Paper SAS4491-2020

Medical Image Analyses in SAS® Viya® with Applications in Automatic Tissue Morphometry in the Clinic

Courtney Ambrozic, Joost Huiskens, and Fijoy Vadakkumpadan, SAS Institute Inc.

ABSTRACT

Imaging and image analytics are indispensable tools in clinical medicine today. Among the various metrics that doctors routinely derive from images, measures of the morphology of tissue structures, including their shape and size, are of key significance. Quantifying tissue morphology and linking those quantities to other clinical data enable clinicians to diagnose diseases and plan treatment strategies. Image segmentation, which classifies image pixels into regions of interest, is an important step in such tissue morphology quantification. However, common segmentation methods involve a process that is either fully or partially manual. Accordingly, these methods can be extremely arduous when you process very large amounts of data. This paper illustrates how to build end-to-end pipelines for automatically deriving clinically significant tissue morphology metrics from raw medical images by using powerful tools that were introduced in SAS® Viya® 3.5. Specifically, it shows how you can load medical images and metadata, preprocess the loaded data, build convolutional neural network models for automatic segmentation, and postprocess the results to compute clinically significant 2-D and 3-D morphological metrics. The examples include colorectal liver metastases morphometry in collaboration with the Amsterdam University Medical Center, and normal spinal cord morphometry with data available from the Cancer Imaging Archive, both based on 3-D CT scans.

INTRODUCTION

Imaging and artificial intelligence have shown many practical applications in the clinic in recent years. SAS® Viya® provides users the building blocks to load, process, and visualize biomedical images. Using these building blocks, users can construct end-to-end pipelines to derive important image-based biomarkers that can be used in the clinic. Segmentation, the process of partitioning an image into sets of pixels, is particularly useful in the derivation of such biomarkers, specifically biomarkers that involve tissue morphometry, i.e., quantification of the size and shape of various tissue structures.

In this paper, we demonstrate how to build fully automatic tissue morphometry pipelines in SAS® Viya® that involves pre-processing, model training, segmentation, and quantification. The motivation for such pipelines is to increase accuracy, decrease subjectivity, and decrease labor of medical professionals. We provide two examples for our demonstration, using different data sets and targeting different clinical biomarkers. Both examples utilize computed topography (CT) scan images in the form of Digital Imaging and Communication in Medicine (DICOM) files. The networks are trained on structural annotations given in the form of DICOM-RT files. The first example uses patient data from Amsterdam University Medical Center and tracks colorectal liver metastasis throughout the duration of a patient's treatment. The second pipeline is built for the segmentation of the spinal cord using data from the Lung CT Segmentation Challenge from 2017 that is publicly available. Using quantification tools in SAS® Viya®, users can derive important biomarkers from the model-predicted contours that have significant impact on clinical insights. The action sets needed to execute these pipelines are the image, biomedimage, fedsql, and

deeplearn action sets. The examples are written in Python, but a user could easily recreate the code in CASL, R, or Lua.

CRLM MORPHOMETRY: AN EXAMPLE USE CASE

MOTIVATION

The first demonstration pipeline that we will highlight will provide a morphometric tissue analysis for patients with colorectal liver metastasis (CRLM). Colorectal cancer is a disease that starts in the colon and often spreads to the liver. The best treatment for CRLM is surgical removal. Currently, the clinical standard for determining a patient's candidacy for surgery lies in the RECIST criterion, which tracks the diameter of two target lesions across successive chemotherapy treatments. To help improve treatment strategies for patients with CRLM through advanced analytics and large amounts of data, SAS joined forces with Amsterdam University Medical Center (AUMC). Through this collaboration, we have access to extensive amounts of data for patients with CRLM.

DATA ACQUISITION AND PREPROCESSING

The data for this example consists of 57 patients and includes medical contours in the form of DICOM-RT files that show medical annotations for the liver and lesions of each image. Data for each patient contains multiple images before chemotherapy, after chemotherapy, and, in some cases, after a continued therapy treatment. Previously, this data was used to show how precisely annotated lesion segmentations can overcome the limitations of the RECIST criteria by quantifying important biomedical metrics such as the volume and contrast of lesions through successive treatments (Vadakkumpadan, Huiskens, 2019). This method, however, requires radiologists to perform an extensive amount of manual contouring for each new patient, which can be both time-consuming and labor intensive. By utilizing deep learning to detect the precise locations of livers and lesions, we overcome the need for these manual tasks on new data. In this use-case example, we propose building an end-to-end segmentation pipeline to identify precise locations of lesions and track the metrics of these lesions over the duration of the patient's treatment process.

Liver lesion segmentation is particularly difficult due to low contrast between lesions and other features as well as lesion shape variability. In order to combat the low contrast within CT scans, several steps are taken in image preprocessing. First, the CT scans are windowed based on the unit of pixel values in CT images, the Hounsfield unit (HU). For the CRLM data, we HU-window the images to the range [-100, 400] to both remove any image features that are not of interest and to highlight important structures. This HU-windowing step is done using the `clamp` step within processBioMedImages. The images are then exported into 2-D slices within the same action call:

```
s.biomedimage.processbiomedimages(
    images=dict(table='ct_scans'),
    steps=[dict(stepparameters=
                dict(steptype='CLAMP',
                     clampParameters=dict(clampType='BASIC',
                                          low=-100, high=400))),
                dict(stepparameters=dict(steptype='export'))],
     copyvars=['_label_','_id_'],
     casout = dict(name='ct_scans_export', replace=True))
```

Next, `hist_equalization` is performed within processImages to enhance the contrast within images. This step distributes the intensity values equally across the image's histogram and therefore allows lesion structures within the image to be highlighted. Figure 1 outlines these preprocessing steps by displaying an original CT scan slice, the same slice

after it is HU-windowed to a range of [-100, 400] and then after its histogram equalization step.

Figure 1. CT Scan Slice, After HU-Windowing, and After Histogram Equalization

The data is then divided into validation, test, and training sets. The training set, which is the data the model will be learning from, consists of 41 patients totaling 18918 slices. The validation set consists of 4 patients (1335 slices) and helps tune hyperparameters of the model. The test set includes 12 patients (6894 slices) and provides unseen images to the model for final evaluation.

SEGMENTATION METHODOLOGY

The convolutional neural network model that will be used for this segmentation task is the U-Net model (Ronneberger et al. 2015), available in SAS® Viya® 3.5. This model is a fully convolutional network that is specifically designed for medical image segmentation and consists of a series of convolutional, concatenation, and max pooling layers. Annotated biomedical images are particularly difficult to obtain due to patient confidentiality and the high level of labor and expertise that is required to create them. The U-Net excels at biomedical image segmentation because it can produce accurate results on smaller annotated image sets. This U-Net is built using layer-by-layer calls within DLPy to create a model with a total of 34512258 parameters. The model uses an Adam Solver with a learning rate of 0.0001.

Two separate U-Net models are trained for liver and lesion segmentation using 2-D slices that have a resolution of 512x512. This two-step model scoring process is done to constrain the solution space for the challenging lesion segmentation task. The first U-Net is trained from scratch for liver segmentation and is shown in Figure 2.

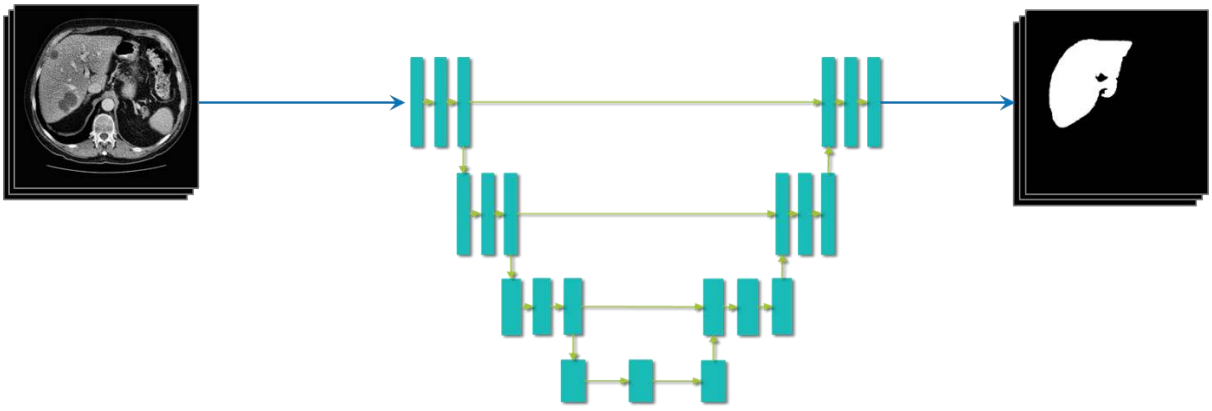

Figure 2. Liver Segmentation Model

The results are cascaded so that the predicted liver segmentations are used as an input test set to the lesion segmentation model. This is accomplished through the `mask_specific` type in the `binary_operation` step of processBioMedImages. Here, the pixel values within the liver region-of-interest (ROI) will be cascaded to the output image and any pixels outside the ROI will have a uniform value. The new `binary_operation` step can read two images from the same input CASTable. The following code demonstrates this process, where the column containing the original CT scan images is named _image_ and the column containing the liver segmentation is labeled "seg":

```
s.biomedimage.processbiomedimages(
    images=dict(table=data_to_be_masked),
    steps=[dict(stepparameters=dict(
                steptype='binary_operation',
                binaryoperation=dict(binaryoperationtype='mask_specific',
                                     image='seg',
                                     outputBackground=-1000,
                                     inputBackground=0)))],
    casout=gray_mask_liver,
    copyvars=['_label_', '_id_'],
)
```

Figure 3 displays the lesion segmentation schematic, where the output of the `mask_specific` type is used as the input to the model.

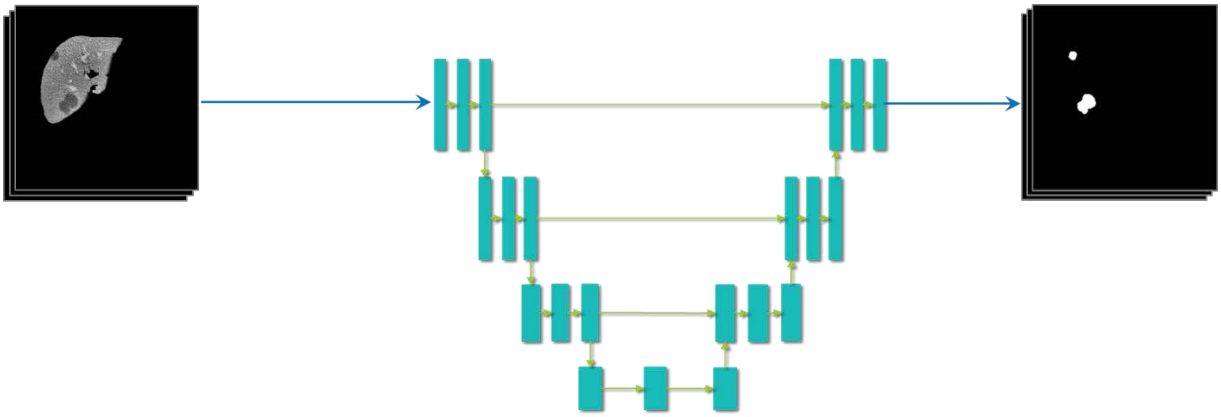

Figure 3. Lesion Segmentation Model

With the solution space constrained to the liver ROI, the model can more accurately predict the lesion regions.

RESULTS

Figure 4 presents the liver segmentation results where blue depicts the liver ROI predicted by the model and the red region is the ground truth liver ROI annotated by the radiologist. These segmentation-overlaid images are created using the annotateImages action.

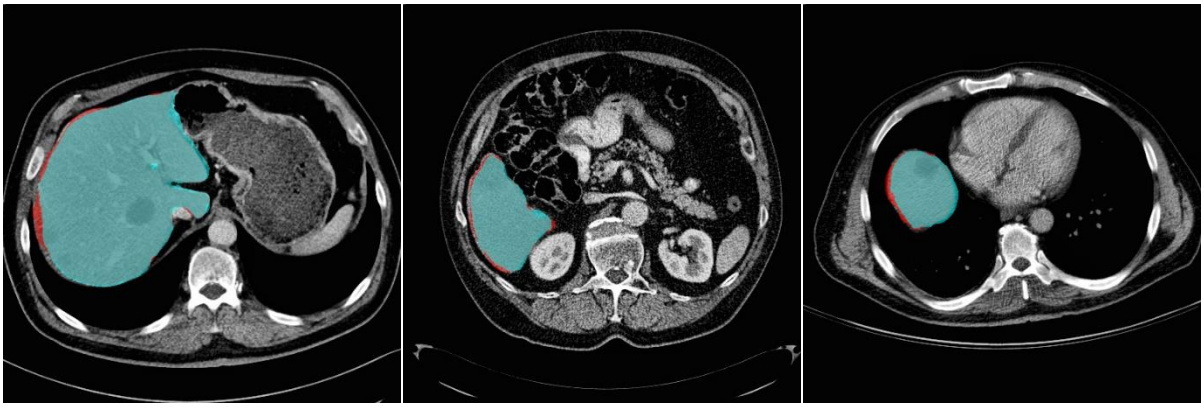

Figure 4. Liver Segmentation Results

The liver segmentation results are then cascaded to be scored for lesion segmentation using the `binary_operation` step outlined previously. In Figure 5, the lesion ROI predicted by the model is shown in blue and the ground truth lesion ROI is displayed in red.

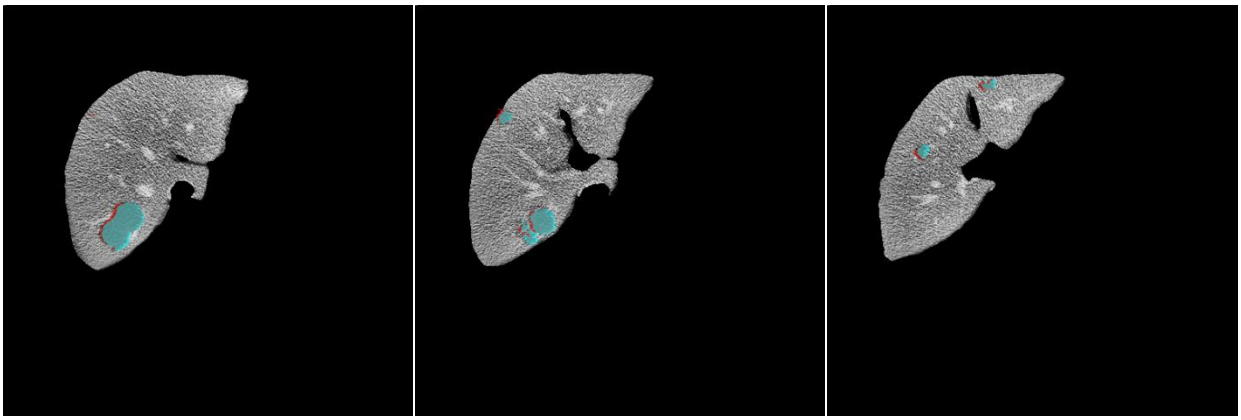

Figure 5. Lesion Segmentation Results

The segmentation predictions are imported back into 3-D using processBioMedImages. These 3-D segmentation images are then built and plotted on the original CT scan using Mayavi software (Ramachandran, 2011). The buildSurface action is utilized for this task to build the surfaces of both the liver and lesion segmentation results. This 3-D visualization for the model prediction is shown in Figure 6. Here, the red surface is the liver ROI and the green surface is the lesion ROI.

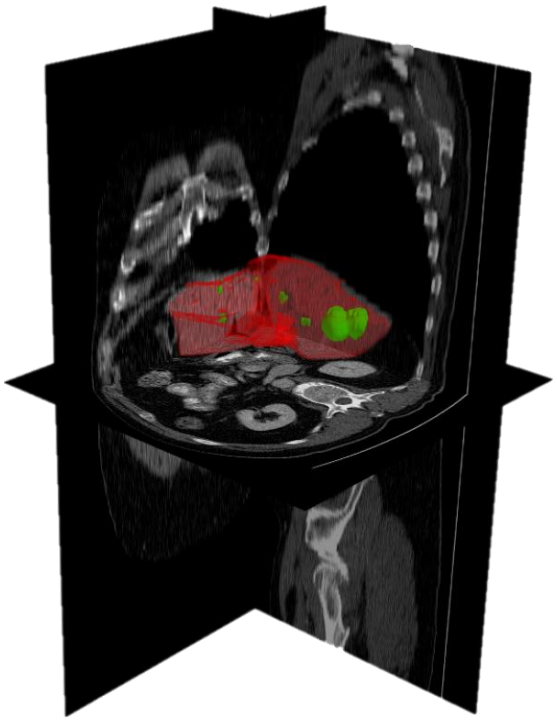

Figure 6. Model-Predicted Liver Lesion 3-D Visualization

The segmentation methods are evaluated against the ground truth for similarity. The main evaluation criteria used for these segmentation images is the DICE coefficient. The DICE score is a much more stringent criterion than misclassification rate and is therefore ideal for quantifying the performance of the model. DICE is defined as two times the area of overlap divided by the total number of pixels:

$$DICE(A,B) = \frac{2|A \cap B|}{|A| + |B|}$$

where a perfect segmentation yields a DICE score of 1. The liver lesion segmentation results are evaluated using the DICE coefficient and the results are shown in Table 1.

	Liver	Lesion
Test set	93.155%	77.703%
Validation set	94.167%	69.975%

Table 1. DICE Coefficients for Liver and Lesion Segmentation

For scoring, we use the DICE global score, which averages the total test set. It is important to note that the global DICE score for lesion segmentation is very dependent on the size of the lesions in the evaluation set. The most competitive models in the Liver Tumor Segmentation Benchmark (LiTS) achieved a DICE score of 96.7% for liver segmentation and 79.40% for lesion segmentation (Bilic et al. 2019). From the predicted contours, metrics can be derived that describe the tissue morphometry. The action quantifyBioMedImages can be used to quantify the lesion segmentation results to analyze volumes and pixel values, both of which are ignored by the RECIST criteria. By specifying the quantify type as `'content'`, the total volume of the lesions is calculated:

```
s.biomedimage.quantifyBioMedImages(
   images=dict(table='bdata_lesion'),
   region='image',
  quantities=[dict(quantityparameters=
```

```
                    dict(quantitytype='CONTENT',usespacing=True)),
                    dict(quantityparameters=dict(quantitytype='MEAN'))],
    inputbackground=-1000,
    labelParameters=dict(labelType='basic', connectivity='vertex'),
    copyvars=['_label_', '_id_'],
    casout=vol)
```

The volume calculations are plotted by patient, ordered by their round of chemotherapy treatment and shown in Figure 7. The first scan before treatment is depicted in blue, the first follow-up scan is depicted in orange, and the second follow-up scan (if it exists) is shown in green.

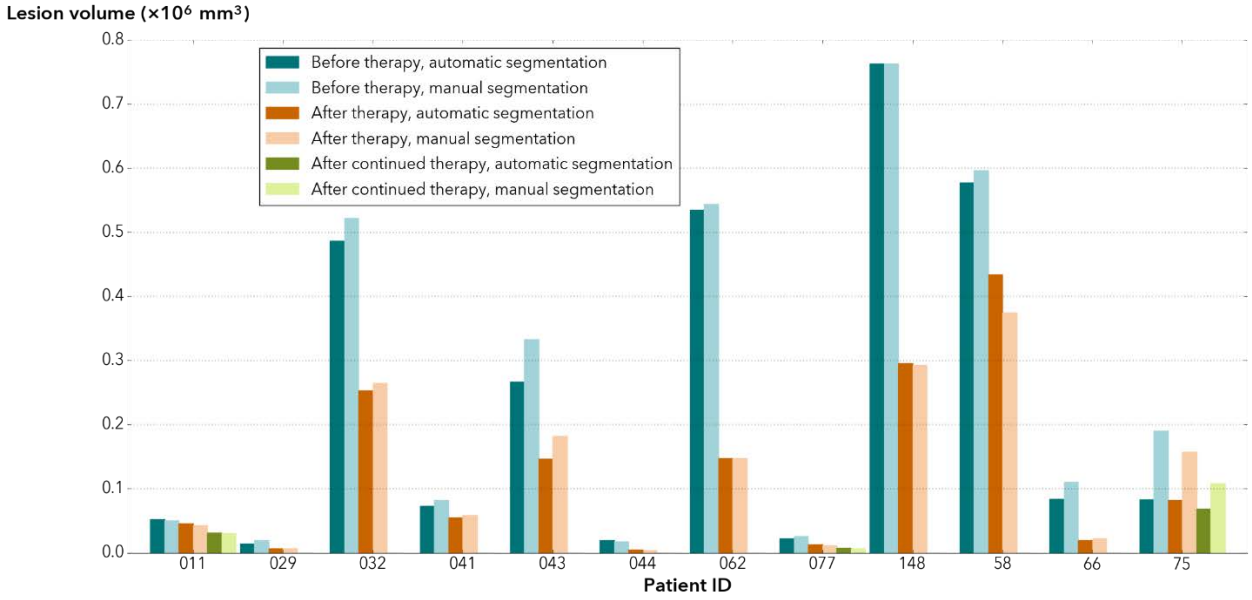

Figure 7. Lesion Volume from Model-Predicted Segmentations

The volumes predicted by the model are plotted against the volumes annotated by the radiologist and used as ground truth. The model-predicted results are displayed in a darker color and the ground truth results are in a lighter color. The model-predicted lesion volumes follow the same trends throughout treatments as the ground truth volumes and therefore verify the effectiveness of the segmentation model. The automatic segmentation volume averaged a 7.713% decrease in comparison to ground truth.

The segmentation predictions are then computed for contrast between the liver and lesions. Contrast is calculated by comparing the mean pixel values within the lesion segmentation to those within the liver segmentation. Figure 8 displays the liver-lesion contrast for each image based on the automatic segmentation from the model. The lesion pixel values are another metric ignored by RECIST that is captured through the automatic segmentation pipeline.

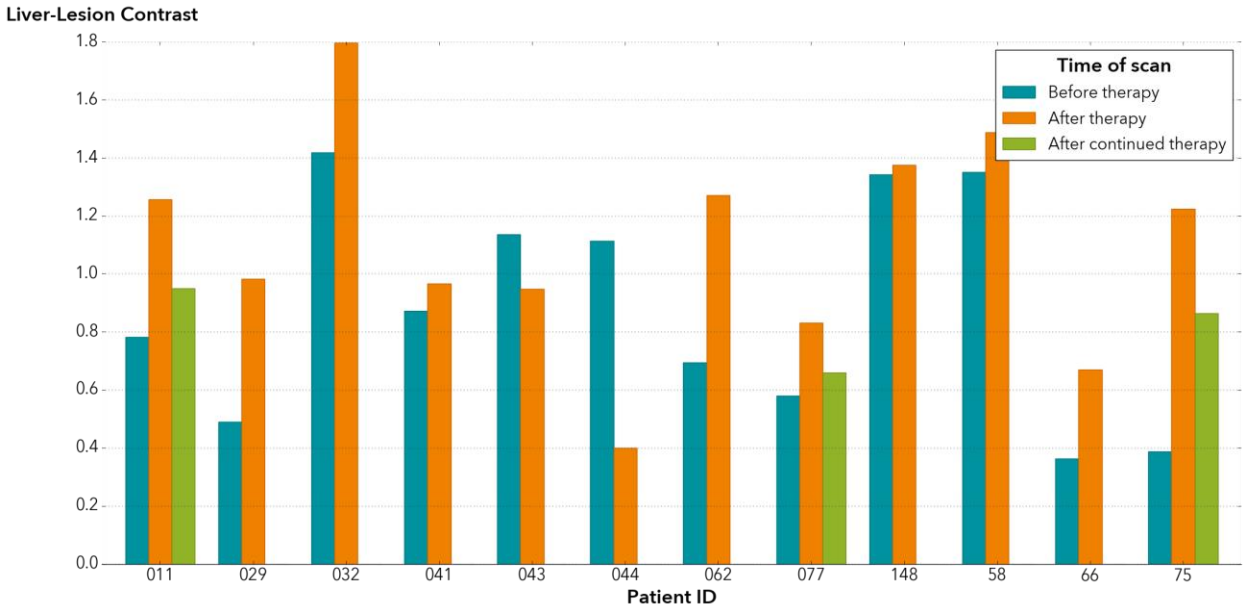

Figure 8. Liver-Lesion Contrast

LCTSC MORPHOMETRY: AN EXAMPLE USE CASE

MOTIVATION

Patients with multiple sclerosis (MS) often develop lesions on their spinal cord and suffer from loss of volume within their spinal cord. This loss of volume can be an important indicator of long-term disability from MS (Andelova et al. 2019). Segmentation of the spinal cord and lesions can provide measures of damage, which are key criteria for the diagnosis and monitoring of patients with MS (Gros, 2018). Automating this contouring process eliminates variability between radiologists. In this example, we apply a similar methodology to construct an automatic pipeline for spinal cord segmentation. The Jupyter notebook for this use-case is available for download here. This notebook gives users the opportunity to run the automatic spinal cord segmentation pipeline using a data set that is publicly available for download and use.

DATA ACQUISITION AND PREPROCESSING

The data used in this experiment is from Lung CT Segmentation Challenge (LCTSC) data set available at the Cancer Imaging Archive (Yang et al. 2017) and consists of 60 patients. The organs-at-risk (OARs) that are included in this challenge consist of annotations for the esophagus, heart, left and right lungs, and spinal cord. The DICOM-RT files contain contours for each of these organs and are displayed in different colors within the image.

The first step of pre-processing the images for spinal cord segmentation is to use the `roi2mask` step to filter out the organs that are not the spinal cord within the DICOM-RT files. This is executed by specifying the color of the spinal cord within the new parameter `roidisplaycolor`:

```
s.biomedimage.processbiomedimages(images=dict(table=imrt),
    steps=[dict(stepparameters=dict(steptype='roi2mask',
          roi2maskparameters=dict(roi2masktype='dicomrt_specific',
                                  roicontoursequence='_ROIContourSequence_,
                                  correctionsensitivity=.25,
                                  pixelintensity=255,
                                  outputbackground=0,
```

```
                                roidisplaycolor=colors[0]))),
        dict(stepParameters=dict(stepType='rescale',
                rescaleparameters=dict(rescaleType="channeltype_8u")))],
    casout=dict(name= 'masks_all', replace=True),
    copyvars=['_id_', 'color', 'RTID', '_label_'])
```

Figure 9 exhibits the contours for a patient before and after this filtering process. With the remaining contours containing only those for the spinal cord, segmentation masks are created for model training.

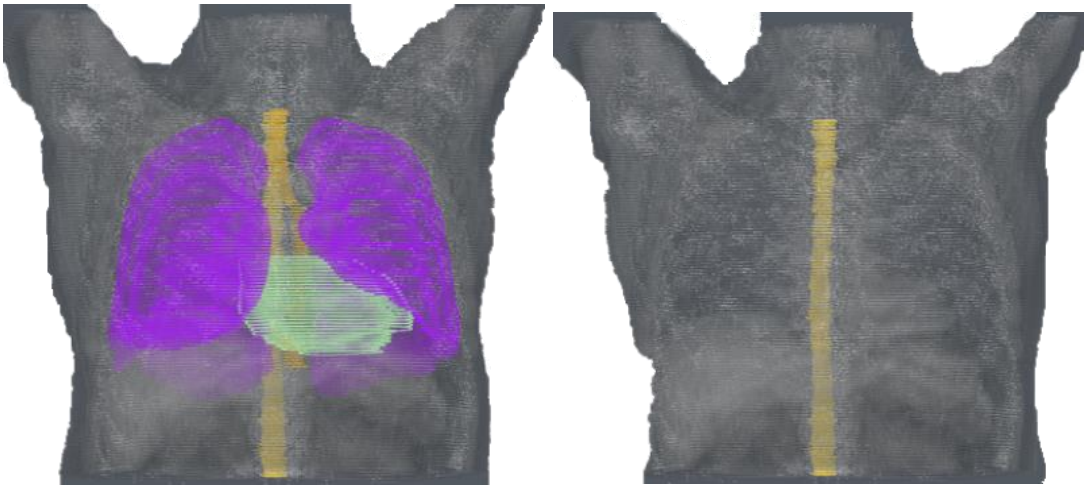

Figure 9. DICOM-RT Contours Before and After Filtering by Color

The image data is divided into training, test, and validation sets. The data consists of 60 patients total where 36 images (2106 slices) are used for training, 7 images (403 slices) are used for validation and 17 images (634 slices) are in the test set. Comparable to the CRLM data, the 3-D images are exported into 2-D slices for training that have a resolution of 512x512.

SEGMENTATION METHODOLOGY

A U-Net is trained for spinal cord segmentation given the segmentation masks created using `roi2mask`. This network is not being cascaded, as the U-Net is trained directly for spinal cord segmentation. The model is built using the U-Net DLPy API:

```
model = UNet(s,
             n_classes=2,
             width=512,
             height=512,
             n_channels=1,
             bn_after_convolutions=False)
```

This default model almost exactly resembles the models trained for liver and lesion segmentation with the main exception being in the kernel size of the last convolutional layer and the lack of activation function in the segmentation layer. This model has 34513282 parameters and uses an Adam solver with a learning rate of 0.0001. This model is trained from scratch for spinal cord segmentation over 50 epochs.

RESULTS

The test set is scored for spinal cord segmentation and these results are imported back into 3-D. The segmentation results for one patient are built and plotted on the original CT scan using Mayavi software. This display is shown in Figure 10, where the red portion of the

image indicates the region that is predicted by the model and the green region indicates the ground truth spinal cord ROI.

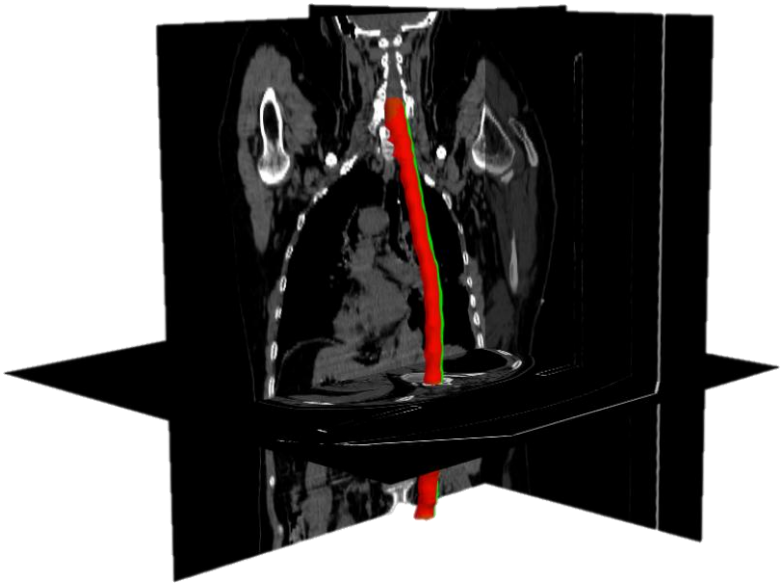

Figure 10. Spinal Cord Segmentation Results

This test set is evaluated using the DICE score coefficient against the ground truth spinal cord contours. The DICE coefficient averaged 71.349% on the test set and 69.547% on the validation set. The results are then quantified using quantifyBioMedImages and the volumes are plotted against the original spinal cord volumes annotated by the radiologist. These predicted volumes along with their ground truth comparison are displayed in Figure 11.

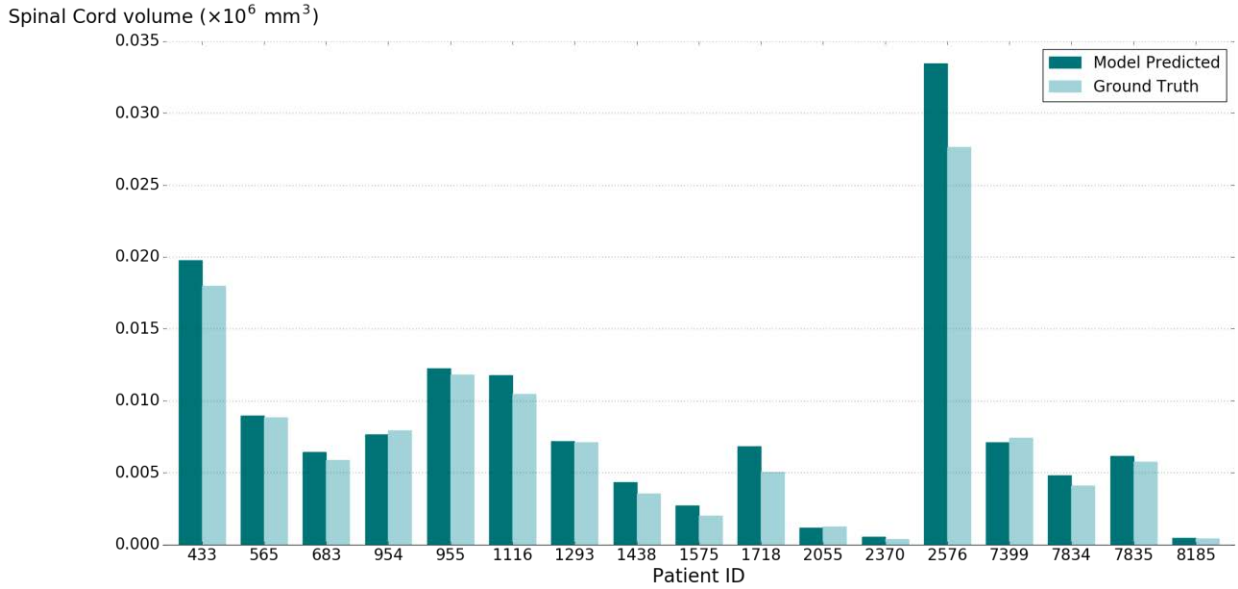

Figure 11. Spinal Cord Volume from Model-Predicted Segmentation Results

The predicted model segmentation averaged a 11.1586% decrease in volume when compared to the ground truth segmentation.

DISCUSSION

The ideas presented in this paper showcase the automation of segmentation and morphometric analysis for biomedical images in SAS® Viya®. This medical contouring application can aid radiologists in the clinic for detection and diagnosis of disease, as well as help track the health of the patient throughout the course of a disease. This automatic process reduces the burden and fatigue placed on medical professionals to perform rigorous manual contouring. By reducing the number of arduous tasks that medical professionals must face, we therefore reduce the risk of human error in the clinic. In addition, visualization and quantification of model predictions can help with personalized medicine for patients within the clinic. These derived measures can aid medical professionals with important decisions about a patient's long-term treatment.

The two use case examples show how automatic segmentation can assist clinicians in deriving important biomarkers. In the CRLM case, we overcame limitations of RECIST by capturing the total volume of the lesions rather than relying on their 1-D representation. A visual analysis of the predicted regions shows that in the before therapy case, the model tends to underpredict the total volume of the lesion. This is most likely due to the fact that before chemotherapy, lesions lack definitive boundaries. After chemotherapy, lesions tend to shrink, darken, and have more defined boundaries. As a result, the model has difficulty capturing the full extension of these pre-chemotherapy lesions within its prediction.

The second example using LCTSC data demonstrates a pipeline that users can run with data that is publicly available. Once again, we show how the automatic segmentation method allows users to derive important clinical biomarkers with little manual effort. Loss of spinal cord volume can be a strong predictor of long-term disability in patients with MS. Therefore, tracking the spinal cord volume loss over a period of time can lead to significant insights about a patient's long-term health within the clinic. It should be noted that the spinal cord images are clipped to a range of slices for each patient. The dramatic difference in volumes between patients corresponds to the number of slices for each patient. Therefore, the patient volumes should not be compared against each other. In contrast to the CRLM case, this model tends to over-predict the spinal cord regions. The predicted spinal cord region from the model often extends outside the ground truth region and, in some cases, misclassifies small regions outside of the target area.

CONCLUSION

New biomedical image analysis features in SAS® Viya® 3.5 provide tools for data preparation, image segmentation, visualization, and quantification. If you wish to download and run the pipeline for spinal cord segmentation, please follow the link provided here. Through the demonstrated segmentation and morphometric analysis pipeline, users can create an efficient detection method for important structures within CT scans. These methods improve efficiency and accuracy of biomedical structure identification, reducing burden and fatigue of medical professionals. Once the models are trained on data, the automatic segmentation can replace or assist manual segmentation tasks by medical professionals in the clinic. The segmentations are evaluated by the DICE coefficient and are shown to be competitive with state-of-the-art methods. Volumetric and pixel analysis are used to track disease progression over time and provide substantial assistance to clinical assessments. In future work, we'll be focusing on the quantification of other important biomarkers, mainly from the IBSI standard (Zwanenburg, 2016), which will involve the expansion of quantifyBioMedImages.

REFERENCES

Andelova, M., et al. 2019. "Additive Effect of Spinal Cord Volume, Diffuse and Focal Cord Pathology on Disability in Multiple Sclerosis." Front. Neurol.

Bilic, P., et al. 2019. ‘‘The Liver Tumor Segmentation Benchmark (LiTS),” arXiv:1901.04056. Available https://arxiv.org/abs/1901.04056

Christ, P. F., et al. 2016. “Automatic Liver and Lesion Segmentation in CT Using Cascaded Fully Convolutional Neural Networks and 3D Conditional Random Fields.” In Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016, ed. Ourselin, S., et al, 415-423. Cham: Springer.

Gros, C., et al., 2018. “Automatic segmentation of the spinal cord and intramedullary multiple sclerosis lesions with convolutional neural networks.” Neuroimage, 184:901–915.

Eisenhauer E.A., et al. 2009. “New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1).” European J Cancer, 45(2):228-247.

Huiskens J., et al. 2015. Treatment strategies in colorectal cancer patients with initially unresectable liver-only metastases, a study protocol of the randomised phase 3 CAIRO5 study of the Dutch Colorectal Cancer Group (DCCG).” BMC Cancer, 15:365.

Ramachandran, P., and G. Varoquaux. 2011. “Mayavi: 3D Visualization of Scientific Data.” IEEE, 13(2):40-51.

Ronneberger, O., et al. 2015. “U-Net: Convolutional Networks for Biomedical Image Segmentation.” MICCAI, 9351: 234–241

Vadakkumpadan, F., and J. Huiskens. 2019. “Medical Image Analytics in SAS® Viya® with Applications in the Treatment of Colorectal Cancer Spread to the Liver” Proceedings of the SAS Global Forum 2019 Conference. Cary, NC: SAS Institute Inc . Available https://www.sas.com/content/dam/SAS/support/en/sas-global-forum-proceedings/2019/3341-2019.pdf.

Vadakkumpadan, F., and S. Sethi. 2018. “Biomedical Image Analytics Using SAS® Viya®.” Proceedings of the SAS Global Forum 2018 Conference. Cary, NC: SAS Institute Inc . Available https://www.sas.com/content/dam/SAS/support/en/sas-global-forumproceedings/2018/1961-2018.pdf.

Yang, Jinzhong; Sharp, Greg; Veeraraghavan, Harini ; van Elmpt, Wouter ; Dekker, Andre; Lustberg, Tim; Gooding, Mark. (2017). Data from Lung CT Segmentation Challenge. The Cancer Imaging Archive. http://doi.org/10.7937/K9/TCIA.2017.3r3fvz08

Yoon, S. H., et al. 2016. “Observer variability in RECIST-based tumour burden measurements: a meta-analysis.”Eur J Cancer, 53:5-15.

Zwanenburg, A. 2016. “Image biomarker standardisation initiative”, 123. EP-1677.

ACKNOWLEDGMENTS

We thank Dr. Geert Kazemier at AUMC for providing us with the image data, and Dr. Nina Wesdorp and Sam at AUMC for annotating the images.

RECOMMENDED READING

- *SAS® Visual Data Mining and Machine Learning 8.5: Programming Guide*

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Courtney Ambrozic
SAS Institute, Inc.
+1 919 531 2125
Courtney.ambrozic@sas.com

Fijoy Vadakkumpadan
SAS Institute, Inc.
+1 919 531 1943
fijoy.vadakkumpadan@sas.com

Joost Huiskens
SAS Institute, Inc.
+31 35 6996 831
joost.huiskens@sas.com