ISP Pipeline Analysis

Sri Maruti Keerthana Ponnuru

1. Introduction

1.1. Introduction and context of the study

This research endeavors to examine Image Signal Processing (ISP) pipelines within the framework of Automotive Cameras, with a focus on their impact on object detection performance. In automobile systems, cameras are essential for gathering visual data for functions like lane tracking, object identification, and collision avoidance. The majority of vision devices are equipped with image signal processing (ISP) pipeline to perform RAW-to-RGB conversions and are embedded into data preprocessing for enhanced image processing[1]. The RAW image is converted into a visually enhanced and highly informative representation through a series of processing stages known as the ISP pipeline. These procedures involve, among other things, adjusting gain, white balance, contrast, and histogram equalization. The particular algorithms used at each pipeline stage have a substantial impact on the final image's quality and interpretability[2]. Understanding adversarial attacks is essential to enhancing the resilience of deep neural network predictions, since these models can be tricked into making a misleading prediction on an image that was accurately predicted without the perturbation by adding subtle disturbances to the input images.

1.2. Research questions and objectives of the project

In the context of this work, some crucial research questions to be addressed include understanding

Which processing pipelines and algorithms are most commonly employed in commercial automobile cameras?

How do the different algorithms impact each stage of the ISP pipeline in terms of image quality and feature enhancement?

What are potential attacks to be simulated? How will the ISP pipelines respond to these attacks?

What are the suitable metrics to evaluate the performance of the ISP pipelines over different datasets?

Existing adversarial attacks aim to manipulate the post-capture images while neglecting potential adversarial patterns introduced by ISP pipeline. The objective of this project is to examine the three ISP pipelines utilized in automotive cameras and study the performance of corresponding object detection following these pipelines by simulating various attack scenarios on the RAW data.

1.3. Objectives and milestones of the task

The objective in the context of the ISP pipeline analysis is to identify suitable attacks, implement the attacks and evaluate the impact of these attacks on the underlying object detection for each of the pipelines.

The milestones associated for my task are:

1. Literature review: Performing literature analysis to understand the ISP pipelines, recognizing potential attacks to be performed on the RAW images, establishing performance metrics for evaluation.
2. Attack estimation: Based on the literature study, finalizing the attack and working on its implementation, its effect on the image quality after object detection.
3. Evaluation criteria: Briefly identifying a set of metrics to evaluate the impact of the attack on object detection.
4. Performance Evaluation: Select and run a suitable object detection algorithm on images generated by each of the three pipelines under various attack scenarios.
5. Final review: Prepare a comprehensive technical report that documents the evaluation process, results, and conclusions.

2. Background and related work

Pertinent research has been done in the area of Image Signal Processing with primary focus on Adversarial Attacks with ISP. One influential work on producing strong adversarial examples is by first proposing the Fast Gradient Sign Method (FGSM) to generate the perturbation δ in a single step [3]. This work explores the Fast Gradient Sign Method, a ∞-bounded adversary can be attacked using FGSM, which computes an adversarial example as $x + \varepsilon \, \mathrm{sgn}(\nabla x L(\theta, x, y))$ [4]. To generate adversarial examples for spoofing, this approach perturbs the data using the sign of the image gradients. This strategy can be understood as a straightforward one-step plan to maximize the inside portion of the formulation of saddle points. FGSM blends the goal of misclassification with a white box method. It causes a neural network model to predict things incorrectly[5].

Another major work on the gradient based attack technique Madry proposed a strong iterative version with a random start based on FGSM, named Projected Gradient Descent (PGD)[6]. This attack can be understood as a straightforward one-step plan to maximize the saddle point formulation's inner portion. The multi-step variant, which is effectively projected gradient descent (PGD) on the negative loss function, is a more formidable opponent. The notion is reinforced by Carlini [7], who shows that PGDadversarial training increases the distortion needed to create adversarial cases by a factor of 4.2. Furthermore, PGD-adversarial training is the only empirical defense among all the white-box defenses that surfaced in ICLR-2018 and CVPR-2018 that hasn't been further challenged (Athalye, Carlini, and Wagner 2018; Athalye and Carlini 2018)[8]. The evaluation metrics commonly used are success rate.The ability of an attack to alter a camera pipeline's prediction to the target class is measured by the pipeline's success rate.

The other attack widely discussed is the image scaling attack. Xiao et al. [9] proposed the image-scaling attack against the scaling operation, which aims to craft an attack

image that becomes a completely different one after scaling.

This attack deliberately modifies a picture so that it appears differently at a given size. Specifically, the attack produces a picture A by moderating the source image S just enough to match a target image T in terms of scale. Other areas of attacks focused on the simple hardware attacks such as an attack on the camera for example a blinding attack. In order to capture better pictures in any lighting situation, cameras are equipped with built-in mechanisms that determine how much light is allowed to pass through their shutter. This kind of attack misuses this feature by beaming a bright light source into the camera to either totally or partially blind it, making it overlook objects. By validating the attack on vehicle cameras, we can carry out blinding attacks in different scenarios, by observing and recording the camera output, one such research done was to cause irreversible damage to a camera by shining a laser light straight at it for many seconds from a distance of less than 0.5 meters, and fully blind the camera for up to 3 seconds [10].

The mentioned research works among others and the knowledge derived from them have influenced the implementation of this project work. This includes the attacks performed, their implementations, selection of the object detection algorithm, the set of evaluation metrics chosen, the reason attributed to this is because they are backed by research and would aid in our analysis of the various ISP Pipelines- traditional, thermal and deep learning based.

3. Methodology

A thorough Literature review is done, to assess the selection of attacks and performance metrics. To perform the attack task, based on the literature survey carried out, mainly five attacks are implemented beginning with perturbing the image by adding random noise, next is the image scaling attack, third attack is the Fast Gradient Sign Method attack, followed by the Projected Gradient Descent and Blinding attack. All the code is written in Python primarily using the PyTorch and the PILlibraryThe details of the implementation for each of the attacks are given below:

1. Adding random noise: This function takes an image and generates a perturbation by adding random noise. The magnitude of the noise is controlled by the epsilon parameter. The resulting perturbed image is then clipped to ensure pixel values remain in the valid range (0 to 255).
2. Image Signaling attack: The code returns a manipulated image using a scaling attack pattern by iterating through each pixel. The attack introduces a visual distortion by changing the color of pixels based on their position.
3. FGSM attack: FGSM targeted attack by perturbing the original image in the direction opposite to the gradient. It aims to maximize the loss, causing misclassification. The epsilon parameter controls the strength of the perturbation. The loss values are computed for two different case scenarios to understand the attack further.
4. PGD attack: A neural network is defined and hyperparameters tuning is done -attack steps, attack learning rate, epsilon. The method iteratively performs the PGD attack for a specified number of steps.

   For each iteration, the gradient of the loss is calculated, and the sign of the gradient is taken into account. The adversarial example is updated by adding the step size or subtracting it based on whether it is a targeted or untargeted attack.

5. Blinding Attack: This is a hardware attack but to simulate it on the input images using the PIL library. In this attack the aim was to mimic the effects of an extreme increase in brightness and contrast, along with the addition of a bright circular shape to simulate a blinding effect.

   For the Blinding attack different case scenarios are experimented such as the brightness, the size or radius of the blind spot.

During the evaluation of adversarial attacks to validate their impact and test the robustness the YOLOv5 object detection model is applied on the images, the implemented attacks were applied to a set of images from the dataset.

The goal is to observe how the YOLOv5 model performs under different adversarial scenarios and compare the different attacks and the confidence scores of the objects detected and the number of objects detected for a given image under the five attack scenarios.

4. Experimental Results

The objective of my task included conducting literature survey to understand the existing ISP pipelines followed by the estimating and implementing of attacks on raw images, selection of evaluation criteria, performance evaluation to summarize results in the context of the ISP pipelines.

To carry out this the setup required was collecting relevant research papers, a RAW dataset each to evaluate the Traditional pipeline and the DeepLearning pipeline and Python frameworks to implement the attack and evaluation code. For the performance metrics we have chosen IOU which stands for Intersection over Union. A metric called Intersection over Union (IoU) indicates how closely the ground truth box and the predicted bounding box match. It assists in differentiating between correct detection and incorrect detection; it is one of the primary metrics for assessing the accuracy of object detection algorithms hence suitable for our work. For the implementation the function takes the coordinates of the first bounding box and the second bounding box as coordinates, it first calculates the intersection area of the two bounding boxes. Then computes the areas of each bounding box individually. The IoU score is calculated by dividing the intersection area by the union area of the two bounding boxes.

The other metric is the success rate of the ISP pipeline. The success rate of a camera pipeline indicates if an attack can alter the pipeline's forecast, in similar lines another metric is the attack success rate (ASR) this refers to whether the attack is successful in making the object detection on the ISP image deviate from the ground truth or misclassify. The metrics like accuracy are not the most suitable for our case because we are performing the attacks for the outputs of the Traditional Pipeline and the DeepLearning pipeline with different datasets for each of the pipelines, hence accuracy would be relative.

The code implementation was successful, the intention was to increase the strength of the attacks going from Attack1 which adds random noise towards Attack 5 which is the simulation of the Blinding attack. For experimental purposes to check the performance of the attacks, I have evaluated the set of attacked images of one of the datasets on the object detection Yolov5 model before sending it through the ISP pipeline as shown in Figure 1. The

The FGSM Attack has two variations 3.1, 3.2 while computing the loss the target is set as a truck to perform a targeted attack and the attack was successful. Attack 5 is simulated such that a spot is drawn on to the image indicating a blind spot and the rest of the image brightness and contrast can be manipulated while implementation portraying the intensity of the blind spot. The variations done include changing the brightness of the image resulting from the spot, changing the dimension which is the radius of the spot while keeping the relative position of the spot as constant. In Attack 5.1 the brightness, contrast, radius of spot and position are set. In Attack 5.2 the brightness is set constant and the size of the attack is altered whereas in Attack 5.3 the size is set while brightness is changed.

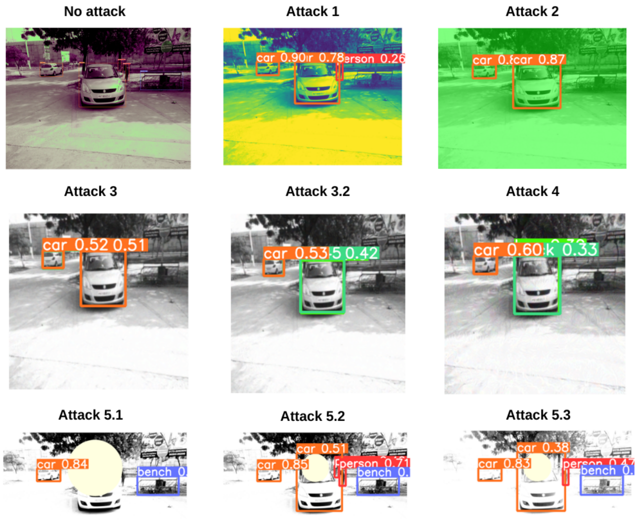

Figure1: The YOLO model detection outputs of the attacked images (before ISP Pipeline)

Attack 3.2 and Attack 4 resulted in misclassification of the cars as truck and bus respectively. While other attacks have changes in the number of objects detected and the confidence scores. Also since I have implemented the IOU code to experiment and check how the attacked images compare to the original, I calculated the IOU scores for the attacked images in Figure 1, the IOU value seemed to decrease from Attack1 towards Attack5. The IOU score for Attack 1 was [0.99, 0.98, 0.99,0.98, 0.95] this indicates that the attack has lesser impact and for Attack 5.1 the IOU scores are [0.79, 0.68].

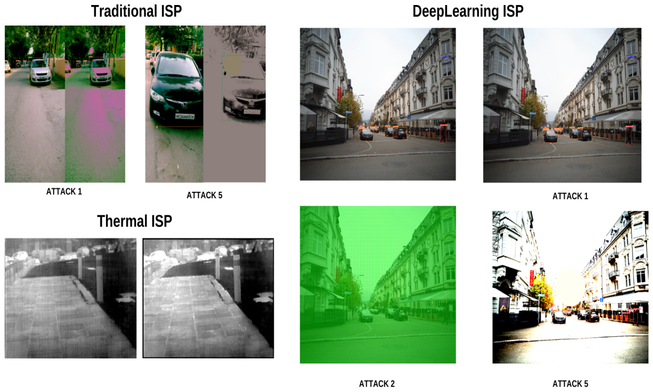

Figure2: The results of model detection outputs of the attacked images (after ISP Pipeline)

In the context of studying the classification and object detection by simulating the spoofing and blinding attacks for Traditional ISP as shown in the Figure 2 gives the Yolov5 outputs for Attack 1 and Attack 5 for a given image in the dataset. It can be observed that the Attack 1 has not much effect on the output since all the objects in the original image are detected even after the attack but with lesser confidence. Whereas in the image given as input after performing Attack 5.3 on it, in the resulting Yolov5 output all the objects are not detected and also the one object detected is misclassified as boat. The IoU for each bounding box is [1.0] for the blinding attack. The bounding boxes overlapped but the object was completely misclassified. Also, the original image had 4 cars but the perturbed image had only 1 car detected. For the Gaussian noise attack, the IoU for each bounding box is [0.97613513, 0.980276, 1.0, 0.12931961]. The metric ASR attack success rate is taken as 1 indicating the attack is successful in misclassifying the result of the model.The Figure 2 shows the Attacks 1,2 and 5 applied to an image in the dataset but it is observed that like Traditional ISP there is not much misclassification although the confidence levels have dropped. The ASR is 0 for most cases in the dataset.

Attack 1	[0.99919736, 0.999532, 0.999216, 0.99713236, 0.9929143, 0.9933053, 0.9930316, 1.0]
Attack 2	[0.9938084, 0.9855927, 0.83310384, 0.9982098, 0.5418314, 0.07891367, 0.8503031, 0, 0]
Attack 5	[0.9902526, 0.93401295, 0.831711, 0, 0.99330354, 0.9667275, 1.0, 0.93081665, 0]

Table1: The IOU values for Attack 1,2,5 for the DeepLearning ISP Pipeline

As observed the IOU values are higher for all attacks and the attack was not as impactful on DeepLearning ISP in comparison to the traditional ISP which caused misclassification and was unable to detect objects in some cases. For the Thermal ISP since no attacks were introduced in this pipeline the evaluation is not done based on the attack success rate or the IoU. Hence, the Thermal ISP performance is evaluated based on the individual results of the blocks implemented.

5. Analysis and discussion

For this project three ISP pipelines have been mainly studied: Traditional ISP -Fast openISP and Infinite-ISP, Thermal ISP and Deep Learning based ISP.

The attacks have been implemented only on the Infinite ISP and the Deep Learning based ISP. The datasets on which the ISP analysis has been done is different for both, so a comparative analysis. The Deep Learning ISP uses a Convolutional Neural Network, whereas the Traditional ISP pertains to the traditional techniques and software employed to adjust and improve digital photos taken by imaging sensors, as those found in cameras and other optical equipment. The attacks implemented provide the scope that adversarial attacks play a critical role in understanding object detection model predictions. ISPs transform RAW measurements to RGB images and traditionally are assumed to preserve adversarial patterns. This was observed in the Analysis of the ISP as well as the ASR was 1 in multiple cases for both the ISP pipelines being evaluated.

5.1. Takeaways

It is observed that the ISP pipelines handling the original raw image undergoes processing, the resultant image is accurately classified with a notably high confidence score. However, upon implementing an attack, the processed image either faces misclassification or experiences a drop in confidence scores. This suggests the case scenario of successful attack, showcasing the model's susceptibility to adversarial techniques. This vulnerability poses a significant security concern in physical systems, as adversaries could exploit it by introducing subtle perturbations to raw images. Consequently, the processed images, pivotal for decision-making in these systems, lose reliability.

A significant security concern was brought to attention through our implementation of various adversarial attacks on raw images, successfully deceiving the object detection system into misclassifying the images. This underscores the need for better image processing to ensure safety against potential security threats.

5.2. Limitations

The main limitation of the project include the following:

Dataset: The project was implemented on a limited number of images, these images do not suffice to all case scenarios that might be available in broader datasets.Instead of using dataset available using a physical camera and collecting datasets can be beneficial since we can tamper the hardware and cause a natural blinding attack instead of the simulation in our case.

Metrics and Evaluation: Since both the ISP pipelines are evaluated on different datasets the scope to form a comparative analysis is not of great importance since the results are relative to the dataset used. Also due to the scope of the Traditional ISP being slower not many results were provided to make an evaluation.

Scope of attacks: Further attack strategies can be implemented such as Carlini and Wagner attack which is an Adversarial Spoofing attack, and the scope of testing these attacks for the Thermal ISP can also pose future scope using even other object detection algorithms.

REFERENCES

[1] Park, Hyun Sang. "Architectural analysis of a baseline isp pipeline." Theory and Applications of Smart Cameras (2016): 21-45.

[2] Buckler, Mark, Suren Jayasuriya, and Adrian Sampson. "Reconfiguring the imaging pipeline for computer vision." Proceedings of the IEEE International Conference on Computer Vision. 2017.

[3] Zhang, Yuxuan, Bo Dong, and Felix Heide. "All you need is raw: Defending against adversarial attacks with camera image pipelines." European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022

[4] Xu, Jin, Zhendong Cai, and Wei Shen. "Using FGSM targeted attack to improve the transferability of adversarial example." 2019 IEEE 2nd International Conference on Electronics and Communication Engineering (ICECE). IEEE, 2019.

[5] Huang, Tianjin, et al. "Bridging the performance gap between fgsm and pgd adversarial training." arXiv preprint arXiv:2011.05157 (2020).

[6] Zheng, Tianhang, Changyou Chen, and Kui Ren. "Distributionally adversarial attack." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 33. No. 01. 2019.

[7] Carlini, Nicholas, and David Wagner. "Towards evaluating the robustness of neural networks." 2017 ieee symposium on security and privacy (sp). Ieee, 2017.

[8] Athalye, Anish, and Nicholas Carlini. "On the robustness of the cvpr 2018 white-box adversarial example defenses." arXiv preprint arXiv:1804.03286 (2018).

[9] Li, Junjian, and Honglong Chen. "Adversarial RAW: Image-Scaling Attack Against Imaging Pipeline." arXiv preprint arXiv:2206.01733 (2022).

[10] Jakobsen, Søren Bønning, Kenneth Sylvest Knudsen, and Birger Andersen. "Analysis of sensor attacks against autonomous vehicles." 8th International Conference on Internet of Things, Big Data and Security. SCITEPRESS Digital Library, 2023