NUS National University of Singapore

NUS AI SUMMER EXPERIENCE 2019

Face Recognition

By Tan Sing Kuang
陈星旷

NATIONAL UNIVERSITY OF SINGAPORE
DEPARTMENT OF ISEM

Face Recognition

- Recap
  - Supervised Learning
  - Deep Neural Network
- What is face recognition?
- Face Recognition Pipeline
- Datasets
  - Performance Evaluation
- Deep Learning Solution

- Theory
  - Verification and Identification
  - Face Alignment
  - Gender and Age Recognition
  - Neural Network Structures
  - Standard Loss Functions and Regularizations
- Open Source Libraries
- Applications

- 3d Face Recognition
- Other Face Related Problems
- History
  - Traditional Face Recognition
    - Eigenface
    - LBPH
  - Traditional Face Detection
  - Traditional Face Datasets

Supervised Learning

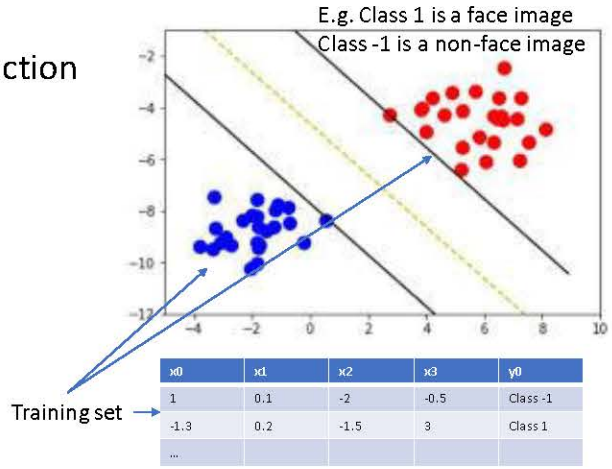
- Learn a function to map inputs to outputs given a set of input-output pairs
- E.g. learn the weights of a linear function

$$y = sign(w^T x)$$

$$\text{Where } sign(a) = \begin{cases} -1 & a < 0 \\ 1 & a \geq 0 \end{cases}$$

- By an optimization function

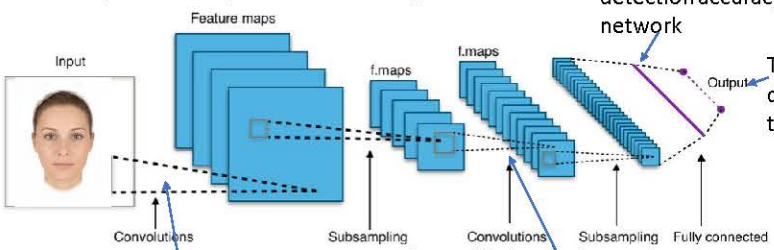
$$\text{Minimize } \|w\|$$
$$\text{subject to } y_i(w^T x - b) \geq 1$$



x0	x1	x2	x3	y0
1	0.1	-2	-0.5	Class -1
-1.3	0.2	-1.5	3	Class 1
...				

Deep Neural Network

- In computer vision, a convolutional neural network is usually used
- E.g. for face detection
  - Input is the face image pixel values
  - Output is a value between 0 and 1
  - Gives us the probability that the image is a face



Fully connected layers before the last layer is used to improve the detection accuracy of the network

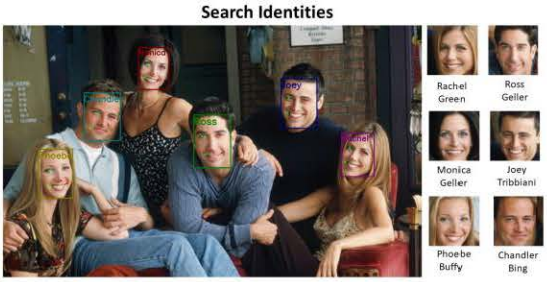
The last layer detects whether there is a face

The early layers detect the low level features such as edges

The middle layers detect the object parts such as eyes, nose and mouth

What is face recognition?

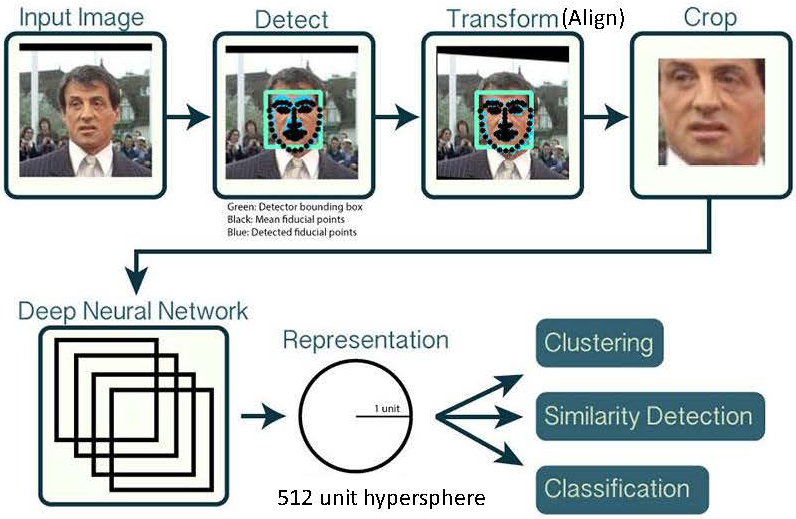
- Face recognition is to identify the person given an image of his/her face
  - Human can do face recognition easily
  - Face recognition algorithm has outperformed human
    - https://medium.com/the-physics-arxiv-blog/the-face-recognition-algorithm-that-finally-outperforms-humans-2c567adbf7fc
    - It can achieve 1 in a million accuracy on the MegaFace dataset
    - And the best thing is computer will never feel tired



Face Recognition Problem

- Face verification, given two faces, whether the two faces belong to the same person
- Face identification, given a probe face, determine whether the face is in the database
- Challenges
  - Face classes is an open set
    - Faces encountered during application are not found in the training set
  - Large number of face classes

Face Recognition Pipeline



https://www.pyimagesearch.com/2018/09/24/opencv-face-recognition/

Popular Face Datasets

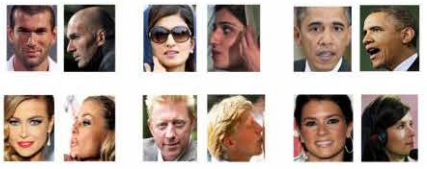
MS-Celeb-1M

- Training set
  - MS-Celeb-1M dataset
    - Celebrities' faces, high quality (suitable for training)
    - Large number of classes (~86000 face ids)
    - Select 1 million identities from the knowledge graph freebase, crawl the faces from the web, manually refine the labels
  - VGG2 dataset
    - large variations in age, pose, illumination, profession and ethnicity
    - ~9000 face ids

Vgg2

Popular Face Datasets

- Test set
  - Faces in the wild (LFW)
    - Unconstrained poses
  - CFP
    - Frontal and profile faces
  - AgeDB
    - Contain faces of different ages
  - MegaFace
    - Millions of faces, for testing the accuracy at false positive rate 10^{-6}
  - IJB-C
    - Occlusion, diverse occupations by avoiding "celebrity-only" data, variability of appearance and environment



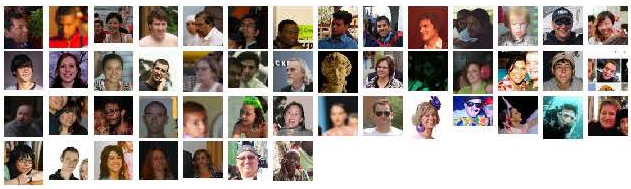
CFP

AgeDB

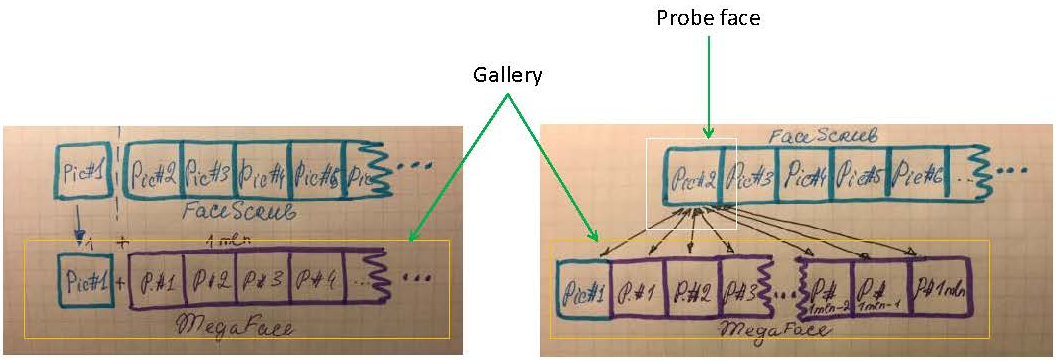
- MegaFace dataset

1 Million Photos
690,572 Unique Users
Collected from Flickr

- Facescrub dataset contains faces with face classes to be identified



- Megaface dataset contains faces to be added as distractors to the Facescrub dataset
- To make face identification difficult for Facescrub dataset



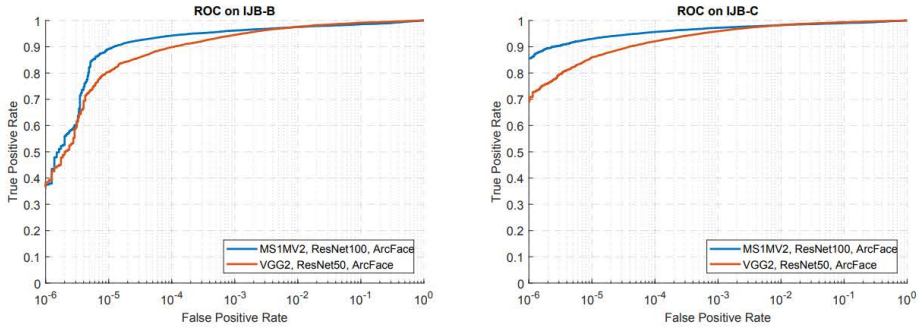
- Pic#1 from Facescrub dataset together with the faces from Megaface is added to the gallery

- Pic#2 from Facescrub is compared to all the faces in the gallery
- The process is repeated for other Pic's of the Facescrub
- An accuracy value is computed base on the identification accuracy of the Pic's against the gallery

https://medium.com/faceter/megaface-benchmark-what-does-it-show-part-i-809b41790de5

Performance Evaluation

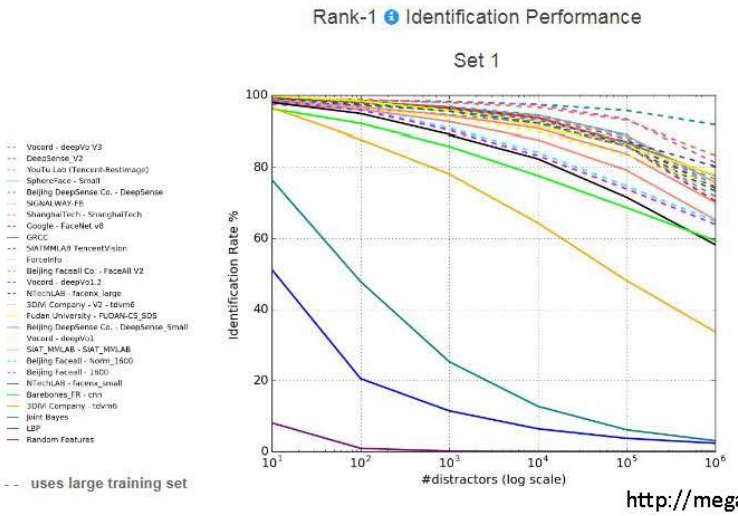
- In each test dataset, there is usually a list of
  - Positive pairs
    - Each pair of faces belong to the same person
  - Negative pairs
    - Each pair of faces belong to the different person
- False positive rate (FPR)
  - Number of wrongly identified negative pairs / total number of negative pairs
- True positive rate (FPR)
  - Number of correctly identified positive pairs / total number of positive pairs
- Accuracy
  - (Number of correctly identified positive pairs + Number of correctly identified negative pairs)/ (Total number of positive pairs + Total number of positive pairs)
- Performance results are usually given in
  - Accuracy at false positive rate of 1e-3, 1e-6, …
- The comparison of two faces will usually return a distance
  - The faces are the same if the distance < threshold
  - Else they are not the same
  - Use of different thresholds will lead to different false positive rates , true positive rates and accuracies



(a) ROC for IJB-B (b) ROC for IJB-C

Deng, J., Guo, J. and Zafeiriou, S., 2018. Arcface: Additive angular margin loss for deep face recognition. *arXiv preprint arXiv:1801.07698*.

Megaface Evaluation Results

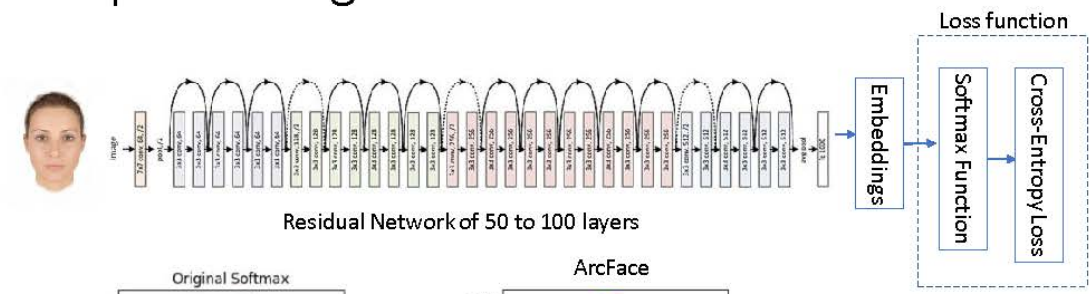


Rank-1 identification results on varying # of distractors

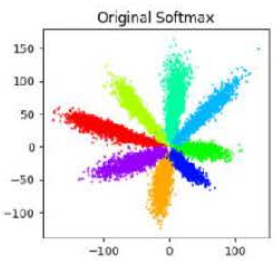
http://megaface.cs.washington.edu/results/facescrub.html

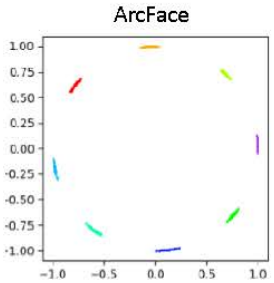
During inference, loss function is removed

Deep Learning Solution



Residual Network of 50 to 100 layers

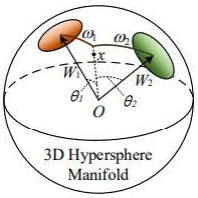




- There are many variations of softmax
- One of it is ArcFace
- The embeddings (face feature vectors) of the faces are L2 normalized (lies on a sphere)
- Two faces are the same person if their embeddings lie on the same cluster (L2 distance is small)

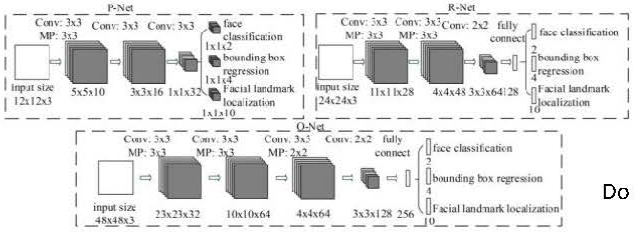
Face Verification and Identification using the Embeddings

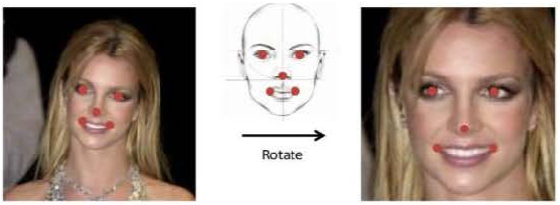
- For each class of faces in the database, find the average embedding c_1 of the class C
  - $c_1 = \frac{\sum_{x \in C} x}{\left\| \sum_{x \in C} x \right\|}$
- Face Verification
  - Face x is the same person from class C if $x^T c_1 > T$
    - Threshold is usually chosen as 0.5
- Face Identification
  - Face x is a person from class C if $\max_i x^T c_i$ and $x^T c_i > T$



Face Alignment using MTCNN

- Three-stage multi-task deep convolutional networks
  - Candidate windows are produced through a fast Proposal Network (P-Net)
  - Refine these candidates in the next stage through a Refinement Network (R-Net)
  - Output Network (O-Net) produces final bounding box and facial landmarks position



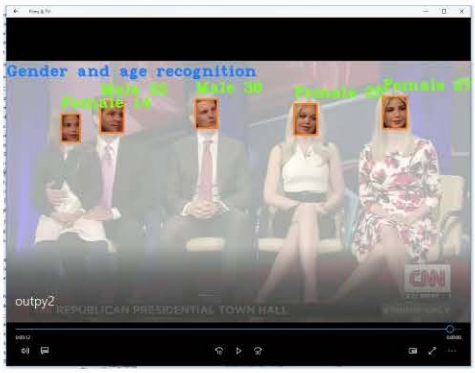


Do similarity transform to align the landmarks to the template face

Zhang, K., Zhang, Z., Li, Z. and Qiao, Y., 2016. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters*, *23*(10), pp.1499-1503.

Gender and Age Recognition

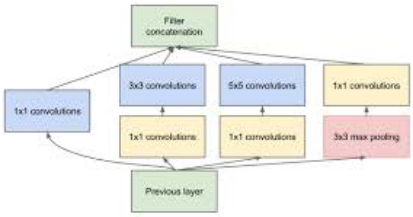
- Gender recognition is a classification problem
  - Put a logistic perceptron at the output layer to return the probability between male and female
  - It is much simple than face recognition
  - Can be trained on a smaller network of 1Mb in size
    - Run in 10ms and has an accuracy of 96%
- Age recognition is a regression problem
  - Put a perceptron with linear activation function at the output layer to return the age
  - Same as gender recognition, it is simple and fast



Neural Network Structure

- There are many predefined networks for learning face recognition
  - Inception network
  - Residual network
  - DenseNet
  - MobileNet
- Each network is an improvement over previous networks due to their drawbacks

- Inception network
  - Multiple filters of different sizes operate at the same level
    - These branches of convolutions help to detect more complex patterns
  - A deep network
    - To reduce the vanishing gradient problem, two auxiliary classifiers are introduced in the middle of the network
  - There are four versions which each version is an improvement over the previous version



- Residual Network
  - Uses skip connections to improve learning
    - Allows us to train a deeper network than a conventional convolutional network
    - Skip connections do not add additional parameters to the network
  - One of the most commonly used network in deep learning
  - Simpler architecture, compared to Inception network
  - Breaks from the traditional convolutional network
    - The lower layers detect local features, middle layers detect object parts and higher layers detect high level object information
    - Residual network layers do not have these specialization

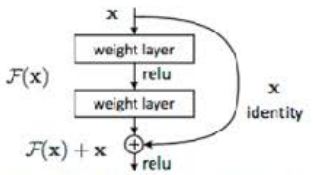
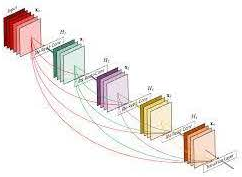
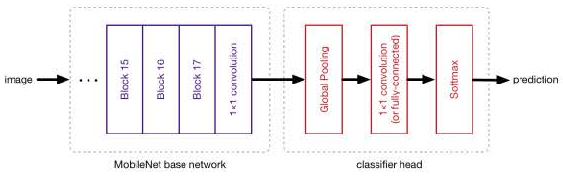


Figure 2. Residual learning: a building block.

- DenseNet
  - Generalized version of the Residual network
  - Uses more skip connections to improve learning
  - Theoretically improves more over the vanishing gradient problem more than a Residual network
  - DenseNet has shown higher performance on some datasets
    - But it is not as popular as Residual network

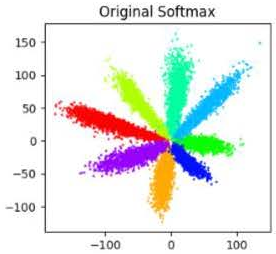


- MobileNet
  - Smaller because it uses depth-wise separable convolutions
  - Made some accuracy trade-off compared to the higher parameters model
- Need for smaller network
  - As it runs faster
  - Takes up less space
  - Can be deployed on a mobile phone



Standard Loss Functions and Regularizations

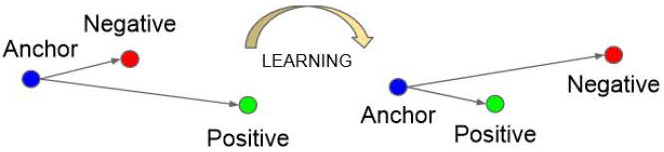
- Softmax
  - Does not enforce a margin between classes
  - Able to separate multiple classes efficiently
  - No need to do batch selection unlike triplet loss
  - Easy to implement
  - Use a lot of memory if there are large number of classes
  - Able to ignore outliers and small amount of noise in data



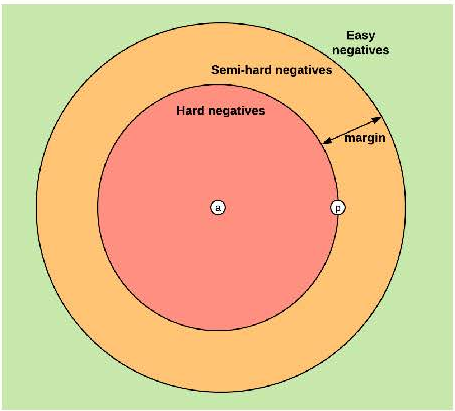
Softmax loss

$$= -\frac{1}{m}\sum_{i=1}^{m}\log\frac{e^{W_{y_i}^T x_i + b_{y_i}}}{\sum_{j=1}^{n} e^{W_j^T x_i + b_j}}$$

- Triple loss
  - An extension from contrastive loss
  - Able to handle variable number of classes
  - Sensitive to outlier faces

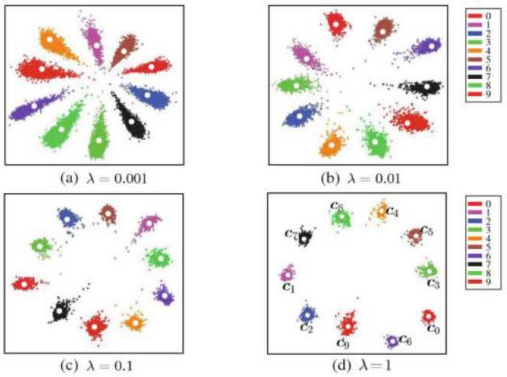


$$\mathcal{L} = max(d(a,p) - d(a,n) + margin, 0)$$



Need a balance selection of triples from each of the three classes of negatives

- Center Loss



Center Loss

$$\mathcal{L} = \mathcal{L}_S + \lambda \mathcal{L}_C$$

$$= -\sum_{i=1}^{m} \log \frac{e^{W_{y_i}^T \boldsymbol{x}_i + b_{y_i}}}{\sum_{j=1}^{n} e^{W_j^T \boldsymbol{x}_i + b_j}} + \frac{\lambda}{2} \sum_{i=1}^{m} \|\boldsymbol{x}_i - \boldsymbol{c}_{y_i}\|_2^2$$

- Additional regularization term that enforces the embeddings to be close to the centroids of their classes
- This creates greater margin between classes and leads to better accuracy

Softmax Variants

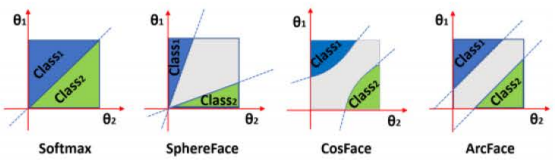
- These variants try to enforce large margin between classes
- Each has their own advantages and they can be combined
- The embeddings and weights are L2-normalized, to make learning easier

Method	m1	m2	m3
W&F Norm Softmax	1	0	0
SphereFace	1.5	0	0
CosineFace	1	0	0.35
ArcFace	1	0.5	0
Combined Margin	1.2	0.4	0
Combined Margin	1.1	0	0.35
Combined Margin	1	0.3	0.2
Combined Margin	0.9	0.4	0.15

Margin: cos(m_1*θ+m_2) - m_3

ArcFace Softmax

$$= -\frac{1}{m} \sum_{i=1}^{m} \log \frac{e^{s(\cos(\theta_{y_i} + m))}}{e^{s(\cos(\theta_{y_i} + m))} + \sum_{j=1, j \neq y_i}^{n} e^{s \cos \theta_j}}$$



- CosFace

$$L_{lmc} = \frac{1}{N}\sum_i -\log \frac{e^{s(\cos(\theta_{y_i,i})-m)}}{e^{s(\cos(\theta_{y_i,i})-m)} + \sum_{j\neq y_i} e^{s\cos(\theta_{j,i})}},$$

subject to

$$W = \frac{W^*}{\|W^*\|},$$

Normalized the softmax weights

$$x = \frac{x^*}{\|x^*\|},$$

Previous softmax variants only normalize the embedding weights

$$cos(\theta_j, i) = {W_j}^T x_i,$$

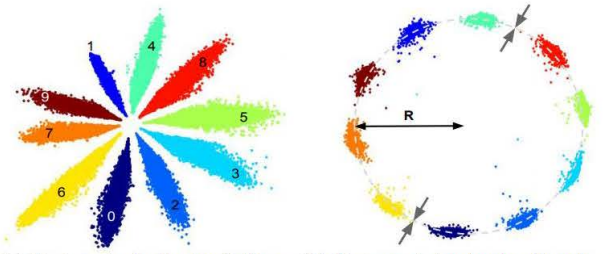
- Ring Loss
  - Extension of L2 normalization of embedding in some softmax variants
  - Use a soft L2 norm constraint on embedding
  - Can be used together with other large margin softmax

Softmax loss with L2 normalization

$$\min L_S(\mathcal{F}(x)) \;\; s.t. \;\; \|\mathcal{F}(x)\|_2 = R$$

Ring Loss

$$L_R = \frac{\lambda}{2m}\sum_{i=1}^{m}(\|\mathcal{F}(\mathbf{x}_i)\|_2 - R)^2$$



(a) Features trained using Softmax (b) Features trained using Ring loss

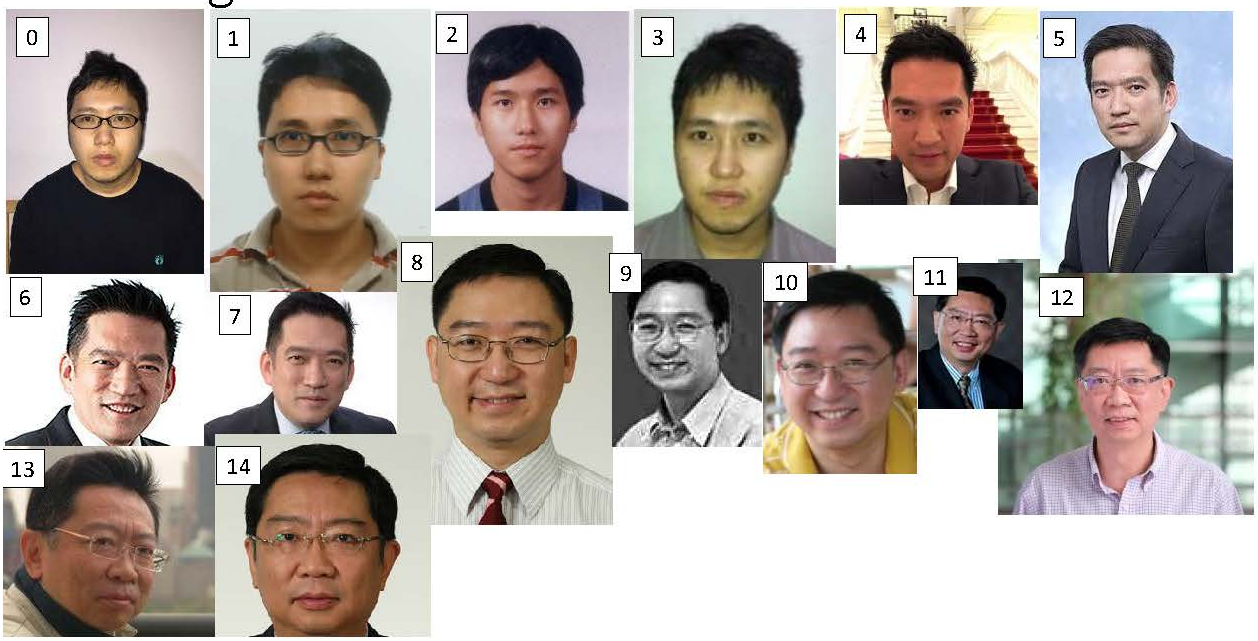
Result of ArcFace pre-trained model

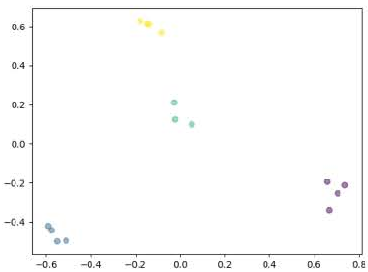
- https://github.com/deepinsight/insightface
- Trained on residual network for 18 epochs
- Align face using MTCNN into 112x112 image
- Do random flip
- Trained on MS Celeb 1M dataset (one of the largest open training dataset)
- Achieved 97.64% accuracy on MegaFace dataset (one of the largest open test dataset) at FPR=1e-3
- embedding size=512

Training and Inference

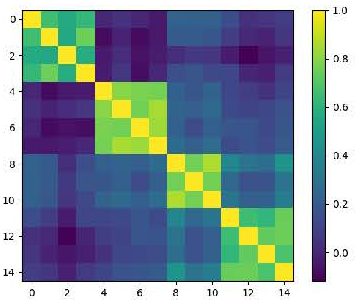
- Face recognition is fast enough to run on a CPU
  - Takes about ~2.5 seconds to extract the face embedding of a face image
  - Including face alignment and detection
- It takes about ~0.25 second to extract face embeddings for a batch of 128 faces on a GPU (including face alignment and detection)
- GPU is much faster
- Training involves alternate iterations of inference and backpropagation
- It takes ~4 days to train a Res-50 network for face recognition on 2 GPUs

Test images

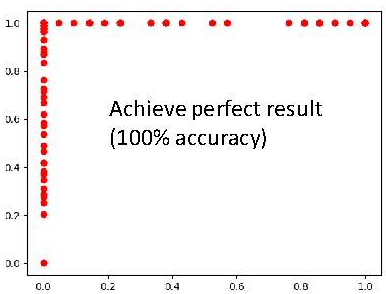




Scatter plot of the face feature vectors of id
1:purple
2:blue
3:green
4:yellow



Dot products between all face feature vectors



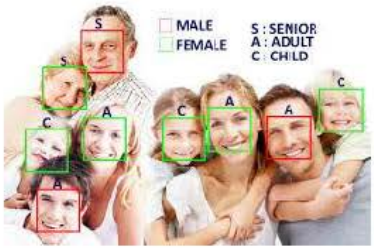
ROC curve

On these 15 images, it has perfect performance

Applications

- Face recognition for airport security
  - Trigger detection of wanted criminals such as terrorist from surveillance camera
  - Automated check-in counter for airport passengers
    - Faster and need less manpower

- Use face recognition for marketing and advertising
  - Face demographics
    - Age and gender
    - Ethnicity
  - Product recommendation based on mood detected from facial expression
  - Recognize customer and target offers based on their previous buying preference
  - Detect loyalty of customers and welcome them



- Use face recognition for login or access control
  - Unlock door, login to app
  - Recognize driver and detect sleepy driver
  - Iphone is using face recognition to unlock the phone using only the camera
  - Login to app without the need to key in username and password

- Use face to make payment
  - Alibaba created 'smile to pay' face recognition payment system at KFC in China
    - Does not require a smartphone, customer need to signed up an Alipay app
    - 3D camera located at the payment kiosk
    - Phone number verification for additional security

https://techcrunch.com/2017/09/03/alibaba-debuts-smile-to-pay/

- Face recognition in healthcare
  - Diagnose genetic disease
    - Based on the image intensity shape of the face
  - Interpret emotions
  - Determine the level of pain
    - To manage chronic pain and medical treatment
  - Quick access to medical records

- Identify genetic disorders
https://www.newscientist.com/article/2189683-ai-can-identify-rare-genetic-disorders-by-the-shape-of-someones-face/

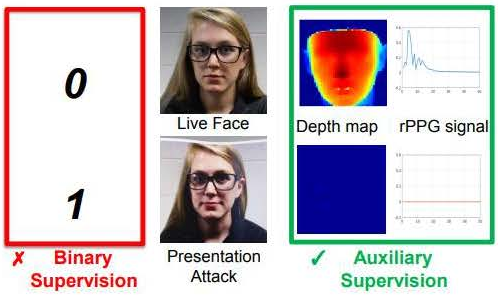
- Identify people from surveillance camera
  - There are many new cameras installed
    - But there are not enough people to monitor the surveillance cameras
  - Can be deployed to monitor event security
  - Modern face recognition algorithm are accurate enough to detect faces with varying poses and lightings
  - Can also be used to identify drivers in the traffic
  - Identify jaywalkers
  - Re-identify people who leave and return to the camera

3d Facial Recognition

- Scan face with 3d scanner
- The pipeline for 3d face recognition is the same as 2d (face alignment, feature vector extraction, feature vector comparison)
- E.g iPhone
  - Project 30,000 invisible infrared dots
  - Use infrared camera to get 3d model of the face

Other Face Related Problems

- Anti-face spoofing
  - From the face video
    - Use estimated depth map of the face
    - And rPPG signal (heart beats from the face)
    - To detect whether the face is a real or fake face



Liu, Y., Jourabloo, A. and Liu, X., 2018, March. Learning deep models for face anti-spoofing: Binary or auxiliary supervision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 389-398).

- Face superresolution
  - Learning a convolutional network to map a low resolution face to a high resolution face
  - Find the high resolution face of a low resolution image from surveillance camera

Target Input (Bicubic) VDSR URDGN SRResNet FSRNet (Ours) FSRGAN (Ours)

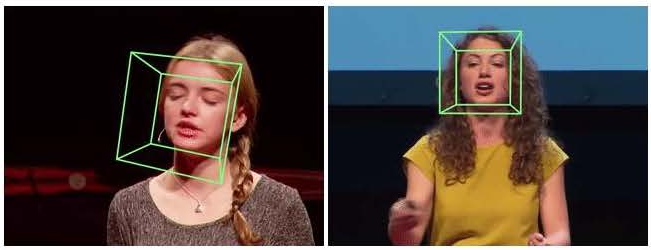
Chen, Y., Tai, Y., Liu, X., Shen, C. and Yang, J., 2018. FSRNet: End-to-End Learning Face Super-Resolution with Facial Priors. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2492-2501).

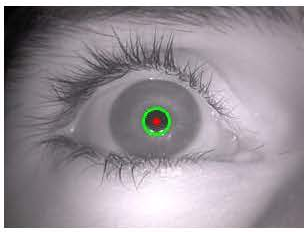
- Facial Expression detection
  - Surprise, happy, sad, angry, neutral, …
- Micro-expression detection
  - Use subtle movements in faces to detect face expressions
  - Difficult even for human

Angry Disappointed Laugh Neutral

https://insights.sei.cmu.edu/sei_blog/2018/01/revealing-true-emotions-through-micro-expressions-a-machine-learning-approach.html

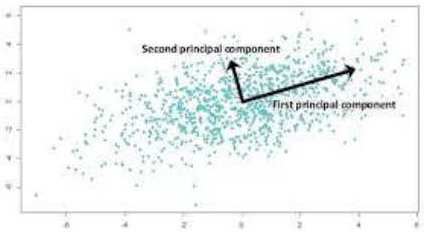
- Head pose estimation
  - Determine the orientation of the face
- Eyes tracking
  - Find out where the person is looking

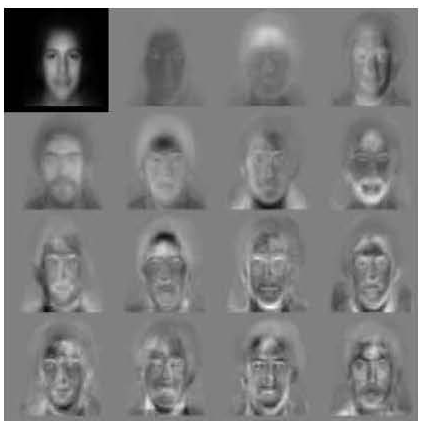




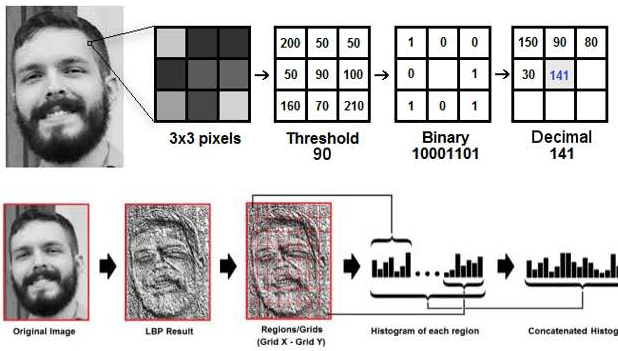
Traditional Method

- PCA is applied to an aligned face image (using eyes and mouths)
  - For dimension reduction
  - Remove some variations in lighting and poses
- A probe face image is projected into the eigenface vectors to get the feature vector
- Face recognition is simply done by comparing the L2 distance of two face feature vectors





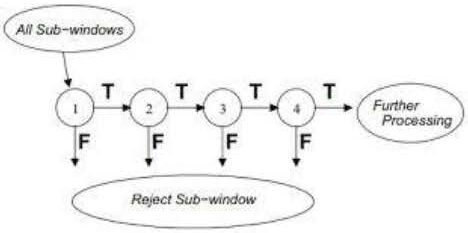
- LBPH can also be used as face recognition features
  - Other features such as SIFT and SURF can also be used together with LBPH

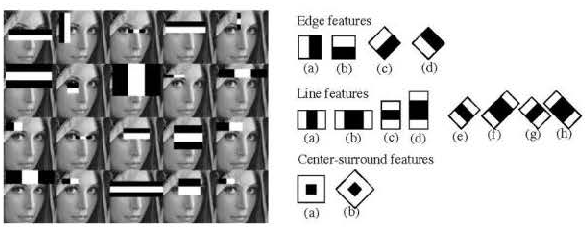


https://towardsdatascience.com/face-recognition-how-lbph-works-90ec258c3d6b

Traditional Face Detection Method

- Face detection using Haar-like features
  - https://towardsdatascience.com/face-detection-for-beginners-e58e8f21aad9
  - Face detection is a rare event detection problem
  - Use sliding window detection
  - Learn an cascade classifier by searching for the most discriminative features
    - Able to exit the classifier early if the early stage of the classifier rejects the region as face
  - Use integral image to speed up feature computation





Traditional Face Datasets

- Traditional face datasets have near-frontal faces with limited lighting variations
- Recent datasets have wider variations of lighting and poses

ORL face database has only 40 subjects

Vgg2 dataset

Lab exercises

- Face detection
  - https://face-recognition.readthedocs.io/en/latest/face_recognition.html
- Face recognition
  - https://www.analyticsvidhya.com/blog/2018/08/a-simple-introduction-to-facial-recognition-with-python-codes/
- Face identification
- Additional materials
  - Enable GPU: https://medium.com/deep-learning-turkey/google-colab-free-gpu-tutorial-e113627b9f5d

Further exercises

- Using LFW dataset
  - Visualize clusters of the face embeddings
  - Visualize PCA on the face embeddings
  - Visualize manifold of the face embeddings
  - Do classification on the face embeddings
    - Using LDA, SVM, Boosting

- How to load LFW dataset
  - https://jakevdp.github.io/PythonDataScienceHandbook/05.10-manifold-learning.html

Tan Sing Kuang 陈星旷

isetsk@nus.edu.sg

TanSingKuang

NUS National University of Singapore

NUS AI SUMMER EXPERIENCE 2019

Thank You

Face Recognition

NATIONAL UNIVERSITY OF SINGAPORE
DEPARTMENT OF ISEM