Accredited Ranking SINTA 2
Decree of the Director General of Higher Education, Research and Technology, No. 158/E/KPT/2021
Validity period from Volume 5 Number 2 of 2021 to Volume 10 Number 1 of 2026

Published online on: **http://jurnal.iaii.or.id**

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 7 No. 2 (2023) 236 - 245 ISSN Media Electronic: 2580-0760

Tajweed-YOLO: Object Detection Method for Tajweed by Applying HSV Color Model Augmentation on Mushaf Images

Anisa Nur Azizah[1], Chastine Fatichah[2]
[1,2]Informatics, Intelligent Electrical and Informatics Technology, Institut Teknologi Sepuluh Nopember
[1]anisaazizah069@gmail.com, [2]chastine@if.ts.ac.id*

Abstract

Tajweed is a basic knowledge of learning to read the Al-Qur'an correctly. Tajweed has many laws grouped into several parts so that only some people can memorize and implement Tajweed properly. Therefore, it is necessary to have an automatic detection system to facilitate the recognition of Tajweed, which can be used daily. This study presents Tajweed-YOLO, which applies the HSV color augmentation model to detect Tajweed objects in Mushaf images using YOLO. The contribution to this study was to compare the three versions of You Only Look Once (YOLO), i.e., YOLOv5, YOLOv6, and YOLOv7, and usage of the HSV color model augmentation to improve Tajweed detection performance. Comparing the three YOLO versions aims to solve problems in detecting small objects and recognizing various forms of Mushaf writing fonts in Tajweed detection. Meanwhile, the HSV color model aims to recognize Tajweed objects in various Mushaf and handle minority class problems. In this study, we collected four different Al-Qur'an mushaf with 10 Tajweed classes. The augmentation process can increase the detection performance by up to 85% compared to without augmentation 6th Class (Mad Jaiz Munfashil) using YOLOv6. The comparison of three YOLO versions concluded that YOLOv7 was better than YOLOv5 and YOLOv6, seen in data with augmentation and without augmentation. The evaluation results of mAP0.5 on 17 test data on the YOLOv7, YOLOv6, and YOLOv5 models are 80%, 69%, and 71%, respectively. These results prove that this research model 's results are suitable for the real-time detection of Tajweed.

Keywords: tajweed; object detection; YOL;, augmentation; HSV color model

1. Introduction

Al-Qur'an is the first fundamental source of the Islamic law that Allah SWT has guaranteed authenticity. Therefore, every Muslim should study and read the Al-Qur'an properly and correctly. Al-Qur'an was written in Arabic, while the Muslim community spread in various countries with different languages. Therefore, it is necessary to have basic rules explaining the rules for correctly reading the Al-Qur'an. Tajweed science studies how to pronounce (makhraj) hijaiyah letters, punctuation, and Arabic grammar, especially in reading the Al-Qur'an. Tajweed is essential to avoid misunderstanding the Al-Qur'an's meaning [1]. The science of Tajweed consists of several parts, such as the laws of Nun Sukun and Tanwin, Mim Sukun, Mad, and others [2]. This section also consists of several sub-parts, such as the Nun Sukun parts, which contain readings of Idzhar Halqi, Idhgam, Iklab, and Ikhfa'. The number of Tajweed is many and varies in pronunciation. Not many people can memorize all of Tajweed and can implement them well.

Introducing Tajweed is a solution to overcome errors in the pronunciation of the Al-Qur'an. The recognition is in the form of an object detection system from several Tajweed, carried out through digital image processing. Several studies on the introduction of Tajweed have been carried out using various methods, including applying typical distance calculations [3]. This research is to get Edge Pattern Detection of Tajweed using Hamming, Manhattan, and Euclidean calculation tests. Another study introduced the three laws of Tajweed, namely Idgham Bighunnah, Idgham Bilaghunnah, and Ikfa' Haqiqi, using image segmentation techniques. Tajweed segmentation uses the Pattern Recognition method with the Speeded-Up Robust Feature (SURF) Algorithm for feature extraction. The features obtained are calculated for their nearness using Euclidean Distance [4]. Furthermore, [5] identified the Tajweed pattern using the Fuzzy Associative Memory (FAM) method. The average detection result reaches 80% on the Tajweed of Idgham Mutajanisain. Another study [6] detected the Tajweed of Mad Lazim Harfi Musyba using the Deep Learning Convolutional Neural

Accepted: 02-12-2022 | Received in revised: 13-02-2023 | Published: 03-03-2023

Network (CNN) method. The average accuracy results reach 93.25%. Based on previous research, the detection of Tajweed objects is carried out with various ideas and methods such as Edge Detection, segmentation, and pattern recognition by calculating feature distance proximity to feature learning in Deep Learning using CNN. However, no Tajweed detection system can be done in real-time. Therefore, this research detects multiclass Tajweed objects using the You Only Look Once (YOLO) method.

Along with the times, technological advances have become the basis for advancing human civilization. Intelligent systems have been developed to increase efficiency in daily activities. The intelligent system utilizes various techniques such as forecasting, classification, and object detection. The YOLO method is a popular object detection method favoring a system that can work in real-time. Several previous studies have proven the extraordinary performance of YOLO. Health field, YOLO is used to identify breast mass [7], glaucoma detection [8], and skin cancer detection [9]. In other fields, YOLO can outperform other object detection methods, such as Mask R-CNN, in identifying fish freshness [10]. In addition, YOLO is used to detect faces in real-time [11]. Furthermore, YOLO is used for the detection of parking locations [12] to detect pain through baby expressions [13]. The detection system results using YOLO have an average accuracy of more than 90%. This research shows that the YOLO system is suitable for object detection. The YOLO object detection method has several versions that differ in model complexity, detection speed, and system performance level. Several versions include [14], YOLOv2 [15], YOLOv3 [16], YOLOv4 [17], YOLOv5, YOLOv6 [18], and YOLOv7 [19]. Determining a suitable object detection method significantly affects system performance. In addition, it is also essential to introduce datasets before the learning process, such as data availability, number of classes, data variants, etc.In the case of detecting Tajweed objects, each sheet of the mushaf of the Al-Qur'an contains various kinds of Tajweed in different amounts. The difference in the amount of data in each Tajweed class becomes a class imbalance problem that affects the creation of a classification model. The problem of class imbalance in detecting Tajweed objects is difficult to overcome because the dataset's source comes from Al-Qur'an's mushaf. Each sheet of mushaf consists of various kinds of Tajweed, which cannot be separated or multiplied only in certain classes. Therefore, in this study, we handle a small number of classes in datasets with augmentation.

Augmentation increases the data variations, which also affects the performance of the classification system. Many augmentation techniques include adding noise, rotation, and reflection. The research [20] performed augmentation to overcome the limitations of the dataset using Multi Augmentation Techniques - Adaptive Gaussian Convolutional Autoencoder (MAT-AGCA). This research results show that the augmentation application improves the system and gets an accuracy value of 96.29%. Furthermore, the research [21] conducted several augmentation processes to improve image classification performance. Some of them are flip (right, left, up, and down), contrast enhancement, brightness, and main crop. The research [22] introduces three augmentation techniques, namely contrast transformation, brightness adjustment, and Gaussian blur, in the case of road damage detection. Based on several previous studies, image data augmentation techniques are very numerous and widely used in various problems. Determination of a good augmentation technique must be adjusted to the problem to be solved.

Contributions to this study are as follows: solving the problem of small object detection and variations in the form of Tajweed by comparing versions of the YOLO method such as YOLOv5, YOLOv6, and YOLOv7; handling minority class problems by utilizing the HSV color model augmentation process on the Tajweed Dataset; recognizing the different colors of the Tajweed in the Mushaf of Al-Qur'an and improving the performance of the detection system using the HSV color model.

This research is expected to be an initial learning method for the general public about introducing Tajweed through digital image processing. This system can clearly show Tajweed objects in real-time as a reminder when reading the Al-Qur'an. Several experiments were conducted to determine the most optimal detection system model by testing the YOLO method and the augmentation process treatment on the mushaf of the Al-Qur'an dataset.

2. Research Methods

This study detected Tajweed with 10 object classes using the YOLO method. We also propose the use of data augmentation using some modifications of the HSV color model to handle a small number of classes in datasets. The research flow can be seen in Figure 1. The research started with data collection, annotation of 10 Tajweed object classes, distribution of training testing data, data augmentation, and the search for the best object detection model using a comparison of three versions of YOLO (YOLOv5, YOLOv6, and YOLOv7).

The initial stage of the study was data collection by scanning several sheets of the Al-Qur'an mushaf and then saving them in the form of images (.jpg). One of the challenges in detecting Tajweed objects is that the model can recognize Tajweed objects even in different fonts and writing styles of the mushaf Al-Qur'an. Therefore, this research creates a system learning model

DOI: https://doi.org/10.29207/resti.v7i1.4739

to overcome this problem. The data collection consisted of four mushaf with different fonts, times, and regions, but this mushaf was widely used by people in Indonesia. The differences between the four mushaf can be seen in Table 1.

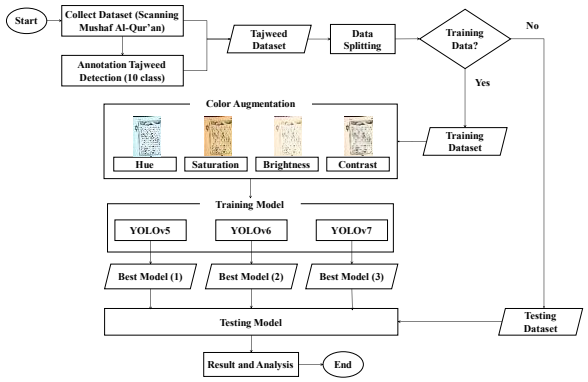


Figure 1. The Flowchat of Tajweed-YOLO system

Table 1. Description of Mushaf Dataset

Mushaf	Publisher	Year	Origin	Image Data
Mushaf 1	PT. Tanjung Mas Majmu'	1980	Semarang	41
Mushaf 2	Khodimain Al Haromain	1992	Madinah	40
Mushaf 3	CV Al-Fatah	2014	Jakarta	26
Mushaf 4	CV Jasa Media	1997	Semarang	62

The next stage is data annotation which is done manually. Annotation creates ground truth objects, which become the target data in the detection system model. Annotation is done based on the Tajweed learning parts with 10 object classes. The annotation results are in the form of a file (.txt) containing class data and the location of the detection object frame. A detailed explanation of the 10 Tajweed class labels is described in Table 2 [23].

Table 2. Description of 10 Tajweed Classes.

Class Number	Tajweed Class	Description
0	Idgham Bighunnah	Idgham Bighunnah, if there is a Nun has the vowel Sukun (نْ) or Tanwin (ً ٍ ٌ) that meets one of the 4 letters, such us ي م ن و.
1	Idgham Bilaghunnah	Idgham Bilaghunnah, if there is a Nun has the vowel Sukun (نْ) or Tanwin (ً ٍ ٌ) that meets one of the 2 letters, such us ر ل.
2	Iqlab	Iqlab if there is a Nun has the vowel Sukun (نْ) or Tanwin (ً ٍ ٌ) that meets ب.
3	Lafdzul Jalalah	Lafdzul Jalalah happens when there is lafadz اللّٰه.
4	Mad Arid LisSukun	Mad Arid Lissukun if there is Mad Thobi'i after which there are letters that are read dead because of Waqof.
5	Mad Iwad	Mad Iwad, if there is a letter that has the vowel Fathahtain (ً) at the end of the sentence (Waqof) other than Ta'marbutho' (ة).
6	Mad Jaiz Munfasil	Mad Jaiz Munfashil when Mad Thobi'i met Hamzah (ء) who was not in one word.
7	Mad Lin	Mad Lin if there is the letter Mad Ya' (ي) or Wawu (و) which has the vowel Sukun, preceded by the letter Fathah (َ) and after it there is Waqof.
8	Mad Wajib Muttashil	Mad Wajib Muttashil when Mad Thobi'i meets Hamzah (ء) in one word.
9	Qalqalah	Qalqalah if there is one of the 6 letters ق ط ب ج د which has the vowel or read dead because it is a Waqaf.

2.1 Dataset for Tajweed Detection

One implementation of the introduction of Tajweed is the reading of the Al-Qur'an mushaf. The mushaf is written in Arabic with various fonts and different writing styles. In this study, we identified two types of writing styles for Al-Qur'an mushaf: Indonesian and Medinan mushaf. There are four different styles of mushaf writing, namely the writing system (Rasm Al-Qur'an), the harakat system (ash-Syakl), the punctuation system (adh-Dhabt), the waqf sign (al-Waqf). In addition to using the Medina Mushaf, this study uses various Indonesian Al-Qur'an Mushaf, where the type of font, sheet color, and differentiating colors indicate the presence of Tajweed. Figure 2 is an example of the Tajweed, Idgham Bighunnah (نْ meets ي) found in each mushaf data.

Figure 2 shows that the Tajweed Idgham Bighunnah is written differently in each mushaf.

In the first mushaf, the font tends to be thicker, and the spacing between letters coincides with that of other mushaf. In the second mushaf, the letter Nun (ن) does not have a vowel, and the letter Ya' (ي) also does not have a tasydid. In the third mushaf, the Tajweed Idgham Bighunnah is marked in red on the letters Nun (ن) and Ya' (ي). The fourth mushaf is written on black and white paper without color. Based on the data analysis, the choice of color augmentation technique using the HSV color model is the right choice. Meanwhile, the use of geometric augmentation techniques in detecting Tajweed objects is considered inappropriate because it can change how the Al-Qur'an is read.

In the Tajweed detection problem, the handling of the imbalance class cannot be solved because the mushaf consists of several Tajweed units with different amounts that cannot be separated. Data augmentation aims to handle a small number of classes in datasets.

DOI: https://doi.org/10.29207/resti.v7i1.4739

This research hopes to increase data variants based on color augmentation to improve the detection system, especially in classes with a small number of classes in datasets.

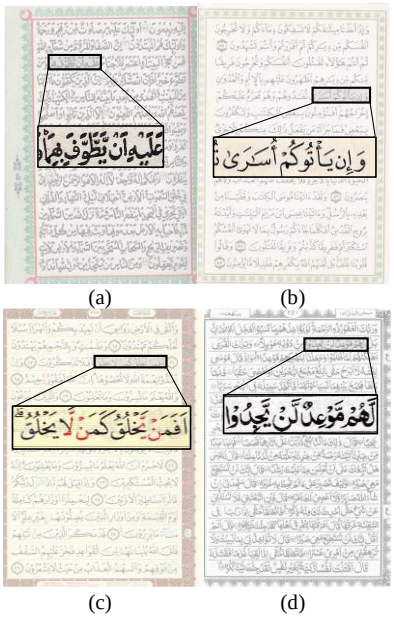


Figure 2. Examples of Idgham Bilaghunnah (a) Mushaf 1, (b) Mushaf 2, (c) Mushaf 3, and (d) Mushaf 4.

2.2 HSV Color Model Augmentation

The Color augmentation technique is an image change based on the HSV (Hue, Saturation, and Value) model, which can be seen in Figure 3.

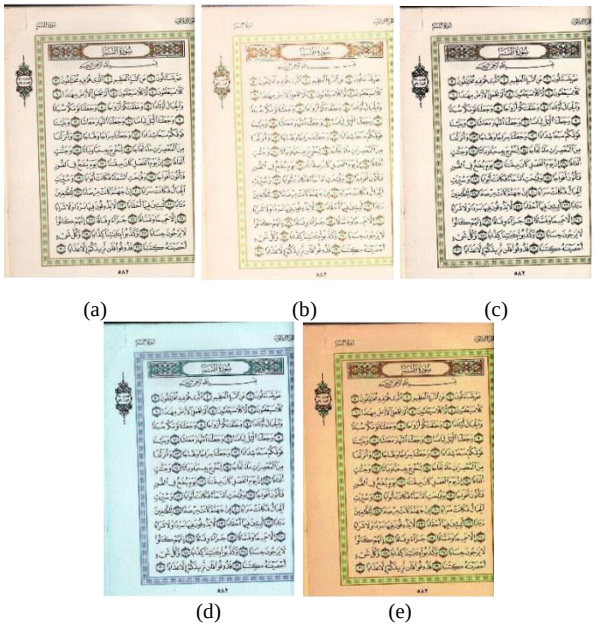


Figure 3. (a) Original Image, (b) Brightness, (c) Contrast, (d) Hue, and (e) Saturation

The HSV color model is widely used in image processing, especially for identifying objects with varied colors [24]. The HSV color model is considered a representation of colors nearly identical to human vision. The HSV color model is modeled in cylindrical coordinates, which change the shape of the RGB color from Cartesian coordinates. RGB consists of three channels: red, green, and blue. Meanwhile, HSV consists of Hue, Saturation, and Value [25].

Brightness is used to change the brightness level of the image. Changes in brightness level can be calculated using Equation (1) and (2).

$$B' = HSV_3 + a \quad (1)$$

$$HSV_3 = \min(\max(0, B'), 1) \quad (2)$$

Equation (1) shows that the image brightness change is done by modifying the Value channel (HSV_3) by entering the parameter a. Then the Value channel (HSV_3) is returned by normalizing the image on the result of adding parameter B' as in Equation (2). If value a is higher, the image quality looks brighter, and vice versa. The value of a can be changed with a value range of -1 to 1 [26].

Contrast is a form of representation of the distribution of dark and light colors in the image. A color will look darker if it is side by side with a light color and vice versa. Changes in the contrast level can be calculated using Equation (3), (4) and (5).

$$C' = mean(HSV_3) \quad (3)$$

$$C = (HSV_3 - C') * a + C' \quad (4)$$

$$HSV_3 = min\,(max(0, C), 1) \quad (5)$$

The first step is to find the mean of the Value channel (HSV_3) as in Equation (3). Furthermore, the value of C' is added to the Value channel (HSV_3) minus $C'$and multiplied by the parameter a as in Equation (4). While Equation (5) above shows the results of normalization in the C result. Contrast changes are made by modifying the Value channel (HSV_3) in the HSV color model [27].

Hue is a color channel that shows authentic tones without any additional white, black, or gray. Changes in the contrast level of a can be calculated using Equation (6).

$$HSV_1 = mod((HSV_1 + a), 1) \quad (6)$$

Equation (6) modifies the Hue channel (HSV_1) according to the channel in the HSV color model. The value of a is filled with a value range of 0 to 1. The calculation then uses the "mod" function as a degree converter between 0^0 to 360^0 to a value with a range of 0 to 1 [26].

Saturation color changes occur in each color (Hue) to gray. The change in the saturation level of a can be calculated using Equation (7) and (8).

$$S' = HSV_2 + a \quad (7)$$

$$HSV_2 = \min(\max(0, S'), 1) \quad (8)$$

DOI: https://doi.org/10.29207/resti.v7i1.4739

Equation (7) modify the saturation according to the HSV color model, namely the Saturation channel (HSV_2). Then the Saturation channel (HSV_2) is returned by normalizing the image on the result of adding parameter S' as in Equation (8). Calculations above the value of a are calculated with values between 0 to 1 [26].

2.3 Object Detection

Object detection is a technology of computer vision related to digital image processing. Object detection consists of 2 processes, namely finding the position of the object and recognizing objects based on the classification results [28]. Each object is studied through the features extracted from the image [29]. These features will provide information on the unique characteristics of each object. The results of object detection in location and object recognition are significant in solving several problems. One widely used object detection method is You Only Look Once (YOLO). The CNN architecture includes the YOLO method because it applies the convolution layer in feature learning. The advantages of YOLO are that it can detect more than one object in one image and has good system performance even though it is done in real-time [30]. YOLO was first introduced by [14]. The first version of YOLO is to introduce an object detection system that has good system performance even though it is done in real-time. In the following year, [15] improved the version of YOLO by experimenting with 9000 different object classes, hence the name YOLO9000. This version has a better and faster system performance than the first version. Furthermore, [16] introduced YOLOv3, which can simultaneously detect large and small objects. Furthermore, YOLOv4 was created by [17] to optimize system performance and object detection accuracy. The YOLO version continued to be developed until YOLOv7.

2.4. YOLOv5

The YOLOv5 method has almost the same architecture as YOLOv4. This model has three important parts: Backbone, Neck, and Head. Architectural details can be seen in Figure 4. The Backbone consists of several convoluted Cross Stage Partial (CSP) processes for feature extraction of the input image. The advantage of CSPNet is optimal feature extraction to enrich information from an image. Furthermore, this feature will enter the Neck to identify the same object but different sizes and scales. The difference in size is usually said to be a feature of the pyramid. The Neck section of YOLOv5 uses PANet to achieve the pyramid feature. The last stage is the Head, where the detection and evaluation of the model are carried out with the results of the object frame and the probability value of each class [31].

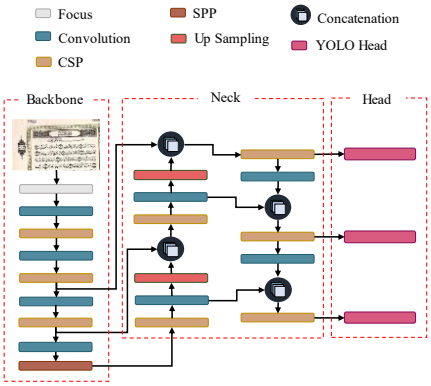


Figure 4. YOLOv5 Archiecture

2.5 YOLOv6

YOLO version 6 was proposed by Meituan [18] with good detection performance results on multi-objects and high inference speed. In YOLOv6, several differences have been improved from the previous version of YOLOv5, namely the implementation of EfficientRep, Rep-PAN, and Head Decoupled. Architectural details can be seen in Figure 5.

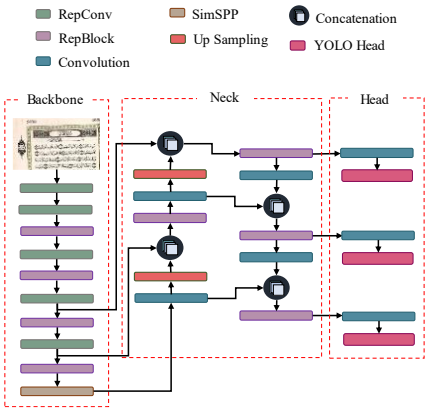


Figure 5. YOLOv6 Archiecture

YOLOv6 uses EfficientRep on the Backbone, which aims to streamline computing work on the GPU to produce a more robust representation than CSP-Backbone.

The Neck section of YOLOv6 implements Rep-PAN to have a good balance between evaluation results and object detection speed. Furthermore, in Head, YOLOv6 improved by proposing Anchor-Free Paradigm, SimOTA Label Assignment, and SioU Bounding Box calculation with Regression Loss [18].

2.6 YOLOv7

The YOLOv7 architecture is different from the previous two methods, which consist of two parts, Backbone and Head. The YOLOv7 Backbone section proposes

DOI: https://doi.org/10.29207/resti.v7i1.4739

"extend" and "multiply scale" in model learning. Architectural details can be seen in Figure 6.

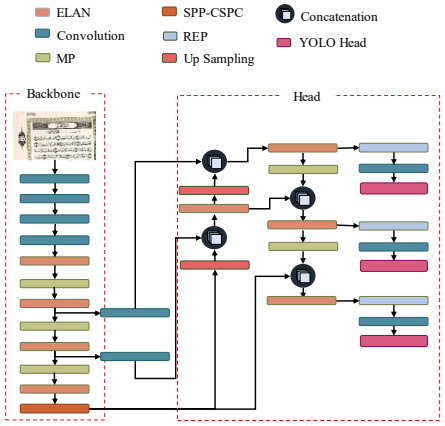


Figure 6. YOLOv7 Architecture

YOLOv7 introduces Extended Efficient Layer Aggregation Networks (E-ELAN), which further optimizes the number of parameters, computation rate, and model density. E-ELAN uses convolution groups to expand the channel and cardinality of compute blocks to learn more features and variations. In addition to E-ELAN, YOLOv7 proposes multiple scaling combining up- and downscaling for a pooling-based model called SPP-CSPC. SPP-CSPC is a combined modification of CSPNet with SPP block. Prior to entering YOLO-Head, this method was proposed by RepConv, which aims to amplify learning outcomes and optimize gradient flow propagation paths. YOLOv7 enables the model to maintain optimal structure and improve the accuracy of object detection models [19].

2.7 Evaluation: mAP

Mean Average Precision (mAP) is an evaluation method used to measure system performance, especially in object detection problems. The results of a suitable detection system will have a high mAP value, indicating a more accurate model. MAP measures the average Average Precision (AP) for each class of objects (N), which can be seen in Equation (10). AP calculation is a combination of recall and precision values, which can be seen in Equation (9). In object detection, the measured evaluation result is the accuracy of the frame or object box against ground truth. This calculation is known as the Intersection over Union (IoU), which describes the overlap between the frame prediction results and ground truth [32].

$$AP = \sum_{k=0}^{k=n-1} (Recall_k - Recall_{k+1}) * Precision(k) \quad (9)$$

$$mAP = \frac{1}{N} \sum_{k=0}^{k=N} AP_k \quad (10)$$

where k is the number of labels for each class and N is the number of Tajweed detection classes.

3. Results and Discussions

In this study, data were collected and labeled manually. Data were collected by scanning the Al-Qur'an mushaf and then annotating them based on 10 Tajweed labels. The annotation results get five variables, namely the Tajweed class, the coordinates of the center of the object (x, y), weight (w), and height (h). These five variables will be the ground truth in the YOLO model. In model training, the dataset is divided into three, namely training, validation, and testing data with a ratio of 70:20:10. Furthermore, the training and validation data are carried out by augmentation processes such as brightness, hue, contrast, and saturation using Equation (1) to Equation (8). The augmentation and without augmentation data results are used to build object detection models using YOLO. The results of data sharing are described in more detail in Table 3.

Table 3. Number of Ground Truth for Tajweed Model Detection

Skenario	Data	Number of Image	Number of Ground Truth in Each Class									
			0	1	2	3	4	5	6	7	8	9
Non-Augmen	Training	118	604	184	88	444	845	276	421	19	399	702
	Validation	34	242	72	41	231	368	87	170	12	130	292
	Testing	17	91	28	29	83	144	31	72	8	49	103
Augmen	Training	532	2966	910	472	2402	4184	1253	2120	113	1889	3437
	Validation	228	1264	370	173	973	1881	562	835	42	756	1533
	Testing	17	91	28	29	83	144	31	72	8	49	103

In this study, we compare the three versions of YOLO, namely YOLOv5, YOLOv6, and YOLOv7. In making the Tajweed object detection model, the number of epochs = 300, batch size = 8, and resizing image = 640. After the learning process, the model's results are tested on data testing with mAP evaluation Equation (10) with IoU threshold = 0.5, which can be seen in Figure 7.

Based on Figure 7, it can be seen that the augmentation treatment greatly influences the performance of the Tajweed detection system. It almost happened that in all three versions of YOLO, the training data evaluation results increased by 50% to 70%, while the test data increased by up to 40% in augmentation data. However, between the training and testing results, there is a significant overfitting of up to 20%. This result proves

DOI: https://doi.org/10.29207/resti.v7i1.4739

that the YOLO architecture needs to be modified to reduce the degree of overfitting of the classification system to the Tajweed detection problem.

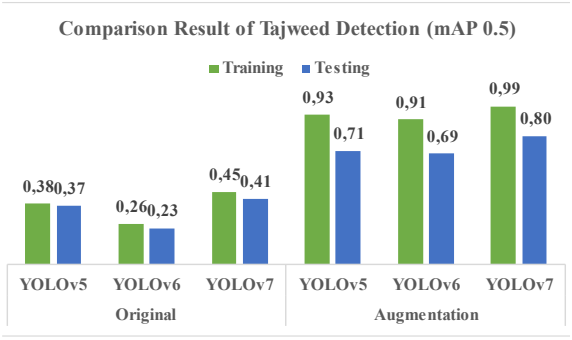


Figure 7. Comparison Result of Tajweed Detection

3.1 Experiments on YOLOv5

In training the tajwid detection model, YOLOv5 requires a learning computation time of 96.60s for the without augmentation dataset and 472.80s for the augmentation dataset. The time difference produces a significant evaluation value. Based on Figure 7, the mAP 0.5 data testing the augmentation model value is 34% better than the without augmentation data. Therefore, data augmentation influences system evaluation and computation time results.

The testing data totaled 17 images with a different number of class labels, as seen in Table 3. Furthermore, Figure 8 compares the evaluation results of mAP, precision, and recall for each class of Tajweed label.

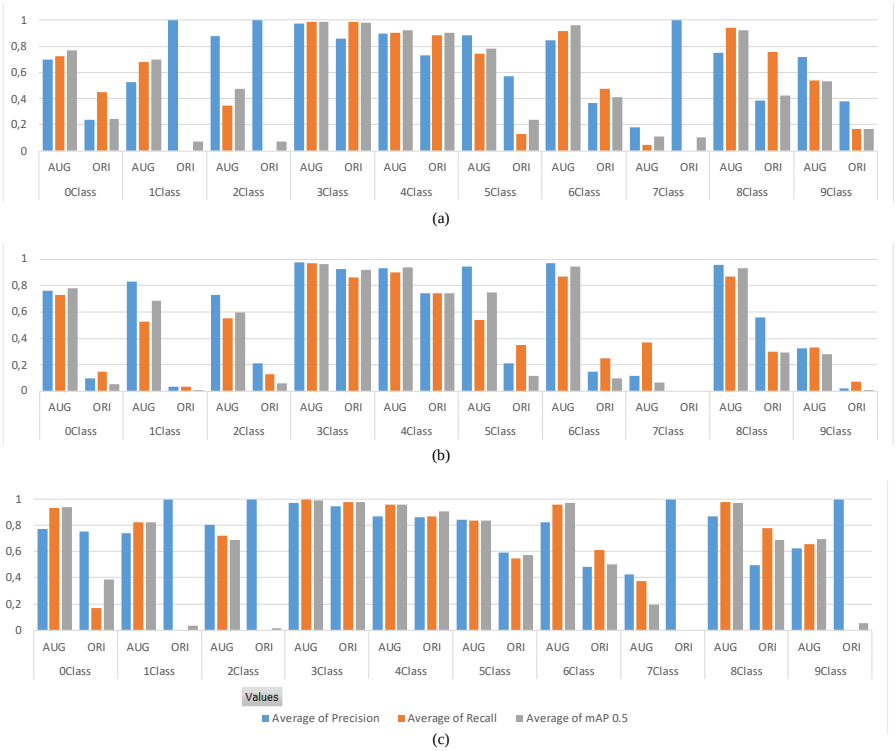


Figure 8. Comparison of Results in Each Tajweed Detection Class (a) YOLOv5, (b) YOLOv6, (c) YOLOv7

Based on Figure 8, it can be seen that the YOLOv5 method is very dependent on the amount of data when creating the detection system. As seen in the results of the without augmentation model, three class labels have low scores: 7th Class, 2nd Class, and 1st Class.

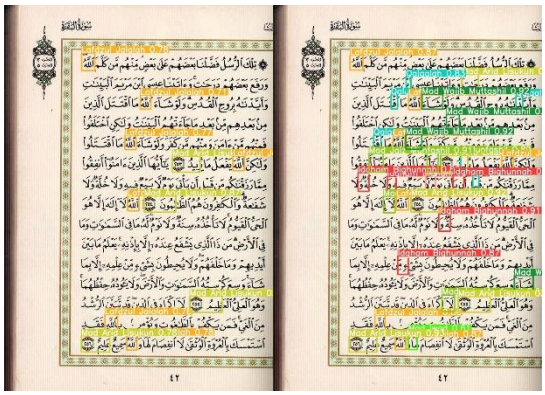
The 3rd Class labels are not small objects in Tajweed detection. Even though small objects also get relatively low results, YOLOv5 considers the number of data objects during training. The results of mAP 0.5, the augmentation model for the entire class label, have increased, where the highest is 62% in the 1st Class.

Next, we analyzed one of the data testing images and compared the results between the without augmentation YOLOv5 model and the augmentation model, as seen in Figure 9.

Based on these results, it can be seen that without augmentation, YOLOv5 can only detect two classes objects: 2nd Class and 4th Class, with a total of objects prediction is 11 labels.

The result of the augmentation model can detect more classes, including tiny objects like the 9th Class. The prediction objects are 33 labels consisting of 3rd Class, 4th Class, 9th Class, 8th Class, 6th Class, 0th Class, 1st Class, and 5th Class. However, there was a detection error on one object in the 5th Class and one object in the 0th Class.

DOI: https://doi.org/10.29207/resti.v7i1.4739



(a) (b)

Figure 9. The Result of Test Data using Yolov5 (a) without Augmentation and (b) with Augmentation Models (IoU=0.5)

3.2 Experiments on YOLOv6

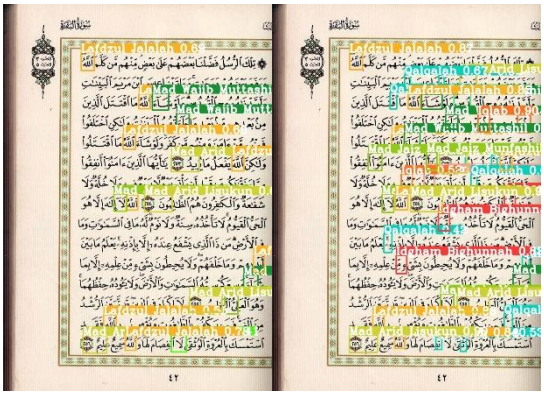
In the training process of the Tajweed detection model using YOLOv6, the difference in computation time is 518.40s longer than the without augmentation model. YOLOv6 requires a long time difference to produce a significant evaluation value between the two.

Based on Figure 7, the value of mAP 0.5 data testing the augmentation model is 46% better than the without augmentation data. Therefore, data augmentation influences system evaluation and computation time results. However, YOLOv6 got the lowest results compared to YOLOv5 and YOLOv7.

Based on Figure 8, it can be concluded that the performance of the YOLOv6 detection system dramatically affects the amount of data. Almost all classes in the evaluation results of the without augmentation model mAP 0.5 model scored below 30% except for the 3rd and 4th Classes. In comparison, the results of the evaluation of the augmentation model experienced a significant increase in all tajwid classes. Therefore YOLOv6 depends not only on the amount of data but also on variations in the shape of objects and small objects. The results of mAP 0.5, the augmentation model for the entire class label, have increased, where the highest is 85% in the 6th Class.

Next, we analyzed one of the data testing images and compared the results between the without augmentation YOLOv6 model and the augmentation model, as seen in Figure 10.

Based on these results, it can be seen that the without augmentation YOLOv6 without augmentation model produces 18 class labels: 3rd Class, 4th Class, 8th Class, 6th Class, and 5th Class. However, there were three label errors, namely in the 3rd Class, 6th Class, and 5th Class. Meanwhile, the YOLOv6 augmented model produces 36 class labels: 3rd Class, 4th Class, 9th Class, 8th Class, 6th Class, 0th Class, 1st Class, and 2nd Class. However, there were nine prediction errors that belonged to the 9th Class, the 2nd Class, and the 0th Class.



(a) (b)

Figure 10. The Result of Test Data using Yolov6 (a) without Augmentation and (b) with Augmentation Models (IoU=0.5)

3.3 Experiments on YOLOv7

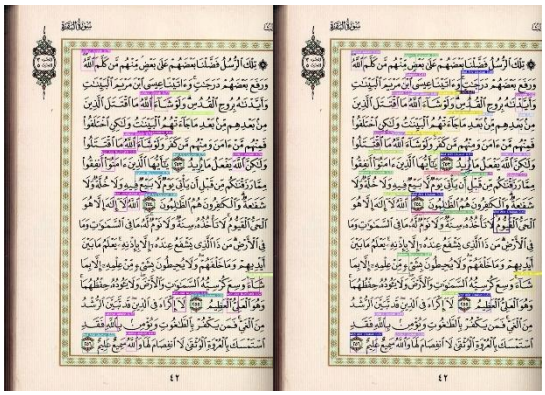
In the YOLOv7 Tajweed detection model's training process, the without augmentation data computation time is 146.40s, while the computation time for the added data is 730.20s. This method requires sizeable computational time and specifications, especially for large amounts of data. However, the detection results obtained the best mAP value compared to the two previous YOLO versions.

In Figure 8, the 7th Class augmentation data evaluation value has increased from the two previous YOLO versions with an mAP value of 0.5 to 0.19. In comparison, the results of the other nine classes get an mAP value of 0.5 between 0.69 and 0.98. The results of testing the tajwid law detection data on YOLOv7 can be seen in Figure 8.

Based on Figure 8, it can be seen that the without augmentation model cannot detect small objects and minority classes, where the 9th Class is wholly undetectable, and other tajwid class labels are. In comparison, the YOLOv7 model with augmentation gets the best results for small and minority object classes and can be appropriately detected.

Based on these results, YOLOv7 is the most optimal method compared to the two previous versions of YOLO in detecting tajwid laws. The results of the detection of the without augmentation YOLOv7 model obtained 22 classes, namely 3rd Class, 4th Class, 8th Class, 6th Class, and 0th Class. However, four labels are misclassified for the 6th Class. While the YOLOv7 augmented model produces 37 class labels consisting of 3rd Class, 4th Class, 9th Class, 8th Class, 6th Class, Id0th Class, 1st Class, 2nd Class, and 7th Class. However, there was only one mislabeled 7th Class out of the entire classes. The results of testing the 17th data testing data can be seen in Figure 11.

DOI: https://doi.org/10.29207/resti.v7i1.4739



(a) (b)

Figure 11. The Result of Test Data using Yolov7 (a) without Augmentation and (b) with Augmentation Models (IoU=0.5)

Based on Figure 12, the results of the confusion matrix in the YOLOv7 augmentation model. The most challenging class to detect is the 7th Class because many of the fundamental truths of these classes are detected as background (undetectable) at 0.50 and detected as 4th Class at 0.33. Therefore it is necessary to modify the model for handling small datasets and the detection model, which focuses on small objects and datasets.

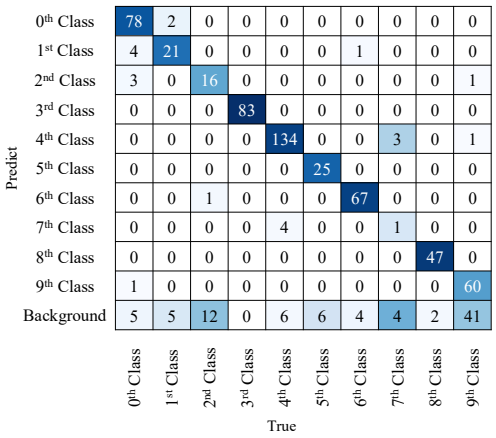


Figure 12. Confusion Matrix of YOLOv7 Augmentation Model

4. Conclusion

This study detected the law of Tajweed with 10 label classes, including the part of Nun Sukun or Tanwin, Mad law, Lafdzul Jalalah, and Qalqalah. A newly constructed dataset is used to build a detection system by comparing the three latest versions of YOLO, namely YOLOv5, YOLOv6, and YOLOv7. In improving the Tajweed detection work system, an augmentation process is carried out based on the HSV color model: Brightness, Contrast, Hue, and Saturation. Based on the test results, the best model for detecting Tajweed is the YOLOv7 method with data augmentation. The results of the mAP 0.5 value are 99% on the training data and 80% on the testing data. However, the computational time required in the training process is the highest (730.20s).

In the problem of detecting Tajweed objects, it is tough to handle a small number of classes in datasets. As depicted in Figure. 7, the three selected YOLO versions trained on the augmented dataset achieved higher mAP compared to the ones that were trained without the augmentation process. Therefore, the results of the augmentation model can improve the detection system. Based on the experiment, the most challenging task for the law of Tajweed detection problem is Mad Lin.

Based on several experiments, the augmentation process can improve the performance of the YOLO model to classify the minor class but still does not get significant improvement. Especially on the 7th Class label (Mad Lin) with an mAP 0.5 of 19%. After further data analysis, the data on the 7th Class label has a high variation with a small amount of data. High variation makes it more difficult for the model to provide predictions for this class label.

Suggestions for further research are to handle variations in data in the Tajweed classes by trying other data augmentation methods, such as the Mosaic or the Cutmix method, to improve imbalanced data. In addition, the detection model could be improved by modifying the Backbone as an image feature extraction to handle data variations. Furthermore, modifying the Anchor parameter as a measurement of the size of the detection object could also be done to improve the model.

Reference

[1] V. Maarif, H. M. Nur, W. Rahayu, M. Informatika, and T. Informatika, "Aplikasi pembelajaran ilmu tajwid berbasis android," *J. Evolusi*, vol. 6, no. 1, pp. 91–100, 2018.

[2] M. A. Amir, *Ilmu Tajwid Praktis*. Pustaka Baitul Hikmah Harun Ar-Rasyid, 2019.

[3] M. Lubis and A. R. Lubis, "Classification of Tajweed Al-Qur'an on Images Applied Varying Normalized Distance Formulas," in *Proceedings of the 3rd International Conference on Electronics, Communications and Control Engineering*, 2020, pp. 21–25.

[4] T. A. Zuraiyah, S. Madenda, R. A. Salim, and R. Noviana, "Tajweed Segmentation Using Pattern Recognition, Extraction and SURF descriptor Algorithms," in *IOP Conference Series: Materials Science and Engineering*, 2020, vol. 846, no. 1, p. 12022.

[5] R. Rizal, B. Bustami, and D. Azzahra, "Pendeteksi Tajwid Idgham Mutajanisain Pada Citra Al-Qur'an Menggunakan Fuzzy Associative Memory (FAM)," *TECHSI-Jurnal Tek. Inform.*, vol. 11, no. 3, pp. 395–407, 2019.

[6] A. Noeman and D. Handayani, "Detection of Mad Lazim Harfi Musyba Images Uses Convolutional Neural Network," in *IOP Conference Series: Materials Science and Engineering*, 2020, vol. 771, no. 1, p. 12030.

[7] G. H. Aly, M. Marey, S. A. El-Sayed, and M. F. Tolba, "YOLO Based Breast Masses Detection and Classification in Full-Field Digital Mammograms," *Comput. Methods Programs Biomed.*, vol. 200, p. 105823, 2021.

[8] R. C. Joshi, M. K. Dutta, P. Sikora, and M. Kiac, "Efficient Convolutional Neural Network Based Optic Disc Analysis Using Digital Fundus Images," in *2020 43rd International*

DOI: https://doi.org/10.29207/resti.v7i1.4739

Conference on Telecommunications and Signal Processing (TSP), 2020, pp. 533–536.
[9] J. Chhatlani, T. Mahajan, R. Rijhwani, A. Bansode, and G. Bhatia, "DermaGenics-Early Detection of Melanoma using YOLOv5 Deep Convolutional Neural Networks," in *2022 IEEE Delhi Section Conference (DELCON)*, 2022, pp. 1–6.
[10] E. Prasetyo, N. Suciati, and C. Fatichah, "A comparison of yolo and mask r-cnn for segmenting head and tail of fish," in *2020 4th International Conference on Informatics and Computational Sciences (ICICoS)*, 2020, pp. 1–6.
[11] W. Chen, H. Huang, S. Peng, C. Zhou, and C. Zhang, "YOLO-face: a real-time face detector," *Vis. Comput.*, vol. 37, no. 4, pp. 805–813, 2021.
[12] E. Tanuwijaya and C. Fatichah, "Penandaan Otomatis Tempat Parkir Menggunakan YOLO Untuk Mendeteksi Ketersediaan Tempat Parkir Mobil Pada Video CCTV," *Briliant J. Ris. dan Konseptual*, vol. 5, no. 1, pp. 189–198, 2020.
[13] T. Abuzairi, N. Widanti, A. Kusumaningrum, and Y. Rustina, "Implementasi Convolutional Neural Network Untuk Deteksi Nyeri Bayi Melalui Citra Wajah Dengan YOLO," *J. RESTI (Rekayasa Sist. Dan Teknol. Informasi)*, vol. 5, no. 4, pp. 624–630, 2021.
[14] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.
[15] J. Redmon and A. Farhadi, "YOLO9000: better, faster, stronger," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 7263–7271.
[16] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv Prepr. arXiv1804.02767*, 2018.
[17] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," *arXiv Prepr. arXiv2004.10934*, 2020.
[18] C. Li *et al.*, "YOLOv6: a single-stage object detection framework for industrial applications," *arXiv Prepr. arXiv2209.02976*, 2022.
[19] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," *arXiv Prepr. arXiv2207.02696*, 2022.
[20] N. P. Sutramiani, N. Suciati, and D. Siahaan, "MAT-AGCA: Multi Augmentation Technique on small dataset for Balinese character recognition using Convolutional Neural Network," *ICT Express*, vol. 7, no. 4, pp. 521–529, 2021.
[21] N. Song and Q. Du, "Classification of cervical lesion images based on CNN and transfer learning," in *2019 IEEE 9th International Conference on Electronics Information and Emergency Communication (ICEIEC)*, 2019, pp. 316–319.
[22] W. Wang, B. Wu, S. Yang, and Z. Wang, "Road damage detection and classification with Faster R-CNN," in *2018 IEEE international conference on big data (Big data)*, 2018, pp. 5220–5223.
[23] R. M. I. Rusyd, *Panduan Praktis & Lengkap Tahsin, Tajwid, Tahfiz Untuk Pemula*. Laksana, 2019.
[24] N. Hassan, K. W. Ming, and C. K. Wah, "A Comparative Study on HSV-based and Deep Learning-based Object Detection Algorithms for Pedestrian Traffic Light Signal Recognition," in *2020 3rd International Conference on Intelligent Autonomous Systems (ICoIAS)*, 2020, pp. 71–76.
[25] V. Popov, M. Ostarek, and C. Tenison, "Practices and pitfalls in inferring neural representations," *Neuroimage*, vol. 174, pp. 340–351, 2018.
[26] W. Chen, W. Shen, L. Gao, and X. Li, "Hybrid Loss-Constrained Lightweight Convolutional Neural Networks for Cervical Cell Classification," *Sensors*, vol. 22, no. 9, p. 3272, 2022.
[27] Q. Al-Jubouri, R. J. Al-Azawi, M. Al-Taee, and I. Young, "Efficient individual identification of zebrafish using Hue/Saturation/Value color model," *Egypt. J. Aquat. Res.*, vol. 44, no. 4, pp. 271–277, 2018.
[28] Z. Zou, Z. Shi, Y. Guo, and J. Ye, "Object detection in 20 years: A survey," *arXiv Prepr. arXiv1905.05055*, 2019.
[29] C. Guo, B. Fan, Q. Zhang, S. Xiang, and C. Pan, "Augfpn: Improving multi-scale feature learning for object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 12595–12604.
[30] J. Du, "Understanding of object detection based on CNN family and YOLO," in *Journal of Physics: Conference Series*, 2018, vol. 1004, no. 1, p. 12029.
[31] Q. Xu, Z. Zhu, H. Ge, Z. Zhang, and X. Zang, "Effective Face Detector Based on YOLOv5 and Superresolution Reconstruction," *Comput. Math. Methods Med.*, vol. 2021, 2021.
[32] J. Yu, Y. Jiang, Z. Wang, Z. Cao, and T. Huang, "Unitbox: An advanced object detection network," in *Proceedings of the 24th ACM international conference on Multimedia*, 2016, pp. 516–520.

DOI: https://doi.org/10.29207/resti.v7i1.4739