Face Recognition & Other Biometrics

CS 640 AI, Fall 2024

Mahir Patel, 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

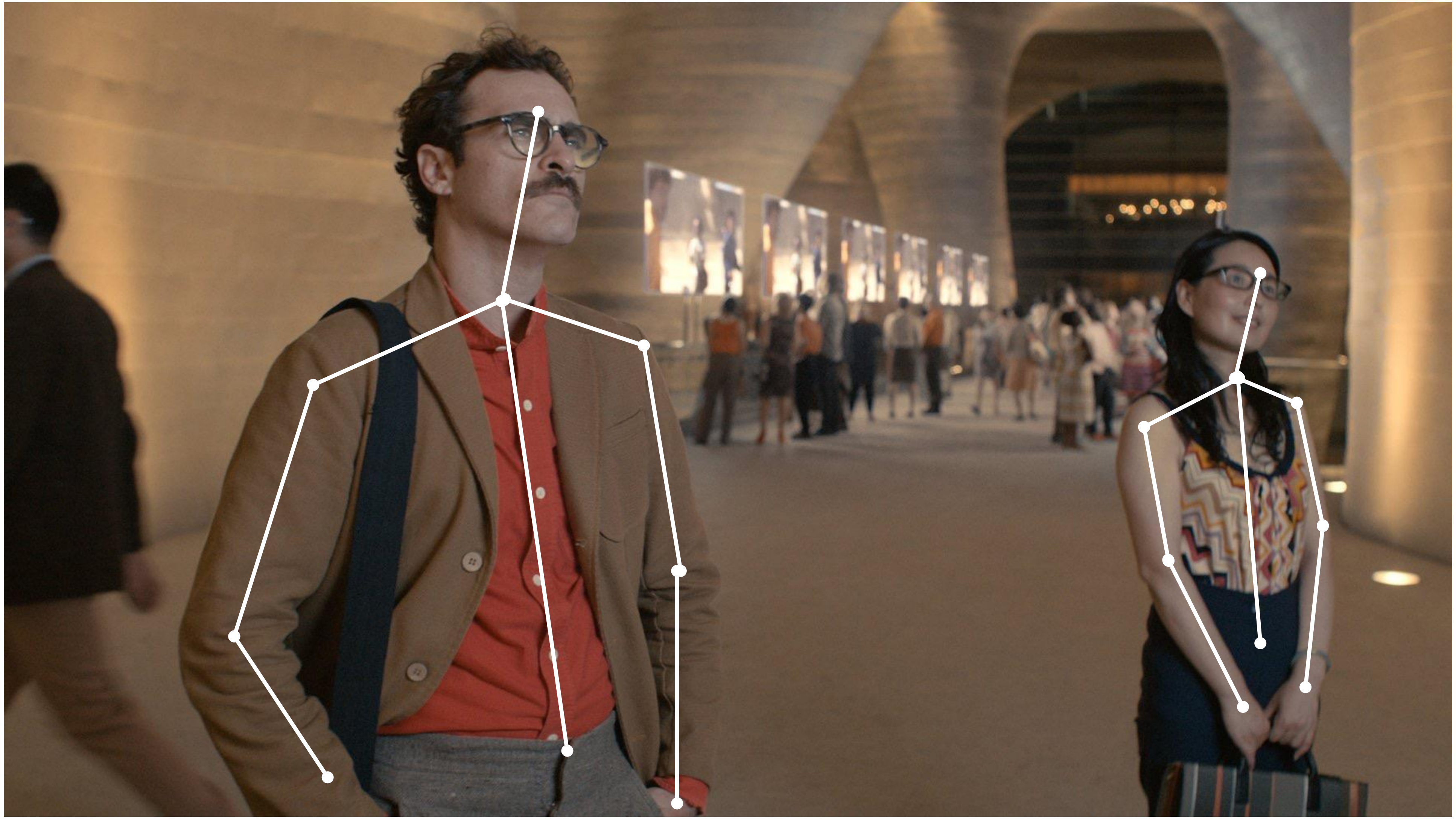


Image source: Her, 2013

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?

Query Image

=

Reference Image

Yes

"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

Gallery of Known Subjects

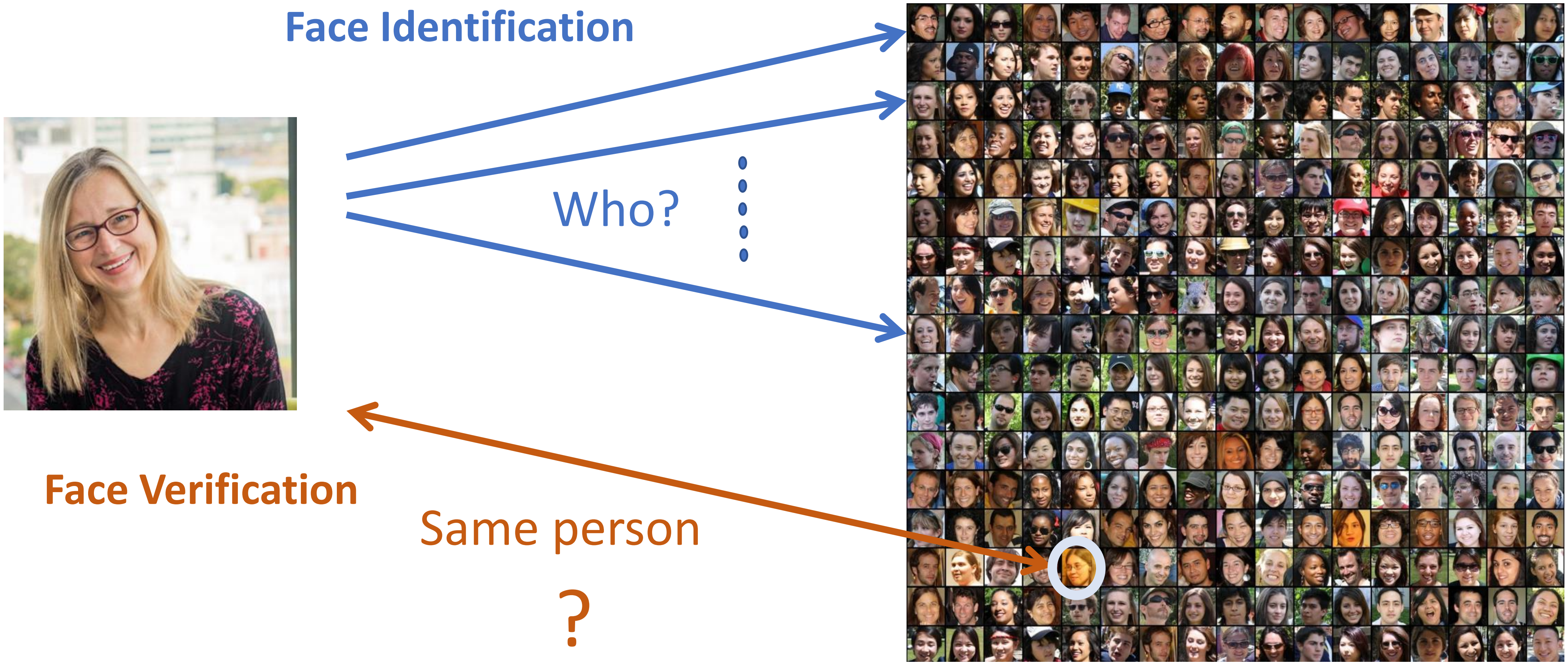


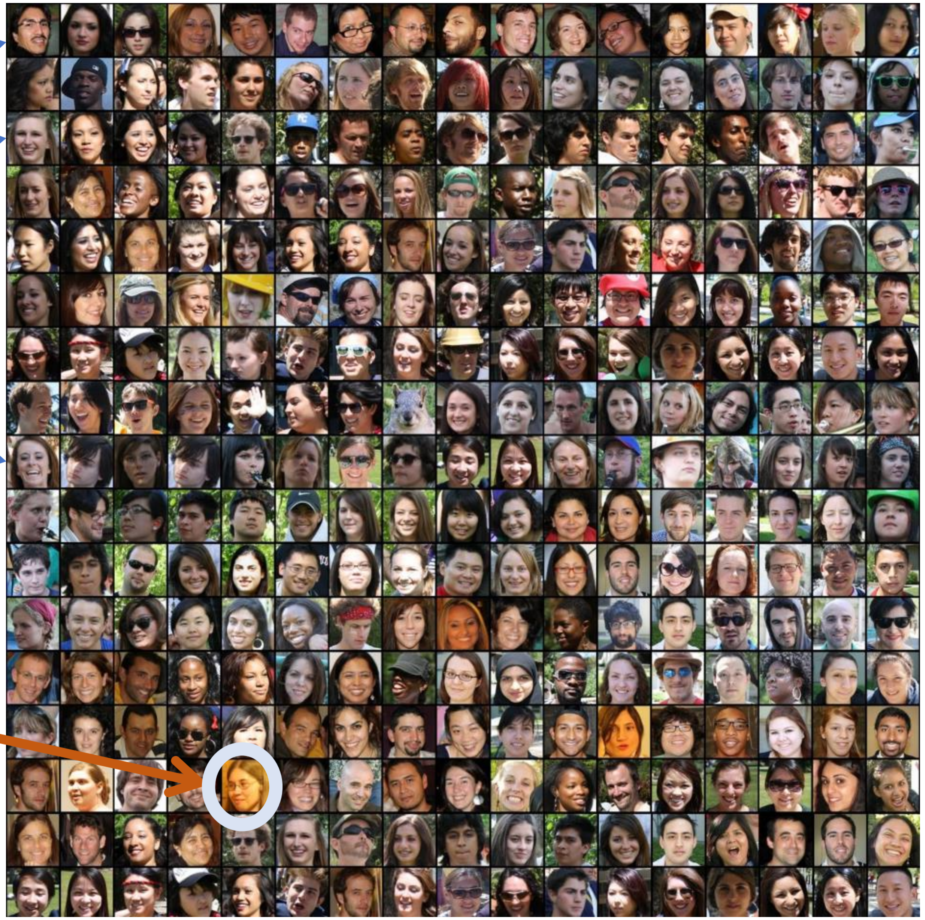
Image credit: Q. Orozco-Ramirez

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

Here:
One picture per person

Better:
Multiple pictures per person

Gallery of Known Subjects

Image credit: Q. Orozco-Ramirez

Face Recognition

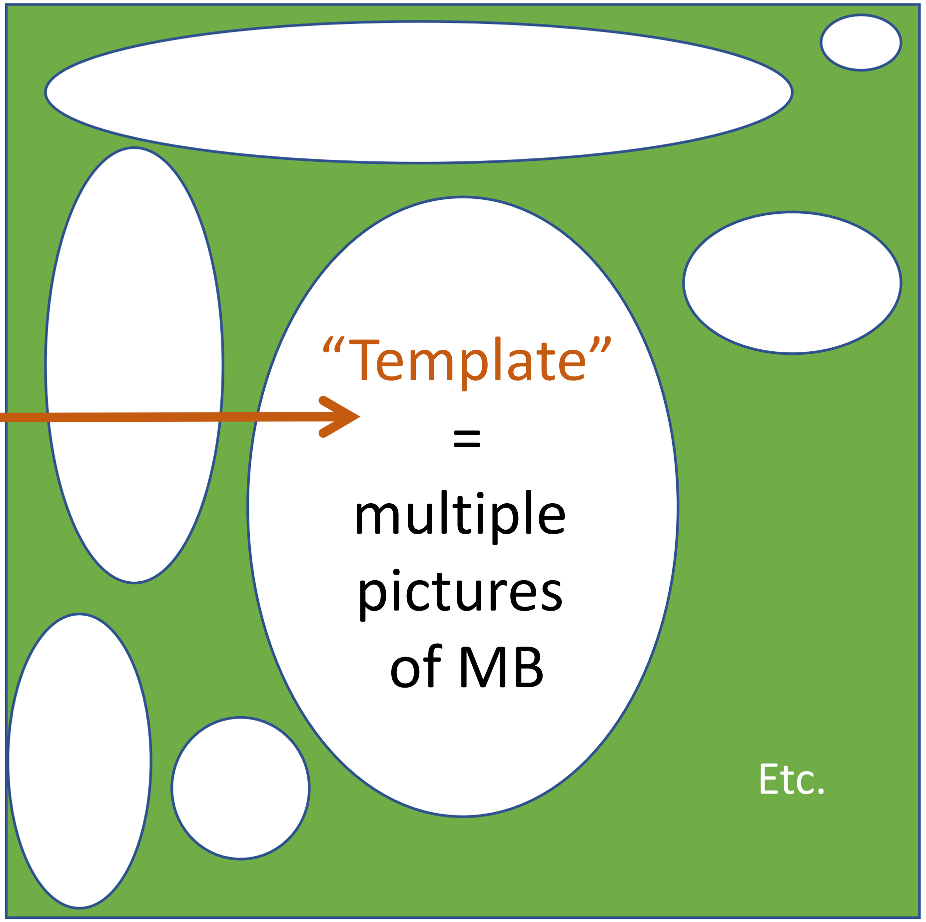
Gallery of Known Subjects

Face Recognition

Probe or **Query** of Unknown Subject

one-to-one
similarity

Gallery of Known Subjects



How does Face Recognition Technology Work?

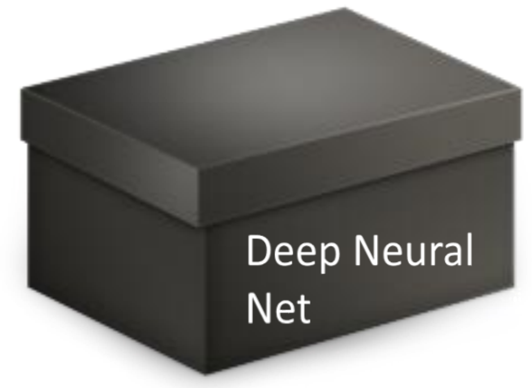
2012 Revolution in Computer Vision:

Deep Neural Networks

Deep Neural Networks used for Face Recognition

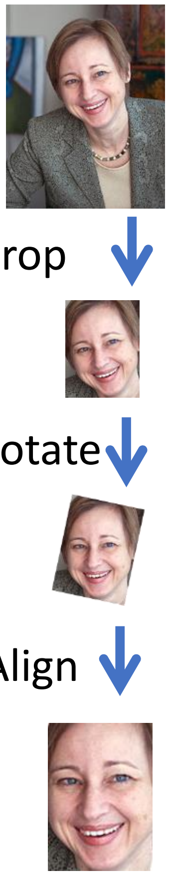
1. Network architecture
2. Training
3. Testing = “use mode”

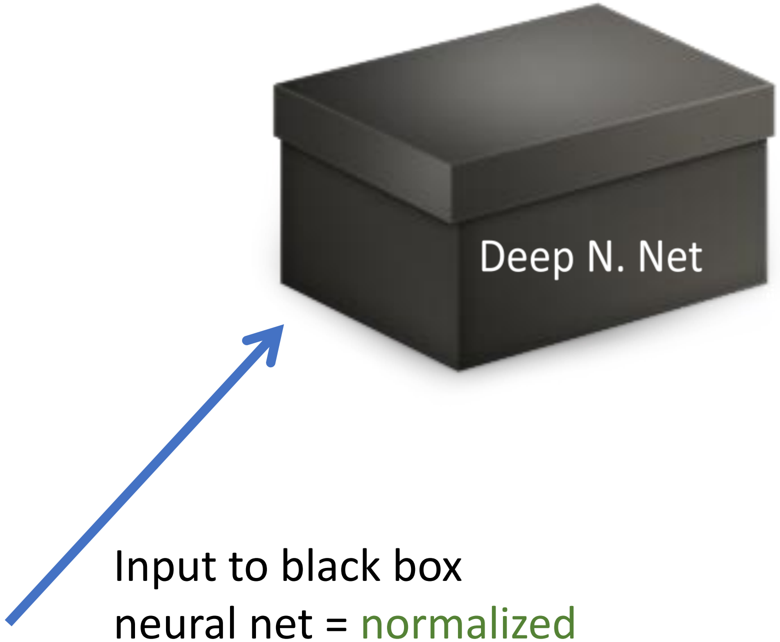
Network Architecture



Training a Face Recognition Network

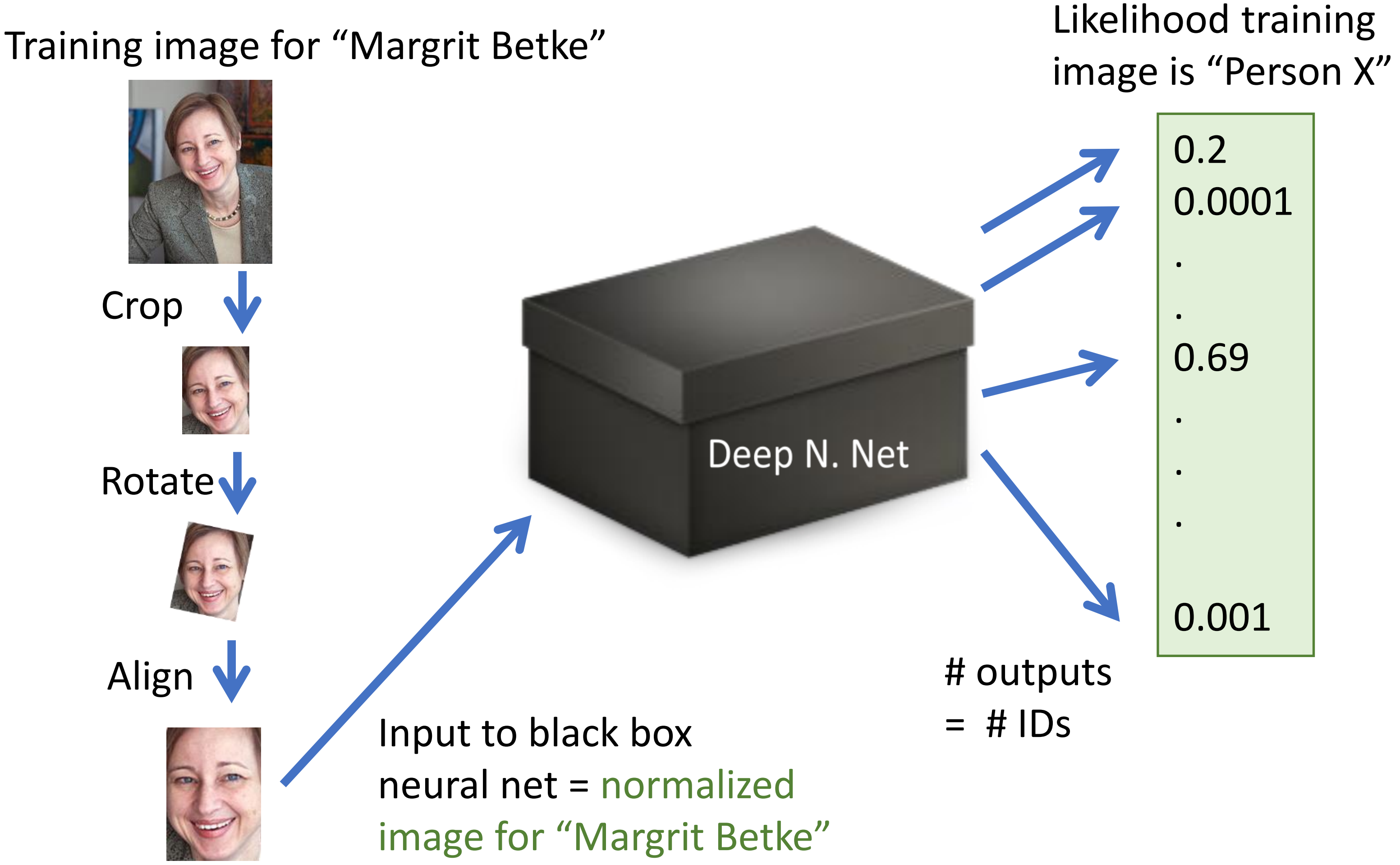
Training image for “Margrit Betke”



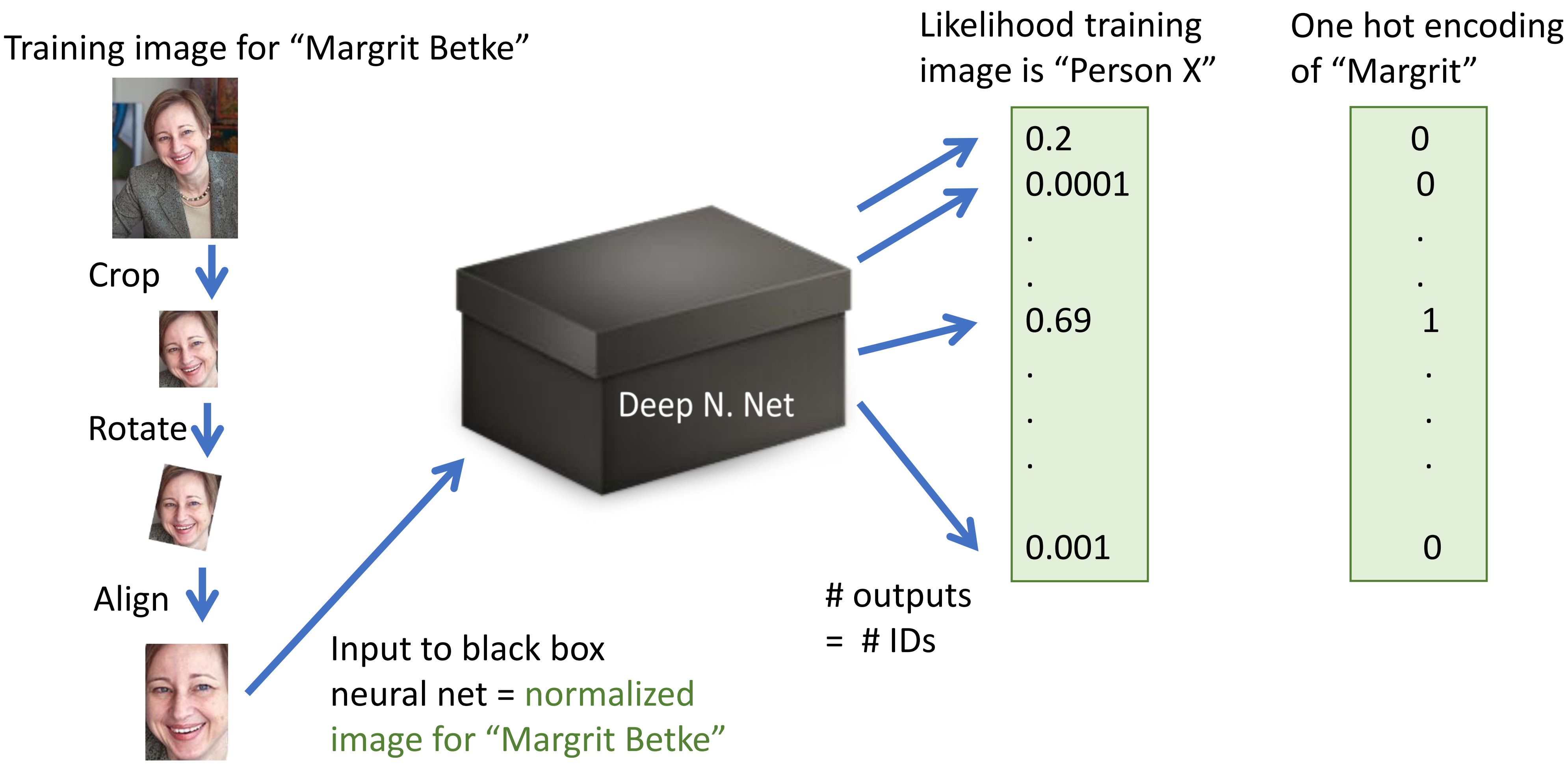


Input to black box neural net = normalized image for “Margrit Betke”

Training a Face Recognition Network



Training a Face Recognition Network



Training a Face Recognition Network

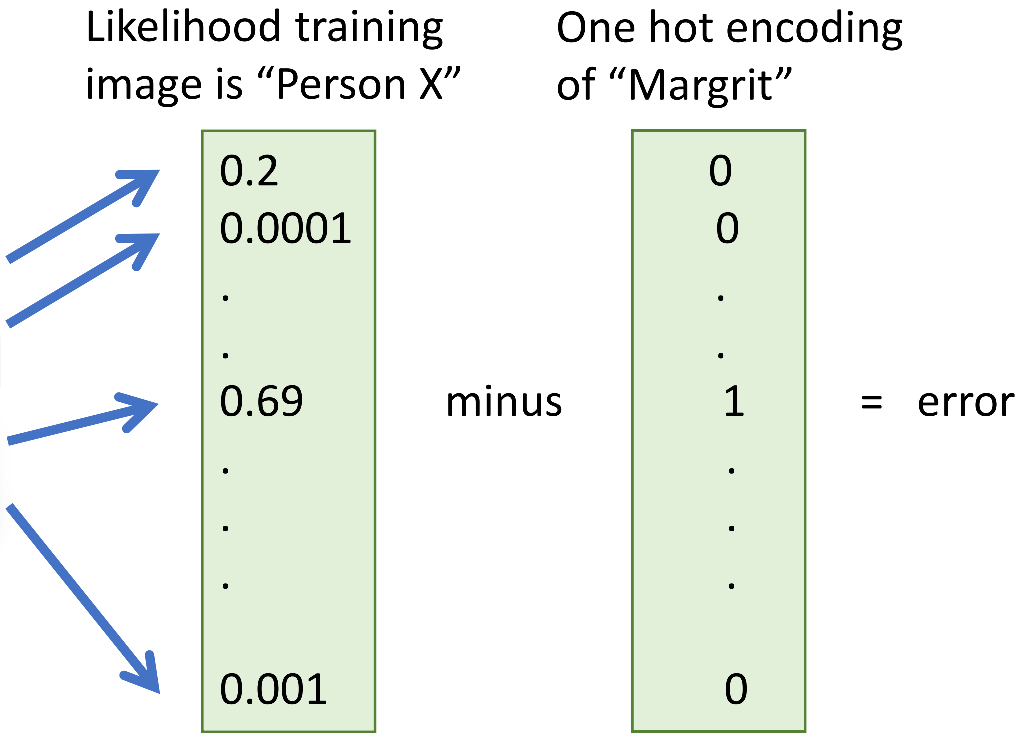
Training image for "Margrit Betke"

Crop

Rotate

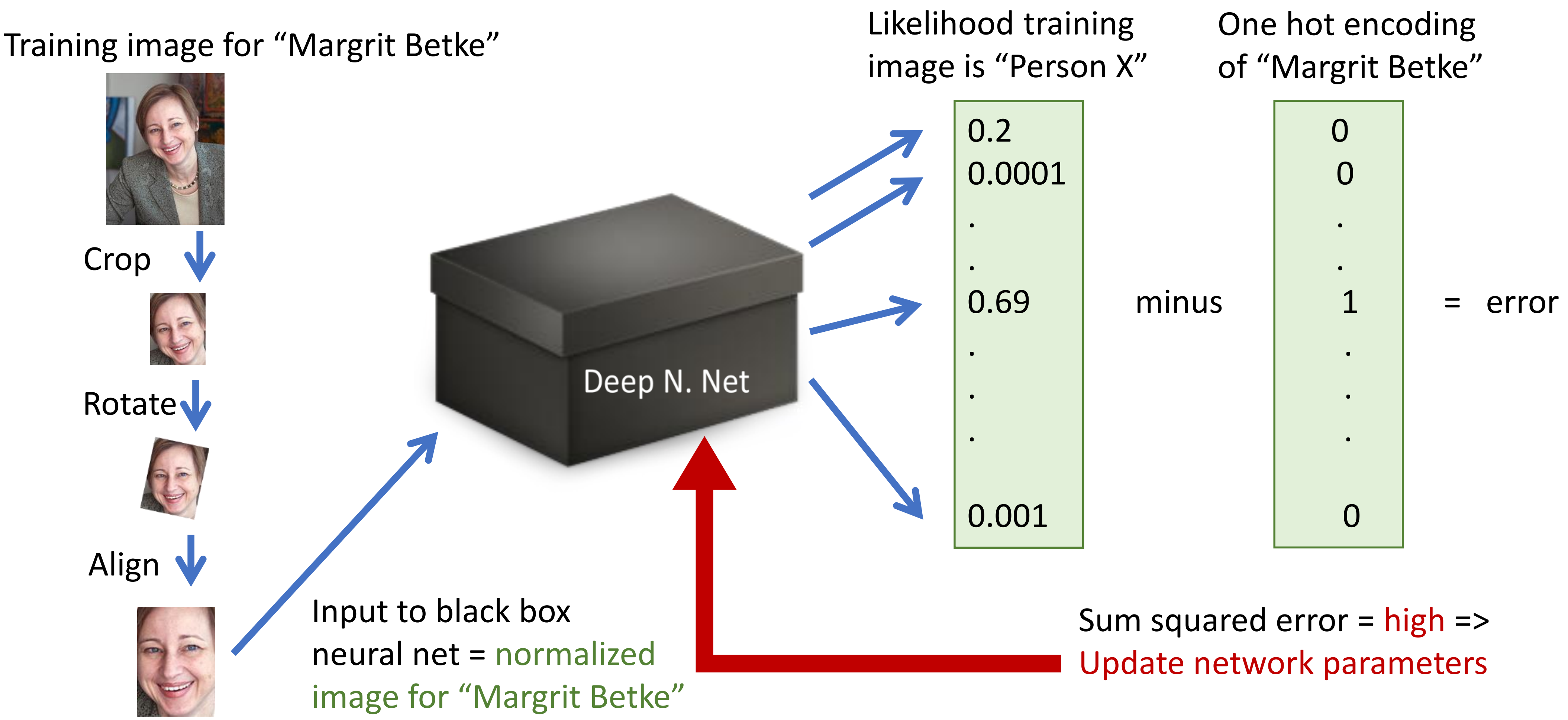
Align

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

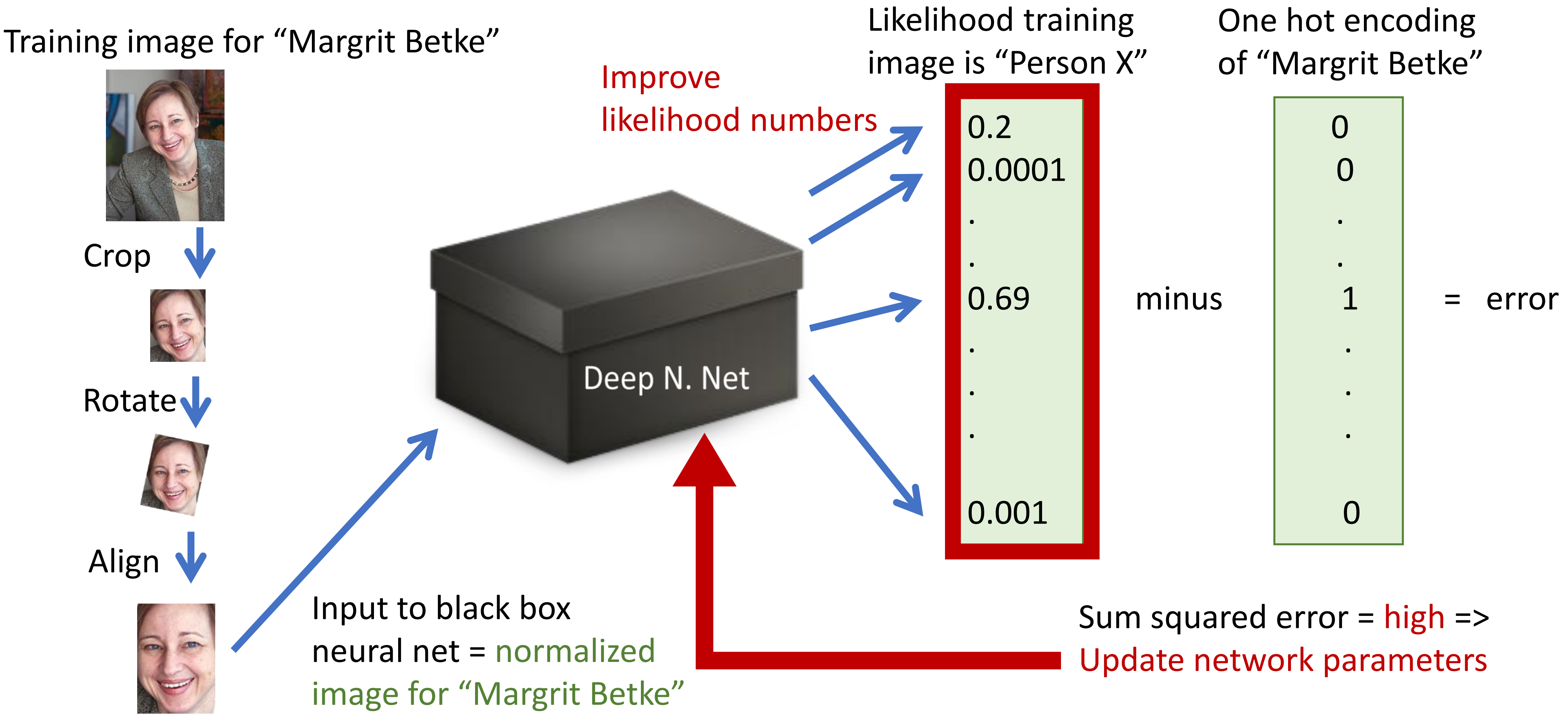


Sum squared error = Measure of training need

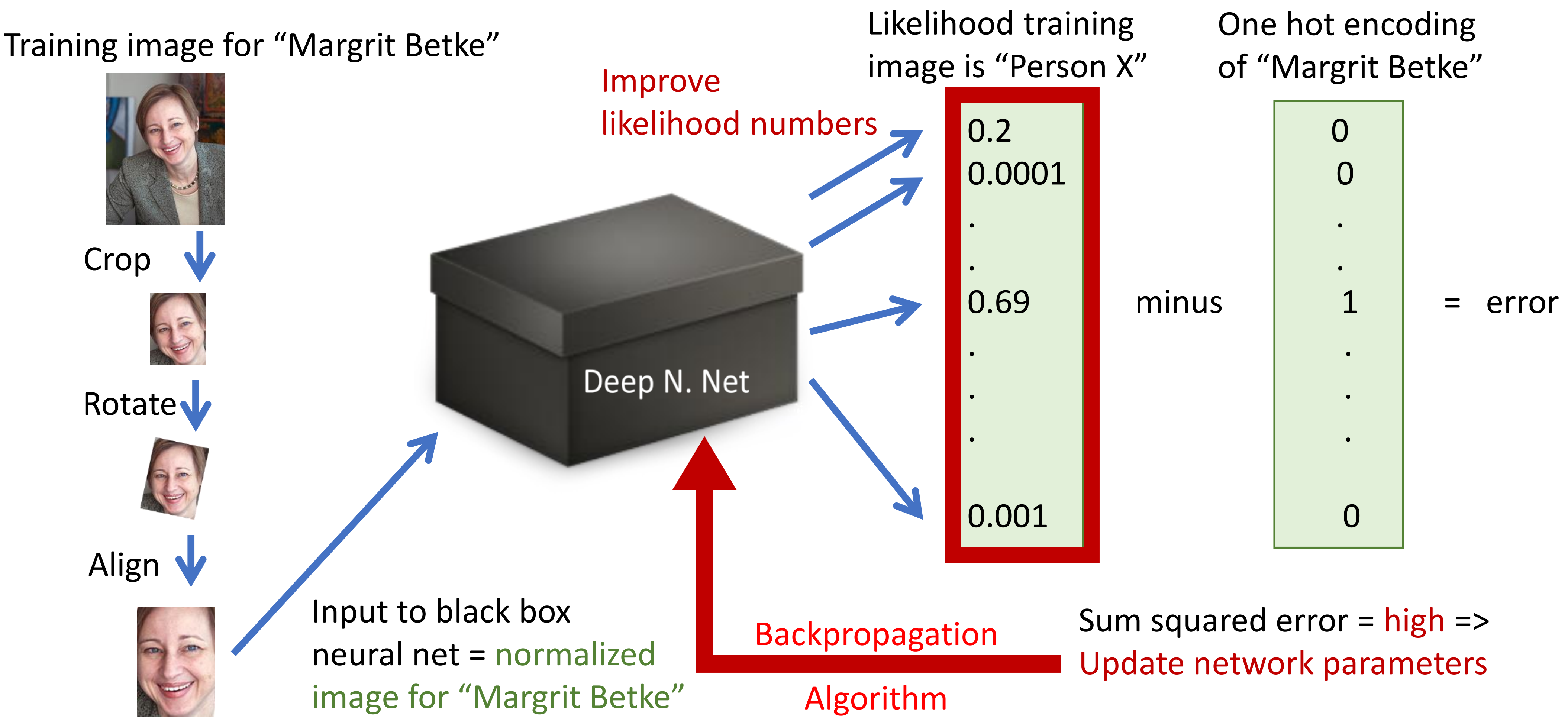
Training a Face Recognition Network



Training a Face Recognition Network



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”

After numerous network updates:

Likelihood training image is “Person X”

0.0002
0.0001
.
.
0.99
.
.
.
0.0001

minus

One hot encoding of “Margrit Betke”

0
0
.
.
1
.
.
.
0

= error

Sum squared error = low =>
Done with training!

Trained network

Training image for “Margrit Betke”

Crop

Rotate

Align

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

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”

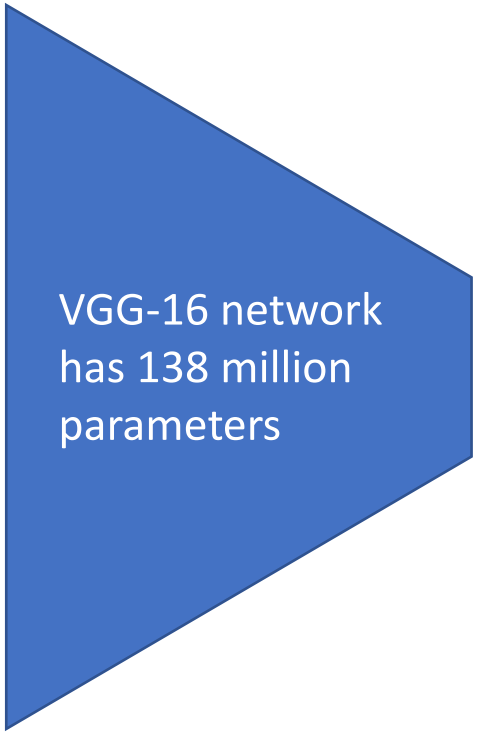
Network Architecture: VGG-16

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

Output:
4,096
numbers

Network Architecture: VGG-16

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



Output:
4,096
numbers

Network Architecture of VGG-16: Two Visualizations

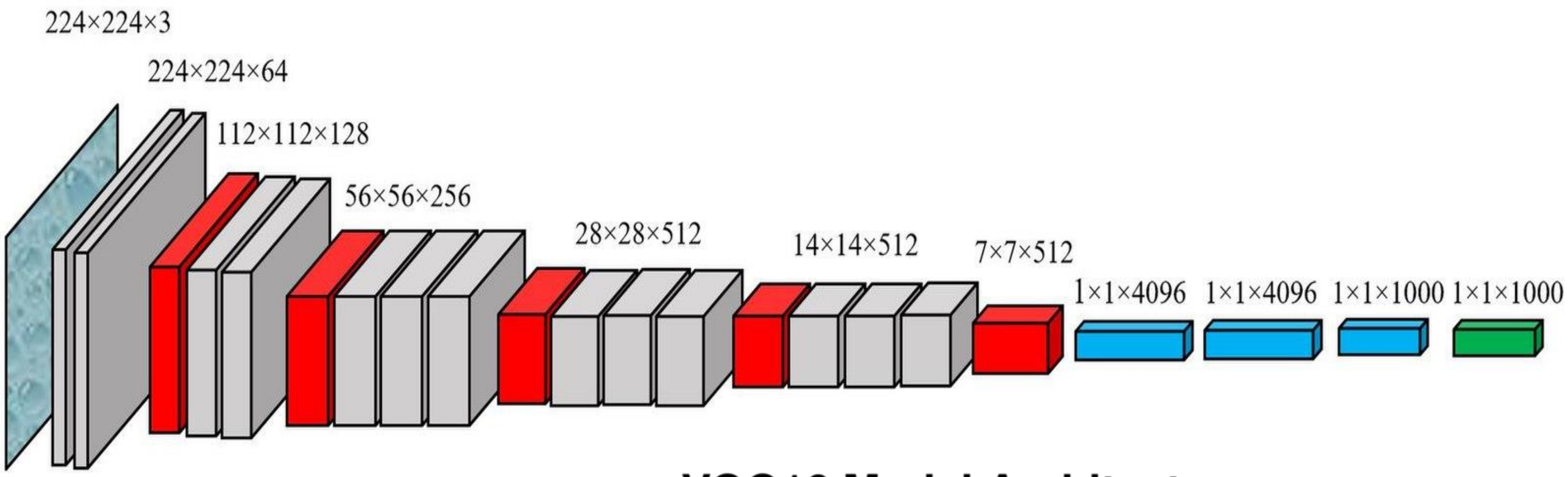


Image credit:
Hamidreza Habibollahi-Najaf-Abadi

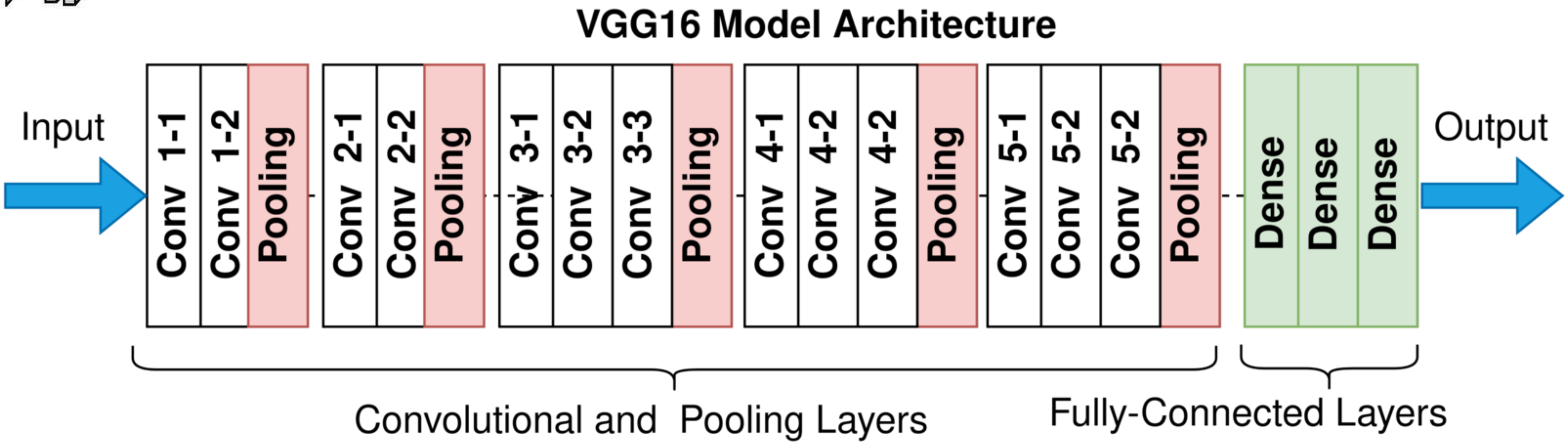


Image credit:
Gorla Praveen

Let’s look at how to use the trained network

Training image for “Margrit Betke”

Crop

Rotate

Align

normalized image for “Margrit Betke”
= Red Green Blue values of
224 x 224 pixels

3 x
224 x 224
numbers

e.g., VGG* Network

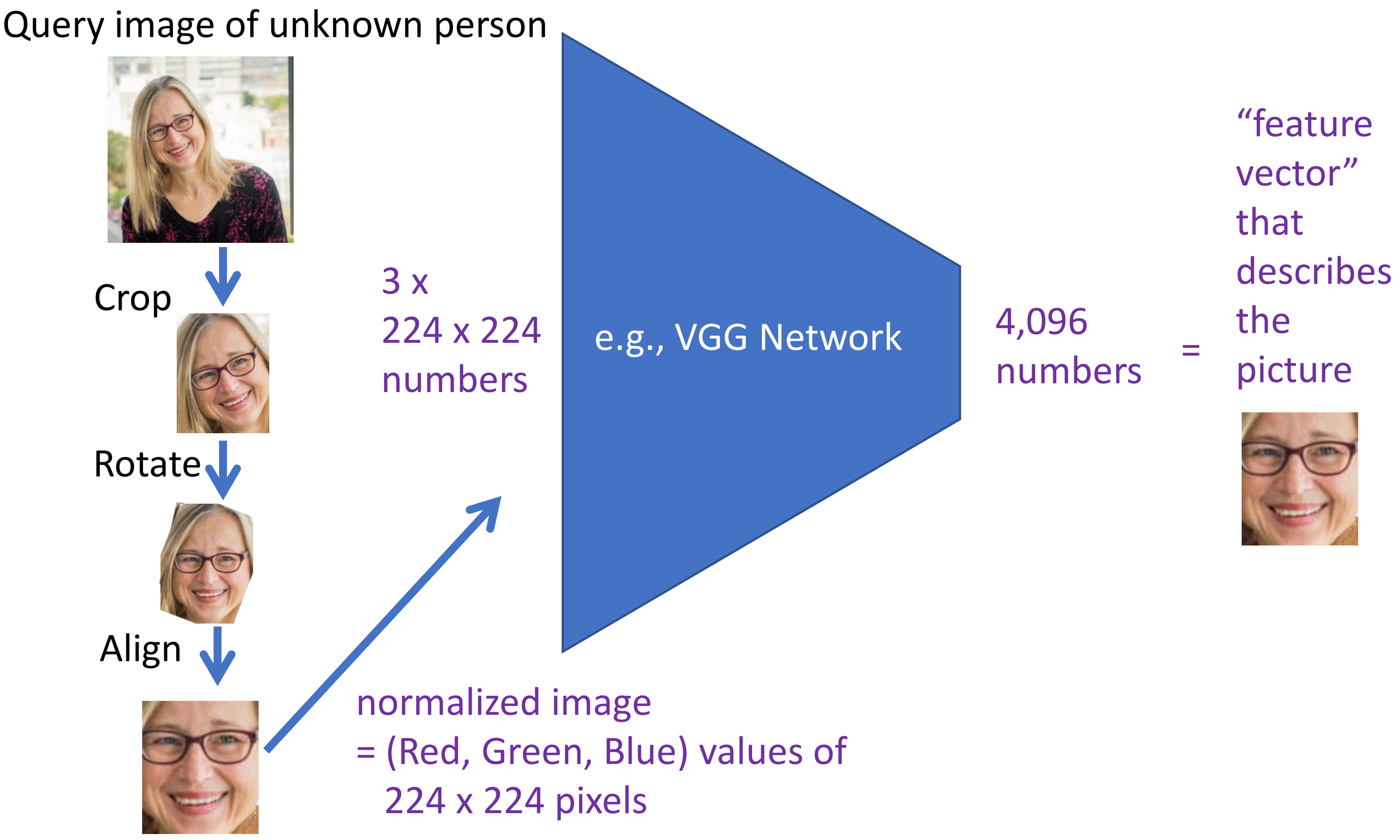
4,096
numbers

=

“feature
vector”
that
describes
my
picture

* 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

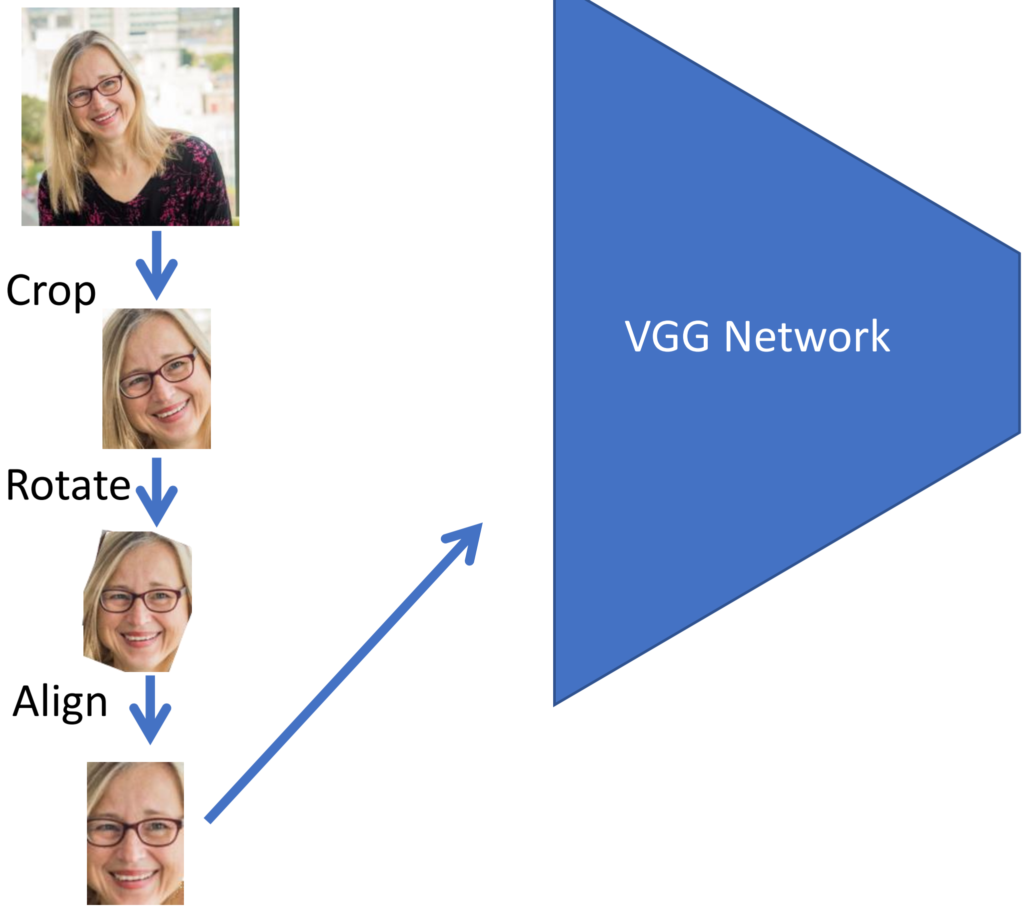
=

“feature
vector”
that
describes
the
picture

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

Face Verification in “Use Mode”

Query image of unknown person



“feature vector” that describes the picture

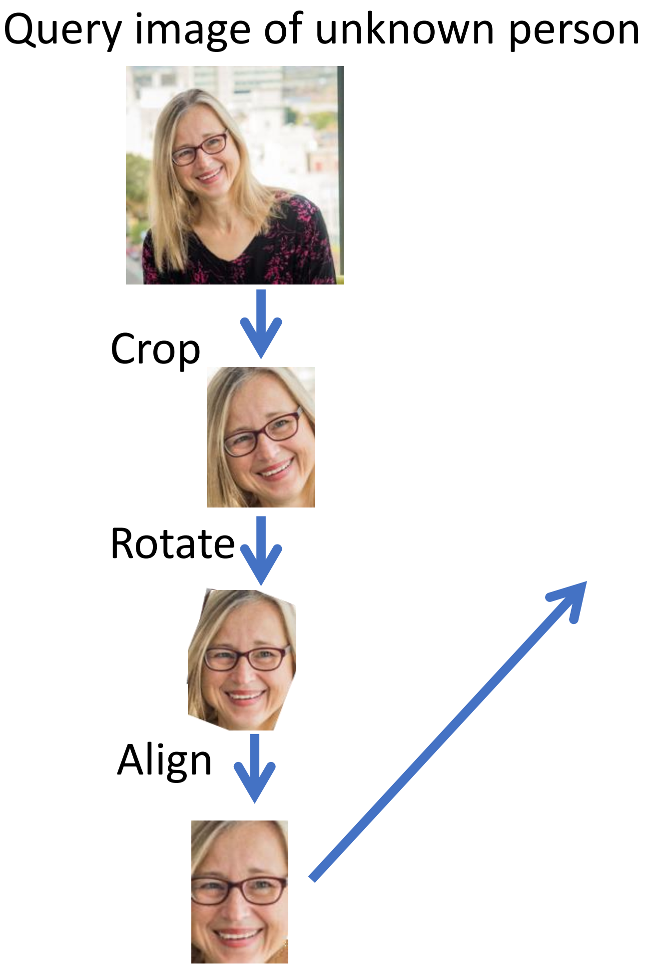
minus

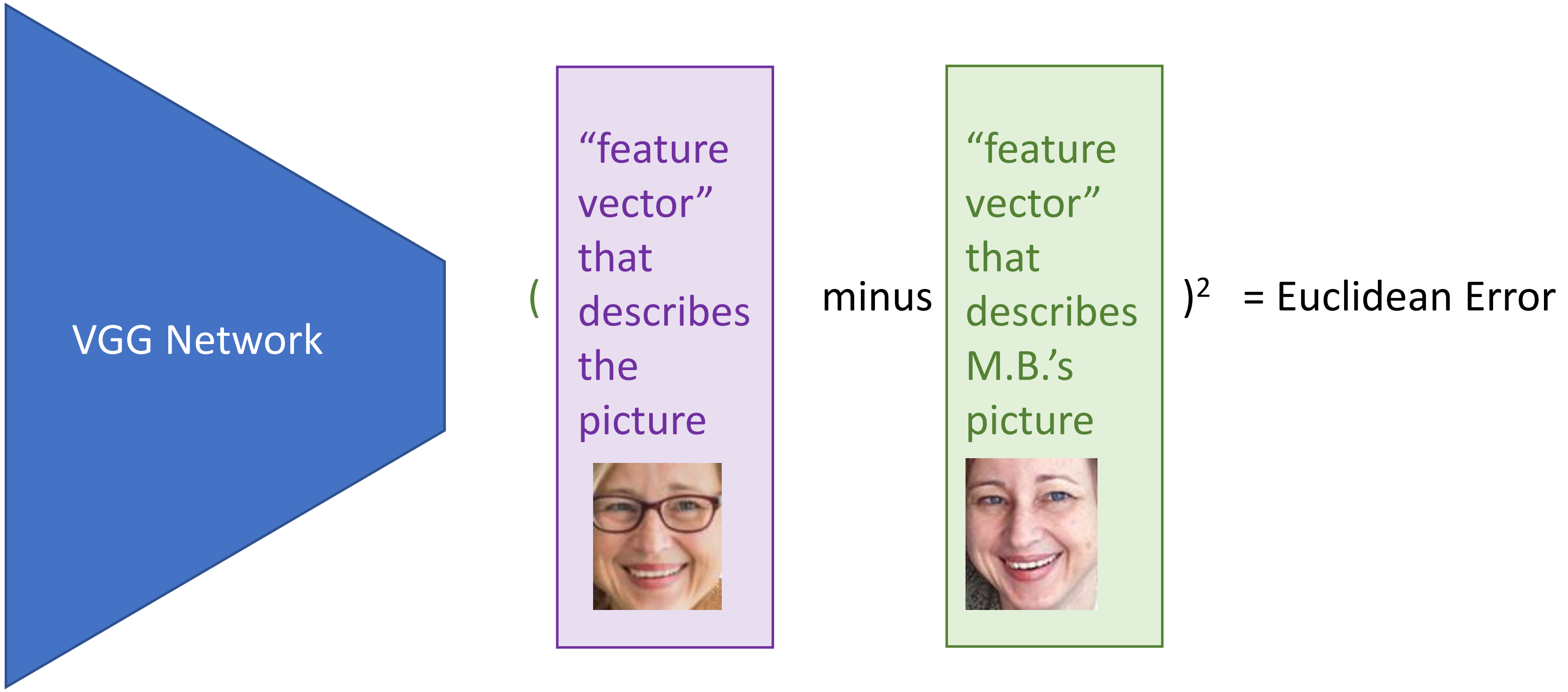
“feature vector” that describes M.B.’s picture

= error

IF error small
THEN unknown person = Margrit Betke

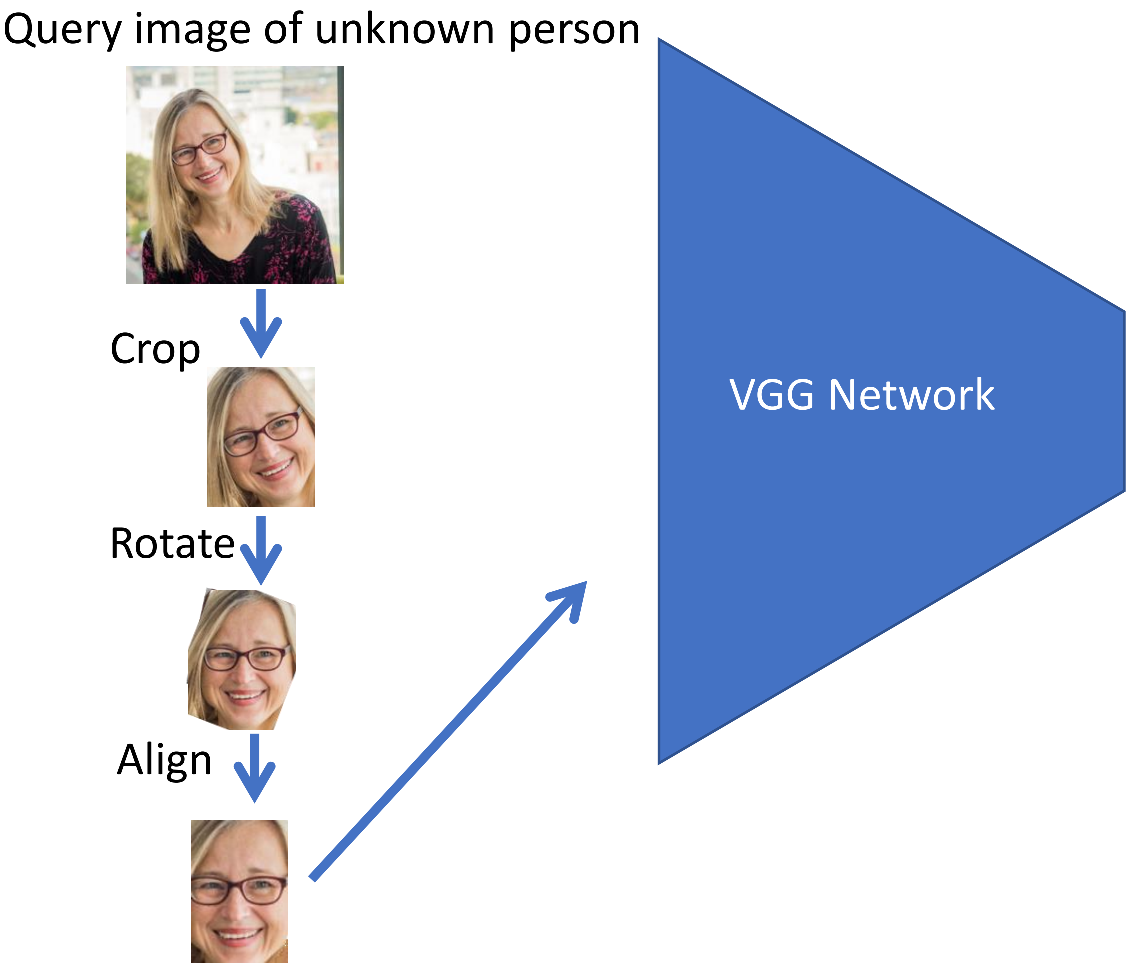
Face Verification in “Use Mode”

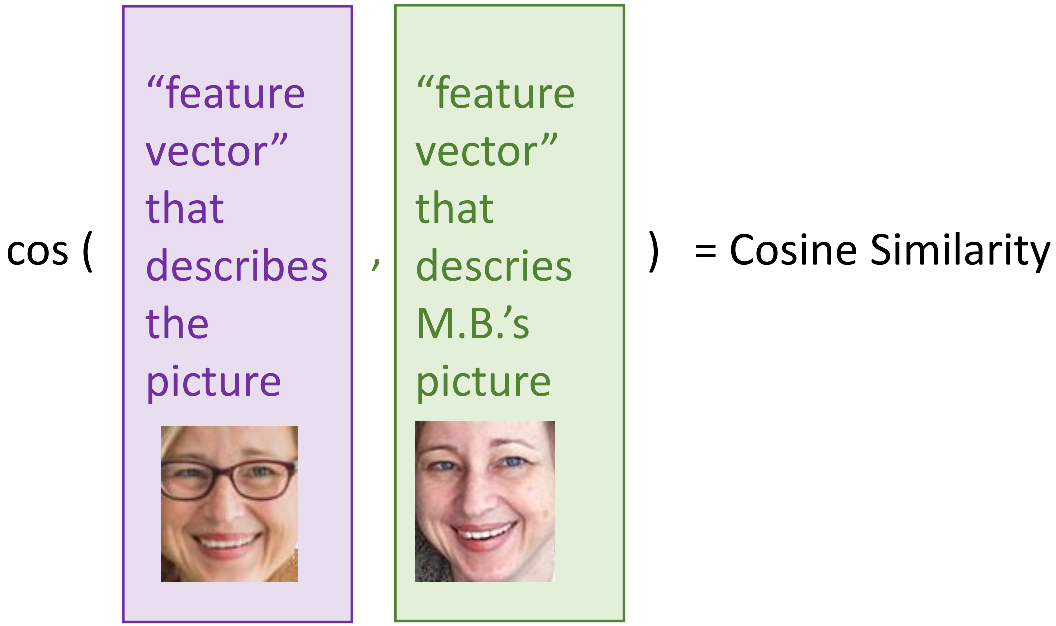




IF Euclidean error small
    THEN   unknown person = Margrit Betke

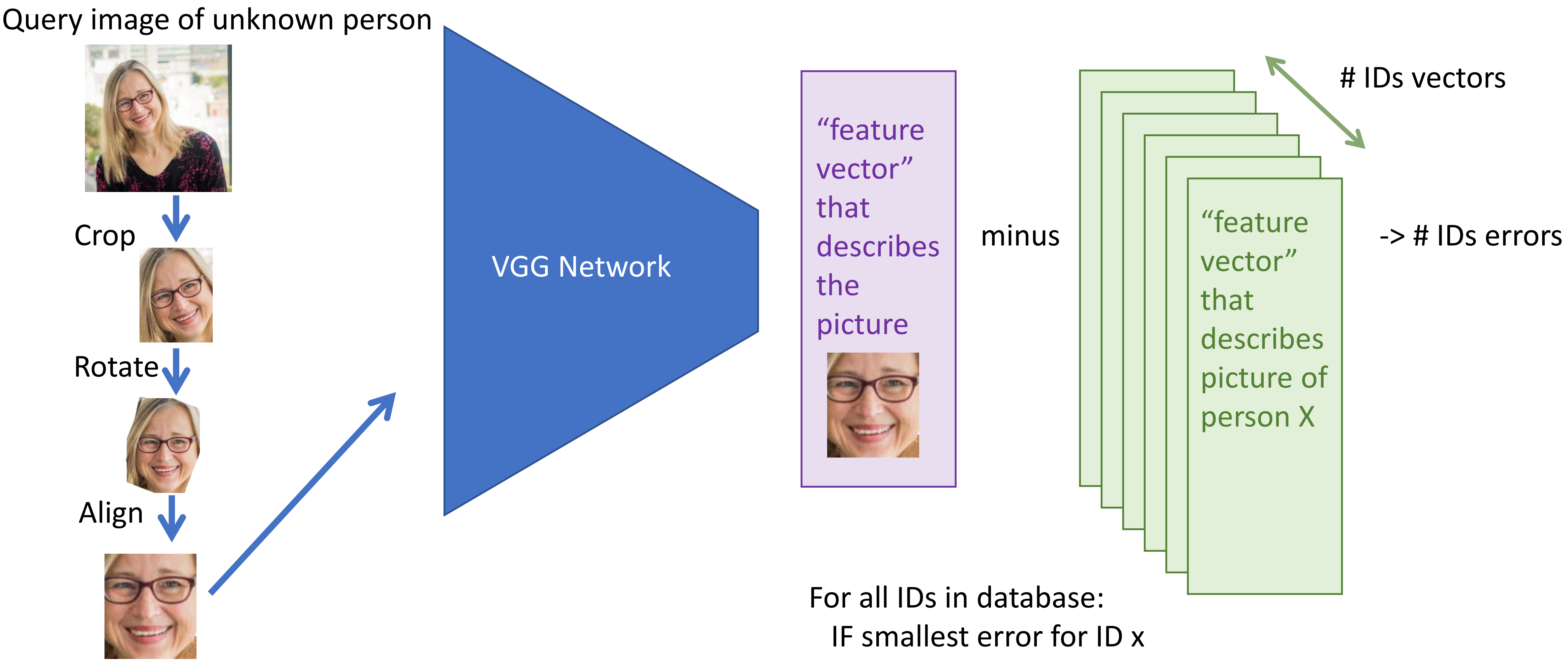
Face Verification in “Use Mode”





IF Cosine Similarity high
THEN unknown person = Margrit Betke

Face Identification in “Use Mode”



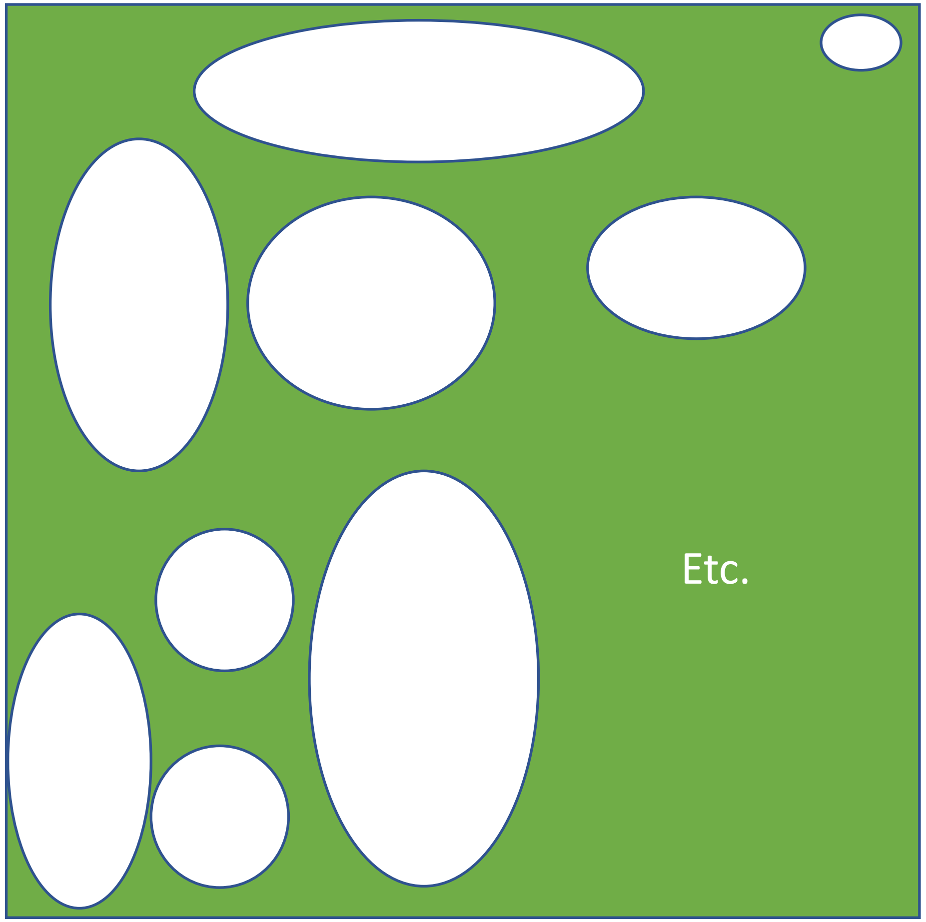
For all IDs in database:
  IF smallest error for ID x
    THEN unknown person = ID x

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)

With 1 million "distractor" face images:

Recognition rates go down a lot!

Gallery of Known Subjects

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%

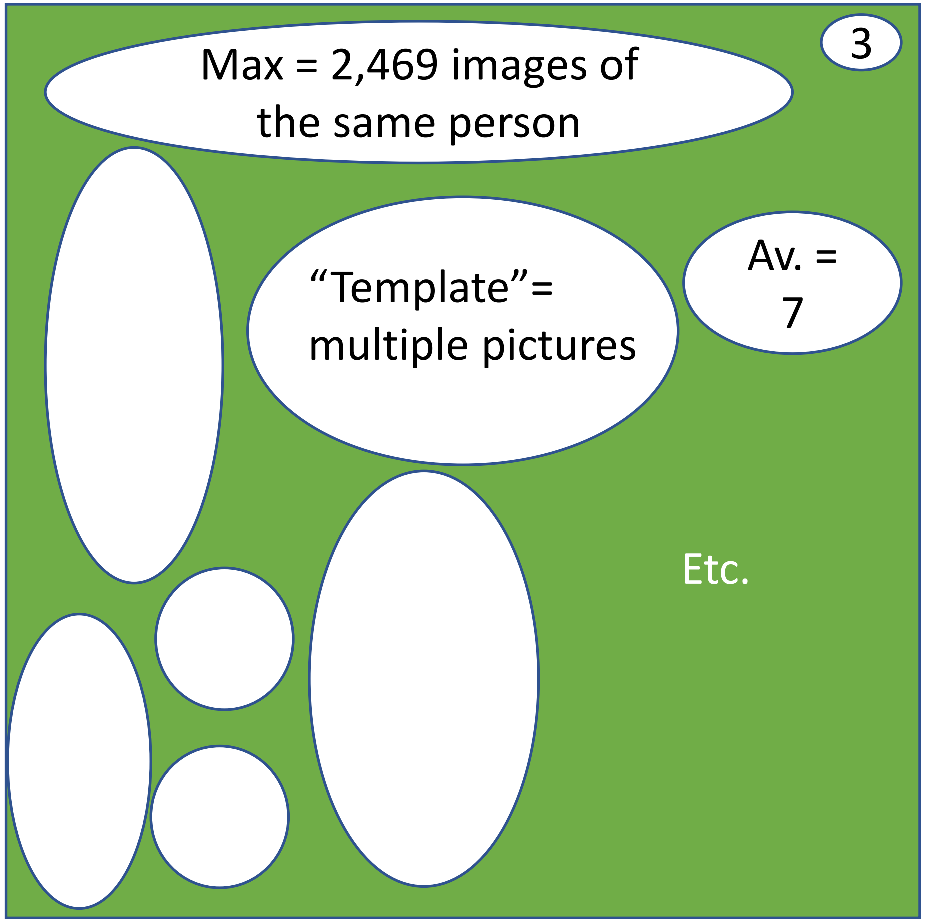
2017 MegaFace2* Dataset

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



2017 Competition

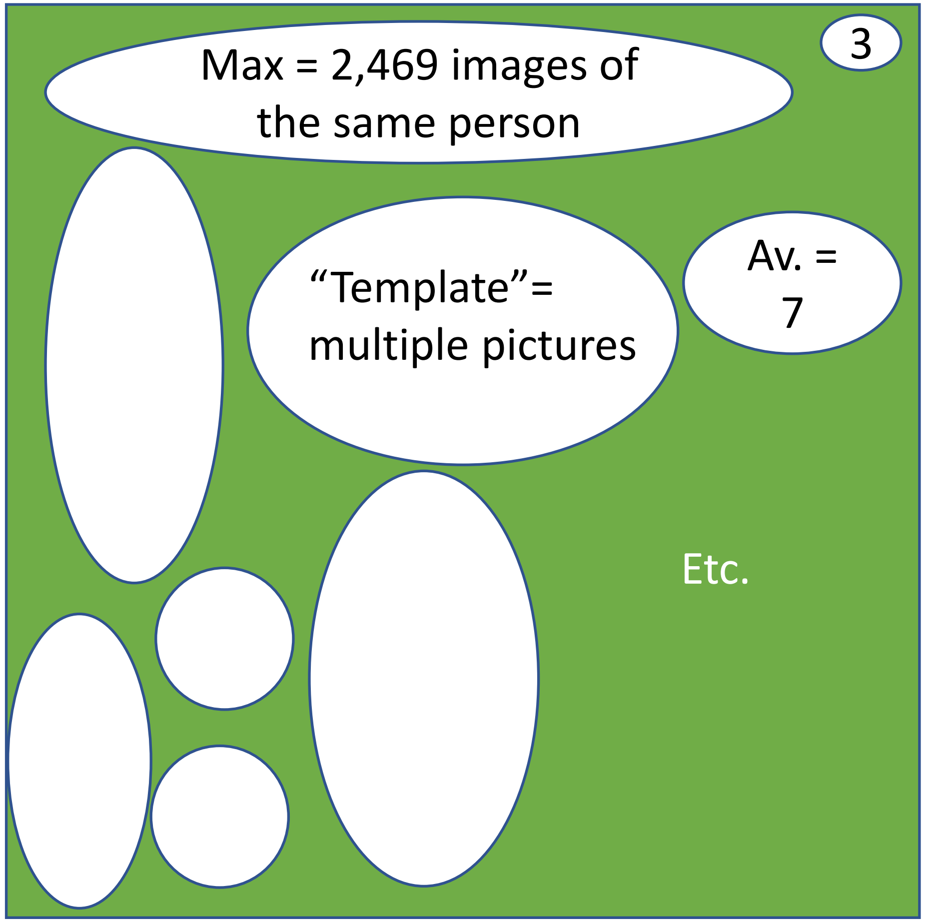
- Train on MegaFace2 (672K IDs, 4 mill. images)
- Test on FaceScrub* (530 IDs, 106K images)
- Add 1 million "distractor" face images

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



* vintage.winklerbros.net

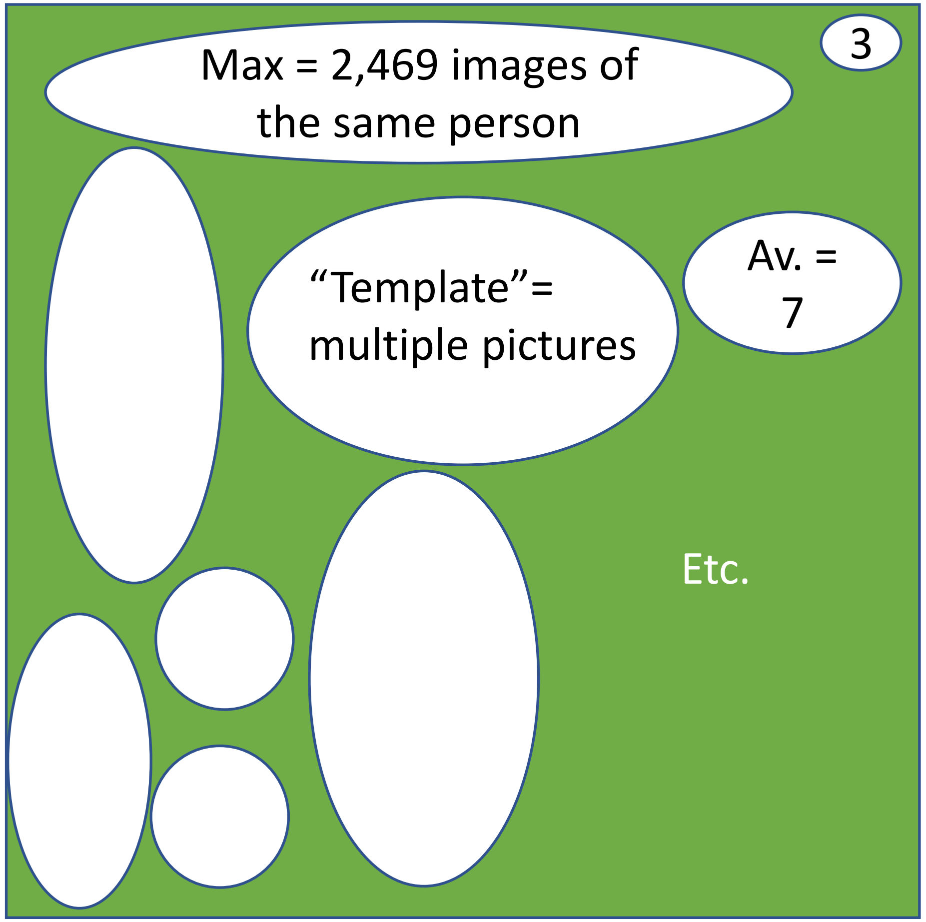
2017 MegaFace2 Dataset

- 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.

0 Same Person?

0 Same Person?

1 Same Person?

Cao et al., CVPR 2018

1 Same Person?

[Cao et al., CVPR 2018]

2 Same Person?

Cao et al., 2018

2 Same Person?

Cao et al., 2018

3 Same Person?

Cao et al., 2018

3 Same Person?

Cao et al., 2018

4 Same Person?

Cao et al., 2018

4 Same Person?

Cao et al., 2018

5 Same Person?

Cao et al., 2018

5 Same Person?

Cao et al., 2018

Cao et al., 2018

Same Person?

1

2

3

4

5

1 Same Person?

Cao et al., 2018

1 Same Person?

Cao et al., 2018

2 Same Person?

Cao et al., 2018

2 Same Person?

Cao et al., 2018

3 Same Person?

Cao et al., 2018

3 Same Person?

Cao et al., 2018

4 Same Person?

Cao et al., 2018

4 Same Person?

Cao et al., 2018

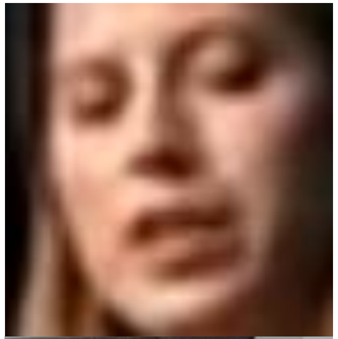
5 Same Person?

Cao et al., 2018

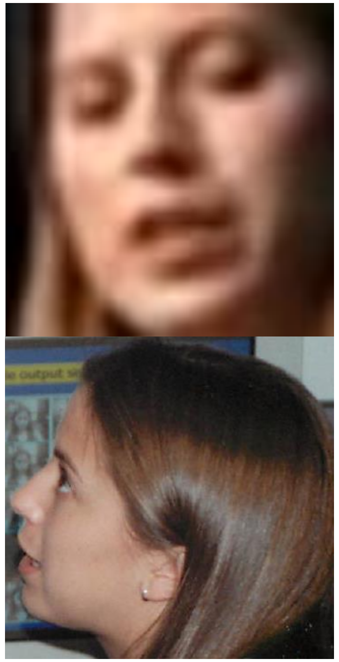
5 Same Person?

Cao et al., 2018

6 Same Person?



6 Same Person?



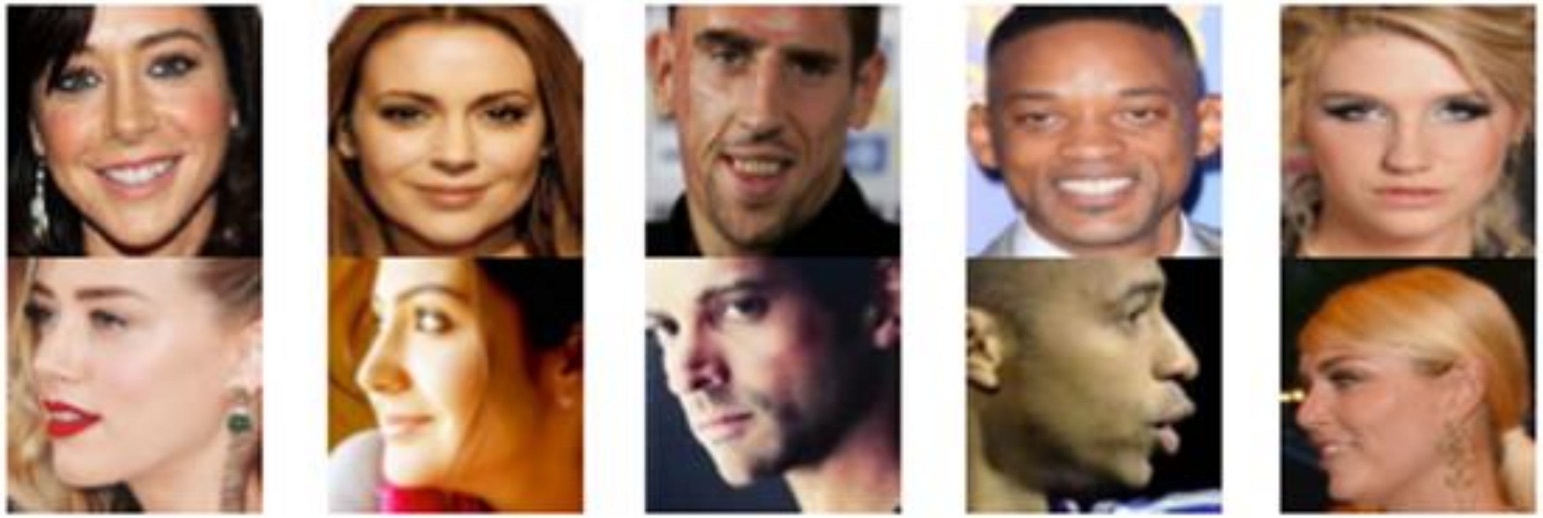
7 Same Person?

7 Same Person?

Frontal vs. Profile

GT: Not same person

False Positives

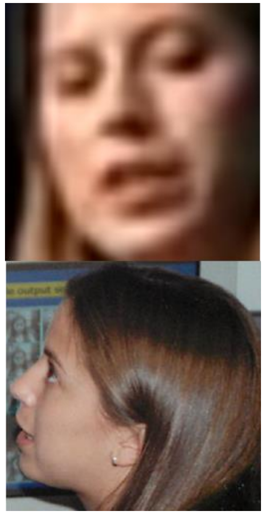


GT: Same person

False Negatives

Cao et al., 2018

Kristen Grauman



Pawan Sinha

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:

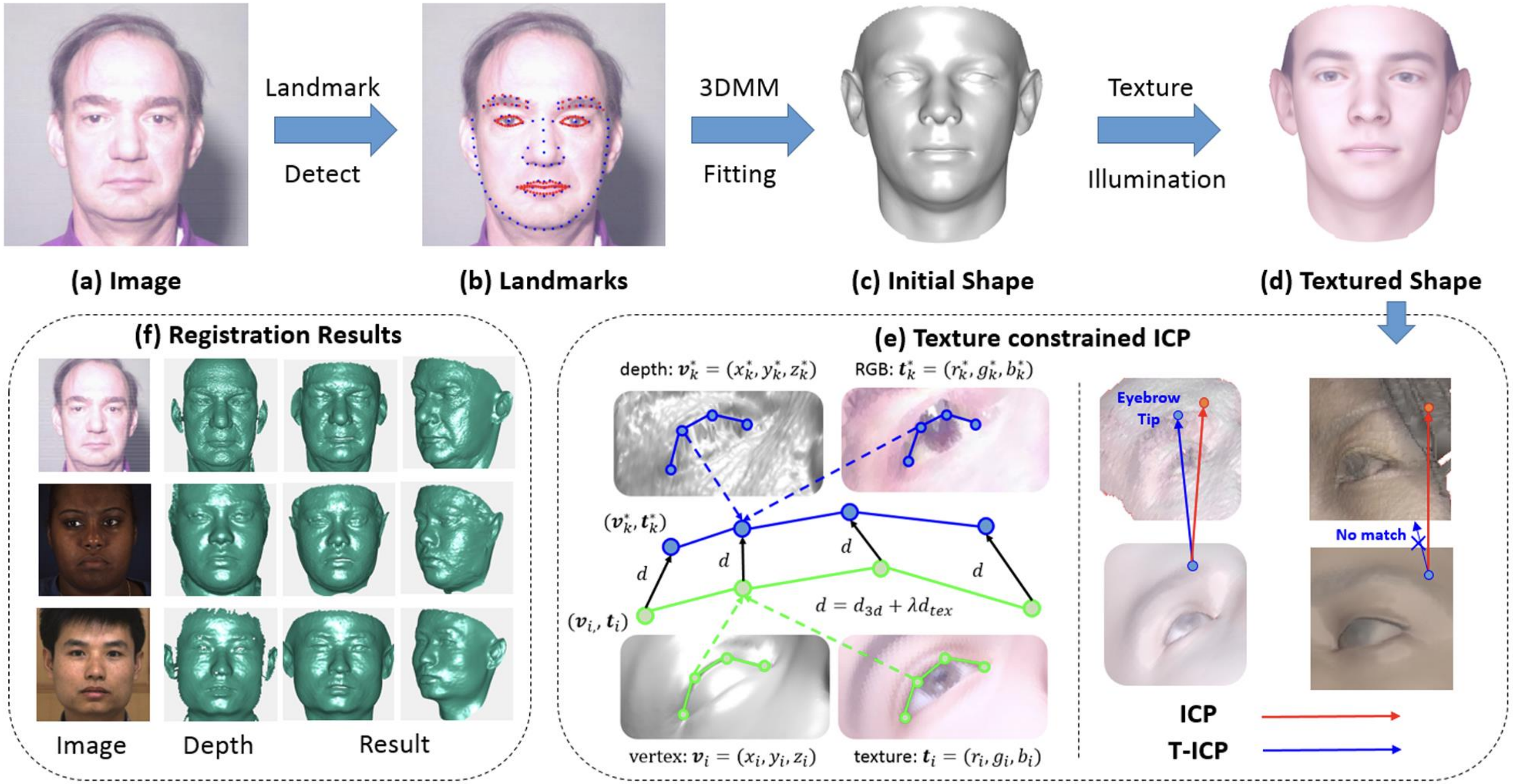
- Pose: 3 Euler angles, translation, scale
- Shape: 50 dimensions
- Expression: 19 dimension

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:



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

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 Age Difference of Images of Faces

A: How Many Years Older?

Best-Rowden and Jain, 2018

B: How Many Years Older?

Best-Rowden and Jain, 2018

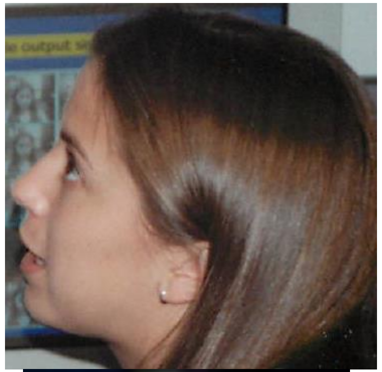
C: How Many Years Older?

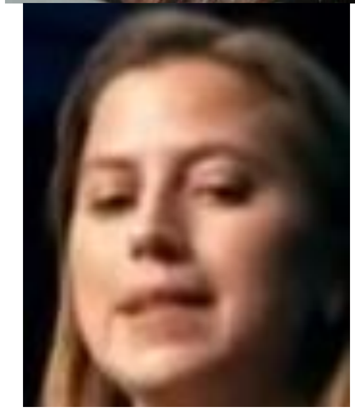
Best-Rowden and Jain, 2018

D: How Many Years Older?

Best-Rowden and Jain, 2018

E: How Many Years Older?





F: How Many Years Older?

G: How Many Years Older?

H: How Many Years Older?

Estimating Aging

A B C D

Best-Rowden and Jain, 2018

Best-Rowden and Jain, 2018

Estimating Aging

Best-Rowden
and Jain, 2018

A B C D

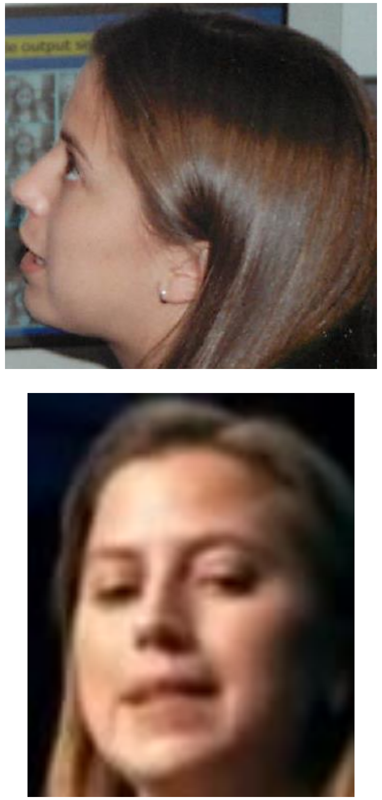
Elapsed time: 9 years 8 years 8 years 9 years

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

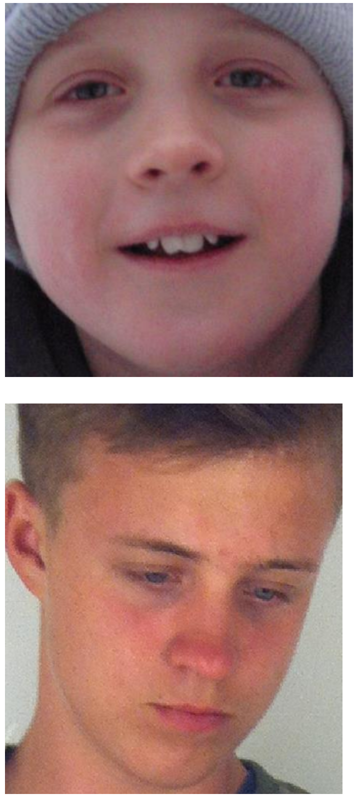
Best-Rowden and Jain, 2018

Estimating Aging

E



F

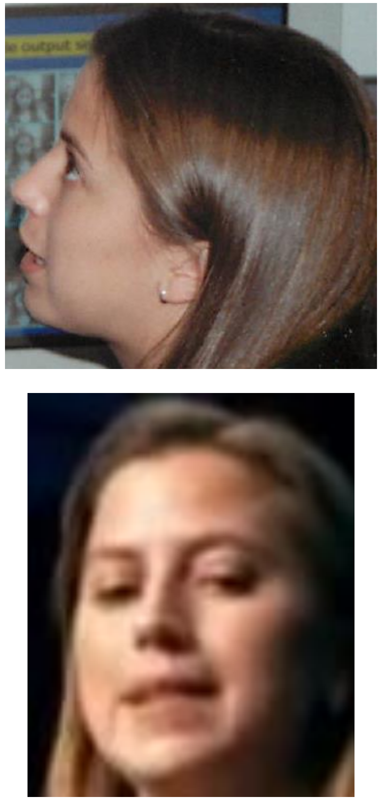


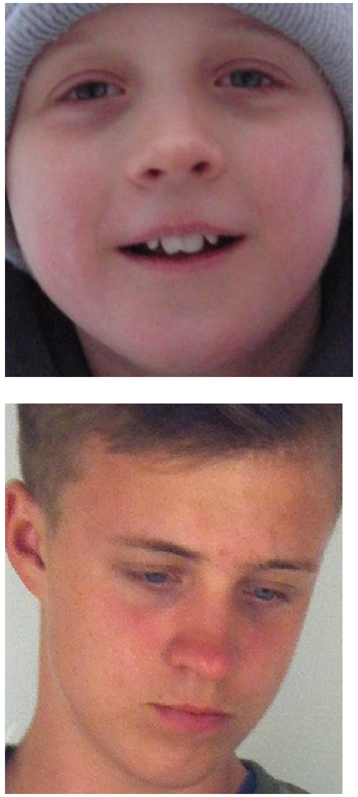
G

H

Estimating Aging

E F G H

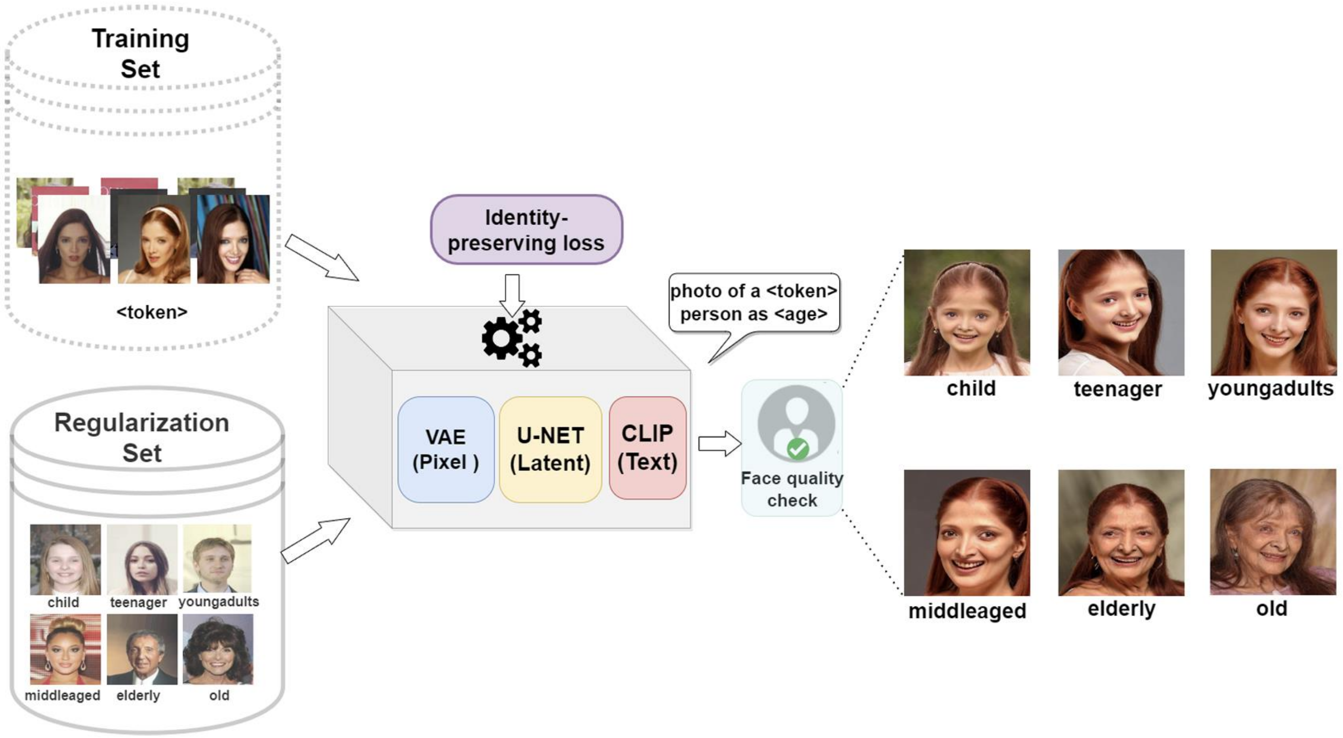




Age Difference: 8 years 12 years 30 years 30 years

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



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

Biometric Datasets from IIT Jodhpur

https://lab-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

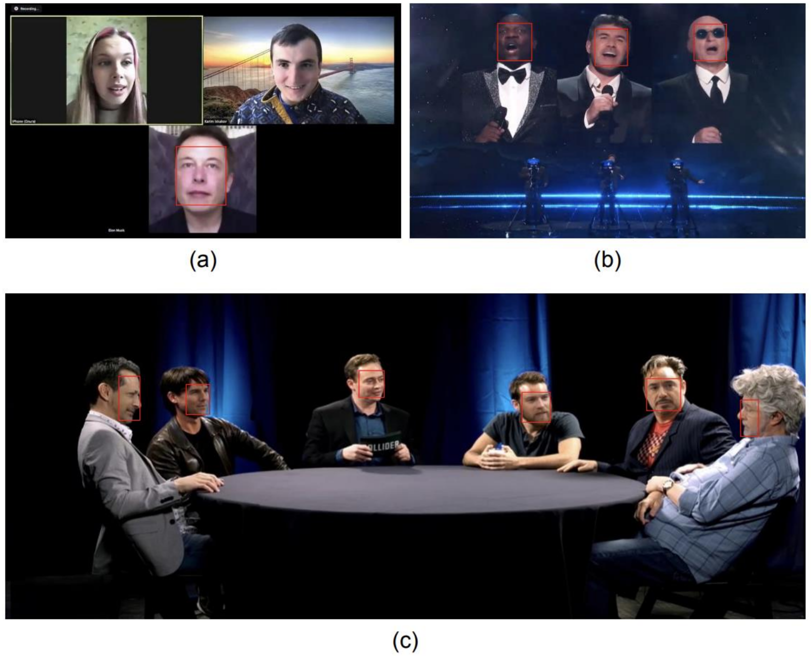
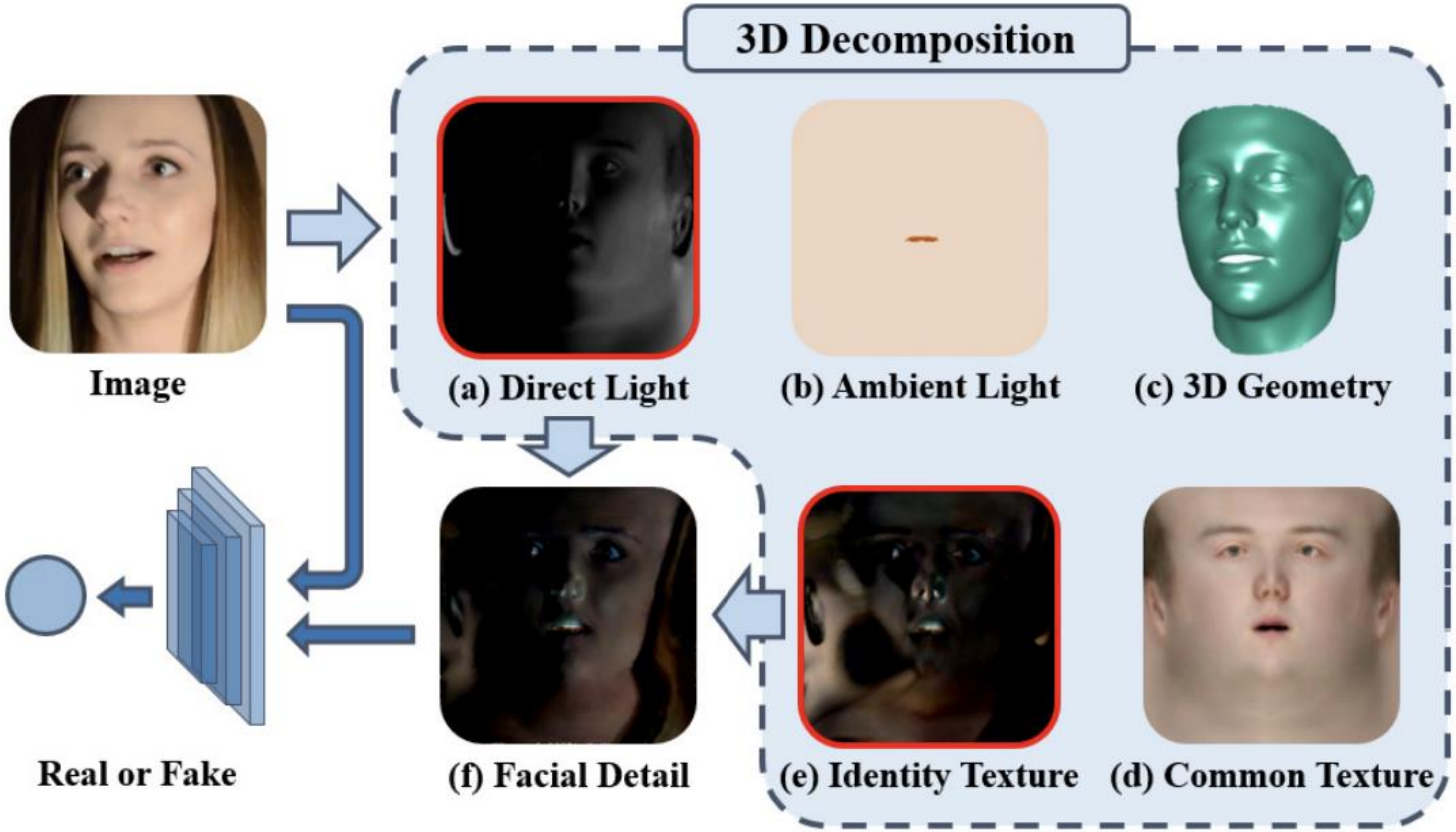


Figure 1. Samples showcasing multi-face deepfakes circulated on social media. (a) A zoom call with a deepfake of Elon Musk [8] (b) Real-time deepfake generation at America's Got Talent [9] (c) Deepfake round-table with multiple deepfake subjects [33].

Face Forgery Detection by 3D Decomposition

Zhu et al., CVPR 2021, PAMI 2023



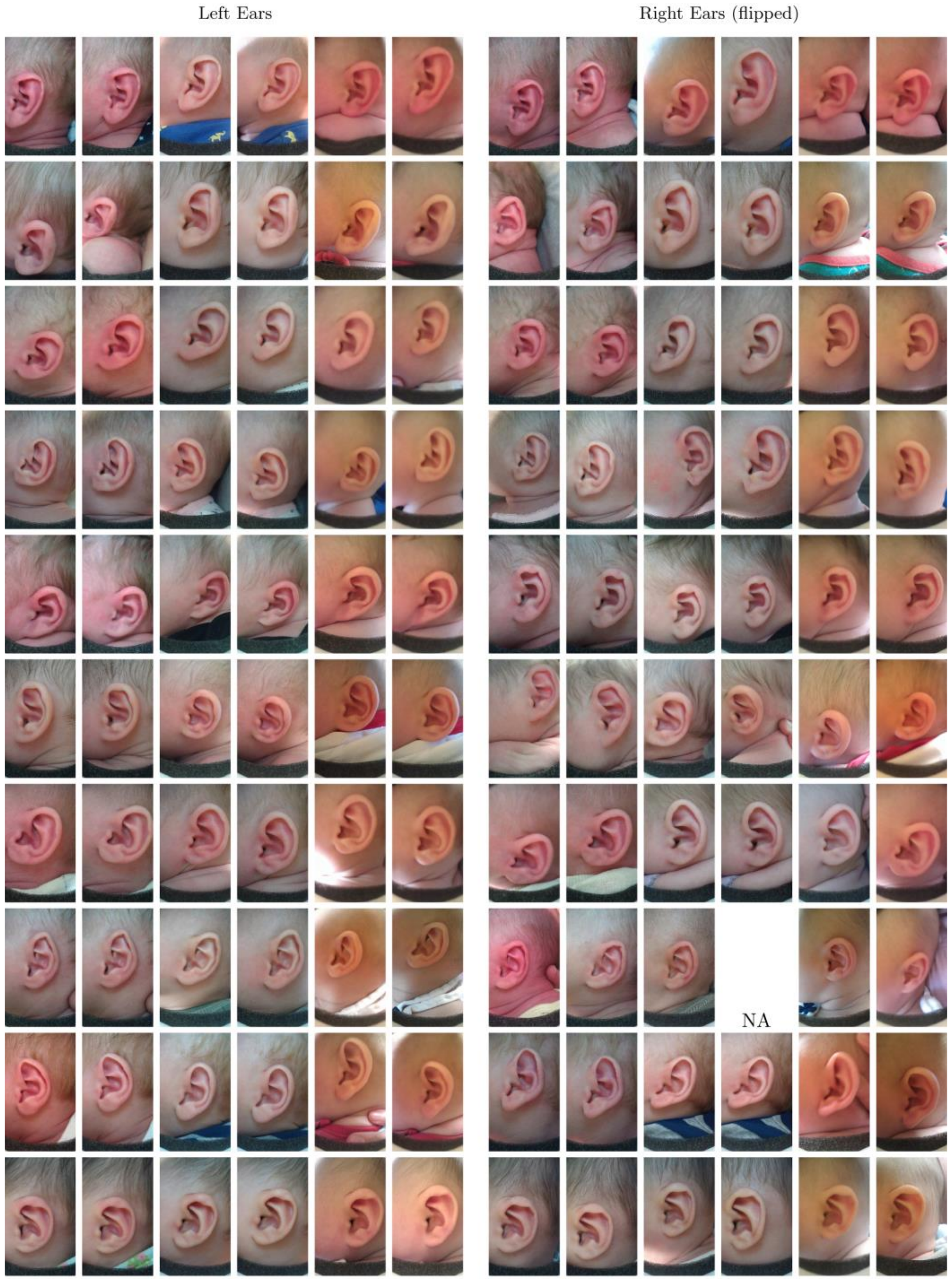
Alternative Biometrics: Ears

Our Task:

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

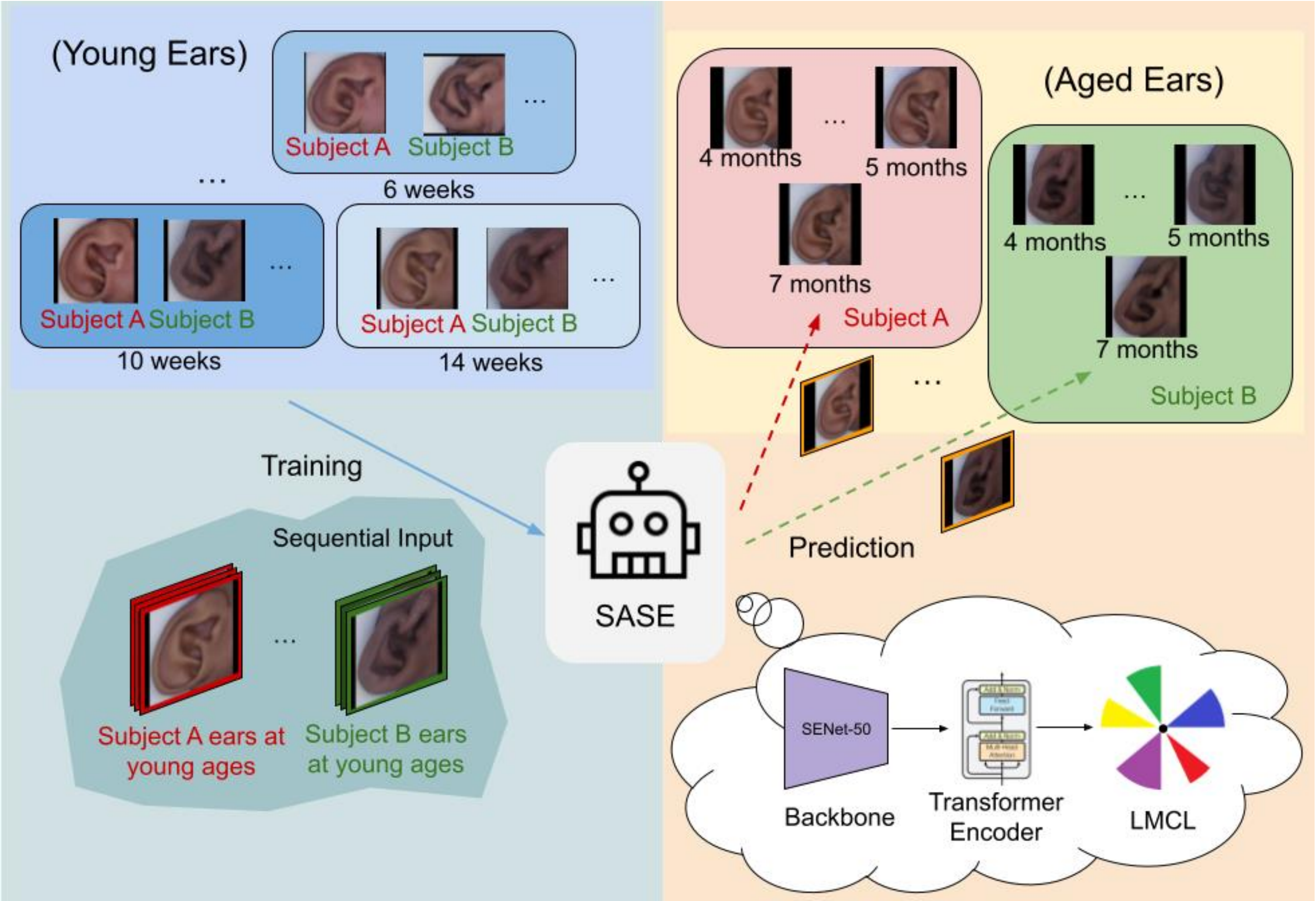
Motivation:

- 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.

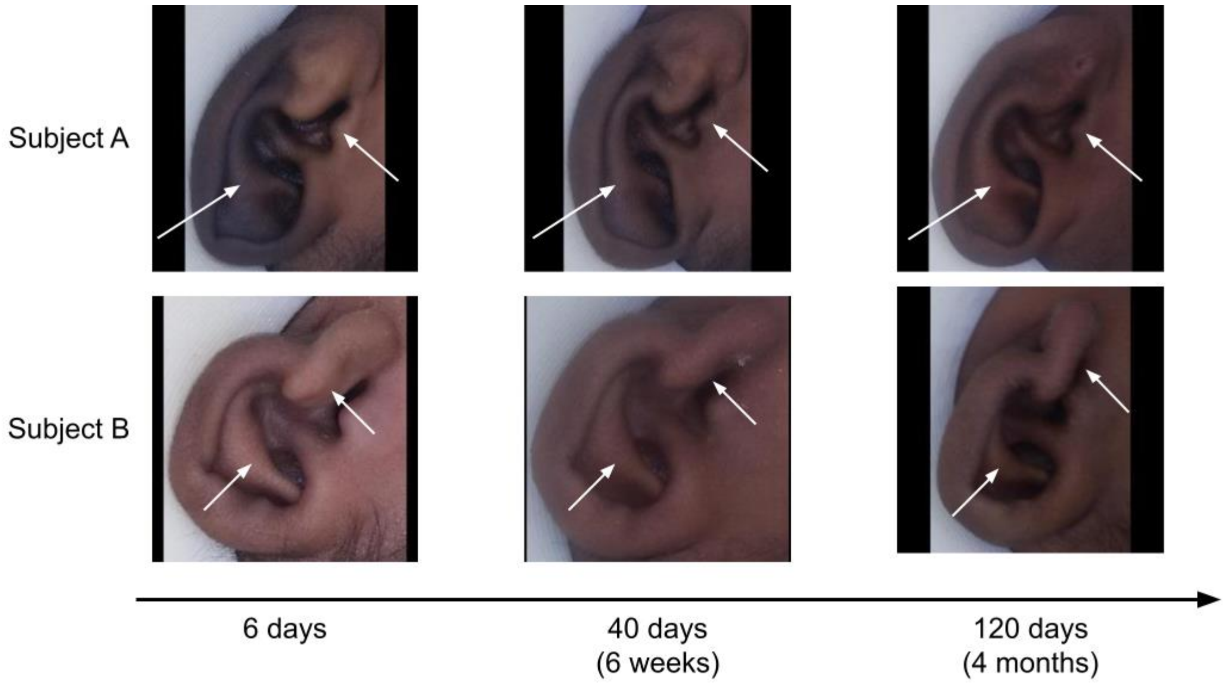


BOSTON UNIVERSITY

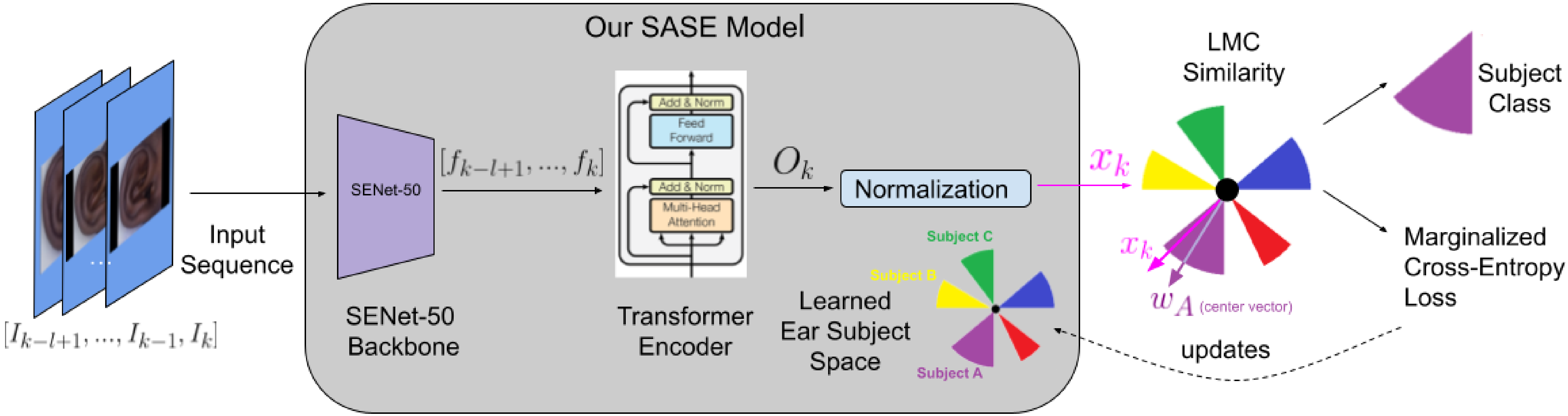
IJCB 2023 Best Poster Award: Qin et al., Age-constrained Ear Recognition: The EICZA Dataset and SASE Baseline Model



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



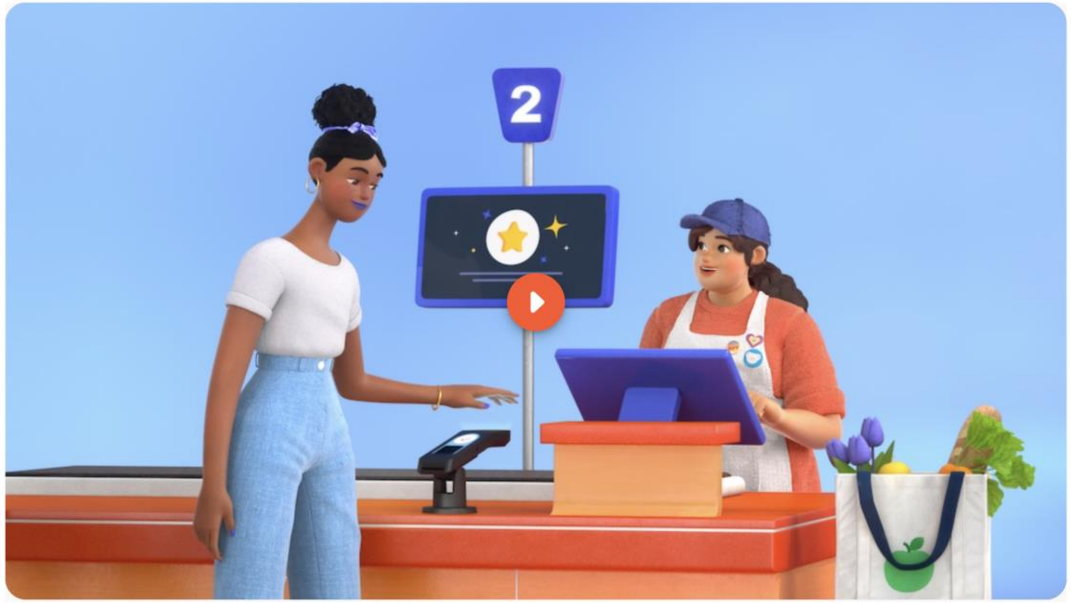
IJCB 2023 Best Poster Award: Qin et al., Age-constrained Ear Recognition: The EICZA Dataset and SASE Baseline Model

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

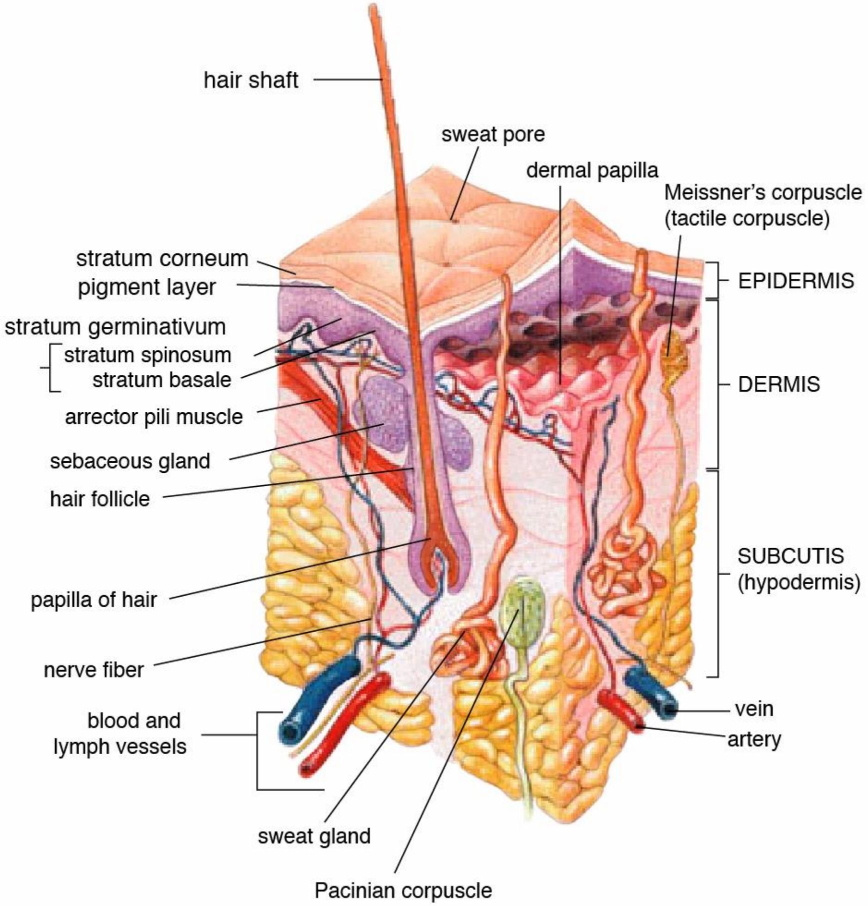
Manoj Aggarwal,
Director of Applied Science,
Amazon One
gave a keynote address
at IJCB 2023, September 28

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