Preprocessed datasets. [70/15/15]

Tile the images using the magnification (20x) and tile size of interest (512x512)

LUAD 518 --> Number of images.

LUAD\_test 75

LUAD\_train 345

LUAD\_valid 98

#/nfs/home/xwang/ReRun\_Feb21th/02\_sort\_tiles\_ReRun

\*-------------------------------------------------------------------------------------------------------------------------

\*-------------------------------------------------------------------------------------------------------------------------

# preprocess 01 step.

# break into tiles

# Tile the images using the magnification (20x) and tile size of interest (512x512 px in this example):

'''

#!/usr/bin/env bash

printf "$PWD \n\n"

# list all files in the image directory:

cd /nfs/home/xwang/Symlink\_image

printf "$PWD \n\n"

for fentry in "$PWD"/\*.svs

do

printf "$fentry\n"

printf "checking\_name\n"

nohup python /nfs/home/xwang/DeepPATH/DeepPATH\_code/00\_preprocessing/0b\_tileLoop\_deepzoom4.py -s 512 -e 0 -j 32 -B 50 -M 20 -o /nfs6/deeppath\_scratch/Pre\_processing\_2nd\_Feb21/ $fentry

done

'''

'''

pip install 'scipy<1.3.0'

pip install scikit-image

pip install pydicom

go to 0b\_tileLoop\_deepzoom4.py, changing import dicom to import pydicom

go to 0b\_tileLoop\_deepzoom4.py, changed the <Object> parameter = 20.

'''

#results:

# at directory:

# /nfs6/deeppath\_scratch/Pre\_processing\_2nd\_Feb21

# -bash-4.2$ ls |wc

# 1040 1040 88400

#

# total: 1040 images.

# for each one, only 20x magnification is contained.

#-----------------

# 0.2: Sort the tiles into train/valid/test sets according to the classes defined.

#Sort the dataset into a test, train and validation cohort for a 3-way classifier (LUAD/LUSC/Normal).

# You need to create a new directory and run this job from that directory

'''

cd /nfs/home/xwang/ReRun\_Feb21th/02\_sort\_tiles\_ReRun

nohup python /nfs/home/xwang/DeepPATH/DeepPATH\_code/00\_preprocessing/0d\_SortTiles.py --SourceFolder=/nfs6/deeppath\_scratch/Pre\_processing\_2nd\_Feb21/ --Magnification=20.0 --MagDiffAllowed=0 --SortingOption=14 --PatientID=12 --nSplit 0 --JsonFile /nfs/home/xwang/Sorting\_option\_JASON\_FILE.txt --PercentTest=15 --PercentValid=15

'''

# <Sorting\_option\_JASON\_FILE.txt>

# nohup python /nfs/home/xwang/DeepPATH/DeepPATH\_code/00\_preprocessing/0d\_SortTiles.py --SourceFolder /nfs6/deeppath\_scratch/ --JsonFile /nfs/home/xwang/Sorting\_option\_JASON\_FILE.txt --Magnification 20 --MagDiffAllowed 5 --SortingOption 14 --PercentTest 15 --PercentValid 15 --PatientID 8 --nSplit 0

#result:

'''

Train / Test / Validation tiles sets for LUAD = 0.698807 % / 0.152109 % / 0.149084 %

Train / Test / Validation slides sets for LUAD = 0.666023 % / 0.144788 % / 0.189189 %

Train / Test / Validation patients sets for LUAD = 0.670103 % / 0.144330 % / 0.185567 %

LUAD 612822

LUAD\_test 93216

LUAD\_train 428244

LUAD\_valid 91362

LUAD\_test 0.1521094216591441

LUAD\_train 0.6988065049883979

LUAD\_valid 0.14908407335245796

LUAD 518 --> Number of images.

LUAD\_test 75

LUAD\_train 345

LUAD\_valid 98

LUAD 485 --> number of patients.

LUAD\_test 70

LUAD\_train 325

LUAD\_valid 90

'''

#bash-4.2$ cd LUAD/

#bash-4.2$ ls | wc

# 612822 612822 59300030

#/nfs/home/xwang/ReRun\_Feb21th/02\_sort\_tiles\_ReRun

\*-------------------------------------------------------------------------------------------------------------------------

\*-------------------------------------------------------------------------------------------------------------------------

#-----------------

# 0.3a Convert the JPEG tiles into TFRecord format for 2 or 3 classes jobs

# Convert data into TFRecord files for each dataset

'''

mkdir r1\_TFRecord\_test

mkdir r1\_TFRecord\_valid

mkdir r1\_TFRecord\_train

python 00\_preprocessing/TFRecord\_2or3\_Classes/build\_TF\_test.py --directory='r1\_sorted\_3Cla/' --output\_directory='r1\_TFRecord\_test' --num\_threads=1 --one\_FT\_per\_Tile=False --ImageSet\_basename='test'

python 00\_preprocessing/TFRecord\_2or3\_Classes/build\_TF\_test.py --directory='r1\_sorted\_3Cla/' --output\_directory='r1\_TFRecord\_valid' --num\_threads=1 --one\_FT\_per\_Tile=False --ImageSet\_basename='valid'

python 00\_preprocessing/TFRecord\_2or3\_Classes/build\_image\_data.py --directory='r1\_sorted\_3Cla/' --output\_directory='r1\_TFRecord\_train' --train\_shards=1024 --validation\_shards=128 --num\_threads=16

'''

'''

test set:

nohup python /nfs/home/xwang/DeepPATH/DeepPATH\_code/00\_preprocessing/TFRecord\_2or3\_Classes/build\_TF\_test.py --directory=/nfs/home/xwang/ReRun\_Feb21th/02\_sort\_tiles\_ReRun/ --output\_directory=/nfs/home/xwang/ReRun\_Feb21th/03a\_convertJPEGtoTFRecord/03a\_TFRecord\_test/ --num\_threads=1 --one\_FT\_per\_Tile=False --ImageSet\_basename='test'

result:

2020-02-22 21:57:55.799083 [thread 0]: Wrote 75 images to /nfs/home/xwang/ReRun\_Feb21th/03a\_convertJPEGtoTFRecord/03a\_TFRecord\_test/test\_0xF425035CD29D98CEAD2A5203DD372E25833C4D32ECB694C29043ED64F9F6BCF9\_1x8ZD0YD4OF7\_1.TFRecord

2020-02-22 21:57:56.339577: Finished writing all 93216 images in data set.

\*----------------------

valid set:

nohup python /nfs/home/xwang/DeepPATH/DeepPATH\_code/00\_preprocessing/TFRecord\_2or3\_Classes/build\_TF\_test.py --directory=/nfs/home/xwang/ReRun\_Feb21th/02\_sort\_tiles\_ReRun/ --output\_directory=/nfs/home/xwang/ReRun\_Feb21th/03a\_convertJPEGtoTFRecord/03a\_TFRecord\_valid/ --num\_threads=1 --one\_FT\_per\_Tile=False --ImageSet\_basename='valid'

result:

(virpy3.6) bash-4.2$ ls |wc

99 99 9516

2020-02-22 22:49:55.781955 [thread 0]: Wrote 98 images to /nfs/home/xwang/ReRun\_Feb21th/03a\_convertJPEGtoTFRecord/03a\_TFRecord\_valid/valid\_0xF2F4BD5B36F481D7EE3970B6EB0BA6D91CC97D8F4049FBD7040D14BAE50956A5\_1x1561JU9WCA\_1.TFRecord

2020-02-22 22:49:56.782040: Finished writing all 91362 images in data set.

\*----------------------

train set:

nohup python /nfs/home/xwang/DeepPATH/DeepPATH\_code/00\_preprocessing/TFRecord\_2or3\_Classes/build\_TF\_test.py --directory=/nfs/home/xwang/ReRun\_Feb21th/02\_sort\_tiles\_ReRun/ --output\_directory=/nfs/home/xwang/ReRun\_Feb21th/03a\_convertJPEGtoTFRecord/03a\_TFRecord\_train/ --num\_threads=1 --one\_FT\_per\_Tile=True --ImageSet\_basename='train'

result:

2020-02-23 03:04:47.904774 [thread 0]: Wrote 428244 images to /nfs/home/xwang/ReRun\_Feb21th/03a\_convertJPEGtoTFRecord/03a\_TFRecord\_train/train\_0xF42156E70CA66CC5ABE61E0E48DFCE826BD7916407E959AED57FD55471EC2FE4\_1xMHPM0WZ0ZC\_9\_50.jpeg\_1.TFRecord

2020-02-23 03:04:48.277915: Finished writing all 428244 images in data set.

'''

"""

# April 24, 2020

train set:

nohup python /nfs/home/xwang/DeepPATH/DeepPATH\_code/00\_preprocessing/TFRecord\_2or3\_Classes/build\_image\_data.py --directory=/nfs/home/xwang/ReRun\_Feb21th/02\_sort\_tiles\_ReRun/ --output\_directory=/nfs/home/xwang/ReRun\_Feb21th/03a\_convertJPEGtoTFRecord/03a\_TFRecord\_train/ --train\_shards=1024 --validation\_shards=128 --num\_threads=16

result:

2020-04-24 11:56:48.026972 [thread 14]: Wrote 419 images to /nfs/home/xwang/ReRun\_Feb21th/03a\_convertJPEGtoTFRecord/03a\_TFRecord\_train/train-00959-of-01024

2020-04-24 11:56:48.027041 [thread 14]: Wrote 26765 images to 26765 shards.

2020-04-24 11:56:49.027243: Finished dealing with all 428244 images in data set.

"""

\*-------------------------------------------------------------------------------------------------------------------------

\*-------------------------------------------------------------------------------------------------------------------------

**Transfer Learning**

* **From DeepPath Saved Model (tune the last layer)**

**Training from scratch [3-way classifier]**

1. Build the model from proper directory, e.g. cd 01\_training/xClasses:

bazel build inception/imagenet\_train

1. After the model is built, train the model for all training images:

bazel-bin/inception/imagenet\_train --num\_gpus=1 --batch\_size=30 --train\_dir='output\_directory' --data\_dir='TFRecord\_images\_directory' --ClassNumber=3 --mode='0\_softmax'

* **num\_epochs\_per\_decay**
* **learning\_rate\_decay\_factor**
* **NbrOfImages – num of tiles for training set.**
* **max\_steps – NbrofImages/batch\_size \* 101**
* **save\_step\_for\_chekcpoint – NbrOfImages / batch\_size**

Example TCGA:

Train the 3-way classifier:

mkdir r1\_results

bazel build inception/imagenet\_train

bazel-bin/inception/imagenet\_train --num\_gpus=4 --batch\_size=400 --train\_dir='r1\_results' --data\_dir='r1\_TFRecord\_train' --ClassNumber=3 --mode='0\_softmax' --NbrOfImages=923893 --save\_step\_for\_chekcpoint=2300 --max\_steps=230001

**Transfer learning – from Inception [3-way classifier]**

1. Build the model (run from proper directory) e.g. cd 01\_training/xClasses:

bazel build inception/imagenet\_train

1. Download the checkpoints of the network trained by google.

> ls inception-v3

README.txt

checkpoint

model.ckpt-157585

1. Run this model for all training images:

```shell

#!/bin/tcsh

#$ -pe openmpi 1

#$ -A TensorFlow

#$ -N rqs\_train

#$ -cwd

#$ -S /bin/tcsh

#$ -q gpu0.q

#$ -l excl=true

module load cuda/8.0

module load python/3.5.3

module load bazel/0.4.4

bazel-bin/inception/imagenet\_train --num\_gpus=1 --batch\_size=30 --train\_dir='output\_directory' --data\_dir='TFRecord\_images\_directory' --pretrained\_model\_checkpoint\_path="path\_to/model.ckpt-157585" --fine\_tune=True --initial\_learning\_rate=0.001 --ClassNumber=3 --mode='0\_softmax'

The second flag --fine\_tune is a boolean that indicates whether the last classification layer should be randomly initialized or restored. You may set this flag to false if you wish to continue training a pre-trained model from a checkpoint. If you set this flag to true, you can train a new classification layer from scratch.

1. Build the exact same model as previously except we change the number of labels in the final classification layer.
2. Restore all weights from the pre-trained Inception-v3 except for the final classification layer; this will get randomly initialized instead.

**Training from scratch [10-way classifier]**

Example TCGA:

Train the model with 10-class sigmoid classifier:

bazel-bin/inception/imagenet\_train --num\_gpus=4 --batch\_size=400 --train\_dir="r3\_results\_train" --data\_dir="r3\_TFRecord\_train" --ClassNumber=10 --mode='1\_sigmoid' --NbrOfImages=326613 --save\_step\_for\_chekcpoint=815 --max\_steps=81501

#--------------------------------------

Transfer learning from DeepPath saved model for 3-way classifier

#--------------------------------------

#--------------------------------------

Inside 01\_training/xClasses

#--------------------------------------

bazel-bin/inception/imagenet\_train --num\_gpus=4 --batch\_size=400 --train\_dir=/nfs/home/xwang/Transfer\_Learning\_April/model\_1a/train\_model\_1a/ --data\_dir=/nfs/home/xwang/ReRun\_Feb21th/03a\_convertJPEGtoTFRecord/03a\_TFRecord\_train/ --pretrained\_model\_checkpoint\_path=/nfs6/deeppath\_models/checkpoints/run1a\_3D\_classifier/model.ckpt-69000 --fine\_tune=True --initial\_learning\_rate=0.001 --ClassNumber=3 --mode='0\_softmax' --NbrOfImages=428244 --save\_step\_for\_chekcpoint=1070 --max\_steps=107001

#--------------------------------------

#--------------------------------------

#--------------------------------------

#--------------------------------------

Wait for 1070 checkpoint occur to run validation

Run validation on the first 1070:

Result:

/nfs/home/xwang/Transfer\_Learning\_April/model\_1a/model\_1a\_r1\_valid/test\_1070k

/nfs/home/xwang/Transfer\_Learning\_April/model\_1a/model\_1a\_r1\_valid/test\_2140k

/nfs/home/xwang/Transfer\_Learning\_April/model\_1a/model\_1a\_r1\_valid/test\_3210k

/nfs/home/xwang/Transfer\_Learning\_April/model\_1a/train\_model\_1a/training\_loss.txt

#--------------------------------------

AUC = 1;

Loss converged. Graphed

#--------------------------------------

To Do List:

- For validation set, heatmap & PB graph.

- After all is done, need to use test set for all task.

* Use the last checkpoint model directory:

/nfs/home/xwang/Transfer\_Learning\_April/model\_1a/train\_model\_1a/

/nfs/home/xwang/Transfer\_Learning\_April/model\_1a/train\_model\_1a/model.ckpt-3210

#--------------------------------------

Transfer learning from DeepPath saved model for 10-way classifier

* Newly sort all LUAD tiles again - Extract probability of LUAD tiles on all LUAD tiles, we will need to run them through the above classifier (3-way).
* Use the last checkpoint model:

/nfs/home/xwang/Transfer\_Learning\_April/model\_1a/train\_model\_1a/

* Sort tiles, assigning them all to “test”.

cd /nfs/home/xwang/Transfer\_Learning\_April/r2\_LUAD\_segmentation

nohup python

/nfs/home/xwang/DeepPATH/DeepPATH\_code/00\_preprocessing/0d\_SortTiles.py

--SourceFolder=/nfs6/deeppath\_scratch/Pre\_processing\_2nd\_Feb21/

--Magnification=20.0

--MagDiffAllowed=0

--SortingOption=14

--PatientID=12

--nSplit 0

--JsonFile /nfs/home/xwang/Sorting\_option\_JASON\_FILE.txt

--PercentTest=100

--PercentValid=0

cd /nfs/home/xwang/Transfer\_Learning\_April/r2\_LUAD\_segmentation

nohup python /nfs/home/xwang/DeepPATH/DeepPATH\_code/00\_preprocessing/0d\_SortTiles.py --SourceFolder=/nfs6/deeppath\_scratch/Pre\_processing\_2nd\_Feb21/ --Magnification=20.0 --MagDiffAllowed=0 --SortingOption=14 --PatientID=12 --nSplit 0 --JsonFile /nfs/home/xwang/Sorting\_option\_JASON\_FILE.txt --PercentTest=100 --PercentValid=0

result:

LUAD 612822

LUAD\_test 612822

LUAD\_train 0

LUAD\_valid 0

LUAD\_test 1.0

LUAD\_train 0.0

LUAD\_valid 0.0

LUAD 518

LUAD\_test 518

LUAD\_train 0

LUAD\_valid 0

LUAD 485

LUAD\_test 485

LUAD\_train 0

LUAD\_valid 0

* Convert to TFRecords:

cd /nfs/home/xwang/Transfer\_Learning\_April/r2\_TFRecord\_test/

nohup python

/nfs/home/xwang/DeepPATH/DeepPATH\_code/00\_preprocessing/TFRecord\_2or3\_Classes/build\_TF\_test.py

--directory=/nfs/home/xwang/Transfer\_Learning\_April/r2\_LUAD\_segmentation/

--output\_directory=/nfs/home/xwang/Transfer\_Learning\_April/r2\_TFRecord\_test/

--num\_threads=1

--one\_FT\_per\_Tile=False

--ImageSet\_basename='test'

nohup python /nfs/home/xwang/DeepPATH/DeepPATH\_code/00\_preprocessing/TFRecord\_2or3\_Classes/build\_TF\_test.py --directory=/nfs/home/xwang/Transfer\_Learning\_April/r2\_LUAD\_segmentation/ --output\_directory=/nfs/home/xwang/Transfer\_Learning\_April/r2\_TFRecord\_test/ --num\_threads=1 --one\_FT\_per\_Tile=False --ImageSet\_basename='test'

result:

-bash-4.2$ ls | wc

519 519 49738

-bash-4.2$ pwd

/nfs/home/xwang/Transfer\_Learning\_April/r2\_TFRecord\_test

* Segment the LUAD tiles using the checkpoint giving the best validation/test AUC:

Result:

/nfs/home/xwang/Transfer\_Learning\_April/r2\_result/test\_3210k

* sort the LUAD tiles identified as LUAD intro a train, valid a test set for mutation analysis:

cd r3\_LUAD\_sorted

nohup python

/nfs/home/xwang/DeepPATH/DeepPATH\_code/00\_preprocessing/0d\_SortTiles.py

--SourceFolder=/nfs6/deeppath\_scratch/Pre\_processing\_2nd\_Feb21/

--Magnification=20.0

--MagDiffAllowed=0

--SortingOption=10

--PatientID=-1

--nSplit 0

--PercentTest=15

--PercentValid=15

--outFilenameStats=/nfs/home/xwang/Transfer\_Learning\_April/r2\_result/test\_3210k/out\_filename\_Stats.txt

nohup python /nfs/home/xwang/DeepPATH/DeepPATH\_code/00\_preprocessing/0d\_SortTiles.py --SourceFolder=/nfs6/deeppath\_scratch/Pre\_processing\_2nd\_Feb21/ --Magnification=20.0 --MagDiffAllowed=0 --SortingOption=10 --PatientID=-1 --nSplit 0 --PercentTest=15 --PercentValid=15 --outFilenameStats=/nfs/home/xwang/Transfer\_Learning\_April/r2\_result/test\_3210k/out\_filename\_Stats.txt

result:

Pre\_processing\_2nd\_Feb21 612819

Pre\_processing\_2nd\_Feb21\_test 91923

Pre\_processing\_2nd\_Feb21\_train 427568

Pre\_processing\_2nd\_Feb21\_valid 9332

Pre\_processing\_2nd\_Feb21\_test 0.1500002447704787

Pre\_processing\_2nd\_Feb21\_train 0.6977068269750122

Pre\_processing\_2nd\_Feb21\_valid 0.15229292825450907

Pre\_processing\_2nd\_Feb21 518

Pre\_processing\_2nd\_Feb21\_test 85

Pre\_processing\_2nd\_Feb21\_train 363

Pre\_processing\_2nd\_Feb21\_valid 70

* Convert to TFRecord:

# valid

python /nfs/home/xwang/DeepPATH/DeepPATH\_code/00\_preprocessing/TFRecord\_multi\_Classes/build\_TF\_test\_multiClass.py

--directory=/nfs/home/xwang/Transfer\_Learning\_April/r3\_LUAD\_sorted/Pre\_processing\_2nd\_Feb21/

--output\_directory=/nfs/home/xwang/Transfer\_Learning\_April/model\_1b/r3\_TFRecord\_valid/

--num\_threads=1 --one\_FT\_per\_Tile=False --ImageSet\_basename='valid'

--labels\_names=/nfs/home/xwang/gene\_mutation\_prediction\_March/labelref\_r3.txt

--labels=/nfs/home/xwang/gene\_mutation\_prediction\_March/labels\_r3.txt

--PatientID=14

cd /nfs/home/xwang/Transfer\_Learning\_April/model\_1b/r3\_TFRecord\_valid/

nohup python /nfs/home/xwang/DeepPATH/DeepPATH\_code/00\_preprocessing/TFRecord\_multi\_Classes/build\_TF\_test\_multiClass.py --directory=/nfs/home/xwang/Transfer\_Learning\_April/r3\_LUAD\_sorted/Pre\_processing\_2nd\_Feb21/ --output\_directory=/nfs/home/xwang/Transfer\_Learning\_April/model\_1b/r3\_TFRecord\_valid/ --num\_threads=1 --one\_FT\_per\_Tile=False --ImageSet\_basename='valid' --labels\_names=/nfs/home/xwang/gene\_mutation\_prediction\_March/labelref\_r3.txt --labels=/nfs/home/xwang/gene\_mutation\_prediction\_March/labels\_r3.txt --PatientID=14

* Only 18/70{VAlid} been converted to TFRecord because only 109 Images -patients have KRAS label. [label file only contain images/patietns with KRAS mutation.]
* I only labeled images that have KRAS mutation. For images that does not have KRAS mutation is not included in the label file. This results in only KRAS mutated images are converted to TFRecord for training/valid/test. so ~ 109/580 images used in total. Since all samples are KRAS mutation class, wouldn’t the trained model will also predict well for KRAS? Should I include other images also? just label them with other genes? If I do that way, we dont know the actual / fact for these labels right?
* First, try with ~109 KRAS mutated labeled images for transfer learning from 10-way classifier.
* Then, try to use all images with 2 class labels for transfer training from model 1b/ 10-way classifier. Class 1 = KRAS mutated; Class 2 = No KRAS mutation.

# test

python /nfs/home/xwang/DeepPATH/DeepPATH\_code/00\_preprocessing/TFRecord\_multi\_Classes/build\_TF\_test\_multiClass.py

--directory=/nfs/home/xwang/Transfer\_Learning\_April/r3\_LUAD\_sorted/Pre\_processing\_2nd\_Feb21/

--output\_directory=/nfs/home/xwang/Transfer\_Learning\_April/model\_1b/r3\_TFRecord\_test/

--num\_threads=1 --one\_FT\_per\_Tile=False --ImageSet\_basename='test'

--labels\_names=/nfs/home/xwang/gene\_mutation\_prediction\_March/labelref\_r3.txt

--labels=/nfs/home/xwang/gene\_mutation\_prediction\_March/labels\_r3.txt

--PatientID=14

cd /nfs/home/xwang/Transfer\_Learning\_April/model\_1b/r3\_TFRecord\_test/

nohup python /nfs/home/xwang/DeepPATH/DeepPATH\_code/00\_preprocessing/TFRecord\_multi\_Classes/build\_TF\_test\_multiClass.py --directory=/nfs/home/xwang/Transfer\_Learning\_April/r3\_LUAD\_sorted/Pre\_processing\_2nd\_Feb21/ --output\_directory=/nfs/home/xwang/Transfer\_Learning\_April/model\_1b/r3\_TFRecord\_test/ --num\_threads=1 --one\_FT\_per\_Tile=False --ImageSet\_basename='test' --labels\_names=/nfs/home/xwang/gene\_mutation\_prediction\_March/labelref\_r3.txt --labels=/nfs/home/xwang/gene\_mutation\_prediction\_March/labels\_r3.txt --PatientID=14

Result:

22/85 {test} converted to TFRecord

# train

python /nfs/home/xwang/DeepPATH/DeepPATH\_code/00\_preprocessing/TFRecord\_multi\_Classes/build\_image\_data\_multiClass.py

--directory=/nfs/home/xwang/Transfer\_Learning\_April/r3\_LUAD\_sorted/Pre\_processing\_2nd\_Feb21/

--output\_directory=/nfs/home/xwang/Transfer\_Learning\_April/model\_1b/r3\_TFRecord\_train/

--train\_shards=1024

--validation\_shards=128

--num\_threads=16

--labels\_names=/nfs/home/xwang/gene\_mutation\_prediction\_March/labelref\_r3.txt

--labels=/nfs/home/xwang/gene\_mutation\_prediction\_March/labels\_r3.txt

--PatientID=14

cd /nfs/home/xwang/Transfer\_Learning\_April/model\_1b/r3\_TFRecord\_train

nohup python /nfs/home/xwang/DeepPATH/DeepPATH\_code/00\_preprocessing/TFRecord\_multi\_Classes/build\_image\_data\_multiClass.py --directory=/nfs/home/xwang/Transfer\_Learning\_April/r3\_LUAD\_sorted/Pre\_processing\_2nd\_Feb21/ --output\_directory=/nfs/home/xwang/Transfer\_Learning\_April/model\_1b/r3\_TFRecord\_train/ --train\_shards=1024 --validation\_shards=128 --num\_threads=16 --labels\_names=/nfs/home/xwang/gene\_mutation\_prediction\_March/labelref\_r3.txt --labels=/nfs/home/xwang/gene\_mutation\_prediction\_March/labels\_r3.txt --PatientID=14

Result:

83/363 {train image} 84916/427568

Transfer learning from model 1b – 10 way classifier

#--------------------------------------

#--------------------------------------

Inside 01\_training/xClasses

#--------------------------------------

nohup bazel-bin/inception/imagenet\_train --num\_gpus=4 --batch\_size=400 --train\_dir=/nfs/home/xwang/Transfer\_Learning\_April/model\_1b/transfer\_learning\_model1b\_onlyKRAS/ --data\_dir=/nfs/home/xwang/Transfer\_Learning\_April/model\_1b/r3\_TFRecord\_train/ --pretrained\_model\_checkpoint\_path=/nfs6/deeppath\_models/checkpoints/run1b\_10way\_MutationClassifier/model.ckpt-81500 --fine\_tune=True --initial\_learning\_rate=0.001 --ClassNumber=10 --mode='1\_sigmoid' --NbrOfImages=84916 --save\_step\_for\_chekcpoint=212 --max\_steps=21201

#--------------------------------------

Run validation set on checkpoint model -212

* AUC = 1, because all training and testing data are only one class. Of course.

#--------------------------------------

#--------------------------------------

#--------------------------------------

Transfer learning from DeepPath saved model for 10-way classifier

* Use all images with 2 class labels for transfer training from model 1b/ 10-way classifier. Class 1 = KRAS mutated; Class 2 = No KRAS mutation.

\*\*\*\*\* Same as previous steps:

* Newly sort all LUAD tiles again - Extract probability of LUAD tiles on all LUAD tiles, we will need to run them through the above classifier (3-way).
* Use the last checkpoint model:

/nfs/home/xwang/Transfer\_Learning\_April/model\_1a/train\_model\_1a/

* Sort tiles, assigning them all to “test”.

cd /nfs/home/xwang/Transfer\_Learning\_April/r2\_LUAD\_segmentation

nohup python

/nfs/home/xwang/DeepPATH/DeepPATH\_code/00\_preprocessing/0d\_SortTiles.py

--SourceFolder=/nfs6/deeppath\_scratch/Pre\_processing\_2nd\_Feb21/

--Magnification=20.0

--MagDiffAllowed=0

--SortingOption=14

--PatientID=12

--nSplit 0

--JsonFile /nfs/home/xwang/Sorting\_option\_JASON\_FILE.txt

--PercentTest=100

--PercentValid=0

cd /nfs/home/xwang/Transfer\_Learning\_April/r2\_LUAD\_segmentation

nohup python /nfs/home/xwang/DeepPATH/DeepPATH\_code/00\_preprocessing/0d\_SortTiles.py --SourceFolder=/nfs6/deeppath\_scratch/Pre\_processing\_2nd\_Feb21/ --Magnification=20.0 --MagDiffAllowed=0 --SortingOption=14 --PatientID=12 --nSplit 0 --JsonFile /nfs/home/xwang/Sorting\_option\_JASON\_FILE.txt --PercentTest=100 --PercentValid=0

result:

LUAD 612822

LUAD\_test 612822

LUAD\_train 0

LUAD\_valid 0

LUAD\_test 1.0

LUAD\_train 0.0

LUAD\_valid 0.0

LUAD 518

LUAD\_test 518

LUAD\_train 0

LUAD\_valid 0

LUAD 485

LUAD\_test 485

LUAD\_train 0

LUAD\_valid 0

* Convert to TFRecords:

cd /nfs/home/xwang/Transfer\_Learning\_April/r2\_TFRecord\_test/

nohup python

/nfs/home/xwang/DeepPATH/DeepPATH\_code/00\_preprocessing/TFRecord\_2or3\_Classes/build\_TF\_test.py

--directory=/nfs/home/xwang/Transfer\_Learning\_April/r2\_LUAD\_segmentation/

--output\_directory=/nfs/home/xwang/Transfer\_Learning\_April/r2\_TFRecord\_test/

--num\_threads=1

--one\_FT\_per\_Tile=False

--ImageSet\_basename='test'

nohup python /nfs/home/xwang/DeepPATH/DeepPATH\_code/00\_preprocessing/TFRecord\_2or3\_Classes/build\_TF\_test.py --directory=/nfs/home/xwang/Transfer\_Learning\_April/r2\_LUAD\_segmentation/ --output\_directory=/nfs/home/xwang/Transfer\_Learning\_April/r2\_TFRecord\_test/ --num\_threads=1 --one\_FT\_per\_Tile=False --ImageSet\_basename='test'

result:

-bash-4.2$ ls | wc

519 519 49738

-bash-4.2$ pwd

/nfs/home/xwang/Transfer\_Learning\_April/r2\_TFRecord\_test

* Segment the LUAD tiles using the checkpoint giving the best validation/test AUC:

Result:

/nfs/home/xwang/Transfer\_Learning\_April/r2\_result/test\_3210k

* sort the LUAD tiles identified as LUAD intro a train, valid a test set for mutation analysis:

cd r3\_LUAD\_sorted

nohup python

/nfs/home/xwang/DeepPATH/DeepPATH\_code/00\_preprocessing/0d\_SortTiles.py

--SourceFolder=/nfs6/deeppath\_scratch/Pre\_processing\_2nd\_Feb21/

--Magnification=20.0

--MagDiffAllowed=0

--SortingOption=10

--PatientID=-1

--nSplit 0

--PercentTest=15

--PercentValid=15

--outFilenameStats=/nfs/home/xwang/Transfer\_Learning\_April/r2\_result/test\_3210k/out\_filename\_Stats.txt

nohup python /nfs/home/xwang/DeepPATH/DeepPATH\_code/00\_preprocessing/0d\_SortTiles.py --SourceFolder=/nfs6/deeppath\_scratch/Pre\_processing\_2nd\_Feb21/ --Magnification=20.0 --MagDiffAllowed=0 --SortingOption=10 --PatientID=-1 --nSplit 0 --PercentTest=15 --PercentValid=15 --outFilenameStats=/nfs/home/xwang/Transfer\_Learning\_April/r2\_result/test\_3210k/out\_filename\_Stats.txt

result:

Pre\_processing\_2nd\_Feb21 612819

Pre\_processing\_2nd\_Feb21\_test 91923

Pre\_processing\_2nd\_Feb21\_train 427568

Pre\_processing\_2nd\_Feb21\_valid 9332

Pre\_processing\_2nd\_Feb21\_test 0.1500002447704787

Pre\_processing\_2nd\_Feb21\_train 0.6977068269750122

Pre\_processing\_2nd\_Feb21\_valid 0.15229292825450907

Pre\_processing\_2nd\_Feb21 518

Pre\_processing\_2nd\_Feb21\_test 85

Pre\_processing\_2nd\_Feb21\_train 363

Pre\_processing\_2nd\_Feb21\_valid 70

#---------------------------------------------------------------------------

* Convert to TFRecord:
* Only 18/70{VAlid} been converted to TFRecord because only 109 Images -patients have KRAS label. [label file only contain images/patietns with KRAS mutation.]
* try to use all images with 2 class labels for transfer training from model 1b/ 10-way classifier. Class 1 = KRAS mutated; Class 2 = No KRAS mutation.
* Need to re-create **<labelref\_r3.txt>** and **<labels\_r3.txt>.** 
  + - /nfs/home/xwang/Transfer\_Learning\_April/model\_1b/ labelref\_r3\_2class.txt
    - /nfs/home/xwang/Transfer\_Learning\_April/model\_1b/labels\_r3\_2class.txt

# valid

cd /nfs/home/xwang/Transfer\_Learning\_April/model\_1b/r3\_2class\_TFRecord\_valid

nohup python /nfs/home/xwang/DeepPATH/DeepPATH\_code/00\_preprocessing/TFRecord\_multi\_Classes/build\_TF\_test\_multiClass.py --directory=/nfs/home/xwang/Transfer\_Learning\_April/r3\_LUAD\_sorted/Pre\_processing\_2nd\_Feb21/ --output\_directory=/nfs/home/xwang/Transfer\_Learning\_April/model\_1b/r3\_2class\_TFRecord\_valid/ --num\_threads=1 --one\_FT\_per\_Tile=False --ImageSet\_basename='valid' --labels\_names=/nfs/home/xwang/Transfer\_Learning\_April/model\_1b/labelref\_r3\_2class.txt --labels=/nfs/home/xwang/Transfer\_Learning\_April/model\_1b/labels\_r3\_2class.txt --PatientID=14

Results:

Wrote 70 images/ 70.

# test

cd /nfs/home/xwang/Transfer\_Learning\_April/model\_1b/r3\_2class\_TFRecord\_test

nohup python /nfs/home/xwang/DeepPATH/DeepPATH\_code/00\_preprocessing/TFRecord\_multi\_Classes/build\_TF\_test\_multiClass.py --directory=/nfs/home/xwang/Transfer\_Learning\_April/r3\_LUAD\_sorted/Pre\_processing\_2nd\_Feb21/ --output\_directory=/nfs/home/xwang/Transfer\_Learning\_April/model\_1b/r3\_2class\_TFRecord\_test/ --num\_threads=1 --one\_FT\_per\_Tile=False --ImageSet\_basename='test' --labels\_names=/nfs/home/xwang/Transfer\_Learning\_April/model\_1b/labelref\_r3\_2class.txt --labels=/nfs/home/xwang/Transfer\_Learning\_April/model\_1b/labels\_r3\_2class.txt --PatientID=14

Results:

Wrote 85 images/85

# Train:

cd /nfs/home/xwang/Transfer\_Learning\_April/model\_1b/r3\_2class\_TFRecord\_train/

nohup python /nfs/home/xwang/DeepPATH/DeepPATH\_code/00\_preprocessing/TFRecord\_multi\_Classes/build\_image\_data\_multiClass.py --directory=/nfs/home/xwang/Transfer\_Learning\_April/r3\_LUAD\_sorted/Pre\_processing\_2nd\_Feb21/ --output\_directory=/nfs/home/xwang/Transfer\_Learning\_April/model\_1b/r3\_2class\_TFRecord\_train/ --train\_shards=1024 --validation\_shards=128 --num\_threads=16 --labels\_names=/nfs/home/xwang/Transfer\_Learning\_April/model\_1b/labelref\_r3\_2class.txt --labels=/nfs/home/xwang/Transfer\_Learning\_April/model\_1b/labels\_r3\_2class.txt --PatientID=14

Results:

427568 tiles/images // 427568

#--------------------------------------

Inside 01\_training/xClasses

<https://www.quora.com/Why-is-it-better-to-use-Softmax-function-than-sigmoid-function>

<used SoftMax in this case vs. sigmoid. >

#--------------------------------------

nohup bazel-bin/inception/imagenet\_train --num\_gpus=4 --batch\_size=400 --train\_dir=/nfs/home/xwang/Transfer\_Learning\_April/model\_1b/transfer\_learning\_model1b\_2CLass/ --data\_dir=/nfs/home/xwang/Transfer\_Learning\_April/model\_1b/r3\_2class\_TFRecord\_train/ --pretrained\_model\_checkpoint\_path=/nfs6/deeppath\_models/checkpoints/run1b\_10way\_MutationClassifier/model.ckpt-81500 --fine\_tune=True --initial\_learning\_rate=0.001 --ClassNumber=2 --mode='0\_softmax' --NbrOfImages=427568 --save\_step\_for\_chekcpoint=1069 --max\_steps=106901

result:

fail.

Invalid argument: Key: Image/class/label/.

#--------------------------------------

1. Try with sigmoid mode. With above commend.
2. Try with 1-5 KRAS; 6-10 NOT\_KRAS class. For transfer learning
3. Try to keep the same class but assign NOT\_KRAS mutation to any of the other gene mutations. -> just have to be careful when calculate the AUC/ final results.

For Final presentation,

1. Go over general pathology. Slices, preprocessing. Why can’t we just through images into model.
2. Go over paper, what I did so far.
3. 总结

For general advice for grad students:

1. GitHub maintenance. – write readme file for notes.
   1. For example, for DeepPath project, make sure to upload the file for documentations for how to run this model, etc.
2. For the ML/DL methods you put down on your resume, make sure you can teach other people about the algorithm. Not just how to use packages but the theoretical.
3. Look into library documentations so that you can adapt the same style for your own code.
   1. E.g. <https://cookiecutter.readthedocs.io/en/1.7.2/>
   2. <https://github.com/audreyr/cookiecutter-pypackage>

May 01, 2020

#--------------------------------------

#--------------------------------------

Inside 01\_training/xClasses

* Try with sigmoid for transfer learning.

nohup bazel-bin/inception/imagenet\_train --num\_gpus=4 --batch\_size=400 --train\_dir=/nfs/home/xwang/Transfer\_Learning\_April/model\_1b/transfer\_learning\_model1b\_2CLass/ --data\_dir=/nfs/home/xwang/Transfer\_Learning\_April/model\_1b/r3\_2class\_TFRecord\_train/ --pretrained\_model\_checkpoint\_path=/nfs6/deeppath\_models/checkpoints/run1b\_10way\_MutationClassifier/model.ckpt-81500 --fine\_tune=True --initial\_learning\_rate=0.001 --ClassNumber=2 --mode='1\_sigmoid' --NbrOfImages=427568 --save\_step\_for\_chekcpoint=1069 --max\_steps=106901

results:

Tuesday- May 5, 2020 => 6 model checkpoints <= run validation.