Identification of Human Sperm based on Morphology Using the You Only Look Once Version 4 Algorithm

Aristoteles[1], Admi Syarif[2]*, Sutyarso[3], Favorisen R. Lumbanraja[4], Arbi Hidayatullah[5]
Doctoral Program of Mathematics and Natural Sciences, Lampung University[1]
Department of Computer Science, Faculty of Mathematics and Natural Sciences[1, 2, 4, 5]
Lampung University, Bandar Lampung, Indonesia[1, 2, 4, 5]
Department of Biology, Faculty of Mathematics and Natural Sciences[3]
Lampung University, Bandar Lampung, Indonesia[3]

***Abstract*—Infertility is a crucial reproductive problem experienced by both men and women. Infertility is the inability to get pregnant within one year of sexual intercourse. This study focuses on infertility in men. Many causes that can cause infertility in men including sperm quality. Currently, identification of human sperm is still done manually by observing the sperm with the help of humans through a microscope, so it requires time and high costs. Therefore, high technology is needed to determine sperm quality in the form of deep learning technology based on video. Deep learning algorithms support this research in identifying human sperm cells. So deep learning can help detect sperm video automatically in the process of evaluating sperm cells to determine infertility. We use deep learning technology to identify sperm using the You Only Look Once version 4 (YOLOv4) algorithm. Purpose of this study was to analyze the level of accuracy of the YOLOv4 algorithm. The dataset used is sourced from a VISEM dataset of 85 videos. The results obtained are 90.31% AP (Average Precision) for sperm objects and 68.19% AP (Average Precision) for non-sperm objects, then for the results of the training obtained by the model 79.58% mAP (Mean Average Precision). Our research show result about identification of human sperm using YOLOv4. The results obtained by the YOLOv4 model can identify sperm and non-sperm objects. The output on the YOLOv4 model is able to identify objects in the test data in the form of video and image.**

***Keywords*—Classification; deep learning; identification; sperm; sperm head; you only look once version 4**

I. Introduction

In the last two decades, the reproductive problem in men that has received much attention is infertility. Infertility is caused by many things, one of which is abnormalities in sperm morphology. Morphological abnormalities experienced by sperm include thin heads, amorphous heads, or bent or asymmetrical neck is of little clinical use [1]. Many studies have reported some analytical disturbances of sperm morphology tests in the details of the sperm sections that were carried out manually. Several studies related to technology to support the diagnosis of infertility in sperm include Computer Assisted Sperm Analysis (CASA), Automatic Assessment of Biochemical Markers of Seminal Plasma, Histopathology Assessment [2]. The technology development that has been carried out still requires further development to get better results, in order to be able to get accurately analyze and identify infertility problems. Currently, there are many studies that predict the cases of infertility. The method that is commonly implemented is the observation method from patient medical record data at a hospital [3]. A number of studies have shown that the factors that cause infertility include age, smoking habits, marijuana use, heroin, hormone disorders, and immunological disorders [4]. These factors can increase the risk of having abnormal sperm so that it becomes infertility.

Identification of sperm is still mostly done manually by humans by observing directly with the aid of a microscope. This requires a high time and cost. To overcome these problems needed a high technology, it is using deep learning. Research on deep learning has been widely carried out [5, 6, 7, 8, 9].

Convolutional Neural Network (CNN) is a development of deep learning that has been developing since 2012. This method can identify sperm morphology accurately based on images [5]. Accurate identification results are influenced by the quality of full image data or with large pixel sizes, so that accuracy and detection performance can be more optimal [6]. In addition to analyzing the morphology of the sperm, CNN can identify sperm based on motility. The dataset used is video [7]. The best results were obtained at the Mean Average Error (MAE) which was 8.786. This shows that the prediction of sperm motility is a fast and consistent process [8]. In the study [9] used Region Based Convolutional Neural Network (R-CNN) to evaluate sperm head motility in video data [7]. The results obtained in this study were 91% and MAE was 2.92.

Research [8, 9] has not been able to identify sperm and non-sperm objects through video. Therefore, our research carried out the latest and most updated breakthrough in the form of YOLO (You Only Look Once). The YOLO algorithm is a floating of CNN which functions as object detection in multiple images [10]. The YOLO algorithm is a more efficient method than object detection algorithms in other machine learning, because it makes everyone can use a 1080 Ti or 2080 Ti GPU to train a super-fast and accurate object detector [11]. Therefore, it is necessary to build a model using the YOLO method to identify sperm and non-sperm based on video.

*Corresponding Author.

II. Materials and Methods

The process conducted in this study is illustrated in Fig. 1.

Fig. 1 shows an illustration of the workflow this research with many steps. The first step is to collect images for the dataset, then resize the image and annotate it based on the object class. The second step is split data to generate train, validation, and test data. The next step is training data based on predetermined hyperparameters. The last step is to evaluate the model from the results of training data and data testing.

A. Dataset

The data used is from Simula Open Dataset with the address https://datasets.simula.no/visem/. This VISEM dataset is a multi-modal dataset that contains data sources such as videos, biological analysis data, and participant data, however in its use for this study only data in the form of videos are used as research datasets. The VISEM dataset contains 85 video recordings of anonymous data from 85 different donor participants with the AVI (Audio Video Interleave) extension and the resolution of each video is 640 x 480 pixels with 50 fps frame rate. Based on the video, 1330 images are produced which will be processed. Fig. 2 shows a piece frame of microscopic video the VISEM dataset.

B. Annotation Data

In this step, two object annotation classes are given for the image to be trained. The classes created in this annotation are sperm and non-sperm. Each frame is annotated in the form of bounding boxes on the morphology of the sperm head and an object that is not sperm. The results of annotations have been made for sperm class is 105.465 bounding boxes. While for the non-sperm class, 22.425 bounding boxes have been annotated. For this annotation use Yolo_mark which is sourced from GitHub https://github.com/AlexeyAB/Yolo_mark.git. The following display of the data annotation can be seen in Fig. 3.

C. Deep Learning

Deep Learning is a branch of machine learning that is inspired by the human cortex by applying an artificial neural network with many hidden layers [12]. There are many types of Deep Learning, such as Deep Auto Encoder, Deep Belief Nets, Convolutional Neural Network, and others. Deep Learning can solve computer difficulties in understanding the meaning of raw input data that is by breaking the desired complex mapping into a series of nested simple mappings.

D. Convolutional Neural Network (CNN)

Convolutional Neural Network is a subdivision of a Deep Learning algorithm used in computer vision to solve certain cases or problems, such as classifying and detecting objects in images, photos, or videos [13]. The characteristics of CNN have a 3D arrangement of neurons (height, width, and depth). The illustration of CNN architecture can be seen in Fig. 4.

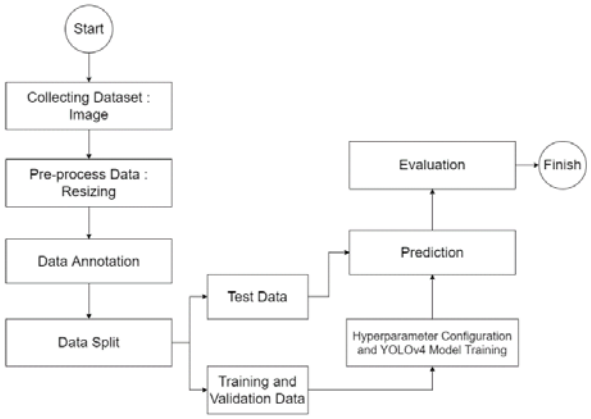

Fig. 1. Research Workflow View.

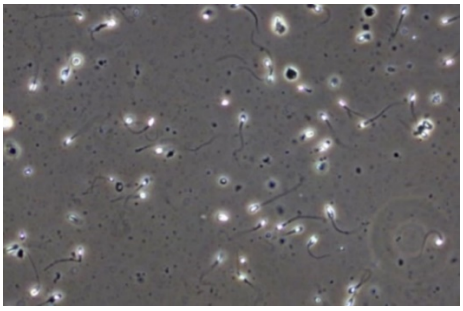

Fig. 2. Frame from Microscopic Video of VISEM Dataset.

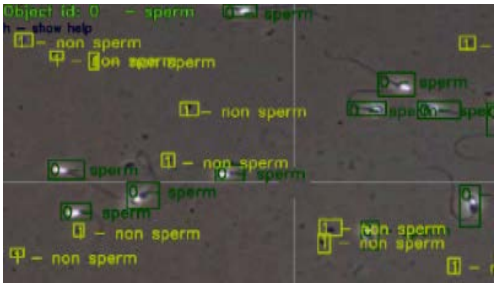

Fig. 3. Display Data Annotation of Image Extraction VISEM Dataset.

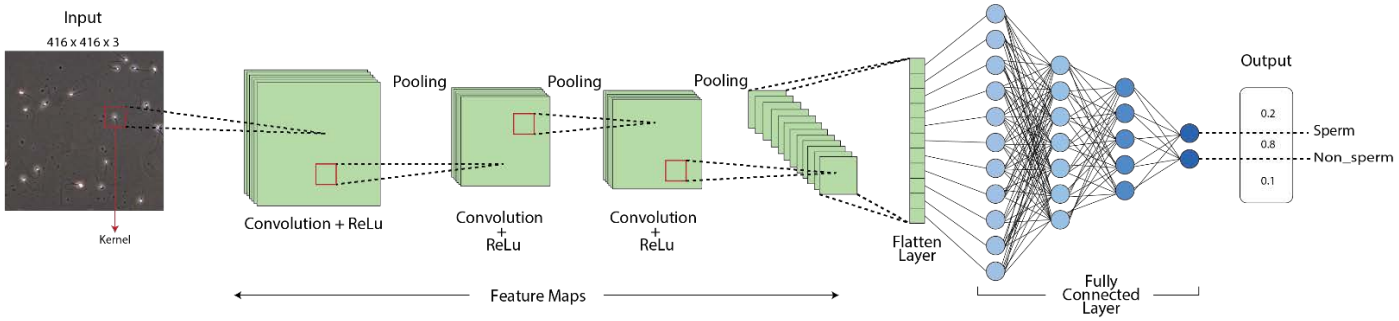

Fig. 4. The Architecture of Convolutional Neural network (CNN) [13].

Fig. 4 shows processes to receive image prediction from CNN. Here we can see the CNN processes are input image, convolution, ReLu, pooling, and fully connected layer.

1) Convolution: The mathematical definition of convolution is the total number of multiplications between the corresponding elements (having the same coordinates) in two matrices or two vectors [14]. Convolution can also be defined as the process of multiplying an image using an external mask or sub-windows to create a new image.

2) Pooling: The pooling layer is a layer that has the function to reduce the spatial size of the convolution process. So as to reduce the computational resources required to process data by reducing the dimensions of the feature map. This pooling layer can make the model training process more effective because the pooling layer is the dominant feature extraction [15].

3) Rectified Linear Units (ReLU): ReLU is the part of the linear function code that removes the negative part to zero and keeps the positive part of the convolution result. Many studies have shown that ReLU outperforms the sigmoid activation function and empirical ground [16]. ReLu activation function is defined as:

$$a_{i,j,k} = max(z_{i,j,k}, 0) \quad (1)$$

Input of the activation function is $z_{i,j,k}$ at location (i, j) on the k-th channel. Simply put, ReLU outputs 0 when $a_{i,j,k} < 0$, and otherwise, it outputs a linier functions when $a_{i,j,k} \geq 0$. Fig. 5 is visual representation of ReLU activation function.

4) Fully connected layer: Fully Connected Layer is a layer that is fully connected; this layer of neurons is connected directly to other neurons by two adjacent layers without being connected to any layer [17]. The Fully Connected Layer processes the output of the final pooling or convolutional layer, which has been flattened. The results of this process will then be continued using the softmax function to get the probability of the input being in a certain class [18].

E. You Only Look Once (YOLO)

You Only Look Once or YOLO is a new approach to object detection and development of CNN. YOLO differs from previous research in that detecting an object reuses its classifier; instead YOLO frames object detection as a regression problem with spatially separated bounding boxes and associated class probabilities [10].

Fig. 6 illustrates the YOLO process in making the input image into an image that has been given a bounding box. The first step is the input image will be resized, and then the second step runs a convolutional neural network, after that it does non-max suppression and produces an image that has been identified with a bounding box.

YOLO is implemented as a convolutional neural network. This architecture is inspired by the GoogleNet model for image classification. The YOLO network has 24 convolution layers followed by 2 fully connected layers. Simply uses 1×1 reduction layer followed by 3×3 convolutional layers. The prediction of the final output of this YOLO network is a tensor of 7×7×30. Fig. 7 shows the YOLO architecture in processing image predictions.

There are several versions YOLO of the development of research that has been carried out, in this study using YOLOv4. YOLOv4 overcomes this problem by creating a CNN that operates in real-time on a conventional GPU, and requires only one conventional GPU for training. The purpose of YOLOv4 is to design the speed of operation of the object detector in producing systems and optimization for parallel computing. YOLOv4 hopes that the designed object can be easily trained and used.

Modern detectors usually consist of two parts consisting of a backbone and a head, for a backbone that has been trained previously with ImageNet. Head used to predict the class and bounding boxes of the object. For detectors running on the GPU platform, the backbone used can be VGG, ResNet, ResNetXt, or DenseNet. For detectors running on the CPU platform, the backbone used is SqueezeNet, MobileNet, or ShuffleNet. The development of object detectors in recent years often inserts several layers between the backbone and the head, these layers are usually used to collect feature maps from different stages. We can call this layer the neck object detector. In general, the neck consists of several bottom-up paths and topdown paths. Networks equipped with this mechanism include the Feature Pyramid Network (FPN), Path Aggregation Network (PAN), BiFPN, and NAS-FPN [11].

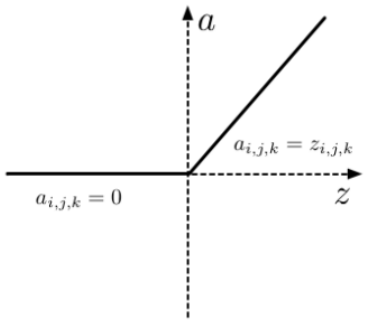

Fig. 5. Display Representation of ReLU Activation Function [16].

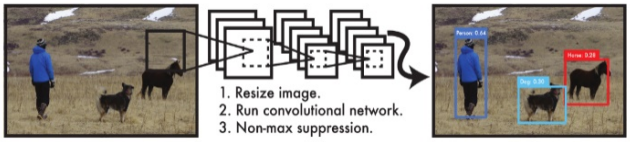

Fig. 6. Processing Images with YOLO Detection System [10].

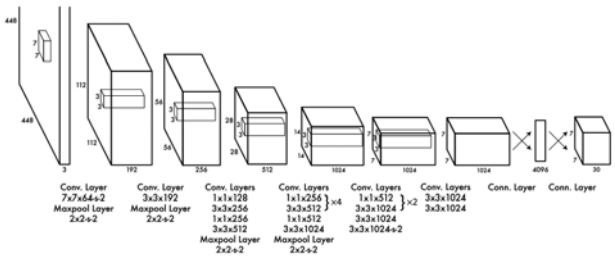

Fig. 7. Display YOLO Architecture [9].

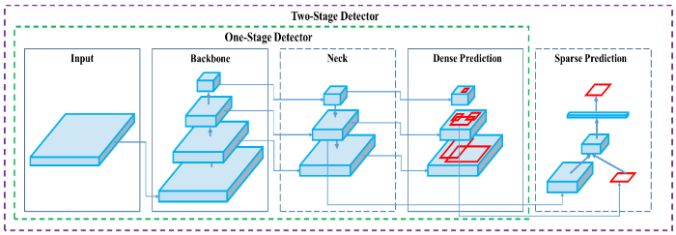

Fig. 8. Structure Modern Detector of YOLOv4 [10].

Based on Fig. 8, YOLO research uses modern detectors, so the researchers make several terms; there are Bag of Freebies (BoF) and Bag of Specials (BoS). The definition of the Bag of Freebies (BoF) is that researchers can perform an optimization to produce better accuracy and not increase inference costs by using training methods. The BoF used for the backbone are DropBlockRegularization, Class Label Smoothing, and CutMix and Mosaic Data Augmentation. BoF for this backbone is useful for increasing the variability in the input image so that the model built has a higher quality for images obtained from different environments. The definition for Bag of Special (BoS) is a set of plugin modules and post-processing methods that only increase inference costs by a small amount, however can significantly improve accuracy in object detection. The BaS used in the backbone include Mish Activation, Cross-Stage Partial Connections (CSP), Multi-Input Weight Residual Connection (MiWRC) [12].

F. Confusion Matrix

Confusion Matrix is a performance measurement or performance in solving machine learning classification problems, where the output results can be in the form of two or more classes. Confusion matrix is a predictive analysis tool that displays and compares the actual value with the predicted model value. Prediction models that can be used to get the results of the evaluation matrix are Accuracy, Precision, Recall and F1 Score [19]. This confusion matrix is shown in Fig. 9.

		Actual Values	
		1 (*Positive*)	0 (*Negative*)
Predicted Values	1 (*Positive*)	TP (*True Positive*)	FP (*False Negative*)
	0 (*Negative*)	FN (*False Negative*)	TN (*True Negative*)

Fig. 9. Display Component of Confusion Matrix.

TP (True Positive): The amount of data that is positive and is predicted to be true as positive.

FP (False Positive): The amount of data that is negative however is predicted to be positive.

FN (False Negative): The amount of data that is positive however is predicted to be negative.

TN (True Negative): The amount of data that has a negative value with a correct prediction as negative.

1) Accuracy: Accuracy is a ratio of selected relevant objects to all selected objects. Accuracy can also be defined as a comparison of an object that is correctly identified with the total number of existing objects and the error rate is an object that is identified incorrectly with the total number of existing objects [20].

$$Accuracy = \frac{TP+TN}{TP+TN+FF+FN} \times 100\% \quad (2)$$

2) Precision: Precision is a level of accuracy of information desired by the user with the prediction results given by the model or system [21].

$$Precision = \frac{TP}{TP+FP} \times 100\% \quad (3)$$

3) Recall: Recall is the ratio of the number of objects that are detected correctly or True Positive compared to all positive data, recall that has a high value means that the system or model created can classify object classes correctly [22].

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (4)$$

4) F1 Score: F1 Score or called the harmonic mean, is a picture of the relative influence between precision and recall [23].

$$F1\ Score = 2 \times \frac{Recall \times Presisi}{Recall + Presisi} \quad (5)$$

G. mAP (Mean Average Precision)

Mean Average Precision or mAP is used as the work evaluation value of the object detection model. Mean Average Precision measures the performance level of the file weights resulting from training mode [24]. The solid mAP equation can be seen in the following equation.

$$mAP = \frac{1}{c}\sum_{t=1}^{c} AP_i \quad (6)$$

III. Result and Discussion

A. Train Model Results

This study uses 3 types of dataset distribution in terms of finding the best level of accuracy. The first division structure is 80% train data, 10% validation data, and 10% test data. The next data division is 70% train data, 25% validation data, and 5% test data. The final data distribution is 60% train data, 20% validation data, and 20% test data. Based on the 3 types of split data, it produces different accuracy values, however the difference in the accuracy values obtained is not too significant. The number of iterations carried out during the training of this model is 6000 iterations. The learning rate was chosen for hyperparameter in model YOLOv4 sperm detection: 0.002, 0.0002, and 0.00002 [9]. The following is a table of accuracy results obtained from training data.

In Table I, testing of all scenarios uses a hyperparameter learning rate of 0.002. We can see that the AP results of each object being trained have the greatest results in two different scenarios. The biggest AP result for sperm objects is in the second scenario with a value of 90.31%. As for the non-sperm object, the biggest result is in the first scenario with a value of

68.35%. The biggest mAP result is in the first scenario of 78.81%.

Table II is a test using a learning rate of 0.0002 from three scenarios. The results obtained from this test are for the biggest AP of the two objects in the second scenario. The AP obtained for sperm objects is 90.37%, while the AP for non-sperm objects is 68.78%. The biggest mAP result is in the second scenario of 79.58%.

Table III uses a learning rate of 0.00002 to get AP results in each scenario. In this training, the biggest AP value for sperm objects is found in the third scenario, with an AP value of 88.42%. As for the AP value of non-sperm objects, the biggest AP result is in the second scenario of 64.40%. The biggest mAP result is in the second scenario of 76.11%.

TABLE I. ACCURACY RESULTS OBTAINED FROM DATA TRAINING WITH A LEARNING RATE OF 0.002

No	Data Composition	AP		mAP
		Sperm	*Non Sperm*	
1	Train 80%, Validation 10%, Test 10%	89.51%	68.13%	78.81%
2	Train 70%, Validation 25%, Test 5%	90.31%	65.35%	77.83%
3	Train 60%, Validation 20%, Test 20%	88.52%	65.24%	76.88%

TABLE II. ACCURACY RESULTS OBTAINED FROM DATA TRAINING WITH A LEARNING RATE OF 0.0002

No	Data Composition	AP		mAP
		Sperm	*Non Sperm*	
1	Train 80%, Validation 10%, Test 10%	89.66%	67.99%	78.83%
2	Train 70%, Validation 25%, Test 5%	90.37%	68.78%	79.58%
3	Train 60%, Validation 20%, Test 20%	89.86%	67.42%	78.64%

TABLE III. ACCURACY RESULTS OBTAINED FROM DATA TRAINING WITH A LEARNING RATE OF 0.00002

No	Data Composition	AP		mAP
		Sperm	*Non Sperm*	
1	Train 80%, Validation 10%, Test 10%	86.94%	62.54%	74.74%
2	Train 70%, Validation 25%, Test 5%	87.82%	64.41%	76.11%
3	Train 60%, Validation 20%, Test 20%	88.42%	63.10%	75.76%

Table I, Table II, and Table III shows that the mAP generated based on several data sharing results in an accuracy range of 74% - 79%. The result of the highest training data accuracy is 79.58% which is found in the distribution of 70% test data, 25% validation data, 5% test data, with a learning rate of 0.0002. The results of the lowest training accuracy are 74.74% which are found in the distribution of 80% test data, 10% validation data, 10% test data, with a learning rate of 0.00002. AP results from Sperm and Non Sperm objects differ greatly. This is due to the fact that the number of datasets used for training is imbalanced data. The best AP result in identifying sperm objects is 90.31% and for non-sperm objects it is 68.13%. The biggest results are obtained from the sharing of data and different learning rates.

B. Graphical Results of Precision, Recall, F1-Score and on Train Model

The results of the tests carried out in this study obtained the values of precision, recall and F1-score of several types of learning rates. The results obtained in the model training are good, because the results of each precision, recall and F1-score do not experience very large differences in values. Based on each learning rate that produces the highest value, there is a learning rate of 0.0002 with an average value of 0.8 of the precision, recall and F1-score values. The results of precision, recall, and F1-score illustrate that the model that has been made can predict and retrieve information well. The graph can be seen in the Fig. 10, Fig. 11, and Fig. 12.

Fig. 10 shows a graph of the values of precision, recall, and F1-score of three scenarios based on a learning rate of 0.002. It can be seen that the values of precision, recall, and F1-score in the second and third scenarios have a slight difference. However, the second scenario has a value that is superior to precision and F1 scores with values of 0.75 and 0.81. While the third scenario outperformed the recall value of 0.88, only 0.01 difference from the recall in the second scenario.

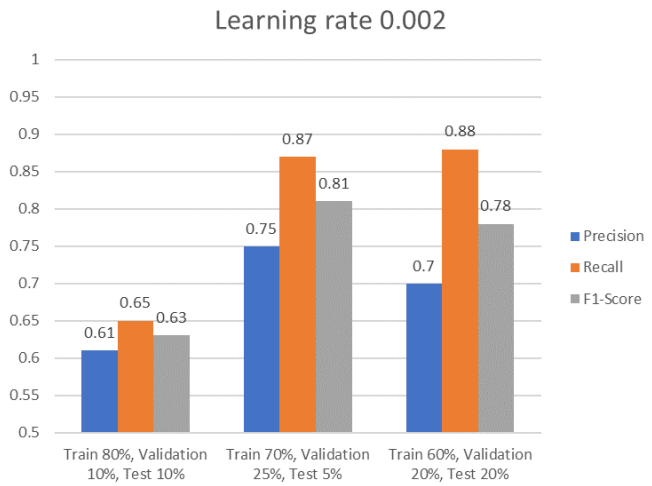

Fig. 10. Graph of Precision, Recall and F1-Score Values at a Learning Rate of 0.002.

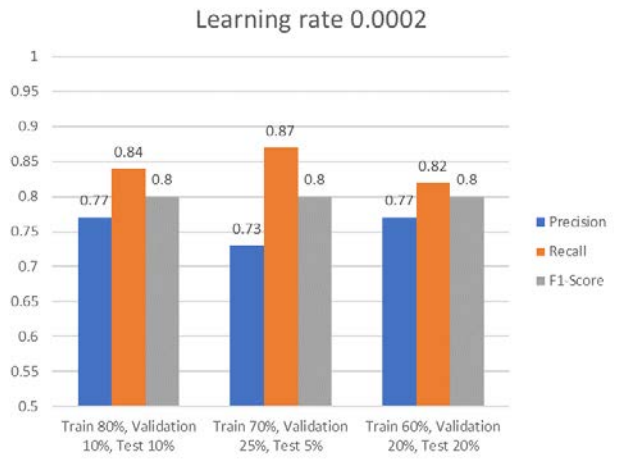

Fig. 11. Graph of Precision, Recall and F1-Score Values at a Learning Rate of 0.0002.

Fig. 11 shows a graph of the values of precision, recall, and F1-score of three scenarios based on a learning rate of 0.0002. It can be seen that the values of precision, recall, and F1-score get a slight difference in the values of the three scenarios. The highest precision value is found in the first and third scenarios with a value of 0.77. then the largest recall value is in the second scenario with a value of 0.87. While the relative F1-Score has the same value in the three scenarios with a value of 0.8.

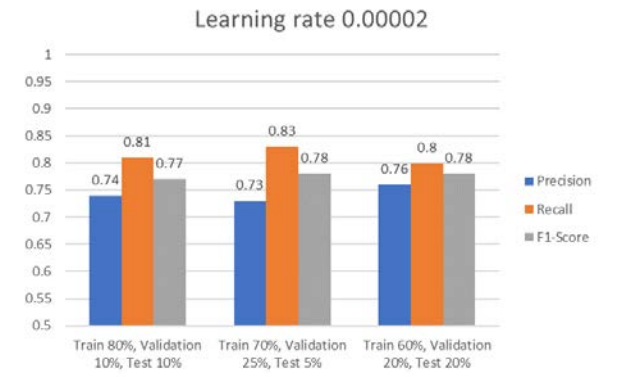

Fig. 12. Graph of Precision, Recall and F1-Score Values at a Learning Rate of 0.00002.

Fig. 12 shows a graph of the values of precision, recall, and F1-scores from three scenarios based on a learning rate of 0.00002. It can be seen that the value of precision, recall, and F1-score in the second scenario has a superior recall value with a value of 0.83. Meanwhile, the highest precision value is found in the third scenario with a value of 0.76. Then for the largest F1-Score value in the second and third scenarios with a value of 0.78.

C. Overfitting Handling on YOLOv4 Model

Overfitting occurs when the amount of training data used a slight variation. Based on the training data in this study using training data that has many variations. Training data is taken from several pieces of video frames on the VISEM dataset. YOLOv4 model in doing overfitting analysis according to Alexey the creator of the YOLOv4 model, validation loss in YOLOv4 should not be too much attention because it will tend to decrease continuously, so only mAP accuracy must be considered. If there is a stagnation in the data training process, it is likely that overfitting has occurred. Fig. 13 shows the results of the data training in this research.

Fig. 13. Graph of Train Data Result from Split Data 70% Train, 25% Validation, and 5% Test with 0.0002 Learning Rate.

Fig. 13 is the result of data training from split data 70% train, 25% validation, and 5% test using a learning rate of 0.0002 which gets the highest mAP results in this research. The graph obtained does not stagnant so that the accuracy of mAP continues to run until the 6000 iteration. The mAP value obtained is 79.58% based on the results of the training data that has been done.

D. Test Results on the YOLOv4 Model

The test on this model is done by using a clip from one of the videos from the VISEM dataset. This is because at certain seconds the appearance of the VISEM video dataset will change. This model detects Sperm and Non-Sperm objects in the form of video so that the results of object checking by the model will be displayed based on the frame detected from the video being tested. The following detection results can be seen in Table IV and Table V.

TABLE IV. THE RESULTS OF THE EVALUATION TEST OBJECT ON THE YOLO MODEL

No	Frame Id	Model Prediction		TP	FP	FN
		Sperm	Non Sperm			
1	1	25	8	31	2	0
2	14	25	9	31	3	0
3	26	18	14	30	1	1
4	38	21	13	33	1	0
5	43	23	12	34	1	0
6	55	17	10	26	1	1
7	67	18	9	26	1	0
8	79	19	9	28	0	0
9	86	19	9	27	1	0
10	97	20	9	28	1	0

Table IV shows the result of an evaluation using confusion matrix to get the number of objects that have been detected or not. Detection is collected by object, the results obtained that almost all objects can be detected accurately. The layer capture taken from the YOLO model experiment in the form of video is taken as much as 10 frames, because there are so many frames generated from 1 video.

Table V describes the results of data processing obtained in the process in Table IV. These values are used to obtain precision, recall, AP, and mAP values. The results obtained will be a benchmark for the accuracy of the model in detecting objects. We can see the results of the experiment stated that the model can detect more than 80% of the number of objects in a video frame.

Based on the description in Table IV and Table V, it can be stated that the model can detect sperm and non-sperm objects with good results. In the sperm and non-sperm sections, a bounding box has been successfully created with a video quality of 50 fps, however there are one or two objects that are not legible, this is due to the object being cut off by the video frame. Fig. 14 is a display of the detection results from the model that has been created.

TABLE V. CALCULATION RESULTS OF PRECISION, RECALL, AP, AND MAP

No	Precision	Recall	AP		mAP
			Sperm	Non Sperm	
1	0.94	1.00	0.94	0.73	0.83
2	0.91	1.00	0.99	0.82	0.90
3	0.97	0.97	0.99	0.96	0.98
4	0.97	1.00	0.98	1.00	0.99
5	0.97	1.00	0.99	1.00	0.99
6	0.96	0.96	0.99	1.00	0.99
7	0.96	1.00	0.99	1.00	0.99
8	1.00	1.00	1.00	1.00	1.00
9	0.96	1.00	0.99	1.00	0.99
10	0.97	1.00	0.99	1.00	0.99

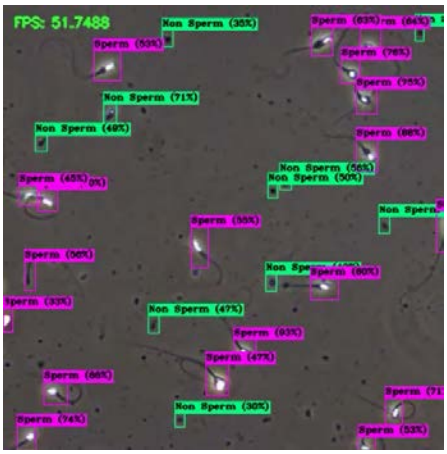

Fig. 14. Display Testing.

IV. CONCLUSION

This research develops an object detection development using the YOLOv4 algorithm to detect sperm and non-sperm objects. Using a dataset derived from the open source Simula Open Dataset, then the data in the form of videos is extracted into 1330 images. In this study, the training process was carried out with 3 different learning rate experiments, namely 0.002, 0.0002, 0.00002. In each of these experiments, 3 data divisions were made for each of the reading rates being tested. The best accuracy results are found in experiments with a learning rate of 0.0002 which has an accuracy value of 79.58% mAP on 70% train data distribution, 25% validation and 5% test. Each trial process for training uses 6000 iterations to create the training data. The test in this study uses video, the results of which are that all objects can be detected properly and have been labeled with a bounding box. In this study there were cases where the model was not able to detect optimally because the video data used contained blurred objects and sperm objects that were cut off by the frame.

ACKNOWLEDGMENT

The experiment in this research used NVIDIA Tesla K80 and Tesla K20 provided by Department of Computer Science, University of Lampung.

REFERENCES

[1] Gatimel, N., Moreau, J., Parinaud, J., & Leandri, R. (2017). Sperm morphology: assessment, pathophysiology, clinical relevance, and state of the art in 2017. ANDROLOGY, 845-862.

[2] Hinting, A., & Agustinus, A. (2021). Technology Updates in Male Infertility Management. Indonesian Andrology and Biomedical Journa, 63-67.

[3] Dhyani, I. A., Kurniawan, Y., & Negara, M. O. (2020). Hubungan Antara Faktor-Faktor Penyebab Infertilitas Terhadap Tingkat Keberhasilan IVF-ICSIi di RSIA Puri Bunda Denpasar Pada Tahun 2017. JURNAL MEDIKA UDAYANA, 2-5.

[4] S.Ningsih, Y. J., & Farich, A. (2016). Determinan Kejadian Infertilitas Pria di Kabupaten Tulang Bawang. Jurnal Kesehatan, 8-5.

[5] Iqbal, I., Mustafa, G., & Ma, J. (2020). Deep Learning-Based Morphological Classification of Human Sperm Heads. Diagnostics, 2-5.

[6] Nissen, M. S., Krause, O., Almstrup, K., Kjærulff, S., Nielsen, T. T., & Nielsen, M. (2017). Convolutional neural networks for segmentation and object detection of human semen. Cornell University, 1-6.

[7] Haugen, T. B., Andersen, J. M., Witczak, O., Hammer, H. L., Hicks, S. A., Borgli, R. J., . . . Riegler, M. A. (2019). VISEM: A Multimodal Video Dataset of Human Spermatozoa. MMSys '19 (ACM SIGMM Conference on Multimedia Systems).

[8] Hicks, t., Andersen, J., Witczak, O., Thambawita, V., Halvorsen, P., Hammer, H., . . . Riegler, M. (2019). Machine Learning-Based Analysis of Sperm Videos and Participant Data for Male Fertility Prediction. Springer Nature, 1-5.

[9] Valiuškait˙e, V., Raudonis, V., Maskeli ¯unas, R., amaševi˘cius, R., & Krilavi˘cius, T. (2020). Deep Learning Based Evaluation of Spermatozoid Motility for Artificial Insemination. Sensors, 2-8.

[10] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. IEEE Xplore, 1-9.

[11] Bochkovskiy, A., Wang, C.-Y., & Liao, H.-Y. M. (2020). YOLOv4: Optimal Speed and Accuracy of Object Detection. Cornell University, 1-17.

[12] Santoso, A., & Ariyanto, G. (2018). Implementasi Deep Learning Berbasis Keras Untuk Pengenalan Wajah. Jurnal Teknik Elektro, 18, 15.

[13] Rahim, A., Kusrini, & Luthfi, E. T. (2020). CONVOLUTIONAL NEURAL NETWORK UNTUK KALASIFIKASI PENGGUNAAN MASKER. Jurnal Teknologi Informasi dan Komunikasi, 10, 110.

[14] Rohim, A., Sari, Y. A., & Tibyani. (2019). Convolution Neural Network (CNN) Untuk Pengklasifikasian Citra Makanan Tradisional. Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, 3, 7038.

[15] Alwanda, M. R., Ramadhan, R. P., & Alamsyah, D. (2020). Implementasi Metode Convolutional Neural Network Menggunakan Arsitektur LeNet-5 untuk Pengenalan Doodle. Jurnal Algoritme , 45-56.

[16] Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Chen, T. (2017). Recent Advances in Convolutional Neural Networks. ELSEVIER, 8.

[17] Artyani, I. (2019). Simulasi Metode Convolutional Neural Network dan Long Short-Term Memory untuk Generate Image Captioning Pada Gambar Lalu Lintas Kendaraan Berbahasa Indonesia. Journal Teknik Informatika UINJKT, 27.

[18] Peryanto, A., Yudhana, A., & Umar, R. (2020). Klasifikasi Citra Menggunakan Convolutional Neural Network dan K Fold Cross Validation. Journal of Applied Informatics and Computing (JAIC), 45-51.

[19] Rahma, L., Syaputra, H., Mirza, A., & Purnamasari , S. D. (2021). Objek Deteksi Makanan Khas Palembang Menggunakan Algoritma YOLO (You Only Look Once). Jurnal Nasional Ilmu Komputer, 214-217.

[20] Arini, Wardhani, L. K., & Octaviano, D. (2020). Perbandingan Seleksi Fitur Term Frequency & Tri-Gram Character Menggunakan Algoritma Naïve Bayes Classifier (Nbc) Pada Tweet Hashtag #2019gantipresiden. KILAT, 103-114.

[21] Hartanti, D., Kusrini , & Taufiq , E. L. (2018). Penerapan Naïve Bayes Dalams Prediksi Ketercapaian Nilai Kriteria Ketuntasan Minimal Siswa. Jusikom PrimA (Junal Sistem Informasi Ilmu Komputer Prima).

[22] Kusuma, T. A., Usman, K., & Saidah, S. (2021). PEOPLE COUNTING FOR PUBLIC TRANSPORTATIONS USING YOU ONLY LOOK ONCE METHOD. Jurnal Teknik Informatika (JUTIF), 2, 60-64.

[23] Fauziah, D. A., Maududie, A., & Nuritha, I. (2018). Klasifikasi Berita Politik Menggunakan Algoritma K-nearst Neighbor. BERKALA SAINSTEK, 106-114.

[24] Fandisyah, A. F., Iriawan, N., & Winahju, W. S. (2021). Deteksi Kapal di Laut Indonesia Menggunakan YOLOv3. JURNAL SAINS DAN SENI ITS, 10, D26-D30.