OFFLINE FACIAL IMAGE ANALYSIS AND RECOGNITION ON MOBILE DEVICES

Andrey V. Savchenko
Dr. of Sci., Prof., Lead Researcher in LATNA,
National Research University HSE, N. Novgorod
Email: avsavchenko@hse.ru
URL: www.hse.ru/en/staff/avsavchenko

October 17-19, 2019

- **Memory!?**

64-256 Gb on Huawei Mate20 Pro, Samsung S10 and Iphone 11

- **Computational complexity!?**

MobileNet inference time:

- 50-60 ms on S9+
- 20 ms on MacBook Pro 2016

- **Power consumption!**

- **Another platform (Android, iOS). Where is Python?**

Outline

1. Face identification task
2. Sequential analysis of high-dimensional features
3. PNN with complex exponential activation functions
4. Maximum likelihood of distances
5. Facial attributes recognition
6. Organizing photo and video albums on mobile device

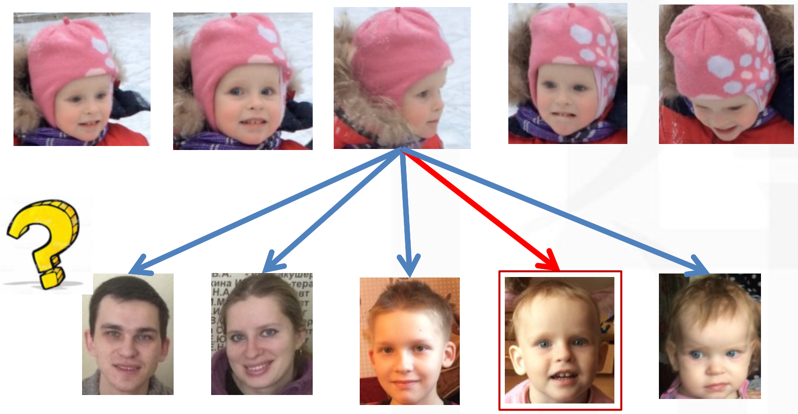
Face identification: main task in this presentation

1. *Still photo recognition*

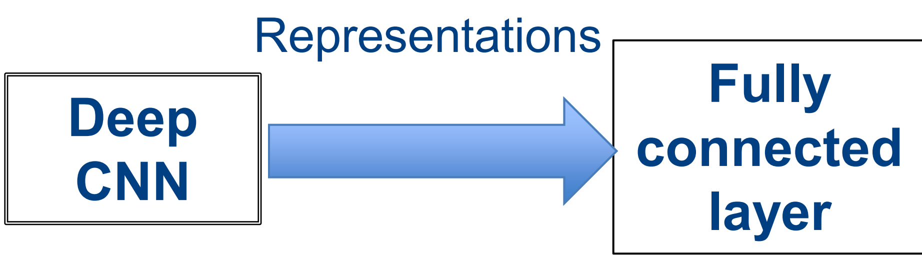
Let the input facial image X be specified. The problem is to assign this image to one of C>>1 subjects (large dataset) specified by R reference facial images $\{X_r\}$.

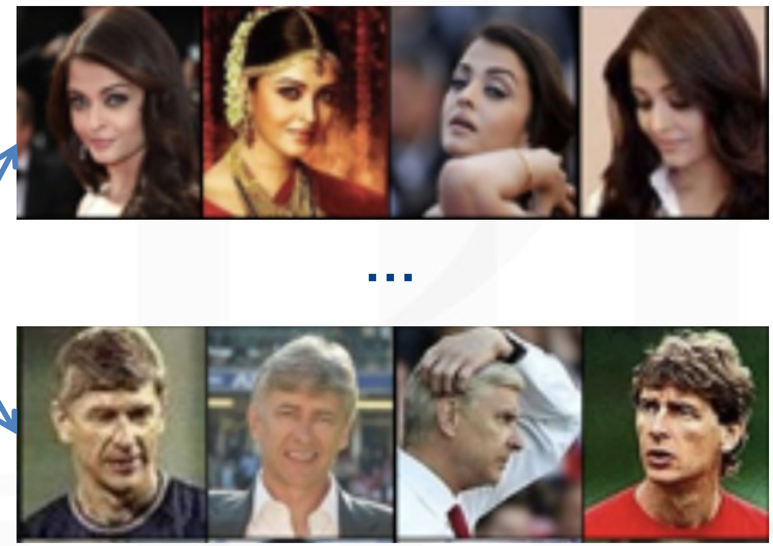
2. *Still-to-video recognition*

The input sequence $\{X(t)\}$ of T>1 observations of one person is assigned to one of C>1 classes specified by R instances $\{X_r\}$

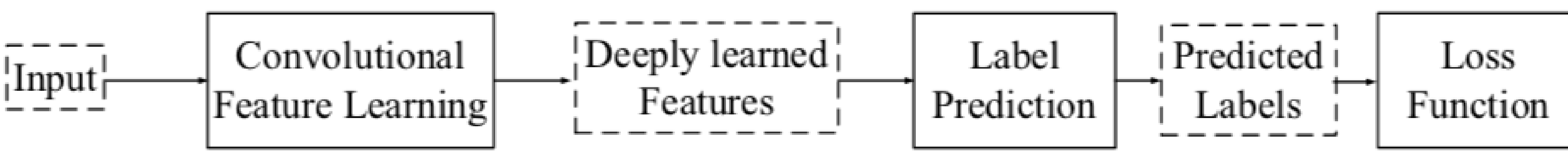


Large training set of celebrities (VGGFace2, CASIA-WebFace, MS-Celeb-1M,…)





…



Each image is described by *D*-dimensional features (*D*=256..4096) by CNN.

[Wen et al, ECCV 2016]

The number of photos per class is small (including one sample per person): $C \approx R$

Nearest neighbor (NN) classifier

Recognition of each frame | **[Learnable] pooling (aggregation)**

$$\min_{r \in \{1,\ldots,R\}} \sum_{t=1}^{T} \rho(\mathbf{x}(t), \mathbf{x}_r)$$

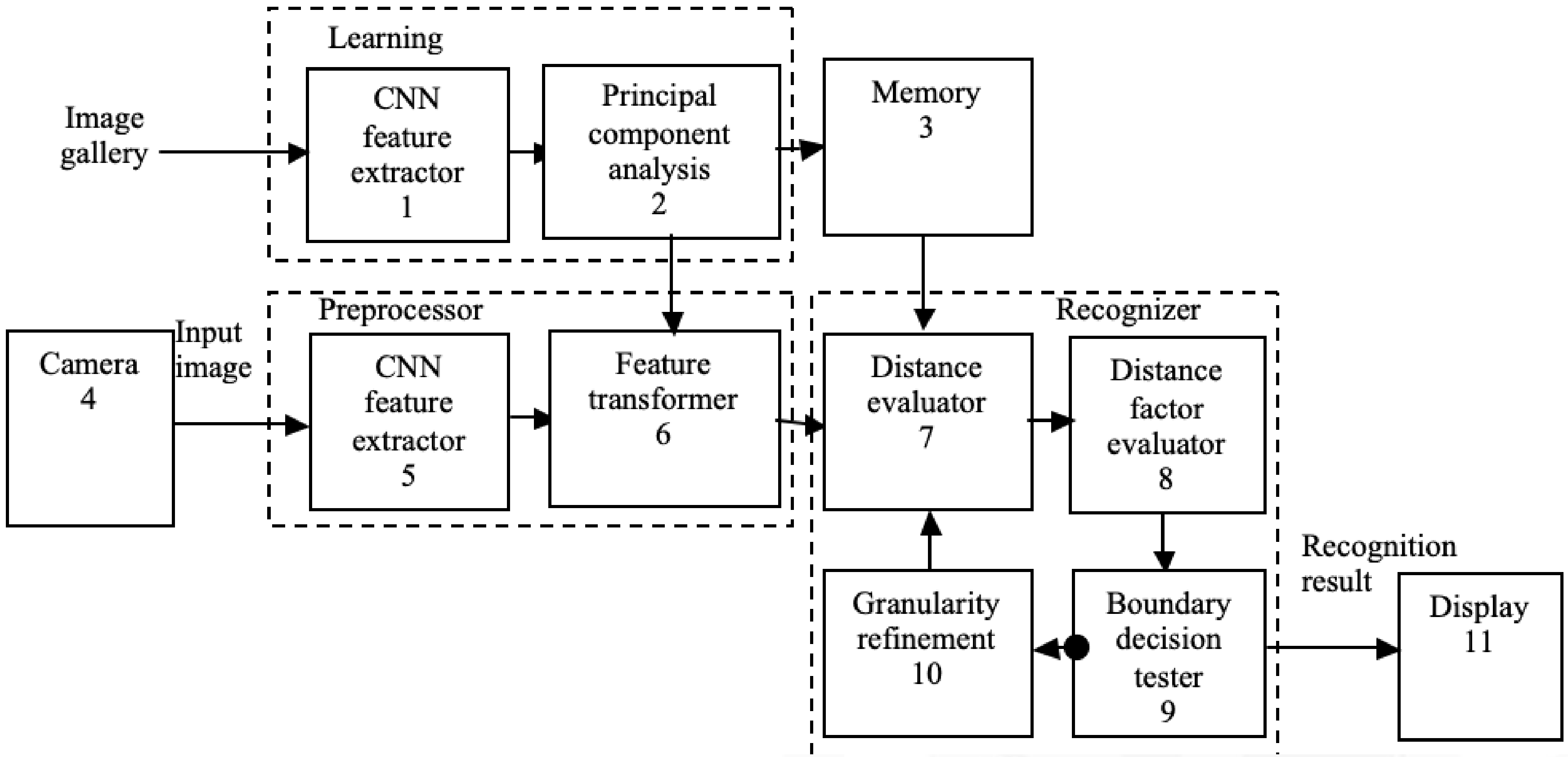
$$\bar{\mathbf{x}} = \sum_{t=1}^{T} w_t \mathbf{x}(t)$$

Brute force of the whole database, run-time complexity: O(*RD*).

- Performance is insufficient, especially for mobile devices
- Multi-objective optimization (minimize error rate with reasonable recognition time): $\overline{\alpha} \to \min \qquad \bar{t} \leq t_0$

Sequential analysis of high-dimensional features

Three-way decisions to choose robust representation of the input image



- Savchenko, A.V. Information Sciences, 2019
- Savchenko, A.V. Knowledge-Based Systems, 2016
- Savchenko A.V., IEEE SACI 2018

[Yao Y., Information Sciences, 2010]: “A **positive** rule makes a decision of **acceptance**, a **negative** rule makes a decision of **rejection**, and a **boundary** rule makes a decision of **abstaining**”

Key question: how to make a decision if the boundary region was chosen? Yao Y. Proc. of RSKT, LNCS, 2013: "Objects with a non-commitment decision may be further investigated by using fine-grained granules"

PCA (principal component analysis), scores are ordered by corresponding eigenvalues

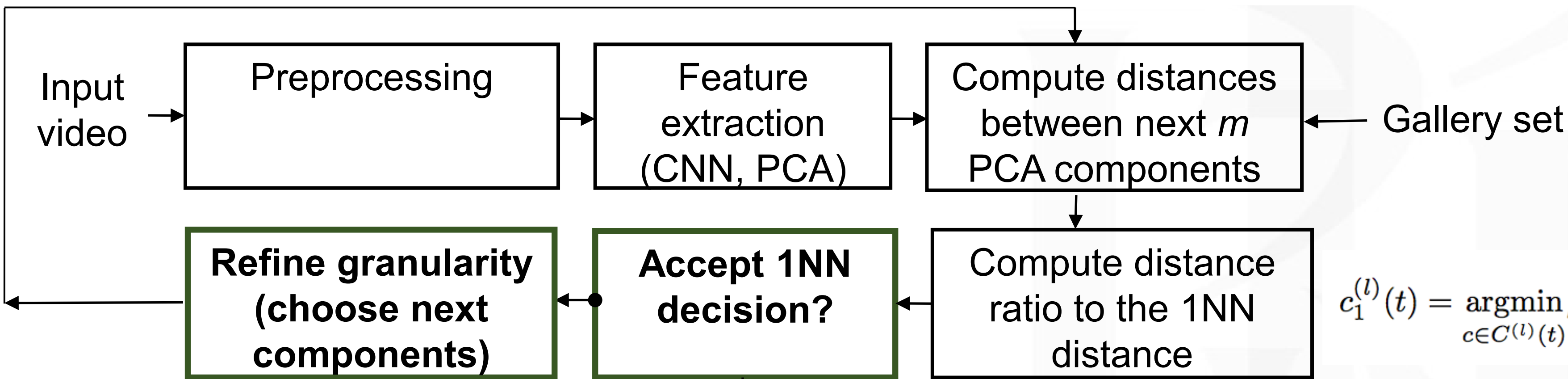
$$\tilde{\mathbf{x}}(t) = [\tilde{x}_1(t), ..., \tilde{x}_D(t)]$$

Proposed: representation of frame at the *l*-th granularity level includes first $d^{(l)}=lm$ principal components. This representation is computationally cheap for additive distances

$$\rho\left(\tilde{\mathbf{x}}^{(l+1)}(t), \tilde{\mathbf{x}}_r^{(l+1)}\right) = \rho\left(\tilde{\mathbf{x}}^{(l)}(t), \tilde{\mathbf{x}}_r^{(l)}\right) + \sum_{d=d^{(l)}+1}^{d^{(l+1)}} \rho(\tilde{x}_d(t), \tilde{x}_{r;d})$$

Distance ratio

$$\rho_c(\tilde{\mathbf{x}}^{(l)}(t)) = \min_{r\in\{1,\dots,R\},c(r)=c} \rho(\tilde{\mathbf{x}}^{(l)}(t), \tilde{\mathbf{x}}_r^{(l)})$$



$$c_1^{(l)}(t) = \operatorname*{argmin}_{c\in C^{(l)}(t)} \rho_c(\tilde{\mathbf{x}}^{(l)}(t))$$

$C^{(1)}(t)=\{1,\dots,C\}$

$$C^{(l+1)}(t) = \left\{ c \in C^{(l)}(t) \,\middle|\, \frac{\rho_c(\tilde{\mathbf{x}}^{(l)}(t))}{\rho_{c_1^{(l)}(t)}(\tilde{\mathbf{x}}^{(l)}(t))} \le \delta \right\}$$

Final Maximum a-posterior (MAP) decision

$$\max_{c\in C^{(L)}} \sum_{t=1}^{T} \frac{\exp(-n\rho_c(\tilde{\mathbf{x}}^{(l)}(t)))}{\sum_{i\in C^{(L)}} \exp(-n\rho_c(\tilde{\mathbf{x}}^{(l)}(t))}$$

Strong theoretical foundations for the Jensen-Snannon and Kullback-Leibler divergences

Here is exactly how our method works in practice

Probe photo Closest gallery photos

(a) (b) (c) (d)

Recognition results, distance factor threshold 0.7

Subject	$l=1$		$l=2$		$l=3$	
	$\rho[r]$	$\rho_{\min}/\rho[r]$	$\rho[r]$	$\rho_{\min}/\rho[r]$	$\rho[r]$	$\rho_{\min}/\rho[r]$
Armstrong (d)	0.0086	0.87	0.0122	1	0.0129	1
Auriemma (b)	0.0074	1.00	0.0170	0.71	0.0195	0.66
McEwen (c)	0.0104	0.72	0.0188	0.65	-	-
Williams	0.0100	0.75	0.0300	0.41	-	-
Wirayuda	0.0103	0.73	0.0217	0.56	-	-
LeBron	0.0105	0.71	0.0200	0.61	-	-

Better recognition performance (no need to process all features)

Higher recognition accuracy

Experimental results: YTF/LFW

	ResFace		VGGFace		VGGFace2_ft	
Classifier	**Accuracy (%)**	**Running time (ms)**	**Accuracy (%)**	**Running time (ms)**	**Accuracy (%)**	**Running time (ms)**
SVM	18.53±0.5	10213±33	22.49±0.6	19840±40	52.85±1.2	10137±35
SVM, AvgPooling	22.15±1.2	10269±39	26.90±0.7	20070±37	55.91±1.3	10252±42
k-NN	30.92±0.1	12.3±0.0	43.44±0.1	23.3±0.1	74.48±0.1	12.7±0.0
MAP	31.49±0.0	12.3±0.0	43.50±0.1	23.7±0.1	74.41±0.1	12.6±0.1
k-NN/32 PCA	26.28±0.3	1.7±0.0	37.69±0.1	2.2±0.1	58.81±0.0	19±0.1
k-NN/256 PCA	31.21±0.1	2.3±0.0	44.64±0.2	2.9±0.1	75.02±0.0	2.5±0.1
Proposed, k-NN	**31.21±0.1**	**0.8±0.0**	**44.64±0.1**	**1.1±0.1**	**74.77±0.1**	**0.7±0.0**
Proposed, MAP	**31.52±0.2**	**0.9±0.1**	**45.46±0.1**	**1.2±0.1**	**74.87±0.2**	**0.7±0.0**

Experimental results: still-to-video recognition

Classifier	ResFace Accuracy (%)	ResFace Running time (ms)	VGGFace Accuracy (%)	VGGFace Running time (ms)	VGGFace2_ft Accuracy (%)	VGGFace2_ft Running time (ms)
SVM	55.98±0.3	117.3±0.9	70.88±0.3	230.3±1.6	84.12±0.4	117.8±1.1
SVM, AvgPooling	57.17±0.5	114.7±0.7	73.62±0.0	233.0±4.7	87.06±0.1	124.4±1.4
k-NN	54.41±0.0	12.9±0.0	68.88±0.3	24.9±0.2	85.67±0.2	12.8±0.1
MAP	55.16±0.2	13.3±0.3	70.50±0.8	25.0±0.0	87.09±0.1	12.8±0.1
k-NN/32 PCA	48.49±0.4	1.2±0.1	66.26±0.3	1.9±0.2	77.91±0.2	1.3±0.1
k-NN/256 PCA	53.91±0.3	1.9±0.1	70.94±0.1	2.5±0.1	85.48±0.1	2.1±0.0
Proposed, k-NN	**54.12±0.2**	**0.8±0.1**	**70.94±0.1**	**1.2±0.0**	**85.65±0.2**	**0.8±0.0**
Proposed, MAP	**56.01±0.1**	**0.8±0.1**	**73.06±0.5**	**1.2±0.1**	**86.89±0.1**	**0.8±0.1**

Probabilistic Neural Network (PNN) with complex exponential activation functions

Statistical approach: empirical Bayesian classifier with naïve assumption about independent features

$$c^* = \arg\max_{c \in \{1,\dots,C\}} \frac{R(c)}{R} f(\mathbf{x}|W_c) \qquad \hat{f}(\mathbf{x}|W_c) = \prod_{d=1}^{D} \hat{f}_d(x_d|W_c)$$

1. We propose to estimate the individual likelihood as the average of the first *J* partial sums. Here the right-hand side is the non-negative Fejér kernel

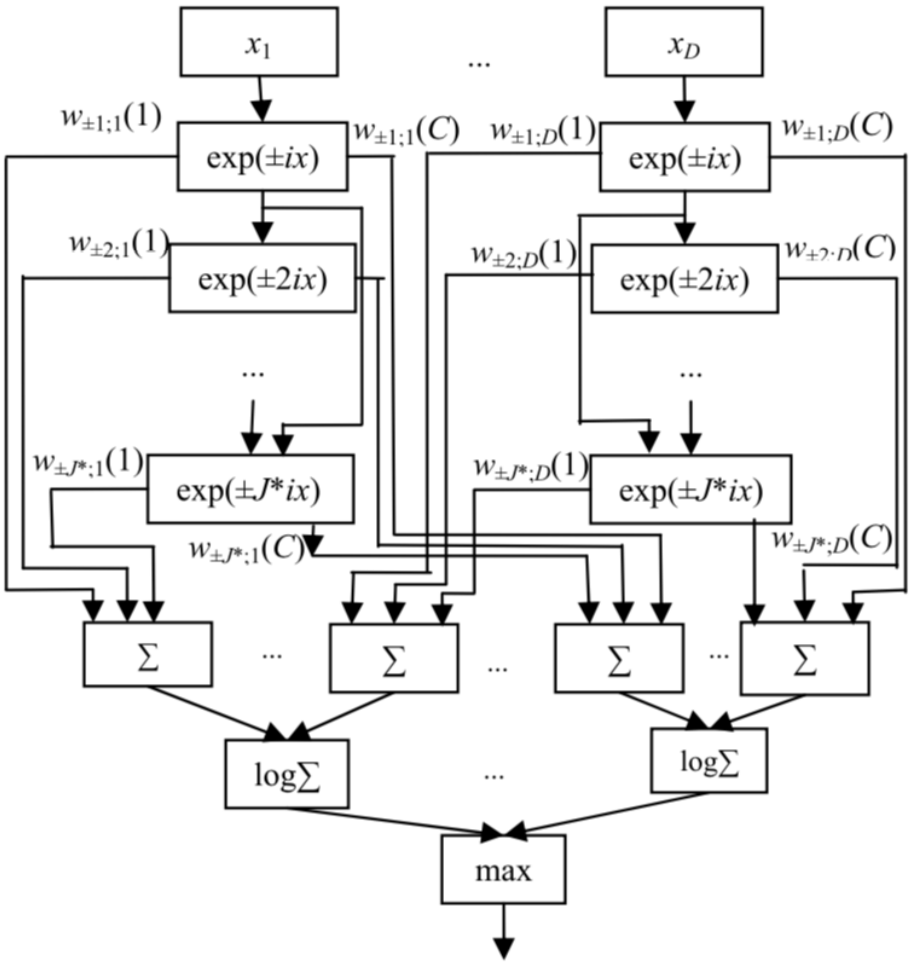
$$\hat{f}_d(x_d|W_c) = \frac{1}{J+1}\sum_{j=0}^{J} \hat{f}_{j;d}(x_d|W_c) \qquad F_{J+1} = \frac{1}{J+1}\left(\frac{\sin((J+1)\pi(x_d - x_{r;d}(c))/2)}{\sin(\pi(x_d - x_{r;d}(c))/2)}\right)^2$$

2. We replace canonical form of density estimate to the equivalent form, **which does not implement the brute force**

$$\hat{f}_{J;d}(x_d|W_c) = \sum_{j=-J}^{J} a_{j;d}(c)\exp(ij\pi x_d) \qquad \psi_j(x) = \exp(ij\pi x).$$

- Savchenko, A.V. IEEE Transactions on Neural Networks and Learning Systems, 2019
- Savchenko A.V., IEEE ICPR 2018

PNN based on trigonometric series



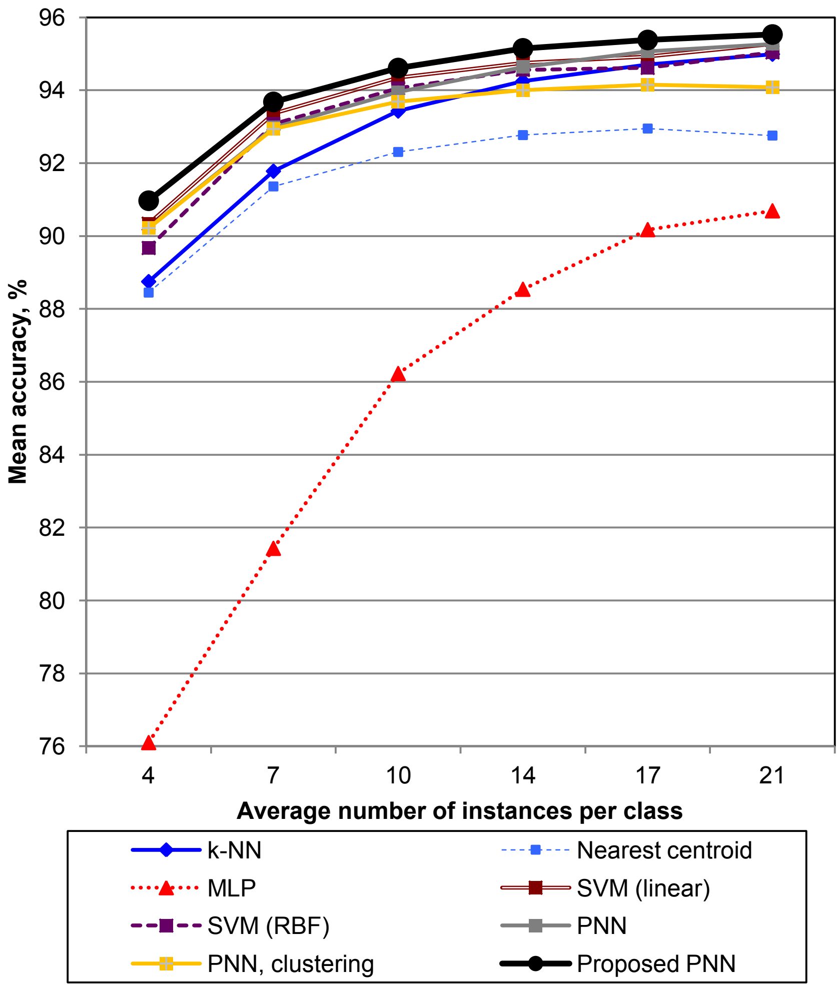
$$c^* = \operatorname*{argmax}_{c\in\{1,\ldots,C\}} \sum_{d=1}^{D} \log \sum_{j=-J}^{J} w_{j;d}(c)\cdot\psi_j(x_d)$$

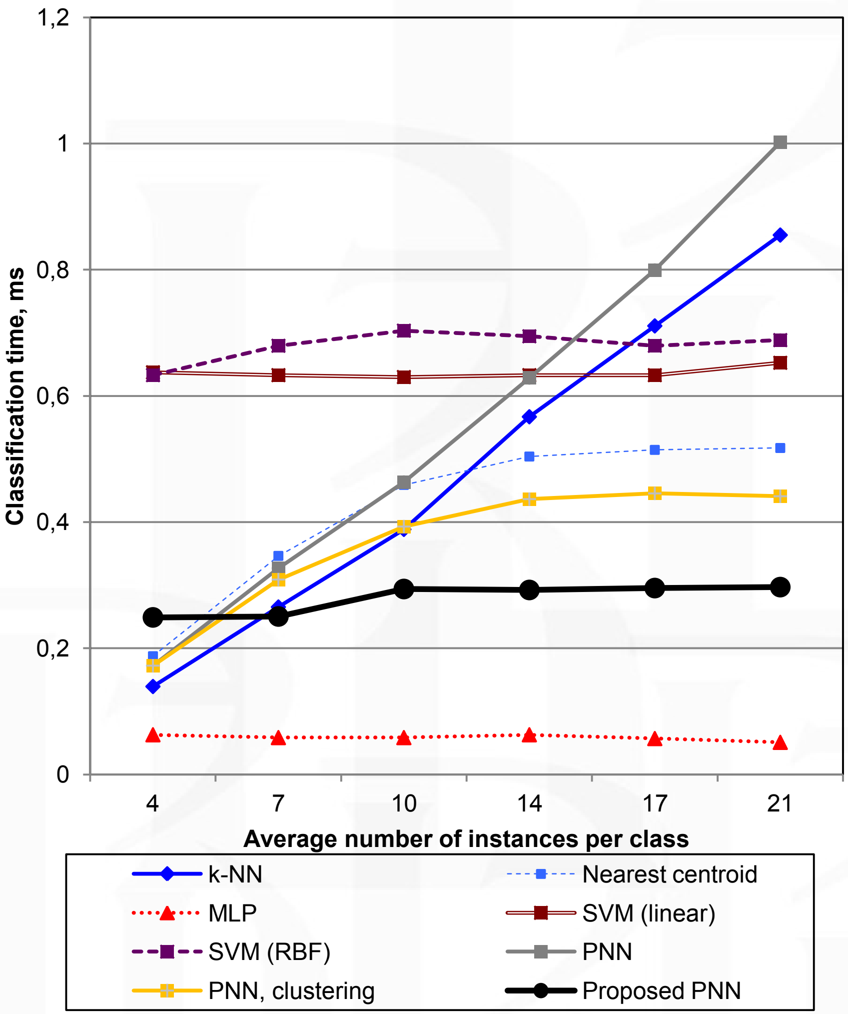
$$\psi_{j+1}(x_d) = \psi_j(x_d)\cdot\psi_1(x_d),$$
$$\psi_{-j}(x_d) = \overline{\psi_j(x_d)}.$$

- **Converges to Bayesian solution**
- **Very high training speed**
- **Runtime complexity and memory space complexity:** $O(DR^{1/3}C^{1/3})$.

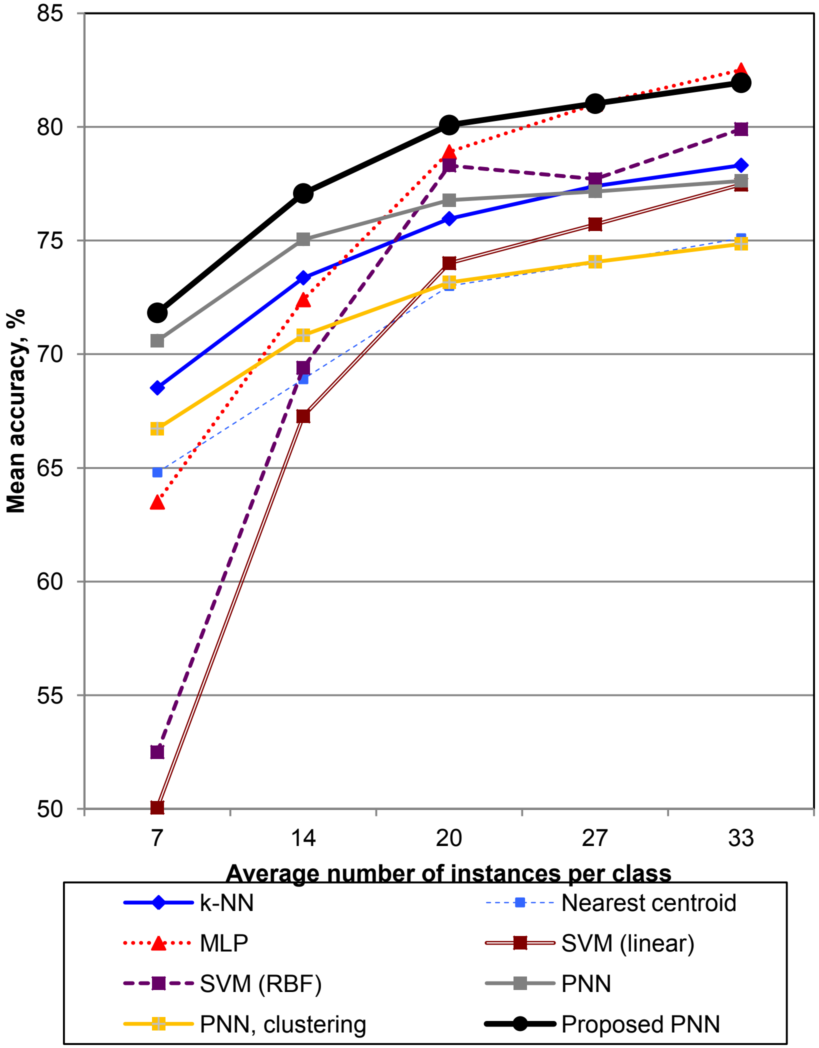
Online classification is approximately $R^{2/3}$–times faster than instance-based learning (PNN, k-NN) if at least 5 photos per subject are available

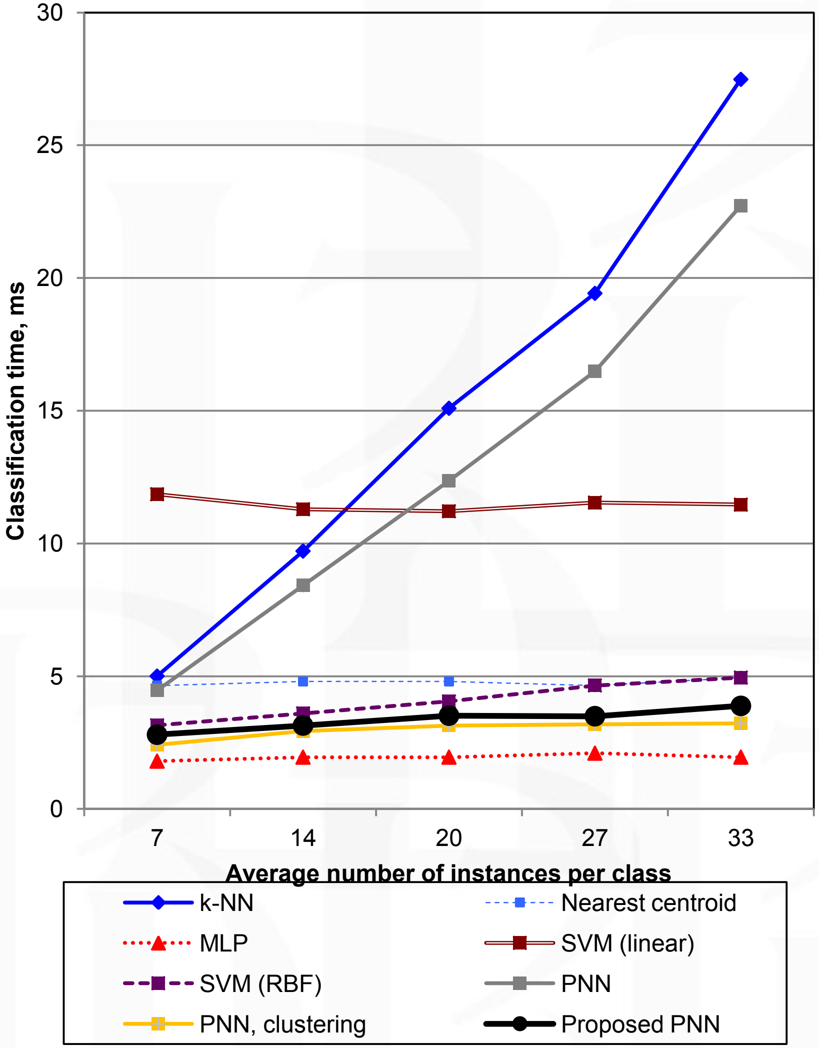
Experimental results (PubFig83 dataset)





Experimental results (subset of Casia WebFaces)





Maximum likelihood of distances

The next reference example (instance) to check is chosen based on the maximal likelihood of *distances*

$$r_{k+1} = \underset{\nu \in \{1,\dots,R\} - \{r_1,\dots,r_k\}}{argmax} \left(p_\nu \cdot \prod_{i=1}^{k} f\left(\rho(X, X_{r_i}) \middle| W_\nu \right) \right),$$

KL divergence is asymptotically distributed as non-central chi-squared, normal approximation

$$\mathcal{N}\left(\rho_{KL}(\mathbf{x}_r, \mathbf{x}_i) + \frac{D-1}{WH}, \frac{4WH\rho_{KL}(\mathbf{x}_r, \mathbf{x}_i) + (D-1)}{2(WH)^2} \right)$$

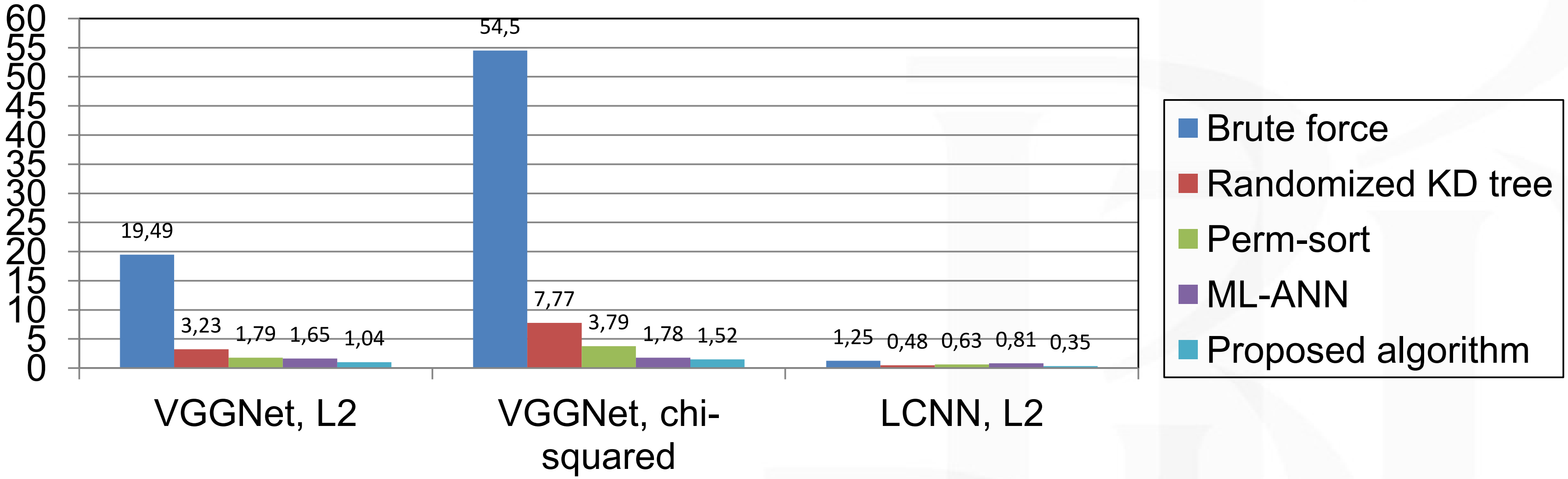
$$r_{k+1} = \underset{\mu \in \{1,\dots,R\} - \{r_1,\dots,r_k\}}{argmin} \left(\sum_{i=1}^{k} \varphi_\mu(r_i) - lnp_\mu \right)$$

$$\varphi_\mu(r_i) \approx \frac{\left(\rho(X, X_{r_i}) - \rho_{\mu,r_i} \right)^2}{\rho_{\mu,r_i}}$$

- Savchenko A.V., Pattern Recognition, 2017,2012
- Savchenko A.V., Optimization Letters, 2017

Experimental Results for face identification in LFW

Recognition time, ms



Error rate, %

	VGGFace, L2	VGGFace, chi-squared	Light CNN, L2
SVM	49.6	-	28.9
1-NN (Brute force)	10.7	10.2	9.6
Randomized KD tree	11.0	10.6	9.9
Perm-sort	10.9	10.6	9.9
ML-ANN	10.9	10.6	10.5
Proposed algorithm	10.9	10.5	9.9

Improve the accuracy by penalizing the classes if their features do not behave like the features of input image in the space of dissimilarities

$$f(\rho_1(\mathbf{x}),\ldots,\rho_C(\mathbf{x})|W_c) = f(\rho_c(\mathbf{x})|W_c)\cdot \prod_{i=1,i\neq c}^{C} f(\rho_i(\mathbf{x})|W_c)$$

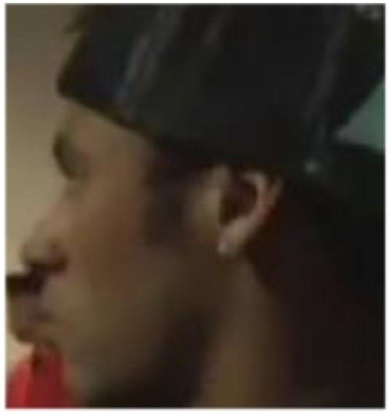
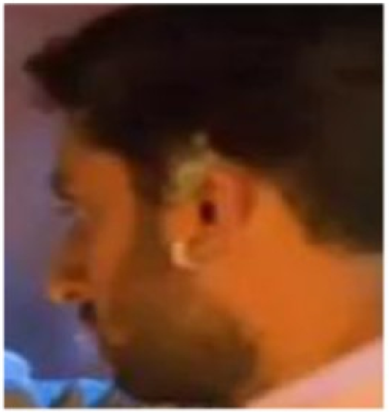
Final criterion

$$\min_{c\in\{c_1,\ldots,c_M\}}\left(\rho_c(\mathbf{x}) + \frac{\lambda}{C}\sum_{i=1,i\neq c}^{C}\frac{(\rho_i(\mathbf{x}) - \rho_{c;i} - \frac{D-1}{WH})^2}{\rho_{c;i} + \frac{D-1}{4WH}}\right)$$

It can be considered as adding regularization term to the nearest neighbor (NN) rule

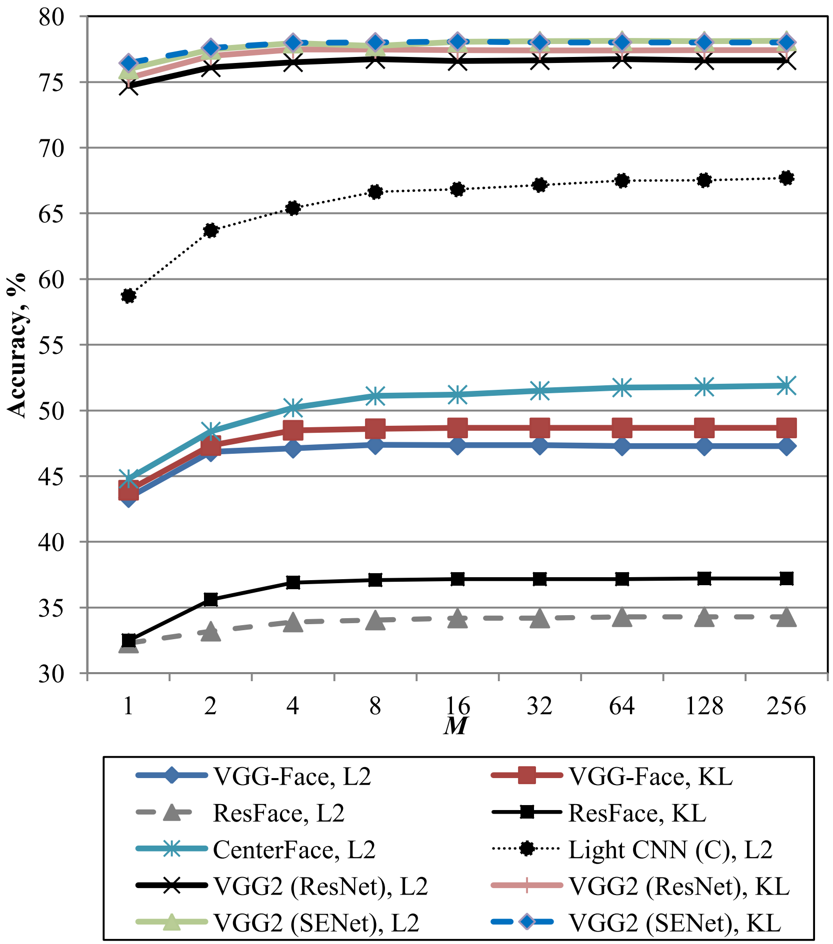
- Savchenko A.V. et al. Expert Systems With Applications, 2018
- Savchenko A.V. et al Optical Memory and Neural Networks, 2018

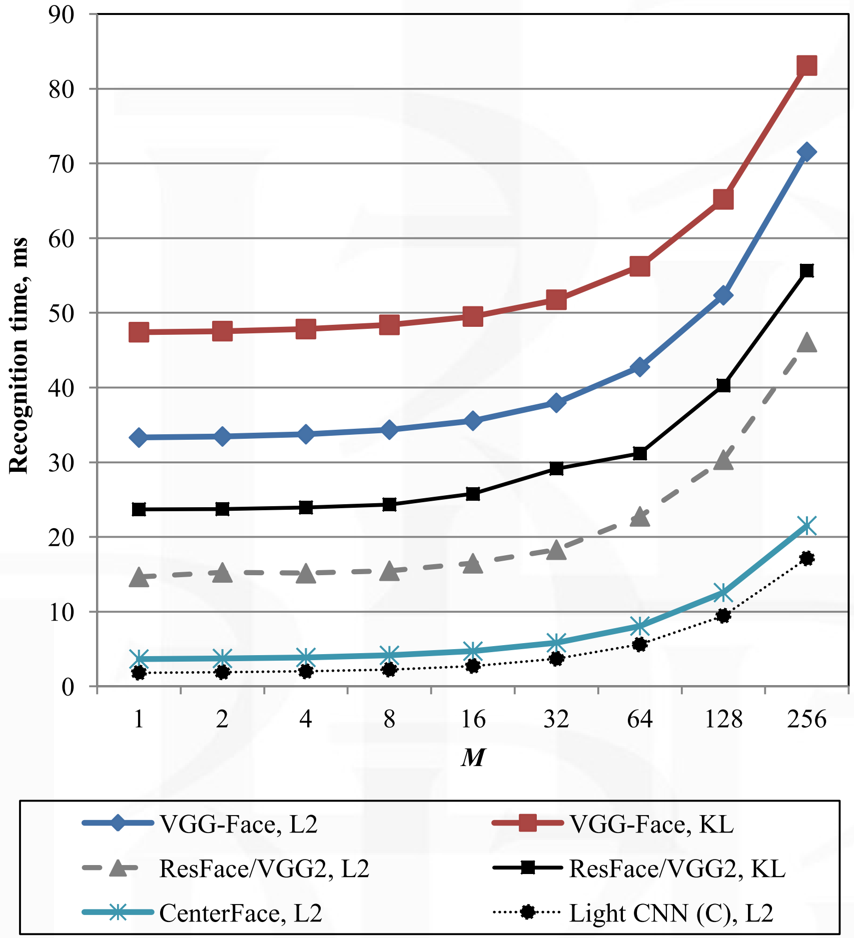
Example

Probe (SubjectId 942)	first NN (SubjectId 1001)	second NN (SubjectId 1303)	Our result (seventh NN, SubjectId 942)
			

SubjectId	L_2 distance	Regularization term	Total objective
1001	0.0236	0.0152	0.0247
1303	0.0238	0.0152	0.0249
314	0.0239	0.0113	0.0247
1980	0.0240	0.0166	0.0252
387	0.0241	0.0300	0.0262
1289	0.0241	0.0140	0.0251
942	**0.0242**	**0.0059**	**0.0246**
1287	0.0242	0.0174	0.0258

Experimental results: YTF/LFW



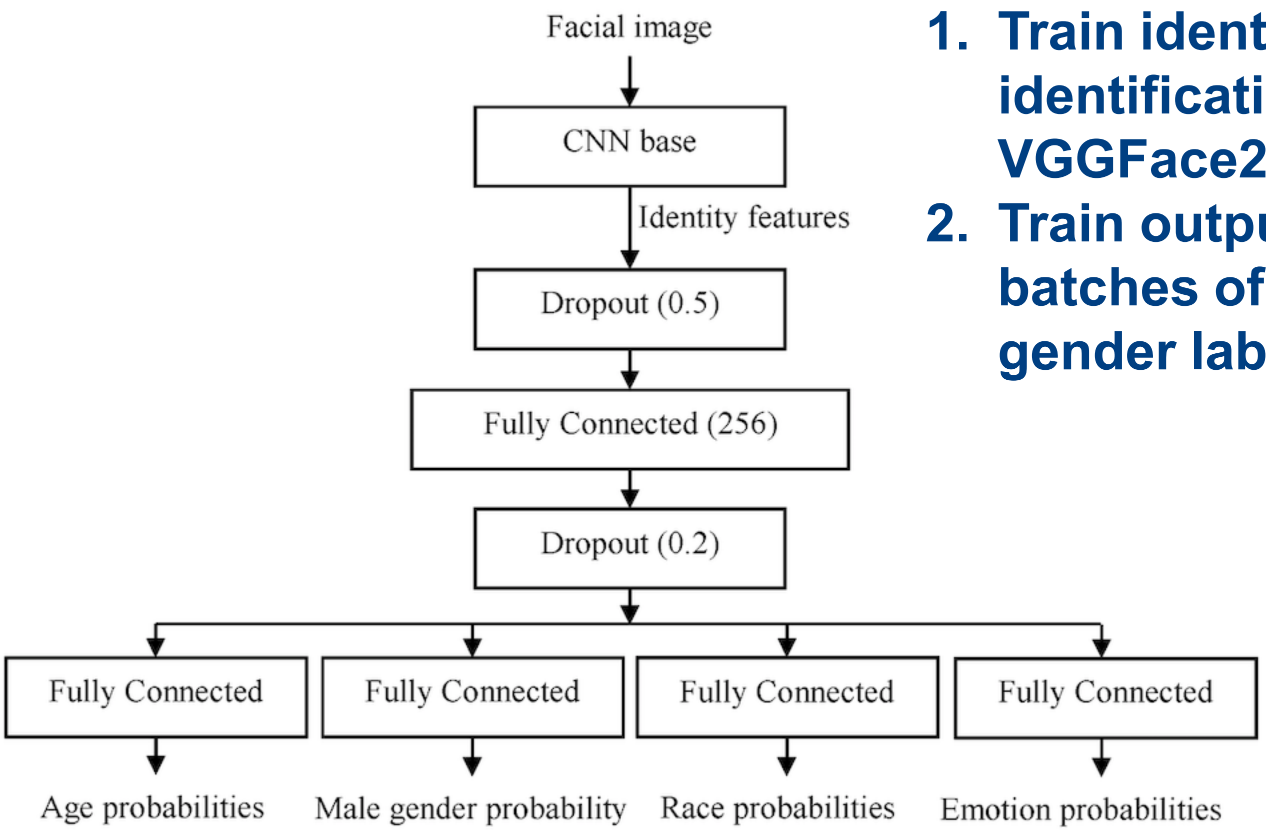


Rank-1 accuracy (%) in 1:N identification

Features	Distance	IJB-A bounding boxes			MTCNN face detection		
		SVM	NN	Proposed	SVM	NN	Proposed
CenterFace	L_2	52.6 ± 1.8	49.5 ± 1.9	54.6 ± 1.7	64.6 ± 1.0	64.9 ± 1.1	66.8 ± 1.1
Light CNN	L_2	57.2 ± 0.5	58.3 ± 0.7	61.8 ± 0.6	79.9 ± 0.4	81.3 ± 0.4	83.3 ± 0.3
ResFace	L_2	79.8 ± 0.7	77.9 ± 0.5	80.7 ± 0.7	86.4 ± 0.5	85.6 ± 0.6	87.6 ± 0.6
ResFace	KL	–	79.2 ± 0.6	81.3 ± 0.7	–	86.0 ± 0.5	88.3 ± 0.6
VGG-Face	L_2	87.7 ± 0.6	87.0 ± 0.5	88.5 ± 0.6	89.2 ± 0.4	89.0 ± 0.4	90.3 ± 0.4
VGG-Face	KL	–	87.5 ± 0.6	88.9 ± 0.6	–	90.0 ± 0.3	91.5 ± 0.4
VGG2 (ResNet)	L_2	95.7 ± 0.3	95.7 ± 0.5	96.1 ± 0.4	96.8 ± 0.4	97.7 ± 0.4	98.0 ± 0.3
VGG2 (ResNet)	KL	–	96.4 ± 0.4	96.7 ± 0.3	–	97.5 ± 0.4	98.0 ± 0.5
VGG2 (SENet)	L_2	95.3 ± 0.3	95.9 ± 0.5	96.8 ± 0.5	97.3 ± 0.5	97.8 ± 0.5	98.1 ± 0.4
VGG2 (SENet)	KL	–	96.0 ± 0.5	96.7 ± 0.3	–	98.0 ± 0.4	98.2 ± 0.6
LBP histograms	L_2	25.8 ± 1.2	24.2 ± 1.1	26.2 ± 1.3	29.0 ± 1.0	28.1 ± 1.2	29.2 ± 1.2
LBP histograms (synthetic images)	L_2	28.2 ± 1.5	27.3 ± 1.2	28.9 ± 1.4	32.3 ± 1.6	31.0 ± 1.3	32.8 ± 1.4
LBP histograms	KL	–	31.0 ± 1.3	33.7 ± 1.5	–	33.6 ± 1.1	34.7 ± 1.4
LBP histograms (synthetic images)	KL	–	33.7 ± 1.4	35.2 ± 1.6	–	35.4 ± 1.2	36.3 ± 1.4
HOG	L_2	11.7 ± 3.4	19.7 ± 2.0	27.9 ± 2.1	12.1 ± 3.3	23.0 ± 1.8	29.1 ± 2.0
HOG (synthetic images)	L_2	12.3 ± 3.5	23.9 ± 3.1	31.9 ± 3.2	12.9 ± 3.1	26.6 ± 2.6	32.3 ± 2.8
HOG	KL	–	29.4 ± 1.9	32.3 ± 2.0	–	31.5 ± 1.6	33.3 ± 1.8
HOG (synthetic images)	KL	–	32.9 ± 1.4	34.4 ± 1.4	–	35.1 ± 1.6	37.2 ± 1.2

Facial attributes recognition

Multi-output model for age, gender and identity recognition



1. **Train identity features for face identification (softmax loss, VGGFace2 dataset)**
2. **Train outputs by alternating batches of data with age and gender labels**

Age prediction

$$l^* = \frac{\sum_{k=1}^{K} l_{(k)} \cdot P\left(l_{(k)} \middle| X(t)\right)}{\sum_{k=1}^{K} P\left(l_{(k)} \middle| X(t)\right)}.$$

$$P\left(l_{(1)} \middle| X(t)\right) \geq P\left(l_{(2)} \middle| X(t)\right) \geq \ldots \geq P\left(l_{(L)} \middle| X(t)\right)$$

[Savchenko, PeerJ Computer Science, 2019]
https://github.com/HSE-asavchenko/HSE_FaceRec_tf/tree/master/age_gender_identity

Experimental results: Age/gender recognition for UTKFace (In-the-wild faces) dataset

Models	Gender accuracy, %	Age MAE	Age accuracy, %	Model size, Mb	Inference time, ms
DEX	91.05	6.48	51.77	1050.5	47.1
Wide ResNet (weights.28-3.73)	88.12	9.07	46.27	191.2	10.5
Wide ResNet (weights.18-4.06)	85.35	10.05	43.23	191.2	10.5
FaceNet	89.54	8.58	49.02	89.1	20.3
BKNetStyle2	57.76	15.94	23.49	29.1	12.5
SSRNet	85.69	11.90	34.86	0.6	6.6
MobileNet v2 (Agegendernet)	91.47	7.29	53.30	28.4	11.4
ResNet-50 from InsightFace	87.52	8.57	48.92	240.7	25.3
"New" model from InsightFace	84.69	8.44	48.41	1.1	5.1
Inception trained on Adience	71.77	-	32.09	85.4	37.7
age_net/gender_net	87.32	-	45.07	87.5	8.6
MobileNets with single head	93.59	5.94	60.29	25.7	7.2
Proposed MobileNet, fine-tuned from ImageNet	91.81	5.88	58.47	13.8	4.7
Proposed MobileNet, pre-trained on VGGFace2	93.79	5.74	62.67	13.8	4.7
Proposed MobileNet, fine-tuned	94.10	5.44	63.97	13.8	4.7

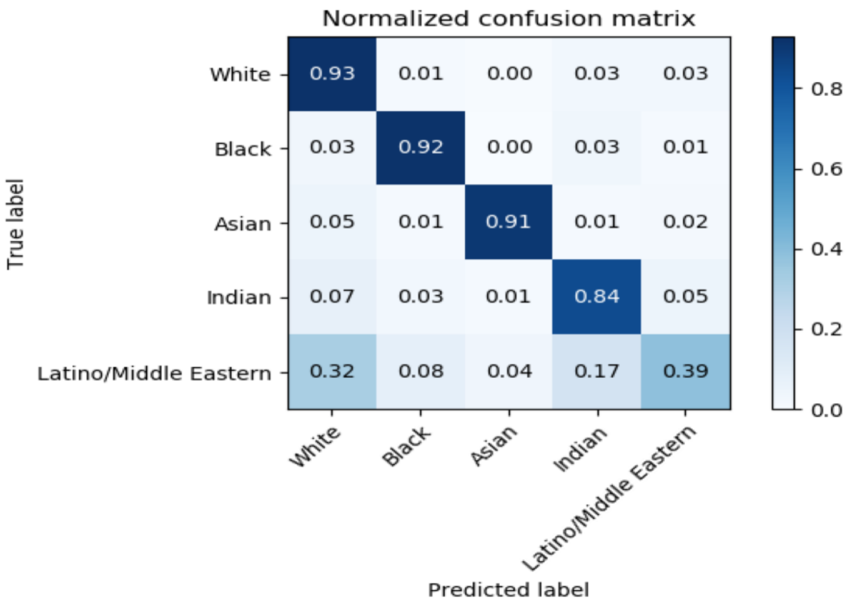
Face extraction and alignment: https://github.com/dandynaufaldi/Agendernet

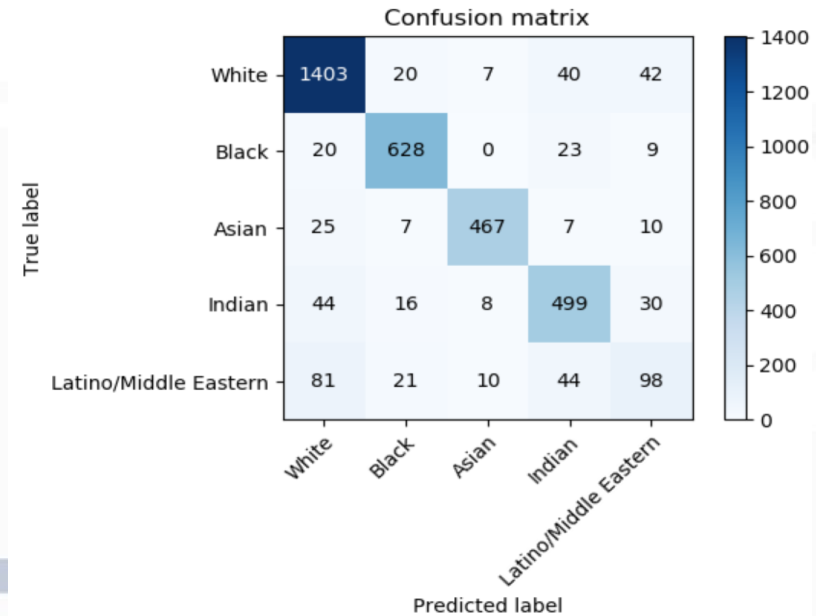
Experimental results: Age/gender recognition for UTKFace (aligned & cropped faces)

Models	Gender accuracy, %	Age MAE	Age accuracy, %
DEX	83.16	9.84	41.22
Wide ResNet (weights.28-3.73)	73.01	14.07	29.32
Wide ResNet (weights.18-4.06)	69.34	13.57	37.23
FaceNet	86.14	9.60	44.70
BKNetStyle2	60.93	15.36	21.63
SSRNet	72.29	14.18	30.56
MobileNet v2 (Agegendernet)	86.12	11.21	42.02
ResNet-50 from InsightFace	81.15	9.53	45.30
“New” model from InsightFace	80.55	8.51	48.53
Inception trained on Adience	65.89	-	27.01
age_net/gender_net	82.36	-	34.18
MobileNets with single head	91.89	6.73	57.21
Proposed MobileNet, fine-tuned from ImageNet	84.30	7.24	58.05
Proposed MobileNet, pre-trained on VGGFace2	91.95	6.00	61.70
Proposed MobileNet, fine-tuned	91.95	5.96	62.74

Ethnicity recognition results, UTKFace

	VGGFace	VGGFace-2	FaceNet	**Our MobileNet**	
				Age/gender features	**Identity features**
Random Forest	83.5	87.8	84.3	80.1	83.8
k-NN	76.2	84.5	84.4	72.2	82.2
SVM (RBF)	78.8	82.4	86.2	82.8	87.7
Linear SVM	79.5	83.1	85.6	80.6	85.6
New Dense layer	80.4	86.4	84.4	80.1	87.0





Dempster-Shafer theory for fusion of decisions for video frames

CNN	Aggregation	Eurecom Kinect	IMFDB	AFEW	IJB-A
gender_net	Simple Voting	0.73	0.71	0.75	0.60
	Product rule	0.77	0.75	0.75	0.59
	Dempster-Shafer	0.83	0.80	0.78	0.63
DEX	Simple Voting	0.84	0.81	0.80	0.81
	Product rule	0.84	0.88	0.81	0.82
	Dempster-Shafer	0.92	0.90	0.90	0.81
Our MobileNet	Simple Voting	0.94	0.98	**0.93**	0.95
	Product rule	0.93	**0.99**	**0.93**	**0.96**
	Dempster-Shafer	0.95	0.98	**0.93**	**0.96**
Our MobileNet, quantized	Simple Voting	0.88	0.96	0.92	0.93
	Product rule	0.86	0.96	0.93	0.94
	Dempster-Shafer	0.89	0.97	0.93	0.94
Our MobileNet, fine-tuned	Simple Voting	0.93	0.95	0.91	0.94
	Product rule	**0.95**	0.97	0.92	0.95
	Dempster-Shafer	**0.96**	0.96	0.92	0.95

[Kharchevnikova , Savchenko, Optical Memory and Neural Networks (Information Optics), 2018] “positive” “neutral” “negative”

EmotiW2017 sub-challenge:

- 3 classes (positive, neutral, negative)
- Training set: 3630 photos
- Validation set: 2065 photos

We got the 4th place (75.4% accuracy on validation set - 23+% better than the baseline, 78.53% on test set - 25+% than the baseline)

Our ensemble based on Dlib face detector and VGGFace was too complicated

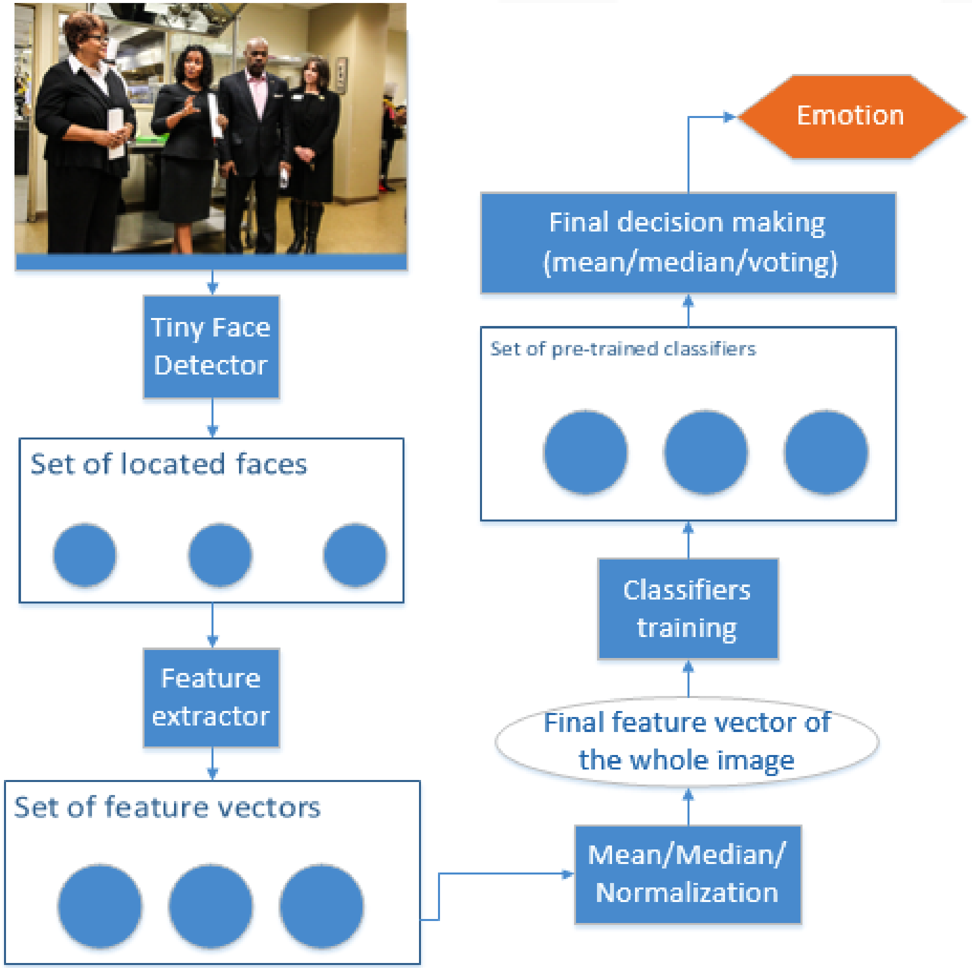
[Rassadin, Gruzdev, Savchenko, ICMI, 2017]

Leaderboard:

Group-level Emotion Recognition	
SIAR	80.89%
UDI-GPB	80.61%
BNU	79.79%
AMD	**78.53%**
AmritaEEE	75.07%

Small convolutional neural networks (CNNs) pre-trained on emotion recognition with FER-2013 (64x64 facial images)

Pipeline



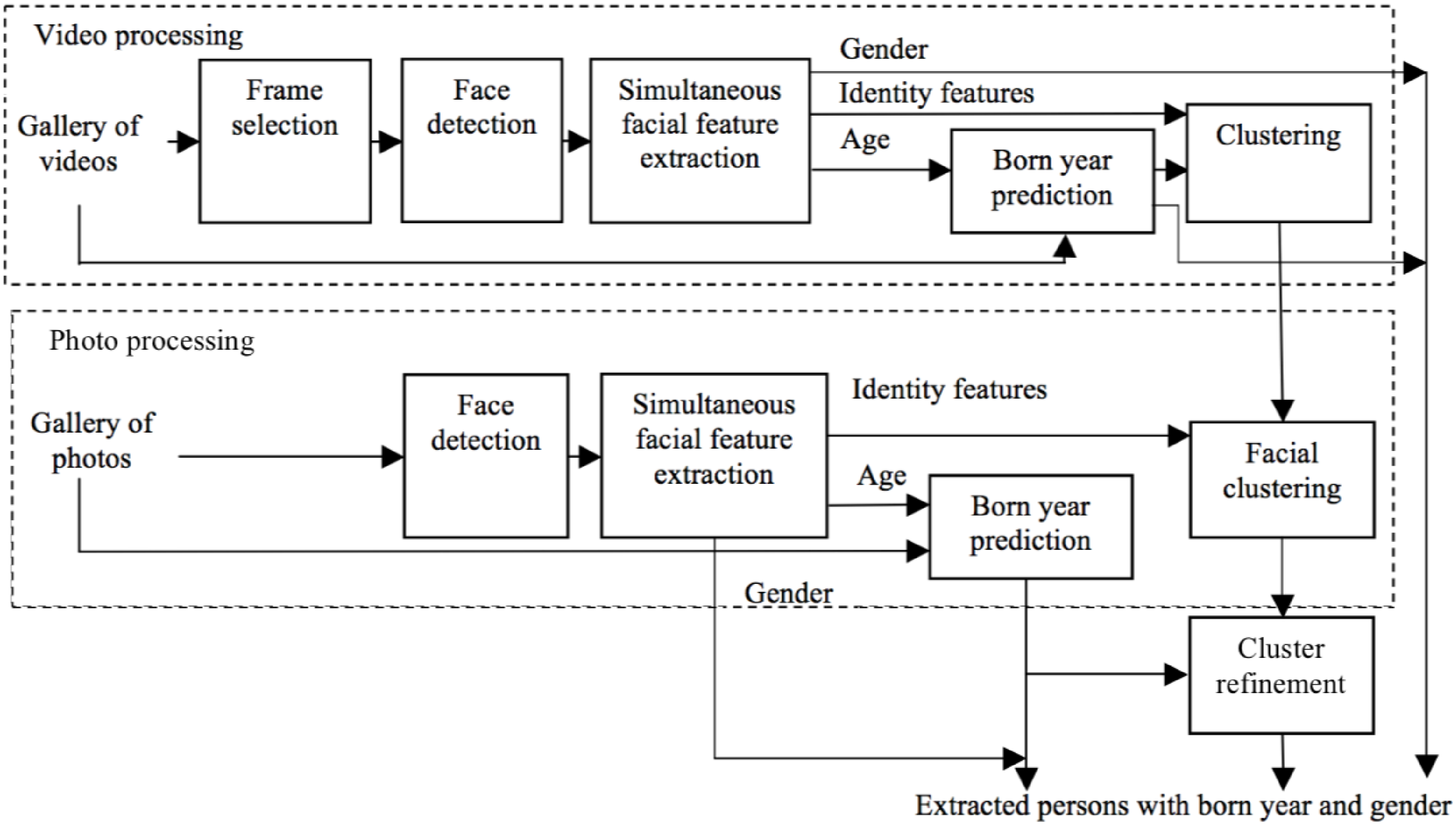
[Tarasov, Savchenko, AIST 2018]

Experimental results on EmotiW 2017 data

	CNN-5		FCN	
Classifier	**7 softmax scores**	**128 embeddings**	**7 softmax scores**	**112 embeddings**
RBF SVM	57.0	60.5	68.5	69.3
Gradient Boosting	61.3	61.9	67.2	70.1
Extra Trees	59.1	61.9	69.6	70.2
Linear SVM	59.6	61.5	69.8	70.3
Bagging	59.0	61.5	68.8	70.7
Random Forest (RF)	58.0	59.9	69.0	70.9
AdaBoost	57.4	58.1	69.5	70.9
Proposed approach (all classifiers)	61.5	62.2	70.7	72.3
Proposed approach (RF, RBF SVM and Linear SVM)	**61.9**	**64.2**	**73.2**	**75.5**

Organizing photo and video albums on mobile device

Organizing Photo and Video Albums



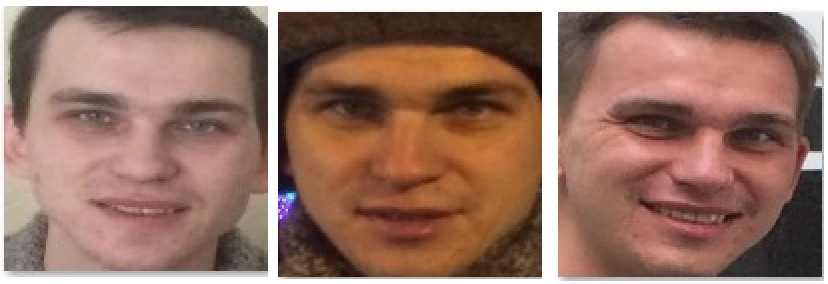
Experimental results: GFW (Grouping Faces in the Wild) dataset

		K/C	ARI	AMI	Homogeneity	Completeness	F-measure
Single	VGGFace	4.10	0.440	0.419	0.912	0.647	0.616
	VGGFace2	3.21	0.580	0.544	0.942	0.709	0.707
	Ours	4.19	0.492	0.441	0.961	0.655	0.636
Average	VGGFace	1.42	0.565	0.632	0.860	0.751	0.713
	VGGFace2	1.59	0.603	**0.663**	0.934	0.761	0.746
	Ours	1.59	**0.609**	0.658	0.917	**0.762**	**0.751**
Complete	VGGFace	**0.95**	0.376	0.553	0.811	0.690	0.595
	VGGFace2	1.44	0.392	0.570	0.916	0.696	0.641
	Ours	1.28	0.381	0.564	0.886	0.693	0.626
Weighted	VGGFace	1.20	0.464	0.597	0.839	0.726	0.662
	VGGFace2	1.05	0.536	0.656	0.867	**0.762**	0.710
	Ours	1.57	0.487	0.612	0.915	0.727	0.697
Median	VGGFace	5.30	0.309	0.307	0.929	0.587	0.516
	VGGFace2	4.20	0.412	0.422	0.929	0.639	0.742
	Ours	6.86	0.220	0.222	**0.994**	0.552	0.411
Rank-Order	VGGFace	0.82	0.319	0.430	0.650	0.694	0.630
	VGGFace2	1.53	0.367	0.471	0.937	0.649	0.641
	Ours	1.26	0.379	0.483	0.914	0.658	0.652

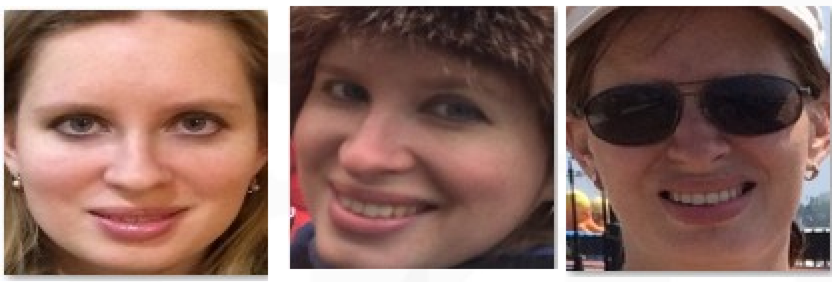
In average, C = 46 subjects

Example (1)

male 1987

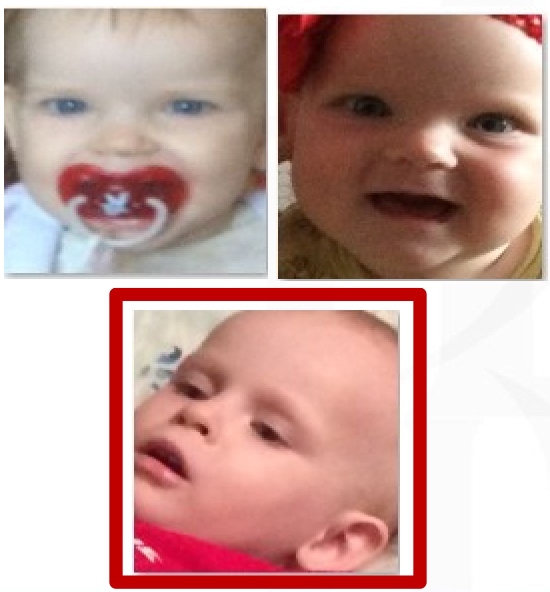


female 1982

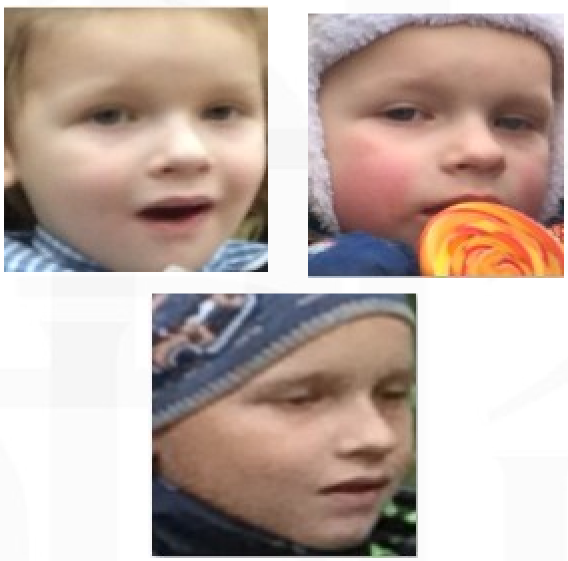


female 2015

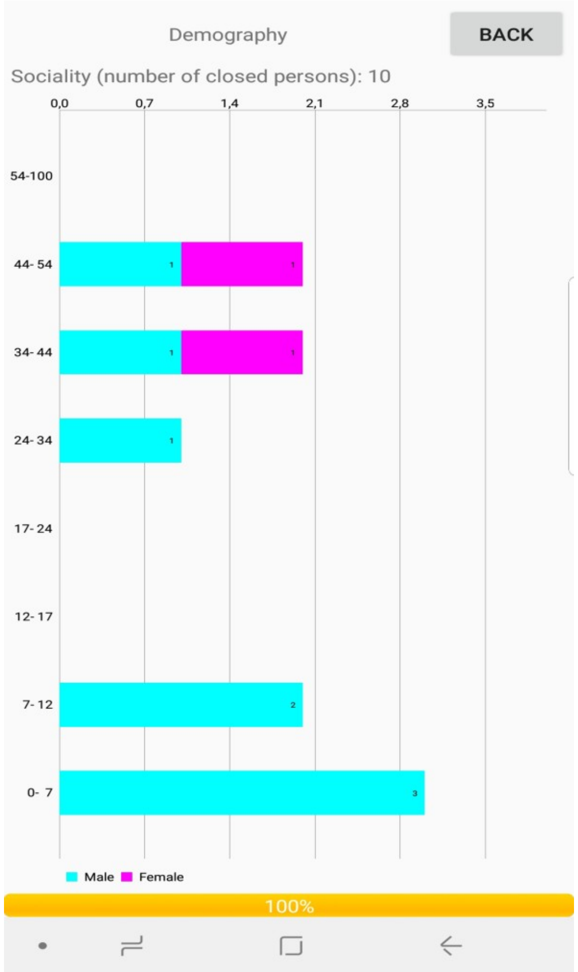
female 2016



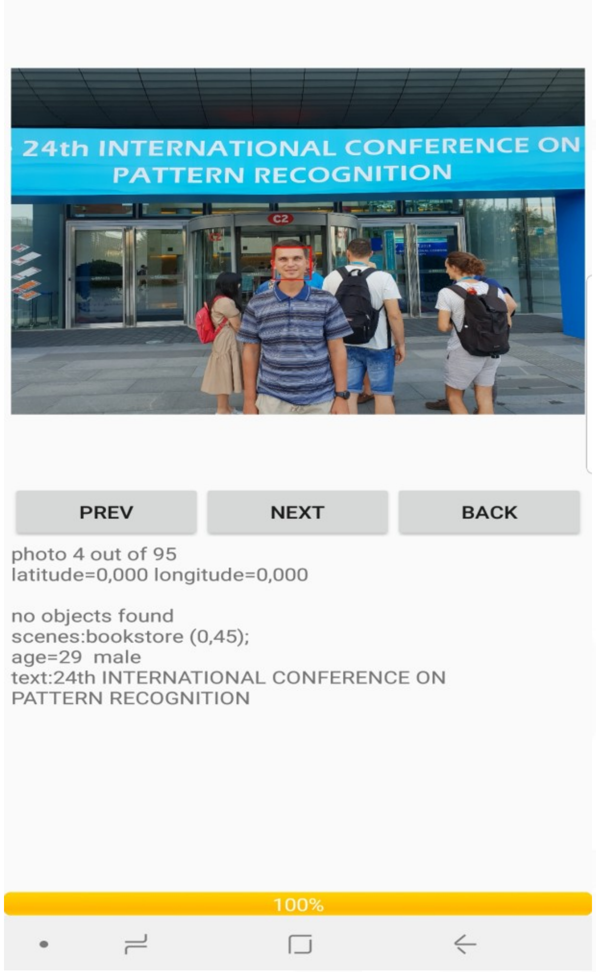
male 2009



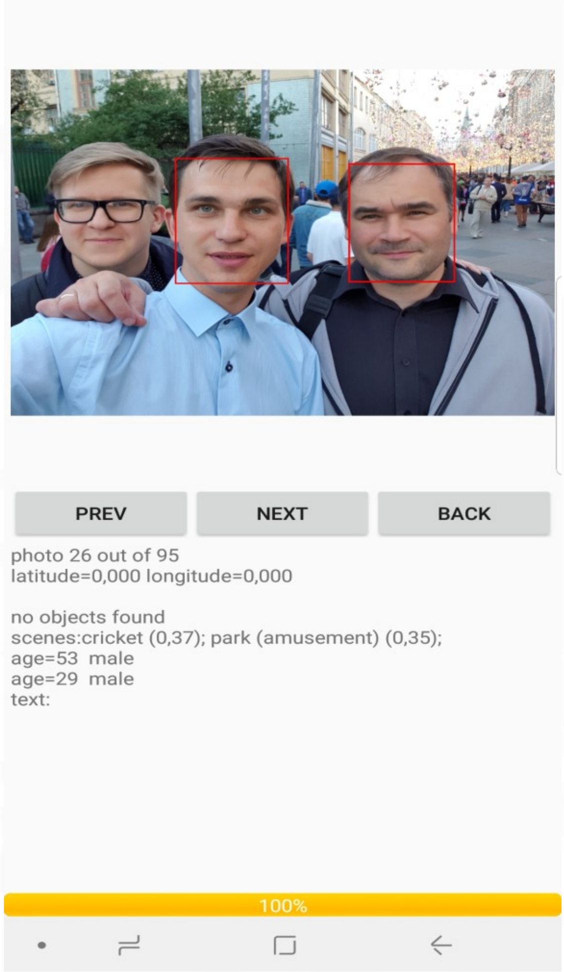
Example (2)



Demography
BACK
Sociality (number of closed persons): 10
0,0 0,7 1,4 2,1 2,8 3,5
54-100
44- 54
34- 44
24- 34
17- 24
12- 17
7- 12
0- 7
Male Female
100%



24th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
PREV
NEXT
BACK
photo 4 out of 95
latitude=0,000 longitude=0,000
no objects found
scenes:bookstore (0,45);
age=29 male
text:24th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
100%



PREV
NEXT
BACK
photo 26 out of 95
latitude=0,000 longitude=0,000
no objects found
scenes:cricket (0,37); park (amusement) (0,35);
age=53 male
age=29 male
text:
100%

Conclusion

1. **State-of-the-art results are obtained with modern deep CNNs, but sometimes their usage is restricted due to performance (low speed) issues**
2. **Sequential analysis of CNN features/layers can potentially provide high performance without losses in accuracy**
3. **Combination of conventional machine learning techniques with face descriptors extracted by contemporary CNNs has proved its potential in approximate nearest neighbor search and modifications of instance-based learning methods**

Thank you!