LXM1

**FYP Progress Report**

**WSSS4LUAD: Weakly Supervised Semantic Segmentation for Lung Adenocarcinoma**

**Diagnosis**

**by**

**Tianqi XIANG, Yiduo YU, and Yiwen ZOU**

**LXM1**

**Advised by**

**Prof. Xiaomeng LI**

**Submitted in partial fulfillment**

**of the requirements for COMP 4981**

**in the**

**Department of Computer Science and Engineering**

**The Hong Kong University of Science and Technology**

**2021-2022**

**Date of submission: October 29, 2021**

Table of Contents

[1 Introduction 4](#_Toc85407910)

[1.1 Overview 4](#_Toc85407911)

[1.2 Objectives 5](#_Toc85407912)

[1.3 Literature Survey 6](#_Toc85407913)

[2 Methodology 7](#_Toc85407914)

[2.1 Design 7](#_Toc85407915)

[2.1.1 Literature reviews on Weakly Supervised Learning 8](#_Toc85407916)

[2.1.2 Find the related datasets 8](#_Toc85407917)

[2.1.3 Classification network modifications 8](#_Toc85407918)

[2.1.4 Bridge the two systems 8](#_Toc85407919)

[2.2 Implementation 8](#_Toc85407920)

[2.2.1 Literature reviews on Weakly Supervised Learning 9](#_Toc85407921)

[2.2.2 Find the related datasets 9](#_Toc85407922)

[2.2.3 Classification network modifications 9](#_Toc85407923)

[2.2.4 Bridge the two systems 10](#_Toc85407924)

[2.3 Testing 11](#_Toc85407925)

[2.4 Evaluation 12](#_Toc85407926)

[3 Project Planning 12](#_Toc85407927)

[3.1 Distribution of Work 12](#_Toc85407928)

[3.2 GANTT Chart 13](#_Toc85407929)

[4 Required Hardware & Software 13](#_Toc85407930)

[4.1 Hardware 13](#_Toc85407931)

[4.2 Software 13](#_Toc85407932)

[5 References 14](#_Toc85407933)

# Introduction

## Overview

The complex environment we live in can be completely digitized, which means that it can be accurately calculated, not only storing the past information, but also making accurate predictions about future development trends. It is the same with disease diagnosis. Based on the huge medical data we have accumulated in the past decades, deep learning in computer science can effectively analyze these data, update our understanding of the disease, and give accurate predictions on the remedy or seriousness of the disease and help us take precautions in advance. Among many notorious diseases, cancer is one of our most concerned topics at the moment, and it is also a major medical problem that has not yet been solved. This is due to not only the technical bottleneck in eradicating cancer absolutely, but also the complex processes of cancer diagnosis. For a human doctor to complete the work of cancer diagnosis and localization, he/she must analyze a large number of tissue samples very carefully to draw a conclusion, and this is an extremely time-consuming and error-prone process. An experienced doctor needs years or even tens of years to train. Now the success of deep learning heralds the arrival of artificial intelligence disease diagnosis.

Lung cancer is the leading cause of cancer deaths worldwide. In this project, our goal is to perform tissue semantic segmentation in Hematoxylin and eosin stain (H&E stained) Whole Slide Image (WSI) of lung adenocarcinoma. H&E stain is one of the principal tissue stains used in histology [1][2][3]. It is the most widely used stain in medical diagnosis [1] and is often the gold standard.[4] For example, when a pathologist looks at a biopsy of a suspected cancer, the histological section is likely to be stained with H&E. The current challenge is that it is extremely difficult and time-consuming to obtain pixel-level annotations for organizational semantic segmentation. Inspired by weakly supervised semantic segmentation (WSSS) in computer vision, we decided to use only image-level annotations to perform organizational semantic segmentation.

## Objectives

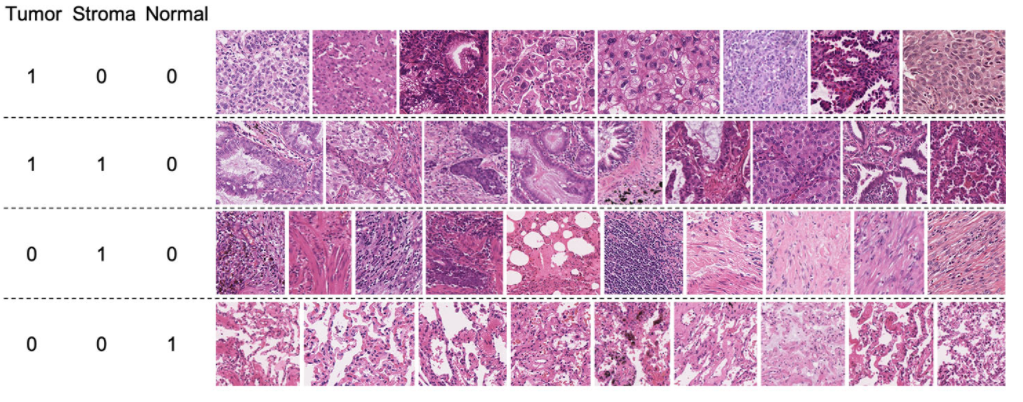
In this project, we obtained 67 H&E-stained glass slides from Guangdong Provincial People’s Hospital (GDPH) and collected 20 WSIs from the Cancer Genome Atlas (TCGA). Only one WSI is extracted per patient. The goal of this project is to use only image-level annotations to achieve pixel-level prediction of three common and meaningful tissue types: tumor epithelial tissue, tumor stromal tissue, and normal tissue.

We want to achieve the state-of-the-art result on this topic by utilizing some novel deep learning technologies related to weakly supervised learning. We believe by only giving the image-level labels, our model could tell the difference between three different tissue types and are able to segment out all the tissues correspondingly in a pixel-level manner.

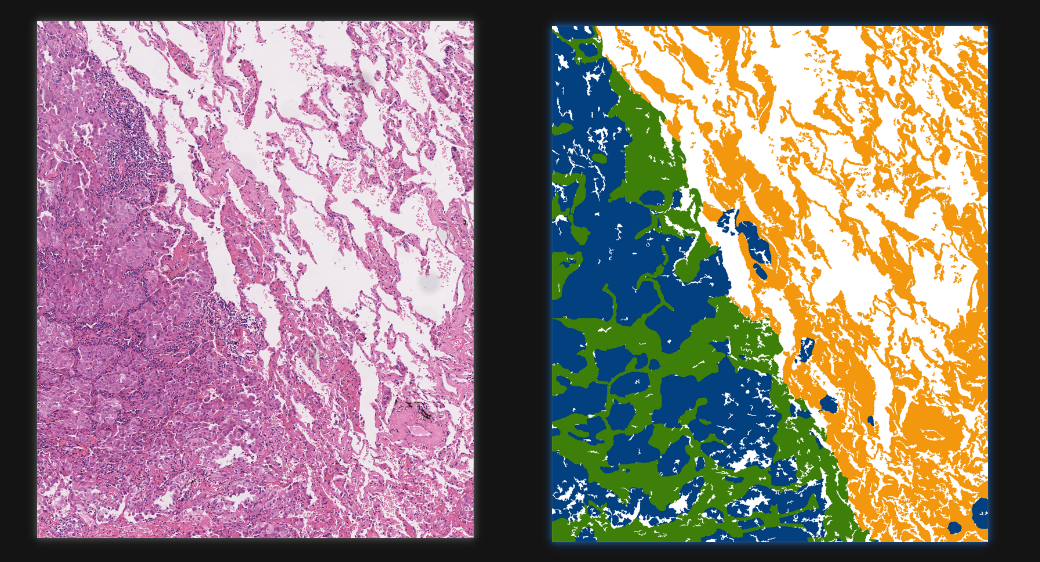
In conclusion, to achieve our goal, we will mainly focus on the following objectives:

1. Build a network and manipulate it so that it could take in multiple labels (image level labels) since our image label could co-exists which is different from previous classification works.

2. Utilizing the weakly supervised learning together with the obtained classification image to tell apart the different tissue types and segment them out.

Below are several training data. For each picture, it will only provide an overall label, stating that the picture contains tumor/stroma/normal cell or not.

Our goal is: given an image with no annotations, predict and give pixel-wise segmentation label for each cell.

Left is the original image, for the right ground truth image, green stands for stroma, blue stands for tumor and yellow represents normal cells.

## Literature Survey

Our tissue segmentation task of three types of tissue is somewhat similar to the k-means clustering based on the RGB value feature [6], which is to cluster points into three types according to their distances in the 3d RGB spaces. The H&S-stained image will render different color to the tumor tissue and normal tissue, so we suggest that despite its simplicity, this approach is meaningful and reasonable.

Further work could be down by DenseCRF [5], a fully connected conditional random field to smooth our clustering boundaries.

Li Yi also proposed a novel method for generating pseudo-mask based on the given image-level labels and class activation map (CAM) [9], the CAM is simply the heatmap that the neural network is focusing on when doing the predictions. It then uses the generated pseudo-mask to train the whole segmentation network by utilizing the pretend-under-fitting strategy. It achieves state-of-the-art result on COCO datasets and VOC12 datasets on the weakly supervised leaderboard.

While class imbalance issue is a hot topic under discussion in supervised training field, research paid less attention to it in the weakly-supervise study. Recently, a class-rebalancing self-training framework Crest [8] is proposed, it argues during traditional training, the majority classes always incur with high recalls while the minority classes gain high precisions. So, with iterative training and weighted probability resampling, by adding more minority classes from unlabeled data to the labeled training set, they can iteratively suppress the drawbacks brought by the data imbalance. This idea motivates our training process when generating the pseudo segmentation label.

Todo: Review ScaleNet

# Methodology

## Design

To accomplish our goals mentioned in the objective parts, we propose our designs as following:

### Literature reviews on Weakly Supervised Learning

Literature reviews on weakly supervised deep learning technologies to find if there’s any other network that is suitable or could motivate our works.

### Find the related datasets

Search online for related datasets that is similar to our H&S-stained lung whole slide image where we could pretrain our segmentation network.

### Classification network modifications

Find ways to modify the typical classification network, modify the loss function and input data format to let the network accept images with two or more labels coexists.

### Bridge the two systems

Incorporates the classification network and our final pseudo label supervised segmentation network together or constructs a pipeline to transmit data between systems. Manipulation required to smooth the training process.

## Implementation

The Implementation Phase will include the following aspects:

### Literature reviews on Weakly Supervised Learning

We started from the survey paper or the one we proposed in the Literature Survey part and look through the papers that cite it or cited by it. Google scholar is a great starting point as it provides us with convenient for the above-mentioned functions.

### Find the related datasets

Based on our design, we will use the key word “lung-cancer”, “WSI”, “H&S-stained” “WSSS” to search the datasets and find related papers and GitHub code for their proposed networks. Currently no related dataset available since the medical data is related to privacy issue and are often too sensitive to put online.

### Classification network modifications

Based on our design, we will think of ways to let the network accept multi label images. One proposed approach that we could try is to give three scores without using SoftMax function at last. Any score that is above a predefined threshold is considered as the label of the image. Another thought is, instead of giving three scores at last (for the three tissue types), we propose seven scores which stand for the combination of any of the three types. For example, score 1 stand for type A only, score 2 and 3 stand for type B and C, relatively. Score 4 stand for A and B coexists in one image, etc. Both ways as well as other manipulations need to be tried out at first.

Currently we modified the loss function to enable training with multi labels. The loss function we use when training our classification network, whether on big patches or cropped small patches are all BinaryCrossEntropy loss (BCE loss), this calculates the loss by taking each label into consideration independently, so it doesn’t care about the co-exists issue and is preferred by our model. For the backbone model, we then could just refer to some famous structures, Resnet101, ScaleNet101 and ScaleNet152 have been tried and we found for the CAM generating task, ScaleNet [10] outperforms Resnet in a large margin but there’s only subtle difference among the 101 and 152 version. For resource utilizing efficiency, we choose ScaleNet101 as out backbone in all cases.

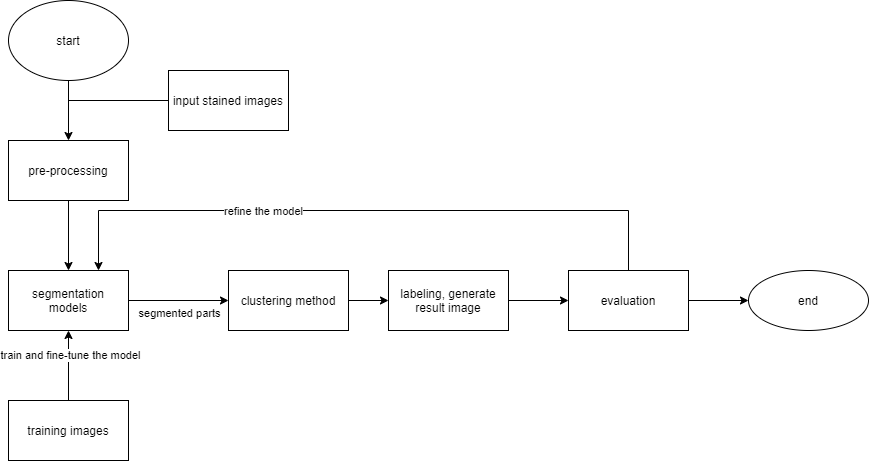
### Bridge the two systems

At last, the two system (classification and segmentation) need to be combined since our task need the information from both sides. Here we also propose two approaches: One way is to incorporate the two networks at the beginning, indicating that we are doing the segmentation and classification at the same time, but in training phrases, we only focus on the classification accuracy. The other way is to separate the two systems and use specific data transfer method to bridge these networks. The data could be in the form of high dimensional vectors which act as a representation of our classification results. One ideal design of our model is like below. Note this is our initial design, and since we are still experimenting different networks and pipelines, the final segmentation system may be quite different.

We now implemented the segmentation network using the structures from [9], and we believe the pretend underfitting strategy could alleviate the pseudo mask noise generated by the CAM, as also indicated by the paper. For the initial classification task, the backbone is ScaleNet101 and we separate the process into two stages:

First, we use the given big image-level labeled patches to train a big model using BCE loss, anticipating that it could give a good classification result on new images. Then we cropped the original images into 96 \* 96 small patches for better CAM details. For original images with only one label, the cropped image must have a single class label which corresponds to the it. But for images with multi-label, we are not sure about the label after cropping, so the trained big-model comes into usage and gives us the predicted pseudo-label.

Secondly, we use all the small patches 96 \* 96 data to train a small network for classification, at last modify the last two fully connected layers to 1 \* 1 convolution layer with same weights to generate the CAM for each image. Now the segmentation pseudo-label is the output and handed to the segmentation part of our model.



## Testing

Our testing procedures is relatively easy, since we are attending the grand challenge: WSSS4LUAD, so we use their provided test set which contains 9 WSIs from GDPH and 3 WSIs from TCGA. Total 80 patches were cropped including 14 large patches (around 1500~5000 \* 1500~5000) and 66 small patches (around 200~500 \* 200~500).

By submitting our model generated results to the grand challenge websites, we could get a final test score and our goal is to maximize this score and surpass other participants.

Currently we got test mIOU 0.7411 and ranked 10 on the leaderboard.

## Evaluation

We use mIOU as the evaluation metrics for model evaluation.

The white background inside the WSI will be excluded when calculating mIOU. The background mask in the validation and testing data is provided. We can directly overlay the white background mask on the prediction results.

The equation for calculating mIOU is attached below for reference.



Where k stands for the total number of classes, P is the prediction results and G is the ground truth.

Currently we got test mIOU 0.7411 and ranked 10 on the leaderboard.

# Project Planning

## Distribution of Work

|  |  |  |  |
| --- | --- | --- | --- |
| **Task** | **Tianqi Xiang** | **Yiduo Yu** | **Yiwen Zou** |
| Do the literature survey | ● | ● | ● |
| Analyze classification models | ○ | ● | ○ |
| Design the classification model | ● | ○ | ● |
| Find related database | ○ | ● | ○ |
| Design the segmentation model | ● | ○ | ○ |
| Design the system bridge | ○ | ● | ● |
| Evaluate the model | ● | ● | ○ |
| Model final fine-tuning of hyperparameters | ○ | ○ | ● |

## GANTT Chart

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Task | July | Aug | Sep | Oct | Nov | Dec | Jan |
| Do the literature survey |  |  |  |  |  |  |  |
| Analyze classification models |  |  |  |  |  |  |  |
| Design the classification model |  |  |  |  |  |  |  |
| Find related database |  |  |  |  |  |  |  |
| Design the segmentation model |  |  |  |  |  |  |  |
| Design the system bridge |  |  |  |  |  |  |  |
| Evaluate the model |  |  |  |  |  |  |  |
| Model final fine-tuning of hyperparameters |  |  |  |  |  |  |  |

# Required Hardware & Software

## Hardware

**Development PC**

**Operating System Type**

**Operating System Name**

**VERSION**

**RAM**

**GPU**

Linux

Ubuntu

20.04.2 LTS

16 GB

3\*NVIDIA GeForce RTX 3090

## Software

**Python**

**VSCode**

**Anaconda**

**Terminus**

**GitHub**

Programming languages

Text editor

Environment manager

Server connecter

Code repository

Here we briefly list our required package for virtual environment:

numpy>=1.18.2

cython>=0.28.5

tensorboard>=1.8.0

tensorboardX>=1.6

cython>=0.28.5

imageio>=2.3.0

scikit-image>=0.14.0

pydensecrf>=1.0

scipy==1.1.0

opencv-python>=3.4.2.17

pandas>=0.23.4

Pillow>=5.2.0

torch>=0.4.1

torchvision>=0.2.1

# References

[1] M. Titford, “The long history of Hematoxylin,” Biotechnic & Histochemistry, vol. 80, no. 2, pp. 73–78, 2005.

[2] Smith C (2006). "Our debt to the logwood tree: the history of hematoxylin". MLO Med Lab Obs. 38 (5): 18, 20–2. PMID 16761865.

[3] R. W. Dapson and R. W. Horobin, “Dyes from a twenty-first century perspective,” Biotechnic &amp; Histochemistry, vol. 84, no. 4, pp. 135–137, 2009. doi:10.1080/10520290902908802. PMID 19384743. S2CID 28563610.

[4] J. Rosai, “Why Microscopy will remain a cornerstone of surgical pathology,” Laboratory Investigation, vol. 87, no. 5, pp. 403–408, 2007.

[5] Krähenbühl, P., & Koltun, V. Efficient inference in fully connected crfs with gaussian edge potentials. Advances in neural information processing systems, 24, 109-117.

[6] Hartigan, J. A., & Wong, M. A. Algorithm AS 136: A k-means clustering algorithm. Journal of the royal statistical society. series c (applied statistics), 28(1), 100-108.

[7] Li, Y., Kuang, Z., Liu, L., Chen, Y., & Zhang, W. Pseudo-mask Matters inWeakly-supervised Semantic Segmentation. arXiv preprint arXiv:2108.12995.

[8] Wei, C., Sohn, K., Mellina, C., Yuille, A., & Yang, F. Crest: A class-rebalancing self-training framework for imbalanced semi-supervised learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 10857-10866).

[9] Li, Y., Kuang, Z., Liu, L., Chen, Y., & Zhang, W. Pseudo-mask Matters in Weakly-supervised Semantic Segmentation. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 6964-6973).

[10] Li, Y., Kuang, Z., Chen, Y., & Zhang, W. Data-driven neuron allocation for scale aggregation networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 11526-11534).