Smart Gym Assistant

Applied AI Course

APPLIED ROOTS

Define & understand the problem:

- screen + Internet
- small form factor
- webcam
- Heart rate sensor (wearable)
- Live or on-demand

$42/month (or) ~ $1,500

APPLIED ROOTS

can we build a $100 solution?

+

$5 per month subscription

⇓

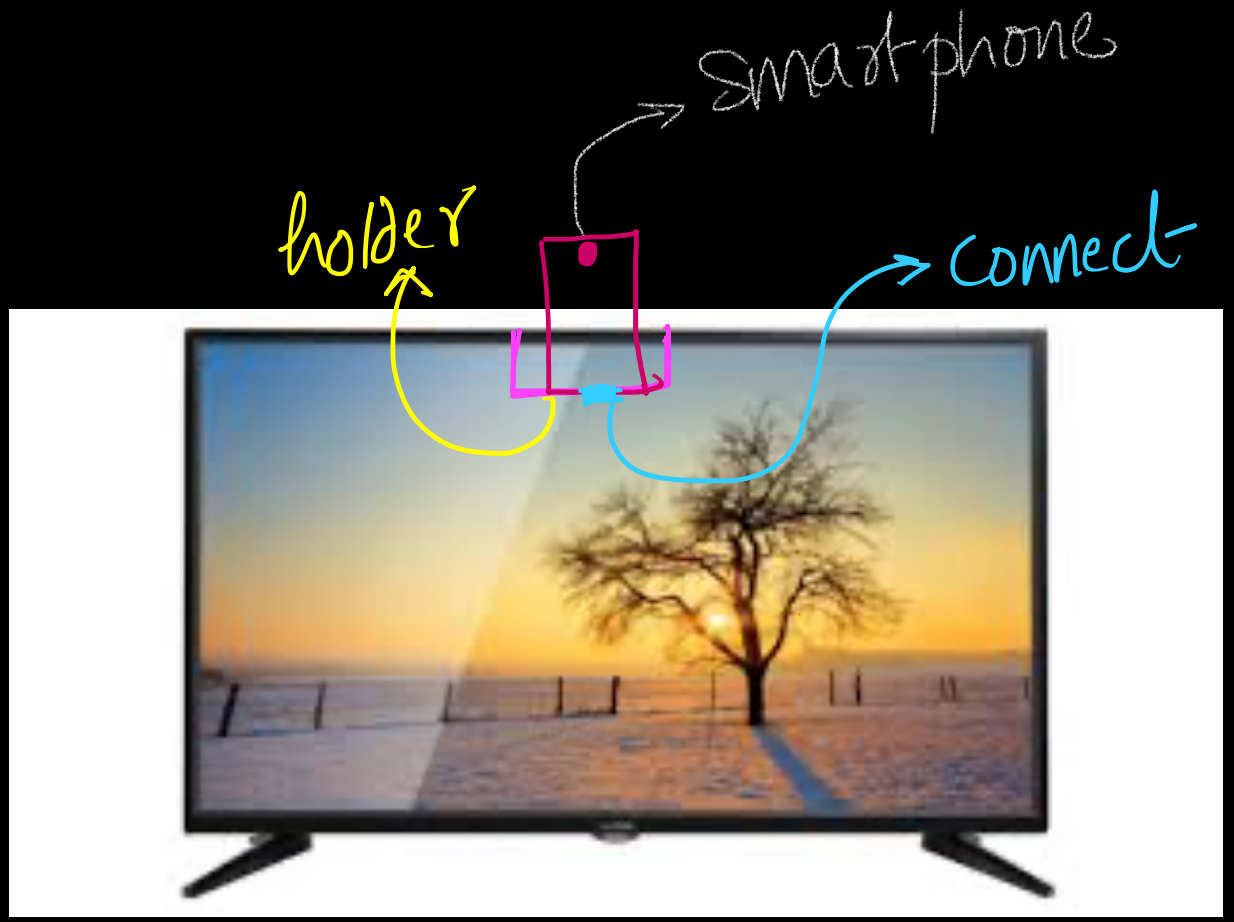
reach more customers

APPLIED ROOTS

- minimal additional hardware
- reuse everything we have

} => low-cost

One idea:



Smartphone

- Camera
- compute / AI
- Wifi / 4G
- app for tracking

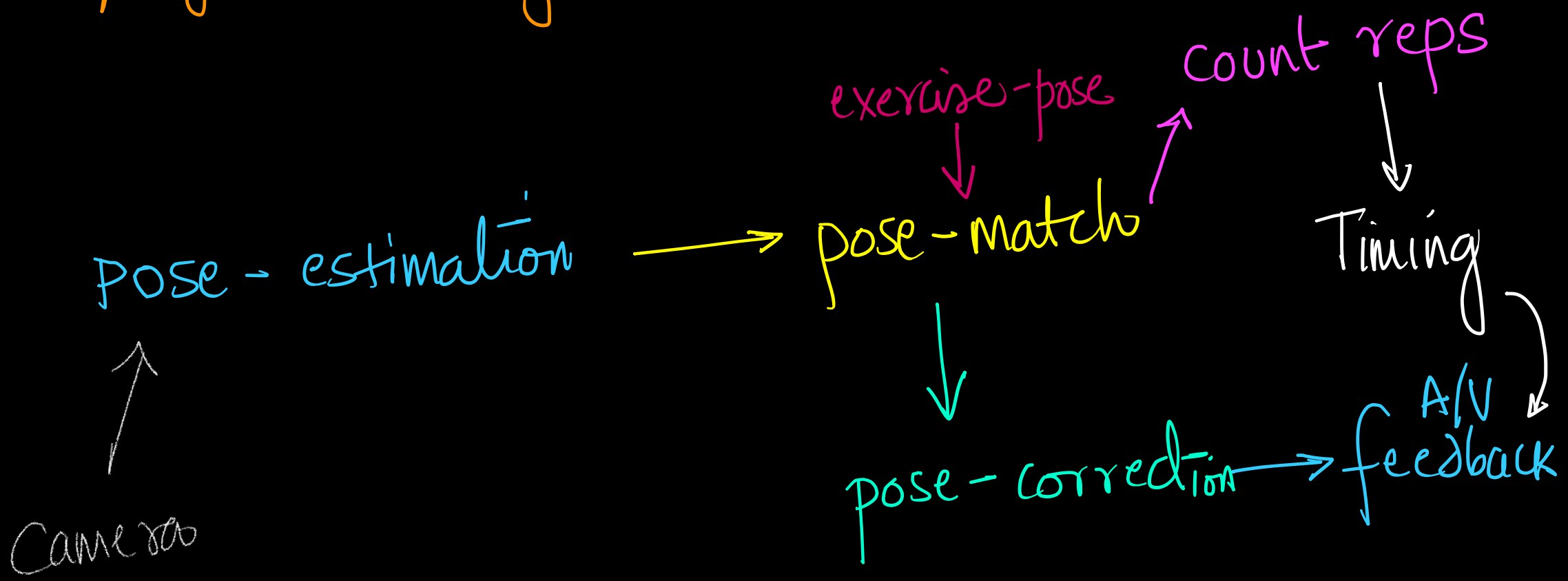
TV

- workout Video
- user's mirrored view
- Music
- Timers / Analytics

Key features:

- Correct exercise pose
- Count reps
- Timed intervals
- estimate fitness level
- Catalogue of workouts
- provide feedback [visual & audio]
- Analytics

System / High level Design:



Key Challenges

- Latency
- Compute [≥30fps, prefably 60 fps]
- Smooth TV-connect

Pose - representation :

Pose estimation :

100s of research papers from 2014 → https://paperswithcode.com/task/pose-estimation

To build a product:

- Experiment with multiple models
- Speed (fps) vs accuracy
- C++ : Android NDK , iOS objective C
- 1000's of hours of engineering & optimization

APPLIED ROOTS

Let us learn : OpenPose from CMU (2017)

- Architecture
- Loss-Metrics
- Performance-measurement
- Code

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

- Train Datasets
- Demo:

https://www.youtube.com/watch?v=pW6nZXeWlGM

- Multiple Variants [2017-2019]

APPLIED ROOTS

Open-pose

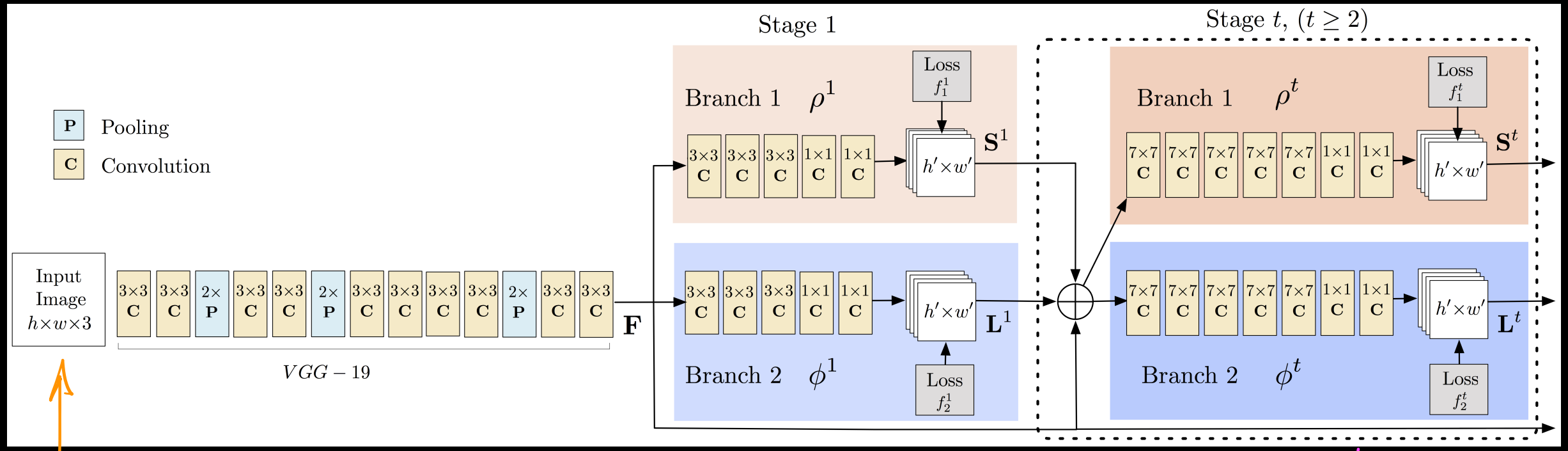
- Multi-person [ours: single person]

- Realtime on Laptop

- C++ source-code

Open-pose : Architecture

https://arxiv.org/pdf/1611.08050.pdf



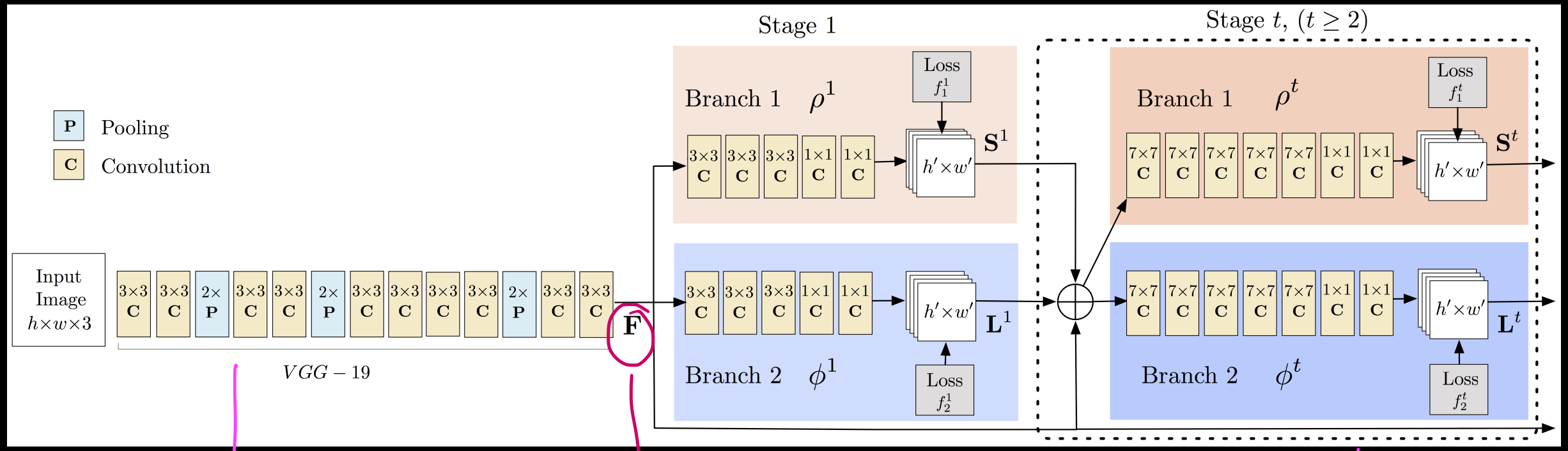
learnopencv.com

input : color-image

APPLIED ROOTS

Open-pose : Architecture

https://arxiv.org/pdf/1611.08050.pdf



first 10 layers of VGG19

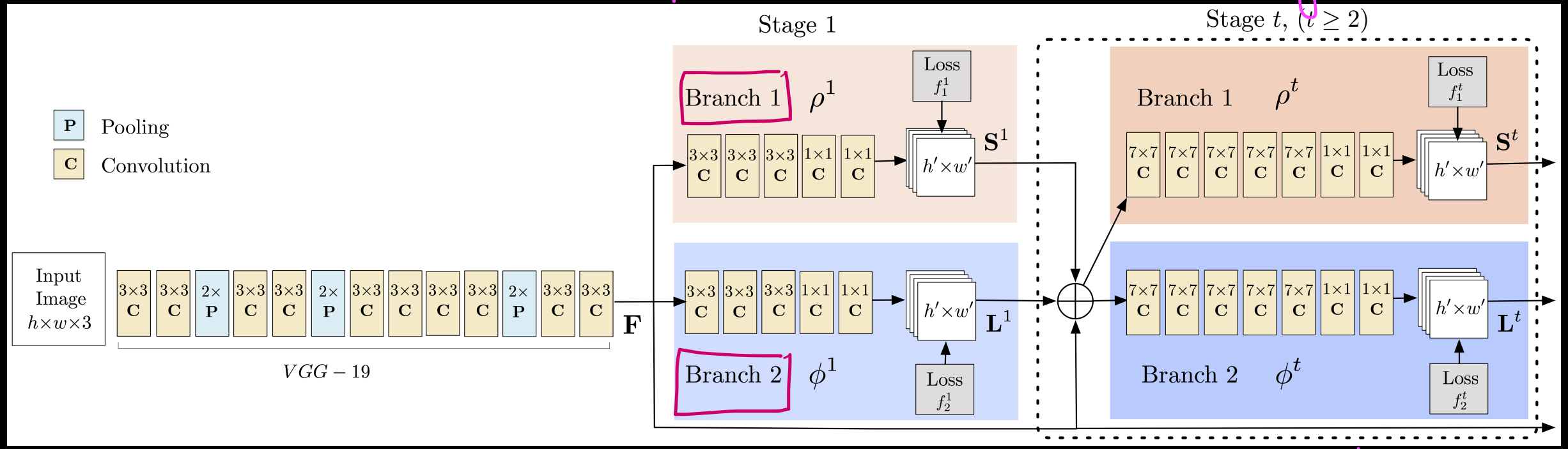
feature-map

APPLIED ROOTS

Open-pose: Architecture

https://arxiv.org/pdf/1611.08050.pdf

⟶ 2 branch multi-stage CNN



learnopencv.com

APPLIED ROOTS

Branch 1:

- Predict set of 2D confidence maps [S] of body parts.

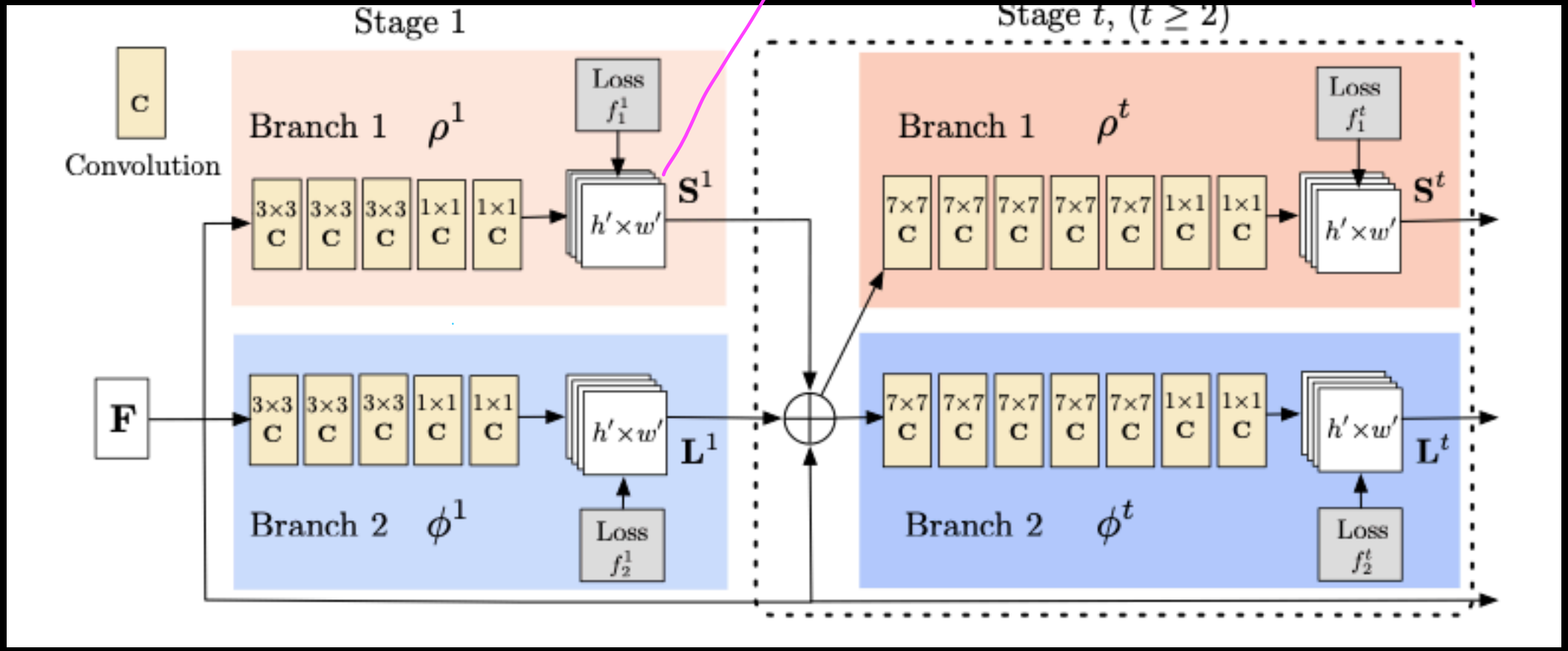
(b) Part Confidence Maps

left-elbow

left-shoulder

Branch: 1

one confidence map per key point



$$\mathbf{S}^t = \rho^t(\mathbf{F}, \mathbf{S}^{t-1}, \mathbf{L}^{t-1}), \ \forall t \geq 2, \qquad (1)$$

APPLIED ROOTS

Branch 1:

Loss:

$$f_{\mathbf{S}}^{t} = \sum_{j=1}^{J} \sum_{\mathbf{p}} \mathbf{W}(\mathbf{p}) \cdot \|\mathbf{S}_{j}^{t}(\mathbf{p}) - \mathbf{S}_{j}^{*}(\mathbf{p})\|_{2}^{2}, \quad (3)$$

= 0 if annotation is missing @ image location p

predicted confidence-map

ground truth confidence-map

ground truth confidence maps from actual positions

$$\mathbf{S}^*_{j,k}(\mathbf{p}) = \exp\left(-\frac{\|\mathbf{p} - \mathbf{x}_{j,k}\|_2^2}{\sigma^2}\right), \qquad (6)$$

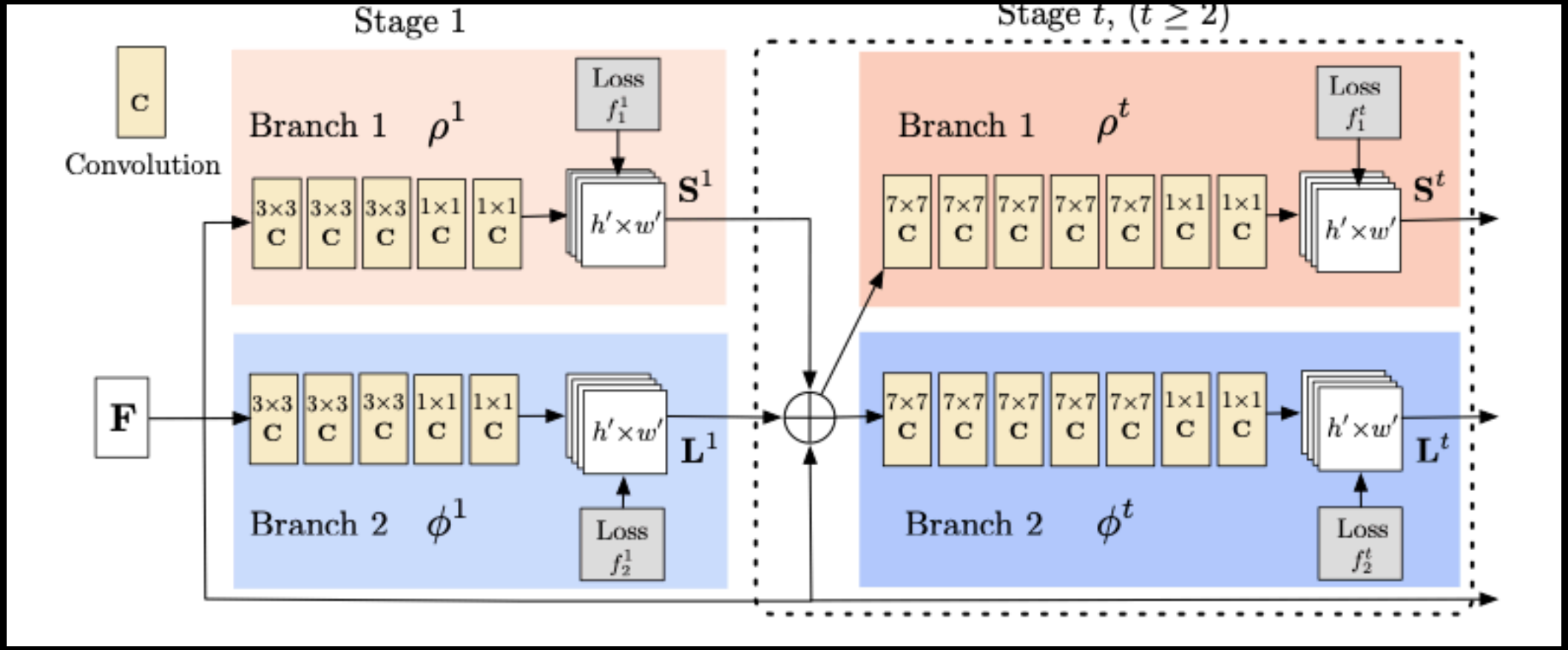
larger ⇒ more scope to err

Branch - 2 :

- predicts 2D - vector fields of part affinities for parts association

(c) Part Affinity Fields

Branch 2 :



$$\mathbf{L}^t = \phi^t(\mathbf{F}, \mathbf{S}^{t-1}, \mathbf{L}^{t-1}), \forall t \geq 2, \qquad (2)$$

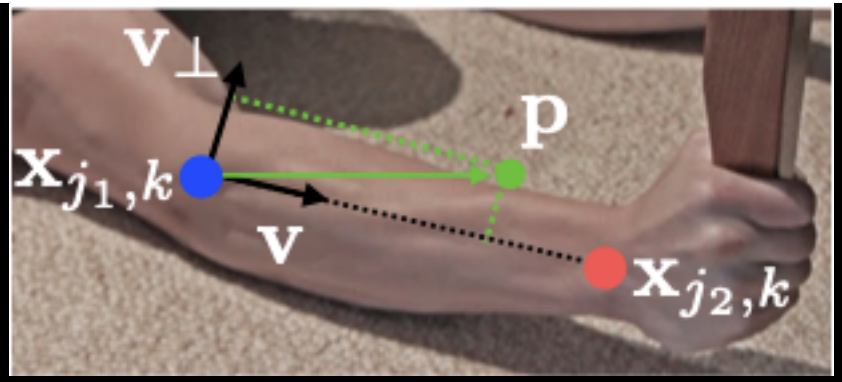
APPLIED ROOTS

Branch 2: Loss

$$f_{\mathbf{L}}^{t} = \sum_{c=1}^{C} \sum_{\mathbf{p}} \mathbf{W}(\mathbf{p}) \cdot \|\mathbf{L}_{c}^{t}(\mathbf{p}) - \mathbf{L}_{c}^{*}(\mathbf{p})\|_{2}^{2}, \quad (4)$$

ground truth part affinity vector field

$$\mathbf{L}^*_{c,k}(\mathbf{p}) = \begin{cases} \mathbf{v} & \text{if } \mathbf{p} \text{ on limb } c, k \\ \mathbf{0} & \text{otherwise.} \end{cases} \tag{8}$$



$$\mathbf{v} = (\mathbf{x}_{j_2,k} - \mathbf{x}_{j_1,k}) / ||\mathbf{x}_{j_2,k} - \mathbf{x}_{j_1,k}||_2$$

Output of the 2-branch multistage CNN

S : Confidence maps of parts

L : Vector fields of part affinities

K-partite graph matching for multi-person pose estimation

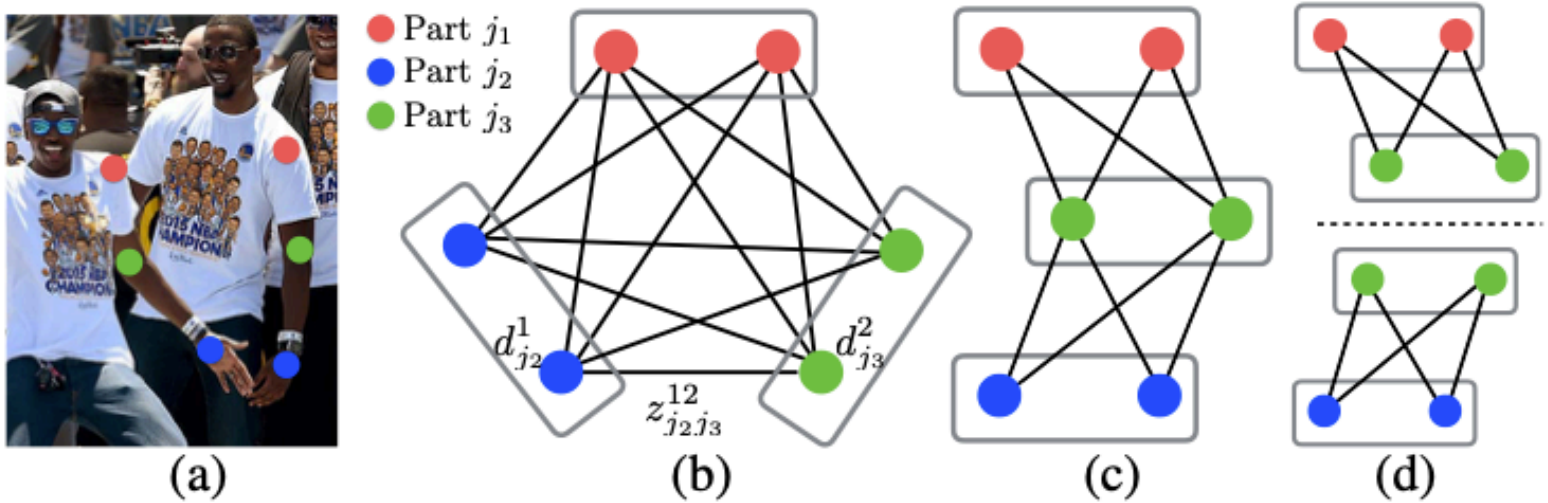


Figure 6. Graph matching. (a) Original image with part detections (b) K-partite graph (c) Tree structure (d) A set of bipartite graphs

Part Affinity fields

K-dimensional matching problem

↳ NP-hard

↳ greedy approximations

Datasets:

https://cocodataset.org/#keypoints-2018

http://human-pose.mpi-inf.mpg.de

https://www.robots.ox.ac.uk/~vgg/data/pose_evaluation/

APPLIED ROOTS

Performance - metrics :

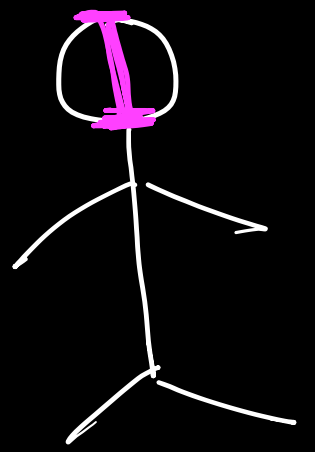
- Average precision (AP) & Average Recall (AR)
- AP^{50} => OKS = 0.5 coco-dataset
  - ↳ dependent on scale of the person
  - &
  - dist (predicted, Groundtruth)

Refer: https://nanonets.com/blog/human-pose-estimation-2d-guide/#hourglass

- Percentage of correct key points (PCK)

PCKh@ 0.5

↳ 50% of headbone link



Code:

- OpenCV can easily load Openpose model
- opencv: belter toolkit for visual-tasks