



OPEN

Canal-Net for automatic and robust 3D segmentation of mandibular canals in CBCT images using a continuity-aware contextual network

Bo-Soung Jeoun^{1,7}, Su Yang^{2,7}, Sang-Jeong Lee³, Tae-Il Kim⁴, Jun-Min Kim⁵, Jo-Eun Kim⁶, Kyung-Hoe Huh⁶, Sam-Sun Lee⁶, Min-Suk Heo⁶ & Won-Jin Yi^{2,6}✉

The purpose of this study was to propose a continuity-aware contextual network (Canal-Net) for the automatic and robust 3D segmentation of the mandibular canal (MC) with high consistent accuracy throughout the entire MC volume in cone-beam CT (CBCT) images. The Canal-Net was designed based on a 3D U-Net with bidirectional convolutional long short-term memory (ConvLSTM) under a multi-task learning framework. Specifically, the Canal-Net learned the 3D anatomical context information of the MC by incorporating spatio-temporal features from ConvLSTM, and also the structural continuity of the overall MC volume under a multi-task learning framework using multi-planar projection losses complementally. The Canal-Net showed higher segmentation accuracies in 2D and 3D performance metrics ($p < 0.05$), and especially, a significant improvement in Dice similarity coefficient scores and mean curve distance ($p < 0.05$) throughout the entire MC volume compared to other popular deep learning networks. As a result, the Canal-Net achieved high consistent accuracy in 3D segmentations of the entire MC in spite of the areas of low visibility by the unclear and ambiguous cortical bone layer. Therefore, the Canal-Net demonstrated the automatic and robust 3D segmentation of the entire MC volume by improving structural continuity and boundary details of the MC in CBCT images.

The mandibular canal (MC) is an important mandibular structure that supplies sensation to the lower teeth, chin, and lower lip¹. Any injury to the MC can lead to temporary or permanent damage resulting in sensory disturbance sequelae such as paresthesia, hypoesthesia, and dysesthesia, which affects speech, mastication, and quality of life²⁻⁵. Therefore, knowing the exact localization of the MC is essential in planning appropriate oral-maxillofacial surgeries such as implant placement and third molar extractions^{6,7}. In preoperative assessments and surgical planning in dental clinics, panoramic radiographs are used as a standard dental imaging tool^{8,9}, but presents limitations in that it is challenging to determine the actual 3D rendering of the entire canal structure as the panoramic radiograph only shows the canal in a single view¹⁰. Therefore, additional investigations using CT may be recommended to verify the exact position of the canal in a 3D view in accordance with the as low as reasonably achievable (ALARA) principle⁸. Due to the advantages of CBCT such as a lower radiation dose, inexpensive image acquisition cost, and high spatial resolution, CBCT has been widely used in dental clinics for 3D diagnosis and treatment planning in the field of oral and maxillofacial surgery¹¹⁻¹³. However, the manual segmentation of the MC that is generally performed using 3D cross-sectional slices in CBCT images is time-consuming and labor-intensive^{10,14}. In addition, the ambiguous cortical bone layer surrounding the canal and the unclear medulla pattern also makes it difficult to distinguish the entire MC because of the lower contrast of

¹Interdisciplinary Program in Bioengineering, Graduate School of Engineering, Seoul National University, Seoul, Korea. ²Department of Applied Bioengineering, Graduate School of Convergence Science and Technology, Seoul National University, Seoul, Korea. ³Vision AI Business Team, LG CNS, Seoul, Korea. ⁴Department of Periodontology, School of Dentistry and Dental Research Institute, Seoul National University, Seoul, Korea. ⁵Department of Electronics and Information Engineering, Hansung University, Seoul, Korea. ⁶Department of Oral and Maxillofacial Radiology and Dental Research Institute, School of Dentistry, Seoul National University, Seoul, Korea. ⁷These authors contributed equally: Bo-Soung Jeoun and Su Yang. ✉email: wjyi@snu.ac.kr

CBCT images¹⁵. Therefore, automatic segmentation of the MC is required to alleviate the workload of dental clinicians by overcoming the limitations of CBCT images.

Among studies for automatic MC segmentation in CBCT images, atlas-based segmentation (ARS) and statistical shape model (SSM) methods have been proposed as two conventional representatives of MC segmentation methods^{16–18}. The SSM method utilized the prior knowledge of shape models to perform MC segmentation^{17,18}. This prior knowledge was required to reconstruct a 3D model of CBCT images, which highly affects the segmentation result^{17,18}. On the other hand, the ARS method only requires the atlas image for MC segmentation, which is independent of prior knowledge¹⁶. However, both the SSM and ARS methods exhibit a limitation in dealing with new forms of data beyond the predefined standard since they depend on prior knowledge or other preprocessing techniques^{16–18}. Recently, deep learning methods have been widely used for the detection^{19–21}, classification^{22–24}, segmentation^{25,26}, and enhancement^{27,28} of medical and dental images. Several convolutional neural networks (CNN) such as 3D U-Net, a type of deep learning method, were used for MC segmentation in CBCT images exhibiting a high accuracy of segmentation^{10,14}. However, these CNNs failed to segment the MC with high consistent accuracy throughout its entire range because of the occasional unclear and ambiguous cortical bone layer caused by the overall lower contrast of CBCT images^{10,14}. CNNs for the segmentation of the entire MC exhibited lower accuracy around the mandibular and mental foramina compared to other parts of the canal^{10,14} since discrimination of the canal from its surroundings became increasingly less clear towards the mental foramen region, and visibility of the MC clearly decreased on cross-sectional images of more distal regions of the MC¹⁵. Precise MC segmentation with high consistent accuracy throughout the entire MC is essential for avoiding nerve injury in oral and maxillofacial surgeries such as mandibular osteotomy and implant surgery²⁹.

We hypothesized that a deep learning model yielded more robust 3D segmentation of the entire MC volume in CBCT images by learning the spatio-temporal features and structural continuity of the MC volume. In this study, we proposed a continuity-aware contextual network (Canal-Net) for the automatic and robust 3D segmentation of the MC with high consistent accuracy throughout the entire MC volume in CBCT images and compared our network with other networks in terms of volumetric accuracy over the entire canal. Our main contributions were as follows: (1) We proposed a continuity-aware contextual network (Canal-Net) that was robust to ambiguous or unclear cortical bone regions of the MC and lower contrast of CBCT images in 3D segmentations of the entire MC. (2) We applied bidirectional convolutional LSTM (ConvLSTM) in order to learn 3D anatomical contextual information of the MC by incorporating spatio-temporal features. (3) We used a multi-task learning framework with multi-planar projection losses (MPL) in three anatomical planes in order to evaluate the global structural continuity of the MC.

Materials and methods

Data acquisition and preparation. We included 50 patients (27 females and 23 males; mean age 25.56 ± 6.73 years) who underwent dental implant surgeries or third molar extractions at the Seoul National University Dental Hospital (2019–2020). The patients had different mandibular canal shapes with various dental conditions including the metallic crowns and implants. The patient data were obtained at 80 kVp and 8 mA using CBCT (CS9300; Carestream Health, New York, USA). The CBCT images had dimensions of $841 \times 841 \times 289$ pixels, voxel sizes of $0.2 \times 0.2 \times 0.2$ mm³, and 16-bit depth. This study was performed with approval from the institutional review board of the Seoul National University Dental Hospital (ERI18001). The ethics committee approved the waiver for the informed consent because this was a retrospective study. The study was performed in accordance with the Declaration of Helsinki.

The mandibular canals including the surrounding cortical bone was manually annotated by an oral and maxillofacial radiologist using a software (3D Slicer for Windows 10, Version 4.10.2; MIT, Massachusetts, USA)³⁰. We used the cropped images consisting of 200 slices of 128×128 pixels that were centered at the left and right mandibular regions in order to reduce the memory requirement. Zero-padding was performed to maintain the input volume of the same length for all patients showing the mandibular canals of different lengths. For deep learning, we prepared 60 volumes from 30 patients for the training dataset, 20 from ten patients for the validation dataset, and 20 from ten patients for the test dataset where the right mandible images were horizontally flipped to match the left. We performed five-fold cross-validation, where each training cycle consisted of 60, 20, and 20 volumes for training, validation and test datasets, respectively.

We estimated the minimally required sample size to detect significant differences in the accuracy between the Canal-Net and the other networks, when both assessed the same subjects (CBCT images). We designed to capture a mean accuracy-difference of 0.05, and a standard deviation of 0.10 between the Canal-Net and the other networks. Based on an effect size of 0.5, a significance level of 0.05, and a statistical power of 0.80, we obtained a sample size of $N = 128$ (*G** Power for Windows 10, Version 3.1.9.7; Universität Düsseldorf, Germany). Eventually, we split the CBCT dataset of 2D images into 10,185, 2546, and 3183 for training, validation, and test datasets, respectively.

Continuity-aware contextual network (Canal-Net). We designed a continuity-aware contextual network (Canal-Net) which had 3D encoder-decoder architecture under a multi-task learning framework consisting of time-distributed convolution blocks, multi-scale inputs³¹, skip connections, and bidirectional convolutional LSTM (ConvLSTM) with side-output layers^{31,32} (Fig. 1). The bidirectional ConvLSTM was used to capture anatomical context information in concatenated feature maps extracted from the corresponding encoding path and the previous decoding up-sampling layer. A multi-task learning approach was adopted to simultaneously output the entire MC volume and its 2D multi-planar projections in three anatomical planes, which helped the network learn the overall MC volume and structural continuity (multi-planar projection outputs and the output

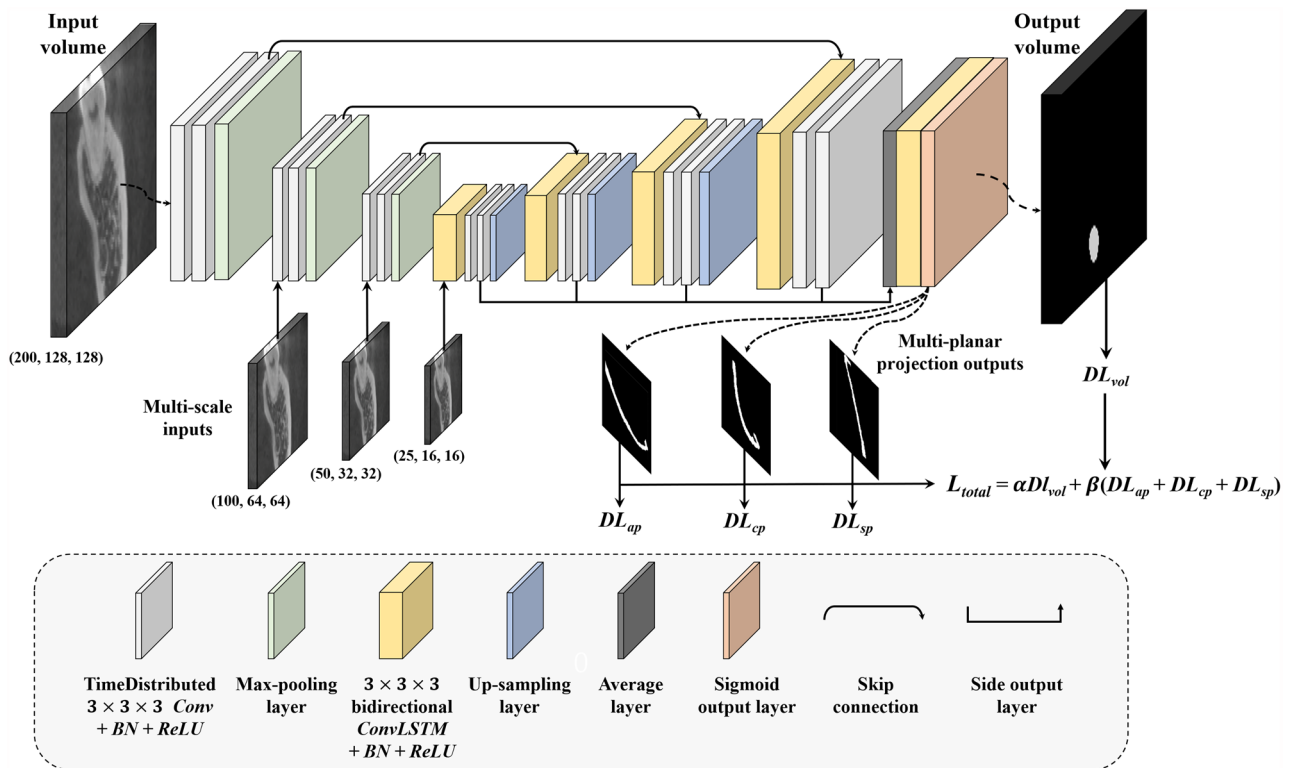


Figure 1. The Canal-Net architecture with a 3D encoder-decoder under a multi-task learning framework consisting of time-distributed convolution blocks, multi-scale inputs, skip connection, and bidirectional convolutional LSTM (ConvLSTM) with side-output layers. The bidirectional ConvLSTM was utilized to capture anatomical context information, and a multi-task learning approach was performed to learn overall MC volume and structural continuity.

volume in Fig. 1). The network under multi-task learning was optimized in an end-to-end manner, where the MC segmentation output was generated directly from the input volumes of the CBCT images.

At the encoder, the time-distributed convolution blocks processed sequential information from 3D volumetric inputs as series of features for 2D slices³³ (white blocks in Fig. 1). It was a typical convolution passed to a time-distributed wrapper that could be applied to every temporal frame of the input independently³³. Convolutional blocks were comprised of two repeated modules of two $3 \times 3 \times 3$ convolutions, batch normalization, ReLU, and $2 \times 2 \times 2$ max-pooling at the encoder path. The number of feature maps gradually decreased from 128 to 64, 32, and 16. To mitigate spatio-temporal information loss caused by max-pooling operations, the multi-scale inputs down-sampled from the original input volume by $2 \times 2 \times 2$ average pooling were concatenated at each level of the encoder (multi-scale inputs in Fig. 1).

At the decoder, the features from time-distributed convolutions at the encoder³³ were concatenated with the corresponding up-sampling layer and fed to bidirectional ConvLSTM blocks (Skip connection and yellow blocks in Fig. 1). Long short-term memory (LSTM), one of the recurrent neural networks (RNN)³⁴, was an efficient network for handling spatio-temporal data and was widely used in contextual processing such as natural language processing³⁵ and video segmentation³⁶. The internal matrix multiplication of the original LSTM was replaced by the convolution operation to maintain the input dimension in ConvLSTM³⁷. The ConvLSTM blocks were composed of two repeated modules of two $3 \times 3 \times 3$ bidirectional ConvLSTMs, batch normalization, ReLU, and $2 \times 2 \times 2$ up-sampling at the decoder path. The number of feature maps gradually increased from 16 to 32, 64, and 128. The ConvLSTM captured 3D local anatomical contextual information more effectively by learning the spatio-temporal features of the 3D volumetric data³⁷.

At the output layer, the averaged side-outputs generated from a local output map from every level of the decoder were merged and fed to the bidirectional ConvLSTM, which mitigated the gradient vanishing problem for encouraging the back-propagation of gradient flow (Side output and average layers in Fig. 1). The 3D volume loss and multi-planar projection losses (MPL) from the 2D projections simultaneously encouraged the network to learn the global structural continuity information of the canal under the multi-task learning framework. The MPL were calculated from the 2D projection maps of the output in three anatomical planes. The Dice similarity coefficient score (DSC) was used for the two loss functions³⁸. The loss function ($L = \alpha DL_{vol} + \beta (DL_{ap} + DL_{cp} + DL_{sp})$) of the Canal-Net consisted of 3D volume loss (DL_{vol}) for the entire canal volume, and the MPL as sum of the 2D projection losses in axial- (DL_{ap}), coronal- (DL_{cp}), and sagittal- (DL_{sp}) planes, where α and β were constant weights for the 3D volume loss and the summation of the 2D projection map losses, respectively (equation of L_{total} in Fig. 1). The weights of α and β were optimized for the best

Loss weight	3D volume	2D axial	2D coronal	2D sagittal
$\alpha=0.1, \beta=0.9$	0.84 ± 0.14	0.92 ± 0.07	0.91 ± 0.17	0.91 ± 0.14
$\alpha=0.2, \beta=0.8$	0.85 ± 0.12	0.91 ± 0.10	0.90 ± 0.12	0.89 ± 0.13
$\alpha=0.3, \beta=0.7$	0.84 ± 0.16	0.90 ± 0.14	0.87 ± 0.17	0.88 ± 0.17
$\alpha=0.4, \beta=0.6$	0.83 ± 0.16	0.90 ± 0.15	0.83 ± 0.19	0.88 ± 0.16
$\alpha=0.5, \beta=0.5$	0.86 ± 0.13	0.91 ± 0.10	0.88 ± 0.16	0.91 ± 0.11
$\alpha=0.6, \beta=0.4$	0.84 ± 0.14	0.89 ± 0.13	0.86 ± 0.17	0.89 ± 0.13
$\alpha=0.7, \beta=0.3$	0.87 ± 0.05	0.93 ± 0.07	0.91 ± 0.14	0.94 ± 0.08
$\alpha=0.8, \beta=0.2$	0.87 ± 0.09	0.92 ± 0.09	0.91 ± 0.14	0.92 ± 0.09
$\alpha=0.9, \beta=0.1$	0.86 ± 0.13	0.91 ± 0.10	0.87 ± 0.16	0.92 ± 0.10

Table 1. Comparison of Dice similarity coefficient scores (DSC) for α , weight of 3D volume loss, and β , weight of the MPL in axial, coronal, and sagittal planes to determine the optimal loss weights for α and β for the Canal-Net.

performance through an ablation study. The weights of 0.7 and 0.3 for the 3D volume loss and MPL, respectively, exhibited the best performance compared to other weight options (Table 1).

The proposed networks were trained using an Adam optimizer, and the learning rate of 0.00025 was reduced on plateau by a factor of 0.5 every 25 epochs in 300 epochs with the batch size of 1. They were implemented with Python3 based on Keras with a Tensorflow backend using a single NVIDIA Titan RTX GPU 24G.

Performance evaluation of Canal-Net for MC segmentation. We compared the performance of the MC segmentation by Canal-Net with those by other networks of 2D U-Net³⁹, SegNet⁴⁰, 3D U-Net⁴¹, 3D U-Net with MPL (MPL 3D U-Net), and 3D U-Net with ConvLSTM (ConvLSTM 3D U-Net). To evaluate the performances quantitatively, we compared the 2D segmentation performance metrics of the Dice similarity coefficient score ($DSC = \frac{2TP}{2TP+FN+FP}$), Jaccard index ($JI = \frac{TP}{TP+FN+FP}$), precision ($PR = \frac{TP}{TP+FP}$), recall ($RC = \frac{TP}{TP+FN}$) among networks, where TP, FP, and FN denoted true positives, false positives, and false negatives, and also 3D volumetric performance metrics of volume of error ($VOE = 1 - \frac{V_{gt} \cap V_{pred}}{V_{gt} \cup V_{pred}}$) and relative volume difference ($RVD = \frac{|V_{gt} - V_{pred}|}{V_{gt}}$), where V_{gt} and V_{pred} represented the number of voxels for the ground truth and for the predicted volume, respectively. We also evaluated the mean curve distance ($MCD = \frac{\sum_{t \in C(V_{gt})} dist(t, C(V_{pred}))}{|C(V_{gt})|}$), where $dist(x, Y) = \min_{y \in Y} \{|x - y|^2\}$, and t denotes coordinates of a ground truth voxel¹⁴, and $C(\cdot)$ is an operation which extracted the center curve line through skeletonization for the set of voxels¹⁴. The higher values of DSC, JI, PR, and RC, and the lower values of VOE, RVD, and MCD indicated better segmentation performance. We used paired two-tailed t-tests to compare performances between Canal-Net and others (SPSS Statistics for Windows 10, Version 26.0; IBM, Armonk, New York, USA). The statistical significance level was set at 0.05. We also performed the Bland–Altman analysis to analyze the bias and agreement limits of the used segmentation models between the number of pixels of ground truth and prediction results.

Results

The performances of Canal-Net, convLSTM 3D U-Net, MPL 3D U-Net, 3D U-Net, SegNet, and 2D U-Net were evaluated for a total of 20 mandibular canals not used for training. Among them, convLSTM 3D U-Net, MPL 3D U-Net, and 3D U-Net were evaluated to demonstrate the effectiveness of the corresponding components in Canal-Net, while the other networks were used for performance comparisons between 2 and 3D CNN-based approaches. In addition, the Canal-Net was evaluated for the impacts of the weights of α and β on 3D volume loss and MPL, respectively. The Canal-Net with loss weights of $\alpha=0.7$ and $\beta=0.3$ achieved the best segmentation performance of 0.87, 0.93, 0.91, and 0.94 DSC for 3D volume, axial, coronal, and sagittal planes, respectively (Table 1).

Table 2 shows the quantitative results of the segmentation performance by the networks. The performances of Canal-Net, ConvLSTM 3D U-Net, MPL 3D U-Net, 3D U-Net, SegNet, and 2D U-Net were compared using 20 total mandibular canals. The Canal-Net achieved the highest values of 0.87 DSC ($p < 0.05$), 0.80 JI ($p < 0.05$), 0.89 PR ($p = 0.05$), and 0.88 RC ($p = 0.05$) in 2D performance metrics, and also the lowest values of 0.14 RVD ($p < 0.05$), 0.10 VOE ($p < 0.05$), and 0.62 MCD ($p < 0.05$) in 3D performance metrics (Table 2). The Canal-Net outperformed all the other networks in DSC, JI, PR, RC, RVD, and VOE, and significantly so in MCD ($p < 0.05$) (Table 2). The performance of the networks is also plotted in boxplots (Fig. 2). The Canal-Net achieves the higher performances than the other networks with a smaller dispersion of data, shorter length of whiskers, and rare existence of outliers (Fig. 2).

In Fig. 3, the Canal-Net exhibited more accurate predictions with more true positives (yellow) and less false positives (red) and false negatives (green) compared to the other networks for MCs with unclear and ambiguous cortical bone layers and metallic objects in CBCT images of lower contrast (Fig. 3a–e). In the 3D segmentation results, the Canal-Net also demonstrated better prediction results with less false positives and false negatives compared to the other networks in the mental foramen area of the various MC shapes (Fig. 4a–e). Furthermore, only a few cases as outliers for the results of the Canal-Net were observed due to other causes such as the presence of a third molar beside the MC (Figs. 3f, 4f). The Canal-Net predicted more accurately the entire MC volume,

	DSC	JI	PR	RC	RVD	VOE	MCD (mm)
Canal-Net	0.87 ± 0.05	0.80 ± 0.06	0.89 ± 0.06	0.88 ± 0.06	0.14 ± 0.04	0.10 ± 0.04	0.62 ± 0.10
ConvLSTM 3D U-Net	0.85 ± 0.08*	0.77 ± 0.08*	0.87 ± 0.08*	0.86 ± 0.09*	0.17 ± 0.05*	0.13 ± 0.05*	0.66 ± 0.12*
MPL 3D U-Net	0.84 ± 0.06 [†]	0.75 ± 0.07 [†]	0.88 ± 0.07 [†]	0.82 ± 0.08 [†]	0.19 ± 0.04 [†]	0.14 ± 0.06 [†]	0.69 ± 0.15 [†]
3D U-Net	0.83 ± 0.07 [‡]	0.74 ± 0.07 [‡]	0.85 ± 0.08 [‡]	0.84 ± 0.09 [‡]	0.19 ± 0.05 [‡]	0.15 ± 0.07 [‡]	0.69 ± 0.13 [‡]
SegNet	0.84 ± 0.06 ⁺	0.77 ± 0.06 ⁺	0.85 ± 0.06 ⁺	0.85 ± 0.07 ⁺	0.18 ± 0.04 ⁺	0.14 ± 0.05 ⁺	0.78 ± 0.19 ⁺
2D U-Net	0.84 ± 0.07 [Ⓟ]	0.77 ± 0.07 [Ⓟ]	0.85 ± 0.07 [Ⓟ]	0.84 ± 0.08 [Ⓟ]	0.18 ± 0.04 [Ⓟ]	0.14 ± 0.05 [Ⓟ]	0.87 ± 0.22 [Ⓟ]

Table 2. Mean (SD) dice similarity coefficient score (DSC), Jaccard index (JI), precision (PR), recall (RC), volume of error (VOE), relative volume difference (RVD), and mean curve distance (MCD) by Canal-Net (ours), ConvLSTM 3D U-Net (ours), MPL 3D U-Net (ours), 3D U-Net, SegNet, and 2D U-Net Net by five-fold cross-validation. *Significant difference between Canal-Net and ConvLSTM 3D U-Net ($p < 0.05$). [†]Between Canal-Net and MPL 3D U-Net ($p < 0.05$). [‡]Between Canal-Net and 3D U-Net ($p < 0.05$). ⁺Between Canal-Net and SegNet ($p < 0.05$). [Ⓟ]Between Canal-Net and 2D U-Net ($p < 0.05$).

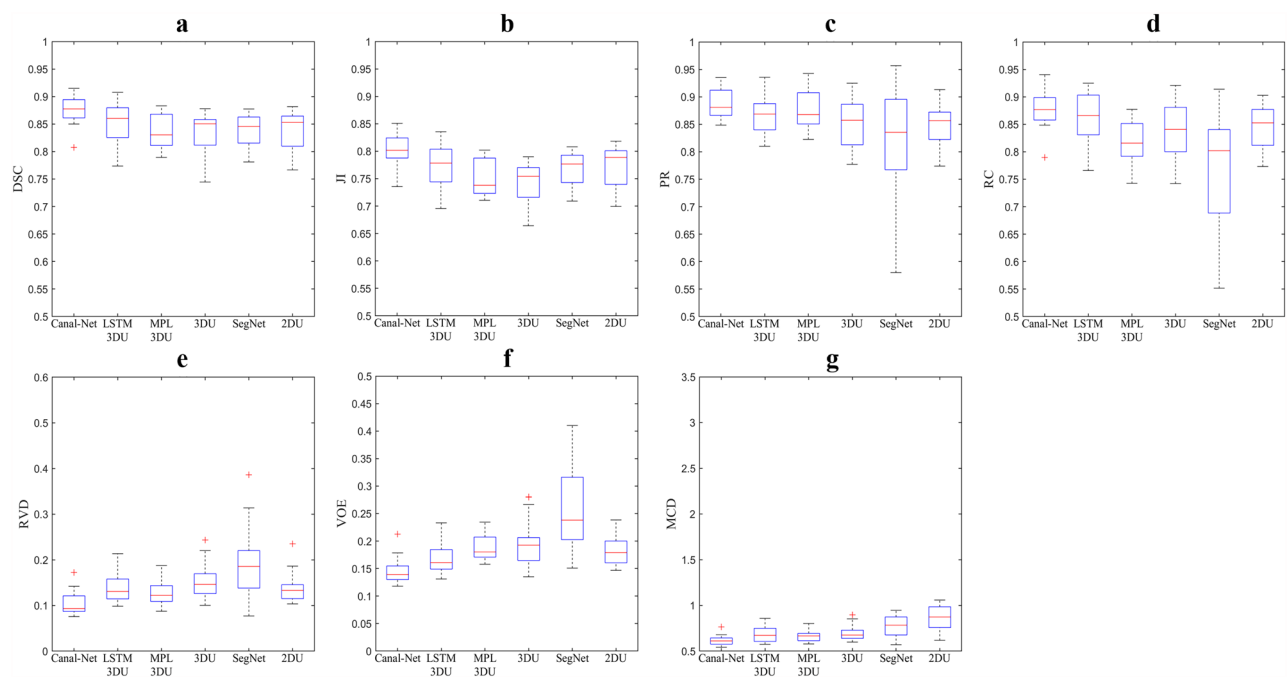


Figure 2. The boxplots of segmentation performance results of the (a) Dice similarity coefficient score (DSC), (b) Jaccard index (JI), (c) precision (PR), (d) recall (RC), (e) relative volume difference (RVD), (f) volume of error (VOE), and (g) mean curve distance (MCD) for the deep learning networks, Canal-Net, ConvLSTM 3D U-Net (LSTM 3DU), MPL 3D U-Net (MPL 3DU), 3D U-Net (3DU), SegNet, and 2D U-Net (2DU). Each box contains the first and third quartile of data. The medians are located inside of the boxes, visualized as red lines. The whiskers are extended above and below each box in ± 1.5 times the interquartile range (IQR), and the outliers are visualized as red + marks defining values 1.5 IQR away from the box.

and demonstrated improved structural continuity and boundary details of the MC from the mental foramen to the mandibular foramen compared to the other networks (Fig. 4a–e).

The DSC and MCD for the whole test dataset were plotted from the mental foramen to the mandibular foramen, and the 3D networks generally exhibited less variations of the performances compared to the 2D networks (Figs. 5 and 6). The Canal-Net demonstrated the most consistent performances with the smallest fluctuations of true segmentation compared to the other networks throughout the entire MC volume (Figs. 5 and 6). As a result, the Canal-Net represented the best 3D segmentation accuracies of RVD, VOE, and MCD throughout the entire MC volume among the networks (Table 2). The Bland–Altman plot between the ground truth and prediction results from the Canal-Net showed higher linear relationships and better agreement limits than those from the other networks (Fig. 7). Therefore, the Canal-Net represented more accurate and robust MC segmentation performance of the entire MC compared to the other networks.

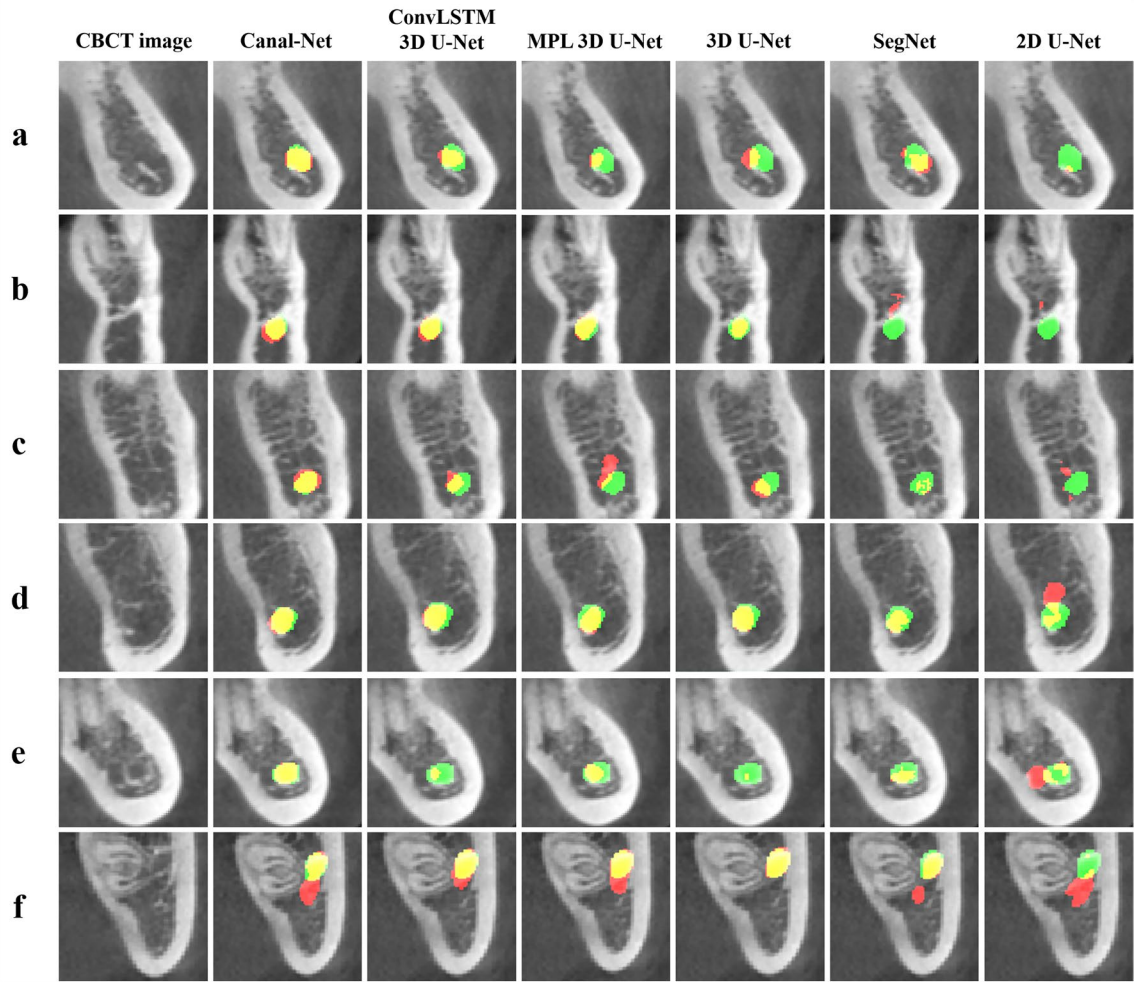


Figure 3. The segmentation results by Canal-Net (ours), ConvLSTM 3D U-Net (ours), MPL 3D U-Net (ours), 3D U-Net, SegNet, and 2D U-Net. The yellow, green, and red areas indicate the true positive, false negative, and false positive, respectively. The segmentation results for the MC with (a–c) low visibility and (d,e) containing the metallic crown and implant fixture, and for (f) the outlier of the Canal-Net in the boxplot in Fig. 2.

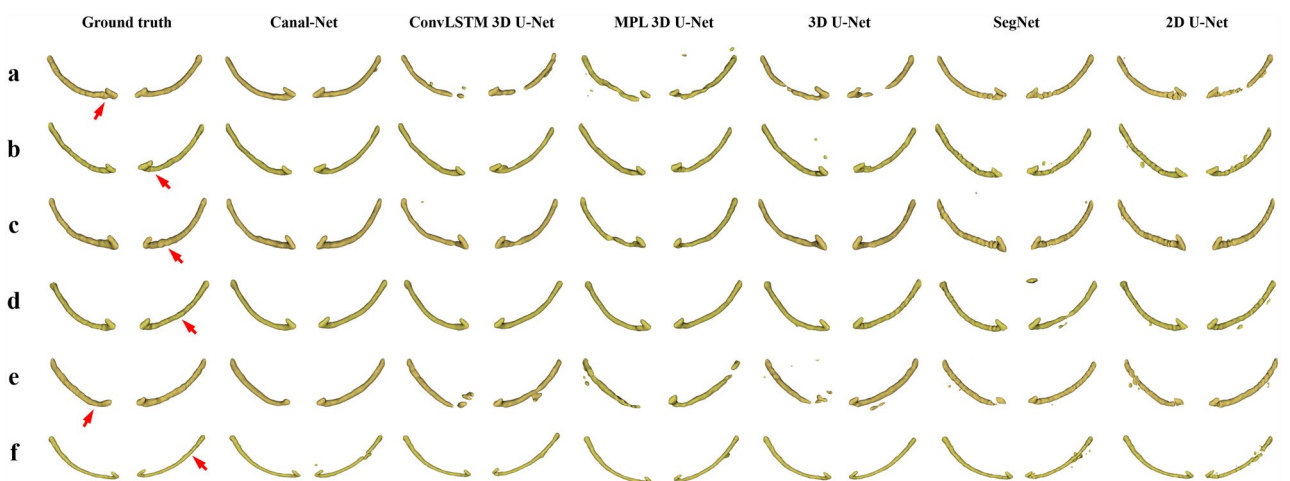


Figure 4. The 3D reconstructed whole canal from the ground truth and segmentation results by Canal-Net (ours), ConvLSTM 3D U-Net (ours), MPL 3D U-Net (ours), 3D U-Net, SegNet, and 2D U-Net from the top to bottom (a–f) from the same canals shown in Fig. 3a–f. The red arrows indicate the position of the corresponding image slice in Fig. 3.

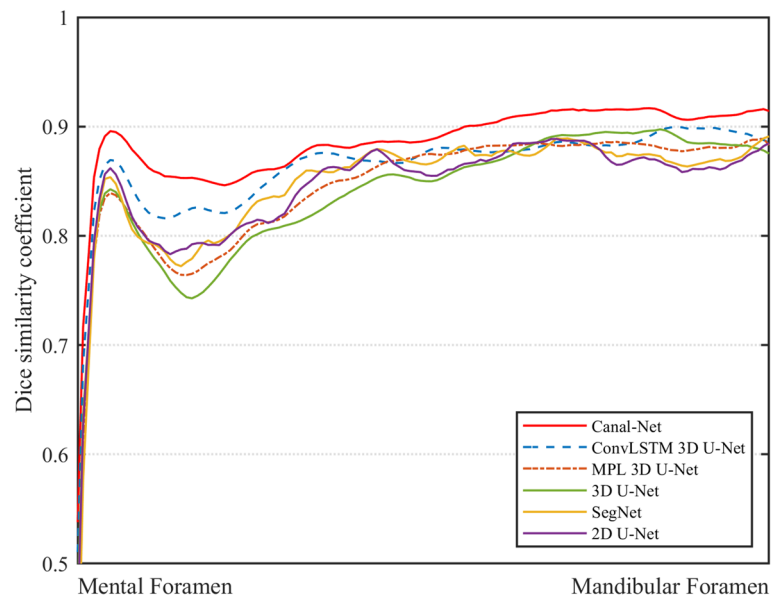


Figure 5. The line plots of dice similarity coefficient score (DSC) from the mental foramen to the mandibular foramen for Canal-Net (ours), ConvLSTM 3D U-Net (ours), MPL 3D U-Net (ours), 3D U-Net, SegNet, and 2D U-Net.

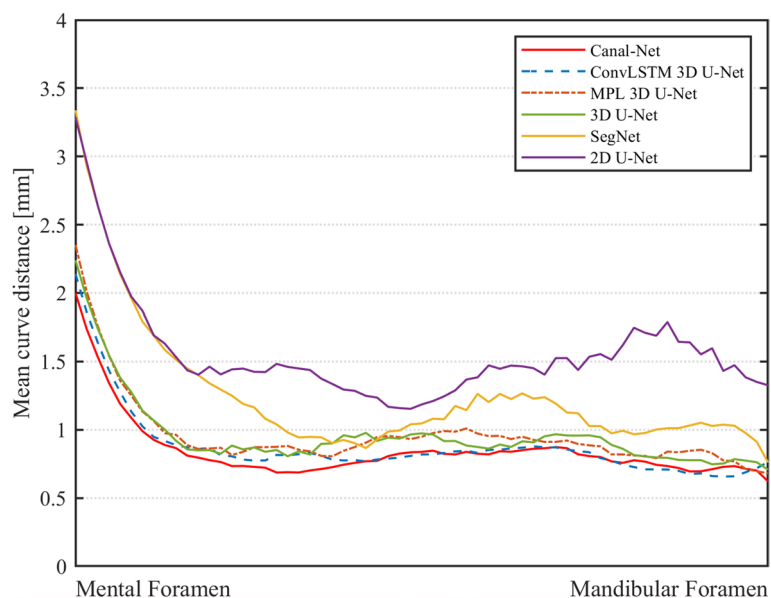


Figure 6. The line plots of mean curve distance (MCD) from the mental foramen to the mandibular foramen for Canal-Net (ours), ConvLSTM 3D U-Net (ours), MPL 3D U-Net (ours), 3D U-Net, SegNet, and 2D U-Net.

Discussion

In this study, we proposed a continuity-aware contextual network (Canal-Net) which learned 3D local anatomical contextual information and the global continuity of the MC complementally in order to segment the MC with high consistent accuracy throughout the entire MC volume in cone-beam CT (CBCT) images. We employed time-distributed convolution layers for handling time-distributed sequential features with multi-scale inputs at the encoder path³³, and bidirectional ConvLSTM layers for extracting spatio-temporal features at the decoder path³⁷. The Canal-Net was able to learn the local anatomical variations of the MC by incorporating the spatio-temporal features effectively, and the global structural continuity information of the MC under the multi-task learning framework, complementally. The Canal-Net used optimized weights for 3D volume loss and multi-planar projection losses in multi-task learning. Therefore, the Canal-Net improved the performance of automatic segmentation of the MC by combining anatomical context information and global structural continuity information, resulting in higher consistent accuracy throughout the entire MC volume in CBCT images.

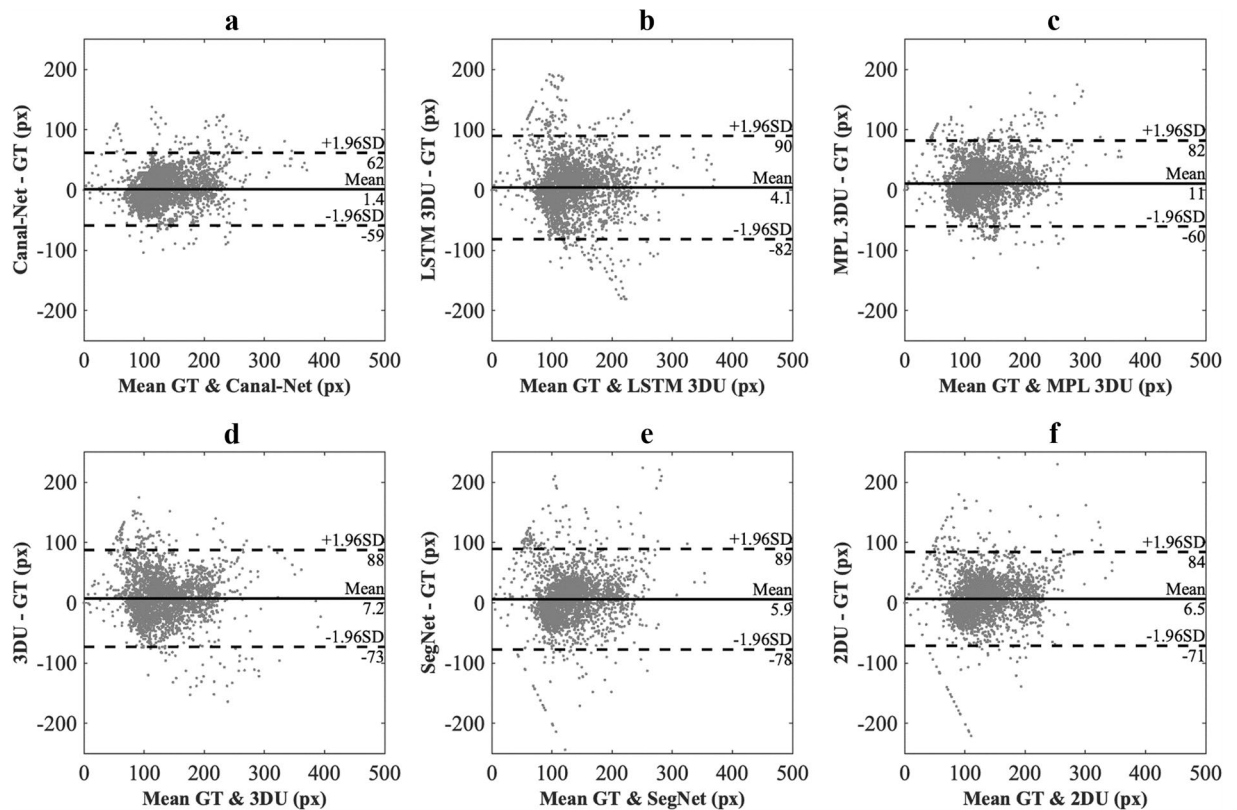


Figure 7. The Bland–Altman plots between the ground (GT) and prediction results from the (a) Canal-Net, (b) convLSTM 3D U-Net (LSTM 3DU), (c) MPL 3D U-Net (MPL 3DU), (d) 3D U-Net (3DU), (e) SegNet, and (f) 2D U-Net (2DU).

We compared the Canal-Net with other popular segmentation networks such as 2D U-Net, SegNet, and 3D U-Net, and also with our MPL 3D U-Net and ConvLSTM 3D U-Net for MC segmentation. In performances of MC segmentation in CBCT images, 2D U-Net and SegNet exhibited lower accuracies compared to the 3D networks, generally. False negatives and positives were observed at a higher rate around the mental foramen area with ambiguous or unclear cortical bone layers. Since the 2D networks were not able to learn the 3D contextual features of the MC volume in CBCT images, the 2D networks exhibited coarser 3D segmentation volumes with more fluctuations of 3D performance accuracy from the mental to the mandibular foramen regions. In terms of learning 3D spatial contextual information between image slices of the 3D anatomical structures, 3D U-Net was generally expected to generate more accurate segmentation results compared to 2D networks⁴¹. In the present study, the 3D U-Net predicted the more accurate segmentation of the MC with fewer false positives and negatives compared to the 2D U-Net and SegNet. However, the 3D U-Net had still limitations in segmenting the MC regions with unclear cortical bone layers accurately by only learning 3D spatial information between image slices, and exhibited inaccurate segmentation results with disconnections around the mental foramen area.

Both MPL 3D U-Net and ConvLSTM 3D U-Net demonstrated better segmentation results than 3D U-Net in different aspects. The MPL 3D U-Net showed an improved travel course of the MC compared to 3D U-Net because its spatial information was complemented with the global structural continuity information by learning through multi-planar projections. Although the structural continuity of the MC volume was improved by multi-task learning, the MPL 3D U-Net exhibited difficulties in producing segmentation boundaries in detail around the mental foramen area. On the other hand, the ConvLSTM learned anatomical context information through spatio-temporal features, and the MC volume showed smooth boundaries with more consistent accuracies even in unclear cortical bone layer regions in the CBCT images. Therefore, the Canal-Net demonstrated the most accurate segmentation of the entire MC volume compared to the other networks by simultaneously learning global structural continuity through MPL, and anatomical context information through ConvLSTM. Compared with previous studies using 3D U-Net^{10,14}, our Canal-Net achieved 0.87 of DSC and 0.80 of the mean intersection of union (IoU) while two previous studies reported 0.58 of DSC^{10,14} and 0.58 of mean IoU^{10,14}. Compared with the previous studies^{10,14}, the Canal-Net showed substantially enhanced performance of the MC segmentation in CBCT images.

In the Canal-Net, the MPL provided global structural continuity from three anatomical projection maps with ConvLSTM anatomical context information by spatio-temporal features, complementally. In the MC areas of low visibility with ambiguous or unclear cortical bone layers in CBCT images, the Canal-Net exhibited the best outcomes with continuous and consistent MC volumes from the mental to mandibular foramens. The Canal-Net especially surpassed other networks by showing continuous MC volumes around the mental foramen area where

the visibility of the MC tended to diminish¹⁵, and in areas affected by metallic objects such as implant fixtures or dental crowns in CBCT images. As a result, the Canal-Net demonstrated the most robust MC segmentation with high consistent DSC throughout the entire MC volume in CBCT images.

The primary reason for improved segmentation performance by Canal-Net was that its network architecture was constructed to complementally learn the 3D anatomical context information of the MC by the spatio-temporal features from the bidirectional ConvLSTM layers and the global structural continuity information by MPL. In the Canal-Net, the complementary context information was successfully learned in the proposed framework, leading to maintaining continuous and consistent MC volumes from the mental to the mandibular foramen areas. The proposed learning process has several advantages. First, it could increase the discriminative capability of intermediate feature representations with multiple regularizations on disentangling subtly correlated tasks⁴⁸, potentially improving the robustness of the segmentation performance. Second, in the application of MC segmentation, the multi-task learning framework could also provide complementary context information that would serve well to segment the MC maintaining overall continuous and consistent volumes. This could improve the performance accuracy of MC segmentations substantially, especially in MC regions with ambiguous or unclear cortical bone layers in lower contrast CBCT images.

The accurate identification of the whole MC structure in the mandible is an essential prerequisite for the preoperative planning of third molar extractions and implant surgeries to avoid any surgical complications⁷. However, the exact recognition of the entire canal structure is considered to be a challenging and delicate task for several reasons¹⁵. CBCT, the most commonly used 3D dental imaging tool, has lower contrast than CT, which negatively affects the ability to distinguish MCs^{10,42}. As a result, the low visibility of MCs, such as in ambiguous or unclear cortical bone regions, affects the structural continuity of MC segmentation in CBCT images^{10,14}. Furthermore, the visibility of the MC itself is low due to variable cortications and bone densities of the canal wall, the diverse travel courses of the canal, and the spread of vessels and nerve branches^{15,43–47}. The Canal-Net could be used in automatic and robust 3D segmentation of the MC structure for the preoperative planning of third molar extractions and implant surgeries to avoid any surgical complications when using CBCT images. The automatic segmentation of the MC volume by the Canal-Net could provide clinicians with accurate identification of the MC structure in the mandible with high consistent accuracy throughout the entire MC volume ranging from the mental foramen to the mandibular foramen while reducing time and effort.

However, our study had several limitations. First, as there was the problem of reducing the memory requirements for dealing with large amounts of data when using deep 3D networks running on the GPU, it was necessary to optimize the way the memory was used in order to maximize GPU utilization. Therefore, we used the cropped images with smaller dimensions than the original, and preprocessing of the images required additional time and labor. Second, our study had a potential limitation of generalization ability due to using internal data from a single organization. Overfitting of training a deep learning model, which resulted in the model learning statistical regularity specific to the training dataset, could negatively impact the model's ability to generalize to a new dataset⁴⁹. Although the proposed network did not show the presence of overfitting for the internal dataset in the five-fold cross-validation, it needs to be trained and evaluated using large datasets from multiple organizations or devices for generalization. Third, the results presented in this study were based on datasets from 50 patients. The proposed method needs to be evaluated for datasets from more patients with various dental restorations and implants. In future studies, we will improve the generalization ability and clinical efficacy of the Canal-Net by using large CBCT datasets acquired under various imaging conditions from multiple organizations or devices.

Conclusions

In this study, we proposed a continuity-aware contextual network (Canal-Net) that was robust to ambiguous or unclear cortical bone regions of the MC and lower contrast of CBCT images in 3D segmentations of the entire MC. The Canal-Net was designed based on a 3D U-Net with the ConvLSTM under the multi-task learning framework using MPL in order to complementally learn anatomical contexts and global structural continuity information. As a result, the Canal-Net achieved substantially enhanced performances compared to other networks such as 2D U-Net, SegNet, 3D U-Net, MPL 3D U-Net, and ConvLSTM 3D U-Net in 2D and 3D performances. Furthermore, Canal-Net demonstrated automatic and robust 3D segmentation of the entire MC volume by improving structural continuity and boundary details of the MC in CBCT images. The Canal-Net could be contributed to accurate and automatic identification of the MC structure for the preoperative planning of third molar extractions and implant surgeries to avoid any surgical complications.

Data availability

The datasets generated and/or analyzed during the current study are not publicly available due to the restriction by the Institutional Review Board (IRB) of Seoul National University Dental Hospital in order to protect patients' privacy but are available from the corresponding author on reasonable request. Please contact the corresponding author for any commercial implementation of our research.

Received: 8 April 2022; Accepted: 25 July 2022

Published online: 05 August 2022

References

1. Ghatak, R. N., Helwany, M. & Ginglen, J. G. *Anatomy, Head and Neck, Mandibular Nerve* (StatPearls, 2020).
2. Shavit, I. & Juodzbalys, G. Inferior alveolar nerve injuries following implant placement—Importance of early diagnosis and treatment: A systematic review. *J. Oral Maxillofac. Res.* 5, e2. <https://doi.org/10.5037/jomr.2014.5402> (2014).
3. Sarikov, R. & Juodzbalys, G. Inferior alveolar nerve injury after mandibular third molar extraction: A literature review. *J. Oral Maxillofac. Res.* 5, e1. <https://doi.org/10.5037/jomr.2014.5401> (2014).

4. Phillips, C. & Essick, G. Inferior alveolar nerve injury following orthognathic surgery: A review of assessment issues. *J. Oral Rehabil.* **38**, 547–554. <https://doi.org/10.1111/j.1365-2842.2010.02176.x> (2011).
5. Loescher, A. R., Smith, K. G. & Robinson, P. P. Nerve damage and third molar removal. *Dent. Update* **30**, 375–380. <https://doi.org/10.12968/denu.2003.30.7.375> (2003) (382).
6. Ai, C. J., Jabar, N. A., Lan, T. H. & Ramli, R. Mandibular canal enlargement: Clinical and radiological characteristics. *J. Clin. Imaging Sci.* **7**, 28. https://doi.org/10.4103/jcis.JCIS_28_17 (2017).
7. Jung, Y. H. & Cho, B. H. Radiographic evaluation of the course and visibility of the mandibular canal. *Imaging Sci. Dent.* **44**, 273–278. <https://doi.org/10.5624/isd.2014.44.4.273> (2014).
8. Ghaemini, H. *et al.* Position of the impacted third molar in relation to the mandibular canal. Diagnostic accuracy of cone beam computed tomography compared with panoramic radiography. *Int. J. Oral Maxillofac. Surg.* **38**, 964–971. <https://doi.org/10.1016/j.ijom.2009.06.007> (2009).
9. Vinayahalingam, S., Xi, T., Berge, S., Maal, T. & de Jong, G. Automated detection of third molars and mandibular nerve by deep learning. *Sci. Rep.* **9**, 9007. <https://doi.org/10.1038/s41598-019-45487-3> (2019).
10. Kwak, G. H. *et al.* Automatic mandibular canal detection using a deep convolutional neural network. *Sci. Rep.* **10**, 5711. <https://doi.org/10.1038/s41598-020-62586-8> (2020).
11. Ludlow, J. B., Davies-Ludlow, L., Brooks, S. & Howerton, W. Dosimetry of 3 CBCT devices for oral and maxillofacial radiology: CB Mercuray, NewTom 3G and i-CAT. *Dentomaxillofac. Radiol.* **35**, 219–226. <https://doi.org/10.1259/dmfr/14340323> (2006).
12. Arai, Y., Tammisalo, E., Iwai, K., Hashimoto, K. & Shinoda, K. Development of a compact computed tomographic apparatus for dental use. *Dentomaxillofac. Radiol.* **28**, 245–248. <https://doi.org/10.1038/sj/dmfr/4600448> (1999).
13. Pauwels, R. *et al.* Variability of dental cone beam CT grey values for density estimations. *Br. J. Radiol.* **86**, 20120135. <https://doi.org/10.1259/bjr.20120135> (2013).
14. Jaskari, J. *et al.* Deep learning method for mandibular canal segmentation in dental cone beam computed tomography volumes. *Sci. Rep.* **10**, 5842. <https://doi.org/10.1038/s41598-020-62321-3> (2020).
15. Oliveira-Santos, C. *et al.* Visibility of the mandibular canal on CBCT cross-sectional images. *J. Appl. Oral Sci.* **19**, 240–243. <https://doi.org/10.1590/s1678-77572011000300011> (2011).
16. Kroon, D.-J. Segmentation of the mandibular canal in cone-beam CT data. *Univ. Twente*. <https://doi.org/10.3990/1.9789036532808> (2011).
17. Abdolali, F. *et al.* Automatic segmentation of mandibular canal in cone beam CT images using conditional statistical shape model and fast marching. *Int. J. Comput. Assist. Radiol. Surg.* **12**, 581–593. <https://doi.org/10.1007/s11548-016-1484-2> (2017).
18. Kainmueller, D., Lamecker, H., Seim, H., Zinser, M. & Zachow, S. Automatic extraction of mandibular nerve and bone from cone-beam CT data. in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 76–83. (Springer, 2009).
19. Ahn, J. M. *et al.* A deep learning model for the detection of both advanced and early glaucoma using fundus photography. *PLoS ONE* **13**, e0207982. <https://doi.org/10.1371/journal.pone.0211579> (2018).
20. Phan, S., Satoh, S. I., Yoda, Y., Kashiwagi, K. & Oshika, T. Evaluation of deep convolutional neural networks for glaucoma detection. *Jpn. J. Ophthalmol.* **63**, 276–283. <https://doi.org/10.1007/s10384-019-00659-6> (2019).
21. Chang, H.-J. *et al.* Deep learning hybrid method to automatically diagnose periodontal bone loss and stage periodontitis. *Sci. Rep.* **10**, 7531. <https://doi.org/10.1038/s41598-020-64509-z> (2020).
22. Shen, W. *et al.* Multi-crop convolutional neural networks for lung nodule malignancy suspiciousness classification. *Pattern Recognit.* **61**, 663–673. <https://doi.org/10.1016/j.patcog.2016.05.029> (2017).
23. Kumar, A., Kim, J., Lyndon, D., Fulham, M. & Feng, D. An ensemble of fine-tuned convolutional neural networks for medical image classification. *IEEE J. Biomed. Health Inform.* **21**, 31–40. <https://doi.org/10.1109/JBHI.2016.2635663> (2016).
24. Yu, Y. *et al.* Deep transfer learning for modality classification of medical images. *Information* **8**, 91. <https://doi.org/10.3390/info8030091> (2017).
25. Cheng, J. Z. *et al.* Computer-aided diagnosis with deep learning architecture: Applications to breast lesions in US images and pulmonary nodules in CT scans. *Sci. Rep.* **6**, 24454. <https://doi.org/10.1038/srep24454> (2016).
26. Christ, P. F. *et al.* Automatic liver and tumor segmentation of CT and MRI volumes using cascaded fully convolutional neural networks. *arXiv preprint: arXiv:1702.05970*. <https://doi.org/10.48550/arXiv.1702.0597> (2017).
27. Yong, T.-H. *et al.* QCBCT-NET for direct measurement of bone mineral density from quantitative cone-beam CT: A human skull phantom study. *Sci. Rep.* **11**, 1–13. <https://doi.org/10.1038/s41598-021-94359-2> (2021).
28. Heo, M.-S. *et al.* Artificial intelligence in oral and maxillofacial radiology: What is currently possible?. *Dentomaxillofac. Radiol.* **50**, 20200375. <https://doi.org/10.1259/dmfr.20200375> (2021).
29. Greenstein, G. & Tarnow, D. The mental foramen and nerve: Clinical and anatomical factors related to dental implant placement: A literature review. *J. Periodontol.* **77**, 1933–1943. <https://doi.org/10.1902/jop.2006.060197> (2006).
30. Fedorov, A. *et al.* 3D Slicer as an image computing platform for the quantitative imaging network. *Magn. Reason. Imaging* **30**, 1323–1341. <https://doi.org/10.1016/j.mri.2012.05.001> (2012).
31. Fu, H. *et al.* Joint optic disc and cup segmentation based on multi-label deep network and polar transformation. *IEEE Trans. Med. Imaging* **37**, 1597–1605. <https://doi.org/10.1109/TMI.2018.2791488> (2018).
32. Yin, P., Yuan, R., Cheng, Y. & Wu, Q. Deep guidance network for biomedical image segmentation. *IEEE Access* **8**, 116106–116116. <https://doi.org/10.1109/ACCESS.2020.3002835> (2020).
33. Novikov, A. A., Major, D., Wimmer, M., Lenis, D. & Buhler, K. Deep sequential segmentation of organs in volumetric medical scans. *IEEE Trans. Med. Imaging* **38**, 1207–1215. <https://doi.org/10.1109/TMI.2018.2881678> (2019).
34. Yu, Y., Si, X., Hu, C. & Zhang, J. A review of recurrent neural networks: LSTM cells and network architectures. *Neural Comput.* **31**, 1235–1270. https://doi.org/10.1162/neco_a_01199 (2019).
35. Sutskever, I., Vinyals, O. & Le, Q. V. Sequence to sequence learning with neural networks. *arXiv preprint arXiv:1409.3215*. <https://doi.org/10.48550/arXiv.1409.3215> (2014).
36. Ventura, C. *et al.* Rvos: End-to-end recurrent network for video object segmentation. in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 5277–5286. (IEEE, 2019).
37. Shi, X. *et al.* Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *arXiv preprint arXiv:1506.04214*. <https://doi.org/10.48550/arXiv.1506.04214> (2015).
38. Sudre, C. H., Li, W., Vercauteren, T., Ourselin, S. & Cardoso, M. J. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. 240–248. (Springer, 2017).
39. Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 234–241. (Springer, 2015).
40. Badrinarayanan, V., Kendall, A. & Cipolla, R. SegNet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**, 2481–2495. <https://doi.org/10.1109/TPAMI.2016.2644615> (2017).
41. Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T. & Ronneberger, O. 3D U-Net: Learning dense volumetric segmentation from sparse annotation. in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 424–432. (Springer, 2016).

42. Moris, B., Claesen, L., Sun, Y. & Politis, C. Automated tracking of the mandibular canal in cbct images using matching and multiple hypotheses methods. in *2012 Fourth International Conference on Communications and Electronics*. 327–332. (IEEE, 2012).
43. Denio, D., Torabinejad, M. & Bakland, L. K. Anatomical relationship of the mandibular canal to its surrounding structures in mature mandibles. *J. Endod.* **18**, 161–165. [https://doi.org/10.1016/S0099-2399\(06\)81411-1](https://doi.org/10.1016/S0099-2399(06)81411-1) (1992).
44. Gowgiel, J. M. The position and course of the mandibular canal. *J. Oral Implantol.* **18**, 383–385 (1992).
45. Monsour, P. A. & Dudhia, R. Implant radiography and radiology. *Aust. Dent. J.* **53**(Suppl 1), S11–25. <https://doi.org/10.1111/j.1834-7819.2008.00037.x> (2008).
46. Wadu, S. G., Penhall, B. & Townsend, G. C. Morphological variability of the human inferior alveolar nerve. *Clin. Anat.* **10**, 82–87. [https://doi.org/10.1002/\(SICI\)1098-2353\(1997\)10:2%3c82::AID-CA2%3e3.0.CO;2-V](https://doi.org/10.1002/(SICI)1098-2353(1997)10:2%3c82::AID-CA2%3e3.0.CO;2-V) (1997).
47. Carter, R. B. & Keen, E. N. The intramandibular course of the inferior alveolar nerve. *J. Anat.* **108**, 433–440 (1971).
48. Ruder, S. An overview of multi-task learning in deep neural networks. *arXiv preprint arXiv:1706.05098*. <https://doi.org/10.48550/arXiv.1706.05098> (2017).
49. Kwon, O. *et al.* Automatic diagnosis for cysts and tumors of both jaws on panoramic radiographs using a deep convolution neural network. *Dentomaxillofac. Radiol.* **49**, 20200185. <https://doi.org/10.1259/dmfr.20200185> (2020).

Acknowledgements

This work was supported by a Seoul National University Research Grant in 2021. This work was also supported by the Korea Medical Device Development Fund Grant funded by the Korean Government (The Ministry of Science and ICT, The Ministry of Trade, Industry, and Energy, The Ministry of Health & Welfare, and The Ministry of Food and Drug Safety) (Project Number: 1711174552, KMDF_PR_20200901_0147 and Project Number: 1711174543, RS-2020-KD000011).

Author contributions

B.-S.J.: Contributed to the conception and design, data acquisition, analysis and interpretation, and drafted and critically revised the manuscript. S.Y.: Contributed to the conception and design, data analysis and interpretation, and drafted and critically revised the manuscript. S.-J.L.: Contributed to the data analysis and interpretation, and drafted the manuscript. T.-I.K.: Contributed to conception and design, data interpretation, and drafted the manuscript. J.-M.K.: Contributed to conception and design, data interpretation, and drafted the manuscript. J.-E.K.: Contributed to the conception and design, data interpretation, and drafted the manuscript. K.-H.H.: Contributed to the conception and design, data interpretation, and drafted the manuscript. S.-S.L.: Contributed to the conception and design, data interpretation, and drafted the manuscript. M.-S.H.: Contributed to the conception and design, data interpretation, and drafted the manuscript. W.-J.Y.: Contributed to the conception and design, data acquisition, analysis and interpretation, and drafted and critically revised the manuscript. All authors gave their final approval and agreed to be accountable for all aspects of the work.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to W.-J.Y.

Reprints and permissions information is available at www.nature.com/reprints.

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



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/>.

© The Author(s) 2022, corrected publication 2022