VisionLabs

MACHINES CAN SEE

Face Recognition

21 / 02 / 2017 Philipp Omenitsch

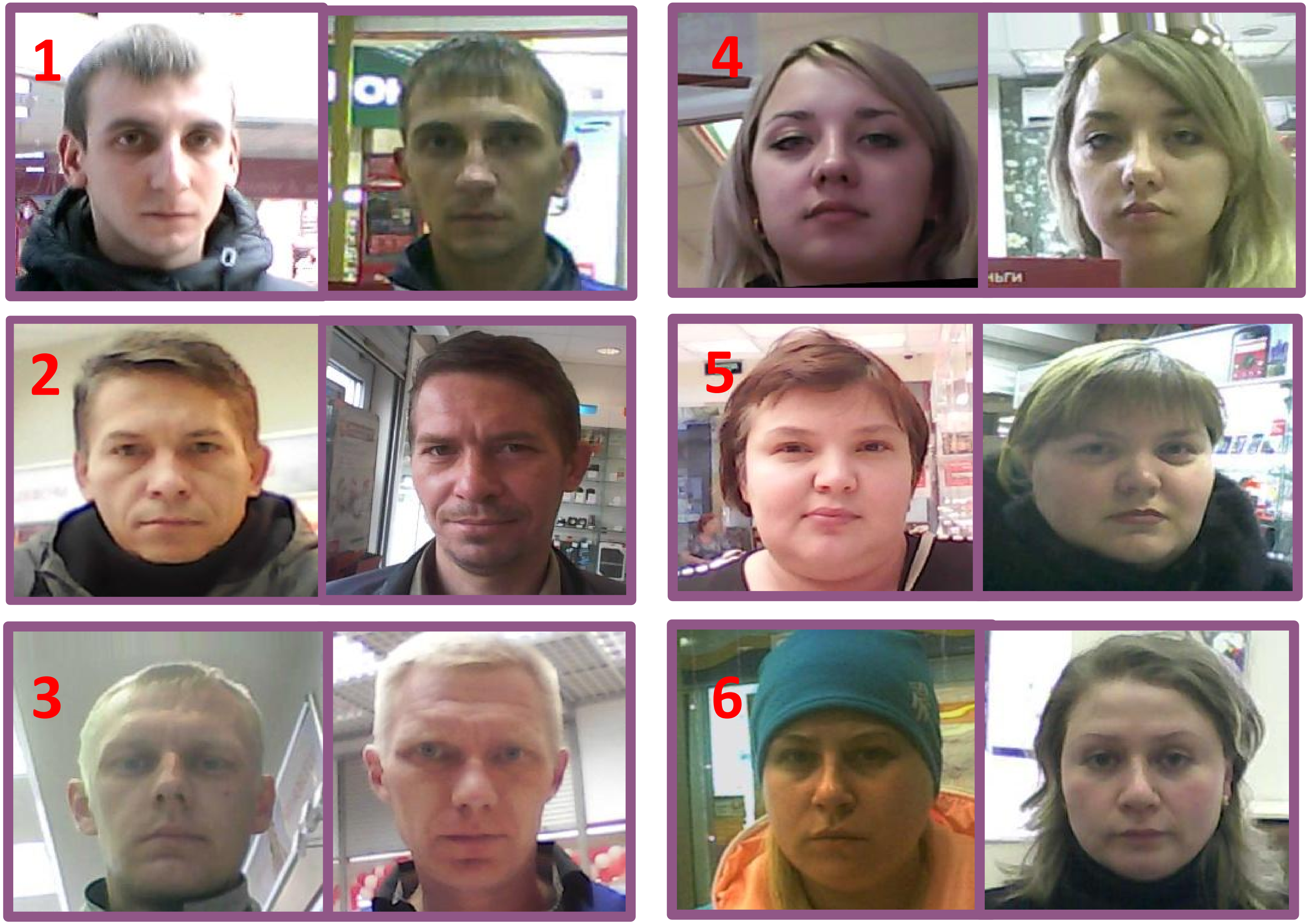
FACE RECOGNITION

Verification (1:1)

Identification (1:N)

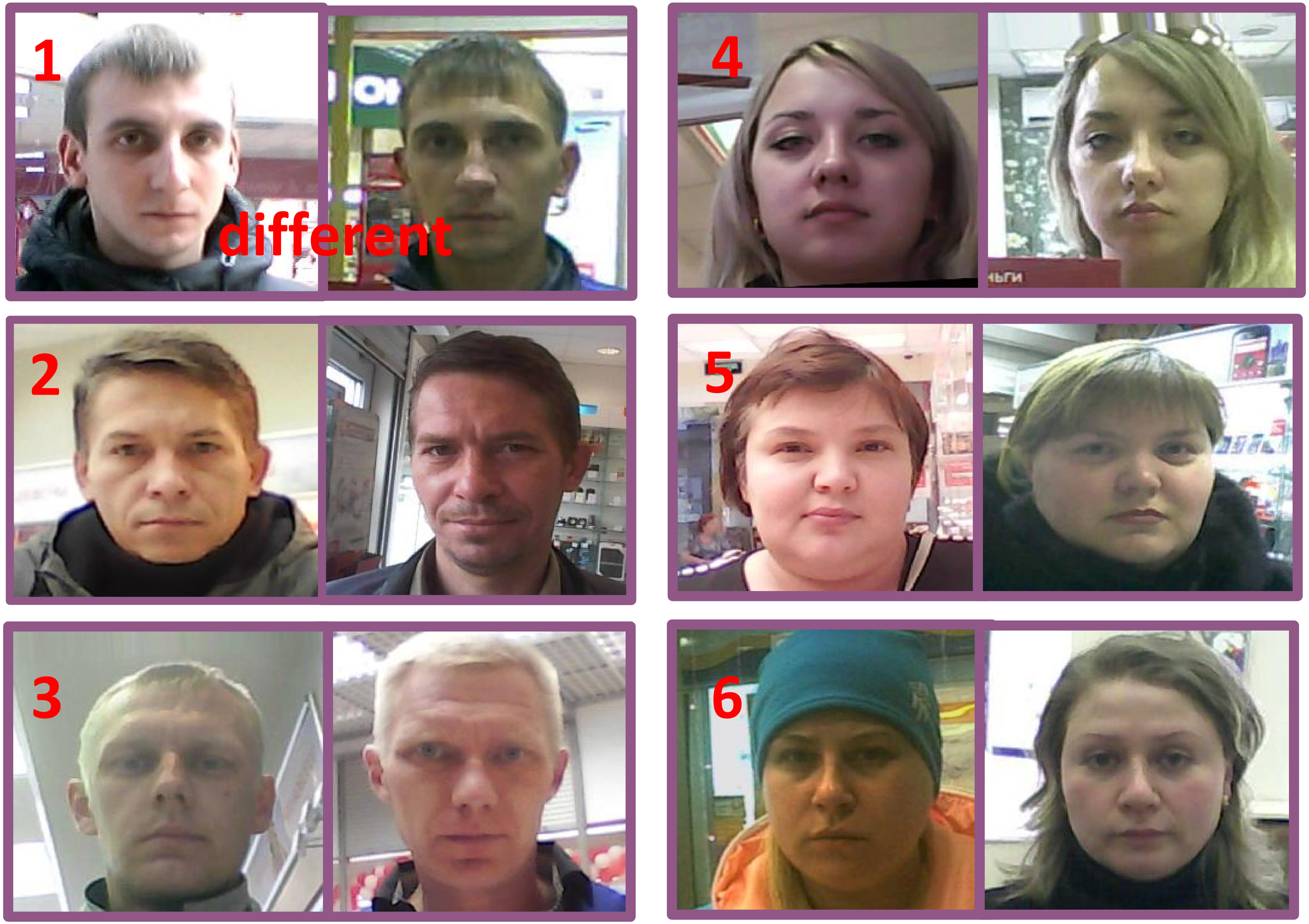
Try to compete with our face recognition engine

VisionLabs
MACHINES CAN SEE



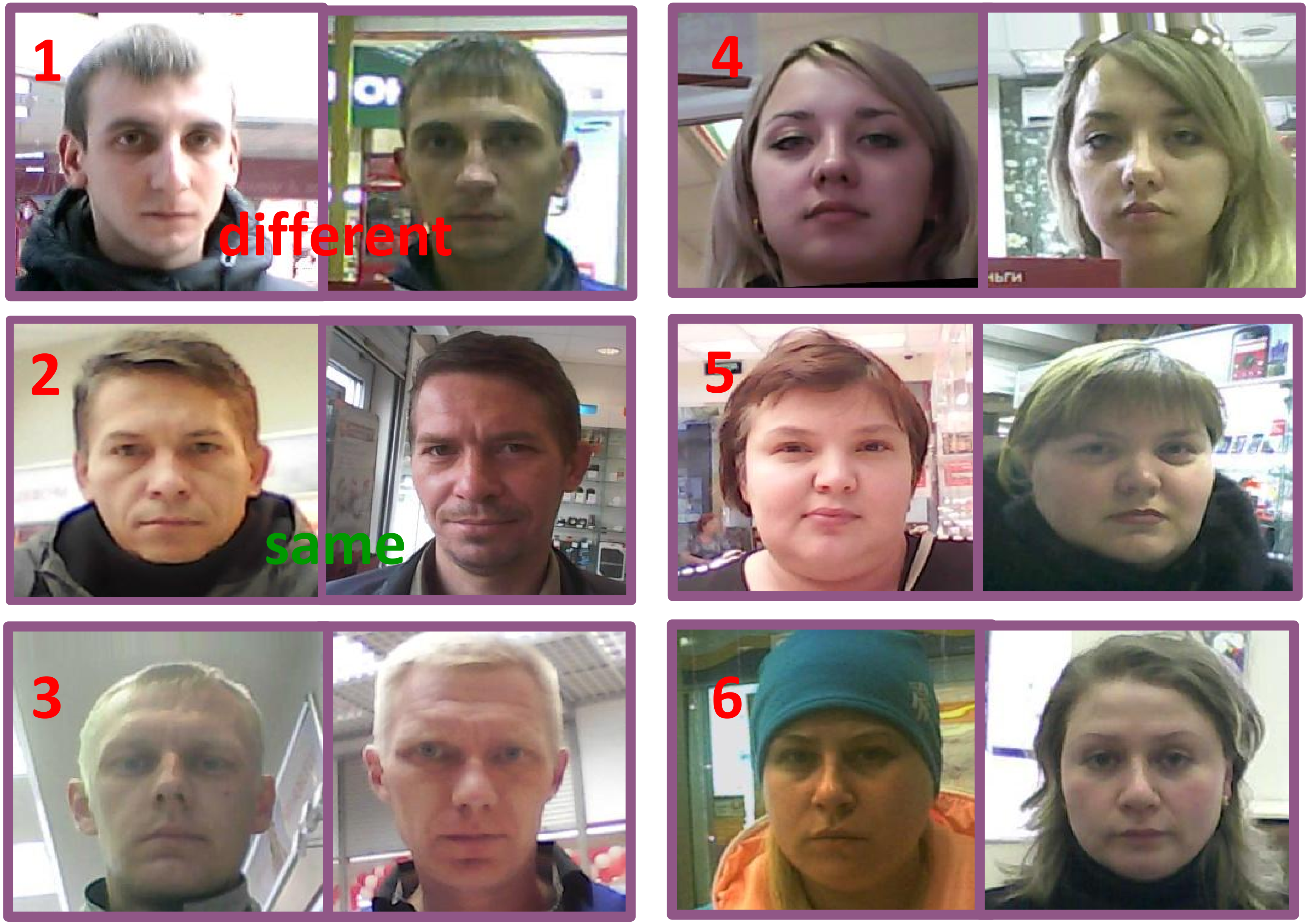
Compare the pairs, same or different?

VisionLabs
MACHINES CAN SEE



Our algorithm gets 6/6 right

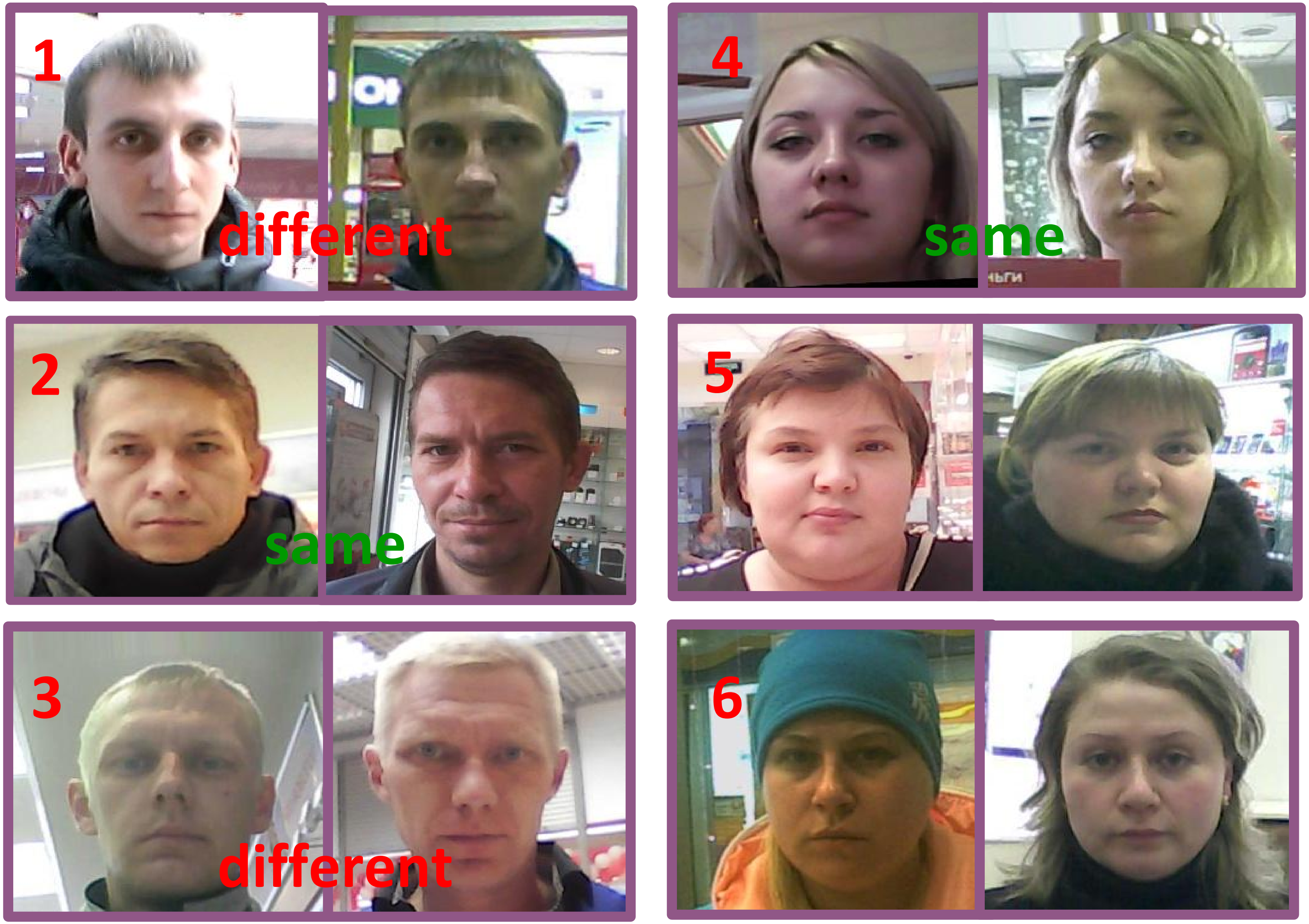
VisionLabs
MACHINES CAN SEE



Our algorithm gets 6/6 right

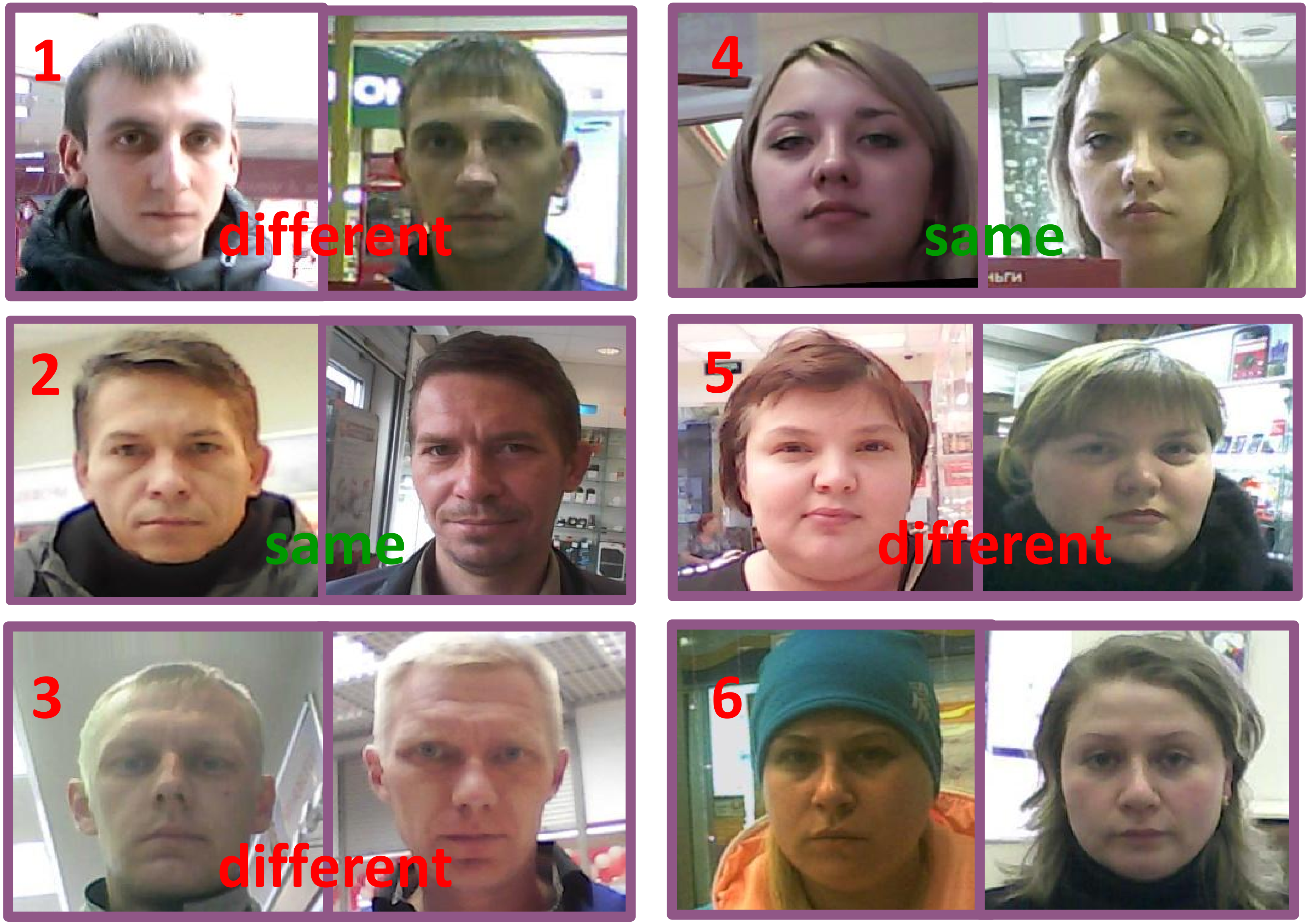
VisionLabs
MACHINES CAN SEE

Our algorithm gets 6/6 right



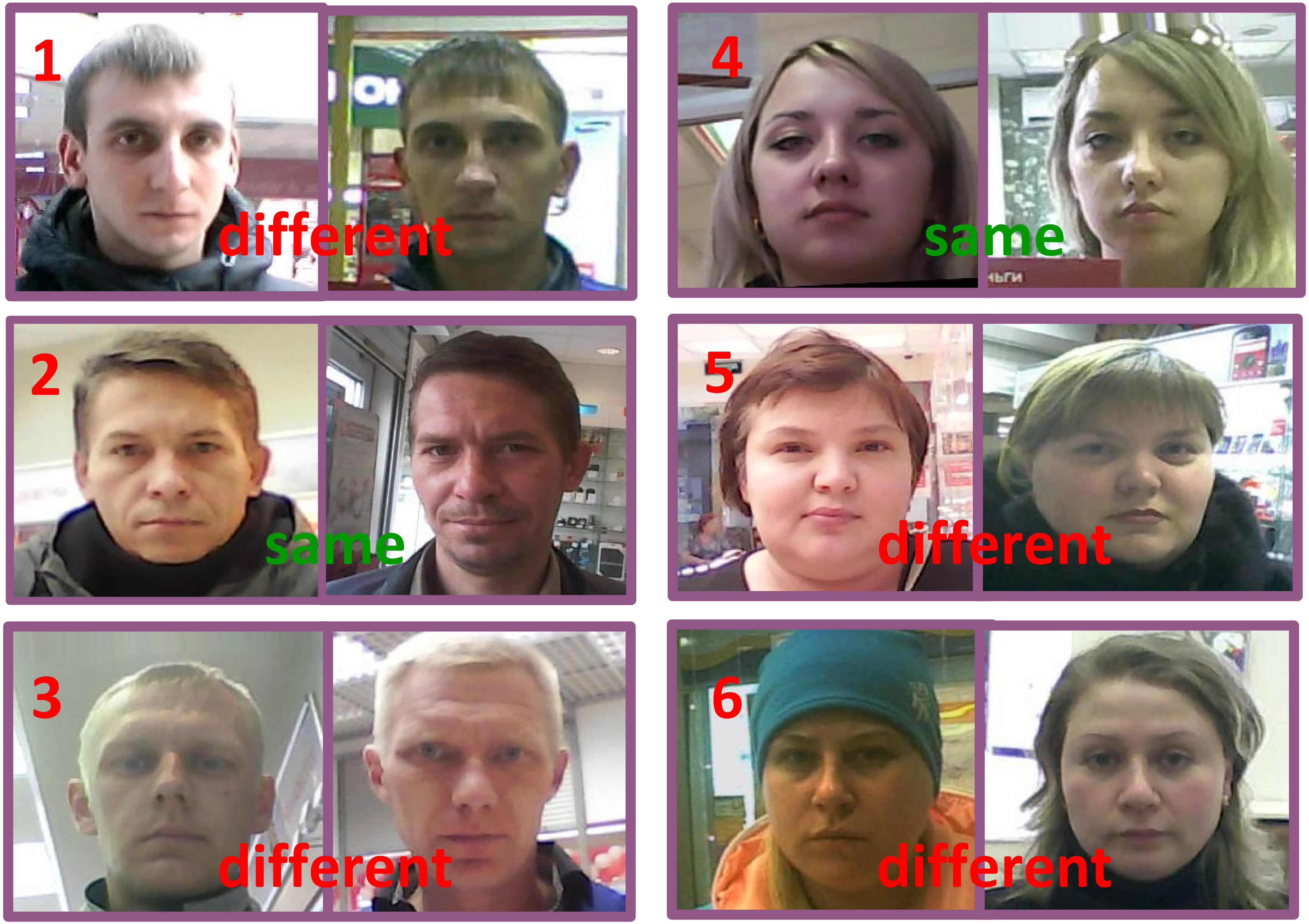
Our algorithm gets 6/6 right

VisionLabs
MACHINES CAN SEE



Our algorithm gets 6/6 right

VisionLabs
MACHINES CAN SEE



Our algorithm gets 6/6 right

VisionLabs
MACHINES CAN SEE

WHY IS IT SO HARD FOR HUMANS?

- Subtle intrapersonal and interpersonal variations
- Intrinsic factors – makeup, hair styles, emotions
- Extrinsic factors – scaling, noise, occlusion, etc.

WHY THE FACE?

- Doesn’t need interaction for capture
- Doesn’t require any special / expensive hardware

VisionLabs
MACHINES CAN SEE

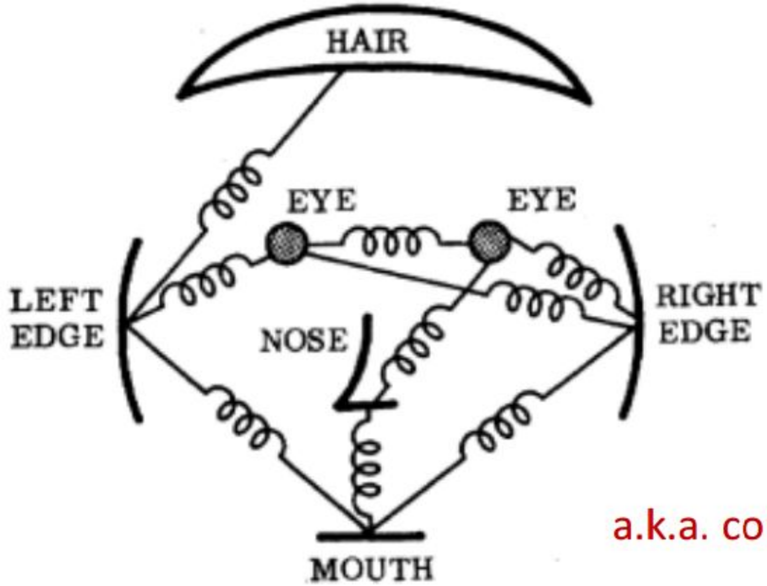
TECHNOLOGY

PREREQUISITES

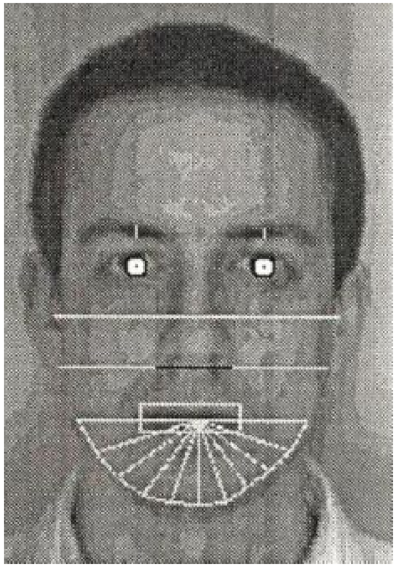
- **Lots of data**
  - access to millions of annotated faces
- **Hardware (GPUs)**
  - currently 12 NVIDIA TITAN X
- **State-of-the-art Neural networks**
  - we are doing our own research and our employees publish papers

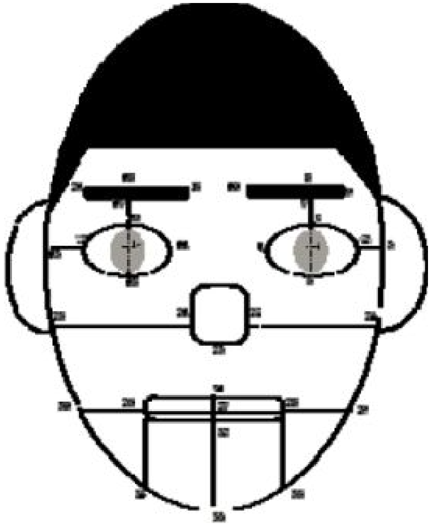
FROM THE VERY BEGINNING…

1973



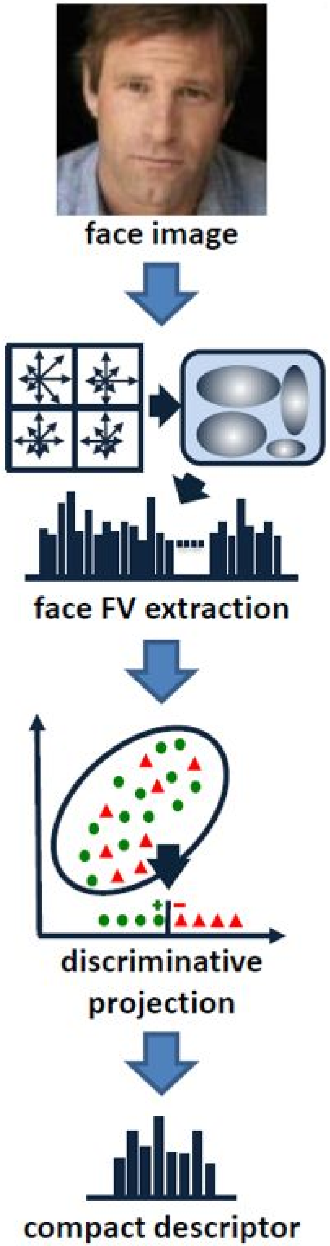
a.k.a. constellation model



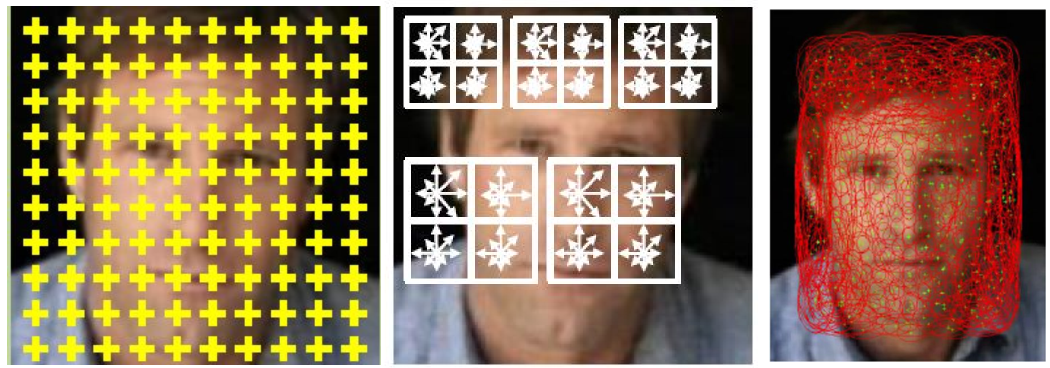


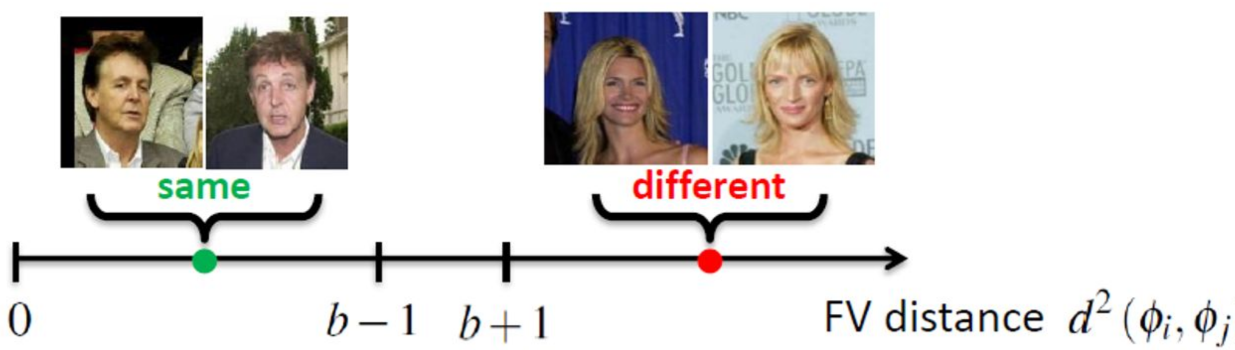
The representation and matching of pictorial structures, Fischler and Elschlager, 1973

OVER HAND CRAFTED FEATURES…

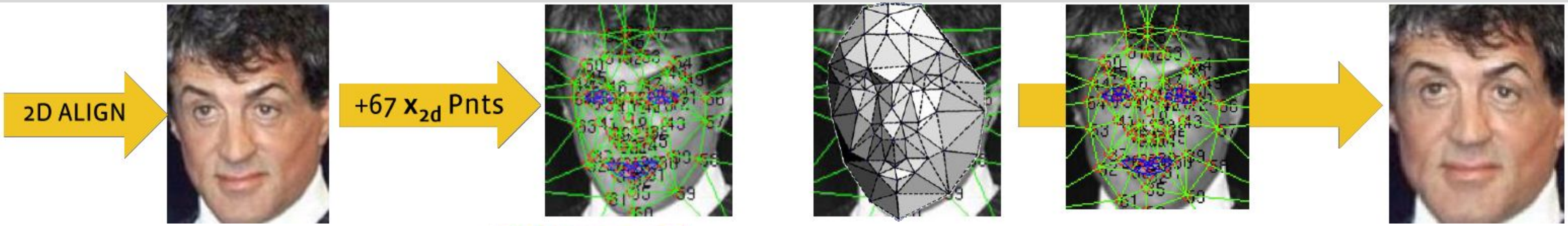


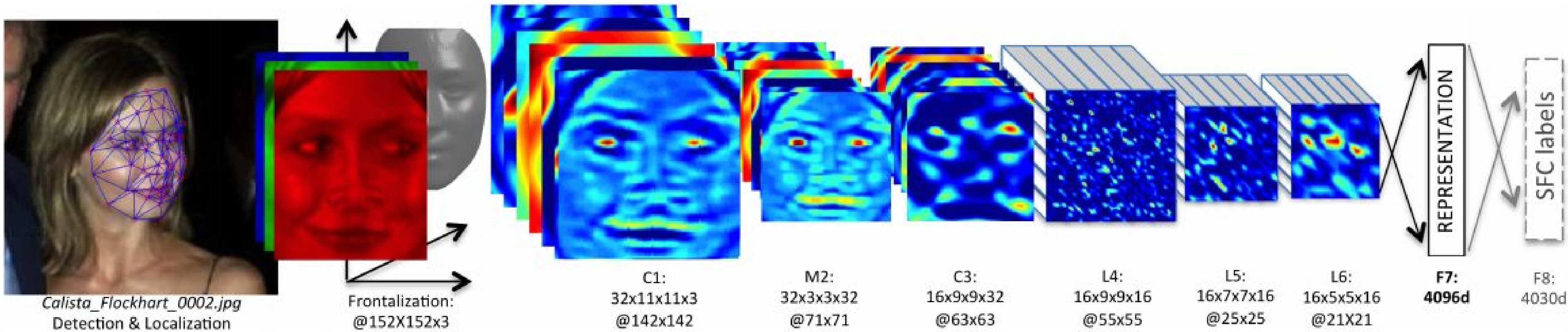
Perronnin 2010, Simonyan 2013 Fisherfaces

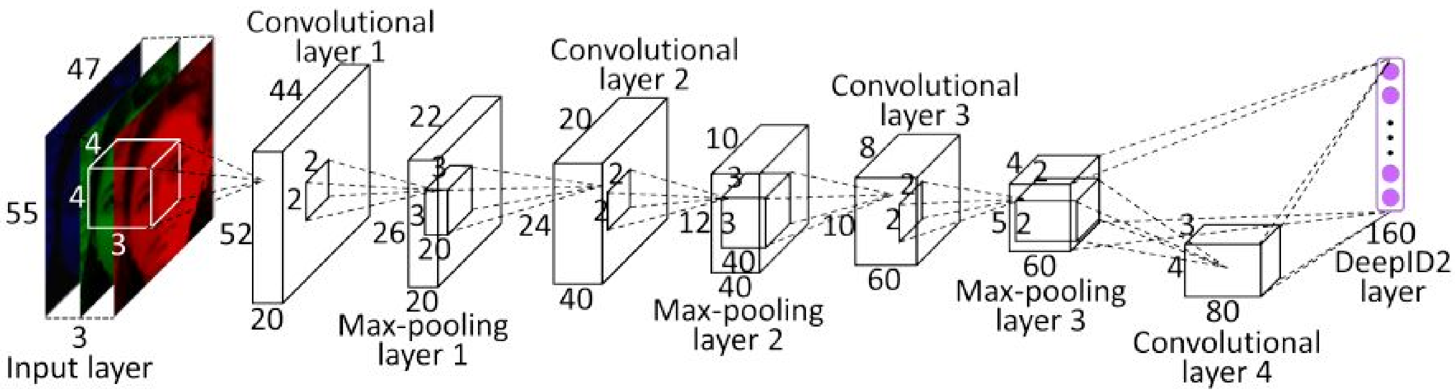




TO NOWADAYS



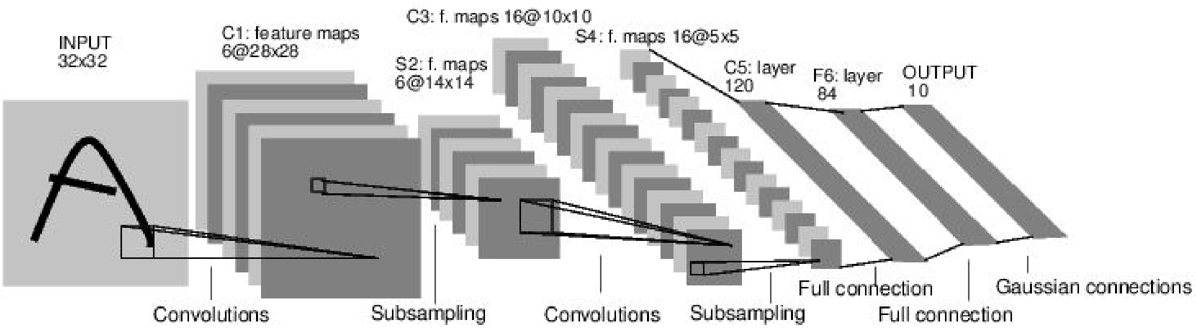


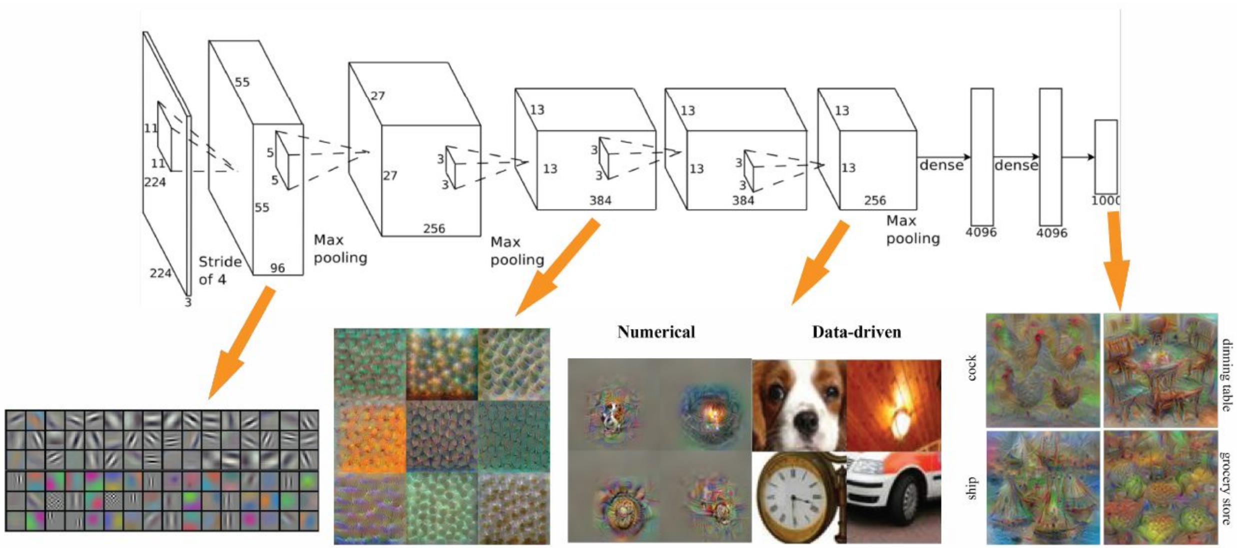


[1-2] Deepface 2015, Facebook
[3] DeepID Sun et al. 2014

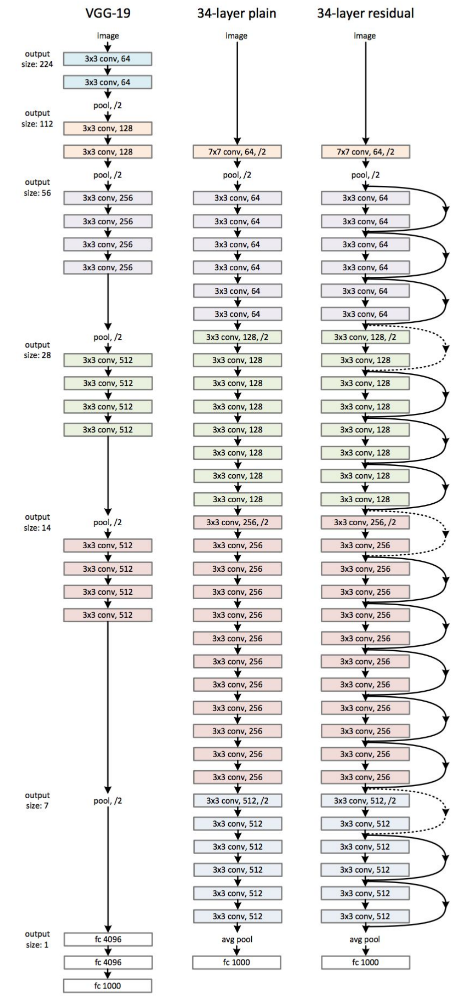
VisionLabs
MACHINES CAN SEE

NETWORK ARCHITECTURES





Conv 1: Edge+Blob Conv 3: Texture Conv 5: Object Parts Fc8: Object Classes



[1] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition Proceedings of the IEEE, 86(11):2278–2324, 1998.
[2] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems, pages 1097–1105, 2012.
[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. arXiv preprint arXiv:1512.03385, 2015.

2015 ResNet (ILSVRC'15) 3.57

Year	Codename	Error (percent)	99.9% Conf Int
2014	**GoogLeNet**	**6.66**	**6.40 - 6.92**
2014	VGG	7.32	7.05 - 7.60
2014	MSRA	8.06	7.78 - 8.34
2014	AHoward	8.11	7.83 - 8.39
2014	DeeperVision	9.51	9.21 - 9.82
2013	Clarifai†	11.20	10.87 - 11.53
2014	CASIAWS†	11.36	11.03 - 11.69
2014	Trimps†	11.46	11.13 - 11.80
2014	Adobe†	11.58	11.25 - 11.91
2013	**Clarifai**	**11.74**	**11.41 - 12.08**
2013	NUS	12.95	12.60 - 13.30
2013	ZF	13.51	13.14 - 13.87
2013	AHoward	13.55	13.20 - 13.91
2013	OverFeat	14.18	13.83 - 14.54
2014	Orange†	14.80	14.43 - 15.17
2012	SuperVision†	15.32	14.94 - 15.69
2012	**SuperVision**	**16.42**	**16.04 - 16.80**
2012	ISI	26.17	25.71 - 26.65
2012	VGG	26.98	26.53 - 27.43
2012	XRCE	27.06	26.60 - 27.52
2012	UvA	29.58	29.09 - 30.04

Microsoft ResNet, a 152 layers network

GoogLeNet, 22 layers network

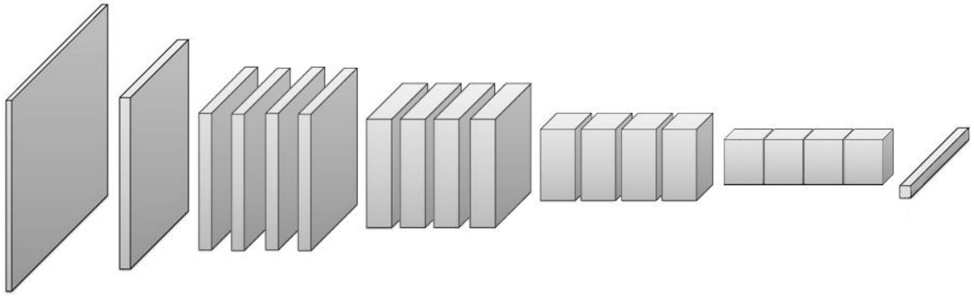
U. of Toronto, SuperVision, a 7 layers network

human error is around *5.1%* on a subset

FACE DESCRIPTOR

Deep CNN model

Millions of face images
with person IDs



+

- Pre-training networks for the face classification task
- Fine-tuning face descriptors with similarity loss
- Network architectures with multiple image resolutions
- Full training with 4 NVIDIA Titan X GPUs done in 2 days enables fast development, optimization and deployment of new models
- Training in this setup on CPU is infeasible for us

HOW TO MEASURE SUCCESS?

Same faces if $D_{i,j} < T$

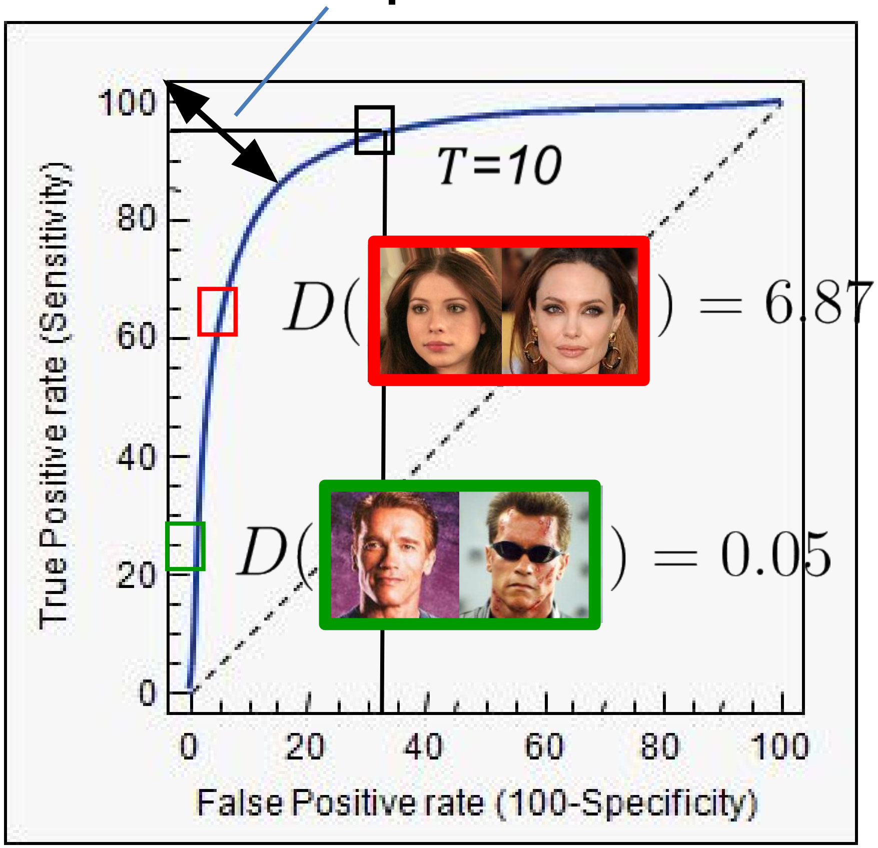
True Positive Rate

$$\frac{\# \; correct \; (T)}{\# \; all \; correct}$$

False Positive Rate

$$\frac{\# \; wrong \; (T)}{\# \; all \; wrong}$$

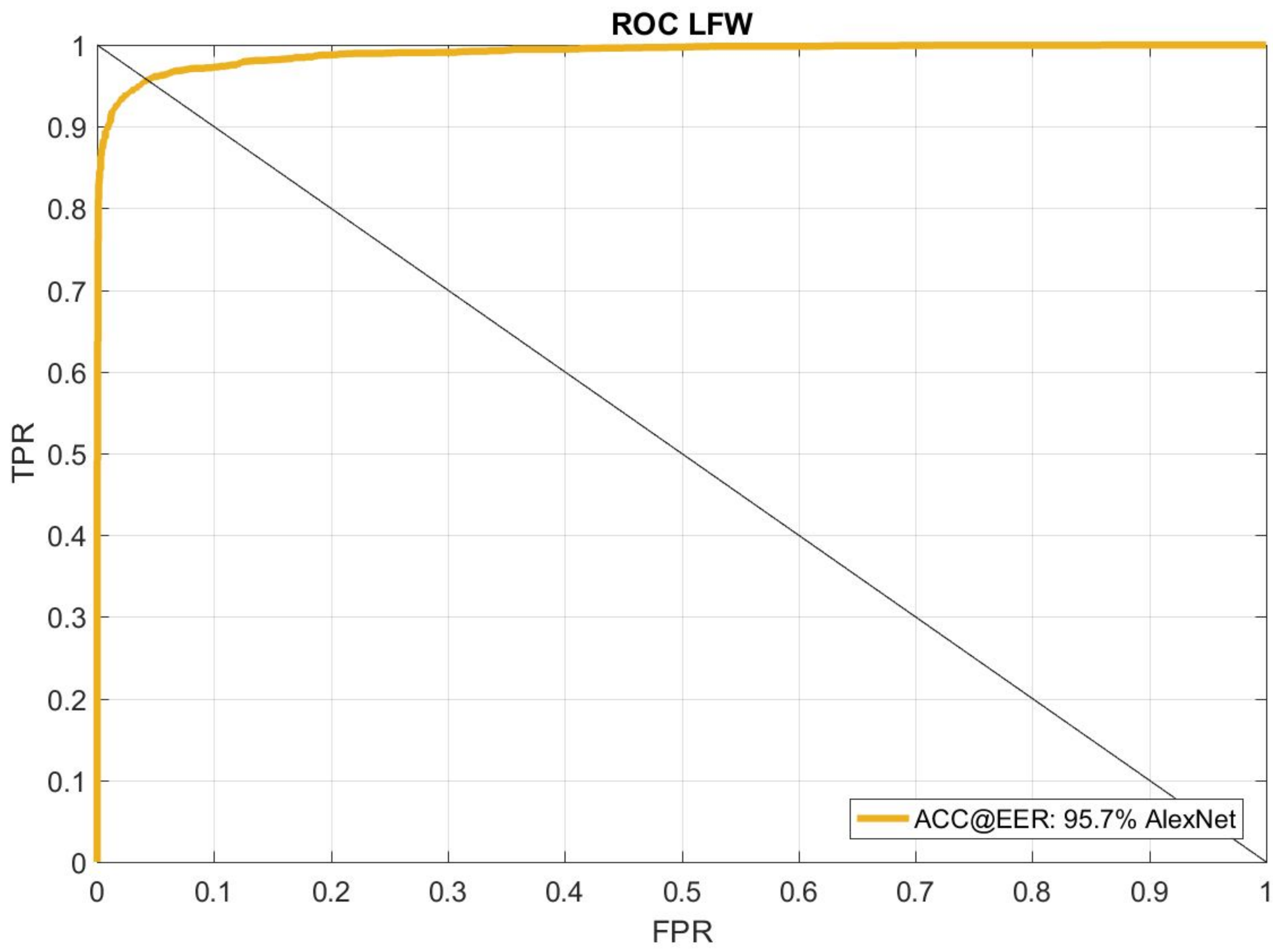
Equal error rate



ROC curve

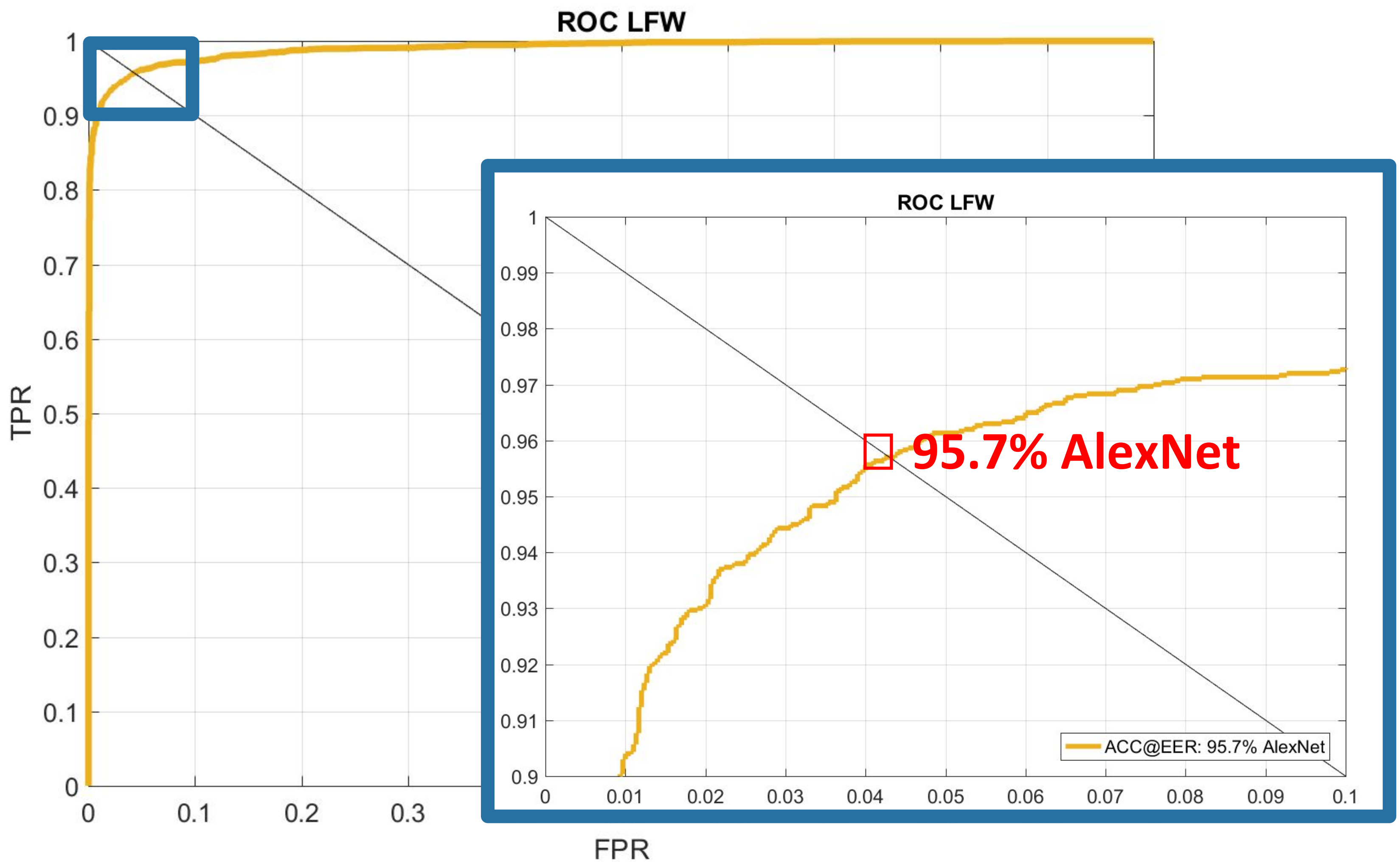
http://en.wikipedia.org/wiki/Receiver_operating_characteristic

TEST ON LFW BENCHMARK



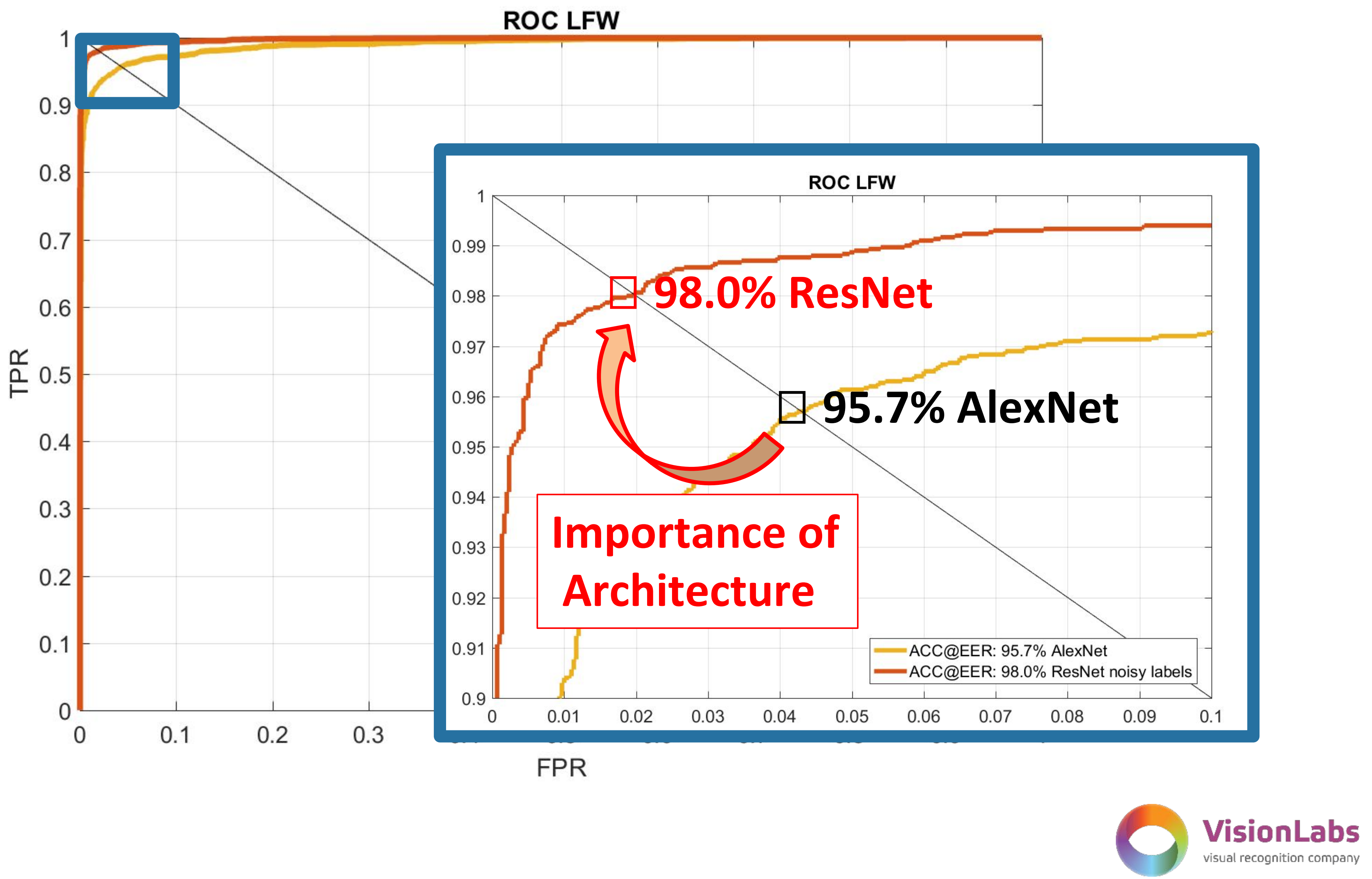
VisionLabs
visual recognition company

TEST ON LFW BENCHMARK



VisionLabs
visual recognition company

IMPROVEMENT ON LFW BENCHMARK

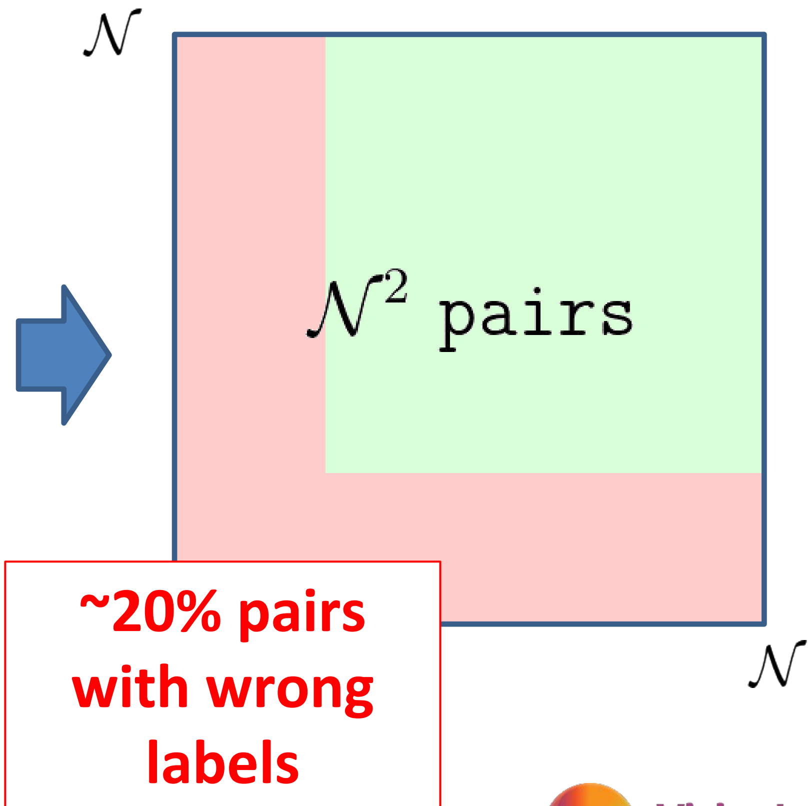


TRAINING DATA

Automatically collected
large-scale face datasets
from the Web

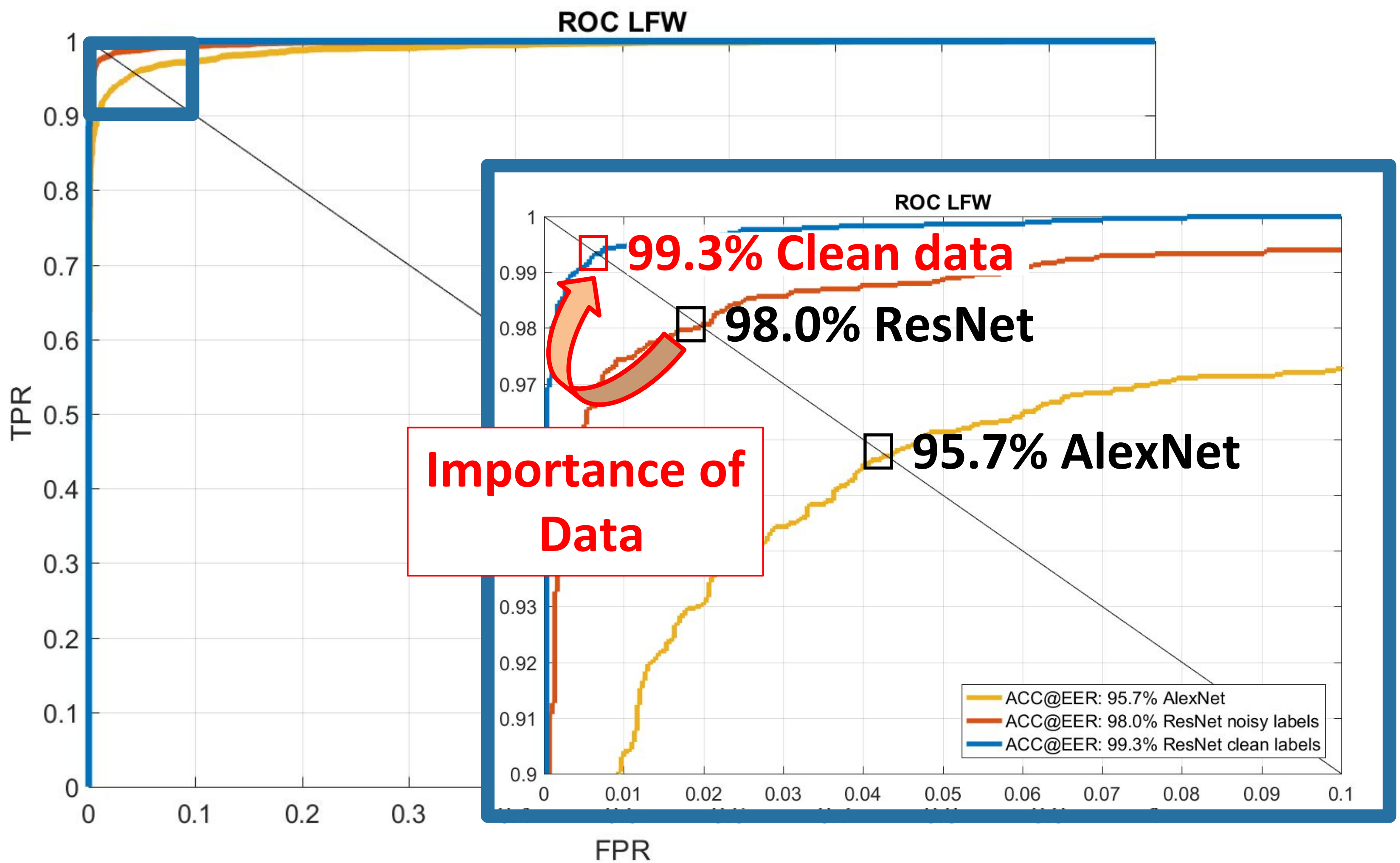
$$\mathcal{N} = 1\ 000\ 000$$

Training with pair-wise
similarity loss
suffers from wrong labels



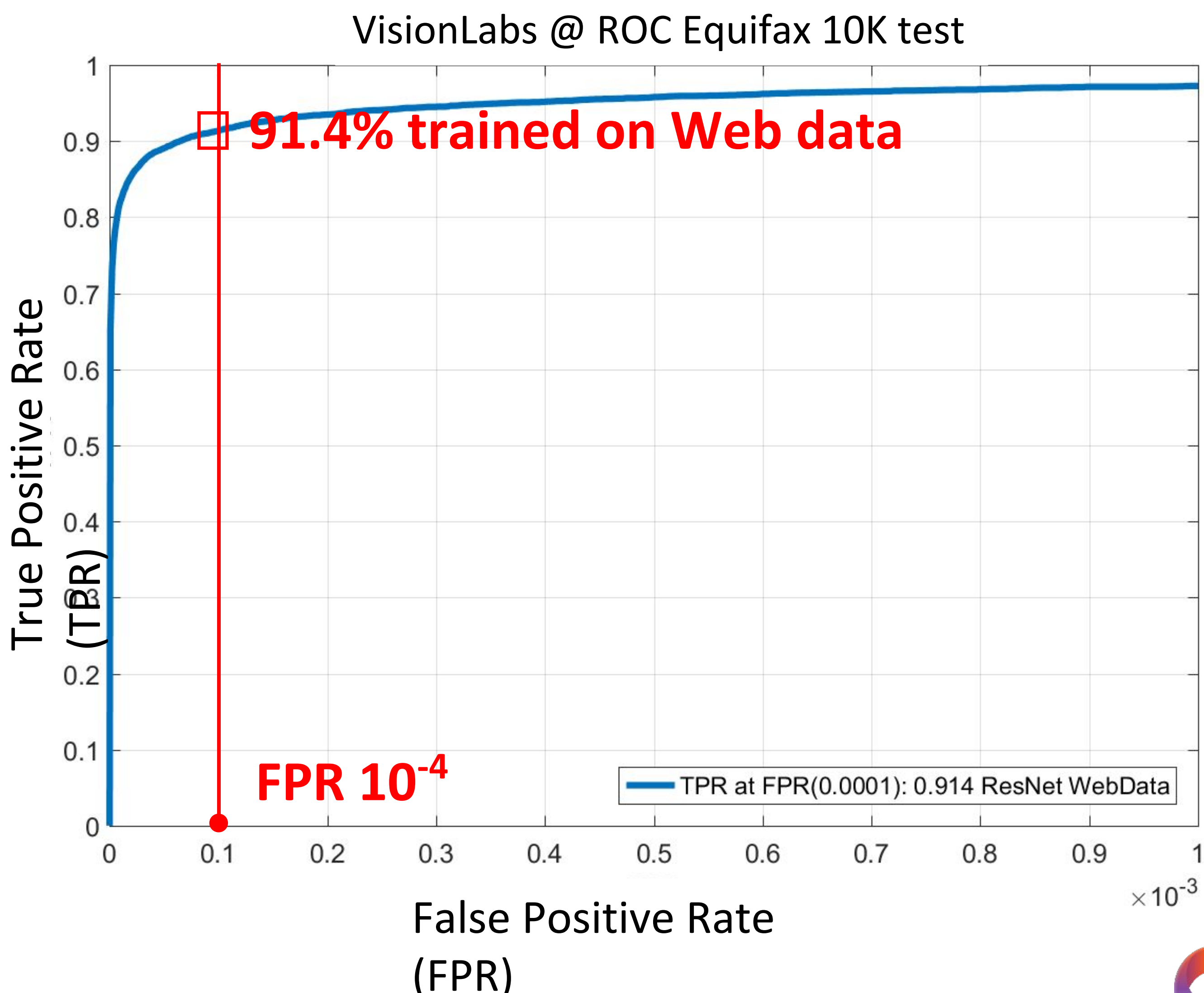
VisionLabs
visual recognition company

IMPROVEMENT ON LFW BENCHMARK



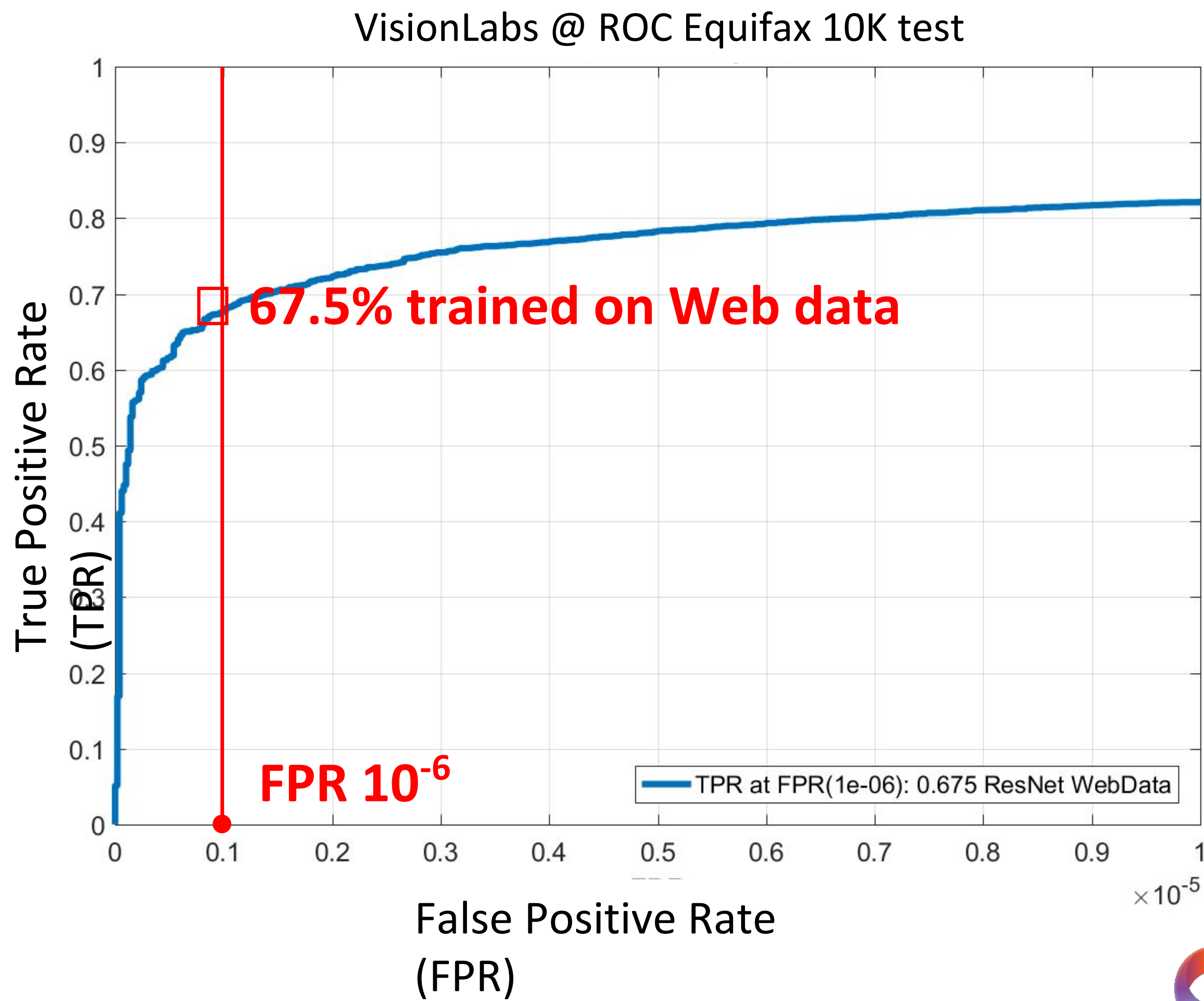
VisionLabs
visual recognition company

LARGE DATASETS: NEED HIGH TPR AT LOW FPR



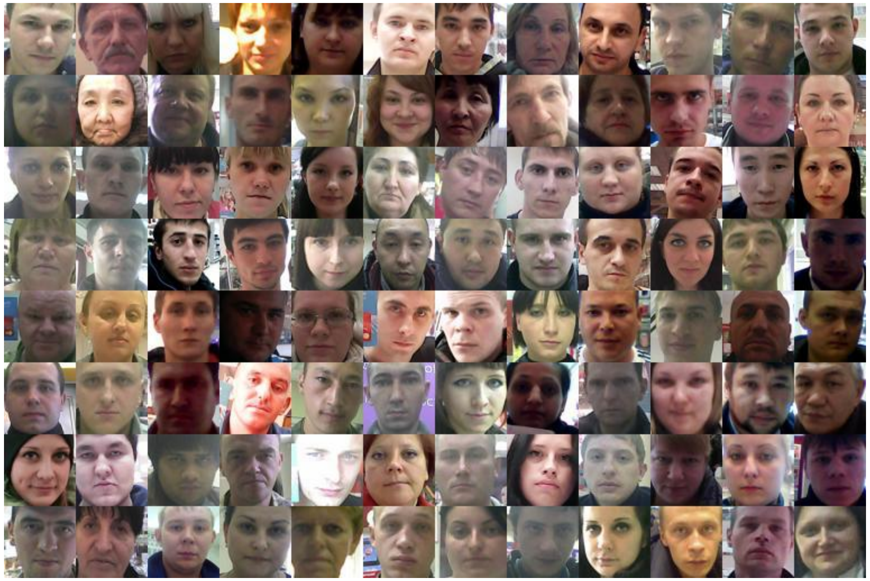
VisionLabs
visual recognition company

LARGE DATASETS:
NEED HIGH TPR AT EXTREMELY LOW FPR



UNIQUE TRAINING DATA

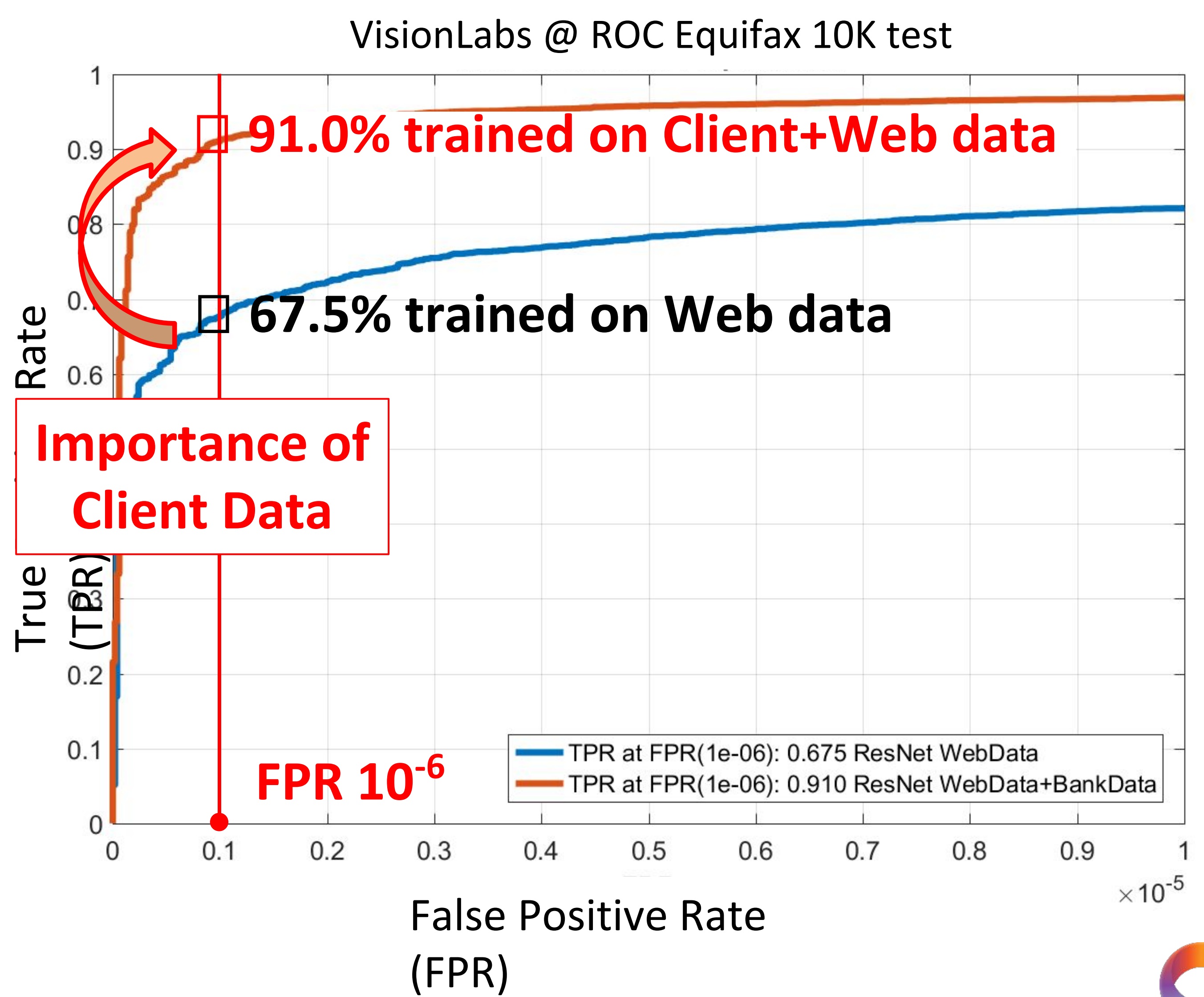
1. Real labeled face images from financial sector provided by our partner - Equifax



2. Celebrities images are collected and labeled by ourselves

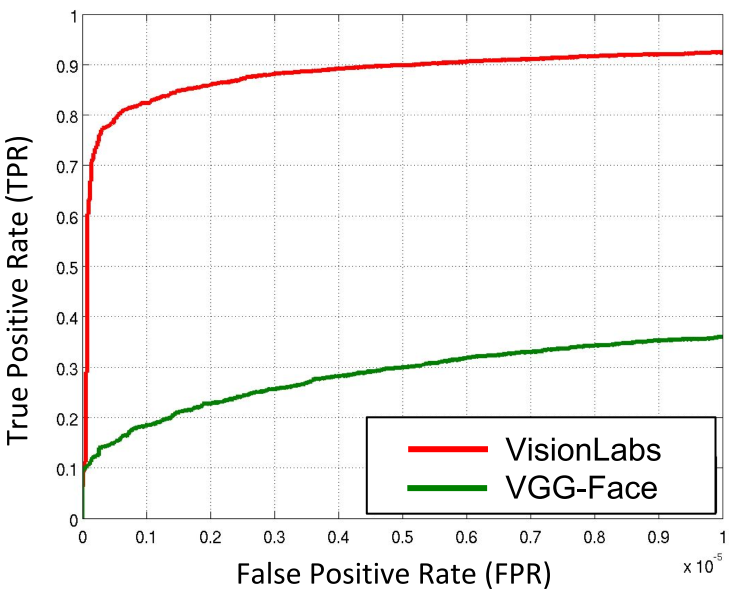
VisionLabs
visual recognition company

LARGE DATASETS:
NEED HIGH TPR AT EXTREMELY LOW FPR

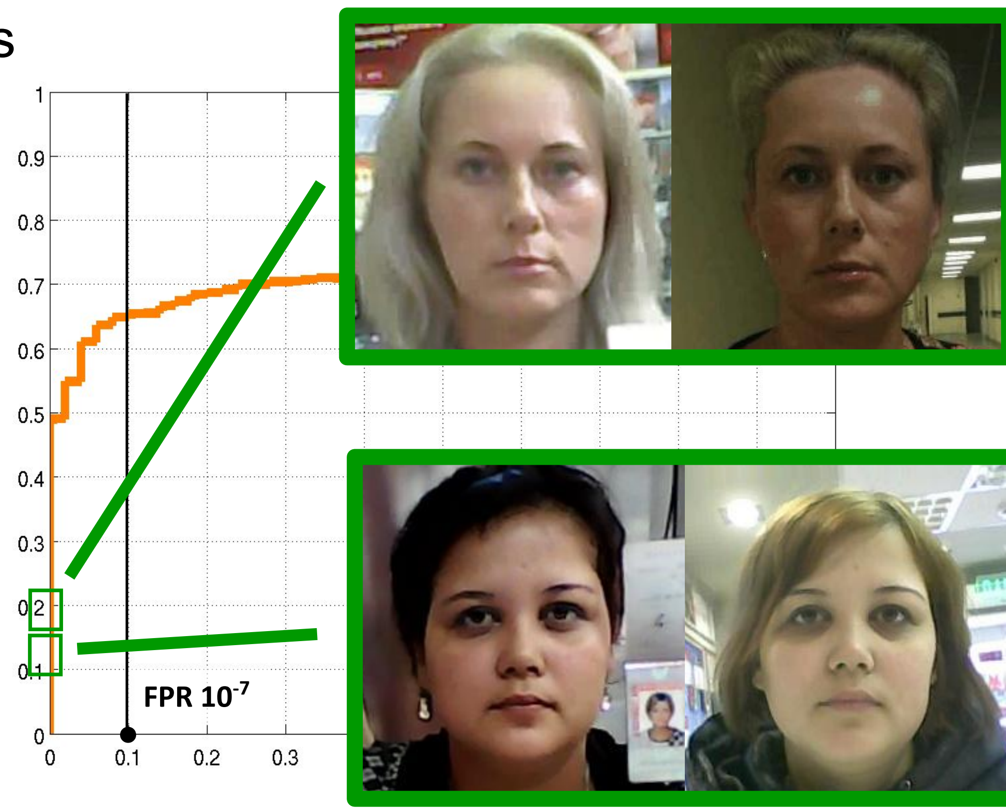


Face descriptors

- Compact: 160 bytes
  binary descriptors
- Fast matching: 40ms for 1M faces
  1 core Intel CPU E5-2650
- Accurate: ~82% TPR at 10^{-6} FPR
  on client image data

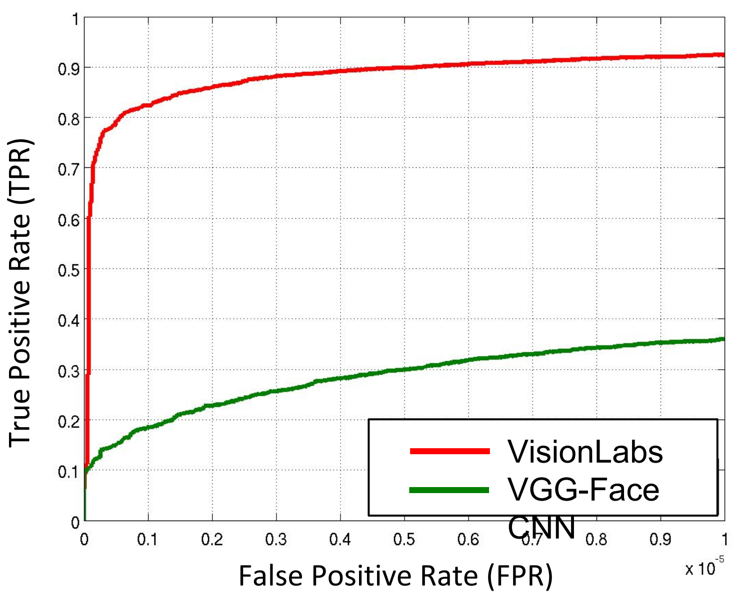


- Robust to hair style

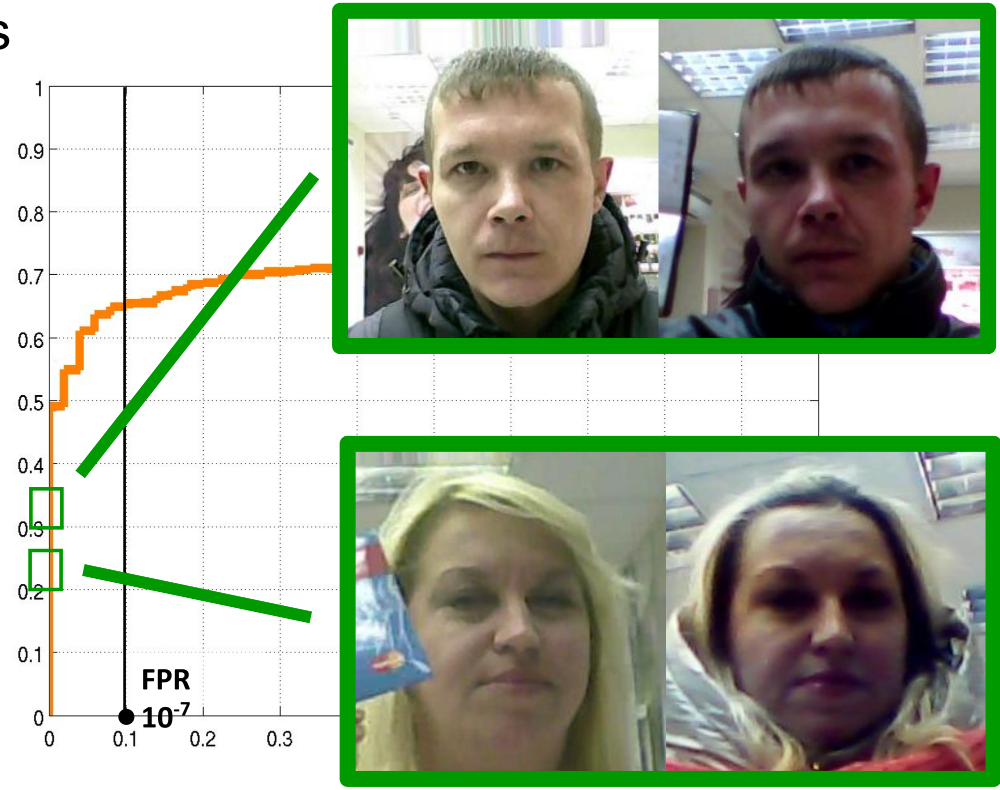


Face descriptors

- Compact: 160 bytes
  binary descriptors
- Fast matching: 40ms for 1M faces
  1 core Intel CPU E5-2650
- Accurate: ~82% TPR at 10^{-6} FPR
  on client image data

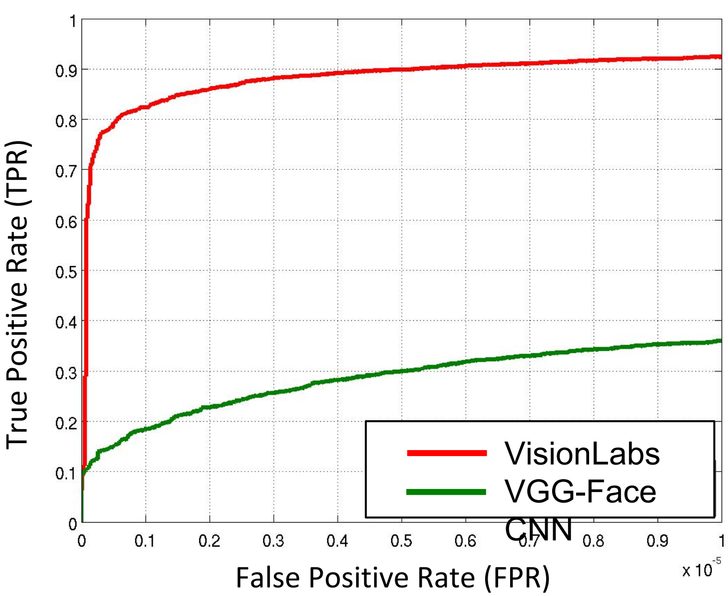


- Robust to poor lighting

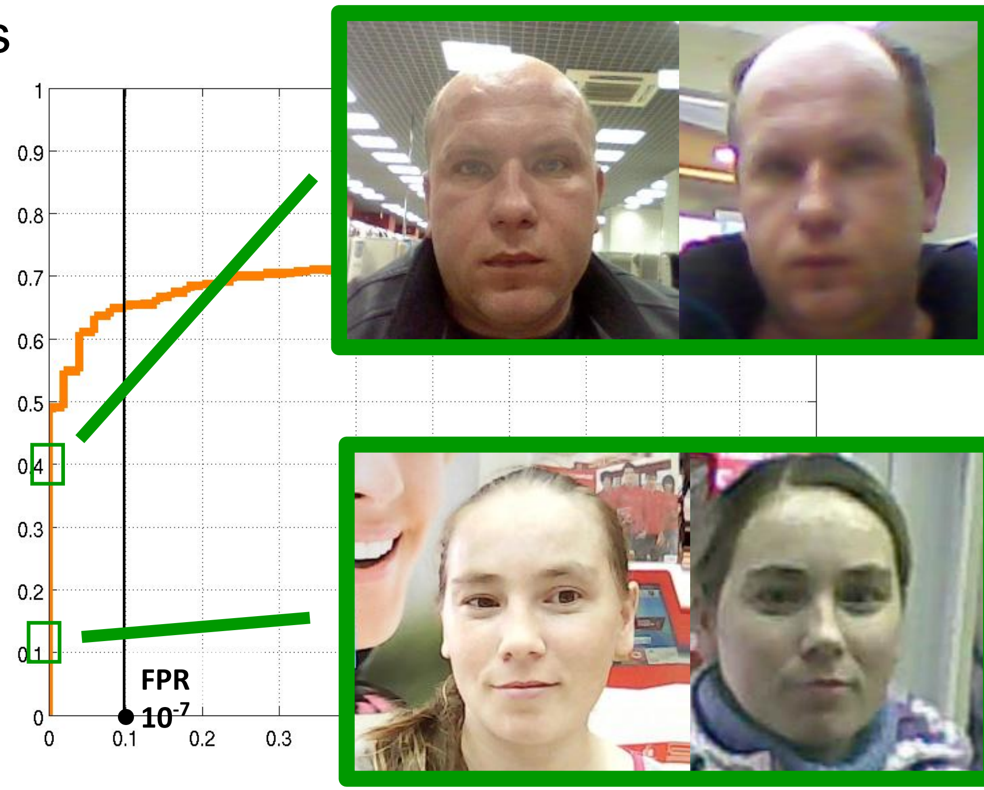


Face descriptors

- Compact: 160 bytes
  - binary descriptors
- Fast matching: 40ms for 1M faces
  - 1 core Intel CPU E5-2650
- Accurate: ~82% TPR at 10^{-6} FPR
  - on client image data

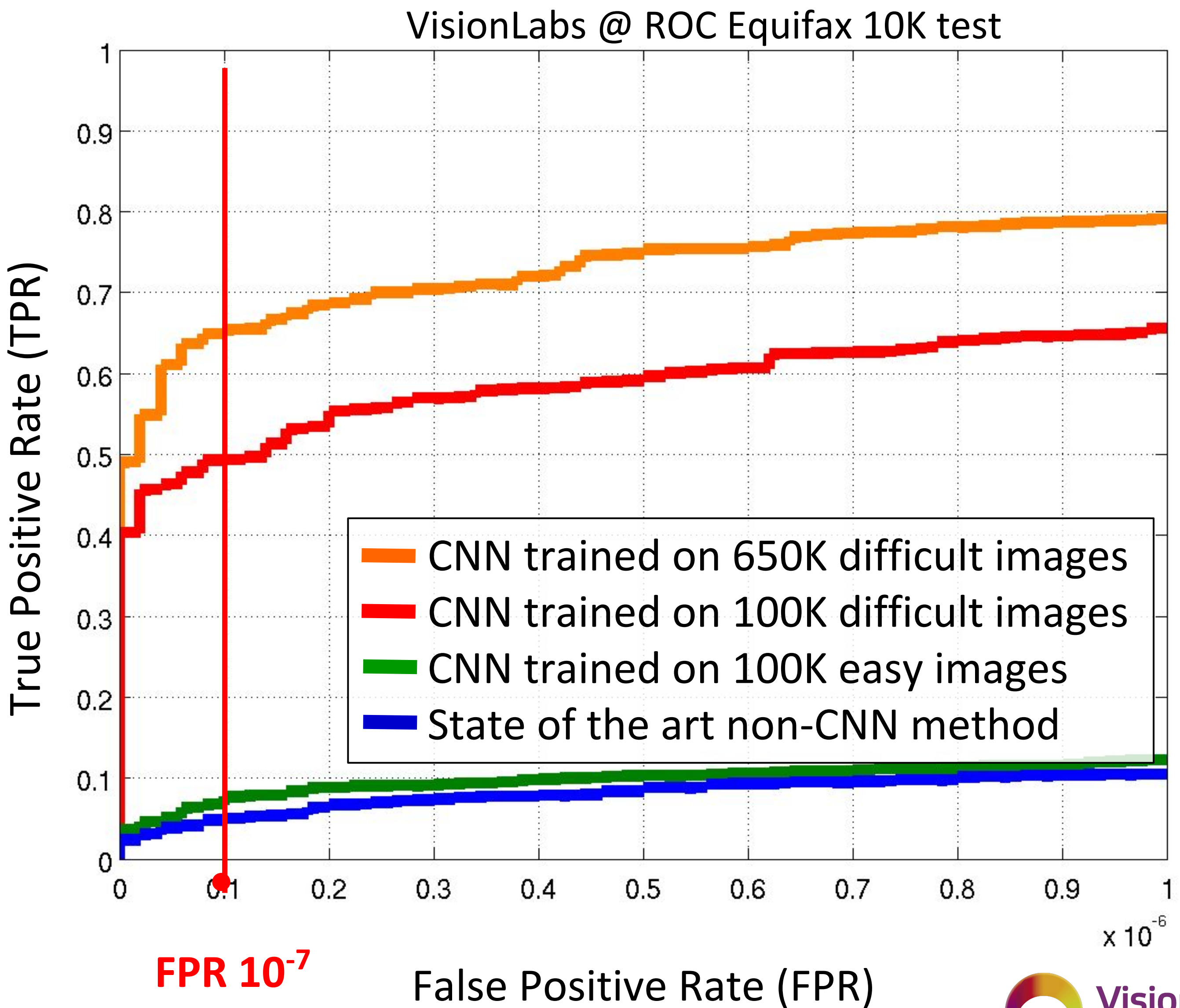


- Robust to low image quality



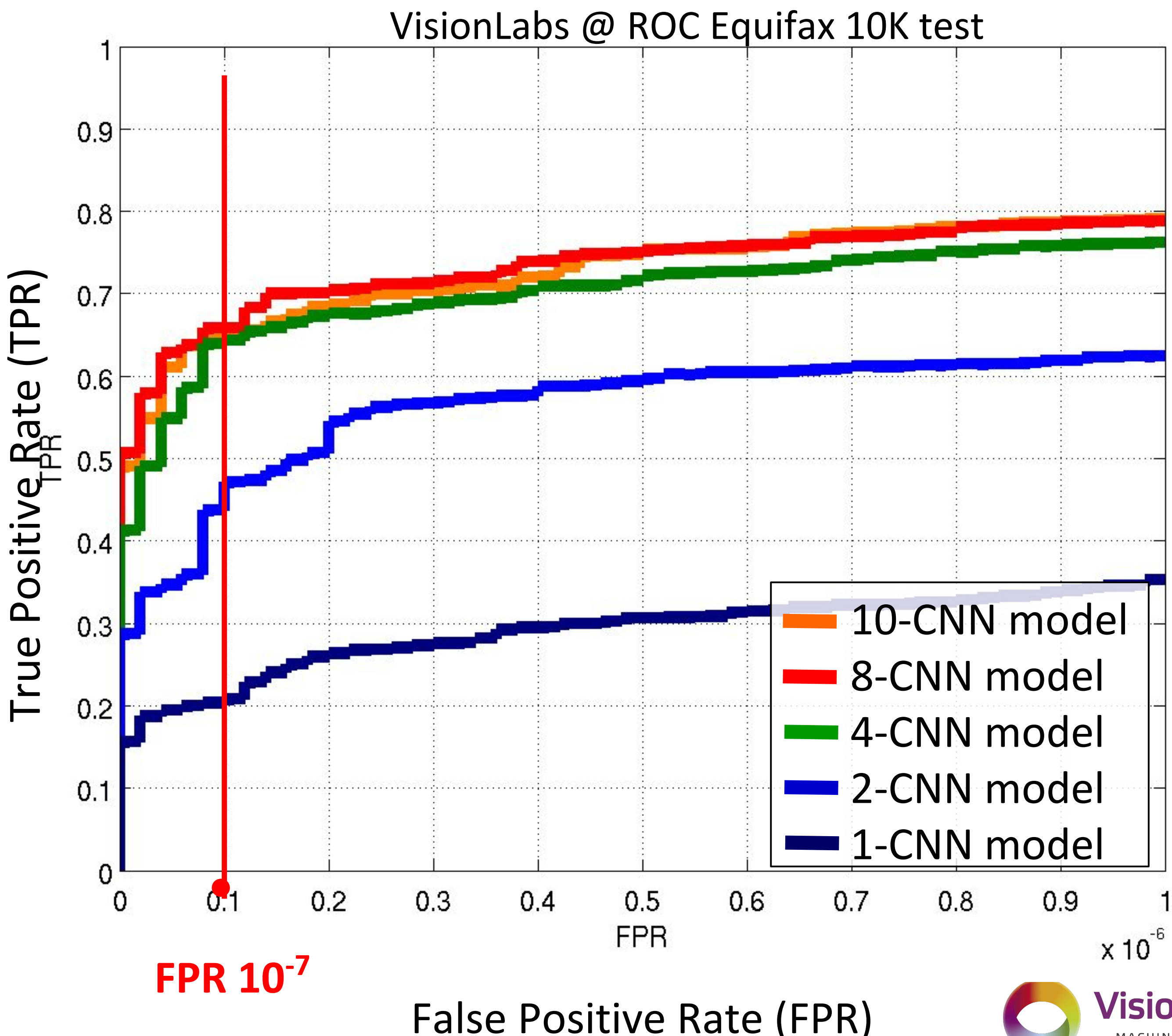
VisionLabs
MACHINES CAN SEE

IMPORTANCE OF TRAINING DATA



VisionLabs
MACHINES CAN SEE

IMPORTANCE OF MODEL COMPLEXITY

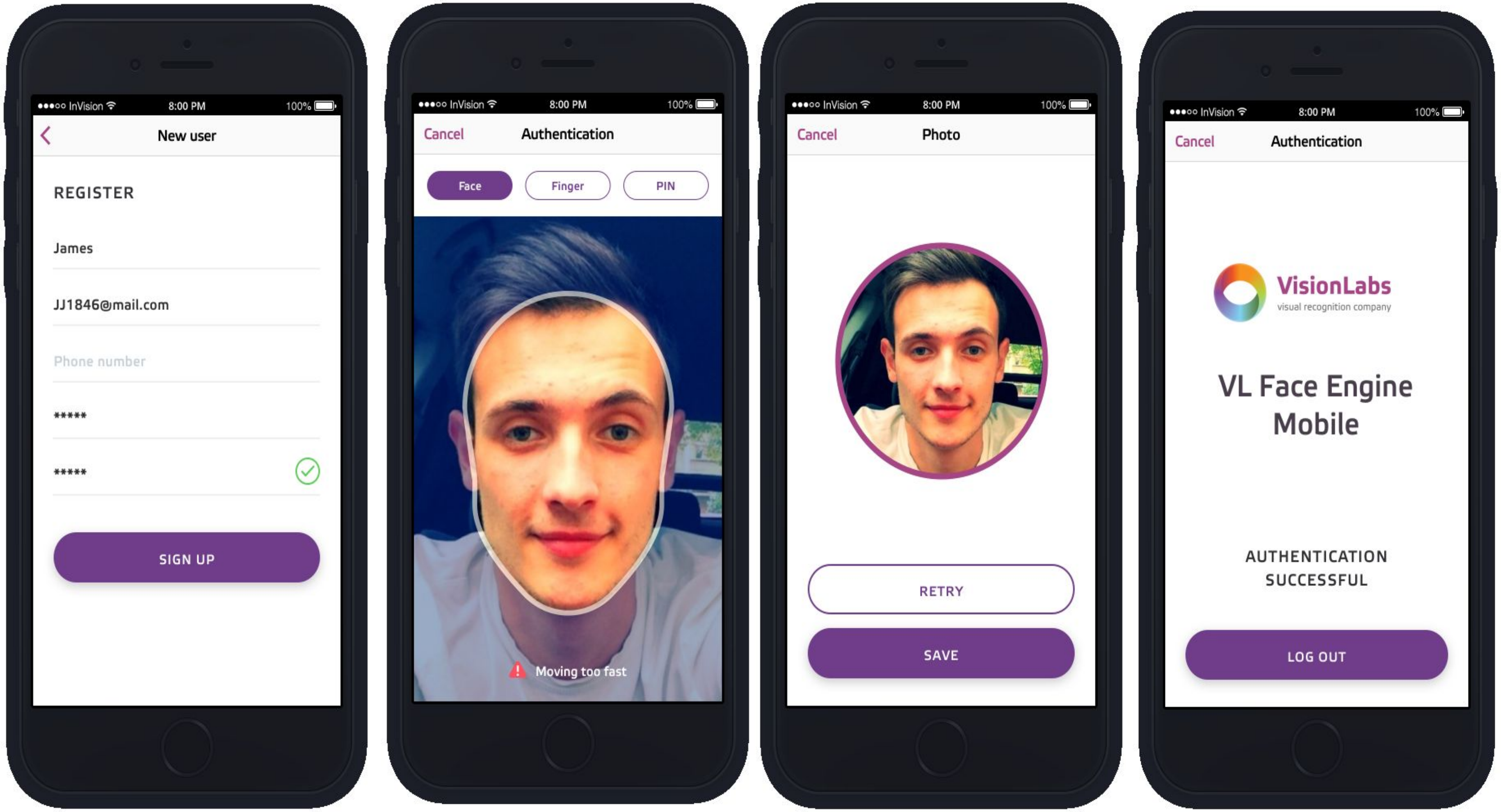


VisionLabs
MACHINES CAN SEE

Challenges

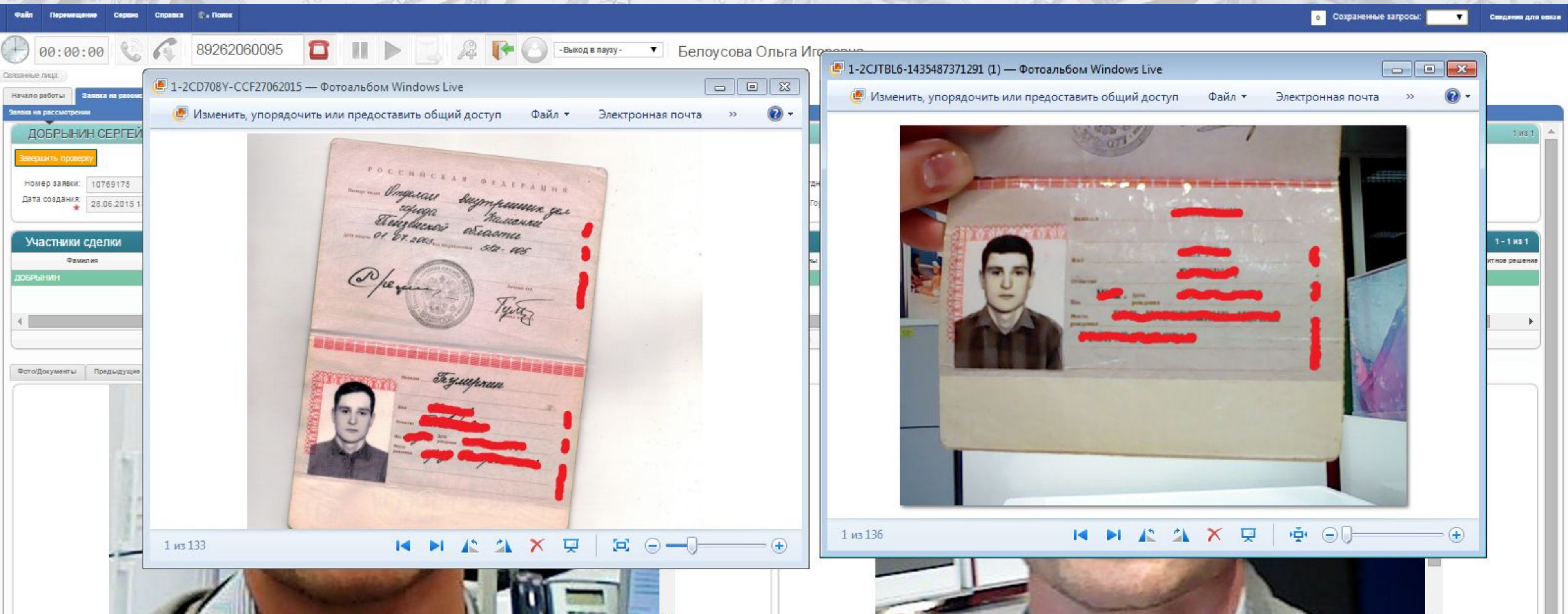
- **Speed**
  - Accuracy is nice, but expensive
- **Mobile devices**
  - Expect up to 100x slower execution and deal with it

MOBILE AUTHENTICATION



3D AVATAR FROM SINGLE IMAGE

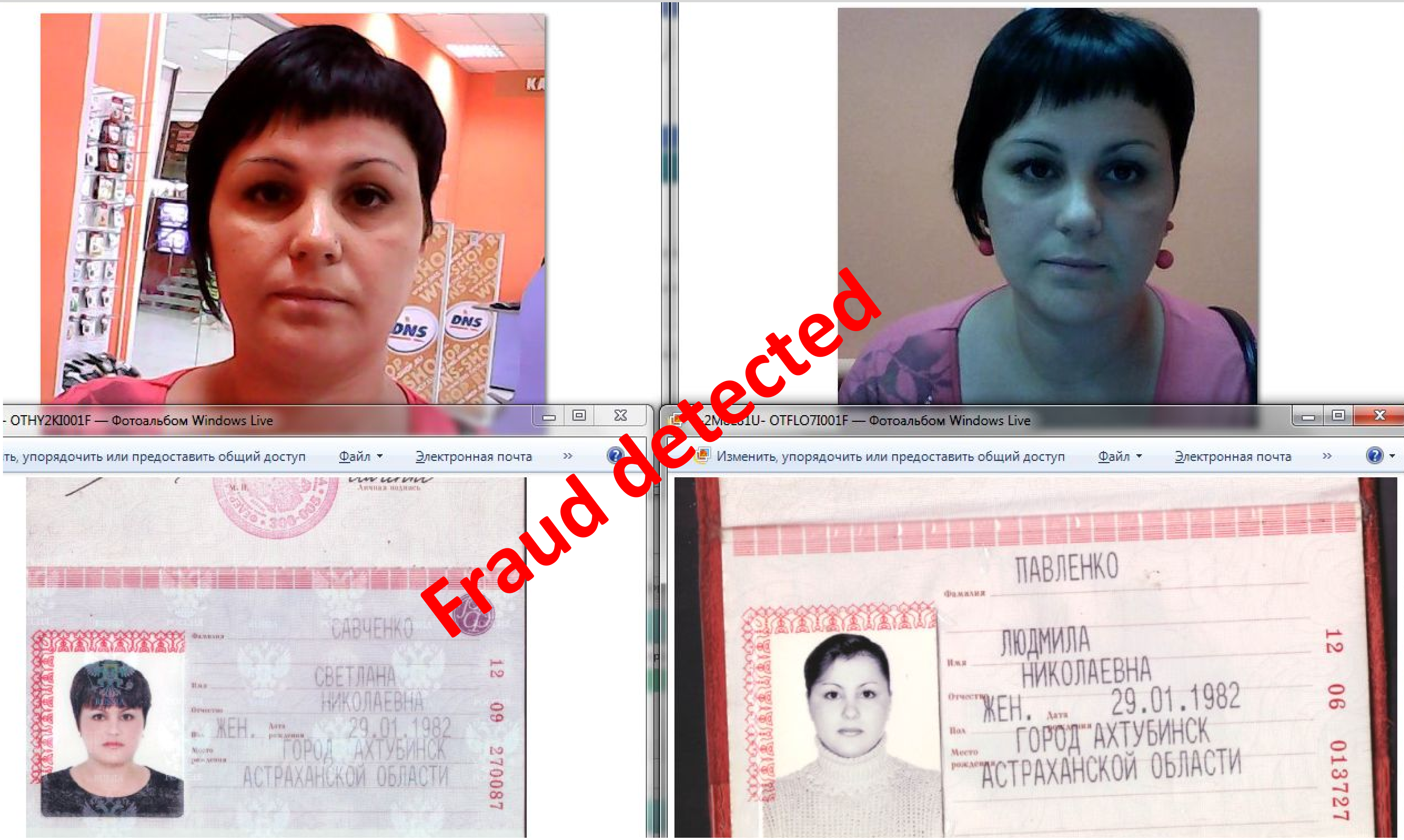
FRAUD PREVENTION



- Cut fraud losses by up to 90%
- Improve the quality of the loan portfolio
- Stop fraudulent employees

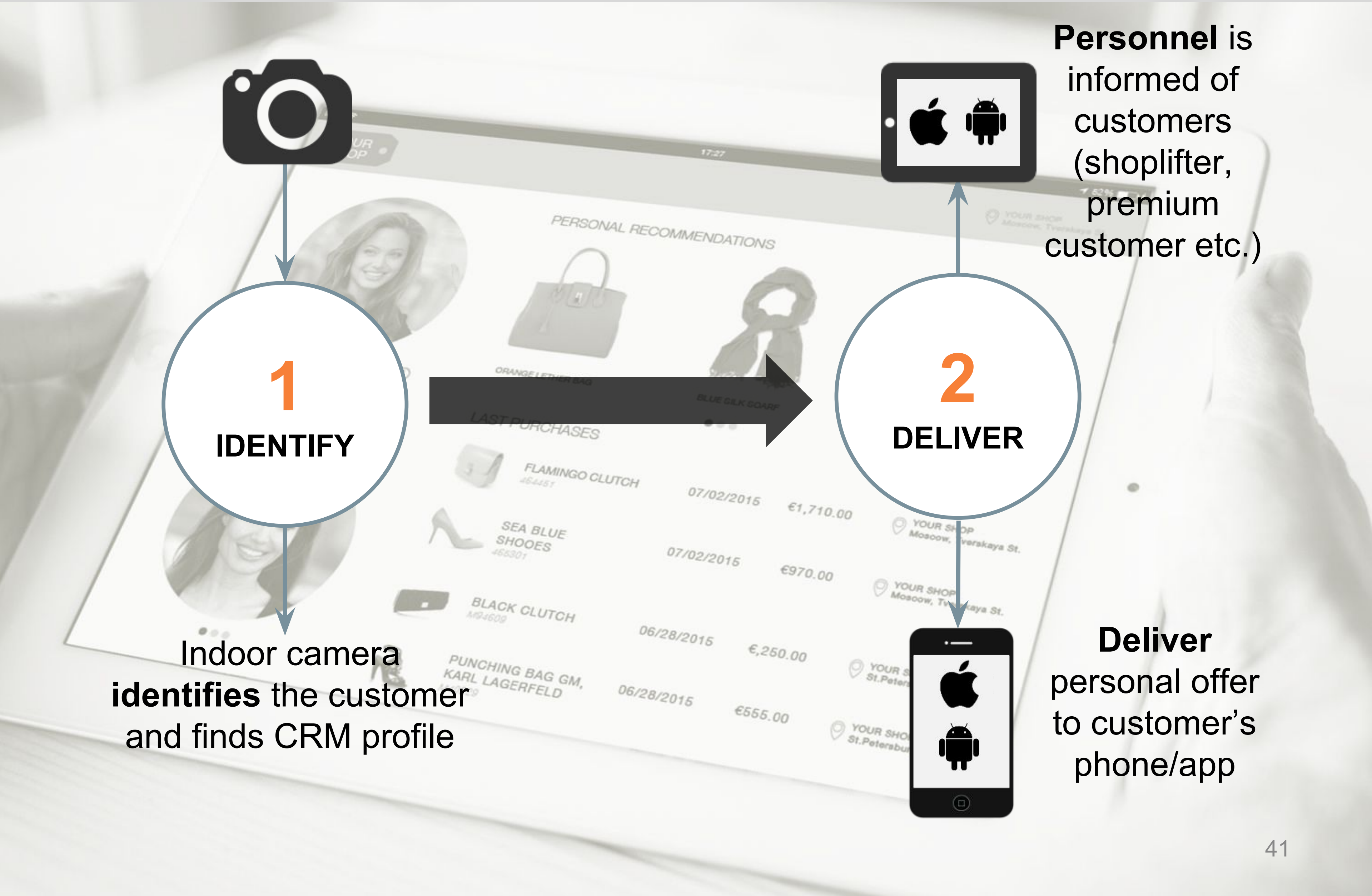
VisionLabs
MACHINES CAN SEE

FRAUD DETECTED



VisionLabs
MACHINES CAN SEE

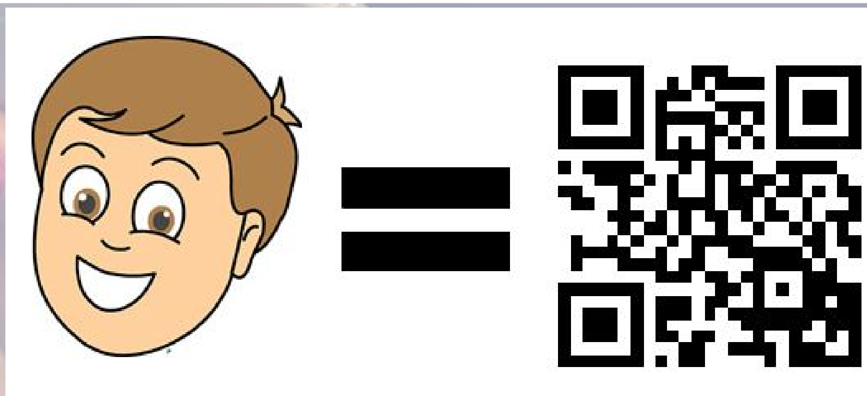
PERSONALIZED SERVICE



MULTIFACTOR AUTHENTICATION

1st FACTOR:
LOGIN + PASSWORD

2nd FACTOR:
LOGIN + FACE



Internal & External Secure Transactions

- "Something one knows" - a secret (PIN)
- "Something one has" - a passport (ID)
- "Something one is" - a biometric key

BLACK LISTING

- Match customers with database
- Inform if blacklisted people are recognized

VisionLabs MACHINES CAN SEE

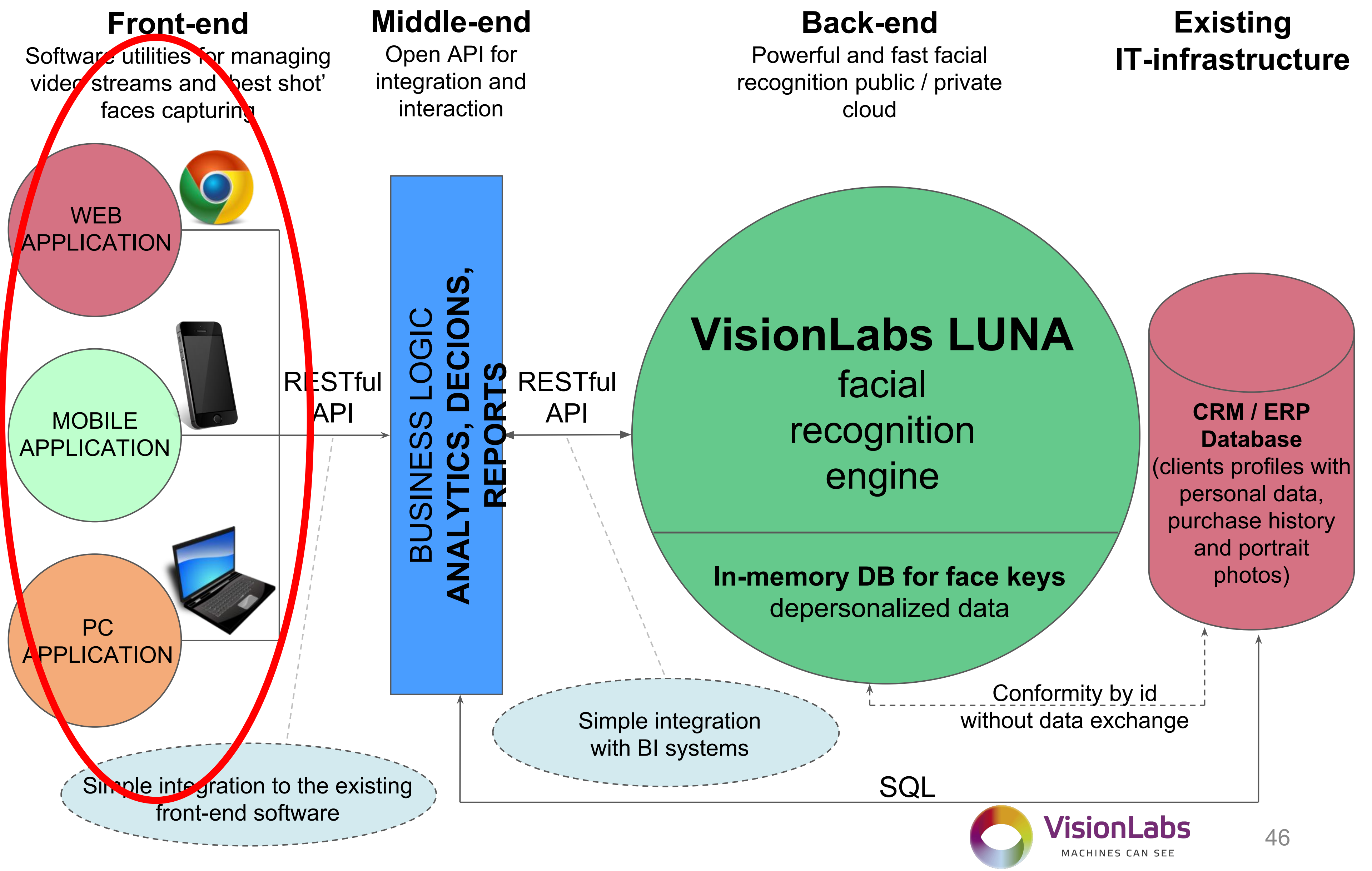
SEARCH PEOPLE

- Match criminals in historic data
- statistics of people behaviour

VisionLabs
MACHINES CAN SEE

APPLICATIONS

FACIAL RECOGNITION PLATFORM

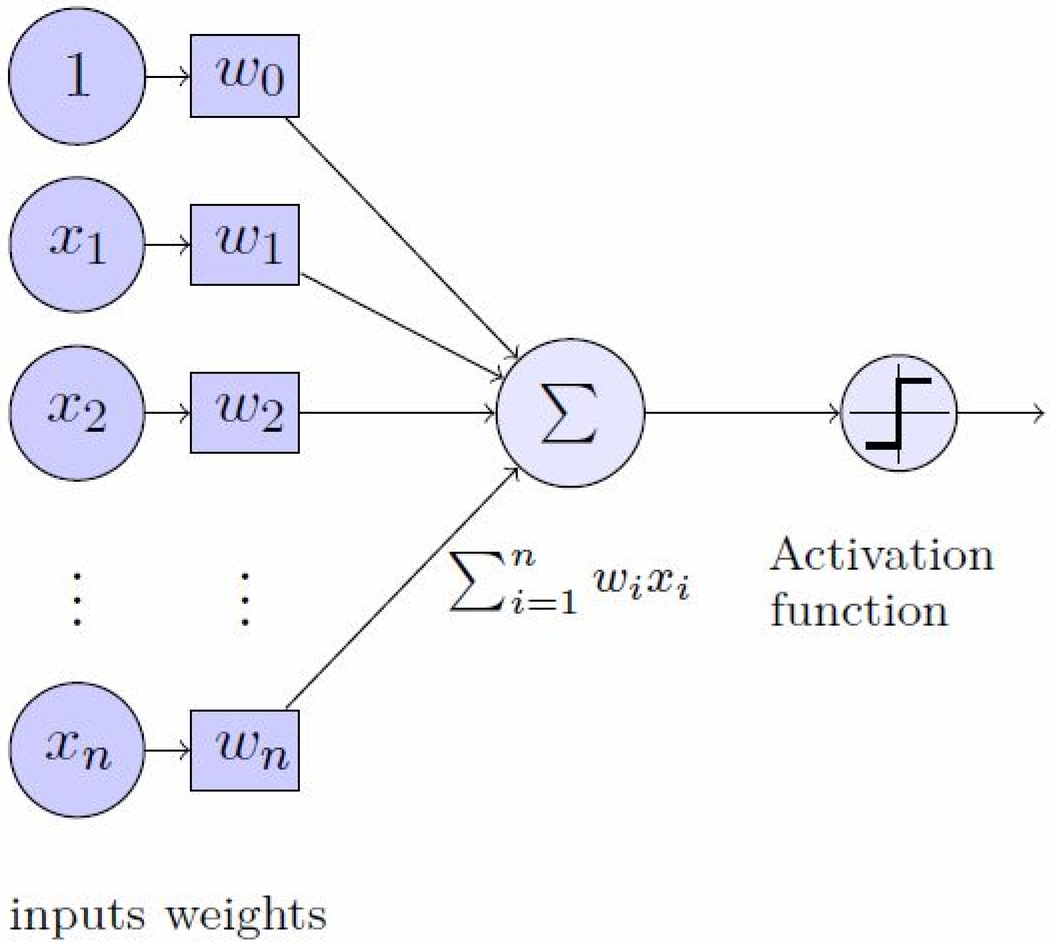


THANK YOU

E-MAIL: p.omenitsch@visionlabs.ru

WEB: www.visionlabs.ru

PERCEPTRON



1
w_0
x_1
w_1
x_2
w_2
x_n
w_n
Σ
$\sum_{i=1}^{n} w_i x_i$
Activation function
inputs weights

NEURAL NETWORK