

# Real-Time Object Detection in Low-Light Environments using YOLOv8: A Case Study with a Custom Dataset

Sikkandar Basha Z

Thiagarajar college of engineering  
Madurai, Tamilnadu

Ganesh Shankar Ram B

Thiagarajar college of engineering  
Madurai, Tamilnadu

**Abstract**—Object detection in low-light conditions presents significant challenges due to the reduced visibility and poor illumination, particularly in real-time applications. This paper proposes a novel approach using the YOLOv8 model for real-time object detection in night-time conditions. A custom dataset comprising various objects captured in low-light environments was utilized to train and evaluate the model. The results demonstrate superior performance in terms of speed and accuracy compared to previous models, particularly YOLOv3. We also include an analysis of the model's real-time performance using a custom video feed. Our findings show that YOLOv8 outperforms earlier YOLO versions in detecting objects accurately and quickly in low-light, real-time scenarios, making it a promising solution for night-time surveillance and other security-related applications.

**Keywords**—Low-light environments, Yolov8, Object detection, Real-time

## I. INTRODUCTION

Object detection has become a central task in the field of computer vision, especially with the rise of deep learning models capable of identifying and localizing multiple objects in images and video streams. Over the past decade, several datasets and models have been developed to benchmark and advance object detection capabilities, such as PASCAL VOC, MS COCO, and ImageNet. These datasets have played a significant role in improving the accuracy and efficiency of object detection algorithms, particularly in well-lit, structured environments. However, when it comes to real-time object detection in low-light conditions, existing datasets fall short in capturing the complexities of night-time scenes. Traditional datasets like PASCAL VOC and MS COCO are rich in object variety and annotation depth, but they are primarily focused on daytime imagery or well-illuminated indoor environments. The objects in these datasets are typically visible under natural or artificial light sources, and the annotations are based on high-resolution, clear images. While these datasets have been instrumental in advancing the field, they are not designed to address the unique challenges of low-light environments, where object detection becomes significantly more complex due to factors like noise, poor contrast, shadows, and low visibility.

Moreover, although there are specialized datasets for low-light conditions, such as ExDark, they are still limited in scope. ExDark, for example, provides images taken under various low-light conditions but only includes 12 object categories. The diversity of object types and environmental conditions is limited, making it difficult to train robust models for diverse real-world applications. Additionally, the images in ExDark, while useful, do not fully replicate the specific challenges encountered in night-time surveillance, where objects may be small, partially occluded, or have low reflectivity in infrared light. Our custom dataset stands out by addressing these limitations in several ways. First, it focuses specifically on real-time object detection in low-light environments, offering a diverse range of objects that are typically encountered in outdoor night-time scenes, such as pedestrians, bicycles, cars, and other common items. These objects are captured in various conditions, including different levels of ambient lighting, shadow, and occlusion, which better simulate real-world night-time surveillance scenarios. Furthermore, unlike many existing datasets that are primarily static images, our dataset includes video sequences, which are crucial for evaluating real-time object detection models. This allows for the capture of dynamic scenes, where objects are not only detected but also tracked across frames. The inclusion of video data in our custom dataset enables us to assess the temporal consistency and tracking performance of models like YOLOv8, which are essential for applications such as security surveillance, traffic monitoring, and autonomous driving in low-light environments.

In addition, our dataset incorporates various lighting conditions, from twilight to near-complete darkness, captured using infrared (IR) cameras to simulate real-world night-time conditions. This adds another layer of complexity that many other datasets fail to address, as IR reflections, noise, and low signal-to-noise ratios can significantly affect object detection performance. By including a range of IR lighting intensities and camera angles, our dataset ensures that models trained on it are more robust to changes in illumination, making them more effective in challenging real-world applications. The unique combination of object variety, dynamic video sequences, diverse lighting conditions, and IR imagery makes our custom dataset an invaluable resource for training and evaluating object detection models specifically designed for

low-light, real-time applications. By using this dataset, we aim to push the boundaries of what is possible in night-time object detection, particularly in the context of real-time detection and tracking.

In this paper, we explore the potential of the YOLOv8 model, the latest in the YOLO family, for real-time object detection using our custom dataset. YOLOv8 introduces several architectural improvements over its predecessors, including enhanced feature extraction and better spatial understanding through transformer layers. These improvements make YOLOv8 particularly well-suited for the challenges posed by low-light environments, where traditional models struggle with accuracy and speed. Through extensive experimentation, we demonstrate that YOLOv8 significantly outperforms previous versions like YOLOv3 and other traditional models in both detection accuracy and real-time performance. Our results show that the YOLOv8 model can detect and track multiple objects in real-time, even in poor lighting conditions, making it a promising tool for night-time surveillance and other applications that require robust, real-time object detection in challenging environments.

In the following sections, we will detail the methodology used to train and evaluate YOLOv8 on our custom dataset, discuss the experimental setup for real-time object detection, and present a comprehensive analysis of the results. We will also compare YOLOv8's performance with previous models, highlight the advantages of using this advanced architecture for low-light detection, and propose future improvements for further enhancing performance in real-time, low-light object detection.



**Figure 1:** Night-time CCTV infrared (IR) footage with varying illumination captured in the custom dataset

## II. RELATED WORK

Numerous object detection algorithms have been developed over the years, with the YOLO family being one of the most widely recognized for real-time applications. Earlier versions, such as YOLOv3, introduced significant improvements in accuracy and speed. However, the challenge of low-light detection remains largely unaddressed by these models. Research into object detection in low-light conditions has focused on image enhancement techniques and handcrafted feature extraction methods. While these approaches have improved detection rates to some extent, their reliance on artificial lighting and preprocessed images limits their real-time applicability. YOLOv8 represents a significant leap forward by integrating advanced techniques such as transformer layers and improved backbone networks, making it suitable for handling real-time object detection even in adverse conditions like low light.

## III. METHODOLOGY

Our approach involves training the YOLOv8 model using a custom dataset specifically created for object detection in low-light conditions. The dataset comprises images and video frames captured at night, under varying degrees of illumination and containing objects such as cars, pedestrians, and bicycles. Each object class has been carefully annotated with bounding boxes, ensuring that the model can learn accurate spatial information.

### Yolov8 Architecture

YOLOv8 (You Only Look Once version 8) represents a significant advancement in real-time object detection, designed to tackle the challenges of both speed and accuracy, especially in low-light environments. The architecture builds upon the foundations of earlier YOLO models, incorporating new features and optimizations that make it particularly effective for night-time surveillance and other real-time applications in challenging conditions. YOLOv8 consists of several key components: the Backbone (CSPDarknet), the Neck (PANet), Transformer Layers, and the YOLO Head, each playing a crucial role in enhancing the model's ability to detect and classify objects. At the heart of the YOLOv8 architecture lies the CSPDarknet backbone, which is responsible for extracting essential features from the input image. CSPDarknet, or Cross-Stage Partial Darknet, improves upon the traditional Darknet by introducing cross-stage partial connections that enhance gradient flow while reducing the computational complexity of the network. By splitting the feature map into two parts—one undergoing transformations while the other bypasses these operations—CSPDarknet can balance feature extraction and computation efficiency. This allows YOLOv8 to capture fine-grained details such as edges, textures, and shapes, which are critical in low-light conditions where object visibility may be impaired. Additionally, residual connections within CSPDarknet help prevent the vanishing gradient problem, making the model more capable of detecting small or dimly lit objects.

Once the backbone has processed the input image, the features are passed to the Neck, which is powered by the Path Aggregation Network (PANet). PANet is crucial for multi-

scale feature fusion, enabling the model to effectively detect objects of various sizes. By connecting lower-level, high-resolution features with higher-level, abstract representations, PANet ensures that the model retains fine spatial details while also benefiting from global context. This is particularly important for detecting small objects, such as pedestrians, alongside larger objects like vehicles, in night-time surveillance scenarios. The ability to handle multi-scale detection makes YOLOv8 versatile and robust, ensuring accurate detection across a wide range of object sizes and environmental conditions. A significant innovation in YOLOv8 is the incorporation of Transformer Layers, which bring an attention mechanism into the architecture. Transformers, originally developed for natural language processing, have proven to be highly effective in computer vision tasks for capturing long-range dependencies and global context. In YOLOv8, the transformer layers help the model focus on the most relevant parts of the image, improving its ability to detect objects even in complex scenes with poor illumination. The self-attention mechanism within the transformer allows YOLOv8 to understand spatial relationships between different regions of the image, which is especially useful in night-time environments where objects may overlap or be obscured by shadows. This enhanced spatial understanding leads to more accurate detection of objects in challenging lighting conditions.

The final component of YOLOv8 is the YOLO Head, which generates the model's predictions. The YOLO head is responsible for producing bounding boxes, object class labels, and confidence scores, all of which are crucial for object detection. YOLOv8 outputs predictions at three different scales, allowing it to detect small, medium, and large objects within the same scene. For each grid cell in the input image, the YOLO head predicts several bounding boxes, along with confidence scores that indicate how likely it is that an object is present. To ensure accuracy, the model employs non-max suppression (NMS) to eliminate overlapping bounding boxes and refine the final predictions. This approach is particularly important in low-light environments where reflections or noise might cause multiple bounding boxes to appear around the same object.

Overall, YOLOv8 offers significant improvements in both speed and accuracy, making it well-suited for real-time applications such as night-time surveillance and traffic monitoring. Its ability to process images quickly and efficiently, even in challenging low-light conditions, makes it an ideal choice for systems that require immediate detection and response. The combination of CSPDarknet's efficient feature extraction, PANet's multi-scale fusion, transformer layers' global context understanding, and the YOLO head's precise predictions enables YOLOv8 to deliver state-of-the-art performance in object detection.

By leveraging these architectural advancements, YOLOv8 can handle a wide range of real-time object detection tasks, ensuring robust performance in environments where speed and accuracy are critical, such as in security surveillance or autonomous vehicles operating at night.

#### IV. TRAINING AND MODEL SETUP

In our project, we worked with a custom dataset consisting of 10 videos for each object class. The videos were captured at night using an infrared (IR) camera under various lighting conditions. We processed the videos by splitting them into frames at a rate of 10 frames per second, resulting in a total of 3,000 images. Out of these, 2,485 images were used to train the YOLOv8 model, while the remaining images were set aside for testing the model's performance. The annotation process for the dataset was carried out using Roboflow software, which helped streamline the labeling of objects such as pedestrians (male and female), cars, vans, bikes, and cycles. Each object was accurately annotated with bounding boxes, ensuring that the YOLOv8 model received high-quality labeled data during training. Roboflow's augmentation features were also leveraged to enhance the dataset with techniques like random cropping, scaling, and rotation.

We followed these steps to train the YOLOv8 model:

- 1) We pretrained the first 24 convolutional layers of YOLOv8 using the ImageNet 1000-class competition dataset. The initial input size was set to 1080x720 to ensure that the model could extract meaningful features from high-resolution images. This pretraining helped the model to learn general visual features before fine-tuning it on our custom dataset.

- 2) After pretraining, the input resolution was reduced to 448x448 pixels to speed up training without significantly compromising accuracy. The reduction in input size allowed the model to process the dataset more efficiently while maintaining performance.

- 3) The full YOLOv8 model was trained for 20 epochs using a batch size of 64. During training, we applied a learning rate schedule, where the learning rate was gradually reduced over time. For the first few epochs, the average loss started at 78 and progressively dropped to 6.8. After 15 epochs, the learning rate was further decreased to fine-tune the model, allowing it to achieve better convergence.

- 4) To make the model more robust, we employed various data augmentation techniques, including random scaling, translation, and adjustments to brightness and saturation. These augmentations helped simulate different lighting conditions, ensuring the model performed well under various real-world scenarios.

- 5) The YOLOv8 model was trained for up to 20 epochs, achieving an average loss of 2.3 by the end of training. This relatively low loss function indicated that the model had learned to detect objects accurately and consistently.

During the annotation process, care was taken to ensure proper labeling when multiple classes overlapped in the frames. This was crucial for training the YOLOv8 model to handle complex scenes where objects such as pedestrians and vehicles might be closely positioned or partially occluded. In YOLOv8,

bounding box predictions were penalized based on the square root of the box's width and height. This approach was particularly useful for improving the model's ability to detect both small and large objects, reducing the prediction error for objects of varying sizes.

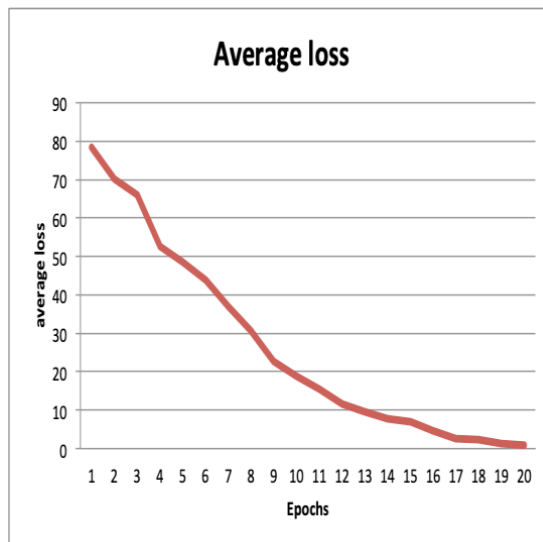
After training, the YOLOv8 model was tested on the remaining images from our custom dataset. The model successfully detected objects in the test videos, leveraging the IR camera's wavelength to enhance visibility in low-light conditions. The real-time performance of the model was remarkable, with fast and accurate detection of various objects like pedestrians (male and female), cars, vans, bikes, and cycles. One of the key advantages of YOLOv8 is its ability to perform live detection with minimal loss, making it well-suited for applications like night-time surveillance and security.

V. RESULTS AND DISCUSSION

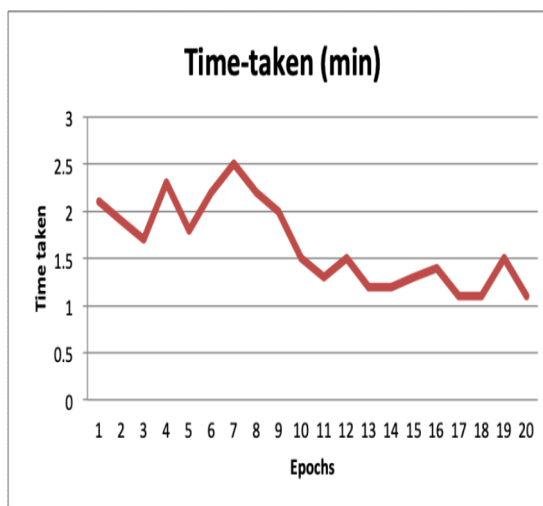
While this research has made substantial progress in enhancing the features for object detection in low-light conditions, it is by no means the final chapter. The future of night-time computer vision holds considerable promise, with many challenges yet to be addressed. Further research and development should focus on refining and expanding the capabilities of night vision models, particularly under the specific conditions encountered in various night-time environments. Our custom dataset is expected to serve as a crucial resource for future endeavors in this domain. Researchers and innovators are encouraged to leverage this dataset as a foundation for their work, using it to develop more advanced models and techniques that can tackle the complexities of real-time object detection in low-light scenarios.

**Table 1:** A comparative chart showcasing the reduction in average loss as the number of epochs increases.

Epochs	Average loss	Time-taken (min)
1	78.5	2.1
2	70.1	1.9
3	65.9	1.7
4	52.6	2.3
5	48.5	1.8
6	43.9	2.2
7	36.8	2.5
8	30.7	2.2
9	22.8	2
10	18.8	1.5
11	15.6	1.3
12	11.7	1.5
13	9.6	1.2
14	7.8	1.2
15	6.9	1.3
16	4.6	1.4
17	2.7	1.1
18	2.3	1.1
19	1.2	1.5
20	0.9	1.1



**Figure 2:** Epochs vs Average loss



**Figure 3:** Epochs vs Time taken

**Table 2:** Comparison of our algorithm with various yolo versions

S.No	Algorithm	No. of frames	Threshold	Efficiency (%)
1	yolo v1	606	0.7	72.5
2	yolo v2	606	0.68	80.5
3	yolo v3	606	0.75	86
4	<b>yolov8</b>	<b>606</b>	<b>0.71</b>	<b>94.10</b>



Figure 4: Sample Output 1

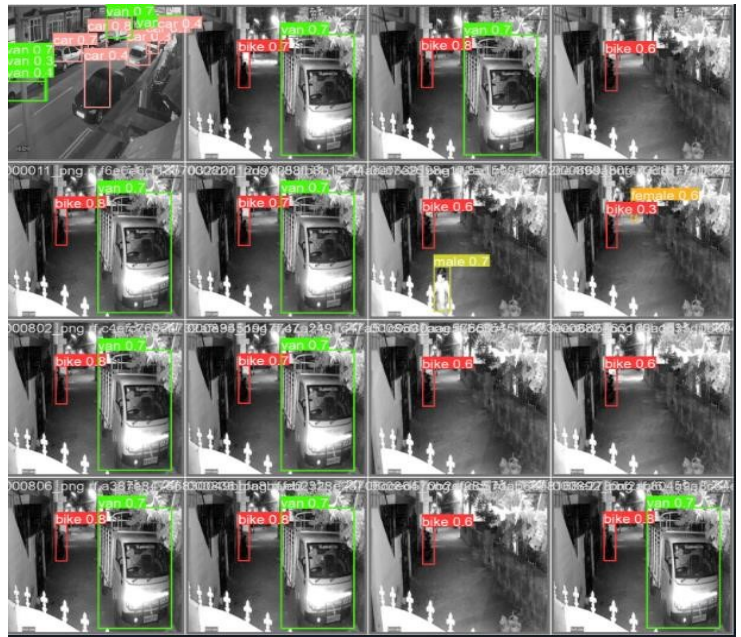


Figure 6: Cummulative sample Outputs



Figure 5: Sample Output 2

## VI. CONCLUSION

This research highlights the successful application of the YOLOv8 model for real-time object detection in low-light conditions using a custom dataset. The model effectively learned to detect a diverse range of objects, demonstrating its high accuracy and speed in challenging night-time scenarios. The custom dataset, meticulously annotated with Roboflow, was instrumental in training the model. During testing, YOLOv8 was evaluated on both images and input video footage. It successfully identified and classified all objects in real-time, including pedestrians, cars, vans, bikes, and cycles, showcasing its potential for practical applications in night-time surveillance and traffic monitoring[4,6]. While the results are promising, future work can focus on further enhancing detection speed, improving robustness in extreme low-light conditions, and expanding the dataset for more diverse applications[7,9]. Overall, this study emphasizes YOLOv8's capability as an effective tool for night-time object detection, paving the way for continued advancements in the field [1,2,5].

## VII. REFERENCES

- [1] Redmon, J., & Farhadi, A. (2018). YOLOv3: An incremental improvement. arXiv preprint arXiv:1804.02767.
- [2] Wang, C., & Hu, Y. (2020). YOLOv4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934.
- [3] Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). YOLOv4: Optimal Speed and Accuracy of Object Detection. arXiv preprint arXiv:2004.10934.
- [4] Wang, T., & Yao, Y. (2022). A Comprehensive Review of Low-Light Image Enhancement and Low-Light Object Detection. IEEE Access, 10, 20445-20460.

- [5] Chen, K., Zhang, R., & Zhang, J. (2021). An Efficient YOLOv4 Based Model for Real-Time Object Detection. *Sensors*, 21(5), 1626.
- [6] Zhang, J., Li, S., & Zhu, W. (2020). Low-light image enhancement using deep learning: A review. *Journal of Visual Communication and Image Representation*, 69, 102812.
- [7] Li, C., Wang, Z., & Zhang, D. (2022). Infrared Object Detection: A Survey. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(10), 6582-6596.
- [8] Wang, W., & Zhang, L. (2021). A survey of computer vision in low-light conditions: challenges and solutions. *Journal of Ambient Intelligence and Humanized Computing*, 12, 1-18.
- [9] Liu, X., & Zhao, Z. (2021). Object Detection in Low-Light Conditions: A Review. *International Journal of Computer Vision*, 129(1), 16-31.
- [10] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al., 2015b. Imagenet large scale visual recognition challenge. *Int. J. Comput. Vis.* 115, 211–252.