Initial Plan - Developing a Computer Vision System to Identify Pulmonary Nodules in Computed Tomography Scans

CM3203 - One Semester Individual Project - 40 Credits

Stuart Clark - C1353032
Supervised by Paul Rosin

1 Project Description

In 2014 lung cancer was responsible for the deaths of 35,895 people in the UK alone (Cancer Research UK 2017a). One of the primary methods for detecting lung cancer is through the use of Computed Tomography (CT) scans of the chest. Such a CT scan produces a set of images displaying cross sections of a patient's chest at different points. Traditionally these images are then given to a radiologist who analyses them in order to identify any issues that may be visible. One of the main indicators a radiologist will use to assert that a patient may potentially have lung cancer is the presence of pulmonary nodule(s). A pulmonary nodule (PN) is a small growth in the lungs that is somewhat round in appearance or form. Figure 1 shows an example of a single cross section obtained using a CT scan with a PN highlighted.

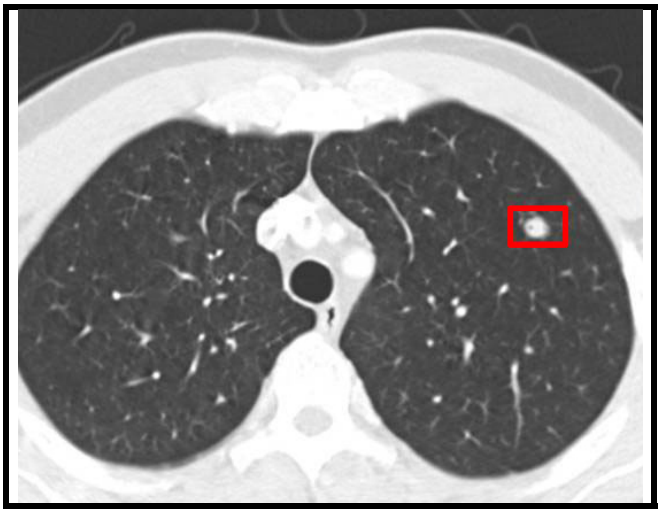

Figure 1: Example cross-section from a CT scan with nodule highlighted

The University of Rochester (2017) state that approximately 40% of PNs detected are reviled to be cancerous using further diagnostic techniques. Furthermore patients who receive treatment for a cancerous PNs survive more than five years beyond the diagnosis in 40% of cases where the PN is >1cm across and in 80% percentage of cases where the PN is ≤1cm across. These statistics demonstrate why it is so important to identify PNs early.

The aim of this project is to use computer vision and machine-learning techniques to create a system that can be used reliably to identify PNs in CT scans and used as an aid for radiologists tasked with examining them. If such a system was extremely successful it could even replace a human radiologist and identify PNs independently, meaning only CT scans that do contain nodules would need to be analysed by a radiologist. This would be advantageous as it would: free up radiologist to perform tasks that cannot be automated, allow for greater numbers to be involved in lung cancer screening schemes and help to aid diagnoses for a growing population as it would likely be much faster than a human radiologist. Furthermore, if such a system was used globally it could share detailed information in a manner that is not possible for humans. This information could be used to improve the machine-learning aspects of the system and hence the reliability. Finally the systems output could be used as the input for another system used to distinguish between benign and cancerous PNs. This would be particularly valuable as many of the diagnostic techniques to do this currently are invasive and dangerous to patients. Having a non-invasive system would mean that screening could be employed more widely and more instances of lung cancer could be detected

in the early stages (Cancer Research UK 2017b).

2 Project Aims and Objectives

When the project is complete the following objectives should be met:

- A well thought out overall design of the system should have been created.
- A system should be implemented that is capable of:
  - Inserting, indexing, and querying meta-data provided with CT scans in a database.
  - Segmenting regions of interest in CT scan images.
  - Comparing segmented regions of interest with a given ground truth.
  - Training a classifier using features of the regions of interest identified.
  - Using a trained classifier to identify PNs in previously unseen CT scans or state that there are non PNs.
  - Highlighting PNs in a suitable GUI so that it can be used as an aid for radiologists.
  - Having additional features added to the classifier easily.
  - Storing the results of features calculated for the training set so that they do not need to be recomputed each time the classifier is trained.
  - Optimising parameters such as thresholds using appropriate algorithms.
  - Unit testing functionality where possible and appropriate.
  - Evaluating the effectiveness of the classifier.
- The system should have be able to identify PNs with a better than chance success rate.
- A report should have been created that:
  - Provides detail on the research performed in order to begin the project.
  - Documents design and implementation choices made during the project.
  - Provides analysis on the effectiveness of the system.

3 Work Plan

Prior to the creation of the following plan some key steps had already been completed and as such were not included in it. For instance, the data source required to train and test the system had already been researched and selected. Additionally, many of the technologies required for the system had also been, researched, selected, installed and set up as project. Finally a simple GUI had also been created that could be used to view stacks of images for debugging.

The planning system bellow is similar to that which is used at Eysys (2017) however the L1 plan gives information at a weekly resolution rather than monthly. Similarly to Eysys's the L2 plan provides detail at a daily resolution showing how many days are allocated for each task. These estimates are liable to error and change, however it is still useful to make them. This is as planning at this level means that a clear idea of where the system is going can be maintained whilst developing. For some weeks the L1 cells are empty. These cells will be completed and others added

to / edited as more information becomes available throughout the project. Some features of this planning system are when a task is complete it is marked with "[DONE]" and when all the tasks for a day are done the cell is marked green. This ensures that tasks are not missed and enables progress to be tracked at the glance of an eye. Another key feature of the plan is that cells that contain milestones are filled yellow with the milestone written in all caps and preceded by a "◇".

	January						
L1	Week 1 Creation of initial plan and supporting document.						
	Mon 23/01/17	Tue 24/01/17	Wed 25/01/17	Thu 26/01/17	Fri 27/01/17	Sat 28/01/17	Sun 29/01/17
L2		• LaTex tutorials [DONE]	• Project description [DONE] • Project aims and objectives [DONE]	• Create initial plan.[DONE]	• Meeting with Paul to discuss initial plan.[DONE] • Make suggested changes[DONE] • Back up images		
L1	Week 2 Begin background research into radiology.						
	Mon 30/01/17	Tue 31/01/17	Wed 01/02/17	Thu 02/02/17	Fri 03/02/17	Sat 04/02/17	Sun 05/02/17
L2	• Make suggested changes ◊ HAND IN INITIAL PLAN	• Watch video tutorials for analysing CT scans • Find some relevant papers to read					

	February						
L1	Week 1 Further research, Investigation of data set suggested in proposal.						
	Mon 30/01/17	Tue 31/01/17	Wed 01/02/17	Thu 02/02/17	Fri 03/02/17	Sat 04/02/17	Sun 05/02/17
L2			• Read papers found and take notes particularly on segmentation techniques, features and classifiers used.	• Import CT image meta-data into MongoDB	• Analyse meta-data using MongoDB to ensure required information is present • Analyse ground truth to ensure required information is present		
L1	Week 2 Investigation of data set suggested in proposal, System design						
	Mon 06/02/17	Tue 07/02/17	Wed 08/02/17	Thu 09/02/17	Fri 10/02/17	Sat 11/02/17	Sun 12/02/17
L2	• Select and Install XML parsing library	• Parse ground truth XML and import to MongoDB	• Establish how to link ground truth data to coresponding images	• System design	• System design		
L1	Week 3 Image Aquasition and Rejection						
	Mon 13/02/17	Tue 14/02/17	Wed 15/02/17	Thu 16/02/17	Fri 17/02/17	Sat 18/02/17	Sun 19/02/17
L2	• Obtaining CT stacks for a given scan in the correct order	• Obtaining CT stacks for a given scan in the correct order	• Rejection of slices that do not show lungs	• Rejection of slices that do not show lungs	• Rejection of slices that do not show lungs		
L1	Week 4 Image Segmentaion						
	Mon 20/02/17	Tue 21/02/17	Wed 22/02/17	Thu 23/02/17	Fri 24/02/17	Sat 25/02/17	Sun 26/02/17
L2	• Implementing chosen segmentaion technique	• Implementing chosen segmentaion technique	• Creating ground truth for optimisation / testing	• Optimising / Testing segmentaion	• Optimising / Testing segmentaion		
L1	Week 5 ROI extraction.						
	Mon 27/02/17	Tue 28/02/17	Wed 01/03/17	Thu 02/03/17	Fri 03/03/17	Sat 04/03/17	Sun 05/03/17
L2	• ROI extraction	• ROI extraction					

	March						
L1	Week 1 First feature						
	Mon 27/02/17	Tue 28/02/17	Wed 01/03/17	Thu 02/03/17	Fri 03/03/17	Sat 04/03/17	Sun 05/03/17
L2			• Feature Framework	• Feature Framework	• First Feature		
L1	Week 2 Training and testing the classifier (Subset of data), inspection GUI						
	Mon 06/03/17	Tue 07/03/17	Wed 08/03/17	Thu 09/03/17	Fri 10/03/17	Sat 11/03/17	Sun 12/03/17
L2	• Testing and training	• Testing and training	• Inspection GUI	• Inspection GUI	◊ FIRST ITERATION COMPLETE		
L1	Week 3 Additional features						
L2	Mon 13/03/17	Tue 14/03/17	Wed 15/03/17	Thu 16/03/17	Fri 17/03/17	Sat 18/03/17	Sun 19/03/17
L1	Week 4 Additional features, Data normalisation to enable use of full set of data						
L2	Mon 20/03/17	Tue 21/03/17	Wed 22/03/17	Thu 23/03/17	Fri 24/03/17	Sat 25/03/17	Sun 26/03/17
L1	Week 5 Optimisation of threshods and parameters, experimenting with differnet classifiers						
L2	Mon 27/03/17	Tue 28/03/17	Wed 29/03/17	Thu 30/03/17	Fri 31/03/17	Sat 01/04/17	Sun 02/04/17

	April						
L1	Week 1 Statisical analisis of effectiveness						
	Mon 27/03/17	Tue 28/03/17	Wed 29/03/17	Thu 30/03/17	Fri 31/03/17	Sat 01/04/17	Sun 02/04/17
L2						• Statisical analisis of effectiveness	• Statisical analisis of effectiveness
L1	Week 2 Lead time for bug fixing						
	Mon 03/04/17	Tue 04/04/17	Wed 05/04/17	Thu 06/04/17	Fri 07/04/17	Sat 08/04/17	Sun 09/04/17
L2						◊ IMPLEMENTATION SHOULD BE COMPLETE	
L1	Week 3 Write up report						
L2	Mon 10/04/17	Tue 11/04/17	Wed 12/04/17	Thu 13/04/17	Fri 14/04/17	Sat 15/04/17	Sun 16/04/17
L1	Week 4 Write up report						
L2	Mon 17/04/17	Tue 18/04/17	Wed 19/04/17	Thu 20/04/17	Fri 21/04/17	Sat 22/04/17	Sun 23/04/17
L1	Week 5 Write up report						
L2	Mon 24/04/17	Tue 25/04/17	Wed 26/04/17	Thu 27/04/17	Fri 28/04/17	Sat 29/04/17	Sun 30/04/17

	May						
L1	Week 1 Write up report						
	Mon 01/05/17	Tue 02/05/17	Wed 03/05/17	Thu 04/05/17	Fri 05/05/17	Sat 06/05/17	Sun 07/05/17
L2					◊ HAND IN FINAL REPORT		

References

Cancer Research UK (2017a). *Lung cancer mortality statistics*. URL: `http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/lung-cancer/mortality` (visited on 01/25/2017).

– (2017b). *Lung cancer screening*. URL: `http://www.cancerresearchuk.org/about-cancer/type/lung-cancer/about/lung-cancer-screening` (visited on 01/25/2017).

Eysys (2017). *Eysys*. URL: `https://www.eysys.com/` (visited on 01/26/2017).

University of Rochester (2017). *Pulmonary Nodules*. URL: `https://www.urmc.rochester.edu/encyclopedia/content.aspx?contenttypeid=22&contentid=pulmonarynodules` (visited on 01/25/2017).