

DeepCat: A Deep Learning Approach to Understand Your Cat's Body Language

Arwa Fawzy¹, Ahmed Hassan², Ahmed Gad², Ahmed B. Zaky^{1,4}, and Haitham El-Hussieny³

Abstract—Pet owners experience difficulty in understanding their pets' body language and its implications for animal welfare. Animals cannot use human speech to communicate their emotions and health conditions. However, previous experiments analyzing cat behavior indicate an overlooked level of expressiveness in cats. DeepCat is a deep-learning approach that translates cats' body language signals, enabling owners to realize their feline's intended messages and emotional states. Our DeepCat was trained on a dataset comprising 10,000 cat images, where automatic labeling was used to track key features, including the tail, eyes, and mouth. Presented as a flutter application, DeepCat could be implemented to function everywhere. It will become easy to monitor the cat and give advice to make decisions in situations that require caution. In this paper, we will discuss the potential benefits and limitations of our DeepCat and provide suggestions for future research on this topic.

Index Terms—YOLOv8, Automatic Labeling, Emotion Recognition, Bounding Box, Voting Classifier.

I. INTRODUCTION

Several cat owners have been curious about understanding the emotions of their feline companions [1]. While animals cannot communicate with humans using verbal language, they do express their feelings through non-verbal cues, including body language and facial expressions [2]. Understanding a cat's body language is crucial for cat owners to establish a strong bond and communicate effectively with them [3]. By deciphering the subtle non-verbal communication cues provided by their ears, eyes, tail, whiskers, and mouth, cat owners can gain valuable insights into their cat's current emotions and attitudes. For example, a cat's ears can indicate more than just listening, and its eyes are highly expressive, revealing a wide range of emotions. Wide-open eyes may indicate fear, interest, excitement, or aggression, while half-closed or drooping eyes convey relaxation and friendliness [4].

Beyond advancing animal welfare, the analysis of cat responses and emotions contribute to economic growth [5].

Surveys and statistics indicate that 70% of employed pet owners claim they would be happier and more productive if they could have their pet at work, either on-site or remotely [6]. The most common ways pets help people de-stress include snuggling (68%), inducing laughter (67%), and alleviating loneliness (61%) [7]. Each year, 2.7 million animals are adopted from shelters, including 1.3 million cats [8]. Regrettably, 2.7 million animals also perish annually, with cats accounting for 1.4 million of these losses [8]. The integration of DeepCat into households could play a pivotal role in detecting and rescuing cats in distress. Depression can also cause cats to overgroom, leading to hair loss, skin irritations, or even bald patches [9]. This research primarily focuses on feline responses due to limited biological investigations concerning other animals. Conversely, it signifies a turning point in understanding emotions across all animals. DeepCat holds paramount significance for pet owners, veterinarians, and researchers alike. Gaining insights into cats' responses aids pet owners and veterinarians in identifying potential health concerns or behavioral irregularities [10].

While a multitude of models exist for analyzing human responses, there are not many studies for detecting cat responses through Multi-Parameter Analysis and Posture Recognition. This endeavor is inherently complex, given the unpredictable nature of feline responses. Whether in life or death, their reactions often exhibit the same expression. Chen, Rung-Ching, et al. (2021) [10] conducted a comprehensive study aimed at identifying and classifying various cats' behaviors using a carefully curated dataset. This dataset consisted of 1,410 images capturing walking, eating, sleeping, sitting, searching trash cans, and using the toilet. The research harnessed the capabilities of the Raspberry Pi 4 Model B v1.2 and the Raspberry Pi Camera Module v2.1. For the object detection task, the YOLOv3-tiny model (an abbreviation for YOLO version 3, a state-of-the-art object detection model) was selected as the primary choice. The evaluation phase of the study was particularly insightful and achieved a mean Average Precision (*mAPs*) ranging from 95% to 98%.

These behaviors are readily discernible to cat owners during their daily interactions with their feline friends. For instance, owners can deduce that the cat is engaged in eating when it responds to the presence of a food container, infer that the cat is in a state of slumber when it closes its eyes, and identify habitual searching behavior when it exhibits recurrent responses to trash. Our innovation lies in our approach, where we emphasize the analysis of feline emotions rather than feline behaviors. Determining the cat's emotional state, such as happiness, sadness, anger, or signs of illness, presents

¹Arwa Fawzy and Ahmed B. Zaky are with the Computer Science and Information Technology (CSIT) program, Egypt-Japan University of Science and Technology, Alexandria, Egypt. arwa.harhash@ejust.edu.eg, ahmed.zaky@ejust.edu.eg

²Ahmed Hassan and Ahmed Gad are with the Department of Computer Science Engineering (CSE), Egypt-Japan University of Science and Technology, Alexandria, Egypt. ahmed.hasan@ejust.edu.eg, ahmed.gad@ejust.edu.eg

³Haitham El-Hussieny is with the Department of Mechatronics and Robotics Engineering, Egypt-Japan University of Science and Technology, Alexandria, Egypt. haitham.elhussieny@ejust.edu.eg

⁴Ahmed B. Zaky is on leave Electrical Engineering Department, Faculty of Engineering at Shoubra, Benha University, Egypt. ahmed.zaky@feng.bu.edu.eg

a complex challenge, yet it holds significant importance in promoting the cat's well-being. Implementing YOLOv8 for the purpose of detecting cats' labels was inspired by [10]. Our primary objective is to enhance the precision of response identification by leveraging the multifaceted postures displayed by cats. DeepCat approach commences with the utilization of a dedicated architecture, implementing YOLOv8 for superior accuracy of cats' label localization: mouth, tail, and eyes. Significant analytical additions were built, considered as our contributions, which are outlined below:

- 1) We utilized Otsu thresholding, which discriminates between different degrees of eye aperture by segregating the white area (foreground) from the dark area (background) based on pixel intensities [11]. This process spans a range from wide openness to narrow openings and closure of the eye.
- 2) We implemented a multi-parametric analysis using the minimax optimization algorithm to discern emotional states with the tail movement to evaluate all possible moves from the current state by recursively simulating the movements until a terminal state is reached.
- 3) We created two new labels for mouth states (open and half-closed), which were not included in the YOLOv8 algorithm. We manually built these labels using the Labellmg tool, and we extracted the coordinates of their locations into a dataset in eXtensible Markup Language (XML) format.
- 4) We yielded two distinct modes of detection: real-time analysis through video streams and offline scans using images.
- 5) We amalgamated YOLOv8 and the voting classifier to ensure a pursuit of heightened accuracy in delineation the cat's emotion.

The former contributions focus on analyzing the emotional state of cats in order to understand their intentions and feelings. Cats are highly expressive if an approach knows how to interpret their body language. Every aspect of their communication, from the way their tail twitches to the movement of their mouths, conveys meaning.

II. PREPROCESSING ARCHITECTURE

In this section, we present the sequence of preprocessing steps through which the input image undergoes as shown in Figure 1. The reasoning behind the selection of YOLOv8 is the implemented enhancements across different versions, spanning from YOLOv3 to YOLOv8. These advancements include adaptations in anchor boxes, and scaling of input resolution [12]. YOLOv8 excels in various visual tasks like object detection, segmentation, tracking, and classification, meeting all necessary requirements of DeepCat model. Primarily, the image is scanned by multiple sliding windows with different sizes. Sub-images are extracted and annotated with (WH) of sliding window width and height and (x,y) of sliding window position [13]. System is trained to implement regression tasks independently and classify each sub-image, extracting multiple classifications in the anchor box probabilities [14]: P_{mouth} , P_{tail} , P_{eyes} .

YOLOv8 has a segmentation attached to each bounding box, and it is localized accurately on the feature borders. This facilitated the exclusion of white space within the eye's enclosure. To clarify, our objective is to segment and focus solely on the white part of the eye, as depicted in points 4 to 6 in the cat of Figure 1 and segmented with green color in Figure 2, while excluding the outer unnecessary regions in the box encompassing the cat's face because it is not utilized for measuring the extent of eye openness. While traditional object detection algorithms produce numerous bounding boxes on the eyes, Non-Maximum Suppression (NMS) filters out redundant and less accurate boxes, keeping only the most relevant ones.

As detailed in Figure 1, YOLOv8 utilizes a backbone similar to YOLOv5, with the addition of the CSPLayer, also known as the cross-stage partial bottleneck, featuring two convolutional layers (C2f module). The C2f module integrates high-level

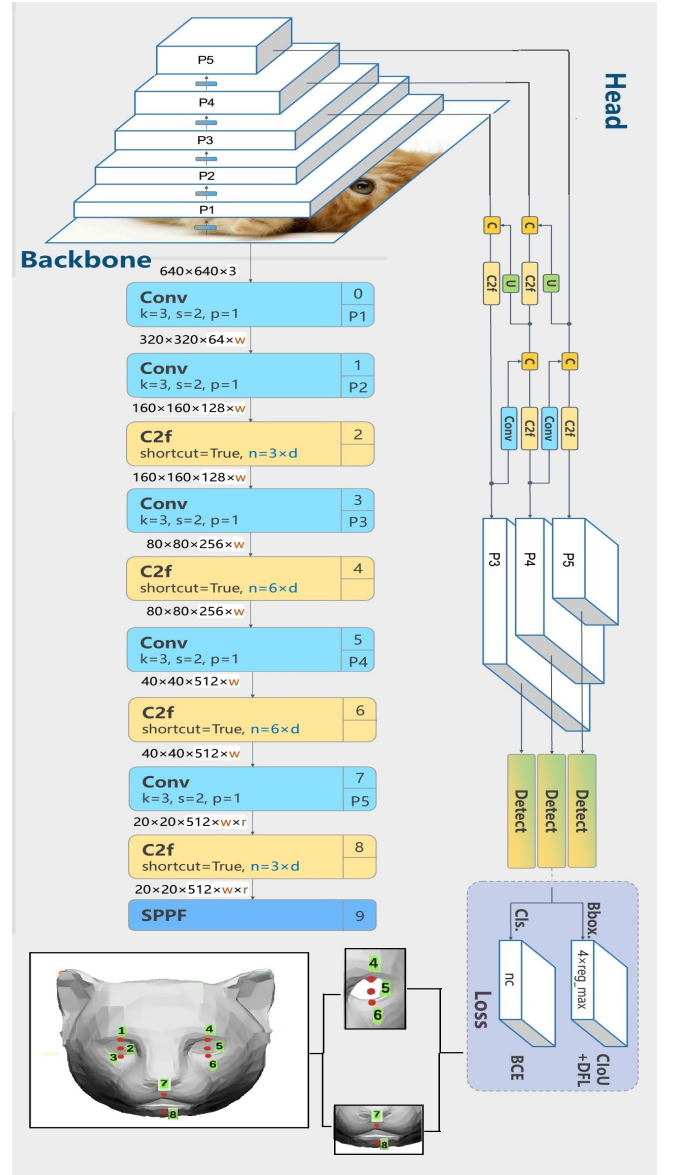


Fig. 1: Preprocessing procedures before implementing the DeepCat algorithms

features with contextual information, enhancing accuracy. In the output layer, a sigmoid activation function is used for the objectness score, indicating the likelihood of detecting the cat's labels. The class probabilities are computed using the softmax function. Finally, YOLOv8 incorporates CIOU and DFL loss functions for bounding box loss and binary cross-entropy for classification loss, resulting in improved cat's features detection performance, particularly for the eyes of cats. Moreover, YOLOv8 introduces a semantic segmentation model known as the YOLOv8-Seg model. The backbone of this model is a CSPDarknet53 feature extractor, followed by a C2f module instead of the conventional YOLO neck architecture. The C2f module succeeded by two segmentation heads responsible for predicting semantic segmentation masks for input images. The YOLOv8-Seg model shares a similar detection head structure with YOLOv8, comprising five detection modules and a prediction layer. Notably, the YOLOv8-Seg model has achieved state-of-the-art results across various features detection and semantic segmentation benchmarks while maintaining exceptional speed and efficiency [12]. The architecture was trained using a dataset comprising 10,000 cat images collected from the Kaggle platform [15]. These images have the three features across various poses, angles, and conditions.

III. DEEPCAT ALGORITHMS

Once the cat's features were recognized and the labels were located, we proceeded to develop distinct algorithms for each feature. This allowed us to monitor the timing of actions and distinguish between various states and body language signals.

In the context of mouth-related indicators, the cat's body language reveals two distinct expressions: an open mouth and a half-closed mouth. When the cat gaps its mouth with a complete opening, it typically signifies displeasure or anger in response to an unpleasant odor. Conversely, when the cat chatters its teeth, it indicates an engagement with captivating stimuli or a heightened state of excitement. The final condition observed is a totally closed mouth, which does not convey any specific indication and is represented as "NaN" [16]. The two states were detected by creating two new labels through the manual creation of bounding boxes to the mouth as either open or half closed through the Labellmg tool. The tool generated XML files containing the class identification number and coordinates of each label to determine their precise locations, resulting in two labels for mouth states. Next, these labels were provided for the training phase of the YOLOv8 architecture.

For the tail algorithm, we initiated the implementation by converting the image from RGB to grayscale. Subsequently, a crucial step involved the adaptive mean thresholding to segregate the tail from the background and the entire cat's body. Furthermore, we used applied min-max optimization to locate the maximum points from the left, right, top, and bottom, effectively delineating the tail's start and endpoints. The resulting land markings simplified the identification of the tail's base and tip, allowing us to construct a right-angled triangle using three of the corresponding coordinates, as shown in Figure 2. To calculate the angle, we inversed

the Pythagorean theorem in Equation 1 where $m1$ and $m2$ represent two gradients, and θ denotes the angle between them [17]. This algorithm determined whether the tail was inclined or not because any deviation from 90 degrees indicated an inclined tail. The tail's position and movement explain the cat's intentions. The cat that is content and engaged with its surroundings typically holds its tail high. In contrast, the tail is bent into a U shape when it suggests a desire to play with other animals or humans. Contrarily, the cat that is frightened or upset often fluffs up its tail and moves the tip side to side in a rhythmic pattern [4].

$$\tan \theta = \left| \frac{m_2 - m_1}{1 + m_2 m_1} \right| \quad (1)$$

Given the coordinates of the cat's eyes obtained from a YOLOv8, our DeepCat model crops the region of interest (ROI) corresponding to the eyes from the image. The cropped eye image is then converted to grayscale and subjected to a pixel-wise division operation and Gaussian blur. Subsequently, we implemented Otsu thresholding to segment the eye region from the rest of the image. It further utilizes morphological operations for refining the eye's boundary. The algorithm calculates the ratio of white pixels, representing the eye area, to the total pixels in the segmented labeled region, effectively quantifying the extent of eye openness, as illustrated in Figure 2. This algorithm detected the subtle changes in a cat's eye, indicative of its emotional state. The wide-open eyes in a cat may signal fear, interest, or excitement, while partially closed eyes often suggest relaxation or drowsiness, and a narrow eye-opening could indicate aggression [4].

In the context of video analysis, frames were sequentially appended to a list to monitor the cat's evolving emotional state. To address the challenge of prioritizing emotional cues that change less rapidly, such as those conveyed through the tail, our algorithm was integrated to track and emphasize the most current and pertinent emotional signals within the video. Infrequently, a potential conflict may arise, for example,

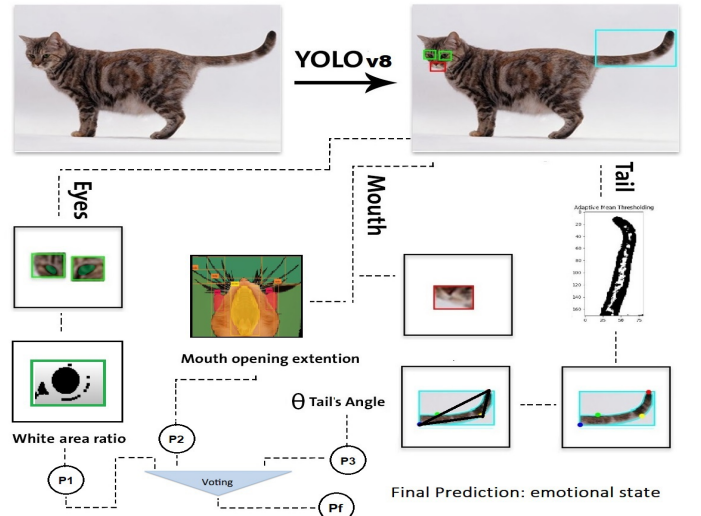


Fig. 2: Overview of the proposed DeepCat algorithms with the three selected features.

when DeepCat simultaneously detects aggression through a cat's tail and happiness through its eyes. To resolve this, we have implemented a voting classifier that takes into account the outputs of diverse feature-based probability models. This classifier intelligently aggregates the inputs and ultimately endorses the most recommended probability outcome, ensuring a balanced and informed decision-making process as illustrated in Figure 2. We integrated the model into a Flutter application interface using Flask, Werkzeug for secure filename handling, and custom image and video processing modules. The Flask application includes an endpoint that, upon receiving a POST request with an image file, securely handles the filename, saves it to a specified directory, performs image detection, and returns the results in JSON format. The system is configured to run on port 1234 as the main module.

IV. RESULTS AND DISCUSSION

The mAP is a metric that measures the model's performance in object detection by considering the precision-recall trade-off at various confidence thresholds. The metric deals with multiple class detection by computing the AP individually for each category and subsequently calculating the mean of these AP values. As demonstrated in Table I, the accuracy associated with eye-related illustrations achieved the highest performance. This can be due to the greater prevalence of images depicting eyes in comparison to those depicting tails, which, in turn, outnumbered the images featuring open and half-open mouth configurations.

Feature Name	Precision	Recall	Accuracy	F1-Score
Tail	0.98	0.74	0.85	0.84
Half-Mouth	0.89	0.76	0.82	0.82
Open-Mouth	0.86	0.85	0.84	0.85
Eye	0.94	0.96	0.97	0.95

TABLE I: Performance metrics of features

Through an iterative process that covers all precisions and recalls, Equation 2 determines the difference between the present and subsequent recalls, which is then multiplied by the current precision. In essence, the AP can be understood

as the cumulative precision values at distinct thresholds, with each value weighted by the corresponding increase in recall [18].

$$\sum_{AP}^{k=n-1} [\text{Recalls}(k) - \text{Recalls}(k+1)] \times \text{Precisions}(k) \quad (2)$$

Where:

$\text{Recalls}(n) = 0, \text{Precisions}(n) = 1, n = \text{Number of thresholds}$

In order to compute the mAP , the process begins with the calculation of the AP for individual classes [18]. The mAP is then calculated as the average of these class-specific AP values using Equation 3.

$$mAP = \frac{1}{n} \sum_{k=1}^n AP_k \quad (3)$$

Where:

$AP_k = \text{the } AP \text{ of class } k, n = \text{number of classes}$

In the training, we observed fluctuations as the DeepCat model explored the loss landscape and adjusted its parameters in the quest for the optimal solution (see Figure 4). We achieved lower loss of training box and classification, and validation box classification avoiding overfitting. Furthermore, we attained lower training and validation delta from ground truth loss, indicating improved alignment with ground truth annotations and enhanced model precision.

Figure I demonstrated a higher precision with lower rate of false positives and a higher recall values with a higher proportion of real positives. The model achieved a high F1-Score, which is a metric that combines precision and recall into a single value, providing a balance between the two as shown in Table I. Figure 4) shows the evaluated mAP at intersection over union (IoU) 50% and mAP from IoU 50% to 95%. There are high mAP values at both IoU thresholds. The confusion matrix evaluated five labels: half closed and open mouth, tail, background label, and eyes. Our model demonstrated impressive performance, achieving 97% accuracy for eyes, 82% to 84% for mouth labels, and 85% for tail. Notably, the background label was excluded from the matrix labels in Figure I, and the sum of values does not equate to a total of 1, reflecting minor errors in detecting the background label, similarly to other labels.

Following our contributions, we concluded the following findings:

- 1) Emotional State recognition: DeepCat facilitated its highly intricate, unlike behaviors such as eating and sleeping, which can be easily inferred by humans.
- 2) Otsu Thresholding: the model could distinguish various degrees of eye aperture.
- 3) Minimax Optimization Algorithm: its multi-parametric analysis could discern emotional states based on tail movements.
- 4) Creation of New Labels: DeepCat introduced two new labels for mouth states (open and half-closed), which were not included in the YOLOv8.

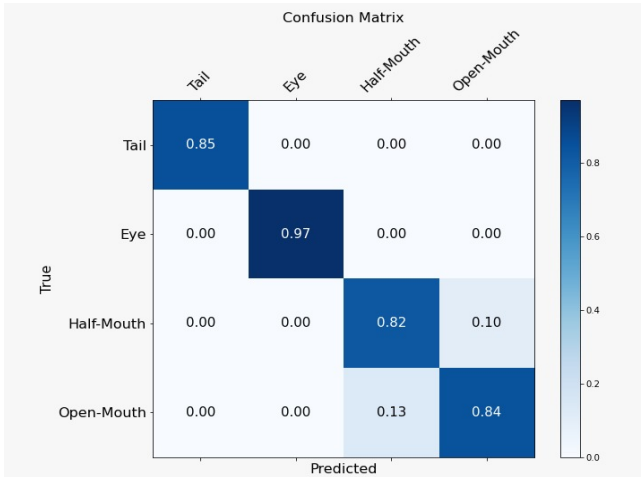


Fig. 3: The resultant confusion matrix analysis.

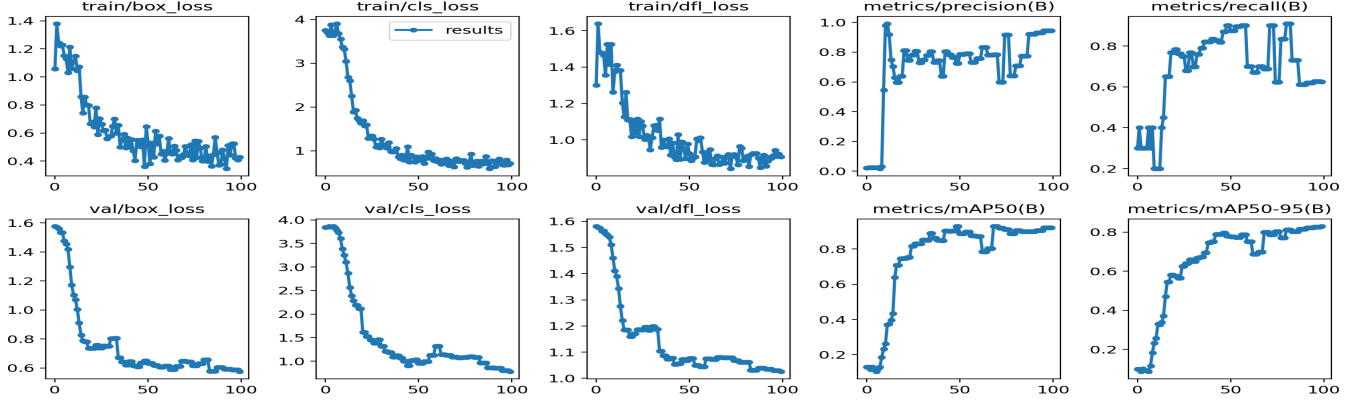


Fig. 4: Training and validation performance analysis of the proposed DeepCat model.

5) Integration of YOLOv8 for with voting classifier: this pursued heightened accuracy in delineating the cat's emotion.

6) Creation of mobile application: DeepCat could function everywhere for real time cat emotional analysis.

In comparison, the results of [10] focuses on different aspects:

- 1) Behavior Monitoring System: YOLOv3 was implemented and could analyze various behaviors of cats at home.
- 2) Recognizing 6 behaviors of cat: the cat was recognized while walking, eating, sleeping, sitting, searching trash cans, and using the toilet.
- 3) Real-Time Monitoring System: a system was proposed to observe the cat's behaviors.

At present, the cat's intentional message is depicted, creating the impression of the cat engaging in a form of dialogue with its owner to express its feelings. Consequently, if any experts in feline emotional state were to take a handbook or medical reference and commence a comparative analysis of each feature's condition, it would align with the outcomes generated by the DeepCat application.

V. CONCLUSION

DeepCat introduces a pioneering avenue for cat owners to effectively communicate with their furry friends. Developing YOLOv8 alongside some algorithms, DeepCat detected the cat's emotional state. It ensures a high accuracy across various positions and lighting conditions. Future iterations may expand its capabilities by incorporating additional features. For instance, cat ears, serving as constant radar systems, constantly scan their surroundings and then reflect emotional states. Another shape of the cat's eye, which is dilated pupils, indicates heightened excitement. DeepCat also has the potential to include whisker indications to further enhance its emotional state recognition. By then, the mobile application will offer worldwide accessibility and include alerts for such cues, promoting harmonious cat-human interactions.

VI. AVAILABILITY

The accessibility of the mobile application associated with the present work is provided through the repository which is

reported in [19].

REFERENCES

- [1] M. Siniscalchi, S. Laddago, and A. Quaranta, "Auditory lateralization of conspecific and heterospecific vocalizations in cats," *Laterality: Asymmetries of Body, Brain and Cognition*, vol. 21, no. 3, pp. 215–227, 2016.
- [2] A. Quaranta, S. d'Ingeo, R. Amoroso, and M. Siniscalchi, "Emotion recognition in cats," *Animals*, vol. 10, no. 7, p. 1107, 2020.
- [3] J. Hulick, "How to speak cat: A guide to decoding cat language by aline alexander newman," *Bulletin of the Center for Children's Books*, vol. 68, no. 8, pp. 414–414, 2015.
- [4] S. Dale, *Decoding Your Cat: The Ultimate Experts Explain Common Cat Behaviors and Reveal how to Prevent Or Change Unwanted Ones*. Houghton Mifflin, 2020.
- [5] D. C. Turner, "The mechanics of social interactions between cats and their owners," *Frontiers in Veterinary Science*, vol. 8, p. 650143, 2021.
- [6] L. R. Kogan, C. Bussolari, J. Currin-McCulloch, W. Packman, and P. Erdman, "Disenfranchised guilt—pet owners' burden," *Animals*, vol. 12, no. 13, p. 1690, 2022.
- [7] B. Fougere, *Pet Lover's Guide to Natural Healing for Cats and Dogs*. Elsevier Health Sciences, 2005.
- [8] L. Mullen, "Rehabilitating canines in animal shelters to reduce euthanasia rates," 2015.
- [9] D. Teich, "Depression in cats: Signs, causes, and how to help," oct 2023.
- [10] R.-C. Chen, V. S. Saravanarajan, H.-T. Hung, *et al.*, "Monitoring the behaviours of pet cat based on yolo model and raspberry pi," *International Journal of Applied Science and Engineering*, vol. 18, no. 5, pp. 1–12, 2021.
- [11] X. Xu, S. Xu, L. Jin, and E. Song, "Characteristic analysis of otsu threshold and its applications," *Pattern recognition letters*, vol. 32, no. 7, pp. 956–961, 2011.
- [12] J. Terven and D. Cordova-Esparza, "A comprehensive review of yolo: From yolov1 and beyond. arxiv," *arXiv preprint arXiv:2304.00501*, 2023.
- [13] W.-Y. Hsu and W.-Y. Lin, "Ratio-and-scale-aware yolo for pedestrian detection," *IEEE transactions on image processing*, vol. 30, pp. 934–947, 2020.
- [14] 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.
- [15] N. Crawford, "Cat dataset kaggle," 2008.
- [16] V. Bennett, N. Gourkow, and D. S. Mills, "Facial correlates of emotional behaviour in the domestic cat (*felis catus*)," *Behavioural processes*, vol. 141, pp. 342–350, 2017.
- [17] C. Frohman, "The full pythagorean theorem," *arXiv preprint arXiv:1001.0201*, 2010.
- [18] C. Caragea and V. Honavar, *Machine Learning in Computational Biology*, pp. 1663–1667. Boston, MA: Springer US, 2009.
- [19] A. Fawzy, "Cat emotional analysis." <https://github.com/Arwa-Fawzy/Cat-Emotional-Analysis>, January 2023.