



Lightweight deep learning methods for panoramic dental X-ray image segmentation

Songyue Lin¹ · Xuejiang Hao² · Yan Liu² · Dong Yan² · Jianwei Liu² · Mingjun Zhong¹

Received: 30 June 2022 / Accepted: 22 November 2022 / Published online: 16 December 2022
© The Author(s) 2022

Abstract

Dental X-ray image segmentation is helpful for assisting clinicians to examine tooth conditions and identify dental diseases. Fast and lightweight segmentation algorithms without using cloud computing may be required to be implemented in X-ray imaging systems. This paper aims to investigate lightweight deep learning methods for dental X-ray image segmentation for the purpose of deployment on edge devices, such as dental X-ray imaging systems. A novel lightweight neural network scheme using knowledge distillation is proposed in this paper. The proposed lightweight method and a number of existing lightweight deep learning methods were trained on a panoramic dental X-ray image data set. These lightweight methods were evaluated and compared by using several accuracy metrics. The proposed lightweight method only requires 0.33 million parameters (~ 7.5 megabytes) for the trained model, while it achieved the best performance in terms of IoU (0.804) and Dice (0.89) comparing to other lightweight methods. This work shows that the proposed method for dental X-ray image segmentation requires small memory storage, while it achieved comparative performance. The method could be deployed on edge devices and could potentially assist clinicians to alleviate their daily workflow and improve the quality of their analysis.

Keywords Image segmentation · Tooth segmentation · Lightweight neural network · Artificial intelligence · Deep learning · X-ray

1 Introduction

Among many medical imaging protocols, panoramic X-ray imaging [1] is a great tool for diagnosing teeth diseases as it requires relatively low dose and low cost. By visualizing panoramic dental X-ray images, dentists can examine whole oral conditions and so could justify whether there are any dental diseases, such as caries/cavities, gum diseases, cracked or broken teeth and oral cancer. Mainly, dentists have to analyze the X-ray images based on their

experience and visual perception [2]. For obtaining dental panoramic images, the X-ray tube needs to rotate around the subject. Consequently, this would make it a challenge for dentists to effectively examine the images due to various reasons, e.g., different levels of noise generated by the machine, low contrast of edges, overlapping of anatomic structures, etc. These issues related to poor quality of the obtained X-ray images would make it hard for dentists to identify diseases. Therefore, automatically analyzing dental X-ray images would be very helpful for assisting practicing dentists to alleviate their daily workflow and improve the quality of their analysis.

For assisting practicing clinicians to analyze dental X-ray images, automatic methods are to be developed to solve those challenges which include identifying and classifying tooth diseases, identifying anatomical landmarks and segmenting tooth structures [2, 3]. Perhaps, among those challenges, segmenting tooth structures is the most elementary task for automatically analyzing dental X-ray images. For example, before identifying tooth

✉ Xuejiang Hao
rmyy@wf.shandong.cn

✉ Mingjun Zhong
mingjun.zhong@abdn.ac.uk

¹ Department of Computing Science, King's College, University of Aberdeen, Aberdeen AB24 3FX, Scotland, UK

² Department of Dentistry, Chang-Le County People's Hospital, Chang-Le County 262400, Shan-Dong Province, China

diseases, it would have to isolate the teeth by using segmentation techniques. Image segmentation is a common technique for analyzing images. The goal of image segmentation is to partition a digital image into various regions which makes it easier to represent an image according to the distinct objects in the image. For example, in dental X-ray image analysis, the goal of segmentation is to isolate each tooth from other objects in the image such as jaws, gums and other details of face. Further analysis, e.g., identifying diseases, could be carried out given an isolated tooth. In this paper, we focus on segmenting the panoramic dental X-ray images.

In the field of image segmentation in general, there are various kinds of approaches to segmenting images. For example, bounding box method is an approach to segment an object from an image. Typically, each object is represented by an axis-aligned bounding box that tightly encompasses the object. This approach could be represented as a classification problem and the task is to classify the image content in the bounding box to a specific object or background [4–6]. Objects as points [7] is an approach to simplify the bounding box method. The idea of objects as points is to represent objects by a single point at the objects' bounding box center, and other features such as object size, dimension, 3D extent, orientation and pose are represented as a regression function of the image features at the center point. Different to bounding box methods, image segmentation could also be presented as a classification problem. The pixels in the image are classified as a predefined object, and so, the task is to assign a class to each pixel of the image. Many deep learning methods treat image segmentation as a classification problem. Experimental results show deep learning methods achieve good performance. Typically, U-Net [8] is a popular deep learning method for image segmentation.

For medical image segmentation, the U-Net would be one of the most popular methods and has been largely modified for the purpose of image segmentation. For example, inspired by the DenseNet [9] and based on the U-Net, the UNet++ proposed an encoder–decoder network where the encoder and decoder sub-networks are connected through a series of nested, dense skip pathways [10]. Such skip pathways could be able to reduce the semantic gap between the feature maps of the encoder and decoder sub-networks. R2U-Net [11] integrates recurrent neural network (RNN) and residual block with U-net architecture. 3D U2-Net [12] can deal with multiple images of different kinds of images using convolution 3D kernel, which means it is not necessary to retrain a new model for different images. RAUNet [13] proposes a new architecture which includes attention and residual block based on UNet. VNet [14] is a special variant designed for 3D data set.

However, all these models require huge computational costs, huge numbers of parameters and large memory. Therefore, cloud computing is required for deploying these algorithms, which would not be applicable in many scenarios. For example, these algorithms must be used with Internet connections; it would not be possible to deploy these algorithms on X-ray devices because of memory requirements. In this paper, for tackling these problems, we investigated a few existing lightweight algorithms for dental image segmentation. In addition, we have proposed a new lightweight algorithm employing knowledge consistency for our segmentation task. In the next section, we reviewed related works for dental image segmentation and knowledge distillation methods which are relevant to our method.

2 Related work

2.1 Dental image segmentation

Deep learning techniques have been experimentally shown as better performance comparing to traditional methods for image segmentation tasks and therefore dental image segmentation. Various traditional methods have been applied to semantic segmentation for dental X-ray images. These algorithms include region growing, splitting and merging, global thresholding, fuzzy method, level set method, etc. For example, the work of [3] provided a comparative study for applying these algorithms to dental X-ray images. Some experiments have shown that deep learning methods perform much better than traditional methods, which was seen as the early study demonstrating the advantages of using deep learning approaches for panoramic X-ray image segmentation [3, 15–20]. The work of [21] systematically reviewed recent segmentation methods for stomatological images based on deep learning. For example, mask R-CNN (MRCNN) was used to make instance segmentation for dental X-ray images [22]; U-Net [8] was another popular method for image segmentation and was applied to semantic segmentation for dental X-ray images as well [16, 23]. These research have shown that data augmentation techniques including horizontal flipping and ensemble technique improved the performance of U-Net [23]. However, the performance of MRCNN and U-Net would be affected by issues in the images such as the low contrast in the tooth boundary and tooth root. This issue could be mitigated by using attention techniques [16].

2.2 Lightweight methods

Generally, the mentioned deep learning methods perform well for dental image segmentation. However, these methods require large memory for storing the trained

model parameters and a lot of floating point operations (FLOPs) which results in long running time. Potentially, it is required to deploy these models to edge devices and thus light weight models are needed for such purpose.

In semantic image segmentation, some lightweight methods have been proposed for the purpose of edge device application. For example, the efficient neural network (ENet) [24] was proposed for real-time semantic segmentation. ESPNetv2 [25] is a typical lightweight convolutional neural network which has shown superior performance on semantic segmentation and requires fewer FLOPs.

Other schemes such as compression methods [26] and knowledge distillation methods [27] could be applied to dental image segmentation to obtain a lightweight model. Typically, knowledge distillation is a scheme to transfer knowledge from cumbersome teacher models to teach a lightweight student model to mimic a teacher model. Originally, the main strategy of knowledge distillation is to use a soft probability output to make a student model to perform similarly to the teacher model. An advanced approach for knowledge distillation is to use the intermediate feature maps to train a student network. For example, adapting intermediate feature maps helps to improve performance of image segmentation [28]. Recently, some novel knowledge distillation methods have been proposed for semantic segmentation tasks. As an example, an efficient and lightweight method using knowledge distillation has been proposed for medical image segmentation such as CT images [29]. Other knowledge distillation methods such as structured knowledge distillation [30] and transformer-based knowledge distillation [31] could also be applied to semantic segmentation. Interestingly, transformer-based knowledge distillation would be an interesting approach for dental X-ray segmentation which learns compact student transformers by distilling both feature maps and patch embedding of large teacher transforms [31].

Rare work has been done for dental image segmentation using knowledge distillation, although it has been used in chest X-ray images [32] and 3D optical microscope image [33]. In this paper, we investigate knowledge distillation methods for semantic dental X-ray image segmentation, which will produce lightweight models for the purpose of deployment on edge devices. In the following section, we will describe how knowledge distillation could be used for dental image segmentation, and propose a knowledge distillation approach for dental image segmentation. Precisely, different to other distillation methods, we propose to use knowledge consistency neural networks for distilling knowledge from teacher to student models. Such knowledge consistency neural network could potentially extract consistent features from the feature maps learnt by using teachers. In the experiments, we will demonstrate that our

method outperforms a number of other knowledge distillation methods for dental X-ray image segmentation.

3 Knowledge consistency neural networks for knowledge distillation

In this section, we propose a knowledge network which aims to extract consistent knowledge learnt by a teacher network. For simplicity, we called it knowledge consistency neural network (KCNet). Simple attention scheme such as the sum of absolute values of the learnt feature maps has been proved useful for extracting spatial activation features in computer vision and therefore image segmentation tasks [34]. These extracted activation features could be used to transfer knowledge from a huge teacher network into a lightweight student network. Such attention strategy has shown a good performance in our semantic segmentation tasks in the experiment results section.

Various attention schemes could be used for transferring knowledge from a large teacher network to a student. For example, these attention schemes could be sum of absolute values of feature maps, sum of the absolute values raised to the power of p . Using this approach, the student network could be designed as a shallower lightweight network. This enables the student network to have similar spatial attention feature maps to those of the teacher network. The idea of this attention scheme is simply to match the sum of the absolute values of the feature maps between teacher and student networks. An advantage of this approach is that it does not introduce extra model parameters, comparing to the Fitnets [35] which requires to learn extra regression parameters.

In this paper, we investigate a nonlinear relationship between the spatial attention feature maps of teacher and student networks. This nonlinear relationship would be able to extract the consistent features from the feature maps. Note that the attention network in [34] could be viewed as a linear relationship. In the literature, it has shown that even different learning tasks for image recognition could have consistent knowledge. For example, when our task was image classification, we can train an ensemble of neural network models and each of them could be performing very well and they could have the similar performance as well. It is believed that these feature maps learnt from different models could have shared consistent knowledge [36]. This means that the feature maps could be mapped to an intrinsic (or consistent) feature map which are shared across all the models by using a nonlinear function. (In the following, we may exchangeably use intrinsic and consistent knowledge.) One possible approach

to learn these intrinsic knowledge is to use the knowledge consistency proposed in [36]. The knowledge consistency was to learn a nonlinear function to map the features from one pretrained model to another. Instead, our target is to train a student neural network with its intermediate feature maps being consistent to the teacher networks. One difference between our approach to attention and Fitnets are that our method does not learn a similarity between teacher and student networks, but instead to learn knowledge that is consistent between the networks which are believed as the intrinsic knowledge learnt by both models.

We now define our model for learning intrinsic/consistent knowledge. The model architecture is shown in Fig. 1. Denote an intermediate feature map from a teacher network and a student network by x_T and x_S , respectively. A function is applied to the teacher feature map to match the student feature map by minimizing the following loss function:

$$\text{loss}_{\text{kc}}(W_{\text{kc}}, W_S) = \|x_S(W_S) - f_{W_{\text{kc}}}(X_T)\|,$$

where W_S and W_T are the parameters in student and teacher networks, and $f_{W_{\text{kc}}}$ is a neural network with parameters W_{kc} similar to the knowledge consistency function defined in

[36]. Typically, we use the architecture described in Fig. 2 for $f_{W_{\text{kc}}}$. The hope is that the knowledge consistency function [36] is able to extract the consistent knowledge from teacher feature maps.

For the purpose of prediction, we used a knowledge distillation loss function (equivalent to Kullback–Leibler divergence), to match the predicted masks between teacher and student networks, assuming there are N samples:

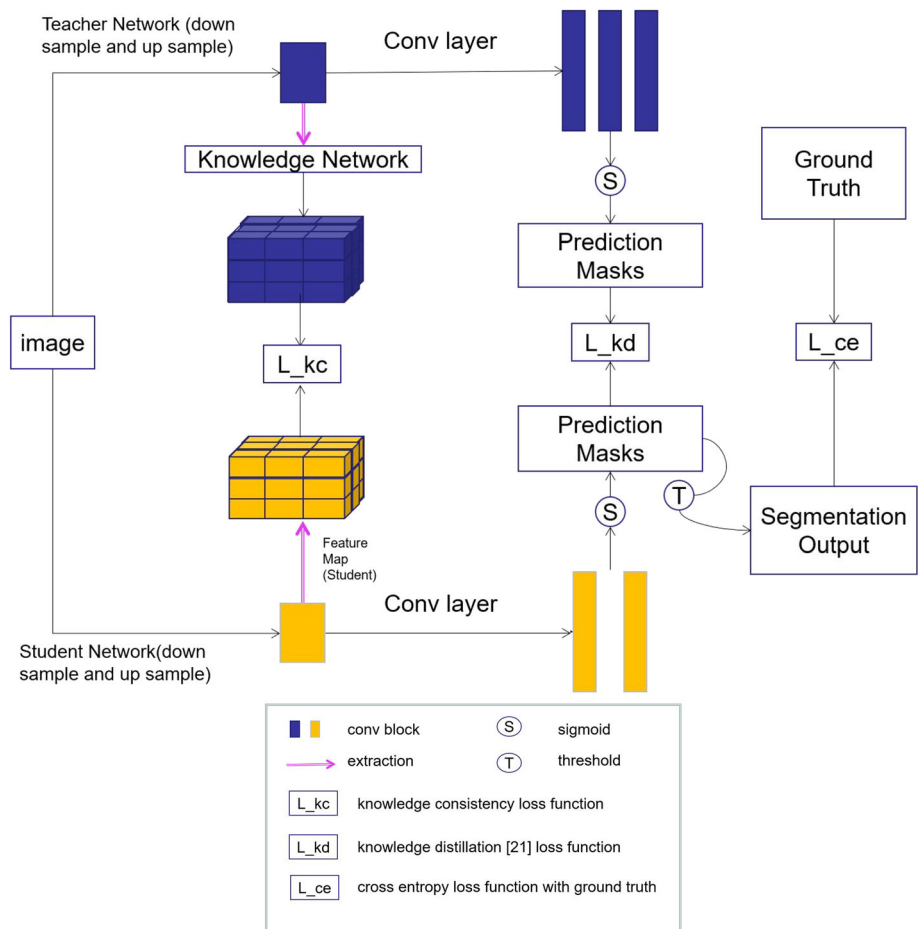
$$\text{loss}_{\text{kd}}(W_S) = \sum_{i=1}^N P(W_T)(\log(P(W_T)) - \log(P(W_S))),$$

where P denotes a softmax output, i.e., $P(W) = \frac{\exp(z_i(W)/\tau)}{\sum_j \exp(z_j(W)/\tau)}$ with a temperature parameter τ , and z_i denotes the output of a network. To compute the distance of the label outputs of teacher and student networks, the cross-entropy loss is used as follows,

$$\text{loss}_{\text{ce}}(y_i^*, y_i(W_S)) = - \sum_{i=1}^N y_i^* \log(y_i),$$

where y^* represents the truth label and y represents the softmax probability output.

Fig. 1 Architecture of our KCNet



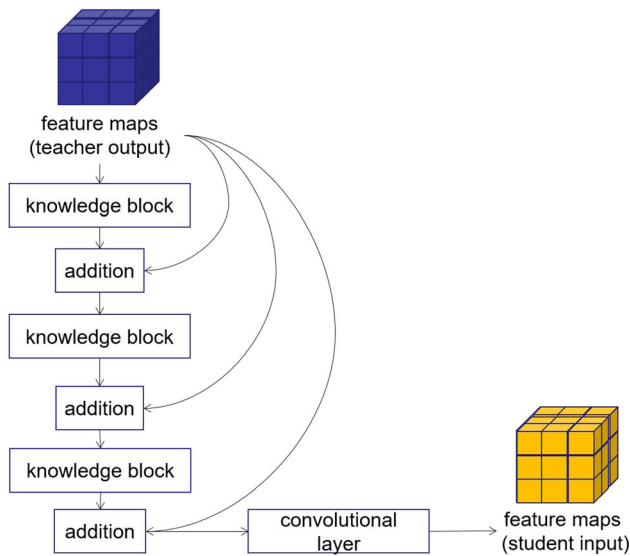


Fig. 2 Knowledge consistent network architecture. The knowledge block is defined by the architecture: *convolutionallayers* → *Normalization* → *ReLU*

Our total loss taking into account all these loss functions has the following form:

$$L = \alpha \sum_{j=1}^J \text{loss}_{\text{kc}}^{(j)}(W_{\text{kc}}, W_S) + \beta \text{loss}_{\text{kd}}(W_S) + \gamma \text{loss}_{\text{sce}}(y_i^*, y_i(W_S)),$$

where J indicates the total number of intermediate layers in the student network to be penalized using knowledge consistency. We found that the parameters $\alpha = 10000$, $\beta = 0.05$ and $\gamma = 1 - \beta$ in the loss function were working well in our model.

4 Experimental results

4.1 The data

The data set in this experiment for evaluating our algorithms contains 1500 panoramic dental X-ray images [3, 37]. There are 10 categories in this data set, but following the suggestions from [37], categories 5 and 6 were not used in our experiments, as they include images with implants and deciduous teeth. Finally, 1321 dental X-ray panoramic images in total were used in our experiments. There are 432 images for training, 111 images for validation and 778 images for testing. For the purposes of our experiments, we only preprocessed the images by normalization and resizing. Typically, the images were resized from 1991×1127 to 256×256 .

4.2 Evaluation metrics

In our experiments, for comparison purposes, five error metrics were used to evaluate these algorithms, which are intersection over union (IoU), Hausdorff distance (HD), Dice coefficient (DC), volumetric overlap error (VOE) and relative volume difference (RVD). They are described as follows.

IoU is a metric to evaluate the accuracy of detecting objects in an image. It is used to measure the correlation between the predicted box and the box containing truth objects.

$$\text{IOU} = \frac{A_{\text{PB}} \cap A_{\text{TB}}}{A_{\text{PB}} \cup A_{\text{TB}}},$$

where A_{PB} represents the area of predicted box and A_{TB} represents the area of truth box containing the targeted objects.

Hausdorff distance can be used to compute the distance between two sets in a metric space. This distance could be interpreted as the maximum value of the shortest distance from a point of set to another point of set

$$d_H(X, Y) = \max \left\{ \sup_{x \in X} d(x, Y), \sup_{y \in Y} d(X, y) \right\},$$

where \sup denotes supremum and \inf denotes infimum, and $d(x, Y) = \inf_{y \in Y} d(x, y)$.

The Dice coefficient is a collective similarity metric defined by

$$\text{Dice} = \frac{2(O_{\text{pred}} \cap O_{\text{truth}})}{O_{\text{pred}} + O_{\text{truth}}},$$

where O_{pred} denotes the segmentation output image of model and O_{truth} denotes the ground truth segmentation image. The range of DC is $[0, 1]$, where ‘1’ indicates that the prediction is identical to ground truth. In DC, the numerator is the intersection of the prediction and the truth, and the denominator is the union. Typically, in the numerator, the intersection between prediction and ground truth is computed twice (Figure 4.2). Some experiments showed that when the intersection between prediction and ground truth was not factored by 2, the Dice value would be fluctuated and hard to be a stable metric.

Similar to DC, the VOE is defined as follows. It denotes the error rate, and $[0, 1]$ is its range, where ‘0’ indicates that there is no mistake.

$$\text{VOE} = \frac{2 * (O_{\text{pred}} - O_{\text{truth}})}{O_{\text{pred}} + O_{\text{truth}}}.$$

RVD is a metric to show the difference of volume between ground truth and prediction. Its value belongs to $[0, 1]$, and

'0' means the model can produce the same segmentation prediction with ground truth. The RVD is defined as

$$\text{RVD} = \frac{\left| |O_{\text{pred}}| - |O_{\text{truth}}| \right|}{|O_{\text{truth}}|}$$

4.3 Settings for experiments

The following are the hyper-parameters for training all the models. We set 200 epochs for training the models. For preventing over-fitting, early stopping was used with 20 epochs as tolerance. We observed that most models converged in around 50 epochs. The learning rate was set to $1e-3$, and the batch size was set to 4. For the purpose of using distillation algorithms [27], the temperature τ in the following class probability is used as a parameter to produce a soft probability distribution over classes:

$$q_i = \frac{\exp(z_i/\tau)}{\sum_j \exp(z_j/\tau)}$$

We set $\tau = 4$ in our experiments. Note that higher value of τ indicates a softer probability distribution over classes.

4.4 Results

In the experiments, deep learning models were applied to the same data set for comparing their performance. These models include UNet [8], UNet++ [10], SegNet [38] and RAUNet [39]. Besides, lightweight models including ENet [24] and ESPNet v2 [25] were applied to the data as well. In our model, we used UNet++ as our teacher network and ESPNet-v2 as our student network. Their architectures are shown in Fig. 3 and Table 6, respectively. We experimentally compared our method to these models. Typically, in our model employing knowledge consistency, the number of layers, i.e., the number of blocks in knowledge consistent network, is a hyper-parameter to tune. We set the number of blocks from one to three, and we chose the

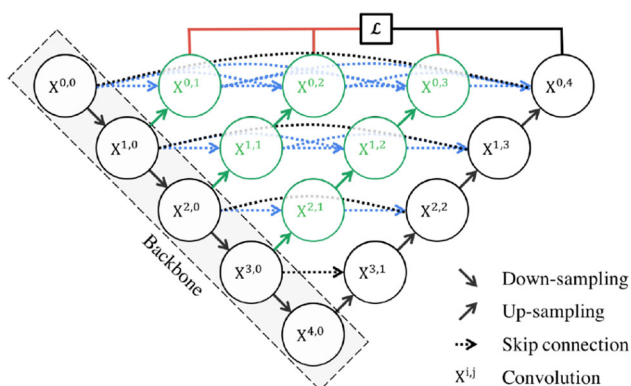


Fig. 3 Architecture of UNet++ [10]

number of blocks with best performance. The results are given in Table 1 which shows that the model achieved the best performance when it has two blocks. Table 2 compares our algorithm to nine other methods. It shows the sizes of these trained models and the FLOPs. It shows that our model has the same size in term of models size and FLOPs to ESPNet-v2, showing that our method requires least number of parameters. In terms of FLOPs, our model is comparable to ENet which has the smallest FLOPs. All these methods are compared in terms of IoU, Dice, HD, VOE and RVD. Across IoU, Dice and HD, the UNet++ method performed the best among all the methods. Our method achieved the best performance in terms of VOE and RVD. It also shows that our model outperforms ESPNet-v2 across all the evaluation metrics, which indicates that the knowledge consistency network may have performed the role to feed the knowledge learned from teacher network into the student network. We also compared our method to other knowledge distillation models which are the knowledge distillation (KD) method proposed in [27], the Fitnet [35] and the attention method [34]. The results are shown in Table 3. In terms of IoU and Dice, both the attention method and ours performed the best. The KD method was the best method in terms of HD, and Fitnet performed best in VOE and RVD. Figure 5 shows some example results for segmentation when lightweight methods were applied. Figure 6 plots some examples of possible improvements of our model compared to the student model. The Dice and FLOPs are plotted in Fig. 4 showing the differences between these lightweight methods and other methods. We can see that ENet, ESPNet-v2 and our method need significantly less FLOPs than other methods, while they achieved comparative performance in terms of Dice (Figs. 5, 6).

The impact of those harmonic factors α , β and γ in loss function over the performance was investigated. The Dice results obtained by using various values of those harmonic factors are shown in Table 4. It indicates that the model would perform poorly if they were not carefully chosen, e.g., $\alpha = 1e4$, $\beta = 5$ and $\gamma = -4$. However, it would not be possible to scan all the possible values for harmonic factors. Optionally, it would be interesting to optimize these hyper-parameters instead. The performance of using

Table 1 Performance of our algorithm KCNet in terms of the metric Dice with various number of knowledge consistency layers

Knowledge consistency layers	KCNet (ours)
1	0.875 ± 0.041
2	0.890 ± 0.038
3	0.889 ± 0.035

Table 2 Comparisons in various evaluation metrics across all the applied algorithms

Methods	Params (M)	FLOPs (G)	IoU	Dice	HD	VOE	RVD
PSPNet [40]	46.7	46.1	0.744 ± 0.050	0.852 ± 0.039	6.838 ± 0.798	0.323 ± 0.160	0.406 ± 0.226
RAUNet [39]	36.8	16.2	0.809 ± 0.069	0.892 ± 0.056	5.561 ± 0.736	0.221 ± 0.185	0.269 ± 0.212
SegNet [38]	29.4	40.1	0.811 ± 0.067	0.894 ± 0.051	5.566 ± 0.787	0.237 ± 0.172	0.288 ± 0.209
BiSeNet [41]	23.1	10.18	0.726 ± 0.062	0.840 ± 0.029	7.252 ± 0.732	0.425 ± 0.169	0.552 ± 0.194
UNet++ [10]	9.2	34.9	0.834 ± 0.056	0.908 ± 0.036	5.100 ± 0.831	0.243 ± 0.146	0.293 ± 0.201
UNet [8]	7.8	13.7	0.829 ± 0.060	0.905 ± 0.040	5.179 ± 0.830	0.221 ± 0.150	0.266 ± 0.199
CgNet [42]	0.49	0.9	0.735 ± 0.055	0.847 ± 0.025	7.087 ± 0.754	0.420 ± 0.181	0.543 ± 0.190
ENet [24]	0.34	0.5	0.810 ± 0.056	0.894 ± 0.037	5.489 ± 0.758	0.233 ± 0.148	0.281 ± 0.198
ESPNet v2 [25]	0.33	0.8	0.799 ± 0.063	0.887 ± 0.045	5.702 ± 0.738	0.225 ± 0.160	0.272 ± 0.205
KCNet (ours)	0.33	0.8	0.804 ± 0.058	0.890 ± 0.038	5.617 ± 0.756	0.220 ± 0.153	0.264 ± 0.206

Best results are highlighted in bold: larger values in Dice and IoU, and smaller values in other metrics are better

Table 3 Comparisons for lightweight neural network methods

Methods	IoU	Dice	HD	VOE	RVD
KD [27]	0.802 ± 0.061	0.889 ± 0.042	5.578 ± 0.744	0.205 ± 0.159	0.246 ± 0.204
Fitnet [35]	0.801 ± 0.065	0.888 ± 0.050	5.589 ± 0.736	0.199 ± 0.173	0.239 ± 0.207
Attention [34]	0.804 ± 0.059	0.890 ± 0.040	5.659 ± 0.760	0.229 ± 0.152	0.277 ± 0.208
KCNet (ours)	0.804 ± 0.058	0.890 ± 0.038	5.617 ± 0.756	0.220 ± 0.153	0.264 ± 0.206

Best results are highlighted in bold: larger values in Dice and IoU, and smaller values in other metrics are better

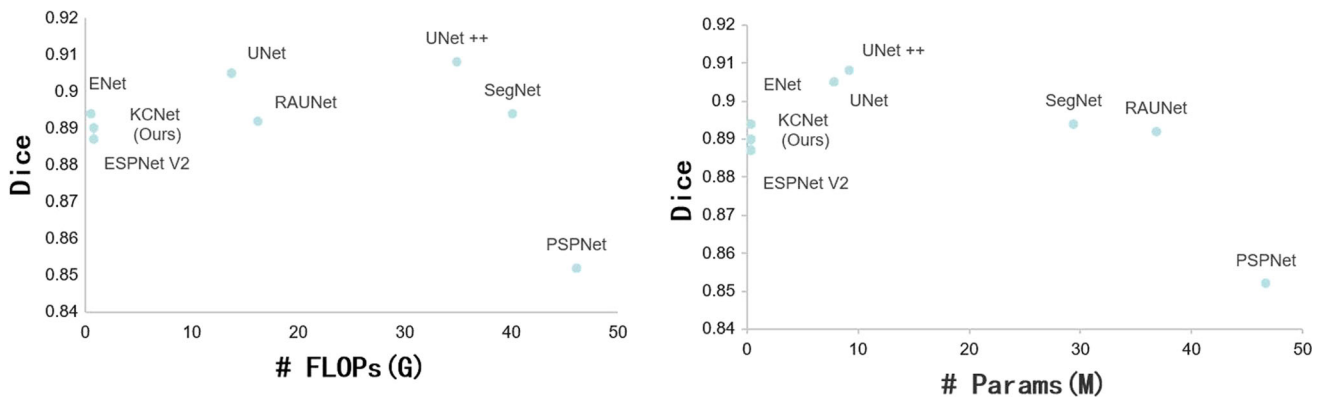


Fig. 4 Evaluation metric Dice (bigger is better) against FLOPs and the number of parameters for various methods

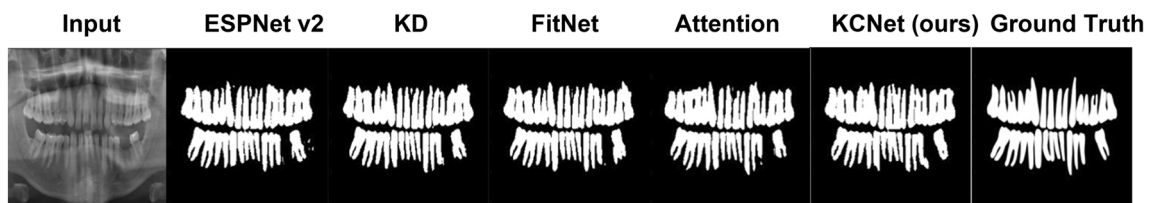


Fig. 5 Examples of the segmented image using lightweight methods

different teacher and student networks was also investigated. Table 5 shows the Dice results when teacher networks were UNet and UNet ++, and students were ENet and ESPNet v2. It shows that when teacher network was

UNet ++ and student ENet, it achieved the best performance. However, in other experiments, we chose UNet++ and ESPNet v2 as our teacher and student networks, respectively, as ESPNet v2 is slighter than ENet.

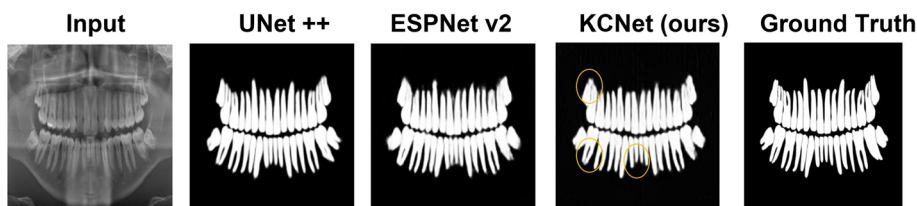


Fig. 6 An example of the segmented images using teaching network UNet++, student network ESPNet v2 and our KCNet. The yellow circles indicate improvements over student network

Table 4 Dice results obtained by using various values of those harmonic factors α , β and γ

α	1e5	1e3	1e2	1e4	1e4	1e4
β	0.05	0.05	0.05	0.5	0.005	5
γ	0.95	0.95	0.95	0.5	0.995	– 4
Mean	N/A	0.898	0.884	0.888	0.895	0.366
Std	N/A	0.033	0.049	0.048	0.042	0.050

The teacher and student models are UNet++ and ESPNet v2, respectively

Table 5 Dice results obtained by using different teacher and student networks

Student	Teacher	
	UNet	UNet++
ENet	0.902 ± 0.025	0.902 ± 0.024
ESPNet v2	0.898 ± 0.025	0.890 ± 0.038

Table 6 Architecture of ESPNet v2 [25]

Layer	Output size	Nos of layers	Nos of output channel
Input	256 × 256	1	–
Convolution	128 × 128	2	32
Strided EESP	64 × 64	2	64
Strided EESP	32 × 32	2	128
EESP	32 × 32	4	128
Strided EESP	32 × 32	2	256
EESP	32 × 32	8	256
Up-sampling	64 × 64	1	128
Up-sampling	128 × 128	1	128
Up-sampling	256 × 256	1	1
Sigmoid	–	–	1

The EESP and Strided EESP blocks can be found in [25]. The convolution layer has filter size 3 × 3 and 2 strides

Table 7 Sensitivity of our algorithm to impulse noise

Noise rate	0	0.01	0.1	0.2	0.3	0.4
Dice	0.912	0.910	0.906	0.882	0.829	0.457

Table 8 Sensitivity of our algorithm to Gaussian noise

Mean	Variance					
	0	1	72	100	120	144
0	0.912	0.910	0.892	0.833	0.729	0.350
5	–	0.909	0.891	0.829	0.675	0.231
10	–	0.908	0.885	0.780	0.554	0.218

For assessing the sensitivity of our proposed algorithm to noise, our algorithm was applied to images with various levels of noise. Two different kinds of noises were considered which are impulse noise and Gaussian noise. For the impulse noise, Table 7 shows the Dice performance on five different noise rates. It shows that the Dice performance was pretty good when noise level was less than 0.3, but reduced dramatically to 0.457 when noise rate was set to 0.4, i.e., 40% data points were interrupted by noise. Similarly, our algorithm was tested on Gaussian noise. Table 8 shows the performance in terms of Dice when various Gaussian noise levels were added to an image. As expected, when either the mean value or the variance was increased, the Dice value was decreased. Figure 7 shows the examples of the input image without noise and the images with the two different types of noises, as well as their segmented results. In addition, the algorithm was tested on flipped and cropped images. Figures 8 and 9 show that our algorithm was working well on such flipped and cropped images.

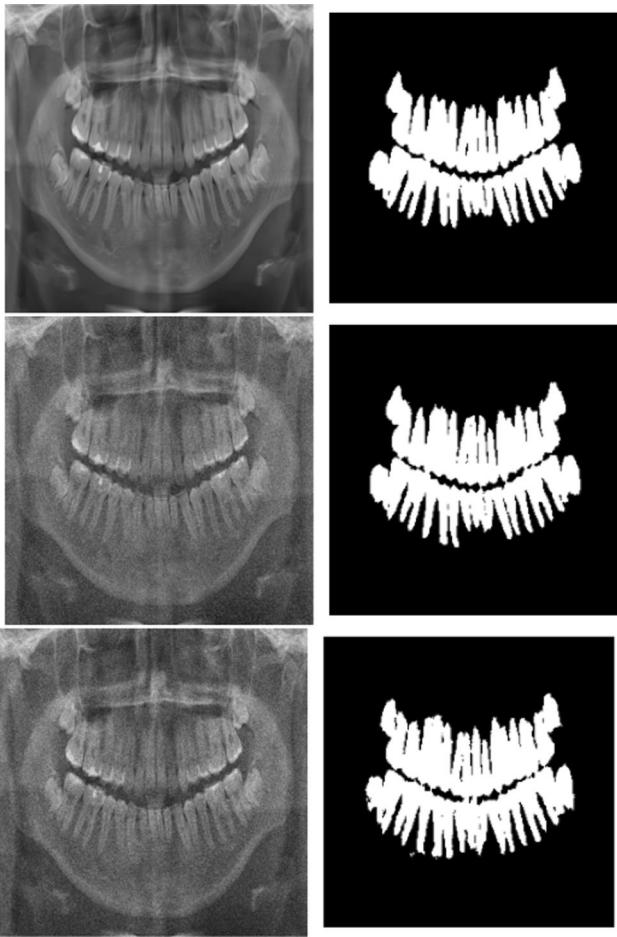


Fig. 7 Input images with no noise (top), impulse noise with noise rate 0.2 (middle) and Gaussian noise with zero mean and variance 72 (bottom). Their segmentation results are on the right

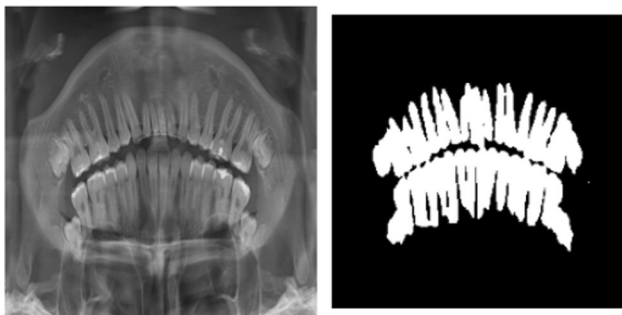


Fig. 8 Flipped image and segmentation results

5 Discussions

There are a few data sets of dental X-ray images available in the literature which are reviewed in the paper [3], but most of them are not accessible. The data set used in this paper had been studied for segmentation in the literature by using various machine learning algorithms. For example, a

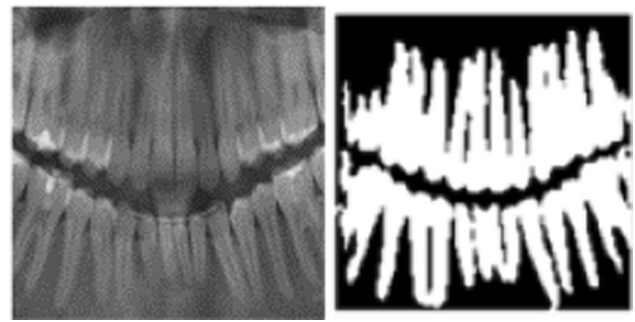


Fig. 9 Center-cropped image (128×128) and segmentation result

number of traditional methods including region-based, threshold-based, cluster-based, boundary-based and watershed-based methods were applied to this data set for segmentation [3]. Besides, a deep learning method, i.e., MASK R-CNN, was also applied to the data, which achieved 0.9208 in terms of classification accuracy. The results clearly indicate that the deep learning algorithm outperforms any traditional methods greatly.

Some other deep learning algorithms were also tested on this data set. For example, the work of [43] proposed a multi-scale location perception network for dental X-ray image segmentation and achieved 0.9301 accuracy in terms of Dice; the work of [16] proposed a two-stage attention network scheme for segmentation which achieved 0.9272 in terms of Dice, and in the same paper, the authors also reported 0.8933 in terms of Dice by using UNet. In our work, Table 2 shows that the Dice was 0.905 by using UNet which is similar to the results reported in other work. The result variability in terms of Dice using UNet may come from the data splitting, which would happen to any machine learning algorithms. There is other work using a different dental X-ray image data for segmentation using UNet which achieved 0.94 in terms of Dice. However, the results may not be comparable because the results were from different data sets, but at least it indicates that the performance of UNet for dental X-ray image segmentation was stable across different data and various research groups.

In addition, we also applied other algorithms to this data set. Our results show that UNet++ had the best performance in a number of evaluation metrics, which drives us to use it as our teacher network. By reviewing the literature, we can see that Dice was a popular evaluation metric for dental X-ray image segmentation. We also applied other metrics for evaluating these algorithms. Besides, all these algorithms applied to the data in this paper were evaluated in terms of the number of model parameters and FLOPs, which are rarely seen in other work on dental X-ray image segmentation. As we have stated in the results section, these metrics in terms of model size and computational efficiency are also important for evaluating a

model specifically for the purpose of deployment. For example, although UNet++ achieved the best performance, the size of the model and FLOPs are high which would prevent to deploy on edge devices. Precisely, the UNet++ requires 9.2 million parameters which is approximately 210 megabytes for memory storage. Of course, an option for deploying these large models would be using cloud computing. However, an obvious obstacle for using cloud computing is that the machine device such as panoramic dental X-ray machine has to be connected to Internet, which would not be possible in many scenarios. Even it is connected to Internet, it would be disrupted if the network was down. A second obstacle would be that the machine must have large memory for storing these trained models, which might not be a problem, but it would need much financial cost. As we have stated, our purpose is to devise efficient models to reduce both the financial and computational costs and maintain high accuracy which is comparable to large models. The results have shown that the sizes of our lightweight models are greatly reduced (~ 7.5 megabytes) and the computational cost. This indicates that these algorithms would be able to deploy on edge devices; for example, these models could be directly deployed on panoramic dental X-ray machines. Dental practitioners could immediately use the results on the machines.

Nevertheless, we would further investigate to improve the methods developed in this paper and develop other methods to achieve improved performances. For instance, ensemble methods based on UNet have improved performance than single methods [23]. A straightforward approach would be an ensemble method which combines those lightweight methods investigated in this paper. In theory, such ensemble method would definitely have better performance while maintaining low memory and computational costs. In addition, other advanced methods such as [30, 31, 44, 45] could be used to improve the performance of dental X-ray segmentation.

Furthermore, the proposed method could be applied to instance segmentation. For example, it would be interesting to investigate the performance of those lightweight algorithms for caries detection [46–48], which is more important and realistic than semantic segmentation. The proposed method could also be applied to segmenting mandible in panoramic X-ray [49]. This would be plausible for eventually deploying these models on edge devices for clinical use.

The proposed lightweight method could potentially be applied to other interesting directions including human face detection in risk situation [50], CT brain tumor detection [51] and IoT-based framework for disease prediction in pearl millet [52]. It would be interesting to compare our method to these methods in these directions.

6 Conclusions

In this paper we have proposed a method for panoramic dental X-ray image segmentation. Compared to various other methods, our algorithm achieved best performance in some metrics and comparative in other metrics. Interestingly, our method is a lightweight method which could be deployed on edge devices. Our method is also compared to three other lightweight methods and achieved the best performance in two out of five evaluation metrics. In our future research, we will interpret our model and investigating what has been learnt by using knowledge consistency networks.

Data availability The data sets analyzed during the current study were published in [3] and are available in the following repository: <https://github.com/IvisionLab/dental-image>.

Declarations

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Terlemez A, Tassoker M, Kizilcakaya M, Gulec M (2019) Comparison of cone-beam computed tomography and panoramic radiography in the evaluation of maxillary sinus pathology related to maxillary posterior teeth: Do apical lesions increase the risk of maxillary sinus pathology? *Imaging Sci Dent* 49(2):115–122
2. Wang C-W, Huang C-T, Lee J-H, Li C-H, Chang S-W, Siao M-J, Lai T-M, Ibragimov B, Vrtovec T, Ronneberger O et al (2016) A benchmark for comparison of dental radiography analysis algorithms. *Med Image Anal* 31:63–76
3. Silva G, Oliveira L, Pithon M (2018) Automatic segmenting teeth in X-ray images: Trends, a novel data set, benchmarking and future perspectives. *Expert Syst Appl* 107:15–31
4. Girshick R, Donahue J, Darrell T, Malik J (2014) Rich feature hierarchies for accurate object detection and semantic segmentation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 580–587

5. Ren S, He K, Girshick R, Sun J (2015) Faster R-CNN: towards real-time object detection with region proposal networks. *Adv Neural Inf Process Syst* 28:91–99
6. Lempitsky V, Kohli P, Rother C, Sharp T (2009) Image segmentation with a bounding box prior. In: 2009 IEEE 12th international conference on computer vision, pp 277–284
7. Zhou X, Wang D, Krähenbühl P (2019) Objects as points. Preprint [arXiv:1904.07850](https://arxiv.org/abs/1904.07850)
8. Ronneberger O, Fischer P, Brox T (2015) U-Net: convolutional networks for biomedical image segmentation. In: International conference on medical image computing and computer-assisted intervention. Springer, pp 234–241
9. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ (2017) Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4700–4708
10. Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J (2018) UNet++: a nested U-Net architecture for medical image segmentation. In: Deep learning in medical image analysis and multimodal learning for clinical decision support, pp 3–11
11. Alom MZ, Hasan M, Yakopcic C, Taha TM, Asari VK (2018) Recurrent residual convolutional neural network based on U-Net (R2U-Net) for medical image segmentation. Preprint [arXiv:1802.06955](https://arxiv.org/abs/1802.06955)
12. Huang C, Han H, Yao Q, Zhu S, Zhou SK (2019) 3D U-Net: a 3D universal U-Net for multi-domain medical image segmentation. In: International conference on medical image computing and computer-assisted intervention. Springer, pp 291–299
13. Ni Z-L, Bian G-B, Zhou X-H, Hou Z-G, Xie X-L, Wang C, Zhou Y-J, Li R-Q, Li Z (2019) Raunet: residual attention U-Net for semantic segmentation of cataract surgical instruments. In: International conference on neural information processing. Springer, pp 139–149
14. Milletari F, Navab N, Ahmadi S-A (2016) V-net: fully convolutional neural networks for volumetric medical image segmentation. In: 2016 fourth international conference on 3D vision (3DV). IEEE, pp 565–571
15. Yang J, Xie Y, Liu L, Xia B, Cao Z, Guo C (2018) Automated dental image analysis by deep learning on small dataset. In: 2018 IEEE 42nd annual computer software and applications conference (COMPSAC), vol 01, pp 492–497. <https://doi.org/10.1109/COMPSAC.2018.00076>
16. Zhao Y, Li P, Gao C, Liu Y, Chen Q, Yang F, Meng D (2020) TSASNet: tooth segmentation on dental panoramic X-ray images by two-stage attention segmentation network. *Knowl-Based Syst* 206:106338
17. Muresan MP, Barbura AR, Nedevschi S (2020) Teeth detection and dental problem classification in panoramic X-ray images using deep learning and image processing techniques. In: 2020 IEEE 16th international conference on intelligent computer communication and processing (ICCP). IEEE, pp 457–463
18. Kong Z, Xiong F, Zhang C, Fu Z, Zhang M, Weng J, Fan M (2020) Automated maxillofacial segmentation in panoramic dental X-ray images using an efficient encoder-decoder network. *IEEE Access* 8:207822–207833
19. Nader R, Smorodin A, De La Fourniere N, Amouriq Y, Autrusseau F (2022) Automatic teeth segmentation on panoramic X-rays using deep neural networks. In: International conference on pattern recognition
20. Cha J-Y, Yoon H-I, Yeo I-S, Huh K-H, Han J-S (2021) Panoptic segmentation on panoramic radiographs: Deep learning-based segmentation of various structures including maxillary sinus and mandibular canal. *J Clin Med* 10(12):2577. <https://doi.org/10.3390/jcm10122577>
21. Luo D, Zeng W, Chen LJ, Tang W (2021) Deep learning for automatic image segmentation of stomatology and its clinical application. *Front Med Technol*:68
22. Chen H, Zhang K, Lyu P, Li H, Zhang L, Wu J, Lee C-H (2019) A deep learning approach to automatic teeth detection and numbering based on object detection in dental periapical films. *Sci Rep* 9(1):1–11
23. Koch TL, Perslev M, Igel C, Brandt SS (2019) Accurate segmentation of dental panoramic radiographs with U-Nets. In: 2019 IEEE 16th international symposium on biomedical imaging (ISBI 2019). IEEE, pp 15–19
24. Paszke A, Chaurasia A, Kim S, Culurciello E (2016) Enet: a deep neural network architecture for real-time semantic segmentation. Preprint [arXiv:1606.02147](https://arxiv.org/abs/1606.02147)
25. Mehta S, Rastegari M, Shapiro L, Hajishirzi H (2019) Espnetv2: a light-weight, power efficient, and general purpose convolutional neural network. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 9190–9200
26. Ba J, Caruana R (2014) Do deep nets really need to be deep? *Adv Neural Inf Process Syst* 27
27. Hinton G, Vinyals O, Dean J, et al (2015) Distilling the knowledge in a neural network. Preprint [arXiv:1503.02531](https://arxiv.org/abs/1503.02531)
28. He T, Shen C, Tian Z, Gong D, Sun C, Yan Y (2019) Knowledge adaptation for efficient semantic segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 578–587
29. Qin D, Bu J-J, Liu Z, Shen X, Zhou S, Gu J-J, Wang Z-H, Wu L, Dai H-F (2021) Efficient medical image segmentation based on knowledge distillation. *IEEE Trans Med Imaging* 40(12):3820–3831
30. Liu Y, Chen K, Liu C, Qin Z, Luo Z, Wang J (2019) Structured knowledge distillation for semantic segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 2604–2613
31. Liu R, Yang K, Liu H, Zhang J, Peng K, Stiefelshagen R (2022) Transformer-based knowledge distillation for efficient semantic segmentation of road-driving scenes. Preprint [arXiv:2202.13393](https://arxiv.org/abs/2202.13393)
32. Ho TKK, Gwak J (2020) Utilizing knowledge distillation in deep learning for classification of chest X-ray abnormalities. *IEEE Access* 8:160749–160761
33. Wang H, Zhang D, Song Y, Liu S, Wang Y, Feng D, Peng H, Cai W (2019) Segmenting neuronal structure in 3d optical microscope images via knowledge distillation with teacher-student network. In: 2019 IEEE 16th international symposium on biomedical imaging (ISBI 2019), IEEE, pp 228–231
34. Zagoruyko S, Komodakis N (2016) Paying more attention to attention: improving the performance of convolutional neural networks via attention transfer. Preprint [arXiv:1612.03928](https://arxiv.org/abs/1612.03928)
35. Romero A, Ballas N, Kahou SE, Chassang A, Gatta C, Bengio Y (2014) Fitnets: hints for thin deep nets. Preprint [arXiv:1412.6550](https://arxiv.org/abs/1412.6550)
36. Liang R, Li T, Li L, Wang J, Zhang Q (2019) Knowledge consistency between neural networks and beyond. In: International conference on learning representations
37. Silva B, Pinheiro L, Oliveira L, Pithon M (2020) A study on tooth segmentation and numbering using end-to-end deep neural networks. In: 2020 33rd SIBGRAPI conference on graphics, patterns and images (SIBGRAPI). IEEE, pp 164–171
38. Badrinarayanan V, Kendall A, Cipolla R (2017) Segnet: a deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans Pattern Anal Mach Intell* 39(12):2481–2495
39. Jin Q, Meng Z, Sun C, Cui H, Su R (2020) RA-UNet: a hybrid deep attention-aware network to extract liver and tumor in CT scans. *Front Bioeng Biotechnol* 1471
40. Zhao H, Shi J, Qi X, Wang X, Jia J (2017) Pyramid scene parsing network. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2881–2890

41. Yu C, Wang J, Peng C, Gao C, Yu G, Sang N (2018) Bisenet: bilateral segmentation network for real-time semantic segmentation. In: Proceedings of the European conference on computer vision (ECCV), pp 325–341
42. Wu T, Tang S, Zhang R, Cao J, Zhang Y (2020) Cgnet: a light-weight context guided network for semantic segmentation. *IEEE Trans Image Process* 30:1169–1179
43. Chen Q, Zhao Y, Liu Y, Sun Y, Yang C, Li P, Zhang L, Gao C (2021) Mslpnet: multi-scale location perception network for dental panoramic x-ray image segmentation. *Neural Comput Appl* 33(16):10277–10291
44. Zhang J, Yang K, Ma C, Reiß S, Peng K, Stiefelhagen R (2022) Bending reality: distortion-aware transformers for adapting to panoramic semantic segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 16917–16927
45. Jaus A, Yang K, Stiefelhagen R (2021) Panoramic panoptic segmentation: towards complete surrounding understanding via unsupervised contrastive learning. In: 2021 IEEE intelligent vehicles symposium (IV). IEEE, pp 1421–1427
46. Casalegno F, Newton T, Daher R, Abdelaziz M, Lodi-Rizzini A, Schürmann F, Krejci I, Markram H (2019) Caries detection with near-infrared transillumination using deep learning. *J Dent Res* 98(11):1227–1233
47. Ying S, Wang B, Zhu H, Liu W, Huang F (2022) Caries segmentation on tooth x-ray images with a deep network. *J Dent* 119:104076
48. Zhu H, Cao Z, Lian L, Ye G, Gao H, Wu J (2022) Cariesnet: a deep learning approach for segmentation of multi-stage caries lesion from oral panoramic X-ray image. *Neural Comput Appl*:1–9
49. Abdi AH, Kasaei S, Mehdizadeh M (2015) Automatic segmentation of mandible in panoramic X-ray. *J Med Imaging* 2(4):044003
50. Wiczorek M, Siłka J, Woźniak M, Garg S, Hassan MM (2022) Lightweight convolutional neural network model for human face detection in risk situations. *IEEE Trans Ind Inf* 18(7):4820–4829. <https://doi.org/10.1109/TII.2021.3129629>
51. Woźniak M, Siłka J, Wiczorek M (2021) Deep neural network correlation learning mechanism for CT brain tumor detection. *Neural Comput Appl*. <https://doi.org/10.1007/s00521-021-05841-x>
52. Kundu N, Rani G, Dhaka VS, Gupta K, Nayak SC, Verma S, Ijaz MF, Woźniak M (2021) IoT and interpretable machine learning based framework for disease prediction in pearl millet. *Sensors* 21(16):5386

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.