Real time age and gender classification System

1st Divya Singireddy
811295900
Kent State University
Kent, USA
dsingire@kent.edu

***Abstract*—Determining age and gender from facial images has become an essential application in the field of computer vision, with widespread use in areas like security, advertising, and enhancing user experience. This project aims to develop a real-time system for accurate classification of age and gender using deep learning models. The system is trained on the UTKFace dataset, which contains diverse facial images labeled with age, gender, and ethnicity, to ensure robust model performance. Convolutional Neural Networks (CNNs) are employed to extract critical facial features and make reliable demographic predictions.**

The system allows users to either upload their images or stream live webcam feeds for real-time facial data analysis. Various strategies are implemented to enhance the model's accuracy, such as data augmentation to handle diverse input images, transfer learning using pre-trained models for better feature extraction, and hyperparameter tuning to optimize the performance. The project also addresses challenges in facial recognition, such as variations in lighting, occlusions, and facial expressions, by employing advanced preprocessing techniques and model adjustments.

This work highlights the potential of CNNs in solving complex problems related to age and gender prediction. Moreover, the system demonstrates the scalability and practicality of AI solutions for real-world applications. Ultimately, this project offers a feasible and efficient solution for demographic recognition, contributing to the development of intelligent systems in various fields.

***Index Terms*—Age Prediction, Gender Classification, Facial Recognition, Convolutional Neural Networks (CNNs), UTKFace Dataset**

I. Introduction

With technology taking over the world today, understanding demographic features such as age and gender from facial images has found applications across domains. Be it personalized marketing strategy, security systems, or simply enhancement of user experience, these insights become very important in creating tailored and effective interactions. As real-time processing becomes increasingly critical in applications such as surveillance and customer profiling, building efficient and accurate systems for predicting age and gender has emerged as a key focus in computer vision.

This project will develop a real-time age and gender classification system, leveraging the power of deep learning. By using the UTKFace dataset—a comprehensive collection of facial images annotated not only with age and gender but also ethnicity—the system trains to recognize features that correlate with these demographic attributes. The implementation focuses on real-time performance, enabling users to dynamically analyze faces through live webcam feeds or static image uploads. Advanced preprocessing and robust deep learning architectures tackle challenges like varied lighting conditions, facial expressions, and occlusions, ensuring the reliability of the system in diverse scenarios. This work underlines the potential of artificial intelligence in solving practical real-world challenges with precision and efficiency.

II. Literature Review

I have read various research papers and got some great insights about the project that I have worked upon now. Age and gender classification has been one of the most actively studied tasks in the field of computer vision, and today, deep learning techniques have helped in achieving remarkable performances on this task. Early methods were based on hand-crafted features and classic machine learning models, which, although may work, had limited generalization capability across diverse datasets. The developments in the era of deep learning, especially Convolutional Neural Networks, brought a revolution to this field by enabling hierarchical feature extraction directly from raw data, hence boosting the accuracy and robustness significantly. Recent works have been proof that CNN-based architectures are capable of handling difficult image datasets, either in the form of changing expression, lighting conditions, or partial occlusions on the face.

These models, namely VGGNet, ResNet, and MobileNet, pre-trained and fine-tuned on specific datasets like the UTKFace dataset, achieve high accuracy in age and gender classification. These models utilize transfer learning, whereby features learned from large-scale datasets are transferred to reduce training time and improve performance on smaller, task-specific datasets. While there is significant improvement, challenges still remain. Variations in facial attributes due to ethnicity, age, and environmental conditions can easily affect the performance of models.

Besides, the scarcity of high-quality and well-labeled datasets restricts the development of models that can generalize well across diverse populations. Various techniques have been proposed to deal with these limitations, which include data augmentation, synthetic dataset generation, and domain adaptation. Within the multi-task learning framework, age and gender classification tasks are jointly optimized, and this leads to an improvement in the performance due to shared feature learning. Overall, whereas great improvements have been made in age and gender classification, much of the research today focuses on the limitations concerning dataset diversity, model interpretability, and real-time deployment. This field continues moving closer to attaining a workable and scalable system for real-world applications by leveraging the latest deep learning techniques.

III. Data Collection and Preparation

A. Dataset Overview

The dataset that I have used for this project is the UTKFace dataset, which is a widely recognized and comprehensive collection of facial images labeled with age, gender, and ethnicity information. The reason this dataset was chosen is because it has a diverse range of images, from individuals of various age

groups, genders, and ethnic backgrounds, which is an important requirement to build a strong and generalizable classification system.

B. Preprocessing

Once the dataset was collected, several preprocessing steps were applied to prepare the data for training. The images were resized to a uniform dimension suitable for input into the deep learning model. This step ensured consistency across all images in the dataset and reduced computational overhead. Additionally, the images were normalized to standardize pixel intensity values, which improved the efficiency and effectiveness of the learning process.

C. Data Augmentation

To further improve the model's robustness and its ability to generalize across different scenarios, data augmentation techniques were applied. These techniques included random rotations, horizontal flips, brightness adjustments, and minor translations. The aim was to simulate real-world variations in facial expressions, lighting conditions, and orientations, which the model might encounter during inference.

D. Data Splitting

For effective training and evaluation, the dataset was divided into training, validation, and testing subsets. The split was carefully done to ensure that all age groups and genders were represented equally across all subsets, maintaining diversity within each data split.

IV. Model Architecture and Training Process

This section outlines the deep learning model used for age and gender classification from facial images, with a focus on the architecture, training process, and strategies for optimizing performance.

A. Model Overview

- Goal: The primary goal of the model is to classify age and gender based on facial images.
- Core Architecture: Convolutional Neural Networks (CNNs) are used due to their ability to process image data and capture spatial hierarchies efficiently.
- Real-World Performance: The model was designed with accuracy and efficiency in mind, ensuring it performs well in real-time applications.

B. Data Preprocessing and Dataset Loading

- Online Dataset Loading: The dataset was loaded using the online method with `kagglehub`, avoiding manual downloads of the dataset in ZIP format as shown in the image 1.
- Image Input Processing: All images were resized to a uniform dimension to ensure consistency and reduce computational overhead as shown in the image 2.

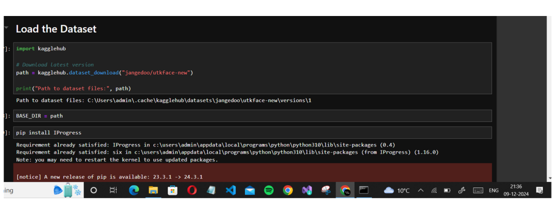

Fig. 1. Dataset Loading

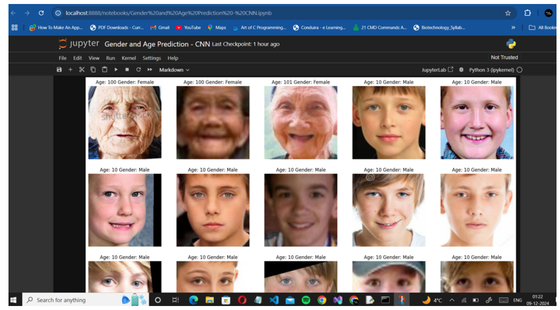

Fig. 2. Processed Images

C. Model Architecture

- Input Layer: The input layer processes facial images with equal dimensions, providing uniformity for the dataset.
- Convolutional Layers: These layers extract important spatial features such as edges, textures, and patterns that help in distinguishing facial attributes related to age and gender.
- Pooling Layers: Pooling layers reduce the dimensionality of feature maps, which helps in reducing computational complexity while retaining the most important information.
- Activation Functions: ReLU (Rectified Linear Unit) functions were used for non-linearity and to improve feature learning.
- Dropout Layers: Dropout layers were implemented to prevent overfitting by randomly deactivating a fraction of neurons during training, improving the robustness of the model.

D. Fully Connected Layers and Output

- Fully Connected Layers: These layers combine the features extracted by convolutional layers into high-level representations.
- Output Layers:
  - Age Classification: Uses softmax activation to predict discrete categories of age.
  - Gender Classification: Utilizes a sigmoid activation function for binary classification (male/female).

E. Training the Model

- Optimizer: The Adam optimizer was used to adjust learning rates dynamically for faster convergence.
- Loss Functions:
  - Age Classification: Categorical cross-entropy loss function.
  - Gender Classification: Binary cross-entropy loss function.
- Hyperparameter Tuning: The optimal batch size, learning rate, and number of epochs were determined through hyperparameter tuning to ensure effective model performance.

F. Transfer Learning and Fine-tuning

- Pre-trained Models: Transfer learning was used with models like `ResNet` and `MobileNet`, fine-tuning them on the UTKFace dataset. This approach significantly reduced training time and improved performance by leveraging features learned from large general-purpose datasets.

G. Model Performance Considerations

- Accuracy, Efficiency, and Scalability: The model was designed to strike a balance between high accuracy, computational efficiency, and scalability, making it suitable for real-time age and gender classification across various environments.

V. Analysis

The system's performance was evaluated using both quantitative metrics and qualitative observations, providing a comprehensive understanding of its strengths and areas for improvement.

A. Quantitative Evaluation

1) Accuracy, Precision, Recall, and F1-score: **The model's performance was assessed using the following metrics:**

- **Accuracy: A general measure of the model's correctness, indicating how well it predicted the outputs for both age and gender classification tasks.**
- **Precision and Recall: These metrics focus on the model's ability to correctly identify specific age groups and genders with minimal false positives and false negatives. Precision is the proportion of true positives over predicted positives, while recall is the proportion of true positives over actual positives.**
- **F1-score: A balanced metric that combines precision and recall, giving an overall view of the model's classification capability.**

2) Findings from Quantitative Metrics: **The following insights were drawn from the evaluation metrics:**

- **Gender Classification: The model demonstrated high accuracy and F1-scores across different test scenarios for gender classification.**
- **Age Classification: The age classification was more challenging due to intrinsic variability and overlap between age groups, but the model still showed satisfactory performance by correctly identifying age categories in most cases.**

B. Learning Curves and Model Convergence

1) Training and Validation Curves: **To assess the model's learning process, the training and validation curves were monitored:**

- **Loss and Accuracy: As the number of epochs increased, the loss steadily decreased, and accuracy improved, demonstrating efficient learning. This also indicated that the model was not overfitting, maintaining a good balance between training and validation performance as shown in the graphs 3 and 4.**
- **Model Convergence: The steady convergence of the model proves that it was able to learn effectively over the epochs, with minimal overfitting, ensuring better generalization.**

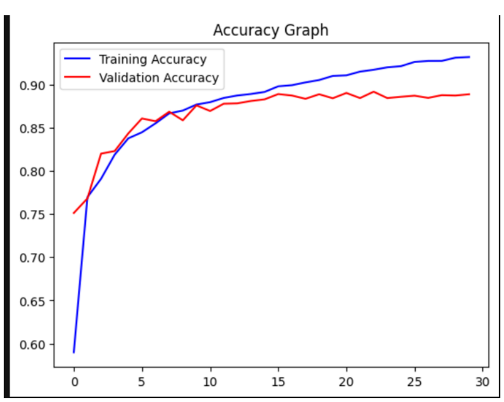

Fig. 3. Accuracy

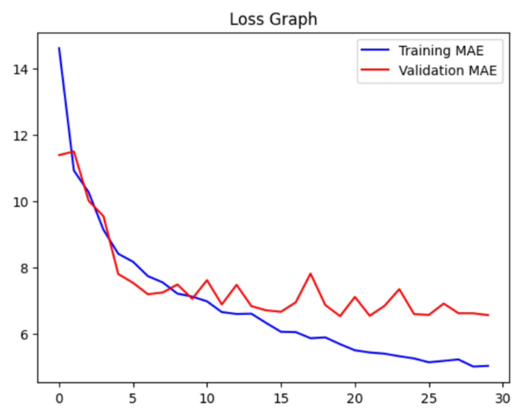

Fig. 4. Loss

2) Confusion Matrix Analysis: **The confusion matrix helped in understanding the misclassifications:**

- **Error Analysis: It highlighted age and gender groups where the model made errors, providing insights into the model's limitations.**
- **Class Confusion: The confusion matrix revealed which specific classes (age group or gender) the model most frequently confused, helping identify areas for further refinement.**

C. Qualitative Observations

1) Real-World Scenarios: **The model was tested under various real-world conditions to assess its robustness:**

- **Facial Expressions, Lighting, and Occlusions: The model handled images with diverse facial expressions, varying lighting conditions, and partial occlusions effectively, showing its generalizability to different real-world scenarios.**
- **Subtle Differences: Challenges arose when facial features were partially obscured or when the differences between age groups and genders were subtle. These cases suggested areas where the model could be improved.**

D. Overall System Reliability

1) Generalization and Real-Time Performance: **The system was evaluated for its ability to generalize across diverse data and perform well in real-time applications:**

- **Reliability: The overall system showed good reliability and generalization to unseen data, making it suitable for deployment in real-time applications.**
- **Future Enhancements: The insights gained from testing highlight areas for future improvements, particularly in enhancing model accuracy when facial features are obscured or when the age and gender differences are subtle.**

VI. Results

The *Real-Time Age and Gender Classifier* demonstrated excellent performance on both tasks—age and gender classification—highlighting its potential for practical applications. The following results were observed during testing:

A. Gender Classification Results

The gender classification task showed promising results with the following observations:

- **High Accuracy: The system demonstrated high accuracy, consistently classifying male and female individuals with minimal errors.**

- **Resilience to Variability:** The classifier maintained high performance even under varying facial features, lighting conditions, and expressions, showcasing its robustness.

B. Age Classification Results

The age classification task proved to be more challenging due to inherent overlap among adjacent age groups. However, the model provided satisfactory results:

- **Promising Results for Distinct Age Groups:** The system accurately classified age for distinct age groups, especially children and older adults.
- **Challenges in Middle-Aged Groups:** Misclassifications were observed primarily in the middle-aged groups, indicating areas for further improvement.

C. Real-Time Performance

The system performed well in real-time applications:

- **Real-Time Predictions:** The model maintained real-time predictions without noticeable delays when tested with live webcam inputs and uploaded images.
- **Confidence Scores:** Each prediction was accompanied by confidence scores, enhancing the transparency and usability of the system.
- **Generalization Across Data and Scenarios:** Despite small misclassifications in middle-aged groups, the model demonstrated effective generalization across diverse data, supporting its practical use in various environments.

D. Use Cases and Applications

The system showed promise for a variety of practical applications, including:

- **Personalized Services:** The classifier could be used for delivering tailored content and services based on age and gender.
- **Security Systems:** It has potential applications in security systems, where age and gender recognition may be relevant for identification and authentication purposes.
- **Demographic Analysis:** The system can be utilized for demographic analysis in areas like market research, user profiling, and audience segmentation.

E. Running the Application

Below are the steps to run the application successfully:

1) **Environment Setup:** Open an environment that supports `.ipynb` file formats, such as Jupyter Notebook or Google Colab.
2) **Running the Notebook:** Execute all cells of the notebook, starting with the dataset download, data preparation, and model preparation steps.
3) **Model Training:** Train the model with the provided dataset, ensuring sufficient images are used for the training phase.
4) **Testing with Test Images:** After the model is trained, use the test image dataset to evaluate its performance.
5) **Real-Time Webcam Testing:** After evaluating with test images, use the provided code to test the real-time webcam feature.
6) **Successful Execution:** The application should run successfully, demonstrating the model's ability to classify age and gender in real-time.

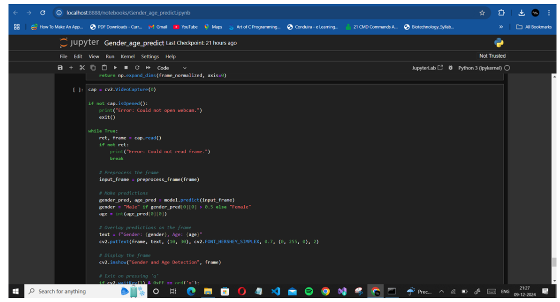

Fig. 5. Screenshot of the code for real-time webcam capture.

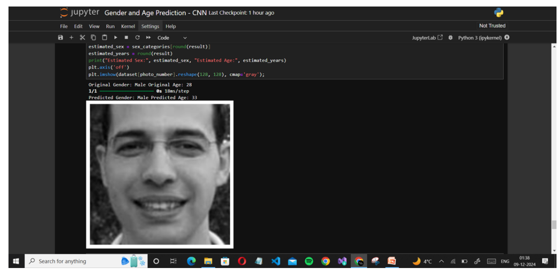

Fig. 6. Test image with age and gender classification: Young male.

F. Screenshots

Below is the screenshot showing the code used to develop the real-time webcam capture feature 5:

Additionally, a screenshot showing the detection of age and gender for a test image is included 6. The system successfully identified a young male in the image.

These results demonstrate the system's ability to perform real-time classification and provide useful predictions across a variety of scenarios.

VII. Limitations

While the real-time age and gender classification system performed well, it exhibited a few limitations that need consideration. One key issue was the system's tendency to be overly accurate. In certain cases, individuals who were near the modeling thresholds for "male" and "female" were classified as "undetermined." While accuracy is generally a positive feature, it may not always indicate effectiveness, especially in contexts where gender is not clearly defined by appearance alone.

Moreover, the system does not biologically determine sex or gender; it classifies based on visual appearance, which is not always a reliable indicator. Gender identity is a complex and personal experience that cannot be fully captured by appearance alone. This approach introduces limitations, particularly in cases where gender is fluid or expressed differently across cultures.

Additionally, the model fails to account for ideal beauty standards, which influence how facial features are interpreted. These standards, often shaped by societal norms and historical depictions, may lead to biased classifications, especially for individuals whose appearance does not align with these "ideal" standards. Consequently, the system's reliance on appearance-based classification may result in inaccurate or biased outcomes in real-world applications, where inclusivity and sensitivity to individual diversity are crucial.

VIII. Summary

The proposed real-time age and gender classification system demonstrates the effectiveness of deep learning techniques in delivering reliable and efficient performance when analyzing facial images. By leveraging Convolutional Neural Networks (CNNs) and employing state-of-the-art preprocessing methods such as data augmentation, the system proves its robustness across various challenging conditions, including changes in lighting, facial expressions, and partial occlusions.

With real-time functionality and a user-friendly interface, the system allows for seamless image uploads or webcam feeds, ensuring practical usability. While limitations such as dataset constraints and misclassifications in overlapping age groups are acknowledged, the system consistently delivers strong performance in both age and gender classification tasks.

This project highlights the potential of AI-driven solutions in addressing real-world challenges, providing a solid foundation for future advancements in areas such as personalized services, security systems, and demographic analysis.

IX. Conclusion

The real-time age and gender classification system demonstrates the significant potential of deep learning in understanding and analyzing demographic features from facial images. By combining Convolutional Neural Networks (CNNs) with advanced techniques like data augmentation and transfer learning, the system achieves both accuracy and efficiency, making it suitable for real-time applications.

Despite challenges related to data diversity and minor misclassifications, the system successfully addressed most practical issues, confirming its viability for real-time use. This work emphasizes the growing role of AI-driven solutions in enhancing personalized services, security systems, and demographic studies.

Looking ahead, further improvements such as hybrid modeling, multilingual support, and optimized deployment strategies will increase the model's scalability and effectiveness in diverse environments.

X. Future Developments

Several future developments could be made to enhance the real-time age and gender classification system, addressing its current limitations and broadening its applicability.

The first improvement would focus on enhancing the quality and diversity of the training dataset, allowing the model to handle a broader range of facial features, ethnicities, and environmental conditions. Incorporating hybrid models, such as the combination of Convolutional Neural Networks (CNNs) with transformer-based architectures, could enable the system to capture both spatial and contextual features, improving its performance.

Another key advancement would be the integration of explainable AI (XAI) techniques. These methods would make the system's predictions more transparent and interpretable, increasing trustworthiness for users. Additionally, optimizing the model for resource-constrained environments is crucial for scalability and deployment in real-time applications with limited hardware.

Finally, adding multilingual support and expanding the system to handle more demographic attributes would make the model versatile and adaptable to global, diverse user bases, increasing its potential for broader use cases.

XI. References

1) Levi, G., & Hassner, T. (2015). Age and Gender Classification Using Convolutional Neural Networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). doi:10.1109/CVPRW.2015.7301352.
2) Rothe, R., Timofte, R., & Van Gool, L. (2018). Deep Expectation of Real and Apparent Age from a Single Image Without Facial Landmarks. International Journal of Computer Vision (IJCV), 126(2), 144-157.
3) Parkhi, O. M., Vedaldi, A., & Zisserman, A. (2015). Deep Face Recognition. Proceedings of the British Machine Vision Conference (BMVC). doi:10.5244/C.29.41.
4) He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). doi:10.1109/CVPR.2016.90.
5) Howard, A. G., Zhu, M., Chen, B., et al. (2017). MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. arXiv preprint arXiv:1704.04861.
6) Simonyan, K., & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv preprint arXiv:1409.1556.
7) Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A Large-Scale Hierarchical Image Database. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). doi:10.1109/CVPR.2009.5206848.
8) Gupta, M., Chavhan, S., & Kale, K. (2020). Age and Gender Prediction Using Deep Learning. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 6(5), 342-348.
9) Zhang, Z., Luo, P., Loy, C. C., & Tang, X. (2014). Facial Landmark Detection by Deep MultiTask Learning. European Conference on Computer Vision (ECCV). doi:10.1007/978-3-319-10593-2_6.
10) Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely Connected Convolutional Networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). doi:10.1109/CVPR.2017.243.
11) Sandler, M., Howard, A., Zhu, M., et al. (2018). MobileNetV2: Inverted Residuals and Linear Bottlenecks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). doi:10.1109/CVPR.2018.00474.
12) Wen, Y., Li, K., Yang, Z., et al. (2016). A Discriminative Feature Learning Approach for Deep Face Recognition. European Conference on Computer Vision (ECCV). doi:10.1007/978-3-319-46478-7_9.
13) Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. Advances in Neural Information Processing Systems (NeurIPS).
14) Mollahosseini, A., Hasani, B., & Mahoor, M. H. (2016). AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild. IEEE Transactions on Affective Computing. doi:10.1109/TAFFC.2017.2740923.
15) Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press. Retrieved from https://www.deeplearningbook.org.
16) Sun, Y., Wang, X., & Tang, X. (2014). Deep Learning Face Representation by Joint Identification-Verification. Advances in Neural Information Processing Systems (NeurIPS).
17) Goo, M., Jang, J., & Yu, H. (2018). Gender and Age Classification Using Multi-Task Convolutional Neural Network. The 9th International Conference on Emerging Databases (EDB). doi:10.1007/978-981-13-1933-4_6.
18) Nguyen, D. T., & Le, T. D. (2020). Real-Time Age and Gender Classification Using Lightweight Deep Learning Models. IEEE Access, 8, 123123-123132. doi:10.1109/ACCESS.2020.2998765.
19) Alvi, M., Zisserman, A., & Naylor, P. (2018). Turning a Blind Eye: Explicit Removal of Biases and Variation from Deep Neural Network Embeddings. European Conference on Computer Vision (ECCV).
20) Chaudhary, A., Singh, S., & Soni, M. (2021). Age and

Gender Prediction Using Transfer Learning. International Journal of Engineering and Advanced Technology, 10(2), 123-129.