IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received September 30, 2019, accepted October 19, 2019, date of publication October 30, 2019,
date of current version November 12, 2019.

*Digital Object Identifier 10.1109/ACCESS.2019.2950286*

# An Early Diagnosis of Oral Cancer based on Three-Dimensional Convolutional Neural Networks

**SHIPU XU[1,2], CHANG LIU[3], YONGSHUO ZONG[4], SIRUI CHEN[4], YIWEN LU[4], LONGZHI YANG[5], (Senior Member, IEEE), EDDIE Y. K. NG[6], YONGTONG WANG[4], YUNSHENG WANG[2], YONG LIU[2], WENWEN HU[2], AND CHENXI ZHANG[1]**

[1]Department of Software Engineering, Tongji University, Shanghai 201804, China
[2]Agricultural Information Institute of Science and Technology, Shanghai Academy of Agricultural Sciences, Shanghai 201403, China
[3]School of Information Engineering, Nanchang Hangkong University, Nanchang 330038, China
[4]Department of Computer Science, Tongji University, Shanghai 201804, China
[5]Department of Computer and Information Sciences, Northumbria University, Newcastle upon Tyne NE1 8ST, U.K.
[6]School of Mechanical and Aerospace Engineering, Nanyang Technological University, Singapore 639798

Corresponding author: Chenxi Zhang (xzhang2000@163.com)

This work was supported in part by the Shanghai Academy of Agricultural Sciences for the Program of Excellent Research Team under Grant 2017[B-09], and in part by the Shanghai Municipal Agricultural Commission Shanghai Agriculture Applied Technology Development Program, China, under Grant G2015 6-4-1.

**ABSTRACT** Three-dimensional convolutional neural networks (3DCNNs), a rapidly evolving modality of deep learning, has gained popularity in many fields. For oral cancers, CT images are traditionally processed using two-dimensional input, without considering information between lesion slices. In this paper, we established a 3DCNNs-based image processing algorithm for the early diagnosis of oral cancers, which was compared with a 2DCNNs-based algorithm. The 3D and 2D CNNs were constructed using the same hierarchical structure to profile oral tumors as benign or malignant. Our results showed that 3DCNNs with dynamic characteristics of the enhancement rate image performed better than 2DCNNS with single enhancement sequence for the discrimination of oral cancer lesions. Our data indicate that spatial features and spatial dynamics extracted from 3DCNNs may inform future design of CT-assisted diagnosis system.

**INDEX TERMS** 2DCNNs, 3DCNNs, CT images, spatial features, spatial dynamics extracted.

## I. INTRODUCTION

Oral cancer is a common type of cancer today, which occurs mostly in people over 40 years of age. At present, the early diagnosis methods of oral cancer include toluidine blue staining, tetracycline fluorescence detection, hematoporphyrin photometry and resonance isotope labeling. Currently, there is a necessary process in the diagnosis of oral cancer, which is to obtain oral CT (Computed Tomography) images.

The oral CT image sequence contains images of different levels of the body. Each slice is stacked and approximated by a reconstruction algorithm to form a three-dimensional volume object [1]–[3]. In the general image classification problem, convolutional neural network (CNN) can process directly on the original input image. Although it is powerful, it usually deals with two-dimensional images or RGB color images, which is limited by two-dimensional input [4]. If the oral CT image is also treated with two-dimensional slices, the three-dimensional related information between the lesion slices may be lost, and the three-dimensional CNN (3DCNN) is used to identify the early oral cancers with 3DCT image data, and the general two-dimensional volume. The neural network is then compared [5], [6].

At present, two-dimensional CNN (2DCNN) are widely used in deep learning, but some scholars have studied 3DCNN and applied them to the recognition of human behavior in video data [7]. In the literature, the authors have proposed a multi-modal 3DCNN feature extraction method which is applied to CT segmentation of brain tumor. This method can tell the difference between brain tumors of different patients at various modes, improving the accuracy of CT segmentation of brain tumor [8].

The associate editor coordinating the review of this manuscript and approving it for publication was Ying Song.

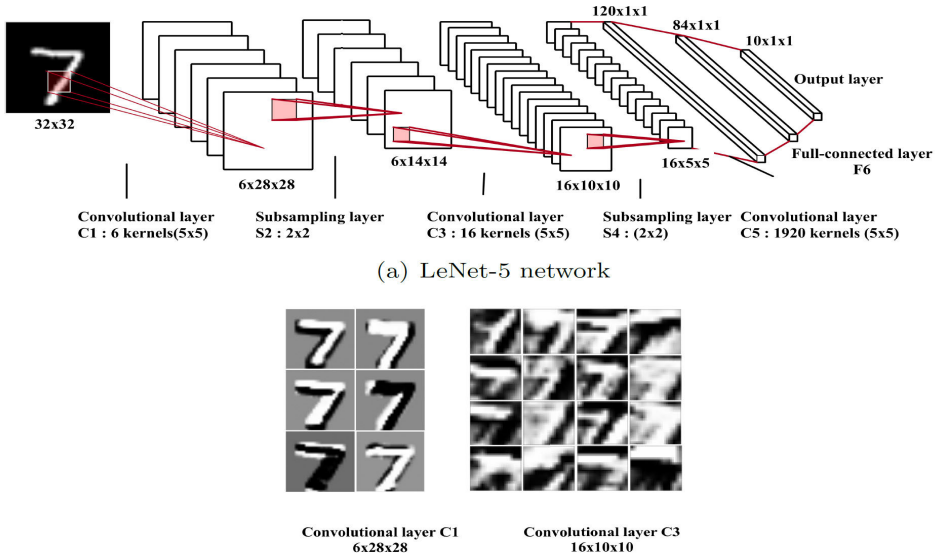


**FIGURE 1.** The architecture of the LeNet-5 network, which works well on digit classification task. (b) Visualization of features in the LeNet-5 network. Each layer's feature maps are displayed in a different block.

Moreover, brain micro-bleeds in MRI images can be automatically detected by 3DCNN. 3DCNN has also been used to automatically detect and segment the CT images of human liver organs. Finally, the images of human liver organs are finely segmented via the Graph Cut algorithm [9], [10].

In this paper, 2DCNN and 3DCNN structures were constructed to judge the benign and malignant oral cancers. In order to compare the advantages and disadvantages of the two methods, the convolutional network used in the experiment has the same hierarchical structure. Because of the individual differences in tumor morphology, a large number of samples are generated by means of data expansion such as translational rotation mirroring, which are used to train these two CNNs.

## II. CONSTRUCTION OF CNN

### A. CONSTRUCTION OF 2DCNN

CNN is a feed-forward neural network with convolutional computation and deep structure. It is one of the representative algorithms of deep learning. CNN have the ability of representation learning, and can get shift-invariant classification of input information according to their hierarchical structure. Therefore, it is also called 'Shift-Invariant Artificial Neural Networks (SIANN)'.

The study of CNN began in the 1980s and 1990s. The time delay network and LeNet-5 were the earliest CNN. After the 21st century, with the introduction of deep learning theory and numerical calculations, and with the improvement of numerical calculation equipment, the CNN has been rapidly developed and applied to fields such as computer vision and natural language processing.

The CNN constructs the visual perception mechanism of the creature, which can perform supervised learning and unsupervised learning. The convolutional kernel parameter sharing and the sparseness of the inter-layer connection in the hidden layer enable the CNN to smaller computational tasks for grid-like topology features such as pixel and audio, have a stabilizing effect and have no additional feature engineering requirements for the data.

LeNet-5 is a CNN applied to image classification problems. Its learning goal is to identify and distinguish 0-9 from a series of handwritten digits represented by $32 \times 32 \times 1$ grayscale images. The hidden layer of LeNet-5 consists of two convolutional layers, two pooled layers, and two fully connected layers, constructed as follows [11]:

(1) $(3 \times 3) \times 1 \times 6$ convolutional layer (step size 1, no padding), $2 \times 2$ mean pooling (step size 2, no padding), **TANH** excitation function

(2) $(5 \times 5) \times 6 \times 16$ convolutional layer (step size 1, no padding), $2 \times 2$ mean pooling (step size 2, no padding), **TANH** excitation function

(3) 2 fully connected layers, the number of neurons is 120 and 84, the process is as shown in **FIGURE 1**.

The 2DCNN hierarchy used in this paper is slightly more complicated than the LeNet network. There are four pairs of convolution pooling layers, and the regularization layer is added. The activation function of the middle layer is the Relu unit, and then uses the Softmax classifier [12], [13].

### B. OVERVIEW OF 3DCNN

Three-dimensional convolution is to form a three-dimensional solid by stacking multiple consecutive slices, and then use 3D convolution kernel in this three-dimensional stereo. In the 3D convolution structure, each feature map is connected with multiple adjacent serial slices in the upper layer. As a result, one can capture three-dimensional spatial information and adjacent layer change information [14]. A 3D convolutional network based on a 3D convolution kernel

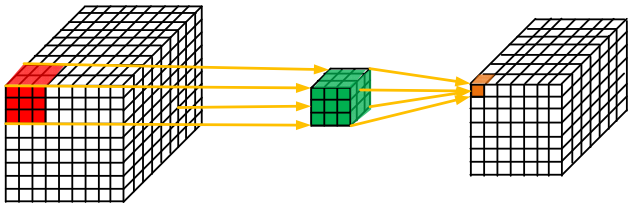

**FIGURE 2.** Three-dimensional convolutional neural network (3DCNN).

feature extractor can generate multi-channel information from successive images, then perform convolution and down-sampling operations on each channel, and finally combine all channel information to obtain the final Feature description. The experimental scheme in this paper is to distinguish the lesions in different data processing situations by constructing the 3DCNN network, and compare it with the traditional 2DCNN experimental results [15].

A three-dimensional convolution is a three-dimensional filter that computes a low-dimensional feature representation from three dimensions (x, y, z), and the output is a three-dimensional convolution space. It is very useful in video event detection, 3D medical image pictures and more. Of note, its use is not limited to three-dimensional space, but can be applied to two-dimensional input as well, as shown in **FIGURE 2**.

### C. CONSTRUCTION OF 3DCNN

Three-dimensional data (or spatial data) has a variety of expressions, such as voxel images, three-dimensional point clouds, and RGB-D images. These information carriers add depth information in three-dimensional space based on the expression of planar image information, so that the three-dimensional data points are uniquely determined in space, which greatly enriches the dimension of acquiring spatial information. The voxel image can visually observe the shape and posture of the three-dimensional object. In addition, it is easier to transplant the processing and analysis method of the ordinary two-dimensional image onto the voxel image than the three-dimensional point cloud data and the RGB-D image. At the same resolution, 3D point cloud data occupy less spatial resources and are easy to express complex textures. RGB-D images simultaneously record RGB images and depth images through RGB-D cameras. RGB images contain object surface color and texture information. Depth images contain spatial shape information of objects. By combining 2 images to detect and identify 3D models. Due to the portability of voxel images, therefore, this paper uses voxel images as a carrier for three-dimensional data.

The complete 3DCNN in this experiment consists of 10 layers and consists of two modules: automatic feature learning module and tumor classification module [16], [17]. The feature extraction module is composed of 8 layers of networks and can be divided into 4 pairs, each of which is followed by a down-sampling layer. We use the 3D convolution kernel as the feature extractor to extract the training volume data space information. The following formula represents a three-dimensional convolution operation:

$$u_{ki}^{1}(\mathrm{x},\mathrm{y},\mathrm{z}) = \sum\nolimits_{m,n,t} h_k^{l-1}(x-m, y-n, z-t) * W_{ki}^{l}(m,n,t) \tag{1}$$

In the above formula, the $W_{ki}^{l}$ represents the kernel in which the 3-dimensional volume space $h_k^{l-1}$ is convoluted. And the $W_{ki}^{l}(m,n,t)$ represents the weights of each voxel in the convoluted kernel. The output value of the corresponding feature space node is:

$$h_i^l = \sigma\left(\sum\nolimits_k u_{ki}^l + b_i^l\right) \tag{2}$$

The entire 3DCNN network model is shown in **FIGURE 3** [18].

### D. DATA SOURCE AND PROCESSING

All oral images in this article were identified and described by a rich oral oncologist and confirmed by a radiologist. The specific information is shown in **FIGURE 4**.

### E. TRAINING OF 3DCNN

The training of 3DCNN is similar to which with 2DCNN. The training process is as follows [18]–[22]:

(1) Using the batch training method, first randomly select N samples from the sample set as a batch group;

(2) Random initialization, setting the weight of the network to a value close to zero in the interval $[-0.5, 0.5]$, initializing the learning rate $\eta$ (generally $\eta = 0.1$ or $0.01$) and the training error threshold $\varepsilon$;

(3) Calculating the error loss for the selected sample set;

(4) Using the batch random gradient descent method to back propagate the error value and update the network parameters;

(5) It is judged whether the total error E of the model after the adjustment of the weight is smaller than $\varepsilon$. If $E < \varepsilon$, proceed to the next step; if not, return to the third step to continue training.

(6) At the end of the training, save the network parameters and get an optimized convolution network. The specific information is shown in **FIGURE 5**.

**I**n order to solve the problem of weak generalization and easy over-fitting of the model trained in the small sample dataset, this paper also uses the 3DCNN network model to be pre-trained on the big dataset, and then migrates the model to our own dataset. A supervised method is to adjust the parameters of the network. There are two main steps: 3DCNN model pre-training and pre-training model migration, fine-tuning network parameters.

The migration learning model migrates 3DCNN models pre-trained on other datasets to other datasets and relearns the characteristics of the target dataset. Training a deep learning model requires a large amount of labeled data, otherwise there will be over-fitting problems, resulting in a high accuracy of the model on the training set and poor performance on the test set. It is very difficult to manually collect a large amount of data and label it for a new task, which requires

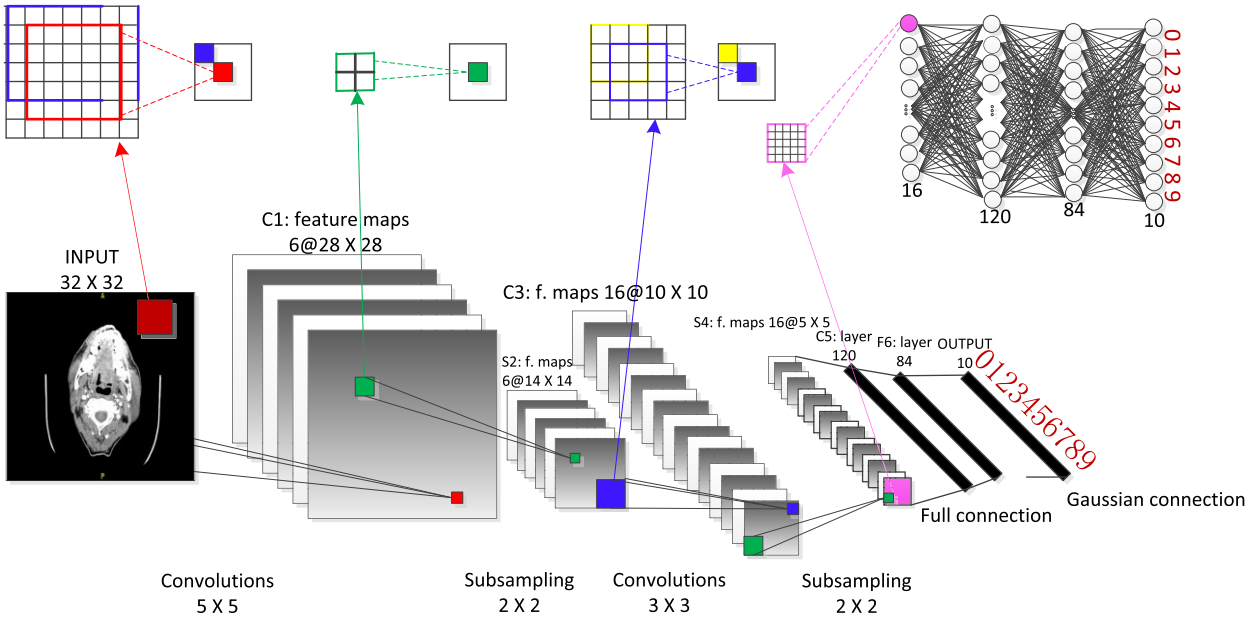


**FIGURE 3.** 3DCNN network mode.

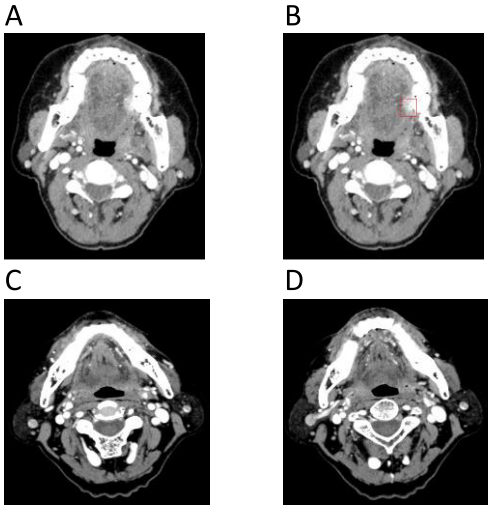


**FIGURE 4.** Samples of our dataset. (A) Original oral image. (B) Oral image with plaque delineated. (C) Healthy control image I. (D) Healthy control image II.

a lot of resources. Migration learning uses the commonality between different learning tasks to transfer knowledge between tasks, and can apply existing data or knowledge learned by the environment to new environments and data.

## III. EXPERIMENTAL RESULTS AND ANALYSIS

### A. DATA EXPANSION METHOD

In general, increasing the number of training sets will result in better training results for an over-fitting network.Data expansion allows us to artificially increase the number of training cases by certain means (Rotation, Scaling, Random translation, and Gamma correction).

There are too few existing dataset samples. The current sample set is about 7,000 CT images of early oral cancers. These images include early oral cancer images of the hospital over the past 20 years, and each sample is labeled with the lableme software. In this paper, before the network training, the data is expanded according to the data processing method in [23], [24] and slightly changed. In the previous study, we have extracted three scale ROIs, of which the smallest scale ROI is the smallest rectangular box that contains the lesion. For irregular lesions, we cut the ROI according to the vertical and horizontal sides of the lesion, and then cut the ROI by 1.5 times and 2.0 times with the primary ROI, respectively. The specific information is shown in **FIGURE 6**.

Because the sample data is too small and the positive and negative samples are unbalanced, we extend the extracted ROI by translational rotation and mirroring. Then, a random number of positive and negative samples are randomly selected from these samples to form a training set for training a two-dimensional convolution network. The expansion of three-dimensional data is similar to the two-dimensional data expansion method, except that the expansion of the three-dimensional data should pay attention to the ROI before and after the largest cross section. The expanded data stack together corresponds to a three-dimensional object.

The specific expansion method: for the ROI or VOI to translate $N_t$ times in any direction, and then rotate the translated ROI around the center at any angle $\alpha = [0°, \ldots, \ldots, 360°]$, then rotate $N_r$ times, then in the translation and the ROI after rotation is adjusted to a different scale $N_m$. So the final result produces a $N = N_t \times N_r \times N_m$ sample for each ROI. In this experiment, we translated the ROI of 1.5 times and 2.0 times by 4 times, rotated by 10 times, and mirrored by 2 times. Finally, each ROI or VOI was expanded by $80 = 4 \times 10 \times 2$ times.

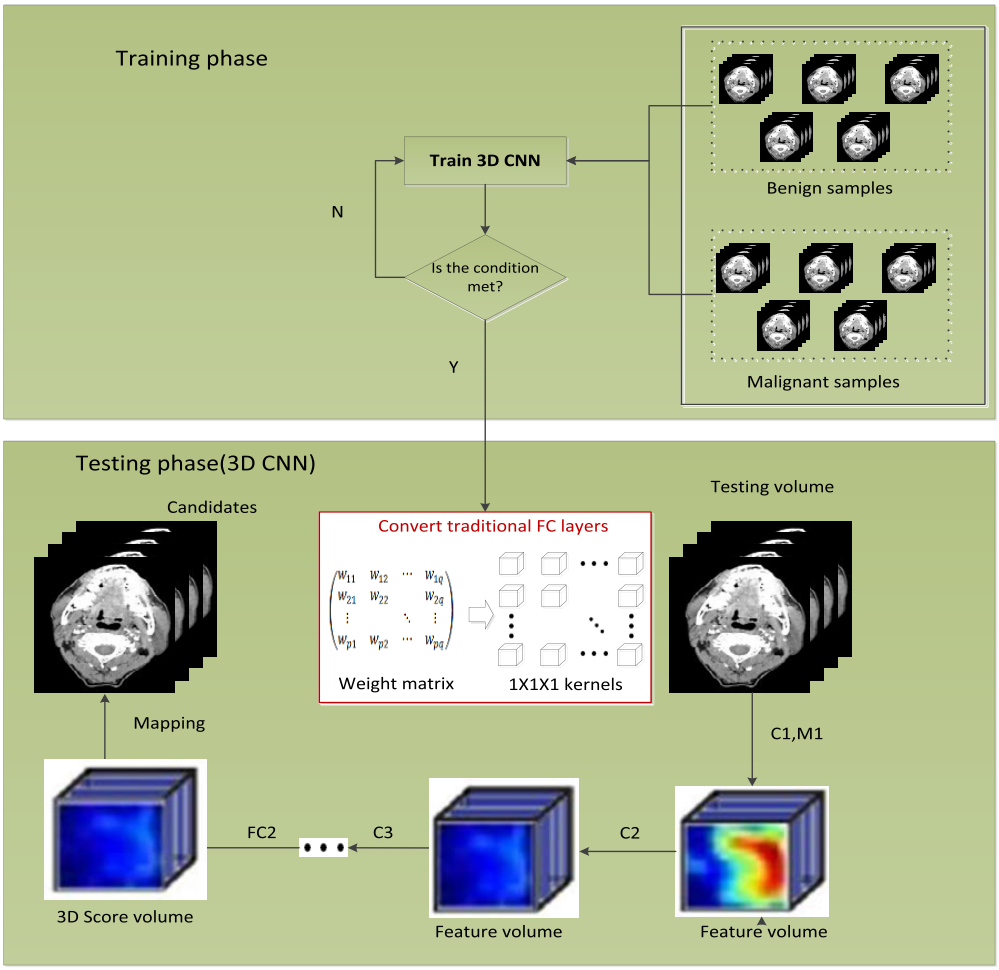


**FIGURE 5.** The Training Process of 3DCNN.

### B. ENHANCEMENT RATE IMAGE CALCULATION

DCE-CT can reflect the dynamic characteristics of the lesion, and the diffusion and absorption characteristics of the enhancer in different types of tumor tissue may be different. According to the dynamic enhancement characteristics of tumors, it has certain reference value for the early diagnosis of oral cancers and the prediction of tumor pathological properties. Based on the acquired oral CT image data, we calculated the dynamic enhancement rate images for three time periods and augmented the data using the same method as above. The following formula represents the enhancement rate:

$$R_{en} = \frac{s_i - s_j}{s_i} \quad i = 1, 2;\ j = 0, 1;\ i \neq j \tag{3}$$

### C. EXPERIMENTAL RESULTS AND ANALYSIS

(1) The Experimental Results and Analysis of Enhanced Image 2DCNN and 3DCNN

The experimental content of this paper mainly studies the ability of 2DCNN and 3DCNN to discriminate early oral tumors in S1 single-sequence original grayscale images. The results of **Table 1** and **FIGURE 7** reveal that the AUC value of the 3DCNN experiment is about 9% higher than which of the 2DCNN experimental and the sensitivity is also improved by nearly 10%.Traditional 2DCNN only uses the structural information of the two-dimensional plane of the image, ignoring the three-dimensional structure of the oral CT image. The 3DCNN combines the slice of different levels of lesion ROI, which not only retains the ability of general CNN to extract two-dimensional planar features, but also takes advantage of the three-dimensional spatial information of CT images to extract the structural features of lesions in three-dimensional space. These indicate that the spatial structure information of oral cancers plays a certain role in distinguishing the benign and malignant early oral cancers. The ROC curves of the two experimental methods are depicted in **FIGURE 8**.

(2) The Experimental Results and Analysis of The Enhancement Rate Image 2DCNN and 3DCNN

The experimental results in **Table 2** and **FIGURE 9** reveal that the three-dimensional enhancement rate image at t1 distinguishes the benign and malignant oral cancers and the performance indexes of the 3DCNN experimental results are significantly better than the 2DCNN network

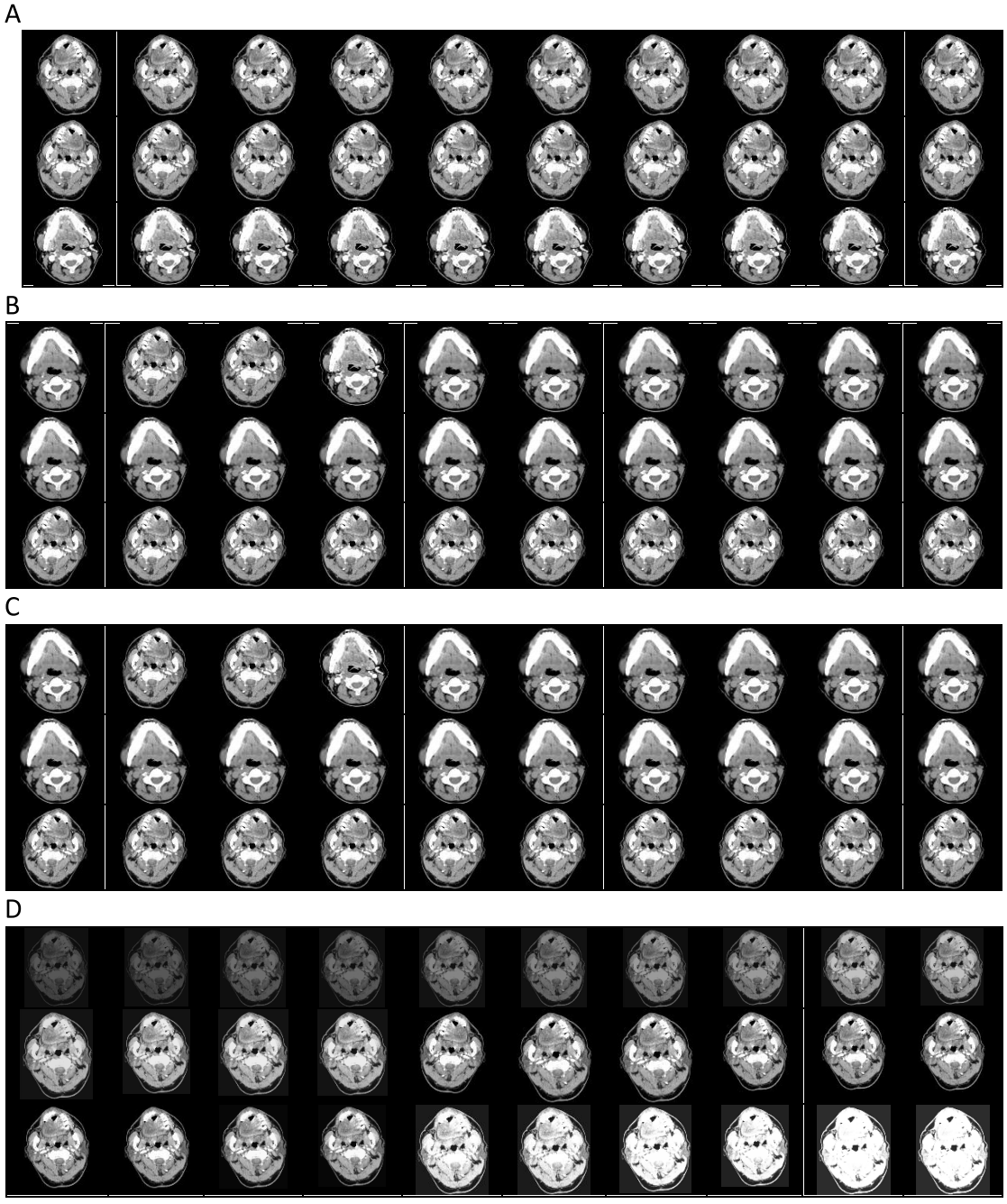


**FIGURE 6.** Results of data augmentation. (A) Rotation. (B) Scaling. (C) Random translation. (D) Gamma correction.

model under the same modal image data. Accuracy ACC and AUC values reached 0.754 and 0.796, sensitivity increased to 0.818, and specificity increased to 0.739. The ROC curve of the two models of the enhanced rate image is shown in **FIGURE 10**. The enhancement rate three-dimensional image reflects the different enhancement characteristics of the lesion in the spatial extent. For different types of oral cancers, the internal space of the tumor is unlikely to be isotropic, and their response to the enhancer is also different. 3DCNN can extract the spatial dynamic characteristics of tumors with enhanced rate imaging. Using this feature can thus significantly improve the ability of the model to distinguish between benign and malignant tumors.

## IV. RESULTS

Through the experiments in this paper, we found that 3DCNN performed better in the diagnosis of benign and malignant oral cancers than the 2DCNN. Since oral CT can acquire the three-dimensional structure of the tumor, 3DCNN uses the oral CT to obtain the three-dimensional spatial information

**TABLE 1.** The experiment results of single sequence enhanced image classification.

| Network Model | AUC | Sens | Spec | ACC |
|---|---|---|---|---|
| 2DCNN | 0.714± 0.133 | 0.752± 0.176 | 0.624± 0.095 | 0.663± 0.189 |
| 3DCNN | 0.798± 0.062 | 0.847± 0.124 | 0.647± 0.089 | 0.759± 0.162 |

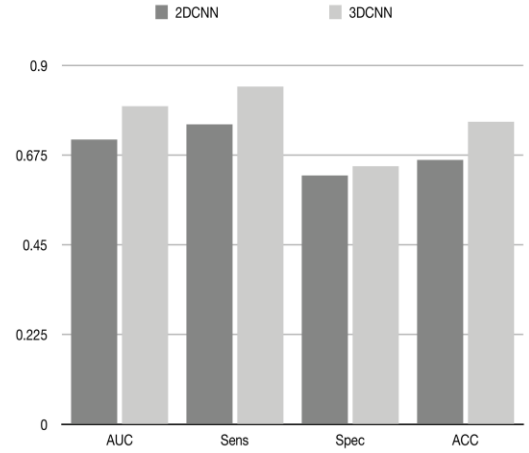


**FIGURE 7.** Histogram of single sequence enhanced image classification experiment results.

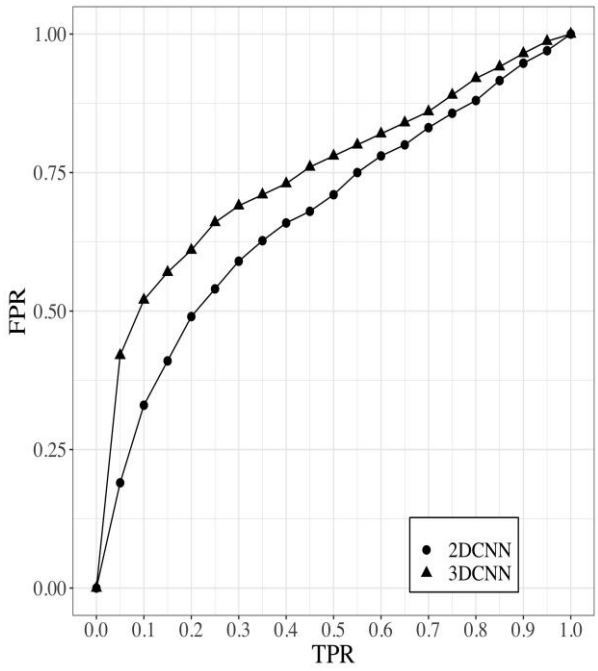


**FIGURE 8.** The Classification Results of Enhanced Image ROI and VOI.

of the tumor, and sequentially processes the single CT image, and can extract the tumor features from multiple angles. The accuracy and sensitivity of the 3DCNN classification results in the single-sequence enhanced image were higher than the 2DCNN by more than five percentage points. Using the multiple sequences of DCE-CT images to calculate the

**TABLE 2.** The classification experiment results of t1-t0 enhancement rate image.

| Network Mode | AUC | Sens | Spec | ACC |
|---|---|---|---|---|
| 2DCNN | 0.734± 0.132 | 0.707± 0.079 | 0.632± 0.169 | 0.687± 0.193 |
| 3DCNN | 0.796± 0.142 | 0.818± 0.165 | 0.739± 0.186 | 0.754± 0.174 |

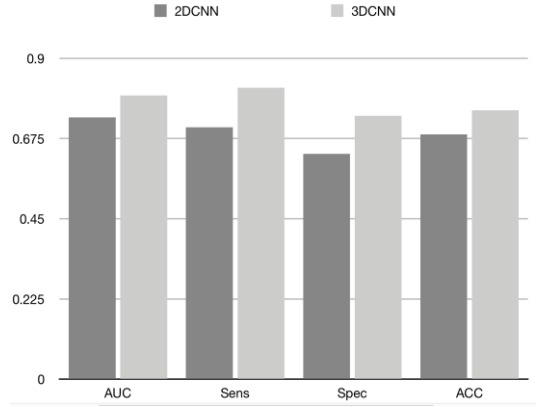


**FIGURE 9.** Histogram of t1-t0 enhancement rate image classification experiment results.

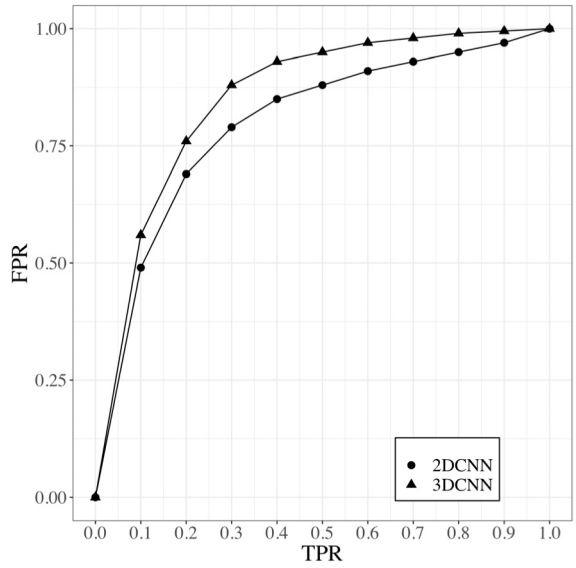


**FIGURE 10.** The Classification Results of Enhanced rate image ROI and VOI.

absorption state of the contrast agent at different times of the tumor; indirectly reflects the difference in the spatial dynamic changes of the tumor. We calculated the ROI of the lesions extracted by the enhancement rate image and combined the 3DCNN network model to identify the early malignant and malignant tumors. From the experimental results, the dynamic characteristics of the enhancement rate image are more distinguishable than the single enhancement sequence. This gives a certain reference value for the design

of oral CT-assisted diagnosis system and feature extraction. By combining the spatial features extracted by 3DCNN and the spatial dynamic change information, the diagnostic accuracy of the oral CT-assisted diagnosis system can be improved. Due to space limit, this paper only discusses a single sequence of images, without combining different imaging modalities.

## V. CONCLUSION

The main research content of this paper firstly proposes the early diagnosis of oral cancers using 3DCNN, and then constructs a deep 2DCNN and 3DCNN. Secondly, because the amount of existing sample data is too small, we carry out image data expansion and calculation. Enhancement rate images of DCE-CT images; then we applied our own 2D and 3D network models to the enhanced image ROI and enhancement rate image ROI experiments, and then evaluated the ability to identify the two models for early lesions on different data modalities. The results show that 3DCNN can better identify benign and malignant lesions of early oral cancers, which is more than 6 percentage points higher than 2DCNN. To diagnose the nature of tumor lesions using information on the dynamic changes of enhancers in tumors of different natures, we applied CNN to the enhancement rate image classification experiment. The results of the study show that the AUC of any convolution model is significantly improved compared to the enhanced image. The 3DCNN network model has an AUC value of 0.801 in the enhancement rate image, which is about 5% higher than the AUC value of the single enhanced image experiment.

## ACKNOWLEDGMENT

The authors thank generous support from the Shanghai Ninth People's Hospital, particularly the Department of Radiology of the hospital.

## REFERENCES

[1] Y. Wen, Y. Lu, J. Yan, Z. Zhou, K. M. von Deneen, and P. Shi, "An algorithm for license plate recognition applied to intelligent transportation system," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 3, pp. 830–845, Sep. 2011.

[2] Z. Liu, J.-Q. Yan, Q.-L. Li, and D. Zhang, "Automated tongue segmentation in hyperspectral images for medicine," *Appl. Opt.*, vol. 46, no. 34, pp. 8328–8334, 2007.

[3] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proc. Int. Conf. Mach. Learn.*, 2010, pp. 807–814.

[4] S. Arora, A. Bhaskara, T. Ma, and R. Ge, "Provable bounds for learning some deep representations," 2013, *arXiv:1310.6343*. [Online]. Available: https://arxiv.org/abs/1310.6343

[5] T. Serre, M. Riesenhuber, T. Poggio, and J. Louie, "On the role of object-specific features for real world object recognition in biological vision," in *Proc. Int. Workshop Biologically Motivated Comput. Vis.* Berlin, Germany: Springer-Verlag, 2002, pp. 387–397.

[6] D. Wazalwar, E. Oruklu, and J. Saniie, "A design flow for robust license plate localization and recognition in complex scenes," *J. Transp. Technol.*, vol. 2, no. 1, pp. 13–21, 2012.

[7] S. Ji, W. Xu, M. Yang, and K. Yu, "3D convolutional neural networks for human action recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 31–221, Jan. 2013.

[8] W. Anjali, B. Anuj, and V. S. Verma, "A review on brain tumor segmentation of MRI images," *Magn. Reson. Imag.*, vol. 61, pp. 247–259, Sep. 2019.

[9] Q. Dou, H. Chen, L. Yu, L. Zhao, J. Qin, D. Wang, V. C. Mok, L. Shi, and P.-A. Heng, "Automatic detection of cerebral microbleeds from MR images via 3D convolutional neural networks," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1182–1195, May 2016.

[10] F. Lu, F. Wu, Z. Peng, D. Kong, and P. Hu, "Automatic 3D liver location and segmentation via convolutional neural network and graph cut," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 12, no. 2, pp. 171–182, 2016.

[11] J. Gu, W. Zhenhua, K. Jason, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai, and T. Chen, "Recent advances in convolutional neural networks," *Pattern Recognit.*, vol. 77, pp. 354–377, May 2018.

[12] W. Dai, Y. Chen, G. Xue, Q. Yang, and Y. Yu, "Translated learning: Transfer learning across different feature spaces," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2008, pp. 353–360.

[13] Y. Zhu, Y. Chen, S. J. Pan, G.-R. Xue, Y. Yu, Q. Yang, and Z. Lu, "Heterogeneous transfer learning for image classification," in *Proc. AAAI Conf. Artif. Intell. (AAAI)*, San Francisco, CA, USA, Aug. 2011, pp. 1304–1309.

[14] C.-C. Chiu, "A novel approach based on computerized image analysis for traditional Chinese medical diagnosis of the tongue," *Comput. Methods Programs Biomed.*, vol. 61, no. 2, pp. 77–89, 2000.

[15] J. J. Hopfield, "Neurons with graded response have collective computational properties like those of two-state neurons," *Proc. Nat. Acad. Sci. USA*, vol. 81, no. 10, pp. 3088–3092, 1988.

[16] T. Tommasi, F. Orabona, and B. Caputo, "Safety in numbers: Learning categories from few examples with multi model knowledge transfer," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 3081–3088.

[17] M. Oquab, L. Bottou, J. Sivic, and I. Laptev, "Learning and transferring mid-level image representations using convolutional neural networks," in *Proc. IEEE Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1717–1724.

[18] Y. Jason, C. Jeff, B. Yoshua, and L. Hod, "How transferable are features in deep neural networks?" in *Proc. Neural Inf. Process. Syst.*, 2014, pp. 3320–3328.

[19] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.

[20] B. Yoshua, "Deep learning of representations for unsupervised and transfer learning," in *Proc. ICML*, 2012, pp. 17–36.

[21] K. He, X. Zhang, J. Sun, and S. Ren, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1026–1034.

[22] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," *J. Mach. Learn. Res.*, pp. 315–323, 2010.

[23] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, R. M. Summers, and D. Mollura, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1285–1298, May 2016.

[24] H. R. Roth, L. Lu, J. Liu, J. Yao, A. Seff, K. Cherry, L. Kim, and R. M. Summers, "Improving computer-aided detection using convolutional neural networks and random view aggregation," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1170–1181, May 2016.

**SHIPU XU** received the M.Sc. degree in computer science from Jiangxi Agricultural University, Nanchang, China, in 2009. He is currently pursuing the Ph.D. degree in software engineering with Tongji University, Shanghai, China. His research interests include image processing, data fusion, machine learning, and the agricultural Internet of Things.

**CHANG LIU** received the Ph.D. degree from the School of Software Engineering, Tongji University, in 2019. He has been with the School of Information Engineering, Nanchang Hangkong University, Nanchang, China, where he is currently a Lecturer, an ACM Member, and a member of the Chinese Computer Federation (CCF). His research interests include medical images analysis, WebVR visualization, virtual reality, and deep learning.

**YONGSHUO ZONG** is currently pursuing the B.Sc. degree with Tongji University, Shanghai, China. His research interest includes image processing.

**SIRUI CHEN** is currently pursuing the B.Sc. degree with Tongji University, Shanghai, China. Her research interests include machine learning, image processing, and reconstruction.

**YIWEN LU** received the B.Sc. degree from the School of Electronic and Information Engineering, Tongji University. Her main research interests include medical image processing and machine learning.

**LONGZHI YANG** (M'12–SM'17) is currently the Director of learning and teaching and an Associate Professor with the Department of Computer and Information Sciences with Northumbria University, U.K. His research interests include computational intelligence, machine learning, big data, computer vision, intelligent control systems, and robotics and the applications of such techniques under real-world uncertain environment. He received the Best Student Paper Award at the 2010 IEEE International Conference on Fuzzy Systems. He is the Founding Chair of the IEEE Special Interest Group on Big Data for Cyber Security and Privacy.

**EDDIE Y. K. NG** received the Ph.D. degree from Cambridge University. He is currently a Faculty Member with the College of Engineering, Nanyang Technological University, Singapore. His main research interests include thermal imaging, human physiology, biomedical engineering, computational fluid dynamics, and numerical heat transfer. He received a Cambridge Commonwealth Scholarship for his Ph.D. degree. He is a Fellow of ASME. He has been the Lead Editor-in-Chief of the Journal of Mechanics in Medicine and Biology, since 2000. He is also the Founding Editor-in-Chief of the *Journal of Medical Imaging and Health Informatics* and an Associate Editor or an EAB of various refereed international journals, such as *Artificial Intelligence*, *Biomedical Engineering Online*, and the *Journal of Advanced Thermal Science Research*.

**YONGTONG WANG** received the B.Sc. degree from Tongji University. She is currently a Software Engineer, her field of work contains data mining and system building and other research interests include deep learning and natural language processing.

**YUNSHENG WANG** was graduated from the Chinese Academy of Agricultural Sciences, in 2007. He was with the Shanghai Academy of Agricultural Sciences. He is currently a Doctor of Crop informatics and a Researcher. His research interests include crop information and agricultural intelligent equipment.

**YONG LIU** received the M.Sc. degree in embedded systems from the Henan University of Technology, in 2013. He is currently with the Shanghai Academy of Agricultural Sciences. His research interests include software and hardware development, and embedded development.

**WENWEN HU** received the M.Sc. degree in computer software and theory from Shanghai Maritime University, in 2017. She is currently with the Shanghai Academy of Agricultural Sciences. Her research interest includes agricultural informationization.

**CHENXI ZHANG** received the Ph.D. degree from the National Defense University, in 1988. He is currently with the School of Software Engineering, Tongji University. His research interests include network engineering and distributed systems.