Vertebral Detection and Labelling Using Deep Learning for Spine MRI Registration

Jonathan Chu[1,2] **Michael Hardisty**[3,4] **Alexander Wong**[2] **Stewart McLachlin**[1]
[1]Orthopaedic Mechatronics Lab, Mechanical and Mechatronics Engineering, University of Waterloo
[2]Vision and Image Processing Lab, System Design Engineering, University of Waterloo
[3]Physical Sciences, Sunnybrook Research Institute
[4]Department of Surgery, University of Toronto
{jh3chu, alexander.wong, stewart.mclachlin}@uwaterloo.ca
{m.hardisty}@utoronto.ca

Abstract

Medical image registration is an important but often challenging aspect for clinical image analysis. It has applications in treatment planning requiring image fusion, or inter-subject atlas based analyses, as well as longitudinal analyses. Spine registration presents extra challenges because of the variability in the field of view (FoV) of the spinal column between different image series and many vertebrae having a similar appearance leading to many local registration minima. To help improve spine registration robustness, we generate a labelled dataset of cervical spine magnetic resonance imaging (MRI) and successfully apply a Mask R-CNN model to localize and label vertebra. An automated method to generate labelled bounding boxes and masks can then be used to seed initial alignment or crop to appropriate FoV for subsequent affine and deformable spine MRI registration.

1 Introduction

Medical imaging is vital for many clinical workflows such as diagnosis, pre-operative planning, and post-operative evaluation. Magnetic resonance imaging (MRI) has revolutionized medical imaging and is used in virtually all medical sub-specialties [1]. This is due to its excellent soft tissue contrast and anatomic detail, in addition to the benefit of not utilizing ionizing radiation. Spine MRI is commonly used for the detection of pathologies such as infections, metastases, nerve root disorders, and disc abnormalities [2].

Image processing methods, such as image segmentation and deformable registration, are commonly used to properly identify and spatially align areas of interest for better pathological understanding [3]. To enhance clinical decision-making, MRI images taken with different weightings, such as anatomical (T1w and T2w) and diffusion weighted (DW) images, are often registered to align the features of interest. However, registration of spine MRI is notoriously difficult often requiring time consuming manual landmarking; improvements to clinical tools are needed.

To improve spine MRI registration, it is beneficial to guide registration algorithms to spatially align the same vertebrae and avoid mismatched vertebrae with poor spatial alignment. However, due to changes in patient position and scanning parameters, resultant image volumes can have different fields of view (FoVs) with varying vertebral levels and number of vertebrae, leading to the requirement of manual or semi-automated image registration.

To assist this process, the objective of this study was to evaluate the use of a Mask R-CNN model to identify and label specific vertebrae in MRI for use in a spine imaging analysis pipeline. This component will allow for the automated detection of individual vertebra within multi-modal images, which can then be used to seed initial alignment or crop to appropriate FoV for subsequent affine and deformable registration of spine MRI images.

1.1 Background

Mask R-CNN is a state-of-the-art object detection and instance segmentation model that builds upon the Faster R-CNN model [4]. The model consists of a region proposal network (RPN) that utilizes a backbone network (usually ResNet). The RPN proposes regions of interest (ROIs) in the image where bounding boxes, classification labels, and segmentation masks of detected objects are generated by three separate heads. This includes a softmax classifier, bounding box regressor, and binary classifier head.

During training, the model utilizes a multi-task loss, L, on each sampled ROI by summing the following: $L_{classifer}$, L_{box}, and L_{mask}. $L_{classifer}$ is a categorical cross-entropy loss responsible for supervising the classification task of labelling the bounding boxes proposed in the ROI [5]. L_{box} is a regression loss of the bounding box coordinates utilizing a smooth L1 loss [5]. Lastly, L_{mask} is an average binary cross-entropy loss responsible for the binary mask generated by the model [4].

2 Dataset

This study utilized the Spine Generic dataset, which consists of T1w, T2w, and DW images from 267 healthy subjects imaged from 42+ centers around the world [6]. The dataset consists of images taken with the spine generic protocol, a quantitative MRI protocol, to ensure consistent imaging for spinal cord research. This includes a consistent voxel size of 1mm^3 with images being 192x260x320 voxels. All images included the entire skull and cervical spine, ended at varying points within the thoracic spine. Figure 1 a) depicts an example subject MRI from the dataset. For the purposes of this study, only the T1w, 2D mid-sagittal slices of the MRI images were utilized.

3 Methodology

The pipeline for vertebrae detection and labelling utilized a Mask R-CNN model to predict each vertebra within the MRI with a bounding box, vertebra class label, and segmentation mask.

3.1 Data Pre-processing

Data pre-processing was required to generate labelled data for the supervised training of the Mask R-CNN model. This included generating class labels, segmentations, and bounding boxes of the vertebrae within the MRI. Of note, the Spine Generic MRI dataset includes the brain and skull, which were not of interest to this investigation and were excluded from the analysis using a crop to a cervical spine only region of interest. Figure 1 shows the data pre-processing pipeline in full detail.

3.1.1 Spinal Cord Toolbox

The Spinal Cord Toolbox (SCT) was utilized as a method to gather ground truth annotations of the Spine Generic dataset in a semi-automated way. SCT is an open source package that was developed for the analysis and processing of spinal cord MRI by establishing standardized templates and analysis procedures [7]. While SCT was developed for spinal cord analysis, it also provides robust tools for spinal vertebrae processing, including spinal cord segmentation and labelling of vertebral levels.

Generation of the spinal cord segmentation and labels of vertebral levels was first performed using SCT's Deepseg algorithm (Figure 1 b)), which utilizes a 2D U-Net convolutional neural network [8]. SCT was then used to register the PAM50 template, an unbiased multi-modal MRI template including the entire spinal column, to the image as seen in Figure 1 c) [9]. This created baseline spine segmentations of the vertebrae. Post SCT processing, the images included segmentations of the spinal cord, labels of vertebral levels on the spinal cord, and warped spine segmentations from the PAM50 template.

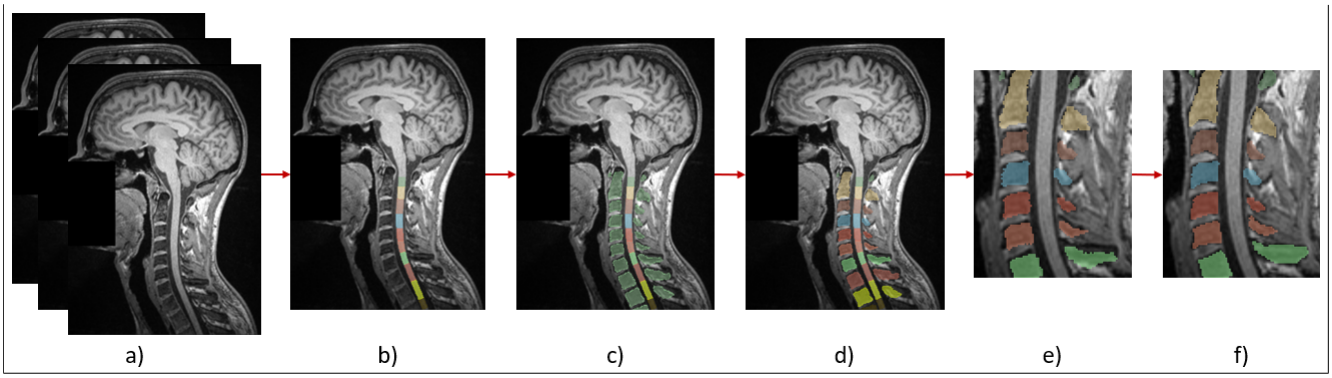

Fig. 1: Data pre-processing pipeline to obtain ground truth annotations from the Spine Generic Dataset. a) The original Spine Generic T1w MRI images of healthy subjects. b) Initial segmentation and labelling of the spinal cord using SCT. The differing coloured segments along the spinal cord correspond to the labelled vertebral levels. This image is labelled from C1-T3. c) Registration of PAM50 template spinal column onto the MRI image using SCT. d) Splitting the PAM50 spinal column into individual vertebrae and relabelling using the spinal cord vertebral level labels in 3DSlicer. e) Cropping to the cervical spine (C1-C7) after quality checks. f) Manual segmentation clean-up and removal of any first thoracic vertebra segmentation if visible.

3.1.2 3D Slicer

3D Slicer is an open source platform for medical image computing [10]. 3D Slicer was used to generate individual vertebral segmentations from the warped PAM50 spinal column segmentation by splitting it into individual islands (or vertebra). Labels of the vertebra were then generated according to the nearest label from the adjacent spinal cord segmentations from SCT's Deepseg as seen in Figure 1 d). Lastly, a quality check of the images was performed to inspect for labelling and segmentation errors and images were cropped to the cervical spine (C1-C7) as seen in Figure 1 e). Of the original 267 images in the Spine Generic Dataset, only 149 were utilized due to substantial segmentation and/or labelling errors. Minor additional manual segmentations were performed to clean up the selected segmentations to ensure robust ground truth segmentation masks and bounding boxes. These segmentation corrections were primarily due to inaccurate posterior element segmentations from the PAM50 warping as seen in Figure 1 f). Removal of the first thoracic vertebra was also performed if included with the cropped cervical spine.

3.2 Mask R-CNN-Based Vertebrae Detection and Labelling

A Mask R-CNN model was evaluated for the detection of cervical spine vertebrae. The PyTorch torchvision implementation of the Mask R-CNN with a ResNet50-FPN backbone (pre-trained on the COCO 2017 dataset) was used to identify eight classes (C1 to C7 and background). The model utilized a ADAMs optimizer with a learning rate of $1e^{-4}$, had a batch size of 5, and a confidence threshold of 0.5. The model was trained on a single NVIDIA Titan V GPU with 12GB of VRAM for 100 epochs.

Data augmentations were performed on the input mid-sagittal slice data for training. This includes a random gaussian blur with a kernel size of 3 and σ between 0.1 and 1, random horizontal flip, random rescaling of the image between 128 and 192 voxels while maintaining the aspect ratio, random rotation $\pm 45°$, and padding to a 256x256 voxel input size. The data was randomly split by subject for training, validation, and testing 70%:20%:10%, which equated to 106:29:14 subjects.

4 Results

The model was evaluated on the 14 test subjects unseen by the model during training and validation.The model was successful in predicting the bounding box, class label, score, and mask of the vertebra in the bounding box as shown in Figure 2. Five metrics (Table 1) were utilized to evaluate the model:

1. The multi-task Mask R-CNN loss, which was used for training. (0 best performance)

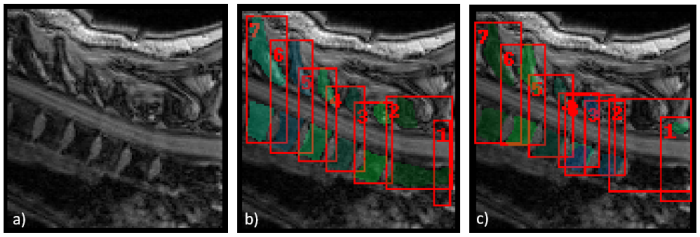

Fig. 2: Example of inference on test set. a) The input cervical spine image for evaluation of the model. b) Ground truth labels, masks, and bounding boxes. c) Predicted labels, masks, and bounding boxes. This test example successfully predicts the bounding box, label, and mask of each vertebra, but also has a false positive detection of C3 on C4 due to vertebrae shape similarities.

2. Mean average precision with a 75% intersection over union (IoU) threshold (mAP@75), a common object detection metric used in state-of-the-art models [4, 5] that considers precision, recall, and IoU with the ground truth bounding boxes. (1 best performance)
3. Localization error (LE) [pixels], the mean absolute error of the predicted and ground truth bounding box centroids. (0 best performance)
4. Mask dice loss, a common segmentation metric. (0 best performance)
5. Identification rate (IDR), a measure of the classification performance of individual vertebra by the model. It is a percentage of vertebrae correctly detected and labelled compared to the total vertebrae in the image. (1 best performance)

Table 1: Model evaluation metric results demonstrate effectiveness of Mask R-CNN for spine vertebrae localization and labelling. Lower is better for all evaluation metrics except mAP@75 and IDR.

Metric	Value
Multi-Task Loss (↓)	0.697
mAP@75 (↑)	0.676
LE [pixels] (↓)	1.573
Dice Loss (↓)	0.111
IDR (↑)	1.0

Figure 2 b) shows the predicted bounding boxes, labels and segmentation masks during testing. The model was capable of identifying and localizing vertebrae within the image demonstrated by the mAP@75 of 0.676. Table 2 shows the AP for each vertebra from the model evaluation. As expected, the C2 and C7 vertebrae were found

Table 2: Average precision of individual vertebral levels used to calculate mAP@75.

	C1	C2	C3	C7	C5	C6	C7
AP	0.446	0.677	0.630	0.654	0.566	0.467	0.669

to have the highest AP of 0.677 and 0.669, respectively, due to the distinct vertebrae features, compared to the other vertebrae which have similar shapes. Specifically, C2 has a elongated vertebral body and rounder posterior element in the mid-sagittal slice, whereas C7 has a similar vertebral body to C3-C6, but the posterior element is elongated. The model's ability to predict C1 and C6 vertebra were lower, with an AP of 0.446 and 0.467, respectively. The poorer performance to detect the C1 vertebra was due to its small size with respect to the other vertebra and difficulty to distinguish within the resolution of the MRI images. The lower AP of C6 may be attributed to the posterior element also being longer than C3-C5, but shorter than C7, leading to an increase in false positive detections as C7. In the example output prediction in Figure 2, the model predicted a false positive C3 on C4. This is due to the close similarities in vertebrae shape or the score threshold used for evaluation. The LE of 1.573 demonstrated good accuracy measuring the voxel-wise error in the centroid locations of the bounding boxes in the image. The segmentation masks were found to be accurate when visually inspected, with a dice loss of 0.111. The largest deviations were visually found to be at the posterior elements, which is due to their complex shapes and sizes depending on the subject anatomy and mid-sagittal slice taken. Lastly, IDR was found to be 1.0, meaning all vertebrae in the input test images were properly detected and labelled by the model.

5 Conclusion and Future Work

This study demonstrates that a Mask R-CNN object detection model is capable of accurately localizing and labelling individual vertebra within the cervical spine in MRI images. The model output bounding boxes and masks can then be used for FoV cropping of images towards improving initial affine alignment prior to deformable registration. Future work will focus on improving model accuracy, expanding its detection capabilities to other regions of the spine, making the model insensitive to the MRI weighting, and generalizing to other novel MRI datasets, including those with pathology. This includes increasing the dataset to include other sections of the spinal column, training with variable fields of view such as C3-C6, the addition of T2w images, and an investigation of a graph or probabilistic models that take advantage of the order of the spine vertebrae to increase labelling performance.

Acknowledgments

We would like to thank NVIDIA for donation of the Titan V GPU through the NVIDIA Academic Hardware Grant Program. Jonathan Chu was supported through the NSERC CREATE training program in Global Biomedical Technology Research and Innovation.

References

[1] E. Bercovich and M. C. Javitt, "Medical Imaging: From Roentgen to the Digital Revolution, and Beyond," *Rambam Maimonides Medical Journal*, vol. 9, no. 4, p. e0034, Oct. 2018. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6186003/

[2] J. G. Jarvik and R. A. Deyo, "Diagnostic Evaluation of Low Back Pain with Emphasis on Imaging," *Annals of Internal Medicine*, vol. 137, no. 7, pp. 586–597, Oct. 2002, publisher: American College of Physicians. [Online]. Available: https://www.acpjournals.org/doi/full/10.7326/0003-4819-137-7-200210010-00010

[3] F. E.-Z. A. El-Gamal, M. Elmogy, and A. Atwan, "Current trends in medical image registration and fusion," *Egyptian Informatics Journal*, vol. 17, no. 1, pp. 99–124, Mar. 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S111086651500047X

[4] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," Jan. 2018, arXiv:1703.06870 [cs]. [Online]. Available: http://arxiv.org/abs/1703.06870

[5] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in *Advances in Neural Information Processing Systems*, vol. 28. Curran Associates, Inc., 2015. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2015/hash/14bfa6bb14875e45bba028a21ed38046-Abstract.html

[6] J. Cohen-Adad, E. Alonso-Ortiz, M. Abramovic, C. Arneitz, N. Atcheson, L. Barlow, R. L. Barry, M. Barth, M. Battiston, C. Büchel, M. Budde, V. Callot, A. J. E. Combes, B. De Leener, M. Descoteaux, P. L. de Sousa, M. Dostál, J. Doyon, A. Dvorak, F. Eippert, K. R. Epperson, K. S. Epperson, P. Freund, J. Finsterbusch, A. Foias, M. Fratini, I. Fukunaga, C. A. M. G. Wheeler-Kingshott, G. Germani, G. Gilbert, F. Giove, C. Gros, F. Grussu, A. Hagiwara, P.-G. Henry, T. Horák, M. Hori, J. Joers, K. Kamiya, H. Karbasforoushan, M. Keřkovský, A. Khatibi, J.-W. Kim, N. Kinany, H. Kitzler, S. Kolind, Y. Kong, P. Kudlička, P. Kuntke, N. D. Kurniawan, S. Kusmia, R. Labounek, M. M. Laganà, C. Laule, C. S. Law, C. Lenglet, T. Leutritz, Y. Liu, S. Llufriu, S. Mackey, E. Martinez-Heras, L. Mattera, I. Nestrasil, K. P. O'Grady, N. Papinutto, D. Papp, D. Pareto, T. B. Parrish, A. Pichiecchio, F. Prados, Rovira, M. J. Ruitenberg, R. S. Samson, G. Savini, M. Seif, A. C. Seifert, A. K. Smith, S. A. Smith, Z. A. Smith, E. Solana, Y. Suzuki, G. Tackley, A. Tinnermann, J. Valošek, D. Van De Ville, M. C. Yiannakas, K. A. Weber, N. Weiskopf, R. G. Wise, P. O. Wyss, and J. Xu, "Generic acquisition protocol for quantitative MRI of the spinal cord," *Nature Protocols*, vol. 16, no. 10, pp. 4611–4632, Oct. 2021. [Online]. Available: https://www.nature.com/articles/s41596-021-00588-0

[7] B. De Leener, S. Lévy, S. M. Dupont, V. S. Fonov, N. Stikov, D. Louis Collins, V. Callot, and J. Cohen-Adad, "SCT: Spinal Cord Toolbox, an open-source software for processing spinal cord MRI data," *NeuroImage*, vol. 145, pp. 24–43, Jan. 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1053811916305560

[8] C. Gros, B. De Leener, A. Badji, J. Maranzano, D. Eden, S. M. Dupont, J. Talbott, R. Zhuoquiong, Y. Liu, T. Granberg, R. Ouellette, Y. Tachibana, M. Hori, K. Kamiya, L. Chougar, L. Stawiarz, J. Hillert, E. Bannier, A. Kerbrat, G. Edan, P. Labauge, V. Callot, J. Pelletier, B. Audoin, H. Rasoanandrianina, J.-C. Brisset, P. Valsasina, M. A. Rocca, M. Filippi, R. Bakshi, S. Tauhid, F. Prados, M. Yiannakas, H. Kearney, O. Ciccarelli, S. Smith, C. A. Treaba, C. Mainero, J. Lefeuvre, D. S. Reich, G. Nair, V. Auclair, D. G. McLaren, A. R. Martin, M. G. Fehlings, S. Vahdat, A. Khatibi, J. Doyon, T. Shepherd, E. Charlson, S. Narayanan, and J. Cohen-Adad, "Automatic segmentation of the spinal cord and intramedullary multiple sclerosis lesions with convolutional neural networks," *NeuroImage*, vol. 184, pp. 901–915, Jan. 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1053811918319578

[9] B. De Leener, V. S. Fonov, D. L. Collins, V. Callot, N. Stikov, and J. Cohen-Adad, "PAM50: Unbiased multimodal template of the brainstem and spinal cord aligned with the ICBM152 space," *NeuroImage*, vol. 165, pp. 170–179, Jan. 2018. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1053811917308686

[10] A. Fedorov, R. Beichel, J. Kalpathy-Cramer, J. Finet, J.-C. Fillion-Robin, S. Pujol, C. Bauer, D. Jennings, F. Fennessy, M. Sonka, J. Buatti, S. Aylward, J. V. Miller, S. Pieper, and R. Kikinis, "3D Slicer as an image computing platform for the Quantitative Imaging Network," *Magnetic Resonance Imaging*, vol. 30, no. 9, pp. 1323–1341, Nov. 2012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0730725X12001816