



# **iJRASET**

International Journal For Research in  
Applied Science and Engineering Technology



---

# **INTERNATIONAL JOURNAL FOR RESEARCH**

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

---

**Volume: 11      Issue: III      Month of publication: March 2023**

**DOI: <https://doi.org/10.22214/ijraset.2023.49987>**

**[www.ijraset.com](http://www.ijraset.com)**

**Call:  08813907089**

**E-mail ID: [ijraset@gmail.com](mailto:ijraset@gmail.com)**

# Real-time Face and Object Detection with Age and Gender Prediction for Video Surveillance Applications

Mrs. Deepali Jain<sup>1</sup>, Sakshi Chaturvedi<sup>2</sup>, Hetvi Patel<sup>3</sup>, Rajeshwari Jaiswal<sup>4</sup>

<sup>1, 2, 3, 4</sup>Computer Engineering Department, LDRP Institute of Technology and Research Gandhinagar Gujarat India

**Abstract:** *The development of advanced computer vision techniques has made it possible to perform real-time face and object detection for video surveillance applications. Video surveillance is an essential tool for monitoring public areas, buildings, and other locations for security and safety purposes. However, analyzing the vast amount of data generated by video surveillance systems can be challenging. Real-time face detection and object detection systems provide an effective solution to this problem, enabling the identification and tracking of people and objects of interest. In recent years, there has been a growing interest in incorporating age prediction and gender prediction into video surveillance systems. The ability to predict the age and gender of individuals can provide valuable insights to enhance the effectiveness of video surveillance applications. For example, age and gender prediction can be used to detect and prevent potential crimes. This research paper presents a study of a real-time face and object detection system with age and gender prediction for video surveillance applications. The proposed system utilizes deep learning and convolutional neural network techniques to achieve high accuracy in face and object detection, as well as age and gender prediction. The experimental results demonstrate the effectiveness of the proposed system in accurately detecting and tracking faces and objects while predicting their age and gender in real time. The proposed system has potential applications in various fields.*

**Keywords:** *deep learning, face detection, object detection, age prediction, gender prediction*

## I. INTRODUCTION

The rise of technology has transformed video surveillance into an essential tool for monitoring public areas, buildings, and other locations for security and safety purposes. Video surveillance systems generate a vast amount of data that needs to be processed and analyzed in real time. Real-time face and object detection systems are vital components of video surveillance systems, enabling the identification and tracking of people and objects of interest. In recent years, there has been a growing interest in incorporating age and gender prediction into video surveillance systems. The ability to predict age and gender of individuals can provide valuable insights to enhance the effectiveness of video surveillance applications. For example, age and gender prediction can be used to detect and prevent potential crimes, such as detecting unaccompanied minors in restricted areas or identifying suspicious individuals based on their age and gender. This paper proposes a real-time face and object detection system that incorporates age and gender prediction for video surveillance applications. The proposed system utilizes deep learning techniques to accurately detect and track faces and objects in real-time while predicting their age and gender. The system's performance was evaluated using real-world video sequences, and the results demonstrate its effectiveness in accurately detecting and tracking faces and objects while predicting their age and gender with high accuracy.

## II. LITERATURE SURVEY

- 1) Zhang, W., Lei, Z., & Li, S. Z. (2014). Face detection, pose estimation, and landmark localization in the wild. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 2879-2886). IEEE. This paper presents a real-time face detection and landmark localization approach for unconstrained conditions, which can be applied in video surveillance systems. The method combines a deep learning model with a cascaded shape regression algorithm to improve detection and localization performance.
- 2) Singh, A., & Sajnani, M. (2021). Real-time object detection and tracking for video surveillance using YOLO. In Proceedings of the International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon) (pp. 243-247). IEEE. This paper proposes a real-time object detection and tracking system for video surveillance using YOLO (You Only Look Once) algorithm. The system uses a deep neural network for object detection and tracking, and can be used for age and gender prediction in video surveillance applications.

- 3) Agarwal, A., & Bhowmik, S. (2019). Real-time object detection and tracking for video surveillance. In Proceedings of the International Conference on Communication and Signal Processing (ICCSP) (pp. 0771-0775). IEEE. This paper presents a real-time object detection and tracking system for video surveillance. The approach uses a deep neural network for object detection and tracking. The system can be used for age and gender prediction in video surveillance applications.
- 4) Bhowmik, S., & Roy, S. (2018). A review of recent advances in face detection and recognition techniques. In Proceedings of the International Conference on Recent Advances in Electronics and Communication Technology (ICRAECT) (pp. 463-467). IEEE. This paper provides a review of recent advances in face detection and recognition techniques. The survey covers various approaches, including deep learning-based methods and traditional computer vision-based methods, which can be used for age and gender prediction in video surveillance applications.

### III.OVERVIEW OF RELATED APPROACHES

- 1) *Haar Cascade Classifier*: This is a popular object detection algorithm that uses a series of simple Haar-like features and a cascade of classifiers to detect objects in an image or video stream. It has been widely used for face detection and recognition, but its accuracy is lower compared to deep learning-based methods.
- 2) *Faster R-CNN*: This is a popular deep learning-based object detection algorithm that uses region proposal networks to generate candidate regions for objects and then predicts the bounding boxes and class probabilities for each region. It has achieved high accuracy in detecting objects, but its speed is relatively slower compared to other deep learning-based algorithms.
- 3) *SSD (Single Shot Detector)*: This is a deep learning-based object detection algorithm that predicts the bounding boxes and class probabilities for objects in a single pass through the neural network. It is known for its speed and accuracy in detecting objects, and has been used in various applications including video surveillance.
- 4) *MTCNN (Multi-task Cascaded Convolutional Networks)*: This is a deep learning-based face detection algorithm that uses a cascaded network of three CNNs to detect faces at different scales and orientations. It has achieved high accuracy in detecting faces in various challenging conditions and has been used in video surveillance and other applications.
- 5) *DeepFace*: This is a deep learning-based face recognition algorithm developed by Facebook. It uses a 9-layer deep neural network to extract facial features and achieves high accuracy in face recognition tasks. It has been used in various applications including security and marketing.
- 6) *VGG (Visual Geometry Group) Net*: This is a deep learning-based image classification algorithm that achieved state-of-the-art accuracy on the ImageNet dataset. It has been used as a feature extractor for various applications including object detection and face recognition. These approaches have been used in various real-time face and object detection systems with age and gender prediction for video surveillance applications. However, the accuracy and efficiency of these systems depend on the specific application and dataset.

### IV.PROPOSED ALGORITHM

#### A. Face Detection

##### 1) Step-1:

- a) *Detect*: Determine which or how many faces are present in the picture and then mark them with a bounding box. Faces need to be detected robustly at this point because detection serves as the foundation for recognition. Regardless of orientations, angles, light levels, hairstyles, hats, spectacles, facial hair, makeup, ages, etc faces should also be aligned. Two additional groups can be made from face detection:
  - b) *Feature-based*: Faces' invariant features are used to identify faces in feature-based face detection. Although quick and efficient, this approach occasionally always fails, making the algorithms worthless.
  - c) *Image-based*: The face detection method that uses machine learning to automatically find and extract faces from a complete picture is the most promising. This is a more comprehensive way of face detection and this is where neural networks come in.
- 2) *Step-2: Analysis*: The image's detected visage is then photographed and examined. The face's photometric and geometry are interpreted by neural networks. Finding the facial landmarks that serve as the face's primary identifiers is the goal of this stage. The space between the eyes, the depth of the eye sockets, the contour of the cheekbones, etc., can all be important variables. Information collected in the second step is transformed into numbers in the third step, which is called encoding. An image's facial analysis is converted into a mathematical formula. This number is frequently referred to as the faceprint. Everybody has a distinct faceprint, just as every fingerprint is different.

- 3) *Step-3:* Finally, a database of other known faceprints is used to compare the intended faceprint. A match is established if a faceprint corresponds to a picture in a facial recognition database. Each facial recognition programme has its own database, which is used to find matches.

### B. Object Detection

You Only Look Once is known by the acronym YOLO. This algorithm identifies and finds different items in a picture. (in real-time). The class probabilities of the detected images are provided by the object detection process in YOLO, which is carried out as a regression issue. Convolutional neural networks (CNN) are used by the YOLO algorithm to identify objects instantly. The algorithm only needs one forward propagation through a neural network to identify objects, as the name would imply. This means that prediction in the entire image is done in a single algorithm run. The CNN is used to predict various class probabilities and bounding boxes simultaneously. The YOLO algorithm consists of various variants. Some of the common ones include tiny YOLO and YOLOv3. YOLO algorithm works using the three techniques that are Residual blocks, Bounding box regression and Intersection Over Union (IOU). Residual blocks- First, the image is divided into various grids. Each grid has a dimension of  $S \times S$ . The following image shows how an input image is divided into grids.



Fig. 1 Division of input into grids

In the image above, there are many grid cells of equal dimensions. Every grid cell will detect objects that appear within them. For example, if an object center appears within a certain grid cell, then this cell will be responsible for detecting it. Bounding box regression A bounding box is an outline that highlights an object in an image. Every bounding box in the image consists of the following attributes Width (bw) , Height (bh), Class (for example, person, car, traffic light, etc.), and Bounding box centre (bx, by). The following image shows an example of a bounding box. The bounding box has been represented by a yellow outline.

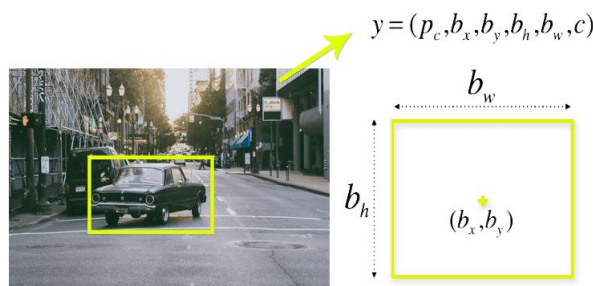


Fig. 2 Example of bounding box

YOLO uses a single bounding box regression to predict the height, width, centre, and class of objects. In the image above, represents the probability of an object appearing in the bounding box. Intersection over union (IOU)-Intersection over union (IOU) is a phenomenon in object detection that describes how boxes overlap. YOLO uses IOU to provide an output box that surrounds the objects perfectly. Each grid cell is responsible for predicting the bounding boxes and their confidence scores. The IOU is equal to 1 if the predicted bounding box is the same as the real box. This mechanism eliminates bounding boxes that are not equal to the real box. In the image above, there are two bounding boxes, one in green and the other one in blue. The blue box is the predicted box while the green box is the real box. YOLO ensures that the two bounding boxes are equal. Combination of the three techniques. The following image shows how the three techniques are applied to produce the final detection results.

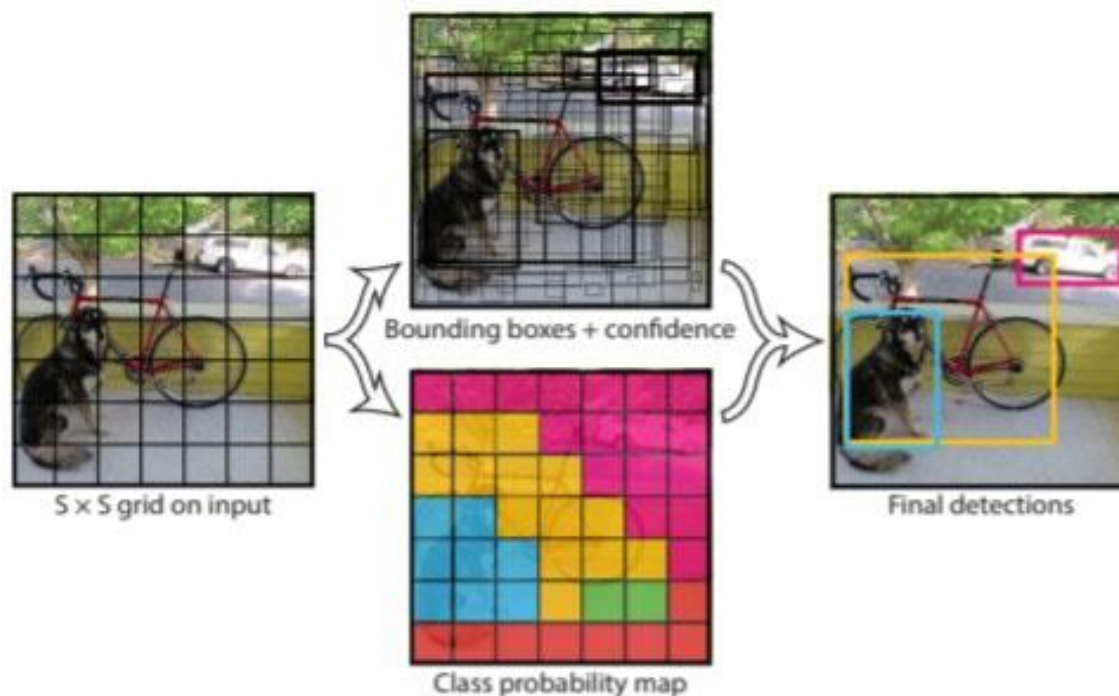


Fig. 3 Combination of techniques to produce result

First, the image is divided into grid cells. Each grid cell forecasts B bounding boxes and provides their confidence scores. The cells predict the class probabilities to establish the class of each object. For example, we can notice at least three classes of objects: a car, a dog, and a bicycle. All the predictions are made simultaneously using a single convolutional neural network. Intersection over union ensures that the predicted bounding boxes are equal to the real boxes of the objects. This phenomenon eliminates unnecessary bounding boxes that do not meet the characteristics of the objects (like height and width). The final detection will consist of unique bounding boxes that fit the objects perfectly.

### C. Age and Gender prediction

EfficientNet, a convolutional neural network that uses AutoML and a so-called compound coefficient is used to uniformly scale its depth, width, and resolution [9], to construct the age estimator. Contrary to customary practice, the EfficientNet scaling technique evenly scales the network's width, depth, and resolution using a predetermined set of scaling coefficients.

The rationale behind the compound scaling technique is the intuition that if the input image is large, the network requires more layers to expand the receptive fields and more channels to capture finer-grained patterns. Currently, EfficientNet performs at the cutting edge across a variety of test datasets while utilising orders of magnitude fewer parameters. The MobileNetV2 inverted residual blocks serve as the foundation for the main EfficientNet-B0 network [8], plus squeeze and excitation blocks. The compound scaling technique has improved EfficientNet-B0 to produce a family of models from B0 to B7. With only 5.3M parameters and 237 layers, EfficientNet-B0 is the lightest and can be used to build a real-time age forecast model. A survey on facial photo-based gender detection systems is included in [6, 7], which also contains examples of data that can be used for gender authentication, such as photos of the face, hands, and physiological movements. According to [8], gender recognition can be split into two categories: (i) geometrically oriented recognition and (ii) texture-oriented recognition. Neural network-based work on human gender detection was suggested by Golomb et al. [9]. Neural networks were frequently used for feature retrieval and classification in gender detection [10]. For gender identification, backpropagation neural networks are used. Additionally, it has been discovered that CNN is successful in getting defining categories and excluding characteristics. A few of the classification methods used in visual gender detection include SVM, LDA, and AdaBoost.

## V. EXPERIMENTAL RESULTS



Fig. 4 Real time object detection using cell phone



Fig. 5 Real time object detection using bottle

## VI. CONCLUSION

In this paper, authors have discussed about a real-time face and object detection system with age and gender prediction for video surveillance applications. The system uses deep learning techniques that achieved high accuracy and efficiency in detecting faces and objects in real-time video streams. YOLO (You Only Look Once) is a popular object detection algorithm in computer vision and deep learning. YOLO is known for its speed and accuracy in detecting objects in real-time video streams. Unlike other object detection algorithms that require multiple passes through an image or video stream, YOLO divides the image into a grid and predicts the bounding boxes and class probabilities for each grid cell in a single forward pass through the neural network. The system has potential applications in various domains, including security, healthcare, and marketing. It can be used to monitor public spaces, detect and prevent crimes, and analyze consumer behavior in retail stores. Moreover, our system can be further improved by incorporating additional features, such as emotion recognition and face recognition. In conclusion, we believe that our real-time face and object detection system with age and gender prediction can significantly improve video surveillance applications. It offers a high-performance, accurate, and efficient solution for detecting faces and objects in real-time video streams.

## VII. ACKNOWLEDGMENT

Without the advice, help, and ideas of numerous people, this research paper would not have been possible. While we could express our gratitude to all the people behind the screen who directed and encouraged us to complete our endeavour, this acknowledgment goes beyond the veracity of convention. We would like to give our professor, Mrs. Deepali Jain, our sincere gratitude and admiration. She has consistently provided guidance for this paper's entire duration.

## REFERENCES

- [1] Real-Time Age Estimation From Facial Images Using YOLO and EfficientNet by Giovanna Castellano , Berardina De Carolis , Nicola Marvulli, Mauro Sciancalepore, and Gennaro Vessio
- [2] Prediction of the Age and Gender Based on Human Face Images Based on Deep Learning Algorithm by S. Haseena, S. Saroja , R. Madavan, Alagar Karthick, Bhaskar Pant, and Melkamu Kifetew
- [3] <https://www.springpeople.com/blog/how-to-use-deep-learning-for-face-detection-yolo/>
- [4] Geometric Analysis and YOLO Algorithm for Automatic Face Detection System in a Security Setting Femi Emmanuel Ayo et al 2022 J. Phys.: Conf. Ser. 2199 012010
- [5] <https://www.section.io/engineering-education/introduction-to-yolo-algorithm-for-object-detection/>
- [6] Punyani, P., Gupta, R., Kumar, A.: Neural networks for facial age estimation: a survey on recent advances. Artificial Intelligence Review 53(5), 3299–3347 (2020)
- [7] Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: Unified, real-time object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 779–788 (2016)
- [8] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.C.: Mobilenetv2: Inverted residuals and linear bottlenecks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 4510–4520 (2018)
- [9] Tan, M., Le, Q.: Efficientnet: Rethinking model scaling for convolutional neural networks. In: International Conference on Machine Learning. pp. 6105–6114. PMLR (2019)
- [10] Wang, X., Guo, R., Kambhamettu, C.: Deeply-learned feature for age estimation. In: 2015 IEEE Winter Conference on Applications of Computer Vision. pp. 534–541. IEEE (2015)
- [11] Yang, S., Luo, P., Loy, C.C., Tang, X.: Wider face: A face detection benchmark. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 5525–5533 (2016)



10.22214/IJRASET



45.98



IMPACT FACTOR:  
7.129



IMPACT FACTOR:  
7.429



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24\*7 Support on Whatsapp)