Real time planetary scale face recognition system

N-TECH. LAB

Artem Kukharenko

N-TECH. LAB

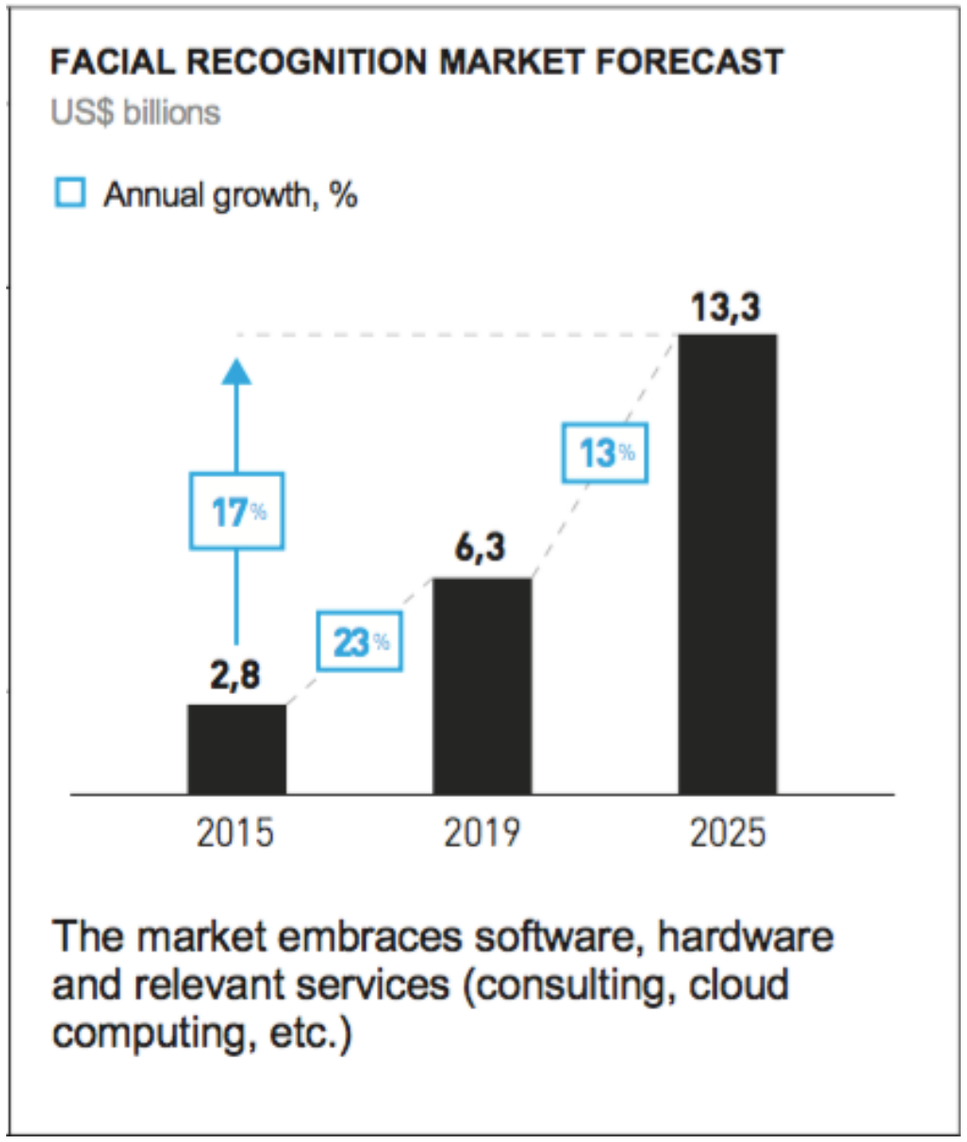
Agenda

- Face recognition market;
- Our algorithm;
- MegaFace challenge;
- FindFace — large scale search engine;
- Accuracy and timing at scale;
- Real-life applications.

N-TECH. LAB

The global facial recognition market

Source: Technavio; Stratistics MRC; Marketsandmarkets



N-TECH. LAB

Applications of face recognition

MAJOR APPLICATIONS		DESCRIPTION
GOVERNMENTAL	Face recognition in a crowd	Identification of blacklisted people in airports and other public places
	Road safety	Detection of traffic violations (e.g. bikers not wearing helmets, etc.)
	Identification of suspects	Identification/Identity verification of suspects in police stations
COMMERCIAL	Night clubs, casinos	Identification of blacklisted people Recognition of visitors' mood, average age and gender ratio to be displayed to prospective visitors
	Shopping	Advising products/services (food, clothing, etc.) based on specific characteristics
	Financial sector*	Advising relevant services (banking) Customer identity verification Payer identity verification and assessment of specific customer parameters
	Dating services	Searching for matches with pre-specified parameters/looking like someone else; photo-based profile search
	Businesses	Identification of employees and time tracking software
	Other	Driver drowsiness detection systems, automatic face sorting in Disneyland photos, etc.

N-TECH.
LAB

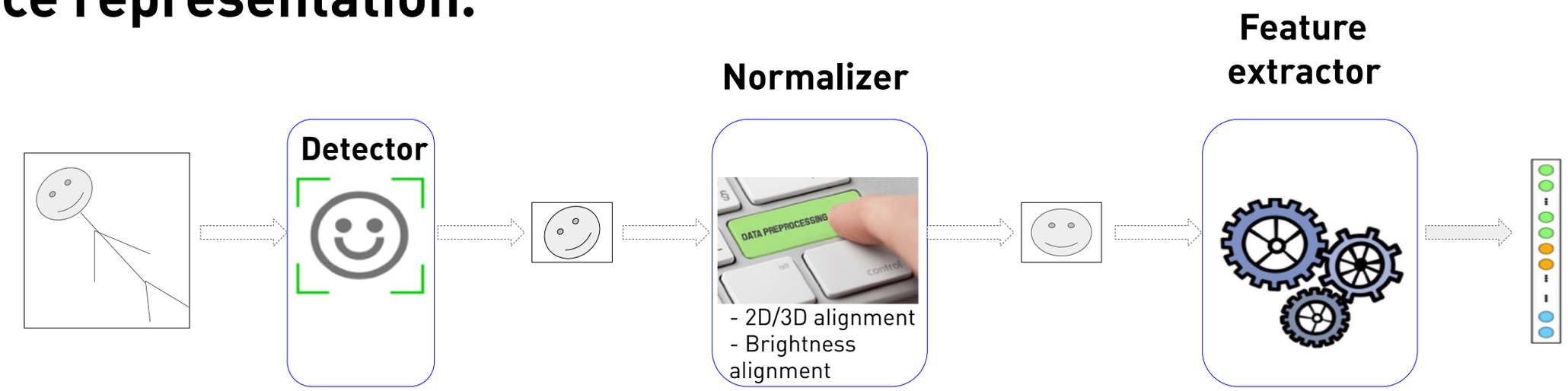
The advantages of the algorithm

- Best result in the Megaface contest — better than Google's FaceNet;
- Proven efficiency on very large databases;
- 73 % rank-1 on 1M dataset;
- 70+% rank-10 accuracy on 300M+ photos from the vk.com social network;
- Extremely low computational requirements (person's features vector takes less than 1 Kb);
- Query time less than a second (based on 300M+ photos).

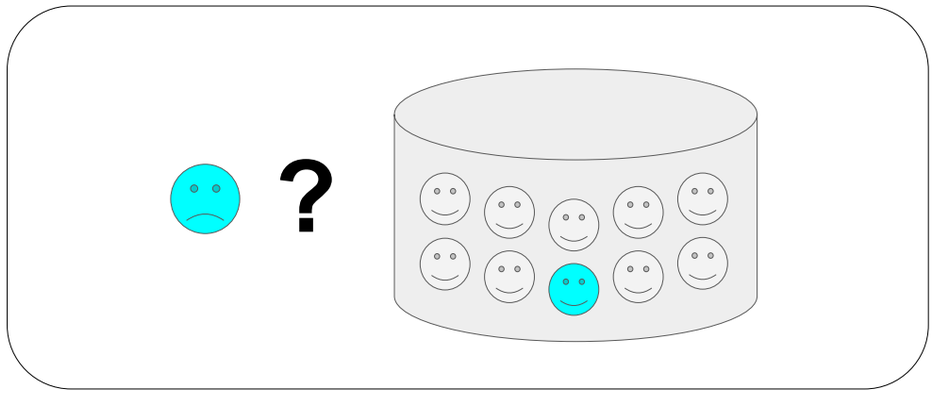
N-TECH. LAB

Face Recognition Pipeline

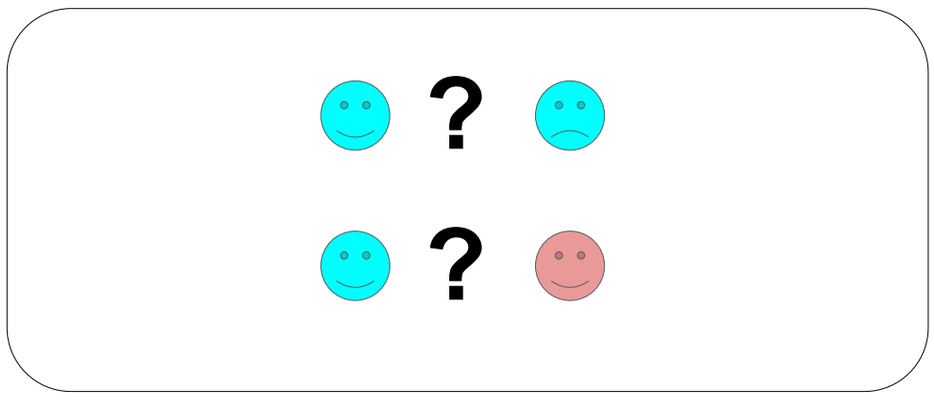
Face representation:



Scenarios:



Identification

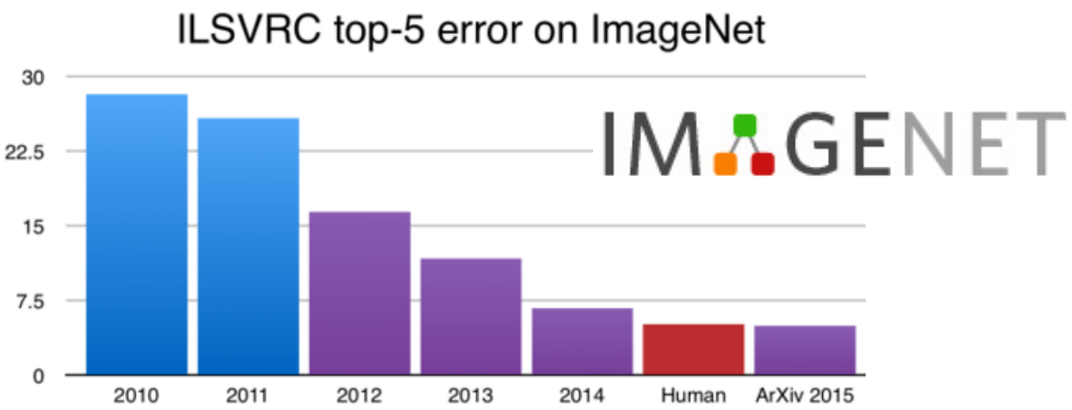


Verification

N-TECH. LAB

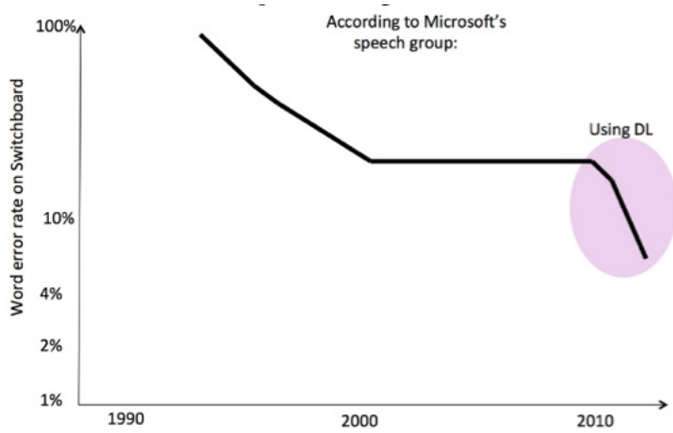
Deep Learning

Image recognition



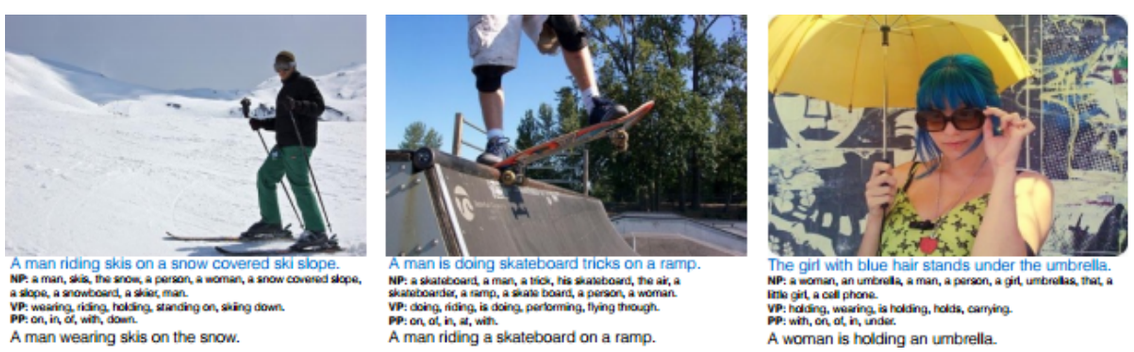
* Source: devblogs.nvidia.com

Speech recognition



* Source: Microsoft's speech group

Image captioning



Source: Rémi Lebret et. al. Phrase-based Image Captioning. ICML 2015.

Text analysis

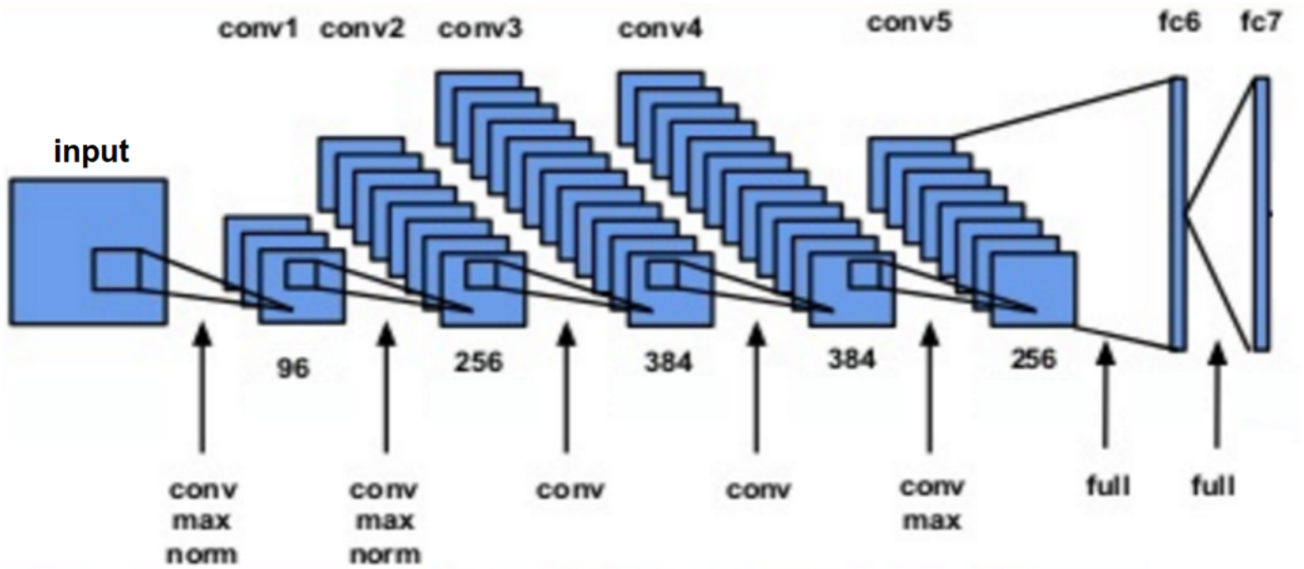
Relationship	Example 1	Example 2	Example 3
France - Paris	Italy: Rome	Japan: Tokyo	Florida: Tallahassee
big - bigger	small: larger	cold: colder	quick: quicker
Miami - Florida	Baltimore: Maryland	Dallas: Texas	Kona: Hawaii
Einstein - scientist	Messi: midfielder	Mozart: violinist	Picasso: painter
Sarkozy - France	Berlusconi: Italy	Merkel: Germany	Koizumi: Japan
copper - Cu	zinc: Zn	gold: Au	uranium: plutonium
Berlusconi - Silvio	Sarkozy: Nicolas	Putin: Medvedev	Obama: Barack
Microsoft - Windows	Google: Android	IBM: Linux	Apple: iPhone
Microsoft - Ballmer	Google: Yahoo	IBM: McNealy	Apple: Jobs
Japan - sushi	Germany: bratwurst	France: tapas	USA: pizza

Source: T.Mikolov. Efficient Estimation of Word Representations in Vector Space

N-TECH. LAB

Deep Learning. Training.

AlexNet architecture



Forward propagation

Backward propagation

Loss function

- **Multinomial logistic regression**

$$J(\theta) = -\frac{1}{m}\left[\sum_{i=1}^{m}\sum_{j=1}^{k} 1\{y^{(i)}=j\}\log\frac{\exp\left(\theta_j^T x^{(i)}\right)}{\sum_{l=1}^{k}\exp\left(\theta_l^T x^{(i)}\right)}\right]$$

- **Triplet loss**

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

- **Contrastive loss**

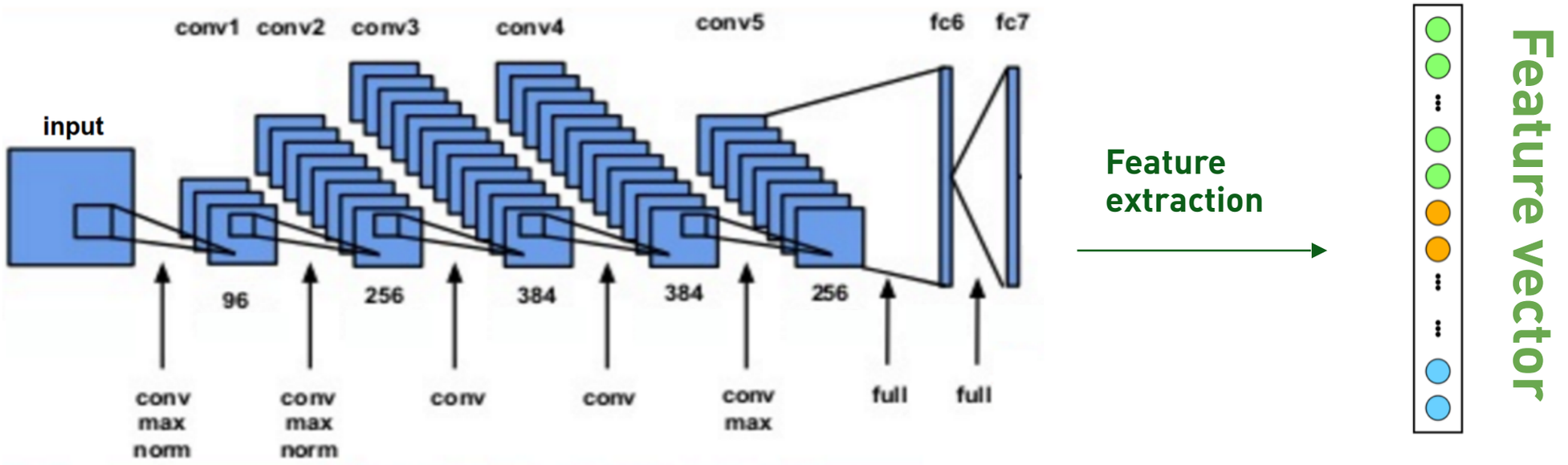
$$\begin{cases}\frac{1}{2}\|f_i-f_j\|_2^2 & \text{if } y_{ij}=1\\ \frac{1}{2}\max\left(0,m-\|f_i-f_j\|_2\right)^2 & \text{if } y_{ij}=-1\end{cases}$$

- 20M face photos for training;
- 3 weeks training on 3 GPUs NVidia Titan Black.

N-TECH.
LAB

Deep Learning. Enrollment.

AlexNet architecture



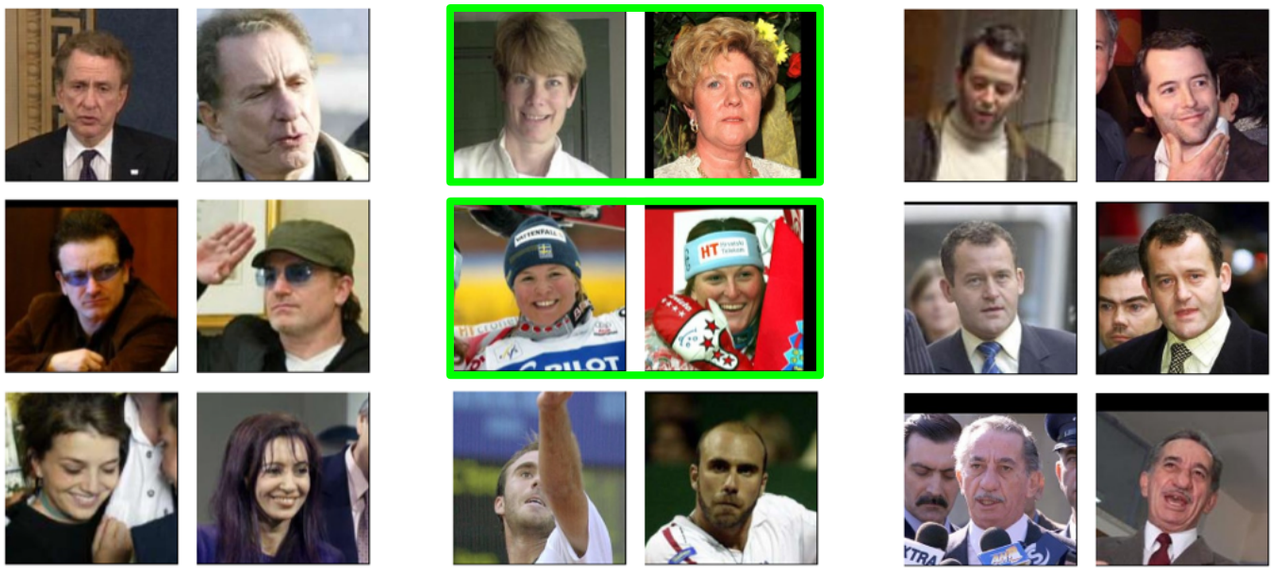
- Robustness to different shooting conditions (perspective, age, emotions so on..);
- Compact face representation (up to 16 floats);
- Ability to reusing for training additional classifier (gender, race, ...).

LFW

- Images collected from the web;
- 13K photos, 5K people;
- The only constraint on faces is that they were detected by the Viola-Jones detector;
- A number of algorithms achieve near to perfect accuracy;
- Need some bigger dataset!

Example errors:

False negative:



False positive:

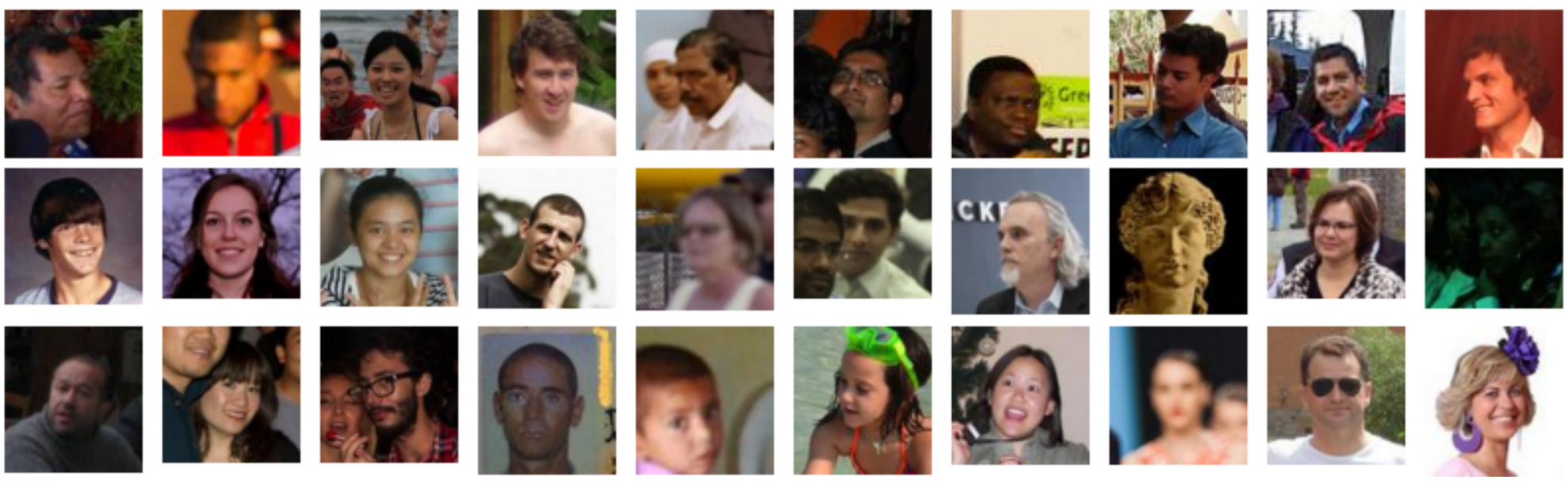
N-TECH. LAB

Megaface challenge

Sponsors

Google SAMSUNG

- Autumn 2015
- 1M identities
- Identification and Verification scenarios
- More than 100 teams participated



N-TECH. LAB

Identification Scenario

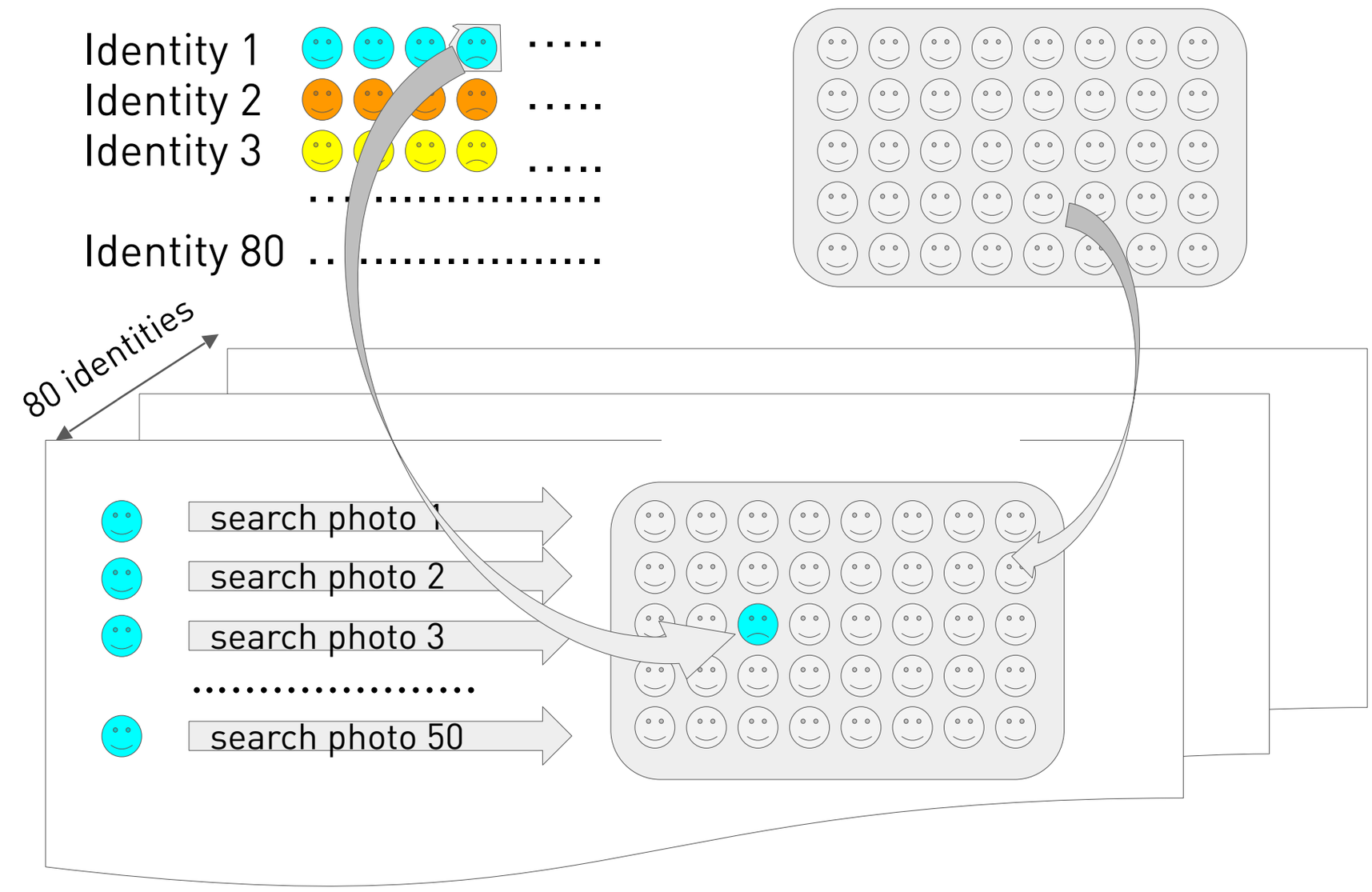
FaceScrub

80x50 photos

Identity 1 .....
Identity 2 .....
Identity 3 .....
.....
Identity 80 .....

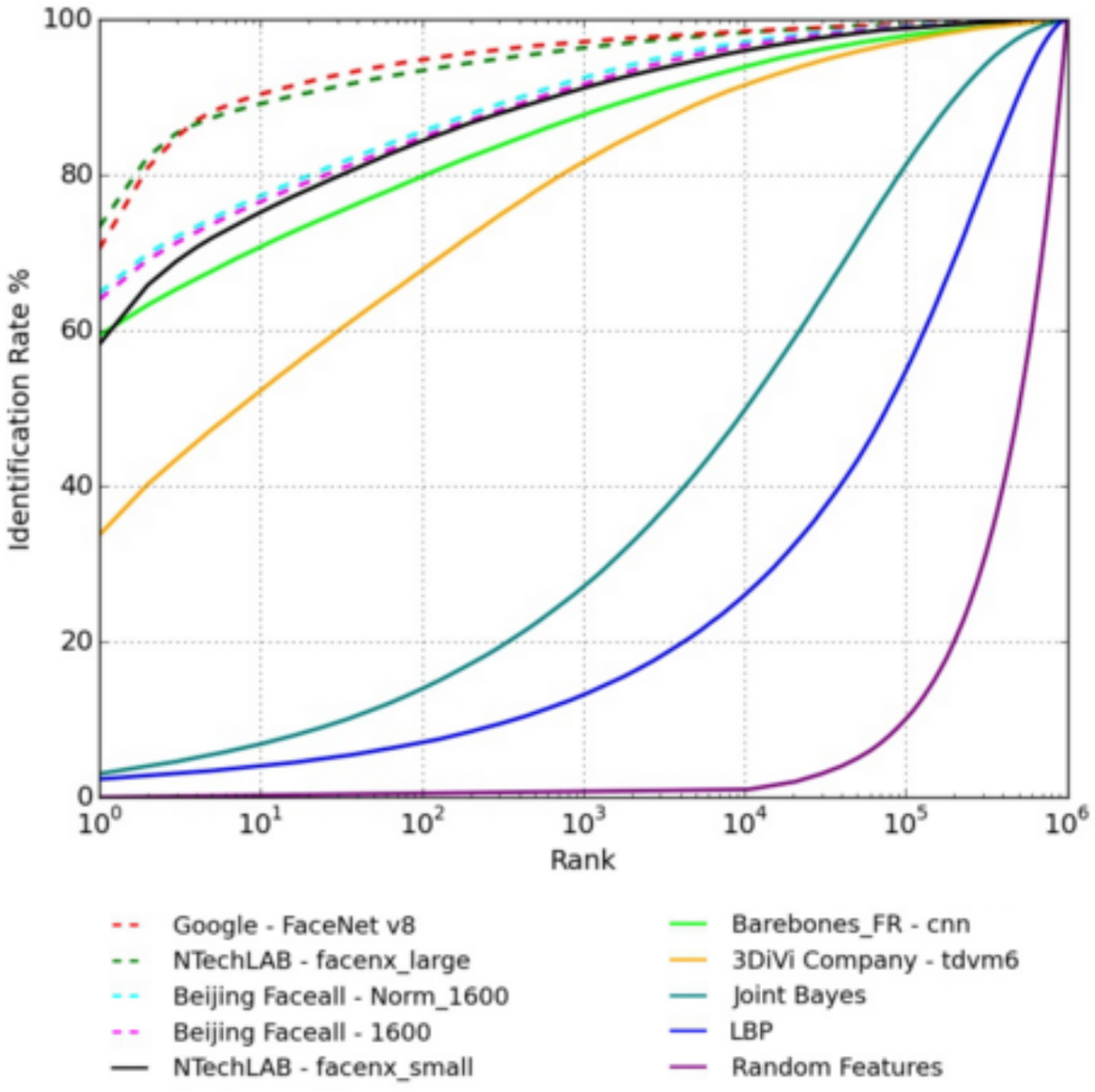
MegaFace

1M unlabeled photos



N-TECH. LAB

Identification Results

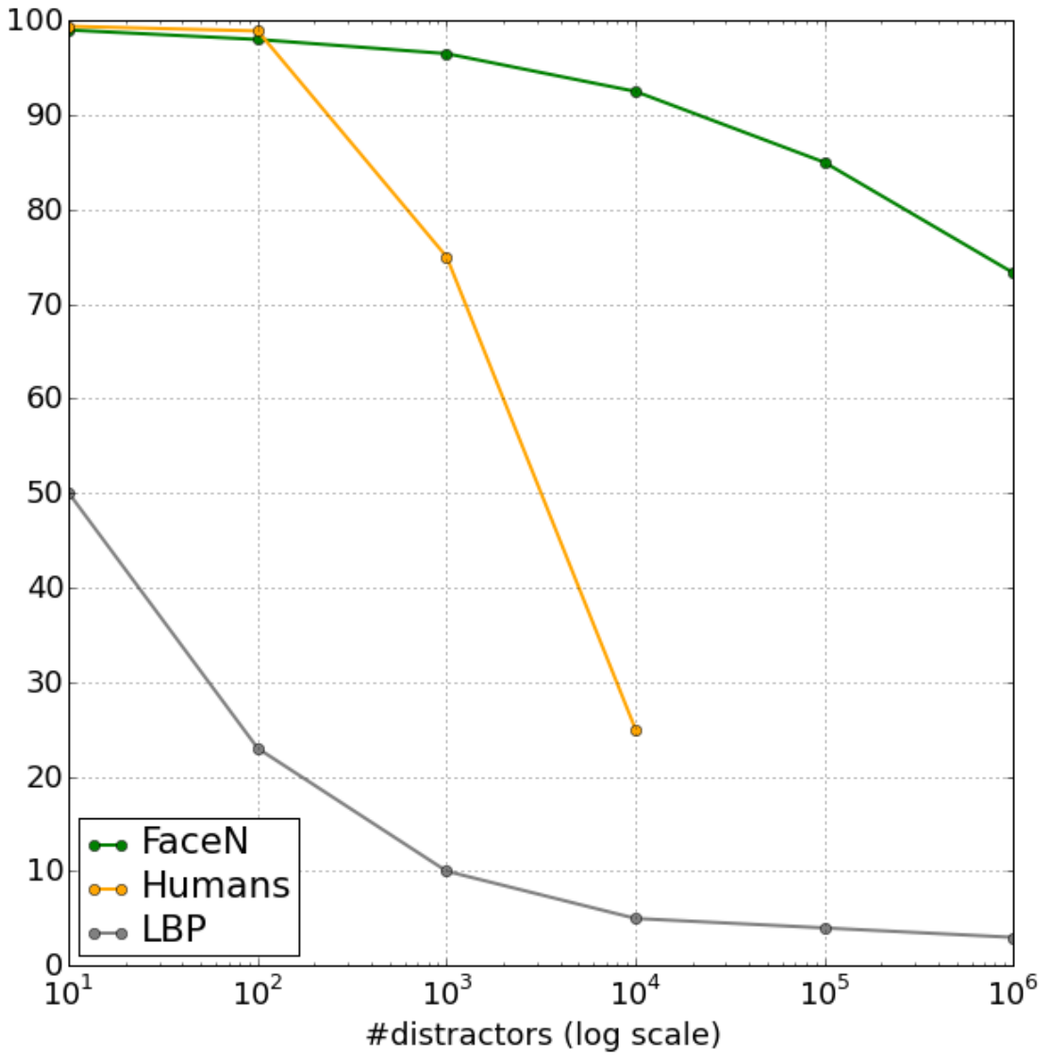


- - uses large training set

Algorithm	Set 1
NTechLAB - facenx_large	73.300%
Google - FaceNet v8	70.496%
Beijing Faceall Co. - FaceAll_Norm_1600	64.803%
Beijing Faceall Co. - FaceAll_1600	63.977%
Barebones_FR - cnn	59.363%
NTechLAB - facenx_small	58.218%
3DiVi Company - tdvm6	33.705%
Joint Bayes	3.021%
LBP	2.326%

N-TECH. LAB

Neural Net vs Human

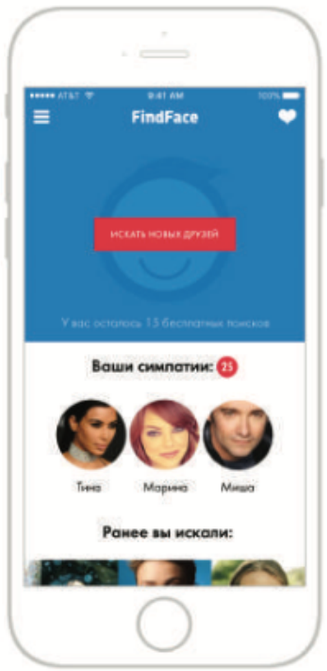


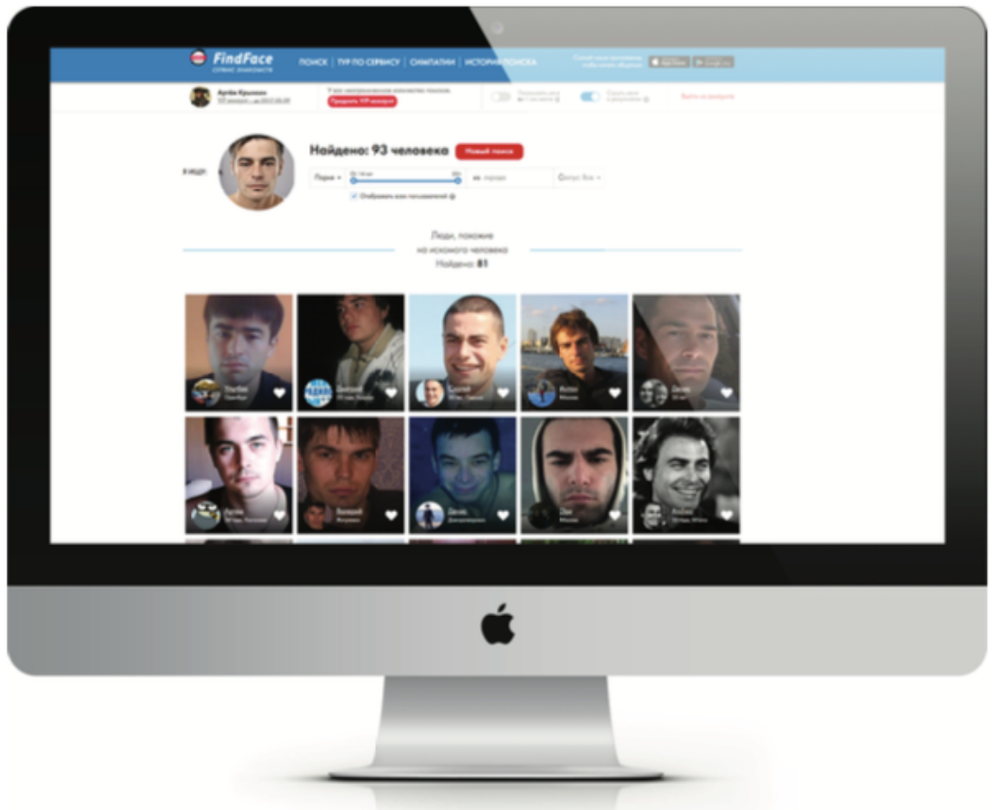
- Comparable to human abilities on small datasets;
- Scales much better and outperforms human in a large scale recognition problem.

N-TECH. LAB

FindFace.ru

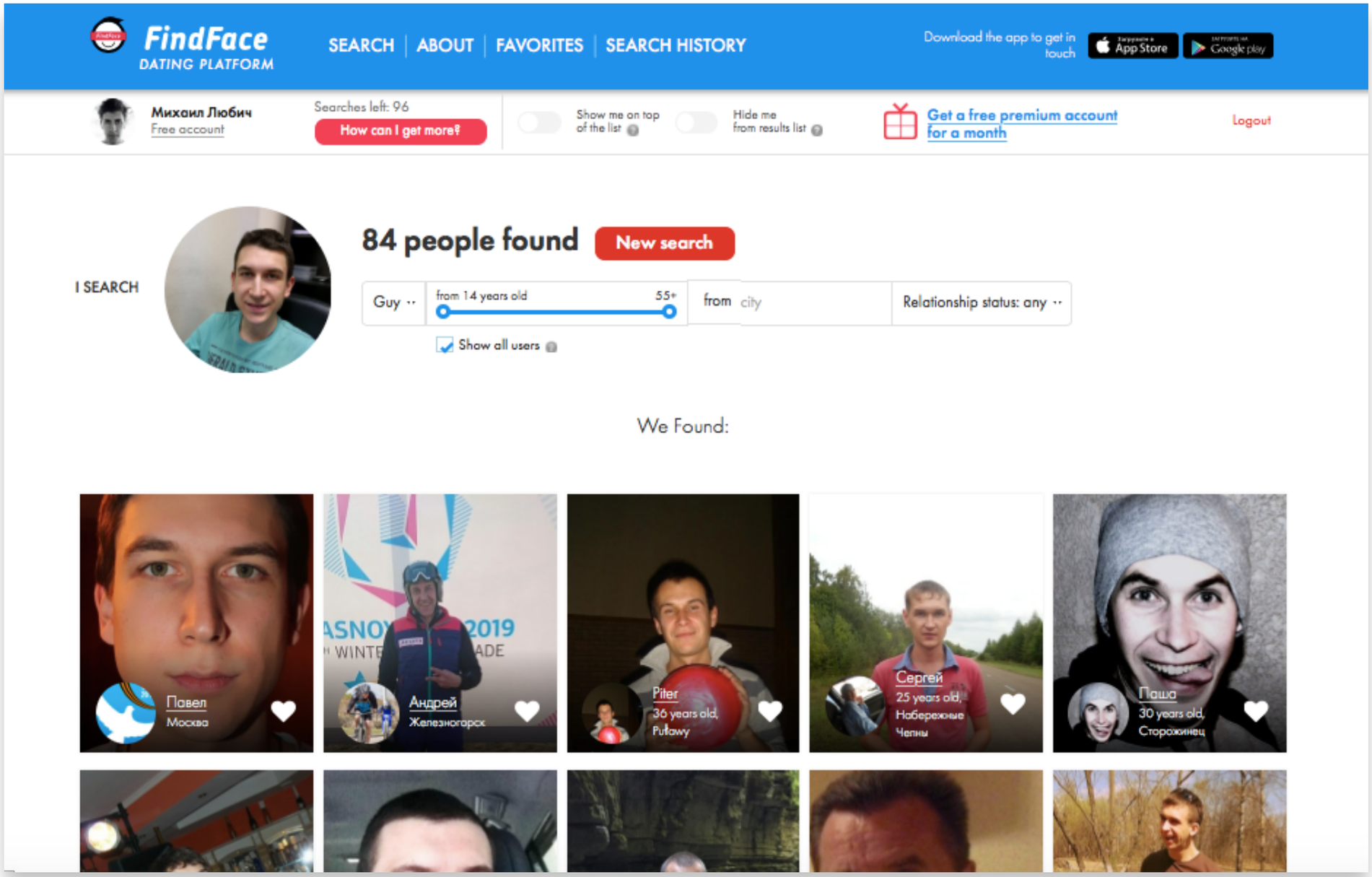
- Search among profile photos in vk.com - largest Russian social network
- 250M photos in index
- 90M people in index
- Search time 0.5 s
- 50 RPS on 5 AWS machines.





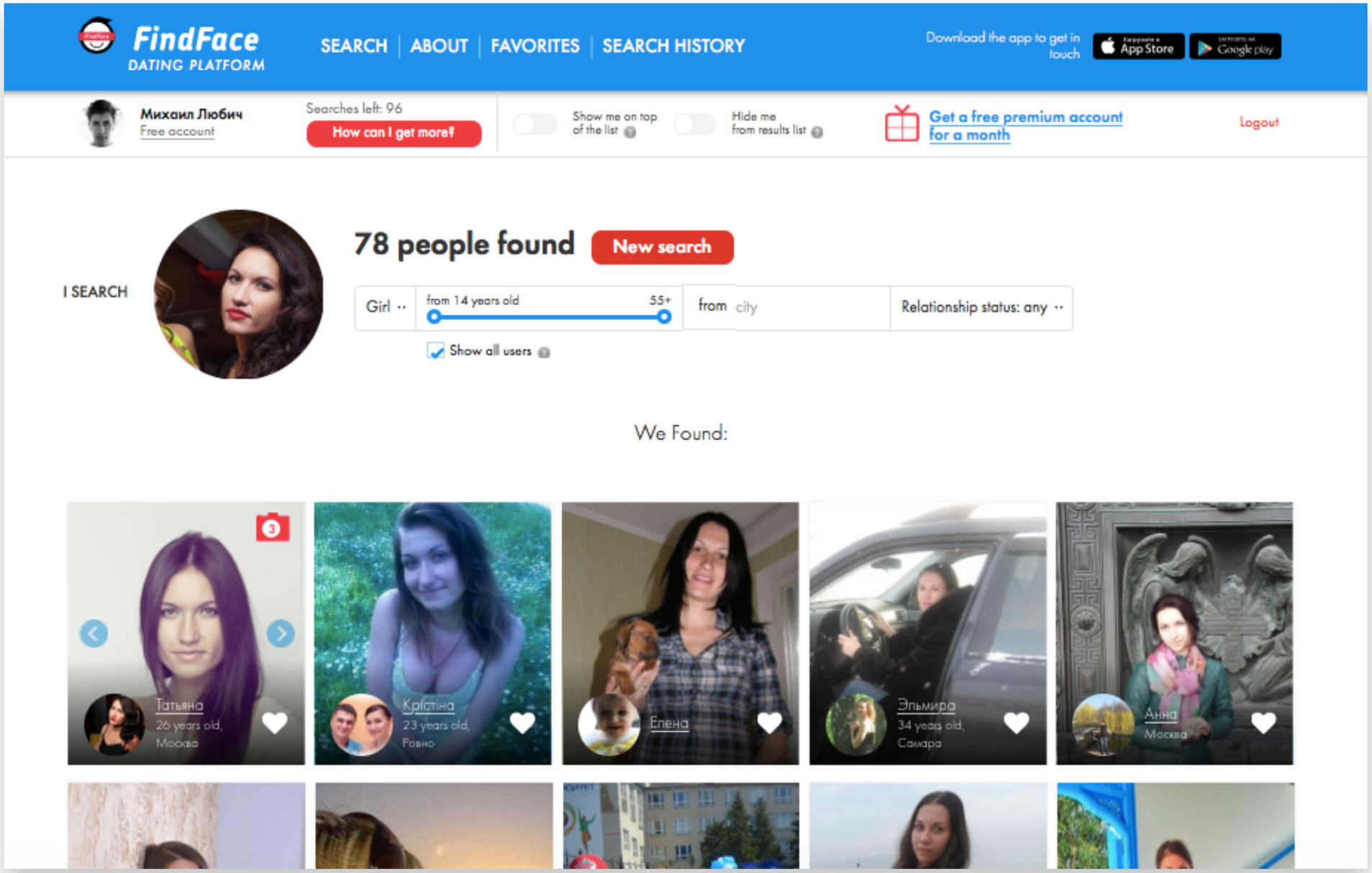
N-TECH.
LAB

FindFace.ru

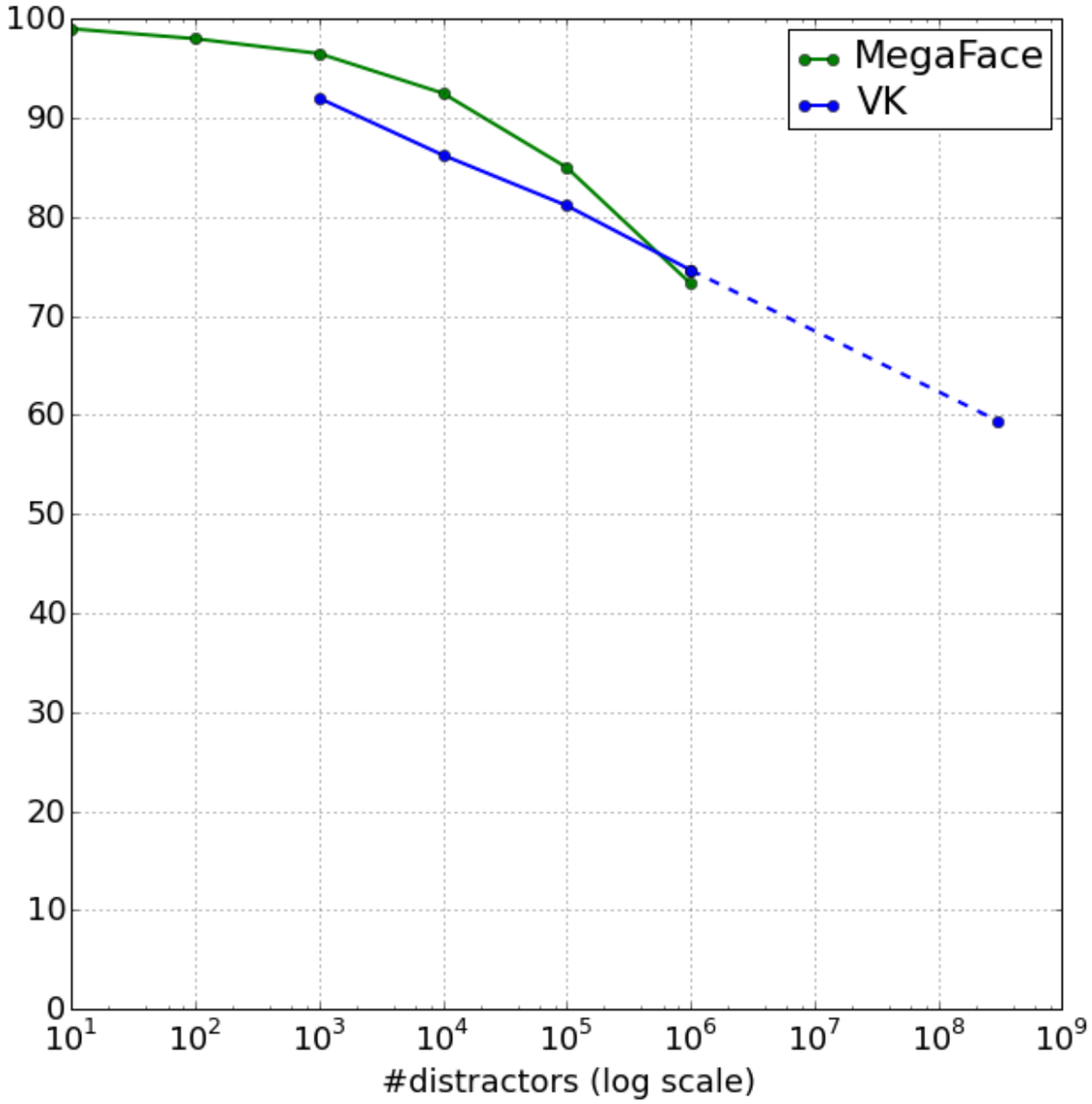


N-TECH.
LAB

FindFace.ru



Accuracy at scale



Rank-1 accuracy

Number of photos	**Accuracy**
1 M	73 %
250 M	60 %

N-TECH. LAB

Timings

	GPU **Nvidia GTX TITAN Black**	**CPU** **Intel Core i7-5930K**
Neural Net Training	514 hours x 3 GPUs	-
Face Detection	-	150 ms
Feature extraction	8.96 ms	143 ms
Search time	-	130 ms

N-TECH.
LAB

FindFace in real-life

- Russian police use findface.ru for searching criminals.
  We've got a lot of emails from them about their experience.
  E.g. police from Udmurt Republic in more than 50% of cold cases found suspect by photo in vk.com.

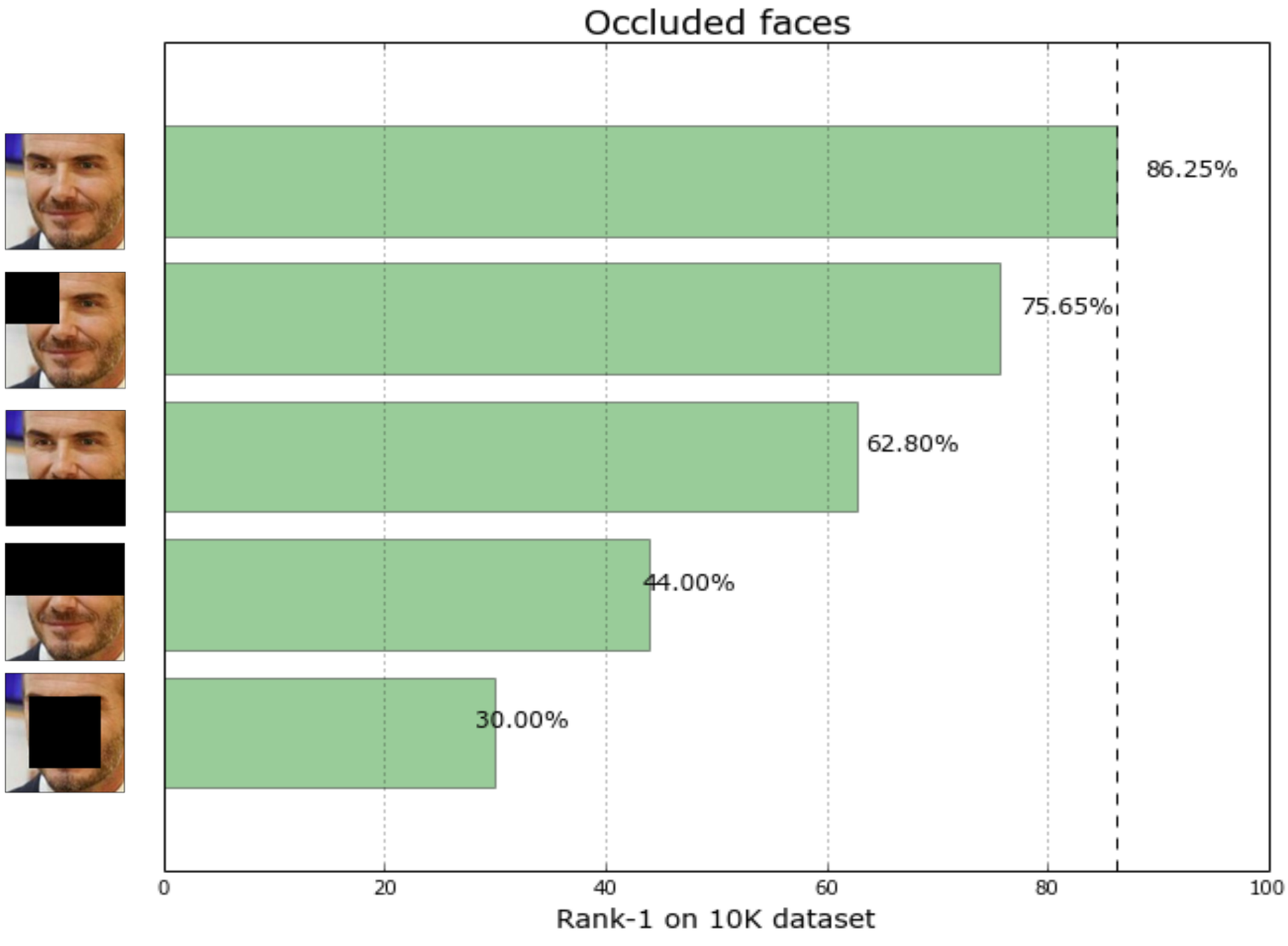
- Findface.ru increases the percentage of solved crimes.

FindFace in real-life

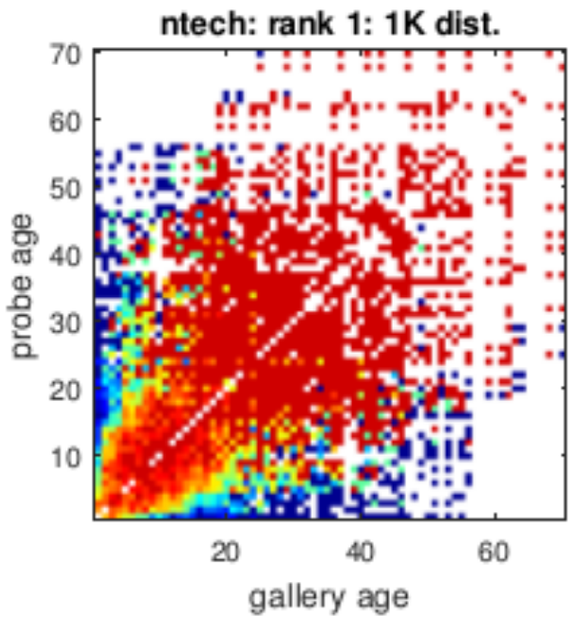
- In St. Petersburg: two teenagers decide to fire a newly built house
- Their faces were filmed on the hidden camera in the elevator
- After the video had appeared in the Internet, people quickly found accounts of these hooligans in the largest Russian social network VK.com with all additional information: where they live, what school they're attending etc.
- All information was transferred to the police.

N-TECH. LAB

Robustness to occlusions



Age and Gender



Gender recognition

- SVM classifier above feature vectors
- 99.5% accuracy

N-TECH. LAB

FINDFACEPRO

FindFace.pro - b2b cloud platform

- Upload your dataset up to 1 billion photos;
- Identification and verification scenarios;
- Gender, Age, Emotion;
- Scale to any number RPS you need;
- Extend to 5B people dataset.

N-TECH.
LAB

Interaction pattern for enterprise

Cameras
Data acquisition

Our software
SDK

Customer's servers

Storing data in customer's Private Cloud.

N-TECH.
LAB

Contacts

Artem Kukharenko

kukharenko@ntechlab.ru

+7 (499) 110-22-54