Real Time Human Gender and Age Prediction Using Machine Learning

[1]Kodi Maheshbabu, [2]Kilari Charan Chandu, [3]Nalagarala Raghava Harshith, [4]Yanamadni Sucharan, [5]Yerra Vishnu Vardhan, [6] Mr. V. Sudheer

[1] UG Scholar, [2] UG Scholar, [3] UG Scholar, [4] UG Scholar,[5] UG Scholar
[6]Assistant Professor, ECE Department, NECN, Nellore

Abstract: *In recent years, significant efforts have been dedicated to the fields of age assessment and gender recognition through the analysis of human faces. Human faces reveal a wealth of information, such as gender and age, which can be deciphered through advanced pattern recognition techniques. This facet of computer vision plays a crucial role in automating the detection of faces and determining demographic details. By leveraging these technologies, systems can accurately predict age and gender, enhancing various applications from security to personalized services. Pattern recognition in computer vision is pivotal for these advancements, as it enables the automated analysis and interpretation of facial features. This technology mimics human perception, allowing machines to identify and classify different facial characteristics with high precision. The automatic detection of faces followed by the analysis to determine age and gender showcases the capability of modern artificial intelligence. Through continuous learning and improvement, these systems become increasingly accurate and reliable.This project focuses on predicting age and gender using live face captures, pushing the boundaries of real-time facial analysis. By capturing live images, the system can dynamically assess the age and gender of individuals, making it highly applicable in various real-world scenarios. This includes areas like social media, retail, and security, where real-time data processing is crucial. The ability to make instantaneous predictions ensures that the system is not only accurate but also efficient in handling large volumes of data.*

Keywords: *Human face, Pattern recognition, Automated Analysis, Facial Analysis, Intelligence, Live Face captures.*

I. INTRODUCTION

Age and gender data are critical for numerous real-world applications, including social understanding, biometrics, identity verification, video surveillance, human-machine interaction, electronic customer engagement, crowd behavior analysis, online advertising, and product recommendations. These applications rely on accurate age and gender predictions to enhance user experience, security, and operational efficiency. However, calculating age and gender from facial images is challenging due to the significant intra-class differences in people's facial features, which can be influenced by factors such as age, gender, ethnicity, and environmental conditions. These variations restrict the applicability of predictive models in real-world scenarios, making it essential to develop robust algorithms that can handle diverse and unconstrained photography.

In the past, differences in facial feature dimensions were used to estimate or classify age and gender attributes from face photos. These methods often relied on handcrafted features and traditional machine learning algorithms, which were limited by their inability to generalize well to new, unseen data. Furthermore, early systems were not designed to handle the plethora of challenges presented by unconstrained photography, such as variations in lighting, pose, and facial expressions. As a result, these models struggled to accurately predict age and gender in real-world conditions, where the diversity of facial appearances is much greater.

Additionally, previous machine learning algorithms did not fully exploit the large volumes of image samples and data available through the Internet to improve classification skills. The vast amounts of data present an opportunity to train more accurate and robust models, but traditional methods were not equipped to handle such large-scale datasets effectively. This limitation hindered the performance of age and gender prediction systems, preventing them from achieving high accuracy in diverse and complex environments. To address these challenges, more advanced machine learning techniques are required, capable of leveraging large datasets and learning from the rich variety of facial features present in the real world.

In this study, we utilize machine learning to consistently predict a person's gender and age from a single facial capture. By leveraging tools like OpenCV and Convolutional Neural Networks (CNNs), we aim to develop an algorithm that can accurately determine age and gender from live frames. CNNs, in particular, have shown great promise in automatically learning hierarchical features from raw data, making them well-suited for this task. These networks can capture subtle differences in facial features, leading to improved prediction accuracy. Moreover, data augmentation and advanced training strategies can further enhance the model's ability to generalize to new and diverse facial images.

Accurately estimating age and gender from single face frames holds significant importance in various intelligence applications, including access control, human-computer interaction, marketing intelligence, and visual surveillance. In social interactions, age and gender play pivotal roles in shaping perceptions and behaviors. Therefore, a reliable algorithm for age and gender prediction can greatly benefit these applications by providing more personalized and effective solutions. As we continue to refine our models and leverage the power of deep learning, we can expect significant advancements in the accuracy and robustness of age and gender prediction systems, ultimately enhancing their applicability in real-world scenarios.

II. RELATED WORK

Research in age and gender prediction has evolved significantly over the past decades, driven by advancements in both hardware and algorithms. Early approaches predominantly relied on handcrafted features, such as texture, shape, and local binary patterns (LBP), which were manually engineered to capture relevant facial characteristics. Traditional machine learning algorithms like Support Vector Machines (SVMs), k-Nearest Neighbors (k-NN), and Random Forests (RF) were employed to classify these features into different age and gender categories. While these methods laid the groundwork, their performance was often limited by the quality and representativeness of the handcrafted features, as well as the complexity of facial variations across different demographics.

The advent of deep learning revolutionized the field by enabling models to automatically learn hierarchical features directly from raw data, eliminating the need for manual feature extraction. Convolutional Neural Networks (CNNs), in particular, have become the cornerstone of modern age and gender prediction systems. CNNs leverage multiple layers of convolutions to detect low-level features like edges and textures in the initial layers, progressively learning more complex patterns and structures in deeper layers. This hierarchical feature learning is crucial for capturing the subtle and diverse characteristics present in facial images, significantly improving prediction accuracy and robustness compared to traditional methods.

Various CNN architectures have been proposed to further enhance performance in age and gender prediction. Deeper networks, such as VGGNet, ResNet, and InceptionNet, have demonstrated superior capabilities in learning detailed and discriminative features due to their increased depth and sophisticated design. Techniques like data augmentation, including random cropping, rotation, and color jittering, are employed during training to increase the diversity of the training dataset, helping models generalize better to unseen data. Additionally, the design of loss functions has been critical; for instance, using weighted loss functions to address class imbalance or employing multi-task learning frameworks to jointly predict age and gender has shown to enhance performance by leveraging shared representations.

Moreover, recent research has emphasized the importance of large-scale datasets and robust evaluation protocols. Publicly available datasets such as IMDB-WIKI, Adience, and UTKFace have facilitated the training and benchmarking of age and gender prediction models. Transfer learning and pre-training on large datasets have also proven effective, allowing models to leverage pre-learned features and adapt them to specific tasks with relatively smaller datasets. Despite these advancements, challenges remain, such as handling occlusions, varying lighting conditions, and extreme facial expressions. Future research is likely to focus on addressing these challenges, further refining model architectures, and exploring novel training strategies to continue advancing the state-of-the-art in age and gender prediction.

III. METHODOLOGY

A. Data Collection

We use the UTKFace dataset, a publicly available and widely recognized collection of facial images labeled with both age and gender. This dataset includes over 26,000 images of faces captured under various conditions, such as different lighting, backgrounds, and poses, making it a robust choice for training and evaluating our model. The diversity in the dataset, spanning age ranges from infants to the elderly and including a balanced gender distribution, ensures that our model is exposed to a wide variety of facial characteristics.

Before feeding the images into our convolutional neural network (CNN), we undertake a meticulous preprocessing stage to ensure uniformity in image size and quality, which is crucial for consistent performance. Each image is resized to a standard dimension of 224x224 pixels. This resizing is necessary to standardize the input size, as CNNs require uniform input dimensions to function correctly. Additionally, we convert all images to grayscale to simplify the computational complexity and focus on the most critical features for age and gender prediction, reducing the noise that might come from color variations.

To further enhance the model's generalizability, we apply a range of data augmentation techniques. These techniques artificially expand the training dataset by creating altered versions of the existing images, thereby exposing the model to a broader variety of scenarios. Specifically, we employ random rotations within a range of ±15 degrees to account for variations in head orientation, and random scaling to simulate different distances from the camera. Horizontal flipping is used to address the asymmetry in facial features, ensuring that the model does not become biased towards any particular facial orientation. Additionally, random cropping is performed to focus on different parts of the face, which helps the model learn to identify key features even when they are not centrally located. Brightness and contrast adjustments are also applied to mimic varying lighting conditions, making the model more robust to different environmental settings.

B. Model Architecture

Our system utilizes pre-trained CNN models for both face detection and age and gender prediction.

1) Face Detection: We use a pre-trained face detection model based on the Single Shot MultiBox Detector (SSD) framework. This model is loaded using OpenCV's deep neural network (dnn) module and is capable of detecting faces in real-time video streams with high accuracy.

2) Age and Gender Prediction: For age and gender prediction, we use two separate CNN models, both based on the Caffe framework. The age prediction model is pre-trained on the UTKFace dataset, while the gender prediction model is trained on the IMDB-WIKI dataset. Both models are fine-tuned to achieve optimal performance.

C. System Implementation

The system is implemented in Python using OpenCV for video processing and Tkinter for the graphical user interface (GUI).

1) Face Detection: The face detection component uses the OpenCV dnn module to load and run the SSD model. Each frame from the video stream is processed to detect faces, which are then cropped and passed to the age and gender prediction models.

2) Age and Gender Prediction: The cropped face images are resized to 227x227 pixels and normalized using the mean values specific to the pre-trained models. These pre-processed images are then fed into the age and gender prediction models. The outputs are the predicted age range and gender, which are overlaid on the original video frame.

3) User Interface: The system includes a user-friendly interface implemented with Tkinter. The GUI allows users to start and stop the video stream, view the predictions in real-time, and terminate the program gracefully.

D. Training and Evaluation

The models are trained using the UTKFace dataset, a comprehensive and publicly available collection of over 20,000 facial images with annotations for age, gender, and ethnicity. This dataset spans a wide range of ages, from 0 to 100 years, and includes a balanced representation of genders, making it an excellent choice for training our models. The diversity in the UTKFace dataset helps in creating a robust model capable of generalizing well across different demographic groups.

1) Data Augmentation

To prevent overfitting and improve the generalizability of our models, we employ various data augmentation techniques during the training process. These techniques include:

- **Rotation:** Random rotations within a range of ±15 degrees to simulate different head orientations and account for minor misalignments.
- **Scaling**: Random scaling to mimic faces at different distances from the camera, ensuring the model can handle variations in image size.
- **Horizontal Flipping:** Applying horizontal flips to address left-right asymmetry in facial features and improve the model's robustness to orientation changes.

- **Brightness and Contrast Adjustments:** Randomly altering the brightness and contrast to simulate different lighting conditions, enhancing the model's ability to generalize to varying illumination.
- **Random Cropping:** Performing random cropping to expose the model to different parts of the face, helping it learn to identify critical features even when they are not centrally located.

These augmentations increase the effective size of the training dataset, exposing the model to a wider variety of scenarios and reducing the risk of overfitting.

2) Model Training

We employ transfer learning by fine-tuning pre-trained CNN models, specifically those trained on large-scale face datasets. Fine-tuning these models involves training them on the UTKFace dataset with a reduced learning rate to adapt the learned features to our specific task of age and gender prediction. The training process involves several key steps:

- **Initialization:** The pre-trained models are initialized with weights obtained from training on large-scale face recognition datasets. These weights provide a strong starting point, capturing essential facial features.
- **Feature Extraction:** The initial layers of the pre-trained models are used as feature extractors. These layers capture low-level features such as edges and textures, which are crucial for accurate face analysis.
- **Fine-Tuning:** The later layers of the models are fine-tuned using the UTKFace dataset. Fine-tuning involves training these layers with a smaller learning rate to adapt the high-level features to the specific tasks of age and gender prediction.
- **Optimization:** We use the Adam optimizer with an initial learning rate of 0.001. The learning rate is dynamically adjusted using a learning rate scheduler, which reduces the learning rate when the validation performance plateaus. This helps in fine-tuning the model towards the end of training, ensuring better convergence and performance.

IV. IMPLEMENTATION

Age and Gender estimation from single face frames is a valuable task with numerous applications in intelligence systems. By leveraging machine learning techniques such as OpenCV. we can develop accurate and efficient algorithms capable of extracting demographic information from facial images in real-time. The proposed methodology encompasses data collection, preprocessing, model training, evaluation, and integration, offering a comprehensive approach to address this important problem.

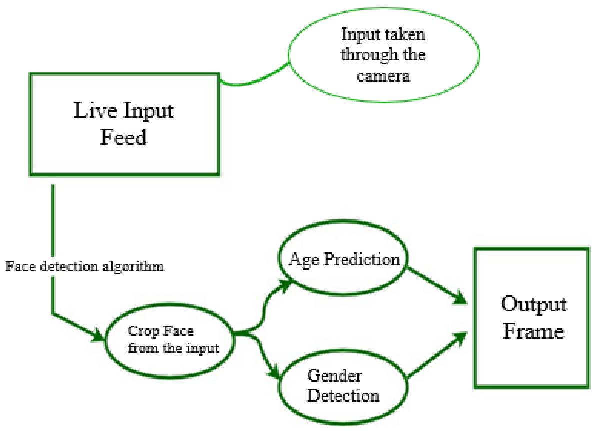

Fig. 1. General flow representation

The flowchart provided illustrates the process of human age and gender prediction using a live input feed from a camera. The process begins with capturing live input through the camera, which continuously streams video frames into the system. This live feed serves as the input for the subsequent steps in the prediction process. The first crucial step is face detection, where an algorithm is applied to identify and locate faces within each frame. This step is essential because the subsequent predictions for age and gender are based on analyzing the facial features of the detected faces.

Once the face detection algorithm identifies a face, the system crops the face from the input frame to focus on the relevant region for analysis. This cropped face is then processed by two separate prediction models. One model is dedicated to age prediction, which estimates the age range of the individual based on the facial features extracted from the cropped image. The age prediction model leverages pre-trained convolutional neural networks (CNNs) fine-tuned for this specific task, ensuring accurate age estimation across various age groups.

Simultaneously, the cropped face is also fed into a gender detection model, which determines whether the individual is male or female. Like the age prediction model, the gender detection model utilizes CNNs trained on large datasets to recognize gender-specific features accurately. Both the age and gender predictions are then combined and superimposed onto the original input frame, resulting in an output frame that displays the predicted age and gender of the detected face. This output frame is subsequently presented to the user, providing real-time feedback on the age and gender predictions derived from the live input feed.

V. RESULTS

The image provided demonstrates a real-time age and gender prediction system at work, showcasing its capability to analyze live video input and produce immediate results. In this particular scene, three individuals are facing the camera, and the system has successfully identified and annotated their faces with the predicted age and gender.

Each face is enclosed within a distinct green bounding box, which is the output of the face detection algorithm. This initial step is crucial, as it isolates the facial regions from the background, enabling accurate analysis of facial features for subsequent predictions.

The bounding boxes around the faces shown in figure 2 are supplemented with text annotations that display the gender and age range of each detected face. In this instance, all three individuals are classified as male, and their ages are estimated to fall within the 21 to 32 years range. These annotations are prominently displayed in yellow text against a green background, ensuring high visibility and readability. The placement of these annotations directly on the live video feed allows for an intuitive and immediate understanding of the predictions being made by the system.

The system's design incorporates color-coded text annotations, which serve multiple purposes. Firstly, the use of contrasting colors like yellow on green helps in maintaining clarity, making the text easily discernible against various backgrounds that might appear in the video feed.

This is particularly important in real-time applications where quick and clear communication of information is essential. Secondly, the choice of colors and their placement enhances the user experience by making the interface more engaging and user-friendly. Users can quickly glance at the screen and grasp the relevant details without any confusion or difficulty.

This real-time prediction system is interactive, evidenced by the hint at the top of the window that reads, "Hint: Press e to exit." This suggests that the system is designed with user convenience in mind, offering a straightforward way to terminate the video feed when necessary. The inclusion of this feature highlights the importance of user-friendly design in real-time applications.

An interactive feature like this is essential for applications requiring flexibility and user control. It is particularly useful during demonstrations, user testing, or dynamic environments where the system might need to be paused or stopped frequently. Providing users with an easy exit option ensures that the system can be managed efficiently, enhancing the overall user experience and adaptability to various situations.

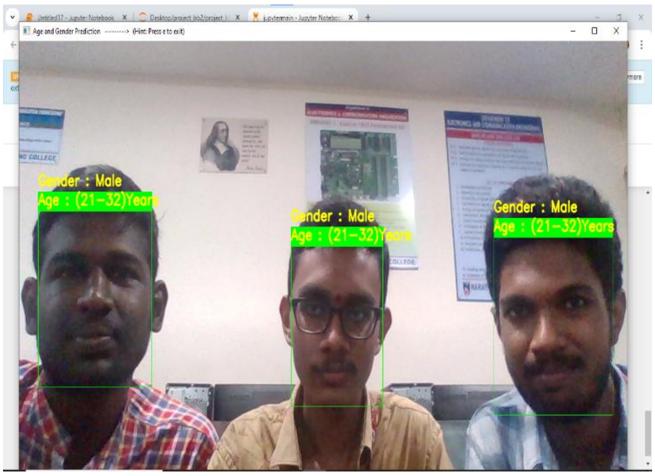

Fig. 2. Final Output

The applications of this technology are vast and varied. In security and surveillance, for instance, the ability to quickly and accurately predict the age and gender of individuals can be critical for monitoring and identifying potential threats or ensuring appropriate responses to different demographic groups. In retail environments, understanding the demographics of customers in real-time can help businesses tailor their marketing strategies, provide personalized services, and improve customer satisfaction. Similarly, in public spaces like airports, train stations, or malls, this technology can assist in crowd management and enhance public safety by providing valuable demographic insights.

Overall, the image exemplifies the practical utility of advanced machine learning models integrated with real-time video processing. The accurate face detection, followed by reliable age and gender predictions, underscores the system's effectiveness and potential impact across various domains. By providing immediate, actionable insights directly within the live video feed, this technology bridges the gap between complex algorithmic predictions and user-friendly applications, paving the way for innovative solutions in numerous real-world scenarios..

VI. REFERENCES

[1] Pattern Recognition and Machine Learning by Christopher M. Bishop.

[2] D. Chen, G. Hua, F. Wen and J. Sun, "Supervised Transformer Network for Efficient Face Detection", European Conference on Computer Vision, pp. 122-138, 2016.

[3] Deep Learning by Ian Goodfellow, Yoshua Bengio, and Aaron Courville.

[4] Z. Li, X. Tang, X. Wu, J. Liu and R. He, "Progressively Refined Face Detection Through Semantics-Enriched Representation Learning", IEEE Transactions on Information Forensics and Security, vol. 15, pp. 1394-1406, 2019.

[5] Machine Learning: A Probabilistic Perspective by Kevin P. Murphy.

[6] D. Li, Y. Li and W. Ji, "Gender Identification via Reposting Behaviors In Social Media", IEEE Access, vol. 6, pp. 2879-2888, 2017.

[7] Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow by Aurélien.

[8] S. Amilia, M. D. Sulistiyo and R. N. Mayawati, "Face Image-Based Gender Recognition using Complex-Valued Neural Network", 2015 3rd International Conference on Information and Communication Technology (ICoICT), pp. 201-206, 2015.

[9] E. S. El-Alfy and A. G. Binsaadoon, "Automated Gait-Based Gender Identification Using Fuzzy Local Binary Patterns with Tuned Parameters", Journal of Ambient Intelligence and Humanized Computing, vol. 10, no. 7, pp. 2495-2504, 2019.

[10] Computer Vision: Algorithms and Applications by Richard Szeliski.

[11] M. H. Rahman, M. A. Bashar, F. H. M. Rafi, T. Rahman and A. F. Mitul, "An Automatic Face Detection and Gender Identification from Color Images Using Logistic Regression", 2013 International Conference on Informatics Electronics and Vision (ICIEV), pp. 1-6, 2013.

[12] Pattern Classification by Richard O. Duda, Peter E. Hart, and David G. Stork.

[13] J. Schmidhuber, "Deep learning in neural networks: An overview", Neural Networks, vol. 61, pp. 85-117, Jan. 2015.

[14] Python Machine Learning by Sebastian Raschka and Vahid Mirjalili.

[15] A. Krizhevsky, I. Sutskever and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks", Adv. Neural Inf. Process. Syst., pp. 1-9, 2012.

[16] Machine Learning Yearning by Andrew Ng.

[17] M. Abadi et al., "TensorFlow: A System for Large-Scale Machine Learning TensorFlow: A system for large-scale machine learning", 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI16), pp. 265-284, 2016.

[18] X. Yang, P. Roop, H. Pearce and J. W. Ro, "A compositional approach using Keras for neural networks in real-time systems," 2020 Design, Automation & Test in Europe Conference & Exhibition (DATE), Grenoble, France, 2020, pp. 1109-1114, doi: 10.23919/DATE48585.2020.9116371.

[19] J. Redmon, "Darknet: Open source neural networks in c", pp. 2013-2016, [online] Available: http://pjreddie.com/darknet/.

[20] R. E. Schapire, "The boosting approach to machine learning: An overview" in Nonlinear estimation and classification, New York:Springer, pp. 149-171, 2003.

[21] Culjak, D. Abram, T. Pribanic, H. Dzapo and M. Cifrek, "A brief introduction to OpenCV," 2012 Proceedings of the 35th International Convention MIPRO, Opatija, Croatia, 2012, pp. 1725-1730.

[22] A. Sharma, J. Pathak, M. Prakash and J. N. Singh, "Object Detection using OpenCV and Python," 2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), Greater Noida, India, 2021, pp. 501-505, doi: 10.1109/ICAC3N53548.2021.9725638.

[23] M. Marengoni and D. Stringhini, "High Level Computer Vision Using OpenCV," 2011 24th SIBGRAPI Conference on Graphics, Patterns, and Images Tutorials, Alagoas, Brazil, 2011, pp. 11-24, doi: 10.1109/SIBGRAPI-T.2011.11.