CHAPTER 2

LITERATURE REVIEW

In recent years, deep learning has shown great significance in medical imaging, especially in segmentation tasks but not limited to dental image segmentation on X-ray images. Conventional and deep learning approaches have been used to accomplish the task of dental image segmentation. The primary objective of this chapter is to provide a detailed survey of the existing methods for dental imaging, understand the current research trends and find out the possible research gaps.

2.1 Literature Review for Dental Image Segmentation

2.1.1 Conventional Approaches

Many traditional approaches have been developed throughout the years, but their usefulness has been restricted by poor image quality and suspicious tooth borders. The conventional approaches for segmenting the dental panoramic images can be roughly categorized into region-based, cluster-based, threshold based, boundary based and watershed-based approaches[2].

Lurie et al.[6] used region-based segmentation approach to detect osteoporosis and osteopenia cancerous bone by segmenting panoramic X-ray images of the teeth. Modi et al. [7] used region growing segmenting and canny edge detector approach for finding the region of interest for tooth isolation and gap valley. Rahma et al. [8] proposes automatic marker applied on region growing algorithm for segmenting bitewing X-ray images to deal with issue of spatial size images. Arifin et al. [9] deals with issue of low contrast dental panoramic images and proposed a region-merging strategy for image segmentation using discriminant analysis. However, these approaches are sensitive to noise and time consuming.

Son et al.[10] applied semi-supervised fuzzy clustering algorithm to dental X-ray image segmentation to determine the common teeth boundaries. The author removes the background by using Otsu method and choose dental structure by fuzzy c-means algorithm. This approach focuses on periapical and bitewing X-ray images. Alsmadi et al. [11] purposed a fully automatic method for detecting jaw lesion based on hybrid fuzzy c-means. The panoramic X-ray images were segmented to detect jaw lesion region. These methods are sensitive to missing teeth.

Ajaz et al.[12] segmented the dental panoramic radiographs for human identification. The author applied mathematical morphological operations followed by thresholding to extract features from panoramic radiographs. Cameriere et al.[13] uses threshold-based segmentation method on periapical images to automatically estimate the age. [14],[15]used hybrid threshold methods to segment the digital dental radiographs for automatic localization of bone loss and wisdom tooth. Bruellmann et al.[16] proposed an algorithm based on fast fourier filter to automatically detect endodontic file in periapical dental X-rays. To detect enamel caries and proximal caries, histogram-based approaches were used to choose grey level threshold on digital periapical radiographs[17]. The filter bank and adaptive thresholding with local thresholding was applied in panoramic X-ray images to exploit statistical features for segmenting teeth[18]. The threshold connected component analysis method for segmenting panoramic radiograph is used by [19] to extract the region of interest of teeth. The threshold-based segmentation methods suffer in identifying root boundaries. Sobel and Canny edge segmentation was applied by[20] on some dental panoramic radiographs to segment the teeth area. Sobel segmentation method was able to segment all teeth while canny suffers on incisors.

The level set methods [21],[22],[23] were constructed using different boundary-based methods to perform tooth segmentation. These methods can deal with sharp curves and corners, variations of teeth shapes but their computation complexity is high and often trapped in local minima. Active contour model without edges was proposed by [24] for accurate segmentation of dental X-ray images. The algorithm was implemented on GPU for fast computation. Hasan et al. [25] proposed an approach for automatic segmenting of jaw form dental panoramic radiograph using gradient vector flow snakes to reduce time. Teeth segmentation was done using mathematical morphology method by[1], to improve the performance grayscale contrast stretching transformation was used. Only bitewing and periapical images were focused in this method. Lira et al.[26] proposed a segmentation method for tooth recognition based on mathematical morphology operators combined with thresholding and snake models. Harandi et al. [27] combined the morphological operations and active contour on teeth X-ray images to segment upper and lower jaws. The performance of this model relies on its initialization.

To summarize, conventional approaches rely primarily on matching between prior image features and are typically limited to certain scenarios. Furthermore, because these approaches incorporate specialised information, generalising them to similar scenarios is not an easy task.

Table 2.1 Summary of conventional approaches for dental image segmentation

Author Year	**Objective**	**Methodology**	**Image Modality**	**Remark**
Li et al. 2012[28]	To separate each tooth and prevent over-segmentation problem	Watershed Algorithm based on mathematical morphology	Bitewing	Improves over-segmentation. Results of segmentation are not perfect. Adjusting parameters is time consuming
Radhiyah et al.[29] 2016	Pre-processing techniques for dental image segmentation.	Gaussian and Histogram equalization for watershed segmentation algorithm.	Panoramic X-ray	Results depends on output of watershed segmentation. Weak segmented mask of tooth roots.
Alsamadi et. Al. 2018 [11]	Automatic method for segmentation of jaw lesions	Hybrid Fuzzy c-means and neutrosophic approach	Periapical	Segment jaw lesion for low-contrast images. Tooth boundaries need to be clearer.
Son et al. 2016 [10]	Segmentation of dental images	Semi supervised fuzzy clustering with Otsu's method	Bitewing/ Panoramic	Clear segmentation results but dental features were not used in clustering process. Parameters values range should be derived properly.
Lin et al. 2012[30]	Gum lesion detection by segmenting dental X-ray	Level-set methods to segment teeth	Periapical	Robust to illumination variation. Tooth boundaries are blurry.
Rad et al. 2013 [31]	Dental X-ray image segmentation and feature extraction of teeth	Level-set and Contour methods	Bitewing	Features were extracted by texture statistics technique. Evaluation of segmentation method is not reported.
Ali et al. 2015 [24]	Fast computation for accurate segmentation of dental X-ray	Active Contour without edges	Bitewing	Used GPU for segmentation algorithm. Settles the issue of weak gradient between teeth and background without boundary. Suffers from slow convergence because of local minima values.

Li et al. 2007[32]	Semi-automatic lesion detection	Level set and SVM	Periapical	Fast and robust segmentation. Sensitive to missing values.
Mao et al.[33] 2018	Segmentation of dental X-ray images	Threshold Segmentation	Periapical/ Panoramic	Considers the grayscale pixels that belongs to target object. Suffers when there is high variation in pixel intensity of dominant object.
Modi et al. 2011 [7]	To find region of interest for tooth isolation and gap valley	Region growing segmentation and Canny edge	Bitewing	Clear isolation of neighbouring tooth from bones and clear segmentation of upper and lower jaws.
Lurie et al. 2012 [6]	To detect osteoporosis and osteopenia cancer bone	Region based segmentation	Panoramic X- rays	Results can be improved with large population. Difficult in dealing with blurry images.
Rahma et al. 2017 [8]	Automatic marker for Dental panoramic radiograph segmentation	Region growing algorithm	Bitewing	Deals with the issue of spatial size images. Sensitive to noise and time consuming.
Arifin et al. 2017 [9]	Interactive image segmentation.	Region Merging using discriminant analysis	Root Canal therapy X-ray	Dealt with the issue of low-contrast dental panoramic images.
Ajaz et al. 2013 [12]	Human Identification using dental features	Morphological operations and Thresholding	Panoramic X-ray	Produces good results while only using dental features
Cameriere et al. [13]	Automatically estimate age.	Threshold based segmentation	Periapical	Modifications of parameter is required for optimal result which should be known priorly.
Lin et al. 2015 [14]	Alveolar bone loss localization	Threshold Segmentation	Periodontitis Radiograph	Hybrid features were properly extracted. Only limited features were extracted. Suffers from blurry root boundaries.
Amer et al. 2015 [15]	Efficient Segmentation of Dental Images	Hybrid Threshold Segmentation	Panoramic radiographs	Extracts wisdom teeth. Heavily rely on pre-processing.
Bruellmann et al. 2016 [16]	Automatically detect endodontic file	Fast Fourier Filter Algorithm	Digital Dental radiographs	Suffers from presence of soft

				tissue, metals and artefacts.
Tikhe et al. 2016 [17]	To detect enamel carries and proximal carries	Gray level threshold	Periapical radiographs	Deals with over-segmentation issue. Rely on pre-processing.
Indraswari et al. 2015 [18]	To exploit statistical features for segmenting teeth	Filter bank and Adaptive Threshold	Panoramic X-ray	Suffers in identifying tooth boundaries.
Majanga et al. 2021 [29]	Dental Image segmentation to extract ROI of teeth	Threshold based segmentation	Panoramic X-ray	Rely on pre-processing which affects the performance.
Razali et al. 2014 [20]	Segmentation for dental age assessment	Canny and Sobel Edge segmentation	Panoramic radiograph	Sobel method was able to segment all teeth while canny suffers on incisors.

2.1.2 Deep Learning Techniques

In recent years, due to the introduction of deep learning in dentistry, automatic dental image segmentation has gain significance attention from researchers. Several deep models have been proposed by researchers for segmenting various dental imaging modalities. In this section, a detailed review of recent deep segmentation models is presented, and they are broadly grouped into five categories according to their deep architecture.

2.1.2.1 Convolutional Neural Network (CNN)

CNN is the most commonly used neural network for many computer vision tasks, including segmentation. In dental image segmentation also, CNN is used by various researchers. CNN consist of multiple layers of convolutional filters followed by nonlinear activation functions, which can capture both local and global spatial information. The final layer of the CNN outputs a segmentation map that assigns a label to each pixel in the input image.

Yang et al.[34] proposed a DCNN (Deep Convolutional Neural Network) with level-set methods for accurate and automatic tooth segmentation. The position of the dental pulp was detected by using a deep convolutional network. Active contours and level set methods were used to provide prior shape information of the tooth. The model was evaluated on 512 CBCT images. Boudraa et al.[35] proposed a deep hybrid model for segmenting 3D dental images. A decimate compression technique was used to reduce data size and faster processing time. The CNN model is then utilized to segment dental images into 15 different classes. The training and testing were done on 95 3D dental images from two different datasets.

Chandrashekar et al.[36] proposed an ensemble model based on collaborative learning for the automatic segmentation of tooth. The authors used the Mask RCNN (Region-based Convolutional Neural Network) and FCNN (Fully convolutional neural network) for segmentation and tooth identification. Initially, both models have trained individually then the aggregation of segmentation and identification was done by using collaborative learning. The experiment was conducted on 1500 panoramic images, and results showed that the collaborative model performs better than individual models. Lin et al.[37] proposed a lightweight deep learning methodology by using the knowledge distillation process for segmenting panoramic dental X-rays. The proposed model was based on conventional CNN with the knowledge consistency. The methodology just required 0.33 million parameters, it has a scope for improvement for dice score and jaccard index score. Li et al.[38] proposed a self-attention network for the automatic segmentation of dental plaque. The main motivation of the authors was to accurately segment dental plaque without using the medical dyeing reagent. CNN architecture with self-attention was developed to obtain local and global features and fuses them to obtain critical fused information. The experiment was conducted on 2884 labelled endoscopic images. The limitation of this approach is that it cannot deal with images having unknown variations.

2.1.2.2 U-Net

U-Net[39] neural network has shown a remarkable performance in the field of medical segmentation, including dental image segmentation. Though U-Net is based on the encoder and decoder structure, in this study, it is considered as a separate category. The encoder performs down-sampling to capture global contextual information, while the decoder performs up-sampling to recover the resolution of the original image. The skip connections enable the decoder to reuse the feature maps from the encoder, allowing the network to better preserve fine-grained details.

Nguyen et al.[40] segmented the alveolar bone in the ultrasound images by using U-Net architecture. The U-Net architecture was trained with VGG16[41] and ResNet16[42] as its backbone, respectively. The total size of the dataset was 1100 ultra sound images, while after data augmentation, the size increased to 2100. Zheng et al.[43] modified the Dense U-Net architecture and integrates oral-anatomical constrained information with the model for segmenting CBCT images. To reduce the computational cost of pixel wise constraints, mean-field approximation and variational inference was used. The experiment was conducted on 20 patients with periapical lesion and slices of 100 CBCT images. Besson et al.[44] utilized two tools for automatic root canal and crown segmentation. The first tool is RootCanalSeg which is based on the U-Net architecture, while the second tool is DentalModelSeg which is based on ResNet[42] architecture. The experiment was carried out on 80 CBCT scans and 80 mandibular digital dental models.

Zhu et al.[45] proposed a deep model named CariesNet for segmenting caries from oral panoramic radiographs. The deep model is based on U-Net architecture with the axial attention module. The objective is to segment three types of caries, shallow, moderate and deep, from the panoramic images. The dataset was developed by the authors and contained 1159 oral images, from which 3127 caries were labelled. The limitation is that for moderate level the performance is low. Nishitani et al.[32] proposed a computer-aided tool for segmentation of dental X-rays. The U-Net architecture was used with loss function focusing on the tooth edges. The model was trained and tested on 162 panoramic X-rays. The loss was calculated on both the entire image and on tooth edge. The results had shown that it performs better than conventional loss function. Li et al.[46] proposed the deep model based on U-Net structure for automatic tooth root segmentation in which the encoder and decoder were replaced by group transformers to reduce computational cost. The deep model also consists hybrid architecture of convolutional layers and transformers to get rid of pre training weights. The Fourier Descriptor was used as loss function to preserve the shape information. The experiment was conducted on 248 root canal therapy X-ray images.

Hou et al.[47] proposed a modified U-Net architecture for segmenting dental panoramic radiographs. The authors addressed the issues of blurred teeth boundaries, low contrast between alveolar bone and teeth. To maintain the flow of semantic information dense skip connections were used. An attention block was introduced to learn shape features. A self- attentive block having dilated convolutional layers was used to capture detailed information of tooth. A total of 1500 digital panoramic X-rays were used to evaluate the performance. Rohrer et al.[48] proposed a methodology for segmenting dental restorations. The U-Net deep model with ResNext50 as the backbone was trained with tiles of the images rather than the full image. The experiment was conducted on 1781 panoramic images. The authors had shown that the tiling helps in overcoming the issue of biasing for frequent and extended classes. Nader et al.[5] proposed a methodology for joint numbering and teeth segmentation. The bounding boxes were used to locate the teeth, and using this information, U-Net was deployed to accurately segment the teeth. The experiment was conducted on 543 dental panoramic images. Caylak et al.[49] used the transfer learning techniques to address the issue of small datasets available in dentistry for deep learning. The authors deployed two different pre-trained models on dental panoramic images. U-Net was pre-trained with chest X-ray images and Inception net was pre-trained with ImageNet. A total of 131 dental panoramic X-ray images were used to conduct the experiment.

2.1.2.3 Graph Convolutional Network (GCN)

Graph Convolutional Networks (GCNs) have emerged as a promising technique for dental image segmentation tasks. GCNs operate on graphs, which can be constructed from dental images by treating image voxels or pixels as nodes and their spatial relationships as edges. By leveraging the connectivity information between image pixels, GCNs can capture both local and global contexts and produce more accurate segmentation results.

Zhao et al.[50] proposed a deep model based on a graph convolutional network for segmenting a single tooth from 3D IOS (Intra-oral scans). Two parallel graph streams were presented to learn single-view geometric features, followed by a self-attention module to learn distinguished multi-view features from mesh cells by avoiding the confusion generated by different raw attributes. The authors evaluated the model on a real patient dataset having 12 intra-oral scans. Zhao et al.[51] proposed an attentional graph neural network for teeth segmentation in 3D IOS. The local features are extracted by a series of attention-based graph layers, while global features are obtained by the global branch. Local and global features are fused to perform the cell-wise segmentation tasks. The model was evaluated on 80 3D dental models.

Wu et al. [3]proposed a two-stage deep network for segmentation and localization of tooth on 3D IOS. In the first stage, an end-to-end graph-based neural net is proposed for tooth segmentation, while in the second stage, PointNet[52] was modified to learn the heatmaps for the localization of landmarks. The experiments were conducted on 136 samples of 3D IOS. Qiu et al.[53] proposed a deep methodology with weak annotations for automatic tooth segmentation in 3D dental models. Tooth centroids were detected to segment each tooth. The deep model was based on GCN, while segmentor was independently trained by using a patch-based strategy. The experiment was conducted on 4773 3D dental models.

2.1.2.4 Encoder-Decoder-based Architecture

The encoder-decoder architecture consists of two main parts: an encoder network that captures high-level features from the input image and a decoder network that produces a segmentation map. The encoder typically consists of several convolutional and pooling layers that down-sample the input image, while the decoder contains up-sampling and deconvolution layers that restore the original resolution of the image.

Zhao et al.[54] addressed the crucial issue of blurriness in tooth boundaries and tooth root segmentation. The two-stage attention network for segmentation of tooth area was proposed. In the first stage, an attention mechanism was adopted to localize tooth regions by embedding local and global attention modules. FCN (Fully Convolutional Network) was used for recognizing precise tooth borders in the second stage. The dataset contains 1500 dental panoramic X-ray images. Cui et al. [55] proposed a deep model based on an encoder-decoder architecture with an attention mechanism to deal with issues of crowding, missing teeth and irregular shape of teeth for segmentation. The model comprises two stages; first, a distance-aware voting scheme is used to localize the tooth. In the second stage, a segmentation module is designed to segment the individual tooth.

The model was evaluated on 2000 dental models, which were categorized into 1000 upper segments and 1000 lower segments. Chen et al.[56] developed a deep network based on encoder-decoder architecture to deal with the issue of fuzzy root boundaries. The network consists of two modules aggregation module for reducing the semantic gap and a location perceptron module to detect tooth pixels from a global level. The model was trained with a hybrid loss function to reduce the boundary loss. The method was tested on 1500 panoramic X-ray images.

2.1.2.5 Others

There are several other deep learning models which are based on different techniques, such as Generative adversarial network (GAN), Transformer, Hybrid models etc. GANs consist of two neural networks: a generator network that produces synthetic images and a discriminator network that distinguishes between real and synthetic images. Transformer consists of a self-attention mechanism that enables the model to attend to different parts of the input image and capture both local and global spatial information. The Hybrid deep models are either the combination of two deep learning techniques or a combination of conventional machine learning methods with deep learning methodology.

Cui et al. [57] proposed a deep segmentation model based on conditioned GAN for precise tooth segmentation from dental panoramic X-ray images. The wide residual blocks with encoder-decoder were introduced in the generator part of the deep architecture. The model was trained with different patches of images to reduce computational cost. A total of 1500 panoramic image dataset was used to conduct an experiment. Kumari et al.[58] proposed a hybrid methodology for the segmentation and detection of dental caries. The hybrid deep model is composed of ResneXt-RNN with meta-heuristics. The preprocessing of the image is done by bilateral filtering and CLAHE. The experiment was carried out on 120 periodic dental X-ray images.

Panetta et al.[59] presented a benchmark multimodal dataset containing 1000 panoramic X-ray images. The categorization of the images was done on five different levels based on peripheral properties, anatomy, abnormality, surrounding structures and density. The efficacy of the dataset was shown by the authors by deploying state-of-the-art deep models on this dataset. Karacan et al.[60] performed the segmentation task on dental panoramic images for segmenting teeth and maxillomandibular region. The authors used three different deep models based on attention mechanism. The two models are based on transformer network namely Vision Transformer and Segmenter while the other model is convNext. The model was trained and tested on Tufts dataset having 1000 dental panoramic radiographs. The authors have shown that these models perform better than U-Net architecture.

Although, these deep networks have shown remarkable performance for segmentation, there is still a need for the improvement of these architectures as some of these models are heavily dependent on pre-processing and post-processing techniques. Also, some deep networks are too complex and require a high number of training parameters which increases the computational cost. Thus, there is a need to develop a lightweight and more generalizable deep neural network for dental image segmentation. The other significant challenge is the limited size of dental image datasets, which often leads to overfitting and reduced generalization performance of deep learning models. Another challenge is the high variability in dental images due to variations in imaging modalities, patient demographics, and dental conditions, which can affect the accuracy of the segmentation. Additionally, the presence of artefacts and noise in dental images can further complicate the segmentation task.

Table 2.2 Summary of deep learning techniques for dental image segmentation

Author Year	**Objective**	**Deep Architecture**	**Image Modality**	**Dataset Size**	**Remark**
Zhao et al. 2020 [50]	Automatic Teeth Segmentation	GCN	3D IOS	12	Performs good for challenging cases but amount of data is too small
Cui et al. 2020 [55]	Accurate and automatic tooth segmentation	Encoder Decoder, Attention	3D dental models	2000	Deals with the problem of crowding, missing and misaligned teeth
Yang et al. 2020 [34]	Precise Tooth segmentation	Hybrid CNN	CBCT	512	Heavily relies on prior shape information
Zhao et al. 2020 [54]	Accurate tooth region segmentation with clear tooth boundaries	Encoder-Decoder, Attention	Panoramic X-ray images	1500	Deals with blurred boundaries but too complex structure
Nguyen et al. 2020 [40]	Alveolar Bone Segmentation	U-Net	Ultrasound	1100	Scope of improvement in terms of dice score
Zhao et al. 2021 [51]	Precise Teeth Segmentation	GCN, Attention	3D dental models	80	Captures complete teeth structure for segmentation
Zheng et al. 2021 [43]	Lesion Detection	Dense U-Net	CBCT	20	Very less amount of data and ground truth was manually developed which can be erroneous
Besson et al. 2021 [44]	Segmentation of Dental root canal and crown segmentation	U-Net	CBCT	80	Used the pre-built software
Zhu et al. 2021 [45]	Multistage Caries lesion segmentation	U-Net, Attention	Panoramic Images	1159	Performs segmentation at different levels, at moderate level performance was low
Chen et al. 2021 [56]	Accurate segmentation of tooth area	Encoder-Decoder, CNN	Panoramic X-rays	1500	Deals with fuzzy tooth root boundaries, too complex architecture with large number of parameters

Nishitani et al. 2021 [32]	Precise Teeth Segmentation	U-Net	Panoramic X-rays	162	Hybrid loss function, parameters are suboptimal
Cui et al. 2021 [57]	Pixelwise Tooth Segmentation	GAN	Panoramic X-rays	1500	Patches of images were used to train which can loose out on small information thus hampers accuracy
Li et al. 2021 [46]	Tooth Root Segmentation	U-Net, Transformer	Root Canal therapy X-ray	248	Fourier loss is shape sensitive thus can affect segmentation accuracy, small dataset size
Wu et al. 2022 [3]	Landmark Localization and Automatic Tooth segmentation	GCN	3D IOS	136	Overlapped teeth results in an incomplete mesh
Hou et al. 2022 [47]	Accurate Tooth Segmentation	U-Net, Attention Block	Panoramic X-ray	1500	Clear tooth root boundaries for precise segmentation
Panetta et al. 2022 [59]	Benchmark Multimodal Panoramic X-ray dataset	Different Models	Panoramic X-ray	1000	Developed a new dental multimodal dataset and executed state-of-the-art deep models on them
Rohrer et al. 2022 [48]	Segmentation of Dental Restorations	U-Net	Panoramic radiographs	1781	Could not be generalized for other deep learning tasks, scope in improving accuracy
Boudraa et al. 2022 [35]	3D dental image segmentation	CNN	3D Images	95	Suffers in segmenting for overlapping teeth
Qiu et al. 2022 [53]	Instance Tooth Segmentation	GCN	3D Dental models	4773	For unseen models may give inaccurate segmentation
Lin et al. 2022 [37]	Dental X-ray image segmentation	CNN	Panoramic X-ray	1321	Has very less parameters but there is a scope of improvement in dice and jaccard index score

Nader et al. 2022 [5]	Joint Teeth Segmentation and Numbering	U-Net	Panoramic X-ray	543	Segmentation depends upon the bounding boxes thus affecting dice score
Kumari et al. 2022 [58]	Caries Segmentation and detection	Hybrid	Periodic dental X-ray	120	Very small dataset
Chandrashekhar et al. 2022 [36]	Tooth segmentation and identification	CNN	Panoramic X-ray	1500	Could not deal with low resolution or small pixel images, Not suited for unusual dental conditions
Li et al. 2022 [38]	Dental Plaque Segmentation	CNN, Attention	Endoscopic Images	2882	Could not deal with images having unknown variations or are obtained from different endoscopes
Caylak et al. 2022 [49]	Automatic Dental Image Segmentation	U-Net	Panoramic X-ray	131	Performance depends upon the pre trained models
Karacan et al. 2022 [60]	Teeth and Maxillomandibular Segmentation	Transformer	Panoramic X-ray	1000	Shown the advantages of attention mechanism, Segmentation results are not precise.

2.1.3 Literature for Capsule Network for medical image segmentation

CNNs, are one of the most effective deep learning techniques which have been extensively used for accurate segmentation in medical imaging and produced satisfactory results. Researchers have deployed different variants of CNN for brain tumour segmentation, lesion segmentation, corneal segmentation, tooth segmentation etc. In spite of the noteworthy advancements, CNNs have primary drawbacks like preserving the relationship between object pieces and whole, losing crucial and relevant information due to pooling layers and requiring large amounts of data to train the network.

The shortcomings of the CNNs were overcome by a capsule neural network called CapsNet, which was proposed by Sabour et al.[61] The capsule network consists of convolutional layers, capsule layers, and fully connected layers. The capsules are a group of neurons which represent the features. The modification of weights in the network is done by a dynamic routing algorithm. The CapsNet has shown a remarkable performance on the MNIST dataset and good results for the CIFAR 10 dataset for classification tasks. Since then, researchers have started to extend the concept of capsule networks for detection and classification purposes in medical imaging. The majority of work using capsule networks has focused on image classification.

Due to complex architecture and dynamic routing, capsule networks are not widely used for segmentation tasks. Recently, LaLonde et al.[62] modified the dynamic routing algorithm to locally constrained dynamic routing and introduced the convolution and deconvolution capsules for object segmentation in the deep capsule model called SegCaps. The authors then extend this encoder-decoder architecture for binary medical image segmentation which can deal with high dimension images[63]. In locally constrained routing children are routed to parents in a local window and the transformation matrix is shared within a capsule type. The encoder encodes the feature vector using convolutional capsules by locally constrained routing while in the decoder deconvolution capsules are introduced for prediction vectors which are formed by transpose operation. The authors perform the segmentation of lungs from CT (Computed Tomography) scans and muscle and adipose tissue from MRIs (Magnetic resonance imaging) for evaluating the network. Bonheur et al.[64] improvised the capsule framework by proposing a new methodology for multilabel segmentation. The authors combined the two matrices that is pose information and appearance feature into a special capsule called MaTwo-CapsNet. Additionally, a novel dual-routing algorithm was introduced to combine information from these two matrices. MaTwo-CapsNet was evaluated on JSRT chest X-ray dataset[65].

Komm et al.[66] proposed a deep-learning method which combines capsule network with the inception architecture for retinal vessel segmentation. The network is based on shallow architecture having fewer parameters. The convolutional layer before the primary capsule is replaced by the inception layer. The model was evaluated on the DRIVE dataset containing retinal fundus images. Koresh et al.[67] proposed a modified capsule network to segment the three major boundaries of the corneal layer from the OCT (Optical coherence tomography) corneal images. The work done by the authors was divided into pre-processing, classification and segmentation tasks. The pre-processed images were given as input to classification capsules, and then these classified images were fed to a modified segmentation capsule network for the segmentation of boundaries. The model was evaluated on an openly available OCT image dataset.

Bonheur et al.[68] proposed the capsule network called OnlyCaps-Net for multilabel semantic segmentation. In the model, the authors adapted the concept of separable depthwise convolution to reduce training time and memory requirements. Two squash functions, soft squash and unit squash were also proposed along with unit routing. The model was evaluated on three medical datasets for segmenting images. Bragsten et al.[69] proposed an optimised capsule network for segmenting intravascular ultrasound images. The model was optimised by changing the number of capsules in the encoder and decoder path and hyper-tuning the parameters of the Matwo Capsule network. The model was evaluated on the IVUS[70] segmentation public dataset containing 435 annotated frames.

Nguyen et al.[71] proposed an efficient deep neural network based on 3D capsule network called 3D-UCaps to segment volumetric medical images. The model contains two paths, the encoder path which is built on the 3D capsules blocks preserving the spatial relationship while the decoder path has the CNN blocks which learn the visual contextual representations. The robustness of the model was demonstrated by the four volumetric medical datasets iSeg17[72] , Hippocampus [73], LUNA16 and Cardiac[73] datasets having MRI images and CT scans of brain, lungs and heart. Moghaddasi et al.[74] proposed a hybrid capsule network for automatically segmenting 3D mandibles. The authors optimised the 3D-Ucaps architecture to take advantage of capsule networks as well as CNN. The authors also proposed the hybrid loss function comprising of focal loss and margin loss to train the network. The model was evaluated on a public domain database for computational anatomy (PDDCA)[75] having 48 CT scans. Tran et al.[76] presented an efficient 3D encoder-decoder based on capsule architecture called 3D ConvCaps for medical image segmentation. The encoder path is designed with convolutional layers to learn low-level features and capsule layers to model high-level features. The authors further extend their work by adding self-supervised learning to the 3D capsule network and proposed SS-3DCapsnet[77]. The models were evaluated on three datasets of infant's brain iseg17[72] dataset , cardiac MRI[73], and Hippocampus MRI [73] datasets.

Huynh et al.[78] proposed a novel model named CapNext, that unifies a capsule architecture and ResNext architecture for medical image segmentation. In the proposed model, the path summation of the ResNext is replaced by the routing-by-agreement algorithm. The model used two datasets for evaluation. The first was 2D SIIM pneumothorax dataset consist of 10,675 chest images and 3D KiTS[79] dataset having 210 CT volumes of kidney.

Zade et al.[80] modified the SegCaps by utilizing the dilation blocks in the primary capsule layer for glioma segmentation. The authors also introduced the curriculum learning approach based on a roulette-wheel selection algorithm to train the model along with the hybrid loss function. The model was evaluated on the BraTS2020[81] dataset. Pawan et al.[82] used the SegCaps to segment sub-retinal fluid from optical coherence tomography images. Further, the authors enhanced the SegCaps by utilizing the concept of dilation, residual connections, inception blocks and capsule pooling. The model reduces the computation complexity of SegCaps by reducing the number of trainable parameters.

Wan et al.[83] proposed an attention guided capsule network for medical image segmentation. The Capsule based high resolution network was formed which was assisted by cross-branch multiscale feature augmentation. Capsule based full attention mechanism improves the feature quality at each scale. The authors evaluated the model on four different medical image datasets[84],[85],[86].

Capsule networks have shown its potential for medical image segmentation but it has not been used for teeth segmentation from panoramic X-rays. Thus, capsule network can be utilized for dental image segmentation.

Table 2.3 Summary of medical image segmentation using Capsule Network

Author / Model Name	**Methodology**	**Task**	**Image Type**
Lalonde et al.[63] SegCaps	Convolutional Capsules in encoder Deconvolutinal Capsules in decoder Locally constrained routing algorithm	Lung Segmentation Muscle and Adipose tissue segmentation	CT Scans , MRI
Bonheur et al. [64] MaTwo-CapsNet	Based on SegCaps Combine Pose and Appearance matrices Dual routing algorithm	Multilabel segmentation of chest X-rays	X-ray
Kromm et al.[66]	Capsule Network with Inception architecture	Segmentation of retinal blood vessels	Fundus Images
Koresh et al.[67]	Modified SegCaps	Segments three major boundaries of corneal layer	OCT images
Bonheur et al. [68] Only-CapsNet	Capsule Network	Semantic segmentation of chest X-ray	X-ray

	Depthwise separable convolutional capsules Squashing function softsquash and unitsquash Unit routing		
Bragsten et al. [69]	Modified MaTwoCapsNet changes number of layer	Lumen and Vessel wall segmentation from Intravascular ultrasound	Ultrasound
Nguyen et al.[71] 3D UCaps	Encoder-Decoder Architecture Encoder – Capsule Layer Decoder- Convolutional layers	Medical Image Segmentation Infant brain, hippocampus segmentation, lung segmentation, heart segmentation	MRI CT Scans
Tran et al. [76] 3D ConvCaps	3D U-Net with convolutional capsules in encoder	Medical Image Segmentation infant brain, hippocampus and heart segmentation	MRI CT scans
Tran et al. [77] SS-3DCapsNet	Self-Supervised learning for 3DConcCaps	Medical Image Segmentation infant brain, hippocampus and heart segmentation	MRI CT scans
Moghaddasi et al.[74]	Hybrid Capsule Network Optimised 3DUCaps	3D Mandible Segmentation	CT Scans
Huynh et al.[78] CapNext	Unifies Capsule Network with ResNext	Pneumothorax Segmentation Kidney tumor segmentation	Chest X-ray Kidney CT Scans
Zade et al.[80]	Modified SegCaps using dilation blocks Curriculum based training algorithm	Glioma segmentation	MRI
Pawan et al.[82]	Improved SegCaps using dilated convolution, Residual connection, Inception block, Capsule pooling	Automatic segmentation of sub-retinal serous fluid	OCT
Wan et al.[83]	Full Attention-guided capsule network	Medical Image Segmentation Lung segmentation, Lesion skin segmentation, Lesion segmentation from fundus images, Polyp segmentation	CT Scans

2.1.4 Advances made by industry for Dentistry

Computer vision has made a significant impact in the field of dentistry by improving diagnostic accuracy, treatment planning and overall patient care. The introduction of artificial intelligence has attracted many industries to develop the advance software based on artificial intelligence for dental care. Pearl.com[87] showcases various innovations and introduce the product like Second Opinion the first AI-powered real time radiologic platform for accurate interpretation of X-rays. Their another product Practice Intelligence is useful for the staff for consistent and optimal performance. Also product like Smart Margin for dental restoration and Prep Assess to assess the margin quality of dental restoration are developed by Pearl.com. Another AI-based product FDA-cleared[88] is developed by Overjet. This platform detects and outlines the decay and quantifies bone loss on radiographs in real time for precise and accurate diagnostic decisions. Manchester-imaging[89] developed product called AssistDent for helping dental practices and OsteoDent which provides AI analysis of dental radiographs. Diagnocat[90] introduced the dental AI software for enhancing transparency and speedy treatment. Another dental AI software is introduced by Adravision[91] which helps in detecting pathologies, measuring bone-levels and colorize dental features. All these technologies utilize the advanced algorithms to help dentists make more precise diagnoses and develop better treatment plans. These advancements are transforming the dental industry, making patient care more effective and accessible.

2.2 Research Gaps

Based on the survey performed on the Dental Image Segmentation approaches the following research gaps have been identified in this field:

- There is the high variability in dental images due to variations in imaging modalities, patient demographics and dental conditions which can affect the accuracy of the segmentation.

- Most of the deep learning models are based on conventional CNN thus leaving scope for different type of CNN which can be utilized for enhancing the spatial features for dental image segmentation.
- Most deep learning approaches focus on teeth regions while ignoring the teeth boundaries in the low contrast images.
- The state of the art-methods is designed with complex network architecture and possess large number of trainable parameters. Thus, there is need to build lightweight deep models.
- The pooling techniques commonly used in CNNs results in loss of precise location and spatial information. Thus, leaving scope for other networks to be explore.

This thesis proposes various methods to address the abovementioned research gaps identified from the existing literature.

2.3 Dataset used for Experimental Analysis

In this thesis for experimental analysis of different models for dental image segmentation two benchmark dental datasets are used Figure 2.1 shows the sample of datasets:

- UFBA_UESC[2] dental images dataset (Dataset 1)
- Tufts dental database[59] (Dataset 2)

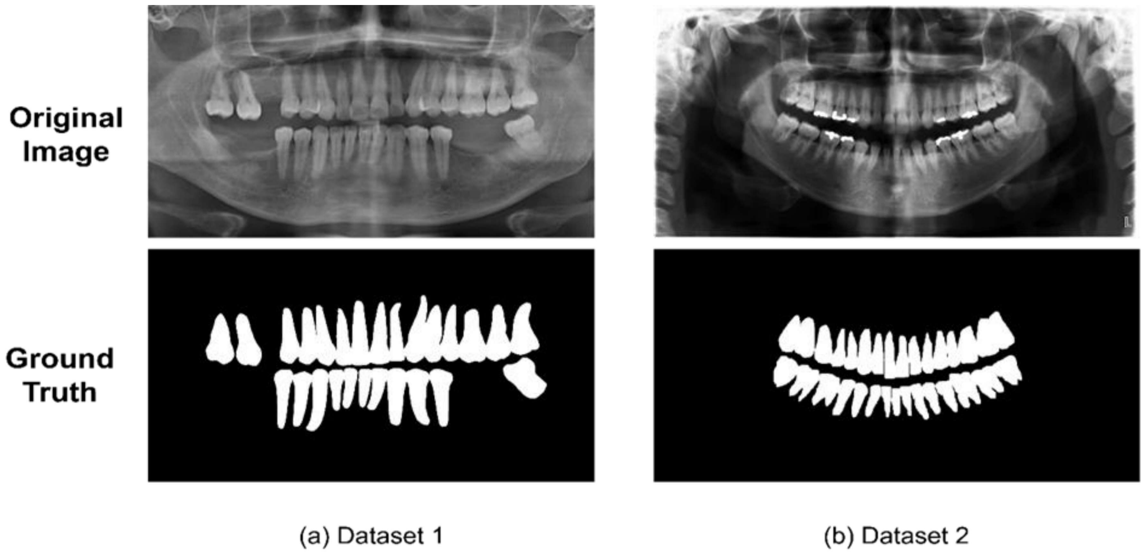

Figure 2.1 Sample from Datasets of dental panoramic X-ray images

2.3.1 UFBA_USEC Dataset

This dataset was developed at Ivision Lab by Silva and Gill[2]. The dataset consists 1500 dental panoramic X-ray images of size [1991 x 1127] pixels, which are categorized in to 10 different categories. The images of the dataset were captured from X-ray camera model Orthophos XG 5/ XG 5 DS/ Ceph, manufactured by Sirona Dental Systems. The ground truth of the respective images is also provided with this dataset. For experiments the size were reduced to [512 x 512] for better memory utilization.

Table 2.4: Category wise detail information of the dataset.

Categories	Cat-1	Cat-2	Cat-3	Cat-4	Cat-5	Cat-6	Cat-7	Cat-8	Cat-9	Cat-10
All Teeth	✓	✓	✓	✓		✓				
Missing Teeth					✓		✓	✓	✓	✓
Dental Appliance	✓		✓				✓		✓	
Restoration	✓	✓					✓	✓		
Dental Implant					✓					
Average Teeth	2	2	2	2	8	7	7	9	8	8
Images	3	20	5	40	20	70	15	57	5	15

2.3.2 Tufts Dental Database

This dataset was developed by Panetta et al[59] at Tufts University. The dataset consists 1000 dental panoramic radiographs of dimensions [1615 x 840] pixels. The radiographs were randomly selected form the electronic patient database. The qualifying criteria for image to be got selected was optimum diagnostic quality of image without any or minimal technical error. The images were acquired by using OP100 Orthopantomograph and Plammeca Promax 2D radiographic units. The ground truth of the respective images was also provided within the dataset. For experiments the size were reduced to [512 x 512] for better memory utilization.

2.4 Performance Metrics

The dice coefficient is commonly used in the field of medical image segmentation field to measure the efficacy of an image segmentation model, is defined as follows:

$$dice\_coef = \frac{2|P_P \cap G_P|}{|P_P| + |G_P|} \qquad (2.1)$$

where P_P is the predicted segmented pixels of teeth region and $G_P$is segmented pixels the ground truth.

Another common metrics used for medical segmentation is Intersection Over Union (IoU) and is defined as follows:

$$Iou = \frac{|P_P \cap G_P|}{|P_P \cup G_P|} \qquad (2.2)$$

where P_P is the predicted segmented pixels of teeth region and $G_P$is segmented pixels the ground truth.

Apart from this the image segmentation is also considered as pixelwise classification task. Thus, three common classification metrics are also utilized to measure the efficacy of the proposed segmentation model, these are described as follows:

$$accuracy = \frac{(T_{TP} + T_{TN})}{T_{TP} + F_{FP} + T_{TN} + F_{FN}} \quad (2.3)$$

$$precision = \frac{T_{TP}}{T_{TP} + F_{TP}} \quad (2.4)$$

$$recall = \frac{T_{TP}}{T_{TP} + F_{FN}} \quad (2.5)$$

where T_{TP} is the number of correctly forecasted pixels of teeth region, F_{FP} is the number of wrongly forecasted teeth region, T_{TN} is the number of correctly forecasted pixels of background region, F_{FN} is the number of wrongly forecasted pixels of background region.

2.5 Conclusion

In this chapter brief literature survey of various methods and models for dental image segmentation was conducted. Both conventional and deep learning approaches were reviewed. Additionally, a review of capsule network for medical image segmentation is also done. Several shortcomings of these approaches were identified and summarized as research gaps for study. After review, various datasets and performance metrics used for evaluation of dental image segmentation is also presented. The forthcoming chapters discuss the main contribution of this thesis.