YOLOv7 Tiny improvement for bull sperm detection

Wafi Khoerun Nashirin*, Azzam Badruz Zaman, Priyanto Hidayatullah, Ardhian Ekawijana
Computer Engineering and Informatics Department, Politeknik Negeri Bandung, Indonesia

Abstract

YOLO (You Only Look Once) is a prominent deep learning model used in object detection due to its high detection accuracy and speed. Nonetheless, in detecting bull sperm, YOLOv7 Tiny performance suffers because of the unique characteristics of bull sperm: its tiny size and the large quantity of sperm. YOLOv7 Tiny's performance can be improved by adjusting based on its unique characteristics. This study proposes a modified YOLOv7 Tiny model to detect bull sperm with higher accuracy. The main objective of this research is to increase the accuracy of YOLOv7 Tiny in detecting and counting bull sperm. The YOLOv7 Tiny architecture will be modified based on the characteristics of the object to be detected, specifically bull sperm. Several architectural parts were deleted, the anchor box's size was changed, and the grid cell's size was changed. The omitted architecture parts are the ones used for detecting large and medium-sized objects. The anchor box and grid cell sizes will be altered to fit the size of the object. Accuracy is measured using mean average precision (mAP). The modified YOLOv7 Tiny will be evaluated in comparison to the original YOLOv7 Tiny. In our experiment, we produced 65.8 mAP with an inference time of 14.4 ms on the test dataset. When detecting bull sperm, the modified model is 1.3 points more accurate and 1.23x faster than YOLOv7 Tiny. The size of the modified model file is likewise decreased by 84.2%.

INTRODUCTION

Beef is one of the most widely consumed types of animal protein in the world. To meet the demand for beef consumption, numerous attempts have been made to increase the number of bulls. To meet the demand for beef, one attempt is to increase the number of bulls through artificial insemination. Artificial insemination in bulls is a method or technique that involves inserting frozen male bull sperm that has been thawed and pre-processed into female animals via a unique method and tool known as an insemination gun [1]. However, artificial insemination failures occur frequently, necessitating repeated insemination for two to three repeats [2].

One of the most important aspects influencing the success of artificial insemination is the quality of the sperm. Bull sperm can be examined both macroscopically and microscopically. Microscopic testing includes motility, numbers, viability (live sperm percentage), and morphological tests (abnormalities in sperm). The test is performed to assess whether the bull sperm fulfills the requirements for insemination. The criteria are the following: at least 40% sperm motility, individual sperm movement has a value of two, and a sperm count of 25 million per dose [3].

Computer Assisted Sperm Analysis (CASA) is a technology that can be utilized in autonomous sperm testing and analysis. CASA is a prominent sperm analysis tool among researchers [4]. CASA, on the other hand, is expensive, and there are various conditions to

be met for good accuracy. As a result, The Lembang Artificial Insemination Agency is increasingly using the expert judgment technique to assess sperm quality, which requires a subjective review by an expert in the field of animal reproduction. In fact, the expert judgment approach still has drawbacks, such as the subjectivity of assessment, which can vary between experts, and the inability to provide more specific quantitative information regarding sperm quality.

The author in [5] carried out a study to aid in the process of assessing sperm quality in order to avoid repeats during the insemination process. A video processing method was employed in the study to detect bull sperm from a dataset taken with a smartphone at a magnification of 200x. The author in [4] suggested a more advanced approach for testing the quality of bull sperm, namely object detection. The DeepSperm model, a modified version of YOLOv3, is introduced in [4]. DeepSperm can detect bull sperm with high accuracy on a 500x magnification dataset. When detecting bull sperm in the dataset used in [5], DeepSperm accuracy falls [4]. As a result, YOLO was modified in this work with a newer YOLO version than DeepSperm (YOLOv7 Tiny) to detect bull sperm in a 200x magnification dataset. Modifications were made to improve accuracy, as described in [4]. In addition to [4], improvements to the YOLO model architecture have been implemented in many other small-size object samples and have been shown to improve the accuracy of each case study [6, 7, 8, 9].

Based on the study of the difficulties outlined above, a more precise approach for detecting bull sperm in the post-thawing phase (sperm containing skim-based diluent) is required. Consequently, it is proposed that YOLOv7 Tiny, which is designed for detecting bull sperm, be utilized. Because of its good detection accuracy and speed, YOLO is a popular deep learning model used in object detection. The accuracy of YOLOv7 Tiny needs to be enhanced by making changes based on the dataset properties. This is possible because the object discovered in this study (bull sperm) has unique properties, such as being small and having a large number in a video frame. The modifications made are adjusting the number of parameters by removing all architectural parts that are related to detect large and medium-sized objects. Furthermore, the size of the anchor box and grid cells were adjusted to match the characteristics of the object and achieve higher accuracy in detecting bull sperm. The modified YOLOv7 Tiny accuracy will be compared to the original YOLOv7 Tiny accuracy to ascertain the effect of the alterations.

Several related publications discuss the topic of object detection. They were discovered using a variety of techniques. The YOLO model and image processing have been used. The YOLO model's accuracy has increased because of several experiments.

The authors in [5] and [10] carried out the process of detecting using image processing. The detected objects are bull sperm [5] and human sperm [10]. The two studies have been able to detect objects using image processing. However, detecting objects using image processing has limitations, such as the fact that there cannot be other objects that have a similar shape to the object to be detected in the sample.

YOLO has been used to detect objects in [11][12], and [13]. Human sperm was discovered by researchers [11][12], and [13]. The YOLO versions used are YOLOv3 [11], YOLOv4 [12], and YOLOv5 [13]. These three experiments demonstrate that the YOLO model can detect human sperm very well.

However, in object detection, tiny object detection is still a difficult task. Usually smaller than 32 by 32 pixels, tiny objects. Thus, in the field of object detection, small item detection has always been a challenging problem to solve [14][15], Tiny items frequently have insufficient visual information to differentiate them from complicated backgrounds when

compared to large or medium-sized objects. The absence of positive samples is typically experienced by little objects, which typically have low IoU with anchors or cover a small number of feature points [16]. In conclusion, small objects are challenging for conventional detectors to precisely identify and localize due to factors such low resolution and considerable form and size variability [17].

Apart from only using the YOLO model, some researchers make modifications to the YOLO architecture to get better accuracy. The authors in [6] and [7] combine YOLO with other modules. Changes and adjustments to the sizes of anchor boxes and bounding boxes are discussed in [8] and [9]. The authors in [18] pruned layers that were not used. These studies were successful in improving detection accuracy in their respective case studies. The most important aspects of the outcomes of modifications are accuracy and speed.

A modified YOLOv7 Tiny model is proposed in this study to detect bull sperm. The bull sperm sample discovered was a post-thawing sample of cattle sperm utilizing a skim-based diluent. The magnification factor is 200. To enhance mean average precision (mAP), improvements will be made to the number of layers and the size factor of anchor boxes and grid cells, among other things. The bounding box of the bull sperm identified will be more precisely defined if the mAP is better, and the number will be high.

METHOD

This study started by identifying real-world issues, specifically those pertaining to bull sperm detection, that can be resolved with information technology (IT). Several papers on the detection of bull sperm through image processing or object detection techniques were examined as part of the literature review. Reading references in the form of books and interviews can also provide general knowledge about topics like spermatozoa and artificial insemination techniques. Numerous studies have concluded that bull sperm detection accuracy based on research results is still subpar. DeepSperm is one model for identifying cow sperm. At very low magnification, DeepSperm can achieve a mAP of 64.45 when detecting cow sperm [18]. To ascertain the boundaries of the study and its ultimate requirements, the problem domain is identified. The dataset used consists of pictures or videos that show objects with bull sperm. At the Lembang Artificial Insemination Agency, data was gathered. A camera-equipped microscope was used to collect the data. Using a skimmer, the camera captures video of bull sperm after they have thawed.

The microscope has 200x magnification. On the obtained dataset, data cleaning and annotation were done. For the results of detecting the number of bull sperm to match the ground truth, a model that can detect bull sperm with good accuracy is required, according to the literature review. To improve accuracy, modifications are made to the model that will be employed. Before beginning the experimental phase, several things must be ready, including an examination of the YOLOv7 model's architecture, specifically the head, neck, and backbone. To ensure that the size matches the bounding box, the anchor box section is analyzed to determine the resolution to be used. As a guide for the techniques used for change and modification, a study of the DeepSperm model was also conducted. The size of the grid cells and feature maps will be determined by analyzing the used dataset. To create a model from the training data, experiments involve adjusting parameters and anchor box sizes. After the results of the experimental analysis are obtained, conclusions are drawn. To help readers comprehend the report's contents and identify areas that require improvement and correction for the case study on bovine sperm detection using the YOLOv7 model, conclusions are drawn. The image Figure 1 illustrates the research process.

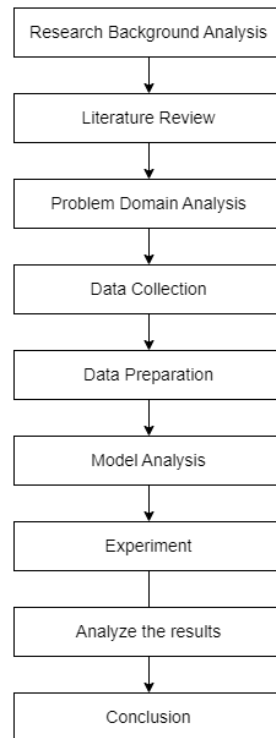Figure 1. Research Process

Dataset

We used secondary data in video format which came from the Lembang Artificial Insemination Agency [5]. The following are the specifications of the hardware and environment when the data were gathered. An optical microscope with a camera was used to acquire the data, whereas a mobile phone camera was used in [5]. After the bull sperm was thawed using a skim-based diluent, the camera recorded video of it. The magnification of the microscope is 200X. There are 50 image frames from the Lembang Artificial Insemination Agency and 50 image frames from [5] study. Pseudo-labeling is used to annotate datasets. The prediction results are used as the bounding box for this operation, and the results are manually rechecked to remove any inaccurate labels. This could hasten the annotation process for images.

Using augmentation, we increase the data in the datasets. Flip, rotate, grayscale, saturation, brightness, and exposure are some of the augmentations used. Data augmentation helps broaden the training sets' variation by producing different versions of the original photos. With a more varied dataset, the model can pick up more patterns and features [19]. The outcomes of the augmentation implementation are represented by 200 photos. When the augmentation results are included, the dataset consists of 300 images. The dataset is divided into 80:10:10 parts, with 80% of the data utilized for training, 10% for validation, and 10% for testing [20]. The proportion of the dataset is shown in Table 1.

Table 1. Dataset Proportion

Dataset	Value
Training	240
Validation	30
Testing	30

YOLOv7 Tiny Architecture Modification

The convolution layer with the LeakyRelu activation function is the first layer in the modified YOLOv7 Tiny architecture. Starting with the LeakyRelu convolution with a value of two times downsampling (2x) twice, a downsampling value of four times (4x) is obtained. Convolution is continued until the max pooling section is reached. Downsampling is repeated twice (2x) at the max pooling layer, for a total of eight times (8x) downsampling. Because the backbone downsampling results are sufficient, no further downsampling is necessary. In the head, convolution is carried out in the spatial pyramid (SP). A max pooling layer with a downsampling value of one is included in the spatial pyramid convolution. Finally, until it reaches the output layer, the LeakyRelu convolution is conducted as usual. LeakyRelu is used since it is one of the best activation functions [21]. LeakyReLU is a generalization or enhancement of Rectified Linear Unit (ReLU) that addresses the issue of ReLU's dead neurons. When the input is negative, ReLU will provide an output that is always 0, which prevents the neuron from updating the parameters. This condition is known as a dead neuron [22]. The modified YOLOv7 tiny architecture is shown in Figure 2 and the hyper parameters used are shown in Table 2.

Anchor Box

Anchor boxes are a set of initial candidate boxes with a fixed width and height. The choice of the initial anchor boxes will directly affect the detection accuracy and the detection speed [9]. The default YOLOv7 K-means clustering calculation results are used to calculate the size of the anchor box. The dimensions of the best anchor boxes for this study dataset are 11x18, 20x12, and 19x22. The clusters generated by K-means can reflect the distribution of the samples in each dataset, which can make it easier for the network to get good predictions [9].

Grid Cell

By using the approach of one one-stage detector YOLO, an input picture is distributed to a system of S×S grids. Each bounding box consists of 5 predictions: x, y, w, h, and confidence. The (x, y) coordinates represent the center of the box relative to the bounds of the grid cell. The width and height are predicted relative to the whole image. Finally, the confidence prediction represents the Intersection over Union (IOU) between the predicted box and any ground truth box [23]. It distributes the given picture to a grid of S×S cells. For every cell in the network, it computes confidence for 'n' bounding boxes. The predicted result is encoded into a tensor as $S \times S \times (n \times 5 + p)$ [23].

This study will put three different input image resolutions to the test. The images used as input are 704x704, 528x528, and 352x352. Grid cells (44x44), (66x66), and (88x88) will be created based on the provided resolution. The grid cell size is calculated by dividing the input image size by the maximum down sampling size, which is 8.

Table 2. Hyper Parameter

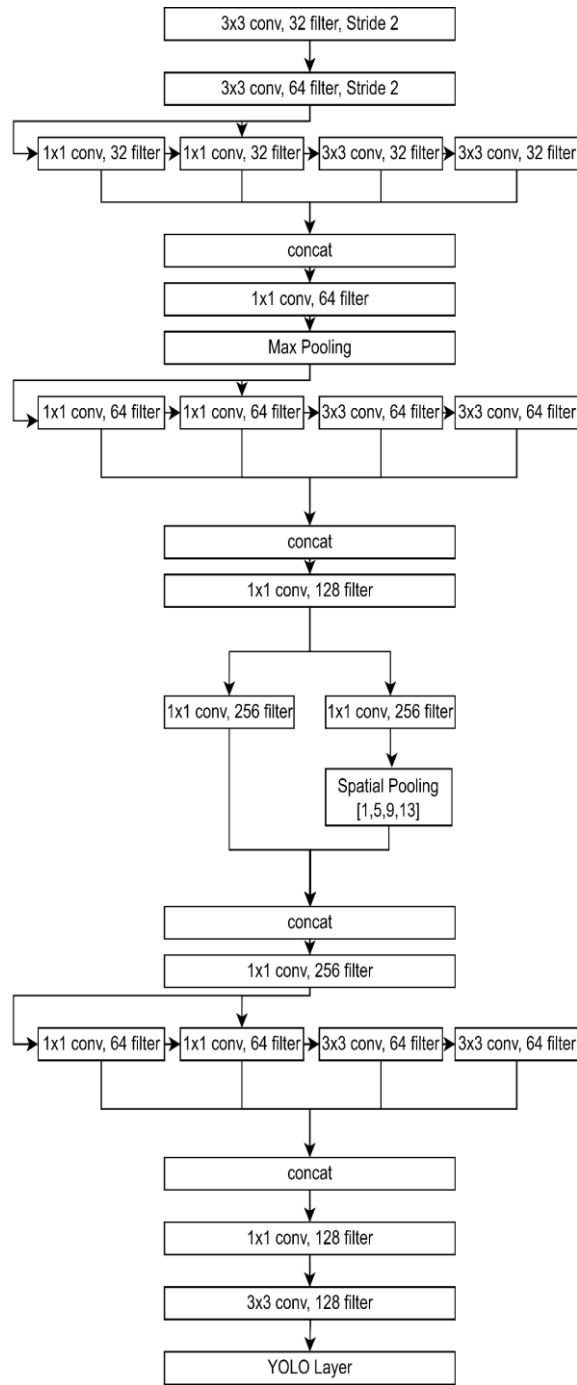| Hyper-parameter	Value
Learning rate	0.01
Epoch	200
Confidence threshold	0.3
Activation	LeakyReLU

Figure 2. Modified YOLOv7 Tiny Architecture

Training

A Tesla T4 Graphic Processing Unit (GPU) was used to perform the training and testing steps of the modified YOLOv7 Tiny model. We utilized the YOLOv7 default learning rate of 0.01 as our learning rate [24]. In this experiment, the epoch was 200. We chose 200 as the number of the epoch as a larger epoch did not provide better accuracy. Table 3 shows the hardware specifications that were used.

Table 3. Hardware Specification

Components	Specification
GPU	Tesla T4
Size	16 GB
Cuda Core	2560 Core
Technology	GDDR6 SDRAM
Effective Clock Speed	5001 MHz
Bus Width	256-bit
OS	Ubuntu 20.04 LTA
CPU	@2.3Ghz i.e(1 core, 2 threads)
RAM	13 GB
Storage	33 GB

Testing and Evaluation

The testing was done with a test dataset containing 30 images. The testing dataset consists of 14 images from the Lembang Artificial Insemination Agency and 16 images from [5]. YOLO's default setting states that 50% of the IOU is used for detection. We used 0.3 as the confidence level when performing detection because sperm cells are considered small objects.

The mean average precision (mAP) and inference time are utilized to evaluate the performance of the modified YOLOv7 tiny architecture in bull sperm detection. The model's detection accuracy for bull sperm is measured using mAP whereas detection speed is measured using inference time. There is a connection between these two matrices. According to deep learning theory, a higher accuracy usually corresponds to a slower detection speed, whereas a lower accuracy corresponds to a faster detection speed [25].

RESULTS AND DISCUSSION

At 200X magnification, YOLOv7 Tiny can detect bull sperm well. Even if the number of parameters is reduced significantly, the accuracy of the modified YOLOv7 Tiny detection is not significantly different. This demonstrates that the model can maintain detection accuracy even after the layer has been reduced. Moreover, after modifying the anchor box and grid cell, we achieved a higher speed and accuracy at the same time. The accuracy obtained can be preserved even though pruning and adjustments are made to the grid cells' and anchor boxes' sizes. Figure 3 shows the detection results using the YOLOv7 Tiny, whereas Figure 4 shows the detection results using the modified YOLOv7 Tiny model.

Figure 5Figure 4 demonstrates the fact that our modified model has better accuracy than the original one. In Figure 5(a) and Figure 5(b), YOLOv7 Tiny recognizes a microscopic artifact as bull sperm. In a blank area, Modified YOLOv7 Tiny, Figure 5(c) and Figure 5(d), detects bull sperm. In Figure 5(e) and Figure 5(f), YOLOv7 Tiny miss detects one bull sperm as two bull sperm. A little artifact in Figure 5(g) and Figure 5(h), is identified as bull sperm by YOLOv7 Tiny. The modified YOLOv7 Tiny ignores it.

Accuracy

During testing, the YOLOv7 Tiny model detected bull sperm with a mAP of 64.5%. On the modified YOLOv7 Tiny model, the value of mAP testing increased by 1.3%, to 65.8%. The accuracy comparison is shown in Figure 6.
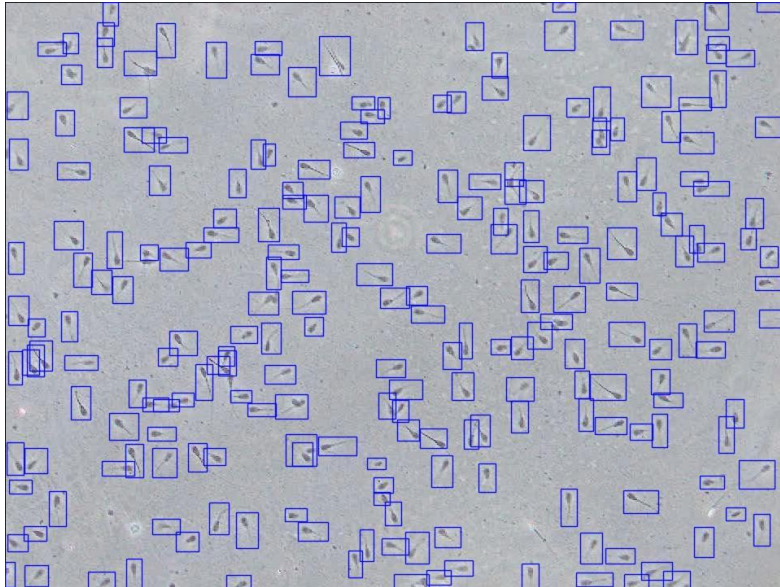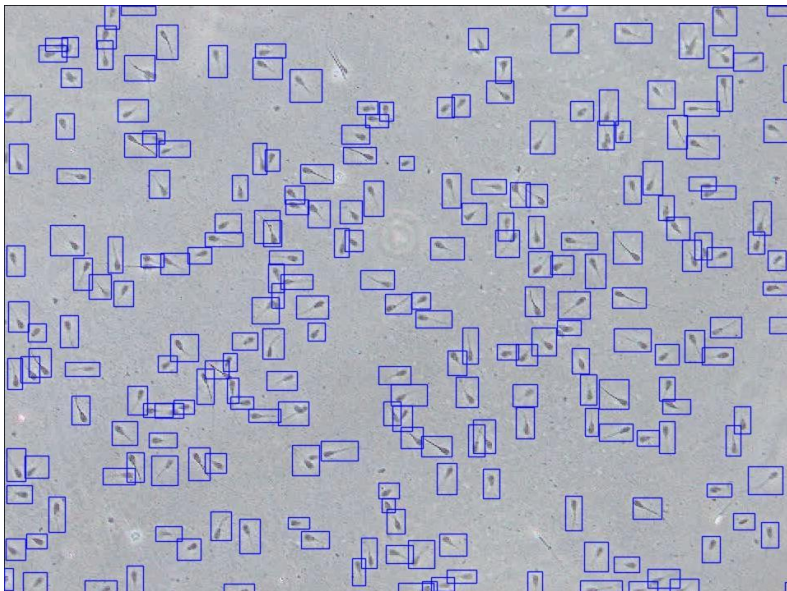
Figure 3. YOLOv7 Tiny

Figure 4. Modified YOLOv7 Tiny

Inference Time

In detecting bull sperm, the inference time on YOLOv7 Tiny is 17.7 ms whereas the inference time in the modified YOLOv7 Tiny model is 14.4 ms. The modified YOLOv7 Tiny is 3.3 milliseconds faster than the original YOLOv7 Tiny. Figure 7 shows a comparison of the inference time.
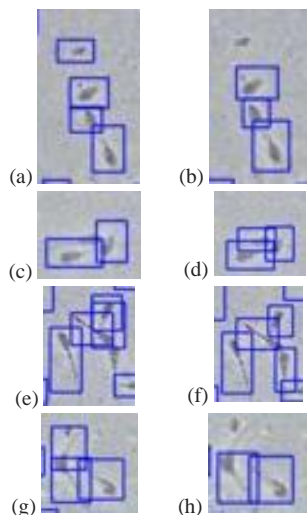
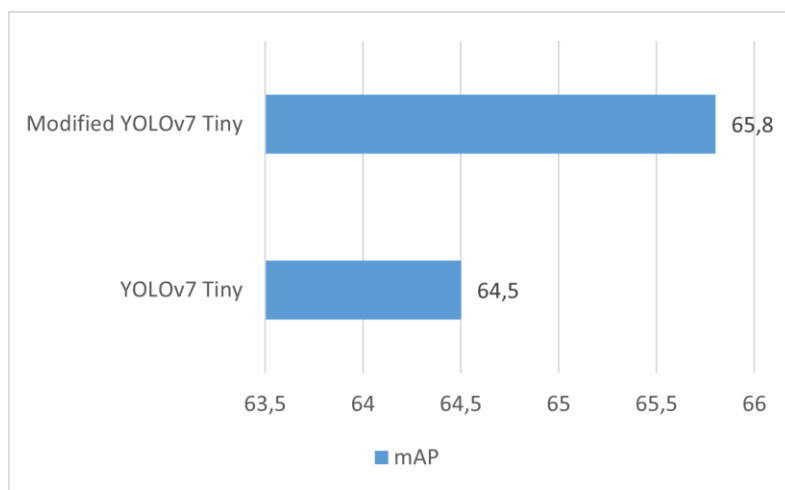Figure 5. Detection result of (a), (c), (e), (g) YOLOv7 Tiny and (b), (d), (f), (h) Modified YOLOv7 Tiny

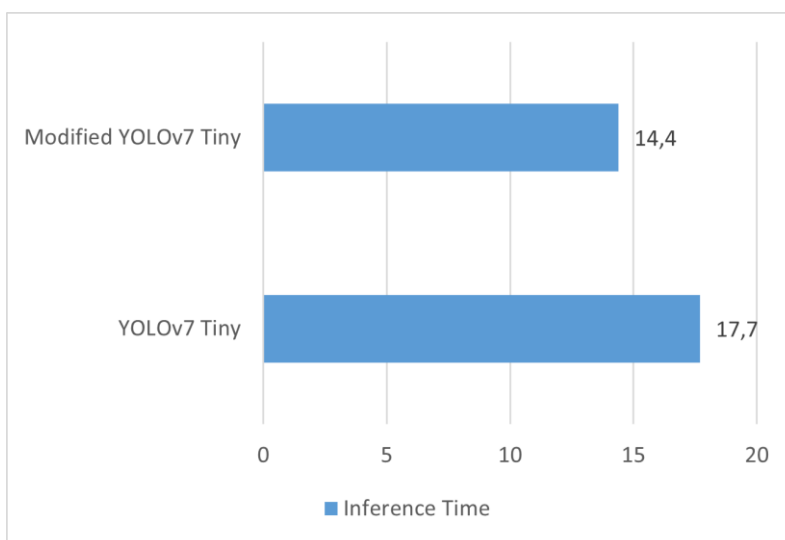Figure 6. Accuracy Comparison on Test Dataset

Figure 7. Inference Time Comparison on Test Dataset

Discussion

The LeakyRELU activation function is utilized in this study, which is similar to [18]. LeakyReLU is a generalization or improvement of Rectified Linear Unit (ReLU) to overcome dead neurons that arise in ReLU [22]. The modified YOLOv7 tiny concentrates object detection on small object detection by removing the architecture components that are designed to detect large and medium objects.

In contrast to DeepSperm [18], which focuses on YOLOv3 improvement to measure bull sperm motility, this research concentrates on YOLOv7 improvement to quantify the number of sperm in a single field of view. Both studies performed architectural changes. Nevertheless, this research's architectural size is 74.2% smaller than DeepSperm and it can employ random input images and videos, whereas DeepSperm cannot. In contrast to DeepSperm, which does not use spatial pooling, this study takes advantage of it to gather feature maps from various backbone stages.

The modified YOLOv7 Tiny has a total of 914,852 parameters. The number of modified YOLOv7 Tiny parameters is 84.2% less than the number of the original YOLOv7 Tiny parameters. It is possible because the architecture for detecting large and medium-sized objects has been removed. The modified YOLOv7 Tiny received a mAP value of 65.8%.

The number of parameters influences the size of the model's weight file. The file grows as the number of parameters increases. The modified YOLOv7 Tiny inference time required to detect bull sperm is 14.4 ms. The modified YOLOv7 Tiny inference time is 3.3 ms faster than the YOLOv7 Tiny inference time. Table 4 shows a comparison of the YOLOv7 Tiny and modified YOLOv7 Tiny model.

The modified YOLOv7 Tiny in this study has improved speed and accuracy in detecting and counting the quantity of bull sperm. However, some limitations apply to this study. This study can only count the number of bull sperm in a single field of view and cannot be used to count the number of bull sperm in one dose of artificial insemination. Other than that, bull sperm motility has not been identified by this research. As a result, the predefined bull sperm quality criteria [3] cannot be accurately applied to this study's evaluation of bull sperm quality. Further research is recommended to count the number of bull sperm for one straw or one dose of artificial insemination so that the quality and quantity of bull sperm can be checked based on predetermined provisions [3].

Table 4. YOLOv7 Tiny and Modified YOLOv7 Tiny Comparison

Architecture	YOLOv7 Tiny	Modified YOLOv7 Tiny
Parameter	6,007,596	**914.852**
mAP	64.5%	**65.8%**
Number of detection layer outputs	3	**1**
Grid Cell (input size 704 pixels)	44x44 66x66 88x88	**88x88**
Weight file	11.73 MB	**1.85 MB**
Inference time	17.7 ms	**14.4 ms**

CONCLUSION

In this study, changes in the size of anchor boxes and grid cells improve the accuracy of YOLOv7 in detecting objects. To increase processing performance while retaining accuracy, all relevant structures for detecting large and medium-sized objects are removed. Changes and removal must be made following the characteristics of the item to be detected. Changing the YOLOv7 architecture based on the features of the item to be detected can improve detection

accuracy. YOLOv7 architectural improvements based on bull sperm characteristics have resulted in improved accuracy. The mAP accuracy of the modified YOLOv7 Tiny is 65.8%, with a detection speed of 17.7 ms. The original YOLOv7 Tiny achieves an accuracy of 64,5% mAP with a detection speed of 14.4ms. According to these data, the modified YOLOv7 Tiny is 1.3% more accurate and 1.23x faster than the original YOLOv7 Tiny. In terms of accuracy and speed, the modified YOLOv7 Tiny model is superior.

ACKNOWLEDGMENT

REFERENCES

[1] Sukentar, "Inseminasi Buatan Pada Sapi." Accessed: Dec. 11, 2022. [Online]. Available: http://cybex.pertanian.go.id/mobile/artikel/90837/Inseminasi-Buatan-Pada-Sapi/
[2] I. Diarmita, "Pedoman Pelaksanaan UPSUS SIWAB 2019," 2019.
[3] BSN, "SNI 4869-1:2017 Semen beku – Bagian 1: Sapi," p. 2, 2017.
[4] P. Hidayatullah, T. L. E. R. Mengko, R. Munir, and A. Barlian, "Bull Sperm Tracking and Machine Learning-Based Motility Classification," *IEEE Access*, vol. 9, pp. 61159–61170, 2021, doi: 10.1109/ACCESS.2021.3074127.
[5] F. Mutia, "Algoritme pemrosesan video untuk pengukuran kualitas semen sapi dengan teknik optical flow," Institut Teknologi Bandung, 2019.
[6] F. Wu, G. Jin, M. Gao, Z. He, and Y. Yang, "Helmet detection based on improved YOLO V3 deep model," *Proc. 2019 IEEE 16th Int. Conf. Networking, Sens. Control. ICNSC 2019*, pp. 363–368, 2019, doi: 10.1109/ICNSC.2019.8743246.
[7] X. Zhou, L. Jiang, C. Hu, S. Lei, T. Zhang, and X. Mou, "YOLO-SASE: An Improved YOLO Algorithm for the Small Targets in Complex Backgrounds," *Sensors*, vol. 22, no. 12, pp. 1–14, 2022, doi: 10.3390/s22124600.
[8] Y. Guo and M. Zhang, "Blood Cell Detection Method Based On Improved YOLOv5," *IEEE Access*, vol. 11, no. July, pp. 67987–67995, 2023, doi: 10.1109/ACCESS.2023.3290905.
[9] M. Ju, H. Luo, Z. Wang, B. Hui, and Z. Chang, "The application of improved YOLO V3 in multi-scale target detection," *Appl. Sci.*, vol. 9, no. 18, 2019, doi: 10.3390/app9183775.
[10] P. Hidayatullah and M. Zuhdi, "Automatic sperms counting using adaptive local threshold and ellipse detection," *2014 Int. Conf. Inf. Technol. Syst. Innov. ICITSI 2014 - Proc.*, no. November, pp. 56–61, 2014, doi: 10.1109/ICITSI.2014.7048238.
[11] T. Sato *et al.*, "A new deep-learning model using YOLOv3 to support sperm selection during intracytoplasmic sperm injection procedure," *Reprod. Med. Biol.*, vol. 21, no. 1, pp. 1–8, 2022, doi: 10.1002/rmb2.12454.
[12] Aristoteles, A. Syarif, Sutyarso, F. R. Lumbanraja, and A. Hidayatullah, "Identification of Human Sperm based on Morphology Using the You Only Look Once Version 4 Algorithm," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 7, pp. 424–431, 2022, doi: 10.14569/IJACSA.2022.0130752.
[13] M. Dobrovolny, J. Benes, J. Langer, O. Krejcar, and A. Selamat, "Study on Sperm-Cell Detection Using YOLOv5 Architecture with Labaled Dataset," *Genes (Basel).*, vol. 14, no. 2, 2023, doi: 10.3390/genes14020451.
[14] Y. Bai, Y. Zhang, M. Ding, and B. Ghanem, "Finding Tiny Faces in the Wild with Generative Adversarial Network," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 21–30, 2018, doi: 10.1109/CVPR.2018.00010.
[15] C. Xu *et al.*, "Dynamic Coarse-to-Fine Learning for Oriented Tiny Object Detection," pp. 7318–7328, 2023, doi: 10.1109/cvpr52729.2023.00707.
[16] H. Lai, L. Chen, W. Liu, Z. Yan, and S. Ye, "STC-YOLO: Small Object Detection Network for Traffic Signs in Complex Environments," *Sensors*, vol. 23, no. 11, 2023, doi: 10.3390/s23115307.
[17] Y. Bai, Y. Zhang, M. Ding, and B. Ghanem, *SOD-MTGAN: Small object detection via multi-task generative adversarial network*, vol. 11217 LNCS. Springer International Publishing, 2018. doi: 10.1007/978-3-030-01261-8_13.

[18] P. Hidayatullah *et al.*, "DeepSperm: A robust and real-time bull sperm-cell detection in densely populated semen videos," *Comput. Methods Programs Biomed.*, vol. 209, p. 106302, 2021, doi: 10.1016/j.cmpb.2021.106302.

[19] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *J. Big Data*, vol. 6, no. 1, 2019, doi: 10.1186/s40537-019-0197-0.

[20] M. Bernico, "Deep Learning Quick Reference: Useful hacks for training and optimizing deep neural networks with TensorFlow and Keras," *Packt Publishing*, 2018.

[21] B. Xu, N. Wang, T. Chen, and M. Li, "Empirical Evaluation of Rectified Activations in Convolutional Network," 2015, [Online]. Available: http://arxiv.org/abs/1505.00853

[22] J. Xu, Z. Li, B. Du, M. Zhang, and J. Liu, "Reluplex made more practical: Leaky ReLU," *Proc. - IEEE Symp. Comput. Commun.*, vol. 2020-July, 2020, doi: 10.1109/ISCC50000.2020.9219587.

[23] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem, pp. 779–788, 2016, doi: 10.1109/CVPR.2016.91.

[24] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," pp. 1–15, 2022, [Online]. Available: http://arxiv.org/abs/2207.02696

[25] I. Goodfellow, Y. Bengio, and Aaron Courville, "Deep Learning," *Foreign Affairs*, vol. 91, no. 5, pp. 1689–1699, 2016.