Examining the Use of Convolutional Neural Networks in Image Registration of Head and Neck Cancer Patients

9615544

School of Physics and Astronomy

University of Manchester

December 2018

Abstract

This experiment aimed to produce an implementation of a convolutional neural network (CNN) that was able to take planning CT scans and PET scans and perform an image registration without the use of labels. The time taken for registration to take place was significantly reduced over those provided by Elastix, with times of 5.76 ± 0.05 seconds for our network and 161.3 ± 5.6 seconds for registrations provided by the Elastix toolkit. However, the loss throughout the experiment remains high and so the output registered images are not accurate enough to be favourably compared to the Elastix registered images despite this improvement in computation time.

1 Introduction

Radiotherapy is one of the most common methods of treatment for all types of cancer in the UK with 27% of patients in 2013 and 2014 using this type of treatment for either palliative or curative therapy [1]. It is usually part of a multi-modality treatment meaning it is used in conjunction with other therapies such as chemotherapy, which uses cytotoxic chemicals to kill cancer cells, and surgery to reduce and remove the tumour [2].

Within radiotherapy itself ionising radiation is used in an attempt to kill cancerous growths in patients. There are several ways in order to deliver this radiation including the utilisation of an external machine, LINAC or linear accelerator, that attempts to deliver the largest amount of radiation to the cancerous region, preserve organ function and reduce side effects caused by the tumour [3]. Other options are to place a radioisotope inside the body either in the form of a radioactive metal, this is called brachytherapy, or to use radiotherapy injections capsules and drinks that are absorbed by the body [4]. The methods used in this experiment used historical externally applied radiotherapy treatments in an attempt to improve upon modern technologies.

Neural networks provide a method of using computers to perform tasks more quickly and potentially with greater accuracy [5]. One potential area that this could be used to improve upon current methodologies is in the area of deformable image registration. The aim of this study is to perform image registration on PET and CT images of head and neck cancer patients, the process by which PET and CT images are transformed and deformed so they match each other [6] [7] .

2 Radiotherapy and Image Registration

Traditionally, in order to plan radiotherapy, the patient has a Computed Tomography (CT) scan of the part of their body where the cancer is located, which is then used for planning. As not only does this improve accuracy of treatment but the intensity of each pixel correlates to the electron density of the tissue making the planning process much simpler [8] . The CT image is a 3D image taken of the body that is built from many 2D X-ray images being taken around a single axis of rotation as the bed moves through the bore of the scanner [9]. An example of a CT scan can be seen in Figure 1.

In addition to this, certain patients may receive a PET scan. PET stands for positron emission tomography and uses radioactive isotopes attached to glucose molecules within fludeoxyglucose (^{18}F) to image the body [11]. When this isotope decays it emits a positron that annihilates with an electron to produces two gamma rays travelling in opposing directions that are detected by the scanner, from this a 3D image of the patient can be built [12]. A PET scanner provides more information on metabolically active tissues meaning it provides better data on soft tissue than the CT scanner. In PET imaging the radioactive isotope used is attached to glucose which is used by muscles, and as these absorb more glucose more gamma rays will be produced at this sight meaning more detail is captured [13]. The tumour in particular draws a high amount

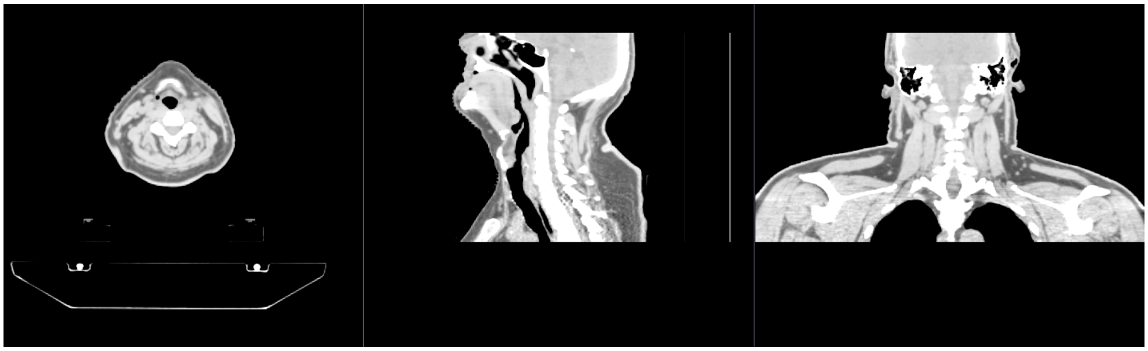

Figure 1: A CT image of a head and neck cancer patient taken from the Cancer Imaging Archive. From left to right we are given an Axial, Sagital and Coronal Image of the patient [10].

of glucose, because tumours undergo mitosis more frequently than other normal tissues in the body, a process that requires lots of energy [12]. This means that cancers are easy to identify when compared to a healthy PET scan. This allows for better staging of cancers thus providing detailed data on extent of disease and location. An example PET scan can be seen in Figure 2.

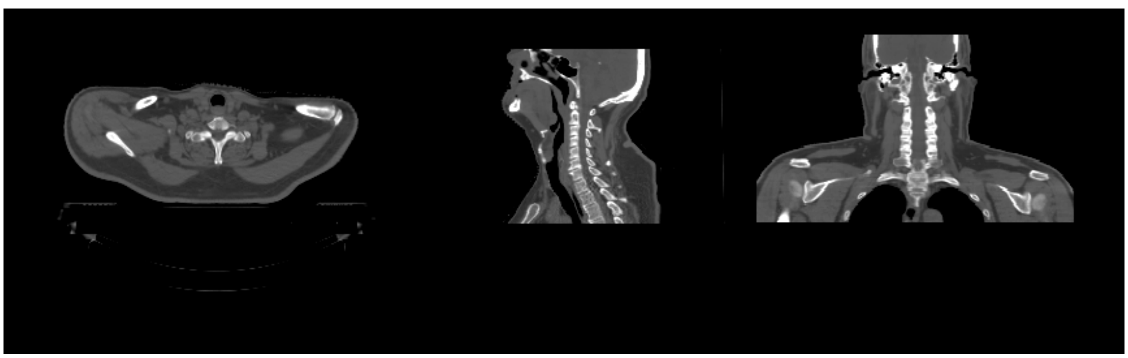

Figure 2: A PET image of a head and neck cancer patient taken from the Cancer Imaging Archive. From left to right we are given an Axial, Sagital and Coronal Image of the patient [10].

After diagnostic and planning images have been taken, contouring of the body takes place; this is a process where structures in the body are identified and outlined on concurrent slices of the CT [14]. For example, in the treatment of head and neck patients areas such as the brain stem, larynx and salivary glands are labelled. These PET and CT images will not however be identical as the patients' position will shift between the two images, in head and neck patients this occurs when in planning images the patients' necks are at a different angle due to the varying amount of cushioning provided for these two images [15]. The images need to undergo registration. This is a process whereby one image is deformed so it matches the other, this provides greater detail of the patient, combining the greater information on the soft tissue in PET imaging with the imaging of bony anatomy provided by CT scanning. During radiotherapy treatment planning three regions are defined in the images of the patient, there is a gross target volume, a clinical target volume, and planning target volume [16]. The gross target volume (GTV) is the total volume of the tumorous region, the clinical target volume (CTV) includes other tissues that may be infected and the planning target volume (PTV) which is a volume that accounts for planning uncertainties such as the movement of a patient during treatment. The combined image of the PET and CT scans reduces the margins (the size difference between the GTV , CTV and PTV) needed during treatment which improves the accuracy of treatment allowing for higher doses to be delivered thus improving patient outcomes [16]. Image registration is done by taking each voxel (an element of volume that constitute a three-dimensional space) in the PET scan and finding the corresponding voxel of the CT image, the transformation that is needed to move this image is then saved as a deformation vector field (DVF). This process is computationally intensive and can take up to two hours to produce the highest accuracy DVFs [6]. After image registration has taken place and an acceptable treatment plan produced treatment of the patient can begin. There are many ways that external radiotherapy doses are commonly administered including Intensity Modulated Radiotherapy (IMRT) which reduces the radiation exposure to surrounding tissues by using several beam angles that are chosen to avoid the most sensitive tissues [17]. One type of IMRT is Volumetric Modulated Arc Therapy (VMAT). This is where during each fraction the patient lies on the bed in the same position, while the beam of radiation is continually swept over them from many different angles, using these different angles means that a greater dose is delivered to the cancerous region and exposure to surrounding tissues is reduced. The rotation of the beam also avoids areas of tissue that are more sensitive to radiation [18] [19]. It is able to do this more accurately and precisely due to contouring. This highlights the need for correct image registration and contouring as more accurate information about the patient can reduce the amount of radiation given to other important structures in the body. For example, in head and neck cancer treatment it is important to limit exposure to the brain stem, as damage from radiation in this area could have severe consequences including necrosis of cells leading to paralysis.

3 Artificial Neural Networks and CNNs

Artificial Neural Networks are a method of using computers in order to perform classification problems such as pattern recognition and data classification. They are named neural networks as they are loosely based on the way the human brain works, though they cannot achieve the complexity that brains have reached [5]. There is a network of interconnected layers between an input and an output. Like mammalian brains these networks learn based on experience and examples fed to them allowing for them to improve at a task [20]. The idea of using artificial neurons for processing of information was proposed in 1943 [21]. This idea was further developed both theoretically and experimentally to recognise patterns throughout the 50s [22].

Neural networks often provide solutions more quickly than is possible using other methods. However not every problem is suitable for analysis using neural networks, certain criteria must be met when determining whether using a neural network an appropriate approach to a problem. One condition that must be met for applications for which neural networks are useful is whether learning from experience will provide a reliable method of predicting future outcomes. They require an abundance of data from which the correct way to find a solution can be learnt. They also require the ability to produce general results from the examples given to it [5]. A correctly implemented network is able to produce correct outcomes from examples that it has not yet been exposed to. One reason neural networks are considered so useful is that they do not require an intricate knowledge of the problem for which they are being used to solve. For instance, in a common example of training a network to identify the difference between cats and dogs the user does not have to know and implement in great detail the ways in which the computer will be able to identify the differences between the two groups, such as hair length and face size. However, a mathematical understanding of the statistical nature of these problems is vital to ensure the utility of the network [5]. One other benefit is the ability of these systems to provide non-linear solutions to problems, meaning they are useful for more complex real-world problems [20].

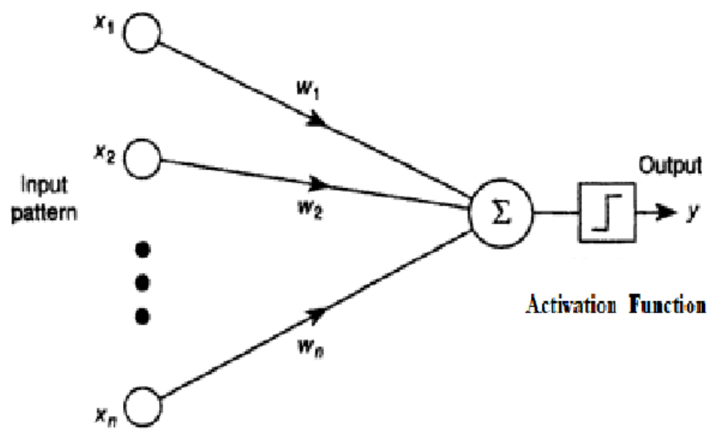

Figure 3: The structure of a neuron in a neural network, with several synapses converging on an adding function. It then passes through an activation function to produce an output [5].

As the name suggests artificial neural networks are comprised of neurons, these compute a weighted summation of its inputs to produce an output. A schematic of a neuron is shown in Figure 3. The neuron has several basic elements. The first of which is a set of links between each other element, these are known as synapses, each have a specified weight. They also have an adder for summing the input signals that have been weighted by the synaptic strengths of the neuron. Then an activation function that acts as threshold the output amplitude of the neuron, meaning little signal is passed through if the summed inputs don't reach this threshold. Also included is an externally applied bias which has the effect of raising the lowering the net input of the activation function [23]. This process is described mathematically by the equation,

$$y = f_a(\sum_{i=1}^{n} w_i x_i + w_0) \tag{1}$$

where y is the output of the neuron, f_a is the activation function of the network, w_i is the synaptic weight(strength) of each synapse, x_i is the input of each node and w_0 is the bias applied to the neuron [5].
Using the outputs generated by these networks can be used alongside a known correct value to produce a loss value, a value of the correctness of a solution. The next step is back propagation which allows the network to alter the weights and biases of each of the connections in its layers and produce an alternative value for the loss. The overall aim over many examples and iterations is to reduce the loss function to a minimum value without over-training, which is where the network becomes very good at solving for the training set of data but becomes unable to generalise to solve for other test data. In this way the computer learns the best way to solve the problem that is given to it.

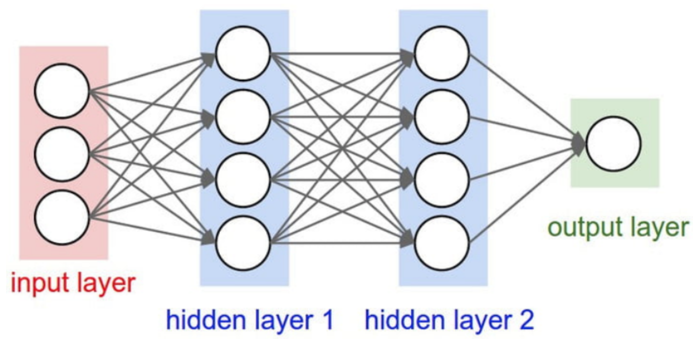

Figure 4: A more complicated neural network built with several layers of neurons. This network is still very basic and is not what would be described as 'deep'. In depictions of more complex neural networks only the layers are shown as showing every synapse is not feasible [23].

In order to solve more complex problems layers of neural networks are joined together as shown in Figure 4. In order to make neural networks more effective at different types of problems they are separated into different classes. Examples of different classes neural networks include the Radial basis function neural network and the convolutional neural network.
The convolutional class of neural networks are used to analyse images as they have a large number of parameters that would be difficult for other types of network to handle. The architecture of a convolutional neural network is shown in Figure 5.
As shown the CNN has three main types of layers; the convolutional layer, activation layer, the pooling layer and the output layer. In the convolutional layer kernels(filters) are used to detect features from the input vector, they are usually small matrices. A convolution of two continuous functions f and g defined by,

$$(f * g)(t) = \int_{-\infty}^{\infty} f(\tau)g(t-\tau)d\tau \tag{2}$$

however in a discrete case where g has support on -M,..,M is defined by the sum:

$$(f * g)(n) = \sum_{m=-M}^{M} f(n-m)g(m). \tag{3}$$

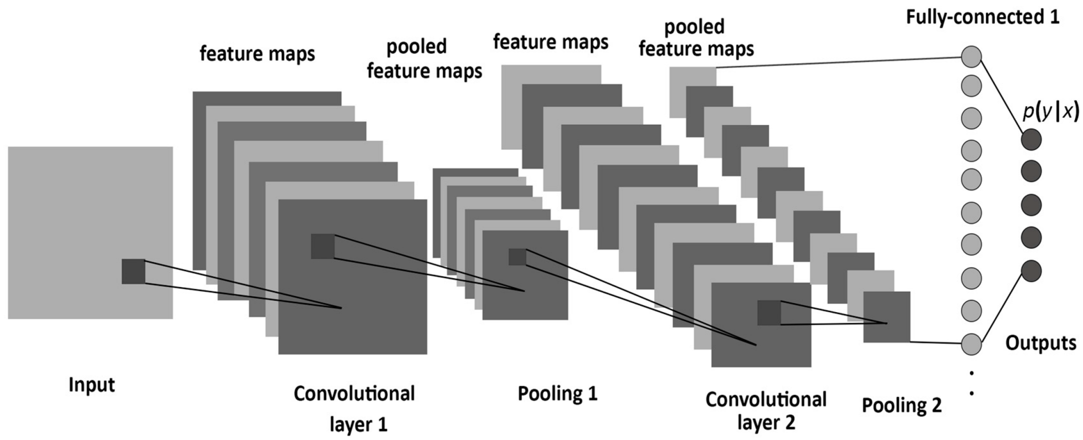

Figure 5: An architecture of a standard neural network which is comprised of the layers discussed in Section 3 [24].

This is the case when the function g is the kernel. The convolutional layer in a neural network performs this across the inputs, usually with many different kernels, as this allows selection of many different features from an input image [20].
Immediately after each convolutional layer an activation layer is applied. The purpose of this layer is to introduce nonlinearity into the system after it has just been computing linear operations. The ReLU layer is the most popular in CNNs as it is computationally efficient, it is shown in Figure 6 [25].

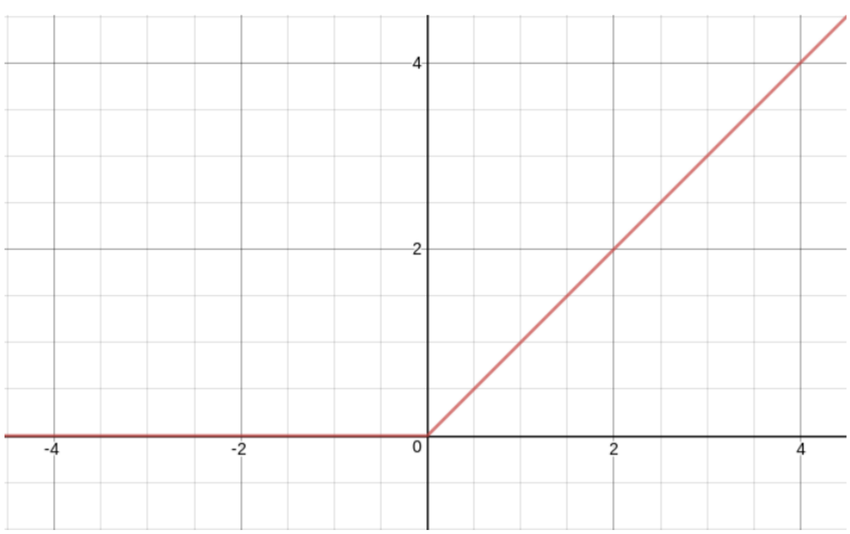

Figure 6: The ReLU activation function. Values are reduced to zero for all negative values and are unchanged for positive values. This activation function has been found to be useful for lots of convolutional neural network implementations [26].

The next layer is the pooling layer (or down sampling layer) which is used to reduce the spatial dimensions which makes the program less computationally demanding, it also reduces the chances of over fitting. The most popular choice is the max pooling layer, in this layer another filter passes across the image and takes the maximum value within that it has [25]. Up sampling is a similar effect but in reverse where vectors are padded in order to increase the output size. The output layer re-joins these convolutions back to the original size of the image. The convolutions that best picked out the features in the input image will have larger weights, meaning that a more accurate solution is chosen [20]. The reason CNNs are so useful for image data is that as the kernels scan over the image, they pick out structures from that image and crucially it extracts information with a strong dependence of the spatial relationship. Image data almost always has a strong emphasis on locality, where parts of an image are in relationship to other parts. Even though a CNN may reduce the size of an image or flatten it in its output layer it does not lose the information of the locality of an image [27].

4 Current Literature Discussing the Use of CNNs in Image Registration

There have been several attempts to use CNNs to perform image registration of cancer patients. One such experiment produced VoxelMorph a learning-based algorithm that is was able to show improvement over conventional methods of deformable image registration [6].
The architecture of the CNN used for this experiment is based on the UNet architecture, a type of encoder-decoder neural network, a network that was originally applied to a similar application in radiotherapy, image segmentation. Image segmentation is the process by which the location of anatomical structures is found along with their boundaries, also known as auto-contouring. The UNet architecture used in this experiment is shown in Figure 7.

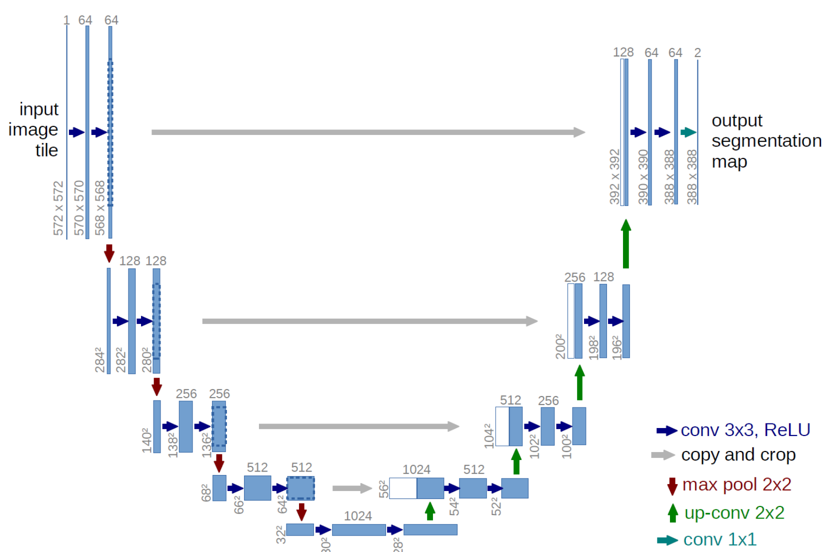

Figure 7: The original implementation of the UNet architecture on which this experiment was based [28].

It has several convolutional and max pooling layers that work to reduce the size of the image, the activation function used is the non-linear ReLU. It then up samples the image to the output size. It has skip connections between each max pooling layer that are used to prevent over fitting. The experiment that used VoxelMorph was used to do brain MRI registration and was performed in two ways, with both supervised and unsupervised learning models. The unsupervised only uses the moving 3D image and the fixed 3D image in order to produce the deformation vector field. The loss function for this method penalises differences of the two images and local spatial variations of the two images. The results for this method were comparable with those used to perform image registration currently. The second method was registration with auxiliary data used anatomical labels obtained by image segmentation to help train the model. The loss function for this method was a DICE function which evaluates the level of overlap of each part of the labelled structure. This method produced results that were more accurate than the traditional method and does so much quicker than it is able to. The authors of this paper suggest that as a general learning model this method of image registration may be applied to other medical image registration tasks.

Similarly, another experiment that was able to perform non-rigid image registration was led by Hu Yipeng [29]. This second paper uses labels to train the CNN but afterwards is able to produce DVFs without labels. This method again uses an encoder-decoder neural network to produce output DVFs. The labels used in this experiment were of the solid organs, ducts, vessels among other structures. This experiment was performed for image registration of the prostate glands for 111 pairs of images. This experiment delivered good results however this experiment did encounter some issues with over fitting around some areas of the prostate for the unaided (registered without the use of labels) tests.

5 Methodology

The main aim of this experiment was to use convolutional neural networks in order to perform the image registration of the planning CT images and PET images. We used open-source data from the cancer imaging archive that was in DICOM format (the standard format for medical imaging files) [10]. This data contained the Planning CT scans, PET scans and DVF files for each of around 300 patients. The aim was initially to train a CNN using the UNet architecture using the DVFs in these files to perform image registration that would achieve high accuracy. However, this method is limited by the accuracy of current DVFs, it would not be possible to surpass the accuracy of modern technologies in this way as the loss function would be designed on the assumption that the DVF file is absolutely accurate. This method avoids using labels as these were not provided with the patient data and introducing and performing image segmentation on the data would increase the time for the overall process.

During the process of the investigation it was decided that the image registration should take place in two parts, a rigid and a non-rigid transform. This is because initially the PET scan and the Planning CT scan are initially very spatially separated. A rigid transform would move the two images so the centre of the images are on top of one another, this is not a computationally difficult procedure as the whole image moves by the same vector. The deformable part of the registration would then be performed by the CNN as it would only have to reduce the loss in the deformable registration allowing it to find a solution much quicker. However, this means that the DVFs that are part of the cancer imaging archive would be unusable as the DVFs in this case do not have a two step registration. Registration of the images was performed using Elastix, a toolkit that is used to perform image registration and segmentation [30]. The output DVFs would then be used to train the network. Having initially decided to use a UNet based architecture, it was decided that we would base our network on a weakly supervised label driven network discussed in Section 4 as this network has been used solely for image registration and can also be changed to not use labels with little difficulty, with changing the loss function being the only major change. The loss function we implemented would flatten the three-dimensional vector field and perform a root mean squared calculation between the output DVF and the DVF that was produced from the output Elastix toolkit.

Another difficulty that was encountered is that many of the different patients' CT and PET scans were a different size and so needed to be re sampled so they had the same dimensions as convolutional neural networks take in only similar inputs. As well as this there were some limitations in the memory available in training the network, as the PET scans, CT scans and the network itself use a large amount of space so again the images had to be down sampled to reduce the amount of memory they would need.

6 Results and Discussion

The CT and PET scans taken from the cancer imaging archive are shown in Figures 1 and 2. As mentioned in Section 5 these images are spatially separated. Within the experiment a non-rigid transformation was performed that would overlay the centre of the two images. The next step was to produce a DVF using Elastix, this would fully register the images so they become identical. An example of rigid and non-rigid transformations are shown in Figure 8.

The CNN would take the non transformed images and attempt to perform the second step. As stated in Section 1 there are two main aims when using CNNs for image registration, that is it should be both faster and at least as accurate as current methods. At this point in the experiment, the first of these aims had been successfully achieved but not in the second. Overall the network is much faster than the Elastix program, the time taken for each of the Elastix toolkit is compared in Table 1.

	CNN	Elastix
Time for deformable image registration(seconds)	5.76 ± 0.05	161.3 ± 5.6

Table 1: These results show that using our convolutional neural networks is much faster than the Elastix toolkit

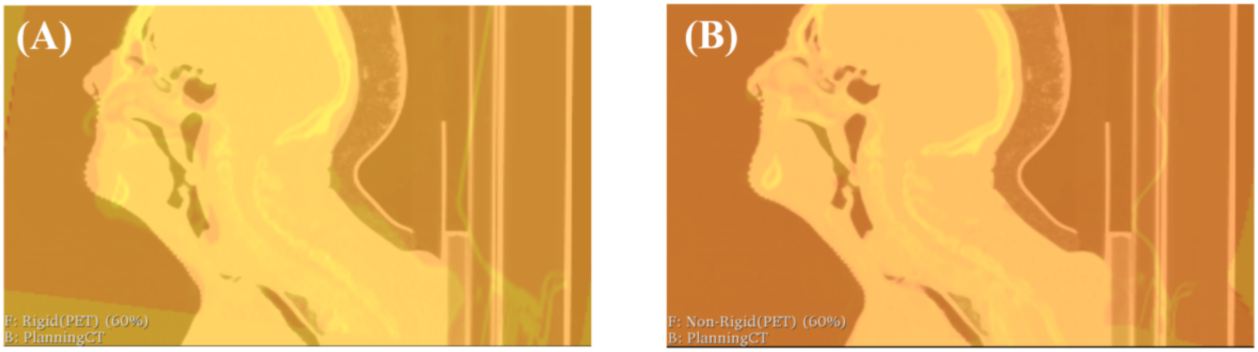

Figure 8: Two images taken from our data at different steps in the registration process. Image A on the left gives a sagittal view of the PET and CT scan after the rigid transformation. Here both the images are centre aligned but are at different angles due to the change of neck position of the patient between the two scans. Image B gives a sagittal view after a non-rigid transformation has been applied using Elastix. The face of the two images are now very closely matched, it is this transformation that was used to train the network.

However, there was less success with the accuracy of the image registrations, an example of an output image from the DFV that was produced is Shown in Figure 9.

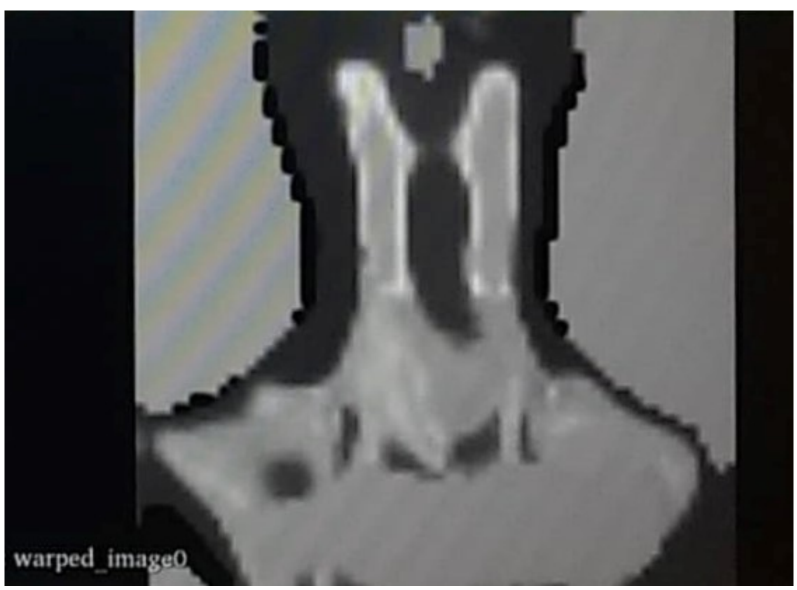

Figure 9: This image shows a coronal view of a PET scan that has undergone a non-rigid transformation using the network. Despite there clearly being some structures that resemble the human anatomy it has not been successfully registered. The loss of the network remains high after training.

While there is some structure within the image it has not been a successful image registration, the loss also remains high throughout the experiment, despite decreasing initially. A graph of the loss against iteration number is shown in Figure 10.

Figure 10: A graph that shows how the loss changes with the iteration number, the number of times the network has attempted a solution. As can be seen the loss function initially decreases but levels of between three and five thousand iterations.

One reason for this may be due to the fact that the amount slice thickness is different for each image is different for each image, so despite down sampling the image to have the same dimensions when they enter the CNN they have different amounts of the persons anatomy in each image, and some detail may be lost that means it is difficult for the CNN to train. One method around this may be to split each image into slices, for example have a slice just for a band near the top, middle and bottom of the head. This would mean that the CNN is being trained on the same anatomical structures. The DVFs from each of these could then be combined to produce a final DVF for the full structure.

7 Conclusion

This network clearly needs some development before it can be described as a successful attempt to produce DVFs that rival the accuracy of modern day technologies. The implementation of using three different parts of the image to ensure that the information fed into the network are of the same part of the anatomy may help in producing an output that will rival the accuracy of alternative methods of performing image registration. Even thought the method described above would require more time, the total time for registration would still be much quicker than that achieved by image registration permed by other methods.

References

[1] Cancer Research UK, *Cancer diagnosis and treatment statistics*, September 1, 2017 (accessed December 20, 2018). [Online]. Available: `https://www.cancerresearchuk.org/health-professional/cancer-statistics/diagnosis-and-treatment#ref-5`.

[2] R. P. Symonds, C. Deehan, C. Meredith, and J. A. Mills, *Walter and Miller's Textbook of Radiotherapy E-book: Radiation Physics, Therapy and Oncology*. Elsevier Health Sciences, 2012.

[3] R. Symonds, "Recent advances: Radiotherapy," *BMJ: British Medical Journal*, vol. 323, no. 7321, p. 1107, 2001.

[4] National Health Service, *How radiotherapy is given*, March 27, 2017 (accessed January 6, 2019). [Online]. Available: `https://www.nhs.uk/conditions/radiotherapy/what-happens/`.

[5] L. Tarassenko, "2 - mathematical background for neural computing," in *Guide to Neural Computing Applications*, L. Tarassenko, Ed., New York: Butterworth-Heinemann, 1998, pp. 5–35, ISBN: 978-0-340-70589-6. DOI: `https://doi.org/10.1016/B978-034070589-6/50002-6`. [Online]. Available: `http://www.sciencedirect.com/science/article/pii/B9780340705896500026`.

[6] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca, "Voxelmorph: A learning framework for deformable medical image registration," *arXiv preprint arXiv:1809.05231*, 2018.

[7] S. Shan, W. Yan, X. Guo, E. I. Chang, Y. Fan, Y. Xu, *et al.*, "Unsupervised end-to-end learning for deformable medical image registration," *arXiv preprint arXiv:1711.08608*, 2017.

[8] L. W. Goldman, "Principles of ct and ct technology," *Journal of nuclear medicine technology*, vol. 35, no. 3, pp. 115–128, 2007.

[9] P. Whiting, N. Singatullina, and J. Rosser, "Computed tomography of the chest: I. basic principles," *Bja Education*, vol. 15, no. 6, pp. 299–304, 2015.

[10] Cancer Imaging Archive, *Head-neck-pet-cts*, November 20, 2018 (accessed December 20, 2018). [Online]. Available: `https://wiki.cancerimagingarchive.net/display/Public/Head-Neck-PET-CT`.

[11] U. Nestle, K. Walter, S. Schmidt, N. Licht, C. Nieder, B. Motaref, D. Hellwig, M. Niewald, D. Ukena, C. M. Kirsch, *et al.*, "18f-deoxyglucose positron emission tomography (fdg-pet) for the planning of radiotherapy in lung cancer: High impact in patients with atelectasis2," *International Journal of Radiation Oncology* Biology* Physics*, vol. 44, no. 3, pp. 593–597, 1999.

[12] A. Berger, "How does it work?: Positron emission tomography," *BMJ: British Medical Journal*, vol. 326, no. 7404, p. 1449, 2003.

[13] H. Yu, C. Caldwell, K. Mah, I. Poon, J. Balogh, R. MacKenzie, N. Khaouam, and R. Tirona, "Automated radiation targeting in head-and-neck cancer using region-based texture analysis of pet and ct images," *International Journal of Radiation Oncology* Biology* Physics*, vol. 75, no. 2, pp. 618–625, 2009.

[14] C. B. Caldwell, K. Mah, Y. C. Ung, C. E. Danjoux, J. M. Balogh, S. N. Ganguli, and L. E. Ehrlich, "Observer variation in contouring gross tumor volume in patients with poorly defined non-small-cell lung tumors on ct: The impact of 18fdg-hybrid pet fusion," *International Journal of Radiation Oncology* Biology* Physics*, vol. 51, no. 4, pp. 923–931, 2001.

[15] D. Thorwarth, X. Geets, and M. Paiusco, "Physical radiotherapy treatment planning based on functional pet/ct data," *Radiotherapy and Oncology*, vol. 96, no. 3, pp. 317–324, 2010.

[16] A. C. Riegel, A. M. Berson, S. Destian, T. Ng, L. B. Tena, R. J. Mitnick, and P. S. Wong, "Variability of gross tumor volume delineation in head-and-neck cancer using ct and pet/ct fusion," *International Journal of Radiation Oncology* Biology* Physics*, vol. 65, no. 3, pp. 726–732, 2006.

[17] A. Taylor and M. Powell, "Intensity-modulated radiotherapy what is it?" *Cancer Imaging*, vol. 4, no. 2, p. 68, 2004.

[18] K. Newbold, M. Partridge, G. Cook, S. Sohaib, E. Charles-Edwards, P. Rhys-Evans, K. Harrington, and C. Nutting, "Advanced imaging applied to radiotherapy planning in head and neck cancer: A clinical review," *The British journal of radiology*, vol. 79, no. 943, pp. 554–561, 2006.

[19] T. Bortfeld, "The number of beams in imrt theoretical investigations and implications for single-arc imrt," *Physics in Medicine & Biology*, vol. 55, no. 1, p. 83, 2009.

[20] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning*. MIT press Cambridge, 2016, vol. 1.

[21] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *Bulletin of mathematical biology*, vol. 52, no. 1-2, pp. 99–115, 1990.

[22] F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain.," *Psychological review*, vol. 65, no. 6, p. 386, 1958.

[23] A. Karpathy, "Cs231n convolutional neural networks for visual recognition," *Neural networks*, vol. 1, 2016.

[24] S. Albelwi and A. Mahmood, "A framework for designing the architectures of deep convolutional neural networks," *Entropy*, vol. 19, no. 6, p. 242, 2017.

[25] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural networks*, vol. 61, pp. 85–117, 2015.

[26] A. F. Agarap, "Deep learning using rectified linear units (relu)," *arXiv preprint arXiv:1803.08375*, 2018.

[27] C. M. Bishop *et al.*, *Neural networks for pattern recognition*. Oxford university press, 1995.

[28] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, Springer, 2015, pp. 234–241.

[29] Y. Hu, M. Modat, E. Gibson, W. Li, N. Ghavami, E. Bonmati, G. Wang, S. Bandula, C. M. Moore, M. Emberton, *et al.*, "Weakly-supervised convolutional neural networks for multimodal image registration," *Medical image analysis*, vol. 49, pp. 1–13, 2018.

[30] S. Klein, M. Staring, K. Murphy, M. A. Viergever, and J. P. Pluim, "Elastix: A toolbox for intensity-based medical image registration," *IEEE transactions on medical imaging*, vol. 29, no. 1, pp. 196–205, 2010.