Face Recognition, Head Pose Estimation, Age Estimation, Facial Expression Analysis, VGG 16 and SSR-Net, Computer Vision and Biometrics

CS 585

March 7 and 19, 2024

Margrit Betke

Department of Computer Science

Boston University

LOOKING AT PEOPLE

Image source: Her, 2013

LOOKING AT PEOPLE: PERSON LOCALIZATION

Image source: Her, 2013

LOOKING AT PEOPLE: HUMAN POSE DETECTION

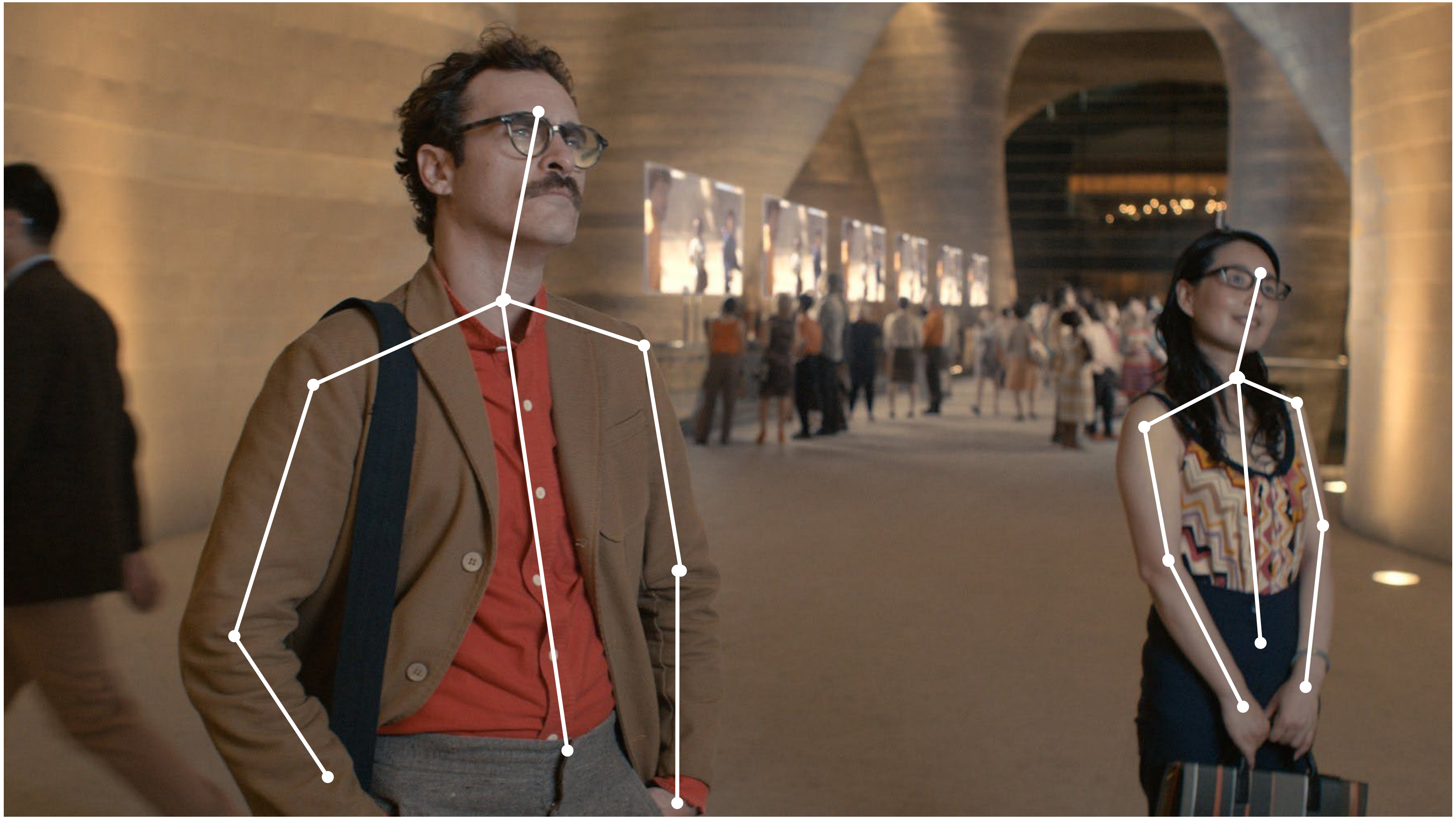


Image source: Her, 2013

LOOKING AT PEOPLE: FACE DETECTION

Image source: Her, 2013

LOOKING AT PEOPLE: FACE RECOGNITION

Image source: Her, 2013

LOOKING AT PEOPLE: FACIAL LANDMARK DETECTION

Image source: Her, 2013

LOOKING AT PEOPLE: FACIAL EXPRESSION RECOGNITION

Image source: Her, 2013

LOOKING AT PEOPLE: FACE RECOGNITION

Image source: Her, 2013

Recognizing Faces

Two Tasks:

- Face Verification

- Face Identification

Face Verification?

Are these two images showing the same person?

=

Yes

Query Image   Reference Image

“One-to-one similarity”

Important for Access Control and Re-identification

Face Identification?

What is the ID or name of this person?

Query Image

= "Margrit Betke"

"One-to-many similarity"

Important for Watch-list Surveillance or Forensic Search

Face Recognition

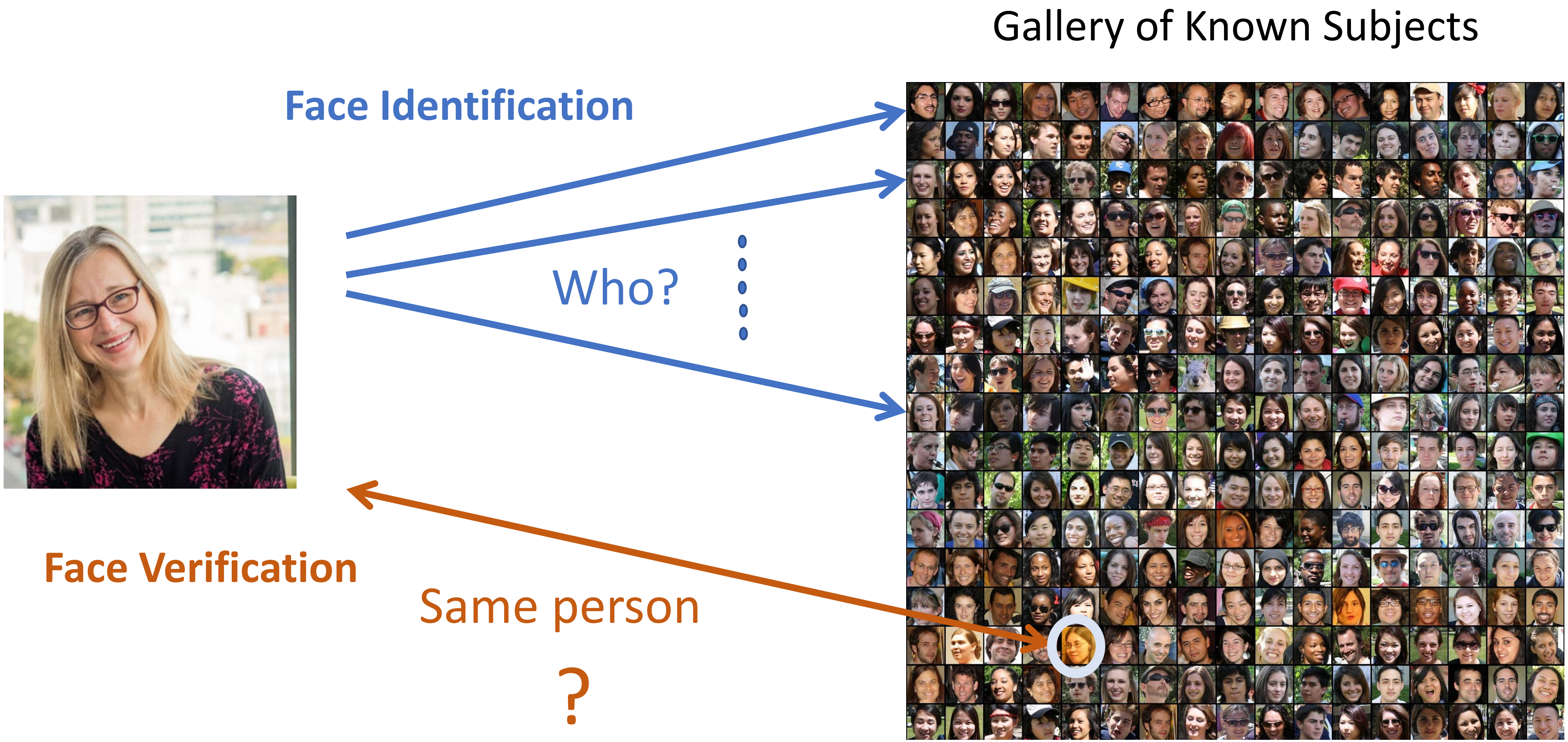


Image credit: Q. Orozco-Ramirez

Face Recognition

Gallery of Known Subjects

Face Identification

one-to-many
similarity

Face Verification

one-to-one
similarity

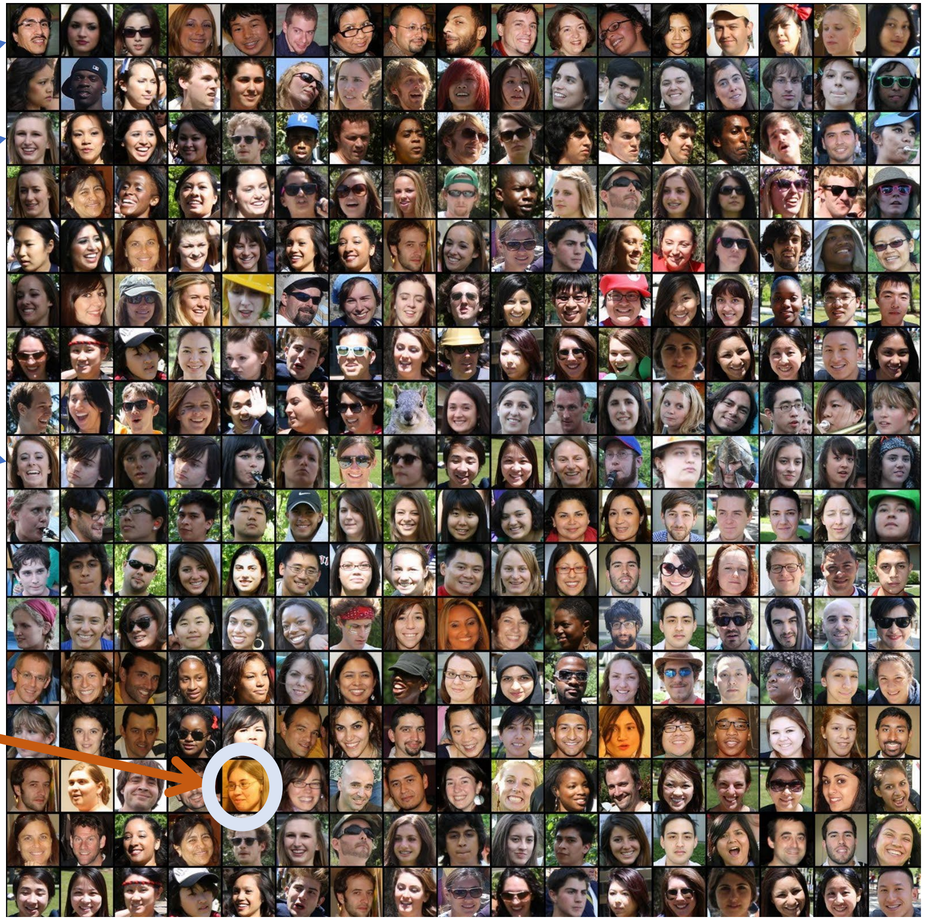


Image credit: Q. Orozco-Ramirez

Face Recognition

Gallery of Known Subjects

Image credit: Q. Orozco-Ramirez

Here:
One picture per person

Better:
Multiple pictures per person

Face Recognition

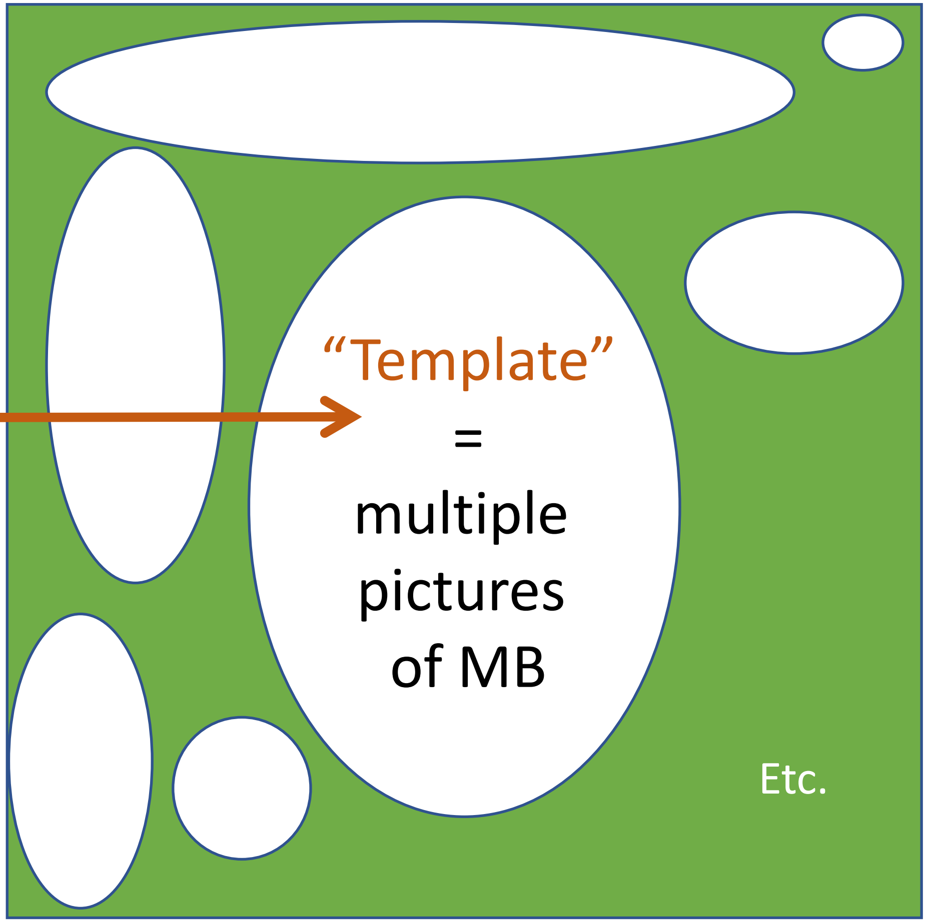
Gallery of Known Subjects

Face Recognition

Gallery of Known Subjects

Probe or **Query** of Unknown Subject

one-to-one similarity



How does Face Recognition Technology Work?

Pre-2012 Revolution in Computer Vision:

MIT's "Eigenfaces"

Now:

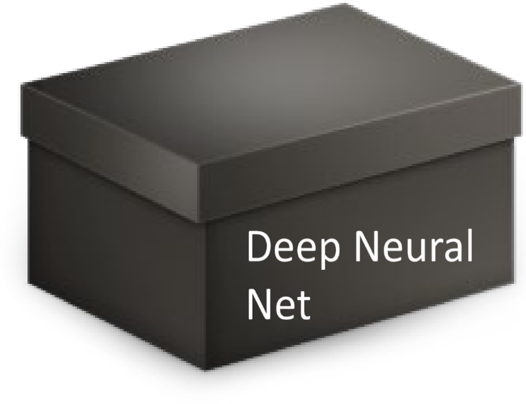
Deep Neural Networks

Except: Pre-processing still uses traditional techniques

Deep Neural Networks used for Face Recognition

1. Network architecture
2. Training
3. Testing = “use mode”

Network Architecture



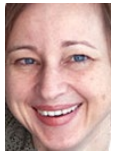
Training a Face Recognition Network

Training image for “Margrit Betke”

Crop

Rotate

Align



Input to black box
neural net = normalized
image for “Margrit Betke”

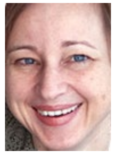
Training a Face Recognition Network

Training image for "Margrit Betke"

Crop

Rotate

Align



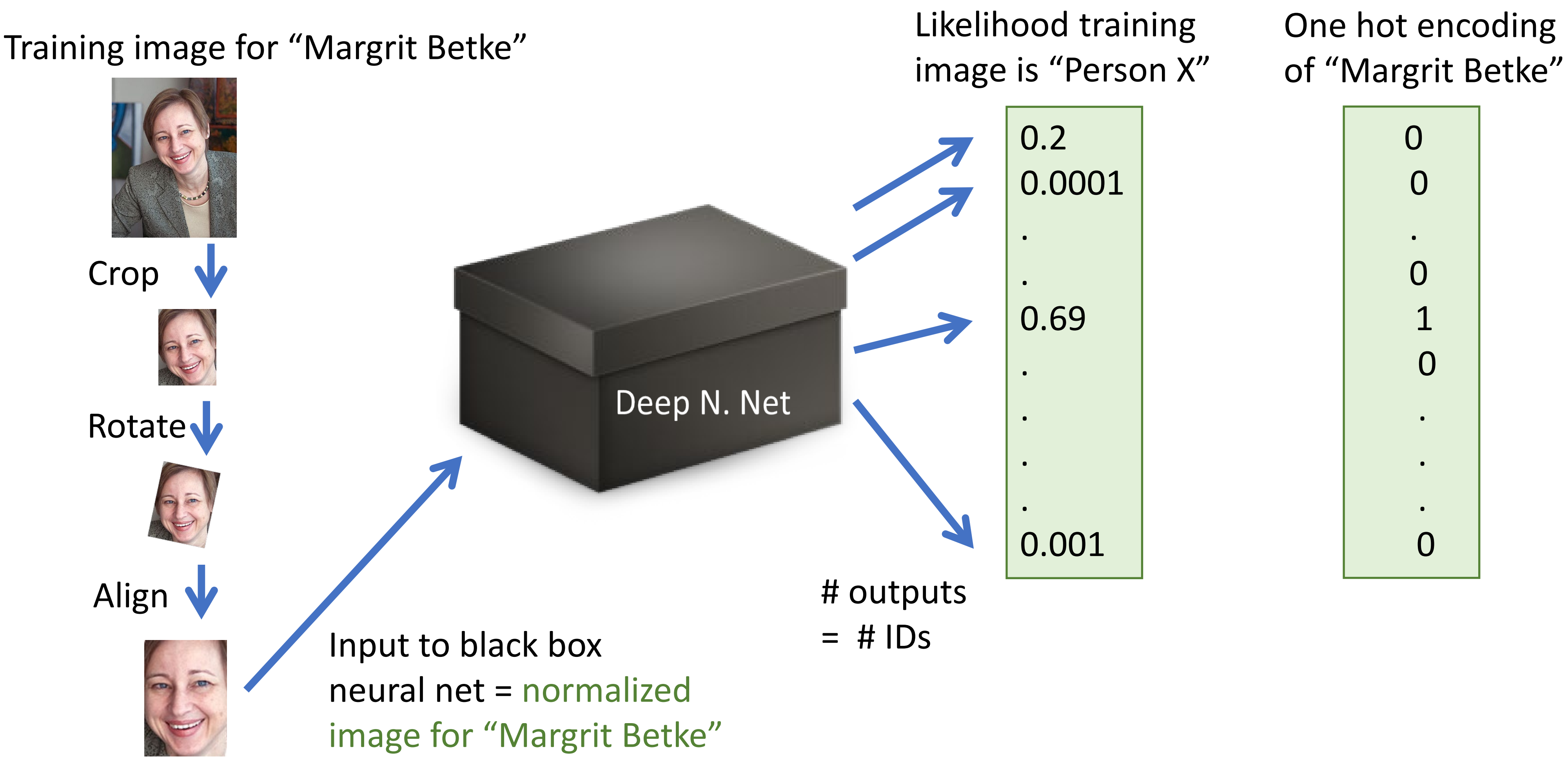
Input to black box neural net = normalized image for "Margrit Betke"

Likelihood training image is "Person X"

0.2
0.0001
.
.
0.69
.
.
.
0.001

outputs = # IDs

Training a Face Recognition Network



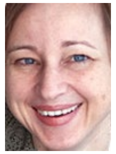
Training a Face Recognition Network

Training image for "Margrit Betke"

Crop

Rotate

Align



Input to black box neural net = normalized image for "Margrit Betke"

Likelihood training image is "Person X"

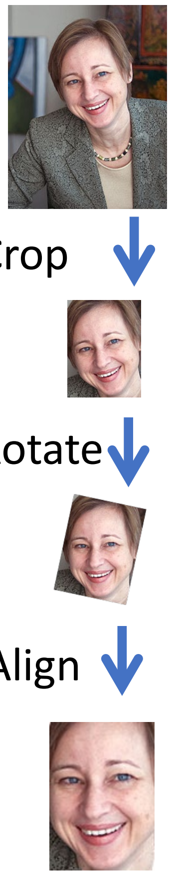
One hot encoding of "Margrit Betke"

$$\begin{bmatrix} 0.2 \\ 0.0001 \\ . \\ . \\ 0.69 \\ . \\ . \\ . \\ 0.001 \end{bmatrix} \text{ minus } \begin{bmatrix} 0 \\ 0 \\ . \\ . \\ 1 \\ . \\ . \\ . \\ 0 \end{bmatrix} = \text{error}$$

Sum squared error = Measure of training need

Training a Face Recognition Network

Training image for “Margrit Betke”



Crop

Rotate

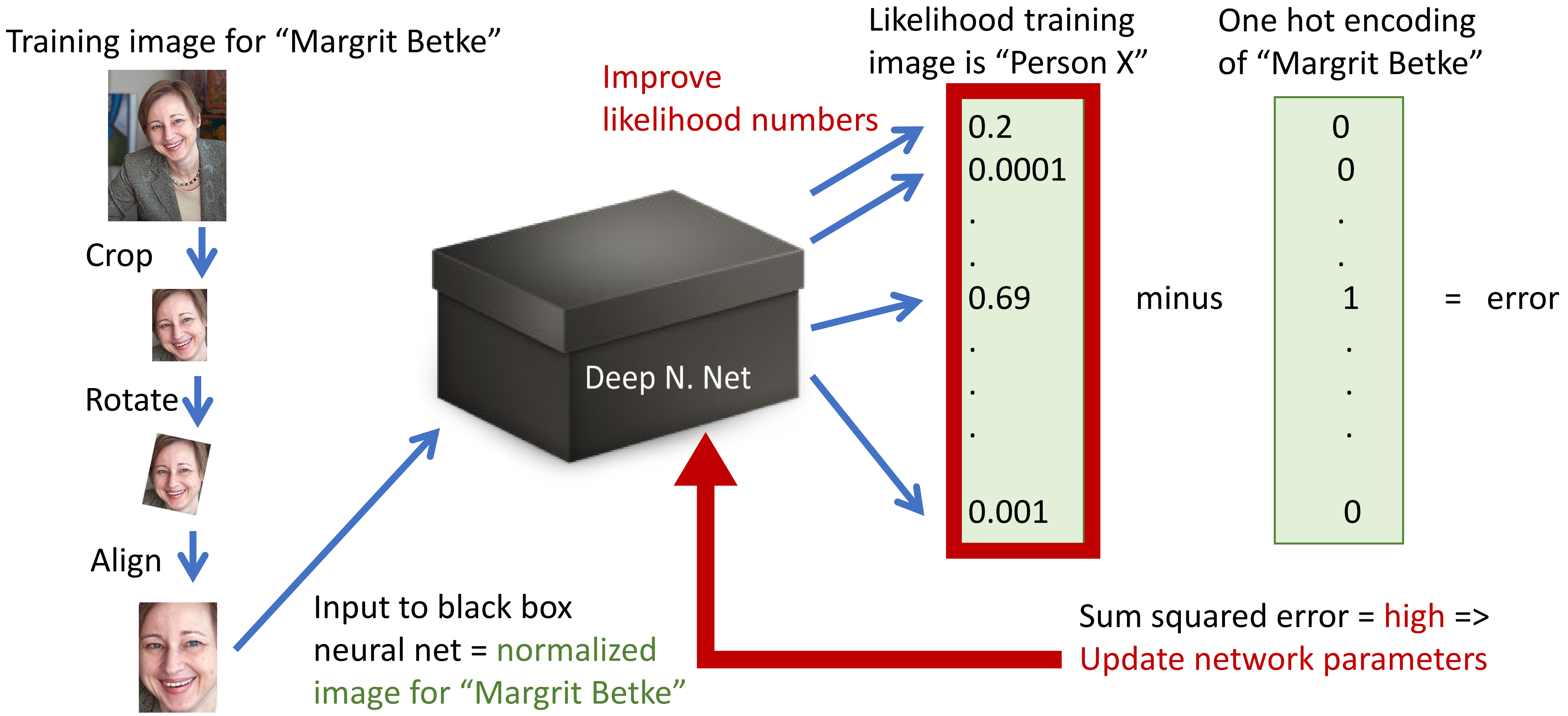
Align

Input to black box neural net = normalized image for “Margrit Betke”

Likelihood training image is “Person X”		One hot encoding of “Margrit Betke”	
0.2		0	
0.0001		0	
.		.	
.		.	
0.69	minus	1	= error
.		.	
.		.	
.		.	
0.001		0	

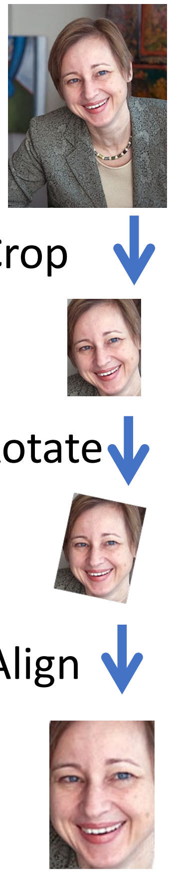
Sum squared error = high =>
Update network parameters

Training a Face Recognition Network



Training a Face Recognition Network

Training image for “Margrit Betke”



Crop

Rotate

Align

Input to black box neural net = normalized image for “Margrit Betke”

Improve likelihood numbers

Backpropagation

Algorithm

Likelihood training image is “Person X”		One hot encoding of “Margrit Betke”	
0.2		0	
0.0001		0	
.		.	
.		.	
0.69	minus	1	= error
.		.	
.		.	
.		.	
0.001		0	

Sum squared error = high =>
Update network parameters

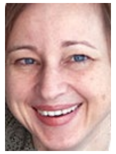
Training a Face Recognition Network

Training image for "Margrit Betke"

Crop

Rotate

Align



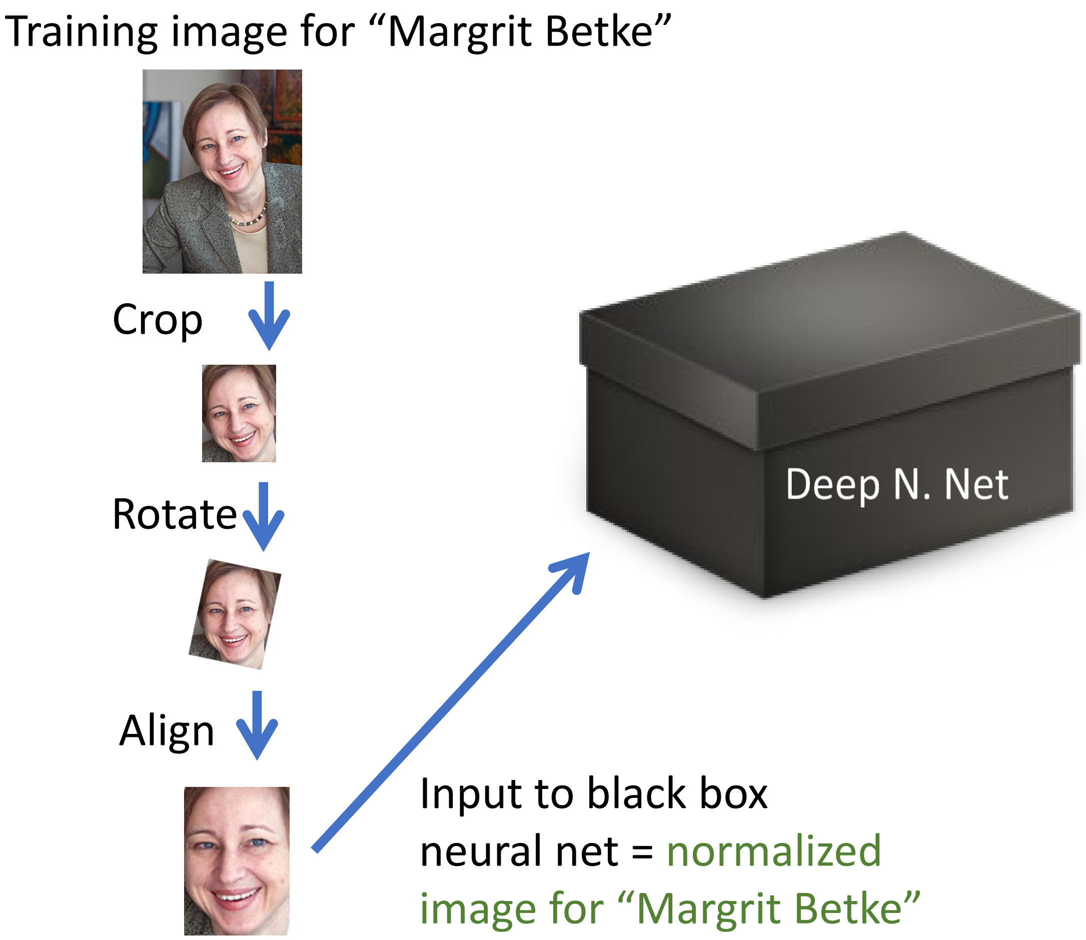
Input to black box neural net = normalized image for "Margrit Betke"

After numerous network updates:

Likelihood training image is "Person X"		One hot encoding of "Margrit Betke"		
0.0002		0		
0.0001		0		
.		.		
.		.		
0.99	minus	1	=	error
.		.		
.		.		
.		.		
0.0001		0		

Sum squared error = low =>
Done with training!

Trained network



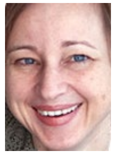
Let’s look at the trained network

Training image for “Margrit Betke”

Crop

Rotate

Align



Input to black box neural net = normalized image for “Margrit Betke”

Network Architecture: VGG-16

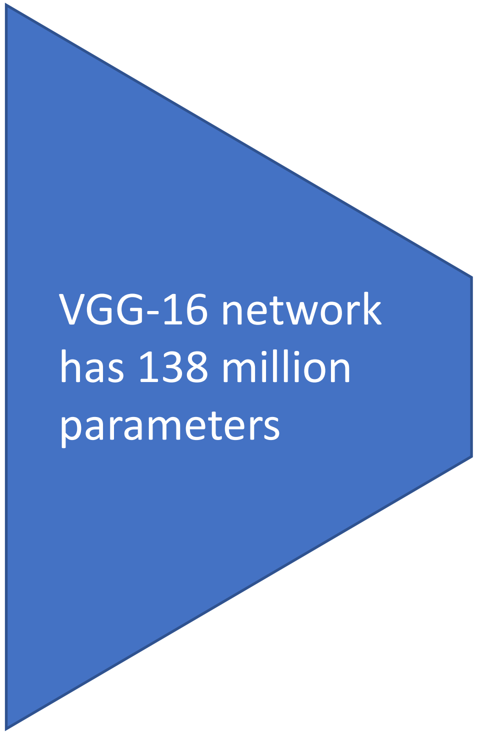
Input:
3 x
224 x 224
=
16,128
numbers

Output:
4,096
numbers

VGG = Visual Geometry Group, Oxford
[Karen Simoyan & Andrew Zisserman, 2014]

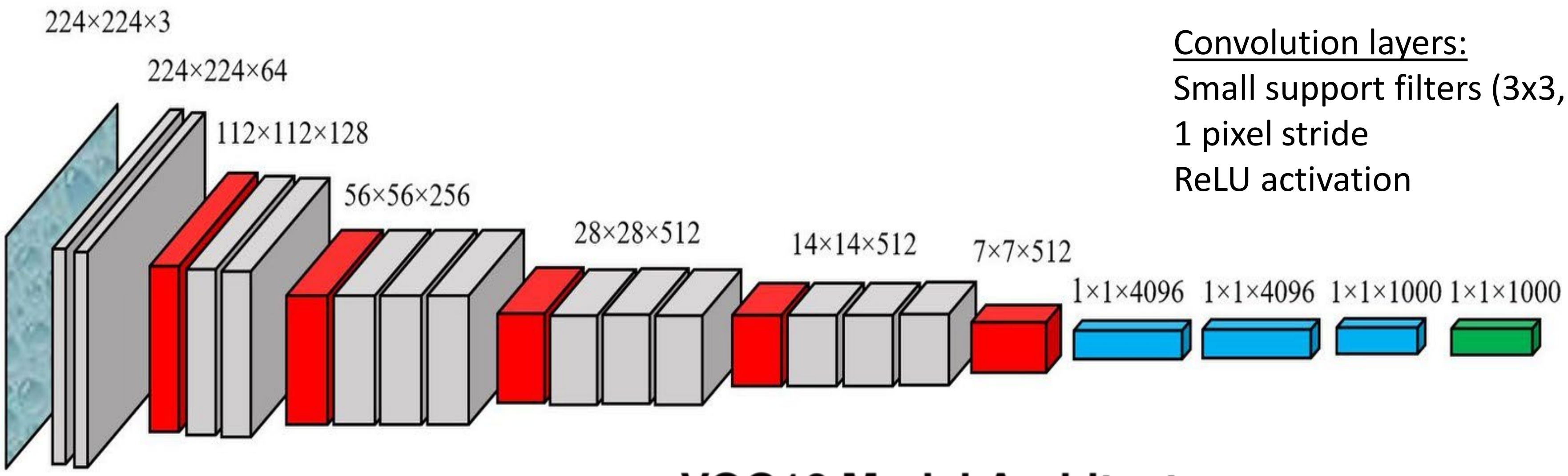
Network Architecture: VGG-16

Input:
3 x
224 x 224
=
16,128
numbers



Output:
4,096
numbers

Network Architecture of VGG-16: Two Visualizations



Convolution layers:
Small support filters (3x3, up/down, left/right, center)
1 pixel stride
ReLU activation

Image credit:
Hamidreza Habibollahi-Najaf-Abadi

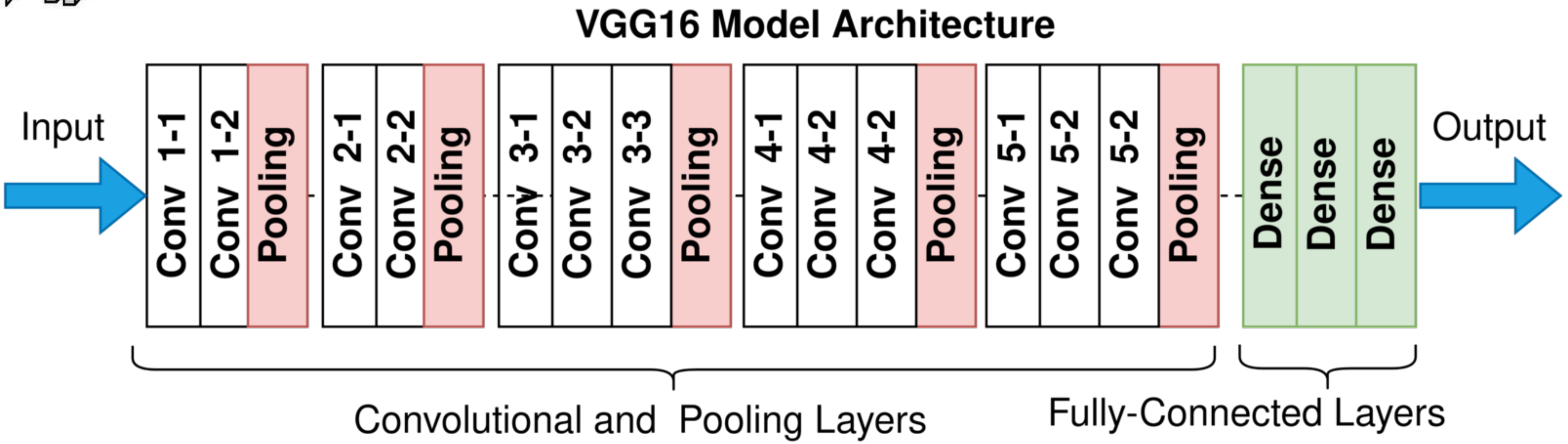


Image credit:
Gorla Praveen

Let’s look at how to use the trained network

Training image for “Margrit Betke”

Crop

Rotate

Align

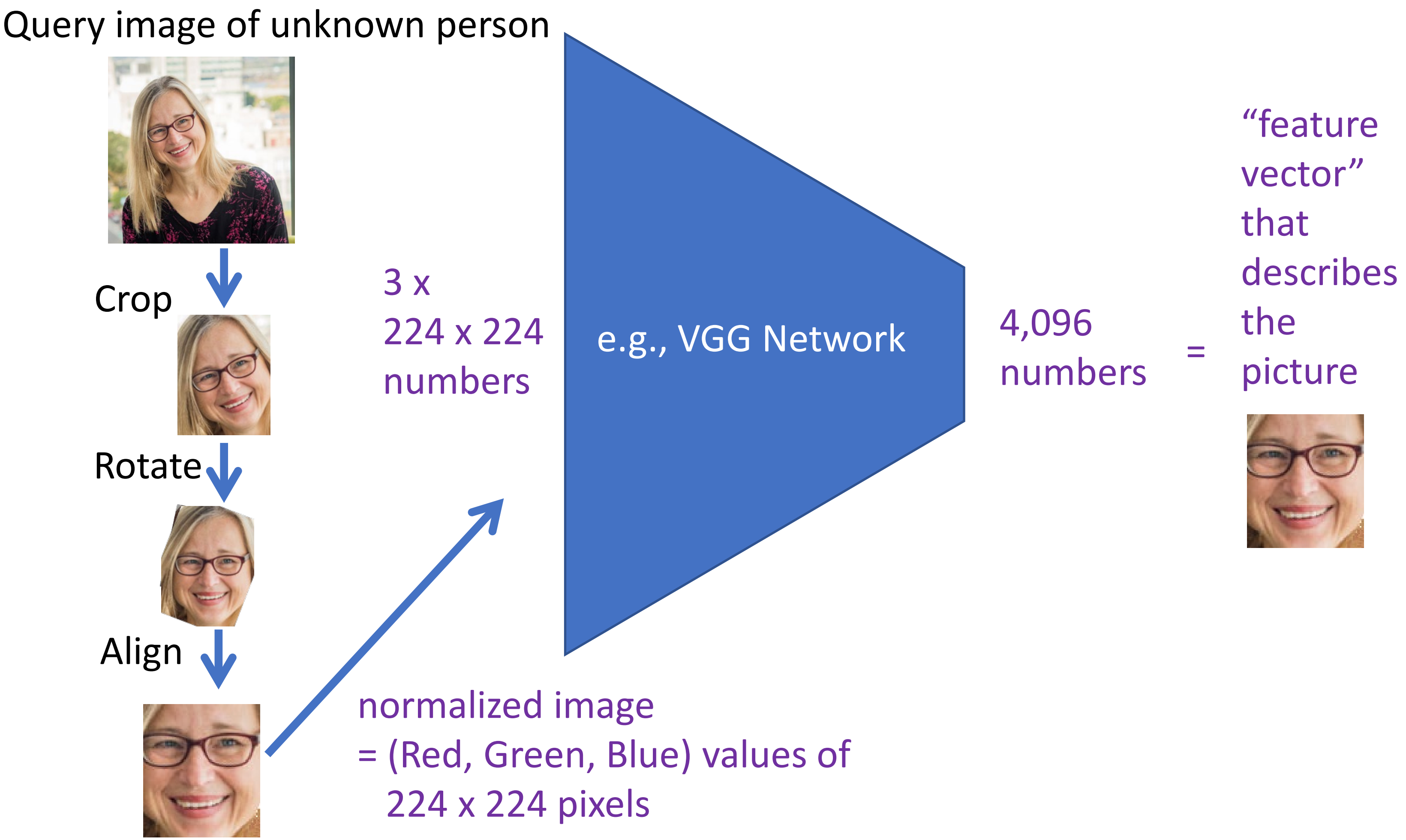
3 x
224 x 224
numbers

4,096
numbers
=
“feature
vector”
that
describes
my
picture

normalized image for “Margrit Betke”
= Red Green Blue values of
224 x 224 pixels
minus average RGB value over training set

* VGG = Visual Geometry Group, Oxford

Face Recognition in “Use Mode”



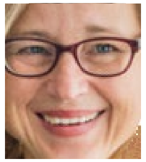
Face Verification in "Use Mode"

Query image of unknown person

Crop

Rotate

Align



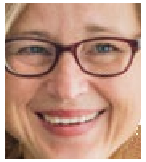
normalized image
= (Red, Green, Blue) values of
224 x 224 pixels

3 x
224 x 224
numbers

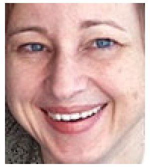
VGG Network

4,096
numbers
=

"feature
vector"
that
describes
the
picture



"feature
vector"
that
describes
M.B.'s
picture



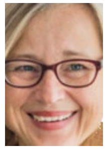
Face Verification in “Use Mode”

Query image of unknown person

Crop

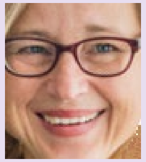
Rotate

Align



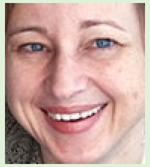
VGG Network

“feature vector” that describes the picture



minus

“feature vector” that describes M.B.’s picture

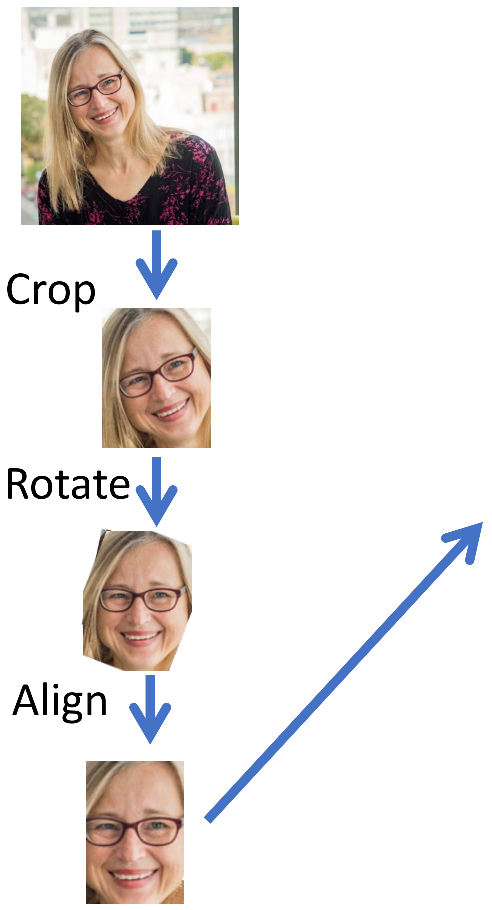


= error

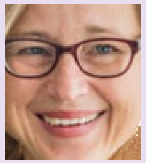
IF error small
THEN unknown person = Margrit Betke

Face Verification in “Use Mode”

Query image of unknown person

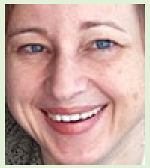


(“feature vector” that describes the picture



minus

“feature vector” that describes M.B.’s picture



$)^2$ = Euclidean Error

IF Euclidean error small
THEN unknown person = Margrit Betke

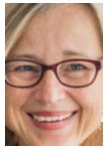
Face Verification in “Use Mode”

Query image of unknown person

Crop

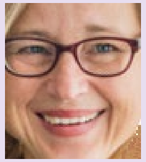
Rotate

Align

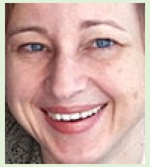


VGG Network

cos (“feature vector” that describes the picture



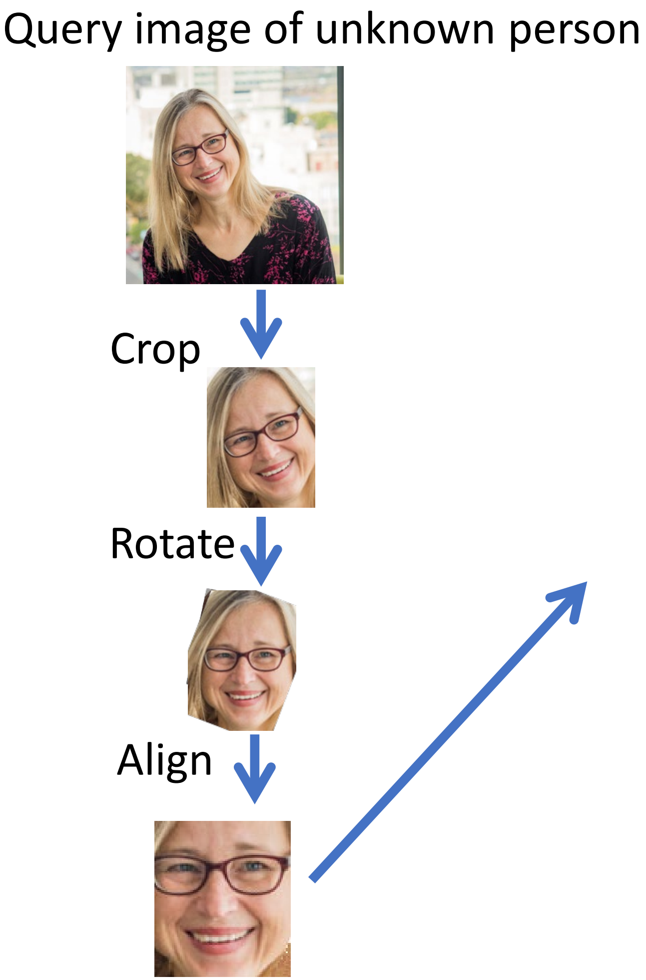
, “feature vector” that descries M.B.’s picture

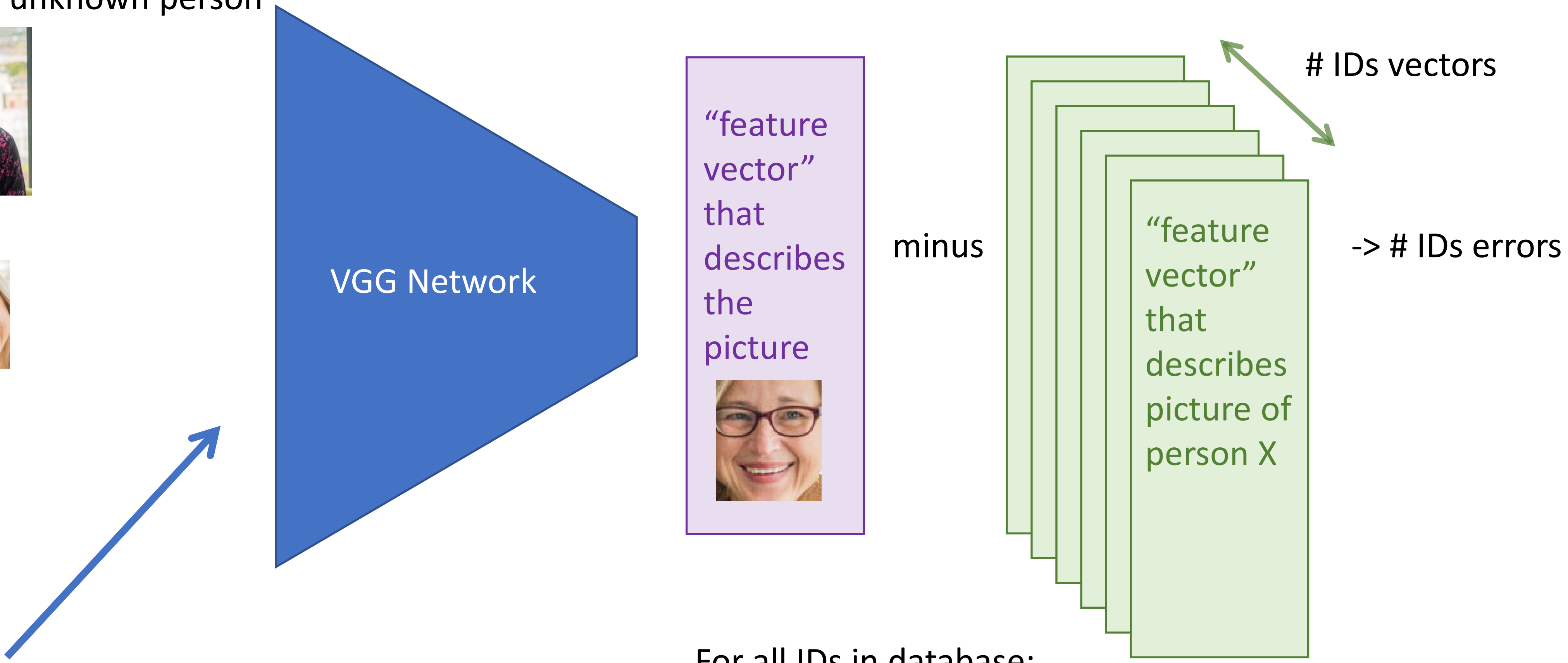


) = Cosine Similarity

IF Cosine Similarity high
THEN unknown person = Margrit Betke

Face Identification in "Use Mode"





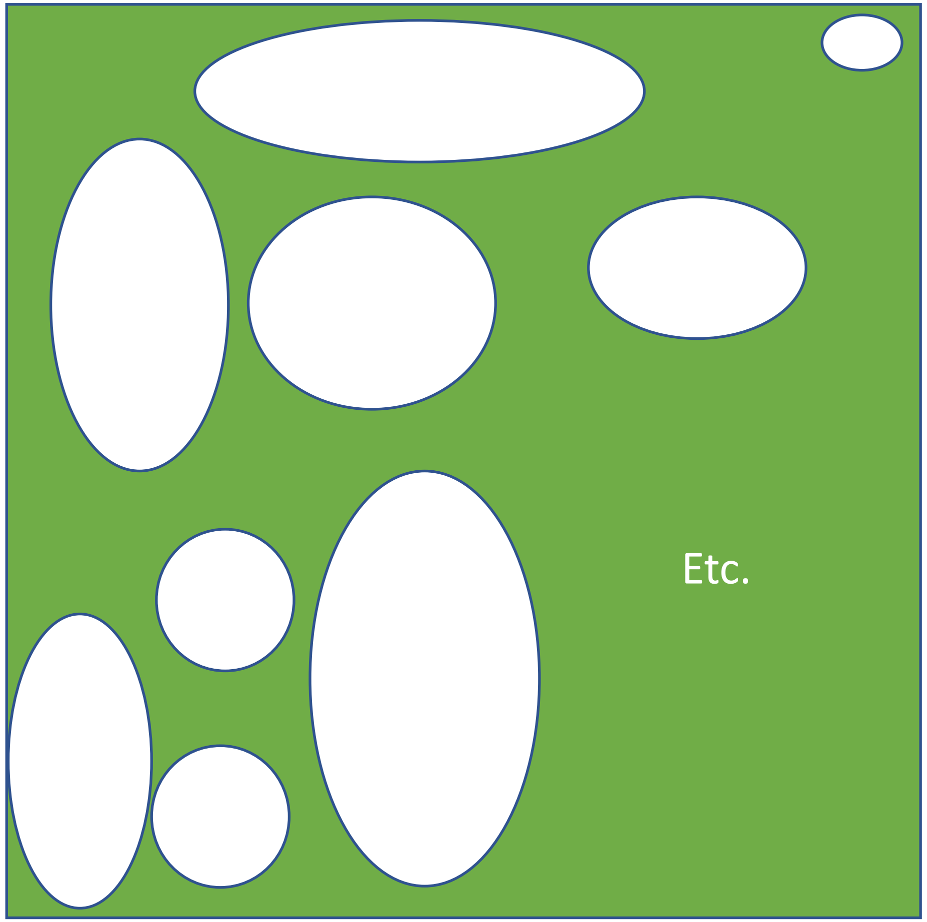
For all IDs in database:
  IF smallest error for ID x
    THEN unknown person = ID x

Benchmark Dataset (from 2007, still used)
Labeled Faces in the Wild (LFW) *

5,749 identities
13,233 face images
1,680 people with two or more images

- Publicly available
- Web data
- Celebrities

Gallery of Known Subjects



* vis-www.cs.umass.edu

Recognition Results on LFW Dataset

	Google	facebook
Neural Net Name:	FaceNet 2015	DeepFace 2014
Number of Photos:	> 500 Million	4.4 Million
Number of Subjects:	> 10 Million	4,000
Accuracy:	99.6%	97.3%

Do these accuracy numbers show that the problem was solved already in 2015?

Do these accuracy numbers show that the problem was solved already in 2015?

No!

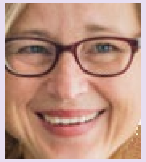
1. Distractor images
2. Training versus testing datasets
3. What is the best network/algorithm?
4. What are the limitations of existing systems?

Benchmark Dataset Labeled Faces in the Wild (LFW)

What if we add
1 million
“distractor”
face images?

5,749 identities
13,233 face images

“feature vector” that describes the picture



“feature vector” that describes picture of person X

5,749

1 Mill

Gallery of Known Subjects

Benchmark Dataset Labeled Faces in the Wild (LFW)

Gallery of Known Subjects

With 1 million "distractor" face images:

Recognition rates go down a lot!

Do these accuracy numbers show that the problem was solved in 2015?

No!

1. Distractor images ✓
2. Size of datasets
3. What is the best network/algorithm?
4. What are the limitations of existing systems?

Size of Training Datasets

	Google	facebook
Neural Net Name:	FaceNet 2015	DeepFace 2014
Number of Photos:	> 500 Million	4.4 Million
Number of Subjects:	> 10 Million	4,000
Accuracy on LFW dataset (5K):	99.6%	97.3%

Do these accuracy numbers show that the problem is solved?

No!

1. Distractor images makes the problem much more difficult ✓
2. Size of datasets does matter a lot ✓
3. What is the best network/algorithm?
4. What are the challenges & limitations of existing systems?

Is the key to success the size of the training data or the network ?

	Google	facebook
Neural Net Name:	FaceNet 2015	DeepFace 2014
Number of Photos:	> 500 Million	4.4 Million
Number of Subjects:	> 10 Million	4,000
Accuracy on LFW dataset:	99.6%	97.3%

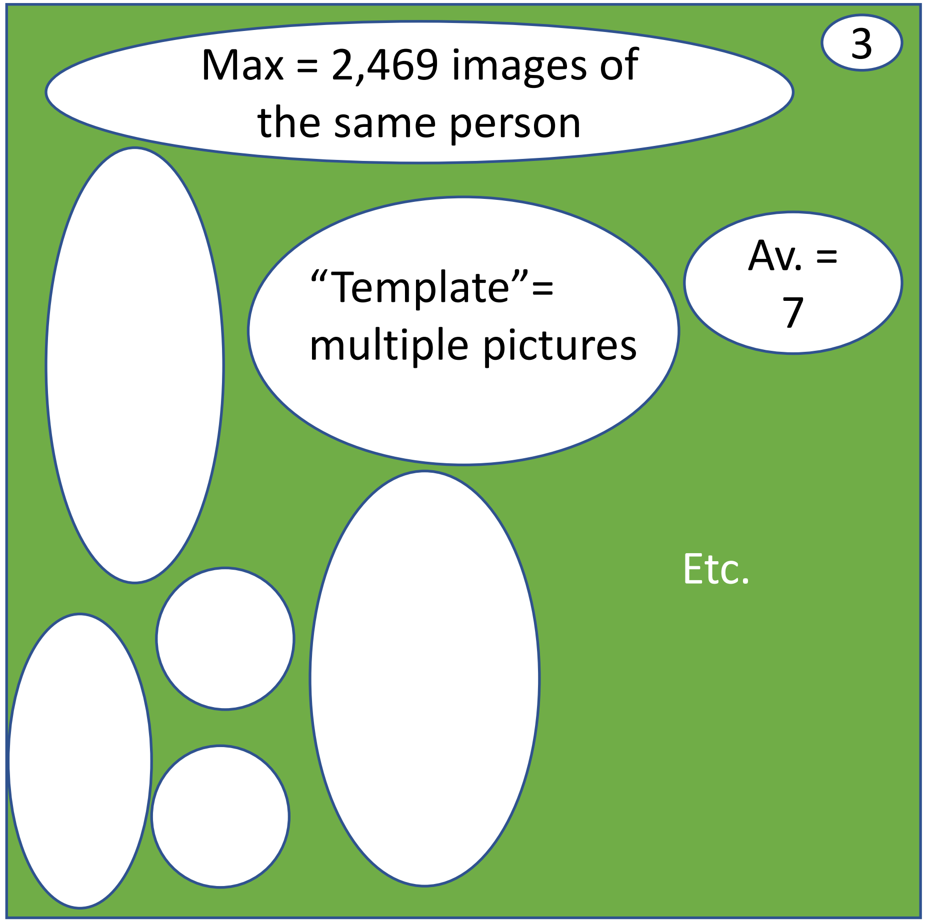
2017 MegaFace2* Dataset

https://arxiv.org/pdf/1705.00393.pdf

672,057 identities
4,753,320 face images

- Publicly available
- No celebrities
- Flickr account data
- Automated labeling
- 59% males, 41% females
- Age range among template images: 16 years

Gallery of Known Subjects



* megaface.cs.washington.edu

2017 Competition

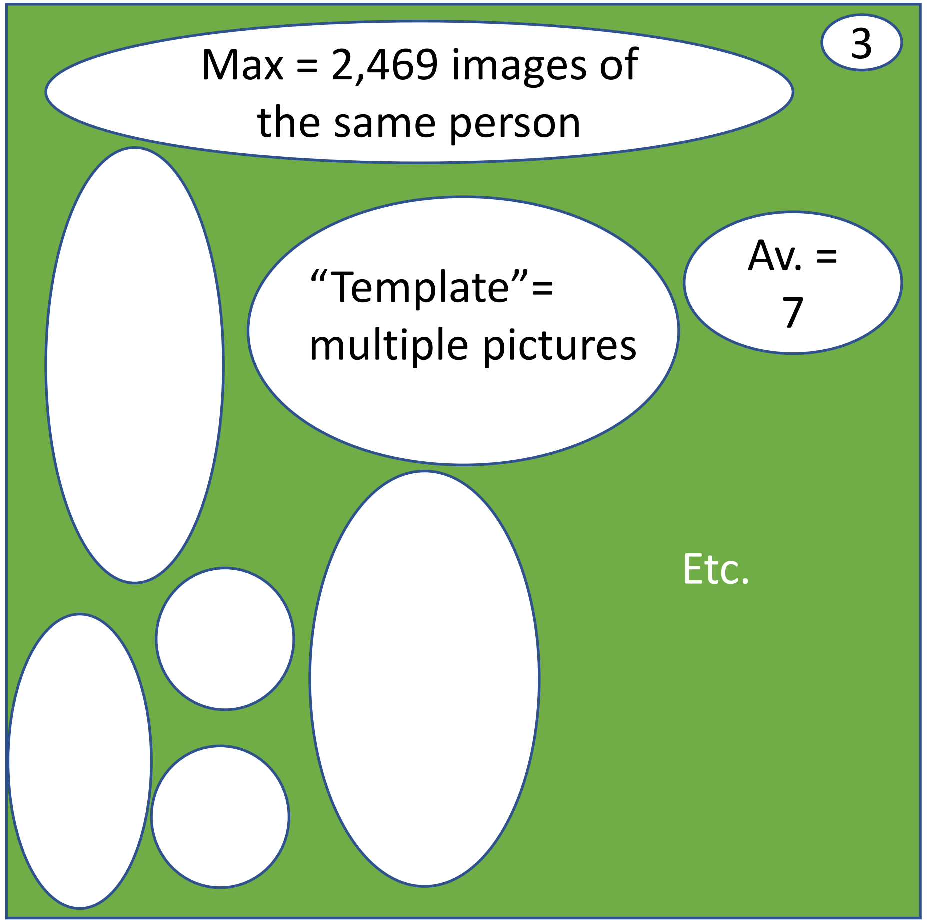
- Train on MegaFace2 (672K IDs, 4 mill. images)
- Test on FaceScrub* (530 IDs, 106K images)
- Add 1 million “distractor” face images

6 teams provided feature vectors to competition organizers on FaceScrub & distractor images

Results varied between 28% to 76%
recognition accuracy

Method matters!

Gallery of Known Subjects



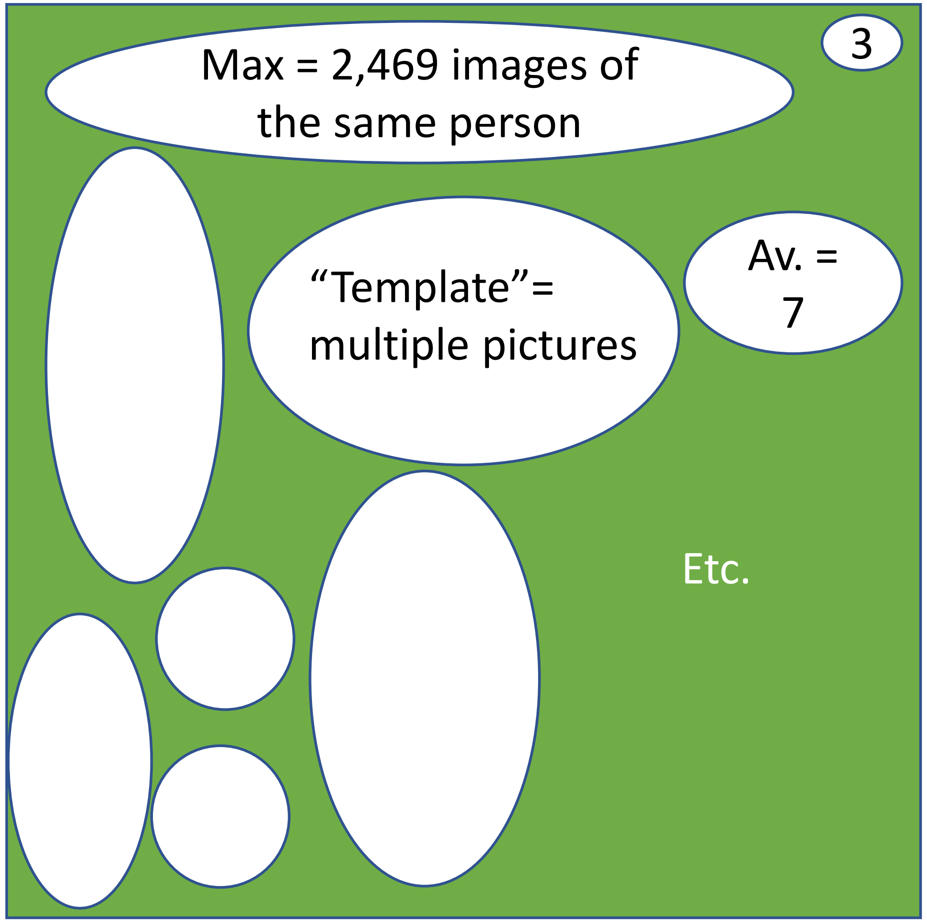
2017 MegaFace2 Dataset

- Train on MegaFace2
- Test on FaceScrub (530 IDs, 106K images)
- Add 1 million “distractor” face images

E.g., NEC’s commercial product:
~100% accuracy without distractors
~60% accuracy with 1 million distractors

Best method: GRCC with 76% accuracy with
1 million distractors

Gallery of Known Subjects



Neural Networks for Biometrics

We don’t know what was under the hood in the 2017 Competition.
Secrets of the trade…

More recently, network structures, loss functions, and training schemes have been published.

Most recent conference: International Joint Conference on Biometric, Ljubljana, Slovenia, September 25-28, 2023

Our research group’s work won a “Best Poster Award”

Do these accuracy numbers show that the problem is solved?

No!

1. Distractor images makes the problem much more difficult
2. Size of datasets does matter a lot
3. Competitions used to determine best network/algorithm
4. What are the challenges & limitations of existing systems?

Do these accuracy numbers show that the problem is solved?

No!

1. Distractor images makes the problem much more difficult
2. Size of datasets does matter a lot
3. Competitions used to determine best network/algorithm
4. What are the challenges & limitations of existing systems?

What are other limitations of existing systems?

System performance degrades due to

Aging

Pose Variations: Frontal vs. Profile

What are other limitations of existing systems?

System performance degrades due to

Aging

Pose Variations: Frontal vs. Profile

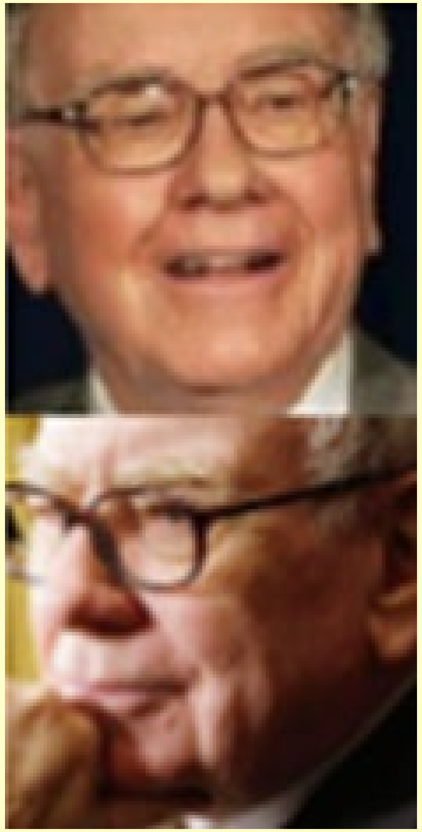
- Klare et al., CVPR 2015: IARPA Janus Benchmark A
- Sengupta et al., WACV 2016: CFP Dataset
- Yu et al., ICCV 2017: AFLW2000 Dataset

Let’s do a Human Experiment on Recognizing Faces in Frontal versus Profile Images

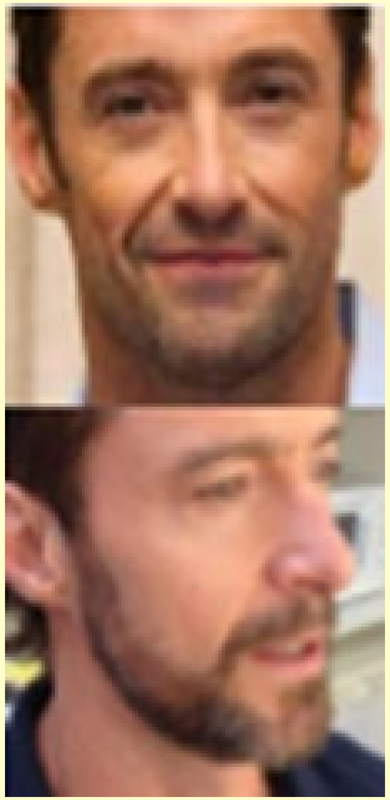
Please determine if the following images show the same person.

Poll: Same Person?

1



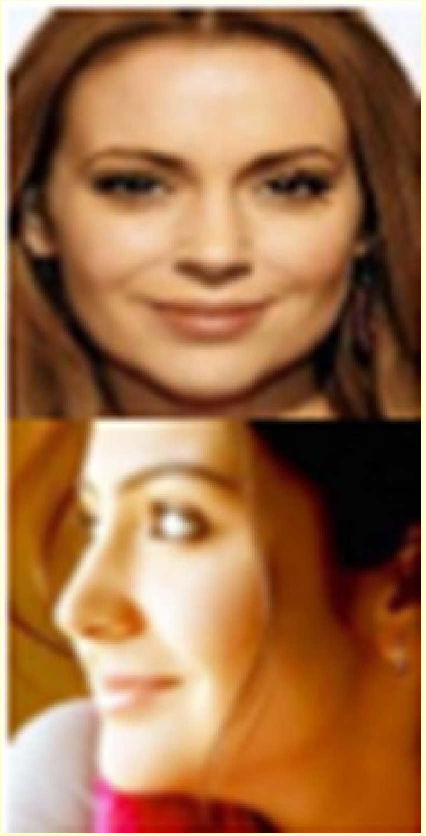
2



3

4

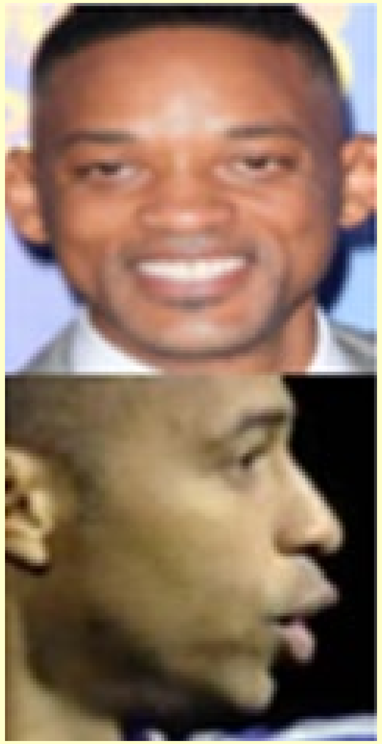
5



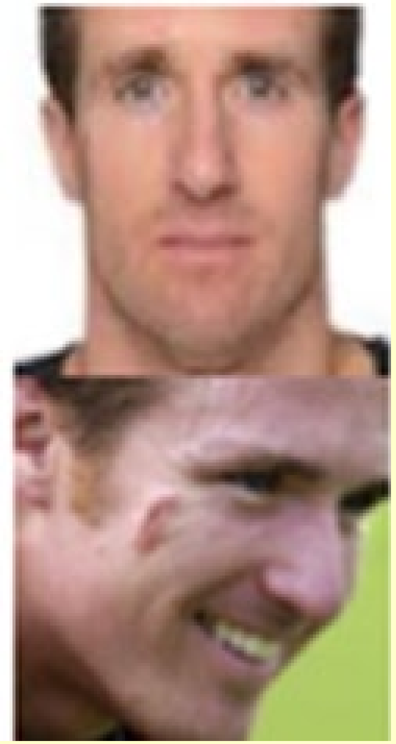
Cao et al., 2018

Same Person?

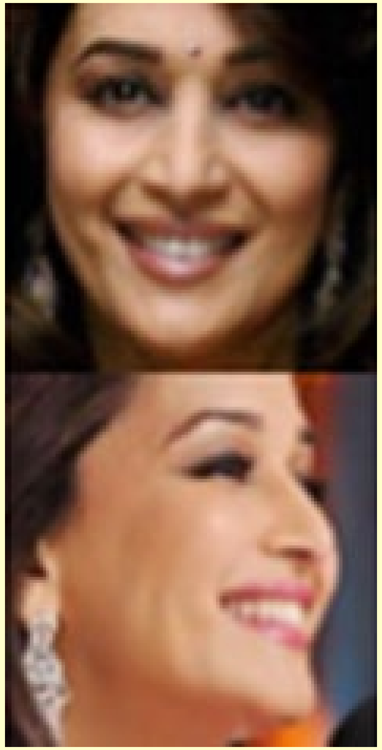
6



7

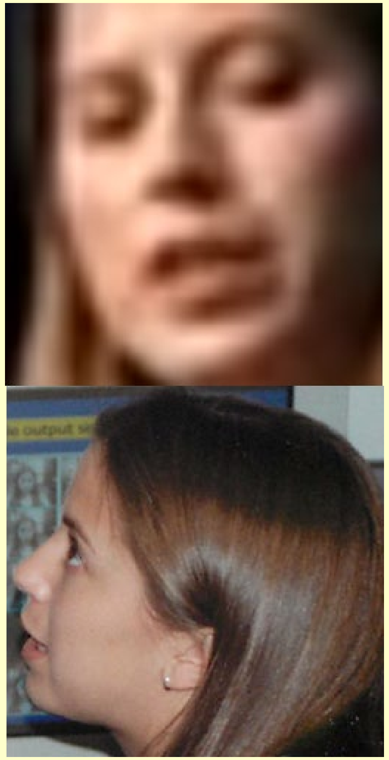


8



9

10



11

Frontal vs. Profile

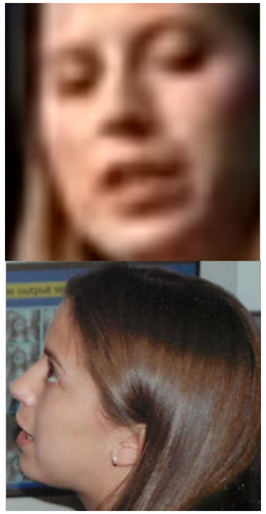
GT: Not same person

GT: Same person

Cao et al., 2018

GT:
Same persons:

Kristen Grauman, UTA



Pawan Sinha, MIT

Research on Face Recognition under Pose Variations

Cao et al., CVPR 2018

Idea: Map profile representations into frontal pose representations

Results:

Verification:

True Acceptance Rate (TAR) at

False Acceptance Rate (FAR) of

0.01: 94%

0.001: 89%

Rank 1 Identification:

96.8%

Zhu et al., PAMI 2019: [pdf]

Idea: 3D Dense Face Alignment (3DDFA) = Use DNNs to estimate 3D Morphable Model (3DMM) parameters:

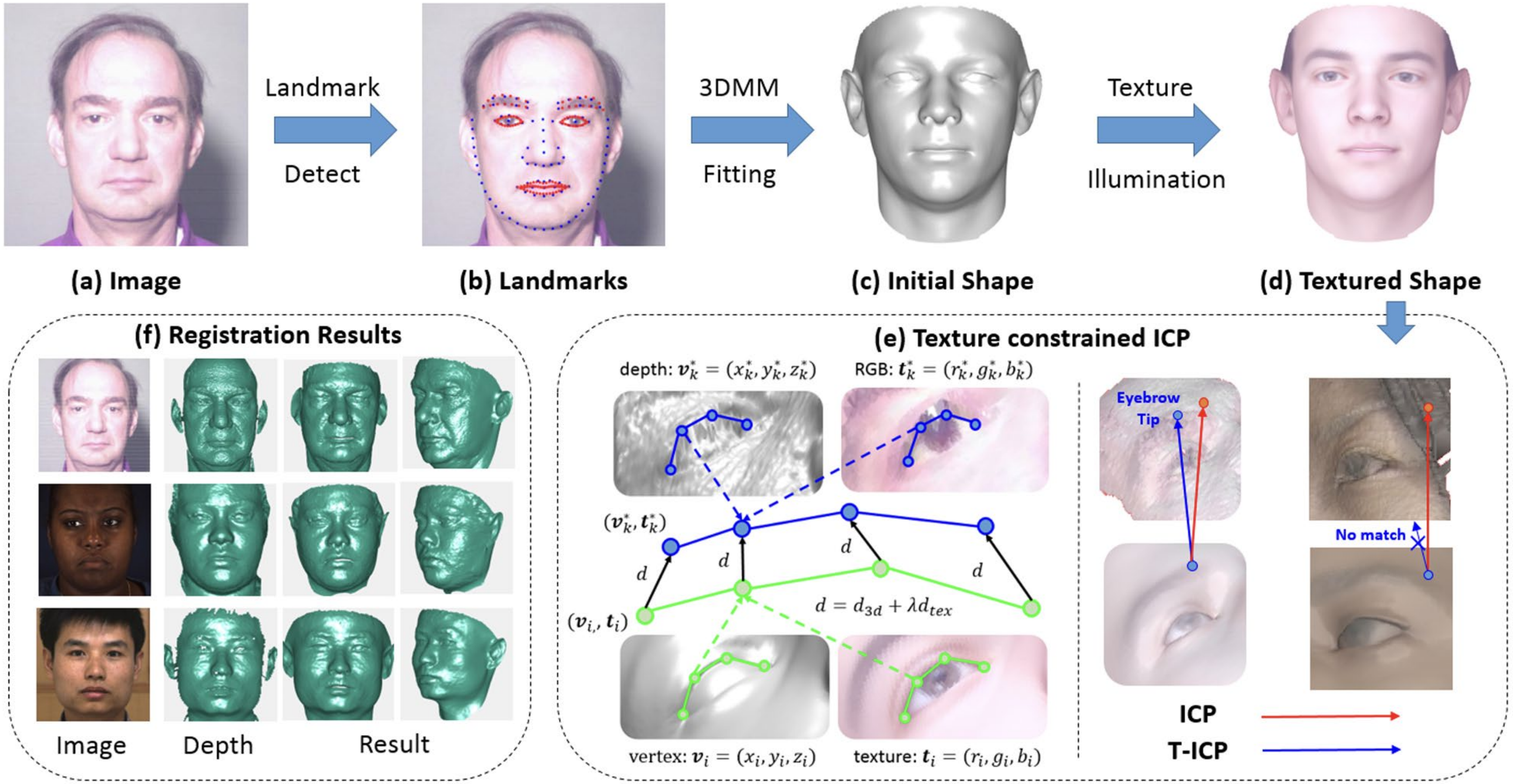
- Pose: 3 Euler angles, translation, scale
- Shape: 50 dimensions
- Expression: 19 dimension

Results:

Better than state-of-the-art but relatively poor (regression by neural network difficult!)

Beyond 3DMM: Learning to Capture High-fidelity 3D Face Shape

Zhu et al., [ECCV 2020](), [PAMI 2022]():



ICP: Iterative Closest Point Algorithm.
We will discuss in detail in a later lecture.

What are other limitations of existing systems?

System performance degrades due to

Aging

- Best-Rowden-Jain-PAMI-2017
- NIST Mugshot Identification Database (MID)
- NIST Multiple Encounter Dataset (MEDS), FBI Biometric Center of Excellence

Let’s do a Human Experiment on Recognizing Age Difference of Images of Faces

Poll: Estimate Aging

A B C D

Credit for Images A-D:
Best-Rowden and Jain, 2018

E F G H

Images E-H: © Betke

Estimating Aging

Best-Rowden and Jain, 2018

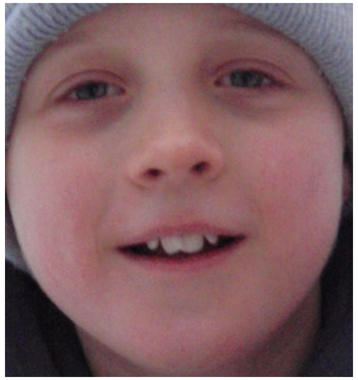
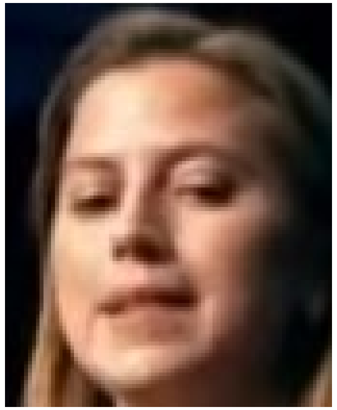
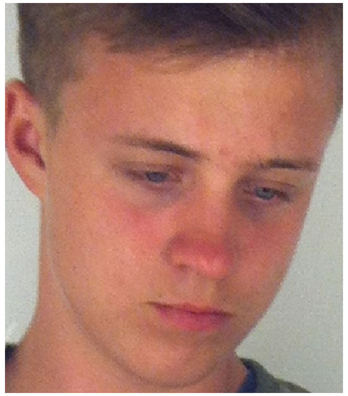
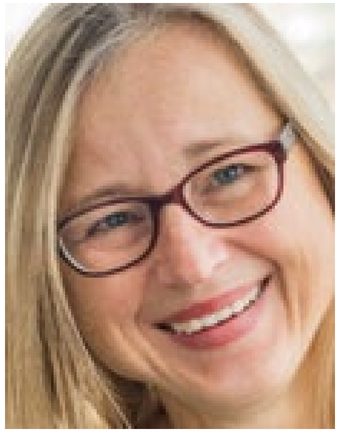
A B C D

Elapsed time: 9 years 8 years 8 years 9 years

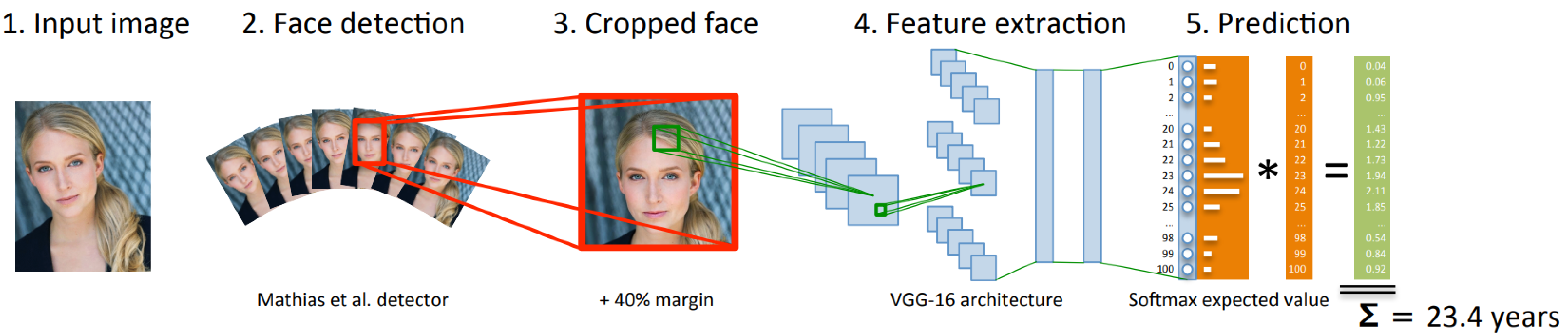
99% of subjects can still be recognized at 0.01% FAR up to approximately 6 years elapsed time

Best-Rowden and Jain, 2018

Estimating Aging

E	F	G	H
			
			
Age Difference: 8 years	12 years	30 years	30 years

Deep EXpectation (DEX): Age Estimation



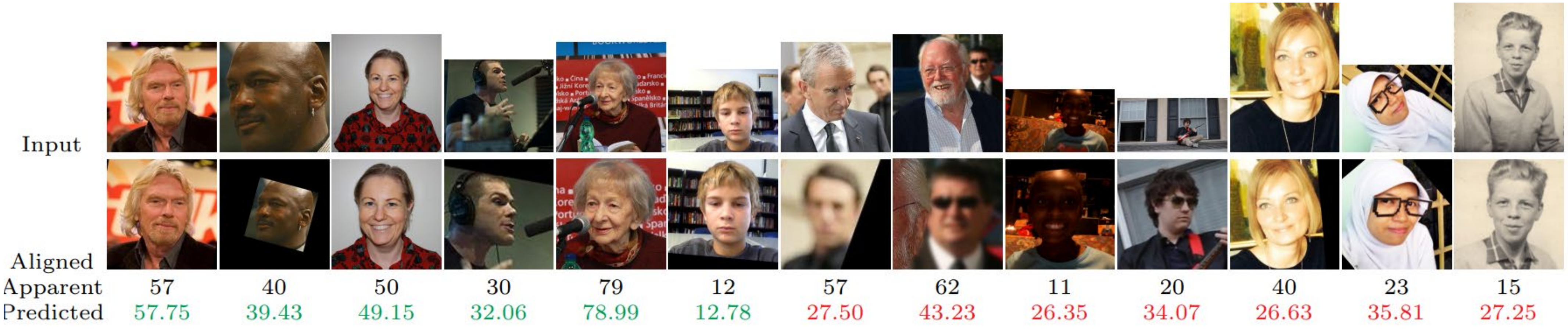


Fig. 7 Examples of face images with good and bad age estimation by DEX.

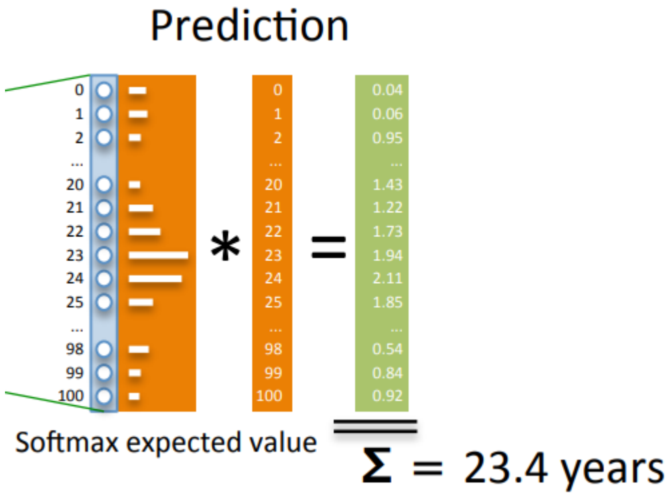
Image Credit: Rothe et al., 2016

BOSTON UNIVERSITY

SSR-Net: A Compact Soft Stagewise Regression Network for Age Estimation

In DEX:

- Regression by conversion into age classes
- #age-groups=100,
- Representative age of *i*th bin is μ_i
- μ_{23} is age 23

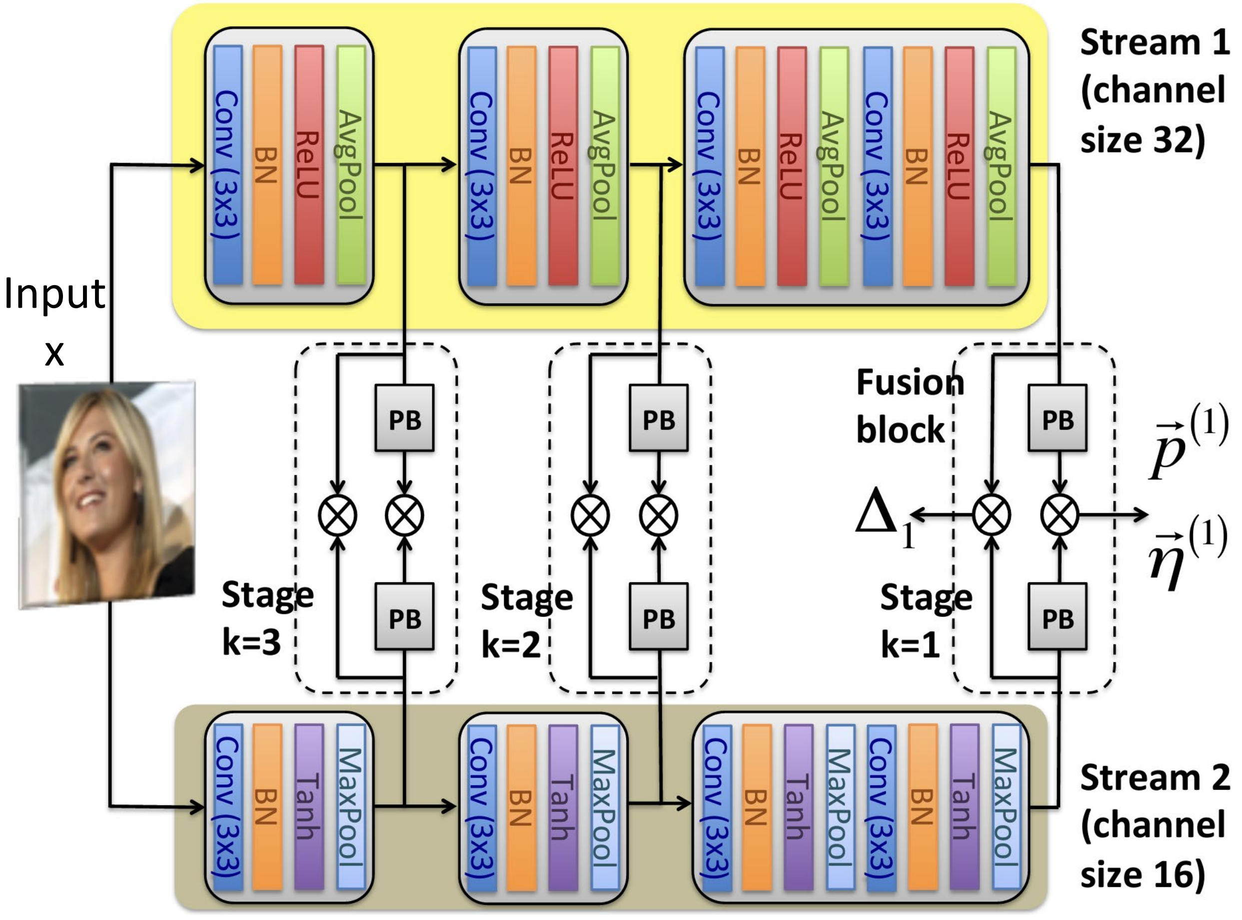


In SSR-Net:

- **Regression by coarse-to-fine hierarchical conversion** into age classes
- Coarse: "relatively younger" "about right age" "relatively older"
- Next stage: Refine decision of previous stage
- Representative of age group: $\mu_i = i\ V/\#\text{age-groups}$
- p(i) = probability x belongs to group *i*
- Estimated age $\tilde{y} = \sum_{k=1}^{K} \boldsymbol{\mu}^{(k)} \cdot \mathbf{p}^{(k)}$ where K = # stages
- To avoid quantization noise & class ambiguity when navigating between stages: Use bin shift η and scale factor Δ
- The shift vector $\boldsymbol{\mu}^{(k)}$ adjusts the center for each bin at kth stage
- The scale factor $\Delta^{(k)}$ scales the widths of all bins at kth stage
- Learned by network: $p^{(k)}, \eta^{(k)}, \Delta^{(k)}$ for all k

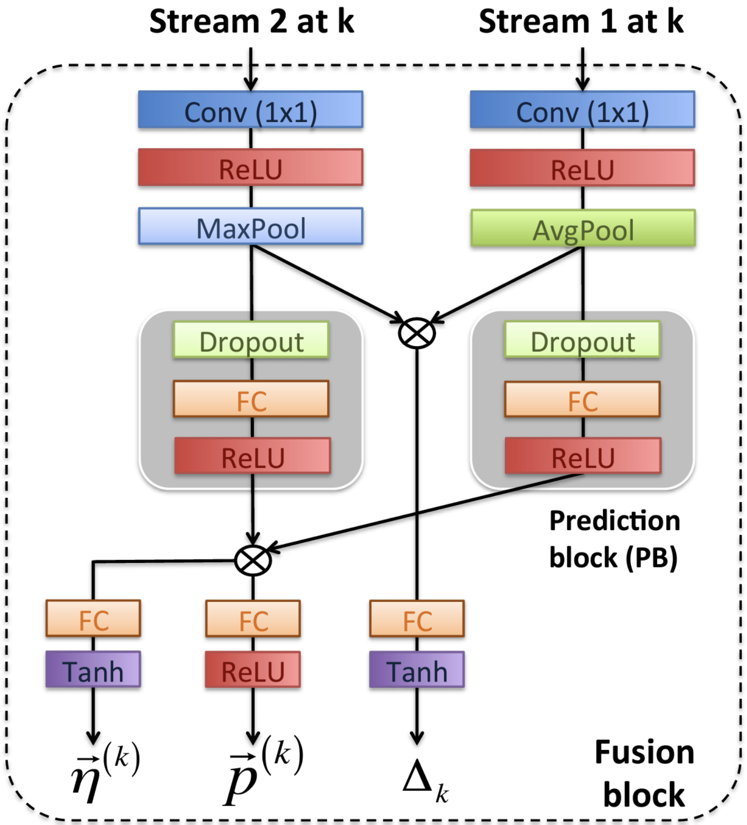
Insight: This is a general regression scheme!

SSR-Net: A Compact Soft Stagewise Regression Network for Age Estimation



μ_i = *i* V/#age-groups

p(i) = probability x belongs to group i



Estimated age $\tilde{y}$

$= \boldsymbol{\mu}^T \mathbf{p}$

Bin shift: η

Scale factor: Δ

Stage #1

$\left\{\vec{p}^{(1)}, \vec{\eta}^{(1)}, \Delta_1\right\}, s_1$

Stage #2

$\left\{\vec{p}^{(2)}, \vec{\eta}^{(2)}, \Delta_2\right\}, s_2$

Stage #3

$\left\{\vec{p}^{(3)}, \vec{\eta}^{(3)}, \Delta_3\right\}, s_3$

Soft stagewise regression

Head Pose Estimation

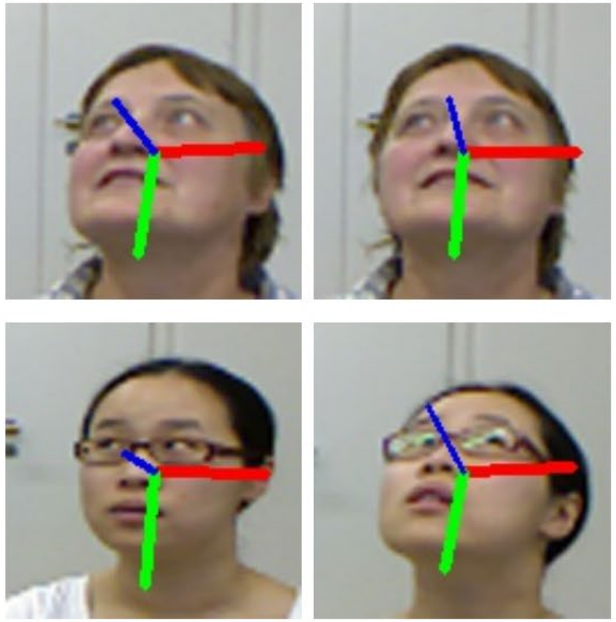
FSA-Net: Learning Fine-Grained Structure Aggregation for Head Pose Estimation from a Single Image

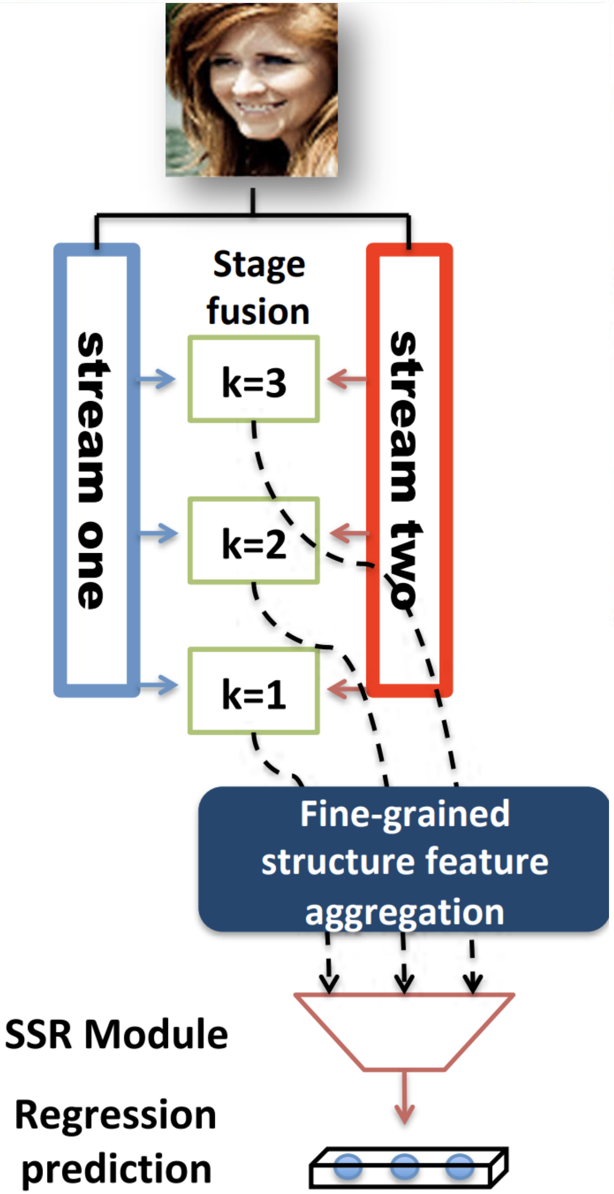
Goal:

Estimate yaw, pitch, roll angles of head

Regression Problem:

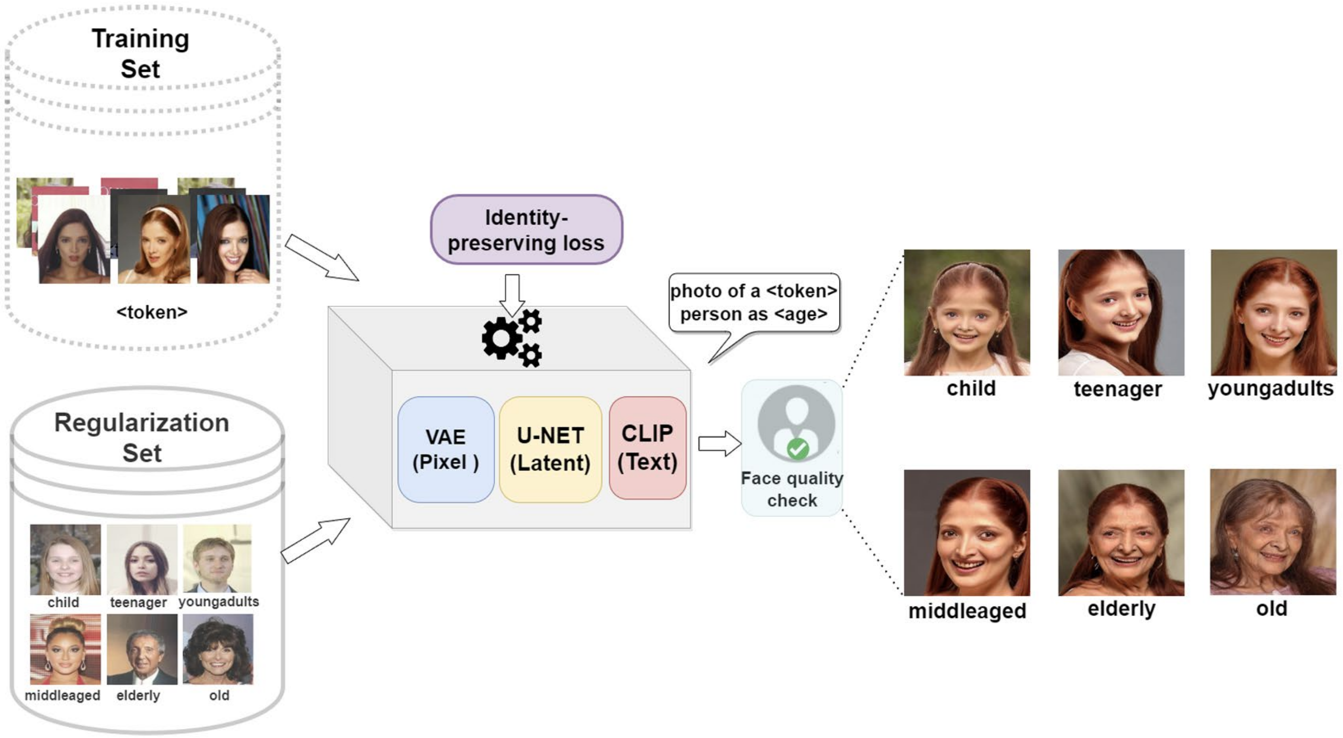
Use SSR-Net but for vector output





Identity-Preserving Aging of Face Images via Latent Diffusion Models*

Banerjee et al., IJCB 2023: Use a latent text-to-image diffusion model to synthetically age and de-age face images



* We will go into Diffusion Model later this semester

International Joint Conference on Biometrics, September 2023: Keynote Address by Mayank Vatsa

Biometric Datasets from IIT Jodhpur

https://iab-rubric.org/resources

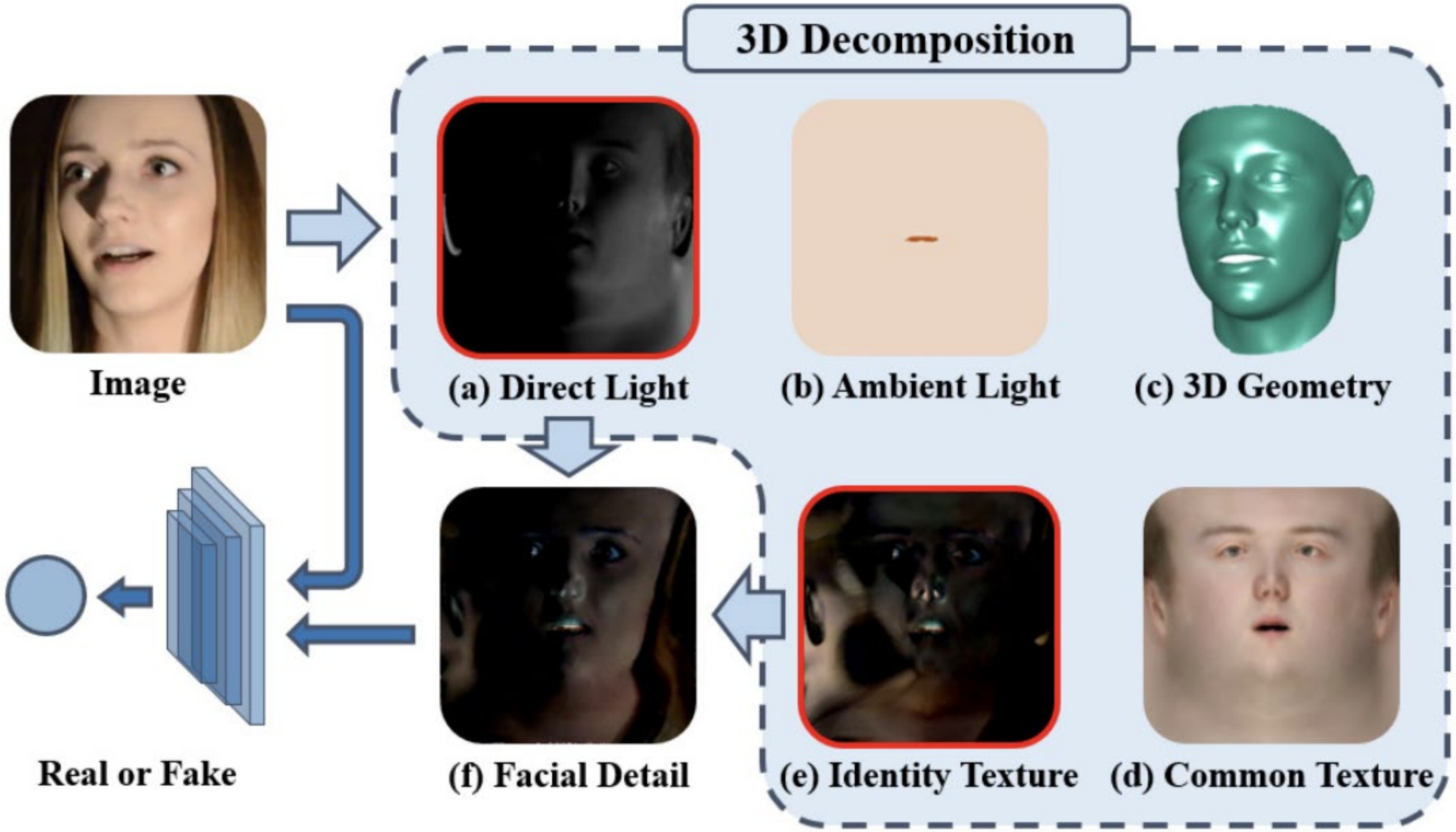
- DroneSurf: face recognition from drones
- [Dhamecha et al., IJCB 2023, Disguise Detection]()
- TIFS 2023: Largest fingerprint dataset
- Injured Face Recognition (120 unclaimed dead identified)

- [Narayan et al., CVPR 2023]()
- [DFPlatter:]() Multi-subject deepfakes

Figure 1. Samples showcasing multi-face deepfakes circulated on social media. (a) A zoom call with a deepfake of Elon Musk [8] (b) Real-time deepfake generation at America's Got Talent [9] (c) Deepfake round-table with multiple deepfake subjects [33].

Face Forgery Detection by 3D Decomposition

Zhu et al., CVPR 2021, Trans PAMI, 2023



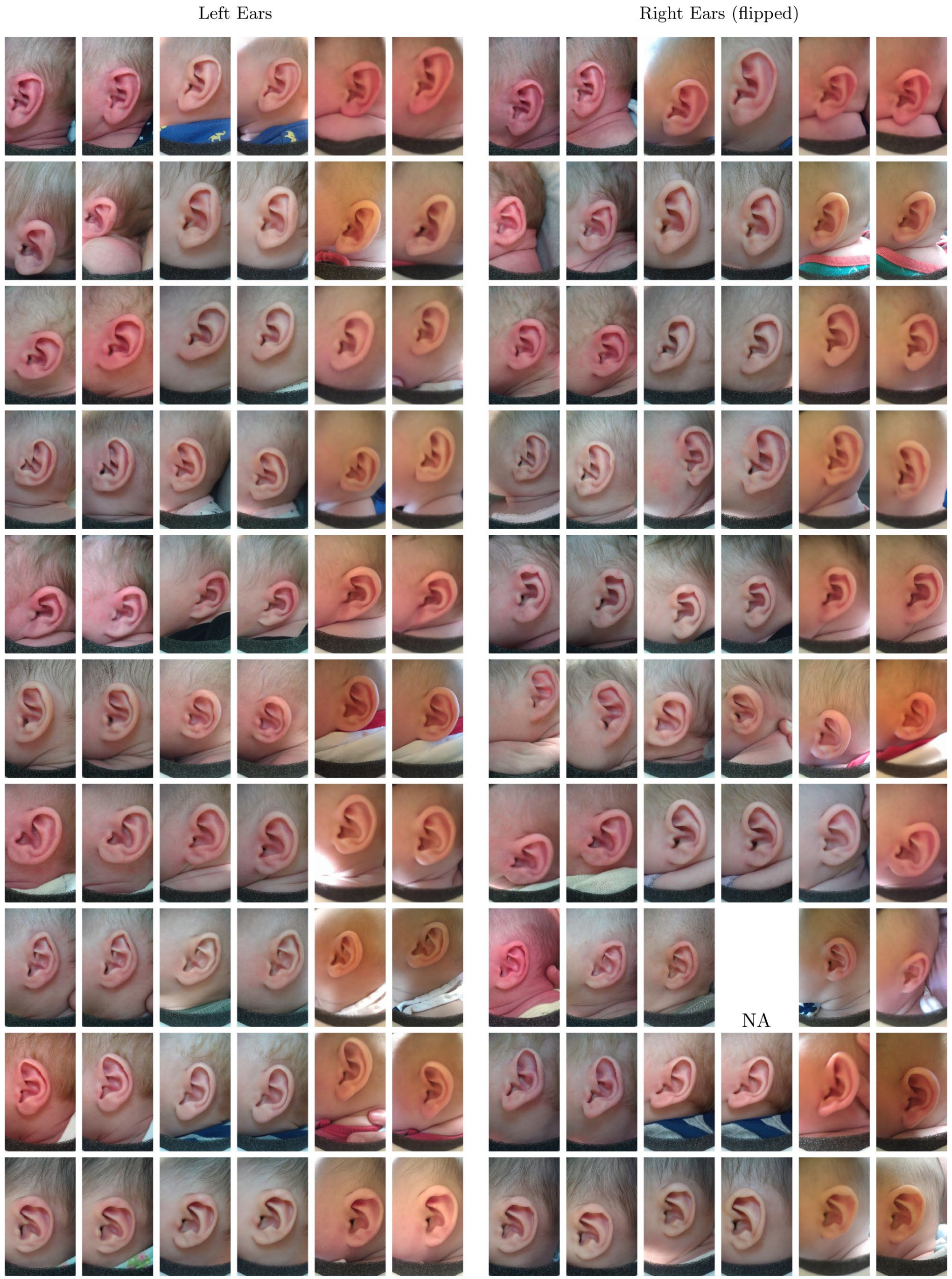
Alternative Biometrics: Ears

Our Task:

Can we identify newborns, during the subsequent months based on images of their growing ears?

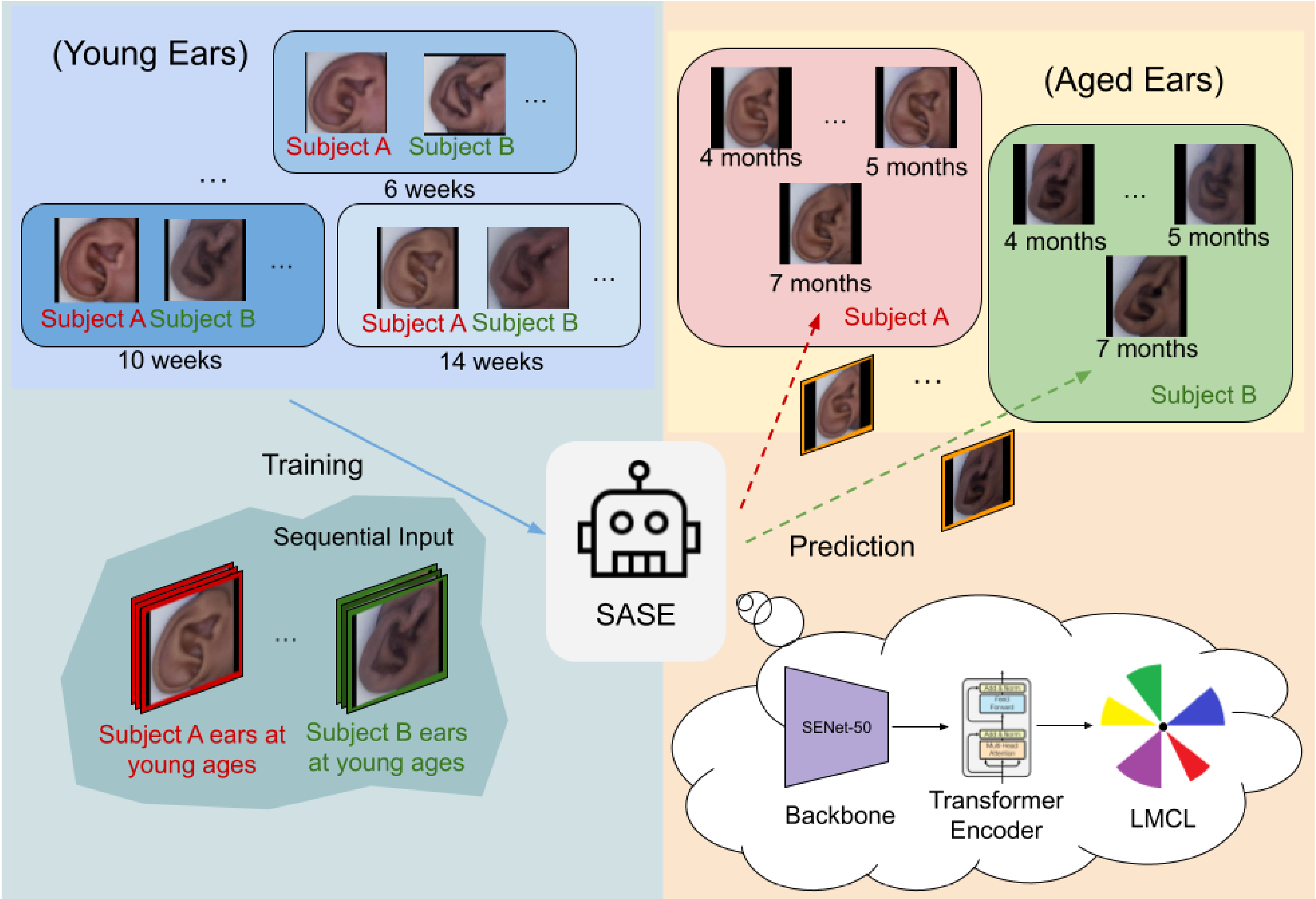
Motivation:

- Infants should be immunized.
- Infants born to women with HIV should receive the HIV-prevention medicine zidovudine.
- Our collaborators in rural Zambia, health care professionals who manage a network of clinics, have difficulties tracking down babies.

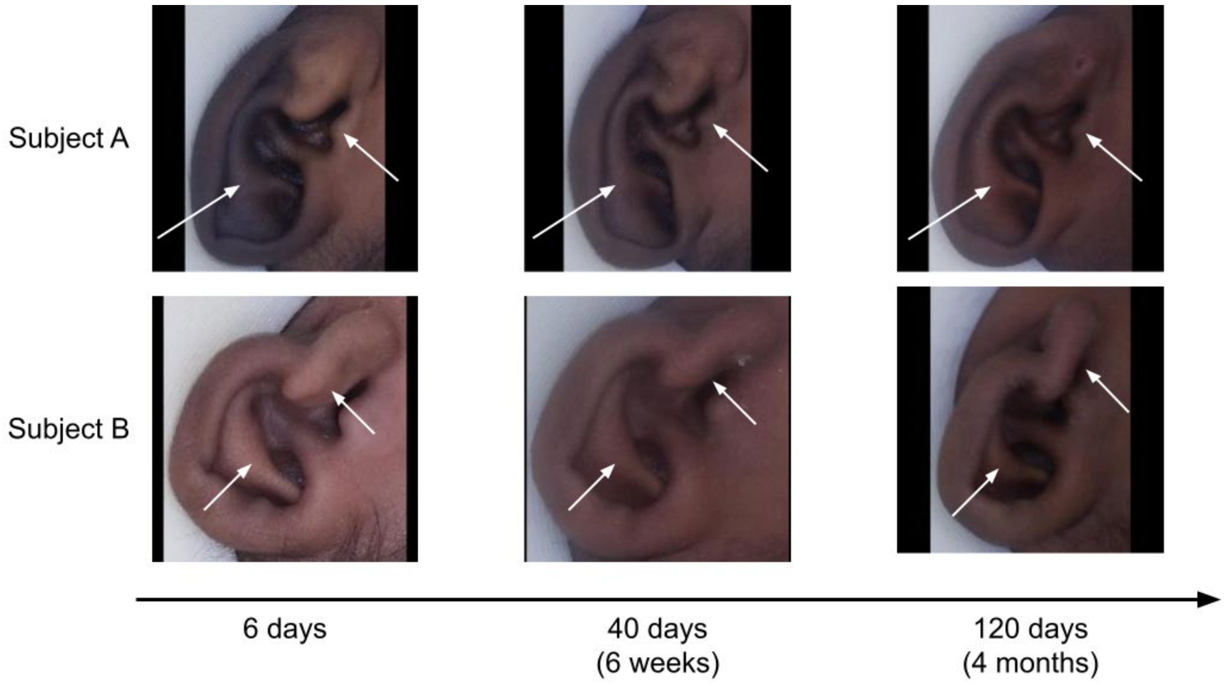


BOSTON UNIVERSITY

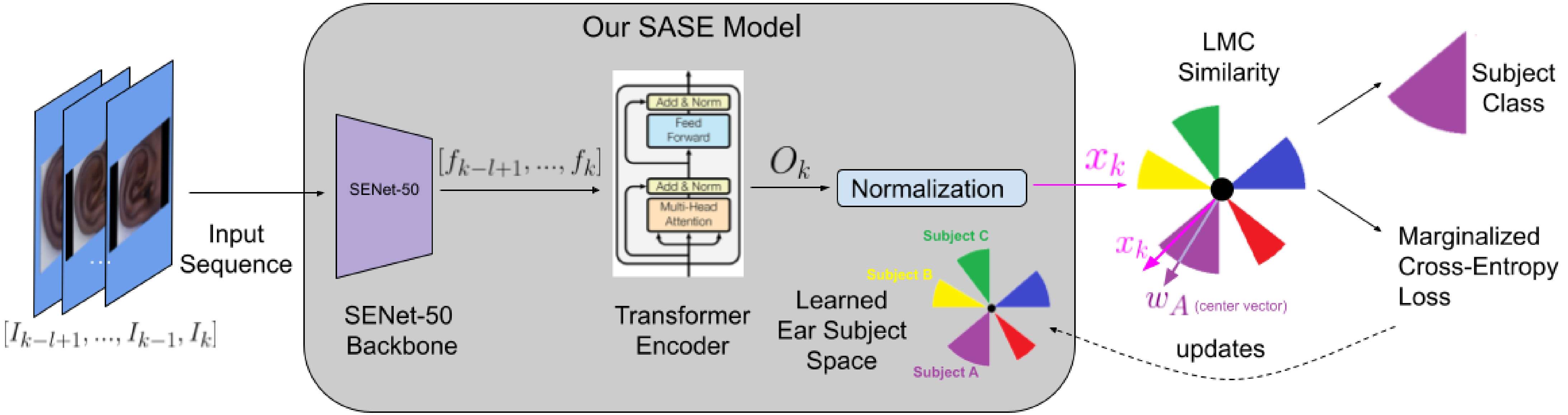
IJCB 2023 Best Poster Award: Qin et al., Age-constrained Ear Recognition: The EICZA Dataset and SASE Baseline Model



IJCB 2023 Best Poster Award: Qin et al., Age-constrained Ear Recognition: The EICZA Dataset and SASE Baseline Model



IJCB 2023 Best Poster Award: Qin et al., Age-constrained Ear Recognition: The EICZA Dataset and SASE Baseline Model



IJCB 2023 Best Poster Award: Qin et al., Age-constrained Ear Recognition: The EICZA Dataset and SASE Baseline Model

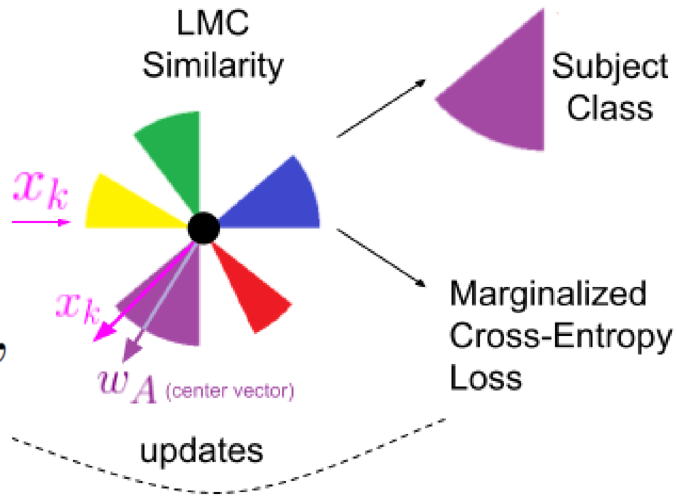
Our SASE Model

The Large

Margin Cosine Loss is defined as

$$\ell_{\mathrm{LMC}} = \frac{1}{N}\sum_{i}^{N} -\log \frac{e^{s(\cos(\theta_{y_i},i)-M)}}{e^{s(\cos(\theta_{y_i},i)-M)} + \sum_{p\neq y_i} e^{s\cos(\theta_{p,i})}},$$

where N is the number of training samples, i stands for the ith sample, y_i stands for the ground truth subject of the sample, $\cos(\theta_p, i) = w_p^T x_i$ (with $\|w^T\| = \|x_i\| = 1$), and the learnable vector w_p of the pth subject, which works as a "center" to represent subject p in latent "ear space."



M = hyperparameter that controls the size of the 'margins', e.g., gaps between colored 'fans'

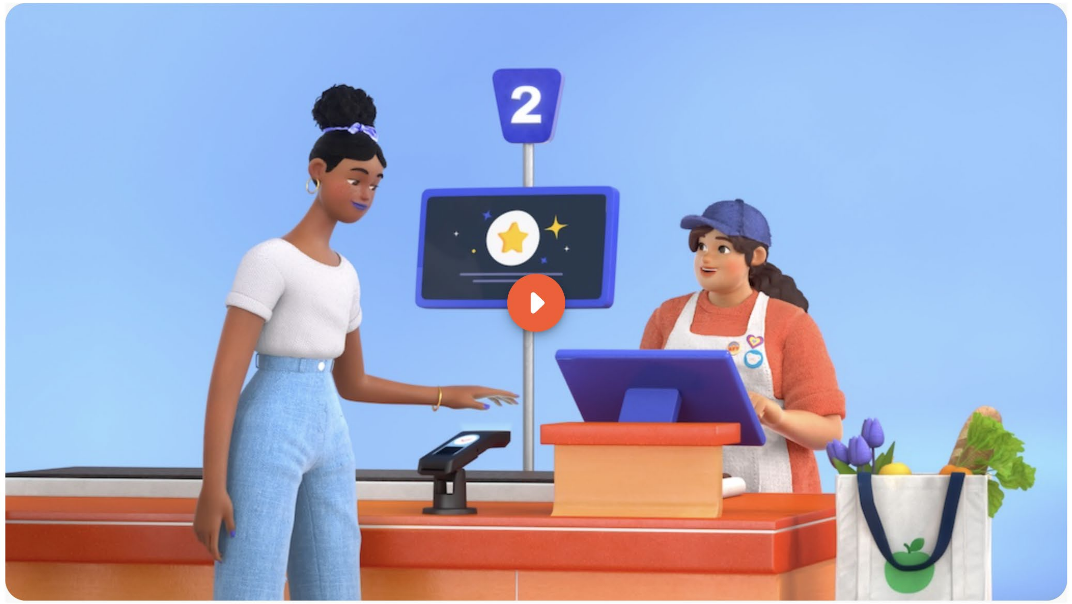
IJCB 2023 Best Poster Award: Qin et al., Age-constrained Ear Recognition: The EICZA Dataset and SASE Baseline Model

Table 3. Average cross-validation recognition accuracy of SASE compared to four baseline models on three datasets

Dataset / Model	UERC [19] without Ear Ages	FG-NET [32] (Aging Faces)		Our EICZA (Aging Ears)		
		Age Neutral Train/Test	Age Constrained Train/Test	Age Neutral Train/Test	Age Constrained Train/Test with Day 6	Age Constrained Train/Test without Day 6
SqueezeNet [28]	26.88%	17.85%	7.24%	52.30 %	8.23%	11.14%
ResNet-50 [25]	36.72%	82.84%	**55.92%**	61.30%	13.84%	22.98%
SENet [27]	41.86%	78.89%	46.05%	68.11 %	18.85%	28.46%
SASE (Our Model)	**42.56%**	**82.90%**	52.96%	**69.49%**	**33.14%**	**49.98%**

The Palm as a Biometric

https://one.amazon.com



Your palm is all you need

Extremely high accuracy needed for financial transactions.

100% accuracy claimed

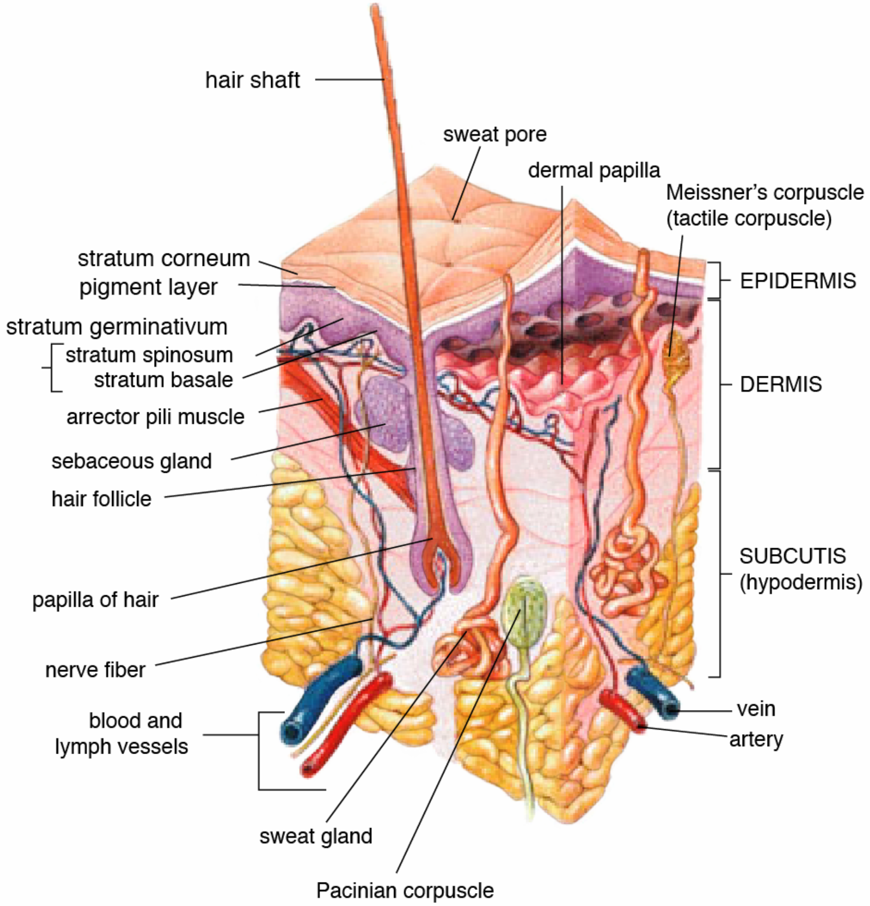
Manoj Aggarwal,
Director of Applied Science,
Amazon One
gave a keynote address
at IJCB 2023, September 28

How does Amazon's Palm Recognition work?

Bi-modal input:

1. RGB image of your palm

2. Subsurface image of your palm illuminated by near infrared light



Subcutaneous image

Computer Vision & Biometrics

Other Biometric Tasks:

- Gait Recognition
- Iris Recognition
- Fingerprint Recognition
- Face Recognition with Face Expressions or micro-expressions

Ethical Concerns:

- Misuse by personal enemies: Fake nude pictures on social media
- Misuse by totalitarian governments: “Big-brother watching you”
- Arms race of fake creation/fake detection

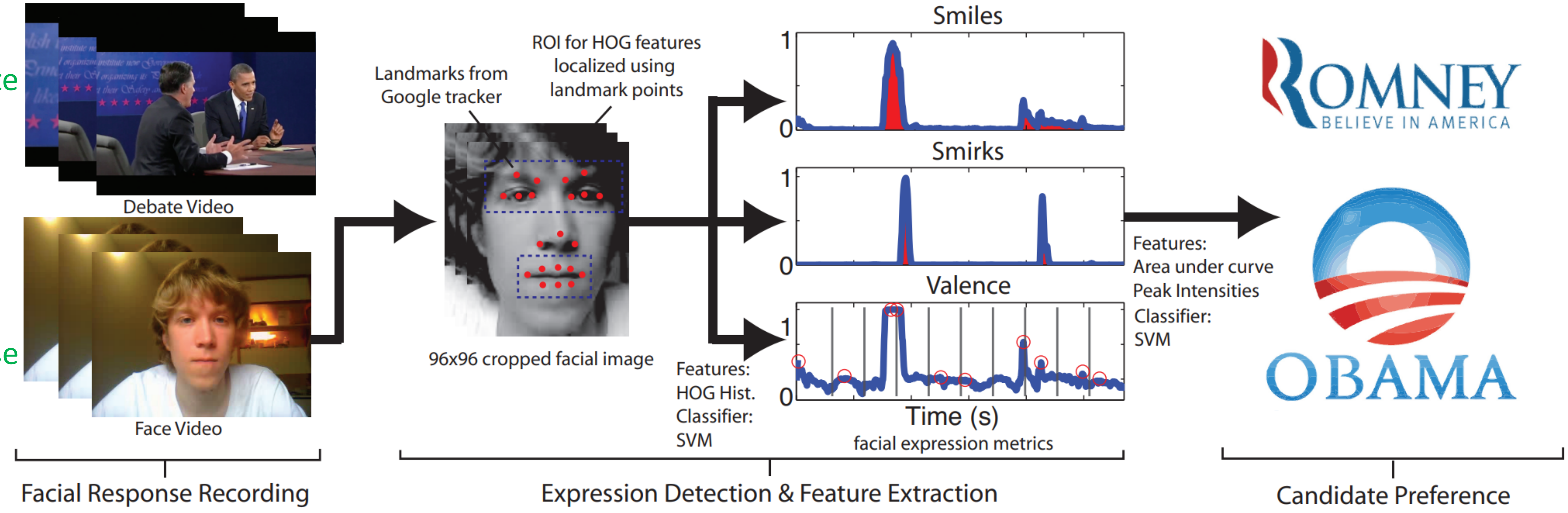
LOOKING AT PEOPLE: FACIAL EXPRESSION RECOGNITION

Image source: Her, 2013

Expression Recognition Reveals Political Preference

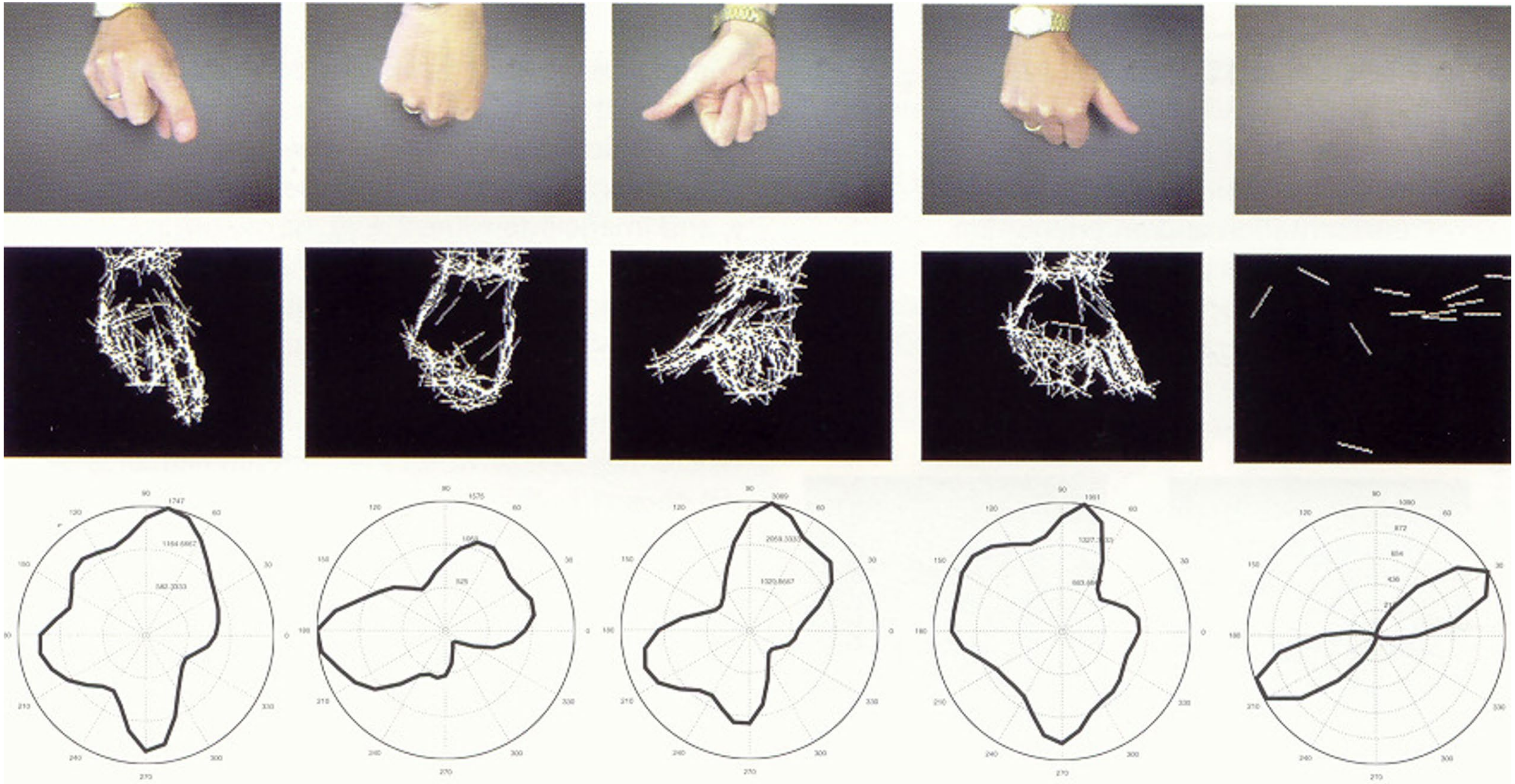
5 debate video clips

611 response videos

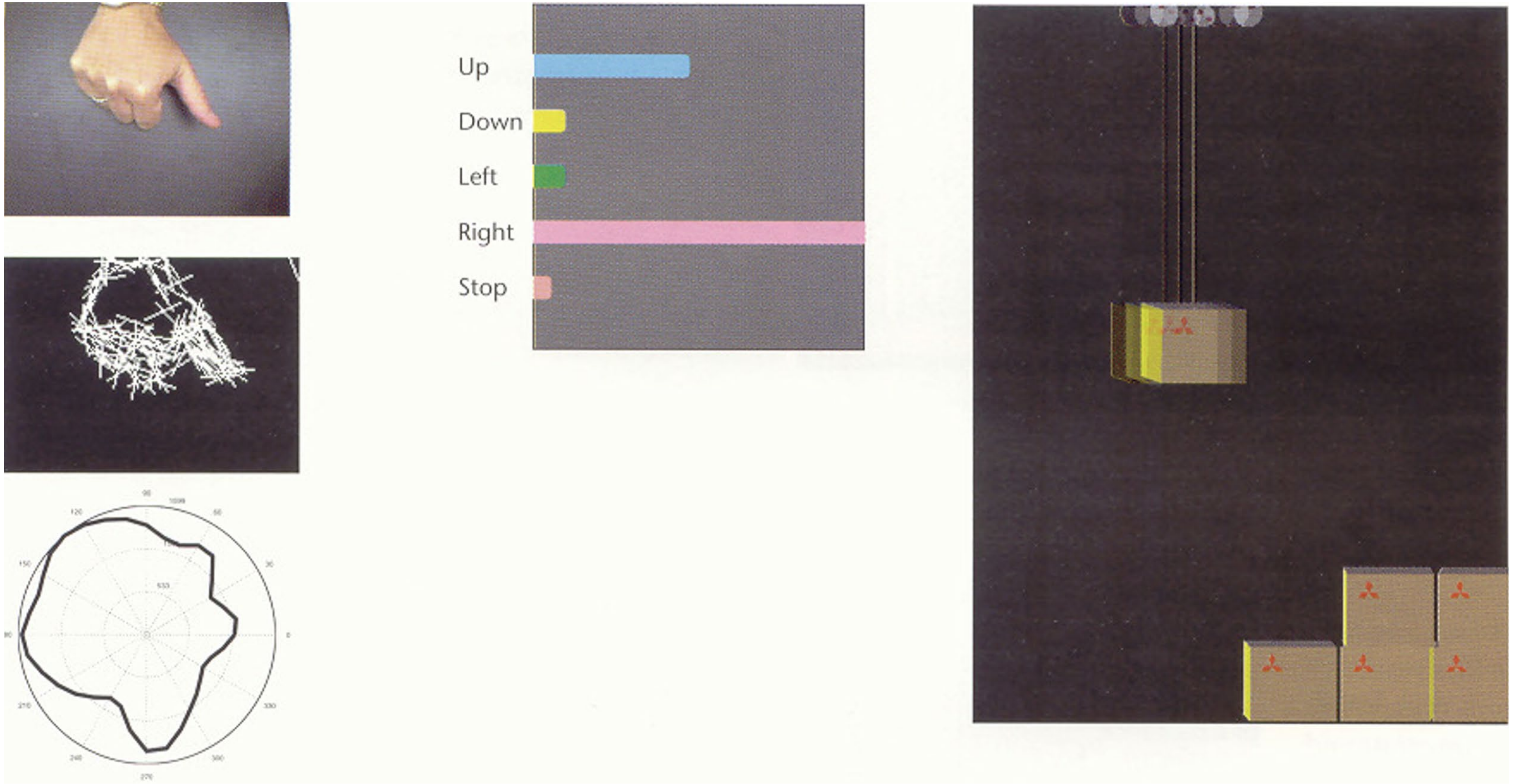


Accuracy ~ 80%

What are HOGs?



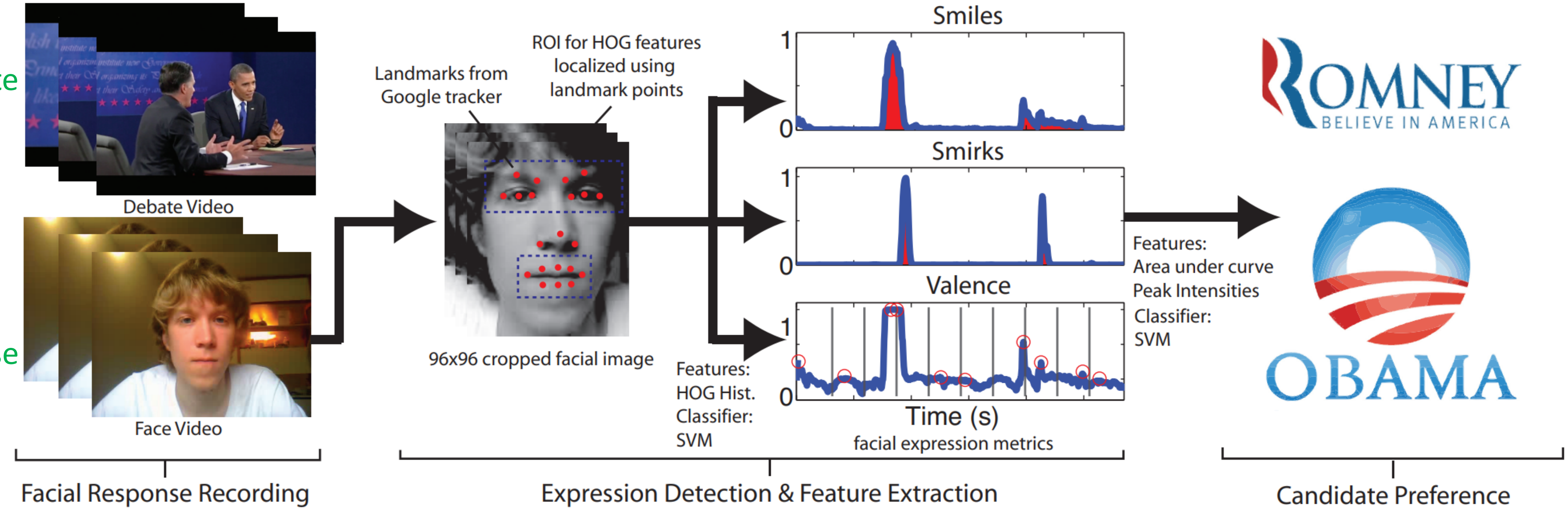
HOGs = Histograms of Gradients



Expression Recognition Reveals Political Preference

5 debate video clips

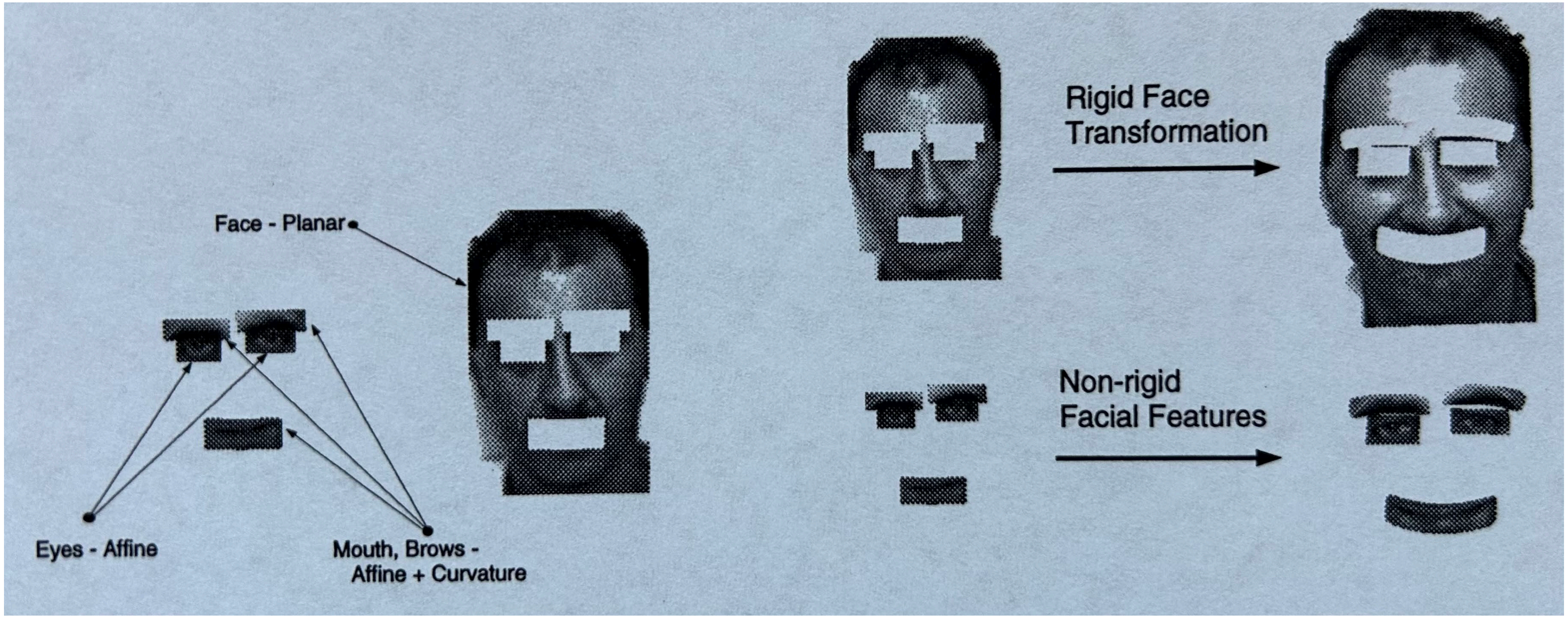
611 response videos

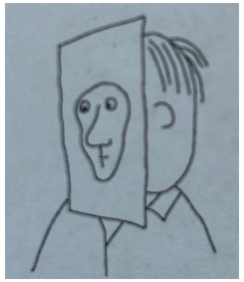


Accuracy ~ 80%

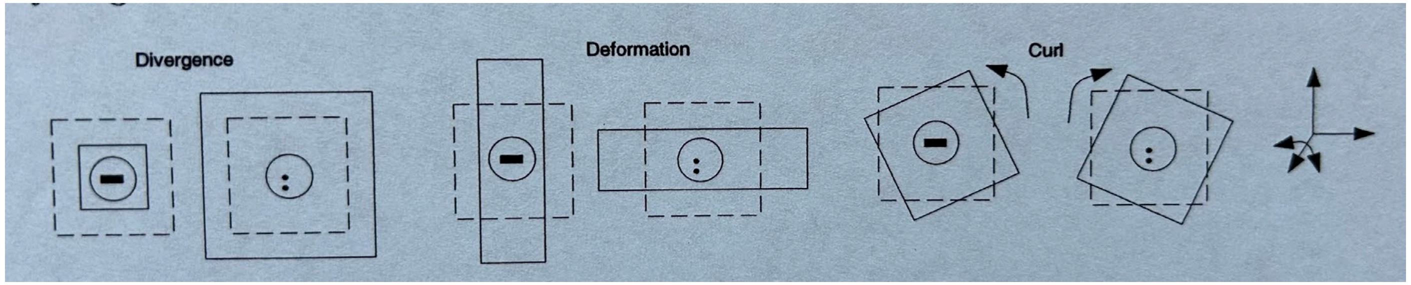
Image Credit: [McDuff et al., 2017]

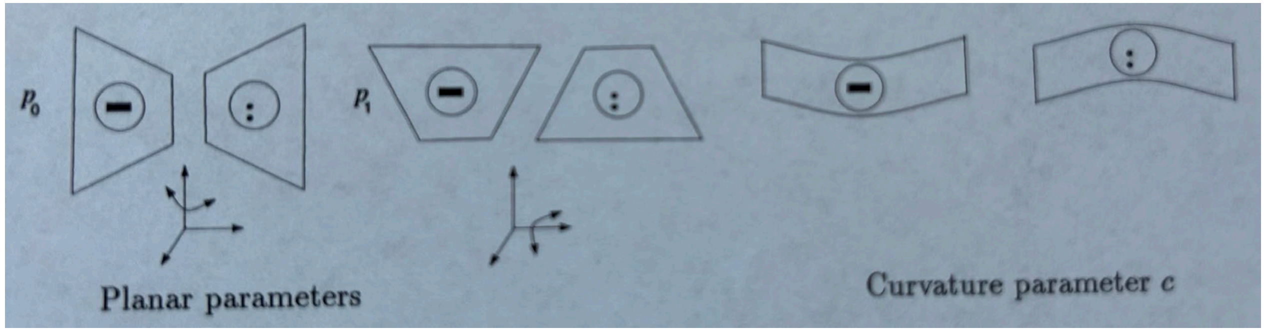
First Computer Vision Work on Facial Expression Analysis: Yaser Yacoob's PhD thesis & IJCV paper with Michael Black





First Computer Vision Work on Facial Expression Analysis: Yaser Yacoob's PhD thesis & IJCV paper with Michael Black





Optical flow (u,v) parameterized:

$u(x,y) = a_0 + a_1 x + a_2 y$
$v(x,y) = a_3 + a_4 x + a_5 y$

Divergence = $a_1 + a_5$
Curl = $-a_2 + a_4$
Deformation = $a_1 - a_5$

Yaw and pitch:
$u(x,y) = a_0 + a_1 x + a_2 y + p_0 x^2 + p_1 xy$
$v(x,y) = a_3 + a_4 x + a_5 y + p_1 y^2 + p_0 xy$

Mouth curvature:
$v(x,y) = a_3 + a_4 x + a_5 y + c x^2$

Credit: Black & Yacoob 1997

First Computer Vision Work on Facial Expression Analysis: Yaser Yacoob's PhD thesis & IJCV paper with Michael Black

Table 1: The cues for facial expressions as suggested by Ekman and Friesen.

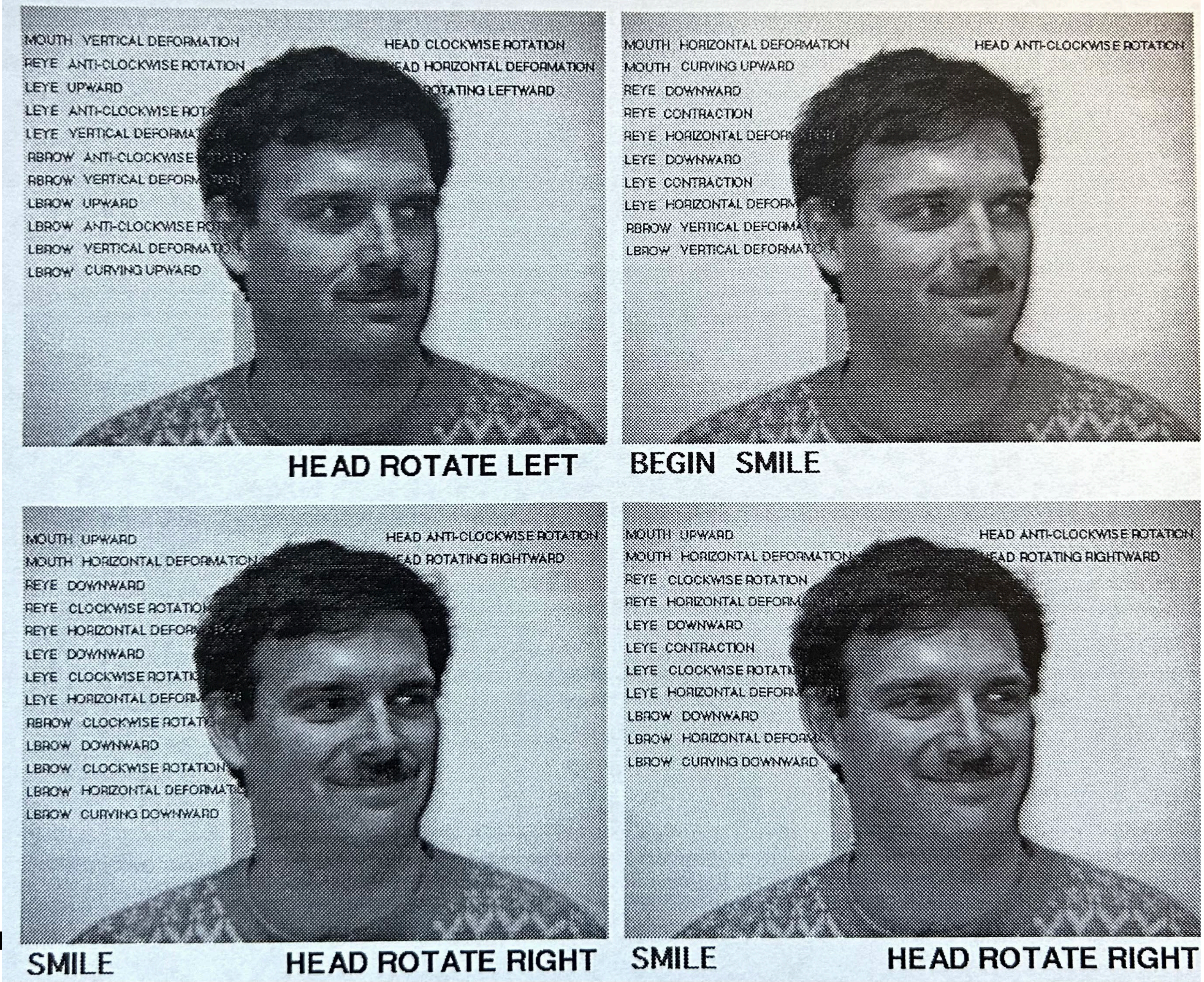
Emotion	Observed facial cues
Surprise	brows raised (curved and high) skin below brow stretched horizontal wrinkles across forehead eyelids opened and more of the white of the eye is visible jaw drops open without tension or stretching of the mouth
Fear	brows raised and drawn together forehead wrinkles drawn to the center upper eyelid is raised and lower eyelid is drawn up mouth is open lips are slightly tense or stretched and drawn back
Disgust	upper lip is raised lower lip is raised and pushed up to upper lip or is lowered nose is wrinkled cheeks are raised lines below the lower lid, lid is pushed up but not tense brows are lowered, lowering the upper lid
Anger	brows lowered and drawn together vertical lines appear between brows lower lid is tensed and may or may not be raised upper lid is tense and may or may not be lowered due to brows' action eyes have a hard stare and may have a bulging appearance lips are either pressed firmly together with corners straight or down or open, tensed in a squarish shape nostrils may be dilated (could occur in sadness too) unambiguous only if registered in all three facial areas
Happiness	corners of lips are drawn back and up mouth may or may not be parted with teeth exposed or not a wrinkle runs down from the nose to the outer edge beyond lip corners cheeks are raised lower eyelid shows wrinkles below it, and may be raised but not tense crow's-feet wrinkles go outward from the outer corners of the eyes
Sadness	inner corners of eyebrows are drawn up skin below the eyebrow is triangulated, with inner corner up upper lid inner corner is raised corners of the lips are drawn or lip is trembling

Table 5: The rules for classifying facial expressions (B=beginning, E=ending).

Expr.	B/E	Satisfactory actions
Anger	B	inward lowering of brows and mouth contraction
Anger	E	outward raising of brows and mouth expansion
Disgust	B	mouth horizontal expansion and lowering of brows
Disgust	E	mouth contraction and raising of brows
Happiness	B	upward curving of mouth and expansion or horizontal deformation
Happiness	E	downward curving of mouth and contraction or horizontal deformation
Surprise	B	raising brows and vertical expansion of mouth
Surprise	E	lowering brows and vertical contraction of mouth
Sadness	B	downward curving of mouth and upward-inward motion in inner parts of brows
Sadness	E	upward curving of mouth and downward-outward motion in inner parts of brows
Fear	B	expansion of mouth and raising-inwards inner parts of brows
Fear	E	contraction of mouth and lowering inner parts of brows

Credit: Black & Yacoob 1997

BOSTON UNIVERSITY



I'll show the video from Yaser's phd defense in class

Credit: Black & Yacoob 1997

Facial Action Coding System: FACS

Wikipedia:

"In 1969, Hjortsjö devised the first system to taxonomize human facial movements by their appearances on the face, along with a description of each change in appearance caused by the action of each facial muscle. American psychologists Paul Ekman and Wallace Friesen later in 1978 formalized the descriptions as the Facial Action Coding System."

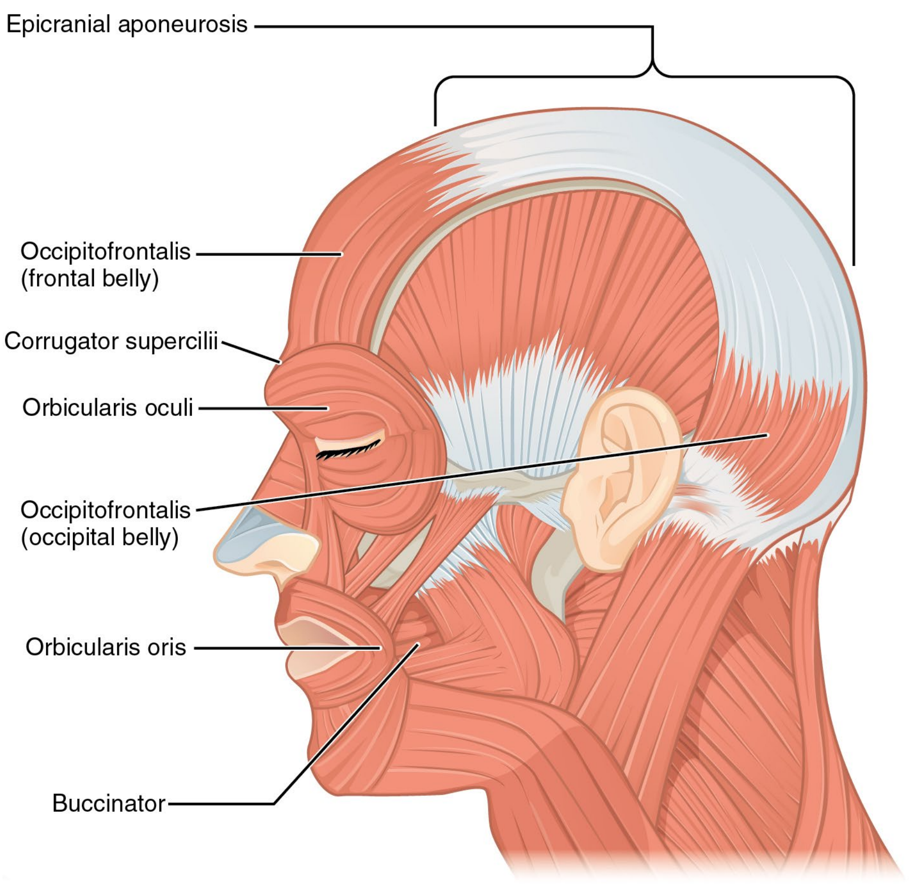


Image Credit - Wikimedia: By CNX Anatomy 213

Expression Recognition based on AUs

Emotion	Action units
Anger	4+5+7+23
Contempt	R12A+R14A
Disgust	9+15+17
Fear	1+2+4+5+7+20+26
Happiness	6+12
Sadness	1+4+15
Surprise	1+2+5B+26

AU number	FACS name	Muscular basis
0	Neutral face	
1	Inner brow raiser	*frontalis* (*pars medialis*)
2	Outer brow raiser	*frontalis* (*pars lateralis*)
4	Brow lowerer	*depressor glabellae*, *depressor supercilii*, *corrugator supercilii*
5	Upper lid raiser	*levator palpebrae superioris*, superior tarsal muscle
6	Cheek raiser	*orbicularis oculi* (*pars orbitalis*)
7	Lid tightener	*orbicularis oculi* (*pars palpebralis*)
8	Lips toward each other	*orbicularis oris*
9	Nose wrinkler	*levator labii superioris alaeque nasi*
10	Upper lip raiser	*levator labii superioris*, *caput infraorbitalis*
11	Nasolabial deepener	*zygomaticus minor*
12	Lip corner puller	*zygomaticus major*
13	Sharp lip puller	*levator anguli oris* (also known as *caninus*)
14	Dimpler	*buccinator*
15	Lip corner depressor	*depressor anguli oris* (also known as *triangularis*)
16	Lower lip depressor	*depressor labii inferioris*
17	Chin raiser	*mentalis*
18	Lip pucker	*incisivii labii superioris* and *incisivii labii inferioris*
19	Tongue show	
20	Lip stretcher	*risorius* with *platysma*
21	Neck tightener	*platysma]*
22	Lip funneler	*orbicularis oris*
23	Lip tightener	*orbicularis oris*
24	Lip pressor	*orbicularis oris*
25	Lips part	*depressor labii inferioris*, or relaxation of *mentalis* or *orbicularis oris*
26	Jaw drop	*masseter*; relaxed *temporalis* and internal *pterygoid*
27	Mouth stretch	*pterygoids*, *digastric*
28	Lip suck	*orbicularis oris*

FFDEX SDK: A Cross-Platform RealTime Multi-Face Expression Recognition Toolkit

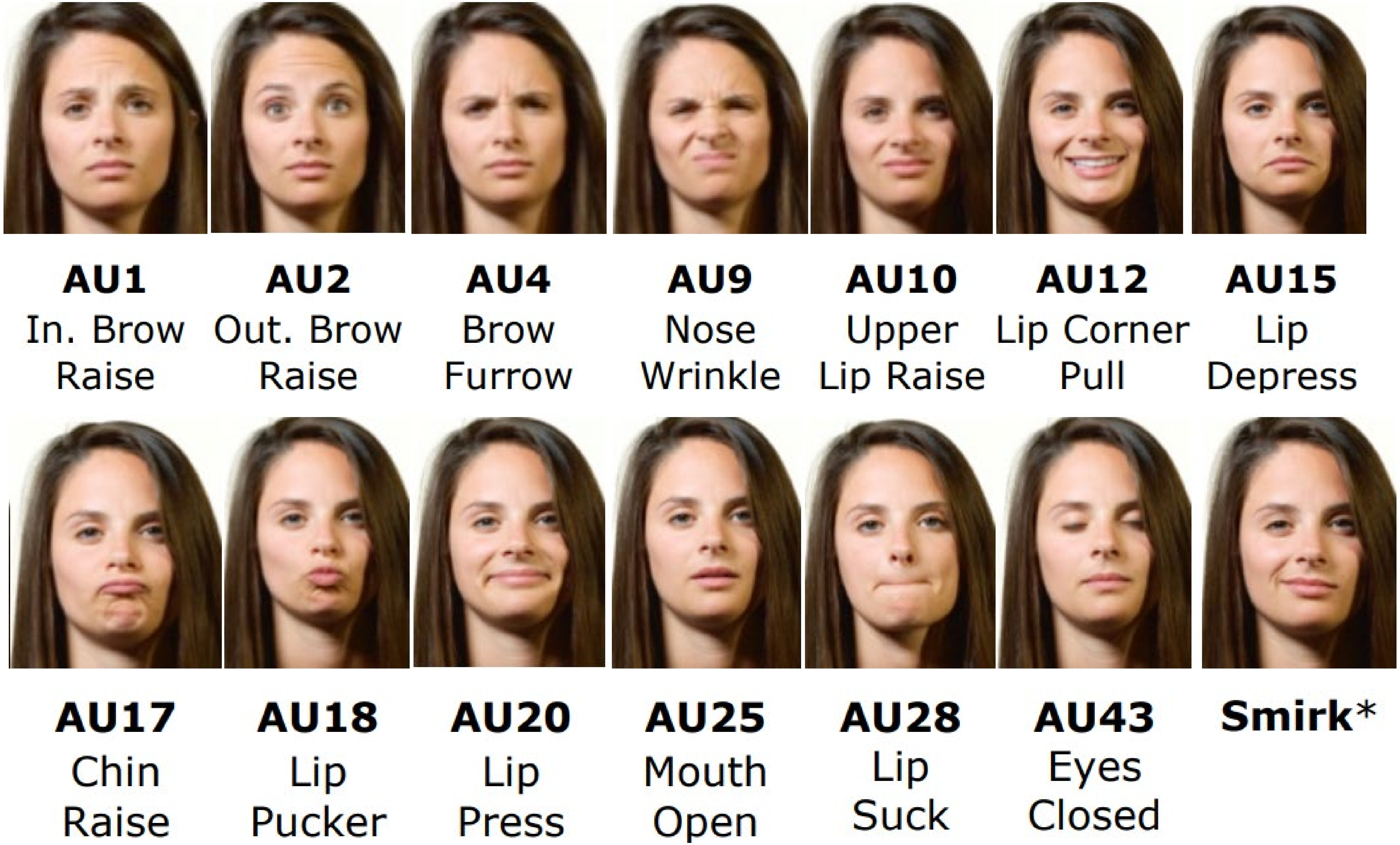


Image Credit: [McDuff et al., 2016]

OpenFace 2.0

Runs in real time
Source code available

68 facial landmarks

SVR-HOG works better for expression analysis than deep methods

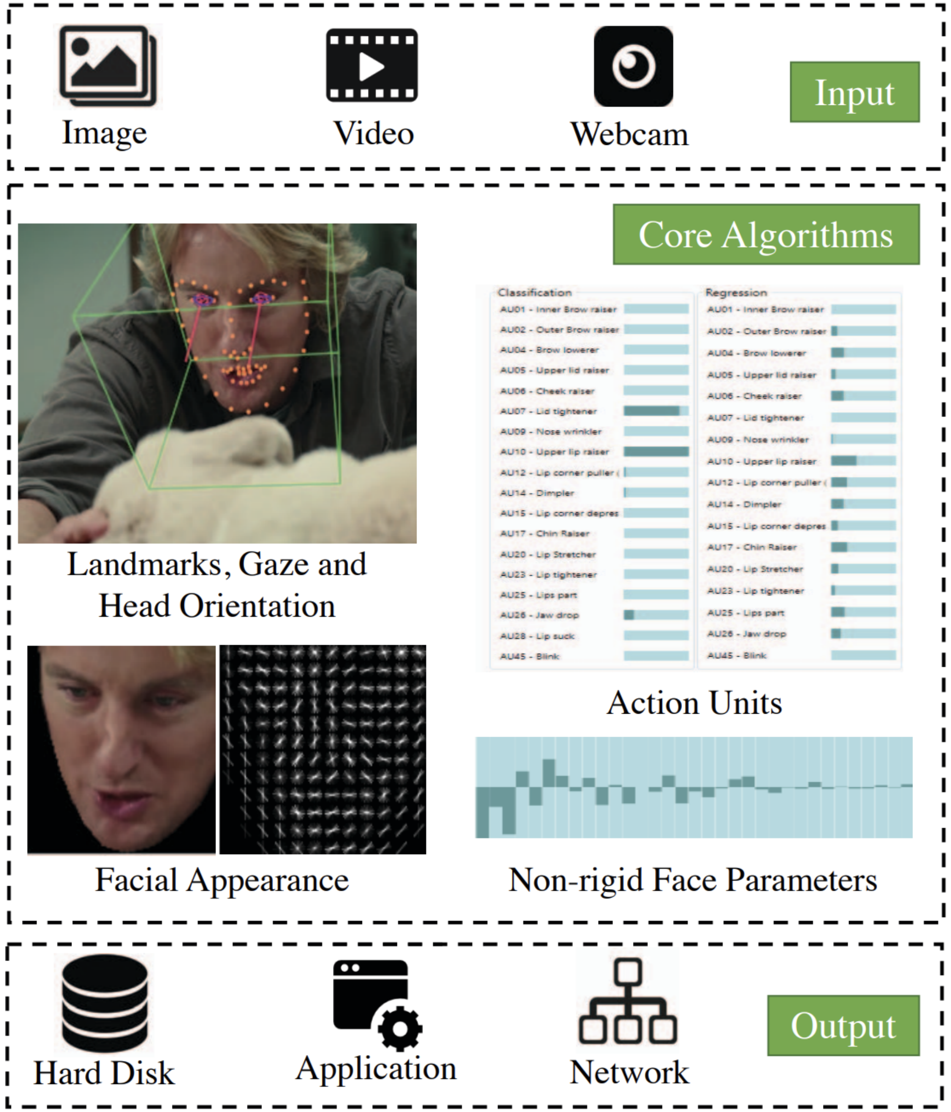


Fig. 1: OpenFace 2.0 is a framework that implements modern facial behavior analysis algorithms including: facial landmark detection, head pose tracking, eye gaze and facial action unit recognition.

Credit: [Baltrusaitis et al., 2018]

Learning Objectives of this Lecture

- Be able to describe computer vision tasks that involve analysis of the humans in images: Person detection, human pose estimation, face detection, facial feature detection, facial expression recognition, face recognition
- Understand the difference between face verification and face identification
- Know biometrics terminology (query, probe, gallery, template, distractors)
- Know mechanism for training and testing face recognition models
- Can describe the VGG-16 architecture
- Know what a HOG feature is
- Know about the FACS
- Know some benchmark datasets for face recognition
- Can discuss the challenges and limitations of face recognition systems (aging, pose) and how they are being addressed by ongoing research
- Can explain how to convert a regression problem into a coarse-to-fine classification problem
- Know about AI for non-face biometrics, e.g., Amazon's palm recognition, ear recognition, gait recognition
- Know about fake ID creation/detection
- Be able to discuss ethical issues with biometrics
- Know about methods for facial expression recognition