BULLETIN OF NATIONAL STUDENTS' EXHIBITION IN THE FIELD OF INFORMATION AND COMMUNICATION TECHNOLOGY
ISSN: xxxxxx

Automatic Election Vote Counting System on Form C1 with Deep Learning Implementation

M. Alif Al Hakim1 , Rakha Abid Nobleman2 ,Muhammad Mariozulfandy3 ,Adila Alfa Krisnadhi4
1Faculty of Computer Science, University of Indonesia, Depok, 16424, email: malif.al@ui.ac.id
2Faculty of Computer Science, University of Indonesia, Depok, 16424, email: rakha.abid@ui.ac.id
3Faculty of Computer Science, University of Indonesia, Depok, 16424, email: muhammad.mariozulfandy@ui.ac.id
4Faculty of Computer Science, University of Indonesia, Depok, 16424, email: adila@cs.ui.ac.id
Corresponding Author: M.Alif Al Hakim

ABSTRACT — The process of calculating or recapitulating the vote count for general elections in Indonesia is still done manually. This of course takes a lot of time and money. Although Indonesia has started implementing an automatic vote counting system using optical recognition text on Sirekap in 2024, there are still accuracy problems with Sirekap. Therefore, a new automatic vote counting system with good performance is needed to help the vote counting or recapitulation process become more effective and efficient. This study aims to develop an automatic vote counting system on Form C1 for the Presidential and Vice Presidential elections using the PaddleOCR and YOLO-NAS fine-tuning methods. This method is used because the method used by Sirekap to determine the Region of Interest (ROI) is still done manually.

In addition, the accuracy of the Sirekap OCR model is also claimed to be 93% on field data. In this study, a dataset was used that was scraped through the general election commission website. The dataset used consists of images of Form C1 that were randomly taken from various Polling Stations (TPS) from all cities/regencies in Indonesia. The dataset consists of 1187 images. Each image in the dataset is then preprocessed, such as Auto Orient, Gray Scale and Data Augmentation. The processed images are then used to train the YOLO-NAS detection model which is useful for detecting ROI, namely the vote counting result box and the PaddleOCR model which is useful for reading text in images. In the evaluation carried out, this method has quite good performance where it has an accuracy of 96.9% in the correctness of reading the vote counting results. Prediction errors that occur are generally caused by images that are very blurry or inconsistent in writing text and marking OMR vote counting results in the Form C1 image data.

Thus, from this study it is concluded that the use of PaddleOCR and YOLO-NAS fine-tuning can improve the efficiency and accuracy of election vote counting. However, it should be noted that the performance of the resulting model depends on the quality of the image.

KEYWORDS — OCR, Paddle, YOLO, Election, Automatic Vote Counting

ABSTRACT — The process of calculating or recapitulating the vote count for general elections in Indonesia is still done manually. This of course takes a lot of time and money. Even though Indonesia has started implementing an automatic vote counting system using optical recognition text in Sirekap in 2024, but there are still accuracy problems in Sirekap. Therefore, a new automatic vote counting system with good performance is needed to help the vote counting or recapitulation process become more effective and efficient. This research aims to develop an automatic vote counting system for Form C1 for the general election of President and Vice President using the PaddleOCR and YOLO-NAS fine-tuning methods. This method is used because the method used by sirekap to determine Region of Interest (ROI) is still done manually. Apart from that, the accuracy of the Sirekap OCR model is also claimed to be 93% on field data. In this research, a dataset was scraped via the general election commission's website. The dataset used consisted of images of Form C1 which were randomly taken from various polling places (TPS) in all cities/districts in Indonesia. The dataset consists of 1187 images. Each image in the dataset is then subjected to preprocessing, such as Auto Orient, Gray Scale and Data Augmentation. The images that have been processed are then used to train the YOLO-NAS detection model which is useful for detecting ROIs, namely boxes resulting from sound calculations and the PaddleOCR model which is useful for reading text on images. In the evaluation carried out, this method performed quite well, having an accuracy of 96.9% in correctly reading the vote count results. Prediction errors that occur are generally caused by images that are very blurry or inconsistent in text writing and marking OMR vote calculation results on Form C1 image data. Thus, from this research it is concluded that the use of fine-tuning PaddleOCR and YOLO-NAS can increase the efficiency and accuracy of election vote calculations. However, it should be noted that the performance of the resulting model depends on the quality of the images.

KEYWORD — OCR, Paddle, YOLO, Election, Automatic Vote Counting

ISSN: xxxxxx

I. INTRODUCTION

A. BACKGROUND

General Election (Pemilu) is an activity held every five years in Indonesia and must be participated by all Indonesian citizens aged over 17 years. Election is an election process to elect most or all members of an elected legislative body and president who are directly elected by the people [1]. Currently, the process of counting and recapitulating election votes in Indonesia is carried out manually starting from the sub-district, city, province, and finally national levels so that this calculation takes a long time, up to at least one month from the voting process and costs a lot of money. The Ministry of Finance (Kemenkeu) has allocated a budget of up to IDR 71.3 trillion for the 2024 Election. This amount of funds has increased by 57.3 percent compared to the budget for the simultaneous democratic party in 2019 which was IDR 45.3 trillion [2].

The automatic vote counting system can handle this. This automatic vote counting system works by automatically counting the votes of each candidate pair with the Form C1 data from the voting results. This system can speed up the vote counting process because it is done automatically by the machine and reduce the operational costs of the Election because the fewer manual vote counting tasks are done (the easier the task, the lower the cost) if this system works well.

The automatic vote counting system with OCR technology has begun to be implemented by the KPU in the 2024 Election under the name Sirekap [3]. However, there are still several problems, such as the accuracy problem of this system, which is mainly caused by digit reading errors and poor image capture. Based on the report published on the Constitutional Court website, Sirekap has three main problems that hinder the vote counting process [4], namely as follows.

1) Handwriting that failed to be detected by OCR on Sirekap. Sirekap claims that their OCR accuracy is 93% in the field so there is still a 7% error in reading the image.

2) Bad image capture. The image data obtained by Sirekap was obtained from a cellphone so the quality of the images obtained varies and of course some are bad.

3) The problem of folded paper. This causes errors in the interpretation of the Sirekap OCR model. Moreover, if the vote counting results section is not visible.

In addition, researchers also conducted an analysis of the process of reading the results of the C1 Form image voice calculation on the Sirekap mobile application.1 Researchers obtained information that Sirekap uses an OCR model that has been trained with MNIST (handwritten dataset of digits 0-9). After the application obtains the image, several preprocessings are carried out on the image, such as grayscale, adjusting the contrast with CLAHE, and applying adaptive thresholding. After that, the vote counting result box for each candidate pair will be detected using the coordinates that have been set (constant) and there is a slight adjustment to the image size mathematically for cropping. After the vote counting result box is obtained, it will be read using the model. The model used is the MNIST OCR model which is used to read each digit in each vote digit box. In addition to using OCR, the application can also perform OMR (Optical Mark Recognition) and is more likely to do that (if OMR fails, the model performs OCR). OMR is implemented using OpenCV by detecting all circles and finding outlier circles (black circles that are not right in the circle area), then Optical Mark Recognition is performed on the circle.

The Sirekap problem gave rise to several other problems, such as public unrest due to the correction of the vote recapitulation of one of the presidential and vice presidential candidates [5].
This error also caused unrest on social media at that time about vote manipulation, even though it was only an error due to a problem with the Sirekap system. These problems prompted researchers to contribute to developing an automatic vote counting system on the C1 Election Form. From the problems and analysis of Sirekap above, researchers will try to implement a vote counting system using object detection (vote counting box) with the YOLO-NAS model and reading with PaddleOCR.

B. PROBLEM FORMULATION

Based on the background that has been explained, the research questions that the research team will propose are as follows.

1) How effective and accurate is the voice counting system on form C1 with PaddleOCR and YOLO optimization?

2) What are the challenges and obstacles faced by the vote counting system on form C1 with PaddleOCR and YOLO optimization?

3) How do the two methods, namely YOLO and PaddleOCR, perform in recognizing voice digits in poor quality images?

C. RESEARCH OBJECTIVES

The following are the objectives to be achieved from this research.

1) To find out the effectiveness and accuracy of the vote counting system on the C1 form with PaddleOCR and YOLO optimization.

2) To know the challenges and obstacles faced by the vote counting system on form C1 with PaddleOCR and YOLO optimization.

3) To find out how PaddleOCR and YOLO perform on poor quality images.

1obtained from Reverse Engineering Sirekap mobile application

ISSN: xxxxxx

D. **BENEFITS OF RESEARCH**

This research is expected to provide benefits to parties who need it, both theoretically and practically, including the following.

1) Theoretical Benefits • As a blueprint to create a vote counting system on form C1 automatically. • Increase the reader's insight and knowledge regarding the vote counting system from the given dataset.

2) Practical Benefits • Assisting the government in its efforts to calculate valid votes from Form C1. • Increasing efficiency of manpower and time in conducting valid vote recapitulation. • Detecting Form C1 image data that has poor quality (sharpness, brightness, retrieval) with the performance of the vote counting model on the image.

E. **RESEARCH LIMITATIONS**

The research limitations of the methods presented in this paper are as follows.

1) The model is made only for voice detection on Form C1 for the election of presidential and vice presidential candidates.
2) The performance of the model depends on the quality (such as sharpness, brightness, and capture distance) of the given Form C1 image.
3) The research only focuses on developing an automatic vote counting system model on the Election Form C1.
4) The number of Form C1 image datasets used in model development was 1187 data.
5) The vote counting box section in the Form C1 image data must be clearly visible to run the vote counting system created.

II. METHODOLOGY

This chapter explains the brief research flow conducted by the research team, the dataset used to train and validate the model, an explanation of the model and data processing performed, and the metrics used to evaluate the model's performance in detecting digits in the calculation of valid votes on the C1 form.

A. **SUMMARY OF RESEARCH FLOW**

The research was conducted to create an automatic vote counting system for each Candidate Pair Form C1. The following is a diagram that can be seen in Figure 1 and a brief explanation of the research flow to achieve this.

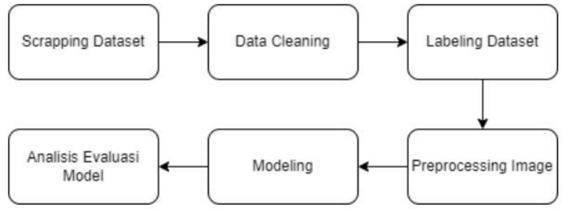

Figure 1. Research Flow

1) Scraping the Election Form C1 Dataset The data that will be detected for the vote results and used for model training is the Form C1 image. This image data was obtained from the page https://pemilu2024.kpu.go.id/pilpres/hitung-suara/. The research team conducted automation scraping with the Selenium 4.22.0 Python package to obtain three random Form C1 images at polling stations in all cities/regencies in Indonesia. We obtained 1194 Form C1 images from this scraping result.

2) Dataset Cleaning From all the scraped data, there are still some data that are formatted incorrectly due to the automatic scraping process or are too dirty to be trained on the model so that cleaning will be carried out to obtain the Form C1 dataset desired by the research team. This very dirty dataset is data that is stated at the model boundary. However, the research team still leaves dirty datasets, as in Figure 2 which will be processed first and trained on the model.

Figure 2 is an example of data that was deleted due to a format error (left) and is very dirty because the vote counting box for one of the Candidate Pairs is not visible (right). The number of datasets after cleaning is 1187 data.

Figure 2. Format Errors When Scraping and Dirty Data (Clipped)

3) Dataset Labeling After cleaning the data, the remaining dataset will be labeled. There are two labeling processes on the dataset that will be performed on the data. The first labeling is labeling to determine the output of the YOLO-NAS model for detecting region of interest (ROI) objects in this case the box.

ISSN: xxxxxx

the results of the vote count as illustrated in Figure 3.

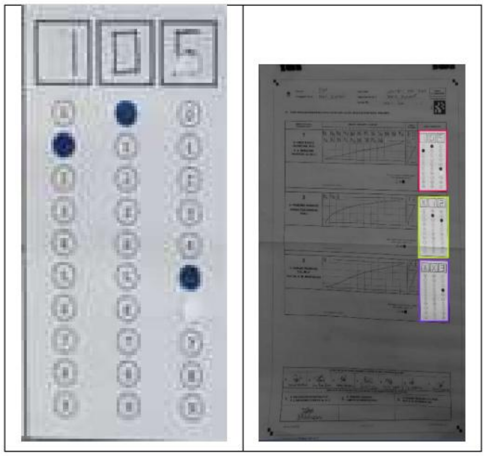

Figure 3. Region of Interest (ROI) in the Dataset

The second labeling is to determine the output of the PaddleOCR model in determining the number of sounds read from the ROI.

4) Dataset Preprocessing

Preprocessing is done on the dataset before it is trained to the model. The dataset that will be detected and used in the training process is not always clean so it is necessary to process the dataset first so that the data is cleaner and can be accepted by the model.

After preprocessing, the dataset will be divided into 70% training data, 20% validation data, and 10% testing data.

5) Modeling

There are two models used by the research team in this study, namely YOLO-NAS and PaddleOCR.
The explanation of both models is in the MODEL section below. Briefly, the YOLO-NAS model will be used for ROI object detection.

Meanwhile, PaddleOCR is for reading the results of voice calculations. The training data that has been obtained previously will be used in the training process of these two models.

6) Model Evaluation

After training the model, the model performance will be evaluated using Mean Average Precision (YOLO-NAS) and Counter Level Accuracy, Digit Level Accuracy, Average Underprediction Error, and Average Overprediction Error.

B. **DATASET**

The dataset used in the study is a collection of C1 form photos scraped from the KPU page. The collected data is cleaned first to obtain a valid dataset according to the C1 form format.

The number of datasets obtained after scraping is 1194.

From the collected datasets, there are some dirty data, such as colored, blurry, dark, taken from a distance, tilted, and bright. Here are some examples of datasets with these conditions in Figure 4.

Figure 4. Dirty Image Samples in the Dataset

C. **PREPROCESSING**

Dirty datasets will be handled by pre-processing the dataset before training it to the model. Here are some preprocessing methods that are carried out.

1) Grayscale: This Grayscale process changes the image from RGB pixels (colored) to Grayscale pixels (black and white) by converting the colors Red (R), Green (G), and Blue (B) to a light intensity scale (light/dark). This processing can help clarify colored, bright, or dark images so that they can be more easily recognized by the model.

2) Auto Orientation: This method is done to automatically change the image orientation to portrait with the correct format. This is needed to handle misoriented data such as landscape or upside down before being trained to the model. This processing can help object detection models such as the YOLO model. Figure 5 is an example of the application of auto orientation preprocessing.

3) Data Augmentation: Data Augmentation is the process of generating new data from existing data by adding some processing to the new data.

In this case, the research team performed several processing, namely Rotation (setting the image rotation to vary), Shear (setting the image perspective to vary), Brightness (setting the image brightness to vary), and Blur (setting the image blur to vary). This was done so that the model

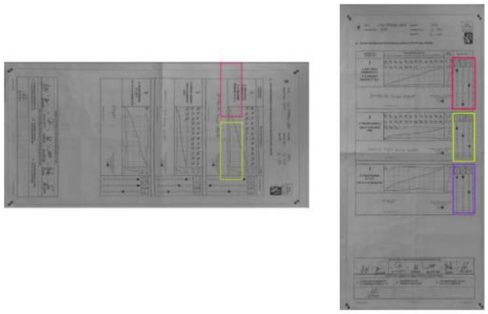

Figure 5. Auto Orientation on Dirty Image

can be trained to be more accustomed to various image data conditions even if the image data is 'dirty'.

D. **MODEL**

The data modeling process will be carried out on two models. The first model is YOLO-NAS which will be used for ROI object detection, namely the vote counting result box in the Form C1 image. After that, cropping will be done on the vote counting result box. The number of vote counting result box images obtained from each Form C1 image will be in accordance with the number of Candidate Pairs. The vote counting result box image will then be read with the second model, namely Paddle-OCR, so that a system will be obtained that automatically counts the votes from each Candidate Pair on Form C1.

1) **YOLO-NAS**

YOLO is a model that is commonly used in the object detection process [6]. YOLO-NAS is an improvement of the YOLO model where optimization is carried out on the YOLO neural network architecture to improve accuracy and speed. The research team used this model to detect the Region of Interest (ROI) in the Form C1 image, in this case the vote counting box. To train this model, the research team first annotated the ROI on the data. After that, the research team will train the model with the annotated data. The trained model can predict the ROI coordinates of the vote counting box on the new Form C1 image. By obtaining the ROI coordinates from the Form C1 image, cropping can be done on the image to obtain the ROI image. This was done because the research team only wanted to focus on the PaddleOCR model used to read the vote results on the ROI. The goal is that the data to be read by the PaddleOCR model is clearer and not complex. The YOLO-NAS Modeling Visualization is shown in Figure 6.

2) **PaddleOCR**

PaddleOCR is one of the pretrained OCR models that has been developed for text detection and recognition in images [7]. This model is known for its accuracy and efficiency in performing OCR tasks, such as

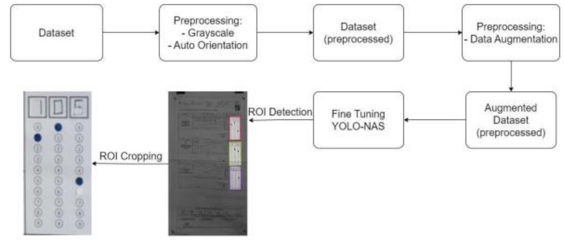

Figure 6. YOLO-NAS Modeling Flow

license plate recognition, handwriting recognition, and others. The research team used this model to read the number of votes from each ROI image received. This model was trained with ROI image data and its labels. The output of this model is the number of votes from the ROI. The research team used the PaddleOCR model with epoch 220, Adam optimizer, Cosine learning rate 0.001, and CTCLoss loss function. The visualization of the PaddleOCR Modeling is shown in Figure 7.

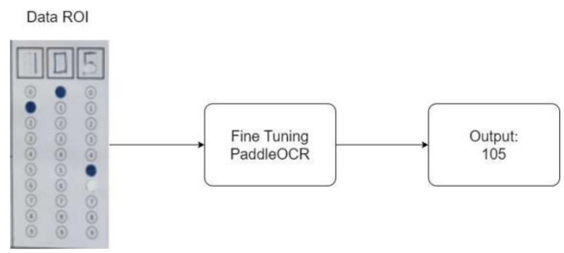

Figure 7. PaddleOCR Modeling Flow

E. **MODEL EVALUATION METRICS**

The following is an explanation of several metrics used to evaluate the model's performance in detecting ROI objects and reading the candidate pair's vote results.

1) **Mean Average Precision**

Mean Average Precision is a metric to evaluate the YOLO-NAS model in detecting objects (ROI).
This metric involves precision and recall values.

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

Average Precision (AP) = Area under the precision-recall curve. The precision-recall curve plots precision against recall for different threshold values. As the detection threshold varies, the precision and recall values change, creating a curve.
This curve is plotted for each class. In this case, there are three vote counting boxes (ROIs) on each image data of Form C1 so that there are three classes and each class forms an AP.

ISSN: xxxxxx

$$\text{Mean AP} = \frac{1}{N}\sum_{i=1}^{N} \text{Fire}$$

N is the number of classes and i is the ith class.

2) **Counter Level Accuracy** Counter Level Accuracy is one of the metrics used to evaluate the PaddleOCR model in reading the sound results on the ROI. In the PaddleOCR model, the data entered is the ROI from the Form C1 image.

This metric works by calculating the percentage of the number of correct predictions (True Positive (TP)) from the total number of testing data. Here is the mathematical formula that states this metric.

$$\text{Count Level Accuracy (CLA)} = \frac{TP}{TP + TN + FP + FN}$$

3) **Digit Level Accuracy**
Digit Level Accuracy is one of the metrics used to evaluate the PaddleOCR model in reading voice results on the ROI. This metric is calculated by extracting digits from the Pad-dleOCR prediction results and comparing each digit with the corresponding digit on the label. The percentage of the number of correct corresponding digits from the total number of digits in the testing data (ROI) is the Digit Level Accuracy (DLA) of the data.

4) **Average Overprediction Error and Average Under-prediction Error**
This is a custom metric created by the research team to measure how far the deviation of the number of votes for each candidate pair predicted by the PaddleOCR model is from the number of votes. Both of these metrics are calculated by subtracting the PaddleOCR prediction results from the data with the data label. If the subtraction result is negative, the research team will categorize it as underprediction.

If the result is positive, the research team will categorize it as overprediction. Each under-prediction and overprediction will be collected and averaged across all testing data to obtain the Average Overprediction Error and Average Underprediction Error.

III. RESULTS AND DISCUSSION

This chapter explains the results and evaluation of the system that has been created in this study. Previously, it has been explained that 1187 Form C1 image datasets were divided into 70% training data, 20% validation data, and 10% testing data and it has also been explained that Form C1 image data will be trained on the YOLO-NAS model to detect the ROI coordinates of the vote counting box. Table I shows the results of the YOLO-NAS model performance evaluation on the testing data with the metrics that have been explained previously.

From these results, it can be said that the performance of the YOLO-NAS model trained to detect box ROIs

Metric	Results
Precision	99.7%
Recall	100.0%
Mean AP	100.0%

TABLE I
YOLO-NAS MODEL EVALUATION RESULTS

the sound calculation results are very good. Although the dataset used only performed two preprocessing for cleaning, namely Grayscale and Auto Orientation, as well as Data Augmentation preprocessing for dirt, the performance of YOLO-NAS is already very good in detecting ROI.
This is a difference or even an improvement from Sirekap where ROI cropping detection is performed at predetermined coordinates (constant) with slight adjustments according to the image size which is done mathematically with OpenCV. According to researchers, the drawback of the Sirekap approach in detecting ROI is that it is susceptible to dirty image data, such as tilted, long-distance shooting, close-up shooting, and paper conditions that may be wrinkled or slightly folded. This problem can be handled with the YOLO-NAS model because it can perform object detection (ROI) automatically. Figure 8 is an example of an image with this problem that can be detected by ROI well by YOLO-NAS.

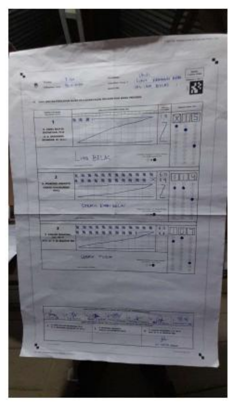

Figure 8. Example of Dirty Image for Detecting ROI

Previously, the research team has explained that the ROI image data that has been obtained will be entered into the PaddleOCR model. This PaddleOCR will produce output in the form of the number of votes read from the ROI image data. Table II shows the results of the Pad-dleOCR model performance evaluation on the testing data with the metrics that have been explained previously.

The results can be said to be quite good even though there are still some errors in the PaddleOCR model in reading. After further analysis of the data that failed to be read, reading errors were generally caused by inconsistent filling (between OMR circles and text) or images that were too blurry. For example, like the two images in Figure 9 that failed to be detected where the left image is inconsistent between text and

Metric	Results		
Count Level Accuracy	0.969		
Digit Level Accuracy	0.983		
	Candidate Pair 1	Candidate Pair 2	Candidate Pair 3
Average Overprediction Error		52.0	1.0
Average Underprediction Error	0 -61.0	-114.0	-44.25

TABLE II
PADDLEOCR MODEL EVALUATION RESULTS

OMR circles, as well as a very blurry right image for read by the model.

Figure 9. Data that PaddleOCR Failed to Read

However, the accuracy of the developed model by the research team is better than the MNIST model Sirekap OCR which is claimed to have 93% accuracy performance in the field. According to researchers, the approach taken by Sirekap is still very vulnerable to inappropriate writing clear, as if there is a crossed out/ex-taped part on the digit thus causing errors in reading the digits. The approach the research team took to address this issue is to create a model to also review the circle OMR at the bottom so that besides focusing on the digits, These dots can also help recognize digits/text. Figure 10 is an example of data with inconsistent writing. which was successfully read by the PaddleOCR model because of the circles also considered.

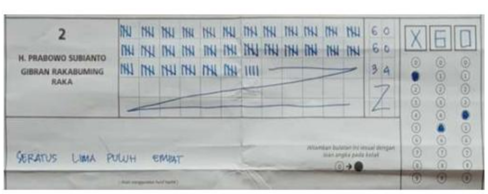

Figure 10. Example of Successfully Detected Image

Apart from accuracy, the research team also conducted an evaluation on the speed performance of the model. The research team conducted measurement of the model prediction speed and obtained the results are as follows.

The model prediction time can be said to be quite fast. This due to the architecture of the YOLO-NAS and PaddleOCR models itself is optimal and efficient.

Model	Inference Time
YOLO NAS	0.13 s
PaddleOCR	0.017 s

TABLE III
MODEL SPEED RESULTS

IV. CONCLUSION

In general, the application of Deep Learning uses YOLO-NAS and PaddleOCR fine tuning techniques have good accuracy and effectiveness performance. However, there are challenges where the system only uses part of the form C1 as a region of interest so that if there is one inconsistencies in the region of interest, the model will have difficulty predicting. This system has also been proven can recognize digits quite well in images with poor quality, such as poor photo angle, photo distance which are too far away, and the photos are relatively dark. In addition, this system also has better performance compared to with Sirekap. This model can provide flexibility which is better in various conditions compared to Sirekap system. This system is expected to help the process better calculation of vote recapitulation for each TPS and efficient.

V. ACKNOWLEDGEMENTS

The research team would like to express its gratitude to Almighty God, and honorable leaders University of Indonesia, head of the Faculty of Computer Science, and Our Supervising Lecturer for his help and support for completing our scientific papers, among others.

1) Badrul Munir, Ph.D. (Director of Student Affairs, University of Indonesia)
2) Muhammad Hafizhuddin Hilman, Ph.D. (Manager Student Affairs and Alumni Relations Faculty of Science Computer)
3) Alfan Farizki Wicaksono, Ph.D. (Program Coordinator) Bachelor of Computer Science and International Class)
4) Adila Alfa Krisnadhi, Ph.D. (Assistant Lecturer)

REFERENCE

[1] "Election political science," accessed: 05-14-2024. [On line]. Available: https://www.britannica.com/topic/election-political-science
[2] "Details on specific regulatory document," accessed: 2024-05-14. [Online]. Available: https://jdih.bpk.go.id/Info/Details?id=dac64143-c74b-4644-aca5-a308c1724717
[3] KPU, "Indonesia: Sirekap app helps to monitor election recapitulation," Asia News Monitor, Jan 2024, accessed: 2024-05-14. [On line]. Available: https://www.proquest.com/newspapers/indonesia-sirekap-app-helps-monitor-election/docview/2916439354/se-2
[4] U. Argawati, "IT expert reveals three sources of problems sirekap," April 2024, accessed: 2024-06-28. [On line]. Available: https://mkri.id/index.php?page=web.Berita&id=20196
[5] "KPU corrects publication: Anies gets 3 million votes in 1 tps, deny the issue of lowering," accessed: 2024-05-14. [On-line]. Available: https://news.detik.com/pemilu/d-7196395/kpu-koreksi-publikasi-anies-dapat-3-juta-suara-di-1-tps-tepis-isu-turunkan
[6] J. Terven, D.-M. Cordova-Esparza, and J.-A. Romero-Gonz alez, "A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas," Machine Learning and Knowledge Extraction, vol. 5, no. 4, p. 1680–1716, Nov. 2023. [Online]. Available: http://dx.doi.org/10.3390/make5040083

[7] C. Li, W. Liu, R. Guo, X. Yin, K. Jiang, Y. Du, Y. Du, L. Zhu, B. Lai,
Available: https://arxiv.org/abs/2206.03001