ICIC Express Letters
Volume **16**, Number **11**, November **2022**

ICIC International ©2022 ISSN 1881-803X
pp. 1147–1155

INDIVIDUAL BALL POSSESSION DETECTION AND TRACKING IN FUTSAL USING MODIFIED DEEP LEARNING

Rianto Setiobudi[1] and Suharjito[2,*]

[1]Computer Science Department, BINUS Graduate Program – Master of Computer Science
[2]Industrial Engineering Department, BINUS Graduate Program – Master of Industrial Engineering
Bina Nusantara University
Jl. K. H. Syahdan No. 9, Kemanggisan, Palmerah, Jakarta 11480, Indonesia
rianto.setiobudi@binus.ac.id; *Corresponding author: suharjito@binus.edu

Received November 2021; accepted January 2022

Abstract. *Ball and futsal player detection is complex because the object detected has high mobility and small size. In deep learning models, datasets have an important role in increasing accuracy and frame rate. The current deep learning object detection method mostly uses augmentation in image merging, color saturation, and image changes but does not use histogram equalization. In this study, the proposed method combines color saturation modification with the Contrast Limited Adaptive Histogram Equalization (CLAHE) method. The image saturation modification process was carried out to change the brightness, contrast, and sharpness levels before applying the CLAHE method. Models were trained using the transfer learning method with pre-training weight darknet53.conv.74 and yolov4.conv.137 in You Only Look Once architecture. Intersection Over Union was used to determine individual ball possession. The dataset in this study amounted to 4,400 images obtained from the converted frames of video futsal. The model trained using a pre-processed dataset with modified image color saturation in the form of brightness, contrast, sharpness, and CLAHE was able to shorten the average training time to 7.9% and increase the average frame per second capability by 1.5% compared to the model trained using only the CLAHE method.*

Keywords: Individual ball possession, Deep learning, You Only Look Once (YOLO), Intersection Over Union (IOU), Contrast Limited Adaptive Histogram Equalization (CLAHE)

1. **Introduction.** Ball possession is a futsal strategy that has a role in determining a winner [1]. Ball possession in futsal will determine the team's adaptation level in facing obstacles that arise in a match. Detecting the ball object's trajectory [2] is not as easy as detecting the movement of a car because a car moves in one direction following the traffic while the ball moves dynamically. Players could also be tracked using existing models and four cameras [3] and detected using the Euclidean calculation method. The weakness of the calculation using the Euclidean distance with the camera intersection is that the camera's position must be fixed, causing some detection objects to be invisible or image bias, resulting in inaccurate calculation of the position of the detection object.

The contribution of this study lies in obtaining a useful dataset to produce a good deep learning model performance and determine individual ball possession. Pre-processing is used to get a useful dataset, namely by modifying image color saturation and increasing the contrast with the CLAHE method. Then, to track a ball and futsal players, YOLO architecture with appropriate activation modification for augmented [4] dataset is used to detect objects' presence. A replica of the bounding box is later generated to calculate the Intersection Over Union (IOU) value.

DOI: 10.24507/icicel.16.11.1147

The paper is organized as follows. In Section 1, we introduce ball possession, and the problems faced. In Section 2 briefly discuss areas related to our work. Section 3 discusses our proposed approach about ball possession and YOLO architecture used. In Section 4, we give the experimental analysis and results. Finally, we conclude in Section 5.

2. **Related Work.** In this study, we will use a deep learning approach to solve the problem, IOU is used for the ball possession approach, and CLAHE is used so that the dataset is more general when applied to the model to be made.

2.1. **Transfer learning.** Among the most widely used models, the one with better reliability than its predecessors is YOLO [5]. YOLOv3 performance had been improved by creating pre-training weight darknet53.conv.74, which had been trained on the darknet53 backbone using the ImageNet dataset. YOLOv3 can detect objects faster than other object detection methods with an image size of 416 having a mean Average Precision-50 (mAP-50) and good speed. Therefore, this study proposed an image size of 416 to detect ball and futsal player objects through the transfer learning method.

2.2. **Intersection Over Union (IOU).** IOU in deep learning is useful in measuring performance of object detection models. In its use, IOU is not only useful in calculating evaluation metrics [6], but many also use it for other purposes. A study uses IOU to calculate the area of a forest [7] and visualize objects in 3D. Individual ball possession could also be calculated using a calculation based on several objects' proximity. IOU was also used to track the ball's whereabouts in soccer. Therefore, we used IOU to calculate individual ball possession in this study.

2.3. **Contrast Limited Adaptive Histogram Equalization (CLAHE).** Adaptive Histogram Equalization (AHE) is a solution to increase image contrast by referring to the image histogram's corresponding pixels and later enhanced with CLAHE [8]. CLAHE is useful when dealing with colors with high haze levels (degrees of grayscale) [9] and low light conditions [10]. This study was conducted because most of the existing deep learning architectures do not use the histogram equalizer image in the data augmentation [11,12]. Therefore, in this study, CLAHE was used as a pre-processing of the dataset.

3. **Proposed Approach.** As shown in Figure 1, the proposed methods are a pre-processing method, where the dataset was modified by image saturation and the CLAHE method. YOLO architecture with appropriate activation modification for augmented dataset was used to compare the model's performance. It was used to obtain bounding box coordinates for the ball and futsal players and calculate individual ball possession.

3.1. **CLAHE with saturation modification.** Images only modified with CLAHE still lost their details in pixels due to the increased contrast. Dataset modification was done first by changing the brightness, contrast, and sharpness of the image. This was done when the CLAHE method was applied. For evaluation metrics, the Absolute Mean Brightness Error (AMBE) and Peak Signal-to-Noise Ratio (PSNR) were used. AMBE can show the modified image's proximity to the original image [13]. AMBE was used to measure the brightness level. Formula (1) shows that the AMBE measurement takes the absolute value from reducing the average brightness value on the original image $E[X]$ and the modified image $E[Y]$.

$$AMBE = |E[Y] - E[X]| \tag{1}$$

$$E[X] = \sum_{i=0}^{K-1} f_i h_i \Bigg/ \sum_{i=0}^{K-1} f_i \tag{2}$$

$$E[Y] = \sum_{i=0}^{K-1} f_i h_i \Big/ \sum_{i=0}^{L-1} f_i \tag{3}$$

K is the maximum gray level image; f_i is frequency of gray level; and h_i is a gray level image.

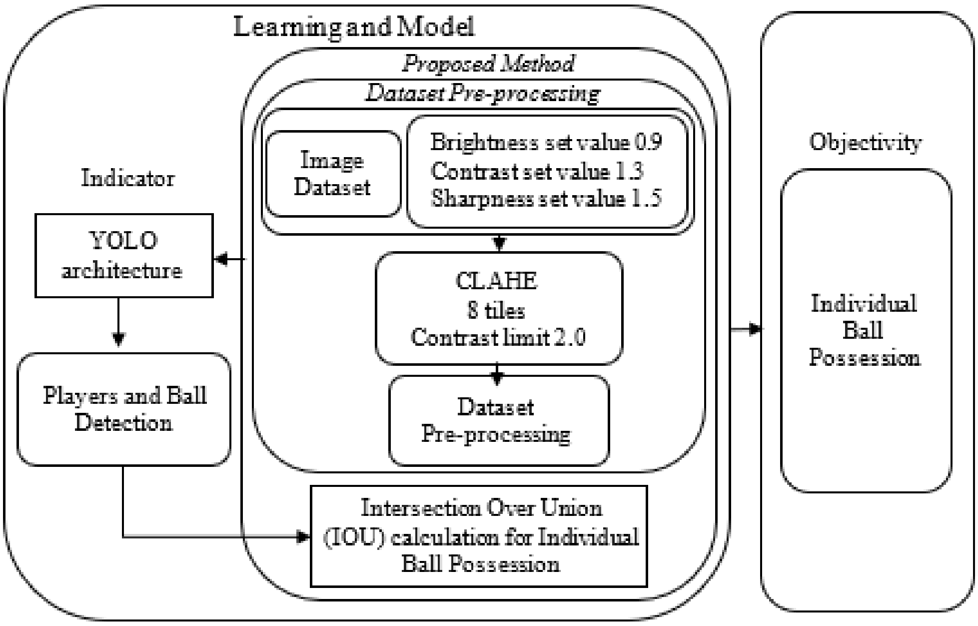


Figure 1. Proposed methods

Peak Signal-to-Noise Ratio (PSNR) was used to measure the contrast level of an image [14]. Formula (5) was used to measure PSNR. The image was calculated by comparing the modified image with the original image. The higher the PSNR value, the better the result of image modification.

$$MSE = \frac{1}{m \times n} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} A(i,j) \tag{4}$$

$$PSNR = 20 \cdot \log_{10} \left(\frac{MAX_1^2}{\sqrt{MSE}} \right) \tag{5}$$

MSE is mean square error; m is maximum gray level image; i is value of gray; and MAX is the pixel maximum value, which is 255.

3.2. **YOLO architecture.** YOLO is a deep learning architecture designed to work fast [15]. However, YOLO has an overfitting problem when learning starts. To minimize this, in this study, we used ReLU as an activation function at the beginning of convolutional replacing leaky as a standard activation. ReLU is good for augmented data [4]. ReLU is not used for all convolutional because, in research, it has a high learning rate [16].

As shown in Figure 2, we replace leaky with ReLU on the second layer before down sample. The model is evaluated by Microsoft Common Objects in Context (MS COCO) dataset as a benchmark to measure the accuracy. Formula (9) shows the mAP calculation.

$$P = \frac{TP}{TP + FP} \tag{6}$$

$$Rc = \frac{TP}{TP + FN} \tag{7}$$

$$AP = \sum_{k=0}^{k=n-1} \left[Rc\,(k) - Rc(k+1)\right] * P(k) \tag{8}$$

$$mAP = \frac{1}{n} \sum_{k=1}^{k=n} AP_k \tag{9}$$

P = Precisions, Rc = Recalls, AP = Average Precision, mAP = mean Average Precisions, TP = True Positive, FP = False Positive, FN = False Negative, k = Number of datasets.

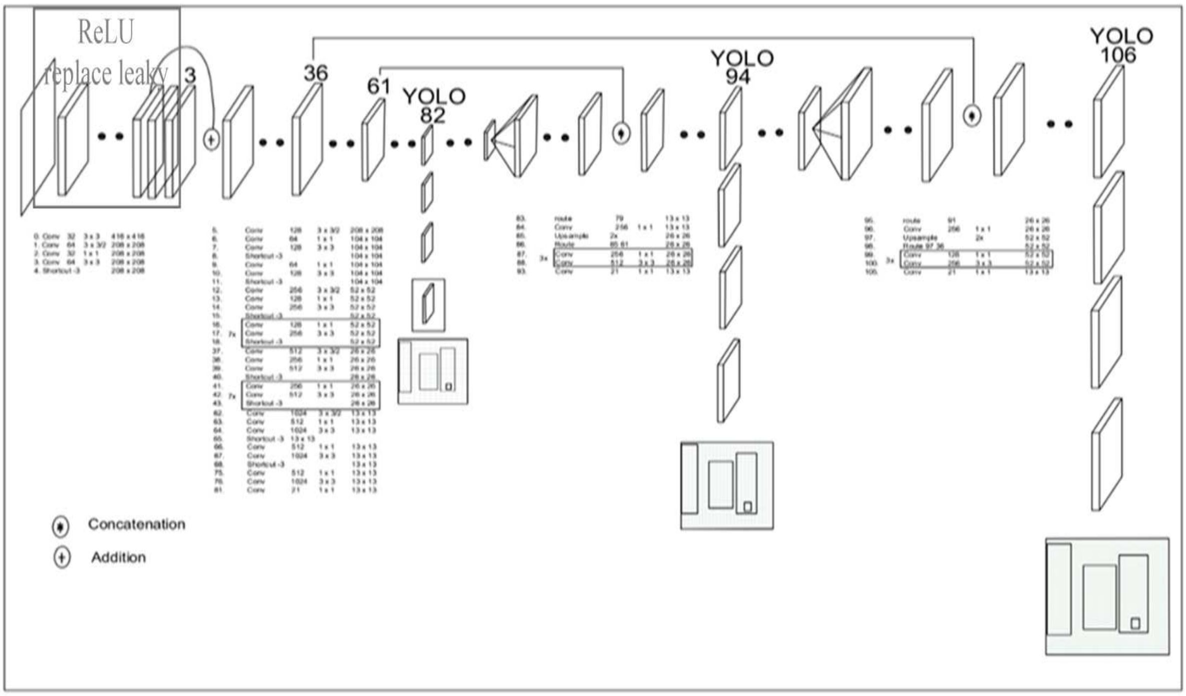


FIGURE 2. Modification YOLO architecture

3.3. **Individual ball possession detection.** In this study, replica IOU was used to calculate the overlapping threshold bounding box values between futsal players and the ball, which aimed as a basis for ball possession. Figure 3 shows the existence of a replica of a ball bounding box. A bounding box replica is a replica of a ball bounding box placed parallel to or right at the bottom corner of the player's bounding box, which is in contact with the ball's bounding box. If more than one player controls the ball, as shown in Figure 3(b), the biggest IOU is the player who owns it.

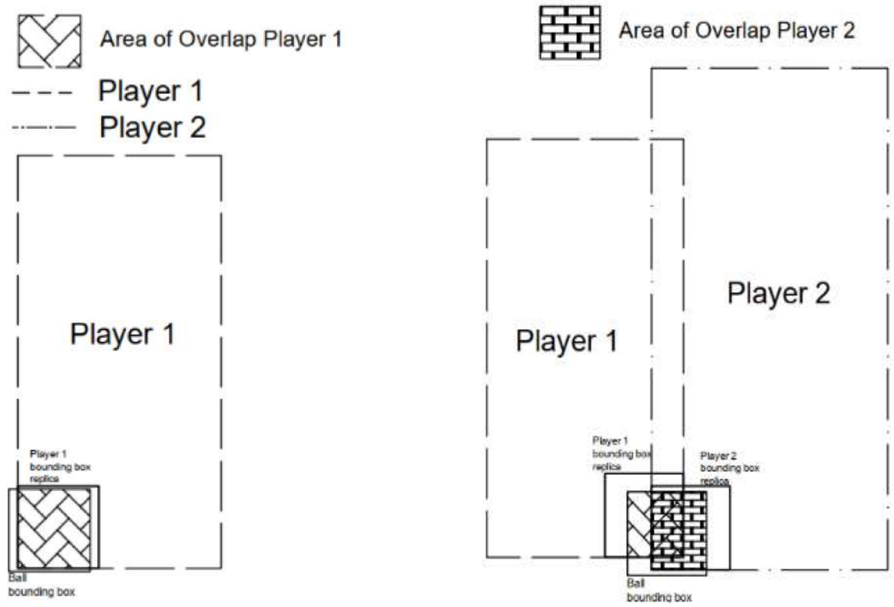
Individual ball possession was only observed in an image. In Formulas (11) and (12), the numbers 70% and 13% were determined to get an optimal IOU above 50% for the ball and player's intersection and provide a threshold value for the ball's bounding box, and 70% is the threshold where the bounding box of the ball is in the player's bounding box. The overlapping threshold bounding box formula is as follows.

$$IOU = \frac{A \cap B}{A \cup B} = \frac{TP}{(TP + FP + FN)} \tag{10}$$

$$(BBW_{Ball} \cap BBW_{Player}) > 70\% \tag{11}$$

$$(BBH_{Ball} \cap BBH_{Player}) < 13\% \tag{12}$$

BBW = Bounding Box Wide, BBH = Bounding Box High. TP, FP and FN for IOU threshold are calculated.



(a) One futsal player for individual ball

(b) Individual ball possession determination on more than one futsal player bounding box that overlaps with the ball bounding

FIGURE 3. Determination of individual ball possession

3.4. **Tracking.** The tracking results will be used as the basis for detecting the presence of the same player in the previous frame of a video. Tracking is done by looking at the bounding box, the results of the corrected model on each video frame, and the bounding box overlaps between the active frame and the previous frame [17]. Euclidean distance will be used as a threshold by measuring the distance of the ball and the distance of the players [18]. Bounding boxes for futsal players who are successfully detected and overlap each frame, then the IOU value is calculated as shown in Formula (13).

$$IOU = \frac{|BBa \cap BBb|}{|BBa \cup BBb|} \tag{13}$$

BBa = Current Frame Player Bounding Box; BBb = Previous Frame Player Bounding Box.

The calculation of evaluation metrics as validation of individual ball possession is based on the IOU approach [19]. The replica ball bounding box will be placed within the player's bounding box. Figure 4 describes the method used to validate individual ball possession. The IOU of the replica bounding box for each model is then calculated for its precision using Formula (10). The individual ball possession intersection in Formula (14) is obtained from the overlapping model trained using MS COCO and the model using the own dataset. Intersection own dataset model is the total area of the intersection own dataset model.

$$Precision\ IBP = \frac{Intersection\ Individual\ Ball\ Possession}{Intersection\ Own\ Dataset\ Model} \tag{14}$$

4. **Results and Evaluation.** In this study, the brightness, contrast, and sharpness levels were carried out before processing the dataset using the CLAHE method and YOLO architecture. The model was trained using an i7 6700K processor, NVIDIA GTX 1070Ti, and 32GB 3000MHz RAM. The model was executed using the TensorFlow framework, YOLO algorithm, OpenCV, and Pillow Library.

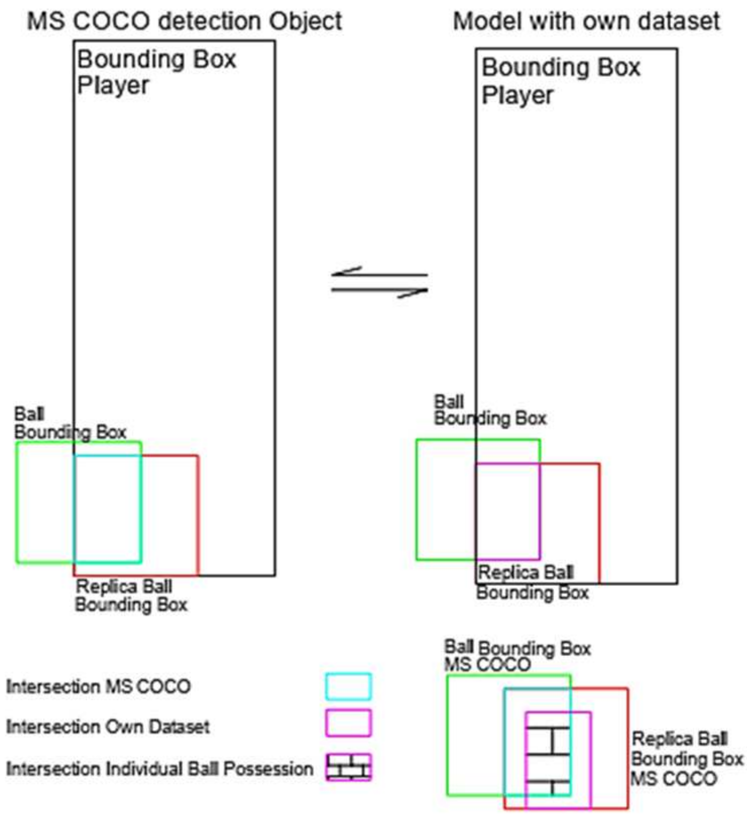


FIGURE 4. How to calculate precision of individual ball possession

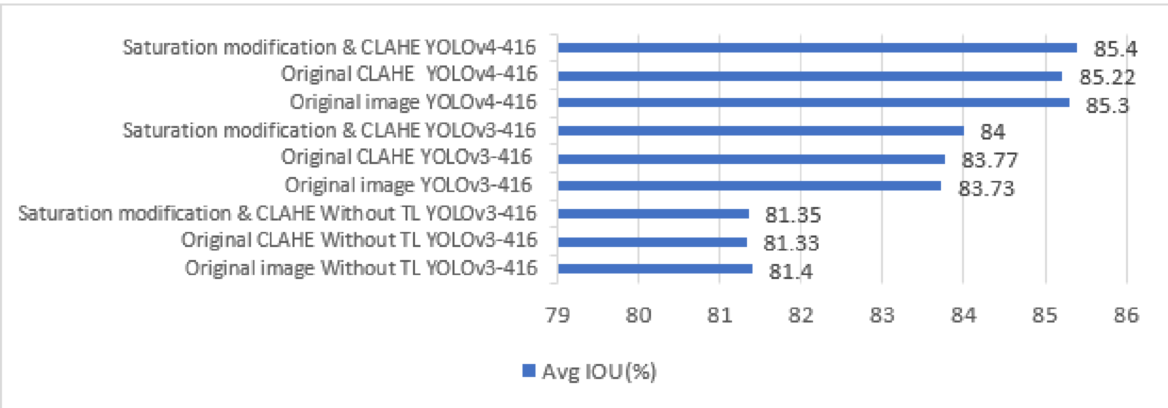
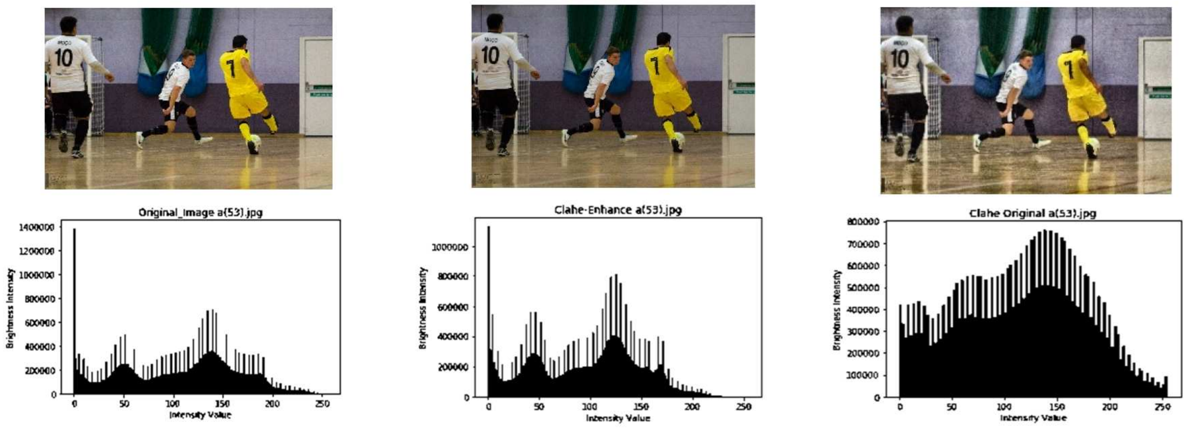


FIGURE 5. Ground truth on several methods

4.1. **Dataset collection.** Dataset is an essential feature in making a good model [20]. In this study, the dataset contains 4,400 images with manual annotations on two classes, ball, and futsal players, collected from frames of video extraction of futsal in several matches from the video-sharing website www.youtube.com. The dataset was released in https://www.kaggle.com/riantosetiobudi/ball-futsalplayer-enhanceclahe.

The useful dataset has a high IOU value in the range of 0.3 (30%) – 0.6 (60%). If higher, then better. Modification of image saturation and CLAHE was carried out first and resulted in a size that was 77.13% smaller than the original dataset, shortening training time and better dataset quality [21]. In this study as seen in Figure 5, all models trained using this dataset could get scores above 80% and improve the deep learning model's performance [6]. YOLOv4 using image saturation modification and CLAHE methods produced the highest average IOU value to 85.4%.

The color composition in the histogram of an image was expected to be filled in all tonal histogram images ranging from 0 to 255. However, the best field is when tonal highlight histogram does not dominate [22]. As shown in Figure 6(b) there has a range of tonal highlight histograms shorter than the range level in the original image in Figure 6(a).



(a) Original image and histogram of image

(b) Saturation modification, CLAHE image and histogram of image

(c) Original CLAHE image and histogram of image

FIGURE 6. Histogram image comparison on the original image, original CLAHE method, and saturation modification with CLAHE

Table 1 shows a comparison of the evaluation metrics on the dataset that has been pre-processed. The AMBE value in the pre-processing dataset using image saturation modification can be seen in the pre-processing dataset. The CLAHE method has a smaller value than the images processed with the CLAHE method. Thus, the brightness condition is closer to the original image. The image has a decrease in brightness quality but does not cause bad quality to be used as a dataset. This loss of quality helps shorten training time.

TABLE 1. Evaluation metrics of the dataset

Method	PSNR	AMBE
Original CLAHE	29.14	10.26
Modification saturation and CLAHE	29.33	10.06

4.2. **Training model.** An overview of the performance comparison of several models that have been trained can be seen in Table 2. Models trained using saturation modification images and CLAHE method as pre-processing dataset spend 7.9% faster training time and increase the frame rate to 1.5% of the CLAHE pre-processed dataset method without modification saturation image. The higher the frame rate, the better the model's performance in detecting the ball and futsal players. It can also be seen that the best model in this study is the model trained using YOLOv3 with saturation modification and CLAHE method, where this model is 20.32% [23] and 10.75% [24] better than previous studies and even though YOLOv4 has a good IOU average value, its mAP accuracy is not the best. Also, its average fps is still far behind YOLOv3.

4.3. **Individual ball possession detection.** As Figure 7 shows that when a player's bounding box that is detected overlaps with the ball bounding box and reaches the IOU threshold, the player's object is shaded, indicating that the player has control of the ball. The individual ball possession calculation method has also succeeded in determining the individual ball possession on two-player bounding boxes.

Table 2. Performance comparison

Preprocessing	Transfer learning	Model CNN	IBP precision (%)	mAP (%)	Avg fps	Train time (min)
Original image	Without TL	YOLOv3-416	97.24	98.74	42.7	1560
Original CLAHE	Without TL	YOLOv3-416	80.78	98.19	42.9	1527
Saturation modification & CLAHE	Without TL	YOLOv3-416	69.10	98.26	43.9	1559
Original image [15]	darknet53.conv.74	YOLOv3-416	83.98	98.92	42.8	3511
Original CLAHE	darknet53.conv.74	YOLOv3-416	79.58	98.93	43.0	3520
Saturation modification & CLAHE	**darknet53.conv.74**	**YOLOv3-416**	**79.81**	**98.95**	**43.3**	**3037**
Original image	yolov4.conv.137	YOLOv4-416	80.30	98.90	39.9	4330
Original CLAHE	yolov4.conv.137	YOLOv4-416	87.84	98.89	40.0	4203
Saturation modification & CLAHE	yolov4.conv.137	YOLOv4-416	87.82	98.93	40.0	4080

Figure 7. Individual ball possession detection

5. **Conclusion.** The best performing model is a model trained using a previously processed dataset using the modified color saturation and CLAHE methods and the darknet53.conv.74 pre-training weight. Modification of image saturation and CLAHE shortened the average model training processing time to 7.9%. Because the size of the dataset becomes smaller, 77.13% than the original dataset. The frame rate increased to 1.5% of the CLAHE pre-processed dataset method without modification saturation image, and have 20.32% and 10.75% better than previous studies.

Acknowledgment. This work is supported by Computer Science Department, BINUS Graduate Program – Master of Computer Science, Bina Nusantara University.

REFERENCES

[1] A. S. Muharram and M. Faruk, Ball possession and shooting analysis of the 2016 AFF Suzuki Cup win (A case study of Thailand and Indonesia), *Jurnal Prestasi Olahraga*, vol.1, no.1, 2019.

[2] X. Yu, C. Xu, H. W. Leong, Q. Tian, Q. Tang and K. W. Wan, Trajectory-based ball detection and tracking with applications to semantic analysis of broadcast soccer video, *Proc. of the 11th ACM International Conference on Multimedia*, Berkeley, CA, USA, pp.11-20, 2003.

[3] E. Morais, A. Ferreira, S. A. Cunha, R. M. Barros, A. Rocha and S. Goldenstein, A multiple camera methodology for automatic localization and tracking of futsal players, *Pattern Recognition Letters*, vol.39, pp.21-30, 2014.

[4] U. Aiman and V. P. Vishwakarma, Face recognition using modified deep learning neural network, *2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, IIT Delhi, India, pp.1-5, 2017.

[5] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, You only look once: Unified, real-time object detection, *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, pp.779-788, 2016.

[6] Z. Huang, J. Zhan, H. Zhao, K. Lin, P. Zheng and J. Lv, Real-time visual tracking base on siamRPN with generalized intersection over union, *International Conference on Brain Inspired Cognitive Systems*, Guangzhou, China, pp.96-105, 2019.
[7] H. Choi, H. J. Lee, H. J. You, S. Y. Rhee and W. S. Jeon, Comparative analysis of generalized intersection over union and error matrix for vegetation cover classification assessment, *Sensors and Materials*, vol.31, no.11, pp.3849-3858, 2019.
[8] A. Vidyarthi, J. Shad, S. Sharma and P. Agarwal, Classification of breast microscopic imaging using hybrid CLAHE-CNN deep architecture, *2019 12th International Conference on Contemporary Computing (IC3)*, Noida, India, pp.1-5, 2019.
[9] L. C. Valeriano, J. B. Thomas and A. Benoit, Deep learning for dehazing: Comparison and analysis, *2018 Colour and Visual Computing Symposium (CVCS)*, Gjøvik, Norway, pp.1-6, 2018.
[10] M. F. Rahman, F. Sthevanie and K. N. Ramadhani, Face recognition in low lighting conditions using fisherface method and CLAHE techniques, *2020 8th International Conference on Information and Communication Technology (ICoICT)*, Yogyakarta, Indonesia, pp.1-6, 2020.
[11] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu and A. C. Berg, SSD: Single shot multibox detector, *European Conference on Computer Vision*, Amsterdam, The Netherlands, pp.21-37, 2016.
[12] A. Bochkovskiy, C.-Y. Wang and H.-Y. M. Liao, YOLOv4: Optimal speed and accuracy of object detection, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Online Virtual Hosting, pp.126-138, 2020.
[13] S. D. Chen and A. R. Ramli, Preserving brightness in histogram equalization based contrast enhancement techniques, *Digital Signal Processing*, vol.14, no.5, pp.413-428, 2004.
[14] S. D. Chen and A. R. Ramli, Contrast enhancement using recursive mean-separate histogram equalization for scalable brightness preservation, *IEEE Transactions on Consumer Electronics*, vol.49, no.4, pp.1301-1309, 2003.
[15] J. Redmon and A. Farhadi, YOLOv3: An incremental improvement, *Computer Vision and Pattern Recognition*, DOI: 10.48550/arXiv.1804.02767, 2018.
[16] Z. Hu, J. Zhang and Y. Ge, Handling vanishing gradient problem using artificial derivative, *IEEE Access*, vol.9, pp.22371-22377, 2021.
[17] P. R. Kamble, A. G. Keskar and K. M. Bhurchandi, A deep learning ball tracking system in soccer videos, *Opto-Electronics Review*, vol.27, no.1, pp.58-69, 2019.
[18] D. Link and M. Hoernig, Individual ball possession in soccer, *PloS One*, vol.12, no.7, e0179953, 2017.
[19] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid and S. Savarese, Generalized intersection over union: A metric and a loss for bounding box regression, *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, Long Beach, CA, USA, pp.658-666, 2019.
[20] S. Oh, A new dataset evaluation method based on category overlap, *Computers in Biology and Medicine*, vol.41, no.2, pp.115-122, 2011.
[21] H. A. Alhaija, S. K. Mustikovela, L. Mescheder, A. Geiger and C. Rother, Augmented reality meets computer vision: Efficient data generation for urban driving scenes, *International Journal of Computer Vision*, vol.126, no.9, pp.961-972, 2018.
[22] J. R. Boyack and A. K. Juenger, *U.S. Patent No.5,724,456*, U.S. Patent and Trademark Office, Washington, D.C., 1998.
[23] Q. C. Mao, H. M. Sun, Y. B. Liu and R. S. Jia, Fast and efficient non-contact ball detector for picking robots, *IEEE Access*, vol.7, pp.175487-175498, 2019.
[24] X. Zhang, T. Zhang, Y. Yang, Z. Wang and G. Wang, Real-time golf ball detection and tracking based on convolutional neural networks, *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Toronto, Canada, pp.2808-2813, 2020.