TRANSFORMTIKA, Vol.19, No.1, July 2021, pp. 1 - 7
ISSN: 1693-3656,
journals.usm.ac.id/index.php/transformatika

YOLO Algorithm for Detecting People in Social Distancing System

Faisal Dharma Adhinata[1], Diovianto Putra Rakhmadani[2], Alon Jala Tirta Segara[3]
[1]Department of Software Engineering, Faculty of Informatics, Institut Teknologi Telkom Purwokerto, Indonesia, e-mail: faisal@ittelkom-pwt.ac.id
[2]Department of Software Engineering, Faculty of Informatics, Institut Teknologi Telkom Purwokerto, Indonesia, e-mail: diovianto@ittelkom-pwt.ac.id
[3]Department of Software Engineering, Faculty of Informatics, Institut Teknologi Telkom Purwokerto, Indonesia, e-mail: alon@ittelkom-pwt.ac.id

ARTICLE INFO

Article history:

Received 21 June 2021
Received in revised form 20 July 2021
Accepted 27 July 2021
Available online 31 July 2021

ABSTRACT

Social distancing is an effort to prevent the spread of the coronavirus. Several systems for monitoring social distancing have been developed. People detection is an essential step in implementing a social distancing system. Failure to detect people causes the social distancing system to be inaccurate. Two people who communicate cannot occur violations of social distancing because one person is not detected. Therefore, we propose a precise person detection method for the social distancing system. The proposed social distancing system uses the YOLOv3 method for people detection and Euclidean Distance for measuring the distance of social distancing. YOLOv3 can detect people's objects precisely, even people who are caught small by the camera. Experiments on two outdoor video datasets result in an F1 value of more than 0.8. This proposed system can serve as a reference for future social distancing research.

Keywords: Social distancing, YOLOv3, Euclidean Distance

1. Introduction

The coronavirus (COVID-19) is an infectious disease with very fast transmission. The coronavirus was first discovered in Wuhan, China. Until now, it has spread widely to many countries, including Indonesia. Therefore, the World Health Organization (WHO) declared the coronavirus outbreak a pandemic [1]. Recently, daily positive cases in Indonesia have reached 4,000 and even 5,000 cases per day [2]. Currently, the government is still conducting studies related to vaccines to be given to Indonesian citizens. The target of vaccines is to stay healthy and become immune to the coronavirus [3]. Besides waiting for a safe corona vaccine, various other preventive efforts are also being carried out. When leaving the house to go outside, there are restrictions on crowding. The public is urged to keep a distance from other people to avoid spreading the coronavirus from the droplets of other people that are infected with the coronavirus. Keep the distance between people to a minimum of one meter [4].

The coronavirus spreads quickly from person to person through sneezing, coughing, direct speech, and even exhaled breath. We need social distancing activities to prevent the spread of the coronavirus. Several applications for social distancing began to be developed this year. Ahamad et al. [5] conducted social distancing research using Region of Interest (ROI) segmentation. In this study, indoor and outdoor social distancing experiments were conducted. The results obtained were 100% accurate in indoor testing, but in outdoor testing on all video experiments, the accuracy was

Received June 21, 2021; Revised July 20, 2021; Accepted July 27, 2021

below 70%. This unfavorable result is due to many false negatives and false positives of people detection.

People detection methods that are often used include a combination of the Histogram of Oriented Gradient (HOG) and Support Vector Machine (SVM) [6]. CNN that deep learning method has the most significant results in image recognition. CNN tries to imitate the image recognition system in human vision so that it can process image information [7]. Recently, deep learning methods have been implemented for processing video data using the You Only Look Once (YOLO) method. YOLO method can detect people accurately, even up to 2 times the ability of other algorithms [8]. This study proposed a method to use YOLOv3 for social distancing cases based on ROI in an outdoor environment. The use of the YOLOv3 method is also faster and more accurate than other people's detection methods [9]. YOLOv3 is the latest deep learning model today and is 3x faster, operating at 22 m/s at 28.2 mAP (mean Average Precision) and fps at Yolo basic 45 frames per second [10].

This paper is organized as follows. Section 2 explains the method that is used in this study. Section 3 contains a discussion of the results and the evaluation of this system. Section 4 contains conclusions and suggestions for future work.

2. Research Method

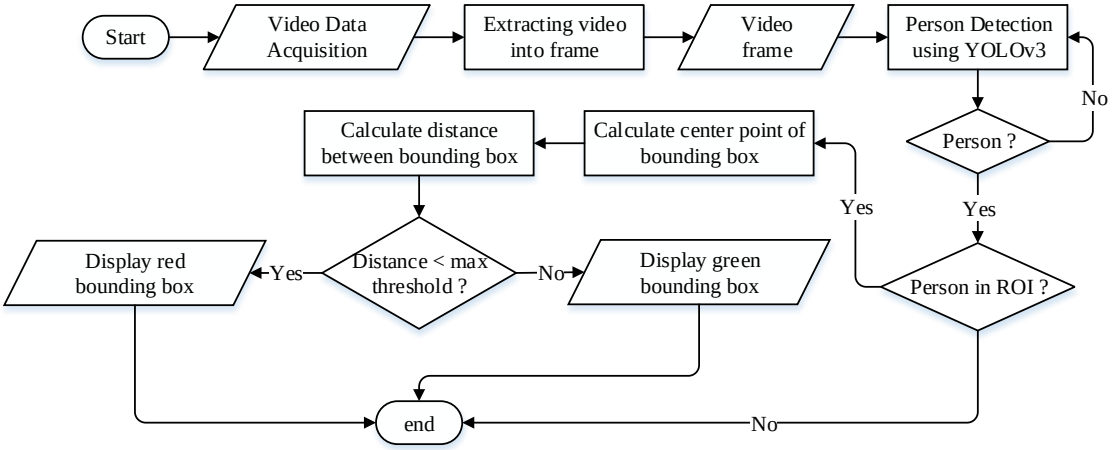


Figure 1. The architecture of the social distancing system

The social distancing system starts with inputting video data. Video data is extracted into video frames. Next, the video frame is checked for people objects or not using the YOLOv3 deep learning technique. If the video frame contains a person object, the person detected is checked whether it is included in Region of Interest (ROI). Making ROI aims to limit social distancing in videos. The movement of people who are too far away from the camera causes a difference in the threshold for social distancing. The center point of the bounding box indicates the people that are detected on the ROI. The center point of the bounding box has two values, namely the x coordinate and the y coordinate.

Furthermore, the distance between the center points of people detection is calculated using the Euclidean Distance method. If the distance value is less than the specified threshold, the system will display a red marker on the bounding box indicating a social distance violation. Conversely, if the distance value is more than the threshold, the system displays a green color on the bounding box, indicating the person's position is safe from the surrounding people. Figure 1 shows the proposed system architecture.

2.1. Video Data Acquisition

The video data used in this study use the same video as the previous study [5], namely the video dataset from PETS2009 [11] and TownCentre [12]. The video obtained is extracted into frames and will be processed for people detection. The use of this video dataset is to see the

accuracy of people's detection based on true positive, true negative, false positive, and false negative values. The process of calculating the accuracy is manually done by looking at the green or red bounding box correctly or not.

2.2. You Only Look Once (YOLO)

YOLO uses a single neural network to predict each bounding box of objects. YOLO can also predict all bounding boxes on video frames at the same time. The purpose of YOLO is to divide the input video frame into an SxS sized grid. If the center point of an object falls into a grid cell, the grid cells will detect the object. Each cell predicts a bounding box, which contains a confidence score to calculate the probability of the object in the bounding box. Each bounding box contains five predictions, namely x, y, w, h, confidence score. The coordinates (x, y) represent the upper left corner of the bounding box, while w and h estimate the width and height of the bounding box. The confidence score represents IOU (Intersection Over Union). Each grid cell will also predict 1 set of probability classes $\Pr(Class_i|Object)$. The confidence score will be calculated using equation (1).

$$\Pr(Class_i|Object) * \Pr(Object) * IOU_{truth\ pred} = \Pr(Class_i) * IOU_{truth\ pred} \quad (1)$$

The resulting score represents the two probability classes of how well the predicted box matches the object [13]. Processing of video frames for people detection is using YOLOv3. The dataset used in this study was downloaded from https://pjreddie.com/. The result of this method is the existence of a bounding box for people detection. YOLOv3 uses the architecture of Darknet-53, which has 53 convolutional layers. Figure 2 is the Darknet-53 layer in the YOLOv3 architecture.

	Type	Filters	Size	Output
	Convolutional	32	3 × 3	256 × 256
	Convolutional	64	3 × 3 / 2	128 × 128
1×	Convolutional	32	1 × 1	
	Convolutional	64	3 × 3	
	Residual			128 × 128
	Convolutional	128	3 × 3 / 2	64 × 64
2×	Convolutional	64	1 × 1	
	Convolutional	128	3 × 3	
	Residual			64 × 64
	Convolutional	256	3 × 3 / 2	32 × 32
8×	Convolutional	128	1 × 1	
	Convolutional	256	3 × 3	
	Residual			32 × 32
	Convolutional	512	3 × 3 / 2	16 × 16
8×	Convolutional	256	1 × 1	
	Convolutional	512	3 × 3	
	Residual			16 × 16
	Convolutional	1024	3 × 3 / 2	8 × 8
4×	Convolutional	512	1 × 1	
	Convolutional	1024	3 × 3	
	Residual			8 × 8
	Avgpool		Global	
	Connected		1000	
	Softmax			

Figure 2. Arsitektur YOLOv3 [14]

2.3. Euclidean Distance

The bounding box detected by the human object is used to calculate the distance to other people's bounding box. The center point of coordinate values in the bounding box is used for social

distancing calculations using the Euclidean Distance method. Equation (2) shows the Euclidean Distance equation.

$$d(C_1, C_2) = \sqrt{(x_{max} - x_{min})^2 + (y_{max} - y_{min})^2} \quad (2)$$

In this study, if the distance between the center point of the bounding box is less than 50 pixels, the two-person objects violate the rules of social distancing.

2.4. System Effectiveness Testing

Testing of a system aims to test the capability of the system according to predetermined research objectives. Junker et al. [15] stated that the system's effectiveness usually uses Information Retrieval (IR) standards, often called recall and precision. The classification table is called the confusion matrix, as shown in Table 1.

Table 1. An indication of the performance of the classification results with confusion matrix

Total population		***Predicted condition***	
		Prediction positive	***Prediction negative***
True condition	***Condition positive***	*True positive (TP)*	*False Negative (FN)*
	Condition negative	*False Positive (FP)*	*True Negative (TN)*

The recall is the system's ability to recall the relevant objects. Precision is the ratio of the number of relevant objects found to the total number by the system. F_1 score is one of the evaluation calculations in information retrieval that combines recall and precision. Equation 3 shows the formulas for recall, precision, and F_1.

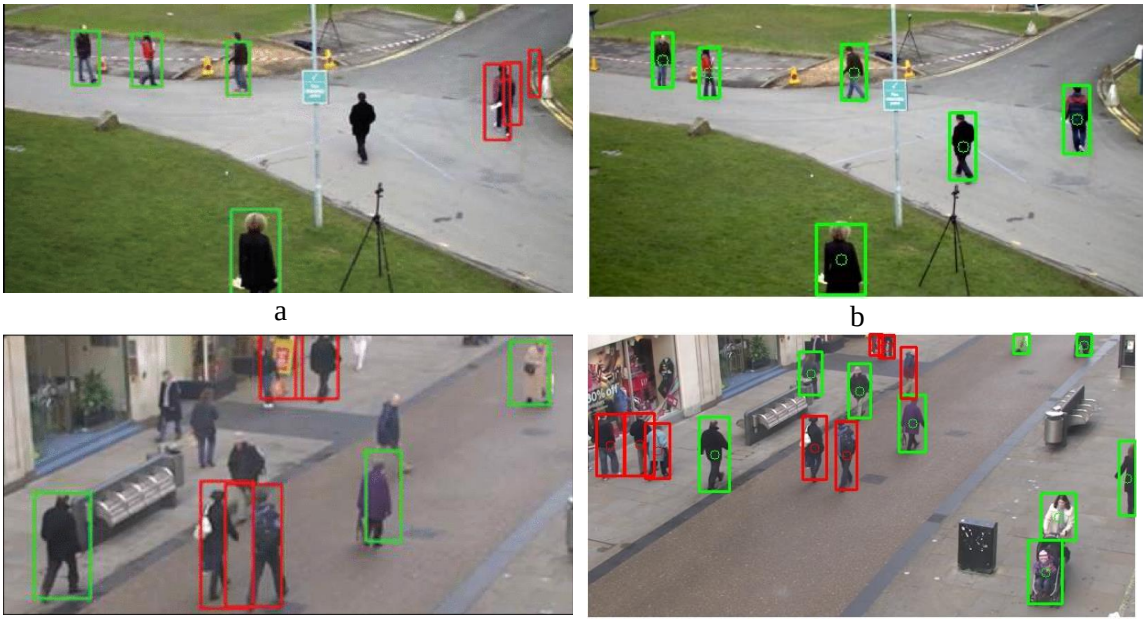
$$\begin{aligned} Recall &= \frac{TP}{TP + FN} \\ Precision &= \frac{TP}{TP + FP} \\ F_1 &= 2 * \frac{precision * recall}{precision + recall} \end{aligned} \quad (3)$$

3. Results and Analysis

A social distancing experiment uses a video dataset as a standard for processing video data. We use two video datasets, namely PETS2009 and TownCentre. Experiments were carried out by examining each video frame that consists of 500 frames on each video. Two experiments will be discussed in this study, namely the person detection experiment and the social distancing experiment. The analysis of experimental results that we propose is carried out by comparing the results of previous research.

3.1. People Detection Experiment

The experiment of person detection for outdoor social distancing was carried out by comparing the same frames with previous studies [5]. Figure 3 shows the experimental results of previous research with our proposed research using YOLOv3 for people detection. In the PETS2009 video dataset experiment, it appears that the person in the middle is not detected as a person. It is a false negative. Then there is one human object detected by three bounding boxes. Even the detection result violates social distancing because it is marked with a red bounding box. When using YOLOv3 for people detection, the detection of people is precisely one bounding box, and every human object is detected.



a b c d

Figure 3. Result of people detection a,c) Previous research [5], b,d) Proposed research

Then in the experiment with the TownCentre video dataset, the previous studies also appeared that some people were not detected. Then in the experiment using YOLOv3, everyone was detected. However, for people detection using YOLOv3 there is a drawback where occlusion human object cannot be detected. People who are far from the camera can still be detected. Figure 4 shows a human occlusion object that cannot be detected.

Figure 4. An occlusion of human object

3.2. Social Distancing Experiment

In the social distancing experiment, the PETS2009 video dataset uses a threshold of 30 pixels to calibrate the 1-meter distance on the video. Then the TownCentre video dataset uses a threshold distance of 50 pixels to calibrate the 1-meter distance on the video. This threshold difference is due to the different camera height installations. This experiment measures true positive, false positive, and false negative values. True positive is the number of social distances that occur correctly in the video frame. False positives are the number of social distances that should not have happened, but there was social distancing in the video frame. Negative false is the wrong number of social distances because it should happen, but the reality in the video frame doesn't happen. Table 2 shows the results of the social distancing experiment.

Table 2. The result of social distancing experiment

Video	True Positive	False Positive	False Negative	F_1
PETS2009	448	32	76	0.89
TownCentre	1672	605	147	0.81

Table 2 shows that the F_1 value is more than 0.8. This result is higher than previous studies [5] for social distancing in outdoor environments. These results are influenced by the results of people detection. The use of ROI also affects the results of social distancing. Objects that are far from the camera causes difficulties in calibrating the social distancing. However, the drawback of our research is that cannot detect human occlusion object as shown in Figure 4. The human object that is blocked can also occur violations of social distancing. For example, two people are communicating, but the first person blocked the second person who causes the camera only one record as a human object.

4. Conclusion

Social distancing is an essential action in preventing the spread of the coronavirus. In public places, prevention has been carried out by placing officers to supervise people to carry out social distancing. Supervision of this officer also limited visibility. Therefore, making an intelligent system for monitoring social distancing violations is made. The important step in making a social distancing system is people detection. The system's accuracy in people detection is a successful measure of the social distancing system. The social distancing system is made using YOLOv3 for people detection with distance measurement using Euclidean Distance. The social distancing system produces resulting an F_1 value more than 0.8. In the experiment using the PETS2009 dataset, the F_1 value was 0.89, while in the TownCentre dataset, the F_1 value was 0.81. However, this study has limitations in people detection. The system cannot detect people objects that are blocked by other objects. For further research, the person detection method can be modified for the case of a human occlusion object.

References

[1] M. Pawar, "The Global Impact of and Responses to the COVID-19 Pandemic," *The International Journal of Community and Social Development*, vol. 2, no. 2, pp. 111–120, 2020, doi: 10.1177/2516602620938542.

[2] WHO Indonesia, "Coronavirus Disease Situation Report World Health Organization," 2020.

[3] I. N. Kandun, "Vaksin Melindungi Masa Depan Generasi Muda Indonesia - Berita Terkini | Satgas Penanganan COVID-19," *covid19.go.id*, 2020. https://covid19.go.id/p/berita/vaksin-melindungi-masa-depan-generasi-muda-indonesia (accessed Nov. 18, 2020).

[4] WHO, "Overview of public health and social measures in the context of COVID-19," in *World Health Organization 2020.*, no. May, 2020, pp. 1–8.

[5] A. H. Ahamad, N. Zaini, and M. F. A. Latip, "Person Detection for Social Distancing and Safety Violation Alert based on Segmented ROI," *Proceedings - 10th IEEE International Conference on Control System, Computing and Engineering, ICCSCE 2020*, no. August, pp. 113–118, 2020, doi: 10.1109/ICCSCE50387.2020.9204934.

[6] F. D. Adhinata, M. Ikhsan, and W. Wahyono, "People counter on CCTV video using histogram of oriented gradient and Kalman filter methods," *Jurnal Teknologi dan Sistem Komputer*, vol. 8, no. 3, pp. 222–227, 2020, doi: 10.14710/jtsiskom.2020.13660.

[7] B. P. G. Pamungkas, B. Nugroho, and F. Anggraeny, "Deteksi Dan Menghitung Manusia Menggunakan YOLO-CNN," *Jurnal Informatika dan Sistem Informasi (JIFoSI)*, vol. 02, no. 1, pp. 67–76, 2021.

[8] M. S. Chauhan, A. Singh, M. Khemka, A. Prateek, and R. Sen, "Embedded CNN based vehicle classification and counting in non-laned road traffic," *arXiv*, 2019.

[9] H. Belhassen, V. Fresse, and E. B. Bourennane, "Comparative Study of Face and Person Detection algorithms: Case Study of tramway in Lyon," *Proceedings of International Conference on Advanced Systems and Emergent Technologies, IC_ASET 2019*, no. August, pp. 154–159, 2019, doi: 10.1109/ASET.2019.8871003.

[10] H. W. B. N, E. Mailoa, and H. D. Purnomo, "Deteksi Buah untuk Klasifikasi Berdasarkan Jenis dengan Algoritma CNN Berbasis YOLOv3," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 4, no. 3, pp. 476–481, 2020.

[11] J. Ferryman and A. Shahrokni, "PETS2009: Dataset and challenge," *Proceedings of the 12th IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, PETS-Winter 2009*, pp. 0–5, 2009, doi: 10.1109/PETS-WINTER.2009.5399556.

[12] B. Benfold and I. Reid, "Stable multi-target tracking in real-time surveillance video," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 3457–3464, 2011, doi: 10.1109/CVPR.2011.5995667.

[13] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2016-Decem, pp. 779–788, 2016, doi: 10.1109/CVPR.2016.91.

[14] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," *Tech report*, pp. 1–6, 2018, [Online]. Available: https://pjreddie.com/media/files/papers/YOLOv3.pdf.

[15] M. Junker, R. Hoch, and A. Dengel, "On the evaluation of document analysis components by recall, precision, and accuracy," *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*, pp. 717–720, 1999, doi: 10.1109/ICDAR.1999.791887.