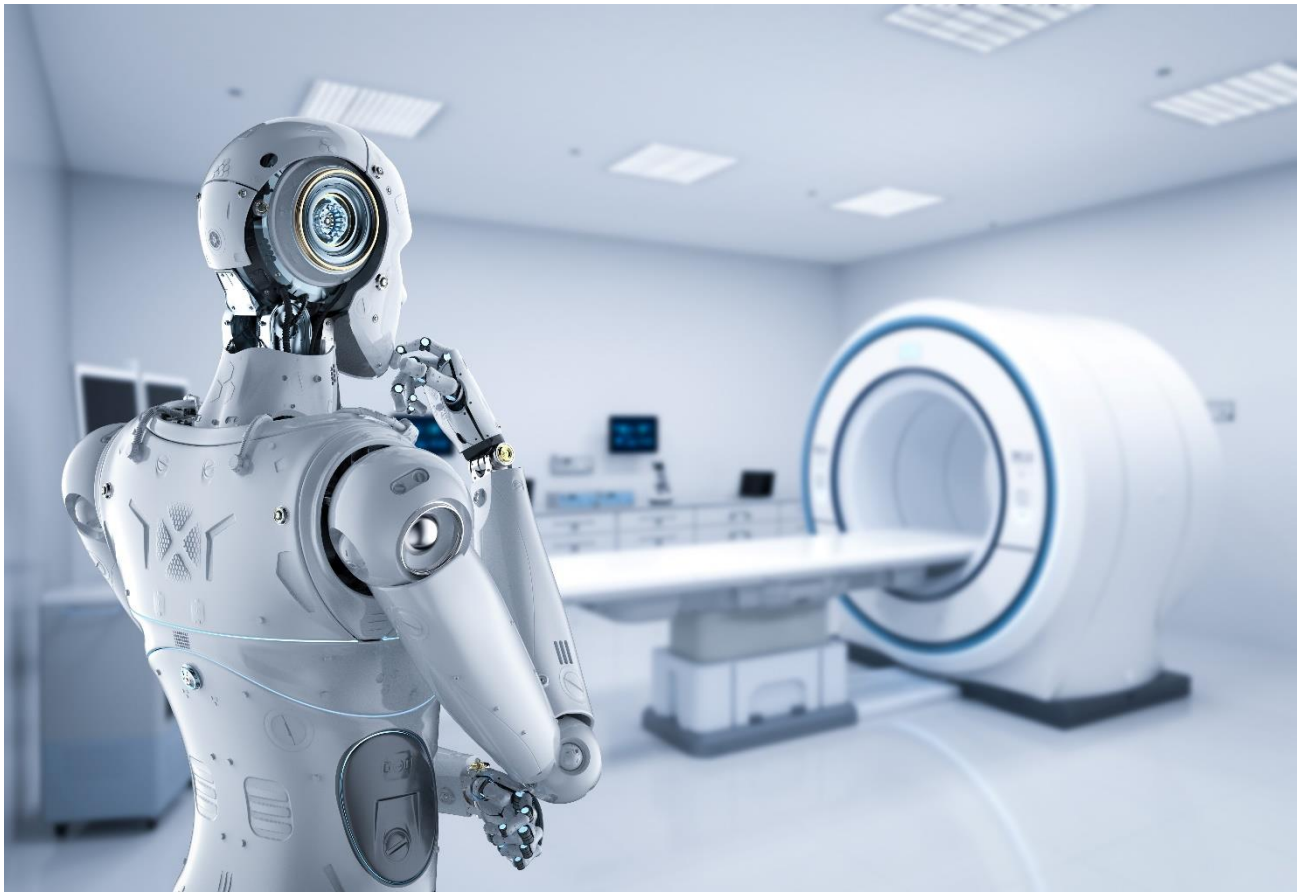
ROKE

AI Comparative CT Scans

Produced for: NHS AI Lab Skunkworks, NHSX

Against C274: AI Comparative CT Scans

Report No: 72/21/R/197/U
October 2021 - 00-003



AUTHORISATION

Author(s): Dan Heath, Tim Freeman, Tim Gaultney, and Richard Potter

Approved By

Richard Potter
Project Manager

Data governance statement

This proof of concept (TRL 4) is intended to demonstrate the technical validity of applying phase correlation, coherent point detection and deep learning techniques to CT scans in order to align multiple scans and detect lesions. It is not intended for deployment in a clinical or non-clinical setting without further development and compliance with the UK Medical Device Regulations 2002 where the product qualifies as a medical device.

This project does not qualify as research as per the UK Policy Framework for Health and Social Care Research and falls out of scope of HRA approval.

Patient data used in this project followed established information governance processes, including data anonymisation.

Data shown in this report does not contain direct or indirect identifiers, and there is no reasonable prospect of a patient being identifiable.

The data anonymisation process and Data Protection Impact Assessment were ratified by the George Eliot Hospital NHS Trust Information Governance Group on 22 Jun 2021.

Ethics approval for the release of this report was provided by the George Eliot Hospital NHS Trust Senior Information Risk Owner and Caldicott Guardian on 10 Nov 2021.

CONTENTS

EXECUTIVE SUMMARY

Background

Roke has developed a proof of concept that has the potential to speed up the analysis of Computerised Tomography (CT) scans to free up radiologists' time and help identify tissue growth. This work was funded by the NHS AI Lab Skunkworks programme in NHSX and was delivered in partnership with George Eliot Hospital (GEH) through the Accelerated Capability Environment (ACE) framework.

This 12-week research project has focussed on how to identify features in a CT scan and automatically align scan "slices" to enable early detection and diagnosis of lesions for patients.

The Challenge

The current process for radiology reviews with oncology patients is that disease progression is evaluated by comparing CT scans. Comparing scans is extremely meticulous job as the radiologist has to go through every quadrant of the radiological anatomy to evaluate increase/decrease of the disease or presence of new features. This means that radiologists can be prone to miss pathologies. The rates of miss vary widely by lesion size and location, but in the abdomen for example, different radiologists may make different interpretations in up to 37% of cases. (Siewert, 2008)

The disadvantages of the current process are as follows:

- **Time Consuming:** It typically takes a radiologist 30 to 40 minutes to assess scans from a single patient. Radiologists can only review scans from one dimension at any one time and lesions do not grow only in one dimension. One dimensional views can give a false perception of ”Static” or “Minimal change”, which in fact might be a ”significant change” if calculated in three dimensions.
- **Not Precise:** The manual alignment of images is not precise due to the variation in the position of the body when the scan is acquired. Where automatic alignment is available it is not perfect. Manual assessment can often lead to inaccurate comparison measurement(s) in any dimension as they are not overlapped.
- **Prone to miss:** There is no automated detection of new lesions and it is not easy to see small or developing lesions. A radiological miss can provide false re-assurance to patient(s) and could be fatal.

The Approach

The approach taken was to build a proof of concept Graphical User Interface (GUI) tool to deliver the following capabilities:

- Fast, automatic overlay of sequential CT scans, enabling users to compare lesion growth, in 2D and 3D easily.
- Deal with changes in body shape caused by, e.g., breathing and time between scans.
- Differentiate between different parts of the body (e.g. bone, organs, and lesions).
- Include automatic measurement of lesion size, both in 2D and 3D.
- Identify new lesions not present in previous scans.

A mixture of computer vision, machine learning, and deep learning techniques were explored, often containing novel methodological steps. The approach included three clear stages: 1) Data ingest, 2) Classical Computer Vision, and 3) Deep Learning.

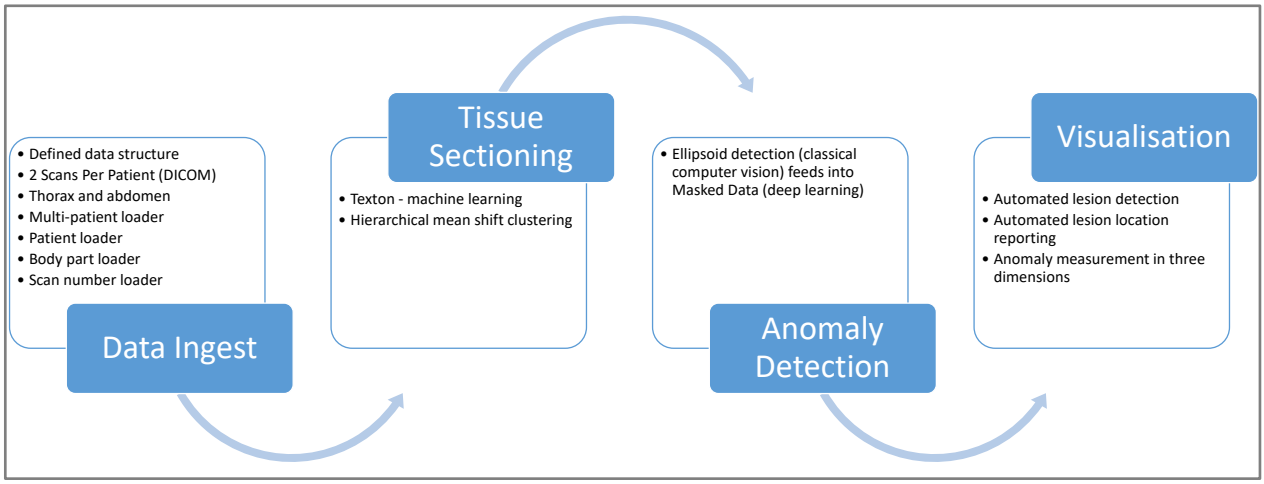
Project Successes

The project showed that AI can be used to support a radiologist in comparing and assessing sequential CT scans. The project delivered the following benefits that have the potential to deliver real value to radiologists:

1. **Alignment & Overlay -** Techniques to automate the alignment, overlay and visualisation of sequential CT scans.
2. **Tissue Sectioning -** The 'texton' machine learning technique was used to provide colour maps for different tissue types.
3. **Automated Lesion Detection & Measurement -** Automated detection, location reporting of potential lesions using AI, and 3D measurement of lesion(s) as shown below.
4. **3D Visualisation -** Overlaying, aligning and displaying sequential CT scans in two different colours.

Unique Application of Multiple AI Techniques

The project included the unique application of multiple techniqies to detect anomolies. This includes: 1) Machine learning using textons for tissue sectioning, 2) Classical computer vision techniques using ellipsoid detection for lesion detection, and 3) Deep learing techniques, including the recently released Facebook self-supervised learning method "DINO", for lesion detection.



Summary

This project has provided the opportunity to showcase the different capabilities available to data scientists to make a difference in healthcare using cutting edge AI techniques. It has shown the potential value AI could bring to the day to day working practices of radiologists by delivering: 1) Automated tools for sequential CT scan overlay, 2) Automated lesion detection, measurement and reporting, and 3) 3D visualistion of sequential CT scans.

Further work is required to develop promising AI techniques further beyond this 12 week proof of concept project. Whilst there is much work still to do to make this a usable solution, it is clear the technology has the potential to provide a much needed 'support tool' needed by Radiologists across the NHS.

1 INTRODUCTION

1.1 DOCUMENT PURPOSE

The purpose of this document is to summarise the work completed to build a proof of concept to demonstrate how Artificial Intelligence (AI) can be used to support a radiologist in comparing and assessing sequential Computerised Tomography (CT) scan images.

1.2 PROJECT BACKGROUND

Roke has developed a proof of concept that has the potential to speed up the analysis of CT scans to free up radiologists' time and help identify lesion growth. This work was funded by the NHS AI Lab Skunkworks programme in NHSX and was delivered in partnership with George Eliot Hospital (GEH) through the Accelerated Capability Environment (ACE) framework.

This 12 week research project has focussed on how to identify features in a CT scan and automatically align their scan "slices" to enhance detection of lesions.

1.3 DOCUMENT STRUCTURE

The structure of this document is as follows:

- **Section 2 - Project Approach:** Provides an overview of the methodology used by the project team to respond to the challenge.
- **Section 3 - Alignment:** Details the methods used to shift, rotate, and otherwise distort the latter of two scans from a particular patient such that it aligns well to the earlier of the two scans.
- **Section 4 - Tissue Sectioning:** Outlines why tissue sectioning is important when viewing and assessing CT scans, and details the methods used to this end.
- **Section 5 - Anomaly Detection:** Provides an overview of the techniques used to identify anomalies in CT scans.
- **Section 6 - Graphical User interface (GUI):** Provides an overview of the GUI that was developed to ingest CT scans from patients and to assess the tools and techniques that were developed.
- **Section 7 - Discussion:** Summarises the results of each of the techniques and the overall results.
- **Section 8 - Future Work:** Outlines work that could be completed to deliver further value to radiologists and patients.

2 PROJECT APPROACH

2.1 CURRENT PROCESS

2.1.1 OVERVIEW

The current process for radiology reviews for oncology patients is that disease progression is evaluated by the comparison of sequential scans. Comparing scans is an extremely meticulous and tedious job as the radiologist has to go through every tissue type of the radiological anatomy to evaluate increase/decrease of any prior lesion, or to detect the presence of a new feature. This means that radiologists can be prone to miss pathologies. The rates of miss vary widely by lesion size and location, but in the abdomen for example, different radiologists may make different interpretations in up to 37% of cases. (Siewert, 2008)

2.1.2 SCAN COMPARISON

The steps for comparing patient scans is as follows:

1. Images acquired via CT scan are sent to radiologists via the Picture Archiving and Communication System (PACS) for assessment.
2. The radiologist opens the patient folder and PACS may automatically display previous scan(s). However, if 3 years have passed between the two scans the previous scan is not automatically displayed and the radiologist must manually find the previous scan.
3. Typically the radiologist has to manually align multiple scans at the same approximate anatomical point, next to each other. In some cases automatic alignment can be achieved. 3D reconstruction of images is possible with some PACS software providers, but not all have this functionality.

2.1.3 DISADVANTAGES OF THE CURRENT PROCESS

The disadvantages of the current process are as follows:

- **Time Consuming:** It typically takes a radiologist 30 to 40 minutes to assess scans from a single patient. Radiologists can only review scans along one axis at any one time (for two-dimensional views) and lesions do not necessarily grow in only two dimensions. Two dimensional views can give a false perception of “Static” or “Minimal change”, which in fact might be a “significant change” if calculated in three dimensions.
- **Not Precise:** The manual alignment of images is not precise due to the variation in the position of the body when the scan is acquired. Where automatic alignment is available it is not perfect. Manual assessment can often lead to inaccurate comparison measurement(s) in any dimension as they are not overlapped.
- **Prone to miss:** There is no automated detection of new lesions and it is not easy to see small or developing lesions. A radiological miss can provide false re-assurance to patient(s) and could be fatal.

2.2 PROJECT METHODOLOGY

2.2.1 APPROACH

The approach taken was to build a ‘Proof of Concept’ (PoC) Graphical User Interface (GUI) tool to deliver the following capabilities that could be used as a decision support tool by radiologists:

- Fast, automatic overlay of sequential CT scans, enabling users to compare lesion growth, in 2D and 3D easily.
- Deal with changes in body shape caused by, e.g., breathing and time between scans.
- Differentiate between different parts of the body (e.g. bone, organs, and lesions).
- Include automatic measurement of lesion size, both in 2D and 3D.
- Identify new lesions not present in previous scans.

A mixture of computer vision, machine learning, and deep learning techniques were explored, often containing novel methodological steps. The approach included three clear sets of techniques: 1) Data ingest, 2) Classical Computer Vision, and 3) Deep Learning, as shown in Figure 1.

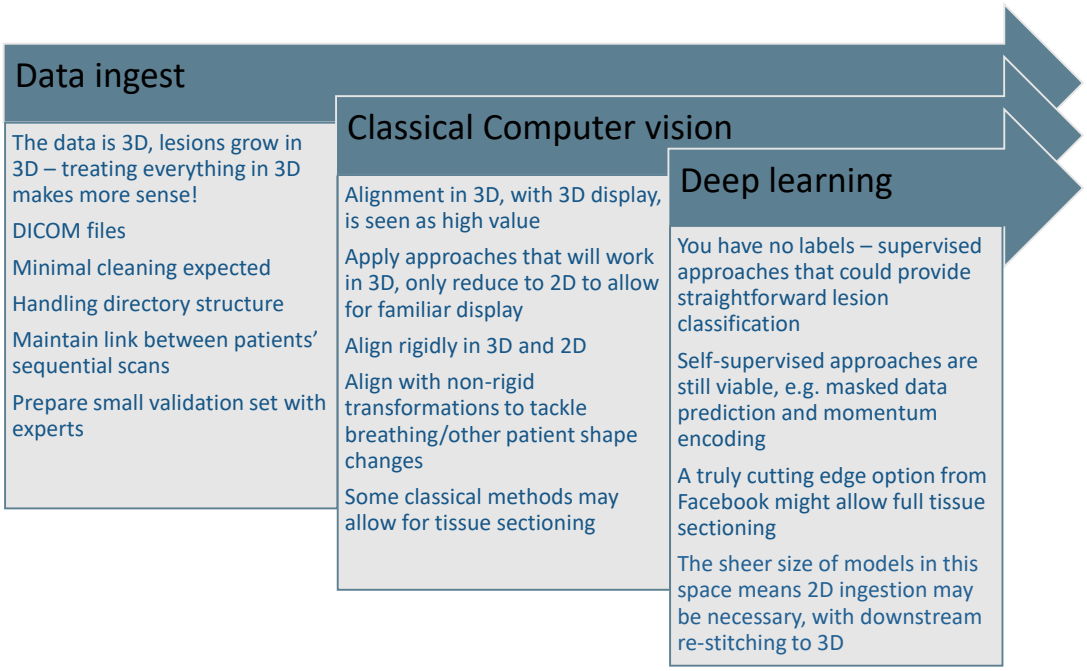


Figure 1 – Project Approach

The features built in each stage fed into the GUI tool, as shown in Figure 2.

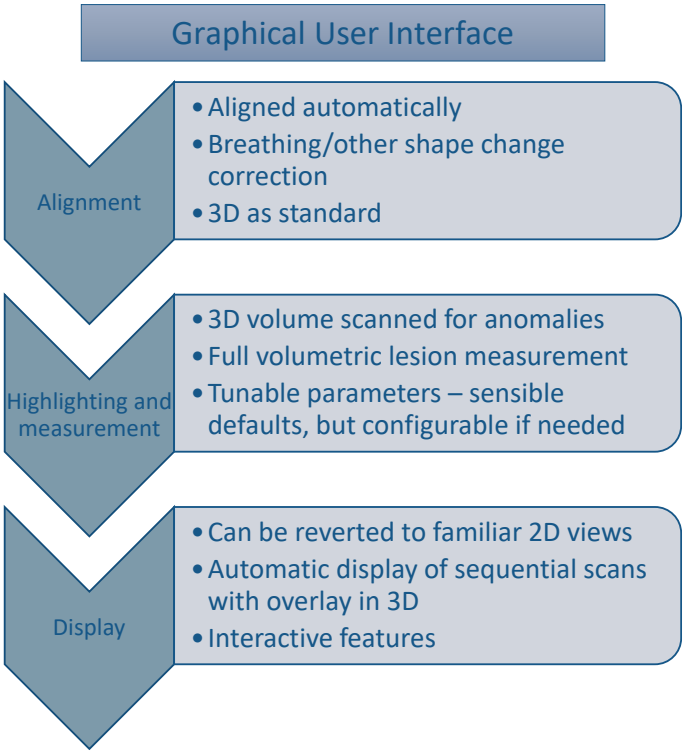


Figure 2 – GUI Tool Overview

2.2.2 Dataset & Data Flow

GEH provided Roke with CT scans from 100 patients in order to develop the PoC GUI tool. All patients in the dataset had developed lesion(s). Radiologist(s) from GEH marked up scans from 9 patients that identified where lesions were identified using existing methods.

The data flow from capture, governance, ingest, processing, and display is shown in Figure 3.

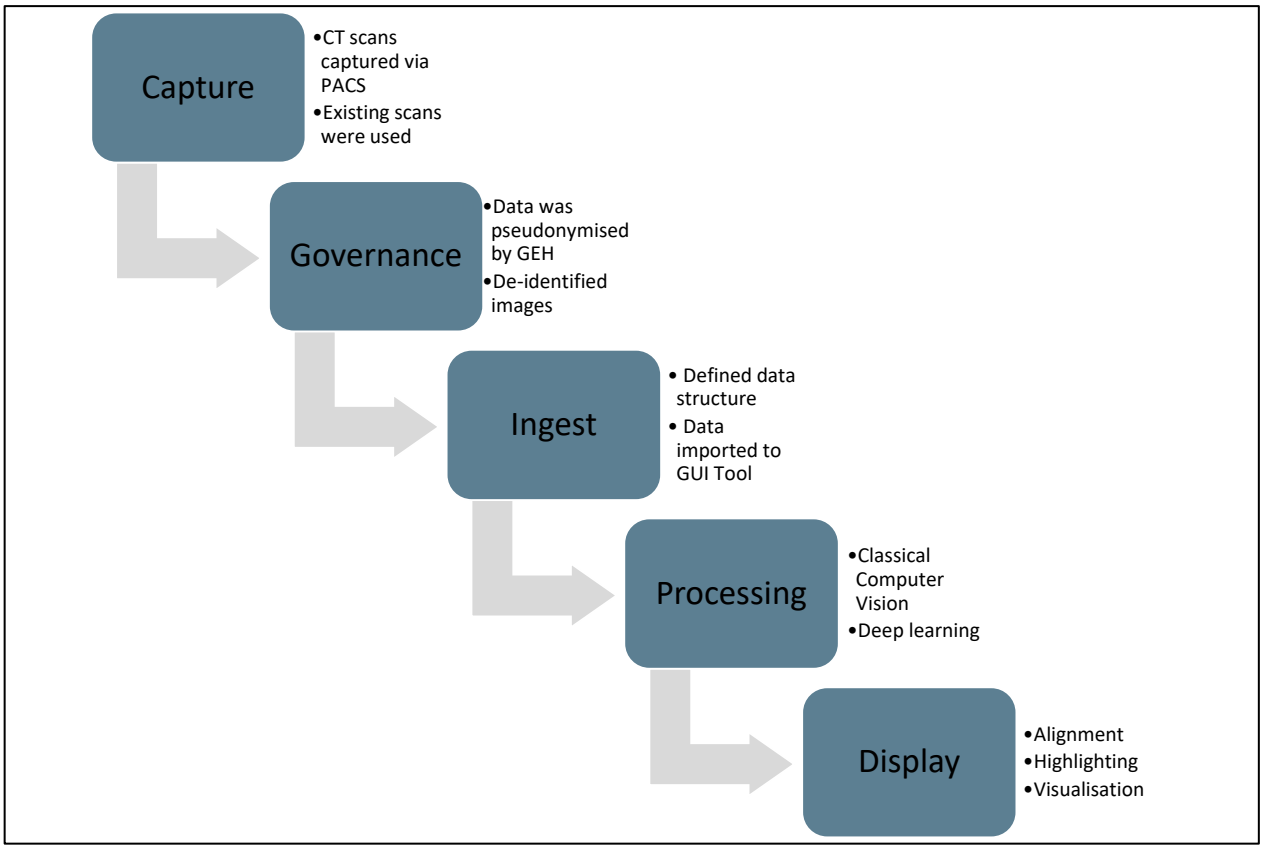


Figure 3 – Data Flow

2.2.3 Validation

Marked up scans from nine patients were provided as 'Ground truth' data near the end of the PoC work. As a result, most of the results lack quantitative metrics of success, and qualitative measures are given, usually in the form of 2D or 3D views of findings. However, where possible the results from the PoC were validated qualitatively against the 'ground truth' data to provide quantitative results (see section 5.3.2 for further information).

2.3 Project Plan & Timeline

An overview of the project plan is provided in Figure 4. The project was delivered in twelve weeks and was split into six, two week sprints.

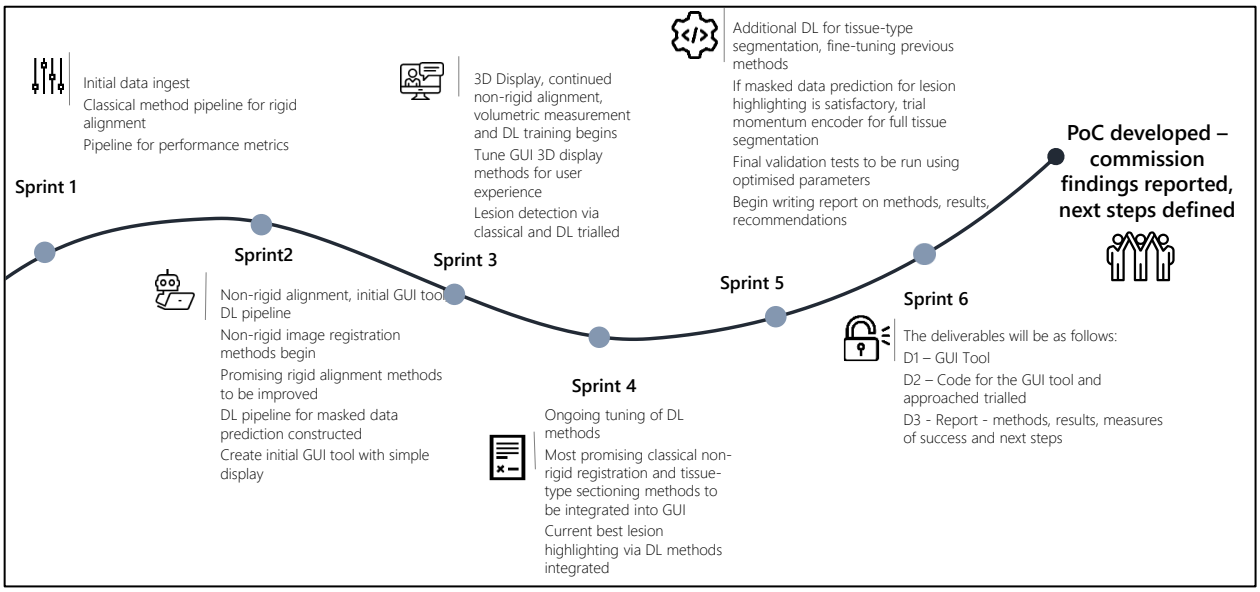


Figure 4 - Project Methodology Overview

A high level overview of each sprint is provided below:

- Sprint 1 - Data Ingest, Initial Classical Approaches 2D.
- Sprint 2 - Non-rigid alignment, initial GUI tool, Deep Learning pipeline.
- Sprint 3 - 3D Display, volumetric measurement and Deep Learning training.
- Sprint 4 - Down-selection of tools for integration into GUI.
- Sprint 5 - Additional Deep Learning for tissue-type segmentation and fine-tuning.
- Sprint 6 - Handover.

2.4 EXPECTED BENEFITS

The expected benefits that could be realised through the development of the GUI Tool are shown in Table 1.

Table 1 – Expected Benefits

Scan Comparison	Current Process	Expected Benefits From the Proof of Concept	Delivered functionality
Alignment	Variable	• Classical Computer Vision methods, in 3D as standard, with 2D representations easily recoverable • Rigid alignment of CT scans • Alignment with breathing and other patient body changes accounted for with more advanced classical methods	• Fast 3D rigid alignment achieved via phase correlation • Breathing and other body changes dealt with via coherent point drift, observed to work in most cases
Overlay	Imprecise	• Precise overlay • Rapid 3D and 2D overlays rapid • Interactive elements such as panning, zooming, rotation	• Precise overlay achieved • 3D and 2D overlays are in the GUI tool • Interactive elements included in GUI tool
Change in lesion size	In one dimension	• Mix of classical and deep learning approaches • Section lesions in 3D which allows fast volumetric measurements	• Novel deep learning approaches explored but classical approaches formed the measurement components • Tissue sectioning was achieved, lesions often sectioned distinctly from surrounding tissue
New pathological lesion	Prone to miss	• Two self-supervised deep learning approaches to automatically identify lesions	• Deep learning approaches trialled, potential for extension on these techniques defined, and highlighted lesions in some validation cases • Computer vision techniques used to identify many lesions, extensions to improve capture rate described
Current process	Tedious and manual	• GUI to aid automation	• Accurate overlay, automated reporting of anomaly locations and properties in the GUI should both speed up inspection
Guidance to Radiologist	Not available	• Potential lesions highlighted • Growth automatically measured • Location of detected lesions presented to radiologist	• Anomaly locations and other properties presented in the GUI • Size of lesions measured, but change in size not completed
Time cost	30-40 min	• Reduced time cost due automation and visualisation	• Not yet measured via radiologist usage
Trust Litigation cost	Immense on miss	• Reduced chance of misses • Decrease risk of litigation	• Fully automated measures unlikely to reduce the chance of miss, but the accurate overlay may help to reduce misses during manual inspections – not yet measured

3 ALIGNMENT

This section details the methods used to shift, rotate, and otherwise distort the latter of two scans from a particular patient and body part such that it aligns well to the earlier of the two scans. These methods all include some combination of shifting, rotating, skewing, or non-linear warping, and aim to align either 2D slices from the full 3D CT scans, a local 3D region, or the full 3D volume of the scans.

Throughout, the first of the two scans from each patient in the dataset will be referred to as 'scan 1', while the latter will be referred to as 'scan 2'. The dataset also contained both abdominal and thorax scans for each time of scan, but typically there was no separate tuning of the algorithms to either of these body regions. Results are usually presented from abdominal scans, unless stated otherwise.

3.1 MOTIVATION

A direct method for the manual discovery and measurement of unexpected features in two images which are expected to be similar is simply to overlay them, potentially in different colour channels, such that any differences are starkly visible. Simply showing 2D slices from the 3D volumes side by side is also common in current CT analysis software. Minor differences in patient position, body shape changes between sequential scans (which can be months or years apart), and differing lung volumes during each of the scans (among other confounding factors) often lead to the failure of automatic alignment methods, such that a simple overlay is not an accurate comparison method. Radiologists often have to manually reposition scans, usually in the axial direction, deciding upon a matched axial position by eye, and having little to no recourse when the patient is misaligned in the coronal and sagittal directions. The problem is displayed in Figure 5.

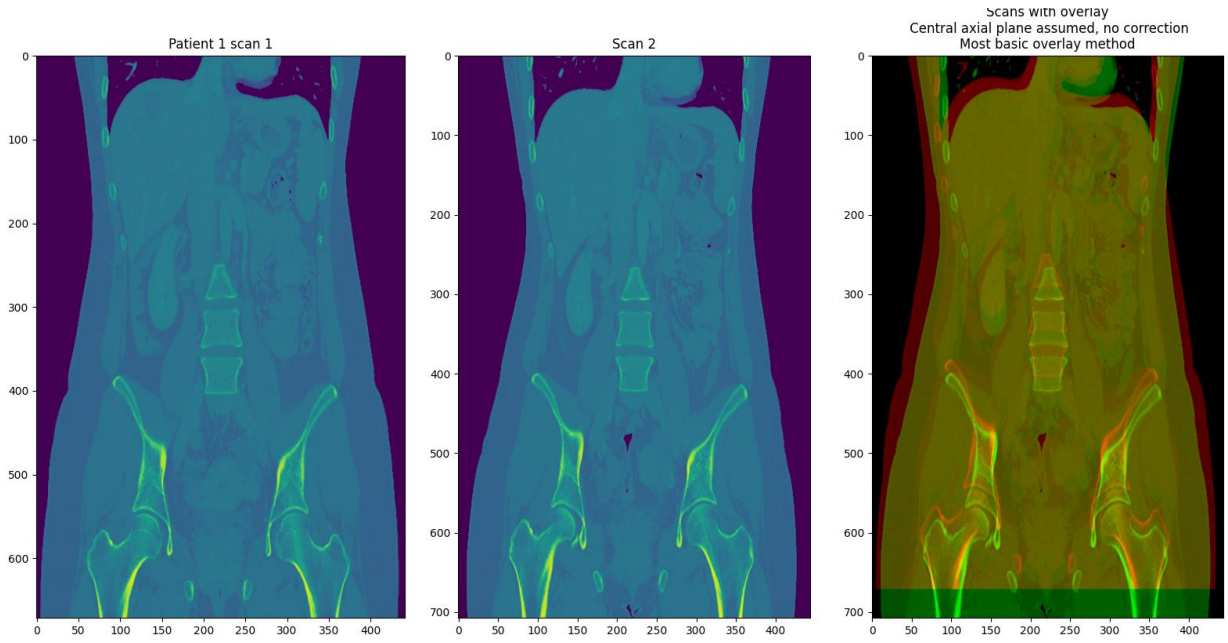


Figure 5 –Two 2D slices from sequential CT scans. In the right-hand image they have been overlaid in different colour channels, scan 1 in the red channel, scan 2 in the green channel. In this case, even with very little change of patient body shape and the same major internal structures visible, a major offset is present when the data is overlaid directly.

It is clear from Figure 5 that even with a vertical shift of scan 2, which would align many major skeletal components, it would not be possible to align all major bodily structures simultaneously with only shifts and rotations (a shift that brings the hip into alignment would bring ribs into non-alignment, for instance).

Additionally, the co-occurrence of many major structures at the same plane was largely coincidental for this patient – in many cases, there is a several millimetre axial, coronal and sagittal shift which leads to disparate structures appearing at the same slice indices. 3D methods that seek out the correct plane during alignment are necessary in the general case.

While 'rigid' techniques that stretch no part of the image might achieve a useful overlay on local regions, the ideal is to align all body parts simultaneously, such that new or growing lesions are easy to identify. Roke has experimented with a mixture of methodologies to both of these ends, which will be explained in the remainder of this section.

3.2 TECHNIQUES

3.2.1 RESCALING

CT scans can be generated with varying axial, coronal and sagittal spacing between pixels. When comparing sequential scans from the same patient, this can cause a simple overlay to be of little value, as major structures rarely align, if at all, as in Figure 6.

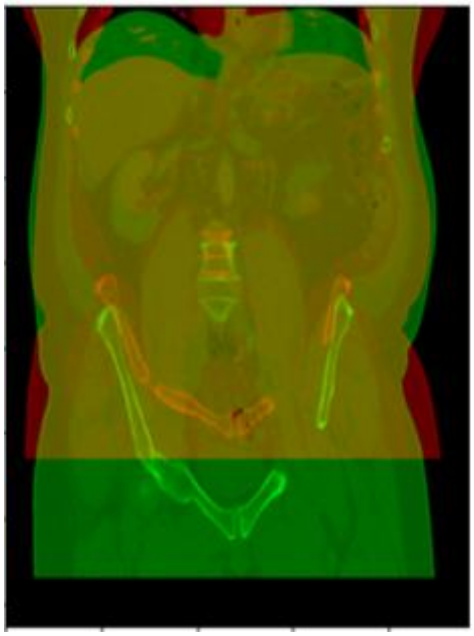


Figure 6 – A naïve coronal overlay from two sequential scans of the same patient, overlaid in red and green channels. The scan in red has been recorded with a different axial spacing, leading to a 'squashed' appearance, and a failure of the overlay to provide useful visualisation.

Metadata in the DICOM files of each scan associated with these scales was extracted and used to reach a common scale in each axis. This rescaling was used in all downstream alignment techniques.

3.2.2 PRE-PROCESSING

In many computer vision problems, the boundaries between objects in an image present useful and robust keypoints upon which to attempt matching to similar keypoints in a second image. Consider two images of a coloured cube under slightly different illumination – different levels of glare and hue due to the lighting lead to confounding keypoints related to these features, but the edges and corners of the cube remain useful keypoints.

Various ‘edge detection’ algorithms exist (Sharifi, 2002), and Roke has found a combination of local mean removal (LMR) and zero crossings detection to be a good general purpose method to use when edge information is required. While all methods in this section were trialled without edge detection pre-processing as well, performance with this pre-processing step showed greatly improved robustness. Outputs at various steps of the LMR and zero crossings procedure are shown in Figure 7.

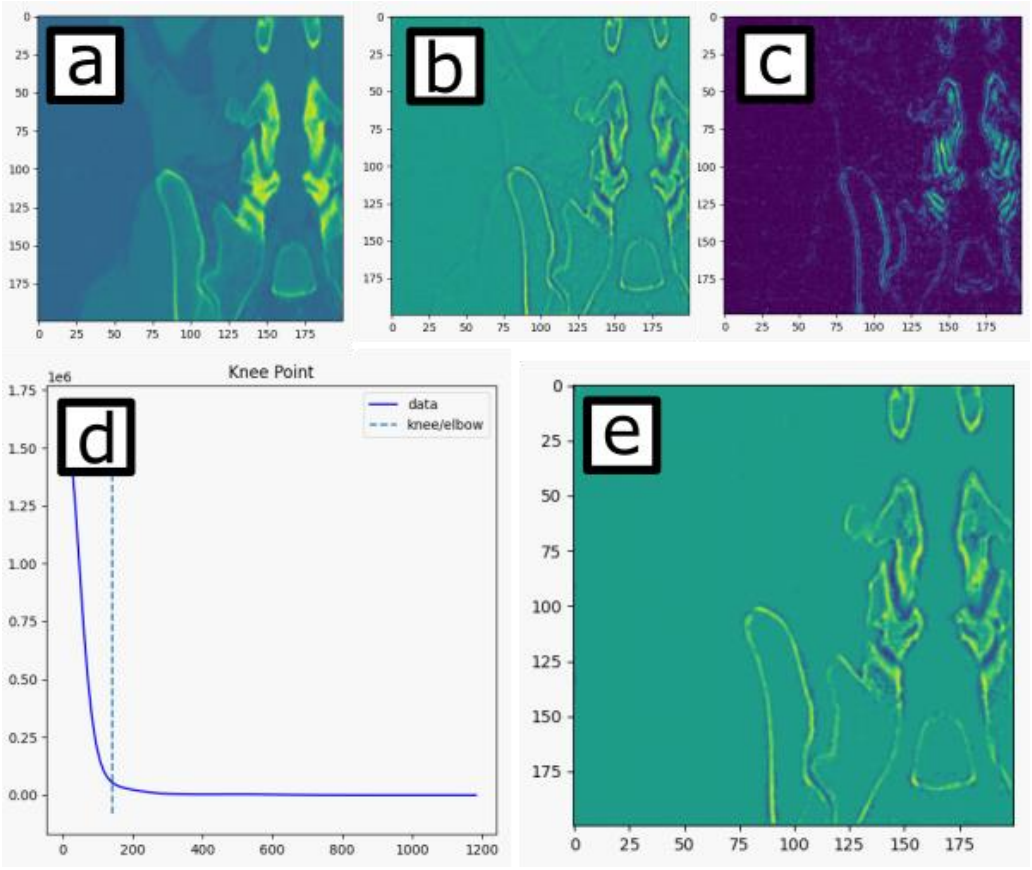


Figure 7 – LMR and zero crossings pre-processing steps. (a) the original data, (b) LMR applied, (c) adjacent pixel differences taken (d) knee-detection on magnitude of local differences (e) LMR points with neighbours above a threshold jump across 0 retained.

The algorithm is presented in Figure 7 in 2D views, but was also applied on the full 3D data of scans. (a) A cropped 2D coronal view from a CT scan. (b) Each pixel from (a) has had the mean value of pixels within a radius of 2 subtracted from it (the LMR step). (c) The magnitude of the difference between pixels and their adjacent neighbours is taken. (d) A ‘knee detection’ (Arvai, n.d.) is performed on the histogram of values in (c), to determine a critical value above which the ‘most prominent’ pixels in (c) possess. (e) the value from (d) is used as a threshold by which adjacent pixels must ‘jump’ across zero, in order to be retained. This results in an image which primarily shows edges of major structures, which can be fed to downstream alignment methods. Many edge detection algorithms in open source libraries exist, e.g. Canny Edge Detection, but Roke has found previously that the above technique requires little manual tuning.

3.2.3 Phase Correlation

Patients are usually aligned in similar orientations when a CT scan is taken, under guidance by the administering radiographer. In computer vision, seeking a transform among possible affine transformations is common when attempting to align similar images, where translations, rotations, reflections, scale changes, and shears are all possible. In the case of CT scans, however, where patients are guided in their orientation by a specialist and scale has already been accounted for as in section 3.2.1, translations (or ‘shifts’) might be expected to be the dominating contributor to a suitable remapping.

Phase correlation (H. Foroosh, 2002) offers the calculation of shifts between images, without regard to other affine transformations. Open source methods are often only applicable for two-dimensional images, while in this case we sought a 3D shift. Roke implemented a fully 3D version of the algorithm, which allowed us to shift scan 2 in 3D such that it overlaid onto scan 1.

Results of the technique are given in section 3.3.1.

3.2.4 KEYPOINT METHODS

Whilst the phase correlation method described in section 3.2.3 was shown to achieve strong alignment in a local region in the majority of cases, it does not provide full global alignment of a 2D slice, often leading to significant misalignment outside the defined local region. A radiologist conducting manual analysis of a scan may have to perform local alignment many times over as different regions are inspected. For this reason, methods of performing global alignment on a 2D slice have been considered, specifically keypoint detection based methods.

The keypoint detection method developed consists of the following steps:

1. Extract keypoints and descriptors from both scan 1 and scan 2
2. Match the keypoints from the two scans based on their descriptors using Flann matching (Muja & Lowe, 2009)
3. Filter the set of matched keypoints using Lowe's ratio test
4. Estimate the homography that maps the keypoints in scan 2 to the matched points in scan 1
5. Apply the homography to the full 2D slice from scan 2 to produce an aligned scan

Two different techniques for keypoint detection have been applied, SIFT (Lowe, 1999) and ORB (Rublee, Rabaud, Konolige, & Bradski, 2011). Key stages from the keypoint alignment method are shown in Figure 8 and Figure 9.

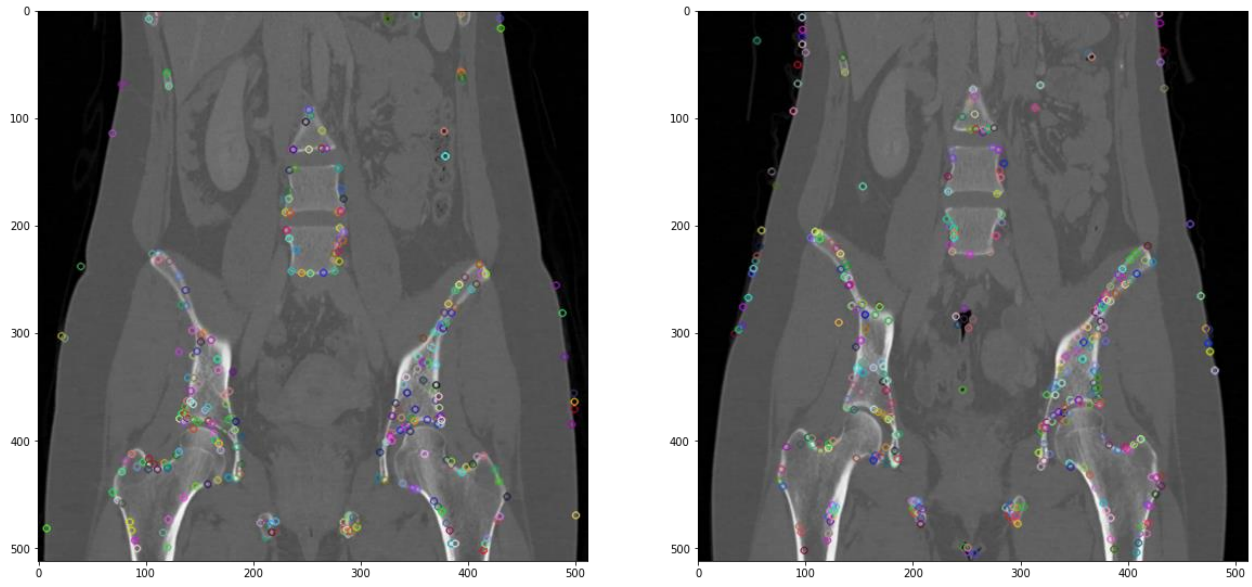


Figure 8 – Keypoints extracted from a pair of scans using an ORB feature detector.

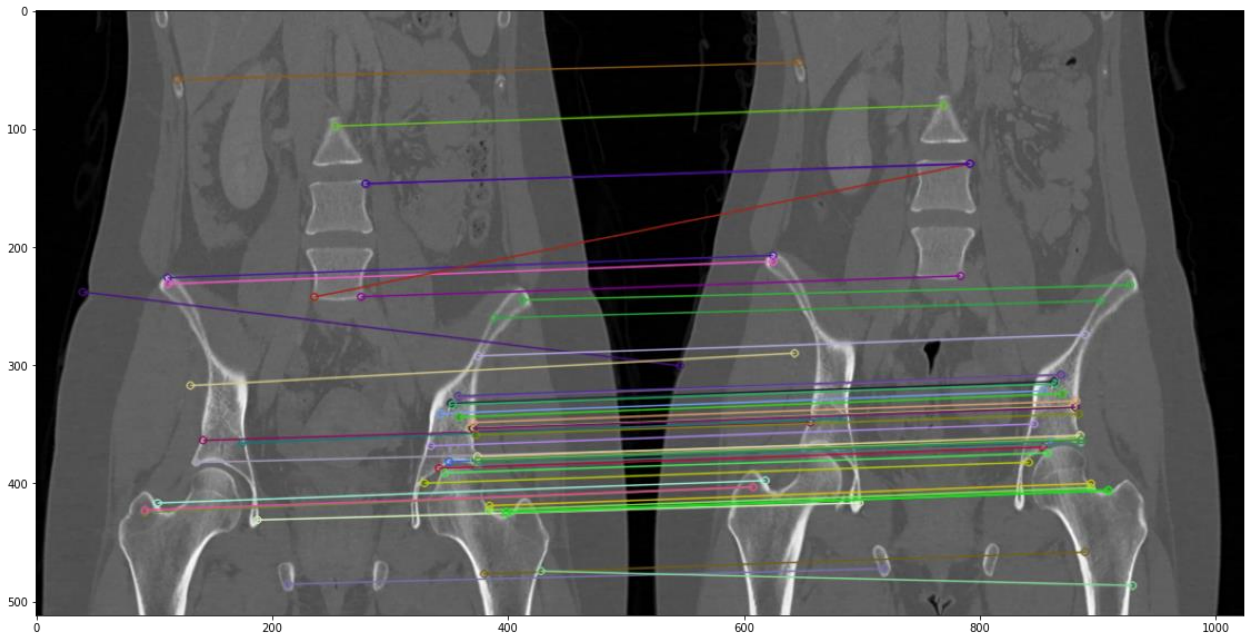


Figure 9 – Result of keypoint matching using FLANN and filtering using Lowe's ratio test on keypoints extracted for a pair of scans.

Global alignment methods that operate over 2D slices are fundamentally constrained by any misalignment that occurs in the third dimension, an example of which is shown in Figure 10. In this example different features can be seen in each scan, such as different parts of the spine, femur and ilium. Whilst this can be partially mitigated by applying an offset between slice indices in each scan, found using either manual or automated methods such as phase correlation, a global offset applied to across the full slice will not fully correct this effect in all cases, such as in the case where the patient does not lie in the same position.

Figure 10 – Example slice from a pair of scans that each display different features.

In order to overcome this limitation, methods of applying keypoint detection based alignment to full 3D volumes have been considered. An iterative method has been trialled in which a 2D alignment is applied in each of the three axis in turn. This approach produced mixed results, suffering from errors being amplified over each iteration, and as such was not pursued further.

3.2.5 COHERENT POINT DRIFT

As has been seen with previous alignment techniques, there will be differing relative transforms between separate structures in the body of a patient between two sequential scans, such that these structures cannot all be aligned ideally with a global affine transformation. A true global alignment requires the use of 'non-rigid' methods, where some regions can be warped differently than others.

Fully 3D methods in the literature to this end are sparse, though Coherent Point Drift (CPD) (Song, 2010) is one popular available technique which offers the required non-rigid transformations. Additionally, an open source Python library is available, utilising Cython for fast execution, which enabled Roke to rapidly trial the method (Gattia, 2021).

CPD can align points in 2D or 3D, using rigid or non-rigid implementations. It is an iterative algorithm, where 'source' points are moved closer to a set of 'target' points on each iteration, with non-linear adjustments made between the source points. Movement is only applied to the source points (typically from scan 2), so that target points (typically from scan 1) remain static throughout. It is robust to missing data in either point set, as well as the presence of extraneous points. The results after 50 iterations for aligning 3D points defining the outer surface of a rabbit shape are shown in Figure 11.

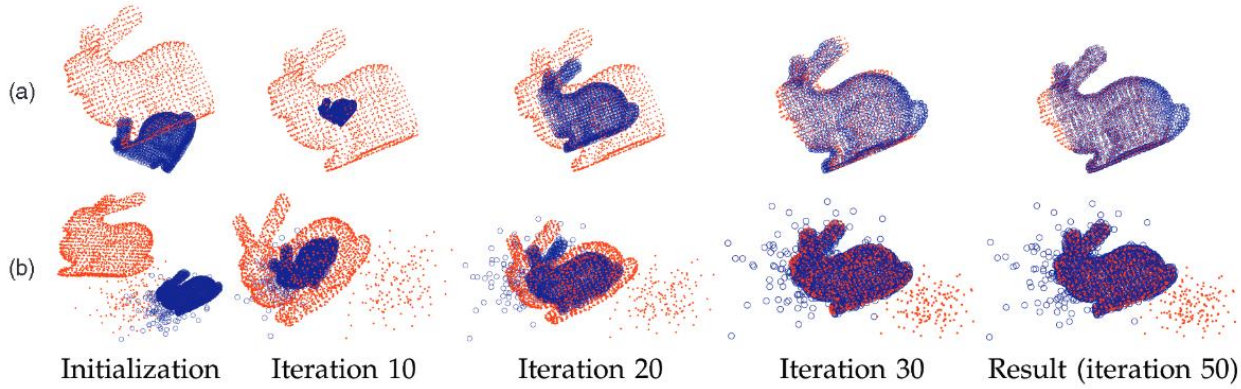


Figure 11 – The CPD algorithm on (blue) source points to align to (red) target points, both defining the surface of a rabbit, (Song, 2010). (a) The result during 50 iterations when a region of the blue source points are absent. (b) The result during 50 iterations when significant extraneous 'noise' points are included in each set.

While CPD then achieves impressive results, it is important to note that it runs natively on *point sets*, rather than full images. It does not explicitly 'match' points between the two sets, though we achieved this with Hungarian matching (Hungarian algorithm, n.d.) on Euclidean distance. Additionally, it is non-trivial to apply the algorithm's transform to new points after a solution has been reached.

To use this method for alignment on our 3D CT scan data, then, suitable point sets must be extracted from scan 1 and scan 2, the CPD algorithm run upon these, and then a transform extracted which can be run on the full 3D image data of scan 2. To this end, the LMR method followed by a thresholding was used to extract points defining strong edges in each scan. CPD was run upon these sets to align the points, and Hungarian matching was used to relate the points in each set to their nearest match. A polynomial fit to these points with ridge loss was applied, resulting in a pipeline that could apply the discovered transformation to the full 3D CT data.

Initial experimentation highlighted the need to detect and remove the table during point extraction, as the alignment of the table and patient across the two scans are not necessarily correlated. An algorithm has been implemented to find the location of the table in a scan based on the observation that the table appears as vertical lines when viewed in the Sagittal plane. The table detection algorithm consists of the following steps:

1. Begin by taking the central Sagittal slice, as shown in the top left of Figure 12.
2. Apply a threshold to the 2D slice to produce a binary image, as shown in the top middle of Figure 12.
3. Apply a Sobel filter (Sobel, 2014) to detect vertical lines, as shown in the top right of Figure 12.
4. Apply Hough transform (Duda & Hart, 1972) to identify the line that correlates to top of the table
5. Repeat the process for slices either side of mid slice until the table is no longer detected, as shown by the red line in Figure 12.fre

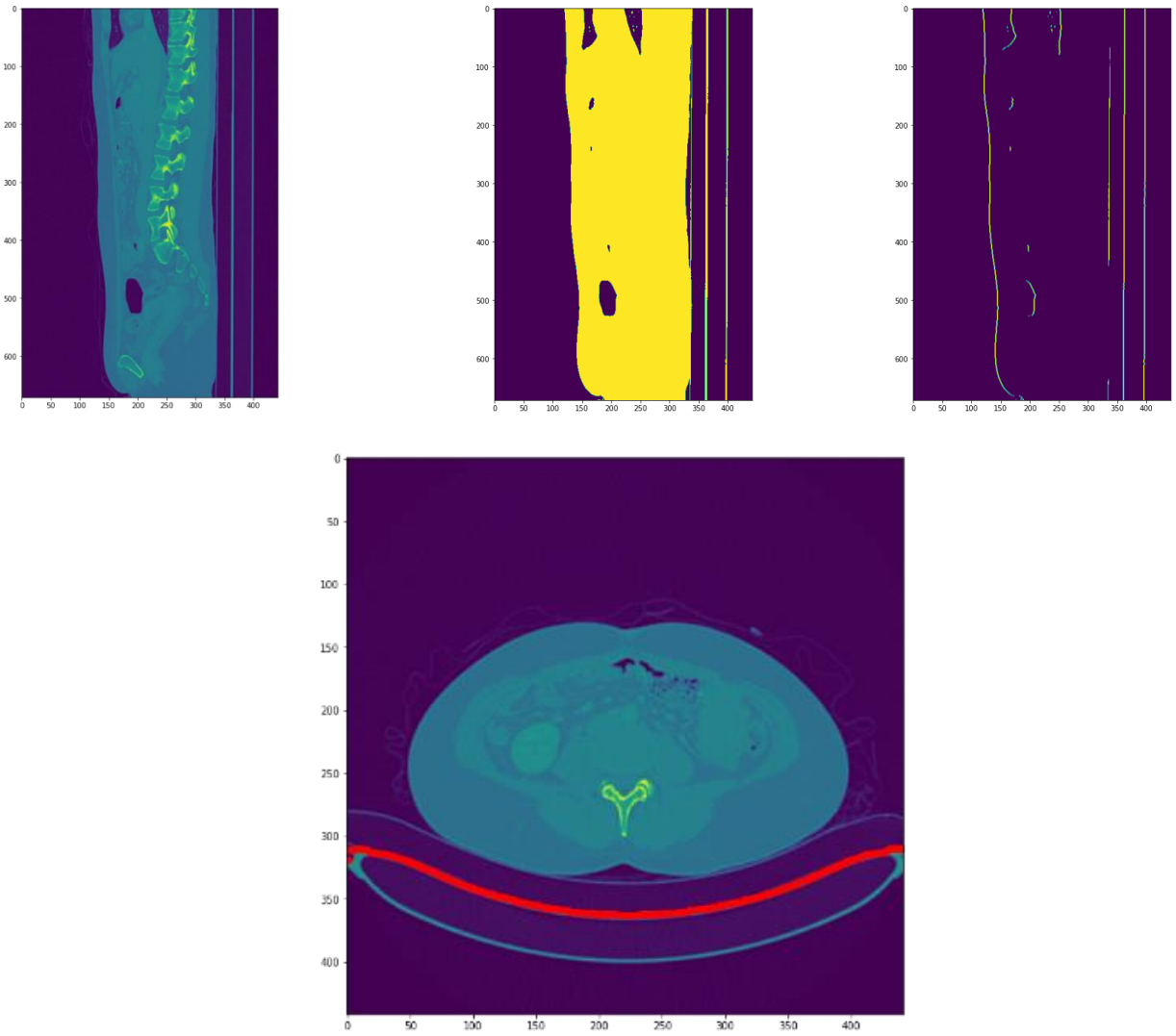


Figure 12 – Illustration of the key steps in the algorithm used to detect the table in a full 3D scan. In the top row, the first image is the original sagittal slice, the second image is a binary version where pixels above a threshold are retained, the third image then relates to a Hough transform being used to find vertical lines. The lower image then shows the surface of the table identified. Pixels vertically below the highlighted red points are then trivially removed, such that the table has been filtered out.

Applying a non-rigid alignment results in a varying magnitude of warping being applied across a scan, causing the size of legions to be distorted to different levels at different locations within the scan. In response, an overlay has been developed which will indicate the level volumetric change that has been applied to any area of an aligned scan, based on the alignment transform that is being applied. The metric that is used to measure the warp is the fractional volumetric change, given by the following equation:

$$Fractional\ Volumetric\ Change = \frac{V_{deformed} - V_{initial}}{V_{initial}}$$

This can be calculated using components of the strain tensor over the full aligned volume.

Figure 13 shows an example of the warp overlay being applied to a slice from an aligned scan. It can be seen from this example that the greatest volumetric change occurs at the edges of the scan, beyond where the patient is located, as no points would be extracted to align in this region. The volumetric change is significantly less across the body of the patient, however there are regions in which the volumetric change is greater than 10%, impacting measurement of legions that appear in those regions.

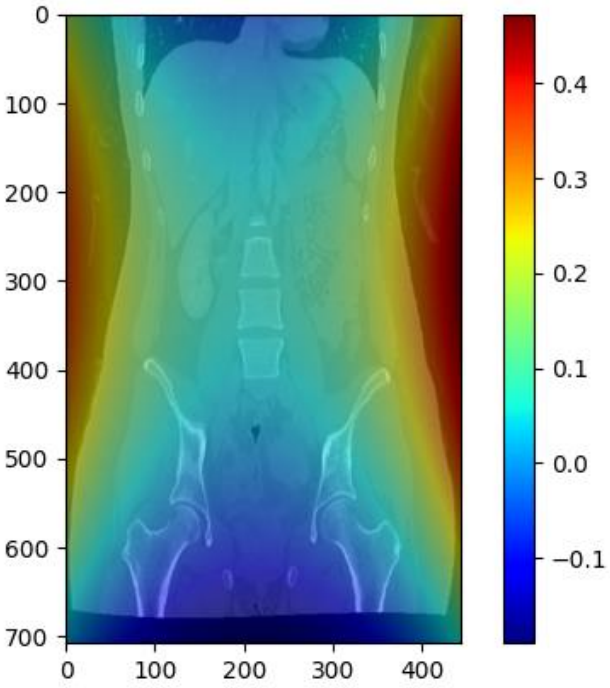


Figure 13 – Example of a warp overlay applied to a slice from a scan aligned using the non-rigid CPD based method. Vertical and horizontal axes are pixel positions. Colour map is volumetric warp fraction.

Further collaboration with expert users is required to fully validate the magnitude of volumetric change that is tolerable and refine the method by which it is presented to the radiologist.

3.3 Results

3.3.1 Phase Correlation

Figure 14 shows the result of a naïve overlay, with no alignment aside from the rescaling step performed. Most regions are offset by many pixels, translating to a several millimetre misalignment between scans.

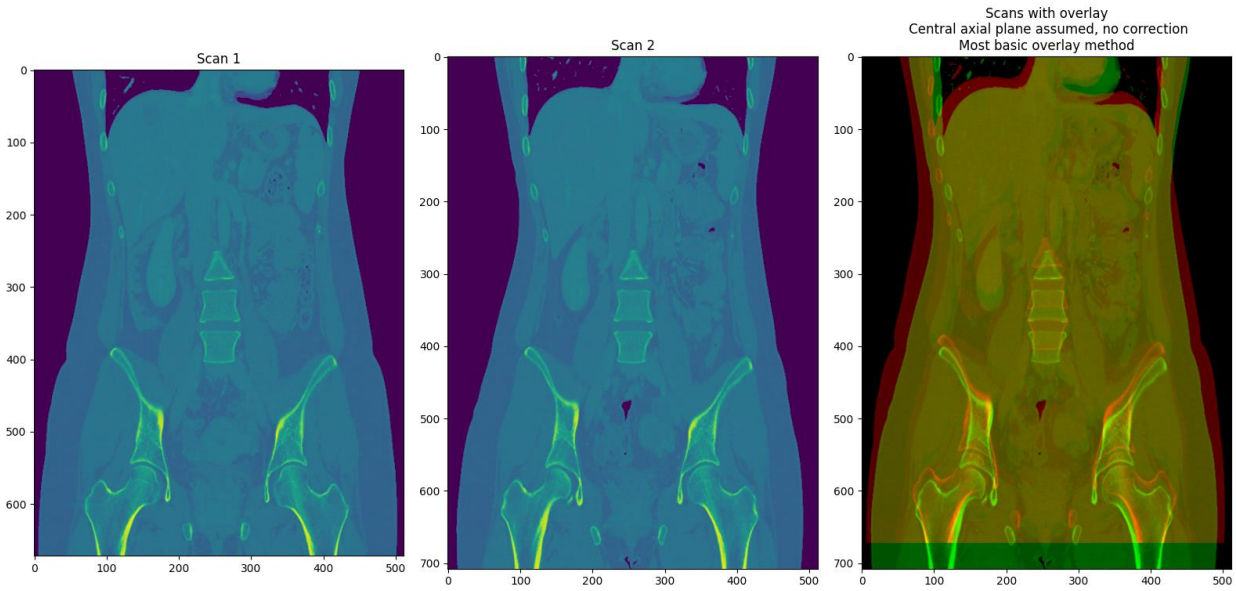


Figure 14 – Scan 1 and Scan 2 overlaid in red and green channels, with no alignment performed. The central coronal plane has been assumed. A misalignment on the order of 10 pixels is evident in most regions. While most of the same internal structure is visible on the selected slices of each scan in this case, i.e. the coronal offset is low for this patient, this is not generally true.

Figure 15 shows the resulting overlay of the scan after phase correlation has been applied (N.B., the slice displayed here was selected by taking a centre of mass calculation of intensity over the full 3D data, rather than assuming the central coronal slice as in Figure 14, leading to slightly different internal structure visible for both slices, but this had no impact on the phase correlation result, which examines every layer regardless).

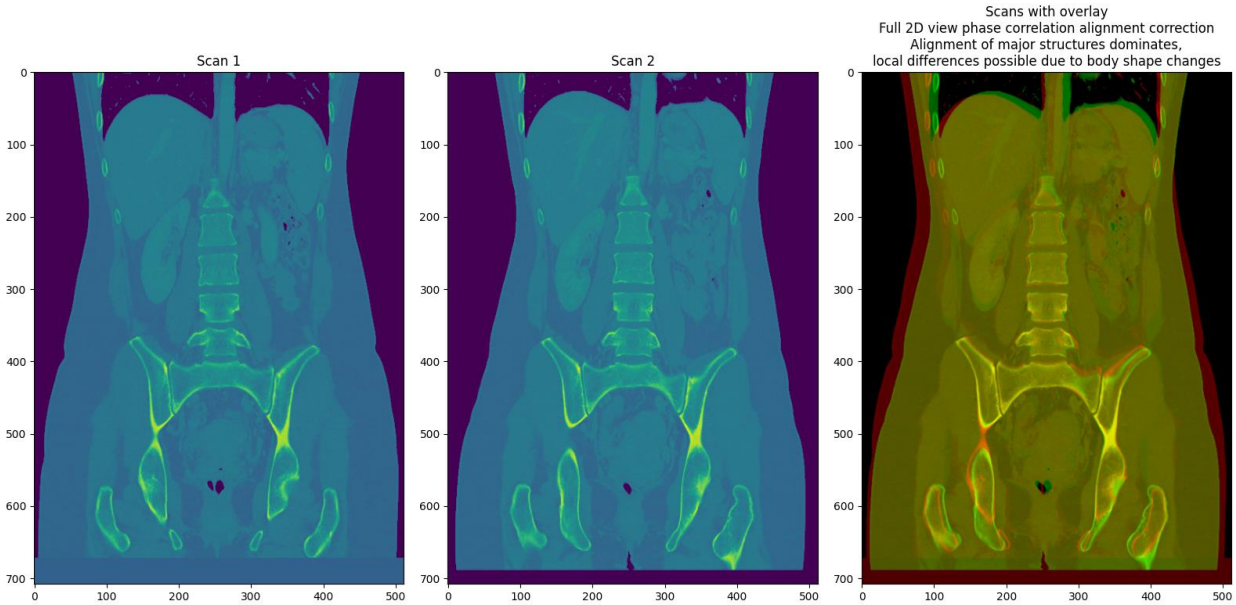


Figure 15 – Scan 1 and Scan 2 aligned via global phase correlation.

The overlay in Figure 15 is a significant improvement. Most major structures have been overlaid well. Some minor differences occur around the lower left quadrant in the hip bones, where presumably the patient is resting in a slightly different position between scans, and around the ribs, where the volume of held breath is different between scans.

Neither of these discrepancies are recoverable while simultaneously matching other global features via only shift-based alignment – while a hip joint may be misaligned by a few pixels, a rib may be misaligned by many more pixels, and hence applying a global shift cannot recover alignment of both structures at once. The results of keypoint and CPD methods to achieve global alignment will be addressed in later sections.

While a true global alignment is desirable, phase correlation quickly showed great promise for achieving alignment over major features, and a phase-correlation based 'local alignment' method was explored. Here, a user selects a pixel position, and a 3D volume around that point is defined, upon which to detect the shift of local features. The detected shift was then applied to the global scan 2 data, resulting in good alignment of the selected feature, as with the hip bone in Figure 16.

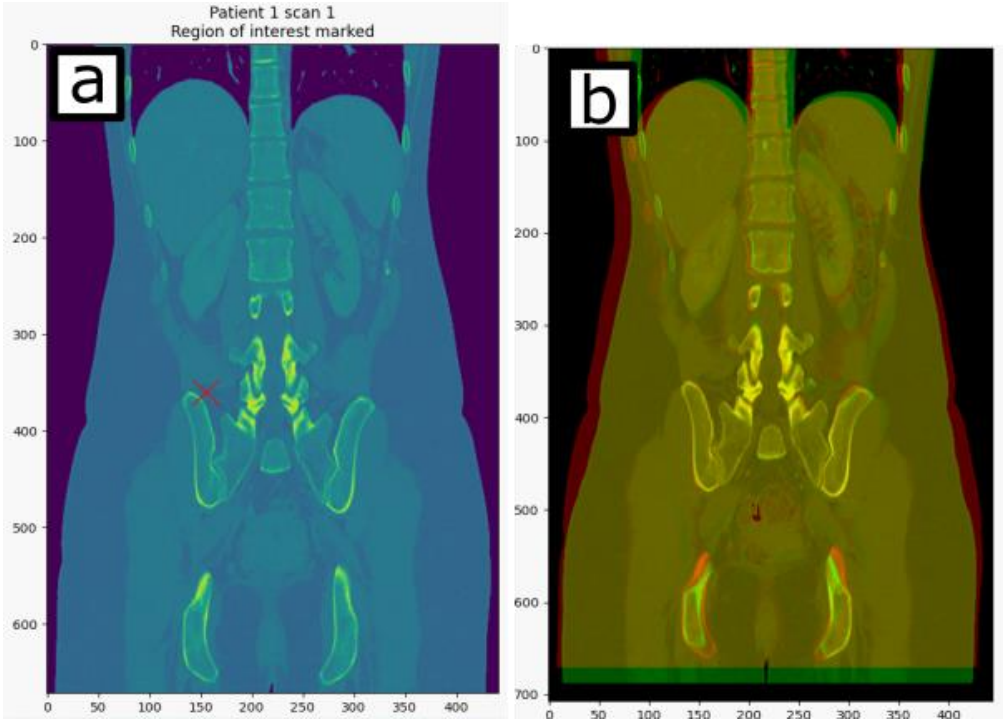


Figure 16 – Local alignment of a hip bone (highlighted with a red X in (a)). (b) shows excellent local alignment around the region

100 pixels (70mm) in positive and negative, in all three axes, defined a local volume around the point at the red X in Figure 16 upon which to align scan 2 using phase correlation. The results show excellent local alignment, and typically takes a few seconds to run from the GUI. A radiologist might use this alignment method upon inspection of a lesion from either scan to generate a well-aligned local overlay on demand, or when automated regions of interest are detected a pipeline to generate a batch of aligned views for each scan pair could be programmed. The local alignment phase correlation method works best when centred upon a clear feature – when a region with little variation is selected, the quality of overlay varies.

The challenge remains to develop a method for achieving global alignment, either in 2D or 3D, but phase correlation-based methods leave only minor room for improvement already.

3.3.2 KEYPOINT METHODS

The keypoint based methods have been predominantly used to perform alignment of 2D slices from scans. Whilst it would be possible to restrict the detection region in order to produce a local alignment capability, there was a significant challenge in ensuring enough matching keypoints are extracted to estimate a homography, and initial results were inferior to those from the phase correlation approach, particularly as it cannot be easily extended into 3D. For these reasons, keypoint methods were always applied to full 2D slices, rather than localised regions.

A number of examples are presented, each showing alignment of the central slice in the coronal plane. In each case scan 1 and scan 2 have been overlaid using the red and green channels. Examples have been selected that show a mixture of strong alignment results, challenging cases and cases where the keypoint alignment approach fails.

Figure 17 shows the results of keypoint alignment applied to slices from scans of patient 1, which represents one of the easier cases in the dataset used throughout the project. The naïve overlay on the left shows significant misalignment over the full scan. The method using both feature detectors produce very similar results for the Coronal plane and slightly differing results for the Axial and Sagittal planes. While there are some regions that display a visible misalignment, particularly around the edge of the skin tissue and bones at the bottom of the Coronal slice, overall the alignment looks strong across the full scan, aligning most of the key features, including the ribs, spine, pelvis and prominent tissue structures, with the possible exception of the ORB alignment in the Sagittal plane. In this case the alignment appears good for some regions, particularly in the middle, but becomes worse at the top and bottom.

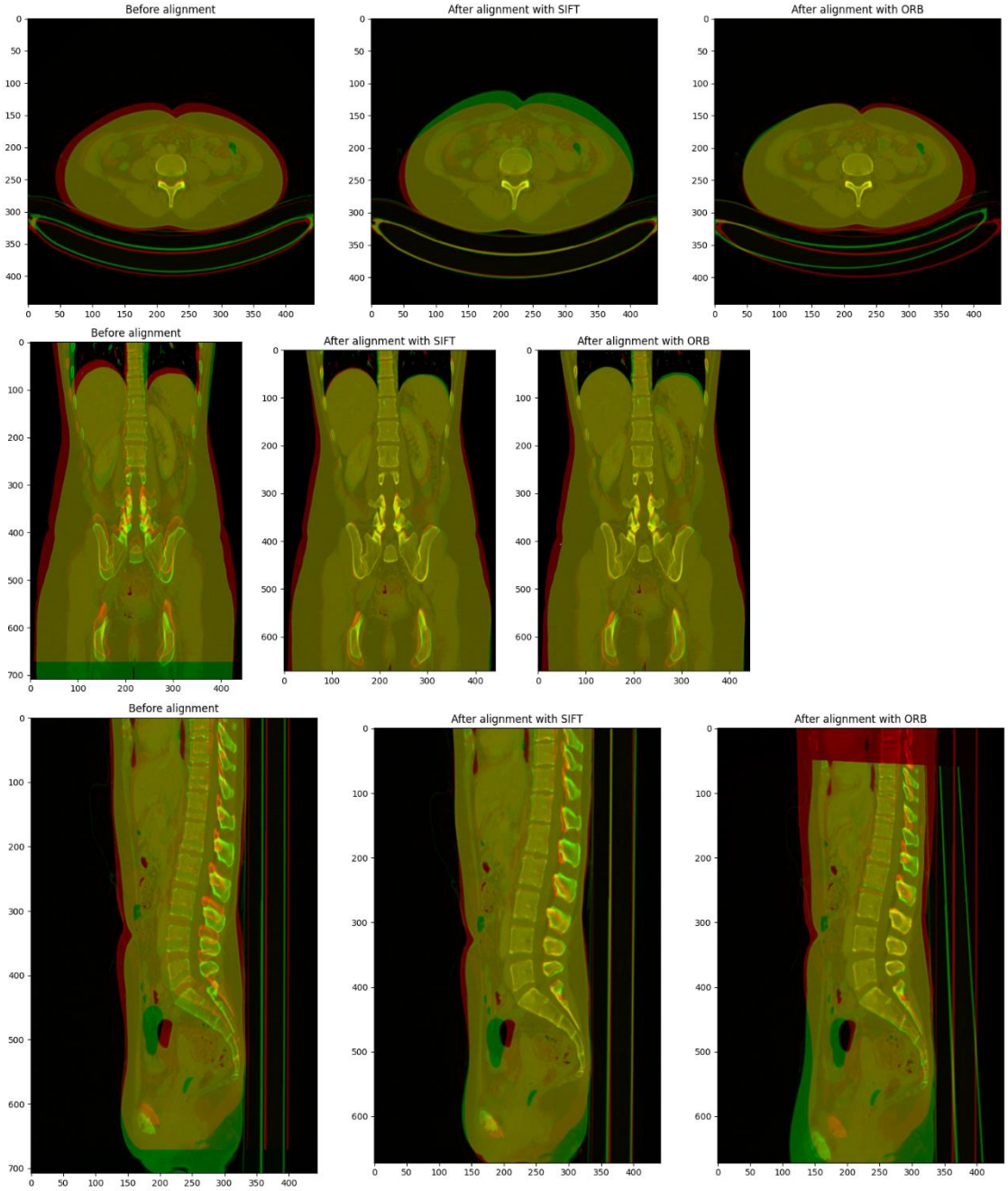


Figure 17 – Keypoint alignment of the 2D central Axial (top) Coronal (middle) and Sagittal (bottom) slice from patient 1 using both SIFT and ORB feature detectors.

Figure 18 shows the results of alignment on slices from scan 2, which represents one of the harder alignment cases within the dataset used, due to the magnitude of the change to the patient between the two scans. In this case there is a significant difference between the results of the SIFT and ORB based approaches in all three planes. For both the axial and sagittal views the ORB based method fails altogether while the SIFT based method produces reasonable alignment, at least over a significant region of the slice. The SIFT based approach also appears to produce a better alignment in the coronal view, particularly on the right hand side of the scan where the ORB based approach performs poorly. Although the alignment shows significant regions of misalignment, particularly around the edge of the skin tissue, generally the SIFT based approach produces a good overall global alignment.

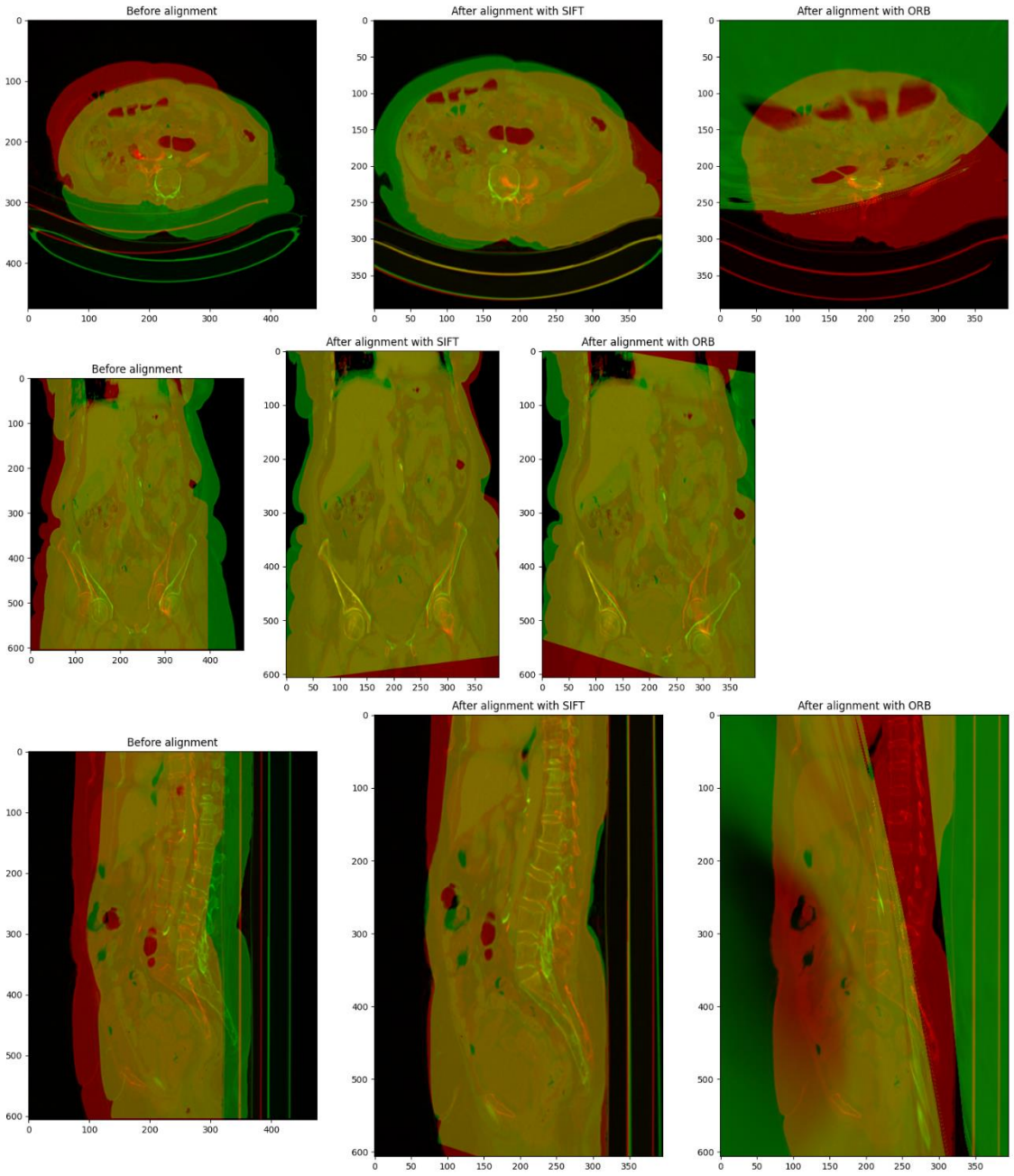


Figure 18 - Keypoint alignment of the 2D central axial (top) coronal (middle) and sagittal (bottom) slice from patient 2 using both SIFT and ORB feature detectors.

Patient 6 represents another challenging case as there is a significant offset in multiple dimensions, causing a particular problem for alignment methods that operate over 2D slices, the results of which are shown in Figure 19. Over all three planes the SIFT based approach appears to result in a strong alignment when looking at the outline of the skin tissue, however there is a significant mismatch around many of the key features in the scan, particularly the parts of the pelvis visible in the coronal view, spine and ribs in the axial view and spine in the sagittal view. This is not a result of the method calculating and applying a poor alignment transform to the slice, but rather the features present in each scan differing, which a purely 2D approach cannot correct for. The ORB based approach fails completely for both coronal and sagittal views, caused by a poorly matched set of feature points being used to compute the homography.

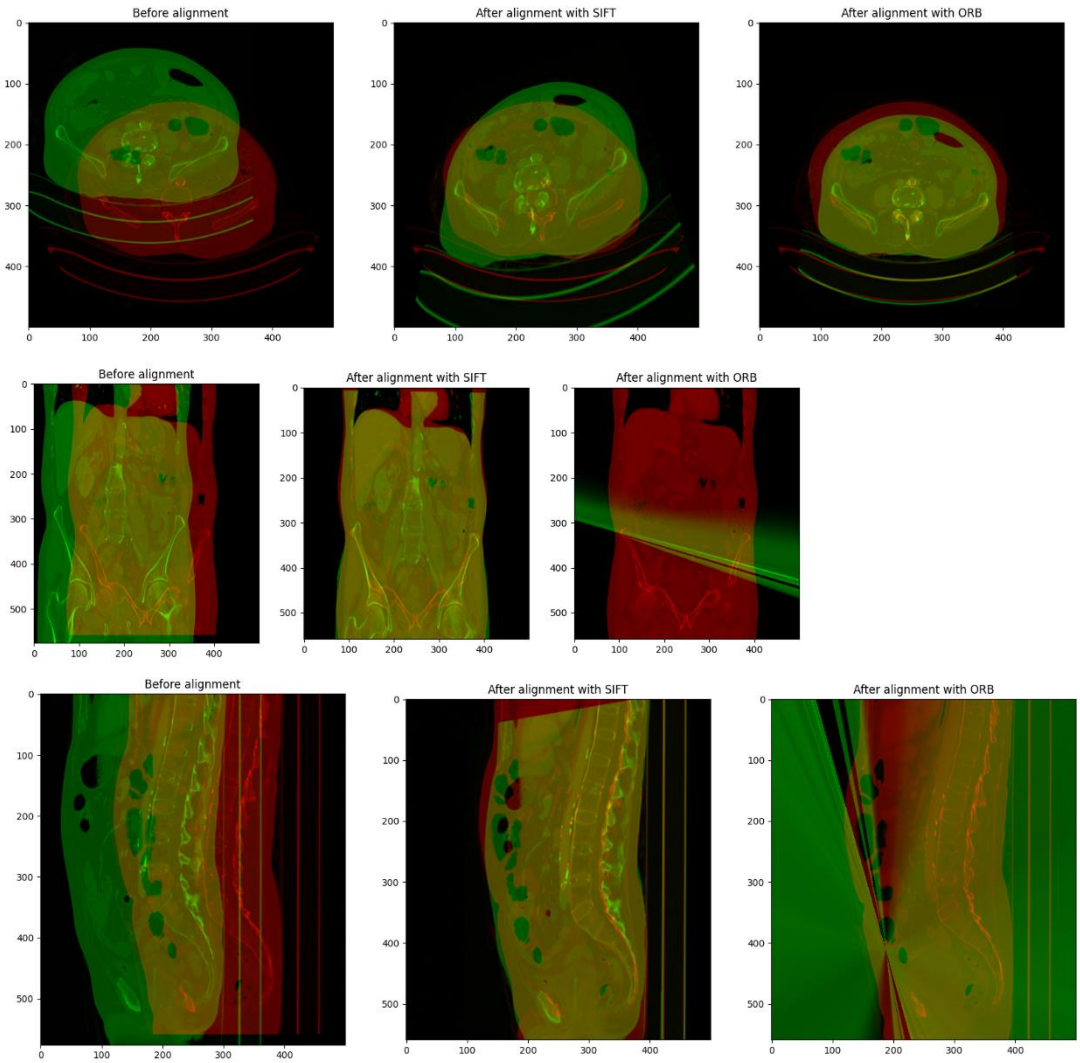


Figure 19 - Keypoint alignment of the 2D central axial (top) coronal (middle) and sagittal (bottom) slice from patient 6 using both SIFT and ORB feature detectors.

The keypoint based approach to alignment of 2D scans is comparatively fast, able to align many slices in a single second. As a result, it is feasible to compute and apply the alignment transform in real time while scrolling through slices of a scan.

3.3.3 Coherent Point Drift

CPD quickly showed promise on 2D slices. The results of the non-rigid algorithm aligning extracted point sets for a case with severe differences in transform between local regions is shown in Figure 20.

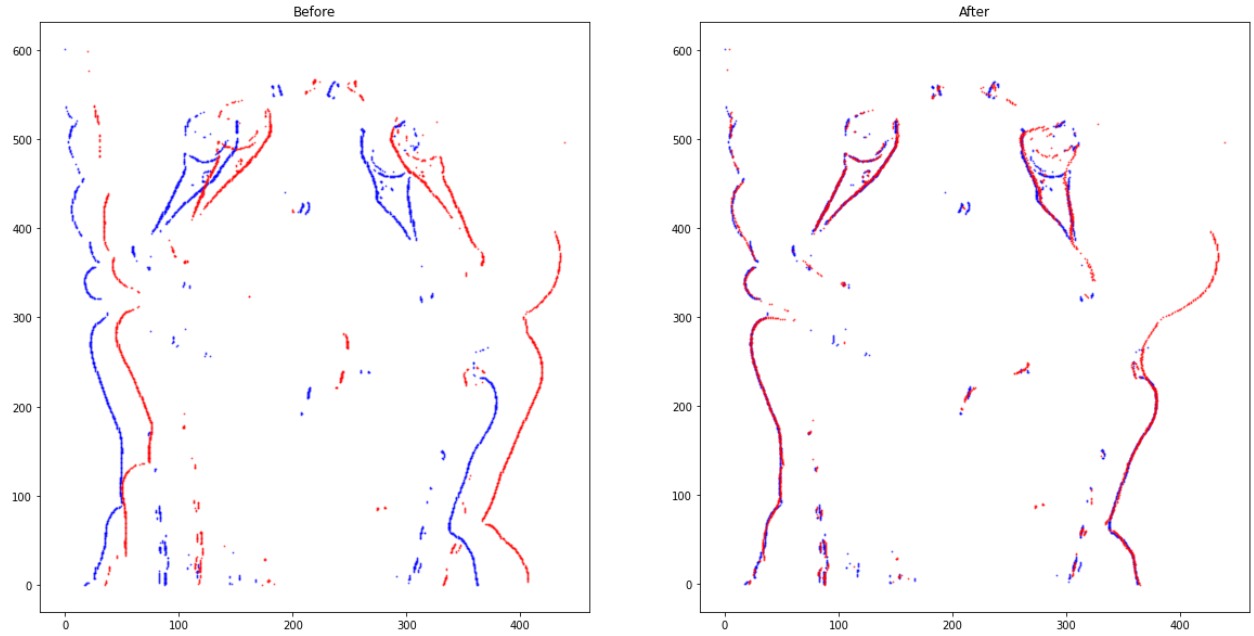


Figure 20 – CPD algorithm aligning source (red) and target (blue) points from a coronal slice, where points have been extracted via LMR and zero crossings.

After extraction of the point matching via polynomial fit, and application to full images from 2D slices, both the rigid and non-rigid CPD methods showed promise. Over the course of the project, Patient 1 was often used as an 'easy' case for alignment, in that little body shape change was observed and even a naïve overlay results in most structures being only a few pixels off of alignment, while Patient 2 was used as a 'hard' case for alignment – severe body shape changes have occurred between sequential scans, and some bones are oriented differently from each other between the two scans. In the 2D case for patient 1, rigid and non-rigid CPD methods achieve excellent global results, and for patient 2 the results are promising. These are shown in Figure 21.

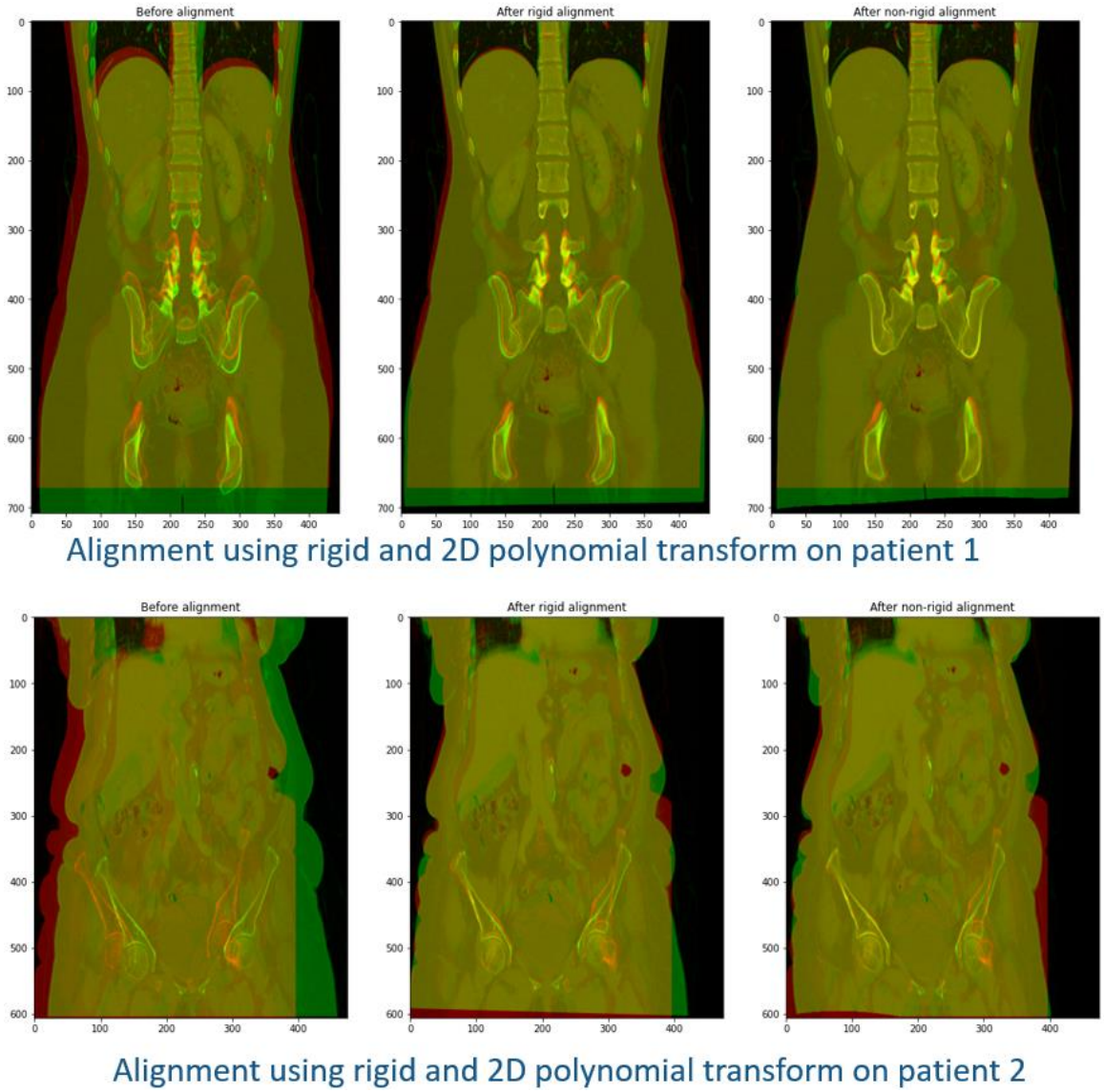


Figure 21 – CPD rigid and non-rigid methods, compared to no alignment for two patients.

Whilst this approach shows promise when applied to 2D slices from a scan, it is still constrained by the fact that it cannot account for misalignment in the third dimension. In order to overcome this, the same approach can be applied to the full 3D data, though the CPD alignment is a processing-intensive step that typically took around 1 hour to complete for the alignment of one scan pair, due to the number of points required to adequately cover the full 3D space ('adequate' here meaning that eventual full scan alignments began appearing 'good' by visual inspection). The run time was variable, as the CPD algorithm can exit early if a close match is reached before the maximum iterations are reached – scans which began highly misaligned typically took longer than ones which began close to alignment. Maxima of 50k target points and 5k source points were allowed in our use of the CPD algorithm. If, after point extraction, the number of points in each set exceeded these values, an integer n was calculated such that taking every nth point in a set would reduce the number of points below the thresholds, with n calculated independently for each point set. Typically the likelihood of a good alignment, as observed by eye, increased with both of these values, but they were limited to allow for reasonable run-time. The workstation used to run the CPD non-rigid alignment and then extract a polynomial fit was Ubuntu 18.04.5, 4 Intel(R) Xeon(R) CPU E5-2630 v4 @ 2.20GHz, 256GB of DDR4 RAM. The extracted point clouds for patient 1, before and after matching, are shown in Figure 22.

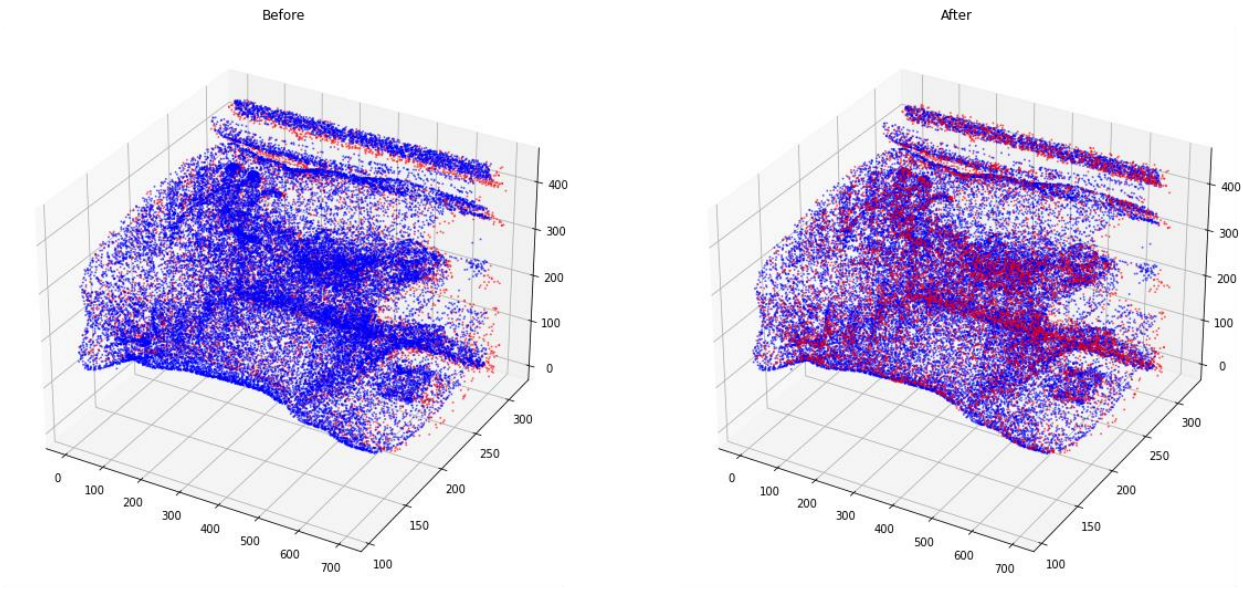


Figure 22 – Before and after CPD non-rigid alignment on extracted 3D point clouds from blue (scan 1) points and red (scan 2) points. In the 'before' image, major structures in the red set are offset behind the blue points. In the 'after' image, major structures in red and blue overlay well, showing that alignment has been achieved.

After running the polynomial fit, and applying to the full 3D data, 2D slices can be extracted for an improved qualitative view of results. These are shown for coronal, sagittal and axial views, before and after full 3D alignment, in Figure 23.

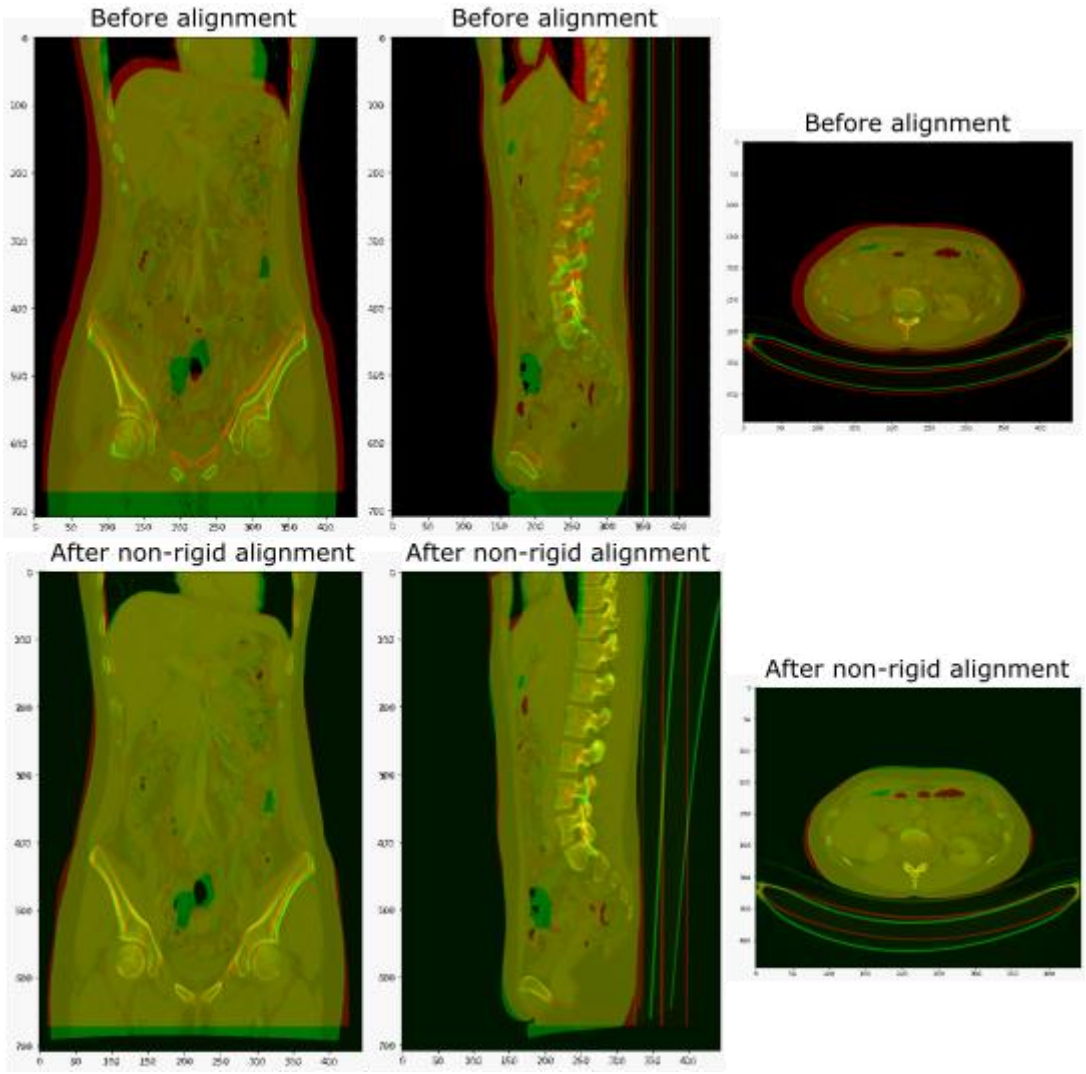


Figure 23 – Before and after full 3D non-rigid CPD-based alignment

Most features appear to overlay the same structures in the lower row of Figure 23. Some minor displacements of ribs remain, and the bright lines associated with the metallic table the patient rests upon (visible in sagittal and axial views) have been warped considerably. This warping of the table is a result of a pre-processing step to section the table out, meaning that it is not considered during the CPD algorithm and hence warps significantly after the polynomial fitting – this was not considered a problem, as the table lends no known diagnostic value. The result in the sagittal view is particularly interesting in this case, where in the 'before' case many vertebrae are not visible in the green layer. It appears that after alignment all vertebrae have been matched well.

While the results for this case are impressive, hyperparameters of the CPD algorithm required tuning for this particular patient, and took approximately 1h to complete. When applied to other patients with the same parameters, results were inconsistent – in many cases, introducing errors significantly greater than a naïve overlay. However, automatic tuning of these parameters to achieve a similar quality of results for all patients may be possible. In the phase correlation vein of work, for instance, results were quite inconsistent before the local mean removal, zero crossings, and knee detection steps were included in the pipeline, but after their introduction the phase correlation pipeline reached a point where there were almost no failure cases seen in the dataset.

4 TISSUE SECTIONING

4.1 MOTIVATION

Tissue sectioning is the process of automatically differentiating between different types of tissue (e.g. 'bone' and 'fat', or simply determining arbitrary classes, e.g. 'class 1', 'class 2'), and assigning a per-pixel label to all 2D and 3D pixels with which tissue type they belong to. Once sectioned, tissues which might originally appear similar in CT intensity values can be made to map to more distinct colouring, which can assist a radiologist with inspecting a scan directly, or feed into downstream analytics which can benefit from the sectioning.

In CT scans where no contrast agent is administered to the patient, the intensities at each volumetric pixel are so well calibrated that a radiologist can often identify tissue type via its Hounsfield unit (Hounsfield scale, n.d.). In our dataset, with a single type of contrast agent applied, the Hounsfield scale is not applicable, but similar tissue types do appear with near-uniform intensity values throughout the dataset. Further, every CT scan was delivered within a small band of intensity differences for the same tissue types. This is in contrast to ordinary computer vision problems, where illumination will vary from one photograph to the next, and texture-based approaches will be confounded by the scale at which the photographed object is recorded.

4.2 TECHNIQUES

4.2.1 TEXTONS

With the highly calibrated nature of the CT recordings, textons (Malik, 2001) form an attractive option for tissue sectioning. As noted by researchers in the literature, (Zhu, 2005) a single definition of 'the texton technique' is elusive, as a specific mathematical model has not been settled upon by the community. However, broadly, texton techniques aim to describe structures in images with reference to the local microstructure around each pixel, where the description can then 'bin' similar pixels to a classification.

In this method, a bank of different textures are convolved with each image, and the responses to these textures are clustered to automatically section differing tissues into separate classes. The nature of the clustering algorithm (here referred to as 'clusterers') may determine how many classes are discovered. Many require a pre-set number of classes to be aimed for, while others will contain procedures for automatically finding a number of distinct classes. Clearly, the latter is preferable when an unknown number of tissue types are present in the dataset, but Roke experimented with both forms of clusterer while evaluating textons for tissue sectioning.

4.2.2 DINO

In May 2021, Facebook released their 'Self-**di**stillation with **no** labels' (DINO) self-supervised model training method, along with open source tools for downloading models that had been pre-trained via the technique, as well as methods to train new models on fresh data sources (Caron, 2021). An example of the sectioning viable under the DINO method is shown in Figure 24.

Figure 24 – An example from the original DINO paper. (Caron, 2021) A vision transformer model trained via Facebook's DINO method learned to section out the pixels associated with 'monkey', into a single class label, without ever being told a monkey was present in the image, or that those particular pixels belonged to a single object.

DINO models whose backbone architecture are Vision Transformers will output a number of 'attention heads', where different objects within an image are expected to be bright in the output of each head. It was the hope that on this commission, a fresh model could be trained with the DINO method on a CT dataset, to separate out different tissue types.

While it is beyond the scope of this report to fully explain the underlying deep learning technique, an understanding of some deep learning concepts explained by (Caron, 2021) will be assumed when discussing the results of DINO for brevity. The open source library for training and applying DINO models required a prescribed data structure, rather than the CT data loading structure that was used for other backend methods and the GUI in this project. DINO datasets took the form of a directory of training images with no patient/body part/scan number separation. Two such datasets were generated, the first taken from the first 10,000 axial slices provided by GEH, and the second was a random 50,000 axial slices taken from across the dataset GEH provided. No performative difference was observed when training with the two different datasets.

4.3 RESULTS

4.3.1 TEXTONS

Gabor filters (Gabor filter, n.d.), a common choice of texture filter within Computer Vision, were the first 'kernel' type trialled for the texton exploration. Later, circular-symmetric exponential-radially decaying cosine filters were also trialled. In both cases, issues were encountered with the size of kernel versus the minimum size of desired detectable feature.

When a kernel is convolved with an image, features in that image around local regions that produce a strong response will be highlighted well, but there will also be an unavoidable blurring of the response as a function of the size of the kernel. The minimum desirable detectable size, following discussions with stakeholders, was on the order of 5 pixels wide, while kernels are typically larger than this to adequately carry texture information. In addition to obscuring small features, use of the kernels blurred the edge boundaries between different tissue types, leading to erroneous tissue classifications by clusterers.

5000 texton pixel-level samples were extracted using Gabor filtering, and various clusterers trained upon these, and the output of each inspected, both by Roke and in discussions with radiologists. When increasing to 10000 samples, no qualitative improvement was observed. An example image, labelled by some of the clusterers, is shown in Figure 25.

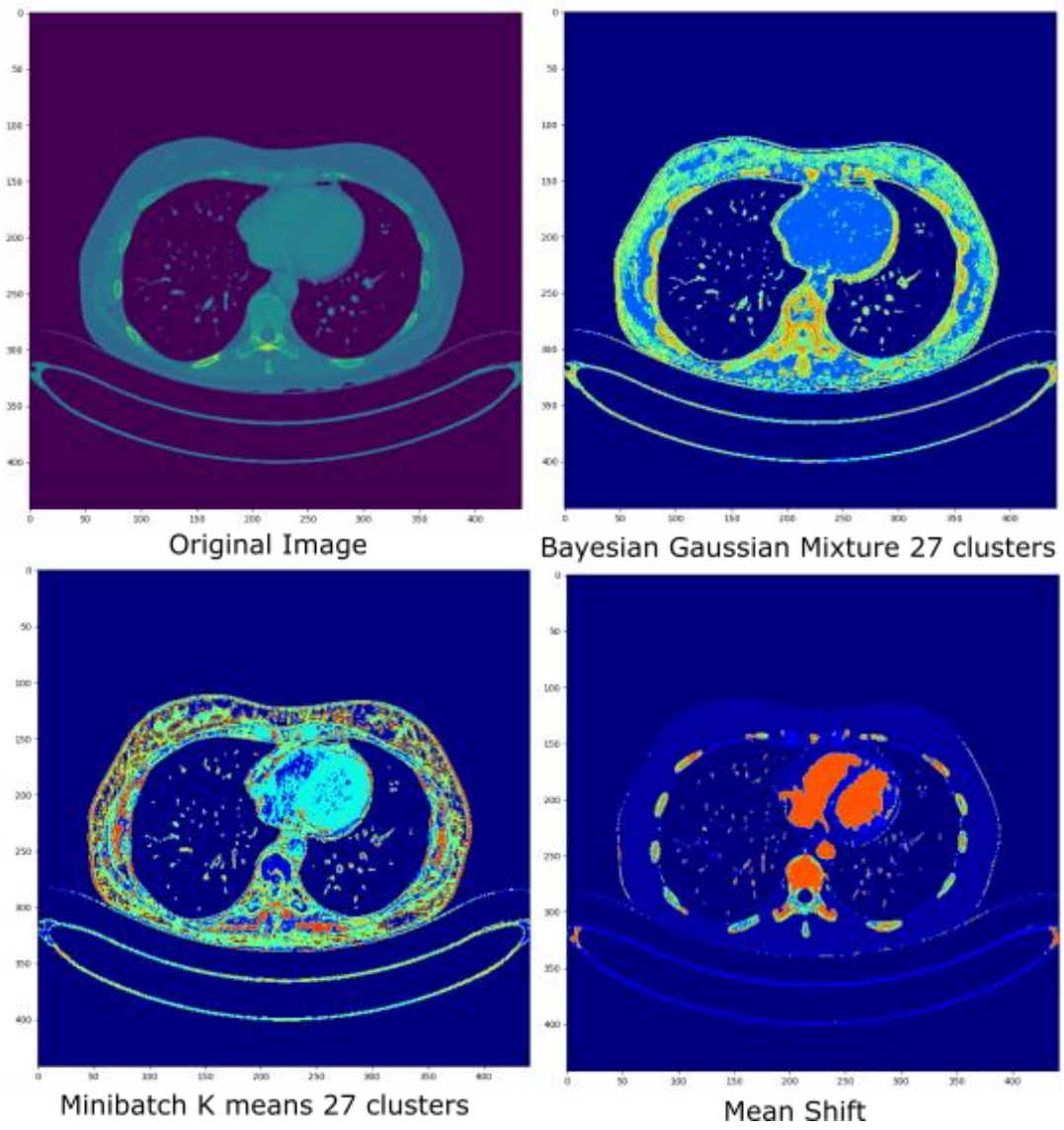


Figure 25 – Tissue class labelled images from some of the clusterers trialled on texton features using Gabor kernels. The colour maps are arbitrary, with a distinct colour only meaning that a distinct class has been found relative to other colours in each image. In the upper right and lower left images, the relevant clusterers were instructed to seek 27 separate tissue classifications, i.e. 27 separate clusters.

In the case of Bayesian Gaussian Mixture (BGM) and Minibatch K means, a user-defined number of classes must be entered. Various values were explored, incrementing from 3 to 27 in steps of 3, and outputs inspected visually by the Roke team. The compute time was a few seconds for the higher values, and exploring higher numbers of desired tissue types would not pose a computational challenge. Though increasing the number of sampled pixels for the dataset did lead to increased computation demands, no qualitative difference was observed when increasing the dataset from 5000 to 10000 sampled points, and sectioning quality was accepted by specialists from the 5000 set. An exact dependence of compute time on size of dataset was not carried out, but typically each clusterer trained in the low tens of seconds.

The investigation was limited below 27 types as the number of tissues listed in reference material that can be sectioned when uncontrasted Hounsfield units are used sent by GEH was on the order of 30. As the materials listed in the Hounsfield reference material included materials that may not be expected in most scans, e.g. ‘windowpane glass, limestone, tarmac’, it could not be expected that an exact match between those listed and those found automatically in our dataset would be found, and insisting upon a particular number may in fact lead to false separations of tissue type, or false conflations. Methods that automatically determined a

suitable number of clusters were preferred for this reason. Typically, when using Gabor-based textons, a lower target number of classes incurred less noise in the output, but this is of little benefit if it has only been achieved by conflating two tissue types which should in fact be distinct. The Mean Shift result appeared to produce the least noisy result, while at the same time not requiring a user-defined number of classes to begin clustering.

It is worth noting that each of these clusterers assumes fairly simple structuring in high dimensional space, and some clusterers, such as DBSCAN (DBSCAN, n.d.), often form a more robust separation of features when less simple structuring is present. However, predicting the value of new points outside the texton set the clusterer is trained on becomes more processing-intensive with many such clusterers, and as the rapid labelling of every pixel in an image was required, clusterers which were fast when labelling new points were preferred for this study.

While some tissue separation in Figure 25 is evident, few-pixel wide features that appear to be distinctly different intensity values in the original image are not separated well, and erroneous classes at the boundaries of different tissue types are introduced, as explained above. Clustering on pure intensity values, rather than applying any texture filtering, was also trialled, and appeared to lead to superior sectioning results, as shown in Figure 26.

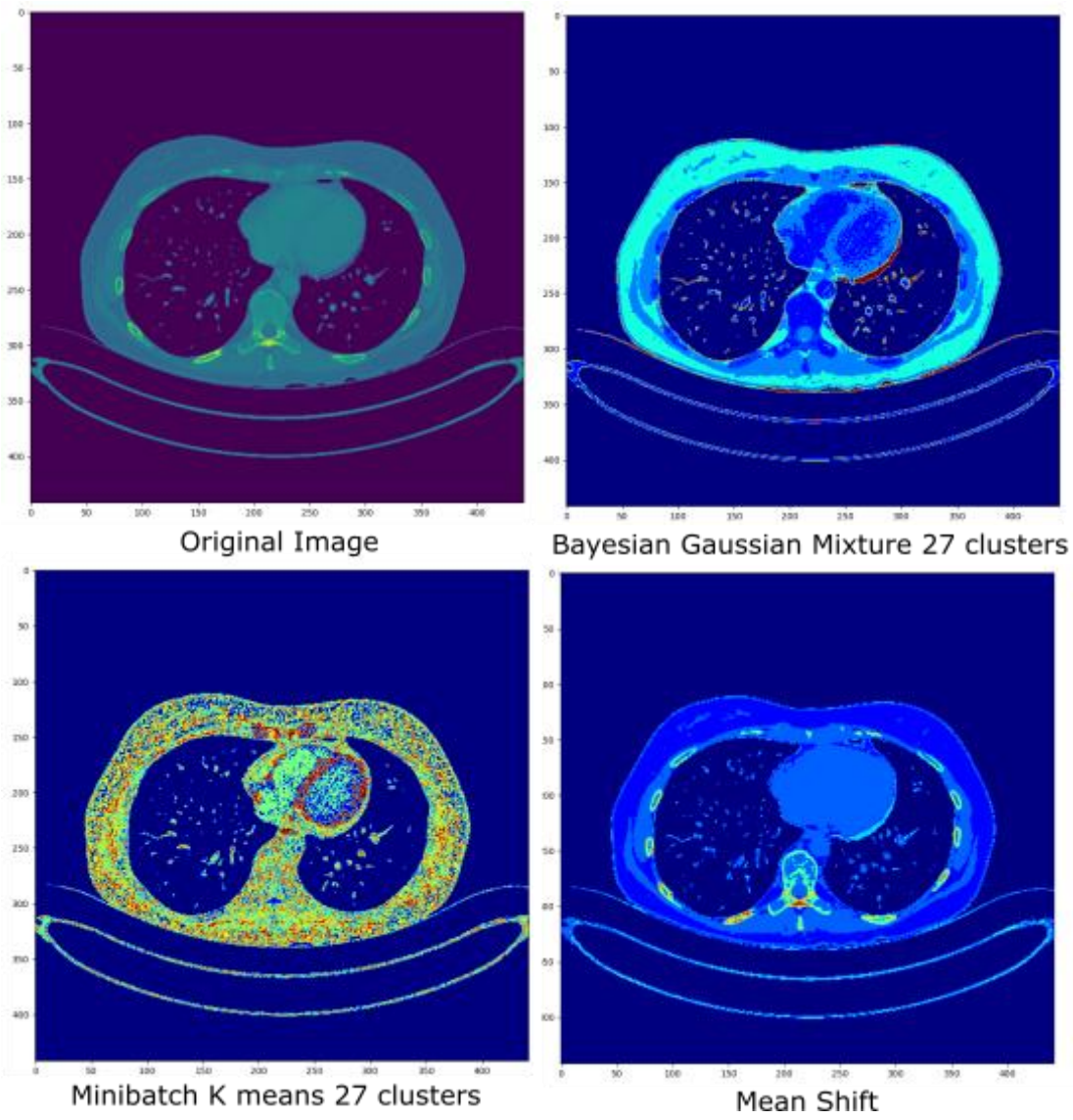


Figure 26 – Tissue class labelled images from some of the clusterers trialled on raw intensity values. The colour maps are arbitrary, with a distinct colour only meaning that a distinct class has been found relative to other colours in each image.

When presented to GEH specialists, both the Bayesian Gaussian Mixture models and Mean Shift clustering were considered to demonstrate low noise, and a robustness against tissue boundary errors when being trained on only raw intensity values, though ground truth sectioned tissues were not available to quantitatively measure this. In this case, the Mean Shift clusterer automatically found 11 classes, while the BGM clusterer required a target number of classes to find, in this case 27 (though increments of 3 were trialled and inspected as well). The BGM model appeared to be separating distinct types of soft tissue in the central mass of the example image, while the non-necessity of hard-coding a number of tissue classes for the Mean Shift clusterer meant that both options had attractive qualities.

The BGM clusterer had the additional option of predicting the *probability* of any pixel belonging to a particular class, as well as outright labelling it as the most probable class. This functionality was lacking in the Mean Shift classifier. During discussions with stakeholders, it was communicated that existing CT analysis software often has preset 'windows' of Hounsfield ranges, which can help to highlight differences within particular tissue types – thresholding values outside of these ranges to the minimum and maximum values, effectively performing contrast enhancement for particular tissues (designed for application when no contrast has been administered). This 'probability view' available with the BGM clusterer posed an interesting path to achieve a similar type of contrast enhancement for a particular tissue type. By sectioning to a particular tissue class, and then displaying the probability that each pixel within that class actually belonged to that class, a strong contrast enhancement method was expected.

An extension of the Mean Shift algorithm to provide a similar probability view was developed , with the probability being given by the exponential decay of the difference of a pixel's value from the centroid value of its assigned class, divided by the standard deviation of pixel values within that class in the training set. An example output when displaying a probability view for 'tissue class 1' is shown in Figure 27.

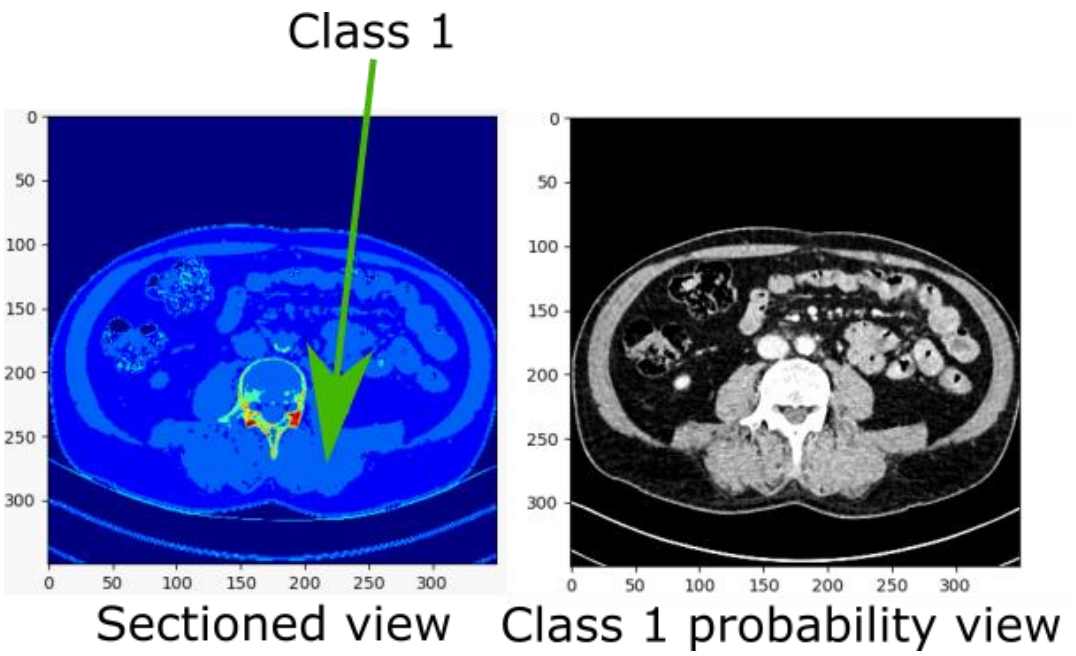


Figure 27 – Probability view of class 1 from the extended Mean Shift clustering algorithm.

While the probability view in Figure 27 clearly enhances the contrast for the related class (when comparing to the raw pixels in an unprocessed 'Original Image', Figure 26), it also highlights that there are sub-regions within class 1 that should be separable by intensity. The reason Mean Shift does not achieve this naturally is because it fits a single 'bandwidth' to find all classes, according to the statistics of the full dataset. The BGM method likely discovered these sub-types naturally, because each Gaussian within the method has its size fitted independently.

To emulate some degree of the BGM variable bandwidth functionality, while not requiring a set number of target classes at input, Roke devleoped a 'Hierarchical Mean Shift' (HMS) clusterer, where a new Mean Shift clusterer is trained on each subset of pixels, as separated by a first order clusterer. A scan then becomes separable by first order classes, or second order classes. 31 possible tissue classes were discovered by the HMS clusterer. A result of all second order classes found in an axial slice is shown in Figure 28.

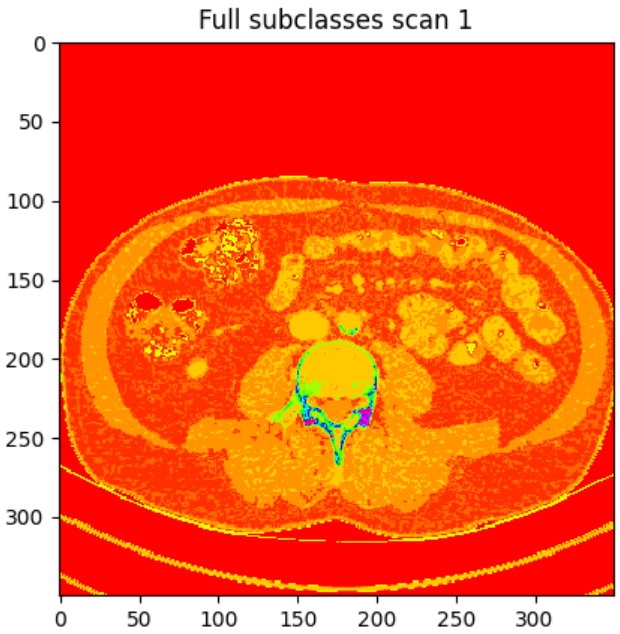


Figure 28 – Subclasses discovered with a HMS clusterer on raw intensity values

While some degree of noise appears to have been reintroduced by the HMS clusterer, GEH experts advised that this fine-grain level of tissue differentiation is likely legitimate. At this point, Roke considered the tissue sectioning adequate, in that the number of tissue types were automatically discoverable, fine grain tissue labelling was achieved, and labelling was rapid enough for real-time usage in the GUI. Moving forward, this HMS clusterer was the tissue sectioner used in the GUI.

4.3.2 DINO

As an initial baseline, the performance of a pre-trained DINO model made public by Facebook was run against axial slices. Poor performance was expected, as DINO models were trained on the ImageNet (ImageNet, n.d.) dataset, which contains no CT imagery, and therefore no separation of tissues based on vision transformer (vit) attention masks was likely. The result of running a pre-trained vit-small backbone on an axial slice from the dataset is shown in Figure 29.

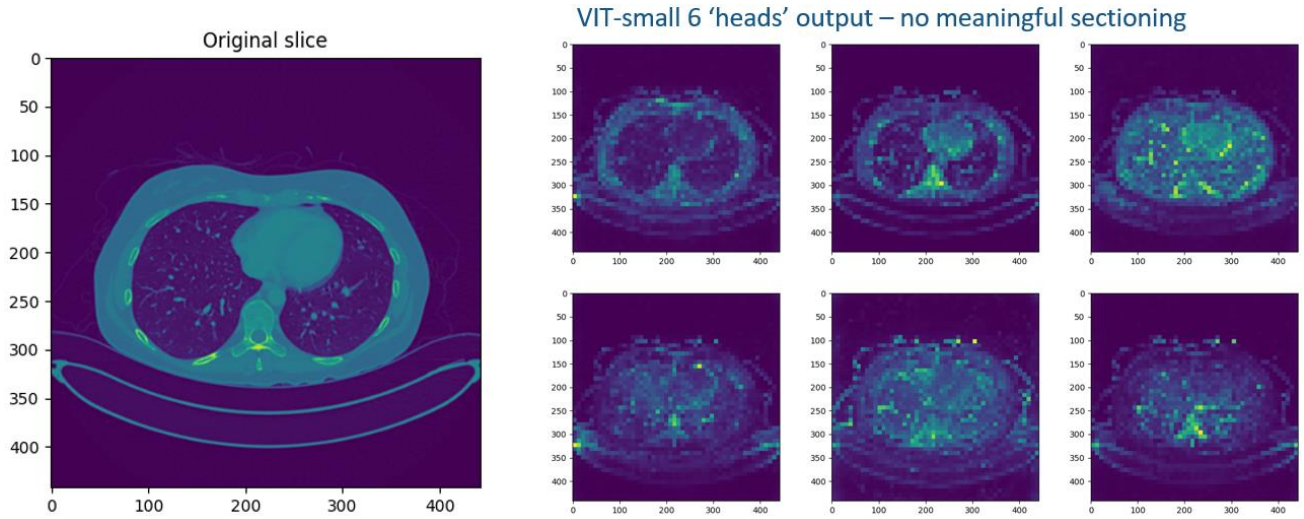


Figure 29 – The output of a DINO pre-trained vit-small network with 6 attention masks, against an axial slice from the CT dataset. No clear segmentation appears to be achieved.

Roke then trained a fresh instance of a vit-small backbone via the DINO tools, with a set of hyperparameters contained in ai_ct_scans/experiments/dino/training_args/vit_small.txt.

An important limitation of the DINO method encountered at this point was the downsampling of input images into 'patches' – when using the vision transformer model backbones (which enable the image segmentation in the original DINO paper), the transformer layers learn to associate square patches of pixels, rather than individual pixels, to deal with memory constraints of the underlying transformer models. Though it would be programmatically possible to define a patch of size 1x1, such that the attention heads return a full pixel-level labelling of input CT slices, such a model was not achievable on the available hardware (and indeed was not attempted by the DINO authors, who had access to greater compute facilities – the default patch size in the library is 16x16 pixels). The attention masks, which form the majority of memory usage in transformer models, scale quadratically with patch size, and with other hyperparameters optimised to allow for smaller patches a vit_tiny model was trainable on a single 8GB GPU, and a total of 200GB of GPU memory then might be expected to be adequate for such an investigation. Though large, this value is within realistic achievable compute capability – the DINO codebase comes with options for distributing the model across multiple GPUs, and GPUs are now available with 80GB on a single card (Newsroom, n.d.).

We trialled 5x5 patches with the vit_tiny backbone, 8x8 patches with the vit_small backbone, and 12x12 pixel patches with the vit_base backbone – balancing memory requirements on batch sizes, patch sizes, and absolute model size. Smaller patches implied higher memory usage, and this was typically prioritised over large batch sizes. Results after training for 100 epochs on the 10,000 image dataset are shown in Figure 30.

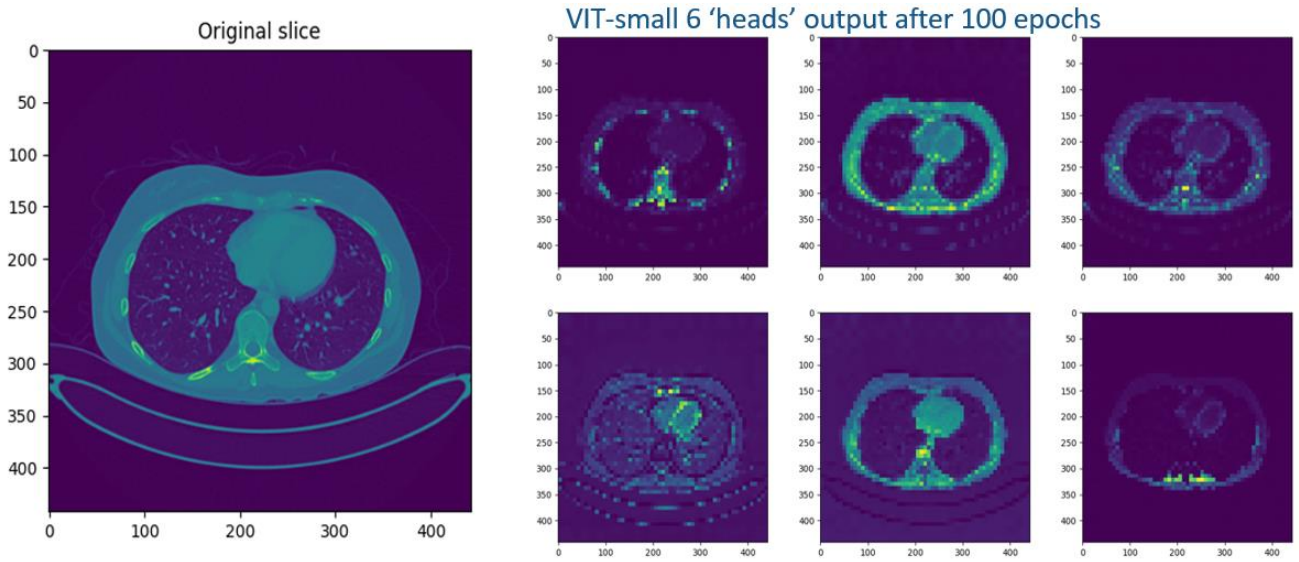


Figure 30 – The output of a vit-small network with 8x8 patches trained against the CT dataset, against an axial slice. The top left attention mask output appears to have found 'bone', while other masks may be focusing on soft tissues.

Some segmentation of different tissue types appears to have occurred with the vit-small Roke-trained model, particularly points associated with ribs and the spine appear to have been separated in the top-left attention head. The behaviour of focus of the other heads is less clear, but they appear to be focusing on soft tissues.

Figure 31 shows a similar investigation using a vit-tiny backbone, trained for 34 epochs (at which point loss had ceased descending) on the 50,000 CT image dataset. Only three attention heads were available from the vit-tiny architecture, but a reduced patch size of 5x5 pixels was possible given available GPU memory (8GB). Some tissue separation was observed, but not to the same degree as the vit-small or vit-base models.

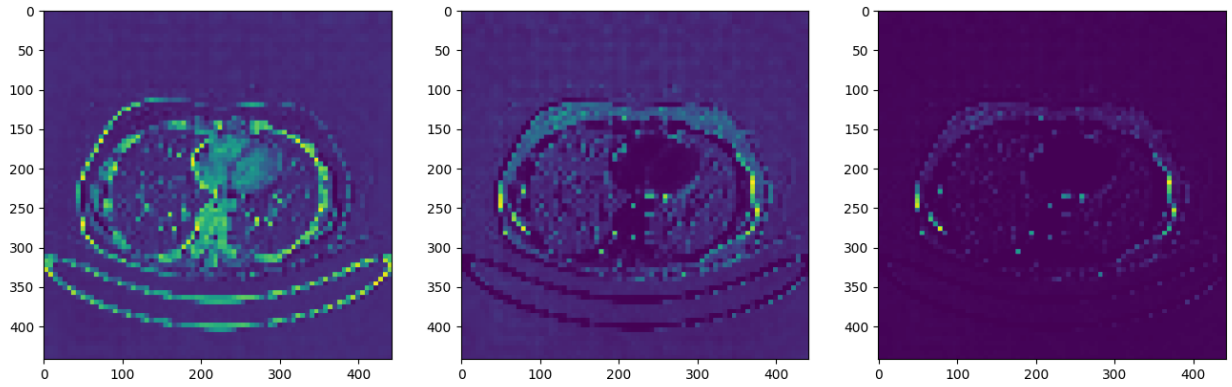


Figure 31 – An output of the vit-tiny network with 5x5 patches trained against an axial slice. Only three attention masks are returned by the vit-tiny backbone, but higher resolution is possible with the smaller model size allowing smaller patches.

The larger vit-base backbone was trained against the 50,000 image CT dataset, for 100 epochs, with patches of size 12x12. The results of the 12 attention heads available from the vit-base architecture are shown in Figure 32.

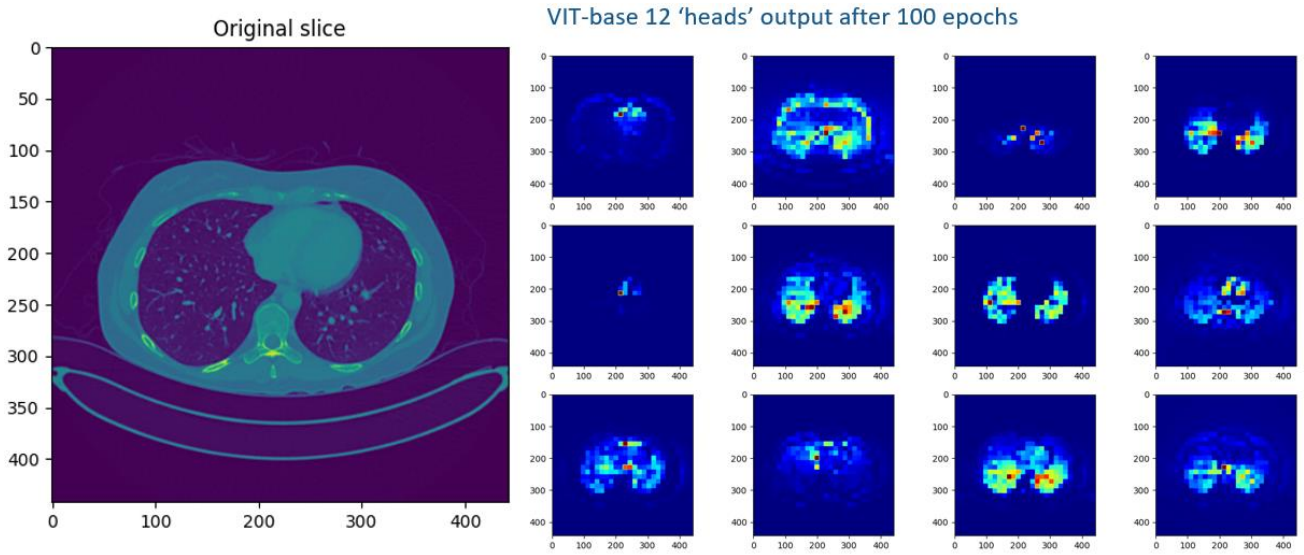


Figure 32 – The output of a vit-base network against an axial slice with 12x12 patches trained against the CT dataset. Though each head is clearly focusing on different tissues, it is unclear whether these segmentations are oncologically relevant.

While the vit-base model appears to be focusing on different tissue types in its 12 attention mask output, it is unclear whether the separation corresponds to a useful separation for radiologists. The limitation of 12x12 patches means that there is severely limited resolution of labelling.

In the original DINO paper, images are sectioned based on the percentage of total intensity contained in pixels in a particular mask, typically 60%. While a valid first pass to demonstrate that segmentation can be performed via attention head masking, the method may be prone to failure when a wide range of intensity is predicted by a vit model on a particular mask, e.g. the lower left attention head output of Figure 30. With the clustering pipeline already constructed for the texton approach, it was a point of interest to use the attention head outputs for clustering in the dimensionality of the number of heads of the given model, potentially forming a more robust method of sectioning images based on DINO models. The result of such a clustering is shown in Figure 33.

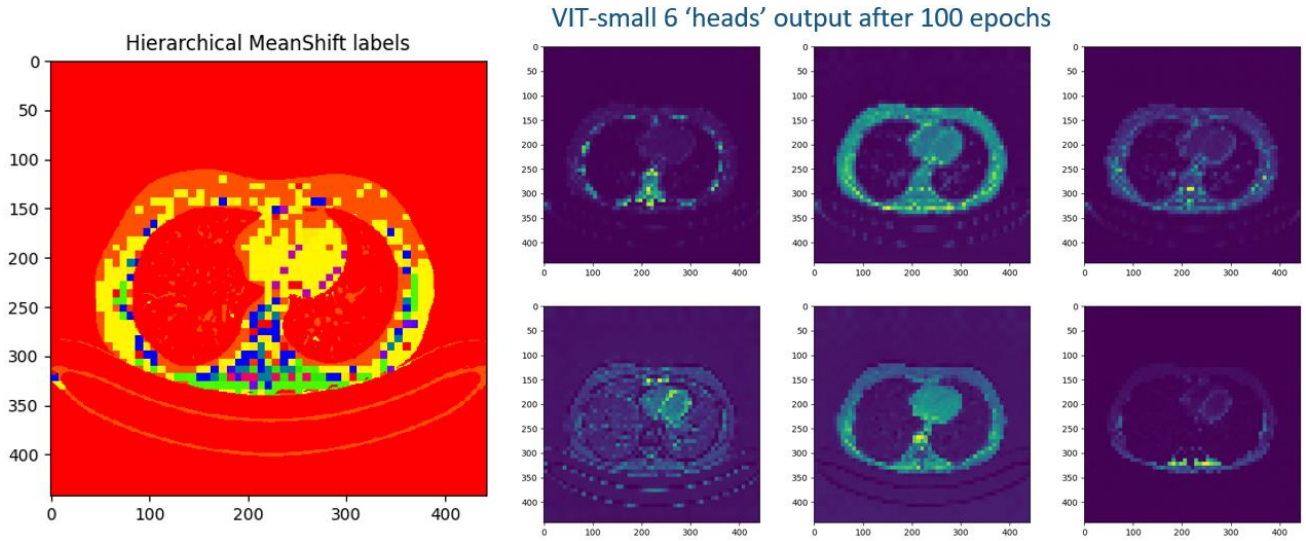


Figure 33 – HMS clustering performed on the DINO vit-small attention head outputs, with any pixels below an intensity of 500 masked to zero, as these were assumed to be associated with air.

Some sectioning related to the structure observable in the original image in Figure 32 is visible, though the limited patch size resolution leads to poor results compared to the texton clustering results. It is important to note that the fine structure (lung tissue, outline of the interior and exterior of the lungs) in the HMS labels in Figure 33 is due to thresholding out pixels whose original intensity was below 500, which was assumed to be due to air, and the actual resolution directly available from the DINO model would remain 8x8 patches.

All models took considerable compute-time to train, between 1.5 days and 8 days, on a GeForce RTX 2080 Super graphics card, though training loss appeared to have stabilised some time prior to this in each case. While all models achieved some level of tissue segmentation, the unsupervised nature of the DINO technique means that it was not possible to guide the model towards a 'good' separation. There is an extremely high level of difficulty in assuring and explaining the behaviour of deep learning techniques, and vision transformer models are no exception – with these various limitations in mind, and the relative success of the texton-based approach, little value was seen in progressing with DINO-based methods for this project.

5 ANOMALY DETECTION

5.1 MOTIVATION

The alignment and tissue sectioning methods presented so far can aid a radiologist in their manual comparisons of CT scans, but inspecting the 3D data for lesions will still be a fairly lengthy process of scrolling through 2D slices of the data. Having computer visions and/or AI methods for directing the radiologist's attention to likely points of interest in a scan may accelerate the rate at which CT inspections can be performed.

5.2 TECHNIQUES

5.2.1 ELLIPSOID DETECTION

During a walkthrough of CT inspections with GEH radiologists, a manual method for identifying lesions from other internal body structures was demonstrated. If a round region of intensity that differs from its surroundings is spotted in a 2D slice, and is not clearly associated with a well-known structure at that point in the body, it draws the attention of the radiologist as potentially anomalous. By scanning in the third dimension around the region, it can be inspected for branching structure, which would indicate it to be a blood vessel and of low interest. This process is repeated in axial, coronal, and sagittal views, inspecting every slice manually and scrolling through in depth, seeking anomalous rounded regions in this way. This description seemed to suggest that true lesions most often presented as 3D ellipsoids, i.e. rounded, and without branching structure. Later, 'spikey' lesions that would not fit an ellipsoidal descriptions were understood to be present in the dataset, but a high fraction of lesions being ellipsoidal was taken as a valid assumption to investigate it as a detection technique.

Roke suggested that triggering any 3D ellipsoid as points of interest may be a valid technique for drawing radiologists' attention, and developed an 'ellipsoid detector' to this end.

Open source computer vision methods from the OpenCV (Bradski, OpenCV, n.d.)allowed for the rapid finding of contours (Bradski, Finding contours in your image, n.d.) in 2D slices around edges, given binary 2D slices. In order to convert from raw CT slices to binary versions that would showed relevant binary edges, the texton-based HMS tissue sectioner was used. Before detecting contours, each image was masked to each distinct tissue class found therein. Checks on whether a contour formed a closed shape, and whether a convex hull (Shrimali, n.d.)drawn around the contour fulfilled minimum and maximum area requirements, eliminated low-interest contours. Ellipses could then be fitted to the remaining contours, and further checks on ellipticity, ellipse axis length limits, and the area enclosed by the ellipse compared to that enclosed by the original contour, then eliminated further low-interest ellipses. Finally, in each 2D slice, a set of ellipses with geometric information for each was returned. The result of such a detection on a 2D axial slice is shown in Figure 34.

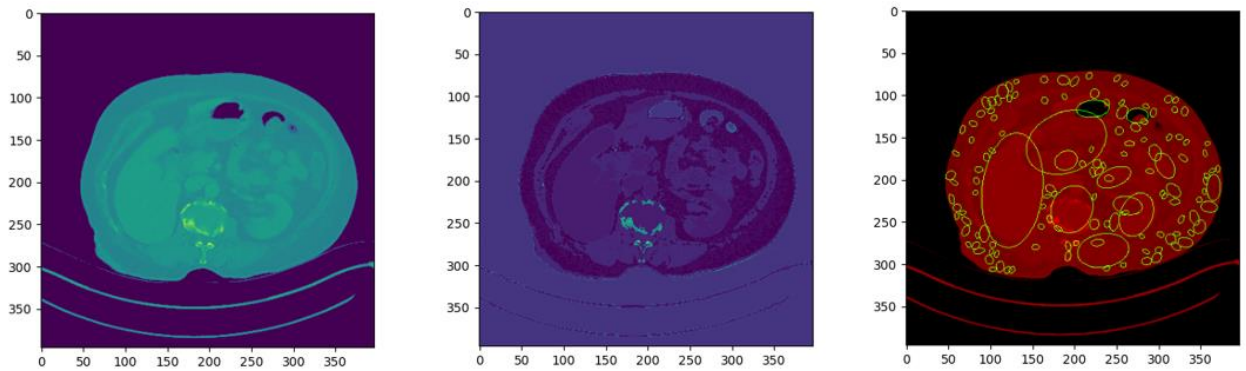


Figure 34 – 2D ellipses detected in a single axial slice. (Left) the original image, (middle) the sectioned image, (right) found ellipses drawn on the image. A high number of non-lesion ellipses are present in most axial views, which motivated moving to full 3D ellipsoidal triggering.

This was repeated for each slice in each axis, returning a large set of valid ellipses from within a full 3D scan. In order to filter down to 3D ellipsoids, the DBSCAN clustering algorithm was run separately on each set of ellipse centres belonging to each tissue class, with a separation epsilon of 2.0, meaning that any ellipse centres of the same tissue class within a distance of 2 pixels of another ellipse centre of that class, as discovered in any of the three axes, would be associated together as a potential ellipsoid. Checks on the ellipsoid being supported by ellipses from more than one axis, or growing across multiple ellipses in a single axis to eliminate false positives due to non-ellipsoidal clusterings of ellipse centres, then added a final screening before an ellipsoid was accepted to be triggered upon.

Each ellipsoid was then returned with information which may be of use to a radiologist, including its centre location in 3D, a 3D bounding box, a volumetric measurement (taken by the sum of 2D ellipse areas in the axis with most elliptical areas enclosed throughout for that ellipsoid), the number of ellipses found in each axis that supported the triggering on the ellipsoid, and the tissue class of the ellipsoid as given by the texton tissue sectioner.

5.2.2 Masked Data Deep Learning

Before the ellipsoid detection was conceived during the course of the project, an approach for flagging the presence of *any* anomaly in a CT scan was desired. As human-level object recognition was expected to be required for this task in the case of lesion detection, where the internal body structure presents a high degree of variation between even healthy patients, a DL approach was deemed a useful avenue for experimentation. Additionally, as the dataset contained no labelled examples of lesions at the start of the project, it was an *unsupervised learning* problem. Any method applied would have to be capable of predicting the presence of anomalies with no human guidance to learn from.

Many deep learning approaches to anomaly detection have been developed in the literature in recent years, (Pang, 2020) and in the case of imagery datasets Roke has previously found masked data prediction methods to be effective.

The steps of a masked data prediction training objective for the CT dataset are as follows:

1. Take a random 256x256 cropped view of any axial, coronal or sagittal slice from within the dataset
2. Blank out a central region
3. Provide edge information in the masked region
4. Have the model attempt to fill in the missing pixel information

Even though the dataset was oncological, and therefore contained more examples of lesions than would be expected in a healthy population, the majority of tissue contained in the dataset was still healthy (only a small region of each patient with a lesion should be truly anomalous). Usually then, it was reasoned that a DL model trained on this objective would begin filling in masked regions with healthy tissue. If a lesion were masked out, and then filled in with healthy tissue instead, subtracting the input from the output should then highlight a found anomaly.

Step 3 in the above objective was added after initial experiments showed poor performance – where Roke had previously experimented with the method on highly uniform industrial data, the natural variation within a human body meant that the masked data prediction performance could not predict the location of non-anomalous structures. Providing edge information of structures within the masked region then would allow the model to fill in realistic, healthy tissue, while hopefully still failing to predict when an outline originally contained a lesion. Figure 35 shows the approach.

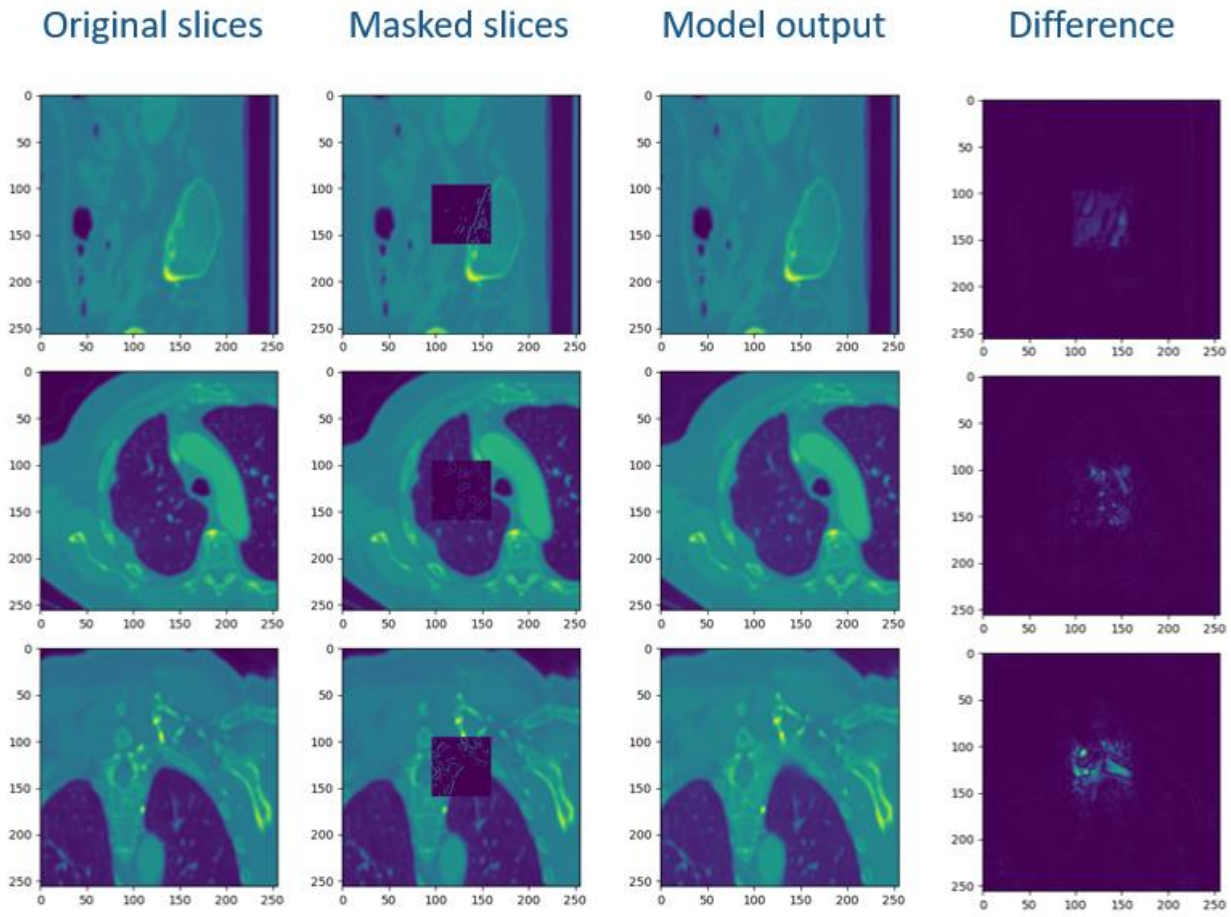


Figure 35 – Results of masked data prediction on a random set of three slices.

The model has achieved realistic infilling of the masked regions, although there is a slight visible blur in the model output in the third row. Note that the colour range in the 'Difference' column is scaled according to the maximum of the difference found – i.e. even though the model's predictions are highly realistic, the small differences within the masked region still show up clearly for the purposes of Figure 35. Unfortunately, determining a valuable way to leverage this difference information to rank the triggered anomalies remained an open task at the end of the project – simple measures, such as the max, mean, or sum of the differences found within the masked region would all lead to misleading measures. Consider a very large lesion, with only subtle intensity difference from the surrounding tissue – though a sum of the differences may rank such a case highly, it would then risk missing the inverse case, a small, starkly different lesion.

Results when the masked data prediction were run against actual lesions, which were labelled later in the project, are given in section 5.3.2.

5.3 RESULTS

5.3.1 ELLIPSOID DETECTION

During backend development, blank volumes of the same scale as input scans had the outlines of ellipses found in each 2D slice drawn, such that the volume could be inspected in 3D for the presence of dense regions where an ellipse outline existed at a volumetric pixel location from more than one axis of observation. A result of this type of investigation is shown for patient 2, who had a relatively clear abdominal lesion, identified to the Roke team by GEH radiologists, in Figure 36.

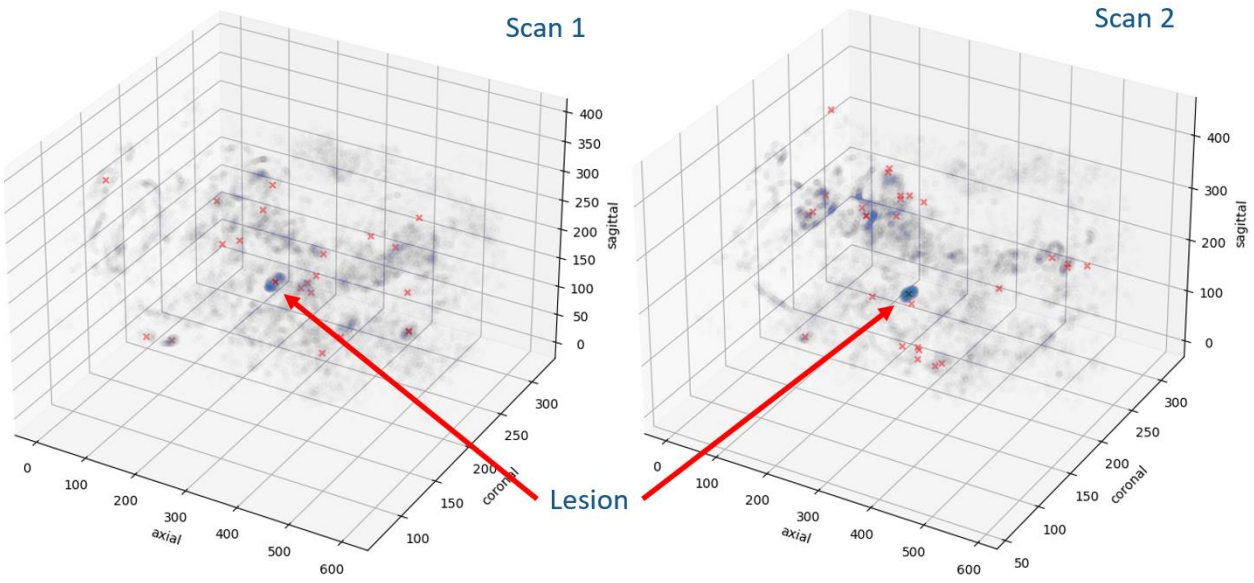


Figure 36 – Ellipsoid centres (red crosses) and 2D ellipse perimeter points supported in more than one axis (blue dots). The expected lesion location showed up with a dense cluster of ellipse perimeter points, and was identified as an ellipsoid.

Around 30 ellipsoids were detected in each scan, one of which corresponded to the true lesion. Although there is clearly a higher density of perimeter points drawn around the true lesion than other ellipsoids, it was unknown whether a metric based on this observation would be robust for thresholding out false detections in wider example lesion cases, and additional elimination of false positives was not undertaken.

While the result in Figure 36 proved the ellipsoid detector could find some lesions, parameters of the ellipsoid detector were tuned against this particular example. A larger set of known lesion locations was sought to validate the approach. A set of 9 patients, often with multiple lesions across different body parts and sequential scans, with 'by eye' lesion centres recorded, was delivered by GEH radiologists for this validation. The results of this validation step will be given in section 5.3.2, where the ellipsoid detector has been used to 'target' the masked data deep learning approach.

While the results were qualitatively validated against GEH radiologist's judgement of whether lesions had been found or not by the two techniques, it is worth stressing that only one set of detection parameters for the ellipsoid detector were used for that stage of validation. The qualitative validation process required a lengthy inspection of results, and time restraints prevented the team from measuring performance against alternate ellipsoid detection parameters. However, when the ellipsoid parameters were 'relaxed', to allow smaller contributing ellipses and using a median 2D filter over tissue sectioning slices before inputting to ellipse detection, the true positive rate appeared to increase (as well as an increase of the false positive rate).

Although these relaxed parameters led to identification of ellipsoids typically numbering in the low hundreds per scan, such a set still may be faster to check through than the current, highly manual inspection of every region of every slice by radiologists. A high false positive rate accompanying a high true positive rate may also be acceptable in an oncological setting, where maximising the true positive rate is of greatest importance.

5.3.2 Masked Data Deep Learning

Initially, it was hoped that the masked data technique would be able to scan across full 3D data for each scan, patching together the centrally predicted output regions and then their differences from the true data, in order to create a one-step solution to finding lesions in CT data. Figure 37 shows the result of such an investigation on patient 1.

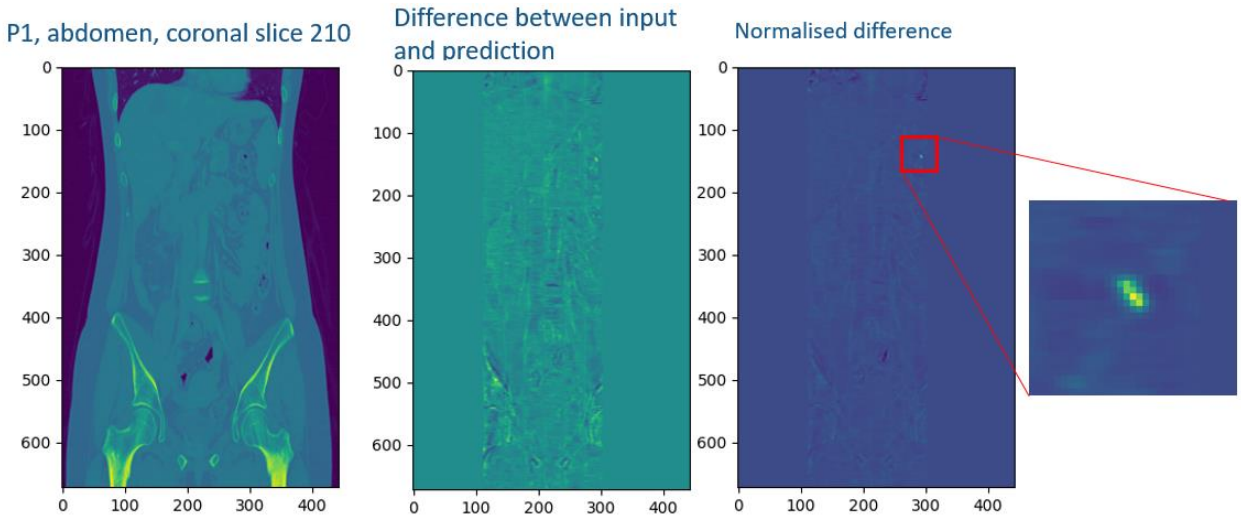


Figure 37 – The masked data model was run against the central column 3D data in axial slices, and stitched together to form predictions in other views. After a normalisation step, it appeared that an anomaly was being clearly found, but this was actually due to a pocket of air in the patient's body, leading to the development of other techniques to 'target' the masked data method.

The method used to produce the columnar results for Figure 37 were later modified to allow input data that overlapped the edges of a scan, such that a masked data model could make predictions about the full volume of the scan. However, it was already becoming clear that further processing steps were necessary to target the infill method at regions of interest. Although at first glance a clear anomaly is found in patient 1 via the masked data method (after a renormalisation step of reweighting the masked data predictions by 1/(0.2+abs(original pixel value – mean of original pixel values)), this anomaly actually corresponds to a small pocket of air in the patient's body, which was not of oncological interest. This development highlighted that all components of the PoC developed in this commission form a *decision support tool*, and cannot replace the expertise of trained radiologists.

While a pocket of air is 'anomalous' in a human body, as one might not expect to find such a pocket at any given location, this highlighted the problem that too many non-lesion anomalies would be flagged when relying on masked data prediction alone, and uninteresting cases would likely be highlighted more strongly than lesions. Consider the case of a lesion with subtly different intensity values to surrounding tissue, versus a pocket of air within the same tissue – a masked data prediction model that is expected to predict neither lesion nor air will usually introduce a larger difference from the original data for the air, simply because the tissue of lesions is qualitatively closer to the surrounding tissue than air is ever likely to be. This motivated bringing the tissue sectioning work into an earlier sprint, than originally planned. Anomalies, eventually flagged in the GUI tool, had tissue classes associated and displayed to help discriminate between found points of interest.

It was for this reason that the ellipsoid detection was added as a pre-processing step prior to the masked data prediction model. The masked data prediction could then provide a second layer of analysis on discovered ellipsoid centres, automatically slicing to axial, coronal and sagittal views around each. The combination of ellipsoid detection, and the masked data model anomaly visualisation for each ellipsoid centre, were presented to GEH specialists to form the qualitative validation of the tools. As a precursor to each detected ellipsoid, the performance of the infill technique was shown on the known lesion location for comparison, and false positive cases were also shown. Example such images are shown in Figure 38.

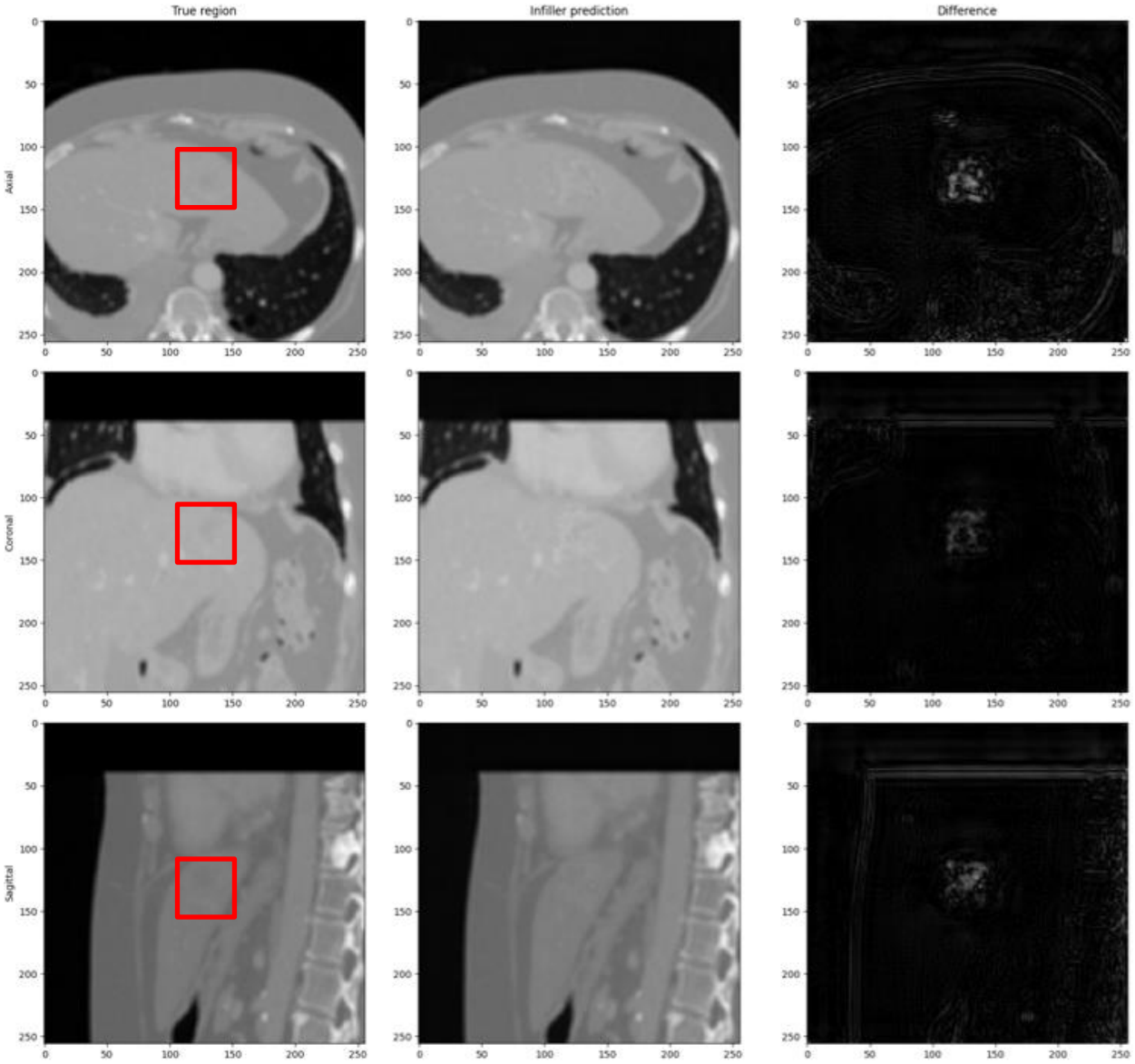


Figure 38 – Masked data prediction on a known lesion in patient 17, abdominal scan 1. The lesion is in the central 32x32 pixels of each image in the first column, highlighted in the figure by a red square. The second column shows the model's predicted, infilled versions. The third column shows the difference between input and output.

In Figure 38, the lesion is a fairly subtle shadow against the surrounding tissue in the centre of each slice shown in the first column, and the infiller appears to have filled in with healthy tissue, which leads to a highlighting in the third column. Note that the colour map in the third column is scaled to the strength of the difference – a clear way to extract a uniform colour map to make comparison between scans was not found during the course of the project, but would be a useful piece of future work.

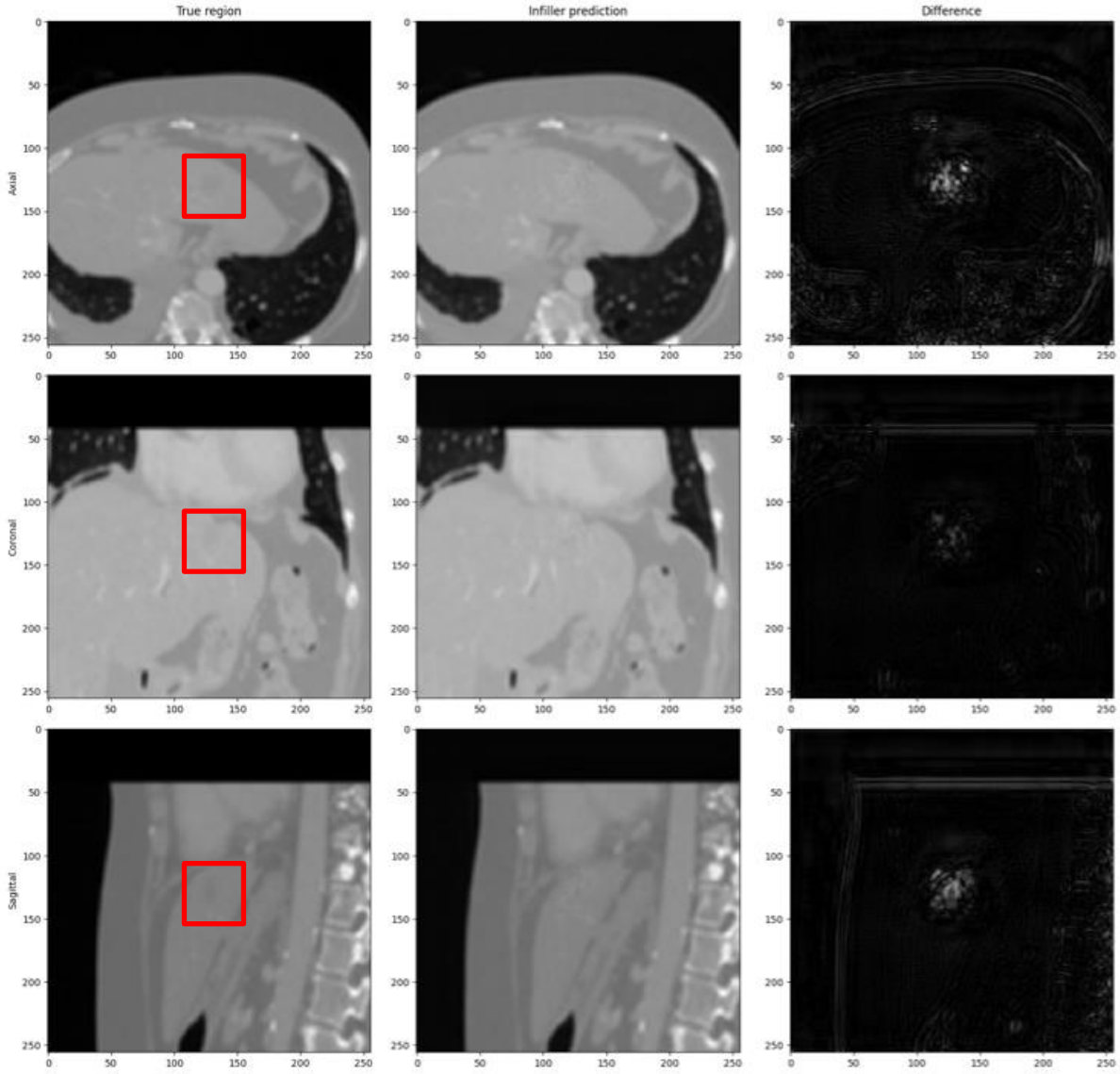


Figure 39 – The masked infill method applied to a detected ellipsoid, which coincided well with the known lesion location on patient 17, abdominal scan 1.

In Figure 39, an ellipsoid at almost the exact expected location has been found, and triggered a masked data prediction on the correct region. As for the known location case, the lesion appears to have been filled in with healthy tissue in the second column, leading to a highlighting via the difference in the third column.

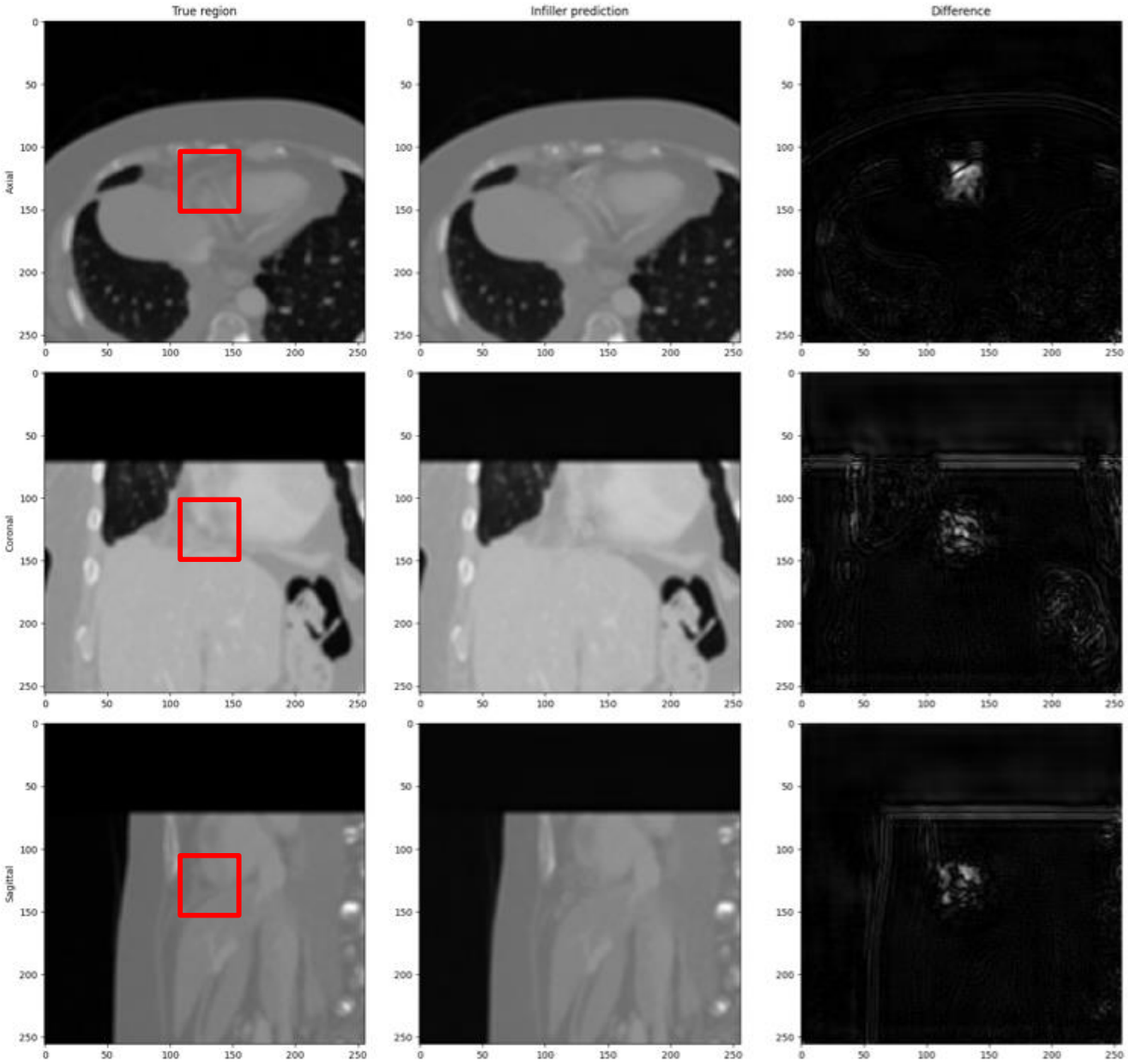


Figure 40 – Performance of the masked infill method on a false positive in patient 17, abdominal scan 1. Although it appears realistic features have been infilled, subtle differences have been automatically scaled in the difference visualisations in the third column such that they appear starkly apparent.

In Figure 40, a remaining issue with the infill prediction on false positives is shown. Even though the infiller has appeared to produce realistic tissues in the centrally masked region, minor variations from the true tissue appearance have been amplified by an automatic colour map scaling in the difference image. This relates back to the issues with using this difference image for a robust metric mentioned at the end of section 5.2.2. During the validation call with GEH radiologists, the colour map was originally scaled to a maximum of the 99th percentile of the intensity in the original image, but this led to the difference images being too dark. Further experimentation may benefit the difference visualisation, and allow a single-number metric to be extracted for the prioritisation of found ellipsoids for radiologists' attention. For now, it serves as a useful indicator of which features within the central region of the displayed slices should be given closer analysis by a radiologist.

For the nine patients examined on a validation call with GEH radiologists, the results of detection via ellipsoid detector and/or highlighting well via the masked data prediction method are shown in Table 2. While 100 patients were provided in the dataset, no pre-existing labelled locations of lesions existed. A validation set of 9 patients was generated by GEH specialists over the course of the project. Each patient inspection requires considerable effort, and so fully labelling the 100 patient dataset was not realistic in the scope of the project.

Table 2 – Result of ellipsoid detection/masked data highlighting against true lesions

Patient	Scan	Slice	Lesion	Detected	Ellipsoid Detection	Masked Data
11	Abdo2	60	Large liver lesion	Yes	Yes	Yes
11	Thorax2	317	Lung nodule	Yes	No	Yes
11	Abdo1	615	Bone lesion	Yes	No	Yes
12	Abdo1	100	Multiple liver lesions	Yes	Yes	No
12	Abdo2	220	Renal Cyst	No	No	No
13	Abdo1	135	Liver cyst	Yes	No	Yes
14	Abdo2	281	Fibroid uterus	Yes	Yes	Yes
15	Abdo1	322	Right ureteral lesion	Yes	No	Yes
16	Thorax1	259	Left hilar nodule	No	No	No
17	Abdo1	88	Liver lesion	Yes	Yes	Yes
17	Abdo1	207	liver lesion	Yes	Yes	Yes
18	abdo1	47	liver lesion	No	No	No
18	Abdo1	142	liver lesion	No	No	No
18	Abdo1	165	liver lesion	No	No	No
18	Abdo1	135	Pancreatic cyst	No	No	No
19	Thorax1	19	lymph node	Yes	Yes	Yes
19	Thorax1	130	thoracic mass	No	No	No

Table 3 – Result of whether the ellipsoid detector *or* masked data highlighting was found to be useful on a per-patient basis in the validation set

Patient	Lesion(s) Detected
11	Yes
12	Yes
13	Yes
14	Yes
15	Yes
16	No
17	Yes
18	No
19	Yes

Table 2 summarises the performance of the ellipsoid detector for finding known lesions, and the qualitative performance of the masked data infiller for highlighting them when the known location was exposed to the masked data model. Table 3 summarises this into a per-patient level, labelled with whether *either* the ellipsoid detection or infiller prediction was useful. It is worth noting that the results here were produced when using a single set of parameters in the ellipsoid detector, with the lowest false positive rate trialled. When the Roke team examined results from a detector with a higher false positive number, the true positive number also appeared to increase – in particular, some misses on the multiple lesions in patient 18 appeared to be found when switching to the less restrictive ellipsoid detector.

Nonetheless, time restrictions meant that a second round of qualitative validation with GEH radiologists was not possible. The outcome of the validation was that, of the nine patients with known lesion locations, either the ellipsoid detector or masked data prediction method appears to be useful for directing a radiologist's attention in seven cases for at least one lesion.

6 GRAPHICAL USER INTERFACE

Throughout the project a GUI based demonstration system has been developed and maintained, primarily to demonstrate algorithms developed and support stakeholder discussion. The demonstrator system has been built as a desktop application using Qt5 with Python binding from PySide2. The demonstrator was primarily developed and tested using Windows, however should be cross platform and able to run on Mac or Linux, although this has not been tested.

It should be noted that there are limitations to the current GUI application, the most significant of which include:

- Threading has not been implemented for any part of the application, causing any significant processing to cause the application to hang, which includes data loading, alignment, and detection computation.
- Numerous capabilities integrated into the GUI each have their own dedicated displays, however the layout of these isn't fully managed automatically, resulting in the need to manually resize the window and position splitters to display the desired display.

6.1 DATA LOADING

The first stage of any workflow using the demonstration tool is to load scan data for one or more patients. This can be done by selecting 'Load data' from the file menu, then selecting the top level directory for a patient using the load file dialog, as shown in Figure 41.

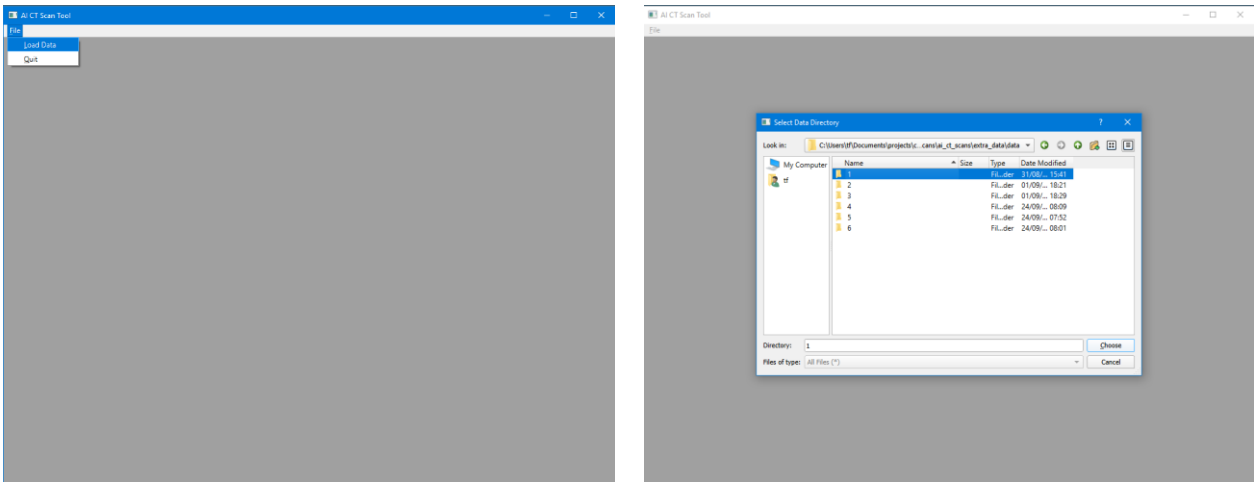


Figure 41 – Process for loading data into the demonstrator application.

If any of the data that is expected cannot be found in the location selected, an error dialog is displayed. This includes scan data, alignment transforms and the tissue sectioning model. Figure 42 shows an example in which the alignment transform was not found during data loading. In this case the remaining data will still be loaded and the tool can still be used, but this feature will be disabled.

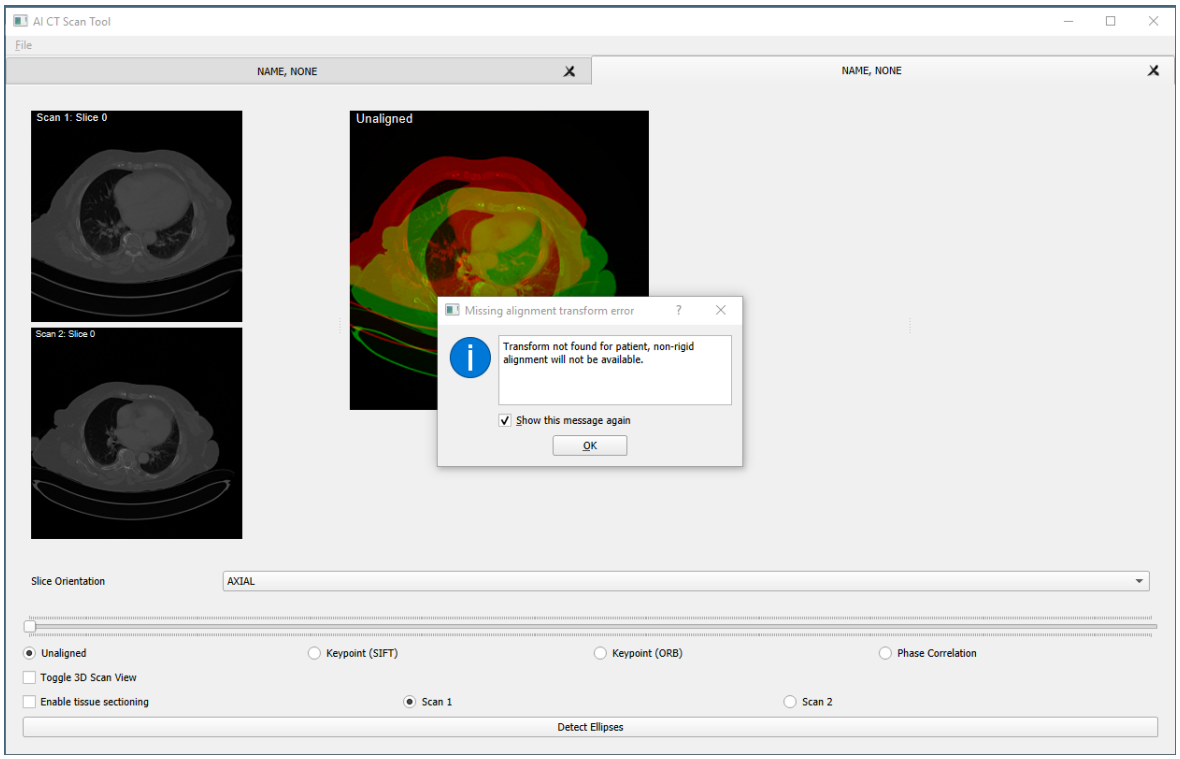


Figure 42 – Error message displayed by the demonstrator application as the alignment transform was not found.

Data can be loaded for multiple patients, each of which is displayed in a separate tab. The example shown in Figure 43 has data for three patients loaded. It should be noted that in all the examples shown throughout this report, the tabs are labelled with the name 'NONE', as the data is anonymised.

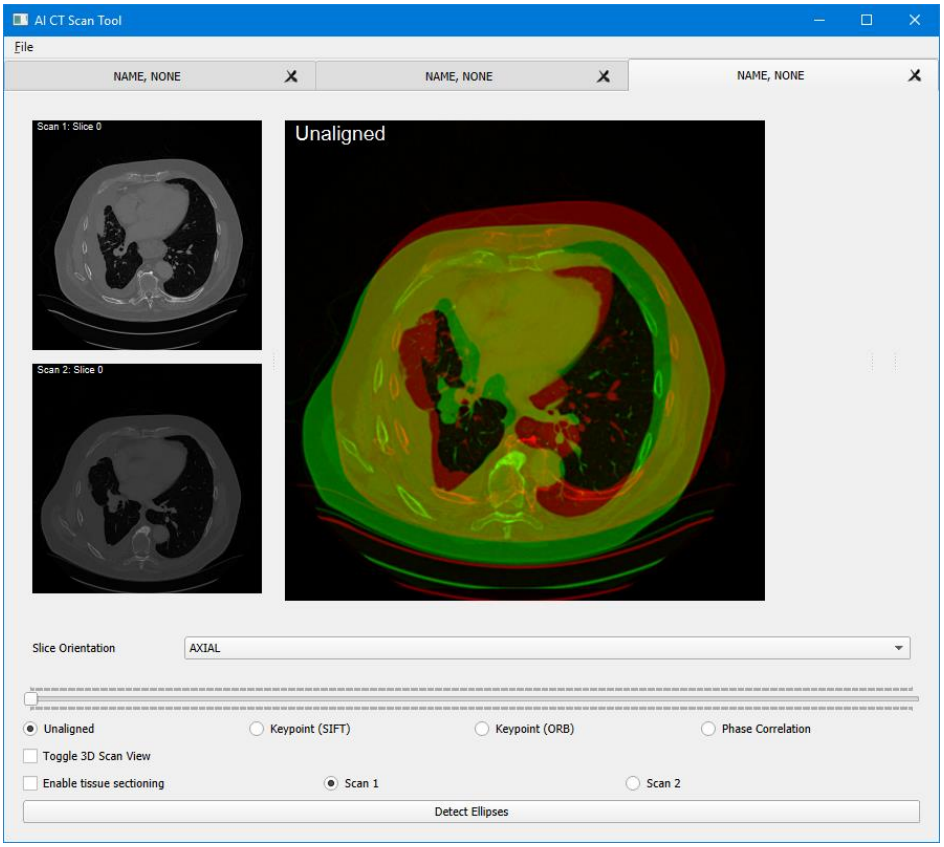


Figure 43 – Scan data for three different patients loaded into the demonstration tool.

Once data has been loaded, slices from both scans are shown on the left hand side of the window. The scan can be sliced in any of the three planes, selected using the 'Slice Orientation' drop down and the slice index can be selected using the slider below.

A 3D view of the current 3D alignment can be toggled on or off, with a plane to indicate current slice location. Methods which rely on changes to the underlying 3D data, e.g. phase correlation alignment, CPD alignment, and slice view position, require 'update 3D scan view' to be pressed for the related changes to be rendered. An example of the 3D view is shown in Figure 44.

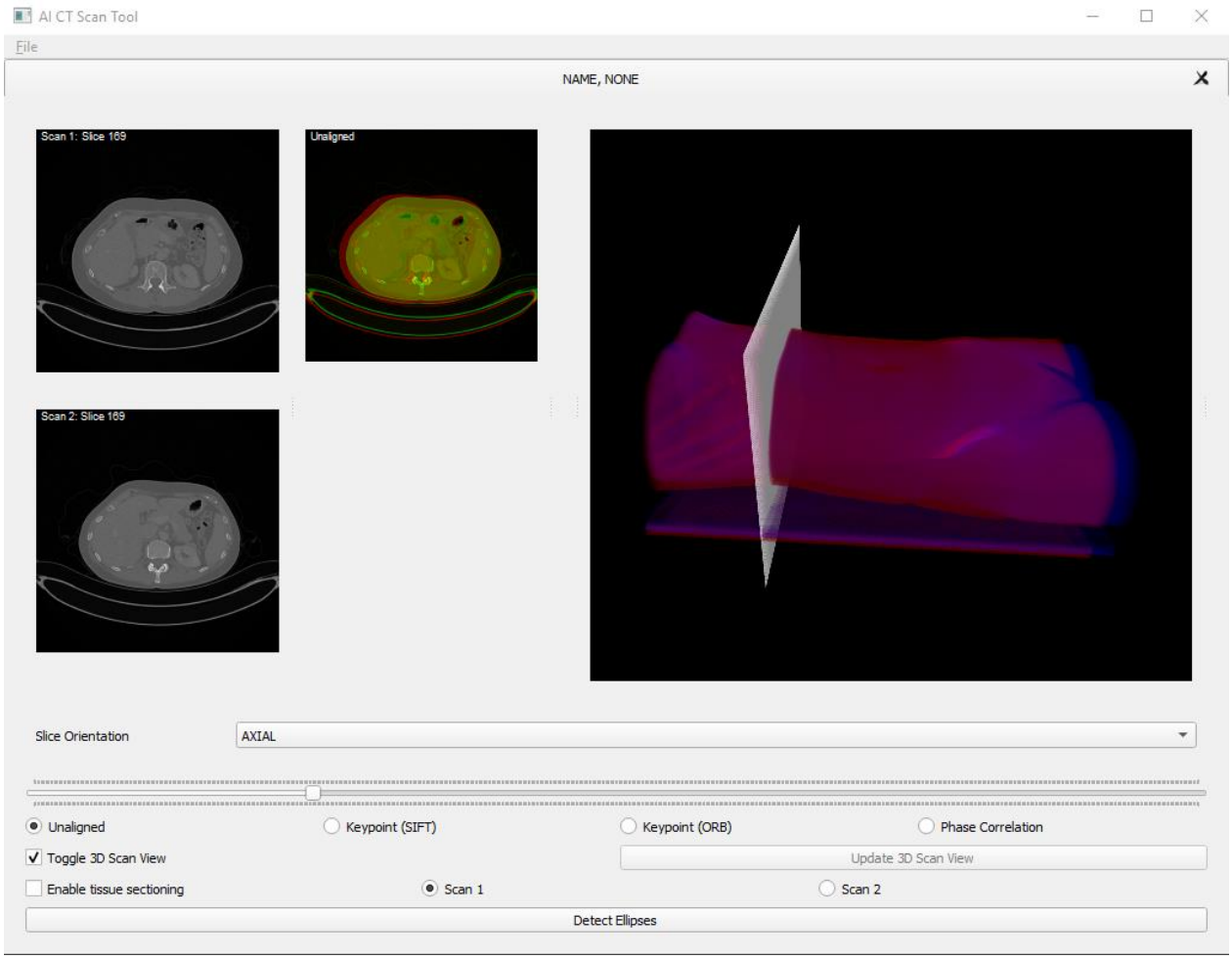


Figure 44 – 3D view rendered, scan 1 in red, scan 2 in blue, overlaying to appear puple. Rotation in 3D can be performed in real time, but alignment methods require 'update 3D scan view' to be pressed, as changes to the underlying 3D data will have been made and this requires a lengthier rendering step.

6.2 SCAN ALIGNMENT TOOLS

The demonstrator system provides five different alignment options, unless no alignment tranform is found, in which case four options are available. The result of whichever alignment option is selected is always displayed in the main results display, shown in the second column from the left. When the 'Unaligned' option is selected a naïve alignment of the currently selected slice is displayed, an example of which can be seen in Figure 43. When either 'Keypoint (SIFT)' or 'Keypoint (ORB)' is selected, a global alignment of the currently selected 2D slice from scan 2 generated using the relevant keypoint method is overlaid on scan 1, as shown in Figure 45.

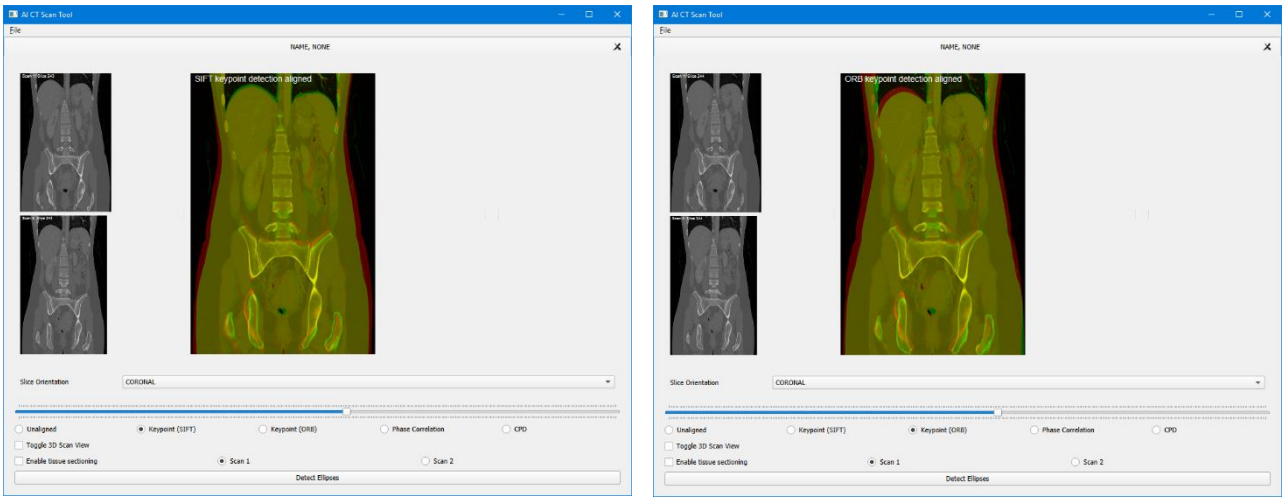


Figure 45 – Keypoint alignment of 2D scan slice using both SIFT and ORB being displayed in the demonstrator system.

Selecting the 'Phase Correlation' option will display an additional set of controls used to define a local region that the phase correlation alignment will be based around, indicated by a blue dot drawn on the scans. In order to calculate and apply the alignment the 'start' button must be pressed. This will calculate a phase correlation alignment based on the local 3D region defined which is applied to the full 3D scan before displaying the selected 2D slice, an example of which is shown in Figure 46. Once a phase correlation alignment has been calculated, moving the slider or orientation will apply the same alignment to the newly selected slice. In order to recalculate the alignment the 'Start' button must be pressed again.

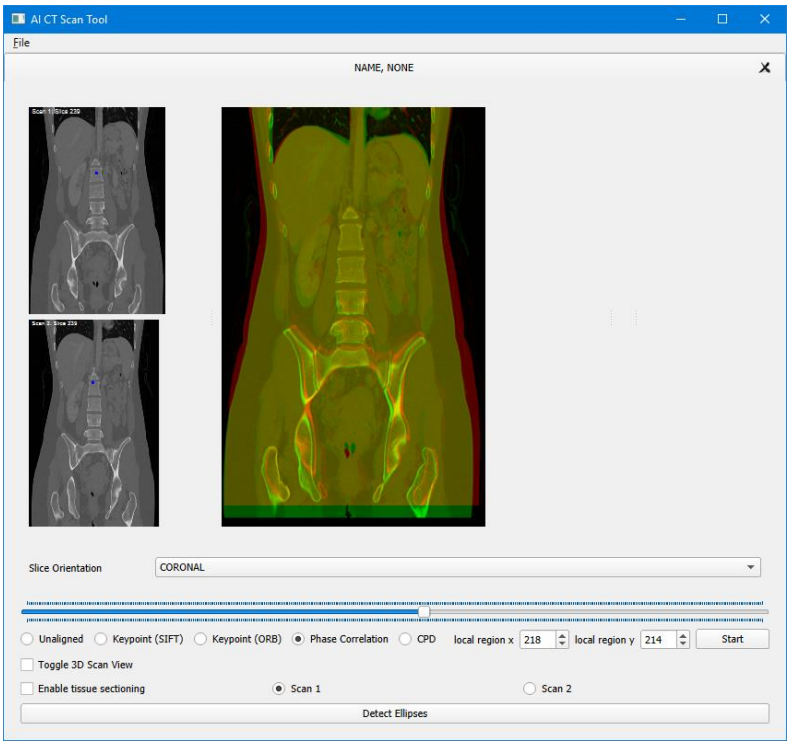


Figure 46 – Phase correlation based local alignment being configured, applied and displayed using the demonstration system.

If an alignment transform is found as the data is loaded then an alignment of the full second scan is calculated and the 'CPD' alignment option is displayed. Selecting this option will display an overlay using the relevant 2D slice from the aligned scan. A second display pane is also displayed which shows the volumetric change at each point in the slice used to produce the alignment. An example is shown in Figure 47.

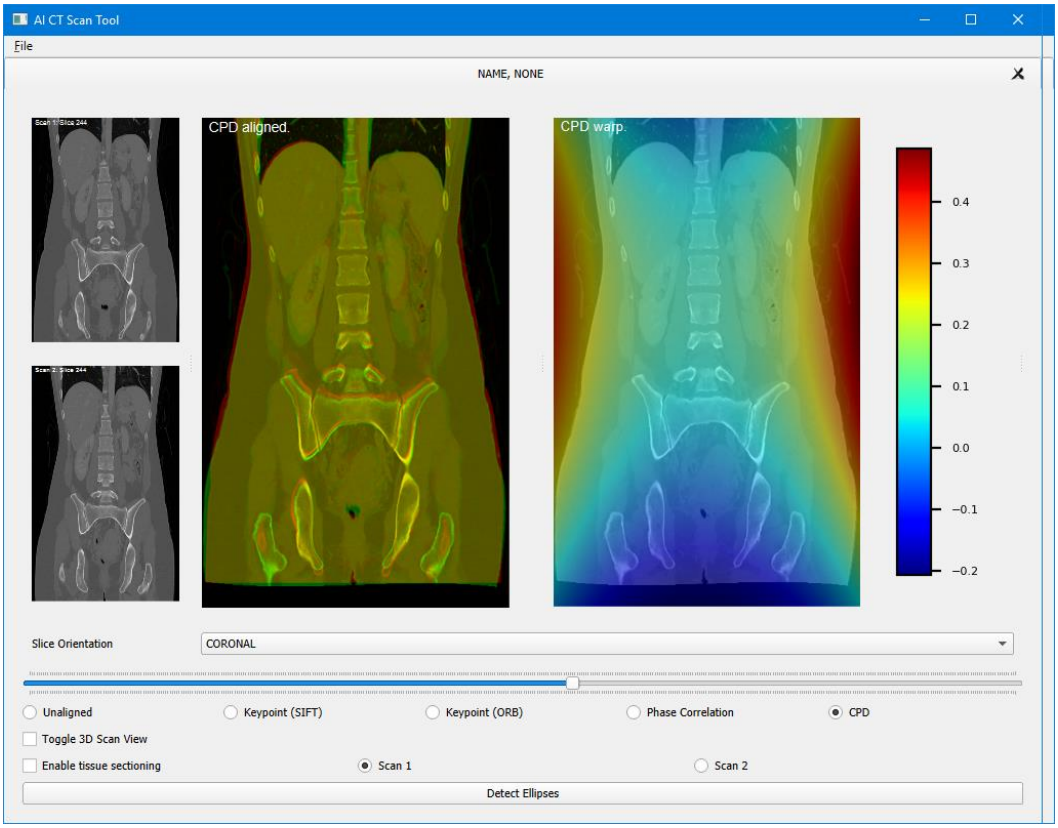


Figure 47 – Coherent Point Drift based non-rigid alignment being displayed in the demonstration system.

6.3 TISSUE SECTIONING TOOLS

Tissue sectioning can be applied to a scan by checking the 'Enable Tissue Sectioning' checkbox. This will show another display, containing the results of tissue sectioning on the 2D slice currently selected. This can be applied to either scan 1 or scan 2 using the relevant radio buttons. An example is shown in Figure 48.

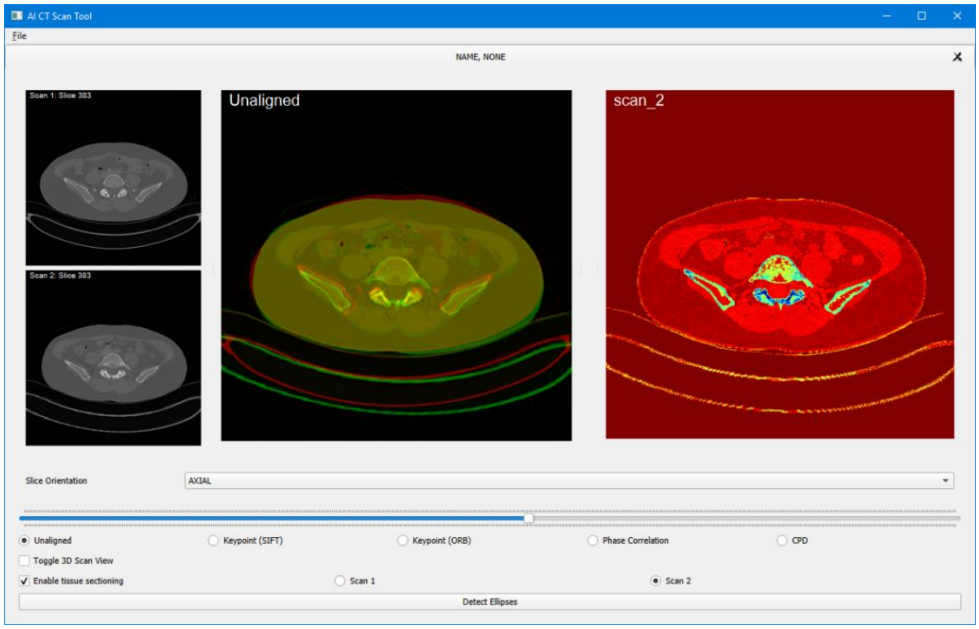


Figure 48 – Tissue sectioning being applied to a 2D scan slice in the demonstration system.

6.4 LESION/ANOMALY DETECTION TOOLS

From the anomaly detection methods investigated, ellipsoid detection was chosen for integration into the demonstration tool. This element is triggered through the button labelled “Detect Ellipses” seen at the bottom of the window (see Figure 49).

Once detection has completed a table will be displayed on the right hand side of the window which details the detected ellipsoids from both scans. A pop up will show when the processing has completed, detailing the total number of ellipsoids detected. The displayed table details the coordinates of the centre of the ellipsoid, the estimated volume in 3D, and the class of the ellipsoid (using the tissue sectioning output).

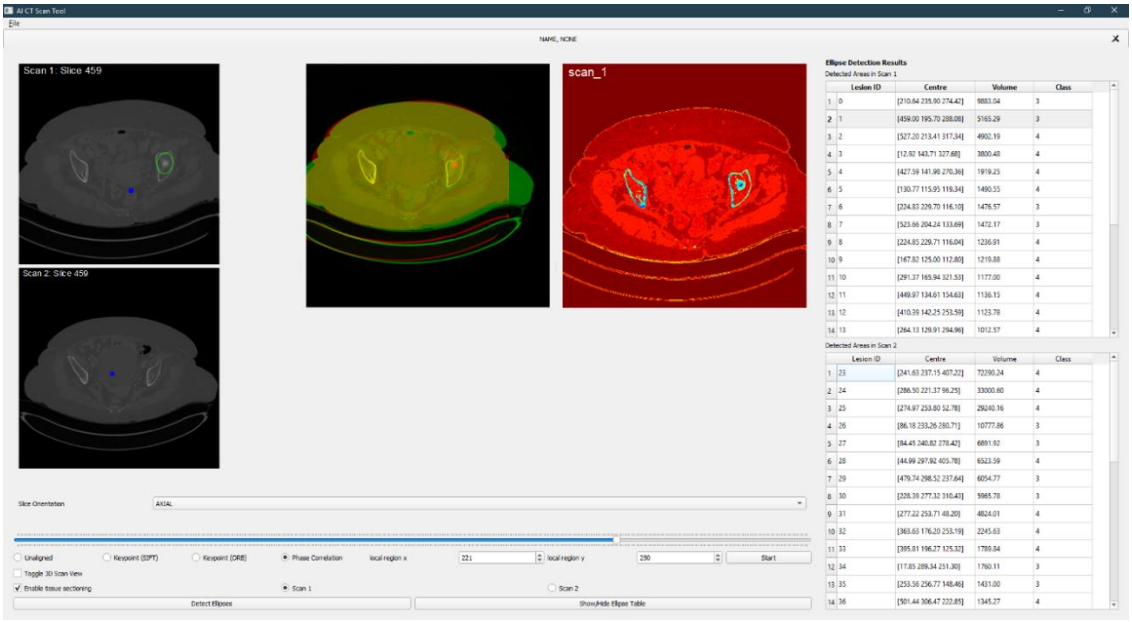


Figure 49 - Ellipse detection being applied to a set of scans. Highlighting the detected ellipsoids in the right-hand table and current selected ellipsoid overlain on the scan view. Tissue sectioning complements the ellipsoid detection features. Volume is measured in pixels, each pixel corresponding to 0.7mm in width.

Selecting an ellipsoid from the table with the mouse button will navigate the scan viewer elements to the relevant slice in the scan. In addition to this the user interface will overlay an ellipse on the slice viewer panes to show the location of the ellipsoid within the 2D slice shown, and if multiple detections have been made on neighbouring slices then moving the slice view forward or backwards allows the user to see other detected ellipses that contributed to the ellipsoid within the current orientation. The ellipse in the current slice is indicated by the green ellipse overlay. If no ellipse is shown in the 2D view then the selected ellipsoid was not detected in the current slice orientation (however is detected in one of the other slice orientations, else would not display in the table).

The user can also hide and show the ellipse table using the provided button (this may be desired to allow more space for other views (such as the sectioned or 3D view) within the window). One of the aims of the project was to measure the *change* in volume of matched lesions, and this was not achieved during the project, mainly due to time constraints. Though the alignment processes should bring this within rapidly achievable extensions, e.g. a simple matching of ellipsoids between scans based on the closest ellipsoid centre of the same class between tables on the two scans may suffice, properly validating such an approach and developing logic to deal with potential false matchings would be required. Such a matching between found ellipsoids then is currently a manual task, though the GUI rapidly displays a highlighted 2D ellipse associated with the ellipsoids, which should allow for fast inspection.

7 DISCUSSION

The alignment strand of work appeared especially strong in Roke's opinion: the phase correlation and coherent point drift techniques in this strand being especially promising. Phase correlation was achieving a reasonably robust local 3D alignment, which ran quickly enough to be used real-time during analysis to realign to new local regions. It did not, however, correct for patient body shape changes, or the change in lung volume problem which is a common problem to tackle.

Non-rigid alignment methods were trialled to tackle the lung volume and body shape problems. Keypoint-based methods showed some promise for global alignment over 2D slices, and were fast enough to use in real time while scrolling through a scan. Their success rate was far lower than the phase correlation technique, however, and would benefit from further work in finding automatic input parameters to help keypoint extraction on a per-scan basis.

The coherent point drift algorithm was found to achieve excellent global alignment for some patients, and was a fully non-rigid method that dealt well with the body shape changes and lung volume problems. As with the keypoint methods, though, it would benefit from automated tuning on a per-patient basis. It succeeded on 16 of 20 patients trialled, using parameters tuned for a single patient. It was a slow algorithm, however, taking around 1h to extract a transform. It would therefore require running in a batch fashion on a hospital's systems, perhaps overnight, before being ready for display to a radiologist the next day.

The tissue sectioning work achieved assurable, explainable results using a custom clustering algorithm with a texton-based approach, and explored the use of cutting edge deep learning models for the same aim, though the results of the DL methods were less convincing, less assurable, and higher engineering effort.

While the automatically discovered tissue class labels were not associated to 'real' tissue types, e.g. correlating 'class 1' with 'bone' and similar pairings, this would not be seen as a difficult extension in the future. Collaboration with radiologists would be necessary to complete this step.

The anomaly detection methods have shown potential, but both the ellipsoid detection and masked data methods would benefit from further fine tuning. Extensions to this end are given in the next section.

Table 1 is recast with delivered functionalities defined against each expected benefit in Table 4.

Table 4 – Outcomes of the commission compared to expected benefits

Scan Comparison	Current Process	Expected Benefits From the Proof of Concept	Delivered functionality
Alignment	Variable	• Classical Computer Vision methods, in 3D as standard, with 2D representations easily recoverable • Rigid alignment of CT scans • Alignment with breathing and other patient body changes accounted for with more advanced classical methods	• Fast 3D rigid alignment achieved via phase correlation • Breathing and other body changes dealt with via coherent point drift, observed to work in most cases
Overlay	Imprecise	• Precise overlay • Rapid 3D and 2D overlays rapid • Interactive elements such as panning, zooming, rotation	• Precise overlay achieved • 3D and 2D overlays are in the GUI tool • Interactive elements included in GUI tool
Change in lesion size	In one dimension	• Mix of classical and deep learning approaches • Section lesions in 3D which allows fast volumetric measurements	• Novel deep learning approaches explored but classical approaches formed the measurement components • Tissue sectioning was achieved, lesions often sectioned distinctly from surrounding tissue
New pathological lesion	Prone to miss	• Two self-supervised deep learning approaches to automatically identify lesions	• Deep learning approaches trialled, potential for extension on these techniques defined, and highlighted lesions in some validation cases • Computer vision techniques used to identify many lesions, extensions to improve capture rate described
Current process	Tedious and manual	• GUI to aid automation	• Accurate overlay, automated reporting of anomaly locations and properties in the GUI should both speed up inspection
Guidance to Radiologist	Not available	• Potential lesions highlighted • Growth automatically measured • Location of detected lesions presented to radiologist	• Anomaly locations and other properties presented in the GUI • Size of lesions measured, but change in size not completed
Time cost	30-40 min	• Reduced time cost due automation and visualisation	• Not yet measured via radiologist usage
Trust Litigation cost	Immense on miss	• Reduced chance of misses • Decrease risk of litigation	• Fully automated measures unlikely to reduce the chance of miss, but the accurate overlay may help to reduce misses during manual inspections – not yet measured

8 FUTURE WORK

As with most short-term proof of concept demonstrations, many avenues for improvement became apparent during the project. We briefly describe some of these for each vein of work below.

8.1 ALIGNMENT AND OVERLAY

- To Roke's knowledge, no other CT analysis product on the market attempts non-rigid global alignment, but coherent point drift has been demonstrated as a viable method to this end. Improving the automated selection of input parameters, and the extraction of the polynomial transform, could progress the method towards a more robust solution. In particular, 3D spline fitting to the output of the CPD algorithm, rather than a single 3D polynomial with ridge loss, may allow different regions of the scan to warp more independently of others, such that shape changes due to the patient's limbs lying in slightly different orientations would be less of a blocker to success.
- The phase correlation algorithm achieved reasonably robust, local 3D alignment. It typically took around ten seconds to run, and so could be used in real-time analysis, but would be tedious to use many times within the same scan. The speed of this alignment depended heavily on the size of the local region in 3D used for the algorithm, which was set as a 200x200x200 pixel volume in the proof of concept. Exploring the accuracy/speed balance in reducing this volume might allow local alignment to occur while scrolling through slices.
- In almost all scans, at least one alignment method succeeded, but currently a user has to select each in turn to find the 'best' overlay. A method of automatically selecting the best alignment method could speed up analysis. Typically, poor overlays were due to catastrophic failure of the alignment, such that scan 2 was warped to an obviously erroneous degree. Various classical computer vision techniques may be able to de-prioritise obvious failure cases, and deep learning methods may be able to differentiate the quality of success cases. The discriminator component of a GAN may be able to rank the most realistic-looking aligned images, which could eliminate many failure cases, but it would be important to investigate how the discriminator performs on images not within the associated generator's output distribution before relying upon this method. A self-supervised classifier approach, where the objective is to judge 'transformed' from 'non-transformed' may also be viable – very well transformed images should be hard to distinguish from non-transformed, and could also eliminate the more obvious failure cases without a radiologist having to inspect them.

8.2 TISSUE SECTIONING

- The HMS clustering and texton-based tissue sectioning appeared to perform well, but no texture information was used in the best model developed. The original texton paper (Malik, 2001) offers some methods to deal with the edge defects that prevented this during the course of this proof of concept, and a study into this may allow higher fidelity tissue sectioning with a similarly assurable, fast final tool.
- The texton clustering backend tools included a 'probability view' for each class, as in Figure 27, which did not make it into the GUI. If this style of contrast enhancement seems valuable to other stakeholders, it may form a useful visualisation in later versions of the GUI.
- The DINO vision transformer attention head outputs showed some tissue separation, but at a reduced resolution which was likely to make the sectioning too coarse. Additionally, the number of distinctly separated tissue types appeared worse than via the texton clustering. To reach full resolution attention masks, a patch size of 1x1 pixels would be required, which would increase compute resources considerably – well beyond what the original DINO researchers used. If the DINO methods were deemed of high interest, scoping and securing this increased compute resource would be the most sensible next step.

8.3 ANOMALY DETECTION

- Early in the project, the masked data model training objective was chosen to look at only a single scan at a time, attempting to fill in the masked region from surrounding context. An obvious alternative, given an aligned dataset of sequential, paired scans, would be to use scan 1 to predict the masked region of scan 2, and vice versa. The reason this was not attempted during this proof of concept was because an aligned dataset was not available. Now that alignment methods are available, such a dataset could be generated, and this alternate training objective for a masked data DL model could be trialled.
- Improved balancing of the true positive and false positive rate from the ellipsoid detector should be achieved through radiologist discussion, and automated hyperparameter grid-search methods.
- The ellipsoid detector might benefit if it were to fit to convex hulls around contours, rather than to contours directly. This may allow for 'spikey' lesions to be detected more robustly.
- In some cases, ellipsoids that related to true lesions were missed because of small tendrils of the same tissue type that were connected. Various computer vision techniques for separating these may lead to improve detection, e.g. erosion (Erosion (morphology), n.d.), watershed segmentation (Watershed, n.d.), or spectral clustering (Spectral clustering, n.d.).
- While ellipsoid volumes were measured on a per-scan basis, they were not linked between scans, meaning that volume changes would still be somewhat manual to measure. Automating this is desirable. Applying the ellipsoid detection to aligned scans, followed by a Hungarian matching method applied to the ellipsoid centres of the same tissue class, may be a quick win to achieve volume change measurements.
- The masked data prediction model appeared to infill lesions with healthy tissue often, and usually filled healthy regions with similar features to the input, but using the difference between the input and output of the model for ranking of anomalies still remains a challenge. Finding a suitable single-number metric for scoring the masked data model's outputs with the ellipsoid detector could enhance its usage.

8.4 GUI

- While the GUI displayed the high-value backend tools developed on the project, it became clear early on that 'quality of life' features of a more mature GUI would be important for adoption by radiologists. These would include double-clicking to maximise particular views, presenting views in familiar colour schemes, faster 3D display updates after alignment steps were made, etc. Integrating the backend tools into an existing product, such as a PAC system already in use by hospitals, may be an attractive option, or gathering a fuller definition of the front-end requirements to build a new GUI environment from the ground up.

- Alignment of sequential scans for comparison was developed on this commission specifically against an oncological dataset, but this technique is not unique to oncology, and the DICOM format is used for many other image sources in a clinical setting. Validating the usefulness of the tool against CT datasets recorded for other diagnostic purposes would be valuable, and the backend functionality is certainly not expected to be limited to oncological CTs. Other sources of 3D and 2D medical imaging data, e.g. MRI scans and x-rays, could also be trialled against the backend methods with fairly rapid data piping tasks for integration.

APPENDIX A REFERENCES AND GLOSSARY

A.1 REFERENCES

Arvai, K. (n.d.). *Kneed*. Retrieved from Kneed: https://kneed.readthedocs.io/en/stable/

Bradski, G. (n.d.). *Finding contours in your image*. Retrieved from OpenCV: https://docs.opencv.org/3.4.15/df/d0d/tutorial_find_contours.html

Bradski, G. (n.d.). *OpenCV*. Retrieved from OpenCV: https://opencv.org/

Caron, M. T. (2021). Emerging properties in self-supervised vision transformers. *arXiv*.

DBSCAN. (n.d.). Retrieved from Wikipedia: https://en.wikipedia.org/wiki/DBSCAN

Duda, R. O., & Hart, P. E. (1972). Use of the Hough Transformation to Detect Lines and Curves in Pictures. *Commun. ACM*, 11-15.

Erosion (morphology). (n.d.). Retrieved from Wikipedia: https://en.wikipedia.org/wiki/Erosion_(morphology)

Gabor filter. (n.d.). Retrieved from Wikipedia: https://en.wikipedia.org/wiki/Gabor_filter

Gattia. (2021, 09 08). *Cycpd*. Retrieved from Github: https://github.com/gattia/cycpd

H. Foroosh, J. B. (2002). Extension of phase correlation to subpixel registration. *IEEE Transactions on Image Processing*.

Hounsfield scale. (n.d.). Retrieved from Wikipedia: https://en.wikipedia.org/wiki/Hounsfield_scale

Hungarian algorithm. (n.d.). Retrieved from Wikipedia: https://en.wikipedia.org/wiki/Hungarian_algorithm

ImageNet. (n.d.). Retrieved from Image-net: https://www.image-net.org/

Lowe, D. G. (1999). Object recognition from local scale-invariant features. *International Conference on Computer Vision*, 1150–1157.

Malik, J. B. (2001). Contour and Texture Analysis for Image Segmentation. *International Journal of Computer Vision*.

Muja, M., & Lowe, D. G. (2009). Fast Approximate Nearest Neighbors with Automatic Algorithm Configuration. *VISAPP*.

Newsroom. (n.d.). Retrieved from Nvidia News: https://nvidianews.nvidia.com/news/nvidia-doubles-down-announces-a100-80gb-gpu-supercharging-worlds-most-powerful-gpu-for-ai-supercomputing

Pang, G. S. (2020). Deep learning for anomaly detection: A review. *ACM Computing Surveys (CSUR)*.

Rublee, E., Rabaud, V., Konolige, K., & Bradski, G. (2011). ORB: An efficient alternative to SIFT or SURF. *International Conference on Computer Vision*, 2564-2571.

Sharifi, M. M. (2002). *A Classified and Comparative Study of Edge Detection Algorithms .*

Shrimali, K. (n.d.). *Convex Hull using OpenCV in Python and C++*. Retrieved from LearnOpenCV: https://learnopencv.com/convex-hull-using-opencv-in-python-and-c/

Siewert, B. S. (2008). Missed lesions at abdominal oncologic CT: lessons learned from quality assurance. *Radiographics*.

Sobel, I. (2014). *History and definition of the sobel operator*. Retrieved from https://www.researchgate.net/publication/239398674_An_Isotropic_3x3_Image_Gradient_Operator

Song, A. M. (2010). Point Set Registration: Coherent Point Drift. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Spectral clustering. (n.d.). Retrieved from Wikipedia: https://en.wikipedia.org/wiki/Spectral_clustering

Watershed. (n.d.). Retrieved from Wikipedia: https://en.wikipedia.org/wiki/Watershed_(image_processing)

Zhu, S. C. (2005). What are textons? *International Journal of Computer Vision*.

A.2 GLOSSARY

If you use any abbreviations or acronyms, list them here. eg:

GUI	Graphical User Interface
LMR	local mean removal
CPD	Coherent Point Drift
BGM	Bayesian Gaussian Mixture
HMS	Hierarchical Mean Shift
DL	Deep learning
vit	Vision transformer

Roke Manor Research Ltd
Roke Manor, Romsey
Hampshire, SO51 0ZN, UK

T: +44 (0)1794 833000
F: +44 (0)1794 833433
info@roke.co.uk
www.roke.co.uk

Approved to BS EN ISO 9001
Reg No Q05609