Lesson-2

Hand Pose Estimation

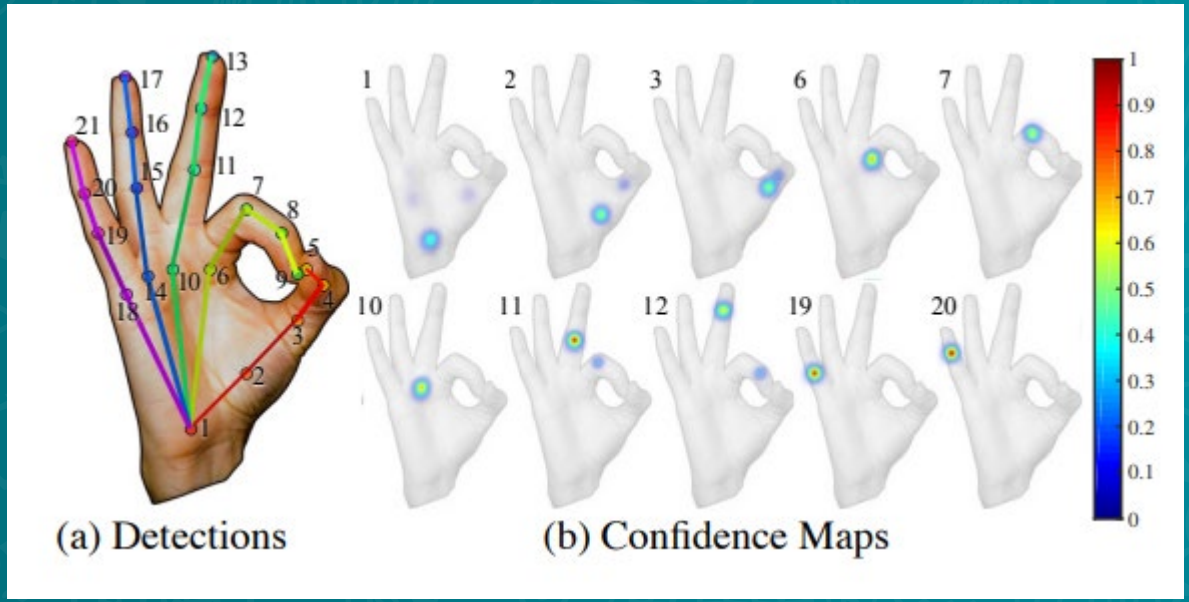
01 Hand Pose Estimation Problem

Let the digital world understand our gestures

Problem definition

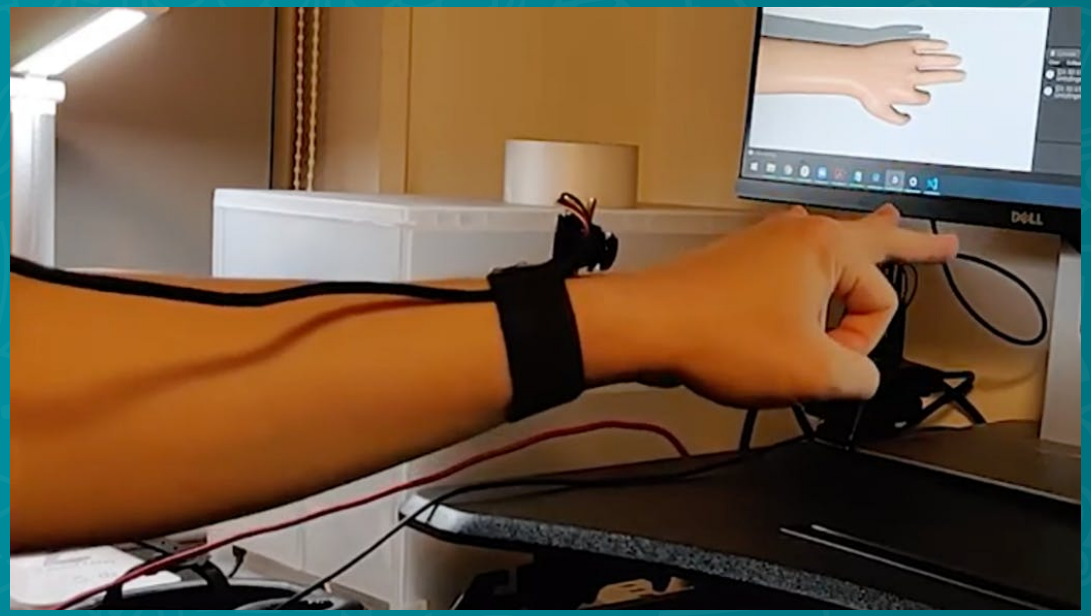
Hand pose estimation is the process of modelling human hand as a set of some parts (e.g. palm and fingers) and finding their positions in a hand image (2D estimation) or the simulation of hand parts positions in a 3D space.

> Let the digital world understand hand pose to enhance the Human Computer Interaction (HCI)



Wide Application :

- Augmented Reality (AR)
- Virtual Reality (VR)
- Mixed Reality (MR)

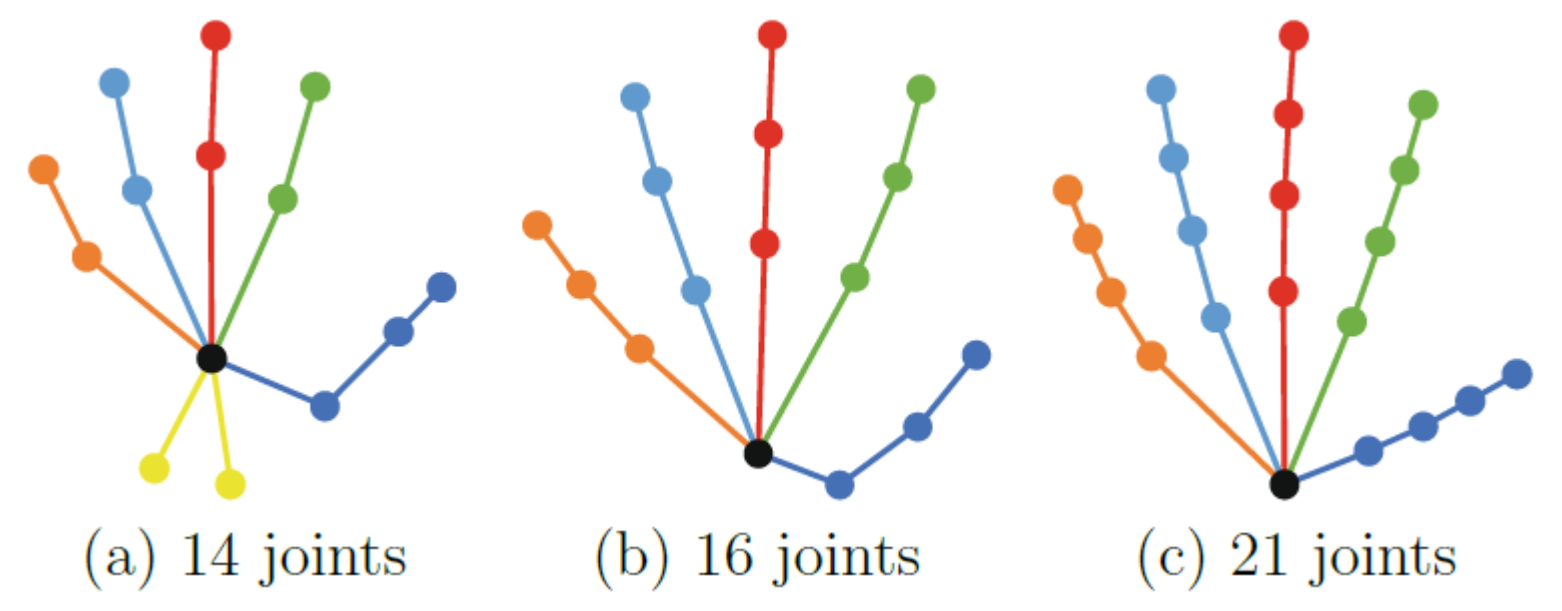


Ref: https://www.hackster.io/news/back-hand-pose-system-uses-wrist-worn-camera-for-3d-hand-pose-estimation-b4c6f2fed061

Estimation gesture

At present, the commonly used gesture estimation method in the industry is to mark the key points of the hand to represent the hand pose.

- **14 joints**
- **16 joints**
- **21 joints (most popular model)**

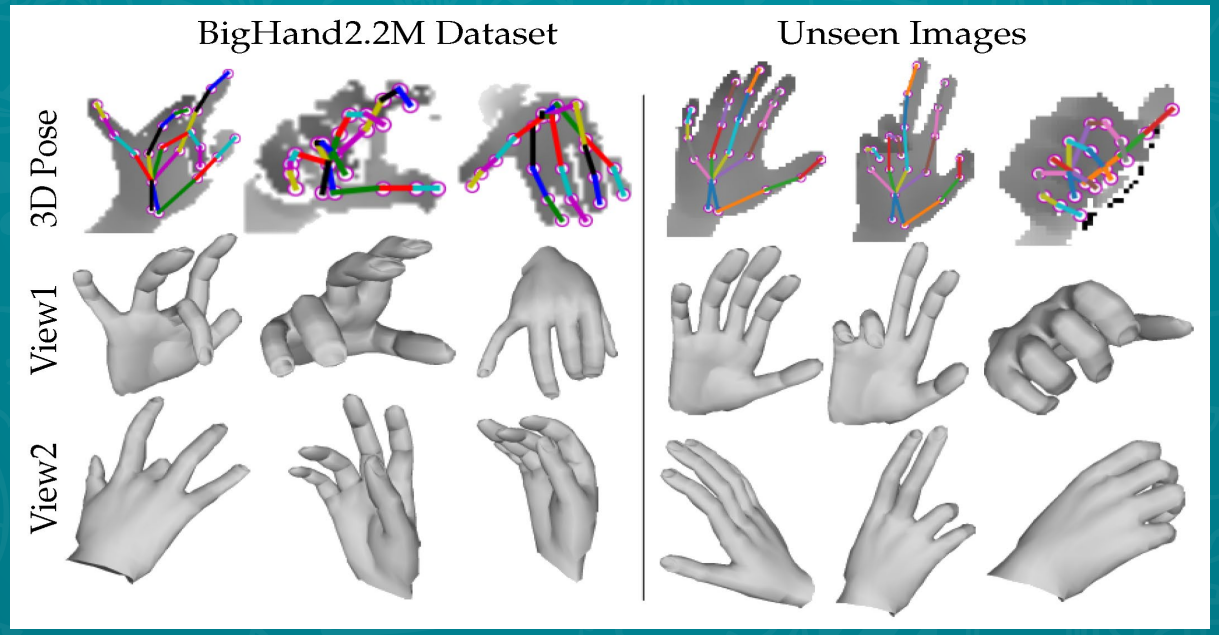


02 Approaches

3D, 2.5D, or 2D?

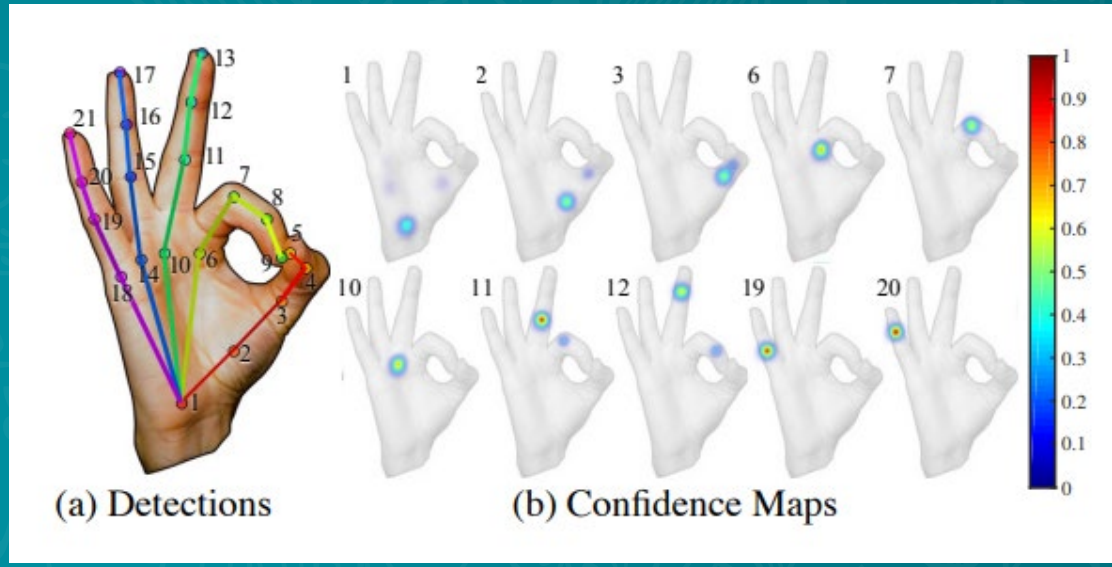
Approaches

Depth map image	RGB image
Depth cameras-not widely-used	RGB camera - common
More operation	Few operations
Easy Semantic segmentation	Difficult Semantic segmentation
Need Few datasets and training	Need more datasets and training
Robust	Not Robust



Detection-based vs. Regression-based

In the detection-based method, the model produces a probability density map for each joint. So, for example, if a network uses 21-joints model for hands, for each image it will produce 21 different probability density maps as heatmaps. The exact location of each joint can be found by applying an argmax function on corresponding heatmap.



In contrast, regression-based method tries to directly estimate the position of each joint. That is, if it uses 21-joints model, it should have **3 x 21** or **2 x 21** neurons in the last layer to predict (x; y; z) or (x; y) coordinates of each joint.

Cons : Due to the high non-linearity, training a regression based network requires more data and training iterations.

Pros: But since producing a 3D probability density function for each joint is a heavy task for a network, regression based networks is used in 3D hand pose estimation tasks.

(0.2,0.4,0.2);

(0.3,0.2,0.9)

(0.7,0.4,0.6)

.

.

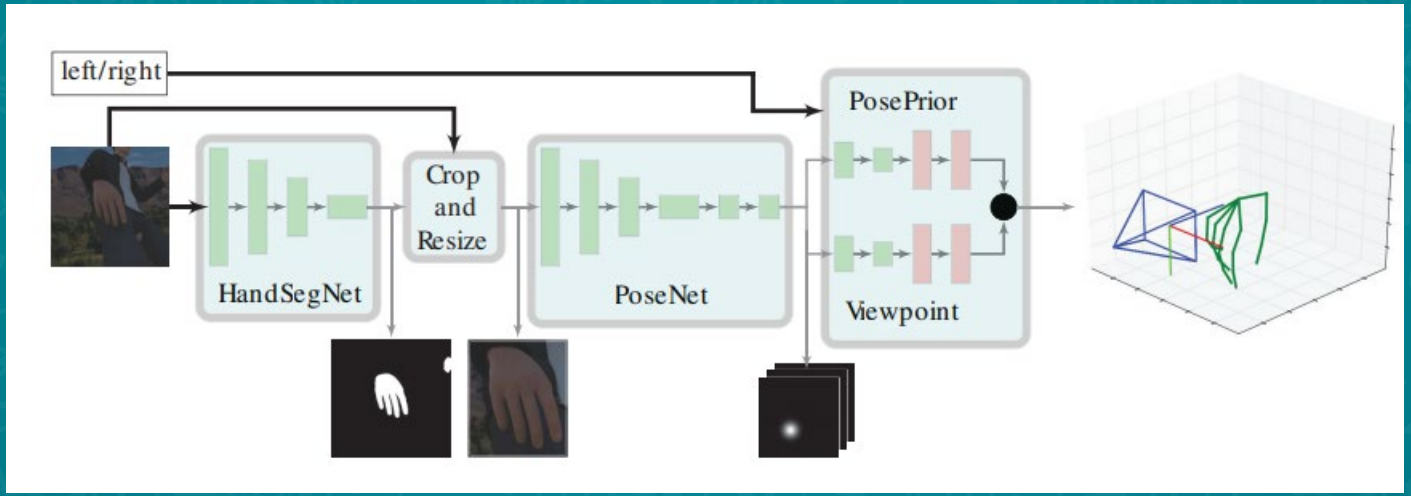
(0.8,0.9,0.1)

(0.4,0.5,0.1)

Image-based Method

Example: Zimmermann et.al.

- HandSegNet is a mask picture showing the hand pixels
- Cropped and resized
- Passed to the PoseNet (probability density function)
- Convert these 2D predictions to 3D hand estimation



Christian Zimmermann and Thomas Brox. Learning to estimate 3d hand pose from single rgb images. In Proceedings of International Conference on Computer Vision (ICCV), volume 10, 2017.

03 Hand Pose Datasets

RGB image dataset for hand pose problems

Hand Pose Datasets

Dataset	Year	RGB/D	Joints	Frames
Occluded Hands	2018	RGB	21	11,840
GANerated	2018	RGB	21	330,000
FHAD	2018	RGB+D	21	105,459
EgoDexter	2017	RGB+D	5	1,485
CMU Panoptic	2017	RGB	21	14,817
Graz16	2016	RGB+D	21	2,166

Hand Pose

THANK YOU
FOR WATCHING