Research on Phased Single Aggregation Age Estimation Algorithm Based on VoVNetv4

Huan Zhang, Tianwei Shi, Wenhua Cui, Ye Tao, Jianfeng Zhang, Yulu Liu

Abstract—**Aiming at the problems of low recognition accuracy and large calculation amount of age estimation, this paper proposes a staged single aggregation backbone network VoVNetv4. In order to improve the feature extraction ability of the network, the Regional One-Shot Aggregation (ROSA) module is proposed in the network for regional encapsulation of the feature module. The classifier adopts the COnsistent RAnk Logits (CORAL) framework, which can solve the problem of inconsistency of classifiers in ordered regression CNNs for multiple binary classification tasks. This paper conducts experimental verification on the Asian Face Age Dataset (AFAD) and compares it with other research methods. The experimental results show that VoVNetv4 can effectively improve the feature extraction and detection ability of the network and improve the detection efficiency of age estimation.**

Index Terms—**AFAD, Age Estimation, ROSA, VoVNetv4**

I. Introduction

With the continuous development and progress of society, age estimation based on deep learning has gradually become the focus of human facial representation detection. Age estimation can be widely combined with applications in other fields, and has a profound impact on the development of society. It has important application value in human-computer interaction and video surveillance [1-2]. Since the degree of maintenance of facial information will greatly affect the discrimination of age, there are often some errors in the detection information. Facial expressions and the shape of the skull will affect the detection results. Age estimation has always been a difficulty in detection [3]. The identification of human age is usually studied in two aspects: classification and regression. Classification is divided into stages according to the age of the human body, so as to identify and classify the age of each stage. Since the age of the human body is an ordered sequence that is constantly changing, age estimation based on regression is also a mainstream research direction [4]. In recent years, domestic and foreign researchers have improved and studied age estimation algorithms in classification and regression, and a large number of research results have emerged.

Li et al. proposed a refinement network LRN [5], which encapsulates the label distribution and regression essence at the same time, and uses the relaxation regression refining field to discriminate age; Dagher et al. pre-trained the CNN network, fine-tuned VGG, Res-Net and other networks, and proposed a classification network composed of two types of CNNs based on the fine-tuned network for age estimation [6]; Based on deep learning and directed acyclic graph support vector machines, Xiao et al. made local adjustments to the algorithm of age estimation, and set a specific neighborhood around the global estimated value to perform accurate age estimation [7]; Guehairia et al. used the Gcforest algorithm to perform age estimation experiments on the basis of images [8], which has the advantage of a cascade structure that allows interaction between trees; Badr et al. proposed a cascaded model system that learns the segmentation of age labels through a classification model, and uses the knowledge learned by the classification model as an auxiliary input to a regression model to achieve age estimation [9]. Zhang et al. [10] used regression forest for age estimation of face images, combined with head pose, and proposed a combined system of age estimation and head pose estimation. Combining the convolutional neural network framework and resnet50 architecture, Pramanik et al. proposed a shortcut strategy for age estimation from face images [11]. Chang et al. proposed an ordinal hyperplane sorting algorithm of OHRank [12], which discriminates human age based on the relative order information between age tags in the database. Many researchers have conducted a lot of experiments on the AFAD. For example, Wang et al. [13] combined course learning with age estimation to improve the training efficiency of neural networks, thereby improving the network's ability to discriminate age; Cao et al. [14] proposed a CORAL framework for age estimation classification, using Resnet as a feature extraction network, which can obtain a rank monotonic threshold model without relying on explicit weighting terms for each training example. Niu et al. [15] made a breakthrough in the traditional research ideas of classification and regression in age estimation. They used sequential regression to simultaneously perform feature

Manuscript received November 28, 2022; revised April 21, 2023.

This work was supported by the Natural Science Foundation project of Liaoning Province (2021-KF-12-06), Department of Education of Liaoning Province (2020FWDF01), and Project of Liaoning BaiQianWan Talents Program.

Huan Zhang is a postgraduate student of School of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan, 114051, China. (e-mail: hzzhbest@163.com).

Tianwei Shi is an associate Professor of School of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan, 114051, China. (corresponding author, phone: 139-9805-3962; e-mail: tianweiabbcc @163.com).

Wenhua Cui is a Professor of School of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan, 114051, China. (e-mail: cwh@systemteq. net).

Ye Tao is a lecturer of School of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan, 114051, China. (e-mail: taibeijack @163.com).

Jianfeng Zhang is a postgraduate student of School of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan, 114051, China. (e-mail: jianfengzhang177@163.com).

Yulu Liu is a postgraduate student of School of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan, 114051, China. (e-mail: lyl66289290@163.com).

learning and regression modeling. On the basis of deep learning, they used convolutional neural network. End-to-end learning specifically analyzes the regression problem, so that the effect of age estimation can be effectively improved. Although these methods can reduce the negative impact of individual differences to a certain extent, due to the complexity of face information and the influence of changes in the environment in which the face is located. Age estimation still has the problems of large amount of calculation and low recognition accuracy.

To solve the above problems, this paper proposes a new object detection backbone network VoVNetv4 for age estimation. VoVNetv4 uses regional splicing to transfer early features, completes the capture of various visual information across latitudes, and realizes staged single aggregation of features. Build an Adaptive Stage module consisting of Conv layer, BN layer and Relu activation function for feature smoothing. The classifier adopts the CORAL framework to divide tasks in binary form and make consistent predictions for each task. This paper has carried out experimental verification on AFAD, its MAE is 3.30, RMSE is 4.64, the detection effect is very obvious.

II. Methodology

A. Architecture

Since the low-level features and high-level features have different detailed information, the low-level features contain relatively more feature information and detailed information, but have relatively low semantics and strong noise. High-level features are less sensitive to details and have lower resolution than low-level features, but have stronger semantic information. The process of feature aggregation is the efficient fusion of low-level features and high-level features, which is an important means to improve model performance. VoVNetv4 uses module splicing to transfer early features, so that the original feature patterns can be preserved. By gathering multiple receptive field blocks, it captures various visual information in a cross-latitude way, so that its extracted features can achieve diversified representation. The VoVNetv4 network consists of 4 ROSA modules and Adaptive Stage modules. In each ROSA module, all features are connected once, and the module output is aggregated once. The output information of ROSA is processed by the Adaptive Stage module, and finally aggregated at one time. VoVNetv4 network can effectively avoid feature redundancy and improve the detection accuracy of the network. The overall architecture of the network is shown in Figure 1.

The features extracted from each layer are first passed through the feature smoothing module, the size of the feature matrix is adjusted, and then spliced to the final output layer of the network. The ROSA module is composed of 12 layers, and uses 12 layers of convolution stacking to process related features in a staged single-aggregation manner. The Adaptive Stage module contains convolutional layers, Batch Normalization (BN) and Recitified Linear Unit (Relu). By adding a BN layer after the convolutional layer to normalize the data, the data can be prevented from being unstable due to the large data before entering Relu. Relu is more expressive for linear functions, especially in deep networks. For nonlinear functions, since the gradient of Relu is constant in the non-negative interval, there is no gradient disappearance problem, so that the convergence rate of the model is maintained in a stable state. Compared with Resnet's residual network, it reduces the repeated operation of features between layers, and the output of each layer does not pass through all subsequent intermediate layers, which keeps the input size of the intermediate layers unchanged.

B. ROSA Stage

In target detection, most networks use ResNet[16] and Densnet[17] as the Backbone network. Densnet uses the ResNet network for reference and densely connects the convolutional layers to ensure the information flow between the convolutional layers. Since the dense connection brings feature enhancement and also brings the drawback of linear growth of the output channel, the VoVNet [18] network has made a lightweight improvement for this drawback, and the dense connection is transformed into all feature aggregation in the last layer. Although the VoVNet network effectively solves the problems of complexity and memory access cost of the original network, the recognition accuracy has not been significantly improved. The VoVNetv4 proposed in this paper draws on the improved idea of VoVNet, and proposes the ROSA module to perform a single aggregation of staged features. The aggregation calculation of VoVNet is shown in Figure 2.

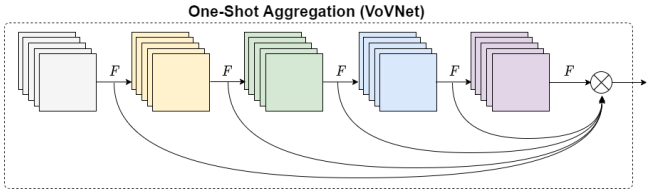

Fig.2: Aggregate computation of VoVNet

The ROSA module retains the single aggregation calculation of VoVNet and encapsulates the VoVNet region. Each VoVNet network is defined as a ROSA region, and VoVNetv4 performs a second single aggregation calculation for these regions.

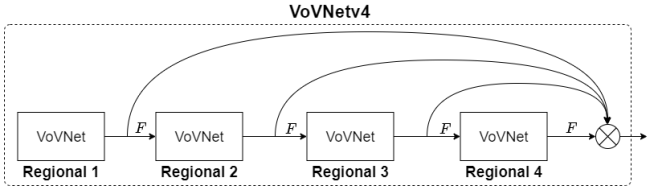

Fig.3: Aggregate computation of VoVNetv4

The ROSA module uses a staged single aggregation design module architecture for aggregation operations. In the ROSA module, regions are divided for different stages of operation, each region is used as a stage of operation, and the ROSA module is composed of four regions. Each region contains 5 identical ConvBlocks with the same input and output channels. ConvBlocks include a 3×3 convolutional layer, a BN layer and a Relu layer, and the stride of the convolution kernel is 1. During the aggregation operation, only one cycle is performed for region 1 and region 2, four cycles are

performed for region 3, and three cycles are performed for region 4. Whenever entering the ROSA Stage of the next stage, the feature map goes through a downsampling with a convolution kernel of 1 and a stride of 1, and a 3×3 max-pooling layer with a stride of 2. Since the high-level semantic information is more important in the target detection task than the low-level semantic information, the ROSA module adds high-level features through different region operations to improve the ratio of high-level features to low-level features.

C. Adaptive Stage

Adaptive Stage is a feature smoothing module, which is used to adjust the size of the feature matrix output by the ROSA module. Adaptive Stage consists of Conv layer, BN layer and Relu activation function. Feature smoothing is a very important link in target detection. Compared with age estimation, the research problem of this paper, the facial data of human body is usually diversified and has a large amount of noise. Using the Adaptive Stage module for smoothing can effectively alleviate the influence of adverse factors on the detection results, thereby improving the detection accuracy of the classifier.

The Conv layer can adaptively adjust the size and stride of the convolution kernel according to the size of the input image and the set output image size. Since the ROSA module adopts a cross-latitude feature fusion method, there are a large number of parameter updates in the network training between each layer, and the BN layer is used to solve the distribution change between the data in the middle layer during the training process. The application of the BN layer can effectively improve the generalization ability of the network model. Because the features are normalized, the network will adapt to a larger learning rate, which plays an important role in improving the training speed of the network model, and can reduce the network to Initialized dependencies. The BN layer has the function of transformation and reconstruction, which can reconstruct the normalized value again. The calculation process is as follows:

$$y^{(k)} = \gamma^{(k)} \hat{x}^{(k)} + \beta^{(k)} \tag{1}$$

Where x is the mini-batch input feature, k is the channel, γ and β are the learnable parameters introduced by BN, respectively. The BN layer needs to perform variance operation and mean operation on all the values in x. For each neuron, there will be γ and β parameters, as shown in formula 2.

$$\gamma^{(k)} = \sqrt{Var[x^{(k)}]}\ ,\ \beta^{(k)} = E[x^{(k)}] \tag{2}$$

Since both γ and β are learnable parameters, the learned features in a certain layer can be recovered, and thus the original feature distribution to be learned can be learned.

$$\hat{x}_i = \frac{x_i - \mu}{\sqrt{\sigma^2 + \varepsilon}} \tag{3}$$

Where μ is the sample mean, σ^2 is the variance of the sample, ε is the operation to prevent the operation from being invalid. When σ^2 is not 0, then ε is 0. The Relu activation function is used to fit the training data. It is a one-sided inhibition function, that is, all negative values are 0, and positive values do not change. The Relu activation function has strong sparsity, so it can greatly promote the mining of relevant features of the model.

D. CORAL framework

The age estimation scheme proposed in this paper includes three stages: image preprocessing, feature extraction and age classification. CORAL framework is an ordered classifier in the age classification stage. How to make the difference between the same categories smaller and the difference between different categories larger is the key to whether a good classification result can be achieved. The CORAL framework converts ordinal targets into binary classification subtasks, which can effectively improve the classification effect. The CORAL framework derivation process is as follows:

$$D = \{x_i, y_i\}_{i=1}^{N} \tag{4}$$

Among them, D is the dataset; x_i is the first picture; N is the number of samples; y_i is the corresponding *rank* value, and *rank* is the ranking of a person's age in a certain area; Among them, y_i is in the set to which the rank belongs, and if the set is Y, the relationship between y_i and Y is shown in formula 5.

$$y_i \in Y = \{r_1, r_2, \ldots r_k\} \tag{5}$$

The elements contained in the Y set are in an ordered arrangement, as shown in formula 6.

$$r_k \succ r_{k-1} \succ \ldots \succ r_1 \tag{6}$$

The main purpose of the ordered regression task is to find an ordering rule, that is, the corresponding relationship between the age picture and the age value in the *rank*, so that its loss function reaches the minimum value. Let C be a $K \times K$ cost matrix, C_{y,r_k} represents the loss value of predicting a sample (x, y) as $rank(r_k)$ .When $C_{y,y} = 0$, the picture information representing the age corresponds exactly to the age value in the *rank*, and the network model presents a perfect prediction state. When $y \neq r_k$, then $C_{y,r_k} > 0$. In ordinal regression, the V-shaped cost matrix is more conducive to the learning and classification of features, that is, when $r_k \leq y$ and $C_{y,r_{k-1}} > C_{y,r_k}$ or $r_k \geq y$ and $C_{y,r_k} < C_{y,r_{k+1}}$, the cost matrix presents a perfect V-shaped state.

In the actual operation process, it is difficult to make the cost matrix reach the V-shaped state. The CORAL framework can produce consistent predictions for each binary task, avoiding the drawbacks of the V-shaped matrix. By performing binary label expansion on y_i, as shown in Equation 7.

$$y_i^{(1)}, \ldots, y_i^{(K-1)} \tag{7}$$

Where $g(x_i, W)$ indicates whether y_i exceeds $rank(r_k)$, and the consistency of prediction is guaranteed according to the response mechanism of binary tasks. The minimized loss function used for model training is shown in Equation 8.

$$L(W,b) = -\sum_{i=1}^{N}\sum_{k=1}^{K-1}\lambda^{(k)}\begin{bmatrix}\log(\sigma(g(x_i,W)+b_k))y_i^{(k)} \\ +\log(1-\sigma(g(x_i,W)+b_k))(1-y_i^{(k)})\end{bmatrix} \quad (8)$$

Where $g(x_i,W)$ is the output of the penultimate layer of the network, $\{g(x_i,W)+b_k\}$ is the input of the corresponding binary classifier in the last layer, and $\sigma(g(x_i,W)+b_k))$ is the empirically predicted probability of task k.

III. Experimental

A. Datasets

The AFAD [19] contains 164,432 images of human faces, each with a corresponding human age and gender label. The AFAD dataset is collected from Renren. Its image information contains 100,752 male pictures and 63,680 female pictures. The minimum age is 15 years old and the maximum age is 40 years old. This dataset is mainly for age estimation of Asians, and the dataset contains different backgrounds and different lighting conditions.

B. Experimental setting

The operating system used in this experiment is Window10 64-bit, and the video memory is 12G. The CPU is Intel(R) Xeon(R) CPU E5-2630 v4 @ 2.20GHz 2.20 GHz (2 processors), and the graphics card is NVIDIA TITAN X(Pascal). Python uses version 3.8 and Pytorch uses version 1.9.

C. Evaluation Metrics

To test the effectiveness of the proposed Backbone network, this paper evaluates the performance of VoVNetv4 on two evaluation metrics, Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). MAE is mainly used to calculate the average value of the difference between the predicted age of the human body and the real age, that is, the average solution of the residuals. The smaller the MAE value, the higher the accuracy of age estimation. The MAE expression is shown in Equation 9.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|y_i - h(x_i)\right| \quad (9)$$

RMSE is used to measure the deviation between the predicted value of age and the true value, and its calculation process is similar to the calculation of standard deviation, as shown in Equation 10.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - h(x_i))^2} \quad (10)$$

Among them, y_i is the age value predicted by the network model, $h(x_i)$ is the real age value of the predicted sample, and N is the total number of samples.

D. Experimental results and analysis

Use the same random seed in training for a fair comparison. The experiment is divided into four parts. The first part is to use the standard Resnet-34 classification network as the performance baseline model. The second part is to integrate the standard Resnet-34 classification network with the CORAL framework to verify the effectiveness of the CORAL framework. The third part is the fusion of VoVNet and CORAL framework to verify the performance of VoVNet. The fourth part is to improve VoVNet, propose VoVNetv4, and integrate with CORAL framework to verify the effectiveness of the proposed method. The random seed is set to 1, the learning rate is 0.0005, the epoch is 200, and the batchsize is set to 128.

In order to verify the effectiveness of the CORAL framework, Resnet-34 is selected as the feature extraction network for verification and analysis. In the first part, the Resnet-34 network is used as a classification network for age estimation, and the convergence of the network is analyzed by analyzing the change trend of loss value. The network training loss diagram is shown in Figure 4, and the network gradually shows a convergence state around 37 rounds.

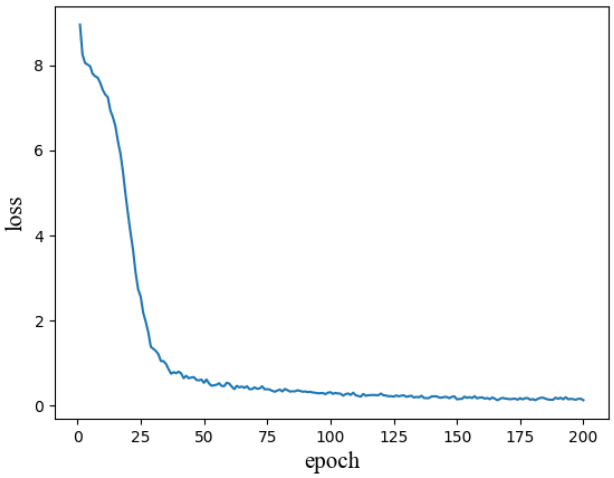

Fig.4: Resnet-34 network training Loss curve

The second part of the experiment is based on the first part of the experiment, and integrates the CORAL framework classifier to verify the effectiveness of the CORAL framework. In this experiment, the cost function is used to optimize the network. The network convergence is shown in Figure 5. The model gradually shows a convergence trend around 25 rounds. The experiment shows that the CORAL framework classifier can effectively improve the generalization ability of the network.

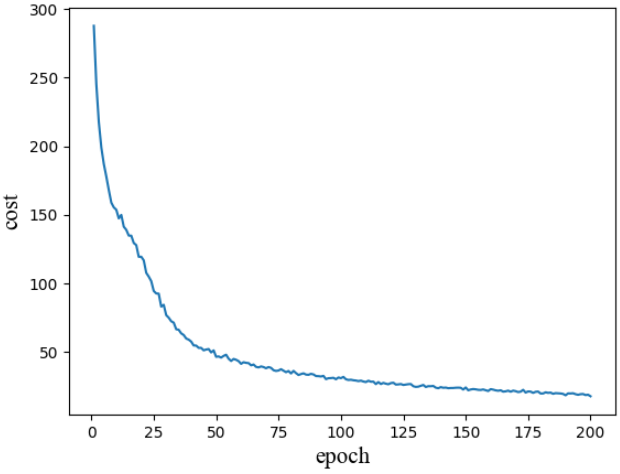

Fig.5: Cost function curve of Resnet-34+CORAL framework

The third part is used to verify the effectiveness of the VoVNet network. In this experiment, VoVNet is used as the feature extraction network, and the CORAL framework is used as the classifier. The network convergence is shown in Figure 6. The experimental results show that compared with Figure 5, the convergence of the network in the 25th round is significantly improved. Compared with Resnet-34 as the feature extraction network, its convergence speed and convergence effect are significantly improved.

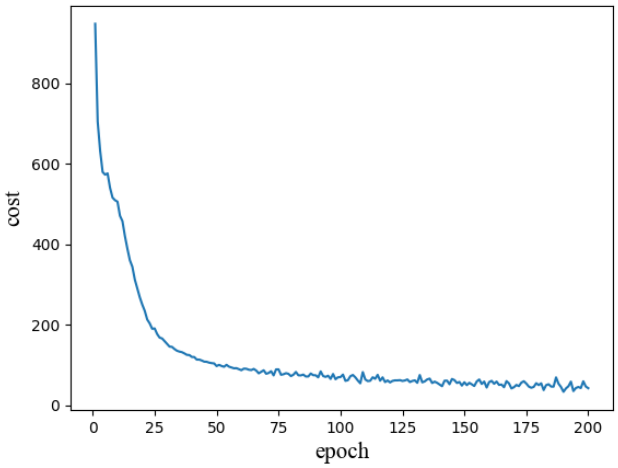

Fig.6: Cost function curve of VoVNet+CORAL framework

In the third part of the experiment, the effectiveness of using VoVNet as a feature extraction network is verified. In the fourth part of the experiment, the VoVNet network is improved and the backbone network VoVNetv4 is proposed. In order to verify the performance of the proposed network, VoVNetv4 is used as the feature extraction network in this experiment, and the CORAL framework is used for the classification network, and its training cost function curve is shown in Figure 7.

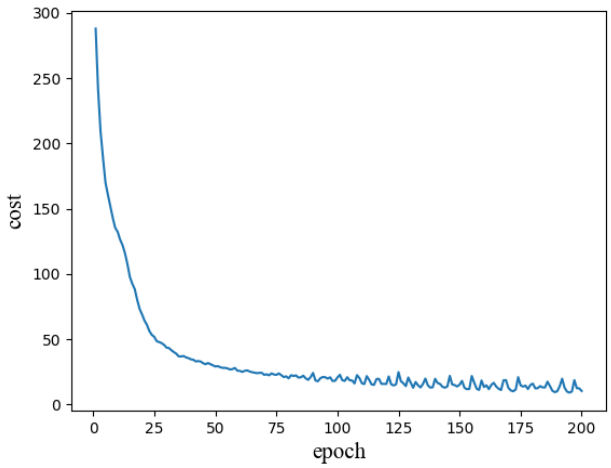

Fig.7: Cost function curve of VoVNetv4+CORAL framework

Table I shows the experimental data of the four stages. From the data in the table, it can be seen that the MAE value of the method proposed in this paper has dropped by 0.41 compared with Resnet-34; compared with Resnet-34 combined with CORAL framework, the MAE value has dropped by 0.14; Compared with VoVNet as a feature extraction network, it has dropped by 0.06. Compared with other methods, the method in this paper has a significant decrease in RMSE value, and the maximum decrease is 0.51. Using VoVNetv4 as the feature extraction network, the final experimental results of the classifier using the CORAL framework are: the MAE value is 3.30, and the RMSE value is 4.64.

TABLE I
EXPERIMENTAL RESULTS ON THE AFAD DATASET

Approaches	MAE	RMSE
Resnet-34	3.71	5.15
Resnet-34+CORAL	3.44	4.69
VoVNet+CORAL	3.36	4.74
Ours: VoVNetv4+CORAL	3.30	4.64

This paper reproduces the research methods of literature [13], literature [14] and literature [15]. That is, under the same AFAD dataset, the number of training rounds is 200, and the evaluation indicators MAE and RMSE are used for performance analysis. The experimental results show that compared with the With Curriculum Learning and CORAL-CNN methods, the MAE value of the method proposed in this paper has decreased by 0.17; compared with the OR-CNN method, the MAE value has decreased by 0.21. Compared with other methods, the method used in this paper has different degrees of reduction in the value of RMSE. The comparison experiment results are shown in Table II.

TABLE II
COMPARISON WITH OTHER METHODS IS BASED ON AFAD DATASET

Approaches	**MAE**	**RMSE**
With Curriculum Learning [13]	3.47	5.03
OR-CNN [15]	3.51	4.75
CORAL-CNN [14]	3.47	4.71
Ours: VoVNetv4+CORAL	3.30	4.64

IV. CONCLUSION

This paper proposes a new Backbone network VoVNetv4 for age estimation. VoVNetv4 is based on VoVNet and proposes a ROSA module for staged single aggregation. The feature extraction capability of the network has been improved while retaining the advantages of the original model in terms of high speed and small parameters. An adaptive stage module for feature smoothing is established in the model, and the feature matrix extracted from each layer is spliced to the output layer through the adaptive stage, so that the feature matrix of each layer reaches the output layer without passing through the intermediate layer. While reducing the input of the middle layer, the input size of the middle layer is kept unchanged. The classifier adopts the CORAL framework to solve the problem of inconsistency of the classifiers in the ordered regression based on multiple binary classification tasks, thereby improving the classification accuracy of the network. This paper conducts

experiments on the AFAD. Compared with other methods, the model has significantly improved in terms of accuracy and speed.

REFERENCES

[1] G. Antipov, M. Baccouche, S.-A. Berrani, and J.-L. Dugelay, "Effective training of convolutional neural networks for face-based gender and age prediction," *Pattern Recognition*, vol. 72, pp. 15-26, 2017.

[2] Jie-sheng Wang, Yan-lang Ruan, Bo-wen Zheng, and Shu-zhi Gao, "Face Recognition Method Based on Improved Gabor Wavelet Transform Algorithm," *IAENG International Journal of Computer Science*, vol. 46, no.1, pp12-24, 2019.

[3] C. Zhang, S. Liu, X. Xu, and C. Zhu, "C3AE: Exploring the limits of compact model for age estimation." pp. 12587-12596.

[4] R. Rothe, R. Timofte, and L. Van Gool, "Dex: Deep expectation of apparent age from a single image." pp. 10-15.

[5] P. Li, Y. Hu, X. Wu, R. He, and Z. Sun, "Deep label refinement for age estimation," *Pattern Recognition*, vol. 100, pp. 107178, 2020.

[6] P. V. Er, and K. K. Tan, "Non-intrusive fall detection monitoring for the elderly based on fuzzy logic," *Measurement*, vol. 124, pp. 91-102, 2018.

[7] C. Xiao, Z. Zhifeng, C. Jie, and Z. Qian, "Combined deep learning with directed acyclic graph SVM for local adjustment of age estimation," *IEEE Access*, vol. 9, pp. 370-379, 2020.

[8] O. Guehairia, A. Ouamane, F. Dornaika, and A. Taleb-Ahmed, "Deep random forest for facial age estimation based on face images." pp. 305-309.

[9] M. M. Badr, A. M. Sarhan, and R. M. Elbasiony, "ICRL: Using landmark ratios with cascade model for an accurate age estimation system using deep neural networks," Journal of Intelligent & Fuzzy Systems, no. Preprint, pp. 1-14.

[10] B. Zhang, and Y. Bao, "Age Estimation of Faces in Videos Using Head Pose Estimation and Convolutional Neural Networks," *Sensors*, vol. 22, no. 11, pp. 4171, 2022.

[11] S. Pramanik, and H. A. B. Dahlan, "Face Age Estimation using Shortcut Identity Connection of Convolutional Neural Network," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 4, 2022.

[12] K.-Y. Chang, C.-S. Chen, and Y.-P. Hung, "Ordinal hyperplanes ranker with cost sensitivities for age estimation." pp. 585-592.

[13] W. Wang, T. Ishikawa, and H. Watanabe, "Facial age estimation by curriculum learning." pp. 138-139.

[14] W. Cao, V. Mirjalili, and S. Raschka, "Rank consistent ordinal regression for neural networks with application to age estimation," *Pattern Recognition Letters*, vol. 140, pp. 325-331, 2020.

[15] Z. Niu, M. Zhou, L. Wang, X. Gao, and G. Hua, "Ordinal regression with multiple output cnn for age estimation." pp. 4920-4928.

[16] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition." *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770-778, 2016.

[17] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks." pp. 4700-4708.

[18] Y. Lee, J.-w. Hwang, S. Lee, Y. Bae, and J. Park, "An energy and GPU-computation efficient backbone network for real-time object detection." pp. 0-0.

[19] O. Agbo-Ajala, and S. Viriri, "Deep learning approach for facial age classification: a survey of the state-of-the-art," *Artificial Intelligence Review*, vol. 54, no. 1, pp. 179-213, 2021.

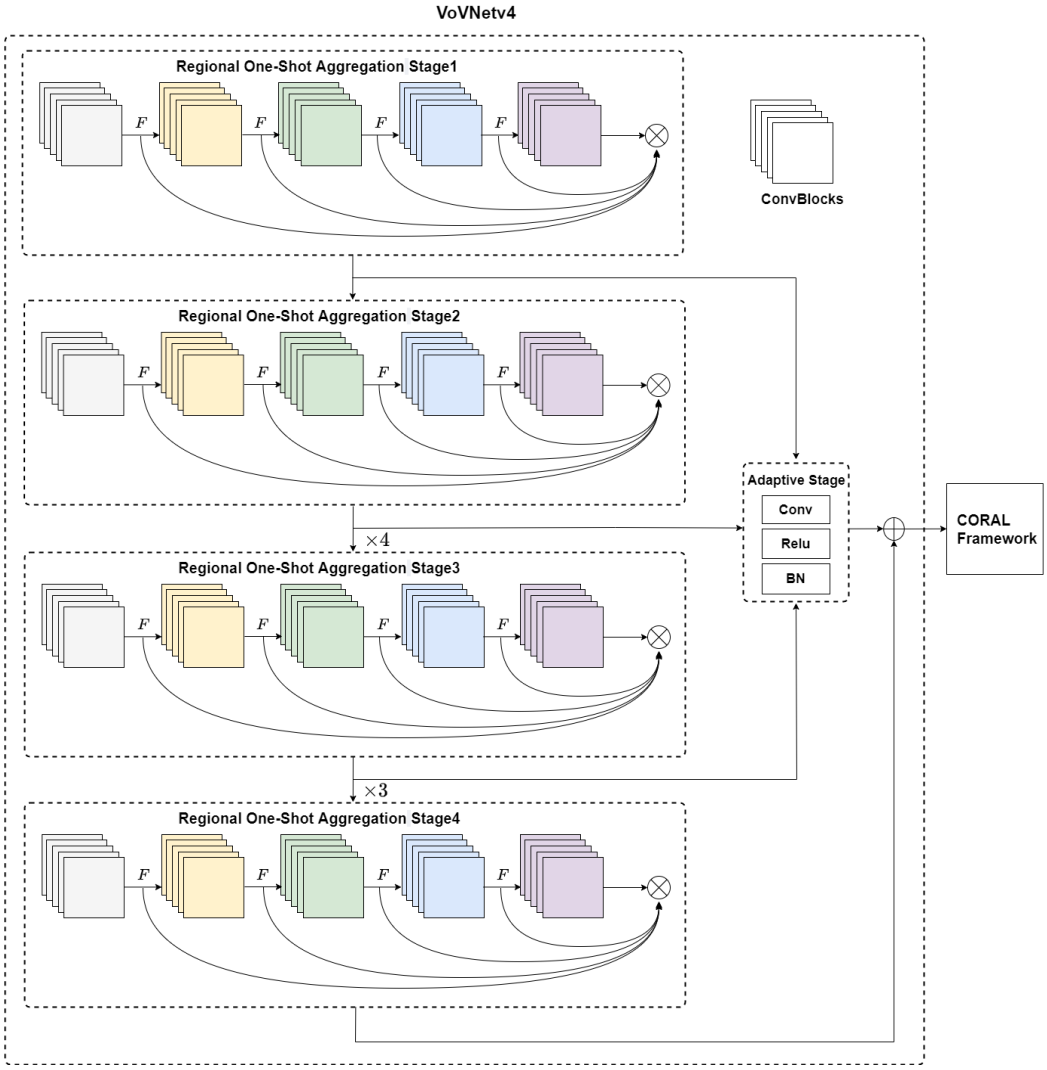

Fig.1: Overall network architecture diagram

HUAN ZHANG was born in Liaoning Province, P. R. China, received the B. Sc degree in Software Engineering from University of Science and Technology Liaoning, Anshan, P. R. China, in 2020.

She is currently pursuing the M. Sc degree in Software Engineering with University of Science and Technology Liaoning, Anshan, P. R. China. Her research interest is computer vision.

WENHUA CUI was born in Liaoning Province, P. R. China, received the M. Sc degree in Electromechanical control and automation from Dalian University of Technology, Dalian, P. R. China, in 1998, received the Ph.D. degree in Control Theory and Control Engineering from Dalian University of Technology, Dalian, P. R. China, in 2014,

She is currently a professor in the School of Computer and Software Engineering, University of Science and Technology Liaoning. She has published more than 20 academic papers and established more than 20 scientific research projects. The main research directions are control theory, sensor measurement and control, intelligent IoT, information security, computer network, machine vision, etc.

TIANWEI SHI was born in Liaoning Province, P. R. China, received the M. Sc degree in Control Theory and Control Engineering from University of Science and Technology Liaoning, Anshan, P. R. China, in 2010, received the Ph.D. degree in Mechatronic Engineering from Northeastern University, Shenyang, P. R. China, in 2016

He is currently an associate professor in the School of Computer and Software Engineering, University of Science and Technology Liaoning, Anshan, P. R. China. The main research directions are brain-computer interface physiological electrical signal analysis and processing, machine vision, etc.

YE TAO was born in Liaoning Province, P. R, received the B. Sc degree in Computer Science and Technology from Anshan Normal University, received the M. Sc degree in Computer Science and Technology from Northwest Normal University.

He is currently a lecturer in the School of Computer and Software Engineering, University of Science and Technology Liaoning, Anshan, P. R. The main research direction is image encryption

JIANFENG ZHANG was born in Heilongjiang Province, P. R. China, received the B. Sc degree in Automation from University of Science and Technology Liaoning, Anshan, P. R. China, in 2020.

He is currently pursuing the M. Sc degree in Control Science and Engineering with University of Science and Technology Liaoning, Anshan, P. R. China. His research interest is computer vision.

YULU LIU was born in Liaoning Province, P. R. China, received the B. Sc degree in Software Engineering from University of Science and Technology Liaoning, Anshan, P. R. China, in 2020.

She is currently pursuing the M. Sc degree in Software Engineering with University of Science and Technology Liaoning, Anshan, P. R. China. Her research interest is computer vision.