Sciatic Nerve Segmentation in Magnetic Resonance (MR) Images of the Upper Leg via 3d Convolutional Neural Networks (CNN)

Matthew Hancock[1], Shashank Manjunath[1], Richard D. Dortch[1 2 3]

Introduction
Patients affect with Charcot-Marie-Tooth disease (CMT), an inherited neurodegenerative neuropathy disorder affecting the peripheral nervous system with a prevalence of 1:2500, commonly suffer from muscle degradation and impaired sensory ability. In the study of CMT1A, the most prevalent form of CMT characterized by the duplication of the peripheral myelin protein (PMP) 22 gene, the level of impairment correlates to the diameter of sciatic nerve measured from standard magnetic resonance (MR) image sequences [how to cite this?]. There is then need to accurately and reliably measure the diameter of the sciatic nerve in the upper leg, a manual labor-intensive process performed by extensively trained expert raters. Due to the size and nature of the sciatic nerve, inconsistencies in the designation of the voxels in the same MR image occur between raters. Our goal was to develop a efficient and consistent method to accurately segment the sciatic nerve in MR volumes voxel wise from the surrounding tissue using convolutional neural networks (CNN), an artificial neural network commonly used in computer vision (for image/video recognition, natural language processing, and medical segmentation) [delete].

Methods
[better introductory sentence] The U-Net CNN network architecture, proposed by Ronneberger et al. for both 2d and 3d applications [citation], diverges into a branched analysis and synthesis paths [reference figure?]. The analysis path down samples the input volume acquiring a larger filter receptive field as the input passes through four hierarchical resolution layers, each consisting of 2 convolutional layers with 3*3*3 [multiplication sign not *] voxel volumetric convolutional kernels with appropriate zero padding (to maintain the existing volumetric dimensions) [delete] followed by rectified linear unit (ReLu) activations. The number of feature channels is doubled in the second convolution increasing the trainable parameters in each layer [citation from Ronneberger], while the volume is down sampled with 2*2*2 stride 2 max pooling operations ensuring non-overlapping feature extraction. The ReLu operation is followed by a batch normalization layer normalizing the batch input to 0 bias inputted into each layer with internally computed mean and variance statistics. (The largest resolution level contains of the most trainable parameters and feature channels, representing the most highly abstracted features from the original input) [delete?]. The synthesis path is implemented similarly with 2*2*2 voxel up sample followed by convolutional layers with concatenations between the analysis to synthesis paths at corresponding resolutions preserving higher level features lost by down sampling [citation from Ronneberger].

Our custom implementation of the U-Net architecture was done in Python 2.7 using Keras 2.0 with Tensorflow backend. Patient data from 72 subjects (34 neuropathies, 38 controls), represented by a 256*256*40 voxel volume with a a voxel-wise hand rated region of interest (ROI) of the sciatic nerve, was randomly distributed between training (64%), validation (16%), and testing (16%). In order to reduce convergence time of the network due to the relative size of the region of interest to overall volume, each 48*48*40 volume fully encapsulating the ROI was cut from the volume. (Slight?) On-the-fly non-linear elastic transformations are applied to each 3d volume in order to reduce overfitting (due to the relatively small amount of available medical data) [delete?]. The cut-volumes are loaded into 5d tensors with user-defined batch size and fed into the network employing stochastic gradient descent (SGD) back propagation minimizing the negation of epoch validation dice coefficient with a learning rate and momentum of .00001/.99. The network trained on a NVIDIA Titan X GPU and when testing images with unknown ROI are inputted into a trained network, a volume with equal dimension is outputted with each voxel containing the probability of it being part of the sciatic nerve.

As implemented above, our experiment tested the validity of 4 distinct implementations of the U-Net network architecture: a 2d CNN trained with non-linear elastic augmented data where batches of MR slices are fed into a 2d representation of the same architecture, a 3d CNN trained with non-linear elastic augmented data, a 3d batch normalizing CNN trained with non-linear elastic augmented data, and a 3d CNN trained with unaugmented data. Three trials were conducted with distinct random partitions of subject data with all 4 nets trained on identical data within each trial. Each CNN was trained over 5000 epochs with a input batch size of 5.

Results
Data from the 3 trials (shown in Fig. X) distinguished by average successful (dice coefficient over .5), overall average dice coefficient, and average CMT patient dice coefficient. The distinction for successful segmentations was necessitated by the occasional complete misclassification (shown in Fig. X) of nerve particularly associated with CMT subjects with more complex neural structures and the CNN with no data augmentation as it was the greatest afflicted from training data over fitting. As shown, the best average network performance from all 3 metrics was the batch normalizing 3d CNN only failing to successfully segment a single image out of all 3 randomly selected testing data sets achieving high average performance on the rest of the control and CMT MR images. All 3 networks utilizing on-the-fly elastic transformations

outperformed the 3d CNN trained on unaugment data, while the 3d CNN trained on augmented data outperformed the 2d CNN trained on augmented data.

Discussion

The data shown above demonstrates the performance advantage of CNN utilizing 3d convolutional filters over 2d as they are able to capture the structural nature of sciatic nerve in 3d feature maps rather than constructing feature maps slice by slice only allowing the network to analyze features of a fragment of MR volumes at once. The improvement in dice coefficient in networks employing on-the-fly non linear elastic deformations also demonstrates the necessity for generating unseen, but structurally similar MR images while training the network to prevent it from becoming dependent on the precise structures of the limited training data set and thus unable to correctly classify the unseen testing MR images (this is captured in Fig. X as the training dice coefficient continues to improve, while the validation dataset, predictive training performance, does not).

Conclusion

CNNs are capable of producing quantitatively and qualitatively high quality segmentations in a reliable and efficient manner. Without the expert rater variability found in hand drawn sciatic nerve ROI, consistent estimations for the diameter of the nerve can be made allowing for predictions on expected patient debilitation.

References

1. Ronneberger, O. *et al*. Medical Image Computing and Computer-Assisted Intervention (MICCAI), 9351, 234-241 (2015).
1. Çiçek, Özgün *et al.* Medical Image Computing and Computer-Assisted Intervention (MICCAI), 9901, 424-432 (2016).
2. Ioffe, S. *et al.* International Conference on Machine Learning (PMLR) **37**, 448-456 (2015).

More…