Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 1, January 2026
DOI : https://doi.org/10.33395/sinkron.v10i1.15804
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Performance Trade-off of Anchor-Based and Anchor-Free Approaches of Faster R-CNN for Face Detection

Arie Satia Dharma[1)*], Herimanto[2)], Niar Fujita Simbolon[3)], Anton Roycar Nababan[4)]
[1,2,3,4)] Institut Teknologi Del, Laguboti, Indonesia
[1)]ariesatia@del.ac.id, [2)]pardzheri@gmail.com, [3)]niarfujita1@gmail.com, [4)]antonnababan214@gmail.com

Submitted :Jan 12, 2026 | **Accepted** : Feb 5, 2026 | **Published** : Feb 9, 2026

Abstract: The face is a unique biometric feature that plays a crucial role in individual identification as it holds essential information for identity recognition. Face detection technology has been experiencing significant advancements in the field of computer vision. However, face detection technology continues to face challenges in balancing high detection accuracy with computational efficiency. While deep learning has advanced this field, there remains a lack of comparative studies that compare the performance trade-offs between anchor-based and anchor-free region proposal mechanisms within a Faster R-CNN framework. This research objective is comparing the performance of face detection using two approaches: anchor-based and anchor-free. The anchor-based approach use anchor boxes to predict bounding boxes, while the anchor-free approach predicts bounding boxes directly from pixel positions oriented around a point. The anchor-based approach is implemented use base line region proposed network method, whereas the anchor-free approach use a centerpoint method. The study utilizes a custom dataset comprising 1,000 formal images of students from Del Institute of Technology, split into 900 training images and 100 testing images. Performance evaluation is conducted based on metrics such as intersection over union, precision, recall, and latency. The results demonstrate that the anchor-based approach achieves superior accuracy with an average IoU of 0.98 but requires a longer detection time of approximately 2.33 seconds per image. Conversely, the anchor-free approach offers significantly faster processing at 0.14 seconds per detection, though with a lower average IoU of 0.78. This study concludes that while anchor-based methods excel in precision, anchor-free architectures provide alternative for time-critical applications, offering a clear reference for optimizing future face detection systems.

Keywords: Face Detection, Faster R-CNN, Anchor-free, Anchor-based, Region Proposed Network

INTRODUCTION

In the rapidly evolving domain of computer vision, biometric identification has established itself as a cornerstone of modern digital security and personal verification systems. Among the various biometric modalities available, the human face is widely regarded as unique feature because involving the identification of features such as eyes, eyebrows, mouth, and nose (Chanda et al., 2024). Consequently, face detection has become a crucial step in a wide array of facial analysis applications, including face recognition, face modeling, verification, and real-time tracking. The output of this process typically involves generating bounding boxes or facial markers that delineate the location and size of detected faces. Despite its extensive utility, robust face detection remains a persistent challenge in computer vision research (Agrwal et al., 2023). While modern detectors can easily identify faces in controlled, frontal-view scenarios, real-world applications often present complex variables (Aditiawarman et al., 2023) (Siahaan et al., 2024). As technology advances, detection systems must contend with significant impediments such as extreme variations in lighting conditions, diverse facial poses, and complex expressions, which often degrade the performance of traditional detection methods (Damarsiwi et al., 2024).

To overcome the limitations of traditional image processing techniques, the field has witnessed a paradigm shift toward machine learning, specifically Deep Learning. Deep learning models utilize multi-layered artificial neural networks to extract abstract and complex features from raw data (Kaur & Singh, 2021), offering superior

*Arie Satia Dharma

e-ISSN : 2541-2019
p-ISSN : 2541-044X

performance in pattern recognition. Within this landscape, Convolutional Neural Networks (CNN) have emerged as the dominant architecture for visual tasks (Amjoud & Amrouch, 2023). The evolution of object detection has been marked by the development of Region-based Convolutional Neural Networks (R-CNN), which utilize selective search to identify Regions of Interest (RoIs) (Ren et al., 2017). However, the original R-CNN faced significant bottlenecks regarding computational speed and efficiency (Sunario Megawan & Wulan Sri Lestari, 2020). This led to the introduction of Fast R-CNN, which improved efficiency by aggregating features in a single forward pass (Li et al., 2018).

Currently, the Faster R-CNN framework represents the state-of-the-art in this lineage. It advances its predecessors by integrating a Region Proposal Network (RPN) directly with the CNN model. The RPN is trained end-to-end to generate high-quality region proposals, which are subsequently processed by the Fast R-CNN module for final detection (Ren et al., 2017) (Salam & Jebaseeli, 2024). Conventionally, Faster R-CNN operates as an anchor-based method. These anchor-based mechanisms have set the benchmark for high accuracy, particularly in detecting small objects.

Despite the success of anchor-based architectures, they introduce substantial computational overhead. The reliance on predefined anchor boxes necessitates complex calculations, such as Intersection over Union (IoU) matching for thousands of candidate boxes, which can slow down the detection process. Furthermore, the performance of these models is often sensitive to the manual design of anchor hyperparameters. To address these inefficiencies, anchor-free detection frameworks have recently gained prominence. While anchor-free models theoretically offer a simpler and faster alternative by removing the anchor matching step (Liu et al., 2020), there remains a critical need to rigorously compare these two paradigms within a unified framework. Existing literature often compares different algorithms entirely (Wang et al., 2023) (Zhang et al., 2020), but fewer studies isolate the specific impact of the region proposal mechanism within the same Faster R-CNN architecture to understand the precise trade-offs between accuracy and latency.

This research aims to bridge this gap by conducting a comparative performance analysis of face detection technology using two distinct approaches: the traditional anchor-based method and the emerging anchor-free method. The primary objective is to implement both approaches within the Faster R-CNN algorithm to evaluate their effectiveness in a controlled context. By keeping the backbone architecture constant, this study seeks to provide a fair and valid comparison, quantifying the specific trade-offs between detection accuracy, measured by IoU, Precision, and Recall, and computational efficiency, measured by Latency.

To achieve these objectives, we propose the development of two face detection models using the Faster R-CNN framework. The first model utilizes the standard Region Proposal Network (RPN) as the anchor-based baseline. The second model integrates a CenterPoint method (Anchor-Free RPN), which modifies the architecture to predict bounding boxes directly from the center of facial features without predefined anchors. Both models are trained and evaluated on a custom dataset comprising 1,000 formal images of students from the Del Institute of Technology, characterized by uniform red backgrounds and consistent lighting. Prior to training, image quality is enhanced using Laplacian sharpening preprocessing to highlight facial edges. This comprehensive evaluation will provide deep insights into which approach offers the optimal balance for future face detection systems.

LITERATURE REVIEW

The field of object detection has undergone a significant transformation with the advent of deep learning, particularly through the development of Convolutional Neural Networks (CNNs). Early methods relied on traditional image processing techniques which often struggled with complex variations in lighting and pose. However, the introduction of Region-based Convolutional Neural Networks (R-CNN) marked a pivotal shift in addressing these challenges. The original R-CNN architecture operated by using a selective search algorithm to identify potential Regions of Interest (RoIs) in an image, which were then individually processed by a CNN to classify objects (Ren et al., 2017). Despite its effectiveness in improving detection accuracy compared to traditional methods, R-CNN faced substantial limitations regarding computational speed and memory efficiency due to the redundant processing of overlapping regions (Sunario Megawan & Wulan Sri Lestari, 2020).

To mitigate the inefficiencies of the original R-CNN, the Fast R-CNN architecture was subsequently developed. This iteration improved upon the predecessor by aggregating feature extraction into a single forward pass through the network for the entire image, rather than processing each region separately. This modification allowed Regions of Interest (RoIs) from the same image to share computation and memory, significantly accelerating the training and detection process (Li et al., 2018).

The evolution continued with the introduction of the Faster R-CNN framework, which established a new state-of-the-art by addressing the bottleneck of region proposal generation. Unlike Fast R-CNN, which still relied on external algorithms for proposals, Faster R-CNN integrates a dedicated Region Proposal Network (RPN) directly into the architecture (Lu et al., 2022). This RPN is trained end-to-end to generate high-quality region proposals, which are subsequently used by the Fast R-CNN detector module (Srikar & K, 2022). This unified approach not

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 1, January 2026
DOI : https://doi.org/10.33395/sinkron.v10i1.15804
e-ISSN : 2541-2019
p-ISSN : 2541-044X

only enhances detection speed but also improves the quality of the proposals, making it highly effective for complex tasks such as face detection (Salam & Jebaseeli, 2024).

Within the domain of modern object detection, particularly in frameworks like Faster R-CNN, the methodology for localizing objects is divided into two primary paradigms: anchor-based and anchor-free approaches.

The anchor-based approach has long been the standard for high-accuracy detection systems. This method involves the creation of predefined anchor boxes or rectangular regions at various scales and aspect ratios to cover the spatial extent of an image (Zhou et al., 2020). During the detection process, the network predicts the probability of an object's presence within these anchors and regresses the coordinates to refine the box location (Faishal et al., 2023). Anchor-based methods are renowned for their robustness and high detection accuracy, particularly when dealing with small or overlapping objects. However, they introduce significant computational overhead. The reliance on dense anchor sampling requires the model to process thousands of candidate boxes, which can slow down the inference time and increase the complexity of hyperparameter tuning (Wang et al., 2023).

Conversely, anchor-free frameworks have emerged as a streamlined alternative aimed at improving efficiency. Instead of relying on predefined boxes, anchor-free models treat object detection as a point prediction problem. These methods directly predict bounding boxes and objectness scores based on pixel locations, often orienting the detection around a centerpoint or key points of the object (Jiao et al., 2020) (Zhang et al., 2020). By eliminating the need for anchor box calculations and matching, anchor-free architectures significantly reduce the model's complexity (Liu et al., 2020), and no need hyper parameters (Zhang et al., 2022). Recent studies indicate that while anchor-free methods, such as FCOS or CenterNet, offer promising results in terms of real-time performance and simplified architectural design, they sometimes face challenges in scenarios requiring extreme precision compared to their anchor-based counterparts (Liu et al., 2022).

Ultimately, the choice between these approaches represents a trade-off. Anchor-based methods generally excel in precision and recall, making them suitable for applications where accuracy is paramount. In contrast, anchor-free methods provide a more efficient solution for time-critical applications where lower latency is required. However, no prior work has explicitly evaluated both strategies under the same Faster R-CNN architecture using identical datasets and evaluation protocols

METHOD

In this research, we proposes a face detection system based on the Faster Region Convolutional Neural Network (Faster R-CNN) framework by comparing two region proposal approaches, namely anchor-based Multi-scale Region Proposal Network (RPN) and anchor-free Region Proposal Network (AF-RPN). The overall methodology consists of dataset preparation, image processing, feature extraction using a convolutional backbone, region proposal generation, object classification, and performance evaluation. The workflow of the research is presented in Figure 1.

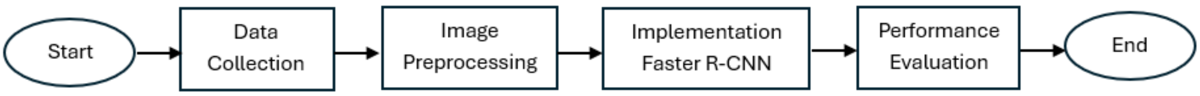

Fig. 1 Research Stages

Data Collection

This dataset consists of formal images of all active students of the Del Institute of Technology, class of 2020-2023 as presented in Figure 2. In this research, we limited to using only IT Del students' facial data and did not use facial data from the public. The collected data consists of images of students with a red background and neatly dressed students, in JPG format, with a total of 1,000 images. For face detection purposes, each image in the dataset has been annotated using a bounding box that marks the location of each student's face. This annotation was done manually by a team of experts to ensure high accuracy. Each bounding box is equipped with the corresponding coordinate information (x, y, width, height).


Fig. 2 Example of Student Dataset

Image Preprocessing

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 1, January 2026
DOI : https://doi.org/10.33395/sinkron.v10i1.15804
e-ISSN : 2541-2019
p-ISSN : 2541-044X

The image processing method used to improve the quality of dataset images is the sharpening technique. In the sharpening stage, a Laplacian sharpening kernel is used, this kernel reduces the intensity value of neighboring pixels and increases the intensity value of the center pixel (Yousif, 2024), resulting in a sharper image. This kernel works by calculating the difference between the center pixel and its neighbors, which results in a sharpening effect on the edges of the image. The right pixel selection is very important in this process. The center pixel is selected based on the highest intensity value in its surroundings, so that the contrast between the center and neighboring pixels can be maximized. This process ensures that the edges of the image are more prominent without adding unwanted artifacts. This pre-processing process aims to produce sharper and clearer images, thereby improving the performance of the face detection model in recognizing important features on the face.

The dataset is divided into two main subsets: a training set of images used to train the face detection model and a testing set of images used to test the model's performance after training is complete. This study implements a hold-out validation method where we experiment with four different partitioning scenarios to achieve the best partitioning accuracy for a CNN model. The dataset is split into training and testing ratios of 90:10, 80:20, 70:30, and 60:40. This systematic variation aims to analyze the impact of training data volume on model convergence and generalization ability, and ultimately identify the optimal partitioning for this specific domain. The best partition will be used subsequently for object detection.

Faster Region-Based Convolution Neural Network

This study using the Faster R-CNN framework for object detection. The core difference in this study lies in the Region Proposal Network (RPN) mechanism as presented in Figure 3.

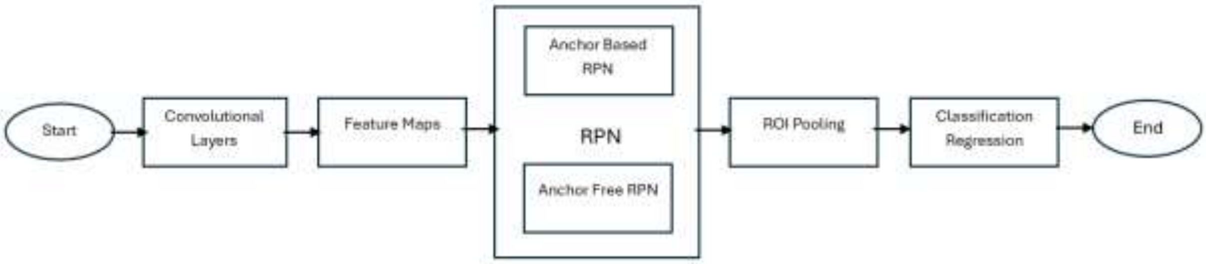

Fig. 3 Faster R-CNN Implementation

In the Faster R-CNN with anchor-based RPN process above, the input image is first processed through a sequence of convolutional layers that function to extract fundamental visual features such as edges, corners, and texture patterns. We utilizes VGG-16 as the backbone network due to its stability and lower resource consumption compared to deeper networks like ResNet-50. Hyperparameters represent configuration settings that regulate the learning behavior of machine learning algorithms and their proper selection critically determines system effectiveness (Franceschi et al., 2025) (Prasetyo et al., 2023). The hyperparameters utilized for training VGG-16 as shown in table 1. The resulting output is a feature map that represents both spatial and high-level abstract characteristics required for object detection.

Table 1. Configuration Hyperparameter

Parameter	Values
Backbone	VGG-16 (conv layers only)
Learning rate	0.001
Batch size	8, 16, 32, 64
Epochs	500
Optimizer	SGD

This feature map is then utilized by the Region Proposal Network (RPN) to identify candidate regions that potentially contain objects by sliding anchors across the feature map, classifying each anchor as object or background, and refining their size and position to better match object boundaries. At each spatial location, RPN generating a set of anchors with predefined scale and aspect ratios as shown in table 2.

Table 2. Ratio of Anchor Box

Scale	Ratio of Anchor Box		
32 x 32	1:1	2:1	1:2

*Arie Satia Dharma

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 1, January 2026
DOI : https://doi.org/10.33395/sinkron.v10i1.15804
e-ISSN : 2541-2019
p-ISSN : 2541-044X

The generated region proposals are then passed to RoI Pooling. Subsequently, Region of Interest (RoI) Pooling is applied to extract and normalize features from each proposed region into a fixed size, enabling consistent processing despite variations in proposal dimensions. Finally, the pooled features are passed through classification and regression layers to determine the object class and to further refine the bounding box coordinates, producing accurate object localization and classification results.

Anchor Free Region Proposal Network

This research used center point–based anchor-free Region Proposal Network (RPN) to eliminates the use of predefined anchors and formulates object proposal generation as a keypoint detection problem. The stages of centerpoint algorithm as shown in figure 4.

Fig. 4 CenterPoint Implementation

In the first stage, the input image is processed by a backbone network to produce dense convolutional feature maps then forwarded to the network predicts a center point heatmap, where each pixel represents the likelihood of being the geometric center of an object. In the second and third stage, additional regression branches predict object attributes at each detected center point, typically including object width and height and center offsets to compensate for feature map downsampling. In the fourth stage, candidate proposals are generated by combining the predicted center locations with the corresponding size regressions to form bounding boxes. Last, a lightweight post-processing step, such as using score thresholding between 0.5-0.8, is applied to retain the most confident proposals, then forwarded to the detection head for classification and further refinement.

Evaluation Method

To assess the performance of the face detection system, several evaluation metrics are used namely Intersection over Union (IoU), Precision and Recall, and Latency. Intersection over Union measures the accuracy of an object detection model in localizing objects. It calculates the degree of overlap between the predicted bounding box (pd) and the ground truth bounding box (gt) provided by human annotation. The IoU value ranges from 0 to 1, a detection is generally considered successful if the IoU score exceeds a predefined threshold. In this research we u used 0.75 as a threshold.

In addition to localization, the model's overall accuracy is evaluated through Precision and Recall. Precision represents the proportion of correctly identified faces out of all positive predictions made, focusing on minimizing False Positives (FP). Recall, however, measures the proportion of actual object faces that were correctly identified, which helps in assessing how well the model avoids False Negatives (FN). Predictions are classified as True Positive (TP), False Positive (FP), or False Negative (FN) based on IoU and a predefined threshold. A model is considered high-performing when it achieves high values for both metrics, indicating that it is both accurate and comprehensive in its detections.

Latency used to evaluate computational efficiency and real-time feasibility. Latency refers to the time a computer system takes to recognize and process faces in an image after receiving input, measured in seconds. It indicates the delay until an output is available.

RESULT

Image Preprocessing

Before entering the modeling phase, the dataset undergoes a critical preprocessing step that is sharpening with Laplacian kernel. This technique utilized to enhance the detail and clarity of the images. This aims to clarify facial features such as eyes, nose, and lips while improving the overall texture of the image. The results show a significant difference between the original and sharpened images, with lines becoming more distinct as shown in figure 5.

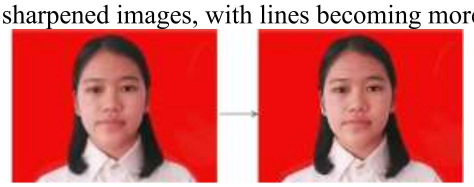

Fig. 5 Before and After Sharpening Image

For dataset division, using 1000 IT Del students' facial data, it was divided into two partitions, namely training data and testing data. Using the hold-out splitting method, we tested the training and testing data in several scenarios and ran the VGG-16 model training. Then we compared the results to determine the best scenario for the data partitioning as shown in table 3.

*Arie Satia Dharma

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 1, January 2026
DOI : https://doi.org/10.33395/sinkron.v10i1.15804
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Table 3. Result

Train : Test Split Partition	Train Acc	Val Acc	Train Loss	Val Loss
90:10	94.8	86.3	0.12	0.24
80:20	93.6	80.1	0.14	0.31
70:30	92.4	75.8	0.16	0.38
60:40	91.2	71.4	0.18	0.46

Based on Table 3 above, it was found that the 90:10 partition scenario was the best. Therefore, the 90:10 data partition was used in the implementation process of object detection using Faster R-CNN.

Convolusional Layer Backbone

Face detection using Faster R-CNN framework started with by specifying a convolutional neural network. The VGG16 model is used with hyperparameters as shown in table 4.

Table 4. Experiment Result of Hyperparameter

Epoch	Learning Rate	Batch Size	Accuracy	Loss
500	0.001	8	0.89	0.03
500	0.001	16	0.90	0.02
500	0.001	32	0.94	0.02
500	0.001	64	0.93	0.01

The table 1 presents the experimental results obtained by varying the batch size while maintaining a constant number of epochs at 500 and a learning rate of 0.001. When a batch size of 8 is applied, the model achieves an accuracy of 0.89. Increasing the batch size to 16 results in a slight improvement in accuracy to 0.90. A more noticeable increase in performance is observed at a batch size of 32, where the highest accuracy of 0.94 is recorded. However, when the batch size is further increased to 64, the accuracy decreases to 0.93. Therefore researcher will use hyperparameters with epoch 500, learning rate 0.001, dan batch size 32 to form the Faster R-CNN's convolutional network. Visualization of accuracy against epochs at each batch size can be seen in figure 6.

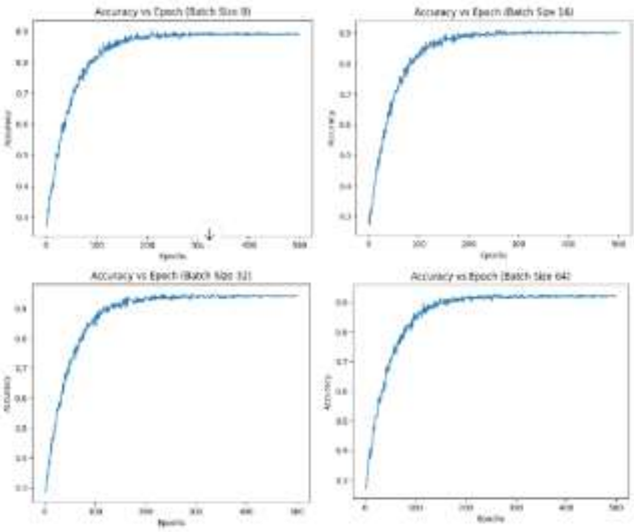

Fig. 6 Accuracy VS Epoch in Graph

Faster R-CNN With Anchor Based RPN

In the Faster R-CNN framework using anchor, researchers first create an anchor box. This anchor box is used to detect the presence of the student's face, by moving it across the image. It operates by sliding a small convolutional network over the shared feature maps and, at each spatial location, generating a set of anchors with scales is 32 and aspect ratios that is 0.5, 1.0, and 2.0 to represent face objects of varying sizes and shapes.

For each anchor, the RPN simultaneously predicts a score for that detected face, indicating whether the anchor corresponds to a face being searched, and a set of bounding box regression offsets to refine the anchor's position and dimensions. These refined anchors are then transformed into candidate region proposals by applying the predicted regression offsets to the predefined anchor boxes. Region proposals generated by RPN with a high probability of containing facial objects will be forwarded to the Classification and Regression stages. At this stage, class labels and face bounding boxes will be determined, with confidence scores as shown in the figure 7.

*Arie Satia Dharma

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 1, January 2026
DOI : https://doi.org/10.33395/sinkron.v10i1.15804
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Fig. 7 Results of Face Detection Using Anchor Based RPN

Faster R-CNN With Anchor Free RPN

In anchor-free approaches with centerpoint, generates region proposals by first predicting facial object center points on dense feature maps. These feature maps serve as the basis for identifying the most representative locations of faces in the image. The network predicts a center-point heatmap in which each pixel represents the likelihood of being the geometric center of a face. High-response points on this heatmap correspond to potential face centers.

For each detected center point, the network simultaneously regresses the facial bounding box attributes, such as width and height, as well as center offsets. The predicted center points and regressed size information are combined to construct face bounding boxes directly. A filtering process using best score is applied to retain the most confident detections.

Fig. 8 Results of Face Detection Using Anchor Free RPN

Figure 8 shows face detection with different boundary boxes. This occurs because the coordinate points used by centerpoint are not in the center of the face, causing the predicted boundary lines to exceed the facial boundaries in the image.

Performance Evaluation

The performance comparison of face detection using the Faster R-CNN with Anchor and Faster R-CNN with Centerpoint is done using the IoU, Precision, Recall, and Latency metrics as shown in table 6. It can be seen that the Anchor RPN method has higher scores in IoU, Precision, and Recall, while the Centerpoint RPN method has higher in Latency.

Table 5. Comparison Two RPN Approaches in Faster R-CNN

Type of RPN	IoU	Precision	Recall	Latency
Anchor	0.98	0.98	0.78	2.33
Centerpoint	0.78	0.81	0.67	0.14

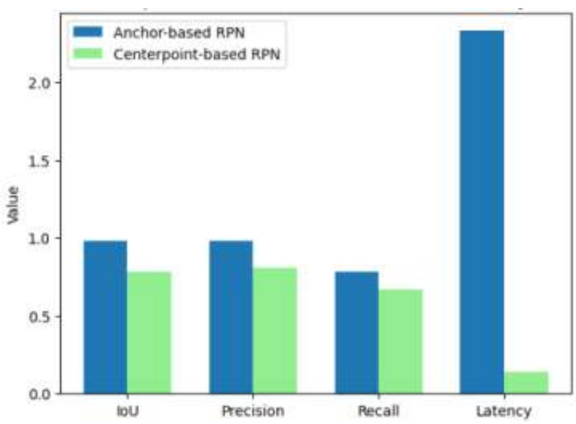

Fig. 9 Comparison Graph of IoU, Precision, Recall, and Latency

*Arie Satia Dharma

Figure 9 show comparison of performance between Anchor RPN and Centerpoint RPN. The comparison graph shows that the anchor-based RPN consistently achieves higher performance in terms of IoU, precision, and recall, indicating more accurate and reliable object localization and detection. However, this improved accuracy comes at the cost of significantly higher latency, reflecting the computational overhead introduced by anchor generation and IoU matching. In contrast, the centerpoint-based RPN demonstrates substantially lower latency, making it more suitable for real-time applications, although with a moderate reduction in accuracy-related metrics.

DISCUSSIONS

This study demonstrates that the choice of region proposal methodology has significant implications for the balance between accuracy and speed. This can be seen in the experimental results demonstrate that both region proposal strategies Anchor RPN and Centerpoint RPN are capable of detecting faces using the Faster R-CNN framework, but they exhibit different performance characteristics. The use of VGG-16 as a constant backbone provides a fair basis for isolating the impact of anchor-based and anchor-free approaches. Furthermore, the Laplacian sharpening preprocessing stage proves methodologically crucial as it clarifies facial features, which in turn produces more informative feature maps for the detection stage. These feature maps provide a strong foundation for face detection.

In terms of face detection performance, the Anchor-Based RPN shows higher localization accuracy, as reflected by its higher IoU, precision, and recall values. The high score of the anchor-based method on RPN is due to the use of anchor boxes with various scales and aspect ratios such as 0.5, 1.0, and 2.0. This mechanism allows the model to densely sample thousands of candidate boxes across the entire image, resulting in a much higher probability of finding a box that is highly precise with the ground truth compared to point-based methods. In contrast, the centerpoint method is able to perform significantly faster, 0.14 seconds latency, because it eliminates the complex IoU matching calculation process for thousands of anchor boxes. Instead of processing thousands of candidates, this method treats detection as a single-point prediction problem, where the bounding box is directly predicted from the facial object center pixel location. This drastically reduces the computational load and simplifies the detection workflow.

The comparison between the two approaches highlights a clear trade-off between detection accuracy and processing speed. While Anchor RPN achieves more accurate and consistent face localization, it requires longer inference time due to the increased number of proposals and anchor evaluations. On the other hand, Centerpoint RPN offers significantly lower latency, making it more suitable for applications where real-time performance is a priority. Architecturally, anchor methods introduce significant computational overhead and are highly sensitive to manual hyperparameter tuning at the anchor scale as shown in table 2. In contrast, centerpoint architectures offer a more compact and efficient design, shifting the burden from manual box matching to directly predicting object attributes on feature maps. However, this simplification sometimes comes at the expense of extreme precision, as it relies heavily on the accuracy of center point prediction as shown in figure 8.

Overall, the results indicate that the choice of region proposal method should be guided by the specific requirements of the application. For scenarios that prioritize detection accuracy and robustness, Anchor RPN is more appropriate, whereas Centerpoint RPN is better suited for time-sensitive applications where faster processing is essential.

Future Development

While the results in this research are excellent, using more complex datasets with extreme lighting variations, diverse facial poses, or overlapping objects will present greater challenges. In such scenarios, anchor-based methods tend to be more robust and excel in precision and recall due to their ability to better handle small or overlapping objects. For future development on complex datasets, the integration of Feature Pyramid Networks (FPN) is recommended to improve multi-scale feature representation without significantly increasing the computational burden. This may improve detection performance for faces of varying sizes without significantly increasing computational cost.

Other research also demonstrate that many anchors in the detection head can be removed without negatively affecting accuracy, and that retraining after anchor pruning may even improve performance becomes up to 44% more efficient while achieving higher accuracy (Bonnaerens et al., 2022). So future development could consider the integration of anchor pruning into this architecture.

CONCLUSION

This study concludes that the Faster R-CNN method with an anchor RPN and a centerpoint RPN can be successfully implemented for student's face detection tasks. Both approaches are capable of detecting faces in the given limited dataset of IT Del students, although the result show different performance characteristics. The results show a clear trade-off between accuracy and efficiency between two approaches of RPN. The anchor approach excels in localization precision show at IoU 0.98 but has high latency that is 2.33 seconds. In contrast, centerpoint approach offers significant inference speedup that is 0.14 seconds with lower accuracy that is IoU 0.78. These results are consistent with the discussion findings and emphasize a clear trade-off between accuracy and efficiency.

The contribution of this study is to conduct a rigorous performance comparison analysis between anchor-based and anchor-free centerpoint region proposal mechanisms within a unified Faster R-CNN framework. By keeping the backbone architecture constant, this study successfully isolates the specific impact of the region proposal mechanism, thus providing a clear reference for face detection system developers in optimizing the balance between accuracy and speed according to application needs.

Despite its positive results, this study has several limitations that need to be addressed, particularly regarding data coverage and technical aspects. The data used relied on a specific dataset consisting of face images of Del Institute of Technology students with a uniform red background and consistent lighting, without involving a more varied public face dataset. Furthermore, the backbone architecture used was limited to VGG-16; although known to be stable, its feature extraction capacity is relatively modest compared to deeper artificial neural networks. Technically, specific weaknesses were also found in the application of the anchor-free centerpoint method, where the bounding box sometimes exceeded the facial area due to the determination of the center point coordinates that were not always precisely in the center of the object.

REFERENCES

Aditiawarman, U., Erlangga, D., Mantoro, T., & Khakim, L. (2023). Face Recognition of Indonesia's Top Government Officials Using Deep Convolutional Neural Network. *Jurnal RESTI*, *7*(1). https://doi.org/10.29207/resti.v7i1.4437

Agrwal, S. L., Sharma, S. K., & Kant, V. (2023). A Review on Unconstrained Real-Time Rotation-Invariant Face Detection. *2023 3rd International Conference on Intelligent Communication and Computational Techniques, ICCT 2023*. https://doi.org/10.1109/ICCT56969.2023.10076222

Amjoud, A. B., & Amrouch, M. (2023). Object Detection Using Deep Learning, CNNs and Vision Transformers: A Review. *IEEE Access*, *11*. https://doi.org/10.1109/ACCESS.2023.3266093

Bonnaerens, M., Freiberger, M., & Dambre, J. (2022). Anchor pruning for object detection. *Computer Vision and Image Understanding*, *221*. https://doi.org/10.1016/j.cviu.2022.103445

Chanda, S., Kumar, Y. N., Srivastava, S., Rani, R., Shree, M., & Mohapatra, A. K. (2024). Optimizing facial feature extraction and localization using YOLOv5: An empirical analysis of backbone architectures with data augmentation for precise facial region detection. *Multimedia Tools and Applications*, *83*(29). https://doi.org/10.1007/s11042-024-19284-8

Damarsiwi, D. K., Pambudi, E. A., Fitriani, M. A., & Wibowo, F. (2024). Face Detection in Complex Background using Scale Invariant Feature Transform and Haar Cascade Classifier Methods. *Sinkron*, *8*(2). https://doi.org/10.33395/sinkron.v8i2.13556

Faishal, M., Sulistiyo, M. D., & Ihsan, A. F. (2023). Javanese Script Letter Detection Using Faster R-CNN. *Indonesian Journal of Artificial Intelligence and Data Mining*, *6*(2), 243. https://doi.org/10.24014/ijaidm.v6i2.24641

Franceschi, L., Donini, M., Perrone, V., Klein, A., Archambeau, C., Seeger, M., Pontil, M., & Frasconi, P. (2025). Hyperparameter Optimization in Machine Learning. *Foundations and Trends in Machine Learning*, *18*(6). https://doi.org/10.1561/2200000088

Jiao, L., Dong, S., Zhang, S., Xie, C., & Wang, H. (2020). AF-RCNN: An anchor-free convolutional neural network for multi-categories agricultural pest detection. *Computers and Electronics in Agriculture*, *174*. https://doi.org/10.1016/j.compag.2020.105522

Kaur, B., & Singh, S. (2021). Object Detection using Deep Learning: A Review. *ACM International Conference Proceeding Series*. https://doi.org/10.1145/3484824.3484889

Li, J., Liang, X., Shen, S., Xu, T., Feng, J., & Yan, S. (2018). Scale-Aware Fast R-CNN for Pedestrian Detection. *IEEE Transactions on Multimedia*, *20*(4). https://doi.org/10.1109/TMM.2017.2759508

Liu, S., Zhou, H., Li, C., & Wang, S. (2020). Analysis of Anchor-Based and Anchor-Free Object Detection Methods Based on Deep Learning. *2020 IEEE International Conference on Mechatronics and Automation, ICMA 2020*. https://doi.org/10.1109/ICMA49215.2020.9233610

Liu, W., Shan, S., Chen, H., Wang, R., Sun, J., & Zhou, Z. (2022). X-ray weld defect detection based on AF-RCNN. *Welding in the World*, *66*(6). https://doi.org/10.1007/s40194-022-01281-w

Lu, J., Li, D., Wang, M., Mi, B., Wang, P., Dai, Z., & Zheng, F. (2022). Object Detection System Based on Faster R-CNN. *Proceedings - 2022 4th International Conference on Artificial Intelligence and Advanced Manufacturing, AIAM 2022*. https://doi.org/10.1109/AIAM57466.2022.00027

Prasetyo, T. A., Desrony, V. L., Panjaitan, H. F., Sianipar, R., & Pratama, Y. (2023). Corn plant disease classification based on leaf using residual networks-9 architecture. *International Journal of Electrical and Computer Engineering*, *13*(3). https://doi.org/10.11591/ijece.v13i3.pp2908-2920

*Arie Satia Dharma

Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-Cnn: Towards Real-Time Object Detection With Region Proposal Networks. *Ieee Transactions on Pattern Analysis and Machine Intelligence*, *39*(6), 1137–1149. https://doi.org/10.1109/tpami.2016.2577031

Salam, N., & Jebaseeli, T. J. (2024). A Comparative Study of Pedestrian Detection Techniques Over the Last Decade. *Lecture Notes in Networks and Systems*, *970 LNNS*. https://doi.org/10.1007/978-981-97-2031-6_15

Siahaan, R. D., Pardede, H., Simbolon, I. N., Simbolon, I., & Gultom, D. J. (2024). Enhancing Real-time Herbal Plant Detection in Agricultural Environments with YOLOv8. *Journal of Information Systems and Informatics*, *6*(4). https://doi.org/10.51519/journalisi.v6i4.889

Srikar, M., & K, M. (2022). A Real Time Object Detection in Integral Part of Computer Vision Using Novel Image Classification of Faster R-CNN Algorithm Over Fast R-CNN Algorithm. *PNR*, *13*(SO4). https://doi.org/10.47750/pnr.2022.13.s04.202

Sunario Megawan, & Wulan Sri Lestari. (2020). Deteksi Spoofing Wajah Menggunakan Faster R-CNN dengan Arsitektur Resnet50 pada Video. *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi*, *9*(3). https://doi.org/10.22146/.v9i3.231

Wang, Z., Bao, C., Cao, J., & Hao, Q. (2023). AOGC: Anchor-Free Oriented Object Detection Based on Gaussian Centerness. *Remote Sensing*, *15*(19). https://doi.org/10.3390/rs15194690

Yousif, M. J. (2024). Enhancing The Accuracy of Image Classification Using Deep Learning and Preprocessing Methods. *Artificial Intelligence & Robotics Development Journal*. https://doi.org/10.52098/airdj.2023348

Zhang, S., Chi, C., Yao, Y., Lei, Z., & Li, S. Z. (2020). Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. https://doi.org/10.1109/CVPR42600.2020.00978

Zhang, T., Jin, B., & Jia, W. (2022). An anchor-free object detector based on soften optimized bi-directional FPN. *Computer Vision and Image Understanding*, *218*. https://doi.org/10.1016/j.cviu.2022.103410

Zhou, B., Li, R., Shang, Z., & Li, X. (2020). Object Detection Algorithm Based on Improved Faster R-CNN. *Laser and Optoelectronics Progress*, *57*(10). https://doi.org/10.3788/LOP57.101009

*Arie Satia Dharma