Using CNNs to detect heart disease in SPECT images.

Michael Smith

9th September 2017

Abstract

A three dimensional two-channel convolutional neural network (CNN) has been used to diagnose simulated single-photon emission computed tomography (SPECT) scans of diseased and healthy hearts [1, 2]. An accuracy of 0.464 is achieved over five classes (Normal, Ischaemic, Infarcted, Mixed, and Artefacted). The ROC AUC of all healthy vs all unhealthy classes is 0.826. Additionally, the trained CNN has been shown to generalise to SPECT scans of real patients. A brief overview of the algorithm used is below.

Preprocess data: cut and reshape incoming data to a two-channel 3D array of shape `[34,34,34,2]`.

Train CNN: train and validate two-channel CNN on simulated SPECT scans.

Finetune CNN: hold all but the final two fully connected layers static, and continue training CNN on real SPECT scans.

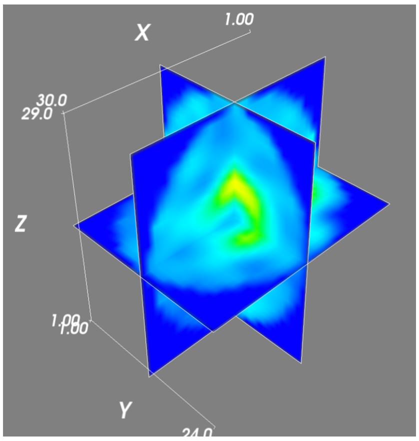
Visualise diagnosticity: generate a diagnostic map by setting a section of a scan to zero and finding the relative diagnosticity of this new scan. The zeroed area is moved, and the process is repeated until the zeroed section reaches the end of the scan.

There is a git repository holding all the code used in this paper at `https://github.com/Smith42/heart-cnn`.

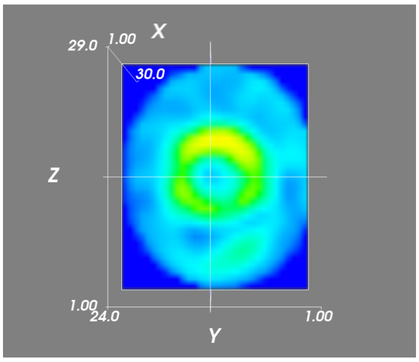
1 Methods

1.1 Data preprocessing

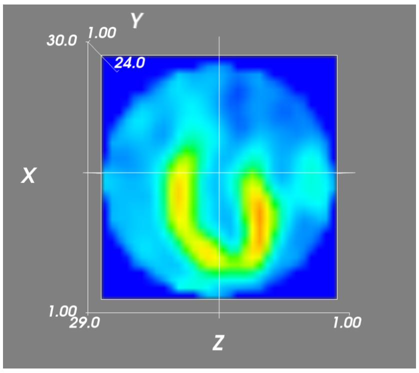
Each patient has two corresponding SPECT scans of their heart. An example of one of these scans is shown in figure 1. One scan is of their heart when the patient is at rest, and the other is when they are physically under stress. Each scan has a large amount of buffer space between useful information, and the edge of the data cube. Therefore, each scan is cut to remove the outer zero valued elements of the array, so that only the useful information remains. Once the cutting is completed, the remaining data is interpolated such that it fits a shape `[34,34,34]`. Finally, the scans taken while the patient is under stress and while calm are stacked to make a dual channelled image, so that the data to be fed into the CNN has shape `[34,34,34,2]`.



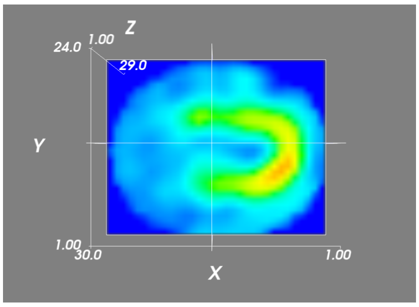
(a) Isometric projection



(b) X projection



(c) Y projection



(d) Z projection

Figure 1: Example images of a SPECT scan. This patient's heart is infarcted, and these scans are taken when they are under stress.

1.2 CNN architecture

The developed CNN has a two-channel architecture, as shown in figure 2, and described in detail in Zagoruyko et al. (2015) [1]. Zagoruyko et al. used this architecture to generate similarity scores between two images. It is also a simple and computationally cheap way of feeding all the required information into the heart CNN. Some other CNN architectures were examined, such as a 4D, pseudo-Siamese, and Siamese CNNs. These alternate CNNs increased the training time significantly, and were not found to have a corresponding accuracy increase, so they were dropped.

The best performing CNN that can diagnose ischaemic, infarcted, mixed, and artefacted hearts found has four convolutional layers interspersed with 3D max pooling layers that all have a kernel size and stride of `[2,2,2]`. The convolutional layers respectively have 32, 64, 128, 256 features, and kernel sizes `[10,10,10]`, `[5,5,5]`, `[2,2,2]`, `[2,2,2]`. After the fourth convolutional layer there are three fully connected layers with 2048, 1024, and 512 neurons respectively. A final softmax layer 5 neurons wide is used as output. All layers apart from the softmax layer use leaky ReLUs. The ADAM optimiser is used with a categorical cross entropy loss, and the learning rate is set at 0.0001 [3]. L2 regularisation has been used in this CNN on the penultimate three fully connected layers at a weight decay of 0.0001 [4]. In addition to L2 regularisation, applying dropout to each of the dense layers may be beneficial [5]. This CNN is trained on 50,000 total scans, evenly distributed over the five classes, for 30 epochs.

Since the CNN described above outputs five predictions, for each possible diagnosis, an ROC curve cannot be generated from the raw output. However, if the healthy predictions are treated as one class, and the unhealthy predictions (ischeamic, infarcted, mixed, and artefacted) as another, it is possible to generate ROC curves. The trained CNN is validated on 5,000 total scans, again evenly distributed over the five classes. The accuracies for each class, along with the ROC curve, and ROC AUC score for healthy and ill group classes are generated. No validation scans have been used in training.

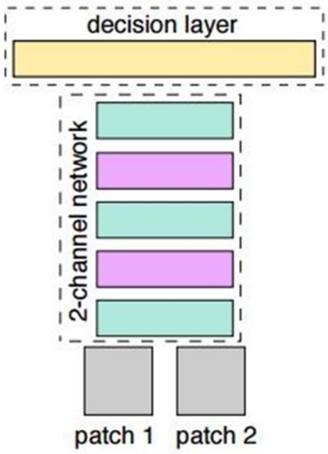


Figure 2: A 2-channel CNN [1]. Patch 1 and patch 2 are the patient's SPECT scans when calm and under stress. These patches have shape `[34,34,34]` and are fed into the CNN as two different channels of the same 3D image that has shape `[34,34,34,2]`. The CNN then convolves as usual over each channel individually. Once the decision layer is reached, the channels are combined into a fully connected (dense) neural layer. Several dense layers are in place before a final dense softmax layer outputs the estimated probability of each diagnosis.

Since there is a very small amount of data from real patients at hand, simulated data is used to train the CNN. The resulting trained CNN can be used as is on the real data, or can be finetuned for use with the real data. This finetuning is carried out by holding all but the final two dense layers of the CNN static, and training only on the last two layers. This reduces the chance of overfitting, which is always a risk when training on small datasets.

A k-fold cross validation is used to test the model on the real data [6]. This validation method maximises the data available for training, while not impacting the data available for testing the completed model.

Visualising the diagnosis is performed by setting a section of the incoming SPECT scan to zero, and comparing the new prediction to the prediction generated from the complete scan. If the two predictions are far apart, then the zeroed section of the scan is very diagnostic. The section of the scan set to zero is moved across the scan, and the diagnosticity is measured after each movement. In this way, we can generate a map of diagnosticity of the scan, and visualise where the CNN believes the issue with this patient's heart is.

2 Results

The main results are detailed below.

Normal	Ischaemic	Infarcted	Mixed	Artefact	Overall acc	AUC
0.45	0.51	0.37	0.47	0.51	0.46	0.83

Table 1: Performance of best multi-diagnosis CNN at 30 epochs of 50,000 total scans evenly distributed over all diagnoses. The first five values are the accuracies found from a validation set of 5,000 total scans. AUC is the ROC AUC for a randomly selected group of 1,000 ill scans, and 1,000 healthy scans. All validation scans have not been used in training.

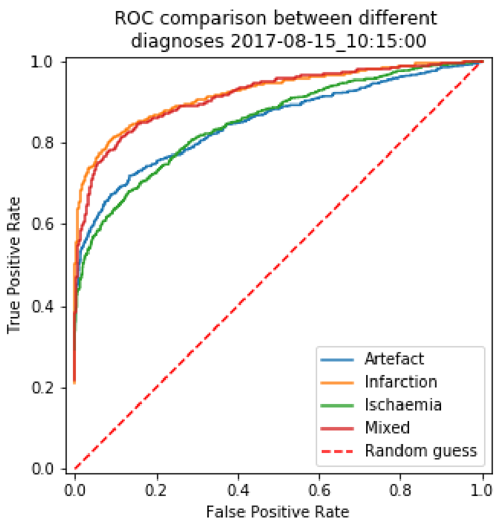


Figure 3: Comparison between ROC curves generated from different healthy/ill pairs. In each case the same set of healthy scans are used. The ill sets are of ischaemic, infarcted, mixed, and artefacted data.

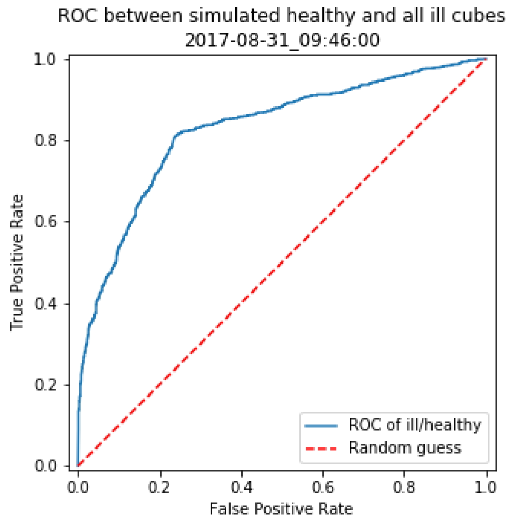


Figure 4: ROC curve for the best performing multi-diagnosis CNN. Generated from 1,000 healthy and 1,000 randomly selected ill scans.

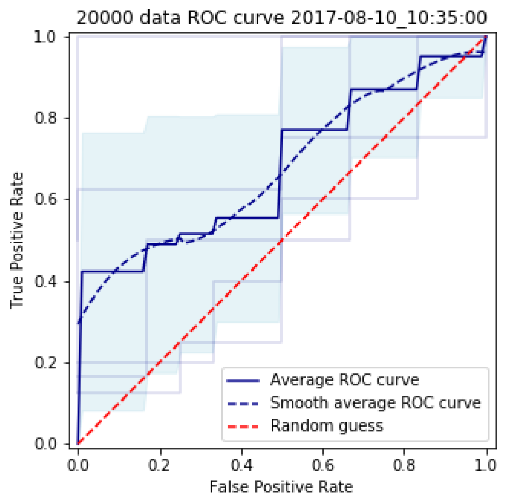


Figure 5: Result of finetuning a CNN trained on 10,000 healthy and 10,000 ischaemic simulated scans. This finetuning used a set of 29 healthy and 29 ischaemic hearts over $k = 5$ folds. It can be seen that there is a notable carryover between the ROC curve for simulated healthy/ischaemic data (figure 3), and this curve.

3 Discussion

Figures 3 and 4 show that the CNN works well on simulated data, and figure 5 shows that a CNN trained on simulated data, and finetuned on real data has some diagnosis capability.

Interestingly, the ROC curves for the artefact/healthy and ischaemia/healthy pairs are lower than those for the mixed/healthy and infarction/healthy pairs in figure 3. This is likely due to the fact that mixed and infarcted scans have artefacts in both the stressed and calm channels, whereas artefacted and ischaemic scans only have artefacts in one channel. The CNN described in this paper uses a two-channel architecture, with convolutions on the stressed and calm channels completely separate until the final dense layers. This separation of convolutions likely means that artefacts present in both stressed and calm channels are picked up easier. The scans that have artefacts on both channels also contrast more with the "clean" healthy scans.

In table 1 we can again see this separation of artefact/ischaemia and mixed/infarction scans. However, this time the scans with artefacts present in both channels are less likely to be diagnosed correctly. This is likely due to this particular CNN training itself to recognise single channel artefacts better. It may also be due to misdiagnoses between the different ill classes that would not be possible in the pairwise comparisons shown in figure 3.

The diagnosticity of the finetuned CNN would likely be improved with more SPECT scan examples. With enough examples ($\sim 10,000$ or so per class) a classifier could be trained with entirely real data, which would result in a drastic improvement in diagnosis power.

It does not take much computing power to use a trained CNN to make a prediction, and it is entirely possible to deploy a centrally trained CNN model onto a local instance. Therefore, a trained model could be incorporated into a SPECT machine, and relay a real-time diagnosis prediction, along with a visualisation of where the issue is.

4 Conclusion

A three dimensional two-channel convolutional neural network (CNN) has been used to diagnose simulated single-photon emission computed tomography (SPECT) scans of diseased and healthy hearts [1,2]. An accuracy of 0.464 is achieved over five classes (Normal, Ischaemic, Infarcted, Mixed, and Artefacted). The ROC AUC of all healthy vs all unhealthy classes is 0.826. Additionally, the trained CNN has been shown to generalise to SPECT scans of real patients. The methods used here could be used to generate a CNN model remotely on a supercomputer, and deploy the model to local SPECT machines. These SPECT machines could then relay diagnostic information and visualisation in real time.

References

[1] Sergey Zagoruyko and Nikos Komodakis. Learning to compare image patches via convolutional neural networks. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, jun 2015.

[2] Magdy M. Khalil, Jordi L. Tremoleda, Tamer B. Bayomy, and Willy Gsell. Molecular SPECT imaging: An overview. *International Journal of Molecular Imaging*, 2011:1–15, 2011.

[3] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2014.

[4] Andrew Y. Ng. Feature selection, l1 vs. l2 regularization, and rotational invariance. In *Twenty-first international conference on Machine learning - ICML 2004*. ACM Press, 2004.

[5] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15:1929–1958, 2014.

[6] M. Stone. Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society. Series B (Methodological)*, 36(2):111–147, 1974.