LEARNING VOLUMETRIC SEGMENTATION FOR LUNG TUMOR

Jhih-Yuan Lin[5], *Min-Sheng Wu*[6], *Yu-Cheng Chang*[4], *Yun-Chun Chen*[2]
Chao-Te Chou[2], *Chun-Ting Wu*[3], *Winston Hsu*[1,3]

[1]Department of Computer Science and Information Engineering, National Taiwan University
[2]Department of Electrical Engineering, National Taiwan University
[3]Graduate Institute of Networking and Multimedia, National Taiwan University
[4] Department of Computer Science and Information Engineering, National Cheng Kung University
[5]Department of Electrical Engineering, National Cheng Kung University
[6]Department of Biomedical Engineering, National Cheng Kung University

ABSTRACT

Convolutional neural networks have recently been applied to various computer vision tasks as well as medical image analysis. Despite their popularity, most of the existing methods are addressing problem where data is presented in the form of 2D images (e.g., semantic segmentation). However, in medical image analysis task, data often contains 3D volumetric contents. In this paper, we present an approach that addresses the challenging 3D image segmentation. Our model is end-to-end trainable on CT images and learns to predict volumetric segmentation outputs. To handle the imbalance distribution between the foreground and background region, we employ the dice loss and the focal loss for optimization. As such, our model can precisely segment sparse objects within dense region. Extensive experimental evaluations on a challenging benchmark demonstrate that out model performs favorably against the state-of-the-art methods.

Index Terms— Image Segmentation, 3D Segmentation, U-Net, V-Net, Lung Cancer Tumor Segmentation, Convolutional Neural Network

1. INTRODUCTION

The vision community has witnessed the success demonstrated by convolutional neural networks (CNNs) in various computer vision tasks such as semantic segmentation, object detection, and image classification and recognition. Among these tasks, data is often given in the form of 2D (planar) images. While numerous methods have shown promising results in dealing with tasks involving 2D data, there are many cases that data inherently contains spatial information and is presented in the form of 3D (volumetric) images. For instance, point clouds generated by LiDAR contain the precise spatial locations. Medical data, in particular, contains spatial information as well.

To tackle the tasks that involve 3D images, several attempts have been made by leveraging image projection techniques in conjunction with off-the-shelf approaches to bypass the large computation efforts required for processing 3D data. Take segmentation task for example, one can project the voxels and form a projected 2D image. The resultant image can then be processed by applying existing 2D semantic segmentation methods. Afterwards, a re-projection operation can be applied on the 2D segmentation output to obtain the desired 3D segmentation result. As such, processing 3D data can be more efficient and computationally tractable. However, projecting data would entail large information loss, which will inevitably result in sub-optimal performance and may limit the generalization capability.

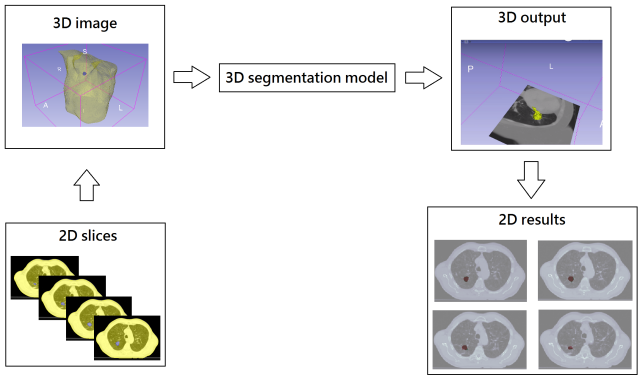

Fig. 1: **Overview of our approach.** We stack the 2D slices into 3D image and use our 3D segmentation model to predict 3D segmentation results. By projecting 3D segmentation results, we can obtain the desired 2D segmentation results. The upper right image is from the internet.

In this competition, our goal is to segment the lung cancer tumor given the 2D data and the corresponding segmentation results for each patient. In light of the information loss in applying 2D segmentation methods, we consider to learn lung cancer tumor segmentation in a 3D fashion as illustrated in Figure 1. For each patient, we cascade all the slices to form

the 3D image. The corresponding 3D ground truth can be obtained similarly. As such, the relationship between each slice can be observed accordingly. To process the 3D data, we propose an end-to-end trainable network that learns lung cancer tumor segmentation. Inspired by Wang *et al*. [1], we propose the dense down-sampling convolutional (DDC) layer and introduce the lateral connections into the proposed network, which reduces the computation burden and encourages faster convergence, respectively. The network then takes the resultant 3D data as input to perform lung tumor segmentation. At test phase, our model produces segmentation results and the 2D segmentation result of a particular slice of image can be obtained by applying an image projection.

The contributions of this work are summarized as follows. First, we propose an end-to-end trainable network that performs lung cancer tumor segmentation in a supervised fashion. Second, inspired by the dense up-sampling layer proposed by Wang *et al*. [1], we propose the the dense down-sampling convolutional (DDC) layer which reduces the computational burden. Third, we introduce the dense lateral connections into the proposed model, which not only facilitate the network training and results in faster convergence. Fourth, extensive experimental results on a standard benchmark demonstrate that the proposed method achieve favorable performance in comparison with the baseline methods as well as the state-of-the-art approaches.

2. RELATED WORKS

As medical image analysis has been extensively studied in the literature. We review several topics pertinent to our approach in this section.

2.1. Image segmentation.

Convolutional neural networks have demonstrated impressive performances in image segmentation, spanning from 2D image segmentation (e.g., object segmentation, semantic segmentation, instance segmentation) to volumetric 3D image segmentation (e.g., cancer tumor segmentation).

2D image segmentation. Long *et al*. [2] propose the Fully Convolutional Networks (FCN) that aims at addressing semantic segmentation. The U-Net presented by Ronneberger *et al*. [3] builds upon the network architecture of the FCN which employs an encoder for feature extraction and a decoder for producing medical image segmentation results. They utilize the skip connection that passes the semantic contents across different network layers to reduce the large information loss in compressing visual contents. In addition, the U-Net utilizes data augmentation techniques, which enables the learned model with good generalization performance. Chen *et al*. [4] develop the Deep Convolutional Nets with atrous convolutional layers and the fully connected conditional random fields (CRFs) to address semantic segmentation. The atrous convolutional layers enables their model to control the resolution of feature responses within deep convolutional networks. While impressive results have been shown by these methods, these models are established based on 2D images.

3D image segmentation. In parallel to 2D image segmentation, several attempts have been made to address the challenging 3D image segmentation in particular for medical image analysis. Çiçek *et al*. [5] extend the idea of U-Net and propose the 3D U-Net that learns dense volumetric segmentation from sparse annotations. The V-Net [6] presented by Milletari *et al*.is end-to-end trainable on MRI volumes depicting prostate. They develop an objective for training their network based on the formulation of Dice coefficient maximization. However, these methods [5, 6] are computationally inefficient making these methods not practical to real-world applications. In light of this, Li *et al*. [7] investigate an efficient and flexible convolutional networks for addressing 3D image segmentation. They employ the dilated convolutional layers and the residual connections and propose a high-resolution and compact convolutional neural network that efficiently and effectively learns 3D representations for volumetric medical data.

Our method. Due to the flexibility and the efficiency, we build our model upon Li *et al*. [7]. However, our model differs from Li *et al*. [7] in three aspects. First, inspired by the dense up-sampling convolutional (DUC) layer proposed by Wang *et al*. [1], we propose the dense down-sampling convolutional (DDC) layer (Figure 4). We introduce the DUC layer and the DDC layer into the proposed model to reduce the computational burden. Second, as employing the dilated convolutional layer in image segmentation often leads to "gridding effect" [1], we integrate the hybrid dilated convolutional (HDC) layer into the proposed model. Third, to facilitate the network training and reduce the information loss, we impose the dense lateral connections that concatenate the low-level feature representations with the high-level semantic contents. As such, the gradients can be propagated efficiently, which results in faster convergence and stable training.

2.2. Imbalanced data distribution.

Imbalanced data distribution is an inevitable issue in computer vision tasks. Several methods are proposed to address the challenge and can be categorized into two aspects: (1) data sampling adjustment and (2) training algorithm adjustment.

Data sampling adjustment Data sampling adjustment is among the most common method adopted in the computer vision task. Pears *et al*. [8] propose the over-sampling method and the under-sampling method, where the probability of the data being sampled depends on its quantity.

Training algorithm adjustment Kulkarni *et al*. [9] develop the incremental learning algorithm that aims at learning from the newly introduced data without forgetting the learned information. Among the recent methods, Chen *et al*. [10] propose the quantity dependent backpropagation and the tree-shaped deep neural network that normalize the gradients with respect to the quantity of each data category and perform classification in and layer-wise fashion to mitigate the effect of imbalanced data distribution.

Our method. In medical image segmentation, the organ (object) of interest often accounts for a very small proportion of volume (Figure 2). Therefore, we simultaneously consider both approaches to specifically handle the large imbalance issue. For data sampling adjustment, we adopt the elastic deformation [3, 6] method. For training algorithm adjustment, we apply the focal loss [11] and the dice loss [6] where both loss terms have demonstrated successful performances in 2D vision tasks.

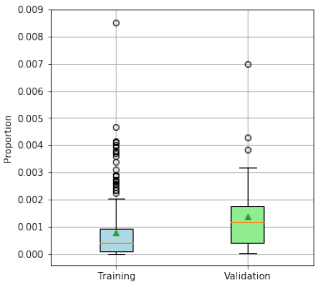

Fig. 2: Label proportion.

3. PROPOSED METHOD

In this section, we first present an overview of our approach. We then describe the proposed framework. Finally, implementation details are provided.

3.1. Algorithmic Overview

Given a 3D image $V \in \mathbb{R}^{h \times w \times d}$ and its corresponding segmentation ground truth $Y \in \mathbb{R}^{h \times w \times d}$, our network predicts a 3D volume $P \in \mathbb{R}^{h \times w \times d}$ where each value $P(i, j, k)$ represents the probability of location (i, j, k) being considered as foreground. To obtain the segmentation result P_t, we binarize the predicted 3D volume P with threshold equals to 0.5.

3.2. The Proposed Models

We apply highres3dnet [7] as our base model since it is shown to perform favorably against V-net [6] and U-net [3] in its experiments. We make the following modifications to improve it and named our model as DenseHighRes3DNet. The model architectures of our two versions DenseHighRes3DNet are provided in Figure 3.

Dense sampling. Due to the large capacity in processing 3D images, we introduce the 3D version of dense up-sampling convolutional (DUC) layer [1] into our model. Inspired by Wang *et al*. [1] who have shown that incorporating DUC layer can restore more information than traditional convolutional layer, we propose the dense down-sampling convolutional (DDC) layer, which can be regarded as DUC's counterpart. The illustration of DUC and DDC are presented in Figure 4. We introduce the DDC layer into our model in substitution of the convolutional layers. More precisely, after incorporating DDC and DUC, the $h_f \times w_f \times d_f$ feature map with output channels c is reshaped to $(h_f/r) \times (w_f/r) \times (d_f/r)$ with output channels $c \times r^3$ and $(h_f \times r) \times (w_f \times r) \times (d_f \times r)$ with output channels c/r^3 respectively.

Gridding effect. Gridding (checkboard) effect often occurs in image generation [12] and medical segmentation [1] tasks. To avoid this problem, we apply hybrid dilated convolution (HDC) proposed by Wang *et al*. [1].

Skip connection Skip connection bypasses feature maps from the early convolution layers to the later ones. Many recent works in medical segmentation [5, 6] also apply this strategy to produce a fine resolution prediction. Since the input data of the competition is high resolution, we also add several dense lateral connections to concatenate low-level and high-level feature maps.

Group normalization (GN) Due to the memory limitation of GPU and the large input volume size, we can only adopt a small number of data in a single batch. However, Batch Normalization (BN) will introduce a higher error with a small batch size as describe in Group Normalization (GN) [13]. Since the computation of GN is independent of batch sizes, and accuracy is stable even with a small batch size campared to BN, we apply GN [13] to substitute the BN in the original network.

3.3. Loss Functions

First we define a notation $p_i \in P$ as the probability volume of the predicted mask, and $y_i \in Y$ as the binary volume of the ground truth mask.

Dice Loss. Since the task is a 3D medical segmentation problem and the data distribution of foreground and background classes is quite imbalance, we apply the version of dice loss introduced by V-Net [6] in our method. Thus, the dice loss can be written as follow

$$L_D = 1 - \frac{2\sum_{i=1}^{N} p_i y_i}{\sum_{i=1}^{N} p_i^2 + \sum_{i=1}^{N} y_i^2} \tag{1}$$

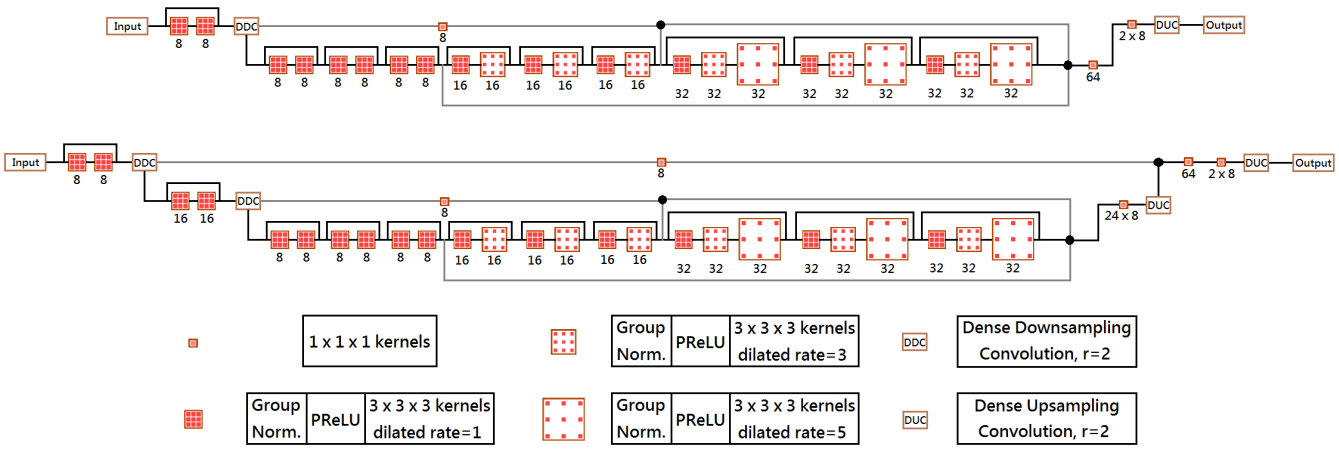

Fig. 3: **The proposed networks. Top**: DenseHighRes3DNet-v1. **Middle**: DenseHighRes3DNet-v2. **Bottom**: illustrations. The number marked under each square block represents the number of the output channels.

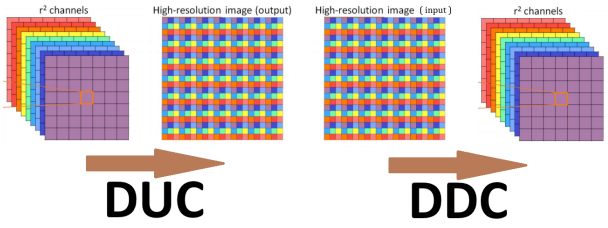

Fig. 4: **The concept figure of DUC and DDC. Left**: DUC proposed by Wang *et al*. [1]. **Right**: our DDC. The figure is derived from [14].

Focal loss. We also apply the focal loss [11] to deal with the class imbalance between foreground and background voxels. For notational simplicity, we fist define a notation q_i as follows

$$q_i = \begin{cases} p_i, & \text{if } y_i = 1 \\ 1 - p_i, & \text{otherwise} \end{cases} \tag{2}$$

Thus, the focal loss can be written as

$$L_{FC}(V) = -\frac{\alpha}{N}\sum_{i=1}^{N}(1-q_i)^{\gamma} log(q_i) \tag{3}$$

where α and γ are two hyperparameters to be decided in the experiments.

3.4. Implementation Details

Our models and baseline model (highres3dnet [7]) [1] are implemented in Tensorflow [15] using NiftyNet [16]. The raw data have the size of $512 \times 512 \times d$ voxels (d is between 75 and 297) and the spatial resolution of $1 \times 1 \times 3\ mm^3$. We first resample them to have the size of $256 \times 256 \times d$ ($2 \times 2 \times 3\ mm^3$) and use random cropping to the size of $192 \times 192 \times 48$. Then we apply elastic deformation as used by U-Net [3] and V-net [6] and spatial scaling with a random scaling factor in the range of [0.9, 1.1]. We use Adam [17] for training with learning rate 0.0003, L2 weight decay 0.0002, batch size 2. We choose α=0.75 and γ=2 in eq 3 for all experiments. We train our models with dice loss for 20k iterations and then focal loss for 10k iterations, as we call it a "round". We found out that this "round" by "round" setting can improve training stability. [2] We use one NVIDIA Tesla V100 with 16GB memory to train our models for two rounds, which takes about 48 hours. All the settings are the same when training the baseline model. When testing, our models generate a $256 \times 256 \times d$ mask. We resize it to $512 \times 512 \times d$ and then compare to its corresponding ground truth.

4. EXPERIMENTS

We report the results of the proposed model as well as the comparisons with the baseline model in this section.

4.1. Datasets

The adopted dataset contains images depicting non-small cell lung cancer (NSCLC) subjects selected from NSCLC-Radiomics dataset [18, 19] provided by The Cancer Imaging Archive (TCIA) [20]. The training set consists of pre-treatment CT scans from 260 subjects while the validation set comprises 40 subjects. The test set is composed of 40 subjects without the manual annotation. Both the training set and the validation are manually delineated by a radiation oncologist.

[1] Because the input volumes are too large, we modify highres3dnet_small on NiftyNet website to have nearly the same parameters as our models.

[2] It is interesting to note that if we train our models with focal loss for more than 10k iterations, the performance degrades severely. So far we still can't explain this phenomenon.

Table 1: Experimental results for training set. For Dice2D, we compute the slices belonging to True Positives and False Negatives.

Model	Iterations	Dice3D				Dice2D			
		Mean	Q1	Q2	Q3	Mean	Q1	Q2	Q3
Baseline	20k	0.601	0.516	0.722	0.807	0.552	0.123	0.717	0.855
	30k	0.626	0.545	0.731	0.812	0.618	0.480	0.758	0.851
	50k	0.691	0.667	0.782	0.845	0.618	0.412	0.789	0.882
	60k	0.724	0.700	0.809	0.853	0.658	0.550	0.813	0.883
DenseHighRes3DNet-v1	20k	0.557	0.398	0.661	0.761	0.563	0.227	0.725	0.849
	30k	0.616	0.524	0.726	0.810	0.618	0.459	0.764	0.863
	50k	0.631	0.520	0.720	0.804	0.648	0.527	0.793	0.875
	60k	0.680	0.645	0.755	0.820	0.675	0.600	0.796	0.873
DenseHighRes3DNet-v2	20k	0.481	0.262	0.549	0.694	0.590	0.342	0.744	0.854
	30k	0.589	0.491	0.693	0.772	0.596	0.404	0.744	0.846
	50k	0.617	0.497	0.714	0.806	0.637	0.500	0.781	0.870
	60k	0.689	0.650	0.780	0.834	0.658	0.561	0.792	0.872

Table 2: Experimental results for validation set. The definition of Dice2D is the same as Table 1

Model	Iterations	Dice3D				Dice2D			
		Mean	Q1	Q2	Q3	Mean	Q1	Q2	Q3
Baseline	20k	0.491	0.174	0.550	0.748	0.421	0.000	0.510	0.793
	30k	0.510	0.309	0.586	0.747	0.475	0.000	0.611	0.815
	50k	0.493	0.241	0.538	0.780	0.422	0.000	0.494	0.807
	60k	0.508	0.324	0.623	0.747	0.437	0.000	0.511	0.826
DenseHighRes3DNet-v1	20k	0.456	0.268	0.568	0.630	0.402	0.000	0.477	0.758
	30k	0.487	0.286	0.546	0.724	0.465	0.000	0.597	0.800
	50k	0.492	0.287	0.572	0.712	0.438	0.000	0.542	0.789
	60k	**0.530**	0.367	0.597	0.751	**0.510**	0.007	0.643	0.825
DenseHighRes3DNet-v2	20k	0.451	0.271	0.514	0.662	0.452	0.000	0.569	0.792
	30k	0.515	0.383	0.589	0.724	0.472	0.000	0.610	0.811
	50k	0.490	0.333	0.562	0.673	0.478	0.000	0.630	0.797
	60k	0.506	0.342	0.587	0.757	0.446	0.000	0.563	0.790

4.2. Evaluation Protocol

We adopt three evaluation metrics including the *dice coefficient*, the *mean surface distance*, and the *Hausdorff distance* to evaluate the effectiveness of the proposed method.

Dice coefficient (D). The dice score measures the similarity between two sets and is defined by

$$D = \frac{2|P_t \cap Y|}{|P_t| + |Y|}, \tag{4}$$

where the $|P_t|$ and $|Y|$ represent the cardinality of P_t and Y, respectively.

Surface distance (SD). For the following two metrics we use, we first introduce the surface distance as the basic function. The surface distance metrics estimate the error between the outer surfaces S and S' of the masks Y and P_t. The distance between a point s on surface S and the surface S' is given by the minimum of the Euclidean norm.

$$\vec{d}(s, S') = \min_{s' \in S'} ||s - s'||_2 \tag{5}$$

Where $s \in S$ and $s' \in S'$. and apply this to all points in S gives the total surface distance $\vec{d}(S, S')$.

Mean surface distance(MSD). The directed mean surface distance is the mean of the surface distance between S and S'.

$$\vec{d}_{MSD}(S, S') = \frac{1}{|S|} \sum_{s \in S} \vec{d}(s, S') \tag{6}$$

Then, the (undirected) mean surface distance is the mean of the two directed mean surface distance.

$$d_{MSD}(S, S') = \frac{\vec{d}_{MSD}(S, S') + \vec{d}_{MSD}(S', S)}{2} \tag{7}$$

Hausdorff distance (HD). The directed percent Hausdorff distance, for a percentile r, is the rth percentile of the total surface distance. Denoting the rth percentile as K_r.

$$\vec{d}_{HD}(S, S') = K_r(\vec{d}(s, S')) \forall s \in S \tag{8}$$

The (undirected) percent Hausdorff distance is defined with the mean of two directed percent Hausdorff distance.

$$d_{HD}(S, S') = \frac{\vec{d}_{HD}(S, S') + \vec{d}_{HD}(S', S)}{2} \tag{9}$$

In this work, we take $r = 95\%$ to measure it.

Table 3: Experimental results for validation set using three evalution criteria. FN and FP means the number of False Negatives and False Positives respectively.

Method	Iteration	D	MSD	95%HD	FN	FP
Baseline	30k	0.599	12.538	24.932	176	314
DenseHighRes3DNet-v1	60k	0.640	9.178	18.620	**173**	363
DenseHighRes3DNet-v2	30k	0.619	9.966	19.759	202	531
Ensemble	-	**0.662**	**8.440**	**17.483**	195	**287**

4.3. Experimental Results

The quantitative results are presented in Table 1 and Table 2. It can be observed that, at the first round of training, the baseline model performs slightly better than the DenseHighRes3DNet-v1. After the second round, however, the DenseHighRes3DNet-v1 model outperforms the baseline model by **2.2** in 3D dice coefficient and **7.3** in 2D dice coefficient. This indicates that the DenseHighRes3DNet-v1 model is effective. On the other hand, we observe that the DenseHighRes3DNet-v2 model is worse than the DenseHighRes3DNet-v1 model but still outperforms the baseline model by **0.9** in 2D dice coefficient. The effectiveness of incorporating the focal loss can be observed from the growth in the first quartile (Q1) from 20k to 30k and 50k to 60k iterations. Additionally, we select the best version of these models based on Dice3D and ensemble them as our final model. We compute their three criteria (dice coefficient D, mean surface distance MSD and 95% Hausdorff distance 95%HD) as mentioned in Sec. 4.2. The results are shown in Table 3. It's worth noting that we use the slices belonging to True Positives when computing these criteria because MSD and 95%HD are undefined excepct True Positives.

5. CONCLUSIONS

We have presented an end-to-end trainable network that learns to segment the lung cancer tumor in a supervised fashion. The core technical novelty lies in the proposed dense down-sampling convolutional (DDC) layer that significantly reduces the heavy burden of processing volumetric data. To facilitate the network training, we introduce the proposed dense down-sampling convolutional layer (DDC) and the dense up-sampling convolutional (DUC) layer into the proposed model, which results in faster convergence and more stable training. We apply the dice loss and the focal loss to handle the inevitable imbalanced data distribution issue in segmenting the region (object) of interest. Extensive experimental evaluations demonstrate the effectiveness of the proposed method. Moving forward, we hope our model could be a fundamental and generic method in addressing imbalanced data issue as well as accelerating 3D image segmentation.

6. REFERENCES

[1] Panqu Wang, Pengfei Chen, Ye Yuan, Ding Liu, Zehua Huang, Xiaodi Hou, and Garrison Cottrell, "Understanding convolution for semantic segmentation," *arXiv*, 2017.

[2] Jonathan Long, Evan Shelhamer, and Trevor Darrell, "Fully convolutional networks for semantic segmentation," in *CVPR*, 2015.

[3] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *MICCAI*, 2015.

[4] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *TPAMI*, 2018.

[5] Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger, "3d u-net: learning dense volumetric segmentation from sparse annotation," in *MICCAI*, 2016.

[6] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *3DV*, 2016.

[7] Wenqi Li, Guotai Wang, Lucas Fidon, Sebastien Ourselin, M Jorge Cardoso, and Tom Vercauteren, "On the compactness, efficiency, and representation of 3d convolutional networks: brain parcellation as a pretext task," in *IPMI*, 2017.

[8] Russel Pears, Jacqui Finlay, and Andy M Connor, "Synthetic minority over-sampling technique (smote) for predicting software build outcomes," in arXiv preprint arXiv:1407.2330, 2014.

[9] Pallavi Kulkarni and Roshani Ade, "Incremental learning from unbalanced data with concept class, concept drift and missing features: a review," in International Journal of Data Mining & Knowledge Management Process, vol. 4, no. 6, pp. 15, 2014.

[10] Yun-Chun Chen, Yu-Jhe Li, Aragorn Tseng, and Tsungnan Lin, "Deep learning for malicious flow detection," in *Personal, Indoor, and Mobile Radio Communications (PIMRC), 2017 IEEE 28th Annual International Symposium on*. IEEE, 2017, pp. 1–7.

[11] Tsung-Yi Lin, Priyal Goyal, Ross Girshick, Kaiming He, and Piotr Dollár, "Focal loss for dense object detection," *TPAMI*, 2018.

[12] Augustus Odena, Vincent Dumoulin, and Chris Olah, "Deconvolution and checkerboard artifacts," *Distill*, 2016.

[13] Yuxin Wu and Kaiming He, "Group normalization," in *ECCV*, 2018.

[14] Wenzhe Shi, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew P Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *CVPR*, 2016.

[15] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al., "Tensorflow: a system for large-scale machine learning.," in *OSDI*, 2016.

[16] Eli Gibson, Wenqi Li, Carole Sudre, Lucas Fidon, Dzoshkun Shakir, Guotai Wang, Zach Eaton-Rosen, Robert Gray, Tom Doel, Yipeng Hu, Tom Whyntie, Parashkev Nachev, Dean C. Barratt, Sebastien Ourselin, M. Jorge Cardoso, and Tom Vercauteren, "Niftynet: a deep-learning platform for medical imaging," *CMPB*, 2018.

[17] Diederik P Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," *arXiv*, 2014.

[18] HJWL Aerts, Emmanuel Rios Velazquez, Ralph TH Leijenaar, Chintan Parmar, Patrick Grossmann, Sara Carvalho, and Philippe Lambin, "Data from nsclc-radiomics. the cancer imaging archive," 2015.

[19] Hugo JWL Aerts, Emmanuel Rios Velazquez, Ralph TH Leijenaar, Chintan Parmar, Patrick Grossmann, Sara Carvalho, Johan Bussink, René Monshouwer, Benjamin Haibe-Kains, Derek Rietveld, et al., "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach," *Nature communications*, 2014.

[20] "The cancer imaging archive (tcia)," http://www.cancerimagingarchive.net/.