Richard Bieck*, David Baur, Johann Berger, Tim Stelzner, Anna Völker, Juliane Neumann, Christoph-E. Heyde, Thomas Neumuth

An interactive system for muscle and fat tissue identification of the lumbar spine using semantic segmentation

Abstract:

We introduce a system that allows the immediate identification and inspection of fat and muscle structures around the lumbar spine as a means of orthopaedic diagnostics before surgical treatment. The system comprises a backend component that accepts MRI data from a web-based interactive frontend as REST requests. The MRI data is passed through a U-net model, fine-tuned on lumbar MRI images, to generate segmentation masks of fat and muscle areas. The result is sent back to the frontend that functions as an inspection tool. For the model training, 4000 MRI images from 108 patients were used in a k-fold cross-validation study with $k = 10$. The model training was performed over 25-30 epochs. We applied shift, scale, and rotation operations as well as elastic deformation and distortion functions for image augmentation and a combined objective function using Dice and Focal loss. The trained models reached a mean dice score of 0.83 and 0.52 and a mean area error tissue of 0.1 and 0.3 for muscle and fat tissue, respectively. The interactive web-based frontend as an inspection tool was evaluated by clinicians to be suitable for the exploration of patient data as well as the assessment of segmentation results. We developed a system that uses semantic segmentation to identify fat and muscle tissue areas in MRI images of the lumbar spine. Further improvements should focus on the segmentation accuracy of fat tissue, as it is a determining factor in surgical decision-making. To our knowledge, this is the first system that automatically provides semantic information of the respective lumbar tissues.

Keywords: Semantic Segmentation, Machine Learning, U-Net, Tissue Diagnostics.

<https://doi.org/10.1515/cdbme-2021-2099>

1 Introduction

In orthopaedic diagnostic of the spine, the grade of fatty infiltration of the muscle tissue evidently correlates with subjective symptoms, e.g., lower back pain, and may be an objective indicator for pathology-induced disability [1]. However, the complex and sometimes irregular distribution of fat and muscle tissue of the lumbar spine is currently only assessable through direct inspection without means for automated measurement. This diagnostic step is time-consuming and requires expert knowledge and routine practice of segmentation tools to generate tissue quantification results in a presentable way. As a response, we introduce a diagnostic tool for the automated detection and interactive exploration of fatty infiltration in the paravertebral musculature. Since deep neural networks (DNN) already provide reliable information of pathologies in the intervertebral discs [2], we adopted a similar approach and implemented a CNN for the identification of fat and muscle tissue from MRI images. In a recent study, Weber et al. [3] combine the V-Net model output of segmented muscle tissue with information derived from fat-water images to reliably produces a fat infiltration score that correlates with patient disability. Similarly, we apply a U-Net model, that is extensively used in medical image segmentation studies [4], for the segmentation task but include both the muscle and fatty tissue masks in the training process of our model. We deploy the model in an interactive environment to enable the exploration of patient-individual MRI data and segmented muscle and fatty tissue in form of a web-based diagnostic tool.

*Corresponding author: **Richard Bieck:** Innovation Center Computer Assisted Surgery, University Leipzig, Semmelweisstraße 14, Leipzig, Germany, richard.bieck@medizin.uni-leipzig.de

David Baur, Anna Völker, Christoph-E. Heyde: Department for Orthopaedics, Trauma Surgery and Plastic Surgery, University Hospital Leipzig AöR, Leipzig, Germany

Johann Berger, Tim Stelzner, Juliane Neumann, Thomas Neumuth: Innovation Center Computer Assisted Surgery, University Leipzig, Leipzig, Germany,

2 Methods

2.1 Data Preparation

The MRI imaging data were chosen from patients with non-traumatic back pain of the lower spine. We excluded patients with traumatic injuries, ages below seventeen, pre-existing spine-implants, or detectable tumors or infections. For each included patient, consecutive axial images of the lumbar spine of two or more lumbar segments were used, resulting in 30 to 60 images per patient. Overall, 3650 axial T2-weighted MRI images from 100 patients acquired with three different MRI systems at one clinical site were included in the training of the segmentation model. The ground truth masks for segmented muscle and fatty tissue areas were created by a spine surgery specialist with 10 years of clinical and radiological experience using Mimics¹ segmentation software. All segmentations were controlled by a second reviewer (certified spine surgeon).

First, the ground truth masks for the complete tissue area comprising muscle and fatty infiltration were acquired. As a segmentation constraint, both the erector spinae muscle and the multifidus muscle had to be included. In a second step, the ground truth mask for muscle tissue was derived through manual threshold-based segmentation of the complete tissue masks leaving out hypersignal pixels primarily corresponding to fat. Lastly, the ground truth fat masks were generated through subtraction of the muscle and complete tissue masks.

2.2 Model Training

For the semantic segmentation task, we used a U-Net model with an initial image feature size of 32. The model performs cascaded convolutions followed by cascaded upsampling steps

to first reduce the latent feature space and then combine throughput data with previous feature representations to generate output masks.

For the model training, a k-fold cross-validation study with $k = 10$ was set up for the 100 patient datasets. The images of left-in datasets were compiled into one dataset, shuffled, and split into training and validation sets with an 85:15 - ratio. We applied augmentation steps to the images and ground truth masks using a sequence of randomly initialized operations of shifting (between - 0.2 to + 0.2 of image height and width), scaling (between - 0.2 to + 0.2), and rotation (between - 30° and + 30°) followed by a nonlinear optical transformation randomly chosen and applied from the following options: (1) elastic deformation, (2) optical distortion, or (3) grid distortion (see fig 1). All augmentation operations were performed using the Albumentations framework [5].

We used a U-Net with weights pre-trained on the brain MRI dataset provided by [6]. The model was then fine-tuned to our data using an Adam optimizer with an initial learning rate of $l_r = 1 \cdot 10^{-3}$ that was decayed every 4 epochs by a factor of 0.5, a batch size of $b = 24$, and an objective function L_c that combines dice loss and focal loss. The combination of a region- and a distribution-based loss function enabled us to penalize both the incorrectly segmented tissue boundaries as well as expected over-fitting behavior due to our relatively small dataset and potential data bias. The optimization process was controlled using the intersection-over-union (IOU) metric and the dice score (DSC). The process was terminated when changes of either metric fell below 1 % between epochs. In this way, models were trained over 27 epochs on average (range 25-30). The stated hyperparameter values represent the final training configuration after semi-automatic parameter search for batch size, learning rate, and optimization step size. All data processing and model computations were performed

Figure 1: Example for augmentation pipeline prior to model training: original image (left), shifted, scaled and rotated image (middle), and distortion (right)

¹ <https://www.materialise.com/en/medical/>

Figure 2: Intersection-over-Union scores for muscle and fat segmentation results of all MRI images.

using Pytorch v1.7.0² with Python v3.6 and on an NVIDIA GTX 2060 with 8GB RAM.

2.3 System Design

The trained segmentation models were incorporated into a server-client system that functions as a diagnostic tool. On the server side, a single backend instantiates the trained segmentation model, provides necessary GPU resources, and performs all model inference steps. The backend workstation receives HTTP requests that contain the target images to be processed through a REST interface implemented using FLASK v2.0.0³ and GUNICORN v20.1.0⁴. The second system component provides a client-side inspection tool through which the physician interacts with the patient data. A patient MRI dataset can be loaded into the tool and inspected in a web-based dashboard with all patient-related parameters retrieved from DICOM meta information. When a patient dataset is loaded, the MRI images are serialized and sent as part of the get-request to the REST interface of the backend. After the deserialized images are passed through the model, the generated tissue masks are sent back to the inspection tool. From the acquired segmentation results, the inspection tool calculates the cross-sectional tissue area (CSA) for muscle and fat and the corresponding fat-to-muscle-area ratio (FMR) as a characterization of the grade of fatty infiltration. On the dashboard, the distribution of both metrics across all loaded

MIR images is then presented to the physician in an ordered fashion.

As an extension to the described dashboard for tissue diagnostics, the tool also incorporates an additional dashboard that comprises all cross-validation study results for the comparison between automatically segmented and ground truth segmentation masks. We included the predicted and ground truth values for the IOU and DSC metrics as well as the FMR and CSA errors. In this way, the patient-individual results for our semantic segmentation tasks can be explored and allow an intuitive investigation of examples for strong and weak model performance. The outlined system is scalable to handle several simultaneous client-side applications and requests to ensure that all diagnostic activities are based on the same inference model.

3 Results

The models reached a mean DSC score of 0.83 and 0.52 as well as an IOU score of 0.77 and 0.40 for muscle and fat tissue segmentation, respectively (see fig 2). When looking at the intra-patient distribution of segmentation results, the DSC and IOU scores were higher towards the cranial direction of the imaging volume and decreased towards the caudal end of the spine. For the CSA, the model exhibited a mean area error of 12.8 % and 36.3 % for muscle and fat tissue and a mean FMR error of 5 %. The distribution of segmentation results shows that 75 % of identified muscle tissue had a DSC score of over 0.75 while 50 % of the identified fat tissue had a DSC score of over 0.55. We observed several cases where no fat tissue was identified in MRI images of the most caudal section of the dataset. Looking at the individual results for segmented tissue, the errors for muscle tissue stemmed from over-segmentation of a broader tissue area, which is also reflected in higher CSA errors. For fat segmentation errors, we observed consistently high CSA errors due to under-segmentation of tissue area occasionally coupled with segmentation holes for tissue areas that were closed in ground truth data.

The web-based inspection tool was successfully implemented with its backend and client-side components. The dashboard offered a fast and intuitive summary view of patient data and can be easily extended by additional functions. The segmentation request that is sent through the REST interface returns tissue information in about 25 ms per processed image. We performed no further limit testing around the backend

² <https://pytorch.org/>

³ <https://flask.palletsprojects.com/en/2.0.x/>

⁴ <https://gunicorn.org/>

server but consider the available GPU memory to be the determining performance factor for simultaneous multi-user access. The introduced web-based diagnostics tool was evaluated by clinicians to be suitable for the presentation and exploration of segmentation results especially when investigating the distribution of tissue parameters over the complete MRI dataset.

4 Discussion

We successfully introduced a web-based diagnostics tool to identify the muscle area and fat infiltration of the erector spinae and multifidus muscle in consecutive axial lumbar MRI slices. The model performance for the segmentation of muscle tissue is comparable to the results of Weber et al. [3]. The over-segmentation of cross-sectional muscle area potentially stems from bias in our training database. We consider the under-segmentation of fatty infiltration in the muscle to also originate from this bias. The lowest DSC and IOU scores for fatty tissue areas are associated with: (1) the most caudal image sections of patient datasets with minimal bone, muscle, and fat areas available and (2) a streaky infiltration pattern that is ignored. The segmentation errors can be attributed model over-fitting and an incomplete representation of fatty infiltration patterns. In further studies, the model architecture should be customized to reflect volumetric information of adjacent MRI images and tissue segments as proposed in Qamar et al. [7]. Alternatively, the down-convolution part of the U-Net could be substituted with an appropriate 3D feature extractor, e.g., ResNext. As outlined in [8], the performance of a U-Net model directly scales with the number of available training data. Increasing the database while most likely also improve the segmentation results. Other architectures that show better performance on smaller datasets, e.g., generative adversarial networks (GAN), could improve model performance as well. Lastly, the employed objective function should be further investigated to better consider smaller image details in the training process.

5 Conclusion

We developed a system that identifies fat and muscle tissue areas in MRI images of the lumbar spine and provides physicians with an interactive web-based diagnostics tool to explore segmentation results. Further improvements should focus on the segmentation performance for fatty tissue. To our knowledge, this is the first system that automatically provides

automatic segmentation of both muscle and fatty tissue of the lumbar region.

Author Statement

Research funding: The author state no funding involved. Conflict of interest: Authors state no conflict of interest. Informed consent: Informed consent has been obtained from all individuals included in this study. Ethical approval: The research related to human use complies with all the relevant national regulations, institutional policies and was performed in accordance with the tenets of the Helsinki Declaration, and has been approved by the author's local ethics committee (025/21-ek).

References

- [1] P. Kjaer, T. Bendix, J. S. Sorensen, L. Korsholm, and C. Leboeuf-Yde, "Are MRI-defined fat infiltrations in the multifidus muscles associated with low back pain?," *BMC Med*, vol. 5, no. 1, p. 2, Dec. 2007, doi: 10.1186/1741-7015-5-2.
- [2] S. Kim, W. Bae, K. Masuda, C. Chung, and D. Hwang, "Fine-Grain Segmentation of the Intervertebral Discs from MR Spine Images Using Deep Convolutional Neural Networks: BSU-Net," *Applied Sciences*, vol. 8, no. 9, p. 1656, Sep. 2018, doi: 10.3390/app8091656.
- [3] K. A. Weber et al., "Deep Learning Convolutional Neural Networks for the Automatic Quantification of Muscle Fat Infiltration Following Whiplash Injury," *Sci Rep*, vol. 9, no. 1, p. 7973, Dec. 2019, doi: 10.1038/s41598-019-44416-8.
- [4] N. Siddique, S. Paheding, C. P. Elkin, and V. Devabhaktuni, "U-Net and Its Variants for Medical Image Segmentation: A Review of Theory and Applications," *IEEE Access*, vol. 9, pp. 82031–82057, 2021, doi: 10.1109/ACCESS.2021.3086020.
- [5] A. Buslaev, V. I. Iglovikov, E. Khvedchenya, A. Parinov, M. Druzhinin, and A. A. Kalinin, "Albumentations: Fast and Flexible Image Augmentations," *Information*, vol. 11, no. 2, p. 125, Feb. 2020, doi: 10.3390/info11020125.
- [6] M. Buda, A. Saha, and M. A. Mazurowski, "Association of genomic subtypes of lower-grade gliomas with shape features automatically extracted by a deep learning algorithm," *Computers in Biology and Medicine*, vol. 109, pp. 218–225, Jun. 2019, doi: 10.1016/j.compbiomed.2019.05.002.
- [7] S. Qamar, H. Jin, R. Zheng, P. Ahmad, and M. Usama, "A variant form of 3D-UNet for infant brain segmentation," *Future Generation Computer Systems*, vol. 108, pp. 613–623, Jul. 2020, doi: 10.1016/j.future.2019.11.021.
- [8] M. Bardis et al., "Deep Learning with Limited Data: Organ Segmentation Performance by U-Net," *Electronics*, vol. 9, no. 8, p. 1199, Jul. 2020, doi: 10.3390/electronics9081199.