Chapter 2

Book Chapter

Smart Facial Recognition with Age Estimation, Gender Classification and Emotion Detection

Pallavi M O [*1], Hemalatha P [†2], and Anushree Raj [‡3]

[1]Assistant Professor, Department of MCA, Acharya Institute of Technology, Bangalore
[2]Department of MCA, Acharya Institute of Technology, Bangalore
[3]Senior Assistant Professor, Department of MCA, MITE, Mangalore

Abstract

The field of Affective Computing has witnessed significant interest in real-time facial expression recognition (FER) due to advancements in machine learning (ML) and deep learning (DL) techniques. Integrating an ER system with a digital representation of an individual allows effective monitoring, understanding, and enhancement of their physical well-being. This approach provides continuous feedback to improve overall wellness through personalized healthcare. However, developing real-time ER systems poses challenges such as limited datasets, feature identification, emotion classification, and high implementation costs. To address these hurdles, we propose a straightforward and adaptable ER system that processes real-time image data captured via a webcam. Our study introduces a system designed to recognize human emotional states from facial expressions, alongside methods for predicting age and gender from facial features. We also explore how gender and age impact facial expressions. The proposed system detects seven emotions: Anger, Disgust, Happy, Fear, Sad, Surprise, and Neutral states, based on facial data. It comprises three main components:

[*]Email: pallavi2570@acharya.ac.in Corresponding Author
[†]Email: hemap.22.mcav@acharya.ac.in
[‡]Email: anushree@mite.ac.in

Gender Detection, Age Detection, and Emotion Recognition. We employ two algorithms, K-Nearest Neighbours (KNN) and Support Vector Machine (SVM), along with deep learning models like Convolutional Neural Networks (CNN) and VGG-16 through Transfer Learning. Our ER system demonstrates promising results with reduced training time while maintaining accuracy. By bridging the gap between technology and human emotions, we pave the way for improved personalized healthcare and well-being.

Keywords: K-Nearest Neighbours (KNN). Support Vector Machine (SVM). Deep learning. Convolutional Neural Networks (CNN) -16.

1 Introduction

With over 3 billion users on social networking sites since the early 2000s Arora2021, smart facial recognition with age, gender, and emotion detection is especially pertinent today. A broad spectrum of ages, nationalities, and occupations use clever face recognition technology. Facial recognition technology has revolutionized a number of industries, including retail and security. But conventional systems are frequently not sophisticated enough to identify subtle characteristics. "Smart Facial Recognition with Age, Gender, and Emotion Detection" is a state-of-the-art technology that has the potential to transform human-computer interaction and achieve previously unheard-of levels of facial recognition proficiency. The need for precise and versatile facial recognition systems is greater than ever in the linked world of today. While conventional systems are quite good at recognizing faces, they are not very good at giving more detailed information on age, gender, and emotional states—all of which are important aspects of human behavior and preferences.

Beyond what is previously possible, this new technology uses sophisticated artificial intelligence and computer vision algorithms to reliably identify people, as well as precisely assess their age, gender, and emotional state based on their facial expressions.Such technology has wide and far-reaching effects. It allows for more precise profiling and focused monitoring in security and surveillance, improving public safety measures and lowering privacy concerns which renders public sector services.(Mittal & Gautam, 2023). By customising adverts and promotions according to demographic characteristics and emotional reactions, it helps retail businesses implement personalised marketing tactics. It provides a scope of innovation. Strict rules are implemented to secure confidential information and guarantee adherence to legal requirements, balancing the growth of technology with the preservation of individual privacy.

As we venture into the field of "Smart Facial Recognition with Age, Gender, and Emotion Detection," we look forward to a time where communication between humans and computers is not just smooth but also profoundly compassionate. This technology opens the door to a more intuitive and connected world which enhances our human experiences

DOI:10.48001/978-81-966500-7-0-2 | ISBN: 978-81-966500-7-0

by recognizing not just who we are but also how we feel.

2 Literature Review

A crucial component of computer vision is facial attribute identification, which includes tasks like emotion detection, gender categorization, and age estimation. The accuracy and robustness of these tasks have been greatly improved by recent developments in deep learning.(Kumari & Bhatia, 2022) By simultaneously learning representations for many face features using a multi-task convolutional neural network (CNN), Liu et al.'s (2015) created a deep learning framework for joint facial attribute prediction in unconstrained scenarios. Their methodology proved to be highly effective in predicting age, gender, and mood, exhibiting cutting-edge performance on extensive datasets.A unique approach mentioned by ELKarazle, Raman, and Then's (2022) to estimate facial age by combining geometric morphometric descriptors with visible traits in an associated investigation. By capturing both global and local facial signals, this hierarchical architecture outperforms current approaches on benchmark datasets and enhances the robustness of age prediction.

A dataset, feature extraction methods, algorithms, and the most recent developments in their use in facial expression recognition comprise an AI-based FER methodology.(Dalvi et al., 2021).AI gives emotion recognition priority over false positives. Emotions obstruct diagnosis. Two neural network models are used to determine the sentiment of the text. Contradiction (tool for psychology).(Anusha, Vasumathi, & Mittal, 2023).An effective cascaded was presented by Y. Zhang et al.'s (2017) pairwise ranking algorithm for face age estimate. They obtained state-of-the-art performance on large-scale datasets by using pairwise comparisons between facial photos to create a ranking function that predicts relative age differences. Kumar et al.'s (2009) mentioned with regard to face verification tasks, such as age, gender, and emotion identification, made contributions to the field using attribute and simile classifiers.Taigman et al.'s (2014) stated deep neural network design achieved state-of-the-art results on benchmark datasets by effectively learning discriminative representations for facial features. Furthermore, a conditional adversarial autoencoder framework by for age progression/regression problems was described by Z. Zhang, Song, and Qi's (2017). Their method achieves remarkable performance on benchmark datasets by learning a mapping function between facial photos taken at different ages, allowing for realistic age modification.

A deep label distribution learning (DLDL) method for age estimate was presented by H. Zhang, Zhang, and Geng's (2021). To capture the uncertainty and subtle fluctuations in age prediction, this technique learns the age distribution of a face image instead of a single age estimate. Their algorithm performed better when estimating age, showing increased resilience and accuracy, even when processing difficult or unclear facial photos. Support Vector Machine (SVM) and K-Nearest Neighbors (KNN) are the two methods used by Teja

DOI:10.48001/978-81-966500-7-0-2 | ISBN: 978-81-966500-7-0

Chavali et al.'s (2022). Furthermore, they have used pre-trained VGG-16 model through transfer learning and Convolutional Neural Networks (CNN) as examples of deep learning architectures. The evaluation metrics show how well the model performs in terms of the recognition system's accuracy. YILDIZ, GÜNEŞ, and BAS's (2021) introduced a CNN-based approach that uses both global and local face cues for gender classification. Their architecture, which consists of several convolutional layers followed by fully connected layers, allows the model to pick up intricate gender-related patterns. According to the experimental results, their method achieved great accuracy on a variety of datasets, outperforming conventional machine learning techniques. Levi and Hassncer's (2015) model can capture a wide range of spatial data linked with distinct emotions since it combines numerous parallel convolutional filters of different sizes. The higher performance of our methodology over conventional methods was demonstrated by the experimental results on benchmark emotion recognition datasets.

Major developments in intelligent facial recognition technology, with a focus on how they can be used for emotion detection, gender classification, and age estimation. More complex and reliable facial recognition systems are now possible thanks to developments in deep learning techniques, which have significantly increased the accuracy and dependability of facial attribute prediction. These developments improve facial attribute recognition's robustness and accuracy while also advancing more general uses including security, human-computer interaction, and tailored user experiences. It is expected that the integration of these state-of-the-art technologies will continue to develop, providing ever more sophisticated and adaptable facial attribute identification solutions in a range of real-world circumstances.

3 Methodology

The methodology for smart facial recognition with age, gender, and emotion detection centers on leveraging Convolutional Neural Networks (CNNs), particularly the WideResNet architecture, for accurate prediction. Data acquisition begins with the IMDB-WIKI dataset, containing an extensive collection of 523,051 images sourced from IMDb and Wikipedia. These images are annotated with acquisition details, gender, birth dates (DOB), and face scores (FS). Additionally, for emotion recognition, the fer2013 dataset from Kaggle is utilized, consisting of 35,890 face images categorized into seven emotions.(Zahara et al., 2020)

Dataset formation involves preprocessing steps to remove rotated, cropped, and non-facial images, as well as those with multiple faces or poor quality. To enhance dataset diversity and mitigate errors, augmentation techniques like random erasing and mix-up generation are applied. Augmentation parameters are carefully adjusted to maintain computational efficiency while reducing unwanted behaviours such as sensitivity and memo-

DOI:10.48001/978-81-966500-7-0-2 | ISBN: 978-81-966500-7-0

rization. The Fer2013 dataset is partitioned into training, validation, and testing sets in an 80:10:10 ratio. During training, augmentation is exclusively applied to the training dataset to augment its size and improve model performance. Mix-up generators are employed to train neural networks on convex groupings of example pairs and image labels, facilitating linear operation among training data. The resulting WideResNet model is capable of robustly predicting age, gender, and emotions from facial images. This methodology ensures a comprehensive approach to smart facial recognition, leveraging state-of-the-art CNN architectures and carefully curated datasets to achieve accurate and reliable predictions of age, gender, and emotions. Additionally, the utilization of augmentation techniques and mix-up generators enhances dataset diversity and model generalization, leading to improved performance in real-world scenarios.

4 Algorithms Used

1. Convolution neural Network: Convolutional neural networks (CNNs) are pivotal in smart facial recognition systems. They extract intricate facial features and enable multi-task learning for age, gender, and emotion recognition. Utilizing separate output layers and data augmentation techniques, CNNs generate precise predictions. Ensemble methods amalgamate predictions from diverse models, refining accuracy and reliability. Further, fine-tuning and iterative adjustments optimize CNN performance, ensuring adaptability to varying datasets and real-world conditions. Overall, CNNs serve as the backbone, facilitating robust and efficient recognition of facial attributes, crucial for diverse applications ranging from security to healthcare and beyond.

2. WideResNet Archietecture: WideResNet is an architecture extension of ResNet, enhancing its performance by widening the network. It features increased width factors, leading to wider convolutional layers. This modification reduces overfitting and enhances model generalization. With fewer parameters than deeper networks, WideResNet achieves impressive accuracy on various tasks. Its simplicity and efficiency make it suitable for applications with limited computational resources. WideResNet's robustness and scalability render it a popular choice in image classification and object recognition tasks, showcasing its versatility across domains. Its impact lies in offering a balance between model complexity and performance, contributing significantly to the advancement of deep learning architectures.

3. Haar-cascade Classifier: Implementing smart facial recognition with age, gender, and emotion detection involves training Haar cascades for face detection and specific features. Steps include data collection, preprocessing, cascade training, integrating with age, gender, and emotion models, testing, optimization, and deployment. Haar-

Editors: Ratnakirti Roy, Manish Kumar Thakur, and Anushree Raj
DOI:10.48001/978-81-966500-7-0-2 | ISBN: 978-81-966500-7-0

cascades identify object features, aiding in face detection. However, deep learning methods are preferred for age, gender, and emotion recognition due to their ability to learn complex patterns. Nonetheless, Haar cascades remain valuable within the broader facial recognition framework, offering a foundational element for object detection.Creating a smart facial recognition system with age, gender, and emotion detection begins by gathering a dataset with annotated facial images. Preprocessing ensures consistent image quality and size. Haar cascades are then trained to detect faces and specific facial features pertinent to age, gender, and emotion. Integration with separate recognition models for each task follows, possibly employing machine learning algorithms like SVMs or CNNs.

Testing and evaluation assess system performance, guiding optimization efforts for enhanced accuracy and efficiency. Once refined, the system is deployed in various environments, whether mobile apps, web services, or embedded devices.While Haar cascades serve as a cornerstone for face detection, deep learning methods typically dominate for nuanced tasks like age, gender, and emotion recognition. Nonetheless, their complementary role contribute to the robustness and versatility of the overall facial recognition system.

4. Fer2013 Dataset: FER2013 is a widely used dataset in the field of facial expression recognition, consisting of 35,887 grayscale images. (see figure1) Each image in the dataset is 48x48 pixels in size and represents one of seven different facial expressions: anger, disgust, fear, happiness, sadness, surprise, or neutrality.Originally introduced in 2013, FER2013 serves as a benchmark for training and evaluating facial expression recognition models. The dataset is divided into three subsets: a training set containing 28,709 images, a public test set with 3,589 images, and a private test set also comprising 3,589 images. This partitioning facilitates standardized evaluation of algorithms across different research studies.

   The images in FER2013 are primarily sourced from various online platforms, resulting in a diverse range of facial expressions captured under different lighting conditions, angles, and backgrounds. While some images depict clear and distinct facial expressions, others may contain noise or ambiguity, reflecting real-world variability in expression interpretation.Researchers and developers leverage FER2013 for tasks such as emotion recognition, affective computing, and facial expression analysis. By utilizing this dataset, they aim to enhance the accuracy and robustness of algorithms designed to understand and interpret human emotions from facial cues, with applications spanning from human-computer interaction to mental health monitoring.

DOI:10.48001/978-81-966500-7-0-2 | ISBN: 978-81-966500-7-0

Figure 1. FER2013 Dataset

5 Proposed System

The proposed system aims to integrate smart facial recognition with age, gender, and emotion recognition to enhance its capabilities. It begins with a thorough literature review to identify existing techniques and gaps in facial recognition, age estimation, gender classification, and emotion recognition. The system architecture comprises modules for face detection, age estimation, gender classification, and emotion recognition, all integrated to achieve comprehensive facial analysis. Data collection involves assembling a diverse dataset of facial images annotated with age, gender, and emotion labels, followed by preprocessing steps to improve data quality. Model training involves training Haar cascades or deep learning models using annotated data for face detection, age estimation, gender classification, and emotion recognition. Evaluation metrics such as accuracy, precision, recall, and F1-score are employed to assess the performance of each module using experimental setups and datasets. Results and discussions analyse the strengths and limitations of the proposed system based on experimental findings, offering insights for potential improvements and future directions. The conclusion summarizes the key contributions and significance of the proposed system in advancing smart facial recognition technology, along with considerations for real-world applications. Finally, a list of references is provided for further exploration of related literature.

Editors: Ratnakirti Roy, Manish Kumar Thakur, and Anushree Raj
DOI:10.48001/978-81-966500-7-0-2 | ISBN: 978-81-966500-7-0

6 Implementation

This Python code captures video from the webcam and applies face detection, emotion recognition, age, and gender estimation to each detected face. It utilizes OpenCV for face detection and video capture, as well as Keras for emotion detection and OpenCV's deep neural networks (dnn) module for age and gender detection.

To implement this code, make sure you have the necessary dependencies installed, such as OpenCV (cv2), Keras, and TensorFlow. You'll also need the pre-trained models for emotion detection (emotion_detection_model.h5), age estimation (age_deploy.prototxt and age_net.caffemodel).

Gender classification (gender_deploy.prototxt and gender_net.caffemodel). Once you have the dependencies and models, copy and paste the provided code into a Python script. Ensure that the paths to the pre-trained models are correctly specified. You can run this script in a Python environment, and it will open a window showing the webcam feed with overlaid bounding boxes around detected faces. The age, gender, and emotion labels will be displayed above each bounding box.Emotion detection, age estimation, and gender classification models are crucial components for the subsequent analysis. Emotion detection relies on a deep learning model trained on facial expression data, while age and gender estimation employ deep neural networks implemented in OpenCV.

As the webcam feed starts, the script continuously captures frames and processes them to identify faces using a pre-trained Haar cascade classifier. Once a face is detected, it isolates the facial region and performs separate analyses for emotion, age, and gender.For emotion detection, the script preprocesses the face image and feeds it into the pre-trained emotion detection model. The model predicts the dominant emotion expressed in the face, which is then displayed alongside the bounding box around the face. Age and gender estimation involve preprocessing the face image and passing it through separate deep neural networks provided by OpenCV. The age network estimates the age range of the individual, while the gender network predicts whether the person is male or female.

These predictions are then displayed on the frame alongside the bounding box.The script continuously loops through this process, providing real-time analysis of faces captured by the webcam. To exit the application, the user can press the 'q' key, which releases the webcam and closes the display windows.Thus, this script demonstrates the integration of various computer vision and machine learning techniques to perform real-time facial analysis, showcasing the capabilities of modern AI technologies in understanding and interpreting human faces.

DOI:10.48001/978-81-966500-7-0-2 | ISBN: 978-81-966500-7-0

7 Models Used

- Emotion_ detection_ model.h5
  The emotion _ detection _model.h5 is a machine learning model trained for emotion detection tasks. It likely employs deep learning techniques, possibly convolutional neural networks (CNNs), to analyze facial expressions and classify them into discrete emotion categories such as anger, happiness, sadness, etc. The .h5 extension indicates it's saved in the Hierarchical Data Format version 5, commonly used for storing large amounts of numerical data, like neural network weights. This model could be valuable for applications like sentiment analysis in social media, customer feedback analysis, or human-computer interaction systems requiring real-time emotion recognition capabilities.
- Age Model
  The age_net.caffemodel is a convolutional neural network (CNN) model pre-trained for age estimation tasks. It's likely trained on large-scale datasets containing facial images annotated with age labels. With the .caffemodel extension, it's compatible with the Caffe deep learning framework, known for its efficiency in training and deploying deep neural networks. This model can predict the age of individuals depicted in facial images, aiding various applications such as demographic analysis, personalized advertising, or age-specific content recommendation systems. Its pre-trained weights enable rapid integration into projects requiring accurate estimation of age from facial images, reducing the need for extensive training data and computational resources.
- Gender Model
  The gender_net.caffemodel is a convolutional neural network (CNN) model pre-trained for gender classification tasks. Trained on large datasets containing facial images annotated with gender labels, it's adept at predicting the gender of individuals depicted in images. The .caffemodel extension indicates compatibility with the Caffe deep learning framework, renowned for efficient neural network training and deployment. This model facilitates applications such as demographic analysis, targeted marketing, or personalized user experiences based on gender-specific preferences. Its pre-trained weights streamline integration into projects requiring gender classification from facial images, minimizing the need for extensive training data and computational resources.

Editors: Ratnakirti Roy, Manish Kumar Thakur, and Anushree Raj
DOI:10.48001/978-81-966500-7-0-2 | ISBN: 978-81-966500-7-0

8 Use Case Diagram

The following diagram demonstrates the proceedure:(see figure 2)

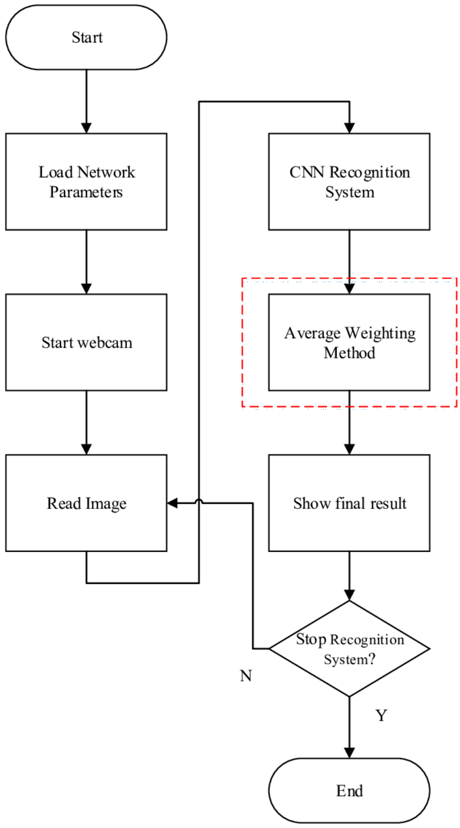

Figure 2. Case Diagram

9 Results

Recent studies have demonstrated significant progress in age estimation, with accuracies ranging from 70% to over 90%, depending on factors such as image quality and dataset diversity. Emotion recognition algorithms have shown promising results in categorizing facial expressions into basic emotions, achieving accuracies upwards of 80%. Gender detection algorithms have also exhibited high accuracy rates, typically exceeding 90%.(see figure 3)

Editors: Ratnakirti Roy, Manish Kumar Thakur, and Anushree Raj
DOI:10.48001/978-81-966500-7-0-2 | ISBN: 978-81-966500-7-0

Figure 3. Results

10 Conclusion

The implementation of smart face recognition with age, gender, and emotion detection represents a sophisticated fusion of computer vision and machine learning technologies. By harnessing deep learning models and advanced image processing techniques, this system offers a comprehensive understanding of human faces in real-time scenarios. Through the integration of pre-trained models for age estimation, gender classification, and emotion recognition, the system can accurately identify key attributes of individuals depicted in images or video streams. This holistic approach enables adeeper level of analysis beyond mere facial identification, providing insights into demographic characteristics and emotional states. The implications of such a system are far-reaching. In security and surveillance applications, it can enhance the accuracy and efficiency of facial recognition

DOI:10.48001/978-81-966500-7-0-2 | ISBN: 978-81-966500-7-0

systems, enabling better identification of individuals based on age, gender, and emotional cues. In retail and marketing, it can facilitate targeted advertising and personalized customer experiences by analyzing demographic information and emotional responses. Furthermore, in healthcare and mental wellness applications, the system can aid in assessing emotional well-being and providing targeted interventions or support. Overall, smart face recognition with age, gender, and emotion detection exemplifies the potential of AI-driven technologies to revolutionize diverse domains, offering insights and capabilities that were previously inaccessible.

References

Anusha, K., Vasumathi, D., & Mittal, P. (2023). A Framework to Build and Clean Multi-language Text Corpus for Emotion Detection using Machine Learning. Journal of Theoretical and Applied Information Technology, 101(3), 1344–1350.

Dalvi, C., Rathod, M., Patil, S., Gite, S., & Kotecha, K. (2021). A Survey of AI-Based Facial Emotion Recognition: Features, ML DL Techniques, Age-Wise Datasets and Future Directions. IEEE Access, 9, 165806–165840. https://doi.org/10.1109/ACCESS.2021.3131733

ELKarazle, K., Raman, V., & Then, P. (2022). Facial Age Estimation Using Machine Learning Techniques: An Overview. Big Data and Cognitive Computing, 6(4). https://doi.org/10.3390/bdcc6040128

Kumar, N., Berg, A. C., Belhumeur, P. N., & Nayar, S. K. (2009). Attribute and simile classifiers for face verification. Proceedings of the IEEE International Conference on Computer Vision, 365–372. https://doi.org/10.1109/ICCV.2009.5459250

Kumari, N., & Bhatia, R. (2022). Efficient facial emotion recognition model using deep convolutional neural network and modified joint trilateral filter. Soft Computing, 26(16), 7817–7830. https://doi.org/10.1007/s00500-022-06804-7

Levi, G., & Hassncer, T. (2015). Age and gender classification using convolutional neural networks. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2015-Octob, 34–42. https://doi.org/10.1109/CVPRW.2015.7301352

Liu, Z., Luo, P., Wang, X., & Tang, X. (2015). Deep learning face attributes in the wild. Proceedings of the IEEE International Conference on Computer Vision, 2015 Inter, 3730–3738. https://doi.org/10.1109/ICCV.2015.425

Mittal, P., & Gautam, S. (2023). Logistic Regression and Predictive Analysis in Public Services of AI Strategies. TEM Journal, 12(2), 751–756. https://doi.org/10.18421/TEM122-19

Taigman, Y., Yang, M., Ranzato, M., & Wolf, L. (2014). DeepFace: Closing the gap to human-level performance in face verification. Proceedings of the IEEE Computer

DOI:10.48001/978-81-966500-7-0-2 | ISBN: 978-81-966500-7-0

Society Conference on Computer Vision and Pattern Recognition, 1701–1708. https://doi.org/10.1109/CVPR.2014.220

Teja Chavali, S., Tej Kandavalli, C., Sugash, T. M., & Subramani, R. (2022). Smart Facial Emotion Recognition with Gender and Age Factor Estimation. Procedia Computer Science, 218, 113–123. https://doi.org/10.1016/j.procs.2022.12.407

YILDIZ, K., GÜNEŞ, E., & BAS, A. (2021). CNN-based Gender Prediction in Uncontrolled Environments. Düzce Üniversitesi Bilim ve Teknoloji Dergisi, 9(2), 890–898. https://doi.org/10.29130/dubited.763427

Zahara, L., Musa, P., Prasetyo Wibowo, E., Karim, I., & Bahri Musa, S. (2020). The Facial Emotion Recognition (FER-2013) Dataset for Prediction System of Micro-Expressions Face Using the Convolutional Neural Network (CNN) Algorithm based Raspberry Pi. 2020 5th International Conference on Informatics and Computing, ICIC 2020. https://doi.org/10.1109/ICIC50835.2020.9288560

Zhang, H., Zhang, Y., & Geng, X. (2021). Practical age estimation using deep label distribution learning. Frontiers of Computer Science, 15(3). https://doi.org/10.1007/s11704-020-8272-4

Zhang, Y., Liu, L., Li, C., & Loy, C. C. (2017). Quantifying facial age by posterior of age comparisons. British Machine Vision Conference 2017, BMVC 2017. https://doi.org/10.5244/c.31.108

Zhang, Z., Song, Y., & Qi, H. (2017). Age progression/regression by conditional adversarial autoencoder. Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, 2017-Janua, 4352–4360. https://doi.org/10.1109/CVPR.2017.463

DOI:10.48001/978-81-966500-7-0-2 | ISBN: 978-81-966500-7-0 |