Article

Improvement in Error Recognition of Real-Time Football Images by an Object-Augmented AI Model for Similar Objects

Junsu Han 1, Kiho Kang 1 and Jongwon Kim 2,*

1 Department of Mechatronics Engineering, Korea University of Technology and Education, Cheonan 31253, Republic of Korea
2 Department of Electromechanical Convergence Engineering, Korea University of Technology and Education, Cheonan 31253, Republic of Korea
* Correspondence: kamuiai@koreatech.ac.kr

Abstract: In this paper, we analyze the recognition error of the general AI recognition model and propose the structure-modified and object-augmented AI recognition model. This model has object detection features that distinguish specific target objects in target areas where players with similar shapes and characteristics overlapped in real-time football images. We implemented the AI recognition model by reinforcing the training dataset and augmenting the object class. In addition, it is shown that the recognition rate increased by modifying the model structure based on the recognition errors analysis of general AI recognition models. Recognition errors were decreased by applying the modules of HSV processing and differentiated classes (overlapped player groups) learning to the general AI recognition model. We experimented in order to compare the recognition error reducing performance with the general AI model and the proposed AI model by the same real-time football images. Through this study, we have confirmed that the proposed model can reduce errors. It showed that the proposed AI model structure to recognize similar objects in real-time and in various environments could be used to analyze football games.

Keywords: object recognition; real-time recognition; object augmentation; improvement in error recognition; structure of AI models

1. Introduction

In the field of vision research, the recognition of target objects using Artificial Intelligence (AI) is a highly active research area. In general, recognizing an object using an image sensor or video camera is the task of processing a set of semantic cases and detecting various features of the target object within the image.

Traditional gradient-based object recognition methods distinguish target objects by detecting characteristic changes that exist in the image through manipulating the local information, such as the target image’s brightness, color, and texture [1]. Previous research has been conducted in the direction of utilizing characteristic changes, such as edge detection [2], blob detection [3], and corner detection [4], to improve the image processing method for object recognition. Recently, various artificial intelligence techniques, based on Convolutional Neural Network (CNN), for automatically recognizing objects in fields of digital image processing have been applied. The high-performance detection models have been implemented in various forms, from model structures such as R-CNN [5], Fast-RCNN [6], Faster-RCNN [7], and RetinaNet [8], to model algorithms such as SSD [9] and YOLO [10].

In general, the object detection model uses a detection algorithm to determine the recognition area (object-box) that contains the recognition object within the detection area and classifies the target object within the recognition area. To conduct the object recognition process using the AI algorithm, select objects from the background image and compare the location (x, y), height, and width (h, w) information in the recognition area with the features of the pre-trained object. Then, determine the target object. When the recognition
process is completed, the target object’s values of the location (x, y), height, and width (h, w) are ensured as feature information to recognize a person, as shown in Figure 1.

![Figure 1. Information of object coordinates within an image.](image1)

The general AI recognition model classifies one object as a person, as shown in Figure 2a, or an area where two or more objects overlap as a person, as shown in Figure 2b.

![Figure 2. Examples of object classification according to recognition range: (a) person recognition of one object; (b) person recognition of overlapped areas.](image2)

Objects can be overlapped or blurred, especially in real-time images of people crowded close together. This is a major cause of the recognition error, making it difficult to accurately classify objects and damaging the feature information of the object. In addition, nonsensical information, such as lines around objects, things, brightness, and shadows, act as negative factors to distinguish objects.

This study uses various real-time player images from football videos as target objects. In the target objects, groups of overlapping people are separately classified as recognizable target objects. With this, we study the performance-enhancing structures and methods of AI models for recognizing people within groups of similar people. Therefore, in image sensing applications using AI models, the performance of object detection can be quantified by examining the recognition errors that occur while individually classifying a person in the recognition area.

When a target object is a person, numerous factors can cause recognition errors. The characteristic of a person in the recognition area is, approximately, a 1.5 fixed aspect ratio, and the range of scale changes, according to perspective, is also substantial. Moreover, the camera’s shooting angle and a person’s behavioral characteristics change the features of a person’s object. Therefore, various object recognition methods for distinguishing object information from surrounding information have been studied.
For the target object’s recognition, in the case of overlap between players, the target objects could be distinguished by recognizing each player’s uniform number through images from various angles, using multiple camera viewpoints [11]. However, in classifying a specific target object in the overlapping area, the recognition method by changing the camera angle is not appropriate, as shown in Figure 3. This is because the higher the similarity of feature information in the overlapping object, the more often the target object cannot recognize individually.

![Figure 3](image-url)

**Figure 3.** Examples of changing camera angle in football video: (a) Angle of view from the audience; (b) Angle of view from behind the goal post.

As a supplementary method, depth information was added to the object feature information (RGB data) collected by multiple camera viewpoints, using cameras such as Kinect or stereo settings [12]. By implementing a lightweight single-pass convolutional neural network architecture with a fused information source, the detection accuracy and location tracking performance are improved compared to single-view camera images. In addition, feature extraction methods utilizing body-worn inertial measurement units (IMU) [13] and LIDAR sensors [14] have frequently been studied. However, the methods mentioned so far are not suitable for real-time object detection environments, such as football games, due to the limitations of the moving speed, distance measurement range, and lighting environment. In addition, the zoom-in-out range manipulation of the camera on the football image changes the size of the target object and the recognition accuracy together. For this reason, when an AI detection model is trained with limited feature information, such as distant objects, as shown in Figure 4a, or small objects, as shown in Figure 4b, various object recognition errors occur.

![Figure 4](image-url)

**Figure 4.** Examples of object recognition errors due to limited feature information: (a) Missing object recognition; (b) Incorrect object recognition.

In real-time football images, correctly recognizing a player as an individual is a valuable issue. However, recognition errors frequently occurred when classifying target objects of one person in a crowded area. In the error case of object recognition, the same identification is assigned to a similar player according to the frame change, as shown in Figure 5.
When various motion changes occur before and after the overlapped target object with similar characteristics in the two-dimensional space, the object recognition model using AI has a high rate of misrecognition and non-recognition errors in the real-time object recognition process. To improve this, we implemented a multi-class object recognition model with the HSV color space conversion processing and compared the recognition performance with the general AI models. In addition, if the target object has a similar shape within the corresponding recognition area or overlap, it becomes the main factor of misrecognition and non-recognition. Therefore, by devising and applying the HSV module to the processing structure of the general AI recognition model, we reduced the misrecognition and non-recognition errors of objects with similar shapes. Then, characteristics within groups of similar objects were added to the HSV model as unique data for learning classes. In this paper, the final AI model for recognizing multi-class objects improved the recognition errors caused by rapid changes and overlaps of similar objects.

2. Methods

2.1. Preparation of the Training Dataset

In general, image preprocessing methods are used to prepare training data to improve the learning effects of AI models. The image data acquired in a limited time is insufficient for model learning, which increases the cost function value and reduces its predictive performance. Image preprocessing methods, such as the standardization of images and clarification of recognized results, are used in the general-purpose, low-performance hardware-based detection model to overcome the environmental limitations of image acquisition.

In this research, we extract the unique feature information of objects from images limited by geometric transformation methods and use it as new data for AI models to learn. Image geometric transformation includes simple data reinforcement methods such as flipping, cropping, rotation, translation, color space, and noise injection. According to Table 1, image cropping is the most accurate geometric transformation method of image manipulation.

Table 1. Results of Taylor and Nitschke’s geometric transformation experiments (units: %) [15].

| Geometric Transformation | Top-1 Accuracy | Top-5 Accuracy |
|--------------------------|----------------|----------------|
| Baseline                 | 48.13 ± 0.42   | 64.50 ± 0.65   |
| Flipping                 | 49.73 ± 1.13   | 67.36 ± 1.38   |
| Rotating                 | 50.80 ± 0.63   | 69.41 ± 0.48   |
| Cropping                 | 61.95 ± 1.01   | 79.10 ± 0.80   |
| Color Jittering          | 49.57 ± 0.53   | 67.18 ± 0.42   |
| Edge Enhancement         | 49.29 ± 1.16   | 66.49 ± 0.84   |
| Fancy PCA                | 49.41 ± 0.81   | 67.54 ± 1.01   |

As shown in the evaluation results [15], reported in terms of Top-1 and Top-5 accuracy, the cropping significantly improves the performance of the CNN tasks. Accuracy is also called Top-1 accuracy to distinguish it from Top-5 accuracy, common in Convolutional Neural Network evaluation [16].
We selected image crop tools as a data preprocessing method to prepare an efficient training dataset. Yolo Mark [17] is the object bounding box cropper from images to extract efficiently object feature information. The experiment datasets were tested through the GEFORCE RTX 3060 D6 12G GPU calculation, based on 1280 × 720 resolution in the K3 Korean national football game video. We also set the COCO [18] mean average precision (mAP50) at 55.3% and 30 FPS to compare the object recognition errors in the implemented AI models.

The proposed AI models were selected randomly within 10% of the 3482 football images as training data, and the remaining 90% were used as test data. We labeled the training data with four type classes (A, B, C, D) in the Yolo Mark. Through this data segmented process, in addition to players (A, B) and referee (C) detected based on the uniform’s colors, overlapped objects were marked as a new class (D) and unlabeled objects were re-marked.

As shown in Figure 6a, the characteristic data is extracted by marking the bounding box according to the color of individual player’s uniforms. Then, various overlapped objects are selected, as shown in Figure 6b, and the obtained reinforced training datasets are denoted by new classes.

![Figure 6. Object bounding box marking samples of training data using Yolo Mark: (a) Samples of Class A (Dark Green box), Class B (Light Green box), and Class C (Red box) marking; (b) Samples of Class D (Sky Blue box) marking.](image)

### 2.2. Modification and Implementation of AI Models

In a football game, players play a complex role as individual performance and tactical team members, and referees play their role as game operator. While they played their part, various errors occurred in the object detection, and this became a topic to be solved in this study. In addition, it can be seen from Table 2 that the detection model based on the YOLO algorithm, which has the highest response speed and accuracy, is suitable when considering frequent changes in the movement of players to recognize objects in real-time.

| Detection Model | mAP (%) | Time (ms) | Detection Model | mAP (%) | Time (ms) |
|-----------------|---------|-----------|-----------------|---------|-----------|
| SSD321          | 28.0    | 61        | RetinaNet-50-500| 32.5    | 73        |
| DSSD321         | 28.0    | 85        | RetinaNet-101-500| 34.4    | 90        |
| R-FCN           | 29.9    | 85        | RetinaNet-101-800| 37.8    | 198       |
| SSD513          | 31.2    | 125       | Yolov3-320      | 28.2    | 22        |
| DSSD513         | 33.2    | 156       | Yolov3-416      | 31.0    | 29        |
| FPN FRCN        | 36.2    | 172       | Yolov3-608      | 33.0    | 51        |

According to the Yolov3 tech report [19], Yolov3-320, 416, and 608 models are fast and accurate compared with other detection models. The three types of Yolov3 detection models have different performance characteristics depending on the application target environment. The selected Yolov3-416 model was the best-performing model in this study. This is because speed, accuracy, and the target image size for recognition are the criteria of choice in real-time object recognition, such as for football games. The Yolov4 and Yolov5 models were released with no significant change in their algorithms and structure. However,
performance differences depend on GPU computing resources at the time of release. In this study, we focused on improving object recognition by revising the model structure and method in limited hardware resources rather than applying the newly released AI model.

Among the various versions of YOLO-based detection models, the Yolov3-416 model structure is shown in Figure 7.

![Figure 7. The architecture of the Yolov3-416 model and YOLO Loss [20].](image)

The YOLO detection model aggregates pixels in the convolution layer to form object-specific features and make predictions based on the loss function output at the network end. We changed this to detect only one person class among 80 class objects. Therefore, the general AI model’s architecture consists of an algorithm that recognizes players and referees as a person.

2.2.1. Structural Modification, Yolov3-HSV Model

In the RGB images, the object information is represented by three unique color values of red, green, and blue properties. In addition, to detect a specific object in the image, all color values of R (0~255), G (0~255), and B (0~255) must be considered. On the other hand, the HSV image is represented by information based on human color perception with three properties: Hue, Saturation, and Value [21]. The expression range of the information for classifying the uniqueness of an object in the HSV image is H (0~360), S (0~1), and V (0~1). This color space conversion improves the object recognition accuracy by making it easier to classify colors than in RGB images.

The Yolov3-HSV model recognizes players with HSV color information by masking three color types of uniforms [22]. It is a similar object recognition model to the Yolov3-416 model’s structure modified, as shown in Figure 8.

![Figure 8. Model structure modified from Yolov3 to Yolov3-HSV [22].](image)

We made it easy to distinguish object information within the image through color mask processing that limits the range of specific colors, as shown in Figure 9.
By uniquely specifying the target objects’ minimum and maximum color ranges within the image only once, we compared whether the players’ H, S, and V color values were in the range. Depending on the presence in the range, the mask matrix element value 1 or 0 is determined correspondingly. Through this process, three color mask matrices were created and applied as a mask to the football images.

The players’ color information accurately represented the pixel value in the image as color and intensity through the color space conversion from RGB to HSV, as shown in Figure 10b. Then, they were divided into three classes based on the uniform’s color. The object color information was extracted by filtering the players with masks for red, blue, and white. As a result, it was classified into three colors (red: Class A, blue: Class B, white: Class C), as shown in Figure 10c.

**Figure 9.** The object’s HSV color mask process.

| Color (Object) | Hue   | Saturation | Value  |
|----------------|-------|------------|--------|
| Red (A)        | 5~179 | 130~255    | 100~255|
| Blue (B)       | 110~179| 130~255    | 100~255|
| White (C)      | 0     | 0          | 0~255  |

**Figure 10.** Examples of filtering red, blue, and white colors from an image: (a) Original RGB image; (b) HSV color conversion image; (c) Color filtered image.

2.2.2. Class Augmentation, Yolov3-Augment Model

In the overlap area, various changes were implemented in the recognition and detection situations, such as front and rear relationships, number of objects, and color contrast occur according to the player’s movement. Consequently, the AI model learning is limited in recognizing and classifying overlapping objects using only the person object, as shown in Figure 11a. Therefore, setting the overlap area as a new single object reduced the uncertainty of the object detection by grouping numerous variables and subdividing them into additional recognition areas.
Figure 10. Examples of filtering red, blue, and white colors from an image: (a) Original RGB image; (b) HSV color conversion image; (c) Color filtered image.

2.2.2. Class Augmentation, Yolov3-Augment Model

In the overlap area, various changes were implemented in the recognition and detection situations, such as front and rear relationships, number of objects, and color contrast according to the player’s movement. Consequently, the AI model learning is limited in recognizing and classifying overlapping objects using only the person object, as shown in Figure 11a. Therefore, setting the overlap area as a new single object reduced the uncertainty of the object detection by grouping numerous variables and subdividing them into additional recognition areas.

Figure 11. Recognition object type of AI Models; (a) Yolov3-416 model (b) Yolov3-Augment model.

In the object class augmentation model shown in Figure 11b, we added the recognition object class to the Yolov3-HSV model by classifying the overlapping areas of the players as class D. As a result, the Yolov3-Augment model improves the recognition performance between similar objects in various object detection situations by supplementing the object’s feature information through recognition class augmentation.

The object recognition procedure of the proposed AI model is shown in Figure 12. The AI models evaluated the recognition results in the process of classifying objects (person, player, and overlap player) via unique training weights with different average loss, as shown in Figure 13. Finally, we compared the recognition error reduction performance of the Yolov3-Augment model, including similar objects with many errors as recognition categories, with the general AI recognition model. We evaluated the object recognition performance of the Yolov3-416, Yolov3-HSV, and Yolov3-Augment models in the same real-time football images.

Figure 12. The object recognition procedure of the AI recognition models.
performance of the Yolov3-416, Yolov3-HSV, and Yolov3-Augment models in the same real-time football images.

Figure 12. The object recognition procedure of the AI recognition models. (a)

Figure 13. Average loss results of AI models training on different iterations: (a) Average loss 1.1770 after training YOLOv3-HSV model; (b) Average loss 0.5875 after training YOLOv3-Augment model.

2.3. Error Criteria and Evaluation Items

There is a generalized measure methodology for evaluating recognition performance, according to a class classification method and class types that constitute a recognition model [23]. However, we do not evaluate the generalized recognition accuracy of the classes themselves. In addition, this study does not include a classification method according to the type of recognition algorithm. The reason for this is that the three AI models with the same recognition algorithm have different procedures and structures for object recognition; therefore, the features of the occurred errors are important.

In this study, we evaluate how many different recognition errors can occur under the same conditions for three types of AI recognition models that have completed model training for an object class with a similar shape to the defined classification method.

In the problem of statistical classification, the error matrix [24] is a classification table layout that evaluates the performance of an object recognition AI model. The unit-object recognition is divided into two stages. The error stage can be classified, as shown in Table 3. It is the result of subdividing each error category into YES or NO, according to the clarity of the object recognition and classification.

| Actual Class | Predicted Class | Object YES | Object NO |
|--------------|----------------|------------|-----------|
| Object YES   | True Positive  | False Negative |
| Object NO    | False Positive| True Negative |

We divide object recognition errors into a False Positive that is incorrectly recognized and a False Negative that is non-recognized in the classification category. However, we did not define True Positive and True Negative categories as recognition errors. The reason for this is that a True Positive is an object correctly recognized, and a True Negative is a non-object that it is not recognized. The experiment includes all of the errors that occur in
the process of recognizing objects (predicted class) and classifying unit objects (actual class) within the object recognition area (object-box).

False Positive Errors are recognition errors in which the object detection model incorrectly predicted the actual object as another object. It is the result of object misrecognition, in which overlap areas or long distances predict target objects differently or additionally. False Negative Errors are recognition errors that do not predict objects because the object detection model cannot detect the objects. It is the result of object non-recognition that does not predict target objects in areas where object overlap, and object separation has occurred or is at a long distance.

The performance of the object detection model was evaluated by the Precision function (1), related to object misrecognition, and the Recall function (2), related to object non-recognition. Subsequently, it comprehensively evaluated the F1 score function (3), the Accuracy function (4), the Error Rate function (5), and the Specificity (6).

These are model evaluation functions:

- **Precision** indicates how accurate the predicted class is.
  \[ Precision = \frac{TP}{TP + FP} \] (1)

- **Recall** indicates how well the actual class was predicted.
  \[ Recall = \frac{TP}{TP + FN} \] (2)

- **F1 Score** is the harmonic mean of Precision and Recall.
  \[ F1 \text{ score} = \frac{2 \times (Recall \times Precision)}{Recall + Precision} \] (3)

- **Accuracy** is the probability that the prediction class is correct in all data.
  \[ Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \] (4)

- The **error rate** is the probability that the prediction class is incorrect in all data.
  \[ Error \text{ Rate} = \frac{FN + FP}{TP + TN + FP + FN} \] (5)

- **Specificity** is known as True Negative Rate (TNR).
  \[ Specificity = \frac{TN}{TN + FP} \] (6)

### 3. Evaluation Results

#### 3.1. Results of Object Recognition

The dataset used in the AI model evaluation selected a total of 900 frames in the three videos of different football games. Then, the prediction accuracy of the proposed AI model is visualized through a classification table.

When augmented from binary classes that classify a person to multi-classes that classify players, the confusion matrix consists of 3-classes in the Yolov3-HSV model and 4-classes in the Yolov3-Augment model. The object classification results of the AI models are shown in Figure 14.

In the model evaluation, when the player was subdivided, based on the uniform’s color, it tended to increase the number of False Positive recognition results. As shown in the False Negative recognition results of the Yolov3-HSV model in Figure 15c, misrecognition errors increased further compared to the Yolov3-416 model. The causes of the various errors are as follows. The left and right upper and lower overlaps of multiple objects were recognized as a single object, and individual objects were identified as different objects after the overlapping object. In addition, small distant objects are recognized as something other than what they were, and sometimes as duplicates.
The error rate is calculated as follows:

\[ \text{Error Rate} = \frac{\text{False Positive} + \text{False Negative}}{\text{True Positive} + \text{True Negative}} \] (1)

The accuracy is calculated as follows:

\[ \text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{True Positive} + \text{True Negative} + \text{False Positive} + \text{False Negative}} \] (2)

The recall is calculated as follows:

\[ \text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \] (3)

The precision is calculated as follows:

\[ \text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \] (4)

The error rate is relatively reduced by 48.39%. The average object recognition results are shown in Figure 14. The Yolov3-HSV model showed improved results in reducing the error rate of object recognition more than the other models.

Yolov3-HSV model increased by 60.38% compared to the Yolov3-416 model. In addition, False Positive Errors of the Yolov3-Augment model were increased by 30.54% compared to the Yolov3-416 model; however, False Positive Errors of the Yolov3-Augment model were improved in Accuracy and Specificity.

The misrecognition of the Yolov3-HSV model increased by 30.54% compared to the Yolov3-416 model. However, False Positive Errors of the Yolov3-Augment model were relatively reduced by 36.59% in the Yolov3-Augment model.

As shown in the False Negative recognition results of the Yolov3-Augment model in Figure 15d, non-recognition errors tended to decrease as the models recognized multiple classes. The reason for this is that overlapping objects were identified as individual objects after separation. In addition, recognition occurred even when small objects were at a long distance or when the spacing between the objects was narrow.

At the total object recognition, shown in Figure 16a, the non-recognition of the Yolov3-HSV model increased by 60.38% compared to the Yolov3-416 model. In addition, False Negative Errors were relatively reduced by 36.59% in the Yolov3-Augment model. The misrecognition of the Yolov3-HSV model increased by 30.54% compared to the Yolov3-416 model. However, False Positive Errors of the Yolov3-Augment model were...
relatively reduced by 48.39%. The average object recognition results of the three AI models are shown in Figure 16b.

![Figure 16](image_url)

**Figure 16.** Summation of objects recognition (a) Total object recognition; (b) Average recognition.

### 3.2. Performance of the AI Models

We compared the relative superiority of each evaluation item according to the recognition result in all AI models, as shown in Figure 17. The Yolov3-Augment model, overall, improved in Accuracy and Specificity to other models. In particular, the Yolov3-Augment model showed improved results in reducing the error rate of object recognition more than the other models.

![Figure 17](image_url)

**Figure 17.** Relative performance comparison of the AI models (relative ratio units: %).

We pre-emptively improved the object recognition of the Yolov3-HSV model compared to the Yolov3-416 model, resulting in the effect of accuracy and error improvement in the Yolov3-Augment model. The macro-average results of the AI models are specified in Table 4.
Table 4. The macro average performance of the AI models (units: %).

| Models          | Items       | Precision | Recall    | F1 Score  | Accuracy | Error Rate | Specificity |
|-----------------|-------------|-----------|-----------|-----------|----------|------------|-------------|
| Yolov3-416      |             | 93.06     | 94.27     | 93.66     | 88.08    | 11.92      | 0.00        |
| Yolov3-HSV      |             | 80.74(−12.32) | 84.00(−10.27) | 81.92(−11.74) | 88.07(−0.01) | 11.93(+0.01) | 90.41(+90.41) |
| Yolov3-Augment  |             | 89.80(−3.26) | 80.85(−13.43) | 82.96(−10.70) | 94.57(+6.49) | 5.43(−6.49) | 95.97(+95.97) |

In the class-specific performance, shown in Table 5, the Yolov3-Augment model is superior to the other AI models. However, the Recall and F1 Score of overlapping objects, named Class D, are lower than in the other AI models.

Table 5. Class-specific performance of the AI models (units: %).

| Models          | Class      | Items       | Precision | Recall    | F1 Score  | Accuracy | Error Rate | Specificity |
|-----------------|------------|-------------|-----------|-----------|-----------|----------|------------|-------------|
| Yolov3-416      | PERSON     |             | 93.06     | 94.27     | 93.66     | 88.08    | 11.92      | 0.00        |
| Yolov3-HSV      | Class A    | 85.75       | 86.38     | 86.06     | 86.45     | 13.55    | 86.53      |
|                 | Class B    | 79.99       | 70.23     | 74.79     | 83.20     | 16.80    | 90.33      |
|                 | Class C    | 76.49       | 95.41     | 84.90     | 94.54     | 5.46     | 94.38      |
| Yolov3-Augment  | Class A    | 94.41       | 95.98     | 95.19     | 95.98     | 4.02     | 95.99      |
|                 | Class B    | 84.07       | 93.28     | 88.44     | 91.59     | 8.41     | 90.70      |
|                 | Class C    | 85.47       | 89.54     | 87.46     | 96.29     | 3.71     | 97.43      |
|                 | Class D    | 95.27       | 44.59     | 60.75     | 94.42     | 5.58     | 99.76      |

As can be seen from the performance results, our experiment confirmed that the Yolov3-Augment model, learned by object specificities such as motion and perspective, effectively limits object recognition errors more than the other models.

3.3. Improvement of the Object Recognition

In football videos, it is hard to recognize target objects when players with similar features repeatedly overlap and separate during the game. In addition, there are limitations that objects can detect only within the viewing angle of a single camera. Furthermore, the characteristic information of players from a long-distance decreased compared to players nearby. As a result, misrecognition and non-recognition errors occurred in the Yolov3-416 model when detecting a person, as shown in Figure 18a.

In this study, we compared the general AI model Yolov3-416 with other AI models, in which objects were subdivided as football players, confirming the significant improvement of the object recognition errors. As shown in Figure 18b, recognition errors are no different in the structural modification model, Yolov3-HSV, compared to Yolov3-416. On the other hand, the Yolov3-Augment model had many improvements in terms of misrecognition and non-recognition errors, as shown in Figure 18c. It recognized a new object (class D) in an area where target objects overlapped through model learning with a reinforced training dataset and an augmented object class. Furthermore, the Yolov3-Augment model detected small distant objects better than the Yolov3-416 model.

We reduced the number of recognition errors caused by a lack of feature information by subdividing the object features used for the model learning into color and motion. This improved the result of recognition errors due to separately subdividing the uncertain elements of object recognition, implicit in a single object and classifying overlapping objects, into similar object groups. As a result, we could accurately recognize objects by improving the prediction performance in the Yolov3-Augment model.
3.3. Improvement of the Object Recognition

In football videos, it is hard to recognize target objects when players with similar features repeatedly overlap and separate during the game. In addition, there are limitations that objects can detect only within the viewing angle of a single camera. Furthermore, the characteristic information of players from a long distance decreased compared to players nearby. As a result, misrecognition and non-recognition errors occurred in the Yolov3-416 model when detecting a person, as shown in Figure 18a.

(a)

(b)

(c)

Figure 18. Comparison of object recognition in three AI models. Class A: red uniform, Class B: blue uniform, Class C: white uniform, Class D: overlapped objects; (a) Objects recognition in Yolov3-416 model; (b) Objects recognition in Yolov3-HSV model; (c) Objects recognition in Yolov3-Augment model.

4. Conclusions

In this study, we compared the general AI model Yolov3-416 with other AI models, in which objects were subdivided as football players, confirming the significant improvement of the object recognition errors. As shown in Figure 18b, recognition errors are no different in the structural modification model, Yolov3-HSV, compared to Yolov3-416. On
As a result, errors such as non-recognition and misrecognition of the general AI model were recognized as detection targets. The reason for this is that we limited specific objects to the classification, detection, and recognition areas. In addition, it became the strategical basis for diverse approaches to changes in time and space according to the target object movement in the overlapped objects, classified as class D. This was also the result of subdividing the research area so that similar objects can be re-recognized. Therefore, we confirmed that the AI recognition model with structural modifications and object class augmentation effectively reduces object recognition errors. In future work, we will propose a method and algorithm for tracking objects individually in areas where objects overlap, and further improve the effectiveness of this study.

Recognizing a player as an individual, and recognizing players as a team in a football game, is an important monitoring task for analyzing players’ performance. After recognizing players without error, the proposed AI model can be extended to include tracking players’ movement changes, analyzing activity, and automatic statistical analysis.

In the future, in football and other field sports, training data augmenting methods designed to reduce recognition errors by improving the uniqueness of similar objects and proposed artificial intelligence models could be used for analyzing player activity and assisting referees’ judgment. We also expect to apply the effectiveness of this study by extending its scope to detect a variety of target objects and minimize loss in real-time (e.g., monitoring and data acquisition on traffic, animal activities, environment monitoring and etc.).

Author Contributions: Data curation, J.H. and J.K.; Formal analysis, J.H., K.K. and J.K.; Methodology, J.H. and J.K.; Software, J.H.; Supervision, K.K. and J.K.; Validation, J.H., K.K. and J.K.; Visualization, J.H.; Writing—original draft, J.H.; Writing—review and editing, J.H., K.K. and J.K.; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Martin, D.R.; Fowlkes, C.C.; Malik, J. Learning to detect natural image boundaries using local brightness, color, and texture cues. *IEEE Trans. Pattern Anal. Mach. Intell.* **2004**, *26*, 530–549. [CrossRef] [PubMed]
2. Canny, J. A Computational Approach to Edge Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **1986**, *8*, 679–698. [CrossRef] [PubMed]
3. Matas, J.; Chum, O.; Urban, M.; Pajdla, T. Robust wide-baseline stereo from maximally stable extremal regions. *Image Vis. Comput.* **2004**, *22*, 761–767. [CrossRef]
4. Rosten, E.; Porter, R.; Drummond, T. Faster and Better: A Machine Learning Approach to Corner Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *32*, 105–119. [CrossRef] [PubMed]
5. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, 23–28 June 2014; pp. 580–587. [CrossRef]
6. Girshick, R. Fast R-CNN. In *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*, Santiago, Chile, 7–13 December 2015; pp. 1440–1448. [CrossRef]
7. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [CrossRef] [PubMed]
8. Lin, T.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, 22–29 October 2017; pp. 2999–3007.
9. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. SSD: Single shot multibox detector. In *Proceedings of the European Conference on Computer Vision*, Amsterdam, The Netherlands, 11–14 October 2016. [CrossRef]
10. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788. [CrossRef]
11. Liang, Q.; Wu, W.; Yang, Y.; Zhang, R.; Peng, Y.; Xu, M. Multi-Player Tracking for Multi-View Sports Videos with Improved K-Shortest Path Algorithm. *Appl. Sci.* **2020**, *10*, 864. [CrossRef]
12. Ophoff, T.; Van Beeck, K.; Goedemé, T. Exploring RGB+Depth Fusion for Real-Time Object Detection. *Sensors* **2019**, *19*, 866. [CrossRef] [PubMed]
13. Kim, H.; Kim, J.; Kim, Y.-S.; Kim, M.; Lee, Y. Energy-Efficient Wearable EPTS Device Using On-Device DCNN Processing for Football Activity Classification. *Sensors* **2020**, *20*, 6004. [CrossRef] [PubMed]
14. Álvarez-Aparicio, C.; Guerrero-Higuera, M.; Rodríguez-Lera, F.J.; Clavero, J.G.; Rico, F.M.; Matellán, V. People Detection and Tracking Using LIDAR Sensors. *Robotics* **2019**, *8*, 75. [CrossRef]
15. Taylor, L.; Nitschke, G. Improving Deep Learning with Generic Data Augmentation. In *Proceedings of the 2018 IEEE Symposium Series on Computational Intelligence (SSCI)*, Bangalore, India, 18–21 November 2018. [CrossRef]
16. Accuracy and Precision. Available online: https://en.wikipedia.org/wiki/Accuracy_and_precision (accessed on 22 September 2022).
17. AlexeyAB. GUI for Marking Bounded Boxes of Objects in Images. Available online: https://github.com/AlexeyAB/Yolo_mark/ (accessed on 27 December 2021).
18. Lin, T.Y.; Maire, M.; Belongie, S.; Bourdev, L.; Girshick, R.; Hays, J.; Perona, P.; Zitnick, C.L.; Dollár, P. Microsoft COCO: Common Objects in Context. In *European Conference on Computer Vision*; Springer: Cham, Switzerland, 2014; pp. 740-755. [CrossRef]
19. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. *arXiv* **2018**, arXiv:1804.02767.
20. Xiang, L.; Deng, K.; Wang, G.; Zhang, Y.; Dang, Q.; Gao, Y.; Shen, H.; Ren, J.; Han, S.; Ding, E.; et al. PP-YOLO: An effective and efficient implementation of object detector. *arXiv* **2007**, arXiv:2007.12099v3.
21. Wikipedia. HSL and HSV. Available online: https://en.wikipedia.org/wiki/HSL_and_HSV#From_RGB (accessed on 16 February 2022).
22. Sokolova, M.; Lapalme, G. A systematic analysis of performance measures for classification tasks. *Inf. Process. Manag.* **2009**, *45*, 427–437. [CrossRef]
23. Sokolova, M.; Lapalme, G. A systematic analysis of performance measures for classification tasks. *Inf. Process. Manag.* **2009**, *45*, 427–437. [CrossRef]
24. Stehman, S.V. Selecting and interpreting measures of thematic classification accuracy. *Remote Sens. Environ.* **1997**, *62*, 77–89. [CrossRef]