ISSN : 2350-0743

*International Journal of Recent Advances in Multidisciplinary Research*

*Vol. 11, Issue 06, pp.10036-10044, June, 2024*

**RESEARCH ARTICLE**

# A NEW DATASET OF DOG BREED IMAGES AND A BENCHMARK FOR FINE GRAINED CLASSIFICATION

***Mohanad A. Al-Askari**

Department Of Information Systems College of Computer Sciences and Information Technology Al-Sidiyah, Baghdad, Iraq

**ARTICLE INFO**

*Article History*
Received 23rd March, 2024
Received in revised form
15th April, 2024
Accepted 27th May, 2024
Published online 21st June, 2024

*Keywords:*
YOLO5, GetHub, Dataset, Training, Classification.

**Corresponding author:*
*Mohanad A. Al-Askari*

**ABSTRACT**

Triage The report describes the YOLO5 algorithm that underpins the approach in question and can be classified as machine learning for object detection and classification conducted to identify different dog breeds. The project entailed curating a huge number of images of dogs, creating a bounding box for each of them similarly to the previous project, and marking the breed of the dog in the image, training the YOLO5 model, releasing the model and using it to achieve real-time detection of the breed of a dog. The report covers the issues encountered during the process, including dealing with class imbalance, the management of computational resources, and data quality assurance. It stresses the significance of appropriate data pre-processing, annotation, and evaluation measures. It also provides and analyzes the results of training the YOLO5 model such as confusion matrix, F1 curve, labels distribution, labels correlogram, precision-recall curves, batch images. The report shows the real-world problem using the YOLO5 algorithm for object detection and classification for pets' identification, animal monitoring, and behavior analysis. It also presents some of the improvement strategies and the future research areas for improvement. On the whole, the given report may be deemed helpful for the comprehension of the problem of applying advanced object detection algorithms, with the focus on data quality, model evaluation, and incremental approach.



***Citation: Mohanad A. Al-Askari. 2024.*** *"A new dataset of dog breed images and a benchmark for fine grained classification", International Journal of Recent Advances in Multidisciplinary Research,* 11, (06), 10036-10044.

## INTRODUCTION

This report will be about my personal experience working with the YOLO5 algorithm to detect and classify different breed types of dogs. The project will start with obtaining a big dataset of dog images, labeling images in a train set and valid ones based on training and testing model parameters. Next, we will be training the YOLO5 model, and test it on a computer camera for dog breed detection and classification in real time.

**Understanding the YOLO5 Algorithm:** The YOLO (You Only Look Once) algorithm has been the biggest breakthrough in real time object detection that has seen subsequent updates including YOLO5 which have additional advancements to improve performance. YOLO5 works by segmenting the input picture into a grid of cells, restricting each one to assigning coordinates and the output class probability to detected objects. YOLO5 has applied the single neural network to accomplish both object detection and classification jobs at the same time, unlike the conventional methods that have to undergo a series of processing stages. Subsequently, efficiency that can not be beat has been achieved. The first part of the algorithm consists in dividing the image into a grid, usually expected to contain a specific number of cells. Each cell, in the meantime, should be responsible for a given number of bounding boxes predictions and confidence scores, each representing the possible presence of an object in each box (Anwar *et al*. 2020). Moreover, the individual probability of class is assigned to every bounding box, reflecting the degree of that object as a member of a particular class. YOLO5 deploys CNN as its underlying base to convert the input image into features, which may be of relevance to object detection. These functions constitute the inputs that drive successive layers of the network that assigns classifieds box bounds and object probabilities (Zou *et al*. 2020). Nonetheless, YOLO5 has an innovative architecture in which the integration of various improvements for speed and accuracy is used to facilitate applications in real-time, across all fields. YOLO5 incorporates transfer learning, extending the effectiveness of the model to its full potential as it uses pre-trained weights from the base system and develops them to achieve specific tasks.

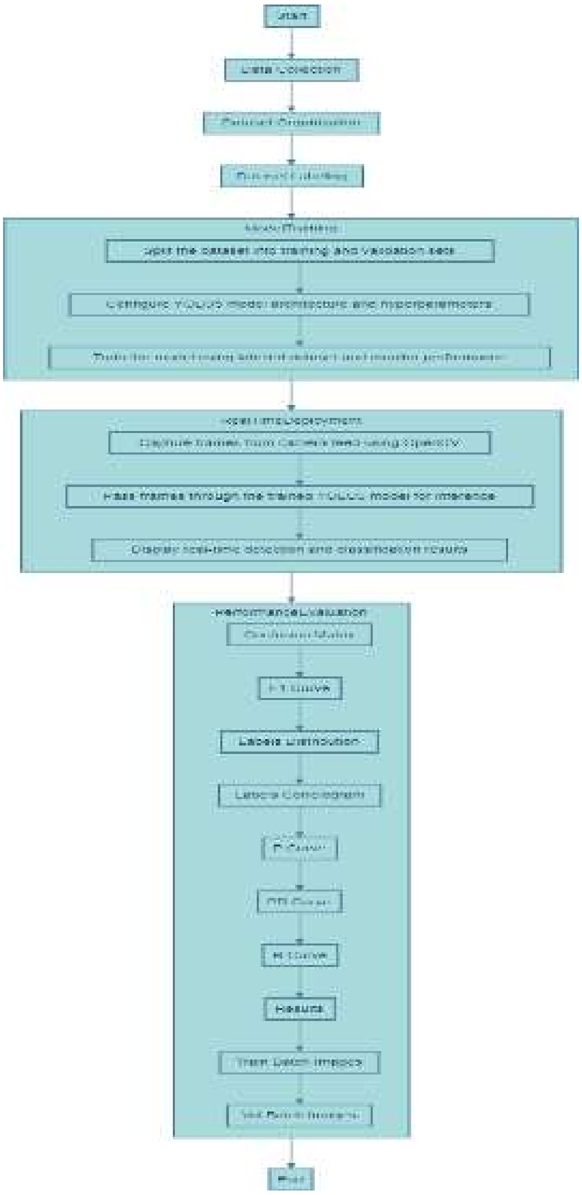
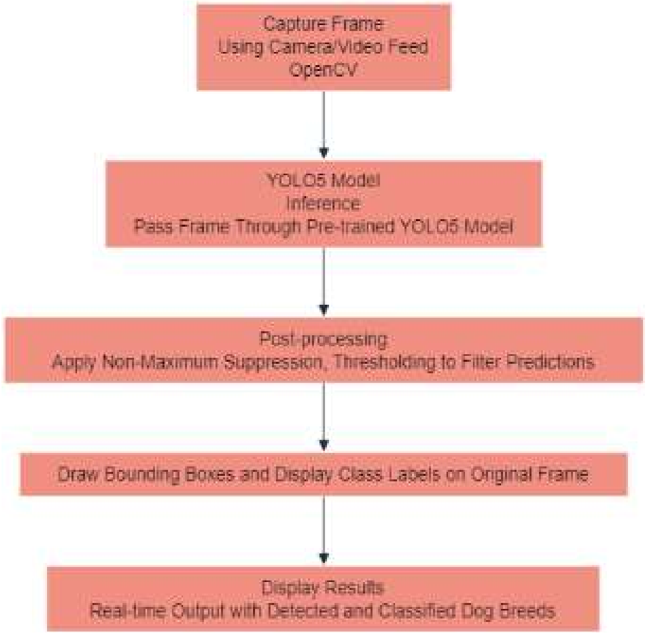


**Figure 1. Flowchart**

Given that well-interpreted, speedy convergence of the model is possible during the training process, the model reaches higher accuracy levels utilizing less annotated data. As a result, YOLO5 is an update to the original anchor box clustering which improves the default bounding box sizes appearance based on the aiming dataset in that category. This outcome leads to a more accurate position of objects and false positives are almost negligible. The YOLO5 algorithm represents a dramatic change in the object detection paradigm by achieving real time accuracy (Wu *et al.* 2021). Utilizing a unified architecture model on top of which transfer learning is applied and anchor box optimization is integrated allows the YOLO5 to set a new level of supporting informed and fast-solving object detection in several scenarios.

**Preparing the Dataset:** Dataset preparation for YOLO5 involves sourcing a new organization and annotating the dog images. To start, at least a sizable number of dog images are consistently scoured from previously known sources, making sure we count all breeds, environments, and poses to emphasize diversity (Pratama*et al.* 2022).



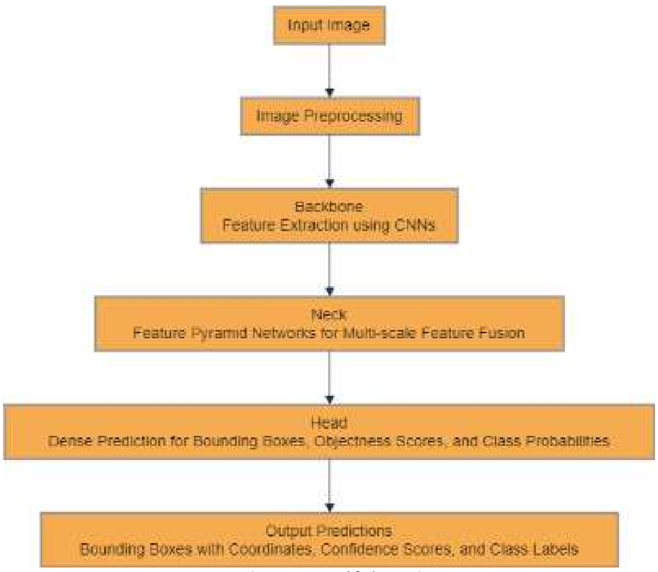

(Source: Self-drawn)

**Figure 2. Real-Time Deployment**

After that, the images are put into a proper format, and the dataset is usually subdivided into separate folders according to the dog breeds. Afterwards comes annotation. Each image was stretched with boundaries of bounding boxes in order to hold the dogs and attribute them class labels as well as class names describing their breeds for hike classification. Concurrently, endeavors are done to pass class biases by using data augmentation or oversampling, which are tools to brace out the dataset. This means that each breed will be well represented in the training data, so that no bias came out during the model training (Jain *et al.* 2020). Primarily as well as limited dataset preparations could make or break any ferment of training the YOLO5 model leading to the correct and unquestionable categorization of dog breeds.

**Labeling the Dataset:** Annotating the dataset for YOLO5 implies marking each image with bounding boxes for most dogs and determining the class labels with the list of their breeds. This is the most important part of training the model for the appropriate task of breed contactable and categorization (Kim *et al*. 2021). The labeling is done using the GetHab annotation tool. Annotators painstakingly traced around dogs in each and every image and ensured that the bounding box included the entire dog(s) without any unnecessary background. Classes are then assigned forms to each bounding box indicating the species of the dog presented in that region. The key to this is to guarantee that the labels are consistent and accurate, thus the need for clear labeling guidelines and quality controls. The annotators may go for training so as to ensure a uniformity in their annotations followed by the resolution of any conflicting or vague labeling situation. Regular review and validation of annotated imaging will be done to identify and rectify error of inconsistency. This, in turn, makes it possible to create a validated labeled set, which finally leads to better model performance during training and inference (Cao *et al*. 2022). The dataset should be carefully labeled because this is the basis for a successful training process and the right object detection and classification.

**Training the YOLO5 Model:** The process of training the YOLO5 model gets done by adjusting its parameters and updating its architecture to precisely detect and classify dog breeding according to the annotated dataset. The evaluation of the model involves a range of performance indicators and graphic representations that give us understanding on when to make changes based on the model performance.

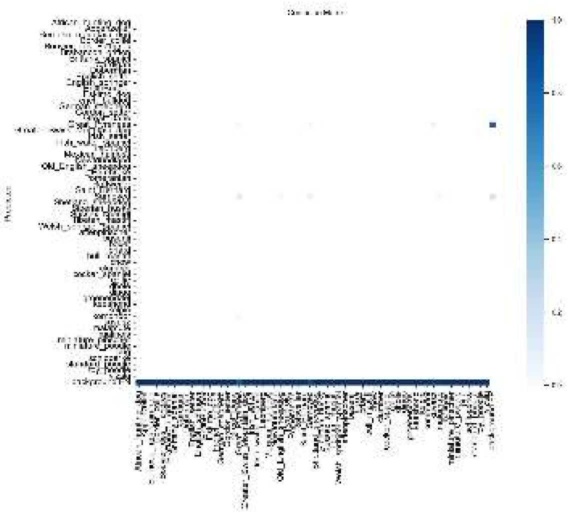



(Source: Self-drawn)

**Figure 2. YOLO5 Model Architecture**

## RESULTS

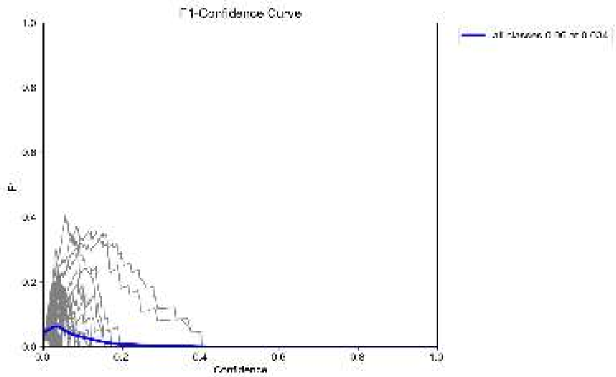
A detailed image of the model classifies performance by showing the number of the dogs classified correctly and wrongly for each breed of a dog. It helps to point out those kinds that are usually over classified and make the model's accuracy perfect wherein adjustments are done afterward if necessary.



(Source: Self-obtained)

**Figure 1. Confusion Matrix**

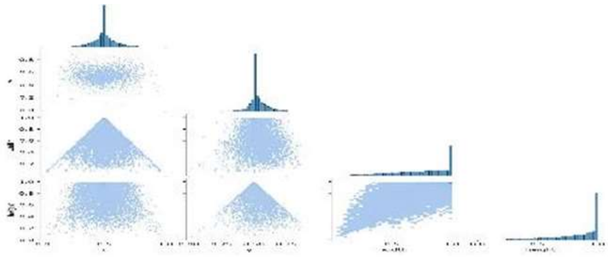
The F1 curve depicts the performance of the model under precision and recall, which uses harmonic mean of both metrics, across different confidence levels.




(Source: Self-obtained)

**Figure 2. F1 Curve**

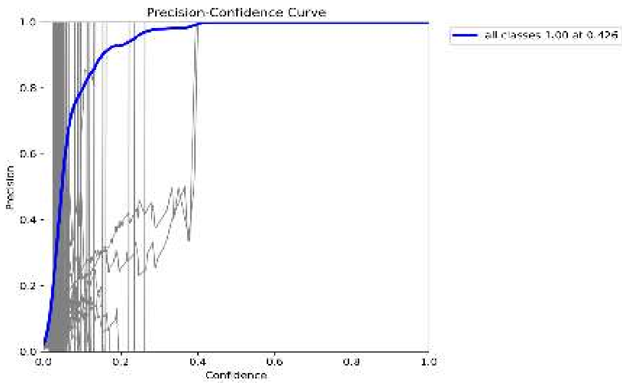
It gives an insight into the decision-making process that involves a precision versus recall trade-off and helps in implementing an effective classifier threshold.



(Source: Self-obtained)

**Figure 4. Labels Correlogram**

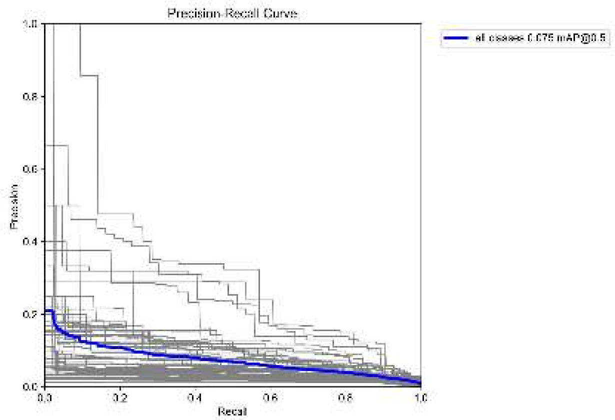
For determining whether class imbalance is present, the class labels are plotted. Balancing classes deals with building a dataset that would reflect a real world population distribution, eradicating bias inferred during classification. This synthetic shows the breeds labels and encourages the breed's character traits or dissimilarity. It plays the role of guiding model training in various ways by reconciling the dissimilarities in the attributes of some breeds.




(Source: Self-obtained)

**Figure 5. P Curve**

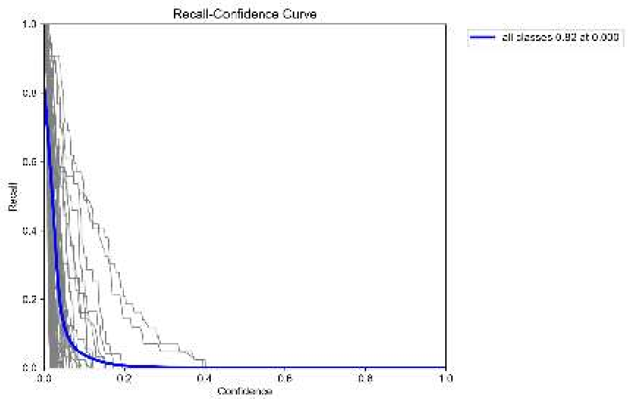
The curve named precision curve plots the model's precision against different confidence thresholds. It also helps in selecting the best tradeoff between precision and recall, which ensures that values of performance are optimized.




(Source: Self-obtained)

**Figure 6. PR Curve**

The Precision-Recall graph is used to gauge the model's performance, which mainly relies on class distribution cases. It shows the empirical view of the overall model's performance which is more inclusive than single threshold interaction, thus helps in the evaluation and selection of the models.




(Source: Self-obtained)

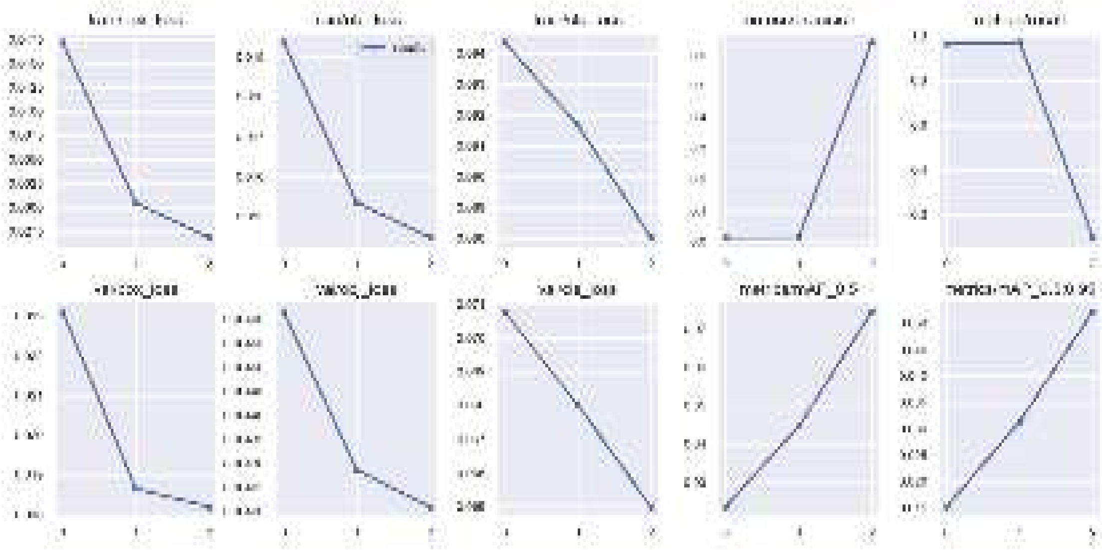
**Figure 7. R Curve**

The Recall curve indicates how recall of the model rises with the chosen threshold of confidence. In the same way, it evaluates the model's capability to get a hold of every possible occurrence within a given class, hence it allows the choice of the threshold to be satisfied. The directory of folders called results contains all the outputs of the model regarding the validation set, including bounding boxes, class, and confidence labels. It enables qualitative testing of the model's capabilities, unearthing possible strong points and weak points. The train batch0 identifies different dogs using different numbers. The train batch1 also identifies different dogs using different numbers. Visualization of training progress using model input images, ground truth, and the predicted outputs at different reminder epochs. They give hints about how the model gets to learn and develop further gradually. The validation batch0 pred shows the images of the dogs without labels. The validation batch0 labels show the labels displaying basenji breed. One dog is of a Vizsla breed.
The validation batch1 pred shows a different set of dogs. The validation batch1 labels identify the dogs of the pug breed. One dog is of the basenji breed. The validation batch2 pred shows a different set of dogs.

The validation batch2 labels identify the dogs as of the Siberian husky breed. Two of the dogs are of affenpinscher breed. The yolo-5 model's performance can be enhanced with the use of these metrics and visualizations such that the developers can iteratively so that they can then tune the model's parameters and architecture and achieve the most accurate and reliable model in detection and classification of dog breeds. The visuals and performance metrics may be analysed to get valuable insights concerning the strengths and weaknesses of the devised YOLO5 model for categorizing dog breeds. Together with these realisations, one can choose the best options for model improvement, which may mean changing hyper parameters, trying different networks, or introducing methods such as data augmentation or domain adaptation. Moreover, it is possible to identify any kind of possible bias or limitations in the database itself, by using the results and the graphics provided. This information can then be used for reference when trying to conduct further attempts at data collecting and curation. It is possible to train the model with a specific number of iterations and then assess its effectiveness on a test sample, and if necessary, increase the number of iterations in subsequent cycles of the model, which will increase its usefulness in such situations in terms of categorizing dog breeds and other possible cases of use.

**Real-time Object Detection and Classification:** Real-time object detection and classification with the YOLO5 model defined is the deployment of the trained model to analyze videos and camera guards at the present time, where it detects and classifies objects, in this case, dogs' breeds, as they appear. Getting libraries like OpenCV into Python and doing their detection and classification, the deployment process has few important steps which could be completed making the detection and classification fast and accurate (Zhuang *et al*. 2020). The first stage of RTDE (Real-Time Data Exchange) is to frame the sequences obtained from the camera feed using OpenCV.

This part of the process involves using the camera device to capture those pictures and giving it commands at an interval that supports the camera hardware, such as how fast it should take the pictures, the usual unit is frames per second (fps). First Frames are taken using the camera and are fed to the YOLO5 trained model for inference. Such operation is concerned with passing the forward pass through the network of model's neurons where at each iteration the frames are processed to recognize as well as give information about the objects that are presented in the frame. Excluding the identification of areas of interest in the camera that may be dogs and surrounding bounding boxes plus the classes and their corresponding. confidence scores which are used during the inference phase by the YOLO5 model (Luo *et al*. 2020). There is an overlay of the predictions of the model on the original frame and this helps to visually show the objects detected and the classifications to which they belong. Tagging objects in real time and recognizing them in frames demands from processing to be speedy so as to ensure the sustained throughput, and the response time as well. YOLO5's lightweight structure together with the streamlined inference pipe makes it possible for real-time operation to be achieved on devices with limited resources as well (Borwarnginn*et al*. 2021).



(Source: Self-obtained)

**Figure 8. Results**

(Source: Self-obtained)

**Figure 9. Train Batch0**

(Source: Self-obtained)

**Figure 10. Train Batch1**

(Source: Self-obtained)

**Figure 11. Train Batch2**

(Source: Self-obtained)

**Figure 12. Validation Batch0 Pred**

(Source: Self-obtained)

**Figure 13. Validation Batch0 Labels**

(Source: Self-obtained)

**Figure 14. Validation Batch1 Pred**

(Source: Self-obtained)

**Figure 15. Validation Batch1 Labels**

(Source: Self-obtained)

**Figure 16. Validation Batch2 Pred**

(Source: Self-obtained)

**Figure 17. Validation Batch2 Labels**

To meet the goal of improved user experience along with the system functionality of the real-time detection system, it is possible to add the functionality to real-time display of classification results, set a confidence threshold for filtering detections and visualizing bounding boxes. These traits give users moment-to-moment updates on the model's efficiency and if there is any need of changing parameters or settings, they become capable of using them. There exist other processing techniques like Non-Maxima suppression which work to clean up the predicted outputs and remove those that give unnecessary or overlapping bounding boxes. This leads to accurate and concise results. Eventually, YOLO5 can be considered an extraordinary and multi solution tool for dealing with numerous tasks such as security monitoring, robotic controlling, and even reality enriched (Valarmathi *et al*. 2023). While using the fast element of YOLO5, people can work to develop a systematic yet interactive structure that can label the objects in a real-time environment with speed and accuracy.

**Challenges and Lessons Learned:** This experience has given me explicit ideas and experiences I cannot trade in. The main drawback I faced throughout was the class imbalance problem in my dataset (Khan *et al*. 2020). Some dog breeds could be over-determined and lead to unbiased predictions while the others could be under-represented which in turn could impair the overall performance. In response to this problem, I delved into oversampling and data augmentation procedures which helped in getting the required number of minute cases for training. The other challenge was how computational resources go for the training of deep learning models like YOLO5. Besides, feeding them with big data can be tiring, in particular, while working with these models (Behera *et al*. 2021). By means of cloud platforms like Google Collab that provide both access to GPU systems and their use for training and consequent experimentation I was able to work effectively. Eventually, I realized the significance of proper data preprocessing and high-quality tagging while working on it. The quality of the training data and the precision of the labels almost determine the overall effect that the model will have (Darvish *et al*. 2022).

Engaging into the process of rigorously collecting and sourcing dataset with sufficient tagging and labeling demonstrated that the investment of time and effort towards that goal led to higher accuracy rate and model reliability. More to this, I developed skill in interpreting and understanding different views in graphs and visualizations as well (Varshney *et al*. 2021). Such instruments demonstrated data visualization on competencies of the model and allowed me to apply correct decisions and adjustments while the model was being trained.

## CONCLUSION

My work on the YOLO5 algorithm for the purpose of recognizing dogs on the basis of their breeds was a great learning experience and engaging activity. From data set collection, labeling, and training to live deployment, I improved my knowledge and skills on different aspects of computer vision deep learning. The project illustrated the necessity of having a wide range of the dataset that is well labeled and the access to the interpretation and analysis of these performance metrics and visualizations. The live object detection and classification showed a hands-on example application of the YOLO5 algorithm and its opportunity for solving genuine world tasks.

As next step, I want to go beyond by adding improvements and enhancements like inference of transfer learning or experimenting with different model network types and possibly useful data augmentation strategies. In the end, it has helped me to add to my knowledge of object detection and classification algorithms and has been a really practical way to improve my practical skills in developing and deploying applications involving computer vision.

## REFERENCES

Anwar, S., Barnes, N. and Petersson, L., 2020. A systematic evaluation: fine-grained CNN vs. traditional CNN classifiers. arXiv preprint arXiv:2003.11154.

Behera, A., Wharton, Z., Hewage, P.R. and Bera, A., 2021, May. Context-aware attentional pooling (cap) for fine-grained visual classification. In Proceedings of the AAAI conference on artificial intelligence (Vol. 35, No. 2, pp. 929-937).

Borwarnginn, P., Kusakunniran, W., Karnjanapreechakorn, S. and Thongkanchorn, K., 2021. Knowing your dog breed: identifying a dog breed with deep learning. International Journal of Automation and Computing, 18, pp.45-54.

Cao, S., Wang, W., Zhang, J., Zheng, M. and Li, Q., 2022. A few-shot fine-grained image classification method leveraging global and local structures. International Journal of Machine Learning and Cybernetics, 13(8), pp.2273-2281.

Darvish, M., Pouramini, M. and Bahador, H., 2022, February. Towards fine-grained image classification with generative adversarial networks and facial landmark detection. In 2022 International Conference on Machine Vision and Image Processing (MVIP) (pp. 1-6). IEEE.

Jain, R., Singh, A., Jain, R. and Kumar, P., 2020. Dog Breed Classification Using Transfer Learning. In Proceedings of the Third International Conference on Computational Intelligence and Informatics: ICCII 2018 (pp. 579-590). Springer Singapore.

Khan, U.A., Din, S.M.U., Lashari, S.A., Saare, M.A. and Ilyas, M., 2020. Cowbree: A novel dataset for fine-grained visual categorization. Bulletin of Electrical Engineering and Informatics, 9(5), pp.1882-1889.

Kim, J., Go, J. and Kwon, C., 2021. Comparison of Fine Grained Classification of Pet Images Using Image Processing and CNN. Journal of Broadcast Engineering, 26(2), pp.175-183.

Luo, W., Zhang, H., Li, J. and Wei, X.S., 2020. Learning semantically enhanced feature for fine-grained image classification. IEEE Signal Processing Letters, 27, pp.1545-1549.

Pratama, N.H., Rachmawati, E. and Kosala, G., 2022. Classification of dog breeds from sporting groups using convolutional neural network. JIPI (JurnalIlmiahPenelitian dan PembelajaranInformatika), 7(4), pp.1080-1087.

Valarmathi, B., Gupta, N.S., Prakash, G., Reddy, R.H., Saravanan, S. and Shanmugasundaram, P., 2023. Hybrid deep learning algorithms for dog breed identification–a comparative analysis. IEEE Access.

Varshney, A., Katiyar, A., Singh, A.K. and Chauhan, S.S., 2021, June. Dog breed classification using deep learning. In 2021 International Conference on Intelligent Technologies (CONIT) (pp. 1-5). IEEE.

Wu, Y., Zhang, B., Yu, G., Zhang, W., Wang, B., Chen, T. and Fan, J., 2021, October. Object-aware long-short-range spatial alignment for few-shot fine-grained image classification. In Proceedings of the 29th ACM International Conference on Multimedia (pp. 107-115).

Zhuang, P., Wang, Y. and Qiao, Y., 2020, April. Learning attentive pairwise interaction for fine-grained classification. In Proceedings of the AAAI conference on artificial intelligence (Vol. 34, No. 07, pp. 13130-13137).

Zou, D.N., Zhang, S.H., Mu, T.J. and Zhang, M., 2020. A new dataset of dog breed images and a benchmark for finegrained classification. Computational Visual Media, 6, pp.477-487.

*******