Int. J. Oral Maxillofac. Surg. 2021; xx: 1–6
https://doi.org/10.1016/j.ijom.2022.10.015, available online at https://www.sciencedirect.com

International Journal of Oral & Maxillofacial Surgery

Research Paper
Deep Learning

Deep-learning-based automatic facial bone segmentation using a two-dimensional U-Net

D. Morita[a], S. Mazen[b], S. Tsujiko[c], Y. Otake[b], Y. Sato[b], T. Numajiri[a]

[a]Department of Plastic and Reconstructive Surgery, Kyoto Prefectural University of Medicine, Kyoto, Japan; [b]Division of Information Science, Nara Institute of Science and Technology, Nara, Japan; [c]Department of Plastic and Reconstructive Surgery, Saiseikai Shigaken Hospital, Shiga, Japan

D. Morita, S. Mazen, S. Tsujiko, Y. Otake, Y. Sato, T. Numajiri: Deep-learning-based automatic facial bone segmentation using a two-dimensional U-Net. Int. J. Oral Maxillofac. Surg. 2021; xx: 1–6. © 2022 International Association of Oral and Maxillofacial Surgeons. Published by Elsevier Inc. All rights reserved.

Abstract. The use of deep learning (DL) in medical imaging is becoming increasingly widespread. Although DL has been used previously for the segmentation of facial bones in computed tomography (CT) images, there are few reports of segmentation involving multiple areas. In this study, a U-Net was used to investigate the automatic segmentation of facial bones into eight areas, with the aim of facilitating virtual surgical planning (VSP) and computer-aided design and manufacturing (CAD/CAM) in maxillofacial surgery. CT data from 50 patients were prepared and used for training, and five-fold cross-validation was performed. The output results generated by the DL model were validated by Dice coefficient and average symmetric surface distance (ASSD). The automatic segmentation was successful in all cases, with a mean ± standard deviation Dice coefficient of 0.897 ± 0.077 and ASSD of 1.168 ± 1.962 mm. The accuracy was very high for the mandible (Dice coefficient 0.984, ASSD 0.324 mm) and zygomatic bones (Dice coefficient 0.931, ASSD 0.487 mm), and these could be introduced for VSP and CAD/CAM without any modification. The results for other areas, particularly the teeth, were slightly inferior, with possible reasons being the effects of defects, bonded maxillary and mandibular teeth, and metal artefacts. A limitation of this study is that the data were from a single institution. Hence further research is required to improve the accuracy for some facial areas and to validate the results in larger and more diverse populations.

Keywords: Deep learning; Facial bones; X-ray computed tomography; Maxillofacial surgery; Artificial intelligence.

Accepted for publication 24 October 2022
Available online xxxx

In recent years, the use of deep learning (DL) has been advancing rapidly in many fields. Various types of DL have been introduced and utilized among the fields of medicine, with DL-based diagnostic imaging support being one of the typical examples.[1–3] DL is a machine-learning technique that enables computers to learn how to emulate the tasks that humans can perform naturally. DL is used in image recognition, speech recognition, and natural language processing, and is also being applied to the analysis of medical images such as computed tomography (CT) images and magnetic resonance images (MRI). Among DL techniques, convolutional neural networks have yielded effective and interesting results in the field of medical image processing, particularly medical image segmentation and detection.[4]

In maxillofacial surgery, computer-assisted surgical techniques such as virtual surgical planning (VSP) and computer-aided design/computer-aided manufacturing (CAD/CAM) have been evaluated widely and are becoming universal.[5,6] In the Department of Plastic and Reconstructive Surgery at Kyoto Prefectural University of Medicine, maxillary and mandibular reconstruction have been performed using these technologies, applying an in-house approach, and their usefulness has been reported.[7–12]

Currently, the application of VSP and CAD/CAM technologies to maxillofacial reconstruction is likely to use free software such as 3D Slicer (https://www.slicer.org/) or Blender (Blender Foundation, https://www.blender.org/) to convert facial-bone CT images in DICOM format (Digital Imaging and Communications in Medicine) into segmentation masks, with subsequent manual segmentation of each region. In addition to the technical aspects, an experienced doctor needs to make sure that the CT data are regionally segmented for each slice and correct the results while constantly adjusting the segmentation parameters. The current manual segmentation method is user-dependent, cumbersome, and time-consuming, leading to inefficiency and inconsistent segmentation. In addition, the presence of metal artefacts increases the workload significantly, leading to possible fatigue in the doctor performing the preparation, with any resulting error in accuracy possibly affecting the outcome of the surgery. Therefore, the current authors have been investigating whether it is possible to automate these tasks, particularly bone segmentation, using DL. These considerations imply that there could be many advantages in accurately automating this complex series of tasks.

The purpose of this study was to develop an algorithm for the automatic segmentation of facial-bone CT images into multiple regions using DL. Automatic segmentation is straightforward for a single anatomical bone such as the mandible, but it is difficult to segment multiple bones such as the maxilla, zygoma, nasal bone, frontal bone, and teeth because of the indistinct inter-bone boundaries and complex articulations. If the segmentation of facial bones could be performed automatically, rapidly, and accurately, it would contribute significantly to the development of surgery involving the maxillofacial region using computer-assisted technology.

Materials and methods

Facial-bone CT data from patients treated in the Department of Plastic and Reconstructive Surgery at Kyoto Prefectural University of Medicine were used in this work. Approval for the study was obtained from the institutional ethics committee (ERB-C-2211). The CT data from 50 patient cases were prepared. The selected data included those images where the imaging range covered the entire face; however, to ensure sufficient training data, some cases where the imaging range missed a marginal part of the face were also included.

Manual segmentation was performed by a single plastic surgeon to generate training data with accurately labelled segmentation. The free open-source software 3D Slicer was used for this labelling. It was decided to perform segmentation into eight areas: nasal bone, maxilla, mandible, left and right zygomatic bones, frontal bone, maxillary teeth, and mandibular teeth. In this study, the zygomatic bone was defined as the zygomatic bone up to the base of the zygomatic arch, and the temporal bone was deleted. The frontal bone was segmented linearly, according to its most likely position, because the actual suture line may be difficult to recognize from the CT image. In this way, a dataset comprising 50 accurately labelled facial-bone CT images was created. The rows labelled ground truth (GT) in Fig. 1 are examples of the training data.

PyTorch (https://pytorch.org/) was used as the DL framework in the experiments. An NVIDIA GeForce RTX 3090 graphics processing unit (GPU) with a 24 GB memory was used for the model training/testing (NVIDIA Corp., Santa Clara, CA, USA).

The DL model used the two-dimensional (2D) U-Net architecture as the convolutional neural network.[13] Five-fold cross-validation was performed to validate the model training. The U-Net was set up with six layers, i.e. one additional deep layer compared with the standard U-Net architecture. This increases the number of parameters in the DL model, thereby improving the flexibility of the analysis and the expressiveness of the results. In the training phase, the batch size was set to three, the loss function to cross entropy, the optimizer to Adam, the learning rate to 0.00001, and the number of iterations to 150,000.

The Dice coefficient and average symmetric surface distance (ASSD) were

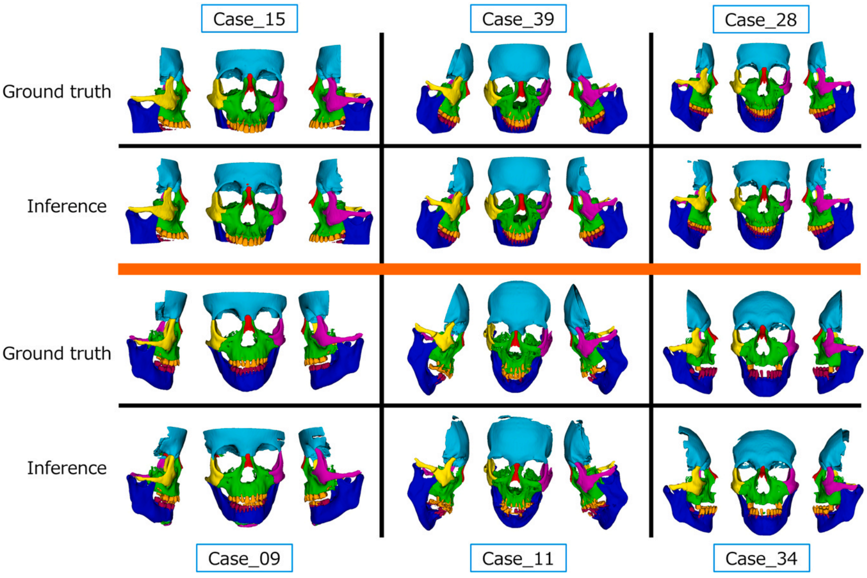

Fig. 1. Comparison of ground truth and inference results. The upper row shows cases with excellent results, with all structures, including the teeth, being segmented accurately. Although some error remains, it should be possible to obtain close to ground truth by manual correction. The bottom row shows cases with inferior results. Case 9 was a paediatric case. The other two cases had large errors in the tooth area, where the influence of teeth defects and metal artefacts is suggested. As a result of the teeth defects, the maxillary and mandibular teeth may have been replaced by inference.

Table 1. Summary of the results for all inference data: the Dice coefficient and ASSD results are shown for each bone region.

Bone region	Dice (mean ± SD)	ASSD (mean ± SD)
Nasal	0.838 ± 0.084	0.755 ± 1.194
Maxilla	0.909 ± 0.036	0.664 ± 0.463
Mandible	0.984 ± 0.017	0.324 ± 0.406
Right zygoma	0.936 ± 0.029	0.486 ± 0.408
Left zygoma	0.926 ± 0.049	0.487 ± 0.441
Frontal	0.858 ± 0.060	1.981 ± 1.547
Maxillary teeth	0.871 ± 0.074	1.781 ± 1.821
Mandibular teeth	0.846 ± 0.102	3.240 ± 4.324
All	0.897 ± 0.077	1.168 ± 1.962

ASSD, average symmetric surface distance; SD, standard deviation.

used to validate the inference results. Inference refers to the output results generated by the DL model. The Dice coefficient is calculated by dividing twice the overlapping area by the sum of both the GT and auto-segmented models and is widely accepted as an indicator of CT segmentation performance. The Dice coefficient takes a value between 0 and 1, with a value closer to 1 indicating a better agreement between the GT and auto-segmentation by DL.[14,15] The ASSD is the average distance between the boundaries of two object areas. This was used to measure the gap between the points of the GT surface and the auto-segmented surface area.[16–18] Mean ± standard deviation values were obtained.

The trained model for segmenting the facial bone is available at https://github.com/d-morita-prs/Facial-Bone-Segmentaion.

Results

The mean age of the 50 patient cases (25 female, 25 male) was 35.88 ± 19.80 years. Among the 50 images, 20 showed metallic artefacts.

The protocol used in this study was five-fold cross-validation, with five sections of training and validation. Each section took approximately 11 h and 10 min. Inference data for 10 cases were output for each section, requiring an average of 10 s to infer each item of CT data.

Table 1 gives a summary of the inference results. The mean Dice coefficient was 0.897 ± 0.077 and the mean ASSD was 1.168 ± 1.962 mm, for the inference results. The Dice coefficients for the maxilla, mandible, right zygoma, and left zygoma were all greater than 0.9, with mean values of 0.909 ± 0.036, 0.984 ± 0.017, 0.936 ± 0.029, and 0.926 ± 0.049, respectively. The ASSD values for the nasal bone, maxilla, mandible, right zygomatic bone, and left zygomatic bone were all less than 1.0 mm, with values of 0.755 ± 1.194 mm, 0.664 ± 0.463 mm, 0.324 ± 0.406 mm, 0.486 ± 0.408 mm, and 0.487 ± 0.441 mm, respectively. The results indicate that precise segmentation was produced for the maxilla, mandible, and two zygomatic bones, whereas the accuracy was somewhat poorer for the other areas.

Fig. 1 shows three examples of cases with good accuracy and three with poor accuracy. Apart from the cases with strong artefacts, the accuracy of the automatic segmentation was sufficient in most cases, and could be made completely accurate with only minor manual corrections.

Discussion

There are many reports on automatic segmentation of the facial bones using machine learning and DL. However, most have only dealt with the segmentation of a limited area, such as the mandible. The present study is novel in reporting on the automatic segmentation of eight regions that are important for surgical planning: the nasal bone, maxilla, mandible, left and right zygomatic bones, frontal bone, maxillary teeth, and mandibular teeth. Maxillofacial surgery involves a wide variety of surgical procedures involving facial bone reconstruction, including the treatment of traumatic injuries such as facial bone fractures, treatment of congenital malformations such as jaw deformities and osteogenesis imperfecta, and bone reconstruction after the resection of malignant tumours. Therefore, it is clinically very meaningful if the facial-bone image can be segmented into separate regions, because this will help the visualization of the surgery in these regions.

The DL model used for this study was U-Net,[19] a model initially proposed for biological image segmentation that has shown promising results in various fields.[13]

There are no clear criteria for an appropriate level of Dice coefficient that would be sufficient for a model to be applicable to VSP. Wallner et al.[20] reported that when different individuals segmented the mandibles of 10 cases using the same CT data, their Dice coefficients averaged 0.9408. In light of this report, the present authors suggest that an appropriate Dice coefficient of about that value would be sufficiently accurate.

Qiu et al.[21] performed a review of studies on the automated segmentation of the mandible. There have been many reports on segmentation using DL, with both two- and three-dimensional neural networks being used. The Dice coefficients reported included 0.9076 ± 0.0245 by Yan et al.,[22] 0.94 ± 0.02 by Xue et al.,[23] and 0.900 ± 0.042 by Lei et al.[24] From the results of the present study experiments, the mean Dice coefficient for the mandible was 0.984, which is at least as accurate as previous reports. As anticipated, the mandible, comprising a single bone, had a higher segmentation accuracy in the present study model when compared to the other more complex facial bones.

Regarding the maxillary bone, Yang and Su[25] reported a Dice coefficient of 0.92. The result in the present study, a mean of 0.909, is slightly inferior. This may be attributed to the fact that strong artefacts in the teeth affected the segmentation of the maxillary bone near the teeth, resulting in inaccuracy in some cases.

The zygomatic bone showed sufficient accuracy, with a mean Dice coefficient of 0.936 for the right side and 0.926 for the left side. The ASSD values were 0.486 mm for the right side and 0.487 mm for the left side. It appeared to be similar between the left and the right sides.

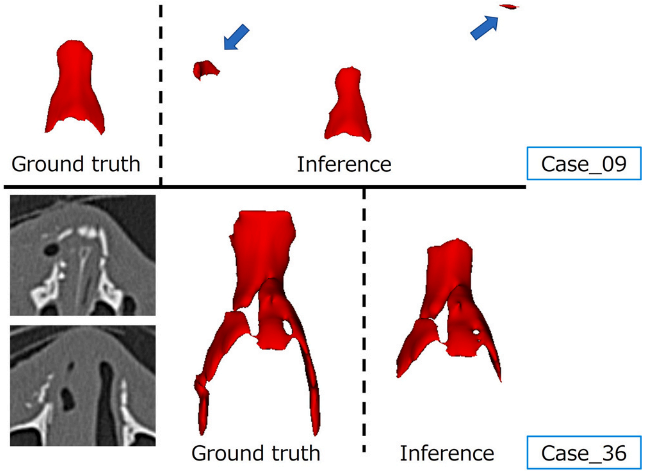

Fig. 2. Inference output of outlier cases—nasal bone. Case 9 was a paediatric patient. As indicated by the blue arrows, wrong output was observed in areas other than the nasal bone. In addition, the output of the nasal bone itself is observed to be smaller. In case 36, there was a complex fracture of the nasal bone with displaced bone fragments. As in case 9, the output of the nasal bone itself is smaller.

The nasal bone and frontal bone showed slightly lower segmentation accuracy than the other regions. This is thought to be because the nasal bone is small and thin, and the suture lines are difficult to recognize. When the output of cases with particularly low accuracy was reviewed, border obscuration caused by nasal bone fractures and poor segmentation in paediatric cases were observed (see Fig. 2). However, the segmentation of another paediatric case was accurate, suggesting that there may have been insufficient training data involving paediatric cases. The frontal bone was similarly difficult to delineate because its suture lines are difficult to see. Its training labels were assigned on the basis of the most likely position of the suture lines, rather than by using anatomical feature points as visual landmarks.

Particularly for the maxillary and mandibular teeth, many cases were outliers in terms of accuracy. Observation of the actual output showed that many images involved strong artefacts, teeth defects, and bonding between the maxillary and mandibular teeth (see Fig. 3). Metal artefacts affected both the maxillary and mandibular teeth, and, in areas with defects or bonded teeth, the maxillary and mandibular dentitions could appear interchanged. In addition, the roots of the teeth often lacked distinct boundaries. Therefore, in cases with poor tooth segmentation, the results for both maxillary and mandibular teeth tended to be less accurate.

In summary, the automatic segmentation algorithm used in this study was able to segment the facial bones into eight regions with high accuracy when there were few artefacts and no serious issues with tooth boundaries. For the mandible and zygoma, the accuracy was sufficiently high to be implemented in VSP and CAD/CAM technology. However, this study revealed three main problems, namely (1) the accuracy is reduced in cases involving strong artefacts, (2) for cases where the maxillary and mandibular teeth are bonded or lack clear boundaries, the accuracy of tooth segmentation is compromised, and (3) the model may lack diversity, because it is based on data from only a single institution.

To address these issues, the next steps should include additional processing to reduce artefacts,[26,27] and the addition of training data with greater variability, along with various data augmentation and spatial normalization approaches, in order to account for inter-patient variations in teeth anatomy and positioning.

In conclusion, using U-Net as a deep learning model, CT images of the facial bones were segmented into eight regions automatically and accurately. Highly accurate segmentation was achieved for the mandible and zygomatic bones, which is sufficient for this approach to be incorporated into VSP and CAD/CAM technology. Problems involving artefacts, tooth condition, and diversity of the training data

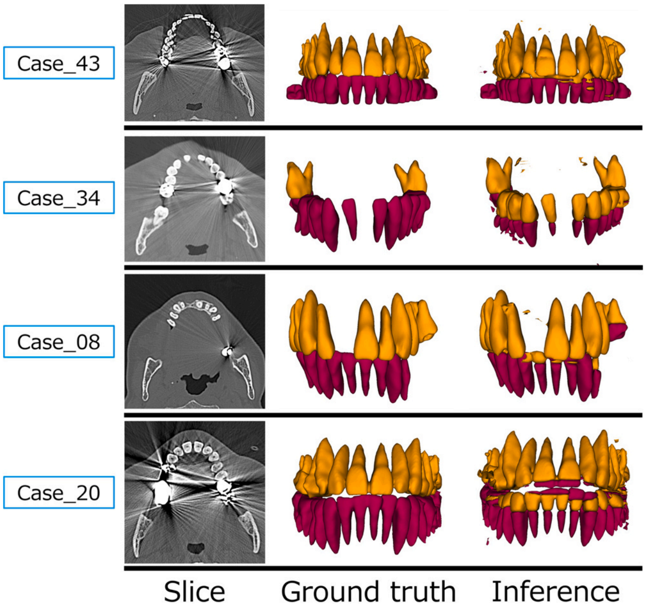

Fig. 3. Examples of cases with inaccurate tooth segmentation. All cases had metal artefacts. It is suggested that there are teeth defects or bonding of the maxillary and mandibular teeth, which causes the maxillary and mandibular teeth to be replaced in the inference output. In addition, noise-like error output is observed in places.

remain, with further research being needed to address these issues.

Funding

None.

Competing interests

None.

Ethical approval

The study was conducted with the approval of the Institutional Ethics Committee of Kyoto Prefectural University of Medicine, Kyoto, Japan (ERB-C-2211).

Patient consent

The study institution applies an opt-out consent process for the use of patient data in research. No patient whose data were included in this study opted out.

References

1. Liu B, Chi W, Li X, Li P, Liang W, Liu H, Wang W, He J. Evolving the pulmonary nodules diagnosis from classical approaches to deep learning-aided decision support: three decades' development course and future prospect. *J Cancer Res Clin Oncol* 2020;**146**:153–85.
2. Ström P, Kartasalo K, Olsson H, Solorzano L, Delahunt B, Berney DM, Bostwick DG, Evans AJ, Grignon DJ, Humphrey PA, Iczkowski KA, Kench JG, Kristiansen G, van der Kwast TH, Leite KRM, McKenney JK, Oxley J, Pan CC, Samaratunga H, Srigley JR, Takahashi H, Tsuzuki T, Varma M, Zhou M, Lindberg J, Lindskog C, Ruusuvuori P, Wählby C, Grönberg H, Rantalainen M, Egevad L, Eklund M. Artificial intelligence for diagnosis and grading of prostate cancer in biopsies: a population-based, diagnostic study. *Lancet Oncol* 2020;**21**:222–32.
3. Tanaka H, Chiu SW, Watanabe T, Kaoku S, Yamaguchi T. Computer-aided diagnosis system for breast ultrasound images using deep learning. *Phys Med Biol* 2019;**64**:235013.
4. Anwar SM, Majid M, Qayyum A, Awais M, Alnowami M, Khan MK. Medical image analysis using convolutional neural networks: a review. *J Med Syst* 2018;**42**:226.
5. Day KM, Kelley PK, Harshbarger RJ, Dorafshar AH, Kumar AR, Steinbacher DM, Patel P, Combs PD, Levine JP. Advanced three-dimensional technologies in craniofacial reconstruction. *Plast Reconstr Surg* 2021;**148**:94e–108e.
6. Padilla PL, Mericli AF, Largo RD, Garvey PB. Computer-aided design and manufacturing versus conventional surgical planning for head and neck reconstruction: a systematic review and meta-analysis. *Plast Reconstr Surg* 2021;**148**:183–92.
7. Morita D, Numajiri T, Nakamura H, Tsujiko S, Sowa Y, Yasuda M, Hirano S. Intraoperative change in defect size during maxillary reconstruction using surgical guides created by CAD/CAM. *Plast Reconstr Surg Glob Open* 2017;**5**:e1309.
8. Morita D, Numajiri T, Tsujiko S, Nakamura H, Yamochi R, Sowa Y, Yasuda M, Hirano S. Secondary maxillary and orbital floor reconstruction with a free scapular flap using cutting and fixation guides created by computer-aided design/computer-aided manufacturing. *J Craniofac Surg* 2017;**28**:2060–2.
9. Numajiri T, Morita D, Nakamura H, Tsujiko S, Yamochi R, Sowa Y, Toyoda K, Tsujikawa T, Arai A, Yasuda M, Hirano S. Using an in-house approach to computer-assisted design and computer-aided manufacturing reconstruction of the maxilla. *J Oral Maxillofac Surg* 2018;**76**:1361–9.
10. Numajiri T, Morita D, Nakamura H, Yamochi R, Tsujiko S, Sowa Y. Designing CAD/CAM surgical guides for maxillary reconstruction using an in-house approach. *J Vis Exp* 2018;**138**:58015.
11. Numajiri T, Morita D, Yamochi R, Nakamura H, Tsujiko S, Sowa Y, Toyoda K, Tsujikawa T, Arai A, Hirano S. Does an in-house computer-aided design/computer-aided manufacturing approach contribute to accuracy and time shortening in mandibular reconstruction? *J Craniofac Surg* 2020;**31**:1928–32.
12. Numajiri T, Nakamura H, Sowa Y, Nishino K. Low-cost design and manufacturing of surgical guides for mandibular reconstruction using a fibula. *Plast Reconstr Surg Glob Open* 2016;**4**:e805.
13. Hiasa Y, Otake Y, Takao M, Ogawa T, Sugano N, Sato Y. Automated muscle segmentation from clinical CT using Bayesian U-Net for personalized musculoskeletal modeling. *IEEE Trans Med Imaging* 2019;**39**:1030–40.
14. Chang HH, Zhuang AH, Valentino DJ, Chu WC. Performance measure characterization for evaluating neuroimage segmentation algorithms. *Neuroimage* 2009;**47**:122–35.
15. Ghafoorian M, Karssemeijer N, Heskes T, van Uden IWM, Sanchez CI, Litjens G, de Leeuw FE, van Ginneken B, Marchiori E, Platel B. Location sensitive deep convolutional neural networks for segmentation of white matter hyperintensities. *Sci Rep* 2017;**7**:5110.
16. Tong N, Gou S, Yang S, Ruan D, Sheng K. Fully automatic multi-organ segmentation for head and neck cancer radiotherapy using shape representation model constrained fully convolutional neural networks. *Med Phys* 2018;**45**:4558–67.
17. Warfield SK, Zou KH, Wells WM. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE Trans Med Imaging* 2004;**23**:903–21.
18. Yeghiazaryan V, Voiculescu I. Family of boundary overlap metrics for the evaluation of medical image segmentation. *J Med Imaging* 2018;**5**:015006.
19. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells W, Frangi A, editors. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Springer; 2015. p. 234–41.
20. Wallner J, Mischak I, Jan E. Computed tomography data collection of the complete human mandible and valid clinical ground truth models. *Sci Data* 2019;**6**:190003.
21. Qiu B, van der Wel H, Kraeima J, Glas HH, Guo J, Borra RJH, Witjes MJH, van Ooijen PMA. Automatic segmentation of mandible from conventional methods to deep learning—a review. *J Pers Med* 2021;**11**:629.
22. Yan M, Guo J, Tian W, Yi Z. Symmetric convolutional neural network for mandible segmentation. *Knowl Based Syst* 2018;**159**:63–71.
23. Xue J, Wang Y, Kong D, Wu F, Yin A, Qu J, Liu X. Deep hybrid neural-like P systems for multiorgan segmentation in head and neck CT/MR images. *Expert Syst Appl* 2021;**168**:114446.
24. Lei W, Mei H, Sun Z, Ye S, Gu R, Wang H, Huang R, Zhang S, Zhang S, Wang G. Automatic segmentation of organs-at-risk from head-and-neck CT using separable convolutional neural network with hard-region-weighted loss. *Neurocomputing* 2021;**442**:184–99.

25. Yang WF, Su YX. Artificial intelligence-enabled automatic segmentation of skull CT facilitates computer-assisted craniomaxillofacial surgery. *Oral Oncol* 2021; **118**:105360.

26. Sakamoto M, Hiasa Y, Otake Y, Takao M, Suzuki Y, Sugano N, Sato Y. Bayesian segmentation of hip and thigh muscles in metal artifact-contaminated CT using convolutional neural network-enhanced normalized metal artifact reduction. *J Signal Process Syst* 2020; **92**:335–44.

27. Nakao M, Imanishi K, Ueda N, Imai Y, Kirita T, Matsuda T. Regularized three-dimensional generative adversarial nets for unsupervised metal artifact reduction in head and neck CT images. *IEEE Access* 2020;**8**:109453–65.

Correspondence to: Department of Plastic and Reconstructive Surgery
Kyoto Prefectural University of Medicine
465
Kajiicho Kamigyoku
Kyoto 602-8566
Japan. Tel/Fax: +81 75 251 5730 (5732).
E-mail: d-morita@koto.kpu-m.ac.jp