Joelle Fitzgerald
HIDS 509 HW #1 Task 1

Homework Task 1: Use 3 whole slide images that were obtained from TCGA collection and located in the same folder that was used in workshop

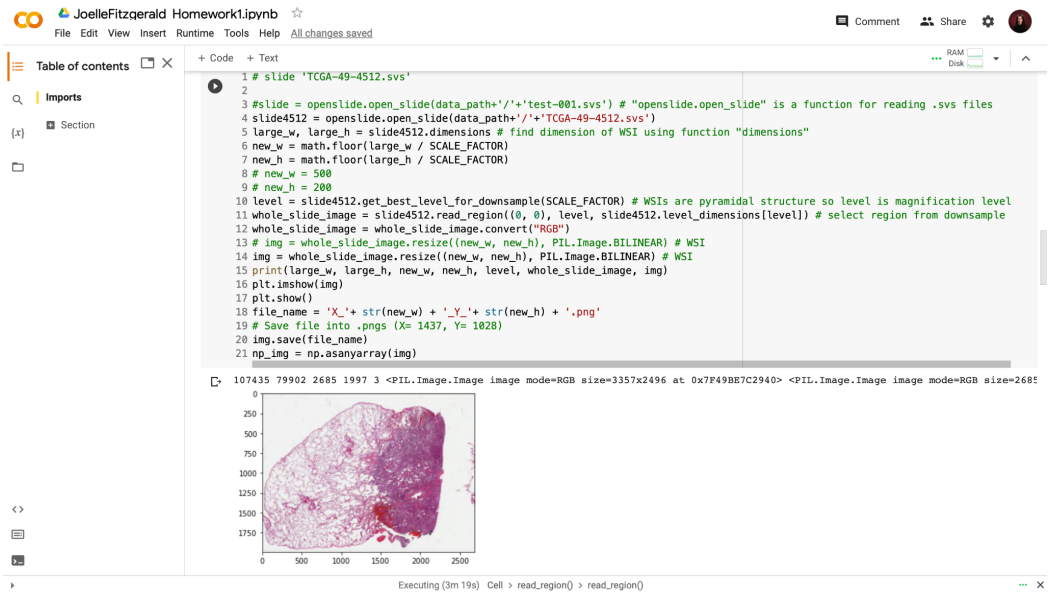
- TCGA-49-4512.svs , TCGA-49-4514svs , and TCGA-78-8640.svs
- Use those WSIs as input and use the same code provided by Dr. Saha during the workshop to process the whole slide images and generate tiles

Summary of steps:

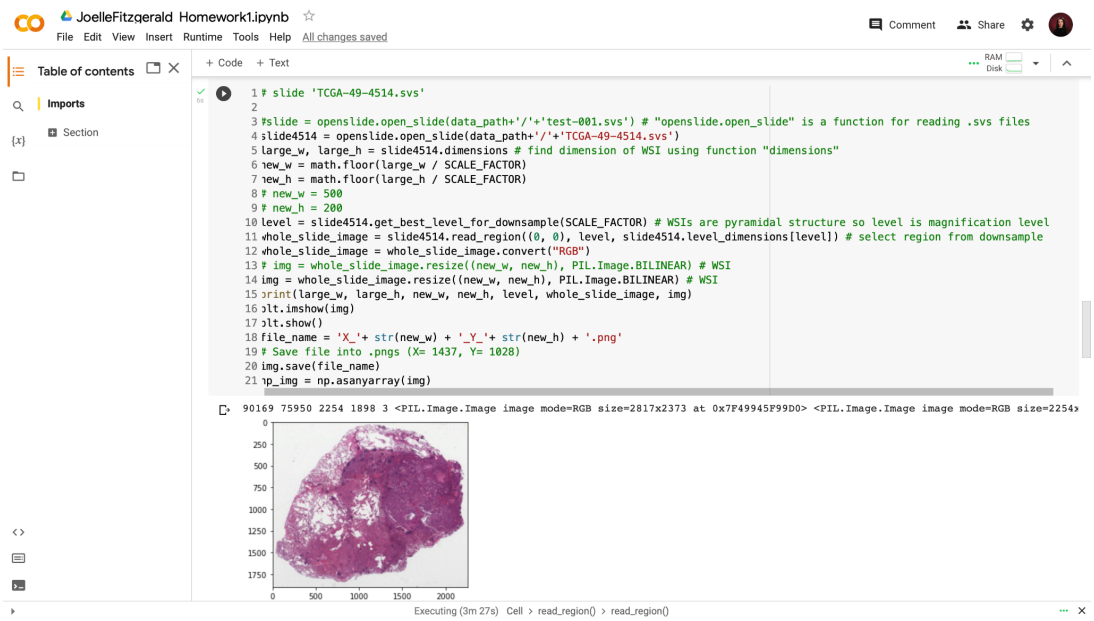
In Task 1, all applicable python library packages are imported into the google collab notebook prior to reading any whole slide images. Next, google drive is mounted to read in the data/images being analyzed. Next, the data is read in and a data path is set to access the images. A scale factor or magnification level for which the images are read under is chosen. In this case, the magnification level is 40x. Next, each slide is read in using openslide for whole slide images (WSI). The dimensions of the whole image are obtained using 'dimensions' and set with a new width and height by dividing the lowest magnification width and height (the original dimensions of the slide) by the set scale factor (40x). The region according to the selected magnification level is obtained by downsampling and is displayed when the code chunk is run. Next, a gray-scale function is used while applying some thresholds to visualize the WSI in black and white to highlight areas of interest (white corresponding to tissue). We then run this gray-scale image on additional functions to calculate the total number of black (background) and white (tissue) pixels and then the total tissue and non-tissue percentage. After the total tissue percentage is obtained and each WSI is run through downsampling, tiles of regions of interest can be downloaded. In the last step, the WSI is iterated over selecting regions of interest at the selected scale factor magnification level (40x). Zoomed-in regions of interest images are created and stored in a new folder 'tiles'.

(see screenshots of generated results below)

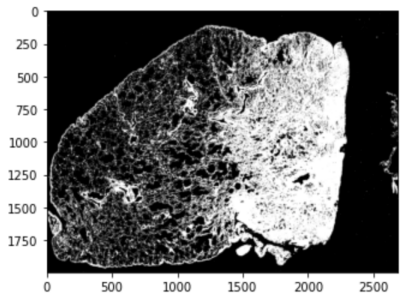
Image



TCGA-49-4512.svs




```
# grey scale, interested in white region (tissue)
r2bgray = filter_rgb_to_grayscale(np_img4512, output_type="uint8")
(thresh, im_bw) = cv2.threshold(r2bgray, 0,255, cv2.THRESH_BINARY | cv2.THRESH_OTSU)
convert_bin = np.invert(im_bw)
plt.imshow(convert_bin, cmap='gray')
plt.show()
```



Image

TCGA-49-4514svs

```
Total tissue is 30.0%
Total Non-tissue is 70.0%
```

JoelleFitzgerald Homework1.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Comment Share

Table of contents

Imports

Section

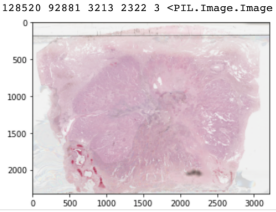
+ Code + Text

RAM Disk

```
# slide 'TCGA-78-8640.svs'

#slide = openslide.open_slide(data_path+'/'+'test-001.svs') # "openslide.open_slide" is a function for reading .svs files
slide8640 = openslide.open_slide(data_path+'/'+'TCGA-78-8640.svs')
large_w, large_h = slide8640.dimensions # find dimension of WSI using function "dimensions"
new_w = math.floor(large_w / SCALE_FACTOR)
new_h = math.floor(large_h / SCALE_FACTOR)
# new_w = 500
# new_h = 200
level = slide8640.get_best_level_for_downsample(SCALE_FACTOR) # WSIs are pyramidal structure so level is magnification level
whole_slide_image = slide8640.read_region((0, 0), level, slide8640.level_dimensions[level]) # select region from downsample
whole_slide_image = whole_slide_image.convert("RGB")
# img = whole_slide_image.resize((new_w, new_h), PIL.Image.BILINEAR) # WSI
img = whole_slide_image.resize((new_w, new_h), PIL.Image.BILINEAR) # WSI
print(large_w, large_h, new_w, new_h, level, whole_slide_image, img)
plt.imshow(img)
plt.show()
file_name = 'X_'+ str(new_w) + '_Y_'+ str(new_h) + '.png'
# Save file into .pngs (X= 1437, Y= 1028)
img.save(file_name)
np_img = np.asanyarray(img)
```

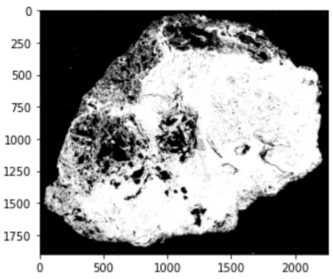
```
128520 92881 3213 2322 3 <PIL.Image.Image image mode=RGB size=4016x2902 at 0x7F498D617EB0> <PIL.Image.Image image mode=RGB size=3213
```



9m 49s completed at 2:39 PM

Image

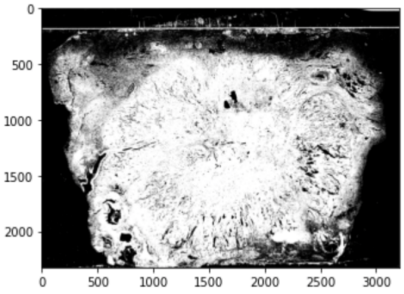
```
[18]  # grey scale, interested in white region (tissue)
      r2bgray = filter_rgb_to_grayscale(np_img4514, output_type="uint8")
      (thresh, im_bw) = cv2.threshold(r2bgray, 0,255, cv2.THRESH_BINARY | cv2.THRESH_OTSU)
      convert_bin = np.invert(im_bw)
      plt.imshow(convert_bin, cmap='gray')
      plt.show()
```



TCGA-78-8640.svs

```
Total tissue is 49.0%
Total Non-tissue is 51.0%
```

```
# grey scale, interested in white region (tissue)
r2bgray = filter_rgb_to_grayscale(np_img8640, output_type="uint8")
(thresh, im_bw) = cv2.threshold(r2bgray, 0,255, cv2.THRESH_BINARY | cv2.THRESH_OTSU)
convert_bin = np.invert(im_bw)
plt.imshow(convert_bin, cmap='gray')
plt.show()
```




```
Total tissue is 55.0%
Total Non-tissue is 45.0%
```