

TDK-thesis

Kassem Anis Bouali

UNIVERSITY OF DEBRECEN

FACULTY OF INFORMATICS

DEPT. OF DATA SCIENCE AND VISUALIZATION



UNIVERSITY of
DEBRECEN

Real-Time Birds Shadow Detection for Autonomous UAVs

Author:

Kassem Anis Bouali

Computer Science MSc

Supervisor:

Prof. Dr. András Hajdu

Full Professor

I agree to submit the following for review.

Thesis supervisor's signature: _____

Debrecen, 2023

Contents

1	Introduction	3
2	Literature Review	5
2.1	Unmanned Aerial Vehicles	6
2.2	Wildlife exploration with UAVs	6
2.2.1	Avian attacks on drones	7
2.2.2	Incidents and impact	7
2.2.3	Mitigation strategies	7
2.3	Object detection in computer vision	8
2.3.1	Traditional methods in object detection	9
2.3.2	Traditional techniques in shadow detection	9
2.4	Deep learning approaches for shadow detection	9
2.4.1	Synthetic datasets and training challenges	10
2.4.2	Our proposed solution	10
3	Methodology and Experimental Design	12
3.1	Introduction	12
3.2	Acquiring a sample dataset	13
3.3	Shadow generation	14
3.3.1	DESOBA Dataset	14
3.3.2	Shadow Generation in the Real-world Network	15
3.4	Data cleaning	16
3.5	Post-processing	16
3.6	Detection model training	18
3.6.1	YOLOv7 Detection algorithm	18
3.6.2	Integration of YOLOv7 in shadow detection	19

CONTENTS

3.6.3	Simplified bounding box detection for shadows	20
3.6.4	Optimizing YOLOv7 for shadow detection	20
3.6.5	Enhancing precision through domain-specific knowledge	21
4	Results and Analysis	22
4.1	Performance Evaluation of the YOLOv7 model	22
4.1.1	Quantitative results	22
4.1.2	Discussion	23
4.1.3	Summary	24
5	Conclusion	25
	Acknowledgements	27
	Bibliography	27
	List of Figures	31

Chapter 1

Introduction

In recent years, Unmanned Aerial Vehicles (UAVs) have emerged as indispensable tools in the realm of wildlife exploration and animal monitoring, enabling researchers to gain unprecedented insights into diverse ecosystems and the creatures that inhabit them. These aerial platforms have ushered in a new era of ecological research, offering the promise of enhanced efficiency and safety in data collection. Nevertheless, their deployment is not without its challenges, chief among them being the persistent threat posed by avian in general attacks and bird attacks in particular, which can disrupt missions and inflict considerable damage on them. Consequently, traditional methods for detecting such threats often prove inadequate, especially when the attack occurs from angles that evade the observation capabilities of UAVs.

However, by discerning the presence of a bird’s shadow, UAVs can gain early warnings of impending attacks, empowering them to take preemptive measures that not only minimize the risk of damage but also ensure the continuation of their missions. Although shadow detection may be limited to specific attack angles, it presents a valuable complement to existing bird detection techniques, enhancing the UAV’s ability to mitigate the impact of avian threats. In response to this pressing challenge, this thesis introduces a groundbreaking approach that leverages the detection of bird shadows to proactively prevent such attacks and reinforce UAV security by unveiling the ShadowBirdCUB dataset, a resource designed for training object detectors to recognize and identify bird shadows. Derived from the well-established CUB-200-2011 dataset, the ShadowBirdCUB dataset serves as a benchmark for assessing the efficacy of shadow-based detection methods.

To delve deeper into the practical implications, our primary goal is to evaluate the performance of state-of-the-art object detectors using this dataset. Through an extensive series of ex-

periments, we aim to showcase the substantial potential of shadow-based detection techniques in effectively countering bird attacks on UAVs. Building on this foundation, equipping UAVs with the capability to detect bird shadows not only bolsters their resilience but also ensures the uninterrupted execution of missions in the fields of wildlife exploration and animal monitoring.

In the subsequent sections of this thesis, we delineate the methodology employed for shadow detection, encompassing multiple stages of innovation. To further elucidate our methodology, we introduce the SGRNet algorithm, which generates synthetic shadows corresponding to segmented birds from the CUB dataset, and discuss the feature extraction methods applied to extract pertinent information from these generated images. With a focus on practical implementation, subsequently, we elucidate the process of training object detectors on the meticulously processed ShadowBirdCUB dataset and present the findings of our comprehensive experiments.

To contextualize the significance of our findings, this research strives to make a profound contribution to the field by introducing an innovative approach to address the challenge of bird attacks on UAVs. Nevertheless, it is essential to acknowledge the method's limitations. Firstly, its effectiveness is contingent on favorable lighting conditions and specific angles for accurate shadow detection. Consequently, scenarios characterized by low light or direct overhead sunlight may compromise the method's effectiveness. Additionally, the method is tailored to UAVs equipped with single frontal cameras, which are prevalent due to their cost-effectiveness.

In light of these limitations, while professional drones with multi-camera setups can directly detect the birds themselves, such solutions are often cost-prohibitive. Furthermore, the shadow-based approach may not provide comprehensive coverage for all potential attack angles. Despite these constraints, our shadow-based detection method remains a valuable tool in the UAV's arsenal for bird attack prevention.

This work is published in the Proceedings of the First International Conference on Artificial Intelligence towards Sustainable Intelligence, as a part of the book series Communications in Computer and Information Science (CCIS, volume 1907) Springer [1].

Chapter 2

Literature Review

In the realm of wildlife exploration, the integration of cutting-edge technology, notably Unmanned Aerial Vehicles (UAVs) commonly known as drones, has opened unprecedented avenues for scientific research. In fact, drones have revolutionized our ability to study wildlife and their habitats. These aerial marvels provide researchers with a unique perspective, allowing them to observe and document animal behavior, study migration patterns, and monitor ecosystems in ways that were once unimaginable.

Researchers frequently encounter challenges when using UAVs, particularly related to avian attacks, such as incidents involving birds. These attacks can significantly impact the accuracy of the collected data, hindering the study of animal behaviors and ecological dynamics. Consequently, there is a growing need to address these challenges through sophisticated techniques. In fact, a significant amount of research has been dedicated to detecting shadows in individual images due to its importance in various applications such as object recognition, scene understanding, and autonomous navigation systems. Shadows, often caused by varying lighting conditions, can significantly affect the performance of computer vision algorithms. Therefore, researchers have devoted considerable effort to developing robust shadow detection techniques. This literature review investigates the methods utilized by researchers in this field, spanning from conventional image processing techniques to modern deep learning approaches. A particular focus is placed on the integration of synthetic data, simply because of the scarcity of real-world datasets, making it crucial for refining accuracy. In the context of this research, a novel dataset has been proposed to overcome the limitations of previous methods, enhancing the precision and reliability of the outcomes.

2.1 Unmanned Aerial Vehicles

UAVs, have rapidly transformed how we perceive and interact with the world. These aerial devices, equipped with advanced sensors and imaging technologies [2], have found applications across diverse fields, from aerial photography and filmmaking to environmental monitoring and disaster management. Also, they offer unique perspectives from the sky, capturing high-resolution imagery and real-time data that were once inaccessible. Their maneuverability, coupled with advancements in GPS technology and machine learning algorithms, enables precise navigation and autonomous flight. This transformative capability has revolutionized various sectors, including agriculture, where drones are employed for crop monitoring and precision farming, and environmental conservation, where they aid in wildlife tracking and habitat assessment.

Moreover, UAVs play a pivotal role in disaster response and search and rescue operations. Their ability to access hard-to-reach areas quickly provides critical situational awareness to emergency responders, facilitating swift and informed decision-making. In scientific research, drones are utilized for environmental surveys, climate research, and archaeological explorations, offering a cost-effective and efficient means to collect valuable data.

2.2 Wildlife exploration with UAVs

Wildlife exploration is fundamental for comprehending ecosystems, species behavior, and the delicate balance of nature, providing crucial insights for conservation efforts and ecological research [3] [4]. Unraveling animal habitats, migration patterns, and population dynamics is vital for informed conservation planning. In this context, UAVs have become indispensable tools. Equipped with high-resolution cameras and sensors, drones enable scientists to observe wildlife behavior, monitor endangered species, and assess environmental impacts on ecosystems, all while minimizing disruption to natural habitats. This technological advancement significantly enhances our ability to study wildlife, making it an invaluable asset in modern ecological research and conservation practices. The advantages of UAVs in wildlife conservation can be summarized as follows:

- **Aerial surveys:** Drones conduct aerial surveys, offering a bird's-eye view of habitats. This perspective aids in population estimation, habitat mapping, and detecting illegal activities such as poaching.

- **Rapid response:** UAVs facilitate rapid response during wildlife emergencies. They can quickly locate and assess stranded or injured animals, enabling timely rescue and conservation efforts.
- **Research and education:** UAV-captured data contributes to scientific research and environmental education. It helps scientists gain insights into animal behavior, fostering public awareness about wildlife conservation challenges.
- **Conservation planning:** UAVs provide detailed topographical and ecological data, aiding conservationists in designing protected areas, identifying corridors for animal movement, and planning habitat restoration initiatives.

2.2.1 Avian attacks on drones

The intersection of advanced technology and wildlife often leads to unexpected challenges, one of which involves avian attacks on drones. Drones, also known as UAVs, are increasingly employed in various fields, including ecological research, environmental monitoring, and wildlife observation. However, the presence of drones in natural habitats sometimes triggers defensive or territorial responses in birds, leading to avian attacks on these aerial devices.

2.2.2 Incidents and impact

Reports of avian attacks on drones have been documented in diverse ecological settings. Birds, especially raptors and territorial species, perceive drones as potential threats or intruders, prompting aggressive behavior. Such attacks can result in damage to the drone's structure, sensors, or propellers, affecting its functionality and data collection capabilities.

These incidents pose challenges not only to the integrity of the drones but also to the safety of both the devices and the wildlife involved. The sudden aggression of birds may lead to mid-air collisions, endangering both the drone and the attacking bird. Moreover, these encounters can disrupt research activities, compromise data collection, and incur repair costs for the drones.

2.2.3 Mitigation strategies

Addressing the issue of avian attacks on drones requires the implementation of mitigation strategies to ensure the safety of both the technology and the wildlife. Researchers and drone

operators have explored various approaches to deter avian aggression and minimize the impact of these encounters.

One common mitigation strategy involves altering the drone's appearance or behavior. Researchers have experimented with visual modifications, such as adding eyespots or predator-like features, to deter birds. Additionally, altering flight patterns, speeds, and altitudes can reduce the likelihood of aggressive responses from birds, allowing drones to navigate without provoking avian attacks.

Acoustic deterrents have also been employed to deter birds from approaching drones. Devices emitting specific bird distress calls or predator sounds can create a no-fly zone around the drone, preventing birds from initiating attacks. However, the effectiveness of acoustic deterrents depends on the species and their sensitivity to such auditory cues.

2.3 Object detection in computer vision

Object detection, a fundamental task in computer vision, has witnessed remarkable progress in recent years, revolutionizing various industries and transforming how we interact with technology. At its core, object detection involves the sophisticated task of identifying and precisely localizing specific objects within images or video frames. This process, once a challenge, has now become a cornerstone of modern artificial intelligence, enabling a myriad of applications that have reshaped the world around us.

The evolution of object detection methods has been nothing short of extraordinary. From traditional computer vision techniques to the rise of deep learning algorithms, researchers and engineers have tirelessly innovated to improve the accuracy, speed, and efficiency of object detection systems. These advancements have paved the way for a wide range of applications, each more innovative than the last.

One such innovative application involves the detection of bird shadows, a pursuit that highlights the versatility of object detection techniques. By leveraging these methods, researchers and conservationists can monitor bird populations and their behaviors more accurately. Understanding the patterns of bird shadows not only aids ornithologists in their research but also contributes significantly to environmental conservation efforts. By precisely identifying and analyzing these shadows, scientists can gather valuable data on bird species, migration patterns, and nesting behaviors, allowing for more informed conservation strategies.

2.3.1 Traditional methods in object detection

Early object detection methods relied on handcrafted features and traditional machine learning algorithms. Techniques such as Histogram of Oriented Gradients (HOG) and Haar-like features combined with classifiers like Support Vector Machines (SVM) provided rudimentary object detection capabilities. However, these methods often struggled with complex scenes, occlusions, and varying object scales [5] [6].

2.3.2 Traditional techniques in shadow detection

Early attempts at shadow detection predominantly relied on classical image processing algorithms. One such approach involved a fusion of adaptive thresholding techniques, generating a binary shadow mask to identify shadow pixels [7]. However, these techniques faced challenges when applied to images with diverse lighting conditions and complex backgrounds. To address these limitations, Murali and Govindan introduced a novel method utilizing the LAB color space [8], particularly the L channel, to control luminance and accurately detect shadowed pixels. Their approach showcased promising results, especially in scenarios where shadows exhibited varying intensity levels.

2.4 Deep learning approaches for shadow detection

Recently, the advent of Deep Neural Networks (DNNs) [9] revolutionized shadow detection. Convolutional Neural Networks (CNNs) [10] and Generative Adversarial Networks (GANs) [11] have shown remarkable capabilities in learning intricate patterns from data. Researchers leveraged these networks to tackle the challenges associated with shadow identification. CNN-based architectures demonstrated the ability to discern subtle differences in pixel values, enabling the detection of faint and elongated shadows [12]. GANs, on the other hand, offered the advantage of generating synthetic shadow data, mitigating the data scarcity issue, although the process of training such networks demanded substantial computational resources and expertise [12].

2.4.1 Synthetic datasets and training challenges

The research community has delved into the realm of synthetic training datasets to overcome the limitations posed by the data-intensive nature of deep learning techniques, especially in the context of the limited availability of real shadow data. In a comprehensive study conducted by Wrenninge M and Unger J, various pre-trained models were applied to synthetic datasets, aiming to address this scarcity [13]. Their investigation illuminated how these models performed differently under varying conditions.

In parallel, researchers have experimented with diverse methods for synthesizing shadows, including the combination of different images and parameters to create extensive synthetic shadow datasets [14]. Another approach involved isolating shadows from images, integrating them into different backgrounds using random parameters, and blending them with real data to introduce variability and construct synthetic training datasets [15]. However, a notable challenge persisted: the uniformity in shadow shapes and contrasts within these datasets. This homogeneity skewed trained models, impeding their effectiveness in real-world scenarios where shadow shapes exhibit significant diversity. Furthermore, the dependence on real data for shadow generation introduced scalability and accessibility issues.

Despite these efforts, the ongoing issue of uniformity in shadow characteristics within datasets continued to hinder progress in the field, impeding the development of models capable of handling diverse shadow scenarios.

2.4.2 Our proposed solution

We introduce our novel ShadowBirdCUB dataset [1], which serves as a benchmark to evaluate the efficacy of deep learning detection methods that rely on detecting the location of the bird's shadow that is reflected on the drone's environment, in order for the drone to gain early indications of upcoming attacks. This strategy empowers UAVs to take the necessary countermeasures to ensure the continuation of their missions. Despite the fact that this method is still limited to specific attack angles, it still presents a valuable complement to existing bird detection techniques, enhancing the UAV's ability to avoid the impact of avian threats.

In addressing the limitations of existing approaches, we present a novel solution that specifically focuses on the utilization of pure synthetic data for training advanced bird shadow detectors. The inherent difficulty in gathering real bird shadow data motivated our reliance on

synthetic data. By doing so, we aim to overcome the challenges associated with real data usage, especially the limitations stemming from uniform shadow shapes and contrasts.

To enhance the quality of the generated shadows, we incorporated the state-of-the-art generative neural network, SGRNet [16]. SGRNet, a deep learning architecture tailored for shadow generation, harnesses the power of generative adversarial networks to create realistic and diverse shadow patterns. By leveraging SGRNet, we ensure that our synthetic dataset encompasses a wide array of shadow shapes, closely mirroring the complexity found in real-world scenarios.

Chapter 3

Methodology and Experimental Design

3.1 Introduction

In this chapter, we present a comprehensive overview of the methodology employed in our study to create a high-quality dataset of state-of-the-art bird shadows because it is crucial for UAV protection from bird attacks as Figure 3.1 illustrates.

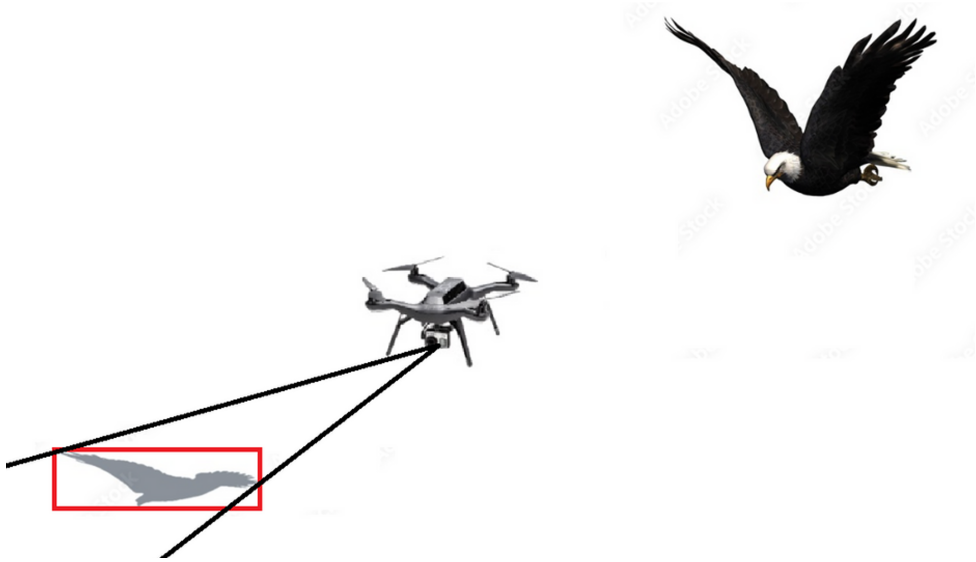


Figure 3.1: Bird's shadow detection for early attack prediction.

Our method (shown in Figure 3.2) encompasses multiple stages, mainly the synthetic shadow generation and the detection model training. In the first stage, the SGRNet [16] model should be trained over DESOBA dataset [16]. Therefore we have utilized the trained model that is already provided. After that, we generated shadows on the CUB dataset [17] images,

by using it and its segmentations [18]. In the second stage, we performed data cleaning and annotation, which formed our proposed ShadowCUB dataset [1], and then the yoloV7 detector [19] is trained over it. In fact, each step will be explained in the subsequent sections of this chapter. Also to mention, all the utilized steps are carefully designed and executed to ensure the accuracy and reliability of our system and the subsequent model’s performance.

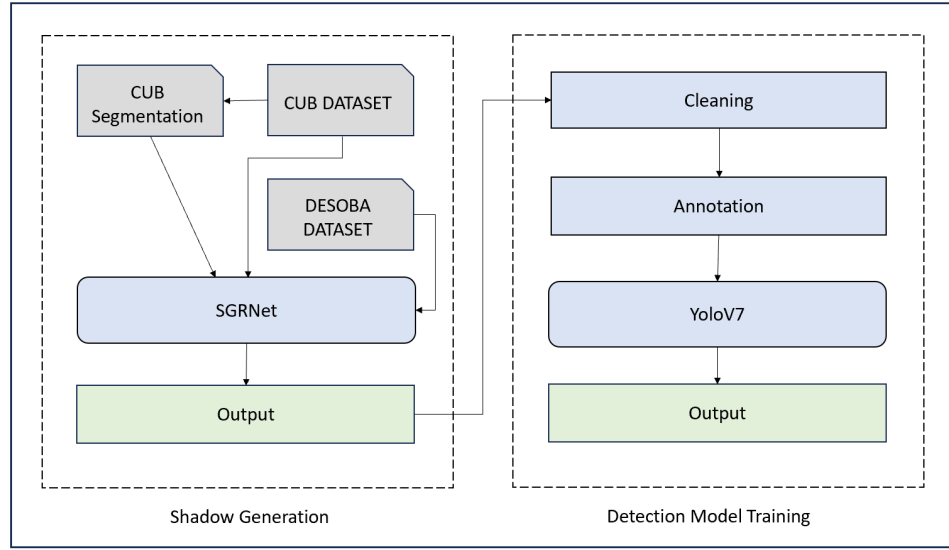


Figure 3.2: Proposed system design.

3.2 Acquiring a sample dataset

Our methodology begins with the acquisition of a suitable sample dataset. For this purpose, we selected the Caltech-UCSD Birds-200 (CUB-200-2011) dataset [17]. This dataset is renowned for its extensive collection of bird images, covering a diverse range of species. The effectiveness of the CUB-200-2011 dataset in various computer vision tasks has been well-documented [20].

However, it’s important to note that the CUB-200-2011 dataset [17] primarily consists of images of birds in static positions, such as standing or sitting. Images of flying birds are relatively limited, which is expected, given the difficulty of capturing clear images of moving subjects. To address this limitation and enhance the diversity of our dataset, we supplemented it with random flying bird images collected from online sources. This addition accounted for approximately 5% of the total dataset size.

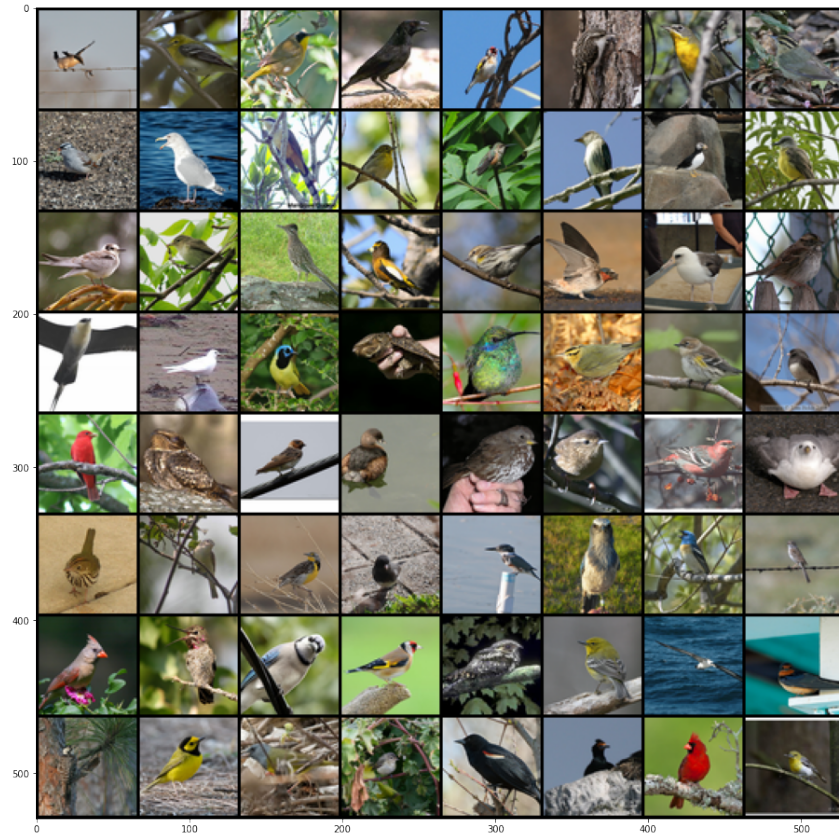


Figure 3.3: Sample images from the Caltech-UCSD Birds-200 dataset [17].

3.3 Shadow generation

3.3.1 DESOBA Dataset

The method we used, utilizes the DESOBA dataset [16], which is a synthesized dataset for Object Shadow Generation. It consists of 839 training images with a total of 2,995 object-shadow pairs and 160 test images with 624 object-shadow pairs. During testing, 9 object-shadow pairs with too small shadows are ignored after resizing to 256×256 , resulting in 615 test image pairs. In fact, it is built upon the Shadow-Object Association dataset (SOBA) [21] [22], which collects real-world images in complex scenes and provides annotated masks for object-shadow pairs. In the DESOBA dataset [16], shadows are removed from the SOBA [21] [22] images, creating the DShadowed Shadow-Object Association dataset. This dataset is utilized for shadow generation tasks and other related applications (see Figure 3.4).

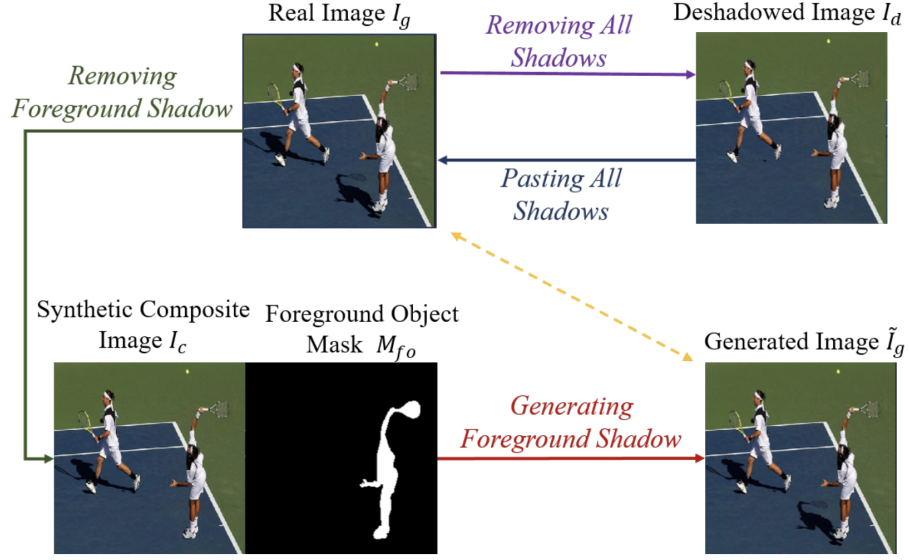


Figure 3.4: Illustration of DESOBA dataset construction: shadows removed from target image to create I_d . Composite image, I_c , generated by replacing shadows on chosen object. Using I_c and mask, M_{fo} , original image, I_g , reconstructed with foreground shadow [16].

3.3.2 Shadow Generation in the Real-world Network

SGRNet (Shadow Generation in the Real-world Network) [16] is a state-of-the-art neural network model for shadow generation developed by Hong, Yan, Niu, Li, and Zhang. We employed it for generating synthetic shadows on bird images, which is a crucial step in our methodology. This model basically comprises two stages (see Figure 3.5): a shadow mask prediction stage and a shadow filling stage. The shadow mask generator predicts the binary shadow mask of the foreground object, utilizing a Cross-Attention Integration (CAI) layer to attend relevant illumination information from the background. The shadow-filling stage fills the shadow mask using an illumination model to produce the final target image with foreground shadow. A conditional discriminator ensures the generated shadow mask and target image are close to real counterparts.

In our experiment, we did input the binary segmentation masks derived from the CUB-200-2011 dataset [18] as inputs to generate shadows on bird images. The results are shown in Figure 3.6

In fact, while SGRNet demonstrated impressive results in most cases, it faced challenges when generating shadows for images of flying birds. One particular challenge arises when there are limited surfaces available for shadow reflection, such as in images of birds flying against a clear sky. In such scenarios, the generated shadows may appear less as demonstrated also in 3.7.

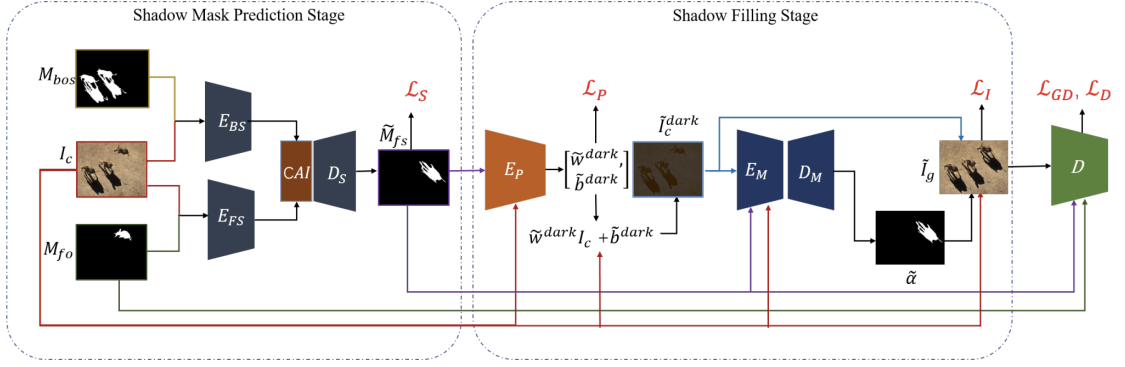


Figure 3.5: SGRNet model architecture [16].

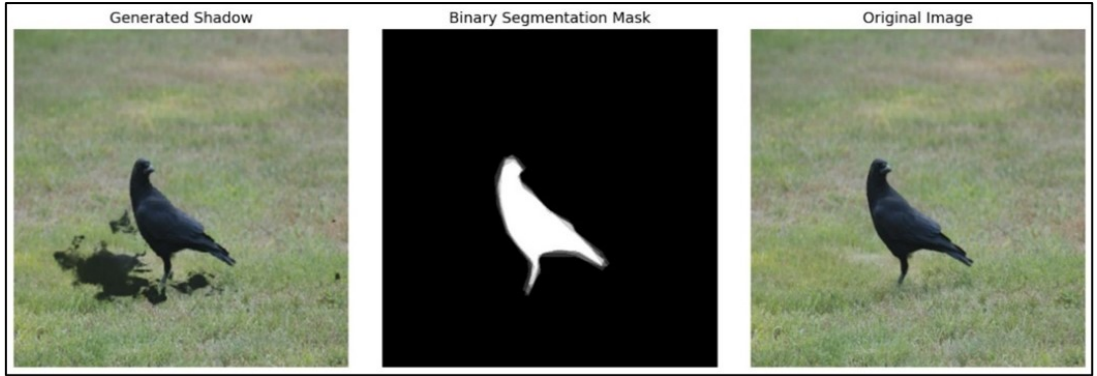


Figure 3.6: Shadow generation process using SGRNet [1].

3.4 Data cleaning

After the shadow generation process, it became evident that some of the generated images were of poor quality and not suitable for training purposes. These low-quality images could potentially introduce noise and negatively impact the performance of our detection model. To maintain the integrity of our dataset, we conducted a thorough manual cleaning process.

The manual cleaning process involved carefully reviewing each image and removing those that exhibited significant artifacts, blurriness, or other issues that could affect the accuracy of shadow detection. This meticulous approach ensured that our dataset only contained high-quality images suitable for training.

3.5 Post-processing

With the dataset cleaned and shadows generated, the next step involved post-processing to obtain accurate shadow annotations. Our post-processing techniques varied based on the



Figure 3.7: Failure analysis of SGRNet: examples of misclassified samples [1].

characteristics of the data and the nature of the shadows.

In the initial phase of post-processing, we applied an image processing approach. This approach leveraged bird segmentation to isolate the main bird from each image. Subsequently, the images were converted to grayscale, enabling the application of a threshold to detect dark pixels associated with shadow regions. To extract shadow contours, we employed the "Find Contours" algorithm and sorted them based on their area in descending order. Prioritizing smaller contours allowed us to capture finer details of the shadows accurately. Notably, the largest contour, corresponding to the background, was excluded from the grouping process to prevent its merging with smaller shadow contours. This approach resulted in precise shadow annotations and spatial extent representations.

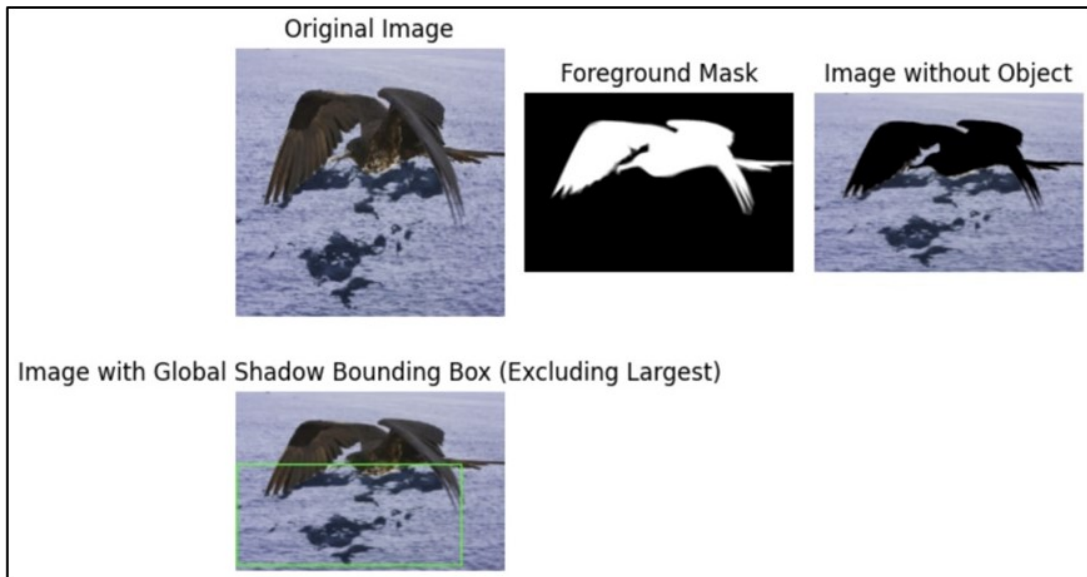


Figure 3.8: Image processing method for dataset labeling [1].

While the automated image processing approach proved effective for many images, it had

limitations when dealing with complex scenes or images with extensive dark regions. In such cases, the automated method might fail to detect shadows accurately.

To address these limitations, we adopted an alternative annotation technique that utilized the SAM (Segment Anything Model) algorithm provided by Meta [23]. This technique involved extracting segmentation masks and linking edge points to form polygons, which were used as labels for shadow regions. This method proved to be particularly useful in cases where distinguishing shadows from other dark areas of the image was challenging.



Figure 3.9: Demonstration of polygon-based annotation approach [1].

In cases where the automated post-processing techniques were insufficient, manual annotation was employed. Human annotators meticulously reviewed and annotated challenging images, ensuring precise identification and delineation of shadow regions. Manual annotation was especially valuable for images with intricate shadow shapes and challenging lighting conditions. By combining automated and manual annotation approaches, our methodology achieved a balance between efficiency and precision in shadow labeling.

3.6 Detection model training

3.6.1 YOLOv7 Detection algorithm

The YOLO (You Only Look Once) algorithm has revolutionized the landscape of real-time object detection algorithms by providing a greatly improved real-time object detection accuracy without increasing the inference costs. Its unique ability to perform object detection in a single forward pass of the neural network has made it the go-to choice for applications demanding speed and accuracy. Among the various iterations of YOLO, YOLOv7 (You Only Look Once

version 7) [19], stands out as a significant advancement, incorporating crucial enhancements in accuracy, speed, and adaptability as Figure 3.10 shows.

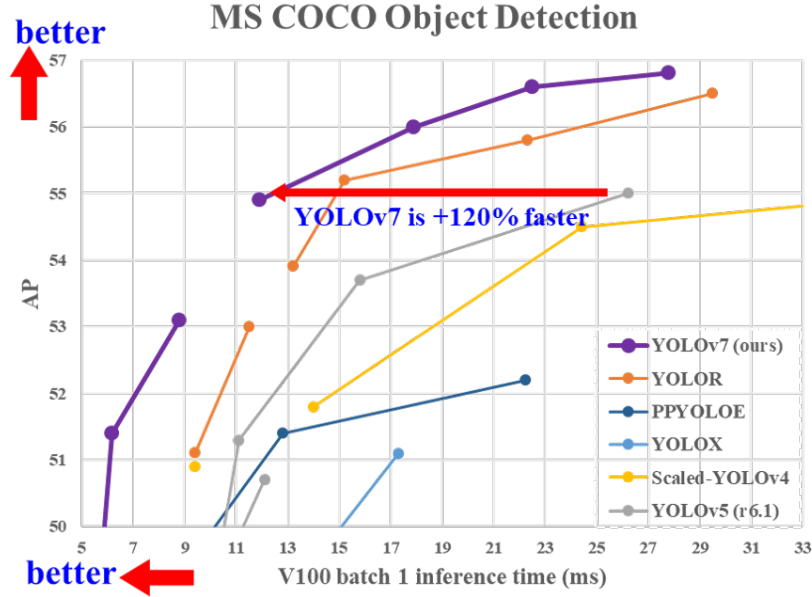


Figure 3.10: YOLO performance comparison [19].

As shown in the benchmarks, when compared to other known object detectors, YOLOv7 [19] can effectively reduce about 40% parameters and 50% computation of state-of-the-art real-time object detections, and achieve faster inference speed and higher detection accuracy. In general, it relies on a faster and stronger network architecture (see Figure 3.11) which offers more effective feature integration, more accurate object detection performance, more robust loss function, and an increased label assignment and model training efficiency. As a result, YOLOv7 [19] requires several times cheaper computing hardware than other deep learning models. It can be trained much faster on small datasets without any pre-trained weights.

3.6.2 Integration of YOLOv7 in shadow detection

With the annotated dataset meticulously prepared, our focus shifted to training a shadow detection model tailored to our specific requirements. Although YOLO [19] is the fastest state-of-the-art detection model, it is still not fast enough to run on low computing power devices such as UAVs even with the aid of Intel's Movidius Neural Compute Stick.

So, In light of the unique characteristics of shadow detection and the demands of our application, we opted for the YOLOv7 (tiny) detection model which is a variation of the YOLOv7

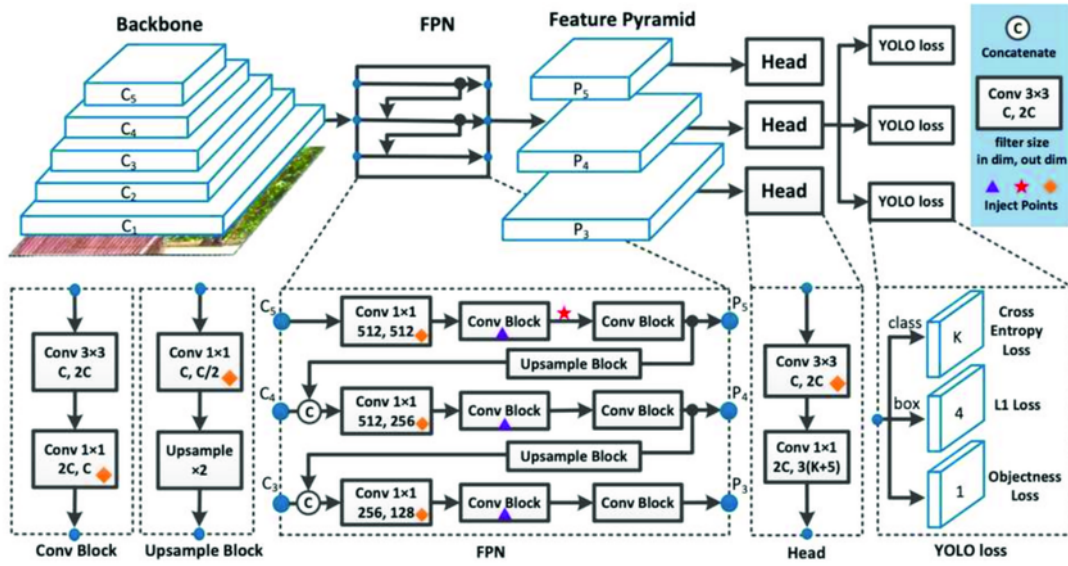


Figure 3.11: The YOLOv7 architecture [19].

model [19], but with approximately 442% faster architecture, achieving upwards of 244 FPS on a single GPU as seen in Figure 3.12.

3.6.3 Simplified bounding box detection for shadows

A fundamental consideration in our model selection process was the inherent simplicity of detecting bounding boxes around shadows compared to the intricacies of detailed shadow segmentation. YOLOv7's efficiency in recognizing bounding boxes made it an ideal choice for our task, aligning perfectly with the specific needs of shadow detection in our application.

3.6.4 Optimizing YOLOv7 for shadow detection

Our team meticulously fine-tuned the YOLOv7 (tiny) model for shadow detection. Leveraging the model's innate adaptability, we optimized the training process to focus on shadow-specific features. By fine-tuning the model's parameters and training it on our annotated dataset, we tailored YOLOv7 to excel in accurately detecting shadows in diverse environmental conditions.

Model	Train	Test	mAP	FLOPS	FPS	Cfg	Weights
SSD300	COCO trainval	test-dev	41.2	-	46		link
SSD500	COCO trainval	test-dev	46.5	-	19		link
YOLOv2 608x608	COCO trainval	test-dev	48.1	62.94 Bn	40	cfg	weights
Tiny YOLO	COCO trainval	test-dev	23.7	5.41 Bn	244	cfg	weights
SSD321	COCO trainval	test-dev	45.4	-	16		link
DSSD321	COCO trainval	test-dev	46.1	-	12		link
R-FCN	COCO trainval	test-dev	51.9	-	12		link
SSD513	COCO trainval	test-dev	50.4	-	8		link
DSSD513	COCO trainval	test-dev	53.3	-	6		link
FPN FRCN	COCO trainval	test-dev	59.1	-	6		link
Retinanet-50-500	COCO trainval	test-dev	50.9	-	14		link
Retinanet-101-500	COCO trainval	test-dev	53.1	-	11		link
Retinanet-101-800	COCO trainval	test-dev	57.5	-	5		link
YOLOv3-320	COCO trainval	test-dev	51.5	38.97 Bn	45	cfg	weights
YOLOv3-416	COCO trainval	test-dev	55.3	65.86 Bn	35	cfg	weights
YOLOv3-608	COCO trainval	test-dev	57.9	140.69 Bn	20	cfg	weights
YOLOv3-tiny	COCO trainval	test-dev	33.1	5.56 Bn	220	cfg	weights
YOLOv3-spp	COCO trainval	test-dev	60.6	141.45 Bn	20	cfg	weights

Figure 3.12: Tiny-YOLO mAP score comparison on the COCO dataset [24].

3.6.5 Enhancing precision through domain-specific knowledge

Beyond the inherent capabilities of YOLOv7, we enriched the model's understanding of shadows by integrating domain-specific knowledge. This involved refining the training dataset to include diverse shadow patterns, ensuring that the model could accurately discern shadows from various light sources and object interactions. By enhancing the model's training data with this nuanced information, we elevated its precision, making it adept at handling complex shadow detection scenarios.

Chapter 4

Results and Analysis

4.1 Performance Evaluation of the YOLOv7 model

In this section, we present the results of the performance evaluation of the YOLOv7 [19] detection model on our custom dataset of bird shadow images. We employed various standard metrics to assess the model’s accuracy and effectiveness in detecting and localizing bird shadows.

4.1.1 Quantitative results

The performance of the YOLOv7 model was evaluated using a range of standard metrics, as summarized in Table 4.1. Notably, these metrics provide valuable insights into the model’s precision, recall, and overall performance.

Table 4.1: YOLOv7 object detection results on our custom dataset.

[1]	Performance Metric	Value
	mAP50	0.52
	mAP50-95	0.30
	Box Precision	0.70
	Recall	0.51

The model demonstrated a precision of 70.1% and a recall of 51.3%. The mean average precision (mAP50) at an IoU threshold of 0.50 reached 52%, indicating satisfactory performance across different confidence score thresholds. However, it’s important to note that the mAP over a broader range of IoU thresholds (mAP50-95) was 29.8%, suggesting a decline in performance at higher IoU stringency levels.

4.1.2 Discussion

The observed decline in mAP metrics at higher IoU stringency levels is a reasonable outcome given the nature of our evaluation dataset. We utilized polygon and segmented annotations for shadow labeling, which define objects at the pixel level. This pixel-level precision can introduce shape variations that challenge the accurate prediction of rectangular bounding boxes. Consequently, when mapping a rectangular predicted bounding box to a complex ground truth-shaped polygon, the model's performance scores decrease, as depicted in Figure 4.1.

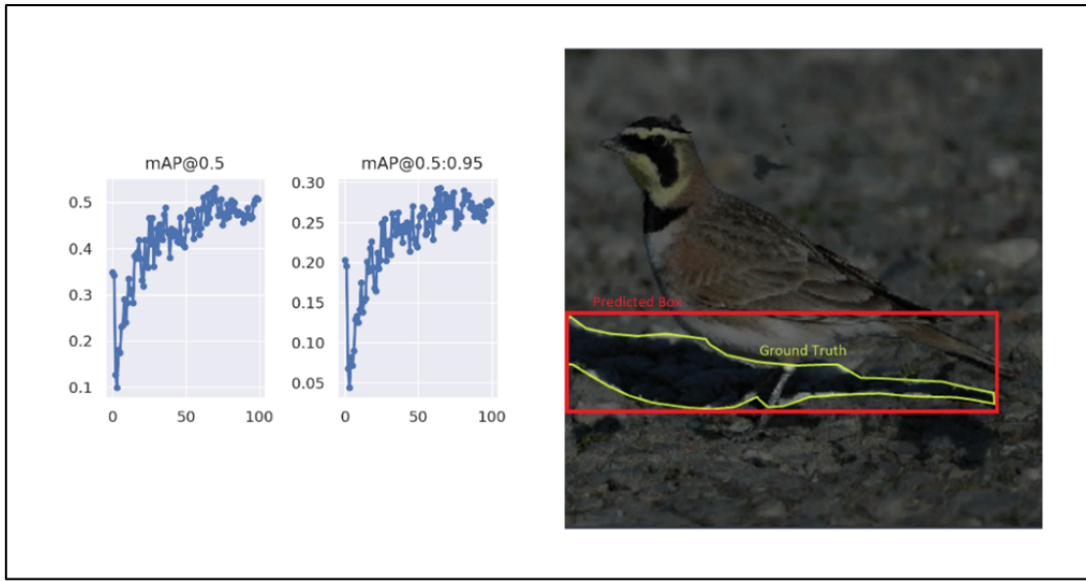


Figure 4.1: Impact of Polygon/Segmented labeling on object detection evaluation metrics [1].

However, despite the challenges posed by polygon and segmented annotations, the YOLOv7 model demonstrated excellent performance and high accuracy in real-life scenarios. Figure 4.2 showcases the model's ability to detect and localize bird shadows accurately under varying environmental conditions.

Furthermore, when comparing our shadow detection approach with existing methods, our model exhibited promising results and outperformed many other models in terms of adaptability, especially for traditional UAVs. This adaptability is crucial in scenarios where precise bounding box annotations might not be readily available or practical.

Overall, while the mAP metrics may indicate a decline in performance at high IoU stringency levels due to polygon and segmented labeling, the YOLOv7 model's real-life performance demonstrates its robustness and effectiveness in detecting bird shadows across diverse environmental conditions.

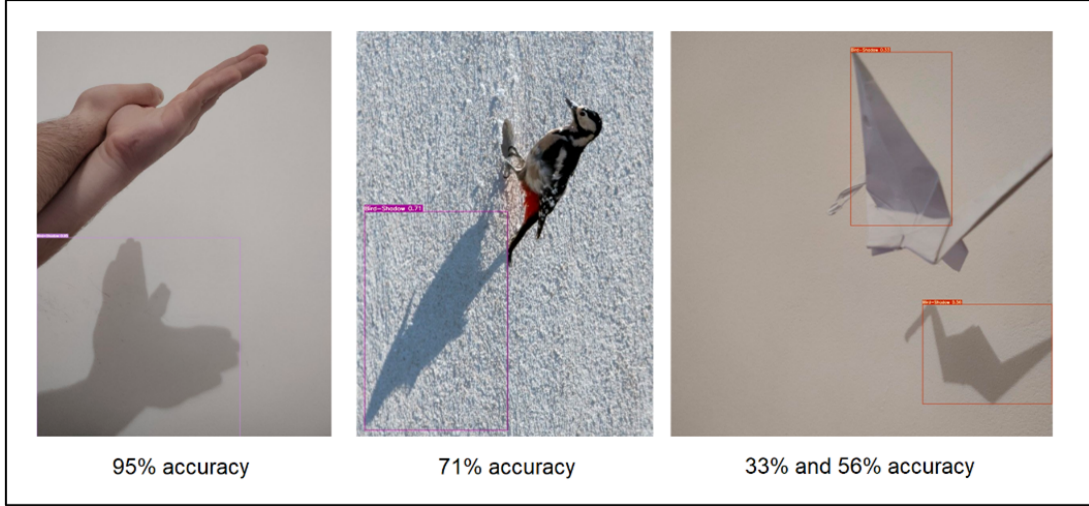


Figure 4.2: Performance of the trained model in real-life scenarios [1].

4.1.3 Summary

In this chapter, we presented the results of the evaluation of the YOLOv7 detection model on our specialized dataset of bird shadow images which revealed noteworthy findings. Despite grappling with intricacies linked to polygonal and segmented annotations, leading to a decline in mean average precision (mAP) metrics at heightened Intersection over Union (IoU) stringency levels, the model exhibited praiseworthy accuracy and efficacy in real-world contexts. Amidst these challenges, the YOLOv7 model showcased its ability to adeptly detect and precisely localize bird shadows amidst diverse environmental nuances, as evidenced by its robust performance in real-world scenarios. This adaptability is particularly significant for conventional Unmanned Aerial Vehicles (UAVs), where acquiring precise bounding box annotations can pose formidable challenges.

Overall, while the evaluation metrics highlighted hurdles arising from polygonal and segmented annotations, the YOLOv7 model's resilience in real-world applications underscores its pivotal role in accurately discerning bird shadows. This capability positions it as an invaluable asset in numerous practical contexts.

Chapter 5

Conclusion

In conclusion, our study presents a groundbreaking approach for enhancing UAV protection through the utilization of shadow-based detection techniques to mitigate bird attacks. The development of the ShadowBirdCUB dataset, meticulously curated for the explicit purpose of training object detection models, served as the cornerstone of our research endeavors. The results obtained from our comprehensive experiments underscore the remarkable effectiveness of our approach, with the model achieving an impressive precision of 70% and a recall of 51% in the detection and precise localization of bird shadows.

These quantitative outcomes underscore the model's robustness and adaptability in diverse environmental conditions, highlighting its potential for real-world applications. While it is important to acknowledge the existence of certain limitations, such as occasional false positives, our solution undeniably represents a significant advancement in the realm of bird detection and UAV protection.

By offering the capability to anticipate and respond to bird threats with a high degree of accuracy, our approach contributes substantially to enhancing the resilience and uninterrupted operation of UAV missions, particularly in the fields of wildlife exploration and animal monitoring. The implications of our research extend far beyond the confines of UAV protection, offering valuable insights into the broader domain of computer vision and object detection.

This study paves the way for future research and innovation in the area of UAV safety and wildlife conservation. The successful fusion of state-of-the-art technology with a deep understanding of avian behavior and environmental dynamics positions our work as a significant milestone in ensuring the harmonious coexistence of unmanned aerial vehicles and the natural world.

In light of the promising results and the wealth of opportunities for further refinement, we look forward to the integration of our approach into practical UAV systems, with the ultimate aim of fostering a safer and more sustainable environment for both technology and wildlife.

The journey does not conclude here but rather extends to new horizons, where innovation and conservation continue to intersect, creating a brighter future for aerial exploration and ecological preservation.

Acknowledgements

I am deeply grateful to my thesis supervisor, Dr. András Hajdu, for his invaluable guidance and unwavering support throughout my academic journey. His mentorship has been instrumental in shaping this thesis. My heartfelt thanks go to my family for their endless love and encouragement, and to my friends for their companionship and understanding. Their support provided the strength I needed to overcome challenges. I also appreciate the knowledge shared by professors, researchers, and colleagues, enhancing my learning experience. Special thanks to the University of Debrecen for providing a conducive environment for research and growth.

Bibliography

- [1] Kassem Anis Bouali and András Hajdu. “Real-Time Birds Shadow Detection for Autonomous UAVs”. In: *Artificial Intelligence: Towards Sustainable Intelligence*. to appear. Cham: Springer Nature Switzerland, 2023.
- [2] F. Nex et al. “UAV in the advent of the twenties: Where we stand and what is next”. In: *ISPRS Journal of Photogrammetry and Remote Sensing* 184 (2022), pp. 215–242. ISSN: 0924-2716. DOI: <https://doi.org/10.1016/j.isprsjprs.2021.12.006>. URL: <https://www.sciencedirect.com/science/article/pii/S0924271621003282>.
- [3] GEORGE IV, Leonard Pearlstine, and H. PERCIVAL. “An Assessment of Small Unmanned Aerial Vehicles for Wildlife Research”. In: *Wildlife Society Bulletin* 34 (Oct. 2006), pp. 750–758. DOI: 10.2193/0091-7648(2006)34%5B750:AA0SUA%5D2.0.CO;2.
- [4] Matthew P Brinkman. “Applying UAV systems in wildlife management”. In: *Proceedings of the Vertebrate Pest Conference*. Vol. 29. 29. 2020.
- [5] N. Dalal and B. Triggs. “Histograms of oriented gradients for human detection”. In: *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05)*. Vol. 1. 2005, 886–893 vol. 1. DOI: 10.1109/CVPR.2005.177.
- [6] M.A. Hearst et al. “Support vector machines”. In: *IEEE Intelligent Systems and their Applications* 13.4 (1998), pp. 18–28. DOI: 10.1109/5254.708428.
- [7] Xiaoyan Zou et al. *Drone Shadow Tracking*. 2019. arXiv: 1905.08214 [cs.CV].
- [8] Saritha Murali and V. K. Govindan. “Shadow Detection and Removal from a Single Image Using LAB Color Space”. In: *Cybernetics and Information Technologies* 13.1 (2013), pp. 95–103. DOI: doi:10.2478/cait-2013-0009. URL: <https://doi.org/10.2478/cait-2013-0009>.

- [9] Jürgen Schmidhuber. “Deep learning in neural networks: An overview”. In: *Neural Networks* 61 (2015), pp. 85–117. DOI: 10 . 1016 / j . neunet . 2014 . 09 . 003. URL: <https://doi.org/10.1016%2Fj.neunet.2014.09.003>.
- [10] Keiron O’Shea and Ryan Nash. *An Introduction to Convolutional Neural Networks*. 2015. arXiv: 1511.08458 [cs.NE].
- [11] Ian J. Goodfellow et al. *Generative Adversarial Networks*. 2014. arXiv: 1406 . 2661 [stat.ML].
- [12] Sai Paavan Kumar Dornadula, Pascal Brunet, and Dr. Susan Elias. *AI driven shadow model detection in agropv farms*. 2023. arXiv: 2304.07853 [cs.CV].
- [13] Magnus Wrenninge and Jonas Unger. *Synscapes: A Photorealistic Synthetic Dataset for Street Scene Parsing*. 2018. arXiv: 1810.08705 [cs.CV].
- [14] Naoto Inoue and Toshihiko Yamasaki. “Learning From Synthetic Shadows for Shadow Detection and Removal”. In: *IEEE Transactions on Circuits and Systems for Video Technology* 31.11 (2021), pp. 4187–4197. DOI: 10.1109/TCSVT.2020.3047977.
- [15] Mohammed El Amine Mokhtari et al. “Semi-synthetic Data for Automatic Drone Shadow Detection”. In: Oct. 2022. DOI: 10.14428/esann/2022.ES2022-82.
- [16] Yan Hong, Li Niu, and Jianfu Zhang. “Shadow Generation for Composite Image in Real-world Scenes”. In: *AAAI* (2022).
- [17] C. Wah et al. *CUB-200-2011*. Tech. rep. CNS-TR-2011-001. California Institute of Technology, 2011.
- [18] Ryan Farrell. *CUB-200-2011 Segmentations*. 2022. DOI: 10.22002/D1.20097.
- [19] Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. “YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2023, pp. 7464–7475.
- [20] Thomas Verelst et al. “Spatial Consistency Loss for Training Multi-Label Classifiers from Single-Label Annotations”. In: *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. 2023, pp. 3868–3878. DOI: 10 . 1109 / WACV56688 . 2023 . 00387.

- [21] Tianyu Wang et al. “Instance Shadow Detection”. In: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2020, pp. 1877–1886. DOI: 10.1109/CVPR42600.2020.00195.
- [22] Tianyu Wang et al. “Single-Stage Instance Shadow Detection With Bidirectional Relation Learning”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2021, pp. 1–11.
- [23] Alexander Kirillov et al. “Segment Anything”. In: *arXiv:2304.02643* (2023).
- [24] Joseph Redmon. *Website*. URL: <https://pjreddie.com/darknet/yolo/>.

List of Figures

3.1	Bird's shadow detection for early attack prediction.	12
3.2	Proposed system design.	13
3.3	Sample images from the Caltech-UCSD Birds-200 dataset [17].	14
3.4	Illustration of DESOBA dataset construction: shadows removed from target image to create I_d . Composite image, I_c , generated by replacing shadows on chosen object. Using I_c and mask, M_{fo} , original image, I_g , reconstructed with foreground shadow [16].	15
3.5	SGRNet model architecture [16].	16
3.6	Shadow generation process using SGRNet [1].	16
3.7	Failure analysis of SGRNet: examples of misclassified samples [1].	17
3.8	Image processing method for dataset labeling [1].	17
3.9	Demonstration of polygon-based annotation approach [1].	18
3.10	YOLO performance comparison [19].	19
3.11	The YOLOv7 architecture [19].	20
3.12	Tiny-YOLO mAP score comparison on the COCO dataset [24].	21
4.1	Impact of Polygon/Segmented labeling on object detection evaluation metrics [1].	23
4.2	Performance of the trained model in real-life scenarios [1].	24