

# Convolutional Neural Networks for Medical Clustering

David Lyndon<sup>1</sup>, Ashnil Kumar<sup>1,3</sup>, Jinman Kim<sup>1,3</sup>, Philip H. W. Leong<sup>2,3</sup>, and Dagan Feng<sup>1,3</sup>

<sup>1</sup> School of Information Technologies, University of Sydney, Australia

<sup>2</sup> School of Electrical and Information Engineering, University of Sydney, Australia

<sup>3</sup> Institute of Biomedical Engineering and Technology, University of Sydney, Australia

dlyn9602@uni.sydney.edu.au

{ashnil.kumar, jinman.kim, philip.leong, dagan.feng}@sydney.edu.au

**Abstract.** A major challenge for Medical Image Retrieval (MIR) is the discovery of relationships between low-level image features (intensity, gradient, texture, etc.) and high-level semantics such as modality, anatomy or pathology. Convolutional Neural Networks (CNNs) have been shown to have an inherent ability to automatically extract hierarchical representations from raw data. Their successful application in a variety of generalised imaging tasks suggests great potential for MIR. However, a major hurdle to their deployment in the medical domain is the relative lack of robust training corpora when compared to general imaging benchmarks such as ImageNET and CIFAR. In this paper, we present the adaptation of CNNs to the medical clustering task at ImageCLEF 2015.

**Keywords:** Deep Learning, Convolutional Neural Networks, Medical Image Retrieval

## 1 Introduction

This paper documents the Biomedical Engineering and Technology (BMET) team from the University of Sydney’s submissions for the ImageCLEF 2015 [1] Medical clustering task [2].

The objective of our experiments was to evaluate the effectiveness of Convolutional Neural Networks (CNNs) for this task. In particular, we propose a deep learning framework that learns high-level representations of anatomical elements contained in each image and uses these to cluster the images.

## 2 Background

Convolutional Neural Networks, a type of deep learning algorithm, have been used to produce state-of-the-art results for a variety of machine learning tasks such as image recognition, acoustic recognition and natural language processing

since 2012 [3–5]. CNNs share the common features of all deep learning algorithms: stacked layers of neuronal subunits that learn hierarchical representations (allowing the data to be understood at various levels of abstraction, in isolation or combination [3]), the ability to perform unsupervised pre-training on unlabeled data and efficient parallelization on multiple core GPUs which can result in improvements of up to 5000% over CPU-only implementations [5].

A more subtle implication of deep learning is that it can automatically extract features from raw data [3–5]. Typically, a key factor in the success of typical machine learning algorithms is extracting salient features from the raw data. Taking image recognition as an example, a feature set such as edges or SIFT [6] would be extracted from the raw data and it is these new features per se or in combination with the original raw data that would be fed into the machine learning algorithm. While some aspects of the process can be automated or implemented with well known algorithms, a major drawback is that it generally requires expert domain knowledge to define which features should be used and evaluate their success.

Deep learning algorithms, however, are able to directly utilise raw data instead of hand-crafted features. By feeding the data sequentially through many successive layers of subunits, the higher levels of the system are able to understand the data in terms of successively abstract representations [3].

Medical Image Retrieval (MIR) tasks, such as the tests devised for ImageCLEF, require learning precisely these kinds of highly abstract representations, i.e. image modality or the anatomical semantics of the image. However, to the best of our knowledge it is not currently a well established method in this domain. This is due to not only the inherent challenges of medical images[7], but also because state-of-the-art deep learning results are typically obtained using huge sets of labelled training data<sup>4</sup> on tasks that are arguably less subtle. As a justification for these claims, consider that the ImageNET general object recognition task corpora consists of millions of robustly labelled images and was created with the assistance of crowdsourcing via Amazon Mechanical Turk [9]. On the other hand, medical imaging datasets require careful labelling by domain experts, often specialists in a particular area [7, 10–12] and as a result are generally much smaller.

Large training sets are a current necessity of very deep systems because they contain many millions of internal parameters that must be estimated from the data. Too little data can result in the the higher-level neurons’ activation being the result of salient features of the training set and not reflecting the high-level representations. If this ’overfitting’ occurs then the system’s ability to generalise on new data is severely impaired [13].

In addition to the issues regarding the volume of data required, it must be mentioned that while deep learning can automatically perform excellent feature extraction, this comes at the significant cost of the larger number of hyperparameters that must be evaluated in order to find an optimal system [14]. For

---

<sup>4</sup> Krizhevsky et. al. [8] used approximately 1.2 million labelled examples for their breakthrough result in ImageNET in 2012.

example, compared to a commonly used machine learning algorithm such as the Support Vector Machine (SVM) that has a basic hyperparameter search space with dimensions of choice of kernel, regularization constant and kernel hyperparameter, even the simplest implementation of a CNN requires fundamental choices about the number and type of layers, filter size and number of filters per layer, and the learning rate. More advanced implementations include factors such as unit activation function and the use of dropout. While there are guidelines for these choices in the literature [14], the difficulty of even a small parameter search is compounded by the increased computational requirements of training the system.

### 3 Methods

#### 3.1 Image Preprocessing

A requirement of our classifiers was uniformly sized input vectors, however, there was some variation in the training data size. The sizes of the images were at least 1600px in one dimension and then between 1600 and 2348px in the other. In order to use consistently sized images and not lose any crucial information, we created a new square image with the dimensions of the largest dimension of the original image, filling any empty space with black pixels.

Even prior to training the CNN, we were aware that the computational requirements were quite demanding and this would be exacerbated by using large images. With that in mind, we resized the images to 256x256px to reduce computational overhead. Good results have been reported in the literature for complex tasks with 48x48px images [15] and Krizhevsky et. al. [8] achieved state-of-the-art general object recognition with 256x256px images (technically, the system had an input of 224x224px, but these were subimages of the original 256x256px images).

After resizing the images were 256x256x3px, the third dimension describing the three colour channels. The images supplied were in actual fact gray scale, i.e. all colour channels were equal-valued, so we simply sliced the array preserving only the 'red' channel.

We chose to train a single run of four models using 100% of the training data with no parameter optimization and use the ImageCLEF results as the test.

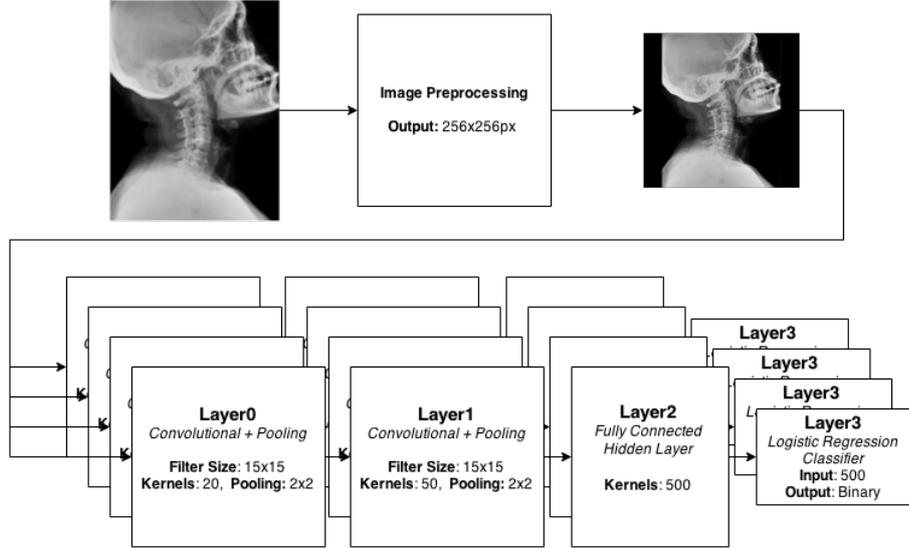
#### 3.2 Convolutional Neural Network

The task requires multi-label classification across four anatomical classes, with a null set indicating that the data is a true-negative image taken with the same camera, but not of the human body. To facilitate this output from our experiments we constructed 4x1 vs. All CNN models.

The architecture for the CNN used for our experimentation was based on a simplified version of Yann LeCun et. al.'s [16] LeNet-5<sup>5</sup>. This basic CNN is capable of correctly classifying the MNIST handwritten digit database with 1.7%

<sup>5</sup> <http://deeplearning.net/tutorial/lenet.html>

test error. We modified the input to account for larger images and output a different number of classifications. The network consists of two convolutional pooling layers, with one fully connected hidden layer. The features that are output by the hidden layer are used for binary classification by a logistic regression classifier. The architecture of the system is shown in Figure 1.



**Fig. 1.** The architecture of CNN used for the experiments

The specifications of the convolutional-pooling layers are detailed in Table 1.

**Table 1.** Details of Convolutional Pooling Layers

Hyperparameter	Layer0	Layer1
Number of Filters	20	50
Size of Filters	15x15px	15x15px
Max Pooling	2x2	2x2
Stride	1	1

Other hyperparameters for the CNN are detailed in Table 2.

**Table 2.** Other details for CNN

Hyperparameter	Value
Number of Units in Fully Connected Layer	500
Batch Size	20
Learning Rate	0.005
Training Epochs	100

As mentioned earlier the CNN requires a great deal of computational resource to run. We initially began training the four models on a CPU-only solution and despite it being a very powerful machine<sup>6</sup>, it took approximately 90 minutes to train a model for a single epoch (albeit, training four models simultaneously). Fortunately, we were given an opportunity to run these models on a system with two Nvidia K20 GPUs. Even training the four models simultaneously (two per GPU), it only took approximately 11 minutes to train a model for a single epoch – an 8-fold speedup. We planned to submit a single run, having trained each of the four classifiers for 100 epochs. This process would have taken over a week on a CPU-only system, instead it took less than a day on the GPU server.

## 4 Results

The test results for our submission as supplied by ImageCLEF are displayed in Table 4.

**Table 3.** Test results as supplied by ImageCLEF.

Exact Match	Any Match	Hamming Similarity
49.7%	59.6%	84.9%

### 4.1 CNN-Learnt Features

The CNNs were able to extract improved representations from raw data without the requirement for domain knowledge. This was done without any hyperparameter tuning suggesting that there are further improvements that could be made. This is an important result both for this task and for MIR generally as it suggests that there is potential in using CNN or other deep learning strategies as a 'black box', whereby we will be able to achieve excellent machine learning performance without the need of expert-designed feature extraction or domain knowledge.

<sup>6</sup> Azure Standard A4 VM: 8-core 2.1GHz CPU, 14GB RAM

## 5 Perspectives for Future Work

We believe that that these results can be significantly improved upon by making use of a variety of techniques. Primarily we would want to continue to explore training the CNNs using GPUs, because as we have demonstrated, the performance increase is non-trivial and allows us to expand our hyperparameter and architecture search. Rectified Linear Units (ReLUs), as opposed the Tanh units used in our network are also known to improve training performance [17, 18].

Although this network is very capable of learning quality representations of the MNIST dataset, it is both less deep and less dense than networks used to achieve state-of-the-art results in more sophisticated tasks [8]. For instance, Krizhevsky et. al. [8] used a network with 2 convolutional-max pooling layers, 3 convolutional layers and 3 fully connected layers, all of which were more neuron-dense than ours, to achieve their result in ImageNET 2012. Improved training performance will allow us to implement a larger and deeper network along these lines.

Larger and deeper networks introduce issues with overfitting, but we believe this can be controlled using well-tried techniques such as dropout [8, 13, 19], data augmentation [8, 20] and unsupervised pretraining [21, 22].

### Acknowledgements

This work was supported in part by a Microsoft Azure for Research grant, which provided the cloud infrastructure to conduct our experiments.

### References

1. M. Villegas, H. Mller, A. Gilbert, L. Piras, J. Wang, K. Mikolajczyk, A. G. S. de Herrera, S. Bromuri, M. A. Amin, M. K. Mohammed, B. Acar, S. Uskudarli, N. B. Marvasti, J. F. Aldana, and M. del Mar Roldn Garcia, General Overview of ImageCLEF at the CLEF 2015 Labs, Springer International Publishing, 2015.
2. M. A. Amin and M. K. Mohammed, Overview of the ImageCLEF 2015 medical clustering task, in CLEF2015 Working Notes, 2015.
3. Y. Bengio, A. Courville, and P. Vincent, Representation learning: a review and new perspectives, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 17981828, Aug. 2013.
4. Y. LeCun, Y. Bengio, and G. Hinton, Deep learning, *Nature*, vol. 521, no. 7553, pp. 436444, May 2015.
5. J. Schmidhuber, Deep learning in neural networks: an overview, *Neural Netw.*, vol. 61, pp. 85117, Jan. 2015.
6. D. G. Lowe, Object recognition from local scale-invariant features, in *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on, 1999*, vol. 2, pp. 11501157 vol.2.
7. A. Kumar, J. Kim, W. Cai, M. Fulham, and D. Feng, Content-based medical image retrieval: a survey of applications to multidimensional and multimodality data, *J. Digit. Imaging*, vol. 26, no. 6, pp. 10251039, Dec. 2013.

8. A. Krizhevsky, I. Sutskever, and G. E. Hinton, ImageNet Classification with Deep Convolutional Neural Networks, in *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 10971105.
9. J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, ImageNet: A large-scale hierarchical image database, in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, 2009, pp. 248255.
10. J. Kalpathy-Cramer, A. G. S. de Herrera, D. Demner-Fushman, S. Antani, S. Bedrick, and H. Mller, Evaluating performance of biomedical image retrieval systems—An overview of the medical image retrieval task at ImageCLEF 20042013, *Comput. Med. Imaging Graph.*, vol. 39, pp. 5561, 2015.
11. H. Mller, N. Michoux, and D. Bandon, A review of content-based image retrieval systems in medical applications—clinical benefits and future directions, *International journal of*, 2004.
12. M. S. S. Faruque, M. S. Shahriar Faruque, B. Shourav, M. Kazi Mohammed, H. Mahady, and M. Ashraf Amin, Teaching & Learning System for Diagnostic Imaging - Phase I: X-Ray Image Analysis & Retrieval, in *Proceedings of the 7th International Conference on Computer Supported Education*, 2015.
13. N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, Dropout: A Simple Way to Prevent Neural Networks from Overfitting, *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 19291958, Jan. 2014.
14. Y. Bengio, Practical recommendations for gradient-based training of deep architectures, *arXiv [cs.LG]*, 24-Jun-2012.
15. D. Ciresan, U. Meier, J. Masci, and J. Schmidhuber, A committee of neural networks for traffic sign classification, in *Neural Networks (IJCNN), The 2011 International Joint Conference on*, 2011, pp. 19181921.
16. Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, Gradient-based learning applied to document recognition, *Proc. IEEE*, vol. 86, no. 11, pp. 22782324, Nov. 1998.
17. V. Nair and G. E. Hinton, Rectified linear units improve restricted boltzmann machines, in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, 2010, pp. 807814.
18. A. L. Maas, A. Y. Hannun, and A. Y. Ng, Rectifier Nonlinearities Improve Neural Network Acoustic Models, *W&CP*, vol. 28, 2013.
19. G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, Improving neural networks by preventing co-adaptation of feature detectors, *arXiv [cs.NE]*, 03-Jul-2012.
20. Classifying plankton with deep neural networks, Sander Dieleman. [Online]. Available: <http://benanne.github.io/2015/03/17/plankton.html>. [Accessed: 30-May-2015].
21. X. Glorot, A. Bordes, and Y. Bengio, Domain adaptation for large-scale sentiment classification: A deep learning approach, in *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, 2011, pp. 513520.
22. Y. Bar, I. Diamant, L. Wolf, and H. Greenspan, Deep learning with non-medical training used for chest pathology identification, in *SPIE Medical Imaging*, 2015, p. 94140V94140V7.