Computers in Biology and Medicine 80 (2017) 24–29

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/cbm

# Classification of teeth in cone-beam CT using deep convolutional neural network

Yuma Miki[a], Chisako Muramatsu[a,*], Tatsuro Hayashi[b], Xiangrong Zhou[a], Takeshi Hara[a], Akitoshi Katsumata[c], Hiroshi Fujita[a]

[a] Department of Intelligent Image Information, Graduate School of Medicine, Gifu University, 1-1 Yanagido, Gifu, Gifu 501-1194, Japan
[b] Media Co., Ltd., 3-26-6 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
[c] Department of Oral Radiology, School of Dentistry, Asahi University, 1851 Hozumi, Mizuho, Gifu 501-0296, Japan

ARTICLE INFO

*Keywords:*
Deep convolutional neural networks
Tooth classification
Dental cone-beam CT
Forensic identification
Dental chart

ABSTRACT

Dental records play an important role in forensic identification. To this end, postmortem dental findings and teeth conditions are recorded in a dental chart and compared with those of antemortem records. However, most dentists are inexperienced at recording the dental chart for corpses, and it is a physically and mentally laborious task, especially in large scale disasters. Our goal is to automate the dental filing process by using dental x-ray images. In this study, we investigated the application of a deep convolutional neural network (DCNN) for classifying tooth types on dental cone-beam computed tomography (CT) images. Regions of interest (ROIs) including single teeth were extracted from CT slices. Fifty two CT volumes were randomly divided into 42 training and 10 test cases, and the ROIs obtained from the training cases were used for training the DCNN. For examining the sampling effect, random sampling was performed 3 times, and training and testing were repeated. We used the AlexNet network architecture provided in the Caffe framework, which consists of 5 convolution layers, 3 pooling layers, and 2 full connection layers. For reducing the overtraining effect, we augmented the data by image rotation and intensity transformation. The test ROIs were classified into 7 tooth types by the trained network. The average classification accuracy using the augmented training data by image rotation and intensity transformation was 88.8%. Compared with the result without data augmentation, data augmentation resulted in an approximately 5% improvement in classification accuracy. This indicates that the further improvement can be expected by expanding the CT dataset. Unlike the conventional methods, the proposed method is advantageous in obtaining high classification accuracy without the need for precise tooth segmentation. The proposed tooth classification method can be useful in automatic filing of dental charts for forensic identification.

## 1. Introduction

Dental records play an important role in forensic identification after large-scale disasters [1–3]. Forensic dentistry is an important field because dental information can be used for identifying a person even when his/her body has been severely damaged; moreover, antemortem (AM) x-ray images are easier to collect than DNA samples. For dental identification, postmortem (PM) dental findings and teeth conditions are recorded in a dental chart. However, most dentists are inexperienced at recording the dental chart for corpses, and it can cause a psychiatric burden. Such a psychiatric stress may also lead to incorrect data recording and psychiatric disorders.

For overcoming these drawbacks, studies have proposed automatically obtaining dental information from the dental x-ray images, creating panoramic-like images from CT data for better image comparison, and/or matching the AM and PM images [4–10]. Jain et al. [4] investigated a computerized method for matching the AM and PM dental images. Each tooth was first isolated from its neighbors and the tooth contour was extracted on the basis of intensity. The corresponding image was then searched for by matching the extracted contours with rigid transformation. Among 38 AM/PM image pairs, 25 were correctly matched while the genuine AM image was selected as the second-best match in 5 of the remaining 13 cases. Zhou et al. [5] proposed a 3-step method to retrieve matched images. Images were first classified into bitewing, periapical, or panoramic images, and the teeth on bitewing images were segmented using a top-hat filter and an active contour method. The corresponding image was searched for by matching the boundary shape.

* Corresponding author.
*E-mail address:* chisa@fjt.info.gifu-u.ac.jp (C. Muramatsu).

http://dx.doi.org/10.1016/j.compbiomed.2016.11.003
Received 7 September 2016; Received in revised form 7 November 2016; Accepted 9 November 2016

Lin et al. [8] proposed a method to classify teeth on bitewing images. After tooth segmentation, the length and width ratio and crown size were used as features to classify each tooth as a molar or premolar by using a support vector machine. They achieved an overall classification accuracy of 95% when using 47 images containing 369 teeth. Hosntalab et al. [10] proposed a multi-stage technique for the classification and numbering of teeth on multi-slice CT images. The 3 step process included the segmentation of tooth regions using a variation level set, feature extraction of the wavelet-Fourier descriptor, and classification of teeth into 4 groups using a supervised classifier. Using this technique, they achieved high classification accuracies above 94% for 804 teeth from 30 CT cases [10].

In this study, as a component of automated dental chart filing system, we investigated an automated method for classifying tooth types on dental cone-beam CT images using a deep convolutional neural network (DCNN). Since the success of Krizhevsky et al. [11] in the ImageNet 2012 competition, DCNNs have shown their outstanding ability in object recognition and natural image classification. The application of DCNNs to medical images has been increasingly investigated by many groups that have achieved certain degrees of success [12–17]. However, a successful application procedure has not yet been established, and to our knowledge DCNNs have only been applied to dental image processing in one study. Wang et al. [18] reported the comparison of dental radiography analysis algorithms for the grand challenges held in IEEE international Symposium on Biomedical Imaging 2015, in which Ronneberger et al. employed u-shaped deep convolution neural network for segmentation of bitewing radiographs for caries detection. In this preliminary study, we investigated the utility of a DCNN in classifying teeth into 7 types by using rectangular regions of interest (ROIs), each of which enclosed a tooth from an axial slice.

## 2. Material and methods

### 2.1. Image dataset

The images used in this study were obtained using two dental CT units, namely Veraviewepocs 3D (J.Morita Mfg, Corp., Kyoto, Japan) and Alphard VEGA (Asahi Roentgen Ind. Co., Ltd., Kyoto, Japan), which were used to acquire images in 33 and 19 cases, respectively. The images were obtained from Asahi University Hospital, Gifu, Japan. The diameter of the field of view ranged from 51 to 200 mm, and the voxel resolution ranged from 0.1 to 0.39 mm. The institutional review boards of Gifu University and Asahi University approved the study protocol.

In general, medical CT images employ Hounsfield units for representing the gray levels. However, in dental cone beam CT images, gray levels are not standardized. Therefore, in this study, the window level and window width were manually adjusted to have an appearance similar to the model image, in which the dental region was clearly visualized. The average window level was 701 ± 461 and the average window width was 1338 ± 973 for 12-bit images. The gray level was then reduced to 8 bits prior to DCNN training and testing. From the 52 cases, 5 cases each from the two imaging systems were randomly selected and used as the evaluation dataset, and the remaining were used as the training dataset. For examining a sampling effect, the test dataset was sampled 3 times, and the training and testing were repeated.

### 2.2. ROI extraction

For both the training and test cases, the smallest bounding box enclosing each tooth was manually placed on the CT volume. From the bounding box, all axial slices excluding the upper and lower 20% were used as the training and test ROIs. As this was a preliminary investigation, the ROIs affected by metal artifacts were not included in this study. The number of ROIs obtained from a single tooth ranged from 19 to 171, with an average of 45. The 7 tooth types included the central incisors, lateral incisors, canines, first and second premolars, and first and second molars. The third molars were excluded from this study because of the small number of samples. Fig. 1 shows an example of the dental chart with different tooth types. The total number of ROIs extracted from the CT volumes was 6653, 6766. 7928, 5794, 3346, 2115, and 2657, respectively, for each of the abovementioned 7 tooth types. For the test cases, all ROIs were used for evaluation. The number of test ROIs for the 7 tooth types in each of the 3 samplings is listed in Table 1. For training the DCNN, the number of samples was balanced to the minimum number of ROIs in the 7 tooth types by random sampling. The number of the training ROIs in the 3 samplings was 11354, 12572, and 12985. Fig. 2 shows the extracted sample ROIs.

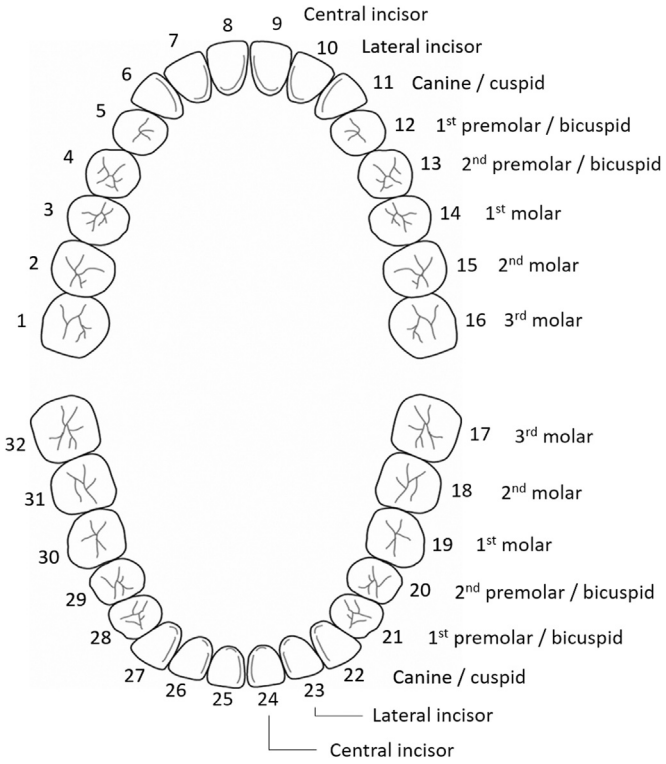


**Fig. 1.** Dental chart and tooth types.

### 2.3. Data augmentation

A small number of training cases can often leads to overtraining. Because the number of training cases in our dataset was limited, the training data were augmented by image manipulation [11,19], and the results with and without data augmentation were compared. One method that we investigated in this study was image rotation. Another method was intensity transformation by gamma correction defined as follow:

$$y = I_{max} \cdot \left( \frac{x}{I_{max}} \right)^{1/\gamma},$$

where x and y are the input and output pixel values, respectively, and Imax is the maximum pixel value, which is 255 for the input image. For the image rotation, the ROIs were extracted by rotating the CT volume from −10 to +10° in 5° steps along the x-y plane and replacing the bounding box. This resulted in a sample size 5 times the original sample size. For intensity transformation, γ values of 0.75 and 1.5 were applied, thereby increasing the number of samples by a factor of 3. The sample images are shown in Fig. 3.

**Table 1**
Numbers of test ROIs for seven tooth types in 3 random samplings.

| | Central incisor | Lateral incisor | Canine | 1st premolar | 2nd premolar | 1st molar | 2nd molar |
|---|---|---|---|---|---|---|---|
| Set 1 | 1169 | 1148 | 1369 | 991 | 716 | 493 | 497 |
| Set 2 | 1277 | 1235 | 1419 | 966 | 639 | 319 | 388 |
| Set 3 | 1208 | 1254 | 1425 | 938 | 401 | 260 | 344 |

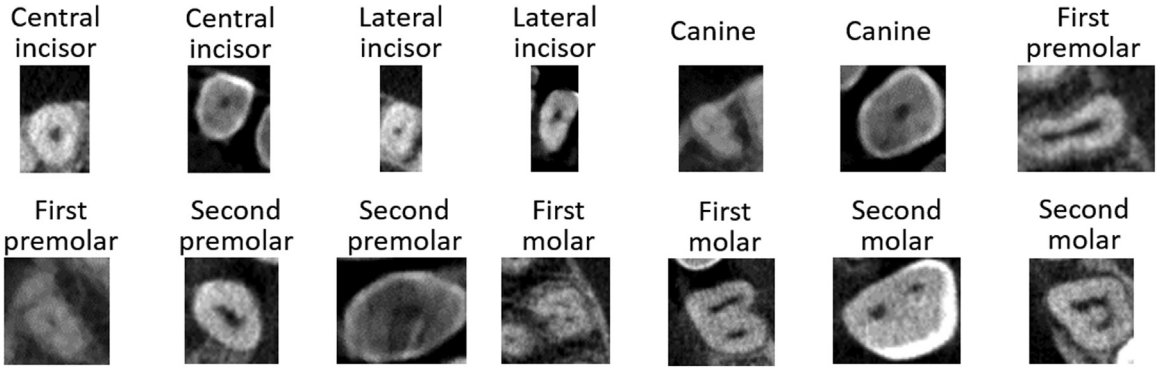


**Fig. 2.** Regions of interest including single tooth extracted from axial slices.

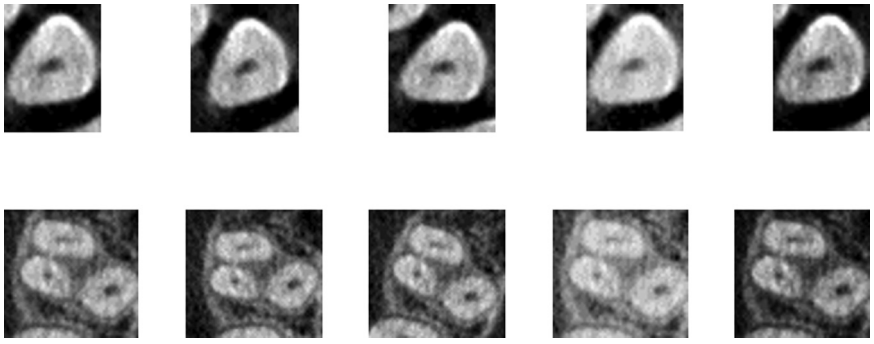

**Fig. 3.** Samples created by rotation and intensity transformation. (from left to right: original, 10° rotation, −10° rotation, high intensity (γ=1.5), and low intensity (γ=0.75) images).

### 2.4. Network architecture

In this study, we employed the AlexNet [11] architecture. The AlexNet, which was proposed by Hinton et al., had won the first place in the ImageNet Large Scale Visual Recognition Challenge 2012 [20], with an exceptional recognition performance. The network can be used from Caffe [21], which is the deep learning framework developed by members mainly working at the Berkeley Vision and Learning Center. The AlexNet architecture is shown in Fig. 4. Our input images were grayscale images with 227×227 pixels, and the network consisted of 5 convolution and 3 max pooling layers. The filter size, the number of filters, and stride for each layer are specified in Fig. 4. Each convolution layer is followed by a rectified linear unit (ReLU). In the last layers, all the units are fully connected to output probabilities for 7 classes using the softmax function. In the actual training, the AlexNet employs dropout [11], in which some unit activations in the fully connected layers are randomly set to zero, so that these connection weights would not be updated to prevent overfitting. As in the original model, we employed a dropout rate of 50%. For increasing the learning speed, the network uses minibatch updates, in which gradient decent computation and updates are carried out for a group of samples. In this study, we set the batch size as 100. The number of training epochs was set to 30, in which one epoch corresponds to the training of all samples once. The base learning rate was set to 0.01, and was decreased by factor of 10 after every 10 epochs. The momentum was 0.9 as in the original network. The training time with 194775 ROIs was approximately 2.5 h when using a GeForce GTX TITAN X GPU (NVIDIA Corporation).

## 3. Experimental results

A DCNN was employed to classify each ROI into 7 tooth types. We compared the classification results using training dataset A without data augmentation, dataset B with image rotation, dataset C with intensity transformation, and dataset D with both image rotation and intensity transformation. The number of training samples in these datasets is listed in Table 2. The original hand-cropped ROIs had different sizes. These images were resized to 256×256 pixels automatically prior to being randomly cropped to 227×227 pixels as they were input to Caffe. We compared the results for the four resizing methods provided by DIGITS, which are (i) crop, (ii) squash, (iii) fill, and (iv) half crop half fill. Fig. 5 shows the images resized using the four methods.

The classification accuracies for the different datasets and resizing methods are presented in Table 3. For all the resizing methods, the classification accuracy improved on including the rotated and window-adjusted samples in the training. Among the resizing methods, the fill method provided the best classification accuracy with dataset D in random sampling 3. However, it was not always the best method for other datasets, and the differences among the different methods were negligible.

Table 4 shows the confusion matrix for the result of dataset D and resizing method (iii) in sample set 3. The most number of misclassifications occurred for neighboring teeth, i.e., a tooth type was misclassified as its neighboring tooth type. For example, the lateral incisors were misclassified as central incisors and the first premolars as the second premolars.

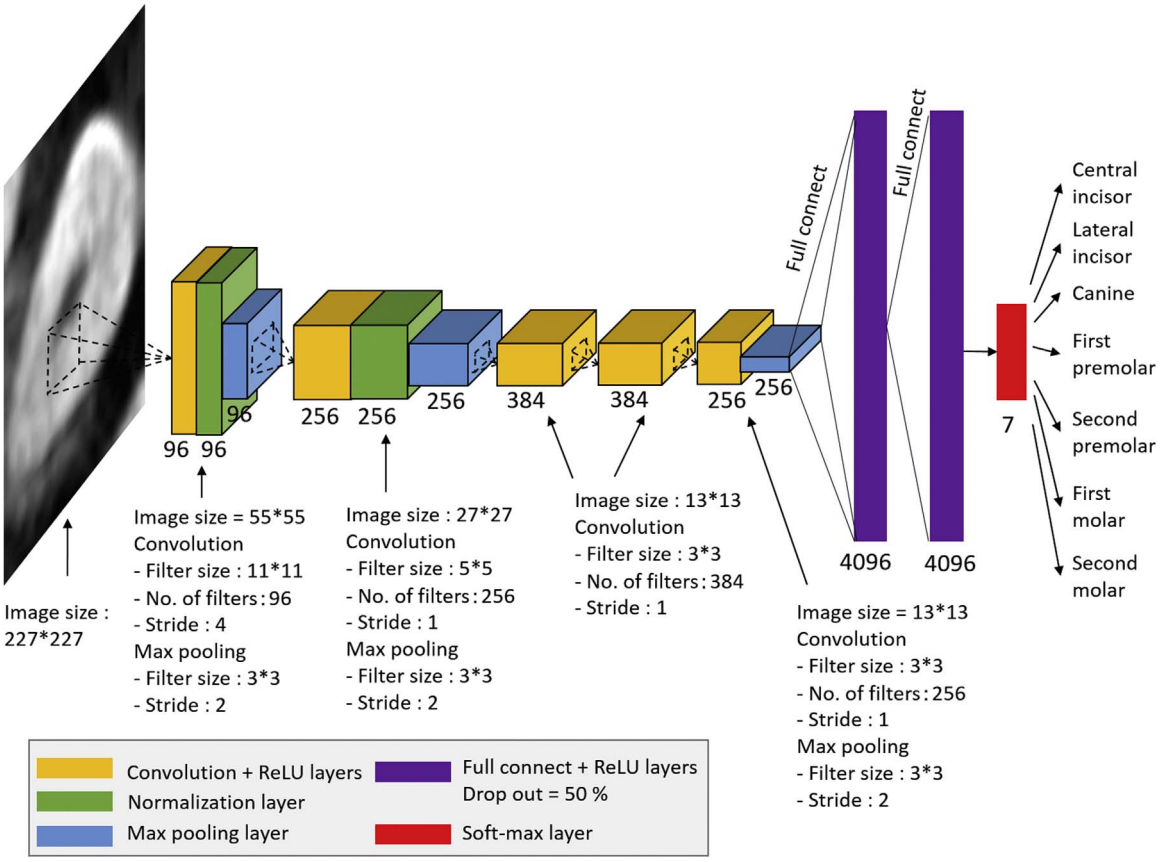


**Fig. 4.** An illustration of AlexNet architecture applied for dental images.

**Table 2**
Numbers of training ROIs by data augmentation.

| | Dataset A (original) | Dataset B (rotation) | Dataset C (intensity transformation) | Dataset D (rotation and intensity transformation) |
|---|---|---|---|---|
| Random set 1 | 11354 | 56770 | 34062 | 170310 |
| Random set 2 | 12572 | 62860 | 37716 | 188580 |
| Random set 3 | 12985 | 64925 | 38955 | 194775 |

**Table 3**
Classification accuracies (%) for with and without data augmentation datasets and different resizing methods.

| | Random set | (i) Crop | (ii) Squash | (iii) Fill | (iv) Half crop half fill |
|---|---|---|---|---|---|
| Dataset A | 1 | 78.5 | 82.0 | 79.9 | 80.7 |
| | 2 | 81.3 | 84.1 | 82.9 | 83.7 |
| | 3 | 78.3 | 84.4 | 81.9 | 82.5 |
| | Average | 79.4 | 83.5 | 81.6 | 82.3 |
| Dataset B | 1 | 79.3 | 84.0 | 82.5 | 82.9 |
| | 2 | 85.0 | 87.4 | 86.9 | 87.1 |
| | 3 | 82.3 | 86.6 | 85.2 | 84.8 |
| | Average | 82.2 | 86.0 | 84.9 | 84.9 |
| Dataset C | 1 | 80.7 | 85.6 | 84.2 | 83.6 |
| | 2 | 84.1 | 87.8 | 87.4 | 88.3 |
| | 3 | 85.7 | 89.1 | 89.1 | 85.2 |
| | Average | 83.5 | 87.5 | 86.9 | 85.7 |
| Dataset D | 1 | 82.3 | 86.0 | 86.8 | 85.5 |
| | 2 | 86.5 | 89.1 | 88.7 | 89.6 |
| | 3 | 86.9 | 90.5 | 91.0 | 90.2 |
| | Average | 85.2 | 88.5 | 88.8 | 88.4 |

## 4. Discussion

Using the DCNN, most teeth could be correctly classified to one of the 7 tooth types. One of the advantages of this method is that it does not require the precise tooth segmentation as might be required for the conventional feature-based classification. Because the convolution with the pooling process is robust enough to withstand image shift, it is possible to automatically recognize and classify tooth type from the larger, whole dental regions.

Deep learning in general is considered to require a large number of training samples. In this study, despite the limited number of cases, the

**Fig. 5.** Images created by different resizing methods. (From left to right: original image, images by crop, squash, fill, and half crop half fill methods).

**Table 4**
Confusion matrix of classification result for Dataset D with fill resizing method.

| | Central incisor | Lateral incisor | Canine | 1st premolar | 2nd premolar | 1st molar | 2nd molar |
|---|---|---|---|---|---|---|---|
| Central incisor | 1160 (96.0%) | 40 (3.6%) | 2 (0.2%) | 1 (0.1%) | 1 (0.1%) | 0 | 0 |
| Lateral incisor | 160 (12.8%) | 1066 (85.0%) | 18 (1.4%) | 4 (0.3%) | 0 | 0 | 6 (0.5%) |
| Canine | 9 (0.6%) | 93 (6.5%) | 1295 (90.9%) | 18 (1.3%) | 2 (0.1%) | 0 | 8 (0.6%) |
| First premolar | 1 (0.1%) | 6 (0.6%) | 19 (2.0%) | 859 (91.6%) | 50 (5.3%) | 2 (0.2%) | 1 (0.1%) |
| Second premolar | 0 | 0 | 0 | 33 (8.2%) | 362 (90.3) | 6 (1.5%) | 0 |
| First molar | 0 | 0 | 0 | 0 | 0 | 252 (96.9%) | 8 (3.1%) |
| Second molar | 1 (0.3%) | 8 (2.3%) | 0 | 11 (3.2%) | 13 (3.8%) | 9 (2.6%) | 302 (87.8%) |

classification accuracy was relatively high (above 80%) even without data augmentation. By increasing the number of samples by using rotation and intensity transformation, the classification performance was further improved. One of the characteristics of dental cone-beam CT images is the variation in image quality. Therefore, including samples with different contrasts was effective in improving the overall classification performance. For the same type of teeth, individual variations in growth directions and spatial relationships with neighboring teeth are large. Including the rotated image samples was, therefore, expected to contribute to increasing the classification accuracy.

Among the four resizing methods, the fill method (iii) provided the best accuracy for dataset D; however, for the other datasets, the squash method (ii) provided the best accuracy. The slight differences in accuracies indicate that the effect of the resizing methods is small, and that the DCNN is robust enough to withstand small image distortion.

In this study, we classified teeth into 7 types for possible application in automated filing of dental charts. In contrast, Hosntalab et al. [10] classified teeth into 4 types by first segmenting each tooth with manual correction, if needed, and applying a 3-layered neural network using wavelet-Fourier features. They obtained a classification accuracy of above 94%. We tested our DCNN for the 4-class classification by using dataset D with the fill method and obtained a classification accuracy of 94.4%. Although the result cannot be directly compared, we infer that a similar classification accuracy can be achieved using the DCNN without the need for precise tooth segmentation.

Wang et al. [18] compared computer algorithms for 2 challenges on dental radiography analysis. One was to identify anatomical landmarks on lateral cephalograms and to classify anatomical types based on the landmarks for diagnosis and orthodontic treatment. The second challenge was to segment seven tooth structures on bitewing radiographs. For the tooth segmentation, Ronneberger et al. employed a U-shaped deep convolutional neural network which includes layers of contracting (downsampling) convolution followed by layers of deconvolution (upsampling). Lee el al. employed a random forest to first predict the probabilities of tooth structure classes for each pixel, and post-processing to output final classes. Although the purpose and image modality are different from those in the current study, such techniques can be used for tooth segmentation and classification for forensic identification purposes.

One of the limitations of this study was the small amount of evaluation data. Another was that the slice images were evaluated independently. Cross-sectional tooth images at certain levels are likely more difficult to classify than those at other levels, such as the ones at the crown and root levels. We expect that the classification accuracy can be improved by combining the results for a tooth, or by applying 3D convolution. Therefore, we plan on investigating the effect of 3D training when additional cases are obtained in the future.

In this preliminary investigation, we excluded the ROIs affected by metal artifacts. We expect that the classification of tooth types for such images would be difficult even if these cases were included in the training samples. An effective preprocessing method for decreasing the effect of metal artifact is needed for such cases.

## 5. Conclusion

We have investigated the utility of the DCNN for classifying of tooth types on dental cone-beam CT. By increasing the number of training samples by rotation and intensity transformation, the classification performance was improved and a high accuracy of 91.0% was achieved. The 7-tooth-type classification result can be effectively used for automatic preparation of dental charts, which may be useful in forensic identification.

## Conflict of interest

None declared.

## Acknowledgement

This study was partly supported by a Grant-in-Aid for Scientific Research (B) (No. 26293402) by Japan Society for the Promotion of Science and a Grant-in-Aid for Scientific Research on Innovative Areas (Multidisciplinary Computational Anatomy, No. 26108005) by Ministry of Education, Culture, Sports, Science and Technology, Japan.

## References

[1] J.D. Wood, Forensic dental identification in mass disasters: the current status, J. Calif. Dent. Assoc. 42 (2014) 379–383.
[2] G. Sable, D. Rindhe, A review of dental biometrics from teeth feature extraction and matching techniques, Int. J. Sci. Res. 3 (2014) 2720–2722.
[3] T.D. Ruder, Y.A. Thali, S.N.A. Rashid, M.T. Mund, M.J. Thali, G.M. Hatch, A.M. Christensen, S. Somaini, G. Ampanozi, Validation of post mortem dental CT for disaster victim identification, J. Forensic Radiol. Imaging 5 (2016) 25–30.
[4] A.K. Jain, H. Chen, Matching of dental X-ray images for human identification, Pattern Recognit. 37 (2015) 1295–1305.
[5] J. Zhou, M. Abdel-Mottaleb, A content-based system for human identification based on bitewing dental X-ray images, Pattern Recognit. 38 (2005) 2132–2142.
[6] S. Tohnak, A. Mehnert, M. Mahoney, S. Crozier. Dental identification system based on unwrapped CT images, in: Proceedings of the International Conference of the IEEE Engineering in Medicine and Biology Society, 2009, pp. 3549–3552.
[7] D.A. Trochesset, R.B. Serchuk, D.C. Colosi, Generation of intra-oral-like images from cone beam computed tomography volumes for dental forensic image comparison, J. Forensic Sci. 59 (2015) 510–513.
[8] P.L. Lin, Y.H. Lai, P.W. Huang, An effective classification and numbering system for dental bitewing radiographs using teeth region and contour information, Pattern Recognit. 43 (2010) 1380–1392.
[9] A.Z. Arifin, M. Hadi, A. Yuniarti, W. Khotimah, A. Yudhi, E.R. Astuti, Classification and numbering on posterior dental radiography using support vector machine with mesiodistal neck detection, in: Joint International Conference on Soft Computing and Intelligent Systems and International Symposium on Advanced Intelligent Systems, 2012, pp. 432–435
[10] M. Hosntalab, R.A. Zoroofi, A.A. Tehrani-Fard, G. Shirani, Classification and numbering of teeth in multi-slice CT images using wavelet-Fourier descriptor, Int. J. CARS 5 (2010) 237–249.
[11] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural network, in: Advances in Neural Information Processing Systems, NIPS, vol. 25, 2012, pp. 1106–1114.
[12] M. Kallenberg, K. Petersen, M. Nielsen, A.Y. Ng, P. Diao, C. Igel, C.M. Vachon, K. Holland, R.R. Winkel, N. Karssemeijer, M. Lillholm, Unsupervised deep learning applied to breast density segmentation and mammographic risk scoring, IEEE Trans. Med. Imaging 35 (2016) 1322–1331.
[13] A.A.A. Setio, F. Ciompi, G. Litjens, P. Gerke, C. Jacobs, S.J. van Riel, M.M.W. Wille, M. Naqibullah, C.I. Sanchez, B. van Ginneken, Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks, IEEE Trans. Med. Imaging 35 (2016) 1160–1169.

[14] M. Anthimopoulos, S. Christodoulidis, L. Ebner, A. Christe, S. Mougiakakou, Lung pattern classification for interstitial lung diseases using a deep convolutional neural network, IEEE Trans. Med. Imaging 35 (2016) 1207–1216.
[15] H.R. Roth, L. Lu, J. Liu, J. Yao, A. Seff, K. Cherry, L. Kim, R.M. Summers, Improving computer-aided detection using convolutional neural networks and random view aggregation, IEEE Trans. Med. Imaging 35 (2016) 1170–1181.
[16] K.H. Cha, L. Hadjiiski, R.K. Samala, H.P. Chan, E.M. Caoili, R.H. Cohan, Urinary bladder segmentation in CT urography using deep-learning convolutional neural network and level sets, Med. Phys. 43 (2016) 1882–1896.
[17] A. Teramoto, H. Fujita, O. Yamamura, T. Tamaki, Automated detection of pulmonary nodules in PET/CT images: ensemble false-positive reduction using a convolutional neural network technique, Med. Phys. 43 (2016) 2821–2827.
[18] C.W. Wang, C.T. Huang, J.H. Lee, C.H. Li, S.W. Chang, M.J. Siao, T.M. Lai, B. Ibragimov, T. Vrtovec, O. Ronneberger, P. Fischer, T.F. Cootes, C. Lindner, A benchmark for comparison of dental radiography analysis algorithms, Med. Image Anal. 31 (2016) 63–76.
[19] G.E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, R.R. Salakhutdinov, Improving neural networks by preventing co-adaptiona of feature detectors, arXiv, 2012, 1207.0580.
[20] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A.C. Berg, L. Fei-Fei, ImageNet large scale visual recognition challenge, Int J. Comput. Vis. 115 (2015) 211–252.
[21] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, L. Jonathan, R. Girshick, S. Guadarrama, T. Darrell, Caffe: Convolutional architecture for fast feature embedding, arXiv, 2014, 1408.5093.