+ What is Face Recognition?

Recognize face and live human vs Picture.

Technologies:

Verification → Input image, name / ID
(1:1) → Output whether the input image is that of the claimed person.

Recognition :→ Has a database of K persons.
(1:K) → Get input image.
→ output ID if the image of the K persons

+ One - Shot learning

Only one image (or opportunity to recognize)

Traditionally Deep Learning algorithms don't work well with one image

Learning needs to happen from just one image!

("learning "similarity" function

$d(img1, img2)$ = degree of difference.

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

$> T$ "different"

+ The Seimese Network

Instead of comparing raw pics, you can compare encoded versions.

$x^{(1)}$ => ConvNet -> [encoding] $f(x^{(1)})$

$x^{(2)}$ => ConvNet => [encoding] $f(x^{(2)})$

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

If they are the same person, the distance (d) will be small.

+ Triplet Loss

Anchor Image (A), Positive (P), Negative (N)

We want the distance between the anchor and the positive value to be smaller than the negative

$$\underbrace{\| f(A) - f(P) \|^2}_{d(A,P)} + \alpha \leq \underbrace{\| f(A) - f(N) \|^2}_{d(A,N)}$$

$$\| f(A) - f(P) \|^2 - \| f(A) - f(N) \|^2 + \alpha \leq 0.$$

α ↓ margin.

The loss function :

Given 3 image A, P, N.

$$L(A,P,N) = \max(\|f(A)-f(P)\|^2 - \|f(A)-f(N)\|^2 + \alpha,\ 0)$$

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

- Choosing the triplets A, P, N.
  + If random, $d(A,P) + \alpha \leq d(A,N)$ is easily satisfies
  + We choose triplets that are "hard" to train on.

+ Face Verification and binary classification.

Another way to approach the problem is to join the binary comparison as a binary problem.

It is a different approach to the tripplet approach.