

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

Aniya Hopson, Chutima Boonthum-Denecke#, Idongesit Mkpong-Ruffin*

#Department of Computer Science, Hampton University, Hampton, VA

*Department of Computer and Information Sciences, Florida A&M University, Tallahassee, FL

Abstract

In this paper, we document our findings from previous research and literature related to adversarial examples and object detection. Artificial Intelligence (AI) is an increasingly powerful tool in various fields, particularly in image classification and object detection. As AI becomes more advanced, new methods to deceive machine learning models, such as adversarial patches, have emerged. These subtle modifications to images can cause AI models to misclassify objects, posing a significant challenge to their reliability. This research builds upon our earlier work by investigating how small patches affect object detection on YOLOv8. Last year, we explored patterns within images and their impact on model accuracy. This study extends that work by testing how adversarial patches, particularly those targeting animal patterns, affect YOLOv8's ability to accurately detect objects. We also explore how untrained patterns influence the model's performance, aiming to identify weaknesses and improve the robustness of object detection systems.

Keywords:

Artificial Intelligence(AI), Object Detection, YOLOv8, Adversarial Patches, Machine Learning

Background on Artificial intelligence in Object Detection

Artificial Intelligence (AI) is still in its early stages but 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].

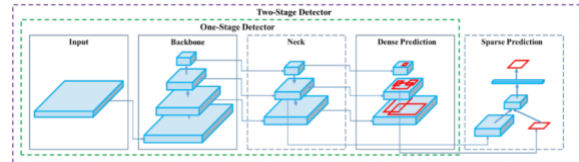


Figure 1: Demonstrates essential mechanic of an object detection model[1]

Growing challenges of 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. Goodfellow et al. (2014) 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. 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].

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.

Significance of the Study

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.

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 can affect YOLOv8's labeling. Animals with patterns, such as cheetahs and some dogs, had the lowest confidence levels in our experiment. As seen in Figure 2 and Figure 3, these animals were sometimes completely mislabeled, even though some of the predictions still had high confidence.

In this study, 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 2 and 3, 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.

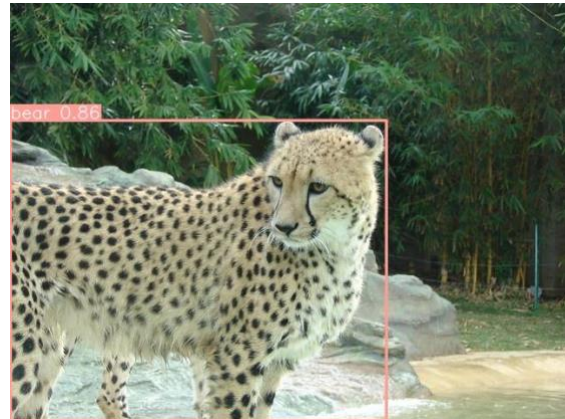


Figure 2: Image results of Cheetah

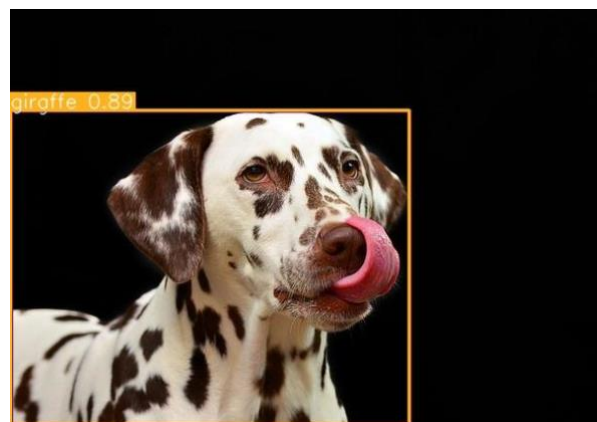


Figure 3: Image results of a Dog

Methodology

For this research, we used images from previous work to analyze how animal

patterns affected object detection using YOLOv8. 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 responds to adversarial patterns and whether certain textures or designs affect its accuracy.

III. Results and Analysis

All of the confidence values retrieved from the output of the algorithms were recorded in

Table 1. 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.

		Cheetah Pattern		Giraffe Pattern	
Animal	Pattern Placement	Confidence Value	Classification	Confidence Value	Classification
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 1: Cow Image Results

		Cheetah Pattern		Giraffe Pattern	
Animal	Pattern Placement	Confidence Value	Classification	Confidence Value	Classification
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 2: Horse Image Results

		Cheetah Pattern		Giraffe Pattern	
Animal	Pattern Placement	Confidence Value	Classification	Confidence Value	Classification
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 3: Dog Image Results

Animal	Pattern Placement	Cheetah Pattern		Giraffe Patter	
		Confidence Value	Classification	Confidence Value	Classi
Elephant	No pattern	0.92			
Elephant	Whole Body	0.91	Elephant	0.91	Elepl
Elephant	Half Body	0.90	Elephant	0.91	Elepl
Elephant	Body Only	0.91	Elephant	0.90	Elepl

Table 4: Elephant Image Results

Table 1-4: Display the confidence values of YOLOv8 ran a subset of images representing animals with different variations of overlay patterns

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 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, in particular, 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 Figure 4 and Figure 5, the horse with a full-body pattern would lead to a classification of another animal.

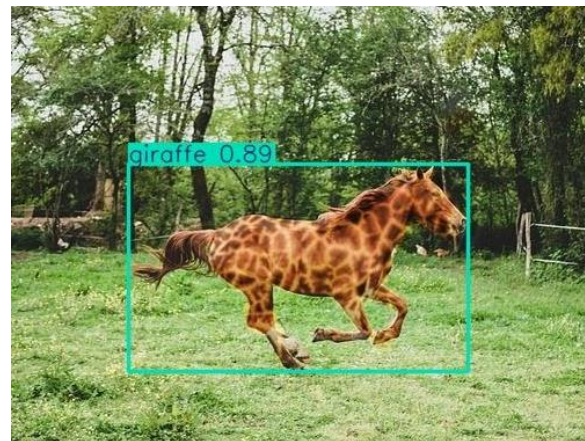


Figure 4: Picture of horse with full giraffe pattern

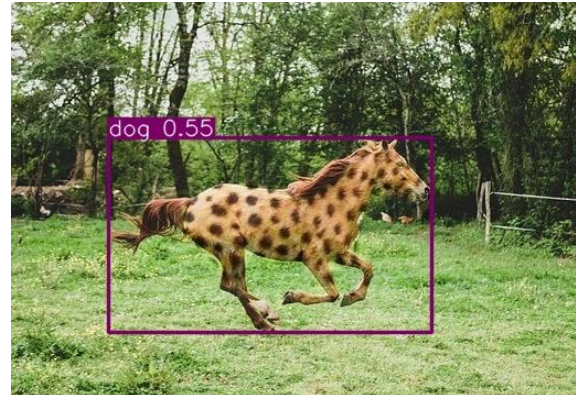


Figure 5: Picture of horse with full cheetah pattern

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.

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 on 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.

Once the study is expanded, we can train AI models to better recognize patterns and animals, even when the placement of those patterns changes. This could help improve YOLOv8's ability to detect objects accurately, even in the face of adversarial attacks. Ultimately, strengthening YOLOv8 systems against these attacks can prevent future misclassification issues, ensuring more reliable AI detection in real-world applications.

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.

References

- [1] Gaudenz Boesch. 2024. YOLOv8: A Complete Guide [2025 Update]. *viso.ai*. Retrieved February 4, 2025 from <https://viso.ai/deep-learning/yolov8-guide/>
- [2] Ian J. Goodfellow, Jonathon Shlens, and Christian Szegedy. 2014. Explaining and Harnessing Adversarial Examples. *arXiv.org*. Retrieved February 4, 2025 from <https://arxiv.org/abs/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,
Lujio 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>