

Artificial intelligence-assisted detection of adhesions on cine-MRI

Master Thesis

Evgenia Martynova - s1038931

Abstract— Adhesive disease, which commonly occurs as a postoperative complication, is a major cause of morbidity and places a substantial burden on healthcare worldwide. Currently, laparoscopy is the only accurate diagnostic technique for abdominal adhesions, which intrinsically involves health risks including the formation of new adhesions. Non-invasive diagnostic methods with similar reliability are lacking. In recent years using cine-MRI scans of the abdomen captured during respiration has demonstrated promising performance in the diagnosis of adhesions. However, correct interpretation of cine-MRI scans requires considerable radiological expertise and this technique has not been widely adopted in clinical practice yet. In this masters thesis, the first fully-automated multi-stage computer-aided diagnosis (CAD) method for adhesion detection is proposed. The method exploits the phenomenon of visceral slide, a pattern of abdominal motion observed during respiration in healthy subjects. Local reduction of visceral slide is a diagnostic criterion of adhesions. Visceral slide that occurs on a cine-MRI slice is quantified using a segmentation mask generated by a deep learning model and a deformation field between cine-MRI frames obtained with an image registration algorithm. Bounding boxes of adhesions are predicted with a region growing method based on the visceral slide values. Additionally, false positives reduction driven by domain knowledge is performed. The impact of using all cine-MRI time points and different normalisation options are investigated and the hyper-parameters of the method are determined with 5-fold cross-validation. When evaluated with cross-validation, the best method configuration yields detection sensitivity of 0.61 and 0.73 at 1 and 2 false positives per slice along with 0.53 AUC in slice-level diagnosis. On the held-out test set, a slightly different configuration is top-performing and achieves detection sensitivity of 0.7 and 0.91 at 1 and 1.89 false positive per slice along with 0.78 slice-level AUC, which indicates the promising potential of the core idea of the method.

I. INTRODUCTION

Adhesions are tough bands of fibrous tissue that form between two or more organs in the abdominopelvic region and/or the inner abdominal wall. The primary cause of the adhesive disease is abdominopelvic surgery, however, they may also form secondary to inflammatory conditions of the abdomen or as a sequela of abdominopelvic radiation [1]. Adhesions make internal organs to be stuck together when they are not supposed to be. Due to the hindrance of normal abdominal motion, various complications may occur. The recently performed meta-analysis [2] has shown that complications of adhesions are frequent and have a large negative effect on patients' health. The most common complications are chronic pain, small bowel obstruction, prolonged operative time, bowel injury during adhesiolysis (a surgical procedure to remove adhesions), and a significantly lower pregnancy rate. The incidence of postoperative adhesions is high. For instance, in a post-mortem study

examining 752 cadavers, adhesions were found in 67% that had undergone a single laparotomy and in 93% of cadavers that had undergone multiple laparotomies [3]. Consequently, adhesion illness increases workload in clinical practice and puts a considerable burden on the economy. For instance, in 1994, the costs for in-patient treatment of adhesion-related morbidity in the United States were estimated to be \$1.3 billion per head [4]. Therefore, improvements in the treatment strategy of the adhesion-related complaints are of interest.

Currently, diagnosis of adhesions requires invasive tools such as diagnostic laparoscopy owing to a lack of effective noninvasive tests. These methods are controversial because they can lead to the formation of new adhesions. This is especially undesirable for the patients whose complaints are not due to adhesions. According to [5] negative-finding rates in studies that have used diagnostic laparoscopy for identifying adhesions is established to be 15%, which is quite high considering the involved risks.

Non-invasively adhesions can be diagnosed with either ultrasound or cine-MRI. Cine-MRI is more powerful than ultrasound because it can detect adhesions in the entire abdomen, whilst ultrasound can detect only adhesions attached to the front abdominal wall due to its limited depth penetration [4]. Cine-MRI is a type of MRI in which a set of consecutive images of the area of interest is acquired at a fixed time interval. The captured images can be merged into a video in which the movements of a certain tissue over time are visible. This imaging modality can be used to visualise the abdominal motion during respiration and by interpreting the recorded motion patterns, radiologists can diagnose adhesions. This method has shown potential for adhesion detection with two studies indicating sensitivities of at least 87.5% when compared with gold standard surgical confirmation [4] [6]. Moreover, usage of cine-MRI for the choice of the treatment strategy for patients with chronic pain caused by adhesions has recently been found beneficial [5]. The approach proposed in the study enabled achieving long-term pain relief in two-thirds of patients as well as a decrease in overall healthcare utilization. Nevertheless, cine-MRI is not widely used in hospitals for making treatment decisions yet since radiological reading is time-consuming and expertise-dependent.

If an accurate computer-aided diagnosis (CAD) system for adhesion detection on cine-MRI was available, that could facilitate and speed up the adoption of cine-MRI usage in clinical practice. First, such a system could assist in the training of radiologists and decrease the learning curve.

Besides, it can improve the accuracy of cine-MRI scans interpretation by decreasing the number of false positives and negatives. False positives can result in unnecessary surgery, whereas false negatives can block surgery for patients who may actually be helped by it.

II. RELATED WORK

Little work has been done to develop a CAD system for adhesion detection on cine-MRI. To our best knowledge, the only semi-automatic method designed so far was proposed in David Randall's PhD thesis [7].

The method is suitable for the detection of adhesions attached to the anterior abdominal wall. It exploits the discontinuity of the abdominal motion during respiration. In healthy subjects, abdominal contents slide smoothly against the surroundings of the abdominal cavity (abdominal wall, back muscles, etc.). Simultaneously, the abdominal wall exhibits a different, anteroposterior mode of motion. This process is called visceral slide. Reduction in visceral slide is a clinical criterion of underlying adhesions. The key idea of the method is that it is possible to quantify the degree of visceral slide captured on a sagittal slice of a cine-MRI scan by computing shear along a boundary separating abdominal contents from its surroundings. The output of the method is a so-called "sheargram" which is obtained by summing shear across all frames of a sagittal slice to include full information about the abdominal motion into the algorithm's outcome. A low degree of shear should indicate the presence of adhesions. The resulting sheargram is meant to be used as a complementary diagnostic aid for medical specialists and requires human judgment whether a reduction of shear is observed.

A pilot study performed to validate the method demonstrated a 84% agreement on the observed reduction in shear with clinical judgment from the specialists [8]. It is noteworthy that interpretation of sheargrams was performed by the same specialists who examined cine-MRI slices to determine the presence of adhesions.

The stages of the method are illustrated in figure 1. To compute a sheargram, the method needs a segmentation map separating the abdominal cavity from its surroundings based on which a masked image registration is performed. Image registration is the process of geometrically aligning two images from various imaging sources. One of the images is referred to as the fixed image and another is referred to as the moving image. Image registration involves spatially transforming the moving image) to align with the fixed image. Registration algorithm outputs a deformation field, which describes this transformation by specifying how each point of the moving image should be shifted to make it as close to the fixed image as possible. In the context of shear computation, the goal is to quantify the motion of abdominal contents and their surroundings across a sagittal cine-MRI slice. The deformation field can be thought of as quantification of this motion and hence shear between frames can be computed from the obtained deformation field. In Randall's method, masked registrations of abdominal

contents and their surroundings are done separately and the complete deformation field is obtained from the addition of two masked deformation fields. Typically registration algorithms do not perform well on the motion that occurs in opposite directions, hence such an approach enables more reliable capturing of the overall motion. For masked registration, an accurate segmentation map of the abdominal cavity is needed.

One of the limitations of the method was the lack of segmentation automation, which is mentioned among the desirable future work directions in the original PhD. Segmentation was performed semi-automatically. For the first cine-MRI frame a segmentation map was drawn by a medical specialist and transferred to the subsequent frames with image registration. Still, manual correction of registration results was required for each frame. Although this method is faster than fully manual annotation, it remains time-consuming and suffers from the annotator's subjectivity. Also, the method assumes the manual interpretation of the sheargram by a radiologist and does not explore ways to predict adhesions based on low shear value automatically.

A. Contributions

This master's thesis proposes the first fully-automated multi-stage CAD method for adhesion detection on cine-MRI designed based on Randall's semi-automated method. The automation comprises deep learning-based segmentation of the abdominal cavity and the region growing algorithm designed to predict adhesions using the value of quantified visceral slide as an input. Apart from automation, a few other contributions summarised in the following list were made:

- The method was extended to be applicable to the detection of pelvic adhesions as well. This means that it is able to detect adhesions located at any position along the abdominal cavity boundary.
- An approach to visceral slide quantification was revised and changed to an alternative one that outputs visceral slide in deformation field units, that can be converted to millimetres, making it more interpretable.
- Two different approaches to visceral slide quantification were explored. The first one uses the two most dissimilar frames of a cine-MRI slice to highlight the differences in the position of the abdominal cavity boundary and abdominal organs at the opposite phases of the respiratory cycle. Another method uses all frames of a cine-MRI slice to exploit temporal information and build a complete picture of abdominal motion.
- Multiple normalization options that account for motion difference between patients and in different areas of abdominal cavity contour were carefully designed and integrated. This enables the same interpretation of a particular visceral slide value regardless of the amplitude of motion that occurs on a cine-MRI slice and the position at which it is observed.
- A false-positives reduction technique was designed based on the domain knowledge about possible adhesion location. For that, an accurate algorithm to detect

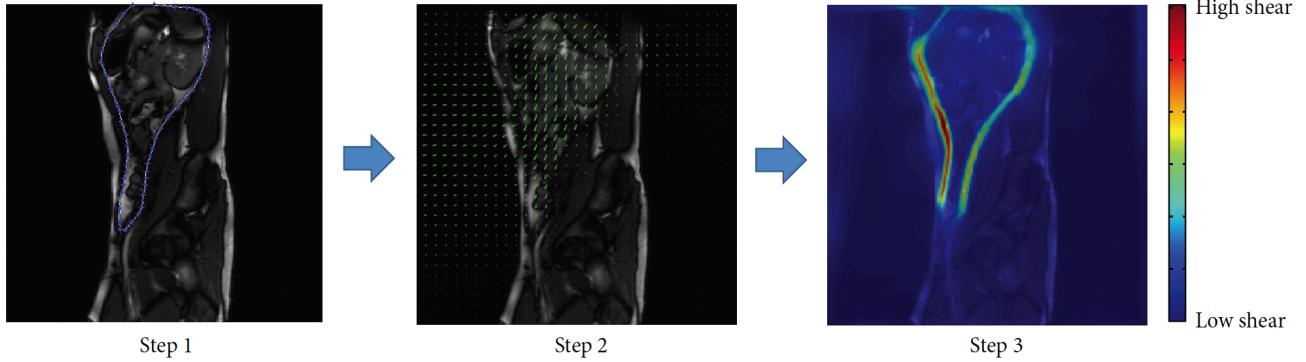


Fig. 1: Flowchart describing Randall’s methodology. Step 1: semi-automated abdominal cavity segmentation (blue line); step 2: depiction of the mathematically quantified movement as a displacement field (green arrows); step 3: depiction of the shear taking place along the boundary in a “sheargram”. The picture is taken from [8].

different parts of the abdominal cavity (top, bottom, anterior wall, posterior wall) was developed, so that it was possible to leave out any subset of abdominal cavity contour during prediction. The resulting algorithm can be reused in other methods for adhesion detection or related tasks.

- As the first automated method for adhesion detection, the method can serve as a baseline for CAD systems that are developed for this task in future.

All possible method configurations with respect to the ways to calculate and normalise visceral slide were evaluated with double cross-validation from the perspective of adhesion detection task as well as the ability to classify cine-MRI slice with respect to adhesions presence. The evaluation yielded a few insights into the applicability of the method to the detection of adhesions attached to the anterior abdominal wall and located in the pelvis area.

III. MATERIALS AND METHODS

A. The entire dataset

The entire dataset available for the project contains 563 studies for 526 patients of which 397 are female and 130 are male. The age of the patients varies from 13 to 84 years. The maximum number of studies per patient is 3 but usually, patients have only one study. Each study was provided with a report containing a radiologist’s diagnosis based on the assessment of a cine-MRI scan. From the conclusions in the report, binary study-level annotations were extracted. Since only patients with relevant complaints are prescribed cine-MRI examination, the dataset is biased towards positive patients and according to reports 408 studies are positive and 155 are negative. Generally, studies contain from 6 to 9 sagittal slices and each slice consist of 30 frames taken with a frame rate of approximately 2.69 frames per second.

All cine-MRI images were taken with 1.5T scanner from Siemens (Siemens, Magnetom Avanto, Erlangen, Germany) at Radboud University Medical Center (RUMC) over the period February 2012 - February 2020.

The method requires two types of data: annotations of adhesions with bounding boxes for detection algorithm and binary masks of the abdominal cavity for segmentation model.

B. Adhesion detection dataset

To be suitable for the visceral slide quantification algorithm, sagittal slices have to meet certain criteria. First, the amplitude of motion performed by a patient should be sufficiently high. Also, midline and left/right paramedian slices are the most informative for diagnosis and are viewed by radiologists in the first place, whereas left/right lateral slices rarely contain enough details. The elaborate discussion of suitability criteria and the used data sampling protocol is given in section VI-A in the appendix.

1) *Adhesion annotations*: To acquire bounding box annotations of adhesions, 197 slices from 85 patients were selected for two reader studies, which were completed by an experienced radiologist. In total 153 adhesions were found on 103 slices in 73 patients, hereby on some cine-MRI slices multiple adhesions are present. It is noteworthy that available annotations are not surgically confirmed and hence not ground truth, but reference standard.

There are three sub-types of adhesions based on typical locations where they occur:

- 1) *Attached to the anterior abdominal wall*, when adhesion is formed between the anterior abdominal wall and intestinal loops.
- 2) *Pelvic adhesions*, when adhesion is developed between the bladder and/or uterus and intestinal loops.
- 3) *Adhesions inside the abdominal cavity*, when an adhesion forms between intestinal loops.

Examples of annotations visualised on an arbitrary frame of a corresponding slice are depicted in figure 2. Adhesions at all possible locations are present in these examples. By design, the proposed method can only detect adhesions that are formed along the boundary of the abdominal cavity. Hence, annotations of adhesions located inside the abdominal

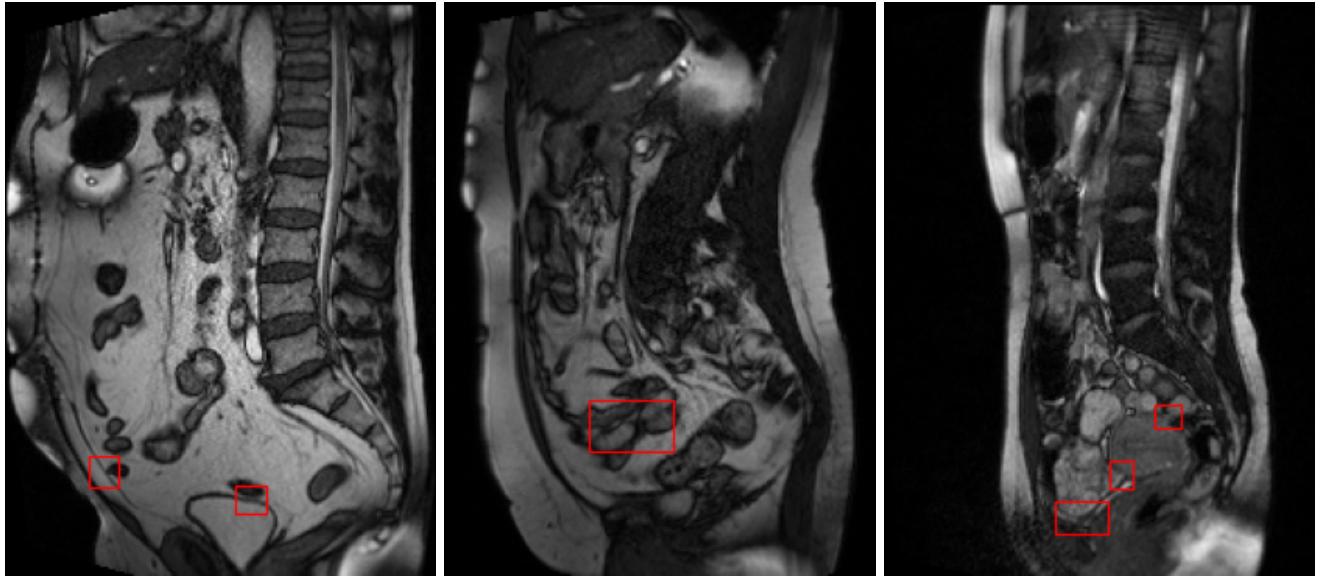


Fig. 2: Examples of frames of sagittal cine-MRI slices with visualised adhesion annotations. On the left slice one adhesion is attached to the anterior front wall and another is formed between the bladder and intestinal loops. On the second slice one large adhesion developed between intestinal loops is present. On the last slice there are three pelvic adhesions between the uterus and intestinal loops.

cavity were disregarded during the method development and evaluation.

2) Training, test and healthy control datasets: The dataset drawn for the development of the method consists of a training set, a held-out test set and a healthy control group to calculate visceral slide statistics. Slices on which adhesions were found in reader studies comprised a subset of positive slices. Negative slices were filtered according to the aforementioned suitability criteria, which gave a negative subset of 110 slices from 99 patients.

Training set: Composed to be balanced by adhesion occurrence at patient level. For the training set annotations from the first reader study were available, that is 95 adhesions on 63 slices in 51 patients. Since the method can detect only adhesions along the abdominal cavity boundary, only annotations at the corresponding locations (attached to the anterior wall and pelvic adhesions) could be included in the subset of positive patients. One slice with suitable annotation was excluded due to artefacts. The resulting positive subset comprised 59 slices for 50 patients and the same number of negative patients and slices was randomly drawn from the negative subset.

Table I presents statistics of annotations from the first reader study filtered by adhesion location. For each location, the number of annotations at this location and the number of slices/patients that contain at least one annotation at this location are given. The table highlights that the majority of adhesion annotations in the training set correspond to pelvic adhesions.

Test set: For the held-out test set 30 patients (15 negative and 15 positive) were selected according to the available binary study-level labels. For each patient, only one study was included, from which three suitable slices were selected:

TABLE I: Statistics of adhesion annotations by location in the first reader study. The data subset corresponding to the row marked with asterisk was used in the training set.

Type	Number of adhesions	Number of slices	Number of patients
All	95	63	51
Anterior wall	18 (18.9%)	18 (28.6%)	12 (23.5%)
Pelvis	65 (68.4%)	50 (79.4%)	48 (94.1%)
Anterior wall & pelvis*	83 (87.4%)	60 (95.2%)	50 (98.0%)
Inside	12 (12.6%)	11 (17.5%)	11 (21.6%)

midline and left/right paramedian.

Bounding box annotations were obtained from the second reader study, in which 58 adhesions were found on 40 slices belonging to 22 different patients (studies) and only 8 patients (studies) were not assigned any annotations. Thereby, binary study-level annotations from radiologists reports are inconsistent with the outcome of the reader study. In two sets of annotations, the binary status of studies with respect to adhesions presence matches for 18 studies, from which 12 are positive and 5 are negative. This yields Cohen kappa of 0.2, which can be interpreted as a slight agreement.

The detailed statistics of adhesion annotations by location in the second reader study is given in table II. Two slices that only have adhesion annotations inside the abdominal cavity were removed from the final test set as well as three slices with artefacts. Hence, eventually, the test set consisted of 38 positive slices and 47 negative slices. It is noteworthy, that positive patients had some slices which were found adhesions free in reader study.

In the test set, the percentages of adhesion annotations

along the anterior wall and in the pelvis area are similar.

TABLE II: Statistics of adhesion annotations with bounding boxes by location in the second reader study. The data subset from the row marked with asterisk was used for the final evaluation of the method.

Type	Number of adhesions	Number of slices	Number of patients
All	58	40	22
Anterior wall	25 (43.1%)	22 (55.0%)	13 (59.1%)
Pelvis	31 (53.4%)	29 (72.5%)	19 (86.4%)
Anterior wall & pelvis*	56 (96.6%)	38 (95.0%)	20 (90.9%)
Inside	2 (3.4%)	2 (5.0%)	2 (9.1%)

Healthy control group: The remaining 34 negative patients, having 36 slices, were used as a healthy control group to calculate visceral slide statistics.

3) *Abdominal cavity segmentation dataset:* For the abdominal cavity segmentation task, we selected a subset of frames from cine-MRI slices and segmented it under the supervision of an experienced radiologist. In total it included 201 segmented frames, of which 140 are midline and 61 are paramedian. Midline slices were segmented for all 68 patients but segmentation of paramedian slices is only available for 28 of 68 patients.

C. Validation data split

During development, the method was evaluated with 5-fold cross-validation. To avoid information leak from the training to validation sets in folds, the data were partitioned by patient. Since two types of annotations for separate sub-tasks are used in the method, both should be taken into account for validation data split. In other words, we do not want the patients used to train the segmentation model to be present in the validation set of the detection algorithm to avoid overestimation of the algorithm performance. In addition, folds should be stratified by the presence of adhesions in patients and, for segmentation task, patients with available paramedian segmentation should be equally distributed across folds.

Therefore validation data split was made with stratification by the presence of adhesions, availability of segmentation and paramedian segmentation in particular. This way, folds are balanced by all the necessary criteria.

D. Automation of Randall's method

As shown in figure 3, the fully automated method consists of three steps: obtaining segmentation map for an input slice, visceral slide calculation and prediction of adhesions with region growing algorithm. Additionally, two types of visceral slide normalisation are integrated: by average horizontal motion and by healthily control statistics.

1) *Automated abdominal cavity segmentation:* To automate segmentation of abdominal cavity nnU-Net [9] model was trained. nnU-Net is a self-configuring deep learning-based method for medical image segmentation that achieves

the state of the art performance on a broad range of medical segmentation tasks. It is based on the popular U-Net architecture [10], which outperformed previously leading models with a large margin when it was introduced. To improve performance and generalisability as well as reduce manual efforts required for U-Net training for a new task, the nnU-Net creators focused on making robust design decisions based on the distilled expert knowledge and input data fingerprint. They separate design choices into three categories: fixed parameters, rule-based parameters and empirical parameters. Fixed parameters are established based on domain knowledge and are data-independent (e.g. learning rate and data augmentation). Rule-based parameters are chosen using the data fingerprint and known interdependencies between model parameters (e.g. patch size and batch size). Lastly, the optimal ensemble of trained models and post-processing are determined empirically with k-fold cross-validation.

For the abdominal cavity segmentation task, the open-source implementation of nnU-Net¹ was used. A 2D version of nnU-Net was trained with standard parameters and 5-fold cross-validation data split described in the section III-C for 1000 epochs. In addition, as custom post-processing, binary holes in the predicted segmentation were filled if present, since we know that the abdominal cavity is a single continuous object.

2) *Visceral slide quantification:* Now, the trained nnU-Net model allows inferring segmentation of the abdominal cavity for any frame of any sagittal cine-MRI slice. The next step of the method is the algorithm for visceral slide quantification. Our algorithm follows the steps of David Randall's approach to a large extent, however, the method to calculate the visceral slide is different. The scheme of the method is present in figure 4.

At first, we give a formal definition of a problem and necessary notations, which are then used to describe the algorithm. Let S be a cine-MRI slice with dimensions (W, H, T) , width, height and time respectively and M is a segmentation mask predicted for this slice with the same dimensions. Then, F_i is the i th frame of a slice and M_i is its segmentation mask, $i = 1, \dots, T$. $C_i = \{c_{il} = (x_{il}, y_{il}), l = 1, \dots, L_i\}$ is the abdominal cavity contour on the frame F_i , where L_i is the number of points in the contour.

Following David Randall's approach, we compute full deformation field D_{ij} between the moving frame F_i and the fixed frame F_j as a sum of two deformation fields that describe the motion of abdominal cavity contents and abdominal cavity surroundings separately. For that, segmentations M_i and M_j of these frames are used to mask frames and compute masked deformation fields. That yields the following formula for the full deformation field:

$$D_{ij} = D(F_i \odot M_i, F_j \odot M_j) + D(F_i \odot (1 - M_i), F_j \odot (1 - M_j)) \quad (1)$$

where the first term describes deformation inside the abdominal cavity and the second is for deformation of the

¹<https://github.com/MIC-DKFZ/nnUNet>

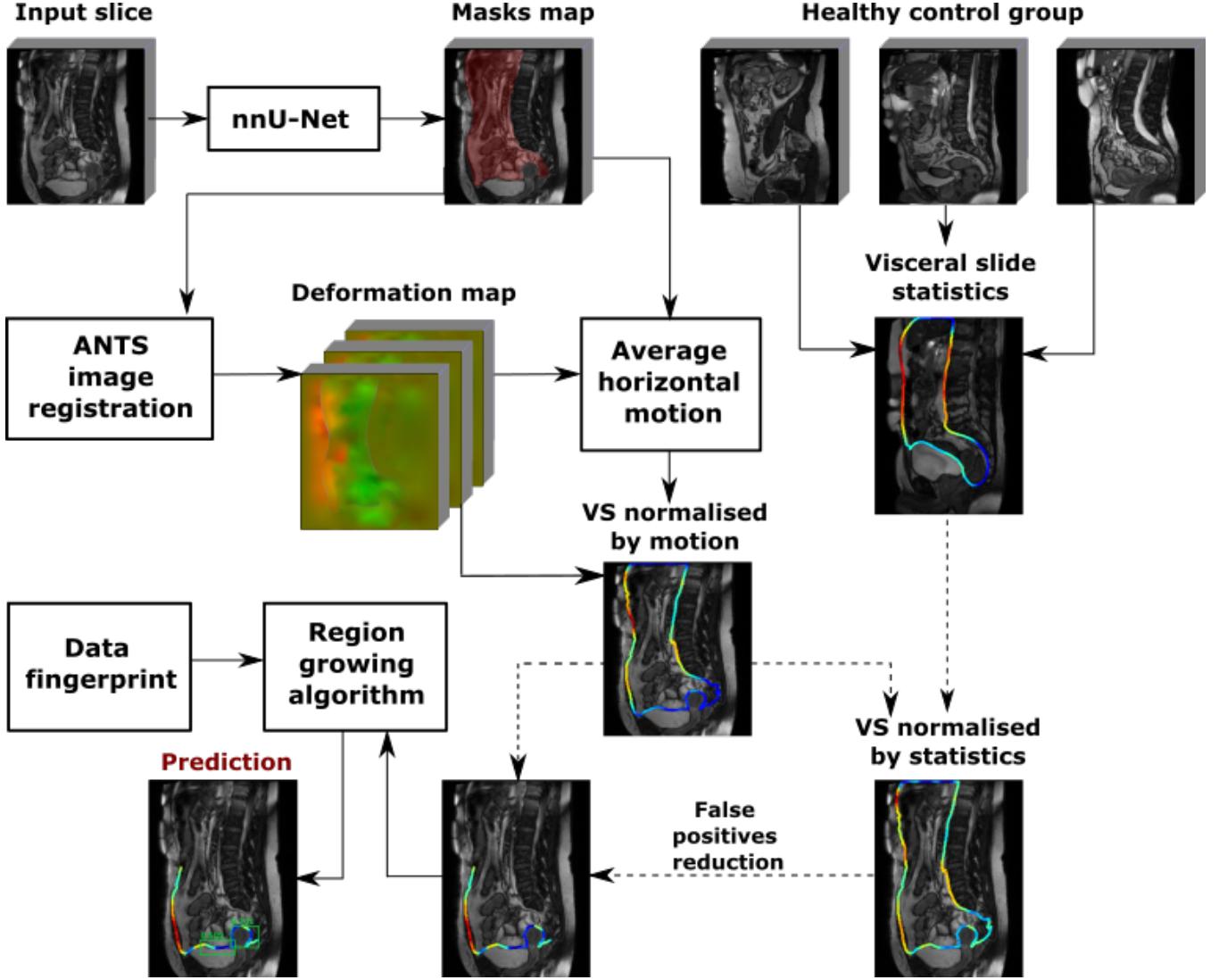


Fig. 3: The proposed fully-automated method for adhesion detection on cine-MRI. For an input slice segmentation mask is predicted by nnU-Net and used to compute a complete deformation field composed of deformation of abdominal cavity contents and deformation of its surroundings. Then, the visceral slide and average horizontal motion that occurs on the slice are computed from the segmentation mask and deformation field. Normalisation by average horizontal motion is always applied to the visceral slide to suppress differences in motion amplitude between patients. Optionally, the visceral slide can be normalised by visceral slide statistics in the healthy control group. After normalisation visceral slide is filtered by the position of the abdominal cavity contour at which it is located and only coordinates where adhesions can form are kept. This way false positives reduction is performed. A filtered subset of the visceral slide is passed to a region growing algorithm that predicts adhesions bounding boxes and assigns confidence. Some parameters of the algorithm that control bounding box growth are estimated from the data.

abdominal cavity surroundings. $D(\cdot, \cdot)$ refers to the deformation outputted by the image registration algorithm.

For image registration, the implementation from ANTS toolkit [11] was used. More details about the ANTS implementation and the selected registration parameters are given in the appendix VI-B.

Since the deformation field describes the transformation that should be applied to the moving frame, the visceral slide is computed along the abdominal cavity boundary on the moving frame. Therefore the visceral slide between

frames F_i and F_j is a function of the deformation field D_{ij} and the contour C_i of the moving frame F_i : $V_{ij} = g(D_{ij}, C_i)$. To compute visceral slide at each contour point $c_{il} = (x_{il}, y_{il})$, we take a clockwise tangent vector \vec{t}_{il} to the contour at this point, the corresponding points inside and outside of the abdominal cavity, $c_{il}^i = (x_{il}^i, y_{il}^i)$ and $c_{il}^o = (x_{il}^o, y_{il}^o)$ respectively, and calculate visceral slide as an absolute difference of dot products between \vec{t}_{il} and the deformation field at c_{il}^i and between \vec{t}_{il} and the deformation field at c_{il}^o . Hereby, the formula for visceral slide at the

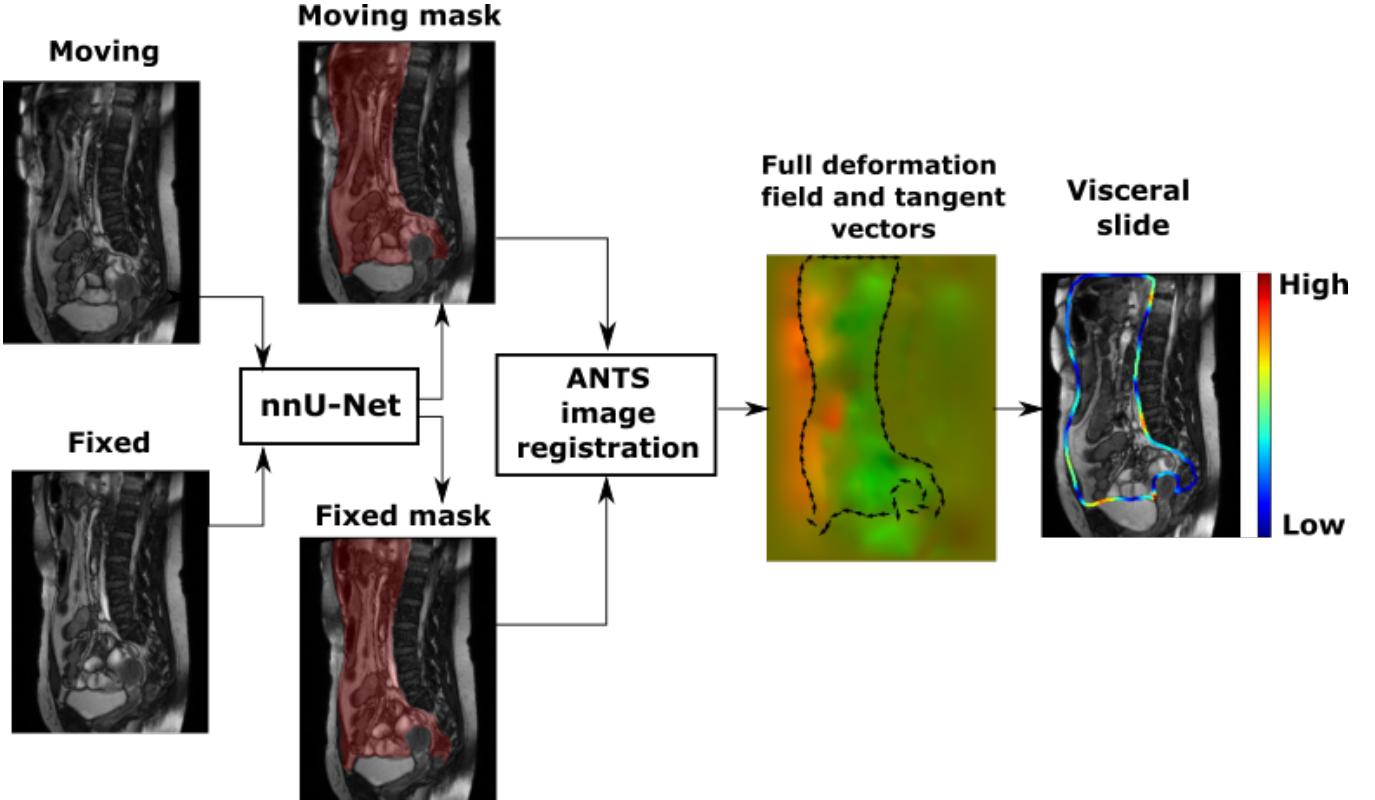


Fig. 4: Flowchart summarising visceral slide calculation between two frames F_i and F_j .

contour point c_{il} is:

$$v_{il} = \text{abs}(D_{ij}(c_{il}^i) \cdot \vec{t}_{il} - D_{ij}(c_{il}^o) \cdot \vec{t}_{il}) \quad (2)$$

The absolute value is taken because we are interested in the magnitude of visceral slide and not in its direction. Then, the visceral slide between the frames F_i and F_j can be expressed as $V_{ij} = \{(x_{il}, y_{il}, v_{il}), l = 1, \dots, L_i\}$, where (x_{il}, y_{il}) are the coordinates of the abdominal cavity contour on the frame F_i and v_{il} is the visceral slide value at (x_{il}, y_{il}) .

An advantage of this method in comparison to quantifying the visceral slide with shear along the abdominal cavity contour, as it is done in David Randall's approach, is that the computed visceral slide is in deformation field units and can be converted into millimetres. This makes its values more interpretable.

In David Randall's work, the visceral slide corresponding to a cine-MRI slice is computed as cumulative shear across all frames. For that, he developed a shear summation procedure that accounts for changes in abdominal cavity boundary position between frames. This summation procedure is applied to shears between each pair of the subsequent frames of a slice.

In this master thesis, we explore two possible ways to quantify the visceral slide that occurs on a cine-MRI slice. The first method is equivalent to David Randall's approach to compute cumulative visceral slide across all frames with an appropriate summation procedure. In another method, only

the two most dissimilar frames are used and the visceral slide computed between these frames is viewed as a sufficient approximation of the total visceral slide on a cine-MRI slice. The pair of the most dissimilar frames correspond to frames at the opposite phases of the respiratory cycle, therefore we call them inspiration and expiration frames.

3) Cumulative visceral slide: In this method information at all cine-MRI slice time points is exploited to compute the visceral slide. Visualisation of the algorithm is given in figure 5. Visceral slide $V_{i,i+1}$ is computed for each subsequent pair of frames F_i and F_{i+1} . To avoid cluttered notation we refer to this visceral slide as V_i onward. This way, $T - 1$ visceral slides are obtained for a slice with T frames.

The cumulative visceral slide is an average of visceral slides computed for all subsequent frame pairs. The main challenge in the summation of visceral slides is the mismatch of abdominal cavity positions between different frames. To address it, at each summation step we warp the contour corresponding to the current cumulative visceral slide to match the contour of the next visceral slide. Note that this way the contour of the cumulative visceral slide always matches the contour of the visceral slide that was added last. For warping of contours, the deformation field D_i^c between contours C_i and C_{i+1} is used. This deformation field is acquired by registering 2D images into which the contours are transformed. An image I_i representing the contour C_i is generated in the following way:

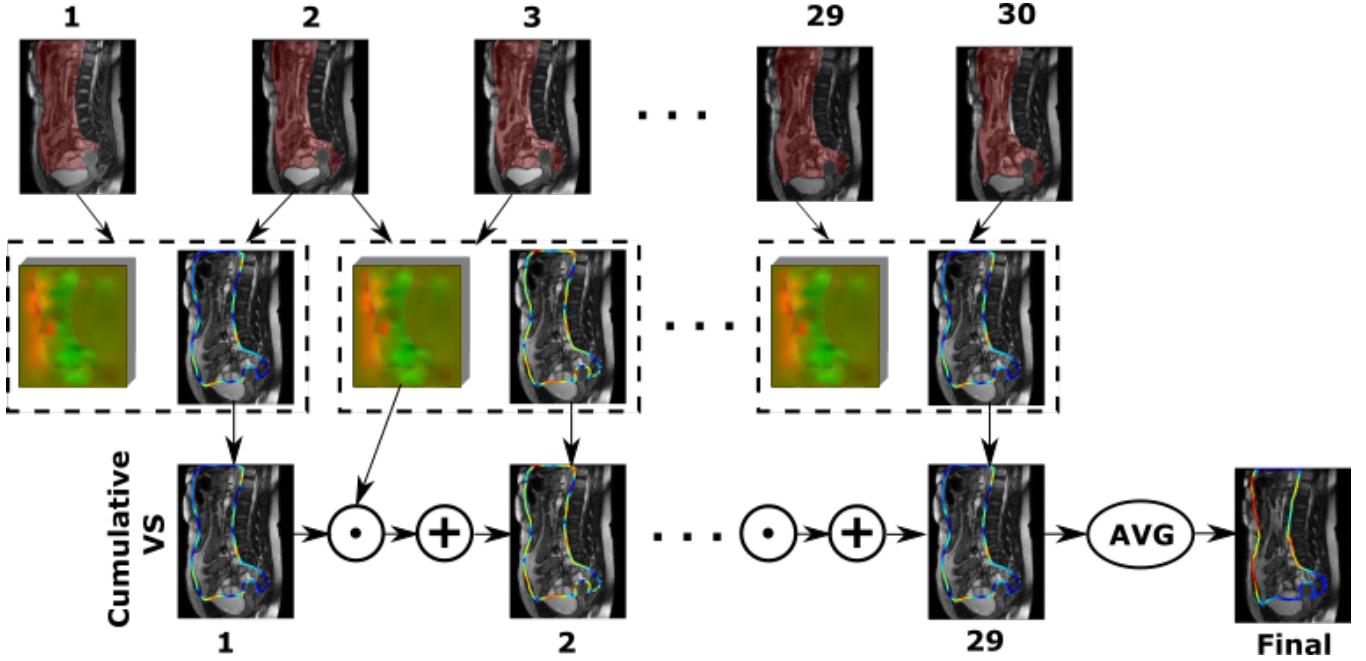


Fig. 5: A scheme that visualises summation of visceral slides across cine-MRI slice to compute the cumulative visceral slide.

$$I_i(x, y) = \begin{cases} c & \text{if } (x, y) \in C_i \\ 0 & \text{otherwise} \end{cases}$$

where c is an integer constant. This way only information about the location of contours is passed to the registration algorithm and the output is not affected by intensities differences. Then, the warped visceral slide is obtained as:

$$\begin{aligned} V'_i &= \{(x'_{il}, y'_{il}, v_{il}), \quad l = 1, \dots, L_i\}, \quad \text{where} \\ x'_{il}, y'_{il} &= \text{round}(x_{il} + u_{il}), \text{round}(y_{il} + u_{il}), \\ u_{il}, v_{il} &= F_i^c(x_{il}, y_{il}) \quad \text{deformation components at } x_{il}, y_{il} \end{aligned} \quad (3)$$

After warping, the cumulative visceral slide may contain points with duplicated coordinates and its contour length does not change. To resolve duplicated coordinates, the visceral slide at a particular coordinate is taken as an average of all visceral slide values that happen to have this coordinate after warping. Then, the contours of the warped cumulative visceral slide and the visceral slide between the next pair of frames still do not match. To add visceral slides, we iterate over each point of a current visceral slide, find the point of the warped cumulative visceral slide closest by euclidean distance and add visceral slide values at these points.

The detailed algorithm of visceral slide summation across frames is described below. V_{cum} denotes the cumulative visceral slide.

- 1) For $i = 1, \dots, T - 1$ compute visceral slides $V_i = \{(x_{il}, y_{il}, v_{il}), \quad l = 1, \dots, L_i\}$ between frames F_i and F_{i+1} and deformation field D_i^c between contours

C_i and C_{i+1} to warp the cumulative visceral slide during summation.

- 2) At the first step, take $V_{cum} = V_1$.
- 3) For $i = 2, \dots, T - 1$ add the cumulative visceral slide obtained at the previous step and the current visceral slide V_i . First, warp V_{cum} with the deformation field F_i^c according to the formula (3) to V'_{cum} . Then, add V'_{cum} and V_i with the following procedure. For $l = 1, \dots, L_i$:
 - a) Find the point $(x'_{cum,l*}, y'_{cum,l*}, v_{cum,l*})$ at V'_{cum} closest to (x_{il}, y_{il}, v_{il}) at V_i by euclidean distance so that $l^* = \arg \min_l (\sqrt{(x_{il} - x'_{cum,l'})^2 + (y_{il} - y'_{cum,l'})^2})$.
 - b) Take $v_{cum,l} = v_{il} + v_{cum,l*}$ as a new cumulative visceral slide value at the coordinate x_{il}, y_{il} .

Finally, the cumulative visceral slide is divided by the number of added visceral slides, that is $T - 1$, to keep it in deformation field units.

Potential advantages of this method are that complete temporal information is exploited and segmentation inaccuracies that may occur at different frames are smoothed during summation. The latter property makes this method more robust to errors a segmentation model makes.

- 4) *Visceral slide with inspiration and expiration frames:*
The main idea behind this method is that if only frames at the opposite phases of the respiratory cycle are used, the deformation that occurs in the abdomen during respiration is highlighted. Hence, regions of higher and lower visceral slide should be more prominent.

The first step of the methods is to find the inspiration and expiration frames F_{insp} and F_{exp} . This is done based on the position of the front abdominal wall since its anteroposterior

movement during respiration can be correlated to the phases of the respiratory cycle well. Therefore, the distance between the anterior wall positions will be the highest for inspiration and expiration frame pair. During inspiration, the diaphragm moves down creating a vacuum that pulls air into the lungs. As the diaphragm moves down, it pushes the abdominal contents down, which forces the abdominal wall to move out. Whereas, during expiration, the diaphragm relaxes and the abdominal muscles contract, causing the abdomen to flatten out. Hence, on sagittal slices, the position of the abdominal wall will be the most anterior for the inspiration frame and the most posterior for the expiration frame.

The position of the anterior abdominal wall on each frame is determined based on the motion perpendicular to the contour of the anterior wall with respect to a fixed frame. The last frame of a slice, F_T is set to be fixed. For each frame F_i of a cine-MRI slice the contour of the anterior abdominal wall $C_i = \{c_{il} = (x_{il}, y_{il}), l = 1, \dots, L_i\}$, vectors $O_i = \{\vec{o}_{il} = (u_{il}, v_{il}), l = 1, \dots, L_i\}$ orthogonal to the contour and deformation field D_{iT} between the current frame (moving) and the last frame (fixed) are determined. The horizontal motion along the anterior wall is calculated as a dot product of deformation at contour point and the corresponding orthogonal vector $H_i = \{h_{il} = D_{iT}(c_{il}) \cdot \vec{o}_{il}, l = 1, \dots, L_i\}$. Then, the motion component with the highest absolute value is found as $h_i^* = \arg \max_l(\text{abs}(h_{il}))$. Inspiration and expiration frames indices are determined as $\arg \min_i(\text{abs}(h_i^*))$ and $\arg \max_i(\text{abs}(h_i^*))$ respectively, since inspiration frame will have the highest displacement to the left in comparison to the fixed frame and the expiration frame will have the displacement to the right. Note, that the choice of the fixed frame to perform registration is not important since we look at the relative displacement.

Then, visceral slide is computed between F_{insp} and F_{exp} as described in the section III-D.2. The inspiration frame is taken as a moving image and the expiration frame a fixed image. Hence, the position of the computed visceral slide matches the abdominal cavity contour on the inspiration frame.

Examples of cumulative and inspiration/expiration visceral slides computed for the same slice are present in figure 7. These examples demonstrate that our expectations about differences in the two approaches to visceral slide computation were valid. Generally, the cumulative visceral slide is smoother and represent the total abdominal motion captured on a cine-MRI slice better: the estimated visceral slide tends to be higher along the anterior and posterior abdominal walls, where it is expected to be maximum, and lower in the pelvis area, where less motion occurs. On the contrary, there are more local changes in inspiration/expiration visceral slide and the magnitude of values correlate with the location at abdominal cavity contour less. Nevertheless, some similarities can often be spotted between cumulative and inspiration/expiration visceral slides computed for the same slice. Also, it is noticeable from the figure that the range of both cumulative and inspiration/expiration visceral slides can

substantially vary for different slices and the possible difference is more drastic for the inspiration/expiration method. Among the given examples, the lowest cumulative visceral slide with the maximum value just under 0.35 is estimated for the first cine-MRI slice and the highest visceral slide with the maximum just over 5.0 for the last one. As for the inspiration/expiration visceral slide, it is the lowest for the first slice too (maximum over 1.75) and highest for the third example (maximum over 14.0). Overall, the difference in visceral slide values distributions between the two methods is striking.

Histograms summarising cumulative and inspiration/expiration visceral slides distributions are given in figure 6.

5) Visceral slide normalization by average motion: Examples of both cumulative and inspiration/expiration visceral slides given in the previous section reveal that the magnitude of the visceral slide can substantially vary across patients. This is because despite receiving the same instructions during a cine-MRI scan acquisition, different patients move with different amplitude. Some patients exhibit high amplitude motion, whereas others show little motion. This affects the magnitude of the visceral slide, which is possible to measure from a cine-MRI slice. Due to the character of abdominal motion during respiration, the higher motion amplitude implies a higher degree of the visceral slide that can occur. These differences hinder the design of automated adhesion detection method because the interpretation of a particular visceral value is patient dependent.

We mitigate these motion differences between patients by normalising the computed visceral slide by the average horizontal motion along the anterior abdominal wall. As discussed in the section III-D.4, this motion represents the overall abdominal motion during the respiration the best. After the abdominal cavity contour C_i of a moving frame F_i and deformation field D_{ij} between a pair of frames are obtained, they are used to calculate the motion we are interested in, as described above: $H_i = \{h_{il} = D_{ij}(c_{il}) \cdot \vec{o}_{il}, l = 1, \dots, L_i\}$. Then, the visceral slide V_{ij} between frames is normalized by division at $h_i = \text{avg}_l(h_{il})$.

Examples of visceral slides normalized by the average horizontal motion are shown in figure 9. The same cine-MRI slices as in figure 7 are selected. Histograms of normalized visceral slide distribution given in figure 8 reveal that for both cumulative and inspiration/expiration visceral slides the distribution of normalized values is more condensed than the distribution of raw values. Besides, the normalisation made ranges of cumulative and inspiration/expiration visceral slides more similar, however, distributions are still noticeably different. The majority of inspiration/expiration visceral slide values are close to 0, whereas in cumulative visceral slide share of very small values is much smaller.

Overall, for both methods the range of the normalised visceral slide is more similar for different scans, however, if the average horizontal motion is low, high values might appear due to normalization. This explains the presence of larger upper outliers in the distribution of the normalized

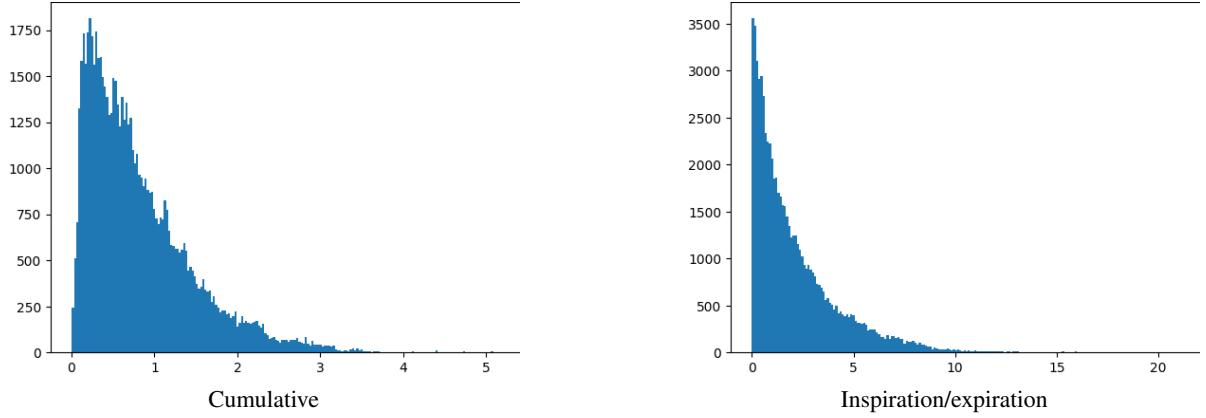


Fig. 6: Distribution of unnormalised cumulative and inspiration/expiration visceral slide (mm) in the training set.

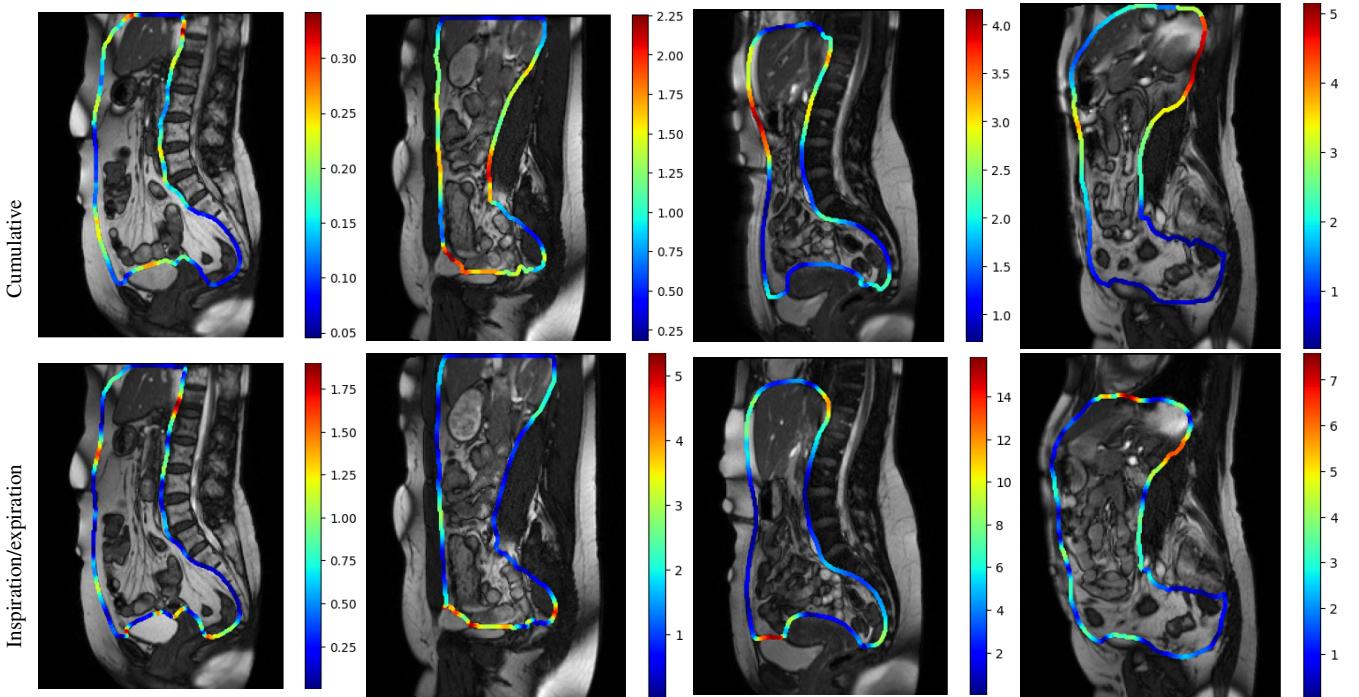


Fig. 7: Examples of raw cumulative and inspiration/expiration visceral slides computed for the same cine-MRI slice. Visceral slide values are converted to millimetres. The cumulative visceral slides are placed in the top row and inspiration/expiration visceral slides are in the bottom. Different frames are used to display cumulative and inspiration/expiration visceral slides since the position of cumulative visceral slide corresponds to the one before the last one frame, whereas the position of inspiration/expiration visceral slide corresponds to the inspiration frame.

cumulative visceral slide. Also in figure 9 this effect is visible for the inspiration/expiration visceral slide computed for the first cine-MRI slice.

Note that in further sections it is implied that visceral slide is always normalised by the average horizontal motion along the anterior wall even if it is not mentioned explicitly.

6) Visceral slide normalization by healthy control statistics: As mentioned before, the quantified visceral slide is generally higher along the abdominal cavity walls and lower in the pelvis area. If we want to predict adhesions, we have to account for the difference in typical visceral slide value in different regions of abdominal cavity contour. Otherwise,

it is likely that the presence of adhesions in the pelvis area will be overestimated and we will miss adhesions along the anterior abdominal wall.

A possible way to adjust for these differences is to normalise the visceral slide by the statistics of the healthy control group. Sampling of the healthy control group is described in the section III-B.2. For each cine-MRI slice, visceral slide normalized by the average horizontal motion is computed and then, its mean and standard deviation are calculated for different regions of abdominal cavity contour. These means and standard deviations are used to normalize visceral slides computed for cine-MRI slices in training and

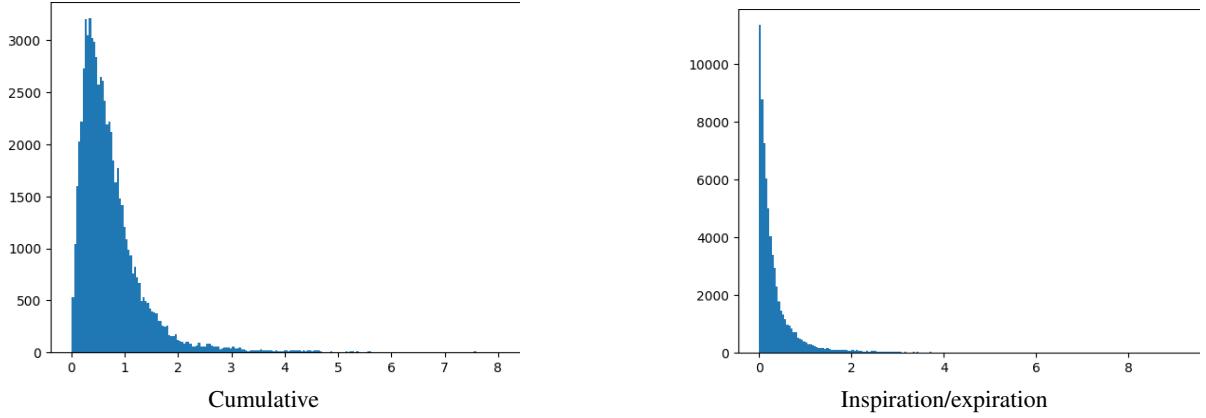


Fig. 8: Distribution of cumulative and inspiration/expiration visceral slide normalised by the average horizontal motion along the anterior abdominal wall in the training set.

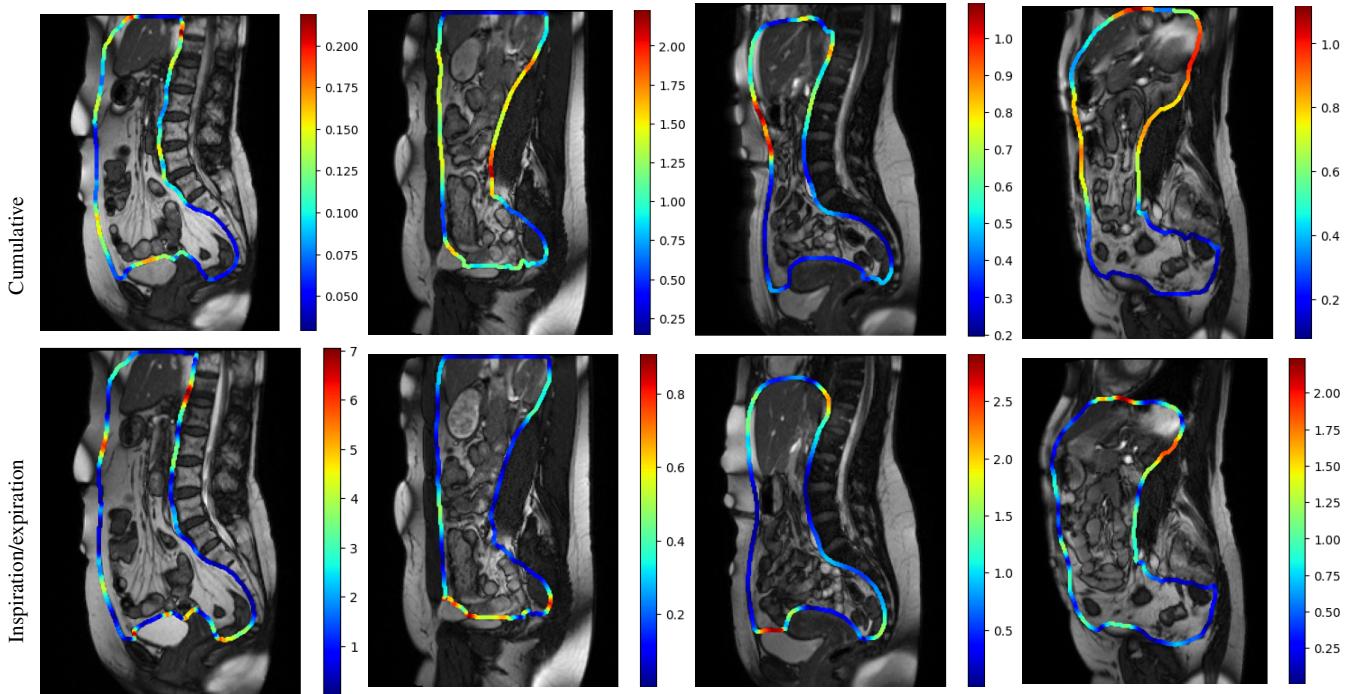


Fig. 9: Examples of cumulative and inspiration/expiration visceral slides normalised by the average horizontal motion along the anterior wall computed for the same cine-MRI slice. The cumulative visceral slides are placed in the top row and inspiration/expiration slides are in the bottom.

test sets.

Since the shape and position of the abdominal cavity highly vary per patient, calculation of the visceral slide statistics by coordinate is infeasible. Instead, contours C_i are spitted into coarse chunks P_{ik} , matched by their relative location. The bottom-left point of the abdominal cavity contour is used as a starting point and the chunks are obtained by moving clockwise along the contour by a next chunk length until the starting point is reached. The lengths of chunks are determined based on the contour length and the number of chunks a contour is split into, which in turn is obtained from the average contour length in the healthy control group (595) and the desired average number of points

per chunk, which we set to 5. That yields 119 unique chunks per contour. When all visceral slides in the healthy control group are split into chunks, visceral slide expectation and standard deviation are computed per chunk for the set of the values which fall into this chunk in all visceral slides. The formal description of the algorithm is the following:

- 1) Compute visceral slides V_i for all cine-MRI slices in the control group, $i = 1, \dots, N$.
- 2) Split each V_i into chunks with the following procedure. Compute length of chunks q_{ik} from the contour length L_i and the target number of chunks so that $\sum_{k=1}^{119} q_{ik} = L_i$. Find the bottom left point (x_{il^*}, y_{il^*}) on the contour C_i . Starting from (x_{il^*}, y_{il^*}) split V_i

- into 119 chunks of the corresponding lengths q_{ik} :
 $P_{ik} = \{v_{il}, l = 1, \dots, q_{ik}\}$
- 3) For each chunk index k , iterate over visceral slides and collect values that fall into it: $P_k = \{v_l, l = 1, \dots, \sum_{i=1}^N q_{ik}\}$
 - 4) For each P_k compute mean and standard deviation.

To normalize a visceral slide with the statistics computed this way, it is split into 119 chunks in the same way and the values in each chunk are normalized with statistics of this chunk. We consider two normalization options:

- 1) **Division by mean:** visceral slide values are divided by mean. The resulting value highlights how visceral slide in a particular region is related to its expected value. Small values indicate less visceral slide than expected and therefore higher likelihood of adhesion in this region.
- 2) **Standardisation:** we subtract mean from visceral slide values and divide by standard deviation. If the distribution of the visceral slide is sufficiently close to normal, we can interpret the resulting values based on the properties of standard normal distributing. Then, conceptually negative outliers indicate adhesions. Or, perhaps we can be less strict about the deviation from the mean required to classify a point as belonging to adhesion and consider all negative values as an indication of a possible adhesion.

Normalization with standardisation requires normality of the visceral slide distribution, however, it is clear from figure 8 that it does not hold. For both cumulative and inspiration/expiration visceral slide normalised by the average horizontal motion, the distribution is heavily skewed to the right and many upper outliers are present. Therefore, to compute sensible statistics, outliers were removed and suitable transformation of visceral slide values was picked. For both cumulative and inspiration/expiration methods squared root transformation worked best. Figure 10 shows the comparison of distributions of raw values of cumulative visceral slide in the control group versus visceral slide with removed outliers and squared root transformation applied. The distribution of transformed values resembles normal distribution rather well. Distribution of cumulative visceral slide in the training set normalised with standardisation is depicted in figure 11 with and without squared root transformation. It is clear from the figure, that distribution of squared root transformed values is much closer to standard normal. For inspiration/expiration visceral slide the trend is similar and is depicted in figures 18 and 19 in the appendix.

Hence, eventually, the visceral slide statistics from a healthy control group is calculated with removed outliers and squared root transformation is applied for both normalisation options. The visceral slide expectation by chunks is depicted in figure 12.

Examples of visceral slides normalised with division by expectation and standardisation (after squared root transformation) are given in figures 21 and 22 in the appendix respectively. Normalisation by the average horizontal motion

is applied prior to normalisation with statistics and the same slices are used as before.

E. Detection of adhesions

To predict adhesions based on the computed visceral slide a region growing algorithm with false positives reduction was developed. The algorithm relies on a few keys assumptions:

- 1) Low visceral slide correlates with the likelihood of adhesion presence.
- 2) The computed visceral slide normalised by the average horizontal motion is comparable among the slices. That is, a particular visceral slide value should have the same interpretation if it belongs to any cine-MRI slice.
- 3) Visceral slide normalised with control group statistics is location-agnostic. In other words, the interpretation of a particular visceral slide value should be independent of its location at the abdominal cavity contour.

1) *Region growing from local minimum:* The algorithm has a few hyper-parameters, which were tuned with 5-fold CV for different experiments if the opposite is not stated in the description:

- *Connectivity threshold* - a maximum distance between two contour points to be treated as belonging to a continuous region at the abdominal cavity contour. Set to 5 for all experiments.
- *Region growing index, RGI* - a constant that is used to determine maximum visceral slide value to be included into the current adhesion prediction based on the visceral slide value from which region growing started. The optimal value depends on the distribution of the visceral slide.
- *Minimum region length, MRL* - minimum number of points added to the region to be treated as a prediction. This parameter aims to prevent the method from generating predictions from visceral slide artefacts (e.g. a single very low visceral slide value surrounded by much higher values). Conceptually we assume that the visceral slide should be lowered in a sufficiently long region to indicate an adhesion.

Two other parameters were derived from data:

- *Limits of the predicted bounding box:* inferred from the statistics of annotations with bounding boxes. Minimum and maximum sizes are taken as the mean bounding box size minus/plus 1.96 standard deviations.
- *Visceral slide range for prediction:* visceral slide tends to have upper outliers. Although some of them might be method artefacts, conceptually it means that adhesions cannot be present in such regions. In both cases, it is sensible to filter out visceral slide regions containing upper outliers. The threshold for upper outliers is defined as $V_{max} = Q_3 + 1.5IQR$, where $IQR = Q_3 - Q_1$ is an interquartile range and Q_1 and Q_3 are the first and the third quartiles. These values are computed for the set of visceral slide values gathered from the whole training set.

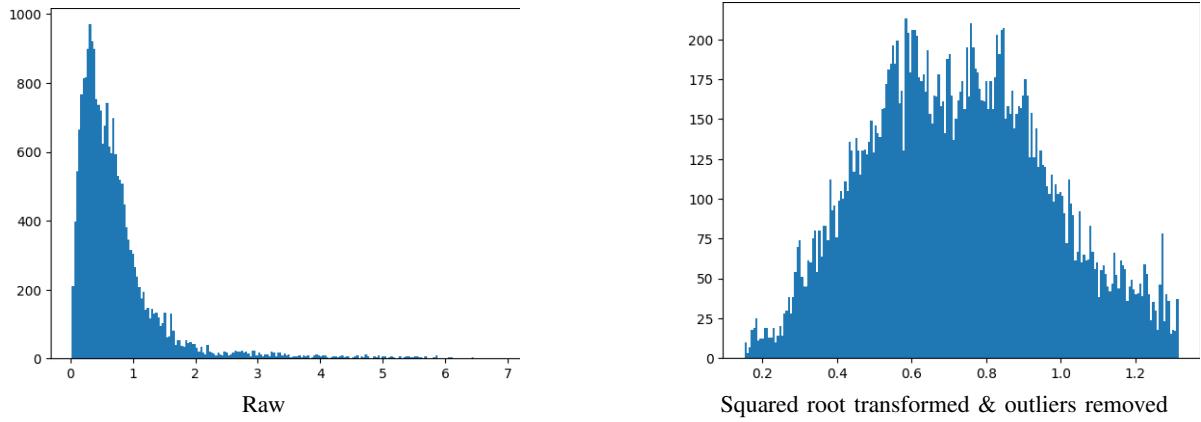


Fig. 10: Distribution of raw and squared root transformed values of cumulative visceral slide normalised by the average horizontal motion computed for the healthy control group. In addition to squared root transformation outliers were removed.

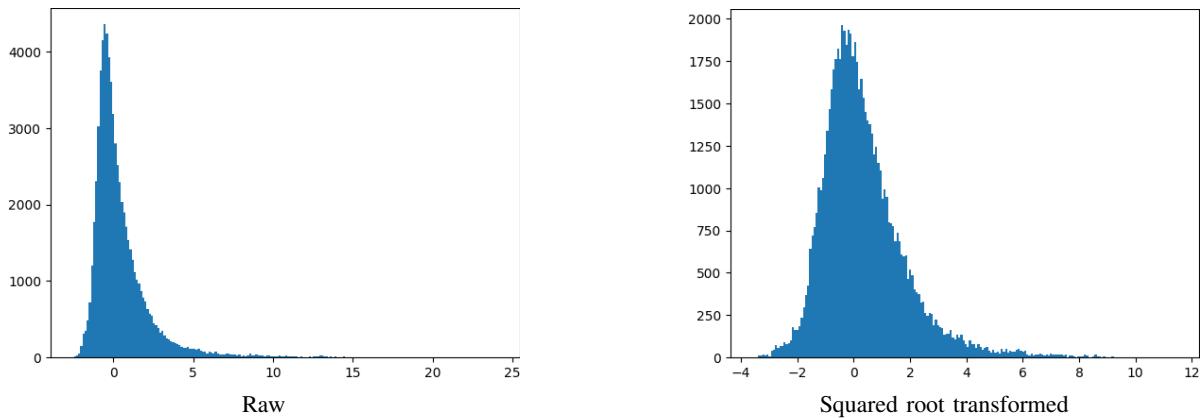


Fig. 11: Distribution of raw and squared root transformed cumulative visceral slide standardised with the control group statistics.

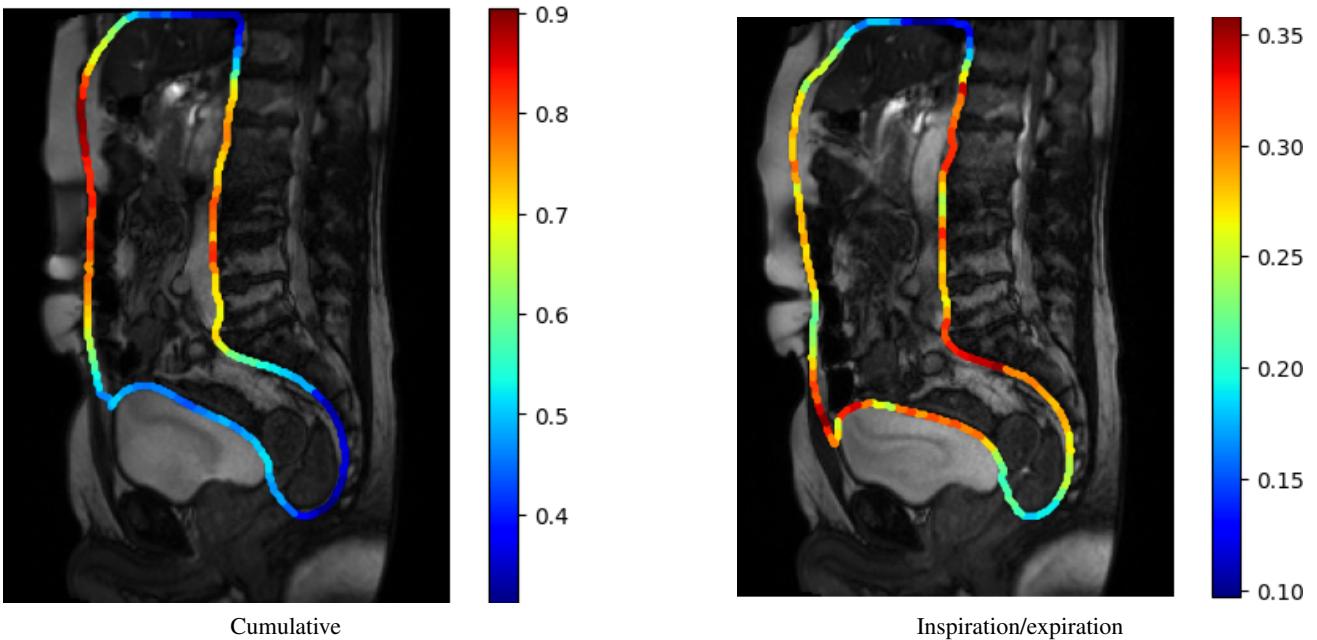


Fig. 12: Expectation of visceral slide normalised by average horizontal motion visualised on the arbitrary frame of a randomly drawn cine-MRI slice.

The method's input is a set of visceral slides $V_i = \{(x_{il}, y_{il}, v_{il}), l = 1, \dots, L_i\}$ computed for cine-MRI slices S_i in the training set. The following algorithm is applied to each visceral slide to predict adhesions:

- 1) Filter out visceral slide outliers: visceral slide points with $v_{il} > V_{max}$ and join the remaining values into connected regions $R_{ij} = \{(x_{jl}, y_{jl}, v_{jl}), l = 1, \dots, L_j\}$ based on the connectivity threshold. Regions which size is smaller than the lower limit of bounding box size are discarded.
- 2) Find global visceral slide minimum across all regions $v_{min} = \arg \min_{jl}(v_{jl})$ and the index j^* of a region that contains it. The corresponding point is used as a starting point for region growing, which is done point by point in both directions. Region growing stops when at least one of the following conditions is reached: a) the value of visceral slide v_{jl} to be added exceeds the maximum determined for the region as $v_{max} = v_{min} * RGI$, b) the size of the bounding box that encloses the region exceeds the upper bounding box size for the prediction, c) an edge of the regions is reached and there are no points left to add. The output of this process is a connected region $Q_k = \{(x_{kl}, y_{kl}, v_{kl}), l = 1, \dots, L_k\}$, which is a subset of R_{j^*} . Based on the coordinates of points in Q_k , an adhesion prediction with bounding box is generated. If the size of a generated bounding box is smaller than the lower size limit, the bounding box and Q_k are enlarged to this size.
- 3) Q_k is cutout from R_{j^*} , which produces two new visceral slide regions $R_{j^*}^1$ and $R_{j^*}^2$ before and after Q_k respectively. $R_{j^*}^1$ and $R_{j^*}^2$ are added to the set of connected regions for prediction if their size is larger or equals to the required minimum bounding box size.
- 4) Steps 2) and 3) are repeated until no suitable regions R_{ij} are left.

Note, that this algorithm assumes that all visceral slide values are positive, which is not the case when normalisation by control group statistics with standardisation is applied. For this normalisation option, a few adjustments to the method are made. First, instead of filtering out upper outliers, only negative visceral slide values are left to form regions. Also, the maximum visceral slide value for the current prediction is defined as $v_{max} = v_{min}/RGI$.

2) *False positives reduction:* By design, the method outputs predictions along the whole abdominal cavity contour. This results in a large number of false positives. It is straightforward to substantially reduce the number of false positives by exploiting the knowledge about possible locations of adhesions, which are the bottom half of the anterior abdominal wall and pelvis area. Therefore, the top half of the anterior wall, the top of the abdominal cavity contour and the entire posterior wall can be removed from the input visceral slides. To achieve this, the algorithm which detects four parts of the abdominal cavity contour (anterior wall, posterior wall, contour top and contour bottom) was developed. As

soon as the location of these parts is determined, irrelevant areas of the abdominal cavity contour can be easily filtered out. Examples of the detection of abdominal cavity parts for different slices are given in figure 23 in the appendix.

3) *Ways to output confidence for a prediction:* Taking into account the properties of the computed visceral slide and the designed region growing algorithm, three ways to output confidence were considered:

- 1) Based on the minimum visceral slide value (v_{min}) inside the predicted bounding box, that is the value from which region growing was started.
- 2) Based on the mean of visceral slide values (v_{mean}) inside the predicted bounding box.
- 3) Prediction of logistic regression trained on the training subset in each cross-validation fold. Bounding box annotations were used as positive examples and negative examples were sampled from negative slices.

To make different confidence values more interpretable and comparable, confidence defined with minimum or mean value is calculated with the formula $c_k = \frac{V_{max}-v_{min}}{V_{max}-V_{min}}$ (similarly for v_{mean}), where V_{max} is the upper outliers threshold and V_{min} is the global visceral slide minimum in the training set. This way higher confidence is assigned to regions with lower visceral slide and its range is $(0, 1)$.

Sampling of negative examples is performed only in the visceral slide region considered for prediction. A center of a bounding box is randomly chosen at the region and bounding box size is drawn from a normal distribution with mean and standard deviation equal to those in bounding box annotations. To obtain enough negative examples, four different bounding boxes were sampled from each negative visceral slide. The following features were tried: bounding box weight and height, visceral slide minimum, mean and maximum in the region, length of the region. Among these features only bounding box height and region length were consistently significant across all folds in the model for different ways to calculate and normalise visceral slide. As explained in the section IV-C they were essential to making the model learn to differentiate between true and false positives. Also, the addition of bounding box width or mean visceral slide value increased the performance of the method.

F. Validation

The method was evaluated from the object detection perspective, that is prediction of adhesions, and as a binary classifier of slices into positive and negative.

For adhesion detection, FROC analysis [12] was a primary evaluation method and average precision (AP) [13] was calculated as an auxiliary metric that summarises method performance. The intersection over union (IoU) threshold to register a hit is set to 0.01. Such a low value is chosen because many bounding box annotations we have are relatively small and clinically, the area affected by adhesion is larger than an adhesion itself. Besides, since it is unknown which frame exactly an annotation corresponds to, there is additional variation due to motion.

A simple way to classify a cine-MRI slice with respect to adhesion presence is to use the highest confidence value among all predicted adhesion bounding boxes as confidence for binary classification of a slice. Then, slice-level ROC and AUC can be computed. One more way to access whether there is a difference in confidences assigned to true and false positives is checking whether average confidences of predictions that fall into these groups are statistically different. The 0.05 significance level was used.

For the training set, all evaluation metrics were computed with 5-fold cross-validation described in the section III-C.

G. Documentation and source code

The method is implemented in Python 3.8 and documented with Numpy docstring format. The source code is located in the private repository of DIAG organisation².

IV. RESULTS

A. Abdominal cavity segmentation with nnU-Net

The standard nnU-Net parameters and simple post-processing yielded reasonably high performance and good segmentation. Cross-validated dice score and Jaccard index were 0.94 and 0.90 respectively. A few examples of good segmentation are given in figure 13. On these frames, abdominal cavity content is accurately segmented out of its surroundings at different positions of sagittal slices. However, rather often the predicted segmentation is less accurate. Most commonly, the model struggles with the pelvis area, where parts of the uterus and/or the rectum are often mistakenly included in the abdominal cavity. Sometimes parts of the bladder can be included as well. Also, difficulties with delineation of the posterior wall and differentiating the organs belonging to the thoracic cavity are common. Nevertheless, precise segmentation in these areas is not important for the algorithm, since it is known that adhesions cannot form in these areas. Generally, segmentation of the front abdominal wall, which is crucial for the task, is very accurate, however, there are a couple of examples on which the model fails, likely due to unusual intensities in the surrounding areas or artefacts. Examples of the listed errors are depicted in figure 24 in the appendix.

The performance of nnU-Net is not evaluated separately on the held-out test set because the goal of this work is the detection of adhesions, not the perfect segmentation of the abdominal cavity.

B. Adhesion detection

In the performed experiments we compare two ways to compute visceral slide (cumulative or inspiration/expiration) and investigate an impact of different normalization by control group statistics (none, division by expectation, standardisation) and a way to output confidence (based on visceral slide min, mean and with logistic regression). This yields 18 different experiments. Initially, the algorithm was run and evaluated for the whole training set.

²<https://github.com/DIAGNijmegen/abdomenmrus-cinemri-cavity-segmentation>

A complete overview of experiments with final hyper-parameters and evaluation metrics is given in table VI in the appendix. For each configuration of the method with respect to the ways to compute and normalise visceral slide, training logistic regression to predict confidence worked best³. For 4 of 6 methods it boosts AP and in 5 of 6 methods slice-level AUC is slightly improved. Also, in 5 of 6 cases mean TPs confidence is significantly higher than mean FPs confidence at 0.05 significance level, whereas when minimum or mean visceral slide value is used to output confidence usually there is no statistically significant difference between TPs and FPs confidence.

FROCs and slice-level ROCs for experiments with logistic regression are depicted in figure 14 and quantitative evaluation metrics are given in table III. From both visual comparison of methods and quantitative metrics, it is apparent that for the adhesion detection task cumulative visceral slide gives consistently better performance than inspiration/expiration visceral slide. At FROCs of methods that use cumulative visceral slide, sensitivity increases faster and AP is higher as well.

TABLE III: Evaluation metrics for experiments with cumulative and inspiration/expiration visceral slide and different normalisation by control group statistics. In all experiments logistic regression is used to output confidence of a predicted adhesion.

Method		Evaluation metrics		
Visceral slide type	Normalised by statistics	AP	Slice AUC	TPs and FPs different
Inspiration/expiration	No	0.1283	0.4818	No
	Division by expectation	0.213	0.5375	Yes
	Standardisation	0.1701	0.5735	Yes
Cumulative	No	0.284	0.5027	Yes
	Division by expectation	0.3198	0.5336	Yes
	Standardisation	0.2073	0.5008	Yes

Also, for both methods of visceral slide calculation, normalisation with division by visceral slide expectation yields the best FROCs. Among the compared methods, cumulative visceral slide normalised with division by expectation has the best performance based on its FROC and average precision (0.32). Sensitivity around 0.6 is reached at 1 FP per image and it rises to 0.73 at 2 FPs per image. The maximum sensitivity of 0.92 is achieved at 3.13 FPs per image. Nevertheless, the overall performance of the best method is rather poor and is far from being suitable for usage in clinical practice.

Slice-level ROCs reveals that neither of explored methods is able to differentiate between slices with adhesions and negative slices. Only ROC of standardised inspiration/expiration visceral slide looks as if the method was better than randomly

³It is noteworthy that features used to fit logistic regression were determined with cross-validation and vary between methods. For each method, the list of used features is given in the caption of figure VI

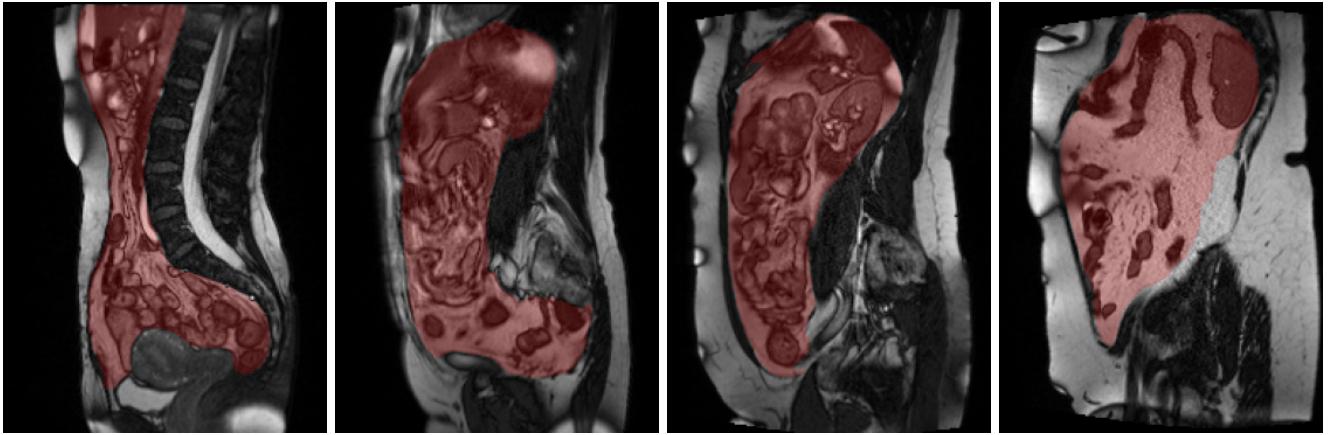


Fig. 13: Examples of very good abdominal cavity segmentation predicted by nnU-Net model for different sagittal slices positions.

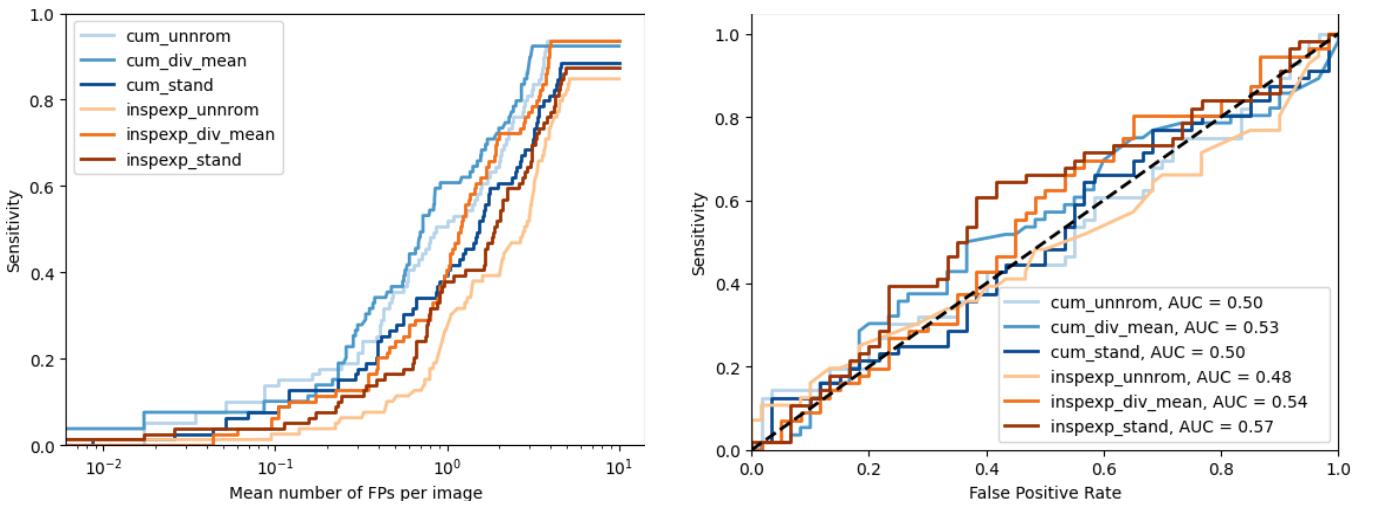


Fig. 14: FROCs (on the left) and slice-level ROCs (on the right) for experiments with cumulative and inspiration/expiration visceral slide and different normalisation options. Contractions used in legend labels: cum = cumulative, inspexp = inspiration/expiration, unnorm = no normalisation by control group statistics applied, div_mean = normalisation with division by expectation, stand = normalisation with standardisation. Curves for cumulative visceral slide are displayed in blue and for inspiration/expiration in orange.

guessing whether adhesions are present. By and large, ROCs of different methods are not much different from the $x = y$ line.

FROCs and slice-level ROCs obtained for different methods highlight that there is a difference in performance gain from the usage of logistic regression to predict confidence for cumulative and inspiration/ expiration visceral slides. In figure 15 FROCs and slice-level ROCs of method configurations that use unnormalised cumulative and inspiration/ expiration visceral slides and vary by the way to output confidence are depicted. For cumulative visceral slide, logistic regression gives noticeably higher sensitivity at almost all levels of FPs per image and there is a slight improvement in slice-level AUC. On the contrary, there is hardly any difference in methods' performance when confidence is derived from minimum or mean visceral slide value in the bounding box. For inspiration/expiration visceral slide, the difference

between methods' performance for different ways to output confidence is negligible.

The trend is similar for other normalisation options, for which FROCs and slice-level ROCs are visualised in figures 25 and 26 in the appendix. The only exception is the inspiration/expiration visceral slide normalised with division by expectation. For this configuration logistic regression noticeably improves both FROCs and slice-level ROCs as well.

Since even the performance of the best method was poor, we decided to look into possible reasons for that. To check whether differences in abdominal motion between these regions can explain low performance, the method was applied separately to the anterior wall and pelvis areas.

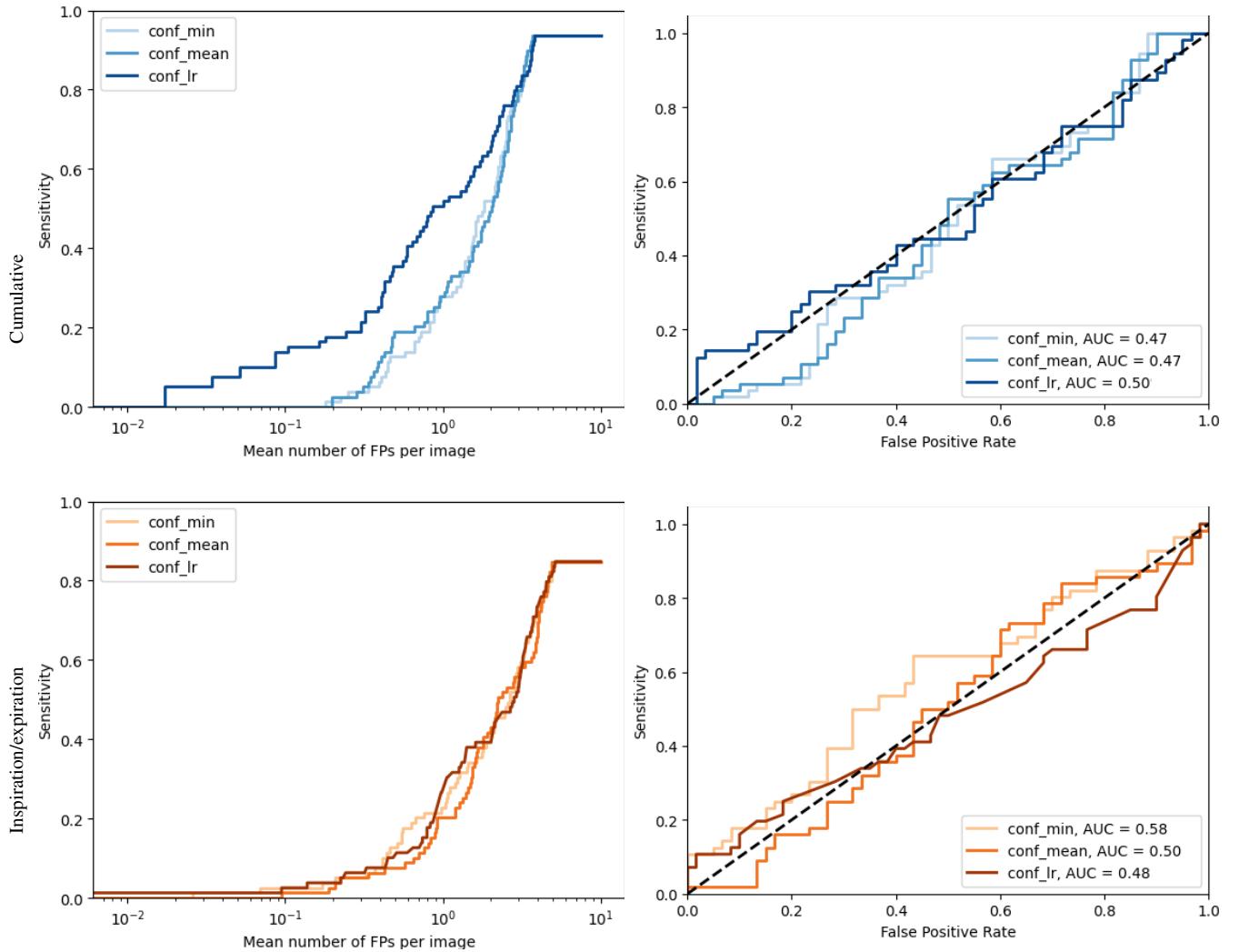


Fig. 15: FROCs (on the left) and slice-level ROCs (on the right) for experiments with unnormalised cumulative (top row) and inspiration/expiration (bottom row) visceral slide plotted separately for different ways to output confidence. Contractions used in legend labels: `conf_min` = confidence based on minimum visceral slide value, `conf_mean` = confidence based on mean visceral slide value, `conf_lr` = confidence predicted by logistic regression.

C. Anterior wall and pelvis separately

Only method configurations that use cumulative visceral slide were evaluated for the anterior wall and pelvis areas separately, since it had previously shown better performance. In these experiments only positive slices with annotations of a considered type (attached to the anterior wall or at the pelvis boundary) were included in the training set and the size of the negative slices subset was adjusted accordingly to make the dataset balanced. As a result, the datasets for the anterior wall and pelvis areas comprised 36 and 96 slices respectively. During prediction only the region of interest, that is the bottom half of the anterior wall or pelvis, was used to generate prediction.

Figure 16 shows comparison of FROCs and slice-level ROCs for separate experiments. For different method configurations the way to output confidence that gives the best performance is chosen as explained in the caption of figure 16. Quantitative metrics of all experiments are given in tables

VII and VIII in the appendix.

A separate evaluation gives a few interesting insights. First of all, for both the anterior wall and pelvis, logistic regression does not yield a performance boost in FROC and AP anymore. Most likely, it happens because there are more annotations in the pelvis area than along the anterior wall in the dataset, 65 versus 18. Then, when the method is used for these two areas simultaneously, logistic regression just learns to differentiate predictions in these areas by the shape of the bounding box and assigns higher confidence for predictions in the pelvis area. This is possible because in all models bounding box height was used as a feature in logistic regression. Moreover, this feature was significant across cross-validation folds, which indicates its importance for the prediction.

The evaluation metrics reveal a striking difference in properties of the visceral slide estimated along the anterior wall and in the pelvis area. For the pelvis area, slice-level AUC

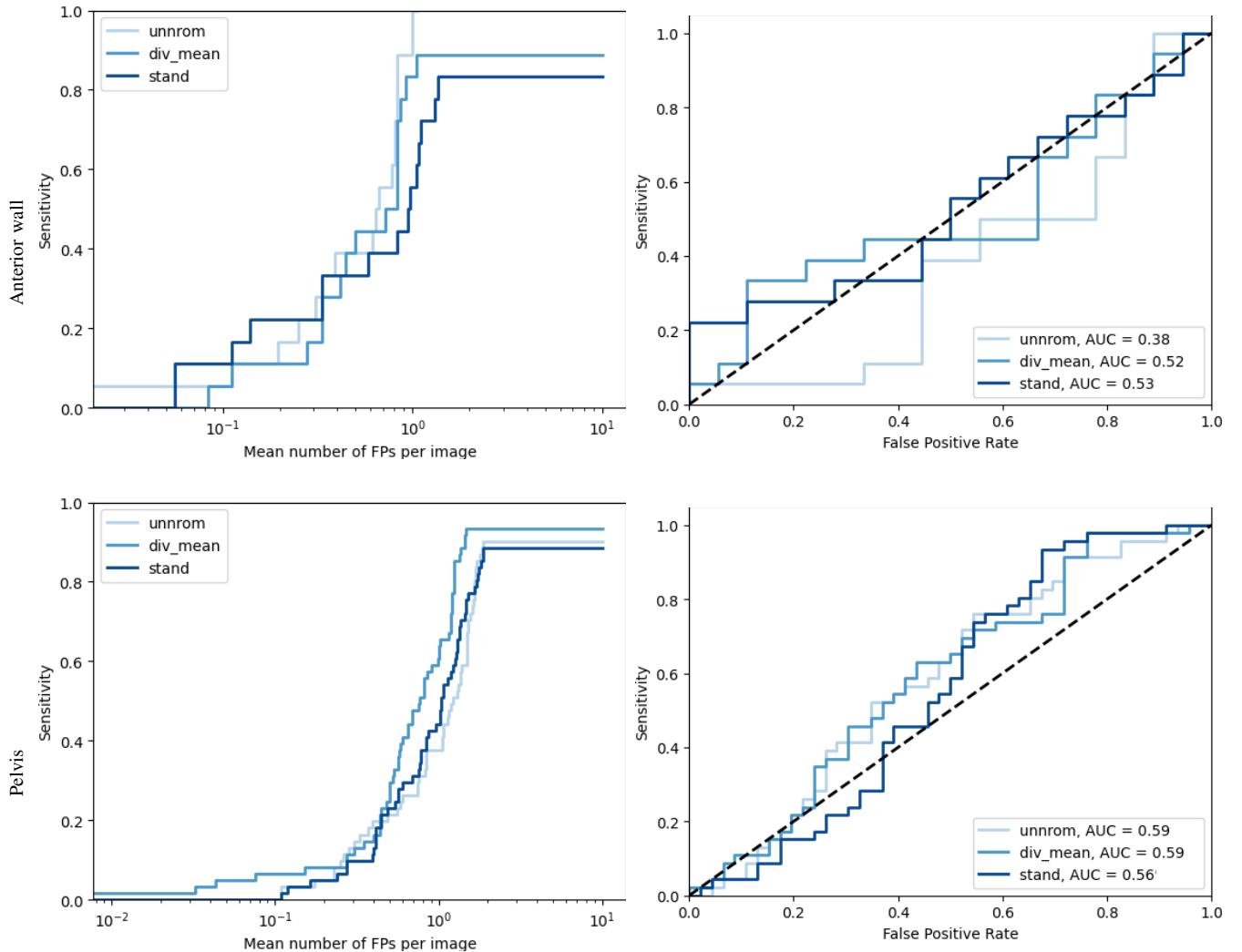


Fig. 16: FROCs (on the left) and slice-level ROCs (on the right) for the best experiments when anterior wall (top row) and pelvis areas (bottom row) are evaluated separately. Ways to output confidence used in the compared experiments: 1) anterior wall, unnorm = based on the mean visceral slide value, div_mean = logistic regression with mean visceral slide values as the last feature, stand = logistic regression with mean visceral slide values as the last feature, 2) pelvis, unnorm = based on the mean visceral slide value, div_mean = logistic regression with mean visceral slide values as the last feature, stand = based on the min visceral slide value

is higher than 0.55 for 7 of 12 experiments, whereas for the anterior wall it is lower than 0.41 for 10 of 12 experiments. What is more, when logistic regression is applied, it tends to decrease the performance for binary classification task for the pelvis area and rise AUC to a random guess level for the anterior wall when the mean visceral slide value is used as a feature. The latter is probably merely because the model learns that along the anterior wall a higher mean visceral slide value is associated with an adhesion presence. To check this hypothesis mean visceral slide value feature was replaced with bounding box width in logistic regression. Figure 27 in the appendix compares the impact of this change by showing FROCs and ROCs for cumulative visceral slide normalised with division by expectation and different ways to assign confidence. When bounding box width is used as a feature in logistic regression slice-level ROC is not improved. The

same comparison is made for the pelvis area in figure 28. The only noticeable impact of using bounding box width as a feature is a worse slice-level ROC.

A random guess level performance of the slice-level binary classification task for the whole dataset can be explained by the fact that in the anterior wall area visceral slide tends to be higher in regions that correspond to adhesion annotations than for adhesions-free regions, while in the pelvis area the trend is the opposite.

As for the adhesion detection task, according to FROCs and AP, for the anterior wall, the best performance is achieved with unnormalised visceral slide and both normalisation options lead to a performance drop. On the contrary, for the pelvis area visceral slide normalised with division by expectation gives the best performance. This means that the interpretation of the visceral slide calculated with the

proposed method should be different for the anterior wall and pelvis. Although performance is still poor, the visceral slide estimated in the pelvis area is in line with the concept of the methods to some extent. Whilst, the correspondence between annotations and visceral slide estimated along the anterior wall contradicts the idea of the method. Apparently, the proposed normalisation options failed to sufficiently account for the differences in motion pattern in these two areas. But crucially, the trend observed in the visceral slide along the anterior wall makes the key idea of the method questionable.

D. Evaluation on the held-out test set

The final evaluation of the method configurations that use cumulative visceral slide with the parameters selected with cross-validation was performed on the held-out test set to check the method's generalisability. For configurations that use logistic regression to output confidence, the whole training set was used to fit the model. The share of pelvic adhesions in the training set is higher than in the test set. Therefore, it was natural to expect a noticeable performance drop based on the differences in the relation of visceral slide value and adhesion annotations in the pelvis and along the anterior wall revealed in the section IV-C. Surprisingly, the performance on the test set was better in all cases, especially for slice-level diagnosis. The list of all evaluated method configurations and quantitative performance metrics is given in table IV. Consistently with the previous experiments, normalisation with division by expectation was superior to other normalisation options. FROC_s and slice-level ROC_s for method configurations that use normalisation with division by expectation and vary by the way to output confidence are depicted in figure 17. On the contrary to the cross-validated evaluation on the training set, using minimum visceral slide value gives better performance for both adhesion detection task and slice-level classification task, whereas logistic regression leads to a performance drop.

On the test set, the top-performing method configuration was the one using cumulative visceral slide normalised with division by expectation and minimum visceral slide value to output confidence. Detection sensitivities of 0.7 and 0.91 were achieved at 1 and 1.89 false positives per cine-MRI slice respectively, and AP was 0.36. Slice-level AUC reached 0.78, which indicates decent potential for binary classification. As for the method configuration found the best with cross-validation, that is cumulative visceral slide normalised with division by expectation and using logistic regression to output confidence, its performance was noticeably lower. Detection sensitivities of 0.5 and 0.7 were reached at 1 and 2 false positives per cine-MRI slice respectively, and the highest sensitivity of 0.91 was achieved at 2.9 false positives. AP and slice-level AUC were 0.28 and 0.59 respectively.

The fact that using visceral slide minimum to assign confidence to predicted bounding boxes indicate that for the test set the assumption that drop in the estimated visceral slide is associated with adhesion presence is more valid than in the training set.

TABLE IV: Evaluation of the method on the held-out test set. Only method configurations with cumulative visceral slide were considered. To train logistic regression the set of features selected for each method configuration with cross-validation was used.

Method		Evaluation metrics		
Normalised by statistics	Confidence	AP	Slice AUC	TPs and FPs different
No	min	0.2411	0.7531	No
	mean	0.2511	0.7576	Yes
	LR	0.2757	0.5991	Yes
Division by expectation	min	0.3556	0.7811	Yes
	mean	0.3056	0.6976	Yes
	LR	0.2818	0.5921	Yes
Standardisation	min	0.2093	0.7296	Yes
	mean	0.2152	0.7234	Yes
	LR	0.1742	0.411	No

V. DISCUSSION

A. Limitations

The main limitation of the method is that the used adhesion annotations were not confirmed surgically. This makes the evaluation and reported results doubtful. However, obtaining a similarly large dataset with ground truth annotations is expensive and time-consuming, thus collection of such a dataset would significantly delay the possible start of the project. Also, since the source code of the method is available, it can be tested on the dataset with ground truth annotations as soon as it is obtained.

By design the proposed method is multi-step. In total it has 6 subsequent steps: 1) segmentation of abdominal cavity, 2) visceral slide calculation, 3) normalisation by the average motion along the abdominal cavity contour, 4) (optional) calculation of the healthy control group statistics, 5) (optional) normalisation of visceral slide with control group statistics and 6) prediction based on the visceral slide value. This makes a possible error space big and as a result, the whole method is error-prone and fragile. The first class of possible errors is implementation errors. The interpretation of the performed experiments and analysis given in the results section assumes that all compounds of the method are implemented correctly, however, we cannot guarantee that it is indeed the case.

Another possible cause of poor performance is insufficient optimisation of the method components. For instance, the typical segmentation errors shown in figure 24 in the appendix are rather common. Most of these errors occur in the pelvis area where most of our annotations are located. Perhaps more accurate segmentation is required to unlock the full potential of the method. Also, the parameters used in the registration algorithm could be sub-optimal and it may have a significant impact on the performance. Finally, the core idea of the method might be unsuitable for tackling the adhesion detection task. When diagnosing adhesions, radiologists look not only at the reduction of visceral slide but also the position of intestines inside the abdominal cavity. Therefore,

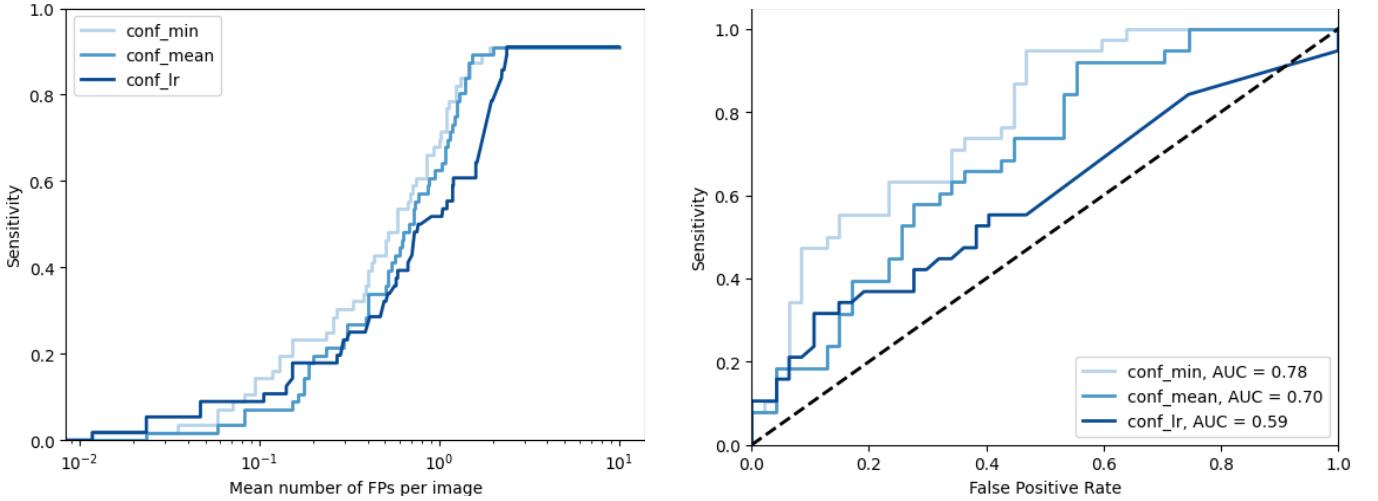


Fig. 17: FROCs (on the left) and slice-level ROCs (on the right) for evaluation of the method on the test set. Compared method configurations use cumulative visceral slide normalised with division by expectation and are different by the way to output confidence.

the proposed method might work only for a fraction of adhesions, which give a noticeable local drop in visceral slide, and miss the others. A more advanced method that takes the position of viscera into account might be required to build an automated CAD system for this task.

Such a wide range of possible reasons of poor method performance makes it hard and time-consuming to determine which improvements should be made in the first place and whether the method has the potential to achieve performance suitable for application in clinics.

Lastly, the method is only able to detect adhesions that intersect the contour of the abdominal cavity. Hence, even if high performance is achieved, adhesions located inside the abdominal cavity between intestinal loops will be missed.

B. Future work

The results of method evaluation with cross-validation and on the test set were controversial due to the difference between test and training sets in the percentage of adhesions at different locations and association of low visceral slide value with adhesion annotations. Hence, for reliable evaluation of the method and perhaps selection of more optimal hyper-parameters, it is necessary to collect a larger and more representative dataset, which is more balanced by the presence of adhesions attached to the anterior wall and pelvic adhesions.

However, even the best performance achieved on the test set is rather poor, therefore an exploration of performance bottlenecks would be a sensible future work direction.

Since there is evidence that in the anterior wall area low visceral slide does not correlate with adhesion annotations, it is worth checking the validity of the key idea of the method. As soon as higher a quality dataset is obtained and there is high confidence in the accuracy of visceral slide calculation and that applied normalisation sufficiently

suppresses differences occurring due to abdominal motion amplitude and position at the abdominal cavity boundary, it should be checked whether lower visceral slide correlates with adhesion annotations. If it does not, it is serious evidence that the method proposed by David Randall is not suitable to form the basics of the automated CAD for adhesion detection. As stated in the pilot study of the method [8], only cine-MRI slices that were accessed as having a sufficiently high amount of respiratory motion were used in the study, which resulted in the exclusion of about half of the available slices. Also, the presence of computed shear drop indicative for adhesions was established by the same specialists who examined the corresponding cine-MR images to confirm adhesions presence. These are two reasons why the reported agreement with clinical judgment might be overestimated. Although in this master thesis cine-MR slices are filtered according to suitability for the method as well, the criteria we use are less strict. Also, human judgment is excluded from the method evaluation, which makes it more objective.

Based on the insights obtained from the evaluation of the method, the following approaches can be tried to improve method performance, listed from the most worthwhile to the least:

- 1) Separate evaluation of the method for anterior wall and pelvis region on the training set indicates that there might be crucial differences in motion patterns in these regions and applied normalisation by control group statistics is insufficient to suppress it. A more detailed examination of these differences on a larger dataset can help to design normalization methods that can align them better and enable using the same algorithm in both areas. Alternatively, it might be sensible to use separate methods to predict adhesions in these areas.
- 2) The segmentation model still frequently fails in the pelvis area and sometimes along the anterior abdom-

inal wall. This may lead to an inaccurate estimate of visceral slide in the areas essential for prediction. Although usage of the cumulative visceral slide should mitigate this problem, visualisation of segmentation on the whole cine-MRI slices shows that segmentation in the pelvis area might be rather unstable even for "easy" slices and hence the computed visceral slide might be noisy. Since abdominal cavity shape and viscera position can drastically vary across patients, we expect that using a larger dataset to train the nnU-Net can give a decent performance improvement. Also, customisation of image pre-processing, data augmentation and nnU-Net parameters can be tried.

- 3) More robust ways to calculate visceral slide statistics by location should be explored. The current way of splitting the abdominal cavity contour into a predefined number of chunks and matching them by relative location is straightforward to implement, however it was not thoroughly tested and it might lack necessary accuracy due to differences in shapes of abdominal cavity among patients and different positions of cine-MRI slices. Also, a larger and more diverse control group is desirable to obtain more accurate expectation and standard deviation estimates.
- 4) Due to time constraints, exploration of registration parameters was not done. However, sub-optimal registration might critically distort the estimated visceral slide, since it is based on the difference in deformation on the opposite sides of the abdominal cavity contour. Now, when the implementation of the automated method is available, it is more feasible, however still time-consuming because image registration with gradient descent optimization used in ANTs is slow.
- 5) To facilitate future experiments and possibly improve the accuracy of image registration, ANTs toolkit can be replaced with a deep learning registration model [14]. It will substantially speed up visceral slide calculation and as a result, investigation of method improvements.
- 6) Since for each point visceral slide is calculated based on the values of the full deformation field at the opposite sides of the abdominal cavity contour, even minor segmentation inaccuracies might lead to an incorrect visceral slide estimate. Checking the effect of a slightly changing segmentation mask on the calculated visceral slide will help to determine which accuracy of a segmentation model is required for an adequate algorithm implementation and how feasible the development of such models is.
- 7) There might be substantial differences in the visceral slide distribution at different positions of a cine-MRI slice. Paramedian and midline slices are quite different in shape, visible organs and motion patterns. That might lead to sub-optimal performance when the method does not account for possible differences between cine-MRI slices at different positions. In the proposed method, calculation of healthy control group statistics might be especially affected by this, since it

relies on the idea that shapes of abdominal cavities are sufficiently similar. Therefore, it is worth investigating whether the visceral slide pattern varies by position at which a cine-MRI slice is taken and whether the calculation of healthy control group statistics by slice position improves the results.

C. Conclusion

This master thesis proposes the first fully-automated multi-step CAD method for adhesion detection on cine-MRI designed based on the semi-automated method introduced by David Randall in his PhD thesis. In the method, the visceral slide that occurs on a sagittal cine-MRI slice is quantified using deep learning and image registration techniques and passed to the region growing algorithm that predicts adhesions. Two ways to compute visceral slide were explored. In the first one, only the two most dissimilar frames of a cine-MRI slice are used to sharpen the differences between regions of lower and higher motion and the second method exploits all time points in a cine-MRI slice to obtain a full picture of motion. To make visceral slide comparable between different cine-MRI slices, normalisation by the average horizontal motion was applied. Also, the differences in typical visceral slice values by location at abdominal cavity boundary were adjusted with normalisation by healthy control group statistics. Finally, three ways to assign confidence to a predicted adhesion bounding box were tried: based on the minimum visceral slide value in the bounding box, based on the mean visceral slide value and by training the logistic regression to output confidence.

The results of the cross-validated evaluation and evaluation on the held-out test set were controversial. Performance on the test set was higher than cross-validated for all method configurations and a slightly different method was top-performing. In both cases, cumulative visceral slide normalised by the average horizontal motion and with division by expectation was found superior for the adhesion detection task and gave one of the highest slice-level AUCs. However, different ways to output confidence yielded further performance improvement. According to cross-validated evaluation, confidence predicted by logistic regression gives a noticeable performance boost. Then, sensitivities of 0.61 and 0.73 at 1 and 2 false positives per slice along with 0.53 AUC in slice-level diagnosis were achieved. On the held-out test set, confidence based on the minimum visceral slide value unlocked the highest performance and resulted in detection sensitivity of 0.7 and 0.91 at 1 and 1.89 false positive per slice along with 0.78 slice-level AUC. These discrepancies can be explained by differences in properties of test and training set. First, in the training set, annotations of pelvic adhesions are over-represented and the test set is more balanced concerning annotations location. Also, there is evidence that in the test set low visceral slide value correlates with adhesion annotations better than in the training set.

The differences in evaluation results on the test and training sets highlight that acquisition of a higher quality dataset is essential for reliable method evaluation and an adequate

choice of the best method and its hyper-parameters. Such a dataset should be balanced by the presence of adhesions attached to the anterior wall and pelvic adhesion and more representative in motion patterns.

Although performance on the held-out test set was better and slice-level AUC of the best method looks promising, it is still insufficient to consider the usage of the method in clinical practice. Nevertheless, the proposed method is the first step on the path towards the creation of the CAD for adhesion detection suitable for clinical usage. The detailed information gathered during the method evaluation and proposed future work directions can guide the further development of the method.

D. Acknowledgements

The author appreciates the collaboration of Frank Joosten, a radiologist at Rijnstate Hospital, who made adhesion annotations for the dataset used in the project and reviewed our abdominal cavity segmentations. Also, the guidance of the project supervisors, Henkjan Huisman, Associate Professor at DIAG, Radboudumc, Twan van Laarhoven, Assistant Professor at iCIS, Radboud University, and Bram de Wilde, PhD candidate at DIAG, Radboudumc, and their contribution to the project was invaluable.

REFERENCES

- [1] N. Tabibian, E. Svehli, A. Boyd, A. Umbreen, and J. Tabibian, “Abdominal adhesions: A practical review of an often overlooked entity,” *Annals of Medicine and Surgery*, vol. 15, pp. 9–13, 2017.
- [2] R. P. Ten Broek, Y. Issa, E. J. van Santbrink, N. D. Bouvy, R. F. Kruitwagen, J. Jeekel, E. A. Bakkum, M. M. Rovers, and H. van Goor, “Burden of adhesions in abdominal and pelvic surgery: systematic review and met-analysis,” *Bmj*, vol. 347, 2013.
- [3] M.-A. Weibel and G. Majno, “Peritoneal adhesions and their relation to abdominal surgery: a postmortem study,” *The American Journal of Surgery*, vol. 126, no. 3, pp. 345–353, 1973.
- [4] R. A. Lang, S. Buhmann, A. Hopman, H.-O. Steitz, A. Lienemann, M. F. Reiser, K.-W. Jauch, and T. P. Hüttl, “Cine-mri detection of intraabdominal adhesions: correlation with intraoperative findings in 89 consecutive cases,” *Surgical endoscopy*, vol. 22, no. 11, pp. 2455–2461, 2008.
- [5] B. A. W. van den Beukel, M. W. J. Stommel, S. van Leuven, C. Strik, M. A. IJsseldijk, F. Joosten, H. van Goor, and R. P. G. Ten Broek, “A shared decision approach to chronic abdominal pain based on cine-mri: a prospective cohort study,” *American Journal of Gastroenterology*, vol. 113, no. 8, pp. 1229–1237, 2018.
- [6] A. Lienemann, D. Sprenger, H. O. Steitz, M. Korell, and M. Reiser, “Detection and mapping of intraabdominal adhesions by using functional cine mr imaging: preliminary results,” *Radiology*, vol. 217, no. 2, pp. 421–425, 2000.
- [7] D. Randall, “Towards a non-invasive diagnostic aid for abdominal adhesions using dynamic mri and image processing,” Ph.D. dissertation, University of Sheffield, 2017.
- [8] D. Randall, F. Joosten, R. P. Ten Broek, R. Gillott, K. D. Bardhan, C. Strik, W. Prins, H. van Goor, and J. W. Fenner, “A novel diagnostic aid for intra-abdominal adhesion detection in cine-mri: pilot study and initial diagnostic impressions,” *The British journal of radiology*, vol. 90, no. 1077, p. 20170158, 2017.
- [9] F. Isensee, P. F. Jaeger, S. A. Kohl, J. Petersen, and K. H. Maier-Hein, “nnu-net: a self-configuring method for deep learning-based biomedical image segmentation,” *Nature methods*, vol. 18, no. 2, pp. 203–211, 2021.
- [10] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.
- [11] B. B. Avants, N. J. Tustison, G. Song, and J. C. Gee, “Ants: Open-source tools for normalization and neuroanatomy,” *HeanetIe*, vol. 10, pp. 1–11, 2009.
- [12] X. He and E. Frey, “Roc, lroc, froc, afroc: An alphabet soup,” *Journal of the American College of Radiology*, vol. 6, no. 9, pp. 652–655, 2009.
- [13] R. Padilla, W. L. Passos, T. L. Dias, S. L. Netto, and E. A. da Silva, “A comparative analysis of object detection metrics with a companion open-source toolkit,” *Electronics*, vol. 10, no. 3, p. 279, 2021.
- [14] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca, “Voxelmorph: a learning framework for deformable medical image registration,” *IEEE transactions on medical imaging*, vol. 38, no. 8, pp. 1788–1800, 2019.

VI. APPENDIX

A. Suitability of slices for the method

The entire data set contains all potentially clinically relevant slices taken according to cine-MRI scan acquisition protocol, however, only slices with particular properties are suitable for the method.

The key idea of the method is to quantify the visceral slide and detect adhesions along the abdominal cavity contour based on the local drop in visceral slide. This implies a few requirements for slices that can be processed by the method.

First, the amplitude of motion exhibited by a patient should be sufficiently high to make the visceral slide visible for both the method and human expert. Generally, there are two motion protocols that patients are asked to follow during a cine-MRI scan acquisition: 1) deep respiration with active movement of the anterior abdominal wall and 2) squeezing pelvis without moving the anterior wall. The second protocol is not suitable for the method because the motion amplitude is lower and the motion pattern is not indicative for visceral slide assessment. Also, not all sagittal slices positions are equally informative. According to a radiologist we consulted, medical specialists check midline slices in the first place and most frequently adhesions are found on such slices. This is in line with the adhesion annotations statistics by slice position in our dataset given in table V, where the majority of slices with adhesion annotations are midline. Left/right paramedian slices are suitable for diagnosis too, however much less is visible on the left/right lateral slices and such slices should be left out while sampling the dataset.

Information about these properties of slices is contained in the metadata of images, which enables automated sampling of suitable slices. However, the precise position of a slice cannot be inferred from metadata since only three position constants are used: middle, right and left. That does not allow differentiating between paramedian and lateral slices. Due to this and better suitability for adhesions diagnosis, it was decided to sample only midline slices to save time on manual check of paramedian slices. The only exception is a sampling of the test set, which consists of 30 studies. We found it sufficiently small to manually filter out lateral slices. Besides, it was unknown on which slices adhesions were present in positive studies.

Lastly, to reduce possible noise which may arise from inaccurate diagnosis, 10 patients who have both positive and negative studies were excluded from sampling.

B. Image registration details

A Python wrapper for ANTS, ANTsPy⁴, was used in the project. The following registration parameters were employed:

- Type of transform (`type_of_transform`) - *SyN-Only*, symmetric normalization without initial transformation and with mutual information as optimization metric,
- The metric for the syn part (`syn_metric`) - mattes,

- Initial transform to prepend (`initial_transform`) - identity,
- Radius for the syn metric (`syn_sampling`) - 8,
- Smoothing for total field (`total_sigma`) - 1,
- Registration iterations (`reg_iterations`) - (1000, 1000, 100). Define smoothing and multi-resolution parameters for the syn part.

For other parameters the default values were kept.

⁴<https://github.com/ANTsX/ANTsPy>

TABLE V: Adhesion annotations statistics by slices position.

Type	Left	Right	Paramedian (L + R)	Middle	Total
All	5 (7.94%)	10 (15.87%)	15 (23.81%)	48 (76.19%)	63
Anterior wall	3 (16.67%)	5 (27.78%)	8 (44.44%)	13 (55.56%)	18
Pelvis	2 (4.00%)	4 (8.00%)	6 (12.00%)	44 (88.00%)	50
Anterior wall & pelvis	5 (8.33%)	9 (15.00%)	14 (23.33%)	46 (76.67%)	60

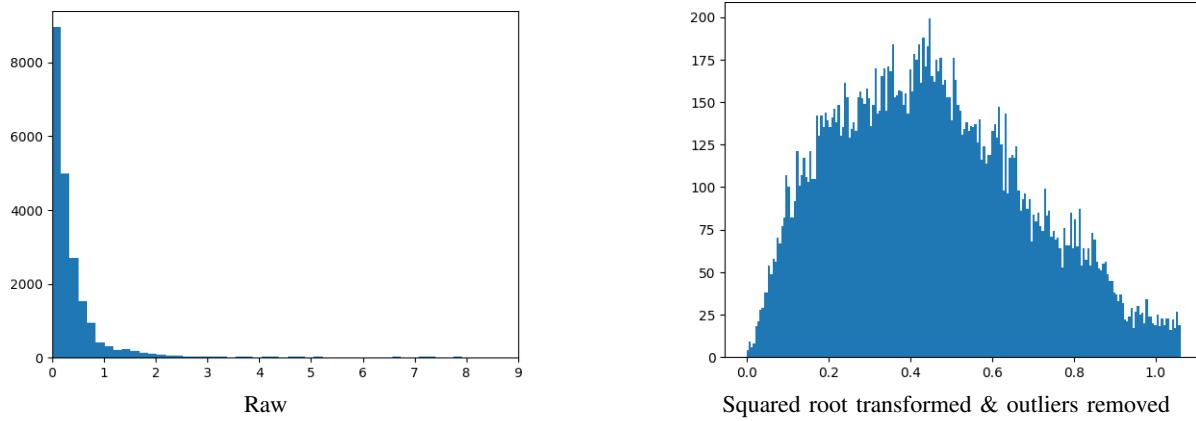


Fig. 18: Distribution of raw and squared root transformed values of inspiration/expiration visceral slide. In addition to squared root transformation outliers were removed.

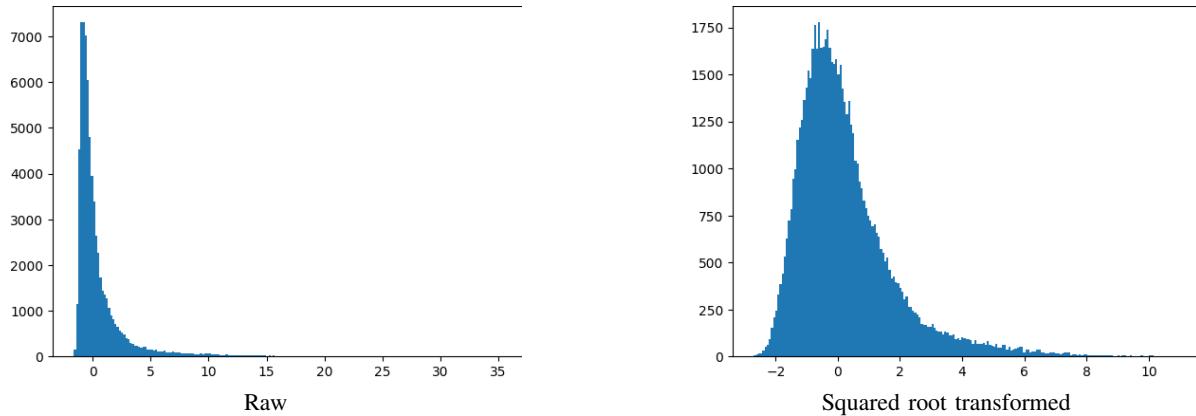


Fig. 19: Distribution of raw and squared root transformed inspiration/expiration visceral slide standardised with the control group statistics.

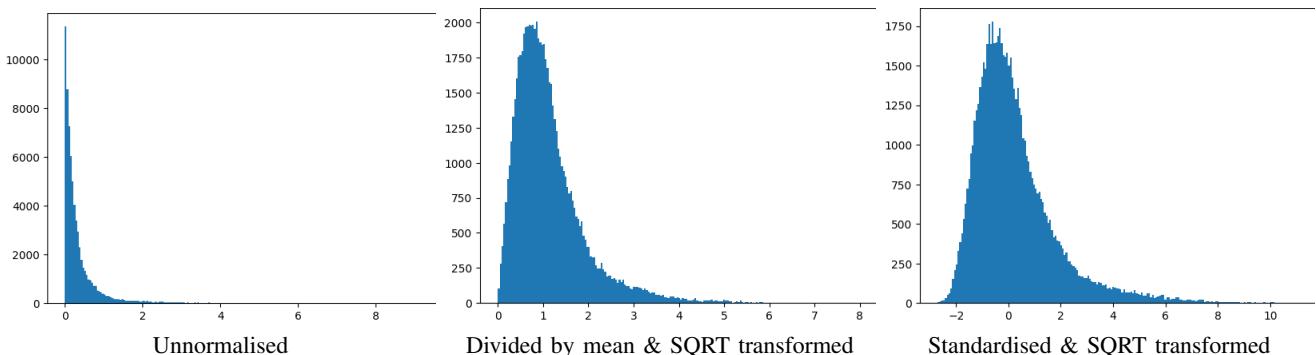


Fig. 20: Distribution of inspiration/expiration visceral slide with different normalisation by the control group statistics.

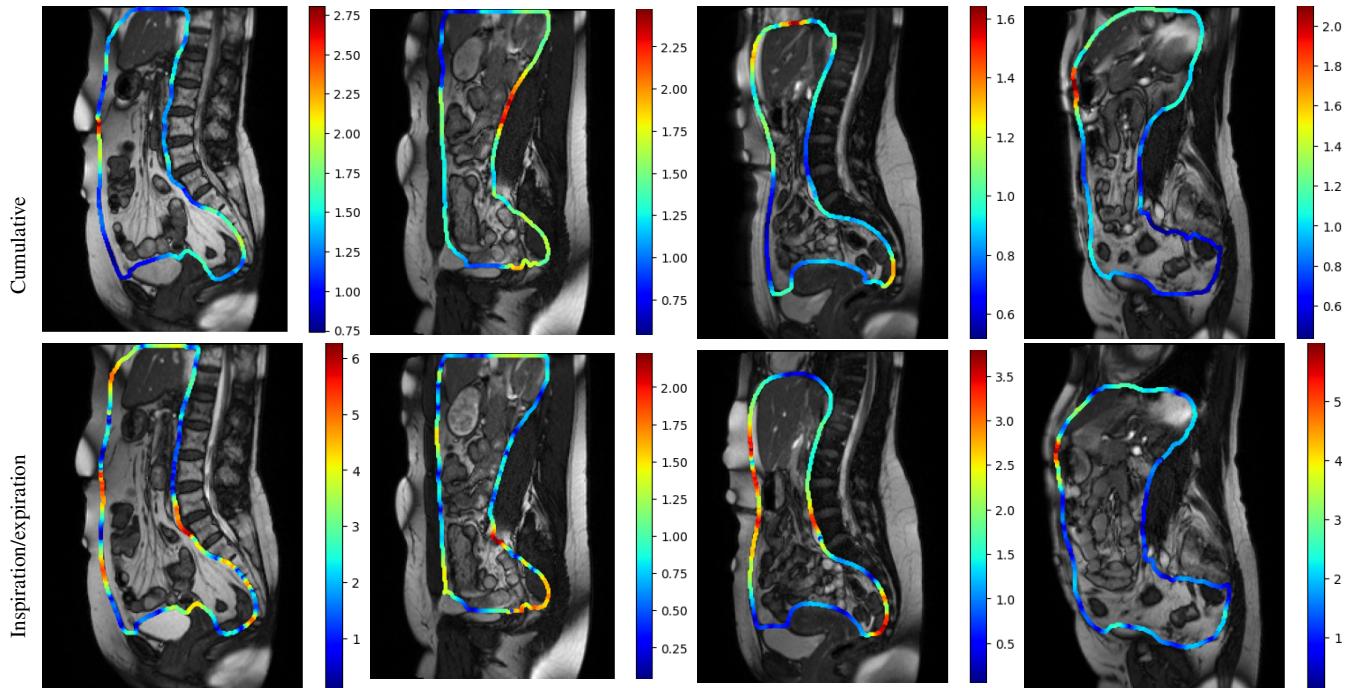


Fig. 21: Examples of cumulative and inspiration/expiration visceral slides normalised by the average horizontal motion and division by expectation after squared root transformation of visceral slide values. The cumulative visceral slides are placed in the top row and inspiration/expiration visceral slides are in the bottom.

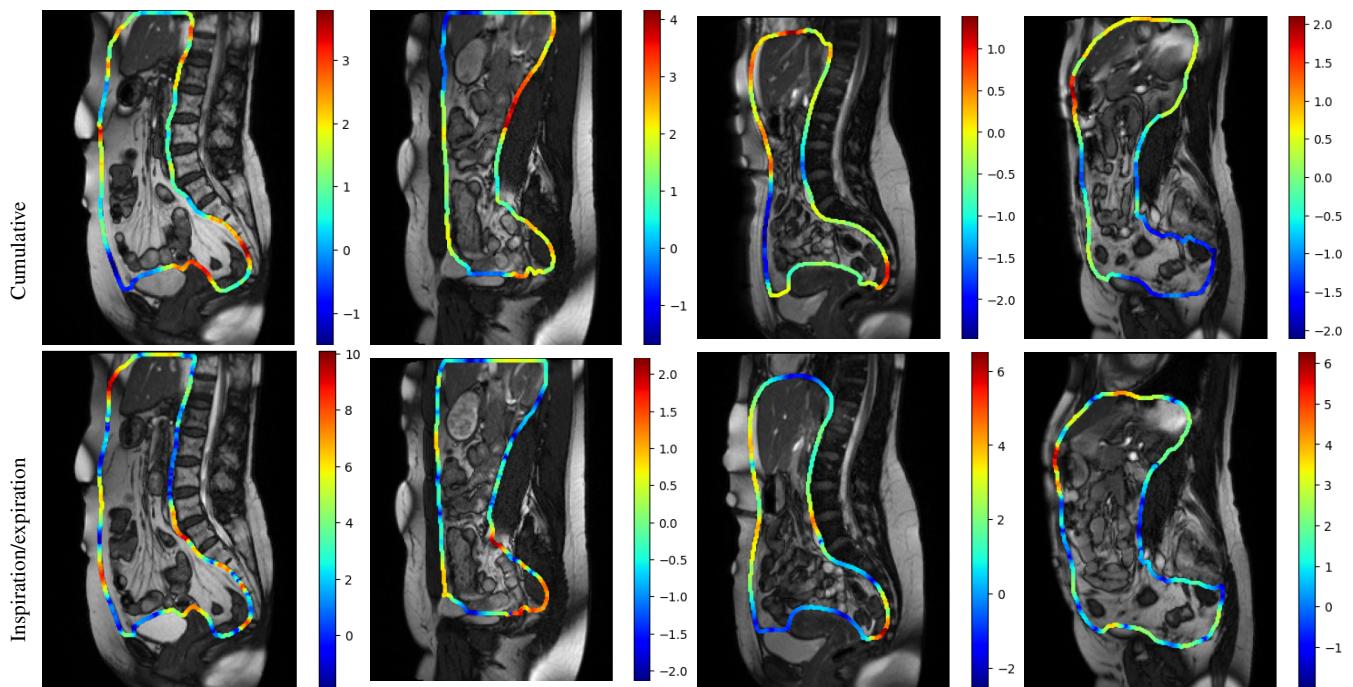


Fig. 22: Examples of cumulative and inspiration/expiration visceral slides normalised by the average horizontal motion and standardisation after squared root transformation of visceral slide values. The cumulative visceral slides are placed in the top row and inspiration/expiration visceral slides are in the bottom.

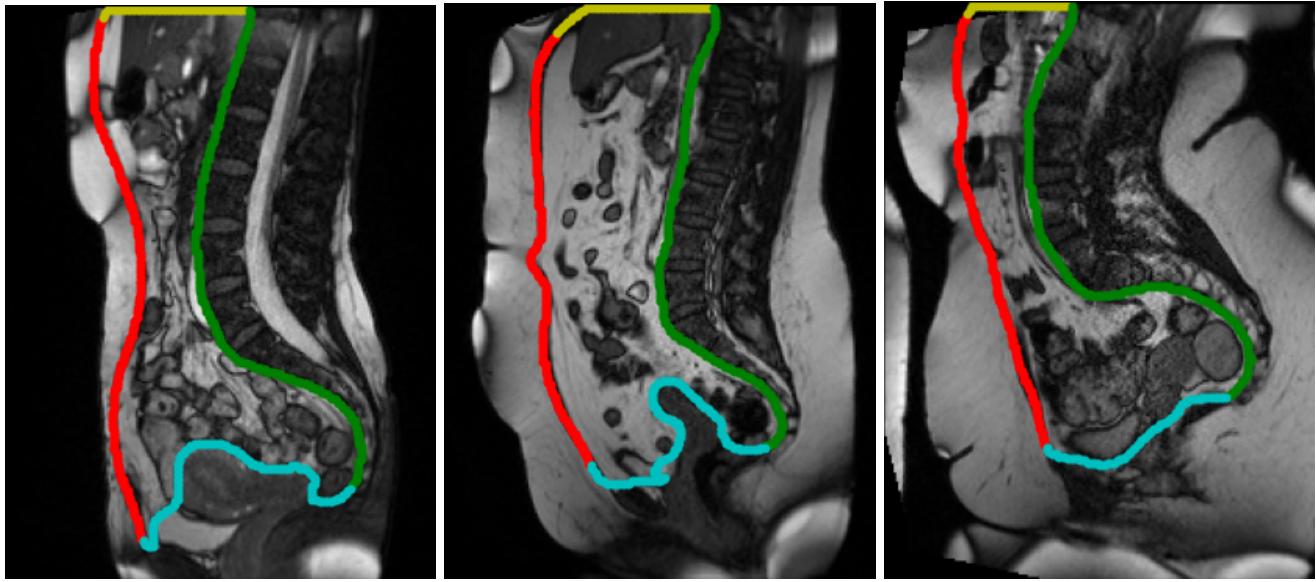


Fig. 23: Examples of detection of four parts (top, bottom, anterior and posterior walls) of abdominal cavity contour visualised with different colors on an arbitrary frame of cine-MRI slices.

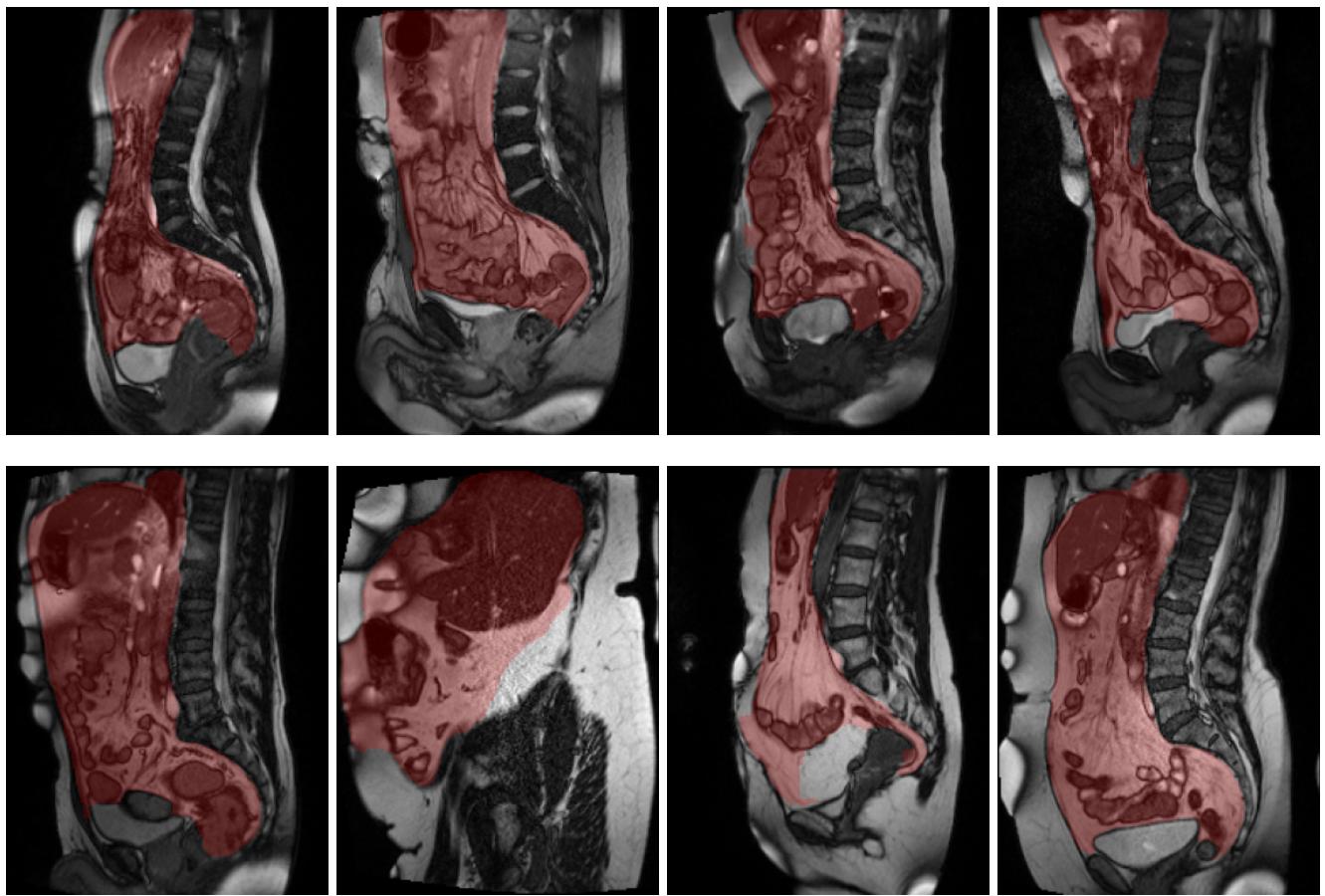


Fig. 24: Examples of typical errors nnU-Net model made in predicted segmentation of abdominal cavity. From top left to bottom right: 1 - top parts of uterus and rectum are included, 2 - rectum is included, 3 - artefacts around the bottom part of the anterior abdominal wall, 4 - a part of the bladder is included and inconsistencies in the posterior wall delineation, 5 - a bowel loop above the bladder is not included and too much is included at the bottom right, 6 - failure with anterior wall detection, 7 - incorrect segmentation of the bottom part, 8 - too much is included at the top right and a small bowel loop at the bottom is not included.

TABLE VI: Performed experiments varied by the used visceral slide computation and normalisation methods and a way to output confidence. *Distribution of inspiration/expiration visceral slide not normalised by the control group statistics has the narrowest IQR and has the highest share of values close to 0. Due to this, a non-linear transformation of the local minimum selected for region growing to obtain the maximum visceral slide allowed in the region gives better performance than using a region growing index. If $v_{min} < 1$, $v_{max} = \sqrt{v_{min}}$ and for $v_{min} > 1$, $v_{max} = v_{min}^2$. Features used to fit logistic regression: 1, 5. bounding box height, bounding box width, region length 2, 3, 6. bounding box height, minimum visceral slide value in a region, region length 4. bounding box height, mean visceral slide value in a region, region length. The best AP and slice-level AUC are given in bold.

Method			Parameters		Confidence	Evaluation metrics		
Visceral slide type	Normalised by statistics	SQRT transformed	MRI	RGI		AP	Slice AUC	TPs and FPs different
Inspiration/expiration	No	No	3	SQRT*	min mean LR ¹	0.1168 0.1059 0.1283	0.5786 0.5012 0.4818	Yes No No
	Division by expectation	Yes	3	7	min mean LR ²	0.144 0.1432 0.213	0.503 0.4854 0.5375	Yes No Yes
	Standardisation	Yes	3	7	min mean LR ³	0.1414 0.1366 0.1701	0.4494 0.4982 0.5735	Yes Yes Yes
Cumulative	No	No	5	2.5	min mean LR ⁴	0.1527 0.1577 0.284	0.4708 0.4661 0.5027	No No Yes
	Division by expectation	Yes	5	2.5	min mean LR ⁵	0.1713 0.1648 0.3198	0.4336 0.4247 0.5336	No No Yes
	Standardisation	Yes	5	5	min mean LR ⁶	0.1076 0.1121 0.2073	0.4682 0.4619 0.5008	No No Yes

TABLE VII: The anterior wall experiments. In all experiments with logistic regression bounding box height and region length are used as features and the last feature is specified in brackets. mean VS = mean visceral slide, min VS = minimum visceral slide, BB width = width of the predicted bounding box. The best AP and slice-level AUC are given in bold.

Method			Parameters		Confidence	Evaluation metrics		
Visceral slide type	Normalised by statistics	SQRT transformed	MRI	RGI		AP	Slice AUC	TPs and FPs different
Cumulative	No	No	5	2.5	min mean LR (mean VS) LR (BB width)	0.3643 0.382 0.3715 0.3774	0.4074 0.3796 0.3426 0.3549	No No No No
	Division by expectation	Yes	5	1.5	min mean LR (mean VS) LR (BB width)	0.2749 0.2648 0.2806 0.274	0.321 0.3117 0.5154 0.3425	Yes Yes No No
	Standardisation	Yes	5	2.5	min mean LR (min VS) LR (BB width)	0.2341 0.2382 0.2655 0.2272	0.3858 0.3395 0.5307 0.4012	No No No No

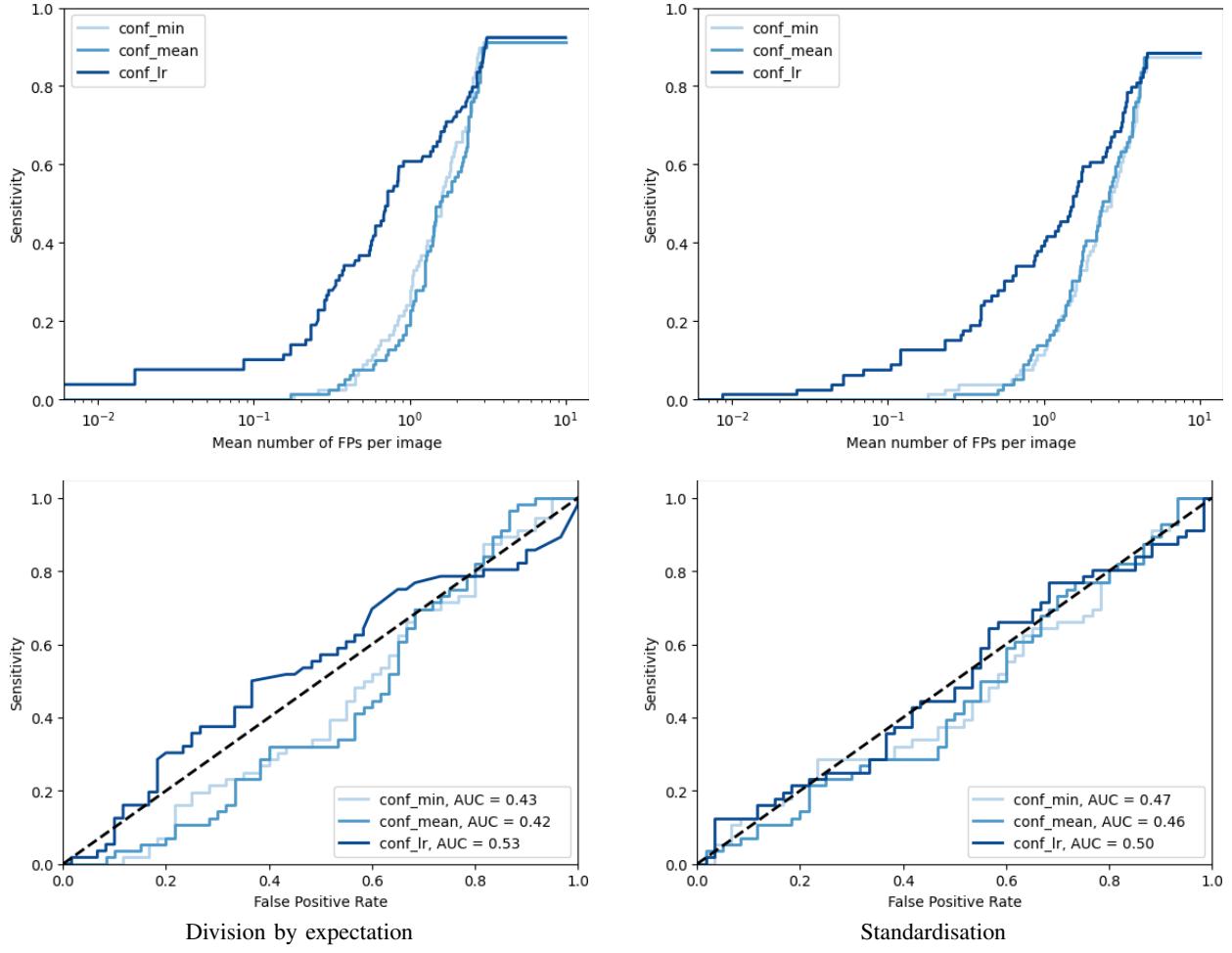


Fig. 25: FROCs (top row) and slice-level ROCs (bottom row) for experiments with cumulative visceral slide plotted separately for different normalisation options with control group statistics.

TABLE VIII: The pelvis experiments. In all experiments with logistic regression bounding box height and region length are used as features and the last feature is specified in brackets. mean VS = mean visceral slide, min VS = minimum visceral slide, BB width = width of the predicted bounding box. The best AP and slice-level AUC are given in bold.

Visceral slide type	Method		Parameters		Confidence	Evaluation metrics		
	Normalised by statistics	SQRT transformed	MRI	RGI		AP	Slice AUC	TPs and FPs different
Cumulative	No	No	5	2.5	min	0.2244	0.5922	No
					mean	0.2255	0.587	No
					LR (mean VS)	0.2197	0.5194	Yes
					LR (BB width)	0.2216	0.4622	Yes
	Division by expectation	Yes	5	2.5	min	0.3102	0.5784	Yes
					mean	0.2981	0.6115	Yes
					LR (mean VS)	0.3078	0.5912	No
					LR (BB width)	0.3252	0.4998	No
Standardisation	Standardisation	Yes	5	5	min	0.2257	0.5581	No
					mean	0.2097	0.578	No
					LR (min VS)	0.2213	0.4887	No
					LR (BB width)	0.207	0.3932	No

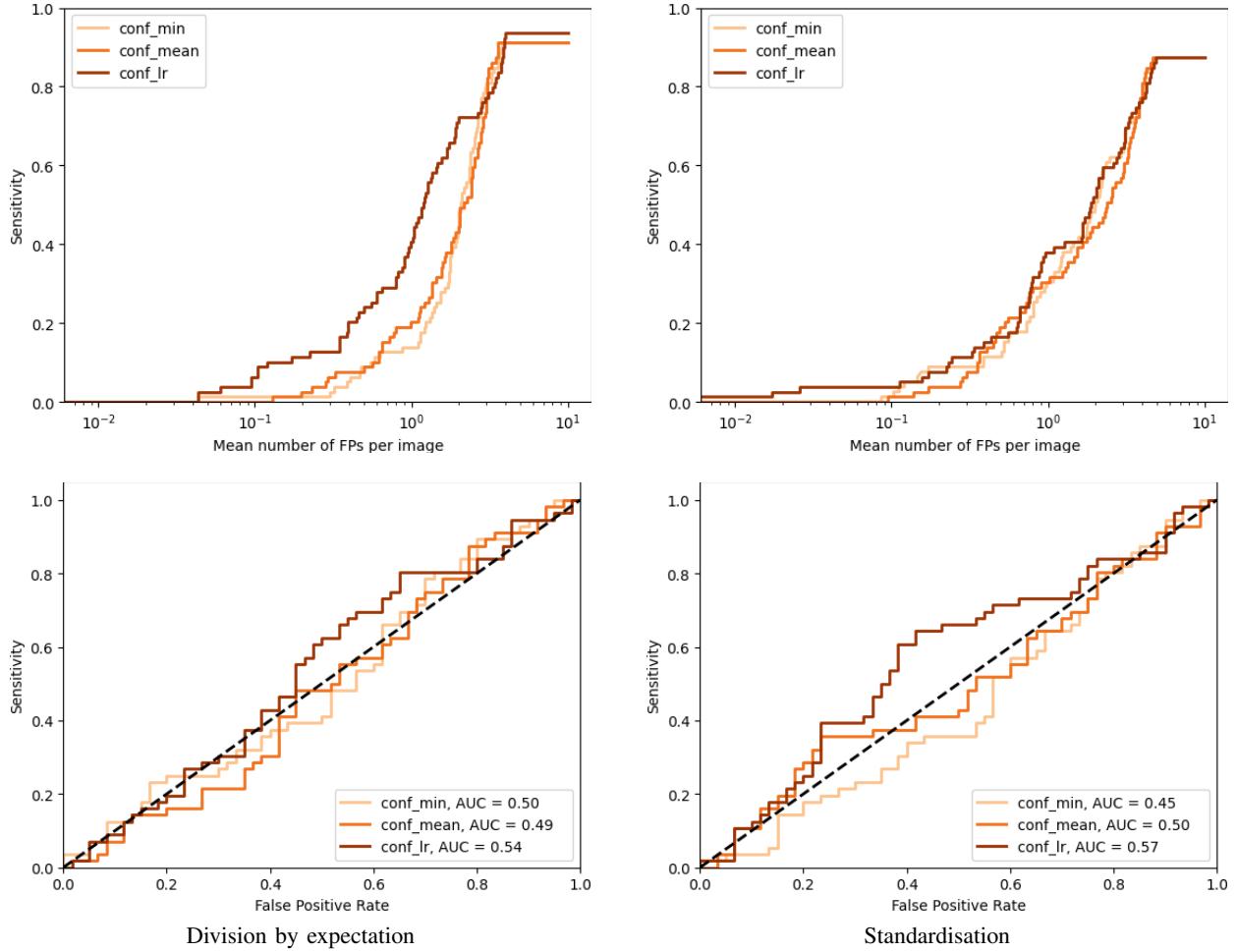


Fig. 26: FROCs (top row) and slice-level ROCs (bottom row) for experiments with inspiration/expiration visceral slide plotted separately for different normalisation options with control group statistics.

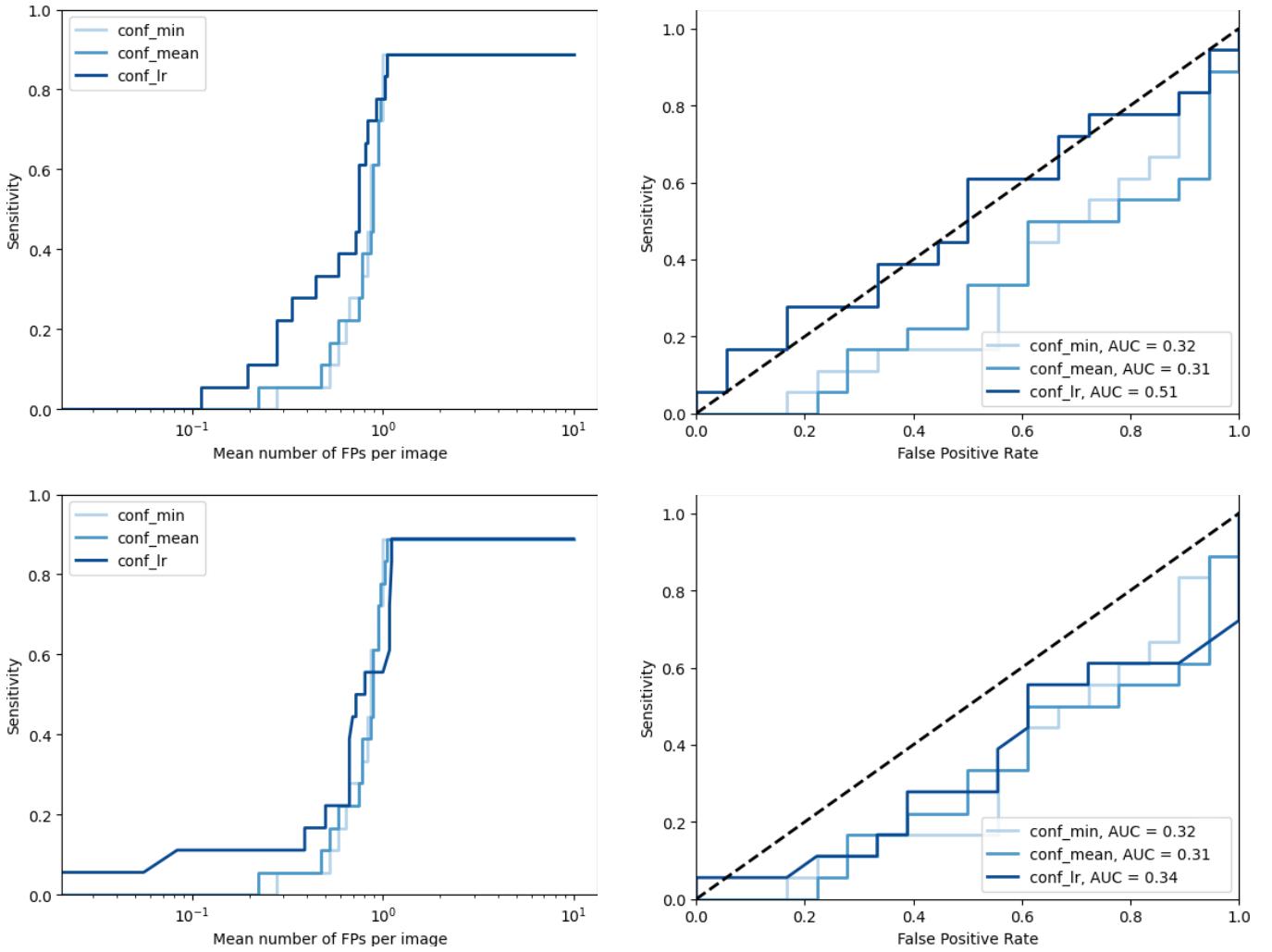


Fig. 27: FROCs (left) and slice-level ROCs (right) for experiments with separate evaluation of the method for the anterior abdominal wall. Cumulative visceral slide normalised with division by expectation is used. In the top and bottom rows only one feature used in logistic regression is different. Top: mean value of visceral slide, bottom: width of the bounding box.

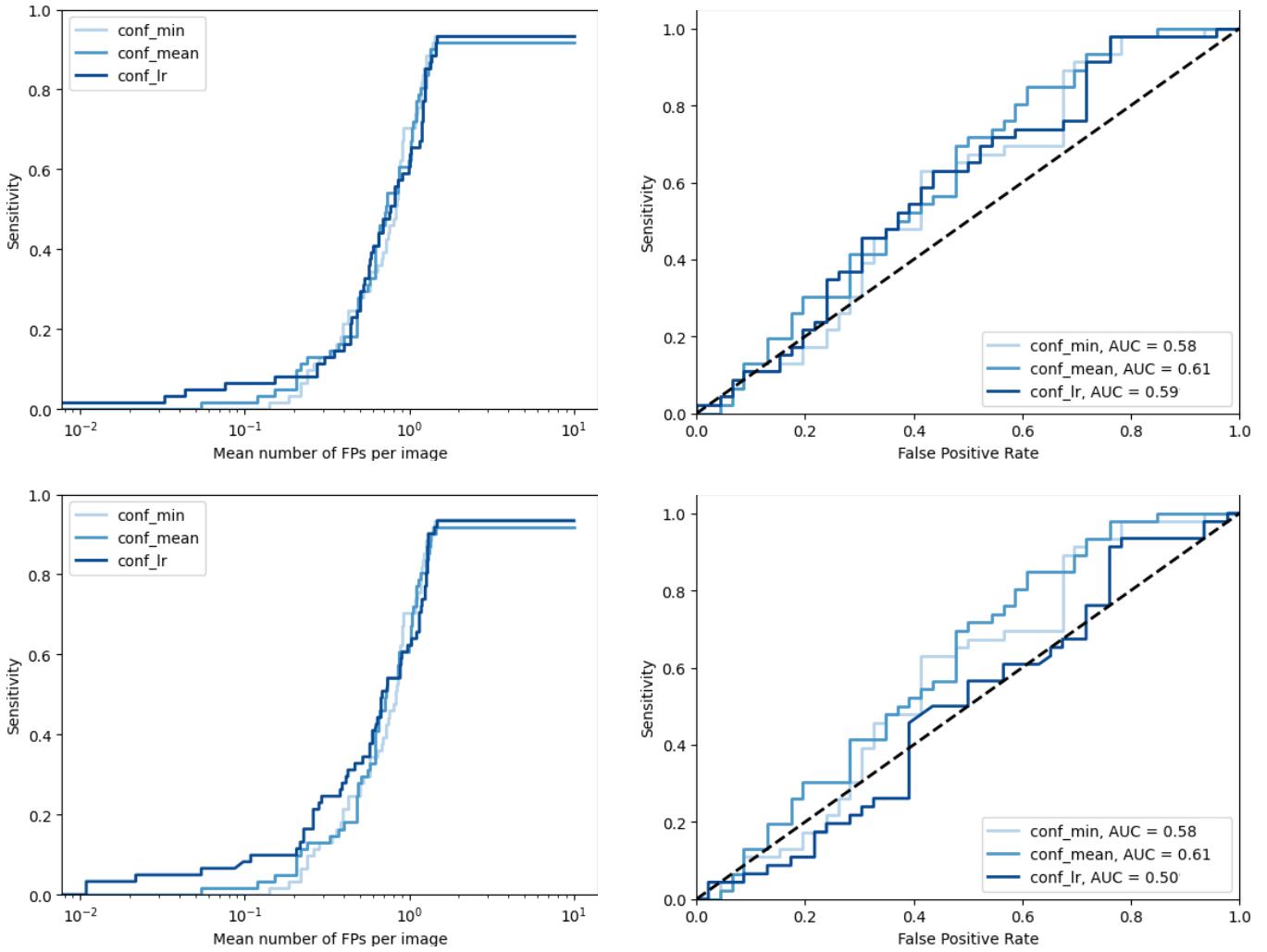


Fig. 28: FROCs (left) and slice-level ROCs (right) for experiments with separate evaluation of the method for pelvis area. Cumulative visceral slide normalised with division by expectation is used. In the top and bottom rows only one feature used in logistic regression is different. Top: mean value of visceral slide, bottom: width of the bounding box.