A Multi-Task Self-Normalizing 3D-CNN to Infer Tuberculosis Radiological Manifestations

Pedro M. Gordaliza[1] PMACIAS@ING.UC3M.ES
Juan José Vaquero[1] JUANJOSE.VAQUERO@UC3M.ES
Sally Sharpe[2] SALLY.SHARPE@HPA.ORG.UK
Fergus Gleeson[3] FERGUS.GLEESON@ONCOLOGY.OX.AC.UK
Arrate Muñoz-Barrutia[1] MAMUNOZB@ING.UC3M.ES

[1] *Dpto.Bioingenieria e Ing. Aeroespacial, Universidad Carlos III de Madrid, Leganés, Spain*
[2] *Microbiology Services Division, Public Health England, Porton Down, England*
[3] *Oncology Department, The Churchill Hospital, Headington, Oxford, UK*

Editors: Under Review for MIDL 2019

Abstract

We propose a learning method well-suited to infer the presence of Tuberculosis (TB) manifestations on Computer Tomography (CT) scans mimicking the radiologist reports. Latent features are extracted from the CT volumes employing the *V-Net* encoder and those are the input to a *Feed-Forward Neural Network (FNN)* for multi-class classification. To overtake the issues (e.g., exploding/vanishing gradients, lack of sensibility) that normally appear when training deep 3D models with datasets of limited size and composed of large volumes, our proposal employs: 1) At the network architecture level, the *scaled exponential linear unit* (SELU) activation which allows the network self-normalization, and 2) at the learning phase, multi-task learning with a loss function weighted by the task *homoscedastic* uncertainty. The results achieve F_1-scores close to or above 0.9 for the detection of TB lesions and a Root Mean Square Error of 1.16 for the number of nodules.

Keywords: Self-Normalizing Neural Networks, multi-task, homoscedastic uncertainty

1. Introduction

Tuberculosis (TB) is an infectious disease which generally affects the lungs and has a high incidence and mortality (World Health Organization and others, 2018). Due to the severity of the pandemic, the World Health Organization (WHO) has launched an ambitious plan to eradicate TB by 2030, for which, the extraction of sensitive radiological biomarkers (Nachiappan et al., 2017) is a clear need. Traditionally, radiologists through visual inspection of x-ray Computed Tomography (CT) volumes generate reports summarizing the presence of TB-related manifestations. However, using this approach for the extraction of robust TB radiological biomarkers, it is unfeasible and automation required. In recent years, the introduction of deep learning techniques has drastically contributed to this task (Wang et al., 2017; Litjens et al., 2017; Hinton, 2018). For deep learning, knowledge is usually injected into the model in the form of manually segmented masks of the lung lesions. Instead, our proposal directly employs the expertise acquired by the radiologist through years of clinical practice as synthesized in tabular reports.

© 2019 P.M. Gordaliza, J.J. Vaquero, S. Sharpe, F. Gleeson & A. Muñoz-Barrutia.

2. Material and Methods

Chest CT scans (56) acquired from 14 male Cynomolgus macaques at 3, 7, 11 and 16 weeks after TB aerosol exposure were employed. First, the CT volumes were cropped (Gordaliza et al., 2018) and resampled to 1mm × 1mm × 2mm. During the training, data augmentation is performed (elastic transformation, addition of Gaussian noise). As labelled data, we employ the tabular reports elaborated by a radiologist (20 years experienced) that contain the number of detected nodules $(0 - 15)$ and binary annotations indicating the presence per lung lobe of the most common TB manifestations (e.g., cavitations, conglomerations, consolidations and trees in bud) (Nachiappan et al., 2017).

2.1. Model Architecture

Our implementation (Figure 1) extracts latent features from the volumes employing the *V-Net* (Milletari et al., 2016) encoder. The extracted features are used as input of the *Feed-Forward Neural Networks* (FNNs) for multi-class classification. The encoder generates $1,376.256$ features, which feed FNN_1 (task-shared). The outputs of FNN_1 are employed as the input of two independent FNNs, FNN_R and FNN_B, corresponding with regression (nodule counting) and binary tasks. Dropout or Batch Normalization (BN) (Ioffe and Szegedy, 2015) layers are included where is needed. When employing SNN, BN is unnecessary.

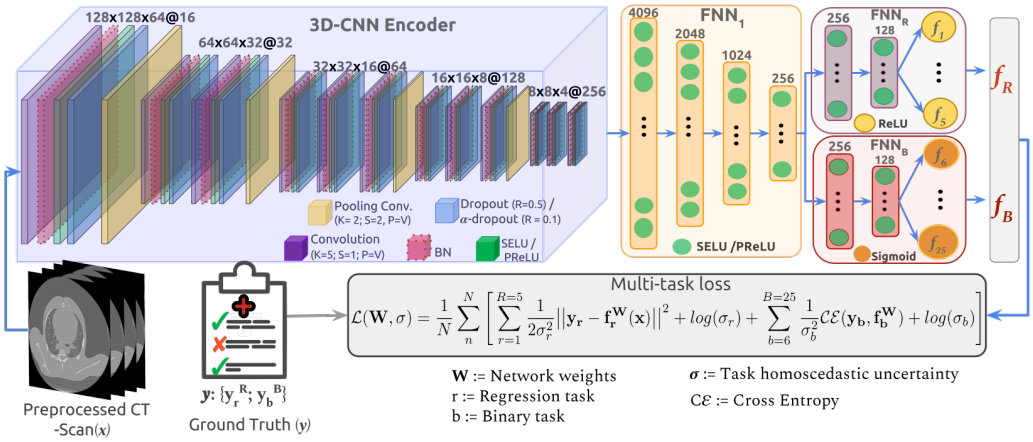

Figure 1: 3D-CNN + three FNNs (*Feed Forward neural networks*): FNN_1 (tasks-shared parameters) and, FNN_R and FNN_B for prediction of regression and binary tasks. BN is not present with SELU.

2.2. Self-Normalizing Neural Networks

For regularization and normalization purposes, we apply the *Self-Normalizing Neural Networks* (SNNs) strategy to our model (Klambauer et al., 2017) that preserves the activation magnitude close to zero mean and unit variance. As activation function, we use the *Scaled Exponential Linear Unit* (SELU) and as dropout, the α-dropout (Klambauer et al., 2017).

2.3. Learning Principle: Uncertainty Weighted Multi-task Loss

When working in multi-task classification, the decision on the influence of each task in the final loss is non-trivial. Traditionally, the loss is computed as $\mathcal{L} = \sum_i w_i \mathcal{L}_i$, where the

weights w_i are selected either manually or after an exhaustive grid search. This approach is highly influenced by the units and the scale of each task and extremely time-consuming. Recently, (Kendall et al., 2018) *et al.* proposed a method to compute the weights guiding each specific task loss by *homoscedastic* uncertainty of the predictions. In this work, we adapt this approach to our multi-label classification problem and derive a loss function for our model (Figure 1).

3. Experiments and Results

For the evaluation, we employed 5-fold Cross-Validation (CV): Validation set (4 CT volumes of one subject); training set (remaining 13 subjects). We compared the proposed model (*SELU*) and a modified version which employs a Parametric Rectified Linear Unit (PReLU), Batch Normalization and standard dropout (*BN+PReLU*). The models are trained with 10.000 iterations of the ADAM optimizer (Kingma and Ba, 2014) (learning rate= 10^{-5}, mini-batch size= 15). For *BN+PReLU*, standard parameters were employed for the ADAM optimizer and a dropout rate of 0.5. For *SELU*: $\beta_2 = 0.9$, $\epsilon = 0.01$ and alpha-dropout= 0.1 (Klambauer et al., 2017). Figure 2 shows for the validation data, the loss at each fold for the BN+PReLU (in red) and the proposed model (in blue). It can be observed that SELU presents improved convergence. The inference error is estimated by the Root Mean Square Error (RMSE) for the nodules count tasks and the F_1-score for the twenty binary tasks. Table 1 presents the results per fold. No significant statistical differences were found for paired t-test ($p \nleq 0.05$). Nevertheless, the proposed model presents better *RMSE* and F_1-score results.

Table 1: Performance measure results the for BN+PReLU and the proposed model (SELU).

Manifestation/	**Nodules** [RMSE]		Cavitations [F_1]		Conglomeration [F_1]		Consolidation [F_1]		**Tree in bud** [F_1]	
Fold	BN+PReLU	SELU	BN+PReLU	SELU	BN+PReLU	SELU	BN+PReLU	SELU	BN+PReLU	SELU
1	$0.73_{0.84}$	$0.85_{0.35}$	$0.88_{0.11}$	$0.88_{0.12}$	$0.90_{0.13}$	$0.92_{0.12}$	$0.88_{0.18}$	$0.83_{0.22}$	$0.83_{0.22}$	$0.79_{0.23}$
2	$1.15_{0.89}$	$1.09_{0.83}$	$0.86_{0.23}$	$0.88_{0.22}$	$0.94_{0.17}$	$0.93_{0.18}$	$0.93_{0.15}$	$0.93_{0.18}$	$0.97_{0.08}$	$0.97_{0.08}$
3	$0.41_{0.34}$	$0.23_{0.39}$	$0.85_{0.12}$	$0.87_{0.11}$	$0.89_{0.19}$	$0.97_{0.11}$	$0.96_{0.11}$	$0.98_{0.03}$	$0.95_{0.14}$	$0.96_{0.12}$
4	$1.22_{0.6}$	$0.78_{0.74}$	$0.94_{0.15}$	$0.9_{0.19}$	$0.9_{0.19}$	$0.92_{0.15}$	$0.88_{0.18}$	$0.87_{0.18}$	$0.87_{0.21}$	$0.94_{0.15}$
5	$0.41_{0.8}$	$0.45_{0.8}$	$0.93_{0.17}$	$0.94_{0.17}$	$0.94_{0.14}$	$0.96_{0.12}$	$0.90_{0.18}$	$0.94_{0.14}$	$0.91_{0.18}$	$0.92_{0.17}$
Total	$0.78_{0.69}$	$\mathbf{0.68_{0.62}}$	$0.89_{0.16}$	$\mathbf{0.90_{0.17}}$	$0.91_{0.16}$	$\mathbf{0.94_{0.16}}$	$0.91_{0.16}$	$0.91_{0.15}$	$0.91_{0.17}$	$\mathbf{0.92_{0.15}}$

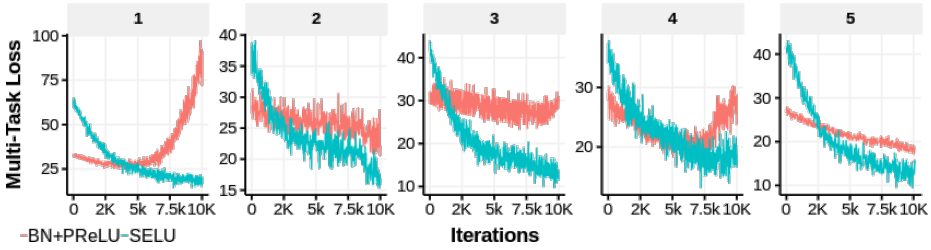

Figure 2: Evolution of the validation loss for each fold of the Cross-Validation.

4. Conclusions

Although further validation on large datasets is needed, the work presents a promising inference of the radiologist reports. This is achieved with a reduced computational complexity by avoiding normalization layers and hyperparameter tuning of the loss weights.

Acknowledgments

The research leading to these results received funding from the Innovative Medicines Initiative (www.imi.europa.eu) Joint Undertaking under grant agreement no. 115337, whose resources comprise funding from EU FP7/2007-2013 and EFPIA companies in kind contribution. This work was partially funded by projects RTC-2015-3772-1, TEC2015-73064-EXP and TEC2016-78052-R from the Spanish Ministry of Economy, Industry and Competitiveness (MEIC), TOPUS S2013/MIT-3024 project from the regional government of Madrid and by the Department of Health, UK.
This material is based upon work supported by Google Cloud Platform.

References

Pedro M. Gordaliza, Arrate Muñoz-Barrutia, Mónica Abella, Manuel Desco, Sally Sharpe, and Juan José Vaquero. Unsupervised CT Lung Image Segmentation of a Mycobacterium Tuberculosis Infection Model. *Scientific Reports*, 8(1):9802, 12 2018. ISSN 2045-2322. doi: 10.1038/s41598-018-28100-x. URL http://www.nature.com/articles/s41598-018-28100-x.

Geoffrey Hinton. Deep Learning—A Technology With the Potential to Transform Health Care. *JAMA*, 320(11):1101, 9 2018. ISSN 0098-7484. doi: 10.1001/jama.2018.11100. URL http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.2018.11100.

Sergey Ioffe and Christian Szegedy. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. 2 2015. URL http://arxiv.org/abs/1502.03167.

Alex Kendall, Yarin Gal, and Roberto Cipolla. Multi-Task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–10, 2018. URL https://arxiv.org/pdf/1705.07115.pdf.

Diederik P. Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization. 12 2014. URL http://arxiv.org/abs/1412.6980.

Günter Klambauer, Thomas Unterthiner, Andreas Mayr, and Sepp Hochreiter. Self-Normalizing Neural Networks. In *Advances in Neural Information Processing Systems 30*, pages 971–980, 2017. URL https://papers.nips.cc/paper/6698-self-normalizing-neural-networks.pdf.

Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen A.W.M. van der Laak, Bram van Ginneken, and Clara I. Sánchez. A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42:60–88, 12 2017. ISSN 1361-8415. doi: 10.1016/J.MEDIA.2017.07.005. URL https://www.sciencedirect.com/science/article/pii/S1361841517301135.

Fausto Milletari, Nassir Navab, and Seyed Ahmad Ahmadi. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. *2016 Fourth International*

Conference on 3D Vision (3DV), 565 - 571:565–571, 2016. doi: 10.1109/3DV.2016.79. URL http://ieeexplore.ieee.org/document/7785132/.

Arun C Nachiappan, Kasra Rahbar, Xiao Shi, Elizabeth S Guy, Eduardo Mortani Barbosa Jr, Girish S Shroff, Daniel Ocazionez, Alan E Schlesinger, Sharyn I Katz, and Mark M Hammer. Pulmonary Tuberculosis: Role of Radiology in Diagnosis and Management. *RadioGraphics*, 37(1):52–72, 2017. doi: 10.1148/rg.2017160032. URL https://doi.org/10.1148/rg.2017160032.

Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu, Mohammadhadi Bagheri, and Ronald M Summers. ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases. In *Computer Vision and Pattern Recognition (CVPR)*, pages 3462–3471, 2017. URL https://uts.nlm.nih.gov/metathesaurus.html.

World Health Organization and others. Global tuberculosis report 2018. Technical report, 2018. URL http://apps.who.int/bookorders.