EE655: Computer Vision & Deep Learning

Lecture 15

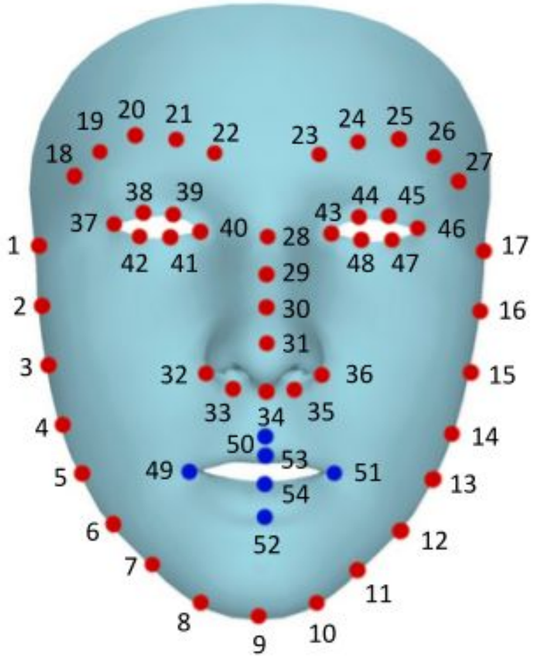
Koteswar Rao Jerripothula, PhD
Department of Electrical Engineering
IIT Kanpur

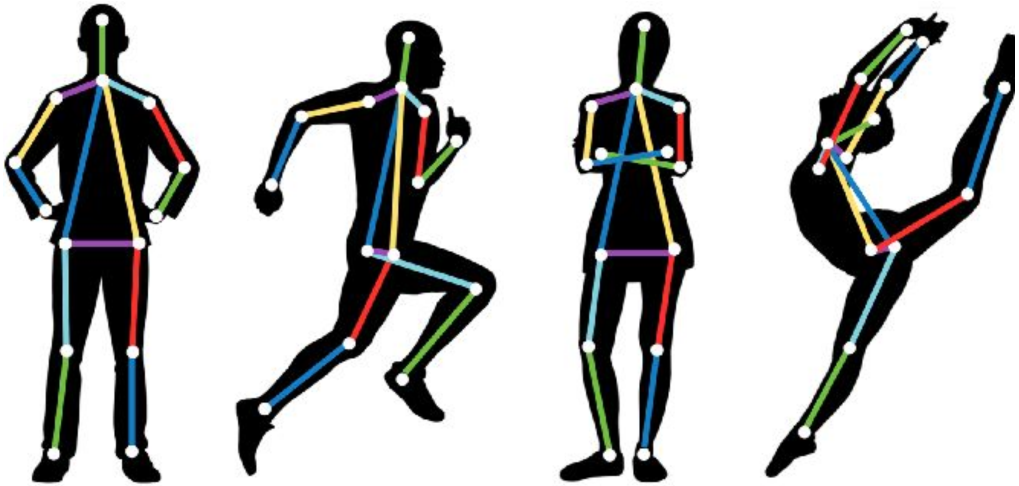
Overview

Landmark Detection

Face Recognition & Verification

Identify important landmarks are represent them as x,y coordinates.



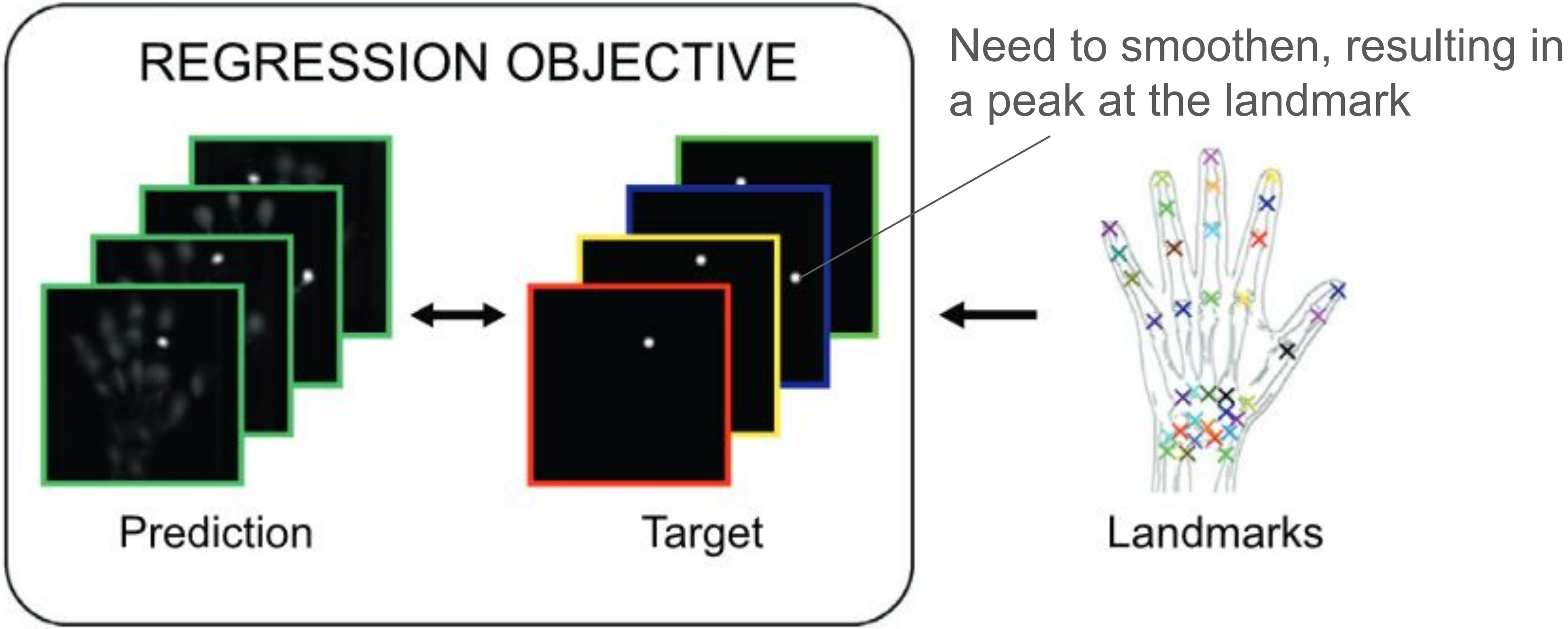


Target Vector:

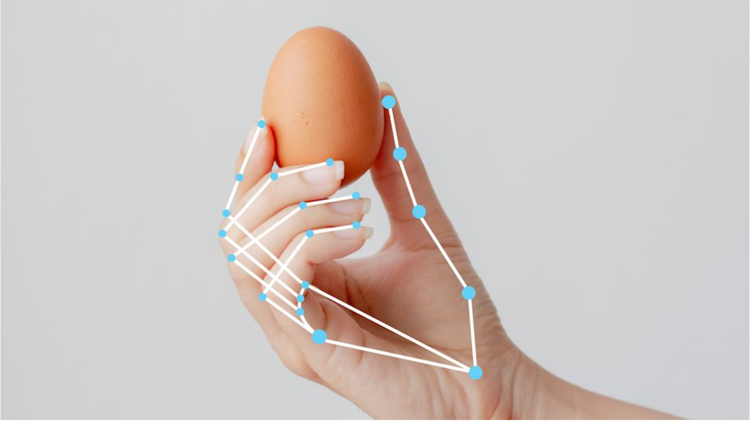
Object?
P1x
P1y
P2x
P2y
P3x
P3y
.
.

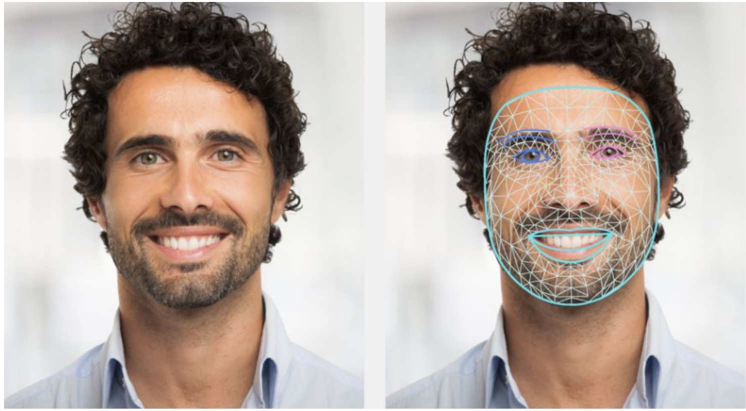
Somewhat similar to classification with localization

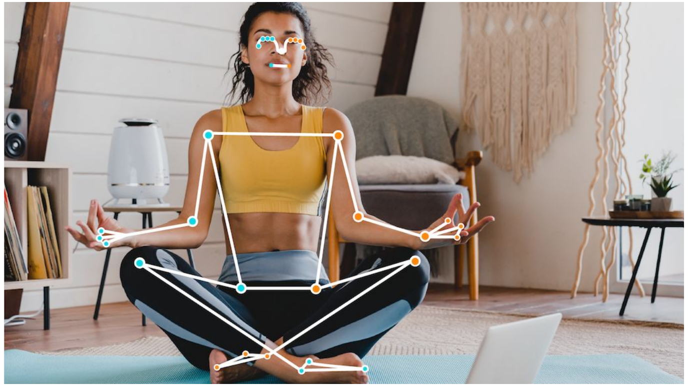
We can also predict the landmarks in the form of heatmaps



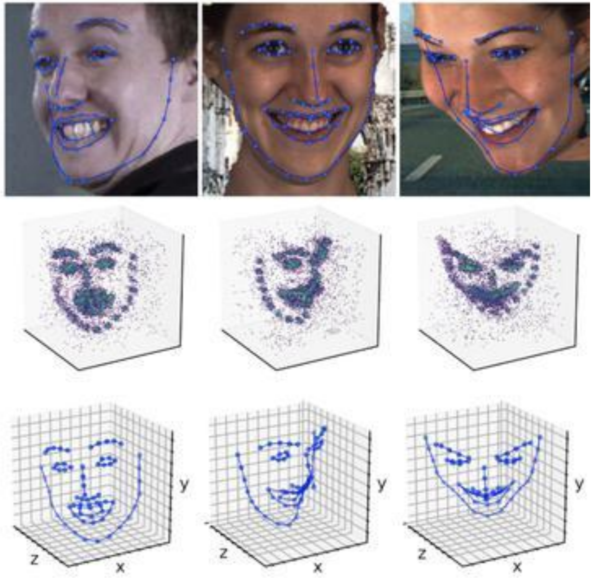
Mediapipe

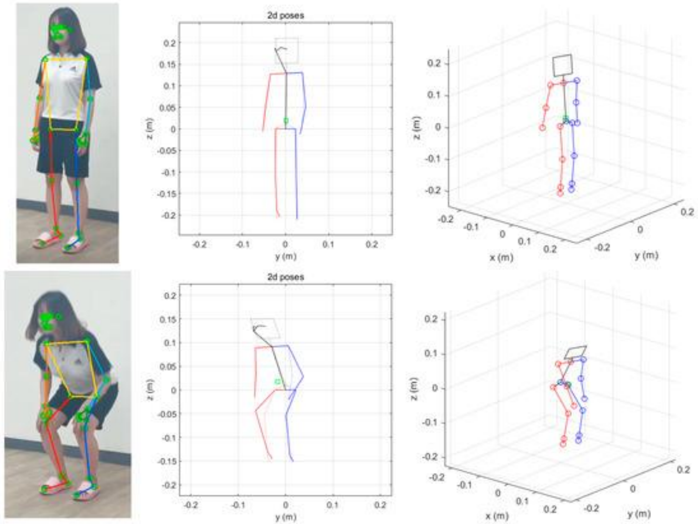






3D Landmarks: more robust





Applications

1. Face Analysis and Recognition

- **Face Recognition**: Helps in aligning and normalizing faces before recognition.
- **Facial Expression Analysis**: Detecting landmarks like eyes, nose, and mouth to analyze emotions.
- **Deepfake Detection**: Identifying inconsistencies in facial landmarks to detect AI-generated faces.
- **Avatar Animation**: Used in virtual avatars and deep learning-based face reenactment.

2. Medical and Healthcare Applications

- **Medical Image Analysis**: Detecting anatomical landmarks in X-rays, CT scans, and MRIs.
- **Surgical Navigation**: Guiding robotic-assisted surgery by recognizing key anatomical points.
- **Ophthalmology**: Landmark detection in retinal images for disease diagnosis.

3. Augmented Reality (AR) and Virtual Reality (VR)

- **Face Filters**: Snapchat, Instagram, and other AR applications use facial landmarks for real-time effects.
- **3D Reconstruction**: Landmark detection is used to create 3D models of faces or objects in VR.
- **Gesture and Gaze Tracking**: Helps in improving user interactions in AR/VR environments.

4. Human-Computer Interaction (HCI)

- **Eye Tracking**: Landmark detection around the eyes is used for gaze estimation.
- **Sign Language Recognition**: Detecting hand and face landmarks for communication aids.
- **Emotion-Based UI Adaptation**: Adapting interfaces based on user facial expressions.

5. Robotics and Autonomous Systems

- **Human Pose Estimation**: Used in robotics for understanding human movements.
- **Hand Gesture Recognition**: Controlling robotic systems through hand movements.
- **Navigation and Mapping**: Landmark detection in aerial images for UAV navigation.

6. Sports and Biomechanics

- **Athlete Performance Analysis**: Tracking body landmarks to study movements.
- **Injury Prevention**: Analyzing posture and joint angles to detect risks of injuries.
- **Physical Therapy**: Monitoring rehabilitation progress using landmark-based motion tracking.

7. Multimedia and Entertainment

- **Video Editing and Effects**: Motion tracking in films and VFX.
- **Gaming**: Motion capture for character animations.
- **Lip Syncing**: Used in animated films and games for realistic lip movement.

8. Forensics and Security

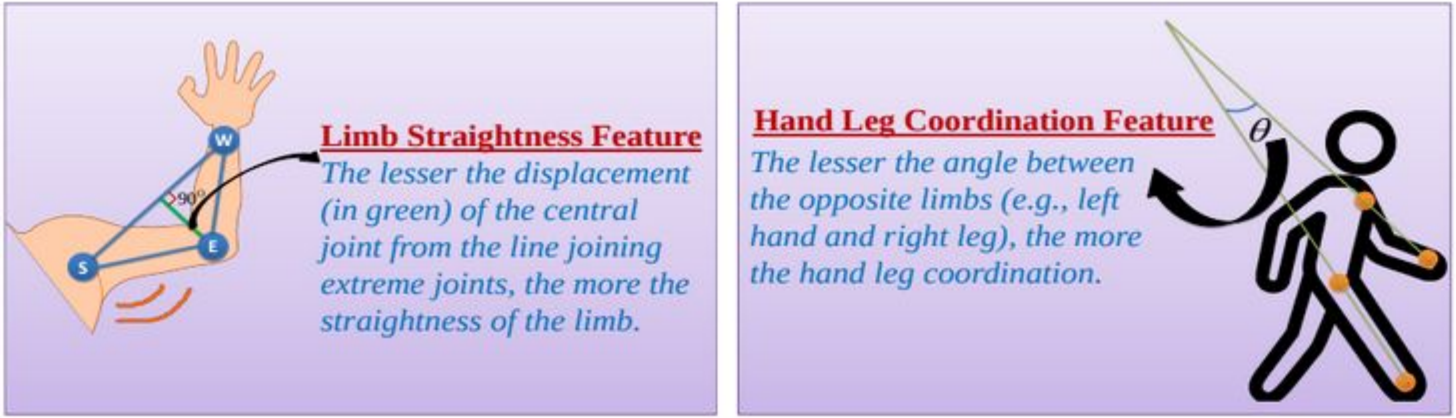
- **Criminal Identification**: Facial landmark detection in surveillance footage.
- **Forgery Detection**: Identifying alterations in images/videos.
- **Biometric Authentication**: Landmark-based fingerprint, iris, and face recognition.

Credits: ChatGPT

A case study: Neurological GAIT

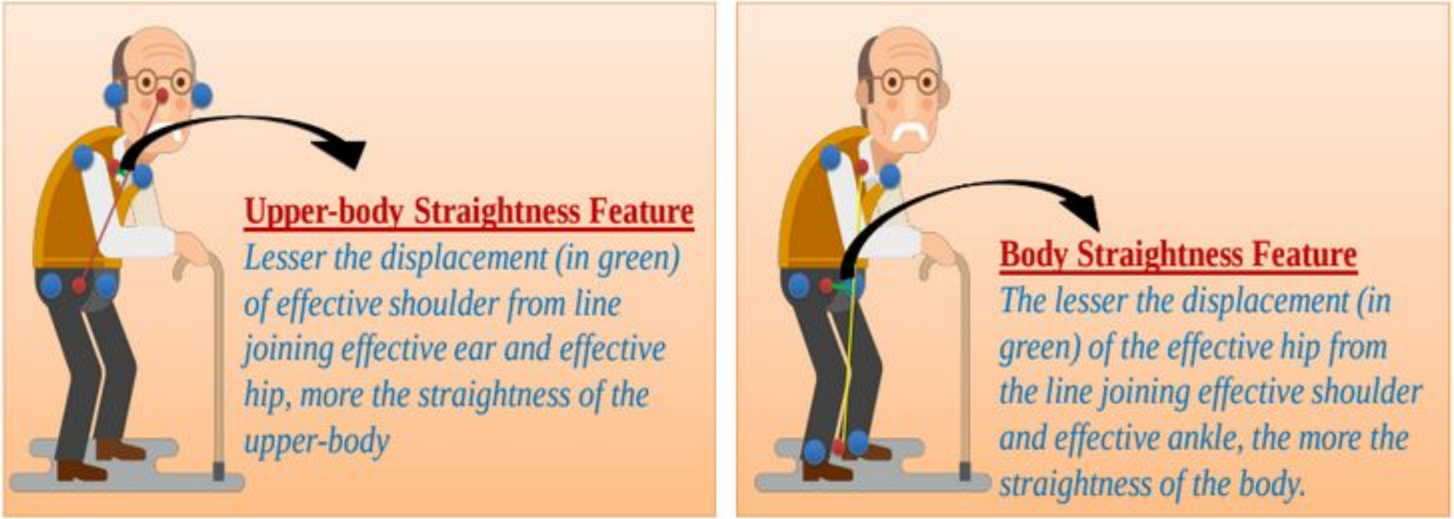
Gait Abnormalities

Features



Limb Straightness and Hand-leg Coordination

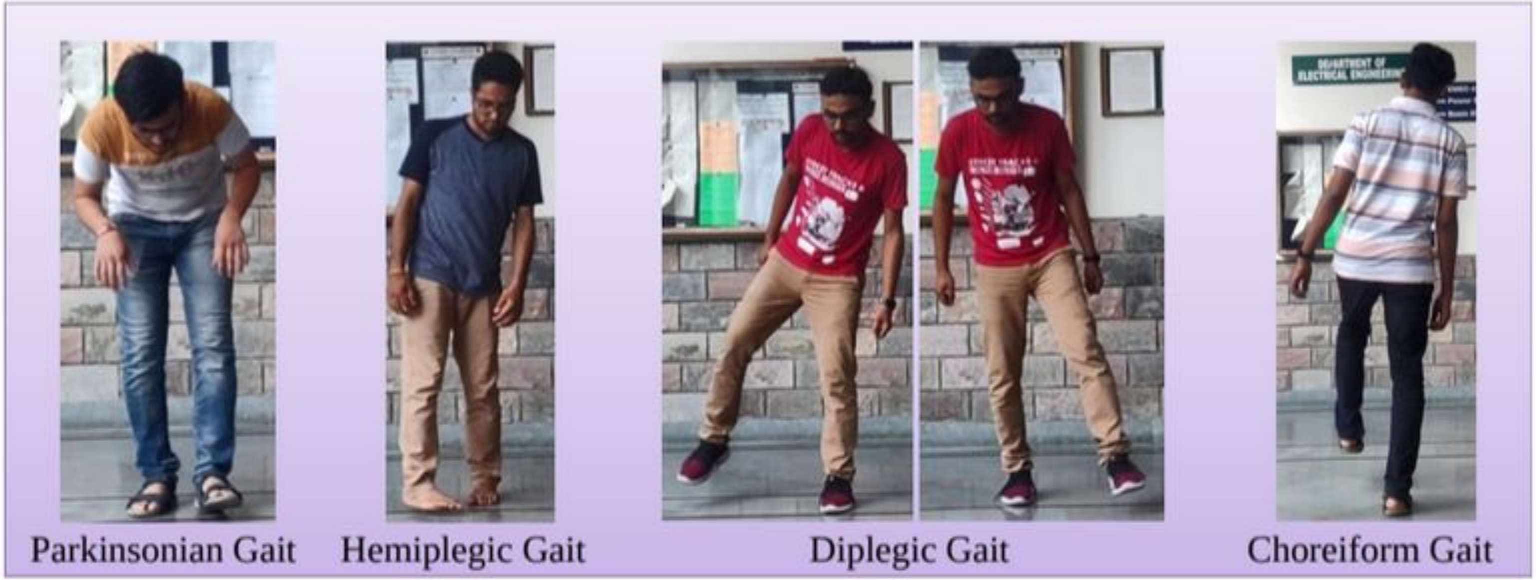
Features (contd..)



Upper-body and Total-body Straightness

Features (contd..)

Central Distances and Mutual Distances



A glimpse of our dataset as we can see our participants mimicking different abnormal gaits

Just 250+ Videos

Video-level Feature: Concatenation of average and standard deviation of frame-level feature vectors

TEST CLASSIFICATION ACCURACIES FOR MULTIPLE ABNORMALITIES DETECTION USING DIFFERENT METHODS AND MACHINE LEARNING ALGORITHMS.

	kNN	Tree	SVM	SGD	Random Forest	Neural Network	Naive Bayes	Logistic Regression	AdaBoost
AlphaPose [13]	0.197	0.591	0.470	0.515	0.470	0.561	0.455	**0.758**	0.576
3D-CNN [25]	0.273	0.379	0.455	0.530	0.379	0.485	0.288	0.591	0.348
AlphaPose [13]+ 3D-CNN [25]	0.424	0.394	0.545	0.606	0.379	0.636	0.470	0.652	0.379
Ours	**0.636**	**0.636**	**0.788**	**0.864**	**0.833**	**0.788**	**0.773**	0.212	**0.712**

Face verification vs. face recognition

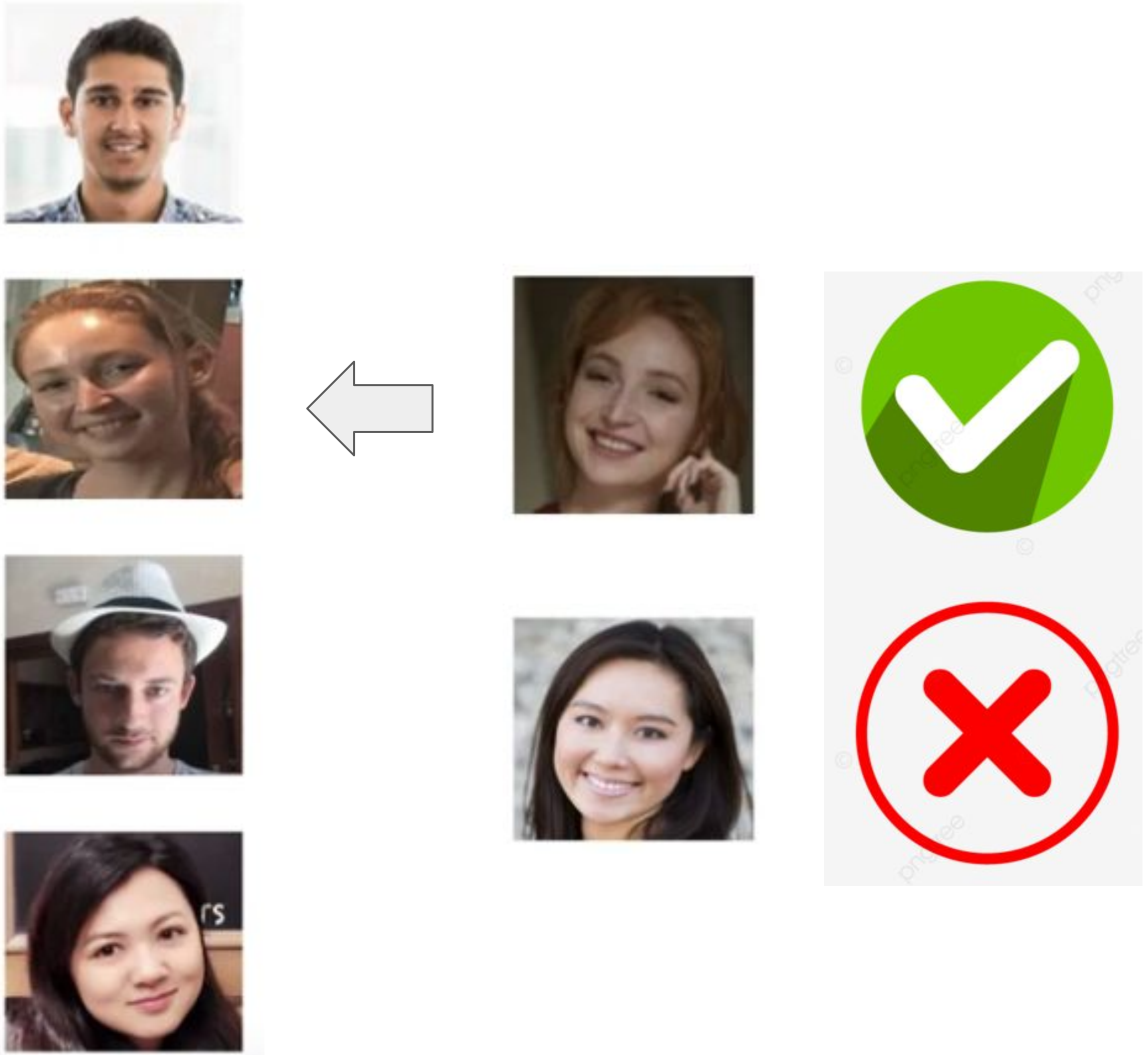
Verification 1:1

- Input image, name/ID
- Output whether the input image is that of the claimed person

Recognition 1:K

- Has a database of K persons
- Get an input image
- Output ID if the image is any of the K persons (or "not recognized")

One shot learning



Learning from one example to recognize the person again

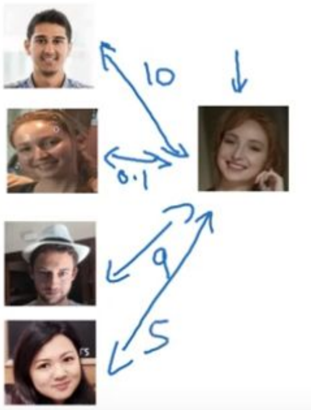
Learning a "similarity" function

d(img1,img2) = degree of difference between images

If d(img1,img2) $\leq \tau$ "same"

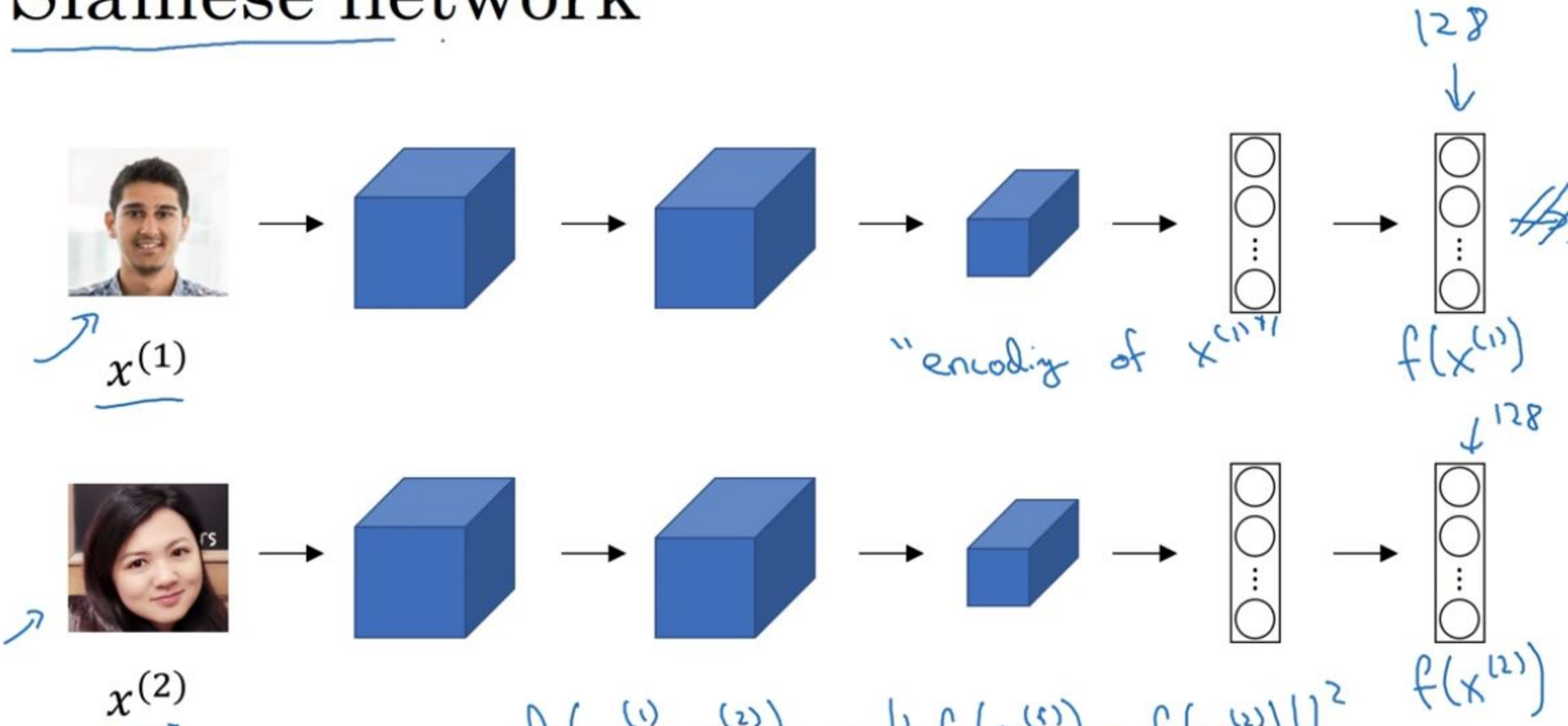
$> \tau$ "different"

Face Verification



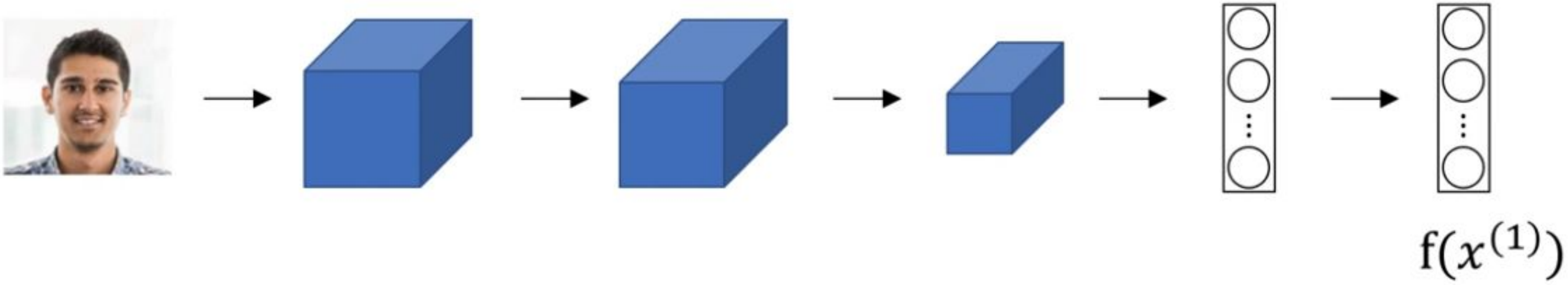
Face Recognition

Siamese network



$$d(x^{(1)}, x^{(2)}) = \| f(x^{(1)}) - f(x^{(2)}) \|_2^2$$

Goal of learning



128

Parameters of NN define an encoding $f(x^{(i)})$

Learn parameters so that:

If $x^{(i)}, x^{(j)}$ are the same person, $\left\|f(x^{(i)}) - f(x^{(j)})\right\|^2$ is small.

If $x^{(i)}, x^{(j)}$ are different persons, $\left\|f(x^{(i)}) - f(x^{(j)})\right\|^2$ is large.

Triplet Loss

Learning Objective

Anchor

Positive

Anchor

Negative

Want: $\underbrace{\|f(A)-f(P)\|^2}_{d(A,P)} \le \underbrace{\|f(A)-f(N)\|^2}_{d(A,N)}$

$$\|f(A)-f(P)\|^2 - \|f(A)-f(N)\|^2 \le 0$$

Looking at three images at a time

Adding margin to avoid trivial solution

$$f(\text{img}) = \vec{0}$$

$$\|\underbrace{f(A) - f(P)}_{0}\|^2 - \|\underbrace{f(A) - f(N)}_{0}\|^2 + \underbrace{\alpha}_{\text{margin}} \leq \underline{0}$$

Loss function

Given 3 images A, P, N:

$$\mathcal{L}(A,P,N) = \max\left(\|f(A)-f(P)\|^2 - \|f(A)-f(N)\|^2 + \alpha,\ 0\right)$$

$$J = \sum_{i=1}^{m} \mathcal{L}(A^{(i)}, P^{(i)}, N^{(i)})$$

Training set: 10k pictures of 1k persons

Notice that you do need multiple pictures of the same person here.

Choosing the triplets A,P,N

During training, if A,P,N are chosen randomly, $d(A, P) + \alpha \le d(A, N)$ is easily satisfied.

$\|f(A) - f(P)\|^2 + \alpha \le \|f(A) - f(N)\|^2$

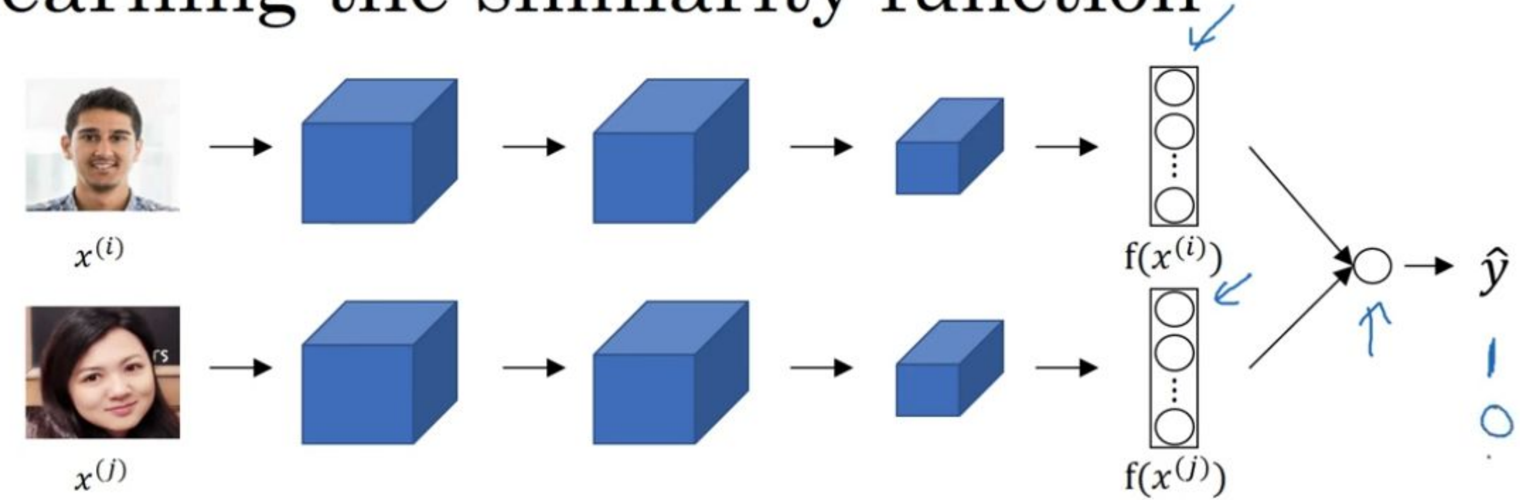
Choose triplets that're "hard" to train on.

$d(A,P) + \alpha \le d(A,N)$

$d(A,P) \approx d(A,N)$

Another Alternative

Learning the similarity function



$$\hat{y} = \sigma\left(\sum_{k=1}^{128} w_i \left| f(x^{(i)})_k - f(x^{(j)})_k \right| + b\right)$$

Important Note

What we have seen is the training process, where there maybe multiple images of the same person.

During inference, we need just a single reference image of a person for face verification.

Reference

https://ai.google.dev/edge/mediapipe/solutions/guide

https://ieeexplore.ieee.org/document/9287163

https://www.youtube.com/watch?v=-FfMVnwXrZ0&list=PLkDaE6sCZn6Gl29AoE31iwdVwSG-KnDzF&index=32

https://www.youtube.com/watch?v=96b_weTZb2w&list=PLkDaE6sCZn6Gl29AoE31iwdVwSG-KnDzF&index=33

https://www.youtube.com/watch?v=6jfw8MuKwpI&list=PLkDaE6sCZn6Gl29AoE31iwdVwSG-KnDzF&index=34

https://www.youtube.com/watch?v=d2XB5-tuCWU&list=PLkDaE6sCZn6Gl29AoE31iwdVwSG-KnDzF&index=35

https://www.youtube.com/watch?v=0NSLgoEtdnw&list=PLkDaE6sCZn6Gl29AoE31iwdVwSG-KnDzF&index=36