicle driving model may be vulnerable to real-time attacks. Researchers propose a powerful image-specific approach as well as a covert image-agnostic approach. Whilst the image-agnostic approach has a lower mean deviation than the image-specific approach, both methods are much more successful over random noise intrusions. The image-agnostic approach causes the car to divert toward the outside of the road in a couple of seconds, whereas the image-agnostic intrusion may cause mishaps at steep turnings. These findings contribute to the growing body of data indicating the susceptibility of safety-critical robotic systems.

DECLARATION

Funding/ Grants/ Financial Support	No, I did not receive.
Conflicts of Interest/ Competing Interests	No conflicts of interest to the best of our knowledge.
Ethical Approval and Consent to Participate	No, the article does not require ethical approval or consent to participate, as it presents evidence that is already publicly available.
Availability of Data and Material/ Data Access Statement	Not relevant.
Authors Contributions	All authors have equal participation in this article.

© Copyright: All rights reserved.

REFERENCES

1. M. Bojarski, D. Del Testa, D. Dworakowski, B. Firner, B. Flepp, P. Goyal, L. D. Jackel, M. Monfort, U. Muller, J. Zhang, et al., "End to end learning for self-driving cars," arXiv preprint arXiv:1604.07316, 2016.
2. I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," arXiv preprint arXiv: 1412.6572, 2014.
3. E. Yurtsever, J. Lambert, A. Carballo, and K. Takeda, "A Survey of Autonomous Driving: Common Practices and Emerging Technologies," IEEE Access, vol. 8, pp. 58 443–58 469, 2020. [CrossRef]
4. Tampuu, T. Matiisen, M. Semikin, D. Fishman, and N. Muhammad, "A survey of end-to-end driving: Architectures and training methods," IEEE Transactions on Neural Networks and Learning Systems, 2020.
5. D. A. Pomerleau, "Alvinn: An autonomous land vehicle in a neural network," in Advances in Neural Information Processing Systems, D. Touretzky, Ed., vol. 1. Morgan-Kaufmann, 1989.
6. U. Muller, J. Ben, E. Cosatto, B. Flepp, and Y. Cun, "Off-road obstacle avoidance through end-to-end learning," in Advances in Neural Information Processing Systems, Y. Weiss, B. Scholkopf, and J. Platt, Eds., vol. 18. MIT Press, 2006.
7. S. Villar, D. W. Hogg, N. Huang, Z. Martin, S. Wang, and G. Scanlon, "Adversarial attacks against linear and deep-learning regressions in astronomy," in Proceedings of Machine Learning Research 2020 1st Annual Conference on Mathematical and Scientific Machine Learning. "Mathematical and Scientific Machine Learning Conference", 2019.
8. A. T. Nguyen and E. Raff, "Adversarial attacks, regression, and numerical stability regularization," arXiv preprint arXiv: 1812.02885, 2018.
9. Y. Deng, X. Zheng, T. Zhang, C. Chen, G. Lou, and M. Kim, "An analysis of adversarial attacks and defences on autonomous driving models," in 2020 IEEE International Conference on Pervasive Computing and Communications (PerCom). IEEE, 2020, pp. 1–10. [CrossRef]
10. K. Ren, T. Zheng, Z. Qin, and X. Liu, "Adversarial attacks and defences in deep learning," Engineering, vol. 6, no. 3, pp. 346–360, 2020. [CrossRef]
11. S.-M. Moosavi-Dezfooli, A. Fawzi, O. Fawzi, and P. Frossard, "Universal adversarial perturbations," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 1765–1773. [CrossRef]
12. S.-M. Moosavi-Dezfooli, A. Fawzi, and P. Frossard, "Deepfool: a simple and accurate method to fool deep neural networks," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 2574–2582. [CrossRef]
13. A. Madry, A. Makelov, L. Schmidt, D. Tsipras, and A. Vladu, "Towards deep learning models resistant to adversarial attacks," arXiv preprint arXiv: 1706.06083, 2017.
14. S. Boyd and L. Vandenberghe, Convex Optimization. Cambridge University Press, March 2004. Science Robotics, vol.. 7, no. 66, p. eabm6074, 2022. [CrossRef]

AUTHORS PROFILE

Dr. Syeda Kausar Fatima, Associate Professor, Department of Electronics and Communications, Deccan College of Engineering and Technology. She has completed her Ph.D. and has over 17 years of teaching experience, with a publication record of more than 40 papers in international journals and two patents related to her research work. Current research interests include cloud-native application security, Automation in AI/ML, and IoT. Orcid id: https://orcid.org/0009-0006-9700-8966

Dr. Syeda Gauhar Fatima, Principal of Deccan College of Engineering and Technology, obtained a B.E. in Electronics and Communication Engineering from Gulbarga University in 1997, an M.Tech in Digital Systems and Computer Electronics from JNTU, Anantapur in 2006 and a Ph.D. in Wireless Communications from JNTU, Hyderabad in 2018. She has 23 years of teaching experience. She has organized National conferences and workshops. She has presented numerous research papers at national and international conferences and Journals. She has published two patents. She is a Member of the International Association of Engineers (IAENG). Her areas of interest include Wireless Sensor Networks, Internet of Things, Artificial Intelligence and machine learning, and Digital Electronics.

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 the Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)/ journal and/or the editor(s). The Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP) 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.

© Copyright: All rights reserved.