



A novel method of human identification based on dental impression image

Jiafa Mao ^{a,*}, Lixin Wang ^a, Ning Wang ^a, Yahong Hu ^a, Weigou Sheng ^b

^a The College of Computer Science and Technology, ZheJiang University of Technology, ZheJiang, Hang Zhou 310023, China

^b The Department of Computer Science, Hangzhou Normal University, Hangzhou 311121, China

ARTICLE INFO

Keywords:

Dental impression image
Tooth print detection
Multi-scale feature extraction
Feature aggregation
Human identification

ABSTRACT

In large-scale natural disasters and special criminal cases, surface features of bodies, such as faces and fingerprints, are easily destroyed. Teeth possess strong high-temperature resistance, corrosion resistance, and high hardness, which can compensate for the shortcomings of the aforementioned situations. This paper proposes an identification method based on the aggregated features of multi-scale dental impression images. Firstly, a method exploiting the adaptive object detection method based on YOLOv8 is proposed to segment toothprints. Next, a novel geometric feature named calibrated offset distance is extracted, combined with the SIFT feature method, to extract multi-scale and multi-dimensional features from the global toothprint, local toothprints, and single-tooth prints. Finally, all features are aggregated to enhance the descriptive ability and robustness. Experimental results indicate that the method proposed in this paper demonstrates good identification performance.

1. Introduction

At present, the widely used identification technologies based on image and vision focus on the face, fingerprint, iris, palm print, etc. However, in natural disasters such as earthquakes, floods, fires, and certain criminal cases, features like the human face can suffer severe damage or deformation, making identification challenging. Although DNA technology is accurate, it is expensive and time-consuming, thereby hindering the identification of numerous unidentified bodies during emergencies [1].

The enamel on the surface of teeth possesses high-temperature resistance, corrosion resistance, and high hardness [2]. In both natural and unnatural disasters, human teeth often retain their original inherent characteristics [3]. Furthermore, the teeth of different individuals exhibit evident differences in terms of occlusal morphology of the upper and lower teeth, shapes of the dental arch, interdental space, and crown size [4]. Additionally, variations can be observed in the shapes of dentition, concave and concave-convex surfaces of the crown [5], as well as in the presence of tooth defects and repairs, enabling the possibility of teeth identification [6]. The existing technology allows for quick and accurate identification of unknown corpses within a specific range. This is achieved by extracting dental impression features from individuals and conducting similarity analysis and comparison with a dental impression features database [7]. Such techniques provide new avenues for judicial institutions in conducting autopsies.

1.1. Related work

Human identification technology has been widely employed. Song et al. [8] proposed an end-to-end network named GaitNet for gait-based human identification. GaitNet consists of two convolutional neural networks: one for gait segmentation and the other for classification.

Zhang et al. [9] introduced VoxelTrack, a method for multi-person 3D pose estimation from a few cameras with wide baselines. It utilizes a multi-branch network to jointly estimate re-identification (Re-ID) for all individuals in the environment. Atta et al. [10] proposed a spatio-temporal gait recognition system to overcome the limitations of existing temporal template approaches like gait energy image.

Cantoni et al. [11] developed a gaze analysis technique called GANT, which employs a graph-based representation of fixation points obtained through eye tracking during human-computer interaction. Wang et al. [12] utilized a graph-based convolutional network and convolutional neural networks, respectively, to recognize gait from event streams. Siarohin et al. [13] addressed the problem of generating person images conditioned on both pose and appearance information, conducting an extensive evaluation using off-the-shelf Re-ID systems. Oh et al. [14] proposed person recognition in personal photo collections and conducted a thorough analysis on the importance of different features in terms of time and viewpoint generalizability.

Existing dental identification methods can be categorized into dental radiograph recognition and non-X-ray tooth image recognition, based on

* Corresponding author.

E-mail address: maojiafa@zjut.edu.cn (J. Mao).

the type of input data they utilize.

In recent years, a relatively comprehensive automatic dental identification system (ADIS) [15] has been proposed, which encompasses tooth segmentation [16], tooth classification [17], tooth numbering, and identification [18]. The system selects high curvature points on the tooth contour as features and compares geometric distances to search for the best matching object and obtain identification results. Nomir et al. [19] suggest that dental works (DWs) such as crowns, bridges, and dental fillings are typically brighter than teeth in X-rays and have less noisy contours. Thus, they offer another significant feature for matching. Chen and Jain [20] utilize area-based indicators for DW matching, along with contour similarity measures to improve accuracy. Hofer and Marana [21] also propose a method based on DW information, including the position, size, and distance between adjacent DWs.

Nomir and Abdel-Mottaleb [22] proposed a feature vector that incorporates prominent points of high-curvature tooth contour points. However, they discovered that the feature was not significant enough for poor-quality images. As a solution, they proposed another method based on prominent points of tooth appearance to overcome this issue. Lin et al. [23] propose an enhanced method that leverages both tooth and dental works (DWs) contours. They suggest trimming the outliers of each tooth contour after the initial calibration and recalibrating it. Additionally, they emphasize the use of additional invariant frequency features when matching with the spatial features of DWs. Rajput et al. [24] propose a method based on the overall skeleton of teeth and employ the level set method to track every contour of teeth. Banday et al. [25] introduce a novel mandibular biometric system that utilizes X-rays to establish an augmented reality (AR) model of personal teeth and extracts information from panoramic images.

In recent years, several biometric models based on deep learning have been proposed, leading to significant improvements compared to traditional biometric methods. In the context of human teeth, Cui et al. [26] were the first to introduce a tooth segmentation and recognition method based on deep learning. They utilized oral cone-beam computed tomography (CBCT) images to learn edge maps, similarity matrices, and spatial relationships between teeth, resulting in accurate outcomes. Lai et al. [27] proposed a method based on a deep convolutional neural network for automatic and precise matching of two-dimensional panoramic dental X-rays. They redesigned the network structure and incorporated an enhanced channel attention module and a learnable connection module to extract superior features.

Although dental impressions may not contain as rich information as X-rays, the imaging equipment for these images is inexpensive, and obtaining a large quantity of images is relatively easy. Dental images can significantly narrow down the scope of potential identities for a deceased individual, thereby facilitating forensic identification. Pushparaj et al. [28] propose using frontal color images of teeth as input. Compared to X-rays, this type of data contains identity information, unique extractable features, and can be conveniently obtained. The results demonstrate that using only the upper jaw, the identification accuracy can reach 70%. After grayscale processing, global segmentation, and connected region detection, Rehman et al. [29] obtain a binary image that solely contains teeth. This approach highlights the structural features of teeth, and the high-intensity regions of each tooth are extracted as matching features. Kumar et al. [30] perform tooth region segmentation and binarization, trim the image matrix, and assess overall tooth similarity to achieve favorable results. Miranda et al. [31] achieve victim identification using six smile selfies taken from different angles in cases where the facial features of the deceased individual were damaged, and no tooth images prior to death were available. They compare the smile arch of the deceased individual's teeth with the smile curve of the exposed teeth in the selfies, drawing the contours of the maxillary anterior teeth on both sets of images and evaluating the degree of overlap.

The network model has found extensive application in the field of pattern recognition. Xin et al. [32] propose a policy iterative algorithm

that can solve coupled algebraic Riccati equations corresponding to multiplayer non-zero-sum games. The effectiveness and feasibility of the design method are demonstrated through a simulation example involving three players.

Xu et al. [33] investigate the exponential stability problem for impulsive systems with double state-dependent delays. They present Lyapunov-based sufficient conditions for the exponential stability of the considered system. Zhang et al. [34] address the fault detection filter design problem for a class of nonhomogeneous higher-level Markov jump systems with uncertain transition probabilities. Stojanovic V. and Nedic, N. [35] propose a joint state and parameter robust estimation approach for stochastic nonlinear systems. The behavior of this new approach to joint estimation of states and unknown parameters of nonlinear systems is illustrated through extensive simulations, particularly in scenarios where measurements have non-Gaussian distributions.

In recent years, several new technologies have emerged in the field of pattern recognition. Wu et al. [36] introduce a pseudo-pair based self-similarity learning approach for unsupervised person re-identification (re-ID) without relying on human annotations. They construct patch surrogate classes as initial supervision and propose a method to assign pseudo labels to images through pairwise gradient-guided similarity separation.

Yang et al. [37] propose QuadNet, a method designed to address the multi-view problem in baggage ReID at three levels. QuadNet extracts view-aware attentional local features from discriminative regions in each view. These local features are then fused with global features to obtain enhanced representations of the quadruplets, leading to improved performance in baggage ReID. Shi et al. [38] propose an Attribute Mining and Reasoning (AMR) framework to address the issues at hand. The AMR framework leverages appearance attributes and consists of two main components. Extensive experiments are conducted to demonstrate the effectiveness of the proposed AMR framework, showcasing its superiority over existing state-of-the-art methods. Additionally, the AMR model exhibits excellent generalization ability.

Han et al. [39] propose the Adaptive Person Super-Resolution (APSR) model. APSR utilizes a joint training and fusion approach for multiple super-resolution (SR) modules based on their generated visual contents. This enables the model to fully compensate for and learn complementary identity features in an end-to-end manner. To enhance robustness against artifacts during fusion, the APSR model also learns informative features by online dividing and integrating the generated body regions. Patruno et al. [40] address the problem of people re-identification by utilizing soft biometric features. Their method operates on RGB-D data to identify the best matching individual from a database of potential users. Experimental results demonstrate an improvement in performance compared to the current state-of-the-art approaches.

1.2. Shortcomings of existing research, our contributions, and the limitations in real applications

The tooth data utilized in many research studies have certain limitations, i.e.,

- (1) Data acquisition is challenging. X-rays involve radiation, making it impossible to collect multiple samples from the one individual. Furthermore, the majority of data obtained from hospitals consists of images of diseased teeth, which may not accurately represent the individuals' overall dental health status.
- (2) Limited feature information. As depicted in Fig. 1, dental radiographs and dental images solely display the longitudinal surface of teeth, providing information about the side view of teeth while lacking details about the internal tooth arrangement. Furthermore, most existing non-X-ray data comprises color images of the front teeth, which also predominantly capture side information only.



Fig. 1. Comparison of the dental X-ray, tooth image, and dental impressions image.

To overcome these limitations, this paper proposes the utilization of dental impression images, as illustrated in Fig. 1. Dental impression images offer several advantages: they can be easily and reliably obtained, and they provide information about the cross-sectional surface of teeth as well as structural features such as the dental arch. Therefore, leveraging dental impression images is of great significance and holds promising potential in advancing dental identification technology.

This paper proposes a multi-scale and multi-dimensional teeth structure feature extraction and recognition method based on dental impression images. The main contributions of this article can be summarized as follows:

- (1) Propose an adaptive threshold strategy for the dental impressions detection model based on the YOLOv8 network. This strategy involves comparing the number of dental impressions detected by the model with the number of dental impressions manually input. The threshold is dynamically adjusted, either increased or decreased, until the desired number of dental impressions is detected. This adaptive thresholding approach ensures accurate and precise detection of dental impressions in accordance with the specified requirements.
- (2) Propose a geometric feature called the calibrated offset distance to describe the dental arch. The calibrated offset distance feature incorporates distance information of the arch polyline point while preserving rotation invariance in dental impression images. Additionally, this feature considers the overall integrity of the dental shape structure and captures whether the tooth is inward concave or outward convex, thereby providing comprehensive information about the dental arch.
- (3) Introduce a multi-scale and multi-dimensional dental impressions aggregation feature. The dental impressions are divided into three scales: global tooth print, local tooth print, and single tooth print. The feature extraction process is conducted separately for each scale, which enhances the diversity of extracted features. To enrich the feature hierarchy and improve robustness, high-dimensional SIFT features are incorporated at both global and local scales. Compared to the SIFT similarity feature, our proposed method exhibits improved dental identification performance.

In specific circumstances, such as natural disasters (earthquakes, floods, fires, etc.) and some crime sites, biological features have been highly damaged or deformed, and it is difficult to complete the identification work using face, fingerprint, iris, palm print and other technologies. Although the technology based on human DNA identification has high accuracy, it is expensive and time-consuming, and it is difficult to apply to many non-living body identifications in emergencies.

The limitations in practical applications are as follows:

Although our proposed method can be applied to the recognition of large-scale deformed anonymous corpses, in real applications, due to the lack of large-scale dental impression databases in the world. Therefore,

it is difficult to implement our technology. Many researchers are needed to conduct extensive research on dental impression, expand the influence of this technology, attract global attention, and make government departments around the world pay attention to this technology, establishing large-scale dental imprint databases. If a large-scale dental impression database is established, then this technology can better serve human society.

2. Dental impression detection model based on YOLOv8

Segmentation plays a crucial role in dental impression images, as its accuracy directly affects the identification results. Due to the complex textures and fuzzy contours present in dental impressions, conventional segmentation methods often fail to yield satisfactory outcomes. Drawing inspiration from object detection algorithms, this paper treats dental impressions as a distinct category, locates their boundaries and coordinates, and divides the regions of interest (ROI) to accomplish dental impression segmentation. This approach aims to improve the accuracy of segmentation for dental impressions.

The YOLOv8 network [41,42] is employed in this paper. Each dental impression image consists of 10 to 14 tooth print objects. Fig. 2 displays the labeled images. Some tooth prints at the end are missing due to the occlusal position being located at the back or the loose arrangement of outer teeth. Consequently, only a portion of the teeth inside the oral cavity is captured. These missing prints fall outside the required scope, as demonstrated in Fig. 2, such as the tooth print on the right side of the second row.

In this paper, the dental impression models from 50 individuals are selected, with each model contributing six images taken from different perspectives. This results in a total of 300 images, with approximately 12 tooth prints per image. Therefore, there are roughly 3600 individual tooth print samples in total. After training, a preliminary detection

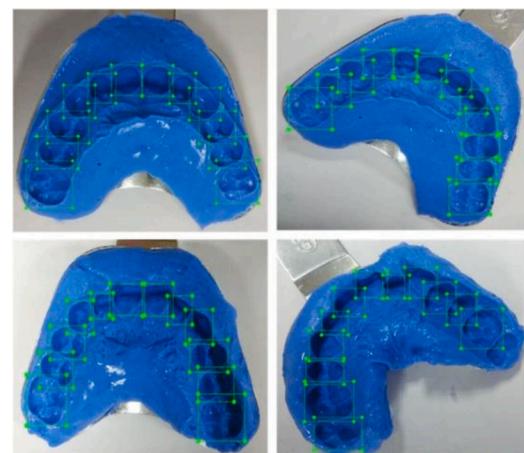


Fig. 2. Labeled sample data.

model is obtained, and the test result is shown in Fig. 3(a). In this result, only 7 out of 12 tooth prints in a dental impression image are successfully detected.

However, to improve the detection performance, a smaller sample set comprising 180 images from 30 individuals is selected, and the same training process is performed. The test result of this refined model is displayed in Fig. 3(b), showing a significant improvement compared to the previous result.

To explain this improvement, we hypothesize that the tooth print features contained in different dental impression images exhibit variability. When too many samples are added with random initial weights, it may become difficult for the model to discern the optimal tooth print feature patterns. As a result, the model may converge to a local optimum, leading to incomplete detection.

During the detection process of the No. 0301 image, it was observed that two tooth prints in the lower right corner were incorrectly detected as one, and repeated attempts to separate them were unsuccessful, as depicted in Fig. 4(a). To address this issue, a "guide" training approach was attempted. Using the existing model as a base, a small number of "problematic" images were included in the training set for several training iterations. The objective was to assess if the problem could be corrected while maintaining the detection capability of the original model. The result of this "guide" training is illustrated in Fig. 4(b), demonstrating that there is no overfitting phenomenon and the issue with the incorrect detection has been resolved.

Based on the above analysis and the inspiration from the "guide" training, finally propose a phased gradient training strategy, as shown in Fig. 5. The gradient training strategy is executed in two stages. In stage 1, a larger learning rate ($\alpha_1 = 10^{-5}$) is set to expedite the convergence of initial network weights. The training is conducted for 200 epochs, where each epoch signifies a complete iteration over the dataset. In the stage 2, a smaller learning rate ($\alpha_2 = 10^{-7}$) is employed to ensure optimal convergence and enhance the detection accuracy of the model. The training continues for 300 epochs, slightly surpassing the duration of stage 1. Fig. 5 illustrates the gradient training scheme, with the y-axis representing the number of individuals included in the training set, and the x-axis denoting the number of training epochs.

- 1) Initially, a small training set comprising a few individuals was utilized to train the model with randomly initialized weights. After a specific number of iterations, the model weights were halted and saved, as depicted in the first column of Fig. 5.
- 2) The training set's size was gradually increased by including more individuals. The training process continued for a specific number of iterations based on the previous weight, and the model weight was saved thereafter.
- 3) The number of individuals in the training set is gradually increased in a gradient manner until all 50 individuals have completed the training. When the training set reaches 30 individuals, the number of training sessions is relatively high, as illustrated in the third column of Fig. 5.

- 4) The training set is reduced from 50 individuals to 30 individuals, and a relatively smaller number of 'guided' training sessions are conducted. This is depicted in column 6 of Fig. 5.
- 5) The training set is further reduced from 30 individuals to 10 individuals, and a smaller number of secondary 'guided' training sessions are conducted. The final dental impression detection model is obtained, as depicted in column 7 of Fig. 5.

The detection result of the gradient model is depicted in Fig. 3(c). Compared to Fig. 3(b), it can be observed that the bounding box is more accurate, the confidence is higher, and there are fewer false detections in the test samples.

The experimental results are presented in Table 1. A total of 916 test images were evaluated with a confidence threshold of 0.5. The non-maximum suppression (NMS) algorithm was employed to exclude bounding boxes with an Intersection over Union (IOU) greater than 0.35. 663 images were accurately detected with all dental prints, accounting for 72.38% of the total. The majority of incorrect results involved missing tooth prints, while only a few cases exhibited incorrect bounding boxes.

To enhance the positioning accuracy of bounding boxes and account for practical application scenarios, the actual number of tooth prints can be specified when inputting the image to be detected. This paper proposes an adaptive threshold strategy, as illustrated in Fig. 6. The strategy involves comparing the number of tooth prints detected with the manually inputted number of tooth prints, and dynamically adjusting the threshold until these two numbers are equal. The details of the adaptive threshold strategy are as follows.

- (1) Input the dental impression image to be detected along with the corresponding number of tooth prints, denoted as n . The default confidence threshold is set to $t = 0.5$.
- (2) Perform tooth print detection by excluding bounding boxes with a confidence lower than t , resulting in n' bounding boxes.
- (3) Evaluate the result. If $n' > n$, it indicates the presence of additional tooth prints with high confidence. In such cases, the threshold should be increased to exclude the extra tooth prints. Set $t = t + 0.1$. Conversely, if $n' < n$, it suggests that the conditions are too strict. In such instances, relax the conditions by setting $t = t - 0.1$. If $n' = n$, the test result is deemed correct, and skip to step (6).
- (4) Re-detect the image using the new threshold value t , resulting in n'' tooth print bounding boxes. If $n'' = n$, it signifies that the correction process is completed, and skip to step (6). If $n' \neq n''$, it implies that the adaptive threshold adjustment was insufficient. Return to step (3) and continue to incrementally increase or decrease t accordingly. If $n' \neq n \neq n''$, it indicates an "over-correction" has occurred, where the detection has shifted from the presence of additional tooth prints to the omission of tooth prints, or vice versa. In such cases, go to step (5).

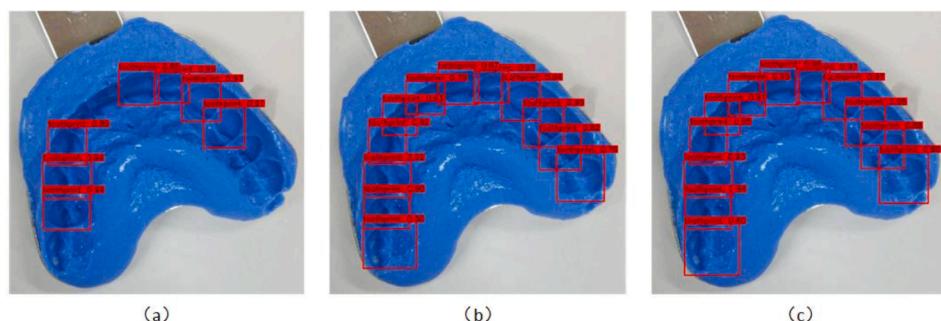


Fig. 3. (a) Test result of the 50-person model. (b) Test result of the 30-person model. (c) Test result of the gradient model.

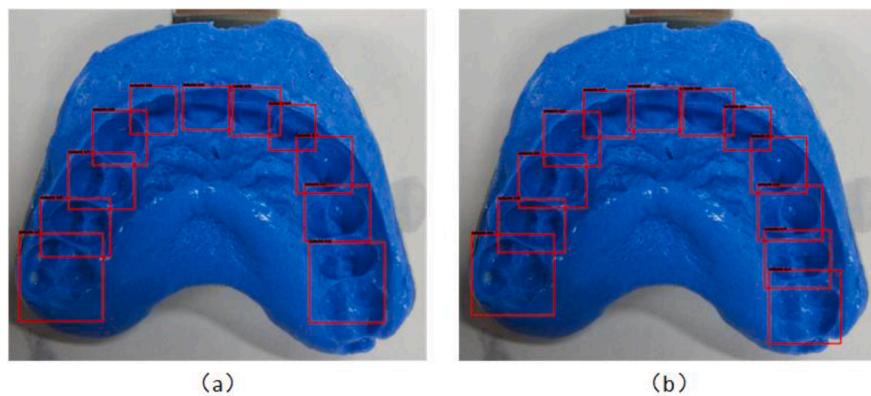


Fig. 4. (a) The result of detecting No.0301 using 30-person model. (b) The result of detecting No.0301 using gradient model.

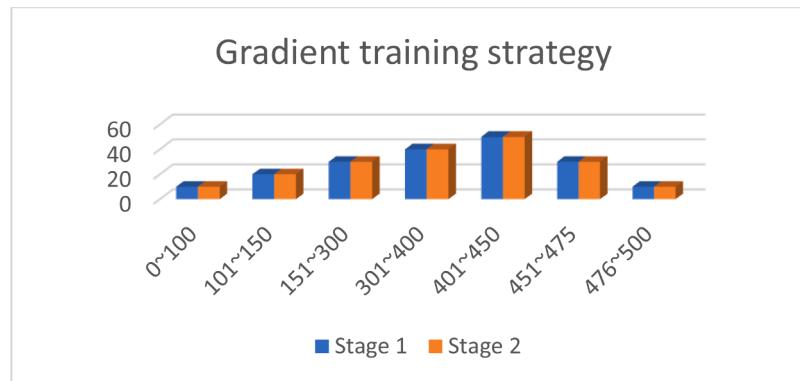


Fig. 5. Gradient training strategy.

Table 1
Accuracy result of tooth print object detection.

	Fully Detected Images	Total Images	Accuracy
Threshold=0.5	663	916	72.38%
Adaptive threshold	842	916	91.92%

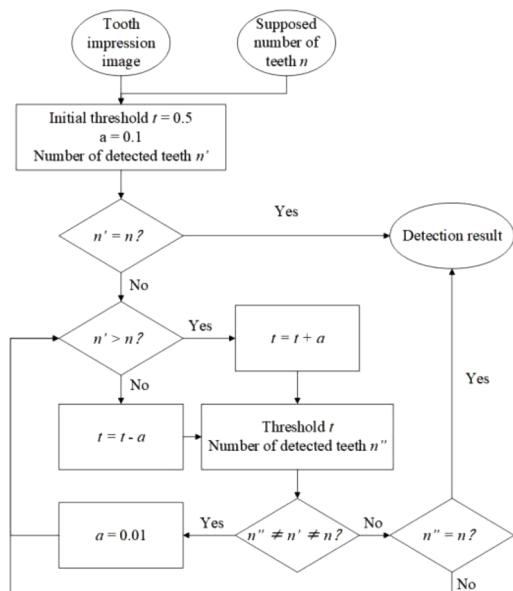


Fig. 6. Adaptive threshold decision process.

(5) Using 0.01 as the increment unit, if $n'' > n$, repeat the process $t = t + 0.01$ until $n'' = n$, and then skip to step (6). If $n'' < n$, repeat the process $t = t - 0.01$. It is important to note that this process is repeated a maximum of 10 times. After 10 iterations, if the threshold reverts back to the value before the initial increment or decrement of 0.1, it indicates that the image is not detected correctly.

(6) Output the result of tooth print detection.

After the adaptive threshold operation, the segmentation results show 842 correct tooth print areas, achieving an accuracy rate of 91.92%. This indicates a successful segmentation of individual tooth print areas. When using a threshold of 0.5, there are 253 images with segmentation errors. However, after the adaptive threshold decision process, 179 of these segmentation errors have been corrected, resulting in a correction rate of 70.75%. Overall, the comprehensive detection effect of our model is highly satisfactory.

In some dental impression images, there are often some "bad" tooth prints. The so-called "bad" tooth prints refer to the unclear or defective contour of the tooth prints, employing gradient training strategies can perfectly detect "bad" tooth prints. Fig. 7 shows detections of some "bad" tooth prints. Therefore, the gradient training strategy has the performance of covering "differences".

3. Multi-scale and multi-dimensional tooth print aggregation feature

3.1. Division of multi-scale tooth print areas

Dividing the original image into multiple scales as input can effectively enhance the stability of features, a technique widely used in object

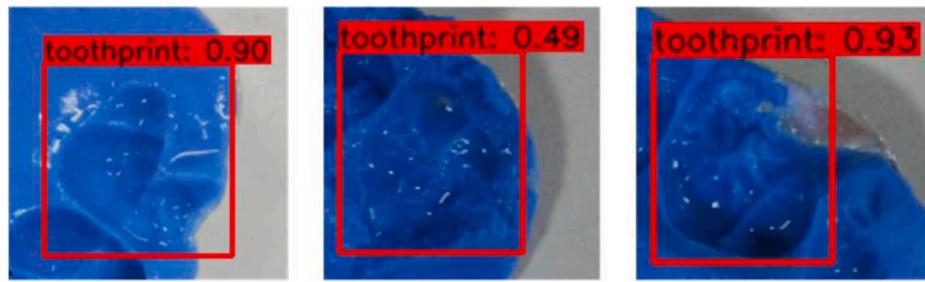


Fig. 7. Detection result of some “bad” tooth prints.

detection networks and certain recognition networks [43]. In this paper, dental impressions are divided into three scales: global, local, and single. The number of scale regions increases from fewer to more, while the number of tooth prints within each region decreases from more to fewer, as depicted in Fig. 8.

The division of the local scale is based on dental classification, which includes incisors, canines, premolars, and molars. The specific areas for each local scale are generated, as illustrated in Fig. 9. It is worth noting that there is only one canine on each side, and its area overlaps with a single tooth print area, thus it is not utilized as a separate scale. Additionally, although there are multiple human molars, most of them are located inside the oral cavity. Only a portion of the molars can be acquired through dental impressions, and the number of acquired tooth prints varies depending on the tightness of an individual’s teeth. Greater tightness in teeth results in more tooth prints, whereas looser teeth tightness corresponds to a lesser number of tooth prints.

3.2. Design of multi-dimensional dental impression image features

In previous work [5], good results have been achieved in identification using tooth print geometric features, which helped narrow down the candidates for identity. However, the uniqueness of its feature recognition is still lacking. To enhance the feature description capability, this paper introduces a novel feature called calibrated offset distance, along with the expansion angle of the dental arch, to form a low-dimensional geometric feature. Additionally, the paper incorporates the Scale-Invariant Feature Transform (SIFT) [44], which is capable of describing fine details as high-dimensional features. Finally, by combining the multi-scale and multi-dimensional features, the tooth print aggregation feature is constructed.

3.2.1. Calibrated offset distance feature (CODF) on single scale

In Fig. 10, the uniqueness of the three dental arches is evident. However, it is challenging to describe these features using general mathematical and geometric expressions. The following reasons contribute to this challenge:

- 1) Fitting the arch curve function is challenging. A dental impression typically contains only 10 to 14 tooth print center points, which is insufficient to accurately fit and obtain a curve function for the dental arch.
- 2) Lack of rotation invariance and scale invariance. When the dental impression image is captured from a rotated angle or at a shorter shooting distance, the arc curve function of the dental arch undergoes significant changes. It becomes challenging to calibrate the curve function to a consistent standard due to the lack of rotation invariance and scale invariance.

To address this challenge, this paper introduces a novel geometric feature to describe two key characteristics of the dental arch: the change of radian and the distance between arc breakpoints. Fig. 11(a) serves as an example to illustrate these concepts. The radian changes can be represented by a series of offset angles between each polyline segment and the next line segment. However, calculating the offset angle is relatively complex, and the variation of angles is often minimal.

To address the complexity of calculating the offset angles and the minimal variation in angles, this paper introduces a calibration approach using the fold line segment. Fig. 11(b) is taken as an example to illustrate this concept. The basic point is defined as the center point No. 1, and the fold line segment $\overline{23}$ is translated accordingly. The endpoint with the smaller number, No. 2, is moved to the position of the basic point, and the corresponding projection point for point No. 3 is denoted as $3'$. The included angle between the fold line segment and the calibrated line segment represents the offset angle, indicating the change in radian.

Furthermore, the length of the green line segment $\overline{23}'$ not only reflects the magnitude of the offset angle but is also influenced by the length of the fold line segment. Therefore, it provides information about the distance between arc breakpoints. This novel feature is referred to as the “calibrated offset distance”. The use of the calibrated offset distance instead of the arc fold line is motivated by the need to capture both the change in radian and the distance information of arc breakpoints effectively.

Performing a similar translation operation on all fold lines allows us to obtain a complete series of calibrated offset distances, as depicted in

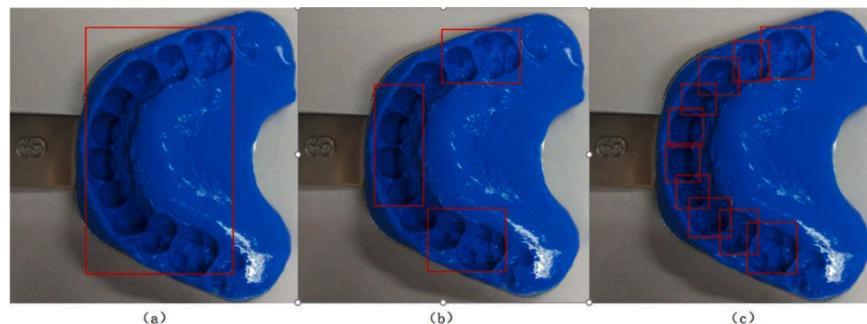


Fig. 8. (a) Global scale area. (b) Local scale areas. (c) Single scale areas.

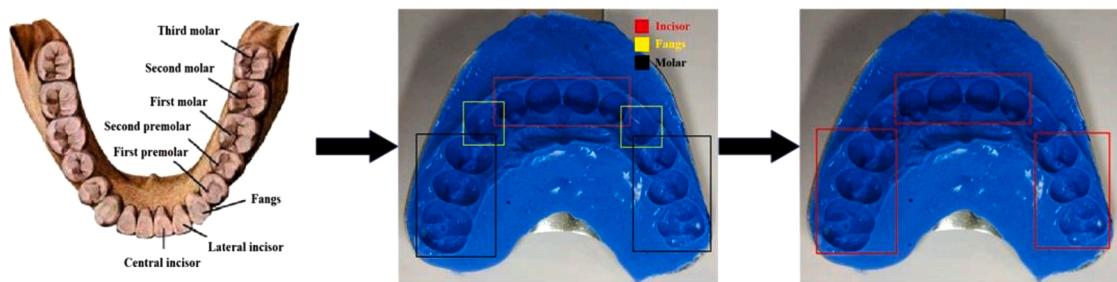


Fig. 9. The process of generating local tooth print areas.

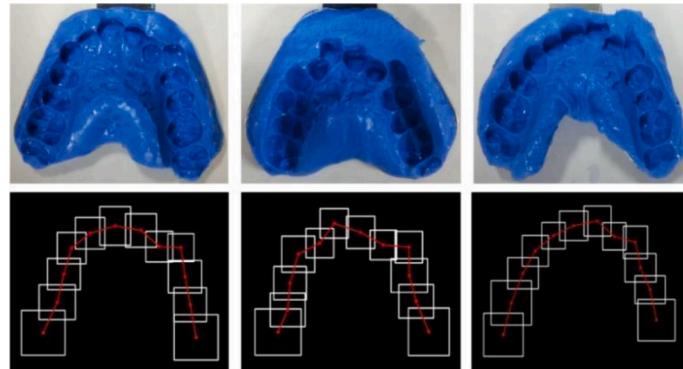


Fig. 10. Overall structures of dental arches.

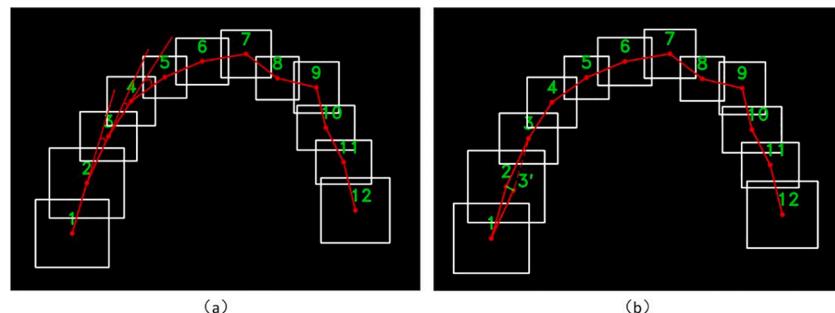


Fig. 11. (a) Offset angle. (b) Calibrated offset distance.

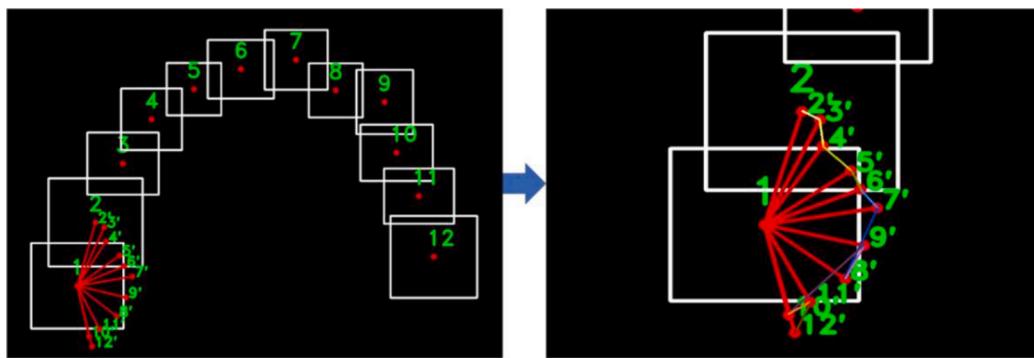


Fig. 12. The calibrated offset distance sequence of a unique individual.

Fig. 12. It is observed that the distance from $7'$ to $8'$ is relatively large, followed by a return to $9'$, indicating that the 8th tooth print in the dental arch is recessed inward. After the 9th tooth print, the arc returns to a smooth shape, but there is another significant shift, resulting in a

relatively large distance from $9'$ to $110'$. This observation highlights the distinguishable performance of the offset distance feature, as it can capture notable variations and patterns within the dental arch.

With the exception of a few tooth prints with sudden curvatures, Fig. 13 illustrates the difference in calibrated offset distance series

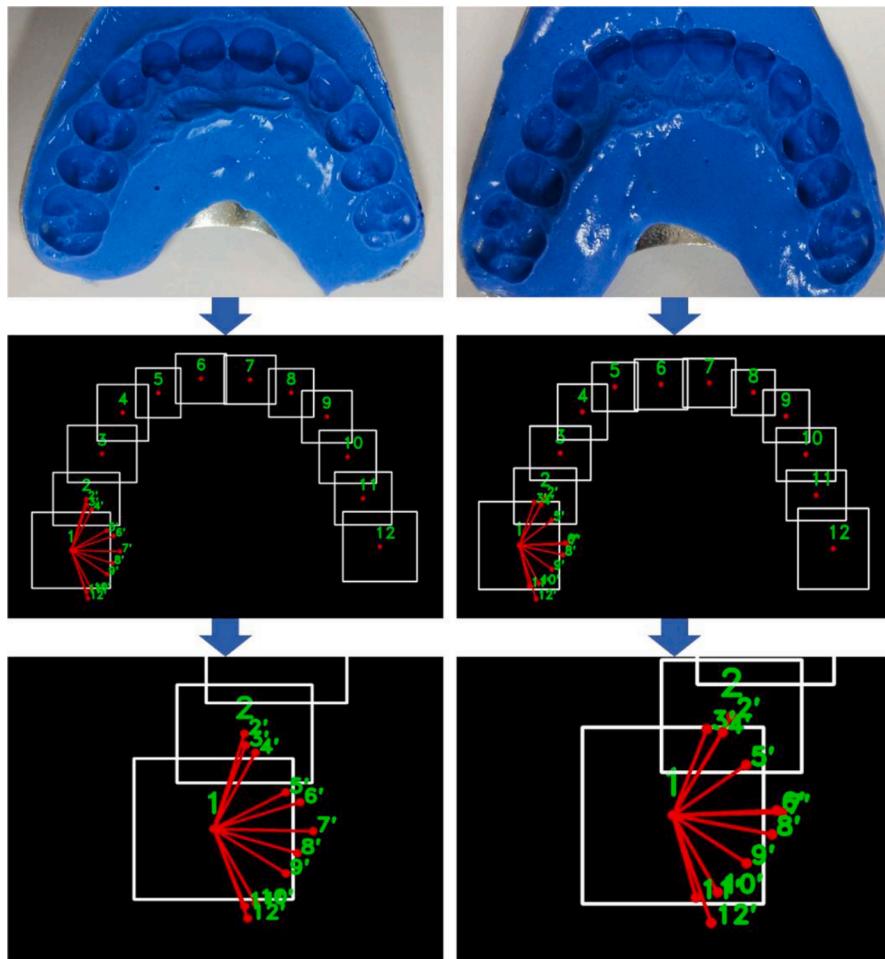


Fig. 13. Calibrated offset distance series of two normal individuals.

between individuals with normal and smooth dental arches. Under naked-eye observation, the two dental arches in the original images appear quite similar. However, after calibration, a significant difference in the calibrated offset distance is observed from 4' to 7', which is reflected in the position of the No. 5 lateral incisor in the right image. The right lateral incisor appears to be more outward and closer to the level of the two front teeth. Consequently, the angular deviation between 45 to 56 is greater, and the offset distance between 5' to 6' is larger compared to the left image.

Furthermore, in the right image, 56 and 67 are almost parallel, resulting in points 6' and 7' being nearly coincident, leading to an extremely small offset distance. This observation demonstrates that our calibrated offset feature effectively captures and describes the nuances of the dental arch.

The detailed calculation steps for generating a calibrated offset distance series are as follows:

- (1) Obtain n tooth print bounding boxes on a single scale of an input dental impression image, and assign numbers to the center points of these boxes as 1 ~ n . The coordinates of each point N_n are as follows:

$$N_n = (x_n, y_n) \quad (1)$$

- (2) Taking point No. 1 (x_1, y_1) as the base point, translate each fold line segment to obtain a series of projections, including $n - 1$ points. The coordinates of each projected point N'_n are as follows:

$$N'_n = (\bar{x}_n, \bar{y}_n) = (x_n - (x_{n-1} - x_1), y_n - (y_{n-1} - y_1)), n \geq 2 \quad (2)$$

(3) It can be calculated the calibrated offset distances \bar{d}_n :

$$\bar{d}_n = \sqrt{(\bar{x}_n - \bar{x}_{n-1})^2 + (\bar{y}_n - \bar{y}_{n-1})^2}, n \geq 3 \quad (3)$$

Calibrate the offset distance series \bar{d}_k ($k = 3, 4, \dots, n$) to form a one-dimensional vector, denoted as $\vec{\bar{d}}$. The distance vector can be obtained as follows:

$$\vec{\bar{d}} = [\bar{d}_3, \bar{d}_4, \bar{d}_5, \dots, \bar{d}_n], n \geq 3 \quad (4)$$

- (4) Divide $(\vec{\bar{d}} - \bar{d}_3)$ by \bar{d}_3 to obtain the scaling factor for calibration. This results in the calibrated offset distance vector, denoted as $\vec{bias_d}$:

$$\vec{bias_d} = \left[\frac{\bar{d}_4}{\bar{d}_3}, \frac{\bar{d}_5}{\bar{d}_3}, \dots, \frac{\bar{d}_n}{\bar{d}_3} \right], n \geq 4 \quad (5)$$

The calibrated offset distance vector, also known as the CODF, has a dimension of $n - 3$, where n represents the number of tooth prints. The

CODF vector is denoted as $\overrightarrow{bias_d}$.

The advantages of the CODF are as follows: (1) Calibration. The CODF combines the characteristics of the dental arch, including the radian and the distance between arc breakpoints, through calibration. (2) Robustness. The CODF feature exhibits translation invariance and rotation invariance, enhancing its robustness in different orientations and positions. (3) Scalability. The CODF reduces the dimensionality of the feature vector while maintaining scale invariance, allowing for scalability and adaptability to different dental arch sizes.

3.2.2. The expansion angle feature of dental arch on local scale

In work [44], three features are extracted, i.e., the number of tooth prints, the ratio of the maximum and the minimum tooth print area, and the included angle between two-fold line of the dental arch. Since the number of tooth prints is an “either this or that” feature, there is no discrimination between dental impression models with the same tooth print number, and there is a 100% discrimination between those with the different numbers. Also, the area ratio feature has the following problems: (1) Contours of tooth prints can be easily deformed with different shooting views; (2) It is difficult to design an efficient, accurate and stable tooth print contour extraction algorithm.

In the previous work [5], three features were extracted: the number of tooth prints, the ratio of the maximum and minimum tooth print area, and the included angle between two-fold lines of the dental arch. However, these features have some limitations.

The number of tooth prints feature is binary, providing discrimination only between dental impression models with the same tooth print number. It does not offer any discrimination within models with the same number of tooth prints. The area ratio feature faces the following challenges:

- (1) Contours of tooth prints can easily be deformed due to variations in shooting views, making the area ratio feature less reliable.
- (2) Designing an efficient, accurate, and stable tooth print contour extraction algorithm is difficult.

These limitations highlight the need for more robust and discriminative features for dental impression analysis.

Therefore, this paper connects the center point of the incisor area in the local scale with the center points of the two end teeth to form an included angle, which is named the expansion angle of the dental arch. As illustrated in Fig. 14(a), this angle reflects the overall expansion of the dental arch. It is observed that when there are more tooth prints, the expansion angle tends to be smaller, while with fewer tooth prints, the expansion angle becomes larger, as demonstrated in Fig. 14(b). The calculation of the expansion angle can be performed using Eq. (6).

$$\theta = \cos^{-1} \left(\frac{a^2 + b^2 - c^2}{2ab} \right) \quad (6)$$

Here, the function $\cos^{-1}(\cdot)$ represents the inverse cosine function. In Eq. (6), θ represents the expansion angle of the two green straight lines

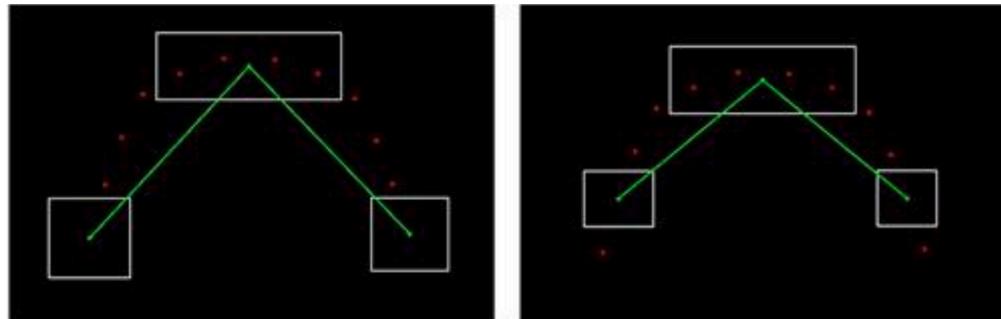


Fig. 14. (a) Expansion angle of 12 tooth prints. (b) Expansion angle of 10 tooth prints.

in Fig. 14. The variables a and b denote the lengths of the two adjacent sides, while c represents the length of the opposite side.

3.2.3. Multi-scale SIFT features

The concept of multi-scale in this paper encompasses both the global scale and the local scale. The low-dimensional geometric features capture the overall information of the teeth structure, providing strong expressiveness. However, these features may not have the ability to distinguish individuals with similar structures. To enhance the accuracy further, it is necessary to incorporate the detailed information from the dental impression. Hence, this paper introduces SIFT [29] as high-dimensional features to capture intricate details.

In this paper, the SIFT features are not directly extracted from the original dental impression image. Instead, they are extracted from the global scale and local scale of the dental impression. The identity information in a dental impression is primarily contained within the tooth prints and their adjacent areas. However, the outer contour of the entire dental impression also contains stable and prominent SIFT feature points. The shape of this contour can be influenced by various factors, such as the amount of material used during the dental impression process and the tilt of the person's head during image acquisition. Therefore, by intercepting the global region as the input, although some prominent feature points may be eliminated, it aligns better with practical application scenarios.

The comparison of SIFT feature points before and after interception is illustrated in Fig. 15. Each circle represents a SIFT feature point, where the center represents the location, the radius indicates the stability of the feature point, and the line segment represents the direction. By transitioning from the original image to the global scale, numerous prominent but irrelevant feature points are excluded, resulting in a focus on the tooth prints and their surrounding areas.

Building upon this concept, the paper further narrows down the feature range to the local scale. Although the local areas are already included within the global area, relying solely on global features without incorporating local features may lead to the exclusion of certain feature

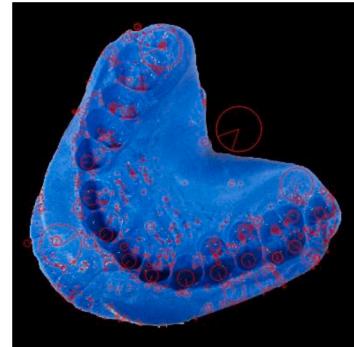
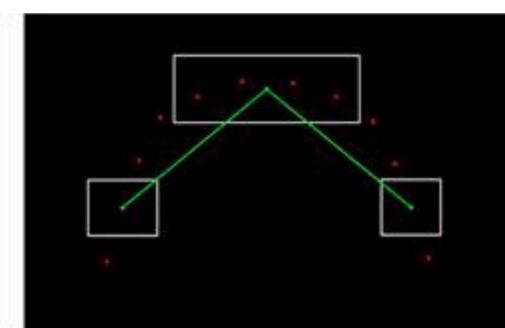


Fig. 15. The SIFT feature points of the dental impression image.



points that rank highly within the local scale but lower within the global scale. This is because the paper only selects the first few pairs of features with the highest matching degree. Consequently, by introducing the local scale, this paper can enhance the diversity of features and provide a more comprehensive representation.

3.3. Similarity quantification

To achieve better feature classification, this paper adopts a different classification approach. Instead of the traditional multi-target classification, where the goal is to determine which person a dental impression image belongs to, this paper employs a two-class classification approach to determine whether two images belong to the same person. This transformation allows for an increased number of positive and negative samples, thus fulfilling the requirements of machine learning classification algorithms and improving the overall classification accuracy.

For instance, let's consider a scenario with 10 dental impression images from person No. 1 and 10 images from person No. 2, resulting in a total of 20 data samples. In the two-class classification scenario, pairwise comparisons are conducted. The comparisons among the 10 images of person No. 1 yield $9 + 8 + 7 + \dots + 1 = 45$ positive samples. Similarly, comparisons between the images of person No. 1 and person No. 2 generate $2 \times 45 = 90$ positive samples. Lastly, comparing an image from person No. 1 with images from person No. 2 leads to $10 \times 10 = 100$ negative samples. Therefore, it is necessary to perform similarity quantification to further analyze and differentiate between positive and negative samples.

3.3.1. Math notations conventions

For ease of understanding and clarity, this paper provides a comprehensive explanation of the mathematical notations used. The meanings of each mathematical notation are summarized in Table 2.

3.3.2. Geometric feature similarity quantification

The similarity value ranges between 0 and 1, where a higher value indicates a higher degree of similarity. The calculation steps for the geometric feature similarity are as follows:

- (1) Perform multi-scale geometric feature extraction on the two dental impression images to be recognized. Obtain the arch expansion angles $bias_a_1$ and $bias_a_2$, representing the overall expansion of the dental arch for each image. Additionally, calibrate the offset distance sequences $\overrightarrow{bias_d}_1$ and $\overrightarrow{bias_d}_2$ for each image.
- (2) Calculate the ratio of the two expansion angles using Eq. (7) to represent the similarity of the arch expansion angles.

$$sim_a = \begin{cases} \frac{bias_a_1}{bias_a_2}, & bias_a_1 < bias_a_2 \\ \frac{bias_a_2}{bias_a_1}, & bias_a_1 > bias_a_2 \end{cases} \quad (7)$$

$\overrightarrow{bias_d}_1$ and $\overrightarrow{bias_d}_2$ are feature vectors. For two images with

Table 2
Notation table.

Notation	Expression
$bias_k$	The arch expansion angle of the k th dental impression image
$\overrightarrow{bias_k}$	The calibrate offset distance feature vector of the k th dental impression image
sim_a	The arch expansion angle similarity
sim_d	The calibration offset distance feature similarity
$geoF$	The geometric feature vector
$siftF$	The SIFT feature vector
$aggF$	The aggregation feature

same number of tooth prints, dimensions of the two feature vectors are also the same, so the cosine similarity (Eq. (8)) can be used to represent the CODF similarity.

- (3) The feature vectors $\overrightarrow{bias_d}_1$ and $\overrightarrow{bias_d}_2$ are obtained. Since the two images have the same number of tooth prints, the dimensions of the two feature vectors are also the same. Therefore, the cosine similarity (Eq. (8)) can be used to represent the similarity of the CODF.

$$sim_d = \frac{\overrightarrow{bias_d}_1 \cdot \overrightarrow{bias_d}_2}{|\overrightarrow{bias_d}_1| \times |\overrightarrow{bias_d}_2|} = \frac{\sum_{i=1}^n bias_d_1^i \times bias_d_2^i}{\sqrt{\sum_{i=1}^n (bias_d_1^i)^2} \times \sqrt{\sum_{i=1}^n (bias_d_2^i)^2}}, \quad n_1 = n_2 \quad (8)$$

In Eq. (8), n_1 and n_2 represent numbers of tooth prints in the two images respectively. When numbers are different, formula (8) cannot be used for the calculation, then take a random number between 0.8 and 0.9 to characterize the similarity.

In Eq. (8), n_1 and n_2 represent the numbers of tooth prints in the two images, respectively. When the numbers are different, Eq. (8) cannot be used for calculation. In such cases, a random number between 0.8 and 0.9 is taken to characterize the similarity.

$$sim_d = \text{random}(0.8, 0.9), \quad n_1 \neq n_2 \quad (9)$$

The effectiveness of this approach can be attributed to the following reasons: 1) The number of tooth prints is a significant feature, and samples with different numbers are clearly indicative of different individuals. Therefore, assigning a distinct value in such cases is reasonable. 2) Negative samples with different numbers of tooth prints constitute only a small portion of all negative samples. The range of 0.8 to 0.9 is determined based on the similarity distribution of negative samples with the same number of tooth prints, most of which have similarities below 0.85. This ensures that the method does not artificially enhance the discrimination of samples and maintains its objectivity and effectiveness.

(4) The features, after undergoing similarity quantification, are already in the same dimension. Therefore, they can be merged to obtain the multi-scale low-dimensional geometric feature of dental impressions.

$$geoF = [sim_a, sim_d] \quad (10)$$

3.3.3. SIFT feature similarity quantification

Similar to geometric features, the comparison of SIFT features between two samples is also expressed by a similarity ratio ranging from 0 to 1. The detailed process is described below.

- (1) Perform the fast nearest neighbor search algorithm FLANN [45, 46] on the SIFT features of the two images to be recognized. Use the k-nearest neighbors' algorithm to find the consistency, which means finding the best-matched K feature points in one set for each SIFT feature point in the other set. In this paper, $K = 2$, meaning that the paper considers the first nearest neighbor point and the second nearest neighbor point. A good match will be close to the first nearest neighbor and far away from the second nearest neighbor. Conversely, a poor match will have very similar distances between the two nearest neighbors. This occurs when the algorithm fails to find a more accurate matching point, and the distances between the two poor neighbors are very close. It is important to note that this distance is a similarity distance, not a physical distance.

- (2) Calculate the ratio r of the distances between the first nearest neighbor and the second nearest neighbor. In contrast to geometric similarity, a larger ratio indicates closer distances between the first nearest neighbor and the second nearest neighbor, resulting in poorer feature point matching. Therefore, matching points with a ratio greater than 0.80 are considered outliers and discarded. To align with the practical interpretation of similarity, the inverse operation of $1 - r$ is performed on all distance ratios to obtain the final SIFT similarity set s_Ratio in this paper.
- (3) Sort the similarity ratios in s_Ratio in descending order. The higher the order, the greater the similarity, the smaller the distance ratio, and the more accurate the matching.
- (4) For the set sd_Ratio of the global scale, select the first 8 similarities. For the three regions of the local scale, select the first 4 similarities. As the region becomes smaller, the effective matching decreases, resulting in a decrease in the number of selections. Use the method shown in Fig. 16 to calculate the average value within the corresponding interval, and finally obtain 4 SIFT-based similarity features in the global scale and 3×3 SIFT-based similarity features in the local scale, respectively.
- (5) Combine the 13 similarity features to obtain the multi-scale high-dimensional SIFT similarity features of the dental impressions.

$$siftF = [sim_{total}^1, \dots, sim_{total}^4, sim_{local1}^1, \dots, sim_{local1}^3, \dots, sim_{local3}^3] \quad (11)$$

3.3.4. Classifiers and aggregation feature

In this paper, SVM (Support Vector Machine) is selected as the classifier, which is a kind of supervised learning model [41]. The strategy is to construct a feature space and find a hyperplane so that the minimum distance between the point set of positive samples and the point set of negative samples to the hyperplane is the largest, and the distance between edge points of the two sets and the hyperplane is the largest, that is, the interval is maximized. As the feature dimension increases, the model trained by SVM will become more accurate. The final multi-scale and multi-dimensional tooth print aggregation feature in this paper is shown in Eq. (12), where "+" represents the concatenation and merging of two feature vectors. The aggregation feature has 15 dimensions, so SVM is suitable for classification.

$$aggF = siftF + geoF = [sim_{total}^1, \dots, sim_d, sim_a] \quad (12)$$

4. Experimental results and comparison

4.1. Experiment preparation

The experimental dataset in this paper consists of 842 dental impression images that have accurately completed tooth print detection. The dataset comprises dental impression images from 50 individuals, with 10 to 25 images per person. The image size is 4608×3456 pixels. Among these images, 10 individuals' images are randomly selected as the outer set, while the remaining 40 individuals' dental impression

images form the test, application, and training database. Within the 40-person dental impression images database, two images are randomly chosen as the test set, one as the application set, and the rest as the training set. The structure of the database set is shown in Fig. 17.

In this paper, the classification task is transformed from multi-classification to binary classification. Specifically, instead of determining the category of a single dental impression image, the task is to determine whether two dental impression images belong to the same category. As a result, the images in the training set cannot be directly used as experimental samples. Instead, they need to be compared with two other dental impression images to generate new samples. Fig. 18 illustrates the process of generating a new sample, using the positive sample as an example.

The purpose of the 8-fold cross-validation strategy is to increase the number of samples. The following steps outline how complete positive and negative samples are obtained:

- (1) Obtain the training set containing data from 40 individuals, with each person having n_i dental impression images.
- (2) Construct positive samples by comparing two dental impression images of the same individual. The aggregation feature is calculated using Eq. (13) to obtain the positive sample feature. A label of "1" is added at the end of the feature vector. After comparing images of all individuals, a set of positive samples is obtained. Specifically, n_i images of the i th individual are compared in pairs, resulting in the generation of N_i^+ positive samples.

$$N_i^+ = (n_i - 1) + (n_i - 2) + \dots + 1 = \frac{n_i \times (n_i - 1)}{2} \quad (13)$$

- (3) Construct negative samples by comparing two dental impression images from different individuals. The process of extracting aggregation features is the same as that of the positive samples, with the addition of a label of "0" at the end. To ensure an equal number of negative samples as positive samples, the value of N_i^- is used during the construction process. For each individual i the n_i images are sequentially selected to be compared with a randomly chosen image from other categories, generating a negative sample. This process is repeated n_i times for one round. The rounds continue until the number of negative samples N_i^- matches the number of positive samples N_i^+ . This ensures that negative samples for individual i are generated. Throughout this process, images from other categories are selected at most once to avoid repetition. All individuals undergo the above operations, resulting in a set of negative samples.

- (4) Compose a complete positive and negative sample set with $2N_i^+$ samples. From the aforementioned steps, it can be observed that the positive samples obtained during each construction are identical, whereas the negative samples are highly randomized. Consequently, during each testing process, the sample library is

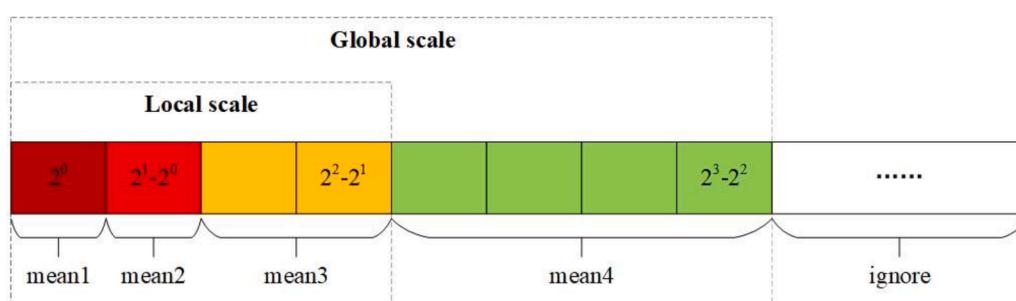


Fig. 16. Multi-scale tooth print features based on SIFT.

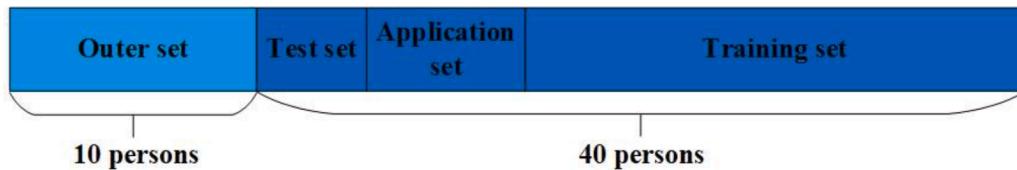


Fig. 17. The structure of the data set.

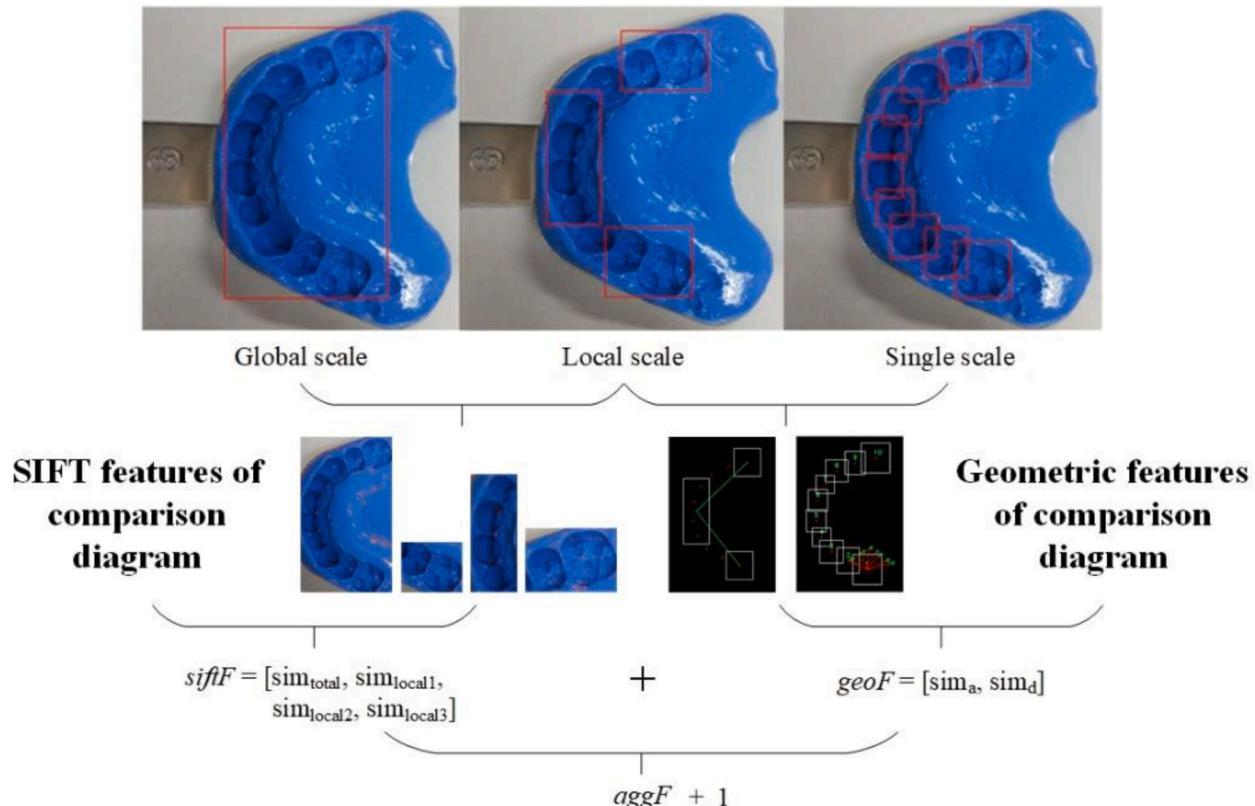


Fig. 18. Construction process of a positive sample.

reorganized to enhance the objectivity of the results and measure the algorithm's robustness.

4.2. Analysis, comparison, and discussion

The experimental results are divided into the following parts.

4.2.1. Training model

In this paper, the classification task is transformed from multi-classification to binary classification. In other words, instead of determining the category of a single dental impression image, the task is to determine whether two dental impression images belong to the same category. As a result, the images in the training set cannot be directly used as experimental samples. Instead, they need to be compared with two other dental impression images to generate new samples.

Using the construction method of positive and negative samples described in Section 4.1, N_i^+ positive samples and N_i^- negative samples are generated. The SVM [49] classifier is then utilized to fit the training data and obtain the hyperplane that maximizes the margin, i.e., the model parameters. In the training process, this paper employed an 8-fold cross-validation strategy to enhance the generalization ability of the model.

Cross-validation [42] plays a crucial role in improving the

generalization ability of machine learning projects, which is beneficial for addressing the issue of limited availability of dental impression images.

K-fold cross-validation is a commonly used technique in machine learning. It involves randomly dividing the original dataset into K subsets, where K-1 subsets are used as the training set in each iteration, and the remaining subset serves as the test set. This process is repeated K times, and the average accuracy across all iterations is taken as the final evaluation metric for the model.

Choosing an appropriate value for K is essential. If K is too small, the number of validations will be insufficient, resulting in less persuasive experimental results. On the other hand, if K is too large, it may lead to high system overhead. Currently, in machine learning, $K = 8$ is commonly used and considered the most popular choice for cross-validation [42]. The advantages of 8-fold cross-validation are as follows: 1) It is used to assess the prediction performance of the model, particularly how well the trained model performs on new, unseen data. This helps reduce overfitting to some extent. 2) It allows for extracting as much effective information as possible from limited sample data.

For the $2N_i^+$ complete positive and negative samples, this paper utilized the 8-fold cross-validation strategy [42] during the training process to evaluate the model's generalization ability. The training set was evenly divided into 8 parts, with one part selected for classification

verification, while the remaining 7 parts were used for training. This process was repeated 8 times, resulting in a total of 8 classification accuracy rates. These accuracy rates help verify the presence of overfitting and objectively assess the model's performance. Finally, the support vector machine (SVM) model m was fitted and saved.

Table 3 presents the 8-fold cross-validation accuracy of the model during training. The SIFT feature exhibits a relatively lower recognition ability because it only captures the global and local areas of tooth prints, thereby excluding certain unique feature points that may lack practical application value. Nevertheless, the accuracy rate still reaches around 80%, indicating its potential to complement low-dimensional features and aggregation features with high-dimensional and detailed features.

The multi-scale geometric features demonstrate a high recognition ability, achieving an average verification accuracy of 91.45%. Furthermore, the accuracy of the multi-scale and multi-dimensional aggregation feature is further improved to 93.38%. This enhancement can be attributed to the fact that geometric features struggle to distinguish between different individuals with highly similar dental arches, while SIFT provides a suitable means to describe the details. By combining aggregation features, these images can be successfully distinguished, resulting in improved accuracy.

4.2.2. Test model

In this paper, a sample consists of two dental impression images. If the two images belong to the same person, the sample is labeled as a positive sample (marked as "1"). Conversely, if the images belong to different individuals, the sample is labeled as a negative sample (marked as "0"). Therefore, it is necessary to define evaluation metrics to assess the performance of the model.

- 1) Positive sample accuracy: This refers to the accuracy rate of correctly recognizing positive samples, which involves accurately identifying that the sample consists of dental impression images from the same person.
- 2) Negative sample accuracy: This refers to the accuracy rate of correctly recognizing negative samples, which involves accurately identifying that the sample consists of dental impression images from different individuals.
- 3) Test accuracy: This refers to the accuracy rate of correctly identifying both positive and negative samples, regardless of their specific category. It measures the overall accuracy of the model in correctly classifying the samples.

A similar test sample generation process was conducted on 80 test images from 40 individuals in the test set. For each person, one positive sample and one negative sample were generated, resulting in a total of 80 samples. The model m obtained from experiment (1) was applied to predict the categories of the sample set. The predicted labels were then compared with the actual labels to calculate the recognition accuracy. This evaluation aimed to test the model's generalization ability within the database.

Table 3
Results of model cross-validation.

	SIFT features	Geometry features	Aggregation features
Rates of 8 verification	0.80668258	0.92362768	0.94033413
accuracy Rates of 8-fold	0.80071599	0.91408115	0.94033413
cross-validation	0.79116945	0.91408115	0.92840095
	0.76968974	0.91885442	0.93078759
	0.80787589	0.91646778	0.94152745
	0.80548926	0.90930788	0.92362768
	0.80071599	0.90334129	0.94033413
	0.81384248	0.91646778	0.92482100
Average verification	0.799522673	0.914528640	0.933770883

Table 4 presents the test accuracy of geometric features and aggregation features. The geometric features exhibit a relatively high positive sample recognition rate, indicating accurate identification of two dental impression images belonging to the same individual. This is mainly attributed to the calibrated offset distance feature that we employed. However, when it comes to negative samples, dental arches of different individuals may exhibit similarities, making it challenging to distinguish them solely based on geometric features. Nevertheless, the aggregation of SIFT features improves the recognition of negative samples, highlighting the effectiveness of our algorithm model in achieving satisfactory recognition results.

Fig. 19 depicts the actual labels of the test samples and the predicted labels obtained through recognition using their aggregation features. The sample grid with an incorrect prediction is represented by a dark blue color, and the label is marked in red.

4.2.3. Application model and analysis comparison

This part evaluates the performance of our model in practical applications by identifying anonymous dental impressions using the database. During application, it is necessary to convert the model's prediction results from binary classification to multi-classification. This can be achieved through the following five steps: 1) Obtain the aggregation feature arrays for all dental impression images in the training set, as well as the aggregation features of the image to be detected. 2) Compare the features to be detected with each feature from the training set individually, resulting in a similarity feature array. The dimensions of this array match the aggregation feature array. 3) Apply the model to identify each similarity feature sample in the array, resulting in predictions. 4) Calculate the ratio of the number of samples predicted as "1" for the i th individual to the total sample number of the i th individual. This ratio represents the probability P_i that the image belongs to the i th individual. 5) Sort the probabilities P_i in descending order. The corresponding individual identities and probabilities P_i provide the sequence of detection results and their respective confidence levels.

Table 5 presents the identification results of the 40 images in the application set. The correct recognition is considered only when the identity probability ranks in the top N (marked as TOP-N) of the result sequence. The proposed aggregation feature model achieved an 85% detection rate in TOP-1 and 100% in TOP-3. Out of the 6 images where TOP-1 failed to detect the correct identity, 2 of them had a detection probability $P_i = 1.00$ for the actual identity, resulting in a best accuracy rate of 90% in TOP-1. However, these two images were misclassified, as they also had another predicted identity with a probability of 1.00.

This paper also conducted a comparison of our model with related methods presented in previous works [5,47,48]. When compared to the classic dental X-ray method [47], our model demonstrates higher accuracy in identification using single jaw teeth. In comparison to the deep learning-based X-ray method [48], our model may be slightly less accurate in TOP-1 results but achieves 100% accuracy faster. Moreover, compared to [5], which also utilizes dental impression images, our model shows a significant improvement in accuracy rate. In fact, the model achieves a TOP-2 detection accuracy rate of 39/40, which translates to 97.5%. Overall, the proposed model in this paper performs well in the task of identification.

Fig. 20 illustrates the detailed probability values of TOP-1 for the

Table 4
Test results and comparison.

	Geometry features	Aggregation features
Test Accuracy	72/80 90.00%	76/80 95.00%
Positive sample accuracy	38/40 95.00%	38/40 95.00%
Negative sample accuracy	34/40 85.00%	38/40 95.00%
5 times average test accuracy	90.75%	93.25%

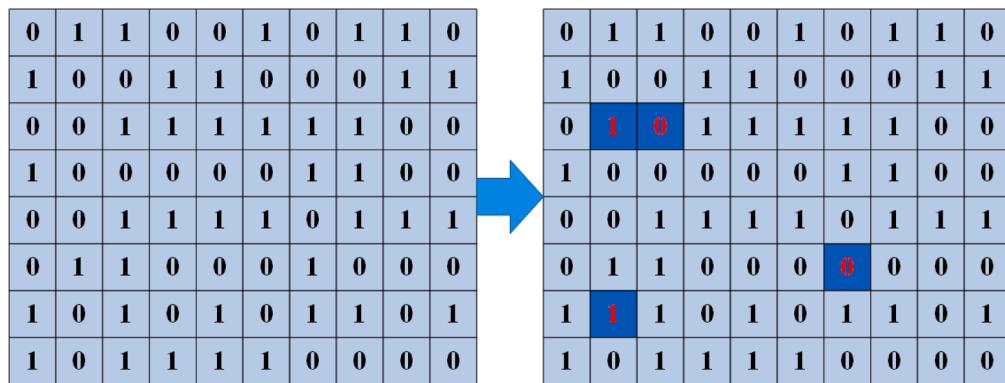


Fig. 19. Test sample labels.

Table 5
Comparison of dental identification accuracy.

Method	Jaw	TOP-N retrievals		
		N = 1	N = 3	N = 5
[44]	upper	130/311	229/311	268/311
		41.80%	73.63%	86.17%
[48]	upper	50/115	77/115	88/115
		43.50%	66.90%	76.50%
[49]	lower	50/105	72/105	83/105
		47.60%	68.60%	79.00%
Ours	both	87.21%	—	95.34%
		34/40	40/40	40/40
		85.00%	100%	100%

identification of the 40 images in the application set. The dark blue grids with red numbers represent the incorrectly recognized samples and their corresponding confidence levels. The dark blue grids with black numbers indicate the presence of two identities with a confidence of 1.00 in the results.

Most of the correctly recognized probabilities are above 0.80, and the majority of the results have a confidence level of 1.00. This indicates that our model's recognition capability is highly effective. However, there is also a relatively high probability of misclassification, suggesting that our model still faces challenges in distinguishing individuals with similar structures. Although this paper has incorporated SIFT as high-dimensional features, there are cases where individuals with similar geometric structures and SIFT detail features are not adequately discriminated. This is also the reason for the two examples with a confidence of 1.00 for both the true and false identities.

It is noteworthy that our model achieves a direct 100% identification in TOP-3, indicating that the actual identity probability of these misidentified samples is only slightly lower than that of the false identity. Thus, the overall recognition effect remains satisfactory.

Fig. 21 displays the TOP-3 probabilities of the application set, where the red markers indicate the true identity probabilities of the corresponding images. In terms of the second probability P-2, some samples exhibit similarities with the actual identity probability, indicating that the model performs reasonably well in distinguishing and identifying individuals with similar characteristics. On the other hand, the third probability P-3 shows a significant gap with the actual identity probability for most samples. This suggests that the model can effectively differentiate between different individuals in such cases. Consequently, the TOP-3 detection successfully identifies all the dental impression images in the application set.

In conclusion, the actual application effect of our model is favorable, aligning with the result analysis in the training and testing process. However, one limitation is observed in the high-dimensional aspect of the aggregation feature, SIFT, which exhibits a slightly inadequate ability to distinguish samples with similar geometric features. As a result, the accuracy is only elevated to the TOP-3 level, and there remains a gap in achieving precise identification at the P-1 and P-2 levels.

4.2.4. Examination of generalization ability in outer set database

For the outer set database, which does not include the identity category of the dental impressions, this paper conducted two experiments to demonstrate two scenarios in practical applications.

Scenario 1: Comparing and identifying two dental impressions that have not been entered in the database to determine whether they belong to the same person. The expected output is either "1" or "0", indicating whether they are from the same person or not.

Scenario 2: Identifying a dental impression image whose real identity is not present in the database. The expected output is "none," indicating that there is no corresponding identity in the database for the given image.

Based on the experimental analysis mentioned above, we classify the detection result as "none" when P-1 is less than 0.5 or P-4 is greater than

0.80	0.55	1.00	0.89	1.00	0.71	0.60	0.90
0.57	1.00	0.64	0.92	1.00	1.00	1.00	1.00
1.00	1.00	1.00	1.00	1.00	1.00	0.78	0.60
0.58	1.00	0.86	1.00	0.60	0.80	0.89	1.00
1.00	0.79						

Fig. 20. Probability value of TOP-1 detection.

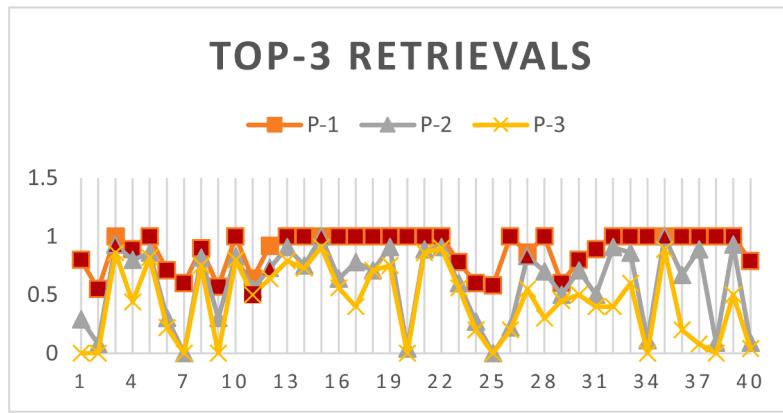


Fig. 21. Probability value of TOP-3 detection.

0.5. This decision is made because, with a 100% recognition rate in TOP-3, if the probability P-4 is still high, it suggests relatively poor detection performance.

Table 6 presents the number of experimental samples and the results for the two scenarios. In the first scenario, positive and negative samples were constructed for all 10 individuals outside the database, following the process shown in Fig. 18. The model m was then used to predict these samples. Ultimately, 1487 out of 1764 positive and negative samples were accurately classified, resulting in an accuracy rate of 84.30%. This is a noteworthy achievement for an outer set database that has not been included in the model training process and lacks entered identities. It demonstrates the model's effectiveness in recognition during practical applications.

In the second scenario experiment, each person contributed 10 images, resulting in a total of 100 images. The model was used to identify these images, out of which 61 were detected as "none." This indicates that the model has a certain ability to exclude identities in outer set data, but there is room for improvement in this aspect.

Based on the aforementioned four experiments, the multi-scale and multi-dimensional dental impression aggregation feature model demonstrates a high level of recognition accuracy. It achieves a verification accuracy of 93.37% and a test accuracy of 93.25%. In the best-case scenario, the TOP-1 identification in the application set reaches 90%. Moreover, the model exhibits a certain degree of generalization ability when dealing with data from an outer set database.

However, there are still limitations to consider. The model's ability to distinguish dental impressions with similar teeth structures is not strong enough. Additionally, its capability to exclude data that lies outside the database requires further improvement.

4.2.5. Comparison of classification models

There is no universally superior classifier, as different classification methods can yield varying recognition results. In order to compare the recognition performance of SVM [48], this paper also conducted a comparative experiment involving three classifiers: Decision Tree [50], Random Forest [51], and Naive Bayes [52]. The comparison results are illustrated in Fig. 22.

In Fig. 22, the horizontal axis represents 10 test groups, while the vertical axis represents the accuracy rate of 10 samples per group. The experimental data was derived from a dental impression database consisting of 50 individuals. For each test group, 10 individuals were

randomly selected, and one dental impression image was chosen for each individual.

Based on Fig. 22, it is evident that the accuracy of SVM generally surpasses the other three classification models. Its highest achievement is successfully identifying all 10 individuals, while even in the worst case, it can still correctly identify 8 individuals. Conversely, the Decision Tree classifier performs the poorest, with its highest record being 8 individuals successfully identified, and in the worst case, only 5 individuals can be recognized. The recognition performance of the Random Forest classifier is second only to SVM, while Naive Bayes exhibits slightly lower accuracy compared to the Random Forest classifier.

5. Conclusions and the further works

This paper proposes an identification method based on dental impression images. In order to detect tooth prints, the tooth print is treated as a kind of object. This paper utilizes the Yolov8 network, design a gradient training strategy, and propose an adaptive threshold strategy, resulting in a 91.92% accuracy rate. Regarding tooth print feature extraction, this paper summarizes two major types of information related to the dental arch. Additionally, this paper proposes a novel feature called the calibrated offset distance feature that considers all the information. This feature not only indicates the distance information of the arc break point, but also maintains the rotation invariance of the dental impression image. It not only considers the integrity of the dental shape structure, but also reflects whether the tooth is inward concave or outward convex.

To enhance the feature representation and compensate for the limited recognition of samples with similar dental arches based solely on geometric features, this paper proposes a novel approach called multi-scale and multi-dimensional tooth print aggregation feature. This approach involves dividing dental impression image into global scale, local scale, and single scale, from which high and low-dimensional features are extracted and aggregated into a comprehensive feature vector. Specifically, the low-dimensional geometric features encompass the expansion angle and calibrated offset distance of the dental arch, and they are distributed across the local and single scales. On the other hand, the high-dimensional detail features are extracted from both global and local scales, with separate calculations of SIFT feature points in each respective area.

The proposed aggregated features in this paper attained a validation accuracy of 93.38% in the training dataset, an accuracy of 93.25% in the test dataset, and a recognition rate of 100% in the application dataset. This demonstrating the effectiveness of our method in successfully completing the recognition task. However, when evaluating the performance in two scenario experiments conducted on the outer set database, the recognition rates were found to be 84.3% and 61% respectively. It is worth noting that the recognition rates in these two

Table 6
Examination of the generalization ability in outer set database.

	Number of samples	Accuracy
Scenario 1	1487/1764	84.30%
Scenario 2	61/100	61.00%

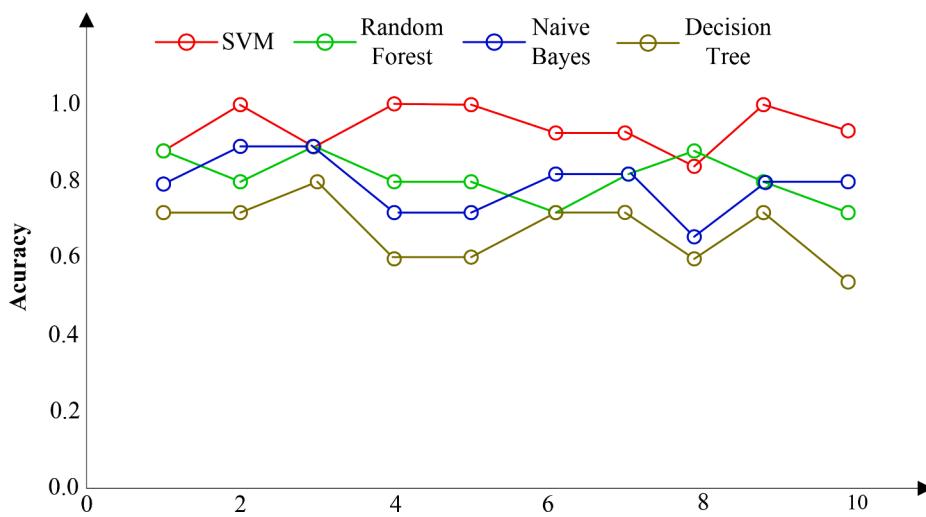


Fig. 22. Recognition accuracy of four classification models.

scenarios were relatively lower, indicating room for improvement in those specific scenarios.

The extracted features are not comprehensive enough, as they fail to capture important characteristics of the teeth, such as the depth and area of the dental hole. These features are not adequately represented by the existing feature set. Additionally, while the calibrated offset distance provides information about the dental arch, it does not encompass the depth or length information of the dental impressions or individual teeth. Consequently, there is a need for further improvement in the feature extraction process. Addressing these limitations represents a crucial direction for future research. These shortcomings of this method are also the direction for further research. In further work, the following areas of investigation will be focused on:

- 1) Future research will be focused on developing methods to accurately measure the depth of tooth holes in dental impression images and will explore various techniques and algorithms to extract the tooth length information from these images.
- 2) In the upcoming research, the calculation method for determining the tooth hole area will be studied in dental impression images. This objective is to develop an effective approach to measure the size of each tooth accurately.
- 3) The tooth structure, depth of the tooth hole, and area of the tooth hole are synergistically integrated to achieve a comprehensive representation of the tooth impression image. This holistic approach aims to enhance the identification rate based on dental impression images.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This work is supported by National Natural Science Foundation of China (No. 62176237), The “Pioneer” and “Leading Goose” R&D Program of Zhejiang Province (Grant No. 2023C01022).

References

- [1] E. Ziętkiewicz, M. Witt, P. Daca, et al., Current genetic methodologies in the identification of disaster victims and in forensic analysis[J], *J. Appl. Genet.* 53 (1) (2012) 41–60.
- [2] Reshma Amin, Pushparaja H Shetty, Angelina Shetty, Reliability of Teeth for Identification after Exposure to varying Degrees of Temperature[J], *World J. Dent.* 8 (2) (2017) 96–103.
- [3] J. Hinchliffe, Forensic odontology, Part 1. Dental identification, *Br. Dent. [J]* 210 (5) (2011) 219–224.
- [4] T. Wu, L.B Zhang, Three-dimension tooth reconstruction using level set active contour model[J], *J. Image Graph.* 21 (8) (2016) 1078–1087.
- [5] S. Jiang, W. Min, L. Liu, et al., Multi-scale multi-view deep feature aggregation for food recognition[J], *IEEE Trans. Image Process.* 29 (2019) 265–276.
- [6] K.P. Khandare, A.A Gurjar, Dental biometric approach for human identification using dental X-ray images of maxillary bone[J], *Int. Res. J. Eng. Technol.* 3 (2016) 1566–1570.
- [7] R.F. Silva, A. Franco, F.F. Picoli, et al., Dental human identification using bitewing radiographs—a case report[J], *Eur. J. Forens. Sci.* 3 (3) (2016) 31–33.
- [8] Chunfeng Song, Yongzhen Huang, Yang Huang, et al., GaitNet: an end-to-end network for gait based human identification, *Pattern Recognit.* 96 (Dec) (2019).
- [9] Yifu Zhang, Chunyu Wang, Xinggang Wang, Wenjun Zeng, VoxelTrack: multi-Person 3D Human Pose Estimation and Tracking in the Wild, *IEEE Trans. Pattern Anal. Mach. Intell.* (2022), <https://doi.org/10.1109/TPAMI.2022.3163709>.
- [10] Randa Atta, Said Port, Samir Shaheen, Mohammad Ghanbari, Human identification based on temporal lifting using 5/3 wavelet filters and radon transform, *Pattern Recognit.* 69 (9) (2017) 213–224.
- [11] Virginio Cantoni, Chiara Galdi, Michele Nappi, Marco Porta, Daniel Riccio, GANT: gaze analysis technique for human identification, *Pattern Recognit.* 48 (4) (2015) 1027–1038.
- [12] Wang Yanxiang, Xian Zhang, Yiran Shen, et al., Event-Stream Representation for Human Gaits Identification Using Deep Neural Networks, *IEEE Trans. Pattern Anal. Mach. Intell.* 37 (1) (2021) 93–104.
- [13] Siarohin Aliaksandr, Lathuiliere Stephane, Sangineto Enver, Sebe Nicu, Appearance and Pose-Conditioned Human Image Generation Using Deformable GANs, *IEEE Trans. Pattern Anal. Mach. Intell.* 43 (4) (2021) 1156–1171.
- [14] Oh Seong Joon, Benenson Rodrigo, Fritz Mario, Schiele Bernt, Person Recognition in Personal Photo Collections, *IEEE Trans. Pattern Anal. Mach. Intell.* 42 (1) (2020) 203–220.
- [15] J. Zhou, Abdel-Mottaleb M. A content-based system for human identification based on bitewing dental X-ray images[J], *Pattern Recognit.* 38 (11) (2005) 2132–2142.
- [16] O. Nomir, Abdel-Mottaleb M. A system for human identification from X-ray dental radiographs[J], *Pattern Recognit.* 38 (8) (2005) 1295–1305.
- [17] O. Nomir, Abdel-Mottaleb M. Fusion of matching algorithms for human identification using dental X-ray radiographs[J], *IEEE Trans. Inform. Forens. Secur.* 3 (2) (2008) 223–233.
- [18] A.K. Jain, H. Chen, Matching of dental X-ray images for human identification[J], *Pattern Recognit.* 37 (7) (2004) 1519–1532.
- [19] O. Nomir, Abdel-Mottaleb M. Hierarchical contour matching for dental X-ray radiographs[J], *Pattern Recognit.* 41 (1) (2008) 130–138.
- [20] H. Chen, A.K. Jain, Dental biometrics: alignment and matching of dental radiographs[J], *IEEE Trans. Pattern Anal. Mach. Intell.* 27 (8) (2005) 1319–1326.
- [21] M. Hofer, A.M Marana, Dental biometrics: human identification based on dental work information[C], in: XX Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI 2007), IEEE, 2007, pp. 281–286.

- [22] O. Nomir, M. Abdel-Mottaleb, Human identification from dental X-ray images based on the shape and appearance of the teeth[J], IEEE Trans. Inform. Forens. Secur. 2 (2) (2007) 188–197.
- [23] P.L. Lin, Y.H. Lai, P.W. Huang, Dental biometrics: human identification based on teeth and dental works in bitewing radiographs[J], Pattern Recognit. 45 (3) (2012) 934–946.
- [24] P. Rajput, K.J. Mahajan, Dental biometric in human forensic identification[C], in: 2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC), IEEE, 2016, pp. 409–413.
- [25] M. Banday, A.H. Mir, Dental biometric identification system using AR model[C], in: TENCON 2019-2019 IEEE Region 10 Conference (TENCON), IEEE, 2019, pp. 2363–2369.
- [26] Z. Cui, C. Li, Wang W. ToothNet, Automatic tooth instance segmentation and identification from cone beam CT images[C], in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 6368–6377.
- [27] Y. Lai, F. Fan, Q. Wu, et al., LCANet: learnable Connected Attention Network for Human Identification Using Dental Images[J], IEEE Trans. Med. Imaging 40 (3) (2021) 905–915.
- [28] V. Pushparaj, U. Gurunathan, B. Arumugam, Dental radiographs and photographs in human forensic identification[J], IET Biometrics 2 (2) (2013) 56–63.
- [29] F. Rehman, M.U. Akram, K. Faraz, et al., Human identification using dental biometric analysis[C], in: 2015 Fifth International Conference on Digital Information and Communication Technology and its Applications (DICTAP), IEEE, 2015, pp. 96–100.
- [30] R. Kumar, Teeth recognition for person identification[C], in: 2016 International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS), IEEE, 2016, pp. 13–16.
- [31] G.E. Miranda, S.G. de Freitas, L.V. de Abreu Maia, et al., An unusual method of forensic human identification: use of selfie photographs[J], Forensic Sci. Int. 263 (2016) e14–e17.
- [32] Xin Xilin, Tu Yidong, Stojanovic Vladimir, et al., Online Reinforcement Learning Multiplayer Non-Zero Sum Games of Continuous-Time Markov Jump Linear Systems, Appl. Math. Comput. 412 (2022) 1–15, 126537Jan2022.
- [33] Xu Zhiliu, Li Xiaodi, Stojanovic Vladimir, Exponential stability of nonlinear state-dependent delayed impulsive systems with applications, Nonlin. Anal. Hybrid Syst. 42 (2021) 1–12, 101088Jul., 2021.
- [34] Zhang Xiang, Wang Hai, Stojanovic Vladimir, et al., Asynchronous Fault Detection for Interval Type-2 Fuzzy Nonhomogeneous Higher-level Markov Jump Systems with Uncertain Transition Probabilities, IEEE Trans. Fuzzy Syst. 30 (7) (Jul. 2022) 2487–2499.
- [35] Stojanovic Vladimir, Novak Nedic, Joint state and parameter robust estimation of stochastic nonlinear systems, Int. J. Robust Nonlin. Contr. 26 (14) (2016) 3058–3074.
- [36] Wu Lin, Liu Deyin, Zhang Wenying, et al., Pseudo-Pair based Self-Similarity Learning for Unsupervised Person Re-identification, IEEE Trans. Image Process. 31 (2022) 4803–4806.
- [37] Yang Hao, Chu Xixiu, Zhang Li, et al., QuadNet: quadruplet loss for multi-view learning in baggage re-identification, Pattern Recognit. (2022), <https://doi.org/10.1016/j.patcog.2022.108546>. Online.
- [38] Shi Yuxuan, Wei Zhen, Ling Hefei, et al., Person Retrieval in Surveillance Videos via Deep Attribute Mining and Reasoning, IEEE Trans. Multimedia 23 (2021) 4376–4387.
- [39] Han Ke, Huang Yan, Song Chunfeng, et al., Adaptive super-resolution for person re-identification with low-resolution images, Pattern Recognit. 14 (2021), 107682. Jun2021.
- [40] Patruno Cosimo, Marani Roberto, Cicirelli Grazia, et al., People re-identification using skeleton standard posture and color descriptors from RGB-D data, Pattern Recognit. 89 (2019) 77–90. May 2019.
- [41] Jyunrong Wang, Huafeng Dai, et al., Toward surface defect detection in electronics manufacturing by an accurate and lightweight YOLO-style object detector, Sci. Rep. 13 (2023) 7062. May 2023.
- [42] Ruiz-Ponce Pablo, Ortiz-Perez David, et al., POSEIDON: a Data Augmentation Tool for Small Object Detection Datasets in Maritime Environments, Sensors 23 (7) (2023), 3691:1–14.
- [43] Amr Noaman Abdel-Hamid, Hamouda Alaa El-deen, C.M.M.I. Lean, An Iterative and Incremental Approach to CMMI-Based Process Improvement, in: 2015 Agile Conference, Washington, DC, USA, 2015, pp. 65–70. Aug., 03–07.
- [44] L. Wang, J. Mao, Y. Hu, et al., Tooth identification based on teeth structure feature [J], Syst. Sci. Contr. Eng. 8 (1) (2020) 521–533.
- [45] S. Gupta, K. Thakur, M. Kumar, 2D-human face recognition using SIFT and SURF descriptors of face's feature regions[J], Vis. Comput. (2020) 1–10.
- [46] Suzuki Ikumi, Hara Kazuo, Centered kNN Graph for Semi -Supervised Learning, in: 40th International ACM SIGIR conference on research and development in Information Retrieval, Shinjuku, JAPAN, 2017, pp. 857–860. Aug., 07–11.
- [47] J. Jo, J. Seo, Fekete J.D. Panene, A progressive algorithm for indexing and querying approximate k-nearest neighbors[J], IEEE Trans. Vis. Comput. Graph. (1) (2018), 1–1.
- [48] D. Meyer, F.H.T. Wien, Support vector machines[J], The Interface to libsvm in package e1071 (2015) 28.
- [49] M. De Rooij, W. Weeda, Cross-validation: a method every psychologist should know[J], in: Adv. Method Pract. Psychol. Sci., 3, 2020, pp. 248–263.
- [50] Hongkuan Sok, Melanieileen Ooi, Yechow Kuang, et al. Multivariate alternating decision tree, Pattern Recognit. 50 (Feb) (2016) 195–209.
- [51] L. Breiman, Random forests, Mach. Learn. 45 (1) (2001) 5–32.
- [52] P. Tsangaratos, I. Iliia, Comparison of a logistic regression and Naïve Bayes classifier in landslide susceptibility assessments: the influence of models complexity and training dataset size[J], Catena 145 (2016) 164–179.



Mao Jiafa received the Ph.D. degree in pattern recognition from East China University of Science and Technology, China, in 2009. Since then, he has worked as a Post-doc Researcher at Beijing University of Posts and Telecommunications, China. In July 2011, he joined Zhejiang University of Technology, China, where he is currently a Professor in the School of Computer Science & Technology. His research interests include computer vision, pattern recognition, digital image processing and information hiding. He has published over 50 papers in the scientific literature.



Lixin Wang was born in WenZhou, Zhejiang, China. He received the M received the B.S. degree from Zhejiang University of Technology in 2017. Graduated from Zhejiang University of Technology with a M.S. degree in computer science and technology in 2021. His research interests include computer vision, pattern recognition and image processing.



Ning Wang received the B.S. degree from Zhoukou Normal University in 2017. Graduated from Zhejiang University of Technology with a M.S. degree in computer science and technology in 2022. From 2017 to 2022, he was a researcher in the School of Computer Science and Technology, Zhejiang University of Technology. His research interests include computer vision, pattern recognition and image processing.



Yahong Hu received her bachelor's degree, master's degree, and Ph.D. all from Xi'an Jiaotong University in 1992, 1995 and 1999 respectively. She now is an associate professor at the College of Computer Science of Technology, Zhejiang University of Technology. Her research interests include image analysis and big data processing.



Weiguo Sheng received the M.Sc. degree in information technology from the University of Nottingham, U.K., in 2002 and the Ph.D. degree in computer science from Brunel University, U.K., in 2005. Then, he worked as a Researcher at the University of Kent, U.K. and Royal Holloway, University of London, U.K. He is currently a Professor at Hangzhou Normal University. His research interests include evolutionary computation, data mining/clustering, pattern recognition and machine learning.