Machine Learning Theory Project

Final Report

By:

Fabian Falck

ID: 2016403399

Russlan Randowar

ID: 2016403381

Gao Cong

ID: 2013011152

Lin Shengjie

ID: 2013011153

Beijing, January 8, 2017

Important notice to the TA

In this note, we shortly want to describe who of the project team worked on what. This is very important to especially Russlan and Fabian since they did they major work (90 %) up to today and are worried that they do not pass the course if their work for this project is not clearly stressed.

Russlan Randowar and Fabian Falck worked on the fully-convoluted neural net which the results in the appendix of this report also allude to. The main part of the report, all efforts of producing results and the leadership in the project belongs to them. Gao Cong and Lin Shengjie worked on the extension-model. This model so far does not work, but they will make it work soon.

Furthermore, we want to stress that 周杉 and 邹旭 left the group without any excuse 2 days before the hand-in of this final project. Russlan and Fabian relied on their work, but they did not do anything and always said they would do something, but then didn't. We are highly disappointed by their behavior and cannot accept that a group structure can be changed so late during the project work without our consultation.

Introduction

In this project, we are faced with the problem of classifying potential lung cancer scans on whether they are dangerous cancel cells or not. This problem refers and its approach refers to one of the large problems of computer vision in health care, namely the detection of hazardous and dangerous objects on the human body being detected by a machine. In the past hundred years of medical research, a deeper and deeper understanding was generated on how cancer detection works. Especially through the advancement of detection machines, such as computer topographies (CT) and x-ray, the medical research advanced to the point that it is nowadays possible for some types of cancers to detect them in an early stage.

However, the vision of detecting and classifying such cancers automatically or at least with a semi-automatic process which is primarily done by a machine and could be conducted just by a machine assistant is not yet reached. Also, the goal of spreading medicine to people around the world, not just the rich people, is not yet reality. All of this would be possible if one could find a reliable algorithm to detect cancer.

In this project, we will – considering the given data set – focus on a particular type of cancer, namely lung cancer. We will do so using the following structure of this report: Starting with a literature overview giving an overview about the state of the art methods in computer vision, we then explain the dataset. Afterwards, we describe our two models (a main model and an extension) and describe the results we achieved with these two models. Finally, we conclude our results and give an overview about the implications of our research.

Literature Review

The problem of lung nodule detection has long been concerned in the area of computer-assisted applications. Traditional image processing methods have been used in this classification work. Chan etc.[1] designed the algorithm to perform a pixel similarity analysis on the appropriate regions within the CT images to detect potential nodules and performs one or more expert analysis techniques using the features of the potential nodules to determine whether each of the potential nodules is or not a lung nodule. As the development of artificial intelligence, neural network is then applied in this problem. B. Lo etc.[3] have developed a double-matching method and an artificial visual neural network technique for lung nodule detection. The technique of which is generally applicable to the recognition of medical image pattern in gray scale imaging. They compare the performance of non-convolution method, convolution method and convolution+ fuzzy training method. And the improvement of using the convolutional neural network can be clearly observed. In the work of G. Penedo etc.[2] , a computer-aided diagnosis system, based on a two-level artificial neural network (ANN) architecture has been developed. They employed two different network structures and combine them together. The output result presents 89%-96% sensitivity and 5-7 false positives/image.

Neural network is proposed as an approach to the development of high performance image classification systems. [4][5] A typical structure of CNN consists of an input layer, convolutional layer, pooling layer, fully-connected layer and an output layer. The feature of the image can be better grasped and presented after passing the deep networks. But traditional CNNs are mostly focused on 2D image classification problems, in the work of Z. Wu etc., 3D shape analysis is firstly proposed using neural network.[6] They proposed to represent a geometric 3D shape as a probability distribution of binary variables on a 3D voxel grid, using a Convolutional Deep Belief Network. The algorithm of ModelNet 40 achieves an mAP of 49.2%, and ModelNet10 68.3%. Several trials of 3D image classification using neural network have been presented later. In our work of this project, we tested the performance of VRN Ensemble, which is the best performing structure in the 3D image classification assignments.

[1] Chan, Heang-Ping, et al. "Lung nodule detection and classification." U.S. Patent Application No. 10/504,197.
[2] Penedo, Manuel G., et al. "Computer-aided diagnosis: a neural-network-based approach to lung nodule detection." IEEE Transactions on Medical Imaging 17.6 (1998): 872-880.
[3] Lo, S-CB, et al. "Artificial convolution neural network techniques and applications for lung nodule detection." IEEE Transactions on Medical Imaging 14.4 (1995): 711-718.
[4] Kanellopoulos, I., and G. G. Wilkinson. "Strategies and best practice for neural network image classification." International Journal of Remote Sensing 18.4 (1997): 711-725.
[5] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems. 2012.
[6] Wu, Zhirong, et al. "3d shapenets: A deep representation for volumetric shapes." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015.

Data description

The data set contains 723 lung cancer scans. Each scan consists of on average 517 images which are the observations. Each observation has an interpolated pixel area of 40 × 40 × 40, i.e. each pixel represents an area of approx. 1mm3. The values of the pixels are given in a linear transformation Hounsfield unit (HU) scale defined as

HU = 1000 (μ – μwater) / (μwater – μair)

where μwater and μair are the linear atenuation coefficients of water and air ?. The linear transformation was applied to guarantee non-negative values which is necessary for training our model. Furthermore, each image was hand-labeled by physicians if the image contains a nodule or not. A nodule is a positive lung cancer observation which can occure in different types. The types are further discussed in section 4.2. Each observation is labeled with either a 1 (true nodule) or a 0 (false nodule). Additional information for each obervation is a nodule ID. Since a true nodule can occure in multiple observations, this ID allows to identify the unique nodules.

Empirical analysis

The total number of obervations (40x40x40) were 374289. The number of obervations ranges from 35 to 1479. We performed our training on in total 12000 observations which we partly created using data augmentation (see below).

Model Description – Convolutional Neural Net

As a first model, with reference to the literature overview above, we decided on a convolutional neural net due to its very convincing results in recent research. Our model consists of in total 5 layers which are listed in the occurring sequence: a 2D convolutional layer, another 2D convolution, a max-pool layer

Since we were not fully satisfied with the stated model above, we studied literature and found another very promising convolutional net structure that we describe in the following. and two fully connected layers. The final probabilities were created using a logit function. Furthermore, we used the famous dropout method to regularize a data. We used a 0.9 keeping-rate in the dropouts. As an activation function, we used the reluctance function. Furthermore, we had to use data augmentation which is further explained down below. We used rotations and transformations of the images since we were not provided with enough positive observations so that our algorithm does work properly.

Data Augmentation

The dataset is separated to two parts for training and self-testing, the ratio of which is 3:1. In order to generate more positive samples, we experiment with treating the different rotations as separate channels for a single instance. For example, a single scan matrix (size 40*40*40) is reflected around its central axis in each dimension. And a rotation operation is also made to produce new samples. Finally, we balanced the two parts of positive and negative samples, which contribute a lot to our final classification result.

Model Description - Extension

Considering the data structure is exactly three-dimension[1], we extended our model to a Voxel-based ConvNets, the Voxception-ResNet (VRN), which is especially applied to 3D object recognition. Key to our approach is the use of Inception-style modules[2], Batch Normalization[3], Residual connections with pre-activation[4][5] and stochastic network depth[6]. In contrast to previous 3D ConvNet approaches which used shallow networks, we train networks with up to 45 layers to take advantage of the increased expressivity that comes with model depth. In order to improve parameter efficiency, the early layers in each path of the block have half as many filters as the final layer. Downsampling is accomplished through Voxception-Downsample blocks. The order of application of rectifying nonlinearities and Batch Normalization to obtain pre-activation blocks is changed. And the non-residual paths of blocks are stochastically dropped.

Fig.1 shows the best-performing architecture and consists of an initial convolutional layer, for main units, each containing three stacked VRN blocks and a Voxception-Downsample block, a final convolution with a residual connection and keep probability of 0.5, then a global pooling layer and two fully-connected layers. The number of filters begins at 32, and is doubled at each downsampling block. The deepest path through the network is 45 layers and the shallowest path is 8 layers deep.

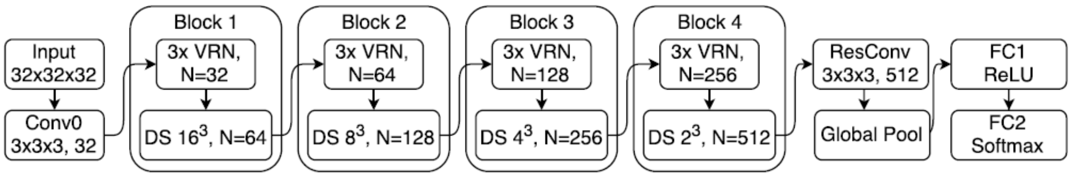

Fig.1

[1] Brock, Andrew, et al. "Generative and Discriminative Voxel Modeling with Convolutional Neural Networks." arXiv preprint arXiv:1608.04236 (2016).
[2] C. Szegedy, S. Ioffe, and V. Vanhoucke. Inception-v4, inception-resnet and the impact of residual connections on learning. arXiv Preprint arXiv: 1602.07261, 2016.
[3] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In ICML 2015.
[4] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In CVPR 2016.
[5] K. He, X. Zhang, S. Ren, and J. Sun. Identity mappings in deep residual networks. arXiv Preprint arXiv: 1603.05027, 2016.
[6] G. Huang, Y. Sun, Z. Liu, D. Sedra, and K. Q. Weinberger. Deep networks with stochastic depth. arXiv Preprint arXiv: 1603.09382, 2016.

Results

The accuracy rate we achieved was 80%. For detailed results, please consider the scripts and run them as explained.

Testing our performance, understanding the file

In order to test the performance of our model, we provide the meta-file of our model. To be able to actually apply a testing, one has to normalize the data manually using the following command:

```
meanNormData=np.divide(np.subtract(data,mean_dataset),std_dataset)
```

In order to understand how we created and pre-processed our data, please read the file "README.odt" which we also provide in the appendix of this paper.