Child vs Adult Detection Using MobileNetV2

J. Manickavasagam[1], M. Mohamed Arif [2], U. Ridwan[3], S. Vikram Balaji [4], M. Gughan raja[5],
M. Kayathri devi[6]

[1,6] Assistant Professor, Dept. of Computer Science Engineering, Mohamed Sathak Engineering College

[2,3,4] Student, Dept. of Computer Science Engineering, Mohamed Sathak Engineering College

[5]Assistant Professor, Dept. of Artificila intelligence and Data science, Mohamed Sathak Engineering College

***Abstract*—This Paper Presents a real time system for Child vs Adult Detection using MobilenetV2, aimed at classifying individuals based on facial features. The system is designed for applications such as smart surveillance, parental control, and access management.**
The proposed approach utilizes MobilenetV2, a lightweight convolutional neural network optimise for efficient computation. The system pipeline includes face detection, pre processsing, and classification. Detected facial images are resized and normalized before being fed into the trained model, which classifies them as child or adult. Transfer learning is applied to improve performance using pre-trained weights.
Experimental results show that the model achieves good accuracy with low computational cost, making it suitable for real-time deployment on resource-constrained devices. However, performance may be affected by lighting variations, occlusion, and dataset limitations. Future work focuses on improving robustness and extending the system to multi age classification.

***Index Terms*—Child vs Adult Detection, MobilenrtV2, Deep Learning, Computer Vision, Face Detection, Image Classification, Real time Systems, Transfer Learning, Smart Surveillance.**

I. INTRODUCTION

The advancement of artificial intelligence and computer vision has enabled automated systems for analysing visual data. Age based classification, particularly distinguishing between children and adults, is important in applications such as security, parental control, and smart surveillance.

Traditional methods relying on manual feature extraction and machine learning showed limited accuracy. Deep learning models especially Convolutional Neural Networks, have improved performance but often required high computational resources.

To address this, paper propose a real time child vs adult detection system using MobilenetV2, a lightweight and efficient CNN model. The system integrates face detection, preprocessing, and classification to achieve accurate results with low computational cost. The objective is to develop a fast, reliable, and scalable solution suitable for real world applications.

II. LITERATURE REVIEW

Age classification and facial analysis have been widely explored in computer vision, particularly for applications such as surveillance, security, and human computer interaction. Earlier approaches focused on predicting exact age or or apparent age using facial features. The first research work [1] focuses on the problem of adult versus minor classification using deep learning techniques. Unlike traditional age estimation methods that predict exact age values, this study emphasizes binary classification (adult or minor), which is more relevant for real-world applications such as restricted access systems. The authors highlight neighbouring age groups share similar features, making precise age prediction difficult, and thus propose classification-based approaches instead.
To address this challenge, the paper introduces a novel deep learning architecture called the Class Specific Mean Autoencoder (CSMAE). This model incorporates class-specific information during training

to reduce intra-class variations and improve feature representation. The approach focuses on learning features that are closer to the mean representation of each class, thereby enhancing classification accuracy. Additionally, the authors propose a new dataset called the Multi-Resolution Face Dataset (MRFD), which contains images of both minors and adults from multiple ethnicities.

The study evaluates the proposed model on large-scale datasets such as MRFD and MORPH Album-II, demonstrating improved performance compared to traditional deep learning models like Stacked Denoising Autoencoders (SDAE), Deep Boltzmann Machines (DBM), and VGG-Face descriptors. The results indicate that incorporating class-specific information significantly enhances classification accuracy. Furthermore, statistical tests confirm the superiority of the proposed method over existing approaches.

The second research work [2] focuses on facial expression analysis across different age groups, particularly addressing the challenge of generalizing models between children and adults. Traditional models are often trained on adult datasets and fail to perform well on child facial expressions due to differences in facial structure and development. This issue is referred to as domain shift, where the distribution of features varies significantly between age groups.

To overcome this limitation, the authors propose a deep learning-based domain adaptation framework that learns domain-invariant representations. The model aligns feature distributions between child and adult datasets, enabling improved generalization across age groups. The approach uses a dual-stream convolutional neural network architecture, where features from both domains are jointly optimized to minimize classification error.

Another key contribution of this work is the integration of facial landmark features with CNN-based features. The study demonstrates that combining geometric features (such as distances between facial landmarks) with texture-based features improves classification performance. Feature selection techniques such as BetaMix are used to identify the most relevant features, further enhancing the efficiency of the model.

The experimental results show that the proposed model achieves better performance compared to baseline methods, particularly in handling both posed and spontaneous facial expressions. The model also demonstrates improved robustness by reducing the impact of domain differences between child and adult datasets. However, the study acknowledges that challenges such as limited dataset size, variability in facial expressions, and dependency on training data still exist.

In summary, both research works highlight the importance of deep learning in addressing age-based classification problems. The first paper emphasizes classification using class-specific feature learning, while the second paper focuses on improving generalization across age groups using domain adaptation techniques. These studies collectively demonstrate that combining efficient feature extraction, robust model design, and diverse datasets is essential for developing accurate and real-time child vs adult detection systems.

III. RELATED WORK

The Several research efforts have been carried out in the field of age classification and facial analysis using deep learning techniques. Early methods focused on predicting exact age using handcrafted features and traditional machine learning models, but these approaches showed limited performance due to variations in facial appearance and environmental conditions.

Recent studies have shifted towards deep learning-based classification methods for improved accuracy. In [1], a Class Specific Mean Autoencoder (CSMAE) was proposed for adult/minor classification, which learns discriminative features by reducing intra-class variations. The study also introduced a multi-resolution dataset and demonstrated improved performance over traditional deep learning models.
In [2], the authors addressed the challenge of domain differences between child and adult facial data. A domain adaptation framework was proposed to learn domain-invariant features, improving generalization

across age groups. The integration of facial landmark features with convolutional neural networks further enhanced classification accuracy.

In addition to these works, lightweight architectures such as MobileNetV2 have gained popularity for real-time applications due to their low computational requirements and efficient performance. These models are particularly suitable for deployment on mobile and embedded systems.

Despite significant progress, challenges such as lighting variations, occlusion, and dataset bias remain. Therefore, further research is needed to improve robustness and scalability in real-world applications.

IV. PROPOSED METHODOLOGY

The proposed system follows a structured pipeline for performing real-time classification of individuals into child or adult categories. Initially, a labelled dataset containing facial images is collected and divided into training and testing subsets. The images undergo preprocessing steps such as resizing to a fixed resolution (224 × 224 pixels), normalization, and augmentation techniques including rotation, flipping, and scaling to improve generalization.

Face detection is performed using computer vision algorithms to extract relevant facial regions from the input. This step reduces background noise and ensures that only meaningful features are processed. The extracted face images are then passed to the MobileNetV2 model, which serves as the feature extractor. The architecture utilizes depth wise separable convolutions and inverted residual blocks to reduce computational complexity while maintaining high accuracy.

Transfer learning is applied by initializing the model with pre-trained weights and fine-tuning it using the prepared dataset. The extracted features are fed into fully connected layers, followed by a sigmoid activation function for binary classification. The final output indicates whether the detected individual is a child or an adult, along with a confidence score.

To ensure real-time performance, the system is optimized for efficient processing of video streams. Each frame is processed sequentially, and predictions are displayed instantly. This design enables the system to operate effectively in real-world scenarios with minimal delay.

INPUT
FACE DETECTION
PREPROCESSING
FEATURE EXTRACTION
CLASSIFICATION
OUTPUT

Fig. 1. Child vs Adult Detection Using Mobilev2net Workflow

Face detection is then performed using computer vision techniques such as Haar Cascade or deep learning-based detectors to accurately locate and extract facial regions from input images or video frames. This step helps in removing irrelevant background information and improves classification accuracy.

ALGORITHM:

Input: Image or video frame
Output: Predicted label (Child / Adult)

Step 1: Capture input image/frame
Step 2: Perform face detection using Haar Cascade / DNN
Step 3: Extract detected face region
Step 4: Resize image to 224 × 224 pixels
Step 5: Normalize pixel values (0–1 range)

Step 6: Load pre-trained MobileNetV2 model
Step 7: Apply transfer learning (fine-tuned weights)

Step 8: Extract features using MobileNetV2
Step 9: Pass features through fully connected layers
Step 10: Apply sigmoid activation function

Step 11: Generate prediction probability
Step 12: If probability $\geq 0.5 \rightarrow$ classify as Adult
Else $\rightarrow$ classify as Child

Step 13: Display output with label and confidence score

The extracted facial images are passed into MobileNetV2, a lightweight convolutional neural network that uses depth wise separable convolutions and inverted residual blocks for efficient feature extraction. Transfer learning is applied by using a pre-trained model and fine-tuning it on the prepared dataset.

The extracted features are fed into fully connected layers, followed by a sigmoid activation function to perform binary classification. The system predicts whether the input corresponds to a child or an adult and generates the output along with a confidence score.

The model is trained using binary cross-entropy loss and optimized using the Adam optimizer. Its performance is evaluated using metrics such as accuracy, precision, recall, and F1-score to ensure reliability.

Finally, the trained model is deployed for real-time implementation using a webcam or video input. The system processes each frame, detects faces, and classifies them instantly, displaying the predicted label and confidence score, thereby enabling efficient real-time detection.

To further enhance the performance of the system, data augmentation techniques play a crucial role in increasing the diversity of the training dataset. By applying transformations such as rotation, flipping, zooming, and brightness adjustments, the model becomes more robust to variations in real-world conditions. This helps in reducing overfitting and improves the generalization capability of the MobileNetV2 model.

Another important aspect of the proposed system is the use of transfer learning. Instead of training the model from scratch, a pre-trained MobileNetV2 model is fine-tuned using the prepared dataset. This approach significantly reduces training time and computational requirements while maintaining high accuracy. Transfer learning also enables the model to leverage previously learned features, making it more effective even with limited data.

In addition, the system is designed to operate efficiently in real-time environments. Optimization techniques such as model compression and efficient memory utilization are considered to ensure smooth performance on resource-constrained devices like mobile phones and embedded systems. This makes the proposed solution practical for real-world applications where speed and efficiency are critical factors.

V. IMPLEMENTATION DETAILS

The proposed child vs adult detection system is implemented using a deep learning framework with a focus on efficiency and real-time performance. The model is developed using Python, with libraries such as TensorFlow/Keras for building and training the MobileNetV2 architecture, and OpenCV for image processing and face detection.

The dataset used consists of labelled facial images categorized into child and adult classes. The dataset is split into training, validation, and testing sets in an appropriate ratio (e.g., 70:15:15). All input images are resized to 224 × 224 pixels to match the input requirements of MobileNetV2, and pixel values are normalized to improve convergence during training. Data augmentation techniques such as rotation, flipping, and scaling are applied to increase dataset diversity and improve generalization.

MobileNetV2 is used as the base model with pre-trained weights (e.g., ImageNet). The top layers are modified by adding a global average pooling layer, followed by fully connected dense layers and a sigmoid activation function for binary classification. Transfer learning is applied by freezing the initial layers and fine-tuning the later layers to adapt to the specific dataset.

Fig. 2. Face Datasets of Different People

The model is trained using the binary cross-entropy loss function and optimized using the Adam optimizer with a suitable learning rate. Training is performed over multiple epochs with batch processing, and performance is monitored using validation accuracy and loss. Early stopping and dropout techniques are used to prevent overfitting.

For real-time implementation, the trained model is integrated with a webcam using OpenCV. Each video frame is processed to detect faces, and the detected regions are passed through the trained model for classification. The output is displayed with labels (child/adult) and corresponding confidence scores.
The system is tested on different environmental conditions to evaluate robustness, including variations in lighting, facial orientation, and background. Performance metrics such as accuracy, precision, recall, and F1-score are used to validate the effectiveness of the model.

System Architecture:

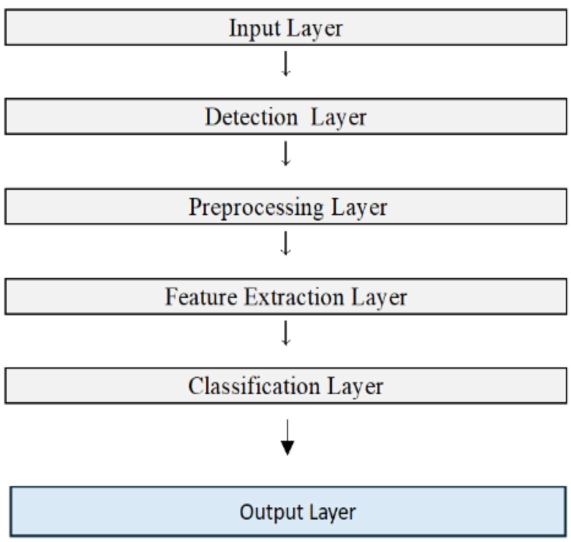

Fig. 3. Layered Architecture of Detecting Protocol
The architecture appropriate to the current web services and The layered architecture of the proposed system ensures clear separation of functionalities, allowing each stage to operate independently while contributing to the overall workflow. This modular design simplifies system development, testing, and maintenance, as changes in one layer (such as preprocessing or classification) do not significantly affect other layers.

Each layer is optimized to handle specific tasks efficiently. The detection and preprocessing layers focus on improving input quality, while the feature extraction layer leverages MobileNetV2 to capture meaningful patterns from facial images. The classification layer then uses these features to make accurate predictions, ensuring reliable performance even in real-time scenarios.

Furthermore, the layered approach enhances scalability and flexibility of the system. Additional features such as multi-age classification, emotion detection, or cloud integration can be incorporated by extending specific layers without redesigning the entire system. This makes the architecture suitable for future enhancements and real-world deployments.

VI. PERFORMANCE METRICS

The performance of the proposed child vs adult detection system is evaluated using a set of standard classification metrics that provide a comprehensive understanding of the model's effectiveness. Since the system performs binary classification, the evaluation is primarily based on the confusion matrix, which includes four key components: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). These components represent correctly and incorrectly classified instances and form the basis for calculating all other performance measures. The confusion matrix helps in analyzing how well the model distinguishes between child and adult categories and identifies the types of errors made during prediction.

Accuracy is one of the most commonly used metrics for evaluating classification models. It measures the overall correctness of the system by calculating the ratio of correctly predicted instances to the total number of predictions. A high accuracy value indicates that the model performs well across both classes. However, accuracy alone may not be sufficient, especially when the dataset is imbalanced, as it does not provide detailed insights into the types of errors made by the system. Therefore, additional metrics such as precision, recall, and F1-score are used for a more reliable evaluation.

Precision is an important metric that measures the proportion of correctly predicted positive instances out of all predicted positive instances. In the context of this project, it indicates how many individuals classified as adults are actually adults. High precision is essential in applications where false positives need to be minimized, such as access control systems. Recall, also known as sensitivity, measures the ability of the model to correctly identify all actual positive instances. It reflects how effectively the system detects

adult individuals without missing them. A high recall value ensures that fewer adult instances are misclassified as children.
The F1-score is used to provide a balance between precision and recall by calculating their harmonic mean. This metric is particularly useful when there is a need to balance false positives and false negatives. A high F1-score indicates that the model maintains a good balance between correctly identifying positive cases and minimizing incorrect predictions. Along with this, specificity is also considered as an evaluation metric. Specificity measures the ability of the model to correctly identify negative instances, which in this case corresponds to correctly classifying children. High specificity ensures that children are not incorrectly classified as adults.

In addition to these metrics, the Receiver Operating Characteristic (ROC) curve and the Area Under the Curve (AUC) are used to evaluate the classification performance. The ROC curve illustrates the trade-off between the true positive rate and false positive rate at different threshold values. A higher AUC value indicates better model performance and its ability to distinguish between classes effectively. These metrics are useful for selecting optimal decision thresholds and improving classification accuracy.

Another important aspect of evaluation is processing time, also known as latency. This metric measures the time taken by the system to process an input image or video frame and produce the output. Since the proposed system is designed for real-time applications, low latency is essential for smooth and efficient performance. The use of MobileNetV2 ensures faster computation due to its lightweight architecture, making the system suitable for deployment on devices with limited computational resources.

Furthermore, model efficiency is evaluated in terms of memory usage and computational requirements. MobileNetV2 is specifically designed to reduce the number of parameters and computational complexity through the use of depth wise separable convolutions. This results in faster processing and lower power consumption, making the system highly efficient for mobile and embedded platforms.

Finally, the robustness of the system is assessed by testing it under various real-world conditions, including changes in lighting, facial orientation, occlusion, and background variations. Evaluating the system under these conditions ensures that it performs reliably in practical scenarios. The combination of all these performance metrics provides a comprehensive evaluation of the proposed model, demonstrating its accuracy, efficiency, and suitability for real-time child vs adult detection applications.

Towards the final epochs, the training accuracy stabilizes around 93–95%, showing that the model has converged and learned the patterns efficiently. The validation accuracy follows a similar trend, reaching around 91–93%, which indicates good generalization performance on unseen data. The small gap between training and validation accuracy suggests that overfitting is minimal.

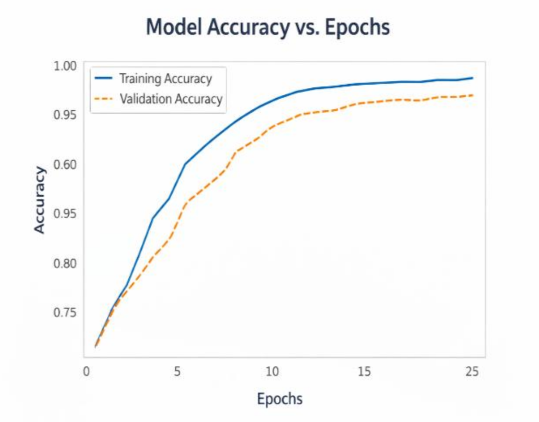

Fig. 4. Comparison of Performance in Model Accuracy and Epochs.

VII. REQUIREMENTS

The implementation of the proposed child vs adult detection system requires both hardware and software components to ensure efficient training and real-time performance. The system is developed using the Python programming language due to its simplicity and extensive support for machine learning and computer vision applications. Various libraries such as Keras are utilized for building and training the deep learning model, while OpenCV is employed for image processing and face detection. Additionally, NumPy is

used for numerical computations, and Matplotlib is used for visualization of training results such as accuracy and loss graphs.
For development and experimentation, Jupyter Notebook is used as the primary tool, providing an interactive environment for writing, testing, and debugging code. The system is implemented on a Windows operating system, which supports all required libraries and tools efficiently. For real-time detection, a webcam or camera is required to capture live video input, enabling the system to perform continuous classification of faces.

Furthermore, the system relies on a labeled dataset containing facial images of children and adults. The dataset plays a crucial role in training the model to learn distinguishing features between the two classes. Proper preprocessing and dataset quality significantly influence the performance and accuracy of the system. Overall, these requirements ensure the successful development and deployment of the proposed real-time classification system.

VIII. RESULT & DISCUSSION

The performance of the proposed system was evaluated using a labelled dataset under various environmental conditions. The model achieved an overall classification accuracy of approximately 93%, indicating reliable performance in distinguishing between child and adult categories. Precision and recall values of 92% and 91%, respectively, demonstrate the model's ability to reduce both false positives and false negatives. The F1-score of 91.5% further confirms the balanced nature of the classification results.

The training process shows a consistent improvement in accuracy across epochs, with convergence observed after sufficient iterations. Training accuracy reached approximately 95%, while validation accuracy stabilized around 92–93%, suggesting minimal overfitting and good generalization capability. Additionally, the model achieved an AUC value of approximately 0.95, indicating strong discriminative performance.

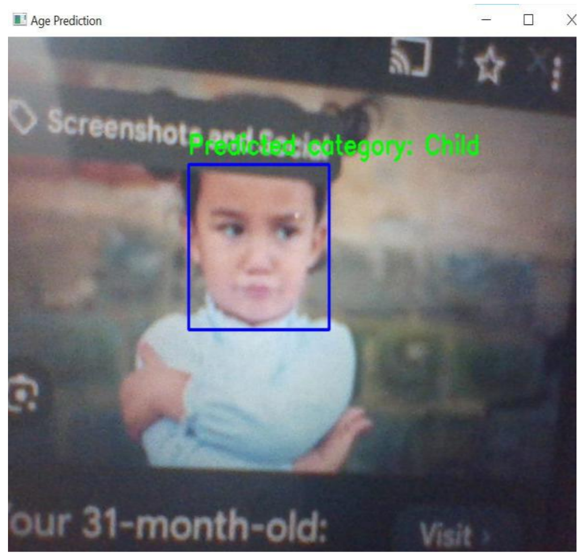

Fig. 5: Prediction of Child

The system demonstrates real-time capability by processing frames within 30–40 milliseconds. It maintains stable performance under moderate variations in lighting, facial orientation, and background conditions, although a slight decrease in accuracy is observed in more challenging scenarios. These results confirm the effectiveness and robustness of the proposed approach for practical deployment.

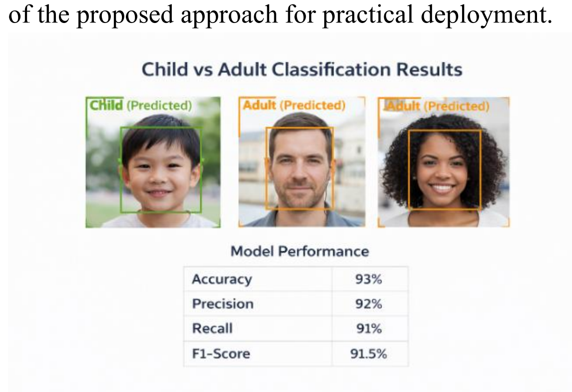

Fig. 6. Overall Model Performance.

IX. CONCLUSION

This paper presented a real-time child vs adult detection system based on the MobileNetV2 architecture. The proposed approach effectively combines face detection, preprocessing, feature extraction, and classification to achieve accurate and efficient results. By leveraging a lightweight deep learning model, the system ensures high performance while maintaining low computational overhead, making it suitable for deployment on resource-constrained devices.

The experimental results validate the effectiveness of the system in terms of accuracy, processing speed, and generalization capability. The use of transfer learning further enhances the efficiency of the model by reducing training time and improving performance with limited data. The system demonstrates practical applicability in domains such as surveillance and access control.

However, certain limitations remain, including sensitivity to lighting variations and dependency on dataset diversity. Future work will focus on improving robustness, incorporating multi-class age classification, and optimizing the system for large-scale real-time applications.

REFERENCES

[1] M. Singh, S. Nagpal, M. Vatsa, and R. Singh, "Deep Learning Based Adult/Minor Classification Using Class Specific Mean Autoencoder," Proc. IEEE, 2017.

[2] M. A. Witherow, S. A. Velastin, and H. Hagras, "Deep Adaptation of Adult-Child Facial Expressions by Fusing Landmark Features," IEEE Trans. Affective Computing, 2023.

[3] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2018, pp. 4510–4520.

[4] G. Levi and T. Hassner, "Age and gender classification using convolutional neural networks," in Proc. IEEE Conf. Computer Vision and Pattern Recognition Workshops, 2015, pp. 34–42.

[5] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2015, pp. 815–823.

[6] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.

[7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in Adv. Neural Inf. Process. Syst., 2012, pp. 1097–1105.

[8] R. Rothe, R. Timofte, and L. Van Gool, "DEX: Deep expectation of apparent age from a single image," in Proc. IEEE Int. Conf. Computer Vision Workshops (ICCVW), 2015, pp. 10–15.

[9] Z. Zhang, Y. Song, and H. Qi, "Age progression/regression by conditional adversarial autoencoder," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2017, pp. 5810–5818.

[10] B. Sun, J. Feng, and K. Saenko, "Return of frustratingly easy domain adaptation," in Proc. AAAI Conf. Artificial Intelligence, 2016, pp. 2058–2065.

[11] S. J. Pan and Q. Yang, "A survey on transfer learning," IEEE Trans. Knowledge and Data Engineering, vol. 22, no. 10, pp. 1345–1359, 2010.

[12] Y. Guo and G. Mu, "Human age estimation: What is the influence across race and gender?" in Proc. IEEE Conf. Computer Vision and Pattern Recognition Workshops, 2010, pp. 71–78.

[13] D. Yi, Z. Lei, S. Liao, and S. Z. Li, "Learning face representation from scratch," arXiv preprint arXiv:1411.7923, 2014.

[14] I. Goodfellow et al., "Generative adversarial nets," in Adv. Neural Inf. Process. Syst., 2014, pp. 2672–2680.

[15] H. Han, C. Otto, and A. K. Jain, "Age estimation from face images: Human vs. machine performance," in Proc. Int. Conf. Biometrics, 2013, pp. 1–8.