Robustness to Domain Shifts in MRI for Deep Learning-based Methods: A Review

LAYMAN'S SUMMARY

Ryan Pollitt

Most large Dutch hospitals have multiple Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) scanners among other medical imaging devices. These allow radiologists and doctors to see inside the human body and often generate a lot of data in the form of 3D scans of the body. To assist radiologists in their analysis of these large amounts of data, computer algorithms are increasingly used to make sense of this data. Specifically deep learning is often used, which is a way of teaching computer models called neural networks to recognize patterns in complex data. Neural networks have shown strong performance across a multitude of tasks, from aligning images to detecting tumours. To train a neural network, we often provide a neural network with an input image, e.g. an MRI scan, and the target output, which is the thing that we want the network to reproduce from the input image, e.g. an image where a brain tumour is coloured in. An optimization process then adjusts the parameters of the neural network in small steps such that the output gets closer and closer to the target output.

Although neural networks are based on biological neural networks as present in the human brain, they are far less advanced and often only excel at one specific task. More specifically, if only one type of data is used, networks will not generalize well to other types of data. An example of this would be a neural network trained to classify images into "cat" or "dog" based on real images of both. While this network could work well on real images, it would fail on drawings or paintings of cats or dogs, because these are too different from real images. This same issue applies to neural networks trained on medical images. It is especially true for MRI scans, because MRI machines allow the user to create images with a wide variety of contrasts, where one image might show fat with a high intensity and fluids with a low intensity and another might show the opposite. These shifts in image appearance can also be smaller and resultant from using MRI scanners from different manufacturers, which might look the same to humans, but different to neural networks. Finally, if a network is only trained on images of adults, it will often fail on images of children, because these also look different. These causes of changes in image appearance are called domain shifts.

We reviewed two groups of methods aimed at reducing the performance loss of neural networks under domain shifts to investigate how the adverse effects of domain shifts can be countered. The first is called data harmonization, which aims at removing domain-specific (e.g. scanner, type of image contrast) information from the images. The second method is called domain generalization, which instead aims to train neural networks in such a way that they work across images from multiple domains. This can be achieved by using advanced training techniques or by creating a large variety of input data. We reviewed five papers performing data harmonization and sixteen papers that used domain generalization.

Based on these papers, we found that data harmonization has more limited uses than domain generalization. Because the former often relies on knowing what domains (what kind of images) you want to apply a neural network to, it is less flexible than domain generalization, which does not have this limitation, theoretically allowing neural networks to be applied to a wide variety of domains. However, data harmonization provides the option of applying neural networks that have already been trained, without retraining on new data, so both methods have their use cases.

Of the various domain generalization methods we found that the relatively simple idea of providing the network with a wide variety of input data actually showed the most promising performance. Researchers often applied augmentations to existing data (stretching, rotating, adding noise to, or adjusting the contrast of real images) to increase data variety or in some cases created completely synthetic data based on segmentation maps, which encode if a location contains e.g. the hypothalamus, grey matter or white matter. By giving each of the anatomical regions in these segmentation maps a random intensity, many often unrealistic synthetic images can be created. Even though neural networks trained with this synthetic data had never been trained on real MRI scans, they still managed to perform well on real data. By training the network with a wide variety of input images, the neural networks were forced to focus more on the shape of anatomical structures, because the intensities were augmented or created randomly. This forced the network to "see" the images more like humans do, who can recognize e.g. the liver regardless of its intensity based on its shape, as long as there is enough contrast with the surroundings.

In conclusion, we found that increasing the variety of the input data is a promising research direction for robustness to domain shifts in neural networks, which is relatively simple to implement.