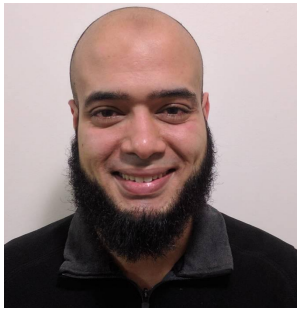
Machine Learning

Large Scale Data Annotation

Mostafa S. Ibrahim
Teaching, Training and Coaching for more than a decade!

Artificial Intelligence & Computer Vision Researcher
PhD from Simon Fraser University - Canada
Bachelor / MSc from Cairo University - Egypt
Ex-(Software Engineer / ICPC World Finalist)



Large Scale Data Annotation

- So far, we learned when we collect data in computer vision for HOW task, we have 2 scenarios
- 1) Fine-grained data collection
  - Specify a huge list of scenarios with wide variance of variables (e.g gender)
  - Start to collect the data in static parking area
  - Pros: We know the scenarios, hence the ground truth is known
    - We can also has more seniors to help annotation
  - Cons: lack actual street driving behaviour
- 2) Free driving
  - Drive and monitor the driver
  - Cons: lack the ground truth

Large Scale Data Annotation

- Assume we collected the following data
- Fine-grained data
  - 125 hours of collection coming from 500 drivers
  - Each driver recording session is 15 minutes for 180 well-defined scenarios
- Free driving
  - 160 hours free driving sessions from 20 driviners
- **Question**: How many frames in total?

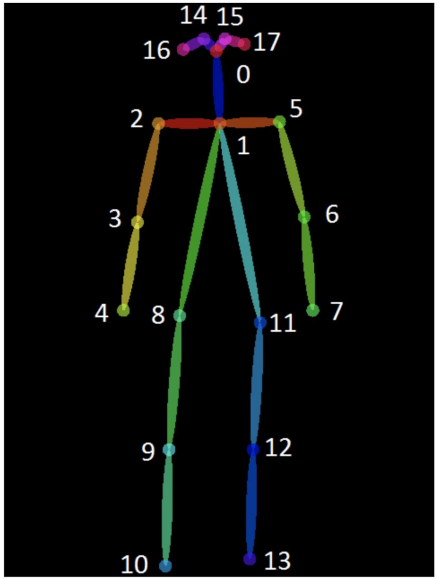
Frame-Filtering Task (CV)

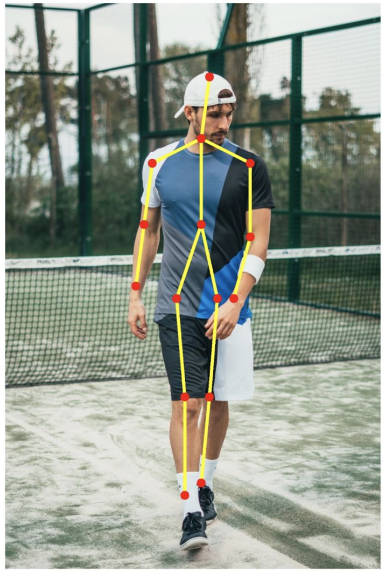
- 285 hours under 30 FPS = 285 * 60 * 60 * 30 = 30,780000 frames
- This is a huge amount of data for
  - Annotation [our goal for now] … we can't ask annotators to annotate 30M frames!
  - Training
- Practically speaking, we may not be able to process all such data
- So we do a process called **'Frame-Filtering'**: Filter (keep) amount of data that you can use and drop the remaining
  - Data you keep may be:
    - some are good from auto-labeling to be used as it is
    - some will be sent to manual annotators
- Given 30M frames, propose an efficient plan to use the data!

Consider

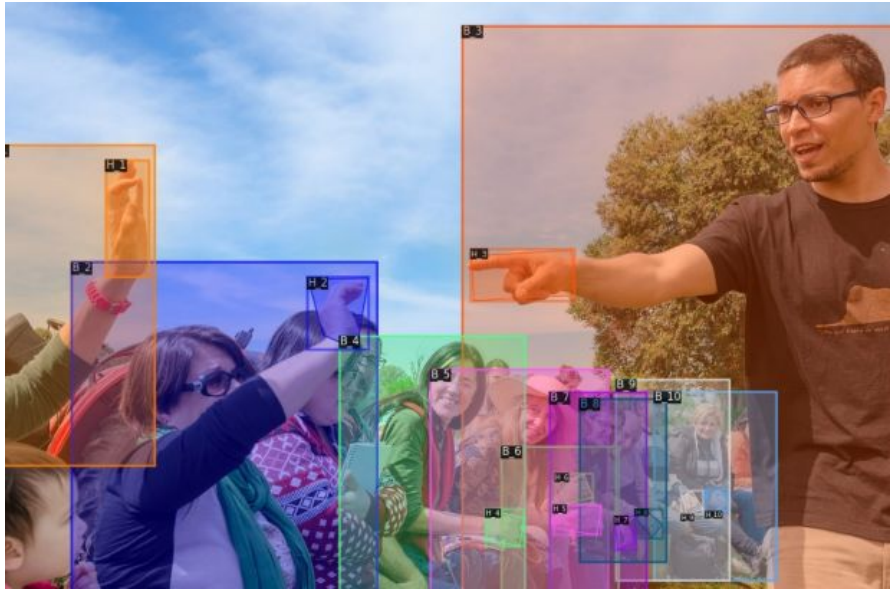
- We have some sub-optimal algorithm that helps us find the hands
  - But 15% of the time, these boxes are in wrong locations
- We already trained our POC model and it has 80% accuracy on test set
- we can utilize some SOTA
  for helping us

Possible SOTA





Given image, find persons pose



Given image, detect [persons and hands]

"Acquire knowledge and impart it to the people."

"Seek knowledge from the Cradle to the Grave."