NARLabs 財團法人國家實驗研究院

國家高速網路與計算中心

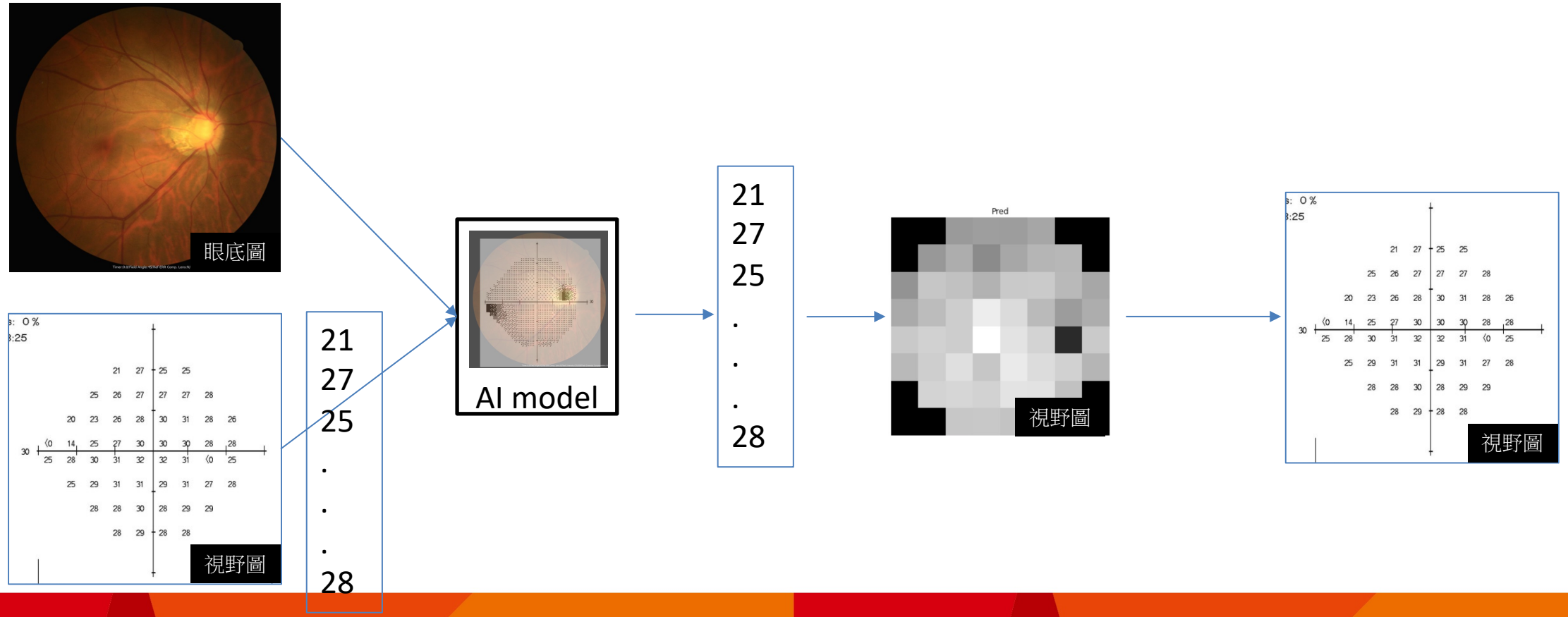
National Center for High-performance Computing

A+DAS組內專案
雲端影像辨識服務API

羅世瑋

Vision filed regression model workflow

- 用途：視野缺損預測模型試圖追溯視網膜神經纖維的弓形會聚，從視盤周圍的視網膜等特徵，遞歸通過所提出的網絡順序產生視覺缺損的估計。

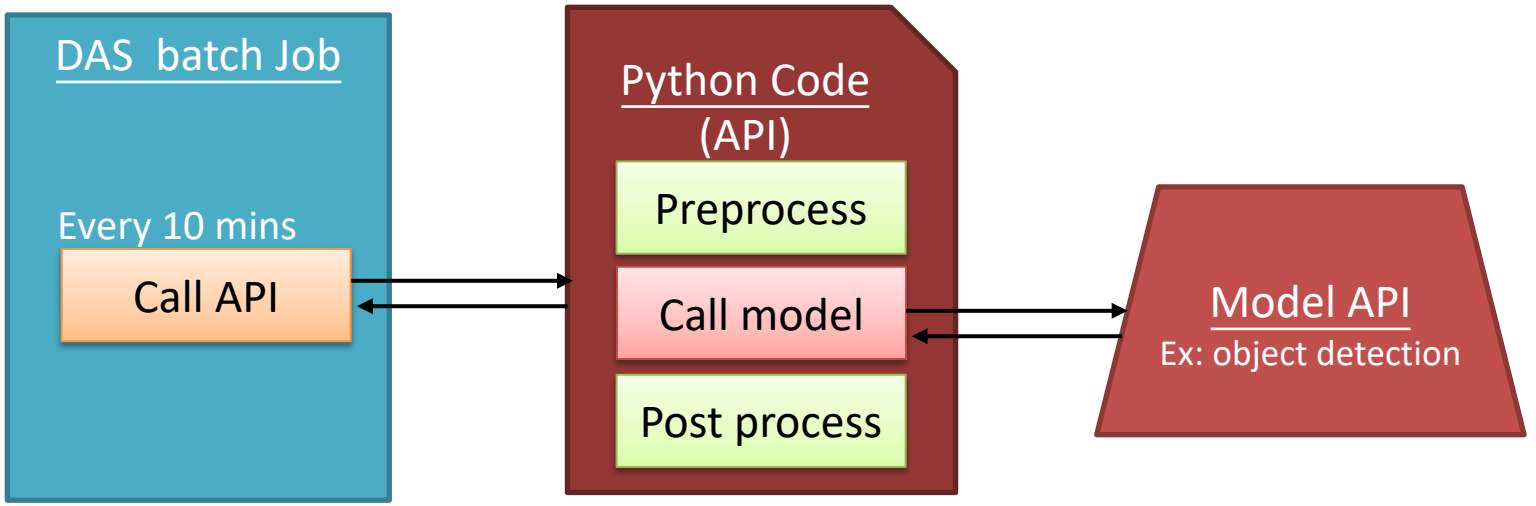


方案一

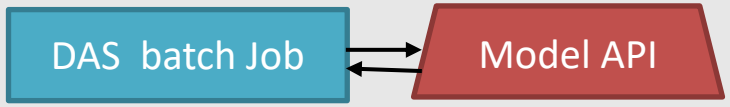
Deployment space

Analysis project job

1. Deploy Model
2. Deploy Python code(call model)
3. Run batch job



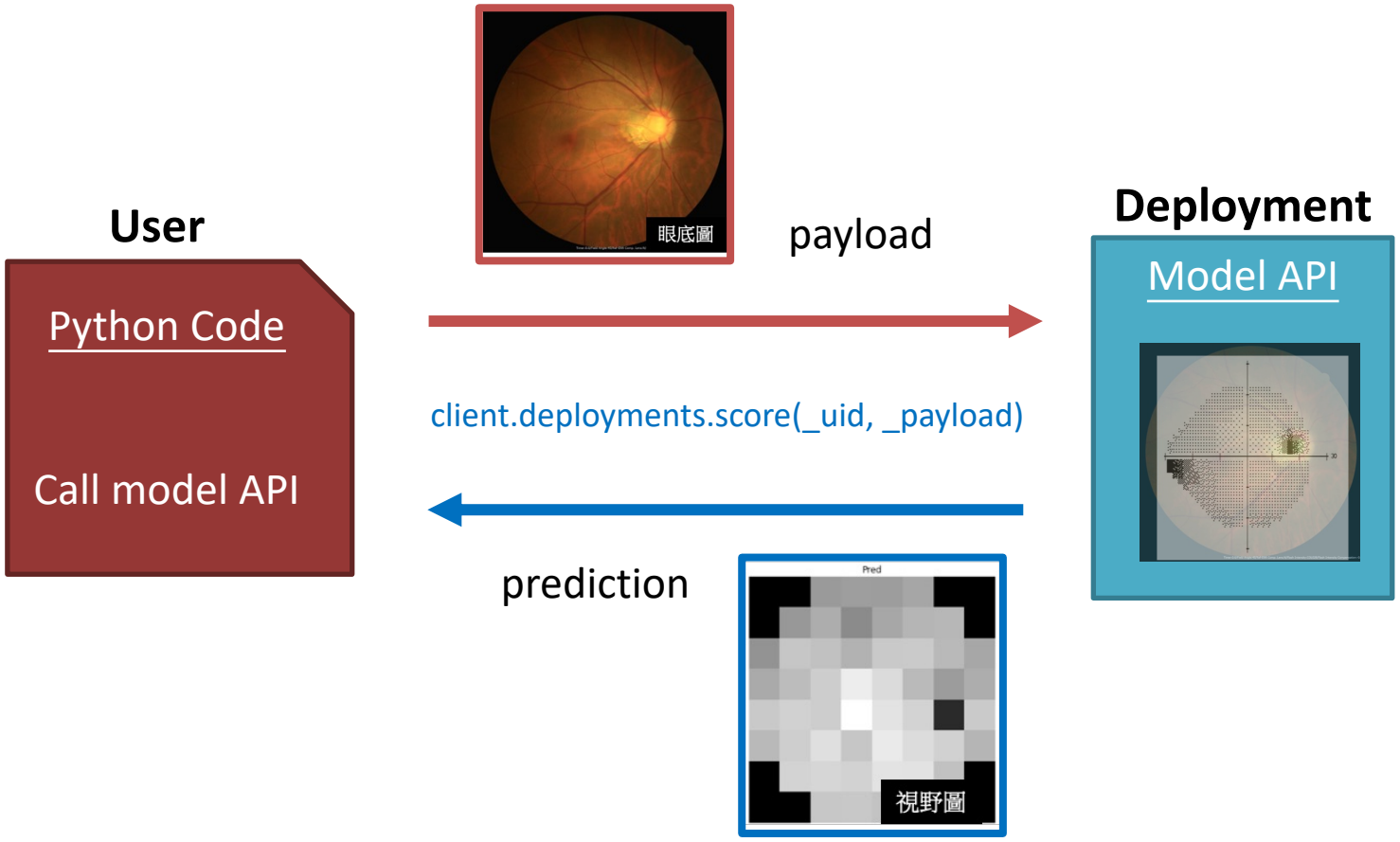
方案比較

方案		loading	run	架構	可彈性拆模組
方案一		API較輕量	Batch	稍複雜	✓
方案二		API較輕量	By user	簡單	✓
方案三		Code包job中，loading較重	Batch	簡單	×

Batch job可以選擇GPU硬體　方案一、二均完成測試

- Deploy python code適合service, 小量即時
- Batch job適合inference大量資料

DAS Model API



Payload

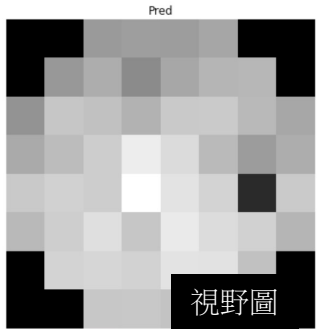
```
# online
# scoring_payload = {"input_data": [{"values": [array]}], "back_data": [{"back_up_int": "This a backup text."}]}

# batch
scoring_payload = {"input_data": [{"values": [array]}], "back_data": [{"back_up_int": "This a backup text."}],
                   "hardware_spec":{"name":"V100", 'num_nodes': 1}}
```

Payload sample

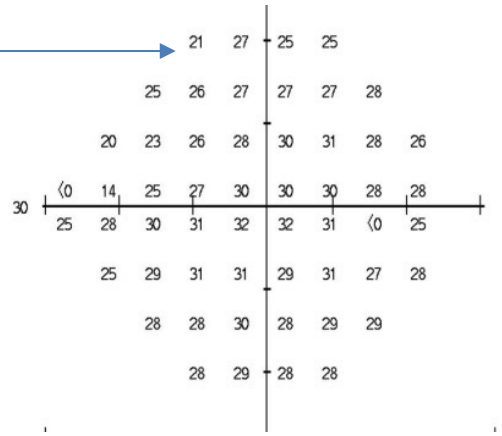
API return

```
print(client.deployments.get_job_details(job_uid)['entity']['scoring']['predictions'])
```


```
[{'id': 'dense_1', 'values': [[17.674304962158203, 17.95292854309082, 17.6146297454834, 16.127347946166992, 19.590438842773438, 21.036605834960938, 20.49911117553711, 20.1116733551
0254, 19.100324630737305, 18.739704132080078, 20.200075149536133, 22.64345932006836, 21.690959930419922, 22.687536239624023, 19.972410202026367, 21.26595687866211, 21.4983348846435
55, 20.0864963531494 14, 20.48937225341797, 19.652311325073242, 23.66286849975586, 24.622997283935547, 22.984596252441406, 22.617801666259766, 18.977439880371094, 21.6691837310791,
21.257341384887695, 12.995844841003418, 24.569049835205078, 27.862613677978516, 29.149593353271484, 24.51448631286621, 11.70016860961914, 22.26041603088379, 21.639036178588867, 23.
94772720336914, 25.25027847290039, 24.84150505065918, 24.181076049804688, 25.50109100341797, 24.461063385009766, 21.980815887451172, 23.014314651489258, 25.902250289916992, 23.0947
322845459, 25.474878311157227, 25.31443214416504, 23.556474685668945, 21.927064895629883, 23.288898468017578, 23.846590042114258, 21.772714614868164]]}]
```

21
27
25
.
.
.
28



(1) 實驗結論

- 現有影像辨識服務與TensorFlow模型可以利用DAS內的API模式進行佈建並提供線上服務，我們也破解了如何調用GPU環境的deployment方式，使得原本的API服務能夠獲得計算效率的加速。然而，由於該辨識服務比較單純僅回傳辨識數值組，在串接資料庫、dashboard上則沒有著墨空間，對於要展示亮點eye candy等用途還需要評估
- 。