NCHC Open Hackathon
Final Day

Smile Lab

OpenACC
More Science, Less Programming

Smile Lab

Leader : Pau-Choo Chung (NCKU)

Mentors : Ken Liao
Yang-Hsien Lin

Team Members:

Yu-Ping Gao (NCKU)

Jia-Xian Jian (NCKU)

Chin-Hua Liu (NCKU)

Yu-Cheng Chang (NCKU)

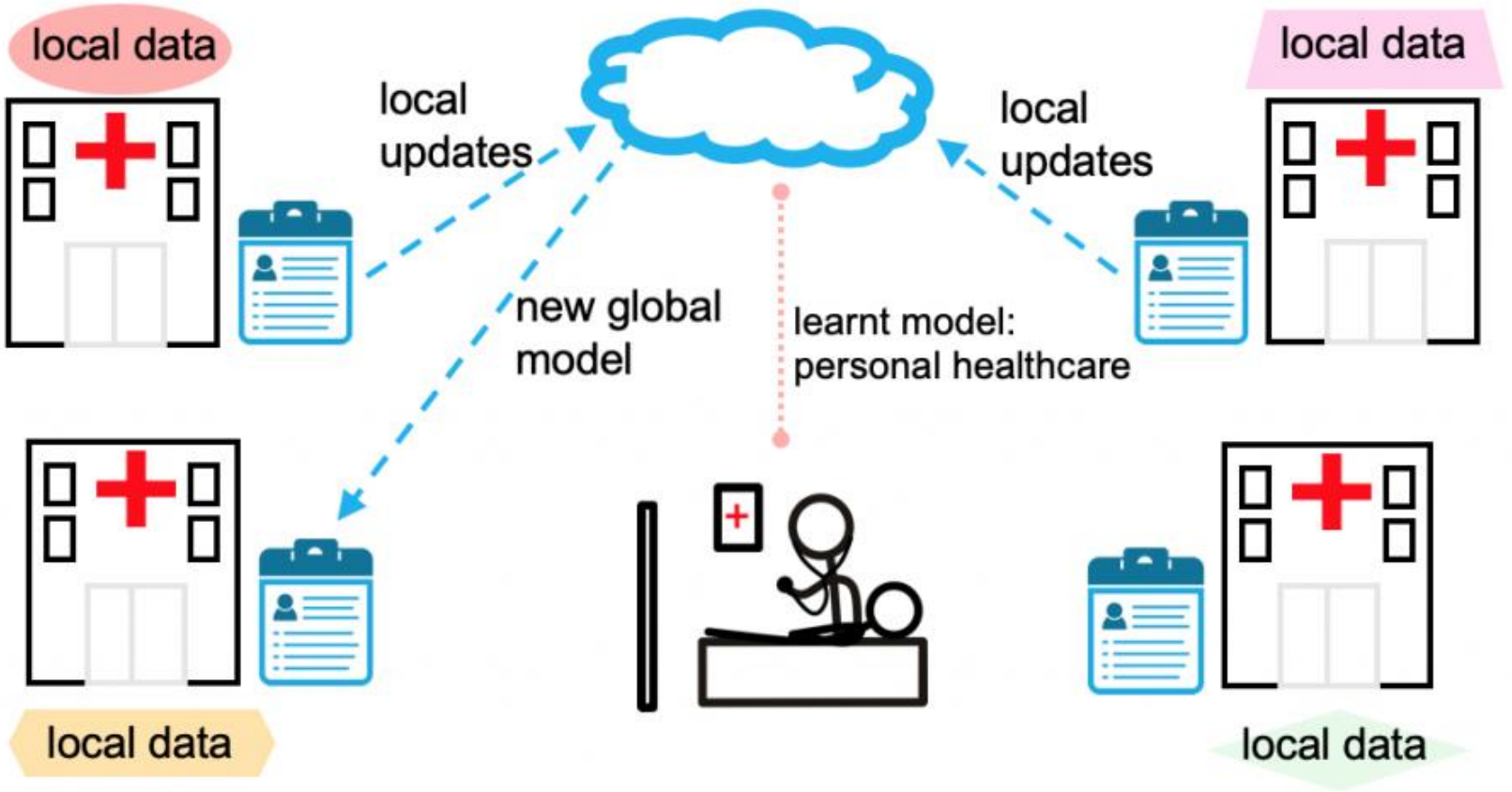
Kai-Xiang Liu (NCKU)

Yen-Jung Chiu (MCU)

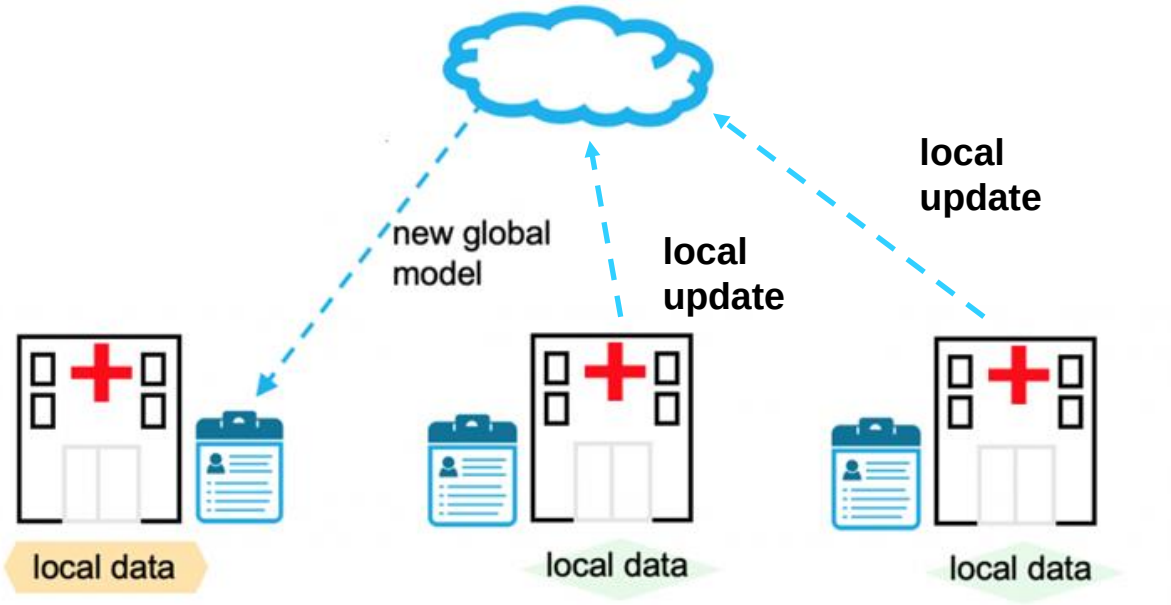
Po-Hao Hsu (NCHC)

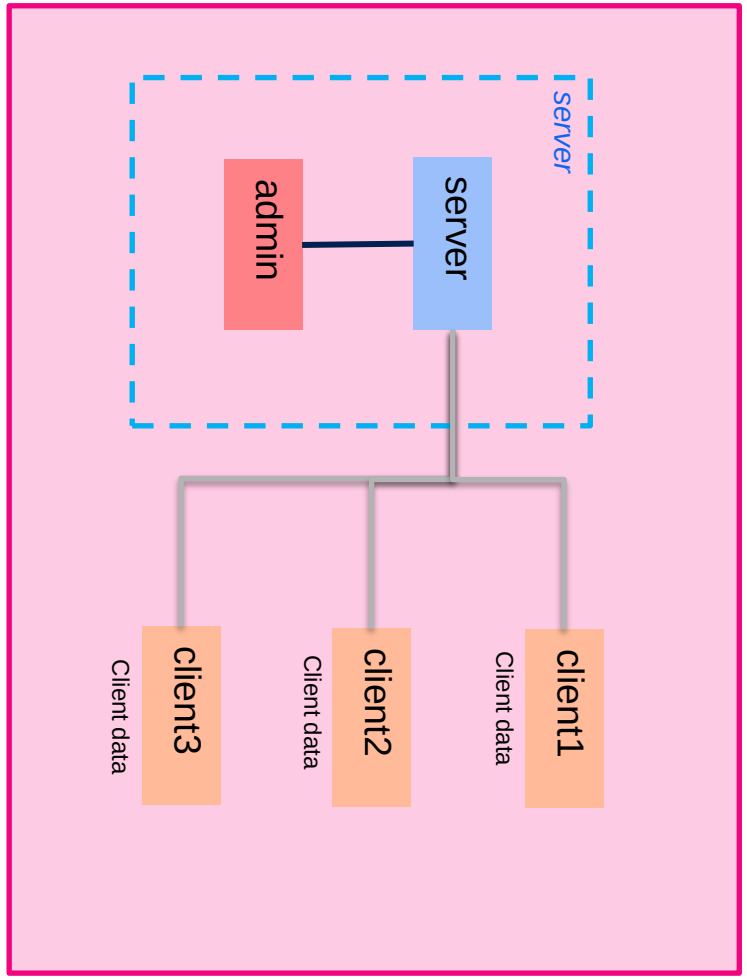
Chao-Chun Chuang (NCHC)

Federated Learning for pathology

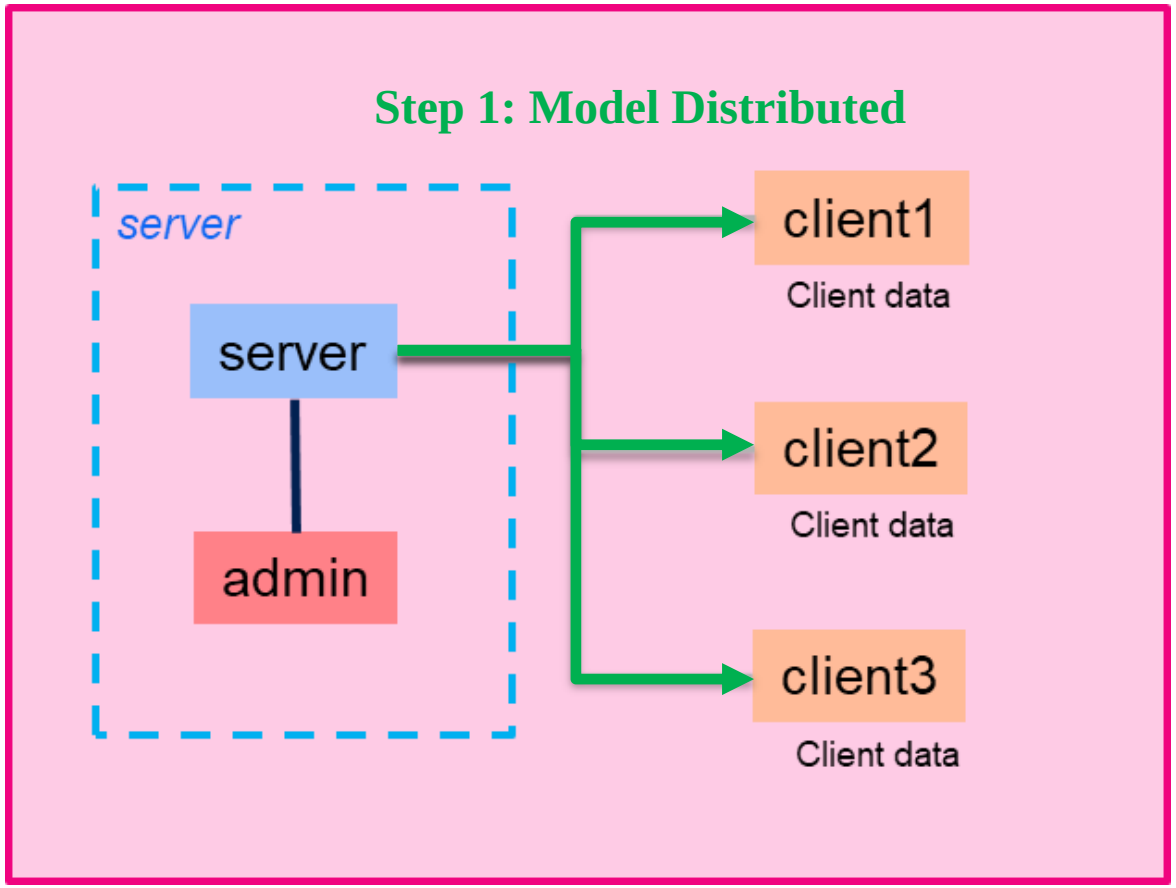


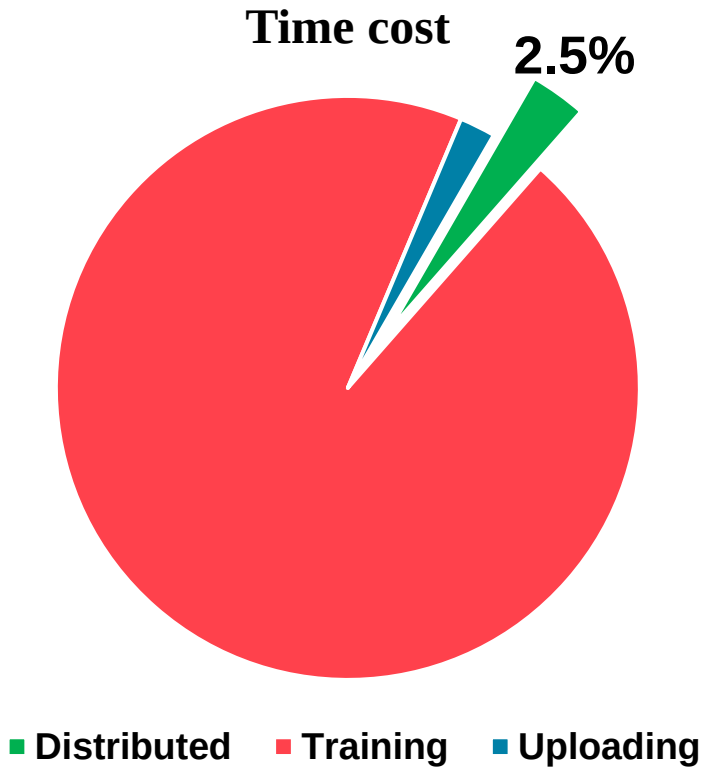
Method Architecture





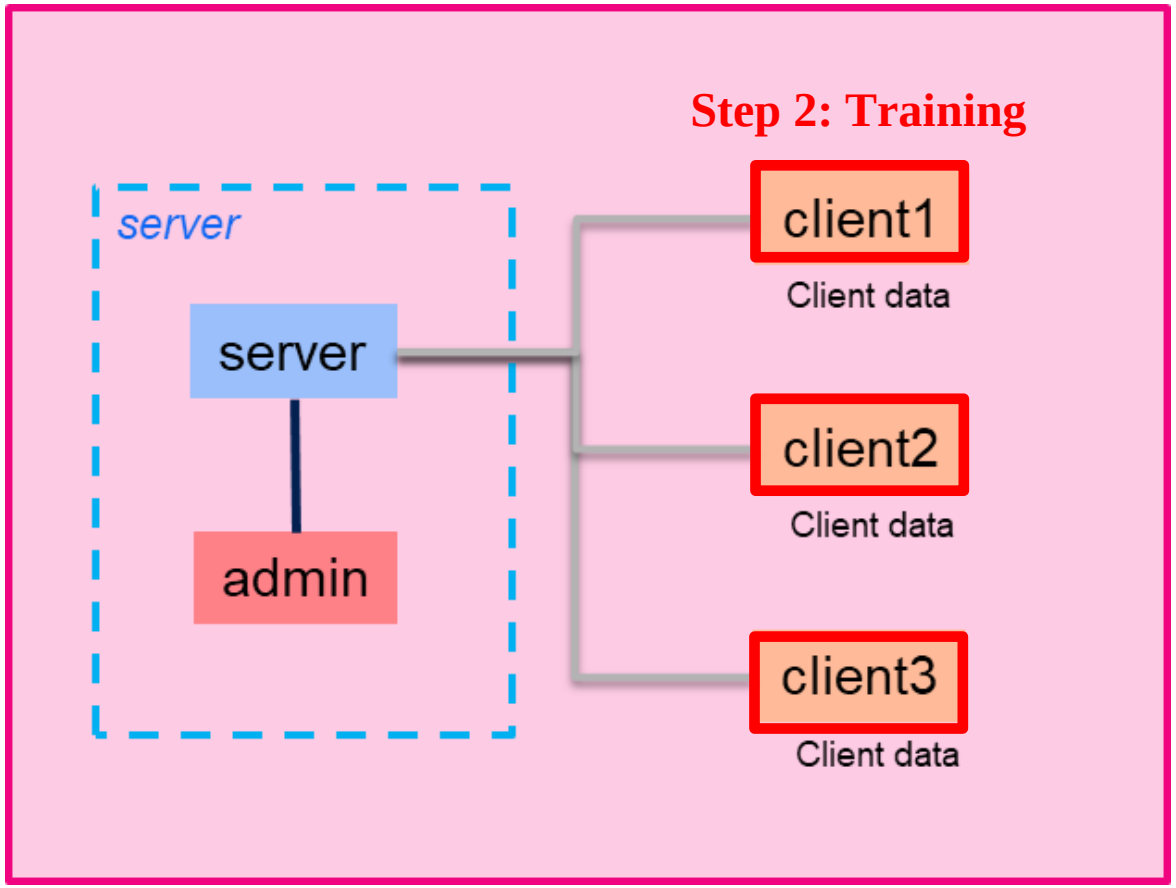
One Round Time Cost

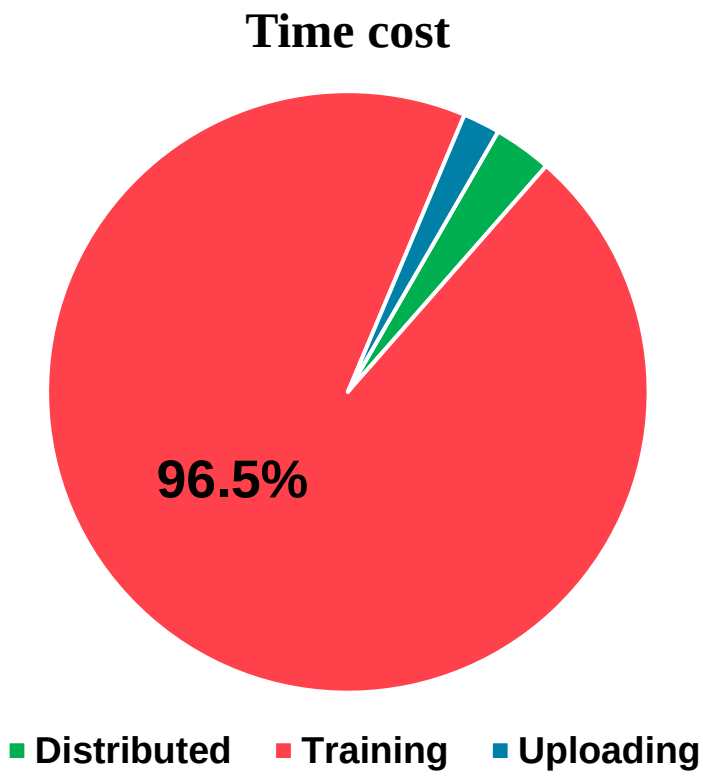




Each round 3 epochs
Each client uses 1 GPU (v100)
batch size: train:8, valid:1
Model size: 275.9MB

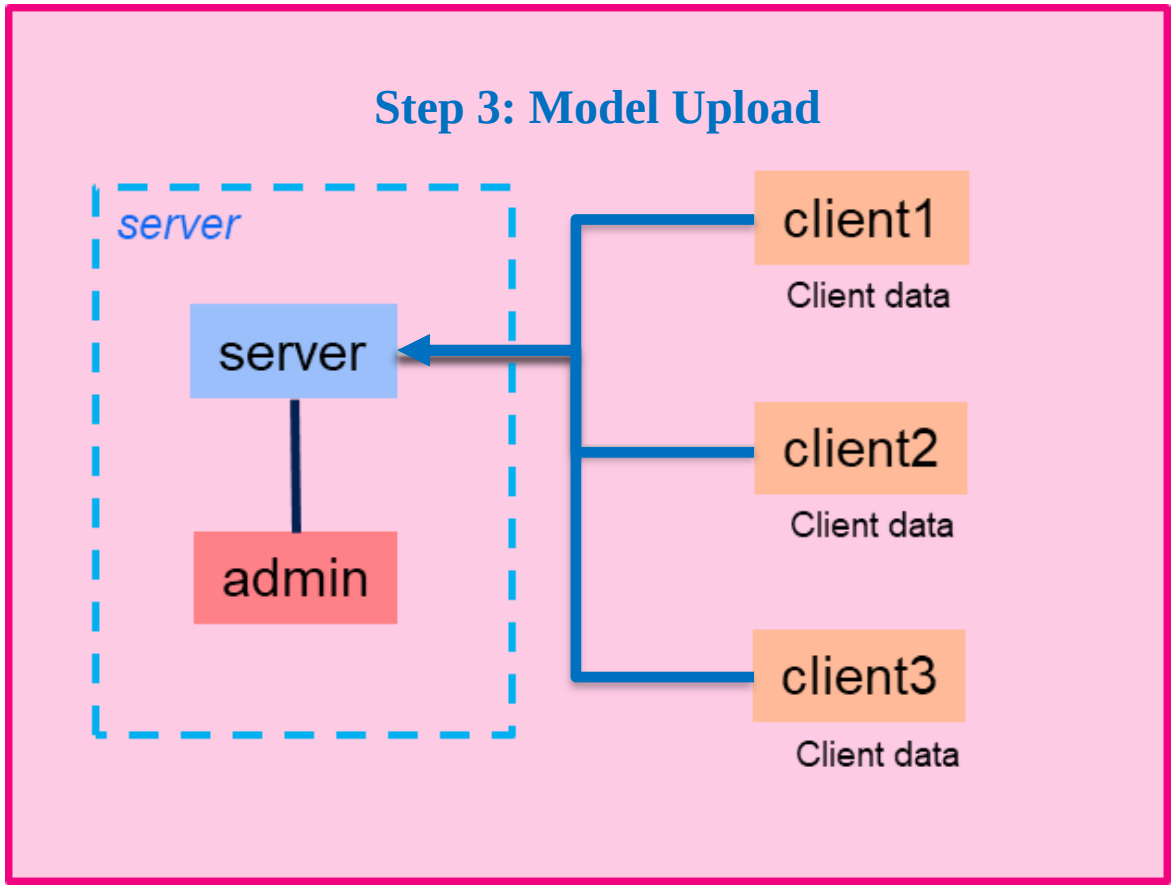
One Round Time Cost

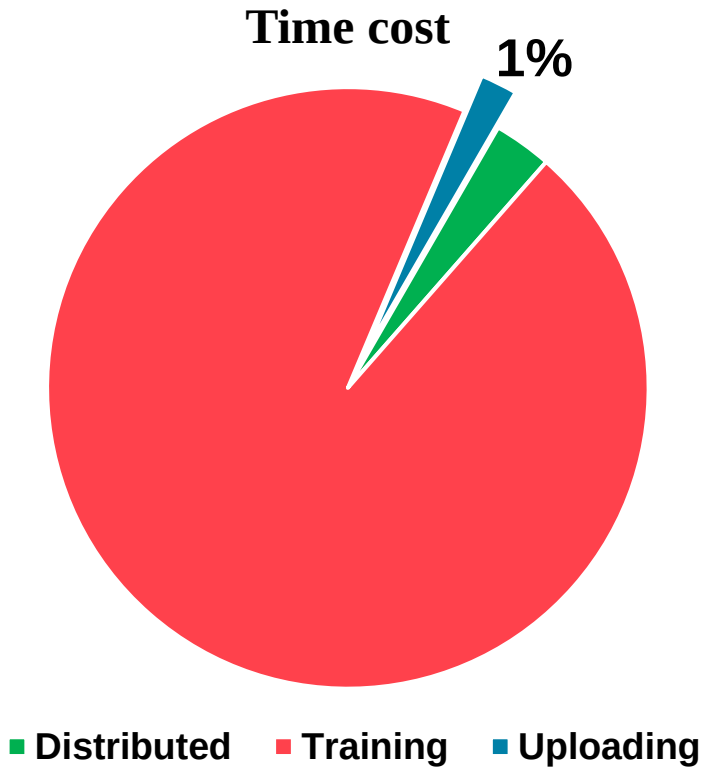




Each round 3 epochs
Each client uses 1 GPU (v100)
batch size: train:8, valid:1
Model size: 275.9MB

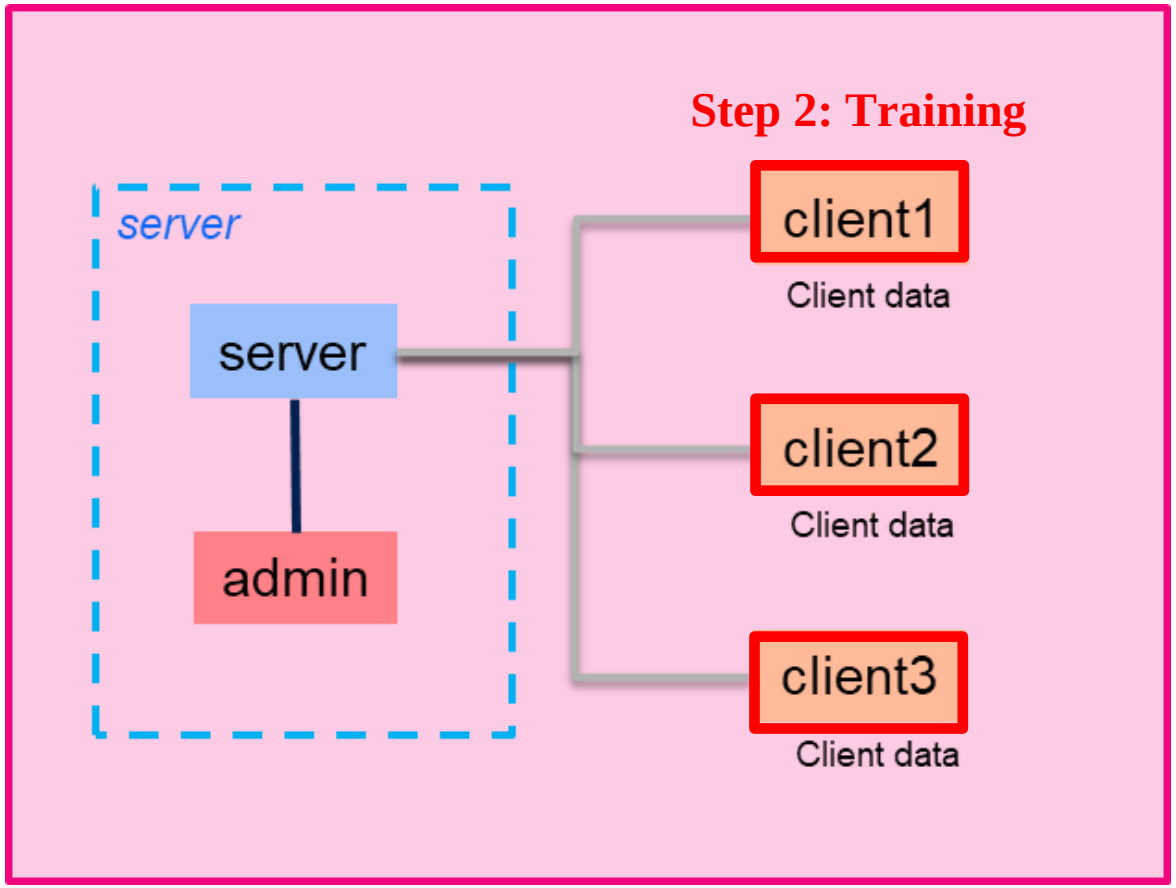
One Round Time Cost

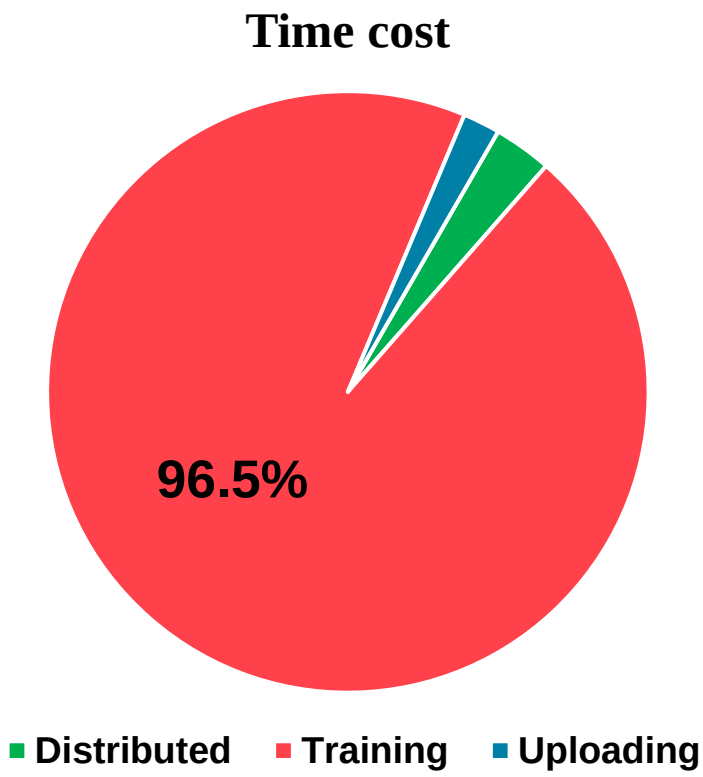




Each round 3 epochs
Each client uses 1 GPU (v100)
batch size: train:8, valid:1
Model size: 275.9MB

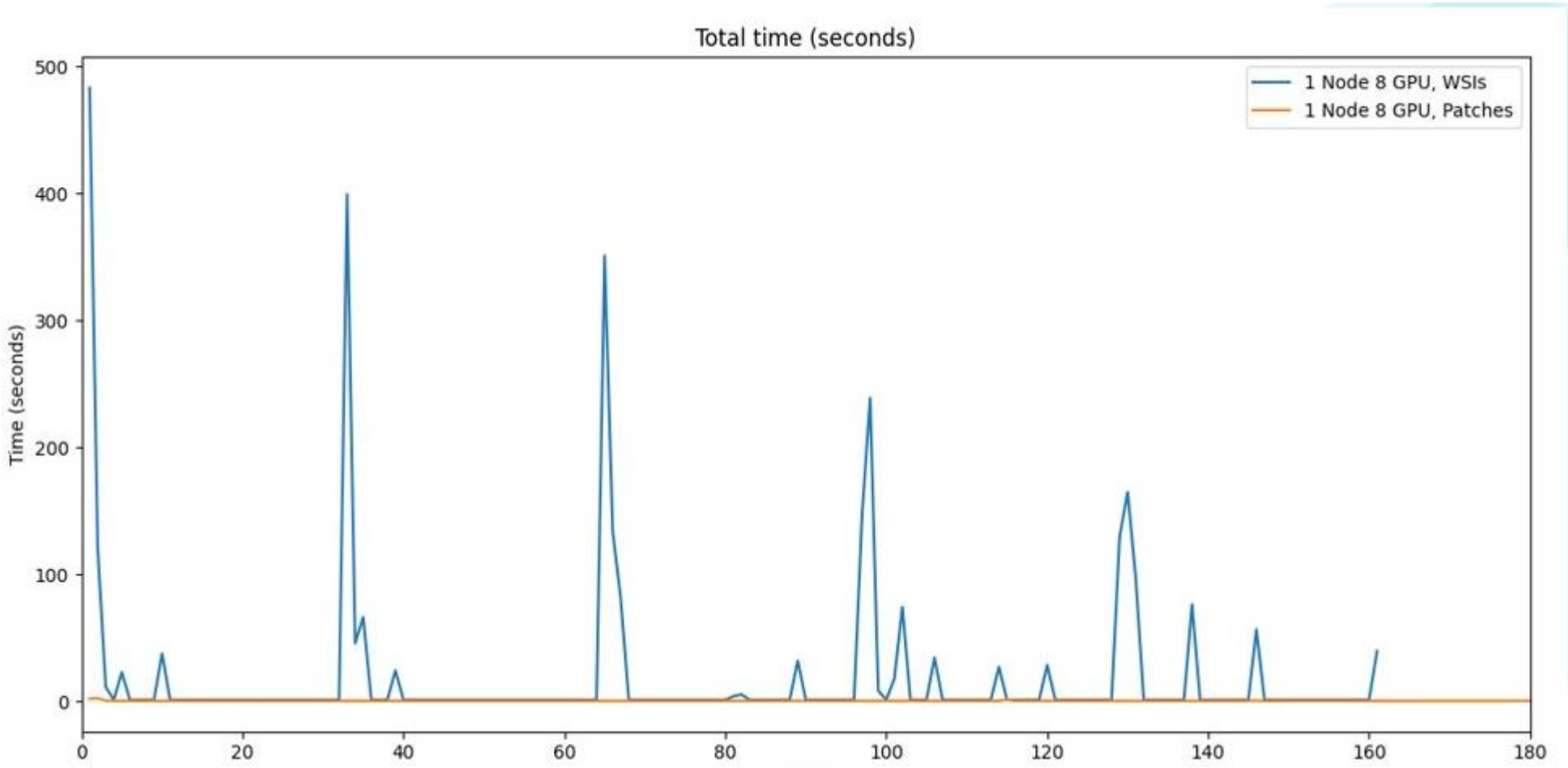
Primary Time Overhead of the Algorithm





Each round 3 epochs
Each client uses 1 GPU (v100)
batch size: train:8, valid:1
Model size: 275.9MB

Training Batch Processing Issue

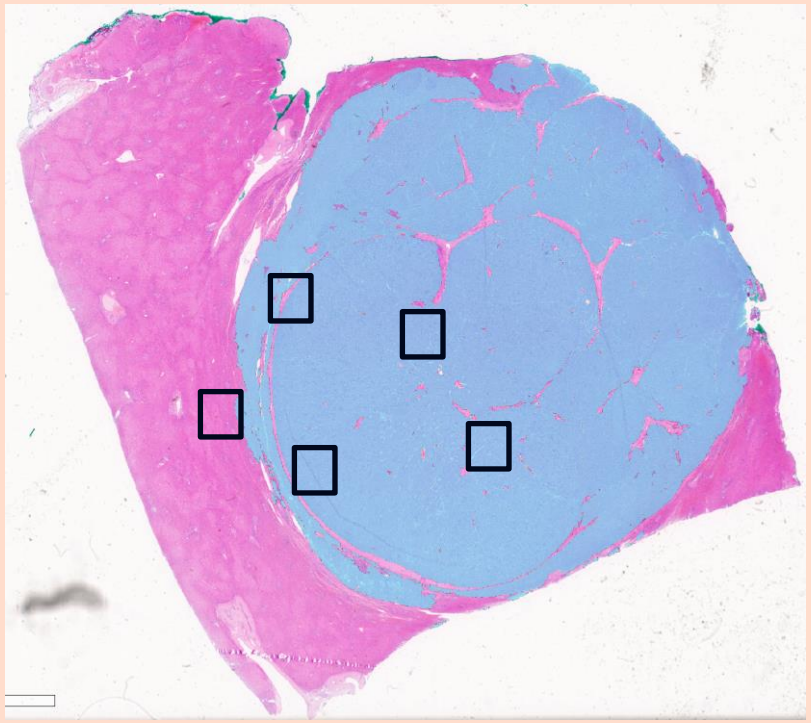


Data Process

WSI extracting patches

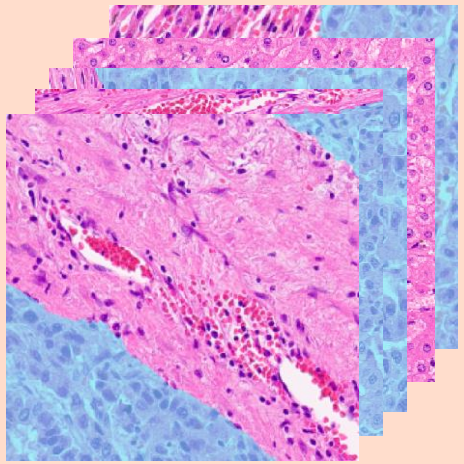
80000 pixels

60000 pixels



Extract patches with pkl file

512 pixels



512 pixels

PKL V.S. Feather

PKL file format

0, 0, white_background
512, 0, white_background
1024, 0, white_background
1536, 0, white_background
2048, 0, white_background
…

1.0613x

Feather file format

	key	label	coordinates
0	white_background	100B	[0, 0]
1	white_background	100B	[512, 0]
2	white_background	100B	[1024, 0]
3	white_background	100B	[1536, 0]
4	white_background	100B	[2048, 0]
...	...	...	...
9387	partial_tissue	100B	[53760, 32256]
9388	partial_tissue	100B	[52736, 32768]
9389	partial_tissue	100B	[53760, 32768]
9390	partial_tissue	100B	[66048, 33280]
9391	partial_tissue	100B	[66560, 33280]

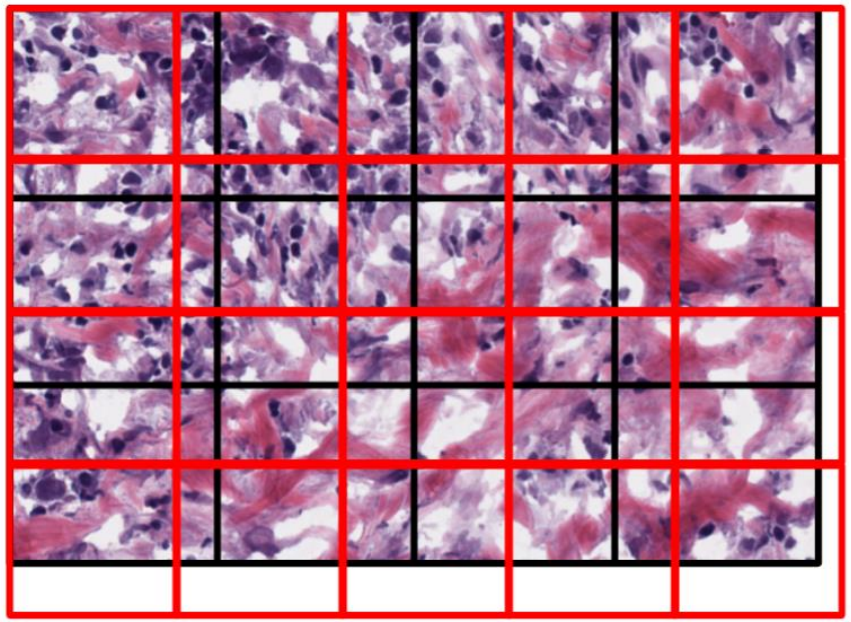
Numpy V.S. CuPy

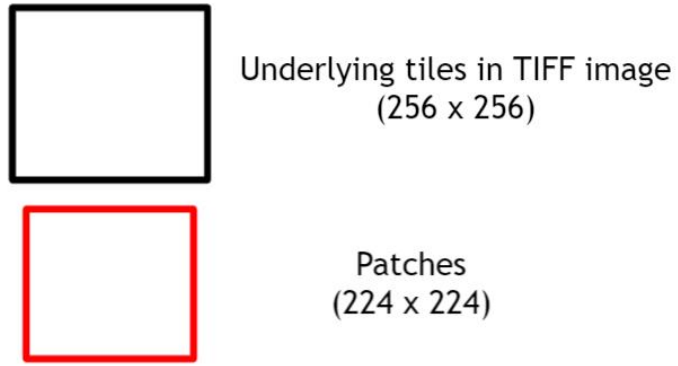
```python
class DINOLoss(nn.Module):
    def __init__(self, out_dim, ncrops, warmup_teacher_temp, teacher_temp,
                 warmup_teacher_temp_epochs, nepochs, student_temp=0.1,
                 center_momentum=0.9, tnum=1):
        super().__init__()
        self.student_temp = student_temp
        self.center_momentum = center_momentum
        self.ncrops = ncrops

        self.register_buffer("center", torch.zeros(tnum, 1, out_dim))
        # we apply a warm up for the teacher temperature because
        # a too high temperature makes the training instable at the beginning
        self.teacher_temp_schedule = np.concatenate((
    np.linspace(warmup_teacher_temp, teacher_temp, warmup_teacher_temp_epochs),
    np.ones(nepochs - warmup_teacher_temp_epochs) * teacher_temp
))
```

```python
import cupy as np  # 使用cupy替代numpy
from PIL import Image
import torch
import torch.nn as nn
import torch.distributed as dist
import torch.backends.cudnn as cudnn
import torch.nn.functional as F
from torchvision import datasets, transforms
from torchvision import models as torchvision_models
```

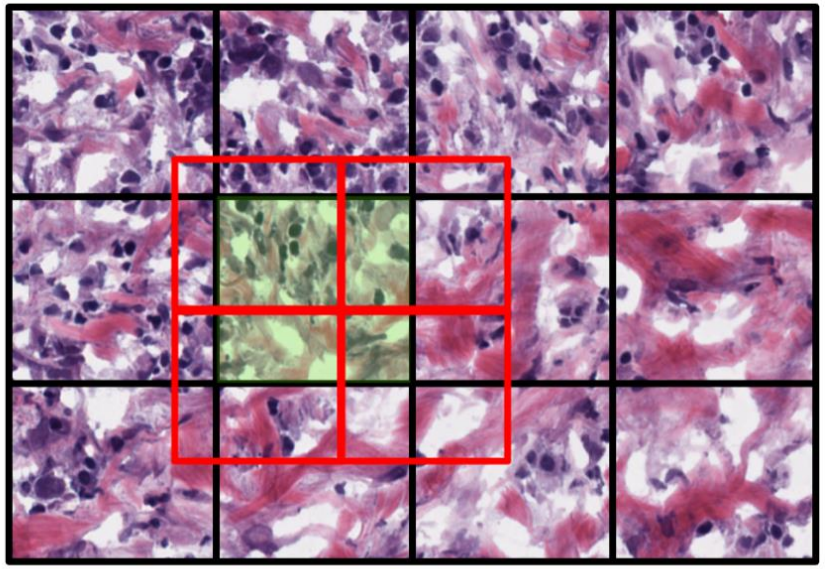
1.1811x

CuCIM Cache



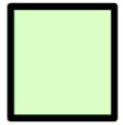


Pyvips V.S CuCIM Cache



Underlying tiles in TIFF image
(256 x 256)

Patches
(224 x 224)



is read four times!

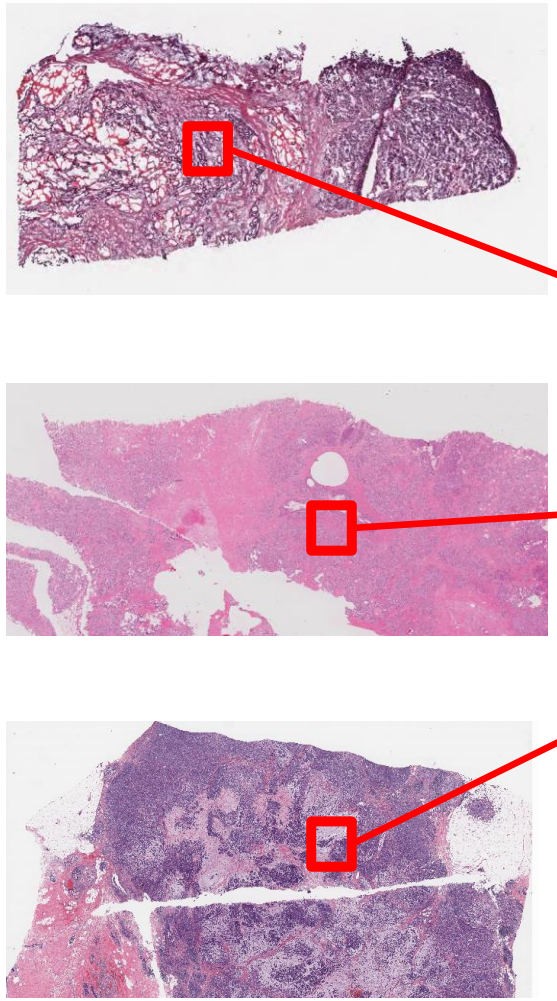
4.69x

Results and Final Profile

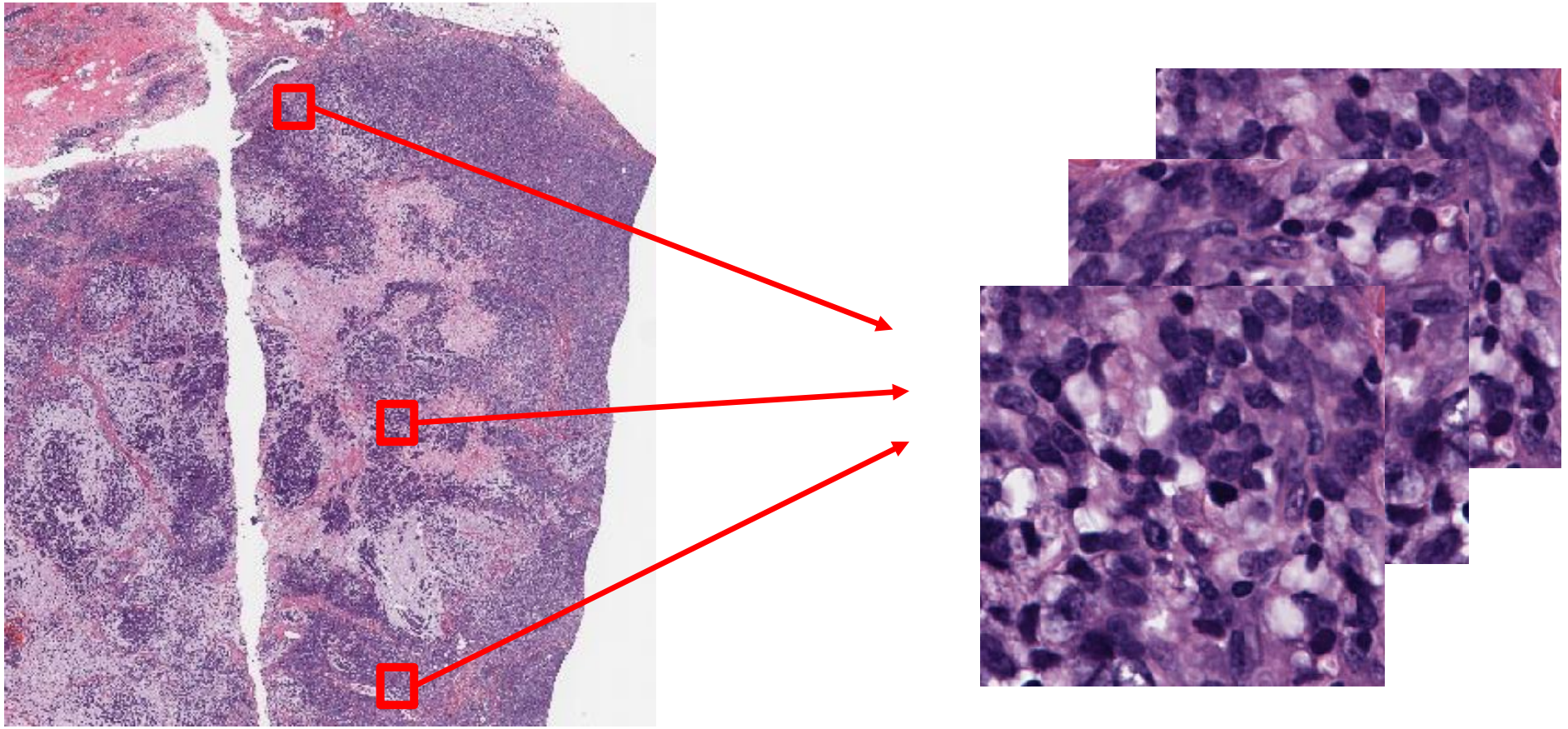
Open File		Coord Fetch			Speedup
Pyvips	CuCIM	pkl	Feather	CuPy	
V		V			-
V			V		1.0613x
V		V		V	1.0449x
V			V	V	1.1811x
	V	V			4.691x
	V		V	V	5.534x

Theoretically

Problems



Problems



Special Thanks

Mentors :

Ken Liao (Nvidia)

Yang-Hsien Lin (Nvidia)

Members :

Yen-Jung Chiu (MCU)

Po-Hao Hsu (NCHC)

Chao-Chun Chuang (NCHC)

OpenACC More Science, Less Programming

OPEN HACKATHONS

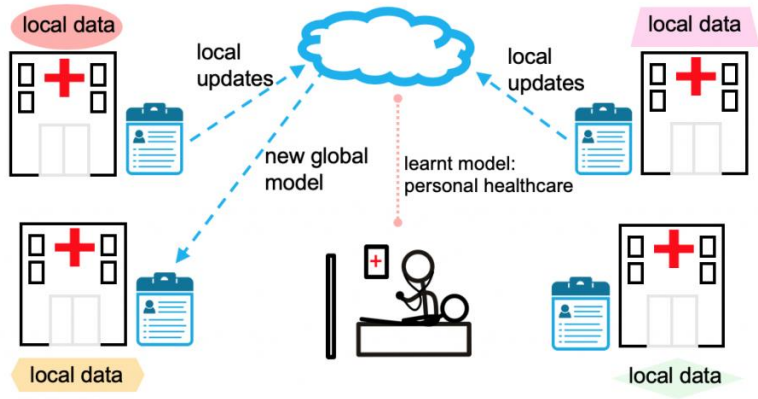
NCHC

NARLabs 財團法人國家實驗研究院 國家高速網路與計算中心 National Center for High-performance Computing

NVIDIA

跨院數據協作的最佳解方——聯邦學習

醫療影像在AI領域上的蓬勃發展，預示著我們即將進入醫療影像智能化的時代。在這個過程中，聯邦學習技術成為一個關鍵工具，它在模型開發中扮演著重要的角色，特別是在處理隱私保護和提升模型準確度方面。由於病人隱私的問題，傳統的影像共享方式受到限制，使得聯邦學習技術在構建多樣化數據集時變得不可或缺。這一技術在當前數據隱私需求較高的時代尤為重要。然而，在分布式環境中執行聯邦式學習對於需要大規模數據整合的問題而言，會遇到訓練速度較慢的挑戰。為了解決這一問題，團隊運用了 GPU 加速來提升模型訓練和數據處理的速度，並結合 Feather 和 CuPy 來優化數據處理過程，提高了數據提取和計算的效率。此外，使用 CuCIM 來驗證和優化影像在GPU上的高效提取，特別是利用 CuCIM 進行 cache 優化。通過這些技術的整合，詹寶珠教授帶領的 Smile Lab 團隊成功實現了聯邦學習四倍的加速效果，顯著提升了整體模型訓練的效率和準確度。



Open File		Coord Fetch			Speedup
Pyvips	CuCIM	pkl	Feather	CuPy	
V		V			-
V			V		1.0613x
V		V		V	1.0449x
V			V	V	1.1811x
	V	V			4.691x
	V		V	V	5.534x

Theoretically