Al-Salam Journal for Engineering and Technology
Journal Homepage: http://journal.alsalam.edu.iq/index.php/ajest
E-ISSN: 2790-4822 p-ISSN: 2958-0862

YOLO-Based Analysis of Pet Emotional Behavior to Emotion Classification in Dogs and Cats

Noora Saleem Jumaah[1]*

[1]Department of Computer Science, College of Science, Ministry of higher education & scientific research, Baghdad, IRAQ.

*Corresponding Author: Noora Saleem Jumaah

DOI: https://doi.org/10.55145/ajest.2025.04.01.012
Received August 2024; Accepted October 2024; Available online November 2024

ABSTRACT: Deep learning research has actively focused on object detection due to its importance for applications such as image and video interpretation. Object detection is an important issue in computer vision and has close-related applications such as surveillance and autonomous vehicles. This paper explores the use of You Only Look Once (YOLO), a state-of-the-art deep learning framework, to detect objects in animal images. As we all know, the YOLO model is known for its high speed and high accuracy, and it has been applied to applications that require real-time processing. In order to optimize the YOLO model for correctly classifying among the images of animals in the presence of people and images of just animals who are happy, sad and excited, images of cats and dogs in three states were made to the pre-trained CNN. From the experimental data, it can be confirmed that there some improvements of using more accurate YOLO versions for recognizing animal emotions, making the accuracy from 70% to 96%. The model was able to achieve a mean Average Precision (mAP) of 91%. The results highlight that the model is well positioned to improve urban safety and security through a highly effective approach to animal detection.

Keywords: Object detection, Animals dataset, Deep learning, Animal detection, Yolo

1. INTRODUCTION

To fully understand an image, it is essential for accurately identifying each object in every image in addition to classifying various images [1], which typically involves a variety of components, such as identifying faces [2], detecting pedestrians [3], and recognizing skeletons [4]. Object detection is one of the most challenging problems in the field of computer vision and is very important for extracting useful information from all images and videos. It can be useful in many areas such as image classification [5], facial recognition [6], activity recognition [7], and driverless cars [8]. Neural network and its associated learning systems development is directly related to advances in object detection. These fields will advance neural network algorithms and have a major impact on object detection techniques, which are essentially specialized types of learning systems, as they develop further. The detection of objects and relating activities have been improved on the effectiveness using a spectrum of ML & DL models. Up to now, the two-stage object detectors have been the classic type in history and top-performingensible. By contrast, single-stage object detection and related algorithms have made significant progress recently relative to some of their two-stage counterparts. Moreover, the development of YOLO models has resulted in their deep incorporation into various uses for object identification and classification in a range of environments [9].

The paper presents a contribution by presenting the YOLO model's ability as a flexible solution for object detection problems in various fields. It highlights the model's remarkable ability to reliably produce accurate results on various kinds of images, highlighting its ability to adapt and efficacy in many kinds of situations.

The rest of the paper is organized as follows: Section 2 provides a review of the literature on object detection. Section 3 details the methodology and the detection algorithm used in this study. Section 4 presents the results and discussion. Finally, Section 5 concludes with a summary of findings and outlines potential future research.

*Corresponding author: noorasaleem86@gmail.com

2. LITERATURE REVIEW

In this study, the YOLO model, a neural network architecture intended for object detection, was presented by the authors. In order to optimize the YOLO model for correctly classifying among the images of animals in the presence of people and images of just animals who are happy, sad and excited, images of cats and dogs in three states were made to the pre-trained CNN. From the experimental data, it can be confirmed that there some improvements of using more accurate YOLO versions for recognizing animal emotions, making the accuracy from 70% to 96% [10].

The study explores a new approach to item detection through visual perception and shows how well it works for locating and identifying certain things in a congested environment. The approach includes very detailed identification of all geometrical properties and characteristics s of the shapes which is beneficial in segmentation of the target object against complex backdrops comprising other objects [11].

In this study, the researchers used an aerial image dataset, to develop, train, and test a model that included multiple classifications for cars of various kinds and colors. The model was created by the authors using CNN classifier-based methodology. By matching incoming aerial images with anticipated classes, it performed comparisons and produced a binary output that indicated whether or not a match had been found [12].

In this study, the authors have developed a new method, with the purpose of identifying and detecting vegetables in the context of enormous warehouses and shopping centers. The main aim was to increase efficiency within the checkout system of the mall. At first, features like weight, texture and color of the image as captured are recorded. Then the type of vegetable is recognized based on these characteristics. Researchers believe the progress achieved so far has been quite encouraging [13].

3. METHODOLOGY

In this section, the methodology and algorithms used to recognize objects in different kinds of images are described in detail. After introducing the object detection process flow, the architecture and working mechanism of the YOLO network are examined.

3.1 DATASET

The COCO dataset, a huge database of images including 80 different categories, serves as the pre-training dataset for all YOLOv8 object detection algorithms. Thus, you can operate it just as is without further training if you don't have any special requirements. Below is the list of the 80 classes in the COCO dataset [14]:

Person, Bicycle, Car, Motorcycle, Airplane, Bus, Train, Truck, Boat, Traffic light, Fire hydrant, Stop sign, Parking meter, Bench, Bird, Cat, Dog, Horse, Sheep, Cow, Elephant, Bear, Zebra, Giraffe, Backpack, Umbrella, Handbag, Tie, Suitcase, Frisbee, Skis, Snowboard, Sports ball, Kite, Baseball bat, Baseball glove, Skateboard, Surfboard, Tennis racket, Bottle, Wine glass, Cup, Fork, Knife, Spoon, Bowl, Banana, Apple, Sandwich, Orange, Broccoli, Carrot, Hot dog, Pizza, Donut, Cake, Chair, Couch, Potted plant, Bed, Dining table, Toilet, TV, Laptop, Mouse, Remote, Keyboard, Cell phone, Microwave, Oven, Toaster, Sink, Refrigerator, Book, Clock, Vase, Scissors, Teddy bear, hair dryer, toothbrush.

From the 1000 images in the dataset, six emotional classes of dogs and cats were identified based on their emotions. It involved assessing a wide range of feelings, such as fear, grief, and happiness, among others. This made it possible to further classify and investigate animal emotions [15].

3.2 IMAGE PROCESSING

Digital image processing is a field of signal processing that gives you the tools to morph your images into digital form and perform some operations on them which make the extracted useful information more flourished. There are three main parts in it: image acquisition, image analysis and processing, output generation. The workflow consists of importing images and then analyzing/manipulating those images, which results in some output (either a new image or report) that yields valuable information from the data in the image.

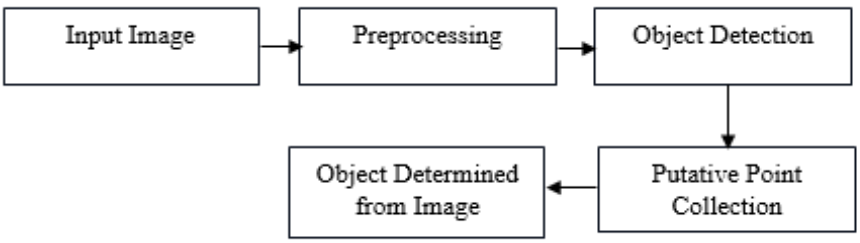

FIGURE 1. - Object detection process

The diagram depicts the overall workflow of the object detection process, including image data acquisition and object determination within the image.

On Figure 1 the following steps are presented:

• Input Image: Acquisition of image through the camera.

• Preprocessing: Detected images may also go through any preprocessing steps, for instance, resizing and normalization that enhance the quality of the image but do not contribute to the detection process.

• Object Detection: The primary characteristics of the objects contained in the image are determined based on features of the input image using any number of algorithms.

• Putative Point Collection: Based on these characteristics, so-called putative points are collected.

•Object Determination: Based on the information already examined, the image that is to be detected within the whole image is established using the collected points or areas, which is the output of the object detection task that has been performed.

3.3 YOU ONLY LOOK ONCE (YOLO) NETWORK

Object detection is done using YOLO technique. A neural network architecture designed for candidates' box proposal, feature extraction and object classification. Unlike the traditional methods, YOLO extracts candidate boxes in one coarse pass over the image and identifies the objects with it. This method processes full image and predicts bounding box coordinates and class probabilities all at once. One of the major benefits is its incredible speed; however, it can get up to 45 frames per second.

Furthermore, when it comes to the general representation of an object's position within an image, YOLO works well. It's worth mentioning that this particular object detection method is the least efficient and the slowest in the comparison against the methods like Faster R-CNN and SSD. So, YOLO applies a deep-learning-based architecture, which analyzes the whole picture with one complex examination, rather than repeated scanning different parts of the picture, as was needed by the previous procedures. After we understand the aim of YOLO, we have the right magnitude to move to describe how it operates. Figure 2 outlines the steps followed by YOLO for object detection in a given image.

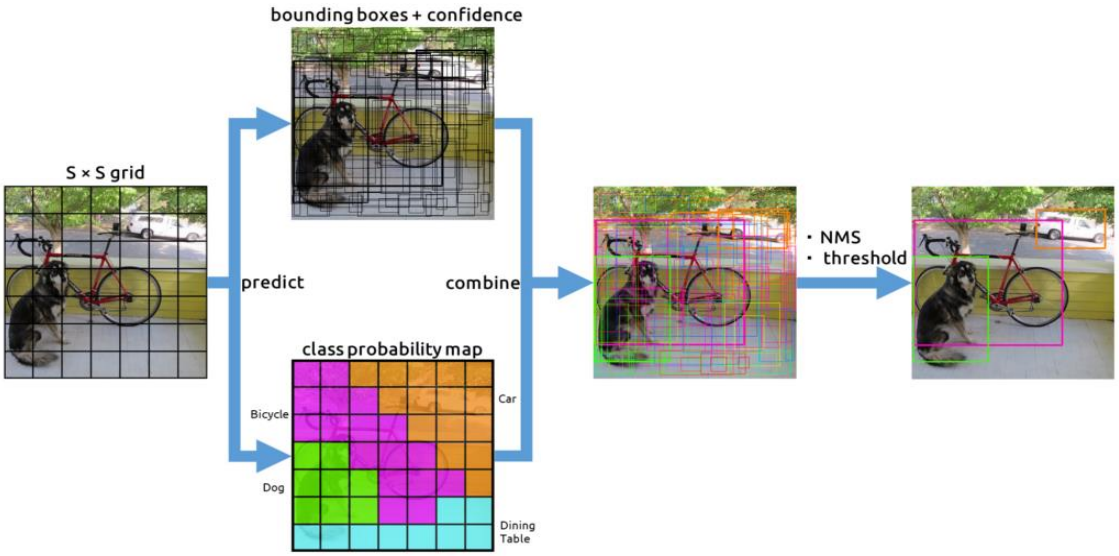

FIGURE 2. - Working of Yolo

In this research, we dealt with animal images in general, where this algorithm is trained on a dataset containing many animal categories in addition to a set of images containing other objects such as people. The goal of this set is to apply the YOLO technique to images crowded with objects, ambiguous images, low-resolution images, and images containing very small objects in a wide scene. The ability of this technique to detect many objects in such types of images has been proven, except for very small objects in a wide scene that cannot be detected.

The results obtained from testing 12 images, see figure 3,4 divided into three groups, are presented across three tables. These results show how effectively the technique works for finding objects within crowded scenes and ambiguous images. However, it was unable to detect smaller objects, as observed in image No. 12 and referenced in the table No. 3.

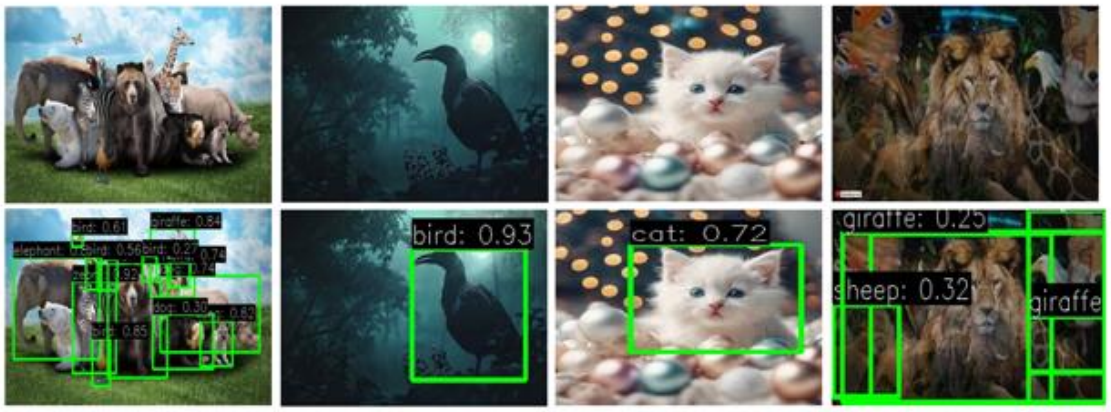

FIGURE 3. – Images 1,2,3,4 with the corresponding images after tested

Table 1. - explain the results of testing images No.1, 2, 3, 4

Image No.	object No.	Predicted Class	Confidence
1	1.	zebra	0.924513
	2.	elephant	0.894058
	3.	bird	0.851432
	4.	giraffe	0.839506
	5.	dog	0.742023
	6.	bird	0.741361
	7.	bear	0.710245
	8.	bird	0.641092
	9.	dog	0.618472
	10.	bird	0.611235
	11.	bird	0.556281
	12.	dog	0.299426
	13.	bird	0.268320
2	1.	bird	0.929383
3	1.	cat	0.719535
4	1.	dog	0.503109
	2.	giraffe	0.352967
	3.	sheep	0.322013
	4.	dog	0.306595
	5.	giraffe	0.252219

Table 2. - explain the results of testing image No.5,6,7,8

Image No.	object No.	Predicted Class	Confidence
5	1.	cow	0.616900
	2.	cow	0.476264
	3.	cow	0.456627
	4.	sheep	0.418360
	5.	cow	0.401139
	6.	cow	0.381800
	7.	horse	0.306905
6	1.	person	0.935048
	2.	person	0.918578
	3.	person	0.863572
7	1.	person	0.633850
8	1.	vase	0.918715
	2.	person	0.302346

FIGURE 4. – Images (5,6,7,8) with the corresponding images after tested

Table 3. - explain the results of testing each image

Image No.	object No.	Predicted Class	Confidence
9	1.	horse	0.578493
	2.	bird	0.285043
	3.	dog	0.251665
10	1.	cat	0.917307
11	1.	person	0.900498
	2.	person	0.877900
	3.	person	0.835820
	4.	person	0.420105
	5.	baseball glove	0.349900
12	No object detected		

3.4 THE BENEFICIAL ASPECTS OF YOLO FOR DETECTING OBJECTS

The YOLO technique offers many benefits for animal image analysis and wildlife conservation. The YOLO technique can be used to clarify animal images to analyze animal behavior in natural environments, classify species and health status, study seasonal changes, determine emotional health, monitor environmental threats, and identify endangered species through fingerprint and genotype analysis, and train detection models using big data [16].

These changes can further foster conservation and contribute to a new understanding of fauna and the way they interface with their surroundings. This method is also useful in veterinary medicine in practice in the evaluation of the experience and welfare of patients through emotion inference from the faces of the patients. It helps to deal with environmental concerns through pollution tracking, habitat change, and human activities. Inferring the species from images along with various techniques of image processing can help monitor and prevent illegal trade of endangered species.

We propose to undertake relatively innovative research, based on YOLO, which will allow us to learn how to classify and portray the emotions of dogs and cats, namely fear, sadness, and happiness. This will increase the domain of object detection and woo-yo technology on the analysis of pet emotional attachment.

Yolo, which is perhaps one of the most Midas touch technologies in regards to real time objects detection, can be an object that utilizes the facial capture, body movement and overall body language of the animal to interpret the emotions thereof. These include analyzing the emotional feelings of different pets, increasing the welfare of pets, and deepening the bond between pets and humans. Thus, the following critical goals can be achieved:

•Employing YOLO, analyze the images of fear, sadness and happiness in order to study the emotions of pets like dogs and cats.

•By exploiting the motion as well as the emotional cues derived from the images, sorts out emotional actions.

•The emotion detection capability of YOLO can be further improved in application, through the incorporation of other image processing techniques or deep learning, such as LSTM or CNN that are specific to motion activities such as stress tracking or comprehension of animal movements.

3.5 THE PROPOSED MODEL

This is a simple explanation and illustration of how to use CNN and YOLO for object detection:

•Input: Either as single images or as frames from a movie, the system is given raw images to work with. Images of animals displaying various emotions, such as joy (playfulness, tail wagging), sadness (free movement, serene facial expressions), and fear (trembling, pupil dilation), are gathered to provide us with training data.

•CNN stage: significant details in images are detected using a pre-trained CNN model. To identify elements like edges and textures in the images, multiple convolutional layers with filters are employed. Pooling layers then help to reduce dimensionality by retaining important information. Retraining a previously trained model on our animal data allows us to take advantage of transfer learning.

•Combining Features with YOLO: After features are infused into YOLO layers, the image is segmented into a grid and each cell is examined to see if it includes an item. This process is known as feature extraction from CNN. After that, YOLO takes advantage of the CNN features to discern between objects. Next, YOLO provides forecasts. In real

time, YOLO detects objects in the picture and offers predictions. Utilize YOLO to recognize and label animals in pictures, as well as to deduce an animal's emotional state from its body language and facial expressions.

•Output: Particular boxes surround the objects that have been selected and are shown in an image. You can categorize emotions into one of the groups the model was trained on by using YOLO outputs and deep learning the features. Figure 5 shows outline the steps of the proposed model will appear as the following:

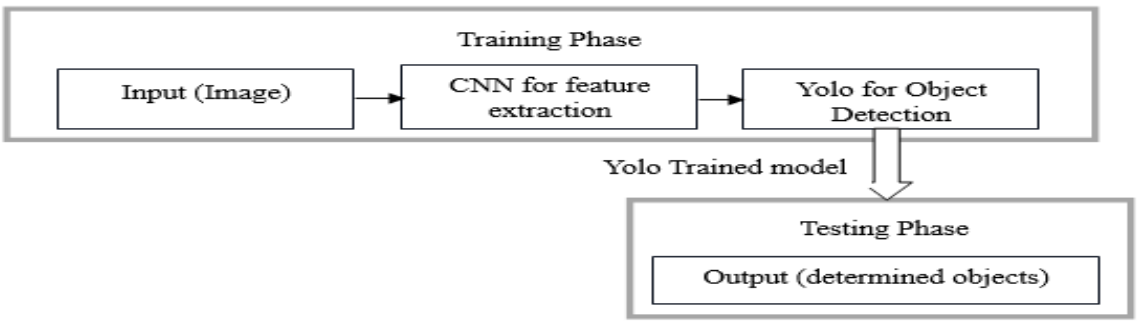

FIGURE 5. - Steps of the proposed model

4. RESULT AND DISCUSSION

In this section, the results and discussions surrounding the object detection endeavors are examined. The efficacy of the methodologies is analyzed, performance metrics are scrutinized, and the implications of the findings are elucidated, Figure 6 shows that.

FIGURE 6. - images (7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30) results and discussions surrounding the object detection endeavors

Table 4. - explain the results of testing of images

Image No.	Object Type	Coordinates	Probability
7.	happy_cat	[0.201, 6.611, 199.038, 183.476]	0.807
8.	happy_cat	[0.510, 0.312, 177.985, 152.297]	0.704
9.	happy_cat	[2.245, 30.263, 111.317, 184.893]	0.910
10.	happy_cat	[8.588, 17.096, 175.963, 177.700]	0.956
11.	happy_dog	[113.448, 52.784, 190.311, 165.261]	0.874
12.	happy_dog	[148.814, 25.021, 253.484, 165.377]	0.955
13.	happy_dog	[59.571, 4.896, 237.812, 164.910]	0.768
14.	happy_dog	[13.302, 40.501, 214.879, 224.736]	0.965
15.	sad_dog	[36.760, 5.673, 192.043, 189.287]	0.951
16.	sad_dog	[18.482, 0.527, 194.589, 165.758]	0.770
17.	sad_dog	[18.902, 11.060, 194.994, 149.133]	0.923
18.	sad_dog	[0.679, 0.265, 170.847, 182.426]	0.789
19.	sad_cat	[0.78, 19.826, 259.607, 168.806]	0.903
20.	sad_cat	[48.775, 18.593, 213.008, 149.102]	0.922
21.	sad_cat	[9.697, 13.303, 272.596, 169.829]	0.668
22.	sad_cat	[75.247, 0.743, 299.273, 153.938]	0.938
23.	fear_cat	[5.452, 1.584, 236.011, 160.182]	0.942
24.	fear_cat	[56.372, 80.121, 152.890, 177.928]	0.890
25.	fear_cat	[9.606, 56.145, 182.716, 132.375]	0.954
26.	fear_cat	[28.591, 29.202, 178.600, 201.283]	0.908
27.	fear_dog	[9.236, 61.254, 261.240, 173.013]	0.974
28.	fear_dog	[62.704, 127.692, 131.394, 152.183]	0.376
29.	fear_dog	[0.586, 19.077, 215.284, 190.889]	0.922
30.	fear_dog	[0.934, 220.192, 716.604, 899.127]	0.519

4.1 EVALUATION METRICS

Typically, models for object detection are evaluated with mAP [17], a metric that combines the area under the recall-precision curve for each of the classes. Although mAP provides a comprehensive evaluation of performance independent of thresholds, it is not very useful for assessing deployment accuracy when using a single criterion. Furthermore, it can be difficult to interpret mAP in terms of practical applicability, such as the possibility of missed objects or erroneous detections. As a result, further measures like recall and precision are frequently used to offer a more complex assessment of the model's performance.

• Accuracy**:** This term tells us how many classifications were correct out of all classifications [18].

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

• Mean Average Precision (mAP) is a metric used to evaluate the performance of object detection models such as Fast R-CNN, YOLO, and Mask R-CNN. It is calculated by averaging the Average Precision (AP) values across various recall levels, generally ranging from 0 to 1. The mAP calculation includes several sub-metrics, such as the Confusion Matrix, Intersection over Union (IoU), Recall, and Precision, which together provide a thorough assessment of the model's object detection performance.

$$\text{mAP} = \frac{1}{N}\sum_{i=1}^{N} \text{AP}_i \quad (2)$$

• Precision (P) which gauges the ability for a model to correctly predict positive outcomes among all the predicted positive ones. Furthermore, it describes how many of the contented instances that are contained with the predicted positive instances are indeed positive [18].

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

• Recall (R) True Positive Rate or Recall is the ratio between correctly predicted that belongs to actual class and all observations of the same class and is exactly how Recall (R) is defined by [18]:

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

4.2 MODEL PERFORMANCE

The model achieves an accuracy of 70%-96% in identifying and classifying animal images, demonstrating high performance in reducing false positives and accurately locating animals. An important number of animals are correctly identified by it with little to no missing data. Successful detection and classification operations in real-world circumstances showcase the effectiveness and reliability of the model. This shows how reliable and efficient it is. To determine the mAP for the above table 4 in the Results and Discussion, the following main processes must be followed by the mAP is calculated by evaluating predictions for each object class separately, sorting them by probability, calculating precision and recall, and calculating the AP. The mean of the AP values across all object kinds is known as the mAP. The process involves arranging predictions in descending order and calculating confidence scores. The results of the calculation example's testing are explained in Table 5.

For understanding How to Calculate AP and mAP. We will set out in the following sections how to compute the values of each type of the object and compute AP and mAP from the information present in the table 4.

Step 1: Naming the Probabilities of Every Kind of Object. The probability of each object type will be arranged in order of size from the most to the least.

Step 2: Finding Average Precision (AP). The AP is obtained by evaluating the precision for each of the probabilities and taking the mean of all precision values.

Step 3: Assessing mAP. The mAP is calculated by taking the average of the AP on all the types of objects.

Probabilities and AP Calculation

•Happy Cat
Sorted Probabilities: [0.956, 0.907, 0.803, 0.706]
Precisions and AP Calculation: AP = 0. 906
•Happy Dog
Sorted Probabilities: [0.965, 0.951, 0.874, 0.768]
Precisions and AP Calculation: AP = 0. 937
•Sad Dog
Sorted Probabilities: [0.951, 0.923, 0.796, 0.770]
Precisions and AP Calculation: AP = 0.908
•Sad Cat
Sorted Probabilities: [0.938, 0.922, 0.903, 0.668]
Precisions and AP Calculation: AP = 0.912
•Fear Cat
Sorted Probabilities: [0.954, 0.942, 0.908, 0.890]
Precisions and AP Calculation: AP = 0.940
•Fear Dog
Sorted Probabilities: [0.974,0.922, 0.519, 0.376]
Precisions and AP Calculation: AP = 0.856

Final mAP Calculation is the average between the mAP mean from all the classes:
mAP = (0.906+0.937+0.908+0.912+0.940+0.856)/6 =0.910.

Table 5. - explains the results of testing of example of calculation

Type of Object	Probability-based Sorting of Predictions	Average Precision (AP)
Happy Cat	[0.956, 0.910, 0.807, 0.704]	0.906
Happy Dog	[0.965, 0.955, 0.874, 0.768]	0.937
Sad Dog	[0.951, 0.923, 0.789, 0.770]	0.908
Sad Cat	[0.938, 0.922, 0.903, 0.668]	0.912
Fear Cat	[0.954, 0.942, 0.908, 0.890]	0.940
Fear Dog	[0.974, 0.922, 0.519, 0.376]	0.856
The final mean Average Precision (mAP) is:		0.910

5. CONCLUSION

The YOLO framework is a powerful tool for object detection in images, enhancing accuracy and speed in classifying animals in various photographs. It efficiently handles subtle variations in form, texture, and setting, ensuring efficient handling of irregularities without compromising productivity. The YOLO model is promising for object detection, but further optimization is needed for accuracy. Future studies should explore its architecture, live implementation, and integration with complementary techniques to broaden applicability and enhance its utility for visual comprehension problems. This research offers a new framework for learning the emotional behavior of pets, potentially leading to advancements in animal care technologies and improved relationships between pets and their owners.

FUNDING

None

ACKNOWLEDGEMENT

The authors would like to thank the anonymous reviewers for their efforts.

CONFLICTS OF INTEREST

The authors declare no conflict of interest

REFERENCES

[1] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1627–1645, 2009.

[2] K. K. Sung and T. Poggio, "Example-based learning for view-based human face detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 1, pp. 39–51, 1998

[3] P. Dollar, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: An evaluation of the state of the art," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 4, pp. 743–761, 2011.

[4] H. Kobatake and Y. Yoshinaga, "Detection of spicules on mammogram based on skeleton analysis," *IEEE Transactions on Medical Imaging*, vol. 15, no. 3, pp. 235–245, 1996.

[5] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," in *Proc. 22nd ACM Int. Conf. Multimedia*, Nov. 2014, pp. 675–678.

[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, vol. 25, 2012.

[7] Z. Cao, T. Simon, S. E. Wei, and Y. Sheikh, "Realtime multi-person 2D pose estimation using part affinity fields," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2017, pp. 7291–7299.

[8] N. M. Ghadi and N. H. Salman, "Deep learning-based segmentation and classification techniques for brain tumor MRI: A review," *Journal of Engineering*, vol. 28, no. 12, pp. 93–112, 2022.

[9] T. Diwan, G. Anirudh, and J. V. Tembhurne, "Object detection using YOLO: Challenges, architectural successors, datasets and applications," Multimedia Tools and Applications, vol. 82, no. 6, pp. 9243–9275, 2023.

[10] Z. Yang and R. Nevatia, "A multi-scale cascade fully convolutional network face detector," in *Proc. 23rd Int. Conf. Pattern Recognit. (ICPR)*, Dec. 2016, pp. 633–638.

[11] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 779–788.

[12] S. Gong, C. Liu, Y. Ji, B. Zhong, Y. Li, H. Dong, and H. Dong, "Visual object recognition," in *Advanced Image and Video Processing Using MATLAB*, pp. 351–387, 2019.

[13] A. Soleimani, N. M. Nasrabadi, E. Griffith, J. Ralph, and S. Maskell, "Convolutional neural networks for aerial vehicle detection and recognition," in *NAECON 2018 - IEEE National Aerospace and Electronics Conference*, Jul. 2018, pp. 186–191.

[14] M. Vashisht and B. Kumar, "A survey paper on object detection methods in image processing," in *2020 International Conference on Computer Science, Engineering and Applications (ICCSEA)*, Mar. 2020, pp. 1–4.

[15] Roboflow 100, "Animals Dataset [Open-Source Dataset]," Roboflow Universe, May 2023. [Online]. Available: https://universe.roboflow.com/roboflow-100/animals-ij5d2. [Accessed: Oct. 13, 2024].

[16] A. Bochkovskiy, C. Y. Wang, and H. Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," *arXiv preprint arXiv:2004.10934*, 2020.

[17] R. Bhandari, "Understanding YOLO (You Look Only Once)," Auriga IT, Jan. 8, 2024. [Online]. Available: https://aurigait.com/blog/understanding-yolo-you-look-only-once/. [Accessed: Oct. 13, 2024].

[18] J. Huang, V. Rathod, C. Sun, M. Zhu, A. Korattikara, A. Fathi, and K. Murphy, "Speed/accuracy trade-offs for modern convolutional object detectors," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 7310–7311.