Research Article

Multipose to detect Airport Visitor Behavior

N.K.E. Iswarawati[1], A. S. Satyawan[2*], H. Puspita[3], P.A. Utomo[4], R.A. Putri[5]

[1,2,3,4,5] Nurtanio University Bandung, Bandung City, West Java, Indonesia
[2] Research Center for Telecommunication, National Research and Innovation Agency (BRIN), Indonesia
[2] Jambi University, Muaro Jambi Regency, Jambi, Indonesia
[2] Telkom University, Bandung, Indonesia

* Corresponding Author: a.s.satyawan@unnur.ac.id and arie021@brin.go.id

Abstract: The airport is a strategic place where it involves important activities, namely airplane flights. Airport activities are greatly influenced by the ongoing security within the airport. The thing that affects security in the airport is the possibility of crimes committed by unexpected visitors. The purpose of this research is to observe the airport area, especially airport visitors so that if there are visitors who have the potential to commit crimes, they can be detected properly and further investigation procedures can be carried out. To be able to observe airport visitors and recognize patterns of visitor behavior that have the potential to commit crimes, an airport visitor gesture recognition system can be used. In this thesis, the gesture recognition of airport visitors is done with the multipose estimation method. This method can detect 17 key points on the human body that are used to detect the behavior of airport visitors who have the potential to commit crimes. To develop this system, deep learning algorithms that are currently developing with the help of TensorFlow and architectural models in multipose estimation, namely MobileNetV2, Feature Pyramid Network, and CenterNet, can be used. The experimental results show that the multipose estimation method can recognize human gestures well under several conditions such as the appropriate distance of the human object from the camera and the lighting conditions around the observed human object. It is also seen that from several scenarios, the crime gesture model can be recognized well.

Keywords: multipose estimation; object detection; tensor flow; deep learning; recoqnition gesture; smart detection

Citation: N.K.E. Iswarawati, A. S. Satyawan, H. Puspita, P.A. Utomo, R.A. Putri, "Multipose to detect Airport Visitor Behavior", *Iota*, **2024**, ISSN 2774-4353, Vol.04, 02. https://doi.org/10.31763/iota.v4i2.723

Academic Editor : Adi, P.D.P

Received : March, 15 2024

Accepted : April, 25 2024

Published : May, 1 2024

1. Introduction

Technological sophistication in the present has developed a lot for services in all fields. Various components are developing in terms of efficiency, effectiveness, and function. Not infrequently the technology created is useful in making it easier for humans to do and do complicated things, even nowadays technological developments can match the basic abilities of humans, such as detecting the behavior of airport visitors.

This aims to identify any behavior that leads to crime. Detecting the behavior of airport visitors who commit crimes will be difficult if monitored only with the human eye. Therefore, a system that can be seen as a whole is needed, one of which is the multi-pose estimation method that works with the help of cameras and computers. To develop this system, deep learning algorithms can be used. Even the implementation of deep learning models can be developed with the help of TensorFlow. In this thesis research, an airport visitor behavior detection system will be developed using the concept of deep learning.

TensorFlow itself is a model of deep learning due to the depth of its network. Deep learning is a branch of machine learning that can teach computers to do human-like work. In general, multipose estimation is used to detect a person's position during exercise by minimizing the possibility of injury. In addition, multipose estimation detection can help a person calculate the number of repetitions that have been performed.

Various methods for human behavior pose detection have been used by many researchers. Some of them are using the Tensorflow Movenet Thunder model method which is focused on pose estimation for yoga sports (Satyam Goyal & Animesh Jain, 2021), DeepPose, Higher Residual Network and OpenPose to determine the evaluation results of each keypoints compared to each architecture (Anant Grover, Deepak Arora & Anuj Grover, 2023), Convolutional Neural Network (CNN) for activity monitoring that can be classified by the designed application is learning, standing, and sleeping in children aged 4 to 6 years (Andrean Lay & Lina, 2022), Human Pose Estimation AI Fitness Counter model used to perform fitness sports activities such as pull ups, push ups, and lifting weights (Irfan, Muchlis Abd. Murhalib, Kartika & Selamat Meliala, 2023), VGG-16 CNN for sign language detection (Amit Moryossef, Ioannis Tsochantaridis, Roee Aharoni, Sarah Ebling & Srini Narayanan, 2020), Convolutional Neural Network (CNN), VGG-16 for facial emotion identification (Lina, Arthur Adhiya Marunduh, Wasino & Daniel Ajienegoro, 2022). The main difference between this research compared to other studies lies in the application of the architecture used there are three namely MobileNetV2, Feature Pyramid Network, and CenterNet for the detection of airport visitor behavior. The dataset used in the experiment is a video recording collected by the author with variations in the number of people involved and variations in the poses of the participants.

This paper discusses the development of human behavior pose detection by demonstrating a scenario that is assumed to be the act of stealing a wallet in a pocket using a hand. Some of the limitations applied are that the model used for object detection is MoveNent multi-pose, this method uses the Python programming language, the object detection system uses video recorded using a camera, and the recorded video is taken using a capture angle such as a surveillance camera. This paper is organized with six structures, namely chapter 1 contains an introduction, chapter 2 contains a theoretical basis, chapter 3 contains research methodology, chapter 4 contains design and realization, chapter 5 contains testing and analysis, and chapter 6 contains conclusions and suggestions. Moreover, research using MobileNetV2 has been investigated in detail in references [1-16], and has been applied to various fields such as health, culinary, and defense and security.

2. Method

2.1 System Design

The airport visitor gesture recognition system is a system that can detect human behavior poses from a video frame offline. Input from the system is in the form of video recordings containing patterns of human behavior that have the potential to commit crimes, then the system automatically detects human body poses containing 17 key points in each joint. If the key points do not match each human body joint, the confidence score and aspect ratio are adjusted to the video. The process of detecting the pose of human behavior contained in the video uses a model that utilizes the best aspects of advanced architecture, namely MoveNet multi-pose.

MoveNet multipose uses the pose detection API in TensorFlow.js which is fast and accurate for pose detection. The MoveNet Multipose architecture consists of two components, a feature extractor and a set of prediction heads. The feature extractor used is MobileNetV2 which generates an image feature representation and Feature Pyramid Network (FPN) which is used to generate feature pyramids with different resolution levels from the MobileNetV2 representation. The flowchart of the system designed for behavioral pose detection from video can be seen in Figure 1.

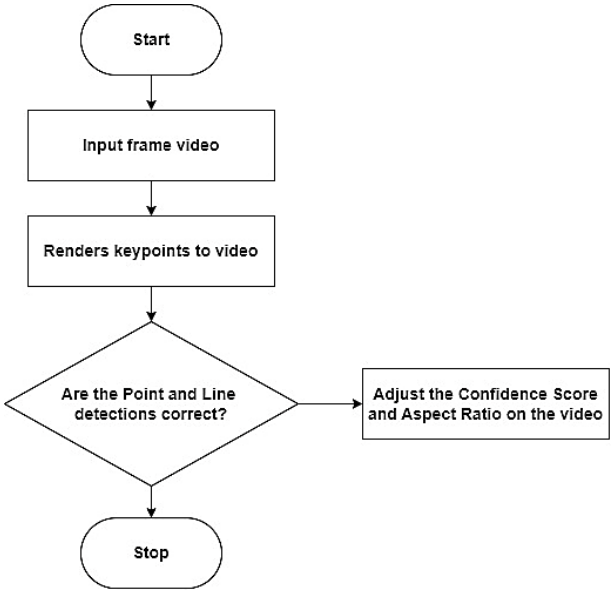

Figure 1. System Design Schematic

2.2 Human Behavior Pose Detection Identification Stage

The stage of identifying human behavior poses that have been successfully detected using feature extractors in the multipose MoveNet model has three stages, namely MobileNetV2, Feature Pyramid Network, and Center Net. The first stage in MoveNet is MobileNetV2 which is one of the architectures of the convolutional neural network (CNN), MobileNetV2 adds two new features namely linear bottlenecks and shortcut connections between bottlenecks. In using the MobileNetV2 model, the image will be inputted, then using convolution and using the ReLU activation function. ReLU is an activation function that is responsible for normalizing the values generated from the convolutional layer. The ReLU activation function equation can be seen in the following equation 1.

$$x \leq 0 \text{ then } x = 0, \text{ if } x > 0 \text{ then } x = x \quad (1)$$

After the ReLU process is complete, it is continued with the bottleneck layer. In the bottleneck residual block, there are three convolution layers used, namely performing 1x1 convolution to reduce the input dimensions, then performing 3x3 convolution to extract features and finally performing 1x1 convolution again to restore the output dimensions to the original input, the use of this bottleneck can reduce the number of parameters and multiplication in the matrix, so that it can run faster and more efficiently, especially on devices with limited resources such as mobile devices. After the process is complete, the Global Average Pooling (GAP) layer will be used to reduce the spatial dimension of the feature map globally so that it becomes one value per channel by performing an average pooling operation on the entire map feature area and producing one average value for each channel (channel) can be seen in Figure 2.

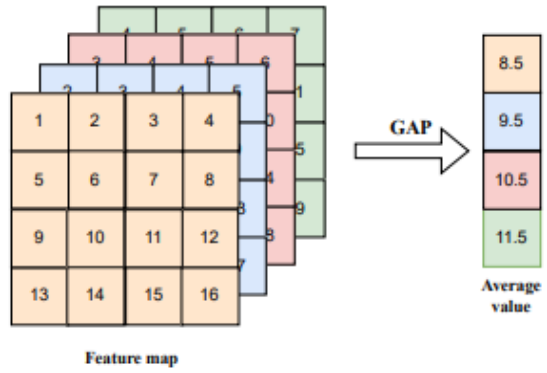

Figure 2. Global Average Point

The last process is using the softmax activation function which will be used in the output layer to calculate the probability of the output result, which occurs in the output layer where the largest probability value will be taken as a prediction. The second stage is the Feature Pyramid Network (FPN) which is used in object detection to generate multi-scale features from input images using feature pyramids to detect small objects. The Feature Pyramid Network architecture can be seen in Figure 3.

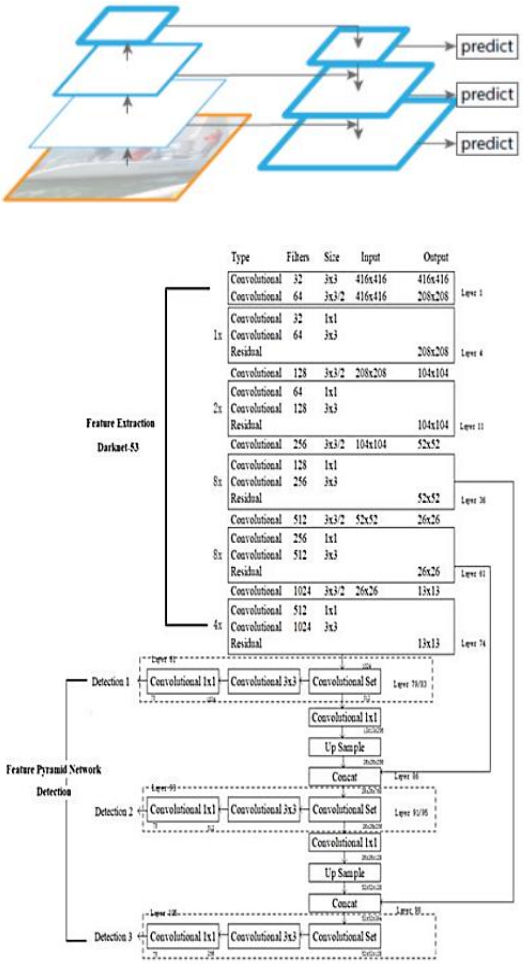

Figure 3. Architecture Feature Pyramid Network

The third stage, CenterNet, is used to predict the center point of the object using a heatmap. CenterNet can overcome several problems such as overlapping bounding boxes and the difficulty of detecting small objects. In Figure 4, it can be seen that the CenterNet architecture uses a Convolutional backbone network by applying cascade corner pooling to the two output corner heatmaps and the center keypoint heatmap, a pair of detected corners is used to detect potential bounding boxes and then the detected keypoints are used to generate the final bounding boxes.

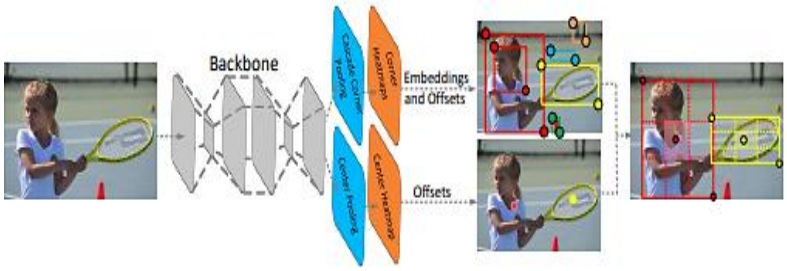

Figure 4. CenterNet Architecture

3. Result and Analysis

3.1. Testing Method

Tests were conducted on the human behavior poses created to ensure that the program can detect the behavior of airport visitors properly and correctly.

There are three tests of airport visitor behavior detection carried out using the movenet multipose model and using three cameras namely Insta 4K, Ricoh Theta S, and conventional Webcam. The first test is intended to determine the maximum and minimum limits of the model's ability to detect poses. The second test was intended to determine the maximum number of people detected in a group of people. Tests one and two were carried out in four stages, namely in an open space in a bright state, in an open space in a dim state, in a closed space in a bright state, and a closed space in a dim state. The third test is intended to detect poses in several videos whose objects demonstrate several people's behaviors with various scenarios that are assumed to be the act of stealing a wallet in a pocket using a hand.

3.2. Testing the MoveNet Multipose Model

Testing is done using a dataset that has been collected in the form of a video. In the first test, one person stands in front of the camera at a predetermined distance, to determine the maximum and minimum limits of the model's ability to detect poses using three types of cameras, namely conventional Webcam, Insta 4k, and Ricoh Theta S. Moreover, Figure 5 shows that testing a distance of 4.5m using a conventional Webcam camera with a light intensity of 3273 Lux gets pretty good detection results because the key points produced are 17 points.

Figure 5. Testing with a Conventional Webcam Camera

Furthermore, Figure 6 shows that testing a distance of 2.5m using an Insta 4k camera with a light intensity of 3273 Lux obtained quite good detection results because the key points produced were 17 points.

Figure 6. Testing with Insta 4k camera

Figure 7 shows that the 0.5m distance test using the Ricoh Theta camera with a light intensity of 3273 Lux obtained poor detection results due to distortion.

Figure 7. Testing with Ricoh Theta camera

Table 1 shows the results of the minimum to maximum distance of object detection obtained from the first test. The table describes poor and good results. The results are not good if the detection is incomplete or imperfect. In conventional webcams the field of view of the camera cannot capture the entire shape of the object, in Insta 4k cameras the field of view is too wide so that the object becomes farther than the actual distance, while the Ricoh Theta camera has a 360-degree field of view and a large enough distortion to cause objects to be detected imperfectly. Good results if the detection is complete and follows the pose or shape of the object.

Table 1. Minimum to Maximum Distance for Object Detection

Camera	Lux	Distance (meters (m))	Results
		2 m	Not Detected
		2,5 m	Not Detected
		3 m	Detected
		3,5 m	Detected
		4 m	Detected
		4,5 m	Detected
		5 m	Detected
		5,5 m	Detected
		6 m	Detected
		6,5 m	Detected
		7 m	Detected
		7,5 m	Detected
		8 m	Detected
		8,5 m	Detected
		9 m	Detected
		9,5 m	Detected
		10 m	Detected
Conventional Webcam	17380	10,5 m	Detected
		11 m	Detected
		11,5 m	Detected
		12 m	Detected
		12,5 m	Detected
		13 m	Detected
		13,5 m	Detected
		14 m	Detected
		14,5 m	Detected
		15 m	Detected
		15,5 m	Detected
		16 m	Detected
		16,5 m	Detected
		17 m	Detected
		17,5 m	Detected
		18 m	Detected

Camera	Lux	Distance (meters (m))	Results
Insta 4k	17380	18,5 m	Detected
		19 m	Not Detected
		1,5 m	Not Detected
		2 m	Detected
		2,5 m	Detected
		3 m	Detected
		3,5 m	Detected
		4 m	Detected
		4,5 m	Detected
		5 m	Detected
		5,5 m	Detected
		6 m	Detected
		6,5 m	Detected
		7 m	Detected
		7,5 m	Detected
		8 m	Detected
		8,5 m	Not Detected
		9 m	Not Detected
Ricoh Theta S	17380	0,5 m	Not Detected
		1 m	Not Detected
		1,5 m	Not Detected
		2 m	Not Detected

The second test was conducted by several people with the aim of knowing the maximum number of people who can be detected using the Insta 4k camera to get pretty good results can be seen in Figures 8 and 9.

Figure 8. Second test

Figure 9. Third test

Furthermore, the Next test was conducted to detect poses in several videos whose objects demonstrate several people's behaviors with various scenarios that are assumed to be the act of stealing a wallet in a pocket using a hand. This test was conducted using an Insta 4K camera.

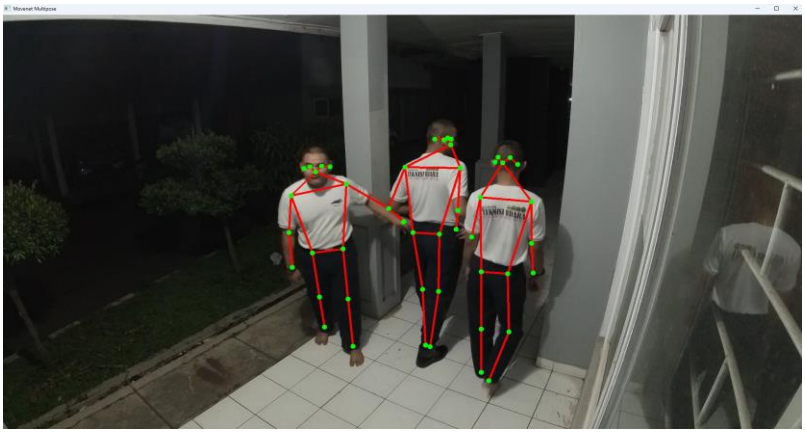

Figure 10. Fourth test

Figure 11. Fifth test

From the theft scenario, a certain object detection distance with the camera can be detected. It is shown in the third test results in Figures 10 and 11 that the skeleton of the arm to the hand when someone takes the wallet in the pocket can be detected.

4. Conclusions

After conducting the testing process, the conclusions that can be drawn from the human behavior pose detection system using the movenet multipose method are as follows [1]. The movenet multipose method can be used to identify criminal behavior that may be committed by visitors using several scenarios in the test. [2] The minimum distance detected using a conventional Webcam camera is 3m to 18.5m, the Insta 4k camera is 2m to 8m, while the Ric Ric camera is 2m to 8m.
to 8m, while the Ricoh Theta S camera is 0.5m to 2m detection is not good. In the second test, the maximum number of detection results was 6 objects. In the third test, the results of testing the behavioral poses of people with various scenarios that are assumed to be the act of stealing a wallet in a pocket using a hand can be detected.

5. Suggestion

For further research, improvements can be made which include: [1] The method used by the object detection system can be developed using more datasets and different scenarios. [2] The object detection system can be developed into real-time object detection. [3] Adding a detection distance indicator on the display to estimate the distance between objects. And [4] Application of software to airport surveillance cameras.

Acknowledgments: We would like to thank all parties, Students, Lecturers, and Researchers, as well as the team at Jambi University, Nurtanio University Bandung, and the National Research and Innovation Agency (BRIN) who helped the publication process so that it could be completed properly, hopefully, this research can be useful for public interest such as airports in particular, as well as public places, or other sectors that require the implementation of this research.

Author contributions: All authors are responsible for building Conceptualization, Methodology, analysis, investigation, data curation, writing—original draft preparation, writing—review and editing, visualization, supervision of project administration, funding acquisition, and have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. K. Dong, C. Zhou, Y. Ruan and Y. Li, "MobileNetV2 Model for Image Classification," 2020 2nd International Conference on Information Technology and Computer Application (ITCA), Guangzhou, China, 2020, pp. 476-480, doi: 10.1109/ITCA52113.2020.00106.
2. T. Adar, E. K. Delice and O. Delice, "Detection of COVID-19 From A New Dataset Using MobileNetV2 and ResNet101V2 Architectures," 2022 Medical Technologies Congress (TIPTEKNO), Antalya, Turkey, 2022, pp. 1-4, doi: 10.1109/TIPTEKNO56568.2022.9960225.
3. U. Kulkarni et al., "Facial Key points Detection using MobileNetV2 Architecture," 2023 IEEE 8th International Conference for Convergence in Technology (I2CT), Lonavla, India, 2023, pp. 1-6, doi: 10.1109/I2CT57861.2023.10126381.
4. N. A. Nainan, J. H, R. Jalan, R. S. N, K. V. C and S. P. Shankar, "Real Time Face Mask Detection Using MobileNetV2 and InceptionV3 Models," 2021 IEEE Mysore Sub Section International Conference (MysuruCon), Hassan, India, 2021, pp. 341-345, doi: 10.1109/MysuruCon52639.2021.9641675.

5. T. M. Fahrudin and I. Z. A. Illah, "SkinMate: Mobile-Based Application for Detecting Multi-Class Skin Diseases Classification Using Pre-Trained MobileNetV2 on CNN Architecture," 2023 IEEE 9th Information Technology International Seminar (ITIS), Batu Malang, Indonesia, 2023, pp. 1-6, doi: 10.1109/ITIS59651.2023.10420370.
6. D. P. P. Javierto, J. D. Z. Martin and J. F. Villaverde, "Robusta Coffee Leaf Detection based on YOLOv3- MobileNetv2 model," 2021 IEEE 13th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM), Manila, Philippines, 2021, pp. 1-6, doi: 10.1109/HNICEM54116.2021.9731899.
7. A. M. Pamadi, A. Ravishankar, P. Anu Nithya, G. Jahnavi and S. Kathavate, "Diabetic Retinopathy Detection using MobileNetV2 Architecture," 2022 International Conference on Smart Technologies and Systems for Next Generation Computing (ICSTSN), Villupuram, India, 2022, pp. 1-5, doi: 10.1109/ICSTSN53084.2022.9761289.
8. M. Xue and L. Gu, "Surgical instrument segmentation method based on improved MobileNetV2 network," 2021 6th International Symposium on Computer and Information Processing Technology (ISCIPT), Changsha, China, 2021, pp. 744-747, doi: 10.1109/ISCIPT53667.2021.00157.
9. E. R. and T. Manoranjitham, "Classification of diseases in Paddy by using Deep transfer learning MobileNetV2 model," 2022 1st International Conference on Computational Science and Technology (ICCST), CHENNAI, India, 2022, pp. 936-940, doi: 10.1109/ICCST55948.2022.10040348.
10. A. Saini, K. Guleria and S. Sharma, "A Pre-trained MobileNetV2 Model for Oral Cancer Classification," 2023 1st DMIHER International Conference on Artificial Intelligence in Education and Industry 4.0 (IDICAIEI), Wardha, India, 2023, pp. 1-6, doi: 10.1109/IDICAIEI58380.2023.10406692.
11. W. Liu, G. Zhou and J. Xu, "S-MobileNetV2+SegNet Model and Rapid Identification of Sugarcane," 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 2021, pp. 3444-3447, doi: 10.1109/IGARSS47720.2021.9553272.
12. Z. He and L. Xiong, "Image Classification of Zinc Dross Based on Improved MobileNetV2," 2021 China Automation Congress (CAC), Beijing, China, 2021, pp. 1503-1508, doi: 10.1109/CAC53003.2021.9728248.
13. Y. Zou, L. Zhao, S. Qin, M. Pan and Z. Li, "Ship target detection and identification based on SSD_MobilenetV2," 2020 IEEE 5th Information Technology and Mechatronics Engineering Conference (ITOEC), Chongqing, China, 2020, pp. 1676-1680, doi: 10.1109/ITOEC49072.2020.9141734.
14. V. Vania, A. Setyadi, I. M. D. Widyatama and F. I. Kurniadi, "Rice Varieties Classification Using Neural Network and Transfer Learning with MobileNetV2," 2023 4th International Conference on Artificial Intelligence and Data Sciences (AiDAS), IPOH, Malaysia, 2023, pp. 165-168, doi: 10.1109/AiDAS60501.2023.10284624.
15. K. Bousbai and M. Merah, "A Comparative Study of Hand Gestures Recognition Based on MobileNetV2 and ConvNet Models," 2019 6th International Conference on Image and Signal Processing and their Applications (ISPA), Mostaganem, Algeria, 2019, pp. 1-6, doi: 10.1109/ISPA48434.2019.8966918.
16. A. B. Handoko, V. C. Putra, I. Setyawan, D. Utomo, J. Lee and I. K. Timotius, "Evaluation of YOLO-X and MobileNetV2 as Face Mask Detection Algorithms," 2022 IEEE Industrial Electronics and Applications Conference (IEACon), Kuala Lumpur, Malaysia, 2022, pp. 105-110, doi: 10.1109/IEACon55029.2022.9951831.
17. DUAN, K. et al. (2019) 'CenterNet: Keypoint Triplets for object detection', 2019 IEEE/CVF International Conference on Computer Vision (ICCV) [Preprint]. doi:10.1109/iccv.2019.00667.
18. GONZALES, L. (2019) A look at mobilenetv2: Inverted residuals and linear bottlenecks, Medium. Available at: https://medium.com/@luis_gonzales/a-look-at-mobilenetv2-inverted-residuals-and-linear-bottlenecks-d49f85c12423 (Accessed: 30 October 2023).
19. TSANG, S.-H. (2019) Review: FPN-feature pyramid network (object detection), Medium. Available at: https://towardsdatascience.com/review-fpn-feature-pyramid-network-object-detection-262fc7482610.

20. GOYAL, S. dan JAIN, A. (2021) 'Yoga pose perfection using deep learning', Journal of Student Research, 10(3), pp. 1–10. doi:10.47611/jsrhs.v10i3.2140.
21. ZAELANI, F. dan MIFTAHUDDIN, Y. (2022) 'Perbandingan metode Efficientnetb3 Dan MobileNetV2 untuk Identifikasi Jenis Buah-Buahan menggunakan fitur Daun', Jurnal Ilmiah Teknologi Infomasi Terapan, 9(1), pp. 1–11. doi:10.33197/jitter.vol9.iss1.2022.911
22. LINA, L. et al. (2022) 'Identifikasi emosi pengguna konferensi video Menggunakan Convolutional Neural Network', Jurnal Teknologi Informasi dan Ilmu Komputer, 9(5), p. 1047. doi:10.25126/jtiik.2022955269.
23. GROVER, ANUJ, ARORA, D. dan GROVER, ANANT (2022) 'Keypoint detection for identifying body joints using tensorflow', Proceedings of the 4th International Conference on Information Management & Machine Intelligence, pp. 1–6. doi:10.1145/3590837.3590948.
24. RAHMAWATI, V.N., YUNIARNO, E.M. dan SUSIKI NUGROHO, S.M. (2023) klasifikasi gerakan pencak silat menggunakan convolutional neural network berbasis body pose. thesis. Institut Teknologi Sepuluh Nopember.
25. TO, T.A., PUSPITASARI, A.A. dan LEE, B.M. (2023) Efficient automatic modulation classification for Next Generation Wireless Networks, pp. 1–12. doi:10.36227/techrxiv.23632308.
26. ABDUL MUTHALIB, M. et al. (2023) 'Pengiraan pose model Manusia Pada repetisi Kebugaran Ai Pemograman Python Berbasis Komputerisasi', INFOTECH journal, 9(1), pp. 11–19. doi:10.31949/infotech.v9i1.4233.