Analysing TB Severity Levels With An Enhanced Deep Residual Learning – Depth-Resnet

Xiaohong Gao, Carl James-Reynolds, Ed Currie

Department of Computer Science, Middlesex University, London, UK

{x.gao, c.james-reynolds, e.currie} @mdx.ac.uk

Abstract. This work responds to the Competition of Tuberculosis Task organised by imageCLEF 2018. While Task #3 appears to be challenging, the experience was very enjoyable. If time had been permitted, it was certain that more accurate results could have been achieved. The authors submitted 2 runs. Based on the given training datasets with severity levels of 1 to 5, an enhanced deep residual learning architecture, depth-ResNet, is developed and applied to train the datasets to classify 5 categories. The datasets are pre-processed with each volume being segmented into twenty- 128×128×depth blocks with ~64 pixel overlaps. While each block has been predicted with a severity level, assembling all constituent block scores together to give an overall label for the concerned volume tends to be more challenging. Since the probability of high severity is not provided from the training datasets, which bears little resemblance to the classification probability, the submission of probability for the first run was manually assigned as 0.9, 0.7, 0.5, 0.3, and 0.1 to severity levels of 1 to 5 respectively. After the deadline was extended, the model was re-trained with frame numbers increased from 1 to 8, which takes much longer to train. In addition, a new measure was introduced to calculate the overall probability of high severity based on the block scores. As a result, with regard to classification accuracy, the 2nd submitted run achieved place 14 over a total of 36 submissions, a significant improvement from position of 35 from the first run.

Keywords: Deep residual learning, classification, severity of Tuberculosis

1. Deep Residual Learning – Depth-Resnet

Since a convolutional neural network (CNN) architecture can be constructed by stacking multiple layers of convolution and subsampling in an alternating fashion, a CNN network can be enhanced into going deeper by piling a large number of layers. However, the increased depth appears to have little contribution to the accuracy of a trained model. This is due to the well-known vanishing gradient obstacle, i.e. as the gradient is back-propagated to earlier layers, repeated multiplication may compose the gradient infinitely small. As a result, as the network becomes deeper, its performance gets saturated or even starts degrading rapidly.

Recently, deep residual networks (ResNet) [1-3] introduce the notion of '*identity shortcut connection*' that bypasses one or more layers. A key advantage of residual units is that their skip connections allow direct signal propagation from the first to the last layer of the network, especially during backpropagation. This is due to the fact that gradients are propagated directly from the loss layer to any previous layer while skipping intermediate weight layers that have potential to trigger vanishing or deterioration of the gradient signal.

In this work, an enhanced ResNet, i.e. depth-Resnet is applied for analysis of the level of severity of tuberculosis from CT lung images, which is built on ResNet-50 model and illustrated in Figure 1.

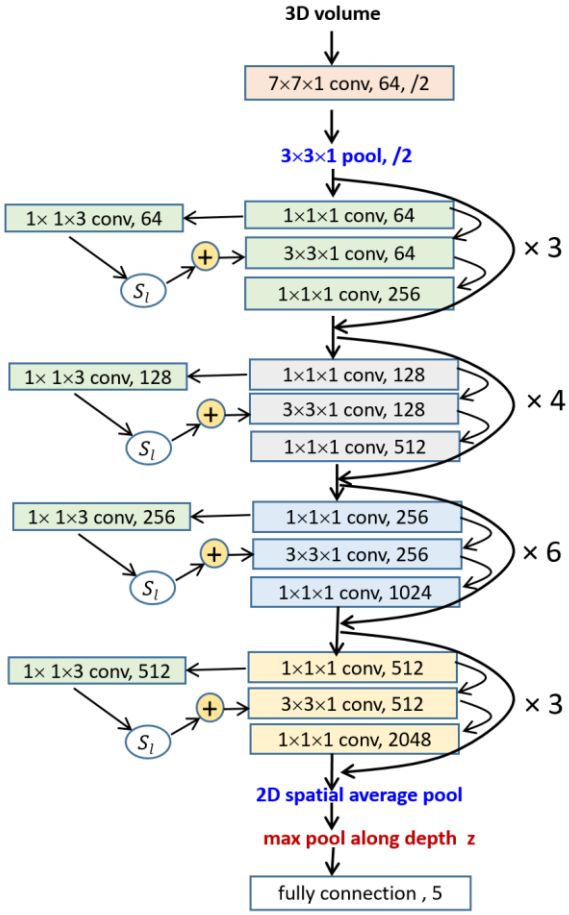

Fig. 1. The depth-ResNet architecture applied in this paper, where ×N at each conv level refers to the block (e.g. conv5_x) repeats N (e.g. 3) times consecutively.

In Figure 1 built on the inception concept, the depth (z) convolution block operates on the dimensionality reduced input, $x_{l,z}$ with a bank of 3D filters, $W_{l,z}$. Biases $b \in \mathbb{R}^C$ are also applied with initial values of 0 as formulated in Eq. (1).

$$x_{l,z} = W_{l,z} x_{l,1} + b \tag{1}$$

Hence the residual unit $\mathcal{F}$ is expressed in Eq. (2).

$$\mathcal{F} = f\left(W_{l,3}\left(S_l f\left(x_{l,z}\right) + f\left(W_{l,2} f\left(W_{l,1} x_{l,1}\right)\right)\right)\right) \tag{2}$$

where S_l is affine scaling along depth direction with a bias between 0 and 0.01. This scaling is adaptive to facilitate generalisation performance and will be learnt during the training of the network. The convolution at each convolution layer along depth (z) direction ($x_{l,z}$) takes place between 3 neighbouring slices or feature maps, i.e. front, current, and back, with a randomly chosen stride (between 1 and 7 in this study). This feature then is added to the block with a scaling factor as a component of the residual unit. The pooling involves two stages. The *avg-pool* occurs for 2D spatial global average pooling whereas *max-pool* is conducted along z direction performing global max pooling upon those feature maps.

On the other hand, to integrate block scores into a volumetric label for each dataset, a support vector machine (SVM) is applied. The system is implemented in Matlab with MatConvNet [4] toolbox by following standard ConvNet training procedures [5, 6]. Upon training, 8 slices is chosen from each block with randomly selected stride between 1 and 7 from 5 categories with a batch of 128 (=16 blocks). At testing time, each dataset

undertakes the same pre-processing procedure to generate 128×128×depth blocks. Then the trained depth-Resnet model (Figure 1) takes each block as a whole, selects 8 slices at equal depth space and propagates these slices through the trained model to produce a single prediction for this block with severity scores labeled between 1 and 5.

2. Datasets

Data are collected from the competition organised by ImageCLEF2018 on Tuberculosis severity scoring task (task#3) [7, 8, 9], with 170 for training and 109 for testing. The training data include chest CT scans of TB from 170 patients with the corresponding severity scores (1 to 5) and the severity levels designated as "*high*" and "*low*", which contains 90 low severity (with scores 4 and 5) and 80 high severity (with scores 1, 2 and 3).

3. Image Pre-processing

The collected data are pre-processed to remove background and to segment into smaller blocks, which is because that most abnormalities occur in small regions and spread over only a few slices. Figure 2 demonstrates the process to remove background by the application of masks that are dilated in advance. As illustrated in Figure 2 (b), some masks [10] over-remove lung information. Hence all the masks are dilated (Figure 2(d)) by a diameter of 30 pixels found empirically to ensure the balance between over- and under- removing of background (Figure 2(e)). Figure 2(f) depicts the final image of removing background, which has a size of $460 \times 340 \times z$ (z is the depth and varies between 50 to 400).

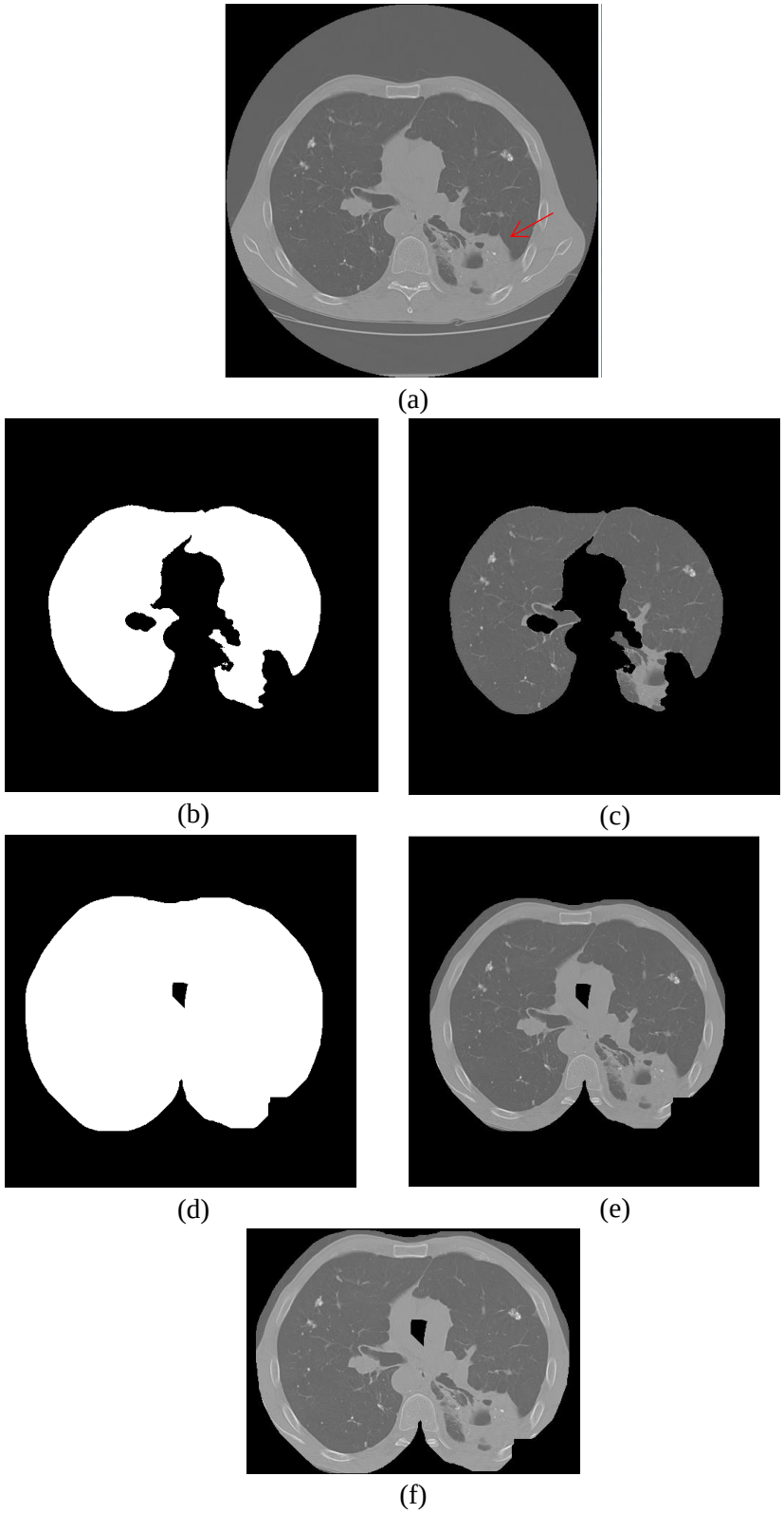

Fig. 2. The process of removing background using dilated mask. (a) an original slice of image; (b) original mask and the masked image (c) between (a) and (b); (d) dilated mask and (e) the result between (a) and (d). (f) final segmented masked image after removing background. The arrow points to the diseased region to be concerned.

Then upon the segmented volume of $460 \times 340 \times z$, 24 blocks of size of $128 \times 128 \times z$ are created with overlapping of ~64 pixels as illustrated in Figure 3.

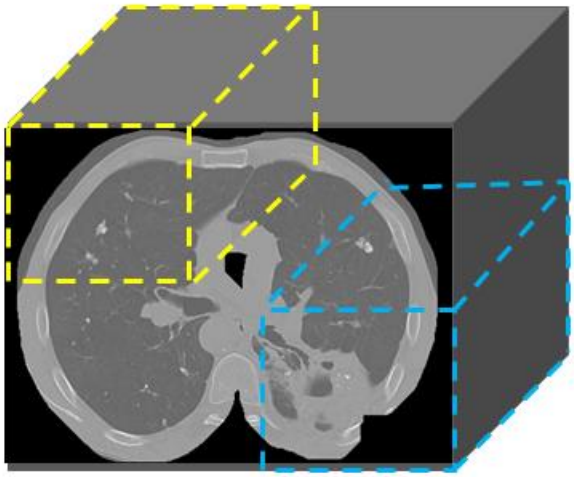

Fig. 3. Segmentation of processed volume into 24 overlapping blocks with each one being 128×128×depth.

Since some corner blocks comprise large amount of background information, i.e. pixel value being 0, these frames, in particular at front and back of a volume along z direction, are removed when its background region occupies more than one third of the total space. Hence the depth (z) of each block varies between 11 and 250 for all datasets after segmentation. As a result, many 3D volume datasets have less than 24 blocks after pre-processing. Each block has been resized to $256 \times 256 \times z$ from $128 \times 128 \times z$ to save training time.

4. Results

Since each volume of 3D dataset contains around 24 blocks with individual severity scores, the final score for this dataset has to be integrated. In principle, the five levels of severity can be treated as 2 classes labeled as 'high' (with scores 1, 2 and 3) and 'low' (with scores 4 and 5). Hence, three measures can then be formulated to convey the inter-relationships between blocks scored 1 to 3, 4 to 5 and 1 to 5 are then calculated using Eqs. (3), (4) and (5) respectively where scores of 1 to 5 are assigned initial probabilities of high severity linearly by 0.9, 0.7, 0.5, 0.3, and 0.1 respectively.

$$prob\_high = \frac{0.9 \times num_{block1} + 0.7 \times num_{block2} + 0.5 \times num_{block3}}{num_{block1} + num_{block2} + num_{block_3}} \quad (3)$$

$$prob\_low = \frac{0.3 \times num_{block4} + 0.1 \times num_{block5}}{num_{block4} + num_{block5}} \quad (4)$$

$$prob\_all = \frac{0.9 \times num_{block1} + 0.7 \times num_{block2} + 0.5 \times num_{block3} + 0.3 \times num_{block4} + 0.1 \times num_{block5}}{num_{block1} + num_{block2} + num_{block_3} + num_{block4} + num_{block5}} \quad (5)$$

Hence the probability of a whole volume dataset can then be decided by these measures, which is in turn utilized to score the overall severity of a volume. For example, in this study, if a dataset has $probability_{high} > 0.7$ and $probability_{low} < 0.20$ and $Num_{block1} > 0$, then this dataset is classified as severity 1. In Table 1, two approaches are applied. One is based on the overall probability (Level-1) as formulated in Eq. (5), which is simple and straightforward. Level 1 approach has been applied to the imageCLEF Tuberculosis 2018 competition [8, 9], which ranks number 14 (out of 36 submissions) in terms of accuracy (AUC=0.6534) by the authors of this paper when using different set of test data (n=109) with unknown severity levels. These results are based on three runs using training datasets where 100 data are randomly selected from 170 training sets and remaining 70 as test.

Table 1. The accuracy performance from both Level-1 and Level-2 calculations.

Severity	1	2	3	4	5	Average
Level-1	0.80 ± 0.00	0.60 ± 0.05	0.75 ± 0.00	0.88 ± 0.02	0.82 ± 0.12	**75.88 ± 3.80%**
Level-2	0.86 ± 0.08	0.70 ± 0.01	0.77 ± 0.02	0.90 ± 0.00	0.84 ± 0.04	**85.29 ± 3.00%**

5. Conclusion

Prediction of probability of high severity level appears to be a challenging task since this information has to be determined from the severity scores of 1 to 5. Due to the limited computer power (with only 1 GPU), each run takes 4 days to train (100 datasets) and 2 days to test, the results from Level-2 approach were only obtained after the deadline. However, the experience gained from this competition was very enjoyable with many lessons learnt in relation to designing deep residual learning network.

References

1. He K, Zhang X, Ren S, Sun J, Identity Mappings in Deep Residual Networks, *European Conference on Computer Vision (ECCV)* (2016).
2. He K, Zhang X, Ren S, Sun J, Deep Residual Learning for Image Recognition, IEEE *Conference on Computer Vision and Pattern Recognition (CVPR)* (2016).
3. Feichtenhofer C, Pinz A, Wildes R, Temporal Residual Networks for Dynamic Scene Recognition, IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017).
4. MatConvNet: http://www.vlfeat.org/matconvnet/. Retrieved in May (2018).
5. *LeCun Y, Bengio Y, Hinton G, Deep Learning, Nature. 521: 436-444 (2015).*
6. Krizhevsky A, Sutskever I, Hinton G. Imagenet classification with deep convolutional neural networks, *Advances in neural information processing systems 2012.* NIPS 2012 (2012).
7. Cappellato L., Ferro N., Nie J, Soulier L, Eds., CLEF 2018 Working Notes, Working Notes of CLEF 2018 – Conference and Labs of the Evaluation Forum, CEUR-WS, Eds. (2018).
8. Dicente Cid Y., Liauchuk V., Kovalev V., Müller H., Overview of ImageCLEFtuberculosis 2018 - Detecting multi-drug resistance, classifying tuberculosis type, and assessing severity score, CLEF working notes, CEUR, 2018., CEUR-WS.org, September 10-14, Avignon, France (2018).
9. Ionescu B., Müller H., Villegas M., de Herrera A., Eickhoff C., Andrearczyk V., Cid Y.D., Liauchuk V., Kovalev V., Hasan S.A., Ling Y., Farri O., Liu J., Lungren M., Dang-Nguyen DT, Piras L., Riegler M., Zhou L., Lux M., Gurrin C., Overview of ImageCLEF 2018: Challenges, Datasets and Evaluation, Experimental IR Meets Multilinguality, Multimodality, and Interaction, Proceedings of the Ninth International Conference of the CLEF Association (CLEF 2018), September 10-14, Avignon, France (2018).
10. Cid YD, Jiménez-del-Toro OA, Depeursinge A, Müller H, Efficient and fully automatic segmentation of the lungs in CT volumes. In: Goksel, O., et al. (eds.) Proceedings of the VISCERAL Challenge at ISBI. No. 1390 in CEUR Workshop Proceedings (2015).