ENHANCING THE EFFECTIVENESS OF THE YOLO MODEL THROUGH CALADIUM LEAF IMAGES GENERATED BY GENERATIVE ADVERSARIAL NETWORKS

Rudy Chandra[1], Tegar Arifin Prasetyo[2*], Akdes Simon Simamora[3], Amanda Artha Regina Simbolon[4], Ester Krismayani Sinaga[5], Lukie Perdanasari[6]
Information Technology, Faculty of Vocational Studies, Institut Teknologi Del, Sumatera Utara, Indonesia[12345]
Department of Information Technology, Politeknik Negeri Jember, Indonesia[6]
arifintegar12@gmail.com

Received: 25 November 2024, Revised: 09 May 2025, Accepted: 25 August 2025
**Corresponding Author*

ABSTRACT
The need for ornamental caladium plants is very popular, but there are several obstacles to recognizing its type. Caladium species classification using AI is needed to overcome the problem of misidentification among enthusiasts. This study uses the Generative Adversarial Network (GAN) algorithm to generate new images from the Caladium dataset: Amazon Caladium, Bicolor Caladium, White Queen Caladium, and Skull Caladium. We combine GAN with YOLOv5 to detect Caladium in real time to improve accuracy. The quality of the generated images is evaluated using the Kernel Inception Distance (KID) method, with the highest scores of 0.2320 for Amazon Caladium, 0.1966 for Bicolor, 0.1713 for Skull, and 0.1857 for White Queen, indicating close similarity to the original images. We chose the best model to generate three datasets: Original Dataset, Mixed Dataset (original images plus GAN-generated images), and Dataset consisting mainly of GAN images. The Mixed Dataset achieved the best results, with a mean Average Precision (mAP) of 0.695 for an Intersection over Union (IoU) of 0.50:0.95 outperforming the GAN dataset and the original Dataset. This training used 50 epochs, a learning rate of 0.0003, and a batch size of 16, to obtain the best model and significantly improve Caladium detection. From this experiment, it was found that the GAN, combined with the original data, was able to support the accuracy of YOLOv5 for real-time caladium classification and was also able to create new images that resembled the original leaves. In the mobile application, this model allows real-time identification of Caladium types, making it easier for users to buy Caladium according to the desired type.
Keywords: *Caladium, GAN, Model, Object Detection, YOLO*

1. Introduction

One of the famous horticultural plants in Indonesia is the ornamental Caladium (Caladium spp.), which is in demand due to its easy maintenance, unique leaf shape and pattern, and diverse colors (Gresinta & Risdiana, 2023). Ornamental plant enthusiasts often have difficulty classifying types of ornamental Caladium because some have similarities in leaf shape, pattern, and color. Recognizing caladium types requires in-depth knowledge of the characteristics of each plant, which is difficult for individuals with limited knowledge, resulting in frequent errors.

The potential of AI in solving various problems, including object recognition, is undeniable(Raees & Al-Tamimi, 2024). In this context, the application of AI , particularly the object detection method(Mishra et al., 2023; Pulungan et al., 2021), holds promise for classifying the type of ornamental Caladium based on the leaves. A single pot of Caladium plants might contain more than one leaf. It's Making multi-class classification inefficient in determining its type. On the other hand, leaves from different plants may overlap with each other so that it can confuse the identification of the Caladium type, so an approach is needed that can classify many leaf objects accurately. Accurate classification is not enough, knowing the location of the leaves and for classification is also necessary so as not to detect other leaves besides Caladium. Object image (with a bounding box), while also classifying the type of object(Liu et al., 2021). The

YOLO framework, known for its quick object detection and high accuracy, is not just an exciting choice for this application, but also a reliable one that instills confidence in the research approach(Abas et al., 2022; Reddy et al., 2024).

You Only Look Once (YOLO) is an algorithm developed to detect an object in real-time directly with a very high level of accuracy (Yunefri et al., 2022; Zanoon et al., 2024). The YOLO framework is used because of its advantage of being able to detect objects quickly (Garcia-Pajuelo & Paiva-Peredo, 2024). mAP stands for mean Average Precision, a commonly used evaluation metric in object detection tasks in computer vision(Xu et al., 2021). In (Abusalim et al., 2024) research, four and seven-class classification problems, the highest mAP value of 87% and 79% was achieved by YOLOv5 respectively. This research also uses a dataset from (Chandra et al., 2024) research that includes four types of Caladiums with similar characteristics (Skull Caladium, Amazon Caladium, Bicolor Caladium, and White Queen Caladium). The previous research implemented the CNN algorithm to classify the Caladium and get 97.5% accuracy.

However, in its application, practicality is needed with real-time object detection and the dataset collected is still relatively small(Jasim & Atia, 2023). In many other areas of artificial intelligence (like car detection, face detection, etc.), we have millions of images. But in horticulture, especially plant phenotyping (analyzing plant traits, like leaf size, health, flower color), data is scarce. The scarcity of data is due to the high cost of labeling, which requires agronomists to label. Generating a good dataset takes a long time because it requires a full growing season to be monitored and high variations such as genetics, environment, and seasons that make the data more diverse. Sometimes there are more images of healthy plants than sick ones (because plants are generally healthy). This makes it difficult for the model to learn to detect rare cases (like minor diseases) (Prasetyo et al., 2024). The consequences of this problem are that limited datasets can affect the YOLO model training process, leading to overfitting, bias in the model, and poor performance on test data during testing. (Li et al., 2023).

Data imbalance is a prevalent issue in diagnostic processes. To address this, the study explores the implementation of Generative Adversarial Networks (GANs) to generate synthetic data based on patterns in small datasets (Jiménez-Gaona et al., 2024; Sowmya et al., 2024). Although there is a connection mentioned earlier between synthesis and analysis, state-of-the-art Generative Adversarial Networks (GANs) are typically trained in an unsupervised manner without utilizing pretrained networks(Kumari et al., 2022). Generative Adversarial Network (GAN) is a type of machine learning architecture that consists of two artificial neural networks that compete in a learning process (Amir et al., 2024). This is useful for increasing data variation and helping to reduce the risk of overfitting (Nugroho et al., 2024).

One of the most common ways to evaluate GANs is by using KID (Kernel Inception Distance), KID has several advantages over FID (Fréchet Inception Distance) in terms of generative model evaluation, especially on small data sets (Darmawan et al., 2024). This method involves a comparison between two sets of images from the GAN and the original image where to perform this comparison, KID utilizes a mathematical function called a kernel with the approach that can be used is a polynomial kernel(Betzalel et al., 2022)

The study's findings, particularly the use of GAN modelling to produce the best dataset, are significant. The generated images, integrated into the YOLO training dataset, pave the way for a comprehensive understanding of the model's performance. The testing metrics will be instrumental in determining the model's best performance.

2. Literature Review

2.1 Caladium

Ornamental caladium plants are one type of horticultural plant that is much loved and has a high selling value in Indonesia. This ornamental plant is popular in Indonesia because of its unique leaf shape and pattern and easy maintenance (Dwitanto & Utami, 2023).

Ornamental caladium plants have 4 types of caladiums that have similarities in leaf shape and pattern, namely Skull Caladium, Amazon Caladium, Bicolor Caladium, and White Queen Caladium. In recognizing these 4 types of caladiums, mistakes often occur, therefore an understanding of the special characteristics of each caladium is needed. Amazon Caladium has green leaves with thick leaf bones. The similarity between the Amazon plant and Skull caladium lies in the similar shape and color of the leaves. Bicolor caladium has a characteristic color variation in the leaf organ, where the green color surrounds the edge of the leaf while the middle is pink. In addition, this type of caladium also has fine green fibers that make this Caladium very beautiful. All the caladiums can be seen in Figure 1 below.

a. Amazon Caladium b. Skull Caladium c. Bicolor Caladium d. White Queen Caladium

Figure 1. All types of Caladiums

2.3 Generative Adversarial Network

The architecture of the GAN model consists of two main components, the generator network (G) and the discriminator network (D) (Berrahal & Azizi, 2022). The G network generates false samples from an input of an arbitrary latent noise vector (a vector containing random values that can be used to generate false samples). Network D then compares the fake samples and the original samples to improve its discrimination ability. Both G and D networks (which are differentiable functions) can be implemented through Multilayer Perception (MLP) which is a type of artificial neural network consisting of multiple layers (Zhao et al., 2020). The learning process of both networks is based on minimax cost where the architecture can be seen in Figure 2.

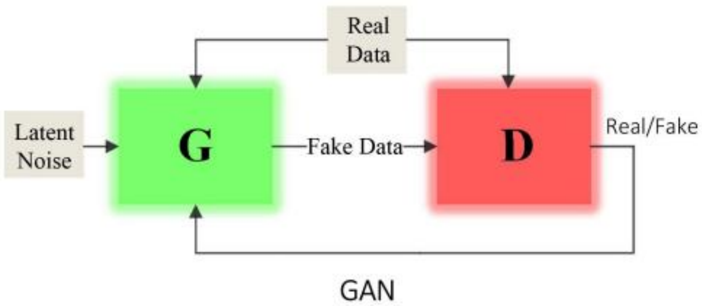


Fig. 2. Architecture of the GAN Model

The optimization process in GAN involves iterating between training the generator to minimize this objective function by improving its ability to produce data that deceives the discriminator and training the discriminator to maximize this objective function by improving its ability to distinguish between real and fake data(Jasim & Atia, 2023). This process continues until an equilibrium is reached where the generator produces data that is very similar to the real data and the discriminator finds it difficult to distinguish between the two(Rguibi et al., 2023).

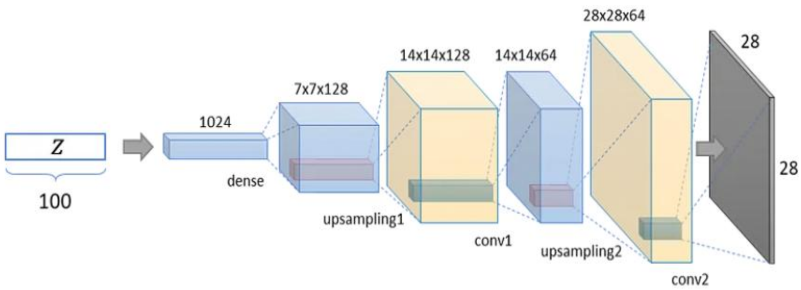


Fig. 3. Architecture of Generator

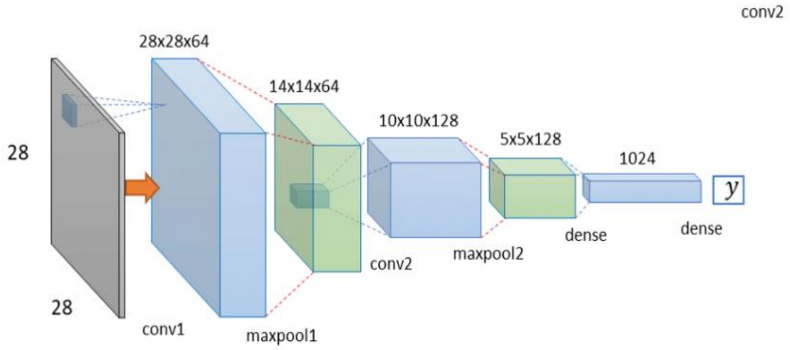


Fig. 4. Architecture of Discriminator

2.4 You Only Look Once (YOLO)

You Only Look Once (YOLO) is a framework for fast and accurate object detection(Mas et al., 2024). YOLO works by processing the whole image at once, then using convolutional neural networks to predict bounding boxes and object classes for each object in the image. By using YOLO, objects can be detected more accurately and efficiently(Gündüz & Işık, 2023). This is because YOLO can learn the relationship between objects in an image, so it can detect objects that overlap or are in proximity(Lim & Kim, 2024). The architecture of YOLO is as shown in Figure 5 below.

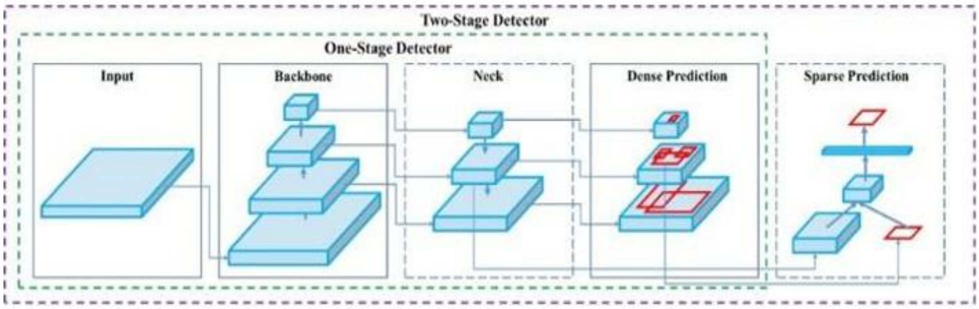


Fig. 5. Architecture of YOLO Object Detection

Fig. 6. Example for Object Detection using YOLO

2.5 Evaluation

Evaluation is used to measure the performance of the algorithm and compare it with other algorithms. This section will discuss the evaluation for each algorithm used in this research.

2.5.1 Kernel Inception Distance

Kernel Inception Distance (KID) is a metric for evaluating the quality of generated images by comparing their feature distributions with real images, similar to Frechet Inception Distance (FID). However, unlike FID—which assumes a Gaussian distribution of features—KID uses a kernel-based approach (specifically, the Maximum Mean Discrepancy, MMD) to measure the discrepancy between real and generated data without restrictive distributional assumptions (Bińkowski et al., 2018; Wang et al., 2023). The KID metric calculates the squared maximum mean discrepancy (MMD) between the Inception features of real and synthetic samples, employing a polynomial kernel. As a non-parametric test, it avoids the rigid Gaussian assumption, relying instead on the kernel's effectiveness as a similar measure. Additionally, it demands fewer samples since estimating the quadratic covariance matrix is unnecessary(Betzalel et al., 2024). Where in KID using a polynomial kernel based on Equation 5 below.

$$k(x,y) = \left(\frac{1}{d}x^T y + 1\right)^3 \quad (1)$$

2.5.2 Evaluation Matrix in YOLO

Performance evaluation can be divided into 2 types, namely detection results and classification performance. The performance evaluation matrix in the context of object detection, to assess the accuracy of the YOLO architecture, often uses Intersection over Union (IoU). IoU is useful for obtaining the error value between the predicted bounding box and the actual bounding box(Cheng et al., 2021; Tran et al., 2023), as shown in Figure 7 below.

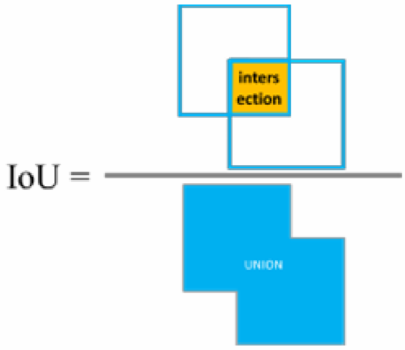


Fig.7. Intersection Over Union

In addition to IoU, there is also an evaluation of the results of the Bounding Box position such as Precision, F1-Score, Recall, mAP@IoU = 0.5 and, mAP@IoU = 0.5:0.95 which can be seen in Equation 6.

- Precision
  Precision is a comparison value to get the result of how often a positive prediction is when the model makes a positive prediction. Precision value can be calculated with Equation 2.
  $$Precision = \frac{True\ Positive(TP)}{True\ Positive(TP) + False\ Positive(FP)} \quad (2)$$
- Recall
  Recall is a description of the successful performance of the model to retrieve information in the form of the success of the model in capturing all positive targets. Recall value can be calculated through Equation 3.
  $$Recall = \frac{True\ Positive(TP)}{True\ Positive(TP) + False\ Negative(FN)} \quad (3)$$
- F1-Score
  F1-Score is calculated to get the average comparison of precision and recall values. To calculate F1-score using Equation 4.
  $$f1 - score = \frac{2 * Recall * Precision}{(Recall + Precision)} \quad (4)$$

Where the Confusion Matrix is a general analysis to represent the performance of the classification. In the classification process there has a condition to determine the value of the object, namely:

- True Positive (TP) if IoU has a value of more than 0.5
- False Positive (FP) if IoU has a value less than 0.5
- False Negative (FN) if the detection performed does not match the trained class and has an IoU value of more than 0.5.

To measure how accurate and complete the YOLO algorithm is in detecting objects, the AP Equation is used in Equation 11 below.

$$AP = \sum_{n=1}^{N} \left(Recalls\ (n) - Recalls\ (n+1)\right) * Precisions\ (n) \quad (5)$$

Then to calculate the average of AP for all object classes the following Equation 12 is used.

$$mAP = \frac{1}{C}\sum_{c=1}^{C} AP_c \quad (6)$$

A common evaluation metric called Mean Average Precision (mAP) gives a single number as the average of the Average Precision (AP) values for all classes. This makes it possible to assess the performance of the model using just one number. As a result, mAP is the most frequently used evaluation metric by object detection algorithms. mAP gives a comprehensive picture of the performance of an object detection model. The higher the mAP (maximum 1.0 or 100%), the more accurate the model. mAP Common Variations:

- mAP@IoU=0.5 → Only consider detections that overlap ≥ 50% with the ground truth.
- mAP@IoU[0.5:0.95] → Average AP at various IoU thresholds from 0.5 to 0.95 (more stringent and realistic).

Then, the results of YOLO evaluation and validation are represented in Precision - Recall curves, and f1-score. In which, the Precision - Recall curve shows the trade-off between precision and recall for various thresholds. The closer to the upper right corner (Precision = 1 and Recall = 1), the better the model(Cook & Ramadas, 2020). The Precision-Recall curve shows the trade-off between precision and recall for various prediction thresholds in the classification model. Therefore, this curve helps in evaluating and finding the optimal balance between precision and recall according to the specific needs of the application. An example image of a precision-recall curve that shows good results is in Figure 8.

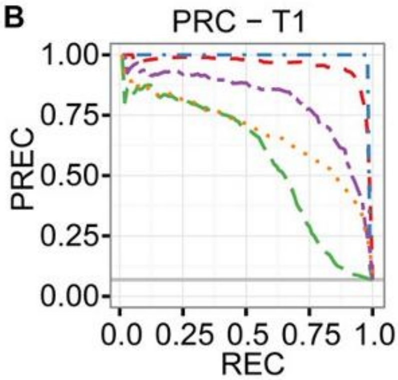


Fig. 8. Precision-Recall Curve

3. Research Methods

This study outlines steps to develop a model for classifying Caladium ornamental plants using GAN and YOLO algorithms. The steps, based on quantitative methods (Kamiri & Mariga, 2021), are shown in Figure 9.

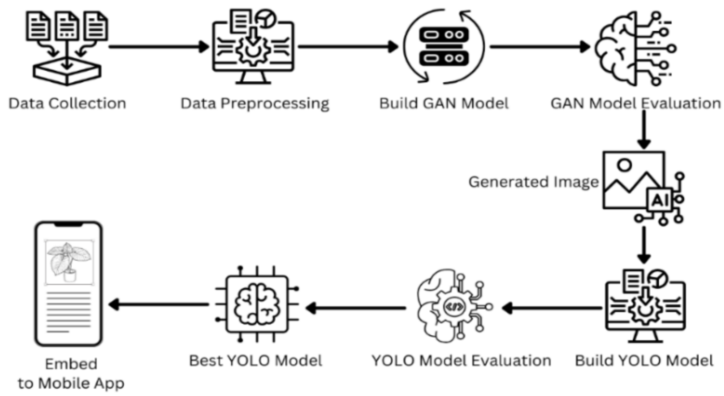


Fig. 9. Research Method

The caladium dataset was obtained from the Kaggle site and then took the type of caladium with the highest intensity of interest. Data was collected in two categories: training data and test data. The training data consists of four varieties of Caladium, namely bicolor, white queen, amazon, and Skull, with 300 images each. Overall, there are 1,200 image data used as the basis of the research.

Preprocessing is an important stage to remove unwanted components from an image, so that the dataset can be processed by the model more effectively. In the GAN Model, preprocessing includes data augmentation, cropping, resizing, and normalization, to optimally prepare the data for further processing. For next step is to build a model using the Generative Adversarial Network (GAN) algorithm to generate new images with similar characteristics to the training data.

After building the GAN model, the next step is to train it. The generator generates a fake image, the discriminator distinguishes between the real and fake images, and the generator is then improved to outwit the discriminator. This process is repeated until it reaches optimal performance.

The GAN model is evaluated using Kernel Inception Distance (KID) to assess the image quality based on the similarity with the original image. KID is an evaluation metric for generative models, specifically Generative Adversarial Networks (GANs). Unlike Fréchet Inception Distance (FID) which assumes a Gaussian distribution on Inception features, KID uses Maximum Mean Discrepancy (MMD) with a polynomial kernel to measure the distance between the feature distributions of real images and generated images. This makes KID non-parametric and unbiased, making it more reliable, especially on small datasets. The polynomial kernel in KID provides a more powerful tool for measuring the distribution distance between the generated images and the original images and helps in making more accurate comparisons, especially when the data being compared has complex patterns (Paik et al., 2023).

KID is more computationally efficient and requires less processing time than FID. KID is also more numerically stable and produces more consistent values. The implementation of KID is easier and does not require complex calculations like FID. The main advantage of KID lies in its performance on small data sets. The expected value of KID does not depend on the number of samples, unlike FID which produces higher expected values on small datasets (Binkowski et al., 2018). The evaluation is performed four times for each GAN model with various parameters, such as epoch, batch size, and learning rate. The best model will be selected based on the KID score. The parameters can be seen in Table 1.

Table 1 - Parameter Combination of GAN

Experiment	Epoch	Learning Rate	Batch Size
1	100	0.00005	2
2	100	0.00005	4
3	100	0.00005	8
4	100	0.00001	2
5	100	0.00001	4
6	100	0.00001	8

Experiment	Epoch	Learning Rate	Batch Size
7	200	0.00005	2
8	200	0.00005	4
9	200	0.00005	8
10	200	0.00001	2
11	200	0.00001	4
12	200	0.00001	8

In the Image Generation step, the best GAN model is used to create new images that are added to the dataset. These images are then tested with the YOLO object detection model. The new dataset will be of three types: the original dataset, a mixture of the original dataset and GAN, and a dataset generated entirely by GAN.

The next process is the image preprocessing stage where the data received by the YOLO model with the YOLOv5 version can be processed effectively by the model, which consists of 3 datasets. Where the three datasets can be seen the number per dataset in Table 2 below.

Table 2 – Three Types of Datasets

Original Dataset	Original and GAN Mixed Dataset	Majority of GAN Dataset
300 original dataset	300 original dataset & 300 GAN Image	420 GAN Image & 180 original dataset

After preprocessing data, the model is built using the YOLOv5 algorithm for object detection. Train dataset with 70% of the training data and the Caladium image dataset. The trained YOLO was then combined with the GAN model to generate synthetic images, increase the variety of datasets, and help YOLO recognize a wide variety of plants.

After training the model with the training data, the next step is to validate the model using 20% validation data. This data helps to understand the patterns of the detected objects and test parameters such as epoch, batch size, and learning rate. This validation aims to optimize the performance of the YOLO model. Parameter combinations for validation can be seen in Table 3.

Table 3 - Parameter Combination of YOLO

Experiment	Epoch	Learning Rate	Batch Size
1	25	0.0003	16
2	25	0.0003	32
3	25	0.00001	16
4	25	0.00001	32
5	25	0.0001	16
6	25	0.0001	32
7	50	0.0003	16
8	50	0.0003	32
9	50	0.00001	16
10	50	0.00001	32
11	50	0.0001	16
12	50	0.0001	32

In this study, two metrics are used that help understand how well the model detects objects, namely mAP 50 and mAP 50-95. mAP 50 measures the average precision at the intersection over union (IoU) threshold of 0.5, while mAP 50-95 measures the average precision at the IoU threshold ranging from 0.5 to 0.95. The mAP itself is used in validating and testing the YOLO model to find the best model to be used for mobile integration (Şirin & Gültekin, 2023).

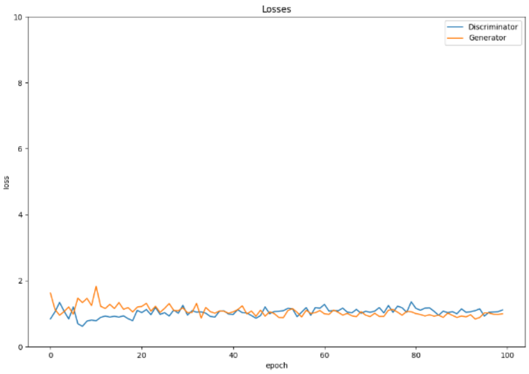
4. Results and Discussion

4.1 GAN Model Evaluation

The KID result will determine the selection of the optimal model to be used in the image generation process, where the best performance is indicated by the lowest KID value of all the experiments conducted. The GAN evaluation graph using the KID score and information related to the KID score value can be found in Table 4, Table 5, Table 6, and Table 7 below.

Table 4 - KID Score in Amazon Caladium

Amazon Caladium			
Epoch	Learning Rate	Batch Size	KID Score
100	0.00005	8	0.2320



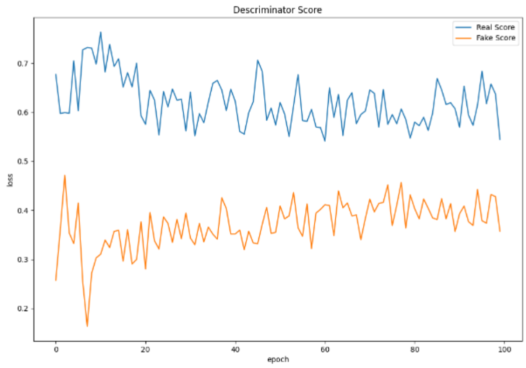
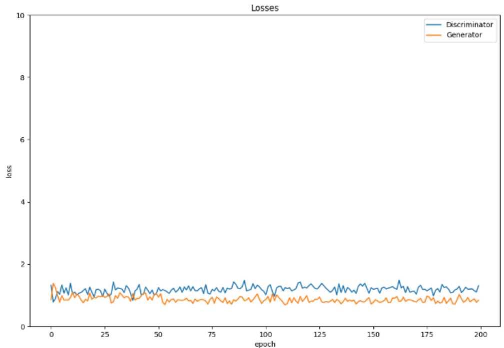
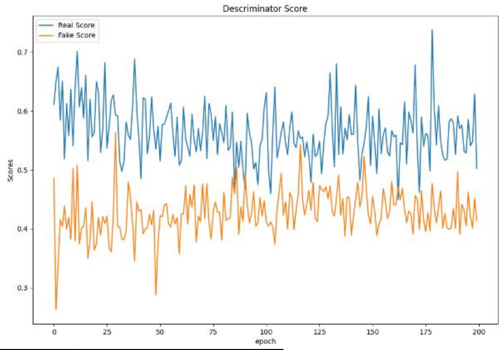


Table 5 - KID Score in Skull Caladium

Skull Caladium			
Epoch	Learning Rate	Batch Size	KID Score
200	0.00005	2	0.1966





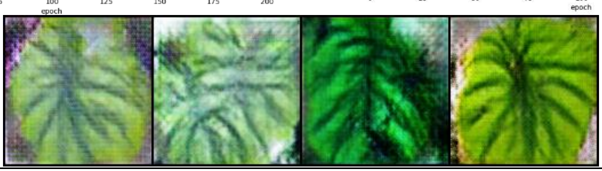


Table 6 - KID Score in Bicolor Caladium

Bicolor Caladium			
Epoch	Learning Rate	Batch Size	KID Score
200	0.00005	4	0.1857

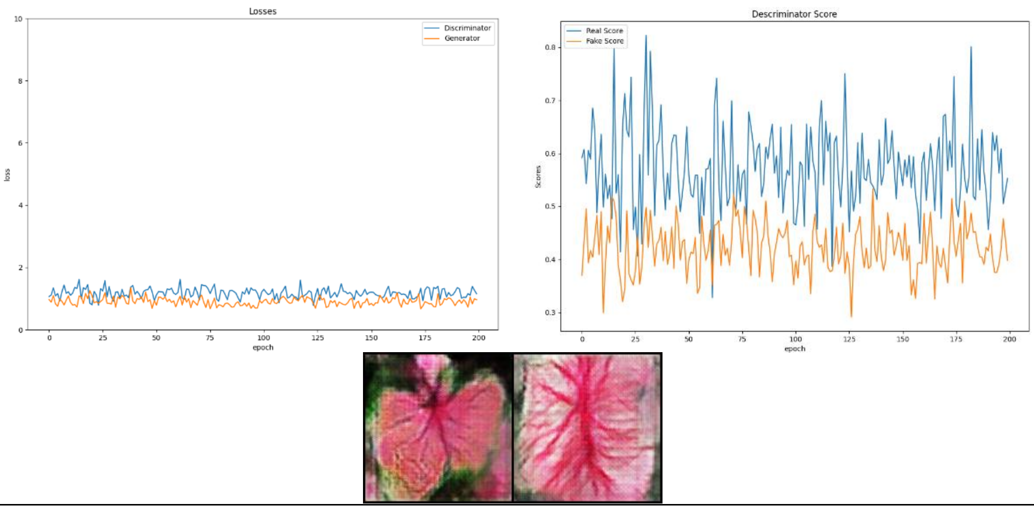
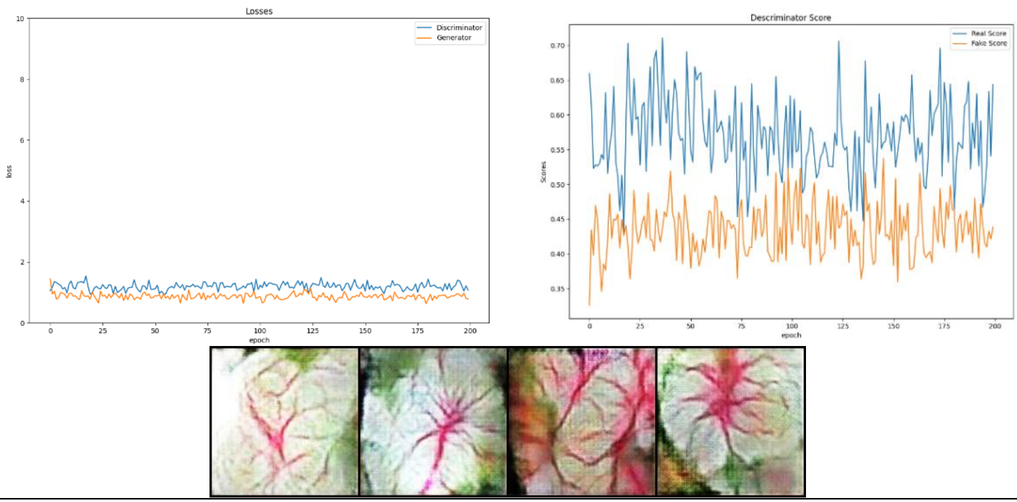


Table 7 - KID Score in White Queen Caladium

White Queen Caladium			
Epoch	Learning Rate	Batch Size	KID Score
200	0.00005	4	0.1713



4.2 YOLO Model Evaluation

After the model is trained using the train data, the next stage is to validate the model using validation data as much as 20% of the total dataset. Validation data is used to understand the patterns and characteristics of the objects to be detected. Validation data is used to experiment with parameter values such as the number of epochs, batch size, and learning rate. Batch size determines the number of data samples processed by the model in one training iteration, while epoch is the number of times the YOLO model sees and processes all training data in one round. Learning rate, as a parameter in model training, is used to calculate weight corrections during the training process. Through the validation and optimization stages of the YOLO model with parameter variations, optimal performance can be achieved. The combination of parameters for the validation stage can be found in Table 8

Table 8 - Parameter Tuning in YOLO Model

Experiment	Epoch	Learning Rate	Batch Size
1	25	0.0003	16
2	25	0.0003	32
3	25	0.00001	16
4	25	0.00001	32
5	25	0.0001	16

Experiment	Epoch	Learning Rate	Batch Size
6	25	0.0001	32
7	50	0.0003	16
8	50	0.0003	32
9	50	0.00001	16
10	50	0.00001	32
11	50	0.0001	16
12	50	0.0001	32

YOLO model validation is carried out after the model has been trained using the available training dataset. In this study, there are three datasets, namely the original dataset, the original and GAN mixed dataset, and the dataset with only GAN training data, where all three datasets are validated on the previously trained YOLO model to obtain the latest evaluation results from the validation data. In choosing the best model, the average of mAP is used, namely mAP 50 and mAP 50 - 95. Where the average of mAP is not an evaluation model but only a basis for determining the best model from the 2 model evaluations obtained. After determining several combinations of epoch parameters, batch size, and learning rate, the three best results can be found in Tables 9 with mAP 50, mAP 50 - 95, and the average of mAP 50 and mAP 50 – 95.

Table 9. Comparison Model Performance of YOLO Model

Dataset Type	mAP50	mAP50-95	mAPavg50and50-95
300 original datasets	0.867	0.509	0.688
300 original dataset & 300 GAN Image	0.861	0.529	0.695
420 GAN Image & 180 original dataset	0.303	0.152	0.2275

The validation results for both the Original Mixture and GAN datasets show the highest average mean Average Precision (mAP) scores of 0.695 for both mAP@50 and mAP@50-90. Among the three models tested, the 300 original dataset and the 300 GAN image dataset models are identified as the best options for achieving optimal results in the mAP metric. Achieving an mAP of 0.695, the mixed dataset confirms conclusions regarding the benefits of hybrid real-synthetic data in improving model resilience. Meanwhile, the pure GAN-generated dataset, with an mAP of only 0.2275, underperformed significantly, suggesting the existence of a threshold beyond which synthetic data no longer provides meaningful utility.

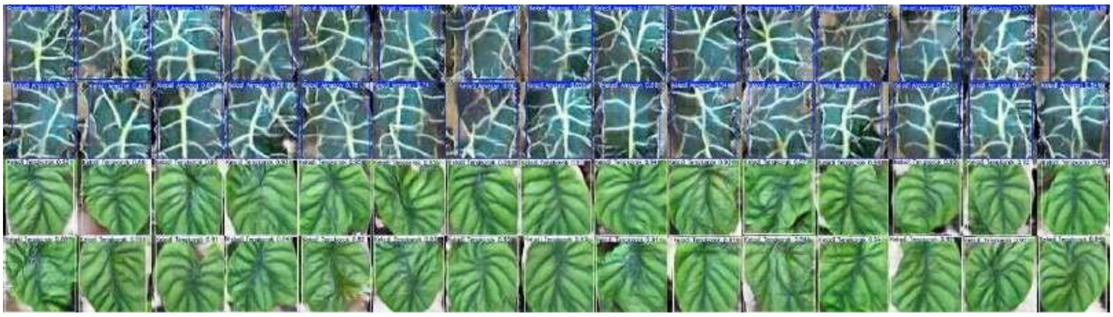
4.3 YOLO Model Testing

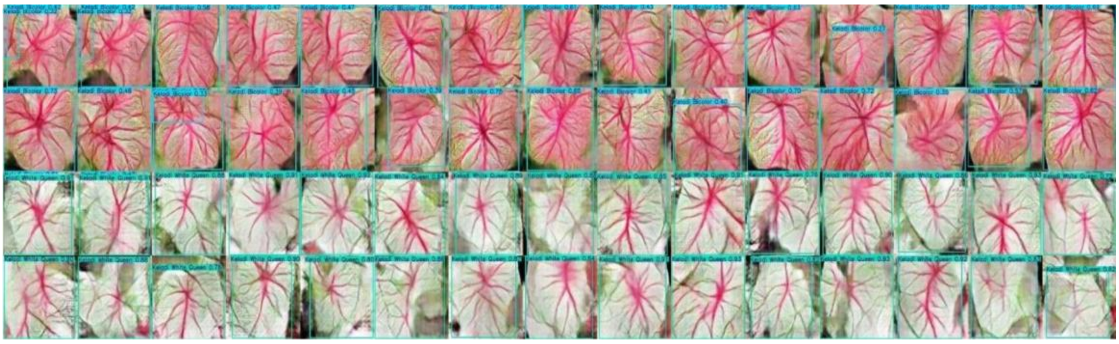
This chapter primarily focuses on developing a mobile-based application for classifying ornamental caladium plants using the YOLO algorithm. Each experiment's results are evaluated using loss and mAP metrics to assess the performance of the trained models. The test results carried out on the YOLO model with 60 data per class each show how well the model makes predictions. Table 10 provides an overview of the number of data that are predicted correctly and incorrectly in each class.

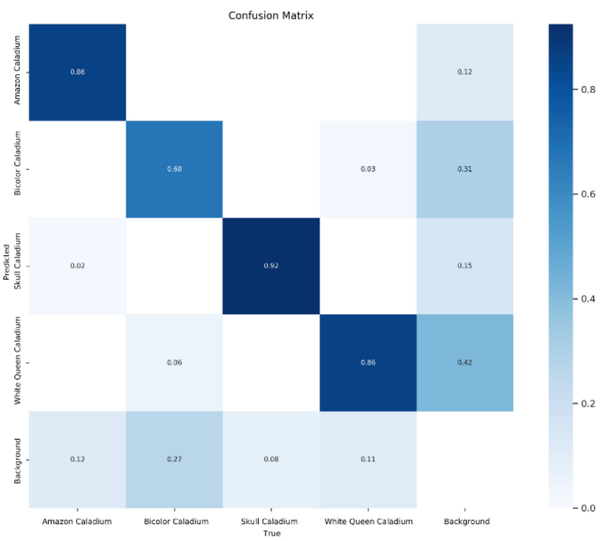
Table 10 - Result of YOLO Model Testing

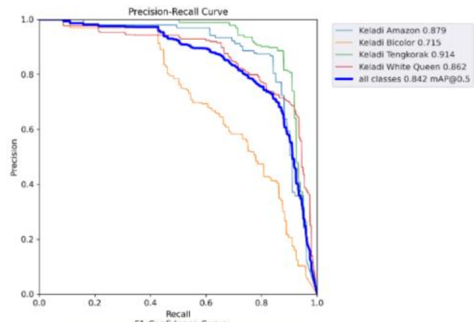
Evaluation Results of the Best Model Testing

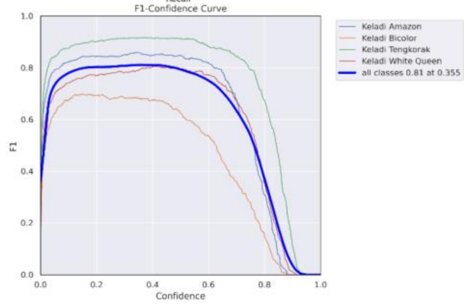
```
val: Scanning /content/drive/MyDrive/Colab Notebooks/Train YOLO/Dataset_WithG/valid/labels.cache... 480 images, 0 backgrounds, 0 corrupt: 100% 480/480 [00:00<?, ?it/s
                 Class     Images  Instances          P          R      mAP50   mAP50-95: 100% 15/15 [00:08<00:00,  1.77it/s]
                   all        480       1257      0.823      0.809      0.861      0.529
         Keladi Amazon        480        217      0.853      0.853      0.884      0.547
        Keladi Bicolor        480        355      0.828      0.635      0.768      0.441
      Keladi Tengkorak        480        318      0.863      0.931      0.946      0.622
    Keladi White Queen        480        367      0.746      0.817      0.847      0.506
Speed: 0.2ms pre-process, 8.7ms inference, 2.3ms NMS per image at shape (32, 3, 416, 416)
Results saved to runs/val/exp7
```











Precision-Recall Curve graph for Amazon Caladium with a precision of 0.879 and recall of 0.76; Bicolor Caladium has a precision of 0.715 and recall of 0.67; Skull Caladium has a precision of 0.514 and recall of 0.52; and White Queen Caladium has a precision of 0.892 and recall of 0.85. The curves in the graphs show the balance between precision and recall for each type of Caladium, where higher values indicate better performance in detecting and classifying objects. Overall, the test results show that the YOLOv5 model can detect and classify caladium plants well

4.4 Comparison with Prior Research

This section will make a comparison with previous research that is used as a reference and development in this research. Comparison using the YOLOv5 algorithm and generative image results from GAN

Research entitled “Optimizing YOLOv5s Object Detection through Knowledge Distillation algorithm” conducted by (Du et al., 2024). They used the COCO dataset with 200,000 images and 80 categories. Testing the YOLOv5 model by measuring mAP50 and mAP50:95. The results obtained at the beginning of modeling without Distillation temperature. Summary of comparison results can be seen in table 11.

Table 11 - Comparison of this study with Du's research

No	Dataset	mAP50 (%)	mAP50:95 (%)
1	COCO dataset with 200,000 images and 80 categories	91.33	67.86
2	Caladium Plant with 300 original dataset + 300 GAN Image and 4 categories	86.1	52.9

From the comparison in Table 11, it can be concluded that large datasets that do not have synthetic data have a higher chance of mAP. This is proven by Du's research with higher results compared to this study which uses a limited dataset and adds synthetic data from GAN. The similarity between these 2 studies is that mAP50:95 is smaller than mAP50 because mAP50:95 includes a tighter threshold that is more difficult to achieve. So, the model must be more precise in its predictions to still be considered correct

The study entitled "Defect Identification of Adhesive Structure Based on DCGAN and YOLOv5" conducted by (Jin et al., 2022). They used test samples of X-ray datasets of multi-layer metal and non-metal bonded tubular specimens in the past two years. A total of 223 original images were acquired, and 442 enhanced images were reutilized in this experiment following augmentation using the improved DCGAN. Testing the YOLOv5 model by measuring mAP50, precision and recall. A summary of the comparison results can be seen in table 12.

Table 12. Comparison of this study with Jin's research

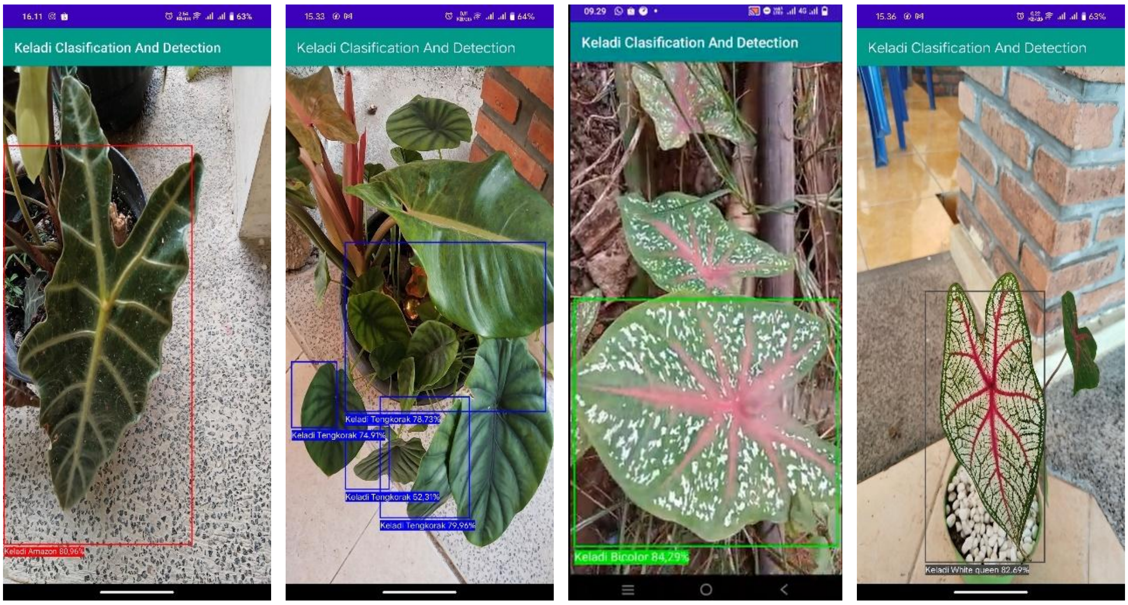
No	Dataset	mAP50	Precision	Recall
1	X-ray datasets of multi-layer metal and non-metal bonded tubular specimens with 442 enhanced images by DCGAN and 3 categories	0.701	0.745	0.691
2	Caladium Plant with 300 original dataset + 300 GAN Image and 4 categories	0.861	0.823	0.809

The object detection model showed higher performance on the Caladium Plant dataset, which consists of a combination of 300 original images and 300 GAN-augmented images, with mAP@50 results of 0.861, precision of 0.823, and recall of 0.809. Meanwhile, the performance on the X-ray dataset with 442 DCGAN-augmented images and 3 categories produced mAP@50 of 0.701, precision of 0.745, and recall of 0.691. This indicates that the model is more effective in recognizing objects on the Caladium dataset, which is likely influenced by more consistent data characteristics or category complexity that is more in line with the model's capabilities.

X-ray images typically have low contrast, faint textures, and variations in internal structure that are more difficult for models to recognize. Object detection performance is heavily influenced by the visual characteristics of the dataset, the balance between real and synthetic data, and the augmentation quality of the GAN.

4.5 Application Interface

This section explains the results of the interface that has been built to help classify mobile-based Caladium ornamental plants using the YOLO model. After the caladium plant is detected, a box is displayed according to the color of the detected Caladium and the level of model conformity (accuracy) with the detected Caladium. The display of the object detection in the application can be seen in Figure 10.



a. Amazon b. Skull c. Bicolor d. White Queen

Fig. 10. Implementation of YOLO in Mobile Object Detection

5. Conclusion

The GAN algorithm utilizes four specialized models tailored for different types of caladiums. KID (Kernel Inception Distance) evaluations revealed strong performance, with the Amazon caladium scoring 0.2320, Bicolor caladium at 0.1966, Skull caladium at 0.1713, and White Queen caladium at 0.1857. These top models will create new images for an enhanced mixed dataset, combining the original data, GAN-generated images, and a dataset of exclusive GAN training data, laying a strong foundation for future research and applications. Evaluation uses two new datasets, a blend of the original dataset and the GAN dataset, to compare the performance of the YOLOv5 model. The comparison data is obtained from these three datasets, with each test conducted under the same parameters. Notably, the mixed GAN dataset with the original data yields the most promising results score 0.695, with an average of the two evaluations, mAP50 and mAP50:90. YOLO has been able to detect objects and classify results with high accuracy through best modelling. This research has succeeded in improving the best YOLOv5 model on mobile applications in real-time classification. For future research, we will expand to rare Caladium variants or 3D leaf modeling to complement the convenience of users who want to know more detailed types of caladiums.

References

Abas, S. M., Abdulazeez, A. M., & Zeebaree, D. Q. (2022). A YOLO and convolutional neural network for the detection and classification of leukocytes in leukemia. *Indonesian Journal of Electrical Engineering and Computer Science*, *25*(1), 200–213. https://doi.org/10.11591/ijeecs.v25.i1.pp200-213

Abusalim, S., Zakaria, N., Maqsood, A., Saboor, A., Yew, K. H., Mokhtar, N., & Abdulkadir, S. J. (2024). Multi-granularity tooth analysis via YOLO-based object detection models for effective tooth detection and classification. *IAES International Journal of Artificial Intelligence*, *13*(2), 2081–2092. https://doi.org/10.11591/ijai.v13.i2.pp2081-2092

Amir, I., Suhaimi, H., Mohamad, R., Abdullah, E., & Pu, C. H. (2024). Hybrid encryption based on a generative adversarial network. *Indonesian Journal of Electrical Engineering and Computer Science*, *35*(2), 971–978. https://doi.org/10.11591/ijeecs.v35.i2.pp971-978

.

Berrahal, M., & Azizi, M. (2022). Optimal text-to-image synthesis model for generating portrait images using generative adversarial network techniques. *Indonesian Journal of Electrical Engineering and Computer Science*, *25*(2), 972–979. https://doi.org/10.11591/ijeecs.v25.i2.pp972-979

Betzalel, E., Penso, C., & Fetaya, E. (2024). Evaluation Metrics for Generative Models: An Empirical Study. *Machine Learning and Knowledge Extraction*, *6*(3), 1531–1544. https://doi.org/10.3390/make6030073

Betzalel, E., Penso, C., Navon, A., & Fetaya, E. (2022). A study on the evaluation of generative models. *arXiv preprint arXiv:2206.10935*. http://arxiv.org/abs/2206.10935

Binkowski, M., Sutherland, D. J., Arbel, M., & Gretton, A. (2018). Demystifying MMD GANs. *6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings*, *2017*, 1–36.

Chandra, R., Arifin Prasetyo, T., Lumbangaol, H. E., Siahaan, V., & Sianipar, J. I. (2024). Development of a Mobile-Based Application for Classifying Caladium Plants Using the CNN Algorithm. *Jurnal Online Informatika*, *9*(1), 111–118. https://doi.org/10.15575/join.v9i1.1296

Cheng, B., Girshick, R., Dollár, P., Berg, A. C., & Kirillov, A. (2021). Boundary IoU: Improving object-centric image segmentation evaluation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 15334-15342).

Cook, J., & Ramadas, V. (2020). When to consult precision-recall curves. *Stata Journal*, *20*(1), 131–148. https://doi.org/10.1177/1536867X20909693

Darmawana, H., Yulianaa, M., Hadia, M. Z. S., & Sudibyoa, R. W. (2024). Fine-tuning GAN Models with Unpaired Aerial Images for RGB to NDVI Translation in Vegetation Index Estimation. *International Journal on Advanced Science, Engineering & Information Technology, 14*(5).

Huang, G., Shen, A., Hu, Y., Du, J., Hu, J., & Liang, Y. (2024). Optimizing yolov5s object detection through knowledge distillation algorithm. *arXiv preprint arXiv:2410.12259*. https://doi.org/10.48550/arXiv.2410.12259

Dwitanto, M. B. F., & Utami, D. (2023). Pola Perilaku Penggemar Tanaman Hias Di Masa Pandemi Covid-19. *Paradigma, 12*(2), 121–130. Retrieved from https://ejournal.unesa.ac.id/index.php/paradigma/article/view/54763

Garcia-Pajuelo, J., & Paiva-Peredo, E. (2024). Comparison and evaluation of YOLO models for vehicle detection on bicycle paths. *IAES International Journal of Artificial Intelligence*, *13*(3), 3634–3643. https://doi.org/10.11591/ijai.v13.i3.pp3634-3643

Gresinta, E., & Risdiana, A. (2023). Identifikasi Penyakit pada Tanaman Keladi Hias (Caladium spp.) dengan Pemanfaatan Sistem Pakar. *EduBiologia: Biological Science and Education Journal*, *3*(2), 131. https://doi.org/10.30998/edubiologia.v3i2.19328

Gündüz, M. Ş., & Işık, G. (2023). A new YOLO-based method for real-time crowd detection from video and performance analysis of YOLO models. *Journal of Real-Time Image Processing*, *20*(1). https://doi.org/10.1007/s11554-023-01276-w

Jasim, R. M., & Atia, T. S. (2023). An evolutionary-convolutional neural network for fake image detection. *Indonesian Journal of Electrical Engineering and Computer Science*, *29*(3), 1657–1667. https://doi.org/10.11591/ijeecs.v29.i3.pp1657-1667

Jiménez-Gaona, Y., Carrión-Figueroa, D., Lakshminarayanan, V., & José Rodríguez-Álvarez, M. (2024). Gan-based data augmentation to improve breast ultrasound and mammography mass classification. *Biomedical Signal Processing and Control*, *94*. https://doi.org/10.1016/j.bspc.2024.106255

Jin, Y., Gao, H., Fan, X., Khan, H., & Chen, Y. (2022). Defect Identification of Adhesive Structure Based on DCGAN and YOLOv5. *IEEE Access*, *10*, 79913–79924. https://doi.org/10.1109/ACCESS.2022.3193775

Kamiri, J. (2023). Research Methods in Machine Learning: A Content Analysis. International *journal of computer and information technology*.

Kumari, N., Zhang, R., Shechtman, E., & Zhu, J. Y. (2022). Ensembling off-the-shelf models for gan training. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 10651-10662).

Li, Q., Li, Y., Duan, H., Kang, J., Zhang, J., Gan, X., & Xu, R. (2023). Improvement and Enhancement of YOLOv5 Small Target Recognition Based on Multi-module Optimization. *arXiv preprint arXiv:2310.01806*. https://doi.org/10.48550/arXiv.2310.01806

Lim, S., & Kim, J. (2024). Advanced Object Tracking through Conditional Online Updates and Noise Suppression. *International Journal on Advanced Science, Engineering & Information Technology*, *14*(5).

Liu, Y., Sun, P., Wergeles, N., & Shang, Y. (2021). A survey and performance evaluation of deep learning methods for small object detection. *Expert Systems with Applications*, *172*, 114602. https://doi.org/10.1016/J.ESWA.2021.114602

Mas, M. S., Saidah, S., & Ibrahim, N. (2024). Detection and counting of wheat ear using YOLOv8. *International Journal of Electrical and Computer Engineering*, *14*(5), 5813–5823. https://doi.org/10.11591/ijece.v14i5.pp5813-5823

Mishra, D. P., Rout, K. K., Mishra, S., & Salkuti, S. R. (2023). Various object detection algorithms and their comparison. *Indonesian Journal of Electrical Engineering and Computer Science*, *29*(1), 330–338. https://doi.org/10.11591/ijeecs.v29.i1.pp330-338

Nugroho, K., Hadiono, K., Sutanto, F. A., Marutho, D., & Farooq, O. (2024). Enhanced multi-ethnic speech recognition using pitch shifting generative adversarial networks. *IAES International Journal of Artificial Intelligence*, *13*(3), 2904–2911. https://doi.org/10.11591/ijai.v13.i3.pp2904-2911

Paik, S., Celentano, M., Green, A., & Tibshirani, R. J. (2023). Maximum mean discrepancy meets neural networks: The radon-kolmogorov-smirnov test. *arXiv preprint arXiv:2309.02422*.

Prasetyo, T. A., Gaol, T. L., Sipahutar, N. F., Siahaan, T., Manik, T. E., & Chandra, R. (2024). Refining tomato disease recognition: hyperparameter tuning on ResNet-101 architecture for precise leaf-based classification. *Indonesian Journal of Electrical Engineering and Computer Science*, *34*(2), 1204–1213. https://doi.org/10.11591/ijeecs.v34.i2.pp1204-1213

Pulungan, A. B., Nafis, Z. ., Anwar, M., Hastuti, H., Hamdani, H., & Myori, D. E. M. (2021). Object Detection with a Webcam Using the Python Programming Language. *Journal of Applied Engineering and Technological Science (JAETS)*, *2*(2), 103–111. https://doi.org/10.37385/jaets.v2i2.247

Raees, S., & Al-Tamimi, M. (2024). The Role of Artificial Intelligence in Providing People With Privacy: Survey. *Journal of Applied Engineering and Technological Science (JAETS)*, *5*(2), 813–829. https://doi.org/10.37385/jaets.v5i2.4013

Reddy, S. S., Rao, V. R. M., Voosala, P., & Nrusimhadri, S. (2024). You only look once model-based object identification in computer vision. *IAES International Journal of Artificial Intelligence*, *13*(1), 827–838. https://doi.org/10.11591/ijai.v13.i1.pp827-838

Rguibi, Z., Hajami, A., Zitouni, D., Maleh, Y., & Elqaraoui, A. (2023). Medical variational autoencoder and generative adversarial network for medical imaging. *Indonesian Journal of Electrical Engineering and Computer Science*, *32*(1), 494–505. https://doi.org/10.11591/ijeecs.v32.i1.pp494-505

Şirin, M., & Gültekin, G. (2023). ORIGINAL RESEARCH PAPER A new YOLO - based method for real - time crowd detection from video and performance analysis of YOLO models. *Journal of Real-Time Image Processing*, *20*(1), 1–12. https://doi.org/10.1007/s11554-023-01276-w

Sowmya, B. J., Meeradevi, & Shedole, S. (2024). Generative adversarial networks with attentional multimodal for human face synthesis. *Indonesian Journal of Electrical Engineering and Computer Science*, *33*(2), 1205–1215. https://doi.org/10.11591/ijeecs.v33.i2.pp1205-1215

Tran, N., Nguyen, M., Le, T., Huynh, T., Nguyen, T., & Nguyen, T. (2023). Exploring the potential of skeleton and machine learning in classroom cheating detection. *Indonesian Journal of Electrical Engineering and Computer Science*, *32*(3), 1533–1544. https://doi.org/10.11591/IJEECS.V32.I3.PP1533-1544

Wang, Z., Farnia, F., Lin, Z., Shen, Y., & Yu, B. (2023). *On the Distributed Evaluation of Generative Models*. http://arxiv.org/abs/2310.11714

Xu, Q., Zhu, Z., Ge, H., Zhang, Z., & Zang, X. (2021). Effective Face Detector Based on YOLOv5 and Superresolution Reconstruction. *Computational and Mathematical Methods in Medicine*, *2021*. https://doi.org/10.1155/2021/7748350

Yunefri, Y., Sutejo, S., Fadrial, Y. E., Anggraini, K., Ramadhani, M., & Hardianto, R. (2022). Implementation of Object Detection With You Only Look Once Algorithm in Limited Face-To-Face Times in Pandemic . *Journal of Applied Engineering and Technological Science (JAETS)*, *4*(1), 400–404. https://doi.org/10.37385/jaets.v4i1.1161

Zanoon, N. I., Alhaj, A. A., & Alkharabsheh, K. (2024). An artificial intelligence approach to smart exam supervision using YOLO v5 and siamese network. *IAES International Journal of Artificial Intelligence*, *13*(4), 3920–3929. https://doi.org/10.11591/ijai.v13.i4.pp3920-3929

Zhao, Y., Li, C., Yu, P., Gao, J., & Chen, C. (2020). Feature quantization improves gan training. *arXiv preprint arXiv:2004.02088*.