

Introduction to Face Recognition

Parkin Alexander
VisionLabs Researcher

Face Recognition tasks

Verification (1:1)



Identification (1:N)



Pipeline



Pipeline



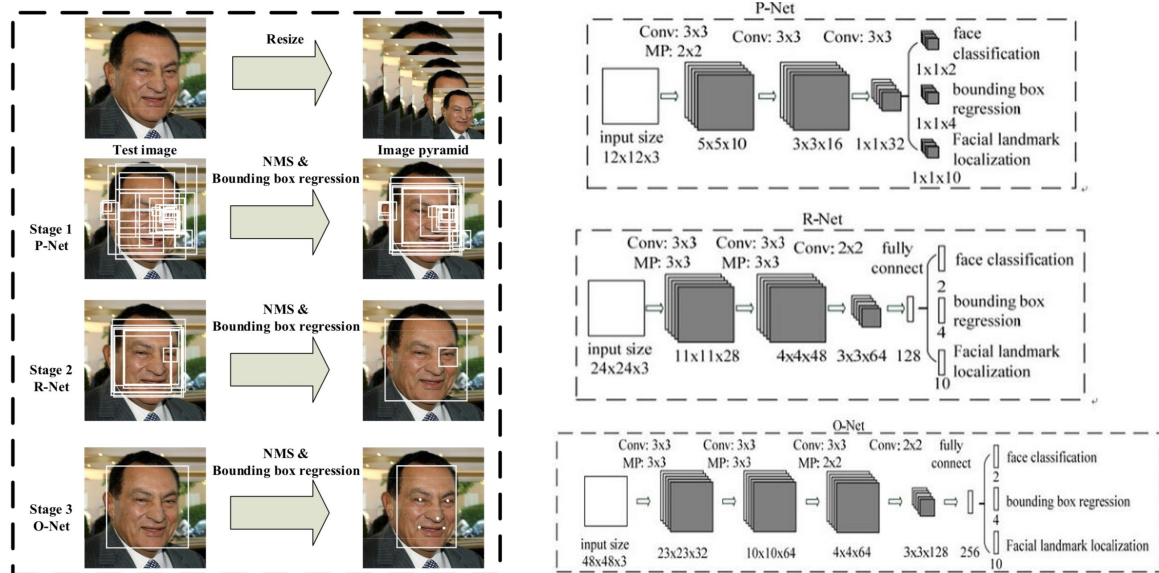
Face detectors

- Multi-task cascade NN (MTCNN)
- Single Shot Scale-invariant Face Detector (S^3FD)



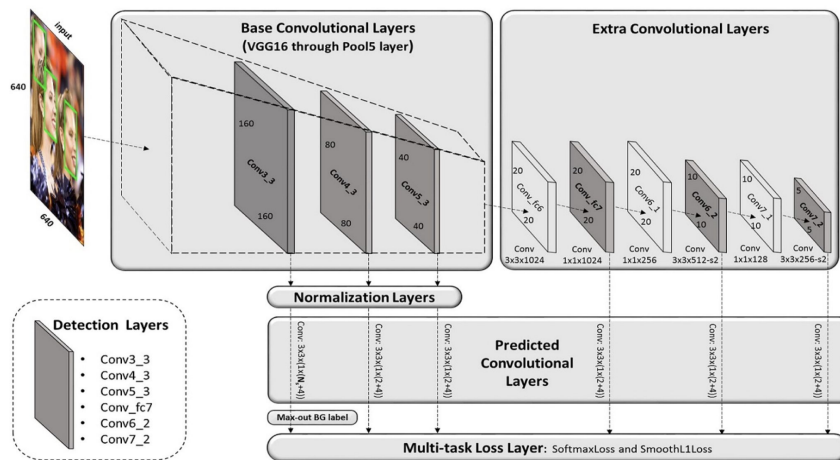
MTCNN

Joint Face Detection and Alignment using Multi-task Cascaded Convolutional Networks, K. Zhang et al., IEEE SPL 2016



S3FD

S3FD: Single Shot Scale-invariant Face Detector, S. Zhang et al., ICCV 2017



Detection Layer	Stride	Anchor	RF
conv3_3	4	16	48
conv4_3	8	32	108
conv5_3	16	64	228
conv_fc7	32	128	340
conv6_2	64	256	468
conv7_2	128	512	724

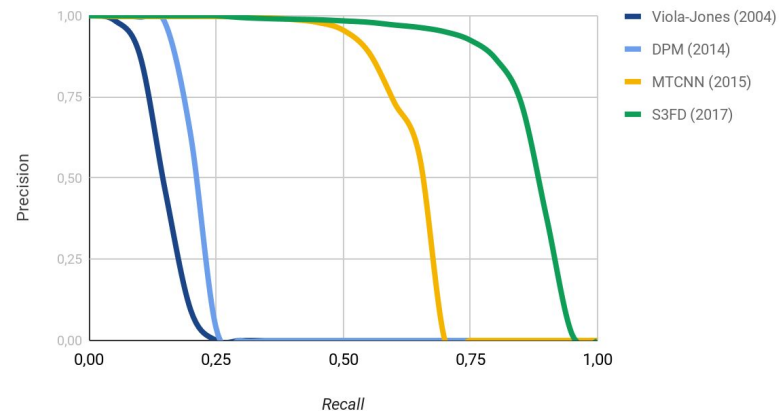
MTCNN vs S3FD

Image size: 640 x 640

Minimal face size: 16 pix

Detector type	CPU detection time, ms
MTCNN	170
S3FD	2500 (!)

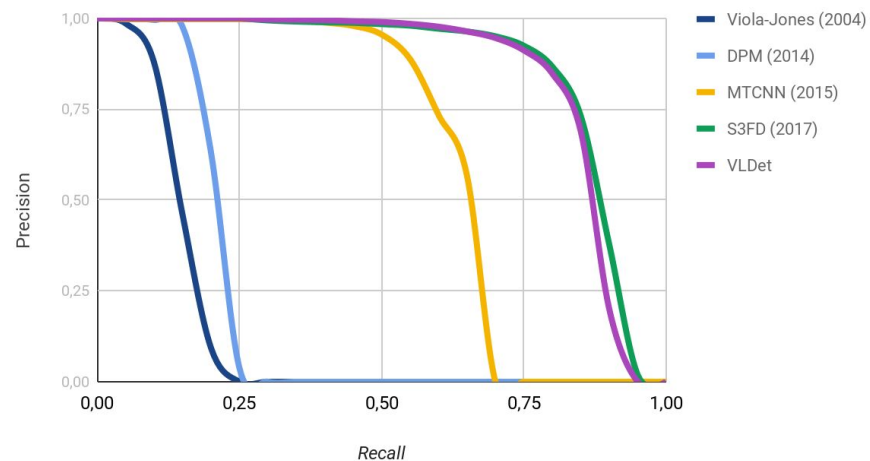
PR curve (WIDER FACE hard)



MTCNN vs S3FD vs VLDet

Detector type	CPU detection time, ms. 1 face	CPU detection time, ms. 40 faces
MTCNN	160	573
S3FD	2500	2500
VLDet	90	91

PR curve (WIDER FACE hard)



Largest selfie example (Viola-Jones)



Largest selfie example (VLDet)



Face Detection datasets

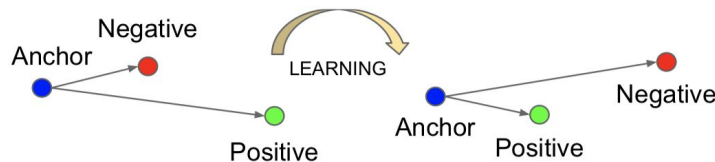
Dataset	#Faces	#Images/Videos
Face Detection Data Set and Benchmark (FDDB)	5171	2845
UTKFace	20K	20K
CelebA	200K	200K
WIDER FACE	394K	32K

Pipeline



Triplet Loss

FaceNet: A Unified Embedding for Face Recognition and Clustering, F. Shroff et al., CVPR 2015

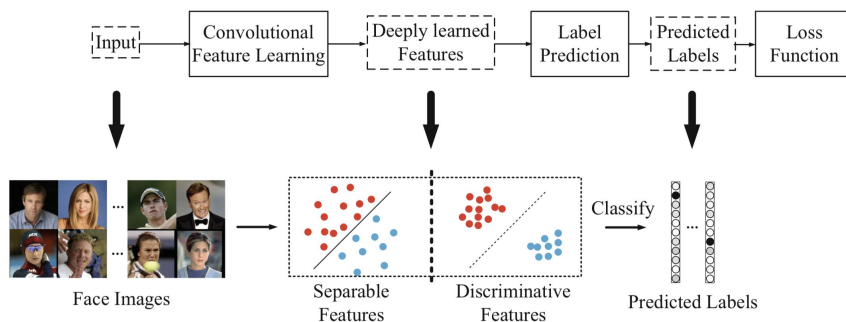


$$\|f(x_i^a) - f(x_i^p)\|_2^2 + \alpha < \|f(x_i^a) - f(x_i^n)\|_2^2$$

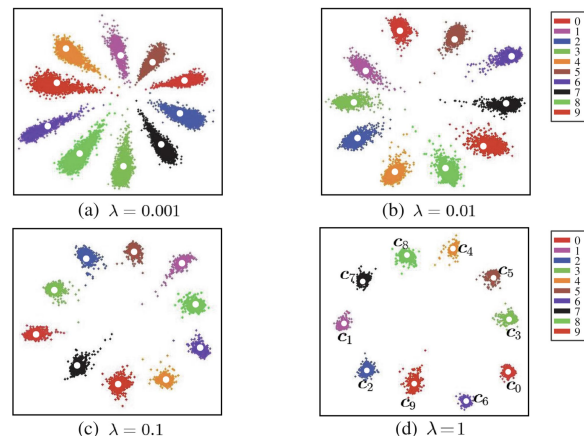
$$\sum_i^N \left[\|f(x_i^a) - f(x_i^p)\|_2^2 - \|f(x_i^a) - f(x_i^n)\|_2^2 + \alpha \right]_+$$

Center Loss

A Discriminative Feature Learning Approach for Deep Face Recognition, Y. Wen et al., ECCV 2016



MNIST features



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

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

SphereFace

SphereFace: Deep Hypersphere Embedding for Face Recognition, W. Liu et al., CVPR 2017

$$p_1 = \frac{\exp(\mathbf{W}_1^T \mathbf{x} + b_1)}{\exp(\mathbf{W}_1^T \mathbf{x} + b_1) + \exp(\mathbf{W}_2^T \mathbf{x} + b_2)}$$

$$p_2 = \frac{\exp(\mathbf{W}_2^T \mathbf{x} + b_2)}{\exp(\mathbf{W}_1^T \mathbf{x} + b_1) + \exp(\mathbf{W}_2^T \mathbf{x} + b_2)}$$

$$\mathbf{W}_i^T \mathbf{x} + b_i \text{ as } \|\mathbf{W}_i^T\| \|\mathbf{x}\| \cos(\theta_i) + b_i$$

$$\|\mathbf{W}_i\| = 1, b_i = 0 \quad p_1 = \|\mathbf{x}\| \cos(\theta_1) \quad p_2 = \|\mathbf{x}\| \cos(\theta_2)$$

SphereFace

SphereFace: Deep Hypersphere Embedding for Face Recognition, W. Liu et al., CVPR 2017

$$p_1 = \frac{\exp(\mathbf{W}_1^T \mathbf{x} + b_1)}{\exp(\mathbf{W}_1^T \mathbf{x} + b_1) + \exp(\mathbf{W}_2^T \mathbf{x} + b_2)}$$

$$p_2 = \frac{\exp(\mathbf{W}_2^T \mathbf{x} + b_2)}{\exp(\mathbf{W}_1^T \mathbf{x} + b_1) + \exp(\mathbf{W}_2^T \mathbf{x} + b_2)}$$

$$\mathbf{W}_i^T \mathbf{x} + b_i \text{ as } \|\mathbf{W}_i^T\| \|\mathbf{x}\| \cos(\theta_i) + b_i$$

$$\|\mathbf{W}_i\| = 1, b_i = 0 \quad p_1 = \|\mathbf{x}\| \cos(\theta_1) \quad p_2 = \|\mathbf{x}\| \cos(\theta_2)$$

$$L_1 = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{W_{y_i}^T x_i + b_{y_i}}}{\sum_{j=1}^n e^{W_j^T x_i + b_j}}$$

$$L_{\text{modified}} = \frac{1}{N} \sum_i -\log \left(\frac{e^{\|\mathbf{x}_i\| \cos(\theta_{y_i, i})}}{\sum_j e^{\|\mathbf{x}_i\| \cos(\theta_{j, i})}} \right)$$

SphereFace

SphereFace: Deep Hypersphere Embedding for Face Recognition, W. Liu et al., CVPR 2017

$$p_1 = \frac{\exp(\mathbf{W}_1^T \mathbf{x} + b_1)}{\exp(\mathbf{W}_1^T \mathbf{x} + b_1) + \exp(\mathbf{W}_2^T \mathbf{x} + b_2)}$$

$$p_2 = \frac{\exp(\mathbf{W}_2^T \mathbf{x} + b_2)}{\exp(\mathbf{W}_1^T \mathbf{x} + b_1) + \exp(\mathbf{W}_2^T \mathbf{x} + b_2)}$$

$$\mathbf{W}_i^T \mathbf{x} + b_i \text{ as } \|\mathbf{W}_i^T\| \|\mathbf{x}\| \cos(\theta_i) + b_i$$

$$\|\mathbf{W}_i\| = 1, b_i = 0 \quad p_1 = \|\mathbf{x}\| \cos(\theta_1) \quad p_2 = \|\mathbf{x}\| \cos(\theta_2)$$

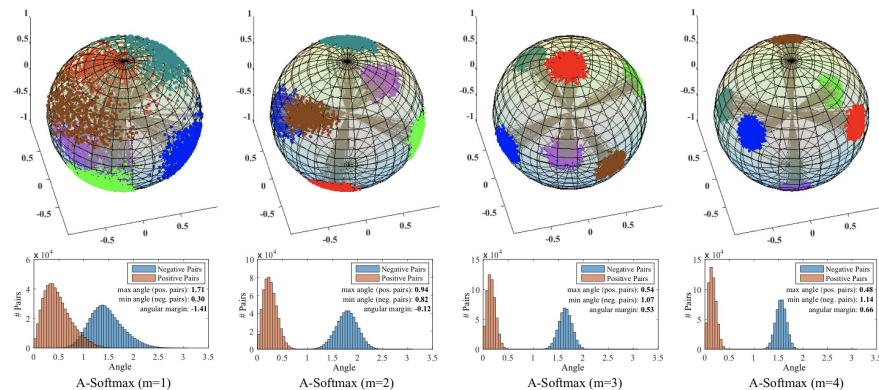
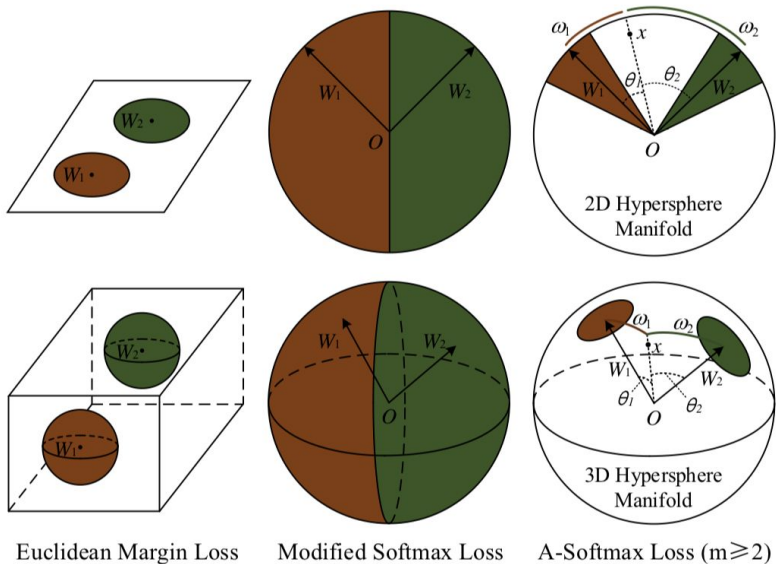
$$L_1 = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{W_{y_i}^T x_i + b_{y_i}}}{\sum_{j=1}^n e^{W_j^T x_i + b_j}}$$

$$L_{\text{modified}} = \frac{1}{N} \sum_i -\log \left(\frac{e^{\|\mathbf{x}_i\| \cos(\theta_{y_i, i})}}{\sum_j e^{\|\mathbf{x}_i\| \cos(\theta_{j, i})}} \right)$$

$$\begin{aligned} \cos(\theta_1) &> \cos(\theta_2) \\ \cos(m\theta_1) &> \cos(\theta_2) \end{aligned}$$

$$L_{\text{ang}} = \frac{1}{N} \sum_i -\log \left(\frac{e^{\|\mathbf{x}_i\| \cos(m\theta_{y_i, i})}}{e^{\|\mathbf{x}_i\| \cos(m\theta_{y_i, i})} + \sum_{j \neq y_i} e^{\|\mathbf{x}_i\| \cos(\theta_{j, i})}} \right)$$

SphereFace



$$\frac{1}{N} \sum_i -\log \left(\frac{e^{\|x_i\| \cos(m\theta_{y_i,i})}}{e^{\|x_i\| \cos(m\theta_{y_i,i})} + \sum_{j \neq y_i} e^{\|x_i\| \cos(\theta_{j,i})}} \right)$$

SphereFace vs CosFace vs ArcFace

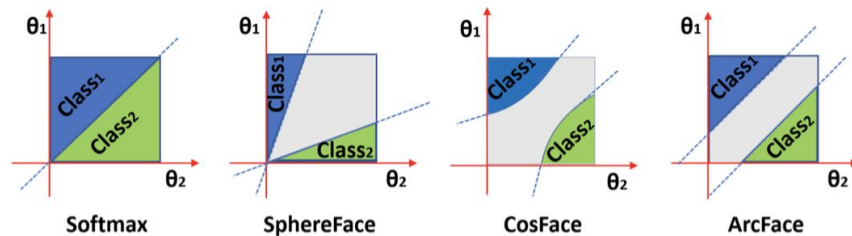
CosFace: Large Margin Cosine Loss for Deep Face Recognition, H. Wang et al., CVPR 2018

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

ArcFace: Additive Angular Margin Loss for Deep Face Recognition, J. Deng and J. Guo et al., arXiv:1801.07698v2

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

SphereFace vs CosFace vs ArcFace



SphereFace

$$\frac{1}{N} \sum_i -\log \left(\frac{e^{\|\mathbf{x}_i\| \cos(m\theta_{y_i,i})}}{e^{\|\mathbf{x}_i\| \cos(m\theta_{y_i,i})} + \sum_{i \neq y_i} e^{\|\mathbf{x}_i\| \cos(\theta_{j,i})}} \right)$$

CosFace

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

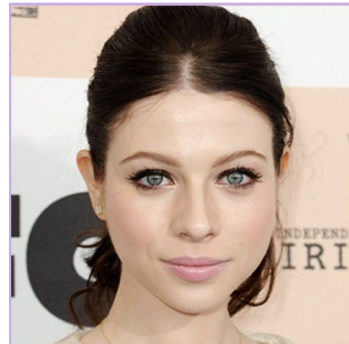
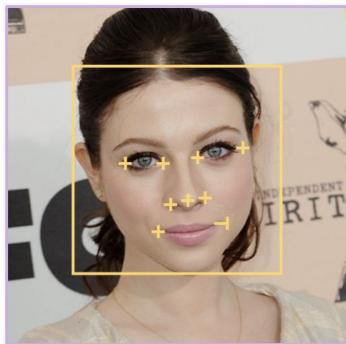
ArcFace

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

Face Datasets

Dataset	#Identity	#Images/Videos
YouTube Faces (YTF)	1595	3425
Labeled Faces in the Wild (LFW)	5,749	13K
MegaFace	672K	4.7M
CASIA WebFace (Casia)	10K	453K
VGGFace2	9.1K	3.3M
MS-Celeb-1M	85K	8.5M

MTCNN
S³FD



state-of-the-art
CNN



1 x n

Parkin Alexander
Researcher

a.parkin@visionlabs.ru

 ODS: @parkin_a