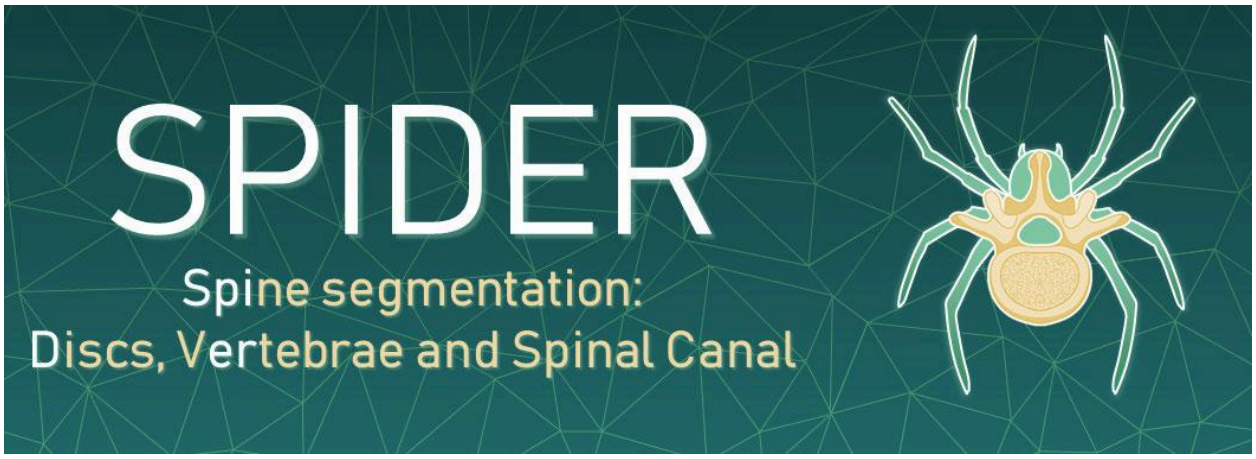
Lumber Spine Segmentation in MR Images (SPIDER)

Deep Learning Course CIT-654

Spring 2024



Abdelrahman Mohamed Amer **231001551**

Abdullah Mohamed Abdelhadi **231002028**

Mohamed Nasr Tamam **231002656**

Abstract

This project presents a large, multi-center magnetic resonance (MRI) image dataset of the lumbar spine with reference segmentations of anatomical structures. The dataset includes 447 T1- and T2-weighted sagittal MR series from 218 patients. An iterative data annotation approach was used to segment the vertebrae, intervertebral discs, and spinal canal in each image. A segmentation algorithm was initially trained on a small subset and then used to provide automatic segmentations of remaining images, which were manually corrected and added to the training data.

The large, multi-center lumbar spine MR dataset with reference segmentations of key anatomical structures aims to advance the development of automatic image analysis methods. The continuous segmentation challenge also encourages collaboration to improve lumbar spine MR interpretation through machine learning and help alleviate the global burden of low back pain.

INTRODUCTION

Low back pain (LBP) represents the most significant global burden of disease, accounting for the highest number of years lived with disability compared to any other disease. Consequently, lumbar spine magnetic resonance imaging (MRI) has become one of the most frequently employed imaging procedures in musculoskeletal diagnostics. In the United States, a substantial 93% of lumbar MRI referrals align with the American College of Radiology guidelines, yet only a mere 13% of these scans impact clinical decision-making. To enhance the diagnostic utility of MRI, the integration of automatic image analysis could prove pivotal by facilitating more objective and quantitative image interpretation. The initial step towards automated lumbar spine MRI assessment involves the segmentation of key anatomical structures, including vertebrae, intervertebral discs (IVDs), and the spinal cord.

Background

Recent strides in machine learning and artificial intelligence (AI) have led to the development of advanced spine segmentation algorithms, predominantly leveraging learning-based techniques that require meticulously curated training data. The evolution of vertebra segmentation algorithms for CT images has significantly benefited from

several extensive publicly available datasets containing CT images and corresponding reference segmentations. However, an equivalent comprehensive, high-quality dataset for lumbar spine MRI is currently unavailable. Existing datasets tend to be limited in scope, either in terms of size, focus solely on vertebral body segmentation, or provide annotations in only the midsagittal slice (2D). Furthermore, many datasets are restricted to a single anatomical structure crucial for evaluating multifactorial disorders like LBP, such as vertebrae or IVDs.

Study Objective

To perform the development of segmentation algorithms and ultimately advance automated image analysis for lumbar spine MRI, this study aims to achieve three primary objectives:

- Dataset Presentation: To introduce an extensive multi-center lumbar spine MRI dataset that includes reference segmentations for vertebrae, IVDs, and the spinal canal, alongside per-level radiological gradings.

- Segmentation Challenge: To establish a continuous lumbar spine MRI segmentation challenge, for three anatomical structures in lumbar spine MRI:
    1. Vertebrae
    2. Intervertebral discs (IVDs)
    3. Spinal canal

  providing a platform for algorithm developers to submit their models for evaluation.

- Performance Metrics: To offer reference performance metrics for MONAI 3D, a widely recognized framework for 3D medical image analysis with publicly available training and inference code, which will be used to segment all three spinal structures automatically.

PROCEDURE

1. Data Collection
2. Data Segmentation
3. Network Archeticture
4. Training
5. Evaluation

Data Collection

In total, 257 lumbar spine studies from patients with a history of LBP were retrospectively collected, with each study consisting of up to three MRI series. Of these 39 patient studies, containing 97 MRI series, were sequestered for public benchmarking. The public data release described in this paper consists of 218 patient studies with 447 series. The study was approved by the institutional review board at Radboud University Medical Center (IRB 2016–2275). Informed consent was exempted, given the retrospective scientific use of deidentified MRI scans. Studies were collected from four different hospitals in the Netherlands, including one university medical center (UMC), two regional hospitals and one orthopedic hospital (data acquired between January 2019 and March 2022). All involved hospitals signed either a data transfer agreement or a public data sharing form in which public sharing of the data under a CC-BY 4.0 license was disclosed.

Data originating from the UMC were all available lumbar spine MRI studies of patients presenting with (chronic) low back pain between January 2019 and November 2020 that included a T2 SPACE sequence. This sequence produces images with almost isotropic spatial resolution (voxel size: 0.90 × 0.47 × 0.47 mm). All studies also contained both a standard sagittal T1 and T2 sequence (voxel size: 3.30 × 0.59 × 0.59 mm). MRI studies were only excluded if the image quality was considered too low for fully manual segmentation (n = 4). Data originating from the other three hospitals were sets of consecutive lumbar spine MRI studies of patients presenting with (chronic) low back pain with at least a sagittal T1 or a sagittal T2 sequence. The voxel size of these images ranged from 3.15 × 0.24 × 0.24 mm to 9.63 × 1.06 × 1.23 mm. At each center, we included a fixed number of consecutive MRI studies that met these inclusion criteria.

There were no other exclusion criteria except for several studies across all four contributing centers being excluded from publication to serve as hidden hold-out test set for an algorithm-development challenge (see the Segmentation data section for further details). Additional dataset characteristics are given in Table 1.

Table 1 Overview dataset.

From: Lumbar spine segmentation in MR images: a dataset and a public benchmark

Hospital	Studies	T1	T2	T2 SPACE	Voxel size range (min – max)(mm)	Sex (% female)
UMC	41	39	39	41	(3.24 × 0.27 × 0.47) - (3.34 × 0.59 × 0.85)*	55%
RH1	43	43	37	0	(0.46 × 0.46 × 4.20) - (9.63 × 1.06 × 1.06)	58%
RH2	44	24	44	0	(0.46 × 0.46 × 4.20) - (5.17 × 1.00 × 1.23)	59%
OH	90	90	90	0	(3.15 × 0.24 × 0.24) - (3.39 × 0.83 × 1.02)	68%
Total	218	196	210	41	(3.15 × 0.24 × 0.24) - (9.63 × 1.06 × 1.23)*	63%

Data PreProcessing

The dataset used in this spine segmentation project is sourced from a spinal cord imaging dataset, formatted for segmentation tasks.We had multiple challenges with the data as Class & region imbalance and Inconsistent image sizes and orientations.

Images and labels undergo several preprocessing steps to ensure consistency and enhance data quality, after reading data using simpleitk, we make statistics about input data to identify issues with the images, after which we made our processing to make it ready for the model.

First, the LoadImaged and EnsureChannelFirstd transforms load the images and labels and ensure they have the correct channel order. The pixel intensities are then scaled to a range of 0 to 1 using ScaleIntensityRanged, standardizing the input data for the neural network. The dataset is divided into training and validation sets as specified in the spider_dataset.json file. Efficient data handling is achieved using CacheDataset, which caches the datasets to accelerate access during training and validation. DataLoaders are configured for both training and validation datasets, supporting batch processing and multi-threaded data loading to optimize performance.

In summary, we do resembling as follows:

- Unified size: (32, 128, 128)
- Unify spacing based on the new size
- Add interpolators

- Set unified orientation
- Save new resampled files to directory

Archeticture

We Considered multiple architecture for this project as follows:

- 3D UNet
- 3D nnUNet
- 3D MONAI

We decided on using 3D Monai for :

- Ease of Use: Simplified coding with many embedded architectures
- Suite of Tools: Integrated data preprocessing, augmentation, model training, evaluation, and deployment
- CacheDataset: Efficient data loading and transformation for faster training
- Debugging Tools: Identify problems with mask labeling

The network architecture employed in this project is the UNETR (U-Net with Transformers), a state-of-the-art model implemented using MONAI (Medical Open Network for AI).

UNETR Archeticture Combines UNet and Transformers for effective medical image segmentation.

UNETR Components:

- **3D Input Volume:** The input is a 3D medical image with dimensions H×W×D and C channels.
- **Patching and Linear Projection:** The 3D input volume is divided into smaller non-overlapping 3D patches, which are then projected into a K-dimensional embedding space using a linear layer. Positional embeddings are added to retain spatial information.
- **Transformer Encoder:** Processes the sequence of embeddings with a stack of transformer layers, each consisting of Multi-Head Self-Attention (MSA) and Multi-Layer Perceptron (MLP) with GELU activation.
- **Skip Connections:** Intermediate representations from the transformer encoder

are reshaped back into 3D tensors and connected to the decoder at corresponding resolutions.
- **Decoder:** Uses 3×3×3 convolutional layers and deconvolution layers for up-sampling and merging features from the encoder to preserve spatial details.
- **Output Layer:** Produces the final output using a 1×1×1 convolutional layer with a softmax activation function for voxel-wise semantic predictions.

Training

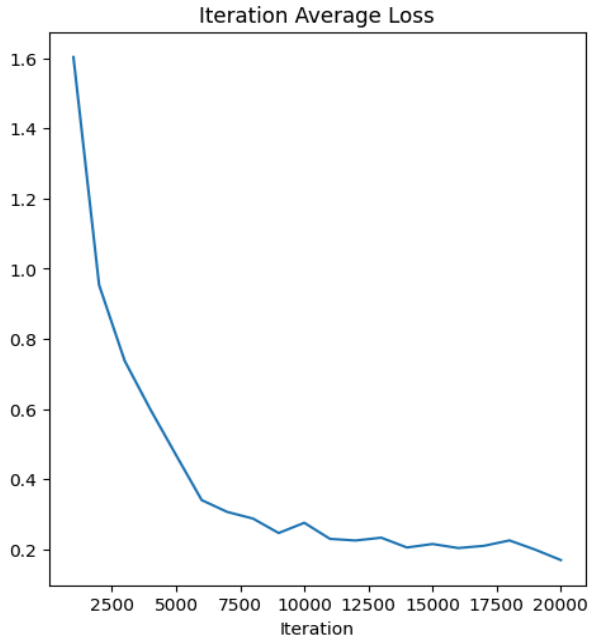
The training process involves optimizing the UNETR model using the AdamW optimizer with a learning rate of 1e-4 and a weight decay of 1e-5. The loss function employed is the Dice Cross Entropy Loss (DiceCELoss), which combines Dice loss and Cross Entropy loss to effectively handle class imbalances and improve segmentation accuracy. The training loop runs for up to 21,000 iterations, with evaluations performed every 1,000 iterations. During each training iteration, the model generates predictions through forward passes, computes the loss, and updates the model parameters via backpropagation. The training progress is visualized using the tqdm library. To ensure proper loss calculation, labels are clamped to a valid range. The best model, determined by the highest validation Dice metric, is saved for future use.

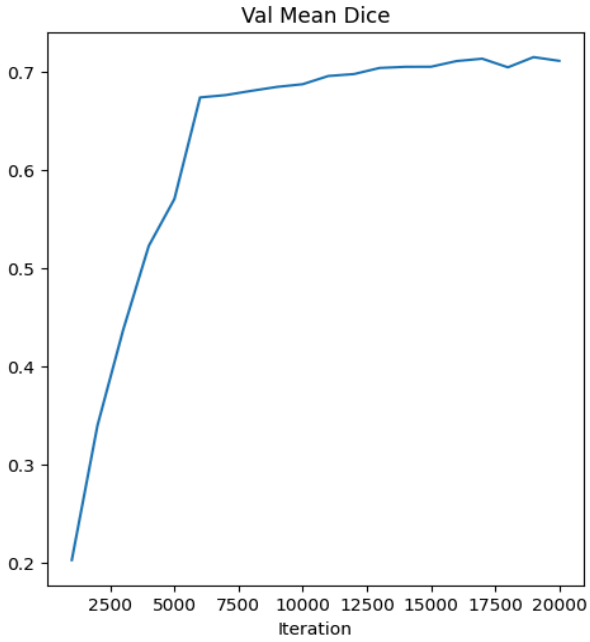
Evaluation & Results

The evaluation of the trained model is conducted using the validation dataset, with segmentation performance quantified by the Dice Metric. The evaluation process utilizes sliding window inference, wherein the model processes smaller patches of the volume sequentially to produce the final segmentation output. Predicted and ground truth labels are post-processed using AsDiscrete transforms to convert them into one-hot encoded and discrete formats suitable for Dice score calculation. The Dice Metric measures the overlap between the predicted and ground truth segmentation masks, providing a metric of accuracy. The average Dice score across the validation set is used to identify the best-performing model, which is then saved.

The result after completing the training and validation can be graphed as follow:

- Lowest loss: 0.15683
- Best metric: 0.7154





CONCLUSION

We successfully developed a spine medical segmentation model with an average Dice Score of 0.72. This model can help healthcare professionals save time and effort while reducing human error.Using the MONAI framework made the development process smoother and improved performance.We also gained a good understanding of data processing and handled computational limitations effectively.

REFERENCES

1. Hatamizadeh, A., Tang, Y., Nath, V., Yang, D., Myronenko, A., Landman, B., Roth, H.R. and Xu, D., 2022. Unetr: Transformers for 3d medical image segmentation. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (pp. 574-584).
2. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for

Biomedical Image Segmentation. arXiv preprint arXiv:1505.04597.
3. van der Graaf, J.W., van Hooff, M.L., Buckens, C.F.M. et al. Lumbar spine segmentation in MR images: a dataset and a public benchmark. Sci Data 11, 264 (2024). https://doi.org/10.1038/s41597-024-03090-w
4. R. Beare, B. C. Lowekamp, Z. Yaniv, “Image Segmentation, Registration and Characterization in R with SimpleITK”, J Stat Software, 86(8), https://doi.org/10.18637/jss.v086.i08, 2018.
5. Z. Yaniv, B. C. Lowekamp, H. J. Johnson, R. Beare, “SimpleITK Image-Analysis Notebooks: a Collaborative Environment for Education and Reproducible Research”, J Digit Imaging., https://doi.org/10.1007/s10278-017-0037-8, 31(3): 290-303, 2018.
6. B. C. Lowekamp, D. T. Chen, L. Ibáñez, D. Blezek, “The Design of SimpleITK”, Front. Neuroinform., 7:45. https://doi.org/10.3389/fninf.2013.00045, 2013.
7. title: "MONAI: An open-source framework for deep learning in healthcare" , doi: https://doi.org/10.48550/arXiv.2211.02701
8. Nyothiri Aung, Tahar Kechadi , Liming Chen and Sahraoui Dhelim , “UNet Architecture for 3D Medical Volume Segmentation,” in JOURNAL OF LATEX CLASS FILES, VOL. 14, NO. 8, AUGUST 2015.