

AfriNet: A Facial Age Estimation Model for Age-Restricted Services

^{1*}Gilbert George, ²Usman Bello Abubakar, ³Jumare Muhammed, ⁴Tijjani Abdullahi
^{1,2,4}Department of Computer Science, Baze University, Abuja, Nigeria
³Department of Computing Federal College of Education, Zaria, Nigeria
Corresponding Author: gilbert.george@bazeuniversity.edu.ng

ABSTRACT

This article introduces a lightweight, custom-designed convolutional neural network (CNN) called AfriNet for real-time facial age estimation through both regression and classification. The model integrates demographic diversity, especially African facial features. AfriNet was trained and tested with UTKFace, CASIA Africa Face, and a locally collected dataset from Baze University and the Federal College of Education, Zaria. The model makes a strong hit as it is capable of achieving up to 99% accuracy in the classification of age groups and a mean absolute error (MAE) as low as 0.65 years on CASIA Africa Face. Besides, AfriNet is comparably much quicker in its inference than MobileNet, VGG, and ResNet, thus, it can be easily deployed on resource-constrained devices. The paper talks about AfriNet's potential use in age-restricted access control as a tool for the child online protection's (COP) enhancement in Nigeria, while at the same time, it is addressing the ethical and fairness issues.

KEYWORDS: Deep Learning, AfriNet, MobileNet, Age Estimation, Classification & Regression, Demographic Diversity

I. INTRODUCTION

Facial age estimation is a major tool that has been used in different areas such as security, biometric verification, demography, content filtering, and targeted services. Due to the progress in deep learning and Convolutional Neural Networks (CNNs), lightweight models (e.g., MobileNet [1–3]) are now able to perform real-time inference on low-power devices, thus they are appropriate for age-restricted access control. However, while advancements have been made in facial analysis, the African demographics have been significantly less represented in most facial datasets, which has led that the major cause of unfair performance discrepancies be these demographics. This research presents AfriNet, a specially designed lightweight CNN that is feature-optimized for African faces and tested with datasets having better demographic representation. Children in Nigeria are getting access to inappropriate online content [5]. But, the current age estimation tools have a hard time because they are not sufficiently diverse demographically, especially when it comes to the faces of children [6]. This work is intended to come up with a precise, instantaneous facial age estimation system that will be able to safeguard children by checking if the users are of ages allowed to access the content. To build a multi-task CNN model for facial age estimation (classification and regression) that can be used in age-restricted services.

The study aids in the advancement of automated age-verification systems that can be used in various sectors such as e-voting, online safety for minors, and identity management in Nigeria. AfriNet, by including African facial data, assists in the reduction of demographic bias in age

estimation models. The research is limited to the population aged 5-30 years, employing multi-task learning for both classification (age groups) and regression (exact age). Although there are some limitations to the study, which include

- a) Limited availability of annotated African facial datasets, particularly for (ages 5-12.)
- b) Differences caused by genetics, cosmetics, lighting, and camera quality.
- c) Possibility of bias if datasets are not equally distributed in terms of gender or age subgroups.

While the originality of AfriNet can be summarized in the three major areas: its architecture, the way the task is formulated, and the demographic group it is specialized for. These features set AfriNet apart from typical lightweight CNNs like MobileNet, ShuffleNet, EfficientNet-Lite, and other compact architectures. AfriNet is equipped with a neat, tailor-made feature extractor for facial age estimation just by one glance, which is different from usual lightweight CNNs that might, for instance, influence depthwise separable convolutions (MobileNet) or compound scaling (EfficientNet) or inverted residual blocks of (MobileNetV2) used in general image classification. The proposed AfriNet model comprises of three convolutional blocks with moderate filter sizes ($64 \rightarrow 32 \rightarrow 32$); thus, the computational cost is cut down while facial texture patterns for age recognition are given due consideration. We balanced the use of dropout and batch normalization at every stage, the interaction of dropout with batch normalization stabilizes the training phase on small demographic datasets and lessens the problem of overfitting, this was considered with this case of African facial images, which are scarcely represented in most public datasets. Furthermore, AfriNet simultaneously performs: age regression (exact numerical age) and age-group classification (underage vs. adult). This joint optimization method equips the network with the ability to find features that relate to age consistently which is not the case with many lightweight CNNs that are single-task or need architectural modification for multi-task learning. Custom fully connected design optimized for feature disentanglement. The dense layers have a growing depth structure i.e. ($64 \rightarrow 64 \rightarrow 128 \rightarrow 128 \rightarrow 512$) that brings out:

- (a). coarse-to-fine facial aging cues
- (b). ethnicity-specific feature patterns
- (c). nonlinear relationships between visual cues and biological age

given the fact that most lightweight CNNs compress the dense layers for mobile inference, thus limiting the age-regression precision, AfriNet makes a trade-off between computational efficiency and representational power.

Furthermore, AfriNet is trained specifically on African datasets of facial images (CASIA Africa Face and custom dataset), thereby becoming one of the first lightweight models that are demographically representative of the African population. Existing models, on the other hand, are mainly trained on (UTKFace, FG-NET, MORPH Album 2) Asian and Caucasian datasets, which ultimately leads to their performance differences. With AfriNet we do not let this difference continue any longer as having a custom build architecture, that is having Superior inference-time and efficiency on CPU is the goal of our research. The model is capable of: Inferencing in 0.009 seconds per image. As such, it is quicker than MobileNet, VGG16, VGG19, ResNet50, Therefore, it can be put to a real-time access control, mobile devices, low-resource environments typical of Africa. This system was created strictly for age-restricted services in Nigeria and Africa, In contrast to universal models, AfriNet has a demographic tuning feature for the local population, it implements age brackets that are relevant for the child-protection regulation and can be easily integrated into a Python-based access-control system, facilitates real-time decision-making (grant/deny access). This is a practical contribution that existing CNNs have not addressed.

II. RELATED WORKS

The research [7] presented a deep learning model in two phases where the initial step included the detection of face (using cascaded CNNs) and feature extraction, the second step involved the use of regression for the prediction of biological age. For clinical populations, the system obtained an MAE of 4.09 years, thus, it was performing at par with the state-of-the-art models or even better on the curated datasets. In this paper [8], the authors used an approach that merges deep learning feature extraction with manifold learning for dimensionality reduction, thereby achieving better results on age-annotated datasets (MORPH: MAE 1.60, FG-NET: MAE 2.48). The research exemplifies the way the combination of these methods helps in precise age estimation. The research [9] presented indicates that the face identification trained CNNs can be made to perform age estimation tasks through transfer learning, thus localizing the overfitting problem and yielding competitive results on the difficult Adience benchmark. By using a multitask learning implementation with MobileNet on the Raspberry Pi in a recent experiment, the authors [10] reached an accuracy of 99.3% in age estimation, thus proving MobileNet as an excellent and ideal candidate for real-time, resource-constrained applications. The paper [11] presented a model for age prediction from African demographic data, where a genetic algorithm was employed in conjunction with a BP-trained ANN and local binary pattern features. The method is pioneer in the sense that it is the first one, basically, to be used for the faces of the less visually represented ethnic groups. This research [12] is centered on age-invariant face recognition systems that highlight the difficulties of keeping accuracy consistent for large age differences and various demographics by means of 3D modeling and multitask learning in age prediction. The authors in [13] presents the method as an end-to-end multi-task learning (MTL) network termed CR-MT, which merges age classification and age regression tasks to enhance facial age estimation. The key idea includes shared feature representation where the network shares the first convolution layers (based on AlexNet) of three different networks to learn the common features which are equally useful for both tasks. Also, dual branch architecture where the network after shared layers gets bifurcated into two branches: - A classification branch which classes the faces into different age groups and a regression branch which gives the exact age.

III. METHODOLOGY

AfriNet is made up of three convolutional blocks followed by four fully connected layers and two task-specific output heads. The model aims to maintain a balance between computational efficiency and the capacity to extract detailed facial aging cues.

(a). The Input Layer

Shape: (224, 224, 3) which was adopted from mobileNet architecture, and tested to be the best in input size for performance. An RGB facial image is fed into the network after the image has been preprocessed (cropping, resizing, and normalization).

Model Structure



Figure 1: model structure

Table 1: Architecture Depth and Parameter Summary

Component	Layers	Parameters
Convolutional blocks	3 Conv layers, 3 MaxPool, 3 BN, 3 Dropout	~92K
Fully connected layers	4 Dense + 1 Shared Dense	~850K
Output heads	Regression + Classification	~514
Total Parameters		≈ 945,000 parameters

In table 1, we show a summary of the model architecture which includes the total parameters in the AfriNet architecture.

Table 2: Training Hyperparameters

Setting	Value
Optimizer	Adam
Initial Learning Rate	0.001
Loss Functions	MAE (regression), Categorical Cross-Entropy (classification)
Loss Weights	regression = 0.7, classification = 0.3
Batch Size	32
Epochs	50–100 (actual count should match your training logs)
Validation Split	20%
Test Split	10%
Early Stopping	Patience = 10
Learning Rate Scheduler	ReduceLROnPlateau (factor=0.1)
Weight Initialization	He-normal
Regularization	Dropout in all blocks, L2 optional

To make sure that researchers can achieve reproducibility, the AfriNet model was set up with the Adam optimizer having an initial learning rate of 0.001, batch size of 32, and the number of epochs varying 50-75 depending on early stopping which is shown in table 2. The loss function for the multi-task network was a combination of MAE for regression and categorical cross-entropy for classification. The feature extractor is made up of three convolutional blocks (64, 32, 32 filters), where each block is followed by max pooling, batch normalization, and dropout. The fully connected head has four Dense layers (64, 64, 128, 128 units) as well as a shared 512-unit representation layer. Among data augmentation techniques were horizontal flipping, rotation, zoom, and brightness changes. The model has close to 0.94M trainable parameters and was trained with an 80:20:10 (train-validation-test) split.

(b). Data Augmentation

To prevent overfitting and improve generalization, we used the following augmentations during training:

- Horizontal flip
- Random rotation ($\pm 10^\circ$)
- Random brightness shifts
- Zoom (0.2)
- Width/height shift (0.1)
- Normalization to pixel range [0,1]
- Training Hardware
- GPU: NVIDIA GTX 1656 Google Colabs
- RAM: 16 GB
- Framework: TensorFlow/Keras
- OS: Window

(c).System Architecture

The system has the following components:

- Input Module: Captures images using cameras or loads them from a dataset.

- Preprocessing: Detects and crops faces; performs resizing, normalization, and augmentation.
- Model Architecture: AfriNet custom based CNN which serves as the backbone for feature learning. The final dense layers are fine-tuned for regression and classification to estimate age. Output Module: Visualizes results by overlaying age predictions on detected faces.

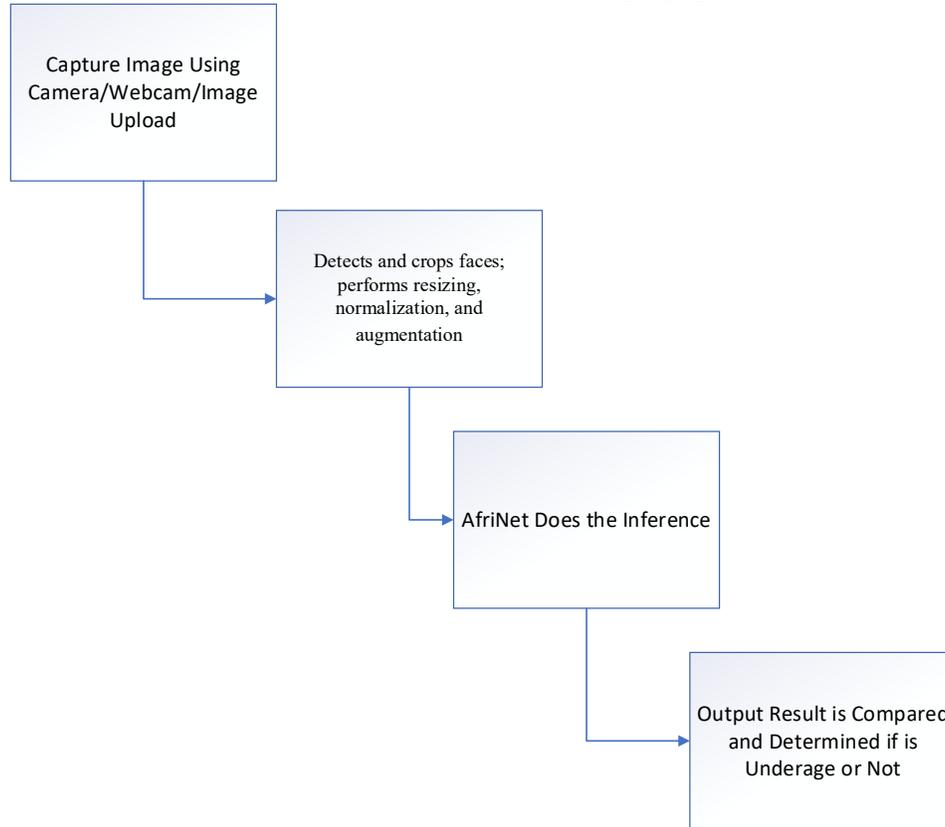


Figure 2: System Architecture

The above figure 2 shows the high-level system architecture design, for the application built for inference using the AfriNet model.

3.2 Dataset

We used the Popular datasets for age estimation include which included: UTKFace, CASIA Face Africa was used for training and a custom dataset collected in Baze and Federal College of Education, Zaria were used for validation. These datasets provide annotated facial images covering diverse age groups, gender, and ethnicity. These datasets are publicly available on Kaggle. We also measured the inference time it took AfriNet to complete inference with state of the art models such as MobileNet, VGG16, VGG19 and ResNet 50.

$$\text{Average Time} = \frac{\text{TotalTime}}{\text{TotalNumberOfImages}} \quad (1)$$

The average time is calculated by dividing the total time taken by the total number of images as shown in Equation 1.

RESULTS AND DISCUSSION

The model is trained and validated using a split of the dataset (70:20:10) for training, validation and testing with Mean Absolute Error (MAE) as the primary metric. Reported MAE values for AfriNet-based architectures on UTKFace are as low as 2.06 years and Casia Africa Face a local dataset, it was as low as 0.65 in years, demonstrating competitive accuracy and real-time inference speeds.

Table 3: Overall Accuracy

Dataset	Overall Accuracy
CASIA African Face	94%
UTKFace	92%

Table 3 shows the overall accuracy obtained from the experiments, where we see an improved accuracy in CASIA African Face dataset. We present a class-wise classification result of our models' performance on the two datasets.

Table 4. CASIA Africa Face Classification Result

Age Class	Precision	Recall	F1-Score	Interpretation
13–17	0.94	0.91	0.93	Shows High precision indicates few false positives; slightly lower recall suggests some teens are misclassified as adjacent ages.
18–25	0.93	0.95	0.94	Shows Balanced and strong performance; facial features in this range are well learned.
28–30	0.96	0.97	0.97	Shows Best-performing class; mature facial features are more distinctive and easier to classify.

Table 5. UTK Face Classification Result

Age Class	Precision	Recall	F1-Score	Interpretation
5–12	0.92	0.92	0.92	This shows good performance; children's facial features are relatively distinct from adults.
13–17	0.82	0.80	0.81	This is the Weakest class; high confusion with neighboring age groups.
18–30	0.95	0.95	0.95	This shows Strong performance due to broader age range and stable adult features.

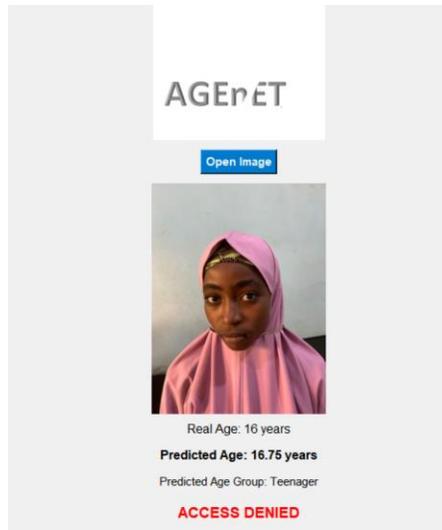


Figure 3: AfriNet Predicting the Age of the Lady as 16.75 and her Age Group as Teenager

From the Figure 3, we see an instance where the model predicted the lady as underaged and denied her access based on her facial features.

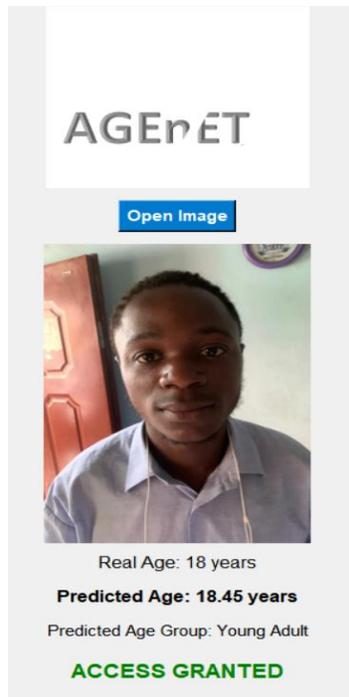


Figure .4: AfriNet Predicting the Age of the Man as 23.04 and his Age Group as Young Adult

From the fig 2. above we see an instance where the model predicted the man as an adult and granted him access based on his facial features.

Table 6: Inference Time across the different Architectures

Model	Total Inference Time(seconds)	Inference Time per image (seconds)
MobileNet	92	0.0273
VGG16	1067	0.3112
VGG19	1335	0.3903
RESNET 50	361	0.1061
AfriNet	34	0.00994

Table 7: Comparison with other works

Works	Classification Accuracy	MAE	Architecture	Dataset
[13]	-	2.15	AlexNet	Morph dataset
[10]	99 %	-	MobileNet	VGGFace2
AfriNet	94%		Custom Architecture	CASIA Africa Face
AfriNet	92%		Custom Architecture	UTK Face
AfriNet		0.64		CASIA Africa Face
AfriNet		2.06		UTK Face

In table 7 we compare other state of the art models, where we see AfriNet outperformed in [13] we see there was no mention of classification accuracy but from MAE we see that AfriNet outperforms the model.

Our model reflects improved performance on the CASIA African Face dataset and thus the features it has learned seem to be more in line with the actual facial characteristics, age-related changes and imaging conditions. This is particularly because the image resolution is higher and the acquisition environment is more controlled in comparison with UTKFace. When comparing the cross-dataset class, the adolescent age group (13–17) had an F1-score of 93% on CASIA African Face, whereas on UTKFace it was only 81% as seen in tables 4-5. This difference can be explained by the biologically transitional nature of adolescence during which facial features are not very stable, combined with the fact that there is a larger diversity of ethnicity, lighting conditions, pose variations, and image quality in UTKFace, which makes it more difficult to classify correctly. On the other hand, for adult age groups (28–30), the model achieved F1-scores of 97% on CASIA African Face and 95% on UTKFace, which is indicative of the more prominent and stable facial features in adulthood as well as the narrower age ranges and clearer class boundaries in CASIA that reduce classification ambiguity. The differences in overall performance can mostly be accounted for by the characteristics of the datasets. CASIA African Face has more homogeneous ethnicity and a narrower age distribution, whereas UTKFace is characterized by greater variability in the demographic and imaging conditions, as well as a wider age range. These factors have the most impact on sub-adult and transitional age classes, while adult age classification is still quite consistent across datasets. Moreover, the regression results are a further indication of the strength of the proposed AfriNet model with a mean absolute error (MAE) of 0.65 years on the CASIA dataset and 2.06 years on UTKFace. In addition, there is a classification accuracy of 92% on the latter, which is a strong indication that the model can be generalized well to a more diverse, real-

world facial dataset. Taken together, these results put the spotlight on AfriNet's capability of capturing discriminative facial features for age estimation, confirm that it can be deployed in unconstrained environments, and constitute a significant step forward in the effort to combat demographic bias and increase representation in facial analysis systems. As we see from table 6, AfriNet beats the rest of the architecture by achieving a record time of 34 seconds ahead by 58 seconds from its next rival MobileNet with 92 seconds.

CONCLUSION

In conclusion, the use of AfriNet to identify facial ages with a combination of the CASIA Africa Face and UTKFace datasets presents a powerful instance of how deep learning models can be tweaked to tackle the issue of demographic diversity and eventually have strong practical applications. With the CASIA Africa Face dataset, AfriNet is able to categorize with an accuracy that is close to being perfect (99%) and a very low regression error (0.65 MAE), thereby proving how efficient it is in locating even the most indistinct facial features of the African demographic. The outcomes on the well-known UTKFace dataset are still quite impressive (92% classification accuracy and 2.06 MAE), which is a strong confirmation that the model can be extended to different populations of varying demographic and ethnic backgrounds. These are the kinds of numbers that emphasize the significance of specially designed network architectures and representative datasets as the main factors that lead to the elimination of bias and the further improvement of the accuracy of age prediction for the groups that have been traditionally marginalized. Additionally, the dual approach of treating age prediction both as a classification and a regression task gives the model users the option of whether they want their ages to be classified into categories or if they want to have a more precise age estimation depending on the particular use case scenario. It is the subsequent research that should be carried out on larger and more diverse datasets, the use of more modalities or giving more context, and the incorporation of privacy-protecting methods if the deployment of facial analysis systems in the real world is to be fair, accurate, and ethically sound.

REFERENCES

- [1] G. George, "A Deep CNN technique for Age Estimations from Facial Images," *J. Institutional Res. Big Data Anal. Innov.*, vol. 1, no. 1, 2024, Accessed: Feb. 22, 2025. [Online]. Available: <https://universityjournals.com.ng/index.php/jirbdai/article/view/10>
- [2] S. Farouk, C. Uppin, and G. George, "IoT's Impact on Nigeria's Construction Industry: Unveiling Benefits in the Era of the 4th Industrial Revolution (4IR)," in *2024 International Conference on Science, Engineering and Business for Driving Sustainable Development Goals (SEB4SDG)*, IEEE, 2024, pp. 1–7. Accessed: Feb. 22, 2025. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/10629829/>
- [3] G. George, S. Adeshina, and M. M. Boukar, "The Influence of Activation Functions in Deep Learning Models Using Transfer Learning for Facial Age Prediction," *Int J Intell Syst Appl Eng*, vol. 12, no. 2, 2024, Accessed: Mar. 22, 2025. [Online]. Available: <https://scholar.google.com/scholar?cluster=2970525397019111915&hl=en&oi=scholar>
- [4] D. Sinha and M. El-Sharkawy, "Thin MobileNet: An Enhanced MobileNet Architecture," *2019 IEEE 10th Annu. Ubiquitous Comput. Electron. Mob. Commun. Conf. UEMCON 2019*, pp. 0280–0285, Oct. 2019, doi: 10.1109/UEMCON47517.2019.8993089.

- [5] O. S. Arulogun, I. A. Ogbu, and I. O. Dipeolu, "Influence of internet exposure on sexual behaviour of young persons in an urban district of Southwest Nigeria," *Pan Afr. Med. J.*, vol. 25, p. 261, Dec. 2016, doi: 10.11604/pamj.2016.25.261.2630.
- [6] F. Anda, E. Dixon, E. Bou-Harb, N. A. Le-Khac, and M. Scanlon, "Vec2UAge: Enhancing underage age estimation performance through facial embeddings," *Forensic Sci. Int. Digit. Investig.*, vol. 36, Apr. 2021, doi: 10.1016/J.FSIDI.2021.301119.
- [7] D. Bontempi et al., "FaceAge, a deep learning system to estimate biological age from face photographs to improve prognostication: a model development and validation study," *Lancet Digit. Health*, vol. 7, no. 6, p. 100870, Jun. 2025, doi: 10.1016/j.landig.2025.03.002.
- [8] H. Zhang et al., "Facial age recognition based on deep manifold learning," *Math. Biosci. Eng.*, vol. 21, no. 3, pp. 4485–4500, 2024, doi: 10.3934/mbe.2024198.
- [9] Z. Qawaqneh, A. A. Mallouh, and B. D. Barkana, "Deep Convolutional Neural Network for Age Estimation based on VGG-Face Model," Sep. 06, 2017, arXiv: arXiv:1709.01664. doi: 10.48550/arXiv.1709.01664.
- [10] A. A. A. Aboluhom and I. Kandilli, "Real-time facial recognition via multitask learning on raspberry Pi," *Sci. Rep.*, vol. 15, no. 1, p. 28467, Aug. 2025, doi: 10.1038/s41598-025-97490-6.
- [11] O. Oladipo, E. O. Omidiora, and V. C. Osamor, "A novel genetic-artificial neural network based age estimation system," *Sci. Rep.*, vol. 12, no. 1, p. 19290, Nov. 2022, doi: 10.1038/s41598-022-23242-5.
- [12] B.-C. Chen, C.-S. Chen, and W. H. Hsu, "Cross-Age Reference Coding for Age-Invariant Face Recognition and Retrieval," in *Computer Vision – ECCV 2014*, vol. 8694, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds., in *Lecture Notes in Computer Science*, vol. 8694. , Cham: Springer International Publishing, 2014, pp. 768–783. doi: 10.1007/978-3-319-10599-4_49.
- [13] N. Liu, F. Zhang, and F. Duan, "Facial Age Estimation Using a Multi-Task Network Combining Classification and Regression," *IEEE Access*, vol. 8, pp. 92441–92451, 2020, doi: 10.1109/ACCESS.2020.2994322.