Lightweight Low-Resolution Face Recognition for Surveillance Applications

Yoanna Martínez-Díaz, Heydi Méndez-Vázquez
Advanced Technologies Application Center (CENATAV), Havana, Cuba

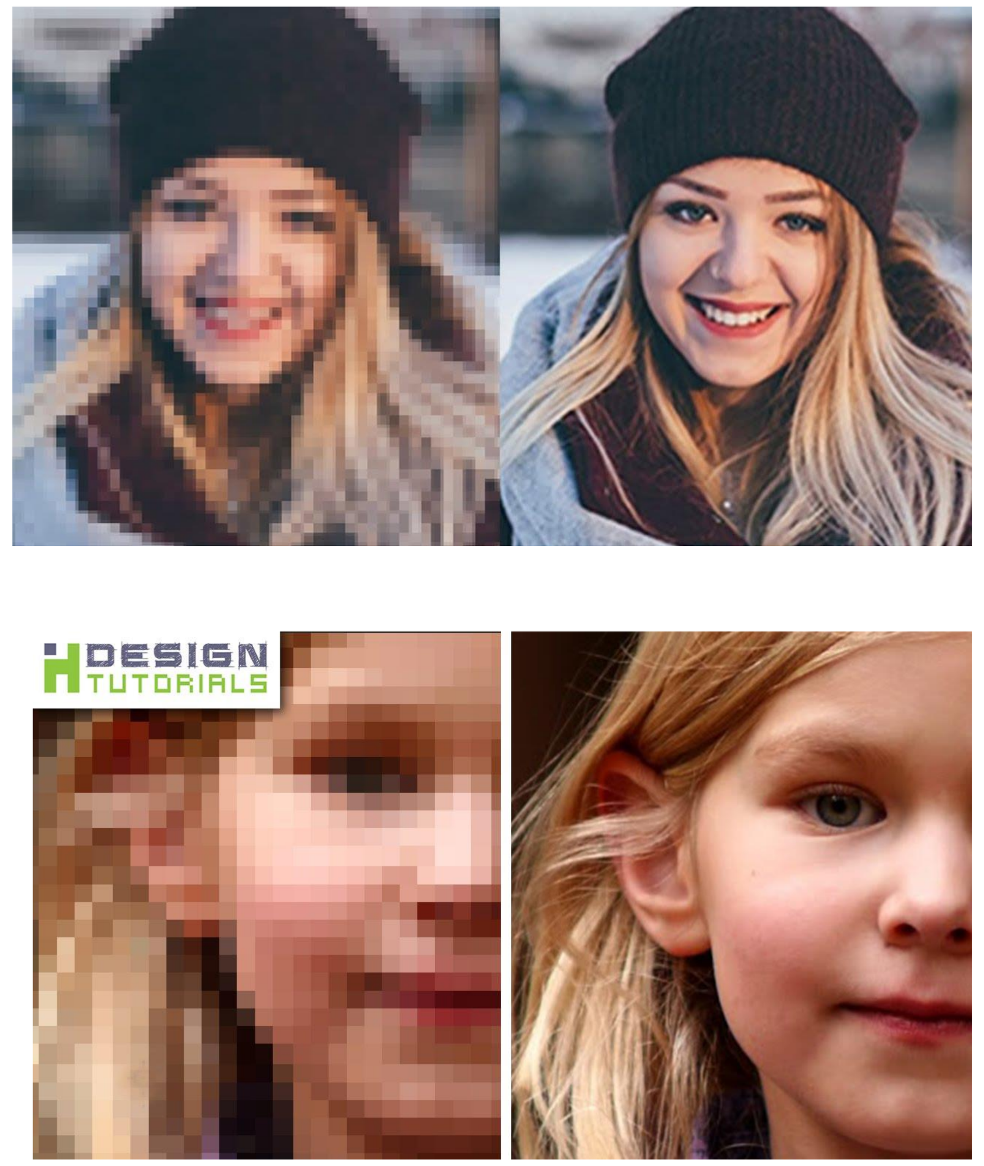
Luis Luevano, Leonardo Chang, Miguel Gonzalez-Mendoza
Tecnologico de Monterrey, Estado de Mexico, Mexico

December, 2020

INTRODUCTION

Surveillance Applications

LR-to-HR matching

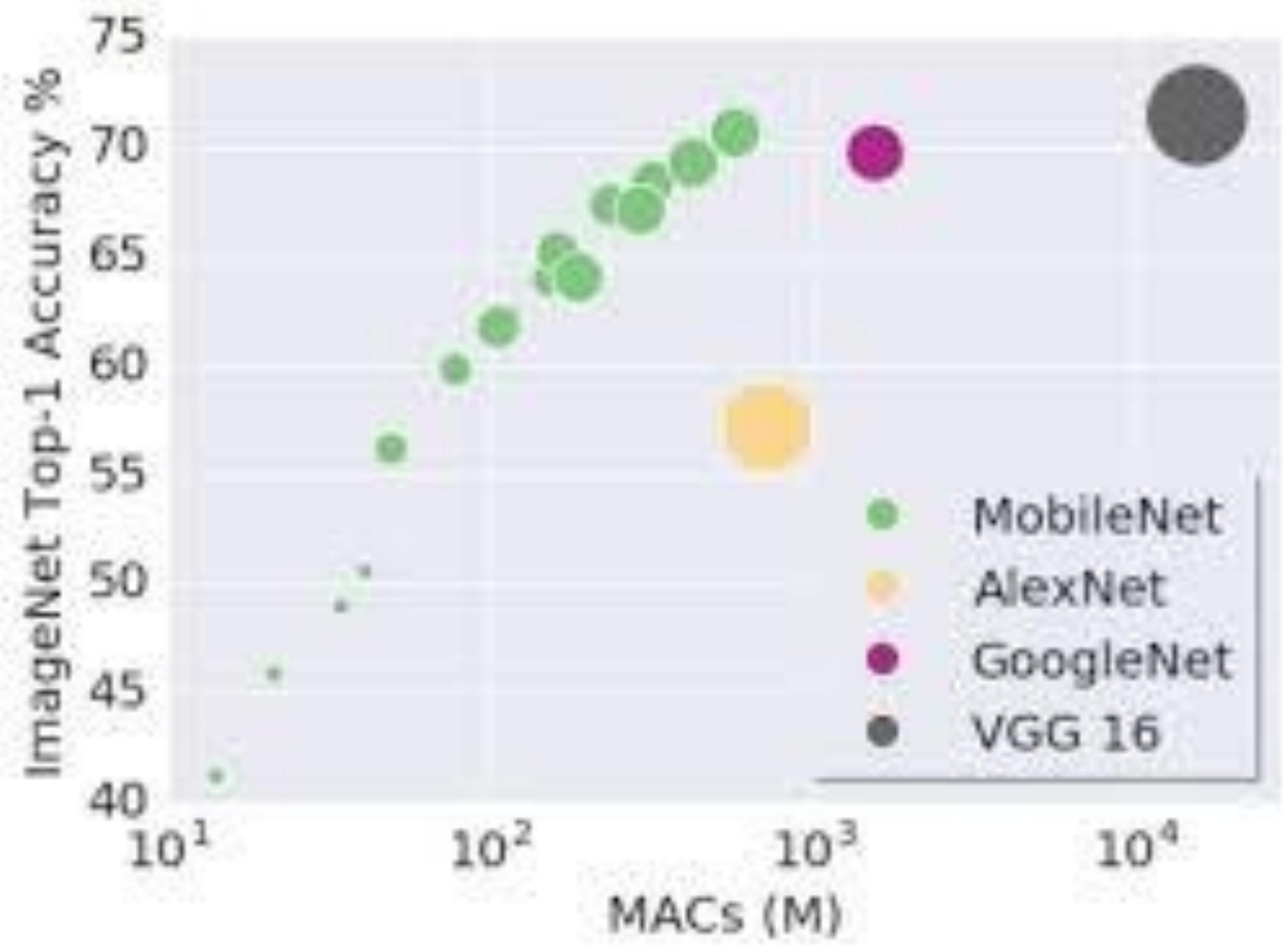


LR-to-LR matching

- Top performing methods are based on very deep CNNs.
- Remarkable success at a high computational cost.
- Demanding powerful computing devices.
- Unfeasible to be employed in practical surveillance applications.

LIGHTWEIGHT ARCHITECTURES

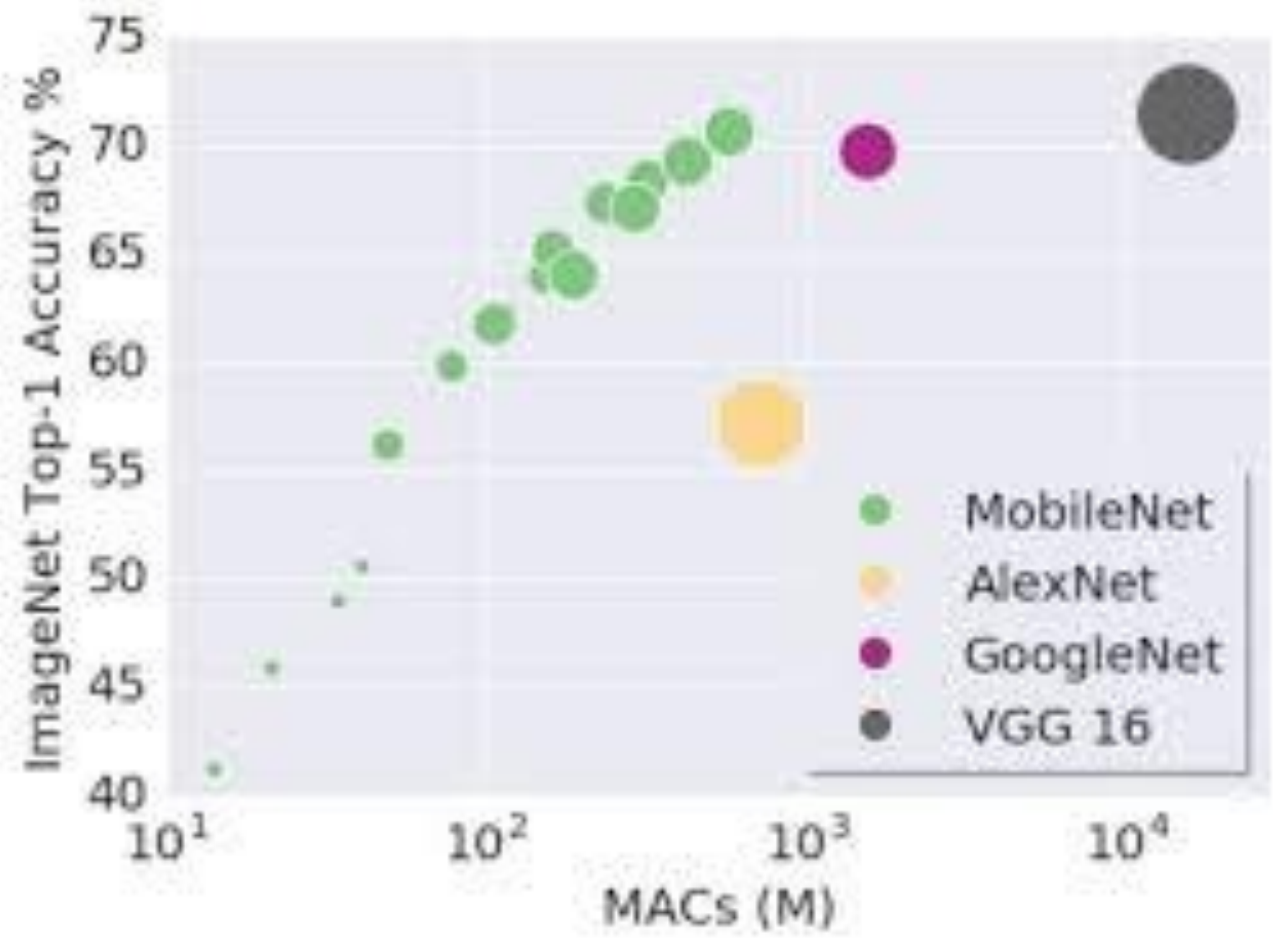
- Better balance between accuracy and efficiency
- More suitable for practical systems



lightweight face architectures: High levels of accuracy on general purpose face verification and identification tasks

LIGHTWEIGHT ARCHITECTURES

- Better balance between accuracy and efficiency
- More suitable for practical systems



lightweight face architectures: High levels of accuracy on general purpose face verification and identification tasks

Baseline lightweight deep face models

MobileFaceNet

ShuffleFaceNet

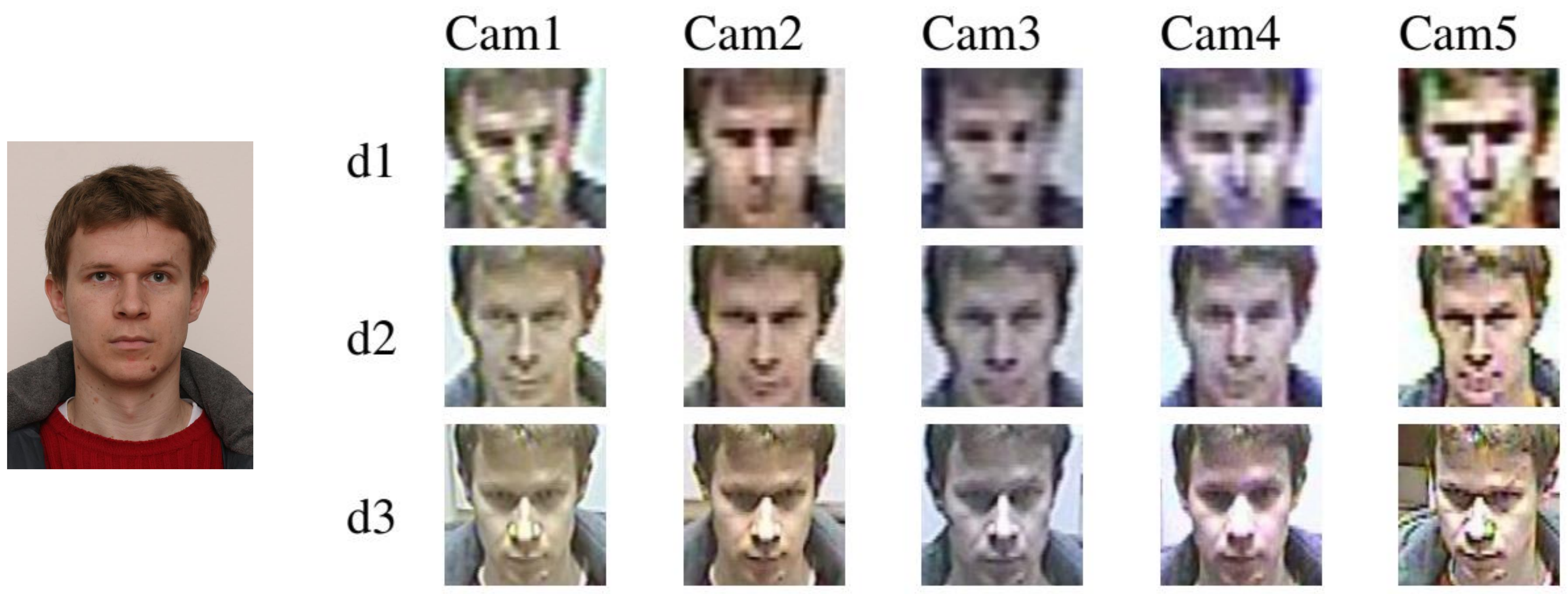
Extremely computation-efficient CNN models.

GDC layer instead of a GAP layer.

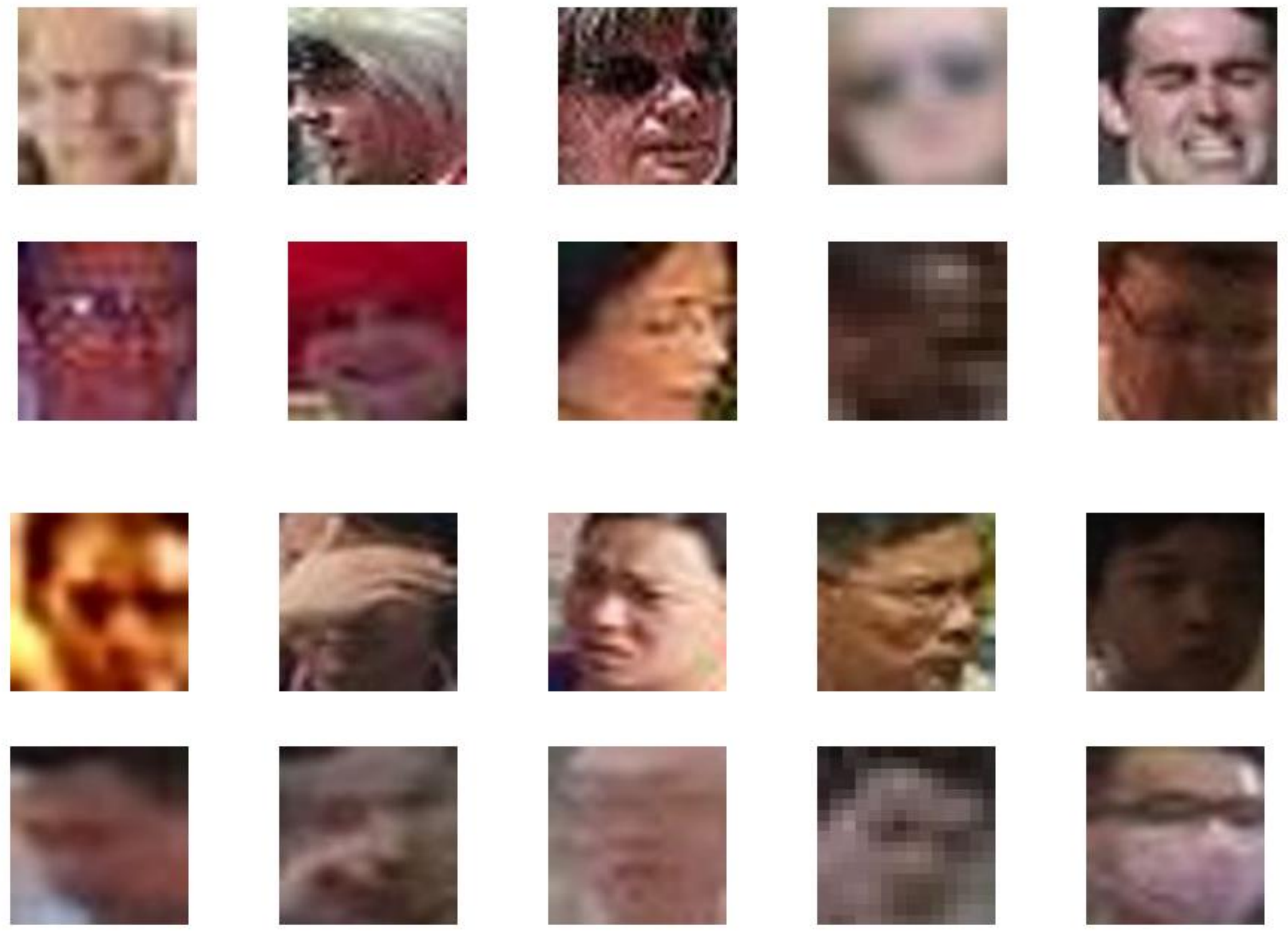
PReLU as non-linear activation function instead ReLU function.

EXPERIMENTAL SETUP

LR-to-HR matching

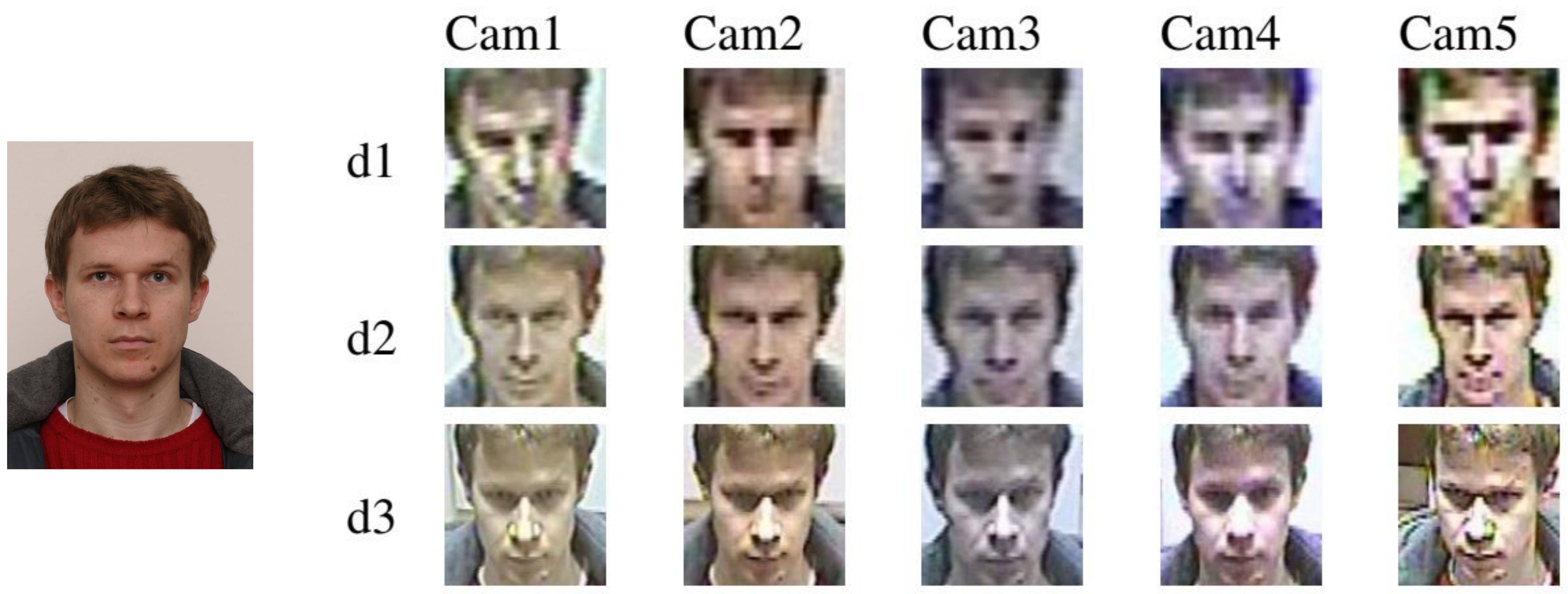


LR-to-LR matching

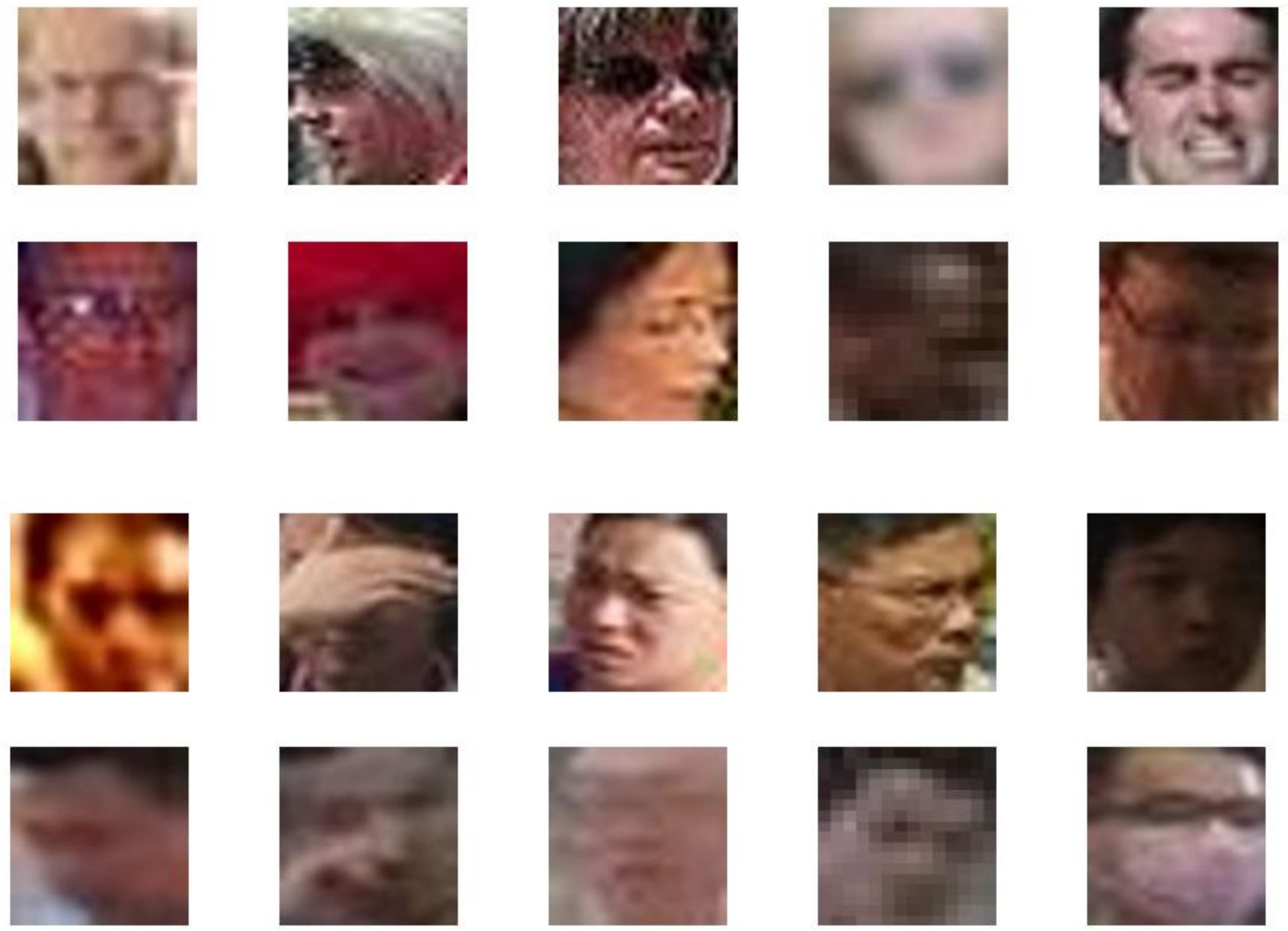


EXPERIMENTAL SETUP

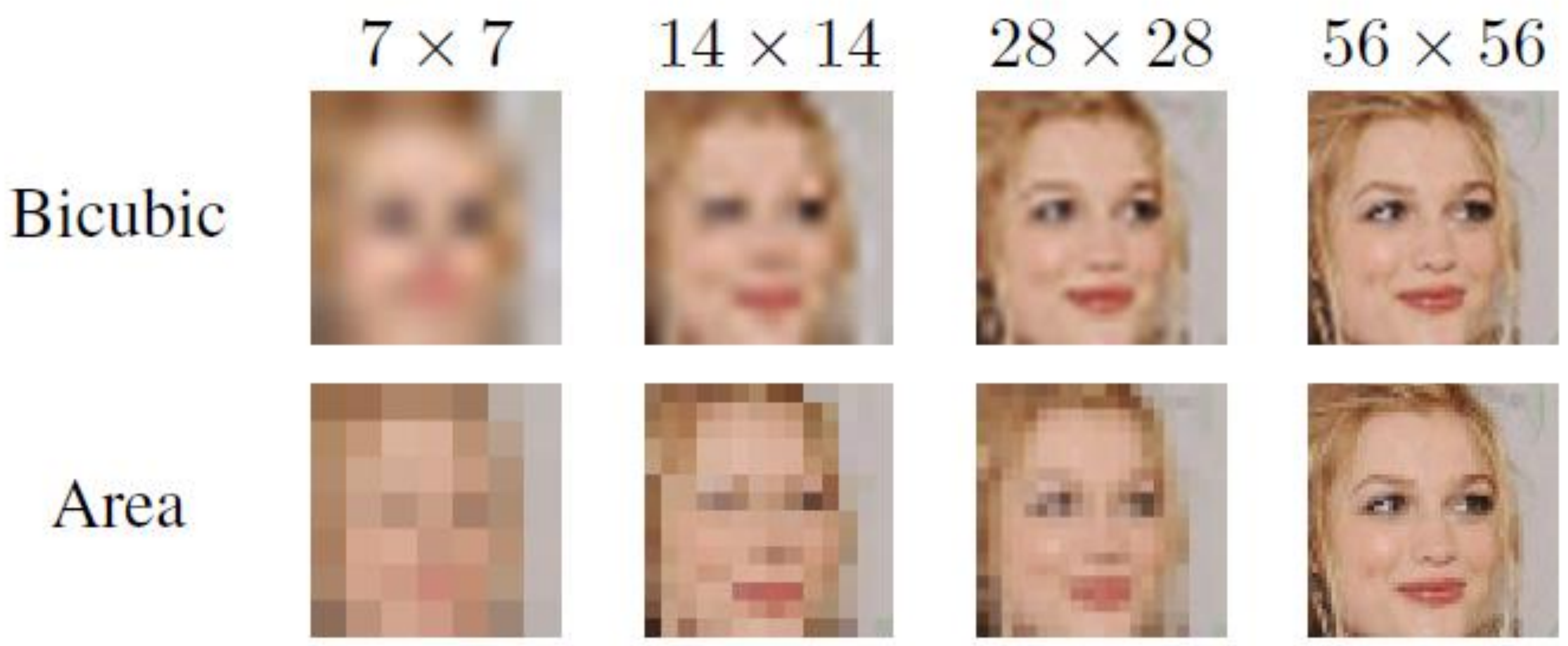
LR-to-HR matching



LR-to-LR matching



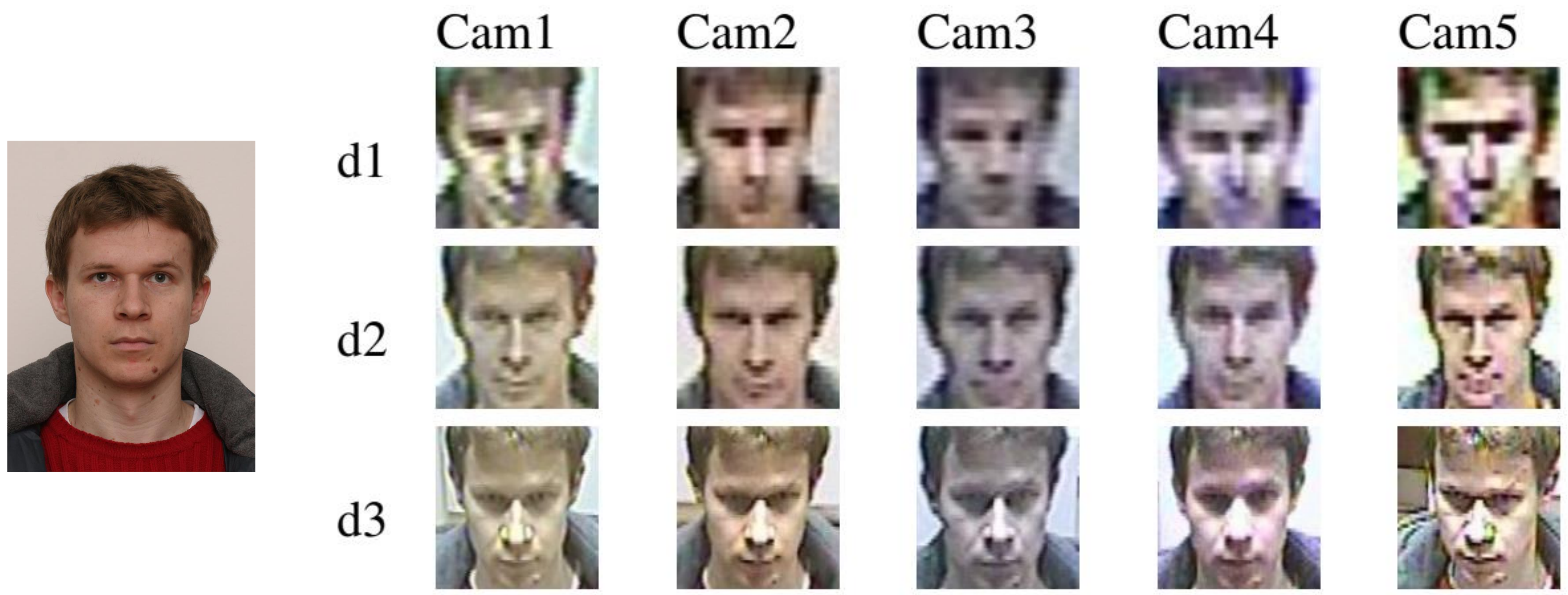
Models trained on down-sampled synthetic data



EXPERIMENTAL RESULTS

LR-to-HR matching

SCface Database



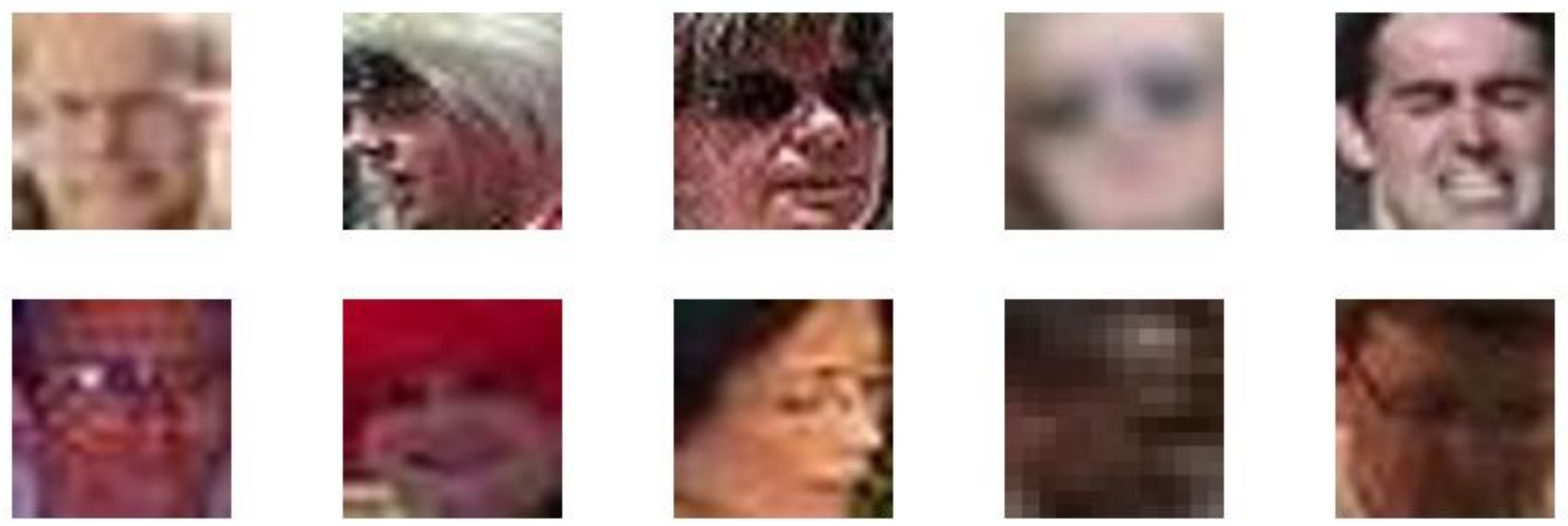
Recognition Rates at Rank-1 on SCface

Method	d1 (4.2m)	d2 (2.6m)	d3 (1.0m)
LightCNN [14]	35.8	79.0	93.8
CenterLoss [35]	36.3	81.8	94.3
VGG-Face [34]	41.3	75.5	88.8
ResNet50-ArcFace [42]	48.0	92.0	99.3
FAN [5]	62.0	90.0	94.8
ShuffleFaceNet	55.5	95.3	99.3
MobileFaceNet	**68.3**	**97.0**	**99.8**
VGG-Face-FT [4]	46.3	78.5	91.5
LightCNN-FT [4]	49.0	83.8	93.5
CenterLoss-FT [4]	54.8	86.3	95.8
ResNet50-ArcFace-FT [5]	67.3	93.5	98.0
DCR-FT [4]	73.3	93.5	98.0
TCN-ResNet-FT [6]	74.6	94.9	98.6
FAN-FT [5]	77.5	95.0	98.3
ShuffleFaceNet-FT	86.0	99.5	99.8
MobileFaceNet-FT	**95.3**	**100.0**	**100.0**

EXPERIMENTAL RESULTS

LR-to-LR matching

TinyFace

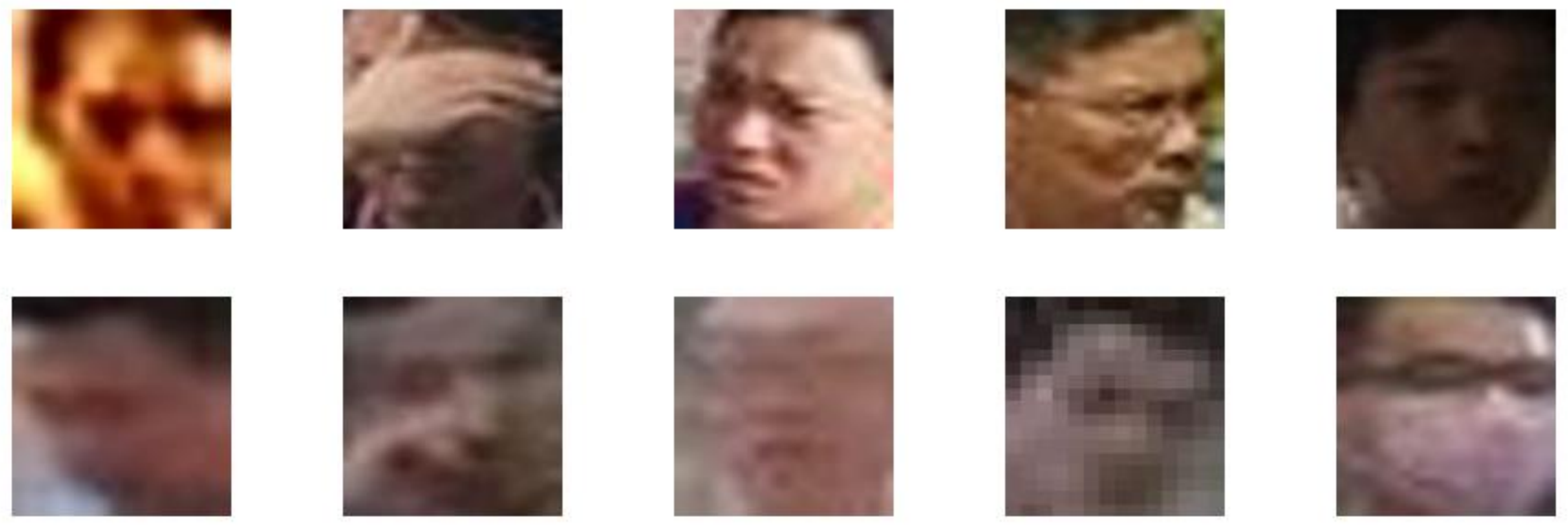


It consists of 169,403 LR face images
(average 20x16 pixels)

Face identification results on TinyFace

Method	Rank-1	Rank-20	Rank-50	mAP
DeepID2 [3]	17.4	25.2	28.3	12.1
SphereFace [3]	22.3	35.5	40.5	16.2
VGG-Face [3]	30.4	40.4	42.7	23.1
CentreFace [3]	32.1	44.5	48.4	24.6
ShuffleFaceNet	43.1	58.9	64.5	34.0
MobileFaceNet	**48.7**	**63.9**	**68.2**	**40.3**

QMUL-SurvFace



It contains 463,507 LR face images
(average 24x20 pixels)

Face verification results on QMUL-SurvFace

Method	TAR@FAR		AUC	Mean Accuracy
	1%	0.1%		
VGG-Face [41]	20.1	4.0	85.0	78.0
DeepID2 [41]	28.2	13.4	84.1	76.1
SphereFace [41]	34.1	15.6	85.0	77.6
FaceNet [41]	40.3	12.7	93.5	85.3
CentreFace [41]	**53.3**	26.8	**94.8**	**88.0**
ShuffleFaceNet	38.5	11.9	89.9	82.3
MobileFaceNet	52.9	**33.1**	89.9	83.2

EXPERIMENTAL RESULTS

Effect of using down-sampled images

Verification Accuracy on different synthetic images from LFW

Method	Bicubic interpolation - LFW			Area interpolation - LFW		
	7 × 7	14 × 14	28 × 28	7 × 7	14 × 14	28 × 28
ShuffleFaceNet	56.7 ± 2.6	82.1 ± 1.7	**98.5 ± 0.5**	50.2 ± 3.2	56.5 ± 1.1	96.4 ± 0.7
ShuffleFaceNet-FT (bicubic)	**68.8 ± 1.9**	**85.7 ± 1.4**	98.2 ± 0.7	50.9 ± 1.9	57.7 ± 2.6	95.9 ± 0.9
ShuffleFaceNet-FT (area)	56.5 ± 2.3	78.6 ± 2.0	97.5 ± 0.7	**77.4 ± 2.1**	**92.4 ± 0.6**	**97.8 ± 0.8**
ShuffleFaceNet-FT (scface)	59.6 ± 1.6	79.1 ± 1.3	95.7 ± 0.9	50.8 ± 1.8	59.8 ± 1.7	93.0 ± 1.1
MobileFaceNet	62.1 ± 2.3	85.3 ± 0.9	98.5 ± 0.5	51.6 ± 1.8	57.4 ± 2.4	96.0 ± 0.7
MobileFaceNet-FT (bicubic)	**66.1 ± 1.4**	**88.8 ± 1.6**	98.3 ± 0.5	49.8 ± 0.8	71.9 ± 1.1	95.3 ± 0.9
MobileFaceNet-FT (area)	59.3 ± 2.6	86.0 ± 1.4	**98.9 ± 0.4**	**74.2 ± 1.9**	**94.6 ± 0.8**	**99.1 ± 0.5**
MobileFaceNet-FT (scface)	62.5 ± 2.2	83.9 ± 1.1	96.2 ± 0.5	55.4 ± 1.6	77.4 ± 1.9	94.7 ± 0.6

Combining down-sampling methods

Method	d1 (4.2m)	d2 (2.6m)	d3 (1.0m)
ShuffleFaceNet-FT (bicubic) + ShuffleFaceNet-FT (area)	58.3	94.3	96.8
MobileFaceNet-FT (bicubic) + MobileFaceNet-FT (area)	71.5	98.0	99.8
ShuffleFaceNet-FT (bicubic+area)	63.8	96.0	98.5
MobileFaceNet-FT (bicubic+area)	75.3	98.0	99.5
MobileFaceNet-FT (area) + ShuffleFaceNet-FT (area)	75.5	98.8	99.5
MobileFaceNet-FT (bicubic+area) + ShuffleFaceNet-FT (bicubic+area)	77.8	99.3	99.5
ShuffleFaceNet-FT (bicubic+area) + ShuffleFaceNet-FT (SCface)	89.8	99.8	99.5
MobileFaceNet-FT (bicubic+area) + MobileFaceNet-FT (SCface)	**96.0**	**100.0**	**100.0**

CENATAV
Tecnológico de Monterrey

EXPERIMENTAL RESULTS

ICPR 2020

Computational Complexity

Computational complexity of different face models

Network	Feat. Dim.	Model Size	Num. Lay.	#Par. (M)	FLOPs (G)
DeepID2	4,500	127	7	10	1.0
VGG-Face	4,096	526	16	138	15.0
ResNet50-ArcFace	512	174	54	43.6	1.6
CenterLoss	512	105	7	27.5	4.2
LightCNN	256	125	28	12.6	3.9
FaceNet	128	95	17	7.5	1.6
ShuffleFaceNet	128	10	58	2.6	0.6
MobileFaceNet	128	8	103	2.0	0.9

- The evaluated models have an accuracy comparable to state-of-the-art models, with a lower computational complexity.
- They require less than 11MB memory and 1GFLOPs, leading with an efficient system deployment.

CONCLUSION

- We presented **a comprehensive evaluation** of lightweight face models for face recognition in LR surveillance imagery.
- **ShuffleFaceNet and MobileFaceNet** were evaluated on three challenging databases, covering the two possible settings: LR-to-HR and LR-to-LR matching.
- Experimental results show that lightweight face models are able to obtain **an accuracy as good as state-of-the-art methods** based on complex deep learning models.
- We show that the low memory footprint and computational complexity of lightweight face models make them **very suitable for practical surveillance applications**.
- We observed that **combining models trained with different degradations** improves the recognition accuracy on low-resolution surveillance imagery, which is feasible due to their low computational cost.