Automatic Segmentation of Lumbar Spine 3D MRI

Using Ensemble of 2D Algorithms.

Nedelcho Georgiev (nedelcho.georgiev@smart-soft.net), Asen Asenov
Smart Soft

Introduction

MRI is considered the gold standard in soft tissue diagnostic of the lumbar spine. Number of protocols and modalities are used – from one hand 2D sagittal, 2D angulated axial, 2D consecutive axial and 3D image types; from the other hand different sequences and contrasts are used: T1w, T2w; fat suppression, water suppression etc. Images of different modalities are not always aligned. Resolutions and field of view also vary. SNR is also different for different MRI equipment. So the goal should be to create algorithms that cover great variety of imaging techniques.

We consider the segmentation as the first step in a 3-step process: 1. Segmentation Fig 1(a); 2. Measurements Fig1(b); 3. Diagnosis (in the case shown in the fig.1 (b) severity of disk herniation and central canal stenosis grading)

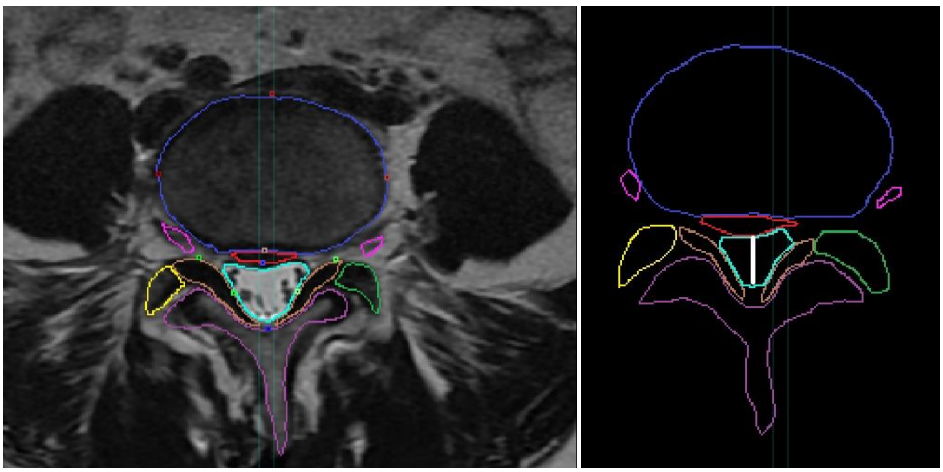

Fig. 1 (a) Segmentation of axial T2w slide; (b) Measurement of dural sac in a different slide (white line)

Methods

In order to cover different protocols a 2D single modality algorithm was developed that in cases of 3D multi-modality data can be used in ensemble of multiple 2D single modality data classifiers. Our 2D algorithms are using CNN [1] and are greatly inspired of ResNet[2]. Dropout regularization [3] and batch normalizations [4] are also used. FCN[5] style network is used for the super-pixel classification as this enable arbitrary field of view input. Super-pixels are 8 times smaller (in all dimensions) than the actual pixels (as stride 8 is used due to 3 2x2 max pooling layers). As a result, fine grained details are lost, so we use Unet-like [6] architecture for up-scaling the low-resolution map into the resolution of the input image.

Separate 2D single modality results are combined into 3D results using ensemble with learnable weighted averaging of the separate probability maps.

To overcome the problem with small data set size, extensive augmentations were used: crop and resize, tilt, rotate, dynamic range changes, random noise in all possible combinations.

The model detects all the IVDs present in the image. To filter the 7 discs (L5S1-Th11Th12) that are needed, a geometrical model of the disc composition was developed. It is scale, position and orientation invariant. The features are the relative orientation of the discs and the relative angle between the centers of the discs. The base composition is calculated as the mean features of the training data. During inference, all the 7-length subsequences of the N discs found by the mask model are compared to the base and the most similar is used as output.

Results

Results achieved on cross validation using the data provided by the organizers [7] are listed in Table 1. We have no access to the data of the previous challenges so we used only 16 3D multi-modality MRI.

2D planes used	DICE
Sagittal	0.84
Axial	0.87
Coronal	0.82
Ensemble	0.92

Table 1 Comparison of single plane results vs 3D ensemble

Test time is 3:10s on a single GPU machine that can be reduced to 1:20 in batch mode. The model supports resolution of 512/512 that is 4 (2*2) times bigger than necessary for this particular set so test time can be further reduced.

Conclusion

Test accuracy was comparable to the previously reported results using 3D convolutions of the previous challenge [8] although the algorithm was designed for 2D single-modality data and we have not used data from previous challenges. Training set accuracy is near 100% which can be expected as the complexity of the model is very big and definitely high variance is the current drawback. Never the less we decided to not reduce the complexity as we believe bigger training set is necessary for reaching human level accuracy. So further test set accuracy improvements can be expected by increasing the training set. Our intention for the future development is to cover great variety of tissues and pathologies by acquiring an annotated training set of 500 patients. Part of them will be released to the scientific community.

References

1. Le Cun, Y., Bottou, L., and Bengio, Y. (1997). Reading checks with multilayer graph transformer networks. In Acoustics, Speech, and Signal Processing, 1997. ICASSP-97., 1997 IEEE International Conference on, volume 1, pages 151–154. IEEE.
2. He, Kaiming & Zhang, Xiangyu & Ren, Shaoqing & Sun, Jian. (2015). Deep Residual Learning for Image Recognition.
3. Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., and Salakhutdinv, R. (2012). Improving neural networks by preventing co-adaptation of feature detectors. Technical report, arXiv:1207.0580.
4. Ioffe, S., Szegedy, C.: Batch normalization: accelerating deep network training by reducing internal covariate shift. CoRR abs/1502.03167 (2015)
5. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation (2014), arXiv:1411.4038 [cs.CV]

6. Ronneberger O., Fischer P., Brox T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab N., Hornegger J., Wells W., Frangi A. (eds) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science, vol 9351. Springer, Cham
7. C. Chen, D. Belavy and G. Zheng, "3D Intervertebral Disc Localization and Segmentation from MR Images by Data-driven Regression and Classificaiton", the 5th International Workshop on Machine Learning in Medical Imaging (MLMI 2014), pp. 50 – 58, 2014
8. Li, Xiaomeng & Dou, Qi & Chen, Hao & Fu, Chi-Wing & Qi, Xiaojuan & Belavy, Daniel & Armbrecht, Gabriele & Felsenberg, Dieter & Zheng, Guoyan & Heng, Pheng-Ann. (2018). 3D multi-scale FCN with random modality voxel dropout learning for Intervertebral Disc Localization and Segmentation from Multi-modality MR Images. Medical image analysis. 45. 41-54. 10.1016/j.media.2018.01.004.