Teeth category classification via 7-layer deep convolutional neural network with max pooling and global average pooling

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

1. Department of Stomatology, Jinling Hospital, School of Medicine, Nanjing University, 305 East Zhongshan Road, 210002 Nanjing, Jiangsu, China
2. School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo, Henan 454000, P R China
3. Nanjing Stomatological Hospital, Medical School of Nanjing University, Nanjing, Jiangsu, China
4. Department of Electrical Engineering, The City College of New York, CUNY, New York, NY 10031, USA
5. Department of Informatics, University of Leicester, Leicester, LE1 7RH, UK

Email: guoting_nj@126.com, yudongzhang@ieee.org

Abstract: Accurately classify teeth category is important in further dental diagnosis. Analyzing huge dental data, i.e., 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 7-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: deep convolutional neural network; cone beam computed tomography; deep learning; max pooling; data augmentation; convolutional neural network; global average pooling;

1 Introduction

Teeth are vital parts of the human body. Teeth can 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 non-invasive 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 image images. Then, diagnostic results can be derived by quantitative image analysis based on preset algorithms and relevant knowledge rules.

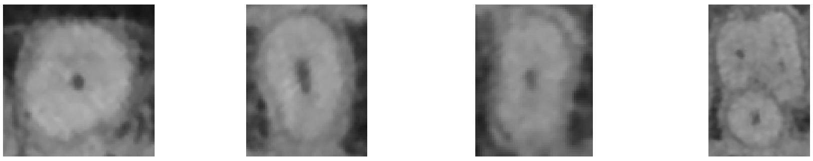
Teeth classification can be solved by many techniques. Carmody, McGrath, Dunn, van der Stelt and Schouten (2001) [5] used machine classification technique to identify diseases of periapical region. Their machine classified image are with 84% accuracy. Veeraprasit and Phimoltares (2011) [6] used a hybridization of local and global features. They provided 25 subjects, which were classified into 25 classes according to the teeth picture. This year, Quinn (2018) [7] proposed a Haar wavelet transform (HWT) method that solves the 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, Farooq, Rafiq and Tariq (2017) [8] used a Genetic Optimization Algorithm (GOA) approach. Vijayalakshmi and Bharathi (2016) [9] employed Legendre moments (LM). Han (2018) [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], etc. 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 molars. Figure 1 shows the samples of our dataset. Each sample will be converted to gray-level image, discarding color information. Afterwards, those images were resampled to the size of 64x64 in sake of ease of following procedures.



(a) Incisor (b) Canine (c) Premolar (d) Molar

Figure 1 Samples of our dataset

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 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. 30 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.

Table 1 Setting of data augmentation

DA Approach	Parameters
Image rotation	Rotation angle θ = -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

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.

Table 2 Dataset and data augmentation

Set	Incisor	Canine	Premolar	Molar	Total
Original Training	50	50	50	50	200
Augmented Training	9,000	9,000	9,000	9,000	36,000
Test	50	50	50	50	200

3 Methodology

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

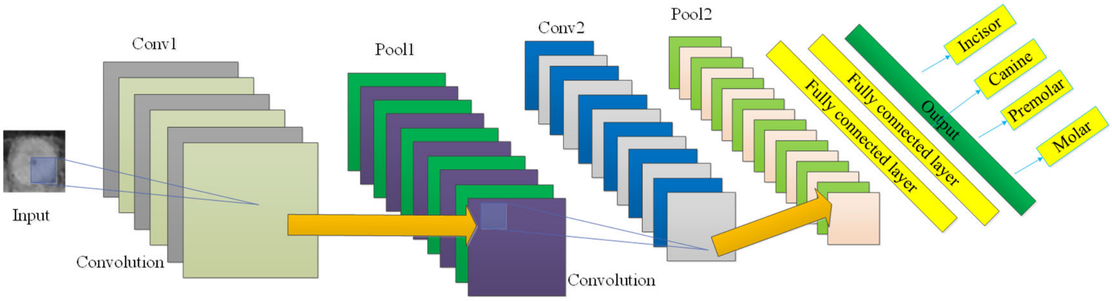


Figure 2 Flowchart of CNN

3.1 Conv 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 (i.e., 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.

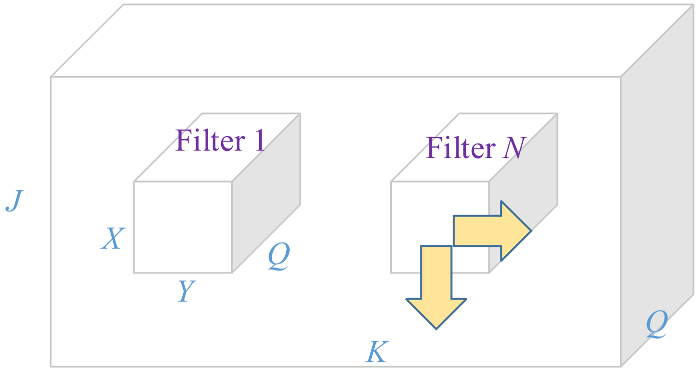


Figure 3 The operation done by convolution layer

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

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

3.2 Pooling Layer

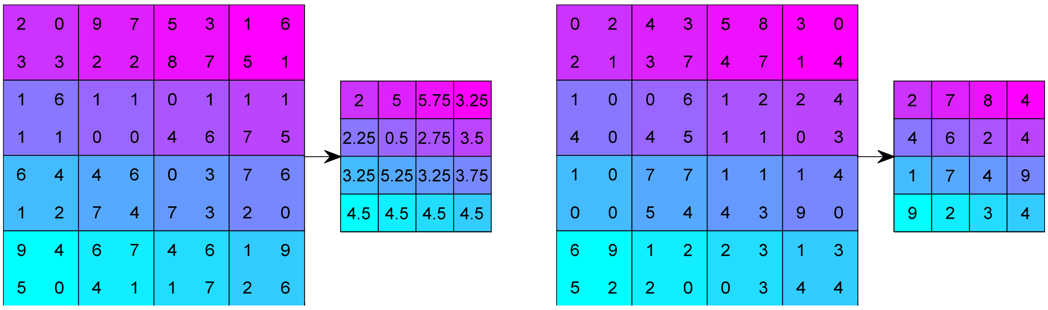
The next is pooling, which helps the activation map less sensitive to precise locations and not vulnerable to slight translation. Two common pooling techniques were widely used. Suppose the pooling region is C, average pooling (AP) is defined as

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

and max pooling (MP) is defined as

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

Figure 4(a-b) shows two toy examples of average pooling and max pooling, 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.



2	0	9	7	5	3	1	6
3	3	2	2	8	7	5	1
1	6	1	1	0	1	1	1
1	1	0	0	4	6	7	5
6	4	4	6	0	3	7	6
1	2	7	4	7	3	2	0
9	4	6	7	4	6	1	9
5	0	4	1	1	7	2	6

2	5	5.75	3.25
2.25	0.5	2.75	3.5
3.25	5.25	3.25	3.75
4.5	4.5	4.5	4.5

(a) AP

0	2	4	3	5	8	3	0
2	1	3	7	4	7	1	4
1	0	0	6	1	2	2	4
4	0	4	5	1	1	0	3
1	0	7	7	1	1	1	4
0	0	5	4	4	3	9	0
6	9	1	2	2	3	1	3
5	2	2	0	0	3	4	4

2	7	8	4
4	6	2	4
1	7	4	9
9	2	3	4

(b) MP

Figure 4 Toy examples of two pooling methods

3.3 Fully-connected Layer

Finally, fully connected layer (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.

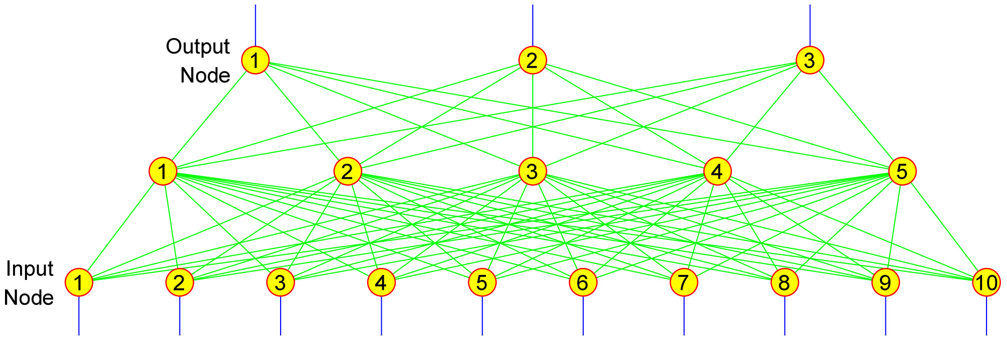


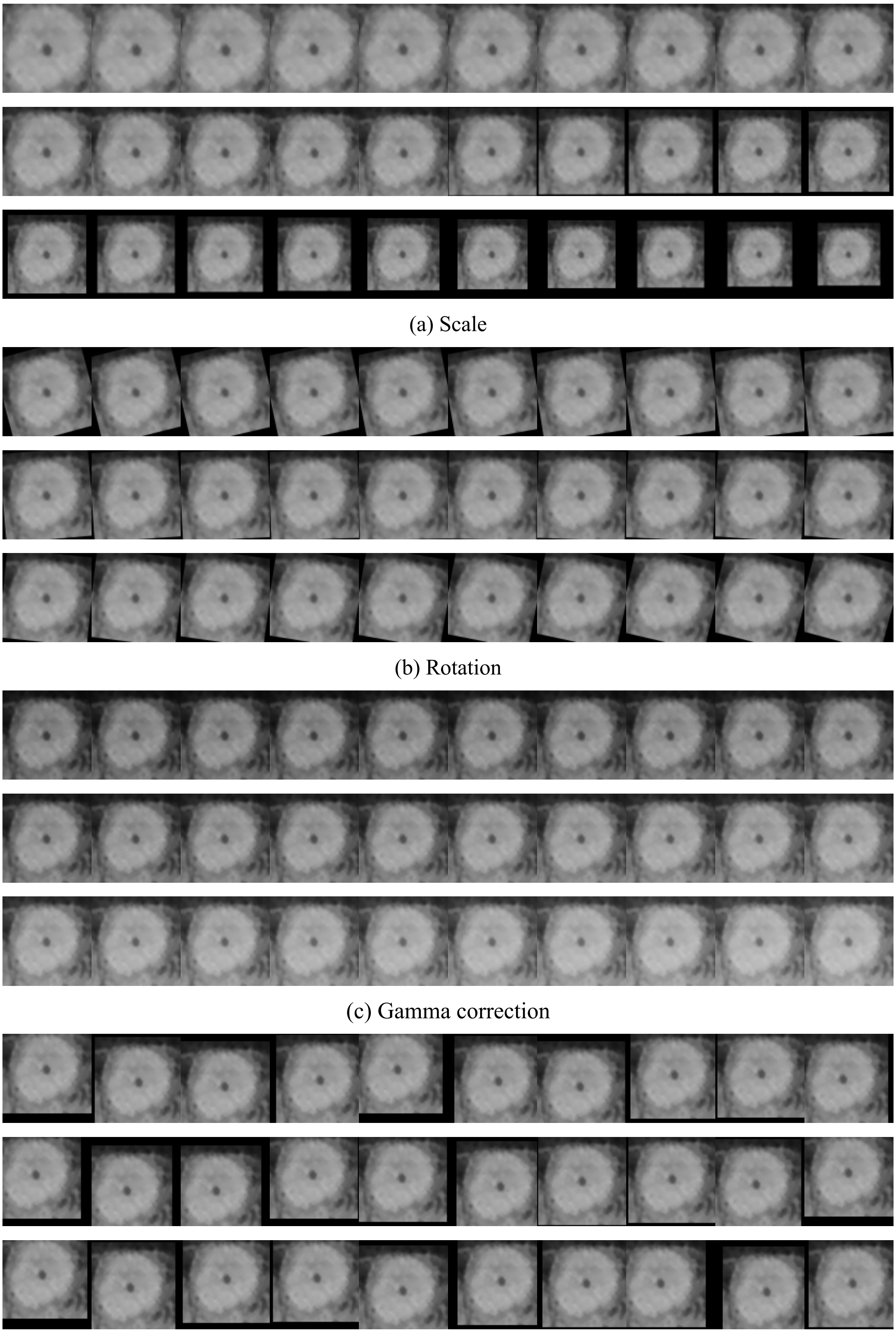
Figure 5 Illustration of 10-5-3 FCL layers

The dropout technique is an important technique associated with fully-connected layer. 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 fully-connected layer, as shown in Table 3.

4 Experiment, Results, and Discussions

4.1 Data Augmentation Results

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

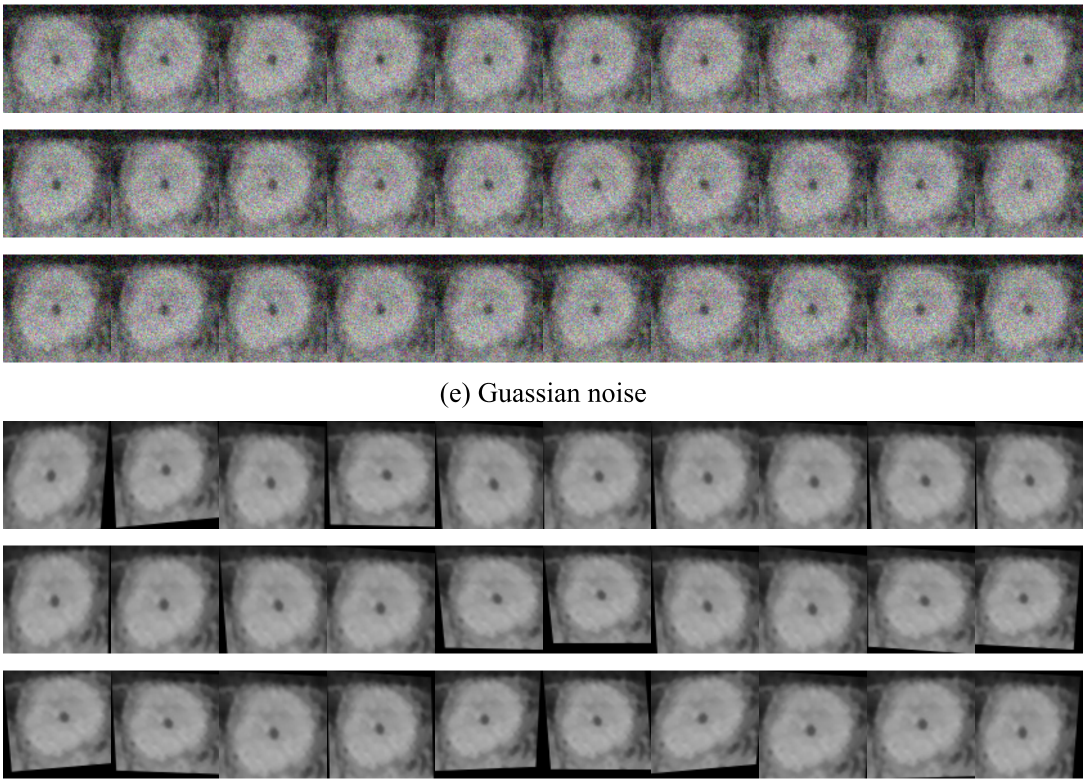


(a) Scale

(b) Rotation

(c) Gamma correction

(d) Random rotation



(e) Guassian noise

(f) affine transform

Figure 6 Data augmentation results

4.2 Neural Network Structure

The data augmented training set are submitted to a 7-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 fully connected layers.

Their characteristics and output activation maps are shown in Table 3. Here every conv layer was followed by a max pooling layer. After passing through all five conv layers, we used a global average pooling layer to shrink the activation map to 1x1. Then, two fully-connected layers with 100 and 2 neurons are attached, respectively. Figure 7 shows the structure of this proposed 7-layer deep neural network.

Table 3 Properties of each layer

Index of Layer	Properties	Activation Map
Input		64x64x1
Conv_MP_1	filter = 5, No. of filters = 32, stride = 3, padding = 2	22x22x32
Conv_MP_2	filter = 3, No. of filters = 64, stride = 3, padding = 1	8x8x64
Conv_MP_3	filter = 3, No. of filters = 128, stride = 1, padding = 1	8x8x128
Conv_MP_4	filter = 3, No. of filters = 256, stride = 1, padding = 1	8x8x256
Conv_MP_5	filter = 3, No. of filters = 512, stride = 1, padding = 0	6x6x512
GAP		1x1x512
FCL_1	No. of neurons = 100	1x1x100

FCL_2	No. of neurons = 2	1x1x2

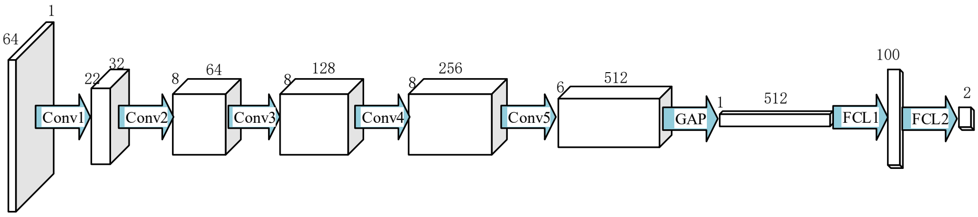
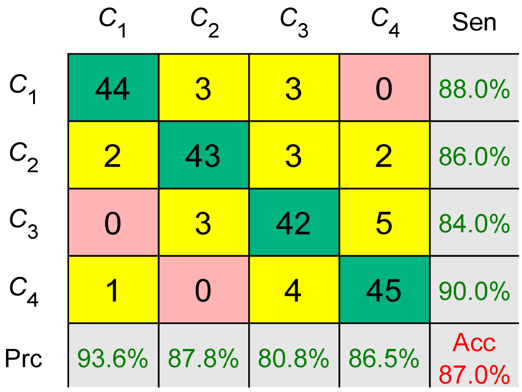


Figure 7 Structure of this proposed 7-layer deep neural network

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 max pooling, where C_1, C_2, C_3, and C_4 represents incisor, canine, premolar, and molar, respectively. Average pooling performed worse than max pooling. Here we can observe from the first row that 44 out of 50 incisors were identified, but 3 were wrongly identified as canine and other 3 were wrongly identified as premolar. The sensitivities of each classes are 88%, 86%, 84%, and 90%, respectively. The overall accuracy (i.e., average sensitivity) is 87.0%.

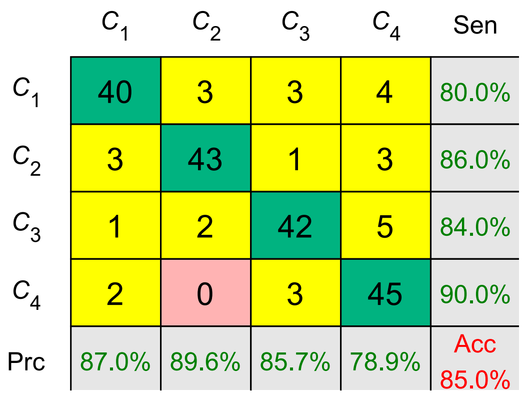


	C_1	C_2	C_3	C_4	Sen
C_1	44	3	3	0	88.0%
C_2	2	43	3	2	86.0%
C_3	0	3	42	5	84.0%
C_4	1	0	4	45	90.0%
Prc	93.6%	87.8%	80.8%	86.5%	Acc 87.0%

Figure 8 Confusion matrix result of our method

4.4 Max Pooling versus Average Pooling

If we replaced all the five max pooling layers with five corresponding average pooling layers, the overall accuracy results will be worsened to merely 85.0%, as shown in Figure 9. This result clearly shows why max pooling is better than average pooling.



	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

4.5 Comparison to State-of-the-art Approaches

In the final experiment, we compared this 7-layer CNN method with state-of-the-art approaches, including genetic optimization algorithm (GOA) [8], Legendre moment (LM) [9], and Haar wavelet transform (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

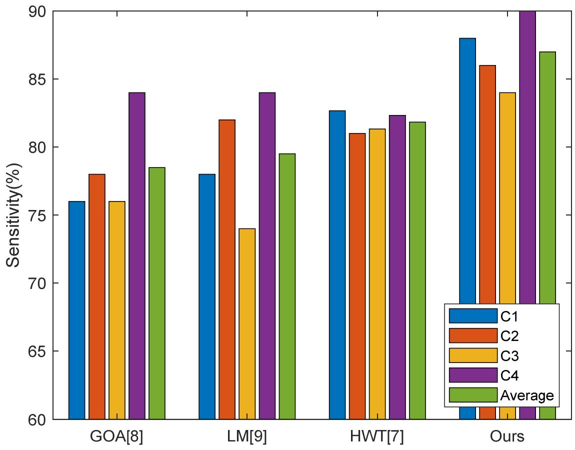


Figure 10 Comparison of our method with three state-of-the-art methods

Machine classification technique allows medical workers to process huge image data of CBCT and recognize quantitative features or areas, which provides 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 paper presents a study using 7-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).

References

1. Gezgin, O. and M.S. Botsali, *Evaluation of Teeth Development in Unilateral Cleft Lip and Palate Patients in Mixed Dentition by Using Medical Image Control Systems.* Nigerian Journal of Clinical Practice, 2018. **21**(2): p. 156-162
2. Martins, J.N.R., D. Marques, H. Francisco, and J. Carames, *Gender influence on the number of roots and root canal system configuration in human permanent teeth of a Portuguese subpopulation.* Quintessence International, 2018. **49**(2): p. 103-111
3. Silva, D.D.E., C.N. Campos, A.C.P. Carvalho, and K.L. Devito, *Diagnosis of Mesiodistal Vertical Root Fractures in Teeth with Metal Posts: Influence of Applying Filters in Cone-beam Computed Tomography Images at Different Resolutions.* Journal of Endodontics, 2018. **44**(3): p. 470-474
4. Baikejiang, R., W. Zhang, D.W. Zhu, A.M. Hernandez, S.A. Shakeri, G.B. Wang, J.Y. Qi, J.M. Boone, and C.Q. Li, *Kernel-based anatomically-aided diffuse optical tomography reconstruction.* Biomedical Physics & Engineering Express, 2017. **3**(5): p. 13: Article ID. Unsp 055002
5. Carmody, D.P., S.P. McGrath, S.M. Dunn, P.F. van der Stelt, and E. Schouten, *Machine classification of dental images with visual search.* Academic Radiology, 2001. **8**(12): p. 1239-1246
6. Veeraprasit, S. and S. Phimoltares. *Hybrid feature-based teeth recognition system.* in *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.* Advances in Computer Science Research, 2018. **80**: p. 249-252
8. Khan, R.A., T. Suleman, M.S. Farooq, M.H. Rafiq, and M.A. Tariq, *Data Mining Algorithms for Classification of Diagnostic*

Cancer Using Genetic Optimization Algorithms. International Journal of Computer Science and Network Security, 2017. **17**(12): p. 207-212

9. Vijayalakshmi, B. and V.S. Bharathi, *Classification of CT liver images using local binary pattern with Legendre moments.* Current Science, 2016. **110**(4): p. 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.* Journal of Real-Time Image Processing, 2018. **15**(3): p. 631-642
12. Muhammad, K., *Image based fruit category classification by 13-layer deep convolutional neural network and data augmentation*. Multimedia Tools and Applications, 2017, doi: 10.1007/s11042-017-5243-3.
13. Lv, Y.-D., *Alcoholism detection by data augmentation and convolutional neural network with stochastic pooling.* Journal of Medical Systems, 2018. **42**(1): Article ID. 2
14. Tang, C., *Twelve-layer deep convolutional neural network with stochastic pooling for tea category classification on GPU platform.* Multimedia Tools and Applications, 2018. **77**(17): p. 22821-22839
15. Ahmad, J., K. Muhammad, and S.W. Baik, *Data augmentation-assisted deep learning of hand-drawn partially colored sketches for visual search.* Plos One, 2017. **12**(8): p. 19: Article ID. e0183838
16. Gonzalez-Garcia, A., D. Modolo, and V. Ferrari, *Do Semantic Parts Emerge in Convolutional Neural Networks?* International Journal of Computer Vision, 2018. **126**(5): p. 476-494
17. Huang, C., *Multiple Sclerosis Identification by 14-Layer Convolutional Neural Network With Batch Normalization, Dropout, and Stochastic Pooling.* Frontiers in Neuroscience, 2018. **12**: Article ID. 818
18. Shin, J., H. Park, and J. Paik, *Fire Recognition Using Spatio-Temporal Two-Stream Convolutional Neural Network with Fully Connected Layer-Fusion*, in *2018 Ieee 8th International Conference on Consumer Electronics - Berlin*, R. Moeller and L. Ciabattoni, Editors. 2018, IEEE: Berlin, GERMANY.
19. Richter, O. and R. Wattenhofer, *TreeConnect: A Sparse Alternative to Fully Connected Layers*, in *2018 Ieee 30th International Conference on Tools with Artificial Intelligence*. 2018, IEEE: Volos, GREECE. p. 924-931.
20. Peng, Y.L., L.J. Li, S.G. Liu, J. Li, and X.L. Wang, *Extended sparse representation-based classification method for face recognition.* Machine Vision and Applications, 2018. **29**(6): p. 991-1007
21. Zhang, K.Y., Y.L. Peng, and S.G. Liu, *Discriminative face recognition via kernel sparse representation.* Multimedia Tools and Applications, 2018. **77**(24): p. 32243-32256
22. Peng, Y.L., L.P. Li, S.G. Liu, and T. Lei, *Space-frequency domain based joint dictionary learning and collaborative representation for face recognition.* Signal Processing, 2018. **147**: p. 101-109
23. Tang, C. and E. Lee. *Hearing loss identification via wavelet entropy and combination of Tabu search and particle swarm optimization*. in *23rd International Conference on Digital Signal Processing (DSP)*. 2018. Shanghai, China: IEEE. p. 1-5