Semi-supervised Method for Image Texture Classification of Pituitary Tumors via CycleGAN and Optimized Feature Extraction

Hong ZHU (✉ zhuhong@xzhmu.edu.cn)

https://orcid.org/0000-0002-1676-940X

Qianhao FANG

Xuzhou Medical University   https://orcid.org/0000-0002-2767-1539

Yihe HUANG

Xuzhou Medical University

Kai XU

Xuzhou Medical College Affiliated Hospital

Research article

Abstract

Background: Accurately determining the softness level of pituitary tumors preoperatively by using their image textures can provide a basis for surgical options and prognosis. Existing methods for this problem require manual intervention, which could hinder the efficiency and accuracy considerably. Methods: we present an automatic method for diagnosing the texture of pituitary tumors using unbalanced sequence image data. Firstly, for the small sample problem in our pituitary tumor MRI image dataset where T1 and T2 sequence data are unbalanced (due to data missing) and under-sampled, our method uses a CycleGAN(Cycle-Consistent Adversarial Networks) model for domain conversion to obtain fully sampled MRI spatial sequence. Then, it uses a DenseNet(Densely Connected Convolutional Networks)-ResNet(Deep Residual Networks) based Autoencoder framework to optimize the feature extraction process for pituitary tumor image data. Finally, to take advantage of sequence data, it uses a CRNN(Convolutional Recurrent Neural Network) model to classify pituitary tumors based on their predicted softness levels. Results: Experiments show that our method is the best in terms of efficiency and accuracy(91.78%) compared to other methods. Conclusions: We propose a semi-supervised method for grading pituitary tumor texture. This method can accurately determine the softness level of pituitary tumors, which provides convenience for surgical selection and prognosis, and improves the diagnostic efficiency of pituitary tumors.

Background

Pituitary tumor is one of the most common diseases in the nervous system [1]. It is the third largest tumor type in brain and is extremely harmful to the human body [2]. Many critical questions, such as whether a surgical procedure is needed, what kind of procedure is most suitable, and what is the expected postoperative effect, are all closely related to the softness of pituitary tumor [3]. It is important to accurately judge the softness level of pituitary tumor preoperatively in a non-invasive manner. This has been a problem for a long time and is still plaguing the clinic. However, due to the closure nature of the cranial cavity, it is often difficult to accurately determine the softness of pituitary tumor before surgery [4]. With the technological advancements in medical imaging, MR, CT and other imaging modality can now provide rich anatomical information non-invasively. It has been shown that such information can be used to improve the treatment planning for 30% to 50% cancer patients, resulting in more accurate treatments for them. Thus, it is extremely valuable to mine deep quantitative information (such as the softness level) from pituitary tumor image data, which is not perceivable by the naked eyes of clinician.

At present, the most commonly used method for evaluating the image texture of pituitary tumor is image omics, which is defined as the conversion of visual image information into deep features for quantitative research. The advantage of such a method is its interpretability, which is based on domain knowledge [5]. Aerts _et al_ [6] extracted 440 CT features for prognosis, and found that imaging histology can reflect tumor phenotype, internal heterogeneity, and the prognostic radiological features of intra-tumoral heterogeneity are related to potential gene expression patterns, which could effectively assess the prognosis of patients. Zhang _et al_ [7] adopted an approach that combines machine learning techniques with imaging

omics. They extracted 970 medical image features, and used six kinds of machine learning phenomenological feature selection methods and nine classification methods to obtain 54 different combinations. They showed that the random forest method (RF) has the best performance in the prognosis analysis of nasopharyngeal carcinoma images. However, since image omics requires accurate extraction of lesions, it is not very efficient. Moreover, the number of deep features that can be extracted by image omics could be as many as thousands, which need to be selected manually. Thus, it is a challenging task to select the best set of features, as it depends largely on the experience of the technician. In general, feature extraction is a computation-intensive and time-consuming process, and thus better solution is needed.

In recent years, artificial intelligence has gained a lot of popularity which propelled a new way for medical imaging processing. The combination of deep learning and medical imaging. It has shown that such an approach is capable of automatically extracting a large number of deep features from large medical image datasets, and yields much improved solutions. For example, Wang et al [8] combined medical imaging with in-depth learning to develop a new generation of image reconstruction theory technology, which enhanced the ability of image analysis and image reconstruction. Xu et al[9] proposed a new network cxnet-m1 to detect abnormal chest X-ray images, which improved the efficiency and accuracy of diagnosis. Wei et al [10] proposed a method called Locality-constrained Sparse Autoencoder (LSAE) which introduces the concept of locality into Autoencoder and can encode similar inputs by similar features. Their method achieves a classification accuracy of 72.7% for CALTECH-101 dataset. Xu et al [11] presented a new Stacked Sparse Autoencoder (SSAE) framework for the diagnosis of high resolution histopathological images of breast cancer. They used a dataset with 500 histopathological images (2200x2200) and 3500 manually segmented cell nucleuses, and showed that their method improves the F value by 84.49% and yields an AVEP of 78.83%.

Despite the aforementioned progresses, deep-learning-based approaches are also facing a number of challenges, such as data unbalancing in small sample, limited reliably labeled data, inaccurate feature extraction, etc. In the case of pituitary tumor, the dataset we collected is unbalanced, e.g., only T1 sequences but lacking of T2 sequences. In addition, more accurate features of pituitary tumor image data are needed for texture classification. In this paper, we proposed a semi-supervised pituitary tumor image classification method based on CycleGAN and optimized feature extraction. Our method first uses CycleGAN to make up the missing T2 sequences, and then adopts a DenseNet-ResNet based Autoencoder-decoder framework to extract pituitary tumor features and optimize adaptively. Finally, the optimized features are inputted to CRNN. It needs only sequence-level label, instead of frame-level label, to complete the training for subtype classification of pituitary tumors.

Methods

Due to technical limitations, the Methods section is only available as a download in the supplementary section.

Results

3.1 EXPERIMENT ANALYSIS

3.1.1 CYCLEGAN-BASED MULTI-SEQUENCE DATA AMPLIFICATION

We use the image data of 374 patients for CycleGAN training, including 280 T1 MRI spatial sequences and 94 T2 MRI spatial sequences. We train a total of 120 times, in which the loss of the generator and the discriminator is shown in Figure 9. When the number of training reaches 90 epochs, the loss of the discriminator reaches its minimum and becomes stable.

We use 152 patient datawith labels(including 112 T1 MRI spatial sequences and 40 T2 MRI spatial sequences) to augment the data using the trained cyclegan model. As a result, there is a multi-sequence of 24 slices (12 T1 slices and 12 T2 slices) for each patient.The result (after 120 times of training) is shown in Figure 10.

Figure 10 shows the original MR image in two domains and the MR image reconstructed after two conversions by the domain converter. Visually, the difference between a real MR image and a transformed MR image is very small.

3.1.2 SEMI-SUPERVISED PITUITARY TUMOR TEXTURE IMAGE CLASSIFICATION BASED ON ADAPTIVELY OPTIMIZATED FEATURE EXTRACTION

After being amplified by CycleGAN, the dataset was then fed to the Auto-Encoder for feature extraction using unsupervised learning. Supervised learning is conducted during the CRNN texture classification stage.

To ensure reliable comparisons, all the models were trained 100 steps in the feature extraction stage. The training process of multi-sequences is shown in Figure 11, and the curve of the single-modal baseline is similar.

It can be seen from the figure that when the model is trained 100 steps, the loss curve reaches its lowest point, which is 0.01, and feature extraction network almost achieves the optimal solution.

The architecture of the experiment can be divided into three models, namely the multi-sequence model, the T1 domain model and the T2 domain model. The multi-sequence (medical image classification) model is compared to two single-modal baseline models:

(1) T1 domain model: We only consider the MRI spatial sequence of T1 domain of all patients, including the MRI spatial sequence generated from another domain converter.

(2) T2 domain model: We only consider the MRI spatial sequence of T2 domain of all patients, including the MRI spatial sequence generated from another domain converter.

(3) Multi-sequence model: We use the trained domain converter to construct an MRI multi-sequence in both T1 and T2 domains, including the MRI spatial sequence generated by the domain converters.

In the texture classification stage, there are many neural network model parameters in the experiment, but a small number of trained samples. This could potentially cause over-fitting. To avoid this issue, we use Dropout and EarlyStopping methods during the training process. The Droupout ratio is set to be 0.5, that is, for all the neural network units in model, they are temporarily discarded from the network with a probability of 50%. We set the patience value of earlyStopping to be 2 and the monitor to be 'val_loss'. That is, if the value of 'val_loss' does not decrease relative to the previous epoch during model training, the model is stopped after 2 epochs. The T1 domain, T2 domain, and multi-sequence model training process are shown in the following figures: **See figures 12-14.**

As can be seen from Figure 12-14, we performed 6 replicate experiments on the T1 domain, T2 domain and the multi-sequence domain. In our experiment, we randomly divide the dataset into training dataset (70%), test dataset (15%), and verification dataset (15%). We repeated this process 6 times, and recorded the average and variance of 6 classification accuracy rates. Table 1 shows the details of classification, and Table 2 shows precision, recall and F1-score of classfication:

TABLE 1 PITUITARY TUMOR CLASSIFICATION ACCURACY

	Multi-sequence(%)	T1 domain(%)	T2 domain(%)
Train	98.8±1.24	97.55±1.40	97.41±1.37
Verification	92.82±1.23	91.70±1.61	91.15±1.13
Test	91.78±1.44	89.24±3.11	88.98±4.23

TABLE 2 PRECISION, RECALL AND F1-SCORE OF PITUITARY TUMOR CLASSIFICATION

	Precision(%)	Recall(%)	F1-score(%)
T1 domain	86.81±3.67	93.33±5.96	89.80±2.64
T2 domain	87.07±3.71	94.44±5.02	90.41±2.15
Multi-sequence	89.89±4.02	95.55±5.44	92.46±1.74

TABLE 3 COMPARISONS OF CLASSIFICATION RESULTS OF DIFFERENT METHODS

Feature extraction	Texture classification	Accuracy(%)	Time(s)
——	VGG	69	113
——	ResNet	78.25	105
——	DenseNet	81.25	97
——	CRNN	73.7	67
ResNet+ ResNet	CRNN	88.76	43
DenseNet+DenseNet	CRNN	90.33	43
DenseNet+ResNet	CRNN	91.78	42
DenseNet+ResNet	RNN	89.12	42

As can be seen from the above table, our proposed DenseNet+ResNet+CRNN architecture significantly outperforms all other methods in terms of running time and classification accuracy. Our method has the fastest convergence rate and thus shortest running time. From the perspective of classification accuracy, we can see that adding an Auto-Encoder-based feature extractor before CRNN can considerably improve the performance. In summary, the comparative experiment suggests that our CycleGAN-based classification model and the adaptively optimized feature extraction has great potential of yielding accurate texture classification results for pituitary tumors.

In order to verify the clinical statistical significance of the experiment, we paired the method proposed in this article with the other methods in Table 3. We use Wilcoxon signed rank test to perform statistical test on paired samples, and the specific data are shown in Table 4.

TTABLE 4 STATISTICS OF WILCOXON SIGNED RANK TEST BASED ON PAIRED SAMPLES

Feature extraction	Texture classification	Z	P
——	VGG	-2.201	0.028
——	ResNet	-2.201	0.028
——	DenseNet	-2.201	0.028
——	CRNN	-2.201	0.028
ResNet+ ResNet	CRNN	-2.201	0.028
DenseNet+DenseNet	CRNN	-2.023	0.043
DenseNet+ResNet	RNN	-2.201	0.028
DenseNet+ResNet	CRNN	——	——

It can be seen from Table 4 that the P values obtained by statistics on various models are all less than 0.05, which is statistically significant. Results have clinical significance.

In order to reflect this contrast more clearly, we have drawn a forest plot, as shown in Figure 15.

As can be seen from the forest plot, our proposed method is more effective compared with other methods.

Discussion

In this study, several experiments were designed to validate our method. Particularly, We first carried out a comprehensive evaluation of the image data generated by CycleGAN, and found that the generated images were great. Subsequently, we list the training curves of feature extraction part to judge extraction effect. Finally, we repeated the experiment six times, calculated the test accuracy, and compared it with other models, and found that our method is the best in terms of accuracy and efficiency. These experiments demonstrate that our method has advantages in grading pituitary tumors. Despite the achievements reported in this paper, several improvements remain possible: On the one hand, the data samples used in the experiment are still insufficient, and it is easy to produce the phenomenon of overfitting. On the other hand, although the loss of feature extraction model training is low and convergence is achieved, the accuracy is still not high enough. Future research in the domain shall address these issues, possibly collecting new data and improving the part of feature extraction.

Conclusion

In this paper, we proposed a deep neural network model for determining the softness level of pituitary tumors, which has the potential to assist clinical diagnosis. Our method first uses CycleGAN to amplify the pituitary tumor dataset to generate multi-sequence samples, which enhances the diversity of pituitary tumor samples and thus helps resolve the under-sampling issue. Then, our method uses an Auto-Encoder architecture, based on ResNet encoding and decoding, to extract the pituitary tumor features, which can improve the classification efficiency of the network to some extent. Finally, the extracted pituitary tumor features are fed to CRNN for classification/grading of the softness level of pituitary tumors. Experiments on a real medical dataset show that our method achieves significantly improved results than some existing popular methods. The experimental results also suggest that our adaptively optimized feature extraction method can better identify deep texture features of pituitary tumor image, and can thus improve the classification accuracy of pituitary tumors.

Declarations

Ethics approval and consent to participate

We have obtained a permission from the Imaging Department of the Affiliated Hospital of Xuzhou Medical University. What we need to declare is that pituitary tumor dataset used in our study has followed all the procedures required by the Chinese government's law. The data has been strictly reviewed by those

in charge of such issues and all sensitive information has been removed. The study is purely for research purpose and does not have any dispute of interest.

Consent for publication

Not applicable.

Availability of data and materials

The data that support the findings of this study are available from [Kai Xu] but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of [Kai Xu].

Competing interests

The authors declare that they have no conflict of interest.

Authors' contributions

Q. F. and Y. H. designed and implemented the machine learning algorithm, and analyzed the pituitary dataset. They contributed equally to this work. K. X. was responsible for the data acquisition and theoretical guidance of medical imaging. H.Z. designed the algorithm and overseen the whole study.

Authors' information

Hong ZHU was born in XU Zhou, JIANG Su, CHINA in 1970. She received her MS and PhD degrees in computer science and technology from China University of Mining and Technology. Hong ZHU, Professor, Master supervisor, is a member of China Computer Federation. Her main research interests include artificial intelligence, medical image processing, machine learning and Granularity calculation.

Qianhao FANG was born in YANG Zhou, JIANG Su, CHINA in 1995. He is currently a student of Xuzhou Medical University pursuing his master degree in medical informatics.

Qianhao FANG, Master graduate student of Xuzhou Medical University. His main research area is deep learning.

Yihe HUANG was born in CHANG Shu, JIANG Su, CHINA in 1996. He has completed his undergraduate degree at Xuzhou Medical University.

Kai XU was born in XU Zhou, JIANG Su, CHINA in 1962. He received his master's degree in diagnostic radiology from the Fourth Military Medical University in 1990 and received his Ph.D. in Imaging Medicine and Nuclear Medicine from Fudan University in 2005.

Kai XU, Professor, chief physician, and doctoral tutor, is the current chairman of the Second Digital Medical Branch Committee of the Jiangsu Medical Association. His main research direction is imaging diagnosis of central nervous system diseases.

Abbreviations

CycleGAN: Cycle-Consistent Adversarial Networks

DenseNet: Densely Connected Convolutional Networks

ResNet: Deep Residual Networks

CRNN: Convolutional Recurrent Neural Network

References

1. Y. M, Pituitary adenomas: an overview of clinical features, diagnosis and treatment, Medical Journal of Chinese People's Liberation Army, 2017; 42(7):576-582.
2. H, et al., Resection of pituitary tumors: endoscopic versus microscopic, J NEURO-ONCOL, 2016; 130(2):1-9.
3. J. D, et al., Magnetic resonance elastography detects tumoral consistency in pituitary macroadenomas, Pituitary, 2016; 19(3):286-292.
4. Hyung. R. A, et al., Natural history of the anterior visual pathway after surgical decompression in patients with pituitary tumors, CAN J NEUROL SCI, 2015; 42(S1):S45-S45.
5. Wenjun, et al., MRI diagnosis and image characteristics of invasive pituitary adenomas, Journal of Modern Oncology, 2014; DOI:10.3969/j.issn.1672-4992.2014.11.64.
6. HJWL, et al., Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach, Nature Communications, 2014; 5:1-8.
7. B, et al., Radiomics Features of Multiparametric MRI as Novel Prognostic Factors in Advanced Nasopharyngeal Carcinoma, CLIN CANCER RES, 2017; 23(15):4259.
8. Wang, and Ge, A Perspective on Deep Imaging, IEEE Access, 2016; 4:8914-8924.

9. S, Wu. H, and Bie R, CXNet-m1: Anomaly Detection on Chest X-Rays with Image-Based Deep Learning, IEEE Access, 2018; 1-1.

10. W, et al., Locality-Constrained Sparse Auto-Encoder for Image Classification, IEEE SIGNAL PROC LET, 2015; 22(8):1070-1073.

11. J, et al., Stacked Sparse Autoencoder (SSAE) for Nuclei Detection on Breast Cancer Histopathology Images, IEEE T MED IMAGING, 2016; 35(1):119.

12. J. Y, et al., Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. Presented at ICCV 2017. [Online]. Avilable: https://arxiv.org/abs/1703.10593.

13. G, et al., Densely Connected Convolutional Networks. Presented at CVPR IEEE Computer Society 2017. [Online]. Avilable: https://arxiv.org/abs/1608.06993.

14. K, et al., Deep Residual Learning for Image Recognition. Presented at CVPR 2016. [Online]. Avilable: https://arxiv.org/abs/1512.03385.

15. B, Bai. X, and Yao. C, An End-to-End Trainable Neural Network for Image-based Sequence Recognition and Its Application to Scene Text Recognition, IEEE T PATTERN ANAL, 2015; 39(11):2298-2304.

Figures
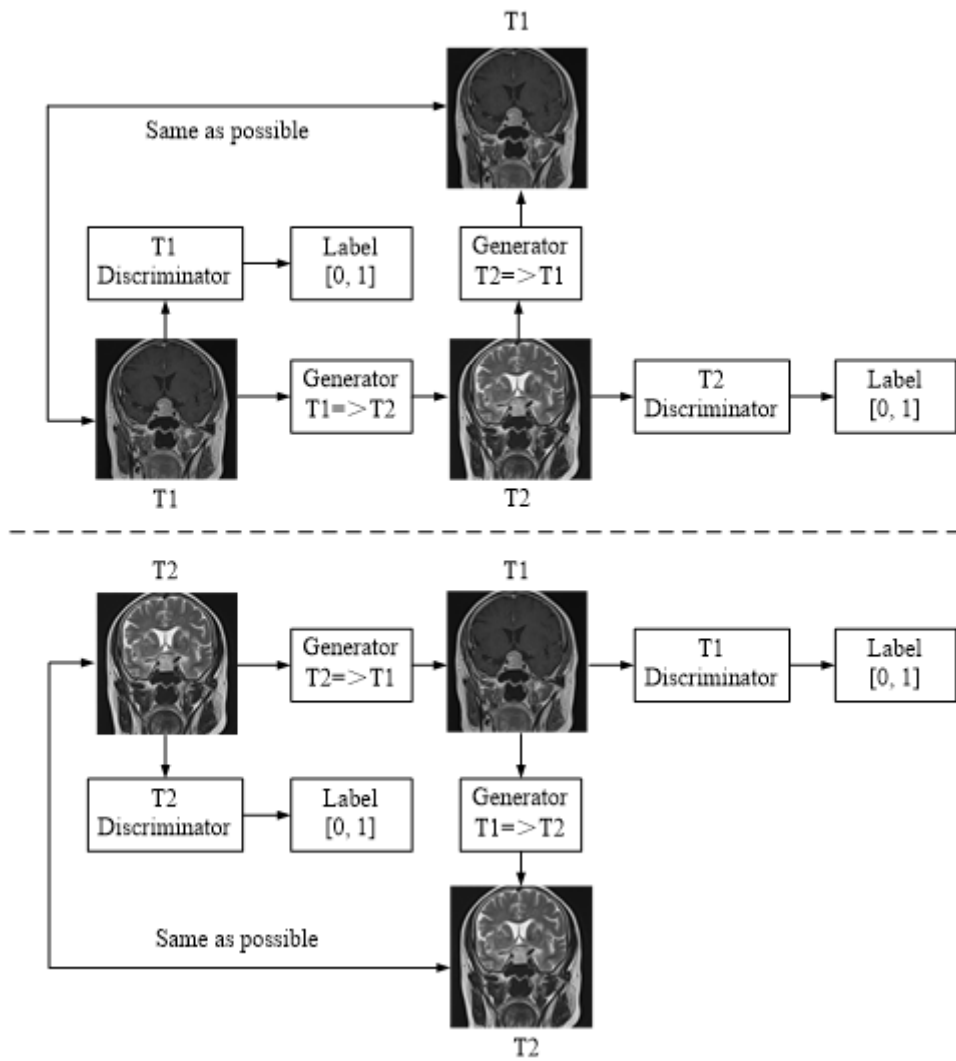
Figure 1

CycleGAN based Deep Neural Network Model for domain conversion

MRI

4*4 Conv2D, 64

4*4 Conv2D, 128

4*4 Conv2D, 256

4*4 Conv2D, 512

4*4 Conv2D, 1

Decision [0, 1]
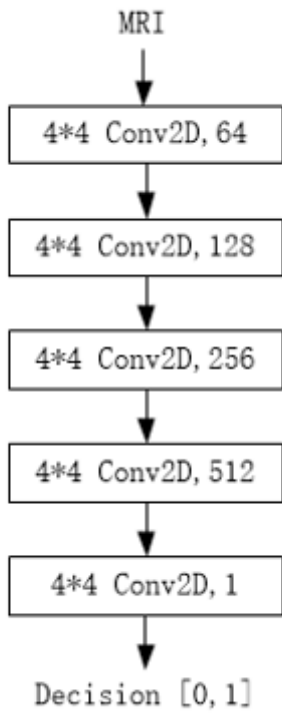
Figure 2

Discriminator architecture

MRI(X domain)

4*4 Conv2D, 32

4*4 Conv2D, 64

4*4 Conv2D, 128

4*4 Conv2D, 256

4*4 Deconv2D, 128

4*4 Deconv2D, 64

4*4 Deconv2D, 32

4*4 Deconv2D, 1

MRI(Y domain)
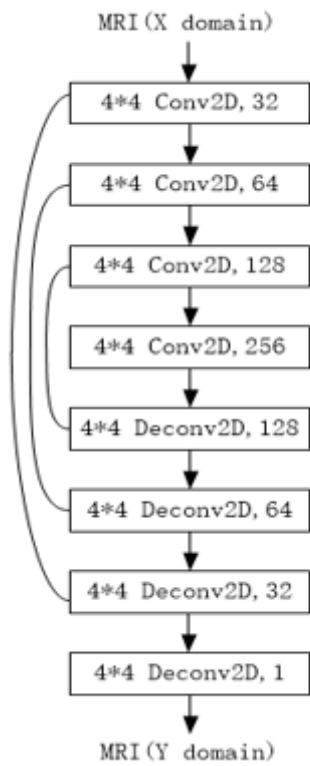
Figure 3

Generator architecture
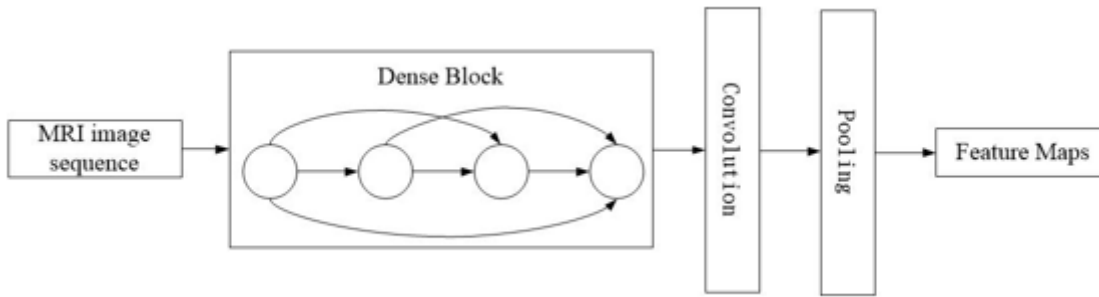
Figure 4

ResNet encoder network architecture
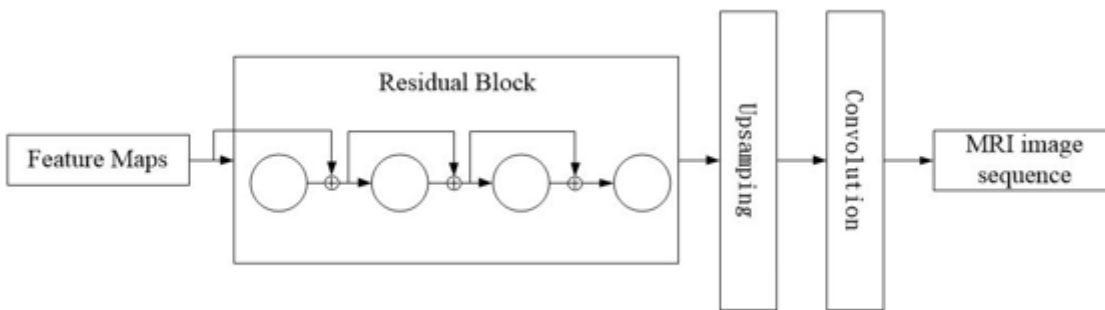
Figure 5

ResNet decoder network architecture
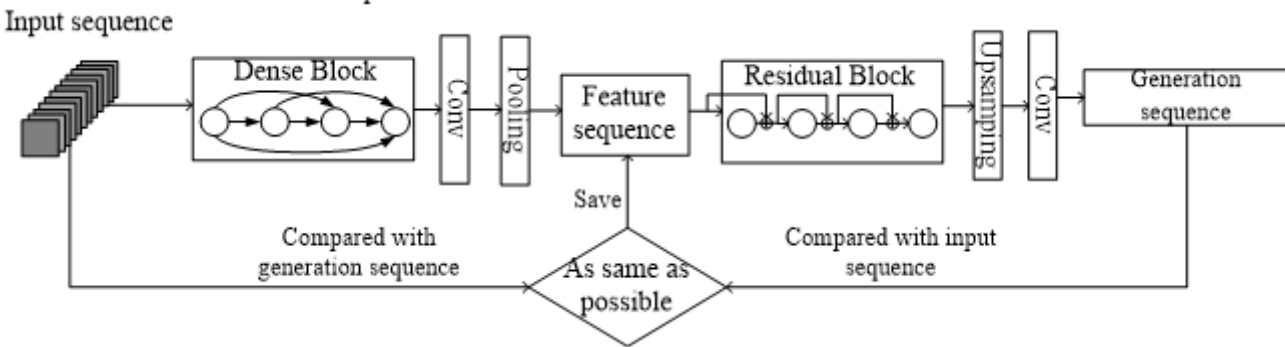
Figure 6

Feature extraction model based on Auto-Encoder
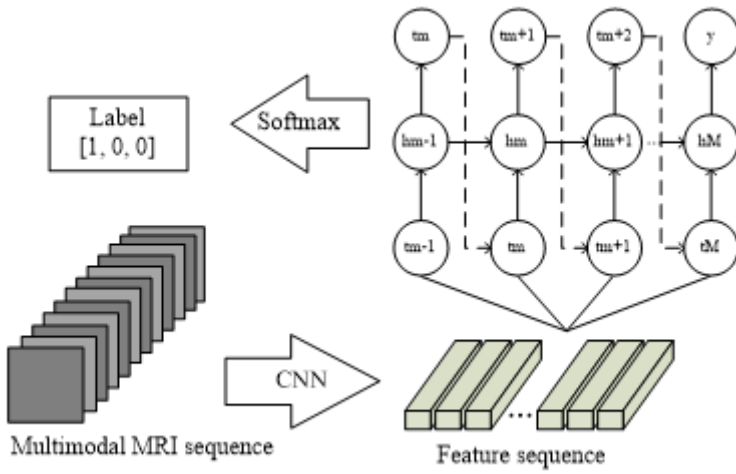
Figure 7

CRNN classifier network architecture
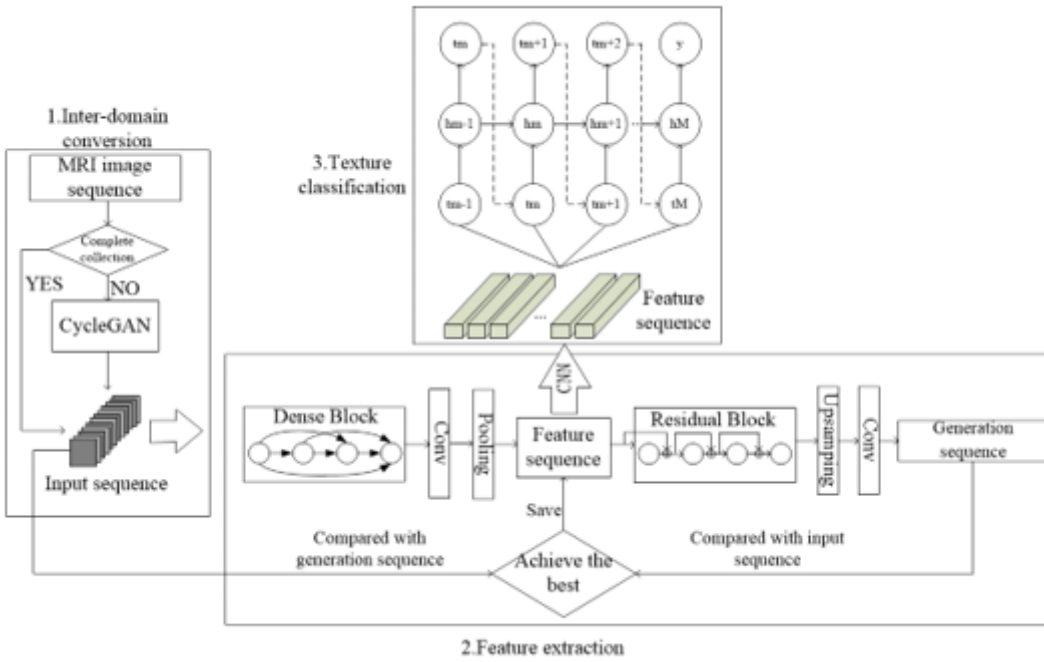
Figure 8

Multi-sequence pituitary tumor grading model
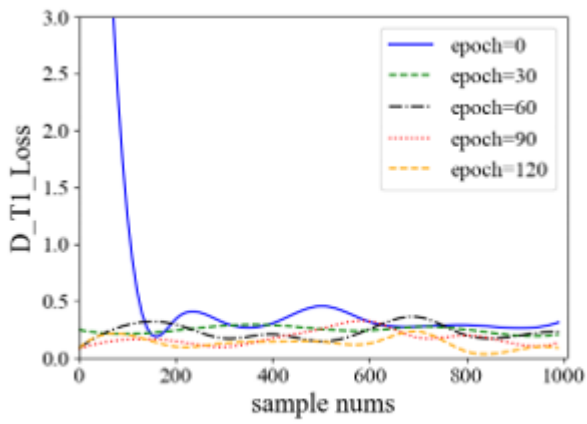
Figure 9

Discriminator loss and generator loss
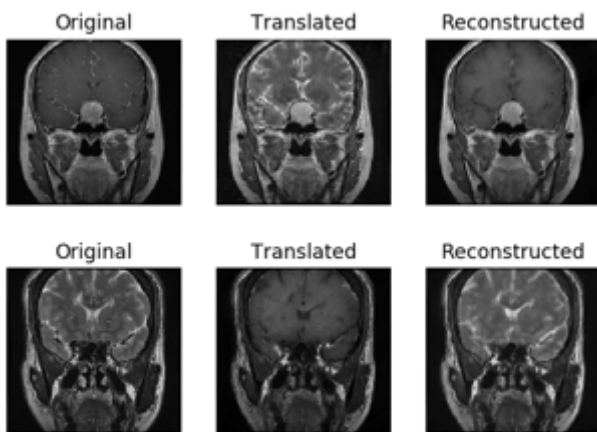
Figure 10

Visualization of training results
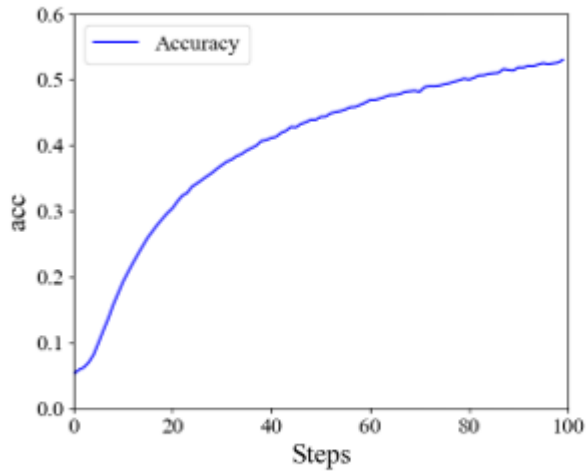
(a)   Feature extraction accuracy
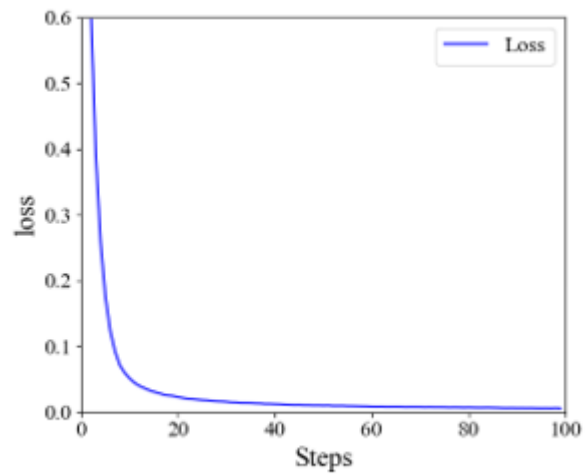
(b)   Feature extraction loss

Figure 11

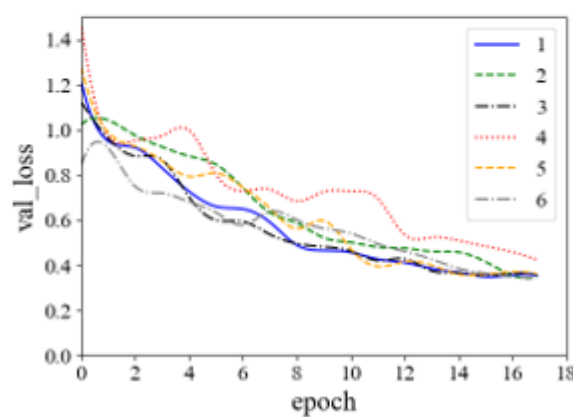Multi-sequence feature extraction model
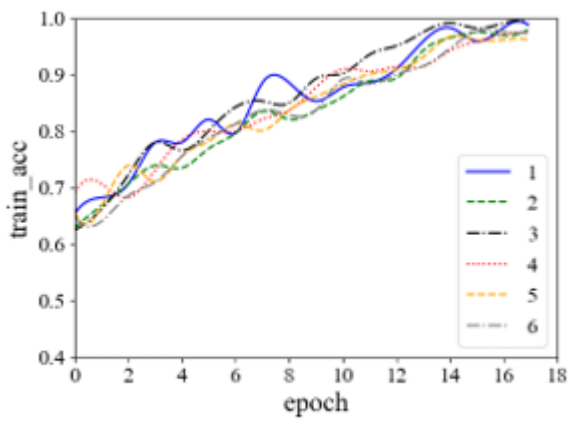
(a)   T1 train accuracy
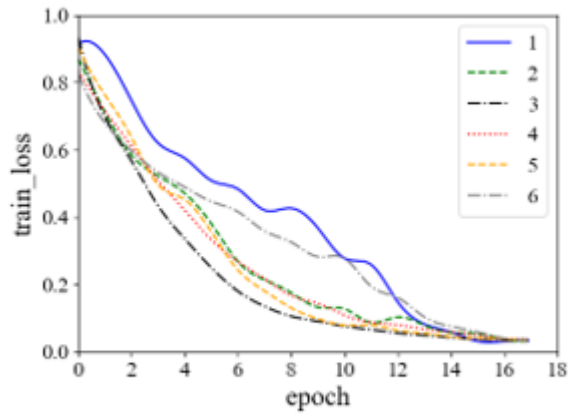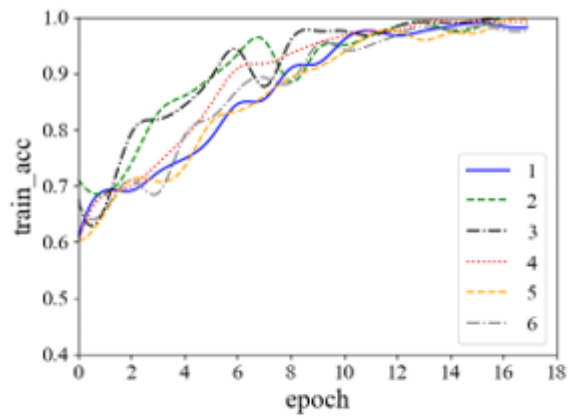
(b)   T1 train loss
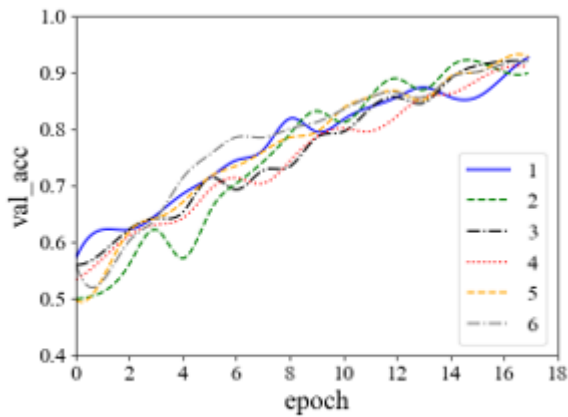
(c)   T1 verification accuracy
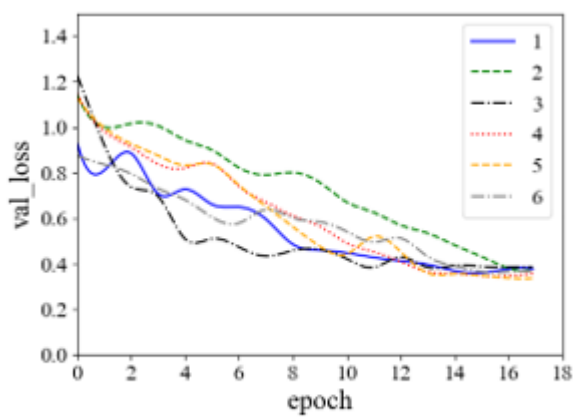
(d)   T1 verification loss

Figure 12

T1 domain image classification model training
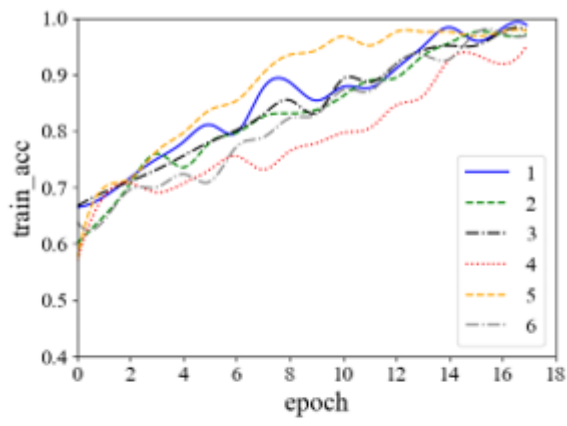
(a) T2 train accuracy
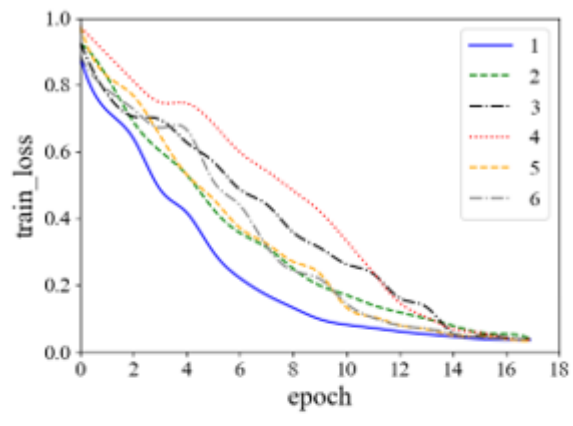
(b) T2 train loss
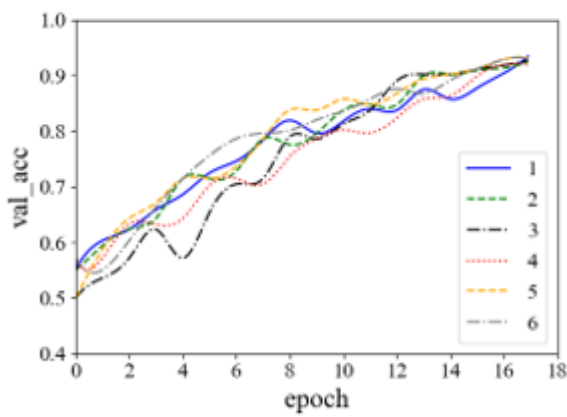
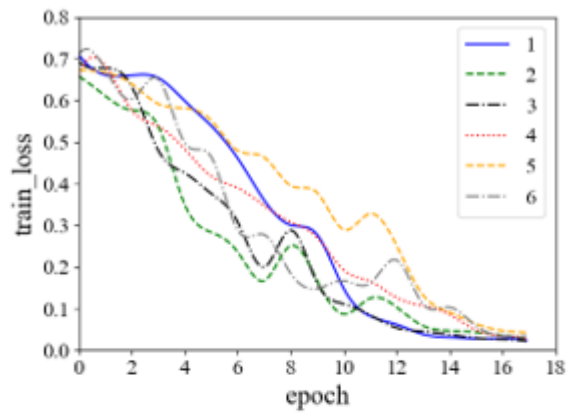(c) T2 verification accuracy

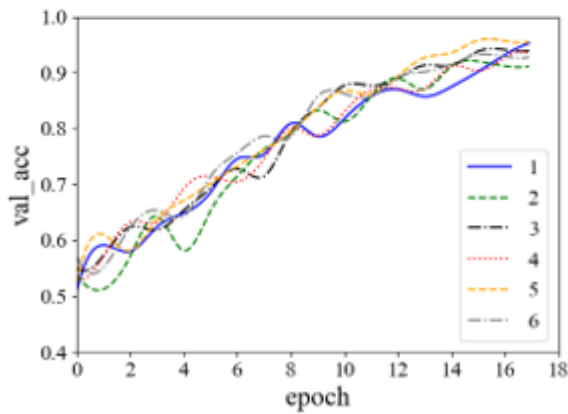(d) T2 verification loss

Figure 13

T2 domain image classification model training
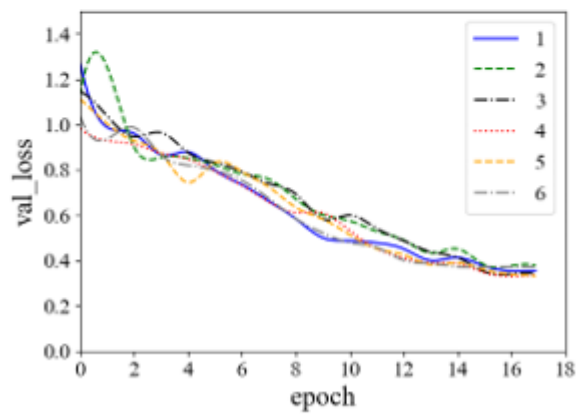
(a) Multi-sequence train loss   (b) Multi-sequence train accuracy

(c) Multi-sequence verification accuracy   (d) Multi-sequence verification loss

Figure 14

Multi-sequence image classification model training

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Methods.docx