Segmentation of a Novel Spine Dataset

Mrigank Raman
Department of Mathematics
IIT Delhi
New Delhi, Delhi 110016
mt1170736@iitd.ac.in

Nikhil Kapoor
Department of Mathematics
IIT Delhi
New Delhi, Delhi 110016
mt1170739@iitd.ac.in

Abstract

In this report we consider the problem of Biomedical Image Segmentation. We consider a novel Spine dataset containing X-Ray images of the spine taken from two different views. We suggest the use of two different segmentation architectures for this problem namely LinkNet and Mask-RCNN. We also classify the spine images as Damaged or Normal.

1 Introduction

The Spinal cord injured patients who have suffered injuries due to some kind of trauma need urgent attention from a medical expert. They must be congregated early in the spinal units to receive better facilities and adequate attention, ultimately improving the outcome of the treatment and rehabilitation. Research indicates that early surgery and comprehensive rehabilitation markedly reduces the overall morbidity of spinal cord injured patients by enabling the patient to lead an independent life [1-3]. Moreover, a higher incidence of neurological deficit if the diagnosis of thoracolumbar spine fracture is delayed[4]. Therefore, to advance our studies in enabling accurate and quick identification of a person's state of damage of the vertebral column, we implement state of the art deep learning techniques for this task.

Convolutional Neural Networks have revolutionized the field of computer vision including tasks such as semantic segmentation and classification [5-6]. These deep learning architectures are extensively being used in biomedical tasks such as predicting whether spine is damaged or normal. The problem of semantic segmentation has been extensively studied over the past few years [8-9]. While convolutional networks have already existed for a long time, their success was limited due to the size of the available training sets and the size of the considered networks [7]. So benchmarking novel datasets like the spine dataset helps us to evaluate the existing models on more data.

2 Dataset Information

The Spine Project Dataset consists of radiological images, X rays, from 978 patients who visited the Indian Spinal Injuries Centre seeking for their treatment. For each patient, we have two X-Ray images highlighting two different views, namely AP and Lateral which gives us a total of 1956 X-Ray images. The labels for our dataset are 'Normal' and 'Damaged' reflecting upon the normal or diseased state of the vertebral column.

The segmentation classes of our dataset are specific to each view. For the AP view, we have 3 classes namely, Vertebra, Spinous Process, and Pedicle. Similarly, for the lateral view, we have 5 classes that include Vertebra, Spinous Process, Disk Height, Anterior Vertebral Line, and Posterior Vertebral Line.

Submitted to 33rd Conference on Neural Information Processing Systems (NeurIPS 2019). Do not distribute.

(a) AP view

(b) LAT view

Figure 1: Two Different Views of the Spine

3 Segmentation Architectures

In this section we discuss about the architectures used by us for the segmentation task. We have tried the Mask-RCNN [10] and a variant of U-Net known as LinkNet [9]. In our Assignment we have submitted results of the LinkNet Architecture as we have got better results using it. Mask RCNN tackles the problem of instance segmentation and we show that we have got comparable results in some of the tasks using Mask RCNN. For this task we found out that the ground truth masks had some values greater than 0 and less than 255. So we considered all the pixels greater than 0 as 255 in the ground truth training as well as testing masks

3.1 LinkNet

This Architecture is similar to U-Net in a way that it leverages the idea of information bottlenecking [11] to achieve state of the art results on the semantic segmentation problem. It uses ResNet18 which helps it perform quite better in this task. There are also connections between an upsampling block and a downsampling block in each layer. These connections help to recover lost spatial information that can be used by the upsampling blocks. In addition, since the upsampling block is sharing knowledge learnt by the downsampling block at every layer, the upsampling block can use fewer parameters. The authors of [9] show that this architecture is quite efficient in terms of operations per frame as well as overall parameters. This was also evident to us during training where training on the COLAB GPU only took 4s per epoch during training. We have taken the code from [15].

3.2 Mask-RCNN

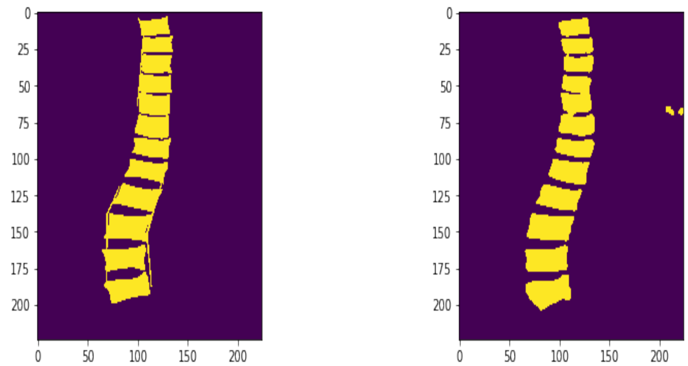
Mask R-CNN [10], extends Faster R-CNN [12] by adding a branch for predicting segmentation masks on each Region of Interest (RoI), in parallel with the existing branch for classification and bounding box regression. The mask branch is a small FCN applied to each RoI, predicting a segmentation mask in a pixel-to-pixel manner. It uses a RoIAlign layer that removes the harsh quantization of RoIPool by properly aligning the extracted features with the input. We have taken the code from [16].

4 Experiments

In this section we discuss the results of the architectures mentioned above on the Spine Dataset for the segmentation task. We report the mean Dice Coefficient, Jaccard Index and the True Positive Rate [13] of the given eight segmentation tasks on test set.

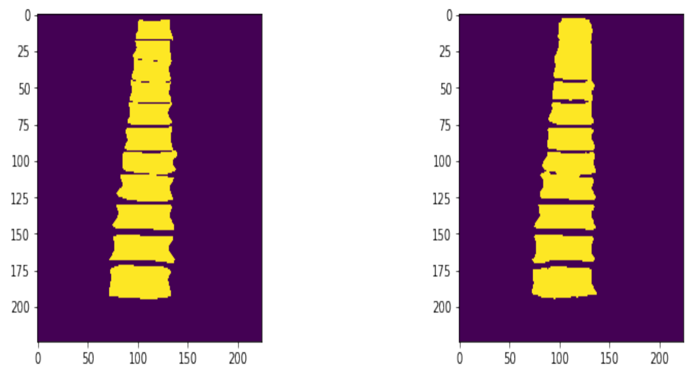
4.1 LinkNet

For training this architecture we used negative Dice Score as our loss function. During training we performed a 90:10 validation split on the training data and trained the model for 200 epochs. We



(a) Ground Truth (b) Predicted

Figure 2: Segmentation for LAT_Vertebra using LinkNet



(a) Ground Truth (b) Predicted

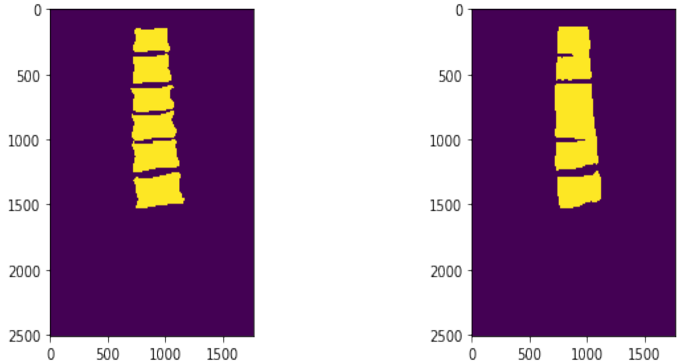
Figure 3: Segmentation for AP_Vertebra using LinkNet

saved the model whenever there was a increase in the validation Dice Score and thus have used the model that performs the best on the validation set. We also used early stopper and a learning rate scheduler to avoid overfitting. We hereby report our results below.

	Dice_Coeff	Jaccard Index	True Positive Rate
Ap_Vertebra	0.8667	0.7711	0.8823
Ap_Spinous_Process	0.5015	0.3481	0.4894
Ap_Pedicle	0.7527	0.6223	0.7541
Lat_Vertebra	0.8717	0.7808	0.8485
Lat_Disk_Height	0.7651	0.6331	0.7177
Lat_Spinous_Process	0.7632	0.6325	0.7319
Lat_Anterior_Line	0.3164	0.1845	0.4494
Lat_Posterior_Line	0.3014	0.1765	0.4394

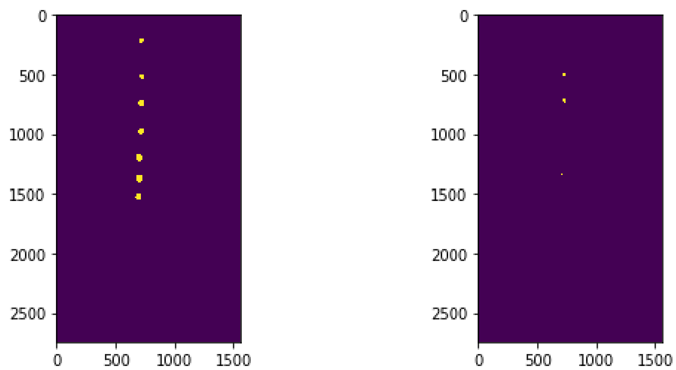
4.2 Mask-RCNN

We used the code given in [10] for implementing our Mask-RCNN. We initialised the model with weights pre-trained on the COCO dataset and ran the model for about 7000 iterations. We observe that it gives comparable results with our LinkNet architecture in some of the segmentations, particularly the ones whose masks have a larger and more continuous white region(Figure 4) and gives poor results for masks which have separated and smaller white regions(Figure 5). We figured this maybe because the model was fine-tuned specifically to the COCO dataset which has more of these type of



(a) Ground Truth (b) Predicted

Figure 4: Segmentation for AP_Vertebra using Mask-RCNN



(a) Ground Truth (b) Predicted

Figure 5: Segmentation for AP_Spinous_Process using Mask-RCNN

segmentations rather than discontinuous ones which justifies the results we get on our dataset. The detailed results are reported below.

	Dice_Coeff	Jaccard Index	True Positive Rate
Ap_Vertebra	0.8509	0.7463	0.8134
Ap_Spinous_Process	0.0932	0.05180	0
Ap_Pedicle	0.5407	0.3943	0.4024
Lat_Vertebra	0.8252	0.7107	0.7705
Lat_Disk_Height	0.5930	0.4361	0.4540
Lat_Spinous_Process	0.7414	0.6082	0.6568
Lat_Anterior_Line	0	0	0
Lat_Posterior_Line	0	0	0

5 Classification

We used the inbuilt Keras ResNet50 model for classification and initialised the weights as given in [14]. We divided the data into two parts- 578 samples for training and 100 samples for validation. The validation set consisted of 50 samples of the Damaged class and 50 samples of the Normal class. We used Model checkpoint to save the Model whenever there was an increase in accuracy and hence are submitting the model which runs best on the validation set.We train two models on our training set, one on the AP images and another on the LAT images. Then we find 3 predictions, first using only first model, second using only second model and third using average of the two models on the validation set. Finally we use the model which gives the highest accuracy on the validation set. We thus obtained an accuracy of 85% on the validation set.

6 Conclusion

We have reported both segmentation and classification results on this novel spine dataset. We compared two different architectures, one a specialized architecture for instance segmentation namely Mask-RCNN and the other an efficient architecture for semantic segmentation namely LinkNet. This assignment helped us to dive deep into semantic segmentation and the architectures used to solve the problem. This has been a learning curve for our team and we hope that our results are good. We sincerely thank Dr. Prathosh AP for giving us the oppotunity to work on this assignment.

References

[1] Scivoletto G, Morganti B, Molinari M. Early versus delayed inpatient spinal cord injury rehabilitation: An Italian study. Arch Phys Med Rehabil. 2005;86:512–6.

[2] Amin A, Bernard J, Nadarajah R, Davies N, Gow F, Tucker S. Spinal injuries admitted to a specialist hospital centre over a 5- year period: A study to evaluate delayed admission. Spinal Cord. 2005;43:434–7.

[3] Kishan S, Vives MJ, Reiter MF. Timing of surgery following spinal cord injury. J Spinal Cord Med. 2005;28:11–9.

[4] Reid DC, Henderson R, Saboe L, Miller JD. Etiology and clinical course of missed spine fractures. J Trauma. 1987;27:980 –986.

[5] Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: NIPS. pp. 1106–1114 (2012)

[6] Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2014)

[7] LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W., Jackel, L.D.: Backpropagation applied to handwritten zip code recognition. Neural Computation 1(4), 541–551 (1989)

[8] Ronneberger, O., Fischer, P., Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham.

[9] Chaurasia, A., Culurciello, E. (2017, December). Linknet: Exploiting encoder representations for efficient semantic segmentation. In 2017 IEEE Visual Communications and Image Processing (VCIP) (pp. 1-4). IEEE.

[10] He, K., Gkioxari, G., Dollár, P., Girshick, R. (2017). Mask r-cnn. In Proceedings of the IEEE international conference on computer vision (pp. 2961-2969).

[11] Tishby, N., Zaslavsky, N. (2015, April). Deep learning and the information bottleneck principle. In 2015 IEEE Information Theory Workshop (ITW) (pp. 1-5). IEEE.

[12] S. Ren, K. He, R. Girshick, and J. Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. In NIPS, 2015.

[13] Taha, A. A., Hanbury, A. (2015). Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. BMC medical imaging, 15(1), 29.

[14] https://www.kaggle.com/donpiano/keras-resnet-50-for-pneumonia-x-ray-images/data

[15] https://github.com/rekon/Smoke-semantic-segmentation/blob/linknet-implementation/LinkNet.ipynb

[16] https://github.com/facebookresearch/detectron2/tree/master/detectron2