

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

Keywords: pituitary tumors, CycleGAN, DenseNet, ResNet, Auto-Encoder, CRNN

Posted Date: December 5th, 2019

DOI: <https://doi.org/10.21203/rs.2.18161/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published on September 9th, 2020. See the published version at <https://doi.org/10.1186/s12911-020-01230-x>.

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 model for domain conversion to obtain fully sampled MRI spatial sequence. Then, it uses a DenseNet-ResNet 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 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 nuclei, 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

2.1 THEORETICAL BASIS

2.1.1 CYCLEGAN

CycleGAN [12] is basically two mirrored GANs that form a ring network. The two GANs share two generators, each with a discriminator, that is, two discriminators and two generators. The discriminator is actually a convolutional neural network that takes an image as input and predicts it as the original image or the output image of generator. The generator consists of an encoder, a converter and a decoder. The specific structure is shown in Figure 1.

See Figure 1

Encoding—The first step uses a convolutional neural network to extract features from the input image.

Transformation—The feature vector of the image in the A domain is converted to a feature vector in the A1 domain by combining the non-similar features of the image. There are 6 layers of Reset modules, each of which is a neural network layer composed of two convolutional layers, which can achieve the goal of retaining the original image features while converting.

Decoding—The work of restoring low-level features from the feature vectors is performed using a deconvolution, and finally the generated image is obtained.

The goal of CycleGAN is to convert image A to another domain to generate image A1 and convert A1 back to A, where output image A1 is similar to the original input image A to form a meaningful mapping that does not exist in the unpaired data set. The advantage of CycleGAN is its ability to train two image sets without pairing.

2.1.2 DENSENET

DenseNet[13] is a convolutional neural network framework with dense connectivity proposed by Huang Gaoren in 2017. In its architecture, there is a direct connection between any two layers of network. The input of each layer of network is a combination of the output of all previous network layers, which enhances the propagation of features. It alleviates the problem of gradient disappearance, reduces network parameters and encourages feature reuse. It has been widely used in medical image field. The specific structure diagram is shown in Figure 2.

See Figure 2

Each Dense Block is a convolution layer with BN layer joined by ReLU layer joined by convolution core (1*1) and BN layer joined by ReLU layer joined by convolution core (3*3). Many basic structures constitute a complete DenseNet. Each block is composed of convolution layer and pooling layer. Densenet uses Block stacking structure to enhance the transfer of features and make more effective use of features, thus reducing the number of parameters to a certain extent and alleviating the phenomenon of gradient disappearance.

2.1.3 RESNET

ResNet[14] is a convolutional neural network framework proposed by He et al in 2015. It adds a shortcut on top of the original architecture to enable direct connection between the mappings of layers. That is, the output of one layer of network is the sum of traditional output and the output of a previous layer, which solves the degradation problem. ResNet alleviates the gradient vanishing and gradient explosion problems caused by the increased depth of the network, and thus protects the entirety of the data. The principle of the model is shown in Figure 3

See Figure 3

As can be seen from Figure 3, the expected network layer relationship mapping output is $H(x)$, where x refers to the input value, and those layers that are increased more are fitted to another mapping, residual, $F(x) = H(x)-x$. Therefore, optimizing the residual map $F(x)$ is easier than optimizing the original map $H(x)$. With the existence of x , the network no longer has the gradient vanishing problem often caused by large depth of the model.

2.1.4 CRNN

CRNN is a model proposed by Shi et al[15] to deal with sequence-like objects in images, which consists of DCNN and RNN. DCNN is used to extract sequence features from the input image. RNN has the advantage of processing sequence data, and can achieve better recognition accuracy from the extracted sequence features. The specific model structure is shown in Figure 4:

See Figure 4

As can be seen from the figure, CRNN consists of three parts, convolutional layer, recurrent layer and transcription layer. The convolutional layer extracts a sequence of features from the input image. Next to the convolutional layer is the recurrent layer which is for predicting each frame using the feature sequence from the convolutional layer. Finally, the predictions of the recurrent layer are converted into a tag sequence using the transcription layer at the end of CRNN.

CRNN abandons the fully connected layer used in traditional neural networks to obtain a more compact and efficient model. It can acquire input images of different sizes, produce predictions of different lengths, and run directly on coarse-grained labels. It is not necessary to specify each individual element in detail during the training phase.

2.2 PITUITARY TUMOR SEQUENCE DATA AMPLIFICATION USING CYCLEGAN

A problem often encountered in MR images of pituitary tumors is under-sampling in a single domain (e.g., T1 or T2). This can be caused by various reasons, such as data missing or simply under sampling. To resolve this issue, our main idea is to use images from other domains (which may come from different image modalities) to generate a set of new images through domain conversion. The set of new and old images forms an augmented set of images which provides a better sample for the domain.

Particularly, we use CycleGAN for data augmentation. First, two domain converters are designed and trained based on the CycleGAN architecture to allow inter-domain conversion from T1 to T2 and from T2 to T1. Then, the generated MR images from domain conversion are added to the original sets of images to form augmented T1 and T2 sequences.

2.2.1 MULTIPLE SEQUENCE OF PITUITARY TUMOR MR IMAGES

As mentioned above, the MR images of one patient usually include spatial sequences from different modalities, such as T1WI, T2WI, T1C and T2FLAIR, etc. In this paper, we mainly use T1 and T2 spatial sequence images.

For each patient i we denote its T1 spatial sequence is as $T1_i = \{t1,1_i, \dots, t1,N_i\}$, where $t1,n_i$ represents the n -th slice/frame in the T1 spatial sequence, and its T2 spatial sequence as $T2_i = \{t2,1_i, \dots, t2,N_i\}$, where $t2,n_i$ represents the n -th slice/frame in the T2 spatial sequence. The number of slices per sequence is N (12 in this paper). To classify the pituitary tumors, we combine the T1 and T2 spatial sequences of each patient i to obtain a spatial sequence of multiple sequences, which is denoted as:

$$T_i = \{t1,1_i, t2,1_i, t1,2_i, t2,2_i, \dots, (t1,N_i, t2,N_i)\} \quad (1)$$

The total number of slices in a multi-sequence spatial sequence is $2N$ (24 in this paper).

2.2.2 TRAINING DOMAIN CONVERTER BASED ON CYCLEGAN

In this paper, we use the CycleGAN framework to design and train the domain converter. CycleGAN is essentially a cyclic network consisting of two mutually symmetric GANs. On top of the original GAN, additional loop constraints are added to force the image to be converted into its original image format so as to reconstruct itself. This allows images to be converted from one domain to another domain without needing to pair them. The architecture of our domain converter is illustrated in Figure 4. In our design, we need to train the T1-to-T2 generator $Trt1,n;\theta_1$ and the T2-to-T1 generator $Trt2,n;\theta_2$, as well as the T1 domain discriminator $Dis(t1,n;\theta_3)$ and T2 domain discriminator $Dis(t2,n;\theta_4)$, where θ_1 , θ_2 , θ_3 and θ_4 are the to-be-determined parameters in the deep neural network.

See Figure 5

The training of the above network mainly consists of two steps:

1) The training of the Discriminator: Fixing the values of θ_1 and θ_2 , update the values of θ_3 and θ_4 . This is for discriminating the authenticity of the image. If the input is MR image data from the real domain, the label is 1, and if the input is an MR image data generated by the generator, the label is 0. In short, the role of the discriminator is to score pictures. If the input pictures are real pictures from the original dataset, they will get high scores. Otherwise (i.e., they are generated fake pictures), their scores will be low. The network of this part is depicted in Figure 6, where the convolution layer is consists of Conv2D, Leaky ReLU, Instance Normalization, and the digits represent the size and number of the convolution kernel.

2) The training of the Generator: Fixing the values of θ_3 and θ_4 , update the values of θ_1 and θ_2 . This is for inter-domain conversion of images. After getting the input MR images from T1 or the T2 domain, the generator sends them to the corresponding domain converter to generate MR images of the other domain. The generated images are then again sent to the corresponding domain converter to generate MR images of the original domain. After being converted twice, the obtained MR images are forced to be as similar as possible to the original ones. The network of this part is shown in Figure 7, where the convolution layer consists of Conv2D, Leaky ReLU, Instance Normalization. The first three de-convolution layers consist of UpSampling2D, Conv2D, ReLU, Instance Normalization, and the last de-convolution layer consists of UpSampling2D, Conv2D, Tanh. The dashed line represents the superimposing operation between the corresponding network layers, and the digits are the size and number of the convolution kernel.

See Figure 6

2.3 SEMI-SUPERVISED CLASSIFICATION METHOD FOR THE IMAGE TEXTURE OF PITUITARY TUMORS BASED ON ADAPTIVELY OPTIMIZED FEATURE EXTRACTION

To improve the efficiency of feature extraction for determining the softness level of pituitary tumor, using ResNet we propose in this paper an Auto-Encoder-based deep neural network model for feature extraction. Since the weight of the features common to all input data could be reduced during the training process, our proposed model can enhance the weight of the features unique to each MRI spatial sequence (i.e., the features of pituitary tumor), and meanwhile reduce the dimensionality of the features of each slice. This can greatly accelerate the operational speed of the subsequent classifier. Therefore, it is essential for our classification method to use the proposed Auto-Encoder-based framework for feature extraction.

2.3.1 ENCODER AND DECODER BASED ON DENSE BLOCK AND RESIDUAL BLOCK

For encoder, we use Dense Block to enhance the feature propagation ability of MRI spatial sequences, rely on the convolutional layer and pooling layer to reduce the dimensionality, and combine them to form an encoder for extracting the common features of MRI spatial sequences. As shown in Figure 8, the encoder uses two dense blocks in the training process (only one is shown in the figure). Due to the fact that the feature maps are superimposed during the training process, it enhances the propagation ability of pituitary tumor features, which consequently improves the accuracy and reliability of feature extraction.

See Figure 8

For decoder, we use Residual Block to compress the dimensionality of the feature map, and rely on the upsampling layer and the convolution layer to increase the dimensionality. These two components together form the decoder which can generate MRI spatial sequences with the same dimensionality as the original input data. The network architecture is shown in Figure 9. It also uses two residual blocks in the decoder (only one is shown in the figure). The decoder uses shortcut to lower the weight of some

features during the training process. Also, the model drift increases due to the added network depth (after adding the decoder). These together improve the effectiveness of MRI spatial sequence reconstruction. It also means that it is quite meaningful to use Residual Block for image reconstruction in the whole model.

See Figure 9

2.3.2 ADAPTIVE OPTIMIZATION OF FEATURE EXTRACTION

The above Dense Block-based encoder and Residual Block-based decoder enable us to adopt an Auto-Encoder model for sequence level feature extraction. Its specific network structure is as follows:

See Figure 10

Firstly, the input image sequence is generated by dense block encoder. Secondly, the feature sequence is decoded by the residual block-based decoder to restore the image sequence. At last, the input image is compared with the corresponding pixels of the generated image. The lower the loss, the more similar the generated image is to the input image, and the more representative the extracted feature sequence is.

2.3.3 SEMI-SUPERVISED CLASSIFICATION OF SPATIAL SEQUENCE IMAGES BASED ON CRNN

The extracted feature map of the pituitary tumor MRI spatial sequence is a three-dimensional matrix using the format of CRNN. An image sequence represents a patient and only one sequence-level label is needed. We first use CNN to extract the spatial feature sequence of the feature map, and then use RNN to train the extracted feature sequence. The neural network architecture is shown in Figure 11:

See Figure 11

2.3.4 MULTI-SEQUENCE PITUITARY TUMOR CLASSIFICATION MODEL

Combining all the above neural network components, we obtain a model for classifying the multi-sequences of pituitary tumors. Its network architecture is shown in Figure 10. The model is capable of augmenting under-sampled T1 and T2 datasets, fusing sequences from multiple modalities, extracting features, and finally obtaining the accurate estimation of the softness level of pituitary tumor by using a CRNN-based classifier.

See Figure 12

Results

3.1 EXPERIMENT ANALYSIS

3.1.1 EXPERIMENT PLATFORM

Our experiments are conducted in the following settings. The operating system is Windows10, the processor is 2.10GHz Intel Xeon (dual core), the memory capacity is 64GB, the development environment

is PyCharm, the deep learning framework is Keras, the programming language is Python, and the graphics card is GeForce RTX 2080Ti (three cores).

3.1.2 CYCLEGAN-BASED MULTI-SEQUENCE DATA AMPLIFICATION

The dataset used in the experiment was pituitary tumors collected in a local affiliated hospital. Each patient had MRI data of OAX, OSAG and OCOR (In this paper, OCOR MRI data are used), and there were T1 and T2 two modes in OCOR MRI data. There are 374 patients in total, 152 of whom are labeled, with each associated with a grading label from the following two grades: soft texture and hard texture.

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 13. When the number of training reaches 90 epochs, the loss of the discriminator reaches its minimum and becomes stable.

See Figure 13

We use 152 patient data with 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 14.

See Figure 14

Figure 14 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.

3.1.3 SEMI-SUPERVISED PITUITARY TUMOR TEXTURE IMAGE CLASSIFICATION BASED ON ADAPTIVELY OPTIMIZED 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 15, and the curve of the single-modal baseline is similar.

See Figure 15

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 Dropout 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 Figure 16

See Figure 17

See Figure 18

As can be seen from Figure 16–18, 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. See Table 1 for details:

TABLE 1 PITUITARY TUMOR CLASSIFICATION ACCURACY

Multi-sequence	T1 domain	T2 domain
91.78%	89.24%	88.98%
±1.44%	±3.11%	±4.23%

3.1.4 COMPARED WITH OTHER MODELS

In the training process of neural network model, the quality of tumor feature extraction is the key to improving the classification accuracy. Compared with the feature extraction method of image omics,

deep learning can adaptively learn the tumor features from big data. Table 2 lists the performance of ours and some commonly used classification models on pituitary tumor MRI images. This further demonstrates the effectiveness of our method.

TABLE 2 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
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.

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 [Hong ZHU] 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 [Hong ZHU].

Competing interests

The authors declare that they have no conflict of interest.

Funding

This work is supported in part by NSFC [grant numbers 81771904 and 61828205].

Authors' contributions

Qianhao FANG and Yihe HUANG contributed equally to this work. Kai XU gave theoretical guidance to this work.

Acknowledgements

Not applicable.

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: An End-to-End Trainable Neural Network for Image-based Sequence Recognition and Its Application to Scene Text Recognition

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]. Available: <https://arxiv.org/abs/1703.10593>.
13. G, et al., Densely Connected Convolutional Networks. Presented at CVPR IEEE Computer Society 2017. [Online]. Available: <https://arxiv.org/abs/1608.06993>.
14. K, et al., Deep Residual Learning for Image Recognition. Presented at CVPR 2016. [Online]. Available: <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

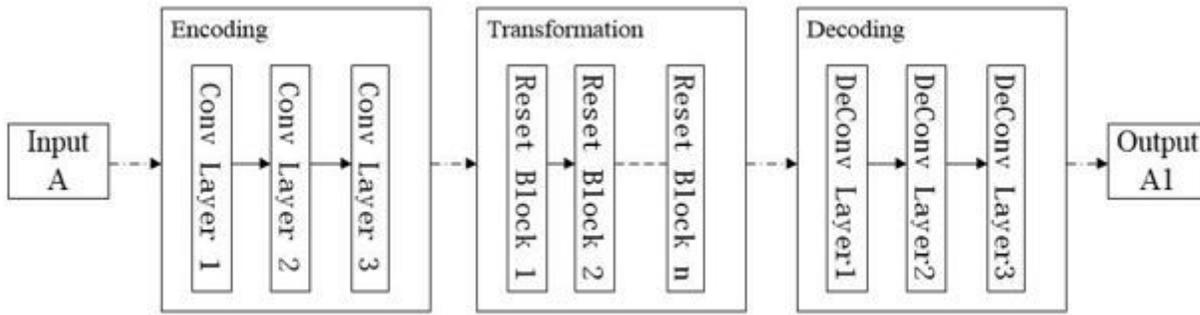


Figure 1

CycleGAN generator structure

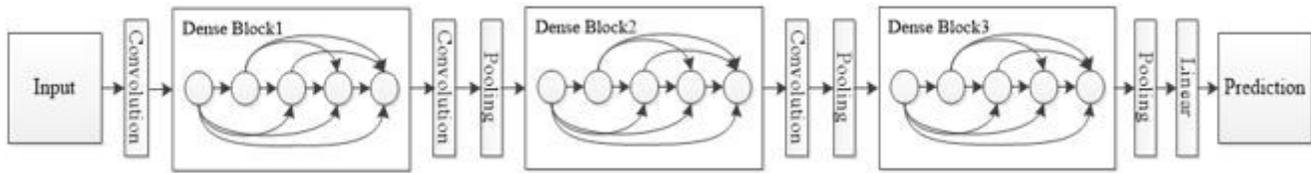


Figure 2

DenseNet structure

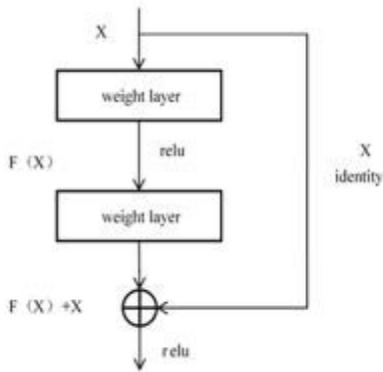


Figure 3

ResNet structure

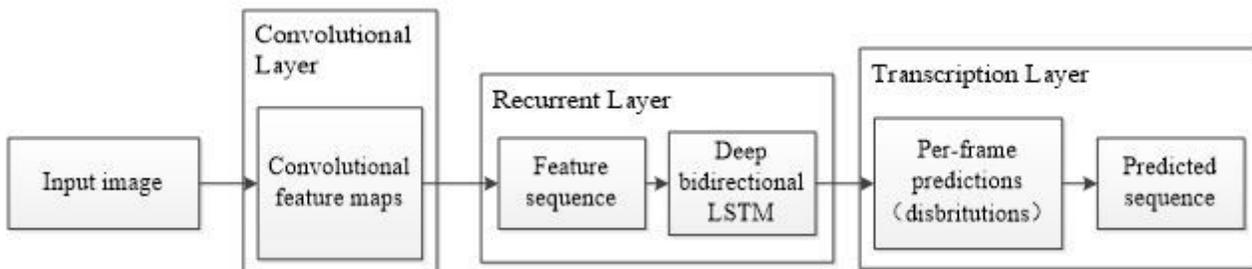


Figure 4

CRNN model structure

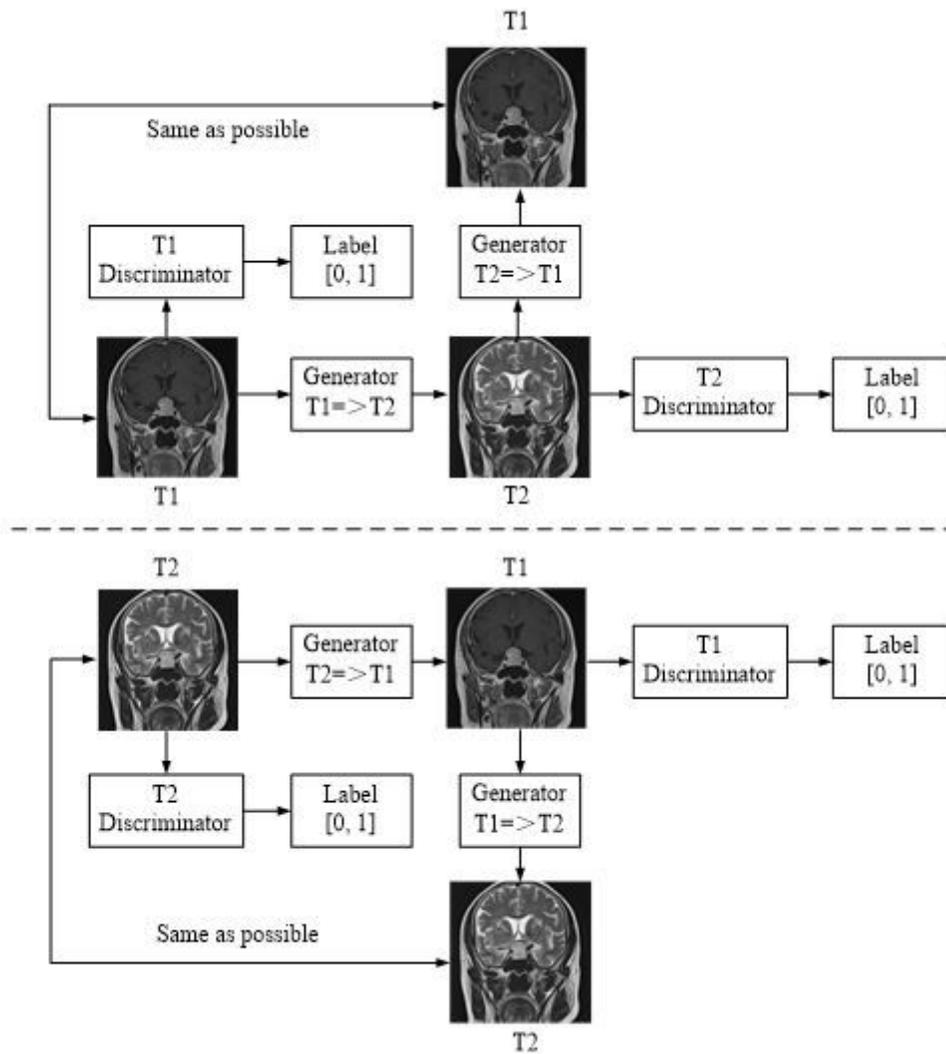


Figure 5

CycleGAN based Deep Neural Network Model for domain conversion

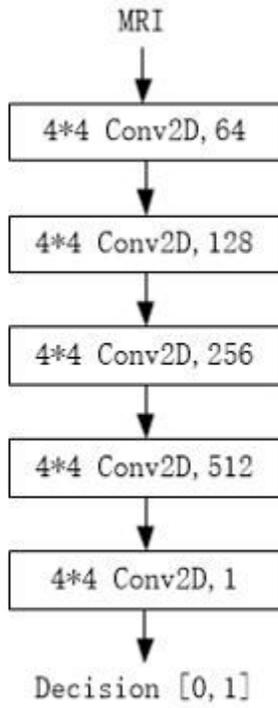


Figure 6

Discriminator architecture

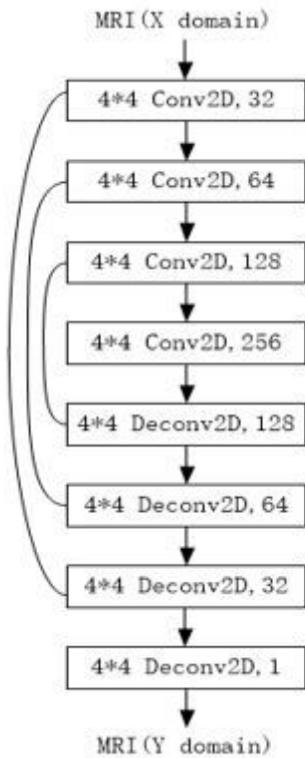


Figure 7

Generator architecture

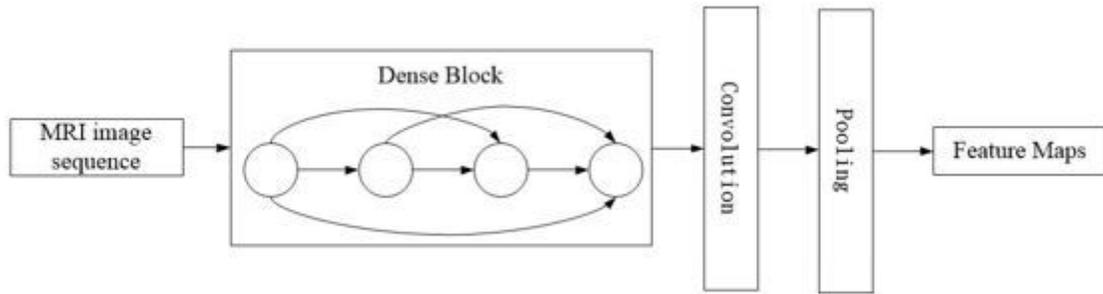


Figure 8

ResNet encoder network architecture

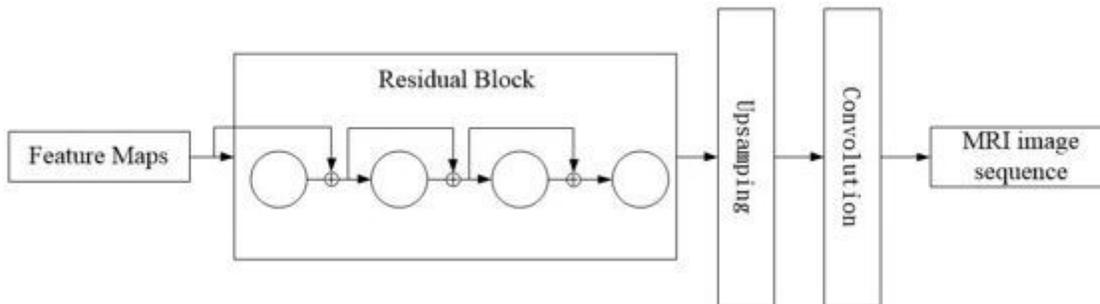


Figure 9

ResNet decoder network architecture

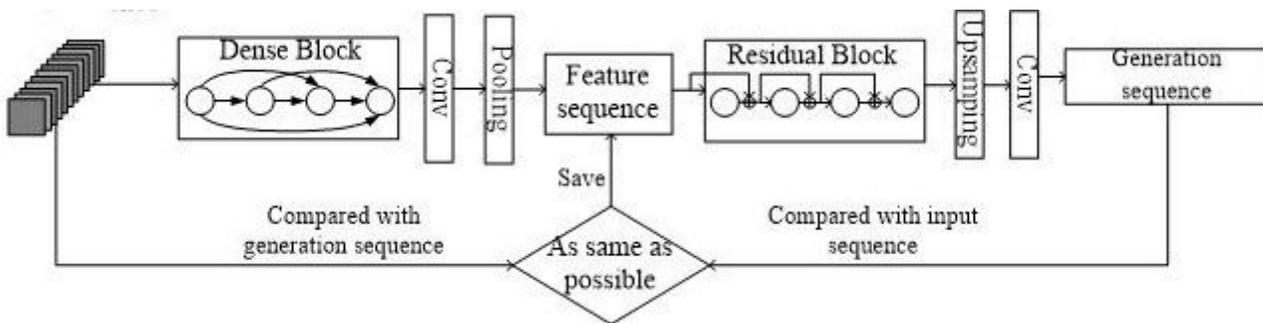


Figure 10

Feature extraction model based on Auto-Encoder

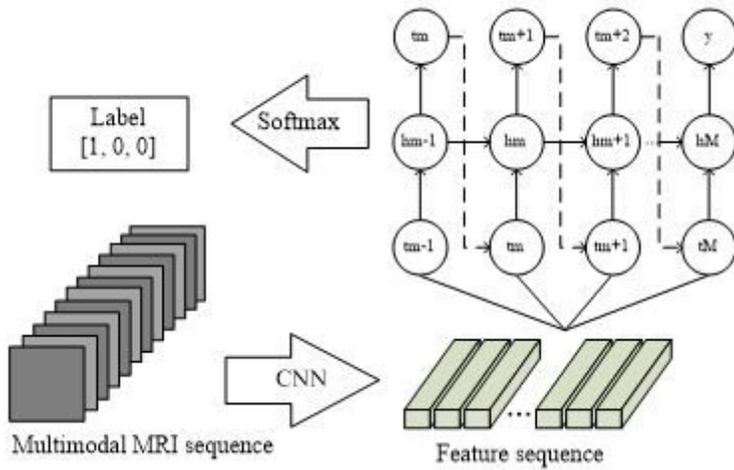


Figure 11

CRNN classifier network architecture

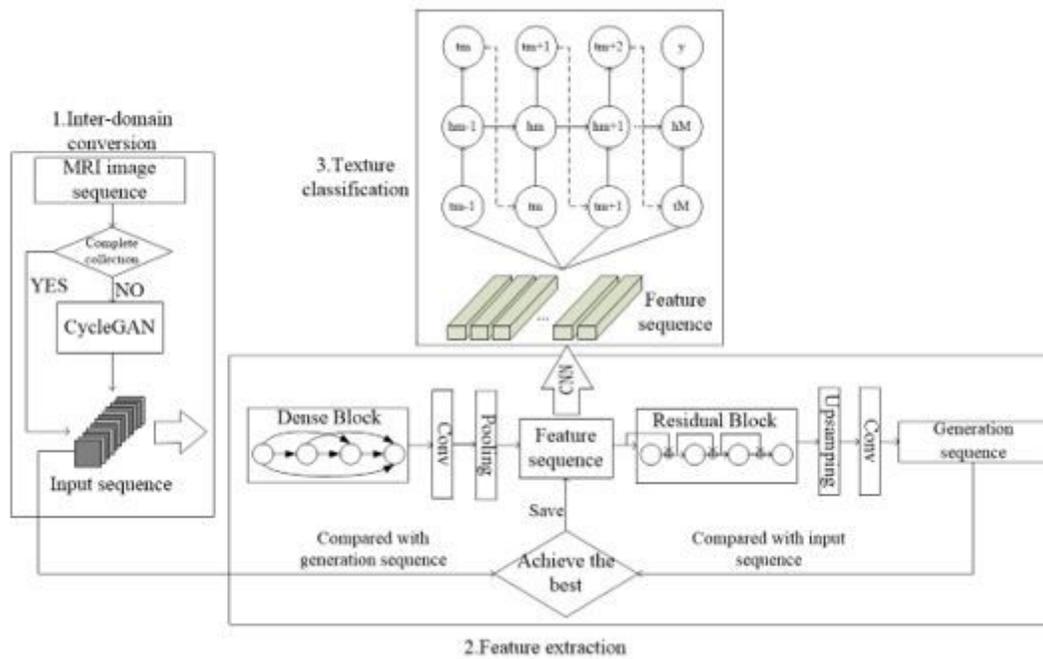
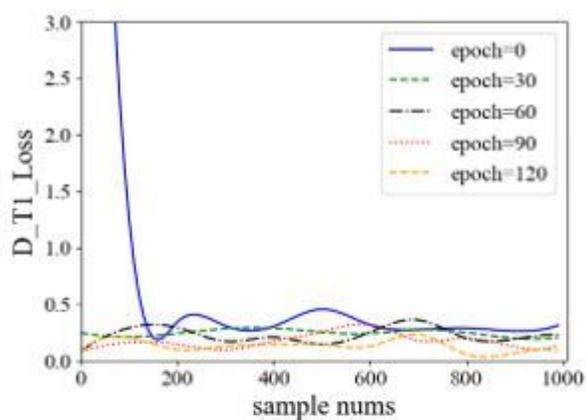
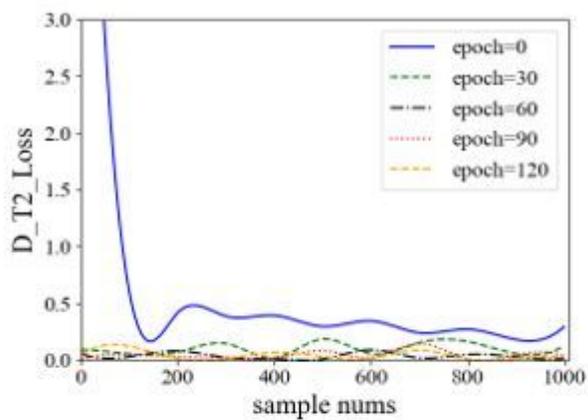


Figure 12

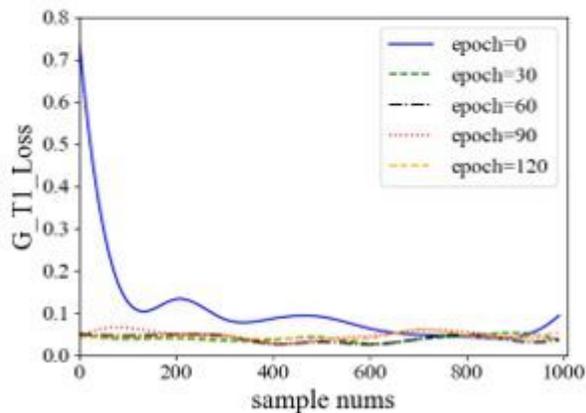
Multi-sequence pituitary tumor grading model



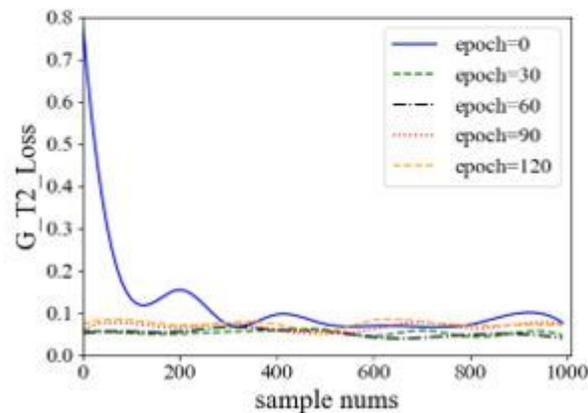
(a) T1 domain discriminator loss



(b) T2 domain discriminator loss



(c) T1 domain generator loss



(d) T2 domain generator loss

Figure 13

Discriminator loss and generator loss

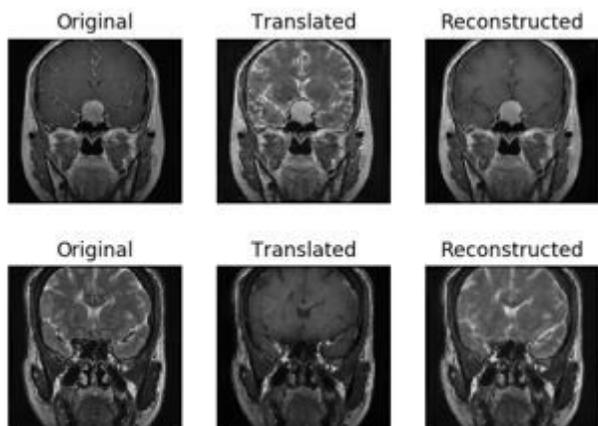
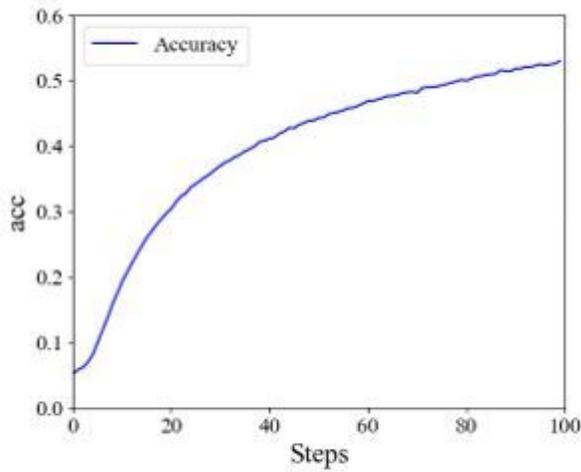
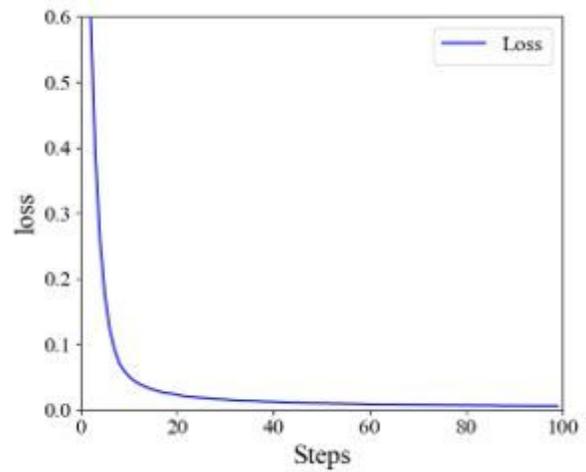


Figure 14

Visualization of training results



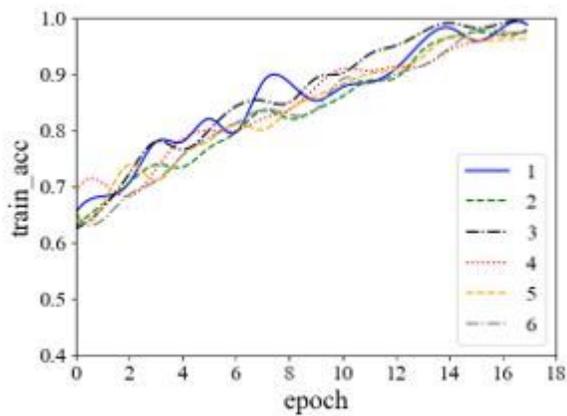
(a) Feature extraction accuracy



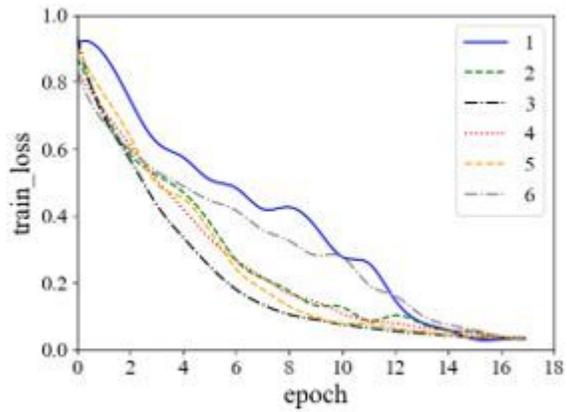
(b) Feature extraction loss

Figure 15

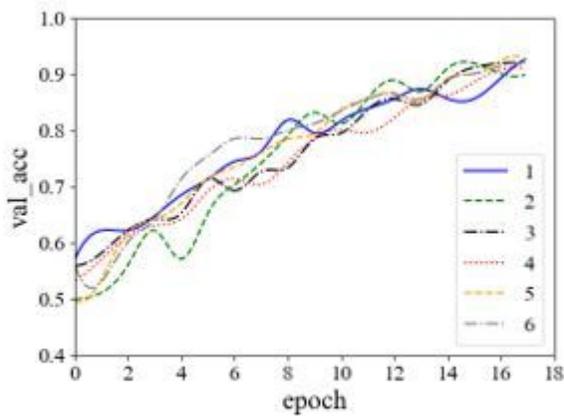
Multi-sequence feature extraction model



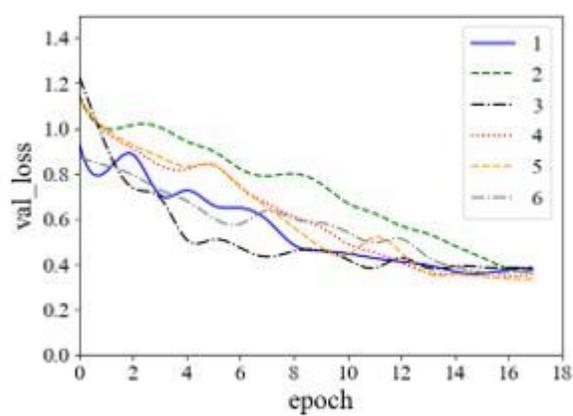
(a) T1 train accuracy



(b) T1 train loss



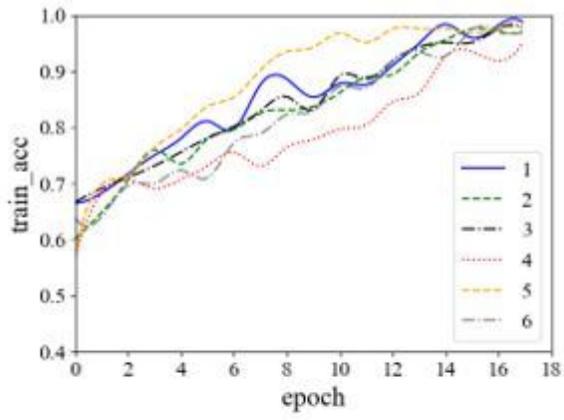
(c) T1 verification accuracy



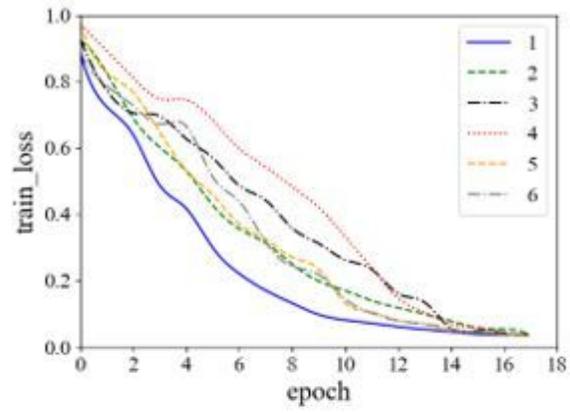
(d) T1 verification loss

Figure 16

T1 domain image classification model training



(a) T2 train accuracy



(b) T2 train loss

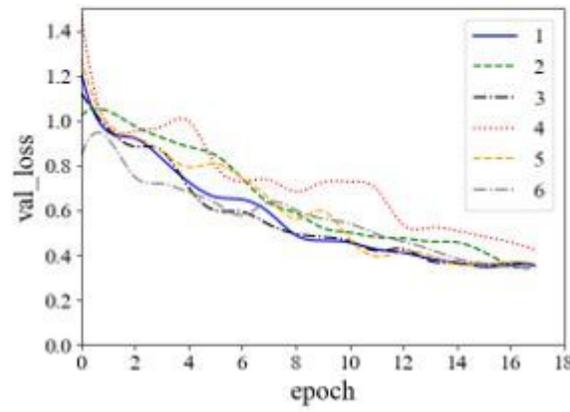
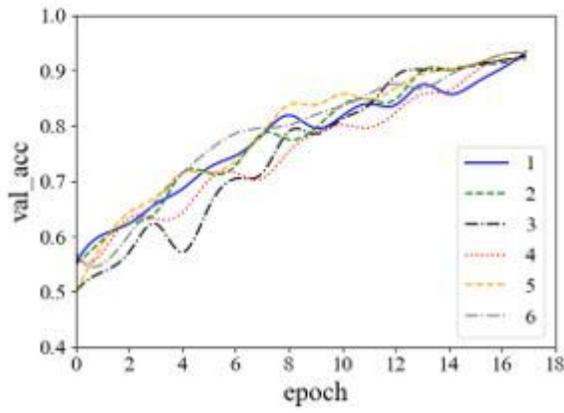
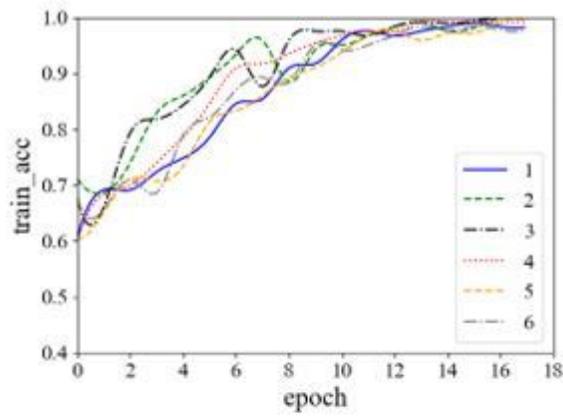
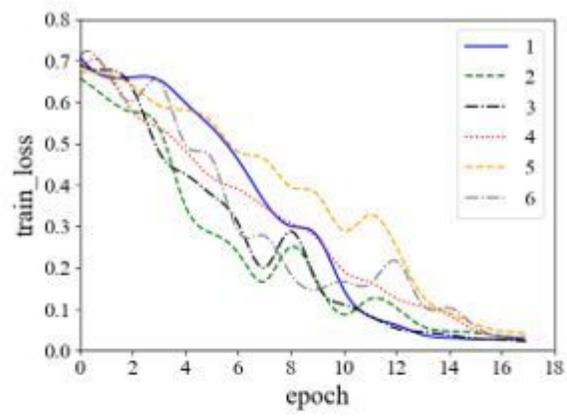


Figure 17

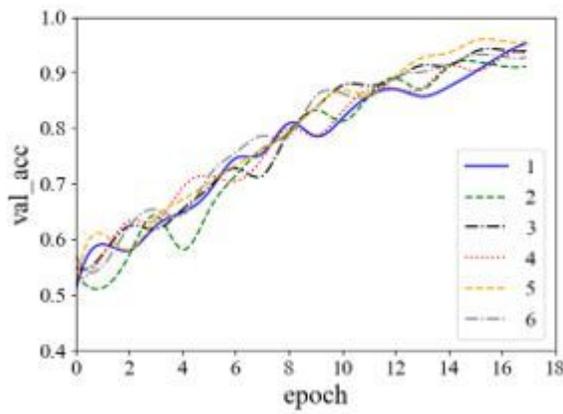
T2 domain image classification model training



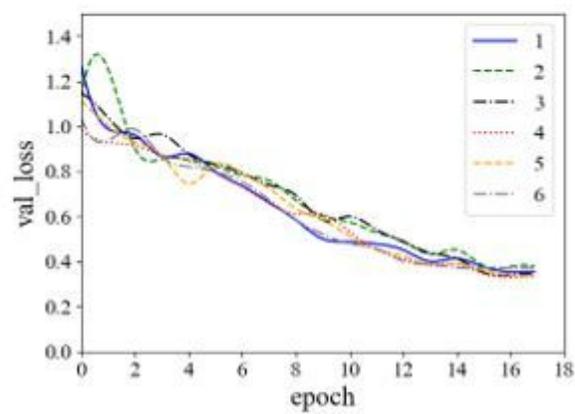
(a) Multi-sequence train loss



(b) Multi-sequence train accuracy



(c) Multi-sequence verification accuracy



(d) Multi-sequence verification loss

Figure 18

Multi-sequence image classification model training