CSCI E-82 Final Project

Abstract

White matter (WM) lesion prediction for Multiple Sclerosis (MS) patients is an important problem in MRI segmentation. It can help diagnose and monitor MS. Even for patients currently experiencing a lesion, the proportion of WM brain volume is very low (~1.5%). Combined with the high costs of labelling MRI data, this leads to a segmentation problem with scarce positive data.
With recent advances in deep learning, it is nowadays possible to sample data from very general distributions using generative adversarial networks (GANs). We approach the problem of lesion sampling using deep convolutional networks, use generative adversarial networks and semi-supervised learning to improve the accuracy of lesion segmentation.

Background

The goal of this work is to be able to localize lesions in multiple sclerosis (MS) patients. These lesions can be found mainly in the *white matter (WM)* tissue of the brain. Magnetic resonance imaging (MRI) is used to diagnose and monitor MS.

The problem in using MRI in this image (or better, volume) segmentation problem is that manual annotation of lesions is expensive and time consuming (a single MRI image consists of 512 x 512 x 512 pixels). Also, there is variation in manual annotations between different physicians.

This fact makes the problem at hand important to solve automatically, but also shows the problem: supervised machine learning usually requires a big number of labelled examples. Furthermore, only a relatively small proportion of the brain volume (~1.5%) is affected by a lesion and thus, the problem of lesion segmentation is *highly imbalanced*.

Our approach to improving on this imbalance uses *Generative Adverserial Networks (GANs)* [Goodfellow, 2014] and *semi-supervised learning* [Weston, Jason, et al., 2012].

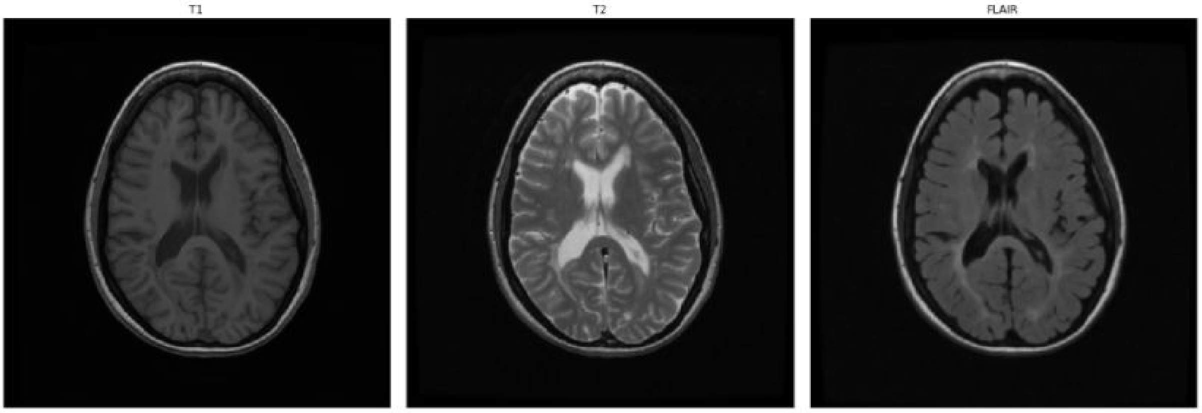
We train a *deep convolutional GAN (DCGAN)* [Radford et al., 2015] on patches around lesion voxels. This allows us to generate volumes of brain mass that look similar to lesions.

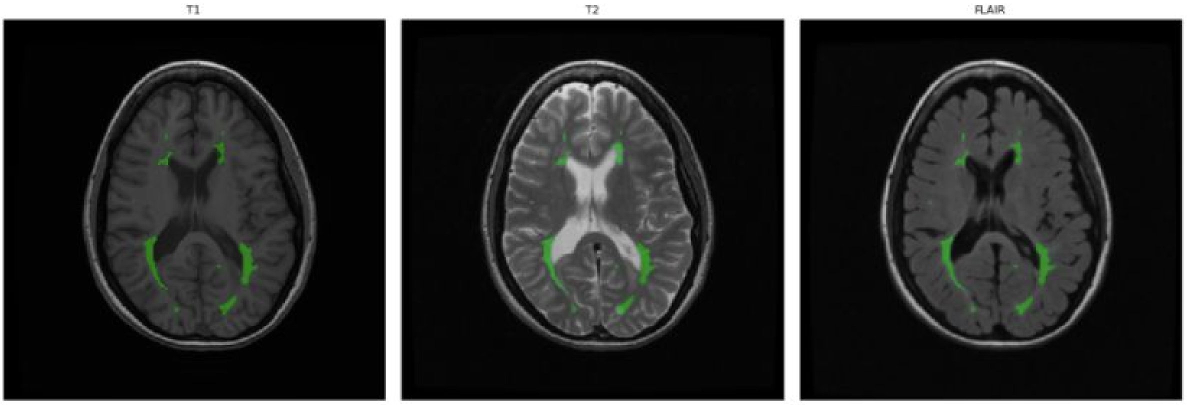
Then, we use *semi-supervised learning*. The general idea of semi-supervised learning is to provide, in addition to the labelled examples, also unlabelled examples. These unlabelled examples usually come from an unlabelled dataset that is "similar" to the dataset of interest. Alternatively, the unlabelled data can also be generated artificially with a GAN.

In our case, we use semi-supervised learning by training a classification network that has three output classes:

1. No lesion: this is the negative class and is sampled from real MRI images.
2. True lesions: this is the positive class and is sampled from real MRI images as well.
3. Fake lesions: these come from the generative adversarial network above.

Using this technique, we can improve the classification accuracy by enhancing the original dataset with sampled new patches.





Methods

Data

We mainly use the 2008 MICCAI MS Lesion Segmentation Challenge dataset. Note that this dataset is rather large (41GB).

The dataset was downloaded from the website above and uploaded to an S3 bucket for easy access.

Deep convolutional networks

Convolutional neural networks[LeCun, Yann, et al, 1998] have recently been shown to be the most powerful machine learning tool for image classification and image segmentation [Long, et. al, 2015].

Deep convolutional neural networks are very similar to more general artificial neural networks in that they consist of a sequence of matrix multiplications followed by a non-linear transformation (like the sigmoid function or a Rectified Linear Unit (RELU)). The main feature that distinguishes convolutional neural networks is that they make heavy use of the **translation symmetry** in image classification. This translation symmetry can be expressed more explicitly using the concept of **weight sharing** in the early layers of the neural network. It leads to a massive reduction in the number of trainable parameters, to much faster convergence and allows one to train much deeper networks (i.e. networks that have more layers) without overfitting the problem.

Generative adverserial networks

Generative adverserial networks (GANs) have been described as one of the most promising ideas in deep learning. One of the main purposes of a GAN is that it allows one to sample data from arbitrary data. This can be achieved by leveraging a special kind of architecture. The main idea of this architecture is that it can be split into two parts:

1. **Generator**: The goal of the generator is to sample data that resemble the original data distribution. It usually starts by sampling random numbers from some distribution (for example, a multi-variate Gaussian in dimension 100). Then this input is upsampled using transposed convolutional layers (in the case of image or MRI generation) to generate output of the right size.
2. **Discriminator**: The discriminator is a classification model. Its goal is to distinguish data from the original data distribution from samples that were generated from the generator. Its architecture can vary from the problem, but usually consists of convolutional layers. It has been shown empirically that **batch normalization**[Ioffe, Sergey, and Christian Szegedy, 2015] can greatly improve the overall performance of the GAN. Also, usually pooling layers are *not used* in the discriminator.

The discriminator has a fairly standard loss, but with a small twist, for the classification problem described above: the discriminator loss is the cross entropy for the discriminator logits and predicting the correct class. One does, however, employ **label smoothing** [Salimans, Tim, 2016|. This means that a real example has a label of 0.9 instead of 1.0. This has been shown empirically to give better results.

Each of these sub-networks has its own loss function: the generator has a cross entropy loss such that being recognised by the discriminator as fake is punished.

Semi-supervised learning

The general idea of **semi-supervised learning** is to use a mixture of labelled and unlabelled data. One of its main use cases it to improve the predictive power of supervised models when labelled data are not abundant.

In this case, one tries to find data that are similar (or maybe they even come from the same distribution). Then, a supervised classification problem is set up where one has an additional class. The model tries to predict the labels from the original supervised problem, but it receives and additional class label: if a sample comes from the unlabelled dataset, it gets the new label.

This approach can be combined with GANs as well, in various ways:

1. Train a GAN beforehand and then sample new data from the generator. The newly sampled examples receive a new class label.
2. The generator can be trained concurrently with the discriminator. In this case, the discriminator component of the GAN is the final result of the GAN training.

Note that the prediction step on real data needs to be modified in this case: only classes that have already been present in the original dataset need to be predicted. Thus, in the prediction step, we only keep the probabilities for those classes that have already been present in the original dataset and renormalize them so that the sum of the predicted probabilities is equal to one.

Results

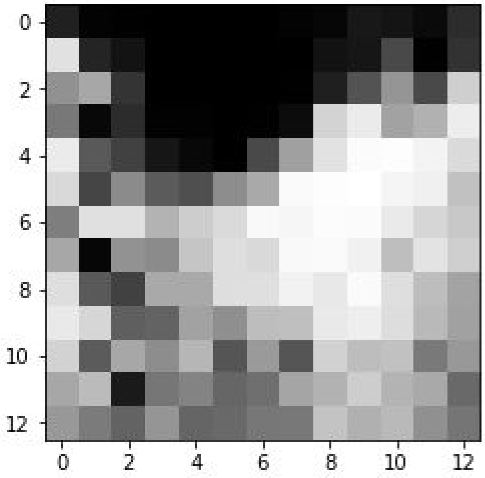
GAN

We successfully trained a GAN to generate patches looking similar to lesions. It has the following architecture:

1. We sample from an uncorrelated normal distribution of dimenison 100.
2. The so created random vector is passed through a fully connected layer of size 81 and reshaped to (3, 3, 3, 3).
3. A batch normalization layer is applied.
4. Afterwards, we pass the output through a leaky relu.
5. This vector is upsampled with a three dimensional transposed convolution to size (3, 6, 6, 6).
6. This output goes through a three-dimensional transposed convolution [Dumoulin, 2016]
7. Batch Normalization.

8. Leaky relu.
9. Another transposed convolution is applied with 'valid' padding.
10. Hyperbolic tangent.

The generated patches were of size 13x13x13. An example of a cross section can be found in the figure below.

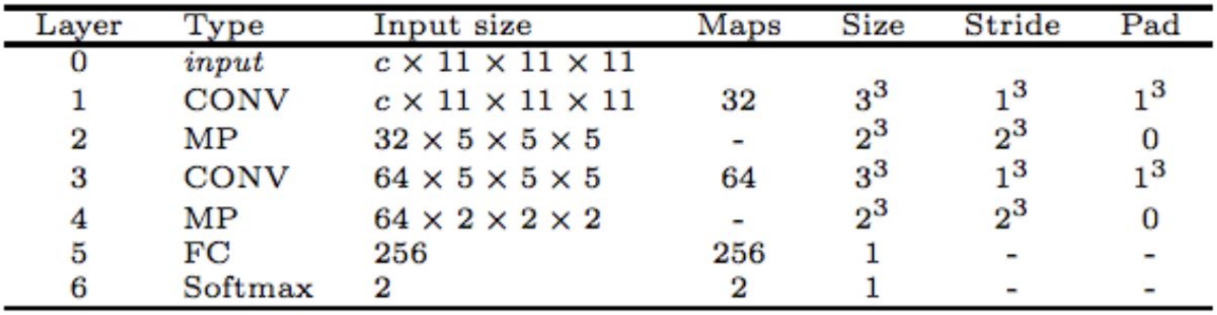


Classification

We implemented two architectures for the segmentation task. The first one took patches of size 13x13x13 as inputs, passed them through three convolutional layers and then two fully connected layers. Batch normalization was applied and the activation functions were leaky relus. The network could not be successfully trained. We subsequently used another network architecture.

The implemented Convolutional Neural Network was based on the approach described by Valverde et al that is the current approach leading the MICCAI challenge.

Each network is composed by two stacks of convolution and max-pooling layers with 32 and 64 filters, respectively. Convolutional layers are followed by a fully connected (FC) layer of size 256 and a soft-max FC layer of size 2 that returns the probability of each voxel to belong to the positive and negative class.



Layer	Type	Input size	Maps	Size	Stride	Pad
0	*input*	$c \times 11 \times 11 \times 11$				
1	CONV	$c \times 11 \times 11 \times 11$	32	3^3	1^3	1^3
2	MP	$32 \times 5 \times 5 \times 5$	-	2^3	2^3	0
3	CONV	$64 \times 5 \times 5 \times 5$	64	3^3	1^3	1^3
4	MP	$64 \times 2 \times 2 \times 2$	-	2^3	2^3	0
5	FC	256	256	1	-	-
6	Softmax	2	2	1	-	-

Discussion

In this project, we have presented a novel lesion segmentation method in brain MRI with application to MS patient images. The use of randomized balanced sampling helped address the large class imbalance issue normally faced with this type of problem.

The use of GANs and semi-supervised approaches (which we believe has not been tried before) helps address the issue of lack of training examples.

Conclusion and future work

Up to this point, we have been able to set up the machinery infrastructure to be able to further tune neural networks for lesion segments. Even though the results have not led to increased performance thus far, we are now in a good position to continue this research.

Additional avenues of future work include:

- Incorporating the GAN directly into the classifier
- Add more data sources
- Test transfer learning from BRATS
- Submit predictions to MICCAI and other competitions
- Test on *real* clinical data from ongoing trials

References

- Dumoulin, Vincent, and Francesco Visin. "A guide to convolution arithmetic for deep learning." *arXiv preprint arXiv:1603.07285*(2016).
- Goodfellow, Ian, et al. "Generative adversarial nets." Advances in neural information processing systems. 2014.
- Ioffe, Sergey, and Christian Szegedy. "Batch normalization: Accelerating deep network training by reducing internal covariate shift." International Conference on Machine Learning. 2015.
- LeCun, Yann, et al. "Gradient-based learning applied to document recognition." Proceedings of the IEEE 86.11 (1998): 2278-2324.
- Long, Jonathan, Evan Shelhamer, and Trevor Darrell. "Fully convolutional networks for semantic segmentation." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015.
- Nair, Vinod, and Geoffrey E. Hinton. "Rectified linear units improve restricted boltzmann machines." Proceedings of the 27th international conference on machine learning (ICML-10). 2010.
- Radford, Alec, Luke Metz, and Soumith Chintala. "Unsupervised representation learning with deep convolutional generative adversarial networks." arXiv preprint arXiv:1511.06434 (2015).
- Salimans, Tim, et al. "Improved techniques for training gans." Advances in Neural Information Processing Systems. 2016.
- Valverde, Sergi, et al. "Improving automated multiple sclerosis lesion segmentation with a cascaded 3D convolutional neural network approach." *NeuroImage* 155 (2017): 159-168.
- Weston, Jason, et al. "Deep learning via semi-supervised embedding." *Neural Networks: Tricks of the Trade*. Springer, Berlin, Heidelberg, 2012. 639-655.
- Zhong, Zhuoyao, Lianwen Jin, and Zecheng Xie. "High performance offline handwritten chinese character recognition using googlenet and directional feature maps." Document Analysis and Recognition (ICDAR), 2015 13th International Conference on. IEEE, 2015.