Cross-age Face Image Similarity Measurement Based on Deep Learning Algorithms

Jing Zhang*, Ningyu Hu
Computer Department, Xinzhou Normal University, Xinzhou, 034000, China

***Abstract*—In this study, a multi-feature fusion and decoupling solution based on the RNN is proposed from a discriminative perspective. This method can address the identity and age information extraction losses in cross-age face recognition. This method not only constrains the correlation between identity and age using correlation loss but also optimizes identity feature restoration using feature decoupling. The model was trained and simulated in CACD and CACD-VS datasets. The single-task learning model stabilized after 125 iterations of training, while the multi-task learning model reached a stable and convergent state after 75 iterations. In terms of performance analysis, the DE-RNN model had the highest recognition accuracy with a mAP of 92.4%. The Human Voting model had a value of 90.2%. The mAP of the Human Average model was 81.8%, whereas the mAP of the DAL model was the lowest at 78.1%. Experiments proved that the model constructed in this study has effective recognition and application value in the cross-age face recognition scenario.**

Keywords*—Cross-age; image recognition; RNN; feature fusion; decoupling; loss function

I. Introduction

Nowadays, the universal facial recognition technology, which refers to facial recognition performed at small age intervals, has reached or even surpassed human performance. However, in specific scenarios such as cross age facial recognition across larger age intervals, the performance of universal facial recognition technology is limited because as individuals age, the facial features, including textures, bones, etc., gradually change. The facial feature changes with age seriously affect the performance of face recognition, resulting in unrecognizability or recognition errors. How to suppress age factors and extract age invariant facial features are the keys to cross age face recognition. The studies of cross age facial recognition algorithms not only compensate for the performance shortcomings of general facial recognition algorithms in large age interval facial recognition, but also has practical significance in criminal investigation, document and other scenarios. Scholar Ali et al. [1] emphasized the main applications, challenges, and trends of facial recognition systems in social and scientific fields in their research. At the same time, this article summarizes recent facial recognition technologies and introduces the future aspects of facial recognition technology and its potential significance in the upcoming digital society. Jin et al. [2] conducted facial image recognition analysis from RGB-D technology, proposing a pseudo RGB-D facial recognition framework and providing a data-driven method for generating depth maps from 2D facial images. The research results indicate that with image fusion technology cooperating, the accuracy of facial recognition is significantly improved. Andrejevic and Selwyn [3] considered applying facial recognition technology to campus safety and emotion detection. And they analyzed the positive and negative impacts of this technology in campus applications. To provide a new approach for current cross age facial recognition and help face recognition overcome the information loss during feature fusion and decoupling, a feature fusion and feature decoupling model based on RNN is proposed.

This study proposes the application of RNN based feature fusion and feature decoupling models in cross age facial recognition. To further separate the age and identity features obtained from the decoupling process, this article also utilizes correlation loss to constrain the correlation between the two. And to ensure the effective identity features restoration during the feature decoupling, this article further designs feature restoration losses, so that the restored features can effectively represent the identity information of the face.

II. Related Work

Chen and Lau [4] proposed an identity-level angle triplet loss and used Euclidean distance to calculate similarity for cross-age FR. They also adopted a moderately positive mining strategy. The results on FR datasets showed the effectiveness of this method. Bahmani and Schuckers [5] studied the impact of short-term changes in children facial features on recognition. They used a quality-aware face matcher (MagFace) to analyze the decay relationship between age gaps and matching performance. This study demonstrated that the accuracy of children’s FR was 98% within 6 months. And the accuracy only dropped to 94% within 36 months, indicating that this method had high accuracy in children's FR within three years. Sajid et al. [6] evaluated different backbone structures of deep convolutional neural networks in FR at different age stages. This method used fine-tuning models for face image feature extraction and transfer learning for matching. Experiments showed that this method had application value. Rizwan et al. [7] combined facial expression and age recognition to develop a recognition system. This system not only considered indoor and outdoor applications but also used landmark positioning to avoid the obstruction of FR by masks. The experiment showed that the system had superior performance in recognition accuracy and computation time. Riaz et al. [8] proposed a 3D aging FR model for facial age-related factors such as line and wrinkle changes. This model also paid attention to gender differentiation, and dataset verification showed that this model had high recognition performance with an accuracy of 83.89% in simulation experiments. Kavita and Chhillar [9] analyzed the application of various technologies in face detection, FR, facial expression, and age estimation. And they investigated the application background of FR technology in recent years. Based on the survey results, they summarized the research directions and future trends of FR technology.

Apart from the RNN used in this study, Long Short-Term Memory (LSTM) is also capable of memorizing sequential information. Therefore, Sepas-Moghaddam et al. [10] applied it in the context of FR, and also conducted multi-task learning using different frameworks. The application scenario was multi-view light field image FR. The experiment showed that the LSTM cell structure had higher accuracy in FR than the existing light field recognition methods. Dubey and Jain [11] used an improved VGG16 model for facial expression recognition and matched facial features by using transfer learning. In the simulation experiment, the recognition rate of this method on the validation dataset reached 93.7%, and the facial expression recognition accuracy was higher. Liu et al. [12] used Markov decision and attention control methods to model unordered images and process FR. Meanwhile, pose-guided methods were used for image recognition optimization. In simulation experiments, the model was proven to be effective. Zhang et al. [13] found that regional differences in facial features and environmental differences had an impact on feature extraction in FR. Therefore, they proposed a model that combined deep CNN network and reinforced attention mechanism for FR. After training the model, this method showed high recognition accuracy in mainstream FR scenarios.

In summary, as more and more researchers have focused on cross-age FR and proposed many solutions, cross-age FR has achieved certain results. However, there are still some issues. How to propose an effective solution that preserves not only facial identity information but also has relatively little correlation with age factors is still a challenging problem. Therefore, this study uses feature fusion and feature decoupling to constrain the correlation between facial identity and age feature information. It is hoped to compensate for the cross-age FR technology shortcomings in specific application scenarios.

III. Multi-Feature Fusion and Decoupling Method for Cross Age FR

A. Cross Age Facial Feature Fusion based on RNN

Cross-age FR is one of the important research branches of general FR. Similar to general FR tasks, it can also be divided into two sub-tasks: cross-age face identification and cross-age face verification. The cross-age face identification task involves inputting a given face query image, and then comparing each image in the database with the input query image, one by one, to determine whether they are the same person. It is worth noting that some images in the database may have a large age gap with the query image, which poses a challenge to the general FR algorithm. In the cross-age face identification task, cross-age face verification involves comparing each image with the query image and then determining whether they are the same person. The cross-age face detection and recognition is shown in Fig. 1.

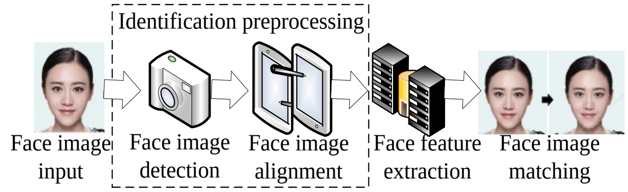


Fig. 1. Process of cross age face detection and recognition.

As shown in Fig. 1, cross-age face detection and recognition usually includes the following steps: face detection, face alignment, face feature extraction, and face matching. Face detection and face alignment are preprocessing steps in FR, which are similar in both general FR algorithms and discriminative-based cross-age FR algorithms. However, the differences are as follows. Firstly, the cross-age FR model uses a cross-age face dataset for training. Secondly, the face features extracted by the discriminative-based cross-age FR algorithm need to have strong anti-interference ability to age factors, that is, only the face features related to the identity need to be extracted. Based on the above steps, each face image can be represented as a feature vector. Face matching is based on the similarity measurement of the extracted face feature vectors, with 1:1 or 1:N matching. And the more similar or exceeding the set threshold, the more likely it is the same person. In face matching, cosine similarity is commonly used. And this study also uses cosine similarity to represent the face similarity. The calculation for cosine similarity is shown in Formula (1).

$$\cos(A,B)=\frac{A\cdot B}{\|A\|_2\cdot\|B\|_2} \tag{1}$$

In Formula (1), A and B represent the feature vectors of two face images, while $\|A\|_2$ and $\|B\|_2$ represent the Euclidean norms of the feature vectors. Deep learning is an important branch of machine learning, which has received widespread attention from researchers in recent years and has made significant progress in applications such as speech recognition, natural language processing, and computer vision [14, 15]. In this study, recurrent neural networks (RNN) in deep learning algorithms are used for feature fusion in cross-age FR. Unlike traditional neural networks, RNN utilizes sequential information within the network. It not only considers the input from the previous moment, but also has a memory of the previous content. So the current sequence output is also related to the previous output. This feature is very important in many application scenarios, as the embedded structure in the data sequence can transmit useful knowledge information. RNN can be regarded as a short-term memory unit including input layer, hidden layer and output layer. And the nodes between hidden layers are no longer connected. The value of hidden layer depends not only on the input at the current time but also on the output value of the hidden layer at the previous time. When training the RNN network for face feature fusion, given that there can be several facial images for each individual in the sample datasets, depicting different ages yet sharing the same identity, such images are initially organized into a sequence of faces, represented by Formula (2).

$$\{F_1^p, F_2^p, ..., F_l^p\} \in \{F_1^p, F_2^p, ..., F_l^p, ..., F_N^p\} \tag{2}$$

In Formula (2), F denotes a face image, l denotes an identity number, P denotes the face sequence length, and N denotes all face images of this identity in a data set. Formula (2) indicates that the sampling strategy for the face sequence is continuous and uninterrupted sampling. The initial face features are extracted from the face images using the RNN backbone network, as shown in Formula (3).

$$x_i^p = G(F_i^p), i \in \{1, 2, ..., l\} \tag{3}$$

In Formula (3), x_i^p denotes the feature extracted from the face image, and G denotes the backbone network of RNN. The RNN network not only considers the input at the previous moment but also has the memory ability for input content. With this characteristic, we treat the face feature sequence of the same identity as a sequence of facial appearance changes with age. And we stack the face feature sequence length and use it as input to the RNN. In order to give different attentions, the attention mechanism of RNN is used to weight the each time step's outcome. The process of RNN face feature fusion is shown in Fig. 2.

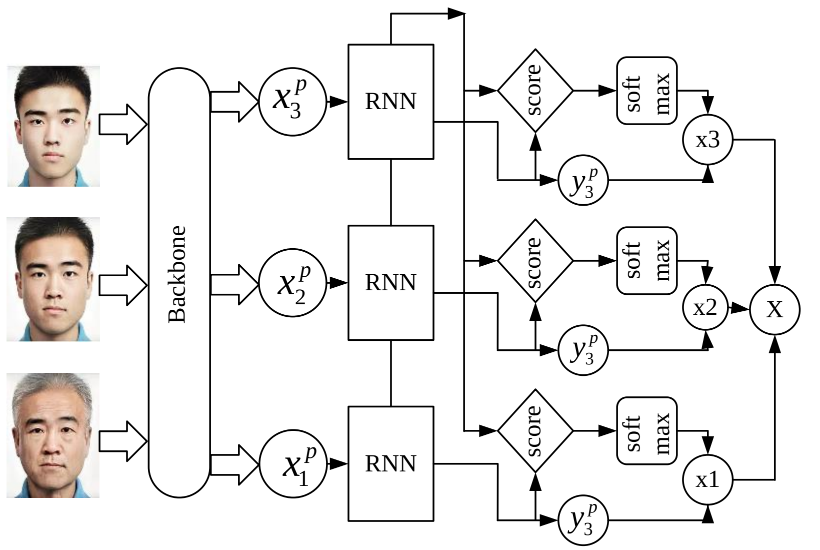


Fig. 2. RNN face feature fusion process.

As shown in Fig. 2, RNN is used to remember features from multiple time periods. X represents the fused face image feature while x1, x2, and x3 represent feature extraction of images at different age stages. This scheme fuses the facial features of the same person at different age stages, thereby achieving feature fusion and making the fused face features have strong robustness against age interference.

B. Optimization Research on Cross Age Face Feature Fusion and Decoupling

Due to the influence of age, a same individual's face is very different at different ages. On occasions, the similarity gap between subjects of the same class surpasses that from different classes [16, 17]. If only the extracted face features are used for cross-age FR, a good performance is hard to obtain. It is necessary to eliminate the age factors influence on FR. Therefore, in this study, feature decoupling is used to optimize the feature fusion. The facial feature decomposition is represented by Formula (4).

$$x^p = x_{age}^p + x_{id}^p \tag{4}$$

In Formula (4), x_{age}^p represents the age feature, while x_{id}^p represents the face identity features. In facial feature decomposition, given the face feature x^p, the RNN fully connected layer is first linearly transformed to obtain age-related features x_{age}^p. In order to reduce the identity information loss in the feature decomposition, this study projects the age-related features in the direction of x^p to obtain the corresponding age projection features, which are used as the final age features, as shown in Formula (5).

$$x_{proj}^p = \frac{x^p x_{r-age}^p}{\left\| x^p \right\|_2} \tag{5}$$

In Formula (5), x_{proj}^p represents the age projection feature in the direction of x^p, while x_{r-age}^p represents the age-related feature. Even after simple feature decoupling, the identity and age feature obtained may still have implicit correlations. Through the above feature decoupling, the study could acquire the features. In order to avoid the recognition degradation caused by the correlation of identity and age features, this study evaluates the correlation between identity and age features, and designs a loss function to minimize the correlation between them, achieving accurate constraint of feature decoupling, as shown in Formula (6).

$$\varphi = \frac{\left\| \frac{1}{m} \sum_{i=1}^{m} (Y_{age}^i - \mu_{age})(Y_{id}^i - \mu_{id}) \right\|}{\sqrt{\sigma_{age}^2 + \varepsilon} \sqrt{\sigma_{id}^2 + \varepsilon}} \tag{6}$$

In Formula (6), φ represents the correlation, μ_{age} and σ_{age}^2 are the mean and variance of Y_{age}^i. μ_{id} and σ_{id}^2 are the mean and variance of Y_{id}^i. ε represents a constant used to ensure numerical stability. Y_{age}^i and Y_{id}^i represent identity variables and age variables, respectively, and their definitions are given in Formula (7).

$$\begin{cases} Y_{age} = w_{age}^T x_{age} \\ Y_{id} = w_{id}^T x_{id} \end{cases} \tag{7}$$

In Formula (7), w represents trainable parameters, x_{age} and x_{id} represent age features and identity features, respectively. Therefore, the correlation loss function between them is represented by $L_c = \exp(|p|)$. FR based on discriminant cross-age is actually a multitask learning method. So besides correlation loss function, this study also needs to construct identity loss function and age loss function for RNN multi-feature fusion and decoupling cross-age FR. The approach taken in this study involves utilizing CosFace Loss as the primary identity loss function, which guides the model in learning essential identity-related information. The calculation is shown in Formula (8).

$$L_{id} = -\frac{1}{N} \sum_i \log\left(\frac{e^{s(\cos(\theta_{yi,i})-c)}}{e^{s(\cos(\theta_{yi,i})-c)} + \sum_{j \neq yi} e^{s\cos(\theta_j, i)}}\right) \tag{8}$$

In Formula (8), N represents the samples number in the training set, while $\cos(\theta_{j,i}) = W_j^T x_{id,i}$ represents the classes number. W represents the weight vector corresponding classifier category j. $x_{id,i}$ represents the identity feature vector corresponding to the face identity label y_i, while $\theta_{j,i}$ represents the angle between the weight vector and the identity feature vector. c, s represent the hyper parameters of the CosFace Loss function. Therefore, this identity loss function improves the model's identity recognition ability by maximizing the inter-class difference and minimizing the intra-class difference. The age loss function is constructed by using the cross-entropy loss function and its calculation is shown in Formula (9).

$$L_{age} = -\log\left(\frac{e^{t_i^p}}{\sum_{j=1}^{m} e^{t_j^p}}\right) \tag{9}$$

In Formula (9), m represents the categories for age labels, and $e^{t_i^p}$ and $e^{t_j^p}$ respectively represent the i-th and j-th term values output by the classifier. Finally, combining the correlation loss function, identity loss function, and age loss function, this study constructs the overall objective function of the supervised learning. It is defined as $L = L_{id} + \alpha L_{age} + \beta L_c$, where α and β are hyper parameters that balance the three loss functions. Finally, to reduce the information loss caused by the feature decoupling process, this study adds the lost face identity information back to the decoupled identity feature. The process of feature recovery is shown in Fig. 3.

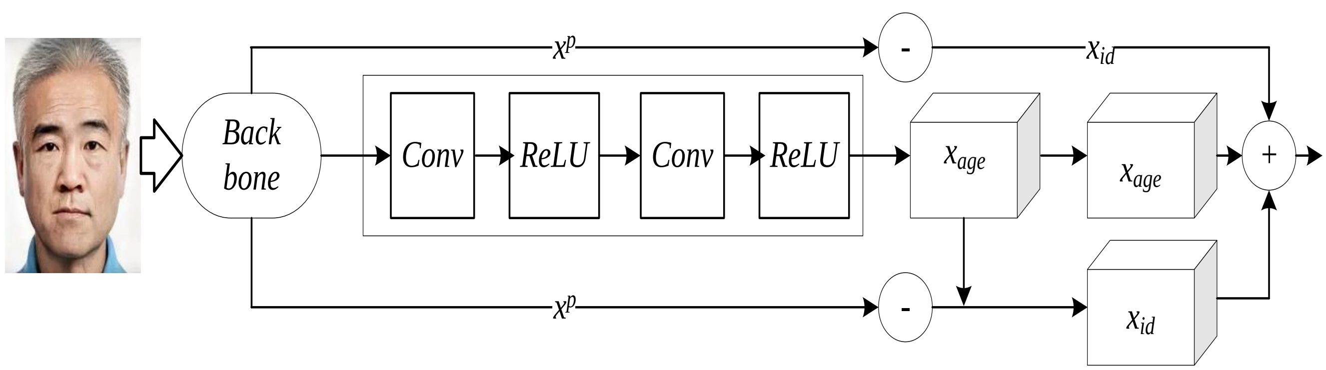


Fig. 3. Facial image feature decoupling and feature restoration.

From Fig. 3, the face features fused by RNN are first reshaped into a feature map. In order to enhance the fitting ability of feature decoupling, this study uses convolution combined with ReLU activation to optimize the decoupling. When decoupling, further recovering the identity feature from the age feature is equivalent to further decomposing the identity-related feature and identity-unrelated feature, which is the final age feature. This process is represented by Formula (10).

$$\begin{cases} \tilde{x}_{id}^p = kx_{age}^p \\ \tilde{x}_{age}^p = x_{age}^p - kx_{age}^p \end{cases} \tag{10}$$

In Formula (10), $\tilde{x}_{age}^p$ and $\tilde{x}_{id}^p$ are the features obtained after decoupling. k represents a learnable vector with a range of 0-1. The calculation of this vector is shown in Formula (11).

$$k = g(\omega_2 f(\omega_1 GAP(x_{age}^p))) \tag{11}$$

In Formula (11), g represents the sigmoid activation function. ω_2 and ω_1 represent the parameters of the fully connected layer. f represents the ReLU activation function. GAP represents global average pooling. The identity-related features extracted from the age feature to the decoupled identity feature are added, and the final recovered identity feature is available, as shown in Formula (12).

$$\hat{x}_{id}^{p} = x_{id}^{p} + \tilde{x}_{id}^{p} \tag{12}$$

The algorithm flow of the face image recognition model constructed in this study is shown in Fig. 4.

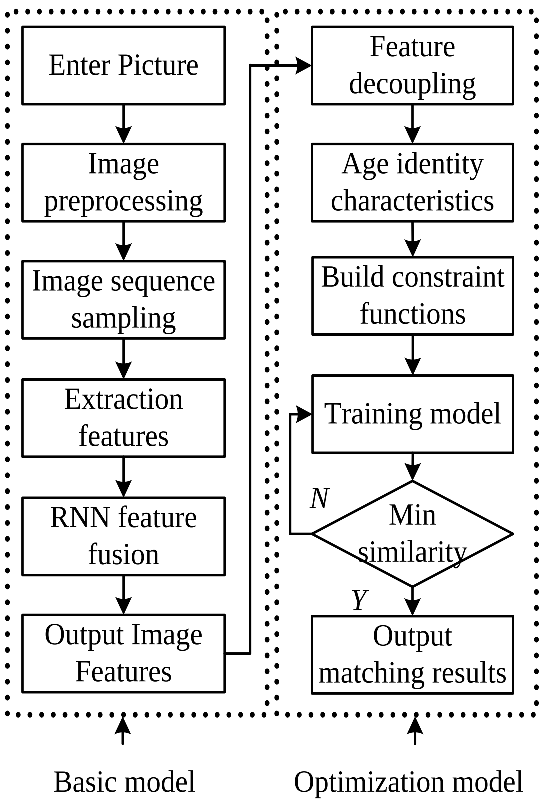


Fig. 4. RNN feature fusion decoupling optimized cross age FR model process.

As shown in Fig. 4, the cross-age face feature fusion and decoupling optimization model constructed in this study mainly uses the RNN and loss function for optimization. The model inputs a sequence of face images obtained by sampling (i.e., face images of the same person at different ages), as well as corresponding identity labels and age labels. Then, the backbone network is used to perform initial feature extraction on the face image sequence to obtain the face feature sequence. After obtaining the face feature sequence, it is fed into the feature fusion module to ensure that the fused face features can contain the adults' identity information to a certain extent. Although the fused features contain rich face identity information, feature decoupling is still necessary to obtain age-related and identity-related features. Finally, the identity recognition loss function is used to supervise the identity feature, the age estimation loss function is used to supervise the age feature. And the correlation loss function is used to constrain the correlation between the age feature and the identity feature.

IV. Application Simulation Analysis of Cross Age FR Model Under RNN Fusion Feature Deconstruction Optimization

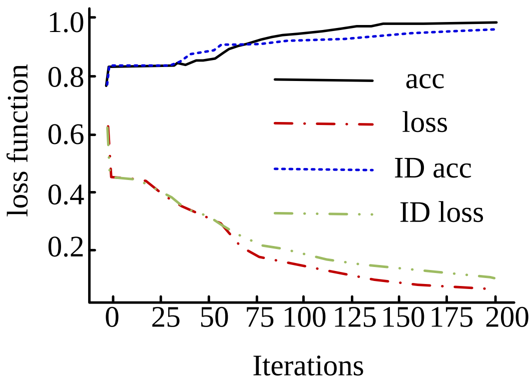
A. *Parameter and Training Analysis of 1DE-RNN Model*

This study used RNN network feature fusion and feature decoupling, as well as related constraint functions optimization, to perform cross-age FR and construct the DE-RNN model. To verify the model, the CACD (Cross Age Celebrity Dataset) and its subset, CACD-VS, were used for simulation validation in this study. The CACD data set is a large-scale face aging data set for cross-age FR tasks. The collected images vary with changes in age, lighting, makeup, and other factors, which improves the simulation. The CACD-VS subset is meticulously annotated by cross-referencing associated images and web content. As for the face verification, each fold of the data set comprises of 200 positive sample pairs and 200 negative sample pairs. In the experiment, the hybrid matrix index and mean Average Precision (mAP) index were used. First, the parameters of the model constructed in this study were set as shown in Table I, and the model was simulated trained on the PyTorch platform.

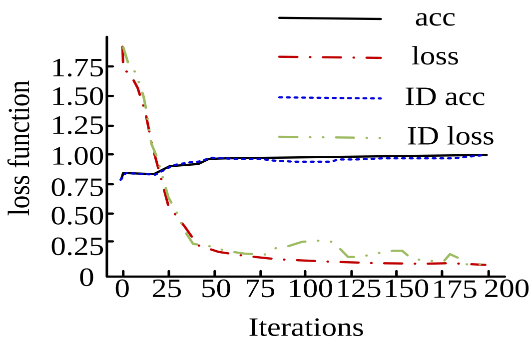
TABLE I. DE-RNN Model Parameter Settings and Training Parameters

Parameter name	Character	Value
Length of face sequence	l	3
CosFace Loss Hyperparameter	c	0.35
	s	64
Loss function equilibrium coefficient	α	0.3
	β	0.3
Pytorch Batch size	/	64
Maximum number of workouts	/	200

After obtaining the above model parameters, this study verified the effectiveness of the proposed method on both cross-age and general face datasets. When testing, Principal Component Analysis (PCA) was used in this study to obtain feature vectors with high discriminability for classification. Cosine distance was used to measure face similarity. The comprehensive loss function during the training process is shown in Fig. 5.



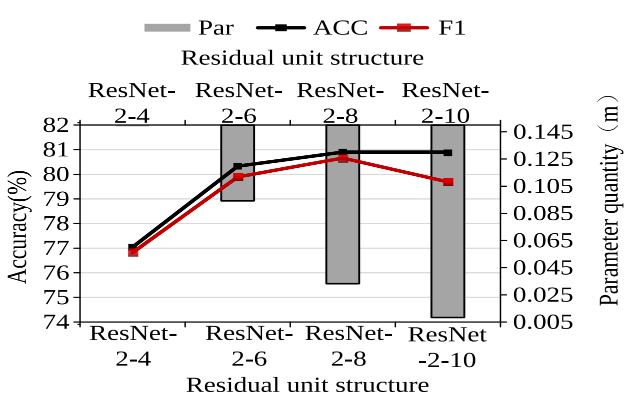
(a)Single-task damage identification model



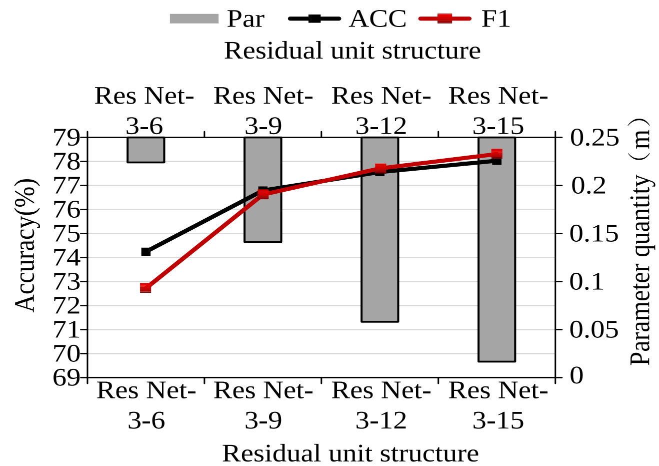
(b)Multi-task damage identification model

Fig. 5. Iterative comparison between single-task model and multi-task learning model.

The training curves in Fig. 5 include the age feature recognition accuracy and identity feature recognition for face matching, represented by acc and ID acc, respectively. From the training curves in Fig. 5, it can be observed that the single-task face image recognition model reaches full convergence after 125 iterations. While the single-task face identity recognition model has not yet reached full convergence. The multi-task learning model reaches a stable convergence state after 75 iterations. It can be seen that by utilizing multi-task learning to simultaneously optimize multiple tasks, the learning efficiency of the model is improved greatly, and the convergence speed of the model is significantly accelerated.



(a)2 Accumulation layer training results



(b)3 Accumulation layer training results

Fig. 6. Effect of different stack layer structure parameters on residual element performance.

In this study, the RNN backbone network was a residual network structure. The simulation tests were conducted to compare the face image recognition capabilities of different residual structures. Fig. 6 shows the effect of the residual blocks on the residual units with 2 and 3 layers of residual network stacks. As shown in Fig. 6, regardless of whether the residual network stack has 2 or 3 layers, the overall performance of the residual unit motion detection accuracy exhibits an increasing trend as the number of residual blocks increases. However, the overall accuracy in the residual network stack with three layers is lower than that in the stack with two layers. The average accuracy and F1 value in the two-layer stack structure were 79.785% and 79.265%, respectively. And in the three-layer structure, they were 76.555% and 76.34%, respectively. In terms of parameter calculation, the computing time of the residual network is negatively correlated with residual blocks. The residual blocks are more, the computation parameters are less. The average parameter for the two-layer stack structure was 0.0711 m, while that for the three-layer structure was 0.109 m. Therefore, the optimal number of residual network stack layers is two. And the best detection performance in the two-layer structure is the unit structure with four residual blocks, with accuracy and F1 values of 80.9% and 80.66%, respectively, which is higher than other structures. The experiment shows that the optimal value for parameter iteration of the RNN backbone network structure is the four residual blocks in the stack with 2 layers. After obtaining the parameters of the locally error-correcting joint residual network algorithm, this study compared the recognition accuracy and training time of 5 traditional CNN network structures. The experimental test data set selected the publicly available CACD data set, and the experimental results are shown in Table II.

TABLE III. TRAINING ACCURACY PERFORMANCE AND TRAINING TIME OF DIFFERENT ALGORITHMS IN CACD DATASET

Network structure	Accuracy (%)	Training time
LeNet5	60.21	50 m 12 s
AlxNet	72.83	47 m 34 s
Vgg16	78.36	24 m 20 s
GoogleNet	82.94	24 m 12 s
ResNet	88.82	22 m 10 s
DE-RNN	91.93	20 m 34 s

The table shows that LeNet5, AlexNet, Vgg16 are classic CNN network structures for image recognition, GoogleNet is the original basic network, ResNet is a single residual network, and DE-RNN is the proposed feature fusion and decoupling RNN model in this study. As shown in the table, the motion image recognition detection accuracy of individual classical CNN networks is decreasing, with accuracy below 80%. The recognition accuracies of GoogleNet and single residual network were 82.94% and 88.82%, respectively. While the proposed algorithm model had a higher accuracy of 91.93% compared to the first five CNN network structures. Moreover, its training time was the shortest, taking only 20 minutes and 34 seconds to train 500 image data.

B. *Performance Analysis Parameters and Training Analysis of Cross Age FR using DE-RNN based on 1DE-RNN Model*

This study analyzed the performance of the DE-RNN cross age FR model on the CACD and CACD-VS datasets. Firstly, this study will compare other face verification methods and use a ten fold cross validation method to report ACC and AUC indicators. The specific experimental results are shown in Table III.

TABLE IV. ANALYSIS OF ACC AND AUC METRICS FOR DIFFERENT CROSS AGE FR MODELS IN CACD-VS DATASET

Methods	ACC	AUC
High-Dimensional LBplia	81.60%	88.80%
HFA	84.40%	91.70%
CARC	87.60%	94.20%
CAN	92.30%	N/A
Center Loss	97.50%	N/A
DAL	99.40%	99.60%
Human Average	85.70%	94.60%
Human Voting	94.20%	99.00%
DE-RNN	97.80%	99.60%

From Table III, the model proposed in this chapter is still competitive compared to other methods. In ACC indicators, DE-RNN performs better, but is lower than DAL. Among the AUC indicators, the AUC value of the DE-RNN model is 99.60%, higher than other algorithm models. In this study, four models with the highest AUC value will be selected for further performance analysis. First, the mAP indicators of the model will be evaluated. The specific results are shown in Fig. 7.

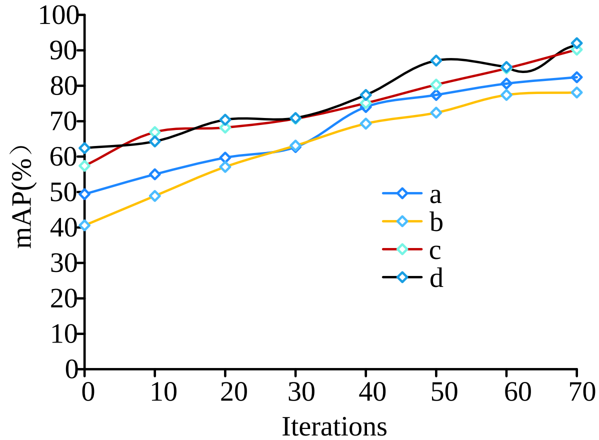


Fig. 7. Comparative analysis of different algorithms for map.

In Fig. 7, a, b, c, and d represent the Human Average model, DAL, the Human Voting model, and the proposed DE-RNN model, respectively. It can be observed from Fig. 6 that the mAP of different models is positively correlated with the number of iterations. When the number of iterations for the test sample was 70, the DE-RNN model had the highest recognition accuracy, with a mAP of 92.4%. The Human Voting model had the second highest recognition accuracy, with a value of 90.2%. The mAP of the Human Average model was 81.8%, while the mAP value of the DAL model was the lowest, at 78.1%. Therefore, in the following experiments, the DAL model will not be considered due to its low mAP value below 80%.

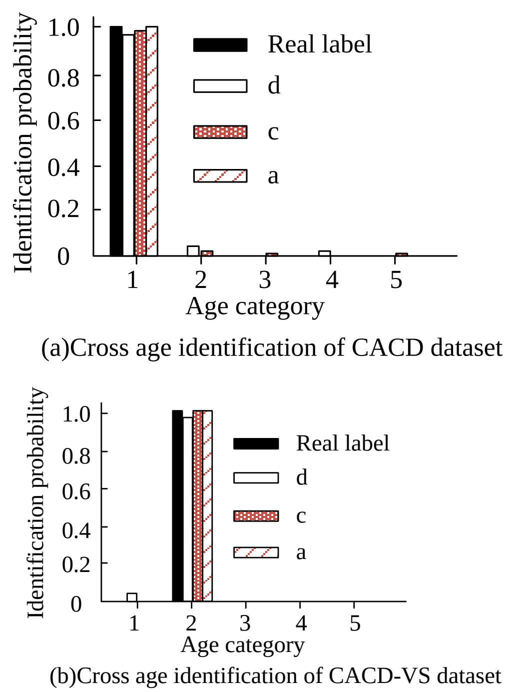
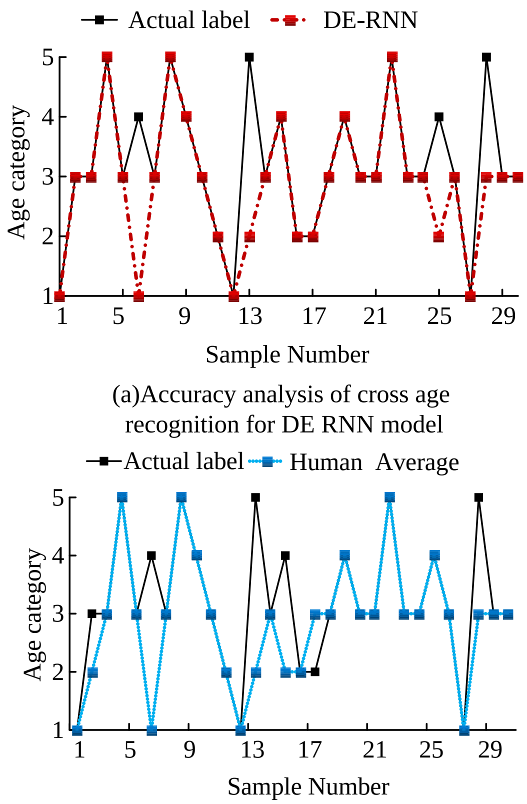


Fig. 8. Application of MTL-1DCNN model for damage identification of double-storey buildings.

This experiment aimed to validate the FR performance of different algorithms for different age groups in the CACD and CACD-VS datasets. The images of different age groups, 0-20, 20-30, 30-45, 45-60, and over 60 years old, in the datasets were used to verify the recognition accuracy. The specific experimental results are shown in Fig. 8, where the horizontal axis represents the age classification, numbered 1-5 from the lowest to the highest. The number of validation samples in both datasets was 200. From Fig. 8, it can be seen that in the face image recognition of the first stage, age 0-20, in the CACD-VS dataset, the DE-RNN model had an accuracy of 96%, higher than other algorithms. In the second stage, age 20-30 of FR in the CACD dataset, the accuracy of the DE-RNN model reached 98%, higher than that of the Human Voting model. Regarding the comparative methods, the Human Average model had better performance than the Human Voting model but was inferior to the DE-RNN model built in this study.



(a)Accuracy analysis of cross age recognition for DE RNN model

(b)Accuracy analysis of cross age recognition for Human Average

Fig. 9. Accuracy of cross age FR using different methods.

Finally, this study combined the CACD and CACD-VS datasets to compare the cross-age face image recognition performance of the two methods in the integrated data set. As shown in Fig. 9, out of the 30 training samples, the model built in this study only had four recognition errors, while the Human Average model had six misjudgments.

V. Result and Discussion

The technology that brings the most convenience to people's daily lives in artificial intelligence is facial recognition technology. As one of the representative artificial intelligence technologies, facial recognition technology has naturally attracted a large number of researchers to conduct research. Since face includes a lot of useful semantic information, such as face, gender, expression, age and identity, face recognition also includes many research directions: face identity recognition, face age estimation, face expression recognition, etc.

In cross age facial recognition, feature fusion and feature decoupling can filter out age independent information in images. Yan et al. [18] proposed a new multi feature fusion and decomposition (MFD) framework for age invariant facial recognition. Based on facial time series, this method could combine feature decomposition and fusion to ensure that the final age independent features effectively represent the face identity information. And it has stronger robustness to the aging process. Huang et al. [19] first utilized attention mechanisms to divide the fused facial information into two parts: identity and age information. Then, they used methods such as multitasking learning and continuous domain adjustment to de correlate the two parts. Unlike conventional single Hot codes, the study aims to use a new identity state model to complete the identity state, and enhances the age smoothness of combined faces through a weighted allocation mechanism. The existing discrimination methods for cross age facial recognition mainly focus on decomposing facial features in images into age related and independent identity features, and then performing facial recognition. In fact, it is inevitable that facial identity information will be lost when feature decoupling. To address this issue, this article first proposes a cross age facial recognition framework based on multi feature fusion and decoupling, which fully learns facial feature representations with high discrimination ability, thereby alleviating intra class differences in cross age facial recognition [20, 21]. Therefore, this study also adopted this approach to optimize the face cross age recognition technology of RNN models. And CACD and CACD-VS datasets are used for validation analysis. In the experiment, it was found that the single task facial image recognition model reached a complete convergence state after 125 iterations. While the multi task learning training model reached a stable convergence state after 75 iterations. Therefore, the multi task optimization project proposed in this study is effective. In addition, through experiments, the research determines the RNN backbone network structure. That is, the best detection performance is the unit structure with two layers of stacked four residual blocks, whose precision and F1 value are 80.9% and 80.66% respectively, higher than other structures.

In performance analysis, the experimental results of facial image recognition between the ages of 0 and 20 in the first stage of CACD-VS data set show that the accuracy of the DE-RNN model is 96%, which is higher than other algorithms. In the second stage of facial recognition between the ages of 20 and 30 in CACD, the accuracy of the DE-RNN model is 98%, which is higher than the Human Voting model. At the same time, this study synthesized CACD and CACD-VS into a data set. And the performance of two methods in cross age facial image recognition was compared in the comprehensive data set. The results showed that out of 30 training samples, the model constructed in the study had only four recognition errors, while the Human Average model had six judgment errors. The experiment shows that the improved recurrent neural network facial cross age recognition model proposed in this study has excellent accuracy and recognition speed.

VI. CONCLUSION

In this study, RNN network was used to construct a face cross-age recognition model with feature fusion and decoupling. The parameters and performance of the model were analyzed on a simulation platform. The experimental results show that the single-task learning model reached convergence after 125 iterations, while the multi-task model built in this study stabilized at 75 iterations. Meanwhile, in the data set analysis, the performance and structure of the backbone networks were compared. And the DE-RNN model had the highest accuracy of 91.93% and the shortest training time, which was 20 minutes and 34 seconds to train 500 image data. Regarding performance analysis, the DE-RNN model had the highest recognition accuracy, with a mAP of 92.4%, followed by the Human Voting model, with a value of 90.2%, and the Human Average model had a mAP of 81.8%. The performance of the three algorithms was compared and analyzed. In the first face image recognition stage, age 0-20, in the CACD-VS dataset, the accuracy of the DE-RNN model was 96%, higher than other algorithms. In the second stage, age 20-30 of FR in the CACD data set, the accuracy of the DE-RNN model reached 98%, higher than that of the Human Voting model. Regarding the integrated data set comparison, out of 30 training samples, the model built in this study only had four recognition errors, while the Human Average model had six misjudgments. The experimental data shows that the model built in this study has superior recognition performance and cross-age face image classification ability. However, the limitation of this study is that the sample classification is based on age stage labels, which still have some noise. For the multi feature fusion in this article, it is necessary to first sample a facial sequence and then fuse the features of the facial sequence. Therefore, this process increases model training and inference time. Therefore, in future experiments, the feature extraction optimization of age attributes will be considered.

ACKNOWLEDGMENT

The research is supported by: Research on virus propagation modeling and optimal control based on node security classification in the Internet of Things environment (No.: 202203021211116).

REFERENCES

[1] W. Ali, W. H. Tian, S. U. Din, D. Iradukunda, and A. A. Khan, "Classical and modern face recognition approaches: A complete review," Multimedia Tools Appl., vol. 80, pp. 4825-4880, October 2021.

[2] B. Jin, L. Cruz, and N. Gonçalves, "Pseudo RGB-D face recognition," IEEE Sens. J., vol. 22, no. 22, pp. 21780-21794, November 2022.

[3] M. Andrejevic and N. Selwyn, "Facial recognition technology in schools: critical questions and concerns," Learn., Media Tech., vol. 45, no. 2, pp. 115-128, November 2020.

[4] X. Y. Chen and H. Y. K. Lau, "The identity-level angular triplet loss for cross-age face recognition," Appl. Intell., vol. 52, no. 6, pp. 6330-6339, September 2022.

[5] K. Bahmani and S. Schuckers, "Face recognition in children: A longitudinal study," 2022 International Workshop on Biometrics and Forensics, pp. 1-6, June 2022.

[6] M. Sajid, N. Ali, N. I. Ratyal, M. Usman, F. M. Butt, I. Riaz, U. Musaddiq, M. J. A. Baig, S. Baig, and U. A. Salaria, "Deep learning in age-invariant face recognition: A comparative study," Comput. J., vol. 64, no. 4, pp. 940-972, April 2022.

[7] S. A. Rizwan, Y. Y. Ghadi, A. Jalal, and K. Kim, "Automated facial expression recognition and age estimation using deep learning," Comput. Mater. & Continuum., vol. 71, no. 3, pp. 5235-5252, January 2022.

[8] S. Riaz, Z. Ali, U. Park, J. Choi, I. Masi, and P. Natarajan, "Age-invariant face recognition using gender specific 3D aging modeling," Multimedia Tools Appl., vol. 78, pp. 25163-25183, May 2019.

[9] K. Kavita and R. S. Chhillar, "Human face recognition and age estimation with machine learning: A critical review and future perspective," Int. J. Elec. Comput. Eng. Syst., vol. 13, no. 10, pp. 945-952, 2022.

[10] A. Sepas-Moghaddam, A. Etemad, F. Pereira, and P. L. Correia, "Long short-term memory with gate and state level fusion for light field-based face recognition," IEEE Trans. Inform. Forensics Secur.,vol. 16, pp. 1365-1379, November 2020.

[11] A. K. Dubey and V. Jain, "Automatic facial recognition using VGG16 based transfer learning model," J. Inform. Opt. Sci., vol. 41, no. 7, pp. 1589-1596, September 2020.

[12] X. F. Liu, Z. H. Guo, J. You, and B. V. K. Vijaya Kumar, "Dependency-aware attention control for image set-based face recognition," IEEE Trans. Inf. Forensics Security, vol. 15, pp. 1501-1512, August 2019.

[13] L. P. Zhang, L. J. Sun, L. N. Yu, X. L. Dong, J. C. Chen, W. W. Cai, C. Wang, and X. Ning, "AR face: Attention-aware and regularization for face recognition with reinforcement learning," IEEE Trans. Biometrics, Behav., Identity Sci., vol. 4, no. 1, pp. 30-42, August 2021.

[14] X. Zhou, Y. Li, and W. Liang, "CNN-RNN based intelligent recommendation for online medical pre-diagnosis support," IEEE/ACM Trans. Comput. Biol. Bioinformatics, vol. 18, no. 3, pp. 912-921, May 2020.

[15] R. Sudharsan and E. N. Ganesh, "A swish RNN based customer churn prediction for the telecom industry with a novel feature selection strategy," Connect. Sci., vol. 34, no. 1, pp. 1855-1876, June 2022.

[16] X. L. Xiao, Y. Y. Chen, Y. J. Gong, and Y. C. Zhou, "Low-rank preserving t-linear projection for robust image feature extraction," IEEE Trans. Image Process., vol. 30, pp. 108-120, October 2020.

[17] M. T. H. Fuad, A. A. Fime, D. Sikder, M. A. R. Iftee, J. Rabbi, M. S. Al-Rakhami, A. Gumaei, O. Sen, M. Fuad, and M. N. Islam, "Recent advances in deep learning techniques for face recognition," IEEE Access, vol. 9, pp. 99112-99142, July 2021.

[18] C. G. Yan, L. X. Meng, L. Li, J. H. Zhang, Z. Wang, J. Yin, J. Y. Zhang, Y. Q. Sun, and B. L. Zheng, "Age-invariant face recognition by multi-feature fusionand decomposition with self-attention," ACM Trans. Multimedia Comput., Commun., Appl. (TOMM), vol. 18, pp. 7-18, January 2022.

[19] Z. Huang, J. Zhang, and H. Shan, "When age-invariant face recognition meets face age synthesis: A multi-task learning framework," Proc. IEEE/CVF Conf. Comp. Vis. Patt. Recognit, pp. 7282-7291, October 2021.

[20] Y. Zhong, W. Deng, J. Hu, D. Zhao, X. Li, and D. Wen, "SFace: Sigmoid-constrained hypersphere loss for robust face recognition," IEEE Trans. Image Process., vol. 30, pp. 2587-2598, January 2021.

[21] L. P. Zhang, L. J. Sun, L. N. Yu, X. L. Dong, J. C. Chen, W. W. Cai, C. Wang, and X. Ning, "ARFace: attention-aware and regularization for face recognition with reinforcement learning," IEEE Trans. Biometrics, Behav., Identity Sci., vol. 4, no. 1, pp. 30-42, August 2021.