

# Study of AI Object Detection: Patterns on Animals with YOLO and Adversarial Patches

Aniya Hopson  
Department of Computer Science  
Hampton University  
Hampton, VA, USA  
0009-0009-2574-0500

Dr. Chutima Boonthum-Denecke  
Department of Computer Science  
Hampton University  
Hampton, VA, USA  
Chutima.Boonthum@gmail.com  
0000-0003-0247-7518

Dr. Idongesit Mkpog-Ruffin  
Department of Computer &  
Information Sciences  
Florida A&M University  
Tallahassee, FL, USA  
0000-0002-0998-3598

**Abstract**—Artificial Intelligence (AI) has become an increasingly powerful tool in various domains, particularly in image classification and object detection. As AI advances, novel methods to deceive machine learning models, such as adversarial patches, have emerged. These subtle modifications to images can lead to misclassification of objects, posing a substantial challenge to their reliability. In this paper, we present our research findings and literature on adversarial examples and object detection.

This research builds upon the previous work by investigating the impact of small patches on object detection using YOLOv8. We started by exploring patterns within images and their influence on model accuracy. Then a follow-up study evaluating how adversarial patches, particularly those targeting animal patterns, affect YOLOv8's ability to accurately detect objects. Additionally, we explore how untrained patterns impact the model's performance, aiming to identify vulnerabilities and enhance the robustness of object detection systems.

**Keywords**—Artificial Intelligence Cyber Security, Object Detection, YOLOv8, Adversarial Patches, Machine Learning

## I. INTRODUCTION

Artificial Intelligence (AI) has been rapidly advancing across many areas of society. Over the years, AI has been adopted for use in various fields, leading to major advancements in technology and the efficiency of autonomous systems. AI has also been integrated into cameras, giving it the ability to process real-time images. AI models are built using Deep Neural Networks (DNNs), which are algorithms that use large datasets to create classifications and make human-like decisions [6]. There are several types of DNNs, each serving different purposes in AI algorithms, such as convolutional neural networks, recurrent neural networks, and deep generative networks.

Convolutional Neural Networks (CNNs) are widely used in computer science for image classification and recognition because they effectively learn complex features and identify objects [3]. Through this neural network, computers can analyze real-world videos and images to "learn." By evaluating various features, CNNs classify images into categories. This process,

known as object detection, aims to detect and label all entities in an image. When an image is processed through an algorithm like You Only Look Once (YOLO), objects are labeled with confidence values, indicating how certain the algorithm is in its classification. However, for AI to effectively categorize images, it must be trained with large datasets. Neural networks and deep learning techniques allow AI to learn from vast amounts of data to identify objects accurately [4].

YOLO is a single-shot object detection algorithm that processes an entire image in one pass, predicting both the locations and categories of objects instantly [1]. This efficiency makes YOLO a popular choice for applications like video surveillance. Figure 1 illustrates the core mechanics of the YOLOv8 object detection model, which consists of a backbone, neck, and head. The backbone, usually a pre-trained CNN, extracts different levels of features from the image. These features are then combined by the neck using methods like the Feature Pyramid Network (FPN) before being passed to the head. The head's task is to identify objects and draw bounding boxes around them using models like YOLO or Single-Shot Detector (SSD) [1].

## II. ADVERSARIAL PATCHES

Adversarial patches are patterns intentionally designed to manipulate object classification in deep learning models, causing them to mislabel or fail to recognize images accurately. Despite extensive training, even state-of-the-art models like YOLOv8 are vulnerable to these small, strategically placed alterations. Some researchers have demonstrated that minor changes to an image could confuse deep learning models, showcasing the susceptibility of AI systems to adversarial examples [2].

These patches can be applied in various forms and sizes, typically noticeable to the human eye, but effective enough to deceive AI models. These patches have been separated into different attacks: physical and digital [8]. In real-world applications, adversarial patches have already had significant impacts. For instance, adversarial glasses designed to distort facial recognition systems have successfully fooled these models, allowing attackers to impersonate others by altering the way the system perceives them [5].

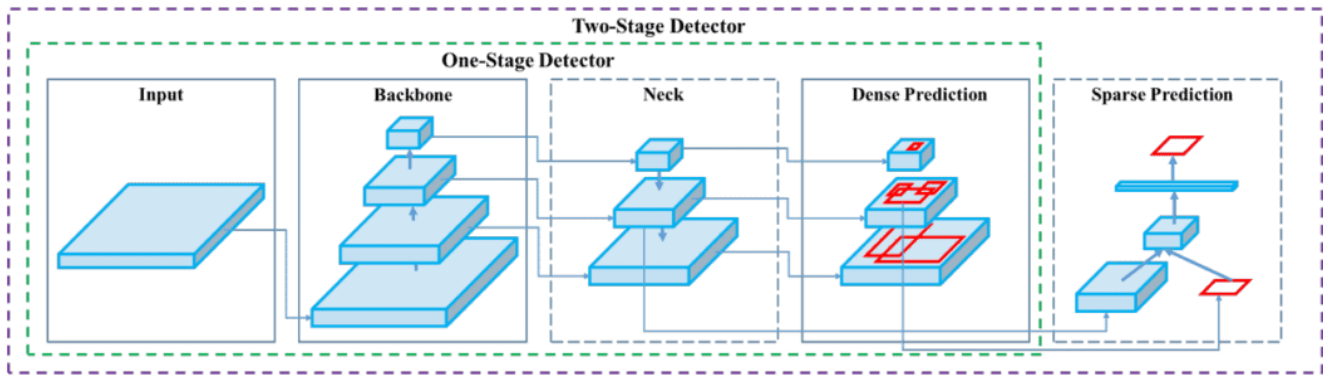


Fig. 1. Demonstrates the essential mechanics of an object detection model. [1]

Through this advancement in adversarial patches there has been research to reduce human error- related accidents. However, research suggests that autonomous vehicles (AVs) pose a significant risk to public safety [11]. Through this analysis it can reflect on the unexpected obstacles which can include animals crossing the roadways. Grosse and Alahi (2024) systemically assessed AI security risk across components, which found object detection to be the highest risk area [12].

Such vulnerabilities pose risks to the future of AI-powered technologies, including autonomous vehicles, video surveillance, and medical imaging. If left unaddressed, these attacks could undermine the reliability of AI systems. Understanding and mitigating the effects of adversarial patches is crucial for strengthening the security of object detection algorithms like YOLOv8, ensuring their robustness in real-world applications.

These most dangerous patches are the ones that look completely normal to the human eye. According to Bai et al. (2022) patches that cover less than 2% of an image can accomplish a 99% success rate and remain undetected. This is why it is important to find concerning categories of

adversarial patches in real world settings where they would be unconscious.

### III. EXPERIMENT

We tested YOLOv8, which had a pre-trained dataset from GitHub, to see how different patches affect object detection. To set up the model, we cloned the YOLOv8 repository from GitHub and downloaded the necessary Python libraries to configure our computers to run the algorithm. The pre-trained model was then used to test our datasets.

#### A. Datasets

For this study, we gathered images from the YOLOv8 open database and Pixabay, focusing on animals like cows, dogs, giraffes, elephants, horses, cheetahs, and birds. We tested 100 images by running them through our code to generate confidence values and classification results. From previous research, we knew that certain patterns could affect YOLOv8’s labeling. Animals with patterns, such as cheetahs and some dogs, had the lowest confidence levels in our experiment. As seen in Figures 2 and 3, these animals were sometimes completely mislabeled, even though some of the predictions still had high confidence.

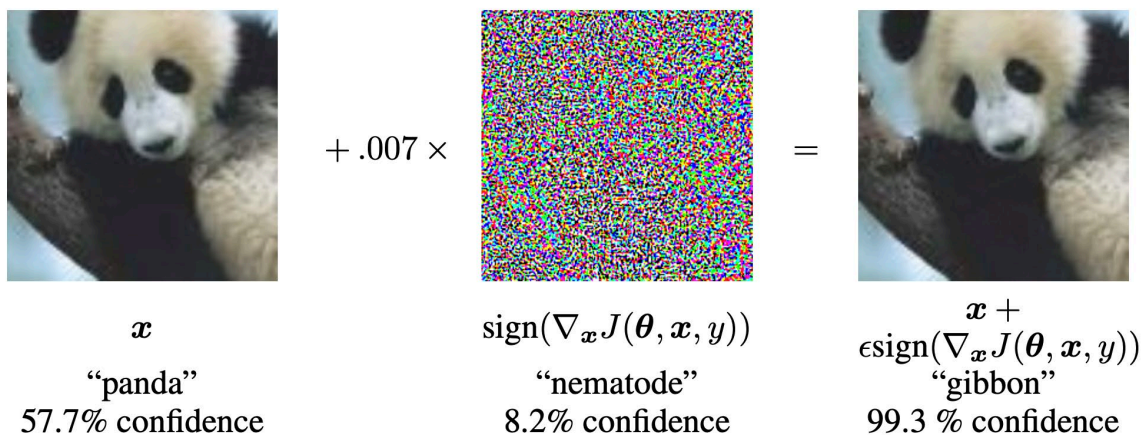


Fig. 2. By adding an imperceptibly small vector, the image classification was changed from “panda” to “gibbon.” [2]



Fig. 3. An example of digital dodging on an image of actor Owen Wilson correctly classified on the left, but not on the right due to an eyeglass frame that fooled the system. [5]

Then, we examine how patterns affect object detection in YOLOv8. We had already tested animal patterns to see how they impact YOLOv8's accuracy. After analyzing 100 different animal images, we found that certain patterns caused mislabeling or low-confidence predictions. As shown in Figures 4 and 5, images of leopards were often mislabeled. This issue occurred with every leopard image tested, and some dog images also had inaccurate classifications. By studying how animal patterns affect YOLOv8's detection, this research could help prevent technology malfunctions like mislabeling in real-time AI detection in the future.

**B. Methodology**

For this research, we used images from previous work to analyze how animal patterns affected object detection using YOLOv8 and the updated version YOLOv11. The goal was to see if certain patterns caused the model to mislabel objects or fail to recognize them correctly.

First, we selected control images of common animals like a horse, cow, dog, and elephant. These images were used as a baseline to compare with altered versions. Then, we focused on patterns from animals that had previously caused mislabeling in YOLOv8, as well as those that led to completely incorrect classifications.

To test the impact of these patterns, we used Photoshop to overlay them onto the control images. This allowed us to see if the AI would still correctly identify the animals or if the new patterns would cause errors. We applied the selected animal patterns in three different ways: covering half of the animal, covering the entire animal, and applying the pattern only to the body while leaving the head and legs unchanged. This setup helped us test whether the model's misclassification was influenced by partial versus full pattern coverage. It also allowed us to determine if the model was associating specific patterns with certain animals instead of focusing on the actual shape and structure of the object.

By comparing the results across different images, we aimed to better understand how YOLOv8 and YOLOv11 respond to adversarial patterns and whether certain textures or designs affect its accuracy.

**IV. RESULTS AND ANALYSIS**

**A. YOLOv8 Results**

All the confidence values retrieved from the output of the algorithms were recorded in Table I. In this study, we tested how various animal patterns, such as a cheetah and giraffe pattern, affect the YOLOv8 algorithm. We analyzed different animal placements, such as whole body, half body, and just the body, to see how they would affect confidence values and classification for different animals.

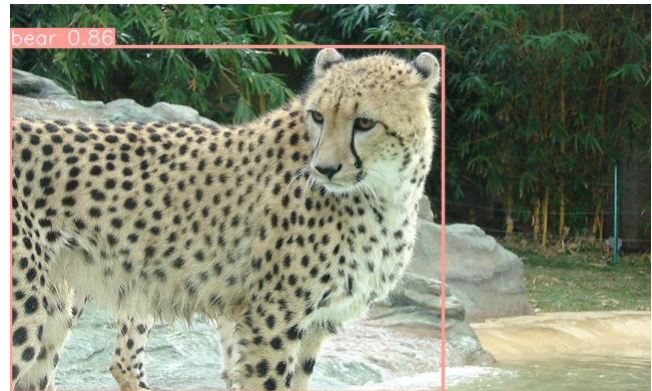


Fig. 4. An image of Cheetah, but was identified as a Bear.



Fig. 5. An image of a Dog, but was identified as a Giraffe.

Tables I - IV: Displays the confidence values (*CValue*) of YOLOv8 when running a subset of images representing animals with different variations of overlay patterns and the classification (*Class*).

TABLE I. Cow Image Results

Animal	Pattern Placement	Cheetah Pattern		Giraffe Pattern	
		<i>CValue</i>	<i>Class</i>	<i>CValue</i>	<i>Class</i>
Cow	No Pattern	0.92			
Cow	Whole Body	0.95	Cow	0.92	Cow
Cow	Half Body	0.90	Cow	0.92	Cow
Cow	Body Only	0.94	Cow	0.95	Cow

TABLE II. Horse Image Results

Animal	Pattern Placement	Cheetah Pattern		Giraffe Pattern	
		CValue	Class	CValue	Class
Horse	No Pattern	0.90			
Horse	Whole Body	0.55	Dog	0.89	Giraffe
Horse	Half Body	0.82, 0.33	Horse, Cow	0.87	Horse
Horse	Body Only	0.63	Horse	0.89	Giraffe

TABLE III. Dog Image Results

Animal	Pattern Placement	Cheetah Pattern		Giraffe Pattern	
		CValue	Class	CValue	Class
Dog	No Pattern	0.83			
Dog	Whole Body	0.50	Dog	0.84	Dog
Dog	Half Body	0.86	Dog	0.91	Dog
Dog	Body Only	0.88	Dog	0.53, 0.47	Dog, Bird

TABLE IV. Elephant Image Results

Animal	Pattern Placement	Cheetah Pattern		Giraffe Pattern	
		CValue	Class	CValue	Class
Elephant	No Pattern	0.92			
Elephant	Whole Body	0.91	Elephant	0.91	Elephant
Elephant	Half Body	0.90	Elephant	0.91	Elephant
Elephant	Body Only	0.91	Elephant	0.90	Elephant

When the cheetah pattern was applied to the whole body, animals like the elephant, horse, and dog resulted in a decreased confidence value. The cow with an animal print resulted in an increased confidence value. However, the cheetah print on the cow caused the confidence value to decrease from 0.92 to 0.90 when the pattern covered just half of the body. Similarly, when the giraffe pattern was applied to the horse, it dropped from 0.90 to 0.87 when it was covered at half the body. However, when the giraffe pattern covered the whole body, the algorithm classified the horse as a giraffe with a 0.89 confidence value. This research proved that the placement of the pattern plays a significant role in confusing the model.

The most vulnerable animals to misclassification with the applied patterns were the horse and the dog. The horse, especially, experienced the most significant shift when the giraffe pattern was added. We assume this is because the

horse and giraffe have very similar body shapes, with the main difference being the length of their necks. This subtle difference, when combined with the giraffe pattern, caused the AI to become confused, especially when the pattern covered the entire body. As seen in Figures 6 and 7, the horse with a full-body pattern would lead to a classification of another animal.

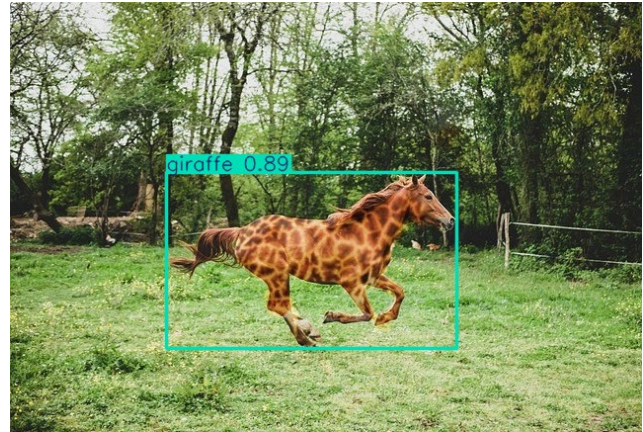


Fig. 6. Picture of a horse with a full giraffe pattern.

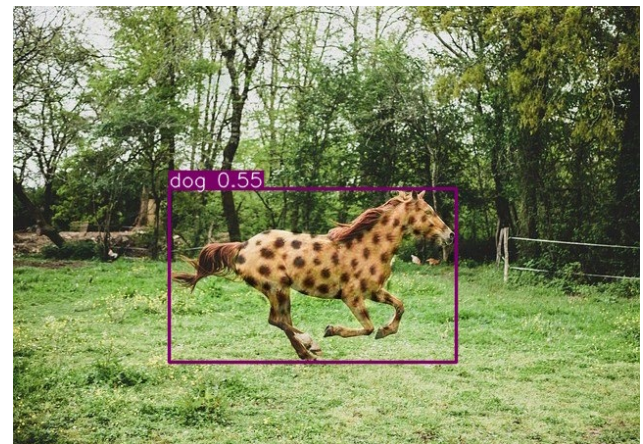


Fig. 7. Picture of a horse with a full cheetah pattern.

The elephant and the cow showed the least amount of change from the original image when patterns were applied. The AI was able to recognize both animals even with the added patterns, with the biggest change being a decrease in the confidence rating by only 0.02. We were not surprised by the giraffe pattern on the cow, as cow patterns can sometimes resemble those of a giraffe. However, previous research has shown that the cow and elephant were correctly identified 98% of the time, which aligns with our findings.

These findings suggest that adversarial patterns, especially when covering the full body of an animal, can greatly reduce the accuracy of object detection in YOLOv8. This resulted in misclassifications and highlighted the vulnerabilities of AI systems to modifications.

**B. YOLOv11 Results**

To get a better understanding of the AI algorithm, I compared YOLOv8 to YOLOv11. YOLOv11 is the newest adaptation within the YOLO AI model. YOLOv11 has research that shows that it has a higher accuracy, efficiency and enhanced detection accuracy [7]. Using the same methods for the YOLOv8 to compare. This model is known for being faster, but to ensure that my research did not have gaps in the model is the reason why we included a further analysis of another model.

Tables V - VIII: Displays the confidence values of YOLOv8 when running a subset of images representing animals with different variations of overlay patterns. Our research showed that even with new updates created by YOLO, the animals' patterns can continue to cause further confusion to the algorithm.

TABLE V. Cow Image Results

Animal	Pattern Placement	Cheetah Pattern		Giraffe Pattern	
		CValue	Class	CValue	Class
Cow	No Pattern	0.94	Cow		
Cow	Whole Body	0.91	Cow	0.90	Cow
Cow	Half Body	0.94	Cow	0.91, 0.42	Cow, Giraffe
Cow	Body Only	0.94	Cow	0.92	Cow

TABLE VI. Horse Image Results

Animal	Pattern Placement	Cheetah Pattern		Giraffe Pattern	
		CValue	Class	CValue	Class
Horse	No Pattern	0.94	Horse		
Horse	Whole Body	0.55	Dog	0.89	Giraffe
Horse	Half Body	0.89, 0.78	Dog, Horse	0.87	Horse
Horse	Body Only	0.76	Dog	0.42, 0.29	Dog, Giraffe

TABLE VII. Dog Image Results

Animal	Pattern Placement	Cheetah Pattern		Giraffe Pattern	
		CValue	Class	CValue	Class
Dog	No Pattern	0.93	Dog		
Dog	Whole Body	0.34	Dog	0.84	Dog
Dog	Half Body	0.91	Dog	0.91	Dog
Dog	Body Only	0.74	Dog	0.78	Dog

TABLE VIII. Elephant Image Results

Animal	Pattern Placement	Cheetah Pattern		Giraffe Pattern	
		CValue	Class	CValue	Class
Elephant	No Pattern	0.93	Elephant		
Elephant	Whole Body	0.92	Elephant	0.86	Elephant
Elephant	Half Body	0.87	Elephant	0.92	Elephant
Elephant	Body Only	0.93	Elephant	0.89	Elephant

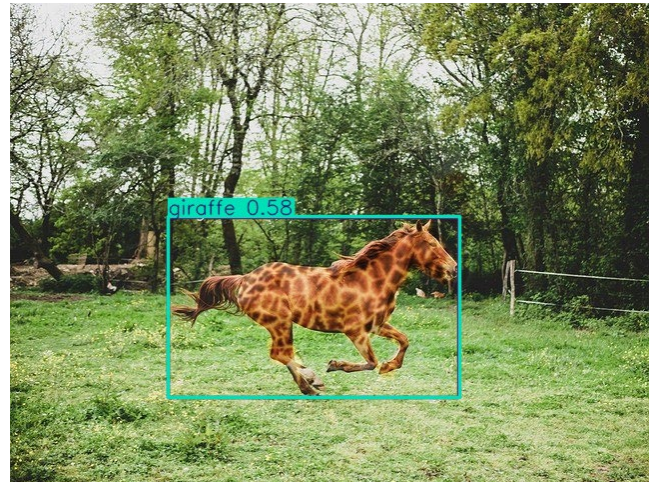


Fig. 8. Picture of a horse with a full giraffe pattern.

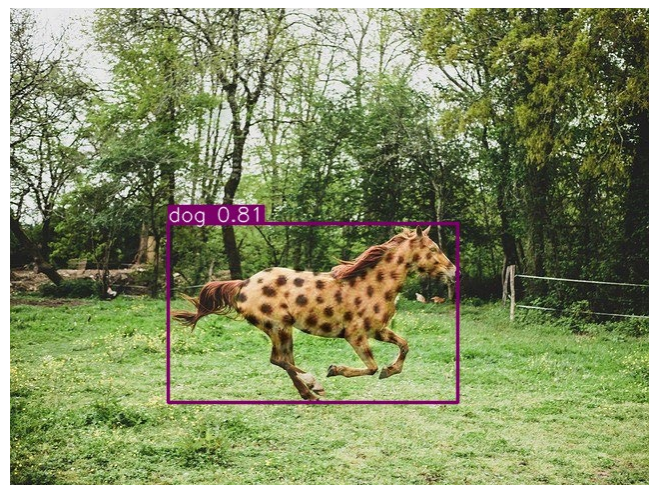


Fig. 9. Picture of a cow with a half giraffe pattern YOLOv11.

While there were a few improvements when it came to animals, such as the horse. When the model observed the horse image with the full body giraffe pattern the confidence level with YOLOv8 was at 0.89 but lower to 0.58 in YOLOv11 (see Figure 8), but the animal was still misclassified (see Figure 9). This demonstrates how vital it is to identify adversarial patches that are natural to the human eye.

When the whole-body pattern was applied to the animals, there were smaller improvements within the object detection phase. The dog image results had changes in accuracy when the giraffe and cheetah patterns were applied. The AI model was able to correctly identify the animal, and if it was unsure, the confidence value increased from a more than the YOLOv8 model. There were mixed results when it came to correctly identifying the object correctly.

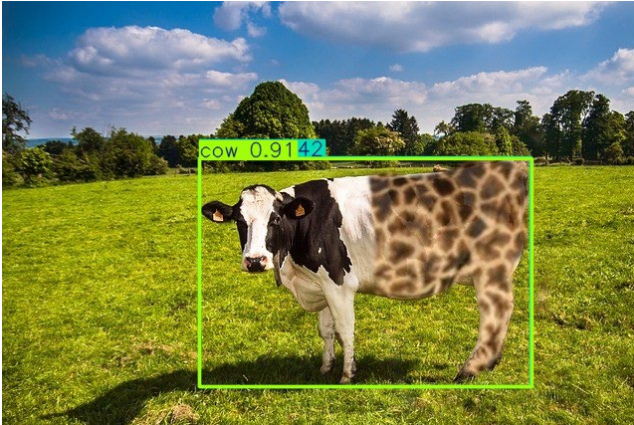


Fig. 10. Picture of a cow with a half giraffe pattern YOLOv11.

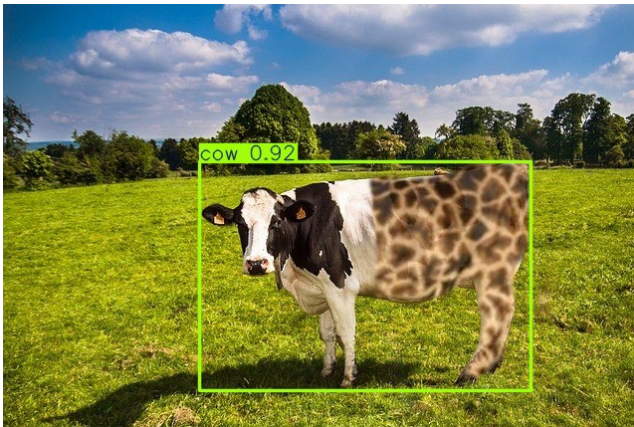


Fig. 11. Picture of a cow with a half giraffe pattern YoloV8.

As shown in Tables V and VIII, in the YOLOv8 model, these were the strongest images that were less affected when it came to adversarial patches. However, in the YOLOv11 model, as seen in Figure 10, a giraffe was also detected at 0.42, while in Figure 11 (YOLOv8) only a cow was detected.

### C. Risk and Implications

AI technologies increase the need for data collection for the improvement of AI algorithms. To prevent negative impacts of using AI technologies it is important to integrate security measures and have risk at an acceptable level.

The findings of this research demonstrate that simple patterns can cause YOLOv8 and YOLOv11 to misclassify or fail to detect objects entirely. While the research might be an

unorthodox approach this is vital for the future of machine learning. These methods were only able to address one of the two types of adversarial attacks, digital attacks. This raises concerns for the future of machine learning technology. This research failed to address the challenges of real-world attacks and if these patterns can also lead to misclassification if printed out.

These concerns are more serious considering that adversarial patches do not need to be visible to be effective. Previous research has demonstrated that there are patches specifically created to be inconspicuous. Animal patterns are patches that can easily blend into the background while still deceiving AI algorithms [10]. This means in the real-world AI detection systems if untrained can be exploited using animal patterns.

Based on research in other literature digital attacks have been proven to successfully cause confusion when put into the physical world [8]. These physical patches can create a different perspective to the AI model by a difference in lighting shadows and camera condition that the patch is displayed on. This causes the images to be distorted. This can further impact AI models being used on autonomous vehicles, surveillance, and aerial drone operations. For example, patterns that are printed on other objects can deceive detection systems the same way that it can in digital overlays.

In addition, through further research adversarial attacks can cause two types of errors: false negatives and false positives [9]. False negatives happen when real world objects are wrongly detected and false positives happen when real world objects have been misidentified. Our results showed both types of errors that happened. When the horse with the full body giraffe pattern was classified as a giraffe it represented a false positive error. Additionally, when confidence values dropped or the animals went undetected it should produce false negative errors. Both errors can cause critical safety concerns to applications leading them to be unreliable.

### V. FUTURE WORK

In the future, we plan to use different control images to further investigate how patterns affect detection within YOLOv8. This will allow us to test various scenarios, including both artificial designs and animal patterns. We also aim to explore how patterns in humans could confuse the AI, especially when the patterns differ. Expanding our dataset will improve the generalization of our study, helping to cover a wider range of patterns and situations.

A vital step to expanding this research would be to use physical testing. Our current experimentation focuses primarily on using photoshop, but as Chen et al. [8] stated, there can be a gap in how adversarial patterns perform in the digital and physical world. Our future work would investigate if these animal patterns would maintain the effect if physically printed or worn by individuals. This would properly fill the gap that is in the research.

Our current patterns cheetah print, and giraffe print are natural and would not attract suspicion if seen on clothes. This connects directly to Bai et al. [10]. They demonstrated the dangers of adversarial patches that follow their Inconspicuous Adversarial Patch (IAP) framework. This framework lowers human detection rates which can lead to a strong attack performance. In the future our research will explore whether animal print patterns could further be used to produce patches that are effective digitally and physically to the human observer. This would reinforce the importance of developing AI systems that are vulnerable to real-world attacks, not just robust ones.

Once the study is expanded, we can train AI models to better recognize patterns and animals, even when the placement of those patterns' changes. In addition to using a framework called Local Interpretable Model-Agnostic Explanations (LIME) to better understand what the AI "sees" and how to improve it. LIME is important so that users interpret AI decision-making and prevent potential issues. Through this framework it can help improve YOLOv8's ability to detect objects accurately, even when faced with adversarial attacks.

A deeper understanding of why AI systems make their decisions will ultimately strengthen YOLOv8 systems against these attacks, reducing misclassification and ensuring more reliable detection in real-world applications.

## VI. CONCLUSION

In this study, we explored how different patterns can impact YOLOv8's ability to detect objects accurately. We found that the placement of patterns on animals plays a key role in confusing the AI algorithm. For example, the most noticeable change occurred when the pattern was applied to the full body of the horse and dog. These results highlight the vulnerabilities in AI algorithms and show where improvements are needed to make them more reliable and resistant to being tricked by patterns.

## ACKNOWLEDGEMENT

This research is funded under NSF CISE-MSI Award # 2131255 (Hampton University) and 2131256 (Florida A&M University).

## REFERENCES

- [1] Gaudenz Boesch. 2024. YOLOv8: A Complete Guide. [2025 Update]. <https://viso.ai/deep-learning/yolov8-guide/>
- [2] Ian J. Goodfellow, Jonathon Shlens, and Christian Szegedy. 2015. Explaining and Harnessing Adversarial Examples. Proceedings of the International Conference on Learning Representation (ICLR 2015). <https://doi.org/10.48550/arXiv.1412.6572>
- [3] Sakshi Indolia, Anil Kumar Goswami, S.P. Mishra, and Pooja Asopa. 2018. Conceptual Understanding of Convolutional Neural Network- A Deep Learning Approach. *Procedia Computer Science* 132, 2018, 679–688. <https://doi.org/10.1016/j.procs.2018.05.069>

- [4] Vidushi Nain, Hari Shankar Shyam, Nitendra Kumar, Padmesh Tripathi, and Mritunjay Rai. 2024. A Study on Object Detection Using Artificial Intelligence and Image Processing–Based Methods. *Mathematical Models Using Artificial Intelligence for Surveillance Systems* (August 2024), 121–148. <https://doi.org/10.1002/9781394200733.ch6>
- [5] Mahmood Sharif, Sruti Bhagavatula, Lujo Bauer, and Michael K. Reiter. 2019. A General Framework for Adversarial Examples with Objectives. *ACM Transactions on Privacy and Security* 22, 3 (June 2019), 1–30. <https://doi.org/10.1145/3317611>
- [6] Emily Sullivan. 2022. Understanding from Machine Learning Models. *The British Journal for the Philosophy of Science* 73, 1 (March 2022), 109–133. <https://doi.org/10.1093/bjps/axz035>
- [7] Gao, X., Cao, C., & Yi, X. 2025. Using the improved YOLOv11 model to enhance computer vision applications for building crack detection algorithms. *Scientific Reports*, 15(1), 38843. <https://doi.org/10.1038/s41598-025-22160-6>
- [8] Chen J., Zhang Y., Liu, C., Chen K., Zou Z. and Shi Z., Digital-to-Physical Visual Consistency Optimization for Adversarial Patch Generation in Remote Sensing Scenes, 2024. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1-17, 2024, Art no. 5623017, <https://doi.org/10.1109/TGRS.2024.3397678>.
- [9] Wang, X., Mei, S., Lian, J., & Lu, Y. 2024. Fooling aerial detectors by background attack via dual-adversarial-induced error identification. *IEEE Transactions on Geoscience and Remote Sensing*, 62, 1–16. <https://doi.org/10.1109/tgrs.2024.3386533>
- [10] Bai, T., Luo, J., & Zhao, J. (2022). Inconspicuous adversarial patches for fooling image-recognition systems on mobile devices. *IEEE Internet of Things Journal*, 9(12), 9515–9524. <https://doi.org/10.1109/ijot.2021.3124815>
- [11] Almaskati, D., Kermanshachi, S., & Pamidimukkula, A. 2023. Autonomous vehicles and traffic accidents. *Transportation Research Procedia*, 73, 321–328. <https://doi.org/10.1016/j.trpro.2023.11.924>
- [12] Grosse, K., & Alahi, A. 2024. A qualitative AI security risk assessment of autonomous vehicles. *Transportation Research Part C: Emerging Technologies*, 169, 104797. <https://doi.org/10.1016/j.trc.2024.104797>