

A Method for Reducing False Negative Rate in Non-Maximum Suppression of YOLO Using Bounding Box Density

Dong-Hyeon Jeon¹, Tae-Sung Kim¹, Jin-Sung Kim^{1*}

Abstract

In the previous non-maximum suppression (NMS) in you only look once (YOLO) v5, false negative error happens even when there are many bounding boxes for an object because all bounding boxes have lower confidence score. This work finds that a lot of bounding boxes near an object of false negative error are removed because of low confidence score. This paper proposes a new modified confidence score that is increased when bounding boxes with the same class prediction are located densely. The proposed method reduces the false negative error caused by low confidence score effectively. Experimental results show that the proposed method detects 25.33% more objects than conventional NMS at mAP@0.5.

Key Words: : Non-Maximum Suppression, Object Detection, False Negative Error, YOLO.

I. INTRODUCTION

The performance of object detection is improved considerably by deep learning, and object detection is widely used in many applications such as auto driving system, aerospace, surveillance and reconnaissance. Object detection can be categorized into a two-stage detector and a one-stage detector. In two-stage detectors, regions with convolutional neural networks (R-CNN) finds candidate regions for objects, and the image regions are put in a neural network for classification [1]. Fast R-CNN and Faster R-CNN improve the speed of R-CNN [2-3]. These two-stage detectors detect objects with high accuracy through two steps which are region proposal and classification. However, high computational complexity degrades the processing speed. One-stage detectors process an image with a single step for determining the location and class. One-stage detectors have lower computational complexity and it is faster than two-stage detectors. Therefore, one-stage detectors can be applied to real-time applications.

In one-stage detectors, OverFeat detects an object in multiple scales of an image [4]. A single shot multibox detector (SSD) detects an object in an image at one time. It uses multiple scales feature map to detect an object [5]. YOLO (You Only Look Once) uses a single neural network to predict both the bounding box and the class simultane-

ously [6]. YOLO has fast speed and high accuracy, which makes it widely used.

In YOLO, multiple bounding boxes are generated for an object. In order to assign a single bounding box to an object, redundant bounding boxes should be removed. Non-maximum suppression (NMS) removes false and redundant predictions, aiming to obtain a single accurate prediction for an object. In NMS, if the confidence score of a predicted bounding box is lower than the confidence threshold, the prediction is considered as a false one and it is ignored. A bounding box that has high overlapping ratio (IoU) with other bounding box is removed because it is considered as a redundant one. Based on a confidence score and overlapping ratio, false and redundant bounding boxes are removed, and each object has a single bounding box as a final detection result. However, in some cases, all bounding boxes for an object have lower confidence score than the threshold, which results in a false negative error even when there are many bounding boxes near the object. In order to reduce the false negative error in NMS, the confidence threshold can be lowered, but the lower threshold increases the false positive error. Thus, simple thresholding method with a single threshold cannot solve the false negative problem in NMS.

This work proposes a method for improving a false negative error which is caused by removing all bounding boxes for an object in NMS because of low confidence score.

Manuscript received November 12, 2023; Revised December 1, 2023; Accepted December 13, 2023. (ID No. JMIS-23M-11-046)

Corresponding Author (*): Jin-Sung Kim, +82-41-530-2232, jinsungk@sunmoon.ac.kr

¹Department of Electronic Engineering, Sun Moon University, Asan, Korea, amuse_dh@naver.com, ts7kim@sunmoon.ac.kr, jinsungk@sunmoon.ac.kr

Based on the empirical observation, there are a lot of true bounding boxes with low confidence score near an object in many cases of false negative errors. It means that the location on which many bounding boxes with the same class label are overlapped densely can be the true position of an object even when confidence scores of the bounding boxes are lower than the threshold.

In this paper, bounding boxes that can be true prediction are selected among the bounding boxes removed by the conventional NMS. Modified confidence scores of the selected boxes are calculated based on the given score and the distribution of bounding boxes.

These selected boxes with the modified score are used for the final detection results, which reduces the false negative errors. Experimental results show that the proposed method reduces false negative errors to 40% compared to the errors in conventional NMS.

The rest of this paper is organized as follows. Section 2 introduces the related work. The proposed method is presented in Section 3. Experimental results are presented in Section 4, and conclusions are drawn in Section 5.

II. RELATED WORK

2.1. You Only Look Once (YOLO)

YOLO is a one-stage detector that performs localization and classification simultaneously, and it processes a whole image at one time. Thus, it gives fast processing time [6].

YOLOv1 detects objects in a whole image at once, which shows the possibility of real-time object detection. YOLOv1 has performance limitations compared to the conventional two-stage detectors. It shows difficulty in detecting small objects accurately.

YOLOv2 improves the network architecture to overcome the limitations of YOLOv1 and it applies anchor box [7]. YOLOv2 shows more effective detection ability with faster processing speeds. However, it still has limitations in performance when compared to two-stage detectors.

YOLOv3 enhances object detection accuracy and processing speeds through the improvement in the network architecture and training methods [8]. Especially, a subdivided anchor box and multistage output are introduced for enhancing the detection ability for objects with variable sizes and shapes.

YOLOv4 utilizes spatial pyramid pooling (SPP) and path aggregation network (PAN) techniques to achieve higher accuracy and faster processing speeds, which shows outstanding performance in object detection [9]. YOLOv5 shows more than a 10% increase in object detection accuracy compared to YOLOv4, and both computation time and memory usage have been reduced compared to the previous version, resulting in better performance .

Each version of YOLO improves the performance of the previous version. YOLO model is attracting a lot of attention in various applications because of the fast processing speed and high accuracy.

2.2. Non-Maximum Suppression (NMS)

NMS is a technique for removing errors or redundant results in computer vision tasks such as object detection. In object detection models, a bounding box is considered as a redundant one when its IoU with other bounding box is larger than a predefined threshold. NMS leaves only a single bounding box among the redundant bounding boxes for a single object. Recently, the accuracy of a bounding box is improved in NMS by utilizing redundant bounding box information [10].

In YOLOv5, the combination of NMS and confidence thresholding improves the performance of object detection. At first, confidence thresholding removes bounding boxes with lower confidence score than the confidence threshold, which removes false positive errors. Then, NMS selects a bounding box with the highest confidence score, and bounding boxes which have higher IoU with the selected box than a IoU threshold are removed because they are considered as redundant ones. The selected box is determined as a final detection result.

The confidence thresholding reduces the false positive error and NMS suppresses redundant detection results, which improves the accuracy of a detection model. However, some objects cannot be detected even when there are many bounding boxes near an object.

When all of bounding boxes for an object have lower confidence score than a confidence threshold, the conventional NMS removes entire bounding boxes and false negative error happens. If the confidence threshold is lowered in this case, some of the bounding boxes would not be removed and false negative error can be avoid. However, the low confidence threshold can lead a false positive error because bounding boxes with wrong prediction would not be removed. Performing confidence thresholding with a single confidence threshold value cannot simultaneously address false negative and false positive errors.

Fig. 1 shows an example of a false positive error. Fig. 1 (a) shows bounding boxes of which the index numbers are given. Table 1 shows the true and predicted classes and confidence score of the bounding boxes in Fig. 1 (a). In this figure, boxes of index 0, 2 and 3 have true class prediction while the scores are lower than the confidence threshold of 0.7. Fig. 1 (b) shows detection result with the confidence threshold of 0.7. In this figure, there is no detection, which results in the false negative error. This example shows that useful bounding boxes are removed by the confidence thresholding.

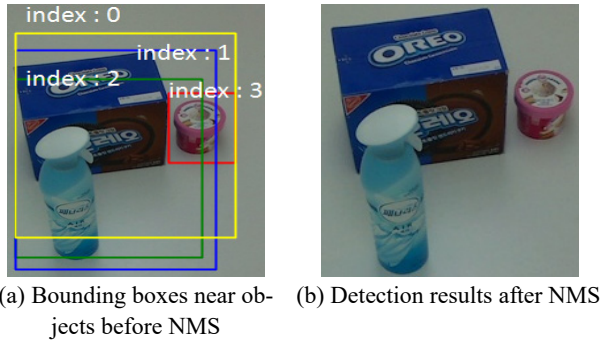


Fig. 1. Example of false negative error in NMS.

Table 1. The predicted class and confidence score of bounding boxes in Fig. 1.

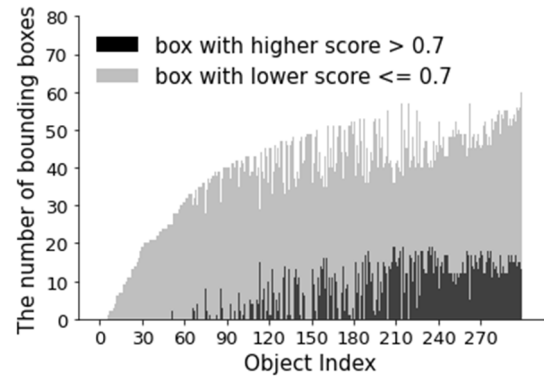
Index	True class	Predicted class	Confidence score
0	Oreo	Oreo	0.62
1	Febreeze	Soda	0.064
2	Febreeze	Febreeze	0.68
3	Ice cream	Ice cream	0.009

In order to reduce the false negative error, the confidence threshold can be lowered. If the threshold is lowered to 0.001, all of three objects are detected. However, the bounding box of index 1 will become a final prediction that is a false positive error. Thus, the problem cannot be solve by adjusting a single threshold value.

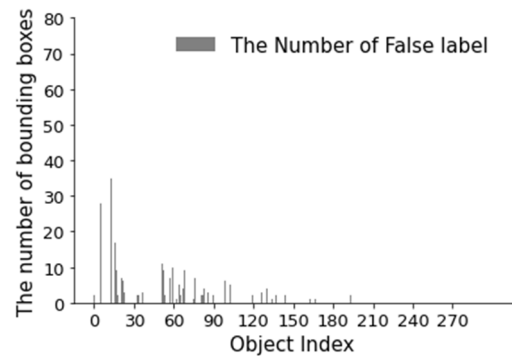
III. PROPOSED WORK

In this work, bounding boxes near an object before NMS step are investigated. Fig. 2 shows the number of bounding boxes near an object before NMS step. Fig. 2 (a) shows the number of bounding boxes with true class prediction, and Fig. 2 (b) shows that with false class prediction. The vertical axis and horizontal axis represent the number of bounding boxes and object index, respectively.

In this experiments, there are 300 objects in 113 images. In Fig. 2 (a), gray bar shows the number of bounding boxes with lower score that the confidence score threshold, and thus they will be removed in NMS. The black bar shows the number of bounding boxes with higher score, and thus they will be used for the final detection results. The average number of bounding boxes with true class prediction near an object is 44, and 38 boxes among them are removed on average because of the low confidence score. It means that there are a lot of bounding boxes with a correct class prediction near an object, and many of them are removed because of low confidence score. Fig. 2 (b) shows the number of bounding boxes near an object with a false class prediction. This figure shows that the number of bounding boxes with false class is small. The average number of bounding



(a) The number of bounding boxes with a true class



(b) The number of bounding boxes with a false class

Fig. 2. The number of bounding boxes near an object.

boxes with false class near an object is 1.03. Fig. 2 indicates that there are a lot of bounding boxes with a true class prediction near an object and the number of bounding boxes with a false class prediction near an object is very small. Therefore, the chance is high that a bounding box that is overlapped with a large number of bounding boxes with the same class is a true prediction even when the confidence score is lower than the threshold.

In this paper, the bounding boxes with higher score than the conventional threshold are fed into the conventional NMS, and additional bounding boxes which can be a true prediction are selected by the proposed scheme. Those additional one are fed into the NMS as well. In this paper, the conventional threshold and the proposed threshold are set to 0.7 and 0.00001, respectively. Fig. 3 shows the maximum confidence score of bounding boxes with true class prediction for an object. The solid line and dotted line depict the threshold value of the conventional method and the proposed method, respectively. When the maximum score is lower than the solid line, all bounding boxes are removed in NMS, resulting in false negative errors. In the proposed method, the threshold is set to 0.00001 to avoid false negatives. However, the lower threshold leads to false positive errors.

Fig. 4 shows an example of detection results for an image in which there are two objects. Fig. 4 (a) shows the detec-

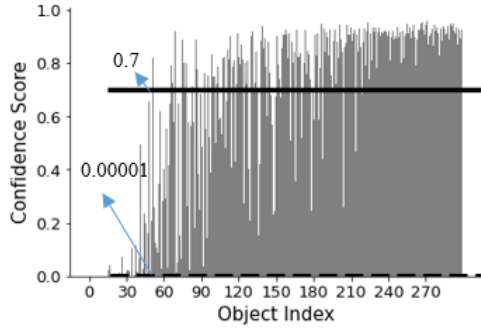


Fig. 3. Maximum confidence scores of bounding boxes for each object. The solid line and the dotted line represent threshold values of the conventional NMS (0.7) and the proposed method (0.00001), respectively.

tion results with a confidence threshold of 0.7. In this figure, the lower object is detected correctly with the score of 0.75, but the upper object is not detected. In Fig. 4 (b), the proposed threshold of 0.00001 is applied. In this figure, there are many bounding boxes. The bounding boxes for the upper object are not removed, which implies that the false negative can be improved. However, it also results in many bounding boxes in the background area and near objects, potentially causing false positive errors.

In order to reduce the false negative error in Fig. 4 (a), more bounding boxes with low score needs to be considered. In this paper, the bounding boxes with higher score than the proposed threshold of 0.00001 are considered. In this case, the bounding boxes for the upper object enables the object to be predicted as shown in Fig. 4 (b). However, there are too many bounding boxes which are redundant or false prediction. The proposed method finds the true one among the bounding boxes of which the score ranges from 0.00001 to 0.7. It can be achieved by considering the spatial density of bounding boxes with the same class prediction. Fig. 2 indicates that many bounding boxes with true class are overlapped near an object. In the proposed method, a modified score is defined for each bounding box considering the number of overlapped bounding boxes with the same class prediction. Additionally, bounding boxes in the background region are ignored.

3.1. The Proposed Method

The proposed method improves the prediction accuracy

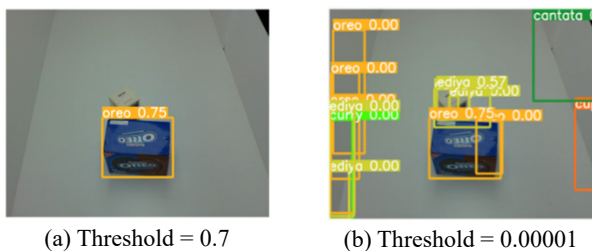


Fig. 4. Remaining bounding boxes after confidence thresholding.

by modifying the NMS step in YOLOv5. Fig. 5 illustrates the overall flow of the proposed method. The thresholds of the conventional NMS and the proposed method are denoted by TH_H and TH_L , respectively. $BBox_H$ represents a set of bounding boxes with higher score than TH_H , and $BBox_L$ represents a set of bounding boxes with score from TH_L to TH_H . In Step 1, the algorithm determines the foreground and removes unnecessary predictions in the background. In Step 2, the remaining predictions are categorized into the $BBox_H$ group and the $BBox_L$ group. In Step 3, the score of a bounding box in $BBox_L$ is updated considering the spatial density of predictions. Some bounding boxes moves to $BBox_H$. Step 4 determines the final prediction. Fig. 6 shows the proposed algorithm in detail.

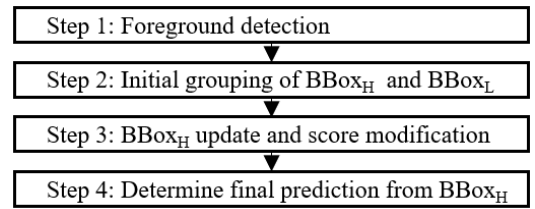


Fig. 5 The overall flow of the proposed algorithm.

Algorithm 1.

TH_H : Conventional NMS Threshold
 TH_L : Proposed Threshold
 $BBox_H$: A set of bounding boxes over TH_H
 $BBox_L$: A set of bounding box between TH_H and TH_L

Step 1: Detect foreground, and remove bounding boxes in background

Step 2: Categorize bounding boxes into $BBox_H$ and $BBox_L$ by using eq (1)

Step 3:
for index in $BBox_L$
 for index in each Class
 - Select the bounding box with maximum score
 - Count overlapped boxes with the selected bounding box ($N_{overlap}$) using IoU calculation.
 - Update the score of the selected box by eq (2)
 - Move the selected box to $BBox_H$.
 - Remove the selected and overlapped bounding boxes from $BBox_L$
 - Repeat Step 3 until all boxes are removed

Step 4:
for index in $BBox_H$
 for index in Class
 - Select a bounding box with maximum score
 - If the confidence score of the selected one is higher than 0.7, it is determined as the detection result.
 - Find the overlapped boxes with the selected bounding box using IoU calculation.
 - Remove the selected and overlapped boxes from $BBox_H$
 - Repeat Step 4 until all boxes are processed.

Fig. 6. The proposed algorithm.

In the first step, the foreground area is determined by difference between background and input images. IoU between the foreground and each bounding box is calculated. When IoU is lower than a predefined threshold, the bounding box is removed because it is not a bounding box for the foreground object.

In the remaining bounding boxes, bounding boxes with confidence scores higher than TH_H are grouped into BBox_H , and those with scores from TH_L to TH_H are grouped into BBox_L , which can be expressed by (1). B_i denotes the i -th bounding box.

$$\begin{aligned} B_i &\in BBox_H, & \text{if } Score(B_i) > TH_H. \\ B_i &\in BBox_L, & \text{else if } Score(B_i) > TH_L. \end{aligned} \quad (1)$$

The bounding boxes in BBox_H are considered as candidates for final prediction in the conventional NMS. On the other hand, those in BBox_L may be true or false predictions because the score is not high enough.

Fig. 7 (a) shows an example of bounding boxes with higher score than TH_L , and Fig. 7 (b) shows bounding boxes in BBox_L in which bounding boxes in background are removed.

In the third step, bounding boxes that can be true prediction are found in BBox_L considering the number of overlapped bounding boxes based on the observation in Fig. 2. When a lot of bounding boxes with the same class prediction are overlapped, there is a high possibility of true prediction. This work proposes a new modified score that considers the spatial density of the bounding boxes.

At first, the proposed method selects a bounding box with the highest score for a class in BBox_L , and it counts the number of overlaps with the selected bounding box. This number of overlaps is denoted by N_{overlap} . Then, the confidence score of the selected bounding box is updated by (2).

$$score_{mod} = 0.06 \times N_{overlap}. \quad (2)$$

The modified score, $\text{score}_{\text{mod}}$, increases as the number of overlapped bounding boxes with the same class prediction increases. The selected bounding box with $\text{score}_{\text{mod}}$ moves

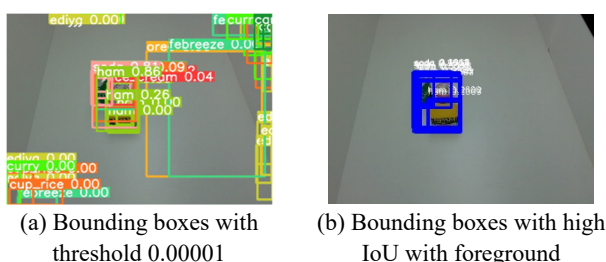


Fig. 7. Bounding boxes of foreground in Step 1 in Fig. 5.

to BBox_H . Then, the bounding boxes overlapped with the selected bounding box are removed from BBox_L . This process is repeated for the remaining bounding boxes within BBox_L . When all bounding boxes in BBox_L are removed, the third step is completed.

The fourth step is almost the same as the conventional NMS in YOLOv5 except for additional bounding boxes with the $\text{score}_{\text{mod}}$. The bounding box with the highest score for a class in BBox_H is selected. If the confidence score of the selected one is higher than 0.7, it is determined as a final detection result. The bounding boxes overlapped with the selected bounding box are removed. This process is repeated until all bounding boxes in BBox_H are processed.

Fig. 8 shows an example of the proposed method. Fig. 8 (a) shows an input image in which there are three objects, ‘Cup rice’, ‘Febreeze’ and ‘Ice cream’. Fig. 8 (b) represents all bounding boxes of which the confidence score is higher than 0.00001. Fig. 8 (c) depicts the result of the conventional NMS. Only a single object of ‘Febreeze’ is detected and two objects are not detected. The bounding boxes for these two object in Fig. 8 (b) are removed because all of their bounding boxes have low confidence score. Fig. 8 (d) shows the foreground region that is determined in the first step in Fig. 6. The bounding boxes with a low IoU with the foreground region are discarded as shown in Fig. 8 (e). The

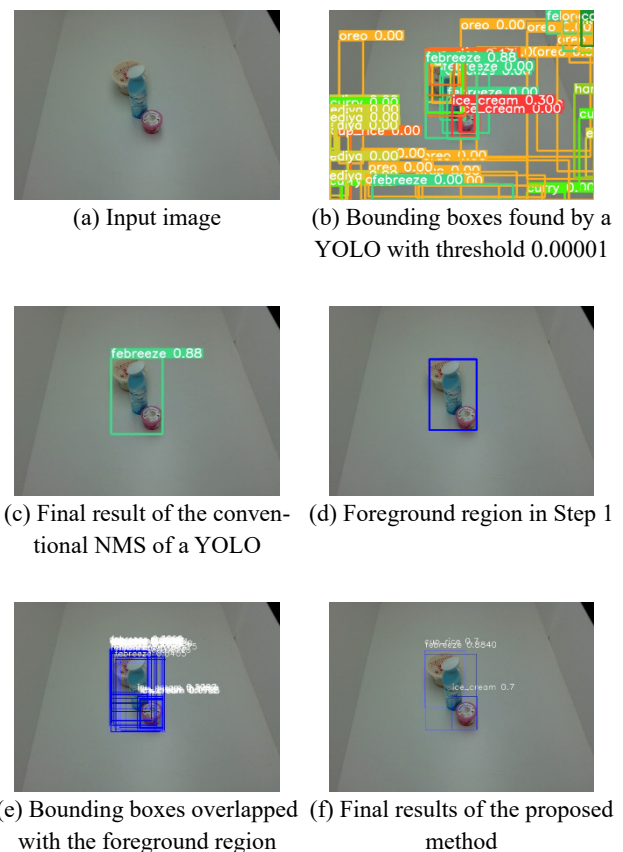


Fig. 8. Comparison of a conventional NMS and the proposed method.

final detection results of the proposed method are shown in Fig. 8 (f). All of three objects are detected correctly with an increased score by (2).

IV. EXPERIMENTAL RESULT

4.1. Datasets

The training dataset has 9 object classes, and 92 images are captured for each class, so 828 images are captured. Additional 2,484 images are generated by applying vertical flips, 90-degree and 270-degree rotations. In total, the training dataset consists of 3,312 images with a resolution of 640×480. The validation dataset has 190 images and the test dataset has 113 images captured from a single view with a resolution of 640×480.

4.2. Environment of Experiment

The experiment is conducted using an NVIDIA GeForce GTX GPU and an Intel(R) Xeon(R) CPU E3-1245 v6 @ 3.70GHz. The entire model is implemented using PyTorch, and it utilizes CUDA 11.6 and cuDNN 8 for computation. YOLOv5 is employed to extract predictions.

4.3. Experimental Results

The performance of the proposed method is evaluated. Fig. 9 shows the mAP for each IoU threshold. In Fig. 9 and Table 2, the proposed algorithm exhibits higher mAP than those of the conventional NMS when IoU is lower than 0.9. Table 3 shows the numbers of TPs, FPs and FNs of the conventional NMS and proposed method when IoU threshold is 0.5. The conventional NMS detects 62.6% of the objects in the ground truth, while the proposed algorithm detects 88.0% of the objects. The number of false negative errors is reduced from 112 to 36. In Table 3, false positive error increases from 8 to 21. Table 4 compares the precision and recall of the proposed method and the conventional NMS. The precision decreases by 3.44% because of the increase of FPs while the recall is improved by 41.94%. It shows that the improvement in recall is significantly higher than the

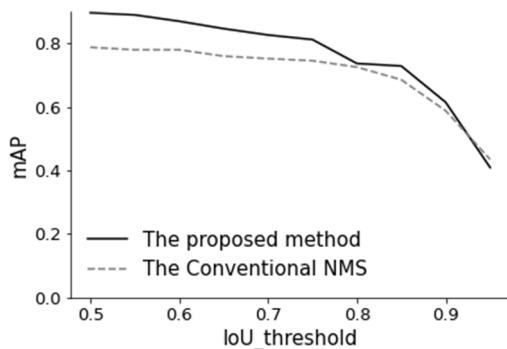


Fig. 9. mAP for each IoU Threshold.

Table 2. mAP@0.5 and mAP@[.50:.95] for the proposed method and the conventional NMS.

Method	mAP@0.5	mAP@[.50:.95]
The conventional NMS	0.787	0.703
The proposed method	0.896	0.763

Table 3. The number of TPs, FPs and FNs in the proposed method and the conventional NMS with 0.5 of the IoU threshold.

Method	True positive	False positive	False negative
The conventional NMS	188	8	112
The proposed method	264	21	36

Table 4. Comparison of precision and recall for the proposed method and the conventional NMS.

	The conventional method	The proposed method	Improvement (%)
Precision	0.959	0.926	-3.44
Recall	0.626	0.880	41.9

degradation in precision.

V. CONCLUSION

This study proposes a method for reducing false negative error caused by the conventional NMS. Among the bounding boxes with lower confidence score than the threshold, the bounding boxes that can be true positive are selected by analyzing the spatial density of the bounding boxes. The proposed method increases the confidence score of the selected bounding boxes in proportion to the number of the overlapped bounding boxes. The experimental results show that false negative error is improved significantly. This improvement can be achieved in the post-processing of which the additional computation cost is minimal. In the proposed method, false positive increases but the degradation in precision is not large. All of the increased false positive errors are caused by a location error. There are previous works for improving the localization accuracy of a bounding box [11,12]. Based on these studies, we will improve the false positive error in the proposed method in future.

ACKNOWLEDGEMENT

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean Government (MSIT) under Grant 2021R1G1A1094961.

REFERENCES

- [1] R. Girshick, J. Donahue, t. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 580-587.
- [2] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1440-1448.
- [3] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *Advances in Neural Information Processing Systems*, vol. 28, 2015.
- [4] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. Lecun, "Overfeat: Integrated recognition, localization and detection using convolutional networks," *arXiv preprint arXiv:1312.6229*, 2013.
- [5] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu, and A. C. Berg, "SSD: Single shot multi-box detector," *Computer Vision—ECCV 2016: 14th European Conference*, pp. 21-37, 2016.
- [6] 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*, pp. 779-788, 2016.
- [7] 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.
- [8] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018.
- [9] A. Bochkovskiy, C. Y. Wang, and H. Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," *arXiv preprint arXiv:2004.10934*, 2020.
- [10] J. Chu, Y. Zhang, S. Li, L. Leng, and J. Miao, "Syncretic-NMS: A merging non-maximum suppression algorithm for instance segmentation," *IEEE Access*, vol. 8, pp. 114705-114714, 2020.
- [11] Y. Zhang, J. Chu, L. Leng, and J. Miao, "Mask-refined R-CNN: A network for refining object details in instance segmentation," *Sensors*, vol. 20, no. 4, p. 1010, 2020.
- [12] W. Lin, J. Chu, L. Leng, J. Miao, and L. Wang, "Feature disentanglement in one-stage object detection," *Pattern Recognition*, vol. 145, p. 109878, 2024.

AUTHORS



Dong-Hyeon Jeon is currently working toward the BS degree in Department of Electronic Engineering at Sun Moon University, Asan, Korea since 2019. His research interests include deep learning algorithms and computer vision algorithms.



Tae-Sung Kim received the B.S degree in electrical electronic engineering from Pusan National University, Pusan, South Korea, in 2010 and the M.S. and Ph.D. degrees in electrical and computer engineering from Seoul National University, Seoul, South Korea, in 2013 and 2017, respectively. He was a senior researcher at Samsung S.LSI from 2018 to 2021. Starting from 2021, he is currently an assistant professor in the Electronic Engineering Department of Sun Moon University. His research interests include the algorithm and architecture design of image/video processing.



Jin-Sung Kim received his BS and MS degrees and his PhD degrees in electrical and computer engineering from Seoul National University, Seoul, Rep. of Korea, in 1996, 1998, and 2009, respectively. From 1998 to 2004, and from 2009 to 2010, he was with Samsung SDI Ltd., Cheonan, Rep. of Korea, as a senior researcher. From 2010 to 2011, he was a post-doctoral researcher at Seoul National University. In 2011, he joined the Department of Electronic Engineering, Sun Moon University, Asan, Rep. of Korea, where he is currently a Professor. His current research interests include deep learning, algorithms and architectures for video compression and computer vision.

