ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

IDENTIFICATION OF PAPAYA VARIETIES IN COMPUTER VISION AND DEEP LEARNING APPROACHES

SHARFINA FAZA[1], ROMI FADILLAH RAHMAT[2*], MERYATUL HUSNA[1], AJULIO PADLY SEMBIRING[1], ANJAS SUBHANUARI[2], FARHAD NADI[3]

[1] Department of Computer Engineering and Informatics, Politeknik Negeri Medan, Indonesia

[2] Department of Information Technology, Universitas Sumatera Utara, Indonesia

[3] School of Information Technology, UNITAR International University, Malaysia

*Corresponding E-mail: romi.fadillah@usu.ac.id

ABSTRACT

Papaya is one of the most popular tropical fruits and is widely cultivated in Indonesia. Papaya plants are easily found in various regions due to their good adaptation to the tropical climate. In addition to being delicious for consumption, papaya also offers many health benefits because of its high fiber content and benefits for the digestive system. There are various popular papaya varieties that are frequently consumed by the Indonesian people. However, the similarity in shape and skin color among these papaya varieties often makes it difficult to distinguish them, especially for laypeople. This poses a challenge in identifying papaya varieties, both for daily consumption purposes and in the marketing process. This research aims to develop a prototype that can identify papaya varieties using computer vision technology with the R-FCN ResNet101 algorithm. The data collection for this research was conducted through direct surveys at several markets that sell various papaya varieties. The data collected was in the form of 1000 papaya fruit images captured using a mobile phone camera. The collected data was then divided into training and testing datasets. The first step was image preprocessing, which consisted of resizing the images, labeling the training data, and converting it into TensorFlow Record files. The next step involved processing the data using computer vision technology with the R-FCN ResNet101 algorithm. The result of this step was a learned model, which was then used to analyze and identify the papaya varieties in the testing data. The test results achieved an accuracy of 97.5%.

Keywords: *Papaya, Varieties, Identify, Computer Vision, R-FCN ResNet101.*

1. INTRODUCTION

Papaya is one of the most popular tropical fruits and is widely cultivated in Indonesia. Papaya plants are easily found in various regions due to their good adaptation to the tropical climate. Papaya fruit is known for its sweet taste and soft flesh texture, making it a favorite among Indonesians. Additionally, papaya is rich in fiber, vitamins, and minerals, providing various health benefits, including aiding digestion and boosting immunity.

There are various papaya varieties that are commonly consumed and cultivated, such as Cibinong, California, Bangkok, Solo, Jinggo, Paris, Black Round, and Bird papaya. The Cibinong, California, and Bangkok varieties have reddish-orange flesh and an exceptionally sweet taste. These varieties also share a similar elongated shape with dark green skin. However, the similarity in shape and skin color among these papaya varieties often makes it difficult to distinguish them, especially for the general public. This poses a challenge in detecting papaya varieties, both for daily consumption purposes and in the marketing process.

Some related research on variety identification includes a study titled 'Identification of apricot varieties using leaf characteristics and KNN classifier'. This research identified 10 apricot varieties, and the data collected was leaf images with different characteristics. The method used in this research was the K-Nearest Neighbor (KNN) algorithm, achieving an accuracy of 79% [1]. Another study titled 'Analysis of the Bayes Method in Identifying Rambutan Fruit Varieties' used the Bayes method to identify rambutan fruit varieties. This research involved 7 varieties and tested 10 data samples, achieving 100% accuracy [2].

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Furthermore, a study titled 'Identification of Rice Varieties Using Machine Learning Algorithms' identified 5 rice varieties using various machine learning algorithms. This research used data on the morphology, color, and shape of rice grains. The best accuracy of 97.99% was achieved using the Random Forest algorithm [3]. Additionally, a study titled 'Computer-vision classification of corn seed varieties using deep convolutional neural network' identified corn seed varieties using the DCNN method, resulting in an accuracy of 98.1% [4].

Computer vision is a field that deals with enabling machines to interpret and understand visual data from the real world. It involves developing algorithms and techniques to automatically extract meaningful information from images, videos, and other visual inputs [5]. By recognizing patterns, detecting objects, and analyzing scenes, computer vision systems can perform tasks such as object recognition, image classification, and scene understanding [6]. The R-FCN ResNet101 algorithm is a deep learning model that combines the Residual Network (ResNet) architecture with the Region-based Fully Convolutional Network (R-FCN) approach [7]. This algorithm is particularly effective for object detection and recognition tasks, making it suitable for identifying different papaya varieties based on their visual characteristics.

The similarity in shape among various papaya varieties in Indonesia often makes them difficult to distinguish, especially for the general public. Therefore, a specific approach is needed to identify papaya varieties both for daily consumption purposes and in the marketing process. Based on several previous studies mentioned above, the author will conduct identification of papaya varieties using computer vision technology with the R-FCN ResNet101 algorithm. The proposed research is expected to identify papaya varieties using computer vision technology with high accuracy.

2. METHODOLOGY

The data used is images of papaya fruits captured using a smartphone camera. These images are then used as the main input to the system for processing. The papaya fruit image files were obtained from the Lau Cih Main Market in Medan Tuntungan. There are three papaya varieties captured in the images, namely California Papaya, Bangkok Papaya, and Cibinong Papaya. The details of the number of papaya fruits captured are presented in Table 1.

Table 1. Distribution of Training and Testing Data

No.	Varieties	Amount of data	
		Train Data	Test Data
1	California Papaya	280	100
2	Bangkok Papaya	245	85
3	Cibinong Papaya	350	140
	Total	875	325

The methodology in this research can be seen in Figure 1.

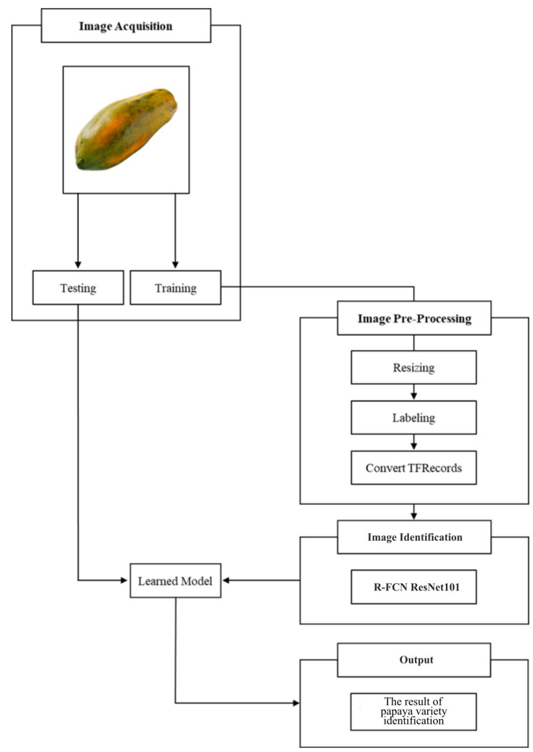


Figure 1. General Architecture

After the image files are combined for system training, they will then go through pre-processing before being entered and processed by the system. The pre-processing will be divided into three processes, which are detailed in the following points. The first step in the pre-processing stage is to resize the image files to a smaller size so that the system can run efficiently. In this case, the images are resized from 4000x4000 pixels to 800x800 pixels. From the initial image capture, a 1:1 aspect ratio was used, so that when the image size was changed, it did not become wider or longer. In carrying out this pre-processing, first, several image files are combined in a separate folder, and then they are collected again in another folder as resized files. Next, in this third stage, classification is carried out on the image files that have gone through pre-processing. The classification process utilizes the R-FCN ResNet101 method. The R-FCN method plays the role of a model that detects image objects based on regions [8], while ResNet101 takes the role of a layer that performs calculations and considers whether it is the object in question or not.

2.1 Image Acquisition

Image acquisition is the process of obtaining images from an external source. In this research, the data used is images of papaya fruits. The image data is captured using a smartphone camera in .jpeg format. Subsequently, the data will be divided into two sets: for training and testing. The division of training and testing data can be seen in Table 1.

2.2 Image Pre-processing

At this stage, resizing, labeling, and converting to TensorFlow Record will be performed. Resizing is the process of changing the size of an image, either by enlarging (upscaling) or reducing (downscaling) the image dimensions, without altering the original format [9]. In this stage, the image size will be adjusted to 800x800 pixels. Next is the labeling stage, which involves naming the objects so that the system can understand the existing objects [10]. This naming process will represent the original objects in a specific file format. The labeling process will be carried out using software called LabelImg. By using this software, naming the objects in the papaya fruit image files can be done easily. The output of the labeling process is a file in XML format. Subsequently, the process of converting to TensorFlow Record files will be performed. The purpose of this conversion is to reduce storage usage on the system, allowing it to run faster and save storage space [11].

filename,width,height,class,xmin,ymin,xmax,ymax
BANGKOK (1).jpg,800,800,Bangkok,294,38,560,744
BANGKOK (17).jpg,800,800,Bangkok,226,132,565,745
BANGKOK (19).jpg,800,800,Bangkok,195,112,558,748
BANGKOK (22).jpg,800,800,Bangkok,178,66,531,732
BANGKOK (25).jpg,800,800,Bangkok,18,256,785,531
BANGKOK (28).jpg,800,800,Bangkok,142,265,702,527
BANGKOK (9).jpg,800,800,Bangkok,276,76,510,764
CALIFORNIA (1).jpg,800,800,California,246,103,539,748
CALIFORNIA (101).jpg,800,800,California,252,82,602,749
CALIFORNIA (107).jpg,800,800,California,249,118,523,765
CALIFORNIA (109).jpg,800,800,California,237,28,555,789
CALIFORNIA (11).jpg,800,800,California,237,31,554,722
CALIFORNIA (115).jpg,800,800,California,70,210,731,569
CALIFORNIA (118).jpg,800,800,California,218,85,575,768

Figure 2. Converted TFRecord File Result

Originally, the named files in XML format were separate from the image files in JPG format. This could lead to increased storage usage and slower system performance. Therefore, after conversion, the result is detailed in Figure 3 and 4, where the image and annotation data are combined into a single TFRecord file format for efficient storage and processing.

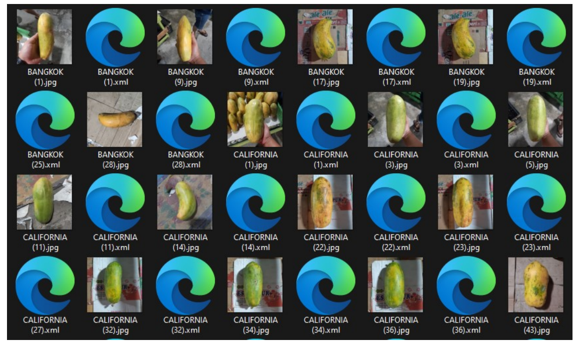


Figure 3. File before converting to TFRecord

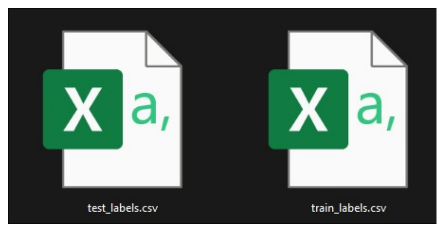


Figure 4. File after being converted to TFRecord

2.3 Image Identification

The identification process utilizes a computer vision algorithm, specifically R-FCN ResNet101.

2.3.1. Region-based Fully Convolutional Network (R-FCN)

The Region-based Fully Convolutional Network (R-FCN) is an advanced computer vision algorithm designed for the task of object detection in digital images. It builds upon the principles of the Fully Convolutional Network (FCN) architecture, which was originally developed for semantic segmentation tasks [12]. While FCN applies convolutional filters across the entire image to make dense predictions, R-FCN introduces a region-based approach. Instead of processing the entire image at once, it focuses on proposed regions of interest that are likely to contain objects. This region-based methodology allows the network to concentrate its computational resources on the most relevant areas, improving efficiency and performance.

By obtaining the output from the FCN, R-FCN adds a position-sensitive Region of Interest (RoI) layer.R-FCN also adopts object detection methods such as the Region Proposal Network (RPN) and

Journal of Theoretical and Applied Information Technology
31st August 2024. Vol.102. No. 16

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Region Classification. The architecture of region-based Fully Convolutional Network can be seen in Figure 5

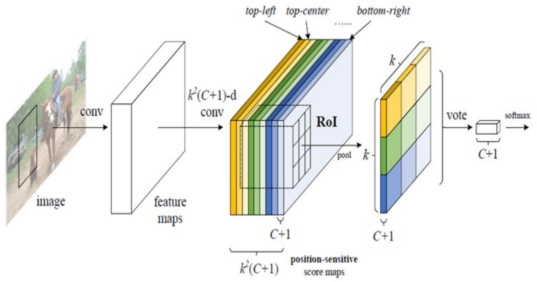


Figure 5. Region-based Fully Convolutional Network Architecture

2.3.2. Region Proposal Network (RPN)

Region Proposal Network (RPN) is a network that serves as the basis for the region-based method utilized for object detection in image research [13]. The Region Proposal Network (RPN) operates by analyzing proposed regions or areas within an image. It performs computations on these proposed regions to determine the likelihood that each one contains an object of interest or just background. The result of these calculations is a value representing the probability of an object's presence for each proposed region.

This probability value is then used to generate a bounding box around the region, essentially "packaging" the possibility of an object's existence into a rectangular box annotation on the image. However, RPN doesn't just blindly propose regions. It utilizes anchor boxes at different scales and aspect ratios to guide the region proposal process.

These anchors act as references to help RPN decide if an object is likely to be present in a particular area of the image. If a proposed region doesn't align well with any of the pre-defined anchors, RPN will discard it. Instead, RPN only retains and processes the proposed region that has the highest intersection over union (IoU) value with one of the anchor boxes within an object's area [14].

This anchor-based approach allows RPN to focus its computational resources on the most promising regions that match common object sizes and shapes, improving overall object detection accuracy and efficiency. RPN will ignore cross-boundary anchors and apply Non-max Suppression (NMS) to obtain one anchor with the highest Intersection of Union value on one of the objects [15].

RPN is the network to identify targets at the pixel level, and the input is a convolution feature map which has been highly abstract, not needing a large number of parameters used for feature extraction [16]. RPN uses sliding windows to generate multiple candidate regions on the convolution feature map. The sliding window is similar to convolution kernels, with the difference being that convolution kernels can change the dimension of feature maps, but sliding windows will not change the dimension of feature maps. The dimensions of each point in the feature map are 1024; these feature points are called anchors [16].

The candidate bounding boxes are generated with the anchors serving as the central points. The goal of this research is to recognize and localize fruits within images. Since fruits can vary significantly in shape and size during the actual detection process, a diverse set of candidate box types is necessary for accurate positioning. To accommodate this variability, multiple anchor scales are defined, specifically:
[64 x 64], [90 x 45], [45 x 90], [128 x 128],
[181 x 90], [90 x 181], [256 x 256], [362 x 181],
[181 x 362], [512 x 512], [724 x 362], [362 x 724].

This range of anchor scales allows the system to effectively capture and propose candidate regions that can potentially contain fruits of different dimensions and aspect ratios, enabling robust fruit detection and localization [16].

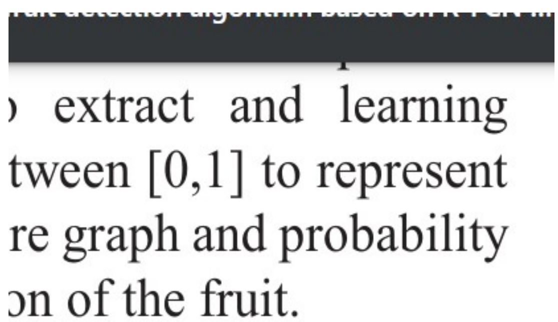


Figure 6. Structure of RPN

The process of generating multiple candidates bounding boxes from a feature map is illustrated in Figure 6. If the feature map contains H×W feature vectors, performing two fully connected operations on each vector is equivalent to applying two 1×1 convolution operations across the entire feature map. For every feature vector, two classification scores and four coordinate offsets are computed. The two scores are used to differentiate foreground from background, while the four coordinates (x, y, w, h) represent adjustments to the original image coordinates. This results in two feature maps of sizes 2×H×W and 4×H×W, corresponding to the

Journal of Theoretical and Applied Information Technology
31st August 2024. Vol.102. No. 16

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

classification layer and position regression layer, respectively. With each feature point treated as an anchor, K bounding boxes are generated. Across the original image, there are a total of K×H×W candidate boxes. These K boxes are pre-defined and can be directly utilized in the calculations, thereby enhancing computational speed and improving recognition accuracy. [16].

R-FCN also uses the RoI produced by RPN. R-FCN and RPN will share features to obtain the region that R-FCN will then classify between background and object. The output of R-FCN is a position-sensitive RoI pooling layer [17]. The architecture of RPN can be illustrated in Figure 7.

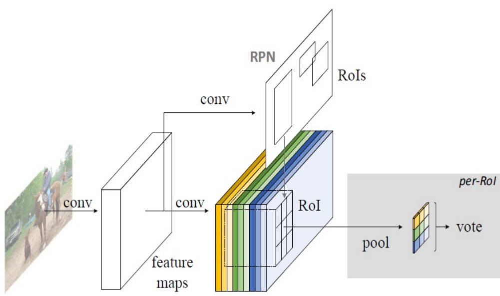


Figure 7. Region Proposal Network Architecture

2.3.3. Residual Network

Residual Network is a network that contains many layers to perform calculations in making decisions that are combined with object detection to determine an object. This Residual Network or ResNet has the characteristic of having many layers ranging from 18 to 152 layers in performing convolution filtering. In this research on papaya fruit classification, the author uses ResNet 101 which initially performs 7x7 convolution, then the layer is set to 64. Next, the calculation shifts two columns. The calculation process continues until the fifth convolution. From the many layers possessed by this ResNet, a sufficient module value is then produced to be able to detect objects according to a predetermined set of data. Details of the ResNet layers are presented in Figure 8.

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
conv2_x	56×56	3×3 max pool, stride 2				
		[3×3, 64; 3×3, 64] ×2	[3×3, 64; 3×3, 64] ×3	[1×1, 64; 3×3, 64; 1×1, 256] ×3	[1×1, 64; 3×3, 64; 1×1, 256] ×3	[1×1, 64; 3×3, 64; 1×1, 256] ×3
conv3_x	28×28	[3×3, 128; 3×3, 128] ×2	[3×3, 128; 3×3, 128] ×4	[1×1, 128; 3×3, 128; 1×1, 512] ×4	[1×1, 128; 3×3, 128; 1×1, 512] ×4	[1×1, 128; 3×3, 128; 1×1, 512] ×8
conv4_x	14×14	[3×3, 256; 3×3, 256] ×2	[3×3, 256; 3×3, 256] ×6	[1×1, 256; 3×3, 256; 1×1, 1024] ×6	[1×1, 256; 3×3, 256; 1×1, 1024] ×23	[1×1, 256; 3×3, 256; 1×1, 1024] ×36
conv5_x	7×7	[3×3, 512; 3×3, 512] ×2	[3×3, 512; 3×3, 512] ×3	[1×1, 512; 3×3, 512; 1×1, 2048] ×3	[1×1, 512; 3×3, 512; 1×1, 2048] ×3	[1×1, 512; 3×3, 512; 1×1, 2048] ×3
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

Figure 8. Details of the ResNet layers

2.3.4. Region of Interest (RoI) Pooling Layer

This is a pre-processing step before training to detect an object in an image. In the previous explanation of the R-FCN theory, it was mentioned that there is an addition of an RoI that is highly position-sensitive. This differentiates it from RPN as the basis of R-FCN. This RoI will be determined based on the position value mapping processed by R-FCN. There will be many RoIs formed and they will be grouped into several object and background categories. After being grouped, the result will be a unified RoI pooling layer. The visualization of when RoIs are accurately and inaccurately placed on objects is presented in Figures 9 and 10 [17].

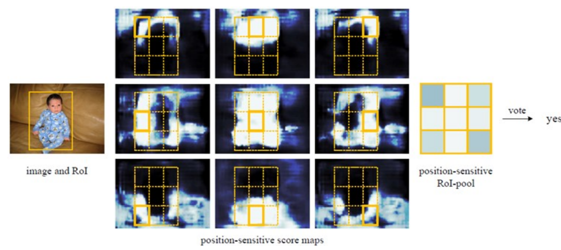


Figure 9. Accurate ROI on Object

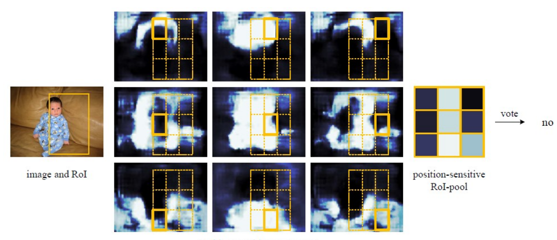


Figure 10. Inaccurate ROI on object

In Figure 9 and Figure 10, the RoIs are shown to be accurately placed on the desired objects, resulting in position-sensitive RoI pools and obtaining a 'yes' vote score. Whereas in Figure 5, the resulting RoI is inaccurately placed on the desired object, resulting in a position-sensitive RoI pool with a low 'no' vote score [17]. This research will be tested using ResNet-101, which consists of 100 convolutional layers.

2.4. Learned Model

During the image data training process, a data model will be produced containing specific values that classify the image as a papaya object according to the specified variety. The trained file is named model.ckpt. From this Learned Model, the success of the testing that will be carried out can be observed. Successful identification is indicated by the detection of the papaya variety, whether it is Cibinong Papaya, Bangkok Papaya, or California Papaya.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

3. TRAINING

The training process plays a crucial role in providing the necessary information and guidance to the object detection system. This stage is where the system learns to recognize and localize objects of interest through exposure to a large number of annotated examples. The training process for this particular implementation is carried out by leveraging the powerful R-FCN (Region-based Fully Convolutional Network) algorithm.

During the training phase, the system ingests a dataset consisting of image files and their corresponding annotations, which are stored in XML format with filenames like "train_labels." These XML files contain bounding box coordinates and class labels for the objects present in each image, serving as the ground truth for the system to learn from.

To facilitate efficient training and take advantage of parallel computing resources, the image and annotation data are converted into the TFRecord format, which is a more compact and optimized representation for efficient input pipeline processing.

The training process itself is built upon the R-FCN ResNet101 architecture, which combines the region-based approach of R-FCN with the powerful ResNet-101 backbone network, known for its deep residual connections and strong feature extraction capabilities. This architecture is implemented using the TensorFlow framework, a widely-used open-source library for machine learning and deep learning applications.

The training process uses R-FCN ResNet101 in TensorFlow environment. The training process is presented in Figure 11.

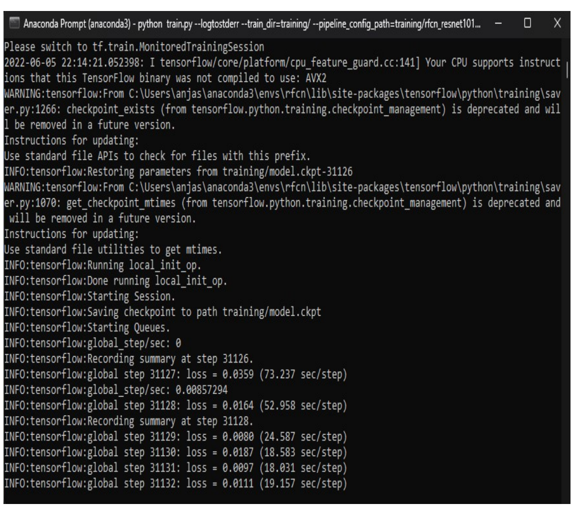


Figure 10. Training Process

4. TESTING AND RESULT

In this research, the data was collected using a smartphone camera by capturing images when the lighting was adequate due to proximity to a light source. The distance between the camera lens and the papaya fruit object was also quite close, no more than 30 centimeters. The captured images can be seen in Figure 12, 13, and 14.

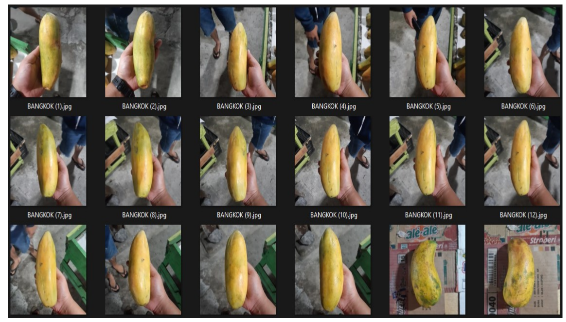


Figure 8. Collection of Bangkok Papaya Image Data

Figure 9. Collection of Bangkok Papaya Image Data

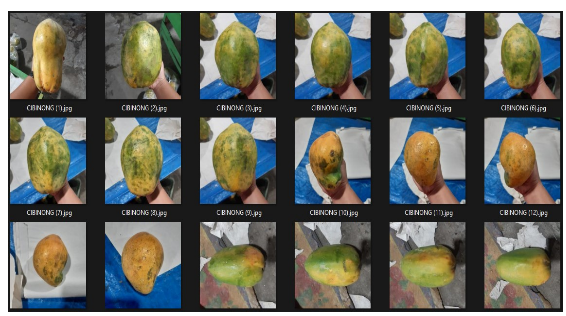


Figure 10. Collection of Cibinong Papaya Image Data

This research presents a desktop-based application prototype as the user interface. When first launched, the program displays a search box to select an image file on the user's computer. After selecting the image file, the user can click the detection button to have the program start identifying and classifying the papaya fruit variety. A new window will appear displaying the image and a box containing the detected object, then at the top it will show the identification result of whether

Journal of Theoretical and Applied Information Technology
31st August 2024. Vol.102. No. 16

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

the object is a California Papaya, Bangkok Papaya, or Cibinong Papaya, along with an accuracy score from 1 to 10. The program display details are shown in Figure 11.

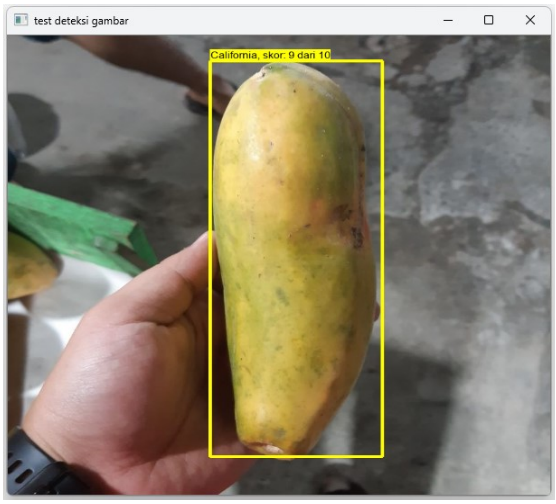
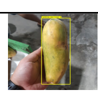
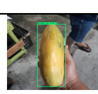
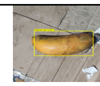
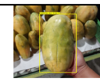
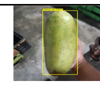
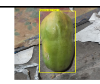
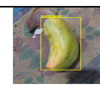


Figure 11. Program Display for Object Detection

The final stage of this research was to test the system in identifying papaya varieties. The goal was to determine whether the system operates according to the specified design and can correctly identify papaya varieties based on the original data.

This testing was conducted on 100 images of California Papaya, 85 images of Bangkok Papaya, and 140 images of Cibinong Papaya. The results of the papaya variety test can be seen in Table 2.

Table 2. The results of the papaya variety test

No.	Papaya Image	Actual Image	Detection Results	Information
1		California	California	Succeed
2		Bangkok	Bangkok	Succeed
3		Bangkok	California	Failed
4		Bangkok	California	Failed
5		California	California	Succeed
6		California	California	Succeed
7		California	California	Succeed
8		California	California	Succeed
9		California	California	Succeed
1		California	California	Succeed

Based on the results of papaya identification and classification tests carried out using the R-FCN method, 6 system errors were found in classifying papaya types from a total of 325 papaya image data used. Total testing data is shown in Table 3 Confusion Matrix. There was an error in the identification of California Papaya with 98 correct data and 2 incorrect data. Identify Bangkok Papaya with 80 correct data and 5 incorrect data. Identification of Cibinong Papaya with 139 correct data and 1 incorrect data.

Table 3. Truth Values of Papaya Identification

Types of Papaya	TP	FP	FN
California Papaya	98	2	0
Bangkok Papaya	80	4	1
Cibinong Papaya	139	1	0
Total	317	7	1

From the observation results, it was also found that system errors in papaya detection were caused by similarities between the image objects of California papaya and Bangkok papaya. In the end, the system experienced difficulty in determining the papaya object correctly.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 4. Confusion Matrix

	California Papaya	Bangkok Papaya	Cibinong Papaya
California Papaya	98	2	0
Bangkok Papaya	5	80	0
Papaya Cibinong	1	0	139

Based on the truth value table for papaya classification, 317 True Positive data are obtained, which is the correct value. This means that the system is able to detect papaya images correctly according to the original. Continues with 7 False Positive data which are wrong values. This means that the system is less accurate in assessing the papaya image with the original. Meanwhile, there is 1 False Negative data, namely wrong detection. Namely a papaya image detection system with double or other values than specified.

Then, based on the training results that have been carried out, accuracy values are obtained from the classification results, namely precision, recall, F-Score [12]. The explanation is in table 5.

Table 4. Confusion Matrix

Types of Papaya	*Precision*	*Recall*	F1-Score
California Papaya	0,98	1	0,98
Bangkok Papaya	0,95	1,03	0,98
Cibinong Papaya	0,99	1	0,99

In the table above there is Precission, namely the comparison between the correct prediction value and all correct results. The formula is:

$$precision = \frac{TP}{TP + FP} \quad (1)$$

Recall is a comparison between the correct predicted value and all correct data. The formula is:

$$recall = \frac{TP}{TP+FN} \quad (2)$$

F1-Score is an average comparison between the precision value and the recall value. The F1-Score formula is:

$$F1\ Score = 2\,x\,\frac{Recall\ x\ Precission}{Recall + Precission} \quad (3)$$

Then, from the test results, the accuracy is calculated using the equation.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

$$accuracy = \frac{317}{325}\,x\,100\%$$

The research entitled Classification of Papaya Fruit Types Using the Desktop-Based R-FCN ResNet101 Method succeeded in detecting and classifying papaya fruit types with an accuracy of 97.5%.

There are some errors amounting to 2.5%. This is due to the similarity between the objects in the image of California papaya and Bangkok papaya. In the end, the system experienced difficulty in determining the papaya object correctly.

5. CONCLUSION

Several conclusions can be drawn from the research conducted to detect image objects and classify their types using the R-FCN ResNet101 method.

1. This study successfully implemented computer vision and deep learning approaches to identify papaya varieties, specifically California, Bangkok, and Cibinong papayas, achieving an accuracy of 97.5%. The developed system demonstrates potential for both consumer use and commercial applications in identifying papaya varieties
2. The R-FCN ResNet101 algorithm proved effective for this task, but faced challenges in distinguishing between California and Bangkok papayas due to their visual similarities, especially in close-up images. Optimal performance required adequate lighting conditions during image capture. To enhance accuracy, future work should focus on expanding the training dataset with more diverse examples and implementing advanced techniques for differentiating visually similar varieties.
3. This research contributes to the field of agricultural technology by showcasing the

potential of computer vision and deep learning in fruit variety identification. The developed system could be adapted for various applications including quality control, inventory management, and automated sorting in the papaya industry.

4. While the achieved accuracy of 97.5% is promising, it leaves room for improvement when compared to similar studies in fruit identification. For instance, previous research using the Bayesian method for rambutan varieties achieved 100% accuracy, and a DCNN method for corn seed varieties reached 98.1% accuracy. Future research directions include exploring these alternative methods to potentially enhance the accuracy of papaya variety identification.

ACKNOWLEDGMENT:

The authors would like to thank profusely for the financial support provided by DIPA funds Polmed (Politeknik Negeri Medan) to carry out research with contract number: B/368/PL5/PT.01.05/2024. The funds received from the DIPA POLMED gives a very important role in the successful completion of this study.

REFERENCES :

[1]. Altuntaş, Y., A.F. Kocamaz, and C. Yeroğlu. *Identification of apricot varieties using leaf characteristics and KNN classifier*. in *2019 International Artificial Intelligence and Data Processing Symposium (IDAP)*. 2019. IEEE.

[2]. Guntur, S. and M.R. Wayahdi. *Analisis Metode Bayes dalam Identifikasi Varietas Buah Rambutan*. in *Semantika (Seminar Nasional Teknik Informatika)*. 2019.

[3]. Cinar, I. and M. Koklu, *Identification of rice varieties using machine learning algorithms.* Journal of Agricultural Sciences, 2022: p. 9-9.

[4]. Javanmardi, S., et al., *Computer-vision classification of corn seed varieties using deep convolutional neural network.* Journal of Stored Products Research, 2021. **92**: p. 101800.

[5]. Khan, A.A., A.A. Laghari, and S.A. Awan, *Machine learning in computer vision: a review.* EAI Endorsed Transactions on Scalable Information Systems, 2021. **8**(32): p. e4-e4.

[6]. Wang, X. and Z. Zhu, *Context understanding in computer vision: A survey.* Computer Vision and Image Understanding, 2023. **229**: p. 103646.

[7]. Wang, J., et al., *Automated diabetic retinopathy grading and lesion detection based on the modified R-FCN object-detection algorithm.* IET Computer Vision, 2020. **14**(1): p. 1-8.

[8]. Xu, C., J. Fan, and L. Liu. *A novel object detection algorithm based on enhanced R-FCN and SVM*. in *2020 IEEE Conference on Telecommunications, Optics and Computer Science (TOCS)*. 2020. IEEE.

[9]. Occorsio, D., G. Ramella, and W. Themistoclakis, *Lagrange–Chebyshev Interpolation for image resizing.* Mathematics and Computers in Simulation, 2022. **197**: p. 105-126.

[10]. Saputra, D.H. and B. Imran, *OBJECT DETECTION UNTUK MENDETEKSI CITRA BUAH-BUAHAN MENGGUNAKAN METODE YOLO.* Jurnal Kecerdasan Buatan dan Teknologi Informasi, 2023. **2**(2): p. 70-80.

[11]. NURSULISTIO, F., *Deteksi Objek Masker Menggunakan Object Detection API dan TensorFlow Lite Model Maker.* 2022.

[12]. Deng, L., et al., *Region-based CNN method with deformable modules for visually classifying concrete cracks.* Applied sciences, 2020. **10**(7): p. 2528.

[13]. Chen, Y.P., Y. Li, and G. Wang, *An Enhanced Region Proposal Network for object detection using deep learning method.* PloS one, 2018. **13**(9): p. e0203897.

[14]. Zou, W., et al., *SC-RPN: A strong correlation learning framework for region proposal.* IEEE Transactions on Image Processing, 2021. **30**: p. 4084-4098.

[15]. Shih, K.-H., et al., *Real-time object detection with reduced region proposal network via multi-feature concatenation.* IEEE transactions on neural networks and learning systems, 2019. **31**(6): p. 2164-2173.

[16]. Jian, L., Z. Mingrui, and G. Xifeng. *A fruit detection algorithm based on r-fcn in natural scene*. in *2020 Chinese Control And Decision Conference (CCDC)*. 2020. IEEE.

[17]. Fadhillah, H., S.A. Wibowo, and Purnamasari, *Investigasi Pengaruh Skema Stride dan Step Training untuk Deteksi Jari pada Region-based Fully Convolutional Network (R-FCN) dalam Teknologi Augmented Reality.* Jurnal Ilmiah Fifo, 2020. **12**(2): p. 138-148.