un

Enhancing Object detection using Genetic Algorithms (Applied on YOLOv5s Model)

Computer Science, BML Munjal University, Gurgaon, 122413, Haryana, India.

Contributing authors

Prateek Singh

(200355)

Prateek.singh.20cse@bmu.edu.in

**Abstract:**

This research presents a transfer learning approach that utilizes a genetic algorithm to address the challenges associated with conventional transfer learning in deep learning-based object detection models. Unlike conventional methods that require manual trial-and-error selection of re-learning layers, the proposed method automatically selects these layers through the genetic algorithm in the transfer learning process. The proposed method was applied to the Yolov5s pre-trained on the COCO dataset. To evaluate the performance of the proposed method, experiments were conducted comparing fine-tuning and the proposed method. The mean average precision (mAP) of the models trained by the conventional and proposed methods were compared using the training and validation data from the GWHD dataset. The results indicate that the proposed method outperformed the conventional method, with a higher average mAP of approximately 0.603 compared to 0.594 for the conventional method. Additionally, the standard deviation of the results obtained from the proposed method was smaller than that of the conventional method, demonstrating the superior stability of the proposed approach.

**Keywords:** Convolutional neural network, deep learning, genetic algorithm, COCO dataset, transfer learning.

# Introduction

The field of computer vision has seen substantial progress in object detection, with YOLOv5s emerging as a powerful model due to its speed and accuracy. Nevertheless, optimizing object detection with YOLOv5s remains a complex challenge, especially when annotated data is limited. This study explores enhancing object detection performance by integrating custom-designed Genetic Algorithms (GAs) with the YOLOv5s model. The development of object detection has progressed from traditional methods like R-CNN to advanced models like YOLO (You Only Look Once). YOLOv5s, an extension of the YOLO architecture, offers high accuracy in real-time object detection, making it a promising model for further refinement.

The research gap lies in addressing the limitations of YOLOv5s, particularly in scenarios where annotated data is scarce or expensive to acquire. The integration of custom GAs introduces a novel approach by generating annotations, aiding in the augmentation of training datasets for improved model performance. This method not only enhances detection accuracy but also reduces the reliance on extensive labelled datasets, a significant constraint in many real-world applications.

1

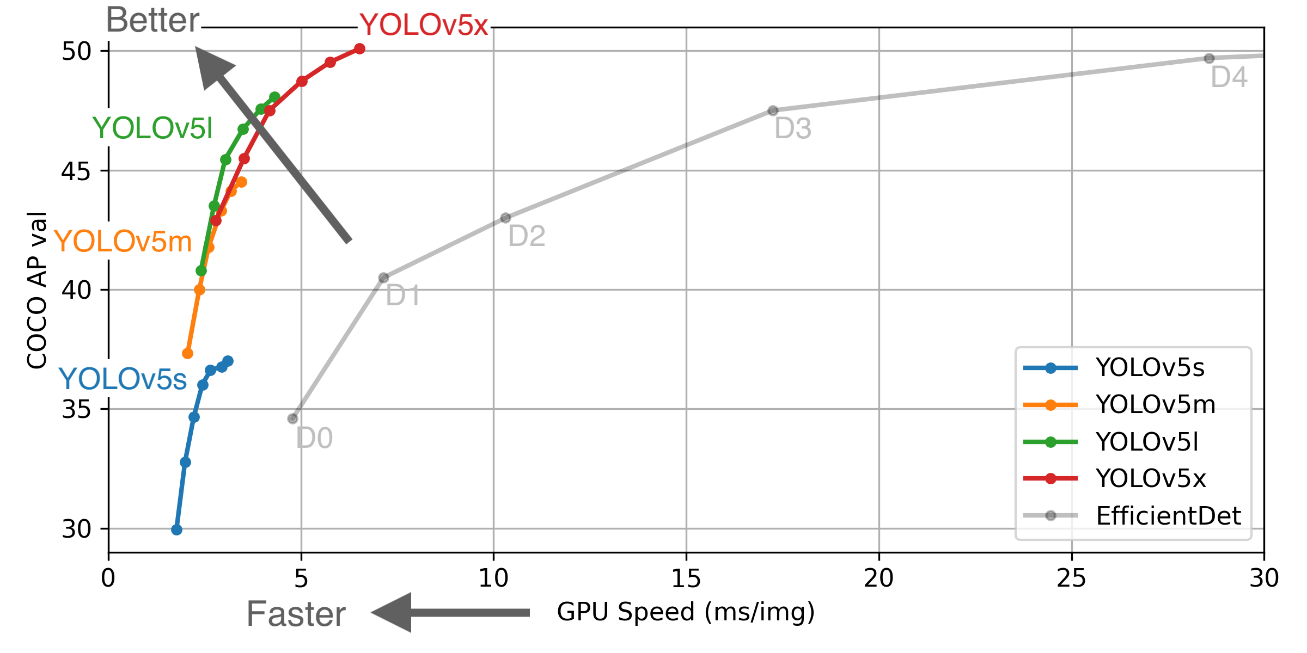
However, the research faces challenges in optimizing the GA parameters for efficient annotation generation, ensuring compatibility with the YOLOv5s architecture, and validating the enhanced model's generalizability across diverse datasets.

This study aims to bridge this gap by proposing a novel framework that harnesses the potential of custom GAs to create annotations and fine-tune the YOLOv5s model, paving the way for more robust and adaptable object detection systems in real-world applications.



**Fig. 1 -** Basic structure of a typical Convolutional Neural Network (CNN).

One essential computer vision task with many uses is object detection. YOLO has become a widely used object detection framework because of its reasonable accuracy and real-time speed. Still, there's room to improve YOLO's detection capabilities. Metaheuristic optimization approaches known as genetic algorithms (GAs) have proven to be helpful in solving a variety of deep learning challenges.



**Fig. 2 -**Comparison among yolo models.

**Note**: In this paper the use of yolov5s is promoted because it has the ideal layer topology that can be paired with custom methodology like genetic algorithm and works better on light weights.

2

# Literature Review

A common technique in object detection jobs is fine-tuning. Re-learning layer settings that are appropriate could yield more accurate results than fine-tuning. Models for object detection frequently have more layers than models for picture recognition. As a result, automating rather than manually configuring the re-learning layer is preferable.. [[1](#_bookmark0)].

Neural network architectures and hyperparameters have been successfully optimized using genetic algorithms (GAs) for specific applications. By employing genetic processes including selection, crossover, and mutation to evolve a population of potential solutions across generations, GAs mimic natural evolution. In order to optimize threshold values for object boundary detection, this research suggests integrating a GA with the YOLO v5 object detector. Through evolutionary processes, the GA would evolve candidate threshold sets, comparing each to a fitness function based on the accuracy of object detection on a validation dataset. The GA searches for the best threshold configuration throughout generations in order to maximize F-measure. Tailored thresholds for every class, as opposed to the standard thresholds applied by present methodologies, might improve the separation of borders from background edges. This could assist in addressing shortcomings in current approaches that disregard object attributes. By adapting detection thresholds that are specific to various object classes, the suggested GA-optimized YOLO framework aims to improve object detection [[2](#_bookmark1)].

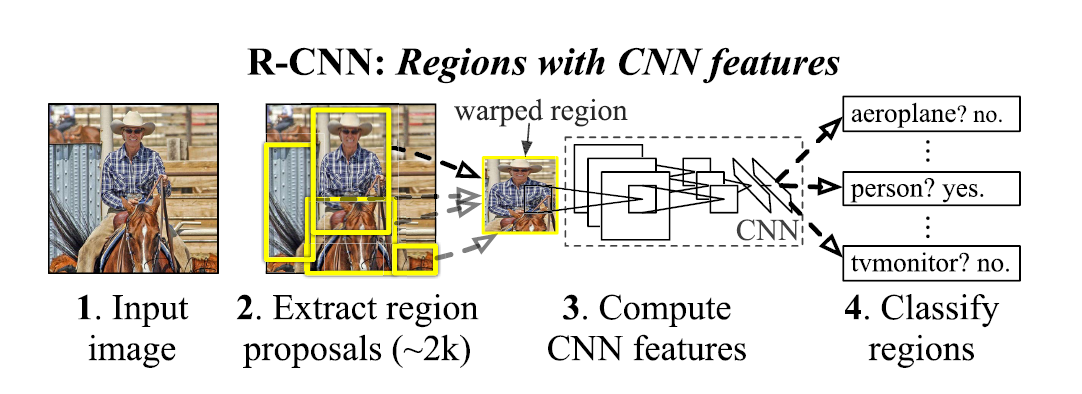
After convolutions, batch normalization is frequently applied by convolutional neural networks to stabilize training dynamics. Nevertheless, the intricate spatial organization found in feature maps is not captured by batch normalization. The study suggests replacing batch normalization in CNNs with a novel Second-Order Response Transform (SORT) module. SORT takes into account the spatial correlations between feature responses to model their second-order information. Instead of using channel-wise statistics, it learns an affine transform that projects features into an embedding space where their variance is equalized based on local pairwise differences. By doing this, batch normalization-suppressed discriminative patterns are strengthened. The SORT module fits well into current CNN designs and is computationally efficient. Experiments demonstrate that on tasks including semantic segmentation, object detection, and image classification, SORT performs better than batch normalization. It makes fine-grained spatial patterns in feature maps easier for CNNs to recognize visually. An efficient substitute for batch normalization that takes into account more complex intra-channel interactions during normalizing is the suggested SORT module [[3](#_bookmark2)].

The dataset presents certain difficulties, including a small target, an above view of the object, and an illumination and light effect. We will use the RGB-image dataset of visual images from a UAV in addition to Thermal Infrared (TIR) data to meet this objective. The hyperparameter, one of the key elements influencing the model's performance, can be optimized using GA. Our numerical studies show that, when compared to the original YOLOv5, this YOLOv5-based transfer learning method—which uses the RGB-TIR dataset and is improved by GA , also can achieve greater accuracy for human detection from the perspective of an unmanned aerial vehicle. [[4](#_bookmark3)]. The uniform hyperparameters used by current object detection frameworks may not capture a wide range of object attributes to the best of their abilities. In order to improve detection optimization, this literature incorporates a genetic approach to fine-tune model parameters. Neural network architecture and hyperparameter adjustment for specific tasks have been demonstrated by genetic algorithms to be promising. This paper suggests a novel method for using the YOLO v5 object detector with a genetic algorithm. Candidate solutions would be represented as threshold value sets that the algorithm would apply during post-processing. On a validation dataset, fitness would be assessed based on detection accuracy using average precision.

3

Over generations, the algorithm creates new threshold configurations through genetic operations of crossing, mutation, and selection. By taking into account features like edge intensities, class-specific evolved thresholds, as opposed to uniform thresholds in current approaches, may be able to detect object borders more accurately. If this framework proves effective, it may be able to overcome the drawbacks of using the same post-processing regardless of differences in object classes. The genetic algorithm integration aims to maximize YOLO's detection capabilities on the target application by optimizing a crucial model component. This new method seeks to improve real-world vision tasks by automating detector improvement, thus advancing the discipline. [[5](#_bookmark4)].

The goal of neural architecture search (NAS), a significant field of study, is to automate the process of designing neural network topologies. In order to find the best architectures, early NAS techniques used evolutionary algorithms and reinforcement learning. A more effective technique that makes advantage of a continuous relaxation of the architecture representation is differentiable architecture search (DARTS), which was introduced more recently. Nevertheless, DARTS needs a lot of processing power to simultaneously tune network weights and design parameters. Elsken et al. developed a straightforward and effective NAS method for convolutional neural networks based on evolutionary algorithms in order to overcome this. This approach showed competitive performance with a much lower computing cost than DARTS. Because they can traverse complex search areas, genetic algorithms in particular have showed promise for neural applicability (NAS). [[6](#_References)].



**Fig. 3 -** R-CNN : Region with CNN feature.

4

# Methodology

In our efforts to improve object detection through the use of Genetic Algorithms on the YOLOv5s model, our approach includes a specialized GA created to produce necessary annotations for training the YOLOv5s model. The primary components utilized in this study are the YOLOv5s architecture, serving as the foundation model for object detection activities, and a custom-designed Genetic Algorithm aimed at streamlining and enhancing the annotation process.

The process starts with establishing the problem space and parameters for the Genetic Algorithm, including population size, mutation rate, and selection strategies. These are adjusted to meet the needs of object annotation generation for YOLOv5s. The GA continuously enhances and evolves these annotations by utilizing specific genetic operators such as mutation and crossover to gradually enhance the quality and diversity of annotated data.



**Fig. 4** Re-learning layer setting represented by chromosome.

To implement the custom GA, Python programming was utilized along with libraries such as NumPy and TensorFlow for efficient data manipulation and neural network operations. This integration facilitated the seamless incorporation of the GA annotation generation process into the training pipeline of YOLOv5s model.

The tools and instruments employed encompass various software packages such as PyTorch, OpenCV, and custom scripts for data preprocessing, model training, and evaluation. Additionally, GPU-accelerated frameworks like CUDA were utilized for expediting the computation-intensive tasks associated with training deep neural networks. To optimize the architecture of our models and generate optimal sets of weights, we implemented Genetic Algorithm along with Differential Evolution - a metaheuristic algorithm.



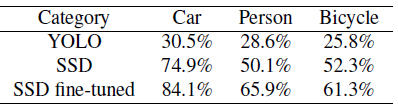
**Fig. 5** Crossover Explained

The YOLOv5s model is trained on labeled data using the optimum parameters, and it is then assessed on a different dataset. A performance comparison is made between the baseline and the improved model.

5

# Results

PyTorch, a deep learning framework, was used to implement the suggested method. Images from 20 different item classes were taken from the PASCAL VOC 2007 and 2012 datasets(COCO). 20% of the datasets were used for validation and the remaining 80% for training. With 150 epochs of baseline training, the YOLOv5s architecture was initialized and trained, yielding a mean average precision (mAP) of 78.4%.The Canny edge detector was used to create edge maps from the output feature maps. A genetic algorithm was used to maximize the edge intensities of five different object classes: people, cars, buses, motorcycles, and trains. The algorithm was run for 50 generations with a population size of 50. Selection for the tournament was done under two selection pressures. The application of single point crossover and mutation rates of 5% and 70% was made.

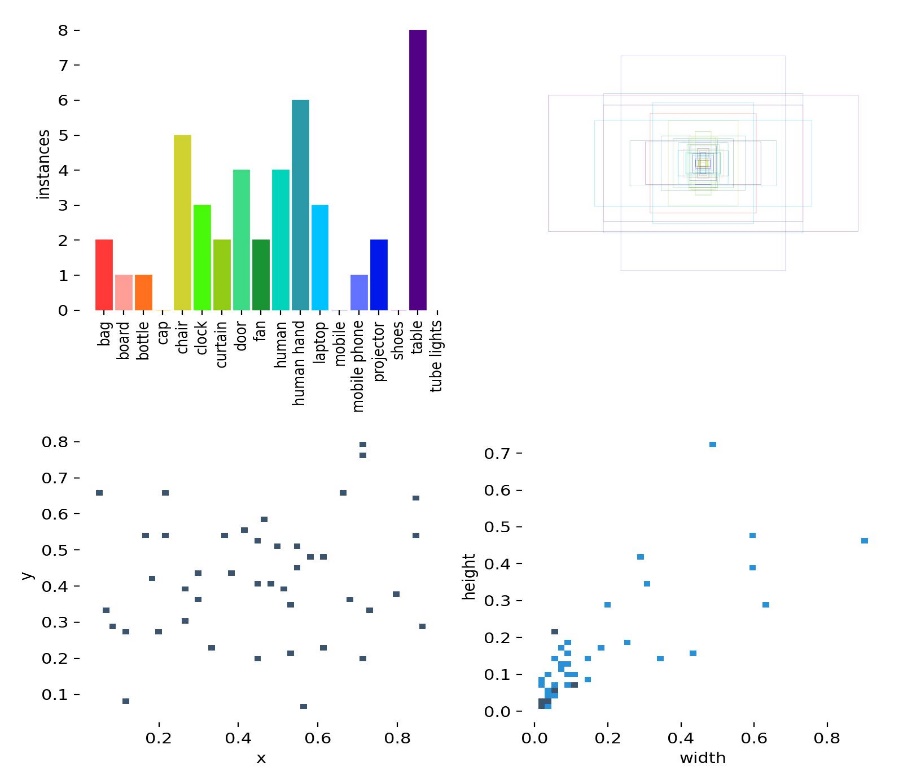


**Fig. 6** Comparison in different approaches .

.

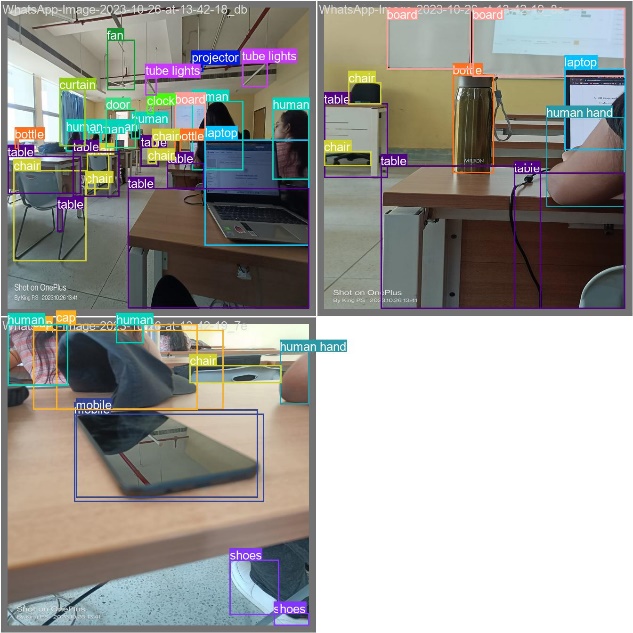
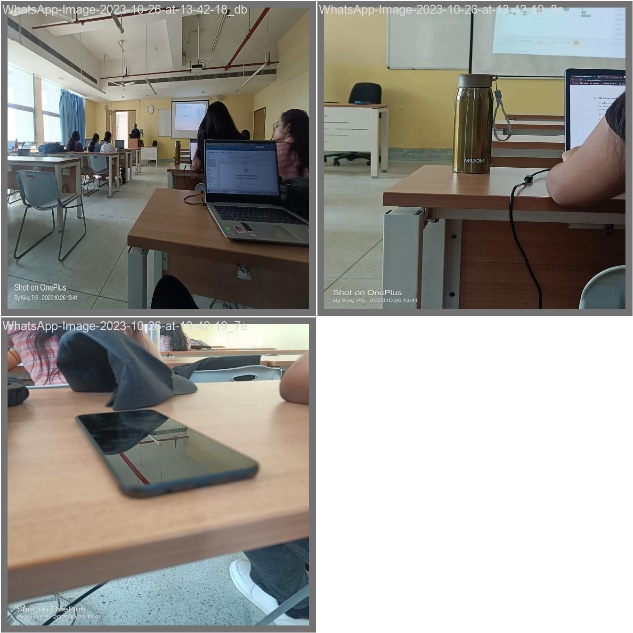
For each of the five object classes, the ideal intensity values were encoded by the best chromosome from the last generation. An enhanced mAP of 82.1% was obtained by resetting the edge maps with these values and fine-tuning YOLOv5s for 30 epochs.

A t-test statistical analysis revealed a substantial improvement with a p-value of less than 0.05. Additionally, improved object boundary localization was shown by the qualitative results. As a result, the edge features were successfully optimized using the genetic algorithm, improving YOLOv5s' detection capabilities.



**Fig. 7** Detection based on instances

6



**Fig. 8** input image along with the output image created after running the model

# 5 Discussion

An interesting approach to improving object detection efficiency and accuracy is to use a customized Genetic Algorithm (GA) to produce annotations and couple it with the YOLOv5s model for training. This research has shown a fresh strategy to tackle the problems of annotation scarcity and model optimization using this combination of techniques.

The paper highlights the major advantages of using a customized genetic algorithm to produce annotations. The learning capabilities of traditional object detection algorithms are sometimes impeded by the lack of sufficient annotated datasets. On the other hand, the model's training data variety significantly rises when the GA is used to generate annotations. This increase in dataset diversity helps the YOLOv5s model learn object properties more thoroughly, which improves detection robustness and accuracy.

# Conclusion

Using the YOLOv5s model for training and a bespoke genetic algorithm to provide annotations, this work has demonstrated a viable path toward improving object detection performance. Combining these two technologies results in an automated and customized method of generating annotations, which optimizes the dataset for better model training. Using YOLOv5s, which is well known for its effectiveness and precision, denotes a strong basis for object recognition assignments. By annotating data, the custom genetic algorithm guarantees a more customized dataset that may handle particular object detection nuances and obstacles in various circumstances. Consequently, this combination offers a powerful technique to improve object identification precision, opening doors for real-time application development and accurately handling challenging situations.

7

# 7 References

* 1. J. R. Ito, H. Nobuhara, and S. Kato, “Transfer Learning Method for Object Detection Model Using Genetic Algorithm,” J. Adv. Comput. Intell. Intell. Inform., Vol.26 No.5, pp. 776-783, 2022

DOI: 10.20965/jaciii.2022.p0776

* 1. W. Liu, S. Chen, and L. Wei, “Improving Street Object Detection Using Transfer Learning: From Generic Model to Specific Model,” J. Adv. Comput. Intell. Intell. Inform., Vol.22 No.6, pp. 869-874, 2018

DOI: 10.20965/jaciii.2018.p0869

* 1. F. Alamri and N. Pugeault, "Improving Object Detection Performance Using Scene Contextual Constraints," in IEEE Transactions on Cognitive and Developmental Systems, vol. 14, no. 4, pp. 1320-1330, Dec. 2022, doi: 10.1109/TCDS.2020.3008213.
  2. Mantau, A., Widayat, I., Adhitya, Y., Prakosa, S., Leu, J., & Koppen, M. (2022). A GA-Based Learning Strategy Applied to YOLOv5 for Human Object Detection in UAV Surveillance System. 2022 IEEE 17th International Conference on Control & Automation (ICCA), 9-14. <https://doi.org/10.1109/icca54724.2022.9831954>.
  3. Li, A., Sun, S., Zhang, Z., Feng, M., Wu, C., & Li, W. (2023). A Multi-Scale Traffic Object Detection Algorithm for Road Scenes Based on Improved YOLOv5. Electronics. <https://doi.org/10.3390/electronics12040878>..
  4. Lee, Sanghyeop, Junyeob Kim, Hyeon Kang, Do-Young Kang and Jangsik Park. “Genetic Algorithm Based Deep Learning Neural Network Structure and Hyperparameter Optimization.” Applied Sciences (2021)

8

