Enhancing Age Estimation from Handwriting: A Deep Learning Approach with Attention Mechanisms

Li Zhao, Xiaoping Wu*, Xiaoming Chen
Huzhou University, School of Information Engineering, Huzhou 313000, P. R. China

***Abstract*—Currently, age estimation is a hot research topic in the field of forensic biology. Age estimation methods based on facial or brain features are easily affected by external factors. In contrast, handwriting analysis is a more reliable method for age estimation. This paper aims to improve the accuracy and efficiency of age prediction using handwriting analysis by proposing a novel method that integrates a coordinate attention mechanism in a deep residual network (CA-ResNet). This method can more accurately capture important features in the input handwritten images while reducing the number of model parameters, thereby improving the accuracy (Acc) and efficiency of the model for age estimation. The proposed method is evaluated on standard handwriting datasets and the created dataset, and it is compared with the current state-of-the-art methods. The results show that the method consistently outperforms others, achieving an accuracy of 79.60% on the IAM handwriting dataset, with a 6.31% improvement over other methods.**

Keywords—Age estimation; coordinate attention mechanism; handwriting analysis; accuracy

I. Introduction

Handwritten images can be used not only for text and signature recognition but also for biometric identification based on demographic features such as age range, gender, handedness, ethnicity, etc. [1–4]. The correlation between these attributes and handwriting has been proposed in study [5], with various applications such as customer identity verification in banks, government institutions, and other financial organizations as well as limiting the investigation scope to a limited group of individuals in forensic science. In recent years, significant progress has been made in improving identity recognition accuracy by studying each attribute separately. For instance, researchers have classified iris images based on age and gender in [6–8]. However, age-related research in handwriting recognition still presents challenges. The reason is that the handwriting is a combination of all elements and qualities unique to the writer. Specific features present in an individual's handwriting form the basis of identification [9], indicating that the existence of critical features that are overlooked, in addition to those considered more important by people. Therefore, this study delves into the exploration and extraction of different features of handwriting. Before deep learning, handwriting recognition relied on the manual extraction of certain features of handwriting such as the length, spacing, direction, and thickness of strokes. These extracted features were then fed into specific classifiers for handwriting classification. However, this manual feature extraction method is time-consuming and tedious, and classifiers may not achieve high accuracy in classifying certain handwriting features. Traditional machine learning models may also be ineffective for large datasets and complex problems.

The morphology, style, and physiological characteristics of handwriting vary among different age groups, making the extraction of handwriting features more challenging. Traditional manual feature extraction methods waste time and effort and the extracted rules are biased and can only learn limited content. In contrast, deep learning neural networks can automatically learn features, and the learned features are more abstract and can extract patterns from a large amount of input.

Presently, deep neural networks have achieved good results in classifying the gender of writers. Nonetheless, research on applying deep learning to the age classification of writers is limited, and existing studies show that although the results of deep learning are better than those of traditional machine learning in terms of age classification, the accuracy is still relatively low. Specifically, opting for the Residual Network (ResNet) presents a viable approach [10]. ResNet offers an effective solution to the vanishing gradient problem by introducing skip connections to reduce the depth of the network. Even so, ResNet is not without its drawbacks, such as high computational costs and susceptibility to overfitting [11]. Furthermore, research indicates that utilizing the GoogleNet and ResNet architectures for automated feature extraction, coupled with SVM classification, results in improved performance for handwritten age classification, although the classification accuracy remains relatively low [11].

In order to overcome these challenges, the introduction of an attention mechanism into the network proves advantageous, enhancing the model's accuracy and performance. This aids in capturing local features more effectively, reducing overfitting, improving model generalization, and concurrently lowering computational costs. Attention mechanisms find wide applications in computer vision tasks and exhibit advantages in reducing classification error rates, particularly in fine-grained datasets [12]. The prevailing focus in existing author age prediction is on single binary classification tasks, wherein the Convolutional Block Attention Module (CBAM) [13] and Squeeze-and-Excitation (SE) [14] attention mechanism modules primarily address high-level features and are less effective at handling low-level features. Consequently, Hou et al. introduced a novel Mobile Network Attention Mechanism known as the Coordinate Attention (CA) mechanism [15]. CA mechanism excels in capturing inter-channel information, enhancing

*Corresponding Author

localization accuracy, and identifying target regions through the fusion of channel relationships and positional information. It boasts strong portability and can be seamlessly integrated into classical mobile networks, all with minimal impact on computational costs. In conclusion, an enhanced ResNet model, CA-ResNet, is proposed in this study, which seamlessly integrates the CA mechanism into the ResNet network. This approach effectively captures local features in images and integrates them with global features, thereby extracting vital handwritten features associated with age from handwritten images, significantly enhancing model accuracy, and making it highly suitable for multi-age classification tasks. Finally, this approach facilitates end-to-end prediction of the authors' ages.

The main contributions of this study are summarized as follows:

1) Propose a novel end-to-end method for recognizing the age of handwriting image writers. This method integrates the attention mechanism into deep neural networks, which is a new technique in the field of handwriting image research.

2) By combining the coordinate attention mechanism with the ResNet-50 [10] model, a new model is created for analyzing handwriting images to recognize the age of the writer.

3) Experimental results on the benchmark IAM Handwriting Database (https://fki.tic.heia-fr.ch/databases/iam-handwriting-database) [16] and the created dataset demonstrate the superior performance of integrating a coordinate attention mechanism in a deep residual network (CA-ResNet) model in terms of accuracy when sufficient training data are available.

The rest of this paper is organized as follows: Section II provides a brief introduction to the related work. Section III outlines the dataset employed for the experiments, the preprocessing techniques applied, and the proposed approach for age classification. Section IV, the experimental setup is described, while in Section V, the experimental results are presented and discussed. Finally, Section VI concludes this paper and proposes future directions.

II. Related Work

There are mainly three types of research methods for classifying handwriting based on its features: clustering analysis, machine learning, and neural networks.

The clustering analysis method for handwriting feature classification can cluster different people's handwriting based on features such as size, density, and direction of strokes. For example, Marzinotto et al. proposed a two-level clustering algorithm [17], where the first level is independent of the writer's word groups, and the second level uses a Bag of Prototype Words generated from each author's words. They used a dataset obtained from the Broca Hospital in Paris, which contained authors ranging in age from 60 to 85 years old. Through supervised learning, they ranked the age of authors of handwritten documents. The study found that there are three different handwriting patterns in terms of dynamics, pressure, and duration for adults over 65, while those over 80 have very similar but slower styles. Basavaraja et al. proposed a new unsupervised age estimation method that utilizes Hu's invariant moments, separation features, and K-means clustering for handwriting analysis to extract discontinuities in the data and estimate the age of the writer [18]. The dataset is divided into four categories, 11-12, 13-16, 17-20, and 22-24 years old, each with 100 images, using K-means clustering. The accuracy achieved on the public English dataset IAM [16] and the public Arabic dataset KHATT [19] was 66.25% and 64.44%, respectively.

In recent years, with the development and application of machine learning technology, handwriting feature classification methods based on machine learning have also rapidly developed. In these methods, feature extraction and analysis are critical steps, and support vector machine (SVM) is a commonly used classifier [20]. Bouadjenek et al. used oriented gradient histogram and gradient local binary pattern features for handwriting feature classification, and SVM was used to classify the extracted features, achieving predictions for writer gender, age range, and handedness [21]. Although the accuracy of this method on the KHATT dataset is only 55%, it achieved 70% accuracy on the IAM dataset. To address the problem of writer age range prediction, Zouaoui et al. [22] proposed a joint training method to predict age range from handwriting analysis. They proposed multiple feature-generated descriptors and used the SVM classifier for prediction. The best descriptor collaboration achieved an accuracy of 78.57% on the IAM dataset. To further improve the classification accuracy of handwritten image recognition, Siddiqi et al. [23] conducted gender classification research. They used local and global features such as slant, texture, curvature, and readability to enhance the features of handwritten text and used artificial neural network (ANN) and SVM classifiers for classification. The experimental results showed that the classification rates reached 68.75% and 73.02% when tested on the Qatar University Writer Identification (QUWI) dataset and a custom-developed Multiscript Handwritten Database (MSHD), respectively. Emran et al. [24] proposed a method for identifying attributes in handwritten documents through three steps: segmentation, feature extraction, and classification using k-nearest neighbor (KNN), SVM, and random forest classifier (RFC) classification algorithms to identify the age, gender, and handedness of the document's writer. This method can simultaneously identify all three attributes in handwritten documents and has high performance in classification accuracy. Al Maaded et al. proposed a novel method for age, gender, and nationality classification using geometric features [25], extracting features through the random forest and kernel discriminant analysis, achieving some results. Mirza et al. studied the influence of handwriting image's visual appearance on the author's gender and used Gabor filters to extract texture information for gender classification [26]. Najla et al. addressed the age detection problem by dividing age into two groups, young adult writers and mature adult writers [1]. They extracted main features including Irregularity in pen pressure (IPP), irregularity in slant (IS), irregularity in text line (TLI), and the percentage of white and black pixels (PWB), and used SVM and neural network (NN) for classification. In experiments on the public Arabic dataset KHATT, the proposed age detection system achieved accuracies of 65.2% and 67%, respectively.

The method of hand-written feature classification based on deep learning technology to establish multi-layer neural networks for feature classification has been widely used. Currently, deep learning technology is widely used in hand-written image gender classification, but there is relatively little research on using deep learning technology for hand-written image age classification. For example, Cha et al. [27] trained an ANN on a capital letter dataset to classify demographic subcategories such as gender, handedness, and age group and used reinforcement learning techniques such as bagging and boosting to extract features and classify them using forward neural networks. Irina Rabaev et al. [28] proposed a deep neural network model called Bilinear Convolutional Neural Network (B-CNN) for automatic age and gender classification of hand-written images. They divided the KHATT dataset into four age groups according to official classification and achieved an accuracy rate of 64.4%. In addition, Najla AL-Qawasmeh et al. [11] extracted hand-written features from a self-built Arabic dataset and used ResNet and GoogleNet to predict the author's age with accuracy rates of 69.7% and 61.1%, respectively [29]. However, simply converting the neural network to hand-written age recognition may lead to lower accuracy because of the complexity of hand-written features, and a single neural network may ignore some important features. In addition, using handwritten text for age estimation has certain limitations, but it is effective for multi-classification problems. Therefore, this study proposes a new model that combines deep learning models with attention mechanisms to extract more comprehensive and important handwritten features for classification across multiple age groups.

III. Methodology

To gain a comprehensive understanding of the age estimation method for handwritten image authors proposed in this study, this section commences with the presentation and introduction of the overall architecture of the age estimation model for handwritten images. Subsequently, a detailed description of the dataset and data preprocessing methods is provided. Finally, the rationale for the age classification model proposed in this study is elucidated, including an introduction to the ResNet and CA Algorithm models.

A. Overall Architecture

In Fig. 1, the processing workflow of the handwritten image age analysis network architecture is depicted. The network takes handwritten text line images as input and, through a series of network layers, ultimately generates the age group to which the author of the handwritten image belongs, using a classification function. Specifically, the network layers comprise three modules: (1) the Convolutional Block Module, which is used to extract low-level features from the input image; (2) the Residual Block Module, incorporating the CA mechanism to enhance the model's focus on important areas within the handwritten image while suppressing irrelevant features; and (3) the Classification Module, responsible for classifying the output features to obtain information about the age of the writer.

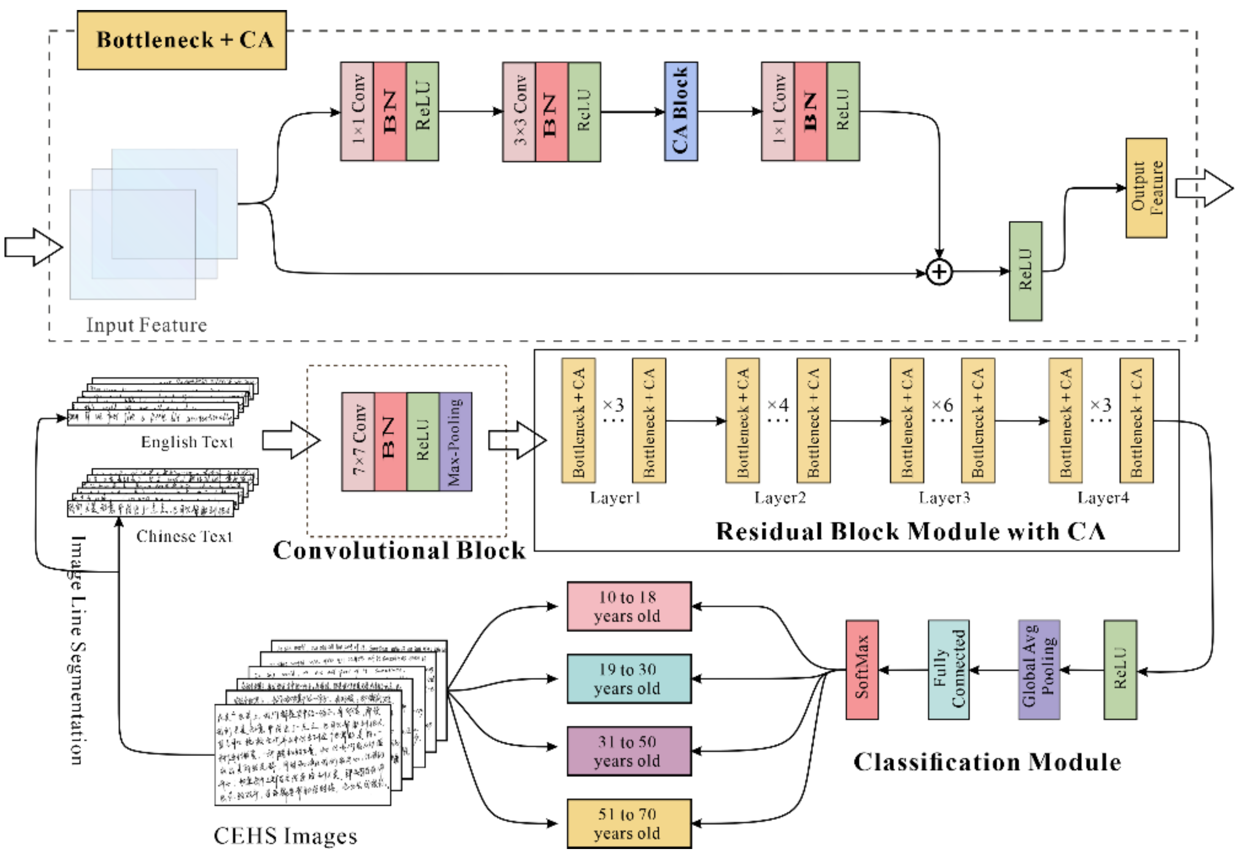


Fig. 1. The CA-ResNet architecture for age recognition of handwritten images.

Specifically, Fig. 1 proposes a CA-ResNet-based handwriting image recognition approach for predicting the age of the writer. The model first extracts feature from the input handwriting image through a 7×7 convolutional layer, followed by Batch Normalization (BN) and Rectified Linear Unit (ReLU) activation function processing [30]. Subsequently, the processed feature maps are fed into the Bottleneck module with a CA mechanism for further processing. Each Bottleneck module consists of a 1×1 convolutional layer, a 3×3 convolutional layer, a CA layer, and another 1×1 convolutional layer. The 1×1 convolutional layer reduces the number of feature map channels to decrease computation and improve the nonlinearity of the network. Then, the 3×3 convolutional layer extracts more detailed and local features of the image. Inserting the CA mechanism enables the network to focus on specific regions of the input image, which improves network performance. The principle of the CA mechanism algorithm will be elaborated in Section III (D). 2. Finally, the 1×1 convolutional layer adjusts the feature map again, including advanced feature extraction techniques such as feature fusion and channel adjustment. This layer is also the crucial level for outputting the feature map. The task of the whole Bottleneck module is to gradually extract low-level image features and transform them into high-level abstract feature representations, which facilitates classification and prediction of the age of the writer of the handwriting image in the classification module.

B. Dataset Creation

Based on the current inadequacy of publicly available handwriting datasets to meet the requirements of age classification tasks for both English and Chinese text, a dataset known as 'Chinese and English Handwriting Samples' (CEHS) has been established. In the dataset construction process, 100 volunteers were invited to transcribe text in both Chinese and English. Specifically, each volunteer transcribed text on two sheets of A4 paper, each bearing a unique identifier. Additionally, individual information about the writers, including age, gender, education level, and handedness, was recorded and matched with their respective unique identifiers. Finally, author information and the unique numbering of paper text were documented in an Excel spreadsheet for ease of subsequent data analysis.

In detail, the CEHS dataset comprises 200 handwritten page samples, with paper images scanned at a resolution of 300 dpi and stored in TIFF format. The age range of the authors falls between 10 and 70 years, with 40% being male and 60% female. Education levels span primary school, middle school, high school, college, and graduate school, and the majority of authors are right-handed. The CEHS dataset can be utilized in the field of handwriting recognition as well as demographic classification tasks, including age, gender, and handedness.

1) A dataset consisting of 1,531 handwritten text line images in both Chinese and English, with four age categories is created. This is the first paper to use a Chinese dataset to recognize the age of handwritten image writers, providing new data sources for handwriting image research.

C. Data Preprocessing

1) Text line segmentation: Compared with using the entire page image as input for deep neural networks, using text lines as input can not only recognize individual characters but also identify whole sentences or even entire articles, which significantly improves classification accuracy. However, owing to the special properties of handwritten documents, text line segmentation remains an important preprocessing stage and one of the most challenging problems in many optical character recognition systems. Therefore, this study adopted a projection-based algorithm to perform text line segmentation on the handwritten page images collected to tackle this problem [31]. The segmented text line sample images are shown in Fig. 2.

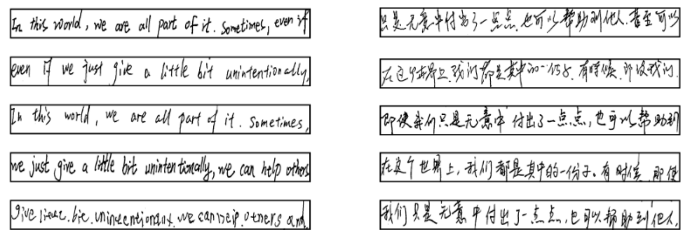


Fig. 2. English and Chinese text line samples.

2) Data augmentation: In classification tasks, the optimal performance of deep learning models is often influenced by the volume of data available. To address this challenge, data augmentation techniques have been introduced, encompassing mean blurring, Gaussian blurring, pooling operations, convolution operations, random cropping, and more [32]. Data augmentation not only enhances the diversity of the training data but also effectively reduces the likelihood of model overfitting [33].

D. Proposed Model

1) ResNet model: To address the problem of degradation in deep neural networks, He et al. proposed the deep residual module [10], which ensures that each newly added layer can easily incorporate the original function as one of its elements. Fig. 3 shows the typical structure of a residual block.

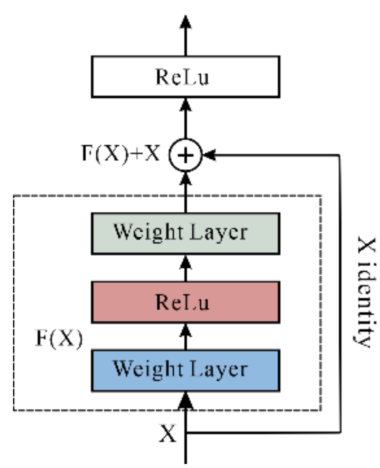


Fig. 3. Residual block structure diagram.

To avoid direct fitting of each stacked layer to the desired underlying mapping, a residual mapping is introduced such that the stacked nonlinear layers fit another mapping $\mathrm{F(X):H(X)-X}$, where $\mathrm{H(X)}$ denotes the desired underlying mapping. The original mapping is then represented by $\mathrm{F(X)-X}$. A feedforward neural network is used to implement shortcut connections for $\mathrm{F(X)-X}$ (see Fig. 3). Shortcut connections [34–36] skip one or more layers and add the output to the output of the stacked layers. In the ideal case where the identity mapping is optimal, it is easier to optimize the residual mapping close to zero than to fit an underlying mapping through stacked nonlinear layers. The entire network can be trained end-to-end via backpropagation and stochastic gradient descent.

2) CA algorithm: The CA mechanism can be viewed as a computing unit designed to enhance the feature representation capabilities of a mobile network, as illustrated in Fig. 4. It takes an intermediate feature tensor $\mathbf{X}=\{\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_C\}\in\mathbb{R}^{C\times H\times W}$ as input and transforms it into an output tensor $\mathbf{Y}=\{\mathbf{y}_1,\mathbf{y}_2,\ldots,\mathbf{y}_C\}\in\mathbb{R}^{C\times H\times W}$ of the same size as $\boldsymbol{X}$, but with enhanced representation capabilities [15].

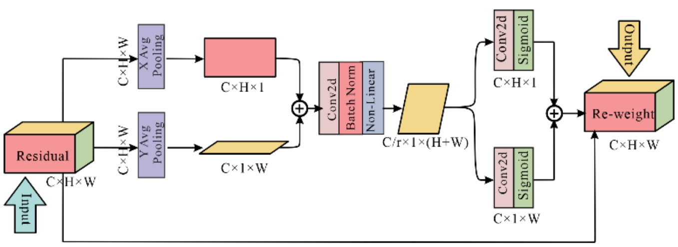


Fig. 4. CA block, where "X Avg Pooling" and "Y Avg Pooling" refer to 1D horizontal global pooling layer and 1D vertical global pooling layer, respectively.

To provide a clearer description of the CA mechanism, the Squeeze-and-Excitation (SE) attention mechanism [14] is first reviewed in this study.

Given an input $\mathbf{X}$, the SE block processes it in two steps: squeeze and excitation. The squeeze step is mainly used to embed global information, while the excitation step readjusts the relationship between channels through adaptive learning to enhance the feature representation capabilities. Specifically, the squeeze step of the c-th channel can be expressed mathematically as follows:

$$z_c=\frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}x_c(i,j) \tag{1}$$

Here, z_c is the output associated with the c-th channel. The input is a set of local descriptors from a convolutional layer with a fixed kernel size, which may only capture local information. To gather more extensive global information, the Squeeze operation endows the model with the ability to aggregate global information [15].

Specifically, to enhance the capability of the module, global pooling is decomposed into a pair of 1D feature encoding operations according to Eq. (1). Given the input $\mathbf{X}$, the horizontal and vertical coordinates are encoded on each channel by using pooling kernels with sizes of $(H,1)$ or $(W,1)$, respectively. The output of the c-th channel with height h, denoted by $z_c^h(h)$, is obtained as follows:

$$z_c^h(h)=\frac{1}{W}\sum_{0\le i<W}x_c(h,i) \tag{2}$$

Similarly, the output of the c-th channel with width w, denoted by $z_c^w(w)$ as follows:

$$z_c^w(w)=\frac{1}{H}\sum_{0\le j<w}x_c(j,w) \tag{3}$$

Next, the transformations in the embedding of information are cascaded and then processed using a convolutional transformation function:

$$\mathbf{f}=\delta(F_1([\mathbf{z}^h,\mathbf{z}^w])) \tag{4}$$

Here, $[.,.]$ denotes the cascading operation along the spatial dimension, δ is a non-linear activation function, and $\mathbf{f}\in\mathbb{R}^{\frac{C}{r}\times(H\times W)}$ is the intermediate feature map encoding spatial information in the horizontal and vertical directions. Here, r is a reduction ratio used to control the block size, similar to the design of SE blocks. After, $\mathbf{f}$ is split along the spatial dimension into two independent tensors, $\mathbf{f}^h\in\mathbb{R}^{\frac{C}{r}\times H}$ and $\mathbf{f}^w\in\mathbb{R}^{\frac{C}{r}\times W}$. Two 1×1 convolutional transformations F_h and F_w are then applied to $\mathbf{f}^h$ and $\mathbf{f}^w$, respectively, to transform them into tensors with the same number of channels. The results are shown below:

$$\mathbf{g}^h=\sigma(F_h(\mathbf{f}^h)) \tag{5}$$

$$\mathbf{g}^w=\sigma(F_h(\mathbf{f}^w)) \tag{6}$$

Here, σ is the sigmoid activation function. To reduce the computational cost and complexity of the model, an appropriate reduction ratio r is usually used to reduce the number of channels in $\mathbf{f}$. For example, r can take the value of 32. Finally, the outputs $\mathbf{g}^h$ and $\mathbf{g}^w$ are expanded and used as attention weights respectively, forming the output $\mathbf{Y}$ of the CA Block, as shown in the following equation:

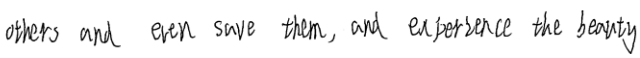
$$y_c(i,j) = x_c(i,j) \times g_c^h(i) \times g_c^w(j) \tag{7}$$

IV. Experiment

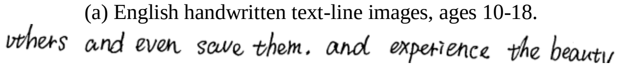
In this section, the main components include preparatory work for the experiments, such as the dataset and model parameters, as well as the visualization and analysis of the experimental results.

A. Datasets

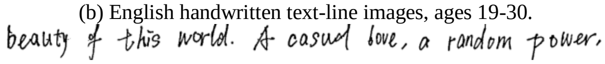
During the experiment, the CEHS dataset created for the study was utilized. The specific details of this dataset can be found in Section III (B). In a previous study [37], the dataset was divided into two age groups, youth, and adulthood, and it was shown through multiple feature experiments that there were differences in handwriting characteristics between these two age groups. For example, adults tend to follow more writing rules and write more neatly, while youth tend to write slower. Thus, to predict the age of the writer reasonably and minimize the difference in sample size for each age group, the text line images of the Chinese and English dataset were further divided into four age categories based on the classification in [37]: the first category representing teenagers aged 10-18, the second category representing young adults aged 19-30, the third category representing middle-aged people aged 31-50, and the fourth category representing elderly people aged 51-70. Fig. 5 and Fig. 6 respectively display handwritten images of English and Chinese text lines from different age groups. Data augmentation techniques were employed to expand the dataset and enhance recognition accuracy. The same experimental setup was applied to both Chinese and English datasets, with samples randomly divided into training sets (70%) and testing sets (30%). Such division is common in data mining practice [38]. Through this experimental setup, the age range could be predicted reasonably, and the sample size for each age category balanced as much as possible.



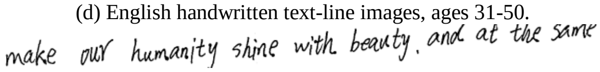
(a) English handwritten text-line images, ages 10-18.



(b) English handwritten text-line images, ages 19-30.

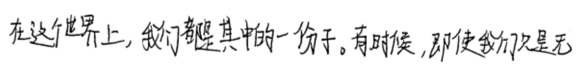


(d) English handwritten text-line images, ages 31-50.

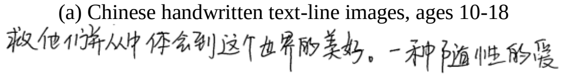


(e) English handwritten text-line images, ages 51-70.

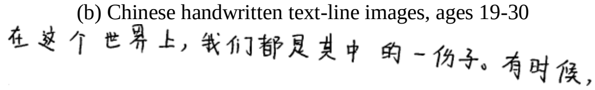
Fig. 5. English handwritten text-line images across different age groups.



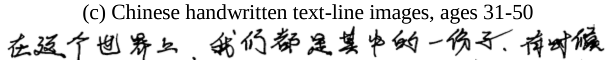
(a) Chinese handwritten text-line images, ages 10-18



(b) Chinese handwritten text-line images, ages 19-30



(c) Chinese handwritten text-line images, ages 31-50



(d) Chinese handwritten text-line images, ages 51-70

Fig. 6. Chinese handwritten text-line images across different age groups.

To validate the objectivity of the proposed CA-ResNet method and compare it with other age detection methods for handwriting, evaluation was conducted on the public IAM dataset. This dataset was collected by the Pattern Recognition and Artificial Intelligence Research Group at the University of Bern and is mainly used for training and testing handwriting recognition systems, and conducting writer identification and verification experiments. The dataset contains unconstrained handwritten text collected using an electron beam system and stored in XML format. It consists of 221 contributors, with a total of 86,272-word instances from an 11,059-word dictionary, including more than 1,700 tables and 13,049 isolated and labeled online and offline text lines. The IAM dataset consists of English text lines written by people of different age groups ranging from 16 to 56 years old, divided into two categories: 25-34 years old and 35-56 years old. Sample images are shown in Fig. 7. This dataset has been used in many papers [18,21,39–41]. For comparison, a similar age prediction dataset as used in [40] was selected, and the relevant information for both the IAM Dataset and the dataset used was listed in Table I.

TABLE I. INFORMATION ABOUT THE DATASETS USED IN THE EXPERIMENT

Datasets	Number of classes	Language	Age group
Ours	4	English and Chinese	10-18years old
			19-30years old
			31-50years old
			51-70years old
IAM-2[40]	2	English	25-34years old
			35-56years old

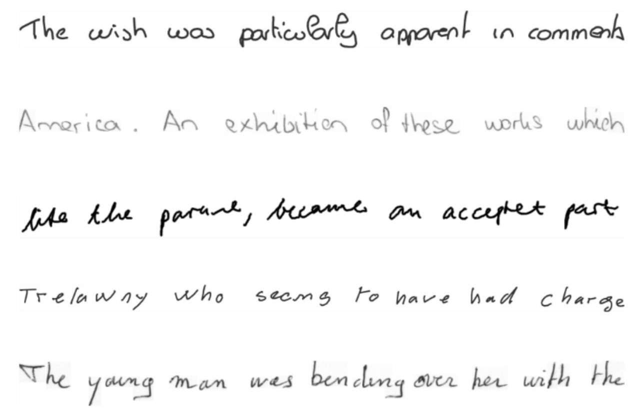


Fig. 7. IAM dataset sample image examples.

B. Setting Model Parameters

In this experiment, the weights learned in the bottom and middle layers were used for classification in the final FC layer, while softmax was used at the top layer to classify the handwritten images. The proposed network has end-to-end training capabilities and separates the images into different age groups for demographic classification. Cross-entropy loss was employed as the loss function, and SGD was used as the optimizer to train the model. The learning rate was set to 0.1, and a StepLR scheduler was implemented for learning rate optimization. The StepLR scheduler was set to adjust the learning rate at the end of the 30th, 60th, and 90th training epochs, by reducing the current learning rate by a certain proportion, which was applied during network training. Additionally, the best weights generated from network training were preserved using validation data to achieve optimal model performance.

To evaluate the performance of the proposed model in the age classification of handwritten images, True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) were computed using the confusion matrix shown in Table II. Performance metrics including precision, recall, and F1 score for each age group were calculated based on these parameters. The specific calculation formulas for Eq. (8) to Eq. (10) are presented. These performance metrics comprehensively and accurately reflect the classification performance of the proposed model in different age groups. By analyzing the performance differences among different age groups, the strengths and weaknesses of the model can be better understood, and further improvements and optimizations can be made accordingly to enhance its performance in practical applications.

$$\text{Precision}=\text{TP}/(\text{TP}+\text{FP}) \tag{8}$$

$$\text{Recall}=\text{TP}/(\text{TP}+\text{FN}) \tag{9}$$

$$\text{F1-Score}=2\times(\text{Precision}\times\text{Recall})/(\text{Precision}+\text{Recall}) \tag{10}$$

TABLE II. CONFUSION MATRIX

		Actual Class	
		Positive	*Negative*
Predicted Class	*Positive*	TP	FP
	Negative	FN	TN

VI, the performance of the proposed CA-ResNet method is compared with that of the baseline ResNet-50 and the existing method [28] in terms of precision, recall, and F1-score, which are three standard evaluation metrics for multiclass classification.

TABLE III. THE AGE CLASSIFICATION RESULTS OF BASELINE RESNET50, B-RESNET, AND THE PROPOSED CA-RESNET METHOD

Net	**Dataset**					
	IAM		*Our Dataset (Chinese)*		*Our Dataset (English)*	
	Loss	*Acc*	*Loss*	*Acc*	*Loss*	*Acc*
ResNet-50 ***[10]***	0.417	77.56%	0.371	83.04%	0.406	81.39%
B-ResNet ***[28]***	0.506	73.29%	0.480	77.81%	0.574	71.87%
CA-ResNet	0.388	79.60%	0.346	83.31%	0.374	82.42%

Table III presents the experimental results conducted on the public IAM dataset and our self-built CEHS dataset. In the IAM dataset, the CA-ResNet method achieves an accuracy rate of 79.60% in the age classification of handwritten image samples, with an improvement rate of 2.04% compared to the baseline ResNet-50 and 6.31% compared to the existing method [28]. In the CEHS dataset, the CA-ResNet method achieves accuracy rates of 83.31% and 82.42% on the Chinese and English

V. RESULTS AND ANALYSIS

In this section, the performance and distinctions of deep learning and machine learning in multi-age group classification tasks are showcased and analyzed.

A. Deep Learning Methods Performance

In terms of performance evaluation, in addition to using classification accuracy (Acc) (see Table III) and the confusion matrix (see Table IV, Table V and Table VI) as evaluation metrics, standard multiclass classification performance metrics (see Table VI), including precision, recall, and F1-score, were adopted to assess the performance of the proposed CA-ResNet method in the age recognition task for handwriting. Furthermore, to demonstrate the practicality of the method, the end-to-end deep learning method B-ResNet and baseline ResNet-50 were used as comparison experiments for age recognition of writers. Specifically, the B-ResNet method replaced the two VGG parallel blocks in the B-CNN [37] model with two identical ResNet models and combined them with the method of truncating the fully connected layer for output processing. Finally, bilinear feature representation was obtained through matrix outer product and average pooling operation, and the softmax function was used for age estimation.

In Table III, the age classification performance comparison results of the baseline ResNet-50, B-ResNet, and the proposed CA-ResNet models are presented, which are displayed in the form of bar charts as shown in Fig. 8, Table IV, Table V and Table VI respectively show the confusion matrices of the proposed method, baseline ResNet-50, and the existing deep neural network model [28] for age recognition of writers applied to the CEHS handwriting dataset and the publicly available English handwriting dataset IAM. In

datasets, respectively. Compared with the baseline ResNet-50 and the existing method [28], the application of the CA-ResNet method results in improvements of 0.27% and 6.21% in the Chinese dataset, and 1.03% and 10.55% in the English dataset, respectively. Our experimental results demonstrate that the CA-ResNet method has significant advantages in the age classification of handwritten images. Additionally, it was observed that the recognition effect is better when using the Chinese dataset compared to the English dataset. This difference may be attributed to the presence of some writers in the collected dataset who have lower education or no exposure to English, making it more difficult for them to write in English but easier to write in Chinese, thus better exhibiting their handwriting characteristics. Therefore, the choice of the dataset should be determined according to its characteristics in practical applications.

From Fig. 8, it can be intuitively seen that the CA-ResNet method achieved higher age recognition accuracy than the B-ResNet on both the handwritten dataset and the publicly available dataset (ACC at 78.17%). In addition, although the improvement of CA-ResNet over baseline ResNet-50 in terms of classification accuracy is not very significant even on the dataset created, it still has important practical significance in the application fields such as recognizing the handwriting author of handwritten images. For example, in medical image diagnosis, accurate age recognition can affect the diagnosis results and treatment plan selection; in face recognition technology, age recognition is also an important factor in determining individual identity. Moreover, it was found that in the field of insurance application review, age is a piece of core information that determines insurance company underwriting costs and risks. In summary, the CA-ResNet method has broad application prospects in these application field.

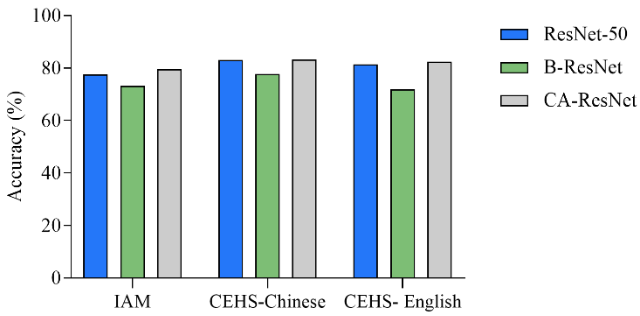


Fig. 8. Accuracy comparison of different methods on the IAM dataset and our created CEHS dataset.

Based on the data analysis in Table IV and Table V, it was found that regardless of whether the handwriting samples were in Chinese or English, the correct classification rates for the first three age groups (10-18 years old teenagers, 19-30 years old young adults, and 31-50 years old middle-aged people) were relatively high, indicating that the proposed method successfully captured the sample characteristics of these age groups and

VII demonstrate that the proposed CA-ResNet method exhibits superior performance in terms of accuracy, recall, and F1 score compared to the other two methods in the task of writer age recognition. Moreover, a comprehensive analysis was conducted by evaluating the ROC curves of each method using the public IAM dataset (see Fig. 9). Each color-coded curve represents the method's performance on two age group categories, namely 25-34 years and 35-56 years. By examining the ROC curves, it was observed that the CA-ResNet method consistently outperformed the other two methods. This observation further substantiates the accuracy, stability, and scalability of our CA-ResNet method for writer age recognition.

TABLE IV. CONFUSION MATRICES (%) OBTAINED BY APPLYING CA-RESNET, RESNET-50, AND THE EXISTING B-RESNET TO OUR CREATED CHINESE DATASET

Method	**Age group(years old)**				
CA-ResNet		*10-18*	*19-30*	*31-50*	*51-70*
	10-18	80.55	13.18	1.77	4.50
	19-30	3.80	93.19	1.27	1.74
	31-50	4.64	9.49	80.79	5.08
	51-70	6.76	11.70	4.39	77.15
ResNet-50 **[10]**		*10-18*	*19-30*	*31-50*	*51-70*
	10-18	81.67	9.97	1.77	6.59
	19-30	3.80	91.60	2.22	2.38
	31-50	4.42	10.60	77.48	7.51
	51-70	7.13	10.42	3.11	79.34
B-ResNet **[28]**		*10-18*	*19-30*	*31-50*	*51-70*
	10-18	78.30	11.74	2.57	7.40
	19-30	7.61	85.58	3.33	3.49
	31-50	9.49	9.93	70.42	10.15
	51-70	8.78	12.43	4.39	74.41

accurately identified their corresponding age categories. However, in the fourth age group (51-70 years old elderly people), the correct classification rates of the created Chinese and English datasets were 77.15% and 75.88%, respectively, which demonstrated a lower accuracy.

After comparing the results from Table IV, Table V, and Table VI, it was found that the existing B-ResNet method did not perform well on the CEHS dataset and IAM dataset. In contrast, the proposed CA-ResNet method performed better in terms of ACC (accuracy) and outperformed both the existing B-ResNet method and the baseline ResNet-50 method. Further analysis of the classification results of the created dataset and the IAM dataset shows that the CA-ResNet method and B-ResNet method performed better on the created dataset, while differences in performance may be due to differences in writing tools in the dataset. Additionally, it was observed that among the three methods, the performance of the B-ResNet method was the worst on both the created dataset and the IAM dataset. This may be because it is based on a traditional deep learning framework, resulting in high model complexity and affecting its performance. In comparison, the proposed CA-ResNet method reduces the frequency of hyperparameter tuning, reduces computational load, and allows the network to focus more on important information by incorporating attention mechanisms, thereby improving performance.

The results presented in

TABLE V. CONFUSION MATRICES (%) OBTAINED BY APPLYING CA-RESNET, RESNET-50, AND THE EXISTING B-RESNET TO OUR CREATED ENGLISH DATASET

Method	**Age group**				
CA-ResNet		*10-18*	*19-30*	*31-50*	*51-70*
	10-18	79.15	13.71	1.74	5.41
	19-30	3.68	92.14	2.34	1.84
	31-50	4.54	11.64	80.87	2.96
	51-70	7.25	13.73	3.14	75.88
ResNet-50 **[10]**		*10-18*	*19-30*	*31-50*	*51-70*
	10-18	74.32	16.02	2.70	6.95
	19-30	3.18	88.80	2.84	5.18
	31-50	2.37	10.06	84.81	2.76
	51-70	7.06	12.94	3.53	76.47
B-ResNet **[28]**		*10-18*	*19-30*	*31-50*	*51-70*
	10-18	66.41	18.34	3.09	12.16
	19-30	6.19	87.46	4.01	2.34
	31-50	9.27	11.64	70.61	8.48
	51-70	17.25	16.08	6.27	60.39

TABLE VI. CONFUSION MATRICES (%) OBTAINED BY APPLYING CA-RESNET, RESNET-50, AND THE EXISTING B-RESNET TO IAM DATASET

Method	**Age group (years old）**	
Our Proposed		*25-34*	*35-56*
	25-34	87.74	12.26
	35-56	29.18	70.82
ResNet-50 **[10]**	*25-34*	85.39	14.65
	35-56	30.85	69.15

B-ResNet **[28]**	*25-34*	82.35	17.65
	35-56	36.49	63.51

TABLE VII. COMPARING EVALUATION RESULTS FOR RESNET-50, B-RESNET, AND CA-RESNET USING PERFORMANCE METRICS

Database		**Age group(years old)**	**ResNet-50** [10]			**B-ResNet** [28]			**CA-ResNet**		
			Precision (%)	*Recall (%)*	*F1-Score(%)*	*Precision (%)*	*Recall (%)*	*F1-Score(%)*	*Precision (%)*	*Recall (%)*	*F1-Score(%)*
IAM		*25-34*	74.91	85.35	79.79	70.90	82.35	76.19	76.45	87.74	81.71
		35-56	81.39	69.15	74.77	76.92	63.51	69.58	84.26	70.82	76.96
Our Dataset	*CEHS-Chinese*	*10-18 19-30 31-50 51-70*	85.96	81.67	83.76	77.80	78.30	78.04	85.93	80.55	83.15
			77.58	91.60	84.01	74.38	85.58	79.59	75.68	93.19	83.52
			89.31	77.48	82.98	83.95	70.42	76.59	89.49	80.79	84.92
			82.82	79.34	81.05	78.12	74.41	76.22	87.19	77.15	81.86
	CEHS-English	*10-18 19-30 31-50 51-70*	85.18	74.32	79.38	66.67	66.41	66.54	83.33	79.15	81.19
			72.64	88.80	79.91	68.91	87.46	77.08	73.37	92.14	81.69
			89.77	84.81	87.22	83.26	70.61	76.41	91.31	80.87	85.77
			82.80	76.47	79.51	71.96	60.39	65.67	87.76	75.88	81.39

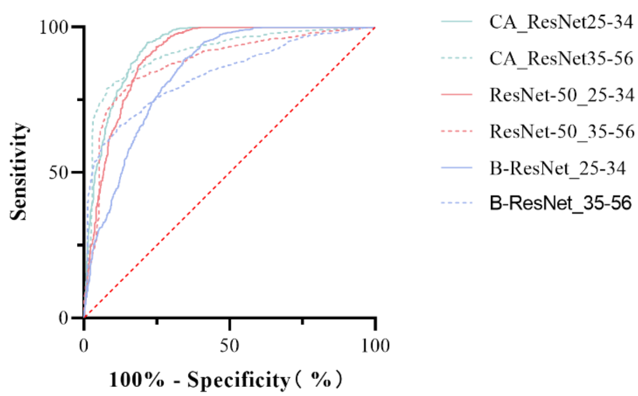


Fig. 9. ResNet-50, B-ResNet, and CA-ResNet ROC curves for age group classification on IAM.

B. Machine Learning Methods Performance

In addition, the performance of several traditional machine learning methods with handcrafted features was also compared. Table VII displays the age recognition performance on the IAM dataset using different features.

Among the several feature extraction methods listed in Table VII. The method that combines pixel density, pixel distribution, and gradient local binary pattern features using fuzzy MIN and MAX rules performed the best, with an accuracy of 78.57%. However, the proposed CA-ResNet method further improved the accuracy based on this method, reaching 79.60%, representing a certain degree of improvement compared to the previous feature combination method. In addition, the proposed CA-ResNet model can automatically learn and extract feature information from handwritten images, thereby greatly reducing the manpower cost required for manual feature extraction.

VI. CONCLUSION

This paper introduces a residual network model enhanced with attention mechanisms for the recognition of age from handwritten images. The proposed approach has been assessed on a diverse dataset comprising English and Chinese handwritten images, including the IAM dataset. Experimental outcomes highlight substantial enhancements in accuracy and efficiency compared to existing advanced methods. Notably, the model achieves an accuracy of 79.60% on the IAM handwriting dataset, marking a 6.31% improvement relative to other methods. Despite these advancements, the study is not without limitations. Future research endeavors are anticipated to broaden in three primary directions: Firstly, augmenting model generalizability through the expansion and diversification of datasets, aiming to encompass a wider spectrum of representative handwriting samples; secondly, investigating combined classifications of handwriting across multiple demographic traits, such as age, gender, and dominant hand; and thirdly, refining model architectures and attention mechanisms to enhance classification accuracy and efficiency, with a particular focus on the design and optimization of attention mechanisms to encapsulate key regions within images pertinent to age.

REFERENCES

[1] A.-Q. Najla, M. Khayyat, C.Y. Suen, Age detection from handwriting using different feature classification models, Pattern Recognition Letters 167 (2023) 60–66. https://doi.org/10.1016/j.patrec.2023.02.001.

[2] P. Maken, A. Gupta, A method for automatic classification of gender based on text-independent handwriting, Multimedia Tools and Applications 80 (2021) 24573–24602.

[3] M. Saraswat, A. Agarwal, Handwriting Recognition for Predicting Gender and Handedness Using Deep Learning, in: International Conference on Applied Machine Learning and Data Analytics, Springer, 2022: pp. 210–221. https://doi.org/10.1007/978-3-031-34222-6_18.

[4] A.P. Choudhury, P. Shivakumara, U. Pal, C.-L. Liu, EAU-Net: A New Edge-Attention Based U-Net for Nationality Identification, in: International Conference on Frontiers in Handwriting Recognition, Springer, 2022: pp. 137–152. https://doi.org/10.1007/978-3-031-21648-0_10.

[5] R.A. Huber, A.M. Headrick, Handwriting identification: facts and fundamentals, CRC press, United States, 1999.

[6] M.R. Rajput, G.S. Sable, Age Group Estimation from Human Iris, in: Soft Computing and Signal Processing: Proceedings of 2nd ICSCSP 2019 2, Springer, 2020: pp. 519–529. https://doi.org/10.1007/978-981-15-2475-2_48.

[7] M. Erbilek, M. Fairhurst, M.C.D.C. Abreu, Age prediction from iris biometrics, in: 5th International Conference on Imaging for Crime Detection and Prevention (ICDP 2013), IET, 2013: pp. 1–5. https://doi.org/10.1049/ic.2013.0258.

[8] M. Da Costa-Abreu, M. Fairhurst, M. Erbilek, Exploring gender prediction from iris biometrics, in: 2015 International Conference of the Biometrics Special Interest Group (BIOSIG), IEEE, 2015: pp. 1–11. https://doi.org/10.1109/BIOSIG.2015.7314602.

[9] S.L. Valiatia, J.A. Velho, A.T. Bruni, Investigation of individual characteristics in handwriting of a Brazilian Amazonian group, Brazilian Journal of Forensic Sciences, Medical Law and Bioethics 5 (2016) 146–170. http://dx.doi.org/10.17063/bjfs5(2)y2016146.

[10] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: pp. 770–778.

[11] A.-Q. Najla, C.Y. Suen, Transfer Learning to Detect Age From Handwriting, (2022).

[12] A.E. Eshratifar, D. Eigen, M. Gormish, M. Pedram, Coarse2Fine: a two-stage training method for fine-grained visual classification, Machine Vision and Applications 32 (2021) 49. https://doi.org/10.1007/s00138-021-01180-y.

[13] S. Woo, J. Park, J.-Y. Lee, I.S. Kweon, Cbam: Convolutional block attention module, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018: pp. 3–19.

[14] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: pp. 7132–7141.

[15] Q. Hou, D. Zhou, J. Feng, Coordinate attention for efficient mobile network design, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: pp. 13713–13722.

[16] M. Liwicki, H. Bunke, IAM-OnDB-an on-line English sentence database acquired from handwritten text on a whiteboard, in: Eighth International Conference on Document Analysis and Recognition (ICDAR'05), IEEE, 2005: pp. 956–961. https://doi.org/10.1109/ICDAR.2005.132.

[17] G. Marzinotto, J.C. Rosales, M.A. El-Yacoubi, S. Garcia-Salicetti, C. Kahindo, H. Kerhervé, V. Cristancho-Lacroix, A.-S. Rigaud, Age-related evolution patterns in online handwriting, Computational and Mathematical Methods in Medicine 2016 (2016). https://doi.org/10.1155/2016/3246595.

[18] V. Basavaraja, P. Shivakumara, D.S. Guru, U. Pal, T. Lu, M. Blumenstein, Age estimation using disconnectedness features in handwriting, in: 2019 International Conference on Document Analysis and Recognition (ICDAR), IEEE, 2019: pp. 1131–1136. https://doi.org/10.1109/ICDAR.2019.00183.

[19] S.A. Mahmoud, I. Ahmad, W.G. Al-Khatib, M. Alshayeb, M.T. Parvez, V. Märgner, G.A. Fink, KHATT: An open Arabic offline handwritten text database, Pattern Recognition 47 (2014) 1096–1112. https://doi.org/10.1016/j.patcog.2013.08.009.

[20] S.R. Sain, The nature of statistical learning theory, Springer science & business media, Germany, 1996.

[21] N. Bouadjenek, H. Nemmour, Y. Chibani, Age, gender and handedness prediction from handwriting using gradient features, in: 2015 13th International Conference on Document Analysis and Recognition (ICDAR), IEEE, 2015: pp. 1116–1120. https://doi.org/10.1109/ICDAR.2015.7333934.

[22] F. Zouaoui, N. Bouadjenek, H. Nemmour, Y. Chibani, Co-training approach for improving age range prediction from handwritten text, in: 2017 5th International Conference on Electrical Engineering-Boumerdes (ICEE-B), IEEE, 2017: pp. 1–5. https://doi.org/10.1109/ICEE-B.2017.8192233.

[23] I. Siddiqi, C. Djeddi, A. Raza, L. Souici-Meslati, Automatic analysis of handwriting for gender classification, Pattern Analysis and Applications 18 (2015) 887–899. https://doi.org/10.1007/s10044-014-0371-0.

[24] M. Al Emran, S.M. Naief, M.S. Hossain, Handwritten character recognition and prediction of age, gender and handedness using machine learning, Dissertation, University of BRAC, 2018.

[25] S. Al Maadeed, A. Hassaine, Automatic prediction of age, gender, and nationality in offline handwriting, EURASIP Journal on Image and Video Processing 2014 (2014) 1–10. https://doi.org/10.1186/1687-5281-2014-10.

[26] A. Mirza, M. Moetesum, I. Siddiqi, C. Djeddi, Gender classification from offline handwriting images using textural features, in: 2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR), IEEE, 2016: pp. 395–398. https://doi.org/10.1109/ICFHR.2016.0080.

[27] S.-H. Cha, S.N. Srihari, A priori algorithm for sub-category classification analysis of handwriting, in: Proceedings of Sixth International Conference on Document Analysis and Recognition, IEEE, 2001: pp. 1022–1025. https://doi.org/10.1109/ICDAR.2001.953940.

[28] I. Rabaev, I. Alkoran, O. Wattad, M. Litvak, Automatic gender and age classification from offline handwriting with bilinear ResNet, Sensors 22 (2022) 9650. https://doi.org/10.3390/s22249650.

[29] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: pp. 1–9.

[30] V. Nair, G.E. Hinton, Rectified linear units improve restricted boltzmann machines, in: Proceedings of the 27th International Conference on Machine Learning (ICML-10), 2010: pp. 807–814.

[31] M. Arivazhagan, H. Srinivasan, S.N. Srihari, A statistical approach to handwritten line segmentation, Document Recognition and Retrieval XIV, Proceedings of SPIE, San Jose, CA (2007) 6500T–1.

[32] M.A. Ponti, L.S.F. Ribeiro, T.S. Nazare, T. Bui, J. Collomosse, Everything you wanted to know about deep learning for computer vision but were afraid to ask, in: 2017 30th SIBGRAPI Conference on Graphics, Patterns and Images Tutorials (SIBGRAPI-T), IEEE, 2017: pp. 17–41. https://doi.org/10.1109/SIBGRAPI-T.2017.12.

[33] O. Stenroos, Object detection from images using convolutional neural networks, Dissertation, Universiy of Aalto, 2017.

[34] C.M. Bishop, Neural networks for pattern recognition, Oxford university press, United Kingdom, 1995.

[35] B.D. Ripley, Pattern recognition and neural networks, Cambridge university press, United Kingdom, 2007.

[36] W.N. Venables, B.D. Ripley, Modern applied statistics with S-PLUS, Springer Science & Business Media, Germany, 2013.

[37] T.-Y. Lin, A. RoyChowdhury, S. Maji, Bilinear CNN models for fine-grained visual recognition, in: Proceedings of the IEEE International Conference on Computer Vision, 2015: pp. 1449–1457.

[38] T.Y. Lin, Y. Xie, A. Wasilewska, C.-J. Liau, Data mining: foundations and practice, Springer, United States, 2008.

[39] Z. Huang, P. Shivakumara, M.A. Kaljahi, A. Kumar, U. Pal, T. Lu, M. Blumenstein, Writer age estimation through handwriting, Multimedia Tools and Applications 82 (2023) 16033–16055. https://doi.org/10.1007/s11042-022-13840-w.

[40] N. Bouadjenek, H. Nemmour, Y. Chibani, Robust soft-biometrics prediction from off-line handwriting analysis, Applied Soft Computing 46 (2016) 980–990. https://doi.org/10.1016/j.asoc.2015.10.021.

[41] N. Bouadjenek, H. Nemmour, Y. Chibani, Combination of topological and local shape features for writer's gender, handedness and age classification, in: Image Analysis and Recognition: 13th International Conference, ICIAR 2016, in Memory of Mohamed Kamel, Póvoa de Varzim, Portugal, July 13-15, 2016, Proceedings 13, Springer, 2016: pp. 549–557.