1. AWS connection:

Use pem file and ipv4 dns for connection

use security-group-…. Add ip address when using new devices

1. Mobaxterm ssh link:

ubuntu@ip-10-0-0-9:~$ ssh-keygen

Generating public/private rsa key pair.

Enter file in which to save the key (/home/ubuntu/.ssh/id\_rsa):

Enter passphrase (empty for no passphrase):

Enter same passphrase again:

Your identification has been saved in /home/ubuntu/.ssh/id\_rsa

Your public key has been saved in /home/ubuntu/.ssh/id\_rsa.pub

The key fingerprint is:

SHA256:CurVynb1qChdXYMVlDRJ3EU4ZKAK14zBZTszKYjUvQI ubuntu@ip-10-0-0-9

The key's randomart image is:

+---[RSA 3072]----+

| .. o..o=BBo+o |

| .E...o\* +=o+ |

| ..o +.Xo . |

| .o.o.+o |

| ....S. . |

| . o..o |

| ....o. o |

| ..ooo. . . |

| .o+... |

+----[SHA256]-----+

ubuntu@ip-10-0-0-9:~$ cat ~/.ssh/id\_rsa.pub

ssh-rsa  ubuntu@ip-10-0-0-9

ubuntu@ip-10-0-0-9:~$ ^C

The last part was added into github SSH to establish the connection

Ensure the SSH key is set up correctly and can authenticate with GitHub:

ssh -T git@github.com

sometimes, after adding the ssh to github, when ‘git fetch’, still get error:

[t18155cb@login2 [csf3] ACIST]$ git fetch Username for 'https://github.com': BianChang Password for 'https://BianChang@github.com': fatal: Authentication failed for 'https://github.com/BianChang/ACIST.git/'

The error indicates that the repository's remote URL is set to use HTTPS instead of SSH. Since you've already added your SSH key, you need to update the remote URL to use the SSH protocol.

Follow these steps to resolve the issue:

**1. Check the Current Remote URL**

Run this command to see the remote URL configuration:

bash

Copy code

git remote -v

You will see output like:

plaintext

Copy code

origin https://github.com/BianChang/ACIST.git (fetch)

origin https://github.com/BianChang/ACIST.git (push)

This confirms that the repository is using HTTPS.

**2. Update the Remote URL to Use SSH**

Update the remote URL to use SSH instead of HTTPS:

bash

Copy code

git remote set-url origin git@github.com:BianChang/ACIST.git

Now verify the change:

bash

Copy code

git remote -v

The output should now look like:

plaintext

Copy code

origin git@github.com:BianChang/ACIST.git (fetch)

origin git@github.com:BianChang/ACIST.git (push)

1. Create Image from docker file:

ubuntu@ip-10-0-0-9:~/chang/infrastructure$ cd docker/

ubuntu@ip-10-0-0-9:~/chang/infrastructure/docker$ docker build -f Dockerfile.spotlight\_base -t spotlight\_base:latest .

check available Images: docker image list

REPOSITORY TAG IMAGE ID CREATED SIZE

spotlight\_base latest 928562bfdd6f 34 seconds ago 9.48GB

1. Create a container from the image:  
   Usage: docker run [OPTIONS] IMAGE [COMMAND] [ARG...]

Create and run a new container from an image

ubuntu@ip-10-0-0-9:~/chang$ docker run -v .:/workspace --gpus all -it spotlight\_base:latest

root@9b4ae2a4bd9f:/workspace# ls

infrastructure

docker ps: check containers

1. When running time is long, use screen to run jobs:

In root (container):

1. screen
2. screen -ls (check screens)
3. screen -r screenname (reactivate screen)
4. Ctrl + A ++D detach the screen
5. Windows VS code connects AWS-EC2

Host EC2

  HostName ec2-13-40-72-34.eu-west-2.compute.amazonaws.com

  User ubuntu

  IdentityFile C:\Users\chang\.ssh\ml\_ec2\_jan\_2024.pem

Note that the pem security: advanced-remove all, disable inheritage, leave only one, the User of the key (enter username (e.g. chang) in the check name box)

Hugging face write token for spotlight: hf\_UMZkjtuIKQizKDtsbBAdkapPvgKjwJADdg

11/7/2024

1. Use git lfs to handle large file commits in GitHub repository:
   1. Install git lfs
   2. git lfs install to initialize
   3. git track large files..:
      1. git lfs track "\*.bin"
      2. git lfs track "\*.safetensors"
   4. commit
2. after create a new branch, when operating on a different machine/server, use: git fetch origin to fetch the new created branch; may required enter the passphrase (bian1219), then you can use git branch -r to see the remote branches, and use git branch -a to see all; use git checkout <branchname > to checkout

11/11/2024

1. Every time into the container, need to pip install . lymph node triage and grout
2. For feature extraction test: python /workspace/chang/lymph-node-triage/lymphtriage/preprocessing/feature\_extraction.py --dataset /workspace/data\_dev\_s3/dev\_dataset\_s3.csv --s3-bucket "spotlight-pathology-lymphnode-datasets" --filepath-col filepath --label-col diagnostic.category.id --artifacts-dir /workspace/data\_dev\_s3/slide-artefacts --logs /workspace/chang/tmp/feature\_extraction.out --artifacts-format png

11/15/2024

1. Feature\_extraction.py of the cohort 1\_2:

python /workspace/chang/lymph-node-triage/lymphtriage/preprocessing/feature\_extraction.py --dataset /workspace/cohort\_1\_2\_partition/cohorts\_1\_2\_database\_20240912\_145230.csv --s3-bucket "spotlight-pathology-lymphnode-datasets" --filepath-col file\_path --label-col lymphoma\_diagnostic\_category\_name --tile 224 --tile-stride 224 --artifacts-dir /workspace/chang/histoqc-slide-artifacts/manchester-2022/ --logs /workspace/chang/tmp/feature\_extraction\_virchow-cohort-1-2.out --artifacts-format png --extractor virchow --dont-extract-if-exists

11/20/2024

1. docker run -v .:/histoqc\_space --gpus all -it histotools/histoqc:master /bin/bash
2. python histoqc\_helper.py --output-dir /histoqc\_space/chang/tmp/histoqc\_Salford\_output --ini-file /histoqc\_space/chang/lymph-node-triage/histoqc\_config\_manchester.ini

11/27/2024

1. set up a new branch:

git branch temp\_vit

git checkout temp\_vit

git push -u origin temp\_vit

12/2/2024

1. python /workspace/chang/lymph-node-triage/lymphtriage/preprocessing/feature\_extraction.py --dataset /workspace/chang/tmp/NIHR\_csv\_test/s3\_file\_list.csv --s3-bucket do --filepath-col filepath --label-col label --artifacts-dir /workspace/chang/tmp/histoqc\_nihr\_test --logs /workspace/chang/tmp/feature\_extraction\_vit-nihr\_test.out--artifacts-format png --extractor vit --tile 224 --tile-stride 224 --dont-extract-if-exists
2. docker run -v .:/leedsvit --gpus all -it spotlight\_base:latest

12/9/2024

1. aws s3 sync /home/ubuntu/chang/tmp/feature\_extraction\_virchow-cohort-1-2.out/ s3://spotlight-pathology-lymphnode-datasets/features/feature-manchester-newcastle-virchow/

copy files (subfolders and contents) in to s3 folder, but not the feature\_extraction\_virchow-cohort-1-2.out itself, the key is to add a ‘/’ to only copy contents, not the folder.

python /workspace/chang/lymph-node-triage/lymphtriage/scripts\_train\_eval/clam/train\_clam.py --train-set /workspace/cohort\_1\_2\_partition/train\_cohorts\_1\_2\_column\_patient\_id\_20240912\_173126.csv --val-set /workspace/cohort\_1\_2\_partition/val\_cohorts\_1\_2\_column\_patient\_id\_20240912\_173126.csv --test-set /workspace/cohort\_1\_2\_partition/test\_cohorts\_1\_2\_column\_patient\_id\_20240912\_173126.csv --filepath-col file\_path --label-col lymphoma\_diagnostic\_category\_name --feature-dirs /workspace/chang/tmp/feature\_extraction\_virchow-cohort-1-2.out/features --extractor vit --instance-loss svm --max-epochs 20 --logs /workspace/chang/tmp/lymphnode-reproduce-virchow\_manchester\_newcastle --plot y --plot-slide-dir /workspace/chang/tmp/lymphnode-reproduce-virchow\_manchester\_newcastle\_plot --max-files-fetch 5 --experiment-name lymphnode-reproduce-virchow\_manchester\_newcastle --tracking-uri-mlflow [http://10.0.0.146:500](http://10.0.0.146:500/) --repeats 1 --augment-mixup n --classes 4

python /workspace/chang/lymph-node-triage/lymphtriage/preprocessing/feature\_extraction.py --dataset /workspace/chang/tmp/Leeds\_csv/s3\_file\_list.csv --s3-bucket "spotlight-pathology-lymphnode-datasets" --filepath-col filepath --label-col label --artifacts-dir /workspace/chang/tmp/histoqc\_Leeds\_output --logs /workspace/chang/tmp/feature\_extraction\_vit-Leeds.out2 --artifacts-format png --extractor vit --tile 224 --tile-stride 224 --dont-extract-if-exists

/NPIC\_24014640\_180308.svs

NPIC\_24013852\_164627

**Option 2: Stash Changes**

If you want to save the changes temporarily without committing, stash them:

bash

Copy code

git stash

After stashing, switch branches:

bash

Copy code

git checkout <branch-name>

You can reapply the stashed changes later with:

bash

Copy code

git stash pop

To create a new branch and move your changes to it without affecting the original main branch, follow these steps:

1. **Check Your Changes**: Ensure you have uncommitted changes on the main branch.

bash

Copy code

git status

1. **Create and Switch to a New Branch**: Create a new branch and switch to it. For example, to create a branch called new-feature:

bash

Copy code

git checkout -b new-feature

1. **Commit Your Changes (Optional)**: If you want to save the changes to the new branch:

bash

Copy code

git add .

git commit -m "Describe the changes here"

Alternatively, if you don’t want to commit yet, your changes will still remain in the working directory of the new branch.

1. **Verify Your Branch**: Confirm that you are now on the new-feature branch:

bash

Copy code

git branch

Now your changes are on the new-feature branch, and the main branch remains unaffected.

2025.01.6

1. MLflow Virchow training test:

root@4e377da96a2d:/workspace/chang/lymph-node-triage# python /workspace/chang/lymph-node-triage/lymphtriage/scripts\_train\_eval/clam/train\_clam.py --train-set /workspace/cohort\_1\_2\_partition/train\_cohorts\_1\_2\_column\_patient\_id\_20240912\_173126.csv --val-set /workspace/cohort\_1\_2\_partition/val\_cohorts\_1\_2\_column\_patient\_id\_20240912\_173126.csv --test-set /workspace/cohort\_1\_2\_partition/test\_cohorts\_1\_2\_column\_patient\_id\_20240912\_173126.csv --filepath-col file\_path --label-col lymphoma\_diagnostic\_category\_name --feature-dirs /workspace/chang/tmp/feature\_extraction\_virchow-cohort-1-2.out/features --extractor vit --instance-loss svm --max-epochs 20 --logs /workspace/chang/tmp/lymphnode-reproduce-virchow\_manchester\_newcastle --plot y --plot-slide-dir /workspace/chang/tmp/lymphnode-reproduce-virchow\_manchester\_newcastle\_plot --max-files-fetch 5 --experiment-name lymphnode-reproduce-virchow\_manchester\_newcastle --tracking-uri-mlflow http://10.0.0.146:500 --repeats 1 --augment-mixup n --classes 4

Traceback (most recent call last):

File "/workspace/chang/lymph-node-triage/lymphtriage/scripts\_train\_eval/clam/train\_clam.py", line 16, in <module>

from lymphtriage.setup\_training import new\_training\_clam, load\_model\_clam

File "/opt/conda/lib/python3.10/site-packages/lymphtriage/setup\_training.py", line 11, in <module>

import topk as topk

ModuleNotFoundError: No module named 'topk'

2025/01/09

1. Record MD5 hash:

Temporary file path: /tmp/tmpo8lf28g2 - NPIC\_24013852\_164627.svs

MD5 hash: 2ccf8963ac00589eb2c35d25d1a34c93

PyVIPS resolution (xres): 3801.8187901091883

grout WSI res\_zero: 0.263032

OpenSlide mpp-x value: 0.26303199999999999

S3 Download: /tmp/tmpo0miz574 - NPIC\_24013852\_164627.svs, size: 270794752, MD5: dbfae04d0840b1d4629259412abe30dc

DOWNLOADED ON LOCAL : MD5: 2ccf8963ac00589eb2c35d25d1a34c93

2025/1/22

1. Training vit with 3 repeats: python /workspace/chang/lymph-node-triage/lymphtriage/scripts\_train\_eval/clam/train\_clam.py --train-set /workspace/cohort\_1\_2\_partition/train\_cohorts\_1\_2\_column\_patient\_id\_20240912\_173126.csv --val-set /workspace/cohort\_1\_2\_partition/val\_cohorts\_1\_2\_column\_patient\_id\_20240912\_173126.csv --test-set /workspace/cohort\_1\_2\_partition/test\_cohorts\_1\_2\_column\_patient\_id\_20240912\_173126.csv --filepath-col file\_path --label-col lymphoma\_diagnostic\_category\_name --feature-dirs /workspace/chang/tmp/feature\_extraction\_manchester\_2022\_vit/features --feature-dirs /workspace/chang/tmp/feature\_extraction\_newcastle\_2021\_vit/features --extractor vit --instance-loss svm --max-epochs 20 --logs /workspace/chang/tmp/lymphnode-reproduce-vit\_manchester\_newcastle --plot n --plot-slide-dir /workspace/chang/tmp/lymphnode-reproduce-vit\_manchester\_newcastle\_plot --max-files-fetch 5 --experiment-name lymphnode-reproduce-vit\_manchester\_newcastle --tracking-uri-mlflow http://10.0.0.146:500 --repeats 3 --augment-mixup n --classes 4

1/27/2025

1. Histoblur detect -f "NPIC\_23051415\_083347.svs" -o blur\_quality\_control -m pretrained\_model/HE\_20X.pth
2. docker run --shm-size=12g --gpus all -t -i -v /workspace/app petroslk/histoblur:latest HistoBlur detect -f 'NPIC\_24013852\_164627.svs' -m /HistoBlur/pretrained\_model/HE\_20X.pth -o results

2/12/2025

Lost ssh mount in docker:

docker run -v .:/workspace -v /home/ubuntu/.ssh:/root/.ssh --gpus all -it spotlight\_base:latest

1. **Run the Container:**  
   Use the updated command to start your container.
2. **Verify the SSH Keys:**  
   Inside the container, run:

bash

ls -la /root/.ssh

You should see your SSH key files along with known\_hosts.

1. **Test SSH Access:**  
   Then test your SSH connection with:

ssh -T git@github.com

A successful output should say:

Hi <your\_username>! You've successfully authenticated, but GitHub does not provide shell access.

Once verified, you can run pip install . in your container and Git should be able to access your private repositories via SSH without any issues.

Can not find installed packages(pip install .): export PYTHONPATH=/workspace/slide-classification:$PYTHONPATH

2/13/2025

S3 manchester prefix: grout

manchester\_2022/HE/CHL/CORES/

manchester\_2022/HE/CHL/NODES/ grout

manchester\_2022/HE/DLBCL//

manchester\_2022/HE/DLBCL/NODES/

manchester\_2022/HE/FL/CORES/

manchester\_2022/HE/FL/NODES/

manchester\_2022/HE/Hyperplastic/CORES/

manchester\_2022/HE/Hyperplastic/NODES/

Newcastle 2021 prefix

newcastle\_2021/CHL/CORES/TEST/

newcastle\_2021/CHL/CORES/TRAIN/

newcastle\_2021/CHL/NODES/TEST/

newcastle\_2021/CHL/NODES/TRAIN/

newcastle\_2021/DLBCL/DLBCL/CORES/TEST/

newcastle\_2021/DLBCL/DLBCL/CORES/TRAIN/

newcastle\_2021/DLBCL/DLBCL/NODES/TEST/

newcastle\_2021/DLBCL/DLBCL/NODES/TRAIN/

newcastle\_2021/FOLLICULAR LYMPHOMA/CORES/TEST/

newcastle\_2021/FOLLICULAR LYMPHOMA/CORES/TRAIN/

newcastle\_2021/FOLLICULAR LYMPHOMA/NODES/TEST/

newcastle\_2021/FOLLICULAR LYMPHOMA/NODES/TRAIN/

newcastle\_2021/Hyperplastic/CORES/TEST/

newcastle\_2021/Hyperplastic/CORES/TRAIN/

newcastle\_2021/Hyperplastic/NODES/TEST/

newcastle\_2021/Hyperplastic/NODES/TRAIN/

export PYTHONPATH=/home/ubuntu/chang/slide-classification:$PYTHONPATH

/home/ubuntu/chang/slide-classification/slideclassification

Code for Newcastle histoblur: python slideclassification/helper\_tools/generate\_wsi\_mask\_histoblur\_from\_s3.py --bucket spotlight-pathology-lymphnode-datasets --prefix newcastle\_2021/CHL/CORES/TEST/ \

--prefix newcastle\_2021/CHL/CORES/TRAIN/ \

--prefix newcastle\_2021/CHL/NODES/TEST/ \

--prefix newcastle\_2021/CHL/NODES/TRAIN/ \

--prefix newcastle\_2021/DLBCL/DLBCL/CORES/TEST/ \

--prefix newcastle\_2021/DLBCL/DLBCL/CORES/TRAIN/ \

--prefix newcastle\_2021/DLBCL/DLBCL/NODES/TEST/ \

--prefix newcastle\_2021/DLBCL/DLBCL/NODES/TRAIN/ \

--prefix newcastle\_2021/FOLLICULAR LYMPHOMA/CORES/TEST/ \

--prefix newcastle\_2021/FOLLICULAR LYMPHOMA/CORES/TRAIN/ \

--prefix newcastle\_2021/FOLLICULAR LYMPHOMA/NODES/TEST/ \

--prefix newcastle\_2021/FOLLICULAR LYMPHOMA/NODES/TRAIN/ \

--prefix newcastle\_2021/Hyperplastic/CORES/TEST/ \

--prefix newcastle\_2021/Hyperplastic/CORES/TRAIN/ \

--prefix newcastle\_2021/Hyperplastic/NODES/TEST/ \

--prefix newcastle\_2021/Hyperplastic/NODES/TRAIN/\

--output-dir /home/ubuntu/chang/tmp/histoblur\_mask\_newcastle\_2021