



NATIONAL RESEARCH
UNIVERSITY

Open-set Face Identification Algorithms with Automatic Detection of Rare Data

Sokolova Anastasiia
adsokolova@hse.ru



INTRODUCTION

- Image Gallery Navigation Systems
- Grouping video data for statistics
- Recognition of objects in the public safety field





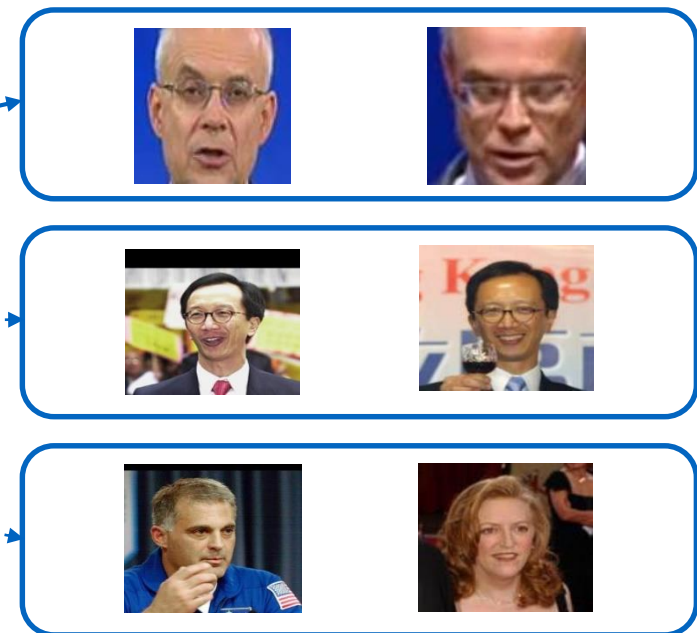
IMAGE RECOGNITION TASK

Input: Face image

Task: Determine the class membership (Open-Set classification)



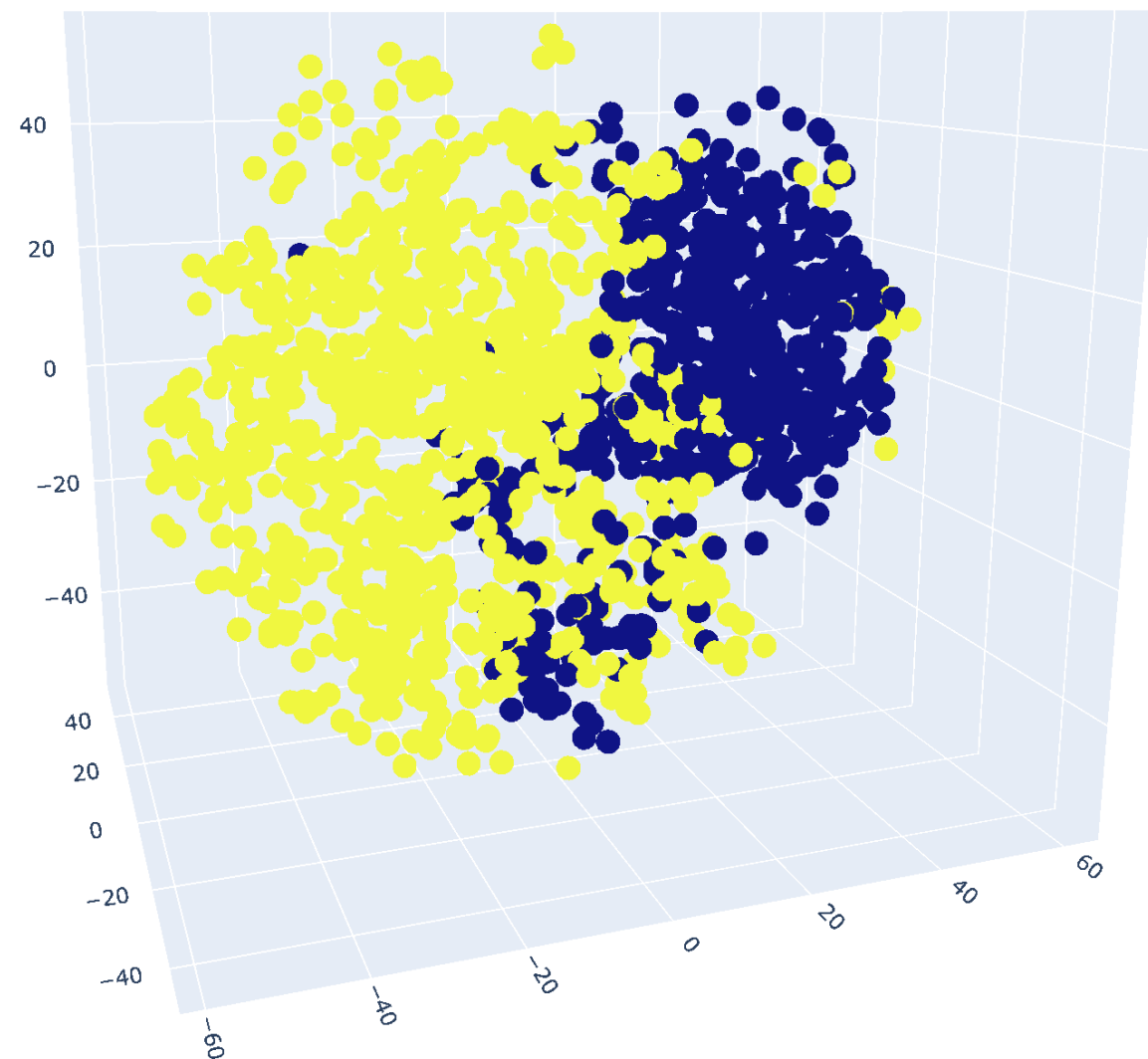
?



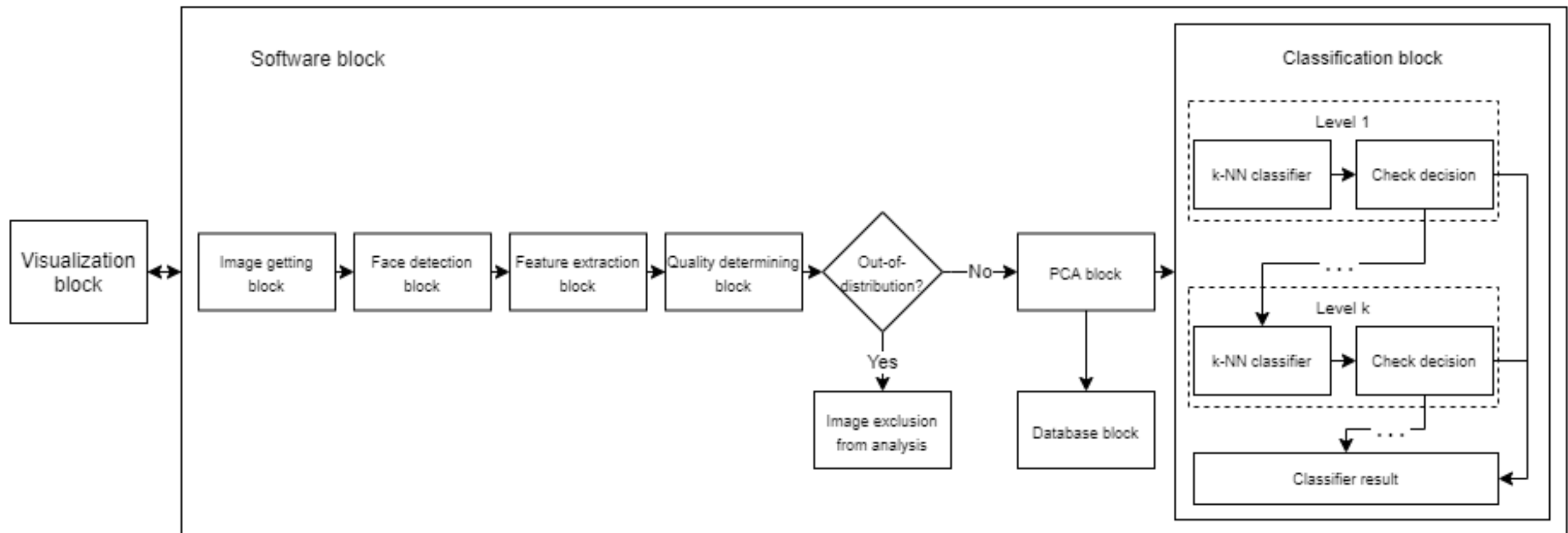
Issue: Images can have different quality, foreshortening, brightness. Moreover, people change their appearance



«OUT-OF-DISTRIBUTION» IMAGES



PROPOSED APPROACH



FEATURE EXTRACTION

- ResNet50 (VGGFace2)

Vector size: 1x2048

Inference time (CPU): 4,39 c

Model size: 92 M6

- MobileNet2

Vector size: 1x1024

Inference time (CPU): 0,98 c

Model size: 12,7 M6

- InsightFace (LResNet100E-IR)

Vector size: 1x512

Inference time (CPU): 1,43 c

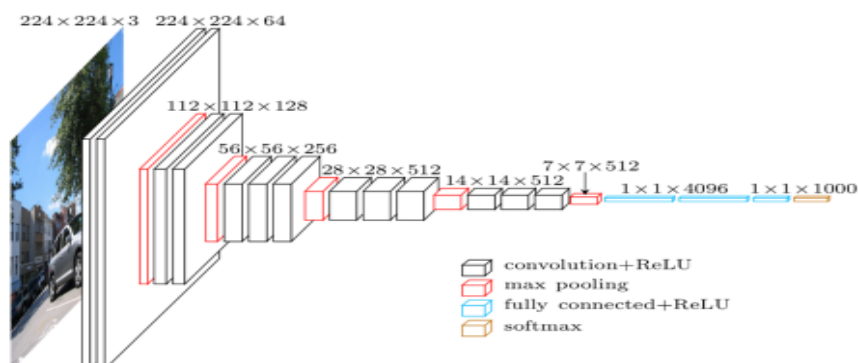
Model size: 170 M6

- FaceNet

Vector size: 1x512

Inference time (CPU): 3,09 c

Model size: 93,5 M6



- VGGFace2
 - 9 000 people
 - > 3.3 billions face images
 - ~362 images per person
- MS-Celeb-1M (Microsoft Celeb)
 - 100 000 people
 - 10 billions face images
 - ~100 images per person
- LWF (Labeled Faces in the Wild)
 - 1680 people
 - 13 000 face images
 - 1-10 images per person
- All ages faces dataset
 - 13 322 people
 - 13 322 face images
 - Ages from 2 to 80
- Large age gap face verification
 - 1 010 people
 - 3 828 face images
 - Per person: current ages images and young ages images

http://www.robots.ox.ac.uk/~vgg/data/vgg_face2/

<https://megapixels.cc/msceleb/>

<http://vis-www.cs.umass.edu/lfw/>

<https://github.com/JingchunCheng/All-Age-Faces-Dataset>

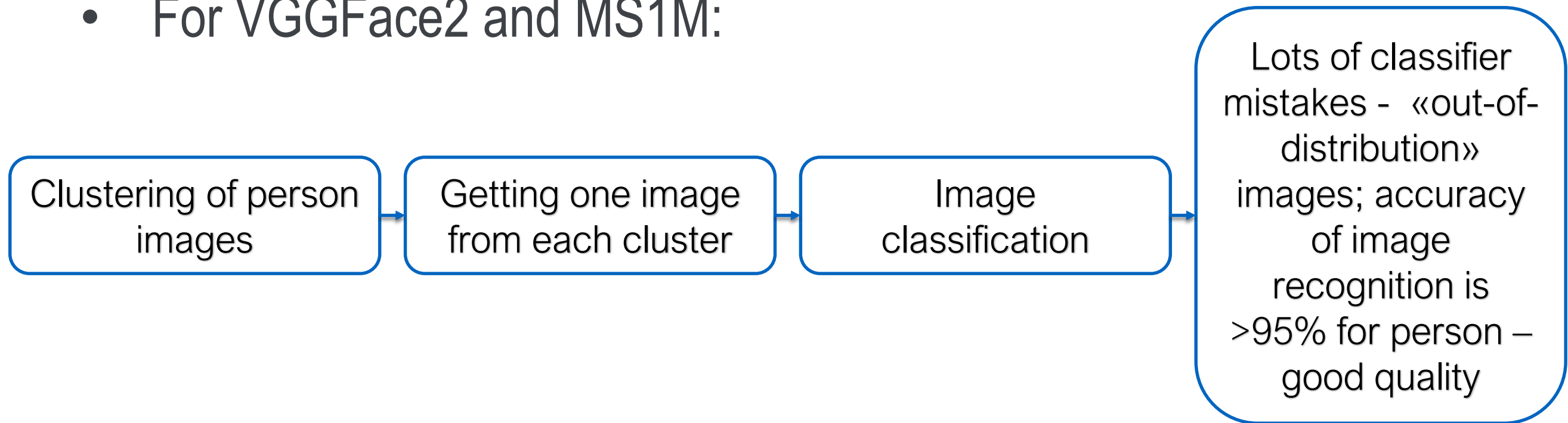
<http://www.ivl.disco.unimib.it/activities/large-age-gap-face-verification/>





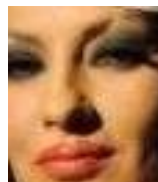
DATASET CREATION (WITH «OUT-OF-DISTRIBUTION» IMAGES)

- For VGGFace2 and MS1M:



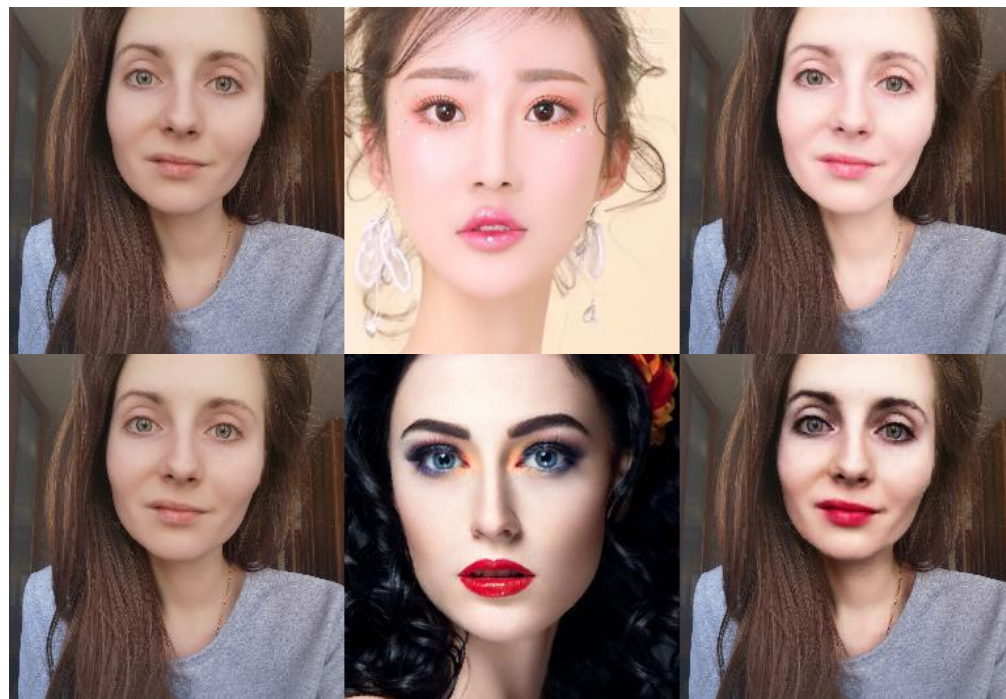
- All ages faces dataset and Large age gap face verification: children and elderly images as «out-of-distribution»
- Additional transformation

TRANSFORMATION





TRANSFORMATION



Zhang H. et al. Disentangled Makeup Transfer with Generative Adversarial Network //arXiv preprint arXiv:1907.01144. – 2019

Chen Y. C. et al. Facelet-bank for fast portrait manipulation //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. – 2018. – C. 3541-3549



«OUT-OF-DISTRIBUTION» IMAGES DETECTION

Data: 26 000 normal + 23 000 «out-of-distribution» from VGGFace2, whole All ages faces dataset and Large age gap face verification, 50 000 normal + 41 000 «out-of-distribution» from MS1M dataset. For model training used 70% of data, for model training – 30%.

- Used convolutional neural networks with pretrained ImageNet weights

CNN	Accuracy
EfficientNetB3	58.9%
MobileNetV2	56.3%
ResNet50	58.8%
Xception	49.7%
InceptionV3	55.4%

Tan M., Le Q. V. Efficientnet: Rethinking model scaling for convolutional neural networks // arXiv preprint arXiv:1905.11946. – 2019

Sandler M., Howard A., Zhu M., Zhmoginov A., Chen L. C. Mobilenetv2: Inverted residuals and linear bottlenecks // In Proceedings of the IEEE conference on computer vision and pattern recognition. – 2018. – P. 4510-4520

He K., Zhang X., Ren S., Sun J. Deep residual learning for image recognition // In Proceedings of the IEEE conference on computer vision and pattern recognition. – 2016. – P. 770-778

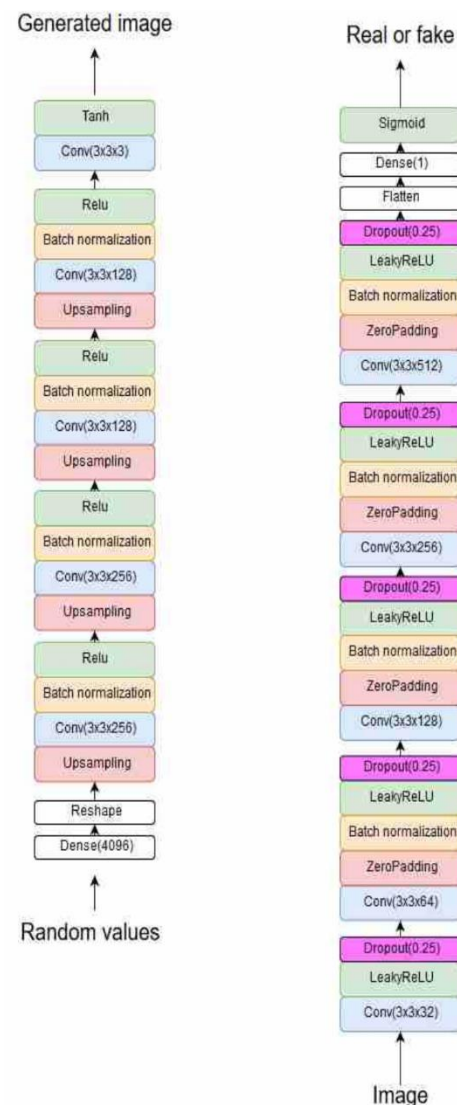
Szegedy C., Vanhoucke V., Ioffe S., Shlens J., Wojna Z. Rethinking the inception architecture for computer vision // In Proceedings of the IEEE conference on computer vision and pattern recognition. – 2016. – P. 2818-2826

Chollet F. Xception: Deep learning with depthwise separable convolutions // In Proceedings of the IEEE conference on computer vision and pattern recognition. – 2017. – P. 1251-1258

«OUT-OF-DISTRIBUTION» IMAGES DETECTION

- Generative adversarial network

Network architecture



Generator results



Accuracy of «out-of-distribution» images detection, using descriptor

69.4%

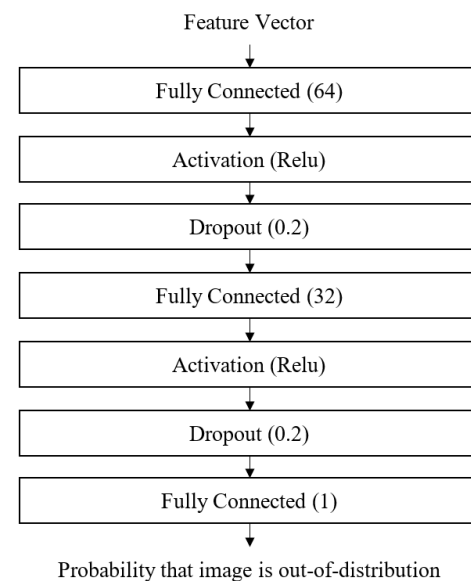


«OUT-OF-DISTRIBUTION» IMAGES DETECTION

- Classifiers (input – feature vector)

Algorithm	ResNet50	MobileNet2	InsightFace	FaceNet
k-NN	89.7%	93.9%	89.6%	90.9%
SVM	69.1%	79.4%	63.0%	64.3%
RandomForest	85.0%	86.5%	80.6%	84.2%
LDA	88.1%	89.0%	83.7%	87.4%
Our NN	92.3%	94.8%	89.7%	90.2%

Our NN:





RESULTS OF IMAGE CLASSIFICATION WITH «OUT-OF-DISTRIBUTION» DATA DETECTION

- Results of image classification (VGGFace2)

Algorithm	Accuracy (%)	Time of one image classification (ms)
k-NN + all	94.9	23.39
k-NN + exclusion	97.9	9.59
k-NN + transform	94.9	24.36
RandomForest + all	21.2	3.80
RandomForest + exclusion	91.6	2.73
RandomForest + transform	21.8	4.02
LDA + all	92.6	1.27
LDA + exclusion	97.1	1.14
LDA + transform	93.0	1.29

CNN	Algorithm	Accuracy (%)	Time of one image classification (ms)
ResNet50	LDA + transform	93.3	1.35
InsightFace	k-NN + transform	83.9	5.94
	LDA + all	75.7	0.23
FaceNet	k-NN + transform	97.7	5.66
	LDA + all	92.5	0.23

Data: 26 172 normal images + 23 653 «out-of-distribution» images, 188 classes

- Results of image classification (MS1M)

Algorithm	Accuracy (%)	Time of one image classification (ms)
k-NN + all	87.6	42.24
k-NN + exclusion	91.9	21.02
k-NN + transform	87.7	43.28
RandomForest + all	50.1	6.34
RandomForest + exclusion	79.8	5.21
RandomForest + transform	50.1	6.35
LDA + all	88.7	1.94
LDA + exclusion	91.5	1.35
LDA + transform	88.6	1.96

Сеть	Алгоритм	Точность (%)	Time of one image classification (ms)
MobileNet2	LDA + transform	91.0	1.12
InsightFace	k-NN + transform	84.1	6.17
	LDA + transform	67.9	0.81
FaceNet	k-NN + transform	89.3	5.68
	LDA + transform	82.3	0.84

Data: 50 597 normal images + 41 665 «out-of-distribution» images, 320 classes



RESULTS OF IMAGE CLASSIFICATION WITH «OUT-OF-DISTRIBUTION» DATA DETECTION

- Confusion matrix of classification «out-of-distribution» images (VGGFace2)

CHC	TPR	TNR	FPR	FNR
ResNet50	97.29%	93.21%	6.79%	2.71%
MobileNet2	98.44%	98.14%	1.86%	1.56%
InsightFace	96.55%	89.57%	10.43%	3.45%
FaceNet	94.73%	90.52%	9.48%	5.27%

- Confusion matrix of classification «out-of-distribution» images (MS1M)

CHC	TPR	TNR	FPR	FNR
ResNet50	96.54%	92.46%	7.54%	3.46%
MobileNet2	97.28%	91.57%	8.43%	2.72%
InsightFace	93.16%	88.95%	11.05%	6.84%
FaceNet	94.55%	90.69%	9.31%	5.45%



SEQUENTIAL CLASSIFICATION

- The distance between feature level to speed-up the matching

$$\rho(\tilde{x}^{(l)}, x^{(l)}) = \rho(\tilde{x}^{(l-1)}, x^{(l-1)}) + \sum_{d=d_{(l-1)}+1}^{d_l} \rho(\tilde{x}_d, x_{r;d})$$

- Nearest neighbor class

$$c_l^* = \operatorname{argmin}_{c \in C_l} \rho_c(\tilde{x}^{(l)})$$

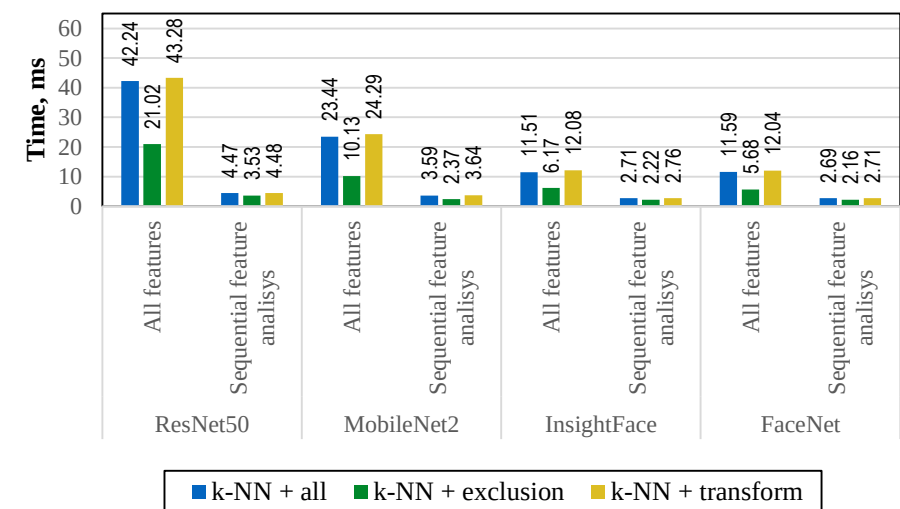
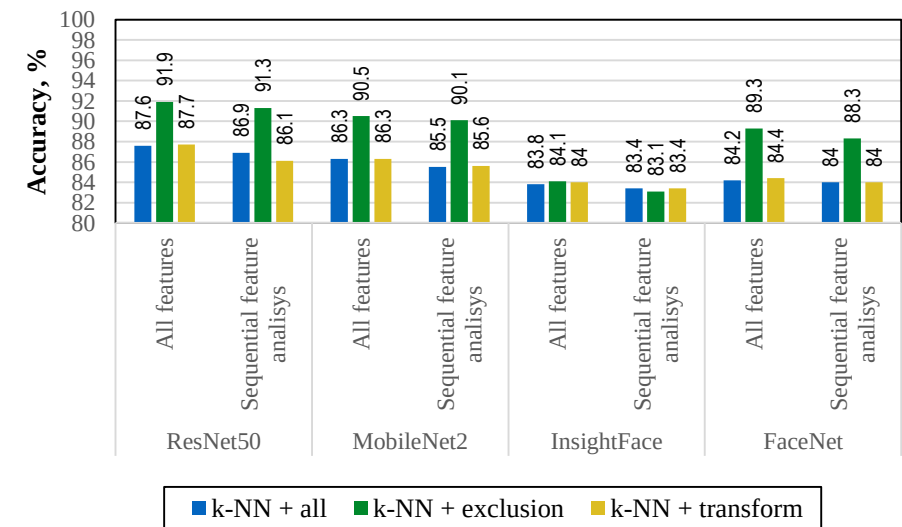
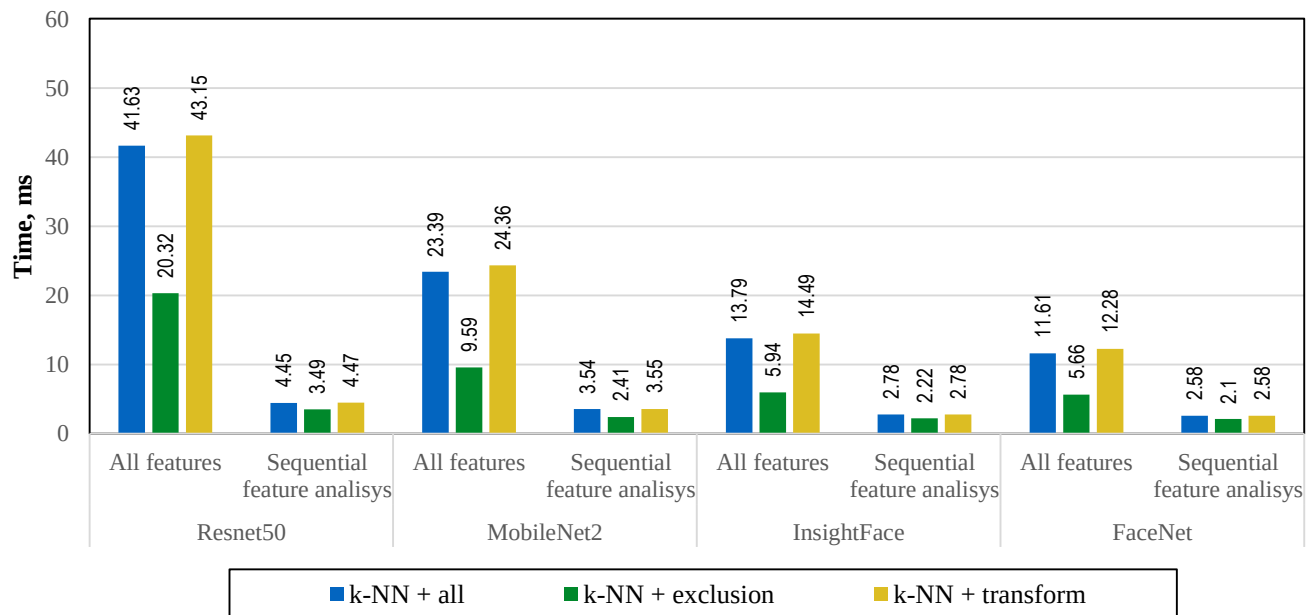
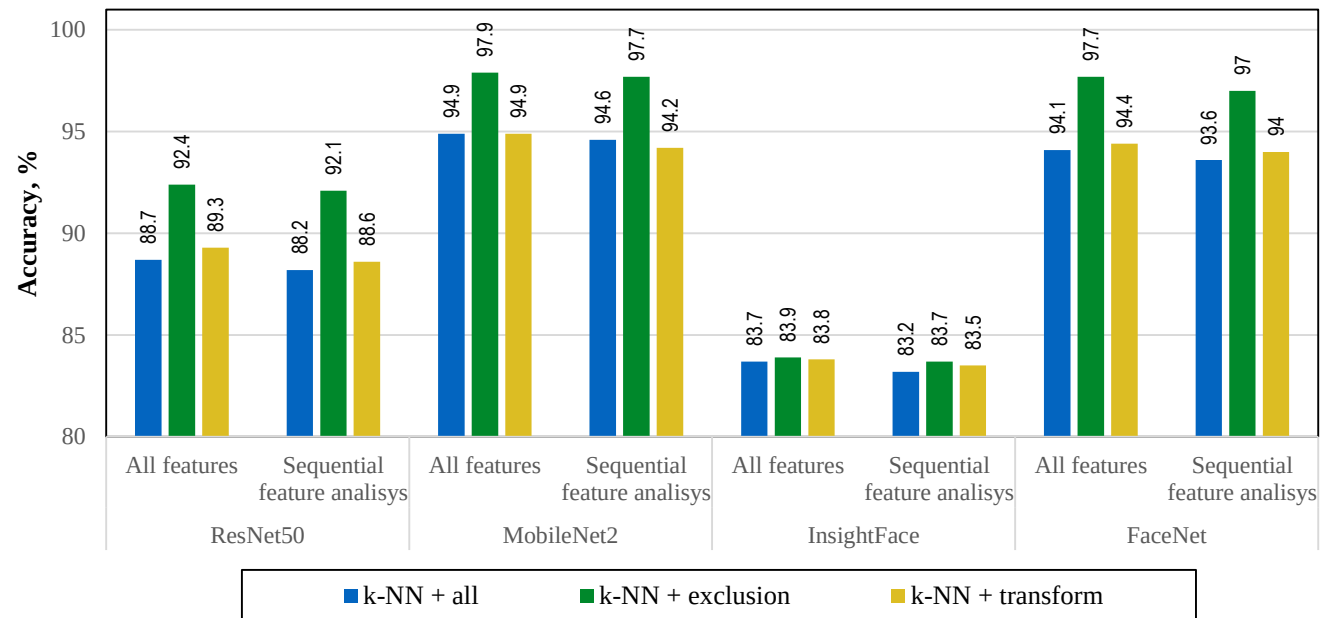
- The set of candidates

$$C_{l+1} = \left\{ c \in C_l \mid \frac{\rho_c(\tilde{x}^{(l)})}{\rho_{c_l^*}(\tilde{x}^{(l)})} \leq \delta \right\}$$



RESULTS OF SEQUENTIAL CLASSIFICATION

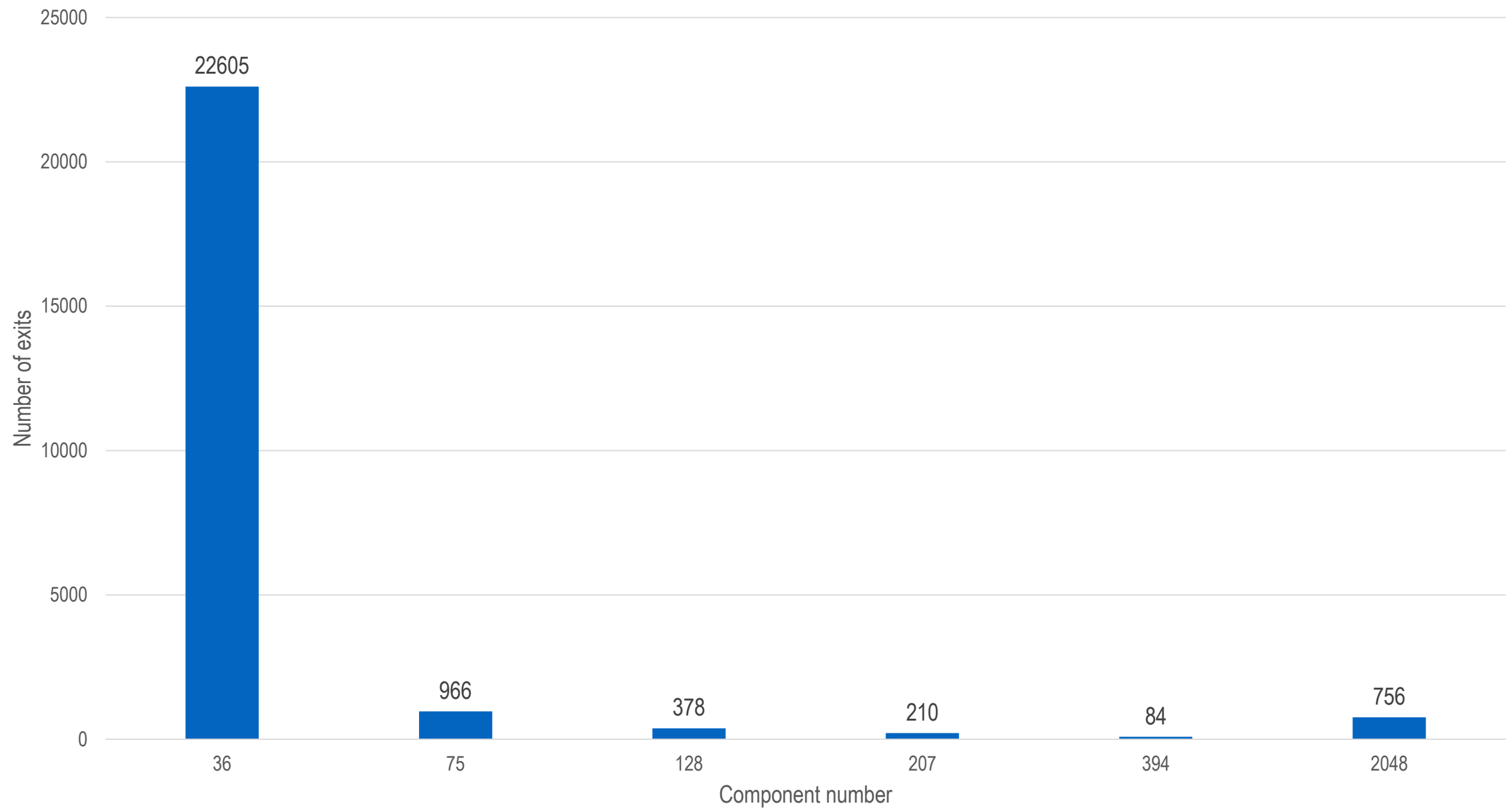
VGGFace2



MS1M



RESULTS OF SEQUENTIAL CLASSIFICATION (VGGFACE2)





RESULTS OF SEQUENTIAL CLASSIFICATION (LFW)

Metrics	Classifier	Image selection	ResNet50	MobileNet2	InsightFace	FaceNet
Accuracy (%)	k-NN, all features	all	98.66	96.58	92.43	98.15
		exclusion	98.71	96.61	92.51	98.21
	Sequential k-NN, fixed number of components	all	98.32	96.24	91.78	98.15
		exclusion	98.37	96.25	91.79	98.19
	Sequential k-NN, variable number of components	all	97.98	96.28	91.88	98.15
		exclusion	98.02	96.32	91.90	98.18
Time (ms)	k-NN, all features	all	34.78	15.32	9.42	9.37
		exclusion	34.72	15.25	9.21	9.19
	Sequential k-NN, fixed number of components	all	2.84	2.73	2.65	2.69
		exclusion	2.79	2.69	2.41	2.42
	Sequential k-NN, variable number of components	all	2.81	2.16	1.89	1.87
		exclusion	2.78	2.14	1.74	1.73

Data: training set - 8774 images of 5749 classes, testing set – 4459 images of 1680 classes



RESULTS OF SCIKIT-LEARN CLASSIFIERS (LFW)

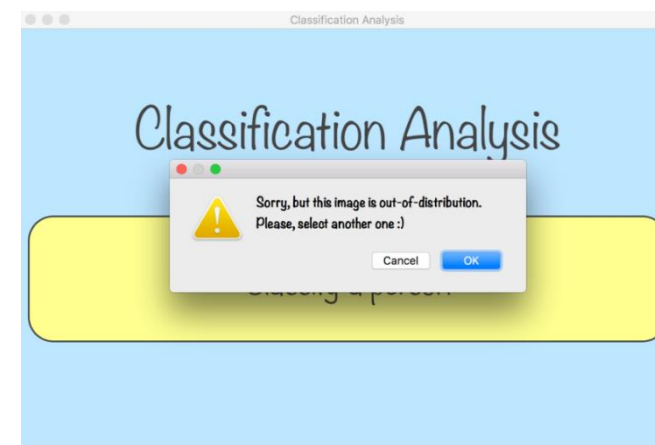
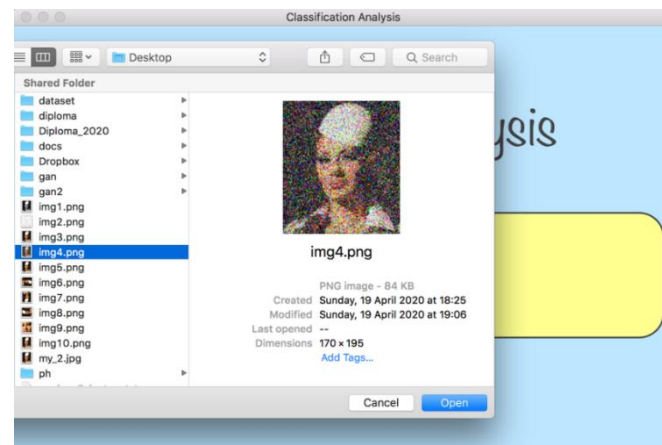
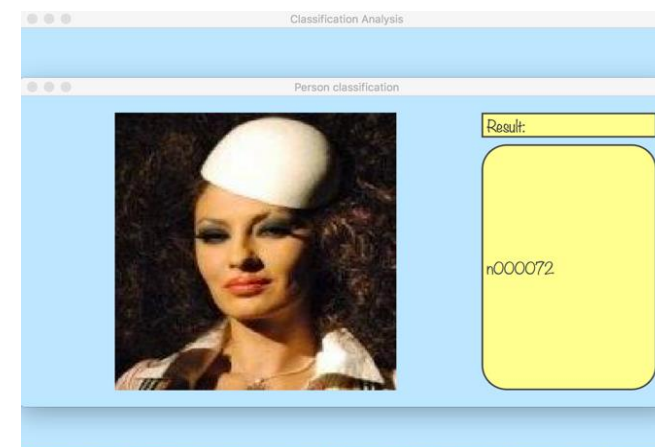
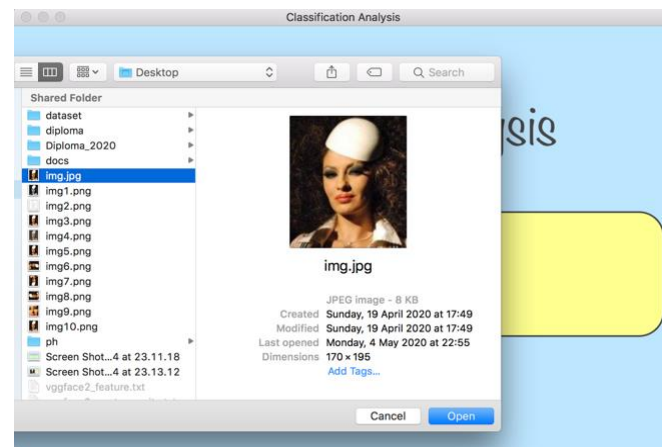
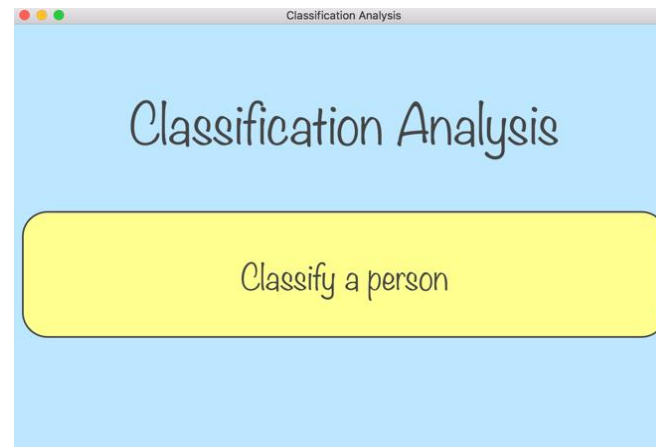
CHC	Component number	SVC	Random Forest	Extremely Randomized Trees	AdaBoost	Gradient Boosting
ResNet50	All features	14.00	81.50	81.50	19.30	22.41
	Fixed number of components	0	79.50	82.10	17.40	21.01
MobileNet2	All features	19.32	82.64	81.95	16.72	22.79
	Fixed number of components	7.46	81.05	80.49	15.39	20.36
InsightFace	All features	10.12	69.50	69.65	19.24	19.41
	Fixed number of components	19.34	64.23	69.98	14.40	19.01
FaceNet	All features	23.88	74.51	75.91	23.84	24.93
	Fixed number of components	44.07	77.87	78.99	24.11	18.80

Accuracy (%)

Time (ms)

CHC	Component number	SVC	Random Forest	Extremely Randomized Trees	AdaBoost	Gradient Boosting
ResNet50	All features	7.17	16.55	1.00	24.98	0.23
	Fixed number of components	3.20	16.14	0.69	21.15	0.16
MobileNet2	All features	4.42	6.55	0.78	18.21	0.35
	Fixed number of components	1.76	6.03	0.51	16.94	0.31
InsightFace	All features	4.12	5.99	0.52	7.99	0.22
	Fixed number of components	2.15	5.34	0.41	7.74	0.19
FaceNet	All features	6.35	6.03	0.42	9.34	0.43
	Fixed number of components	4.63	5.7	0.33	9.01	0.17

SYSTEM PROTOTYPE





CONCLUSION

- Our algorithm is based on the ways to efficiently process «out-of-distribution» data and compute image dissimilarity in order to classify people images
- The most accurate and computationally cheap technique involves «out-of-distribution» images exclusion and sequential feature vector processing



THANK YOU FOR ATTENTION



NATIONAL RESEARCH
UNIVERSITY