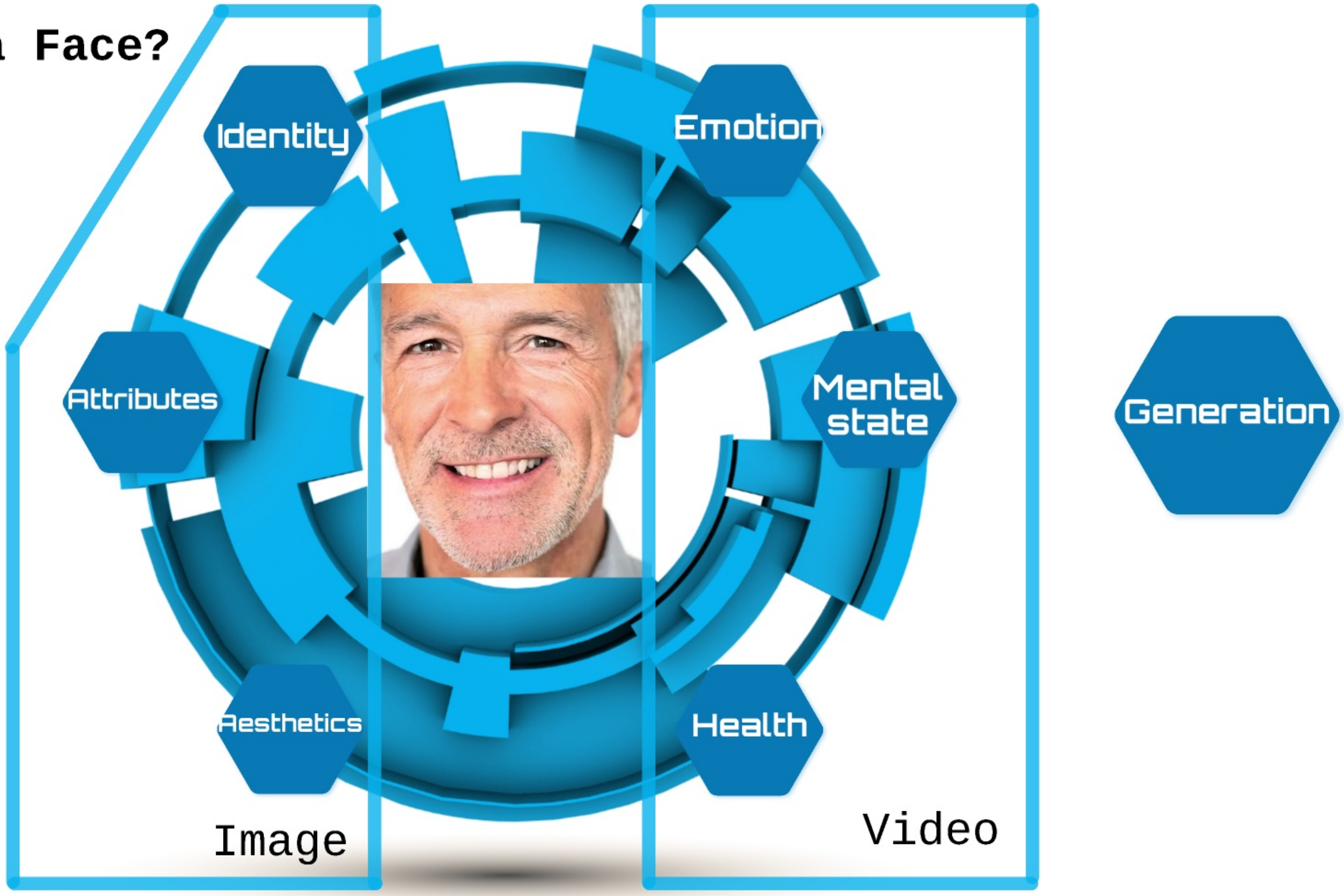
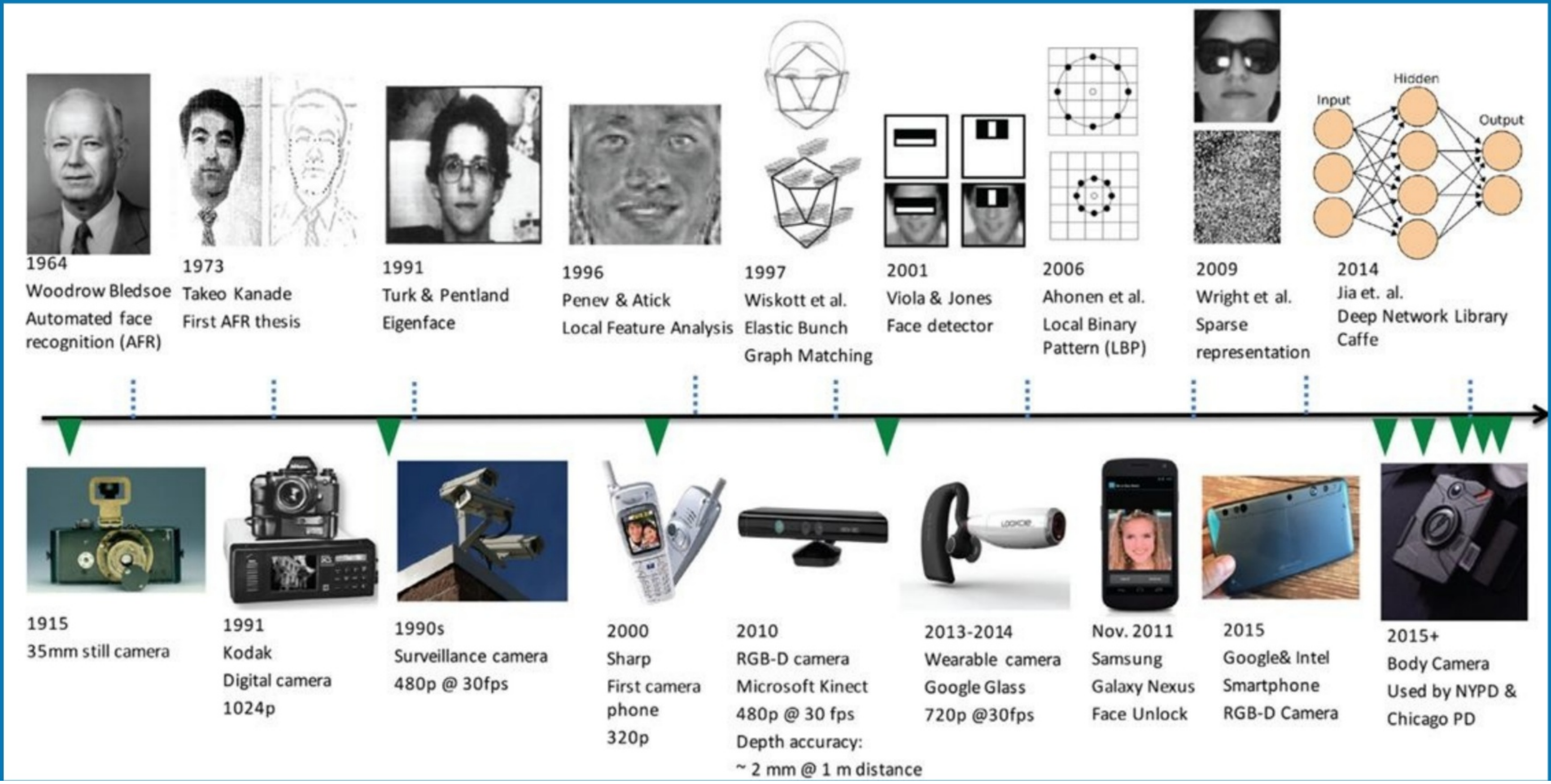
What's in a Face?

Machine Learning and Computer Vision: decipher appearance and dynamics of faces

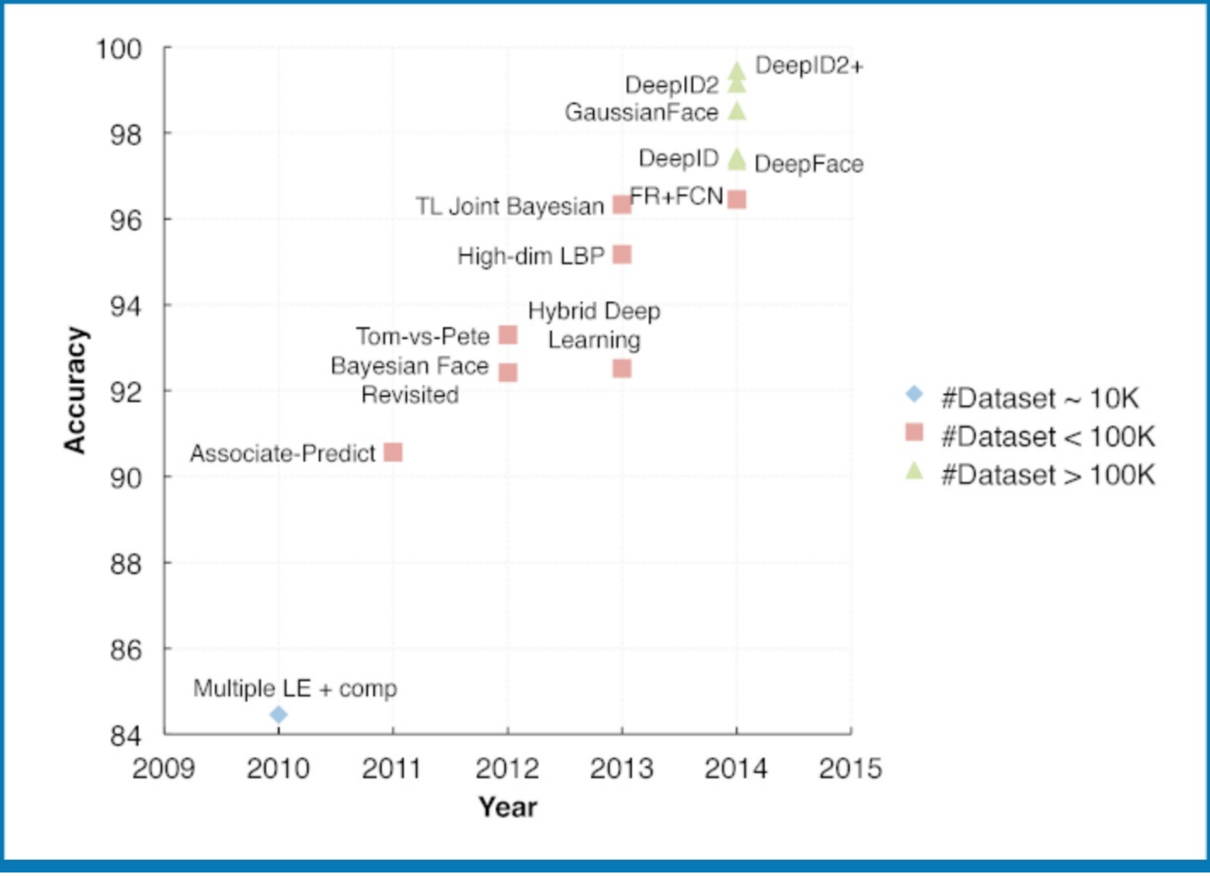


Antitza Dantcheva - Inria, STARS Team

Face Recognition: Timeline

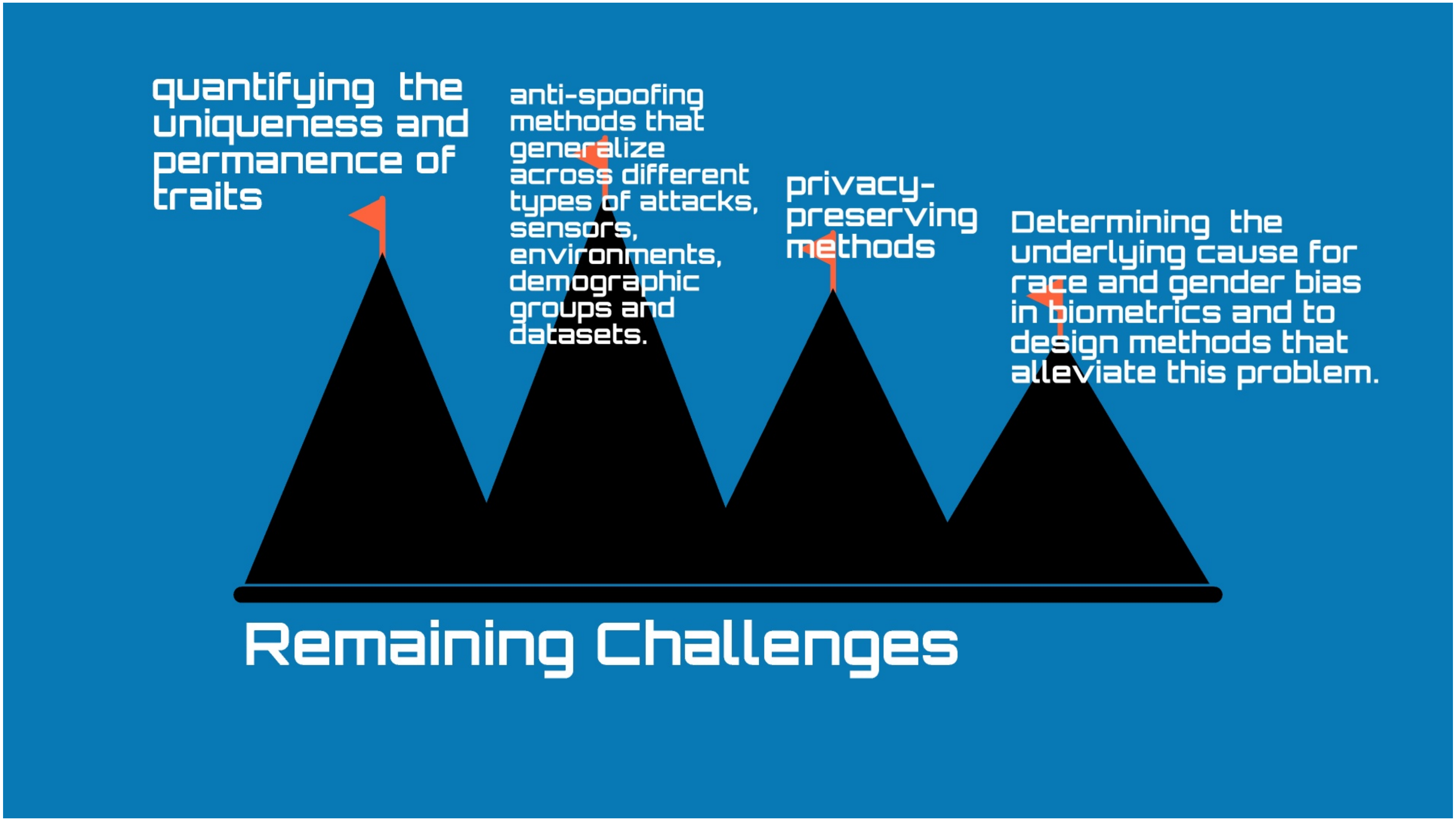


Accuracy: timeline



Accuracy of diff. algorithms on LFW dataset

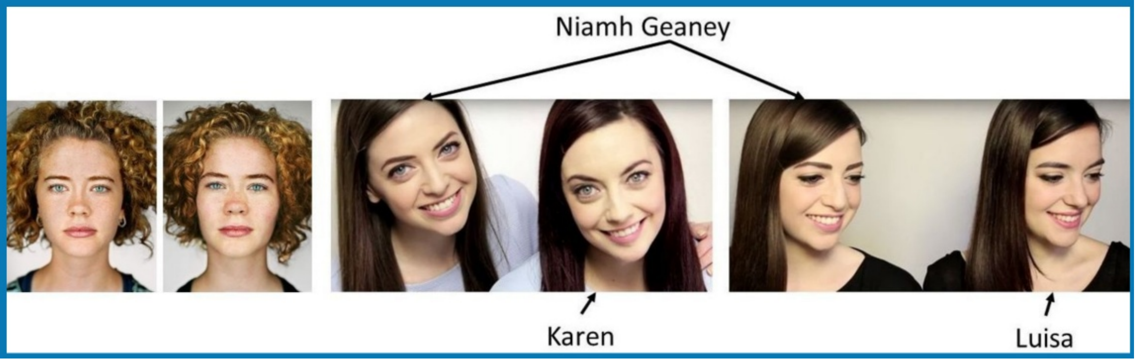
Method	Public. Time	Loss	Architecture	Number of Networks	Training Set	Accuracy±Std(%)
DeepFace [195]	2014	softmax	Alexnet	3	Facebook (4.4M,4K)	97.35±0.25
DeepID2 [187]	2014	contrastive loss	Alexnet	25	CelebFaces+ (0.2M,10K)	99.15±0.13
DeepID3 [188]	2015	contrastive loss	VGGNet-10	50	CelebFaces+ (0.2M,10K)	99.53±0.10
FaceNet [176]	2015	triplet loss	GoogleNet-24	1	Google (500M,10M)	99.63±0.09
Baidu [124]	2015	triplet loss	CNN-9	10	Baidu (1.2M,18K)	99.77
VGGface [149]	2015	triplet loss	VGGNet-16	1	VGGface (2.6M,2.6K)	98.95
light-CNN [225]	2015	softmax	light CNN	1	MS-Celeb-1M (8.4M,100K)	98.8
Center Loss [218]	2016	center loss	Lenet+-7	1	CASIA-WebFace, CACD2000, Celebrity+ (0.7M,17K)	99.28
L-softmax [126]	2016	L-softmax	VGGNet-18	1	CASIA-WebFace (0.49M,10K)	98.71
Range Loss [261]	2016	range loss	VGGNet-16	1	MS-Celeb-1M, CASIA-WebFace (5M,100K)	99.52
L2-softmax [157]	2017	L2-softmax	ResNet-101	1	MS-Celeb-1M (3.7M,58K)	99.78
Normface [206]	2017	contrastive loss	ResNet-28	1	CASIA-WebFace (0.49M,10K)	99.19
CoCo loss [130]	2017	CoCo loss	-	1	MS-Celeb-1M (3M,80K)	99.86
vMF loss [75]	2017	vMF loss	ResNet-27	1	MS-Celeb-1M (4.6M,60K)	99.58
Marginal Loss [43]	2017	marginal loss	ResNet-27	1	MS-Celeb-1M (4M,80K)	99.48
SphereFace [125]	2017	A-softmax	ResNet-64	1	CASIA-WebFace (0.49M,10K)	99.42
CCL [155]	2018	center invariant loss	ResNet-27	1	CASIA-WebFace (0.49M,10K)	99.12
AMS loss [205]	2018	AMS loss	ResNet-20	1	CASIA-WebFace (0.49M,10K)	99.12
Cosface [207]	2018	cosface	ResNet-64	1	CASIA-WebFace (0.49M,10K)	99.33
Arcface [42]	2018	arcface	ResNet-100	1	MS-Celeb-1M (3.8M,85K)	99.83
Ring loss [272]	2018	Ring loss	ResNet-64	1	MS-Celeb-1M (3.5M,31K)	99.50


quantifying the uniqueness and permanence of traits
anti-spoofing methods that generalize across different types of attacks, sensors, environments, demographic groups and datasets.
privacy-preserving methods
Determining the underlying cause for race and gender bias in biometrics and to design methods that alleviate this problem.
Remaining Challenges

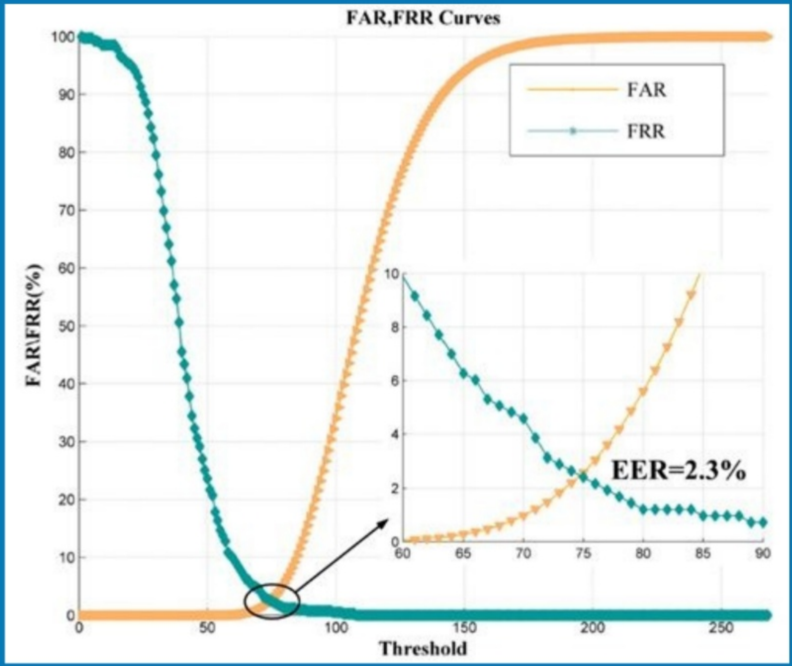
Facial Uniqueness

Facial entropy is a function of
(a) image resolution
(b) feature representation
(c) (unknown) feature distribution
(d) limited sample size
(e) database (f) pose (g) noise
(h) blur as well as (i) expression

[1] Balazia, Happy, Bremond, Dantcheva
An investigative Study on Facial Uniqueness
Under submission



Uniqueness measure



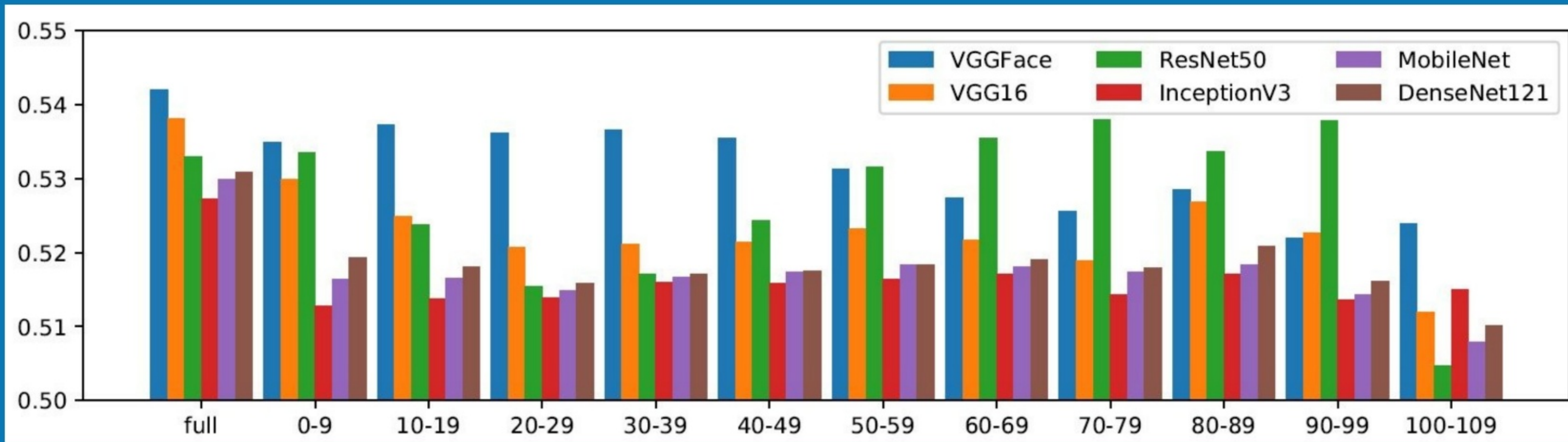
Features / Datasets

	AT&T	LFW	IMDb	TWINS
VGGFace	0.6710	0.5637	0.5420	0.5591
VGG16	0.6417	0.5364	0.5381	0.5339
ResNet50	0.6650	0.5364	0.5330	0.5574
InceptionV3	0.5915	0.5293	0.5272	0.5242
MobileNet	0.6204	0.5353	0.5299	0.5301
DenseNet121	0.6338	0.5302	0.5309	0.5268

Resolution

	AT&T	LFW	IMDb	TWINS
224×224	0.6710	0.5637	0.5420	0.5591
112×112	0.6725	0.5636	0.5418	0.5590
64×64	0.6620	0.5635	0.5405	0.5581
48×48	0.6485	0.5637	0.5385	0.5566
36×36	0.6241	0.5603	0.5345	0.5522

Age and gender



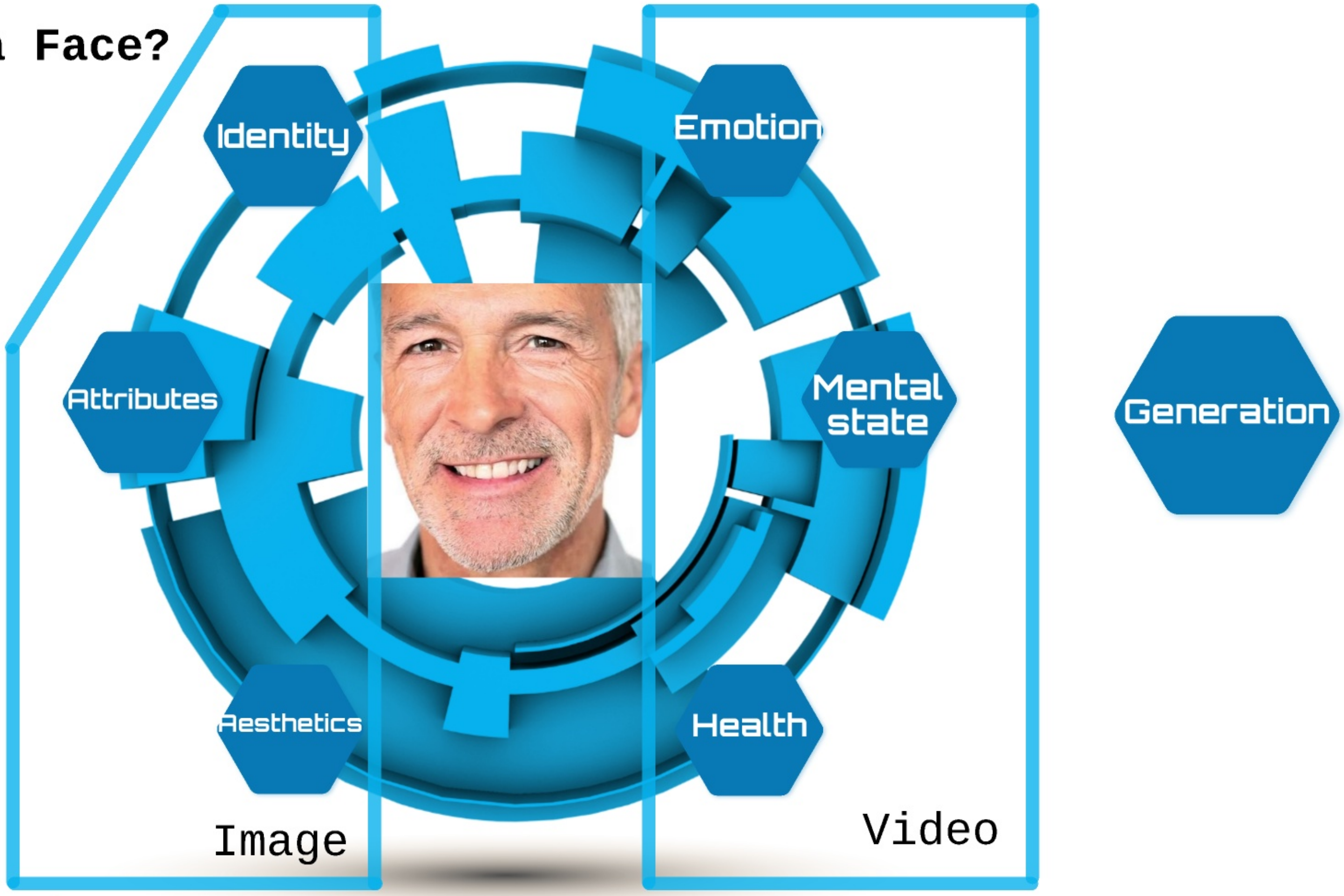
	full	female	male
VGGFace	0.542	0.535	0.532
VGG16	0.538	0.520	0.521
ResNet50	0.533	0.519	0.525
InceptionV3	0.527	0.511	0.514
MobileNet	0.530	0.514	0.516
DenseNet121	0.531	0.515	0.517

Dataset#2

	U based on $\bar{\mathrm{D}}$		U based on $\bar{\mathrm{D}}^{\mathrm{MIN}}$	
	AT&T	TWINS	AT&T	TWINS
VGGFace	0.671	0.559	0.632	0.507
VGG16	0.642	0.534	0.582	0.491
ResNet50	0.665	0.557	0.542	0.449
InceptionV3	0.592	0.524	0.542	0.485
MobileNet	0.620	0.530	0.570	0.498
DenseNet121	0.634	0.527	0.577	0.494

What's in a Face?

Machine Learning and Computer Vision: decipher appearance and dynamics of faces



Antitza Dantcheva - Inria, STARS Team

Attributes

	Age	Gender	Ethnicity	Height	Weight	Hair	Clothing	Proportions
Face	√	√	√	√	√	√	√	√
Iris	√	√	√	?	?	√		?
Fingerprint	√	√	√	?	√	?		√
Gait	√	√	√	√	√	√	√	√
Ear	√	√	√	?	?	√		?
Voice	√	√	√	√	√	?		?
Tattoos	√	√	√	√	√	√	√	√
Body	√	√	√	√	√	√	√	√

Bias in face recognition

An algorithm is considered biased, if statistically different outcomes (decisions) are produced by it for different demographic groups.

Face recognition algorithms have been labeled as "racist" or "biased".

[2] P. Drozdowski, C. Rathgeb, A. Dantcheva, N. Damer, C. Busch, "**Demographic Bias in Biometrics: An Emerging Challenge**" to be submitted

Bias in face recognition

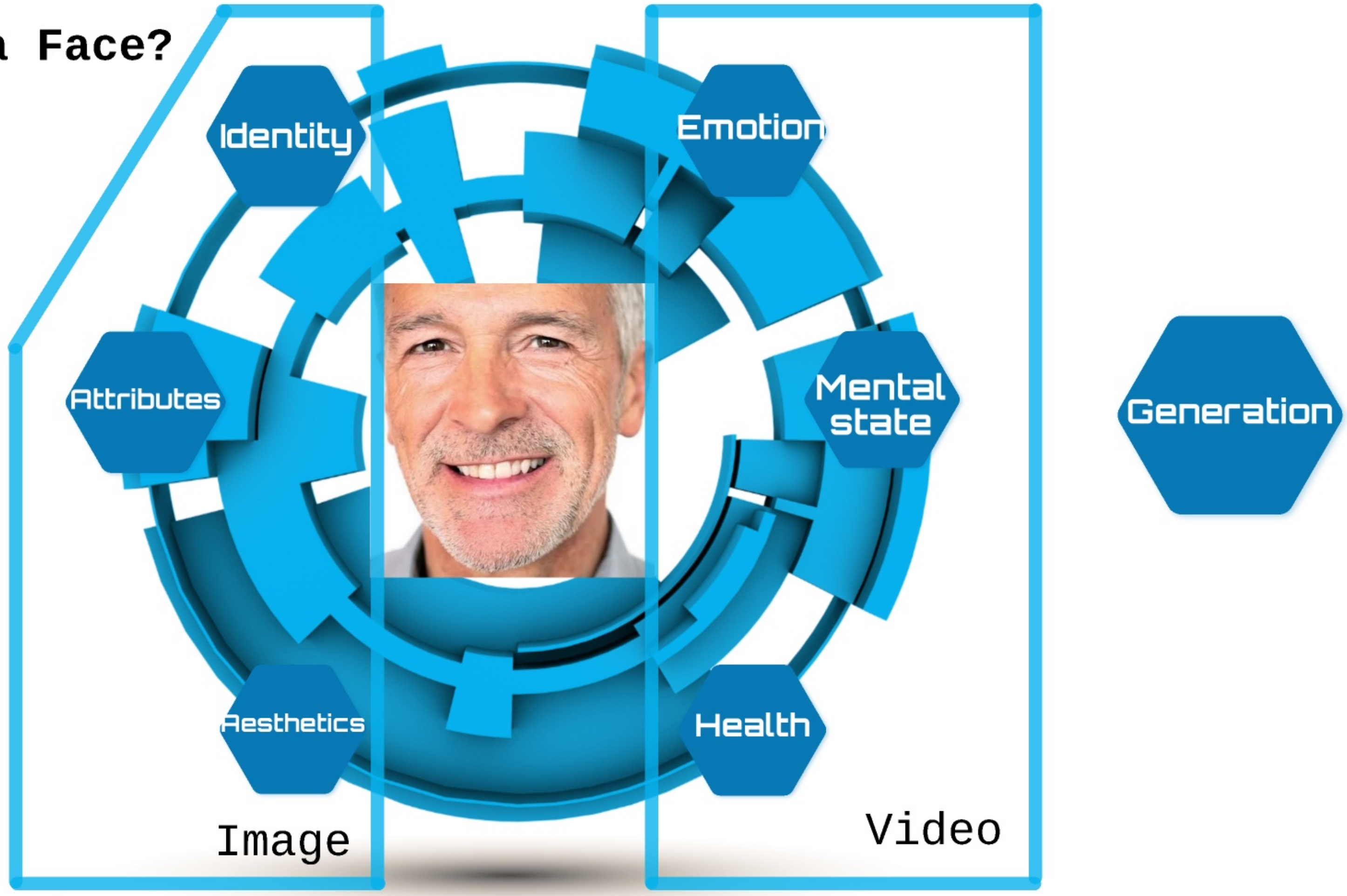
Model	LFW	RFW			
		Caucasian	Indian	Asian	African
Microsoft	98.22	87.60	82.83	79.67	75.83
Face++	97.03	93.90	88.55	92.47	87.50
Baidu	98.67	89.13	86.53	90.27	77.97
Amazon	98.50	90.45	87.20	84.87	86.27
mean	98.11	90.27	86.28	86.82	81.89
Center-loss [218]	98.75	87.18	81.92	79.32	78.00
Sphereface [125]	99.27	90.80	87.02	82.95	82.28
Arcface [42]	99.40	92.15	88.00	83.98	84.93
VGGface2 [22]	99.30	89.90	86.13	84.93	83.38
mean	99.18	90.01	85.77	82.80	82.15

Possible reasons for bias

- Training data bias
- Algorithmic processing bias
- Facial expressions
- Zero pitch angle
- Hair occlusion
- Makeup

What's in a Face?

Machine Learning and Computer Vision: decipher appearance and dynamics of faces



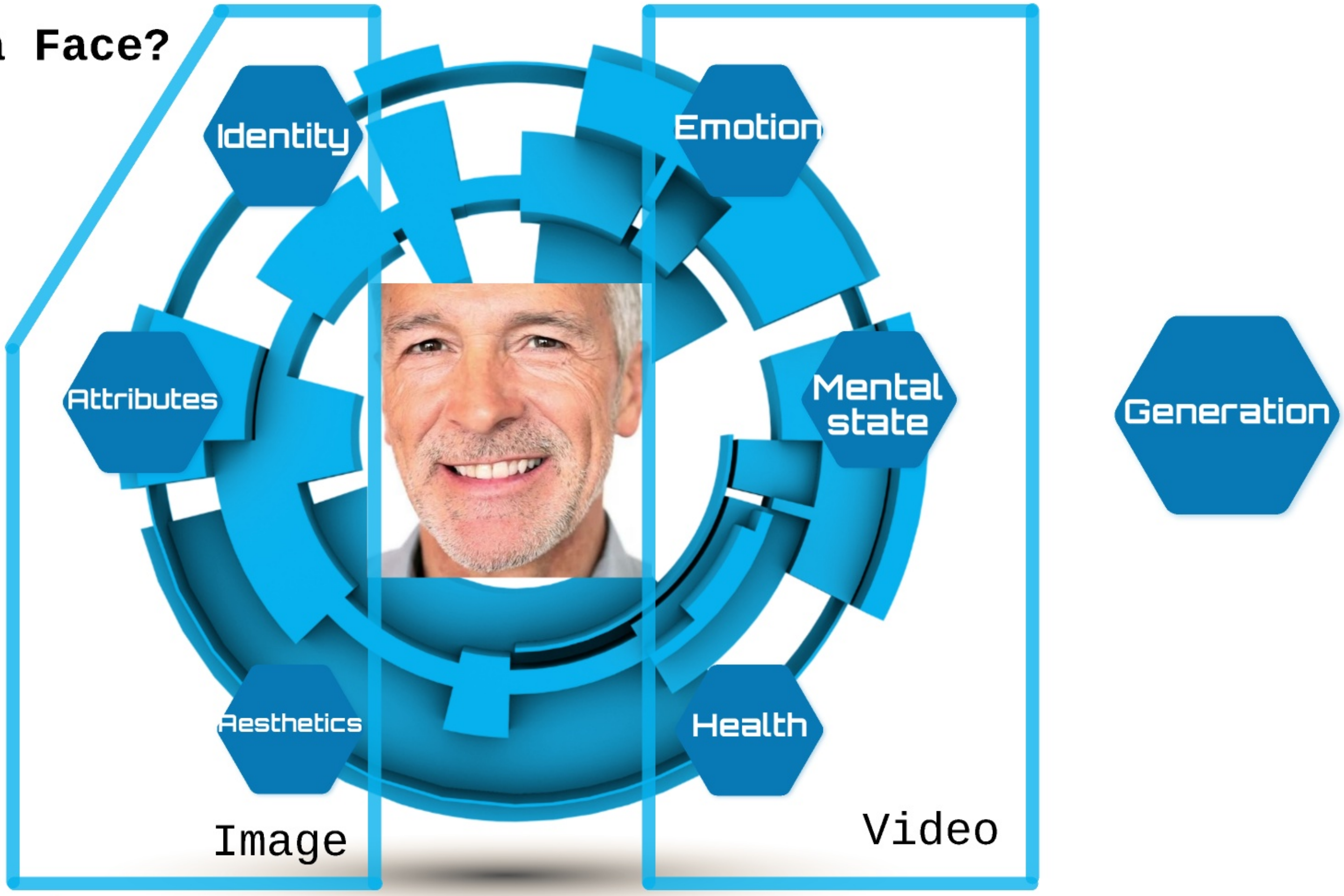
Antitza Dantcheva - Inria, STARS Team

Aesthetics

[3] Rathgeb, Christian; Dantcheva, Antitza; Busch, Christoph
Impact and detection of facial beautification in face recognition: An overview
IEEE Access, vol. 7, no. 1, December, 2019.

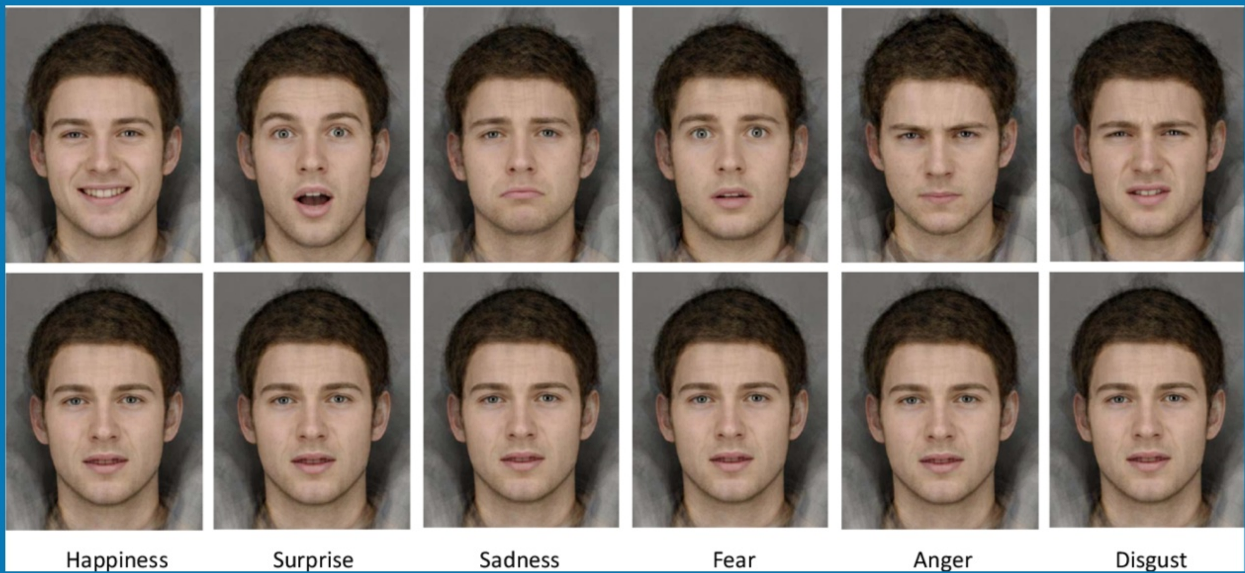
What's in a Face?

Machine Learning and Computer Vision: decipher appearance and dynamics of faces



Antitza Dantcheva - Inria, STARS Team

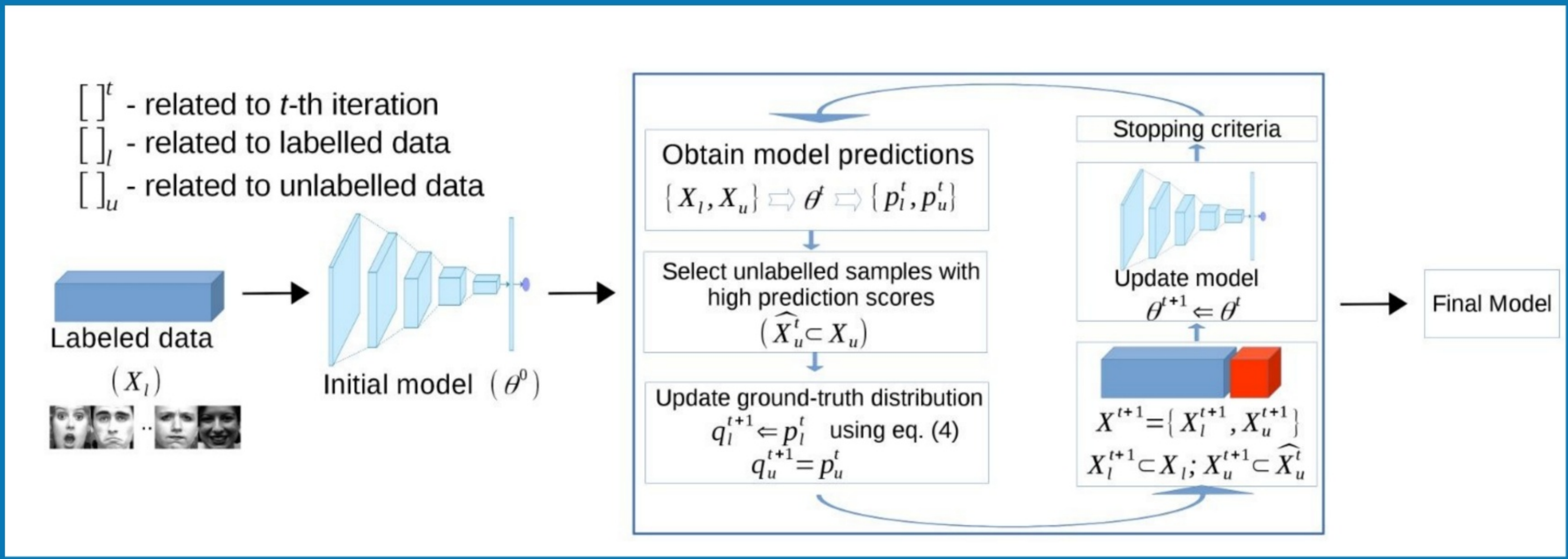
Emotion Recognition



State of the Art

Author Name	Technique Used	Database Used	Emotion Recognition Accuracy	Emotions Considered	Drawbacks
Mutsugu [10]	Convolution neural network (CNN) Curveletface with LDA Curveletface with PCA	Still images own database (5600 images)	97.6%—CNN 83.5%—Curveletface + LDA 83.9%—Curveletface + PCA	Happy, Neutral and Talking	System insensitive to individuality of facial expressions mainly by virtue of the rule-based facial analysis
Zhang [42]	Patched based 3D Gabor filters SVM and Adaboost	JAFFE C-K	92.93%—JAFFE 94.48%—C-K	Happy, Neutral, Sadness, Surprise and Anger, Fear and Disgust	JAFFE DB requires larger sizes of patches than the CK DB to keep useful information
Hayat [43]	SVM with clustering algorithms	BU 4DFE	94.34%	Anger, Disgust, Happiness, Fear, Sadness, and Surprise	——
Hablani [44]	Local binary patterns	JAFFE	Person dependent—94.44% Person independent—73.6%	Happy, Neutral, Sadness, Surprise and Anger, Fear and Disgust	Manual detection of face and its components, experimented only on JAFFE DB
Zisheng [45]	PHOG (Pyramid Histogram of Oriented Histogram) descriptors	C-K	96.33 %	Happy, Neutral, Sadness, Surprise and Anger, Fear and Disgust	——
Lee [46]	Sparse Representation	JAFFE BU 3DFE	Person dependent - 94.70%—JAFFE Person independent - 90.47%—JAFFE 87.85%—BU 3DFE	Happiness, Disgust, Angry, Surprise, and Sadness	The face images used in the experiment were cropped manually
Zheng [47]	Group sparse reduced-rank regression (GSRRR) + ALM	BU 3DFE	66.0%	Happiness, Fear, Angry, Surprise and Sadness	Implementation of new method with less accuracy
Yu [37]	Deep CNN, 7 hidden layers with minimization of hinge loss	SFEW	61.29%	Happiness, Disgust, Fear, Angry, Surprise and Sadness	Less accuracy through more networks

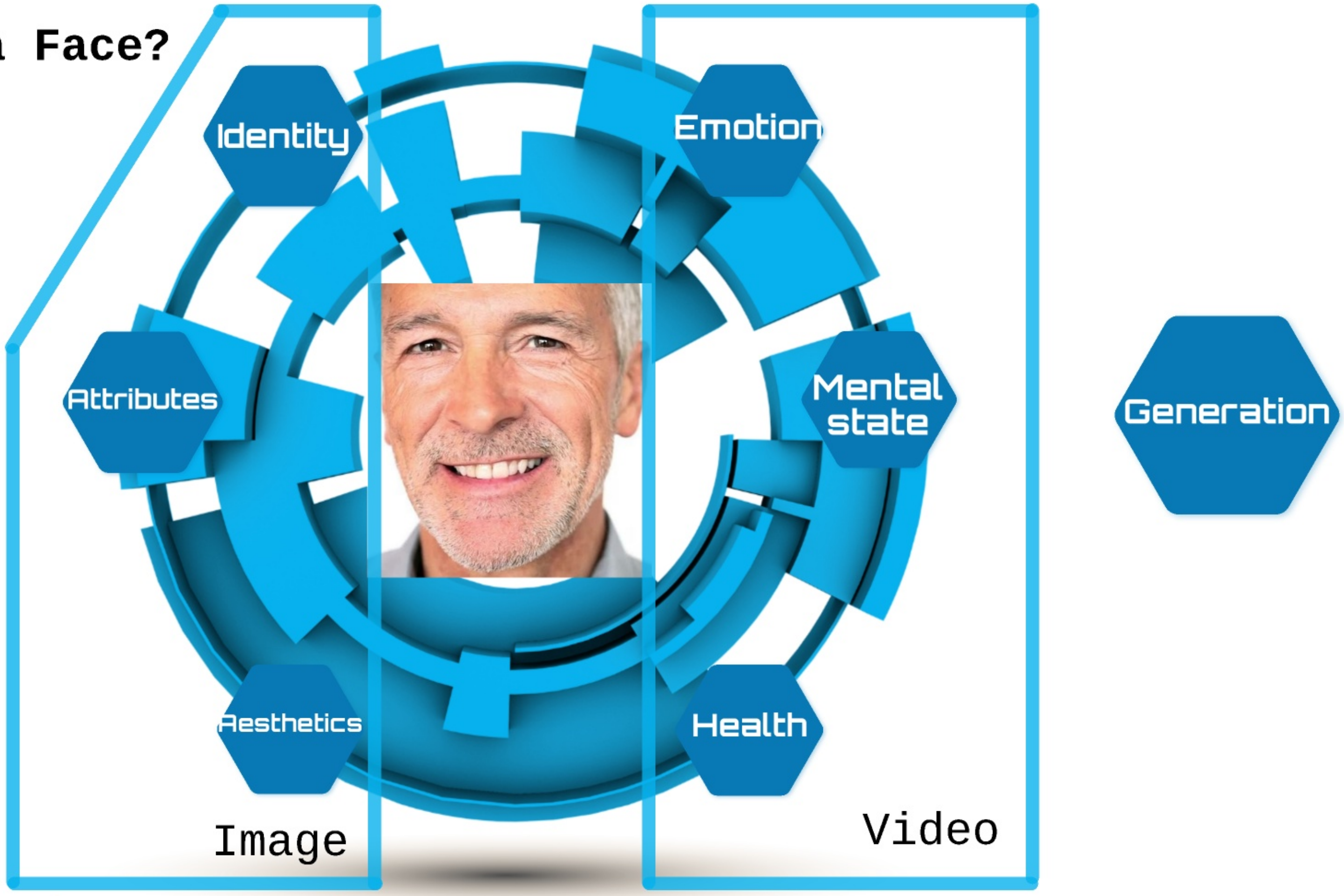
Limited Training Data



[4] Happy, SL; Dantcheva, Antitza; Bremond, Francois
A weakly supervised learning technique for classifying facial expressions
Pattern Reognition Letters, 2019.

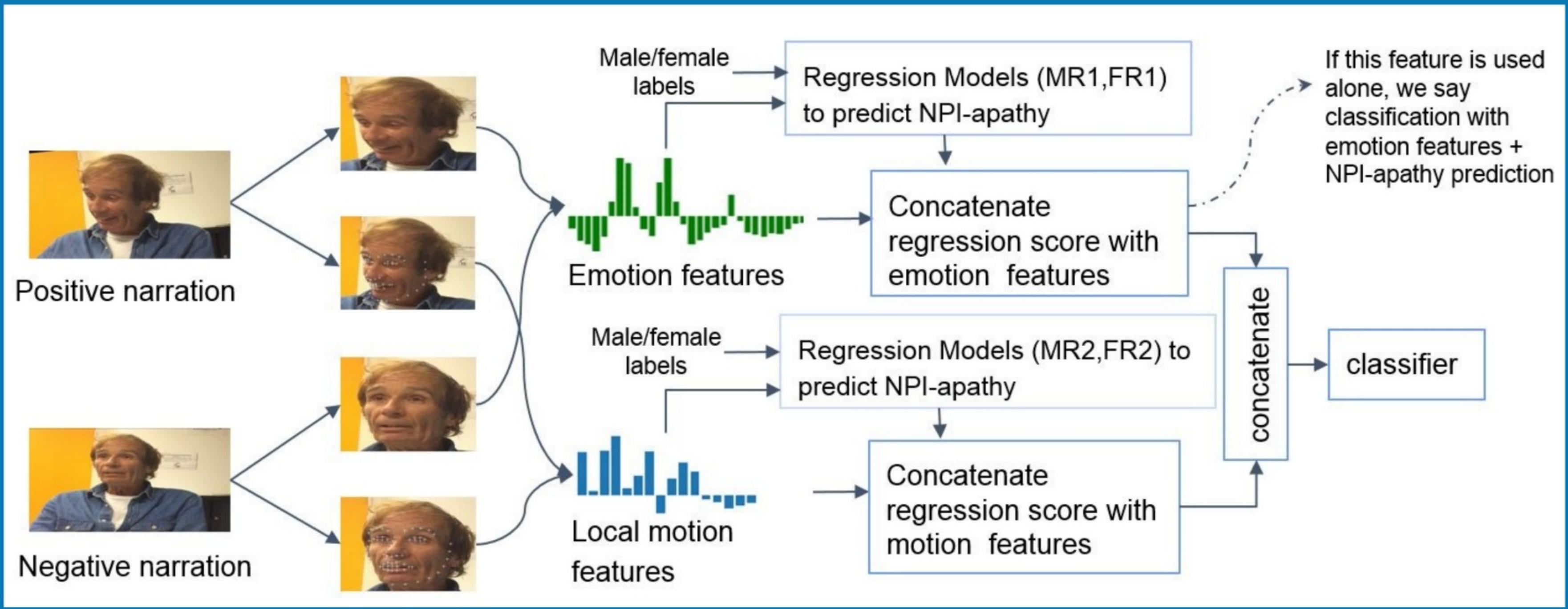
What's in a Face?

Machine Learning and Computer Vision: decipher appearance and dynamics of faces



Antitza Dantcheva - Inria, STARS Team

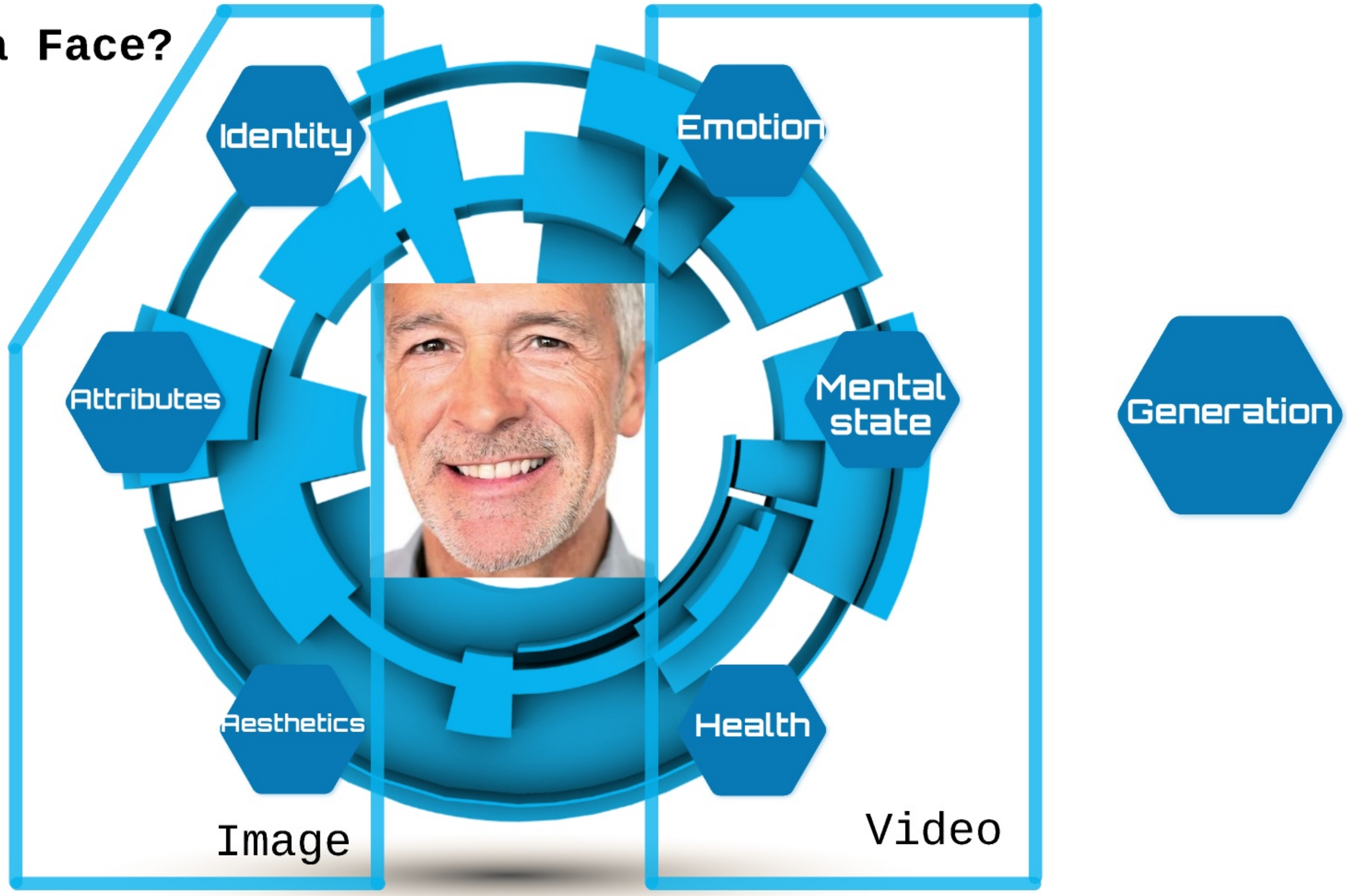
Apathy classification



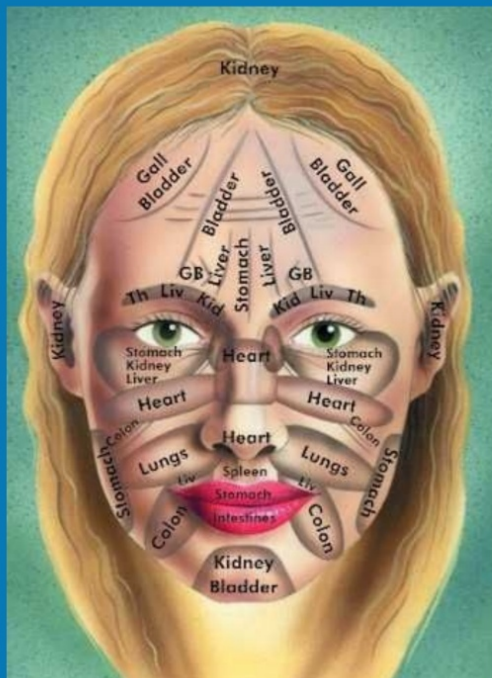
[5] Happy, Dantcheva, Das, Zeghari, Robert, Bremond, “Characterizing the state of apathy with facial expression and motion analysis” FG, 2019

What's in a Face?

Machine Learning and Computer Vision: decipher appearance and dynamics of faces



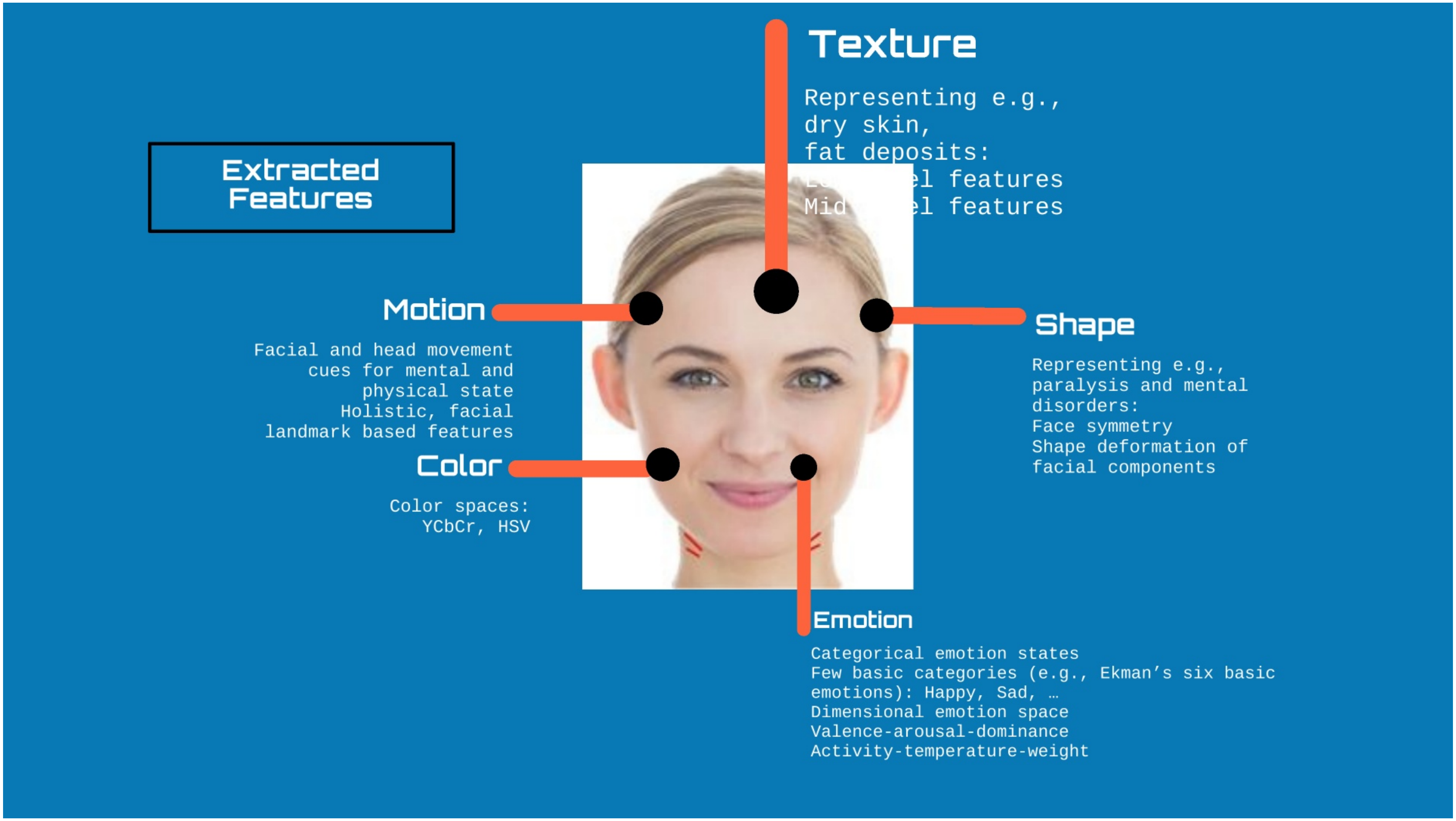
Antitza Dantcheva - Inria, STARS Team



Facial appearance depends on: Skull morphology, muscles, innervation of blood flow, fat deposit, etc.

Health clues:

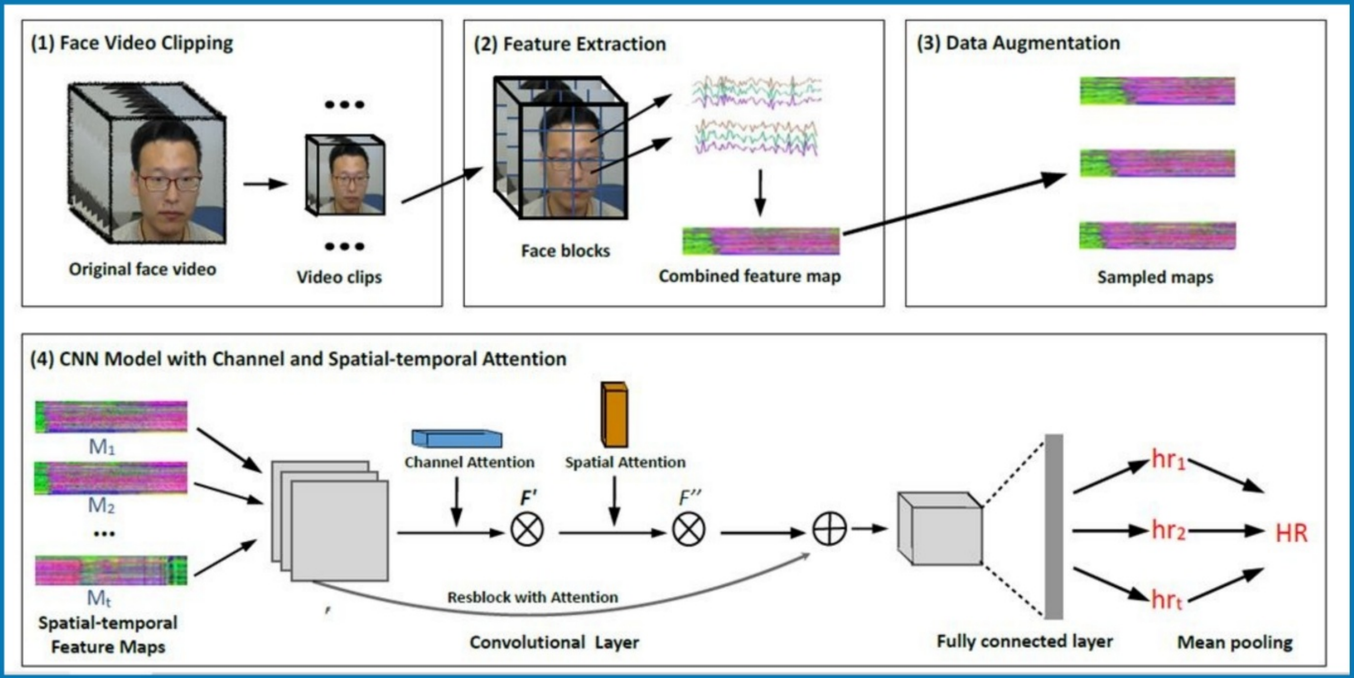
- dry skin
- eye bags
- facial asymmetry
- paleness
- facial behavior
- Gestures
- Body appearance


Extracted Features
Texture
Representing e.g.,
dry skin,
fat deposits:
el features
Mid el features
Motion
Facial and head movement cues for mental and physical state
Holistic, facial landmark based features
Shape
Representing e.g., paralysis and mental disorders:
Face symmetry
Shape deformation of facial components
Color
Color spaces:
YCbCr, HSV
Emotion
Categorical emotion states
Few basic categories (e.g., Ekman's six basic emotions): Happy, Sad, …
Dimensional emotion space
Valence-arousal-dominance
Activity-temperature-weight

State of the Art

Conditions	Observation cues	Imaging Modalities	Datasets	Analysis Methods
Psychological disorder	Facial expression and behavior; eye tracking; gesture	image, video, depth-images, thermal images, others (EEG, MRI)	KOOMA [Jaiswal *et al.*, 2017];	MONOVA, SVM [Jian, *et al.*, 2017]; expectation maximization [Galgani *et al.*, 2009]; CNN [Jaiswal *et al.*, 2017]
Depression (bipolar disorder)	facial expression disturbances (AU), face pose, head movement	images and video-s	Depression analysis corpus [Ringeval and et al., 2017]; Distress analysis interview corpus (DAIC) [Gratch and et al.,]; Bipolar Disorder Corpus [Ciftci, 2018]	geometry and appearance features, motion, support vector machines (SVM) [Pampouchidou A., 2015]; support vector regressor [Williamson *et al.*, 2016]; random forest regression[Ringeval and et al., 2017]; AU detection, bag of words [Ringeval and et al., 2017]; pose and head movement [Dibeklioğlu *et al.*, 2018]
Neurological disorder	facial motion; gesture; gait and finger tapping patterns	image, video, depth images; others (EEG, ECG, MRI, speech, accelerometer) [Ramgopal and *et al.*, 2014]	Pediaditis et al.[Pediaditis *et al.*, 2012]	facial motion [Tsiknakis *et al.*, 2012]; CNN[Pereira *et al.*, 2018]
Genetic and developmental disorder	facial geometry and texture	images and videos	Hammond's database[Hammond, *et al.*, 2005]; Loos et al.[Loos, *et al.*, 2003]; Shukla et al.[Shukla, *et al.*, 2017];	Gabor filter[Loos, *et al.*, 2003]; CNN[Shukla, *et al.*, 2017]
Facial paralysis	facial asymmetry	image and video; infrared thermal imaging	UPFP[Guo *et al.*, 2017]; private facial paralysis data[Wang *et al.*, 2016]	geometry features [Wang *et al.*, 2016]; Gabor filters [Ngo *et al.*, 2016]; LBP[Wang *et al.*, 2016]; VAE[Wang, *et al.*, 2018]; CNN[Storey and Jiang, 2018]
Pain	facial muscle movements and expressions	image and video; others (EEG, ECG, GSR)	STOIC[Roy *et al.*, 2007]; COPE[Brahnam *et al.*, 2007b]; Pain Archive Dataset [Lucey *et al.*, 2011]; EmoPain[Aung *et al.*, 2016]	PCA, LDA, SVM, NN - 95% [Brahnam *et al.*, 2007b]; CNN, RCNN, and LSTM [Rodriguez *et al.*, 2017]
Skin temperature and mandibular disorder	change of skin temperature	thermal imagery	Souza et al. [Souza *et al.*, 2016]	LDA, classification tree [Graft, *et al.*, 1994]; short time Fourier transform, information content [Nhan and Chau, 2009]
Other disorders	facial geometry, eye gaze	images; thermal images; others (MRI,speech)	Acromegaly data [Gencturk *et al.*, 2013]; Turner syndrome data [Chen *et al.*, 2018]	face geometry[Chen *et al.*, 2018], Gabor filter [Chen *et al.*, 2018]; LBP [Gencturk *et al.*, 2013]; eye gaze [Liu *et al.*, 2018];

Heartrate



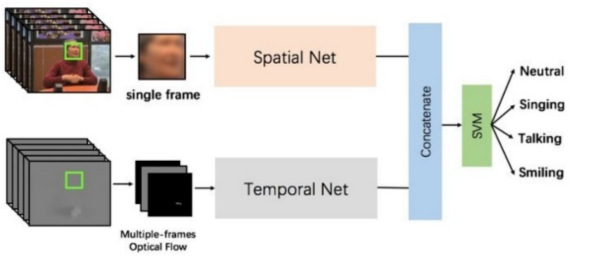
Niu, X.; Zhao, X.; Han, H.; Das, A.; Dantcheva, A.; Shan, S.; Chen, X. Robust remote heart rate estimation from face utilizing spatial-temporal attention, FG'19

Facial Behavior

Wang, Y.; Dantcheva, A.; Broutart, J.-C.; Robert, P.; Bremond, F.; Bilinski, P. **Comparing methods for assessment of facial dynamics in patients with major neurocognitive disorders** ECCVW'18

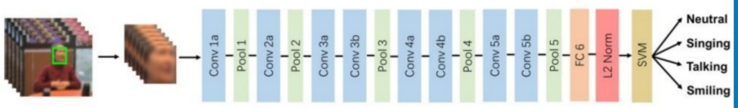
Spatio-temporal analysis methods

2streams Convnets

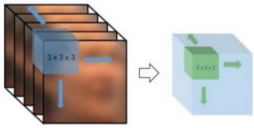


- *Very Deep ConvNets (VGG-16, ResNet152)*
- *RGB features + Optical Flow features*

C3D



- *3D ConvNet (3D Conv kernel + 3D MaxPooling)*



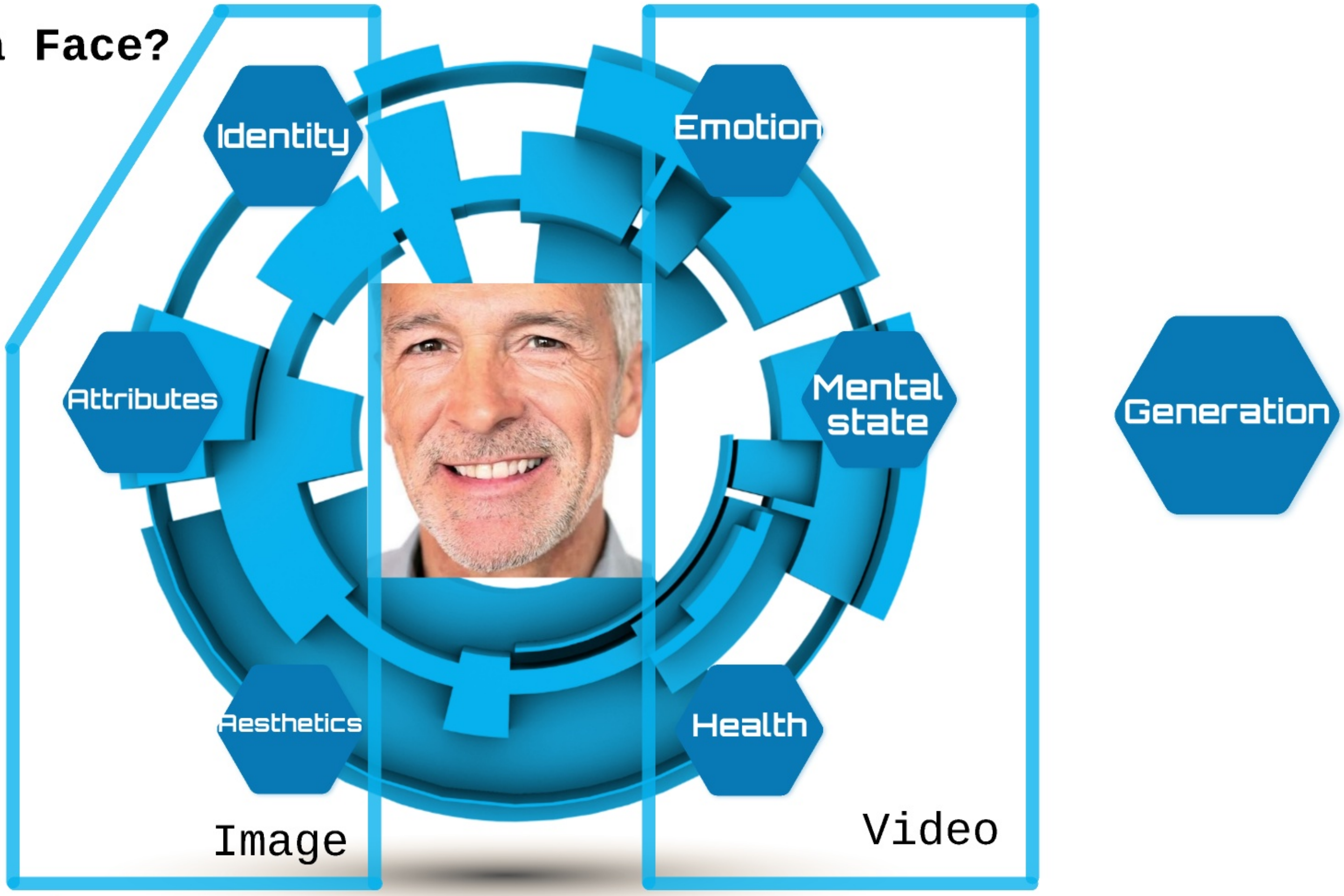
- *Spatial and Temporal features from RGB frames*

Mean accuracy of each method

Method	MA (%)
C3D	67.4
SN of Two-Stream ConvNets (VGG-16)	65.2
TN of Two-Stream ConvNets (VGG-16)	69.9
Two-Stream ConvNets (VGG-16)	76.1
SN of Two-Stream ConvNets (ResNet-152)	69.6
TN of Two-Stream ConvNets (ResNet-152)	75.8
Two-Stream ConvNets (ResNet-152)	76.4
iDT	61.2
C3D + iDT	71.1
Two-Stream ConvNets (VGG-16) + iDT	78.9
Two-Stream ConvNets (ResNet-152) + iDT	**79.5**

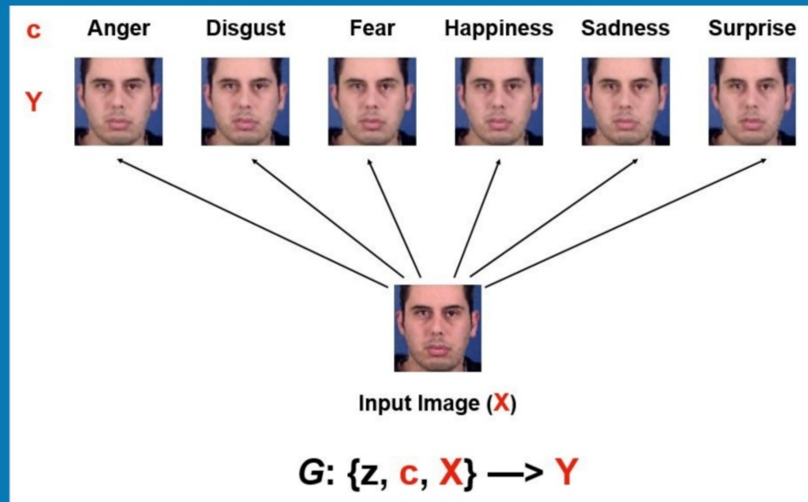
What's in a Face?

Machine Learning and Computer Vision: decipher appearance and dynamics of faces



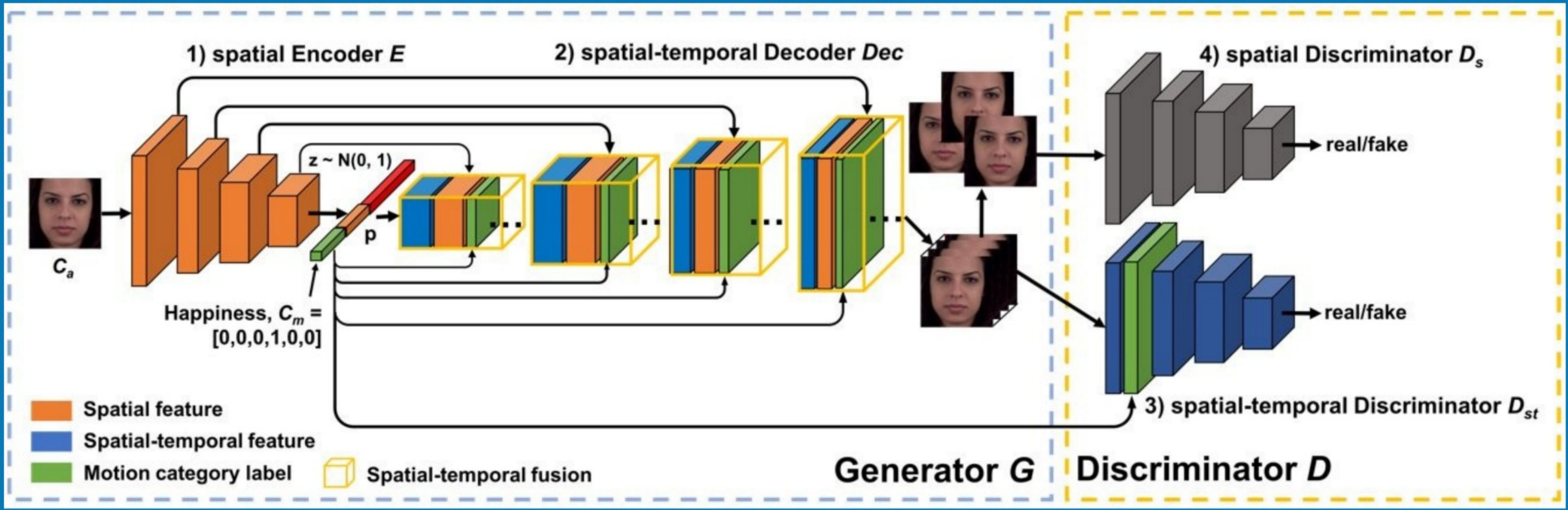
Antitza Dantcheva - Inria, STARS Team

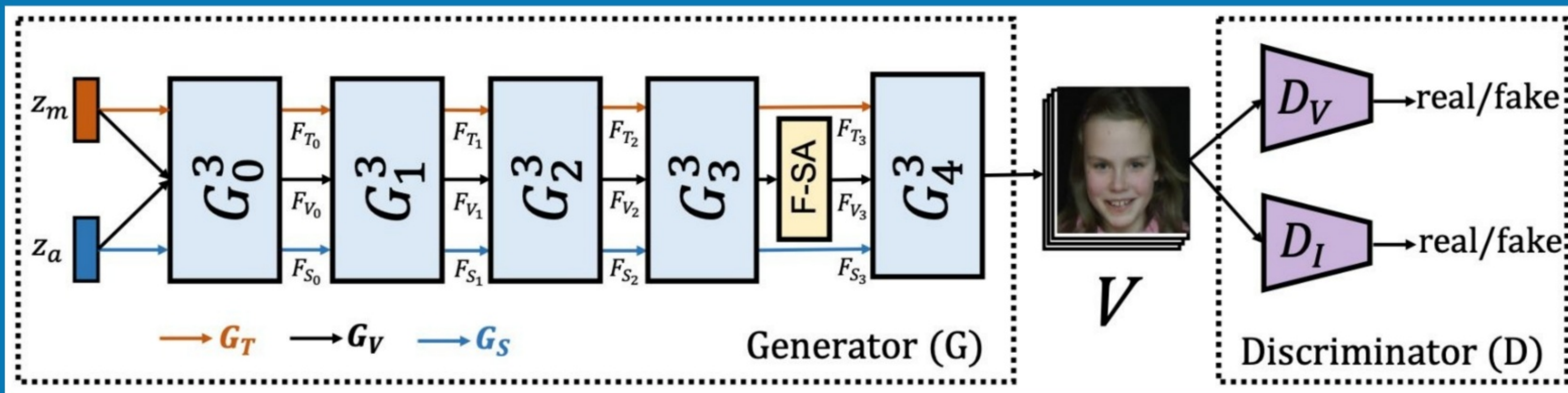
Problem Formulation



[6] Wang, Yaohui; Bilinski, Piotr; Bremond, Francois; Dantcheva, Antitza
ImaGINator: conditional spatio-temporal GAN for video generation
WACV'20, Winter Conference on Applications of Computer Vision, March 2-5, 2020, Aspen, USA.
[7] Wang, Yaohui; Bilinski, Piotr; Bremond, Francois; Dantcheva, Antitza
G^3AN: This video does not exist. Disentangling motion and appearance for video generation
arXiv preprint arXiv:1912.05523, 2019

Proposed architecture





[6] Wang, Yaohui; Bilinski, Piotr; Bremond, Francois; Dantcheva, Antitza
ImaGINator: conditional spatio-temporal GAN for video generation
WACV'20, Winter Conference on Applications of Computer Vision, March 2-5, 2020, Aspen, USA.
[7] Wang, Yaohui; Bilinski, Piotr; Bremond, Francois; Dantcheva, Antitza
G^3AN: This video does not exist. Disentangling motion and appearance for video generation
arXiv preprint arXiv:1912.05523, 2019

Deepfakes in FaceForensics++

Original Videos

4 Manipulation techniques

Manipulation techniques

1) **face-swap** represents a graphic approach transferring a face region from a source video to a target video. Using facial landmarks, a 3D template model employs blendshapes to fit the transferred face. FaceSwap6.
2) **deepfakes** has become the synonym for all face manipulations of all kind, it origins to FakeApp and faceswap.
3) **face2face** is a facial reenactment system that transfers the expressions of a source video to a target video while maintaining the identity of the target person.
4) **neuraltextures** incorporates facial reenactment as an example for a NeuralTextures-based rendering approach. It uses the original video data to learn a neural texture of the target person, including a rendering network.

Deepfake concerns

(a) such manipulations can fabricate animations of subjects involved in actions that have not taken place
(b) such manipulated data can be circumvented nowadays rapidly via social media
(c) deepfake techniques have become open to the public via apps

(i) our society relies heavily on the ability to produce and exchange legitimate and trustworthy documents,
(ii) sound and images have recorded our history, as well as informed and shaped our perception of reality, e.g., axioms and truths such as "I'll believe it when I see it." "Out of sight, out of mind." "A picture is worth a thousand words.",
(iii) social media has catapulted online videos as a mainstream source of information;

Threats:

- deepfakes being perceived as real
- real videos being misdetected for fake

Detection Results each manipulation technique individually

Algorithm	DF	F2F	FS	NT
Steg. Features + SVM [13]	73.64	73.72	68.93	63.33
Cozzolino et al. [15]	85.45	67.88	73.79	78.00
Bayar and Stamm [16]	84.55	73.72	82.52	70.67
Rahmouniet al. [17]	85.45	64.23	56.31	60.07
MesoNet [18]	87.27	56.20	61.17	40.67
XceptionNet [34]	96.36	86.86	90.29	80.67
3D ResNet	91.81	89.6	88.75	73.5
3D ResNeXt	93.36	86.06	92.5	80.5
I3D	95.13	90.27	92.25	80.5

[8] Yaohui Wang, Antitza Dantcheva
A video is worth more than 1000 lies. Comparing 3DCNN approaches for detecting deepfakes
Under submission, 2020.

Detection of all 4 manipulation methods

Algorithm	Train and Test	TCR
Steg. Features + SVM [13]	FS, DF, F2F, NT	55.98
Cozzolino et al. [15]	FS, DF, F2F, NT	58.69
Bayar and Stamm [16]	FS, DF, F2F, NT	66.84
Rahmouniet al. [17]	FS, DF, F2F, NT	61.18
MesoNet [18]	FS, DF, F2F, NT	70.47
XceptionNet [34]	FS, DF, F2F, NT	81.0
3D ResNet	FS, DF, F2F, NT	83.86
3D ResNeXt	FS, DF, F2F, NT	85.14
I3D	FS, DF, F2F, NT	87.43

Cross-manipulation techniques

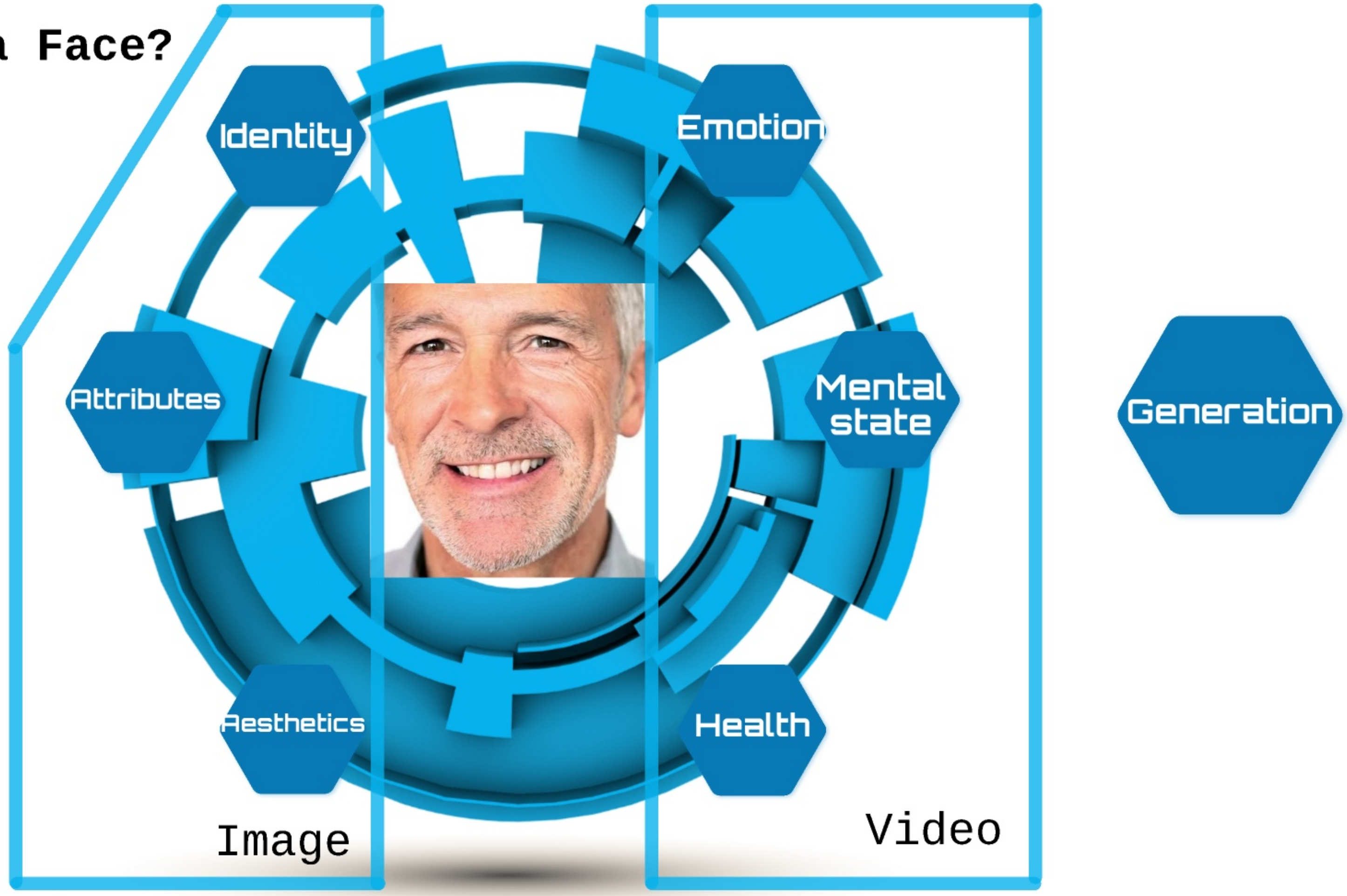
Train	Test	3D ResNet	3D ResNeXt	I3D
FS, DF, F2F	NT	64.29	68.57	66.79
FS, DF, NT	F2F	74.29	70.71	68.93
FS, F2F, NT	DF	75.36	75.00	72.50
F2F, NT, DF	FS	59.64	57.14	55.71

Future Work

Detect deepfakes:
- better discriminators (cat and mouse game)
- analyzing appearance preservation and motion consistency
- generated noise
- behavior-based

What's in a Face?

Machine Learning and Computer Vision: decipher appearance and dynamics of faces



Antitza Dantcheva - Inria, STARS Team