Object Detection For Home Security Using YOLO v7 Model

1st Christopher Edmund Haryanto
School of Computer Science
Bina Nusantara University
Jakarta 11480, Indonesia
christopher.haryanto@binus.ac.id

2nd Christabel Dior Poluan
School of Computer Science
Bina Nusantara University
Jakarta 11480, Indonesia
christabel.poluan@binus.ac.id

3rd Bhismar Firdaus Putra
School of Computer Science
Bina Nusantara University
Jakarta 11480, Indonesia
bhismar.putra@binus.ac.id

4th Said Achmad
School of Computer Science
Bina Nusantara University
Jakarta 11480, Indonesia
said.achmad@binus.edu

5th Rhio Sutoyo
School of Computer Science
Bina Nusantara University
Jakarta 11480, Indonesia
rsutoyo@binus.edu

***Abstract*—In owning a property such as a house or residence, security is one of the most important things that we cannot ignore. This is because the level of crime such as robbery, house burglary, etc. is still relatively high in Indonesia. In overcoming this problem, several solutions have been implemented, such as installing CCTV cameras. However, in some cases CCTV cameras have not been effective in preventing these problems from occurring. Therefore, we created a system capable of detecting faces and objects, using CCTV cameras that apply the YOLO (You Only Look Once) Object Detection algorithm. Object Detection is the process of finding object instances of a certain class, such as faces, cars, etc. in images or videos. We hope that the information obtained from the detection of these objects will help us catch the perpetrators of criminal acts**

***Index Terms*—Object Detection, Home Security, YOLO**

I. Introduction

Security is the number one problem when owning a house or residence, especially in Indonesia where the crime rate is still quite high. According to the data from the Ministry of Home Affairs in the year 2020, there are 108.673 cases of theft by force and 68.768 cases of house theft [1]. The high crime rate in Indonesia will be a threat to property or residential security.

This problem could be solved by installing CCTV in their properties or homes. But there are still a few weaknesses by just using CCTV for security. One of the weaknesses of CCTV cameras is that sometimes they have low resolution, so people often have difficulties distinguishing whether the object recorded on the CCTV camera is a person, animal or something else.

To solve this problem this paper implemented an object detection system using YOLO (You Only Look Once) algorithm to detect and to recognize the object captured by the CCTV camera. By first dividing the image into N grids, YOLO aims to complete object detection in a single step. Their grids are of the same size. It is utilized to find and locate any objects that might be present in any of these areas. Bounding box coordinates, B, for any prospective objects are predicted for each grid along with their object labels and a likelihood rating for their presence. This results in a considerable overlap of expected objects from the grids' cumulative predictions. The YOLO algorithm employs Non-Maximal Suppression to suppress all the bounding boxes with significantly lower probability scores in order to handle this redundancy and narrow the predicted objects down to those of interest. To achieve this, YOLO examines the probability of scores attached to each decision and selects the one with the highest score [2] [3].

Identify applicable funding agency here. If none, delete this.

II. Related Works

A. Object Detection

In the past, video monitoring tasks were often performed by supervisors who had to manually observe the camera system. This resulted in harmful health effects for the supervisor's eyesight as they had to spend enormous amounts of time monitoring multiple surveillance cameras. Today, thanks to advances in science and technology, security surveillance systems have become smarter. Human and object detection is a major task for various surveillance applications including abnormal action recognition, vandalism deterrence, etc [4]. Object detection is performed to check the existence of objects in a video frame and to detect that object. Then detected objects can be classified in various categories such as humans, vehicles and other moving objects [5]. The compact size and quick calculation speed of the model form the basis of the YOLO target identification technique. Through the neural network, YOLO can output the position and category of the bounding box immediately. YOLO employs the global image directly for detection, allowing it to encrypt the information at the global level and minimize errors caused by mistaking the background for an item. [6]. The YOLO algorithm uses the entire image as its input and gets the bounding box and detection result right away, the YOLO algorithm breaks the original image into small, non-conforming squares that correspond to the original images. The target of those central points in the little square can then be predicted using each element [7] [8] [9]

B. Comparison of YOLO v7 with Previous Models for Home Security

YOLO v7 is the latest YOLO version with the highest performance and efficiency in terms of real-time application. The network architecture of YOLO v7 is faster and more reliable, and it provides a better feature integration method, more accurate object detection performance, a more reliable loss function, and more effective label assignment and model training. [9]. Before YOLO v7 there are other previous models of YOLO, among others are YOLO v5 and YOLO v6, Yunus EĞİ conducted an experiment to compare the YOLO v7 and the previous YOLO models. Where Yunus EĞİ uses Precision, Recall, mAP@.5 and mAP@.95 values as parameters for each model. And based on these experimental results it can be concluded that the YOLO v7 model has the best performance among the previous YOLO models with higher values for each parameter than the previous YOLO model [15].

Paper	Objectives	Method	Results	Conclusions
[1]	Provide statistics for crime rates of force theft and house theft in Indonesia	Survey	From the survey of year 2020, there are about 108.673 cases of forced theft and 68.768 cases of house theft	The article may explore the implications of the declining crime rates for government policies and law enforcement strategies. It could discuss the need for continued efforts to sustain this positive trend, address any remaining challenges, or allocate resources effectively to further improve public safety.
[2]	Provide explanation of how using YOLO algorithm is effective whilst also having high accuracy	Literature Study and Comparison	YOLO algorithm processes images in real time, which makes it suitable for applications where speed is critical and is also relatively easy to train	YOLO algorithm follows a step-by-step process to detect objects in an image. It divides the input image into a grid then predicts bounding boxes with the highest class probabilities.
[3]	Provide explanation of how YOLO algorithm works	Literature Study	YOLOv3 is pretty good. In terms of COCOs weird average mean AP metric it is on par with the SSD variants but is 3 x faster. It is still quite a bit behind other models like RetinaNet in this metric though	YOLOv3 is a good detector. It's fast, it's accurate. It's not as great on the COCO average AP between .5 and .95 IOU metric. But it's very good on the old detection metric of .5 IOU
[4]	Provide a practical solution for real-time human detection in smart video surveillance systems at the edge	Experiment	The proposed model can detect humans with high accuracy of 95.05 % and 96.81 % while testing on the INRIA and PENNFUDAN datasets	The experimental results show that the network model employing Residual blocks and multiple SPP modules can fit the human detection task better than the model which doesn't use them.
[5]	Provide a comprehensive review of various methods for object detection and tracking in computer vision	Literature study and Comparison	All the methods described for object detection and tracking have some pros and cons,	One method cannot be used for all types of photographs, nor can all methods be effective for all types of images. The background subtraction method produces inaccurate results when trying to distinguish objects with noise. By integrating various approaches and using them in concert, it is possible to tackle the aforementioned issues and add some accuracy and richness, depending on the application.
[6]	Provide a review of YOLO algorithm developments	Literature Study	Weekly trends for YOLO versions in the past five years Top ten queries for each TOLO version	The YOLO version has a lot of differences. However, they still have some features in common. Hence, they are still similar The YOLO versions are still quite young and offer lots of potential for further study. For scenario implementations in particular.
[7]	Provide Evaluation of YOLO Object Detection Methods with Precision, Recall paper	Literature Study and Comparison	The higher the version of the YOLO algorithm, the better the performance and accuracy	Higher versions of the YOLO algorithm will outperform older versions, the performance of YOLO can also be improvised by increasing the number of training epochs and by using larger batch size.
[8]	Provide comparison on YOLO v5,v6, and v7	Comparison	The performance n on the MS COCO datasets resulted in YOLOv5 mAP@0.5 scoring 42 mAP@0.95 scoring 20.1, YOLOv6 mAP@0.5 scoring 43 mAP@0.95 scoring 21.3, whilst YOLOv7 mAP@0.5 scoring 44.2 mAP@0.95 scoring 22.5	YOLOv7 uses a new backbone network (CSPDarknet53), object detection head (CenterNet), and loss function (GloU) hence outperforms YOLOv5 and YOLOv6
[9]	The paper describes the methodology used to train and optimize the YOLOv7 and Deep SORT models, as well as implementation details of the proposed system	Propose a deep learning model for tracking vehicles and detecting their movements in real time, combining the YOLOv7 and Deep SORT models to create an efficient and accurate system for tracking and detecting vehicles in complex environments, such as traffic intersections and highways.	Compared to YOLOv5-DeepSORT, YOLOv7-DeepSort has a greater tracking accuracy	Compared to YOLOv5- DeepSORT, YOLOv7-DeepSort has greater tracking accuracy, according to experiments. The experimental findings on a difficult data set evaluated at various camera heights, Rear View of the Vehicles, and angles showed flexibility and The suggested method's accuracy is good. The classification of automobiles is one upcoming project that will increase the system's dependability. It is beneficial to include a classification procedure because it further enhances the performance of detection.

Paper	Objectives	Method	Results	Conclusions
[10]	rovide comprehensive review of the state-of-the-art techniques for object detection using deep learning methods. Specifically YOLO (You Only Look Once) and CNN (Convolutional Neural Network)	Review the literature on the models, highlighting their strengths and weaknesses, and discussing various approaches to improve their performance	YOLO provides speedier yet high accuracy object detection whilst CNN provides better accuracy but is slower and computationally intensive to train.	YOLOv7 is a newer object detection algorithm that is based on YOLOv5 and is more effective than CNN
[11]	Talk about how Faster-YOLO is faster than YOLOv3 on average and has a 1.1 percent detection accuracy advantage over the original YOLOv2.	Experiment, Comparison	Faster-YOLO have a marginally greater detection accuracy than YOLOv2, 1.1% higher than Faster SSD, 1.2% higher than Faster R-CNN, 1.5% higher than Faster R-CNN, and 0.3% lower than YOLOv3; The fastest detection speed is Faster-YOLO, which is 34 and 22 frames per second faster than YOLOv2 and YOLOv3, respectively. This is due to the Faster-YOLO algorithm framework's network topology, which is much simpler than the DarkNet structure and requires less rounds.	The testing findings demonstrate that Faster-YOLO offers a balance between speed and accuracy and is quicker than previous detection methods. Faster-YOLO is well suited for real-time applications that require quick, effective object recognition because of its strong generalization performance. The Faster R-CNN algorithm framework generally has great accuracy but falls short of real-time needs.
[12]	Discuss about how Yolov7-DeepSORT has a better performance in tracking accuracy than the previous versions in Multiple Object Tracking (MOT)	Experiment, Comparison	The tracking accuracy of YOLOv7-DeepSORT (hereinafter referred to as YOLOv7) is higher than that of YOLOv5-DeepSORT (hereinafter referred to as YOLOv5s/m/l) in MOTA, MOTP, and IDF1. YOLOv7 is superior to YOLOv5l in terms of ID switch. YOLOv7 has a good tracking impact on the target and performs somewhat worse for ML and MT than YOLOv5l, better than YOLOv5s, and worse than YOLOv5m.	Compared to YOLOv5-DeepSORT, YOLOv7-DeepSORT considerably increased tracking accuracy.
[13]	Discuss about how YOLOv7 is the most advance model for real-time object detectors with its leading speed and accuracy of 5-160 fps with a 56.8% accuracy	Experiment, Comparison	Yolov7 achieved an average precision(AP) of 55.9% on the COCO test with 56 FPS on a v100 GPU whilst SWIN-L Cascade-Mask R-CNN achieved 53.9% AP with 56 FPS with the same specification	YOLOv7 has improved with new features that includes new backbone network, head network, and training method thus outperforming most algorithms to date.
[14]	Discuss about using object detection for home security (smart alarm system) with YOLO algorithm	Experiment	The system captures images of the environment then the computer processes the image and identifies objects in the image, if the computer considers an object as a threat it will send an alarm and notify the user	Using the smart alarm system may improve security, reduce false alarms, increase convenience, while also being affordable
[15]	Propose a system that uses deep learning and computer vision techniques to detect people who are not wearing masks in crowded public area	Experiment	The YOLO V7 demonstrates the best performance compared to the previous YOLO models	based on these experimental results it can be concluded that the YOLO V7 model has the best performance among the previous YOLO models with higher values for each parameter than the previous YOLO model

TABLE I: Table of Comparisson

III. SYSTEM OVERVIEW

A. *Yolov7 Custom Dataset Training*

In this experiment the researcher used the dataset from roboflow to train the YOLO v7 model. The dataset consist of 899 images of a variety of knives with different angles. Where 672 images from the dataset were used as the training set and 148 for the validity set and 79 for the test set. Some of the images used in this experiment are shown in Fig. 1.

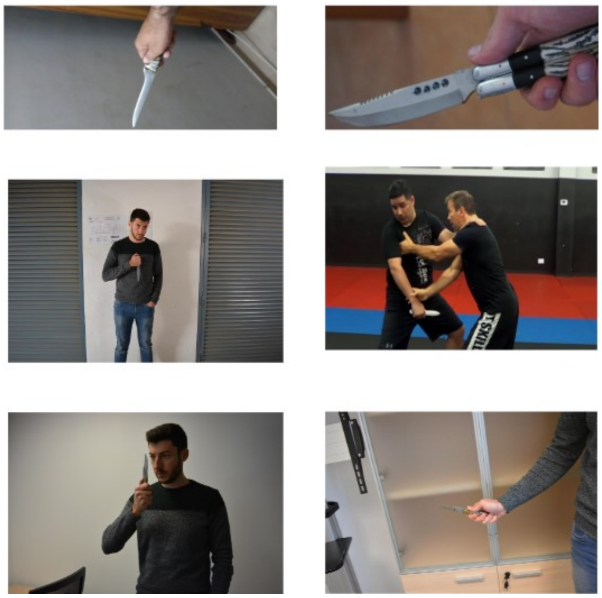

Fig. 1: variety of images with a knife

The images used in the experiment vary from one another, some images show a different knife, some show different angles of the knives, and some are a little unclear. Labeled data, i.e. the class-label and location (coordinates) of all ground truth bounding boxes in training photos, are necessary for YOLO detection model training [16]. All of the images used in this experiment had already been labeled in roboflow and all the labels and images formats are compatible with YOLO detection model training. The dataset construction and YOLO v7 model training can be seen in Fig. 2.

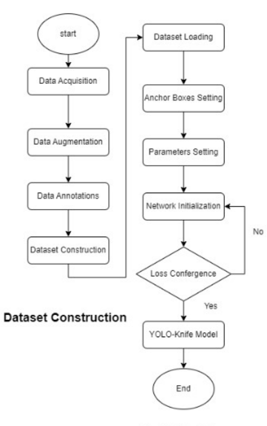

Fig. 2: Model Training

1) Data Acquisition: This step involves gathering the necessary data to train the YOLO v7 model, it includes collecting a set of images or videos that represent the objects to be detected.

2) Data Augmentation: This step is the process of applying various transformations to the acquired data to increase its diversity and improve model performance. Common augmentation techniques include rotation, scaling, flipping, and adding noise to the images.

3) Data Annotations: This step involves manually labeling the acquired data by marking the objects within the images or videos. These annotations provide ground truth information for the model during training, specifying the location and class labels of the objects.

4) Data Constructions: This step involves organizing the annotated data into a structured dataset format that can be fed into the training pipeline. This includes creating image or video datasets with corresponding annotations, ensuring proper file organization and data splitting for training, validation, and testing.

5) Dataset Loading: In this step, the constructed dataset is loaded into the training pipeline, allowing the model to access and process the training samples during the training iterations.

6) Anchor Box Setting: This step involves determining the appropriate anchor box configurations based on the dataset and the target objects' characteristics. In YOLO v7, anchor boxes, which are predefined bounding boxes with various sizes and aspect ratios, are used to forecast item placements and sizes.

7) Parameters Setting: The YOLO v7 model has various parameters that need to be set before training. These include network architecture configurations, input image dimensions, learning rate, batch size, and other hyperparameters that affect the training process and model performance.

8) Network Initialization: This step involves initializing the neural network model architecture, including the layers, connections, and weight parameters. Proper network initialization is crucial for enabling the model to learn effectively during the subsequent training iterations

9) Loss Convergence: Loss convergence refers to the process of iteratively adjusting the model's parameters to minimize the loss and improve the model's performance on the training data.

10) YOLO-Knife Model: The YOLO-Knife model is a specific variant or modification of the YOLO v7 model that incorporates additional enhancements or improvements. It could refer to architectural changes, modified loss functions, or specialized techniques tailored for specific applications or performance enhancements.

B. Yolov7 Custom Dataset Test

After training the Yolov7 knife model this paper also tested the model to test the capability of the model. Fig. 2. illustrates the research method that this paper uses to conduct the experiments. In this experiment this paper uses datasets that consist of images to test the capability of the trained model.

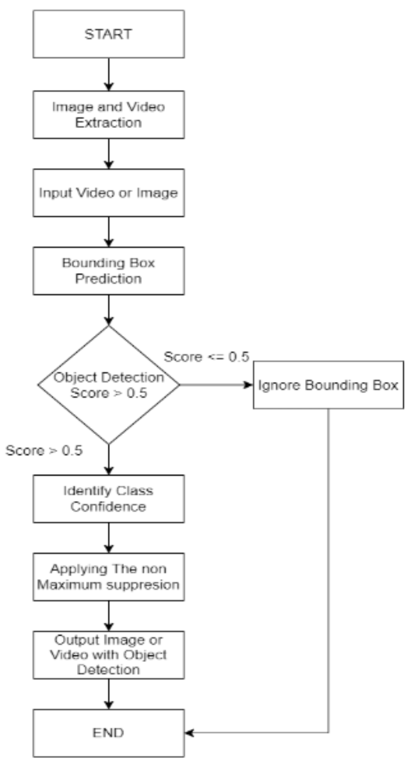

Fig. 3: Knife Detection Test Flow

In this experiment the researcher used Google Colab to conduct the experiment, where the researcher implemented the trained model using the Python programming language. The researcher also used google drive to store some of the needed files to compile the codes used in the YOLO v7 algorithm. After the codes were compiled, The researcher uploaded the images to apply object detection to. The images that are uploaded then are going to be predicted and determined whether the object detected in the images or videos have an object detection score of more than 0.5 or not. If the score is below or equal to 0.5 then the bounding box of the object will be ignored and the object detected by the system will not be considered as an object. But if the score is above 0.5 then the object will be identified and the system will predict what the object is.

IV. Result and Discussion

The trained model was tested using the image of 640 x 640 pixels set at batch size 16 in order to maintain consistency with the training image resolution. The training is initialized with an epoch of 44. The result of the training can be seen in Fig. 4. After training is completed, all classes' best precision and recall values are obtained at 0.913 and 0.99.

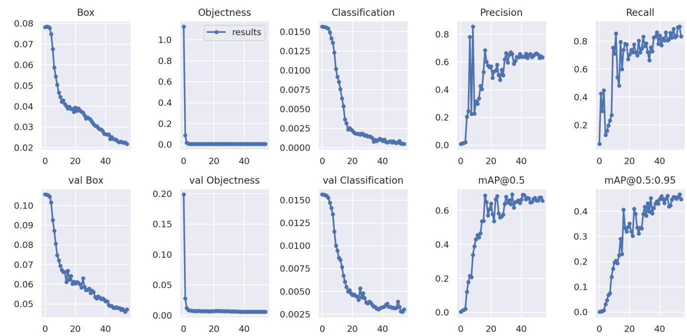

Fig. 4: Training Result

mAP@.5 (mean Average Precision at 5) is a performance metric that measures the average precision of retrieving or detecting objects across a set of queries or images, considering the top 5 results for each query or image. In this paper the highest mAP@.5 detection accuracy is obtained at 0.904 for the "Knife" class. YOLO V7 detection results are represented in Table 2.

	All	Knife	Person
mAP@.5	0.657	0.904	0.411

TABLE II: Table 2. The YOLO v7 detection results

For a thorough model evaluation, the training effectiveness of individual classes based on confidence level is also essential. The F1 score and the PR graph are also plotted. According to Fig. 6 the best precision for all classes is obtained at the confidence level of 0.913.

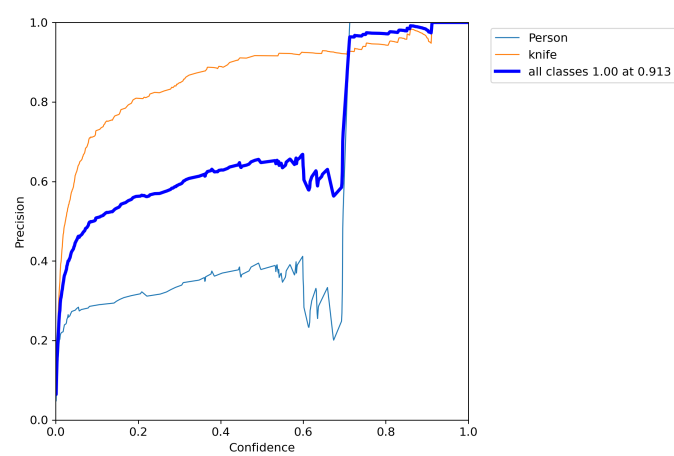

Fig. 5: Precision-Confidence level graph

Fig. 7 plots the recall values versus confidence level to show

how well the YOLO V7 model describes the data. The model has a significant level of identification of the data, as seen by the recall value of 0.99 for all classes at 0 confidence.

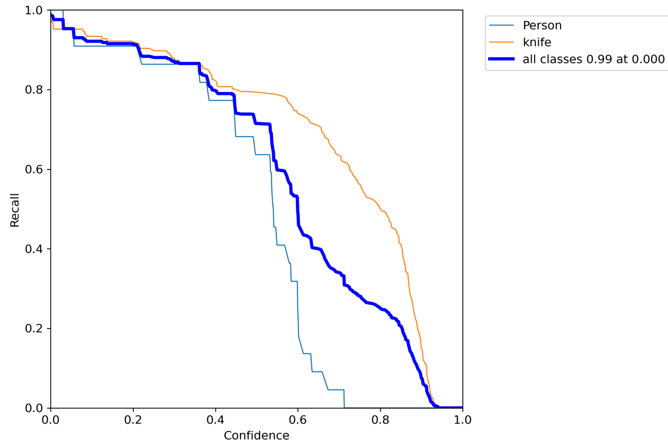

Fig. 6: Recall-Confidence level graph

The result of the experiment with an F1 score of 0.69 for all classes at a 0.378 confidence level as seen in Fig. 8, shows that there is a decent amount of scattering between classes. The F1 score for the knife class is higher, which means there are low false positives and false negatives in this class. On the contrary, for person class the F1 score is lower than the average class score, meaning there are high false positives and high false negatives.

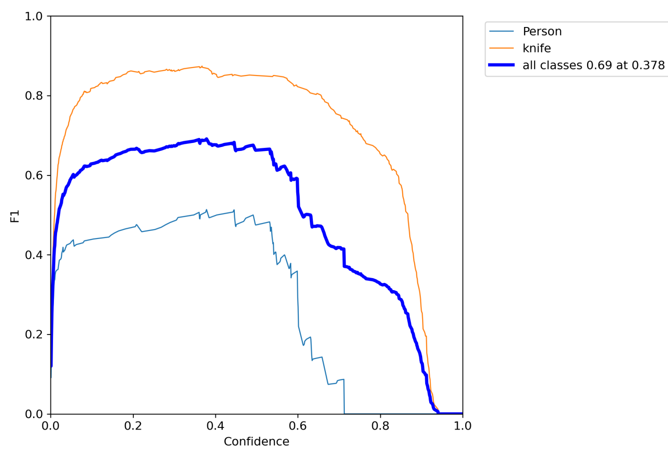

Fig. 7: F1-Confidence level graph

The PR graph is then shown to show the degree of loss between precision and recall values at various thresholds from 0 to 1. As shown in Fig. 9.

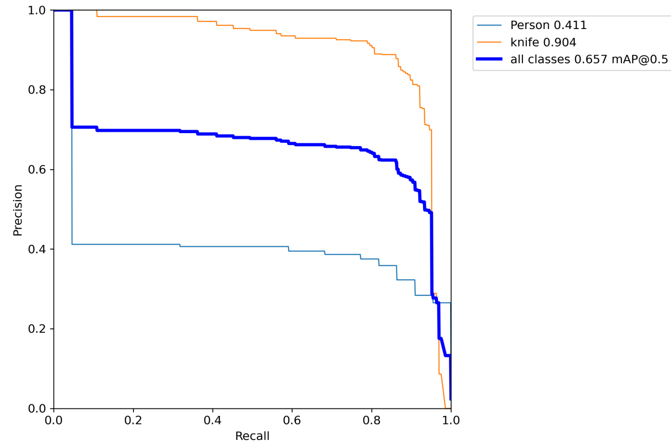

Fig. 8: precision-Confidence level graph

From the figure it is known that the minimum loss is obtained from the knife class and the class with the biggest loss is obtained from the person knife. This result can be attributed to the use of a small dataset. Some of the best detection results with the final model are represented in Fig. 10.

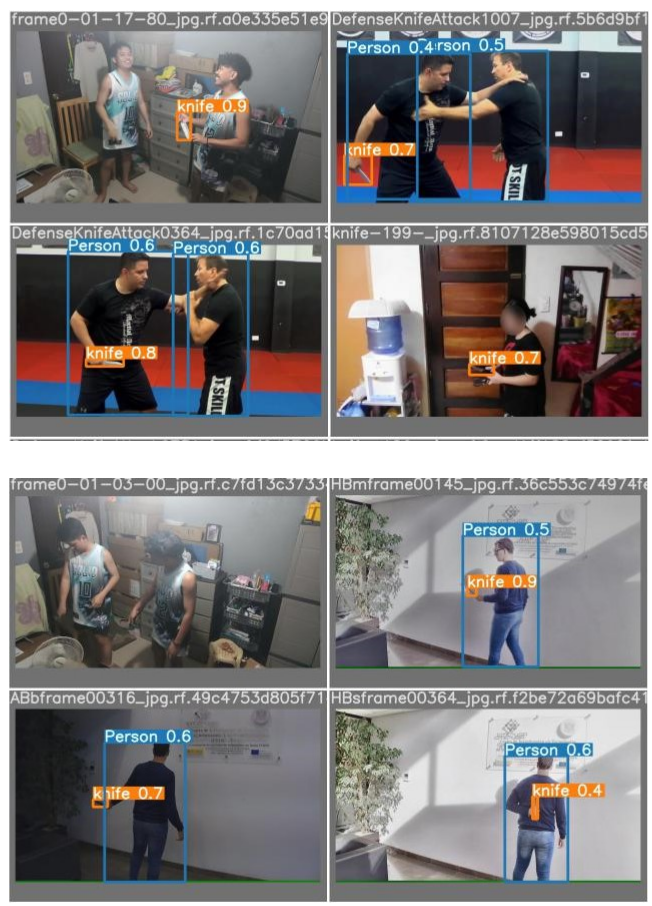

Fig. 9: Custom Training Result

V. Conclusion

In this paper, the researcher conducts an experiment using the YOLO v7 model for object detection. This paper used a dataset for knife detection, where the dataset used in this experiment was obtained from roboflow. This experiment used Google Colab to compile the python code used in this experiment. After the model training was done the results were, 0.913 for precision, 0.99 for recall, and an F1 score of 0.69. From the custom training result there are some knife detection results that have a precision score of 0.3 or lower. This result can be attributed to the use of a small dataset. The result of the knife detection model could be higher if the dataset used in the experiment is bigger. With promising results, this paper has successfully determined that the YOLOv7 is one of the best algorithms to use for object detection as of now for object detection to help with home security. The authors hope that with these results, can contribute to lessen crime rates of theft by force and house theft in Indonesia.

References

[1] Jumlah Tindak Kejahatan di Indonesia turun dalam 1 Dekade Terakhir: Databoks (no date) Pusat Data Ekonomi dan Bisnis Indonesia. Available at: https://databoks.katadata.co.id/datapublish/2022/12/13/jumlah-tindak-kejahatan-di-indonesia-turun-dalam-1-dekade-terakhir (Accessed: 18 June 2023).

[2] Skelton, J. (2022) How to train and use a custom YOLOV7 model, Paperspace Blog. Available at: https://blog.paperspace.com/yolov7/ (Accessed: 18 June 2023).

[3] Redmon, J., & Farhadi, A. (2018). YOLOv3: An Incremental Improvement. arXiv preprint arXiv:1804.02767.

[4] Nguyen, H. H., Ta, T. N., Nguyen, N. C., Bui, V. T., Pham, H. M., & Nguyen, D. M. (2021). YOLO Based Real-Time Human Detection for Smart Video Surveillance at the Edge. 2020 IEEE Eighth International Conference on Communications and Electronics (ICCE). doi:10.1109/icce48956.2021.9352144

[5] 'A review of object detection and tracking methods' (2017) International Journal of Advance Engineering and Research Development, 4(10). doi:10.21090/ijaerd.45913.

[6] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[7] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[8] Yunus Eği. (2021). Comparison of YOLO Models: YOLOv5, YOLOv6, and YOLOv7. arXiv preprint arXiv:2106.05840

[9] Yadav, V. K., Yadav, P., Dr, & Sharma, S., Dr (2022). An Efficient Yolov7 and Deep Sort are Used in a Deep Learning Model for Tracking Vehicle and Detection. Journal of Xi'an Shiyou University, Natural Science Edition, 18(11), 759-763. https://doi.org/10.3969/j.issn.1673-064X.2022.11.009

[10] V,V., K, C. R., & C, R. A. (2022). Real Time Object Detection System with YOLO and CNN Models: A Review. Journal of Xi'an University of Architecture & Technology, XIV(7), 144-151. https://arxiv.org/ftp/arxiv/papers/2208/2208.00773.pdf

[11] Yin, Y., Li, H., & Fu, W. (2020). Faster-YOLO: An accurate and faster object detection method. Digital Signal Processing, 102(102756), 1-11. https://doi.org/10.1016/j.dsp.2020.102756

[12] Yang, F., Zhang, X., & Liu, B. (2022). Video object tracking based on YOLOv7 and DeepSORT. 14(8), 1-4. https://doi.org/10.48550/arXiv.2207.12202

[13] Wang, C. Y., Bochkovskiy, A., & Liao, H. Y. M. (2022). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detector. https://doi.org/10.48550/arXiv.2207.02696

[14] Aman, Babu & Saxena, Rai & Nikhil, Madhuri & Godha, Varun & Dev, & Sabri, Mustafa. (2021). SMART ALARMING SYSTEM USING OBJECT DETECTION. SSRN Electronic Journal. 9. 3707-3713.

[15] EĞİ, Y. (2023). YOLO V7 and Computer Vision-Based Mask-Wearing Warning System for Congested Public Areas. Bilgisayar MüHendisliği, 13(1), 22-32.https://doi.org/10.21597/jist.1243233

[16] Lawal, M. O. (2021). Tomato detection based on modified YOLOv3 framework. Scientific Reports, 11(1). doi:10.1038/s41598-021-81216-5