Vision Transformer for Biomedical Image Segmentation

Bedionita Soro (20205677), Ying Hui Tan (20204871), Oh Joon Kwon (20213044)

Abstract

Segmentation is crucial in biomedical images analysis for detecting diseases such as breast cancer and pulmonary diseases. However, manually detecting a specific cell or organ can be laborious and difficult because it requires professional medical knowledge. Moreover, the complexity of images (artifacts, other cells or organs) can add to errors that can be potentially harmful in medical practices. There have been numerous attempts to automate this procedure, but most recently, deep learning based automatic segmentation methods such as U-Net and its variants have gained attention. Nevertheless, it requires a large amount of training data, and histological images contains much noise that degrades the performance. In this paper, we exploit recent advanced methods in deep learning and propose a multi domain approach that can semantically segment histology images for breast cancer detection and extend it to lesion segmentation in COVID-19 CT scan, exploiting domain adaptation techniques. The proposed method makes use of the recent state-of-the-art computer vision methods such as Vision Transformer (ViT).

1. Introduction

Semantic segmentation is crucial in biomedical image analytics, yet requires heavy workload for manual annotation. Segmentation task can become very difficult for the pathologists even with high-expertise due to other organs, noise and artifacts[10]. Moreover, human annotators tend to be fatigued because they often have to screen all the whole slide images (WSI) to find regions of interest with abnormal tissues or organs [13]. Recently, several semantic segmentation methods exploiting deep learning have been developed for histo-pathological images or organ segmentation [11]. However, most of these existing methods not only focus on a specific organ or tissue [2], but they have not largely investigated domain adaptation techniques with organs from different domains.

In this paper, we propose a semantic segmentation method with Transformer architecture [15] to semantically segment and identify different type of cancer regions in the WSI for breast cancer detection. We then extend this approach to segment COVID-19's lesion in lung [16] using domain adaptation. This allows us a model to generalize over totally different domains with less effort.

1.1. Problem and Contributions

Organ segmentation is crucial for disease detection and treatment. Breast cancer and COVID-19 are among the common diseases that can be helped with automatic segmentation. Even though there exists some academic works on classifying cancer diseases in the WSI of breasts, those models perform below par compared to deep learning methods in computer vision. On the other hand, it is very challenging to build a model specific to that new disease due to the lack of data for COVID-19. Therefore, there is a need for domain adaptation in medical image segmentation.

The first objective of this study is to propose a deep neural network model that will efficiently and accurately identify different regions of interest in the WSI of breast for different type of cancer diseases detection. The second objective is to exploit the knowledge obtained on the breast cancer dataset to perform COVID-19 lesion segmentation through domain adaptation techniques. The performance of the proposed method will be compared with other state-of-the-art models [2, 3, 7, 9] on publicly available datasets. The contributions of this work can be summarized as follows:

- We propose a semantic segmentation method for breast cancer to identify type of regions of interest,
- We also investigate the problem COVID-19's lung lesion segmentation which is a hot topic with less available data. In order to address the lack of data, domain adaptation will be explored as an option, adapting histological images to the COVID-19 patient CT scans.
- The proposed method exploits the advantage of Vision Transformer [6], which makes this the first work that uses Transformer in histological images segmentation for breast cancer detection,
- This approach will reduce the burden of manually detecting the regions of interest in histopathology images. It will provide a way to mitigate the problem of limited data in images segmentation for new diseases.

- We expect to compare the proposed method against some recent state-of-art methods [6, 14].

- This work aims to contribute to the development of automatic healthcare systems that is becoming essential with the ongoing pandemic.

2. Method

The proposed method integrates recent state-of-the-art deep learning model in image classification known as Vision Transformer (ViT) [6]. Transformer, proposed by Vaswani et al.[15], makes use of multi-head self-attention mechanism that can easily be scaled with large datasets. While the widely used methods in image segmentation uses convolutional neural networks (CNN) as their building blocks, parameter sharing nature of CNNs often introduces inductive bias when adapted to different data domains [4]. This leads to degraded model performance in inexperienced domains.

Pretrained Transformer models have been shown to be flexible when fine-tuned for down-stream tasks, both in natural language processing and computer vision [5, 12]. While it can be computationally expensive to pretrain the models on large datasets, Transformers often excel at adapting to different domains which makes it ideal for transfer learning and possibly domain adaptation. Moreover, the nature of semantic segmentation task requires the contextual understanding of the relationship among image regions. The self-attention mechanism that is built into Transformers could boost the model performance in such a task.

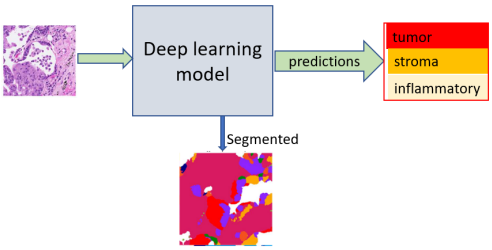

Figure 1. Block diagram of the model for WSIs segmentation and classification

As shown in Figure 1, the first step of in this work is to build a deep learning model with Vision Transformers as input that will receive patches of WSIs and encodes the image, then outputs the probabilities of the dominant diseases as well as the segmented version of the WSIs where each tissue type is mapped to the region of interest. Both segmentation and classification objective functions will be combined to achieve this task. For the lung lesion segmentation, the same model will be used as a baseline. To that end, different techniques for domain adaptation will be explored such as self-supervised learning and transfer learning.

2.1. Evaluation metric

To evaluate the proposed method, we aim to use two different metrics for prediction and segmentation. The prediction level metric consists of: *F1*, *precision*, and the *sensitivity* as defined in 1

$$\begin{aligned} \text{F1} &= \frac{2TP}{2TP + FP + FN}, \\ \text{Sensitivity} &= \frac{TP}{TP + FN}, \\ \text{Precision} &= \frac{TP}{TP + FP}, \end{aligned} \tag{1}$$

where TP, FP and FN are true positive, false positive and false negative, respectively.

For segmentation level metric, the Jacard index or DICE's coefficient D defined in equation 2,

$$D(\hat{Y}, Y) = 2\frac{|\hat{Y} \cap Y|}{|\hat{Y} \cup Y|}, \tag{2}$$

where $\hat{Y}$ and Y are the predicted region from the model and the ground truth region of input image, respectively.

3. Data

The model will be trained and assessed on several datasets. The first dataset used in this study was introduced in [1]. It consists of 151 hematoxylin and eosin strained whole-slide images (WSIs) from histologically confirmed cases of breast cancer. The dataset is available on the Cancer Genome Atlas. The mean and the standard deviation of the region of interest are respectively 1.18mm^2 and 0.80mm^2. The second dataset considered is from [17] which consists of 349 COVID-19 CT images from 216 patients and 463 non-COVID-19 CTs. In addition to that we will used another publicly available dataset published by China National Center for Bioinformation [8] consisting of 650 scans across 150 patients with various stages of COVID.

4. Tentative plan

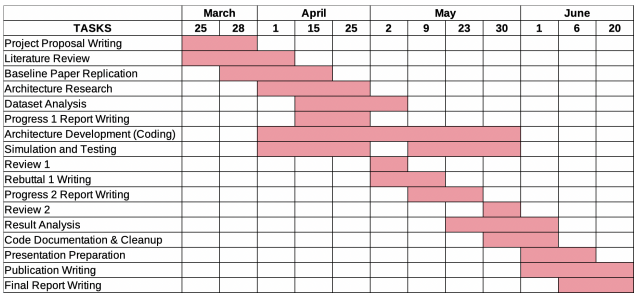

Figure 2. Tentative schedule

References

[1] Mohamed Amgad, Habiba Elfandy, Hagar Hussein, Lamees A Atteya, Mai A T Elsebaie, Lamia S Abo Elnasr, Rokia A Sakr, Hazem S E Salem, Ahmed F Ismail, Anas M Saad, Joumana Ahmed, Maha A T Elsebaie, Mustafijur Rahman, Inas A Ruhban, Nada M Elgazar, Yahya Alagha, Mohamed H Osman, Ahmed M Alhusseiny, Mariam M Khalaf, Abo-Alela F Younes, Ali Abdulkarim, Duaa M Younes, Ahmed M Gadallah, Ahmad M Elkashash, Salma Y Fala, Basma M Zaki, Jonathan Beezley, Deepak R Chittajallu, David Manthey, David A Gutman, and Lee A D Cooper. Structured crowdsourcing enables convolutional segmentation of histology images. *Bioinformatics*, 35(18):3461–3467, 02 2019. 2

[2] Lyndon Chan, Mahdi S. Hosseini, Corwyn Rowsell, Konstantinos N. Plataniotis, and Savvas Damaskinos. Histosegnet: Semantic segmentation of histological tissue type in whole slide images. In *The IEEE International Conference on Computer Vision (ICCV)*, October 2019. 1

[3] Jieneng Chen, Yongyi Lu, Qihang Yu, Xiangde Luo, Ehsan Adeli, Yan Wang, Le Lu, Alan L. Yuille, and Yuyin Zhou. Transunet: Transformers make strong encoders for medical image segmentation, 2021. 1

[4] Nadav Cohen and Amnon Shashua. Inductive bias of deep convolutional networks through pooling geometry, 2017. 2

[5] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding, 2019. 2

[6] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale, 2020. 1, 2

[7] Xiaohang Fu, Tong Liu, Zhaohan Xiong, Bruce H. Smaill, Martin K. Stiles, and Jichao Zhao. Segmentation of histological images and fibrosis identification with a convolutional neural network. *Computers in Biology and Medicine*, 98:147–158, Jul 2018. 1

[8] Hayden Gunraj, Linda Wang, and Alexander Wong. Covidnet-ct: A tailored deep convolutional neural network design for detection of covid-19 cases from chest ct images. *Frontiers in Medicine*, 7:1025, 2020. 2

[9] Xuehai He, Xingyi Yang, Shanghang Zhang, Jinyu Zhao, Yichen Zhang, Eric Xing, and Pengtao Xie. Sample-efficient deep learning for covid-19 diagnosis based on ct scans. *medRxiv*, 2020. 1

[10] Elizabeth A Krupinski, Allison A. Tillack, Lynne Richter, Jeffrey T. Henderson, Achyut K. Bhattacharyya, Katherine M. Scott, Anna R. Graham, Michael R. Descour, John R. Davis, and Ronald S. Weinstein. Eye-movement study and human performance using telepathology virtual slides. implications for medical education and differences with experience. *Human Pathology*, 37(12):1543–1556, 2006. 1

[11] Tao Lei, Risheng Wang, Yong Wan, Bingtao Zhang, Hongying Meng, and Asoke K. Nandi. Medical image segmentation using deep learning: A survey, 2020. 1

[12] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9, 2019. 2

[13] Rhys Thomas & Darren Treanor Rebecca Randell, Roy A. Ruddle. Diagnosis at the microscope: a workplace study of histopathology. *Cognition, Technology & Work*, 14:319–335, 2014. 1

[14] Hugo Touvron, Matthieu Cord, Matthijs Douze, Francisco Massa, Alexandre Sablayrolles, and Hervé Jégou. Training data-efficient image transformers & distillation through attention, 2021. 2

[15] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need, 2017. 1, 2

[16] Q. Yao, L. Xiao, P. Liu, and S. Kevin Zhou. Label-free segmentation of covid-19 lesions in lung ct. *IEEE Transactions on Medical Imaging*, pages 1–1, 2021. 1

[17] Jinyu Zhao, Yichen Zhang, Xuehai He, and Pengtao Xie. Covid-ct-dataset: a ct scan dataset about covid-19. *arXiv preprint arXiv:2003.13865*, 2020. 2