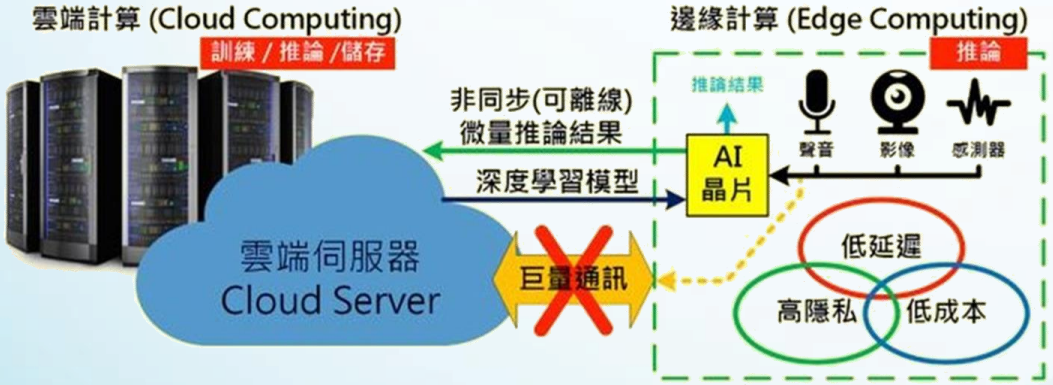
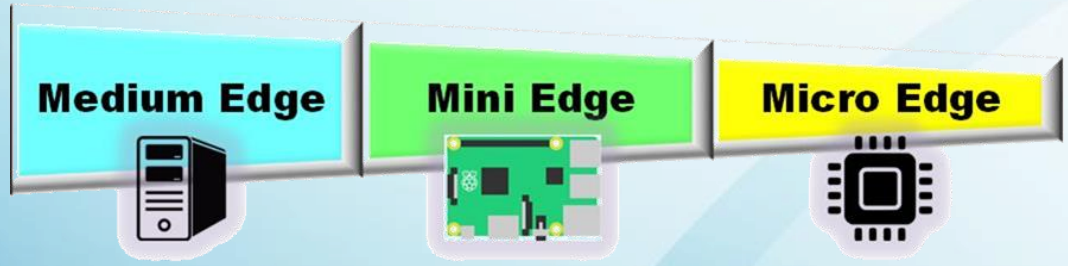
OmniXRI's Edge AI & TinyML 小學堂





【第11講】
案例實作 — 姿態估測

歐尼克斯實境互動工作室 (OmniXRI Studio)
許哲豪 (Jack Hsu)

簡報大綱

- 11.1. 姿態估測簡介
- 11.2. 姿態估測模型
- 11.3. 姿態估測評量
- 11.4. 姿態估測實作

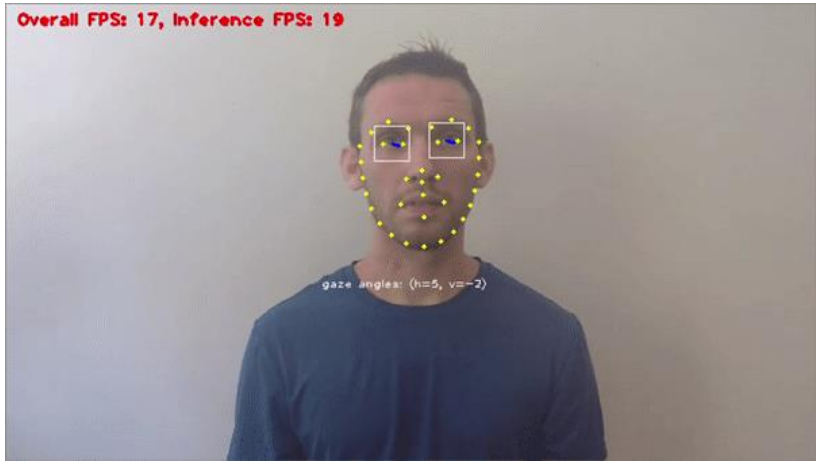
本課程完全免費，請勿移作商業用途！
歡迎留言、訂閱、點讚、轉發，讓更多需要的朋友也能一起學習。

完整課程大綱： https://omnixri.blogspot.com/2024/02/omnixris-edge-ai-tinyml-0.html
課程直播清單： https://www.youtube.com/@omnixri1784/streams

姿態估測(Pose Estimation)

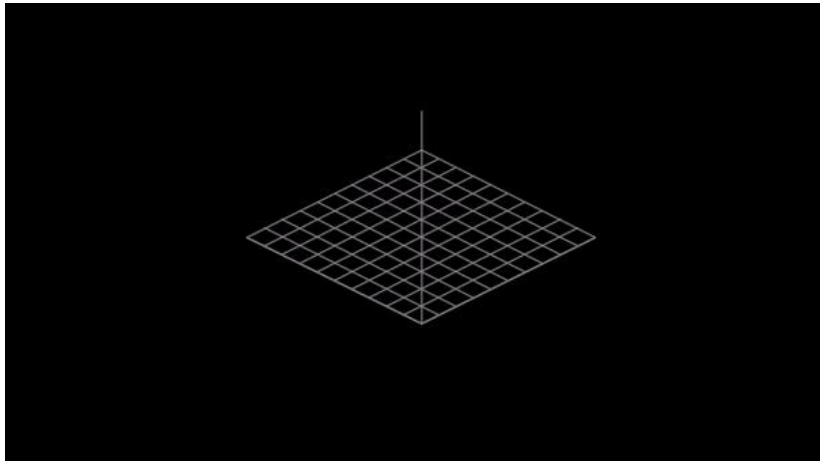
主要部位

- 頭部
- 注視點
- 手部
- 全身



估測方式

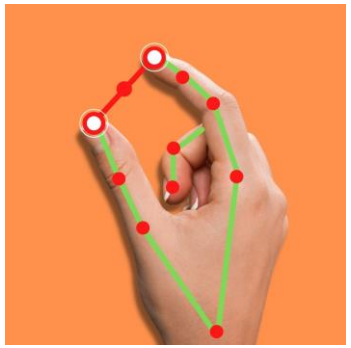
- 2D（照片、攝影機）
- 3D（深度攝影機）
- 2.5D（假3D，輸入2D，輸出3D，以平面估測立體深度）

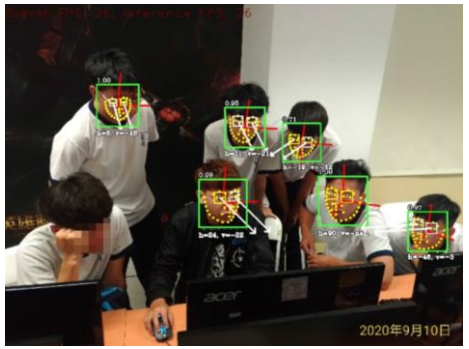


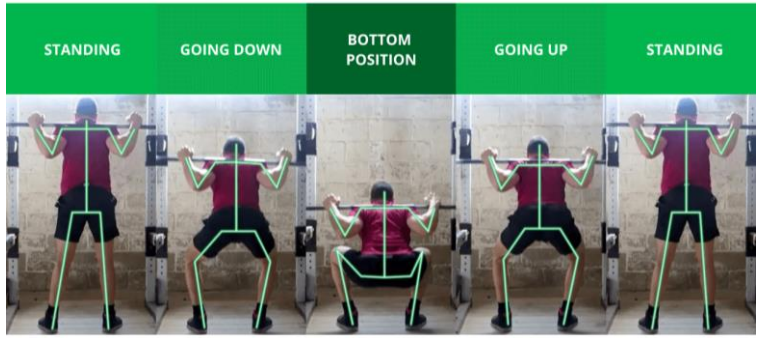
姿態估測常見用途

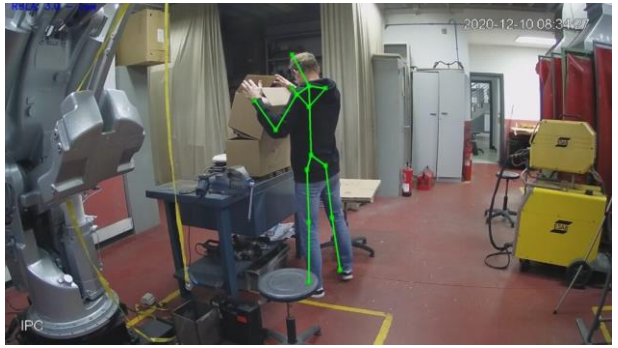
常見用途

- 人機互動
- 運動分析
- 組裝確認
- 智慧零售
- 智慧安防
- 智慧教室
- 安全駕駛
- …







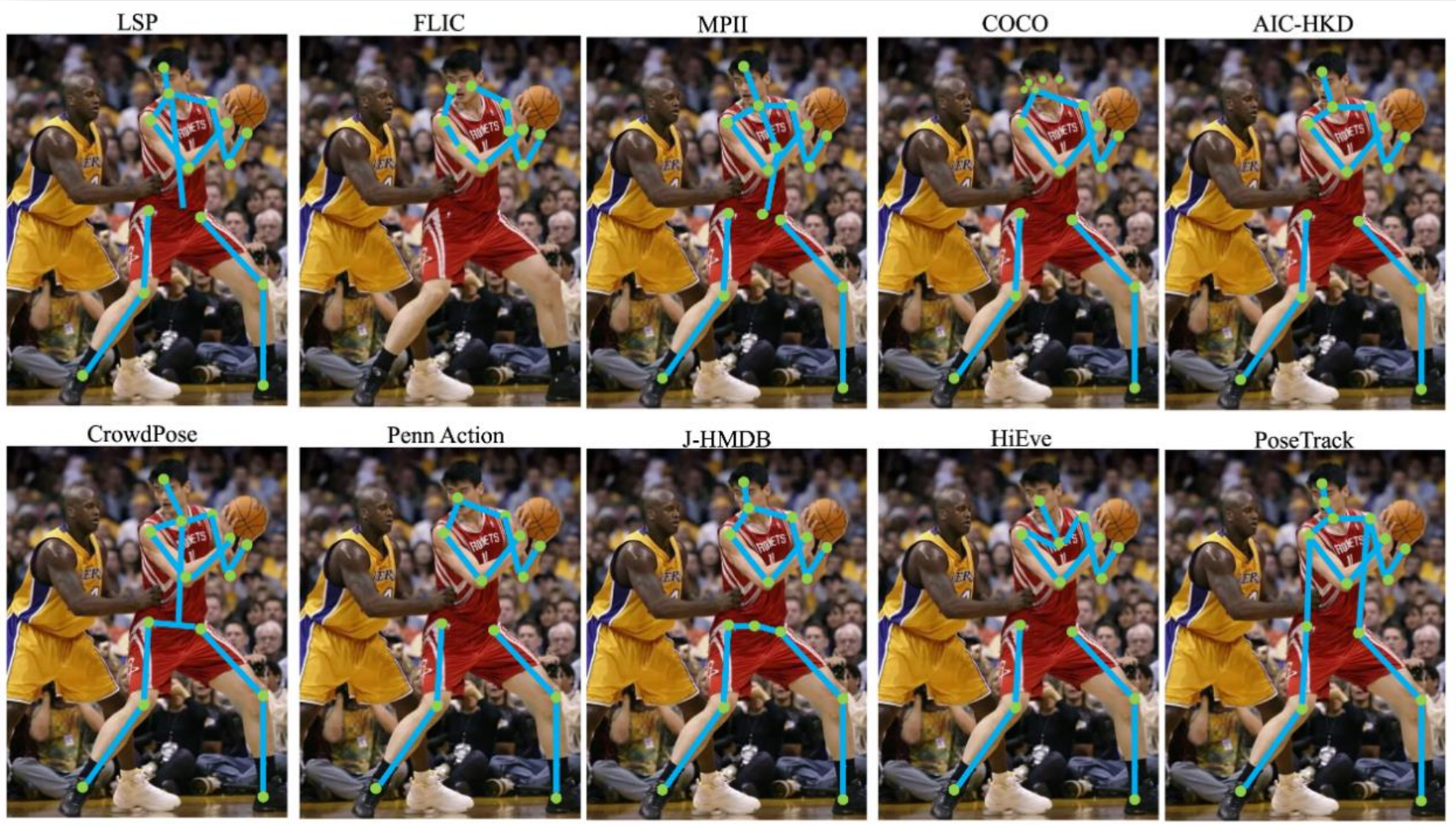


姿態估測（人體關鍵點）資料集

常見的2D關鍵點開放資料集

- **微軟COCO (17點)**
- **CMU OpenPose (18/25點)**
  - 人體骨架 (18/25點)
  - 臉部動作 (70點)
  - 手部動作 (22點)
- MPII (16點)
- AI Challenge (14點)
- LSP (14點)
- FLIC (9點)

姿態估測標註比較



資料來源：https://arxiv.org/abs/2204.07370

MS COCO 關鍵點資料集

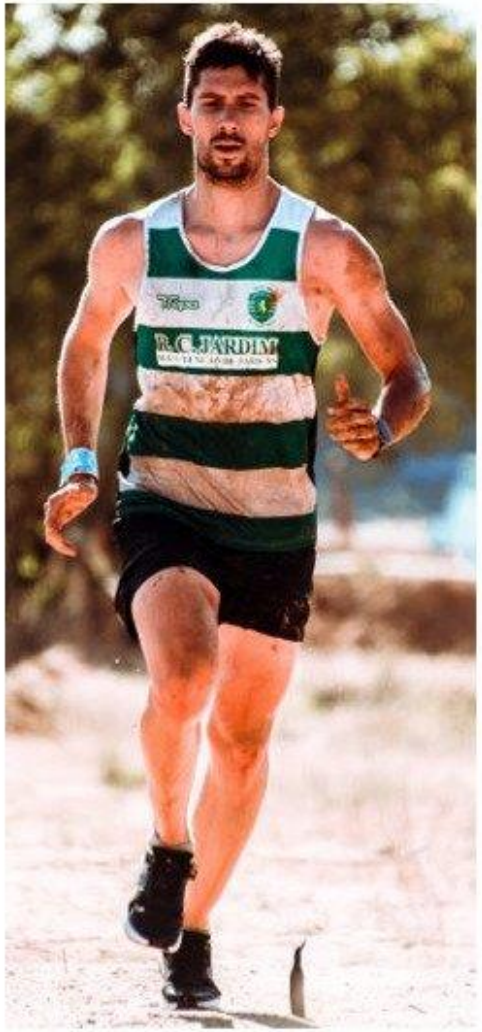
Person_keypoints_val2017 範例 (JSON)

```
annotation{
    "keypoints": [x1,y1,v1,...],
    "num_keypoints": int,
    "id": int,
    "image_id": int,
    "category_id": int,
    "segmentation": RLE or [polygon],
    "area": float,
    "bbox": [x,y,width,height],
    "iscrowd": 0 or 1,
}
```

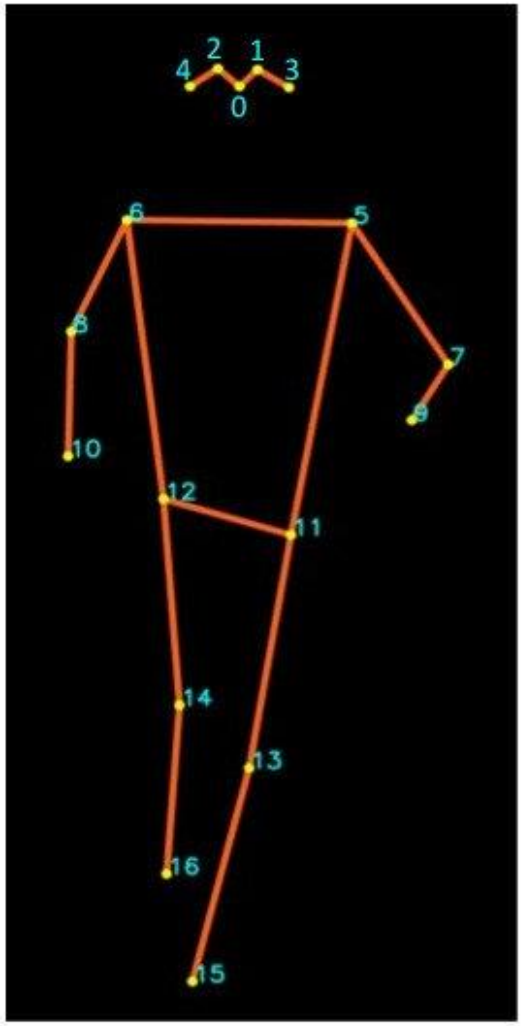
v=0 未標註, xy須為0
v=1 已標註但不可見
v=2 標註且可見

資料來源：https://cocodataset.org/#keypoints-2020, https://cocodataset.org/#keypoints-eval

MS COCO 關鍵點定義



Index	Key point
0	Nose
1	Left-eye
2	Right-eye
3	Left-ear
4	Right-ear
5	Left-shoulder
6	Right-shoulder
7	Left-elbow
8	Right-elbow
9	Left-wrist
10	Right-wrist
11	Left-hip
12	Right-hip
13	Left-knee
14	Right-knee
15	Left-ankle
16	Right-ankle

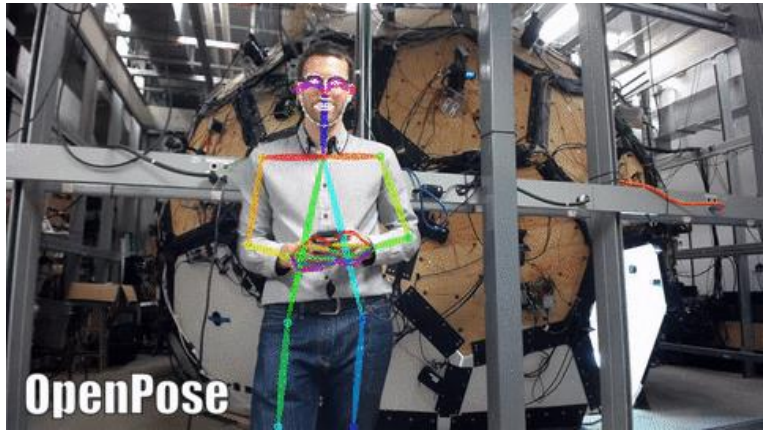


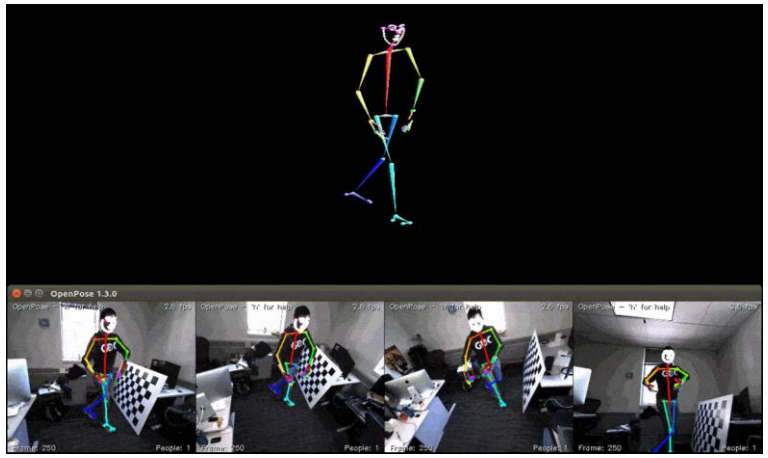
Categories字串

```
{
    "id": int,
    "name": str,
    "supercategory": str,
    "keypoints": [str],
    "skeleton": [edge]
}
```

關鍵點名稱 (17)
骨架連結

OpenPose 基本介紹





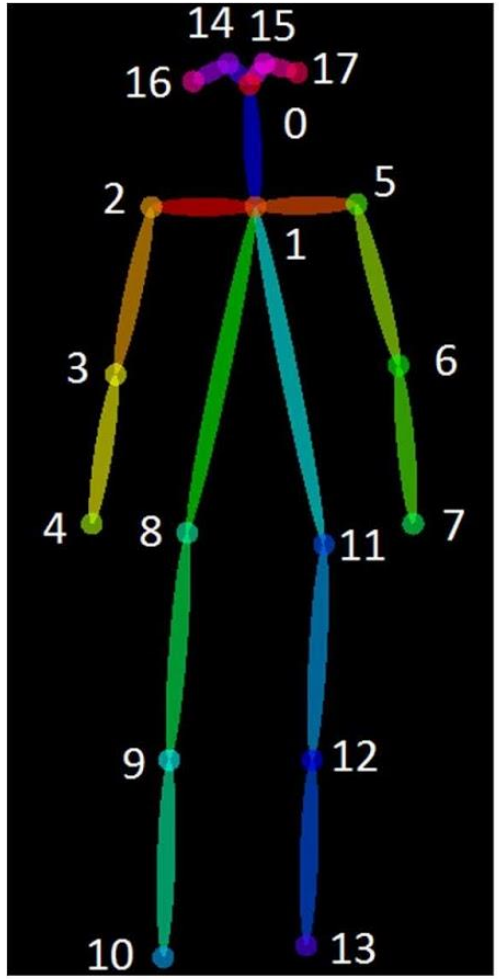
卡內基美濃大學(CMU) 2017年提出。

開源人體姿態專案，包含即時人體動作(2D/3D)、面部表情及手指運動偵測等。

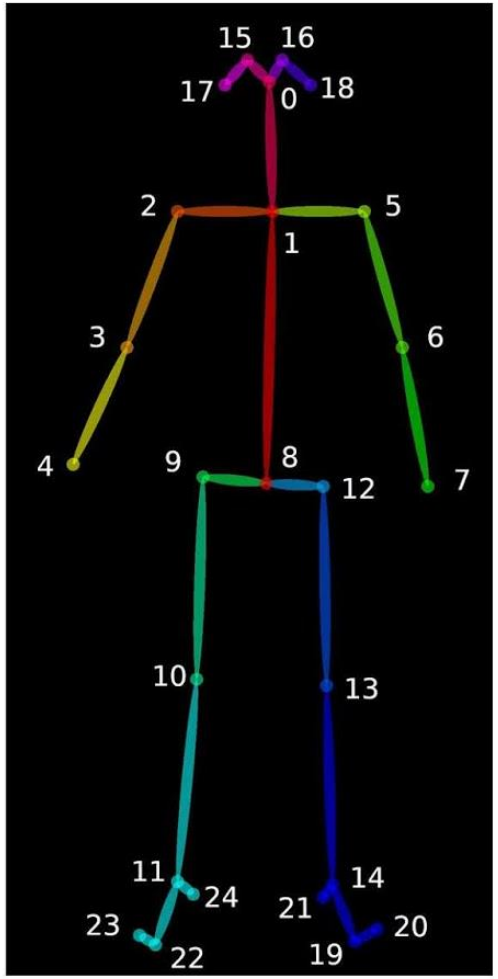
支援Windows, Linux(Ubuntu), Mac OS。支援CPU及GPU(CUDA, OpenCL)加速。

https://github.com/CMU-Perceptual-Computing-Lab/openpose

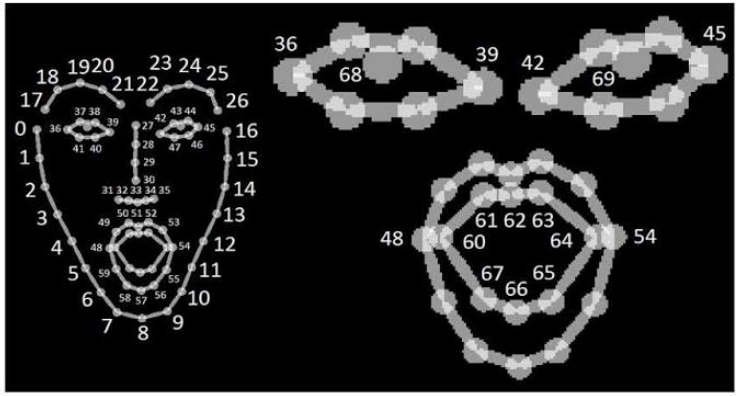
OpenPose 關鍵點定義



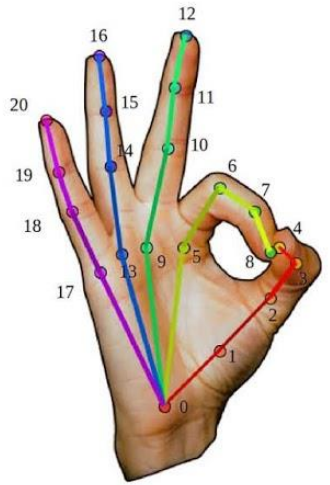
(a)



(b)



(b)



(d)

OpenPose 支援輸出格式：

(a) 類COCO 18點

(b) BODY 25點

(c) 臉部70點

(d)手部22點

姿態關鍵點轉換

MS COCO: (17點)

{0:鼻子, 1:左眼, 2:右眼, 3:左耳, 4:右耳, 5:左肩, 6:右肩, 7:左肘, 8:右肘, 9:左腕, 10:右腕, 11:左臀（腰）, 12:右臀（腰）, 13:左膝, 14:右膝, 15:左踝, 16:右踝 }

CMU OpenPose: (18點)

{0:鼻子, 1:脖子, 2:右肩, 3:右肘, 4:右腕, 5:左肩, 6:左肘, 7:左腕, 8:右臀（腰）, 9:右膝, 10:右踝, 11:左臀（腰）, 12:左膝, 13:左踝, 14:左眼, 15:右眼, 16:左耳, 17:右耳 }

人體姿態估測模型(OpenVINO OMZ)

Legacy Features\ Open Model Zoo\

Intel's Pre-Trained Model

Model Name	Complexity (GFLOPs)	Size (Mp)
human-pose-estimation-0001	15.435	4.099
human-pose-estimation-0005	5.9393	8.1504
human-pose-estimation-0006	8.8720	8.1504
human-pose-estimation-0007	14.3707	8.1504

OpenPose

EfficientHRNet

https://docs.openvino.ai/2024/omz_models_group_intel.html#human-pose-estimation-models

Public Pre-Trained Model

Model Name	Implementation	OMZ Model Name	Accuracy	GFlops	mParams
human-pose-estimation-3d-0001	PyTorch*	human-pose-estimation-3d-0001	100.44437mm	18.998	5.074
single-human-pose-estimation-0001	PyTorch*	single-human-pose-estimation-0001	69.0491%	60.125	33.165
higher-hrnet-w32-human-pose-estimation	PyTorch*	higher-hrnet-w32-human-pose-estimation	64.64%	92.8364	28.6180

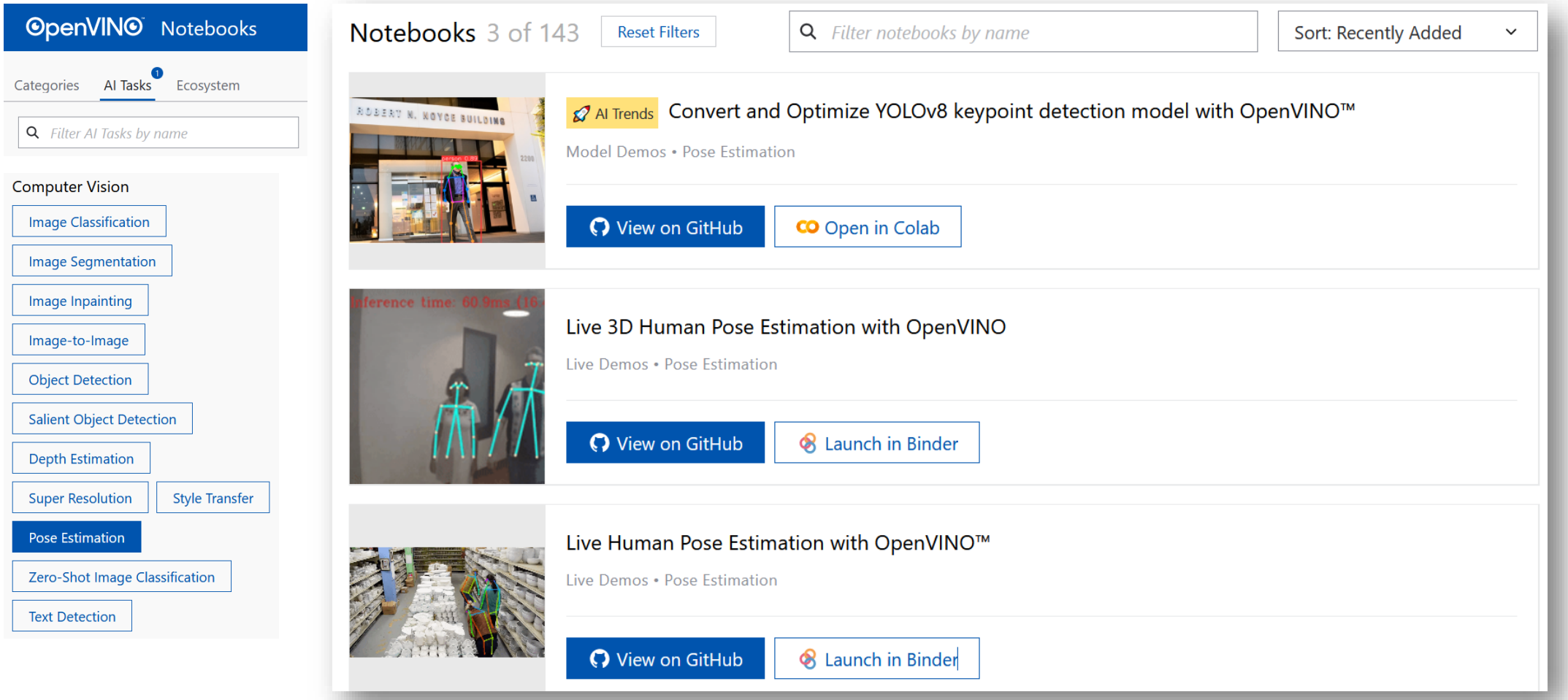
Lightweight OpenPose

Convolutional Pose Machines

HigherHRNet

https://docs.openvino.ai/2024/omz_models_group_public.html#human-pose-estimation-models

人體姿態估測模型(OpenVINO Notebooks)



https://openvinotoolkit.github.io/openvino_notebooks/

YOLOv8 姿態估測模型及資料格式

Model	size (pixels)	mAP^{pose} 50-95	mAP^{pose} 50	Speed CPU ONNX (ms)	Speed A100 TensorRT (ms)	params (M)	FLOPs (B)
YOLOv8n-pose	640	50.4	80.1	131.8	1.18	3.3	9.2
YOLOv8s-pose	640	60.0	86.2	233.2	1.42	11.6	30.2
YOLOv8m-pose	640	65.0	88.8	456.3	2.00	26.4	81.0
YOLOv8l-pose	640	67.6	90.0	784.5	2.59	44.4	168.6
YOLOv8x-pose	640	69.2	90.2	1607.1	3.73	69.4	263.2
YOLOv8x-pose-p6	1280	71.6	91.2	4088.7	10.04	99.1	1066.4

Format with Dim = 2

```
<class-index> <x> <y> <width> <height> <px1> <py1> <px2> <py2> ... <pxn> <pyn>
```

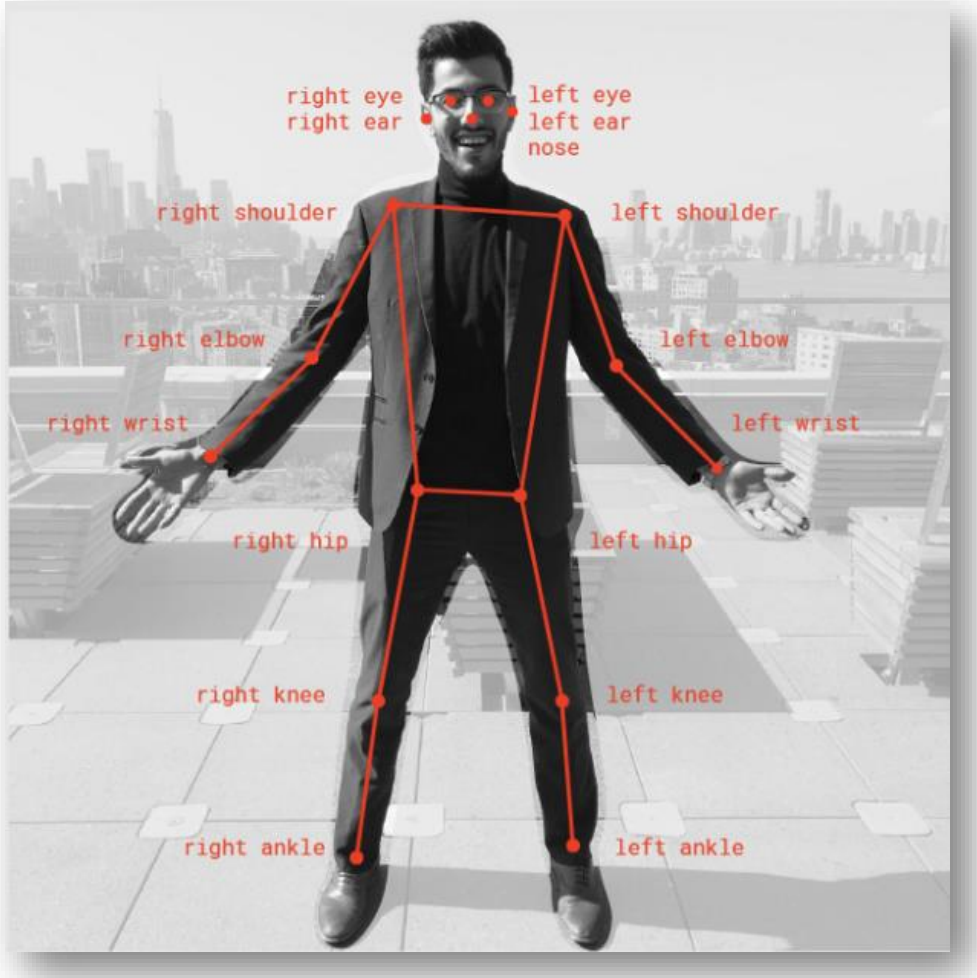
Format with Dim = 3

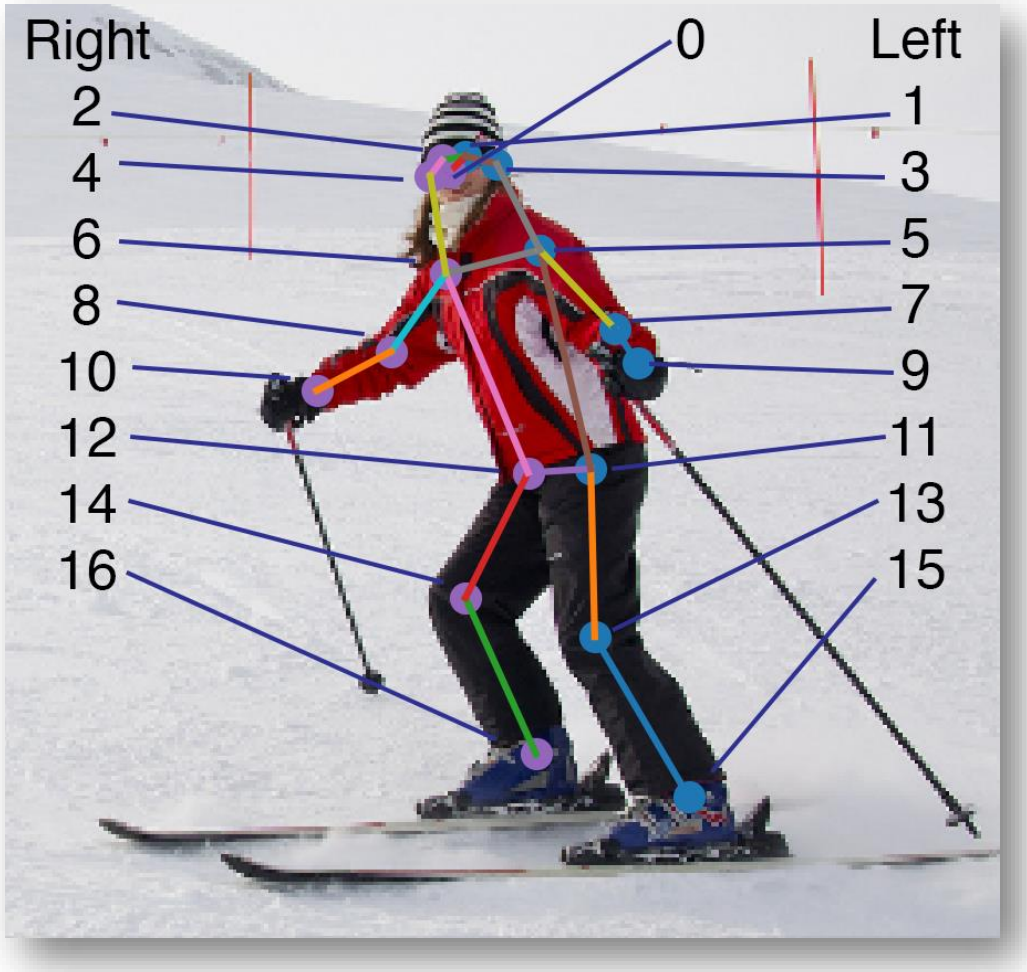
```
<class-index> <x> <y> <width> <height> <px1> <py1> <p1-visibility> <px2> <py2> <p2-visibility> <pxn> <p
```

和 MS COCO Keypoints 定義類似

資料來源：https://docs.ultralytics.com/tasks/pose/ https://docs.ultralytics.com/datasets/pose/

Yolov8 關鍵點定義

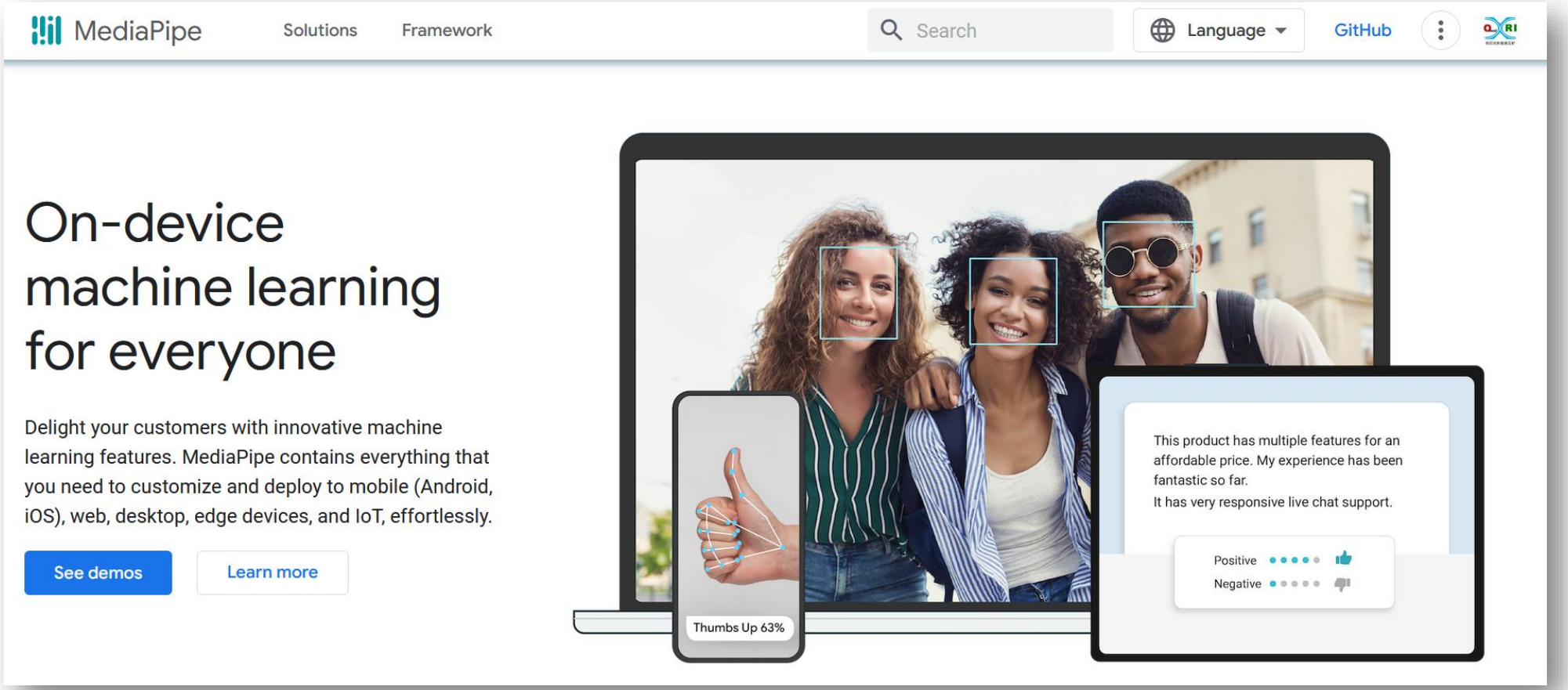




和 MS COCO Keypoints 定義類似

資料來源：https://chtseng.wordpress.com/2023/07/14/yolov8-pose-estimation/

Google MediPipe



2019/6 首次於CVPR 展示預覽版本

2023/4/3 起由 https://google.github.io/mediapipe/ 遷移至 https://developers.google.com/mediapipe

MediaPipe 主要特色

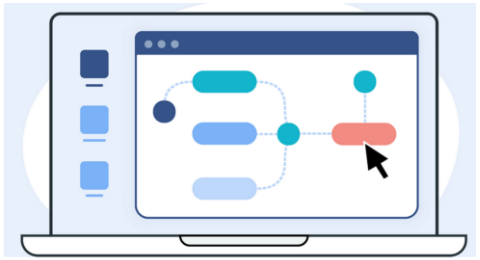
MediaPipe

免費及開源
支援CPU, GPU加速計算

多種解決方案
- 視覺
- 文字
- 聲音
- 生成式

多種開發框架
- Android
- iOS
- Python
- Web

Low-Code APIs



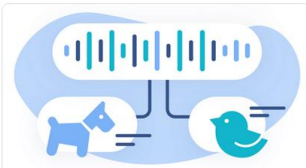
No-Code Studio

Vision

Analyze things in images and videos.

Natural language

Understand meanings behind text.

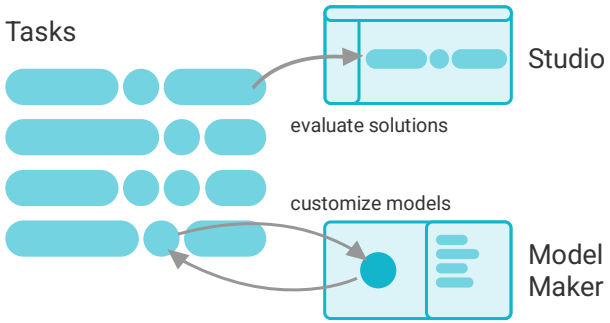


Audio

Listen and recognize sounds.

Others

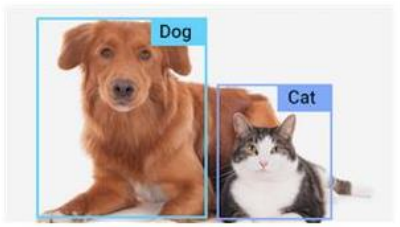
More solutions coming soon.



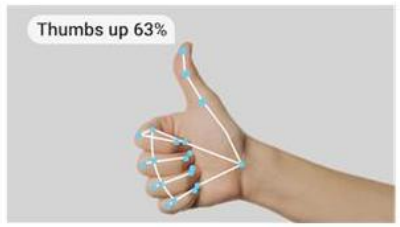
https://github.com/google/mediapipe

MediaPipe 解決方案

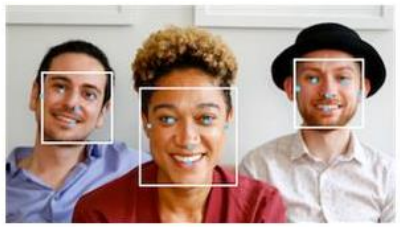
影像辨識



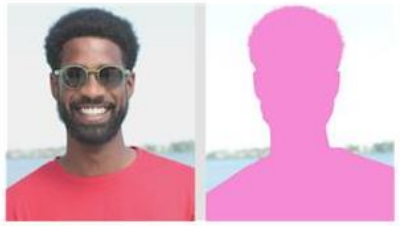
物件偵測



手勢辨識

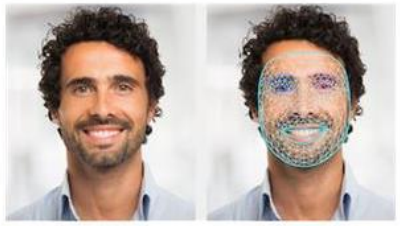

人臉偵測


影像相似度


影像分割


交互式分割


手部特徵點


臉部特徵點


姿態特徵點

文字情緒

文字相似度

語系辨識

語音辨識

MediaPipe

DEMO: https://mediapipe-studio.webapps.google.com/home

姿態估測指標

Object Keypoint Similarity (OKS) 關鍵點相似度 (COCO)

$$OKS_p = \frac{\sum_i exp\{-d_{pi}^2/2S_p^2\sigma_i^2\}\delta(v_{pi} > 0)}{\sum_i \delta(v_{pi} > 0)}$$

Average Precision (AP) 平均精準度

（單人、多人）

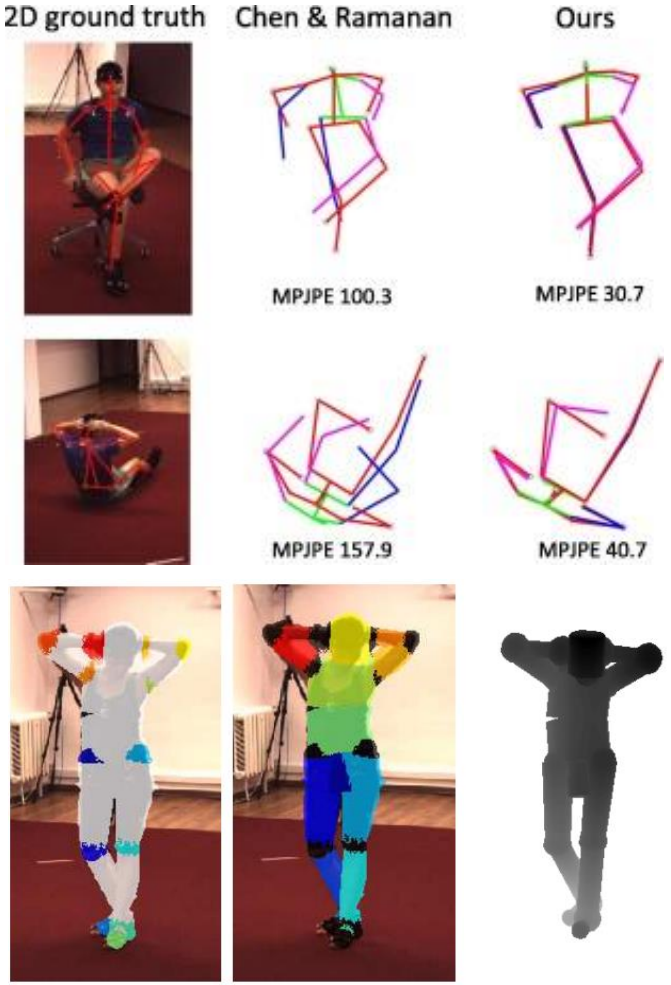
$$AP = \frac{\sum_p \delta(oks_p > T)}{\sum_p 1}$$

Mean Per Joint Position Error (MPJPE)

每個關節位置誤差。

$$E_{MPJPE}(f, \mathcal{S}) = \frac{1}{N_{\mathcal{S}}} \sum_{i=1}^{N_{\mathcal{S}}} \|m_{\mathbf{f},\mathcal{S}}^{(f)}(i) - m_{\mathbf{gt},\mathcal{S}}^{(f)}(i))\|_2$$

資料來源：http://vision.imar.ro/human3.6m/pami-h36m.pdf



實作範例1 – Yolov8 姿態估測 (OpenVINO Colab)

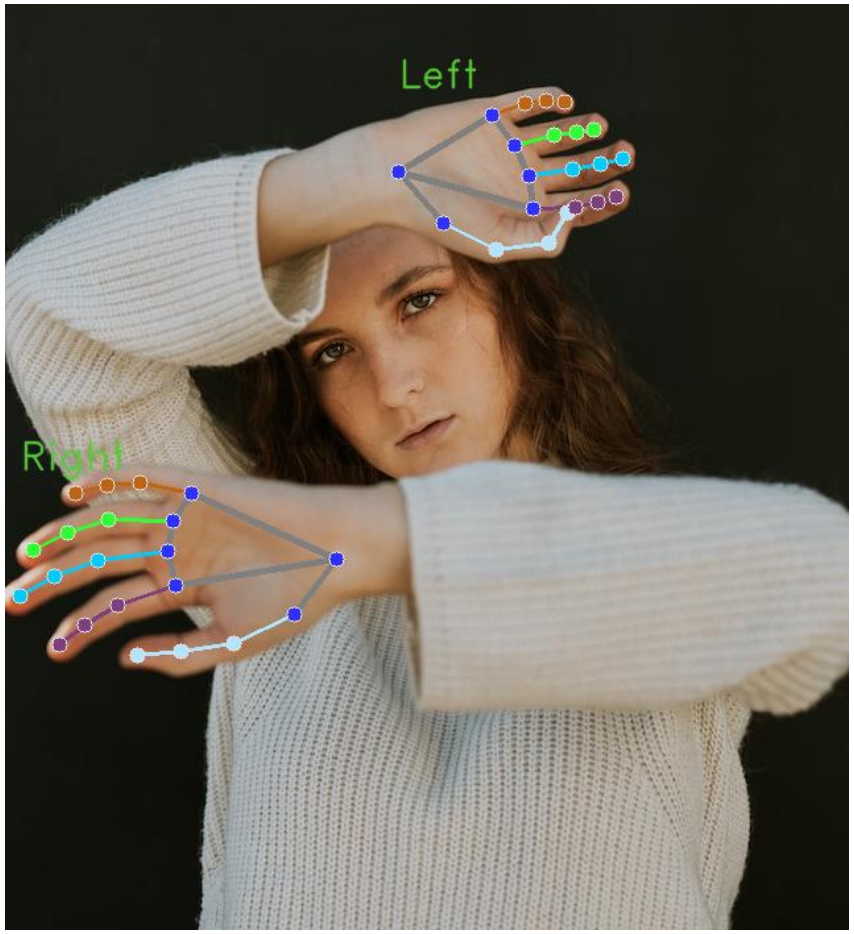
Convert and Optimize YOLOv8 keypoint detection model with OpenVINO

程式導讀：

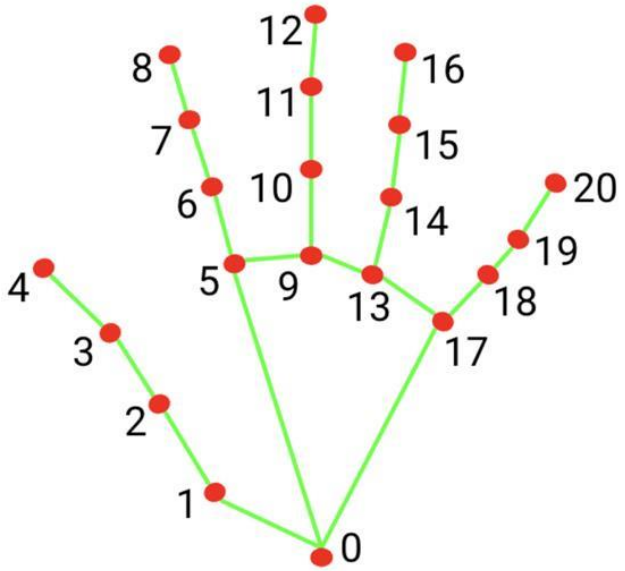
- Prepare the PyTorch model.
- Download and prepare a dataset.
- Validate the original model.
- Convert the PyTorch model to OpenVINO IR.
- Validate the converted model.
- Prepare and run optimization pipeline.
- Compare performance of the FP32 and quantized models.
- Compare accuracy of the FP32 and quantized models.
- Live demo

https://colab.research.google.com/github/openvinotoolkit/openvino_notebooks/blob/latest/notebooks/yolov8-optimization/yolov8-keypoint-detection.ipynb

實作範例2 ─手部特徵點偵測 (MediaPipe Colab)



Model name	Input shape	Quantization type
HandLandmarker (full)	192 x 192, 224 x 224	float 16

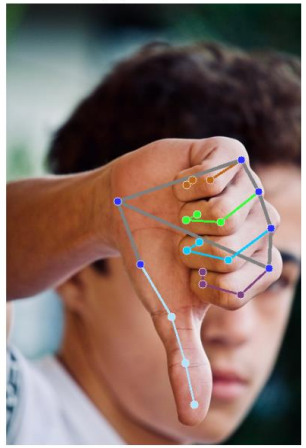


0. WRIST
1. THUMB_CMC
2. THUMB_MCP
3. THUMB_IP
4. THUMB_TIP
5. INDEX_FINGER_MCP
6. INDEX_FINGER_PIP
7. INDEX_FINGER_DIP
8. INDEX_FINGER_TIP
9. MIDDLE_FINGER_MCP
10. MIDDLE_FINGER_PIP
11. MIDDLE_FINGER_DIP
12. MIDDLE_FINGER_TIP
13. RING_FINGER_MCP
14. RING_FINGER_PIP
15. RING_FINGER_DIP
16. RING_FINGER_TIP
17. PINKY_MCP
18. PINKY_PIP
19. PINKY_DIP
20. PINKY_TIP

https://colab.research.google.com/github/googlesamples/mediapipe/blob/main/examples/hand_landmarker/python/hand_landmarker.ipynb

實作範例3－手勢辨識 (MediaPipe Colab)

Thumb_Down (0.77)

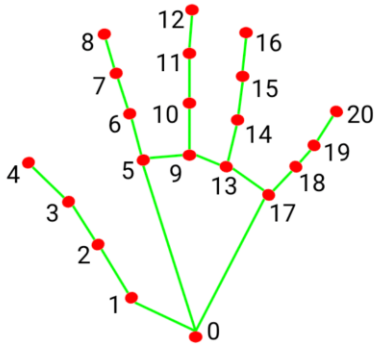


Victory (0.91)

Thumb_Up (0.73)

Pointing_Up (0.82)

Model name	Input shape	Quantization type
HandGestureClassifier	192 x 192, 224 x 224	float 16

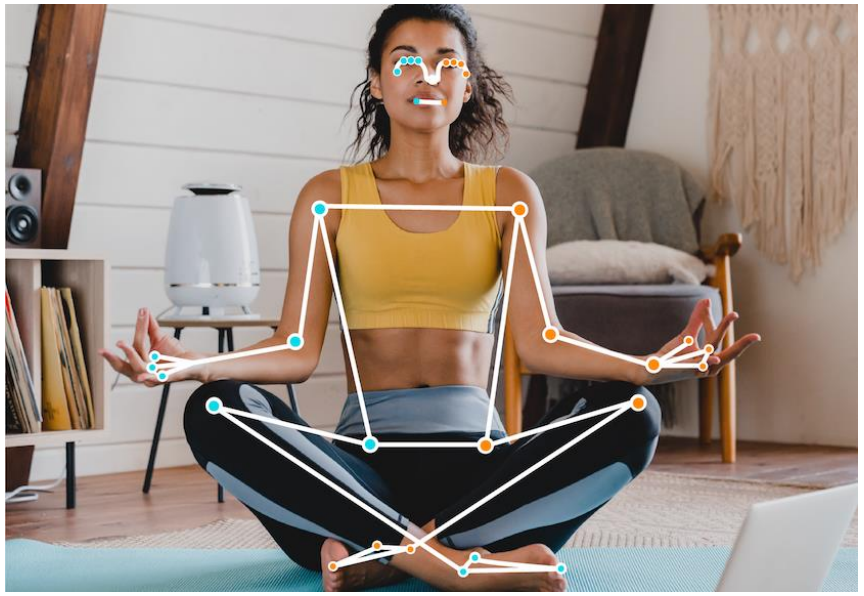


0. WRIST
1. THUMB_CMC
2. THUMB_MCP
3. THUMB_IP
4. THUMB_TIP
5. INDEX_FINGER_MCP
6. INDEX_FINGER_PIP
7. INDEX_FINGER_DIP
8. INDEX_FINGER_TIP
9. MIDDLE_FINGER_MCP
10. MIDDLE_FINGER_PIP
11. MIDDLE_FINGER_DIP
12. MIDDLE_FINGER_TIP
13. RING_FINGER_MCP
14. RING_FINGER_PIP
15. RING_FINGER_DIP
16. RING_FINGER_TIP
17. PINKY_MCP
18. PINKY_PIP
19. PINKY_DIP
20. PINKY_TIP

```
0 - Unrecognized gesture, label: Unknown
1 - Closed fist, label: Closed_Fist
2 - Open palm, label: Open_Palm
3 - Pointing up, label: Pointing_Up
4 - Thumbs down, label: Thumb_Down
5 - Thumbs up, label: Thumb_Up
6 - Victory, label: Victory
7 - Love, label: ILoveYou
```

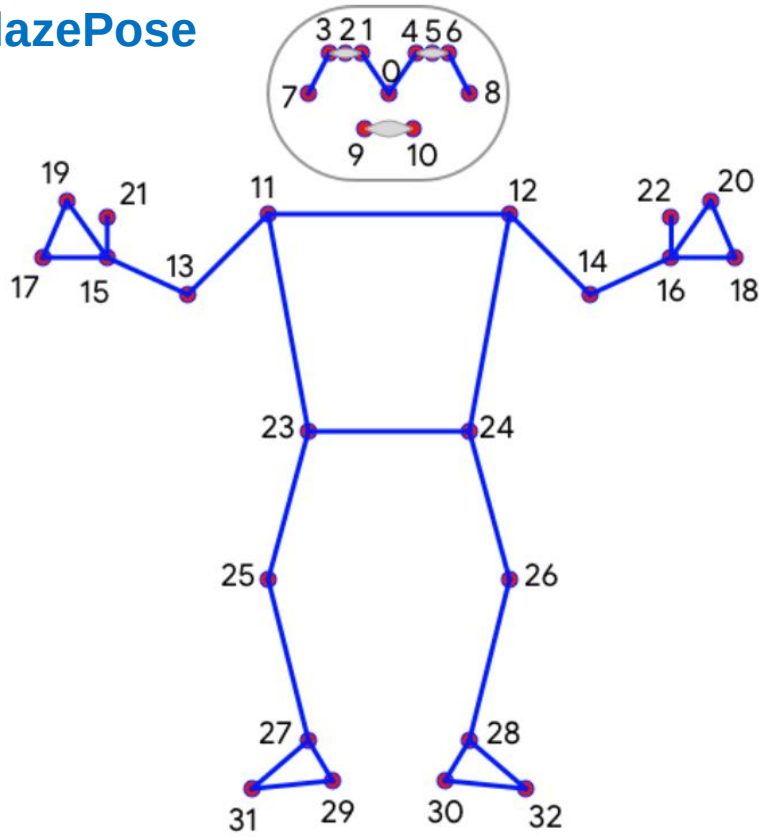
https://colab.research.google.com/github/googlesamples/mediapipe/blob/main/examples/gesture_recognizer/python/gesture_recognizer.ipynb

實作範例4 ─姿態特徵點偵測 (MediaPipe Colab)



Model bundle	Input shape	Data type	Model Cards	Versions
Pose landmarker (lite)	Pose detector: 224 x 224 x 3 Pose landmarker: 256 x 256 x 3	float 16	info	Latest
Pose landmarker (Full)	Pose detector: 224 x 224 x 3 Pose landmarker: 256 x 256 x 3	float 16	info	Latest
Pose landmarker (Heavy)	Pose detector: 224 x 224 x 3 Pose landmarker: 256 x 256 x 3	float 16	info	Latest

BlazePose



0 - nose
1 - left eye (inner)
2 - left eye
3 - left eye (outer)
4 - right eye (inner)
5 - right eye
6 - right eye (outer)
7 - left ear
8 - right ear
9 - mouth (left)
10 - mouth (right)
11 - left shoulder
12 - right shoulder
13 - left elbow
14 - right elbow
15 - left wrist
16 - right wrist
17 - left pinky
18 - right pinky
19 - left index
20 - right index
21 - left thumb
22 - right thumb
23 - left hip
24 - right hip
25 - left knee
26 - right knee
27 - left ankle
28 - right ankle
29 - left heel
30 - right heel
31 - left foot index
32 - right foot index

https://colab.research.google.com/github/googlesamples/mediapipe/blob/main/examples/pose_landmarker/python/%5BMediaPipe_Python_Tasks%5D_Pose_Landmarker.ipynb

參考文獻

- 許哲豪，臺灣科技大學資訊工程系「人工智慧與邊緣運算實務」（2021~2023）

https://omnixri.blogspot.com/p/ntust-edge-ai.html

- Haoming Chen etc., 2D Human Pose Estimation: A Survey

https://arxiv.org/abs/2204.07370

- Microsoft, COCO Keypoints Dataset https://cocodataset.org/#keypoints-2020
- OpenPose, Github https://github.com/CMU-Perceptual-Computing-Lab/openpose
- Ultralytics, Yolov8 Docs – Pose Estimation
  - Overview: https://docs.ultralytics.com/tasks/pose/
  - Dataset: https://docs.ultralytics.com/datasets/pose/
- Google, MediaPipe
  - Home: https://developers.google.com/mediapipe
  - Demo: https://mediapipe-studio.webapps.google.com/home
  - Github: https://github.com/google/mediapipe

延伸閱讀

- CH.Tseng, YOLOV8 Pose Estimation

https://chtseng.wordpress.com/2023/07/14/yolov8-pose-estimation/

- Roboflow, How to Train a Custom Ultralytics YOLOv8 Pose Estimation Model

https://blog.roboflow.com/train-a-custom-yolov8-pose-estimation-model/

- 許哲豪，【vMaker Edge AI專欄 #05】Google MediaPipe快速上手 ─ 浮空手勢也能用來當作簡報播放器
  - https://omnixri.blogspot.com/2023/05/vmaker-edge-ai-05google-mediapipe.html
  - https://github.com/OmniXRI/PPT_Gesture_Demo

歐尼克斯實境互動工作室
(OmniXRI Studio)
許哲豪 (Jack Hsu)

Facebook : Jack Omnixri
FB社團： Edge AI Taiwan邊緣智能交流區
電子信箱：omnixri@gmail.com
部落格：https://omnixri.blogspot.tw
開　源：https://github.com/OmniXRI

YOUTUBE 直播： https://www.youtube.com/@omnixri1784/streams