Received: 5 January 2019 | Revised: 9 April 2019 | Accepted: 2 May 2019

DOI: 10.1002/ima.22337

RESEARCH ARTICLE

WILEY

Teeth category classification via seven-layer deep convolutional neural network with max pooling and global average pooling

Zhi Li[1] | **Shui-Hua Wang**[2,3] | **Rui-Rui Fan**[4] | **Gang Cao**[1] | **Yu-Dong Zhang**[2,5] | **Ting Guo**[4]

[1]Department of Stomatology, Jinling Hospital, School of Medicine, Nanjing University, Nanjing, Jiangsu, China

[2]School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo, Henan, China

[3]Department of Electrical Engineering, The City College of New York, CUNY, New York, New York

[4]Nanjing Stomatological Hospital, Medical School of Nanjing University, Nanjing, Jiangsu, China

[5]Department of Informatics, University of Leicester, Leicester, UK

Correspondence
Yu-Dong Zhang, School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo, Henan, China.
Email: yudongzhang@ieee.org

Ting Guo, Nanjing Stomatological Hospital, Medical School of Nanjing University, Nanjing, Jiangsu, China.
Email: guoting_nj@126.com

Funding information
Henan Key Research and Development Project, Grant/Award Number: 182102310629; National key research and development plan, Grant/Award Number: 2017YFB1103202; National Postdoctoral Foundation, Grant/Award Number: 2016M593040; Natural Science Foundation of Jiangsu Province, Grant/Award Number: BK20161389; Open fund for Jiangsu Key Laboratory of Advanced Manufacturing Technology, Grant/Award Number: HGAMTL-1703; Open Fund of Guangxi Key Laboratory of Manufacturing System & Advanced Manufacturing Technology, Grant/Award Number: 17-259-05-011K; Open fund of Key Laboratory of Guangxi High Schools Complex System and Computational Intelligence, Grant/Award Number: 2016CSCI01; Six talent peaks project in Jiangsu Province, Grant/Award Number: 2016-WSW-093; the National Natural Science Foundation of China, Grant/Award Number: 81500872; Young Medical Talent Foundation of Jiangsu Province, Grant/Award Number: QNRC2016906

Abstract

Accurately classify teeth category is important in further dental diagnosis. Analyzing huge dental data, that is, identifying the teeth category, is often a hard task. Current automatic methods are based on computer vision and deep learning approaches. In this study, we aimed to classify the teeth category into four classes: incisor, canine, premolar, and molar. Cone beam computed tomography was used to collect the data. We proposed a seven-layer deep convolutional neural network with global average pooling to identify teeth category. Data augmentation method was used to enlarge the size of training dataset. The results showed the sensitivities of incisor, canine, premolar, and molar teeth are 88%, 86%, 84%, and 90%, respectively. The average sensitivity is 87.0%. We validated max pooling gives better results than average pooling. Our method is better than three state-of-the-art approaches.

KEYWORDS

cone beam computed tomography, convolutional neural network, data augmentation, deep convolutional neural network, deep learning, global average pooling, max pooling

Zhi Li and Shui-Hua Wang contributed equally to this study.

1 | INTRODUCTION

Teeth are vital parts of the human body. Teeth are not only chew food and assist with pronounce, but also have a great influence on the beauty of human face. According to the morphological function of the teeth, it can be divided into four categories: incisor (incisor, incisor), canine, premolar (first premolar, second premolar), molar (first molar, second molar, third molars).[1] However, teeth are under risk to caries and periodontal disease without proper protection, which can deeply threaten human health and life quality.[2] Cone Beam Computed Tomography (CBCT) can produce high-quality images for hard tissues, especially dental tissues, and reveal carious or periodontal lesions that are not directly visible to naked eyes as a noninvasive diagnostic method.[3] With CBCT, we can observe the tooth shape, size, location, and the relationship with the adjacent teeth at any angle, with low radiation and high spatial resolution. Thus, CBCT achieves ubiquity in the diagnosis of oral diseases.

As the population and related medical needs increase, analyzing huge clinical data is often a hard task. This requires an auxiliary means to reduce the burden of physicians. Machine learning technology can deal with huge CBCT data and obtain computerized quantitative features or areas, which can provide references and theoretical basis for the diagnosis of medical workers.[4] It converts digital medical images into data, these data and experience can be used to optimize the performance of computer programs. With high-throughput calculations, many quantitative features can be quickly extracted from digital images. Then, diagnostic results can be derived by quantitative image analysis based on preset algorithms and relevant knowledge rules.

Teeth classification can be solved by many techniques. Carmody, McGrath[5] used machine classification technique to identify diseases of periapical region. Their machine classified image is with 84% accuracy. Veeraprasit and Phimoltares[6] used a hybridization of local and global features. They provided 25 subjects, which were classified into 25 classes according to the teeth picture. Quinn[7] proposed a Haar wavelet transform (HWT) method that solves exactly the same task as in this paper. Their method achieves an overall accuracy of 81.83%. Let us view medical images not limited to teeth images, Khan, Suleman[8] used a genetic optimization algorithm (GOA) approach. Vijayalakshmi and Bharathi[9] employed legendre moments (LM). Han [10] utilized wavelet Renyi entropy to identify alcoholism patients.

Above methods obtained promising results; nevertheless, those methods need to manually design features, a procedure like selecting biomarkers. The feature design is tedious. In this study, we try to apply a new convolutional neural network (CNN) method to implement the teeth classification in cross-section image of CBCT. CNN is the most successful tool in machine learning and gained successes in many fields, including remote-sensing image segmentation,[11] fruit category classification,[12] alcoholism detection,[13] tea type classification,[14] and so forth. Those papers reported CNN had gained significant improvement compared to conventional machine learning approaches. The most advantage of CNN lies in it can automatically generate the features suitable for the cognate tasks, while traditional computer vision methods need to manually design and test features or their clusters.

In this study, we aimed to make a tentative application using CNN to identify teeth category, by which teeth are recognized as four categories: incisor, canine, premolar, and molar. Our experiments results show that this proposed method is superior to state-of-the-art teeth identification methods.

2 | DATASET AND DATA AUGMENTATION

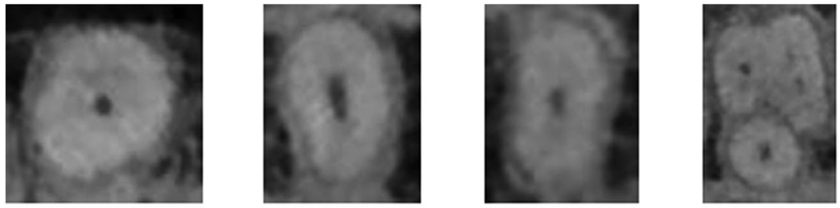
In our experiment, the CBCT images of teeth are used for reducing the damage to the human body in the process of imaging. In total, we have collected from local hospitals a 400-image dataset, which contains 100 incisors, 100 canines, 100 premolars, and 100 M. Figure 1 shows the samples of our dataset. Each sample will be converted to gray-level image, discarding color information. Afterward, those images were resampled to the size of 64×64 in sake of ease of following procedures.

The dataset is divided into training set and test set by hold out method. The training set contains 200 images, with 50 images for each type. The test set also contains 200 images, with 50 images for each type. Data augmentation (DA) method[15] was used over training set to create "*fake*" training samples, in order to avoid overfitting.

The first DA method was image rotation. The rotation angle θ was set from $-15°$ to $15°$ in step of $1°$. The second DA method was gamma correction. The gamma-value r varied from 0.7 to 1.3 with step of 0.02. The third DA method was noise injection. The zero-mean Gaussian noise with variance of 0.01 was employed. Thirty new noise-contaminated images were created for each original image. The fourth DA method used random translation by 30 times for each original image. The fifth DA method was scaling. The scaling factor s varied from 0.7 to 1.3 with step of 0.02. The final DA method was random affine transform, in which 30 new randomly affined images were created for every original image. Table 1 lists the setting of data augmentation.

In total, 180 new images were generated for each original image in training set. The original training set and augmented training set were combined, and now we have a 36 200-image dataset for training. The evaluation was performed over the test set, as shown in Table 2.

FIGURE 1 Samples of our dataset



(A) Incisor (B) Canine (C) Premolar (D) Molar

TABLE 1 Setting of data augmentation

DA Approach	Parameters
Image rotation	Rotation angle $\theta = -15{:}1{:}15$
Gamma correction	Gamma value $r = 0.7{:}0.02{:}1.3$
Noise injection	Variance = 0.01
Random translation	30 times
Scaling	Scaling factor $s = 0.7{:}0.02{:}1.3$
Random affine	30 times

TABLE 2 Dataset and data augmentation

Set	Incisor	Canine	Premolar	Molar	Total
Original Training	50	50	50	50	200
Augmented Training	9000	9000	9000	9000	36 000
Test	50	50	50	50	200

3 | METHODOLOGY

The CNN is composed of (a) alternating layers of convolution and pooling, serving as feature extraction; and (b) repeating layers of fully connected layers (FCL), serving as classification.[16] Different from conventional machine learning (ML) methods, the CNN learned the features while traditional ML methods design features manually. Figure 2 presents the flowchart of CNN.

3.1 | Convolution layer

In CNN, the convolution layer carries out a 2D convolution along width and height direction, for given 3D input and 3D filter. The convolution operation is not performed along depth (ie, channel) direction, because the 3D input and 3D filter have the same depth.[17]

Assume we have N filters with different learnable weights, size of each filter is $X \times Y \times Q$, here X and Y represent the height and width of each filter, Q the depth of channel. Assume the size of the input from previous layer is $J \times K \times Q$, here J and K represent the height and width of the input. The convolutional layer operates as shown in Figure 3.

Suppose the stride size is Z, the padding size at each margin is G. The height O and width P of output can be gained as

$$O = 1 + \frac{J - X + 2G}{Z} \tag{1}$$

$$P = 1 + \frac{K - Y + 2G}{Z} \tag{2}$$

The output is also 3D with size of $O \times P \times N$. The output will pass through a rectified linear unit (ReLU) with function of

$$\mathrm{ReLU}(t) = \begin{cases} t & t \geq 0 \\ 0 & t < 0 \end{cases} \tag{3}$$

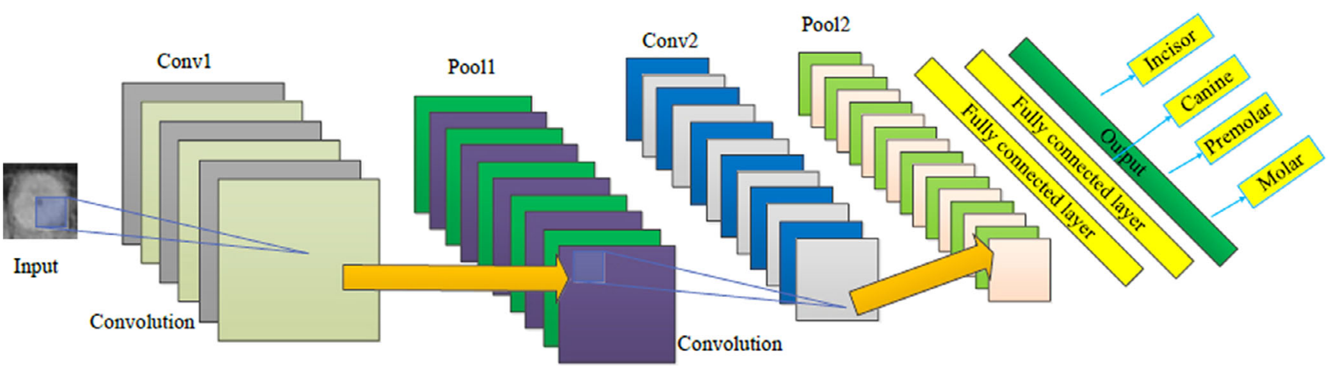


FIGURE 2 Flowchart of CNN. CNN, convolutional neural network [Color figure can be viewed at wileyonlinelibrary.com]

3.2 | Pooling layer

The next is pooling, which helps the activation map less sensitive to precise locations and not vulnerable to slight translation. Two common pooling techniques were widely used. Suppose the pooling region is C, average pooling (AP) is defined as

$$\mathrm{AP} = \frac{\sum C}{|C|} \tag{4}$$

and max pooling (MP) is defined as

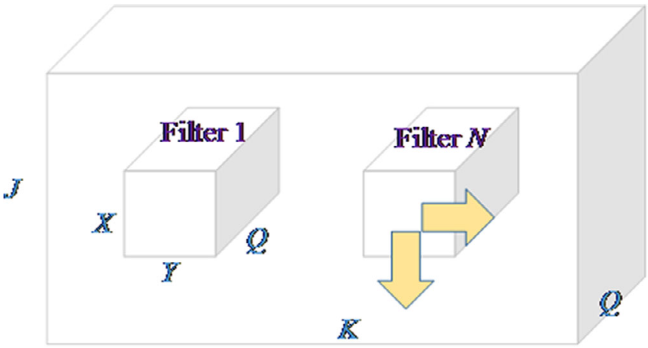


FIGURE 3 The operation done by convolution layer [Color figure can be viewed at wileyonlinelibrary.com]

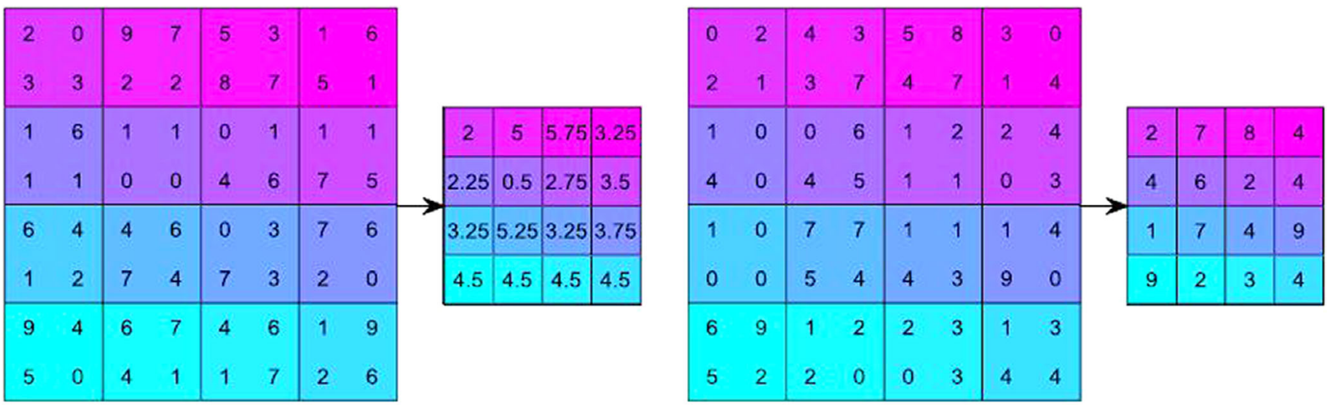
$$\mathrm{MP} = \max(C) \tag{5}$$

Figure 4a-b shows two toy examples of AP and MP, respectively. Note that global average pooling is an extreme of average pooling that can reduce a tensor with size of $w \times w \times d$ to $1 \times 1 \times d$. It just simples take the average of all w^2 values in each channel.

3.3 | Fully connected layer

Finally, FCL and softmax layer serve as role of classification, which the same as conventional deep learning.[18] FCL receives input from all previous layer and thus densely connected.[19] Each FCL layer has a weight matrix W, bias vector b, and an activation function a. Figure 5 shows the structure of 10-5-3 FCL layers.

The dropout technique is an important technique associated with FLC. It can freeze a portion of neurons and only train the survivor neurons within one training iteration, and randomly assign the freezing/survivor neurons in next iteration. In this study, we did not use dropout technique, since our neural network only contains two small FLC, as shown in Table 3.



(A) AP (B) MP

FIGURE 4 Toy examples of two pooling methods [Color figure can be viewed at wileyonlinelibrary.com]

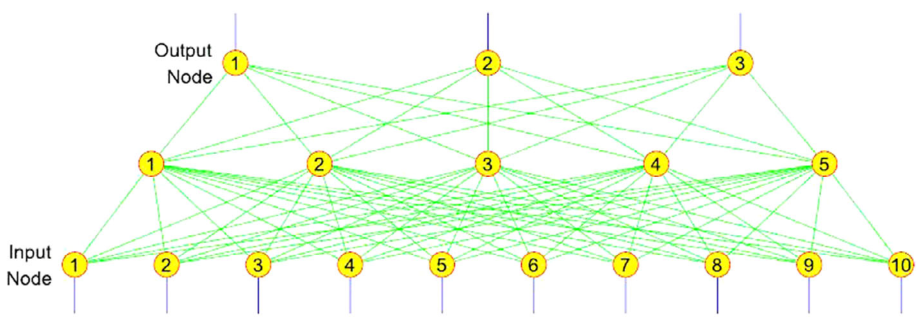


FIGURE 5 Illustration of 10-5-3 FCL layers. FCL, fully connected layer [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 3 Properties of each layer

Index of layer	Properties	Activation map
Input		64 × 64 × 1
Conv_MP_1	filter = 5, No. of filters = 32, stride = 3, padding = 2	22 × 22 × 32
Conv_MP_2	filter = 3, No. of filters = 64, stride = 3, padding = 1	8 × 8 × 64
Conv_MP_3	filter = 3, No. of filters = 128, stride = 1, padding = 1	8 × 8 × 128
Conv_MP_4	filter = 3, No. of filters = 256, stride = 1, padding = 1	8 × 8 × 256
Conv_MP_5	filter = 3, No. of filters = 512, stride = 1, padding = 0	6 × 6 × 512
GAP		1 × 1 × 512
FCL_1	No. of neurons = 100	1 × 1 × 100
FCL_2	No. of neurons = 2	1 × 1 × 2

4 | EXPERIMENT, RESULTS, AND DISCUSSIONS

4.1 | Data augmentation results

The original image is in Figure 1A. After preprocessing, Figure 6 shows the 180 new data augmentation results from six types of augmentation methods.

4.2 | Neural network structure

The data augmented training set is submitted to a seven-layer deep convolutional neural network. Here the number of layers only counts those layers associated with learnable weights/biases. ReLU layer, pooling layer, and softmax layer are not counted. We have in total five convolution layers and two FCL.

Their characteristics and output activation maps are shown in Table 3. Here every convolution layer was followed by a MP layer. After passing through all five convolution layers, we used a global AP layer to shrink the activation map to 1 × 1. Then, two FCL with 100 and two neurons are attached, respectively. Figure 7 shows the structure of this proposed seven-layer deep neural network.

4.3 | Confusion matrix

The results over the 200-image test set were shown in the form of a confusion matrix. Figure 8 lists the confusion matrix using MP, where C_1, C_2, C_3, and C_4 represents incisor, canine, premolar, and molar, respectively. AP performed worse than MP. Here we can observe from the first row that 44 of 50 incisors were identified, but three were wrongly identified as canine and other three were wrongly identified as premolar. The sensitivities of each class are 88%, 86%, 84%, and 90%, respectively. The overall accuracy (ie, average sensitivity) is 87.0%.

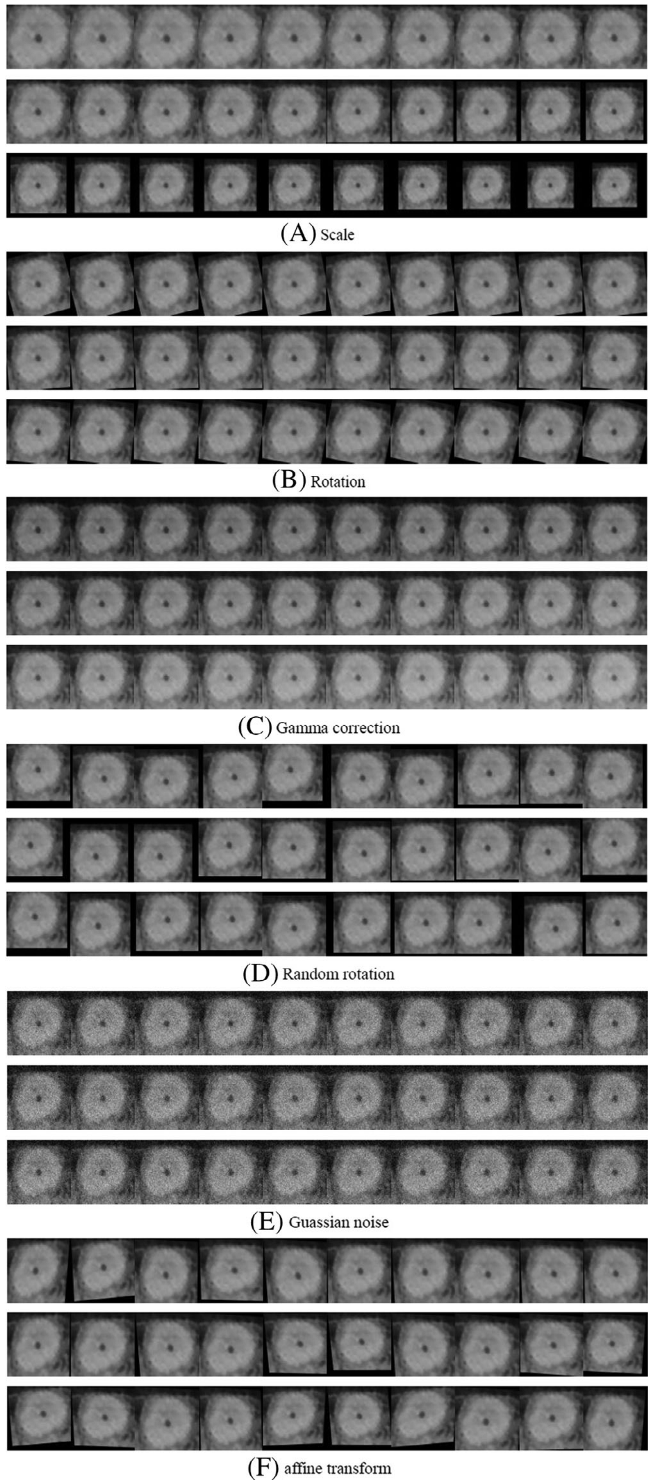


FIGURE 6 Data augmentation results

4.4 | Max pooling vs average pooling

If we replaced all the five MP layers with five corresponding average pooling layers, the overall accuracy results will be

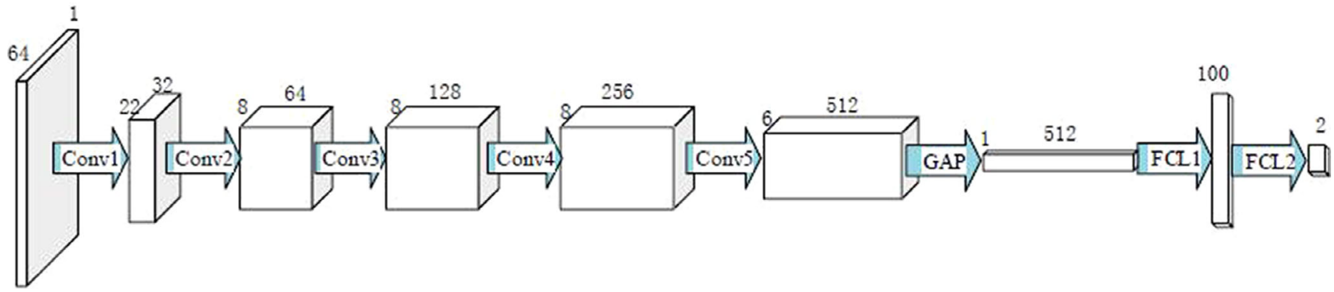


FIGURE 7 Structure of this proposed 7-layer deep neural network [Color figure can be viewed at wileyonlinelibrary.com]

	C_1	C_2	C_3	C_4	Sen
C_1	44	3	3	0	88.0%
C_2	2	43	3	2	86.0%
C_3	0	3	42	5	84.0%
C_4	1	0	4	45	90.0%
Prc	93.6%	87.8%	80.8%	86.5%	Acc 87.0%

FIGURE 8 Confusion matrix result of our method [Color figure can be viewed at wileyonlinelibrary.com]

	C_1	C_2	C_3	C_4	Sen
C_1	40	3	3	4	80.0%
C_2	3	43	1	3	86.0%
C_3	1	2	42	5	84.0%
C_4	2	0	3	45	90.0%
Prc	87.0%	89.6%	85.7%	78.9%	Acc 85.0%

FIGURE 9 Confusion matrix result of replacing max pooling with average pooling [Color figure can be viewed at wileyonlinelibrary.com]

worsened to merely 85.0%, as shown in Figure 9. This result clearly shows why MP is better than AP.

4.5 | Comparison to state-of-the-art approaches

In the final experiment, we compared this seven-layer CNN method with state-of-the-art approaches, including GOA,[8] LM,[9] and HWT.[7] All the results were obtained on our dataset. The results are listed in Table 4 and Figure 10. It indicates that our 7-layer CNN outperforms those two methods significantly. F1 score is not calculated since it is for measuring a binary classification, and ours is a four-class classification.

TABLE 4 Comparison to state-of-the-art approaches in terms of sensitivities

Approach	C_1	C_2	C_3	C_4	Average
GOA (2017)[8]	76.0	78.0	76.0	84.0	78.5
LM (2016)[9]	78.0	82.0	74.0	84.0	79.5
HWT (2018)[7]	82.67	81.00	81.33	82.33	81.83
7-layer CNN (Our)	88.0	86.0	84.0	90.0	87.0

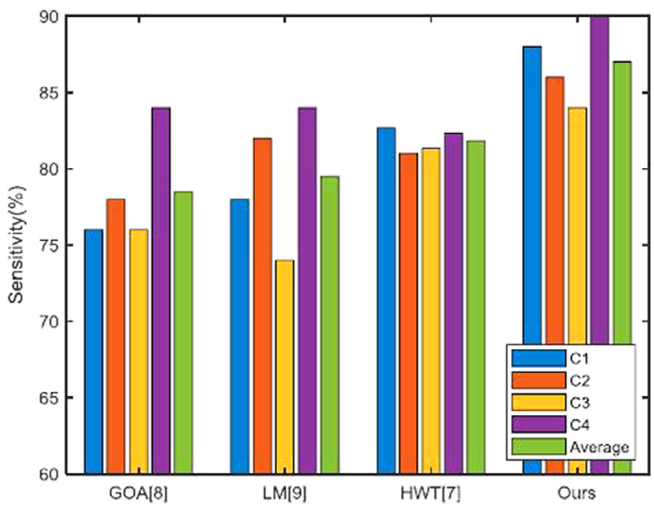


FIGURE 10 Comparison of our method with three state-of-the-art methods [Color figure can be viewed at wileyonlinelibrary.com]

Machine classification technique allows medical workers to process huge image data of CBCT and recognize quantitative features or areas, which provide reference and theory for diagnosis, and improve diagnosis efficiency, accuracy, and repeatability. In the future, artificial intelligence may help medical workers in prognosticating the disease and guiding in clinical diagnosis and treatment.

5 | CONCLUSIONS

This article presents a study using seven-layer convolutional neural network to classify teeth category. The results showed our method is superior to both GOA,[8] LM,[9] and HWT.[7] It indicates the CNN is a promising method in identifying teeth category, with an overall accuracy of 87.0%.

In the future, we shall try to collect more data, and use advanced variants of CNN. Transfer learning may be used, since it is a pre-trained deep neural network. Besides, autoencoder is also a successful tool in deep learning, which will be tested in future studies.

The CNN method may be applied to other fields, such as face recognition,[20-22] hearing loss,[23] gingivitis detection, and other stomatological disease grading.

ACKNOWLEDGMENT

The authors sincerely appreciate the support of Young Medical Talent Foundation of Jiangsu Province (QNRC2016906), and Six talent peaks project in Jiangsu Province (2016-WSW-093), Natural Science Foundation of China (61602250), Natural Science Foundation of Jiangsu Province (BK20150983), National key research and development plan (2017YFB1103202), Henan Key Research and Development Project (182102310629), Open fund for Jiangsu Key Laboratory of Advanced Manufacturing Technology (HGAMTL-1703), Open fund of Key Laboratory of Guangxi High Schools Complex System and Computational Intelligence (2016CSCI01), Open Fund of Guangxi Key Laboratory of Manufacturing System & Advanced Manufacturing Technology (17-259-05-011K).

ORCID

Zhi Li https://orcid.org/0000-0002-2795-0693
Yu-Dong Zhang https://orcid.org/0000-0002-4870-1493

REFERENCES

1. Gezgin O, Botsali MS. Evaluation of teeth development in unilateral cleft lip and palate patients in mixed dentition by using medical image control systems. *Niger J Clin Pract*. 2018;21(2):156-162.
2. Martins JNR, Marques D, Francisco H, Carames J. Gender influence on the number of roots and root canal system configuration in human permanent teeth of a Portuguese subpopulation. *Quintessence Int*. 2018;49(2):103-111.
3. Silva DDE, Campos CN, Carvalho ACP, Devito KL. Diagnosis of Mesiodistal vertical root fractures in teeth with metal posts: influence of applying filters in cone-beam computed tomography images at different resolutions. *J Endod*. 2018;44(3):470-474.
4. Baikejiang R, Zhang W, Zhu DW, et al. Kernel-based anatomically-aided diffuse optical tomography reconstruction. *Biomed Phys Eng Exp*. 2017;3(5):13 Article ID. Unsp 055002.
5. Carmody DP, McGrath SP, Dunn SM, van der Stelt PF, Schouten E. Machine classification of dental images with visual search. *Acad Radiol*. 2001;8(12):1239-1246.
6. Veeraprasit, S, Phimoltares S. Hybrid feature-based teeth recognition system. Paper presented at: International Conference on Imaging Systems and Techniques. 2011. Penang, Malaysia: IEEE. p. 302–305
7. Quinn W. Teeth classification based on Haar wavelet transform and support vector machine. *Adv Comput Sci Res*. 2018;80:249-252.
8. Khan RA, Suleman T, Farooq MS, Rafiq MH, Tariq MA. Data mining algorithms for classification of diagnostic cancer using genetic optimization algorithms. *Int J Comput Sci Network Sec*. 2017;17(12):207-212.
9. Vijayalakshmi B, Subbiah Bharathi V. Classification of CT liver images using local binary pattern with Legendre moments. *Curr Sci*. 2016;110(4):687-691.
10. Han, L., Identification of Alcoholism based on wavelet Renyi entropy and three-segment encoded Jaya algorithm. Complexity, 2018. 2018: Article ID. 3198184
11. Zhao G. Polarimetric synthetic aperture radar image segmentation by convolutional neural network using graphical processing units. *J Real-Time Image Proc*. 2018;15(3):631-642.
12. Muhammad K. Image based fruit category classification by 13-layer deep convolutional neural network and data augmentation. *Multimed Tools Appl*. 2017;78:3613-3632. https://doi.org/10.1007/s11042-017-5243-3.
13. Lv Y-D. Alcoholism detection by data augmentation and convolutional neural network with stochastic pooling. *J Med Syst*. 2018;42(1):2.
14. Tang C. Twelve-layer deep convolutional neural network with stochastic pooling for tea category classification on GPU platform. *Multimed Tools Appl*. 2018;77(17):22821-22839.
15. Ahmad J, Muhammad K, Baik SW. Data augmentation-assisted deep learning of hand-drawn partially colored sketches for visual search. *PLoS ONE*. 2017;12(8):19, e0183838.
16. Gonzalez-Garcia A, Modolo D, Ferrari V. Do semantic parts emerge in convolutional neural networks? *Int J Comput Vis*. 2018;126(5):476-494.
17. Huang C. Multiple sclerosis identification by 14-layer convolutional neural network with batch normalization, dropout, and stochastic pooling. *Front Neurosci*. 2018;12:818.
18. Shin J, Park H, Paik J. Fire recognition using spatio-temporal two-stream convolutional neural network with fully connected layer-fusion. Paper presented at: 2018 IEEE 8th International Conference on Consumer Electronics - Berlin, R. Moeller and L. Ciabattoni, Editors. 2018, IEEE: Berlin, Germany.
19. Richter O, Wattenhofer R. Treeconnect: a sparse alternative to fully connected layers. Paper presented at: 2018 IEEE 30th International Conference on Tools with Artificial Intelligence. 2018, IEEE: Volos, Greece. p. 924–931.
20. Peng YL, Li LJ, Liu SG, Li J, Wang XL. Extended sparse representation-based classification method for face recognition. *Mach Vis Appl*. 2018;29(6):991-1007.
21. Zhang KY, Peng YL, Liu SG. Discriminative face recognition via kernel sparse representation. *Multimed Tools Appl*. 2018;77(24):32243-32256.
22. Peng YL, Li LP, Liu SG, Lei T. Space-frequency domain based joint dictionary learning and collaborative representation for face recognition. *Signal Process*. 2018;147:101-109.
23. Tang C, Lee E. Hearing loss identification via wavelet entropy and combination of Tabu search and particle swarm optimization. Paper presented at: 23rd International Conference on Digital Signal Processing (DSP). 2018. Shanghai, China: IEEE. p. 1–5

How to cite this article: Li Z, Wang S-H, Fan R-R, Cao G, Zhang Y-D, Guo T. Teeth category classification via seven-layer deep convolutional neural network with max pooling and global average pooling. *Int J Imaging Syst Technol*. 2019;29:577–583. https://doi.org/10.1002/ima.22337