Visual Attribute Learning: From STL to MTL

Hu Han
VIPL group @ICT, CAS
2017/08/30
hanhu@ict.ac.cn
http://www.escience.cn/people/hhan/index.html

中国科学院计算技术研究所
Institute of Computing Technology,Chinese Academy of Sciences

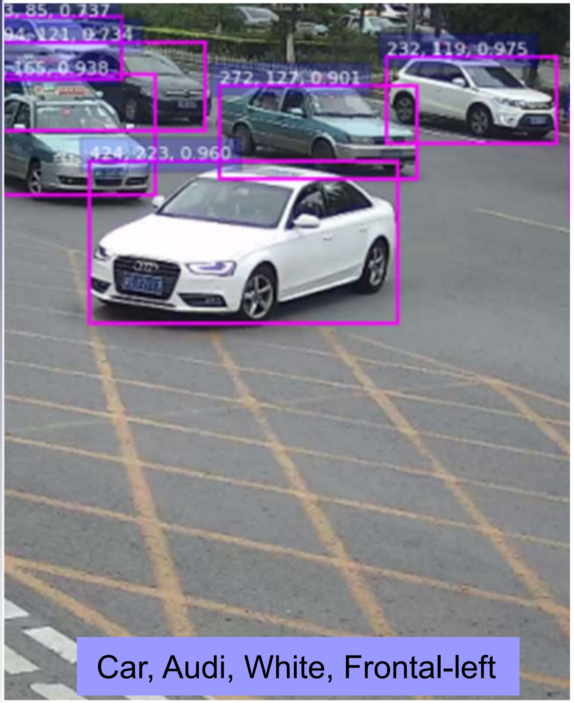
Outline

- Background
- Related work
- Attribute learning via STL
- Attribute learning via MTL
- Conclusion and discussion
- Data, demo, etc.

中科院计算所
Institute of Computing Technology, Chinese Academy of Sciences

Background

- What can an image tell us?



Vehicle

Pedestrian

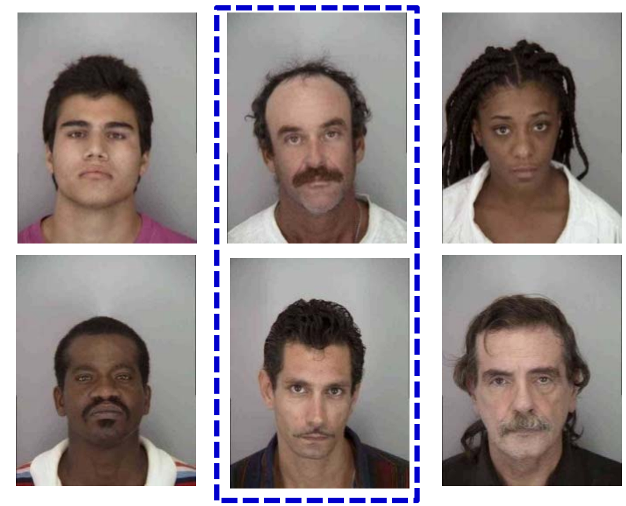
Identity: ABC
Age: ~ 40
Gender: Male
Race: White
Hair: Short, Brown
Moustache: Yes
Beard: Yes
Mole: Yes
Scar: Yes

Face

Background

- Wide applications of face attributes

Filtering: 30-40 yrs old, white, male



Access control: age estimation can prevent minors from purchasing alcohol or cigarette from vending machines

Retail advertisement: advertisements (e.g., smart shopping cart), can be changed dynamically based on customer demographics

Face retrieval: demographic information can be used to filter mugshot databases

http://www.ubergizmo.com/2011/12/krafts-pudding-dispensing-machine-is-child-proof/
http://www.selfserviceworld.com/article/166151/From-RFID-World-Media-Cart-deploys-smart-shopping-cart

Background

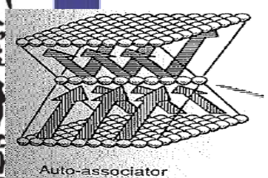
- Face visual attribute learning is nontrivial, particularly under real application scenarios
  - *Unconstrained* sensing and *uncooperative* subject: large pose, non-uniform illumination, occlusion, etc.
  - A wide variety of attributes are both *correlated* and *heterogeneous*
  - The number of face attributes can be large, requiring efficient models for attribute learning

Outline

- Background
- Related work
- Attribute learning via STL
- Attribute learning via MTL
- Conclusion and discussion
- Data, demo, etc.

Related work

- Major milestones of face attribute learning methods



1990

MIT: Cottrell & Metcalfe
把基于Auto-Endoder的特征降维用于性别和表情识别

1999

塞浦路斯学院：Lanitis构建了FGNET年龄估计数据库(82人,1002张图像)

PCA特征

2006

PCA特征

北卡：Ricanek & Tesafaye
构建了首个大规模年龄、性别、种族数据库MORPH (1.3万人，5.5万图像)

2008

哥大: Kumar等人
构建了包含10个属性的大规模名人数据库PubFig（6万图像，200人）仅部分公开

手工设计特征+SVM

2010

ICA + FFT

MIT: Pho等人
首次研究了基于普通摄像头的非接触式心率估计

“由表及里”

2015

生物启发特征+SVM

MSU: Han & Jain
首次研究了人与机器在属性识别上的性能差异(可控)，并发现机器在年龄、性别和种族的识别上已经可以超过人类

NIST组织了年龄和性别预测方面的评测竞赛

港中文: Liu等人
构建了大规模互联网名人的40属性数据集(20万图像)

深度特征+SVM

Related work

- Feature representations in AL
  - Holistic appearance
    - Intensity *[Lanitis TPAMI2002]*
    - PCA *[Lanitis et al. TPAMI02, Geng et al. TPAMI07, ...]*
    - Gabor, LBP *[Choi et al. PR11]*
    - BIF (Biologically Inspired Features) *[Guo et al. CVPR09, CVPR11]*
  - Wrinkle, skin color, and 2D shape, etc.
    - Wrinkle *[Hayashi et al. ICPR02]*
    - Skin color *[Suo et al. TPAMI10]*
  - Deep feature
    - MS-CNN *[Yi et al. ACCV14]*
    - ANet *[Liu et al. ICCV15]*
    - VGG *[Rothe IJCV16]*

Related work

- Classification methods in AL
  - Single task learning (STL)
    - One classifier (e.g., SVM) per attribute *[Kumar et al. ECCV08, TPAMI11, Geng TPAMI07, TPAMI13, Guo et al. CVPR08, Han ICB13, TPAMI15, Liu et al. ICCV15 ...]*
  - Multi-label learning
    - *[Guo and Mu ICV14, Yi et al. ACCV14]*
  - Hierarchical classifier
    - Coarse-to-fine *[Choi et al. 11, Thukral et al. 12, Han TPAMI15]*
  - Multi-task learning
    - Multi-task Restricted Boltzmann Machines *[Ehrlich CVPRW16]*
    - Multi-task CNN *[Chellappa Arxiv16]*
    - DMTL *[Han TPAMI17]*

中科院计算所

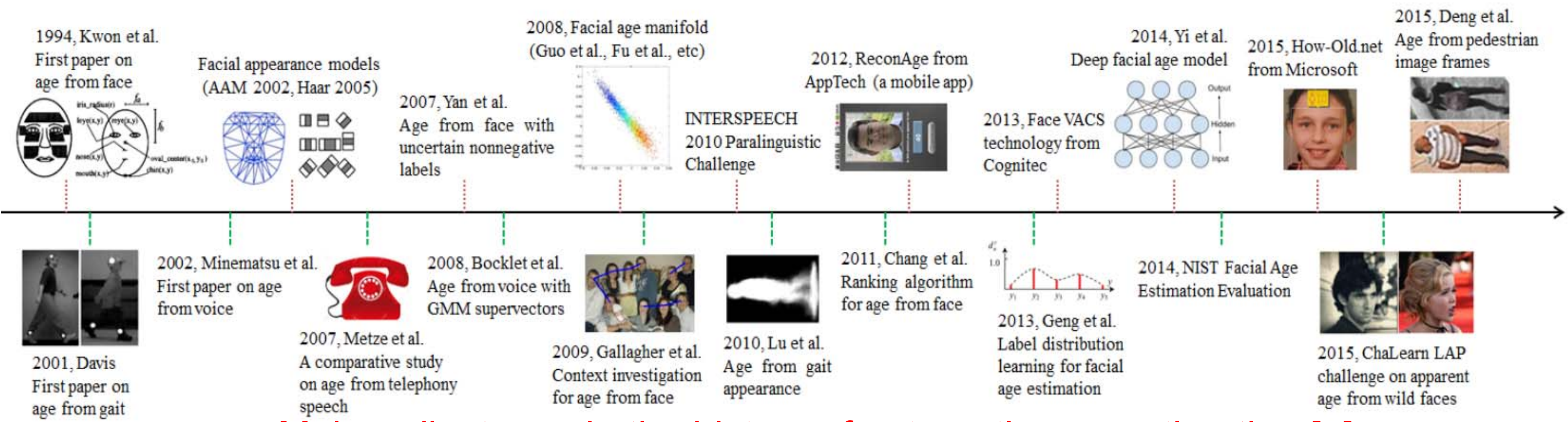
Institute of Computing Technology, Chinese Academy of Sciences

Related work

中科院计算所

Institute of

- Trend
  - **From hand-crafted features to deep features**
  - **From step-by-step to end-to-end**
  - **From STL to MTL**
    - STL methods for face attribute learning have been very popular, e.g., age estimation



Major milestones in the history of automatic age estimation [a]

[a] Yunlian Sun et al., Demographic Analysis from Biometric Data: Achievements, Challenges, and New Frontiers, TPAMI, 2017

Outline

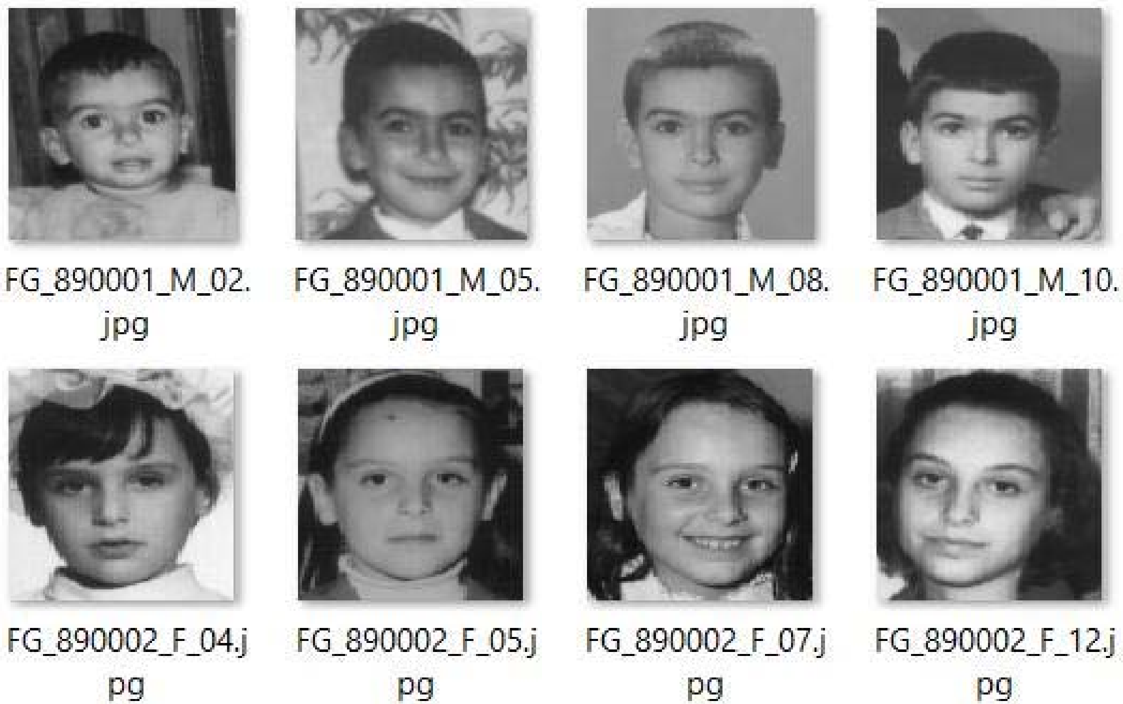
- Background
- Related work
- Attribute learning via STL
- Attribute learning via MTL
- Conclusion and discussion
- Data, demo, etc.

中科院计算所

Institute of Computing Technology, Chinese Academy of Sciences

Attribute learning via STL

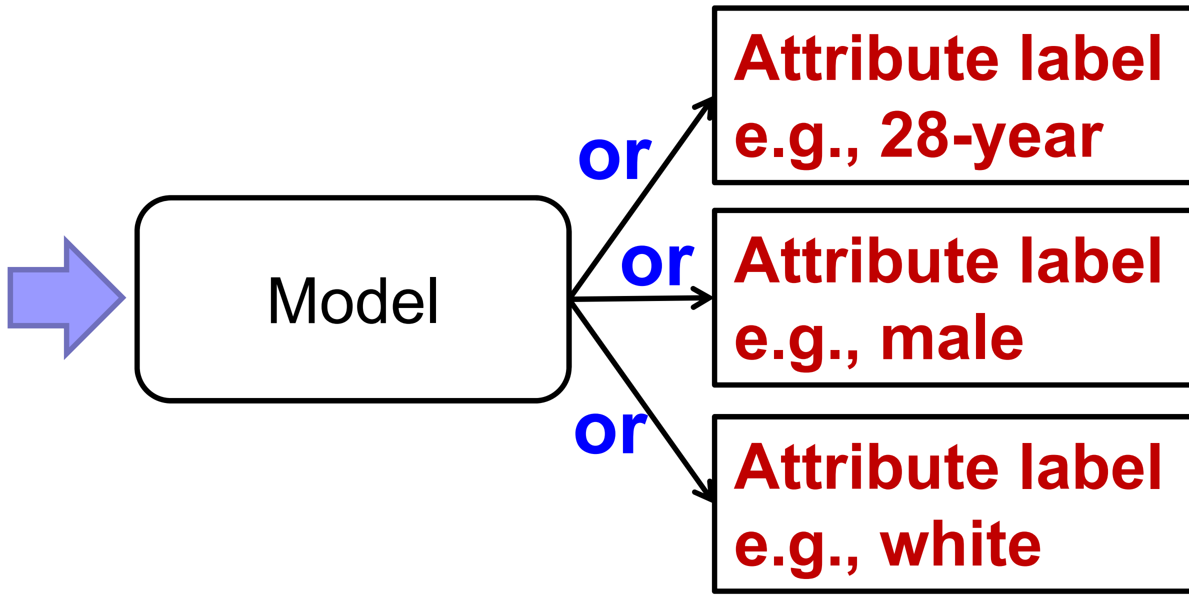
- Early databases for attribute learning are usually annotated with a single attribute



FG-NET, consisting of 1002 images of 82 subjects, has been widely used for age estimation since 1999

Attribute learning via STL

- Label a face image automatically with a label of a particular attribute, e.g., age/age group

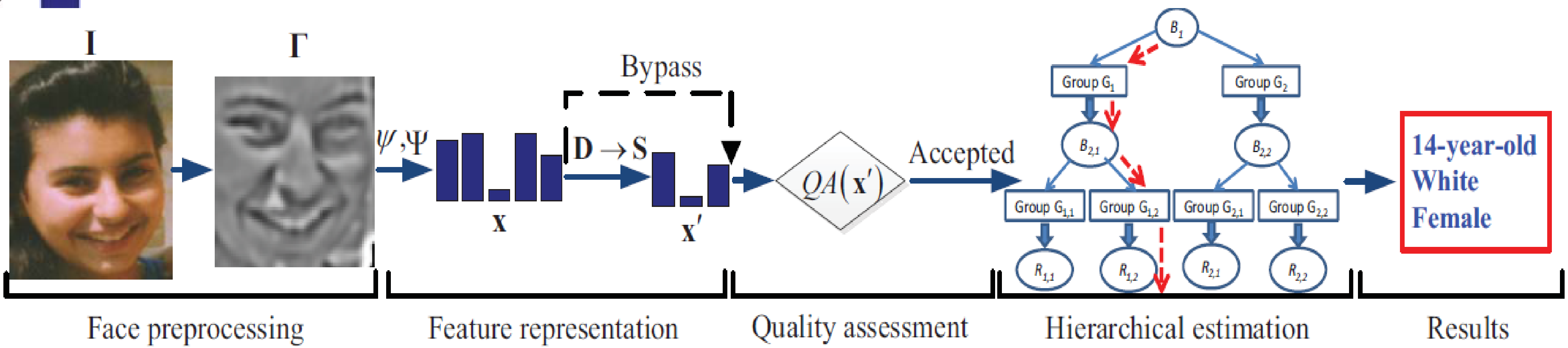


Attribute learning via STL

Publication	Face representation	Face database (#subjects, #images)	Human perception of age	Performance measure and accuracy (MAE†, 5-year CS‡)
Lanitis *et al.* (2002) [17]	2D shape, Raw intensity	FG-NET (NA, 565)	Studied on 32 face images	MAE: 4.3
Geng *et al.* (2007) [19]	Active appearance model (AAM)	FG-NET (82, 1002) MORPH (NA, 433)	Studied on 51 face images from FG-NET	FG-NET / MORPH MAE: 6.8 / 8.8 CS: ∼65% / ∼46%
Yang and Ai (2007) [7]	LBP, Haar-like features	FERET (1196, 3540) CMU-PIE (68, 696) Snapshot (NA, 9000)	Not studied	FERET / CMU-PIE / Snapshot Age group: 92.1% / 87.5% / 93.2%
Fu and Huang (2008) [21]	Manifold of raw intensity	YGA (1600, 8000)	Not studied	MAE: 5∼6 CS: F: ∼55%, M: ∼50%
Suo *et al.* (2008) [22]	Holistic and local topology, 2D shape, color, and gradient	FG-NET (82, 1002) Private (NA, 8000)	Studied on 500 images from a private database	FG-NET / Private MAE: 6.0 / 4.7 CS: ∼55% / ∼66%
Guo *et al.* (2009) [23]	Biologically inspired features (BIF)	FG-NET (82, 1002) YGA (1600, 8000)	Not studied	FG-NET / YGA MAE: 4.8 / F: 3.9, M: 3.5 CS: 47% / F: 75%, M: 80%
Guo and Wang (2011) [25]	BIF	MORPH II (NA, 55132)	Not studied	MAE: 4.2
Choi *et al.* (2011) [26]	AAM, Gabor, LBP	FG-NET (82, 1002) PAL (NA, 430) BERC (NA, 390)	Not studied	FG-NET / PAL / BERC MAE: 4.7 / 4.3 / 4.7 CS: ∼73% / ∼70% / ∼65%
Luu *et al.* (2011) [27]	Holistic contourlet appearance model	FG-NET (82, 1002) PAL (NA, 443)	Not studied	FG-NET / PAL MAE: 4.1 / 6.0 CS: ∼74% / NA
Chang *et al.* (2011) [28]	AAM	FG-NET (82, 1002) MORPH II (NA, 5492)	Not studied	FG-NET / MORPH II MAE: 4.5 / 6.1 CS: 74.4% / 56.3%
Wu *et al.* (2012) [29]	Grassmann manifold of 2D shape	FG-NET (82, 1002) Passport (109, 233)	Not studied	FG-NET / Passport MAE: 5.9 / 8.8 CS: 62% / 40%
Thukral *et al.* (2012) [30]	Grassmann manifold of 2D shape	FG-NET (82, 1002)	Not studied	MAE: 6.2
Chao *et al.* (2013) [32]	AAM with distance metric adjustment	FG-NET (82, 1002)	Not studied	MAE: 4.4
Lu and Tan (2013) [31]	Manifold of raw intensity	MORPH II (NA, 20000)	Not studied	White / Black MAE: 5.2 / 4.2 CS: 67% / 59%
Hadid and Pietikäinen (2013) [8]	Raw intensity, volume LBP	Internet videos (NA, 2000)	Not studied	Age group classification: 83.1%
Guo and Mu (2013) [33]	BIF	MORPH II (NA, 55132)	Not studied	MAE: 4.0
Geng *et al.* (2013) [34]	AAM, BIF	FG-NET (82,1002) MORPH II (13000, 55132)	Studied on 51 and 60 images, respectively, from FG-NET and MORPH II	FG-NET / MORPH II MAE: 4.8 / 4.8
Proposed method	Demographic informative features	FG-NET (82, 1002) MORPH II (20569, 78207) PCSO (81533, 100012) LFW (4211, 4211)	Studied on 1002, 2000, 2200 and 4211 images, respectively, from FG-NET, MORPH II, PCSO, and LFW	FG-NET / MORPH II / PCSO / LFW MAE: 3.8 / 3.6 / 4.1 / 7.8 CS: 78.0% / 77.4% / 72.6% / 42.5%

Demographic informative feature

- Overview



- Highlight
  - Demographic informative features (DIF)
  - Hierarchical classification
  - Human vs. machine performance

Hu Han et al., "Demographic Estimation from Face Images: Human vs. Machine Performance," TPAMI 2015.
Hu Han et al., "Age Estimation from Face Images: Human vs. Machine Performance," ICB, 2013. (Oral)

Demographic informative feature

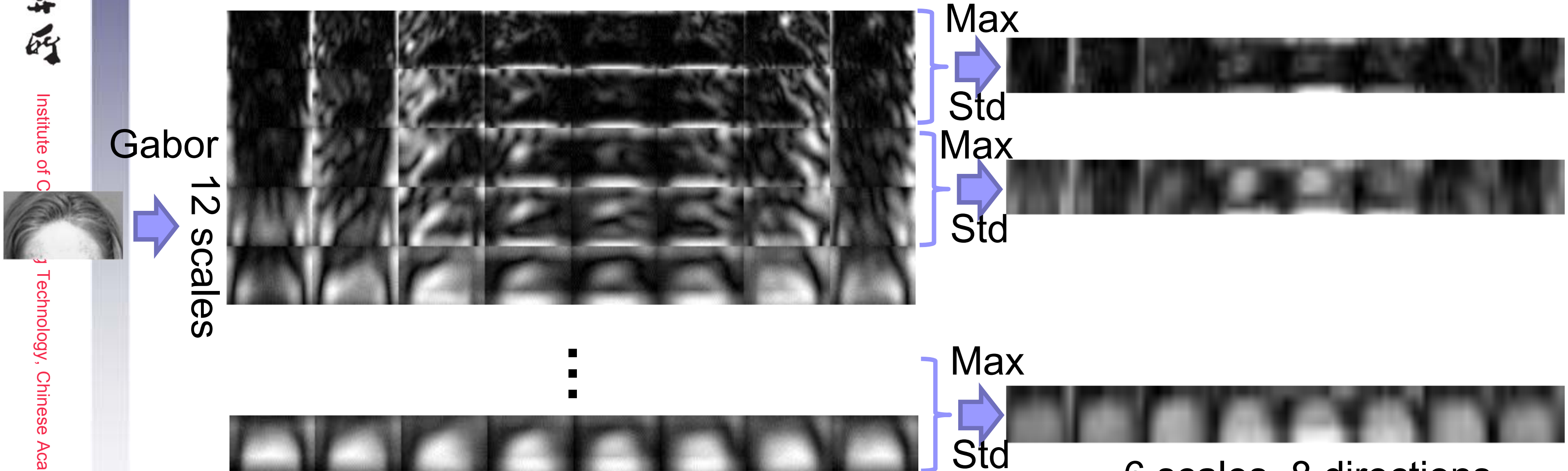
中科院计算所

Institute of C… Technology, Chinese Academy of Sciences

- Demographic informative features
  - Based on BIF, but introduced boosting feature selection

S1 layer: Simulate the simple (S) cell units

C1 layer: Simulate the complex (C) cell units



8 directions

6 scales, 8 directions

All C1 layer features are concatenated into a 4280D feature vector

BIF: Biologically Inspired Features

Demographic informative feature

- Demographic informative features
  - BIF is computed in an unsupervised way
  - Some dimensions of feature can be redundant or irrelevant to the attribute learning task
    - Learn a new feature subspace, e.g., LDA
    - Feature selection via boosting

General features $$\mathbf{D} = \{(\mathbf{x}_i, y_i) : \mathbf{x}_i \in \mathbb{R}^d, y_i \in \mathbb{N}, i \in [1, m]\}$$

Specific features $$\mathbf{S} = \{\mathbf{x}'_i : \mathbf{x}'_i \in \mathbb{R}^{d'}, \mathbf{x}'_i \subset \mathbf{x}_i, i \in [1, m]\}$$

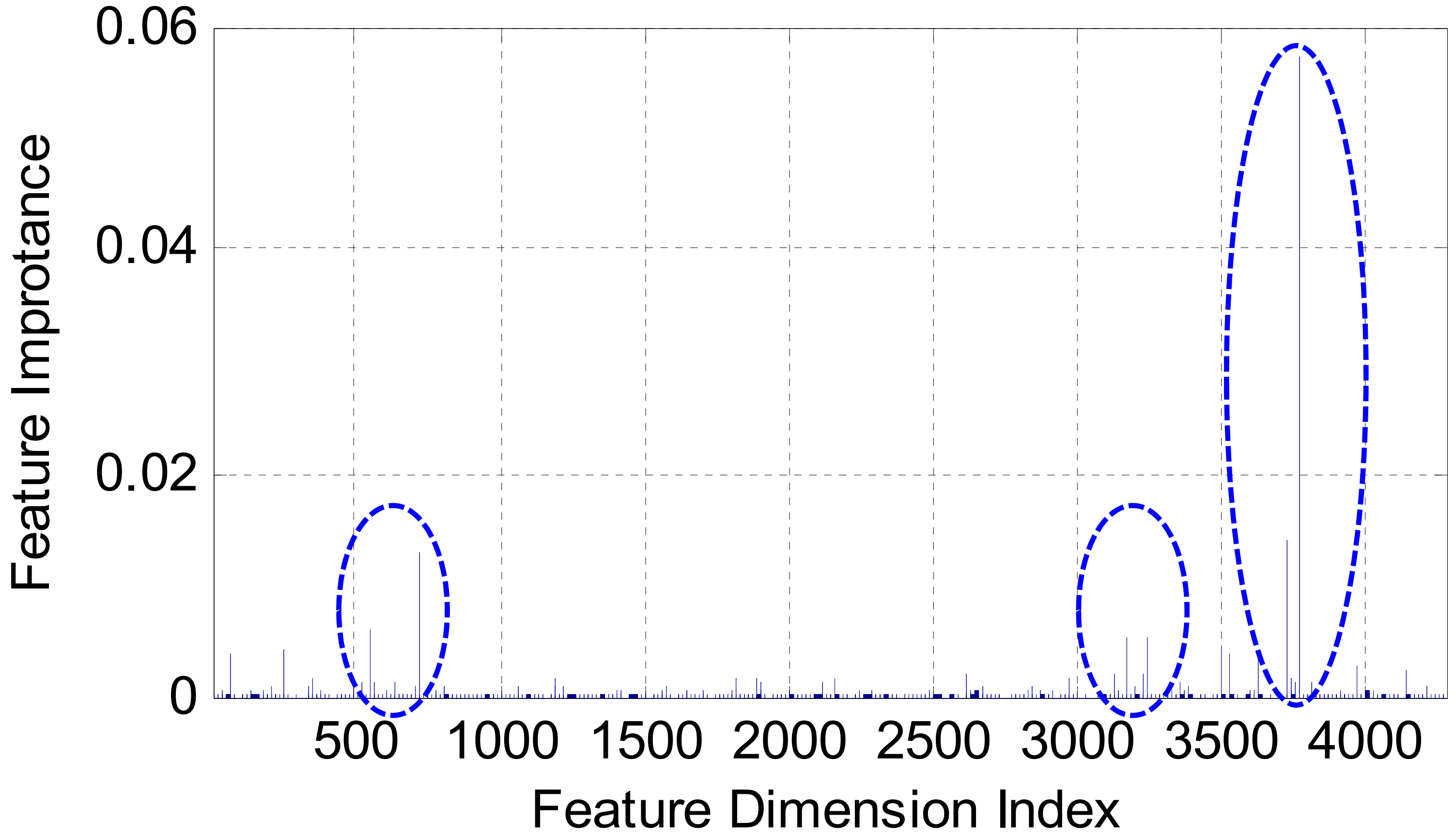
$$d' \ll d$$

中科院计算所

Institute of Computing Technology, Chinese Academy of Sciences

Demographic informative feature

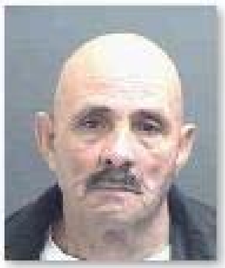
- Demographic informative features
  - Feature selection via boosting



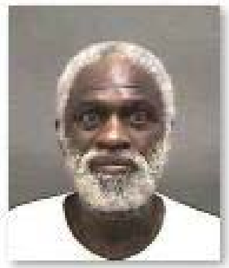
Selected 800 out of 4280 dimensions

Attribute learning via STL

- Face databases with several attribute annotations

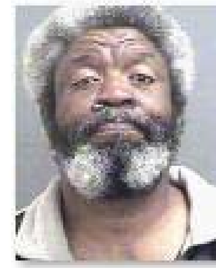


MO_028189_M_65_04.JPG



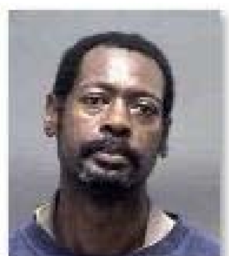
MO_029485_M_58_01.JPG

MO_035620_M_55_02.JPG

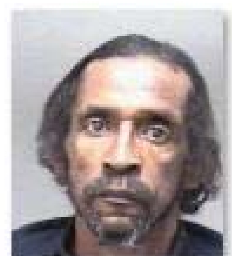


MO_035625_M_51_03.JPG

MO_047413_M_49_01.JPG



MO_047427_M_47_00.JPG



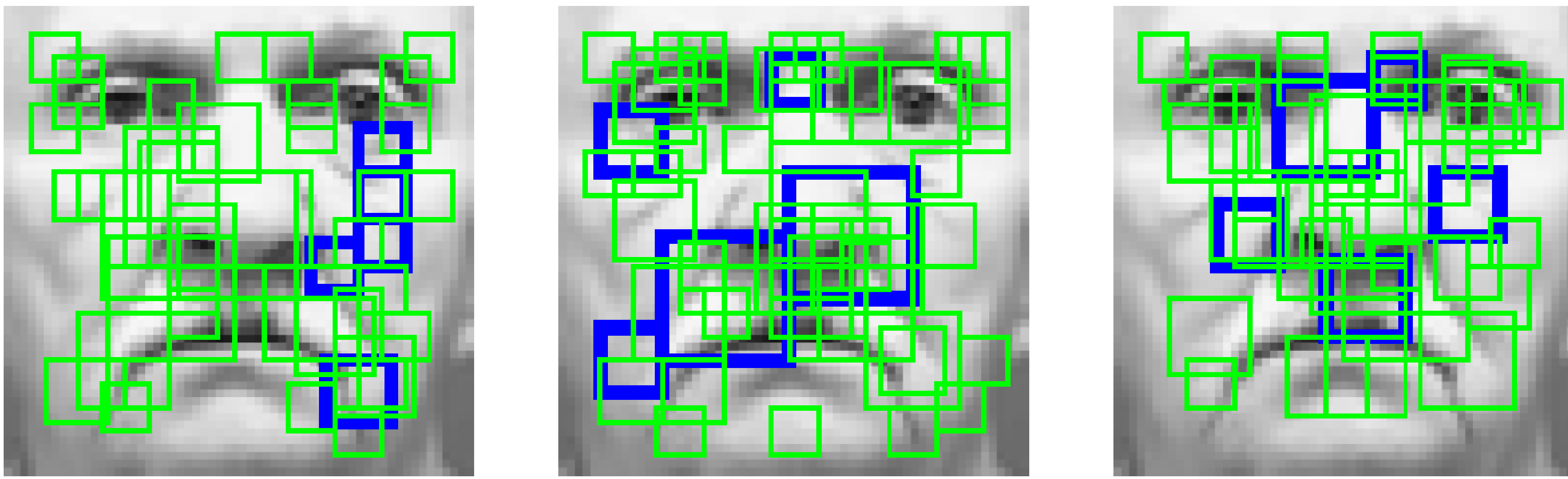
MO_047429_M_54_02.JPG

MO_047936_M_50_02.JPG

MORPH (2006), consisting of ~55,000 images with age, gender, and race information

Attribute learning via STL

- Demographic informative features
  - Visualization of feature selection



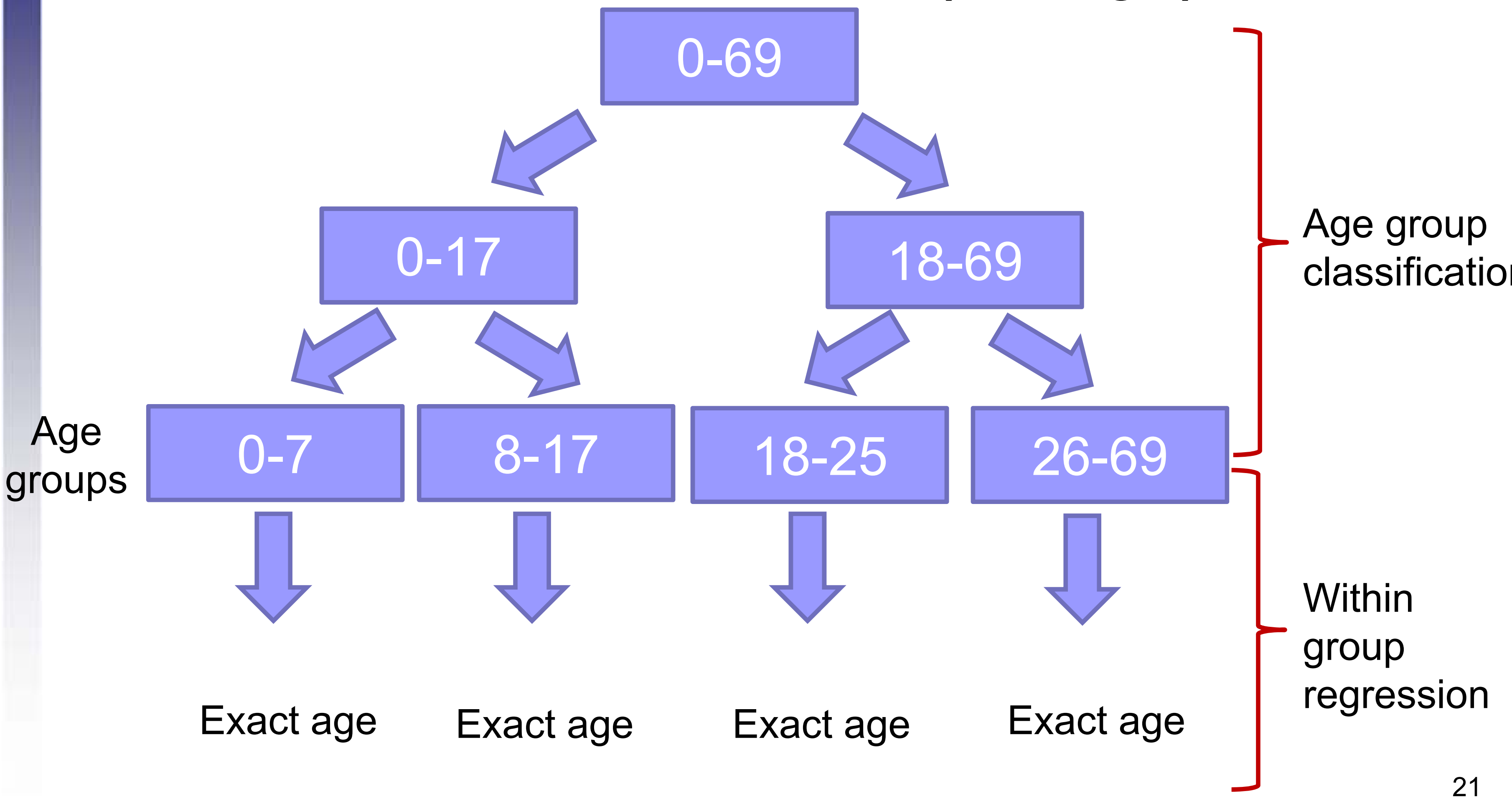
For Age | For Gender | For Race

Blue boxes: top 5 features
Green boxes: top 6-50 features

The selected featured are used by age, gender, and race estimation tasks, but the a classifier is learned for each task separately; so overall the method is STL

Attribute learning via STL

- Demographic informative features
- Hierarchical classification (for age)

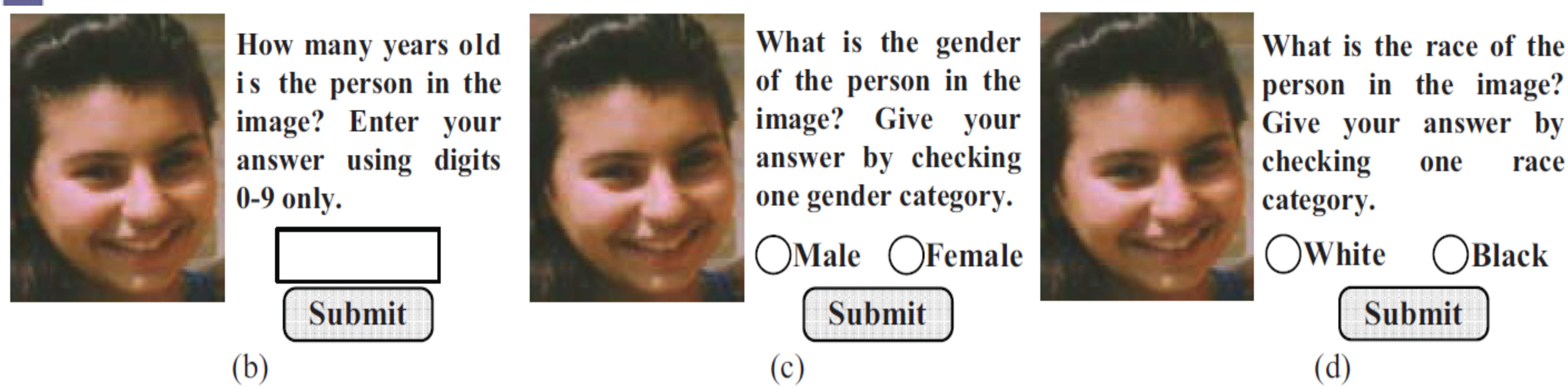


Attribute learning via STL

- Demographic informative features
- Hierarchical classification (for age)
- Human vs. machine performance
  - Compiled and released the first large-scale dataset for measuring the performance of human and machine (algorithm)
  - Human age estimates for FG-NET
  - Human age, gender, and race estimates for a MORPH set with 2000 images
  - Human age, gender, and race estimates for a PCSO set with 2000 images

Attribute learning via STL

- Human vs. machine performance
  - Data collection for measuring human performance

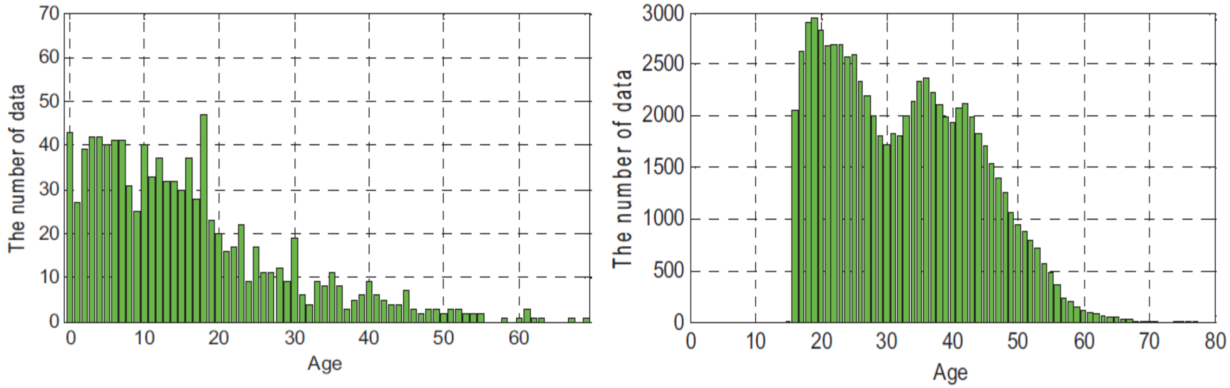


GUI shown to Amazon MTurk workers
Three cents per HIT;
Three workers per image;
Voting based on 3 workers' responses;

Age estimates by MTurk workers for FG-NET:
http://biometrics.cse.msu.edu/pub/databases.html

Demographic informative feature

- Results of age estimation on FG-NET and MORPH II



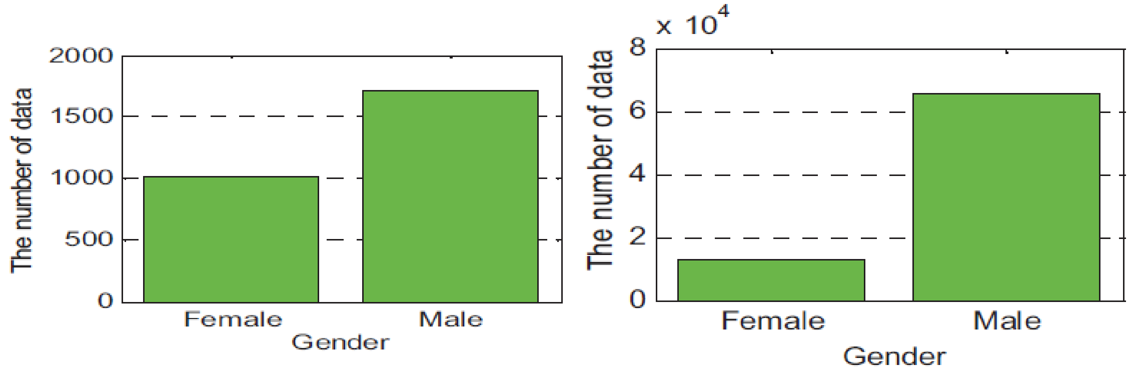
(a) FG-NET (b) MORPH

Dataset	Mean absolute error (in years)				
	Geng07	Chang11	Chao13	Guo13	Proposed
FG-NET	6.8	4.5	4.4	n/a	3.8
MORPH	8.8	6.1	n/a	4.0	3.6

Institute of Computing Technology, Chinese Academy of Sciences

Demographic informative feature

- Results of gender classification estimation on FERET and MORPH II

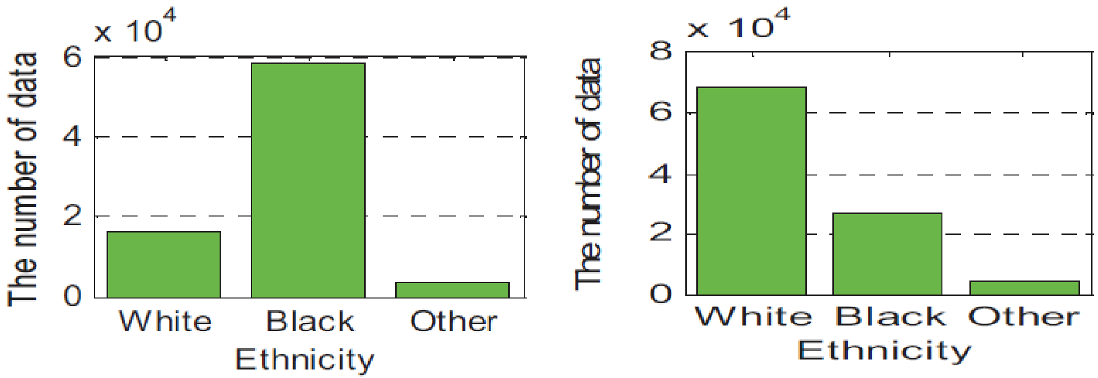


(d) FERET (e) MORPH

Dataset	Accuracy (in %)		
	Baluja07	Guo13	Proposed
FERET	94.4	n/a	96.8
MORPH	n/a	96.0	97.6

Demographic informative feature

- Results of race classification estimation on MORPH II and PCSO



(g) MORPH (h) PCSO

Dataset	Accuracy (in %)		
	Ross13	Guo13	Proposed
MORPH	98.7	98.9	99.1
PCSO	n/a	n/a	98.7

Demographic informative feature

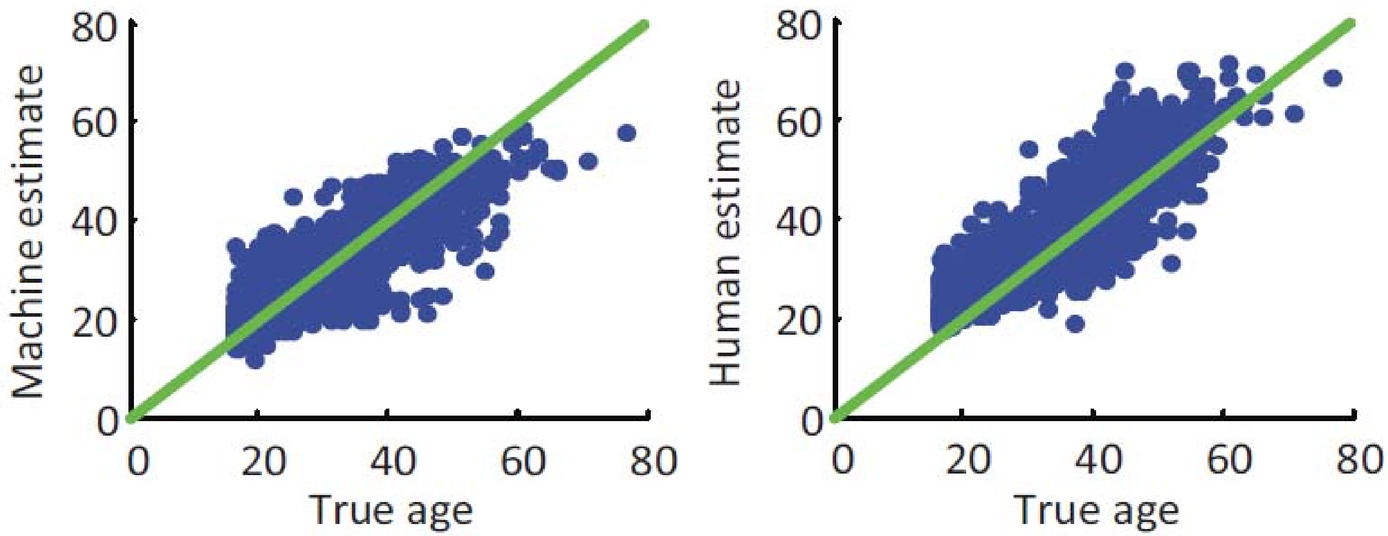
- Comparisons between human and machine

Task	Dataset	Machine	Human
Age estimation	FGNET	3.8 yr.	4.7 yr.
	MORPH	3.6 yr.	6.3 yr.
Gender classification	FERET	96.8%	n/a
	MORPH	97.6%	96.9%
Race classification	MORPH	99.1%	97.8%
	PCSO	98.7%	96.5%

Machine outperforms human!

Demographic informative feature

- Comparisons between human and machine

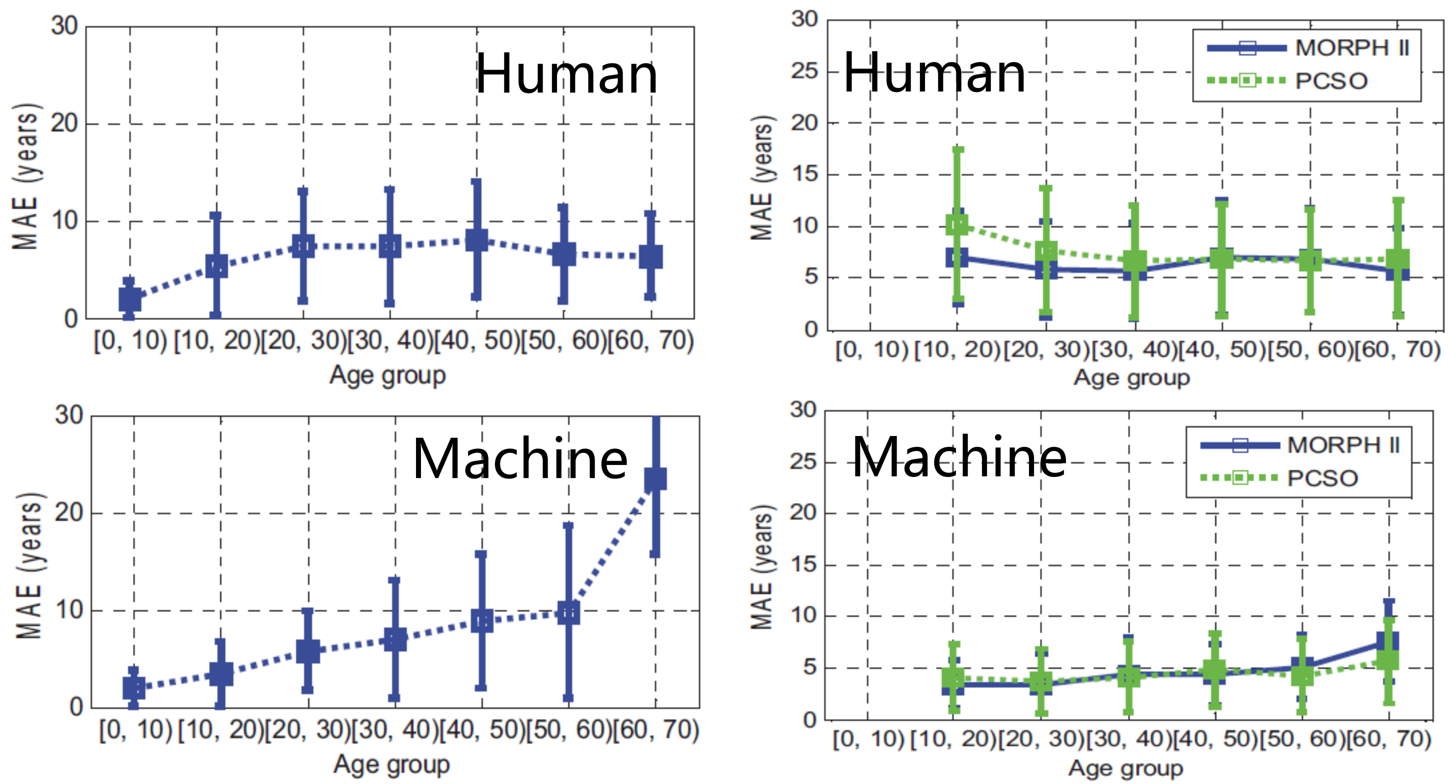


(a) (b)

On average, human tend to overestimate the age

Demographic informative feature

- Comparisons between human and machine

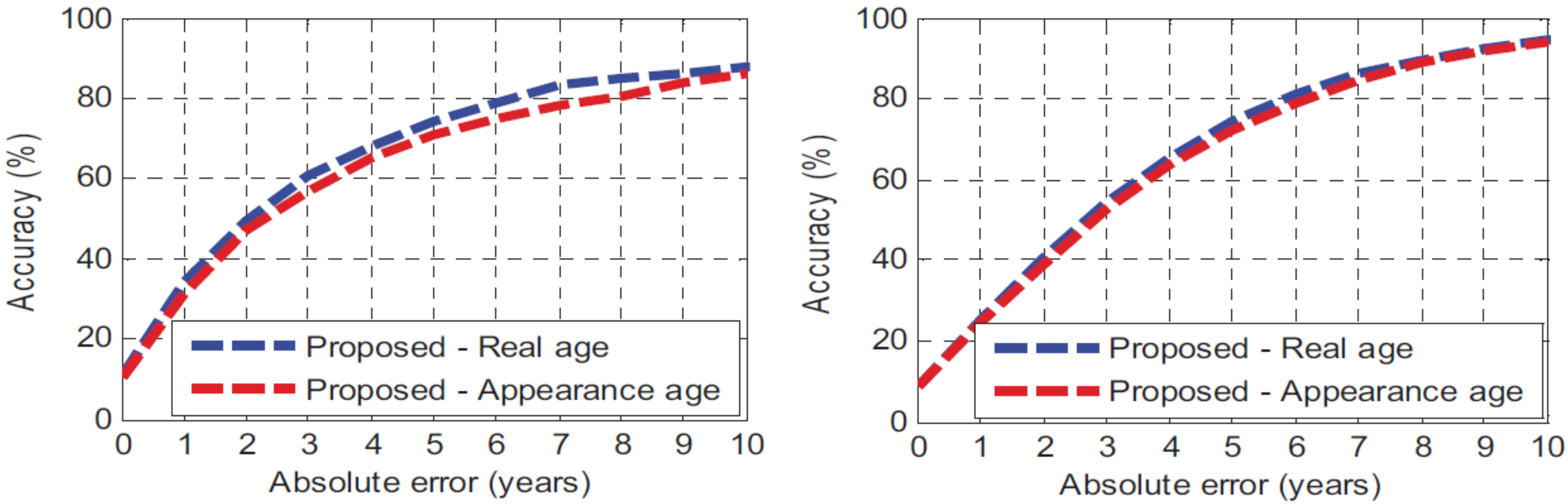


(a) FG-NET (b) MORPH II & PCSO

Machine can perform better than human, but human is more stable

Demographic informative feature

- Estimating the real age vs. apparent age

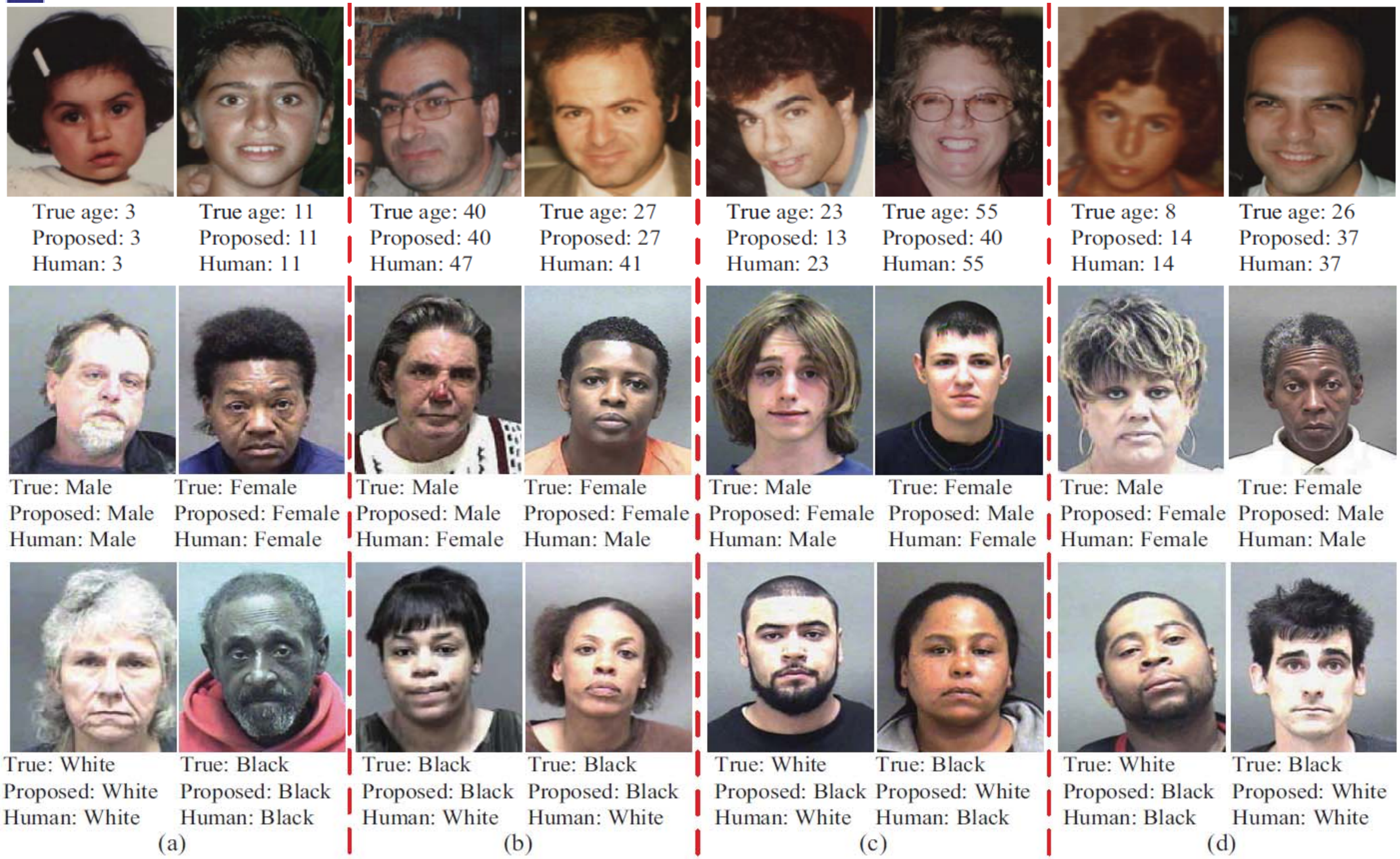


(a) FG-NET (b) MORPH II

Real age or apparent age makes minor differences to machine's (algorithm's) accuracy

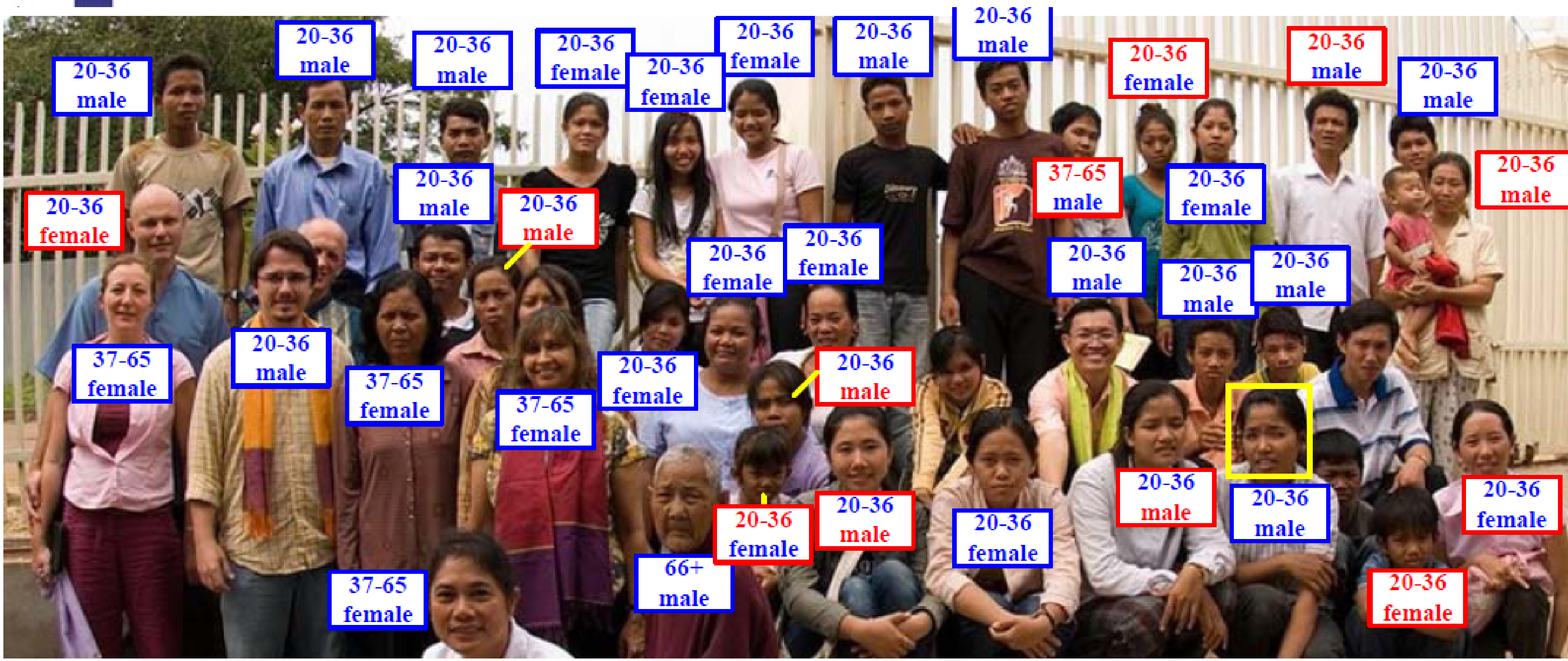
Demographic informative feature

- Examples of attribute learning results



Demographic informative feature

- Examples of attribute learning results



An image from the Images of Groups database.

Attribute learning via STL

- A short summary
  - Learned shared DIF features that are informative for age, gender, and race estimation tasks simultaneously
  - A hierarchical classification model for coarse-to-fine age estimation
  - Compiled and released the first large-scale dataset for measuring the performance of human and machine (algorithm)
  - Estimates by MTurk workers: http://biometrics.cse.msu.edu/pub/databases.html

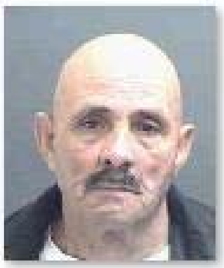
Hu Han et al., "Demographic Estimation from Face Images: Human vs. Machine Performance," TPAMI 2015.
Hu Han et al., "Age Estimation from Face Images: Human vs. Machine Performance," ICB, 2013. (Oral)

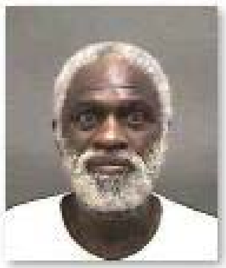
Outline

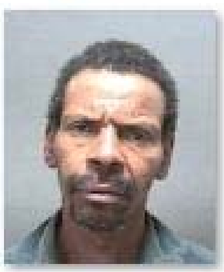
- Background
- Related work
- Attribute learning via STL
- Attribute learning via MTL
- Conclusion and discussion
- Data, demo, etc.

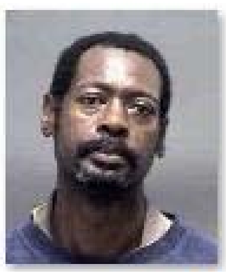
Background

- Recent face databases with several attribute annotations


MO_028189_M_65_04.JPG


MO_029485_M_58_01.JPG


MO_047413_M_49_01.JPG


MO_047427_M_47_00.JPG

MORPH has age, gender, and race attributes


000002.jpg


000004.jpg


000028.jpg


000029.jpg

CelebA has 40 binary attributes: hair, eyebrow, nose, beard, gender…

Background

- Goal: Label a face image automatically with a set of attribute labels

Model

28-year

male

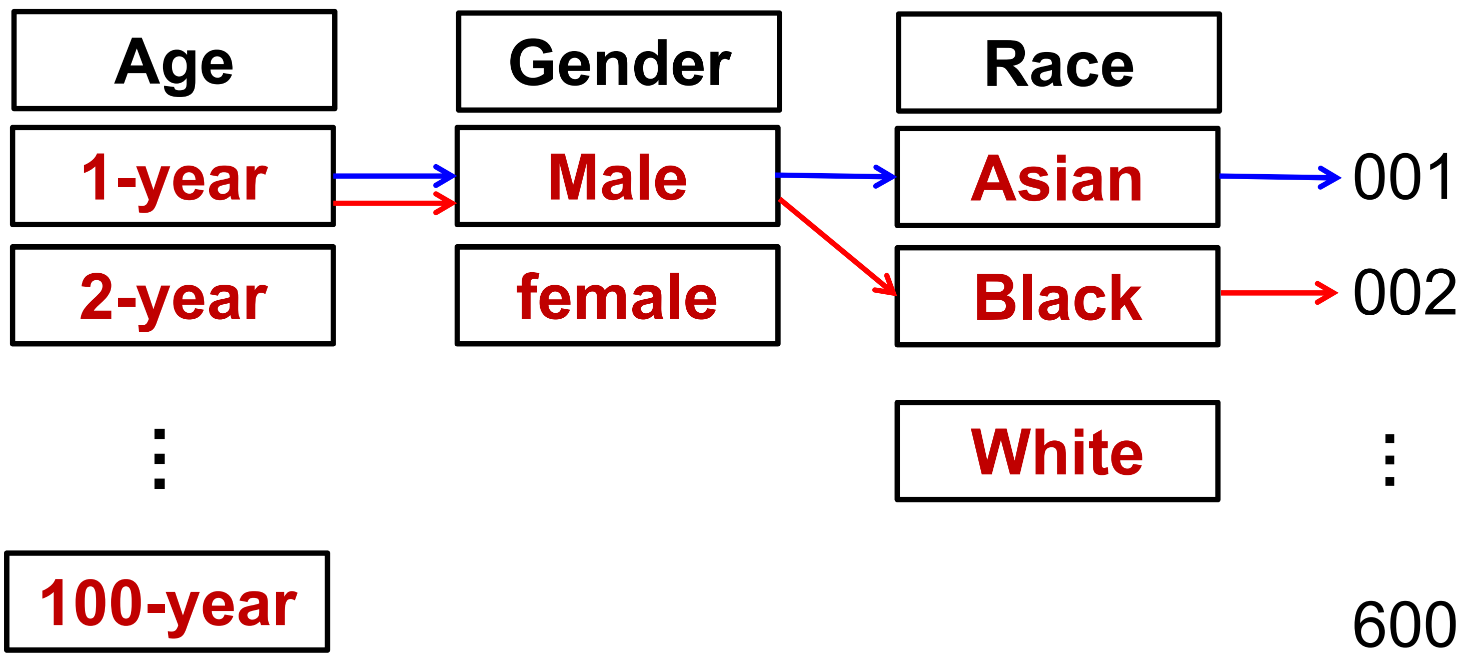
white

eye glasses

short hair

Solution (1): Label coding

- A simple solution: label coding



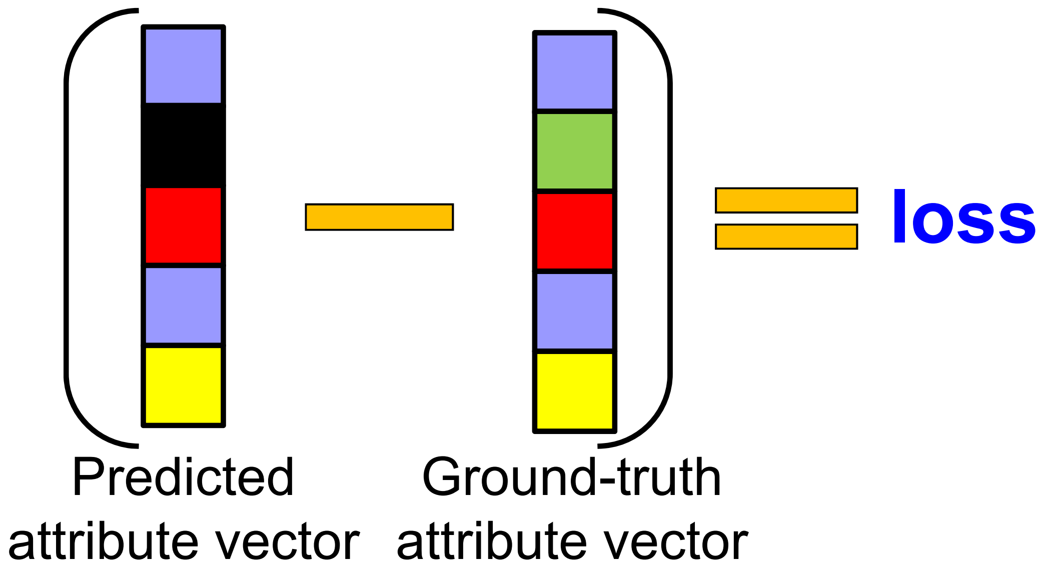
Converted from multi-attribute into single-attribute

Cons: difficult to handle a large number of attributes

H. Han and A. K. Jain, "Age, Gender and Race Estimation from Unconstrained Face Images," MSU Technical Report, MSU-CSE-14-5, 2014

Solution (2): Multi-label regression

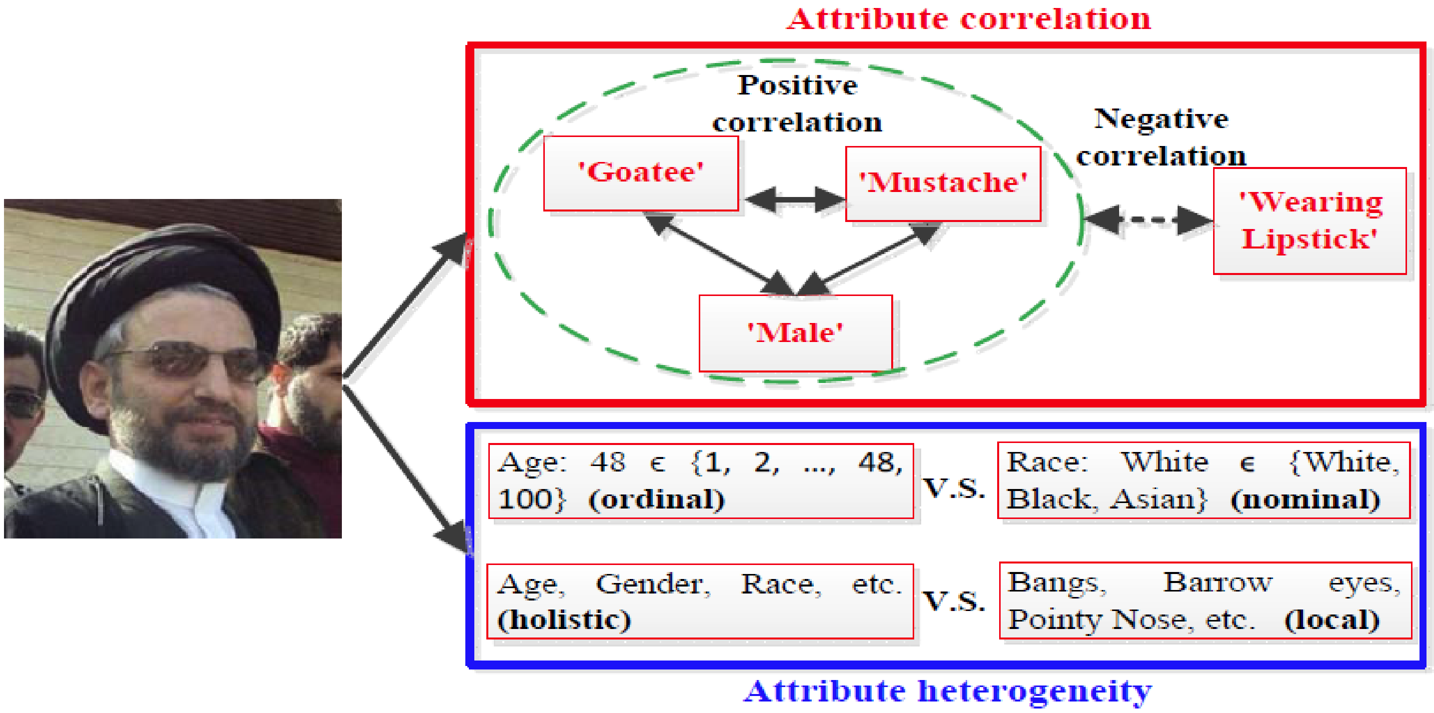
- Regression of a attribute vector with each element denoting one attribute [Yi et al. ACCV14, Chellappa arXiv16]



Cons: the same feature is used for multiple attribute learning tasks; which is not optimal

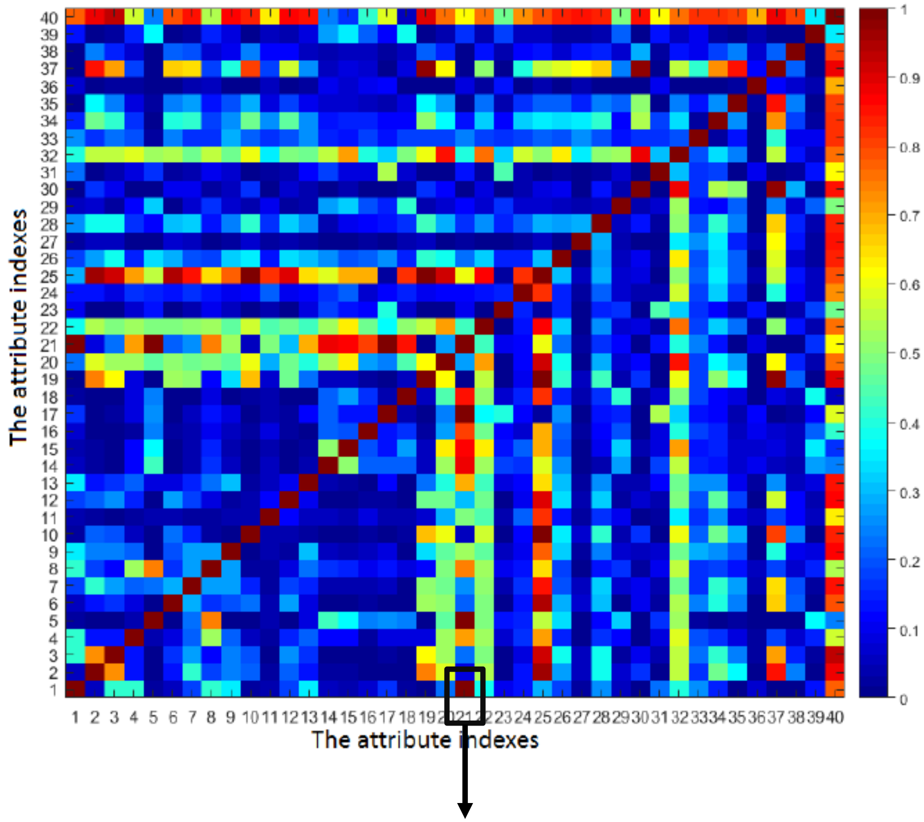
Attribute learning via MTL

- Joint learning of features and classifiers that are optimal for individual tasks
  - How to model the attribute correlations and attribute heterogeneities?



Attribute learning via MTL

- Attribute correlation



(5 O’ClockShadow, Male)

Pair-wise co-occurrence matrix of the 40 face attributes provided with the CelebA database

Attribute correlation is helpful for learning informative and robust feature representations.

Attribute learning via MTL

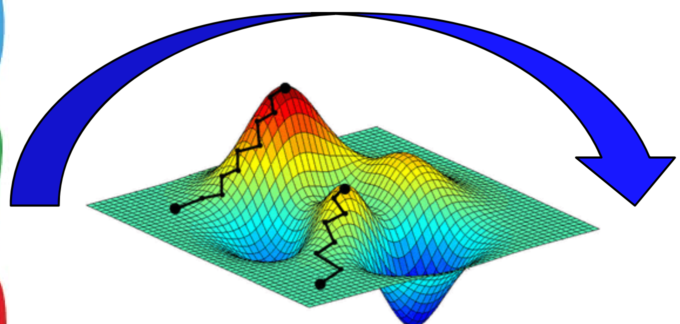
- Attribute heterogeneity
  - Data type and scale of individual attribute
    - Ordinal vs. local
      - Ordinal attribute, such as, age [0, 1, 2, ..., 100] (has a clear ordering of its variables)
      - Nominal attribute, such as, race {Asian, Black, White} (no intrinsic ordering)
    - Holistic vs. local
      - Age, gender, and race describe the whole face's characteristic, while pointy nose and big lips describe the local facial components' characteristics

Attribute heterogeneity can be handled in a divide and conquer way.

Deep multi-task learning

- Formulation

Image space



Non-linear
High dimensional

Attribute space

age
gender
⋮
hair

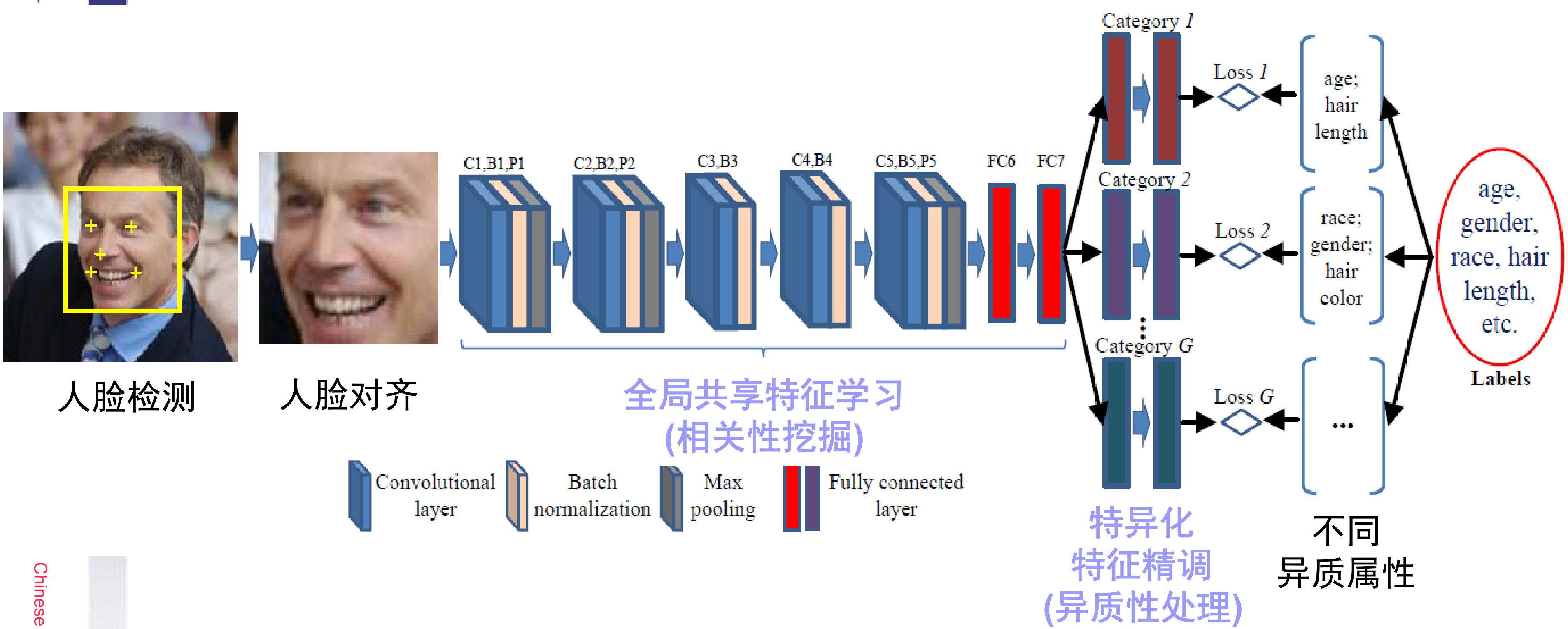
$$\mathbf{X} = \{X_i\}_{i=1}^{N} \qquad \mathbf{D} = \{\mathbf{X}, \mathbf{Y}\} \qquad \mathbf{Y} = \{\{y_i^j\}_{j=1}^{M}\}_{i=1}^{N}$$

N images

Each image has M attributes

Deep multi-task learning

■ Overview of Deep Multi-task Learning



Hu Han et al., "Heterogeneous Face Attribute Estimation: A Deep Multi-Task Learning Approach," TPAMI 2017.8.
Wang et al., "Deep Multi-Task Learning for Joint Prediction of Heterogeneous Face Attributes", FG, 2017.5.
Han & Jain, "Age, Gender and Race Estimation from Unconstrained Face Images," MSU TR, 2014.

Deep multi-task learning

- MTL loss

$$\underset{\{W^j\}_{j=1}^{M}}{\arg\min} \sum_{j=1}^{M} \sum_{i=1}^{N} \underbrace{\boxed{\mathcal{L}}}_{\text{Loss function}}\left(y_i^j, \mathcal{F}(X_i, \underbrace{\boxed{W^j}}_{\text{Network weights}})\right) + \underbrace{\boxed{\gamma \Phi(W^j)}}_{\text{Regularization term}}$$

Learn the same features and classifiers for M different tasks

Deep multi-task learning

- MTL loss considering attribute heterogeneity

Loss function for each of the heterogeneous attributes

Subnetwork weight

Shared network weight

$$\arg\min_{W_c,\{W^j\}_{j=1}^{M}} \sum_{g=1}^{G} \sum_{j=1}^{M^g} \sum_{i=1}^{N} \lambda^g \boxed{\mathcal{L}^g}\left(y_i^j, \mathcal{F}(X_i, \boxed{W^g} \circ \boxed{W_c})\right)$$

$$+ \boxed{\gamma_1 \Phi(W_c) + \gamma_2 \Phi(W^g)}$$

Regularization term

Learn task-specific features and classifiers for M different tasks, while sharing features at the early stage.

中科院计算所

Institute of Computing Technology, Chinese Academy of Sciences

Deep multi-task learning

- Evaluations

(a) Face images from the MORPH II database

(b) Face images from the LFW+ database

(c) Face images from the CelebA database

(d) Face images from the LFWA database

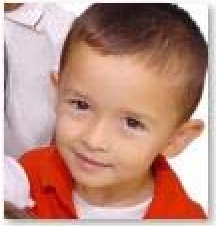
(e) Face images from the ChaLearn LAPAge15 and FotW databases

Five databases in public domain

Deep multi-task learning

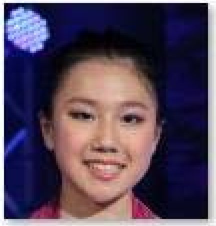
- LFW+ database (~15,699 images)
  - Extend the LFW database with 2,466 unconstrained face images of young subjects (0–20 years)
  - Three MTurk workers were asked to provide their estimates of age, gender, and race for each image
  - Will be available here: http://biometrics.cse.msu.edu/pub/databases.html

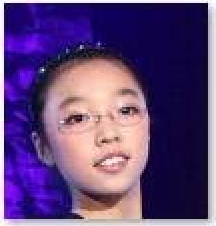

IN_004148_0007.jpg


IN_004148_0006.jpg


IN_004148_0005.jpg


IN_004094_0004.jpg


IN_004091_0005.jpg


IN_004090_0003.jpg


IN_004089_0011.jpg

Deep multi-task learning

- Accuracy for nominal and ordinal attributes

Approach	MORPH II			LFW+		
	Age[2]	Gender	Race	Age[2]	Gender	Race
Guo and Mu [19]	3.92/70.0	98.5	99.0	NA	NA	NA
Yi *et al.* [20]	3.63/NA	98.0	99.1	NA	NA	NA
DIF [16]	3.8/75.0	97.6	99.1[3]	7.8/42.5[4]	94[4]	90[3,4]
DEX [17]	3.25/NA	NA	NA	NA	NA	NA
DEX [17][1]	2.68/NA	NA	NA	NA	NA	NA
Proposed	3.0/85.3	98.0	98.6	4.5/75.0	96.7	94.9

1 The IMDB-WIKI database was used for network pre-training.

Deep Multi-task Learning

- Accuracy for binary attributes (CelebA, LFWA)

	Approach	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
CelebA	FaceTracker [27]	85	76	80	78	76	89	88	70	80	60	90	64	74	81	86	88	98	93	85	84
	PANDA [57]	88	78	86	81	79	96	92	85	93	77	94	67	75	86	86	88	98	93	90	86
	LNets+ANet [23]	91	79	90	81	79	98	95	88	95	80	97	68	78	84	91	92	99	95	90	87
	CTS-CNN [34]	89	83	87	82	79	96	94	87	93	79	95	70	79	87	88	89	99	94	91	87
	MCNN-AUX [33]	**95**	83	**93**	83	85	**99**	**96**	**90**	**96**	89	**98**	71	85	**96**	96	96	**100**	97	**92**	**88**
	Proposed	**95**	**86**	85	**85**	**99**	**99**	**96**	85	91	**96**	96	**88**	**92**	**96**	**97**	**99**	99	**98**	**92**	**88**
LFWA	FaceTracker [27]	70	67	67	71	65	77	72	76	88	62	78	68	73	73	67	70	90	69	88	77
	PANDA [57]	**84**	79	79	81	80	84	84	87	94	74	81	73	79	74	69	75	89	75	93	86
	LNets+ANet [23]	**84**	82	82	83	83	88	88	90	**97**	77	84	75	81	74	73	78	**95**	78	95	88
	CTS-CNN [34]	77	83	83	79	83	91	**91**	90	**97**	76	87	78	83	**88**	75	80	91	83	95	88
	MCNN-AUX [33]	77	82	**85**	80	83	92	90	**93**	**97**	81	**89**	79	**85**	85	77	82	91	83	**96**	88
	Proposed	80	**86**	82	**84**	**92**	**93**	77	83	92	**97**	**89**	**81**	80	75	**78**	**92**	86	**88**	95	**89**

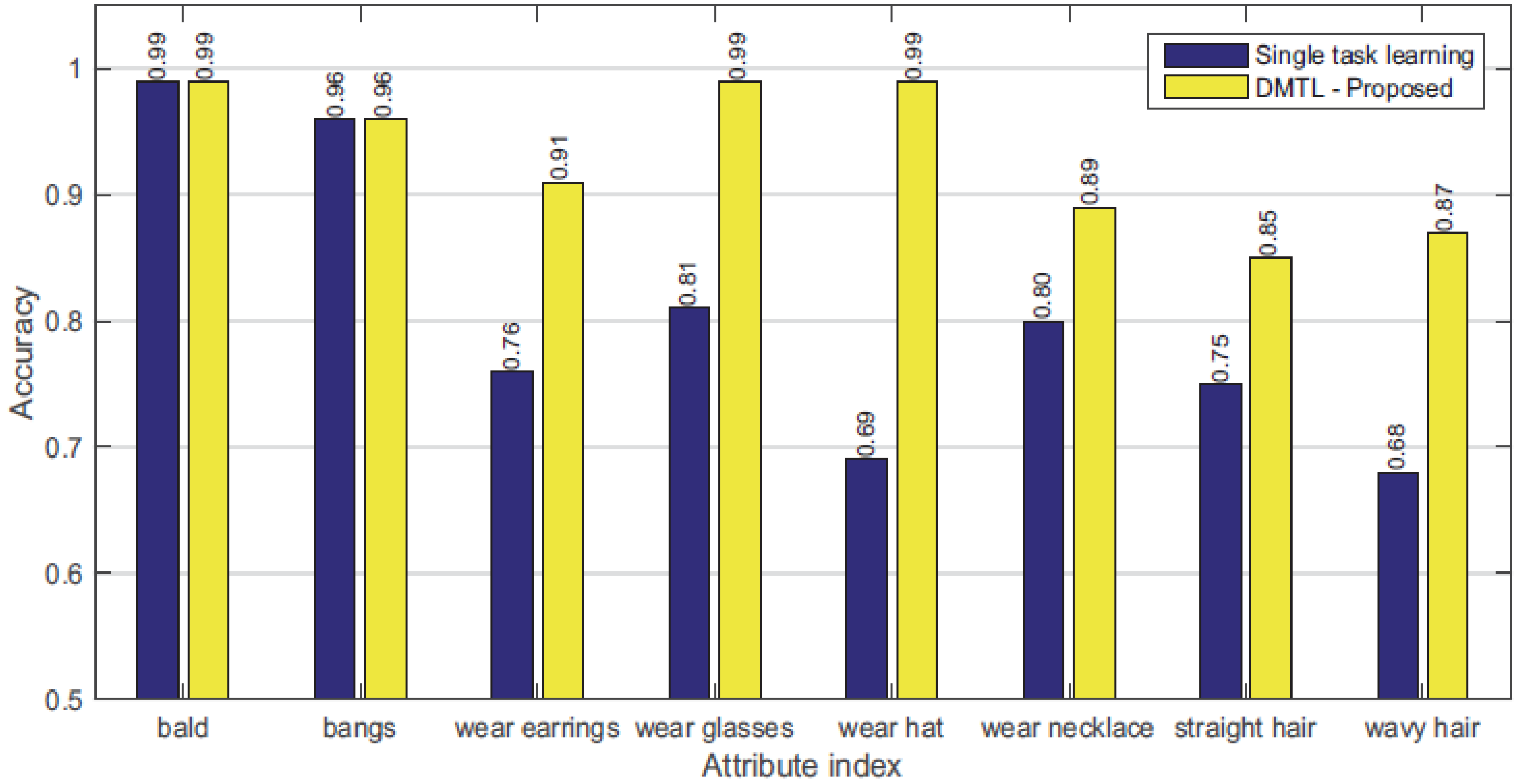
	Approach	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
CelebA	FaceTracker [27]	91	87	91	82	90	64	83	68	76	84	94	89	63	73	73	89	89	68	86	80
	PANDA [57]	97	93	93	84	93	65	91	71	85	87	93	92	69	77	78	96	93	67	91	84
	LNets+ANet [23]	98	92	95	81	95	66	91	72	89	90	96	92	73	80	82	**99**	93	71	93	87
	CTS-CNN [34]	**99**	92	93	78	94	67	85	73	87	88	95	92	73	79	82	96	93	73	91	86
	MCNN-AUX [33]	98	**94**	**97**	87	96	76	**97**	77	**94**	95	**98**	93	84	84	90	**99**	**94**	87	**97**	88
	Proposed	98	**94**	**97**	**90**	**97**	**78**	**97**	**78**	**94**	**96**	**98**	**94**	**85**	**87**	**91**	**99**	93	**89**	**97**	**90**
LFWA	FaceTracker [27]	84	77	83	73	69	66	70	74	63	70	71	78	67	62	88	75	87	81	71	80
	PANDA [57]	92	78	87	73	75	72	84	76	84	73	76	89	73	75	92	82	93	86	79	82
	LNets+ANet [23]	**94**	82	92	81	79	74	84	80	85	78	77	91	76	76	94	88	**95**	88	79	86
	CTS-CNN [34]	**94**	81	94	81	80	75	73	83	**86**	82	82	90	77	77	94	90	**95**	90	**81**	86
	MCNN-AUX [33]	**94**	84	93	**83**	**82**	**77**	**93**	**84**	**86**	**88**	**83**	**92**	**79**	**82**	**95**	90	**95**	90	**81**	86
	Proposed	93	**86**	**95**	82	81	75	91	**84**	85	86	80	**92**	**79**	80	94	**92**	93	**91**	**81**	**87**

中科院计算所

Institute of Computing Technology, Chinese Academy of Sciences

Deep Multi-task Learning

- MTL vs. STL on 9 common attributes in CelebA



Deep Multi-task Learning

- Generalization ability to single attribute (ChaLearn2016 FotW database)

Method	Hat	Headband	Glasses	Earrings	Necklace	Tie	Scarf	Avg.
SIAT MMLAB	94.7	94.9	94.7	91.0	88.2	97.3	93.7	93.5
IVA NLPR	92.2	95.1	93.9	85.3	87.4	96.1	94.0	92.0
Proposed	94.7	96.1	96.1	89.1	89.5	97.4	95.1	94.0

(a) FotW - accessory classification

Method	Smile	Gender	Avg.
SIAT MMLAB	92.7	85.8	89.3
IVA NLPR	91.5	82.5	87.0
VISI.CRIM	90.2	82.1	86.1
SMILELAB NEU	90.0	81.5	85.7
Proposed	84.9	87.3	86.1

(b) FotW - smile and gender classification

Deep Multi-task Learning

- Cross-database testing
  - Cross-database testing could provide insights of the system's performance under real application scenarios
  - We have called on the use of cross-database testing on several problems, including
    - Attribute learning [Han TPAMI 2015, Han TPAMI 2017]
    - Face liveness detection [Wen TIFS 2014, Patel TIFS 2016]

Deep Multi-task Learning

- Cross-database testing

Database		Accuracy		
Training	Testing	Age[1]	Gender	Race
MORPH II	MORPH II	3.0/85.3	98.0	98.6
LFW+	MORPH II	7.0/60.1	89.0	85.7
LFW+	LFW+	4.5/75.0	96.7	94.9
MORPH II	LFW+	9.4/52.6	77.4	70.5
		Avg. accuracy of 40 attributes		
CelebA	CelebA	93.0		
LFWA	CelebA	70.2		
LFWA	LFWA	86.0		
CelebA	LFWA	73.0		

中科院计算所

Institute of Computing Technology, Chinese Academy of Sciences

Outline

- Background
- Related work
- Attribute learning via STL
- Attribute learning via MTL
- Conclusion and discussion
- Data, demo, etc.

Conclusion and discussion

- The performance of attribute learning has also been improved significantly, benefited from deep learning methods
- Modeling attribute correlation and heterogeneity via MTL is an efficient way to handle a large number of visual attribute
- Unsolved
  - Attribute learning from incompletely data [Chang AAAI17]
  - Attribute learning from noisy data
  - …

Outline

- Background
- Related work
- Attribute learning via STL
- Attribute learning via MTL
- Conclusion and discussion
- Data, demo, etc.

Data, demo, etc.

- LFW+ dataset
  - Extend LFW with 2,466 unconstrained face images of subjects in age range 0 – 20
  - Age, gender, and race labels of each image provided by MTurk workers: http://biometrics.cse.msu.edu/pub/databases.html
- The human age estimates for FG-NET
  - Apparent age for FG-NET, provided by MTurk workers: http://www.cse.msu.edu/rgroups/biometrics/pubs/databases.html

Data, demo, etc.

- Demo

Attribute learning from face

Heart rate estimation from face

Ground-truth

http://ddl.escience.cn/f/Ndme

http://ddl.escience.cn/f/Ndme

Xuesong Niu, el al., Continuous Heart Rate Measurement from Face: A Robust rPPG Approach with Distribution Learning, IJCB, 2017.10

References

- **H. Han**, A. K. Jain, S. Shan, and X. Chen. "Heterogeneous Face Attribute Estimation: A Deep Multi-Task Learning Approach," To appear in *IEEE Trans. Pattern Analysis and Machine Intelligence (T-PAMI)*, pp. 1-14, 2017. (CCF-A, IF: 8.3) [arXiv:1706.00906, DOI: 10.1109/TPAMI.2017.2738004]
- **H. Han**, C. Otto, X. Liu, and A. K. Jain. "Demographic Estimation from Face Images: Human vs. Machine Performance," *IEEE Trans. Pattern Analysis and Machine Intelligence (T-PAMI)*, vol. 37, no. 6, pp. 1148-1161, Jun. 2015. (CCF-A, IF: 8.3, GS: 80+ citations)
- F. Wang, **H. Han**, S. Shan, and X. Chen. "Deep Multi-Task Learning for Joint Prediction of Heterogeneous Face Attributes," in Proc. IEEE FG, May 2017.(CCF-C)
- **H. Han**, C. Otto and A. K. Jain. "Age Estimation from Face Images: Human vs. Machine Performance," in Proc. ICB, 2013. (Oral, CCF-C, GS: 100+ citations)
- **H. Han** and A. K. Jain, "Age, Gender and Race Estimation from Unconstrained Face Images," MSU Technical Report, MSU-CSE-14-5, 2014. (GS: 31 citations)
- LFW+ dataset: http://biometrics.cse.msu.edu/pub/databases.html
- Demo: DMTL-FaceAttribute (http://ddl.escience.cn/f/FOrq), rPPG-HeartRate (http://ddl.escience.cn/f/Ndme)

中科院计算所

Institute of Computing Technology, Chinese Academy of Sciences

中科院计算所
Institute of Computing Technology, Chinese Academy of Sciences

Thank You!