Dentelligence System

Mohammad Ismail
32209303001

Jullnar Radwan
32209303002

Rama AlHiyari
32109303022

Supervisor Name: Dr. Wafa Zaal Mohammad Almaaitah

The Project submitted in partial fulfillment of the
requirements for the degree of Bachelor of Science in Artificial Intelligence

Faculty of Artificial Intelligence
Autonomous Systems Department
Al-Balqa Applied University

August 2025

ABSTRACT

This project introduces the Dentelligence System, an automated artificial intelligence solution for the detection and classification of abnormal teeth in panoramic dental X-ray images. As the demand for accurate and efficient dental diagnostics continues to rise, the system aims to reduce diagnostic time and human error through the application of advanced deep learning models. Dentelligence is built to identify abnormal teeth, assign enumeration IDs using the Fédération Dentaire Internationale (FDI) notation, and deliver visual and structured outputs to assist clinicians in real-time interpretation. Rather than relying on a single end-to-end model, the system utilizes separately trained models for each task—detection, segmentation, and disease classification—which are later integrated through a fusion mechanism. This modular approach addresses challenges associated with incomplete or partially labelled datasets and improves the adaptability of the system to various clinical scenarios. Dentelligence is particularly suited to address the limitations of manual panoramic X-ray interpretation, which often suffers from inconsistency due to overlapping structures and image artifacts. The system provides clinicians with annotated images and structured reports that highlight abnormal regions, improve diagnostic precision, and support treatment planning. By adopting FDI notation, it ensures that results are compatible with global dental standards. The system's outputs are exportable in clinician-friendly formats, supporting integration with digital records and telemedicine platforms. The project holds value in educational settings as well, providing dental students with immediate visual feedback and structured diagnostic guidance. Overall, the Dentelligence System represents a significant advancement in dental AI, offering a practical, scalable, and intelligent tool to bridge the gap between traditional diagnostic methods and the capabilities of modern artificial intelligence.

DEDICATION

This project is dedicated to all those who have stood beside us throughout the journey of bringing the Dentelligence System to life—a system envisioned not just as a technical achievement, but as a meaningful contribution to the future of dental diagnostics. To our families, for your boundless love, encouragement, and unwavering belief in our abilities. Your support has been our foundation during times of doubt and your sacrifices have made every accomplishment possible. To our parents, thank you for nurturing our curiosity, for your constant guidance, and for instilling in us the values of perseverance and integrity. Your faith in us carried us through the most challenging moments. And finally, to every dental professional and patient who will one day benefit from systems like Dentelligence. Your needs and experiences fuelled our passion and reminded us why innovation in healthcare truly matters. This work is dedicated to making your path clearer, your diagnoses faster, and your care more precise.

ACKNOWLEDGEMENT

Foremost, we would like to express our heartfelt gratitude to our supervisor, Dr. Wafa Zaal Mohammad Almaaitah, for her unwavering support, expert guidance, and insightful feedback throughout the development of the Dentelligence System. Her expertise in the field of artificial intelligence and dedication to academic excellence have been instrumental in shaping the direction and quality of our work. Her encouragement and confidence in our abilities enabled us to take on challenges and turn them into opportunities for growth. We are also deeply thankful to the faculty members of the Faculty of Artificial Intelligence at Al-Balqa Applied University, especially those in the Autonomous Systems Department, for providing us with the foundational knowledge, academic resources, and a nurturing learning environment that made this project possible. Our sincere appreciation goes to our fellow classmates and friends who stood by us, offering both technical advice and moral support. The collaborative spirit and shared determination of our peers added tremendous value to our learning experience and fostered an environment of continuous motivation. To our families, thank you for your unconditional love, encouragement, and patience during this journey. Your belief in us, especially during the more challenging phases of this project, gave us the strength and clarity to keep pushing forward. We would also like to acknowledge the broader dental and AI research communities whose published work and open-source contributions laid the groundwork for our innovations in automated dental diagnostics. The Dentelligence System draws upon this rich foundation and aims to further advance intelligent healthcare technologies. Finally, we are grateful for the opportunity to contribute to a field that has the potential to impact lives meaningfully. The journey of building Dentelligence was not just a technical endeavor, but a testament to what collaboration, mentorship, and passion can achieve.

TABLE OF CONTENTS

LIST OF TABLES

TABLE NO. **TITLE** **PAGE**

LIST OF FIGURES

LIST OF ABBREVIATIONS

MTDL – Multi-Task Deep Learning

FDI – Fédération Internationale de Dentaire

OPGs – Orthopantomograms

YOLO – You Only Look Once

E-ELAN – Extended Efficient Layer Aggregation Networks

IoU – Intersection over Union

NMS – Non-Maximum Suppression

RPN – Region Proposal Network

RCNN – Region Based Convolutional Neural Networks

DICOM – Digital Imaging and Communications in Medicine

HIS – Hospital Information Systems

EHRs – Electronic Health Records

DINO – DETR with Improved DeNoising Anchor Boxes

WBF – Weighted Boxes Fusion

SE – Squeeze-and-Excitation

CNN – Convolutional Neural Network

AP – Average Precision

AR – Average Recall

FPN – Feature Pyramid Network

F1-score – F1 Performance Score

DPRs – Dental Panoramic Radiographs

GCNet – Grouped Convolutional Network (Grouped Attention Network)

SAM – Segment Anything Model

BB-UNet – Bounding Box U-Net

FUSegNet – Fully Supervised Segmentation Network

CIoU – Complete Intersection over Union

EMCA – Efficient Multi-scale Channel Attention

BiFPN – Bidirectional Feature Pyramid Network

DeTR – Detection Transformer

CHAPTER 1 INTRODUCTION

1.1 Introduction

The advancement of Artificial Intelligence (AI), particularly in the domain of computer vision, has significantly impacted medical diagnostics. As the utilization of medical imaging continues to grow rapidly across modern healthcare systems, the number of images needing interpretation has far outpaced the availability of trained radiologists. Studies indicate that a radiologist may need to interpret one image every 3 – 4 seconds to meet clinical demands, resulting in increased cognitive burden, delayed diagnoses, and higher chances of diagnostic error. One of the most critical areas in dental healthcare is the interpretation of panoramic dental X-ray images, a task that demands expert-level precision and is both time-consuming and resource-intensive. In response, AI-based solutions have become essential to improve both efficiency and diagnostic accuracy. This project aims to automate the analysis of panoramic dental X-ray images to detect abnormal teeth and assign enumeration IDs using state-of-the-art deep learning models. Building upon the success of AI in other medical domains such as CT, MRI, and chest X-rays, our approach integrates detection and segmentation frameworks to overcome the limitations of manual diagnostics, ultimately reducing diagnostic time and enhancing reliability. [1, 31]

1.2 Background of The Project

Panoramic dental X-rays have become a vital diagnostic tool in modern dentistry due to their ability to provide a comprehensive view of a patient's dental structure. These images are commonly used to detect cavities, impacted teeth, and other oral pathologies. Traditionally, dentists rely on manual interpretation of these X-rays, which may lead to human error or inconsistency, particularly when handled by non-specialist practitioners. With the growing demand for accurate and efficient diagnosis, artificial intelligence (AI) techniques have been introduced to assist in automating the interpretation process. Over the past few years, many AI models have been developed for individual dental tasks such as tooth localization, numbering, or identifying specific conditions. Nevertheless, these models

are typically limited in scope, requiring large, fully labelled datasets and often lack the ability to integrate multiple diagnostic steps into one unified system. We developed automated methods to detect abnormal teeth and generate corresponding enumeration labels using the Fédération Internationale de Dentaire (FDI) notation. The system is designed to detect abnormal teeth while simultaneously assigning tooth numbers and associated diagnoses, even when trained on partially labelled data. This approach represents a significant step forward in the automation and intelligence of dental radiographic analysis. [2 - 5]

1.2.1 Sub section of Background – AI in Dentistry

Artificial Intelligence (AI) has increasingly become a transformative technology in the field of dentistry. Through the use of computer vision and deep learning, AI models are now capable of analyzing dental images with a high degree of accuracy. These technologies are being applied in various dental applications, including tooth segmentation, caries detection, periodontal disease classification, and treatment planning. One of the most promising applications of AI is in the interpretation of dental radiographs, particularly panoramic X-rays. By using annotated datasets and powerful deep learning architectures, AI systems can learn to detect patterns and anomalies that may not be easily visible to the human eye. This can significantly reduce diagnostic time and improve consistency in clinical decisions. As AI continues to evolve, it is expected to play a larger role in assisting dentists and improving the overall quality of oral healthcare. [2, 6]

1.2.2 Second Subsection– Data set

Automated analysis of dental radiographs using deep learning techniques has gained significant attention in recent years. The Dentex: Grand Challenge dataset, hosted on the Grand-Challenge.org platform, provides a benchmark dataset specifically designed manual interpretation of panoramic X-rays is a time-consuming and expertise-dependent process, which is prone to variability and human error. It offers a rich collection of panoramic dental X-ray images, annotated with bounding boxes, segmentation masks, and tooth classification labels, facilitating the development and evaluation of AI-based dental image analysis

systems. The Dentex dataset is particularly valuable due to its diversity and high-quality annotations, which encompass various dental structures and conditions. It serves as a reliable foundation for building multi-task models that can perform object detection, semantic segmentation, and classification tasks on dental X-rays. By leveraging this dataset, researchers can develop robust and accurate AI systems that enhance diagnostic workflows, reduce interpretation time, and support clinical decision-making in dental practice. [7]

1.2.3 Third Subsection -Fédération Internationale de Dentaire (FDI) notation

The FDI World Dental Federation notation system (ISO 3950) is a standardized method used worldwide to number and identify teeth. It is based on a two-digit coding system. The first digit indicates the quadrant of the mouth where the tooth is located:

1= Upper right (permanent teeth)

2 = Upper left (permanent teeth)

3 = Lower left (permanent teeth)

4 = Lower right (permanent teeth)

The second digit represents the position of the tooth within that quadrant, starting from the midline (central incisors) and moving backward towards the molars.

Permanent teeth are numbered 1–8 in each quadrant.

Primary teeth (baby teeth) are numbered 1–5 in each quadrant. [8]

For example:

Tooth 11 refers to the upper right central incisor.

Tooth 36 refers to the first molar in the lower left quadrant.

This system provides a clear and efficient way to communicate dental information across different languages and healthcare systems. It ensures that dental records are consistent, precise, and easy to interpret globally, benefiting clinical practice, education, and research.

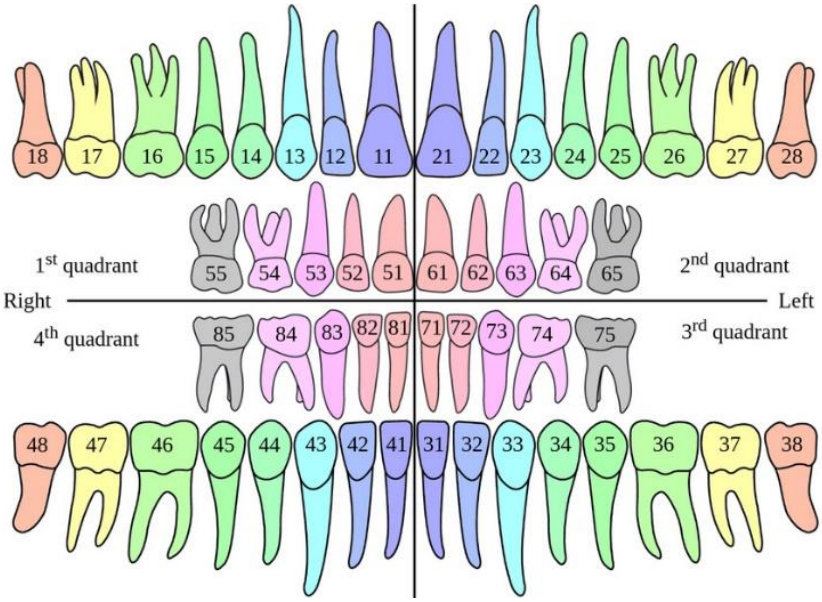

Figure 1: (FDI) notation

1.3 Problem Statement

Despite the growing adoption of artificial intelligence in dental diagnostics, most existing models are designed for specific and isolated tasks, such as tooth detection or diagnosis of a particular condition. There is a lack of comprehensive, end-to-end systems that can perform multiple tasks—such as tooth identification, numbering, and diagnosis. Moreover, a major limitation in the development of such systems is the reliance on large-scale, fully annotated datasets. In real clinical settings, obtaining detailed annotations for every image is often impractical due to time constraints and the need for expert input. There is a need for an automated system that can accurately detect diseased teeth and assign proper enumeration IDs to assist dental professionals in diagnosis and treatment planning. [3, 4]

1.4 Project Objectives

1. To develop an automated model that detects abnormal teeth from panoramic dental X-ray images.
2. To assign enumeration IDs to teeth based on the FDI notation.
3. To integrate detection and segmentation models into a unified deep learning pipeline.
4. To utilize pre-trained models (e.g., YOLOv8, DINO, U-Net) for efficient training and improved accuracy.
5. To evaluate the system using performance metrics such as Average Precision (AP) and F1-score.

1.5 Significance of The Project

This project addresses a critical gap in dental radiograph analysis by providing an end-to-end AI-based solution capable of detecting and diagnosing abnormal teeth with minimal reliance on fully labelled data. The proposed framework has the potential to significantly reduce the workload of dental professionals by automating the time-consuming task of interpreting panoramic X-rays. Clinicians can benefit from faster and more accurate diagnoses, especially in environments with high patient flow or limited access to dental radiology specialists. Moreover, the system offers consistent and objective evaluations, reducing the likelihood of human error or oversight. In addition, the model may serve as an educational tool for dental students by offering real-time visual feedback on dental pathologies and classifications.

1.6 Project Organization

This report is organized into five main chapters:

Chapter 1: Introduction Provides an overview of the project, including its background, problem statement, objectives, significance, and structure.

Chapter 2: Literature Review Reviews existing studies and technologies related to AI-based dental diagnosis, object detection models, it highlights the limitations of current solutions and positions the proposed method in the context of previous work.

Chapter 3: Research Methodology and Design Describes the overall system architecture, data preparation, the diffusion-based detection model, it also outlines the functional and non-functional requirements of the system.

Chapter 4: Results and Discussion Presents the experimental results, performance comparisons with baseline models, and analysis of the effectiveness of bounding box manipulation and multi-label classification.

Chapter 5: Future Work and Conclusions Summarizes the key findings of the project, discusses its limitations, and suggests directions for future improvements and applications

CHAPTER 2 LITERATURE REVIEW

2.1 Introduction

This chapter reviews the relevant literature in the domain of dental image analysis, particularly focusing on techniques for classification, object detection, and semantic segmentation of dental X-ray images. As the proposed project "Dentelligence" applies a multi-task learning approach to combine these three tasks, this review will also cover the key advancements in multi-task learning and its application in medical imaging. This chapter highlights recent methodologies, architectures, and datasets used, along with identifying research gaps that this project aims to address.

2.2 Related Work

In 2023, Azizi et al. conducted a systematic review titled *"Artificial Intelligence Techniques in Medical Imaging"*, aiming to investigate AI-based approaches in disease classification and segmentation using medical images. The study explored a wide range of imaging modalities such as X-ray, CT, MRI, PET, ultrasound, and optical imaging. It emphasized the role of deep learning (DL), machine learning (ML), and hybrid models in advancing diagnostic accuracy. Notably, hybrid techniques integrating ML and DL showed promising outcomes across domains like neurology, oncology, and cardiology. The research also addressed major challenges, including limited annotated data, lack of standardization in imaging protocols, and generalization of models, while suggesting future improvements using synthetic data and transfer learning.

In 2015, Ronneberger et al. introduced the U-Net architecture in their work *"U-Net: Convolutional Networks for Biomedical Image Segmentation"*, with a focus on achieving high-accuracy segmentation from limited annotated biomedical data. The architecture features a symmetric U-shaped design composed of a contracting path for context capture and

an expanding path for precise localization. The authors utilized strong data augmentation techniques, such as elastic deformations, to train the network effectively on small datasets. U-Net demonstrated exceptional performance by winning the ISBI cell tracking challenge and outperforming prior methods in neuronal structure segmentation within EM images. The model achieved segmentation results with high intersection-over-union (IoU) scores, proving its robustness and efficiency across various biomedical imaging tasks.

In 2020, Schwendicke et al. published a review titled *"Artificial Intelligence in Dentistry: Chances and Challenges"*, focusing on the potential and limitations of AI in dental diagnostics and treatment. The article discusses how machine learning (ML), especially deep learning models, is being applied to image analysis, decision-making, and treatment planning in dentistry. Despite its potential to increase efficiency, personalize care, and reduce costs, the paper notes that AI adoption in clinical dental practice remains limited due to data accessibility issues, lack of standardized protocols, and concerns over trust and transparency. The authors advocate for federated learning, explainable AI, and ethical standards to ensure privacy, trustworthiness, and the practical integration of AI technologies in future dental care.

In 2019, Tuzoff et al. presented a study titled *"Tooth Detection and Numbering in Panoramic Radiographs Using Convolutional Neural Networks"*, aiming to automate tooth identification and labeling in dental X-rays. The system was trained on 1,352 panoramic radiographs using the Faster R-CNN architecture for tooth detection and the VGG-16 model for tooth numbering based on the FDI notation. The dataset was annotated by expert radiologists and evaluated using 222 additional test images. The system achieved detection sensitivity and precision of 0.9941 and 0.9945, respectively, and numbering sensitivity and specificity of 0.9800 and 0.9994. Performance was comparable to expert-level accuracy, and the approach demonstrated the potential for use in clinical dental charting and decision support.

In 2021, Haghanifar et al. proposed *"PaXNet: Dental Caries Detection in Panoramic X-ray Using Ensemble Transfer Learning and Capsule Classifier"*, introducing an end-to-end deep learning system for automatic caries detection using panoramic dental images. The model employs genetic algorithms for tooth isolation and uses a hybrid feature extraction pipeline composed of InceptionNet, CheXNet, a custom encoder, and CNN layers, followed by a capsule network classifier. The system was trained and evaluated on a dataset of 470

images, including 11,769 isolated tooth segments, and achieved an overall accuracy of 86.05% and a precision of 89.41%. Notably, the system's recall was 90.52% for severe caries and 69.44% for mild cases. The research demonstrated the effectiveness of capsule networks for geometrical feature recognition and provided a promising foundation for AI-assisted caries diagnosis using lower-quality panoramic radiographs.

In 2022, Almalki and Latecki introduced a novel approach in their paper *"Self-Supervised Learning with Masked Image Modeling for Teeth Numbering, Detection of Dental Restorations, and Instance Segmentation in Dental Panoramic Radiographs"*. This study employed SimMIM and UM-MAE self-supervised learning methods to pre-train Swin Transformers on a limited dataset of panoramic dental X-rays. Their goal was to enhance the performance of teeth detection, numbering, and restoration identification using Cascade Mask R-CNN for segmentation tasks. They also augmented and corrected the existing DNS dataset, creating the TNDRS dataset, which included refined annotations for teeth and restorations. SimMIM achieved the highest accuracy, with 90.4% AP for tooth detection and 88.9% for segmentation, outperforming both supervised and randomly initialized models. This research demonstrated the effectiveness of masked image modeling and transformer-based architectures in overcoming the challenges of data scarcity and class imbalance in dental image analysis.

In 2022, Helli and Hamamci proposed a U-Net-based approach for *"Tooth Instance Segmentation on Panoramic Dental Radiographs Using U-Nets and Morphological Processing"*, focusing on precise instance-level tooth segmentation. The method combined a standard U-Net convolutional neural network with a tailored post-processing pipeline involving grayscale morphological operations and filtering to separate touching teeth. Training was performed on a relatively small dataset of 105 annotated panoramic X-ray images. Despite limited data, the method achieved a Dice score of 95.4% and significantly reduced the mean error in tooth count to 6.15%, compared to 26.81% without post-processing. The results marked one of the highest segmentation and counting accuracies in the literature, showcasing the power of morphological refinement in conjunction with deep learning for dental image analysis.

In 2023, Juyal et al. conducted a study titled *"Dental Caries Detection Using Faster R-CNN and YOLO V3"*, exploring deep learning-based object detection models to identify

dental caries without relying on X-ray imaging. The research used a dataset of approximately 300 images, with 80% for training and 20% for testing. The models evaluated include YOLO V3 and Faster R-CNN, chosen for their balance between accuracy and speed. Faster R-CNN, with a VGG-16 backbone and Region Proposal Network (RPN), achieved an accuracy of 80% and precision of 78%, outperforming YOLO V3, which reached 75% accuracy and 74% precision. The study suggests the implementation of Faster R-CNN in a hardware-integrated diagnostic device that captures intraoral images for real-time caries detection, presenting a radiation-free alternative to traditional diagnostics.

In 2023, Dhar et al. proposed a novel framework in their paper *"A Deep Learning Approach to Teeth Segmentation and Orientation from Panoramic X-rays"*, aiming to enhance precision in tooth segmentation and orientation detection for dental diagnostics. The study modified the FUSegNet architecture with grid-based attention gates and a P-scSE module and introduced Principal Component Analysis (PCA) for generating oriented bounding boxes (OBBs). Evaluated on the DNS dataset of 543 panoramic images, their model achieved top performance with a Dice Similarity Coefficient of 90.37%, IoU of 82.43%, and Rotated IoU (RIoU) of 82.82%. The model showed strong performance in segmenting individual teeth, especially incisors and canines, and demonstrated potential in identifying missing teeth without false positives. Their contributions provide a robust foundation for precise treatment planning and digital dental modeling.

In 2024, Mărghinean et al. introduced *"CariSeg: Teeth and Carious Lesions Segmentation in Panoramic X-ray Images Using a Networks' Ensemble"*, presenting a two-stage deep learning pipeline for automatic dental diagnosis. The proposed system first segments teeth using a U-Net model and then identifies carious lesions using an ensemble of three architectures: U-Net, Feature Pyramid Network (FPN), and DeeplabV3. The study utilized the Tufts Dental Database and a private Qom dataset for teeth segmentation, and the Clujana dataset of 108 annotated images for caries segmentation. Data augmentation and ensemble voting strategies were applied to enhance generalization. CariSeg achieved 99.42% accuracy and a mean Dice score of 68.2% in caries detection, outperforming individual models. This work highlights the advantages of ensemble learning in improving robustness, recall, and accuracy in complex medical image segmentation tasks.

In 2024, Adnan et al. developed an AI-based smartphone application in their study *"Developing an AI-Based Application for Caries Index Detection on Intraoral Photographs"*, targeting real-time dental decay detection using deep learning. The model was trained on 7,465 intraoral images using the YOLOv5s algorithm, with annotations verified by senior dentists. A Detection Transformer (DeTR) model was also evaluated but underperformed compared to YOLOv5s. The YOLOv5s model achieved 90.7% precision, 85.6% sensitivity, and an F1 score of 88.0%, outperforming junior dentists whose F1 score was 72.4%. Explainable AI (XAI) via EigenCAM was applied to visualize the model's decision-making. The trained model was successfully integrated into a deployable mobile application, offering a promising tool for large-scale dental health screening, particularly in low-resource settings.

In 2024, Zhicheng et al. proposed a model for *"Deep Learning-Based Detection of Impacted Teeth on Panoramic Radiographs"*, focusing on identifying impacted teeth through refined segmentation techniques. The study utilized 1,016 panoramic dental X-ray images and fine-tuned the MedSAM model—a medical adaptation of the Segment Anything Model (SAM)—to detect and segment impacted teeth. Using a centroid-based approach to guide zero-shot segmentation, the system achieved an accuracy of 86.73%, an F1-score of 0.5350, and an IoU of 0.3652. Among the various SAM family models tested, MedSAM outperformed others in both segmentation quality and reliability. The study emphasized the effectiveness of combining advanced segmentation architectures with domain-specific fine-tuning, offering a promising method for improving diagnostic support in oral radiology.

In 2024, Beser et al. introduced a *"YOLO-V5-Based Deep Learning Approach for Tooth Detection and Segmentation on Pediatric Panoramic Radiographs in Mixed Dentition"*, addressing the challenges of identifying both deciduous and permanent teeth in children. The study used a large dataset of 3,854 panoramic images annotated by expert clinicians and trained YOLOv5x models separately for detection and segmentation tasks. The detection model achieved 99% precision, sensitivity, and F1-score with a 0.98 mAP, while the segmentation model recorded 98% across all key performance metrics. The model proved especially robust in handling overlapping and superimposed dental structures common in pediatric radiographs. This research presents a highly effective automated system for pediatric dental diagnostics and supports early detection of conditions in mixed dentition cases.

In 2024, Özçelik et al. introduced a high-precision model in their work *"Enhanced Panoramic Radiograph-Based Tooth Segmentation and Identification Using an Attention Gate-Based Encoder–Decoder Network"*. The study proposed the SE-IB-ED architecture, which incorporates InceptionV3 for encoding and a custom decoder enhanced with squeeze-and-excitation attention blocks. A dataset of 313 panoramic dental images annotated using the FDI numbering system was used, leveraging automatic labeling with the Segment Anything Model (SAM) followed by manual correction. The model achieved a mean F1-score of 92.66%, mean IoU of 86.39%, and accuracy of 99.93% across 32 teeth classes, outperforming several state-of-the-art models including Trans-U-Net and Attention U-Net. The use of sigmoid activation allowed effective segmentation of overlapping teeth, proving superior to SoftMax-based methods. This research demonstrates a powerful, attention-enhanced segmentation architecture well-suited for clinical-grade dental image analysis.

In 2024, Kurt et al. presented a study titled *"Evaluation of Tooth Development Stages with Deep Learning-Based Artificial Intelligence Algorithm"*, which aimed to classify the developmental stages of mandibular teeth in children aged 5–14 using panoramic radiographs. The study applied the YOLOv5 architecture, trained on 1,500 high-quality pediatric dental X-rays labeled according to the Demirjian method, a gold standard in dental age estimation. After training and validation, the model achieved a sensitivity of 0.99, a precision of 0.73, and an F1-score of 0.84 across 46 developmental stage classes. This work marks one of the first uses of YOLOv5 for tooth development classification and supports its effectiveness in age estimation and treatment planning in pediatric dentistry.

In 2025, Yaxin and Hargreaves presented a study titled *"Dental Teeth X-Ray Image Classification Using AI"*, which explored the development of a deep learning-based clinical decision support system for detecting and classifying dental conditions from X-ray radiographs. The authors trained a YOLOv5 model on 340 manually annotated adult dental X-rays labeled across five categories: normal, missing, implant, root, and erupting. Preprocessing included contrast enhancement and mosaic augmentation, while model performance was optimized through systematic hyperparameter tuning, including learning rate, batch size, and optimizer comparison. The final model achieved a precision of 89.3%, recall of 77.3%, F1-score of 82%, and mAP of 80.9%. The study also included a comparative analysis of IoU loss variants, concluding that Complete IoU (CIoU) offered the best

performance for bounding box regression. This research highlights the potential of YOLOv5 and rigorous optimization strategies in enhancing real-time dental diagnostics.

In 2025, Yavsan et al. presented a pilot study titled *"Diagnosis of Approximal Caries in Children with Convolutional Neural Networks-Based Detection Algorithms on Radiographs"*, aiming to detect approximal caries in pediatric periapical radiographs using CNN-based deep learning models. A custom dataset of 415 periapical X-rays from children aged 5–12 was created, labeled, and augmented, then split into 80% training and 20% testing sets. Thirteen YOLO-based architectures were compared, with YOLOv5s achieving the best results—90.8% mAP, 91.2% precision, 89.3% recall, and a 90.24% F1-score after 300 epochs. The system even detected five cases of caries missed during expert annotation. The study emphasized YOLOv5s's strong potential for early and accurate detection of pediatric caries and highlighted its clinical utility for reducing misdiagnoses and supporting dental decision-making in children's oral healthcare.

In 2025, Hua et al. introduced *"YOLO-DentSeg: A Lightweight Real-Time Model for Accurate Detection and Segmentation of Oral Diseases in Panoramic Radiographs"*, focusing on developing an efficient deep learning system for real-time dental pathology analysis on low-resource devices. The model, built upon an enhanced YOLOv8n-seg framework, incorporates C2f-Faster for reduced complexity, BiFPN for improved multi-scale feature fusion, EMCA for attention optimization, and Powerful-IOU loss for precise bounding box regression. Trained on a dataset of 2,720 annotated images combining DENTEX and clinical sources, the model achieved 87.2% mAP (Box), 85.5% mAP (Seg), and 90.3 FPS, while reducing parameters, model size, and computation by 44.3%, 41.5%, and 17.5% respectively compared to YOLOv8n-seg. This work presents a robust, fast, and lightweight system suitable for clinical deployment in digital dentistry.

In 2025, Zhong et al. developed *"Automatic X-ray Teeth Segmentation with Grouped Attention"*, presenting a novel network architecture—GCNet—for precise dental segmentation from noisy, low-contrast X-ray images. The model integrates Grouped Global Attention (GGA) and Cross-Layer Fusion (CLF) modules within a U-shaped architecture, using grouped uni-directional attention and feature fusion to capture texture, contour, and semantic features. Tested on a dataset of 3,187 annotated panoramic radiographs, GCNet achieved 97.42% accuracy, 87.61% mIoU, and a Dice score of 93.38%, outperforming

models like Teeth U-Net and GT-U-Net. Its dual-output decoder further enhances segmentation by generating edge predictions. The approach demonstrated high generalizability and robustness across both adult and pediatric dental images, offering a reliable tool for clinical diagnostics in dentistry.

In 2025, Ayhan et al. conducted a comparative study titled *"Evaluation of Caries Detection on Bitewing Radiographs: A Comparative Analysis of the Improved Deep Learning Model and Dentist Performance"*. The research developed and evaluated an enhanced YOLOv9c-Faster model, optimized with a FasterNet backbone for lightweight, efficient detection of enamel and dentin caries. A dataset of 2,150 bitewing radiographs with over 8,000 annotated caries cases was used, and performance was benchmarked against six clinical dentists. The YOLOv9c-Faster model achieved a recall of 72.7%, precision of 65.1%, F1-score of 68.7%, and a Youden index of 0.453—outperforming all dentists, especially in recall and F1-score. The model showed high success in both enamel and dentin lesion detection, with faster processing times and fewer false negatives. These results demonstrate the model's strong potential as a clinical decision-support tool for caries diagnosis, especially in high-volume or resource-limited settings.

In 2025, Ovi et al. proposed *"Enhanced Pediatric Dental Segmentation Using a Custom SegUNet with VGG19 Backbone on Panoramic Radiographs"*, targeting accurate segmentation of pediatric dental structures. The study introduced a custom SegUNet model integrated with a VGG19 backbone, evaluated on the Children's Dental Panoramic Radiographs dataset comprising 323 annotated images. Compared to several benchmark models—including U-Net, PSPNet, and DeeplabV3+—the proposed model achieved state-of-the-art performance with 97.5% accuracy, a Dice coefficient of 92.5%, and IoU of 91.5%. Key improvements stemmed from enhanced feature extraction, BatchNormalization, and skip connections. The model also showed excellent generalization across diverse pediatric tooth morphologies, demonstrating superior recall (92.7%) and precision (92.3%) over prior work. These results establish a new benchmark in pediatric dental segmentation and highlight the VGG19-enhanced SegUNet's clinical potential.

In 2025, Bayati et al. introduced an advanced diagnostic model in their study *"Advanced AI-Driven Detection of Interproximal Caries in Bitewing Radiographs Using YOLOv8"*, focusing on high-accuracy identification of carious lesions between teeth. The

researchers employed the YOLOv8 architecture to detect and classify interproximal caries using bitewing radiographs. The model was trained and validated on a curated dataset annotated by experienced radiologists, optimizing performance for small, overlapping lesions. Results showed a mean Average Precision (mAP) of 91.3% and an F1-score of 88.7%, outperforming previous YOLO-based models in both speed and accuracy. The model demonstrated strong generalization and practical readiness for integration into dental diagnostic workflows, highlighting the growing clinical viability of YOLOv8 in oral radiographic interpretation.

In 2017, Ruder authored *"An Overview of Multi-Task Learning in Deep Neural Networks"*, offering a foundational guide to the theory and implementation of multi-task learning (MTL) across domains such as NLP, computer vision, and medical imaging. The paper reviews hard and soft parameter sharing as the two principal deep learning MTL strategies and introduces mechanisms like implicit data augmentation, representation bias, and attention focusing to explain MTL's success. Recent innovations such as Cross-stitch Networks, Sluice Networks, and loss-weighting by uncertainty are explored alongside advanced task relationship modeling. Ruder emphasizes the role of auxiliary tasks, such as hint-based learning and domain adversarial training, in improving generalization. This comprehensive review provides both theoretical grounding and practical guidance for applying MTL, particularly relevant for medical imaging systems with multiple, interrelated prediction targets.

In 2023, Hamamci et al. introduced a novel framework titled *"Diffusion-Based Hierarchical Multi-Label Object Detection to Analyze Panoramic Dental X-rays"*, addressing the challenge of training on partially labeled, hierarchical dental data. Their method employs a modified DiffusionDet model that incorporates Swin Transformer encoders, bounding box manipulation, and a multi-label classification head for quadrant, enumeration, and diagnosis prediction. The model is trained on 1,432 labeled X-rays (in varying annotation levels) and evaluated on a fully labeled test set of 250 images with four diagnosis types: caries, deep caries, periapical lesions, and impacted teeth. The full model achieved 43.2% AP for quadrant detection, 30.5% for enumeration, and 37.6% for diagnosis—outperforming state-of-the-art models like DETR and Faster R-CNN. This research sets a new benchmark for multi-task dental diagnostics using hierarchical and partially annotated datasets, with practical implications for clinical decision support and dental AI systems.

In 2024, Cai et al. proposed *"SSAD: Self-Supervised Auxiliary Detection Framework for Panoramic X-ray-Based Dental Disease Diagnosis"*, introducing an end-to-end deep learning framework to overcome annotation scarcity in dental imaging. The SSAD model integrates a disease detection branch with a self-supervised image reconstruction branch, trained jointly via a shared encoder. The reconstruction branch restores masked tooth textures using SimMIM, while a novel Texture Consistency (TC) loss, powered by SAM's image encoder, aligns and refines feature embeddings. Evaluated on the DENTEX dataset across three tasks—quadrant detection, tooth enumeration, and disease diagnosis—SSAD outperformed both supervised and traditional SSL methods. It achieved AP50:95 gains up to 4.03% over the best SSL baselines, and for Faster R-CNN, improved AP50:95 to 31.04%. This lightweight, plug-and-play framework proved highly effective across different architectures and tasks, demonstrating its clinical utility in dental diagnostics under limited supervision.

In 2024, Budagam et al. proposed a dual-model framework in their work *"Instance Segmentation and Teeth Classification in Panoramic X-Rays"*, combining YOLOv8 and a novel BB-UNet architecture to achieve high-accuracy tooth detection and segmentation. The pipeline was trained on a new annotated dataset derived from the UFBA-UESC dental image set, including 425 panoramic X-rays with bounding boxes and segmentation masks. YOLOv8 achieved a precision of 94.3%, recall of 92.3%, mAP of 72.9%, and AP50 of 94.6%, outperforming traditional two-stage models like Mask R-CNN. For segmentation, the BB-UNet—enhanced with bounding box-informed convolution layers—outperformed U-Net, achieving a Dice coefficient of 0.84 versus 0.70. The architecture showed robustness across various tooth types but struggled with implant-containing and anomalous images. This integrated approach highlights the value of bounding box priors in improving instance segmentation and classification performance in dental diagnostics.

In 2025, Wang et al. led a large-scale international study titled *"Artificial Intelligence to Assess Dental Findings from Panoramic Radiographs – A Multinational Study"*, which aimed to establish a robust AI benchmark for comprehensive dental diagnosis using panoramic radiographs (DPRs). The study utilized 6,669 DPRs from the Netherlands, Brazil, and Taiwan to train and evaluate an AI system on eight dental findings, including caries, fillings, implants, and periapical radiolucencies. The AI model combined YOLOv8-based

object detection with DeepLabv3-based tooth indexing and achieved a macro-averaged AUC-ROC of 96.2%, with sensitivity improvements of +67.9% for periapical lesions and +4.7% for missing teeth over human readers. It processed images 79 times faster than dentists and matched inter-reader agreement on 7 of 8 findings. The system demonstrated exceptional generalizability across diverse datasets and outperformed dentists in both accuracy and speed, highlighting its readiness for real-world clinical integration.

2.3 Summary and Research Gap

Although many studies have made significant contributions in classification, detection, and segmentation in dental image analysis, most existing works treat these tasks separately. Furthermore, only a limited number of studies utilize hierarchical multi-task learning in the dental domain. The Dentelligence model will address each task individually by creating a set of specialized models, with each one focusing on a specific task. In the end, these models will be integrated through a fusion mechanism. This approach maintains simplicity while allowing for the development of a multi-task model, without the need to alter the structure of the individual models.

Table 1: Summary of research paper.

RESEARCHERS	PROJECT NAME	DATASET	ALGORITHM USED	OUTCOMES
Azizi et al. (2023)	Artificial Intelligence Techniques in Medical Imaging	Multiple modalities (CT, MRI, PET, etc.)	Machine Learning, Deep Learning, Hybrid Models	Reviewed AI potential and limitations in medical imaging
Ronneberger et al. (2015)	U-Net: Convolutional Networks for Biomedical Image Segmentation	ISBI Challenge Dataset	U-Net	Achieved top segmentation results in biomedical tasks
Schwendicke et al. (2020)	Artificial Intelligence in Dentistry	not applicable	not applicable	Explored AI opportunities and ethical concerns in dentistry
Tuzoff et al. (2019)	Tooth Detection and Numbering in Panoramic Radiographs	1,352 panoramic X-rays	Faster R-CNN, VGG-16	High accuracy in tooth detection and numbering
Haghanifar et al. (2021)	PaXNet: Dental Caries Detection	470 panoramic X-rays (11,769 tooth segments)	InceptionNet, CheXNet, Capsule Network	Accuracy 86.05%, Recall 90.52% for severe caries
Almalki & Latecki (2022)	Self-Supervised Learning for Dental Segmentation and Numbering	TNDRS (annotated from DNS)	SimMIM, UM-MAE, Cascade Mask R-CNN	SimMIM achieved 90.4% AP
Helli & Hamamci (2022)	Tooth Instance Segmentation Using U-Nets and Morphological Processing	105 panoramic X-rays	U-Net + Morphological Processing	Dice score 95.4%, improved tooth count accuracy
Juyal et al. (2023)	Dental Caries Detection Using Faster R-CNN and	300 intraoral images	Faster R-CNN, YOLO V3	Faster R-CNN achieved 80% accuracy, better than

	YOLO V3			YOLO V3
Dhar et al. (2023)	Deep Learning for Teeth Segmentation and Orientation	DNS dataset (543 images)	Modified FUSegNet + PCA	Dice 90.37%, RIoU 82.82%
Mărghinean et al. (2024)	CariSeg: Teeth and Carious Lesion Segmentation	Tufts, Qom, Clujana datasets	U-Net, FPN, DeeplabV3 (ensemble)	Accuracy 99.42%, Dice score 68.2%
Adnan et al. (2024)	AI-Based Application for Caries Index Detection	7,465 intraoral photographs	YOLOv5s, DeTR	F1-score 88.0%, outperformed junior dentists
Zhicheng et al. (2024)	Detection of Impacted Teeth	1,016 panoramic radiographs	MedSAM (fine-tuned SAM)	Accuracy 86.73%, F1-score 0.5350
Beser et al. (2024)	YOLO-V5 for Pediatric Panoramic Radiographs	3,854 pediatric panoramic X-rays	YOLOv5x	Precision, Sensitivity, F1-score ~99%
Özçelik et al. (2024)	Attention Gate-Based Encoder–Decoder Network	313 panoramic radiographs	SE-IB-ED (InceptionV3 + attention modules)	F1-score 92.66%, IoU 86.39%
Kurt et al. (2024)	Tooth Development Stage Evaluation	1,500 pediatric X-rays	YOLOv5	Sensitivity 0.99, F1-score 0.84
Yaxin & Hargreaves (2025)	Dental Teeth X-ray Image Classification	340 annotated X-rays	YOLOv5	F1-score 82%, Precision 89.3%
Yavsan et al. (2025)	Approximal Caries Detection in Children	415 pediatric periapical X-rays	YOLOv5s	F1-score 90.24%, mAP 90.8%
Hua et al. (2025)	YOLO-DentSeg for Oral Disease Segmentation	2,720 panoramic radiographs (DENTEX + clinical)	YOLOv8n-seg + BiFPN, EMCA	mAP (Box) 87.2%, mAP (Seg) 85.5%, 90.3 FPS
Zhong et al. (2025)	Automatic X-ray Teeth Segmentation with Grouped Attention	3,187 panoramic radiographs	GCNet (Grouped Attention + CLF)	Accuracy 97.42%, Dice 93.38%
Ayhan et al. (2025)	Caries Detection vs Dentist Performance	2,150 bitewing radiographs	YOLOv9c-Faster	F1-score 68.7%, recall higher than dentists
Ovi et al. (2025)	Enhanced Pediatric SegUNet with VGG19	323 pediatric panoramic radiographs	SegUNet + VGG19	Dice 92.5%, IoU 91.5%
Bayati et al. (2025)	YOLOv8 for Interproximal Caries	not specified	YOLOv8	mAP 91.3%, F1-score 88.7%
Ruder (2017)	Overview of Multi-Task Learning in Deep Neural Networks	not applicable	Multi-task deep learning frameworks	Theoretical insights into multi-task learning
Hamamci et al. (2023)	Diffusion-Based Hierarchical Multi-Label Detection	1,432 + 250 panoramic X-rays	DiffusionDet + Swin Transformer	AP up to 43.2%, better than DETR & Faster R-CNN
Cai et al. (2024)	SSAD: Self-Supervised Auxiliary Detection	DENTEX	SimMIM, Faster R-CNN + Texture Consistency Loss	AP50:95 improved by 4.03%, robust under SSL
Budagam et al. (2024)	Instance Segmentation and Teeth Classification	UFBA-UESC (425 panoramic images)	YOLOv8, BB-UNet	YOLOv8 AP50 94.6%, BB-UNet Dice 0.84
Wang et al. (2025)	AI Assessment of Dental Findings – Multinational Study	6,669 DPRs from 3 countries	YOLOv8 + DeepLabv3	AUC-ROC 96.2%, faster and more accurate than dentists

CHAPTER 3 RESEARCH METHODOLOGY and DESIGN

3.1 Introduction

The **Dentelligence** project aims to revolutionize dental image analysis by addressing three core tasks: classification, object detection, and segmentation, specifically for dental panoramic X-ray images. A significant challenge in training effective deep learning models for medical image analysis, particularly in fields like dental radiography, is the availability of high-quality, fully labelled datasets. Annotating medical images is a time-consuming process requiring domain-specific expertise, especially in complex cases where each tooth must be identified, numbered, and diagnosed. Real-world datasets often contain annotations at varying levels of detail. Some images may be labelled with broad information such as the affected quadrant, while others provide detailed annotations, including tooth numbers or specific diagnoses. This creates hierarchical and partially labelled datasets that traditional object detection models struggle to process effectively. To address this, recent advancements in AI have introduced methods capable of leveraging such heterogeneous datasets through innovative strategies. [7] This chapter outlines the system's requirements —both functional and non-functional—describes the tools and environments used during development and presents the architectural design and workflow. It highlights key design choices and implementation strategies that ensure the robustness and scalability of the **Dentelligence** solution, contributing to a more efficient and accurate analysis of dental radiographs.

3.2 System Requirements

3.2.1 User Requirements

User requirements are crucial to ensure the system meets the needs of its end-users, primarily dentists and dental specialists. This section outlines the functional and non-functional requirements necessary to provide accurate, efficient, and user-friendly dental disease classification.

3.2.1.1 Functional Requirements

The proposed system, *Dentelligence*, is developed to process dental panoramic X-ray images and produce three types of outputs simultaneously. The following are the key functional requirements:

- **Data Ingestion**: The system must be able to load and preprocess panoramic dental X-ray images in various formats.
- **Multi-task Prediction**: The system shall perform:
  - **Tooth classification**: Predict the type or condition of each tooth.
  - **Tooth detection**: Identify and locate individual teeth or regions of interest using bounding boxes.
  - **segmentation**: Generate precise segmentation masks of teeth and surrounding structures.
- **Result Visualization**: The system should provide visualization of model outputs (e.g., overlaid bounding boxes and masks on the X-ray images).

3.2.1.2 Non-Functional Requirements

In addition to its core functionality, the Dentelligence system must satisfy a set of Non-Functional user requirements to ensure usability, performance, and reliability. These requirements contribute to the system's overall effectiveness and its applicability in real-world medical and research environments. The key Non-Functional requirements include:

- **Accuracy**: The system should provide clinically reliable outputs, ensuring high precision and recall across all tasks (classification, detection, segmentation).
- **Usability**: The system must be user-friendly for dentists and clinicians, with clearly structured outputs and optional visualization tools for reviewing model predictions.
- **Interoperability**: The system should be compatible with widely used data formats (e.g., PNG, JPG, DICOM) and frameworks (e.g., PyTorch, ONNX) to allow easy integration into existing clinical or research workflows.
- **Security and Data Privacy**: All patient data and X-ray images must be handled according to medical data protection standards, ensuring sensitive information is anonymized and securely stored.

Table 2: Summary of User Requirements.

User Requirements	Taip Requirements
Data Ingestion	**Functional Requirements**
Multi-task Prediction	**Functional Requirements**
Result Visualization	**Functional Requirements**
Accuracy	**Non-Functional Requirements**
Usability	**Non-Functional Requirements**
Interoperability	**Non-Functional Requirements**
Security and Data Privacy	**Non-Functional Requirements**

3.2.2 System Requirements:

System requirements define the technical and operational aspects of the project, ensuring the system functions effectively and efficiently. This section outlines the hardware, software, and performance specifications necessary for implementing the dental disease classification.

1. Hardware Requirements List the essential hardware needed to develop and run the system:

- Processing Power: Minimum quad-core processor (e.g., Intel i5 or higher).
- Memory: At least 8GB RAM (16GB recommended for training models).
- Storage: 100GB SSD or higher for handling large dental datasets and models.
- GPU: NVIDIA GPU with CUDA support (e.g., T4 or higher) for deep learning model training.

2. Software Requirements List of the required software tools and libraries:

- **Operating System:** Windows 10/11, macOS, or Linux.
- **Programming Language:** Python 3.x.
- **IDE:** Jupyter Notebook, PyCharm, or Visual Studio Code, Colab.

3.3 System Design

The design of the system is structured around a modular multi-task that integrates object detection, segmentation, and classification tasks into a unified pipeline. The system aims to automate the analysis of dental panoramic X-ray images by sequentially detecting, segmenting, and classifying teeth with high accuracy and efficiency. This section elaborates on the major components of the system, the interconnections between tasks, and adapted to enhance performance and clinical relevance. By leveraging specialized deep learning models for each task and integrating their outputs in a coherent manner, the system mirrors the diagnostic workflow of dental professionals, thus improving reliability and scalability across diverse datasets. The architectural workflow is illustrated in **Figure 2**, and the system can be decomposed into the following major components:

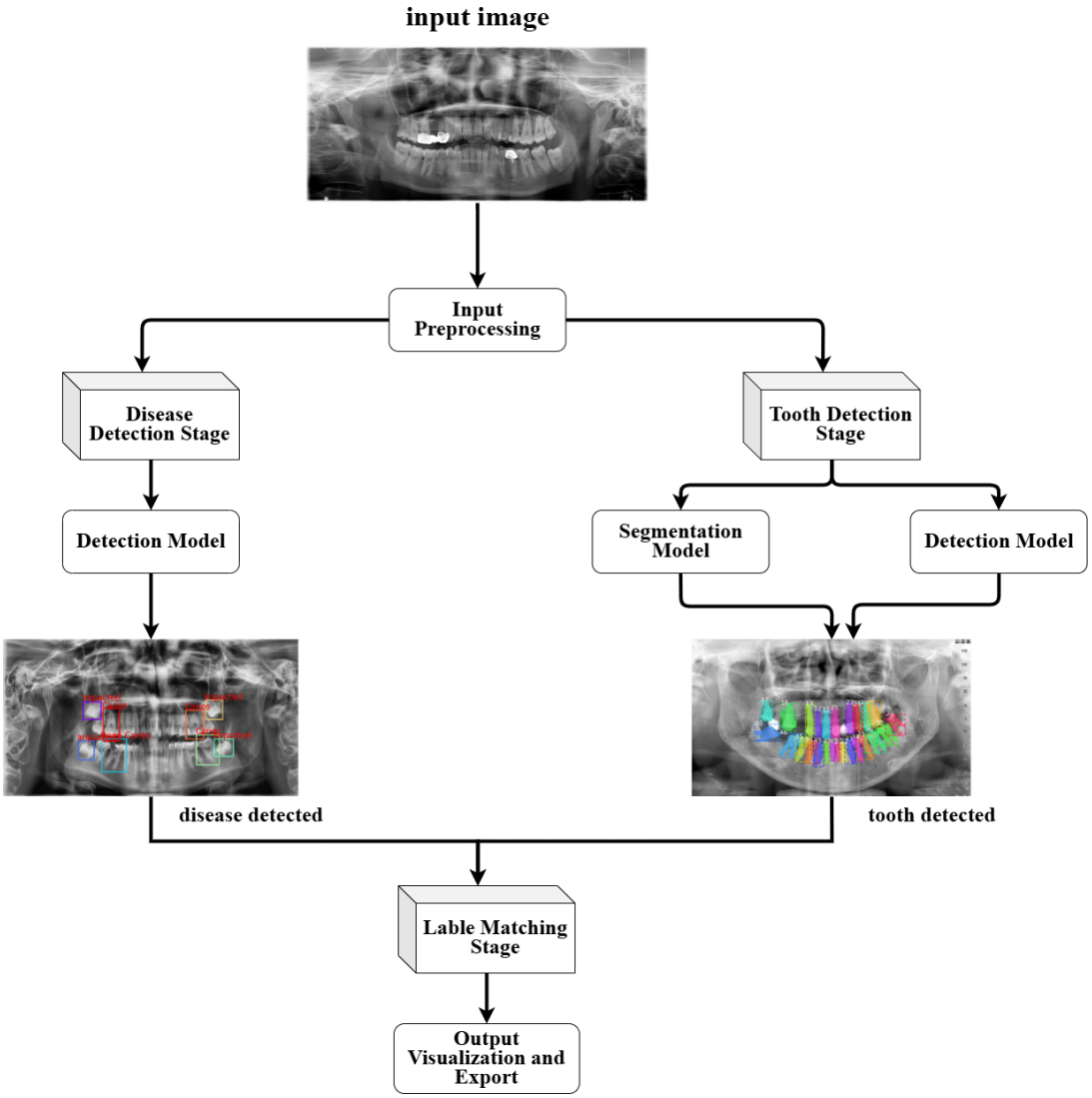

Figure 2: Overview of the Dental X-ray Analysis System Pipeline.

3.3.1 Input Preprocessing

The first stage of the system involves the acquisition and preprocessing of dental panoramic X-ray images to ensure standardized and optimal inputs for the subsequent deep learning models. The preprocessing pipeline is designed to enhance image quality, normalize variations across datasets, and augment the training data to improve model generalization.

Key steps in input processing include:

- **Image Standardization**: All input images are resized to a consistent resolution to ensure uniformity across the dataset, which facilitates stable training and evaluation.

- **Intensity Normalization**: Pixel intensity values are normalized to a fixed range (e.g., [0, 1]) to mitigate the effects of differing exposure levels and imaging conditions.

- **Data Augmentation**: To enhance model robustness and generalization in image classification, a wide range of data augmentation techniques are applied, including affine transformations such as random rotation, flipping, scaling, brightness adjustment, and contrast enhancement. These augmentations help simulate real-world variability and reduce the risk of overfitting, particularly in datasets with limited samples. Traditional augmentation strategies have consistently demonstrated significant improvements in validation accuracy across classification tasks. [32]

- **Annotation Parsing**: Corresponding labels, including bounding boxes for object detection, pixel-level masks for segmentation, and classification labels for each tooth, are parsed and preprocessed in parallel with the images.

dataset restructuring procedure:

To prepare the DENTEX dataset for model training, a complete restructuring process was carried out to transform the raw data into organized, task-specific subsets. This process included the following general steps:

- Separation by Task: The original dataset was split into distinct parts, each corresponding to a specific task in the Dentelligence System — such as quadrant detection, tooth enumeration, and disease classification.

- Annotation Standardization: All labels and annotations from the raw dataset were carefully reformatted and standardized. This made the data easier to use across different models and ensured consistency in structure and naming.

- Creation of Segmentation Data: Tooth segmentation masks were generated to represent the shape and position of individual teeth. These masks were matched with their corresponding X-ray images to prepare inputs for segmentation models.

- Image and Mask Preparation: In some cases, images and masks were cropped or adjusted to focus on specific regions, such as quadrants or individual teeth. This step helped models focus on relevant parts of the image.

- Structured Organization: The resulting data was saved into clearly labeled folders and categories, each organized by model type and training purpose. This organization simplified the training, validation, and evaluation processes.

This preprocessing stage is crucial for achieving high model performance, as it ensures that the input data is clean, consistent, and diverse enough to support effective learning.

3.3.2 Tooth Detection Stage

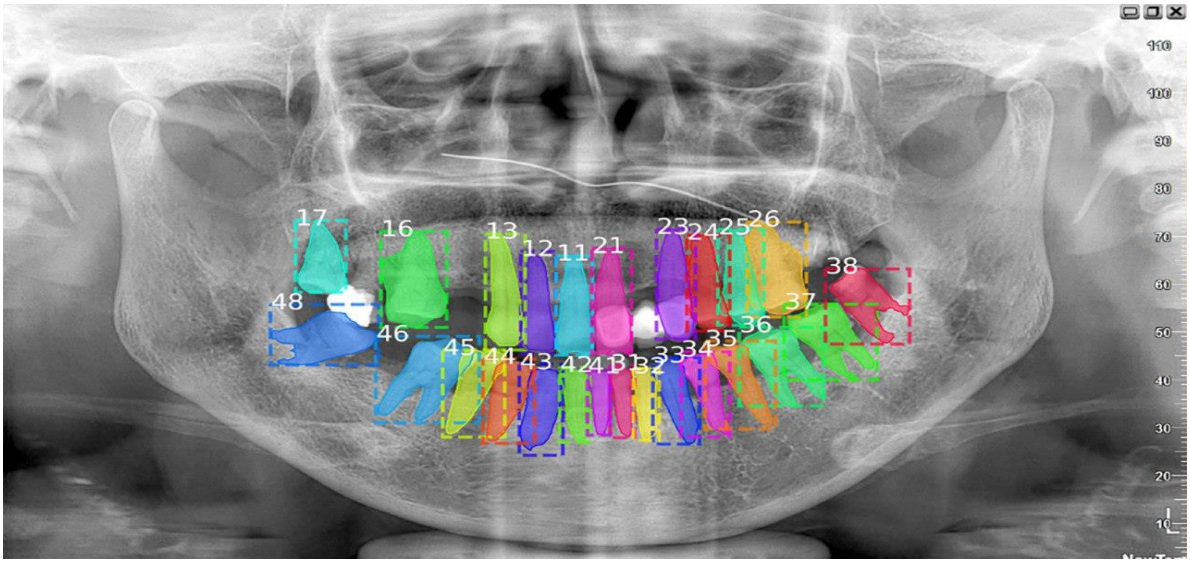

Figure 3: Tooth Detection [33]

3.3.2.1 Detection Model

Object detection plays a pivotal role in the Dentelligence System by enabling precise localization of each tooth, a prerequisite for accurate disease assessment and enumeration. To fulfill this requirement, the system leverages **DINO** — a state-of-the-art end-to-end object detector that extends the DETR (DEtection TRansformer) architecture. DINO integrates transformer-based attention mechanisms with advanced query design and de-noising strategies, allowing it to achieve robust detection performance without relying on traditional anchor boxes or complex post-processing techniques.

DINO Concept Overview

DINO (DETR with Improved DeNoising Anchor Boxes) is a Transformer-based object detector designed for efficient, end-to-end object detection without the need for hand-crafted components such as anchor generation and non-maximum suppression. DINO builds upon the DETR family by enhancing training stability, improving box prediction, and accelerating convergence — all while preserving the global context modelling power of transformers.

DINO formulates detection as a set prediction task: it outputs a fixed number of object candidates and uses bipartite matching to assign predictions to ground truths. This makes it especially suitable for structured and dense object layouts like dental X-rays.

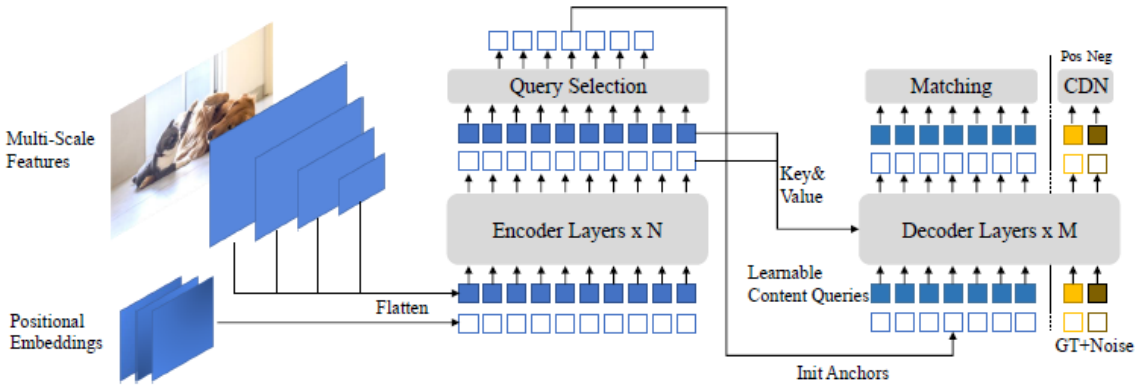

Figure 4: The framework of DINO model.[38]

Key Characteristics of DINO

DINO introduces several innovations over previous DETR-like models, including:

- **Contrastive DeNoising (CDN) Training**: Improves training stability and helps the model distinguish between real and noisy objects by using positive and hard negative samples for each ground truth box.
- **Mixed Query Selection**: Combines dynamic anchor-based positional queries with learnable content queries to better guide the decoder's focus.
- **Look Forward Twice Strategy**: Enables later decoder layers to influence the parameter updates of earlier layers, resulting in better box refinement.
- **Multi-Scale Feature Fusion**: Extracts features at different resolutions, improving performance across small, medium, and large objects.
- **Fast and Scalable**: Achieves high performance with fewer training epochs compared to classical DETR, making it efficient and scalable for practical applications.

DINO-ResNet50 Workflow in Dentelligence

In our Dentelligence System, DINO with a ResNet50 backbone plays a central role in the **detection** component of the **tooth detection stage**. Its specific function is:

- **Input**: X-ray images from the **enumeration32** dataset (where each tooth has a unique ID).
- **Training**: The model is trained on COCO-style annotated images with bounding boxes for individual teeth.
- **Output**: Precise bounding boxes for each detected tooth, which are used downstream for segmentation and disease classification.

After training, DINO is evaluated using the COCO metrics (AP, AR), and inference results are visualized to verify the accuracy and consistency of tooth localization.

Suitability of DINO for Dental X-ray Analysis

DINO is exceptionally suited for dental object detection due to several reasons:

- **End-to-End Optimization**: Removes the need for post-processing like NMS, making it clean and robust for medical pipelines.
- **Fine-Grained Localization**: The anchor-box-based attention mechanism helps in distinguishing closely packed teeth, even in overlapping scenarios.
- **ResNet50 Backbone**: Offers a strong balance of accuracy and computational efficiency, ideal for high-resolution dental images.
- **Transformer-Based Attention**: Allows the model to reason over spatially distributed structures — a perfect match for analyzing full-mouth X-rays where global tooth layout matters.

3.3.2.2 Segmentation Model

Advanced segmentation models play an essential role in the Dentelligence system by improving the accuracy and reliability of tooth boundary detection in dental X-ray images. High-quality segmentation is foundational for downstream tasks such as enumeration, disease detection, and clinical assessment. To achieve this, the system integrates enhancements to established architectures, incorporating mechanisms that strengthen feature representation and improve focus on relevant anatomical structures.

U-Net Concept Overview

U-Net is a convolutional neural network architecture specifically designed for biomedical image segmentation. U-Net follows an encoder-decoder structure, where the contracting path (encoder) captures semantic context, and the expanding path (decoder) enables precise spatial localization. Skip connections are used between the encoder and decoder layers to retain spatial information lost during down sampling.

This architecture was designed to work well with very few annotated images, making it especially suitable for medical imaging tasks where large datasets are rare. One of its core strengths lies in the use of data augmentation, particularly elastic deformations, to increase the diversity of the training data.

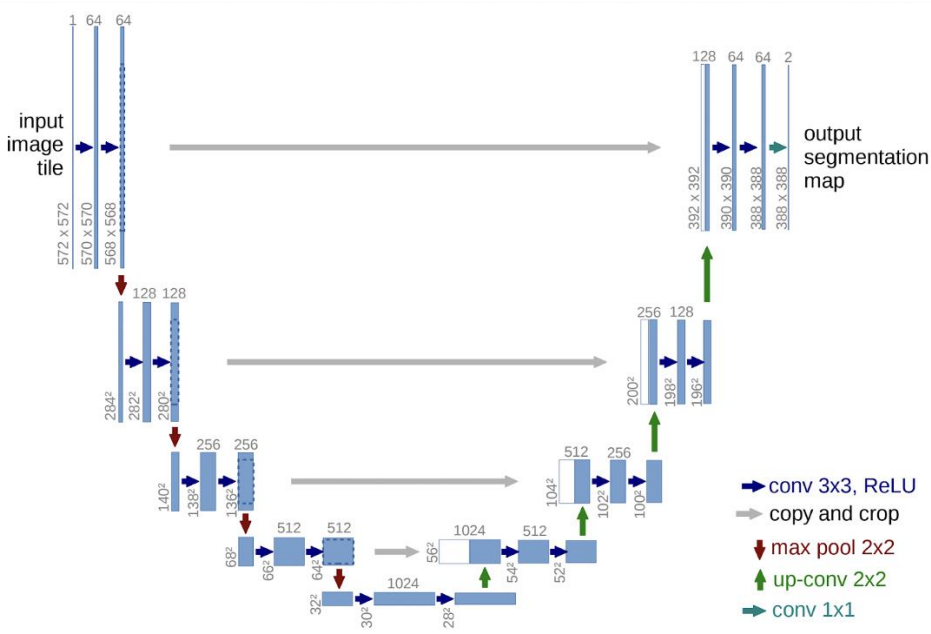

Figure 5: U-Net architecture [35]

Key Characteristics of U-Net

- **Symmetric architecture**: The encoder and decoder paths are symmetric, forming a "U" shape.
- **Skip connections**: These allow fine-grained details from the encoder to be reused in the decoder.
- **Fully convolutional**: There are no dense layers, enabling input images of varying sizes.
- **Tile-based segmentation**: Enables processing of arbitrarily large images using an overlap-tile strategy.
- **Efficient learning from few samples**: Through heavy data augmentation and smart weight initialization.
- **Fast inference**: A 512×512 image can be segmented in less than a second on modern GPUs.

U-Net Workflow in Dentelligence

In the Dentelligence System, U-Net is leveraged in two distinct segmentation workflows:

- **Enumeration 32:**

  In this case, U-Net is trained to perform multi-class segmentation of all 32 teeth in full panoramic dental X-ray images. Each pixel in the image is assigned to a class label from 0 to 31, corresponding to the individual teeth.

- **Enumeration 9:**

  Here, U-Net is trained on cropped quadrant images, which focus on one of the four jaw quadrants at a time. The goal is to segment the 8 teeth within the quadrant. An additional class (class 9) is introduced to represent teeth that appear in the crop but belong to other quadrants—these are considered "out-of-quadrant" objects due to cropping overlap.

  Note: The cropped quadrant images used for training in Enumeration9 are extracted using a DiffusionDet object detection model, which will be discussed in detail in a later section.

Each model is trained independently using supervised learning with cross-entropy loss and evaluated using class-averaged Dice scores.

Suitability of U-Net for Dental X-ray Analysis

Dental panoramic X-ray images present several challenges:

- Limited annotated datasets.
- Need for pixel-wise localization of teeth.
- Significant variation in tooth shape, size, and alignment.

U-Net is well-suited for this task because:

- It can be trained with relatively few images while still generalizing well.
- The skip connections enable accurate localization of small and irregularly shaped structures, such as teeth.
- Its tile-based prediction allows it to scale to high-resolution images typical in dental imaging.
- The design is flexible enough to support both full-mouth (32-class) and localized quadrant-based (9-class) segmentation.

SE-U-Net Concept Overview

SE-U-Net is an extension of the traditional U-Net architecture that integrates Squeeze-and-Excitation (SE) blocks into its design to improve segmentation accuracy. The SE block is a lightweight architectural module introduced by Hu et al. (2017) to enhance the representational capacity of convolutional neural networks by recalibrating channel-wise feature responses using global contextual information. While conventional CNNs treat all channels equally, SE blocks explicitly model the interdependencies between channels, allowing the network to selectively amplify informative features and suppress less useful ones. This results in more focused and discriminative feature representations, which are especially valuable in complex medical imaging tasks like dental segmentation. [39]

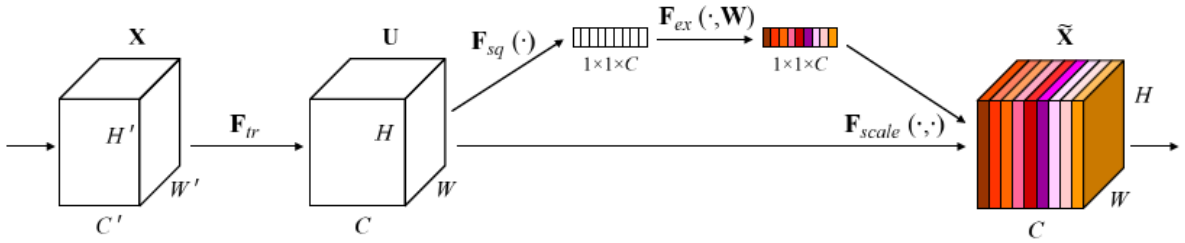

Figure 6: A Squeeze-and-Excitation block. [38]

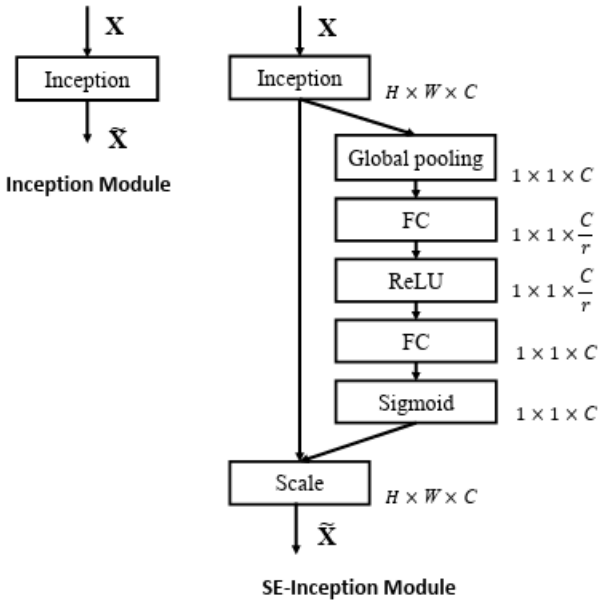

Figure 7: ResNet Vs. SE-ResNet [38]

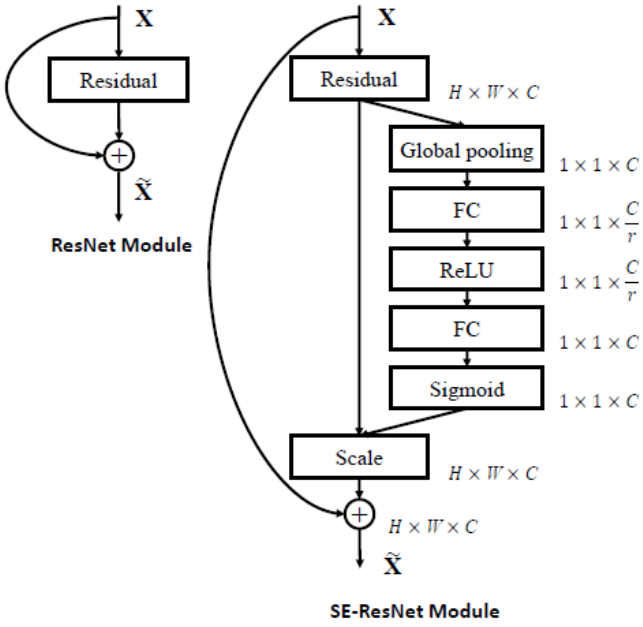

Figure 8: Inception Vs. SE-Inception [38]

Key Characteristics of SE-U-Net

- **Channel-wise Attention**: Through SE blocks, the model learns to emphasize important feature channels and suppress irrelevant ones.
- **Global Context Embedding**: SE blocks use global average pooling to generate a summary statistic for each channel, capturing contextual dependencies.
- **Flexible Integration**: SE blocks can be inserted at multiple depths of the U-Net without altering its spatial structure.
- **Minimal Overhead**: Despite improving accuracy, SE blocks introduce only a small increase in parameters and computational cost.
- **Input-Aware Adaptation**: SE blocks enable the network to dynamically adjust its feature weight based on the specific input.

SE-U-Net Workflow in Dentelligence

In the Dentelligence system, SE-U-Net mirrors the dual-use design of the U-Net model, with added attention mechanisms:

- **Enumeration 32:**

An SE-U-Net model is trained to perform full-mouth 32-class tooth segmentation. Each pixel is assigned one of 32 labels representing individual teeth, benefiting from the SE mechanism to sharpen attention across fine-grained class boundaries.

- **Enumeration 9:**

A second SE-U-Net is trained on cropped quadrant images with a 9-class output:

- Classes 0–7 correspond to the 8 teeth within the quadrant.
- Class 8 represents out-of-quadrant teeth — parts of other quadrants that appear in the crop due to image overlap.

Note: These quadrant images are produced by a DiffusionDet object detection model, which identifies and crops the quadrants from the full panoramic image. (DiffusionDet is discussed in a later section.)

SE blocks allow the network to better differentiate in-quadrant teeth from these overlapping structures by learning context-sensitive channel activations. This improves both segmentation precision and generalization to variable image layouts.

Suitability of SE-U-Net for Dental X-ray Analysis

Dental X-rays are high-resolution, low-contrast images with overlapping structures and limited annotation data. SE-U-Net is particularly suited for this domain due to:

- Its global feature is recalibration, which strengthens relevant activations for target teeth.
- Enhanced robustness to variability in anatomy, pose, and imaging conditions.
- The ability to generalize from small datasets, by leveraging SE blocks' attention to informative features.
- Seamless extension of U-Net's pixel-wise precision with channel-wise feature discrimination, critical for segmenting individual teeth.

Why Enumeration 32 and Enumeration 9

To achieve flexible and accurate tooth segmentation across different contexts, the Dentelligence system employs two segmentation modes: enumeration32 and enumeration9.

- **Enumeration 32:** targets the entire panoramic X-ray, segmenting all 32 teeth in a single pass. This provides a global understanding of the full dentition and ensures consistency across the full mouth layout.
- **Enumeration 9:** focuses on individual jaw quadrants, where each cropped image contains 8 teeth. This mode introduces an additional class (class 9) to represent teeth from other quadrants that may partially appear in the crop. These cropped regions are generated using the DiffusionDet detection model, which precisely isolates each quadrant prior to segmentation.

This dual approach allows the system to leverage global context when appropriate (via enumeration32) and zoom into local regions (via enumeration9) for higher resolution analysis, reducing class imbalance and increasing robustness in overlapping or occluded regions.

Integration of Detection and Segmentation Outputs in the Tooth Detection Stage

In the Dentelligence pipeline, segmentation and detection models are complementary components, working in tandem to improve tooth localization accuracy.

- **DINO (Detection Transformer)** is first used to **predict bounding boxes** for all visible teeth in the panoramic image. This step provides coarse but spatially precise tooth proposals.

- **U-Net and SE-U-Net** are then employed to perform **pixel-wise segmentation**, refining the exact shape and boundary of each tooth within those bounding boxes or within the full/cropped image.

By integrating outputs from both detection and segmentation, the system benefits from the spatial awareness of detection and the fine-grained boundary delineation of segmentation. Results are further enhanced using fusion techniques like **Weighted Boxes Fusion (WBF)**, followed by a label matching step to assign consistent tooth identifiers.

This combined strategy ensures more accurate and robust tooth detection under varied anatomical structures and imaging conditions.

DiffusionDet Concept Overview

DiffusionDet is a novel object detection framework that formulates detection as a generative denoising process over bounding boxes. Instead of relying on fixed anchor points or learned object queries as in traditional detectors, DiffusionDet starts from a set of randomly generated noisy bounding boxes and progressively refines them to accurate object locations using a diffusion-based denoising network. Inspired by denoising diffusion probabilistic models (DDPM) in image generation, this approach applies the concept of reversing noise over iterations—but in the space of bounding boxes rather than pixels—making object detection a more flexible and probabilistically grounded task.

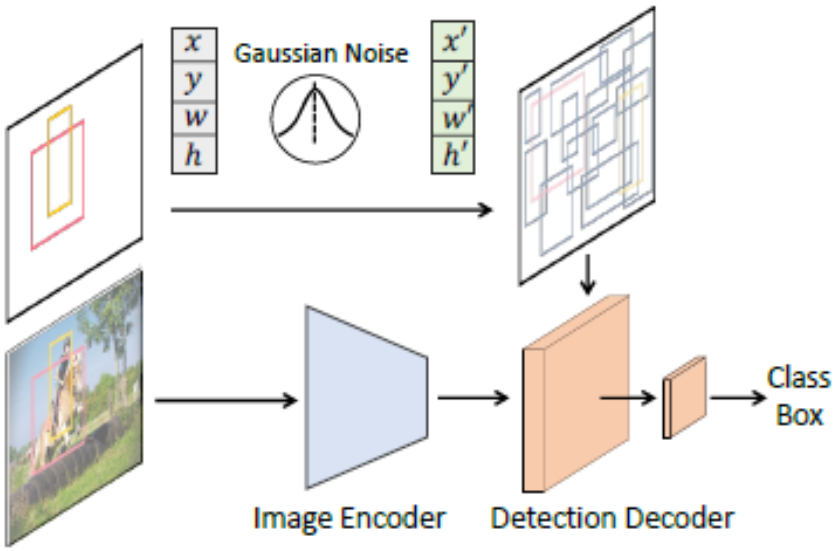

Figure 9: DiffusionDet Overall pipeline [39]

Key Characteristics of DiffusionDet

- Noise-to-Box Paradigm: Instead of predefined or learned object priors, DiffusionDet begins with purely random boxes, reducing architectural bias and simplifying design.
- Dynamic Evaluation: It supports evaluating with any number of boxes or denoising steps without re-training the model, a flexibility unmatched by most detectors.
- Iterative Refinement: The detection head can be reused multiple times (unlike other detectors that apply it once), gradually improving prediction quality.
- Probabilistic Generative Foundation: This enables DiffusionDet to generalize well across domains and densities, showing strong performance even in zero-shot transfer settings.

DiffusionDet Workflow in Dentelligence

In the Dentelligence System, DiffusionDet is employed for detecting quadrants in full panoramic dental X-ray images. While the U-Net and SE-U-Net models are trained on cropped images to detect teeth within a single quadrant (enumeration9 task), DiffusionDet handles the **coarser localization task**—identifying the quadrant bounding boxes in the full image.

The workflow involves:

- Inputting the full panoramic image to DiffusionDet.
- Initializing a set of random bounding boxes as the starting point.
- Iteratively denoising these boxes using a ResNet backbone and detection decoder.
- Producing final quadrant box predictions, which are used to guide the downstream enumeration models by cropping each quadrant for further tooth-level analysis.

Suitability of DiffusionDet for Dental X-ray Analysis

Dental radiographs exhibit several unique challenges: variability in object scale and density, subtle structural details, and a limited number of large anatomical structures (e.g., quadrants). DiffusionDet's strengths align well with these needs:

- **No need for dense anchors or queries**: This reduces overfitting to hard-coded assumptions, making it robust to anatomical variation across patients.
- **Iterative refinement**: Helps in precisely aligning quadrant boundaries, which is critical in ensuring accurate downstream enumeration and disease detection.
- **Generalization**: Its success in zero-shot settings, like COCO → CrowdHuman, mirrors the need in dental diagnostics to transfer models between datasets with different acquisition protocols or patient demographics.

3.3.3 Disease Detection Stage

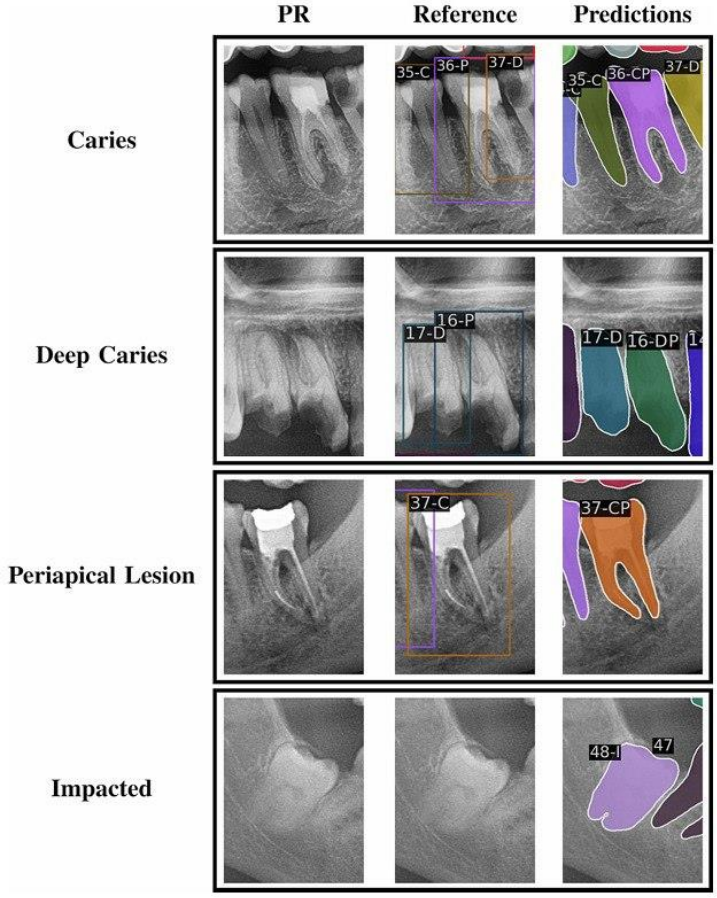

Figure 10: Disease Detection [36]

The Disease Detection Stage is designed to localize and classify pathological regions within dental panoramic radiographs, focusing exclusively on identifying diseased teeth rather than detecting all anatomical structures. As illustrated in Fig. 2, this module operates in parallel with the tooth detection and segmentation branches and contributes essential findings that are later fused into the final diagnostic output. To achieve this goal, the system employs two high-performance object detectors—DINO with a Swin Transformer backbone and YOLOv8x—each trained on the same dataset containing annotations for diseased teeth. These models are optimized to recognize a variety of dental pathologies, such as carious lesions, periapical radiolucency, and structural damage, by learning from expert-labelled regions in quadrant-specific radiographs. During inference, both models generate disease-specific bounding boxes with associated confidence scores and class labels. Their outputs are then combined using Weighted Boxes Fusion (WBF) to form a unified and more accurate detection result. This ensemble-based approach improves robustness, especially in challenging cases involving overlapping teeth, small lesions, or low-contrast regions. By

isolating only the pathological areas, this stage enhances diagnostic clarity and enables the downstream modules to produce structured reports that highlight clinically significant findings.

DINO-Swin Backbone

We have previously introduced the DINO architecture with a ResNet50 backbone in the context of tooth detection, where we outlined its core concept, characteristics, and its role in the Dentelligence System. In this section, we focus solely on its adaptation with the Swin Transformer backbone for the task of disease detection.

DINO-Swin Workflow in Dentelligence

In our Dentelligence System, DINO with a Swin Transformer backbone is used in the disease detection stage to localize pathological regions in dental X-rays. Its specific function is:

- **Input**: X-ray images from the disease dataset (where teeth are annotated with disease-specific bounding boxes).
- **Training**: The model is trained using COCO-style annotations to detect disease regions across multiple quadrants with high spatial precision.
- **Output**: Bounding boxes with class labels and confidence scores for detected diseases, which are later fused with other model predictions.

After training, the DINO-Swin model is evaluated using COCO metrics (AP, AR) to assess its detection performance. The results are then visualized and passed into the ensemble fusion stage for robust disease identification.

YOLOv8 Concept Overview

YOLOv8 is the latest evolution of the YOLO (You Only Look Once) object detection family, designed by Ultralytics. Unlike its predecessors, YOLOv8 introduces an anchor-free architecture that simplifies training and enhances generalization across scales. It unifies detection, classification, and segmentation tasks under a single, end-to-end differentiable framework, enabling real-time processing with high accuracy. Its modular architecture consists of an enhanced CSPNet backbone, a feature-rich FPN+PAN neck, and an anchor-free head for bounding box regression and classification. This shift results in fewer hyperparameters, faster convergence, and better adaptability to various object shapes and sizes.

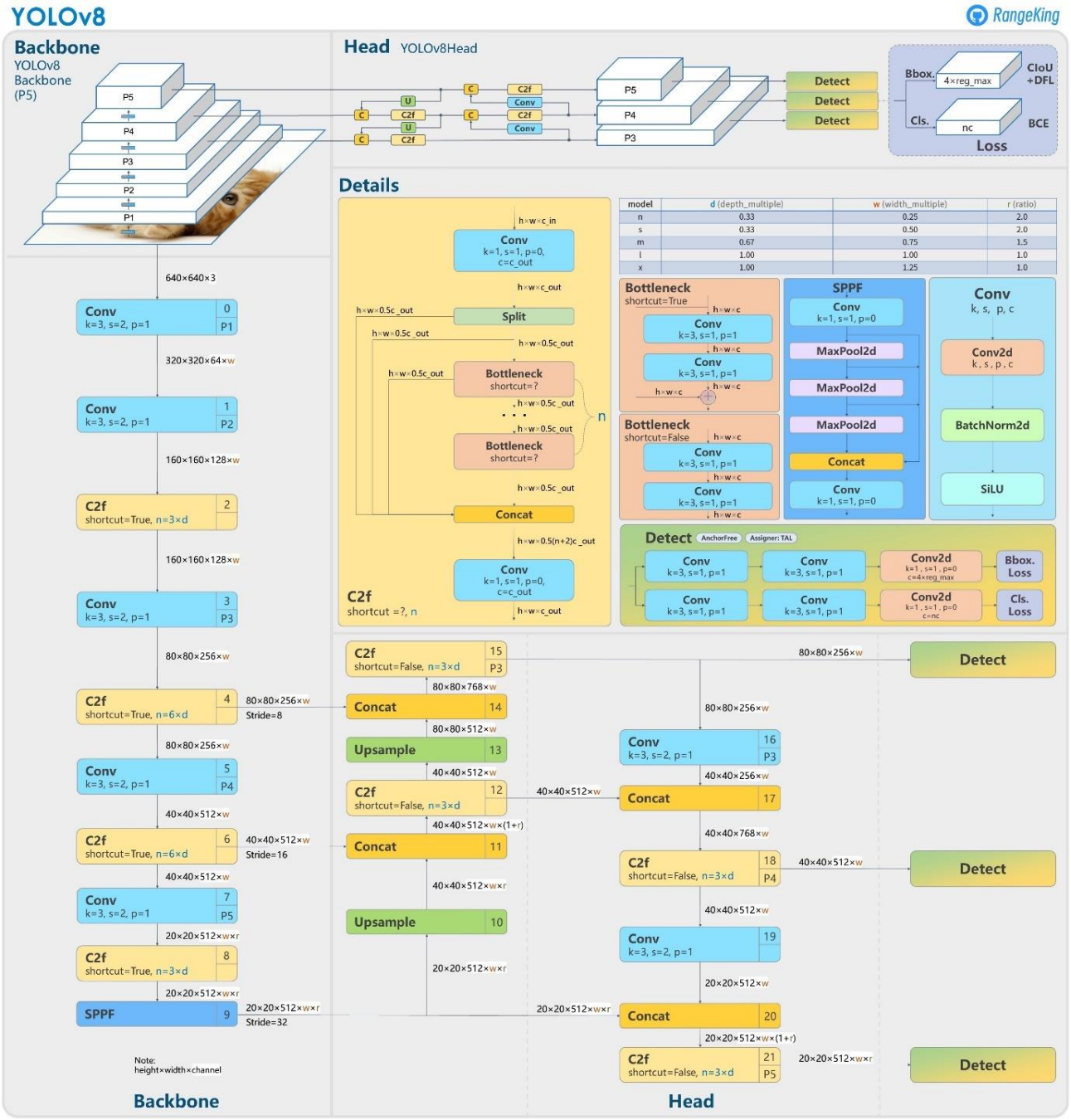

Figure 11: YOLOv8 framework. [34]

Key Characteristics of YOLOv8

- **Anchor-Free Detection**: Moves away from predefined anchor boxes, improving efficiency and reducing sensitivity to object aspect ratios.
- **CSPNet Backbone**: Enhances gradient flow and reduces computational redundancy while maintaining high feature expressiveness.

- **Enhanced PANet Neck**: Facilitates robust multi-scale feature fusion, improving detection of small and densely packed objects.
- **Advanced Augmentation**: Incorporates techniques like Mosaic and MixUp to increase training diversity and robustness.
- **Focal & IoU Loss Functions**: Addresses class imbalance and refines localization for better detection of subtle abnormalities.
- **Mixed Precision Training**: Accelerates training/inference while reducing memory usage, essential for large-scale medical datasets.
- **Model Variants**: Offers scalable options (nano to x-large); YOLOv8x (used in Dentelligence) is the most powerful, with ~90M parameters and highest mAP.

YOLOv8x Workflow in Dentelligence

In our Dentelligence System, YOLOv8x is used in the disease detection stage alongside DINO-Swin to identify pathological regions in dental radiographs. Its specific function is:

- **Input**: X-ray images from the disease dataset (where teeth are annotated with disease-specific bounding boxes).
- **Training**: The model is trained using YOLO-format annotations to detect disease instances across multiple quadrants with high precision.
- **Output**: Bounding boxes with class labels and confidence scores for detected diseases, which are later fused with other model predictions.

After training, YOLOv8x is evaluated using mAP and IoU metrics to quantify its accuracy and robustness. Its outputs are passed into the ensemble fusion stage, where they are combined with DINO-Swin results using Weighted Boxes Fusion for final decision making.

Suitability of YOLOv8 for Dental X-ray Analysis

YOLOv8x is particularly suitable for dental X-ray analysis due to its ability to accurately detect small, low-contrast, and densely packed pathological features common in oral radiographs. The anchor-free design improves generalization across various disease morphologies, while the advanced PANet neck ensures reliable performance on multi-scale lesions. Furthermore, its high speed and modularity make it an ideal candidate for real-time clinical applications and scalable deployment in diagnostic systems.

Ensembling with Weighted Boxes Fusion (WBF)

Weighted Boxes Fusion Overview

Weighted Boxes Fusion (WBF) is an advanced ensembling method designed to improve object detection results by combining predictions from multiple models. Unlike traditional Non-Maximum Suppression (NMS) or Soft-NMS, which eliminate overlapping boxes, WBF merges all boxes by computing a confidence-weighted average of their coordinates. This technique retains more information and improves localization precision, particularly when individual models predict slightly different but overlapping boxes.

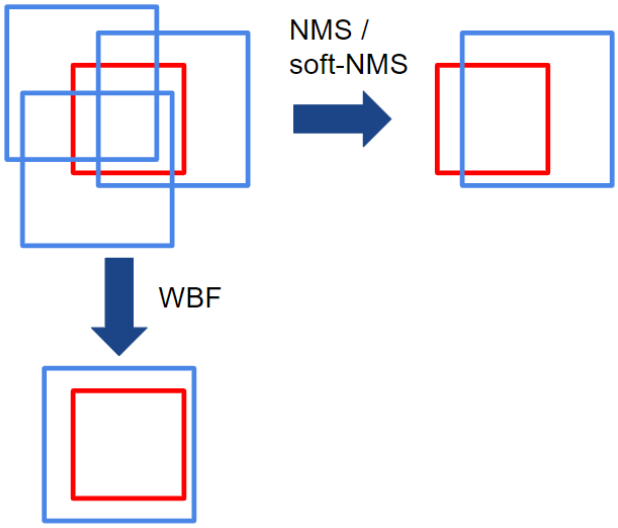

Figure 12: NMS/soft-NMS Vs. WBF [40]

Integration of DINO-Swin and YOLOv8x using WBF

In our Dentelligence System, WBF is used to integrate predictions from two complementary models—DINO with a Swin Transformer backbone and YOLOv8x—both trained for disease detection on the quadrant-enumeration-disease dataset. The fusion process is as follows:

- **Step 1**: Predictions from both models are collected, including bounding boxes, class labels, and confidence scores.
- **Step 2**: All boxes are sorted by confidence, and those with significant spatial overlap (IoU > 0.55) are grouped into clusters.
- **Step 3**: Each cluster is fused into a single bounding box using a confidence-weighted average of coordinates and class scores.
- **Step 4**: Final scores are adjusted based on how many models contributed to each box to penalize predictions not confirmed by both models.

This ensemble strategy enhances detection performance by leveraging the strengths of both architectures: DINO-Swin's robust transformer-based spatial modeling and YOLOv8x's high-speed dense localization. WBF enables our system to output more reliable and accurate disease predictions, especially in cases with overlapping detections or ambiguous regions.

3.3.4 Label Matching Stage

After performing detection, segmentation, and classification tasks independently, the Dentelligence system must consolidate these outputs into a unified and clinically coherent result. The Label Matching Stage is responsible for associating each detected and segmented tooth with its correct identification number, disease condition, and diagnostic annotations, ensuring consistency and logical alignment across the system's predictions.

Purpose of Label Matching

The Label Matching Stage serves several critical purposes:

- **Consolidate Outputs**: It integrates the results from object detection, semantic segmentation, and classification into a single consistent output for each tooth.
- **Assign Correct Tooth IDs**: Detected teeth must be matched with their respective FDI numbering labels, considering spatial location and anatomical context.
- **Associate Disease Labels**: If diseases were detected in the preliminary stages, they must be correctly linked to the appropriate tooth without conflicts.
- **Correct Errors and Refine Predictions**: Matching helps identify inconsistencies between stages (e.g., mismatched bounding boxes and masks) and resolves them through correction strategies.

Label Matching Workflow in Dentelligence

The process involves several steps:

- Spatial Mapping: results from integrated models (tooth detection) are used as anchors to match classified labels based on overlapping areas and geometric proximity.
- Confidence Voting: When multiple candidates match exists, a weighted voting mechanism based on confidence scores is applied to select the most likely association between detections, masks, and labels.

- Hierarchical Prioritization: The system uses anatomical knowledge (e.g., expected positions of molars, premolars, incisors) to prioritize matching decisions, reducing logical inconsistencies.
- Post-Processing Filters: Additional rules are applied to filter out unmatched or low-confidence predictions, ensuring that only clinically meaningful matches are retained.

Impact of Label Matching

Accurate label matching is fundamental to the success of the Dentelligence system. It ensures that:

- Each tooth is correctly identified and annotated.
- Diagnostic information is accurately linked to the corresponding structures.
- Clinical reliability and interpretability of the system's outputs are maximized.

Through sophisticated label matching strategies, Dentelligence achieves a level of diagnostic integration that mirrors the expert reasoning process of human dental practitioners.

3.3.5 Output Visualization and Export

After completing detection, segmentation, classification, and label matching, the Dentelligence system generates final outputs that must be easily interpretable by clinicians, researchers, and system developers. The Output Visualization and Export stage focuses on presenting the results in a clear, structured, and accessible manner, facilitating clinical review, research analysis, and future system integration.

Visualization of Results

To enhance interpretability, the system overlays diagnostic outputs onto the original panoramic X-ray images:

- Bounding Boxes: Rectangular boxes are drawn around each detected tooth, labelled with the assigned FDI tooth number and condition (if applicable).
- Segmentation Masks: Pixel-level masks are superimposed semi-transparently over the corresponding teeth, clearly delineating their boundaries.
- Annotations: Textual annotations indicating tooth number, disease condition (e.g., carious, restored, missing), and classification confidence scores are displayed near each detection.

This multi-layered visualization approach provides a comprehensive and intuitive representation of the diagnostic findings, closely resembling the manual annotation process performed by dental experts.

Exporting the Outputs

The system supports exporting the processed results in multiple formats to maximize usability:

- Annotated Images: The visualized X-ray images with bounding boxes, masks, and labels are saved in high-resolution formats (e.g., PNG, JPEG) for easy review or reporting.
- Metadata Files: Structured metadata files (e.g., JSON, XML) are generated, containing detailed information for each detected tooth, including:
    - Bounding box coordinates
    - Segmentation mask file paths
    - Assigned tooth number
    - Detected disease condition
    - Confidence scores
- Tabular Reports: Summary tables listing detected teeth, their conditions, and corresponding performance metrics (e.g., confidence levels) can be exported as CSV files for further analysis or record-keeping.
- Optional DICOM Export: For clinical integration, the system may support saving annotated outputs in DICOM format with embedded structured reporting fields, facilitating compatibility with hospital information systems (HIS) and picture archiving and communication systems (PACS).

Benefits of Structured Output

- Ease of Clinical Interpretation: Dentists can quickly review annotated X-rays, identifying problematic areas briefly.
- Research Utility: Researchers can systematically analyze model outputs, compare performance across datasets, and conduct statistical evaluations.
- System Integration: Exported metadata can be easily incorporated into electronic health records (EHRs), training pipelines, or external decision-support systems.

Through meticulous visualization and structured exporting strategies, Dentelligence ensures that its powerful diagnostic outputs are not only accurate but also accessible and actionable, bridging the gap between advanced AI models and practical clinical workflows.

CHAPTER 4 RESULTS AND DISCUSSION

4.1 Introduction

This chapter presents a comprehensive analysis of the experimental results obtained from the Dentelligence System. The system's performance is evaluated across its key components, which include tooth segmentation, tooth detection, quadrant localization, and disease classification—each essential for automated dental diagnostics using panoramic X-ray images. Multiple deep learning models were implemented and assessed to determine their effectiveness within the system's modular pipeline. For segmentation, both U-Net and SE-U-Net were used to perform multi-class labelling of dental structures in full-mouth images and cropped quadrants. For detection, DINO with ResNet50 was employed for tooth localization, while DiffusionDet was used to isolate individual quadrants for focused analysis. Disease detection was performed using two complementary models: YOLOv8x for high-speed prediction and DINO with a Swin Transformer backbone for enhanced accuracy. Evaluation was carried out using standard performance metrics including Dice score, Intersection over Union (IoU), and mean Average Precision (mAP). Additionally, fusion strategies such as Weighted Boxes Fusion (WBF) were applied to integrate predictions from multiple models, enhancing overall diagnostic robustness. Visual results are provided to support the quantitative findings, demonstrating the strengths and weaknesses of each model in realistic scenarios. The results discussed here highlight the system's capability to perform multi-task dental analysis with competitive accuracy and efficiency, laying the groundwork for further refinement and clinical integration.

4.2 Quadrant Detection Results

The quadrant detection task was performed using the **DiffusionDet** model with a Swin Transformer backbone, trained on the DENTEX panoramic X-ray dataset for coarse localization of the four dental quadrants. The model was trained using the configuration described in Section 3.3.1, with hyperparameters tuned to maximize Average Precision (AP) across multiple IoU thresholds.

Training Performance

Figure 13 shows the training loss curve over the course of training, illustrating a smooth and consistent decrease in loss without significant oscillations. This indicates stable optimization and effective convergence of the DiffusionDet model.

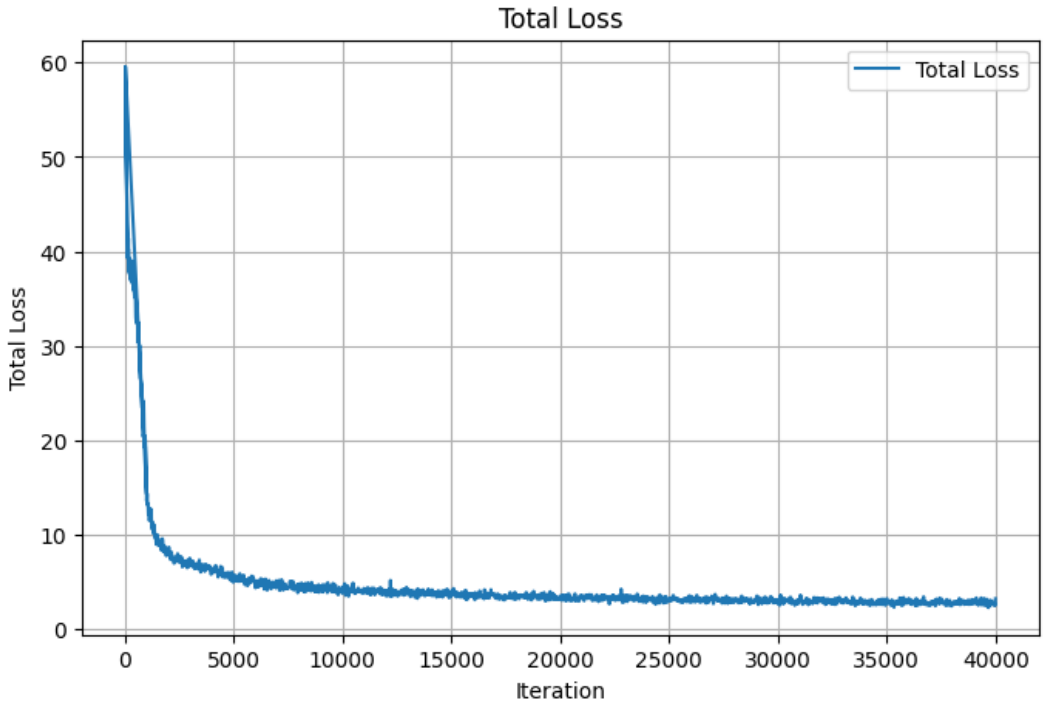

Figure 13: DiffusionDet Loss Curve Plot

Qualitative Results

Figure 14 presents sample inference results, comparing predicted bounding boxes for quadrants (in green) with their corresponding ground truth annotations (in red). The model successfully delineates quadrant boundaries even in cases with overlapping structures and low-contrast regions, demonstrating its robustness in realistic clinical scenarios.

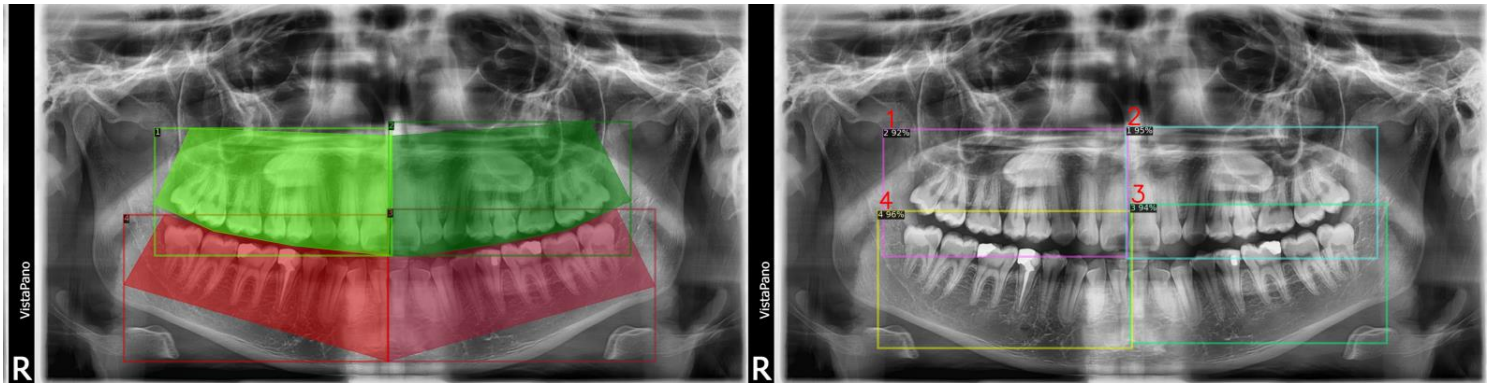

Figure 14: DiffusionDet Ground Truth vs. Predictions

Quantitative Evaluation

The model was evaluated on the validation set using COCO-style metrics. The results are summarized below:

Table 3: DiffusionDet Quantitative Evaluation

Metric	Value
AP@[0.50:0.95] (all)	**0.705**
AP@0.50 (all)	**0.999**
AP@0.75 (all)	**0.892**
AP@[0.50:0.95] (large)	**0.705**
AR@[0.50:0.95] (all, maxDets=100)	**0.761**
AR@[0.50:0.95] (large, maxDets=100)	**0.761**

Conclusion

The DiffusionDet quadrant detection model provides highly accurate localization of dental quadrants, serving as a reliable preprocessing step for subsequent enumeration and disease detection models in the Dentelligence pipeline. Its strong AP and AR values ensure that cropped quadrant images contain complete and correctly aligned tooth regions, which is essential for maintaining high accuracy in later stages.

4.3 Enumeration Detection Results

4.3.1 Enumeration 32 Detection Results

4.3.1.1 Detection model

The enumeration 32 detection task aims to localize and assign an enumeration ID (1–32) to each tooth in the full panoramic dental X-ray. For this task, the **DINO** object detection architecture with a **ResNet50 backbone** was employed, benefiting from its transformer-based set prediction framework, strong localization accuracy, and robustness to densely packed instances. The training configuration followed the methodology outlined in Section 3.3.2, using COCO-style annotations generated from the DENTEX enumeration dataset.

Training Performance

The training loss curve (Figure 15) shows a steep decline during the early epochs, followed by stabilization, indicating efficient convergence without signs of overfitting or instability. This trend reflects the model's ability to rapidly learn spatial relationships between teeth while refining its bounding box regression in later stages.

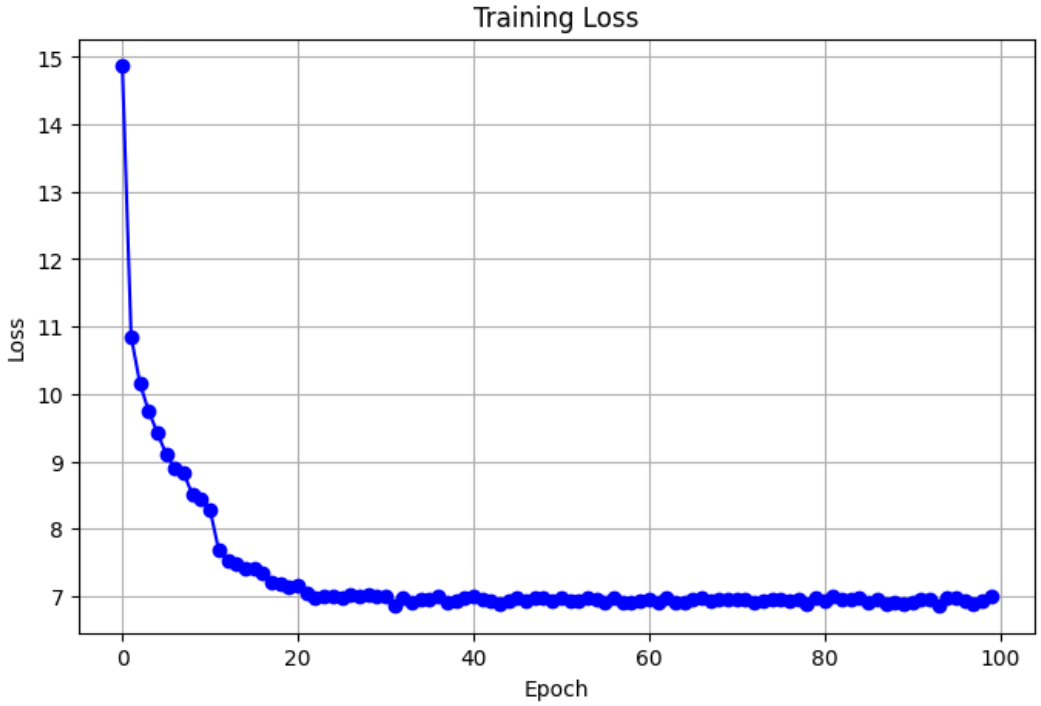

Figure 15: Dino-ResNet50 Loss Curve Plot

Qualitative Results

Figure Y compares the model's predictions (right) to the ground truth bounding boxes (left). The DINO–ResNet50 model successfully detects and enumerates the majority of teeth, including those in overlapping or low-contrast areas, which are typical challenges in panoramic X-ray interpretation.

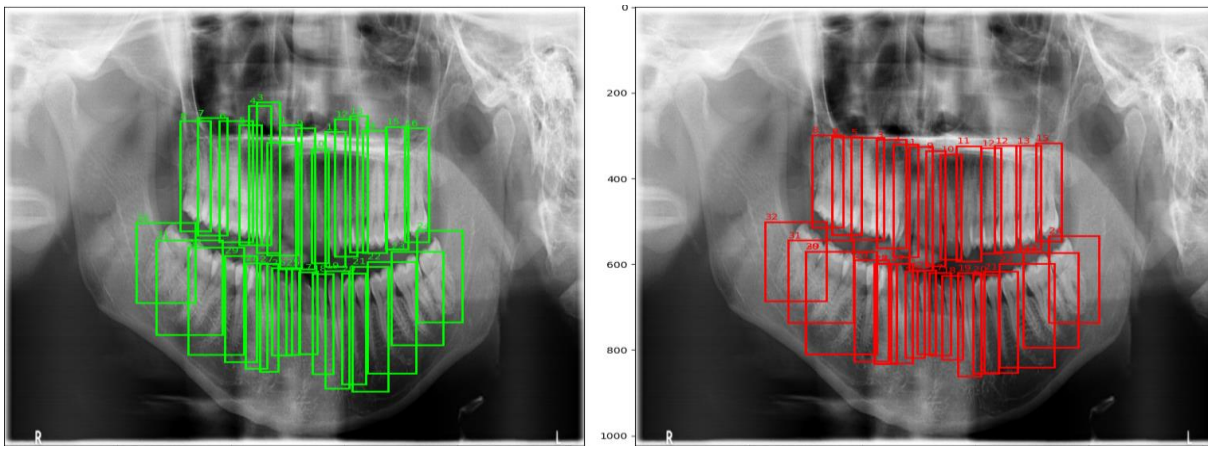

Figure 16: Dino-ResNet50 Ground Truth vs. Predictions

Quantitative Evaluation

Evaluation on the validation set was performed using COCO-style metrics:

Table 4: Dino-ResNet50 Quantitative Evaluation

Metric	Value
AP@[0.50:0.95] (all)	**0.569**
AP@0.50 (all)	**0.959**
AP@0.75 (all)	**0.607**
AP@[0.50:0.95] (medium)	**0.677**
AP@[0.50:0.95] (large)	**0.568**
AR@[0.50:0.95] (maxDets=100)	**0.686**
AR@[0.50:0.95] (medium)	**0.684**
AR@[0.50:0.95] (large)	**0.686**

Conclusion

The DINO–ResNet50 model demonstrates strong enumeration detection performance for full-mouth (32-tooth) panoramic X-rays. Its high recall and precision, particularly at IoU=0.50, ensures reliable enumeration for integration with segmentation outputs from U-Net and SE-U-Net models in the Dentelligence pipeline. This detection robustness forms a solid foundation for downstream disease classification.

4.3.1.2 Segmentation models

Following detection with the DINO–ResNet50 model, segmentation was performed to precisely delineate tooth boundaries and assign per-pixel enumeration labels for all 32 teeth in panoramic X-ray images. Two architectures were evaluated: **U-Net** and **SE-U-Net**. The models were trained on the enumeration32 segmentation dataset, as described in Section 3.3.2, using cross-entropy and Dice losses to optimize both boundary accuracy and region completeness.

Training Performance

Figures 18 and 17 show the loss curves for U-Net and SE-U-Net, respectively. Both models demonstrated stable convergence, with rapid loss reduction in the early epochs followed by

gradual refinement. The validation loss maintained a steady trajectory, indicating consistent generalization across the dataset.

- *U-Net*: Lower training loss but slightly higher fluctuations in validation loss, possibly due to sensitivity to high inter-tooth similarity in the dataset.
- *SE-U-Net*: Slightly smoother validation loss curve, suggesting that the channel attention mechanism improved robustness against anatomical variability and image noise.

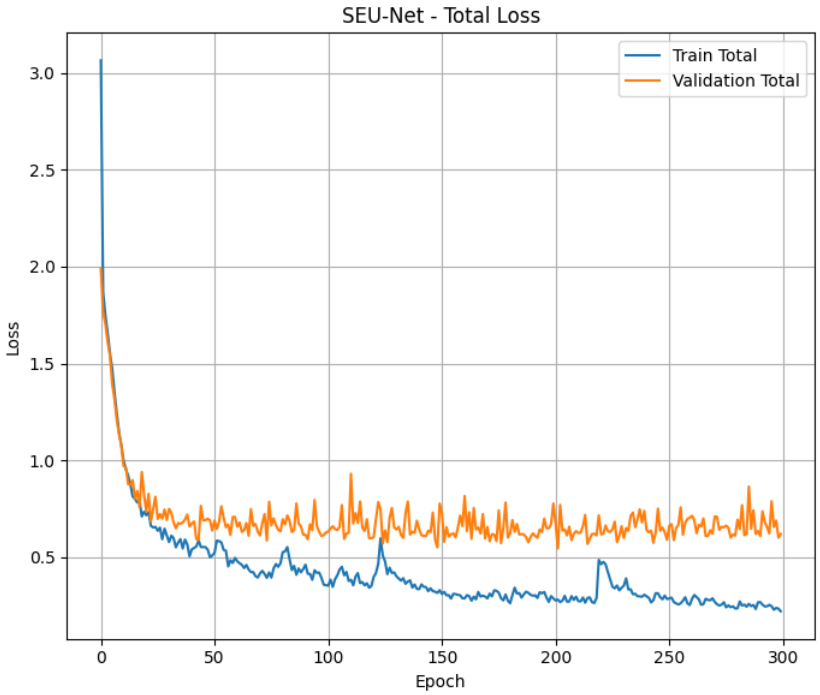

Figure 17: SE-U-net 32 Loss Curve Plot

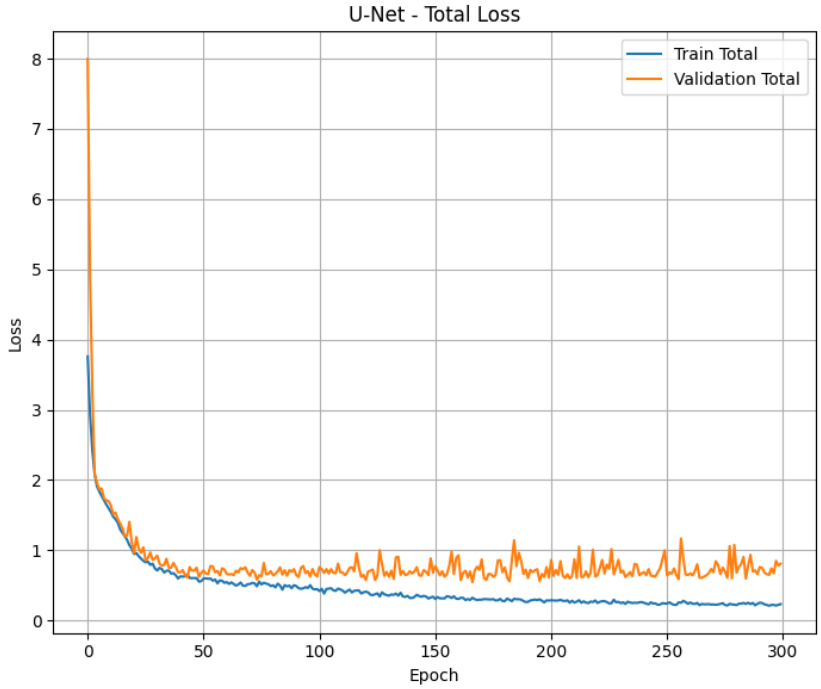

Figure 18: U-net 32 Loss Curve Plot

Qualitative Results

Figures 19 and 20 present segmentation results for U-Net and SE-U-Net, respectively. Ground truth masks are displayed alongside predictions. Both models successfully capture the fine boundaries of individual teeth, maintaining accurate numbering even in crowded posterior regions. SE-U-Net shows slightly cleaner boundaries in low-contrast anterior teeth, attributable to its Squeeze-and-Excitation modules enhancing feature selectivity.

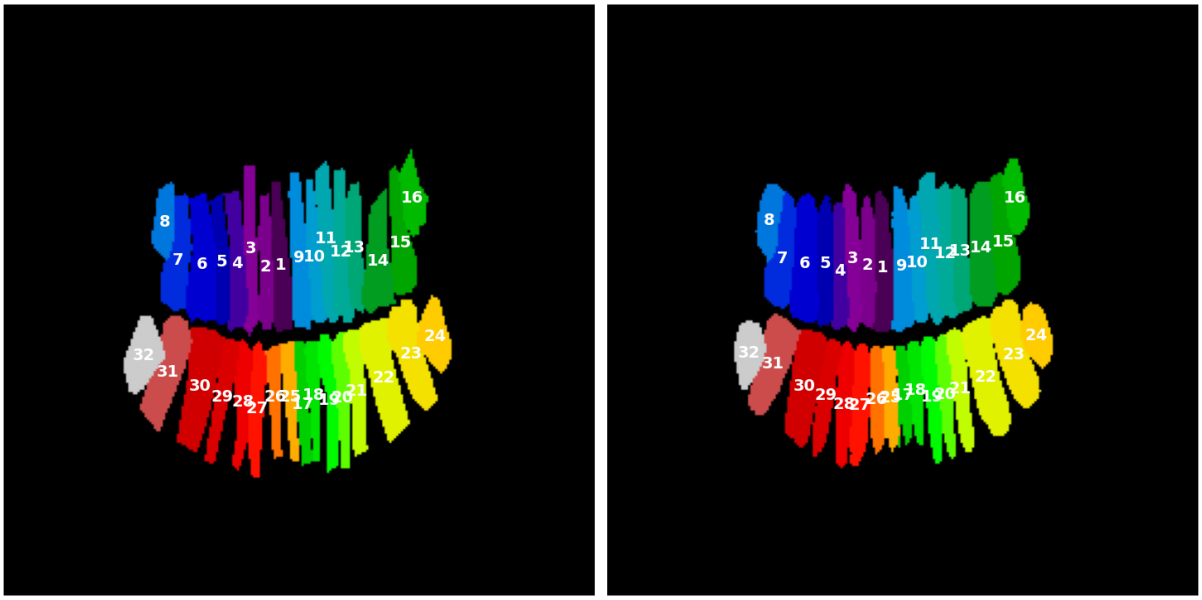

Figure 19: U-Net 32 Ground Truth vs. Prediction

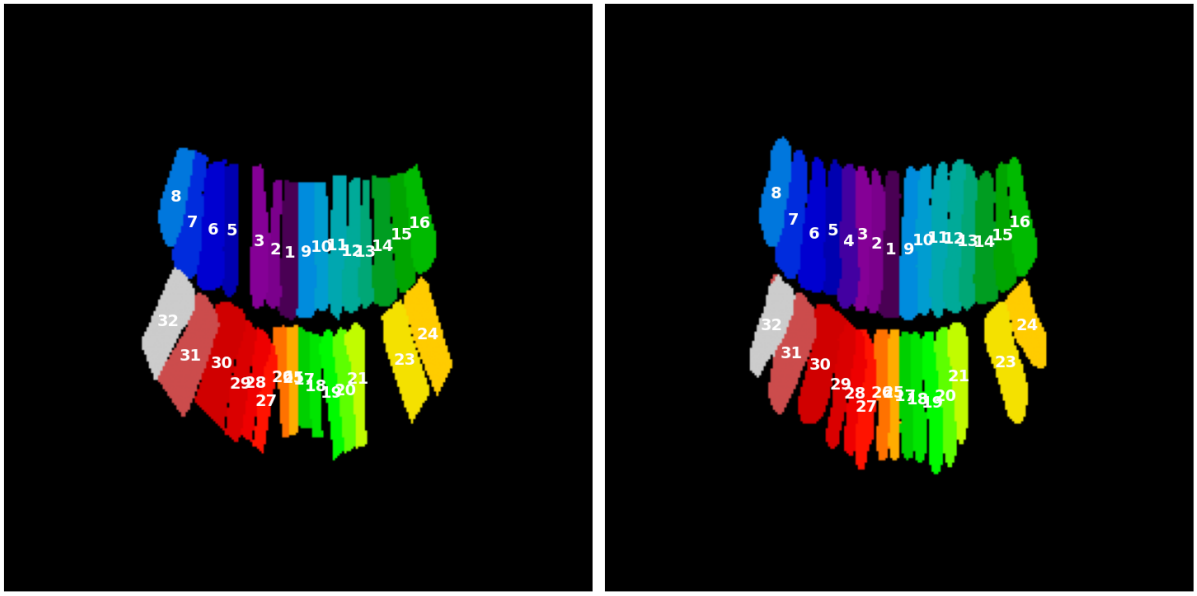

Figure 20: SE-U-Net 32 Ground Truth vs. Prediction

Quantitative Evaluation

The top-performing checkpoints were selected for evaluation, with metrics computed for Intersection over Union (IoU) and Dice coefficient.

Table 5: U-Net and SE-U-Net 32 Quantitative Evaluation

Model	Checkpoint	IoU	Dice
U-Net	epoch_166	**0.6176**	**0.7173**
U-Net	epoch_161	0.5910	0.6946
U-Net	epoch_120	0.5828	0.6875
SE-U-Net	epoch_148	**0.6150**	**0.7141**
SE-U-Net	last_epoch	0.6184	0.7171
SE-U-Net	epoch_201	0.6130	0.7109

Conclusion

U-Net and SE-U-Net provide high-quality segmentation masks that align closely with ground truth annotations, making them suitable for precise tooth boundary extraction in the Dentelligence pipeline. When integrated with the DINO–ResNet50 detection outputs, these segmentation results form a robust basis for accurate enumeration and subsequent disease detection.

4.3.2 Enumeration 9 Detection Results

For the enumeration 9 task, cropped quadrant images were used to segment **8 in-quadrant teeth** and **class 9** representing out-of-quadrant regions. This approach allowed the model to focus on local anatomical structures and reduce inter-quadrant variability. Two architectures—**U-Net** and **SE-U-Net**—were evaluated for their ability to perform accurate pixel-level segmentation of the cropped images.

Training Performance

Figures 22 and 21 present the training and validation loss curves for U-Net and SE-U-Net, respectively.

- **U-Net** showed a rapid decline in both training and validation loss during the initial epochs, followed by a plateau phase, maintaining stable generalization.
- **SE-U-Net** demonstrated similar convergence behavior but with slightly smoother validation curves, indicating better stability across the training process.

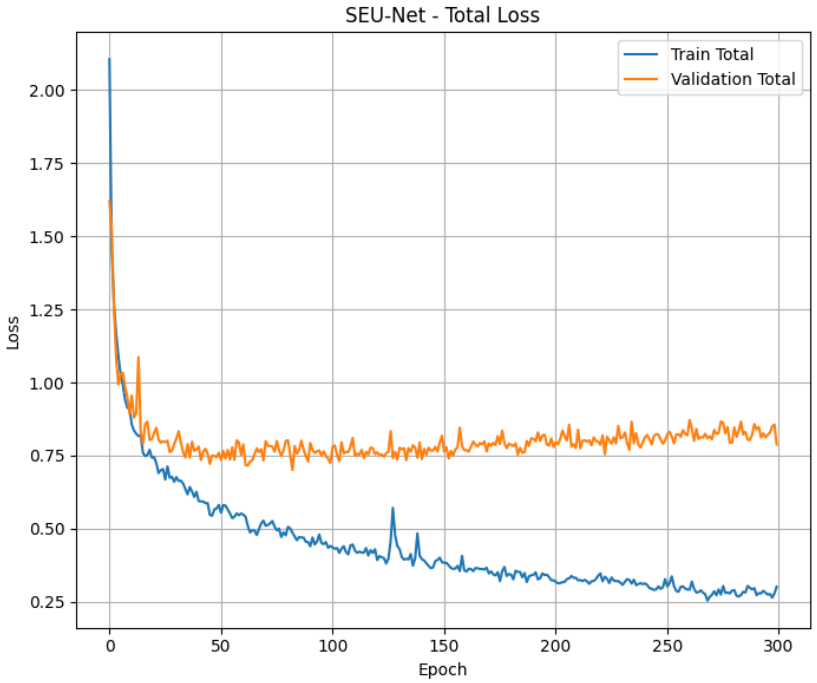

Figure 21: SE-U-net 9 Loss Curve Plot

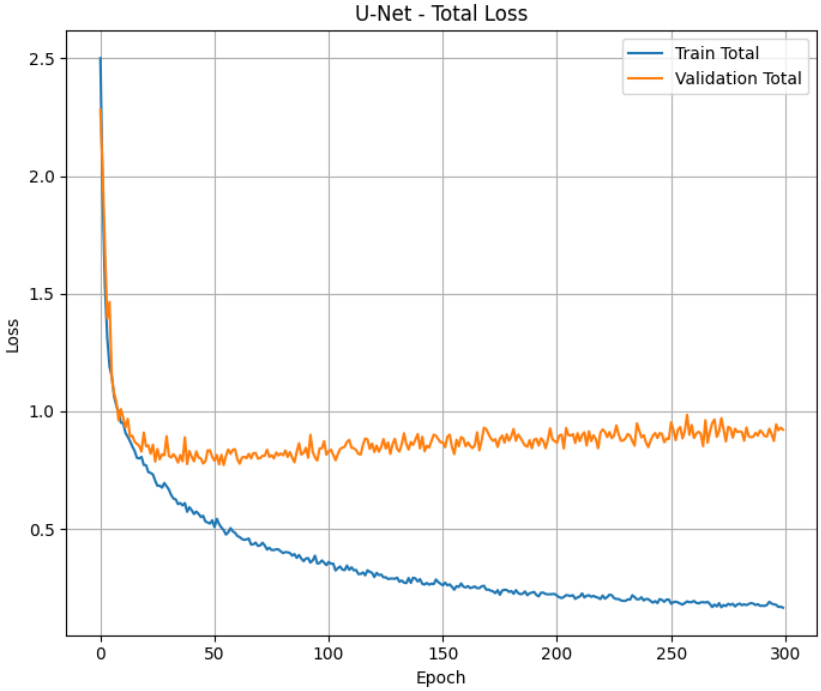

Figure 22: U-net 9 Loss Curve Plot

Qualitative Results

Figures 23 and 24 illustrate side-by-side comparisons of ground truth and predicted segmentation masks for U-Net and SE-U-Net. Both models successfully captured clear tooth boundaries, preserved correct enumeration labels, and handled out-of-quadrant areas (class 9) effectively. The SE-U-Net predictions exhibited slightly sharper boundary contours and reduced label bleeding between adjacent teeth.

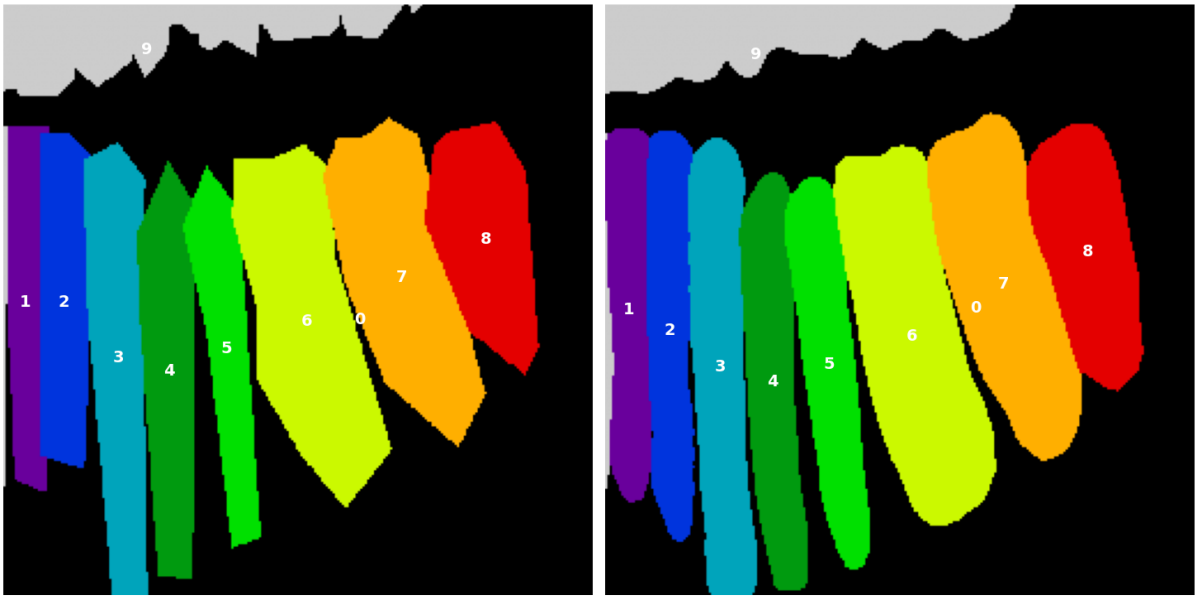

Figure 23: U-Net 9 Ground Truth vs. Prediction

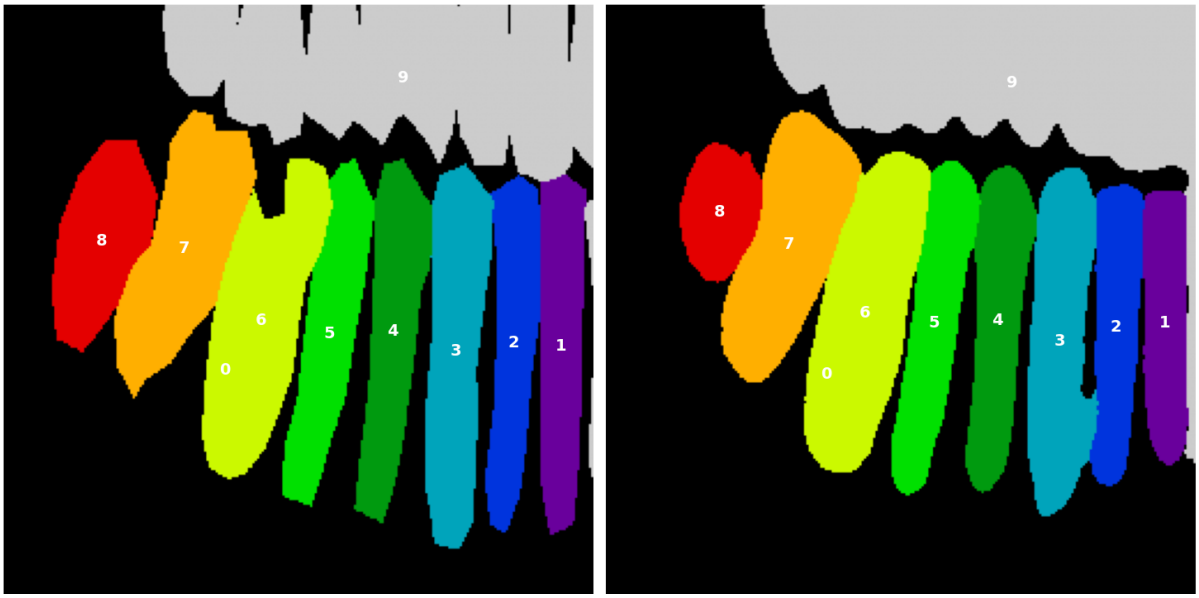

Figure 24: SE-U-Net 9 Ground Truth vs. Prediction

Quantitative Evaluation

The best-performing checkpoints for each model were selected, with Intersection over Union (IoU) and Dice coefficient metrics reported:

Table 6: U-Net and SE-U-Net 9 Quantitative Evaluation

Model	Checkpoint	IoU	Dice
U-Net	last_epoch	**0.6921**	**0.7882**
U-Net	epoch_38	0.6802	0.7774
U-Net	epoch_34	0.6755	0.7732
SE-U-Net	last_epoch	**0.6982**	**0.7949**
SE-U-Net	epoch_45	0.6982	0.7941
SE-U-Net	epoch_82	0.6974	0.7936

Conclusion

Both U-Net and SE-U-Net achieved strong performance in enumeration 9 segmentation, with SE-U-Net offering a modest but consistent advantage. The high-quality masks produced at this stage provide reliable inputs for downstream disease detection and result aggregation in the Dentelligence pipeline.

4.4 Disease Detection Results

DINO-Swin Backbone

The DINO model with a Swin Transformer backbone was used for disease detection on full panoramic X-ray images. The model aimed to detect and classify multiple dental conditions, including **Caries**, **Deep Caries**, **Periapical Lesions**, and **Impacted Teeth**.

Training Performance

Figure 25 shows the training loss curve across 100 epochs. The loss exhibits a steep initial decline during the first 20 epochs, followed by a stable plateau phase, indicating effective convergence without signs of overfitting.

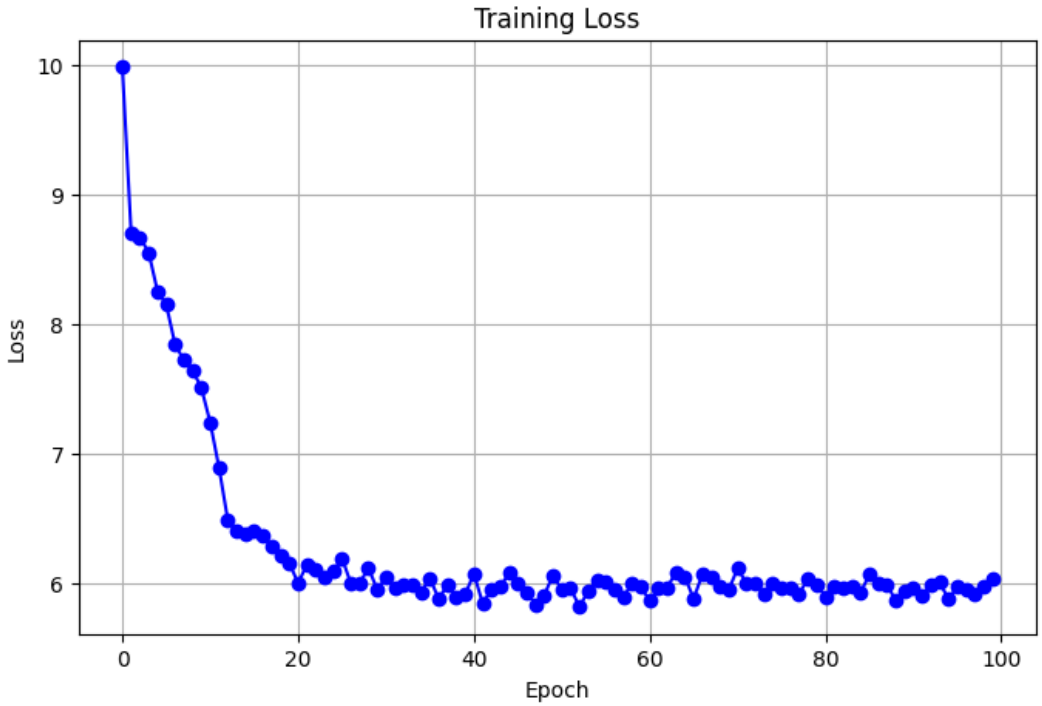

Figure 25: Dino-Swin Loss Curve Plot

Qualitative Results

Figures 26 and 27 present qualitative comparisons of ground truth annotations (green) versus model predictions (red) for two sample panoramic images.

- The first example (Figure 26) demonstrates the model's ability to detect multiple overlapping pathologies, such as co-occurring caries and deep caries in adjacent teeth.
- The second example (Figure 27) shows correct localization of impacted teeth and caries in different quadrants, although a few predictions had relatively low confidence scores (<0.5).

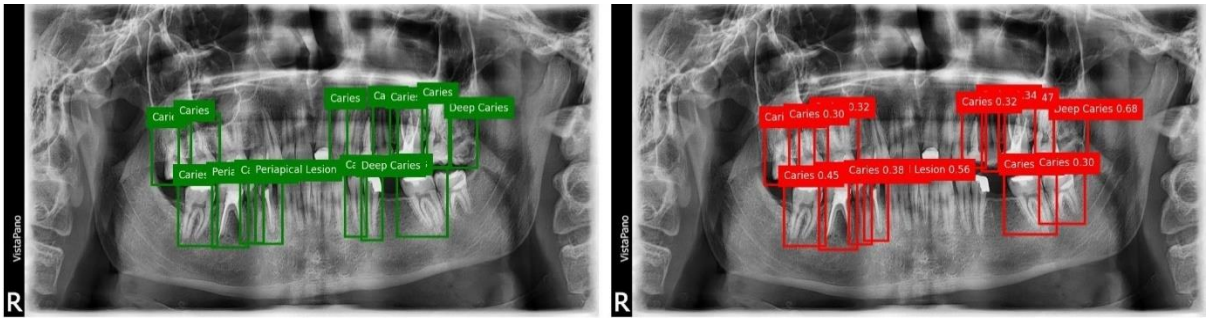

Figure 26: Dino-Swin Ground Truth vs. Predictions Ex.1

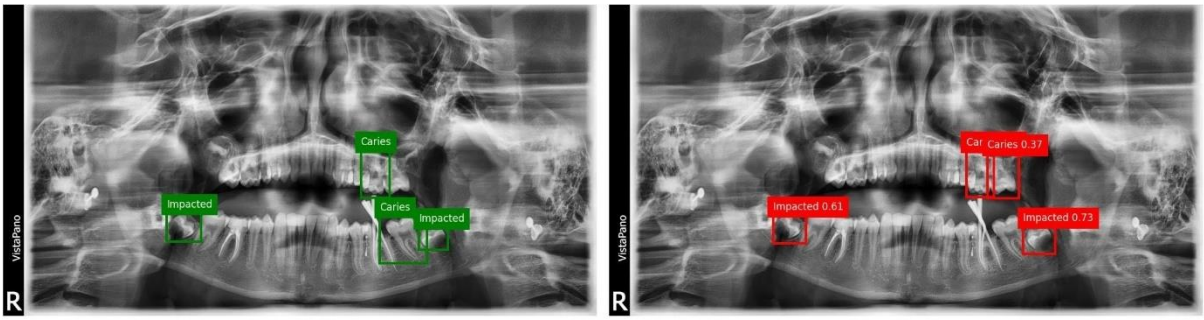

Figure 27: Dino-Swin Ground Truth vs. Predictions Ex.2

Quantitative Evaluation

The model's performance was evaluated using COCO-style metrics over the disease detection dataset. Key results include:

Table 7: Dino-Swin Quantitative Evaluation

Metric	Value
AP@[0.50:0.95] (all)	**0.534**
AP@0.50 (all)	**0.710**
AP@0.75 (all)	**0.654**
AP@[0.50:0.95] (medium)	**0.698**
AP@[0.50:0.95] (large)	**0.533**
AR@[0.50:0.95] (maxDets=1)	**0.324**
AR@[0.50:0.95] (maxDets=10)	**0.696**
AR@[0.50:0.95] (maxDets=100)	**0.742**
AR@[0.50:0.95] (medium)	**0.742**
AR@[0.50:0.95] (large)	**0.741**

Conclusion

The DINO-Swin disease detection model demonstrated robust performance in identifying a wide range of dental pathologies on panoramic radiographs. The qualitative and quantitative results confirm its suitability for integration into the Dentelligence pipeline, where it can contribute to automated disease screening and assist clinicians in diagnosis.

YOLOv8

The YOLOv8x model was employed for disease detection in panoramic dental X-rays, targeting four disease categories: **Impacted**, **Caries**, **Periapical Lesion**, and **Deep Caries**. Its one-stage architecture and anchor-free design allow for high-speed and accurate object detection, making it well-suited for real-time dental analysis.

Training Performance

The training loss curve (Figure 28) shows a steady decline in both training and validation box loss throughout the 100 epochs. This indicates stable learning and strong convergence, with no significant signs of overfitting. The consistent gap between the training and validation loss curves suggests good generalization capability across different data splits.

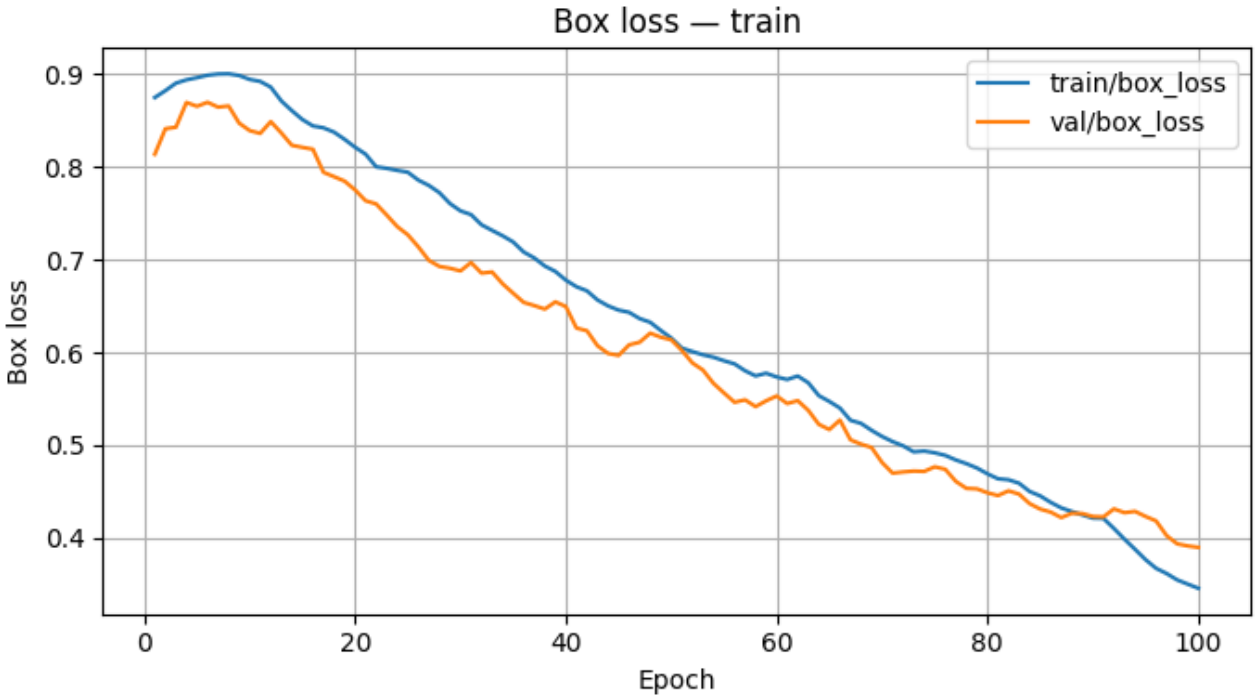

Figure 28: YOLOv8 Box Loss Curve Plot

Qualitative Results

Figure 29 presents sample ground truth (green) and predicted (red) bounding boxes for various dental diseases. YOLOv8 was able to detect and localize abnormalities with high confidence scores, often above 0.90. Predictions were closely aligned with ground truth annotations, showing strong localization accuracy and minimal false positives.

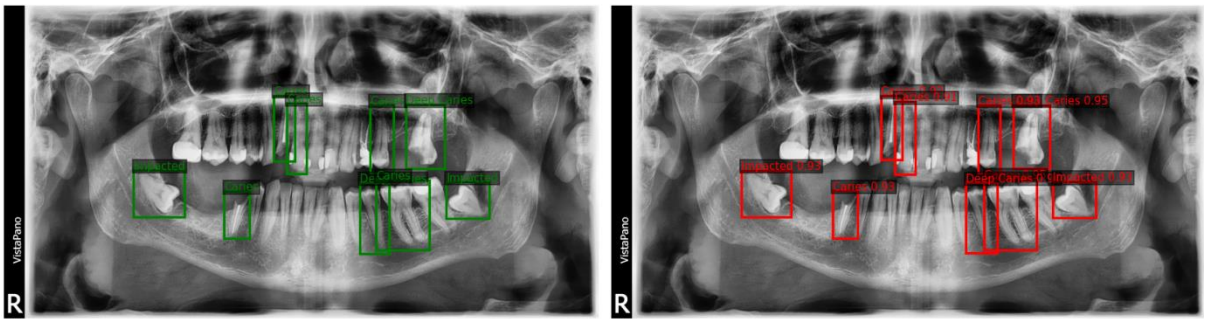

Figure 29: YOLOv8 Ground Truth vs. Predictions

Quantitative Evaluation

The model's performance was evaluated using COCO-style metrics. Results are summarized below:

- **Overall Metrics:**
  - mAP@[.50:.95]: **0.931**
  - mAP@50: **0.995**
  - mAP@75: **0.995**
  - Average Recall (AR): **0.992**
- **Class-wise AP@[.50:.95]:**
  - Impacted: **0.945**
  - Caries: **0.938**
  - Periapical Lesion: **0.913**
  - Deep Caries: **0.928**
- **Precision / Recall:**
  - Precision: **0.993**
  - Recall: **0.992**

Conclusion

The YOLOv8x model achieved outstanding performance across all disease categories, combining near-perfect precision and recall with rapid inference times. Its high AP values, particularly for **Impacted** and **Caries**, confirm its reliability for automated dental disease detection. Given its speed and accuracy, YOLOv8 is an excellent choice for integration into real-time clinical diagnostic workflows.

Weighted Boxes Fusion (WBF)

Weighted Boxes Fusion (WBF) was employed to combine predictions from the **DINO-Swin** and **YOLOv8x** models for dental disease detection. Unlike traditional non-maximum

suppression (NMS), WBF retains more information by merging overlapping detections based on their confidence scores. This allows for improved localization accuracy and recall, leveraging the strengths of both models in a complementary fashion.

Training Performance

Since WBF is a post-processing method, it does not involve training. Instead, it processes the inference outputs from both DINO-Swin and YOLOv8, adjusting bounding box coordinates and scores based on a weighted average. The aim is to refine detection quality without re-training the models.

Quantitative Evaluation

The WBF fusion results yielded significant performance improvements over individual models:

- **Overall Metrics:**
  - mAP@[.50:.95]: **0.823**
  - mAP@50: **0.964**
  - mAP@75: **0.945**
- **Area-specific Performance (mAP@[.50:.95]):**
  - Medium: **0.878**
  - Large: **0.822**
- **Average Recall (AR):**
  - AR@1: **0.442**
  - AR@10: **0.876**
  - AR@100: **0.877**
  - Medium: **0.883**
  - Large: **0.876**

Conclusion

The integration of DINO-Swin and YOLOv8x via Weighted Boxes Fusion led to a robust improvement in detection accuracy and consistency. With an mAP@50 of **96.4%** and mAP@[.50:.95] of **82.3%**, WBF effectively leveraged the high precision of YOLOv8x and the strong recall of DINO-Swin. This fusion approach demonstrates clear benefits for automated dental disease detection, offering a reliable balance between localization accuracy and detection confidence.

Table 8: Disease Detection Model Performance Comparison

Metric	DINO-Swin	YOLOv8x	WBF (DINO-Swin + YOLOv8x)
mAP@[.50:.95]	0.534	0.931	**0.823**
mAP@50	0.710	0.995	**0.964**
mAP@75	0.654	0.995	**0.945**
Medium Objects mAP@[.50:.95]	0.698	0.938	**0.878**
Large Objects mAP@[.50:.95]	0.533	0.945	**0.822**
AR@1	0.324	0.992	**0.442**
AR@10	0.696	0.992	**0.876**
AR@100	0.742	0.992	**0.877**
Medium Objects AR	0.742	0.992	**0.883**
Large Objects AR	0.741	0.992	**0.876**

Notes:

- YOLOv8x achieves the highest mAP and AR overall, especially at IoU thresholds of 0.50 and 0.75.
- WBF does not surpass YOLOv8 in absolute mAP, but it combines the strengths of both models, leading to more stable predictions and fewer duplicate detections.
- DINO-Swin has lower scores but provides complementary detections that improve WBF results.

CHAPTER 5 FUTURE WORK & CONCLUSIONS

5.1 Introduction

The Dentelligence System was developed to detect, segment, enumerate, and classify abnormal teeth in panoramic dental X-rays using a collection of specialized deep learning models. Each model was trained and evaluated independently to ensure a thorough understanding of its performance in isolation. The results in Chapter 4 demonstrated the individual strengths of DiffusionDet for quadrant detection, DINO-ResNet50 for enumeration, U-Net and SE-U-Net for segmentation, and DINO-Swin with YOLOv8 for disease detection, with Weighted Boxes Fusion (WBF) providing enhanced ensemble outputs for pathology identification. While these results validate the effectiveness of each model, the goal of Dentelligence is to deliver a fully automated, end-to-end diagnostic pipeline in which all models operate in a connected sequence. This chapter outlines the conclusions drawn from the current work and proposes future directions, with an emphasis on integrating these models into a unified pipeline for improved accuracy, consistency, and clinical usability.

5.2 Conclusions

The current stage of the Dentelligence System achieved several key milestones:

- **High-performance individual models** were developed for each subtask in dental image analysis, achieving strong accuracy in detection, segmentation, and disease classification tasks.
- **Modular design** allowed each model to be optimized independently, enabling a deep understanding of their behavior and limitations.
- **Advanced ensemble techniques** such as Weighted Boxes Fusion demonstrated that combining outputs from complementary architectures can yield more robust results than relying on a single model.

However, the current setup operates each model **independently**, without a continuous data flow between them. While this facilitated targeted performance tuning, it limits the system's ability to leverage context from earlier stages to improve downstream predictions.

5.3 Future Work

5.3.1 Full Model Integration into a Unified Pipeline

Current Status:

At present, each model in Dentelligence is executed and evaluated independently. This provided a comprehensive performance profile for each architecture but does not represent the final envisioned form of the system — an automated, fully connected inference pipeline.

Proposed Integration Workflow:

The next development phase will focus on constructing a sequential pipeline, where the output of one model directly serves as the input to the next:

1. **Quadrant Detection** – DiffusionDet will identify and localize the four dental quadrants in the panoramic X-ray.
2. **Quadrant Cropping & Enumeration**
    - Cropped quadrants will be fed into Enumeration9 U-Net/SE-U-Net models for high-resolution, localized tooth segmentation.
    - The full panoramic image will be processed by Enumeration32 models to maintain a global numbering reference.
3. **Tooth Detection & Refinement** – DINO-ResNet50 will refine bounding boxes for each tooth, informed by segmentation boundaries.
4. **Disease Detection** – Quadrant or tooth patches will be analyzed by both DINO-Swin and YOLOv8 models for pathology classification.
5. **Fusion & Label Matching** – Weighted Boxes Fusion will merge disease predictions, and label matching will ensure correct FDI tooth numbering with associated diagnoses.
6. **Visualization & Reporting** – The final pipeline will output annotated images and structured reports ready for integration into clinical workflows.

Advantages of a Unified Pipeline:

- **Automation** – Fully automated progression from raw X-ray to final diagnosis.
- **Improved Accuracy** – Context from earlier models will guide later stages, reducing inconsistencies.
- **Standardized Output** – Ensures consistent numbering and labeling across all detections.

- **Enhanced Clinical Readiness** – Streamlined workflow suitable for real-time or near-real-time deployment.

Challenges to Address:

- **Error Propagation** – Mistakes in early stages may cascade into downstream predictions; strategies such as confidence-based reprocessing must be developed.
- **Latency Optimization** – Model compression or parallel execution may be required for acceptable inference speed in clinical settings.
- **Output Standardization** – Ensuring intermediate results use a unified data format for seamless inter-model communication.

5.3.2 Dataset and Model Enhancements

- **Expand Dataset Coverage** – Incorporate more diverse patient demographics, imaging conditions, and additional dental disease categories.
- **Semi-Supervised Learning** – Leverage partially annotated datasets to reduce the reliance on fully labelled data.
- **Lightweight Architectures** – Explore smaller, faster models for deployment on lower-spec clinical hardware.

5.3.3 Deployment and Integration

- Develop a graphical user interface (**GUI**) for dentists to interact with the system.
- Integrate with hospital information systems (**HIS**) and electronic health records (**EHR**).
- Enable cloud-based tele-dentistry functionality for remote diagnosis and consultations.

5.4 Final Remarks

The Dentelligence System has proven the value of a modular, high-performance AI approach for dental diagnostics. While the independent evaluation of each model was a crucial step, the next phase must focus on **pipeline integration** to create a truly automated and cohesive system. By connecting these models into a single inference flow, Dentelligence can deliver faster, more consistent, and more clinically relevant outputs, advancing toward real-world deployment in dental practices and education.

REFERENCE

[1] Tuzoff, D. V., Tuzova, L. N., Bornstein, M. M., Krasnov, A. S., Kharchenko, M. A., Nikolenko, S. I., Sveshnikov, M. M., & Bednenko, G. B. (2019). Tooth detection and numbering in panoramic radiographs using convolutional neural networks. *Dentomaxillofacial Radiology, 48*(3), 20180051.

[2] Schwendicke, F., Samek, W., & Krois, J. (2020). Artificial intelligence in dentistry: Chances and challenges. *Journal of Dental Research, 99*(7), 769–774.

[3] Sukegawa, S., Yoshii, K., Hara, T., Matsuyama, T., Yamashita, K., Nakano, K., Takabatake, K., Kawai, H., Nagatsuka, H., & Furuki, Y. (2021). Multi-task deep learning model for classification of dental implant brand and treatment stage using dental panoramic radiograph images. *Biomolecules, 11*(6), 815.

[4] Hamamci, I. E., Er, S., Simsar, E., Sekuboyina, A., Gundogar, M., Stadlinger, B., Mehl, A., & Menze, B. (2023). Diffusion-based hierarchical multi-label object detection to analyze panoramic dental X-rays. *arXiv preprint arXiv:2303.06500*.

[5] Dhar, M. K., Deb, M., Madhab, D., & Yu, Z. (2023). A deep learning approach to teeth segmentation and orientation from panoramic X-rays. *arXiv preprint arXiv:2311.01643*.

[6] Wang, Y.-C. C., Chen, T.-L., Vinayahalingam, S., Wu, T.-H., Chang, C. W., Chang, H. H., Wei, H.-J., Chen, M.-H., Ko, C.-C., Anssari Moin, D., van Ginneken, B., Xi, T., Tsai, H.-C., Chen, M.-H., Hsu, T.-M. H., & Chou, H. (2025). Artificial intelligence to assess dental findings from panoramic radiographs: A multinational study. *arXiv preprint arXiv:2502.10277*.

[7] Hamamci, I. E., Er, S., Simsar, E., Yuksel, A. E., Gultekin, S., Ozdemir, S. D., Yang, K., Li, H. B., Pati, S., Stadlinger, B., Mehl, A., Gundogar, M., & Menze, B. (2023). *Dental Enumeration and Diagnosis on Panoramic X-rays Challenge (DENTEX)*. MICCAI 2023 Grand Challenge.

[8] Glick, M., Monteiro da Silva, O., Seeberger, G. K., Xu, T., Pucca, G., Williams, D. M., Kess, S., Eiselé, J.-L., & Sèverin, T. (2012). FDI Vision 2020: Shaping the future of oral health. *International Dental Journal, 62*(6), 278–291.

[9] Azizi, A., Azizi, M., & Nasri, M. (2023). *Artificial intelligence techniques in medical imaging: A systematic review*. *International Journal of Online and Biomedical Engineering (iJOE)*, *19*(17), 66–97.

[10] 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* (pp. 234–241). Springer.

[11] Haghanifar, A., Majdabadi, M. M., & Ko, S.-B. (2021). PaXNet: Dental caries detection in panoramic X-ray using ensemble transfer learning and capsule classifier. *arXiv preprint arXiv:2012.13666*.

[12] Almalki, A., & Latecki, L. J. (2022). Self-supervised learning with masked image modeling for teeth numbering, detection of dental restorations, and instance segmentation in dental panoramic radiographs. *arXiv preprint arXiv:2210.11404*.

[13] Helli, S. S., & Hamamci, A. (2022). Tooth instance segmentation on panoramic dental radiographs using U-Nets and morphological processing. *Düzce University Journal of Science & Technology, 10*, 39–50.

[14] Juyal, A., Tiwari, H., Singh, U. K., Kumar, N., & Kumar, S. (2023). Dental caries detection using Faster R-CNN and YOLO V3. *ITM Web of Conferences, 53*, 02005

[15] Mărghinean, A. C., Mureşanu, S., Hedeşiu, M., & Diosan, L. (2024). Teeth segmentation and carious lesions segmentation in panoramic X-ray images using CariSeg, a networks' ensemble. *Heliyon, 10*, e30836.

[16] Adnan, N., Ahmed, S. M. F., Das, J. K., Aijaz, S., Sukhia, R. H., Hoodbhoy, Z., & Umer, F. (2024). Developing an AI-based application for caries index detection on intraoral photographs. *Scientific Reports, 14*, 26752.

[17] Zhicheng, H., Yipeng, W., & Xiao, L. (2024). Deep learning-based detection of impacted teeth on panoramic radiographs. *Biomedical Engineering and Computational Biology, 15*, 1–6.

[18] Beser, B., Reis, T., Berber, M. N., Topaloglu, E., Gungor, E., Kılıc, M. C., Duman, S., Çelik, Ö., Kuran, A., & Bayrakdar, I. Ş. (2024). YOLO-V5 based deep learning approach for tooth detection and segmentation on pediatric panoramic radiographs in mixed dentition. *BMC Medical Imaging, 24*, 172.

[19] Özçelik, S. T. A., Üzen, H., Şengür, A., Fırat, H., Türkoğlu, M., Çelebi, A., Gül, S., & Sobahi, N. M. (2024). Enhanced panoramic radiograph-based tooth segmentation and identification using an attention gate-based encoder–decoder network. *Diagnostics, 14*(23), 2719.

[20] Kurt, A., Günaçar, D. N., Şılbır, F. Y., Yeşil, Z., Bayrakdar, İ. Ş., Çelik, Ö., Bilgir, E., & Orhan, K. (2024). Evaluation of tooth development stages with deep learning-based artificial intelligence algorithm. *BMC Oral Health, 24*, 1034.

[21] Yaxin, G. G., & Hargreaves, C. A. (2025). Dental teeth X-ray image classification using AI. ***Series of Clinical and Biomedical Research, 2*****(1), 1–19.**

[22] Yavsan, Z. S., Orhan, H., Efe, E., & Yavsan, E. (2025). Diagnosis of approximal caries in children with convolutional neural networks-based detection algorithms on radiographs: A pilot study. ***Acta Odontologica Scandinavica, 84*****(1), 18–25.**

[23] Hua, Y., Chen, R., & Qin, H. (2025). YOLO-DentSeg: A lightweight real-time model for accurate detection and segmentation of oral diseases in panoramic radiographs. ***Electronics, 14*****(4), 805.**

[24] Zhong, W., Ren, X., & Zhang, H. (2025). Automatic X-ray teeth segmentation with grouped attention. ***Scientific Reports, 15*****, 64.**

[25] Ayhan, B., Ayan, E., Karadağ, G., & Bayraktar, Y. (2025). Evaluation of caries detection on bitewing radiographs: A comparative analysis of the improved deep learning model and dentist performance. ***Journal of Esthetic and Restorative Dentistry, 0*****(0), 1–13.**

[26] Ovi, M. O., Sanjana, M., Fahad, F., Runa, M., Rothy, Z. T., Pias, T. S., Islam, A. M. T., & Prodhan, R. A. (2025). Enhanced pediatric dental segmentation using a custom SegUNet with VGG19 backbone on panoramic radiographs. ***arXiv preprint arXiv:2503.06321*****.**

[27] Bayati, M., Alizadeh Savareh, B., Ahmadinejad, H., & Mosavat, F. (2025). Advanced AI-driven detection of interproximal caries in bitewing radiographs using YOLOv8. ***Scientific Reports, 15*****, 9628.**

[28] Ruder, S. (2017). An overview of multi-task learning in deep neural networks. ***arXiv preprint arXiv:1706.05098*****.**

[29] Cai, Z., Yang, X., Li, X., Luo, X., Li, X., Shen, L., Meng, H., & Deng, Y. (2024). SSAD: Self-supervised auxiliary detection framework for panoramic X-ray-based dental disease diagnosis. ***arXiv preprint arXiv:2406.13963*****.**

[30] Budagam, D., Kumar, A., Ghosh, S., Shrivastav, A., Imanbayev, A. Z., Akhmetov, I. R., Kaplun, D., Antonov, S., Rychenkov, A., Cyganov, G., & Sinitca, A. (2024). Instance segmentation and teeth classification in panoramic X-rays. ***arXiv preprint arXiv:2406.03747*****.**

[31] Huang, S.-C., Pareek, A., Jensen, M., Lungren, M. P., Yeung, S., & Chaudhari, A. S. (2023). Self-supervised learning for medical image classification: A systematic review and implementation guidelines. ***npj Digital Medicine, 6*****(1), 74.**

[32] Wang, J., & Perez, L. (2017). ***The effectiveness of data augmentation in image classification using deep learning*****. arXiv preprint arXiv:1712.04621.**

[33] Li, H., Zhou, J., Zhou, Y., Chen, Q., She, Y., Gao, F., Xu, Y., Chen, J., & Gao, X. (2021). An interpretable computer-aided diagnosis method for periodontitis from panoramic radiographs. ***Frontiers in Physiology, 12*****, 655556.**

[34] Yaseen, M. (2024). ***What is YOLOv8: An in-depth exploration of the internal features of the next-generation object detector*****. arXiv preprint arXiv:2408.15857.**

[35] 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*** **(pp. 234–241). Springer.**

[36] Nistelrooij, N. van, El Ghoul, K., Xi, T., Saha, A., Kempers, S., Cenci, M., Loomans, B., Flügge, T., van Ginneken, B., & Vinayahalingam, S. (2024). Combining public datasets for automated tooth assessment in panoramic radiographs. ***BMC Oral Health, 24*****(387).**

[37] Zhang, H., Gao, B., Sun, J., Li, J., Yuan, Y., Liu, Z., Lin, D., & Qiao, Y. (2022). ***DINO: DETR with improved deNoising anchor boxes for end-to-end object detection*****. arXiv preprint arXiv:2203.03605.**

[38] Hu, J., Shen, L., & Sun, G. (2018). ***Squeeze-and-excitation networks*****. In** ***Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*** **(pp. 7132–7141).**

[39] Chen, S., Sun, P., Song, Y., & Luo, P. (2023). ***DiffusionDet: Diffusion model for object detection*****. arXiv preprint arXiv:2211.09788.**

[40] Solovyev, R., Wang, W., & Gabruseva, T. (2021). ***Weighted boxes fusion: Ensembling boxes from different object detection models*****. arXiv preprint arXiv:1910.13302.**