E-ISSN: 2717-3453 | P-ISSN: 2783-1442

Journal of Fuzzy Extension and Applications

www.journal-fea.com

J. Fuzzy. Ext. Appl. Vol. 3, No. 3 (2022) 219–230.

Paper Type: Research Paper

Age and Gender Classification from Facial Features and Object Detection with Machine Learning

Mehmet Karahan[1,*], Furkan Lacinkaya[1], Kaan Erdonmez[1], Eren Deniz Eminagaoglu[1], Cosku Kasnakoglu[1]

[1] Department of Electrical and Electronics Engineering, TOBB University of Economics and Technology, Ankara, Turkey; m.karahan@etu.edu.tr; flacinkaya@etu.edu.tr; kerdonmez@etu.edu.tr; eeminagaoglu@etu.edu.tr; kasnakoglu@etu.edu.tr.

Citation:

Karahan, M., Lacinkaya, F., Erdonmez, K., Eminagaoglu, E. D., & Kasnakoglu, C. (2022). Age and gender classification from facial features and object detection with machine learning. *Journal of fuzzy extension and applications, 3*(3), 219-230.

Received: 06/02/2022 Reviewed: 03/03/2022 Revised: 03/04/2022 Accepted: 16/04/2022

Abstract

In recent years, development of the machine learning algorithms has led to the creation of intelligent surveillance systems. Thanks to the machine learning, it is possible to perform intelligent surveillance by recognizing people's facial features, classifying their age and gender, and detecting objects around instead of ordinary surveillance. In this study, a novel algorithm has been developed that classifies people's age and gender with a high accuracy rate. In addition, a novel object recognition algorithm has been developed that detects objects quickly and with high accuracy. In this study, age and gender classification was made based on the facial features of people using Convolutional Neural Network (CNN) architecture. Secondly, object detection was performed using different machine learning algorithms and the performance of the different machine learning algorithms was compared in terms of median average precision and inference time. The accuracy of the age and gender classification algorithm was tested using the Adience dataset and the results were graphed. The experimental results show that age and gender classification algorithms successfully classify people's age and gender. Then, the performances of object detection algorithms were tested using the COCO dataset and the results were presented in graphics. The experimental results stress that machine learning algorithms can successfully detect objects.

Keywords: Face detection, Facial feature extraction, Convolutional neural network, Gender classification, Age classification, Machine learning, Object detection.

1 | Introduction

Machine learning defines the study of computer algorithms which can be automatically improved through experience and by the use of data [1]. In our era, machine learning has two main missions, one is to classify data according to developed models and the other is to make predictions for future outcomes based on these models. Thanks to classification algorithms, people's faces, age, gender, human behavior and objects can be classified [2]. In this study, an algorithm that recognizes facial features of people and classifies their age and gender has been developed with the help of machine learning. In addition, in the second part of the study, algorithms that recognize objects with machine learning have been developed. Face detection and facial feature extraction are significant for the face tracking, facial expression recognition and face recognition. Facial feature extraction plays a crucial

Corresponding Author: m.karahan@etu.edu.tr

http://dx.doi.org/10.22105/jfea.2022.328472.1201

role in the areas of human computer interaction [3], video monitoring and person identification [4]. Age and gender classification algorithms are mainly based on the identification of facial features [5].

Viola et al. [23] developed the Viola Jones algorithm with the aim of face detection. This algorithm consists of four steps. In stage 1, face images are determined using Haar features. In step 2, the speed of calculating Haar features is reduced by using integral images. In step 3, using a cascaded AdaBoost classifier, a face dataset is trained. In step 4, a trained detection classifier is used to detect the final face image [6].

Marčetić et al. [7] developed a two stage model with the aim of face detection. While stage 1 depends on normalized pixel stage and its aim is reduce false negative face detection, stage 2 based on the deformable part model and decreases false positive detections.

Gupta et al. [8] combined image processing and pattern recognition methods to extract facial features. K-mean clustering and morphological techniques are used to determine facial landmarks. Chowdhury et al. [9] proposed a system that detects facial features by analyzing color components in the images of human faces.

Ko et al. [10] used facial features to classify age and gender of people. Local Binary Patterns is used classify gender of young and adult people and Euclidean distance among facial feature points is used to classify gender of old people. Higashi et al. [11] convolved images with Gabor filters and encoded with local directional pattern to classify age and gender of people. The dimensions of the image is reduced by the principal component analysis and support vector machine is used to classify feature vector.

Tang et al. [12] compared the performance of models based on region proposals and the models based on regression with the aim of object detection. Redmon et al. [13] focused on real-time object recognition problem and used You Only Look Once (YOLO) architecture for object detection. Li et al. [14] proposed a light-weight RetinaNet structure with the aim of reducing computation for RetinaNet for object detection.

In this study, novel algorithm has been developed that extracts facial features of people and classifies people's age and gender with a high accuracy rate. In this study, face detection, facial features extraction, age estimation and gender classification are presented using Python and OpenCV programming languages. Viola-Jones algorithm is used for face detection. Dlib machine learning toolkit is used with Python and OpenCV for facial features extraction. Convolutional Neural Network (CNN) architecture is used with the aim of age and gender classification [15]. Secondly, novel object detection algorithms detect objects quickly and with a high accuracy rate have been developed. The main contributions of this paper are summarized as follows:

- *Machine learning based facial features extraction is achieved.*
- *A novel algorithm based on CNN has been developed that accurately classifies people's age and gender.*
- *Novel object detection algorithms have been developed and time response and accuracy of these algorithms have been compared.*

The rest of this study is organized as follows. In Section 2, the methods used in the research are explained under 3 subheadings. Section 3 gives experimental results of our algorithms. Section 4 discusses the conclusion of this study.

Karahan et al. | J. Fuzzy. Ext. Appl. 3(3) (2022) 219-230

2 | Methods

2.1 | Face Detection and Facial Feature Extraction Algorithm

Face detection depends on the Viola Jones algorithm. The most important features of this algorithm are stable and real time. Thanks to these features, the algorithm works quickly and with high accuracy. Viola Jones algorithm consists of 4 subcomponents. The first of these components is the Haar Cascade. In this section, a picture placed in the system is scanned with rectangles and operations are performed on the pixels of the picture. Another subtitle is integral images. In this part of the algorithm, the process of collecting the pixels from the previous section is performed here in an accelerated manner. The third stage is the AdaBoost algorithm. The identification and learning of the regions that are candidates for finding faces in the picture given to the system are done in this section. In the last step of the algorithm, the cascade classifier operation is performed. In this section, it gives a positive or negative response depending on whether there is a face in the regions that are determined with the previous steps and are candidates for face. After all these steps are completed, the face detection process will be completed [16].

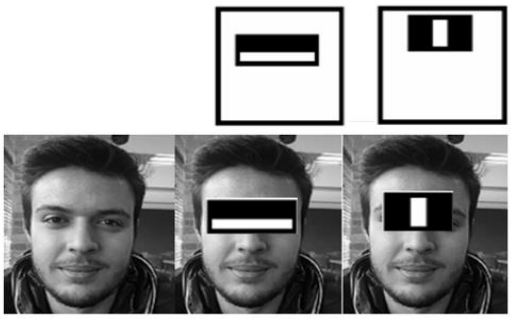

Fig. 1. Haar features on a person's face.

In this study, Dlib library is used to extract facial features of people. Dlib is an independent software library developed in C++ language and it includes machine learning algorithms. Dlib can also be used in the Python programming language. This library contains a face pointer data set. This data set contains 68 pointers expressing the human facial features and the boundaries of parts such as mouth, eyes, eyebrows and nose on the face. Each pointer represents a specific point on the face. Each pointer has (x, y) coordinates and can be expressed in *Eq. (1)* [17].

$$P_1(x_1, y_1), P_2(x_2, y_2) \dots \dots . P_{68}(x_{68}, y_{68}). \quad (1)$$

2.2 | Age and Gender Classification Algorithm

CNN is a deep learning computer vision algorithm that can detect, classify and reconstruct images with high accuracy. CNN consists of five steps and takes images as input. The first step is the convolutional layer step. In this step, low and high level features in the image are extracted with the help of some filters and thus, the features of the picture given to the system as input are determined. Afterwards, the transition to the non-linearity layer step is made. The reason for the application of this step is that since all layers applied before can be a linear function, the neural network can act as if a single perception is made. In this case, the result can be calculated by the linear combination of outputs. Thus, non-linearity is introduced to the system. The third step of the algorithm is called the Pooling Layer. In this step, the process of reducing the slip size of the representation and the number of parameters and calculations in the network is performed, and the appropriate and unsuitable parts are checked. The next step is Flattening Layer and its aim is to prepare the data at the input of the fully connected layer. The last and most important step is called fully connected-layer, in this step, data from the previous layer is taken and the learning process is performed [18].

In our study, CNN algorithm performs age and gender classification with 3 convolutional layers. These are 2 fully connected layers and a final output layer. The first convolutional layer has 96 nodes with kernel size 7. The second convolutional layer has 256 nodes with kernel size 5 and the third convolutional layer has 384 nodes with kernel size 3. Each of the two fully connected layers has 512 nodes. The output layer in the gender classification network is of the softmax type with 2 nodes specifying two classes, "Male" and "Female". Adience dataset is used to train the model. This should be approached as a regression problem, as a true and clear number is expected as a result of the gender estimation. In this solution, estimation is made by classifying age groups. There are 8 classes in the Adience data set divided into the eight different age groups. These groups are (0-3), (4-6), (7-12), (13-19), (20-30), (31-45), (46-59) and (60-100). Therefore, the age classification network has 8 nodes that show the age ranges specified in the last softmax layer [19].

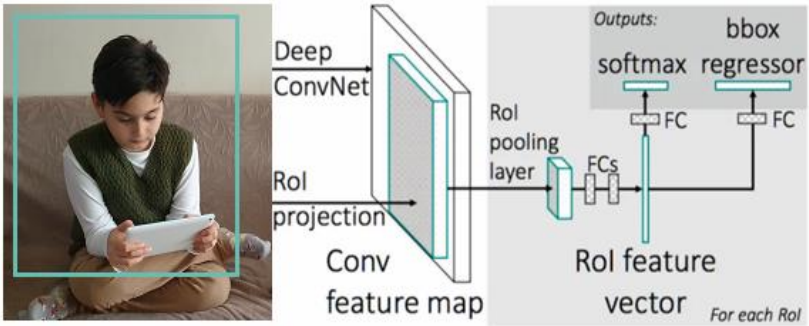

Fig. 2. Illustration of the CNN.

2.3 | Object Detection Algorithm

In this part of the study, object detection is performed using RetinaNet and YOLOv3 architectures. The results obtained were compared in speed and accuracy. *Fig. 3* gives a schematic representation of the RetinaNet architecture. RetinaNet architec-ture consists of 4 basic layers [20].

Bottom-Up pathway. The ResNet backbone network which computes the feature maps at various scales, regardless of the size of input image or the backbone.

Top-Down pathway and lateral connections. The top-down path upsamples spatially coarser feature maps from higher pyramid levels, and lateral links join top-down layers and bottom-up layers in the same spatial dimension.

Classification subnetwork. It estimates the probability of finding an object at each spatial location for each junction box and object class.

Regression subnetwork. Regresses the offset of bounding boxes from junction boxes for each ground reality object.

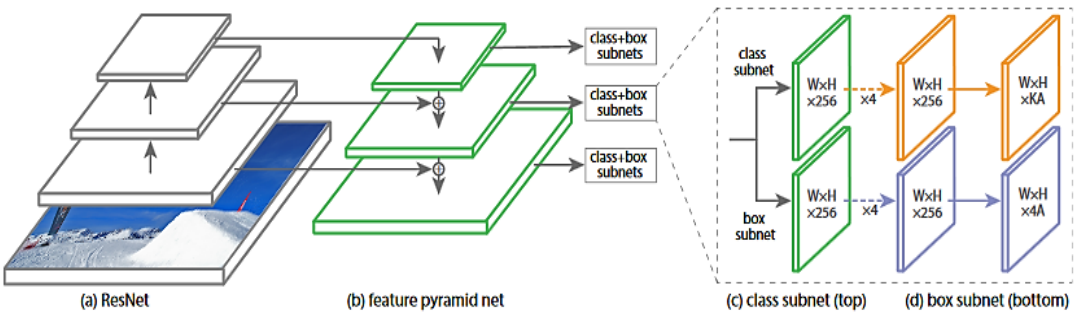

Fig. 3. RetinaNet architecture.

Karahan et al. | J. Fuzzy. Ext. Appl. 3(3) (2022) 219-230

In this study, the YOLO algorithm was also used for object detection. YOLO is an algorithm that can detect objects using CNN. It stands for "YOLO". This name was chosen because the algorithm is fast enough to detect objects in one go. When the YOLO algorithm starts working, it simultaneously detects objects in images or videos and their coordinates.

The most important feature that distinguishes YOLO from other algorithms is its real-time object detection. Although programs capable of real-time object detection such as YOLO have existed before, the Mean Average Precision (MAP) value is not sufficient for these programs. The reason why the YOLO algorithm is so fast is that it can predict the class and coordinates of all objects in the picture by passing the picture through the neural network at once.

There are more advanced versions of the YOLO algorithm, YOLOv2 and YOLOv3. In the YOLOv2 version, the success rate in detecting nearby objects has been increased compared to the previous version. One of the most important innovations is the integration of the "anchor boxes" feature. Since an output vector can recognize an object, it becomes a serious problem if there are multiple objects in the same cell. As a solution, the cell was expanded in width/length with "anchor boxes". These improvements made in the YOLOv2 version have increased the accuracy rate. The YOLOv3 version was used in this study. This version is the latest and most advanced version of the YOLO algorithm. The main innovation with this version is the determination of bounding boxes for different sizes. The base layer structure has been increased to 53. YOLOv3 has models trained in different layouts. The COCO dataset was used as the dataset. With the trained models, 80 objects can be detected. *Fig. 4* shows the structure of the YOLOv3 architecture [21].

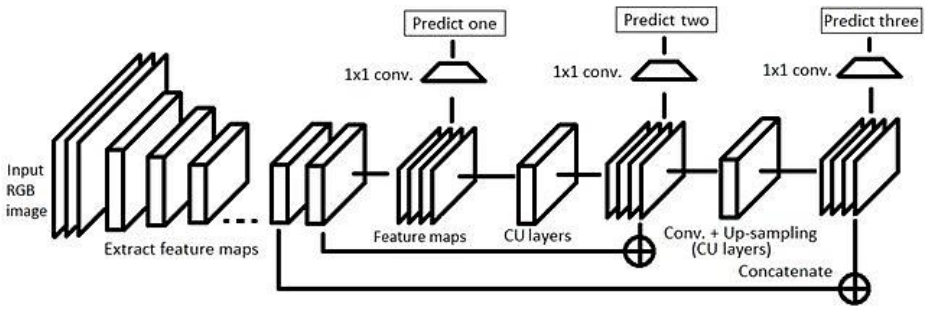

Fig. 4. The structure of the YOLOv3 architecture.

3 | Results

3.1 | Experimental Results of Age and Gender Classification

In this section, experimental results of the face detection, facial feature extraction and age and gender classification algorithms are presented. *Fig. 5* shows the face detection of a single person and a group of people. *Fig. 6* shows the result of the facial feature extraction algorithm. Facial feature algorithm successfully detects a person's eyes, mouth, nose and eyebrows. *Fig. 7* through *Fig. 12* gives the examples of machine learning based age and gender classification.

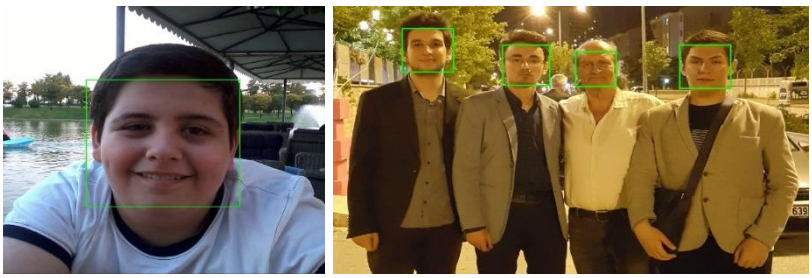

Fig. 5. Detecting a single person's face and a group of people's faces.

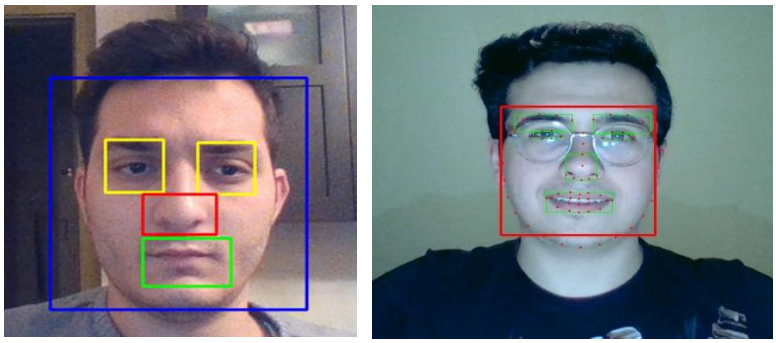

Fig. 6. Facial feature extraction to detect a person's mouth, nose and eyes; the face on the right illustrates the detection of eyebrows and the jawline as well.

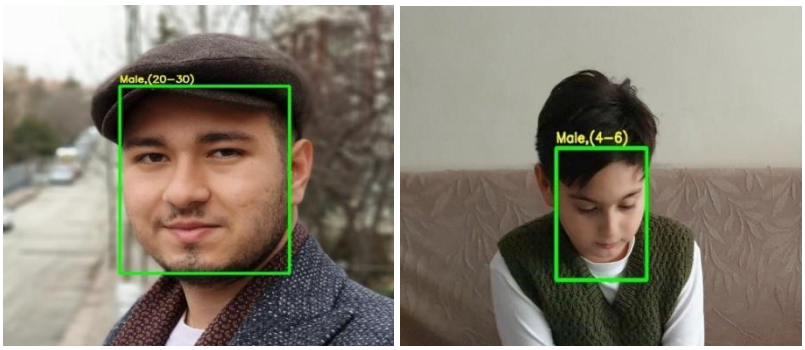

Fig. 7. Age and gender classification of a young adult and a child via CNN architecture.

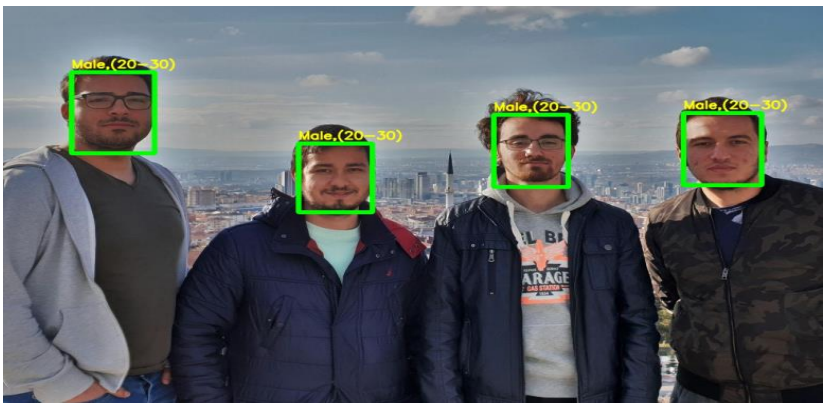

Fig. 8. The age and gender classification of four young men via CNN architecture.

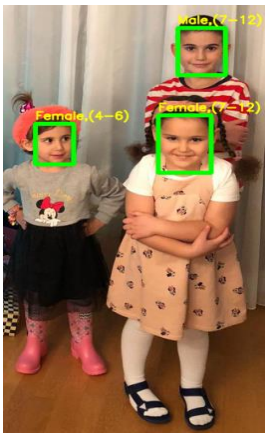

Fig. 9. The age and gender classification of two girls and a boy via CNN architecture.

Karahan et al.|J. Fuzzy. Ext. Appl. 3(3) (2022) 219-230

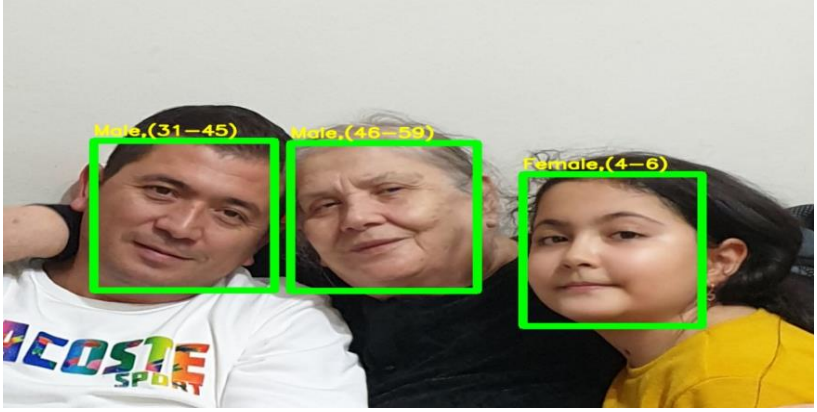

Fig. 10. Age and gender classification of a middle aged man, an elderly, and a child.

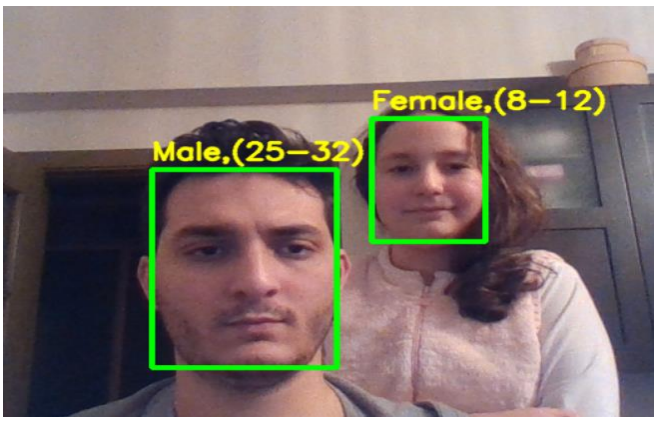

Fig. 11. Age and gender classification of a young man and a girl.

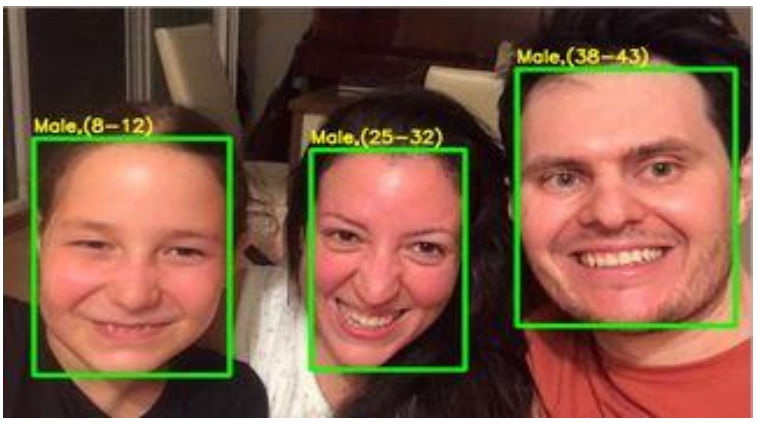

Fig. 12. Age and gender classification of a girl, a young woman and a middle aged man.

The accuracy of age and gender classification algorithm is tested using Adience dataset [19]. *Fig. 13* represents the accuracy of age classification algorithm. The graphic shows accuracy percentage of the algorithm according to age groups. The situation of estimating the age group of the person with 1 less or 1 more accuracy is shown in the correct & 1-off column on the graphic. One hundredth of the number of samples used for each age group is shown in the sample space/100 column in the graph. *Fig. 14* shows the accuracy of the gender classification algorithm.

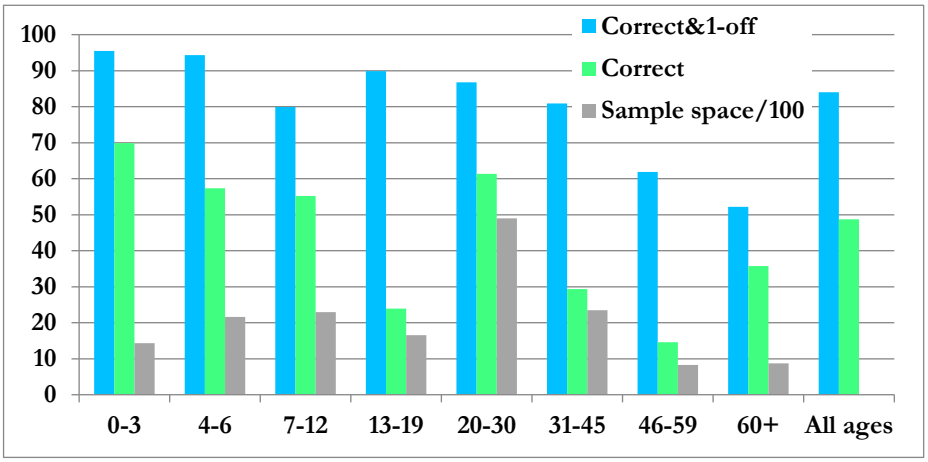

Fig. 13. Accuracy of the age estimation algorithm.

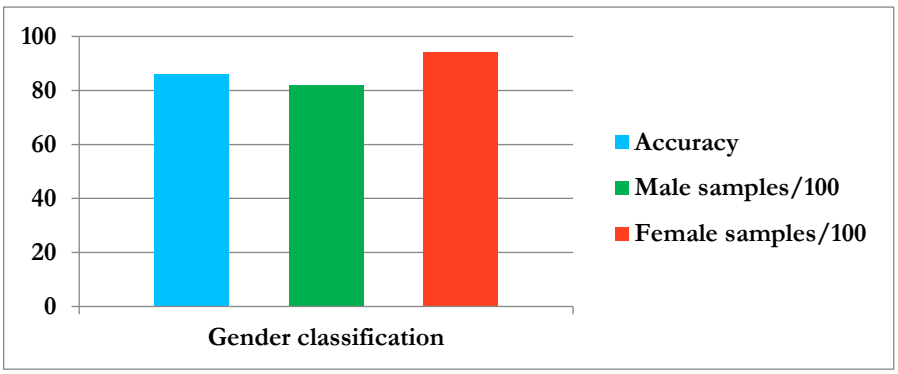

Fig. 14. Accuracy of the gender classification algorithm.

3.2 | Experimental Results of Object Detection

Fig. 15 shows the detection of objects on a study table using the RetinaNet architecture. In *Fig. 16*, object detection was performed using the YOLOv3 architecture in the same photograph. When the two figures are compared, it is seen that the YOLOv3 algorithm recognizes more objects. While the YOLOv3 algorithm could recognize the books on the shelf above the study table, the RetinaNet algorithm could not recognize these objects.

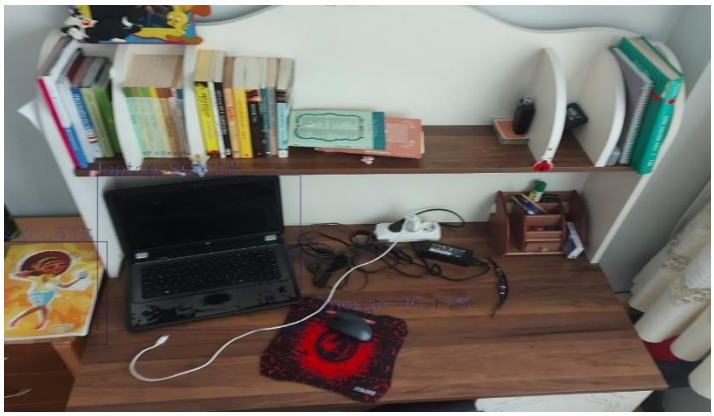

Fig. 15. Objects recognized using RetinaNet architecture on the study table.

Karahan et al.|J. Fuzzy. Ext. Appl. 3(3) (2022) 219-230

Fig. 16. Objects recognized using YOLOv3 architecture on the study table.

Fig. 17 shows the detection of objects in the living room with RetinaNet. *Fig. 18* shows the detection of the same objects with YOLOv3. When the figures are examined, it is understood that the YOLOv3 architecture is more successful. RetinaNet mistakenly recognized the basket in the photograph as a chair. Also, RetinaNet could not recognize a chair in the photo. YOLOv3 did not misidentify any objects. Also, YOLOv3 recognized the chair that RetinaNet could not.

Fig. 17. Objects recognized with RetinaNet architecture in the living room.

Fig. 18. Objects recognized with YOLOv3 architecture in the living room.

Fig. 19 shows the detection of objects on the street with the RetinaNet architecture. *Fig. 20* shows the detection of the same objects with YOLOv3. RetinaNet was able to recognize only one car. The YOLOv3 algorithm, on the other hand, was able to recognize all 3 cars in the photograph.

Fig. 19. Objects recognized with RetinaNet architecture on the street.

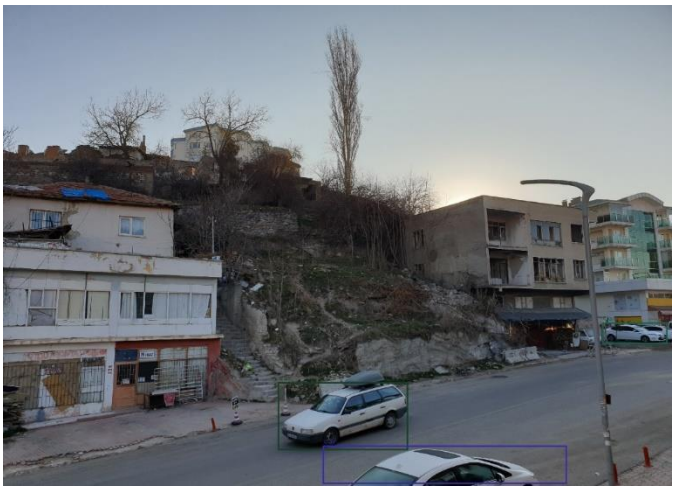

Fig. 20. Objects recognized with YOLOv3 architecture on the street.

Fig. 21 shows a graph giving the MAP and inference time of different object detection algorithms for the COCO dataset. When the graph is examined, it is understood that the YOLOv3 algorithm recognizes faster and with higher precision than other algorithms [22].

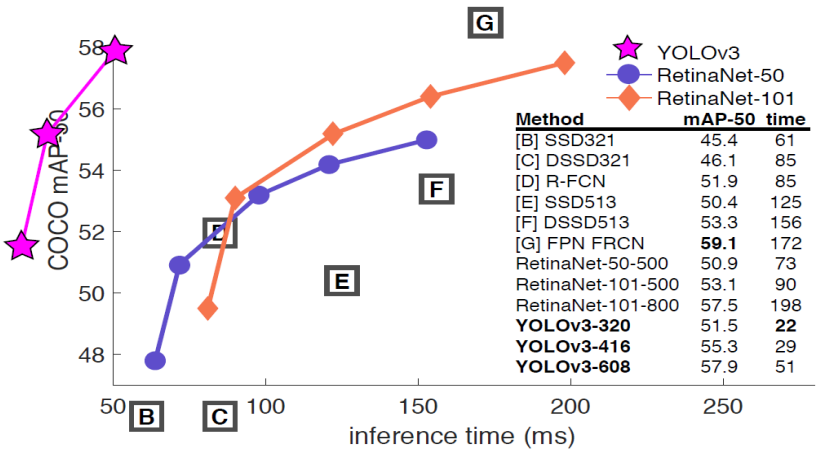

Fig. 21. Performance comparison of object detection algorithms.

4 | Conclusion

In this study, face detection, facial feature extraction, age and gender classification and object detection are presented using Python and OpenCV programming languages. Face detection algorithm can detect

Karahan et al.|J. Fuzzy. Ext. Appl. 3(3) (2022) 219-230

human faces and machine learning based novel facial feature extraction algorithm successfully extracts mouth, nose, eyebrows and eyes. Our novel CNN architecture performs age and gender classification with high accuracy. The data on the accuracy rate of age and gender classification are presented on graphics. Experimental results stress that face detection, facial feature extraction algorithms successfully detect facial landmarks and age and gender classification algorithm work with high accuracy. RetinaNet and YOLOv3 architectures are used for object detection. The obtained results were compared in terms of MAP and inference time. As a result of the comparisons, it has been observed that the YOLOv3 algorithm is faster and works with a higher accuracy rate.

Conflicts of Interest

All co-authors have seen and agree with the contents of the manuscript and there is no financial interest to report. We certify that the submission is original work and is not under review at any other publication.

References

[1] Eyo, I. J., Adeoye, O. S., Inyang, U. G., & Umoeka, I. J. (2022). Hybrid intelligent parameter tuning approach for COVID-19 time series modeling and prediction. *Journal of fuzzy extension and applications*, *3*(1), 64-80.

[2] Javanbakht, T., & Chakravorty, S. (2022). Prediction of human behavior with TOPSIS. *Journal of fuzzy extension and applications*, *3*(2), 109-125.

[3] Benedict, S. R., & Kumar, J. S. (2016). Geometric shaped facial feature extraction for face recognition. *2016 IEEE international conference on advances in computer applications (ICACA)* (pp. 275-278). IEEE.

[4] Wu, Y. M., Wang, H. W., Lu, Y. L., Yen, S., & Hsiao, Y. T. (2012). Facial feature extraction and applications: a review. *Asian conference on intelligent information and database systems* (pp. 228-238). Springer, Berlin, Heidelberg.

[5] Karahan, M., Kurt, H., & Kasnakoglu, C. (2020). Autonomous face detection and tracking using quadrotor UAV. *2020 4th international symposium on multidisciplinary studies and innovative technologies (ISMSIT)* (pp. 1-4). IEEE.

[6] Lu, W. Y., & Ming, Y. A. N. G. (2019). Face detection based on viola-jones algorithm applying composite features. *2019 international conference on robots & intelligent system (ICRIS)* (pp. 82-85). IEEE.

[7] Marčetić, D., Hrkać, T., & Ribarić, S. (2016). Two-stage cascade model for unconstrained face detection. *2016 first international workshop on sensing, processing and learning for intelligent machines (SPLINE)* (pp. 1-4). IEEE.

[8] Gupta, S., & Singh, R. K. (2015). Mathematical morphology based face segmentation and facial feature extraction for facial expression recognition. *2015 international conference on futuristic trends on computational analysis and knowledge management (ABLAZE)* (pp. 691-695). IEEE.

[9] Chowdhury, M., Gao, J., & Islam, R. (2016). Fuzzy rule based approach for face and facial feature extraction in biometric authentication. *2016 international conference on image and vision computing New Zealand (IVCNZ)* (pp. 1-5). IEEE.

[10] Ko, J. B., Lee, W., Choi, S. E., & Kim, J. (2014). A gender classification method using age information. *2014 international conference on electronics, information and communications (ICEIC)* (pp. 1-2). IEEE.

[11] Higashi, A., Yasui, T., Fukumizu, Y., & Yamauchi, H. (2011). Local Gabor directional pattern histogram sequence (LGDPHS) for age and gender classification. *2011 IEEE statistical signal processing workshop (SSP)* (pp. 505-508). IEEE.

[12] Tang, C., Feng, Y., Yang, X., Zheng, C., & Zhou, Y. (2017). The object detection based on deep learning. *2017 4th international conference on information science and control engineering (ICISCE)* (pp. 723-728). IEEE.

[13] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: unified, real-time object detection. *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 779-788). IEEE.

[14] Li, Y., Dua, A., & Ren, F. (2020). Light-weight retinanet for object detection on edge devices. *2020 IEEE 6th world forum on internet of things (WF-IoT)* (pp. 1-6). IEEE.

[15] Karahan, M., Lacinkaya, F., Erdonmez, K., Eminagaoglu, E. D., & Kasnakoglu, C. (2021). Face detection and facial feature extraction with machine learning. *International conference on intelligent and fuzzy systems* (pp. 205-213). Springer, Cham.

[16] Viola, P., & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition, CVPR 2001* (pp. I-I). IEEE.

[17] Ayvaz, U., & Gürüler, H. (2017). The detection of emotional expression towards computer users. *International journal of informatics technologies*, *10*(2), 231-239.

[18] Lin, S. D., & Chen, K. (2019). Illumination invariant thermal face recognition using convolutional neural network. *2019 IEEE international conference on consumer electronics-Asia (ICCE-Asia)* (pp. 83-84). IEEE.

[19] Eidinger, E., Enbar, R., & Hassner, T. (2014). Age and gender estimation of unfiltered faces. *IEEE Transactions on information forensics and security*, *9*(12), 2170-2179.

[20] Jung, H., Kim, B., Lee, I., Yoo, M., Lee, J., Ham, S., ... & Kang, J. (2018). Detection of masses in mammograms using a one-stage object detector based on a deep convolutional neural network. *PloS one*, *13*(9), e0203355. https://doi.org/10.1371/journal.pone.0203355

[21] Punn, N. S., Sonbhadra, S. K., Agarwal, S., & Rai, G. (2020). *Monitoring COVID-19 social distancing with person detection and tracking via fine-tuned YOLO v3 and Deepsort techniques*. Retrieved from https://doi.org/10.48550/arXiv.2005.01385

[22] Redmon, J., & Farhadi, A. (2018). Yolov3: an incremental improvement. Retrieved from https://doi.org/10.48550/arXiv.1804.02767

[23] Viola, P., Jones, M. J., & Snow, D. (2005). Detecting pedestrians using patterns of motion and appearance. *International journal of computer vision, 63*(2), 153-161.