A coarse-to-fine framework with ordinal regression for facial age estimation

Yongfeng Yan[a], Wenhao Li[b], Lulu Zhao[a], Wanyong Tian[a], Zheng Tang[a], Guobao Hui[a], Yin Ye[*d], Jianjun Li[#b,c]

[a]China Electronics Technology Group Corporation (CETC), 20th Institute, Xi'an, Shaanxi, China; [b]School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou, Zhejiang, China; [c]School of Information Science and Technology, Hangzhou Normal University, Zhejiang, China; [d]CEC Huada Electronic Design Co., Ltd., Beijing, China

ABSTRACT

This study introduces a novel coarse-to-fine framework combined with ranking regression designed to capture the ordinal nature of age progression. Our approach initially categorizes ages into broader groups, utilizing the inherent order of age labels to refine age estimation hierarchically. A ranking regression model then meticulously fine-tunes the predictions, resulting in a more accurate age estimate. We present a multi-stage neural network architecture that first differentiates between broad age categories and then hones in on more specific age distinctions. Our evaluation of multiple benchmark datasets indicates a substantial reduction in prediction error over current leading models. The empirical findings highlight the effectiveness of our methodology in addressing the complex, non-linear patterns of facial aging. The proposed method propels the domain of age estimation forward and provides a versatile framework for other ordinal regression tasks.

Keywords: Facial age estimation, coarse-to-fine, ordinal regression

1. INTRODUCTION

Facial age estimation is vital in computer vision, with applications in advertising and healthcare monitoring, where accurately gauging age from facial features is transformative[1]. Yet, factors like genetics and lifestyle, plus image occlusions, complicate the task[2,3]. Age prediction benefits from the sequential nature of age labels, a detail multi-class classifiers miss, but metric regression methods use[4]. Facial aging varies by age group, causing non-stationary aging patterns that make developing accurate regression models difficult[5,6]. This perspective frames age estimation as an ordinal regression challenge[4,6-8]. For instance, Cao et al. perceived age estimation through the ranking lens, introducing a method anchored on Rank-SVM[9].

This ordinal regression paradigm can be deconstructed into a series of binary classification tasks[10,11]. As an illustration, a decomposition strategy suggests training a binary classifier for each rank k within the range $\{1,2, ..., K\text{-}1\}$, determining if a sample's rank surpasses k. Subsequently, a sample's rank is deduced from the collective outcomes of these K-1 classifiers. Yet, this strategy often grapples with inconsistencies across the binary classifiers when implemented via neural networks. The CORAL framework[12] was developed to address this, offering a method for ordinal regression through extended binary classification with theoretical assurances for classifier consistency[13]. CORAL can be easily integrated into existing convolutional neural network (CNN) architectures to handle ordinal regression tasks effectively.

In this paper, a coarse-to-fine learning strategy with ranking regression has been proposed to address the inherent ordinal nature of age labels. The proposed approach consists of global and local regressors. The global regressor is responsible for initial broad age bracketing, followed by multiple specialized local regressors, each dedicated to refining age predictions within its designated age range. The global network first categorizes a facial image into a general age bracket, after which the corresponding local network fine-tunes the age prediction, ensuring both efficiency and precision in the estimation process.

Our coarse-to-fine approach is inherently flexible, accommodating the varied aging patterns across different age groups.

[*]yeyin@hed.com.cn; [#]lijjcan@gmail.com

International Conference on Optics, Electronics, and Communication Engineering (OECE 2024), edited by Yang Yue, Proc. of SPIE Vol. 13395, 1339531 · © 2024 SPIE · 0277-786X · Published under a Creative Commons Attribution CC-BY 3.0 License · doi: 10.1117/12.3049387

This flexibility ensures that the model remains sensitive to subtle age-related changes within narrow ranges while also being able to generalize across more comprehensive age categories. The dual-layered structure, featuring a global regressor followed by targeted local regressors, balances swift age categorization and meticulous refinement. By incorporating ranking regression, we reinforce the model's capacity to recognize the relative order of age labels, which is essential for accurate age estimation.

The contributions of this paper are as follows.

(1) We present a dual-layered ordinal regression framework with a coarse-to-fine approach. The initial network layer conducts a broad age categorization and is then refined by a more specialized sub-network within a narrower scope.

(2) We provide empirical evidence of our methodology's effectiveness across various age estimation datasets, demonstrating its advanced performance in ordinal regression tasks.

2. RELATED WORK

2.1 Ordinal regression

In the domain of ordinal regression, the primary objective is to determine the rank of a particular item. A prevalent approach in the field is to transform ordinal regression into a series of binary classification tasks[14]. Li et al. innovatively employed a soft ordinal label for training their ordinal regression model and devised a method that integrated probabilistic embedding techniques within ordinal regression[11]. A notable contribution introduces a deep ordinal regression mechanism tailored for datasets of smaller sizes[15]. In subsequent work, they incorporated a multi-class classification loss into an ordinal regressor[11]. Fu et al. ventured into monocular depth estimation, leveraging the principles of ordinal regression[16].

2.2 Coarse-to-fine strategy

The coarse-to-fine strategy, inspired by human learning patterns, is an approach in machine learning where challenging tasks are incrementally tackled by first addressing broader, more straightforward concepts and then progressively refining them to more specific ones. In classification contexts, this involves leveraging an automatically constructed label hierarchy, allowing models to transition from general categories, like distinguishing between broad classes of animals and objects, to more nuanced classifications, such as differentiating specific entities like cats and dogs[17].

Adopting the coarse-to-fine strategy offers several benefits. It aligns with the natural human learning process, starting with foundational knowledge and building upon it incrementally[18]. This approach effectively addresses challenges arising from similarities in the output space, reducing classification errors between closely related classes[17]. Moreover, it provides flexibility in curriculum design, emphasizing the learning tasks rather than just the sequence of training data presentation[19]. This strategy facilitates more efficient and practical learning by breaking down complex tasks into simpler sub-tasks, especially when datasets lack examples for intermediate goals[20].

3. APPROACH

The proposed method is illustrated in Figure 1. The model operates in a dual-phase process: initial coarse categorization followed by fine-grained prediction. In the first phase, the model leverages deep convolutional networks to discern broad age categories, utilizing the ordinal nature of age data. The resultant features are then refined through a ranking regression approach, where a fine-tuning mechanism is applied to achieve precise age estimates. This hierarchical processing ensures the model effectively captures the complex patterns of facial aging. The final output is a refined age prediction, offering improved accuracy over conventional methods.

For any input image $x \in X$, we begin by cropping the human face to eliminate the background and then align the face. The aligned face image is fed into a deep convolutional neural network to extract features. These features are subsequently linked to a global regressor, producing an initial coarse estimation. This preliminary result is further refined by mapping it to specific local regressors. The final age estimation is obtained as a weighted combination of all the local regressors.

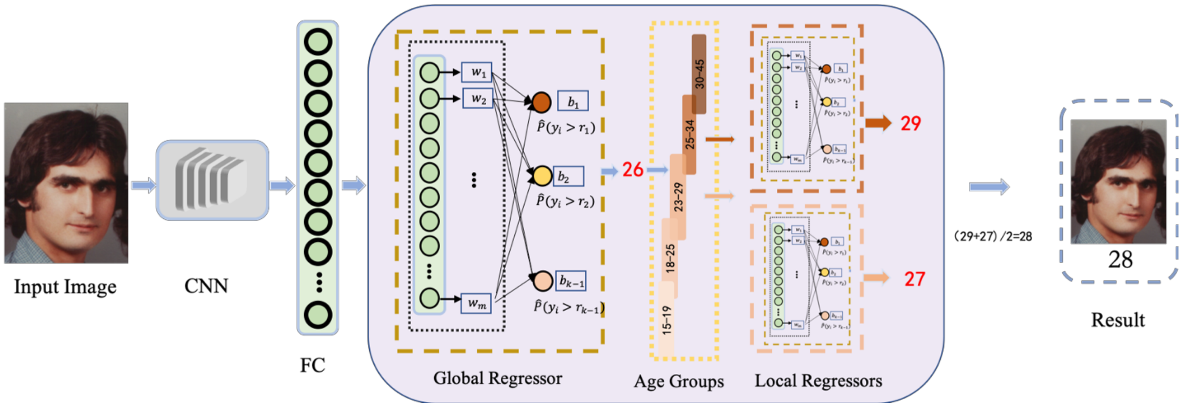

Figure 1. Architectural overview of the coarse-to-fine age estimation framework with ordinal regression. The model operates in a dual-phase process: initial coarse categorization followed by fine-grained prediction.

3.1 Age group clustering

Facial age estimation is a complex task due to the inherent nonlinearity of the facial aging process. This nonlinearity arises because different age groups exhibit distinct aging characteristics, which are influenced by a combination of genetic, environmental, and lifestyle factors[2]. In youth, facial changes predominantly revolve around growth, with features maturing and becoming more defined. As one transitions into adulthood, the emphasis shifts to skin texture alterations, influenced by environmental factors and lifestyle choices. In the later years, skin elasticity diminishes, leading to pronounced wrinkles and changes in facial contour. These distinct phases, influenced by genetics, environment, and habits, underscore the complexity of accurately estimating age based solely on facial features[19].

Utilizing a singular regression network to predict across an expansive age spectrum may not yield optimal precision due to the diverse aging characteristics inherent to different age groups. By segmenting the age spectrum and deploying dedicated regression networks for each segment, we can harness the specificity of aging patterns within that bracket. This segmented approach enables each network to learn intricately and adapt to its designated age range's unique facial features and aging nuances, thereby enhancing prediction accuracy within that segment. The training would also be more accessible because only the patterns in the smaller range should be learned[16].

Hence, we divide the entire age range into multiple groups. We first embarked on feature extraction from the facial images to segment the comprehensive age dataset into distinct age groups. Leveraging convolutional neural networks ensures the extraction of salient and discriminative features that encapsulate the intricate aging patterns present in the dataset. We utilized the K-means algorithm to cluster the extracted features, and then we visualized the clustering results using t-SNE. Figure 2 shows the clustering result. The result revealed a natural partitioning into five overlapping age groups. The rationale behind choosing five intersecting age groups stems from the continuous and transitional nature of facial aging. By allowing these age groups to overlap, we accommodate individuals who may exhibit facial features characteristic of adjacent age brackets, ensuring a more nuanced and accurate age estimation. Therefore, the entire range [7, 45] in the AFAD dataset is classified into five groups [7, 24], [20, 29], [25, 34], [30, 39], and [35, 45].

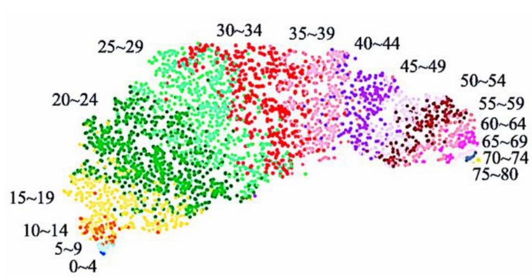

Figure 2. t-SNE visualization of the feature space in the AFAD and FG-NET datasets. Note that feature vectors are aligned roughly.

3.2 Ranking regression

$D = \{xi, yi\}_{i=1}^{N}$ represents the training dataset which consists of N training examples. Here, x_i denotes the $i\text{-}th$ training example and y_i the corresponding rank, where $y_i \in Y = \{r_1, r_2, \ldots, r_k\}$ with ordered rank $r_k > r_{k-1} > \cdots > r_1$. The ranking regression task can be achieved by solving a ranking function $f: X \rightarrow Y$ while minimizing a loss function $L(f)$.

Li and Lin introduced a comprehensive framework that transforms an ordinal regression challenge into multiple binary classification tasks[11]. This method necessitates a cost matrix exhibiting convexity across each row to derive a consistently ranked threshold model. However, given that the weightage associated with each binary task varies for individual training samples, this strategy is often deemed impractical due to its intensive computational demands, as highlighted by Niu et al.[4].

In the training dataset $D = \{xi, yi\}_{i=1}^{N}$, the rank y_i is first extended into K-1 binary labels: $y_i^1, y_i^2, \ldots, y_i^k$ such that $yi\ k \in \{0, 1\}$ indicates whether yi exceeds rank rk. Utilizing these extended binary labels during the training phase trains a singular CNN equipped with K-1 binary classifiers in its output layer. The predicted rank label for an input x_i is derived as $h(x_i) = rq$, where the rank index q is defined by $q = 1 + \sum_{k=1}^{K-1} f_k(x_i)$. Here, $f_k(x_i) \in \{0, 1\}$ represents the prediction from the k-th binary classifier in the output layer. Notably, while the K-1 binary classifiers share identical weight parameters, their bias units remain distinct. For model training, the loss function is

$$L(W, b) = -\sum_{i=1}^{N}\sum_{k=1}^{K-1} \lambda^{(k)}[\log(\delta(g(x_i, W) + b_k))y_i^k + \log(1 - g(x_i, W) + b_k))(1 - y_i^k)] \tag{1}$$

which is the weighted cross-entropy of $K-1$ binary classifiers. In equation (1), $\lambda^{(k)}$ refers to the weight of the loss associated with the k-th classifier (assuming $\lambda^{(k)} > 0$). In this paper, we indicate $\lambda^{(k)}$ as the most critical parameter for task k. For simplicity, all experiments will be implemented with uniform task weighting, that is, $\forall k$: $\lambda^{(k)} = 1$.

3.3 Global and local regressor

Based on the discussion in the age group clustering section, we proposed global and multiple local regressors. All these regressors employ the ranking regression concept as elucidated in Section 3.2. The global regressor is used in the entire age range, and is tasked with providing a coarse estimation. The input image will be mapped to one or two predefined age groups based on the coarse prediction age. Subsequently, the specific local regressor will undertake a more precise age estimation. We refer to this prediction process as the coarse-to-fine approach, where initial age estimation is refined through subsequent specialized prediction. The flow chart of our proposed coarse-to-fine approach is illustrated in Figure 3.

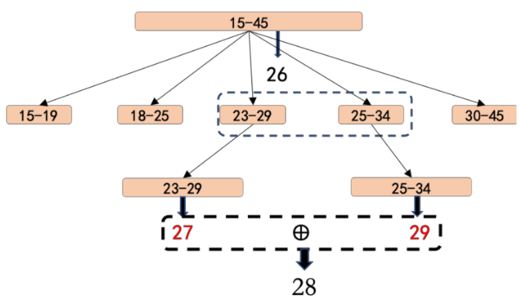

Figure 3. Age prediction process flows within the coarse-to-fine framework. This illustration details the sequential steps from the initial age range classification to the final refined age prediction.

Due to the overlap, the previous estimate may be involved in two rank groups in each iteration. In this case, both groups will be selected, and the estimated ages from the corresponding local regressors will be averaged. For instance, we consider a xi, if the global regressor predicts $G(x_i) = 23$, based on the group partitioning detailed in Section 3.1, the age 23 falls within the subsets D_2 and D_3. Consequently, the final prediction for xi should be jointly determined by the local regressor L_2 and L_3. If $L_2(x_i) = 22$ and $L_3(x_i) = 26$, the ultimate estimation for xi is given by $T(x_i) = \frac{L2(xi)+L3(xi)}{2} = \frac{22+26}{2} = 24$.

4. EXPERIMENT

4.1 Datasets

The MORPH II dataset[21], one of the most extensive open-access facial datasets, is frequently utilized for age estimation tasks. It comprises 55,608 facial images. These images underwent preprocessing, which involved aligning each face based on the average eye position determined through facial landmark detection. Subsequently, the images were adjusted to position the nose tip at the center. The age labels from 16 to 70 years old have been used in this study.

FG-NET[22] consists of 1002 color or greyscale face images of 82 individuals ranging in age from 0 to 69 and the CACD dataset[23] contains 159,449 face images in the age range of 14 to 62.

The Asian Face Database (AFAD)[24], used in this study, contained 165,501 faces between 15 and 40 years old. Since these images were already centered, no additional preprocessing steps were necessary.

4.2 Experimental settings

In alignment with the methodology presented by Niu[4], each image collection was split, allocating 80% for training and 20% for testing. All images were resized to 128×128×3 pixels at first and then randomly cropped to 120×120×3 pixels. The face images of 128×128×3 RGB were center-cropped to a model input size of 120×120×3 for model evaluation. To avoid overfitting, especially when the training data is insufficient, we also augment training images by horizontal flipping and random cropping.

ResNet-34[25] serves as the backbone network for the proposed method. We refer to the original ResNet-34 CNN with standard cross-entropy loss as CE-CNN. The network is initialized with weights pre-trained on the ImageNet dataset, then further pre-trained on the IMDB-WIKI dataset. We employ mini-batch stochastic gradient descent (SGD)[26] with Adam optimizer to optimize the network. The exponential decay rates are set to $\beta_0 = 0.90$ and $\beta_1 = 0.99$, with a batch size of 256. The initial learning rate is 0.0005, and λ_k is set to 1. The learning rate is adaptively adjusted during training. The algorithm is implemented using the PyTorch framework[27] and accelerated on a GeForce GTX 2080Ti GPU.

4.3 Evaluation metrics

To evaluate and compare model performance, we calculated two key metrics on the test set after the final training epoch:

1) Mean Absolute Error (MAE)

2) Root Mean Squared Error (RMSE)

These metrics provide a comprehensive assessment of the model's accuracy in age estimation.

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

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_i - h(x_i)\right)^2} \tag{3}$$

4.4 Results and analysis

Comparisons with the State-of-the-art: We performed a series of experiments on four independent face image datasets to evaluate and compare our age estimation method. All implementations utilized the ResNet-34 architecture, as detailed in Section 4.2. For benchmarking purposes, we included the standard ResNet-34 classification network with cross-entropy loss as a baseline for performance comparison.

Table 1 shows the performance of our coarse-to-fine with ranking regression compared to other SOTA methods on the MORPH II, FG-NET, and CACD datasets. Our approach consistently outperforms other methods in terms of MAE. This underscores the effectiveness of integrating the coarse-to-fine strategy with ranking regression, which captures the ordinal nature of age labels and adapts to diverse aging patterns.

Table 1. The comparisons between the proposed method and other state-of-the-art methods on MORPH II dataset, FG-NET dataset and CACD dataset.

Method	MORPH II	FG-NET	CACD
DRFs[28]	2.91	3.85	5.77
DEX[29]	2.68	3.09	5.68
OR-CNN[4]	2.83	-	5.38

Method	MORPH II	FG-NET	CACD
Ranking-CNN[7]	2.96	-	5.45
AGEn[5]	2.52	-	2.68
C3AE[30]	2.75	2.95	-
BridgeNet[31]	2.38	2.56	-
CORAL-CNN[9]	2.64	-	5.25
AVDL[32]	2.37	2.51	-
DRC-ORID[33]	2.26	2.48	-
MWR[34]	2.13	2.23	5.21
Ours	**2.11**	**2.21**	**5.16**

Table 2 shows the performance of our coarse-to-fine with ranking regression in comparison to other state-of-the-art methods in the AFAD datasets. The AFAD dataset, known for its demographic diversity and age variance, serves as an ideal benchmark for assessing age estimation models. Finally, we compared our proposed method against several state-of-the-art approaches in both Mean Absolute Error (MAE) and Root Mean Square Error (RMSE), two widely recognized metrics for evaluating the performance of age estimation systems.

Table 2. The comparisons between the proposed method and other state-of-the-art methods on AFAD dataset.

Method	MAE	RMSE
LDLF[35]	3.78	5.09
CORF[36]	3.66	4.83
OR-CNN[4]	3.23	4.48
CORAL-CNN[9]	3.30	4.52
DRFs[28]	3.48	4.85
coGOL[37]	3.20	4.45
Ours	**2.92**	**4.25**

Ablation Study: We embarked on a series of ablation experiments to elucidate the individual contributions of the various components within our proposed methodology. These experiments were designed to systematically remove or substitute key components, thereby revealing their respective impacts on the overall performance.

Initially, we assessed the influence of the coarse-to-fine strategy. A noticeable decrement in performance was observed by bypassing this strategy and directly employing the global regressor for predictions. This unequivocally underscores the pivotal role the coarse-to-fine strategy plays within the model, particularly in offering a structured learning trajectory when confronted with a broad age distribution.

Subsequently, the contribution of the ranking regression was scrutinized. Reverting to conventional regression techniques in lieu of the ranking regression led to a discernible dip in predictive accuracy. This further attests to the superiority of ranking regression in capturing the ordinal nature of age labels and the additional robustness it imparts to the model.

Table 3 shows the result of ablation study result with/without coarse-to-fine strategy and ordinal regression. Through these ablation studies, a conclusion was drawn: each component bestows unique value to our approach, collaboratively ensuring the model's exemplary performance in facial age estimation tasks. Notably, the coarse-to-fine strategy and ranking regression emerge as quintessential, ensuring not just accurate predictions but also adeptly capturing the ordinal relationships between age labels.

Table 3. Ablation study result with/without coarse-to-fine strategy and ranking regression.

Experiment variant	MAE	RMSE
Baseline model	2.92	4.25
Without coarse-to-fine strategy	3.30	4.48
Without ranking regression	3.34	4.57

Success and failure cases: In the experimental validation, we present two sets of results to demonstrate our model's predictive capabilities. Figure 4 exemplifies accurate age estimations, reflecting the model's strength in capturing diverse aging features. These successes are attributed to the effective hierarchical feature interpretation by our coarse-to-fine framework.

Figure 4. Successful predictions by the coarse-to-fine age estimation model. These examples highlight instances where the model accurately estimates ages across a range of demographics and conditions.

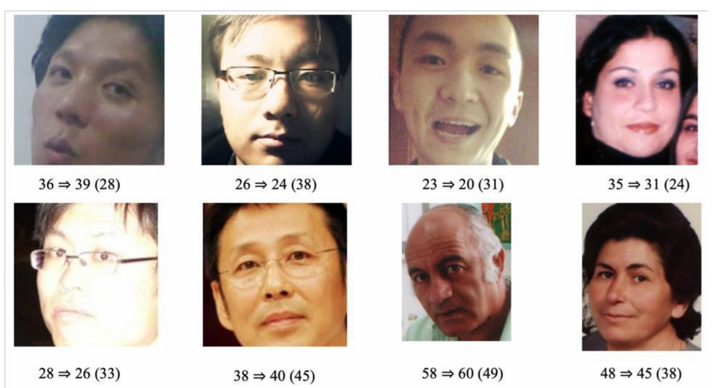

Figure 5. Challenges in age estimation with the coarse-to-fine model. This set illustrates scenarios where the model's predictions deviate from the actual ages, providing insights into conditions that affect prediction accuracy.

Contrastingly, Figure 5 displays cases with less accurate predictions, often due to challenging factors such as poor lighting, occlusions, or atypical facial features. Notably, performance dips were observed with extreme age groups where training data was limited. Furthermore, the model appeared to struggle with age estimations at the extremities of the age spectrum. The youthful and elderly faces displayed a higher discrepancy between predicted and actual ages. This is partly due to the sparsity of representative training samples in these age groups, leading to less robust feature learning.

5. CONCLUSION

In this paper, a coarse-to-fine learning strategy has been proposed by combining it with ranking regression for enhanced facial age estimation. This method, characterized by a global and subsequent local regressor, efficiently addresses the ordinal nature of age labels and diverse aging patterns. Empirical tests across various datasets confirmed its robustness and superiority in age estimation tasks. While primarily tailored for age estimation, the approach's principles show promise for other regression-based vision challenges. Future work will probe its adaptability in pose estimation and crowd-counting areas.

ACKNOWLEDGMENTS

This work was partly supported by the Key Research and Development Plan of Zhejiang: No.2021C03131; National Science Fund of China no.61871170; Opening Fund of Key Laboratory of Data Link Technology: CLDL-20202207.

REFERENCES

[1] Zhang, Z., Song, Y. and Qi, H., "Age progression/regression by conditional adversarial autoencoder," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 4352-4360 (2017).
[2] Rothe, R., Timofte, R. and Van Gool, L., "Deep expectation of real and apparent age from a single image without facial landmarks," International Journal of Computer Vision, 126(04), (2018).
[3] Neyshabur, B., Li, Z., Bhojanapalli, S., LeCun, Y. and Srebro, N., "Towards understanding the role of over-parametrization in generalization of neural networks," arXiv:1805.12076, (2018).
[4] Niu, Z., Zhou, M., Wang, L., Gao, X. and Hua, G., "Ordinal regression with multiple output CNN for age estimation," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 4920-4928 (2016).
[5] Ramanathan, N., Chellappa, R. and Biswas, S., "Age progression in human faces: a survey," Visual Languages and Computing, 15(01), (2009).
[6] Liu, X., Zou, Y., Kuang, H. and Ma, X., "Face image age estimation based on data augmentation and lightweight convolutional neural network," Symmetry, 12(1), 146 (2020).
[7] Yang, P., Zhong, L. and Metaxas, D., "Ranking model for facial age estimation", International Conference on Pattern Recognition, 3404-3407 (2010).
[8] Li, C., Liu, Q., Liu, J. and Lu, H., "Learning ordinal discriminative features for age estimation," IEEE Conference on Computer Vision and Pattern Recognition, 2570-2577 (2012).
[9] Cao, D., Lei, Z., Zhang, Z., Feng, J. and Li, S. Z., "Human age estimation using ranking SVM," Biometric Recognition, 324-331 (2012).
[10] Frank, E. and Hall, M., "A simple approach to ordinal classification," Proceedings of the 12th European Conference on Machine Learning, 145-156 (2001).
[11] Li, L. and Lin, H. T., "Ordinal regression by extended binary classification," 19, 865-872 (2006).
[12] Cao, W., Mirjalili, V. and Raschka, S., "Rank consistent ordinal regression for neural networks with application to age estimation," Pattern Recognition Letters, 140, 325-331 (2020).
[13] Chang, K. Y., Chen, C. S. and Hung, Y. P., "Ordinal hyperplanes ranker with cost sensitivities for age estimation," IEEE Conference on Computer Vision and Pattern Recognition, 585-592 (2011).
[14] Guo, G., Mu, G., Fu, Y. and Huang, T. S., "Human age estimation using bio-inspired features," IEEE Conference on Computer Vision and Pattern Recognition, 112-119 (2009).
[15] Liu, H., Lu, J., Feng, J. and Zhou, J., "Ordinal deep feature learning for facial age estimation," IEEE International Conference on Automatic Face & Gesture Recognition, 157-164 (2017).
[16] Fu, Y. and Huang, T. S., "Human age estimation with regression on discriminative aging manifold," IEEE Transactions on Multimedia, 10(4), 578-584 (2008).
[17] Krizhevsky, A., [Learning Multiple Layers of Features from Tiny Images], University of Toronto, (2012).
[18] Warrington, E. K., "The selective impairment of semantic memory," Quarterly Journal of Experimental Psychology, 27, 635-657 (1995).
[19] Geng, X., Zhou, Z. H. and Smith-Miles, K., "Automatic age estimation based on facial aging patterns," IEEE Transactions on Pattern Analysis and Machine Intelligence, 29(12), 2234-2240 (2007).
[20] Guo, G. and Mu, G., "Joint estimation of age, gender and ethnicity: CCA vs. PLS", 1-6 (2013).
[21] Ricanek, K. and Tesafaye, T., "MORPH: a longitudinal image database of normal adult age-progression," International Conference on Automatic Face and Gesture Recognition, 341-345 (2006).
[22] Chen, B. C., Chen, C. S. and Hsu, W. H., "Face recognition and retrieval using cross-age reference coding with

cross-age celebrity dataset," IEEE Transactions on Multimedia, 17(6), 804-815 (2015).
[23] https://bcsiriuschen.github.io/CARC/.
[24] https://github.com/John-niu-07/afad-dataset.github.io/commit/7207b12d6583b2bcd14eef3165fe63359ba88ed5.patch.
[25] He, K., Zhang, X., Ren, S. and Sun, J., "Deep residual learning for image recognition," IEEE Conference on Computer Vision and Pattern Recognition, 770-778 (2016).
[26] Kingma, D. and Ba, J., "Adam: a method for stochastic optimization," International Conference on Learning Representations, (2014).
[27] Paszke, A., et al., "PyTorch: an imperative style, high-performance deep learning library," Neural Information Processing Systems, (2019).
[28] Shen, W., Guo, Y., Wang, Y., Zhao, K., Wang, B. and Yuille, A., "Deep regression forests for age estimation," IEEE Conference on Computer Vision and Pattern Recognition, 2304-2313 (2018).
[29] Rothe, R., Timofte, R. and Van Gool, L., "DEX: Deep expectation of apparent age from a single image," IEEE International Conference on Computer Vision Workshop, 252-257 (2015).
[30] Zhang, C., Liu, S., Xu, X. and Zhu, C., "C3AE: exploring the limits of compact model for age estimation," IEEE Conference on Computer Vision and Pattern Recognition, 12579-12588 (2019).
[31] Li, W., Lu, J., Feng, J., Xu, C., Zhou, J. and Tian, Q., "BridgeNet: a continuity-aware probabilistic network for age estimation", 1145-1154 (2019).
[32] Wen, X., et al., "Adaptive variance based label distribution learning for facial age estimation," 379-395 (2020).
[33] Lee, S. H. and Kim, C. S., "Deep repulsive clustering of ordered data based on order-identity decomposition," International Conference on Learning Representations, (2021).
[34] Shin, N. H., Lee, S. H. and Kim, C. S., "Moving window regression: a novel approach to ordinal regression," IEEE Conference on Computer Vision and Pattern Recognition, 18739-18748 (2022).
[35] Wei, S., Zhao, K., Guo, Y. and Yuille, A. L., "Label distribution learning forests," Advances in Neural Information Processing Systems, 30 (2017).
[36] Zhu, H., Shan, H., Zhang, Y., et al., "Convolutional ordinal regression forest for image ordinal estimation," IEEE Transactions on Neural Networks and Learning Systems, 33(8), 4084-4095 (2021).
[37] Lu, F., Ferraro, F. and Raff, E., "Continuously generalized ordinal regression for linear and deep models," Proceedings of the 2022 SIAM International Conference on Data Mining (SDM), 28-36 (2022).