Final Report, EE209AS

Rahul Shenoy, Dhruva Nathan

Submitted: March 28, 2019

Abstract

This work presents a convolutional neural network trained for image registration using a weakly supervised training scheme. While previous deep learning methods for image registration require image datasets with ground truth spatial transformations, the present work relies only on anatomical labeling as the ground truth, datasets of which are more readily available. Anatomical labels may be of organs, blood vessels, tumors, or other ad hoc structures. The proposed method requires anatomical labels for training, but inference may be performed with unlabeled image pairs of the same modality as the training pairs. A key advantage is the speed of inference in this method, because previous neural network-based methods required training and optimization for each image pair in order to do registration. Magnitude-based parameter pruning and quantization are applied to the final model to make it more amenable to hardware-based acceleration.

Introduction

Image registration is a common and important task, especially in the area of medical diagnosis. Registration allows physicians to use non-invasive imaging procedures to generate complete views of a patient's anatomy which would normally only be available by invasive methods. Traditional methods for image registration have attempted to maximize mutual information, or minimize intensity difference, between image pairs by use of affine transformations and scaling. Finding a similarity metric which is sufficiently robust for clinical use has been challenging. Alternative methods rely not on pixel/voxel intensity difference but on automatic feature extraction and comparison for registration. In this work, brain magnetic resonance images (MRIs) with multiple anatomical labels are registered using a convolutional neural network (CNN). This work uses a modified implementation of the CNN model presented in Hu et. al [1], one of several recent works which attempt to use unsupervised learning schemes with less stringent data labeling requirements than traditional methods [2]. While previous CNN-based image registration methods require independent optimization of parameters for each image pair to be registered, the CNN in this work can be trained once and then deployed for immediate inference of new image pairs. We also adapted the TensorFlow model available from the same authors [5] for a new medical image dataset.

Methods

The CNN architecture is shown in Figure 1. The network calculates the optimal dense displacement field (DDF) based on Dice Loss, as described in [1].

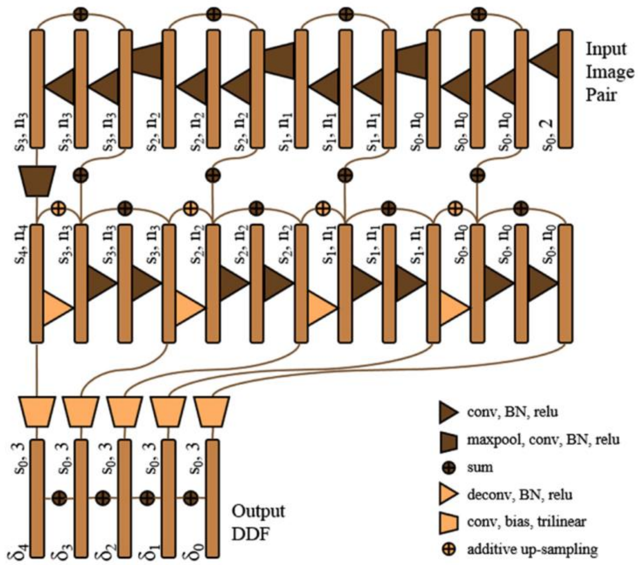

Figure 1: convolutional neural network for weakly supervised image registration. Adapted from [1]

The dataset used for training and images was obtained from Internet Brain Segmentation Repository (IBSR) [4]. 18 magnetic resonance imaging (MRI) images of a single patient brain with up to 32 anatomical labels. The anatomical labels in the datasets highlight morphological, anatomical, or physiological features present in both images. The IBSR dataset contains 32 such labels. These datasets are relatively straightforward to acquire because they rely simply on the knowledge of medical professionals, unlike spatial transformations which may require multiple iterations to obtain an accurate registration. Due to hardware and time limitations, only 5 labels of 32 were used to train the network.

The model was trained using 13 images, with 5 other images reserved for testing. During training, an image is randomly chosen as the fixed image and another is randomly chosen as the moving image for registration (so that there are 13*12 = 156 possible combinations). For each combination, the dice loss is computed for one randomly selected label. Using a mini-batch size of 4 image pairs (and 4 labels), the loss gradients are calculated and weights are updated. The training and inference strategy is outlined in Figure 2. During training, the fixed and moving images are passed through the CNN to generate the output DDF. Then, the DDF is applied to the moving label and the loss between the fixed and moving labels (not images) is computed. Parameters are updated according to the gradient of the losses, and the parameters are maintained for the next image pair. Although there are only 156 possible image combinations in the dataset, the training procedure requires many passes over the entire dataset, because only one of the labels is used in one training iteration, and the network must learn to recognize all 32 labels.

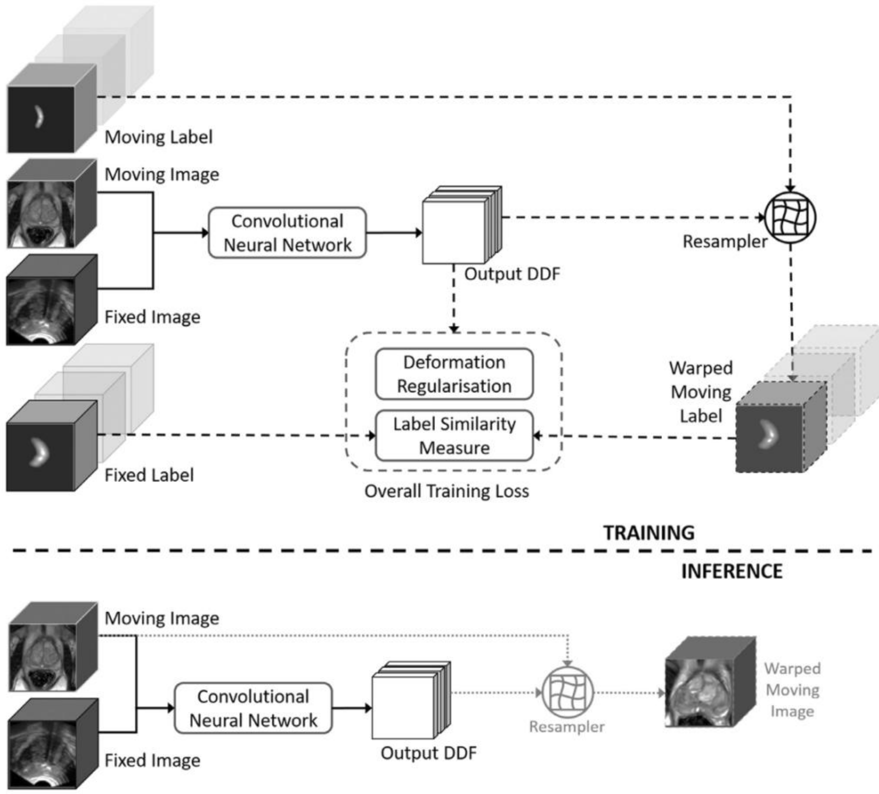

Figure 2: Training and inference schemes. Adapted from [1].

During inference, no labels are required. The fixed image is used to calculate the pre-trained DDF, which is then applied to the moving image and combined to create the final registered image.

For acceleration, A global magnitude-based pruning scheme was applied to the weights and biases in the model. After one iteration, the model was fine-tuned using a pruning mask to maintain the zero-weight of the pruned connections. Lastly, the model was quantized so that the parameters could be stored in 8-bit length. 25% of the parameters were pruned in one iteration, and the quantization from 32-bit to 4-bit resulted in a 4x compression in data size. Since the pruning and quantization was simulated only in software, the effective time required for inference and training was the same. These acceleration techniques are expected to improve the speed if implemented in hardware.

Results & Discussion

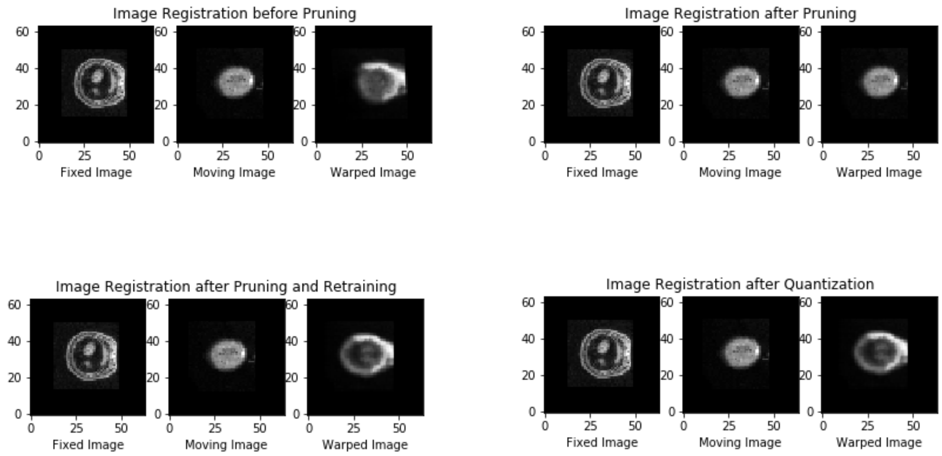

Figure 3: Example registered images after model optimization and acceleration

Model Type	Target Registration Error	Dice Similarity Coefficient
Trained Model	1.87	0.57
Pruned Model	1.01	0.62
Pruned, re-trained model	2.52	0.51
Quantized model	2.51	0.51

Table 1: Accuracy measurements after optimization and acceleration

In Figure 3 above, the registration of a randomly selected image pair is shown in three cases: after training, after pruning, after fine-tuning and after quantization. Qualitatively, the registration accuracy decreases slightly after both compression techniques

In Table 1, quantitative measures of the error are presented. We present the Dice similarity coefficient (DSC) and the target registration error (TRE). TRE is defined as the root-mean-squared distance between the fixed and warped labels for all the images in the dataset. DSC is defined as the overlap between the binary warped and fixed labels, using only one label. These errors increase an acceptable amount after the compression and acceleration techniques applied to the model.

References

[1] Hu, Y., Modat, M., Gibson, E., Li, W., Ghavami, N., Bonmati, E., ... & Ourselin, S. (2018). Weakly-supervised convolutional neural networks for multimodal image registration. *Medical image analysis*, *49*, 1-13.

[2] Li, H., & Fan, Y. (2017). Non-rigid image registration using fully convolutional networks with deep self-supervision. *arXiv preprint arXiv:1709.00799*.

[3] Hu, Y., Modat, M., Gibson, E., Ghavami, N., Bonmati, E., Moore, C. M., ... & Vercauteren, T. (2018, April). Label-driven weakly-supervised learning for multimodal deformarle image registration. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)* (pp. 1070-1074). IEEE.

[4] https://www.nitrc.org/projects/ibsr

[5] https://github.com/YipengHu/label-reg