EasyChair Preprint

№ 11335

A Deep Learning Framework with Pruning RoI Proposal for Dental Caries Detection in Panoramic X-Ray Images

Xizhe Wang, Jing Guo, Peng Zhang, Qilei Chen, Zhang Zhang, Yu Cao, Xinwen Fu and Benyuan Liu

November 19, 2023

A Deep Learning Framework with Pruning RoI Proposal for Dental Caries Detection in Panoramic X-ray Images

Xizhe Wang[1], Jing Guo[2,3], Peng Zhang[4], Qilei Chen[1], Zhang Zhang[1], Yu Cao[1], Xinwen Fu[1], Benyuan Liu[1]

[1] Miner School of Computer & Information Sciences, University of Massachusetts Lowell
[2] Department of Dental General and Emergency, The Affiliated Stomatological Hospital of Nanchang University
[3] The Key Laboratory of Oral Biomedicine, Jiangxi Province Clinical Research Center for Oral Diseases
[4] Department of Oral and Maxillofacial Imaging, The Affiliated Stomatological Hospital of Nanchang University
{xizhe_wang, zhang_zhang}@student.uml.edu, {ndkqgj, ndfskqyy363}@ncu.edu.cn, {qilei_chen1, yu_cao, xinwen_fu, benyuan_liu}@uml.edu

Abstract. Dental caries is a prevalent noncommunicable disease that affects over half of the global population. It can significantly diminish individuals' quality of life by impairing their eating and socializing abilities. Consistent dental check-ups and professional oral healthcare are crucial in preventing dental caries and other oral diseases. Deep learning based object detection provides an efficient approach to assist dentists in identifying and treating dental caries. In this paper, we present a deep learning framework with a lightweight pruning region of interest (P-RoI) proposal specifically designed for detecting dental caries in panoramic dental radiographic images. Moreover, this framework can be enhanced with an auxiliary head for label assignment during the training process. By utilizing the Cascade Mask R-CNN model with a ResNet-101 backbone as the baseline, our modified framework with the P-RoI proposal and auxiliary head achieves a notable 3.85 increase in Average Precision (AP) for the dental caries class within our dental dataset.

Keywords: Object detection · Pruning RoI proposal · Label assignment · Dental caries

1 Introduction

Oral health plays an important role in healthcare. Oral diseases, such as tooth decay, gum disease, and oral cancer [20], can have a seriously negative impact on the quality of life by impairing patients' abilities to eat and socialize. According to the "Global Oral Health Status Report" of World Health Organization (WHO), oral diseases are the most widespread noncommunicable diseases,

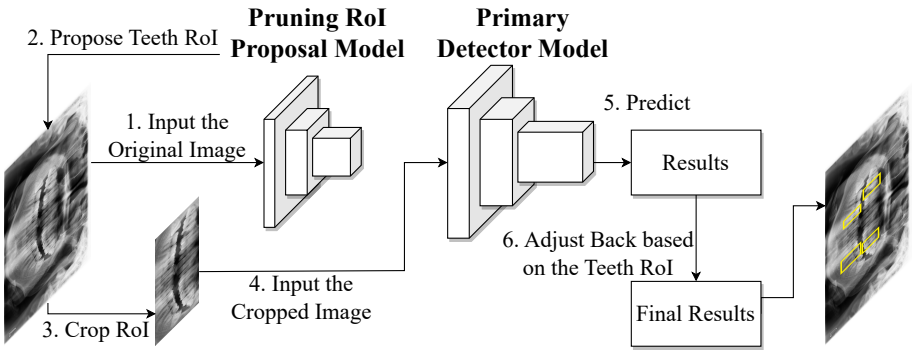


Fig. 1. The workflow of the proposed framework during the inference stage.

affecting approximately half of the population worldwide. Particularly, tooth decay, also referred to as dental caries, is the most prevalent oral disease, affecting more than one third of the world's population throughout individuals' lifetime [1]. Regular dental check-ups and professional oral healthcare are essential preventive measures to combat dental caries and other oral diseases.

The detection of tooth and dental caries [5], [11], [17], [18] in X-ray images has been proven to assist dentists in improving the identification and treatment of tooth decay. The choice of deep learning models and associated techniques plays a crucial role in achieving high performance in this domain. Among the state-of-the-art models for object detection, the You Only Look Once (YOLO) series stands out as a notable choice. These one-stage models have gained significant recognition for their performance. Notably, YOLO-X [3] and YOLOv7 [19] incorporate label assignment methods to improve classification accuracy. On the other hand, when segmentation labels are available in the dataset, the Cascade Mask R-CNN [2], [6] emerges as one of the leading two-stage models for object detection.

In this study, we propose a deep learning framework for the detection of dental caries in panoramic dental X-ray images. Our framework integrates a lightweight pruning region of interest (P-RoI) proposal model, equipped with an enhanced regression loss function, to accurately identify the teeth RoI. The cropped RoI is then used for detecting dental caries. To enhance the label assignment, we introduce an auxiliary head that utilizes the AutoAssign model [22] during the training stage. The workflow of proposed framework is depicted in Figure 1, illustrating the sequential steps during the inference stage. Initially, the P-RoI proposal module identifies the teeth RoI. Subsequently, the image is cropped according to the RoI, which serves as the input to the primary model for dental caries detection. Since the results are predicted within the cropped image, their coordinates are finally adjusted to align with the original image based on the RoI. Through experimentation, our enhanced framework demonstrates remarkable performance, surpassing the baseline model with an increase of 3.85 in the average precision (AP) metric specifically for the dental caries class.

The contributions of this study encompass various aspects, as outlined below:

1. In the P-RoI proposal module, a lightweight model is introduced to ensure computational efficiency. Additionally, an enhanced regression loss function is incorporated, leading to improved accuracy in the P-RoI proposals.
2. During the training phase, the integration of an auxiliary head within the primary head aids in label assignment, particularly in the classification task. Notably, the auxiliary head can be excluded during the inference phase, without introducing additional model complexity in practical applications.
3. An alternative approach is explored, wherein the P-RoI proposal is utilized for cropping the extracted features generated by the primary detection model, instead of cropping the original image. While this exploration did not yield favorable outcomes, an extensive investigation was conducted to assess its potential for the specific task.

2 Related Work

Singh and Raza introduced TeethU^{2}Net [18], an innovative framework specifically designed for tooth saliency detection in panoramic X-ray images. Their work significantly improved upon the original U^2-Net by refining the loss function and training scheme, thereby enhancing the accuracy of tooth boundary segmentation. The proposed framework achieved exceptional performance with an accuracy of 0.9740, specificity of 0.9969, precision of 0.9880, recall of 0.8707, and an F_1-score of 0.9047. Moreover, in their study [17], the researchers also presented an optimized single-stage anchor-free deep learning model capable of detecting teeth and distinguishing their respective treatment types. Their proposed work demonstrated a remarkable AP of 91%.

Haghanifar *et al.* introduced PaXNet [5], a deep learning model designed for dental caries detection in panoramic X-ray images. The PaXNet pipeline involves a series of steps, including jaw extraction, jaw separation, tooth extraction, and classification of the cropped tooth images. According to the evaluation results, PaXNet achieved an accuracy of 86.05%, precision of 89.41%, and recall of 50.67% on their dataset. In addition, Imak *et al.* proposed an approach that employed a multi-input deep convolutional neural network ensemble model with a score fusion technique [15]. The model takes both raw and enhanced periapical images as inputs and fuses the scores in the Softmax layer. Evaluation on a dataset of 340 periapical images demonstrates the effectiveness of the proposed model, achieving an accuracy of 99.13% in diagnosing dental caries.

For the backbone architecture of the P-RoI proposal model, several lightweight models such as SqueezeNet [10], ShuffleNet series [21], [13], and MobileNet series [9], [16], [8] can be considered as potential candidates. The neck architecture commonly employed is the Feature Pyramid Network (FPN) [12]. To effectively handle the RoI proposals, the Region Proposal Network (RPN) [14] is adopted. These models are characterized by their compact and efficient design, making them suitable foundations for the framework due to their small-scale structure and computational efficiency. In terms of the regression loss function

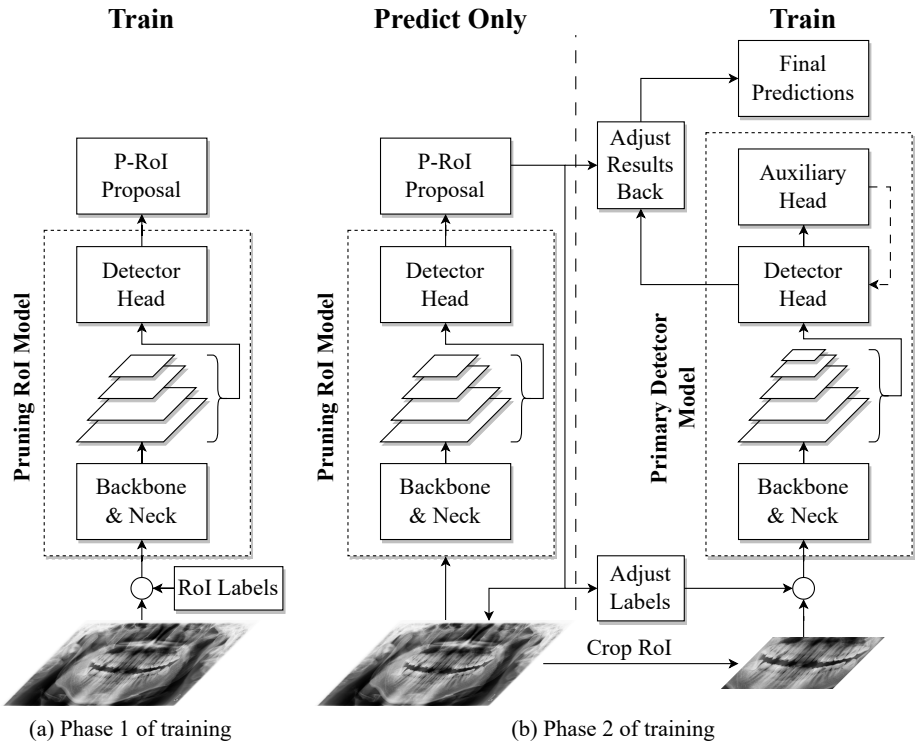


Fig. 2. The two training phases of our framework. "◯" implies data combination.

for the P-RoI proposal, our observations indicate that the Smooth L_1 Loss [4] outperforms the intersection of union (IOU) loss series.

3 Method

3.1 The Overview of Framework

The training of the framework is structured as a two-phase process, as illustrated in Figure 2. The framework includes two detection models: a primary detector model specifically designed for dental caries detection and an additional detection model used for the pruning region of interest (P-RoI) proposal. The teeth RoI ground truth in each image can be obtained by finding the smallest top-y and left-x coordinates and the greatest bottom-y and right-x coordinates of all the bounding box labels. In phase 1, the P-RoI proposal model undergoes independent training for multiple epochs. Notably, the training of the primary detector model is temporarily deactivated during this phase. This precaution is taken to prevent the occurrence of loss exceptions that may arise from simultaneously training both models. The duration of phase 1 is empirically determined through experimentation to ensure effective training of the P-RoI proposal model.

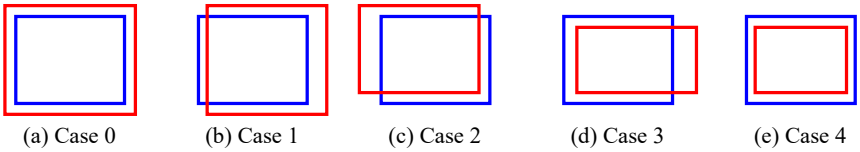


Fig. 3. Five typical IOU cases. The blue bounding boxes denote ground truths. The red bounding boxes denote predictions.

Phase 2 of the training process initiates after reaching a pre-defined number of epochs, determined by the satisfactory performance of the P-RoI proposal model. During this phase, the primary detector model is trained while the training of the P-RoI proposal model is deactivated. Hence, this model only generates the teeth RoI proposals required for cropping the original images for the primary detector model. The localization of predicted results will be translated back based on the P-RoI proposal. The significance of the dashed arrow connecting the auxiliary head to the primary detector head will be elucidated in section 3.3.

3.2 The Pruning RoI Proposal

In the second phase of the training process, the performance of the P-RoI proposal plays a vital role in influencing the final predictions of dental caries. Insufficient P-RoI proposals can lead to incomplete information being cropped, thereby deteriorating the performance of the primary detector model. In our implementation, we integrated the MobileNet-v3-small, FPN, and RPN with a single convolutional layer preceding the proposal head, which encompasses classification and regression branches, to form the P-RoI proposal model.

Figure 3 showcases several common scenarios of IOU cases, depicting the relationships between the bounding boxes of ground truths and predictions. The ground truths are represented by blue bounding boxes, while the predictions are denoted by red ones. The sub-figures demonstrate the following cases:

- (a) "Larger" prediction: This scenario displays a prediction with its top-y and left-x being equal or smaller than those of the ground truth, and bottom-y and right-x being equal or greater than the ground truth's. In other words, the ground truth bounding box is entirely contained within the prediction.
- (b) Insufficient on one side: In this case, only one side of the prediction bounding box is inadequate to achieve the "larger" prediction.
- (c) Insufficient on two sides: Here, two sides of the prediction bounding box are inadequate compared to the "larger" prediction.
- (d) Insufficient on three sides: This case showcases a prediction bounding box with three sides that are insufficient to reach the "larger" prediction.
- (e) Insufficient on all four sides: Finally, this scenario represents that a prediction bounding box is completely contained within the ground truth.

The performance of the P-RoI proposal is significantly influenced by the choice of the regression loss function. Through experimental observations, we

find that the Smooth L_1 Loss function outperforms all variations within the IOU Loss series. While the selection of the Smooth L_1 Loss has yielded promising outcomes, there is still potential for further improvements by fine-tuning specific aspects of the loss function. When modifying the regression loss function, it is essential to adhere to several key principles that consider the characteristics and objectives of the RoI proposal process:

1. If an RoI proposal perfectly matches the ground truth, the loss should be 0.
2. When comparing multiple predictions, the loss associated with the more suitable prediction for detecting the ground truth should be smaller.
3. Assuming all the IOU values in the five cases of Figure 3 are equal, the loss function should demonstrate a relative inclination to accommodate Case 0, the "larger" prediction. This inclination facilitates the preservation of valuable information within the RoI.

The Smooth L_1 Loss function satisfactorily fulfills principles 1 and 2. Thus, our modification primarily focuses on addressing principle 3 by introducing an additional loss penalty specifically for the "non-larger" predictions, which can be represented by cases 1 to 4 in Figure 3. This modification aims to further refine the loss function and improve the overall performance of the P-RoI proposals. The original loss function of the P-RoI proposal model can be succinctly expressed as $\mathcal{L}_{P\text{-}RoI} = \mathcal{L}_{cls} + \mathcal{L}_{reg} = \mathcal{L}_{BCE} + \mathcal{L}_{SmoothL_1}$, where $\mathcal{L}_{cls}$ represents the classification loss implemented using Binary Cross Entropy Loss $\mathcal{L}_{BCE}$, $\mathcal{L}_{reg}$ denotes the regression loss achieved through Smooth L_1 Loss $\mathcal{L}_{SmoothL_1}$.

The modified regression loss function is defined as $\mathcal{L}_{reg} = \lambda_p^{\hat{n}} \cdot \mathcal{L}_{SmoothL_1}$, and the modified loss function of the P-RoI proposal model can be expressed as:

$$\mathcal{L}_{P\text{-}RoI} = \mathcal{L}_{BCE} + \lambda_p^{\hat{n}} \cdot \mathcal{L}_{SmoothL_1} \tag{1}$$

where λ_p is a weight used to fine-tune the regression loss, and $\lambda_p \in (0, +\infty)$, the exponent $\hat{n}$ indicates the number of sides of a prediction that are inadequate to reach the "larger" prediction case, and $\hat{n}$ can be 0, 1, 2, 3, or 4. Consequently, $\lambda_p > 1$ implies an additional penalty for "non-larger" predictions, and $0 < \lambda_p < 1$ indicates an additional penalty for "larger" predictions.

3.3 The Auxiliary Head for Label Assignment

The incorporation of an auxiliary head into the detector architecture plays a pivotal role in optimizing the label assignment process. There are two approaches to implementing the auxiliary head: one involves applying a label assignment algorithm to all of the extracted and cropped features provided by the proposed RoI, while the other involves globally performing the label assignment process on the entire set of features. In our framework, AutoAssign is selected as the auxiliary head to perform a global label assignment process, since its efficiency in distinguishing positive and negative targets with respect to the label assignment.

Figure 4 illustrates an example of the Cascade Mask R-CNN architecture with three cascaded bounding box regression stages, showcasing the integration

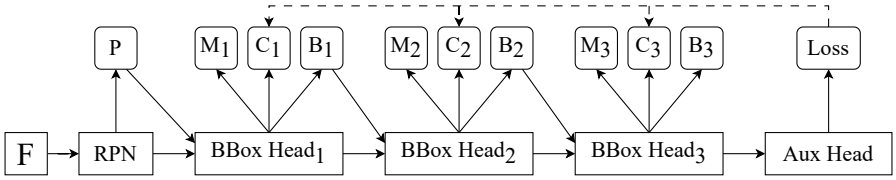


Fig. 4. Cascade Mask R-CNN detector head with an auxiliary head. "F" is the features, "P" denotes the proposals generated by the RPN, "M" represents the mask predictions, "C" signifies the classification results, and "B" indicates the bounding box predictions.

of an auxiliary head at the conclusion of the primary head. The auxiliary head receives the complete set of features as input. The dashed arrow connecting the auxiliary head to the primary head signifies that the AutoAssign module influences the label assigning through exerting its impact on the loss function, rather than directly modifying the classification results within the primary head. Considering the loss function of Cascade Mask R-CNN, the loss function of the primary detector model can be expressed as follows:

$$\mathcal{L}_{Primary} = \mathcal{L}_{RPN} + \sum_{i=1,2,3} \lambda_i \cdot (\mathcal{L}_{cls,i} + \mathcal{L}_{reg,i} + \mathcal{L}_{mask,i}) + \mathcal{L}_{AutoAssign} \quad (2)$$

where $\mathcal{L}_{RPN}$ is the loss of RPN, λ_i, $\mathcal{L}_{cls,i}$, $\mathcal{L}_{reg,i}$, and $\mathcal{L}_{mask,i}$ are the weight, classification loss, regression loss, and segmentation loss of the i-th stage of the cascaded bounding box heads, $\mathcal{L}_{AutoAssign}$ denotes the loss of the auxiliary head.

However, it is important to note that the final predictions are still obtained from the primary head, which is the last stage of the cascaded bounding box heads in the Cascade Mask R-CNN architecture, rather than directly from the auxiliary head itself. This design decision is driven by three main reasons. Firstly, the cascaded bounding box heads exhibit superior performance in terms of bounding box prediction. Secondly, the losses provided by AutoAssign are highly effective in fine-tuning the positive and negative of labels generated by the primary head. Lastly, the auxiliary head is solely utilized during the training stage to provide additional losses and can be safely omitted during the inference stage, thus not increasing the model's scale in practical applications.

4 Experiments

4.1 Dataset

A dataset consisting of 1,008 panoramic dental X-ray images, with each image exhibiting a resolution of 2,918 × 1,435 pixels, was created. The dataset incorporates a total of 31,296 annotations of teeth in both bounding box and segmentation, which can be classified into ten distinct dental categories, as specified in Table 1. Using a random and exclusive approach, we split the images, along with their corresponding labels, into training and test sets in a ratio of

Table 1. Statistical information of our dental dataset.

	Class	Training	Test	Total
1	Normal	19,688	6,638	26,326
2	Caries	2,889	876	3,765
3	Implant	9	1	10
4	Inlay	4	4	8
5	Crown	183	49	232
6	Retainer	21	12	33
7	Pontic	21	5	26
8	Teeth Filling	355	118	473
9	Carie and Tooth Filling	163	55	218
10	Residual Root	146	59	205
	Others (Merged by Class 3 - 10)	902	303	1,205
	Sum (Class 1 - 10)	23,479	7,817	31,296
	Number of Images	756	252	1,008

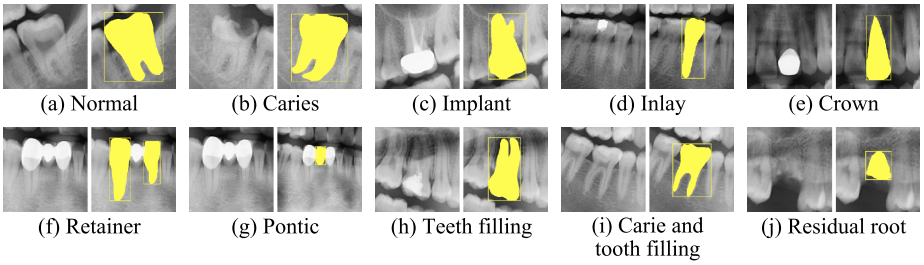


Fig. 5. Tooth samples of ten dental classes.

3:1. Figure 5 illustrates representative samples of teeth within the ten classes. In each sub-figure, the left one exhibits a tooth of a particular class, and the right one shows its related bounding box and segmentation annotations.

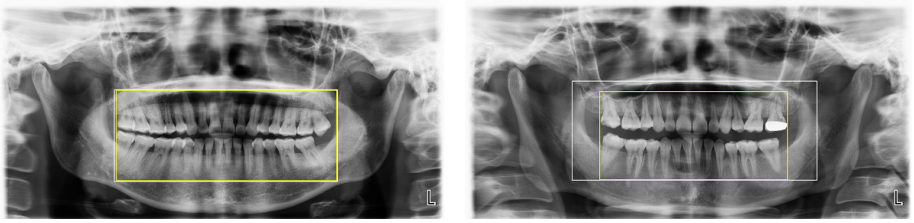
Furthermore, we performed a class reorganization process on our dataset, resulting in four distinct scenarios, based on the usage of categories. In Scenario I (SCE I), we only employed the labels from the "Caries" class. By default, the AP metric refers to the evaluation for the "Caries" class only. Therefore, the mean average precision (mAP) and AP metrics for the test set of SCE I are identical. In the SCE I dataset, where only the "Caries" class is included, the mAP and AP values are identical. In Scenario II (SCE II), we utilized the labels of both the "Normal" and "Caries" classes. In Scenario III (SCE III), we reorganized all labels into a three-class scenario, consisting of "Normal," "Caries," and "Others," which is merged by labels from classes 3 to 10 in Table 1. Finally, in Scenario IV (SCE IV), we employed all labels present in the original ten classes.

4.2 Performance of the Pruning RoI Proposal

We have introduced a modified version of the regression loss function for the P-RoI proposal, as described in equation (1). In our implementation, λ_p is set

Table 2. Evaluation of the P-RoI proposal.

	Modified $\mathcal{L}_{reg}$	Avg IOU	Best IOU	Worst IOU	IOU $\geqslant 0.7$	IOU $\geqslant 0.8$	IOU $\geqslant 0.9$	"Larger" RoI
Train		0.8790	0.9796	0.6489	99.34	94.18	40.78	32.28
Train	✓	0.9054	0.9843	0.6794	99.74	95.11	51.59	50.93
		+0.0264	-	-	+0.40	+0.93	+10.81	+18.65
Test		0.8623	0.9847	0.6517	99.60	92.46	32.94	34.52
Test	✓	0.8843	0.9823	0.6834	99.60	94.05	41.27	49.60
		+0.0220	-	-	+0.00	+1.59	+8.33	+15.08



(a) Best teeth RoI: IOU=0.9823 (b) Worst teeth RoI: IOU=0.6834

Fig. 6. The best and worst cases of the P-RoI proposals in the test set. The RoI ground truths are labeled in yellow color, and the predictions are labeled in white color.

to 1.05. To illustrate the impact of this modified loss function, we present the evaluation results on the training and test sets in Table 2. For the evaluation on the test set, we observe that the average IOU of the teeth RoI proposals increases by 0.0220 to 0.8843 when using the modified loss function. Furthermore, with the modified loss function, 99.60% of the RoI proposals achieve an IOU value greater than 0.7, 94.05% of them achieve an IOU value greater than 0.8, and 41.27% of them achieve an IOU value greater than 0.9. Notably, there is a significant increase of 15.08 percentage points in the number of "larger" prediction cases benefiting from the modified regression loss. Figure 6 shows the best and the worst teeth RoI predictions in the test set with respect to IOU metric.

4.3 Fundamental Experiments

As part of our investigation, we conducted experiments using several contemporary models. In all of these models, we applied multi-scale image input with a ratio range of $[0.8, 1.2]$ and random flip with a probability of 0.5 as data augmentation techniques. All images underwent a keeping-ratio resize operation. Furthermore, for the YOLOv7 series, we employed the Mosaic data augmentation approach, and the input images were padded to achieve a square shape. In terms of the Cascade Mask R-CNN, λ_1, λ_2, and λ_3 are set to 1, 0.5, and 0.25.

A notable finding derived from the presented results in Table 3 is the remarkable performance of the YOLOv7-W model, particularly when trained and

Table 3. Evaluations of selected models. The class-wise AP refers to "Caries" class.

Detector	Back-bone	SCE I	SCE II		SCE III		SCE IV		Input	# of Param.	FLOPs
		mAP	mAP	AP	mAP	AP	mAP	AP			
YOLOv7	L	53.83	66.46	**56.85**	63.41	55.18	30.70	56.26	640^2	36.50M	51.74G
	L	**57.28**	67.56	56.53	64.14	56.44	29.11	54.35	1280^2	36.50M	0.207T
	L	57.65	70.09	60.06	67.64	**60.61**	29.35	59.26	2560^2	36.50M	0.828T
	W	**56.89**	67.67	56.80	65.10	56.49	29.36	56.39	1280^2	79.39M	0.176T
	W	58.73	72.49	63.22	69.37	61.36	34.47	**63.51**	2560^2	79.39M	0.706T
Cascade Mask RCNN	R50	**60.86**	64.34	50.46	62.41	46.43	35.30	46.54	Orig.	75.42M	1.052T
	X50	**58.99**	63.98	49.93	64.10	50.86	38.75	48.16	Orig.	75.12M	1.090T
	R101	**60.94**	64.68	50.81	61.17	43.60	36.58	43.54	Orig.	94.41M	1.360T
	X101	**58.12**	64.91	52.05	63.99	51.06	39.23	49.46	Orig.	94.27M	1.408T

evaluated on the SCE IV dataset, surpassing all other models in terms of class-wise AP for the "Caries" class. In contrast, among the Cascade Mask R-CNN series, all models exhibit the highest class-wise AP for "Caries" when trained and evaluated on the SCE I dataset. The exceptional performance of YOLOv7-W can be attributed to its utilization of Mosaic data augmentation and the incorporation of an auxiliary head for label assignment. Consequently, the Cascade Mask R-CNN model with a ResNet-101 [7] backbone was selected as the baseline for subsequent improvements. The parameter scales and FLOPs denote the model configurations in the inference phase.

4.4 Performance of the Improved Models

Table 4 displays the AP and best F_1-score values for training and evaluation on the SCE I dataset. An additional experiment was conducted, referred to as the case of "baseline$^{\boxplus}$", where the extracted features generated by the primary detection model, instead of the original image, were cropped using the P-RoI proposal. The results indicate a decrease of 3.01 in the "Caries" AP compared to the baseline. Alternatively, the framework proposed in this study, denoted by "baseline$^{\square}$", demonstrates an increase in AP by 2.62 when utilizing the P-RoI proposal to crop the original image. Moreover, incorporating the AutoAssign auxiliary head solely in the baseline model, signified by "baseline*", leads to an AP increase of 2.21. Ultimately, combining both the P-RoI proposal and the AutoAssign auxiliary head into the baseline model, indicated by "baseline$^{\square,*}$", results in a comprehensive improvement of 3.85 in the "Caries" class AP, reaching 64.79. This improvement also surpasses the performance of the YOLOv7-W model by 1.28.

It is worth highlighting that the parameter scales and FLOPs presented in Table 4 specifically pertain to the inference phase. Thus, the models equipped with the auxiliary head maintain the equivalent parameter scale and FLOPs as the models without the auxiliary head. When considering the training phase, the "baseline*" model is characterized by a parameter scale of 99.15M, while the "baseline$^{\square,*}$" model exhibits a parameter scale of 102.58M.

Table 4. Performance of Cascade Mask R-CNN R101 baseline and modified models.

	AP	AP_{50}	AP_{75}	conf.	TP	FP	FN	Pre	Rec	Best F_1	Param.	FLOPs
baseline	60.94	75.04	72.76	0.81	540	138	336	79.65	61.64	69.50	94.41M	1.360T
baseline$^{\boxplus}$	57.93	72.82	69.63	0.85	568	253	308	69.18	64.84	66.94	97.84M	1.410T
	-3.01	*-2.22*	*-3.13*	-	-	-	-	-	-	*-2.56*	-	-
baseline$^{\square}$	63.56	75.87	73.86	0.86	578	192	298	75.06	65.98	70.23	97.84M	1.410T
	+2.62	+0.83	+1.10	-	-	-	-	-	-	+0.73	-	-
baseline*	63.15	76.14	74.12	0.87	566	173	310	76.59	64.61	70.09	94.41M	1.360T
	+2.21	+1.10	+1.35	-	-	-	-	-	-	+0.59	*99.15*M	-
baseline$^{\square,*}$	64.79	77.34	75.52	0.86	589	172	287	77.40	67.24	71.96	97.84M	1.410T
	+3.85	+2.30	+2.76	-	-	-	-	-	-	+2.46	*102.58*M	-

5 Conclusion and Future Work

In this paper, we propose a deep learning framework for the detection of dental caries in panoramic X-ray images. Our framework integrates a lightweight pruning region of interest proposal with an enhanced regression loss function to improve its accuracy. On the head of the primary detector, we introduce an auxiliary head, utilizing the AutoAssign to promote the accuracy through the label assignment during the training phase. Ultimately, our framework achieves a notable increase of 3.85 in the AP with respect to the dental caries class.

There are several aspects that can be further explored and improved upon. 1) The development and refinement of regression loss functions for the P-RoI proposal module could be investigated. 2) It would be beneficial to explore and experiment with additional label assignment methods. 3) The Mosaic data augmentation can be implemented with the mask to apply to the Cascade Mask R-CNN model.

References

1. Global oral health status report. World Health Organization
2. Cai, Z., Vasconcelos, N.: Cascade r-cnn: high quality object detection and instance segmentation. IEEE transactions on pattern analysis and machine intelligence **43**(5), 1483–1498 (2019)
3. Ge, Z., Liu, S., Wang, F., Li, Z., Sun, J.: Yolox: Exceeding yolo series in 2021. arXiv preprint arXiv:2107.08430 (2021)
4. Girshick, R.: Fast r-cnn. In: Proceedings of the IEEE international conference on computer vision. pp. 1440–1448 (2015)
5. Haghanifar, A., Majdabadi, M.M., Ko, S.B.: Paxnet: Dental caries detection in panoramic x-ray using ensemble transfer learning and capsule classifier. arXiv preprint arXiv:2012.13666 (2020)
6. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask r-cnn. In: Proceedings of the IEEE international conference on computer vision. pp. 2961–2969 (2017)
7. 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)

8. Howard, A., Sandler, M., Chu, G., Chen, L.C., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V., et al.: Searching for mobilenetv3. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 1314–1324 (2019)
9. Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H.: Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861 (2017)
10. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 7132–7141 (2018)
11. Imak, A., Celebi, A., Siddique, K., Turkoglu, M., Sengur, A., Salam, I.: Dental caries detection using score-based multi-input deep convolutional neural network. IEEE Access **10**, 18320–18329 (2022)
12. Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S.: Feature pyramid networks for object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2117–2125 (2017)
13. Ma, N., Zhang, X., Zheng, H.T., Sun, J.: Shufflenet v2: Practical guidelines for efficient cnn architecture design. In: Proceedings of the European conference on computer vision (ECCV). pp. 116–131 (2018)
14. Ren, S., He, K., Girshick, R., Sun, J.: Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in neural information processing systems **28** (2015)
15. Saini, D., Jain, R., Thakur, A.: Dental caries early detection using convolutional neural network for tele dentistry. In: 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS). vol. 1, pp. 958–963. IEEE (2021)
16. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.C.: Mobilenetv2: Inverted residuals and linear bottlenecks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 4510–4520 (2018)
17. Singh, N.K., Faisal, M., Hasan, S., Goshwami, G., Raza, K.: Dental treatment type detection in panoramic x-rays using deep learning. In: 22nd International Conference on Intelligent Systems Design and Applications (ISDA 2022) Held December 12-14, 2022-Volume 3. pp. 25–33. Springer (2023)
18. Singh, N.K., Raza, K.: TeethU^{2}Net: A deep learning-based approach for tooth saliency detection in dental panoramic radiographs. In: Neural Information Processing: 29th International Conference, ICONIP 2022, Virtual Event, November 22–26, 2022, Proceedings, Part VII. pp. 224–234. Springer (2023)
19. Wang, C.Y., Bochkovskiy, A., Liao, H.Y.M.: Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7464–7475 (2023)
20. Welikala, R.A., Remagnino, P., Lim, J.H., Chan, C.S., Rajendran, S., Kallarakkal, T.G., Zain, R.B., Jayasinghe, R.D., Rimal, J., Kerr, A.R., et al.: Automated detection and classification of oral lesions using deep learning for early detection of oral cancer. IEEE Access **8**, 132677–132693 (2020)
21. Zhang, X., Zhou, X., Lin, M., Sun, J.: Shufflenet: An extremely efficient convolutional neural network for mobile devices. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 6848–6856 (2018)
22. Zhu, B., Wang, J., Jiang, Z., Zong, F., Liu, S., Li, Z., Sun, J.: Autoassign: Differentiable label assignment for dense object detection. arXiv preprint arXiv:2007.03496 (2020)