Imperial College London

IMPERIAL COLLEGE LONDON

DEPARTMENT OF COMPUTING

Background and Progress Report

Author:
Tianyang Sun

Supervisor:
Name of supervisor

September 2015

Contents

1 Task description **1**
1.1 Motivation . . . 1
1.2 Feasibility . . . 2
1.3 Challenge . . . 2

2 Background **3**
2.1 Medical Segmentation . . . 3
2.1.1 Unet & Friends . . . 3
2.1.2 Loss functions for medical segmentation . . . 3
2.2 Data augmentation . . . 5
2.2.1 Traditional Data Augmentation . . . 5
2.2.2 Learning for augmentation . . . 6
2.3 Few shot learning in non-medical domain . . . 6
2.4 Few-shot learning in medical application . . . 7
2.4.1 Transfer learning method . . . 7
2.4.2 Special network design . . . 8

3 Progress Section **10**
3.1 Medical Segmentation data . . . 10
3.2 Data preprocessing . . . 10
3.3 Basic result . . . 10
3.4 Brief plan . . . 10

Chapter 1

Task description

1.1 Motivation

Computed Tomography(CT) is widely used for diagnosing Covid-19 outbreak and tremendous studies[1] reported deep learning can potentially provide fast and accurate image analysis while most of them train on massive dataset with more than 1000 annotated CT scans. However, until today (May 29^{th}, 2020) non of those huge dataset is available for public research.

To compensate the lack of annotated dataset on Covid-19 Lung CT scans, researchers developed a small Covid-19 CT infection segmentation benchmark [1][2] consists of 20 annotated CT scans available to public. Meanwhile they also proposed three segmentation tasks:

1. Learning with limited annotations

2. Lerning to segment Covid-19 CT from non Covid-19 CT scans

3. Learning with both Covid-19 and non Covid-19 CT scans

Task 1 and 3 are few shot segmentation tasks because only 4 volumes are allowed for training and the rest are for testing to report a 5-fold cross validation result according to the proposal [2]

We aims to explore several few shot segmentation methods in medical domain in this project, specifically 3D CT images. One of the target is to provide solutions to the few shot segmentation challenge proposed in the COVID19 segmentation benchmark. Given this is a relatively small dataset and might not present good evaluation of robustness analysis, we also plan to evaluate the performance on other lung CT segmentation datasets.

---

[1] https://github.com/HzFu/COVID19_imaging_AI_paper_list#technical_CT
[2] http://medicalsegmentation.com/covid19/

Subtask	Lung				Infection	
	Left Lung		Right Lung		NSD	
	DSC	NSD	DSC	NSD	DSC	7021.3
Fold-0	84.98 .2	68.713 .3	85.213 .0	70.615 .8	68.120 .5	70.923 .0
Fold-1	80.314 .5	61.815 .1	83.99 .6	68.39 .0	71.320 .5	71.823 .0
Fold-2	87.112 .1	74.316 .0	90.38 .2	78.512 .0	66.221 .7	71.724 .2
Fold-3	88.47 .0	75.28 .8	89.96 .3	78.58 .0	68.123 .1	70.827 .1
Fold-4	88.376 .0	75.811 .0	90.27 .0	78.310 .2	62.726 .9	64.928 .2
Avg	**85.810.5**	**71.213.8**	**87.99.3**	**74.811.9**	**67.322.3**	**70.024.4**

Table 1.1: Dice score and Normalised Surface Distance reported in work [2]

1.2 Feasibility

We first explain the feasibility. The proposer of the benchmark have done some of the basic training for task 1 with 3D Unet model and reported acceptable performance (see table 1.1). Based on the reported Dice score and Surface distance, we assume this few-shot segmentation problem is feasible, and decide to investigate further into few-shot segmentation in medical domain.

1.3 Challenge

In non-medical domain, both few shot classification and few shot segmentation has been explored and provide good performance on some of the segmentation benchmarks. In medical application, however, few-shot learning mainly focus on classification tasks while dense segmentation does not provide promising results as far as we know.

We so far believe the task is doable based on the previous analysis. We aim to implement methods that provide comparable performance compared to U-net and V-net, which are good models on small sample medical image segmentation. The task however might not provide satisfactory results given the current dataset. In the ”future work” section in the review of current deep learning methods related to COVID19 in paper [3], the author mentioned the explainability of network prediction. We might use class activation maps followed his suggestion to see why the the implemented method give the result, in both good and less desirable cases **if time permits**.

Chapter 2

Background

2.1 Medical Segmentation

2.1.1 Unet & Friends

This section we introduce several well known methods in medical segmentation. Unet [4] and its variations plays a dominant role in current medical segmentation tasks, and it is often used as a baseline model for performance evaluation in the literature.

Unet [4] is among one of the most widely used medical segmentation models since the day it was proposed. The original Unet consists of a contracting path followed by an expansive path that gives a "U" shaped architecture. The network architecture is shown in figure 2.1. Later this 2D model was extended to 3D version in [5] for for Kidney segmentation tasks so that the model learn features from information implicated between slices.

VNet [6] is another 3D variation of Unet, and the network performed evaluation on prostate dataset. Each block of convolution has a residual feature that the input of the block is added to the last convolutional layer. The author argued that leverage residual structure enables network convergence in a fraction of the amount of time other network used.

2.1.2 Loss functions for medical segmentation

Loss function or objective function is a crutial component in neural network training. Segmentation tasks usually make use of *Distribution loss*, *Region based loss* and *boundary-based* loss for training and evaluation of segmentation performance. Recent work in [7] summarized some common loss functions for segmentation.

Cross entropy (CE) measures the dissimilarity between the learned distribution and

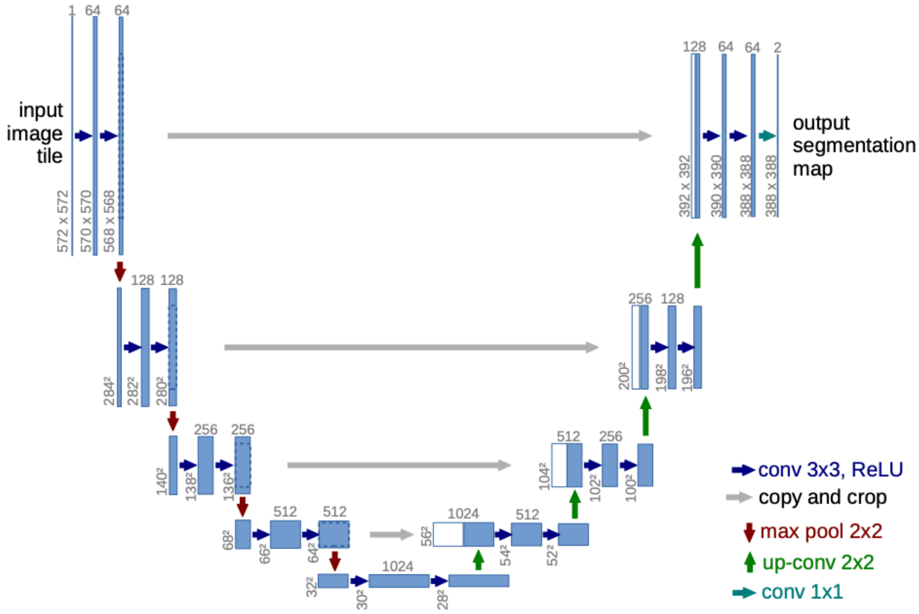

Figure 2.1: Original Unet architecture in [4]

target distribution.

$$CE = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} w_c \cdot (y_i^c \log p_i^c)$$

where y_i^c indicates the prediction result (correct or wrong) and p_i^c denotes the predicted probability of pixel i for class c, w_c is now 1 for original cross entropy loss. Unet [4] training extend the CE by adding weight w_c. A common example for weight measurement is through the inverse proportion of observed class frequency. This modification potentially deal with imbalance class which is very common in medical domain.

Dice loss is a region based loss function that learn to optimize the Dice Coefficient (D). Vnet [6] first brought the Dice loss into machine vision community to solve the problem of highly biased prediction towards dominant area (e.g background) in medical segmentation.

$$D = \frac{2\sum_i^N p_i g_i}{\sum_i^N p_i^2 + \sum_i^N g_i^2}$$

Assume we segment N samples, p_i denotes the prediction volume and g_i denotes the ground-truth volume. Dice loss usually require the label to be one hot encoded during training. One benefit is that Dice loss does not requires class balance methods such as weighting method in CE loss.

Hausdorff Distance loss(HD) aims to minimize the boundary distance between prediction and ground-truth segmentation masks. Similar to Dice loss, it also alleviate the class imbalance issue during training. However, paper [] showed that directly minimizing Hausdorff Distance is intractable while an approximation (HD

Loss) through distance transform (gray-level intensities of points inside foreground regions are changed to show the distance to the closest boundary from each point [1]).

$$L_{HD_{DT}} = \frac{1}{N}\sum_{i=1}^{N}\left[(s_i - g_i)\cdot\left(d_{Gi}^2 + d_{Si}^2\right)\right]$$

Paper [7] proposed that so far none of the papers in the literature provide a comprehensive comparison of the loss functions for segmentation task. Selecting loss function is still based on empirical comparison. Several works [] used compound loss function that combined several loss together as training objectives overall gives good performance compared to individual loss functions.

2.2 Data augmentation

In medical domain, huge dataset consists of large numbers of carefully labelled samples is rarely available due to heavy workload for annotations, rarity of disease, ethic issues of data acquire process and data privacy. Furthermore, different data acquire protocols (i.e. CT machines used by different hospitals) brings difficulties to clinical practice for good accuracy of existing pre-trained models. As a result, few shot learning and/or few shot segmentation has been explored in recent years. In this section, we focus on a few approaches that has been used in current literature which aim to explore the potential of existing training samples through various augmentation methods to alleviate the insufficient training samples in medical imaging. We discuss data augmentation method as well as the amount of data used in each work. Table **??** provide an overview of each method.

2.2.1 Traditional Data Augmentation

Traditional Data augmentation method in imaging domain refers to the process that does not require such training data to learn a transformation.

[8] investigated data augmentation methods under 3D medical domain of MR and ultrasound images. The data augmentation process consists of a sequence of traditional transformation techniques. The paper argued that sharpness in medical images during training process limits model generalization thus applying gaussian filter to images take noise into consideration. Brightness and contrast difference caused by variations in scanning protocols brings potential domain shift thus a sequenced random shift followed gamma correction and random linear transform in intensity are reasonable data augmentation methods. Finally spatial transformations including rotation, scaling and deformation is added to the augmentation process. The source domain in this method is Prostate dataset [2] consists 48 4D volumes. We argue here that the stacked transformation is physical transformation process independent to the size of dataset because no learning or training process is required in

[1] https://homepages.inf.ed.ac.uk/rbf/HIPR2/distance.htm
[2] http://medicaldecathlon.com/index.html#tasks

this augmentation method thus might bring benefits to our task. However we doubt the improvement due to the difference between CT and MR images.

Another method "mixup" by [9] based on generic vicinal distribution, which generates new samples through interpolation between two existing data. The author argued that this method works as a regularizer which encourages linear behavior between training samples. In terms of imaging, the augmentation is applied to CIFAR 10 (2D non medical) Dataset. The work was originally implemented on training GANs. Later in medical domain, [10] further investigated mixup method on Knee MR images on OIA database [3] that consists of 88 3D MRI scans. The work showed mixup improves generalization under their experiment setup while having risk of slight underfitting due to to the strong regularization. The author further mentioned not using weight decay in the experiment solve the underfitting issue. [11] summarized mixup augmentation method gives "soft labels". Variations of the mixup method utilize asymmetric is further explored in [12] trained on Brain MR images, and the method reports huge gain under their experiment setup. We cannot guarantee the effectiveness of this augmentation method due to the domain difference between MRI and CT scans.

2.2.2 Learning for augmentation

Following the traditional augmentation methods, we here want to discuss a few medical image augmentation based on deep learning. Specifically we focus on those who used one or a few data during the augmentation which is close to the problem we are facing.

Adversarial defense was deployed in [13] for the augmentation of small samples in Brain MR segmentation. Adversarial samples generated though Fast Gradient Sign Method[] and then added as training data to improve the robustness, which according the study in [] without decreasing the performance. So far we believe this method worth trying because the experiment was trained on 7 brain volumes, which is close to our task.

Paper	Method	Dataset	Number of samples
[8]	Stacked traditional transform	Prostate dataset	48 4D volumes
[9]	mixup	CIFAR 10	Huge
[10]	mixup	Knee MR images	88 3D MRI
[12]	Asymmetric mixup	BraTS ?	
[13]	Adversarial defense	MRBrainS18 challenge	7 train, 14 test

2.3 Few shot learning in non-medical domain

Recent work in non-medical few shot segmentation have explored several network structures that tend to provide correct segmentation for few labelled samples from

[3] http://www.oai.ucsf/.edu/

dense to sparse annotations.

Although few-shot learning for classification has been wide explored, and segmentation, in some sense, can be viewed as classification task, a direct extension from those networks might not give good results.

Experiment of extending Siamese Network in meta learning classification to pixel-wise segmentation failed to scale well enough, but this idea inspire the work in [14] that proposed a two-branch network for few-shot semantic segmentation. The work assumes that semantic class labels for training and testing have no overlap, that is $L_{train} \cap L_{test} = \varnothing$. Large dataset updates a pre-trained VGG network on ImageNet in the ”conditioning branch”, and a query sample updates the ”segmentation branch” including a FCN-32 and a conv-fc7. The authors further utilize masking and weight hashing to combine the two branches. The experiment reported promising results in non medical dataset (PASCAL-5) the work further removed the effect of ImageNet and still show good performance. However, in our CT segmentation task, we might want to take 3D features into consideration.

Another work in [15] relax the restriction of requiring dense segmentation masks to sparse annotations with similar backbone of VGG net. They utilize the guidance network that provide guiding inference to extract visual features through an encoder structure. The work is evaluated on both 2D images and video sets while the structure used in video sets.

2.4 Few-shot learning in medical application

2.4.1 Transfer learning method

In medical domain, few shot learning mainly focus on transfer learning from pre-trained networks that leverage both medical and non-medical datasets.

In Lung CT segmentation area, Sports-1M dataset has been used as source domain to train a multi-task learning model for nodule malignancy prediction and rating [16]. The author reported significant improvement in the prediction accuracy, however, did not mention the proportion of data used for transfer training.

People tend to choose datasets from closer domain for transfer learning. It is reasonable to consider methods that transfer across disease in the same structure under the same modality. In our case, we might want to investigate transfer learning from NSCLC Dataset to Covid segmentation set given that both of them are lung CT scans.

Recent work explored several across disease transfer learning training techniques under MRI domain [17]. The paper evaluated three transfer learning methods trained on 3D U-Net by Fine tuning the last three layers, Fine tuning the decoder and Fine tuning all model parameters. The Source Dataset: Multiple Sclerosis Dataset consists 3630 MRI volumes and used Brain Tumor Dataset as Target dataset including 210 high-grade glioma (HGG) and 75 low-grade glioma (LGG) Brain MRI scans. The training target is a decaying weighted categorical cross entropy loss weighted by relative voxel. Their best validation performance of pre-trained network achieved validation performance AUC 0.77. Experiment result on 20, 50, 100 and 150 samples during Fine tuning respectively showed that Fine tuning all parameters out performed the rest methods in most cases.

One potential drawback is that compared to the our task, the target training set is relatively larger, the performance is expected to be less ideal when using ”fine tune all” method using 4 or less volumes in our case.

Paper [13] trained segmentation of Brain MR image on 7 brain volumes after Adversarial defense augmentation. The work first split the segmentation from easy to hard into 2 individual classifiers then joint learning with dense pixel segmentation. The author reported that the result outperformed Unet and Vnet method trained from scratch. We argue that the augmentation provide good result in the segmentation accuracy while the segmentation part is not well explained in the paper. We have emailed the author for further details.

2.4.2 Special network design

Transfer learning methods usually require small samples to update millions of parameters that take the risk of overfitting [14]. The design of multiple branch network, usually includes conditioner arm and segmentation arms inspired the deep learning in medical domain to go beyond transfer learning while encourage a stronger between-arms interaction to compensate the lacking in pre-trained model [18]. The proposed method perform 3D volume segmentation at test time while use 2D images during training. However the method requires start and end slice to be indicated for each query volume and still not achieving good dice score.

Paper	Method	Domain Details	Task
[14]	Design Conditional Branch	Target PASCAL-5	Few shot segmentation
[15]	Guidance network	Target PASCAL VOC	Few shot segmentation
[16]	Non medical to medical transfer	Source: Sports-1M; Target: Lung nodule	Multi-task learning: prediction and rating
[17]	Across domain transfer	Source: MSD; Target: Brain Tumor Dataset	Segmentation
[13]	Augmentation+pixel dense segmentation	Only trained on 7 brain Volumes	Dense segmentation
[14]	Branch network design	–	Segmentation

Table 2.1: Small sample methods in medical and non-medical domain

Chapter 3

Progress Section

3.1 Medical Segmentation data

Due to restrictions in medical data, one of the problem we faced is finding some potentially useful segmentation dataset. We found it reasonable to start with same domain dataset so we focus the dataset on Lung CT volumes.

- NSCLC (Non small cell Lung Cancer) Dataset consists of 402 CT scans where 78 cases has Pleural Effusion(PE). Besides the PE label, left and right lung mask is provided for segmentation training.
- MSD Tumor segmentation dataset consists of 63 labelled volumes non small cell lung cancer infected area. Unfortunately lung area are not labeled.
- Strucseg Dataset consists of 50 lung cancer patients of Lung cancer. Both organ segmentation and gross target segmentation data is annotated. However, the dataset require some documentation work and we have signed the data agreement to their email, and we cannot guarantee they will reply.

3.2 Data preprocessing

3.3 Basic result

3.4 Brief plan

Bibliography

[1] M. Jun, G. Cheng, W. Yixin, A. Xingle, G. Jiantao, Y. Ziqi, Z. Minqing, L. Xin, D. Xueyuan, C. Shucheng, W. Hao, M. Sen, Y. Xiaoyu, N. Ziwei, L. Chen, T. Lu, Z. Yuntao, Z. Qiongjie, D. Guoqiang, and H. Jian, "COVID-19 CT Lung and Infection Segmentation Dataset," Apr. 2020. [Online]. Available: https://doi.org/10.5281/zenodo.3757476 pages

[2] M. Jun, W. Yixin, A. Xingle, G. Cheng, Y. Ziqi, C. Jianan, Z. Qiongjie, D. Guoqiang, H. Jian, H. Zhiqiang, N. Ziwei, and Y. Xiaoping, "Towards efficient covid-19 ct annotation: A benchmark for lung and infection segmentation," *arXiv preprint arXiv:2004.12537*, 2020. pages

[3] F. Shi, J. Wang, J. Shi, Z. Wu, Q. Wang, Z. Tang, K. He, Y. Shi, and D. Shen, "Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for COVID-19," pp. 1–1. [Online]. Available: https://ieeexplore.ieee.org/document/9069255/ pages

[4] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation." [Online]. Available: http://arxiv.org/abs/1505.04597 pages

[5] . iek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3d u-net: Learning dense volumetric segmentation from sparse annotation," in *Medical Image Computing and Computer-Assisted Intervention MICCAI 2016*, S. Ourselin, L. Joskowicz, M. R. Sabuncu, G. Unal, and W. Wells, Eds. Springer International Publishing, vol. 9901, pp. 424–432, series Title: Lecture Notes in Computer Science. [Online]. Available: http://link.springer.com/10.1007/978-3-319-46723-8_49 pages

[6] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation." [Online]. Available: http://arxiv.org/abs/1606.04797 pages

[7] J. Ma, "Segmentation loss odyssey." [Online]. Available: http://arxiv.org/abs/2005.13449 pages

[8] L. Zhang, X. Wang, D. Yang, T. Sanford, S. Harmon, B. Turkbey, H. Roth, A. Myronenko, D. Xu, and Z. Xu, "When unseen domain generalization is unnecessary? rethinking data augmentation." [Online]. Available: http://arxiv.org/abs/1906.03347 pages

[9] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "mixup: Beyond empirical risk minimization." [Online]. Available: http://arxiv.org/abs/1710.09412 pages

[10] E. Panfilov, A. Tiulpin, S. Klein, M. T. Nieminen, and S. Saarakkala, "Improving robustness of deep learning based knee MRI segmentation: Mixup and adversarial domain adaptation," in *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*. IEEE, pp. 450–459. [Online]. Available: https://ieeexplore.ieee.org/document/9022164/ pages

[11] N. Tajbakhsh, L. Jeyaseelan, Q. Li, J. N. Chiang, Z. Wu, and X. Ding, "Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation," vol. 63, p. 101693. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S136184152030058X pages

[12] Z. Li, K. Kamnitsas, and B. Glocker, "Overfitting of neural nets under class imbalance: Analysis and improvements for segmentation." [Online]. Available: http://arxiv.org/abs/1907.10982 pages

[13] X. Ren, L. Zhang, D. Wei, D. Shen, and Q. Wang, "Brain MR image segmentation in small dataset with adversarial defense and task reorganization," in *Machine Learning in Medical Imaging*, H.-I. Suk, M. Liu, P. Yan, and C. Lian, Eds. Springer International Publishing, vol. 11861, pp. 1–8, series Title: Lecture Notes in Computer Science. [Online]. Available: http://link.springer.com/10.1007/978-3-030-32692-0_1 pages

[14] A. Shaban, S. Bansal, Z. Liu, I. Essa, and B. Boots, "One-shot learning for semantic segmentation," in *Procedings of the British Machine Vision Conference 2017*. British Machine Vision Association, p. 167. [Online]. Available: http://www.bmva.org/bmvc/2017/papers/paper167/index.html pages

[15] K. Rakelly, E. Shelhamer, T. Darrell, A. Efros, and S. Levine, "Few-shot segmentation propagation with guided networks," p. 10. pages

[16] S. Hussein, K. Cao, Q. Song, and U. Bagci, "Risk stratification of lung nodules using 3d CNN-based multi-task learning," vol. 10265, pp. 249–260. [Online]. Available: http://arxiv.org/abs/1704.08797 pages

[17] B. Kaur, P. Lematre, R. Mehta, N. M. Sepahvand, D. Precup, D. Arnold, and T. Arbel, "Improving pathological structure segmentation via transfer learning across diseases," in *Domain Adaptation and Representation Transfer and Medical Image Learning with Less Labels and Imperfect Data*, Q. Wang, F. Milletari, H. V. Nguyen, S. Albarqouni, M. J. Cardoso, N. Rieke, Z. Xu, K. Kamnitsas, V. Patel, B. Roysam, S. Jiang, K. Zhou, K. Luu, and N. Le, Eds. Springer International Publishing, vol. 11795, pp. 90–98, series Title: Lecture Notes in Computer Science. [Online]. Available: http://link.springer.com/10.1007/978-3-030-33391-1_11 pages

[18] A. G. Roy, S. Siddiqui, S. Plsterl, N. Navab, and C. Wachinger, "'squeeze & excite' guided few-shot segmentation of volumetric images." [Online]. Available: http://arxiv.org/abs/1902.01314 pages