TOOTH SEGMENTATION IN 3D CONE-BEAM CT IMAGES USING DEEP CONVOLUTIONAL NEURAL NETWORK

S. Khan[*], *A. Mukati*[*], *S.S.H. Rizvi*[*], *N. Yazdanie*[†]

Abstract: Segmentation of an individual tooth in dental radiographs has great significance in the process of orthodontics surgeries and dentistry. Machine learning techniques, especially deep convolutional neural networks can play a key role in revolutionizing the way orthodontics surgeons and dentists work. Lately, many researchers have been working on tooth segmentation in 3D volumetric dental scans with a great degree of success, but to the best of our knowledge, there is no pre-trained neural network available publicly for performing tooth segmentation in 3D cone-beam dental CT scans. The methods which so far have been proposed by the researchers in this domain are based on complex multistep pipelines. This lack of the availability of a pre-trained model blocks the path for further explorations in this domain. In this research, we have produced a deep learning model for tooth segmentation from CBCT dental radiographs. The proposed model can segment teeth in CBCT scans in a single step. To train the proposed model, we obtained a dataset consisting of 70 3D CBCT volumes from a local health facility. We labeled the ground truth through a semi-automatic method and trained our neural network. The training yielded a validation accuracy of 95.57 % on a binary class semantic segmentation of the 3D CBCT volumes. The model is successfully able to segment teeth, regardless of their type from the background in a single step. This eliminates the need of having a complex and lengthy pipeline which many researchers have been proposing. The proposed model can be extended by incorporating labeling schemes. The custom labeling schemes will help healthcare professionals to perform the labeling as per their needs. The produced model can also provide a basis for further research in this domain.

Key words: *cone-beam computed tomography (CBCT), deep-learning, volumetric semantic segmentation, deep convolutional neural network*

Received: January 21, 2021 **DOI:** 10.14311/NNW.2022.32.018
Revised and accepted: December 31, 2022

[*]Shahid Khan – Corresponding author; Altaf Mukati; Syed Sajjad Hussain Rizvi; Shaheed Zulfikar Ali Bhutto Institute of Science and Technology, M-43, 100, SZABIST, Block 5, Clifton, Karachi, Pakistan 75500, E-mail: shahid.khan@szabist.edu.pk altaf.mukati@szabist.edu.pk dr.sajjad@szabist.edu.pk

[†]Nazia Yazdanie; FMH College of Dentistry, Lahore, E-mail: nazia508@gmail.com

1. Introduction

The application of machine learning models in the medical field has become popular. Machine learning has been used for the analysis, diagnosis, and understanding of the medical conditions in various scanning methods. The reason behind it is the exceptional results produced by deep convolutional neural networks in the recent past. Common applications of semantic segmentation through deep convolutional neural networks in the field of medical imaging are tooth segmentation [2, 14, 20, 4], Alzheimer's disease detection [5], breast cancer diagnosis [7], osteoporosis detection in panoramic jaw radiographs [13], detection of COVID-19 in chest X-ray radiographs [9], and more [3, 21]. Transfer learning from popular deep convolutional neural networks has been producing exceptional results in classification and segmenting teeth in dental radiographs too [13].

Isolation of a single tooth is important in dentistry for critical analysis and diagnosis of the patient's dental condition. In dentistry, generally, three types of scans are used by the practitioners.

- 2D dental radiographs [10]
- 3D surface mesh optical scanning [28]
- CBCT dental radiographs [2, 20, 4]

The use of the type of scan depends on the condition of the patient and the need of the practitioner. The 2D dental radiographs provide basic X-ray imaging [10]. The 3D surface mesh is obtained through optical scanners. The 3D surface mesh optical scans provide detail of the surface of the teeth in 3D space [28]. The state-of-the-art in dental radiography is the cone-beam CT scanning (CBCT). The CBCT scans provide the volumetric scan of the jaw [4]. The inside detail of the teeth and jawbones, including soft tissues, gingiva, and soft bones are visible in form of 3D volume.

Though CBCT scans are currently the most advanced form of dental scanning, the other forms of scanning still have significance. Dentists and orthodontics surgeons prescribe the type of scanning based on the condition of the patient and based on the type of procedure to be performed on the patient. For example, a panoramic 2D X-ray is advised by the dentist to examine a broken or cracked tooth. The 3D dental surface optical scanning is advised before scaling of the teeth or when alignment of the teeth is to be performed. The CBCT dental scans provide inside detail of the teeth. A CBCT scan is referred by the dentists when there are nonerupted and overlapping teeth inside the gingiva of the patient. CBCT scanning also helps in conditions such as broken teeth, cracked teeth, rotten teeth, root canal treatment, alignment, filling and filing of the teeth, and extraction of the teeth.

2. Related work

Medical image processing started as soon as the computer experts found a way to acquire and load medical imagery into the computers. They started to work on

easing and automating the analysis and diagnostic process through medical image processing [16]. The methods of medical image processing started with basic imaging processing techniques. For example, edge detection through geometrical methods, and evolved to the state-of-the-art deep convolutional neural networks based methods [16]. Today a variety of diseases and disorders detection have been made through machine learning. From the detection of thin hairline bone fracture in X-ray to the detection of novel coronavirus infection, that emerged in 2019, in chest X-rays are being done through deep machine learning models. The segmentation and classification of human teeth are also not an exception. It has been done through computer-aided systems for decades. Like any other medical image processing, tooth segmentation has also been evolved from basic geometrical detection methods through the state-of-the-art deep convolutional neural networks [28]. Following is a brief introduction of the similar work carried out by the researchers.

It is established among the researchers in the field of the segmentation of the teeth in CBCT volumetric radiography, that segmenting and identifying an individual tooth through a single-stage multiclass neural network is unrealistic [2]. A group of researchers in the year 2020 proposed a pipeline for segmenting and labeling of teeth in two stages [2]. They first used a fully connected neural network to identify tooth surface and tooth region and in the second stage, they used a marker-controlled watershed algorithm for reconstructing the individual tooth in 3D space.

In 2019, a group of researchers from The University of Hong Kong proposed a two-stage methodology for segmentation and identification of teeth in CBCT volumetric radiographs [4]. They proposed a two-stage network for achieving highly accurate tooth segmentation and labeling for identification. In the first stage of the proposed method, an edge map is extracted and in the second stage, the edge maps are passed to a region proposal network. A novel-learned similarity matrix is used to eliminate the redundant proposals. The elimination of the redundant proposal helps them in speeding up the processing by reducing the workload significantly.

In late 2019, a novel coronavirus emerged in China which caused SARS like diseases. The disease caused by this novel coronavirus was called COVID-19. The infection caused by the novel coronavirus first infects the upper respiratory system but eventually ends up in the lungs, causing pneumonia. A group of researchers developed a machine learning model to detect the COVID-19 in chest CT scans. The model is based on learning of structured latent, multiview representations [9]. The proposed model can successfully discriminate COVID-19 from regular pneumonia despite having similar symptoms and similar appearance. In their experiments, they used a set of 2,522 chest CT scans of variety of patients including male and female and of different age groups. Since the conditions and symptoms of pneumonia and COVID-19 are similar, and also the COVID-19 turns into pneumonia if not treated properly, the researcher has included an inverse dataset as well. The total number of 2,522 scans they used had 1,495 COVID-19 patients and the rest of the 1,027 CT scans from pneumonia patients. The researchers have reported that their model produced 95.5 % accuracy, 96.6 % sensitivity, and 93.2 % specificity.

With the use of deep learning in tooth segmentation, researchers have been using deep learning for detecting bone diseases like osteoporosis in dental radiographs. Most segmentation problems are solved by researcher by retraining pre-trained

networks such as GoogLeNet, VGG and ResNet. A group of researchers have successfully performed transfer learning from VGG16 to achieve high accuracy in detecting osteoporosis in dental radiographs [12]. To improve the accuracy, the researchers finetuned VGG16 by modifying weights and achieved an accuracy of 85.8 %.

Depending on the application, equipment used, and the kind of dataset in hand, the methods of segmentation and classification are divided into the following three broad categories.

- Segmentation in 2D radiographs [1, 25]
- Segmentation in 3D volumetric radiographs [27, 17, 19]
- Segmentation in 3D surface mesh dataset [28]

The literature review suggests that the above types of tooth segmentation are attempted by the researchers using the following methods.

- Manual segmentation [29]
- Semi-automatic segmentation [15, 6]
- Fully automatic segmentation [18, 28]

While the manual segmentation usually involves an operator performing the whole segmentation manually using pointing, dragging, outlining, and clicking, the other two i.e. semi-automatic, and fully automatic methods have been attempted by the researchers using the following models for general purpose or tooth segmentation.

- Geometrical models [26, 22]
- Active contour based models [30]
- Machine learning-based models [18, 28]

The machine learning-based models include the following methods:

- Unsupervised learning [23]
- Supervised learning [18, 28]

Since the research in hand is focused on a supervised machine learning model and especially on the use of deep convolutional neural networks for segmenting and classifying individual tooth, we will look deeper into the research found in the literature related to this domain.

In the research presented in [18], the researchers have performed transfer learning from AlexNet [18] to perform segmentation, classification, and labeling in a 3D CBCT dataset. The objective of this research is to automate the process of forensic dental identification for identifying individuals in cases of mass disasters like earthquakes, tsunamis or fire outbreaks, etc. The automation involves the comparison of the postmortem dental CT report with the antemortem dental CT scan report.

To better understand the 3D CBCT images and to simplify the identification of the individuals, they transformed the 3D CBCT volumetric radiographs into 2D panoramic images. Total of 52 original volumes were used as the training dataset. For more simplification, they excluded teeth with artificial installations such as braces, fillings, implants, etc. Due to the shortage of dataset of wisdom teeth, they also excluded the wisdom teeth from segmentation and classification.

Since it is known that the machine learning models need a huge amount of data to be able to train and avoid overfitting, they performed common data augmentation methods to supplement the dataset. The common technique they used included rotating the original images +10 and −10 degrees and intensity transformation using the gamma correction method.

They also exploited the AlexNet [11] to avoid drafting a deep CNN from scratch. With transfer learning from the AlexNet and with the above-mentioned data augmentation techniques, they achieved an overall accuracy of 91.0 % in segmentation and classification, which was favorable for identifying individuals in cases of mass disasters.

In another research [28], deep CNN is used to segment and classify individual teeth. The dataset used in this research is different from the one used in the research earlier. The scientists used 3D surface mesh to experiment with tooth segmentation and classification through deep convolutional neural network. A novel model has been presented which achieved 99.06 % accuracy generally and achieved favorable accuracy in condition of rotten, missing, crowded, and featureless teeth.

2.1 Gap analysis

The literature review suggests that CNN based models are state-of-the-art in this domain. However, the methods we found in the literature are mostly based on multiple steps. Also, there is no publicly available pre-trained neural network for evaluation. It is important to have an efficient pre-trained model for automatically segmenting teeth in CBCT radiographs. This can pave the path towards effective utilization of the CNN based model in practical dentistry.

3. Methodology

In our previous work, we proposed a deep learning model for segmenting 2D panoramic dental radiographs into 9 classes. We proposed a feedforward deep convolutional neural network with 5 sets of convolutional layers by batch normalization and ReLu layers. We trained the model with 116 2D panoramic dental radiographs [10]. The model yielded an accuracy of 80 % accuracy, surpassing VGG16 [24] which yielded 72 % accuracy in similar training conditions.

For this research, we modified our network to add the capability to handle 3D volumetric data. After testing several different configurations, we finally obtained a stable version of the network with 38 layers that yielded promising results.

3.1 Dataset and statistics

It is a common understanding among researchers that a deep learning model will always need a huge amount of training data. It is not always easy to obtain a huge dataset that is enough to train a deep learning model. Especially in the domain of medical image processing, it is even difficult to manage a huge dataset. The other problem with the research at hand that we faced is to find an annotated dataset that can be used to train our model for performing semantic segmentation. The CBCT radiography is relatively a new technology in Pakistan. We did not have many options. Luckily, a local dental college and hospital agreed to provide us the required dataset. We obtained 70 3D CBCT volumetric radiographs in DICOM format from this facility.

The figures show a raw instance from the 3D CBCT dataset. Fig. 1 shows a side-to-side view of the 3D volume, Fig. 2 shows the same volume from the front

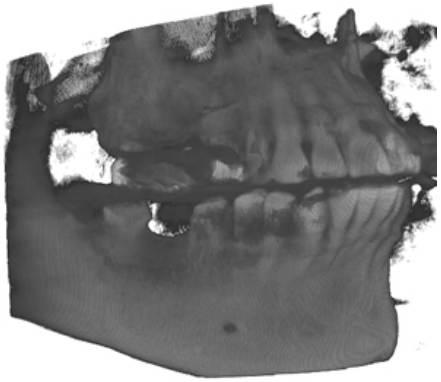

Fig. 1 *Side view of a raw CBCT volume.*

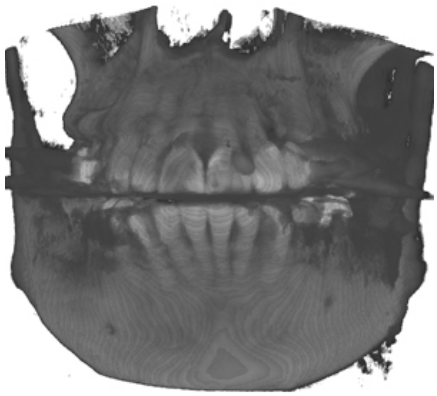

Fig. 2 *Front view of a raw CBCT volume.*

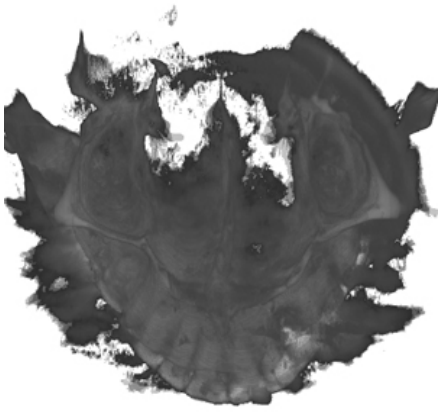

Fig. 3 *Top-bottom view of a raw CBCT volume.*

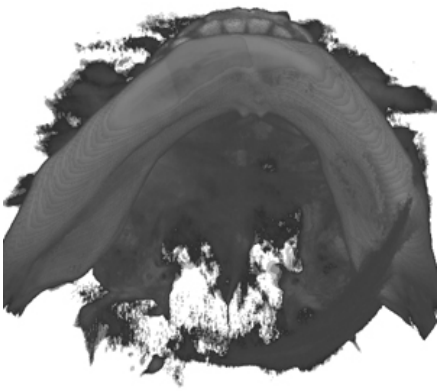

Fig. 4 *Bottom-up of a raw CBCT volume.*

side. In Fig. 3, we can see the same volume from the top and in Fig. 4 it is shown from the bottom-top side.

Although, we obtained a dataset that was good for our research, the problem that we faced was that the dataset was not annotated. Annotating a 3D volumetric dataset is a very tedious and laborious job and it is not possible to annotate many 3D volumes manually. So, we selected a subset of the volumes and semi-automatically labeled them. We used the interpolation technique for labeling the dataset.

In Fig. 5, a plane from the maxilla is seen. The categorical labels of the same slice can be seen in Fig. 6. Similarly, Fig. 7 is a slice from the mandible. The categorical labels for the same are seen in Fig. 8.

Fig. 5 *A slice of maxilla.*

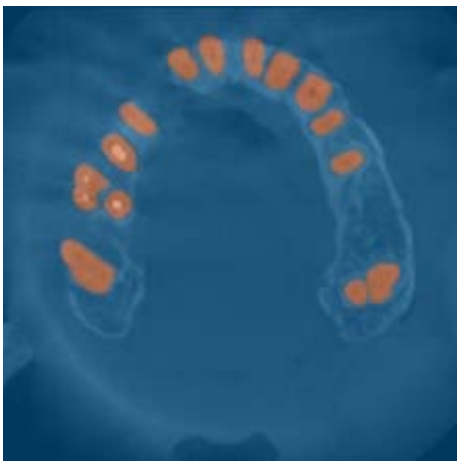

Fig. 6 *Ground truth for the slice of maxilla.*

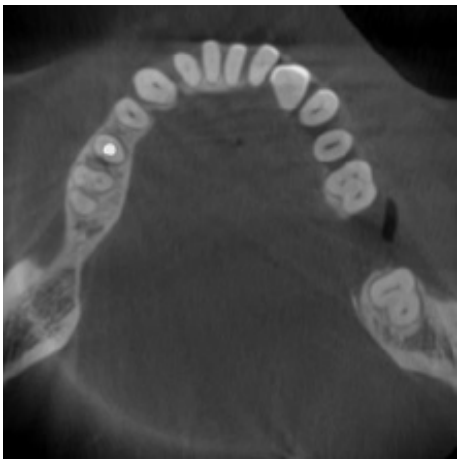

Fig. 7 *A slice of mandible.*

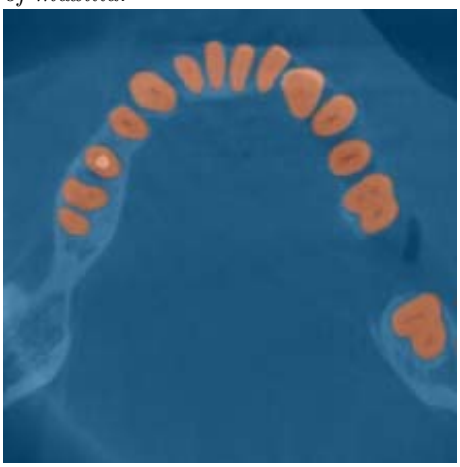

Fig. 8 *Ground truth for the slice of mandible.*

In Fig. 9, Fig. 10, Fig. 11, and Fig. 12, we can see the 3D volume with the ground truth annotation for the teeth from different angles.

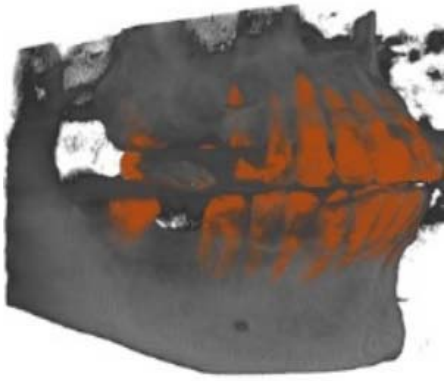

Fig. 9 *Ground truth volume side view.*

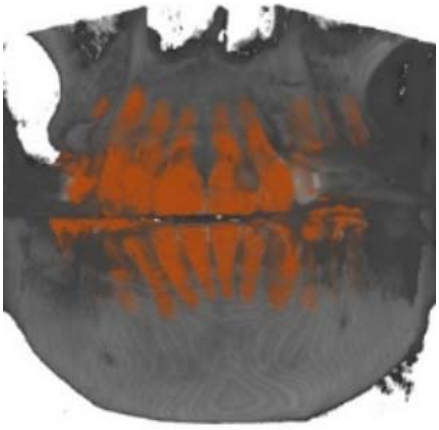

Fig. 10 *Ground truth volume front view.*

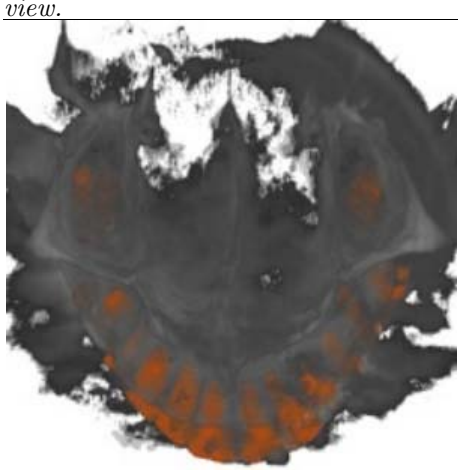

Fig. 11 *Ground truth volume top-bottom view.*

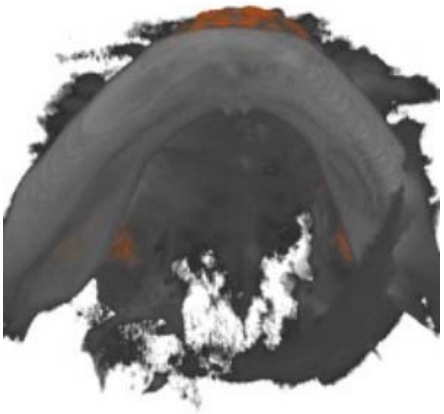

Fig. 12 *Ground truth volume bottom-up view.*

3.2 Preprocessing

Our dataset is obtained through a variety of CBCT equipment. The manufacturers of the radiography equipment do not follow any common standard grey level intensities for their radiographs. The setting in the software packages that work with the CBCT equipment allows the operators to select the dimensions and size of the scans. The volumes in our dataset varied from $450 \times 450 \times 450$ to $550 \times 550 \times 550$. To obtain a balance between the size of the volume and the quality of the volume,

we resized the volumes to $500 \times 500 \times 500$ in the first step and then center cropped the volumes in $496 \times 496 \times 496$ size. In addition to resizing and cropping, we applied histogram adjustments to balance the grey level intensities.

3.3 Model pipeline

The pipeline of our proposed methodology is illustrated in Fig. 13. The pipeline of our proposed methodology starts by inputting a raw CBCT dental radiograph straight from the CBCT scanner. In the first step, we remove any singleton dimensions from the volume. The input size of the volume that our method expects is $496 \times 496 \times 496$. To transform the input volume into the target input size, we first resize the volume to $500 \times 500 \times 500$, and then center crop the volume of the size $496 \times 496 \times 496$ to make the input volume uniform while retaining maximum information. The manufacturers of the radiography equipment do not follow any common or standard greylevel intensities for their radiographs. Histogram equalization is a technique to make the greylevel intensities uniform of a greyscale image. Since our dataset is obtained from more than one CBCT scanners, we employed histogram equalization on our dataset to make it uniform.

To perform supervised training of the model, the raw data must be annotated. To annotate the dataset to produce ground truth, we labeled our dataset into two classes. In our binary class problem, one class for a tooth, regardless of its label, and everything else is labeled as background.

Given the large size of the CBCT dental radiographs and given the computation power and time available, it was not possible to train the model on the full dataset

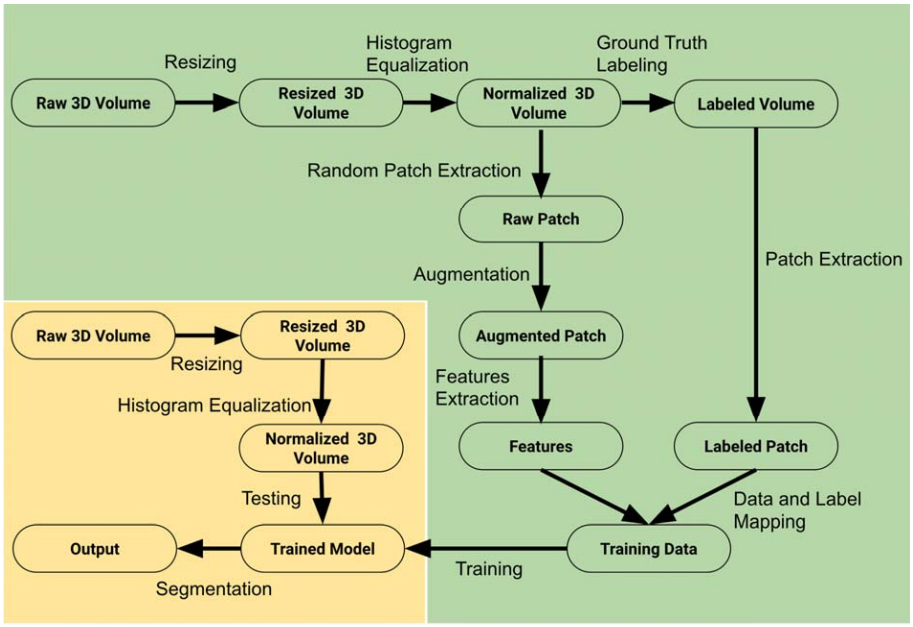

Fig. 13 *Pipeline of the proposed methodology.*

with full size. Keeping in mind these constraints, we broke down the volumes into small patches through the random patch extraction mechanism.

It is a common understanding among researchers that machine learning models need a huge amount of data for training. It is also widely accepted that the augmentation of the visual datasets usually increases the accuracy of the deep learning models by fulfilling the requirements of large datasets. To enlarge our dataset, we applied traditional dataset augmentation techniques to surplus our dataset. The visual data augmentation techniques include rotating left and right to 10 degrees, horizontal flipping, and resizing [10]. For this research, we employed only the flipping technique. The dataset size after augmentation is 140 volumes. We splitted the dataset into 60 %, 20 % and 10 % for training, validation and testing respectively.

In the next step of the proposed pipeline, we extract 16 random patches from each volume from random places with corresponding ground truth patches to pass it to the neural network for feature extraction and training. At the end of the proposed neural network, a $44 \times 44 \times 44$ set of labels are predicted. Backpropagation is applied for validating the results and weights are adjusted accordingly.

In the testing phase, a raw CBCT volume is first squeezed to eliminate any singleton dimensions. In the next step, the test volume is resized to $500 \times 500 \times 500$ voxels, and then center cropped to $496 \times 496 \times 496$ voxels, which is a suitable size for our deep network. Histogram adjustment is then applied as the last preprocessing step to balance the gray level intensities. The test volume is then passed to the network which spits out a label for each voxel in the volume to predict the teeth and background for semantic segmentation.

3.4 Network design

Our proposed network design consists of 38 layers having 11 blocks of 3D convolutional layers followed by batch normalization layers and Relu layers. The first layer of our network is a 3D input layer of size $132 \times 132 \times 132 \times 1$. The structure of the proposed models can be seen in Fig. 14. The input layer of our network does not perform any kind of normalizations on the input data. The input layer of our network is followed by 11 blocks of 3D convolutional layers which downsample the volumes to $44 \times 44 \times 44$.

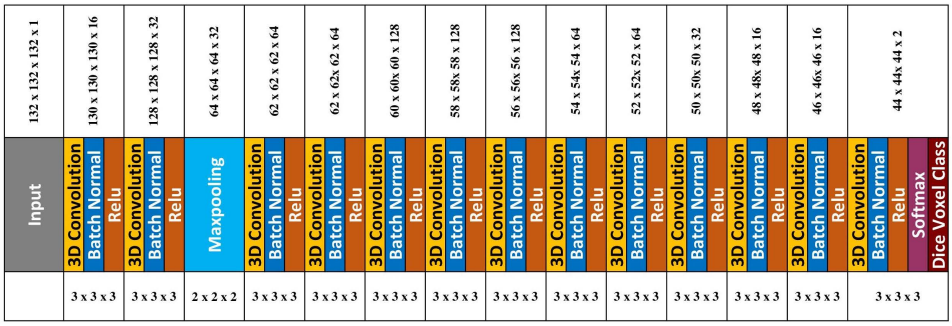

Fig. 14 *Network design of the proposed model.*

In the 3D convolutional layers, we used filters of $3 \times 3 \times 3$ voxels with a stride of 1 voxel at a time, i.e. $1 \times 1 \times 1$ with a dilation factor of $1 \times 1 \times 1$. In each convolutional block, we increased the number of filters starting from 16 filters to up to 128 filters, and then reduced the number of filters eventually to 2 to match the number of classes to the binary classification. The ratio of tooth voxels in any CBCT dental radiographs is 3 % as compared to the background voxels. To solve this huge class imbalance, we used Dice pixel classification layer. The Dice pixel classification layer provides a categorical label for each voxel using generalized Dice loss (GDL). Which eliminates the class imbalance problem. The GDL controls the influence that each class contributes to a loss through weighting the classes by the inverse size of the expected area.

3.5 Training

We carried the training through the adaptive moment estimation (Adam) optimizer for a fast convergence with the computing power constraints. The input to our network is a fixed $496 \times 496 \times 496 \times 1$ 3D volumetric image. The first layer of the network, that is the input layer extracts 16 random 3D patches from the volume of $132 \times 132 \times 132$ size for convolutions. We used the random patch extraction mechanism to extract the 3D patches because it is impossible to perform training on full-size 3D volume due to the limited computation power and available time. In parallel with the extraction of patches from the volume for training, we also extracted ground truth labels for the same patches. We used a small batch size of 8 for the training with an initial learning rate of 0.0001 with a drop factor of 0.95 on every 5 epochs. Our model yielded validation accuracy of 95.57 %. Fig. 15 shows that how the training process converged.

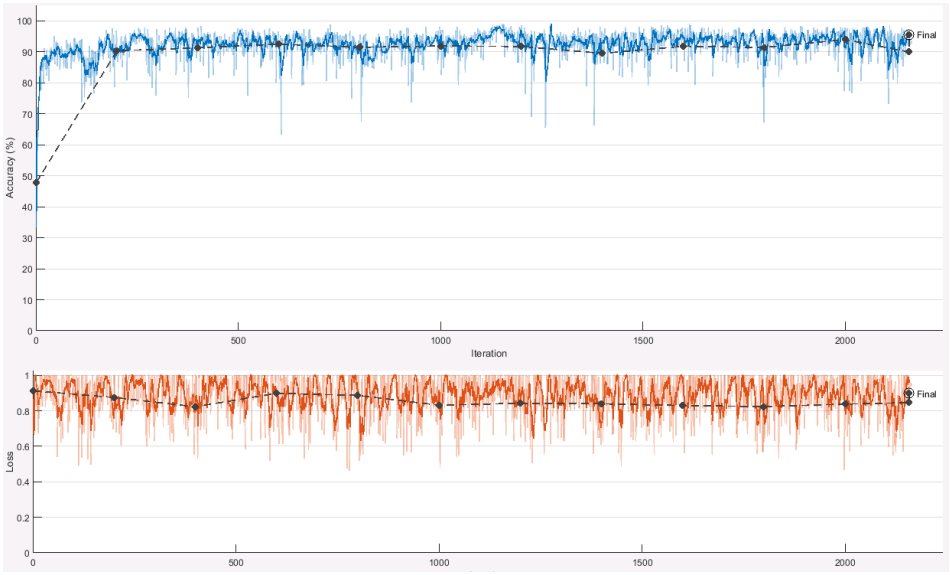

Fig. 15 *Training plot.*

3.6 Results

We achieved promising results with the proposed methodology as can be seen in Fig. 16 and Fig. 17. The arrow marker "1" in figure Fig. 16 and Fig. 17 indicates the accuracy of tooth segmentation, however, certain parts of the mandibula bone are also misclassified as teeth, which can be seen in Fig. 16 and Fig. 17 where marked with the arrow marker "2".

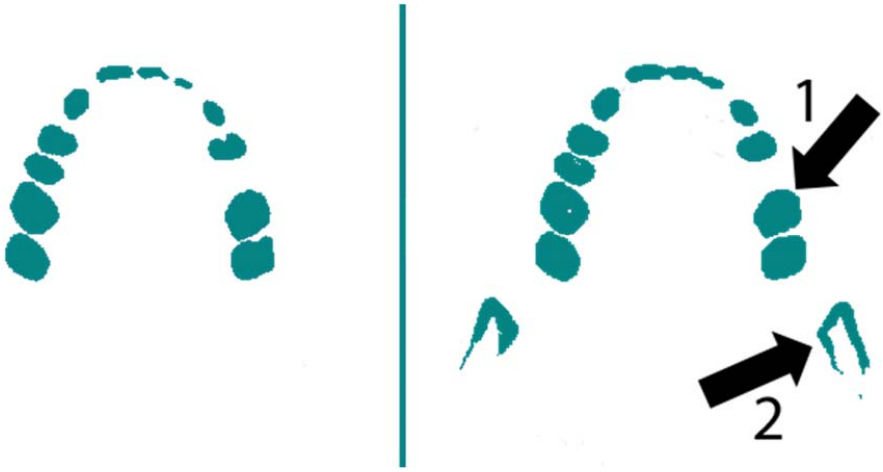

Fig. 16 *Ground truth (left) vs network predicted (right).*

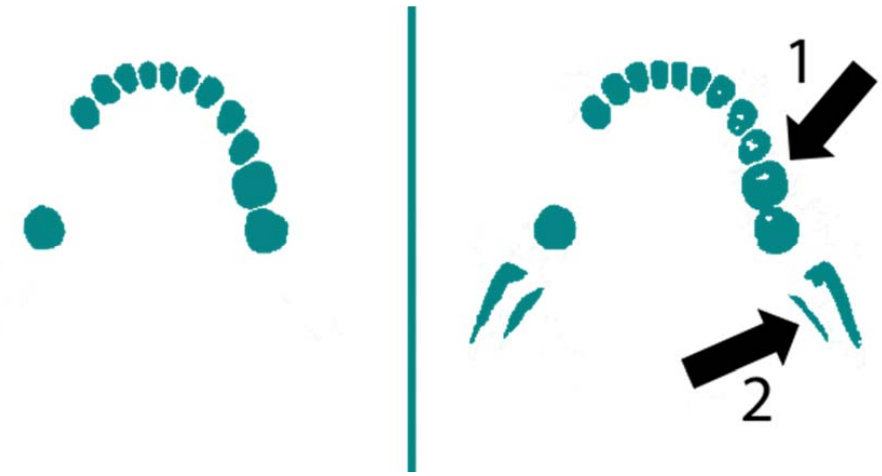

Fig. 17 *Ground truth (left) vs network predicted (right).*

It is noted that background voxels inside the teeth voxels are also correctly segmented. This indicates the exceptional segmentation performance by the proposed method as indicated by the arrow marker "1".

Fig. 18 shows the 3D reconstruction of the segmented slices. It is to be noted that the color labels as shown in Fig. 18 are not network-predicted. The colors are used for visibility purposes only. There are various labeling schemes followed by the researchers for classification of the teeth [28, 18, 4]. This suggests that the dental surgeons and medical practitioners follow different labeling schemes. We believe that tying the proposed model with any particular labeling scheme may reduce the usefulness of the proposed model. A custom labeling scheme can be incorporated through a secondary model.

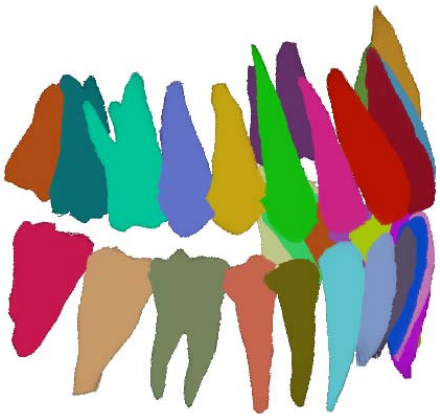

Fig. 18 *3D reconstruction of the segmented slices.*

3.7 Transfer learning for comparison

To compare the results of the model that we propose, we also performed transfer learning from 3D versions of popular deep learning models within the fair conditions.

3.7.1 3D U-Net

U-Net [21] is a deep convolutional neural network specially designed for working with biomedical images. U-Net [21] is essentially a 2D deep convolutional neural network for processing biomedical images in two dimensions. To process the volumetric data through U-Net, a third dimension is added to its layers with the same parameters as used in two-dimensional U-Net implementation. For our research, we adopted the 3D version of U-Net from this [3] research.

For retraining the 3D U-Net on our dataset, we used similar training conditions and with the same parameters as used in our proposed model. Similar to our proposed network, the 3D U-Net also accepts a 3D volume of $132 \times 132 \times 132 \times 1$. Since the dimension of our dataset is $496 \times 496 \times 496 \times 1$, we used the Random Patch Extraction mechanism to randomly extract 16 patches from each of our 3D volumes along with its ground truth and passed it to the input of the 3D U-Net as input. The 3D U-Net has 70 layers with several downsampling and upsampling layers.

The network eventually spits out a $44 \times 44 \times 44 \times 1$ response with categorical labels for each voxel. For retraining the 3D U-Net, we used the same parameters that we used to train our proposed network. As the optimizer, we used the adaptive moment estimation (Adam) with an initial learning rate of 0.0001 with a drop factor of 0.95 on every 5 epochs. Retraining the 3D U-Net on our dataset in fair conditions yielded an accuracy of 95.71 %, which is fractionally higher than the accuracy yielded by our proposed network. More detail on the comparison, that how our proposed model is still favorable is given in the results comparison section later in this report.

3.7.2 3D ResNet

ResNet [8] is a popular deep neural network that essentially performs classification on ImageNet 1000 class problem. In its essence, ResNet is a two-dimensional image classification neural network. To train the ResNet, we adopted the 3D version of the ResNet pre-trained model, used by the researchers to detect Alzheimer's disease in their research [5]. As compared to the promising results rendered by the model we proposed, and the results produced by the 3D U-Net, the results produced by the 3D ResNet are on the lower side. We achieved an accuracy of 80 % in similar training conditions.

3.8 Results comparison

Tab. I shows result comparison achieved by training the proposed method, 3D U-Net and 3D ResNet. The training is performed using Adam optimizer. The mini-batch size is 8 and the initial learning rate that we used is 0.0001 with a drop factor of 0.95 after every 5 epochs. The results show that the proposed model is by far favorable as compared to the other models.

A comparison of the 3 models is given in Tab. I.

	3D U-Net	DRNet	3D ResNet
Layers	58	38	71
Mean intersection over union (IoU)	0.60	0.70	0.48
Mean Dice score	0.86	0.90	0.36
Time to train [hours]	62	23	18
Validation accuracy [%]	95.75	95.54	80
Model size [MB]	70	4.3	121.4

Tab. I *Training results comparasion.*

The validation accuracy yielded by the 3D U-Net, the proposed model, and the 3D ResNet are respectively 95.75 %, 95.54 %, and 80 %. If we compare the validation accuracy of the proposed model with the 3D U-Net, we do not see any huge difference. However, if we compare the size of both the networks, the proposed network model is significantly smaller in size. We infer that the proposed model with a comparatively small size is as effective as 3D U-Net in terms of

validation accuracy. However, with a small size and less time to compute, the proposed method is favorable for this CBCT dental dataset. It is also noted that the proposed model size after training is very small as compared to the competitors. This makes the proposed model a better candidate for embedding directly into the CBCT equipment and supporting hardware.

The performance of the proposed model is compared with the competitor models in terms of mean intersection over union (IoU) and mean Dice score. Tab 1. shows that the proposed model has outperformed the competitor models by a considerable margin.

4. Conclusion

Segmentation of teeth in the computer-aided system in orthodontics and dentists' practice is a useful tool for critical analysis of the patient's condition. It helps dentists in many ways. To the best of our knowledge, there is no pre-trained model available publicly for segmenting teeth in 3D CBCT volumetric dataset. There is no dataset available publicly for experimenting on 3D volumetric dental data. We aim to produce a quality annotated 3D volumetric dental dataset and a pre-trained model for segmenting teeth in 3D CBCT volumes. We obtained a dataset of 70 3D CBCT volumes from a local health care facility and produced an annotated dataset. The accuracy of the annotation is certified by the health care facility for accuracy. To perform the segmentation through deep learning on our dataset, we designed a 3D deep convolutional neural network consisting of 38 layers and trained it. Our model yielded a validation accuracy of 95.54 %. To compare the results that we achieved, we retrained 3D U-Net and 3D ResNet on our dataset in the same scenario. The 3D U-Net produced an accuracy of 95.75 %, which is fractionally higher than the accuracy produced by our proposed model. But, the proposed model is favorable because of the relatively smaller size, less time required to train, and small model file size. The accuracy that the 3D ResNet produced is 80 %, which is relatively lower than both the proposed model and 3D U-Net.

Acknowledgement

We highly appreciate the cooperation and contribution of the Fatima Memorial Hospital and Zulfiqar Ahmad Atiq for providing expert domain knowledge related to dentistry and orthodontics that we required. Fatima Memorial Hospital also provided a dataset for this research. We also appreciate Ali Mobin for helping in choosing the hardware for the training of our model.

References

[1] ABDI A.H., KASAEI S., MEHDIZADEH M. Automatic segmentation of mandible in panoramic x-ray. *Journal of Medical Imaging*. 2015, 2(4), pp. 044003–044003, doi: https://doi.org/10.1117/1.JMI.2.4.044003.

[2] CHEN Y., DU H., YUN Z., YANG S., DAI Z., ZHONG L., FENG Q., YANG W. Automatic Segmentation of Individual Tooth in Dental CBCT Images From Tooth Surface Map by a Multi-Task FCN. *IEEE Access*. 2020, 8, pp. 97296–97309, doi: https://doi.org/10.1109/ACCESS.2020.2991799.

[3] ÇIÇEK Ö., ABDULKADIR A., LIENKAMP S.S., BROX T., RONNEBERGER O. 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19*, 2016, pp. 424–432. doi: https://doi.org/10.48550/arXiv.1606.06650.

[4] CUI Z., LI C., WANG W. ToothNet: Automatic Tooth Instance Segmentation and Identification From Cone Beam CT Images. In: *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 6361–6370. doi: https://doi.org/10.1109/CVPR.2019.00653.

[5] EBRAHIMI A., LUO S., CHIONG R. Introducing Transfer Learning to 3D ResNet-18 for Alzheimer's Disease Detection on MRI Images. In: *2020 35th International Conference on Image and Vision Computing New Zealand (IVCNZ)*, 2020, pp. 1–6. doi: https://doi.org/10.1109/IVCNZ51579.2020.9290616.

[6] EVAIN T., RIPOCHE X., ATIF J., BLOCH I. Semi-automatic teeth segmentation in cone-beam computed tomography by graph-cut with statistical shape priors. In: *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, 2017, pp. 1197–1200. doi: https://doi.org/10.1109/ISBI.2017.7950731.

[7] HAN Z., WEI B., ZHENG Y., YIN Y., LI K., LI S. Breast Cancer Multi-classification from Histopathological Images with Structured Deep Learning Model. *Scientific Reports*. 2017, 7(1), pp. 4172, doi: https://doi.org/10.1038/s41598-017-04075-z. ISSN 2045-2322.

[8] HE K., ZHANG X., REN S., SUN J. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778. doi: https://doi.org/10.48550/arXiv.1512.03385.

[9] KANG H., XIA L., YAN F., WAN Z., SHI F., YUAN H., JIANG H., WU D., SUI H., ZHANG C., et al. Diagnosis of coronavirus disease 2019 (COVID-19) with structured latent multi-view representation learning. *IEEE transactions on medical imaging*. 2020, 39(8), pp. 2606–2614, doi: https://doi.org/10.1109/TMI.2020.2992546.

[10] KHAN S., MUKATI A., ZULFIKAR S, BHUTTO A. Dataset augmentation for machine learning applications of dental radiography. *Int. J. Adv. Comput. Sci. Appl.* 2020, 11(2), pp. 453–456, doi: https://doi.org/10.14569/IJACSA.2020.0110258.

[11] KRIZHEVSKY A., SUTSKEVER I., HINTON G.E. ImageNet Classification with Deep Convolutional Neural Networks. In: F. PEREIRA, C. BURGES, L. BOTTOU, K. WEINBERGER, eds. *Advances in Neural Information Processing Systems*, 2012. Available also from: https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf.

[12] LEE J.-S., ADHIKARI S., LIU L., JEONG H.-G., KIM H., YOON S.-J. Osteoporosis detection in panoramic radiographs using a deep convolutional neural network-based computer-assisted diagnosis system: a preliminary study. *Dentomaxillofacial Radiology*. 2019, 48(1), pp. 20170344, doi: https://doi.org/10.1259/dmfr.20170344.

[13] LEE K.-S., JUNG S.-K., RYU J.-J., SHIN S.-W., CHOI J. Evaluation of transfer learning with deep convolutional neural networks for screening osteoporosis in dental panoramic radiographs. *Journal of clinical medicine*. 2020, 9(2), pp. 392, doi: https://doi.org/10.3390/jcm9020392.

[14] LEE S., WOO S., YU J., SEO J., LEE J., LEE C. Automated CNN-Based Tooth Segmentation in Cone-Beam CT for Dental Implant Planning. *IEEE Access*. 2020, 8, pp. 50507–50518, doi: https://doi.org/10.1109/ACCESS.2020.2975826.

[15] LI Z., WANG H. Interactive tooth separation from dental model using segmentation field. *PloS one*. 2016, 11(8), pp. e0161159, doi: https://doi.org/10.1371/journal.pone.0161159.

[16] LITJENS G., KOOI T., BEJNORDI B.E., SETIO A.A.A., CIOMPI F., GHAFOORIAN M., VAN DER LAAK J.A., VAN GINNEKEN B., SÁNCHEZ C.I. A survey on deep learning in medical image analysis. *Medical image analysis*. 2017, 42, pp. 60–88, doi: https://doi.org/10.1016/j.media.2017.07.005.

[17] MARIN I., PĂVĂLOIU I.-B., GOGA N., VASILĂŢEANU A., DRĂGOI G. Automatic contour detection from dental CBCT DICOM data. In: *2015 E-Health and Bioengineering Conference (EHB)*, 2015, pp. 1–4. doi: https://doi.org/10.1109/EHB.2015.7391424.

[18] MIKI Y., MURAMATSU C., HAYASHI T., ZHOU X., HARA T., KATSUMATA A., FUJITA H. Classification of teeth in cone-beam CT using deep convolutional neural network. *Computers in biology and medicine*. 2017, 80, pp. 24–29, doi: https://doi.org/10.1016/j.compbiomed.2016.11.003.

[19] PAVALOIU I.-B., GOGA N., VASILATEANU A., MARIN I., UNGAR A., PATRASCU I., ILIE C. Neural network based edge detection for CBCT segmentation. In: *2015 E-Health and Bioengineering Conference (EHB)*, 2015, pp. 1–4. doi: https://doi.org/10.1109/EHB.2015.7391414.

[20] RAO Y., WANG Y., MENG F., PU J., SUN J., WANG Q. A Symmetric Fully Convolutional Residual Network With DCRF for Accurate Tooth Segmentation. *IEEE Access*. 2020, 8, pp. 92028–92038, doi: https://doi.org/10.1109/ACCESS.2020.2994592.

[21] RONNEBERGER O., FISCHER P., BROX T. U-net: Convolutional networks for biomedical image segmentation. In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, 2015, pp. 234–241. doi: https://doi.org/10.48550/arXiv.1505.04597.

[22] SEPEHRIAN M., DEYLAMI A.M., ZOROOFI R.A. Individual teeth segmentation in CBCT and MSCT dental images using watershed. In: *2013 20th Iranian Conference on Biomedical Engineering (ICBME)*, 2013, pp. 27–30. doi: https://doi.org/10.1109/ICBME.2013.6782187.

[23] SIDI O., VAN KAICK O., KLEIMAN Y., ZHANG H., COHEN-OR D. Unsupervised co-segmentation of a set of shapes via descriptor-space spectral clustering. In: *Proceedings of the 2011 SIGGRAPH Asia Conference*, 2011, pp. 1–10. doi: https://doi.org/10.1145/2070781.2024160.

[24] SIMONYAN K., ZISSERMAN A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*. 2014, doi: https://doi.org/10.48550/arXiv.1409.1556.

[25] WONGWAEN N., SINTHANAYOTHIN C. Computerized algorithm for 3D teeth segmentation. In: *2010 International Conference on Electronics and Information Engineering*, 2010, pp. V1–277. doi: https://doi.org/10.1109/ICEIE.2010.5559877.

[26] XIA Z., GAN Y., CHANG L., XIONG J., ZHAO Q. Individual tooth segmentation from CT images scanned with contacts of maxillary and mandible teeth. *Computer methods and programs in biomedicine*. 2017, 138, pp. 1–12, doi: https://doi.org/10.1016/j.cmpb.2016.10.002.

[27] XU X., LI L., ZHANG L., WANG Q. A metal projection segmentation algorithm based on Random walks for dental CBCT metal artifacts correction. In: *2013 IEEE Nuclear Science Symposium and Medical Imaging Conference (2013 NSS/MIC)*, 2013, pp. 1–4. doi: https://doi.org/10.1109/NSSMIC.2013.6829361.

[28] XU X., LIU C., ZHENG Y. 3D tooth segmentation and labeling using deep convolutional neural networks. *IEEE transactions on visualization and computer graphics*. 2018, 25(7), pp. 2336–2348, doi: https://doi.org/10.1109/TVCG.2018.2839685.

[29] YAQI M., ZHONGKE L. Computer aided orthodontics treatment by virtual segmentation and adjustment. In: *2010 International Conference on Image Analysis and Signal Processing*, 2010, pp. 336–339. doi: https://doi.org/10.1109/IASP.2010.5476100.

[30] ZHANG S., XIE J., SHI H. Jaw Segmentation from CBCT Images. In: *2018 IEEE 23rd International Conference on Digital Signal Processing (DSP)*, 2018, pp. 1–5. doi: https://doi.org/10.1109/ICDSP.2018.8631819.