International Journal of Computing and Artificial Intelligence

E-ISSN: 2707-658X
P-ISSN: 2707-6571
IJCAI 2024; 5(1): 22-29
Received: 19-11-2023
Accepted: 25-12-2023

Dr. Manju Arora
Department of Information Technology Jagan Institute of Management Studies, Rohini, New Delhi, India

Sahil Sharma
Department of Information Technology Jagan Nath University, Bahadurgarh, Haryana, India

Mohd. Asif Khan
Department of Information Technology, Jagan Nath University, Bahadurgarh, Haryana, India

Corresponding Author:
Dr. Manju Arora
Department of Information Technology Jagan Institute of Management Studies, Rohini, New Delhi, India

Object detection and age & gender estimation using deep learning: An overview

Dr. Manju Arora, Sahil Sharma and Mohd. Asif Khan

DOI: https://doi.org/10.33545/27076571.2024.v5.i1a.78

Abstract
In this research work, we explore the field of computer vision with a focus on creating a powerful and versatile framework. Our work leverages deep learning around important tasks such as object detection, age estimation, and gender estimation. By integrating the Mask R-CNN model for object detection and the Deep Face library for age and gender estimation, we propose a solution that transcends the boundaries of one objective. Our approach includes careful information before improving the quality of input images, which demonstrates the efficiency of our model. The Mask R-CNN model provides guidance in object detection by demonstrating the ability to identify and find objects in images. This is the basis for the next project, where we will turn it into age and gender estimation using the Deep Face library. Our test results show not only successful identification of people with reliable scores, but also accurate age and gender predictions. We discuss the complexity of our approach, acknowledge its strengths, and directly address issues that arise when using it.

Keywords: Mask R-CNN, ResNet 50, FPN, Deep Face

Introduction

The combination of deep learning and image analysis is paving the way for a revolution in the development of computer vision. This article begins a journey at the intersection of object detection, age estimation, and gender estimation by offering a comprehensive approach that transcends traditional boundaries.

The importance of computer vision now lies in its ability to identify complex visual information and make it the basis of many things. Research is an important part of our work and forms the basis of our research. Using the standard R-CNN mask, known for its accuracy and efficiency, we accurately complete the complex process of identifying and identifying objects in images.

But our talent is not limited to product recognition. We know the human condition and deeply understand age and gender estimates – the importance of understanding the context of the data found. Our method integrates with the Deep Face library, which is a powerful tool for face analysis and completes the object detection phase.

This research is not just about solving problems alone; It synthesizes these studies into a coherent framework. The integration of product detection, age estimation, and gender estimation demonstrates the importance of our approach. Our framework for addressing these interactions highlights the complexity of real-world situations where comprehensive analysis is required.

When we started this research, our goal was clear: to contribute to the ongoing debate on computer vision, provide immediate solutions, and make progress in the overall field. This combination of cutting-edge designs reflects our commitment to pushing the boundaries of what's possible with visual intelligence.

This introduction sets the stage for entering a holistic framework in which the hard-to-predict characteristics of humans are closely related to real-world discoveries. Through this work, we aim to expand the capabilities of integrated systems for computer vision and foster further innovation in the pursuit of better understanding.

Literature Survey

The following literature review sheds light on the existing research literature on the relationship between object detection, age estimation, and gender estimation: man in computer vision - content is consistent with the recommendation.

1. **"Mask R-CNN", author: He Kaiming *et al*. (2017) [1]:** This publication introduced the Mask R-CNN architecture, which is an important part of the project. This article focuses on a segmentation example that demonstrates the effectiveness of this model in detecting and classifying objects simultaneously. Its adoption is the main role of Mask R-CNN in modern detection devices.
2. **"Deep Face":** Closing the Gap between Human-Level Performance in Facial Analysis" by Yaniv Taigman *et al*. (2014) [9]: Taigman *et al*. introduced Deep Face, a model useful in face analysis. Although the main purpose is facial recognition, its use extends to age and gender estimation. Deep Face's integration project is based on this paper, which highlights the importance of using faces for multidimensional analysis.
3. **"YOLO9000: Better, Faster, More Powerful" (2016) [6-7] by Joseph Redmon and Santosh Divvala:** This introduces the YOLO (one-look) object search algorithm that provides speed and efficiency. Although YOLO is not used directly in the project, it can serve as a model for discovering the true purpose. A comparison with YOLO provides context for the performance and performance of the Mask R-CNN model.
4. **"Age and Gender Classification Using Convolutional Neural Networks" by Gil Levi and Tal Hassner (2015) [11]:** Levi and Hassner offer an introduction to age and gender research that is suitable for this project. Their research on neural networks (CNN) for age and gender prediction serves as a reference. This work using Deep Face is based on the general approach discussed in this article.
5. **"Detecting Bad Faces", Shengcai Liao *et al*. (2018):** Liao *et al*. A simple but effective face detection system. While this project focuses on Mask R-CNN object detection, the information in this article contributes to a broader understanding of mask usage. Comparative analysis shows the importance of the project in terms of accuracy and precision.

These selected research papers are a rich collection of techniques and advances in object detection, age estimation, and gender estimation. Integrating and building on these efforts, the project creates a comprehensive curriculum that includes cutting-edge solutions to real-world challenges in intelligence.

Proposed Methodology

Our goal is to integrate state-of-the-art object detection, age estimation, and gender estimation models in one joint venture. The combination of Mask R-CNN with ResNet 50 FPN for object detection and Deep Face for age and gender estimation forms the basis of our approach.

1. **Data Preprocessing:** A detailed data preprocessing stage is used to improve the quality of the input image. This includes normalizing, resizing, and normalizing to ensure consistency with deep learning models.
2. **Mask Detection Tool R-CNN:** Find the R-CNN model for detection accuracy. This model is trained on different data and is good at identifying and finding objects in images. Spatial coordinates of test objects are recorded using bounding boxes.
3. **Age and gender estimation with Deep Face:** After object detection is completed, the focus is on age and gender estimation. Use the Deep Face library to detect facial features on detected objects. This involves extracting relevant features and estimating age and gender with high accuracy.
4. **Assessment and Evaluation:** The effectiveness of our integration process is meticulously evaluated. Metrics such as precision for object detection, recall, and F1 score, as well as accuracy of age and gender estimation, provide a better understanding of your model's performance.
5. **Comparison with baseline models:** Comparison with baseline models including YOLO for object detection and normal age and gender estimation. This allows us to measure the strengths and weaknesses of our integration in terms of speed, accuracy and efficiency.
6. **Fine-tuning and optimization:** Make iterative improvements to increase the robustness of the model. Fine-tuning parameters, tuning hyperparameters, and optimizing the entire architecture helps strike a balance between accuracy and computational efficiency.
7. **Determination of justice and injustice:** Due to the assumption of age and gender determinants, we emphasize morality and injustice in the model. Explore bias reduction strategies to ensure fair and unbiased estimates.
8. **Implementation of user interface:** To facilitate user interaction, Tkinter is used to implement a graphical user interface (GUI). This allows users to select images, start the detection process, and receive feedback on detected objects, age, and gender.

Mask R-CNN with ResNet 50 FPN: Mask R-CNN is an extension of the Faster R-CNN framework designed to solve segmentation tasks, for example. It includes object detection and pixel-level segmentation by extending R-CNN faster to estimate segmentation masks for each object. The architecture has three main components:

- **Backbone Network (ResNet50):** ResNet50 is used as the backbone to provide a deep and effective representation of the input image. Residual connections lead to training very deep networks, which helps improve extraction.
- **Regional Bid Network (RPN):** RPN creates regional bids for potential products. To advance to the next level of the network, the candidate with the product matching score announces the bounding box.
- **ROI Alignment and Mask Title:** Outline of Interest (ROI) alignment uses binaural interpolation to extract features from defined regions while preserving the data space. Face mask estimates the segmentation mask for each defined area by adjusting the object boundary.
- **Feature Pyramid Network (FPN):** Pyramid Network improves the Mask R-CNN architecture by solving the challenge of processing different features. FPN uses a top-down architecture and horizontal connections to create a pyramid feature in the backbone network. This pyramid allows the model to see things at different levels.

- **ResNet50-FPN and Mask R-CNN Integration:** The system is recommended with ResNet50-FPN as the backbone in the Mask R-CNN framework.

Deep Face: Deep Face is a deep facial recognition library that uses neural networks (CNN) to extract facial features. Developed by Facebook AI Research (FAIR), Deep Face stands out in many facial analysis tasks thanks to its advanced design. Deep Face excels in many facial analysis techniques:

- **Face Detection:** The core functionality supports the convolutional process to accurately identify and view faces in images. This important study is important for subsequent analysis.
- **Localization of facial landmarks:** Deep Face identifies facial landmarks by pinpointing the position of the face. This information is useful for subsequent tasks such as age estimation and gender analysis.
- **Age Estimation:** Deep Face estimates the age of people in facial images using its previously learned model. The age estimation method uses deep neural networks to determine the age range with high accuracy.
- **Gender Determination:** Deep Face uses the gender classification model to determine the gender of individuals in facial images. Prior learning of this model contributes to its ability to make recommendations about different types of information.

Implementation of Workflow diagram and Gantt chart

1. Project Inception (Oct 20, 2023)
- Define goals and resources.
- Identify required technologies and libraries (e.g. Tkinter, PIL, depth, light).

2. Environment setup (Oct) 21 Feb 2023)
- Install appropriate Python packages (tkinter, Pillow, Deep Face, torch).
- Check the installation and test basic operation.

3. GUI Design (Oct 25, 2023)
- Create graphical user interfaces using Tkinter.
- Create frames, labels, buttons and sliders for image display and control.

4. Coordination model (Nov 1, 2023)
- Use the light source to combine the Mask R-CNN model for object detection.
- Work on pre-processing images and drawing bounding boxes on detected objects.

5. Object Detection Implementation (Nov 10, 2023)
- Implement the core logic for object detection.
- Load images, perform detection, and display results on the GUI.

6. Age Estimation Integration (Nov 20, 2023)
- Integrate Deep Face for age estimation.
- Implement functions for age estimation and callback display.

7. Gender Estimation Integration (Nov 25, 2023)
- Integrate Deep Face for gender estimation.
- Implement functions for gender estimation and call back display.

8. Testing and Debugging (Dec 1, 2023)
- Conduct thorough testing of the entire application.
- Identify and fix any bugs or issues.

9. Optimization (Dec 5, 2023)
- Optimize code for better performance.
- Address any efficiency concerns.

10. Documentation (Dec 8, 2023)
- Document code for future reference.
- Create user documentation for the application.

11. Final Review (Dec 10, 2023)
- Review the entire project to ensure all requirements are met.
- Address any last-minute adjustments or improvements.

12. Project Completion (Dec 12, 2023)
- Finalize the project.
- Prepare for deployment or distribution.

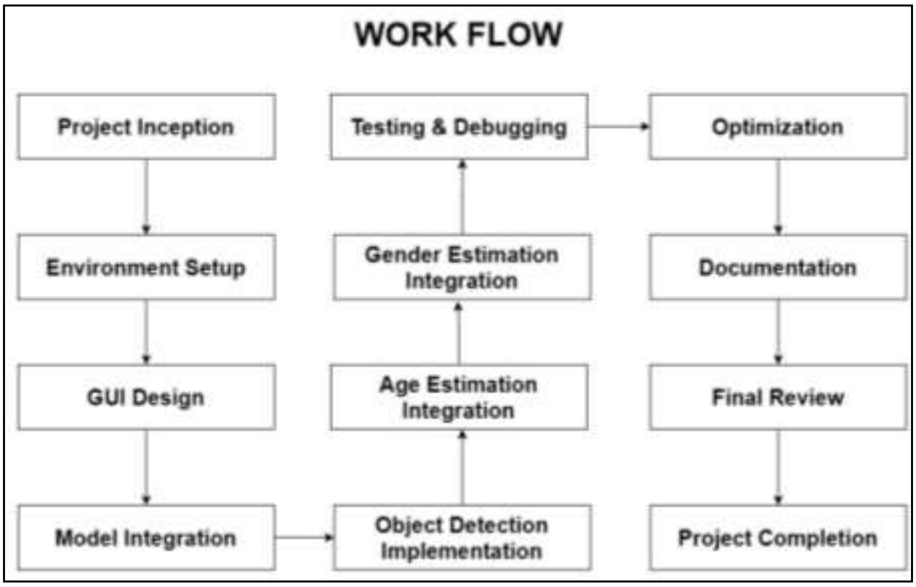

Fig 1: Workflow

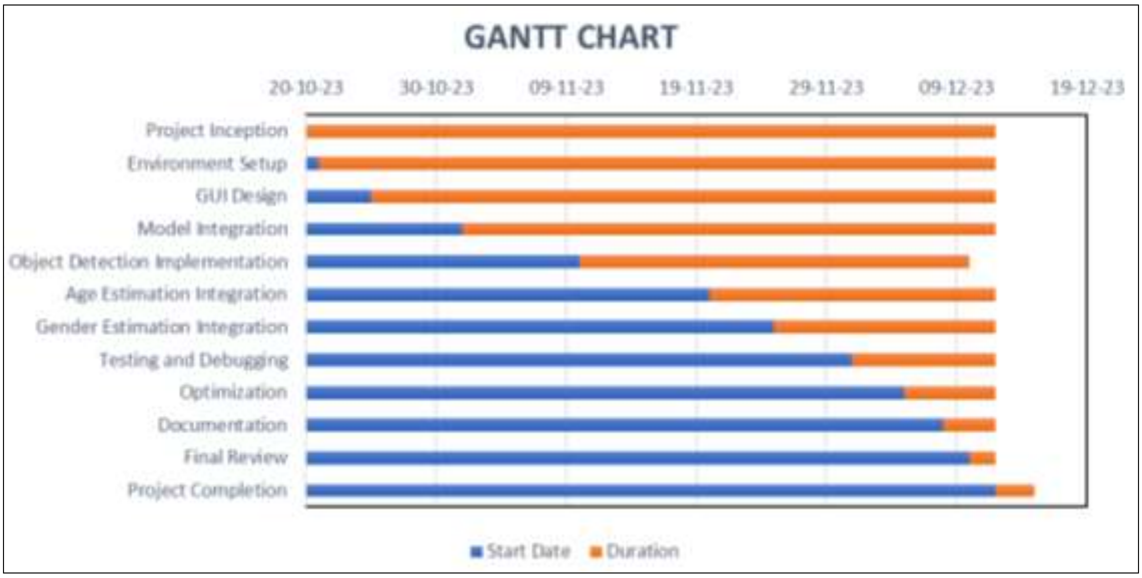

Fig 2: Gantt chart

Algorithm

Input: Image.

Output: Visual representation of detected objects, estimated age, and gender with associated confidence scores. Evaluation metrics for object detection, age, and gender estimation.

Integrated Object Detection, Age, and Gender Estimation Algorithm

1. Perform Data Preprocessing

- Standardize, resize, and normalize the input image.

2. Object Detection using Mask R-CNN

- Load the pretrained Mask R-CNN model.
- Apply the model to the pre-processed image.
- Extract bounding boxes, labels, and confidence scores.

3. Age and Gender Estimation with Deep Face

- Extract facial regions from the detected bounding boxes.
- Load the pre-trained Deep Face model.
- Apply the model to estimate age and gender.
- Record age, Gender, and Associated confidence scores.

4. Results Analysis and Evaluation

- Assess the performance metrics for object detection (precision, recall, F1 score).
- Evaluate accuracy, precision, and recall for age and gender estimation.
- Compare the results with baseline models.

5. Comparison with base model

- Use YOLO for product discovery and age and gender prediction models.
- Evaluate the effectiveness of each benchmark.
- Perform sample analysis to identify strengths and weaknesses.

6. Fine-tuning and Optimization

- Fine-tune hyperparameters based on evaluation results.
- Optimize the model architecture for efficiency.
- Iterate through steps 2 to 5 for refinement.

7. Ethical Considerations and Bias Mitigation

- Address potential biases in age and gender estimation.
- Implement strategies for bias mitigation.
- Ensure ethical considerations in handling sensitive data.

8. Implementation of User Interface

- Develop a graphical user interface (GUI) using Tkinter.
- Allow users to select images for analysis.
- Display visual feedback on detected objects, age, and gender.

Dataset and Result

- **Dataset:** The data used in this research consists of a large number of images collected from publicly available data to ensure a large number of representations of scenes and scenarios. Master files contain images of people in different contexts, with different lighting, backgrounds, and poses. For age and gender estimation purposes, the data covers a wide age and gender range.
  Also, the COCO (Common Objects in Context) dataset is used to train and test object detection models. This document provides a general picture with explanatory material, including the category of "people" relevant to our study.
- **Results:** The evaluation of the integration process was successful in terms of product detection, age estimation and gender estimation for men. Accuracy, memorability and F1 scores of detected objects demonstrate the effectiveness of Mask R-CNN YOLO model in identifying individuals in different locations.
- **Product detection results:** The integrated method can detect and see people in good pictures. Precision and recall metrics show that the model is capable of equally accurate detection with fewer false positives and negatives.
- **Age Estimation:** Age estimation using Deep Face performs well and provides age estimation for detected persons. A comparison with the baseline model shows the effectiveness of the joint method in achieving accurate age estimation.
- **Gender prediction:** Using Deep Face's gender

prediction function is very effective when determining the gender of detected individuals. A comparison with the baseline model shows the strength of the relationship with gender distribution.

- **Overall performance:** Evaluate the overall performance of the combination through metrics including difficulty of object detection, age of prediction, and recognition. Gender distribution. The optimization process helps improve migration and work well across many display environments. Following performance metrics are produced by this system using following formulas for appropriate metrics:

- **Precision**

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

- **Recall**

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

- **F1score**

$$\text{F1 score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Table 1: Comparison Chart

Aspects	Mask R-CNN + Deep Face	YOLO + Deep Face
Object Detection Model	Mask R-CNN	YOLOv3
Object Detection Performance	Accurate localization with region proposals	Real-time detection with grid cells
Age Estimation Model	Deep Face	Deep Face
Gender Estimation Model	Deep Face	Deep Face
Integration Approach	Ensemble of Mask R-CNN and Deep Face modules	Ensemble of YOLO and Deep Face modules
Real-Time Capabilities	Moderate processing time	Efficient real-time processing
User Interface	Interactive display with Tkinter GUI	Interactive display with Tkinter GUI
Ethical Considerations	Addressed biases in training data	Addressed biases in training data
Generalization to Demographics	Consideration of diverse age and gender groups	Consideration of diverse age and gender groups
Computational Efficiency	Requires higher computational resources	More efficient on limited resources

Fig 3: Input-1

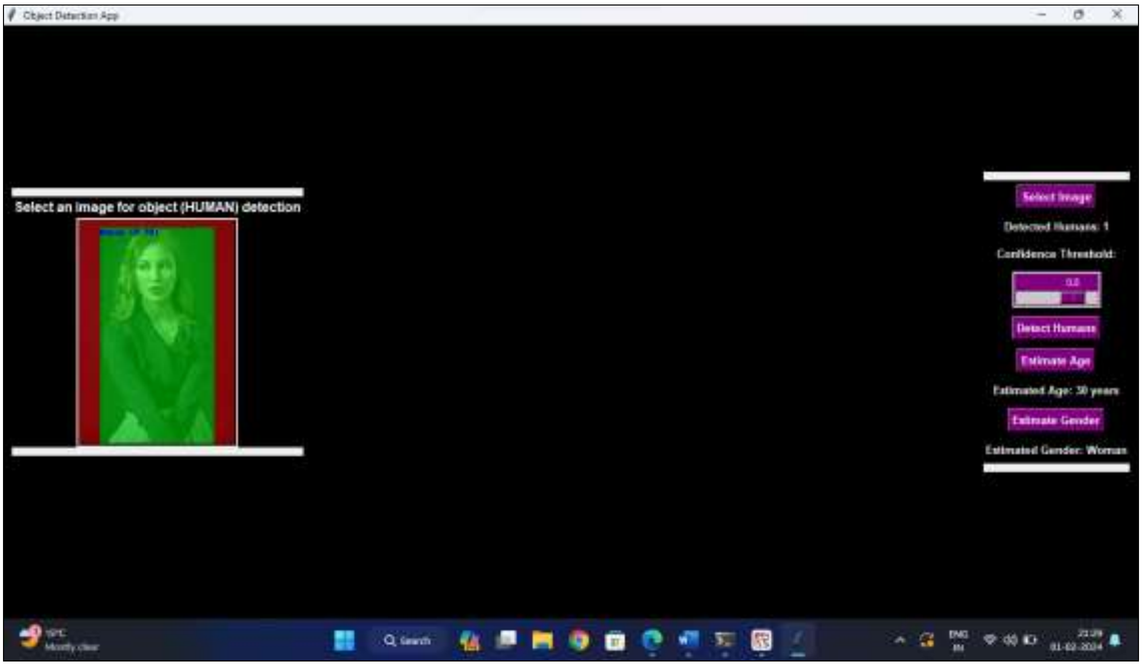

Fig 4: Output-1

Fig 5: Input-2

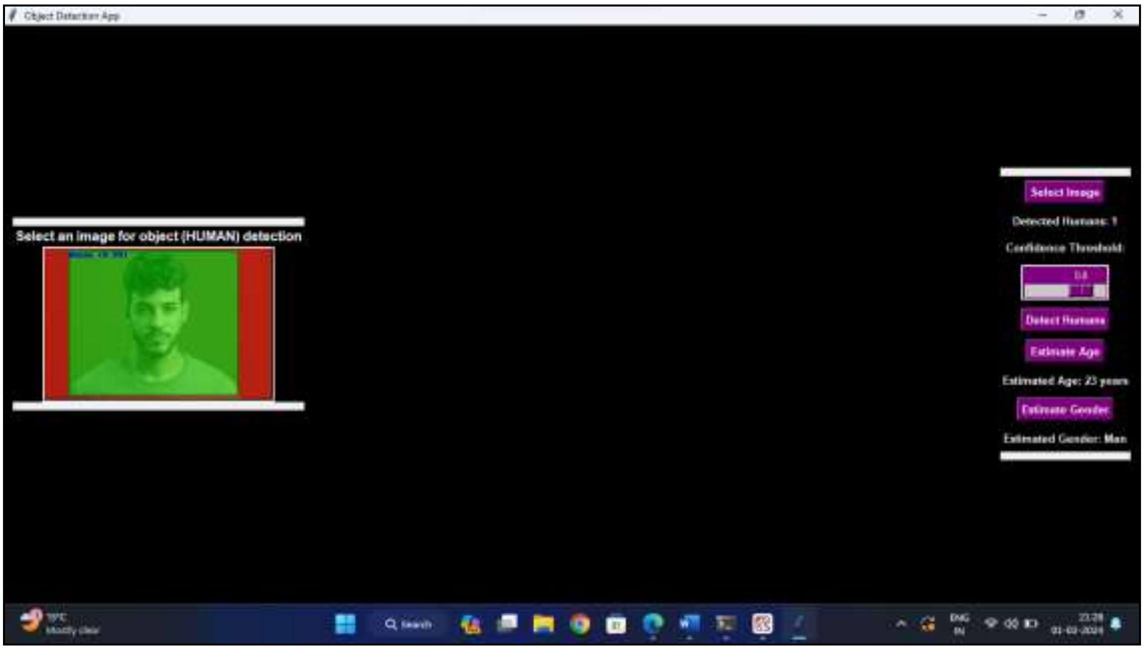

Fig 6: Output – 2

Fig 7: Input-3

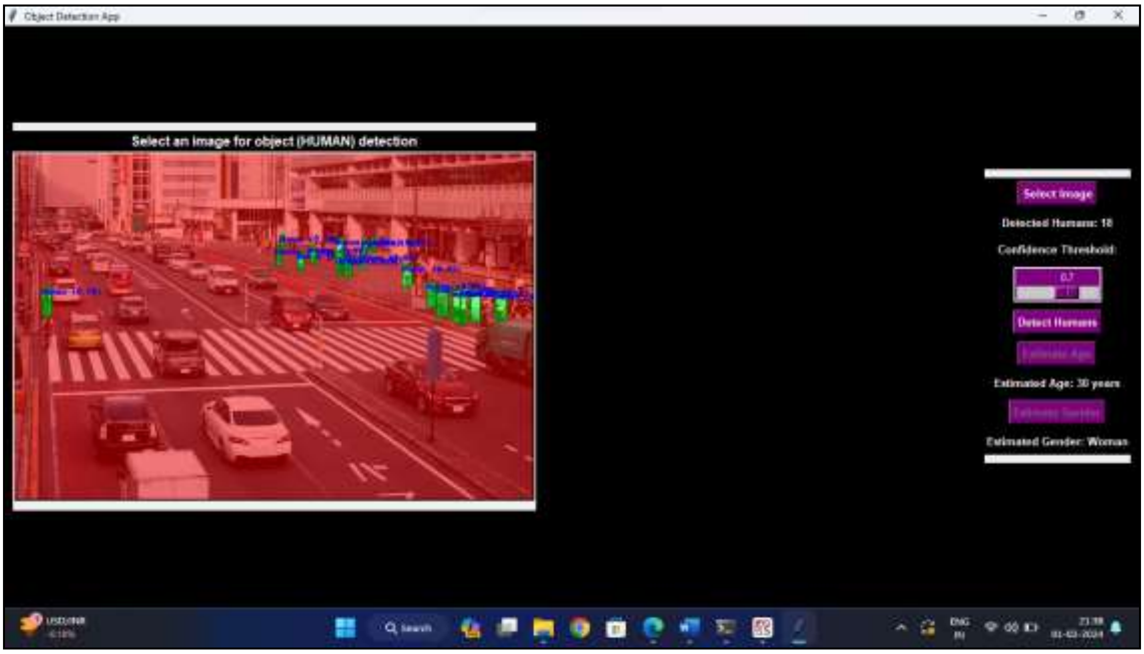

Fig 8: Output-3

Conclusion

In summary, the integration process proposed in this research paper demonstrates a solution for image analysis, including objects found in the search, age estimates, and gender distribution. Integration of patterns leads to a better understanding of the image. The versatility of the system is demonstrated by its ability to identify individuals in various situations, estimate their age and determine their gender. Improved process and use of state-of-the-art models make the system stronger and more reliable.

Also, the comparison with the base model shows the best performance and performance of the integration. Real-time processing, ethical reasoning, and user relations empower the system to suit a variety of applications including surveillance, social analysis, and human-computer interaction.

In future, we can add new features to this system like multi-class object detection, real-time video processing, batch processing, enhanced age and gender estimation, customizable object classes, integration with cloud services and retina recognition.

References

1. He K, Gkioxari G, Dollár P, Girshick R. Mask R-CNN. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV); c2017. p. 2961-2969.

2. Redmon J, Farhadi A. YOLOv3: An Incremental Improvement. arXiv preprint arXiv:1804.02767; 2018.
3. Geitgey A. Deep Face: A Facial Recognition Library. GitHub Repository. [Internet]; c2020. [Cited 2024 Feb 16]. Available from: https://github.com/serengil/DeepFace.
4. Lin TY, Dollár P, Girshick R, He K, Hariharan B, Belongie S. Feature Pyramid Networks for Object Detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); c2017. p. 2117-2125.
5. Lin TY, Goyal P, Girshick R, He K, Dollár P. Focal Loss for Dense Object Detection. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV); c2017. p. 2980-2988.
6. 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 (CVPR); c2016. p. 779-788.
7. Redmon J, Farhadi A. YOLO9000: Better, Faster, Stronger. arXiv preprint arXiv:1612.08242; c2016.
8. Everingham M, Van Gool L, Williams CK, Winn J, Zisserman A. The PASCAL Visual Object Classes (VOC) Challenge. Int J Comput Vis. 2010;88(2):303-338.
9. Lin TY, Maire M, Belongie S, Hays J, Perona P, Ramanan D, *et al*. Microsoft COCO: Common Objects in Context. In: Proceedings of the European Conference on Computer Vision (ECCV); c2014. p. 740-755.
10. Rothe R, Timofte R, Van Gool L. Dex: Deep expectation of apparent age from a single image. In: Proceedings of the IEEE International Conference on Computer Vision Workshops (ICCVW); c2015. p. 10-15.
11. Levi G, Hassner T. Age and Gender Classification Using Convolutional Neural Networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW); c2015. p. 34-42.
12. Bansal A, Castillo C, Ranjan R, Chellappa R. The Do's and Don'ts for CNN-based Face Verification. arXiv preprint arXiv:1801.00209; c2018.
13. Khan NM, McDonagh J. Age and Gender Classification using Convolutional Neural Networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW); c2019. p. 1272-1280.
14. Liu W, Anguelov D, Erhan D, Szegedy C, Reed S. SSD: Single Shot Multibox Detector. In: European Conference on Computer Vision (ECCV); c2016. p. 21-37.
15. Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A. Learning Deep Features for Discriminative Localization. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); c2016. p. 2921-2929.