Automatic Classification of MRI Contrasts Using a Deep Siamese Network and One-Shot Learning

Yiyu Chou*[a], Samuel W. Remedios[b], John A. Butman[c], Dzung L. Pham[a]

[a]Center for Neuroscience and Regenerative Medicine, Henry M. Jackson Foundation, Bethesda, MD, USA
[b]Department of Computer Science, Johns Hopkins University, Baltimore, MD, USA
[c]Radiology and Imaging Sciences, National Institutes of Health, Bethesda, MD, USA

ABSTRACT

Fully automatic classification of magnetic resonance (MR) brain images into different contrasts is desirable for facilitating image processing pipelines, as well as for indexing and retrieving from medical image archives. In this paper, we present an approach based on a Siamese neural network to learn a discriminative feature representation for MR contrast classification. The proposed method is shown to outperform a traditional deep convolutional neural network method and a template matching method in identifying five different MR contrasts of input brain volumes with a variety of pathologies, achieving 98.59% accuracy. In addition, our approach permits one-shot learning, which allows generalization to new classes not seen in the training set with only one example of each new class. We demonstrate accurate one-shot learning performance on a sixth MR contrast that was not included in the original training.

Keywords: magnetic resonance imaging, Siamese neural network, contrast classification, one-shot learning

1. INTRODUCTION

The rapid advancement in medical imaging technology has led to a large growth of data generated for disease diagnosis and research trials. Proper contrast identification is often a requirement in multi-contrast image processing pipelines for defining parameters and associating the appropriate training data. Furthermore, retrieval of clinical cases from medical archives is sometimes required for instruction or interpretation. It is challenging to index the associated clinical cases automatically, efficiently, and accurately due to the diversity and heterogeneity of MR sequences and naming conventions. In addition, because the task of labeling these images is performed manually, it can be prone to errors or even removed due to anonymization procedures[1,2]. For MR images in particular, differing acquisition protocols during a scan result in different image contrast properties. The ability to automatically distinguish between these contrasts allows large image archives from multiple sites and scanners to be organized into broad categories for efficient indexing and/or processing, especially when image meta-data can be inconsistent between sites and scanners.

Standard approaches to contrast identification use textual identifiers that are typically assigned within the DICOM metadata of medical images[4]. As an alternative, several approaches have been proposed in the literature to directly classify modalities or MRI contrast based only on the image content. In our previous study[3], a 3D convolutional neural network called PhiNet was presented to classify different contrasts of MR brain images and achieved a mean 97.57% accuracy across 3 tasks, including T1 vs T2 vs FLAIR, pre-contrast T1 (pre-T1) vs post-contrast T1 (post-T1) and pre-contrast FLAIR (pre-FLAIR) vs post-contrast FLAIR (post-FLAIR). High levels of accuracy were also reported in Pizarro[5]. Chiang[6], demonstrated the ability for deep learning approaches to also distinguish anatomical location as well as modality.

In this paper, we propose an efficient and robust approach that can automatically classify pre-T1, post-T1, T2, pre-FLAIR and post-FLAIR brain images based on a deep Siamese neural network with a triplet loss. Because this approach performs metric learning rather than direct classification, it can generalize to new categories, unseen in the training process, with limited examples (one-shot learning). Siamese networks are a type of deep learning network composed of a set of two or more sub-networks having the same architecture with shared weights. The parallel network architecture defines an embedding space such that images of the same contrast cluster closely within the space, while different contrasts are distant. Siamese networks have been applied to various problems, including image recognition and verification, visual tracking,

* - yiyu.chou@nih.gov

Medical Imaging 2022: Image Processing, edited by Olivier Colliot, Ivana Išgum, Proc. of SPIE Vol. 12032, 120320G · © 2022 SPIE
1605-7422 · doi: 10.1117/12.2613052

novelty and anomaly detection, one-shot and few-shot learning[7,8,9,10,11,12]. The source code, as well as a trained model for MRI contrast classification, is available on Github: http://github.com/chouyiyu/deepMRImgContrast.

2. METHOD

2.1 Dataset

In total, 3916 image volumes with various resolutions were obtained retrospectively and from public sources, representing 4 different sites and 5 different scanners: GE 3T, GE 1.5T, Philips 3T, Siemens 3T, and Siemens 1.5T. Five different contrasts of MR brain images including pre-T1, post-T1, T2, pre-FLAIR and post-FLAIR were acquired from not only healthy volunteers but also patients with traumatic brain injury, hypertension, multiple sclerosis, and Alzheimer's disease. Examples of these five contrasts of brain MR images and the distribution of training and testing data were shown in Figure 1. To preprocess the images, the neck regions were first removed from each image volume using FSL robustfov. The images were then resampled to 2x2x2 mm^3 to improve the processing speed for the neural network. Finally, each image was rigidly registered to MNI space using the ANTs software package[13]. In addition, 442 proton density (PD) weighted images were also available but were held out to demonstrate the one-shot learning capabilities of this approach.

MR contrast	Pre-FLAIR	Post-FLAIR	Pre-T1	T2	Post-T1	Total

Training	659	480	978	714	687	3518
Testing	51	55	116	90	86	398

Figure 1: Example image of each MR contrast and the distribution of number of images for training and testing

2.2 Siamese Neural Network Architecture

A Siamese neural network with a triplet loss is a type of deep learning network structure that contains three identical subnetworks (encoders) used to generate feature vectors for each input and compare them[4]. As illustrated in Figure 2, in each iteration of training, the input triplet (A,P,N) is sampled from the training set, where a baseline (**anchor**) input **A** is compared to a **positive** input ***P*** (same class as the anchor) and a **negative** input ***N*** (a different class from the anchor). Then the triplet (A, P, N) is fed into the encoder network simultaneously to obtain their latent embeddings. The loss of a triplet (A, P, N) can be formulated as[14]:

$$L(A, P, N) = max\,\{0, d(A, P) - d(A, N) + m\},$$

where $d(x, y) = ||f(x) - f(y)||_2$ is the Euclidean distance between the latent vectors of image x and image y; $\boldsymbol{m}$ is the hyperparameter that controls the separation between similar and dissimilar vectors in the latent embedding. The triplet loss function encourages large distances between anchor and negative images while minimizing the distances between anchor and positive images, thereby learning to differentiate similar images from non-similar ones. With the triplet loss function, not only are inter-class features differences enlarged, but also the intra-class feature variations are reduced, allowing the discriminative power of the deeply learned features to be enhanced. As we will show, these features are sufficiently generalized even for distinguishing new unseen classes.

At inference time, the input image (**query**) of an unknown class is processed by the encoder to compute a feature vector in the latent space. This embedding is then compared with other vectors representing different image contrast clusters, known as the **support set**. This provides us with similarity scores or relative distances between the image with an unknown contrast and all of the existing clusters. To obtain a classification result, the image contrast with the highest similarity (shortest distance) is selected.

3. RESULTS

The proposed Siamese network was implemented on a Linux server using Keras[15] with sixteen 32GB NVIDIA Tesla V100 graphics processing units and trained for 200 epochs with a batch size of 32. To stabilize the training process, the Adam optimizer was used with a learning rate 0.0001. The total training time was approximately 30 minutes and run time was

0.1 seconds. Given a support set with 3 reference images per class, the proposed model achieved 98.59% classification accuracy outperforming a deep learning method (PhiNet[3], 98.34%) and a template matching method (92.16%) where each test image is deformably registered[13] to a template pre-T1, post-T1, T2, pre-FLAIR and post-FLAIR image. Pearson correlation coefficients were computed between each registered test image and the five templates; the template having the highest correlation was labeled as the contrast of the test image. Figure 3 shows the per-class sensitivity for each contrast. Figure 4 shows all the incorrect classifications by the proposed method, including 5 pre-FLAIR (image A to E) and 9 post-FLAIR (image F to N) images. No pre-T1, post-T1 and T2 images were misclassified.

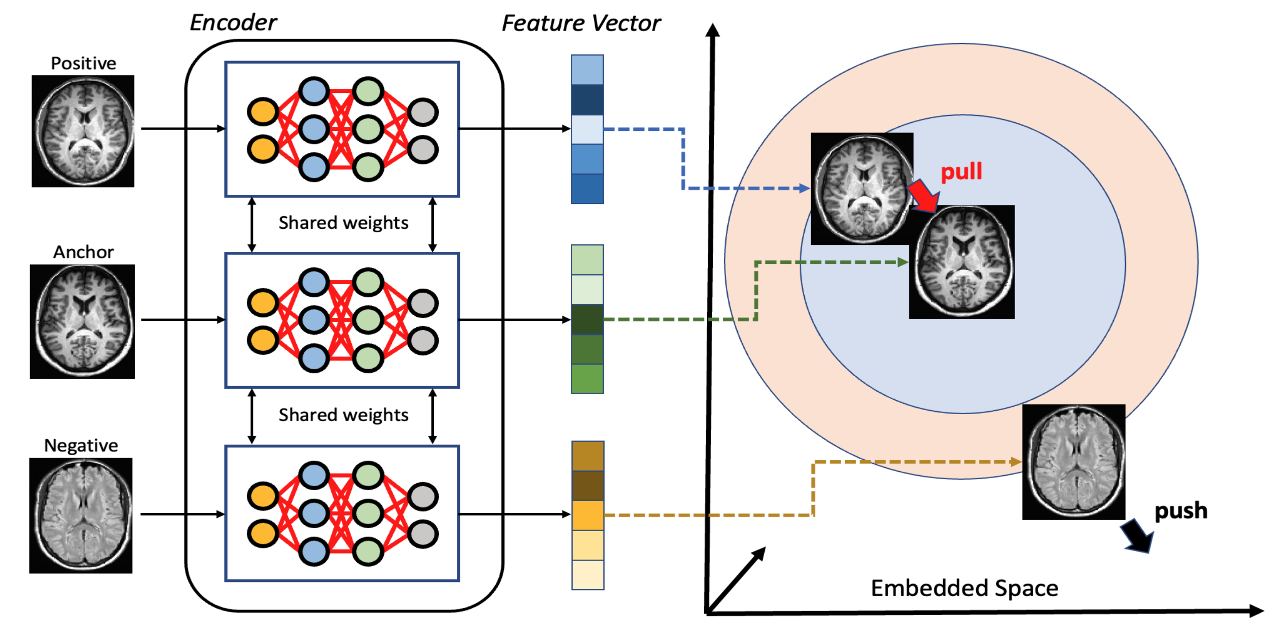

Figure 2. Siamese neural network with triplet loss function during training. A triplet of images (**Anchor**, **Positive** and **Negative**) are given to three identical encoder networks (one input layer + 3 hidden layers with 16 filters and 3x3x3 kernels + one fully-connected output layer with 32 dimensions). Training pulls embedded images of the same class closer together and pushes different classes further apart.

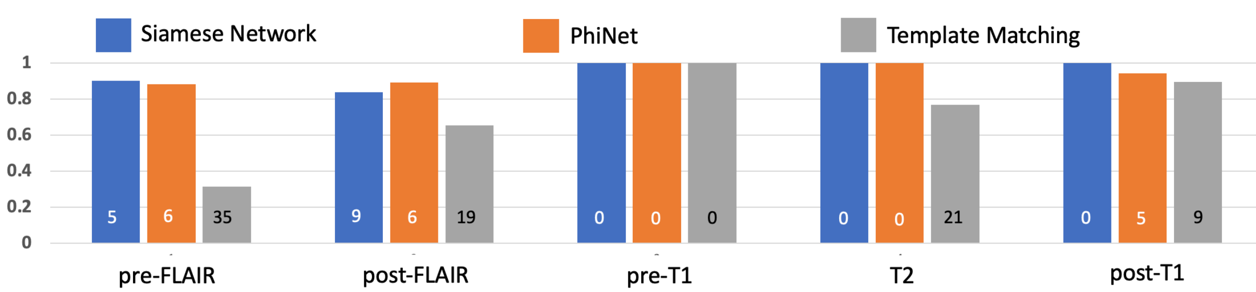

Figure 3: Classification sensitivity of the proposed Siamese neural network, PhiNet and the template matching method. The values denote the number of incorrect predictions for each MR image contrast.

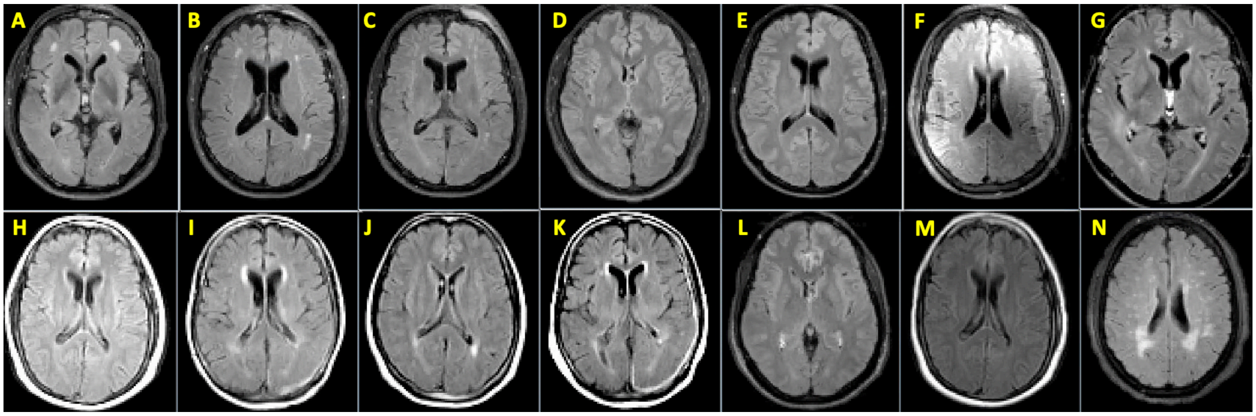

Figure 4: Incorrect predictions; Images A-D were pre-FLAIR but misclassified as post-FLAIR; image E was pre-FLAIR but misclassified as pre-T1. Images F to N were post-FLAIR but all misclassified as pre-FLAIR.

For data visualization, all the embedded vectors were projected from 32 dimensions into a 2-dimensional space using PCA with each color representing a distinct class as shown by the legend (Figure 5). We can see the embeddings of different classes are mixing before training since the model has not learned to separate the classes out. After training, we can see clear clustering of the intra-class images and better separation of the inter-class images. These plots indicate the model has learned to cluster the MR images for all 5 MR imaging contrasts even after reducing the dimensionality of the image features. The greatest overlap exists between the pre-FLAIR and post-FLAIR images, which is not surprising because these two classes are difficult to distinguish, even by an expert.

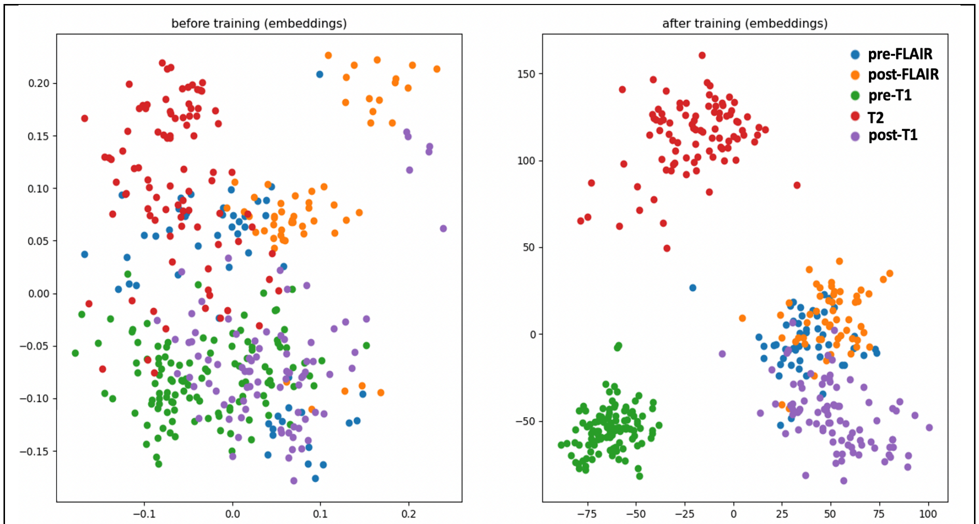

Figure 5: Plots of the embedding vectors projected down to 2-dimensions using PCA before and after the training.

The Siamese network can be used for one-shot classification by learning to discern a new class given only a single example without re-training. To demonstrate the discriminative potential of the learned feature mappings at one-shot classification, we reduced the support set for each MR contrast to 1 image and included PD images, with one of the images serving as the reference image. All other 441 PD images were used for evaluation. The overall classification accuracy remained high at 98.49%, and the classification sensitivity for PD was 100%.

4. NEW OR BREAKTHROUGH WORK TO BE PRESENTED

In this study, we present a deep learning approach to MR brain contrast classification based on a Siamese neural network demonstrating superior performance. The architecture creates a low-dimensional embedding space for MR contrasts by mapping images with the same class to nearby points in a low-dimensional space using a triplet loss function. The proposed method outperforms the traditional convolutional neural network method, and template matching method, only misclassifying 14 pre-contrast and post-contrast FLAIR out of 398 images that can be very difficult for a human expert to identify. Although the performance gain over traditional deep learning was small, an important advantage of the proposed method is that its discriminatory power can be generalized without any retraining for new image contrasts. We showed that given sufficient initial training data to define the embedding space, only one example was required as a reference for classifying a completely new MR image contrast. This approach is referred to as one-shot classification, drastically reducing the need for labeled datasets. In addition, we expect the model is more robust to class imbalance as it can be used on a dataset where very few examples exist for some classes.

5. CONCLUSIONS

The proposed approach achieved high accuracy in the classification of MRI contrasts. Future work will further examine the performance in data sets with differing pathologies and acquisition protocols, and the ability to use one shot learning to address data sets where the algorithm performs inaccurately. We will also examine the sensitivity of the approach to the selection of hyperparameters, such as the dimensionality of the embedding space and the size of the support set.

REFERENCES

(1) Kalpathy-Cramer J, Hersh W. Automatic image modality based classification and annotation to improve medical image retrieval. Stud Health Technol Inform. 2007; 129:1334-1338.

(2) Gueld MO, Kohnen M, Keysers D, Schubert H, Wein BB, Bredno J, Lehmann TM. Quality of DICOM header information for image categorization. Proceedings of SPIE- The International Society for Optical Engineering. 2002; 4685.

(3) Remedios S, Pham DL, Butman JA, Roy S. Classifying magnetic resonance image modalities with convolutional neural networks. Proceedings of SPIE- Medical Imaging 2018.

(4) Gauriau R, Bridge C, Chen L, Kitamura F, Tenenholtz NA, Kirsch JE, Andriole KP, Michalski MH, Bizzo BC. Using DICOM Metadata for Radiological Image Series Categorization: a Feasibility Study on Large Clinical Brain MRI Datasets. J Digit Imaging. 2020 Jun;33(3):747-762.

(5) Pizarro R, Assemlal HE, De Nigris D, Elliott C, Antel S, Arnold D, Shmuel A. Using Deep Learning Algorithms to Automatically Identify the Brain MRI Contrast: Implications for Managing Large Databases. Neuroinformatics. 2019 Jan;17(1):115-130.

(6) Chiang CH, Weng CL, Chiu HW. Automatic classification of medical image modality and anatomical location using convolutional neural network. PLoS One. 2021 Jun 11;16(6):e0253205.

(7) Schroff D, Kalenichenko D, Philbin J. FaceNet: A Unified Embedding for Face Recognition and Clustering, 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

(8) Bromley J, Guyon I, LeCun Y, Sackinger E, Shah R, Signature verification using a “siamese” time delay neural network. Advances in Neural Information Processing Systems: 737-744 (1994)

(9) Rana S, Kisku DR, Face Recognition Using Siamese Network, Proceeding of International Conferences on Frontier in Computing and Systems. Advances in Intelligence Systems and Computing, vol 1255: 469-376 (2020)

(10) Bertinetto L, Valmadre J, Henriques JF, Vedaldi A, Torr PHS, Fully-convolutional Siamese networks for object tracking. European Conference on Computer Vision (ECCV 2016):850-865

(11) Koch G, Zemel R, Salakhutdinov R, Siamese Neural Networks for One-Shot Image Recognition, Computer Science 2015.

(12) Chou Y, Chang C, Butman JA, Chan L, Pham DL, Automated classification of resting-state fMRI ICA components using a deep Siamese Network. Machine Learning for Quantitative Neuroimaging Analysis, 2021 submitted.

(13) Avants BB, Tustison NJ, Song G, Cook PA, Klein A, Gee JC, A reproducible evaluation of ANTs similarity metric performance in brain image registration," NeuroImage 54(3), 2011.

(14) Wang J, Song Y, Leung T, Rosenberg C, Wang J, Philbin J, Chen B, and Wu Y. Learning fine-grained image similarity with deep ranking. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 1386–1393, 2014

(15) Chollet F, Keras, https://github.com/fchollet/keras, 2016.