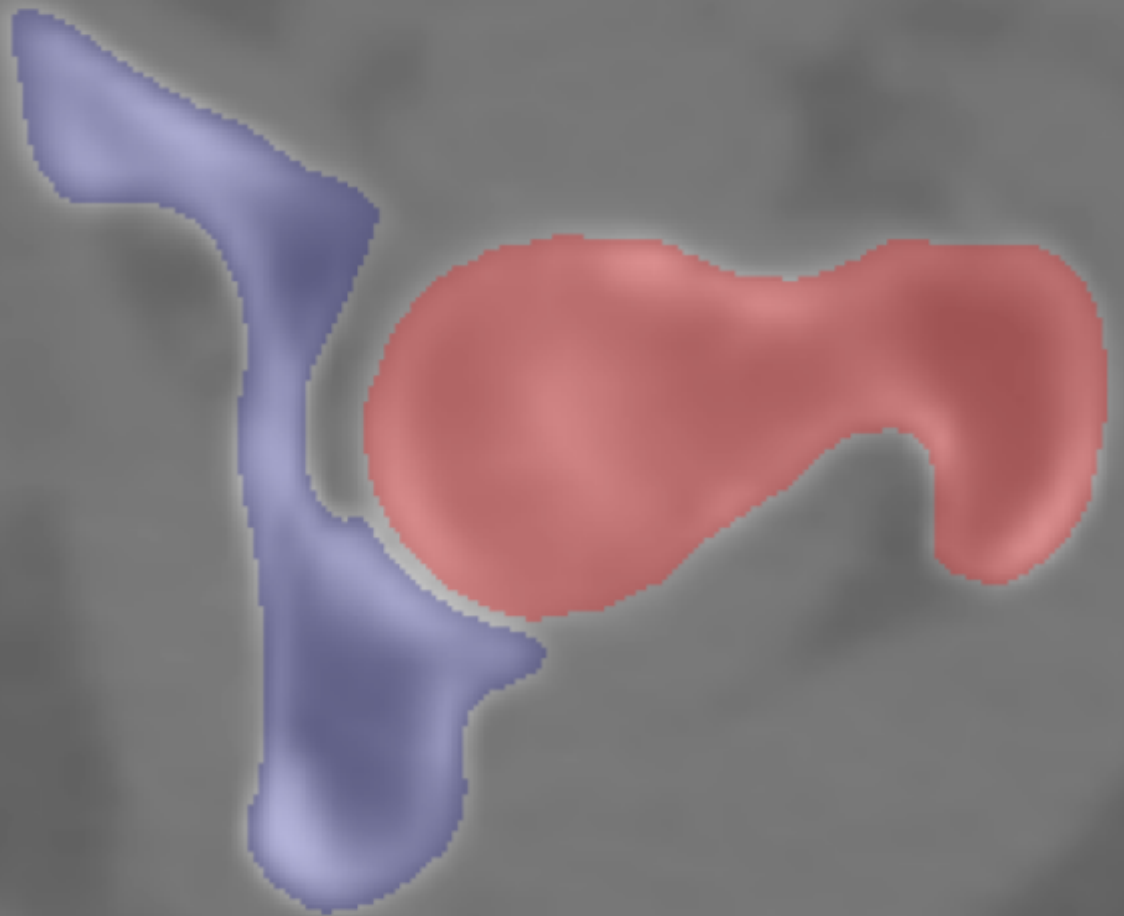


# Interactive 3D Segmentation of CT-scans of the Hip Joint, Using Active Learning to Select Delineation Planes.

Thera E. Pals

Master of Science Thesis





# Interactive 3D Segmentation of CT-scans of the Hip Joint, Using Active Learning to Select Delineation Planes.

by

**Thera E. Pals**

in partial fulfillment of the requirements for the degree of

**Master of Science**  
in Biomedical Engineering

at the Delft University of Technology,  
to be defended publicly on Tuesday January 17, 2017 at 9:30 AM.

Student number:	1524267	
Supervisor:	Dr. M. Loog	
Thesis committee:	Prof. dr. ir. B. P. F. Lelieveldt,	TU Delft
	Dr. F. M. Vos,	TU Delft

An electronic version of this thesis is available at <http://repository.tudelft.nl/>.



# Preface

During the final project of my bachelor in Applied Physics I got interested in medical imaging. While I did not fully want to leave physics I was interested in doing something else than the master Applied Physics. Luckily the Biomedical Engineering masters program has a specialization that combines knowledge of physics with medical imaging. During the courses of this specialization my interest for image processing and machine learning grew. During my internship at the company Clinical Graphics, they were in contact with Marco Loog about a master research project on combining ideas of Active Learning to make their interactive segmentation more efficient. Because of my interest in both machine learning and image processing I decided to take on this project for my master thesis, with Marco Loog as a supervisor. At Clinical Graphics Krijn van Golen helped me as my day to day supervisor. I want to thank Marco Loog and Krijn van Golen for all their help and feedback during this project.

For this project I implemented a method that intelligently suggests planes for correction of a segmentation. This thesis consists of a paper and a workbook. The paper describes the theory of the implemented method, what experiments I performed and the results of these experiments. The paper gives an overview of the most important findings of my thesis project. The workbook is a collection of documents which I used during the project to keep track of what I was doing.

*Thera E. Pals  
Delft, January 2017*

# Interactive 3D Segmentation of CT-scans of the Hip Joint, Using Active Learning to Select Delineation Planes.

Thera E. Pals\*

*Delft University of Technology*

---

## Abstract

This paper evaluates a method that improves segmentation efficiency by intelligently suggesting planes where correction is most valuable. An existing method is extended to work for segmentation of multiple bones simultaneously. This method is evaluated because in clinical practice it is often necessary that scans are segmented very accurate. When automatic segmentation algorithms are not accurate enough, a user needs to provide correction. The method described in this paper uses active learning to alleviate the user from picking a plane to provide correction on. We describe a series of experiments that compare the accuracy of the segmentation using the implemented method with a method that corrects random planes and with an upper limit. The implemented method performs better than random plane suggestion, however the accuracy does not reach the upper limit. Out of the two bones that were evaluated the more complex shaped bone takes longer to reach a high accuracy than the less complex bone.

*Keywords:* Segmentation, CT-scans, Active Learning

---

## 1. Introduction

### 1.1. Importance

Image segmentation is the partitioning of an image into segments by assigning pixels as belonging to objects in the image. To be able to analyze the shape of bones in medical images, the contour of a bone can for instance be segmented in the image. Segmenting bones by hand gives accurate results, but it is time-costly work. Automatic segmentation methods are faster but often less accurate and robust. In clinical practice, where it can be necessary that every scan is segmented accurately, combining user interaction with an automatic segmentation method can be very helpful. [1]

Active learning is an area within machine learning, where users are actively queried for input. The idea behind active learning is that a machine learning algorithm can learn faster by asking better questions, with the goal of minimizing the needed user input. This can be useful in situations where manually labeling data is expensive but a lot of unlabeled data is

---

\*I am corresponding author

available. A user can interact with the model by answering the questions, and thus labeling a part of the data. This newly obtained labeled data is used to further expand the model. An often used query framework is uncertainty sampling; the user is queried to label the points about which the current model is least certain. Ideas from Active Learning can be used in interactive segmentation to decrease the necessary user interaction and to help the user with the decision of where to provide corrections.

The project is performed in collaboration with the Delft based company Clinical Graphics. The company processes patient scans of hip joints, with the goal of visualizing the joint mobility with a digital 3D bone model. Because manual correction is still necessary, it is beneficial for the company if the scans can be processed more efficiently and less time consuming.

## *1.2. Related Work*

Interactive segmentation is usually posed as an iterative process. An automatic segmentation is corrected by a user, this correction is then used to guide the computation update. A new segmentation is presented to the user who makes additional corrections if necessary. Active learning can be combined with interactive image segmentation, to increase the efficiency of interaction. By learning from the user input, segmentations can be improved. And by using the already labeled data smart suggestions for regions where to provide corrections in a segmentation can be made.

This work focuses on active learning in the segmentation of multiple 3D objects. Previous literature on active learning in segmentation has not addressed segmenting multiple bones simultaneously. It has discussed active learning in a 2D segmentation context and in the past few years also in a 3D segmentation context. Iglesias et al. in 2011 [2] propose a method where the user is queried to manually segment a subset of scans from a large set, called a pool, of unlabeled scans. They try to find a subset of the pool of scans to obtain an optimal training set. Even though this does reduce the needed manual segmentation, it still requires the complete manual segmentation of multiple scans.

The most promising use of active learning in segmentation seems to be a method that suggests planes for correction proposed by Top et al. [3][4], for faster segmentation in 3D medical scans. Using uncertainty sampling a user is queried to correct the segmentation on a 2D plane that runs through the region the segmentation model is least confident about. They show that their method significantly reduces the required user input.

Konyushkova et al. [5] en Yifrah et al. [6] expand on the method by Top et al. however they fail to properly evaluate the method as it is. Konyushkova et al. do not require the user to delineate the entire object in the plane, only part of the plane. In doing so they get less labelled data and they need more user input than the method by Top et al. to get a high performance. Yifrah et al. only make small changes to the method by Top et al. and get similar results, however they only evaluate the method on two real-world examples. Which gives us no information about the consistency of the method.

### 1.3. Research question

In joints, like the hip joint, the distance between two bones is very small. When segmenting the bones in a joint separately it can occur that part of one bone is included in the segmentation of the other bone and the two segmentations can overlap. This problem does not occur when segmenting the two bones at the same time. Previous literature about active learning in interactive segmentation has not addressed the problem of segmenting multiple objects simultaneously. *The objective of this work is to show that interactive multi-class segmentation in medical volume scans becomes more efficient using active learning ideas to recommend where to provide correction.*

To address this objective this work has been split into five parts that are described below. The method as proposed by Top et al. has been implemented and extended to work for multiple bones.

1 . Because the method starts with segmenting a random plane, it is important to know *how the first random plane influences the performance.*

2 . Previous literature showed that correcting planes based on the uncertainty of the volume leads to a higher accuracy compared to correcting random planes [3][4][6][5]. To investigate the efficiency, these papers use random sampling as a baseline that does not rely on active learning. *This paper compares the segmentation accuracy of planes suggested based on the uncertainty with random planes.*

3 . However even if the performance is better compared to random planes it is still not known whether this is the best possible plane to correct. To get an idea of how well a method that suggests planes can perform it is investigated if *the most uncertain plane leads to the highest increase in accuracy.*

4 . Because the total uncertainty of the proposed segmentation is based on different terms it is important to know *what the influence of each term is.* Does changing the weights of the terms lead to a difference in accuracy.

5 . With the extension to multiple bones a new multi-class uncertainty term is introduced, which focuses on the boundary between the two bones. To determine if this term is useful the *influence of the multi-class uncertainty term on the accuracy of the segmentation is investigated.*

1 and 2 are required to show that active learning does in fact make the process more efficient. 3 gives a theoretical upper bound of what active learning could provide to the efficiency of the segmentation, showing how far the proposed method lies from the optimum. Because the method has not been used for multi-class segmentation as far as the author knows, 5 is used to investigate this new area. Finally, 4 will try to reduce the possible bias that results from using the Random Walker segmentation technique. Together, they provide an answer to every part of the main research objective as given in the beginning of this section.

## 2. Materials and Methods

This section will give an overview of the data and method used for this paper. First a description of the available data is presented. Then a description of the segmentation



method is given. After which the theory of the evaluated method is described. Lastly the additions to that method are explained.

### *2.1. Data Description*

The focus of this project is on segmentation of the hip joint. The hip joint consists of two bones, the hip bone and the femur. The data consists of a set of annotated CT scans. Annotation of the dataset was done interactively, where an automatic segmentation is checked and corrected by an expert. After correction each scan is also checked by two different experts to minimize inter-expert variability. The annotations by the experts are taken as the ground truth segmentation. The hip bone and the femur are segmented as two different bones. The difficulty in this dataset lies in the region of the joint gap. In this region the hip bone and the femur are close together and the boundary between the two based on the scan data can be difficult to determine.

For the CT scans in this research the spacing is set to 0.5 mm/pixel in the x- and y-direction and 1.5 mm/pixel in the z-direction. The scans are of the area around the left or right hip joint, containing the hip bone and the upper extremity of the femur. The dataset is made up of 13 CT scans of both male and female hips of different ages.

### *2.2. Segmentation Method*

The method described in this paper is an interactive segmentation method. Instead of letting the user pick where to correct the segmentation, the computer determines where the user should provide corrections. To determine where input is most valuable uncertainty sampling is used. Uncertainty sampling is a query framework in active learning, which requires a user to label the points about which the current model is least confident. In this case the uncertainty is calculated based on the image intensity gradient, the confidence of the segmentation algorithm and the image intensity of fore- and background.

An overview of how the implemented algorithm works is given in Figure 1. First the computer proposes a segmentation, which is done by classifying each voxel in the scan as femur, hip bone or background. Then based on this segmentation the uncertainty of the proposed segmentation is calculated. Next as is done in Active Learning the computer queries the user to segment or label the plane where user input is most valuable. This process is repeated until the user accepts the segmentation.

To determine the uncertainty, the described method requires a user to first segment a random plane. Then, using that as seed, the segmentation of the entire volume is calculated, shown in Figure 2. Based on the uncertainty of the segmentation a new plane is suggested and corrected by a user, this is then used as seed in addition to the first plane. After which the automatic segmentation is calculated again. This is repeated until the user is satisfied with the result. The segmentation for multiple bones is done simultaneously. The segmentation algorithm labels each voxel as background or one of the bones. A voxel can only have one label.

Top et al. propose to use the random walker algorithm for segmentation. In this paper the random walker algorithm as implemented by scikit-image [7] is used. The random walker segmentation algorithm is a probabilistic classification algorithm. The collection of voxels

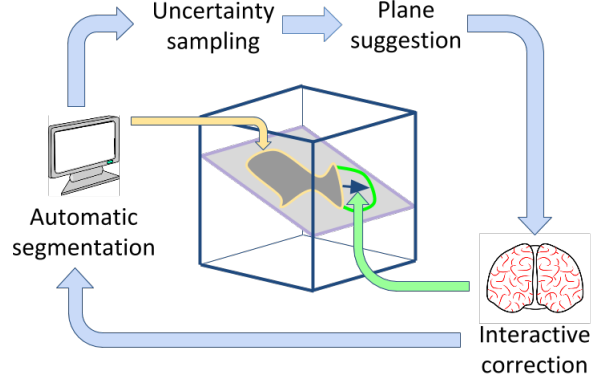


Figure 1: Overview of the plane suggestion method, automatic segmentation is alternated with human input on a plane picked by the algorithm.

in the scan are modeled as a weighted graph, each voxel is a node connected to its neighbors by an edge. The weights of the edges represent the similarity in image intensity between the voxels. Each unlabeled voxel can be imagined to release a random walk and the probability is calculated that the walk reaches a seed bearing each label. The unlabelled voxels are then classified as belonging to the seed corresponding to the walk with the highest probability. An advantage of a probabilistic segmentation algorithm in an Active Learning approach is that the uncertainty can be assessed by using the entropy of the segmentation result.

### 2.3. Active Learning in Interactive Image Segmentation by Suggesting Planes for Correction

Top et al. [3][4], suggest a method for more efficient segmentation in 3D medical scans using uncertainty sampling. Usually when correcting segmentations in 3D scans a user has to choose a plane on which to make corrections. Top et al. propose to suggest planes, on which a user makes corrections, based on regions of uncertainty. This is done to alleviate the user from choosing what the best place for corrections is. The input is used to expand the model and the automatic segmentation is repeated. The method as implemented by Top et al. only works when segmenting a single bone. For all voxels the random walker segmentation algorithm returns a proposed label  $y$  with value zero for background and one for bone. The uncertainty at a voxel at position  $\mathbf{x}$  is a weighted sum, with weights  $w_i$ , of

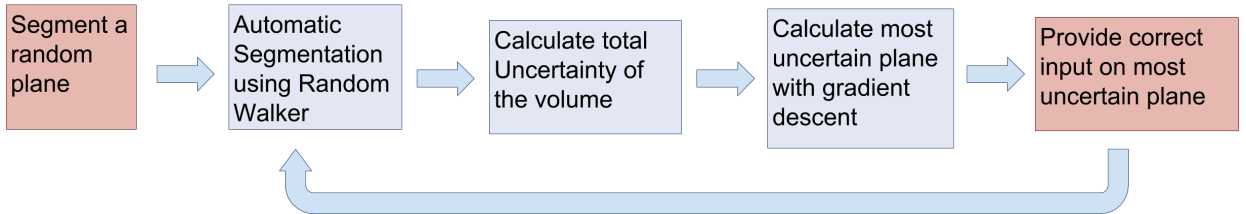


Figure 2: Diagram of the implemented method, in blue computations by a computer and in red segmentations made by a user.

four heuristics that each try to model the uncertainty in the scan,

$$U(\mathbf{x}) = w_E U_E(\mathbf{x}) + w_B U_B(\mathbf{x}) + w_R U_R(\mathbf{x}) + w_S U_S(\mathbf{x}), \quad (1)$$

$U_E$  depends on the probability of a point having a certain label, and is called the entropy energy. The boundary uncertainty  $U_B$  depends on the image intensity gradient, the regional uncertainty  $U_R$  depends on how well the voxels in- and outside the current distribution conform to the distribution of fore- and background from the training data, and  $U_S$  is the uncertainty due to smoothness of the scan in that point. The entropy uncertainty depends on the probability of a voxel at  $\mathbf{x}$  having the proposed label  $y(\mathbf{x})$  equal to the label of the foreground  $y = 1$ ,  $p_1(\mathbf{x}) = p(y(\mathbf{x}) = 1)$ .  $p_1(\mathbf{x})$  is the probability of label one as determined by the random walker algorithm. For two labels the entropy uncertainty can be calculated with,

$$U_E(\mathbf{x}) = -p_1(\mathbf{x}) \log_2(p_1(\mathbf{x})) - (1 - p_1(\mathbf{x})) \log_2(1 - p_1(\mathbf{x})). \quad (2)$$

The calculation for more than two labels is explained in Subsection 2.4, where the regional uncertainty for more than two labels is also defined. The logarithm in base two is denoted by  $\log_2$ . The voxels closer to the decision boundary will have a higher entropy uncertainty than voxels further away from the decision boundary, because the random walker algorithm is least certain about the voxels closer to the decision boundary.

The boundary uncertainty is inversely proportional to the image intensity gradient  $\nabla I(\mathbf{x})$ ,

$$U_B(\mathbf{x}) = \delta(D_s(\mathbf{x})) \frac{1}{1 + |\nabla I(\mathbf{x})|^2}, \quad (3)$$

where  $D_s(\mathbf{x})$  gives the distance from voxel  $\mathbf{x}$  to the segmentation boundary.  $\delta$  is the function  $\delta(z) = e^{-\frac{z^2}{2}}$ . This is used here so that only points close to the segmentation boundary are considered and not all points in the image. Because the boundary uncertainty is inversely proportional with the intensity gradient, the boundary uncertainty will be high for low gradients and weak edges, and low for high gradients and strong edges.

The regional uncertainty is a measure for how well a voxel inside or outside the proposed segmentation conforms to a respective intensity distribution. The intensity distributions are derived using the intensity of the already labelled voxels in the scan. The foreground and background distributions are given as Gaussian distributions:  $p(I(\mathbf{x})|Y = n) \sim \mathcal{N}(\mu_n, \sigma_n)$ .  $\mu_n$  and  $\sigma_n$  are the mean and standard deviation of the intensity of the voxels with label  $n$  in the set of manually labelled voxels. The regional uncertainty is the probability that trained label  $Y$  of a voxel is equal to the proposed label of that voxel  $y(\mathbf{x})$  given the intensity of that voxel  $I(\mathbf{x})$ ,

$$U_R(\mathbf{x}) = p(Y = y(\mathbf{x})|I(\mathbf{x})) = \frac{p(I(\mathbf{x})|Y = y(\mathbf{x}))}{p(I(\mathbf{x})|Y = 0) + p(I(\mathbf{x})|Y = 1)}. \quad (4)$$

The smoothness uncertainty is based on the surface area of the volume of the segmentation in the neighborhood  $N_{\mathbf{x}}$  of voxels surrounding point  $\mathbf{x}$ . It is assumed that smooth surfaces have

less area and therefore the surface area will be lower in that neighborhood. The smoothness uncertainty is given by

$$U_S(\mathbf{x}) = \iiint_{N_{\mathbf{x}}} \delta(D_s(\mathbf{x})) dV. \quad (5)$$

The total uncertainty of a plane is then calculated by taking the surface integral of the combined uncertainty. An estimation of the plane of highest uncertainty is determined by using the gradient descent method, which maximizes the surface integral. The gradient descent method is an optimization method that optimizes an objective function. Here the objective function is the function for the total uncertainty of a plane. It depends on the coordinates and rotation of the plane and is optimized to get a plane with a high uncertainty. The gradient descent method takes a randomly oriented plane as input. Then a step is taken proportional to the gradient of the function for the total uncertainty of a plane. Because there can be many optimal planes, this is repeated with different randomly oriented planes as input. The gradient descent method as implemented by SciPy [8] is used.

#### 2.4. Extension to Multi-Class Segmentation

The problem as described by Top et al. is only formulated for situations where you want to segment one object in a scan. To be able to work with more bones, the equations that make up the total uncertainty should be adapted. In the case of two bones this means that a voxel can have one of the labels background, first bone, or second bone. The entropy uncertainty, Equation 2, can be rewritten for  $m$  labels as

$$U_E(x, p_m(x)) = - \sum_{k=0}^m p_k(x) \log_m p_k(x). \quad (6)$$

Using Bayes' theorem the regional uncertainty, Equation 4, can be rewritten into

$$U_R(x, y) = p(Y = y(x) | I(x)) = \frac{p(I(x) | Y = y(x))}{\sum_{k=0}^m p(I(x) | Y = k)}. \quad (7)$$

The other two uncertainty terms do not depend on the label of the segmented object but on the distance to the boundary of the object. This distance is taken as the distance to the closest object.

#### 2.5. Additional Multi-Class Uncertainty

When segmenting two bones in a volume a difficult part of the segmentation is the region where the two bones are closest to each other. Therefore a new uncertainty term is introduced which focuses on the boundary between the two objects. For all voxels in the scan the uncertainty  $U_{multi-class}$  counts the number of voxels of the first bone,  $n_a$ , and the number of voxels of the second bone,  $n_b$ , in a neighborhood of 5 by 5 by 5 voxels.  $U_{multi-class}$  is equal to the sum of  $n_a$  and  $n_b$  minus their absolute difference and divided by the total number of voxels  $n_{total}$  in the neighborhood.

$$U_{multi-class} = \frac{(n_a + n_b) - |n_a - n_b|}{n_{total}} \quad (8)$$

The multi-class uncertainty can have a value between zero and one and is highest when  $n_a$  and  $n_b$  are equal and there is no background in the neighborhood of the voxel. It is lowest when the neighborhood only contains background or one of the bones. This term cannot easily be extended to work for more than two bones.

### 3. Experimental Setup

To reduce effects caused by users getting to know a system, a human user is simulated by a computer. Instead of presenting a user with a slice to segment, that slice is taken from the ground truth segmentation and used as corrected input.

It was noted early on in the project that the initial random plane has to go through both bones. If this does not happen it can take a few planes before a plane is sampled that goes through the second bone, this is illustrated with an example in the appendix see Figure A.10.

The weights used for the different terms are determined by scaling with the average voxel value per uncertainty term. The average voxel value is calculated for 20 random planes per volume and for 10 volumes. These are different CT-scans than the ones used in evaluation. The uncertainty terms depend on the segmentation, therefore in this case the uncertainty is calculated after the first random plane suggestion. As can be seen in Table A.1 the average voxel value per plane differs a lot for the different uncertainty terms. Scaling with these values results in the following weights,  $w_E = 3$ ,  $w_R = 1$ ,  $w_S = 52$ ,  $w_B = 27$ . These weights are used in the evaluations described in the result section, except in the comparison with the multi-class uncertainty where  $w_{MC} = 1300$ , which was determined at the same time as the other weights. Using these weights gave much better results than using the weights as described in the plane suggestion paper [4].

The results for the femur and hip bone are presented separately, however the segmentation is done for both bones in the same scan at the same time. The overlap accuracy  $\Omega$  is calculated as the intersection over the union of the ground truth segmentation  $S_g$  of femur or hip bone and the segmentation result  $S_r$  of femur or hip bone. The error  $\epsilon$  is calculated as one minus the overlap accuracy.

$$\Omega = \frac{S_g \cap S_r}{S_g \cup S_r}, \quad \epsilon = 1 - \Omega \quad (9)$$

The overlap accuracy is used to inspect if the segmentation is globally correct. It is calculated per bone to be able to assess in which bone mistakes are made.

In addition to the overlap accuracy, the Hausdorff distance is used to inspect if there are local mistakes. The Hausdorff distance is defined as the maximum of the closest distances between the segmentation result and the ground truth. It is calculated two-sided, and for the hip bone and femur combined.

#### 3.1. Experiments

To address the five points formulated in Subsection 1.3, a set of experiments is performed. The influence of the initial plane, the first point, is already addressed in the previous part

of this section. The first three experiments address points 2 and 3, they create a window of comparison for the uncertainty sampling method. Point 4 addresses the influence of the different terms and is handled in the fourth experiment. The last experiment deals with the multi-class uncertainty term and addresses point 5.

#### *Comparison to a Baseline*

As a baseline that does not rely on active learning the overlap accuracy for segmentation using random sampling is calculated. Random sampling means that, instead of selecting planes based on their uncertainty, planes with a random orientation are selected. Uncertainty sampling and random sampling are compared based on the overlap accuracy versus the number of corrected planes.

#### *Comparison to a Upper Bound*

As an upper bound the overlap accuracy is calculated for planes that lead to the segmentation with the highest increase in accuracy. The implemented method can present the most uncertain plane, however this does not have to be the plane that leads to the highest increase in accuracy. The plane that leads to the highest increase in accuracy is calculated in a brute-force kind of way. First a random plane is suggested that passes through both bones, this is done to have a fair comparison with the implemented method. Then the overlap accuracy is calculated for a plane that is rotated and translated through the volume. In total the overlap accuracy is calculated for 225 planes per volume.

The plane that leads to the highest average overlap accuracy is used to update the seed-volume for the next plane suggestion. The overlap accuracy of the implemented method is compared to the overlap accuracy of the best plane suggestion. The best possible suggestion is calculated for three consecutive plane suggestions due to the high computation time necessary. The overlap accuracy versus the number of planes suggested for the best suggestion will be compared to that of using uncertainty sampling.

#### *Irregularities in the Segmentation*

In addition to the hip bone and femur segmentation the ground truths also contain part of the sacrum, which is supposed to be labeled as background. This can be seen in the left part of Figure 7a. In the previous experiments this part is cropped out of the scan. To see how the method reacts with these types of inconsistencies the comparison to random sampling and the best suggestion is repeated without this part cropped out.

#### *Local Errors*

The boundary and smoothness terms are only defined locally in the neighborhood of the segmentation surface, while the entropy and regional terms are defined globally over the entire volume. Therefore, even though the terms are scaled to have values in the same range, since the total uncertain of a plane is calculated as the sum over all values in that plane, the global terms will have a larger influence. Figure 3 compares the values of the different uncertainty terms on a plane, it shows that the regional and entropy terms cover a larger area. To analyze if local errors, like spikes, are also reduced when more planes are corrected the Hausdorff distance is calculated and compared to the error of uncertainty sampling.

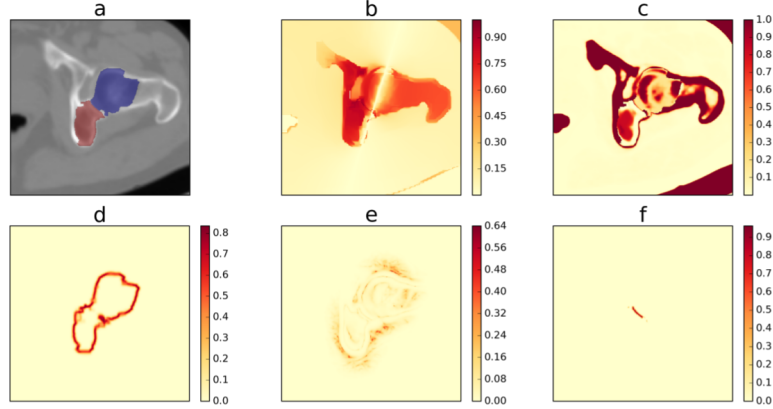


Figure 3: Comparison of the values of the uncertainty terms on the plane with suggested segmentation in (a). (b), (c), (d), (e), (f) are respectively the entropy, regional, smoothness, boundary and multi-class uncertainty terms.

#### *Multi-Class Uncertainty Term*

To see if the multi-class uncertainty term influences the accuracy, the segmentation is calculated and planes are suggested based on the total uncertainty with the multi-class uncertainty term added,  $w_{mc} = 1300$ . The overlap accuracy has been observed to converge after approximately four plane suggestions. Because of that the overlap accuracy of the segmentation with and without the multi-class uncertainty will be compared after four plane suggestions.

## 4. Results

To evaluate the presented method a series of experiments were performed. The following section describes the results obtained from those experiments. Most importantly the results show that the presented method of uncertainty sampling outperforms random sampling.

In Figure 4 uncertainty sampling is compared to random sampling. The overlap accuracy using uncertainty sampling increases faster than random sampling. For the hip bone the difference between random sampling and uncertainty sampling is larger than for the femur. After five plane suggestions the overlap accuracy of the femur shows only a small difference between random sampling and uncertainty sampling.

Figure 4 also compares with corrected planes that lead to the highest increase in accuracy. The overlap accuracy of uncertainty sampling is very close to the best suggestion. Again the difference for the hip bone is larger than that for the femur. For all sampling methods the median overlap accuracy is lower for the hip bone than for the femur.

Figure 5 shows the Hausdorff distance of the segmentation of hip bone and femur together and the average error of the femur and hip bone segmentation. The error, one minus the overlap accuracy, converges at five corrected planes. The Hausdorff distance shows a large drop at seven corrected planes. Furthermore even though the average error is reduced to 0.075 without outliers, the Hausdorff distance drops to 10 mm with outliers of 40 mm.

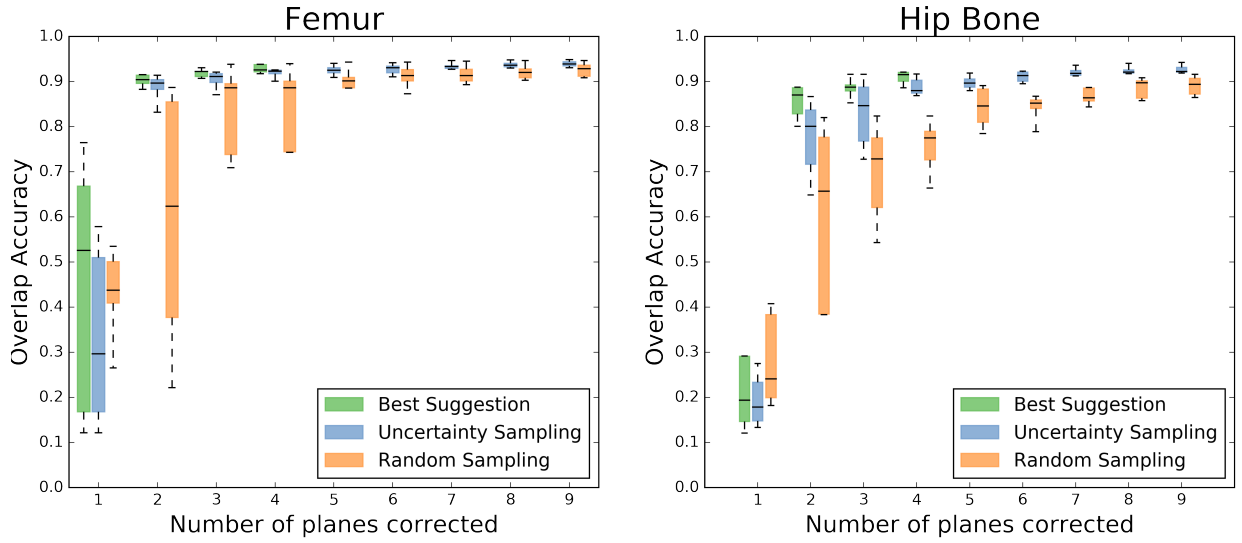


Figure 4: Accuracy of the segmentation, as a function of the number of planes suggested for multiple volumes. Uncertainty sampling in blue performs better than random sampling in orange and performs worse than the best suggestion in green. The ends of the whiskers represent the 5<sup>th</sup> and 95<sup>th</sup> percentile, the black middle line represents the median and the bottom and top of the box the first and third quartiles.

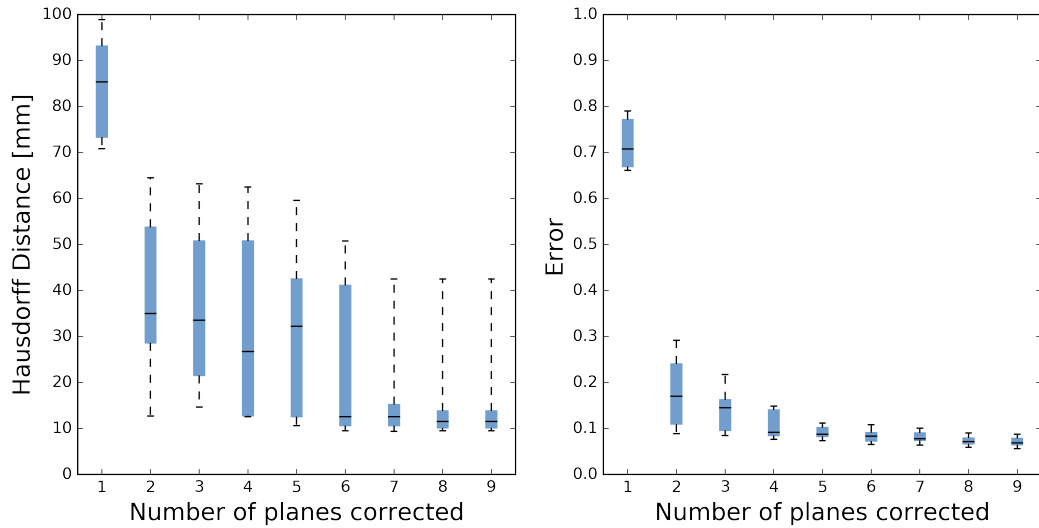


Figure 5: On the left the overall Hausdorff distance of both femur and hip bone is compared to the number of planes corrected for multiple volumes. On the right the error (one minus the overlap accuracy) versus the number of planes corrected. The error converges faster than the Hausdorff distance. The ends of the whiskers represent the 5<sup>th</sup> and 95<sup>th</sup> percentile, the black middle line represents the median and the bottom and top of the box the first and third quartiles.



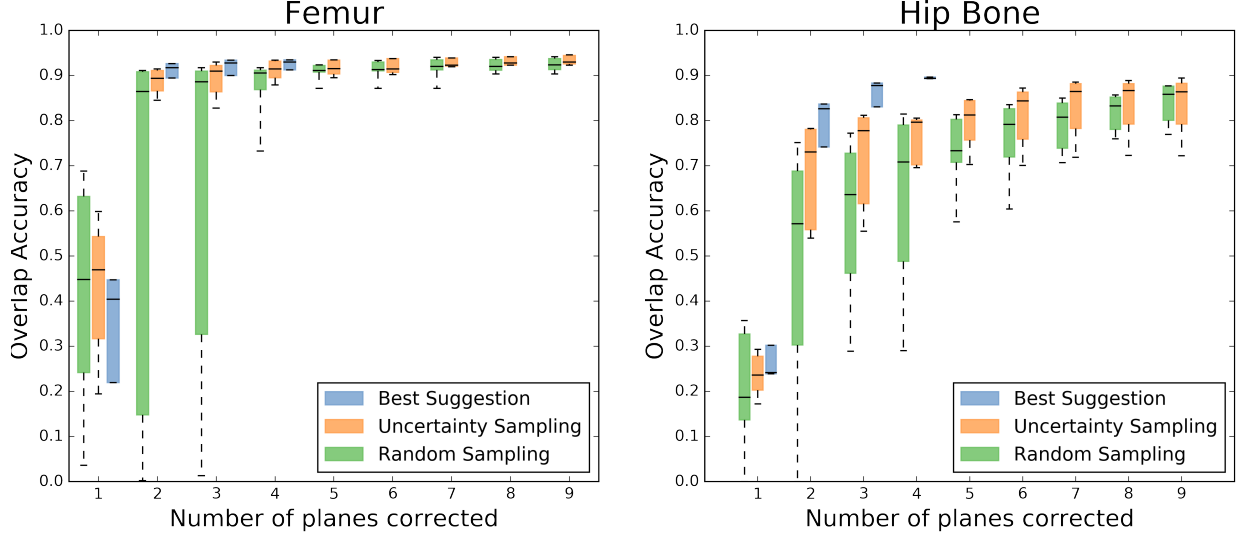


Figure 6: Overlap accuracy of the segmentation as a function of the number of planes suggested. Uncertainty sampling in orange is compared to random sampling in green with a lower performance, and the best suggestion in blue with a higher performance. The best suggestion is only shown for 3 suggested planes, random and uncertainty sampling are shown for 8 suggested planes. The ends of the whiskers represent the 5<sup>th</sup> and 95<sup>th</sup> percentile, the black middle line represents the median and the bottom and top of the box the first and third quartiles.

Figure 6 shows the overlap accuracy versus the number of planes suggested, for volumes where part of the sacrum is included in the hip bone segmentation. Uncertainty sampling is compared to random sampling and the best suggestion. The figures show a faster increase in overlap accuracy when comparing uncertainty sampling with random sampling. The best suggestion shows that it is still possible to get a faster increase in accuracy. The segmentation of the femur reaches a higher overlap accuracy than that of the hip bone, and the difference between the three sampling methods is lower for the femur than for the hip bone.

The difference between hip bone and femur, in Figure 6, is partly caused by the addition of part of the sacrum in the hip bone segmentations. Initially the random walker classifies the area of the sacrum as background. However uncertainty sampling does end up sampling a plane in that region, and the final segmentation includes that region in the hip bone segmentation, which can be seen in Figure 7 b-c. This illustrates why in Figure 6 the hip bone needs more plane suggestions to reach a high overlap accuracy.

Inspecting the segmentations it could be seen that the large Hausdorff distances often occur in the joint gap, as spikes. It is suspected that when only using local uncertainty terms the Hausdorff distance will decrease further. To see if this suspicion has any ground a small experiment was performed. On the left of Figure 8 the segmentation after four iterations is shown with the same weights used in all evaluations. Here you can see a large spike at the top, this is in the joint gap. On the right the result is shown for adding four more plane suggestions to the segmentation in (a), but with the weights of the regional and entropy terms set to zero.

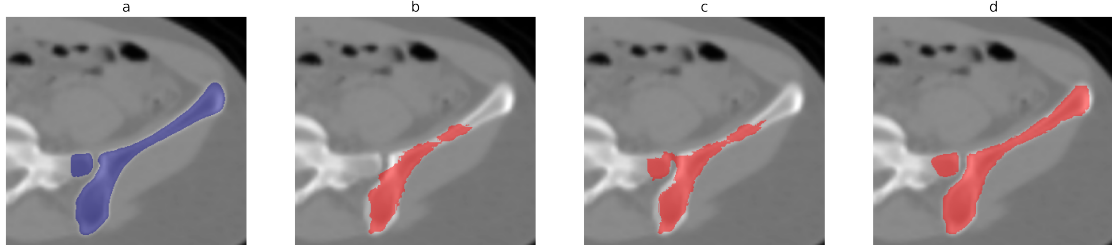


Figure 7: Slice of a CT scan where, in addition to the hip bone, part of the sacrum on the left is also labeled as ground truth. (a) shows the ground truth in blue. (b), (c) and (d) show the calculated segmentation after respectively 4, 6, and 8 corrected planes.

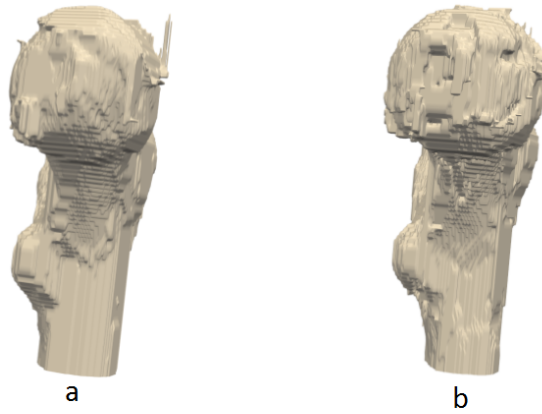


Figure 8: (a) is the segmentation of the femur after four iterations, the sharp spike sticking out at the top gives a high Hausdorff distance without much difference in the overlap accuracy. (b) is the segmentation after four more iterations with only local uncertainty terms, the sharp point disappears decreasing the Hausdorff distance.

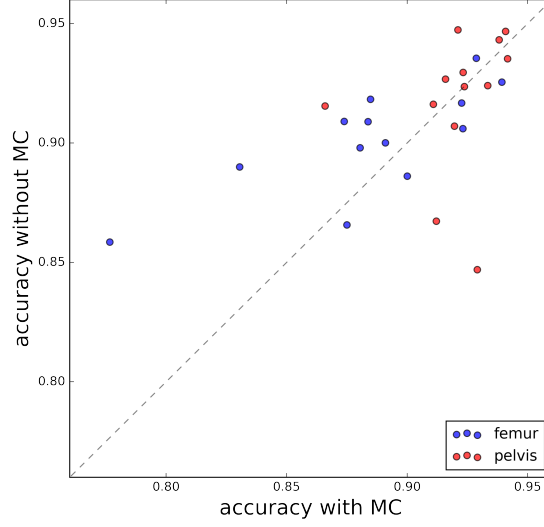


Figure 9: Overlap accuracy of the region of interest of the segmentation after four plane suggestions with multi-class uncertainty weight set to 1300 on x-axes and with multi-class weight set to zero on the y-axis. The different dots are different volumes.

Lastly, Figure 9 shows the influence of introducing the multi-class uncertainty term. The overlap accuracy with the multi-class weight set to zero and with the multi-class weight set to 1300 are compared after five corrected planes for multiple scans. The overlap accuracy is not calculated over the entire scan but only in the region of interest, around the joint gap. Because the multi-class uncertainty term focuses on the boundary between two bones, it is expected to lead to better results in the region of interest. Overall the figure shows that adding the term does not increase or decrease the overlap accuracy, there are some outliers on both sides.

## 5. Discussion

This research has evaluated a method for suggesting planes for interactive segmentation and expanded it to be able to segment multiple bones at the same time. The main objective of this paper is to show that interactive multi-class segmentation in medical volume scans becomes more efficient using Active Learning ideas to recommend where to provide correction. From the results it can be concluded that the segmentation is more efficient compared to random sampling, however a better performance is possible.

### 5.1. Discussion of results

We cover the five parts of the main objective as described in Subsection 1.3.

1 . It was noted in Section 3 that when the initial random plane does not go through both bones it takes long before the second bone is included in the segmentation. It is expected that the cause of this is that the uncertainty after each plane suggestion is determined compared

to the proposed segmentation at that moment. Points closer to the segmentation surface will have a higher boundary and smoothness uncertainty, and therefore uncertainty sampling is not likely to sample far away from the current segmentation surface. *Initialization using a random plane has a large influence on the performance when segmenting multiple bones.*

2 . Figure 4 and Figure 6 show that for the femur after eight corrected planes random sampling reaches the same overlap accuracy that uncertainty sampling reaches after three corrected planes. This reflects the results in the paper by Top et al., where in most cases the overlap accuracy for random and uncertainty sampling are comparable after eight corrected planes. For the hip bone the overlap accuracy is still lower at eight corrected planes. This is because the hip bone has a more complex shape and a plane in the right place will lead to a higher increase in overlap accuracy. However the question can be raised if comparing to random sampling makes sense, since a normal user would not select random planes. In random plane suggestion there is a chance that two subsequently suggested planes are very close to each other. In that case the newest plane will not add much information for the segmentation algorithm and the overlap accuracy will not increase much. Instead of comparing to random sampling, uncertainty sampling could also be compared to a user study or a comparison with predetermined equally distributed planes. *Uncertainty sampling leads to an increase in efficiency compared to random sampling; fewer plane suggestions are needed to reach the same overlap accuracy.*

3 . Figure 4 and Figure 6 compare uncertainty sampling with the best suggestion. Both figures show a large difference in overlap accuracy between the hip bone and the femur. The femur reaches a higher overlap accuracy for fewer corrected planes. This can be explained by the fact that more planes are required to describe the shape of the hip bone. The femur can get a lot of information from a plane that goes through the shaft from top to bottom, while for the hip bone it is much more difficult to find such a plane.

The different areas in the hip bone can be very thin, sometimes the size of one voxel, and have smoother edges than the femur. Because areas in the hip bone can be very thin the random walker algorithm will have a harder time than in the solid simple shaped femur. The intensity in a thin part of the bone has less resemblance to the average intensity of bone, this results in a lower regional uncertainty for wrongly classified voxels in the hip bone than for wrongly classified voxels in the femur. The femur has sharper edges, therefore the intensity gradient is stronger on the edge of the femur and thus the boundary uncertainty is higher. Using uncertainty sampling suggested planes are more likely to pass through the less complex bone, while a plane through the more complex bone will often lead to a higher increase in accuracy. This method sets out to correct the plane with the highest uncertainty, while this does not necessarily have to be the best plane to correct. A plane through the femur can have a higher uncertainty, while a plane through the hip bone might lead to a higher increase in accuracy.

For the non-cropped scans, Figure 6, the difference between uncertainty sampling and the best suggestion in the hip bone is larger than for the cropped scans, Figure 4. This is because using uncertainty sampling it takes time to include the sacrum, in the best suggestion it is included faster. This does not mean that there is no benefit in using uncertainty sampling; the results are still close to the best suggestion and by better tuning of the weights and

parameters of the random walker the gap between the best suggestion and uncertainty sampling can be decreased for both types of volumes. With the current parameters *correcting the most uncertain plane does not always lead to the highest increase in overlap accuracy*.

4 . There is a large difference in the average voxel values of each uncertainty term. The values of regional and entropy uncertainty are much higher than those of boundary and smoothness, as described in Section 3. Therefore if the weights of the uncertainty terms are set equal, regional and entropy uncertainty will have more influence on the total uncertainty. This corresponds with Top’s findings that the regional and entropy uncertainty have the largest influence. Scaling the uncertainty terms so that they all have the same influence on the total uncertainty, shows a fast increase in overlap accuracy.

Figure 5 shows that even though the average error converges the Hausdorff distance is still high at six corrected planes. A segmentation can have large Hausdorff distance and at the same time a low error, for instance when there is a large spike in the segmentation. From Figure 5 it can be concluded that even though the overall segmentation is accurate, using this method there are still large local mistakes. For the segmentation to be corrected in that area a plane needs to be suggested exactly there. But, because the regional uncertainty dominates the total uncertainty, the total uncertainty is not high in that area. The regional uncertainty is based on the probability that the suggested label is the correct label given the intensity of the voxels. Because the intensity in the joint gap is similar to that of bone, the regional uncertainty is lower than in other regions and therefore the total uncertainty is low as well. It is suspected that these large Hausdorff distances are reduced by focusing on the local uncertainty terms. This speculation is illustrated by the example in Figure 8. *The different uncertainty terms can be divided in ones that influence the segmentation globally and ones that influence the segmentation locally*.

5 . Figure 9 shows the effect of adding the multi-class uncertainty. It was expected that the extra term would yield a higher accuracy in the joint gap area. However the figure shows that *adding the multi-class term has little impact on the accuracy*. The multi-class term adds extra uncertainty in the area where the two bones meet, therefore it is more likely that a plane is suggested in that area. In CT-scans the intensity in the joint gap is very similar to that of the hip bone and the femur. The edge between bone and joint gap can therefore be very soft. The multi-class term does succeed in suggesting more planes through the joint gap. However the random walker segmentation does not succeed in correctly segmenting the entire edge between the bone and joint gap, just the area in the vicinity of the corrected plane.

## 5.2. Limitations

The method as it is described here requires every voxel in the scan to have a value for the uncertainty. Furthermore to calculate the entropy as it is defined in this paper each voxel needs a probability of having a certain label, as determined by the segmentation algorithm. The entropy uncertainty can only be calculated if the segmentation algorithm is a probabilistic classification algorithm. If this active learning method is implemented with a different segmentation algorithm it needs to be able to return a probability for every voxel.

One of the limitations of using a voxel labelling segmentation algorithm is that by labelling voxels you will get a more jagged segmentation. The ground truth segmentations are made manually and are very smooth. This means that a jagged segmentation is compared with a smooth segmentation, therefore the accuracy will never reach one. If you want to get a segmentation that is accurate enough for clinical use, a shape based segmentation will most likely be an improvement. But shape based algorithms do not return a probability per voxel. Which means that the entropy uncertainty cannot be used. If the entropy uncertainty is left out in the total uncertainty the regional uncertainty is the only global term, and will likely dominate the total uncertainty.

This method gives the most advantage when the initial segmentation is poor. When initially just a small part of both bones is segmented only a few planes are needed to segment a larger part of the bones, which will lead to an large increase in accuracy. However after the first big jump the accuracy converges fast. If the segmentation only has multiple small errors left this active learning method will not help much anymore, because a plane needs to go through all locations where errors are.

## 6. Conclusions

In this paper a method for suggesting planes for correction in 3D image segmentation was evaluated and extended to work for segmenting two objects instead of one. It was found that by suggesting planes for correction based on the uncertainty of the segmented volume the segmentation becomes more efficient. If there are no irregularities in the ground truth the largest increase in accuracy is after a user has corrected three planes suggested based on the uncertainty. After three plane suggestions only a slight increase in accuracy can be observed. The overlap accuracy does not reach one, which means that compared to a ground truth the segmentation is not 100% correct. When segmenting two bones, the first randomly picked plane must go through both bones to lead to a fast and accurate result. It was also found that correcting planes based on uncertainty sampling resulted in a better performance than correcting planes based on random sampling. However there are planes that lead to a higher increase in accuracy than those based on uncertainty sampling do. Because the boundary and smoothness uncertainty terms are defined locally in the neighborhood of the segmentation surface they have less influence on the most uncertain plane than the regional and entropy uncertainty terms. The addition of the multi-class uncertainty term does not positively influence the accuracy of the segmentation.

## Appendix A. Appendices

Table A.1: Mean and standard deviation of the average voxel values of random planes, for different uncertainty terms. Calculated for 12 volumes and 20 random planes per volume.

	Mean	Standard Deviation
Entropy	0.091	0.043
Regional	0.30	0.118
Boundary	0.0057	0.0058
Smoothness	0.0107	0.013
Multi-class	0.00023	0.00077

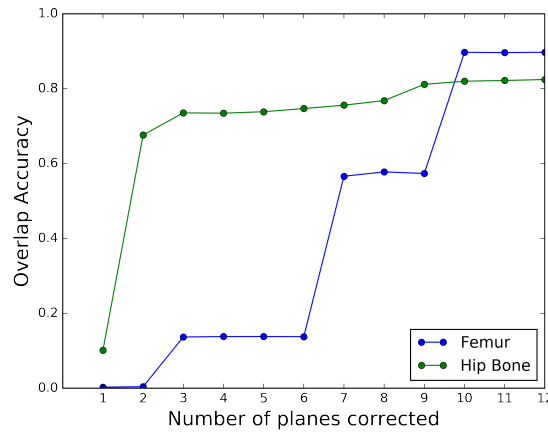


Figure A.10: Overlap accuracy versus the number of planes corrected for one volume where the first plane did not pass through the femur. It takes 10 corrected planes for the femur to pass the overlap accuracy of the hip bone.

## References

- [1] S. D. Olabarriaga, A. W. Smeulders, Interaction in the segmentation of medical images: A survey, *Medical image analysis* 5 (2) (2001) 127–142.
- [2] J. Iglesias, E. Konukoglu, A. Montillo, Z. Tu, A. Criminisi, Combining generative and discriminative models for semantic segmentation of CT scans via active learning, in: *Information Processing in Medical Imaging*, Springer, 2011, pp. 25–36.  
URL <http://www.springerlink.com/index/j44865300544r462.pdf>
- [3] A. Top, Automated confidence-based user guidance for increasing efficiency in interactive 3d image segmentation, Thesis, Applied Science: School of Computing Science (Jan. 2012).  
URL <http://summit.sfu.ca/item/12668>
- [4] A. Top, G. Hamarneh, R. Abugharbieh, Active Learning for Interactive 3d Image Segmentation, in: *Medical Image Computing and Computer-Assisted Intervention MICCAI 2011*, no. 6893 in *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, 2011, pp. 603–610, doi: 10.1007/978-3-642-23626-6\_74.  
URL [http://link.springer.com/chapter/10.1007/978-3-642-23626-6\\_74](http://link.springer.com/chapter/10.1007/978-3-642-23626-6_74)

- [5] K. Konyushkova, R. Sznitman, P. Fua, Introducing Geometry in Active Learning for Image Segmentation, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 2974–2982.  
URL [http://www.cv-foundation.org/openaccess/content\\_iccv\\_2015/html/Konyushkova\\_Introducing\\_Geometry\\_in\\_ICCV\\_2015\\_paper.html](http://www.cv-foundation.org/openaccess/content_iccv_2015/html/Konyushkova_Introducing_Geometry_in_ICCV_2015_paper.html)
- [6] S. Yifrah, E. Zadicario, T. Ju, D. Cohen-Or, An algorithm for suggesting delineation planes for interactive segmentation, in: 2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI), 2014, pp. 361–364. doi:10.1109/ISBI.2014.6867883.
- [7] S. van der Walt, J. L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, N. Yager, E. Gouillart, T. Yu, the scikit-image contributors, scikit-image: image processing in Python, PeerJ 2 (2014) e453. doi:10.7717/peerj.453.  
URL <http://dx.doi.org/10.7717/peerj.453>
- [8] E. Jones, T. Oliphant, P. Peterson, et al., SciPy: Open source scientific tools for Python (2001–).  
URL <http://www.scipy.org/>



# Workbook

Thera Pals

January 7, 2017

## 1 Introduction

This workbook is a collection of documents which I used during my thesis to keep track of the different aspects of this project.

Manually segmenting data in medical scans can be difficult and time consuming. In the paper Active Learning for Interactive 3D image segmentation by Top et al. a method for suggesting planes where correction of the segmentation of volumes is most valuable. By determining the uncertainty of a segmented volume the most uncertain volume can be calculated. The uncertainty is calculated as a weighted sum of four sub uncertainties. This method is for medical scans where only one bone needs to be segmented, I extended it to work for multiple bones. The plane suggestion method is evaluated for scans of different bones, while I am interested in the situation where you want to segment the same bone in multiple patients.

### Research Question:

- Can interactive multi-class segmentation in medical volume scans become more efficient using Active Learning ideas to recommend where to provide correction?

### Sub Questions:

- (a) What is the influence of the different uncertainty terms on the efficiency of the method?
- (b) Does uncertainty sampling lead to an increase in efficiency compared to random sampling?
- (c) Does uncertainty sampling lead to an increase in efficiency compared to fixed sampling?
- (d) Is the performance of the plane suggestion method consistent over multiple scans?
- (e) Does initialization using a random plane influence the performance?
- (f) Does correcting the most uncertain plane lead to the highest increase in accuracy?

### Contributions:

1. Extension of Top's binary uncertainty terms to support multi-class classification.
2. New uncertainty term that focuses on the boundary between two classes.
3. Validation of plane suggestion method for this specific dataset.

## 2 Algorithms

Below algorithm 1 describes the general algorithm, where a computer decides on what plane a user needs to give input to update the segmentation model.

Algorithm 1: Segmentation Method
<p><b>Data:</b> CT-Scan Volume <b>Result:</b> Segmentation of two bones A user delineates the bones on a random plane; Delineated plane is stored in empty seed volume; Random Walker classifies voxels as background, femur or hip bone; <b>for</b> <math>i = 1</math> to 8 <b>do</b>     Calculate each uncertainty term for the voxels in the volume;     Sum the uncertainties weighted, to get the total uncertainty;     <b>for</b> <math>k = 1</math> to 10 <b>do</b>         Select random plane through total uncertainty;         Find plane with high uncertainty using gradient descent optimization;         Store plane with the highest summed uncertainty;     <b>end</b>     User delineates plane with the highest summed uncertainty;     User input is added to the seed volume;     Random Walker classifies voxels as background, femur or hip bone; <b>end</b></p>

Figure 1 gives an overview of how this algorithm is implemented in this project. Instead of letting a user give input, the input is taken from the ground truth segmentation, this step is shown in the top left of the figure.

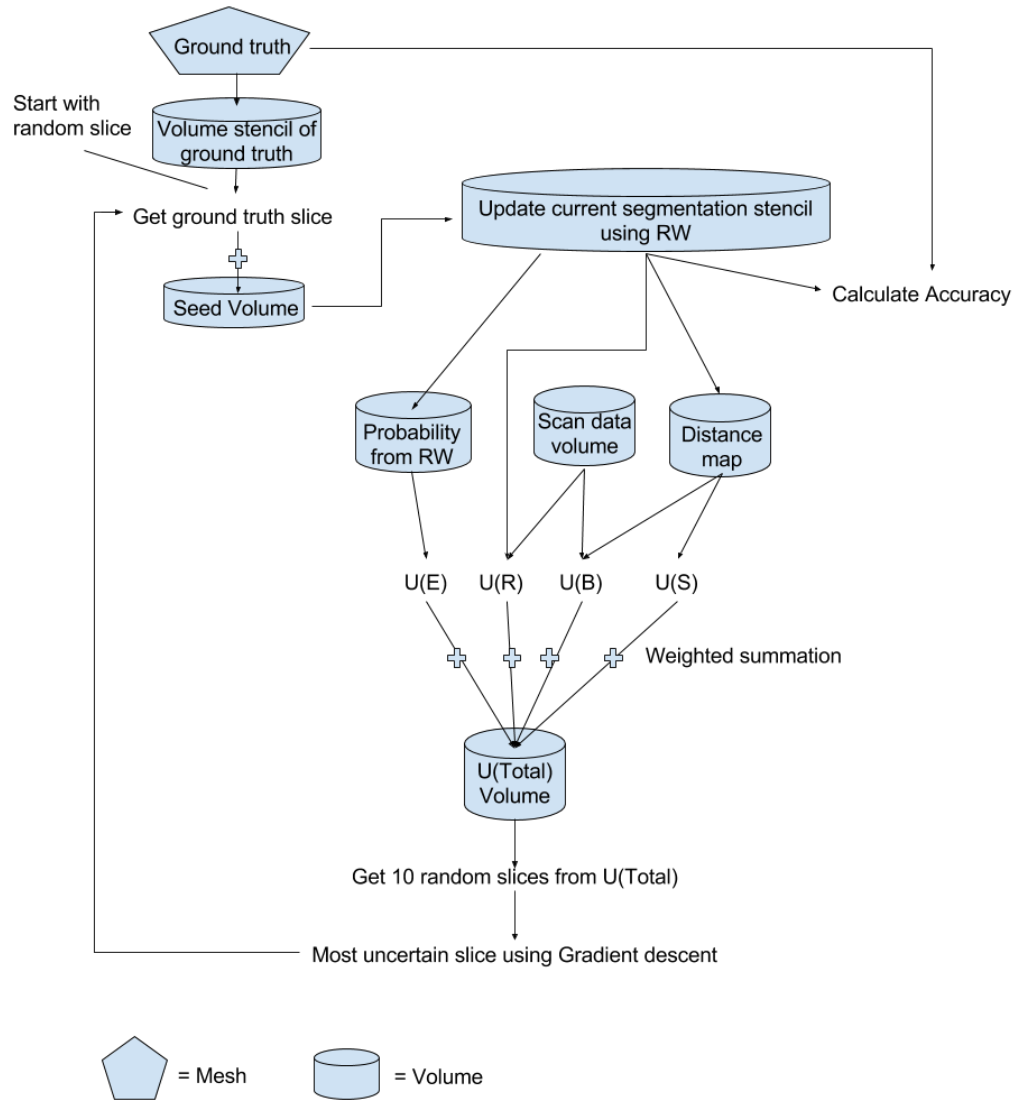


Figure 1: Overview of how the algorithm is implemented in this project. A mesh is a collection of vertices, edges and faces that make up the shape of an object. A volume is a 3D matrix where each element in the matrix represents a voxel in the volume.

## 2.1 Uncertainty terms

Next the algorithms used to calculate the uncertainty terms are given:

**Algorithm 2:** Calculation of Entropy Uncertainty.  $p_i$  is the probability of having label  $i$  according to the Random Walker algorithm.

**Data:** For all voxels the probability of having all labels from Random walker

**Result:** The entropy uncertainty for all voxels.

Entropy Uncertainty =  $-p_0 \log_3(p_0) - p_1 \log_3(p_1) - p_2 \log_3(p_2)$

**Algorithm 3:** Calculation of Regional Uncertainty.

**Data:** CT-scan, seed volume, proposed segmentation

**Result:** The regional uncertainty for all voxels.

From the CT-scan get the gray values of the voxels corresponding to the labelled voxels in the seed volume;

**for** *Voxels manually labelled with label = 1 to 3* **do**

    Collect gray value of all voxels corresponding to the label;

    Calculate the mean and standard deviation of the gray values;

**end**

**for** *All voxels in CT-scan* **do**

    Get proposed label from proposed segmentation;

    Calculate the probability of the intensity of the voxel given the proposed label of the voxel;

    prior probability = 0;

**for** *label= 1 to 3* **do**

        Calculate the probability of the intensity of the voxel given the label;

        Add to prior probability;

**end**

    To get the regional uncertainty divide the probability of the gray value given its proposed label with the prior probability;

**end**

**Algorithm 4:** Calculation of Boundary Uncertainty

**Data:** CT-scan, Proposed segmentation

**Result:** The boundary uncertainty for all voxels.

Calculate the Gaussian gradient magnitude for the CT-scan;

For all voxels determine closest distance to the proposed segmentation;

Determine the boundary uncertainty of all voxels using equation 3;

**Algorithm 5:** Calculation of smoothness uncertainty

**Data:** CT-scan, Proposed segmentation

**Result:** The smoothness uncertainty for all voxels.

**for** *All voxels* **do**

    Determine closest distance to the proposed segmentation;

    Average  $\delta(D_s(\mathbf{N}))$ , with  $\mathbf{N}$  a neighborhood of 5x5x5 voxels around the voxel;

**end**

## 3 Segmentations

As can be seen in the figure below the largest step in the accuracy of the segmentation is after the correction of the second plane. At first the objects look like random blobs and at the second step it already looks like two bones. The second corrected plane is the first intelligently picked plane, because the first plane to correct is picked randomly. After the initial rise in accuracy it takes six more corrected planes to fill up the gap in the hip bone.

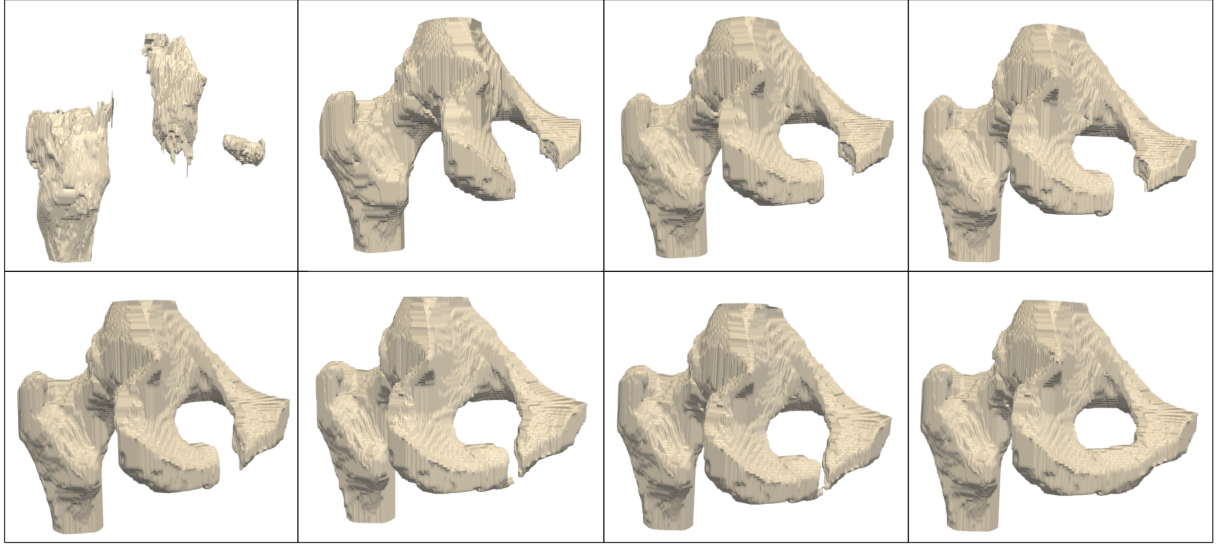


Figure 2: The segmentation of femur and hip bone after different amounts of corrected planes. The top row shows the segmentation after 1 to 4 corrected planes, and the bottom row after 5 to 8 corrected planes.

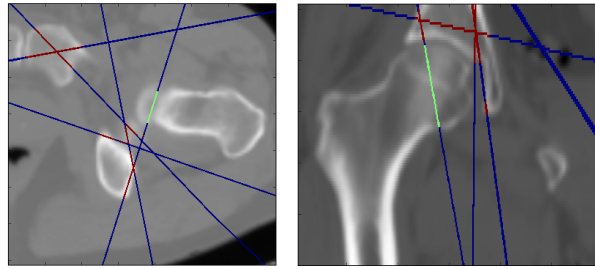


Figure 3: Slice of a CT scan of the hip joint in the xy-direction on the left and in the yz-direction on the right, with the seed-volume laid over it. The different lines are different planes in the seed-volume, blue is background, green is femur and red is hip bone.

Figure 3 shows the seed volume after five suggested planes, laid over the corresponding CT-scan. On the right of the figure it can be seen that the planes tend to pass through the joint gap. Overall the planes tend to go a bit more through the center of the scan, this is good because that is where the bones are.

## 4 Experiments

This section of the workbook describes some of the experiments that were done, some of these are also included in the paper. For each experiment it is listed what research question it helps answer, how the experiment works and what the results are.

### 4.1 Evaluation of uncertainty terms

**What research question does it help to answer?**

- (a) What is the influence of the different uncertainty terms on the efficiency of the method?

Table 1: Mean and standard deviation of the average voxel values of random planes, for different uncertainty terms. Calculated for 12 volumes and 20 random planes per volume.

	Mean	Standard Deviation
Entropy	0.091	0.043
Regional	0.30	0.118
Boundary	0.0057	0.0058
Smoothness	0.0107	0.013
Multi-class	0.00023	0.00077

## Method

To get a more general idea of how the weights relate to each other, the average uncertainty value in a plane is compared for all different uncertainty terms. All terms are normalized so that the value for individual pixels is between zero and one. The average value over a plane however can differ a lot for different terms. The sum of the boundary and smoothness terms over the plane will probably have a lower value than that of the entropy and regional terms, because the boundary and smoothness terms are evaluated in the neighborhood of the segmentation surface and can only have value one on the segmentation surface. To see how the different terms relate to each other, for random planes average voxel value on a plane per term is calculated. This is done for 20 random planes per volume and for 12 volumes.

## Results

As can be seen in table 1 the average voxel value per plane differs a lot for the different uncertainty terms.

### 4.2 Evaluation of weights

#### What research question does it help to answer?

(a) What is the influence of the different uncertainty terms on the efficiency of the method?

## Method

Once the relation between the uncertainty terms is known we explore how the weights influence the segmentation accuracy. For different weight configurations the accuracy is compared to the number of planes suggested. Keeping all weights equal will be compared with normalized weights, where the weight is equal to 1 divided by the mean uncertainty in table 1. Evaluations will be run with and without adding the new multi-class uncertainty term weighted to the rest. Evaluation is done on one volume and the same random initialization of the start plane.

## Results

Figures 4 and 5 show the accuracy versus the number of planes suggested separately for pelvis and femur. These results were obtained during the same segmentation run. After each plane suggestion the segmentation is calculated using the random walker algorithm for the femur and pelvis simultaneously.

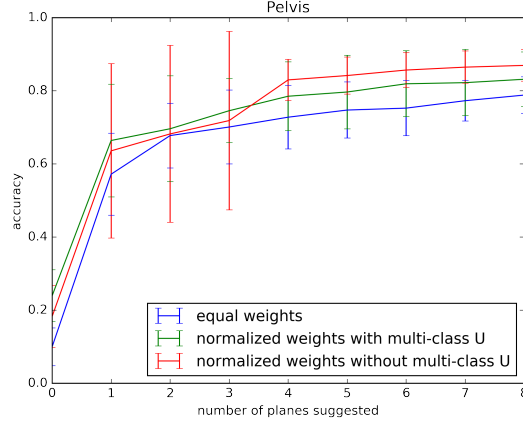


Figure 4: Accuracy vs number of planes suggested, the average and standard deviation of multiple volumes is compared for different weights in the uncertainty sampling equation.

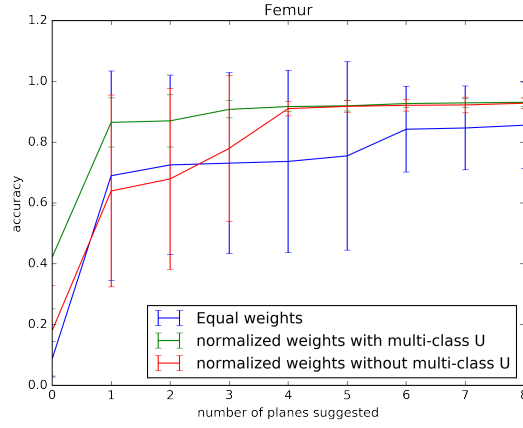


Figure 5: Accuracy vs number of planes suggested, the average and standard deviation of multiple volumes is compared for different weights in the uncertainty sampling equation.

### 4.3 Random plane suggestion

**What research question does it help to answer?**

(b) Does Active Learning in image segmentation lead to an increase in efficiency compared to random sampling.

#### Method

Compare accuracy vs number of planes suggested for random plane suggestion and uncertainty sampling. Examples of random planes are given in figure 6. Random planes will be suggested for different volumes.

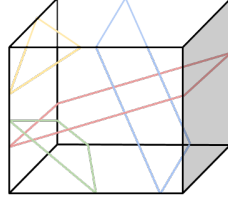


Figure 6: Example of randomly oriented planes, in black the volume and in color the planes.

## Results

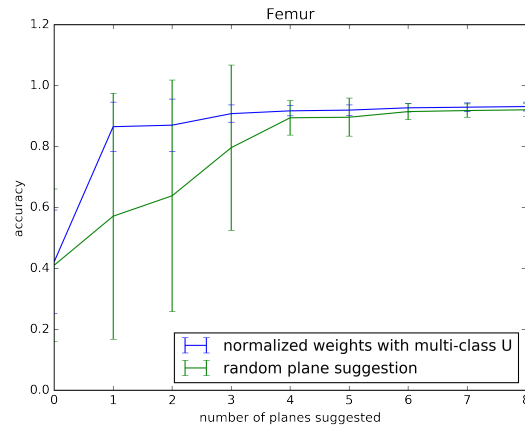


Figure 7: Accuracy vs number of planes suggested, the average of multiple volumes for random plane suggestion is compared to uncertainty sampling.

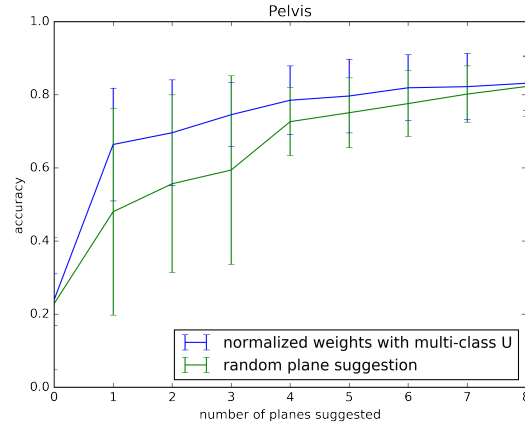


Figure 8: Accuracy vs number of planes suggested, the average of multiple volumes for random plane suggestion is compared to uncertainty sampling.



## 4.4 Systematically Fixed Plane Suggestion

What research question does it help to answer?

(c) Does Active Learning in image segmentation lead to an increase in efficiency compared to fixed sampling?

### Method

For different volumes determine the accuracy versus number of planes suggested, for fixed planes. With fixed planes I mean planes with a predetermined orientation, as shown in figure 9. First the transverse, sagittal and frontal planes are suggested and then the planes that make  $0.25\pi$  angles with these planes are suggested. This is repeated for different volumes.

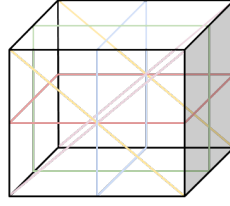


Figure 9: Example of predetermined planes, in black the volume and in color the planes.

### Results

Figures 10 and 11 show the accuracy vs number of planes suggested for fixed planes compared to uncertainty sampling.

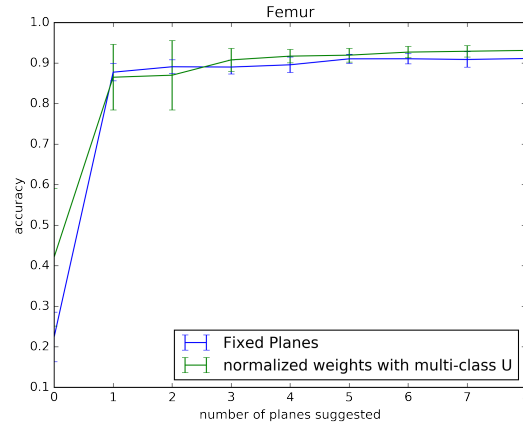


Figure 10: Accuracy vs number of planes suggested, the average of multiple volumes for fixed plane suggestion is compared to uncertainty sampling.

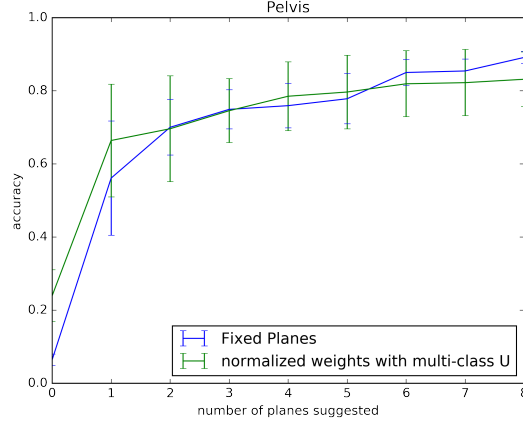


Figure 11: Accuracy vs number of planes suggested, the average of multiple volumes for fixed plane suggestion is compared to uncertainty sampling.

#### 4.5 Evaluation on different volumes, for fixed weights

What research question does it help to answer?

(d) Is the performance of Top's method consistent over multiple scans?

##### Method

For a fixed weight compare accuracy with number of planes suggested for different volumes.

##### Results

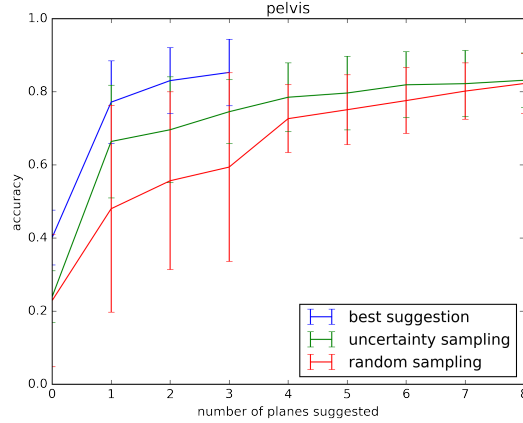


Figure 12: Accuracy vs number of planes suggested, the average of multiple volumes. Best suggestion is compared to uncertainty sampling.

#### 4.6 Evaluation of different initializations for 1 volume.

What research question does it help to answer?

(e) Does initialization using a random plane influence the performance?

## Method

Compare accuracy vs number of planes suggested for different random initializations on the same volume. Also compare with a fixed initial plane though pelvis and femur.

## Results

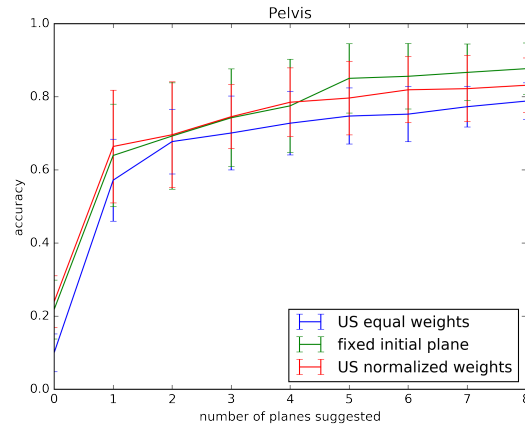


Figure 13: Accuracy vs number of planes suggested, the average and standard deviation of multiple volumes is compared for different weights in the uncertainty sampling equation.

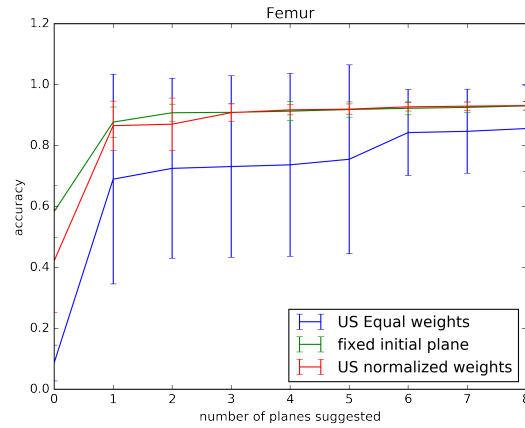


Figure 14: Accuracy vs number of planes suggested, the average and standard deviation of multiple volumes is compared for different weights in the uncertainty sampling equation.

## 4.7 Find upper limit.

**What research question does it help to answer?**

(f) Does correcting the most uncertain plane lead to the highest increase in accuracy?

## Method

Using the uncertainty the most uncertain plane can be presented. However the most uncertain plane does not necessarily have to be the plane that leads to the highest increase in accuracy. To see if this method

can perform better each plane suggestion will be compared to the plane that leads to the highest increase in accuracy. This is calculated in a brute-force kind of way.

First suggest a random plane that passes through both bones and calculate the accuracy. Then a plane is rotated through the origin in equal steps, for each step the accuracy is calculated. The coordinates of the plane with the highest increase in uncertainty are stored.

## Results

### 4.8 Influence of amount of random planes as initializations for gradient descent?

What research question does it help to answer?

#### Method

To find the most uncertain plane a gradient descent optimization is performed on the uncertainty volume, where the variable parameters of the plane are its origin and the three angles. However gradient descent has the risk of getting stuck in a local optimum, especially when many local optimums exist. To explore if gradient descent is a reasonable choice a plane fixed at the middle of the volume is rotated around the y- and z-axis, and the sum of the uncertainty over the plane is calculated for the rotations.

## Results

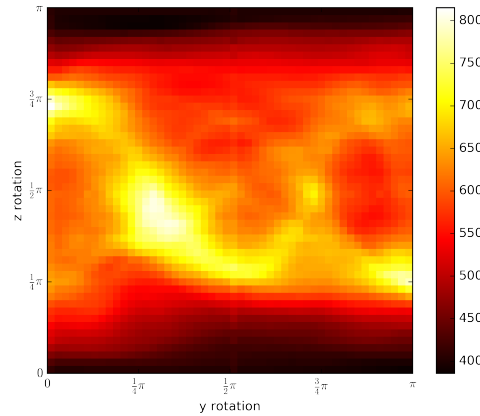


Figure 15: Summed uncertainty over a plane for rotation around the y- and z-axis.