Spinal Stenosis Detection in MRI using Modular Coordinate Convolutional Attention Networks

Uddeshya Upadhyay
Dept. of Computer Science and Engineering
Indian Institute of Technology-Bombay

Badrinath Singhal
Synapsica Technologies

Meenakshi Singh
Synapsica Technologies

***Abstract*—Spinal stenosis is a condition in which a portion of spinal canal narrows and exerts pressure on nerves that travel through it causing pain and numbness that might require surgery. This narrowing can be caused by pathologies in bony structures (vertebrae) or soft tissue structures (intervertebral discs) that comprise the spine. Radiography, particularly Magnetic Resonance Imaging (MRI) is the modality of choice to evaluate stenosis and intervertebral disc pathology. Radiologists examine axial MRI scans at various levels along the spine to detect stenosis. Further, they evaluate the diameters of spinal canal and bulging in nearby discs which can indicate narrowing and compression on nerves. Hence measuring various diameters in a scan is a crucial step in diagnosis. However, affected regions occupy a very small fraction of the scan and there is virtually no room for error as a deviation of few pixels will also lead to discrepancies in measured and original lengths which makes it a very difficult and laborious task to measure the length of such intricate structures accurately. This paper proposes a novel deep learning based solution to tackle this problem. Proposed method attempts to solve it in two independent modules and makes prediction on the enlarged section of the scan which also makes it easier to measure various lengths. Human radiologists focus on certain parts of the scan rather than attending to the entire scan which largely consists of irrelevant background. Proposed modular approach is designed to mimic this attention mechanism. Both modules are built using coordinate convolutional networks, comparisons with baseline method empirically demonstrate superiority of the proposed approach.**

***Index Terms*—spinal stenosis, magnetic resonance imaging, deep learning, coordinate convolutions, attention**

I. Introduction

Spinal stenosis is the most common reason for lumbar spine surgery in patients over 65 years [1]. Given that many patients are asymptomatic, radiologists rely primarily on *Magnetic Resonance Imaging (MRI)* to provide objective evidence of neurovascular compromise. However, there is recognized variability in description and reporting of spinal stenosis among radiologists and other physicians making any analysis of surgical outcomes less insightful [2]. This variability inspires the need for quantitative methods in forming diagnosis. The most common quantitative method used for this purpose is measurement of anteroposterior (AP) diameter of the spinal canal [2]. This measure is not affected by ageing [3] once spine reaches adult size and hence is generically applicable. Stenosis is indicated by an AP diameter of canal less than 10 mm in cervical spine or 12 mm in lumbar spine.

MRI scans consist of images taken at various sections along 3 planes - Coronal (from front), Sagittal (from side) and Axial (from top down). They create a 3D impression of the region by stacking 2D images along different planes. In MR images AP diameter of canal is typically calculated on axial images corresponding to mid-disc levels along the spine. Similarly, *foraminal* stenosis is established by observing narrowing of intervertebral *foramen*, a small hole through which nerves exit the spinal canal and travel through the body. Diagnosis takes into account characterization and measurement of AP diameter of spinal canal and foramen over different discs and vertebrae positions.

Measuring canal diameters on MR images is a tedious task, requiring radiologists to accurately place markers provided by DICOM viewer at appropriate locations to get measurement. Multiple measurements corresponding to discs present in the scan are made for a typical report. This paper proposes a method to automate the process of measuring the diameter of spinal canal in MRI scans using *Convolutional Neural Network* architectures. Proposed method divides the problem statement into two parts, first part is building attention network which decides which region to look in the image and the second part works as the regression model which calculates the diameter of the spinal canal.

Existing work [4] builds similar automation by providing stenosis grading for lumbar spine in 3 classes - Normal, Mild/Moderate, and Severe; and reports class accuracy of approximately 80.4. However, medically the general standard is to provide actual measurements in radiology reports, and practitioners have differing opinions about range of measurements that form a particular stenosis grade. Our approach not only provides measurements of AP diameter of canal but also is generalized for both lumbar and cervical regions of spine thus giving more freedom to radiologist for diagnosis. Our best performing model predicts the diameter of the spinal canal accurately within 2mm for 85.5% of the test dataset and diameter of foraminal gaps within 2mm for 77.5% of the test dataset.

II. Related Work

A. Medical Image Analysis with deep learning

Deep learning methods are increasingly used to analyse medical images and related data. Typically deep convolutional networks are used for tasks such as classification, segmentation, regression [5]–[9] to achieve some state of the art performances. Recent works have shown successful

applications of deep learning methods to biomedical image analysis including counting [10], [11], segmentation [12], [13], super-resolution [14], [15]. One such landmark method demonstrating successful application of deep learning over a variety of biomedical image analysis task with relatively small amount of tagged data was proposed as *U-Net* [12]. The network has been instrumental in tasks such as [16]–[19]. *U-Net* was built upon *Fully Convolutional Networks (FCN)* [20] which utilizes only convolutional nets allowing for processing of arbitrary sized input, it was proposed to perform segmentation. Deep neural networks have also been used to learn important features and segment out tumor in various scans [21], [22]. Recently, deep neural networks have also been employed to reduce required gadolinium dose in contrast-enhanced brain MRI [23]. Cardiac MRI analysis has also benefited by deep learning [24], [25]. Deep learning has also been used to analyse CT scans and ultrasound images [26]–[29].

In context of spinal MRI analysis there has been significant work before rise of deep learning which involved the use of classical hand crafted features and various classifiers like *Support Vector Machines (SVM)* [30]–[32]. Most of the prior work has focused on segmentation in MRI scans [33]–[35]. Recently proposed [4] performs the segmentation of the discs which allows them to process the axial and sagittal scans of a disc on various levels simultaneously using a pair of deep neural network to *grade* the cross sectional region of discs on the basis of seriousness of stenosis.

B. Visual Attention

Retina in human visual system intercepts light coming from a wide range of directions covering a vast area yet humans implicitly pay attention to a very specific region in the field of view, i.e, at any given instant specific region is viewed in high-resolution where as the rest of the field of view remains in low-resolution. As we shift the focus of our eyes, we shift the high-resolution region in our field of view. This allows for efficient computation by visual cortex system as it only processes information at a few fixation region, rather than processing the entire scene.

Some of the methods proposed to capture this mechanism in deep neural networks include *recurrent attention model (RAM)* [36] which formulates the vision problem as a reinforcement learning problem. Other methods such as *Deep recurrent attention model (DRAM)* [37] was designed to overcome the limitations of RAM by introducing additional components like *context network*, *classification network* and *emission network*. *Enriched deep recurrent attention model (EDRAM)* [38] was proposed as an attempt to eliminate reinforcement learning as used in RAM and DRAM. EDRAM augments the attention network with spatial transformer [39] and uses a differentiable loss function.

Proposed method takes inspiration from this concept to design a module capable of extracting "important region" from the MRI scan and further process it to predict the lengths of various structures. However, this method is not using recurrent networks or reinforcement learning but only convolutional blocks which have shown remarkable feature extraction capability from complex images.

C. Coordinate Convolutions

Recent advancements in convolutional networks have shown that on the simple task of learning *coordinate transforms*, i.e., transforms from Cartesian coordinates (*x,y*) to one-hot vector or vice versa, convolutional neural networks fail [40]. However, in the same work it has been shown that altering the input to the convolutional layers by including additional coordinate related channels help rectify the problem.

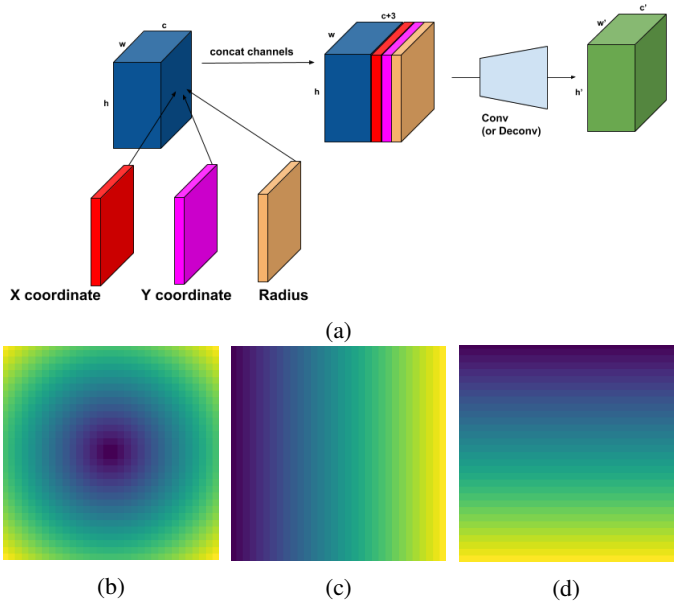


Fig. 1: **(a)** Coordinate convolutional layer taking c channel input and appending 3 different coordinate related channels and producing output with c' channels. **(b)** Radius channel. **(c)** X channel. **(d)** Y channel

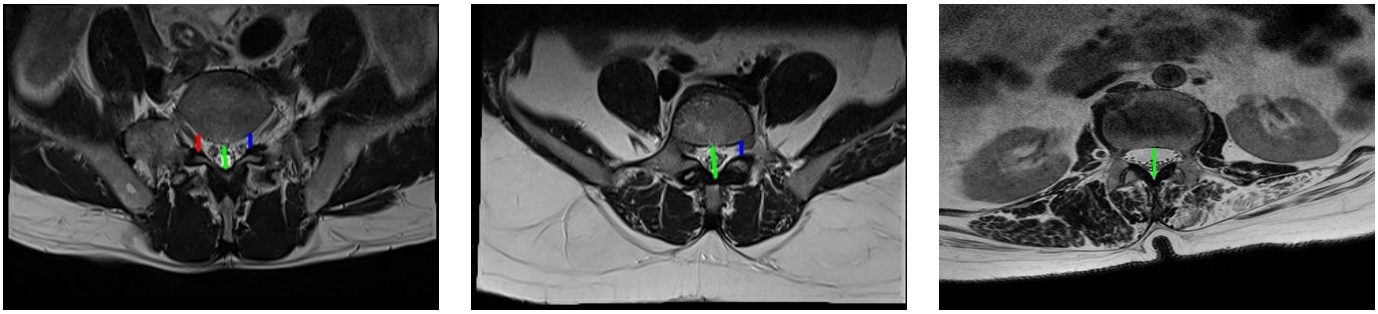
Various experiments have shown significant improvement in vision tasks such as detection, segmentation etc by including *coordinate convolutions* in the pipeline appropriately. This method proposes to generate binary mask to highlight each of the specific regions in the scan whose length is to be measured, i.e, spinal canal and foramina. The length of these structures can be inferred from the predicted masks by measuring it's length in a particular direction.

This work produces binary masks to calculate the length for several regions and *coordConv* layers are better suited for this task. Therefore, we design the *coordConv* version of baseline model and show that there is significant improvement in the performance of the network by comparing appropriate results.

III. Methods

A. Dataset

The dataset consists of spinal MRI scans from multiple sources with different MRI machines. In this study a total



(a) (b) (c)

Fig. 2: Different samples from training dataset, representing possible input. **(a)** shows the case where the diameter of left foramen, right foramen and spinal canal are clearly visible and labelled. **(b)** shows the case where the left foramen has fused. **(c)** shows the case where both the foramina have fused

of 944 unique T2-axial MRI scans are used, out of which 653 have been used for training and remaining 291 have been used for testing. There are different MRI machine models from different manufacturers, specifically, scans come from *Siemens* with models *C!, Magntom-essenza, Sempra, Skyra*; *Phillips Medical Systems* with models *Achieva, Multiva*; *GE Medical Systems* with models *Signa-creator, Signa-excite, Signa-explorer*.

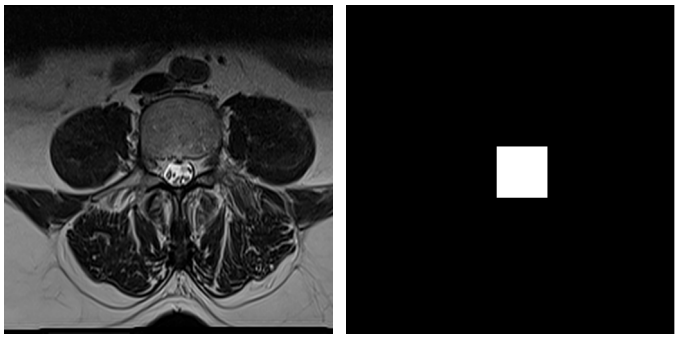
Train set consists of 554 scans at magnetic field strength of *1.5T* and 99 scans at *3T*. 372 scans are of female patients and 281 scans are of male patients. 83% of scans are from *Siemens* machines, 15.2% of scans are from *Phillips Medical Systems* and remaining from *GE Medical Systems*.

Test set consists of 257 scans at magnetic field strength of *1.5T* and 34 scans at *3T*. 168 scans are of female patients and 123 scans are of male patients. 80.8% of scans are from *Siemens* machines, 17.2% of scans are from *Phillips Medical Systems* and remaining from *GE Medical Systems*.

These MRI scans were annotated by a group of professional radiologists who marked out the diameter for spinal canal and the narrowest gap between the intervertebral foramen on either side of spinal canal (if foramen was visible). Spinal canal is always visible in every valid scan; however, the narrowing in foramen may not be visible in every scan. Figure 2 shows a few samples from the dataset and various types of annotation which are possible.

B. Architecture

Proposed approach to this problem can be split into two separate models. First model works as a explicit attention network and we refer to this part of the pipeline as *module 1*, this model is used to segment out the "important" region in the scan which encloses spinal canal and foramina (if present). We use this output to create a square crop-section. This allows us to focus the downstream computations on "important" region and avoid possibly noisy signals from the rest of the scan to affect the learning. This is possible because using the available annotations it is easy to create required dataset for the training, by creating a sufficiently large square bounding box. Fig 3 shows a training sample with the prepared ground truth (obtained via bounding box enclosing all the lines with a small amount of padding vertically and horizontally).



(a) (b)

Fig. 3: **(a-b)** example of an input and binary-mask pair used for training *module 1*

This paper proposes a new deep neural network, *CoU-net* as shown in figure 4, based on *U-Net* [12] and *coordinate convolution* [40]. *CoU-Net* makes use of coordinate convolutions by appending the *x, y, radius* coordinate channels as shown in figure 1, at the beginning and after every max-pooling or up-scaling in the network. The last convolutional layer in *module 1* outputs a 1 channel binary masks for each of the input scan representing the "important" region in the scan.

Model trained in *module 1* learns to output a mask representing important region, this mask is used to generate a square bounding box to crop the "important" region from the scan. Interpolation of the smaller cropped output to a larger square image is done before passing it downstream for further computations.

The interpolated output from *module 1* is passed to a second network to localize various regions, which will eventually lead to the required measurement as explained later. This part of pipeline is referred as *module 2*. In this module, we pose the learning task as *multi-channel segmentation*. Ground truth annotations which are available in form of lines as shown in figure 2, are used to create tight bounding box around each line. Padding is done in horizontal direction to give

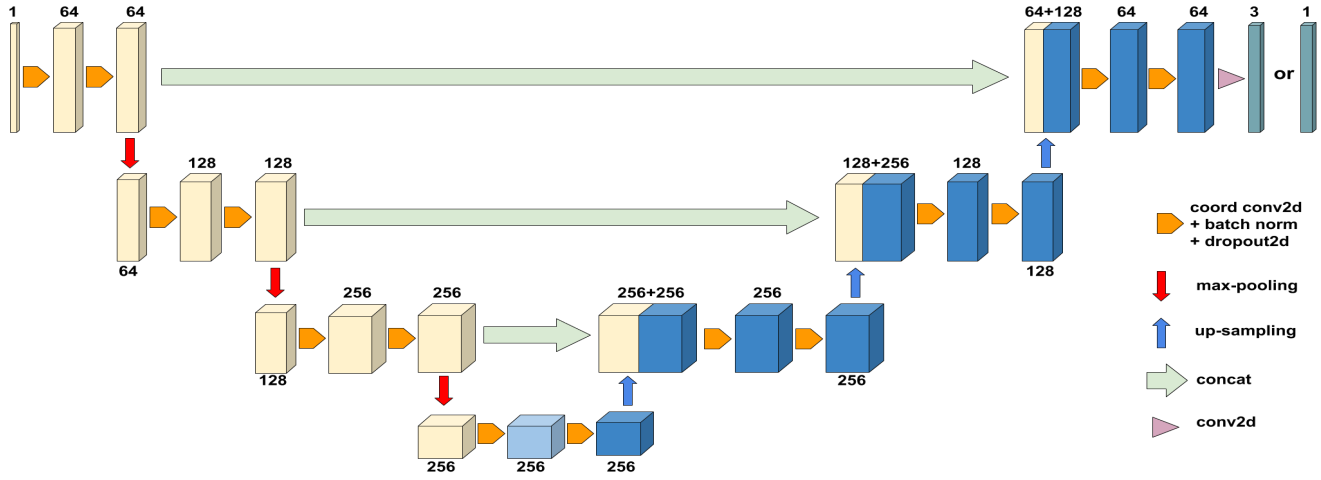
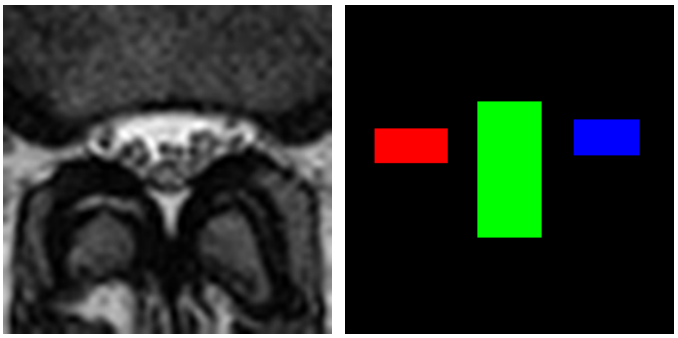


Fig. 4: *CoU-Net*, a modification inspired from *U-Net* and *coordinate convolution*.

each of the boxes significant width. No padding is done in the vertical direction as the length in the vertical direction is crucial for accurate calculation (length of most of the markings can be estimated as the difference in coordinates in the vertical direction). Multi-channel (in this case 3, because there can be a maximum of three boxes) segmentation masks are created, such that each channel corresponds to the binary mask of only one annotated line. Channel is zeroed out if a line is missing from annotation. Figure 5 shows an example of the training sample and corresponding *multi-channel* masks created for training purpose.



(a) (b)

Fig. 5: **(a-b)** example of an input crop-section and multi-channel mask (represented in RGB) pair used for training *module 2*

Since the final reported length depends directly on the prediction of model in *module 2*, it is essential for this model to output the maps with high precision. Again *CoU-Net* is used to predict the output precisely. However, the last convolutional layer in *module 2* outputs a 3 channel tensor for every input scan. These 3 channels represents the binary mask for left foramen, spinal canal and right-foramen. We compare the performance of *CoU-Net* with a similar *U-Net* (without any coordinate convolution) and demonstrate the superiority of proposed method.

C. Loss Functions

Let $M_i(.;\Theta_{M_i})$ represent deep neural network for *module* i with trainable parameters Θ_i. Let X_k represent input to the model M_i which produces output $\hat{Y_k}$, i.e

$$\hat{Y_k} = M_i(X_k, \Theta_{M_i}) \tag{1}$$

Let Y_k be the ground truth corresponding to X_k. The following loss function is used for both *module 1* and *module 2*

$$L(\hat{Y_k}, Y_k) = \frac{||\hat{Y_k} - Y_k||^2}{d} - \lambda \frac{2((\hat{Y_k} \cdot Y_k) + \epsilon)}{(\sum_j \hat{Y}_{jk} + \sum_j Y_{jk}) + \epsilon} \tag{2}$$

where flattened Y_k is a d dimensional vector, Y_{jk} is j^{th} coordinate in Y_k and $\epsilon = 1e^{-5}$ is a small constant introduced for numerical stability. The first term represents the *Mean Square Error* (MSE) between the ground-truth and the predicted output, whereas the second term represents the dice score between prediction and the ground-truth. We set $\lambda = 1e^{-4}$ a small positive number (obtained empirically). At the start of the training loss is dominated by the MSE term, as training progresses this term reduces in magnitude and in later epochs of training, loss has significant contribution from the dice score as well. In these experiments, dice score alone took longer to converge where as the MSE did not generalize well for the pixels at the boundary. We found this combination to provide much faster convergence and improved performance.

D. Training strategy

Training dataset is augmented by applying five different level of contrasts as well as flipping the images left to right.

Both *module 1* and *module 2* are trained independently and use images and corresponding ground-truth masks of dimensions 256×256 for training. Required length l (in millimeters) is calculated using

$$l = s_1 \cdot s_2 \cdot p_s \cdot h \tag{3}$$

where s_i refers to the scaling factor introduced due to cropping and re-scaling in *module i*, p_s is the pixel-spacing specified in

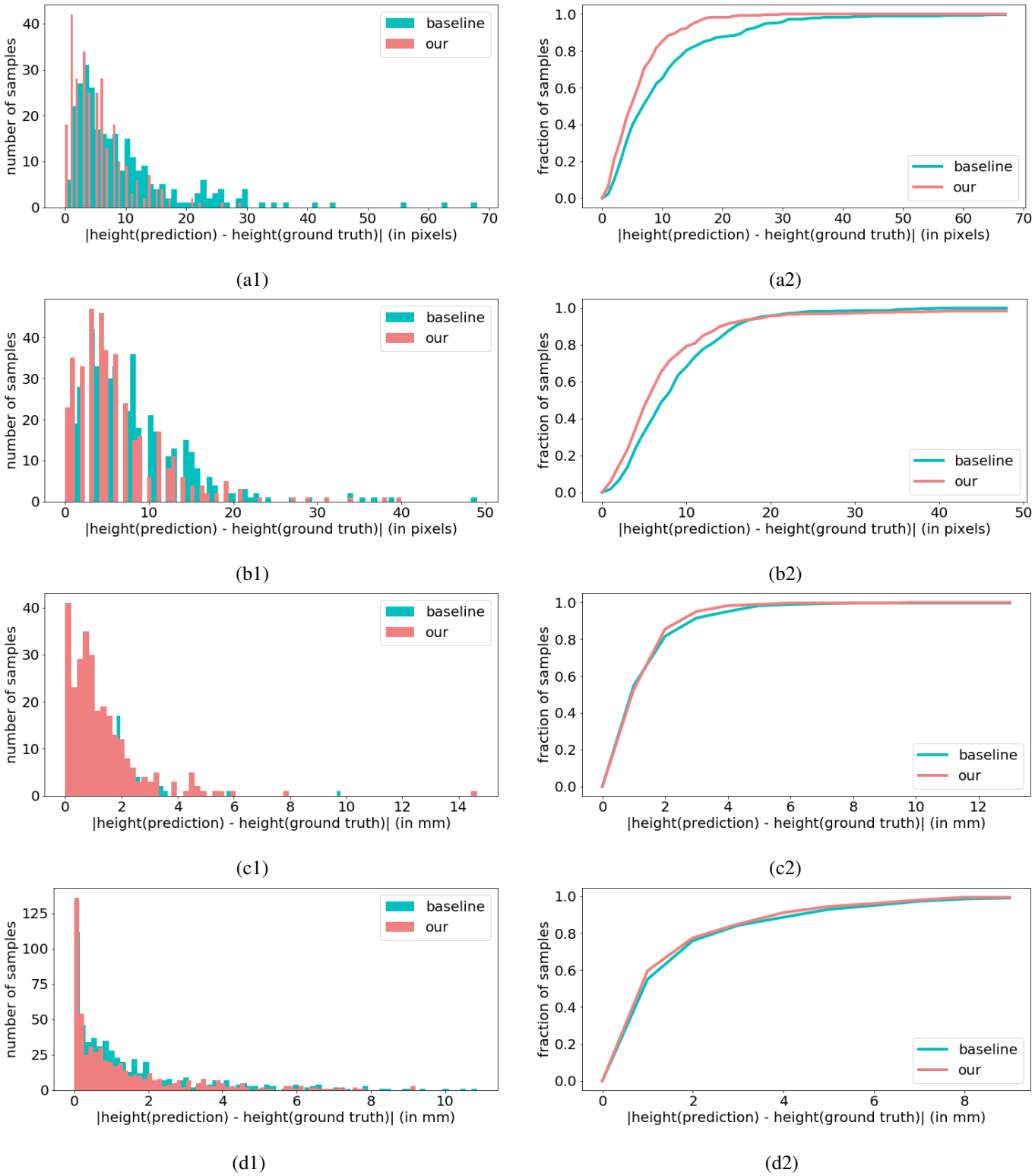


Fig. 6: **(a1-a2)** Histogram and cumulative distribution function respectively, for deviation in measurement (in pixels) of spinal canal. **(b1-b2)** Histogram and cumulative distribution function respectively, for deviation in measurement (in pixels) of foramina. **(c1-c2)** Histogram and cumulative distribution function respectively, for deviation in measurement (in mm) of spinal canal. **(d1-d2)** Histogram and cumulative distribution function respectively, for deviation in measurement (in mm) of foramina.

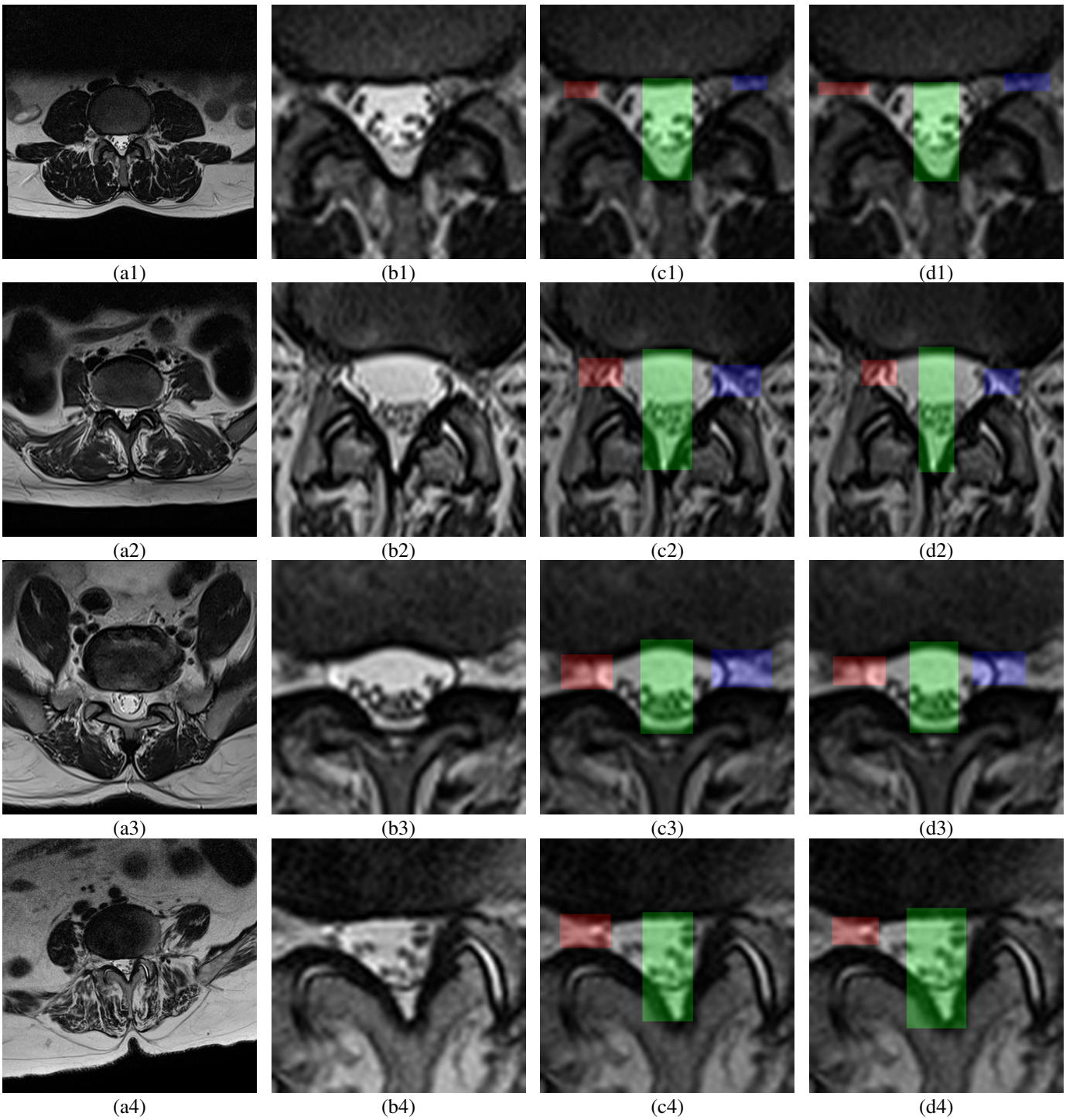


Fig. 7: Some results of our model. **(a1-a4)** input scan to *module 1*. **(b1-b4)** square cropped section using output of *module 1*. **(c1-c4)** predictions from *module 2* overlaid on cropped section for visualization. **(d1-d4)** ground-truth labels overlaid on cropped section for visualization

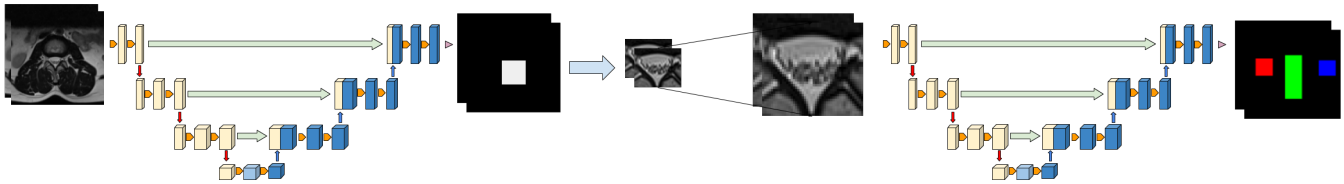


Fig. 8: Schematic diagram of our pipeline consisting of two different modules *module 1* and *module 2*

the DICOM file and h is the final measurement inferred by our *module 2*. We center crop the original input scan to create a square scan with side l_s, where l_s is the shorter side of the scan and re-scale it to 256×256 ($s_1 = \frac{l_s}{256}$), this ensures that the aspect ratio of the scan does not changes through out the pipeline. Output of *module 1* is used to crop a square bounding box of size $m \times m$ in the original scan, this crop section is re-scaled to 256×256 ($s_2 = \frac{m}{256}$), which is used as input to the network in *module 2*.

To train the network in *module 1* batch size of 10, with Adam optimizer [41] is used. Initial learning rate is set to $1e^{-4}$ and cosine annealing is used to decay the learning rate over epochs, to avoid over-fitting we use L2 regularization on the weights of the neural network with weight decay parameter set to $1e^{-3}$. A similar configuration for model in *module 2* is used for training. Figure 8 shows the schematic for our entire pipeline.

IV. Results and Conclusions

To measure the performance of proposed networks, various metrics including *IoU*, *dice score*, *precision*, *recall* between the predicted and ground-truth masks are compared. The required length of various structures can be approximated by the length in vertical direction because the orientation of spinal canal ans foraminal gaps is *almost vertical*. To quantify the performance for this metric, difference in the height of predicted mask and the ground-truth label is also measured. Table I shows various performance metric of two approaches, the baseline model and the coordinate convolution version that is proposed. Results show that incorporating coordinated convolutions improved the performance of the network.

	baseline	our
Accuracy	95.5	**96.31**
Dice score	78.81	**83.00**
IoU	66.15	**71.61**
Recall Score	78.41	**81.2**
Precision Score	81.36	**86.6**

TABLE I: Comparing performances of baseline and our model

Figure 6-a1 and figure 6-b1 shows the distribution of the predictions from *module 2* on test dataset. The horizontal axis of histogram represents the bins of deviation in the length of spinal canal or foraminal gaps inferred from prediction and the ground-truth masks (in pixels) respectively, the vertical axis shows the number of scans in test dataset which fall into a particular bin. Figure 6-c1 and figure 6-d1 shows the similar distribution but the deviation is measured in *millimeters* (on horizontal axis) for spinal canal and foraminal gaps respectively.

Introducing coordinate convolutional blocks in *CoU-Net* reduces the deviation between the predicted length and the ground truth, this is evident from figure 6-(a2, b2, c2, d2) which represents the cumulative distribution function which shows that a greater fraction of scans have smaller amount of deviation. Cumulative distribution in 6-(c2, d2) shows that our model can predict the length of the spinal canal and foraminal gaps accurately within 2mm for 85.5% and 77.5% of the test dataset respectively.

Narrowing of the foramina is much harder to detect and measure due to small opening compared to typical diameter of the spinal canal, this is evident from figure 6-(b1,b2) which shows the similar histogram and cumulative distribution function of deviation in heights, the plot shows the distribution which combines both left and right foraminal gaps. The fraction of scans which have greater deviation is much larger compared to the spinal canal deviation, yet the improvement due to *CoU-Net* over baseline is evident. It was also observed that for this task, under identical hyper-parameters setting coordinate convolution version of the model converges faster than the corresponding baseline. Figure 7 shows some results on test dataset obtained by *CoU-Net* along with the ground-truth, different masks obtained can be used to calculate the length of the required region accurately.

Acknowledgment

We are grateful to Dr. Kumar Rahul and other radiologists for their guidance and help with tagging our datasets. We thank all radiology centers who readily agreed to share their valuable data with us. We thank the entire team of *Synapsica Technologies* for their continuous support throughout the project.

References

[1] E. S. M. Kiran S. Talekar MD, Mougnyan Cox MD and A. E. F. MD, "Imaging spinal stenosis," 2017.

[2] M. Brant-Zawadzki, M. C Jensen, N. Obuchowski, J. S Ross, and M. Modic, "Interobserver and intraobserver variability in interpretation of lumbar disc abnormalities: A comparison of two nomenclatures," *Spine*, vol. 20, pp. 1257–63; discussion 1264, 07 1995.

[3] K. C. K. C. Kim, Park, "Changes in spinal canal diameter and vertebral body height with age," vol. 54. Yonsei University College of Medicine, 2013.

[4] J.-T. Lu, S. Pedemonte, B. Bizzo, S. Doyle, K. P. Andriole, M. H. Michalski, R. G. Gonzalez, and S. R. Pomerantz, "Deep spine: Automated lumbar vertebral segmentation, disc-level designation, and spinal stenosis grading using deep learning," in *Machine Learning for Healthcare Conference*, 2018, pp. 403–419.

[5] R. LaLonde, D. Zhang, and M. Shah, "Clusternet: Detecting small objects in large scenes by exploiting spatio-temporal information."

[6] H. Hu, J. Gu, Z. Zhang, J. Dai, and Y. Wei, "Relation networks for object detection."

[7] L. W. X. B. Z. L. Zhang, Shifeng and S. Z. Li., "Single-shot refinement neural network for object detection," in *Computer Vision and Pattern Recognition (CVPR), IEEE*, 2018.

[8] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "Pointnet: Deep learning on point sets for 3d classification and segmentation," *Proc. Computer Vision and Pattern Recognition (CVPR), IEEE*, vol. 1, no. 2, p. 4, 2017.

[9] S.-W. Kim, H.-K. Kook, J.-Y. Sun, M.-C. Kang, and S.-J. Ko, "Parallel feature pyramid network for object detection," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 234–250.

[10] W. Xie, J. A. Noble, and A. Zisserman, "Microscopy cell counting and detection with fully convolutional regression networks," *Computer methods in biomechanics and biomedical engineering: Imaging & Visualization*, vol. 6, no. 3, pp. 283–292, 2018.

[11] T. Chen and C. ChefdHotel, "Deep learning based automatic immune cell detection for immunohistochemistry images," in *International Workshop on Machine Learning in Medical Imaging*. Springer, 2014, pp. 17–24.

[12] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[13] D. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Deep neural networks segment neuronal membranes in electron microscopy images," in *Advances in neural information processing systems*, 2012, pp. 2843–2851.

[14] J. Kim, J. Kwon Lee, and K. Mu Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 1646–1654.

[15] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1874–1883.

[16] T. Falk, D. Mai, R. Bensch, Ö. Çiçek, A. Abdulkadir, Y. Marrakchi, A. Böhm, J. Deubner, Z. Jäckel, K. Seiwald *et al.*, "U-net: deep learning for cell counting, detection, and morphometry," *Nature methods*, p. 1, 2018.

[17] P. Esser, E. Sutter, and B. Ommer, "A variational u-net for conditional appearance and shape generation," in *CVPR*, 2018.

[18] K. Sirinukunwattana, J. P. Pluim, H. Chen, X. Qi, P.-A. Heng, Y. B. Guo, L. Y. Wang, B. J. Matuszewski, E. Bruni, U. Sanchez, A. Bhm, O. Ronneberger, B. B. Cheikh, D. Racoceanu, P. Kainz, M. Pfeiffer, M. Urschler, D. R. Snead, and N. M. Rajpoot, "Gland segmentation in colon histology images: The glas challenge contest," *Medical Image Analysis*, vol. 35, pp. 489 – 502, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1361841516301542

[19] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3d u-net: learning dense volumetric segmentation from sparse annotation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2016, pp. 424–432.

[20] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.

[21] V. A. Srgio Pereira, Adriano Pinto and C. A. Silva, "Brain tumor segmentation using convolutional neural networks in mri images," in *IEEE Transactions on Medical Imaging*. IEEE, 04 March 2016, pp. 1240 – 1251.

[22] M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P.-M. Jodoin, and H. Larochelle, "Brain tumor segmentation with deep neural networks," *Medical image analysis*, vol. 35, pp. 18–31, 2017.

[23] W. M. Gong E, Pauly JM and Z. G, "Deep learning enables reduced gadolinium dose for contrast-enhanced brain mri." in *J Magn Reson Imaging 2018*, 2018, p. 330340.

[24] W. Bai, M. Sinclair, G. Tarroni, O. Oktay, M. Rajchl, G. Vaillant, A. M. Lee, N. Aung, E. Lukaschuk, M. M. Sanghvi, F. Zemrak, K. Fung, J. M. Paiva, V. Carapella, Y. J. Kim, H. Suzuki, B. Kainz, P. M. Matthews, S. E. Petersen, S. K. Piechnik, S. Neubauer, B. Glocker, and D. Rueckert, "Automated cardiovascular magnetic resonance image analysis with fully convolutional networks," *Journal of Cardiovascular Magnetic Resonance*, vol. 20, no. 1, p. 65, Sep 2018. [Online]. Available: https://doi.org/10.1186/s12968-018-0471-x

[25] O. Bernard, A. Lalande, C. Zotti, F. Cervenansky, X. Yang, P.-A. Heng, I. Cetin, K. Lekadir, O. Camara, M. A. G. Ballester *et al.*, "Deep learning techniques for automatic mri cardiac multi-structures segmentation and diagnosis: Is the problem solved?" *IEEE Transactions on Medical Imaging*, 2018.

[26] P. Lobo and S. Guruprasad, "Classification and segmentation techniques for detection of lung cancer from ct images," in *2018 International Conference on Inventive Research in Computing Applications (ICIRCA)*. IEEE, 2018, pp. 1014–1019.

[27] A. M. Rossetto and W. Zhou, "Deep learning for categorization of lung cancer ct images," in *Proceedings of the Second IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies*. IEEE Press, 2017, pp. 272–273.

[28] H. Sugimori, "Classification of computed tomography images in different slice positions using deep learning," *Journal of healthcare engineering*, vol. 2018, 2018.

[29] L. J. Brattain, B. A. Telfer, M. Dhyani, J. R. Grajo, and A. E. Samir, "Machine learning for medical ultrasound: status, methods, and future opportunities," *Abdominal Radiology*, vol. 43, no. 4, pp. 786–799, 2018.

[30] S. Ghosh, M. R. Malgireddy, V. Chaudhary, and G. Dhillon, "A new approach to automatic disc localization in clinical lumbar mri: combining machine learning with heuristics," in *Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on*. IEEE, 2012, pp. 114–117.

[31] Z. Peng, J. Zhong, W. Wee, and J.-h. Lee, "Automated vertebra detection and segmentation from the whole spine mr images," in *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*. IEEE, 2006, pp. 2527–2530.

[32] M. Lootus, T. Kadir, and A. Zisserman, "Vertebrae detection and labelling in lumbar mr images," in *Computational methods and clinical applications for spine imaging*. Springer, 2014, pp. 219–230.

[33] D. Gaweł, P. Główka, T. Kotwicki, and M. Nowak, "Automatic spine tissue segmentation from mri data based on cascade of boosted classifiers and active appearance model," *BioMed research international*, vol. 2018, 2018.

[34] C. Ling, W. M. Diyana, W. Zaki, A. Hussain, and H. A. Hamid, "Semi-automated vertebral segmentation of human spine in mri images," in *Advances in Electrical, Electronic and Systems Engineering (ICAEES), International Conference on*. IEEE, 2016, pp. 120–124.

[35] P. D. Barbieri, G. V. Pedrosa, A. J. M. Traina, M. H. Nogueira-Barbosa *et al.*, "Vertebral body segmentation of spine mr images using superpixels," in *International Symposium on Computer-Based Medical Systems, 28th*. Institute of Electrical and Electronics Engineers–IEEE, 2015.

[36] V. Mnih, N. Heess, A. Graves *et al.*, "Recurrent models of visual attention," in *Advances in neural information processing systems*, 2014, pp. 2204–2212.

[37] J. L. Ba, V. Mnih, and K. Kavukcuoglu, "Multiple object recognition with visual attention," *context*, vol. 2, no. l3, p. l4.

[38] A. Ablavatski, S. Lu, and J. Cai, "Enriched deep recurrent visual attention model for multiple object recognition," in *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2017, pp. 971–978.

[39] M. Jaderberg, K. Simonyan, A. Zisserman *et al.*, "Spatial transformer networks," in *Advances in neural information processing systems*, 2015, pp. 2017–2025.

[40] R. Liu, J. Lehman, P. Molino, F. P. Such, E. Frank, A. Sergeev, and J. Yosinski, "An intriguing failing of convolutional neural networks and the coordconv solution," in *Advances in Neural Information Processing Systems*, 2018, pp. 9628–9639.

[41] P. Kingma and J. Ba., "Adam: A method for stochastic optimization," in *In International Conference in Learning Representations*, 2015.