# Towards Automated Pediatric Dental Development Staging: A Dataset and Model

Peng Wang[1,2], Along He[4], Anli Wang[5,6], Zhenhuan Zhou[1], Xiaohang Guan[5,6(✉)], and Tao Li[1,3(✉)]

[1] College of Cryptology and Cyber Science, Nankai University, Tianjin, China
[2] College of Engineering, Yanbian University, Yanji, China
[3] Haihe Lab of ITAI, Tianjin, China
[4] School of Artificial Intelligence, National Engineering Laboratory for Big Data System Computing Technology, Shenzhen University, Shenzhen, China
[5] Tianjin Stomatological Hospital, Tianjin, China
[6] Tianjin Key Laboratory of Oral and Maxillofacial Function Reconstruction, Tianjin, China
guanxiaohang@aliyun.com,litao@nankai.edu.cn

**Abstract.** Dental development assessment (DDA) is crucial for orthodontic diagnosis and treatment planning. Recent advances in deep learning have shown promising results in dental image analysis tasks. However, the study of dental development staging, particularly in pediatric dental development, remains underexplored. This is primarily attributed to the scarcity of publicly available datasets. In this paper, we present a pediatric Dental Development Staging Dataset(DentalDS). To the best of our knowledge, this is the first publicly available dataset for pediatric DDA. It comprises **2,583** orthopantomogram (OPG) images, with a total of **18,081** annotated teeth. Furthermore, we propose a dental development staging network (DDSNet) designed to address the classification of tooth development stages. In DDSNet, we propose a Region-Instance Cross-Attention (RICA) block and a Multi-Expert Collaborative Classification (MECC) block to enhance the fine-grained feature fusion and classification accuracy of dental development stages. To evaluate the effectiveness of the proposed DDSNet, we conducted experiments on the DentalDS. Our proposed method achieves the state-of-the-art accuracy of 76.3% and an F1-score of 77.1%, outperforming the existing approach method by 1.9% in accuracy and 3.8% in F1-score. To facilitate further research in pediatric orthodontic treatment, code and dataset will be available at `https://github.com/ybupengwang/DDSNet`.

**Keywords:** Dental development staging · Multi-expert collaborative · Cross attention.

## 1 Introduction

The World Health Organization (WHO) considers malocclusion one of the major oral health issues, following caries and periodontal disease. Its prevalence shows significant variation across different countries, sexes, and age groups, with an estimated range of 39% to 93% in the pediatric population [2]. Orthodontic treatment is the primary method for correcting malocclusion. In clinical practice, dental development assessment (DDA) is essential for formulating appropriate treatment plans [15]. Orthodontists assess the corresponding developmental stage of each tooth using orthopantomogram (OPG) images, according to the Demirjian method [6]. This method divides tooth development levels into eight stages. However, determining the dental development stage is a challenging and time-consuming process, heavily relying on the orthodontist's experience. Therefore, it is necessary to develop a computer-aided diagnosis method to assist clinicians in DDA, so as to carry out accurate and reliable treatment planning.

Recently, deep learning models have been extensively utilized in the field of digital dentistry, including dental region segmentation [5,16,10,31], tooth detection and numbering [25,11], teeth reconstruction [20,29] and dental disease detection [30,3]. Deep learning based methods [8,26] have also been applied in DDA, and they have achieved promising results, further demonstrating the significant potential of deep learning approaches in pediatric DDA tasks. However, there are still several limitations when applying these methods to pediatric DDA tasks. **Firstly**, most previous works were trained and evaluated on in-house datasets and there is no publicly available dataset for model development. **Secondly**, the majority of them focused on dental development for all age groups, rather than focused on DDA during the pediatric developmental period. **Finally**, the dental structures in pediatric OPG images are more complex than those in adults, as they include unerupted permanent tooth buds. These tooth buds are very similar in shape, which can easily cause semantic confusion.

To address these issues, we first constructed a dental development staging dataset called DentalDS. It includes 2,583 images with annotations for 18,081 teeth of patients aged between 3 and 15 years. Further, we propose an end-to-end dental development staging network called DDSNet, which consists of three components: tooth localization, feature encoding, and dental development stage classification. For tooth localization, we adopt the traditional DETR[1] model, which leverages the Transformer architecture to directly predict bounding boxes without relying on region proposals. For feature encoding, we introduce the Region-Instance Cross-Attention (RICA) block to capture rich and fine-grained features. For dental development stage classification, we propose a Multi-Expert Collaborative Classification (MECC) block, which enhances classification performance by leveraging the collaboration of multiple experts. The main contributions in this paper can be summarized as follows:

- We create a pediatric dental development assessment dataset called DentalDS, which is the first public dataset specifically for pediatric DDA.
- We propose an end-to-end network called DDSNet for DDA. In the DDSNet, we introduce the RICA and MECC, which enhance the fusion of region and tooth instance features, improving classification accuracy.
- Extensive experiments were conducted on the proposed DentalDS to establish benchmarks for pediatric DDA. The results show that our model outper-

forms previous state-of-the-art (SOTA) networks and can serve as a strong baseline for the pediatric DDA task.

## 2 Dataset

As shown in table 1, compared to mainstream dental datasets, our dataset addresses the limitations of existing datasets in pediatric dental development staging. Notably, the dataset in Ong et al. [19] is not publicly available and excludes cases with orthodontic treatment or apical lesions, whereas ours includes such complexities to better reflect real-world clinical scenarios. The acquisition process of the DentalDS is illustrated in Figure 1.

**Table 1.** Comparison with existing dental datasets. ①, ②, ③ and ④ represent the detection, segmentation, age estimation and dental development staging, respectively.

| Reference | Modality | Image Number | Year | Task | Age | Available |
|---|---|---|---|---|---|---|
| Dental X-ray [27] | X-ray | 520 | 2016 | ① | - | ✓ |
| Vila et al. [26] | OPG | 2289 | 2020 | ③ | 4.5-89.2 | ✗ |
| CTooth+ [4] | CBCT | 168 | 2022 | ② | - | ✓ |
| DENTEX [12] | X-ray | 1005+1571 | 2023 | ① ② | >12 | ✓ |
| Zhang et al. [30] | OPG | 193 | 2023 | ① ② | 2-13 | ✓ |
| Dong et al. [8] | OPG | 673 | 2023 | ① ④ | 3-14 | ✗ |
| Ong et al. [19] | OPG | 5133 | 2024 | ① ④ | 4-16 | ✗ |
| DentalDS(ours) | OPG | 2583 | 2025 | ① ④ | 3-15 | ✓ |

DentalDS was collected from patients who visited the pediatric orthodontics department between January 2021 and May 2024. Due to privacy concerns, we eliminate all personal information except for gender and age. During data collection, radiologists first screen and exclude unqualified images based on the IQSC. After the initial selection, the chosen images will be reviewed by two orthodontists to ensure their quality. Finally, a total of **2,583** images were selected.

The data annotation was divided into two steps: (a) localization of each left permanent mandibular tooth with a bounding box, and (b) annotation of tooth development stages based on the Demirjian method [6]. All images in DentalDS were manually annotated in detail using LabelMe [23] by three dental experts, each with at least ten years of experience. The annotation was carried out in the following steps: Two experts were randomly assigned anonymous images each to complete the image annotation independently. Then, the third expert reviewed the annotation quality. When ambiguities were identified during the review, they voted on the annotation results, and the one with the most votes was the final result. Finally, the dataset was reviewed and verified by all experts to ensure the accuracy of the annotation. A total of **18,081** teeth were annotated. Cohen's kappa showed substantial agreement. This study was approved by the Ethics Committee of Tianjin Stomatological Hospital (Approval No. YPH2024-S-024)

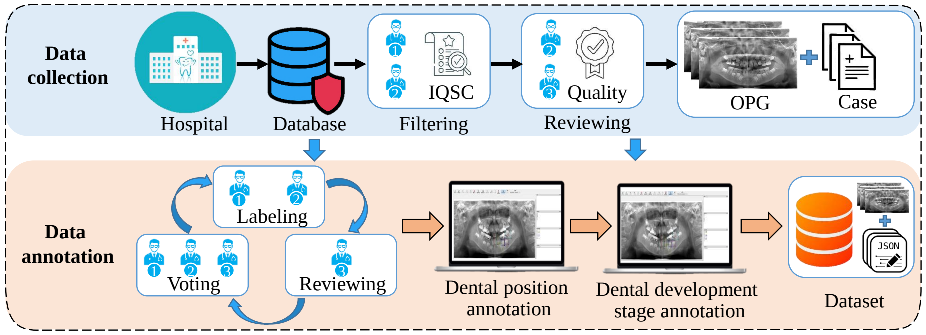



**Fig. 1.** Production process of DentalDS: 1) use the pediatric image quality scoring criteria(IQSC) to manually screen and obtain the target images; 2) Label and review the tooth positions and developmental stages. The dataset was divided into training set (2,048 images) and testing set (535 images).

## 3 Method

We propose an end-to-end network called DDSNet for pediatric DDA, which consists of three components, as illustrated in Fig. 2. In this section, we present each component of the proposed model in detail.

### 3.1 Tooth Localization

We adopt DETR [1] with ResNet-50 pre-trained on ImageNet-1K [7] as the tooth localization model. The input image $I \in \mathbb{R}^{H\times W\times C}$ is fed into the DETR model, where $H, W$ and $C$ represent the height, width, and channel number, respectively. The model then extracts seven bounding boxes, along with an additional region of interest (ROI) box. Seven bounding boxes correspond to the seven tooth instances, ranging from the central incisor to the second molar. The ROI is the smallest region that covers the seven teeth. These bounding boxes are used to crop the tooth instances from the original image. After cropping, each cropped tooth instance is resized to $S\times$ S, denoted as $\{I_t, t = 1, 2, \ldots, 7\}$. The ROI image is resized to $M \times M$, denoted as $I_r$.

### 3.2 Region-Instance Cross-Attention Block

Both local and global features are crucial for classification tasks. CNNs [14,24] are limited by the convolutional kernel, which prevents them from capturing global features. Transformer-based networks [9,13] can model global information and long-range dependencies, but they are unable to capture the local features of the image. To enable the model to focus on the ROI of the image and improve the integration of instance and region features, we propose a RICA block to enhance the dental development stage classification performance by effectively leveraging both region and instance features.

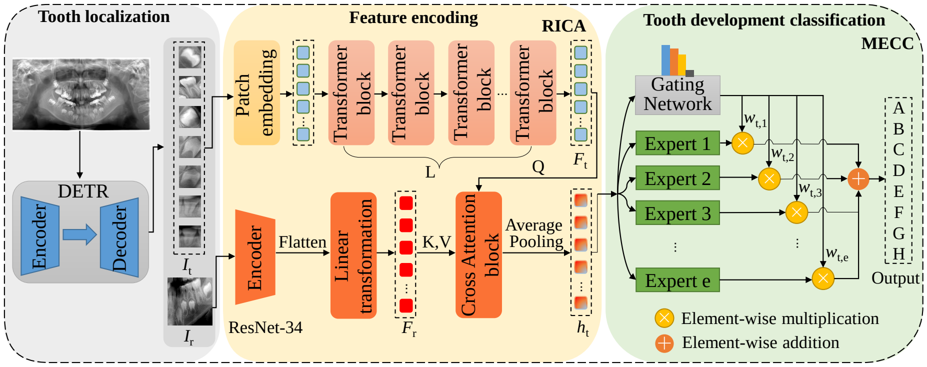



**Fig. 2.** Overall structure of the proposed DDSNet, which consists of three components: tooth localization; feature encoding and tooth development classification.

Each tooth instance $I_t$ is divided into patches and embedded into feature vectors. We use $P_t \in \mathbb{R}^{N\times D}$ to denote the patch embedding with positional encoding of $t$-th tooth, where $N$ is the number of patches, $D$ is the dimension of the embedding. Then, $P_t$ is processed through multiple Transformer blocks, which consist of multi-head self-attention and a feed-forward network [9]. The output of $l$-th Transformer block is $F_t^l \in \mathbb{R}^{N\times D}$. Next, the image $I_r$ is passed through ResNet-34 [14], which extracts the region features. To facilitate subsequent feature fusion, these features are flattened and transformed by a linear layer to obtain $F_r \in \mathbb{R}^{\frac{M}{2}\times D}$. To capture both tooth region and instance features, we employ a cross-attention mechanism to perform the region-instance feature fusion. Specifically, instance features $F_t^l$ are used as queries, regional features $F_r$ are adopted as keys and values. We perform the cross-attention between $F_t^l$ and $F_r$ as follows:

$$\mathrm{Z}_t = \mathrm{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)\cdot V \tag{1}$$

where $Z_t \in \mathbb{R}^{N\times D}$ is the final fused embedding , $Q$ is the instance features $F_t^l$, $K$ and $V$ are the regional features $F_r$, $d_k$ is the dimension $D$. By combining region features extracted through CNN and instance features captured through ViT, the hybrid features are expected to effectively enhance the performance of tooth development staging.

### 3.3 Multi-Expert Collaborative Classification Block

Inspired by V-MOE [22] in image classification tasks, we propose a MECC block. It adopts a gate network $G$ that dynamically selects expert based on the current task $t$ (i.e., each tooth instance classification is a task), while also allowing other experts to assist when necessary. The difference between our MECC block and V-MoE is that MECC does not replace the MLP layer in the Transformer block. Instead, we adopt the multi-expert structure at the classification head, and it fuses the outputs of experts during the classification stage based on the gating weights.

Specifically, the final fused embedding $Z_t$ is average pooled to obtain hidden state $h_t \in \mathbb{R}^D$. Then, the expert weights $w_t$ are formulated as follows:

$$w_t = \mathrm{Softmax}(W_t \cdot h_t) \tag{2}$$

where $W_t \in \mathbb{R}^{E \times D}$ denotes the weight matrix of $G$ for task $t$, and $E$ is the number of experts. Each element $w_{t,e}$ in $w$ represents the gate weight for task $t$ and expert $e$, which reflects the expert's confidence in handling a given task $t$.

Each expert $\mathcal{E}_e$ is a simple fully connected layer. Let $P_t^e \in \mathbb{R}^{cls}$ denote the output of the $\mathcal{E}_e$ for a given input $h_t$, where $cls$ refers to the number of tooth development classification grades. To handle uncertainty in classification, we introduce a threshold $\alpha$ to measure the confidence level of the model's classification decision. If the highest value in $w_t$ falls below $\alpha$, all experts contribute to the decision-making process. Otherwise, the final decision is based on the prediction of the main expert, and the index of the main expert is defined as:

$$e_{\text{main}} = \arg\max_e w_{t,e} \tag{3}$$

The $G$ allows the model to be more flexible in handling uncertainty, especially when the current task is complex. The final output for task $t$ is:

$$y_t = \begin{cases} P_t^{e_{\text{main}}} & \text{if } w_{t,e_{\text{main}}} > \alpha \\ \sum_{e=1}^{N} w_{t,e} \cdot P_t^e & \text{otherwise} \end{cases} \tag{4}$$

## 4 Experiment

### 4.1 Implementation Details

In the experiments, the input images $I$, tooth instances $I_t$ and ROI images $I_r$ are resized to 1200×1200, 224×224 and 512×512, respectively. The number of experts and the threshold $\alpha$ are set to 5 and 0.7, respectively. We use the AdamW optimizer [18] with an initial learning rate of $10^{-4}$, which is decayed using a cosine annealing schedule, and cross-entropy as the loss function. Training is conducted for 400 epochs with a batch size of 16. The parameters of the DETR model are updated during the first 300 epochs. We adopt ViT [9] pre-trained on ImageNet-21K [21] as the backbone for tooth instance feature extraction. During the last 100 epochs, the layer normalization of the ViT, RICA and MECC blocks are updated, while the DETR is frozen. Our framework was implemented using Pytorch, all experiments were conducted on a single Nvidia RTX A6000.

### 4.2 Comparisons With Other Methods

To evaluate the effectiveness of our proposed DDSNet, we conducted experiments with several classification methods, including ResNet-34 [14], EfficientNet

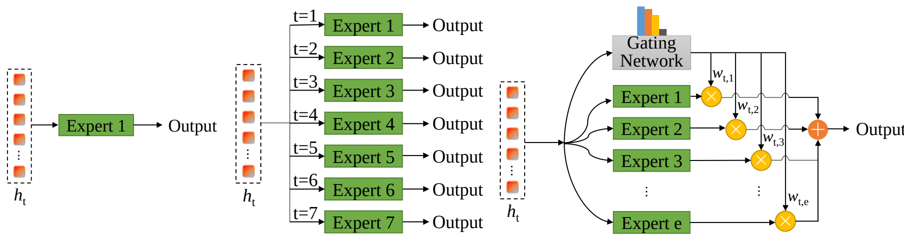
**Table 2.** Performance comparisons with other SOTA methods on DentalDS.

| Method | Acc(%) | Precision(%) | Recall(%) | F1-score(%) |
|---|---|---|---|---|
| Resnet34 [14] | 68.6 | 72.7 | 68.6 | 67.7 |
| EfficientNet [24] | 65.4 | 71.5 | 65.4 | 64.7 |
| Swin Transformer [17] | 73.3 | 75.9 | 73.3 | 72.3 |
| TransFG [13] | 73.9 | 76.8 | 73.9 | 73.4 |
| VDPF[28] | 74.4 | 77.5 | 74.4 | 73.3 |
| DDSNet(ours) | **76.3** | **78.0** | **76.3** | **77.1** |

[24], Swin Transformer [17], TransFG [13], and VDPF [28]. These models are widely used in image classification tasks and have demonstrated strong performance. We evaluate performance using the following metrics: accuracy (Acc), precision, recall, and F1-score. We use `average='weighted'`; thus, the recall is mathematically equivalent to accuracy. The comparison with state-of-the-art methods is presented in Table 2. Our DDSNet achieves 76.3% Acc and 77.1% F1-score on DentalDS. The Acc and F1-score surpass the second best method VDPF [28] by 1.9% and 3.8%, respectively. Our method effectively combines region feature extraction with global context modeling through a cross-attention fusion mechanism. This approach enhances fine-grained feature extraction and improves classification performance, demonstrating its effectiveness in handling the pediatric dental development staging.

### 4.3 Ablation Studies

**Analysis of RICA.** To analyze the RICA block, we conduct ablation experiments in Table 3. When RICA is solely used, there occur performance enhancements of 1.2%. Additionally, we explore the impact of feature region selection and fusion methods on performance based on our model. We compare the performance of models using ROI-based features with those utilizing features from the entire image (EI). As shown in Table 4, ROI features play a more important role, because ROI features enable the model to focus on key details for accurate tooth development assessment while avoiding distractions from irrelevant background. We also investigate different feature fusion methods, including Concatenation Fusion (CF), Cross-Attention Fusion (CA), and Summation Fusion (SF) in Table 4. The results indicate that the CA method outperforms the others in enhancing classification accuracy. The CA method focuses on the most relevant features through the attention mechanism.
**Analysis of MECC.** As shown in Table 3, incorporating the MECC block brings a 1.3% improvement in Acc. When both the MECC and RICA block are combined together, it produces the best Acc of 76.3%, surpassing the baseline by 3.4%. To validate the effectiveness of expert collaboration in the MECC block, we designed three experimental setups: single expert block, independent experts block (seven experts), and MECC block. The block structures are shown in Fig. 3. It can be seen from Table 5 that MECC improves classification accuracy




(a) Single expert (b) Independent experts (c) MECC

**Fig. 3.** Ablation study of contribution of collaboration in the MECC block.

**Table 3.** Ablation results of the proposed blocks, RICA and MECC.

| Method | Acc(%) | Precision(%) | Recall(%) | F1-score(%) |
|---|---|---|---|---|
| DETR+ViT(baseline) | 72.9 | 74.9 | 72.9 | 71.8 |
| +RICA | 74.1 | 76.7 | 74.1 | 73.2 |
| +MECC | 74.2 | 76.4 | 74.2 | 73.4 |
| +RICA+MECC | **76.3** | **78.0** | **76.3** | **77.1** |

**Table 4.** Analysis of different region selection and feature fusion method in RICA on DentalDS. EI, SF, CF, and CA denote the entire image, summation fusion, concatenation fusion, and cross-attention, respectively.

| Region Selection | | Feature Fusion Method | | | Acc(%) | Precision(%) | Recall(%) | F1-score(%) |
|---|---|---|---|---|---|---|---|---|
| EI | ROI | SF | CF | CA | | | | |
| ✓ | - | ✓ | - | - | 73.4 | 75.6 | 73.4 | 72.6 |
| ✓ | - | - | ✓ | - | 74.1 | 76.7 | 74.1 | 73.2 |
| ✓ | - | - | - | ✓ | 74.3 | 76.1 | 74.3 | 73.2 |
| - | ✓ | ✓ | - | - | 74.5 | 76.7 | 74.5 | 73.7 |
| - | ✓ | - | ✓ | - | 75.1 | 77.7 | 75.1 | 74.6 |
| - | ✓ | - | - | ✓ | **76.3** | **78.0** | **76.3** | **77.1** |

by leveraging the strengths of both the main expert and auxiliary experts. It dynamically adjusts expert collaboration based on task complexity and achieves the best performance.

**Table 5.** Comparison of model performance across different expert blocks.

| Block | Acc(%) | Precision(%) | Recall(%) | F1-score(%) |
|---|---|---|---|---|
| Single expert block | 72.9 | 74.9 | 72.9 | 71.8 |
| Independent experts block | 74.1 | 76.7 | 74.1 | 73.2 |
| MECC block | **76.3** | **78.0** | **76.3** | **77.1** |

## 5 Conclution

In this paper, we first collected and annotated an OPG dataset called DentalDS, which is the first publicly available dataset for dental development staging, specifically designed for pediatric dental development assessment. Further, we proposed DDSNet, which leverages two key components: RICA block and MECC block. The combination of these advanced blocks provides DDSNet with a significant ability in handling the challenges of pediatric dental development staging. Through extensive experiments, we demonstrated that DDSNet outperforms other SOTA models, achieving superior performance in classifying dental development stages.

**Acknowledgments.** This work was supported by the Scientific Research Project of the Jilin Provincial Department of Education (No. JJKH20230622KJ), the National Natural Science Foundation of China (No. 62272248), and the Natural Science Foundation of Tianjin (No. 23JCZDJC01010)

**Disclosure of Interests.** The authors have no competing interests to declare that are relevant to the content of this article.

## References

1. Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S.: End-to-end object detection with transformers. Computer Vision–ECCV 2020 pp. 213–229 (2020)
2. Cenzato, N., Nobili, A., Maspero, C.: Prevalence of dental malocclusions in different geographical areas: scoping review. Dentistry Journal **9**(10), 117 (2021)
3. Chen, B., Fu, S., Liu, Y., Pan, J., Lu, G., Zhang, Z.: Cariesxrays: Enhancing caries detection in hospital-scale panoramic dental x-rays via feature pyramid contrastive learning. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 38, pp. 21940–21948 (2024)
4. Cui, W., Wang, Y., Li, Y., Song, D., Zuo, X., Wang, J., Zhang, Y., Zhou, H., Chong, B.s., Zeng, L., et al.: Ctooth+: A large-scale dental cone beam computed tomography dataset and benchmark for tooth volume segmentation. In: MICCAI Workshop on Data Augmentation, Labelling, and Imperfections. pp. 64–73. Springer (2022)
5. Cui, Z., Fang, Y., Mei, L., Zhang, B., Yu, B., Liu, J., Jiang, C., Sun, Y., Ma, L., Huang, J., et al.: A fully automatic ai system for tooth and alveolar bone segmentation from cone-beam ct images. Nature communications **13**(1), 2096 (2022)
6. Demirjian, A., Goldstein, H., Tanner, J.M.: A new system of dental age assessment. Human biology pp. 211–227 (1973)
7. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. pp. 248–255. Ieee (2009)
8. Dong, W., You, M., He, T., Dai, J., Tang, Y., Shi, Y., Guo, J.: An automatic methodology for full dentition maturity staging from opg images using deep learning. Applied Intelligence **53**(23), 29514–29536 (2023)

9. Dosovitskiy, A.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020)
10. Gamal, M., Baraka, M., Torki, M.: Automatic mandibular semantic segmentation of teeth pulp cavity and root canals, and inferior alveolar nerve on pulpy3d dataset. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 14–23. Springer (2024)
11. Hamamci, I.E., Er, S., Simsar, E., Sekuboyina, A., Gundogar, M., Stadlinger, B., Mehl, A., Menze, B.: Diffusion-based hierarchical multi-label object detection to analyze panoramic dental x-rays. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 389–399. Springer (2023)
12. Hamamci, I.E., Er, S., Simsar, E., Yuksel, A.E., Gultekin, S., Ozdemir, S.D., Yang, K., Li, H.B., Pati, S., Stadlinger, B., et al.: Dentex: an abnormal tooth detection with dental enumeration and diagnosis benchmark for panoramic x-rays. arXiv preprint arXiv:2305.19112 (2023)
13. He, J., Chen, J.N., Liu, S., Kortylewski, A., Yang, C., Bai, Y., Wang, C.: Transfg: A transformer architecture for fine-grained recognition. In: Proceedings of the AAAI conference on artificial intelligence. vol. 36, pp. 852–860 (2022)
14. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
15. Kokomoto, K., Kariya, R., Muranaka, A., Okawa, R., Nakano, K., Nozaki, K.: Automatic dental age calculation from panoramic radiographs using deep learning: a two-stage approach with object detection and image classification. BMC Oral Health **24**(1), 143 (2024)
16. Lin, Z., Yang, Y., Huang, A., Shou, Z., Zhang, Q.: Self-training based semi-supervised learning and u-net with denoiser for teeth segmentation in x-ray image. In: MICCAI Challenge on Semi-supervised Tooth Segmentation, pp. 124–132. Springer (2023)
17. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.: Swin transformer: Hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 10012–10022 (2021)
18. Loshchilov, I.: Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101 (2017)
19. Ong, S.H., Kim, H., Song, J.S., Shin, T.J., Hyun, H.K., Jang, K.T., Kim, Y.J.: Fully automated deep learning approach to dental development assessment in panoramic radiographs. BMC Oral Health **24**(1), 426 (2024)
20. Park, S., Kim, S., Song, I.S., Baek, S.J.: 3d teeth reconstruction from panoramic radiographs using neural implicit functions. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 376–386. Springer (2023)
21. Ridnik, T., Ben-Baruch, E., Noy, A., Zelnik-Manor, L.: Imagenet-21k pretraining for the masses. arXiv preprint arXiv:2104.10972 (2021)
22. Riquelme, C., Puigcerver, J., Mustafa, B., Neumann, M., Jenatton, R., Susano Pinto, A., Keysers, D., Houlsby, N.: Scaling vision with sparse mixture of experts. Advances in Neural Information Processing Systems **34**, 8583–8595 (2021)
23. Russell, B.C., Torralba, A., Murphy, K.P., Freeman, W.T.: Labelme: a database and web-based tool for image annotation. International journal of computer vision **77**, 157–173 (2008)

24. Tan, M., Le, Q.: Efficientnet: Rethinking model scaling for convolutional neural networks. In: International conference on machine learning. pp. 6105–6114. PMLR (2019)
25. Tuzoff, D.V., Tuzova, L.N., Bornstein, M.M., Krasnov, A.S., Kharchenko, M.A., Nikolenko, S.I., Sveshnikov, M.M., Bednenko, G.B.: Tooth detection and numbering in panoramic radiographs using convolutional neural networks. Dentomaxillofacial Radiology **48**(4), 20180051 (2019)
26. Vila-Blanco, N., Carreira, M.J., Varas-Quintana, P., Balsa-Castro, C., Tomas, I.: Deep neural networks for chronological age estimation from opg images. IEEE transactions on medical imaging **39**(7), 2374–2384 (2020)
27. Wang, C.W., Huang, C.T., Lee, J.H., Li, C.H., Chang, S.W., Siao, M.J., Lai, T.M., Ibragimov, B., Vrtovec, T., Ronneberger, O., et al.: A benchmark for comparison of dental radiography analysis algorithms. Medical image analysis **31**, 63–76 (2016)
28. Xie, X., Ye, Y., Yang, T., Huang, B., Huang, B., Huang, Y.: Vdpf: Enhancing dvt staging performance using a global-local feature fusion network. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 744–753. Springer (2024)
29. Xu, C., Liu, Z., Liu, Y., Dou, Y., Wu, J., Wang, J., Wang, M., Shen, D., Cui, Z.: Teethdreamer: 3d teeth reconstruction from five intra-oral photographs. In: MICCAI (7) (2024)
30. Zhang, Y., Ye, F., Chen, L., Xu, F., Chen, X., Wu, H., Cao, M., Li, Y., Wang, Y., Huang, X.: Children's dental panoramic radiographs dataset for caries segmentation and dental disease detection. Scientific Data **10**(1), 380 (2023)
31. Zhou, Z., Chen, Y., He, A., Que, X., Wang, K., Yao, R., Li, T.: Nkut: Dataset and benchmark for pediatric mandibular wisdom teeth segmentation. IEEE Journal of Biomedical and Health Informatics (2024)