Several axes of improvement can be imagined regarding our research work.

When performing a hand-crafted or a deep radiomics study, we still believe that the key part will be the area selection preceding the computation of the radiomics features.

*Semantic Segmentation*

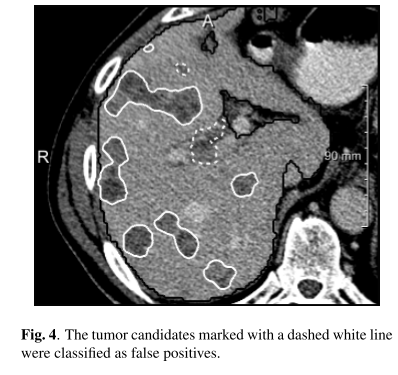
As discussed before, the HCR paradigm requires manual experts segmentations, and several studies already discussed the inter- and intra-observer variability present among the obtained delineations. This variability is translated in the quality of the retained features, that tend not to be robust enough when performing the final predictions.

The DLR paradigm allows the segmentation to be made automatically, thus reducing the number of potential biases. The automatic segmentation, as performed in our research work, however requires a high number of training cases to offer a realistic delineation of the targeted tissues.

In the case of liver tumors, a cascaded architecture stacking a first network responsible for the segmentation of the liver and a second dedicated only to detect the tumor within the obtained liver mask is currently the method allowing to obtain the most accurate results. State-of-the-art liver segmentation methods currently allow us to reach annotations similar to those obtained by the experts with mean DSC often above 0.95-0.96. We believe that there is no real need to improve the liver segmentation in the future, and that more interest needs to be shown in tumor and other liver tissues segmentation.

Regarding the segmentation of the tumors, current state-of-the-art results were obtained thanks to post-processing steps such as false positives (FP) filtering post-processing method to reduce the number of object misclassified as lesions.

Therefore current semantic segmentation architectures still lack the ability to discriminate between real lesions and other areas sharing the same textural properties such as regions close to the vessels as illustrated in the following figure.

 **Chlebus et al**.

One reason explaining this low recall might be the poor quality of the currently available segmentation datasets. Liver annotations often contain non-hepatic areas (such as air most of the time) or imperfections close to the organ borders.

When performing the automatic liver segmentation, the deep neural networks will most of the time be able to avoid these regions. However, in case they are integrated in the liver mask, they will often be misclassified as tumors by the second network in the cascade (responsible for the tumor segmentation).

To overcome this issue, the publicly available datasets such as the LIST-dB need to be properly resegmented. They usually only contain annotations for the parenchyma or the tumors, but a third class (or more) could be considered to incorporate tissues belonging to none of these groups such as the cysts or the blood vessels.   
As explained in the DL sem. Seg. section, current state-of-the-art tumor segmentation results are often obtained with 2D or 2.5D networks, but the post-processings steps usually help to improve the obtained accuracy (3D CRF or even simple morphological operations). Several studies implemented a U-Net like architecture, where the information is compressed before being decoding to obtain the final segmentation map. One of the key concepts in these architectures is the reinjection of features learned in the earlier layers later in the network, either thanks to skip-connections, residual units or even densely connected layers. Another way to perform the segmentation could be to implement a multi-scale pyramidal architecture where tumor-related features can be learned directly from the raw images at multiple scales instead of being learned from previously obtained features, which is often the case in the aforementioned architectures.

It might also be interesting to perform the segmentation with a full 3D architecture. Some studies already investigated this solution (Dou et al. for example), but they could not yet overtake state-of-the-art results.

Some other concepts could be incorporated in new architectures. Jin et al. for example proposed a network that integrated both U-Net and attention residual mechanism to proceed the segmentation of both the liver and the lesions. The residual attention mechanism has been introduced in 2017 by Wang et al. [cite Wang2017] to perform image classification, with the idea that the attention mechanism can help the network focusing on specific parts of the image. The study from Jin et al. [cite Jin2018] was the first to use the attention mechanism for semantic segmentation purposes. On the hidden test set of LITS, they outperformed a lot of 2D-based methods, but were still far from the top-ranked teams. However, new paradigms such as the self-attention mechanism, in combination with state-of-the-art 2D and 3D architectures are certainly an avenue for the improvement of the automatic liver and tumors segmentation tasks [cite Chen2019].

*Dynamic contrast-enhanced images*

In our research work, the key element is the extraction of the features from dynamic contrast-enhanced images. However, it has been very difficult to collect dynamic CECT based images, since only a few publicly available datasets contain this type of volume.

There is a huge room of improvement regarding this specific type of dataset. It has been mentioned in the Medical Context Section, that the use of images obtained after the injection of contrast medium largely improves the visual and the automatic quality of the diagnosis. Nonetheless, we realized that multiphasic images could be better used in either the semantic segmentation and/or the radiomics field.

The first issue is related to the acquisition protocol required to obtain this type of images. Most of the time, the different retained phases, namely arterial, portal venous or delayed phases are acquired following either visual inspection of the radiologists \footnote{The arterial phase can sometimes be acquired with the help of a bolus tracking technique}, or after a specific duration following the injection moment that can differ from one study to the other. Another criteria to consider is the physiological characteristics of the patient such as its weight that will determine how the contrast medium is diffused through the liver.

When dealing with multiphase databases, we believe that these key elements need to be considered, at least the moment of injection and/or the duration between the injection and the acquisition since the weight of the patient can be inferred from the liver volume (which can be predicted thanks to semantic segmentation). Currently, only the DICOM format allows the addition of such metadata (only the exact acquisition moment but not the injection moment) but they are barely reported and are often left to default values.

Obtaining such data seems really challenging, hence, some techniques need to be implemented to deal with current multiphase images.

For example, we believe that a “phase correction system” could be designed that would potentially normalize images belonging to the same phase. The automatic detection of the amount of contrast medium still in the liver could also help to predict the exact duration between the injection of the CM and the acquisition.

In our research work, we have decided to implement an histogram specification algorithm to “normalize” images from a given phase. The first step requires to compute a mean histogram per phase (requires to have a sufficient amount of patients images correctly labeled), before transforming the volumes of a given patient to fit the obtained mean histogram.

|  |
| --- |
| Slice before mapping |
| Mapping function from the current patient to the average distribution |
| Slice after mapping |
|  |

In the case of multiphase images, the expert performing the segmentation will usually try to obtain only one segmentation for the registered volumes. Being coherent for the liver, we believe that this technique needs to be modified for internal liver tissues. The lesions for example will have a different behavior from one phase to the other (in case for example of hypo- or hyper-dense tumors), several segmentation should therefore be performed. This might potentially help the deep neural networks to better understand the dynamic of the lesions, and can also address the problem of small mis-registration when only one ground truth segmentation map is shared by all the available phases.

One long-term objective could be to create, with the help of radiological experts, a dataset containing registered multiphase CT volumes with expert annotations performed following the aforementioned protocol.

*Deep Radiomics*

Regarding our DLR study, we believe that the prediction of the histological grade can be improved.

One lead of improvement could be the prediction of the grade at a finer scale, but this will require to know the exact position of the extracted sample, or to obtain a map of the heterogeneous regions after surgical removal of liver tissues through liver resection or liver transplant for example.

In our research work we investigated a patch-wise prediction of the histological grade but the results were not as promising as those obtained through our slice-wise approach, especially because our patch-wise semantic segmentation network needs to be improved.

We are currently considering the relevant imaging features as being the ones extracted from the bottleneck part of our U-Net network but other techniques such as the auto-encoder could be investigated to better extract the relevant information from the raw images, or from the retained features. We could have also potentially incorporated clinical data (AFP levels, age, …) in our pipeline, however such data are often difficult to retrieve, and difficult to combine with imaging features, consequently, we decided to focus only on images in the current work.

If more interest is shown in the future towards the prediction histological grade, we believe that the existing grading systems have to be standardized by considering more criteria than just the worst or the most frequent grade present in the histological slices, which is currently the case.

Our DLR architecture obtained promising results for the prediction of the histological grade, but we believe that the same architecture can be adapted to focus on other prediction tasks, such as the prediction of recurrence after treatment using longitudinal studies.