Electronic Research Archive

ERA, 31(11): 6585–6599.
DOI: 10.3934/era.2023333
Received: 29 June 2023
Revised: 26 September 2023
Accepted: 27 September 2023
Published: 13 October 2023

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

Research article

Portrait age recognition method based on improved ResNet and deformable convolution

Ji Xi[1,*], Zhe Xu[1], Zihan Yan[1], Wenjie Liu[1] and Yanting Liu[2]

[1] School of Computer Information Engineering, Changzhou Institute of Technology, Changzhou 213022, China
[2] School of Software and Big Data, Changzhou College of Information Technology, Changzhou 213032, China

* **Correspondence:** Email: xiji@czust.edu.cn; Tel: +8613861060088; Fax: +86051988510366.

Abstract: ResNet-based correlation models excel in age recognition algorithms, but specific age recognition research is currently limited and often plagued by substantial errors. We introduce an enhanced portrait age recognition algorithm based on ResNet, using CORAL (consistent rank logits) rank consistent ordered regression instead of traditional classification to predict precise ages. We further improve this approach by incorporating DCN (deformable convolution), resulting in the DCN-R model. DCN dynamically adjusts convolution kernels for diverse faces, improving accuracy and robustness. We tested DCN-R34 and DCN-R50 against the SOTA model, achieving better results with the same complexity. This reduces the computational load while maintaining or enhancing performance.

Keywords: ResNet; deformable convolution; portrait age recognition; rank-uniform ordered regression

1. Introduction

Portrait age recognition is a technique that utilizes computer image recognition technology to accurately determine a person's age based on their facial features. Various algorithms are employed to predict the specific age range of the identified individual. Huang et al. introduced a comprehensive multi-task framework called MTLFace to mitigate the impact of age changes on face recognition [1]. Abu et al. proposed the use of a CNN (convolutional neural network) to detect gender and estimate age from a single person's photo, followed by a double-checking layer verifier [2]. Rafique et al. made

enhancements to CNN in order to improve age recognition [3]. Othmani et al. presented a face recognition method based on the AAE knowledge migration assessment framework, which utilized manually designed features to encode the aging model [4]. Sharma et al. improved CNN by replacing manual features to enhance the computational performance of the model [5]. However, these methods have limitations as they can only predict age groups, rather than specific ages. Traditional age recognition methods employ multi-classification techniques to categorize age into specific groups, such as 20–30 years old or 30–40 years old. Sakata et al. proposed a gait-based CNN estimation method [6]. Hsu et al. further suggested a data augmentation technique using random occlusion based on CNN [7]. On the other hand, Mamatkulovich et al. applied age identification grouping to practical problems [8]. This study aims to enhance age identification by transforming it into a regression problem, thereby enabling the prediction of accurate ages instead of age groups.

In the realm of multiple binary classification problems, Li and Lin put forward the idea that the extended binary classification method demonstrates superior classification accuracy and faster training speed in comparison to the conventional binary classification approach [9]. Nevertheless, this method frequently leads to discrepancies in the predictions made by a single binary classifier. To tackle this problem, Cao et al. introduced the CORAL framework for ordered regression algorithms. This algorithm has the capability to merge multiple binary classification tasks into an ordered regression problem, guaranteeing a consistent ranking of multiple binary classifications and facilitating the prediction of individuals' age in face images [10].

In addition, ResNet series neural networks have been widely utilized in age recognition tasks. However, due to the complexity of facial features, changes in facial age can result in alterations in facial shape. Traditional convolution operations may not accurately capture these subtle changes, necessitating the use of higher layers for feature extraction. For instance, ResNet50 and OR-CNN, recently proposed by Paplham et al. (2023) [11], have achieved promising outcomes in portrait age recognition. Nevertheless, this comes at the cost of increased model complexity. Deformable convolution can adapt to different face shapes by slightly modifying the shape of the convolution kernel, enabling more accurate extraction of facial features with a relatively minor impact on model complexity [12]. This approach can enhance the accuracy of age identification in portraits.

Based on the aforementioned analysis and considerations, we introduce a novel age recognition model called DCN-R, based on the ResNet neural network. Simultaneously, the traditional multi-classification method is transformed into multiple binary classification methods within the CORAL framework, aiming to enhance classification accuracy and minimize the average absolute error (*MAE*) value. Additionally, deformable convolution is incorporated and its structure is innovatively improved to enhance the model's generalization ability and robustness.

2. Improved description of network structure

2.1. DCN-R network architecture

Let us examine the DCN-R34 model, a derivative of the ResNet34 network architecture, as an illustrative example. Figure 1 provides a schematic representation of the ResNet34-based DCN-R34 network design. This design incorporates distinctive modifications. Primarily, the conventional softmax activation function is substituted with the Sigmoid activation function. While the softmax activation generates a probability distribution for each category, the Sigmoid activation yields binary

classification outcomes for each category, thereby altering the output format. Additionally, an enhancement is introduced by replacing the initial convolution of the fourth network layer in the ResNet34 model with DCN deformable convolution to augment performance.

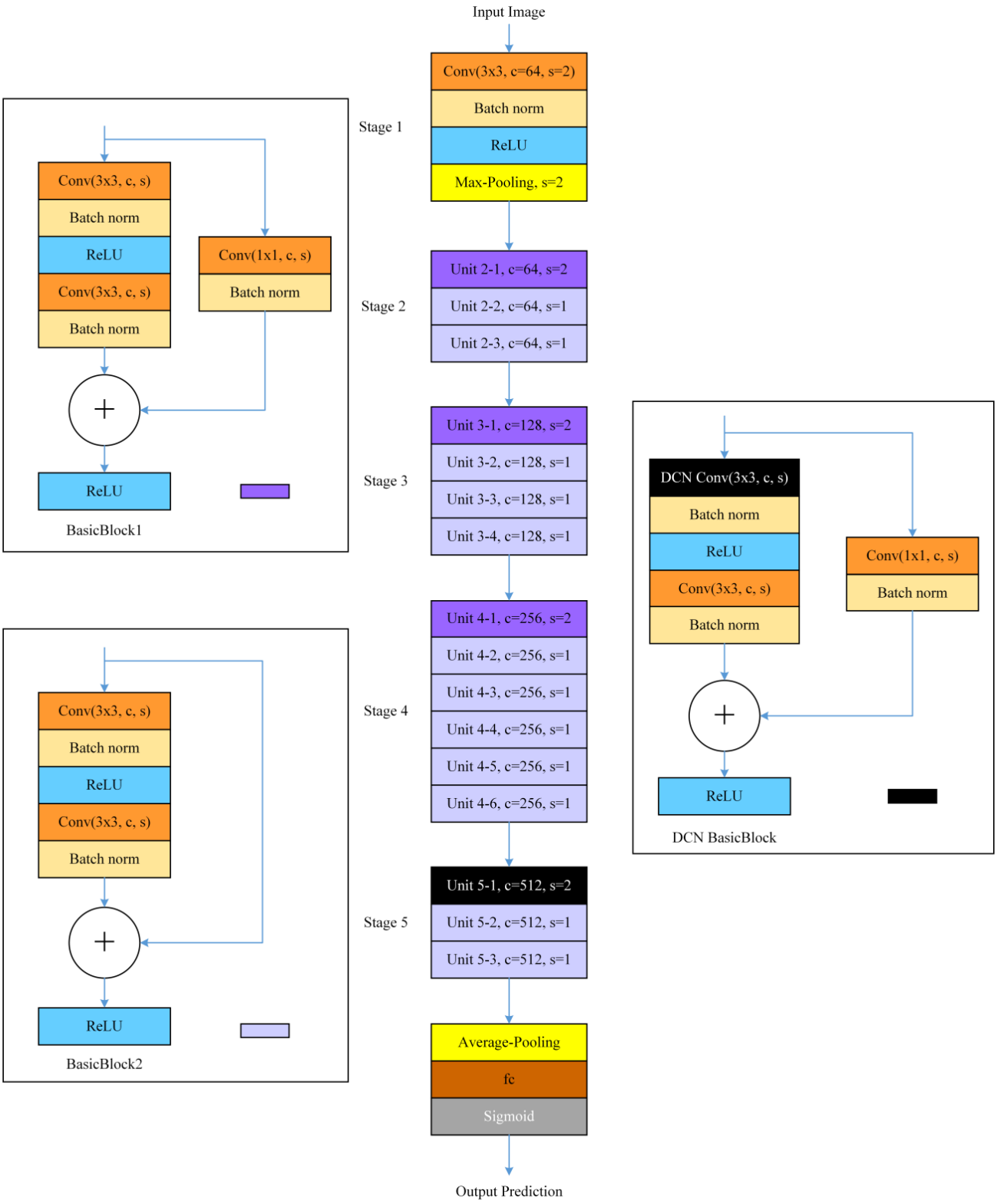


Figure 1. Overall architecture of DCN-R34.

Figure 1 delineates the specific structural components employed in the overall architecture. The top-left block, denoted as BasicBlock1 and shaded in dark purple, is deployed in the initial layer of stages 2–4. The bottom-left block, designated as BasicBlock2 and represented in light purple, finds use in all layers except the first layer of stages 2–4. On the right-hand side, the DCN BasicBlock denotes the adapted block featuring DCN deformable convolution, depicted in black. The yellow block signifies the pooling method.

For further elucidation, each module within ResNet34 encompasses the following categories of residual blocks:

Module 1 encompasses 3 residual blocks intended for the extraction of low-level features from the image domain, such as edges and textures.

Module 2 is comprised of 4 residual blocks employed to extract intermediary features from the image, encompassing aspects such as contours and shapes.

Module 3 incorporates 6 residual blocks, strategically applied to capture more intricate image features, including object components and overall structure.

Module 4 comprises 3 residual blocks, primarily serving the purpose of extracting the most advanced image features and facilitating tasks such as classification or regression.

Notably, the DCN-R34 network model adheres to the architectural principles of ResNet34 within the initial three modules. The substantial alteration lies in the replacement of the initial convolution operation within the first residual block with deformable convolution in the fourth module, constituting a fundamental adaptation for performance enhancement.

2.2. Improved classification method

Although ResNet, as a deep convolutional neural network, has powerful feature extraction capabilities. It effectively extracts features from portrait images in the training data, offering high model accuracy and stability. This inherent strength helps avoid overfitting and underfitting issues. However, the original classification method of ResNet34 is limited to perform a single classification task. It is unable to simultaneously identify multiple attributes such as gender and race, which limits its application in portrait age recognition. Additionally, the original classification method of ResNet can only identify a fixed age range, lacking the ability to perform more detailed age recognition. Furthermore, a large amount of annotated data is required for training, therefore, for portrait age recognition, it may be necessary to acquire more annotated data and employ more sophisticated annotation methods, which would increase the difficulty and cost of data collection and processing.

CORAL (Correlation Alignment) is a technique used for aligning features between source and target domains in domain adaptation. CORAL rank-uniform ordered regression is an extension of CORAL specifically designed to address the ordered output problem. The main idea behind CORAL rank-uniform ordered regression is to align the features of the source and target domains, convert the multi-classification problem into a binary classification problem and ensure that the predicted results are ordered consistently with the actual labels by sorting the samples. To illustrate this process, we use ResNet34 from the ResNet series as an example. Figure 2 depicts the specific steps involved.

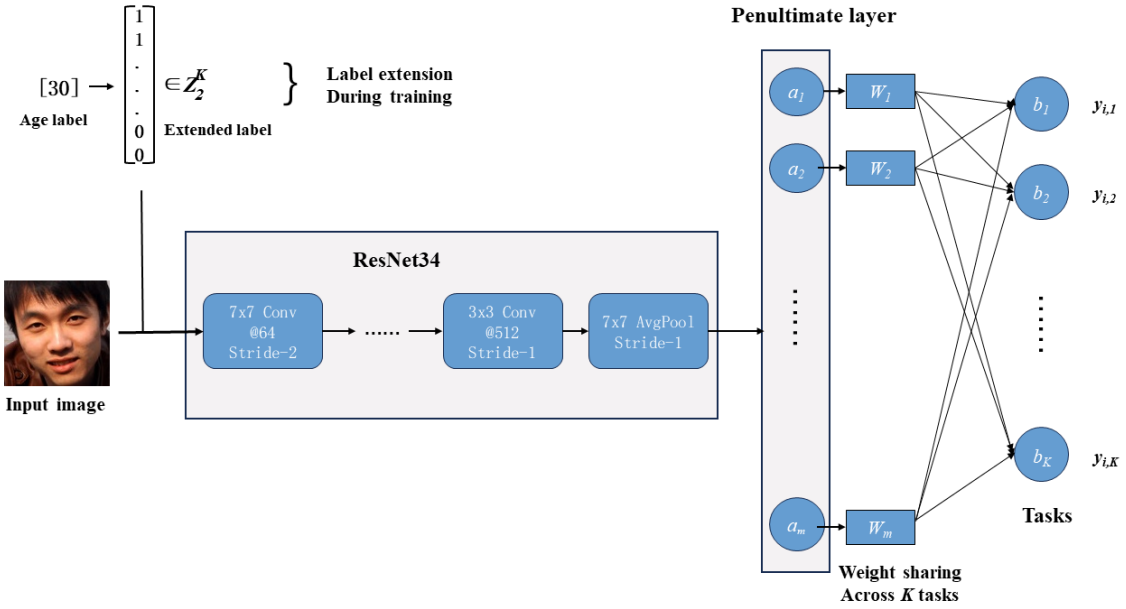


Figure 2. CNN's rank-consistent logic diagram for age prediction.

Specifically, we assume that we have n samples, each with a label $L_i \in \{1,2,\ldots,K\}$, where K represents the total number of categories. For each sample i, we assign it a K-dimensional vector $y_i = (y_{i,1}, y_{i,2}, \ldots, y_{i,K})$, where $y_{i,K}$ denotes the probability of sample i belonging to category K.

And then we arrange the samples in a descending order as $y_{i,1} > y_{i,2} > \cdots > y_{i,K}$. The sorted order of $y_{i,K}$ is then used to generate a sorting matrix $P = (p_{i,j})_{n\times K}$, where $p_{i,j} = 1$ indicates that sample i precedes category j, and $p_{i,j} = 0$ indicates that sample i follows category j.

Furthermore, we propose a rank consistency loss function $J(P)$ to quantify the disparity between the predicted order of the model's outcomes and the actual ordering of the labels.

$$J(P) = \sum_{i=1}^{n}\sum_{j=1}^{K-1}\sum_{k=j+1}^{K} max(0, \Delta_{i,j,k}) \tag{1}$$

where $\Delta_{i,j,k} = p_{i,j} - p_{i,k}$ represents the sequence difference between the sample i in the category j and the category k. If $p_{i,j} > p_{i,k}$, so $\Delta_{i,j,k} = 1$, otherwise $\Delta_{i,j,k} = 0$.

Lastly, we employ the rank consistency loss function $J(P)$ as the objective function to minimize the loss in CORAL. Minimizing this objective function enhances the model's predictive results to align more closely with the actual label order, thereby enhancing the model's generalization capability.

In this paper, we propose a modified version of the ResNet34 model, referred to as ACR34, which incorporates the CORAL framework. This modification is done to facilitate the subsequent discussions. The modifications made to the ResNet34 model are depicted in Figure 3, resulting in the ACR34 model.

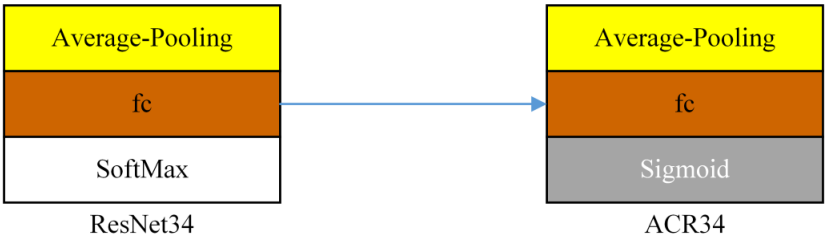


Figure 3. Modified model structure (ResNet34 on the left and ACR34 on the right).

2.3. Deformable convolution

Deformable Convolution is a convolutional neural network architecture that has been improved [12]. It has the ability to learn the deformation of each convolutional kernel during the convolution process. The structure of deformable convolution is typically illustrated in Figure 4.

In standard convolution, the convolution operation is fixed in position. However, by introducing an offset, the position of the convolutional operation can be selected based on supervised information. This enables better adaptation to the various sizes and shapes of the target, resulting in richer and more concentrated features extracted specifically from the target itself.

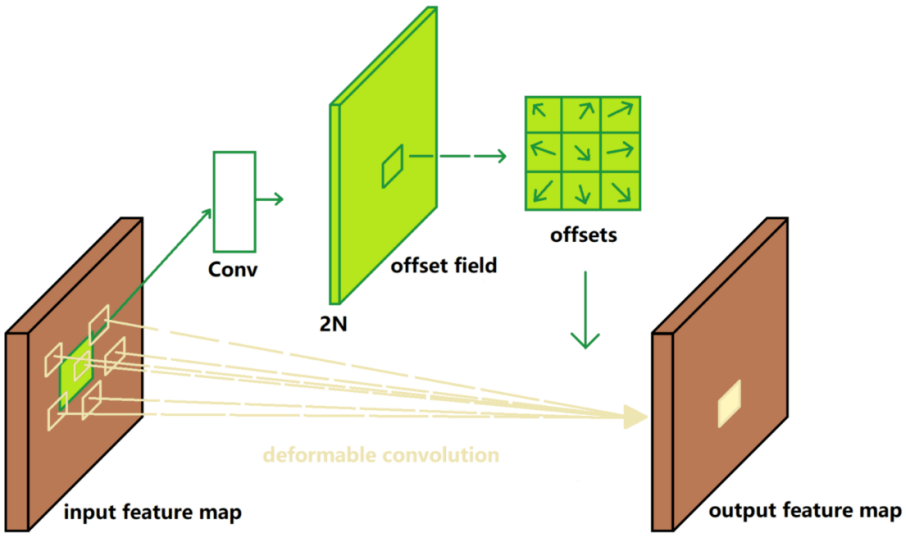
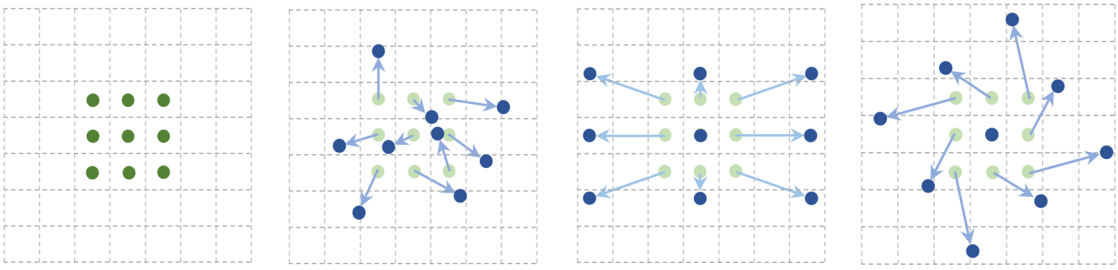


Figure 4. Deformable convolution 3×3.

In deformable convolution, each convolution kernel is not a fixed matrix but a deformable matrix controlled by a set of control points. These control points can be learned automatically during the training process to adjust the shape of the convolution kernel and adapt to different image structures. This is illustrated in Figure 5.



(a) (b) (c) (d)

Figure 5. Ordinary convolution (a) and deformable convolutions (b)–(d) are shown.

In Figure 5, (a) represents a standard convolution operation, while (b)–(d) depict adaptive deformable convolutions. Deformable convolution refers to the displacement of sampling positions in the standard convolution operation, allowing the convolution kernel to cover a larger range during training. (c) and (d) are specific instances of (b), demonstrating that deformable convolution encompasses various transformations, such as scaling, aspect ratio adjustment and rotation.

Two-dimensional convolution involves two steps: First, sampling the input feature map x with a regular grid R, and second, summing the sampled values with weighted w. The grid R determines the size and expansion of the receptive field. For instance, it can be defined as a 3×3 kernel with a dilation of 1.

$$R = \{(-1,-1),(-1,0),\dots,(0,1),(1,1)\} \tag{2}$$

The Eq (3) demonstrates the position p_0 on the output feature graph y.

$$y(p_0) = \sum_{p_n \in R} w(p_n) \cdot x(p_0 + p_n) \tag{3}$$

p_n enumerates the positions in R. In deformable convolution, regular grid R is subject to augmentation offset $\{\Delta p_n | n = 1,\dots,N\}$, when $N = |R|$, Eq (3) becomes

$$y(p_0) = \sum_{p_n \in R} w(p_n) \cdot x(p_0 + p_n + \Delta p_n) \tag{4}$$

Now, the sampling is conducted at irregular and displaced locations $p_n + \Delta p_n$. As the displacement Δp_n is typically a fraction, Eq (4) becomes

$$x(p) = \sum_q G(q,p) \cdot x(q) \tag{5}$$

where p represents any position, that is $p = p_0 + p_n + \Delta p_n$, q in Eq (4) enumerates all integral space positions in feature mapping x, $G(.,.)$ Expressed as bilinear interpolation kernel. However, notice that G is two-dimensional. It is divided into two one-dimensional kernels.

$$G(q,p) = g(q_x,p_x) \cdot g(q_y,p_y) \tag{6}$$

The Figure 6 demonstrates the distinct disparity in performance between the receiving field and the conventional convolution.

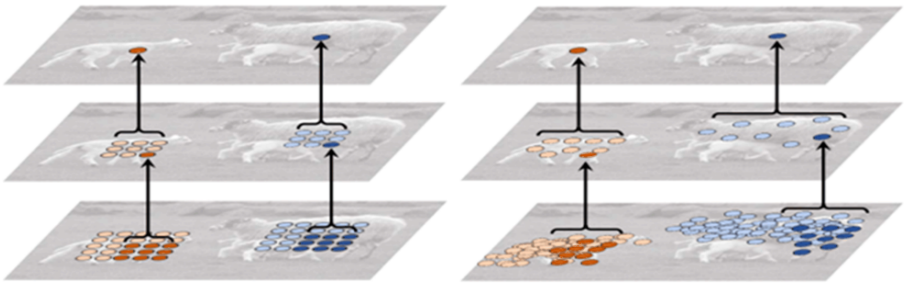


Figure 6. Sampling contrast: convolution (left) vs deformable convolution (right).

Since deformable convolution introduces additional computational cost, the placement of deformable convolution needs to be carefully considered. In terms of location, the first network block of ResNet34 consists of only a 7×7 convolutional layer. Applying deformable convolution to this layer would not adequately capture facial features, resulting in inefficient sampling. Placing deformable convolution in the second network block improves the results slightly compared to the first block, but it is still limited. The third and fourth network blocks serve as transitional blocks from shallow to deep layers, making it difficult to capture deep layer characteristics. However, in the fifth network block, where deep image information is obtained, deformable convolution can be effectively utilized for information extraction. Therefore, based on these reasons, it is decided to place deformable convolution in the fifth network block, resulting in the modification of ACR34 to obtain the DCN-R34 model.

3. Experiments

3.1. Experimental setting

As we primarily focus on analyzing the age of Asian individuals in portraits, we have chosen to utilize the AFAD Asian Open Face dataset for our experiment. This dataset comprises face images of over 20,000 Asian individuals, with an equal distribution of males and females, spanning ages 0 to 70 years. Currently, the dataset is widely used in various fields of research and application, such as face recognition, face detection and face attribute analysis.

To compare the performance of different models, namely ResNet34, ACR34, DCN-R34 and DCN-R50, we have conducted our experiments using this dataset. It is worth noting that the AFAD dataset contains a substantial amount of data for individuals aged between 15 and 40 years, while the data for other age groups is relatively limited. Therefore, we have chosen to train our models using the age range of 15 to 40 years in order to achieve higher accuracy and enhance the reliability and robustness of the models.

Since the faces in the dataset were already centered, no further preprocessing was necessary. Figure 7 displays a partial view of the dataset.

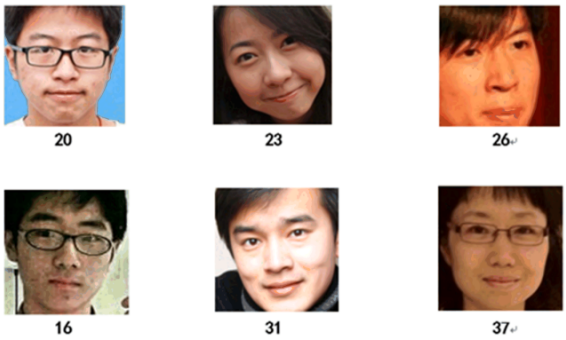


Figure 7. Part of the AFAD dataset is shown.

The information regarding the experimental software and hardware is presented in Table 1.

Table 1. Experimental software and hardware information.

Name	Version
CPU	Inter (R) Core (TM) i7-9750H CPU @ 2.60 GHz
GPU	NVIDIA GeForce RTX 2060 6 GB
RAM	32 GB
Operating system	Windows 11 professional edition 64 bit (10.0, Version 22621)
Software platform	PyCharm community edition 2022.2.3
Python	3.6.13
PyTorch	1.10.2
CUDA	12.0

In terms of training strategies, we utilize the SGD (Stochastic Gradient Descent) optimization algorithm and the Adam (Adaptive Moment Estimation) as the optimal optimization algorithm. The learning_rate was adjusted to 0.001 in order to ensure a more stable training process and prevent issues such as gradient explosion or disappearance.

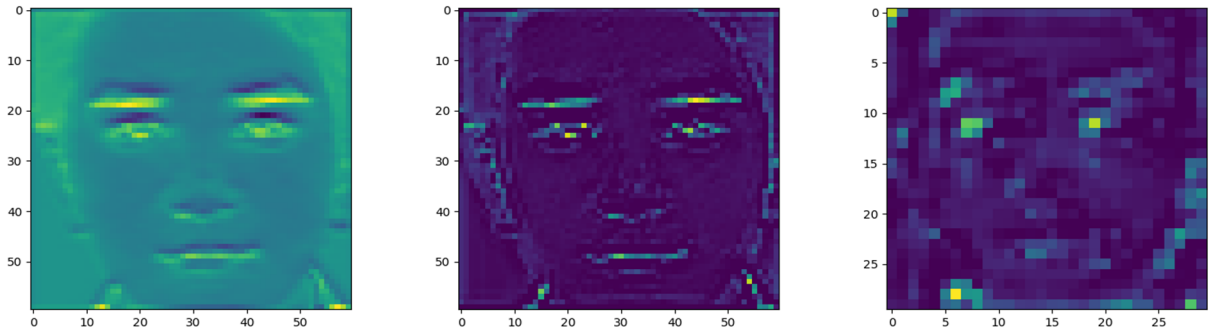
The batch size for training was set to 400. This decision was made considering the large amount of data available for individuals between the ages of 15 and 40 in the AFAD dataset. Using a larger batch size, the training process can be accelerated and the GPU resources can be fully utilized through hardware parallelization. Additionally, a larger batch size helps reduce variance in the stochastic gradient descent algorithm, allowing the model to better approximate the distribution of real data. This ultimately improves the stability and generalization ability of the model.

Moreover, we employ two methods to enhance the data:

1) Resizing each picture to a dimension of 128×128 pixels can enhance the visual quality and sharpness of lower resolution images. Additionally, this reduces computational load, thereby enhancing the speed and efficiency of image processing.

2) The image is promptly cropped to exclude a section of 120×120 pixels. This cropping technique enables the model to be exposed to a wider range of images during training, thereby enhancing the model's ability to generalize.

Taking one of the data graphs as an example, let's consider the feature extraction of the first three layers in the model. The results can be seen in Figure 8.



(a) (b) (c)

Figure 8. (a)–(c) show the feature extraction results of the first, second and third layers of DCN-R34.

3.2. Model quality measure

In this study, the model quality will be evaluated using *MAE* (Mean Absolute Error) and *RMSE* (Root Mean Square Error). Both *MAE* and *RMSE* are metrics used to quantify the discrepancy between the model's predicted results and the actual results. *MAE* stands for mean absolute error, while *RMSE* stands for root mean square error.

MAE is computed by summing the absolute differences between the predicted values and the actual values of each sample, and then dividing by the total number of samples.

$$MAE = \frac{(|y_1-\hat{y}_1|+|y_2-\hat{y}_2|+\cdots+|y_n-\hat{y}_n|)}{n} \tag{7}$$

RMSE is computed by summing the squared differences between the predicted and actual values for each sample. This sum is then divided by the total number of samples, and the square root of the quotient is taken.

$$RMSE = \sqrt{\frac{[(y_1-\hat{y}_1)^2+(y_2-\hat{y}_2)^2+\cdots+(y_n-\hat{y}_n)^2]}{n}} \tag{8}$$

In Eqs (7) and (8), y_i refers to the true value of sample i, $\hat{y}_i$ refers to the predicted value of sample i and n represents the total number of samples.

The utilization of *MAE* and *RMSE* as metrics for evaluating the quality of models primarily relies on the following rationales:

1) Compared to basic accuracy metrics, such as simple accuracy indicators, *MAE* and *RMSE* are more effective in assessing the predictive capability of a model. This is because they consider the magnitude of the difference between the predicted and actual results.

2) *MAE* and *RMSE* are less sensitive to outliers compared to other evaluation metrics. This is because they measure the average difference between predicted and actual values, either by taking the absolute value or the square value of the error, respectively.

3) *MAE* and *RMSE* are quantitative measures that offer a more comprehensive analysis of the model's performance and potential issues, thus aiding in the enhancement of the model.

3.3. Experimental result

First, the ResNet34 architecture was employed to train the testability model. Subsequently, the AFAD dataset was used to evaluate the model's performance, yielding an optimal *MAE* value of 3.52 and *RMSE* value of 5.01. These metrics provided a fundamental evaluation measure.

On this basis, the pre-existing model was adjusted, and then the ACR34 network architecture was employed for training. The deformable convolution was incorporated into the fifth layer of the network. By modifying the model structure, the DCN-R34 model was obtained. The hyper-parameters were set with a learning rate of 0.001 and a seed of 0, and subsequently, the training process was conducted.

The cost loss value of DCN-R34 after training with 100 epochs is shown in Figure 9. It is evident that the cost cross entropy loss function consistently decreases during training, without significant fluctuations. This observation suggests that the model exhibits stable performance throughout the training process.

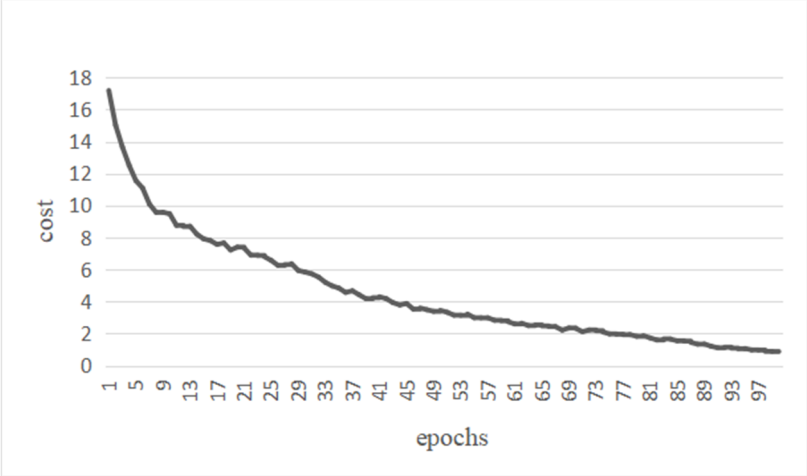


Figure 9. Cost loss function diagram of DCN-R34 model.

Furthermore, we consider that a deeper network architecture can potentially yield greater advantages. Therefore, we proceeded to incorporate CORAL and DCN deformable convolution techniques into the ResNet50 network structure, resulting in the creation of DCN-R50. Throughout our exploration process, we compared the performance of these four models.

Finally, the experimental results of the four models can be obtained and are presented in Table 2.

Table 2. The results of the four models.

Model	*MAE*	*RMSE*
ResNet34	3.52	5.01
ACR34	3.44	4.81
DCN-R34	3.34	4.54
DCN-R50	**3.23**	**4.32**

3.4. Model comparison and result analysis

We compare the suggested approach that combines CORAL and DCN (DCN-R34) with the ordered regression method (OR-CNN) proposed by Niu et al. (2016) [13] and the Ranked consistent ordered regression method (CORAL) proposed by Cao et al. (2020) [10].

Furthermore, we have included the most recent OR-CNN method proposed by Paplham et al. (2023) [11], which is based on the ResNet50 architecture, for the purpose of comparison. It is important to note that this method utilizes ResNet50 rather than ResNet34. The specific results of this comparison can be seen in Figures 10 and11.

Upon analyzing the results presented in Figures 10 and 11 and Table 3, it becomes clear that DCN-R34 outperforms ResNet34, OR-CNN and CORAL-CNN in terms of regression metrics such as *MAE* and *RMSE*. This superiority is particularly evident when ResNet34 is used as the base network layer. These findings suggest that combining DCN deformable convolutions with the CORAL rank uniform ordered regression framework can enhance the model's efficiency for complex image tasks like age recognition. Furthermore, when comparing the higher network layers of the ResNet50-based OR-CNN approach, the *MAE* value is reduced by approximately 1.8% and the *RMSE* value is reduced by around 2% compared to DCN-R34. However, the DCN-R50 model, also based on ResNet50, outperforms the ResNet50-based OR-CNN method. Our DCN-R50 model achieves a 1.5% decrease in *MAE* value and a 3% decrease in *RMSE* value.

Table 3. Comparison of *MAE* and *RMSE* values of each model and the underlying network structure.

Method	*MAE*	*RMSE*	ResNet
OR-CNN (Niu et al., 2016)	3.51	4.75	ResNet34
CORAL-CNN (Cao et al., 2020)	3.47	4.71	ResNet34
DCN-R34 (our proposed)	3.34	4.54	ResNet34
OR-CNN (Paplham et al., 2023)	3.28	4.45	ResNet50
DCN-R50 (our proposed)	**3.23**	**4.32**	**ResNet50**

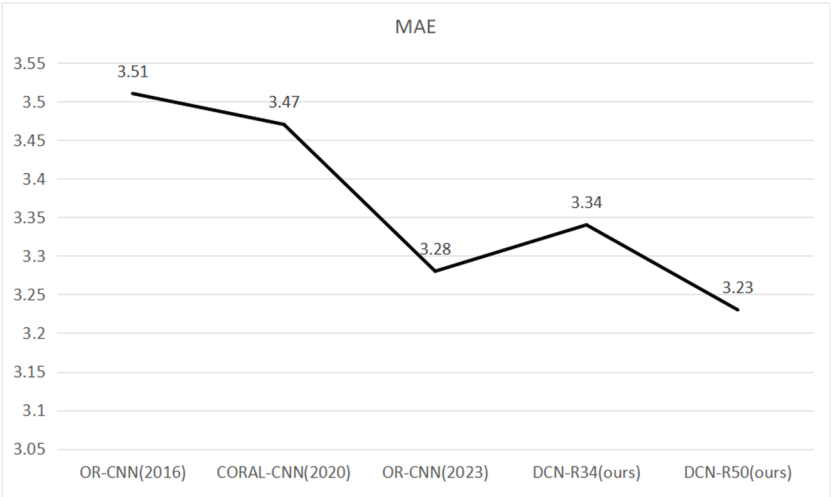


Figure 10. Comparison of *MAE* and *RMSE* values of each model.

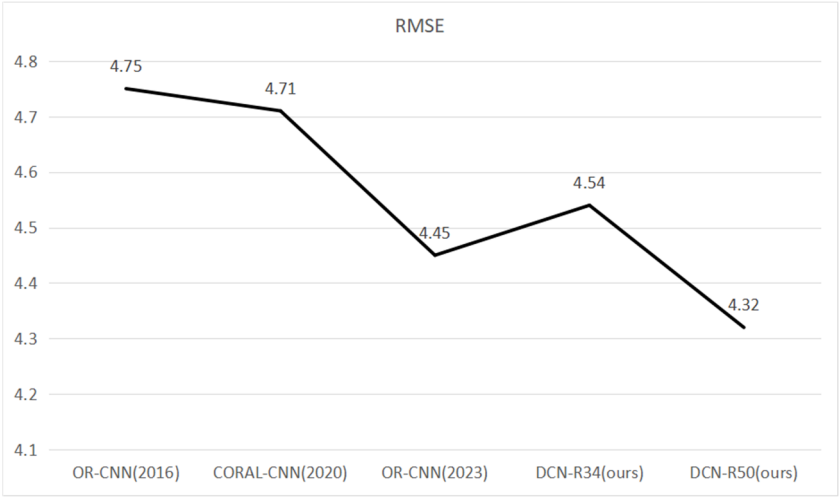


Figure 11. Comparison of *MAE* and *RMSE* values of each model.

Next, we analyze the parameter count and FLOPs (floating point operations) of DCN-R34 and the latest ResNet50 model, comparing them with OR-CNN (2023). The comparative analysis is presented in Table 4.

Table 4. Comparison of parameters and FLOPs.

Method	Parameters	FLOPs
OR-CNN (Niu et al., 2016)	22.8×10^6	4.1×10^9
CORAL-CNN (Cao et al., 2020)	22.2×10^6	3.8×10^9
DCN-R34 (our proposed)	$\mathbf{22.2 \times 10^6}$	$\mathbf{3.7 \times 10^9}$
DCN-R50 (our proposed)	26.4×10^6	4.1×10^9
ResNet50 OR-CNN (Paplham et al., 2023)	26.6×10^6	4.4×10^9

As evident from the data presented in Tables 3 and 4, within the same ResNet34-based model framework, DCN-R34 outperforms the other two ResNet34-based models in terms of performance, parameter count and FLOPs. While the ResNet50 model combined with OR-CNN, as proposed by Paplham et al. (2023), exhibits superior model recognition performance compared to DCN-R34, it is crucial to note that this combination significantly escalates model complexity. Specifically, there is a 20% increment in parameter count and a 19% rise in FLOPs. This underscores that the ResNet50 with OR-CNN approach substantially elevates model complexity to achieve only a marginal enhancement in recognition performance. Consequently, the improvement in recognition performance does not warrant the substantial increase in model complexity.

To facilitate an objective comparison between the CORAL and DCN deformable convolution methods and the combined OR-CNN approach introduced by Paplham et al. (2023), we applied both CORAL and DCN deformable convolution techniques to the ResNet50 architecture, yielding the DCN-R50 model. The results from the DCN-R50 model surpassed those obtained by merging ResNet50 with the OR-CNN model proposed by Paplham et al. (2023).

In accordance with the findings presented in Tables 3 and 4, our proposed CORAL and DCN deformable convolution methods exhibit superior performance when applied to the ResNet50 network architecture in comparison to the combined OR-CNN method suggested by Paplham et al. (2023). Our

approach achieves a notable 1.5% reduction in the *MAE* value and a significant 3% reduction in the *RMSE* value. Furthermore, the DCN-R50 variant of our approach maintains lower parameter counts and computational complexity.

4. Conclusions

In this paper, we proposed an algorithm for age recognition in portraits, utilizing an improved version of the ResNet model and deformable convolution. First, we explained the differences between traditional classification methods and the CORAL rank-uniform ordered regression approach. Then, we also introduced deformable convolution (DCN) and explored the potential of combining CORAL rank-uniform ordered regression with DCN in portrait age recognition. Finally, we proposed the innovative DCN-R34 model architecture.

The AFAD Asian face open dataset was used for training and testing purposes. The results demonstrate that the combined model architecture can enhance the accuracy of the model and reduce the error rate, as evidenced by the decrease in Mean Absolute Error (*MAE*) and Root Mean Squared Error (*RMSE*) values.

It was found that the DCN-R34 model outperformed the SOTA (State-of-the-art) model in terms of the same network structure of ResNet34, showing a decrease of 3.7% in *MAE* value and 3.6% in *RMSE* value. Similarly, when compared to the SOTA model based on ResNet50, the DCN-R50 model presented a reduction of 1.5% in *MAE* value and 3% in *RMSE* value. These findings confirm that the proposed approach, combining ResNet with CORAL and DCN deformable convolution, can achieve improved results in portrait recognition.

Use of AI tools declaration

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

Acknowledgments

The work was supported by Science and Technology Plan Project of Changzhou (CJ20210155, CJ20220151, CJ20220174), Natural Science Foundation of the Jiangsu Higher Education In stitutions of China (23KJA520001, 21KJD520002), and Jiangsu Province Vocational College Teacher Professional Leader High and Training Project under Grant (2023TDFX003).

Conflict of interest

The authors declare there is no conflict of interest.

References

1. Z. Huang, J. Zhang, H. Shan, When age-invariant face recognition meets face age synthesis: A multi-task learning framework and a new benchmark, *IEEE Trans. Pattern Anal. Mach. Intell.*, **45** (2023), 7917–7932. https://doi.org/10.1109/TPAMI.2022.3217882

2. A. M. Abu Nada, E. Alajrami, A. A. Al-Saqqa, S. Abu-Naser, Age and gender prediction and validation through single user images using CNN, *Int. J. Acad. Eng. Res.*, **4** (2020), 21–24.
3. I. Rafique, A. Hamid, S. Naseer, M. Asad, M. Awais, T. Yasir, Age and gender prediction using deep convolutional neural networks, in *2019 International Conference on Innovative Computing (ICIC)*, 2019, 1–6. https://doi.org/10.1109/ICIC48496.2019.8966704.
4. A. Othmani, A. R. Taleb, H. Abdelkawy, A. Hadid, Age estimation from faces using deep learning: A comparative analysis, *Comput. Vision Image Understanding*, **196** (2020). https://doi.org/10.1016/j.cviu.2020.102961
5. N. Sharma, R. Sharma, N. Jindal, Face-based age and gender estimation using improved convolutional neural network approach, *Wireless Pers. Commun.*, **124** (2022), 3035–3054. https://doi.org/10.1007/s11277-022-09501-8
6. A. Sakata, N. Takemura, Y. Yagi, Gait-based age estimation using multi-stage convolutional neural network, *IPSJ Trans. Comput. Vision Appl.*, **4** (2019), 1–10. https://doi.org/10.1186/s41074-019-0054-2
7. C. Y. Hsu, L. E. Lin, C. H. Lin, Age and gender recognition with random occluded data augmentation on facial images, *Multimedia Tools Appl.*, **80** (2021), 11631–11653. https://doi.org/10.1007/s11042-020-10141-y
8. B. B. Mamatkulovich, H. A. Alijon o'g'li, Facial image-based gender and age estimation, *Eurasian Sci. Her.*, **18** (2023), 47–50.
9. L. Li, H. T. Lin, Ordinal regression by extended binary classification, *Adv. Neural Inf. Process. Syst.*, **19** (2006).
10. W. Cao, V. Mirjalili, S. Raschka, Rank consistent ordinal regression for neural networks with application to age estimation, *Pattern Recognit. Lett.*, **140** (2020), 325–331. https://doi.org/10.1016/j.patrec.2020.11.008
11. J. Paplham, V. Franc, Unraveling the age estimation puzzle: Comparative analysis of deep learning approaches for facial age estimation, *arXiv preprint*, (2023), arXiv:2307.04570. https://doi.org/10.48550/arXiv.2307.04570
12. J. Dai, H. Qi, Y. Xiong, Y. Li, G. Zhang, H. Hu, et al., Deformable convolutional networks, in *Proceedings of the IEEE International Conference on Computer Vision*, (2017), 764–773.
13. Z. Niu, M. Zhou, L. Wang, X. Gao, G. Hua, Ordinal regression with multiple output CNN for age estimation, in *Proceedings of the IEEE conference on computer Vision and Pattern Recognition*, (2016), 4920–4928.

AIMS Press