A regression network age estimation method with ordered anchor replacement

[1st] Tian Tan[a], [2nd] Chunlong Hu[b]

School of Computer Science and Engineering, Jiangsu University of Science and Technology, Zhenjiang 212100,Jiangsu,China

[a]954077441@qq.com, corresponding author: [b]huchunlong@just.edu.cn

Abstract. Face age estimation is often regarded as a regression problem, and anchored regression network is a representative regression algorithm. It selects a facial feature from each age group as an anchor point for linear regression, and then combines all the weighted linear regression results to obtain the predicted age. During training, to ensure the comprehensiveness of anchor points, the regression results of all anchor points are included in the calculation. However, some selected anchor points may exhibit significant differences in appearance from the samples, leading to inaccurate age predictions.To address this issue, an improved anchor point selection method is proposed. This method initially employs k-means clustering to output initial anchor points, calculates the Euclidean distance between the initial anchor points and samples to construct a distance vector. Subsequently, based on the distance vector, it identifies several nearest anchor points and replaces the initial anchor points, termed as ordered anchor point replacement.This method enhances the acquisition of anchor points, and experiments conducted on multiple datasets show promising results. The experimental outcomes yield a lower average absolute error (MAE) of 2.54, demonstrating comparability with similar algorithms. Thus, this validates the effectiveness of the proposed method in improving the accuracy of age estimation.

Keywords: Age estimation; Anchored network; Regression network; Convolutional neural network

1 Introduction

Facial age estimation aims to infer a person's age range or precise age based on facial appearance and other relevant information, making it a hot topic in the field of computer vision. Feature extraction methods based on deep learning can capture feature information in facial images at the pixel level, rendering them more robust compared to traditional manual feature extraction methods. As a result, deep convolutional neural networks dominate the feature extraction aspect of age estimation tasks. Following feature extraction, appropriate algorithms or models are employed for age estimation, with common methods including regression, classification, and others.

K. Subramaniyam et al. (eds.), *Proceedings of the 3rd International Academic Conference on Blockchain, Information Technology and Smart Finance (ICBIS 2024)*, Atlantis Highlights in Computer Sciences 21,
https://doi.org/10.2991/978-94-6463-419-8_25

The Anchored Regression Network (ARN) aims to define a set of age anchors, typically representative age values like 10 years, 20 years, 30 years, etc. These anchors serve as a reference framework to calculate the offset of facial images from each age anchor. The offset values are then used to compute weights, which are applied to the regression values for age prediction.

The selection of anchors is a crucial component of ARN[1]. The choice of age anchors can introduce bias into prediction results, with insufficiently comprehensive or inaccurate anchors leading to deviations in predictions. Existing anchored regression methods often include the regression results of all anchors after initialization, even those with significant distance-related differences, introducing bias into the process.

To address this issue, an improved anchor point selection method called the ordered replacement anchor point method is proposed. This method addresses biases caused by distant anchors in existing ARN models by selecting anchor points closest to facial images from the initial set. These selected anchor points are then used to train the model, eliminating the regression results of distant anchors. The advantages of this method include:

1. Improved Prediction Accuracy: By eliminating the regression results of distant anchors, the method reduces the impact of error terms on prediction results, thereby improving the accuracy of facial age estimation.

2. Enhanced Robustness: The ordered replacement anchor point method involves two rounds of anchor selection, ensuring that the anchor points used for regression training are more representative and enhancing the robustness of the model.

3. Reduced Training Parameters: The method filters out a significant portion of anchor points, simultaneously eliminating the corresponding regression layers. This results in a lighter model that retains the advantages of anchor points covering the entire age range.

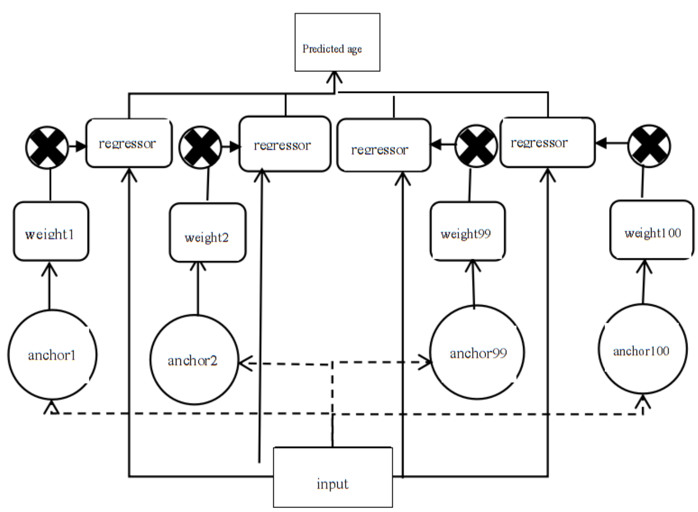

Fig. 1. Anchoring network processes

2 Related Work

2.1 Facial feature extraction

With the emergence of large training datasets and improved computational capabilities, Convolutional Neural Networks (CNNs) have demonstrated outstanding performance in various fields such as image classification, semantic segmentation, and object detection. In the task of age estimation, CNNs learn age features that are more robust than traditional manually crafted feature descriptors, leading to higher accuracy in non-restricted environments.

2.2 Age estimation

Based on the different ways of handling age labels, age estimation methods can be categorized into classification, regression, ranking, and distribution learning. Traditional approaches treat age estimation as either a classification or regression problem. In regression, Liu et al[2]. proposed AgeNet, utilizing regression instead of traditional classification methods to avoid information loss caused by age segmentation. This approach enables the model to learn more advanced features from facial images, thereby improving age estimation accuracy.

Apuandi et al[3]. introduced a method for facial age estimation using Convolutional Neural Networks (CNNs) and Extreme Learning Machine (ELM). CNNs are employed for feature extraction, while ELM is utilized for regression tasks. The combination of ELM and CNN takes advantage of the superior feature learning capabilities of CNNs and achieves efficient regression through ELM, resulting in accurate and efficient age estimation.

Wang et al[4]. innovatively applied the concept of soft stage-wise regression in SSR-Net. By progressively refining age estimation, the model becomes more stable throughout the entire regression process, effectively enhancing age estimation accuracy.

Rothe et al[5]. combined regression and classification tasks, leveraging label information to improve the accuracy and robustness of age estimation. They also introduced techniques such as data augmentation and ensemble learning to enhance the model's generalization capability.

3 Ordered Replacement Anchor Network

3.1 Anchor Network

The given set$\{(x_1, y_1), ..., (x_n, y_n)\}$ consists of n samples, where features $x_i \in$ Rd, xi is the output of a neural network, and label values $y_i \in R^{d'}$.

The fundamental assumption of the anchor regression network is that the relationship between xi and yi can be approximated as a linear mapping on specific partitions in the feature space. There exists a partition set $U_1, ..., U_m \in R^d$, $U_i = R^d$, and if $i \neq j$, then U_i

$\cap$ U_m = $\varnothing$. The space is partitioned using a similarity measure s into m anchor point sets C = {c_1, ..., c_m}, satisfying:

$$U_i = \left\{x \in R^d \middle| \forall j \neq i : s(x, c_i) \rangle s(x, c_j)\right\} \tag{1}$$

Based on the anchor regression network, the age estimation method involves having a corresponding regressor for each anchor point. The function of each regressor is defined as:

$$y_m \approx W_m x_i + b_m, \left(W_m \in R^{d' \times d}\right) \tag{2}$$

In this scenario, the entire regression function can be written as:

$$f_s(x) \approx W_{\gamma_s(x)} x + b_{\gamma_s(x)}, \gamma_s(x) = \mathrm{argmax}_m s（x, c_m） \tag{3}$$

Where γs(x) represents the anchor point closest to x.

In this method, the number of anchor points m is typically set as a tunable parameter. To fully utilize information from nearby age intervals in other anchor point partitions, the Anchor Regression Network (ARN) employs "soft assignment." The input features X undergo a soft assignment module, and the weighted coefficients obtained are used to perform a weighted sum of regression results from k partitions, resulting in the predicted age. As shown in Figure 1,All linear regression results for each anchor point partition regarding the input features X contribute to the weighted sum calculation.

Although ARN adjusts the regression results for different anchor points based on the distance through weighted coefficients, it is noted that distant anchor points still contribute to the calculation of the final predicted age.

3.2 Ordered Replacement Module

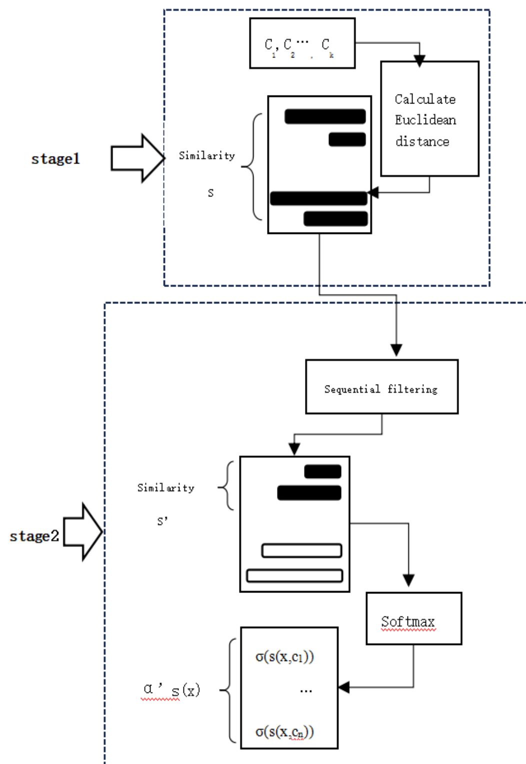

Fig. 2. Flow of orderly replacement module

In response to this drawback, the paper proposes an improved anchor selection method, namely "Ordered Replacement," which consists of two stages, as illustrated in Figure 2.

In Stage One, the Euclidean distance between anchors and samples is computed to represent the "similarity" between anchors and samples. The selection of anchors is determined based on the similarity measure s(x). First, calculate the similarity S of the samples with respect to k anchors, obtaining a set of similarities that lacks regularity and only associates anchors with samples:

$$S\ (x, c_i) = \left\{ s_{x,c_1}, \cdots, s_{x,c_k} \right\} \tag{4}$$

In Stage Two, introduce a tunable parameter n. Arrange the elements in the set S obtained from the previous step in ascending order, and extract the first n elements to obtain a new set of n elements, representing the closest n anchors selected through the second screening:

$$S'\ (x, c_n) = \left\{ s'_{x,c_1}, \cdots, s'_{x,c_n} \right\} \tag{5}$$

The similarity sets S and S' depict black and white line segments, representing the Euclidean distances between samples and each anchor point in the anchor set C. A longer length indicates a greater distance, with black indicating preservation and white indicating exclusion.

Apply softmax to all elements in the set S' obtained after the second screening to obtain weight coefficients $\alpha_{s'(x)}$, used for the weighted sum of n linear regression results. The weight coefficients $\alpha_{s'(x)}$ are defined as:

$$\alpha_{s'}(x) = \sigma(\begin{pmatrix} s'(x, c_1) \\ \vdots \\ s'(x, c_n) \end{pmatrix}) = \frac{1}{\sum_{i=1}^{n} e^{s'(x,c_i)}} \begin{pmatrix} e^{s'(x,c_1)} \\ \vdots \\ e^{s'(x,c_i)} \end{pmatrix} \tag{6}$$

The final predicted age is obtained by weighting the results of the selected n regressors using the calculated weight coefficients:

$$\sim f_{s'(x)} = \sum_{i=1}^{n} \alpha_{s',i}(x)(W_i x + b_i) \tag{7}$$

where $\alpha_{s',i(x)}$ represents the i-th coordinate of $\alpha_{s'(x)}$.

In summary, the proposed Ordered Replacement Anchor Regression Network involves replacing the initial anchors with the nearest n anchors selected through a second screening. These anchors are used to compute weight $\alpha_{s'(x)}$ for linear regression around them. The calculated weight coefficients are then introduced for weighted summation, combining multiple regression quantities in a restricted manner. This approach aims to enhance the anchor regression network by preserving the most similar anchors through a two-step selection process. As shown in Figure 3, the anchored regression network process with the addition of ordered replacement modules

For the loss function of the regression task, the paper adopts the widely used Mean Absolute Error (MAE) as the evaluation metric to ensure consistency between training and testing objectives. The complete loss function under this framework is as follows:

$$L = \sum_{i=1}^{n} | \left| \tilde{f}_s(x_i) - y_i \right| |^2 + \lambda \sum_{j=1}^{m} | \left| W_j \right| |^2 \tag{8}$$

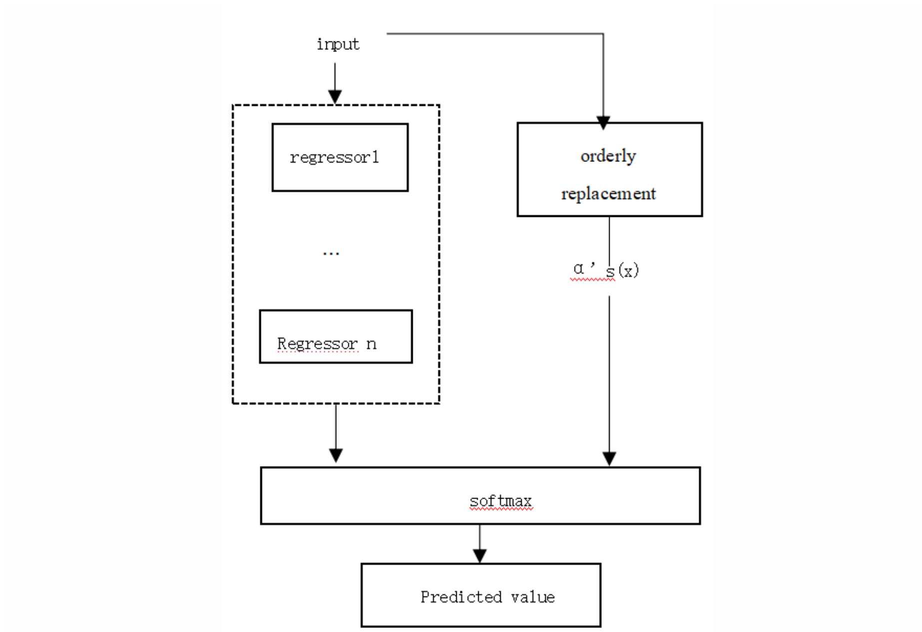

Fig. 3. Anchored Regression Networks under Ordered Replacement Modules

4 Experiments

4.1 Datasets

MORPH-II[6]dataset, as one of the largest publicly available real age datasets to date, comprises over 55,000 facial images of approximately 1,300 individuals of various genders and ethnicities.Wiki-imdb[7] dataset draws its images from Wikipedia and IMDB, widely used resources that provide real-world facial images.With around 500,000 facial images, it covers various crucial aspects.MegaAge-Asian is a large-scale facial recognition dataset containing over 1 million images of Asian faces.FG-NET consists of 1002 facial images from 82 subjects, with significant variations in background, color tones, and lighting conditions. FG-NET is a relatively small-scale facial age dataset.

4.2 Experimental environment

CPU:AMD Ryzen7 5700X, GPU:NVIDIA RTX 3060, and 16GB of memory. The software environment was set up on a Windows 10 system, using Python 3.8 as the deployment environment, PyTorch 1.13 framework for deep learning architecture, and CUDA 12.0.89 for accelerated computing.

The experimental parameters are as follows: the neural network optimization algorithm utilizes the Adam optimization algorithm with an initial learning rate of

0.001, momentum of 0.9, and weight decay of 2×10^{-4}. The number of epochs is set to 200, and the batch size is 32.

4.3 Evaluation Metrics

This paper adopts commonly used evaluation metrics in the field of age estimation, namely Mean Absolute Error (MAE) and Cumulative Score (CS). MAE calculates the absolute value error between the true age and the estimated age, and its computation is as follows:

$$\mathrm{MAE} = \frac{1}{N}\sum_{k=1}^{N}\left| l_k - \hat{l}_k \right| \tag{9}$$

Where l_k is the actual age of the k-th image, and $\hat{l}_k$ is the estimated age of the k-th image.A lower MAE indicates lower average error and better model performance.Cumulative Score (CS) is calculated as follows:

$$\mathrm{CS}(j) = \frac{N_{e \le j}}{N} \times 100\% \tag{10}$$

Where N is the total number of images in the test set, and N_e<=j represents the number of images with absolute age estimation errors within the range j. A higher CS value indicates better model performance.

4.4 Results

4.4.1 Results on MORPH-II.

To validate the effectiveness of the Ordered Replacement Module under the anchor regression network, this study conducted ablation experiments focusing on anchor point selection, and the specific experimental results are shown in Table 1.

In Table 1, when the parameter k is set to 100, the network structure is equivalent to ARN. Under this parameter, the performance is poor when using only initial anchor points for regression, with a MAE of 3.0, influenced by too many anchor points. Experimental results indicate that ARN's performance significantly declines due to the impact of errors caused by excessive anchor points. When the Ordered Replacement Module is introduced, the MAE improves to 2.548. This improvement is attributed to filtering out most anchor points from the initial set, eliminating errors caused by less relevant anchor points, and enhancing the weights of results from effective anchor point regression. Consequently, the network's ability to accurately represent the target is strengthened, leading to a 10% increase in testing accuracy.

The use of the Ordered Replacement Module for filtering initial anchor points effectively enhances the prediction accuracy of the network on the Morph II dataset, outperforming traditional anchor regression networks. This demonstrates the strong effectiveness of this method in improving anchor point selection.

Furthermore, since the network needs to set a corresponding regression layer for each anchor point, the original ARN method's regression layer quantity is consistent with the initial anchor point quantity. In contrast, this method's anchor points are obtained after further filtering the initial set, significantly reducing the required regression layer quantity compared to ARN. This results in a significant reduction in the number of parameters during training, leading to a notable improvement in training efficiency.

Table 1. Comparison of experimental results with ARN on the MegaAge Asian dataset

method	Adjustable parameters k	Adjustable parametersn	MAE
ARN	100	-	3.00
ARN+OR(Orderly replacement)	50	10	2.572
		9	2.582
		8	2.588
		7	2.593
		6	2.566
		5	2.548
		4	2.556
		3	2.637
		2	2.688

In order to evaluate the predictive accuracy of the network, this study conducted a comparison with current mainstream age estimation methods across multiple aspects, including MAE, parameter count, and CS. The specific experimental data is presented in Table 2.

As shown in Table 2, the proposed method achieved a MAE accuracy of 2.54. Compared to heavyweight methods with parameters several hundred times larger, this method outperformed DEX (Deep EXpectation)[5], RankingCNN[8], Deng et al[10]., Posterior[11], and other methods. While methods like Posterior exhibit high accuracy, they suffer from drawbacks such as large network parameters, long training times, and significant storage space consumption. In contrast, the proposed method demonstrated competitive performance with short training times and minimal space requirements.

In comparison to the lightweight C3AE model, the MAE of the proposed Ordered Anchor Replacement Network showed a 7% improvement. This improvement was achieved while attempting to minimize parameter count, resulting in a lower MAE. Additionally, integrating the well-tuned secondary screening module into the anchor network led to a 15.3% increase in the model's MAE.

Therefore, the adoption of the Ordered Replacement Module for reselecting anchors in the anchor regression network demonstrates the effective enhancement of the network's predictive accuracy on the Morph II dataset.

Table 2. Comparison of experimental results of different methods on the MORPH-II dataset

method	Pre-training set	MAE	Parameters/10^3	CS(5)/%
RankingCNN[8]	Audience	2.96	500000	83.7
C3AE[9]	IMDB-WIKI	2.75	39.7	65.0
SSR-Net[20]	IMDB-WIKI	3.16	40.9	76.3
ARN[1]	IMDB-WIKI	3.00	138000	83.1
DEX[5]	ImageNet	3.25	138000	66.9
DEX[5]	IMDB-WIKI	2.68	138000	78.1

Posteriror[11]	IMDB-WIKI	2.52	138000	80.4
Deng et al.[10]	ImageNet	2.97	138000	84.1
ARN+OR(Orderly replacement)	IMDB-WIKI	2.54	6900	82.1

In order to further validate the effect of ordered anchor points on the anchor regression network, we conducted ablation experiments on the parameters k and n in Equation (4). Here, k represents the number of initial anchor points, and n represents the number of anchor points after replacement. On the MORPH-II dataset, we tested various values of k and n, and the change trend of MAE is shown in Figures 4 and 5.

From the graphs, it can be observed that when the replaced anchor point number is 50, MAE decreases initially with an increase in the number of initial anchor points, as adding anchor points significantly improves the model's performance when the number of anchor points is low. However, as the number of anchor points becomes sufficiently comprehensive, further increasing anchor points has a diminishing impact on task improvement compared to the introduction of errors, resulting in a subsequent increase in MAE. The model performs optimally when k is 50.

When the initial anchor point number is 50, MAE shows a trend of decreasing initially and then increasing as the number of secondary screening anchor points (n) decreases. During the replacement process, when n is small, removing too many effective anchor points can lead to a decline in model performance. However, when n is 5, the model performs optimally.

By adjusting parameters k and n, we can find the optimal number of initial and replaced anchor points to achieve the best model performance. These experimental results further validate the effectiveness of anchor point replacement for the anchor regression network.

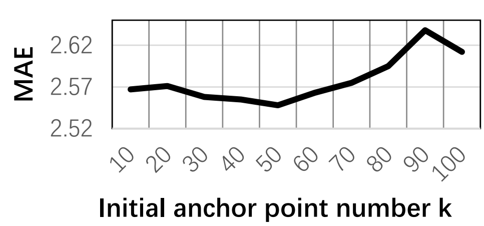

Fig. 4. MAE changes under different initial anchor points

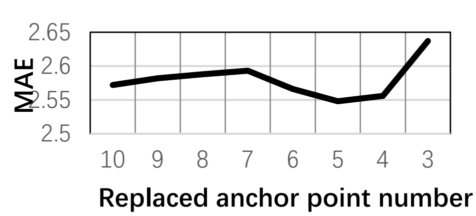

Fig. 5. MAE changes under different replaced anchor points

4.4.2Results on FG-Net.

On the FG-Net dataset, we employed the same experimental setup as in the MORPH-II dataset. As shown in Table 3, the proposed method achieved a MAE accuracy of 3.47, ranking slightly lower than DOEL3groups[16] (3.44) and ALD-Net[19] (3.25) among the methods presented. However, it's important to note that both DOEL3groups and ALD-Net require substantial additional training. While they exhibit higher accuracy, they suffer from drawbacks such as large network parameters and long training times. In comparison, our proposed method has a 50% reduction in parameters compared to DOEL3groups, with only a 0.87% increase in MAE. When compared to ALD-Net, our method reduces parameters by 68%, with only a 6.7% increase in MAE. This indicates that our approach significantly enhances age estimation accuracy while notably reducing parameter count.

Table 3. Comparison of experimental results on the FG-Net datase

method	MAE
DOEL3groups[16]	3.44
MV-Loss[17]	4.10
DEX[5]	4.30
Group-n[18]	2.96
ALD-Net[19]	3.25
ARN+SP(Secondary Partition)	3.47

4.4.3Results on Megaage-Asian.

Due to differences in facial age features among different ethnicities, using datasets trained on facial images of Caucasians and Africans to estimate the age of other ethnicities may introduce biases. To address this issue, we utilized the Ordered Replacement Module and retrained our age estimation model on the MegaAge-Asian and FG-Net datasets.

On the MegaAge-Asian dataset, we employed the same experimental setup as in the Morph II dataset and set the initial anchor point and post-filtering anchor point values to the optimal values of 50 and 10, respectively. As shown in Table 4,the experimental results showed that our method achieved results of 63.4% and 82.1% on the CS3 and CS5 metrics, respectively. Compared to lightweight methods such as CEN[12]and LRN[13], our approach demonstrated competitive accuracy. Additionally, when compared to the large-scale method Posterior, our method achieved comparable performance. This suggests that our approach exhibits good comprehensiveness and accuracy in handling age prediction tasks.

Table 4. Comparison of experimental results on the MegaAge

method	Pre-training set	CS(3)/%	CS(5)/%
Posteriror[13]	IMDB-WIKI	62.1	80.4
CEN[12]	IMDB-WIKI	63.7	82.9
LRN[13]	IMDB-WIKI	64.4	82.9
SSR-Net[20]	IMDB-WIKI	54.9	74.1
Dense-Net[14]	IMDB-WIKI	51.7	69.4

MobileNet[15]	IMDB-WIKI	44.0	60.6
ARN+SP(Secondary Partition)	IMDB-WIKI	63.4	82.1

5 Conclusions

This paper introduces a regression network method for age estimation using ordered anchor point replacement. Building on the advantages of segmented linear models in anchor regression networks, this approach mitigates interference from distant anchor points in predicting results. Experimental results validated the superiority of this method on the Morph II, FG-Net, and Megaage-Asian datasets, but there is still room for improvement. For instance, the strategy combining anchor points and regression may be influenced by factors like facial expressions and lighting, leading to some uncertainty in prediction outcomes. To further enhance the model's predictive accuracy and robustness, future research will delve into more in-depth studies and discussions, exploring effective strategies that are better suited for complex environments.

References

1. Agustsson E, Timofte R, Van Gool L. Anchored regression networks applied to age estimation and super resolution[C]//Proceedings of the IEEE international conference on computer vision, 2017: 1643-1652.
2. Liu X, Li S, Kan M, et al. Agenet: Deeply learned regressor and classifier for robust apparent age estimation[C]//Proceedings of the IEEE International Conference on Computer Vision Workshops, 2015: 16-24.
3. Apuandi I, Rachmawati E, Kosala G. ConvELM: Exploiting Extreme Learning Machine on Convolutional Neural Network for Age Estimation[C]//2023 International Conference on Artificial Intelligence in Information and Communication (ICAIIC). IEEE, 2023: 407-412.
4. Yang T Y, Huang Y H, Lin Y Y, et al. Ssr-net: A compact soft stagewise regression network for age estimation[C]//IJCAI, 2018, 5(6): 7.
5. Rothe R, Timofte R, Van Gool L. Dex: Deep expectation of apparent age from a single image[C]//Proceedings of the IEEE international conference on computer vision workshops, Santiago, Chile, 2015: 10-15.
6. Ricanek K, Tesafaye T. Morph: A longitudinal image database of normal adult age-progression[C]//7th international conference on automatic face and gesture recognition (FGR06). IEEE, 2006: 341-345.
7. Nga C H, Nguyen K T, Tran N C, et al. Transfer learning for gender and age prediction[C]//2020 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-Taiwan). IEEE, 2020: 1-2.
8. Chen S, Zhang C, Dong M, et al. Using ranking-CNN for age estimation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition, 2017: 5183-5192.
9. Zhang C, Liu S C, Xu X, et al. C3AE: exploring the limits of compact model for age estimation[C]//Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2019: 12587-12596.
10. Deng Y, Teng S, Fei L, Zhang W, Rida I. A multifeature learning and fusion network for facial age estimation. Sensors, 2021.

11. Zhang Y, Liu L, Li C. Quantifying facial age by posterior of age comparisons. Proceedings of British Machine Vision Conference (BMVC), 2017.
12. Li P P, Hu Y B, He R, et al. A coupled evolutionary network for age estimation［EB/OL］,［2022-03-26］. https：//arxiv. org/pdf/1809. 07447. pdf..
13. Li P, Hu Y, Wu X, et al. Deep label refinement for age estimation[J]. Pattern Recognition, 2020, 100: 107178.
14. Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition, 2017: 4700-4708.
15. Howard, Andrew G, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications, 2017.
16. Xie J C, Pun C M. Deep and Ordinal Ensemble Learning for Human Age Estimation From Facial Images[J].IEEE Transactions on Information Forensics and Security, 2020, 15:2361-2374.
17. Liu N, Su Z Y, Chai Y T, et al. Feedback neural network for constrained bi-objective convex optimization [J]. Neurocomputing, 2022, 514(Dec.1):127-136.
18. Tan Z, Wan J, LeI Z , et al. Efficient Group-n Encoding and Decoding for Facial Age Estimation.[J].IEEE transactions on pattern analysis and machine intelligence, 2018, 40(11):2610-2623.
19. He J H, Hu C L, Wang L J. Facial age estimation based on asymmetrical label distribution[J]. Multimedia Systems, 2023, 29:753–762.
20. Yang T Y, Huang Y H, Lin Y Y, et al. Ssr-net: A compact soft stagewise regression network for age estimation[C]//IJCAI, 2018, 5(6): 7.