NPTEL ONLINE CERTIFICATION COURSES

# Course Name: Deep Learning

Faculty Name: Prof. P. K. Biswas
Department : E & ECE, IIT Kharagpur

## Topic

Lecture 51: Face Recognition

# CONCEPTS COVERED

Concepts Covered:

- Face Recognition System
- One shot learning
- FaceNet
- Triplet Loss
- Triplet Selection


# Face Recognition System

Face recognition system has become a integral part of our modern day to day life. Various application of face recognition system are:

- Payments
- Access and security
- Criminal identification
- Advertising
- Healthcare

# Face Recognition System

Challenges:

- ❑ Different illumination condition.
- ❑ Different Pose and orientation of image.
- ❑ Other variational conditions.
- ❑ Limited Dataset for training.

Image Source: Schroff, Florian, Dmitry Kalenichenko, and James Philbin. "Facenet: A unified embedding for face recognition and clustering." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 815-823. 2015.

# One Shot learning

- One-shot learning is an object categorization problem, found mostly in computer vision.
- Most machine learning based object categorization algorithms require training on hundreds or thousands of samples/images and very large datasets,
- One-shot learning aims to learn information about object categories from one, or only a few, training samples/images.

# One Shot learning

Face Recognition as a One shot Learning:

- Consider a facial recognition system which is used by a small organization for security purpose.
- It has one image of every person working in that company.
- The network needs to be train using those few images,
- It can identify a person who is not working in the company and also the verify who is working currently in the company.
- This problem becomes one shot or few shot learning problem.

# FaceNet

FaceNet learns a embedding function $f(x); \quad \|f(x)\|_2 = 1)$

$$f: x \in R^{M \times N} \rightarrow R^d; \quad d < M \times N$$

Take two images $x_i$ and $x_j$

$$\|f(x_i) - f(x_j)\|^2$$

Small if $x_i$ and $x_j$ are same person

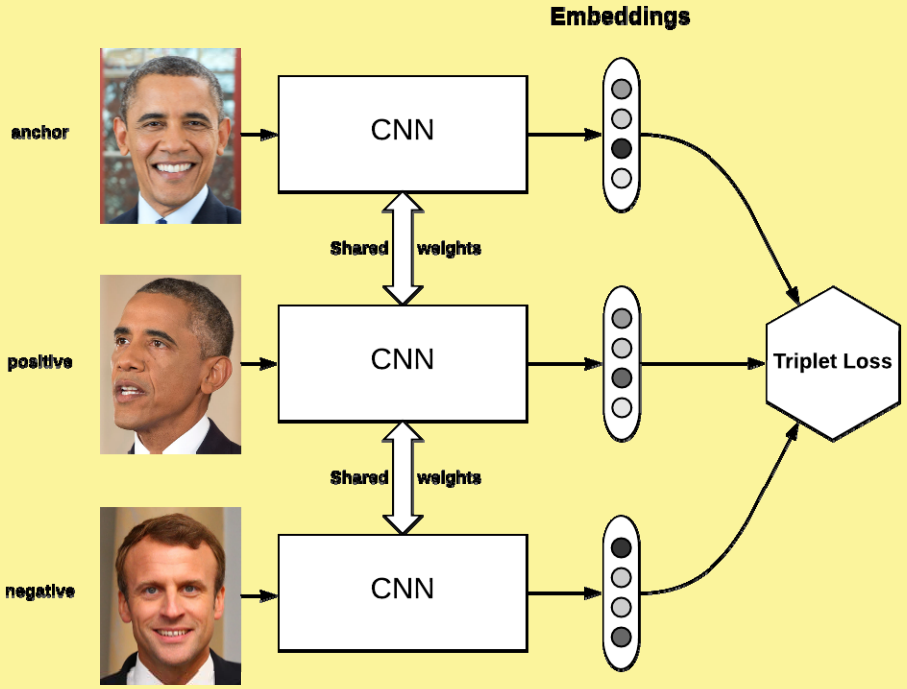
Large otherwise

# FaceNet


- FaceNet uses a deep CNN model to learn the embedding function $f(x)$.
- It consists of a batch input layer and a deep CNN
- Followed by L2 normalization, which results in the face embedding.

Image Source: Schroff, Florian, Dmitry Kalenichenko, and James Philbin. "Facenet: A unified embedding for face recognition and clustering." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 815-823. 2015.

# Training/Triplet Loss




- Minimize **triplet loss function** :-loss function using **three** images
- An anchor image A, a positive image P (same person as the anchor), and a negative image N (different person than the anchor).
- Distance d(f(A), f(P)) must be less than or equal to the distance d(f(A), f(N))

Image Source: https://omoindrot.github.io/triplet-loss

# Triplet Loss

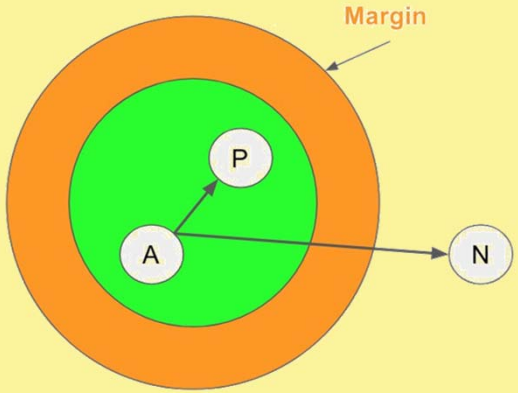
Desired:- $\|f(x^a) - f(x^p)\|_2^2 < \|f(x^a) - f(x^n)\|_2^2$

The problem:

- The model can learn to make the same encoding for different images of the same person.
- That makes $\|f(x^a) - f(x^p)\|_2 = 0$
- Unfortunately, it will satisfy the triplet loss function.
- Model stops learning.
- Solution: add a margin $\alpha$ to always have a gap between A and P versus A and N.

# Triplet Loss

Thus : $\|f(x^a) - f(x^p)\|_2^2 + \alpha < \|f(x^a) - f(x^n)\|_2^2$

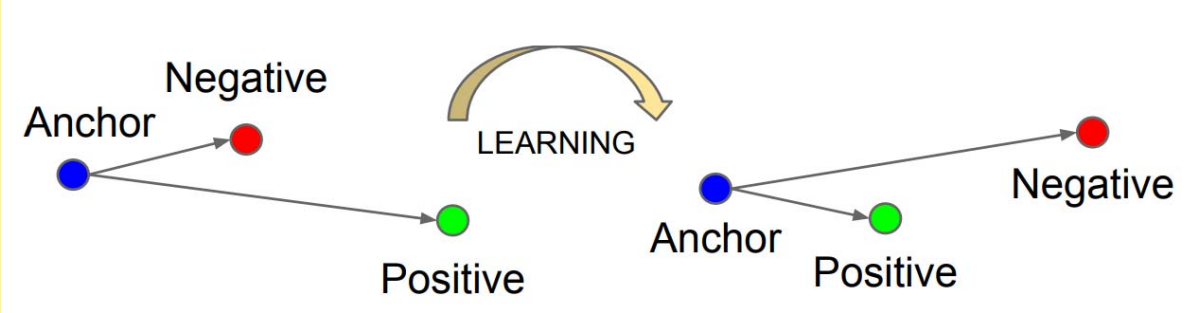



The loss that is being minimized is then

$$L = \sum_{i=1}^{N} [\|f(x^a) - f(x^p)\|_2^2 - \|f(x^a) - f(x^n)\|_2^2 + \alpha]$$

Image Source :https://medium.com/@ahmdtaha/facenet-a-unified-embedding-for-face-recognition-and-clustering-7d34abde9

# Triplet Loss




- The Triplet Loss minimizes the distance between an anchor and a positive.
- Maximizes the distance between the anchor and a negative.
- Compact clusters of embedding of same person.
- Pictures of the same person become close to each other.
- Pictures of different persons are far from each other.

Image Source: Schroff, Florian, Dmitry Kalenichenko, and James Philbin. "Facenet: A unified embedding for face recognition and clustering." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 815-823. 2015.

# Triplet Selection

- Selecting all possible triplets would result in many triplets that satisfy $\|f(x^a - x^p\|_2^2 + \alpha < \|f(x^a - x^n\|_2^2)$.
- These triplets would not contribute to the training and result in slower convergence, as they would still be passed through the network.
- It is crucial to select hard triplets, that are active and can therefore contribute to improving the model.

Image Source: https://omoindrot.github.io/triplet-loss

# Triplet Selection

- In order to ensure fast convergence it is crucial to select triplets that violate the triplet constraint .
- This means that, given $x_i^a$ , we want to select an $x_i^p$ (hard positive) such that $argmax_{x_i^p} \left\| f(x_i^a) - f(x_i^p) \right\|_2^2$
- Similarly select$x_i^n$ (hard negative) such that $argmin_{x_i^n} \| f(x_i^a) - f(x_i^n) \|_2^2$


Image Source: https://omoindrot.github.io/triplet-loss

# Triplet Selection

- Selecting hardest negative may collapse the model: f(x)=0.
- Select semi-hard negative

$$\|f(x^a) - f(x^p)\|_2^2 < \|f(x^a) - f(x^n)\|_2^2$$

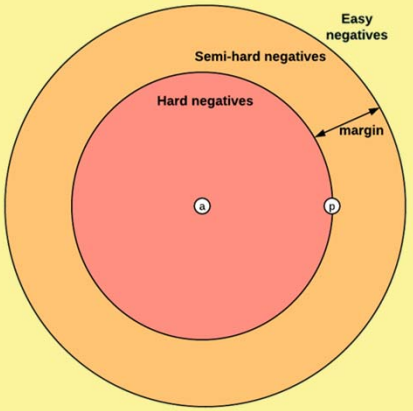



Image Source: https://omoindrot.github.io/triplet-loss

# Face Verification

- Pass the reference image and the query image through the embedding network.
- Use the distance between them for verification.

$$d(reference, query) = \|f(refernce) - f(query)\|_2^2$$

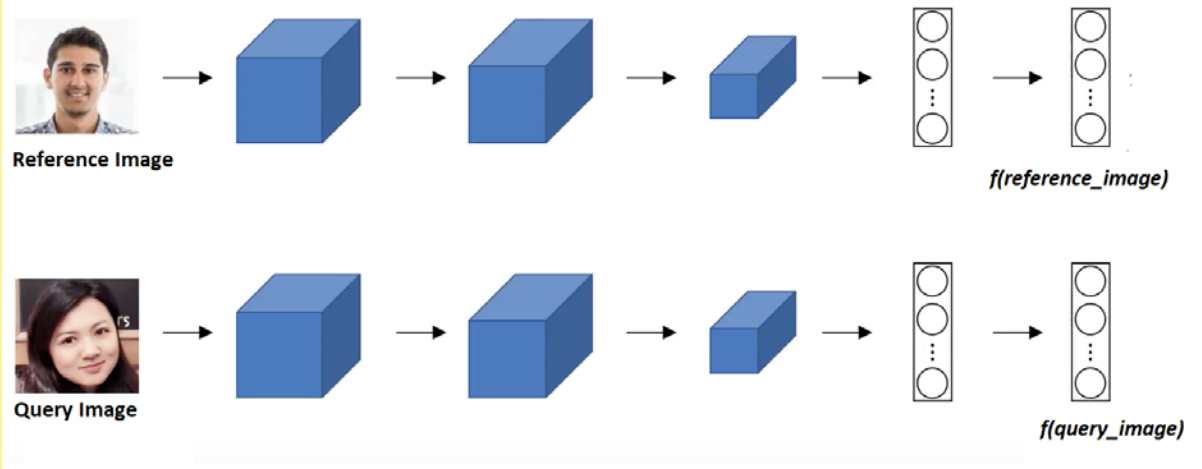



Image Source :
https://www.coursera.org/learn/convolutional-neural-networks?specialization=deep-learning

## References:


- Schroff, Florian, Dmitry Kalenichenko, and James Philbin. "Facenet: A unified embedding for face recognition and clustering." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 815-823. 2015.
- https://omoindrot.github.io/triplet-loss
- http://bamos.github.io/2016/01/19/openface-0.2.0/

NPTEL ONLINE CERTIFICATION COURSES

Thank you