www.visceral.eu

Data set for first competition

Deliverable number	*D2.3.1*
Dissemination level	*Public*
Delivery date	*03 May 2013*
Status	*Final*
Author(s)	*Marianne Winterstein, Marc-André Weber, Katharina Grünberg, Bjoern Menze, Georg Langs*

This project is supported by the European Commission under the Information and Communication Technologies (ICT) Theme of the 7th Framework Programme for Research and Technological Development.

Grant Agreement Number: 318068

Executive Summary

VISCERAL aims at distributing a substantial amount of medical imaging data, together with high quality expert annotations, which is resulting in the “Gold Corpus”. This document describes the different types of computed tomography (CT) and magnetic resonance imaging (MRI) data sets, provided by UKL-HD that will be used for creating the gold corpus. The document also lists the anatomical structures visible from these gold corpus images that will be annotated through medical experts as well as in the VISCERAL challenge. We also comment shortly on technical aspects of the annotation process, such as anonymization and data transfer from UKL-HD and upload to the VISCERAL cloud infrastructure.

Table of Contents

List of abbreviations

MRI	Magnetic Resonance Imaging
CT	Computed tomography
ce	contrast-enhanced
wb	whole-body
STIR	short tau inversion recovery
tse	turbo-spin-echo
flash	fast low-angle shot
fs	fat saturated

1 Introduction

For creating the Gold Corpus, a large data set of whole-body (wb) imaging data sets or data sets comprising a large scale of the whole trunk are needed. In the provided data sets, as many organs as possible should be visible and depicted in a resolution as high as needed to reliably detect an organ and delineate its borders, so that a sufficiently large number of organs can be segmented in one data set. In order to evaluate the segmentation algorithms on several modalities, contrasts and sequence directions, we will create the Gold Corpus with un-enhanced and enhanced MRI and CT data sets; in detail MRI in three different contrasts (T1-weighted, T2-weighted and contrast-enhanced T1-weighted MRI) and two different directions (coronal and axial), CT in axial and coronal direction and in two different contrasts (un-enhanced and contrast media-enhanced).

2 Data Set for Gold Corpus

The data sets of the Gold Corpus have been acquired during the daily clinical routine work. Thus, not all wb imaging data sets display all organ structures with high image contrast, or have a sufficient isotropic resolution (> 1 mm) leading to problems in the segmentation of smaller organs, such as the adrenal glands. To this end, we amend the wb imaging data sets with images showing the wb trunk, displaying anatomical structures at high image resolution and with improved image contrast due to the administration of specific contrast-enhancing agents both in CT and MRI. Having both un-enhanced and enhanced data sets also allows us to test segmentation algorithms on high as well as on low contrast.

For the VISCERAL project, CT and MRI scans that were acquired in the years 2004 – 2008 were used. Data sets of children (< 18 years) were not included due to the recommendation of the local ethical committee (number S-465/2012, approval date February 21th, 2013). For the first competition (i.e. creating the gold corpus by segmentation of organs) wb MRI and CT scans or examinations of the whole trunk were used. Furthermore, imaging of the abdomen in MRI and contrast-enhanced CT for oncological staging purposes will be used, since there is a higher resolution for segmentation especially of smaller inner organs, such as the adrenal glands.

For creation of the gold corpus, an overall of 400 data sets are used, which comprise different CT and MRI scans described in the following.

2.1 CT Scans

2.1.1.1 Unenhanced Whole-Body CTs

For the Gold Corpus of the first benchmark we chose 100 wb CT data sets. These images have been acquired in patients with bone marrow neoplasms, such as multiple myeloma, in order to detect focal bone lesions (osteolysis). The field of view of these CT scans start at the head and end at the knee, as shown in Figure 1.

The wb CT scans are unenhanced, since these examinations are focusing on the bone marrow and the maintenance of the trabecular bone structure. For this purpose, no contrast media is needed. However, all other soft tissue organs may also be seen in these examinations, albeit with sometimes a lower contrast, e.g. lower contrast of vessels from bowl loops.

The in plane resolution for these examinations is between 0.977/0.977 – 1.405/1.405 mm, and the in-between plane resolution at least 3 mm or higher.

2.1.1.2 Contrast Enhanced CT Scans of the Trunk

In order to include contrast-enhanced data sets with large field of view in the Gold Corpus, we chose another 100 CT scans of the trunk. These CT scans have been acquired primarily in patients with malignant lymphoma. Their field of view starts at about at the corpus mandibulae, i.e. in between the skull base and the neck and ends at the pelvis, as seen in Figure 2.

The CT scans of the trunk are usually enhanced by an iodine-containing contrast agent that is administered, for example, to improve tissue contrasts for detecting pathological lymph nodes or organ affection of the lymphoma.

The in plane resolution for these examinations is between 0.604/0.604 – 0.793/0,793 mm, and the in-between plane resolution at least 3 mm or higher.

2.2 MRI scans

2.2.1.1 Unenhanced Whole-Body MRIs

Similar to the wb CTs, a set of 100 whole body MRIs are included for the first benchmark. About 50 % of the wb MRIs will be of the same patient as the wb CT scans in order to provide an optimal data set with the maximum number of tissue contrasts for the image segmentation algorithms. These studies are usually acquired in patients with multiple myeloma in order to detect affection (either as diffuse infiltration or as (multi-)focal infiltration or both) of the bone marrow and to detect extra osseous involvement, e.g. soft tissue masses. The field of view of these MRI scan starts with the head and ends at the feet, as shown in Figure 3.

Again, these studies are unenhanced. Nevertheless, most organs can be seen in these MR-images. All of these examinations include a coronal T1-weighted and fat-suppressed T2-weighted or STIR (short tau inversion recovery) sequence of the whole body, plus a sagittal T1-weighted and a sagittal T2*-weighted sequence of the entire vertebral column. An overview of the contained sequences in wb MRI is given in Table 1. Examples of wb-MRI scans are shown in Figure 3.

The in plane resolution for these examinations is 1.250/1.250 mm, and the in-between plane resolution is 5 mm.

2.2.1.2 Contrast Enhanced MRI of the Abdomen

Since the in plane resolution in wb MRIs is not suitable enough for identification and segmentation of smaller organs (such as the adrenal glands), MRI studies of the abdomen are also included. These images are acquired in oncological patients, who most likely have metastases within the abdomen.

The examinations are contrast-enhanced by a gadolinium-chelate. The scan starts at the top of the diaphragm (thus the lower parts of the lung are also visible in these scans) and ends at the pelvis. The sequences that are acquired in all of these studies are listed in table Table 2. An example of an abdomen-MRI is given in Figure 4.

The abdominal MRI images have a higher resolution compared to the whole body MRIs: The in plane resolution for these examinations is between 0.840/0.840 – 1.302/1.302 mm, and the in-between plane resolution ranges from 3 – 8 mm.

Sequence	Orientation	Slice Thickness
T1 tse (turbo-spin-echo)	coronal	5 mm
T2-weighted STIR (short tau inversion recovery)	coronal	5 mm
T2* flash (fast low-angle shot) 2D	sagittal	6 mm
T1 tse	sagittal	6 mm

Table 1 : Sequences in Whole Body MRI

Sequence	Orientation	Slice Thickness
T2-weighted STIR (short tau inversion recovery)	coronal	7 mm
T2 haste (half fourier-acquired single shot turbo-spin-echo)	axial	8 mm
T1 flash (fast low-angle shot) 2D	axial	8 mm
T1 flash 3D fs (fat saturated)	axial	3 mm
T1 flash 3D fs ce (contrast-enhanced) (arterial phase), upper abdomen	axial	3 mm
T1 flash 3D fs ce (portal venous phase), upper abdomen	axial	3 mm
T1 flash 3D fs ce (venous phase), upper abdomen	axial	3 mm
T1 flash 3D fs ce (venous phase), abdomen	axial	5 mm
T1 flash 3D fs ce (venous phase)	coronal	3 mm

Table 2 : MRI of the Abdomen

3 Organs and Landmarks for Annotation

3.1 Landmarks

Anatomical landmarks are the locations of selected anatomical structures that – ideally – can be identified in different image sequences, such as CT or MRI, unenhanced or enhanced scans, whole body images or with limited field of view. Their universal nature makes them important as a first step in parsing image content, or for triangulating other more specific anatomical structures. Being invariant against the field of view they are of particular importance for image retrieval tasks as evaluated in our second competition. For the Gold Corpus of the first competition, twelve landmarks were defined. A list of the landmarks, and an overview, in which studies they are best found, is given in Table 3. The definition of the landmarks is pointed out in Figure 5.

wb MRI, wb CT, CT of the trunk	MRI abdomen
Lateral end of clavicle (2)	Crista iliaca, at the top (2)
Crista iliaca, at the top (2)	Symphysis below (1)
Symphysis below (1)	Trochanter major, at the tip (2)
Trochanter major, at the tip (2)	Aortic bifurcation (1)
Trochanter minor, most medial part (2)	
Tip of aortic arch (1)	
Tracheal bifurcation (1)	
Aortic bifurcation (1)	

Table 3 : Landmarks

3.2 Organ segmentation

In our list of organs to be annotated in the first competition we focus on a representative selection of major and minor structures. We avoid structures such as joints, since they consist of several parts. Also, lymph nodes are often too small to be seen reliably in all examinations. Moreover, we also avoid anatomical structures that are gender specific, thus are not frequently definable. So, for the first competition we defined a list of organs that can be detected in a large set of examinations (Table 4).

When delineating the organs, we face the problem of dealing with defining the outer extensions, requiring a definition of what part of a connected structure is still “within” the organ and what is already another structure “outside” of our organ delineation. Some organs on this list have a hilum, that means a depression or indentation, where vessels and nerves enter (Fig. 6). The hilum has to be cut off during the segmentation process. This means we segment the organs without the external structures outside the hilum and include the vessels within the hilum, by taking the shortest distance

within the hilum like demonstrated on the images for each concerned organ: lung, liver, and kidney (Figures 7 – 9). For the spleen, it is not necessary to do that, because it is feasible to segment the spleen parenchyma apart from the vessels within the hilum. For the lung, we segment the lung up to the main bronchus and up to the first division of the pulmonary artery and vein.

Organ	Comment
Kidneys (2)	see Figure 7
Spleen	
Liver	see Figure 8
Lungs (2 lobes)	see Figure 9
Urinary bladder	
Rectus abdominis muscle (2)	
Lumbar Vertebra #1	
Thyroid Gland	
Pancreas	
Psoas major muscle (2)	The psoas muscle consist of the psoas major and minor muscle, but the psoas minor muscle is an inconstant muscle.
Gallbladder	without cystic duct
Sternum	
Aorta	From ascending aorta beginning at the aortic valve, until the aortic bifurcation; cut of the exiting vessels.
Trachea	Beginning from the larynx until the tracheal bifurcation.
Adrenal glands (2)	Hard to see on wb MRI, prefer MRI abdomen; well seen on CT scans.

Table 4 : List of organs to be segmented

Not all of the structures can be detected precisely in all MRI sequences. Also, for some structures segmentation in contrast-enhanced scans is easier than in those without administered contrast media. In unenhanced MRI, we have two contrasts (T1-weighting and T2-weighting). In native unenhanced CT, we have real density correlating to tissue composition, the so called Hounsfield units with air having a Hounsfield unit of -1000, fat tissue a Hounsfield unit of about −100, pure water of 0, and bone depending on mineralization and density Hounsfield units between 500 and 1500. And we also have contrast-enhanced MRI or CT (in different time-depending phases: e.g. arterial and portal-venous. In general, the defined structures are constantly depicted in the contrast-enhanced CT

examinations of the trunk (portal venous phase). In order to test how good segmentation algorithms work, the segmentation will be made in the MR sequences or CT enhancement phases most suitable for the annotated structure. After that, the algorithms can then be tested in sequences with lower contrast for the structure of interest. Table 5 gives an overview of the MRI sequence or CT enhancement phase where the structures are best to be seen.

		wb MRI, unenhanced		**wb CT, unenhanced**	**MRI Abdomen**						**CT trunk**
	Organ	T1 cor	T2 fs cor	CT ax	T1 ax, nat.	T1 fs ax ce, Liver dynamics (opt.)	T2 ax	T2 fs cor	T1 fs ax, ce	T1 fs cor, ce	CT ax, ce
1	Kidney		x	x			x	x	x	x	x
2	Spleen	x		x	x		x		x	x	x
3	Liver	x		x	x	x	x		x	x	x
4	Lung		x	x	----------						x
5	Urinary bladder		x	x			x	x	x	x	x
6	Rectus muscle			x			x		x	x	x
7	Lumbar vertebra #1	x		x			x		x	x	x
8	Thyroid gland	x	(x)	x	----------						x
9	Pancreas	x		x			x		x	x	x
10	Psoas muscle	x		x			x		x	x	x
11	Gallbladder		x	x			x	x	x		x
12	Sternum	x		x	---------- -						x
13	Aorta	x		x	x		x		x	x	x
14	Trachea	x		x	---------- --						x
15	Adrenal gland	x		x			x		x	x	x

Table 5 : List of Organs in first competition and where best to be seen for segmentation.
x = good to see,
nat = unenhanced, ce = contrast-enhanced, ax = axial, cor = coronal, fs = fat suppressed, opt.= optional

4 Data Transfer and Anonymization

All 400 data sets have been stored from the local GE PACS of the University Hospital of Heidelberg with all sequences on a local external hard disc. To guarantee the utmost security (according to legal obligations), someone of the responsible VISCERAL study group from Heidelberg (UKL-HD) will ensure that correct anonymization and defacing will take place before the data set is uploaded onto the cloud servers. A workflow for effectively anonymizing, storing and transferring the data has been thoroughly discussed with all partners, and the data transfer to HESSO and MUW is currently in progress.

For anonymization, the following items are removed from the DICOM headers: Date of birth, (but the separate item: age will be preserved), Institution name (e.g. DI-Radiologie Orthopädie UKL-HD or Diag. & Interv. Radiologie UKL HD), patient's name, patient-ID, examination number and study date. For further and on-going comparison and analysis, the age of the patient will be preserved. Also, studies of one patient shall be stored in one separate data folder per patient to enable the algorithmic develops the means to make joint use of all image contrasts.

5 Conclusion

For the first competition, it is important to choose a homogenous data set with as many landmarks and organs to be visible. Therefore, large field-of-view MRI and CT scans are used (i.e. whole body CT or MR imaging, or CT and MR imaging of the trunk). For smaller structures in the abdomen a higher resolution in MRI and contrast-enhanced CT is needed, so that these data sets of the abdomen were also included. For expert annotation we chose landmarks and organs that can be seen best in the image data sets.

Annex

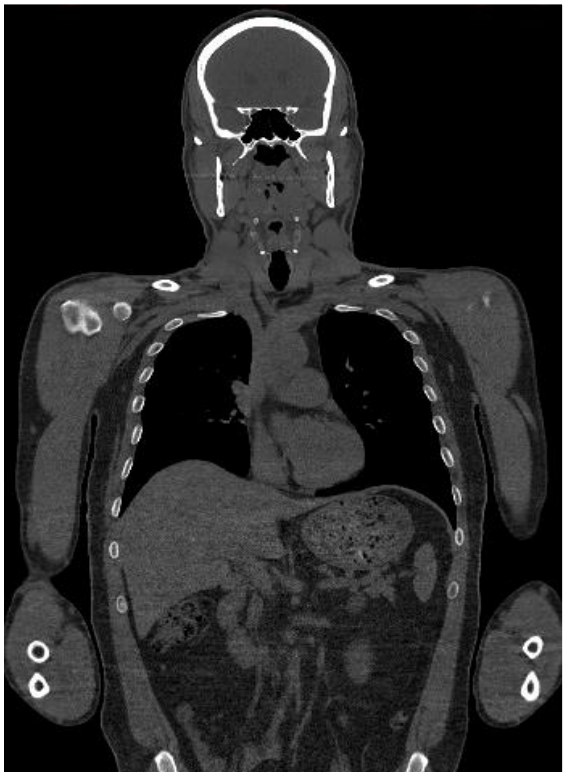

Figure 1a: Wb CT, upper body.

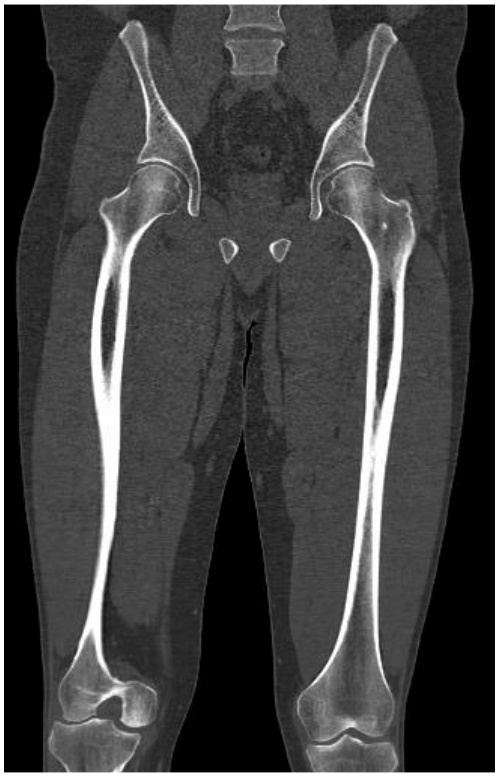

Figure 1b: Wb CT, lower body.

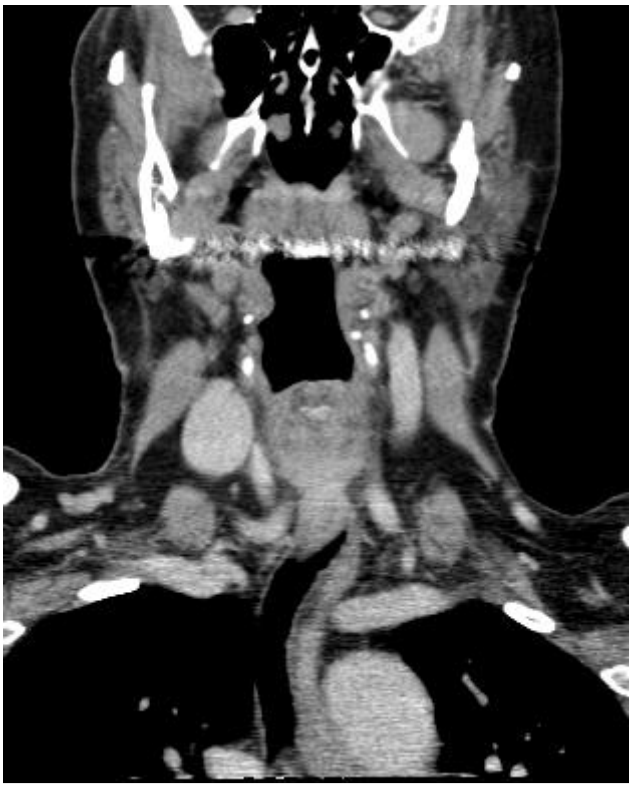

Figure 2a: Ce CT of the trunk, neck.

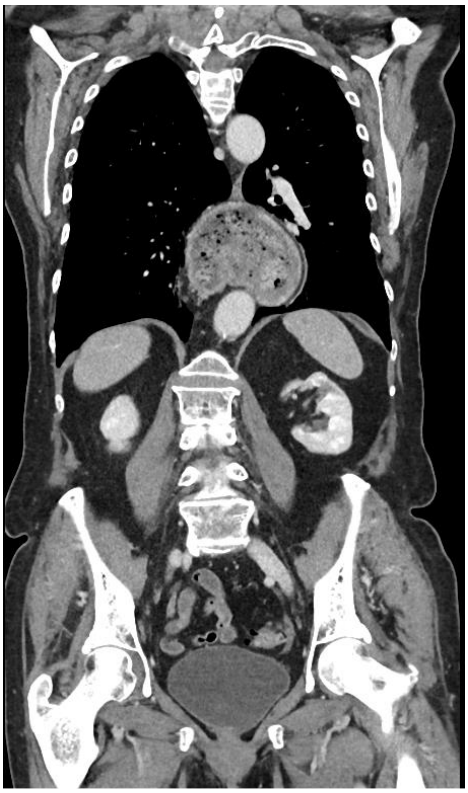

Figure 2b: Ce CT of the trunk, thorax/abdomen.

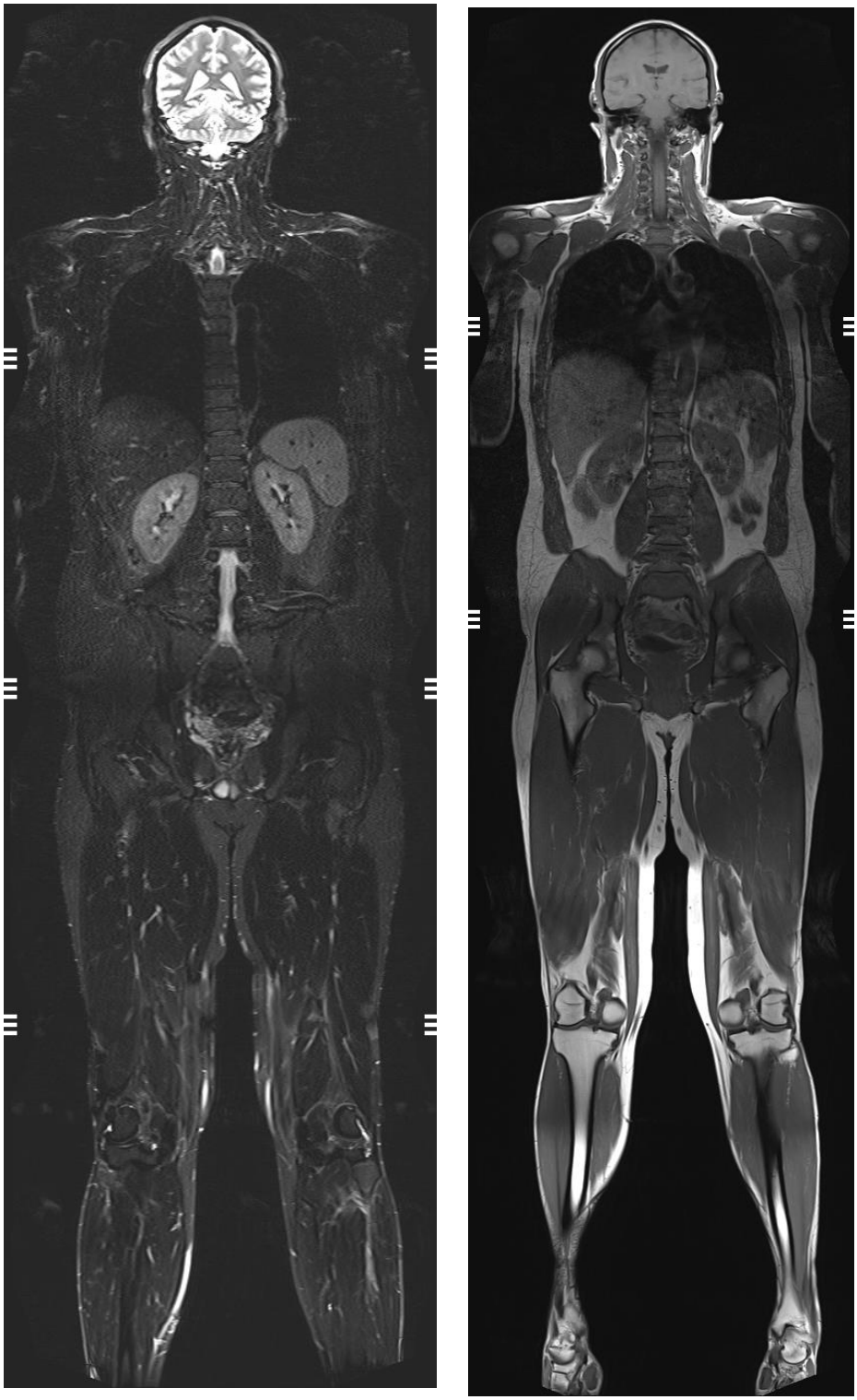

Figure 3: wb MRI. T2 STIR (left), T1w tse (right).

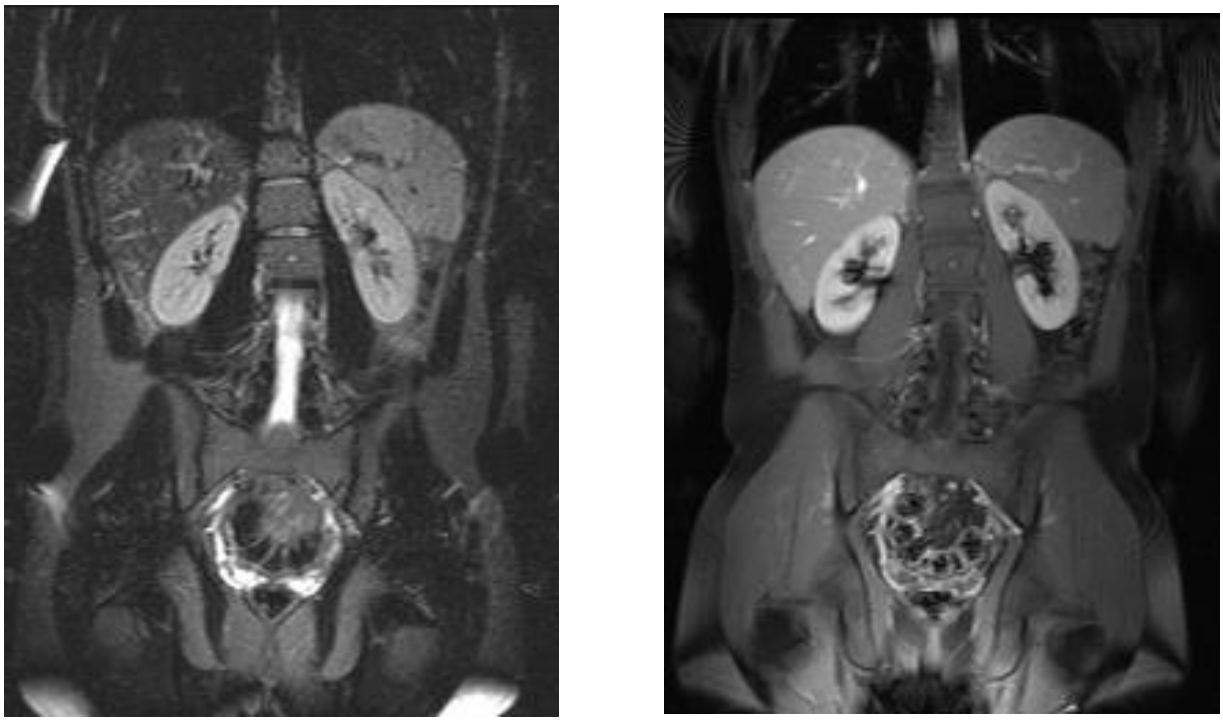

Figure 4: MRI abdomen, coronal, T2 STIR (left), T1 tse ce (right).

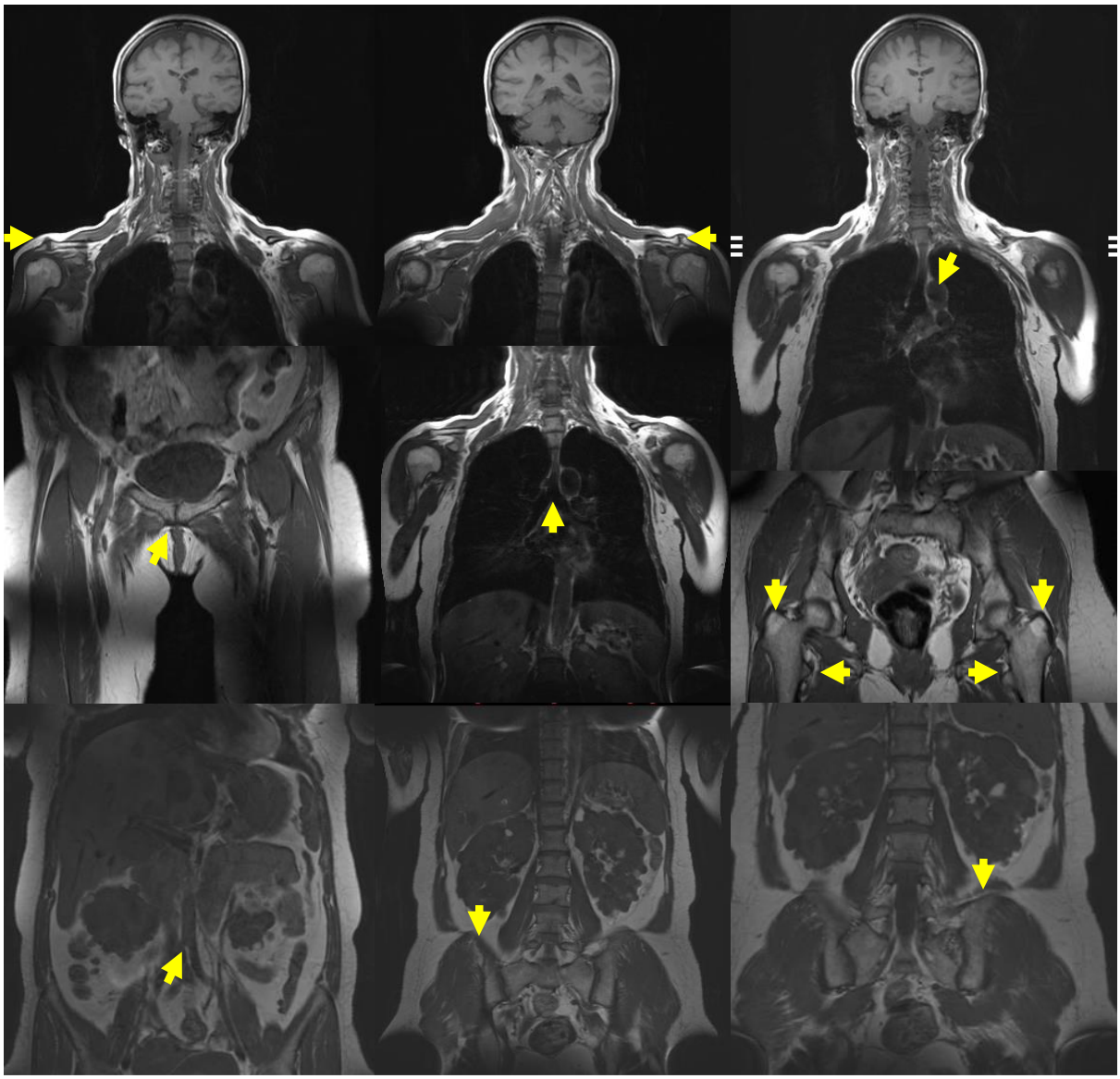

Figure 5: Landmarks for first competition.

From right top to left bottom: right and left lateral end of the clavicle, tip of aortic arch, symphysis below, tracheal bifurcation, trochanter major at the tip and trochanter minor (most medial part), aortic bifurcation, Crista iliaca (at the top).

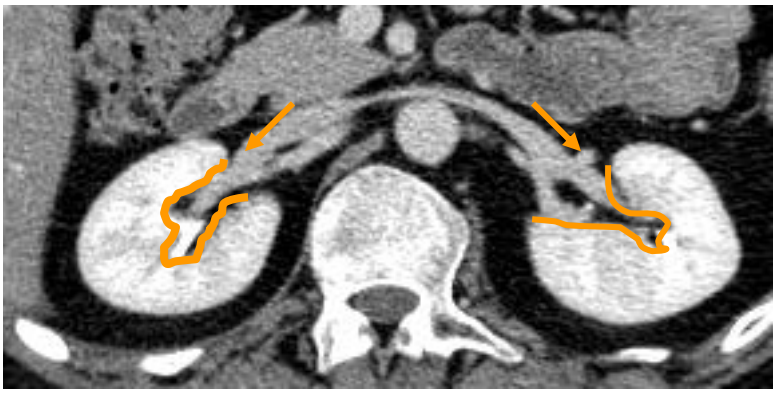

Figure 6: Definition of Hilum.
Marked with the orange line is the hilum of the kidneys, where the vessels enter.

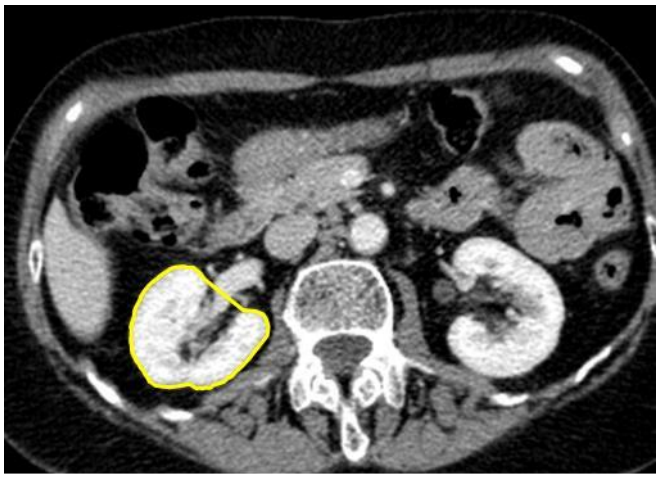

Figure 7: Segmentation of the right kidney (ce CT).

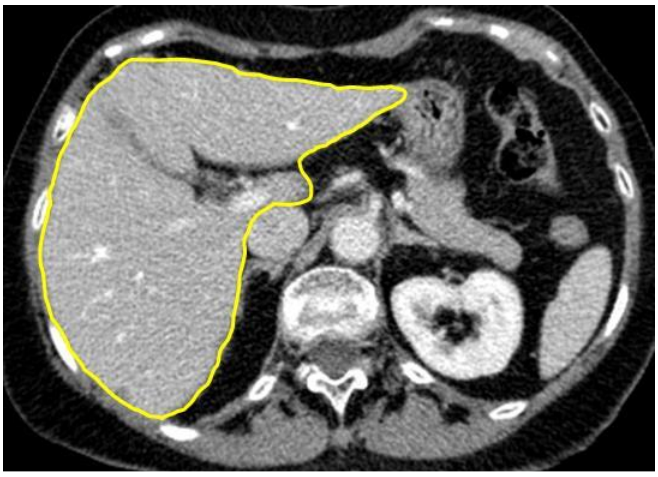

Figure 8: Segmentation of the liver (ce CT).

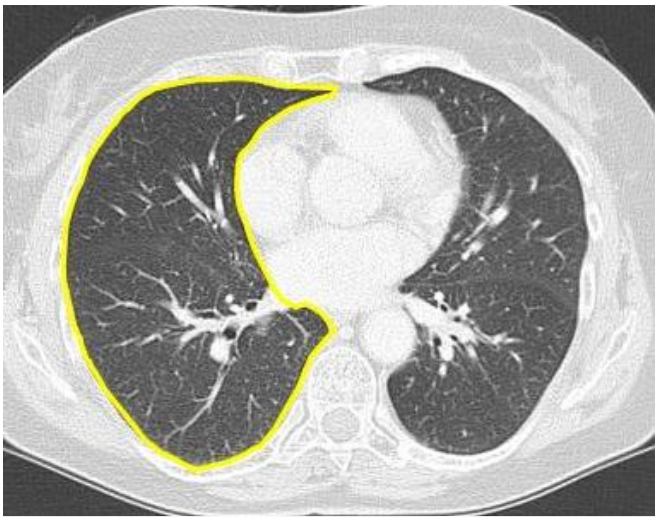

Figure 9: Segmentation of the right lobe of the lung (ce CT).