Electronic Research Archive

ERA, 31(8): 4907–4924.
DOI: 10.3934/era.2023251
Received: 17 April 2023
Revised: 16 June 2023
Accepted: 22 June 2023
Published: 11 July 2023

http://www.aimspress.com/journal/ERA

Research article

Age estimation algorithm based on deep learning and its application in fall detection

Jiayi Yu, Ye Tao*, Huan Zhang, Zhibiao Wang, Wenhua Cui and Tianwei Shi

School of Computer Science and Software Engineering, University of Science and Technology Liaoning, China

* **Correspondence:** Email: taibeijack@163.com; Tel: +8613304224928; Fax: +8604125929818.

Abstract: With the continuous development and progress of society, age estimation based on deep learning has gradually become a key link in human-computer interaction. Widely combined with other fields of application, this paper performs a gradient division of human fall behavior according to the age estimation of the human body, a complete priority detection of the key population, and a phased single aggregation backbone network VoVNetv4 was proposed for feature extraction. At the same time, the regional single aggregation module ROSA module was constructed to encapsulate the feature module regionally. The adaptive stage module was used for feature smoothing. Consistent predictions for each task were made using the CORAL framework as a classifier and tasks were divided in binary. At the same time, a gradient two-node fall detection framework combined with age estimation was designed. The detection was divided into a primary node and a secondary node. In the first-level node, the age estimation algorithm based on VoVNetv4 was used to classify the population of different age groups. A face tracking algorithm was constructed by combining the key point matrices of humans, and the body processed by OpenPose with the central coordinates of the human face. In the secondary node, human age gradient information was used to detect human falls based on the AT-MLP model. The experimental results show that compared with Resnet-34, the MAE value of the proposed method decreased by 0.41. Compared with curriculum learning and the CORAL-CNN method, MAE value decreased by 0.17 relative to the RMSE value. Compared with other methods, the method in this paper was significantly lower, with a biggest drop of 0.51.

Keywords: fall detection; age estimate; module processing

1. Introduction

1.1. Research background and significance

In computer science, artificial intelligence has been dominant and has gained increased attention, as it is a necessary means for the development of social intelligence. Artificial intelligence simulates human thought information through machine learning and performance. There have been major breakthroughs in artificial intelligence technology, and it is widely used in production learning. In every area of everyday life, it is of driving significance to the development of society [1]. Computer vision is a prominent area of machine learning. According to the information such as pictures and videos, the computer has the ability to recognize itself. Computers and people have similar methods for information acquisition. Computers work through cameras, massive data, and related algorithms [2]. With the continuous development of society, computer vision has been gradually applied to various industries. In picture classification, there has been excellent performance in detection and segmentation, though there is still plenty of room for expansion.

Because of the evolution of society and the ubiquity of videos, there was an explosion of data. Image-based visual detection and classification has been gradually replaced by videos, using video surveillance to obtain information. Focusing on the analysis based on a large number of historical information sets has gradually become an inevitable trend of social development [3]. Population aging has become a major concern of society. An increasing amount of elderly people live alone. Relative to other abnormal detection behaviors, monitoring falls in the elderly is getting more and more attention and is significant to promote the healthy development of human beings.

Along with the social structure of population aging, you can see exactly how the population is aging. At the same time, population aging is still increasing. Compared with the previous decade, there was a significant increase in the number of individuals over 60 years old [4]. Based on a survey of the causes of accidental injuries suffered by the elderly, injuries from falls account for a large proportion of deaths. Since the damage caused by a fall is irreversible, the fall detection of the elderly has become a major concern of the society [5]. Nowadays there are more and more elderly people living alone. It is particularly necessary to test the physical health of the elderly in the aspect of old-age care, including monitoring those who live in nursing homes. Monitoring human fall information according to video and visual information has become an inevitable trend of social development.

In the process of human fall detection, the fall is usually taken as the detection center and the judgment of the age of the person who fell is ignored. Falls in different age groups tend to have different consequences; therefore, the measures taken after a fall are also different. Focused monitoring of falls based on age is especially necessary. In this paper, human fall behavior is thoroughly studied. Additionally, a specific analysis of the problems faced are shown, such as a data explosion which occurs when too many videos are utilized as the input; this subsequently has a great influence on the real-time performance of fall detection. Moreover, fall detection for the elderly needs to be graded in the form of gradients. Therefore, on the basis of solving these problems, this paper constructs a fall prediction mechanism for people of different ages.

1.2. Research status at home and abroad

Human fall detection is a classification problem based on human posture information and age.

Information about the human face and biological characteristics of the human body needs to be gathered to make predictions about how old people are through the use of an age estimator. In recent years, researchers at home and abroad have performed numerous experiments on age estimation.

Age is the most important characteristic information about humans. In human-computer interaction, age estimation has always been difficult to detect through the use video surveillance and other important application values. Depending on how much facial information is maintained, detection information often contains some errors. Facial expressions and skull shape also play a role in the results. Many scholars and researchers have made a lot of improvements on the age-based estimation algorithm.

Li et al. [6] proposed a refined network Local Response Normalization (LRN) simultaneously utilizing packaging label distribution and a regression essence, and a relaxation regression refining field was used for age discrimination. Yu et al. [7] solved the problem of insufficient data volume and differences in data distribution. Based on a fine-tuned network, they composed a classification network of two types of Convolutional Neural Networks (CNN) for age estimation. Li et al. [8] provided some novel ideas and prospects for the application of CNN. This work provided an overview of various convolutions and outlined some rules of thumb for function and hyperparameter selection. Guehairia et al. [9] used the Gcforest algorithm to perform age estimation experiments based on images. The algorithm had the advantage of a cascading structure that allowed for interactions between trees. Badr et al. [10] proposed a cascaded model system. The division of age labels was understood through a classification model, and the knowledge learned from the classification model was used as the auxiliary input for a regression model to achieve age estimation. Zhang and Bao [11] used a regression forest to estimate the age of face images alongside head posture. Pramanik and Dahlan [12] combined a convolutional neural network framework and Resnet50 architecture to propose a fast method for age estimation from face images. Chang et al. [13] proposed the OHRank's ordinal hyperplane sorting algorithm based on the relative sequence information of age labels in the database; subsequently, the age of the human body is determined. Many researchers conducted experiments on the Asian face age dataset (AFAD). For example, Wang et al. [14] combined curriculum learning with age estimation, which was used to improve the training efficiency of the neural network and enhance the ability of the network to discriminate age. Santos et al. [15] proposed a fall detection system based on commodity mmWave sensors along with body-feature estimation algorithm to overcome the low-resolution deficiency. They compared several body-feature estimation algorithms and selected the best one. Then, they illustrated the potential of body-feature estimation algorithms via a case study of a threshold-based fall detection system. Niu et al. [16] made a breakthrough in the traditional classification and regression research in age estimation, in which sequential regression was adopted to carry out feature learning and regression modeling at the same time. Based on deep learning, the end-to-end learning of convolutional neural networks was used to concretely analyze the regression problem. The age estimation effect has been effectively improved. Age estimation is a key part of human-computer interaction and can be widely combined with other fields of application. It has a profound impact on the development of society.

2. Age estimation algorithm based on VoVNetv4

2.1. Algorithm design

2.1.1. VoVNetv4 overall network analysis

Because the bottom feature has unique details from the top feature, they can provide more detailed richness; however, its semantics are poor. With strong noise, high level features are usually less keen to capture fine nodes. Therefore, they usually have a lower resolution than the underlying features, though they have much more robust semantic information. The goal of feature aggregation is to efficiently integrate low-level features with high-level features. It is an important means to improve the model performance. VoVNetv4 adopts the mode of module splicing to deliver early features, thereby the original characteristic pattern is preserved by aggregating multiple receptive fields to capture all kinds of visual information in a trans-latitude manner and to extract features to achieve a diversified representation. The VoVNetv4 network is composed of four ROSA and adaptive stage modules. All features are connected once in each ROSA module. First, it performs a single aggregation at the module output. Then, the ROSA output information is respectively processed by the adaptive stage module. Finally, a sexual polymerization is performed. The VoVNetv4 network can effectively avoid feature redundancy and improve the detection accuracy of the network. The overall architecture of its network is shown in Figure 1.

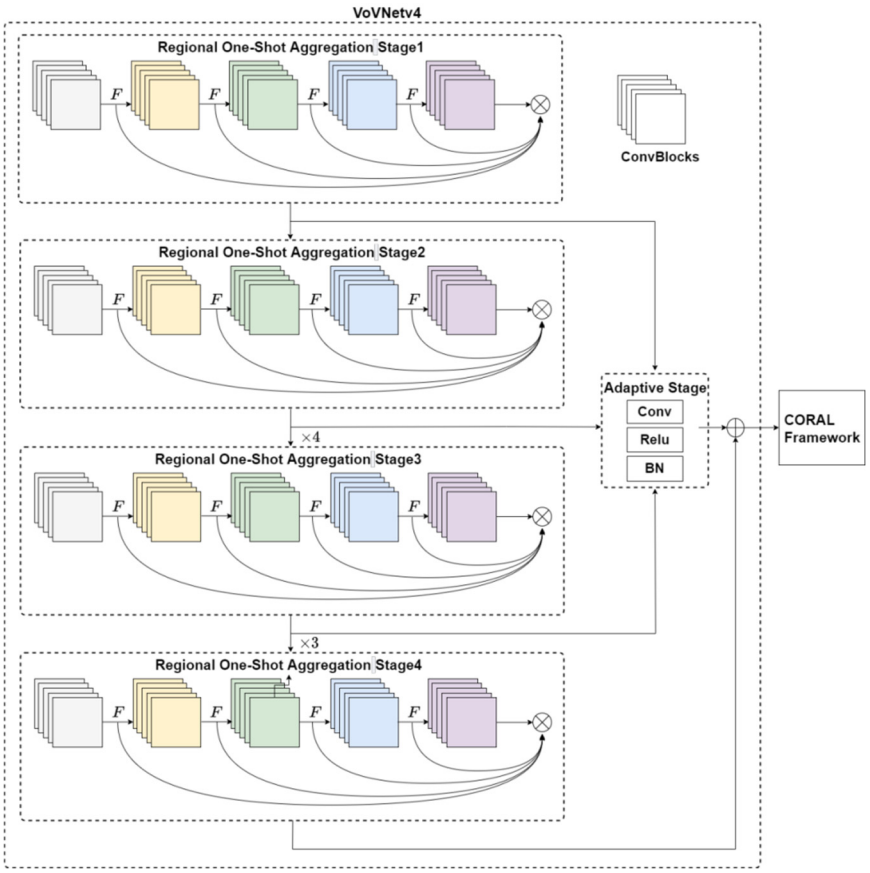

Figure 1. Overall architecture diagram.

2.1.2. Rosa module

In object detection, most networks use Resnet [17] and Densnet [18] as Backbone networks. Densnet is based on the Resnet network, which densely joins the convolution layers to ensure the flow of information between volumes and layers because dense connections not only bring feature enhancement, but also brings the disadvantage of linear growth of the output channel. The VoVNet [19] network has made a lightweight improvement from the dense connections to all feature aggregations in the last layer to address this shortcoming. Although the VoVNet network effectively solves the problems of complexity and memory access cost of original network, the recognition accuracy has not been significantly improved. The VoVNetv4 proposed in this paper draws on the improvement idea of VoVNet. The ROSA module is proposed for a single aggregation with phased characteristics. The aggregation calculation of VoVNet is shown in Figure 2.

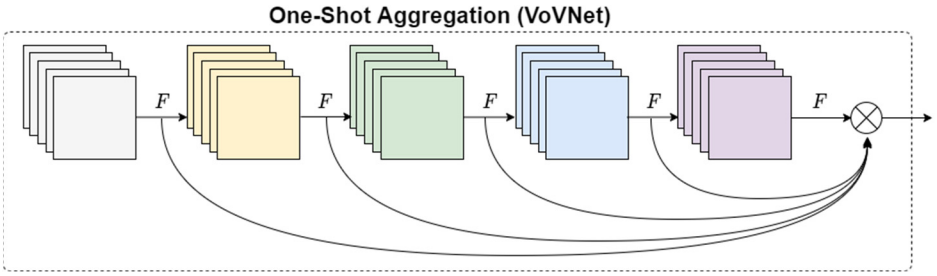

Figure 2. Aggregate computation of VoVNet.

The ROSA module is designed with a phased single aggregation module architecture, and is used for aggregation operations. In the ROSA module, the region is divided into different stages of operation, where each region represents a stage of operation. The ROSA module consists of four areas. Each area contains 5 identical ConvBlocks and has the same input and output channels. ConvBlocks consists of a 3 root 3 convolution layer, a BN layer and a Relu layer. The convolution kernel has a step size of 1. In the process of the aggregation operation, region 1 and region 2 are looped only once, region 3 is looped four times, and region 4 is looped three times. Every time you enter the next Stage of the ROSA stage, the feature graph goes through a convolution kernel of 1, downsampling with step size 1 and a 3 root 3 convolution layer. The maximum pooling layer has a step size of 2 because high-level semantic information is more important than low-level semantic information in target detection tasks. The ROSA module adds high-level features through different regional operations and is used to improve the ratio of high-level features to low-level features.

2.1.3. CORAL framework classifier

The age estimation scheme proposed in this paper includes image preprocessing, three stages of feature extraction and age classification. The CORAL framework is an ordered classifier for the age classification stage. In terms of sorting task, it is important to reduce the differences between the same classes, as there is more variation between categories of different species. Whether good classification results can be achieved is a critical task. The CORAL framework converts ordinal targets into binary classification subtasks. This framework can effectively improve the classification effect. The

derivation process of the CORAL framework is shown in Eq (1):

$$D = \left\{ \mathrm{x}_{i,} \mathrm{y} \right\}_{\mathrm{i}=1}^{N} \tag{1}$$

where D is the data set, x_i is the i-th picture, N is the number of samples, and y_i represents the corresponding rank value. The rank is the age rank of a human body within a certain region, where y_i is the set that rank belongs to. Let's call this set Y. Then, the relation between y_i and Y is shown in Eq (2):

$$y_{\mathrm{i}} \in Y = \left\{ r_1, r_2, \ldots r_{\mathrm{k}} \right\}. \tag{2}$$

The elements contained in set Y are arranged in order, as shown in Eq (3):

$$r_{\mathrm{k}} \succ r_{\mathrm{k\text{-}1}} \succ \ldots \succ r_1. \tag{3}$$

The main purpose of the ordered regression task is to find a sorting rule, that is, the corresponding relationship between the age picture and the age value in rank minimizes its loss function. Let *C* be a cost matrix of $K \times K$. C_{y,r_k} represents the loss value of rank (r_k) when a sample (x, y) is predicted. When $C_{y,r_k} = 0$, the picture that represents the age corresponds exactly to the age in the rank. The network model presents a perfect prediction state when$y \neq r_k$, $C_{y,r_k} > 0$. In ordinal regression, the V-shaped cost matrix is more conducive to feature learning and classification. In the actual computation, it is hard to obtain the cost matrix to a V state. The CORAL framework can produce consistent predictions for each binary task and avoids the disadvantages of a V matrix. One can extend y_i with binary labels, as shown in Eq (4):

$$\mathrm{y}_i^{(1)} \succ \ldots \succ \mathrm{y}_i^{(K-1)} \tag{4}$$

where $y_i^k \in \{0,1\}$ indicates whether $\mathrm{y_i}$ exceeds rank (r_k). For example: $y_i^k \in 1\{y_i > r_k\}$, if the internal condition is true, it indicates that the function 1 is 1; otherwise, it is 0. Given a response mechanism based on binary tasks, the predictive rank label of input x_i is obtained by $h(x_i) = r_q$. The Rank index q is obtained by $q = 1 + \sum_{k=1}^{K} f_k(x_i)$ where$f_k(x_i) \in \{0,1\}$ is the prediction of the KTH binary classifier in the output layer. $\{f_k\}_{k=1}^{K-1}$ reflects the ordinal information and is rank monotone, that is: $f_1(x_i) \geq f_2(x_i) \geq \cdots \geq f_{K-1}(x_i)$. This guarantees a consistent prediction.

2.1.4. Loss function

The weight parameter of the neural network is defined as *W*. Currently, the network structure does not include the bias unit of the last layer. The penultimate layer output of the network is expressed as $g(x^i, W)$ and shares a weight with all nodes in the final output layer. Then, one can add *K-1* independent bias units to $g(x^i, W)$. Put the corresponding binary in the last layer. The input to the classifier is defined as $\left\{ g(x^i, W) + b_k \right\}_{k=1}^{K-1}$. The activation function is shown in Eq (5):

$$\sigma（z）=1/(1+\exp(-z)) . \tag{5}$$

Here, the empirical prediction probability of task k is defined in Eq (6):

$$\hat{P}（y_i^{(k)}=1）=\sigma(g(x_i,W)+b_k) \tag{6}$$

where sigma is the activation function. For model training, one can use the loss minimization function *L(W, b), L(W, b)*, as shown in Eq (7):

$$L（W，b）=-\sum_{i=1}^{N}\sum_{k=1}^{K-1}\lambda^{(k)}\begin{bmatrix}\log(\sigma(g(x_i,W)+b_k))y_i^{(k)}\\+\log(1-\sigma(g(x_i,W)+b_k))(1-y_i^{(k)})\end{bmatrix} \tag{7}$$

where $\lambda^{(k)}$ is the loss weight associated with a k classifier when you assume that the value is greater than zero. The model is trained by extending binary tags. One can assign the $K-1$ binary classifier to the output layer to train individual convolutional neural networks.

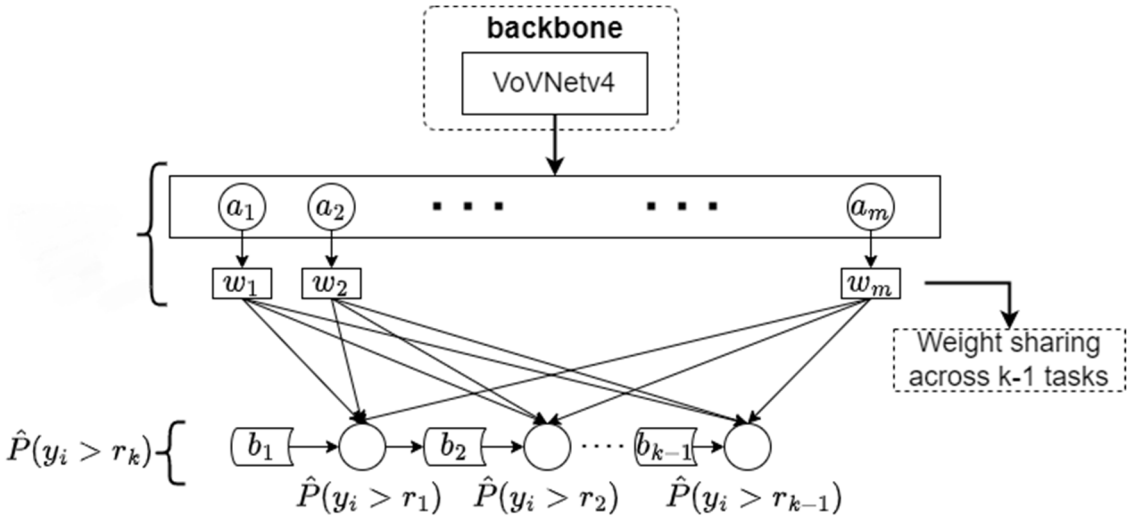

Figure 3. Calculation process based on CORAL age classification.

2.2. Experimental results and analysis

2.2.1. Experimental environment and parameters

This section evaluates the proposed age estimation methods according to different experimental procedures to prove the validity of experimental data. Meanwhile, it is compared with other methods. This experiment was conducted on a server with an Intel(R) Xeon(R) CPU E5-2630 v4 @ 2.20GHz 2.20 GHz (2 processors) and a NVIDIA TITAN X(Pascal). The operating system is Window10 64-bit. The video memory is 12 GB. The software utilized included Python (version 3.8) and Pytorch (version 1.9).

2.2.2. Experimental data set

The Asian face age dataset (AFAD) [20] is the largest age estimation dataset so far, containing 164,432 images of human faces. Each image contains a corresponding human age and gender label. The AFAD data set was collected on Renren. Its image information contained 100,752 images of men and 63,680 pictures of women. The data set focuses on Asian age estimates and contains different backgrounds and lighting conditions.

2.2.3. Evaluation index

In this paper, to test the effectiveness of the proposed backbone network, the performance of VoVNetv4 was evaluated on two evaluation indexes: Mean Absolute Error (MAE) and Root Mean Square Error ($RMAE$). MAE is mainly used to calculate the average difference between a person's predicted age and their actual age, that is, the average solution of residual error. The smaller the *MAE* value, the higher the accuracy of age estimation, as shown in Eq (8):

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|y_i - h(x_i)\right| \tag{8}$$

.

$RMAE$ is used to measure the difference between the predicted age and the actual age. The calculation process is similar to that of the standard difference, as shown in Eq (9):

$$RMAE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_i - h(x_i)\right)^2} \tag{9}$$

where y_i is the age 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.

2.2.4. Analysis of experimental results

i. Experimental results

In order to make a fair comparison, using the same random seeds in the training, the experiment was divided into four parts: the first part is to take the standard Resnet-34 classification network as the performance baseline model; the second part is to integrate the standard Resnet-34 classification network and the CORAL framework and to verify the effectiveness of the CORAL framework; the third part is a fusion using VoVNet and the CORAL framework to verify the VoVNet performance; and the fourth part is to improve VoVNet. Therefore, VoVNetv3 is fused with the CORAL framework. To verify the validity of the proposed method, the random seed is set to 1, the learning rate is set to 0.0005, the epoch is 200, batchsize is set to 128.

1) To verify the effectiveness of the CORAL framework, Resnet-34 was selected as the feature extraction network for verification analysis. Primarily, the Resnet-34 network is used as a classification network for age estimation. The convergence of the network is analyzed by examining any changing trend of loss value. The network training loss figure is shown in Figure 4. The network gradually

converges at approximately round 37.

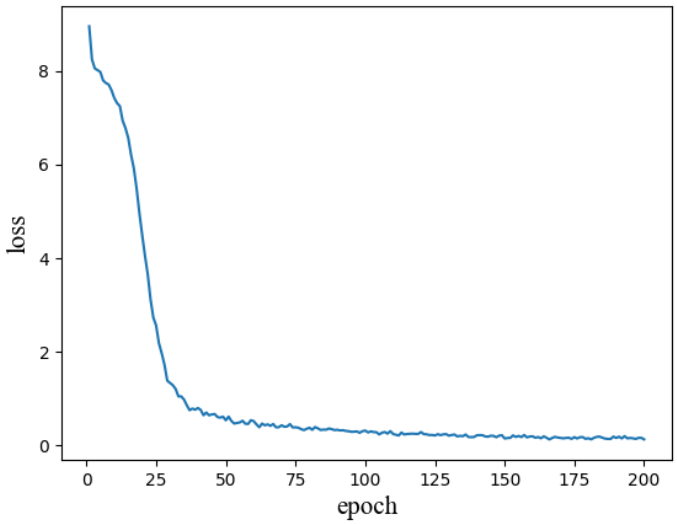

Figure 4. VoVNetv4 +CORAL framework cost function curve.

2) Based on the first part of the experiment, the fusion of the CORAL framework classifiers is used to verify the effectiveness of the CORAL framework. In this experiment, the cost function is used for network optimization analysis. The network convergence is shown in Figure 5. The model gradually converges at approximately round 25. Experiments show that the CORAL framework classifier can effectively improve the generalization ability of the network.

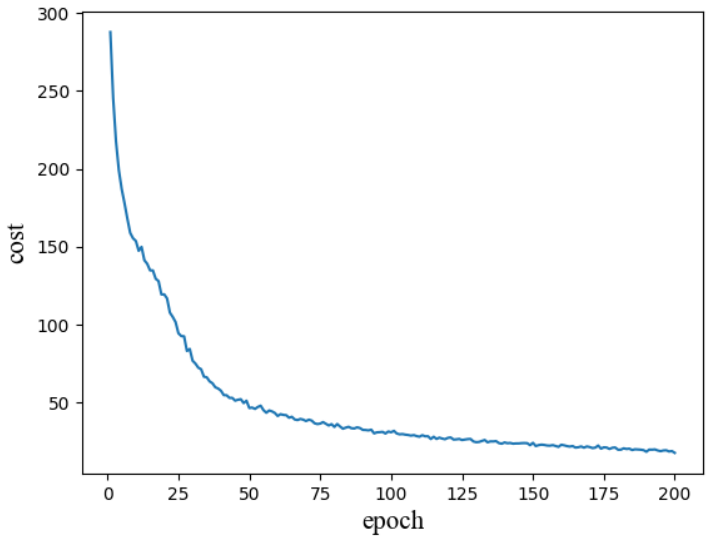

Figure 5. Resnet-34+CORAL framework cost function curve.

3) In this experiment, VoVNet is used as a feature extraction network and the CORAL framework is used as a classifier. The network convergence is shown in Figure 6. The experimental

results show that the convergence of the network is significantly improved in the 25th round. Compared with Resnet-34 as a feature extraction network, its convergence speed and effect have significantly improved.

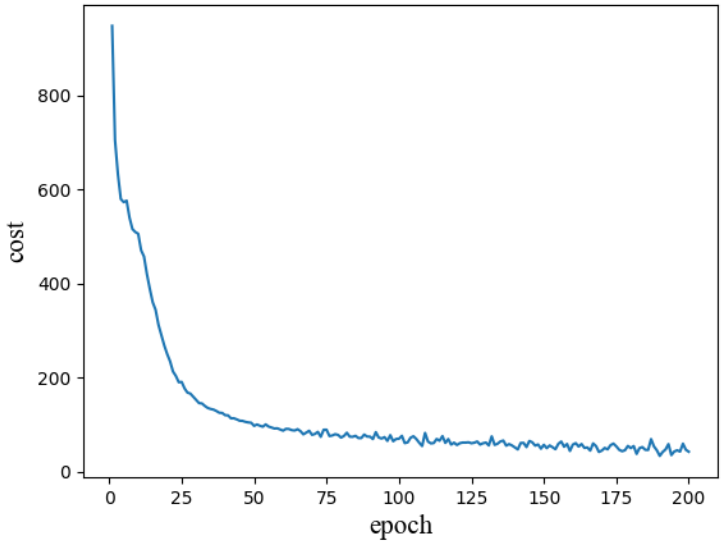

Figure 6. VoVNet +CORAL framework cost function curve.

4) From the third part of the experiment, the effectiveness of using VoVNet as a feature extraction network was verified. In the fourth part of the experiment, the VoVNet network was improved. The backbone network VoVNetv4 was proposed. In order to verify the performance of the proposed network. In this experiment, VoVNetv4 is used as the feature extraction network. The classification network uses the CORAL framework. Its training cost function curve is shown in Figure 7. The experimental results show that the VoVNetv4 network has a higher convergence ability and speed.

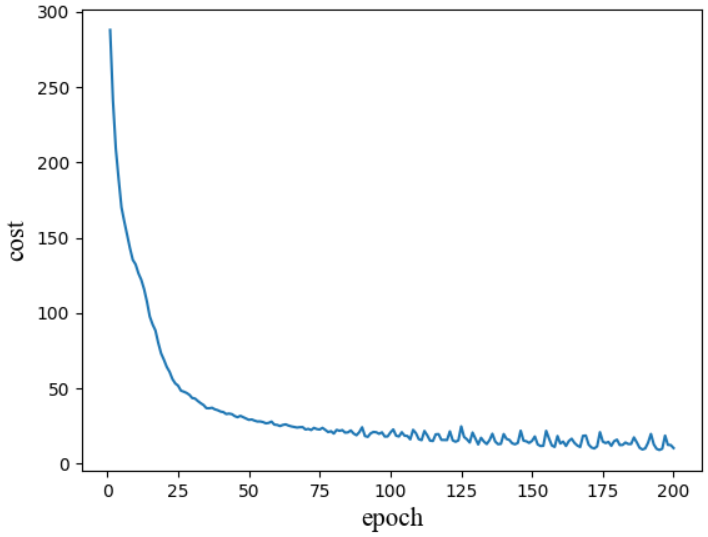

Figure 7. VoVNetv4 +CORAL framework cost function curve.

ii. Result analysis

Other research methods are replicated in this chapter, that is, under the same AFAD data set. Training rounds of 200, MAE and RMSE were used to analyze the performance. The experimental results show that compared with curriculum learning and the CORAL-CNN method, the MAE value decreased by 0.17. Compared with the OR-CNN method, relative to the value of RMSE, the MAE value decreased by 0.21. The methods used in this paper have different degrees of decline compared with the comparison methods. The comparative experimental results are shown in Table 1.

Table 1. Comparison with other methods is based on AFAD data set.

Approaches	MAE	RMSE
With Curriculum Learning	3. 47	5. 03
OR-CNN	3. 51	4. 75
CORAL-CNN	3. 47	4. 71
Ours: VoVNetv4+CORAL	3. 30	4. 64

iii. Ablation experiment

The results of the ablation experiment are shown in the Table 2. The VoVNetv4 in the table is a phased single aggregation backbone network VoVNetv4 that uses the ROSA module proposed in this paper for regional encapsulation. The CORAL framework is used as the classifier for the CORAL substitute experiment. The evaluation index with a downward arrow indicates that the lower the value of the evaluation index, the better the performance of the algorithm to achieve the effect. The number below the evaluation index is the value of the decline of the optimized laughter method used as compared with the original algorithm.

Table 2. Experimental results on the AFAD dataset.

VoVNetv4	CORAL	MAE↓	RMSE↓
	√	0.27	0.46
√		0.35	0.41
√	√	0.41	0.51

As shown in the Table, MAE and RMAE both decline when using either the proposed VoVNetv4 as the backbone network or the CORAL framework as classifiers alone, and the combination of the two produces an improved effect. Therefore, we finally choose VoVNetv4 as the feature extraction network and the CORAL framework as the classifier.

3. Application of age detection algorithm in fall detection

A gradient-style two-node fall detection framework combined with age estimation is designed. In order to achieve the detection priority of human falls in different age groups, the framework adopts a dual-node method to divide the detection into a primary node and a secondary node. In the first-level node, the age estimation algorithm based on VoVNetv4 is used to classify people of different age groups. At the same time, a face tracking algorithm is constructed based on the combination of the key point matrix of the human body processed by OpenPose and the center coordinates of the face. In the second-level node, human body fall detection based on the AT-MLP model is performed based on the age gradient information of the human body in order to improve the real-time performance in the fall detection process [21].

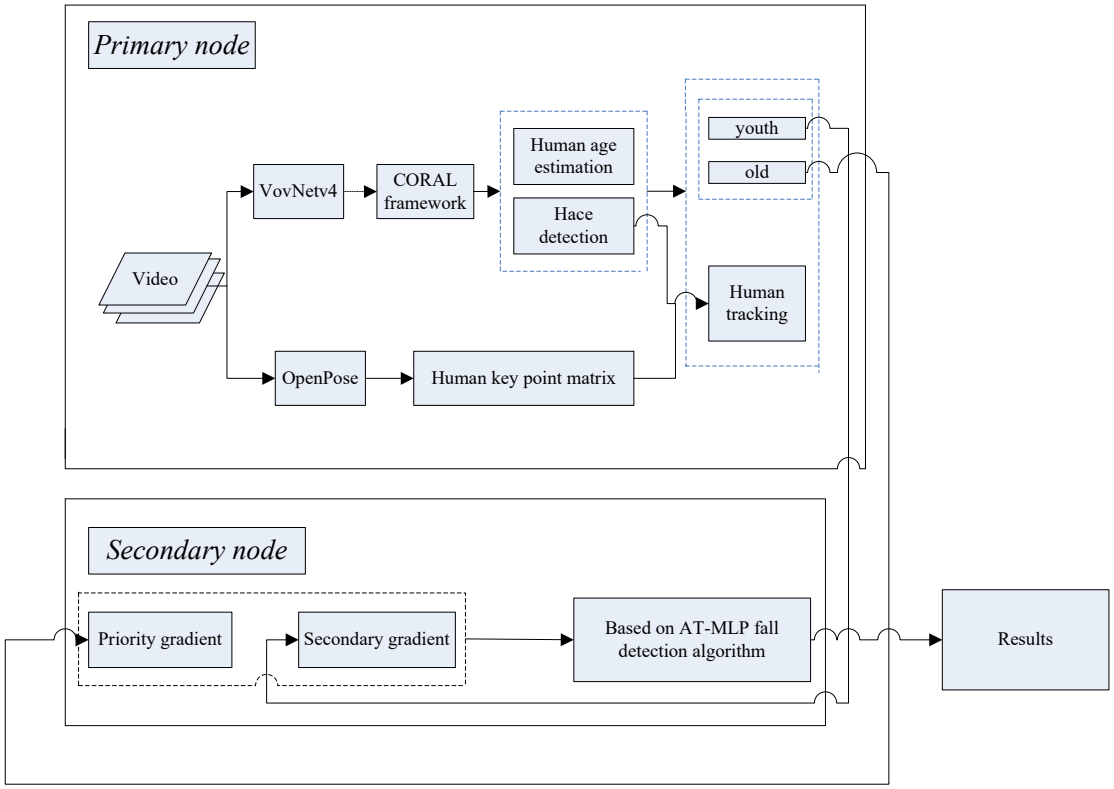

Figure 8. First level node frame diagram.

3.1. Primary node

The framework adopts a two-node task approach for task analysis. The primary node is divided into two stages: information processing and priority allocation. The information processing stage seeks to understand relevant video information. the first, the video is processed by the VoVNetv4 age estimation algorithm to obtain the age information of the human body and the central position of the face. The age information is classified into stages by the age group. If the age information is 60 years or older, it is classified as elderly. Those under 60 are divided into young adults and teenagers. It is used for task deployment in the next phase. At the same time, for the video, it was passed into the

OpenPose network to carry out the key point coordinates of human body, that is, the output key point matrix information of human body. First, it combines the central position of face obtained in the age estimation algorithm with the key point matrix of human body output by OpenPose. Next, it constructs the face center point tracking algorithm to locate the age of the body and the key points of the body; thus, the tracking process is realized. Then, it sets the distance between the central point coordinates of face and the central point coordinates of human body recognized by OpenPose as D. The calculation of D is shown in Eq (10):

$$D = \sqrt{(x_{\mathrm{o}} - x_{\mathrm{v}})^2 + (y_{\mathrm{o}} - y_{\mathrm{v}})^2} \quad . \tag{10}$$

Among them, $(x_o,\ y_v)$ is the key point coordinates of human face center recognized by OpenPose. $(y_o,\ y_v)$ are the key point coordinates of human face center recognized by the age estimation algorithm. According to D, the corresponding relationship between the target and the key points is determined. The priority allocation stage mainly prioritizes the tracked human body information: information older than or equal to 60 years of age is classified as an advanced resource; the rest is passed to the next node as low-level resources.

3.2. Secondary node

The secondary node includes two parts: resource allocation and human fall detection. It is used to further process the information of the first node. The gradient two-node fall detection framework combined with age estimation allocates human information of different age groups in two directions of priority gradient and secondary gradient. Let's say one has n videos to process. Then, n videos are passed into the level 1 node for processing. After tracking the center of the face, the corresponding relationship between the age of human body and the key point matrix of human skeleton is obtained. Equation (11) shows the storage mode of the resource allocation phase:

$$\mathrm{m} = \begin{cases} S_h(v_j, age_j); & when \quad age \geq 60 \\ S_l(v_k, age_k); & others \end{cases} \tag{11}$$

where m is the information passed into the gradient sequence and is a high-level resource. The age of the human body is greater than or equal to 60. Key point matrix information of human skeleton, it is a low-level resource, that is, the human age is less than human skeleton key point matrix information. k and j are the video sequences in the incoming gradient to set the threshold information in the incoming gradient stack respectively. Suppose that the threshold of key bone information of the elderly is set to 5. For others, the threshold for key bone information is set to 1. In the resource allocation phase, five falls of the elderly will be performed. One can perform another fall of a young person or juvenile and enter the loop detection process in this way.

3.3. Visualization of detection effect

In order to reflect the detection results and practical application effects of the two-node fall detection framework combined with age estimation, this section demonstrates on the LFDD data set. The LFDD dataset contains multiple fall scenarios. There are different light and color differences and

associated occlusion conditions. It feeds the human fall video into the two-node fall detection framework through the information processing and priority allocation of the first node. Then, it obtains the age label of the human body. Then, according to the age information gradient into the second level node, they mark the human body for falls. Now, visualization results of part of the model detection are presented, as shown in Figure 9. These results include the elderly, young people, and young people with special conditions.

In Figure 9(a), a group of images showing fall detection of human subjects aged 65 is presented. On the left is the visual information that has not fallen, including information about the person's age and whether they fell and marked the priority distribution channels. The figure on the left and the figure on the right are of high priority. The picture on the right shows the status information of the elderly when they fall. It includes the special case of occlusion. In Figure 9(b), a group of images showing visual information about falls among young people is presented. All of them are assigned with low priority. In Figure 9(c), a group of images shows no information about the human face. The mark at this time does not show the age label because of the realistic facial occlusion factor. It treats such cases as high priority assignments. The image on the right shows a human squat that looks like a fall. At this time, the model can correctly identify whether the human body is in a falling state.

Since there are no multiplayer scenarios in the LFDD data set, to demonstrate the applicability of the two-node fall detection model combined with age estimation,other multi-person scenarios are experimentally verified. The visualization results are shown in Figure 10.

(a) Fall detection in human aged

(b) Visual information on falls among young people

Continued on next page

(c) Fall detection without human face information

Figure 9. Visualization of detection effect in LFDD data set.

(a) Fall scenes at different ages

(b) High priority age fall scenario

Figure 10. Visualization of multi-scene detection effect. (a) Group of images shows the fall scenes at different ages. (b) Group of images shows falls in the high priority age group. And includes confusion in which the human body performs a squat. The visualization results can be analyzed. The model in this paper can be adapted to multi-person scenarios. And can accurately identify the fall according to the priority.

4. Conclusions

In this paper, the proposed human age estimation algorithm is introduced in detail. First, the backbone network VoVNetv4 is proposed for overall analysis. It consists of the ROSA module and the adaptive stage module. The second is a concrete analysis of the two modules. The ROSA module is encapsulated by the region of the feature module. In order to achieve a phased single aggregation of features, the mode of aggregation calculation and the architecture of ROSA module are explained. The adaptive stage consists of Convolutional layer, BN layer, and Rectified Linear Unit activation function. It is used to smooth the feature information of the ROSA module. The three network layers covered are described in detail in this section. In this chapter, the classifier and loss function used in age estimation are introduced and explained in detail. The principle and calculation process of CORAL framework classifier are shown. The use of Loss function is explained and analyzed. Finally, the age estimation algorithm based on VoVNetv4 is analyzed experimentally. The data set of the experiment is introduced. The experiment was divided into three parts. The effectiveness of the proposed method is verified by a series of comparative experiments. This paper proposes VoVNetv4, a new backbone network for object detection. VoVNetv4 uses regional splicing to transfer early features, completes the capture of various visual information across latitudes, and realizes a single feature aggregation in stages; at the same time, it builds an adaptive stage module composed of Conv layer, BN layer, and Relu activation function. Feature smoothing; the classifier uses the CORAL framework to divide tasks in binary form and make consistent predictions for each task. In this paper, the experimental verification is carried out on the AFAD data set of the Asian face age data set. The effectiveness of the proposed method is verified through four sets of ablation experiments, and compared with other detection methods. The experimental results show that the average absolute error is 3.30, and the root mean square error is 4.64, and the detection effect is very obvious. Realize the hierarchical detection of the elderly and young people, effectively improve the real-time monitoring ability, avoid the phenomenon of data explosion caused by too many input videos, and realize the gradient detection of human falls.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This work was supported by Joint Fund Project of the National Natural Science Foundation of China (U1908218), the Natural Science Foundation project of Liaoning Province (2021-KF-12-06), and the Department of Education of Liaoning Province (LJKFZ20220197).

Conflict of interest

The authors declare there is no conflict of interest.

References

1. A. F. Bekhit, Introduction to computer vision, in *Computer Vision and Augmented Reality in iOS*, **1** (2022), 1−20. https://doi.org/10.1007/978-1-4842-7462-0_1
2. J. Han, L. Shao, D. Xu, J. Shotton, Enhanced computer vision with microsoft kinect sensor: a review, *IEEE Trans. Cybern.*, **43** (2013), 1318−1334. https://doi.org/10.1109/TCYB.2013.2265378
3. Y. O. Sharrab, I. Alsmadi, N. J. Sarhan, Towards the availability of video communication in artificial intelligence-based computer vision systems utilizing a multi-objective function, *Cluster Comput.*, **25** (2022), 231−247. https://doi.org/10.1007/s10586-021-03391-4
4. N. Haering, P. L. Venetianer, A. Lipton, The evolution of video surveillance:an overview, *Mach. Vision Appl.*, **19** (2008), 279−290. https://doi.org/10.1007/s00138-008-0152-0
5. Y. Zhang, H. Liu, Constraints and countermeasures of the new situation of population on the future development of higher vocational education−based on the analysis of the seventh national population survey (in Chinese), *Educ. Vocation*, **6** (2022), 12−20. https://doi.org/10.13615/j.cnki.1004-3985.2022.06.016
6. P. Li, Y. Hu, X. Wu, R. He, Z. Sun, Deep label refinement forage estimation, *Pattern Recognit.*, **100** (2020), 107178. https://doi.org/10.1016/j.patcog.2019.107178
7. Y. Yu, K. Tang, Y. Liu, A fine-tuning based approach for daily activity recognition between smart homes, *Appl. Sci.*, **13** (2023). https://doi.org/10.3390/app13095706
8. Z. Li, F. Liu, W. Yang, S. Peng, J. Zhou, A survey of convolutional neural networks: analysis, applications, and prospects, *IEEE Trans. Neural Networks Learn. Syst.*, **33** (2022), 6999−7019. https://doi.org/10.1109/TNNLS.2021.3084827
9. O. Guehairia, A. Ouamane, F. Dornaika, A. Taleb-Ahmed, Deep random forest for facial age estimation based on face images, in *2020 1st International Conference on Communications, Control Systems and Signal Processing (CCSSP)*, IEEE, (2020), 305−309. https://doi.org/10.1109/CCSSP49278.2020.9151621
10. M. M. Badr, A. M. Sarhan, R. M. Elbasiony, ICRL: using landmark ratios with cascade model for an accurate age estimation system using deep neural networks, *J. Intell. Fuzzy Syst.*, **43** (2022), 72−79. https://doi.org/10.3233/JIFS-211267
11. B. Zhang, Y. Bao, Age estimation of faces in videos using head pose estimation and convolutional neural networks, *Sensors*, **22** (2022), 4171. https://doi.org/10.3390/s22114171
12. S. Pramanik, H. A. B. Dahlan, Face age estimation using shortcut identity connection of convolutional neural network, *Int. J. Adv. Comput. Sci. Appl.*, **13** (2022), 515−521. https://doi.org/10.14569/IJACSA.2022.0130459
13. K. Y. Chang, C. S. Chen, Y. P. Hung, Ordinal hyperplanes ranker with cost sensitivities for age estimation, in *CVPR 2011*, IEEE, (2011), 585−592. https://doi.org/10.1109/CVPR.2011.5995437
14. W. Wang, T. Ishikawa, H. Watanabe, Facial age estimation by curriculum learning, in *2020 IEEE 9th Global Conference on Consumer Electronics (GCCE)*, IEEE, (2020), 138−139. https://doi.org/10.1109/GCCE50665.2020.9291929
15. G. L. Santos, P. T. Endo, K. H. de Carvalho Monteiro, E. da Silva Rocha, I. Silva, T. Lynn, Accelerometer-based human fall detection using convolutional neural networks, *Sensors*, **19** (2019), 1644. https://doi.org/10.3390/s19071644

16. Z. Niu, M. Zhou, L. Wang, X. Gao, G. Hua, Ordinal regression with multiple output cnn for age estimation, in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2016), 4920−4928. https://doi.org/10.1109/CVPR.2016.532
17. A. Schmeling, G. Geserick, W. Reisinger, A. Olze, Age estimation, *Forensic Sci. Int.*, **165** (2007), 178−181. https://doi.org/10.1016/j.forsciint.2006.05.016
18. K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2016), 770−778. https://doi.org/10.1109/CVPR.2016.90
19. G. Huang, Z. Liu, L. van der Maaten, K. Q. Weinberger, Densely connected convolutional networks, in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2017), 4700−4708. https://doi.org/10.1109/CVPR.2017.243
20. O. Agbo-Ajala, S. Viriri, Deep learning approach for facial age classification:a survey of the state-of-the-art, *Artif. Intell. Rev.*, **54** (2021), 179−213. https://doi.org/10.1007/s10462-020-09855-0
21. Y. Ma, Y. Tao, Y. Gong, W. Cui, B. Wang, Driver identification and fatigue detection algorithm based on deep learning, *Math. Biosci. Eng.*, **20** (2023), 8162−8189. https://doi.org/10.3934/mbe.2023355

AIMS Press