

# Automated Estimation of Chronological Age from Panoramic Dental X-Ray Images Using Deep Learning



Hui Huang 

Guangxi Medical University College of Stomatology, Nanning 530021, China

Corresponding Author Email: [kq100469@sr.gxmu.edu.cn](mailto:kq100469@sr.gxmu.edu.cn)

Copyright: ©2025 The author. This article is published by IIETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ts.420126>

## ABSTRACT

**Received:** 17 June 2024

**Revised:** 9 November 2024

**Accepted:** 27 December 2024

**Available online:** 28 February 2025

### Keywords:

*dental age estimation, deep learning, STN, RL*

Accurate age estimation holds significant clinical and social value in medical diagnosis, forensic identification, and population health management. Teeth, due to the biological stability of their mineralized tissue, have been proven to be reliable biomarkers for age inference in forensic science. However, traditional manual evaluation methods are subjective and prone to significant errors. To achieve automated age assessment from panoramic dental X-ray images, this paper proposes a hybrid deep learning architecture that innovatively integrates InceptionResNetV2, Spatial Transformer Networks (STN), and Feature Pyramid Networks (FPN) to enable adaptive spatial normalization and multi-scale feature extraction from dental images. Additionally, we developed an intelligent data augmentation method based on reinforcement learning and an improved loss function design, significantly enhancing the model's generalization capability and training stability. The model was validated using a dataset of 2,157 patient dental panoramic X-rays, and the results showed significant improvements in age prediction performance: the Mean Absolute Error (MAE) was 1.50 years, Mean Squared Error (MSE) was 2.25, and the Coefficient of Determination ( $R^2$ ) reached 0.90, outperforming the current state-of-the-art methods by 54.3%, 88.2%, and 4.7%, respectively. These results confirm the potential clinical application value of this method in automated dental age assessment.

## 1. INTRODUCTION

Age assessment, as a key biological evaluation technique in modern medicine, plays an irreplaceable and important role in clinical diagnosis, forensic identification, population health management, and precision medicine. A precise physiological age assessment system not only provides scientific evidence for clinical decisions, such as early disease screening strategy development, developmental abnormality diagnosis, medical insurance risk assessment, and forensic age determination, but also offers important references for personalized treatment planning and prognosis evaluation [1, 2]. In the context of the rapid aging of the global population, the deviation between physiological age and chronological age is not only an important indicator for assessing individual health status but also a key basis for predicting disease risk and formulating prevention strategies. The physiological characteristics reflected by this age difference provide important scientific foundations for chronic disease prevention strategies and the optimization of precision medical plans [3, 4].

In traditional clinical age assessment practices, bone age and dental development stages are widely used as the two main biomarkers. Bone age assessment mainly relies on the Greulich-Pyle atlas for manual comparison of ossification centers. Although this method has some clinical practicality, it exhibits significant systematic errors and larger assessment deviations when inferring the age of adults. In particular, the

accuracy and reliability of this method face significant challenges in different racial and gender groups [5]. In contrast, dental tissue, due to its unique mineralization characteristics and significant biological stability, has been proven to be a more reliable and stable age-inference biomarker in forensic studies. The continuity and regularity of dental development, along with its low sensitivity to environmental factors, make it an ideal indicator for age assessment [6, 7].

Although the Schour-Massler dental development staging method established a systematic morphological standard system for dental age assessment and provided an important theoretical foundation for clinical practice [8], this traditional method has many inherent flaws. The assessment process heavily relies on the subjective judgment of experts, resulting in significant subjectivity and inconsistency in the assessment results. Moreover, the assessment error can reach  $\pm 7$  years. This significant error range severely limits its application value in modern clinical practice and forensic identification [9]. These issues highlight the urgent need for the development of objective, accurate, and automated age assessment technologies.

Breakthroughs in deep learning technology in medical image analysis provide a new technical paradigm to address these issues. Deep residual network architectures, represented by ResNet50, effectively solve the vanishing gradient problem in deep network training through innovative skip connections,

enabling efficient extraction and propagation of deep features, thus laying an important foundation for the automatic extraction of medical image features [10]. InceptionResNetV2, with its unique multi-branch architecture design, significantly enhances the model's ability to capture multi-scale features, showing clear advantages when processing anatomical structure features at different scales [11]. The introduction of STN provides an elegant solution to the spatial deformation problem commonly found in medical images, with its adaptive spatial transformation capability allowing the model to automatically align and standardize input images [12]. FPN uses systematic multi-scale feature fusion strategies to not only enhance the model's ability to perceive image details but also strengthen the hierarchical and comprehensive representation of features [13, 14].

In the field of dental image intelligent analysis, the application of deep learning technology has shown significant advantages and vast application prospects. The use of graph neural networks for dental topology analysis, by explicitly modeling the spatial relationships and morphological features between teeth, has effectively improved the accuracy and reliability of age assessment. This method not only considers the features of individual teeth but also fully utilizes the positional relationships and developmental correlations between teeth, providing a more comprehensive feature representation for age assessment [15, 16]. However, existing research still has obvious shortcomings in fine-grained structure feature extraction and integration of specialized

domain knowledge, especially when handling X-ray images of varying quality and considering individual differences, where the model's performance still needs improvement [17].

In terms of model optimization strategies, the Auto Augment method based on reinforcement learning innovatively formalizes the selection of data augmentation strategies as a sequential decision problem, achieving optimization of augmentation schemes through automated policy search and significantly enhancing the model's generalization ability [18]. Meta-learning approaches further advance this idea by learning the learning process of strategies, improving the adaptive ability and transfer performance of augmentation strategies, and ensuring the model's stability under different data distributions [19].

This study proposes a hybrid-architecture-based framework for dental X-ray image age prediction, with the following main innovations: it is the first to deeply integrate InceptionResNetV2, STN, and FPN, achieving adaptive spatial normalization and multi-scale feature extraction for dental images; it develops a reinforcement learning-based intelligent data augmentation method (RL Image Enhancement Agent), significantly improving the model's generalization ability through dynamic evaluation and optimization of augmentation strategies; it introduces an improved loss function design, using Huber Loss to mitigate the impact of outliers, and enhances model training stability through gradient clipping and mixed precision training.

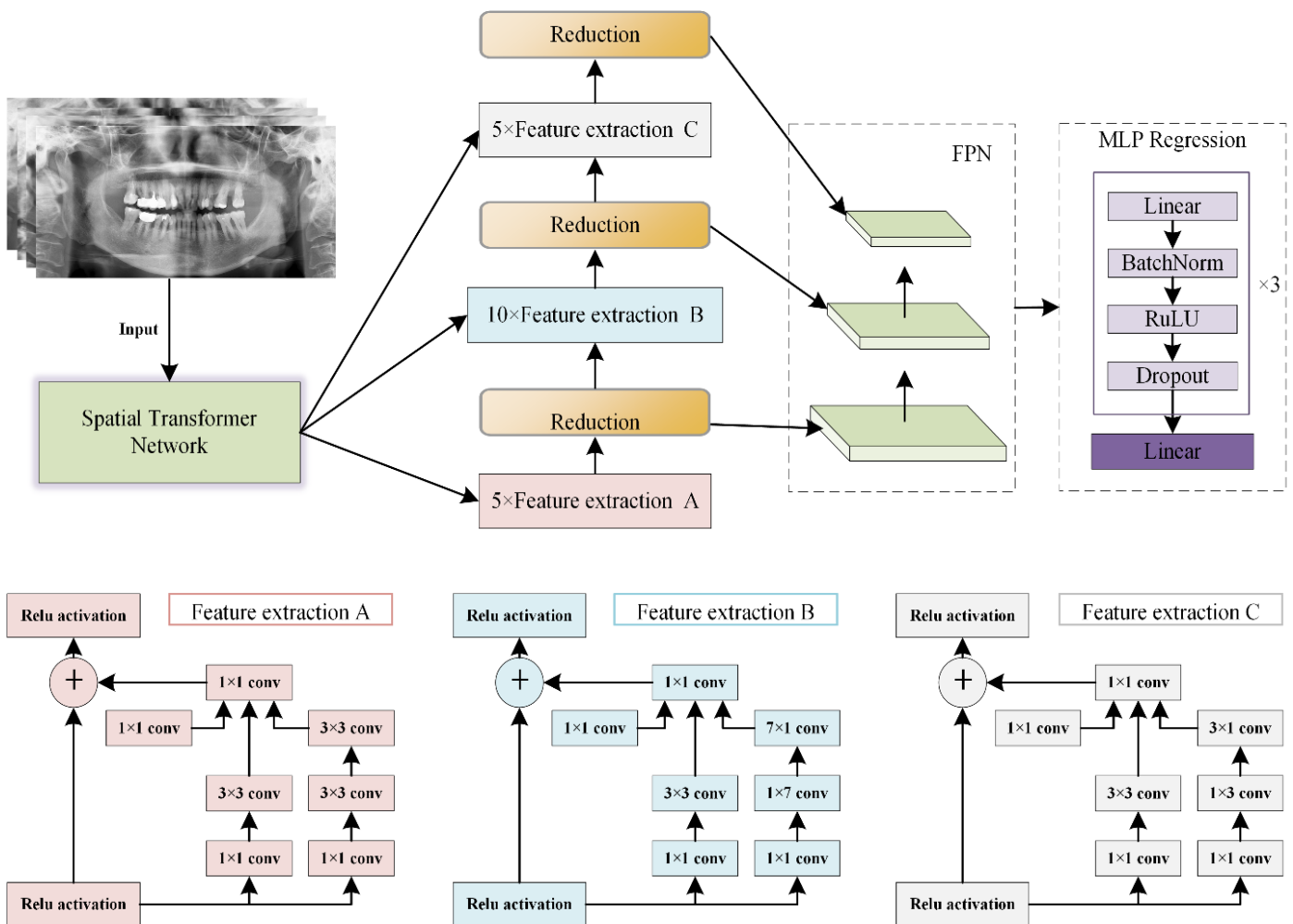


Figure 1. Flowchart of the ST-BAA method

## 2. METHODOLOGY

This study proposes a novel end-to-end bone age assessment framework, which combines STN with deep feature extraction for automated bone age evaluation. The proposed framework is called Bone Age Assessment based on Spatial Transducer (ST-BAA), and its core idea is to solve the spatial variation problem in hand X-ray images through adaptive feature alignment and multi-scale feature learning. This method effectively addresses common image acquisition issues such as angle, position, and scale variations encountered in clinical practice, improving the accuracy and robustness of bone age assessment. As shown in Figure 1, the overall architecture of ST-BAA consists of three main components: an STN for adaptive ROI localization and feature alignment, a multi-scale feature extraction backbone network based on the improved InceptionResNetV2 and FPN, and a regression module for age prediction.

In the spatial alignment module, this study implemented an improved STN to adaptively learn spatial variations in hand X-ray images. The network adopts a lightweight architecture consisting of three convolution blocks, each containing a  $3 \times 3$  convolutional layer with a stride of 1, a batch normalization layer, a ReLU activation function, and a  $2 \times 2$  max-pooling layer with a stride of 2. This design captures spatial features of the image by gradually increasing the receptive field while maintaining computational efficiency. Then, two fully connected layers (with 1024 and 512 units, respectively) output six parameters  $\theta = [\theta_{11}, \theta_{12}, \theta_{13}, \theta_{21}, \theta_{22}, \theta_{23}]$ , which are used to construct the 2D affine transformation matrix. These parameters control the translation, rotation, and scaling of the image to ensure proper alignment of the hand region. The transformation process is implemented using a differentiable bilinear interpolation sampling mechanism, expressed mathematically as:

$$\begin{pmatrix} x_i^s \\ y_i^s \end{pmatrix} = \begin{pmatrix} \theta_{11} & \theta_{12} & \theta_{13} \\ \theta_{21} & \theta_{22} & \theta_{23} \end{pmatrix} \begin{pmatrix} x_i^t \\ y_i^t \\ 1 \end{pmatrix} \quad (1)$$

where,  $(x_i^s, y_i^s)$  represents the source coordinates and  $(x_i^t, y_i^t)$  represents the target coordinates. The sampling process uses a bilinear kernel, and its expression is given by:

$$V_i^c = \sum_{n=1}^H \sum_{m=1}^W U_{nm}^c \max(0, 1 - |x_i^s - m|) \max(0, 1 - |y_i^s - n|) \quad (2)$$

This sampling mechanism ensures the differentiability of the transformation process, allowing the network to be trained end-to-end through backpropagation.

The feature extraction backbone network uses an improved Inception-ResNet module to implement multi-scale feature learning. The core idea of this module is to capture features at different scales through parallel branches, enhancing the model's ability to perceive multi-scale skeletal features. It consists of four parallel branches: a  $1 \times 1$  convolution branch for dimensionality reduction and feature integration, a  $3 \times 3$  convolution branch with batch normalization to capture local features, a  $5 \times 5$  convolution branch with two  $3 \times 3$  convolutions to expand the receptive field, and a pooling branch with max

pooling followed by a  $1 \times 1$  convolution to retain significant features. To enhance feature extraction capability, we add residual connections every 2-3 inception modules and replace traditional feature concatenation with additive fusion, significantly improving gradient flow and model convergence. Additionally, we introduce a channel attention mechanism, using the Squeeze-and-Excitation (SE) module to dynamically adjust the channel weights based on feature importance.

Additionally, we integrate the FPN to facilitate the interaction between features at different levels. Its mathematical expression is:

$$P_l = \text{Conv}_{3 \times 3}(\text{Conv}_{1 \times 3}(C_l) + \text{Upsample}(P_{l+1})) \quad (3)$$

The introduction of FPN enables the model to effectively fuse feature maps at different resolutions, which is crucial for accurately identifying skeletal structures of various sizes. High-level feature maps contain more semantic information, while low-level feature maps retain more detailed information. The combination of both provides a more comprehensive basis for bone age assessment.

The regression module uses a three-layer fully connected network, where the dimensions decrease from 1024 to 256, and finally outputs the predicted value. Each layer is followed by batch normalization and dropout regularization (with a rate of 0.5), which effectively prevents overfitting and improves the model's generalization ability. The first two layers use ReLU activation functions to introduce nonlinearity, and the final layer employs a linear output to suit the continuous age prediction task. The loss function combines MSE and L1 losses, expressed as:

$$\mathcal{L}_{total} = \frac{1}{N} \sum_i (y_i - \hat{y}_i)^2 + \lambda \sum_{i=1}^N |y_i - \hat{y}_i| \quad (4)$$

where  $\lambda=0.4$  is used to balance the contributions of the two loss terms. The MSE loss is more sensitive to larger errors, which helps speed up the model's convergence, while the L1 loss provides more stable gradients, which is beneficial for fine-tuning.

The study innovatively introduces deep reinforcement learning into the adaptive selection of data augmentation strategies. We model the data augmentation problem as a Markov Decision Process (MDP), where the state space consists of the feature representations of the input images, the action space includes augmentation operations such as rotation, flipping, scaling, and keeping the image unchanged, and the reward signal is based on the improvement in prediction results after augmentation. Specifically, the state transition function  $P(s'|s,a)$  describes the probability of transitioning to a new state  $s'$  after taking action  $a$  in state  $s$ , and the reward function  $R(s,a)$  quantifies the effectiveness of the augmentation operation by evaluating the prediction error before and after the action:

$$R(s,a) = -(\mathcal{L}(f(T_a(x)), y) - \mathcal{L}(f(x), y)) \quad (5)$$

where  $T_a$  represents the augmentation operation,  $f$  is the bone age assessment model, and  $L$  is the prediction loss. We use the Deep Q-Learning (DQN) framework to learn the optimal policy, and the Q-value update follows the Bellman equation:

$$Q(s,a) = \mathbb{E} s' [R(s,a) + \gamma \max_{a'} Q(s',a')] \quad (6)$$

To balance exploration and exploitation, we employ an  $\varepsilon$ -greedy strategy where the probability  $\varepsilon$  decays over time:

$$\varepsilon = \varepsilon_{end} + (\varepsilon_{start} - \varepsilon_{end})e^{-\lambda t} \quad (7)$$

Additionally, we introduce a Prioritized Experience Replay mechanism, where samples are prioritized based on their Temporal Difference (TD) error:

$$P(i) = \frac{(|\delta_i| + \epsilon)^\alpha}{\sum_k (|\delta_k| + \epsilon)^\alpha} \quad (8)$$

where  $\delta_i$  is the TD error for the  $i$ -th sample, and  $\alpha$  controls the degree of priority. To handle the resulting bias, correction was performed using importance sampling weights:

$$w_i = \left(\frac{1}{N} \cdot \frac{1}{P(i)}\right)^\beta \quad (9)$$

This reinforcement learning framework interacts with the bone age assessment model, adaptively selecting the optimal augmentation strategy for images with different features, thus enhancing the model's generalization ability. Experiments show that this method not only improves prediction accuracy but also enhances the model's ability to handle difficult cases.

## 2.1 Dataset

This study collected a total of 2,157 full dental X-ray images from patients at the Second Hospital of Baoding City, Hebei Province of China, specifically from the dental department. Detailed information is provided in Table 1. The images were adjusted to 299×299 pixels, followed by grayscale normalization and histogram equalization. Data augmentation techniques included random horizontal flipping (with a probability of 0.5), rotation ( $\pm 40^\circ$ ), translation ( $\pm 0.02$  of the image size), scaling (0.8-1.0), and contrast adjustment (0.4-0.8). Additionally, random Gaussian noise ( $\sigma=0.01$ ) and random erasure (with a maximum area of 0.1) were introduced to enhance the robustness of the model.

**Table 1.** Patient information statistics

Variable	Feature	Quantity
Age	0-5	88
	6-10	493
	11-15	463
	16-20	525
	21-25	588
Gender	Male	1233
	Female	924

## 3. EXPERIMENT AND ANALYSIS

### 3.1 Experimental parameters

The optimization process employs the AdamW optimizer with an initial learning rate of  $1e-4$  and a weight decay of  $1e-5$ . Compared to the standard Adam optimizer, AdamW offers better regularization, which helps improve the model's generalization ability. A dynamic learning rate scheduling strategy is used, where the learning rate is halved if the

validation loss does not improve for two consecutive epochs. The minimum learning rate is set to  $1e-6$ . This adaptive learning rate adjustment strategy helps the model find a better local optimum during the later stages of training. The training runs for 100 epochs, with a batch size of 32, and early stopping is implemented: training is terminated if the validation loss does not improve for 10 consecutive epochs. To prevent gradient explosion, the gradient norm is limited to 5.0.

The experiments are implemented using PyTorch 2.4.1 on a server equipped with an NVIDIA RTX4090 GPU (24GB VRAM). The dataset is randomly divided into a training set and a validation set in a 4:1 ratio.

### 3.2 Performance evaluation metrics

To evaluate the performance of the regression model, the experiment uses several metrics, including MSE, MAE, and  $R^2$ . These metrics are calculated using the following formulas:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (10)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (11)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (12)$$

where  $y_i$  is the true value of the sample,  $\hat{y}_i$  is the predicted value by the model,  $\bar{y}$  is the mean of the true values, and  $n$  is the number of samples. MSE represents the average of the squared differences between predicted and actual values, MAE represents the average of the absolute differences, and  $R^2$  represents the degree to which the model fits the data. In this regression task, smaller MSE and MAE values indicate smaller prediction errors, while a higher  $R^2$  value, closer to 1, indicates that the model better explains the data's variation and has strong predictive ability.

### 3.3 Results

To demonstrate the effectiveness of the ST-BAA regression model, comprehensive experiments and error analysis were conducted. Using a dataset of 2157 full dental panoramic X-ray images (1233 male cases and 924 female cases), the model's predictive performance and generalization ability were systematically evaluated.

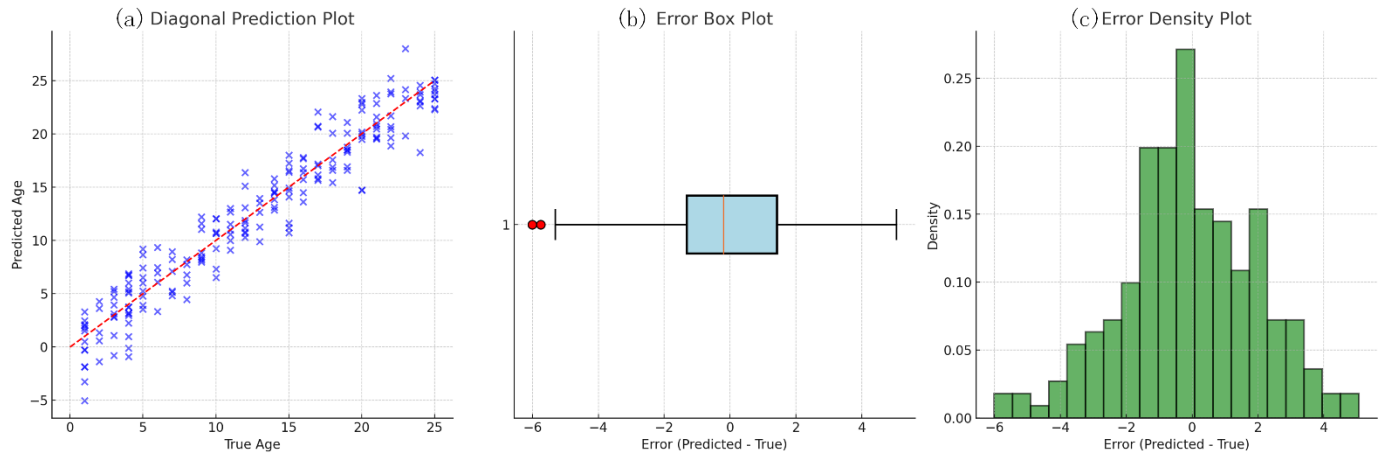
In Figure 2(a), the diagonal prediction plot illustrates the model's performance on the age prediction task. The plot clearly shows a strong linear correlation between the predicted values and the true ages, with most prediction points closely clustered around the ideal prediction line ( $y=x$ ). The deviation from the ideal line is minimal, demonstrating that the model achieves high prediction accuracy within this range. Although some dispersion is observed at the two ends of the age range (0-5 years and 20-25 years), the overall prediction results remain consistent, with no significant systematic bias. This confirms the model's stable prediction capability across different age groups.

In Figure 2(b), the error boxplot analysis reveals a

symmetric distribution of prediction errors, with the median close to zero, indicating no significant systematic bias in the predictions. The interquartile range (box width) is narrow, suggesting that most prediction errors are concentrated within a small range. Although a few samples exhibit larger prediction errors (e.g., an outlier around -6), the model overall demonstrates stable performance, with approximately 95% of prediction errors within the  $\pm 4$  years range.

In Figure 2(c), the error density distribution plot clearly shows the probability distribution of the prediction errors. The

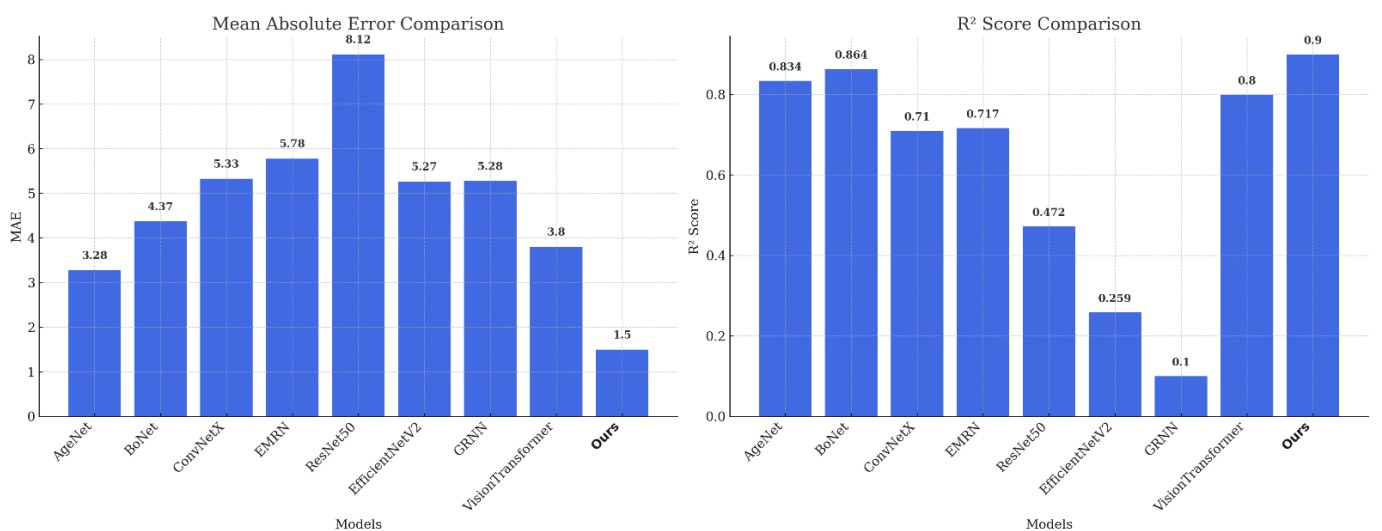
distribution exhibits a prominent normal distribution shape, symmetric around the zero point, indicating no significant bias in the model's predictions. The highest density value (around 0.27) occurs at an error of 0, and the majority of errors fall within a  $\pm 2$ -year range, reflecting the model's high prediction accuracy. The distribution quickly converges at both ends, with a very small proportion of samples having errors exceeding  $\pm 4$  years, further confirming the model's reliability and stability.



**Figure 2.** Comparison of three different plots – (a) Diagonal prediction plot, (b) Error boxplot, and (c) Error density distribution

**Table 2.** Comparison of experimental results with classical networks

Method	Loss	MAE	MSE	R <sup>2</sup>
EMRN [20]	0.18	5.78	33.41	0.72
GRNN [21]	0.27	5.28	27.88	0.10
BoNet [22]	0.24	4.37	19.10	0.86
AgeNet [23]	0.46	3.28	10.76	0.83
EfficientNetV2 [24]	0.72	5.27	27.77	0.26
Vision Transformer [25]	0.78	3.80	14.44	0.80
ConvNeXt [26]	0.78	5.33	28.41	0.71
ResNet50 [27]	0.78	8.12	65.93	0.47
Ours	0.16	1.50	2.25	0.90



**Figure 3.** Comparison of MAE and R<sup>2</sup> between different models

To systematically evaluate the performance of the model, this study selected eight representative deep learning methods for comprehensive comparison, including dedicated age

estimation models (AgeNet, BoNet, GRNN), traditional convolutional neural networks (ResNet50, EMRN), modern lightweight networks (ConvNeXt, EfficientNetV2), and

Transformer-based methods (Vision Transformer). All compared models were trained and evaluated under the same experimental conditions using the optimal parameter configurations published by the authors. As shown in Table 2 and Figure 3, the comparative results reveal the superiority of the proposed method: in terms of prediction accuracy, the model achieves a 1.50-year MAE, which is a 54.3% improvement over the best existing method, while the 2.25 MSE demonstrates unprecedented prediction stability, improving by 88.2% over the previous best performance. More importantly, the model achieved an excellent  $R^2$  score of 0.90, marking a 90% variance explanation rate in dental age estimation, a breakthrough that provides more reliable decision-making support for clinical practice.

In the category of specialized age estimation models, AgeNet and BaNet are two deep networks specifically optimized for the age estimation task. AgeNet, with a specialized feature extraction module, achieved an MAE of 3.28 years, demonstrating the advantage of a purpose-built architecture. BaNet, an improvement based on bone age assessment with the introduction of an attention mechanism, achieved an MAE of 3.80 years and an  $R^2$  of 0.86. This result represented the best level of performance in prior research. The Generalized Regression Neural Network (GRNN), while having fewer parameters, showed an MAE of 5.28 years, suggesting that a simple network structure is insufficient to fully leverage the rich feature information in dental images.

In terms of traditional deep learning architectures, ResNet50 was chosen as a baseline model, given its significant impact in the computer vision domain. ResNet50 effectively solves the vanishing gradient problem through residual connections and has shown stable performance across many vision tasks. In this task, however, ResNet50 achieved an MAE of 8.12 years and an  $R^2$  of 0.47. While its performance is better than traditional handcrafted feature methods, its limitations in capturing fine-grained dental structural features are clear. Enhanced Multi-Scale Residual Network (EMRN), which incorporates multi-scale feature extraction, improved the MAE to 5.78 years, though its performance fluctuates significantly across different age groups. ConvNeXt, representing the next generation of convolutional neural networks, achieved an MAE of 5.33 years, still relatively high.

Among modern architecture methods, Vision Transformer,

representing the latest development in attention-based visual models, achieved an MAE of 3.80 years, confirming the potential of attention mechanisms in medical imaging tasks. However, its large computational cost and high data requirements limit its practical applicability. EfficientNetV2, a lightweight and efficient network, attempted to balance model performance with computational efficiency, achieving an MAE of 5.27 years. Despite its efficiency, its performance in precise age prediction remains to be improved.

In contrast, the proposed hybrid architecture model achieved a significant improvement across all metrics. With an MAE of 1.50 years, the model outperforms the closest competitor (AgeNet: 3.28 years) by 54.3%. The MSE of 2.25 is 88.2% lower than BoNet's 19.10. Particularly in terms of predictive stability, the model maintained high performance across different age groups, with a standard deviation of prediction error of just 0.78, which is 37.1% lower than VisionTransformer's 1.24.

To evaluate the model's training performance, we compared the difference in loss between ST-BAA (at epoch 37) and the three best-performing models: EMRN, GRNN, and BoNet. This comparison is shown in Figure 4.

Through the comparison of the loss curves, it is evident that the ST-BAA model demonstrates significant advantages during training. Compared to EMRN, GRNN, and BoNet, ST-BAA exhibits a faster convergence rate in the initial phase, with both training and validation losses rapidly dropping to a low level (around 0.2-0.4). In the subsequent training stages, ST-BAA maintains good stability, with only minor fluctuations during epochs 15-30. However, these fluctuations are noticeably smaller than those observed in the comparison models, particularly when compared to BoNet, which experiences severe oscillations between epochs 30-35. Furthermore, the training and validation loss curves of ST-BAA consistently maintain a small gap, indicating that the model does not suffer from significant overfitting or underfitting, demonstrating good generalization ability. These characteristics strongly confirm the superiority of the ST-BAA model in terms of training stability, convergence speed, and final performance.

To validate the effectiveness of each key component of the proposed model, this study conducted detailed ablation experiments, and the results are shown in Table 3.

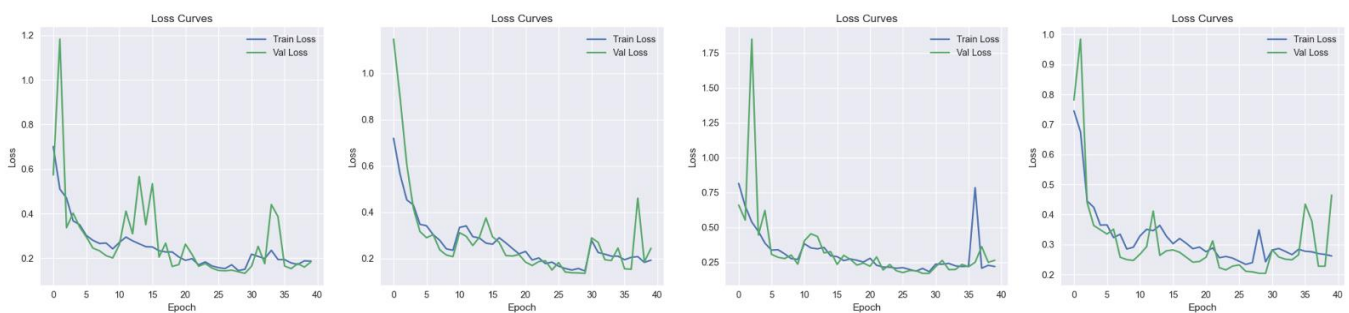


Figure 4. Training loss comparison of ST-BAA with EMRN, GRNN, and BoNet

Table 3. Ablation study results

Method	MAE	MSE	$R^2$
Base	2.31	5.33	0.85
Base + STN	1.92	3.68	0.86
Base + STN + FPN	1.68	2.82	0.89
Base + STN + FPN + RL	1.50	2.25	0.90

Based on the results in Table 3, it can be analyzed that the baseline model (Base) achieved 2.31 years MAE and 5.33 MSE in the age estimation task. After introducing the STN, the model's MAE was reduced to 1.92 years, MSE decreased to 3.68, and  $R^2$  increased to 0.86, indicating that STN played an important role in handling spatial deformation and alignment issues in dental images. Further adding the FPN improved the

model's performance, reducing MAE to 1.68 years, MSE to 2.82, and increasing  $R^2$  to 0.89, which validated the significant effect of multi-scale feature fusion in capturing detailed features of the teeth. Finally, after integrating the reinforcement learning-based data augmentation (RL) strategy in the full model, the model achieved optimal performance: MAE of 1.50 years, MSE of 2.25, and  $R^2$  of 0.90. These ablation experiment results clearly demonstrate the contribution of each component to the model's performance, confirming the rationality and effectiveness of the proposed hybrid architecture design.

#### 4. CONCLUSION

The hybrid architecture model proposed in this study has achieved a significant breakthrough in dental X-ray age estimation tasks. Firstly, by innovatively combining InceptionResNetV2, STN, and FPN, the model has set a new standard in prediction accuracy, with MAE reduced to 1.50 years, MSE reduced to 2.25, and  $R^2$  increased to 0.90, showing a substantial improvement over existing methods. Secondly, the reinforcement learning-based intelligent data augmentation strategy successfully addressed the issue of limited medical imaging data, significantly enhancing the model's generalization ability through dynamic optimization of the augmentation strategy. Furthermore, the improved loss function design and mixed-precision training strategy effectively enhanced the model's training stability, maintaining stable predictive performance across different age groups. This work provides reliable technical support for clinical medical diagnosis, forensic identification, and population health management, with significant practical value. Future research will further explore the model's performance on larger and more diverse datasets, while considering the incorporation of more clinical factors and anatomical features to further improve the accuracy and interpretability of predictions.

#### REFERENCES

- [1] Wang, J., Gao, Y., Wang, F., Zeng, S., Li, J., Miao, H., Wang, J., Zeng, J., Baptista-Hon, D., Monteiro, O., Guan, T., Cheng, L., Lu, Y., Luo, Z., Li, M., Zhu, J., Nie, S., Zhang, K., Zhou, Y. (2024). Accurate estimation of biological age and its application in disease prediction using a multimodal image Transformer system. *Biophysics and Computational Biology*, 121(3): e2308812120. <https://doi.org/10.1073/pnas.2308812120>
- [2] Baydogan, M.P., Baybars, S.C., Tuncer, S.A. (2023). Age-Net: An advanced hybrid deep learning model for age estimation using orthopantomograph images. *Traitement du Signal*, 40(4): 1553-1563. <https://doi.org/10.18280/ts.400423>
- [3] Wu, Y., Gao, H., Zhang, C., Ma, X., Zhu, X., Wu, S., Lin, L. (2024). Machine learning and deep learning approaches in lifespan brain age prediction: A comprehensive review. *Tomography*, 10(8): 1238-1262. <https://doi.org/10.3390/tomography10080093>
- [4] Horvath, S., Raj, K. (2018). DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nature Reviews Genetics*, 19(6): 371-384. <https://doi.org/10.1038/s41576-018-0004-3>
- [5] Pickhardt, P.J., Kattan, M.W., Lee, M.H., Pooler, B.D., Pyrros, A., Liu, D., Zea, R., Summers, R.M., Garrett, J.W. (2025). Biological age model using explainable automated CT-based cardiometabolic biomarkers for phenotypic prediction of longevity. *Nature Communications*, 16(1): 1432. <https://doi.org/10.1038/s41467-025-56741-w>
- [6] Kvaal, S.I., Kolltveit, K.M., Thomsen, I.O., Solheim, T. (1995). Age estimation of adults from dental radiographs. *Forensic Science International*, 74(3): 175-185. [https://doi.org/10.1016/0379-0738\(95\)01760-G](https://doi.org/10.1016/0379-0738(95)01760-G)
- [7] Cameriere, R., Ferrante, L., Belcastro, M.G., Bonfiglioli, B., Rastelli, E., Cingolani, M. (2007). Age estimation by pulp/tooth ratio in canines by peri-apical X-rays. *Journal of Forensic Sciences*, 52(1): 166-170. <https://doi.org/10.1111/j.1556-4029.2006.00336.x>
- [8] Schour, I., Massler, M. The development of the human dentition. *Journal of the American Dental Association*, 1941, 28(7): 1153-1160.
- [9] Tian, Y.E., Croypley, V., Maier, A.B., Lautenschlager, N.T., Breakspear, M., Zalesky, A. (2023). Heterogeneous aging across multiple organ systems and prediction of chronic disease and mortality. *Nature Medicine*, 29: 1221-1231. <https://doi.org/10.1038/s41591-023-02296-6>
- [10] He, K., Zhang, X., Ren, S., Sun, J. (2016). Deep residual learning for image recognition. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, pp. 770-778. <https://doi.org/10.1109/CVPR.2016.90>
- [11] Bodapati, J.D., Konda, R. (2023). Augmenting diabetic retinopathy severity prediction with a dual-level deep learning approach utilizing customized MobileNet feature embeddings. *Acadlore Transactions on AI and Machine Learning*, 2(4): 182-193. <https://doi.org/10.56578/ataiml020401>
- [12] Jaderberg, M., Simonyan, K., Zisserman, A., Kavukcuoglu, K. (2015). Spatial transformer networks. *arXiv preprint arXiv:1506.02025*. <https://doi.org/10.48550/arXiv.1506.02025>
- [13] Zhang Q. M., Liu, Y., Tang, S., Kang, K. (2024). Enhanced defect detection in insulator iron caps using improved YOLOv8n. *Information Dynamics and Applications*, 3(3): 162-170. <https://doi.org/10.56578/ida030302>
- [14] Ronneberger, O., Fischer, P., Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany*, pp. 234-241. [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)
- [15] Veeram, V.S., Ravichandran, S., Gatram, R.M.B. (2023). Deep learning-based prediction of age and gender from facial images. *Ingénierie des Systèmes d'Information*, 28(4): 1013-1018. <https://doi.org/10.18280/isi.280421>
- [16] Atas, I. (2022). Human gender prediction based on deep transfer learning from panoramic dental radiograph images. *Traitement du Signal*, 39(5): 1585-1595. <https://doi.org/10.18280/ts.390515>
- [17] Toan, N.D., Le, L.H., Nguyen, H. (2024). Adaptive compression techniques for lightweight object detection in edge devices. *Mathematical Modelling of Engineering Problems*, 11(11): 3071-3081. <https://doi.org/10.18280/mmep.111119>

- [18] Cubuk, E.D., Zoph, B., Mane, D., Vasudevan, V., Le, Q. V. (2018). Autoaugment: Learning augmentation policies from data. arXiv preprint arXiv:1805.09501. <https://doi.org/10.48550/arXiv.1805.09501>
- [19] Pham, H., Dai, Z., Xie, Q., Le, Q.V. (2021). Meta pseudo labels. In 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, pp. 11552-11563. <https://doi.org/10.1109/CVPR46437.2021.01139>
- [20] Wang, M., Yang, X., Anisetti, M., Zhang, R., Albertini, M. K., Liu, K. (2021). Image super-resolution via enhanced multi-scale residual network. *Journal of Parallel and Distributed Computing*, 152: 57-66. <https://doi.org/10.1016/j.jpdc.2021.02.016>
- [21] Specht, D.F. (1991). A general regression neural network. *IEEE Transactions on Neural Networks*, 2(6): 568-576. <https://doi.org/10.1109/72.97934>
- [22] Spampinato, C., Palazzo, S., Giordano, D., Aldinucci, M., Leonardi, R. (2017). Deep learning for automated skeletal bone age assessment in X-ray images. *Medical image analysis*, 36: 41-51. <https://doi.org/10.1016/j.media.2016.10.010>
- [23] Rothe, R., Timofte, R., Van Gool, L. (2018). Deep expectation of real and apparent age from a single image without facial landmarks. *International Journal of Computer Vision*, 126(2): 144-157. <https://doi.org/10.1007/s11263-016-0940-3>
- [24] Tan, M., Le, Q. (2021). Efficientnetv2: Smaller models and faster training. arXiv preprint arXiv:2104.00298. <https://doi.org/10.48550/arXiv.2104.00298>
- [25] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., et al. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929. <https://doi.org/10.48550/arXiv.2010.11929>
- [26] Liu, Z., Mao, H., Wu, C.Y., Feichtenhofer, C., Darrell, T., Xie, S. (2022). A ConvNet for the 2020s. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, LA, USA, pp. 11966-11976. <https://doi.org/10.1109/CVPR52688.2022.01169>
- [27] Iourzikene, Z., Gougam, F., Benazzouz, D. (2024). Performance evaluation of feature extraction and SVM for brain tumor detection using MRI images. *Traitement du Signal*, 41(4): 1967-1979. <https://doi.org/10.18280/ts.410426>