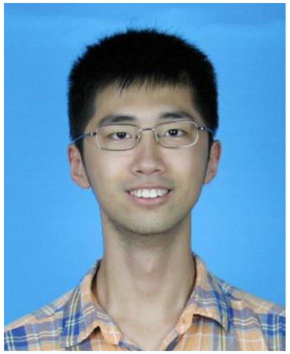
IEEE FG 2018 Tutorial on

Representation Learning for Face Alignment and Recognition

Tutors: Hao Liu, Yueqi Duan, and Jiwen Lu



Outline

- Part 1: Introduction
- Part 2: Representation Learning for Face Alignment

------------------- Short Break: 30 minutes -------------------

- Part 3: Representation Learning for Face Recognition
- Part 4: Discussions

Part 1: Introduction

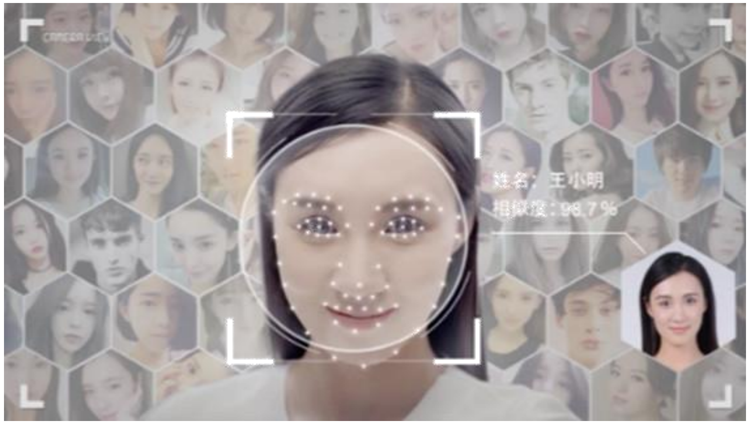
Why Understanding Human Faces

Biometrics (visual authorization and identification)

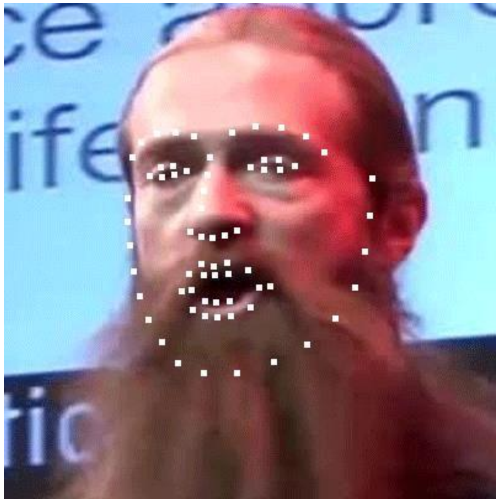
Computer Control (visual surveillance and multimedia)

Face Analysis Tasks

- Face Detection (control/surveillance)



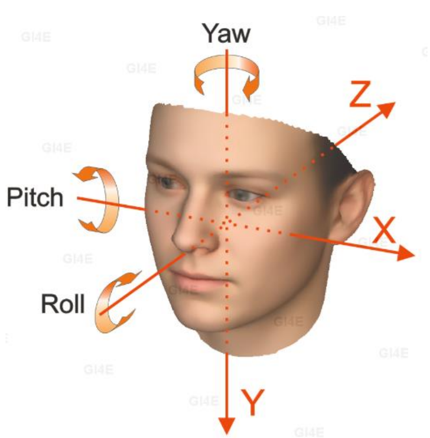
- Face Tracking (control/surveillance)

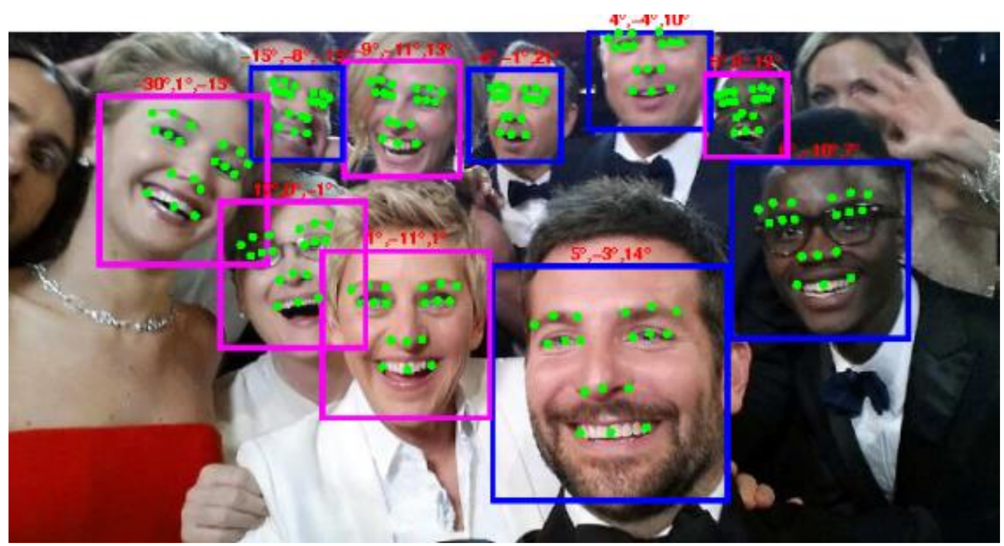



Face (Attribute) Analysis Tasks

- Expression Analysis (human-computer interaction)

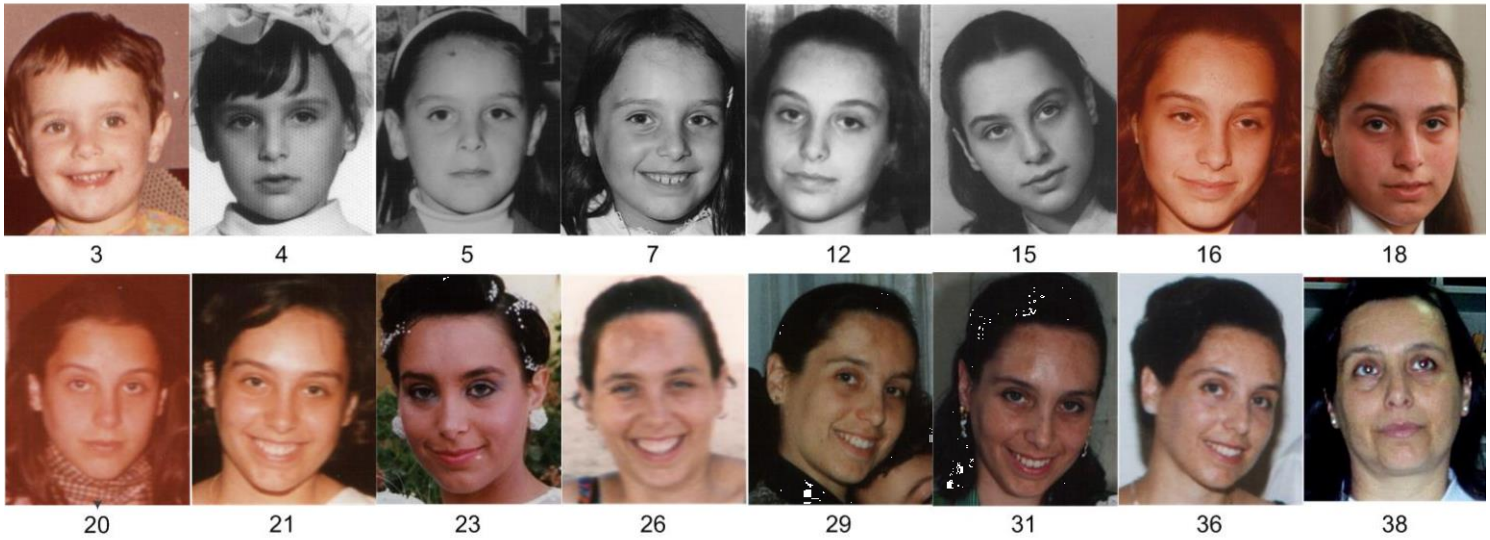
- Facial Pose Estimation (human-computer interaction)



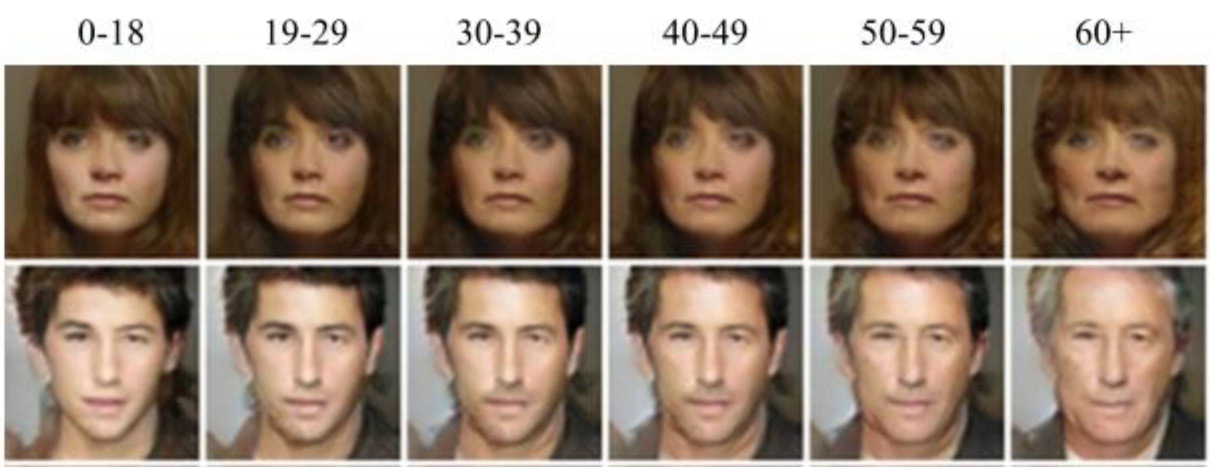


Face (Attribute) Analysis Tasks

Facial age estimation (visual advertisement)



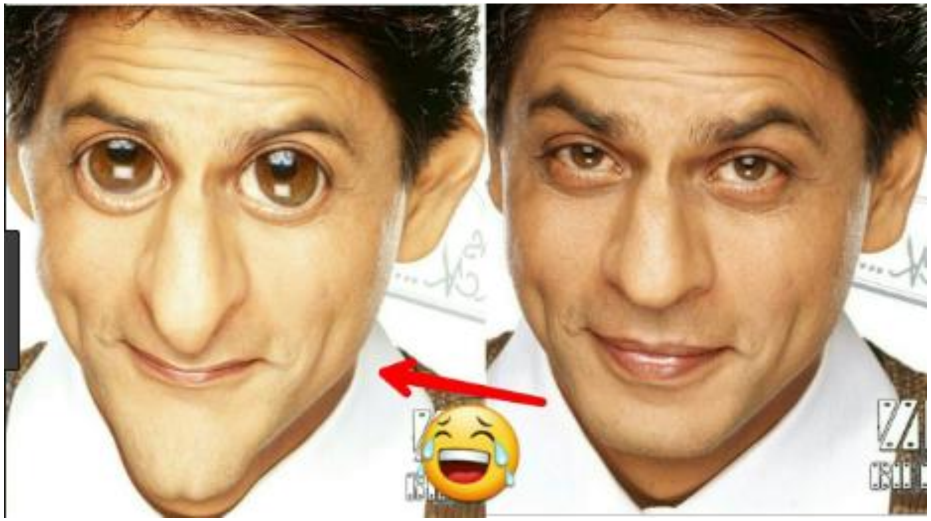
Face Aging (finding lost children, visual advertisement)



Face Analysis Tasks

- Face Reconstruction (visual animation)

- Face Editing (visual entertainment)



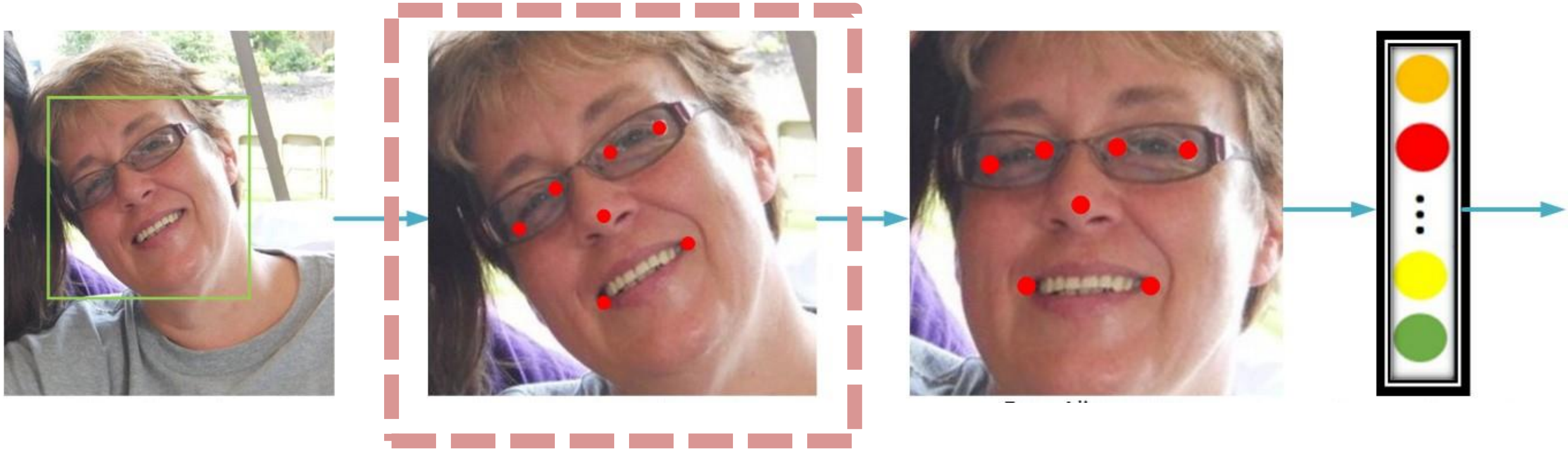
Face Recognition

Face identification (1 : n)

Face verification (1 : 1)

Face Analysis Pipeline

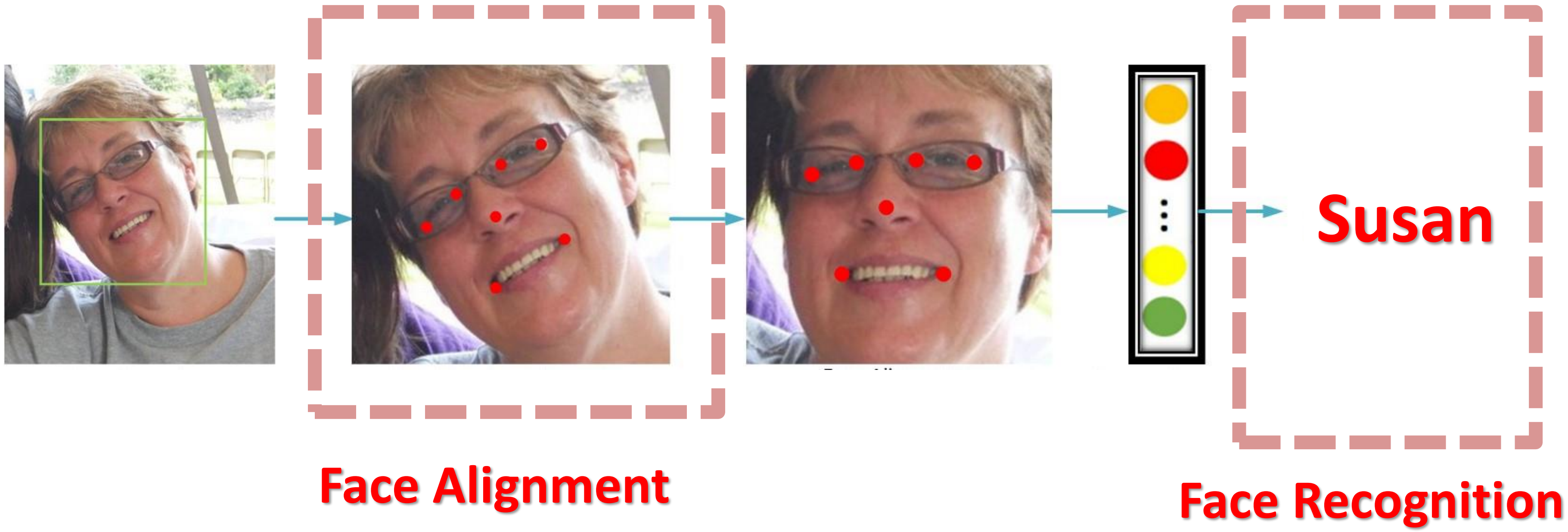
- Face Detection
- Face Alignment
- Face Feature Extraction
- Face Analysis (e.g. Face Recognition)



- **expression recognition**
- **pose estimation**
- **reconstruction**

Face Recognition Pipeline

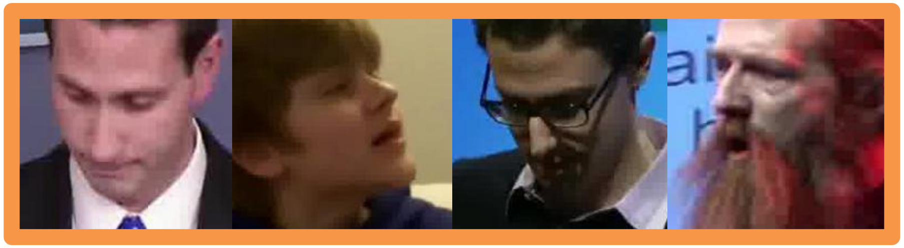
- Face Detection
- Face Alignment (Facial Landmark Localization)
- Face Feature Extraction
- Face Recognition



Challenges

☐ Face Alignment

- Lightness\low-resolution
- Large poses
- Facial expressions
- Partial occlusions
- Face changes/ motion



Large Pose

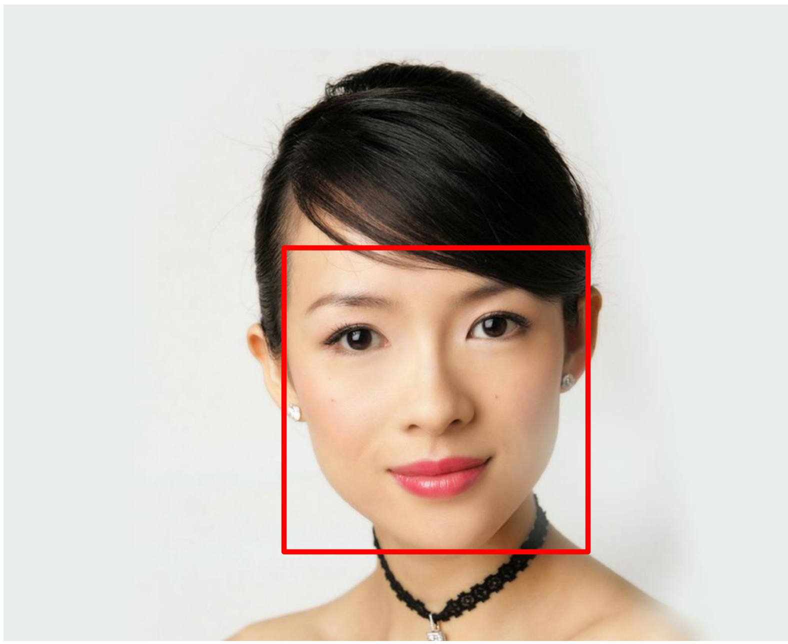
Various Expressions

Partial Occlusions

Challenges

Face Recognition

- High-dimensional data



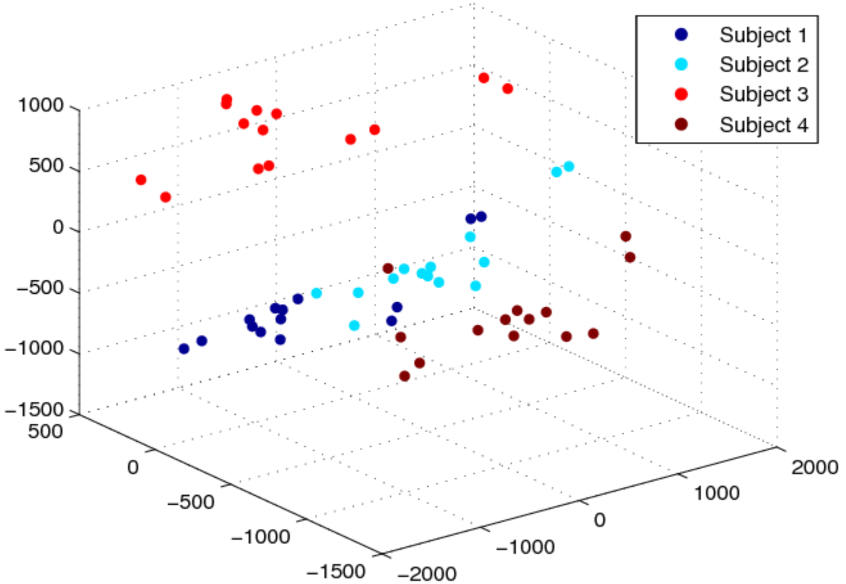
Feature vector

- Deteriorate the performances of classifiers
- High computational complexity

Challenges

Face Recognition

- Large intra-class variances



Conventional Solutions

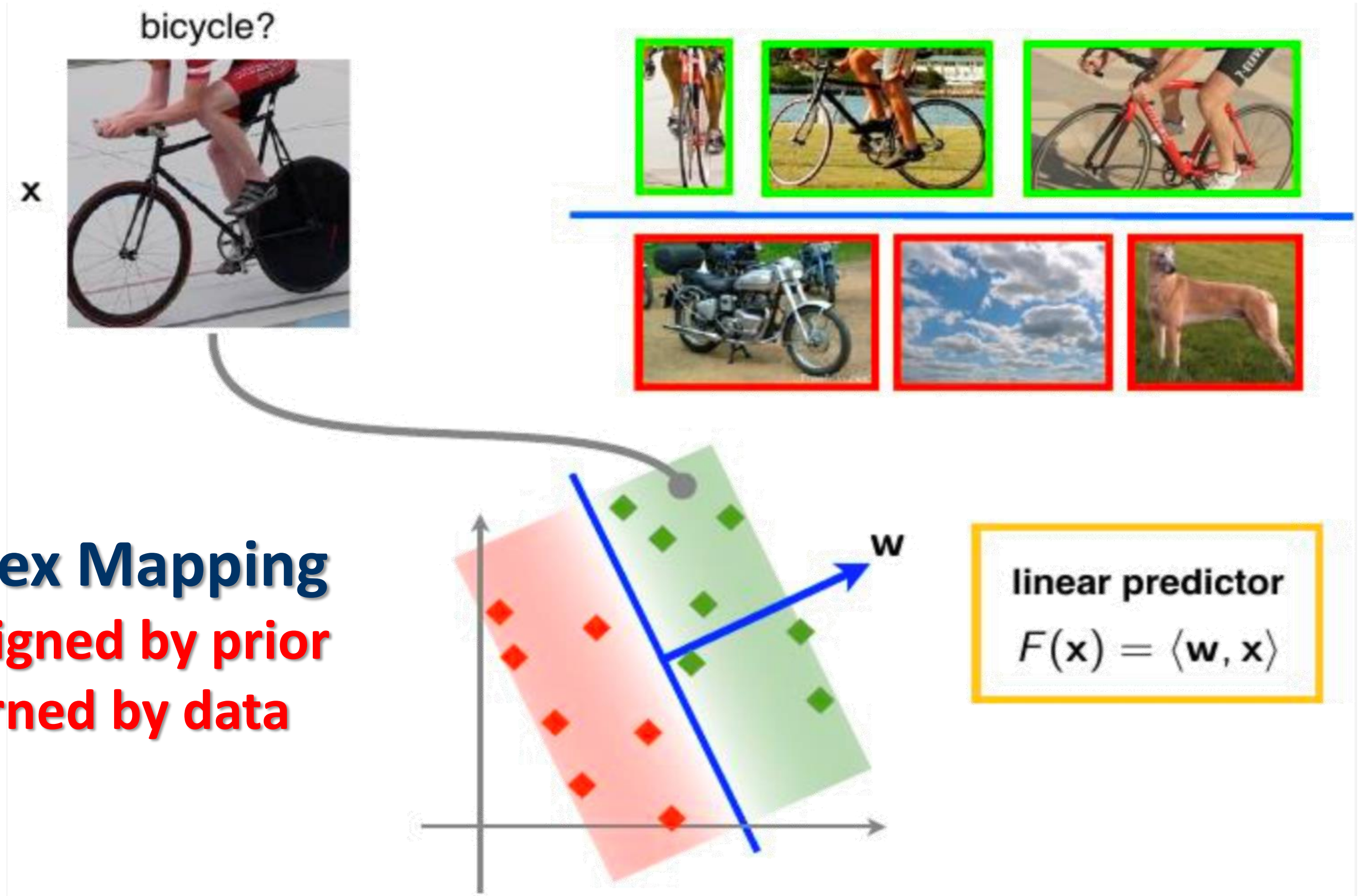
Feature Extraction

- LBP/TPLBP/FPLBP/CSLBP
- Gabor/LGBP/HGPP
- HOG
- SIFT
- POEM
- LE

Model Learning

- Support Vector Machine
- Manifold Learning
- Metric Learning
- Active Learning
- Random Forests
- Neural Networks
- Cascade Regression

Feature Representation for Classification



- **Complex Mapping**
  - **Designed by prior**
  - **Learned by data**

Designed / Learnable Feature Representation

Hand-Designed Representation (Rule-based)

Designed Feature (SIFT, HOG, etc)

Adaboost, SVM,etc.

Tiger

Data-Driven Representation (Learning-based)

Learnable Features **W**

SVM or End-to-End

Tiger

Jointly optimize representation learning and classification/regression

Representations Matter

Robust, compact and informative representation.

- Hand-crafted
- Learning based

Hand-crafted

- LBP/TPLBP/FPLBP/CSLBP
- Gabor/LGBP/HGPP
- HOG
- SIFT
- POEM

Learning based

- Eigenface/Fisherface
- LE
- CBFD/CA-LBFL/SLBFLE
- CNN
- GAN

Part 2: Representation Learning for Face Alignment

Face Alignment in a Nutshell

- Input： Image pixels
- Output： Facial landmarks
- Point distribution model

$$\mathbf{S} = [p_1, p_2, \cdots, p_l, \cdots, p_L] \in \mathbb{R}^{2L}$$

- Objective：

$$J = \|\hat{\mathbf{S}} - \mathbf{S}^*\|$$

Shape Prior

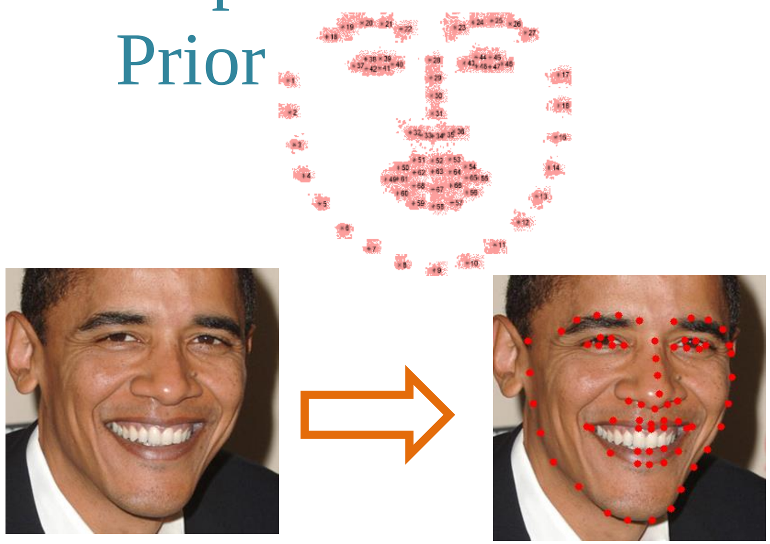
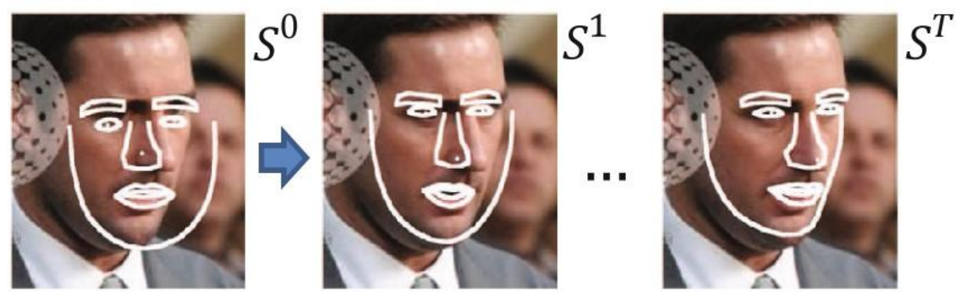


image coordinates

Existing Works

- Model-based Optimization
  - PCA shape model
  - holistic and local appearance
  - active shape and appearance fitting
- Cascaded Shape Regression
  - shape refinement
  - shape-index features
  - cascaded/coarse-to-fine



- ASM [Coots et al., CVIU 1995]
- AAM [Coots et al., PAMI 2004]
- CLM [Coots et al., BMVC 2006]

- ESR, [Cao et al., CVPR 2012]
- SDM, [Xiong et al., CVPR 2013]
- CFSS, [Zhu et al., CVPR 2015]

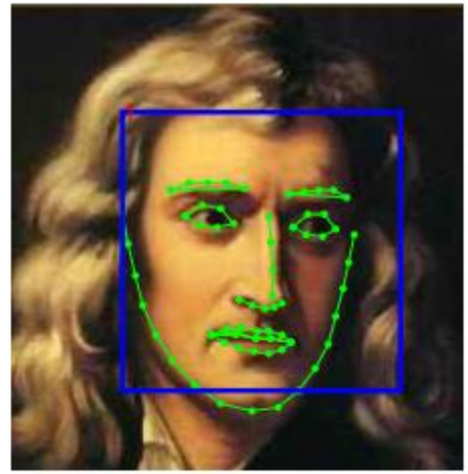
Key Points for Alignment Representation

- Hand-crafted Representation
  - ✓ HOG, SIFT, geometric-based (2D-3D projection)
- Shape-informative Representation
  - ✓ Local and global → Structural Learning
  - ✓ Robustness → Hierarchical Learning
- Knowledge-sharable Representation
  - ✓ Correlated Attributes → Multi-task Learning
  - ✓ Video-based → Spatial-temporal modeling

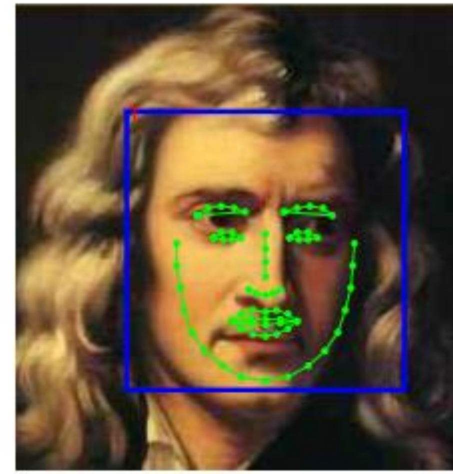
2.1 Hand-Crafted Representation

2D Hand-Crafted Representation

- Supervised Descent Model
  - texture-based representatior (SIFT, HOG, etc.)
  - cascaded linear regression

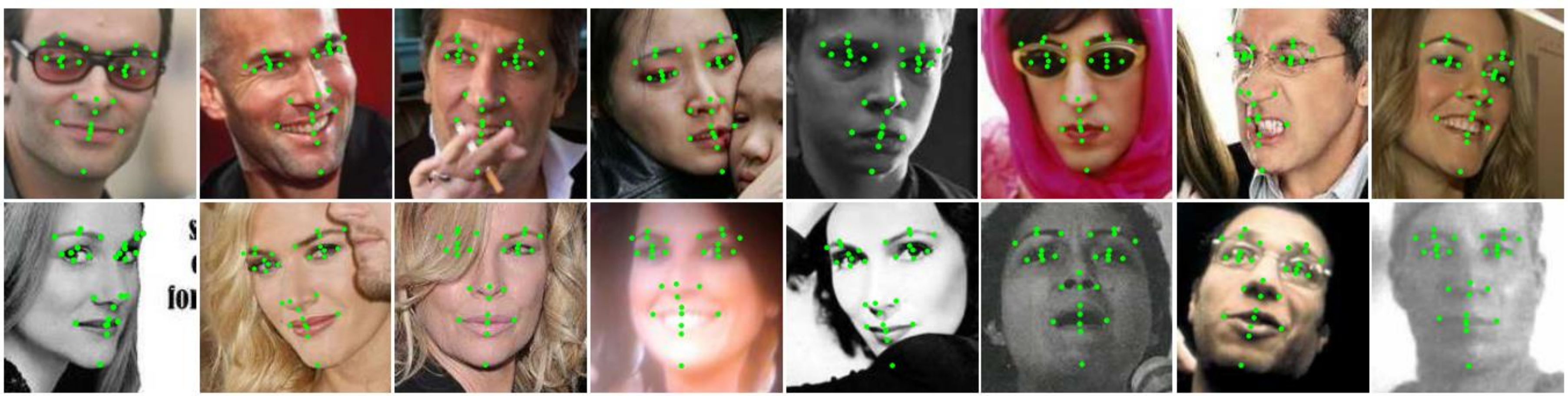


Final prediction



Initialized shape

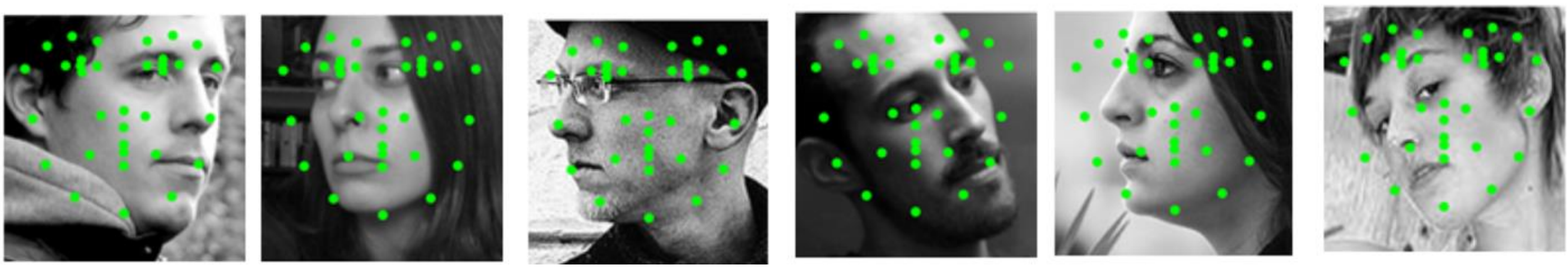
Example results on LFPW dataset



[Xiong et al., CVPR’13]

2D Hand-Crafted Representation

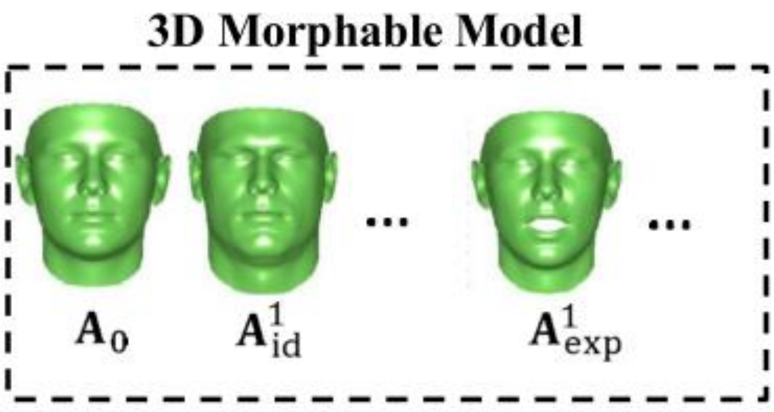
Performance degrades largely for large-pose faces!!!

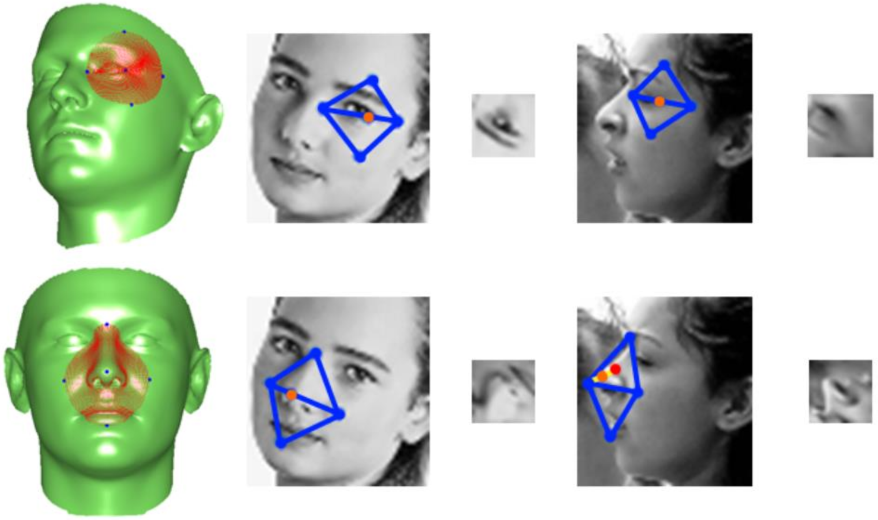


- Solution
  - Introduce 3D geometric information
  - Preserve spatial structure

2D-3D Hand-Crafted Representation

- Basic Idea
  - 3D surface (geometric)
  - 2D-3D projection



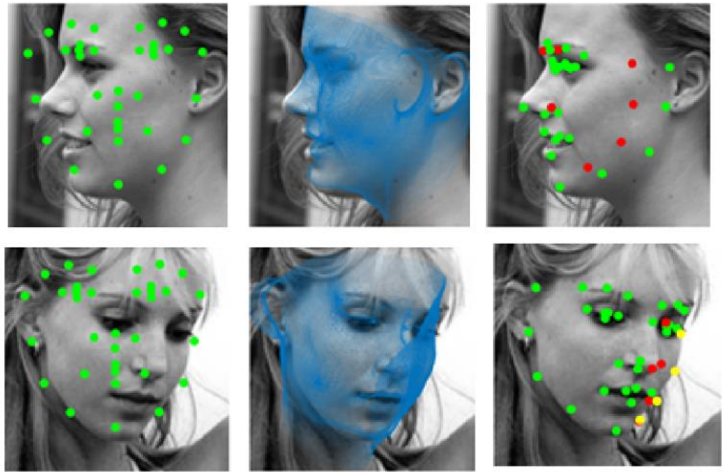


[Liu et al., CVPR'16]

2D-3D Hand-Crafted Representation

- Basic Idea
  - large-pose faces
  - profile (self-occlusion)
  - Solution: 3D fitting

[Liu et al., CVPR’16]



The result of the proposed method across stages, with the extracted features (1st row) and alignment results (2nd row).

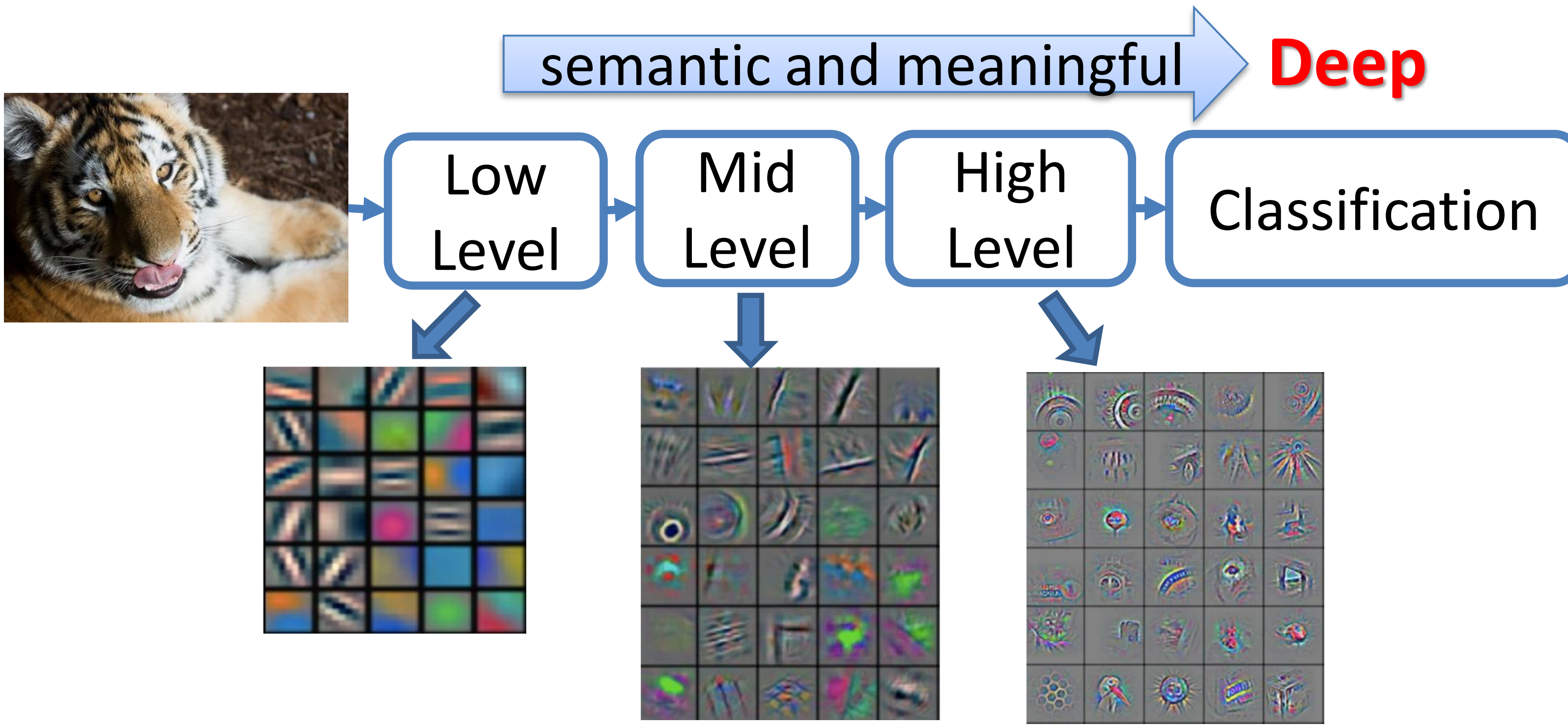
Limitations of Hand-Crafted Representation

- Require strong prior knowledge
- May not work in *Domain Adaptation*
- Separated learning leads to local optima

2.2 Hierarchical Representation Learning

Hierarchical Representation: Deeply Learned Features

- Visual Recognition
  - pixel→edge→texture→pattern→component→object



Hierarchical Representation: Deeply Learned Features

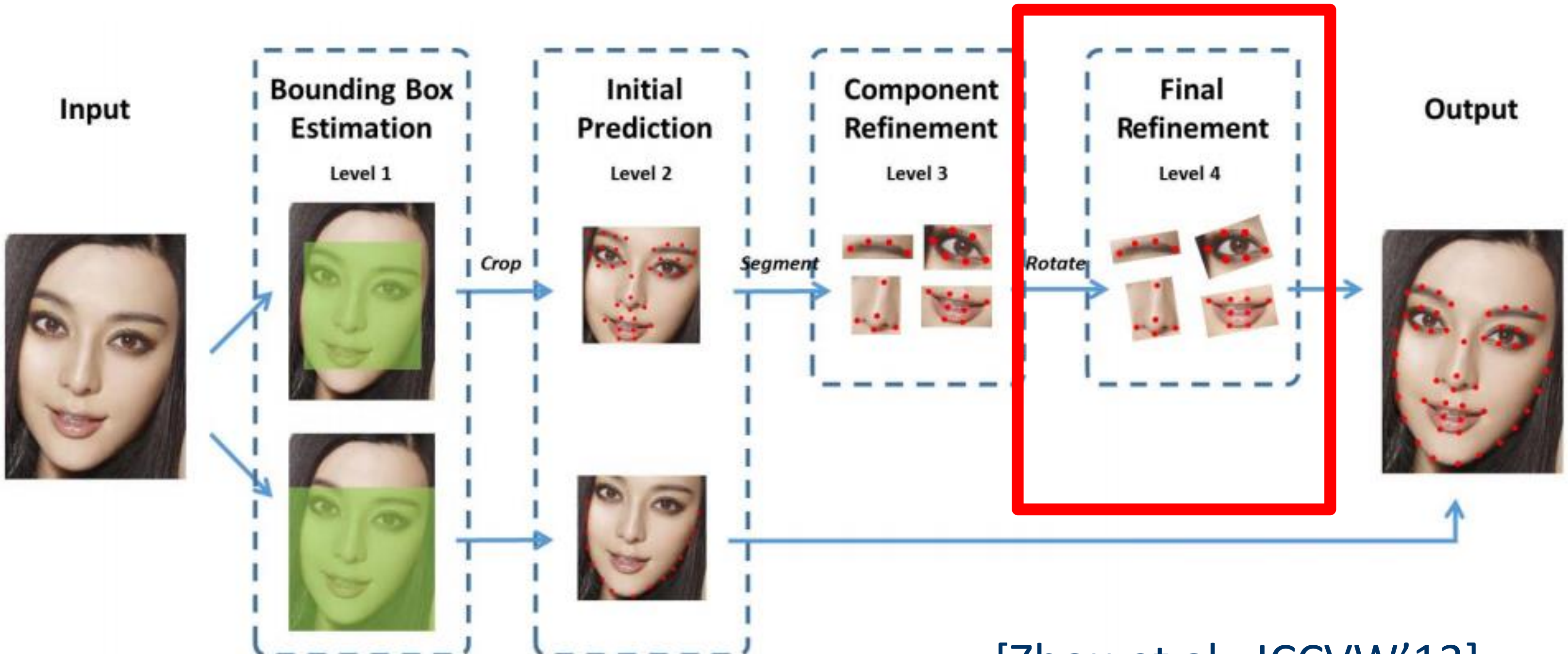
- Visual Recognition
  - pixel→edge→texture→pattern→component→object

semantic and meaningful → **Deep**

→ Low Level → Mid Level → High Level → Classification

- Text Classification
  - character→word→word group→clause→sentense→story

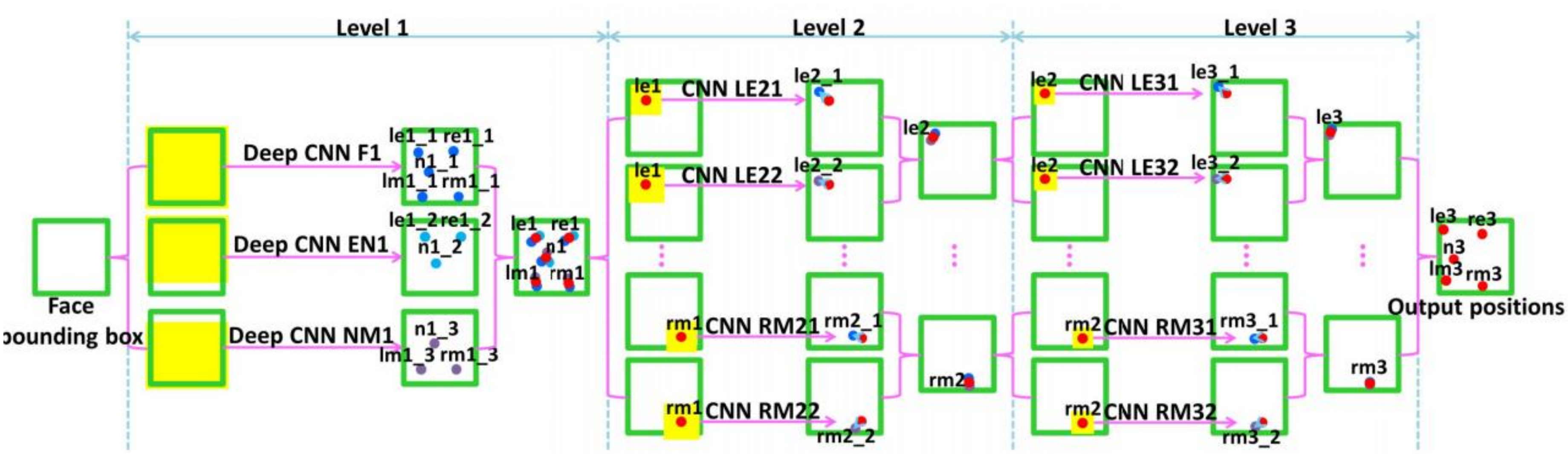
Learning Hierarchical Representation by Facial Landmarks Partition



[Zhou et al., ICCVW'13]

- Basic Idea
  - Face partition based on different facial components
  - Learning local features hierarchically by a set of CNNs

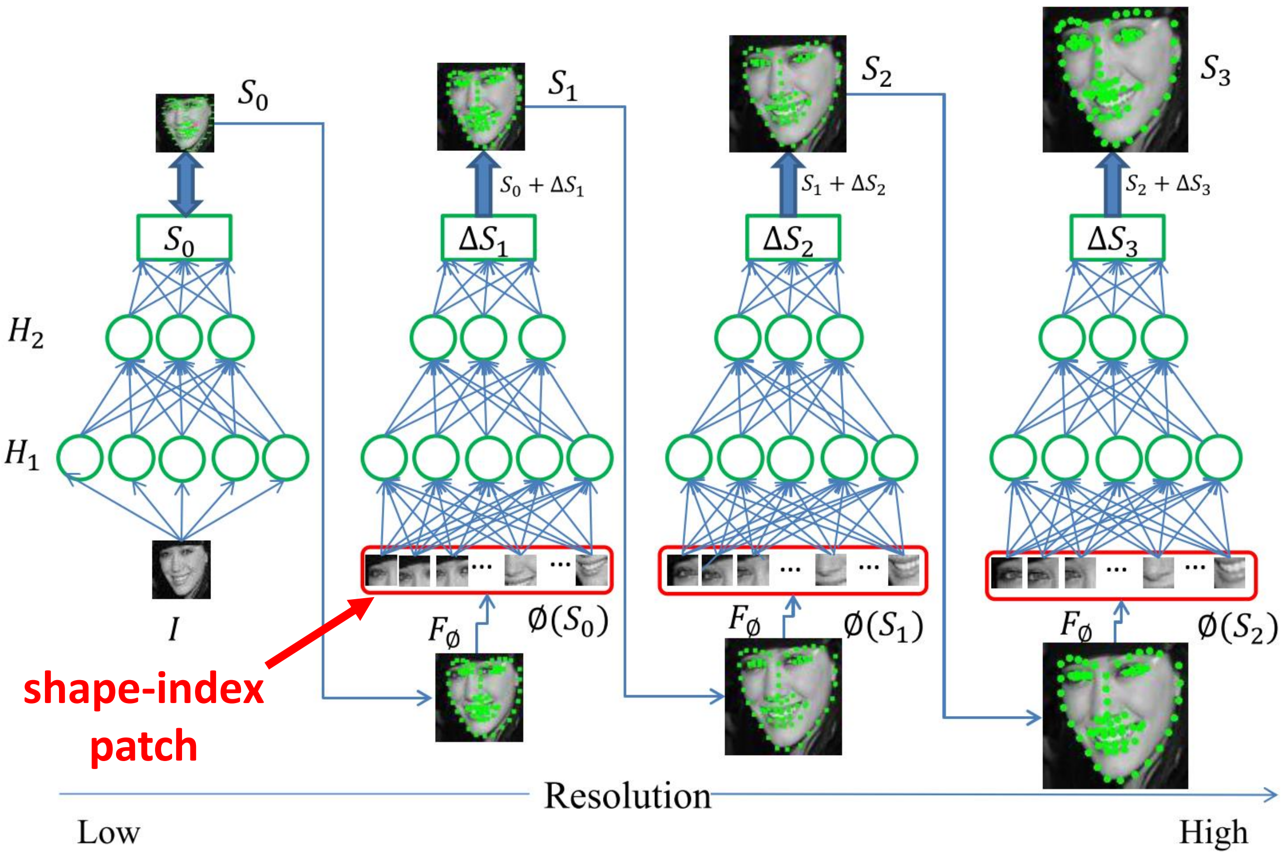
Learning Hierarchical Representation by Convolutional Networks



[Sun et al., CVPR'13]

- Level 1
  - image->landmark
  - rough prediction
- Level 2 & 3
  - landmark update
  - Coarse-to-fine
- Limitations
  - Correlation of landmarks and neighbors
  - Global shape constraint

Learning Hierarchical Representation by Auto-Encoder Networks

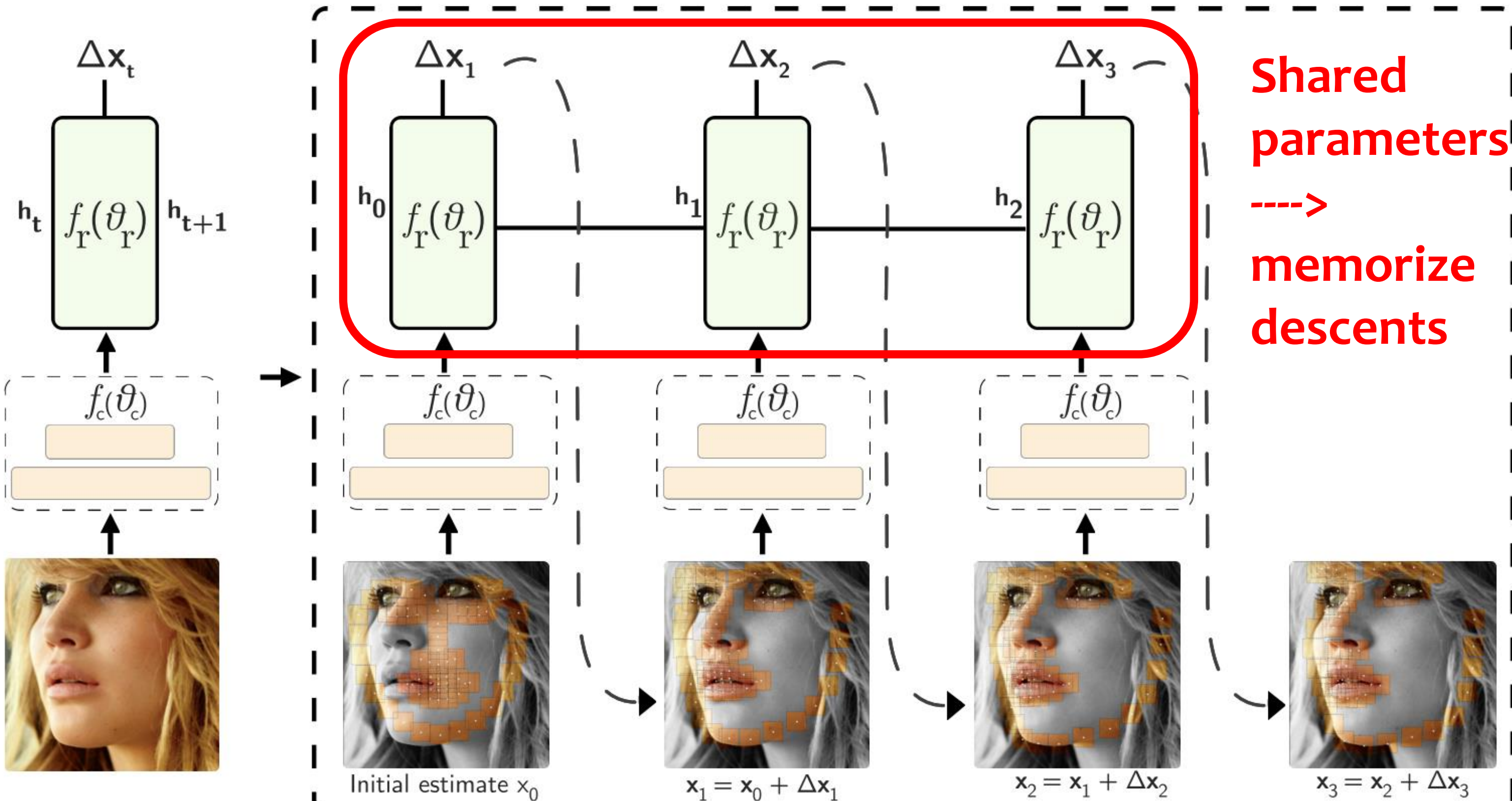


[Zhang et al., ECCV'14]

- Shape Initialization
  - image->landmark
  - rough prediction
- Shape Refinement
  - landmark update
  - Coarse-to-fine
- Representation
  - Shape-index patches
  - Raw pixel input

Auto-encoder only for parameter initialization of deep neural networks! (Use CNN instead)

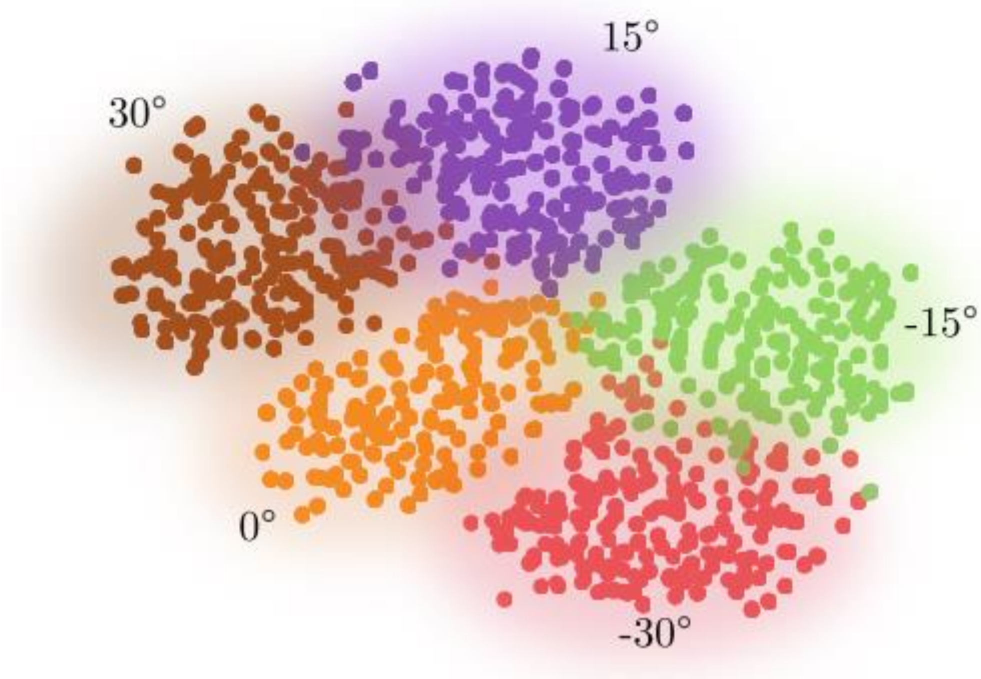
Learning Hierarchical Representation by Feedback Neural Networks



Shared parameters ----> memorize descents

[Trigeorgis et al., CVPR'16]

Learning Hierarchical Representation by Feedback Neural Networks



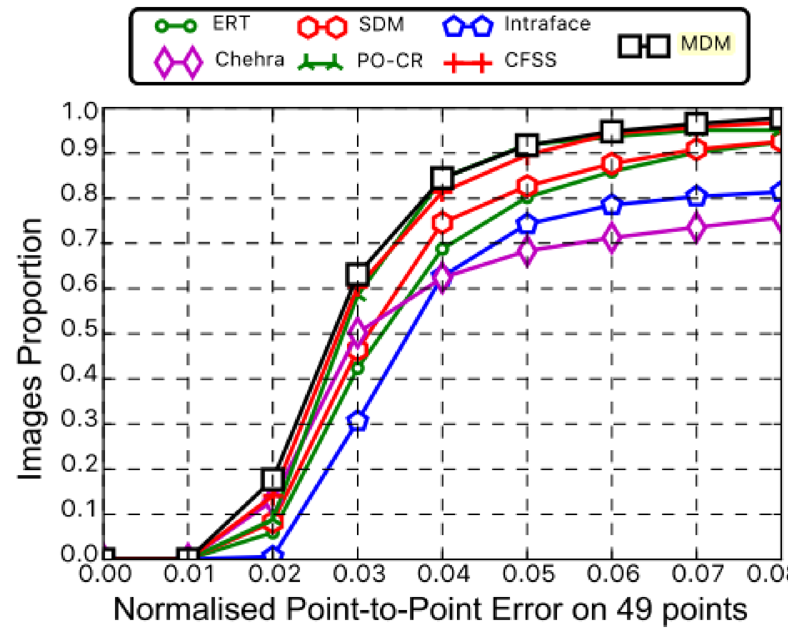


Figure 6: Results on the full testing set of the 300W competition, which was used as a validation set (49-points).

[Trigeorgis et al., CVPR'16]

- A t-SNE depiction of the internal states (T = 1)
- Each color corresponds to a cluster of head pose.
- MDM learn to partition the input data based on the head pose.

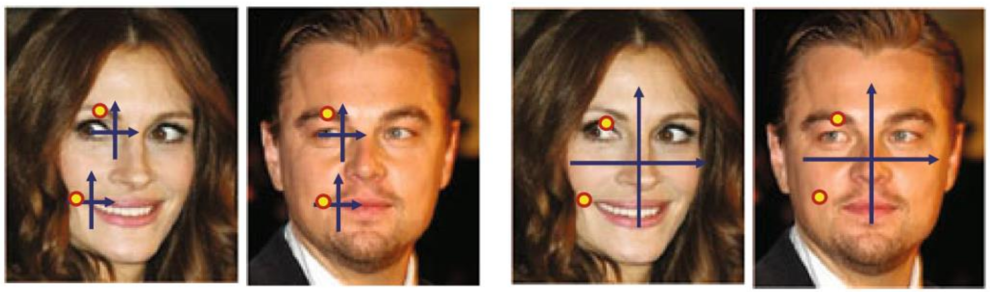
Improvements

- Explicit shape-index local features
- Coarse-to-fine facial shape constraint
- Correlation of neighbouring landmarks

2.3 Structural Representation Learning

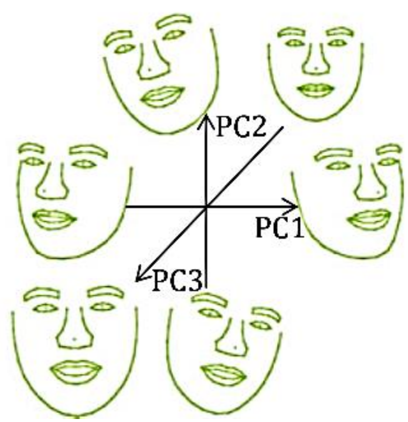
Shape-Index Representation

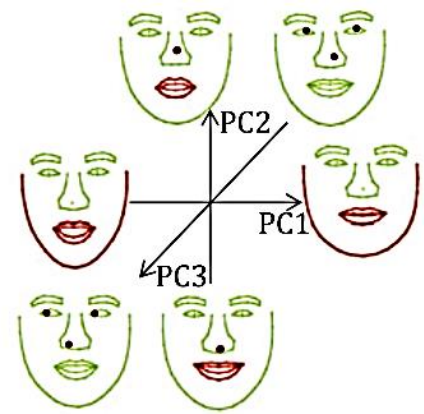
- Motivation
  - Exploit shape-sensitive structure (local and global)
- Explicit shape-index local feature



Benefit from the spatial locality.

- PC coefficients exploit different facial components

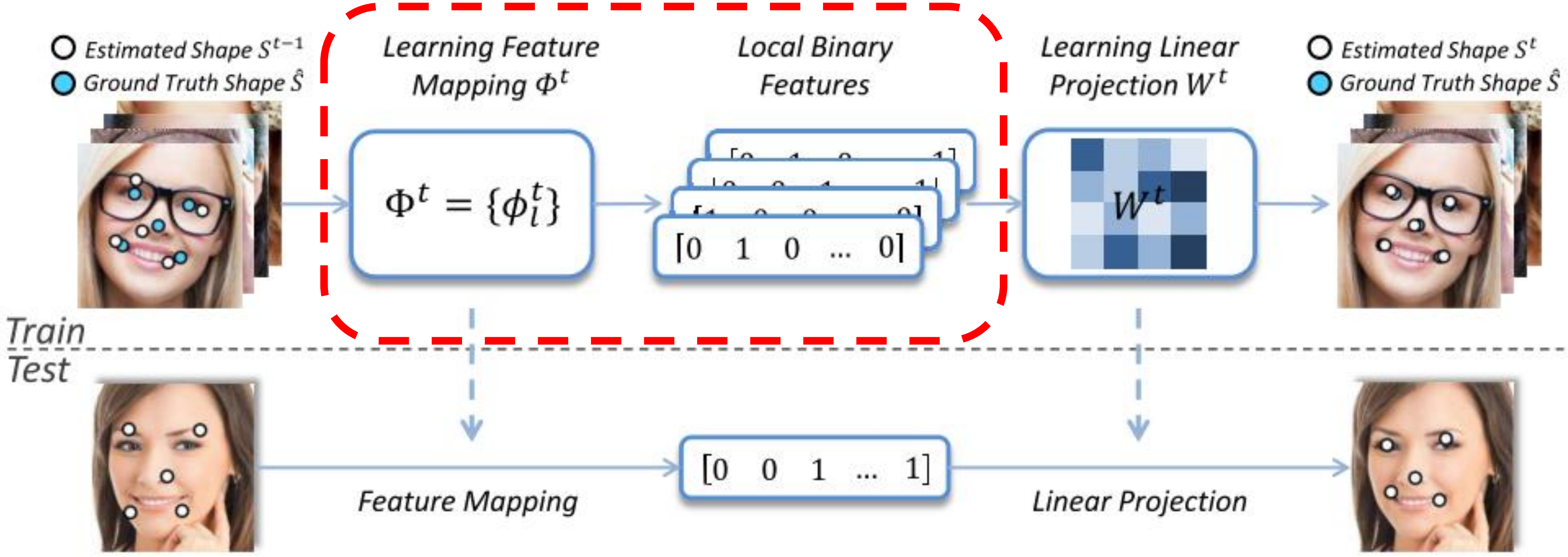




[Cao et al., CVPR'12]

Local Binary Representation

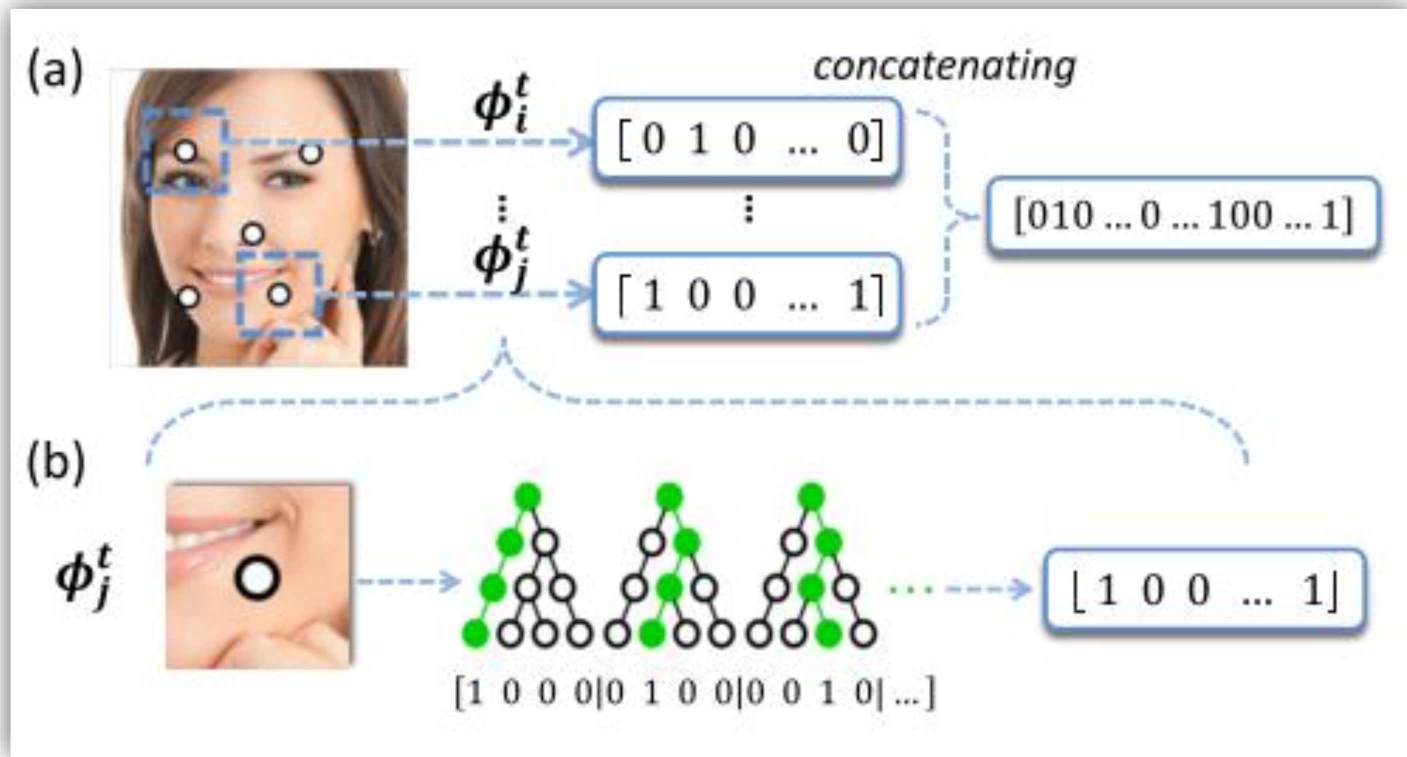
Framework



[Ren et al., CVPR'14]

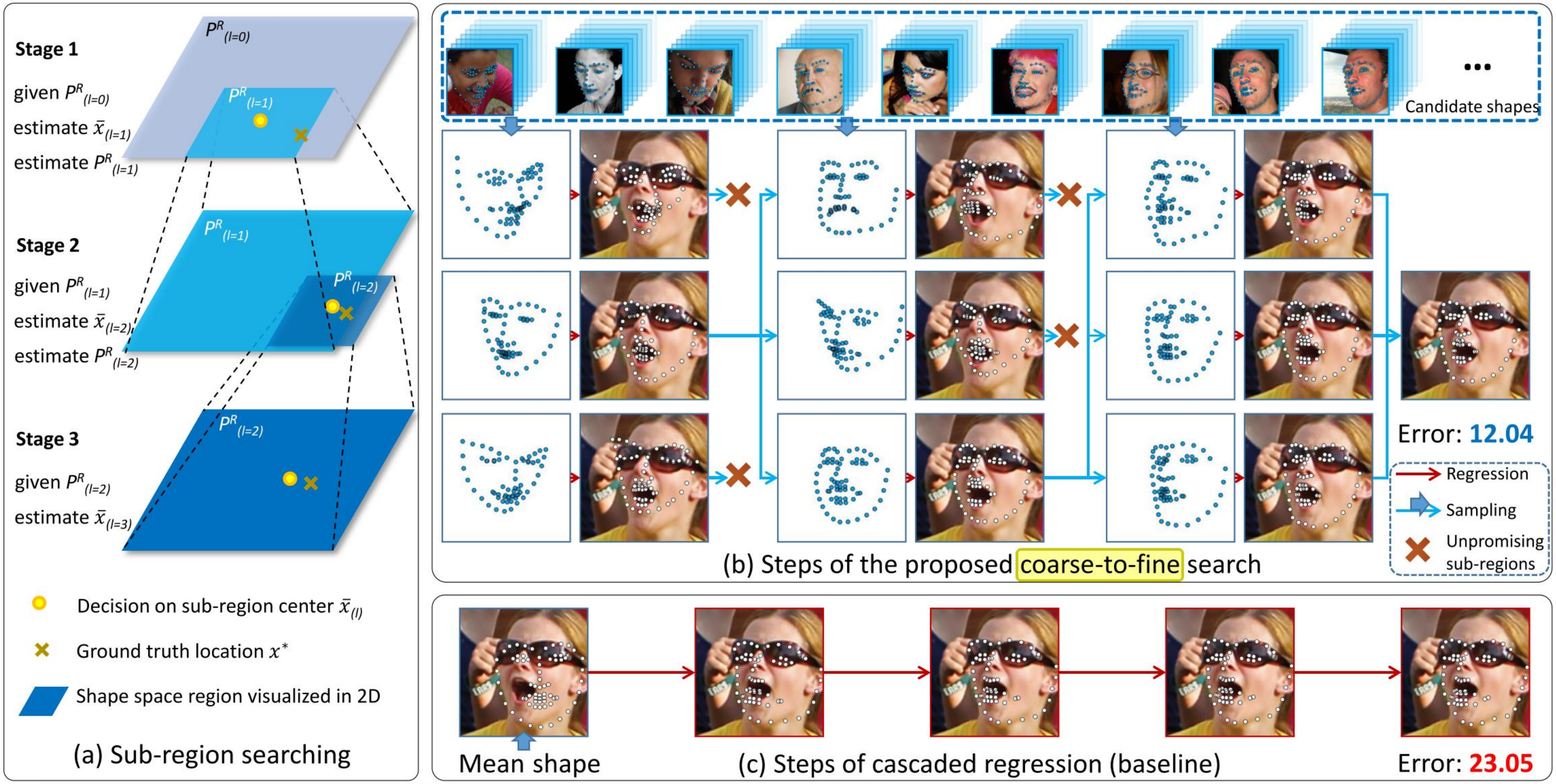
Local Binary Representation

Shape-index feature → local binary feature



[Ren et al., CVPR'14]

Structural Learning by Coarse-to-Fine Shape Searching



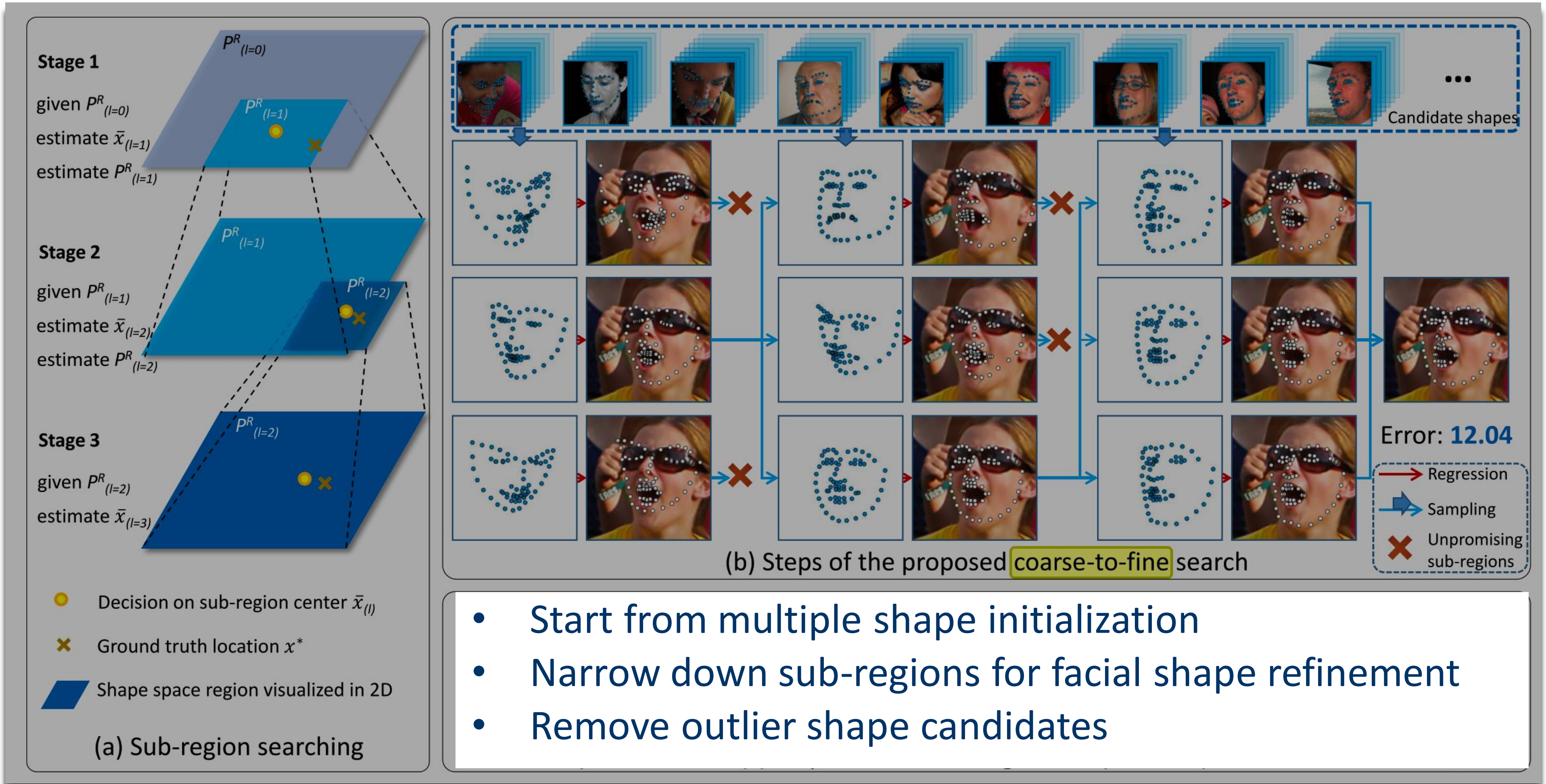
(a) Sub-region searching

(b) Steps of the proposed coarse-to-fine search

(c) Steps of cascaded regression (baseline)

[Zhu et al., CVPR'15]

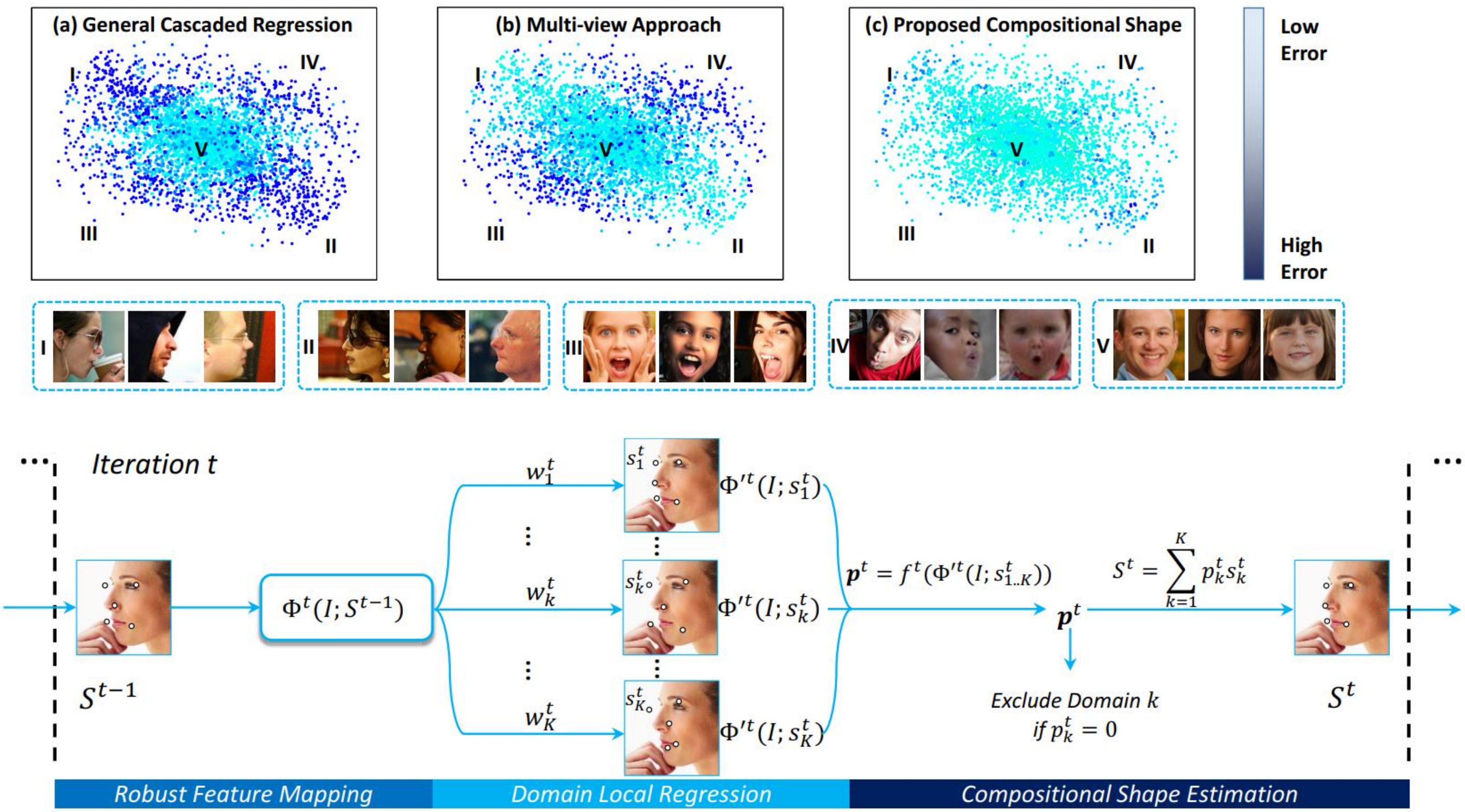
Structural Learning by Coarse-to-Fine Shape Searching



- Start from multiple shape initialization
- Narrow down sub-regions for facial shape refinement
- Remove outlier shape candidates

[Zhu et al., CVPR'15]

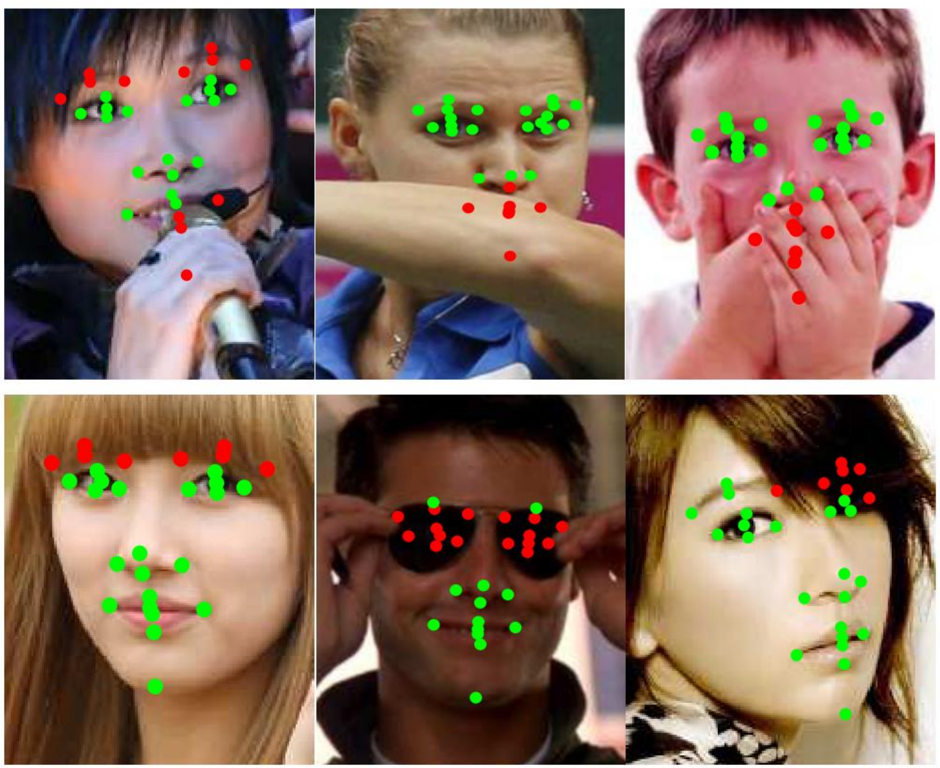
Structural Representation by Cascade Compositional Learning

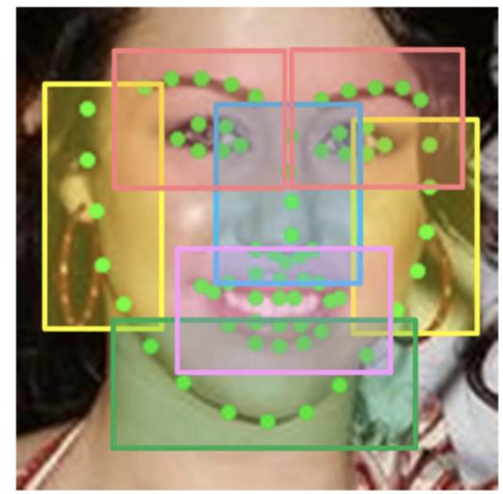


[Zhu et al., CVPR'16]

Learning Structural Representation for Robust Face Alignment

Real-world conditions present large variations in use of accessories such as sunglasses and hats and interactions with objects (e.g. food).



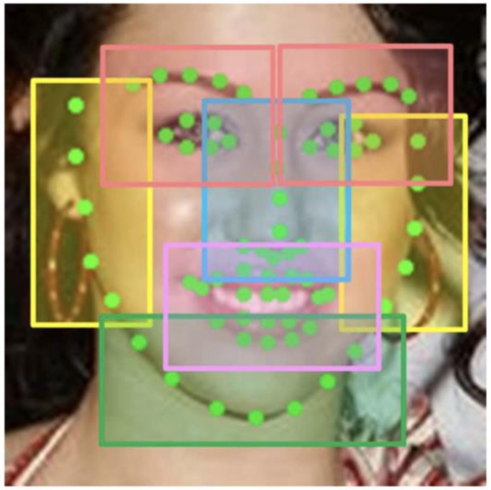


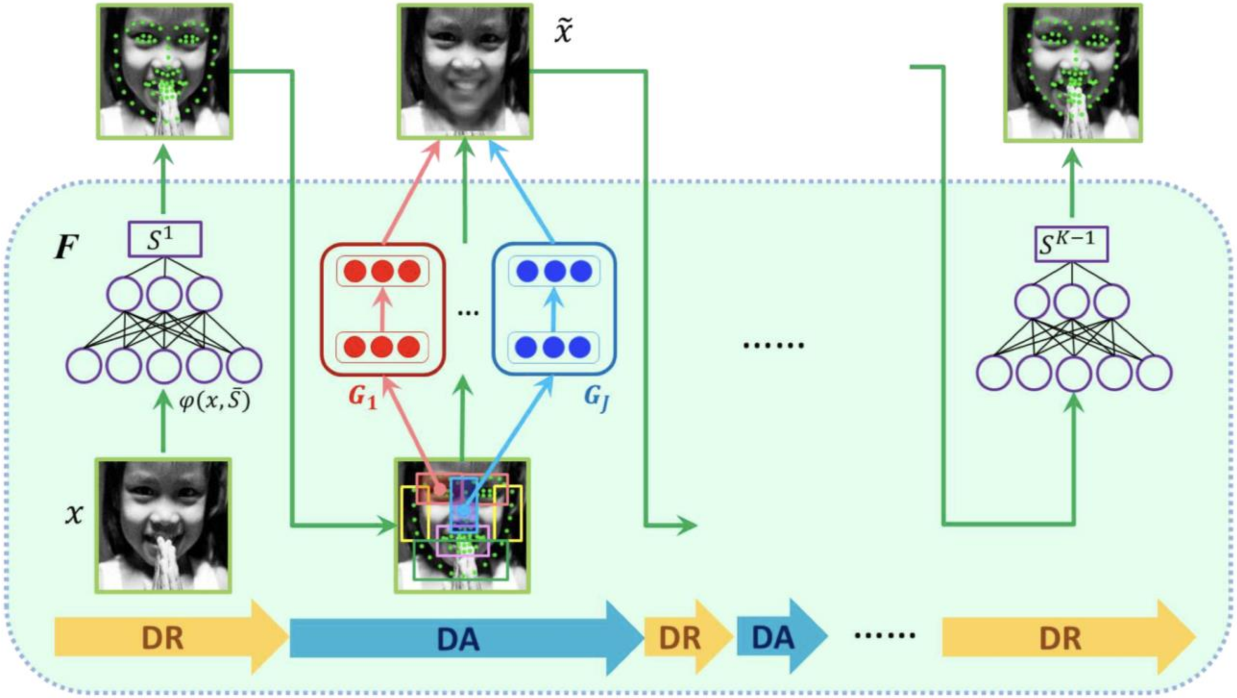
Face Partition

- Deformable Part-based Model
- Infer occluded part via non-occluded part

[Burgos-Artizzu et al., ICCV’13]

Learning Structural Representation for Robust Face Alignment





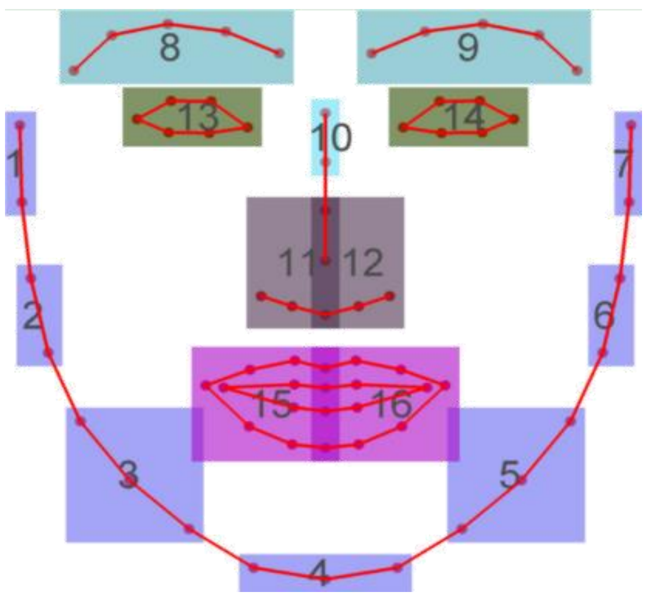
[Zhang et al., CVPR'16]

Main Idea:

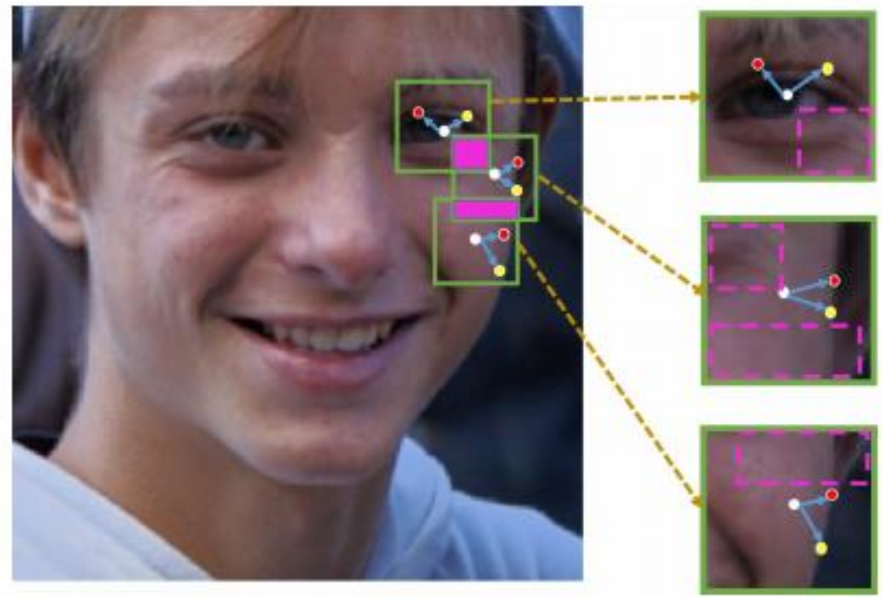
- Use auto-encoder network to recover the occluded-part
- Face alignment performs on the recovered face cascaded

Learning Deep Structural Representation

Motivation



Semantic Facial Parts

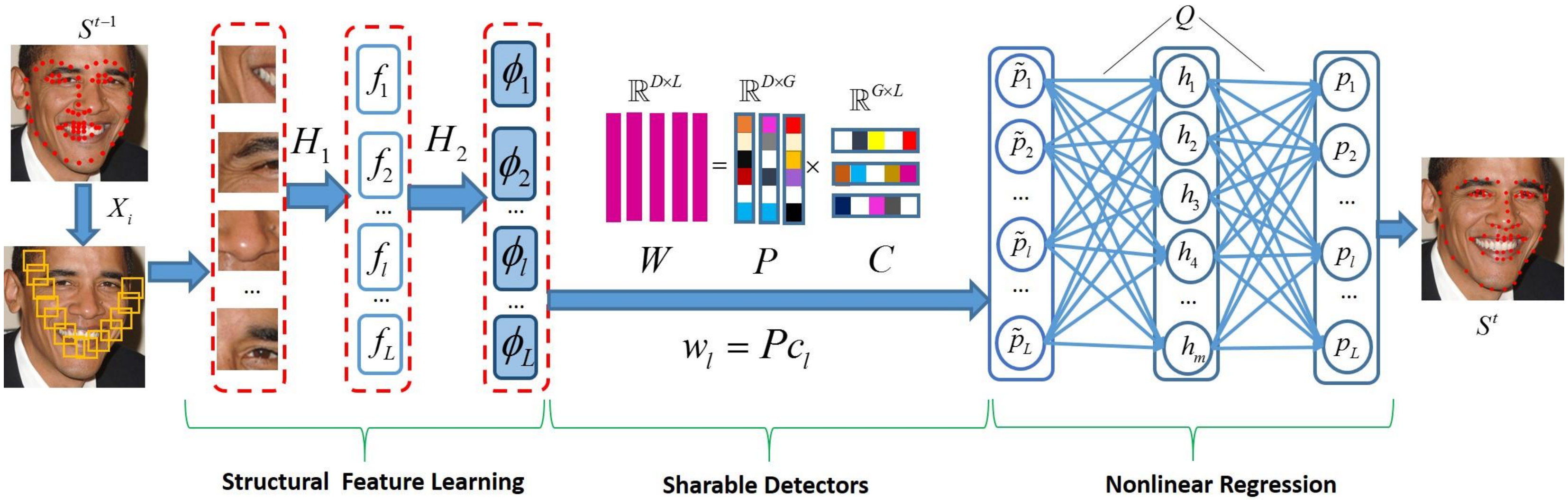


Structural learning from neighbouring landmarks

Hao Liu, **Jiwen Lu**, jianjiang feng and Jie Zhou, Learning deep sharable and structural detectors for face alignment, *IEEE Transactions on Image Processing (T-IP)*, 26(4):1666-1678, 2017.

Learning Deep Structural Representation

Basic Idea:



$$\min_{\{\mathbf{P},\mathbf{C},\mathbf{H},\mathbf{Q}\}} J = J_1(\mathbf{P},\mathbf{C},\mathbf{H},\mathbf{Q}) + J_2(\mathbf{P},\mathbf{C})$$

$$= \sum_{j}^{G}\sum_{i}^{N}\frac{1}{2}\left\|\mathbf{S}_i^* - \mathbf{S}_i^0 - \mathbf{Q}\left[(\mathbf{P}c_j)^T\mathbf{\Phi}_i\right]\right\|_2^2$$

$$+ (\gamma\,\|\mathbf{C}\|_1 + \beta\,\|\mathbf{P}\|_1)$$

Hao Liu, **Jiwen Lu**, jianjiang feng and Jie Zhou, Learning deep sharable and structural detectors for face alignment, *IEEE Transactions on Image Processing (T-IP)*, 26(4):1666-1678, 2017.

Experiments on Benchmark

Robustness to various poses

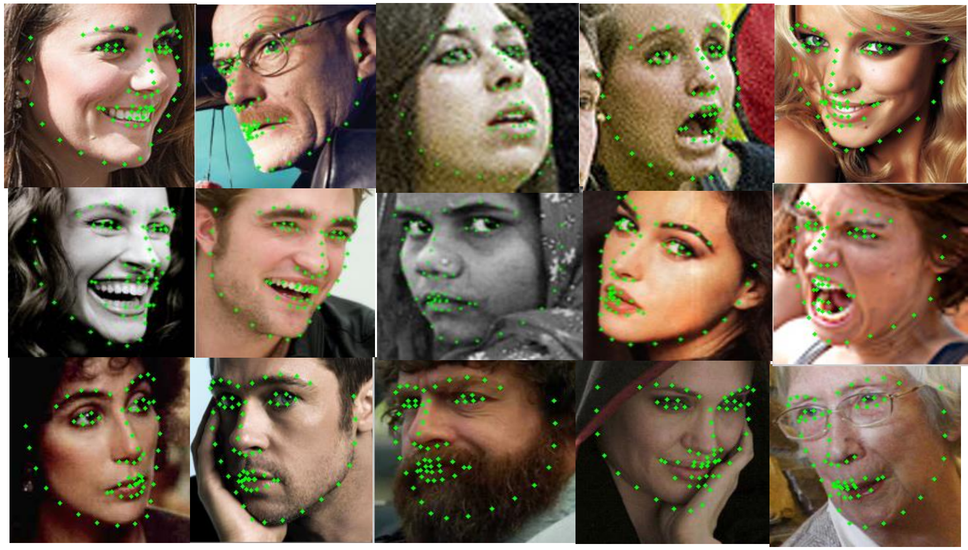
Method	LFPW 68-pts	HELEN 68-pts	HELEN 192-pts	Common Set 68-pts	Challenging Set 68-pts	Full Set 68-pts
FPLL	8.29	8.16	-	8.22	18.33	10.20
DRMF	6.57	6.70	-	6.65	19.79	9.22
RCPR	6.56	5.93	6.50	6.18	17.26	8.35
GN-DPM	5.92	5.69	-	5.78	-	-
SDM	5.67	5.50	5.85	5.57	15.40	7.50
CFAN	5.44	5.53	-	5.50	-	-
ERT	-	-	4.90	-	-	6.40
BPCPR	-	-	-	5.24	16.56	7.46
ESR	-	-	5.70	5.28	17.00	7.58
LBF	-	-	5.41	4.95	11.98	6.32
LBF fast	-	-	5.80	5.38	15.50	7.37
Deep Reg	-	-	-	4.51	13.80	6.31
CFSS	4.87	*4.63*	*4.74*	*4.73*	*9.98*	*5.76*
CFSS Practical	4.90	4.72	4.84	4.73	10.92	5.99
TCDCN	*4.57*	4.60	*4.63*	4.80	8.60	*5.54*
DCRFA	*4.57*	*4.25*	-	*4.19*	**8.42**	**5.02**
R-DSSD*	4.77	4.31	4.95	4.57	10.86	5.91
R-DSSD	**4.52**	**4.08**	**4.62**	**4.16**	*9.20*	5.59

Hao Liu, **Jiwen Lu**, jianjiang feng and Jie Zhou, Learning deep sharable and structural detectors for face alignment, *IEEE Transactions on Image Processing (T-IP)*, 26(4):1666-1678, 2017.

Evaluation on Landmark Density

Robustness to density, expressions and poses

HELEN 192-pts



IBUG 68-pts

Hao Liu, **Jiwen Lu**, jianjiang feng and Jie Zhou, Learning deep sharable and structural detectors for face alignment, *IEEE Transactions on Image Processing (T-IP)*, 26(4):1666-1678, 2017.

2.4 Multi-Task Representation Learning

Motivation

- Facial landmarks are correlated with facial expression, facial 3D pose and partial occlusion
- Sharing knowledge in representation learning with multiple related tasks

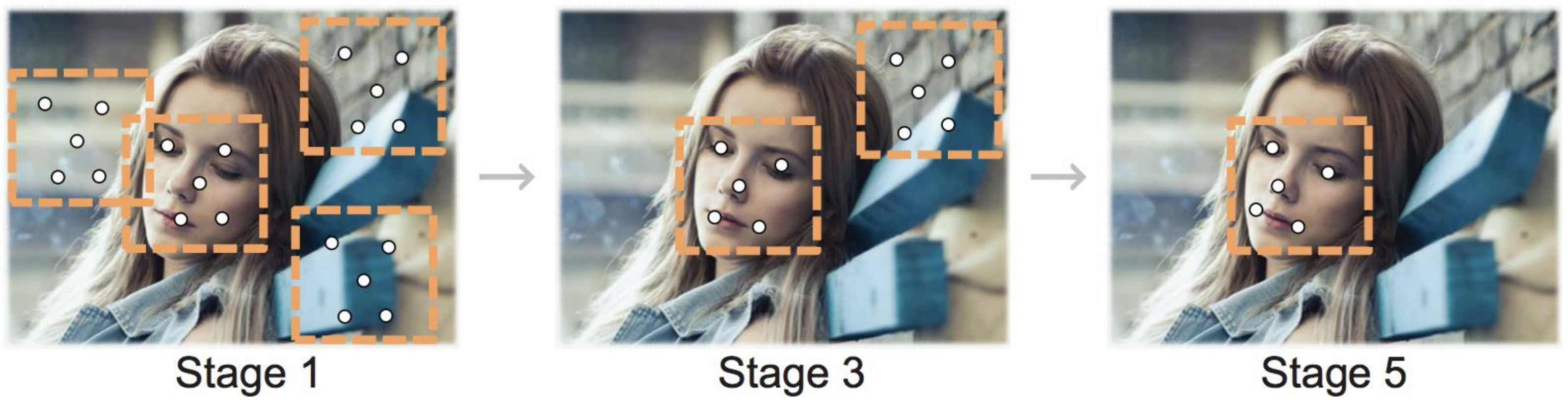
Face Detection and Alignment

- The main goal of multi-task learning

$$\min_{\mathbf{W},\mathbf{b}} \sum_{i=1}^{m}\sum_{j=1}^{n_i} l(y_j^i, (\mathbf{w}^i)^T\mathbf{x}_j^i + b_i) + \lambda\|\mathbf{W}\|_{2,1}.$$

where x denotes training samples and y specifies each sample's label for each task.

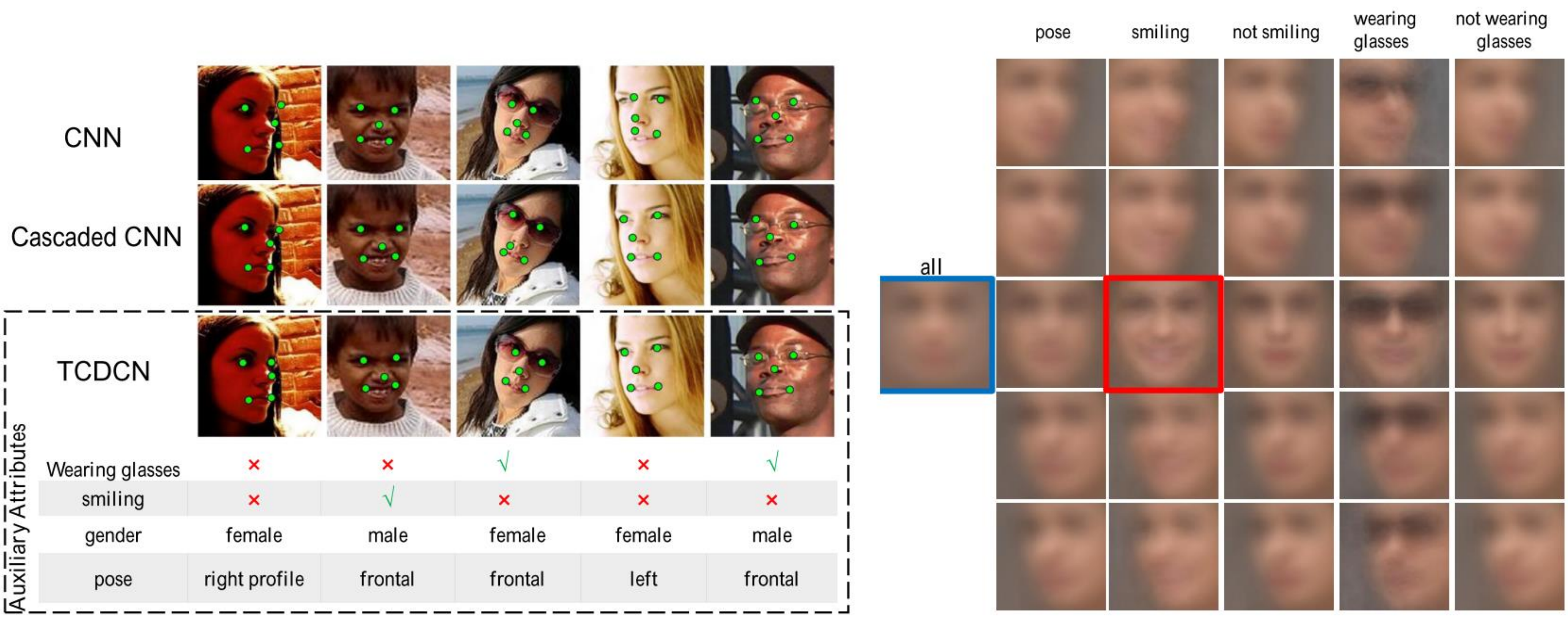
- Multi-task learning for face detection and alignment



[Chen et al., ECCV'14]

Learning Multi-Task Representation with Auxiliary Facial Tasks

- Face alignment by auxiliary tasks

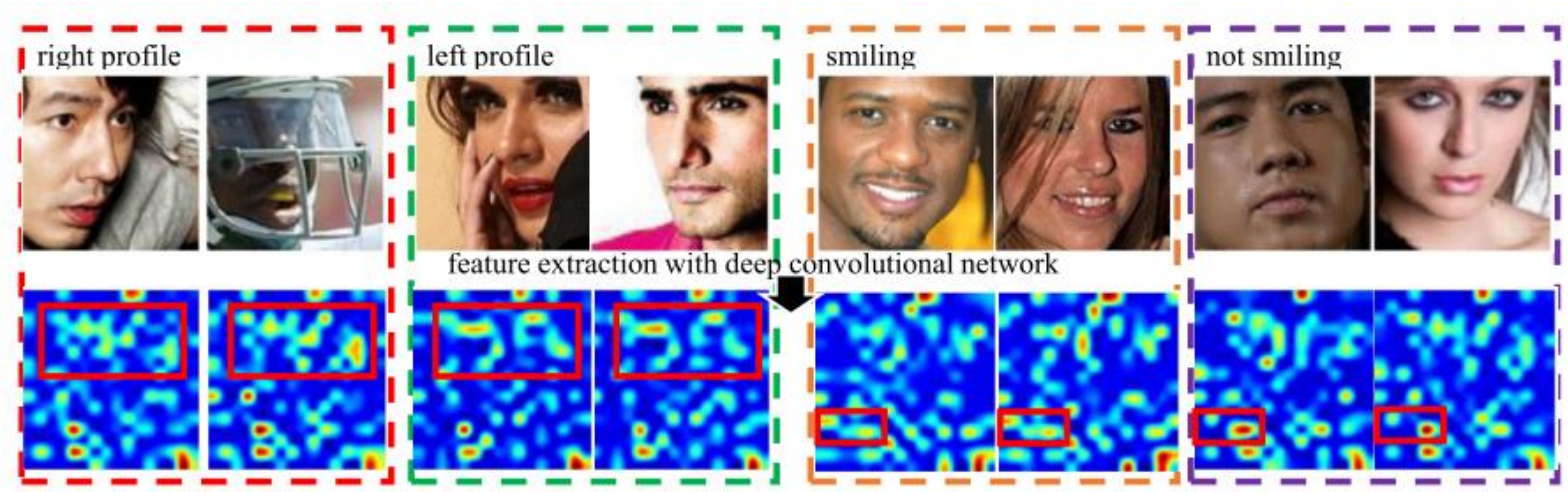


(a) (b)

Auxiliary Attributes：Gender、Expressions、Pose、Wearing Glasses

[Zhang et al., PAMI’16]

Learning Multi-Task Representation with Auxiliary Facial Tasks

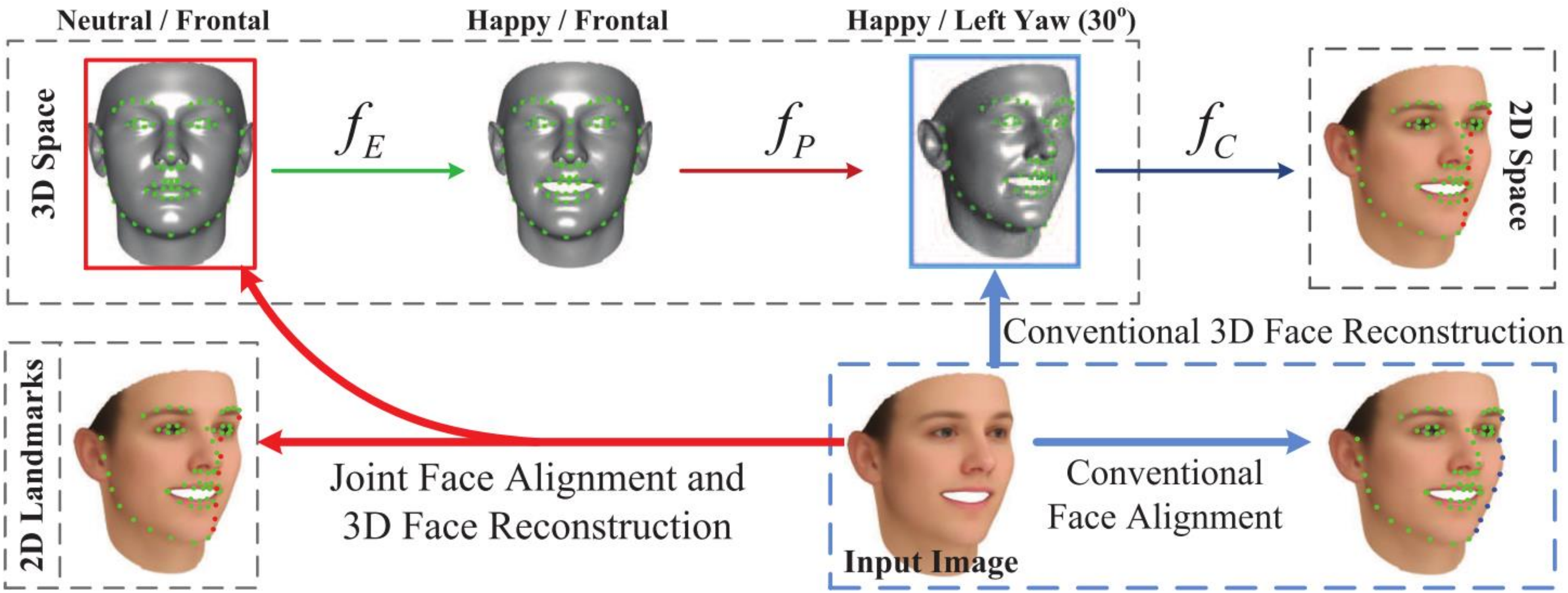


- The first row-face images/ the second row-corresponding features
- The face images with similar poses and attributes are close with each other.
- Learned feature space is robust to pose, expression, and occlusion.

[Zhang et al., PAMI'16]

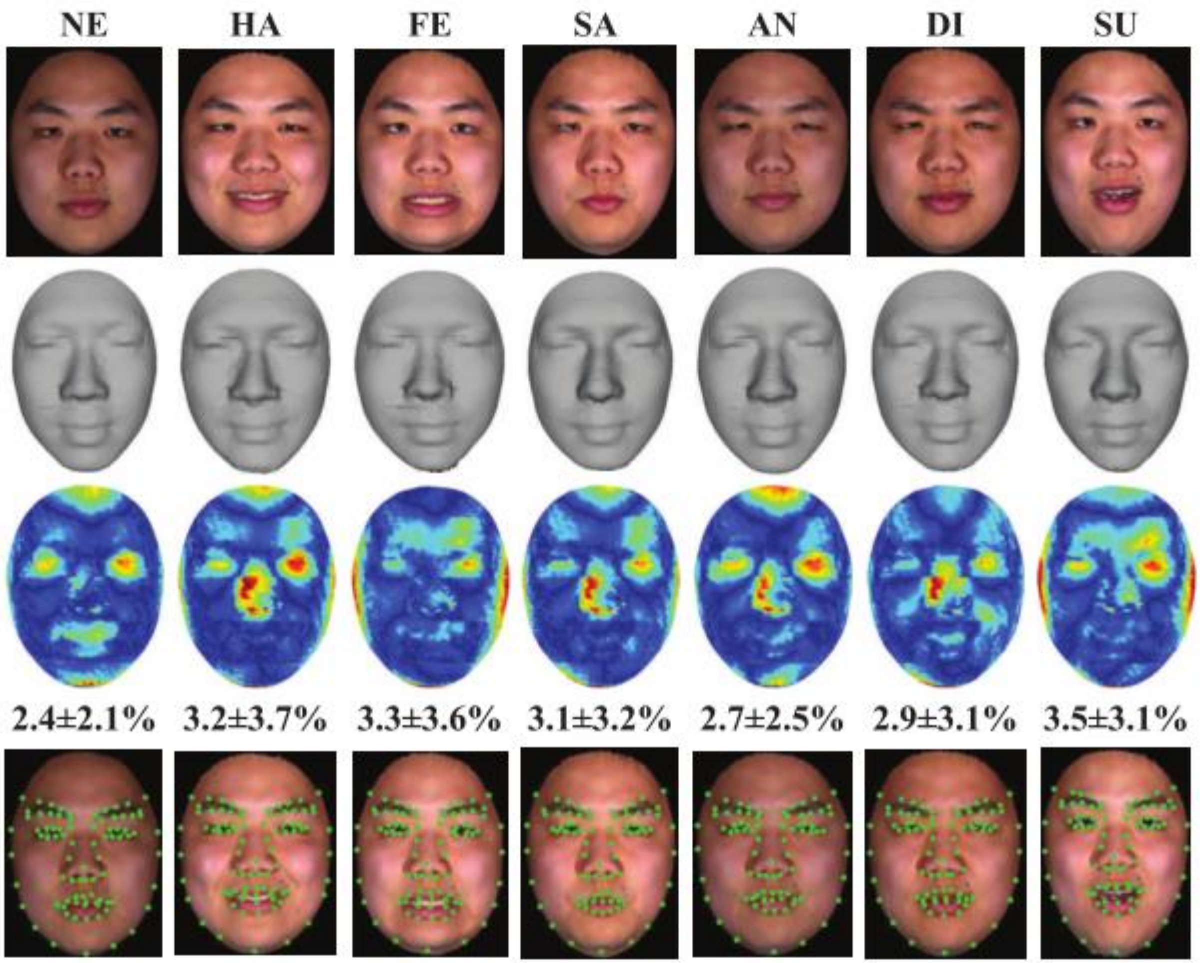
Learning Multi-Task Representation with 3D Surface Reconstruction

- Joint face alignment and 3D face reconstruction
  - 2D landmark contributes to 3D surface reconstruction



[Liu et al., ECCV'16]

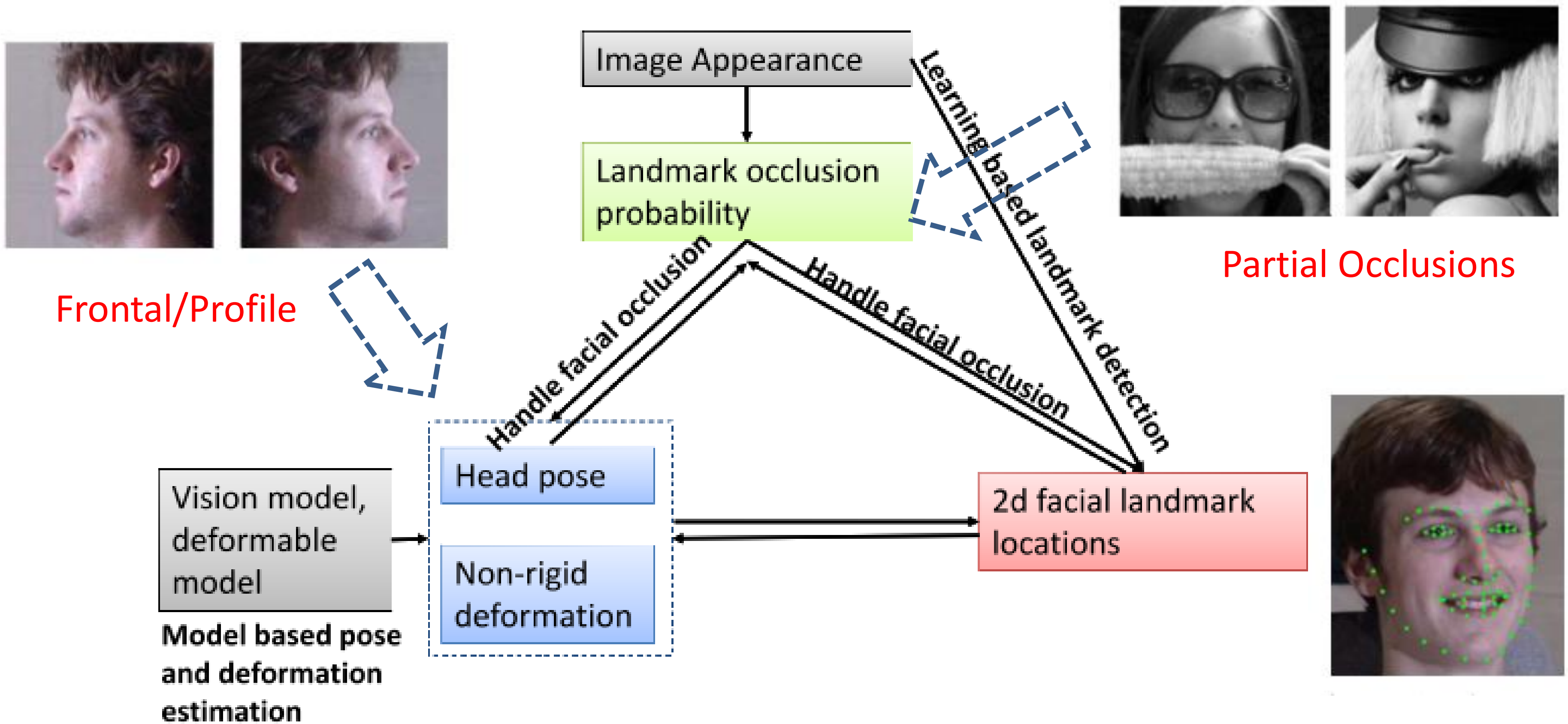
Learning Multi-Task Representation with 3D Surface Reconstruction



[Liu et al., ECCV'16]

Learning Multi-Task Representation with Pose/Deformation/Occlusion

- Coupling tasks for face alignment



[Wu et al., CVPR'17]

Performance effects of multi-task learning

Table 1. Comparison of the head pose estimation methods (mean absolute errors) on BU database.

Method	Pitch	Yaw	Roll	Average
Rigid model [21]	11.9	5.2	2.8	6.6
Cylindrical [15]	6.6	3.3	9.8	6.4
Cylindrical+AAM [20]	5.6	5.4	3.1	4.7
Deformable model [24]	4.3	6.2	3.2	4.6
3D CLM [2]	6.0	3.9	3.7	4.5
ours	5.3	4.9	3.1	4.4

Table 5. Comparison of landmark detection (average pixel errors) on MultiPIE database (51 points).

near-frontal							all poses
CLM [18]	FPLL [31]	Pose-free [28]	Deep3D [30]	3D CLM [2]	Chehra [1]	ours	ours
4.75	4.39	7.34	5.74	5.30	4.09	3.51	3.50

[Wu et al., CVPR'17]

Performance effects of multi-task learning

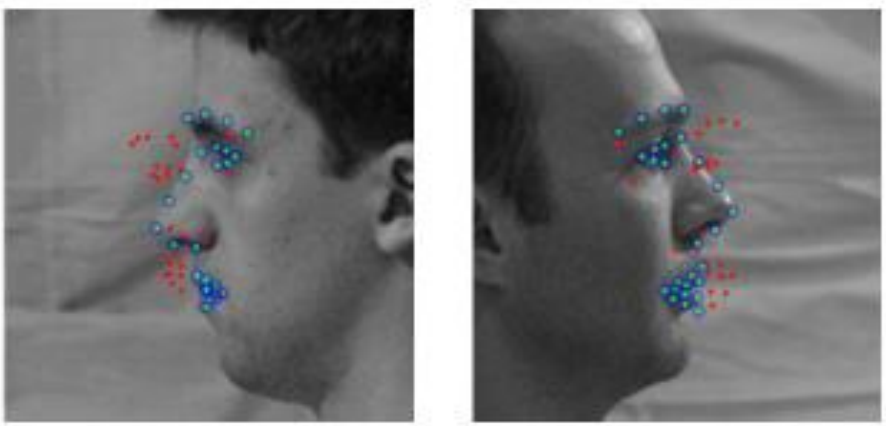


Table 2. Comparison of facial landmark detection errors (normalized errors w.r.t. inter-ocular distance) and occlusion prediction results on COFW database (29 points) [3].

Method	Landmark error	Occlusion (precision/recall)
Human	5.6 [3]	-
CRC [8]	7.30	-
OC [9]	7.46	80.8/37.0%
RCPR [3]	8.50	80/40%
ESR [4]	11.20	-
FPLL [31]	14.40	-
SDM [26]	7.70	-
ours	6.40	80/44.43%

[Wu et al., CVPR’17]

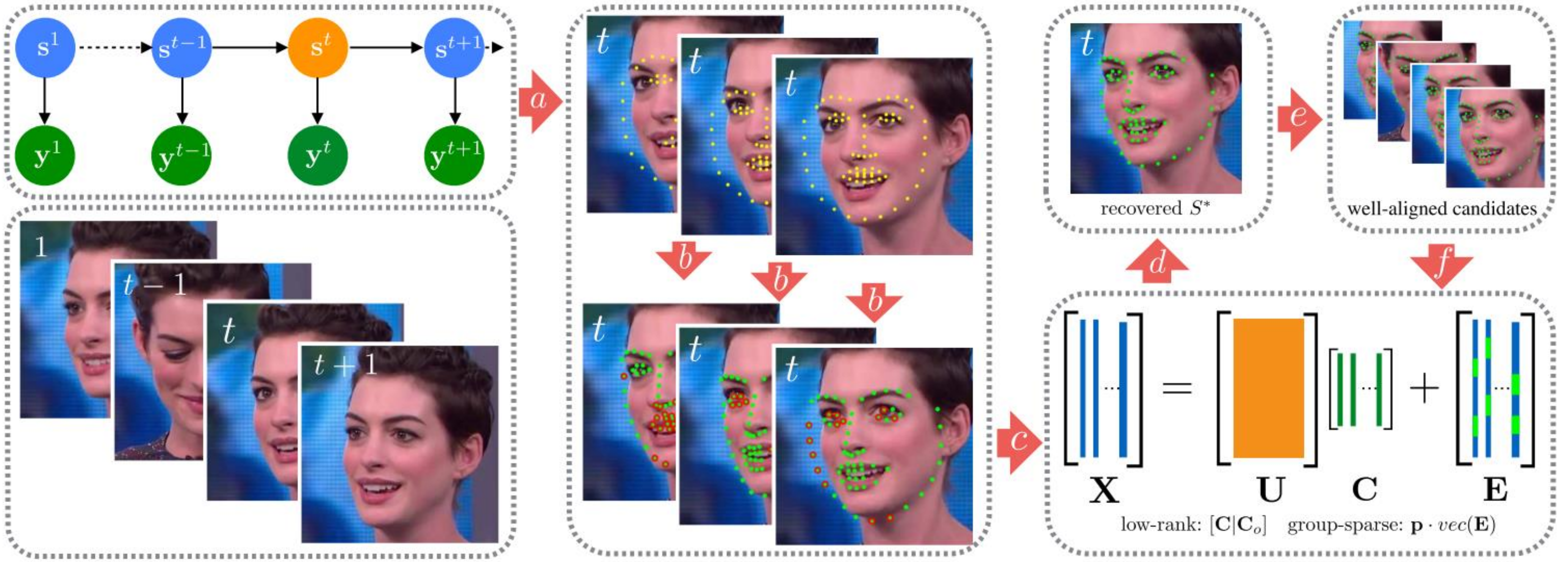
2.5 Spatial-Temporal Representation Learning

Video-based Face Alignment

Time-Stamps

- Problem Setting
  - Input： face sequence $\mathbf{x}_i^{1:T} = \{\mathbf{x}_i^1, \mathbf{x}_i^2, ..., \mathbf{x}_i^t, ..., \mathbf{x}_i^T\}$
  - Output： landmarks for t-th frame $\mathbf{p}_i^t = [p_1, p_2, \cdots, p_l, \cdots, p_L]_i^{t'}$
  - Goal: sequential face alignment $\{\mathbf{x}^t\}^{t=1:T} \longrightarrow \{\mathbf{p}^t\}^{t=1:T}$

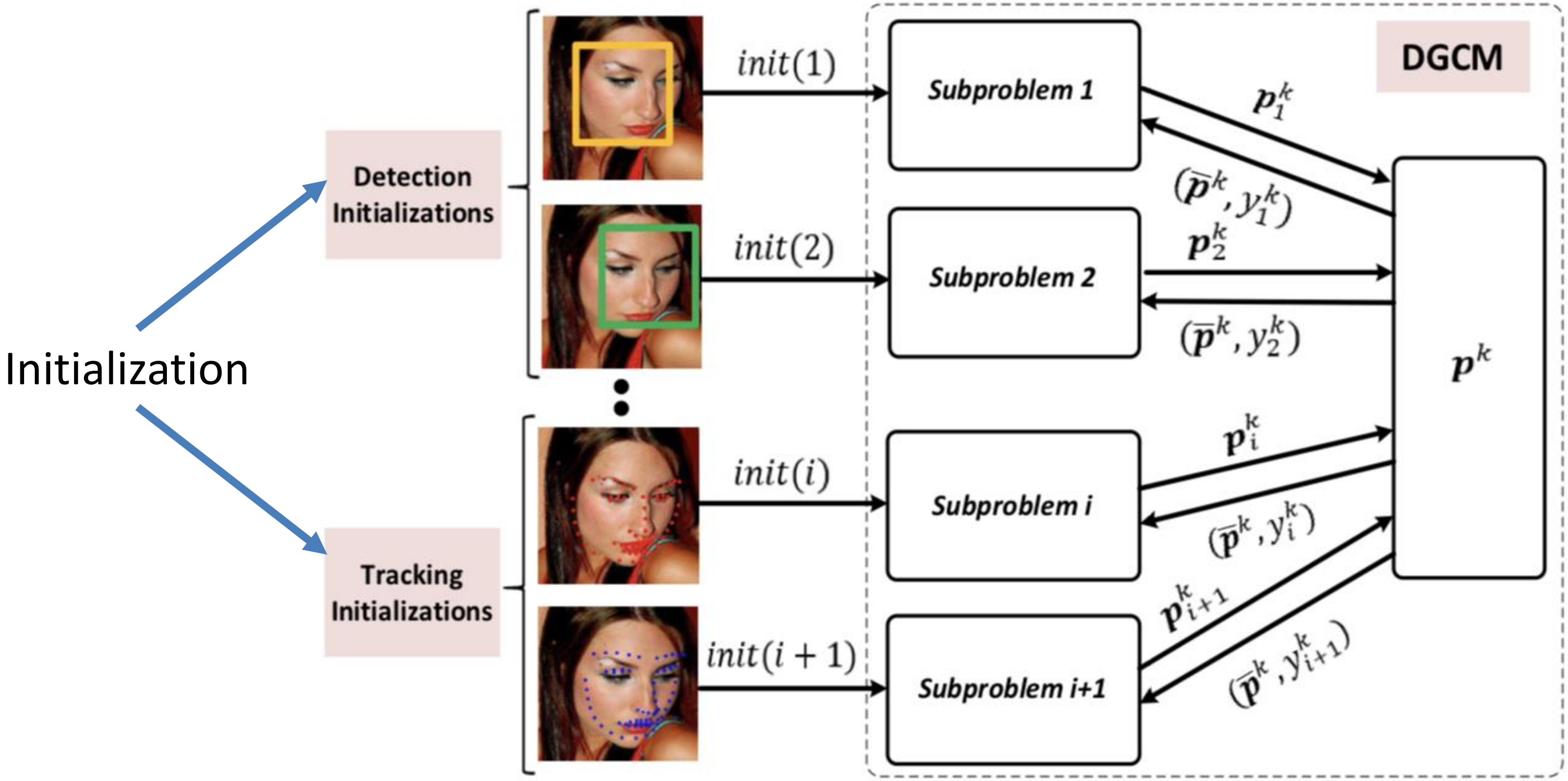
Learning Temporal Representation for Video-based Face Alignment



- Challenges for Video-based Face Alignment
  - Consistency over time steps
  - Robustness to initialization of both spatial and temporal dimension

[Peng et al., ICCV'16]

Learning Spatial-Temporal Representation for Video-based Face Alignment



[Khan et al., ICCV'17]

Learning Spatial-Temporal Representation for Video-based Face Alignment

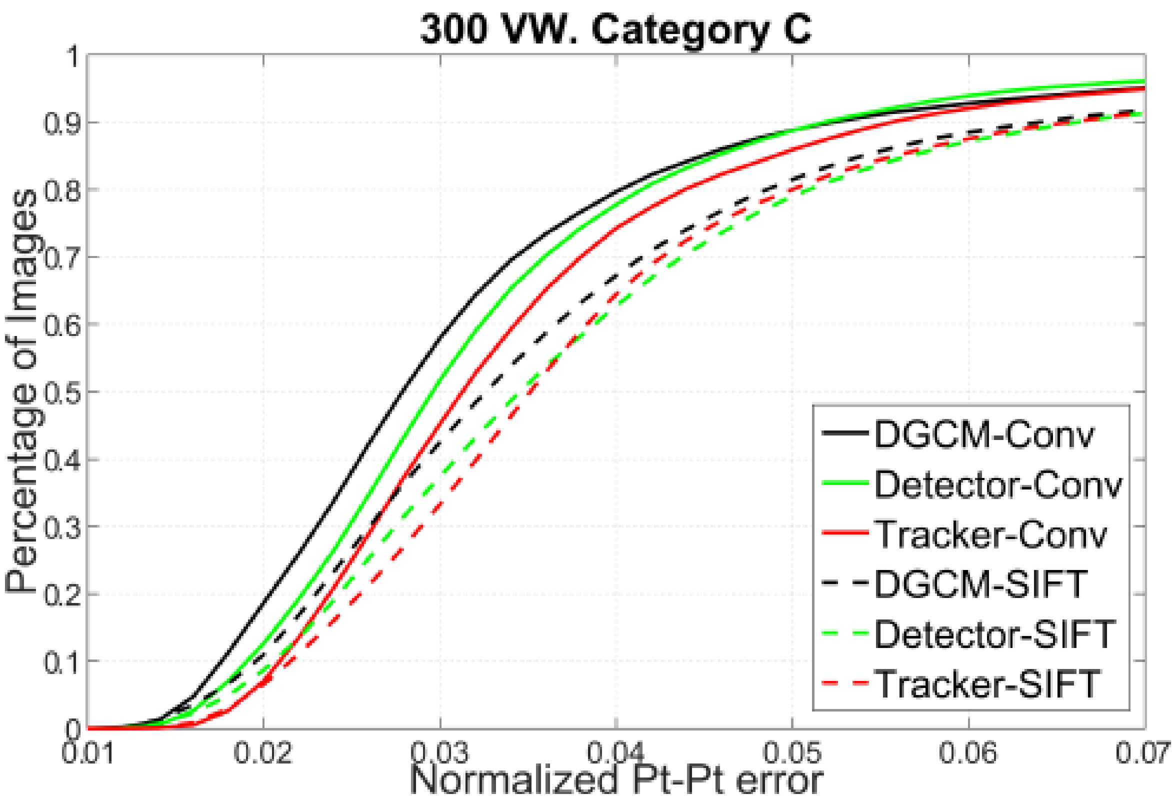


Figure 3. Comparison between DGCM with Detection (alone) and Tracking (alone) on category C of 300-VW.

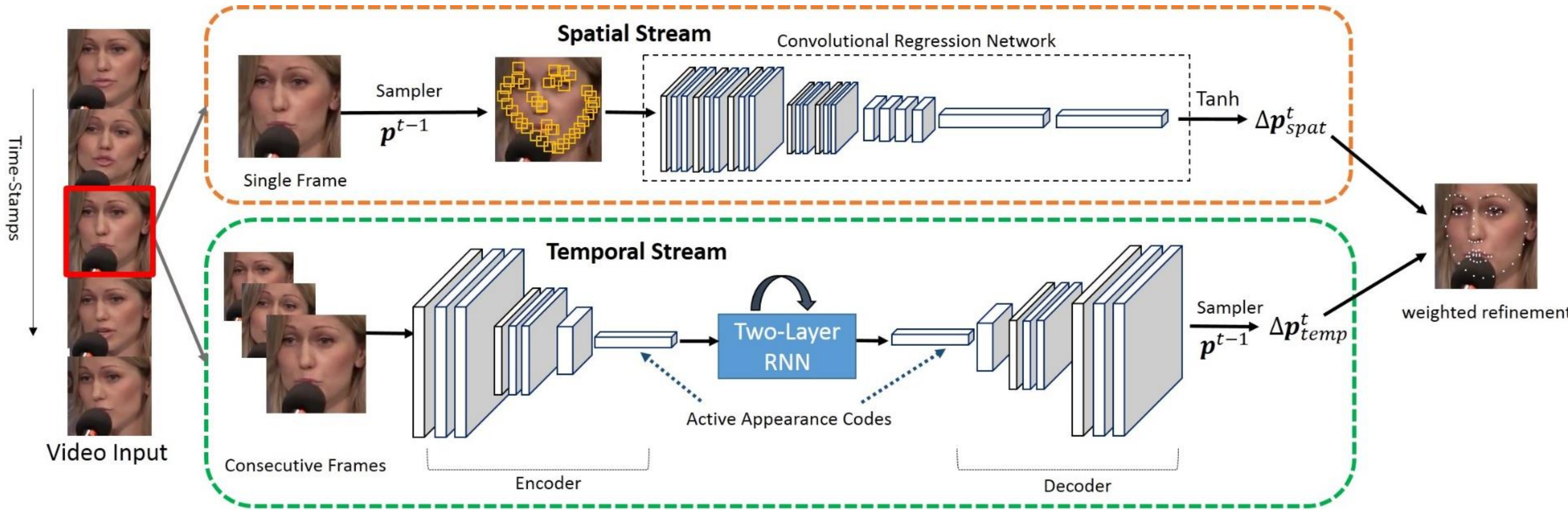
Performance benefits from both detection and tracking!

[Khan et al., ICCV'17]

Two-Stream Transformer Networks for Video-based Face Alignment

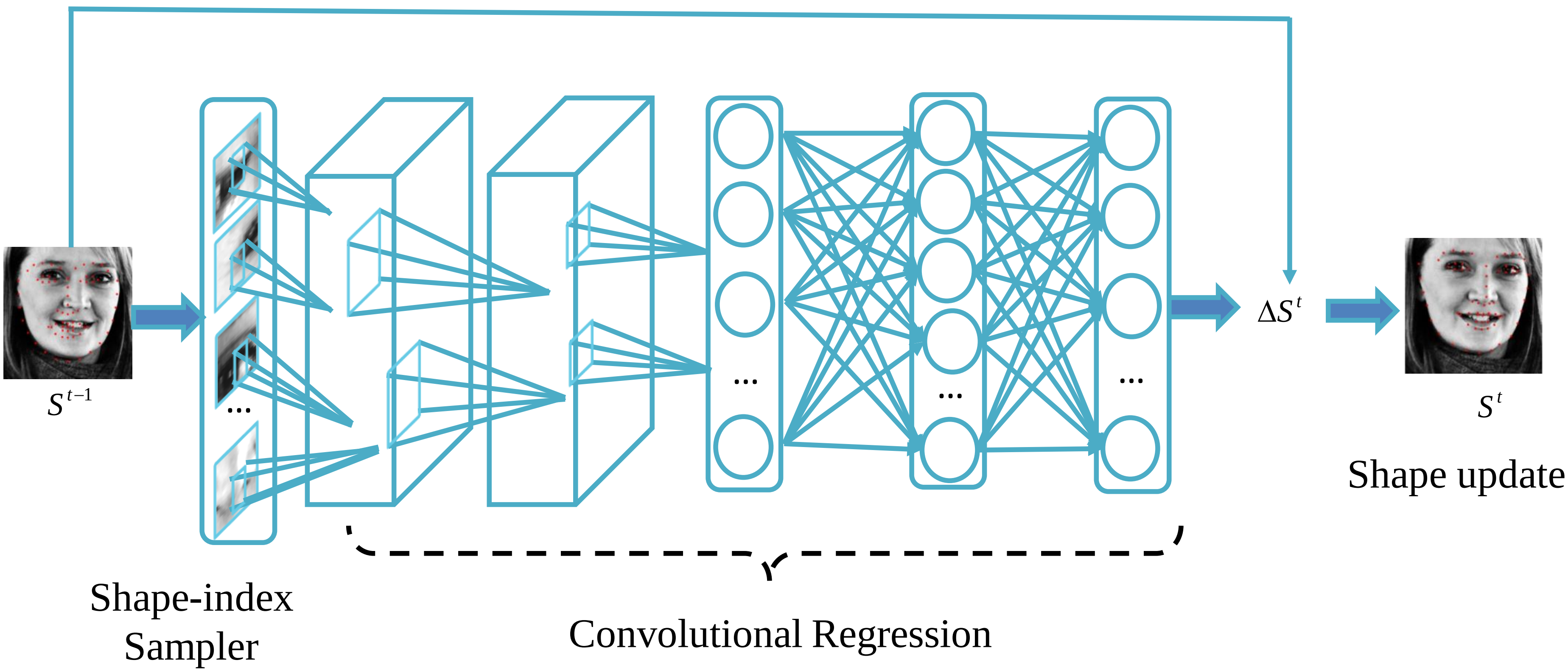
- Basic Idea
  - Complementary information of both streams
  - Spatial stream: spatial appearance in still images
  - Temporal stream: consistency across frames
- Objective:

$$\min_f J = \sum_i^N \sum_t^T \frac{1}{2} \left\| \Delta \mathbf{p}_i^t - \beta_1 f_{\text{spat}}\left(\mathbf{x}_i^t\right) - \beta_2 f_{\text{temp}}\left(\mathbf{x}_i^t\right) \right\|_2^2,$$
$$\text{subject to} \quad \beta_1 + \beta_2 = 1.$$



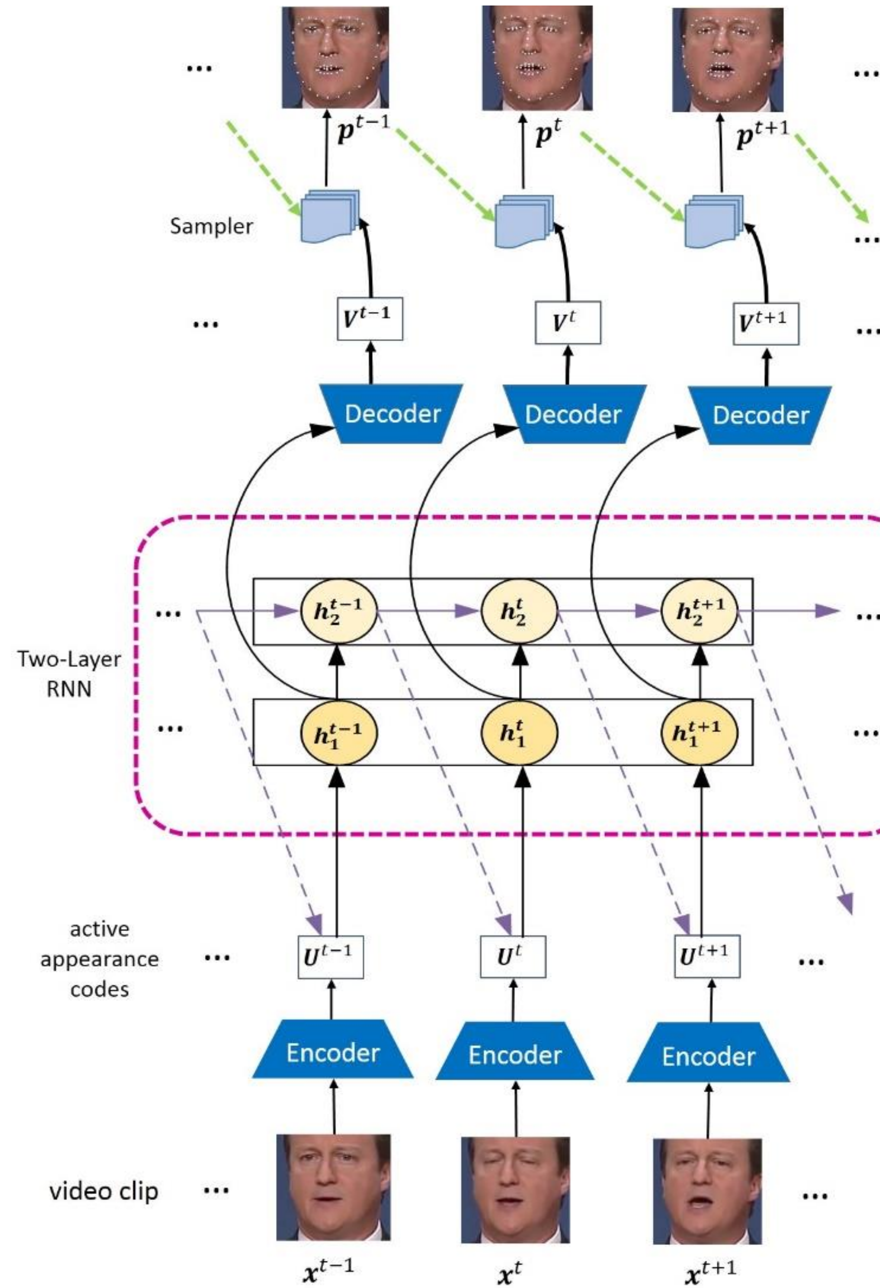
Hao Liu, **Jiwen Lu**, jianjiang feng and Jie Zhou**,** Two-Stream Transformer Networks for Video-based Face Alignment. IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI), 2017, in Press

Spatial Stream



Hao Liu, **Jiwen Lu**, jianjiang feng and Jie Zhou**,** Two-Stream Transformer Networks for Video-based Face Alignment. IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI),, 2017, in Press

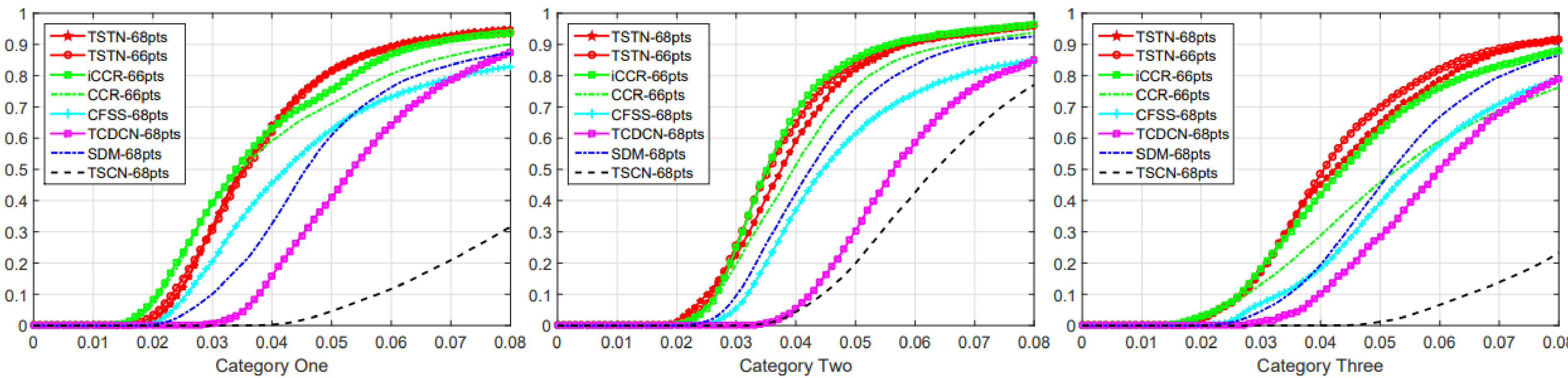
Temporal Stream



Hao Liu, **Jiwen Lu**, jianjiang feng and Jie Zhou**,** Two-Stream Transformer Networks for Video-based Face Alignment. IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI), , 2017, in Press

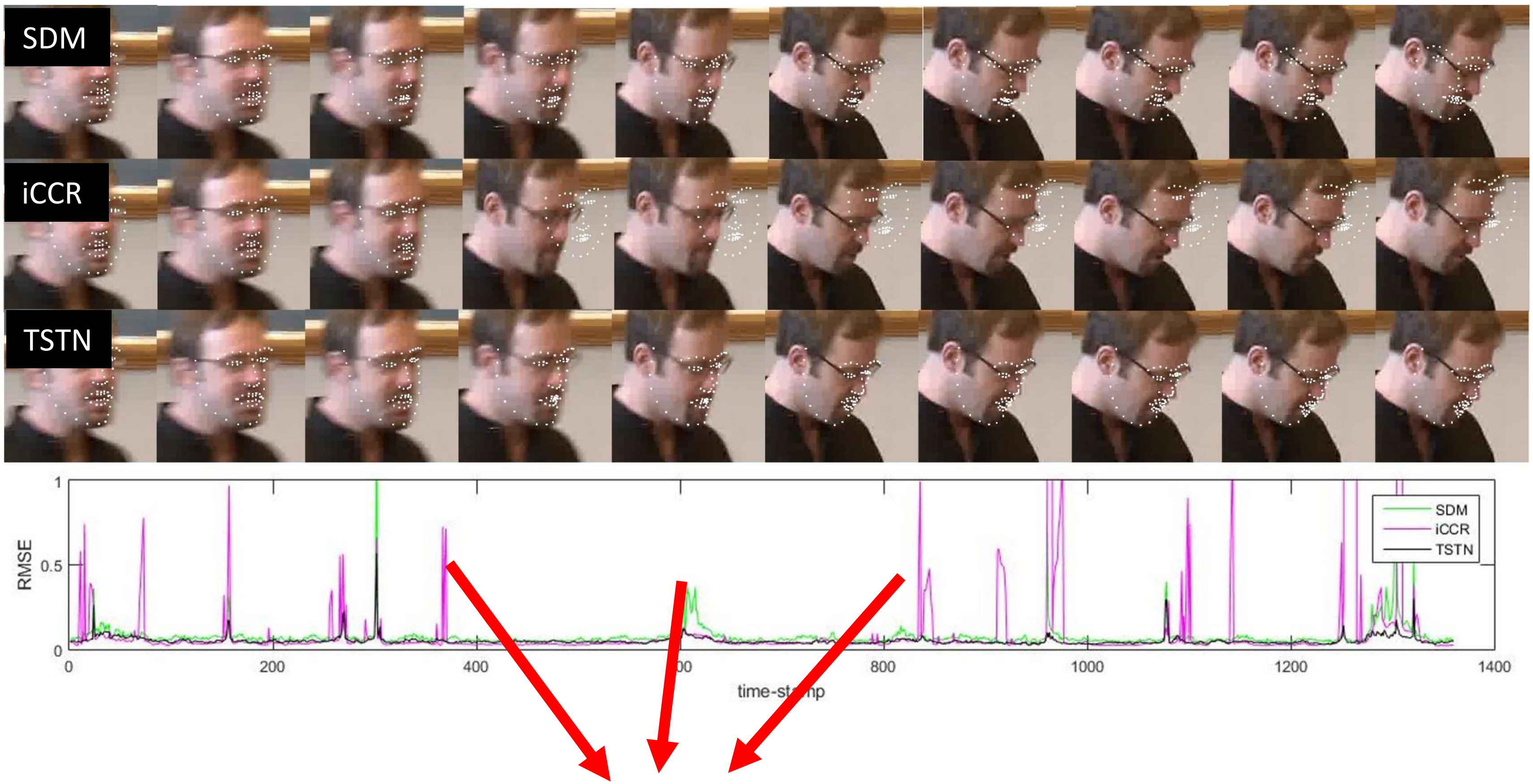
Quantitative Evaluation

Methods	Model Description	Category 1	Category 2	Category 3	Challset [25]	-pts	Year
SDM [46]	Cascaded Linear Regression	7.41	6.18	13.04	7.44		2013
TSCN [35][1]	Two-Stream Action Network	11.61	11.59	17.67	-		2014
TSCN [35][1,2]	Two-Stream Action Network	12.54	7.25	13.13	-		2014
CFSS [50]	Coarse-to-Fine Shape Searching	7.68	6.42	13.67	5.92	68	2015
PIEFA [26]	Personalized Ensemble Learning	-	-	-	6.37		2015
REDN [25]	Recurrent Auto-Encoder Net	-	-	-	6.25		2016
TCDCN [49]	Multi-Task Deep CNN	7.66	6.77	14.98	7.27		2016
TSTN	Two-Stream Transformer Net	**5.36**	**4.51**	**12.84**	**5.59**		-
CCR [32]*	Cascaded Continuous Regression	7.26	5.89	15.74	-		2016
iCCR [32]*	Cascaded Continuous Regression	*6.71*	**4.00**	*12.75*	-	66	2016
TSTN	Two-Stream Transformer Net	**5.21**	*4.23*	**10.11**	-		-



Hao Liu, **Jiwen Lu**, jianjiang feng and Jie Zhou**,** Two-Stream Transformer Networks for Video-based Face Alignment. IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI), , 2017, in Press

Qualitative Evaluation



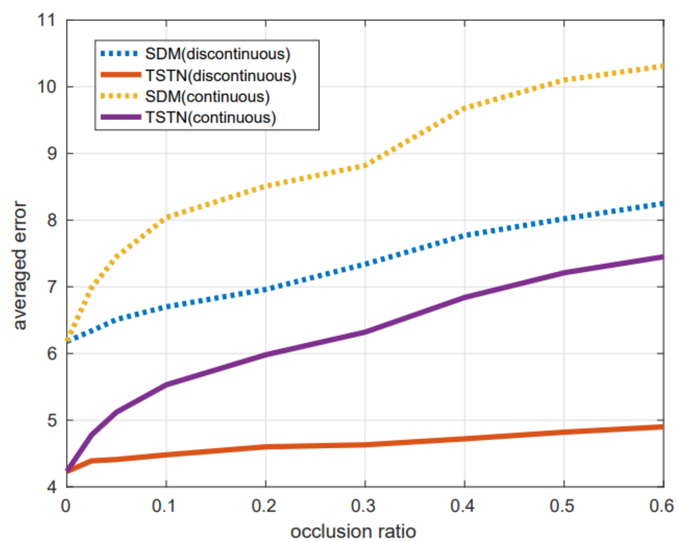
Drifting

Hao Liu, **Jiwen Lu**, jianjiang feng and Jie Zhou**,** Two-Stream Transformer Networks for Video-based Face Alignment. IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI), , 2017, in Press

Robustness to Temporal Occlusions

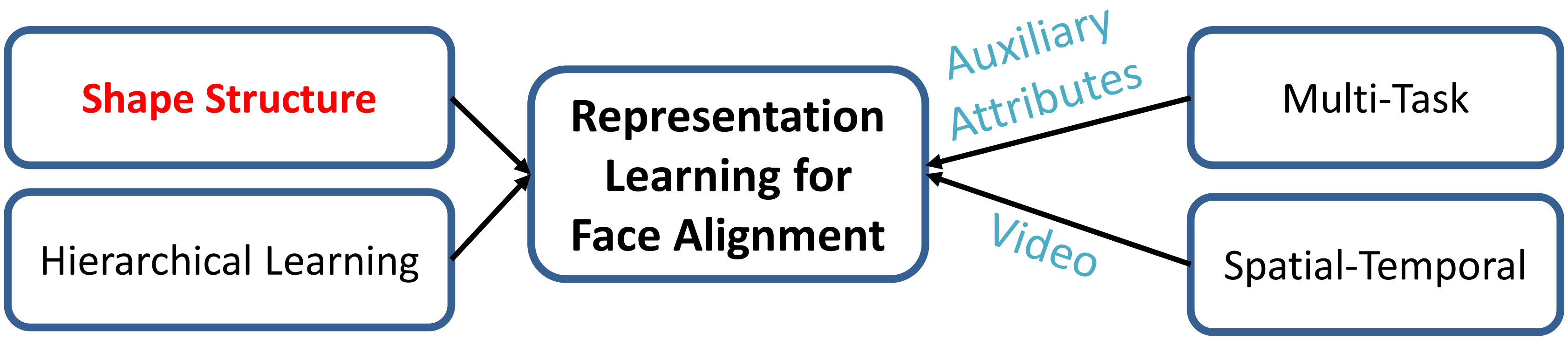
Continuous Occlusion across the Video Clip

Discontinuous Occlusion across the Video Clip

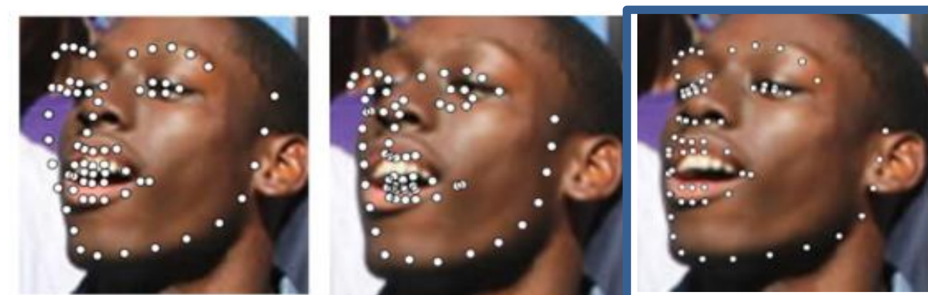


Hao Liu, **Jiwen Lu**, jianjiang feng and Jie Zhou**,** Two-Stream Transformer Networks for Video-based Face Alignment. IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI), , 2017, in Press

Summary

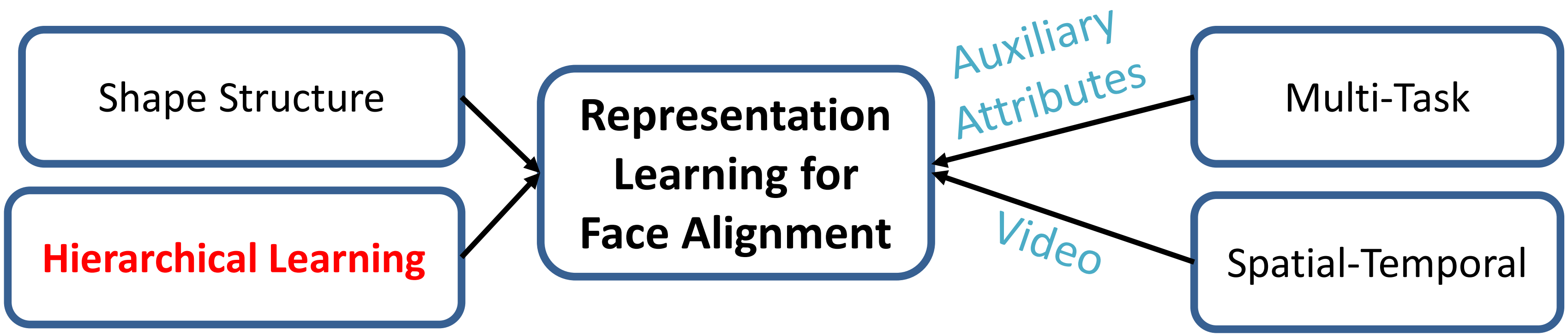


- Shape Structure
  - correlation for neighboring landmarks (locality)
  - holistic shape constraint

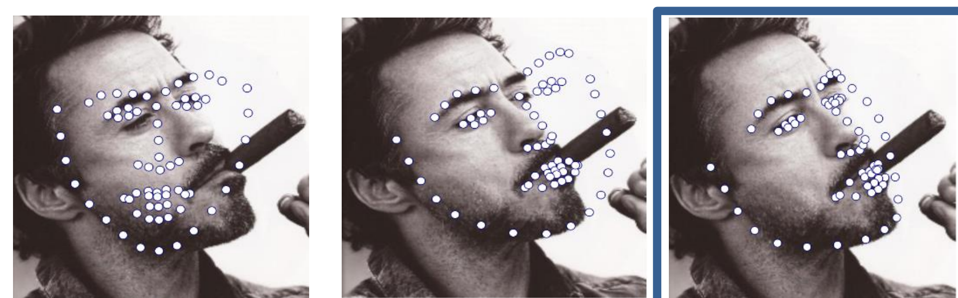


Designed | Local Patches | **Structural Representation**

Summary

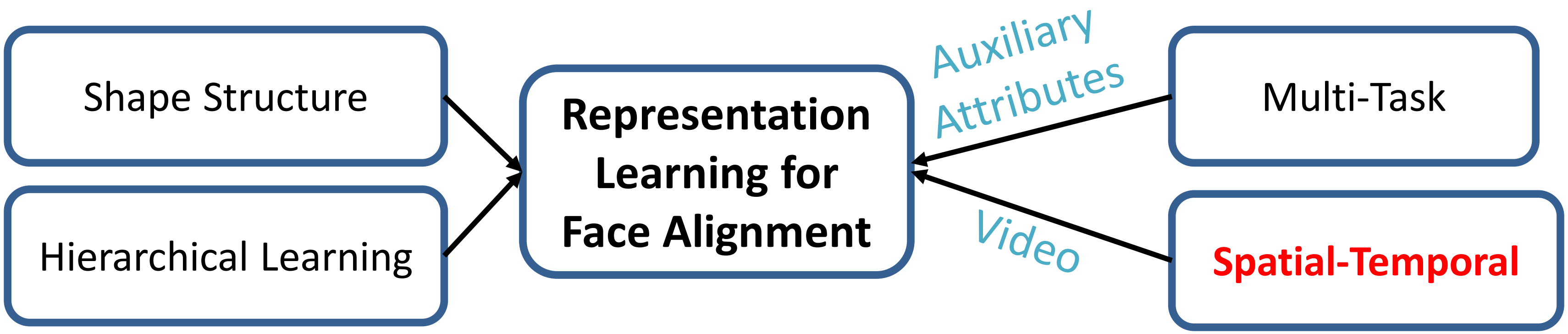


- Hierarchical Learning
  - nonlinear relationship
  - end-to-end optimization



Global Appearance | Local Patches | **Hierarchical Representation**

Summary



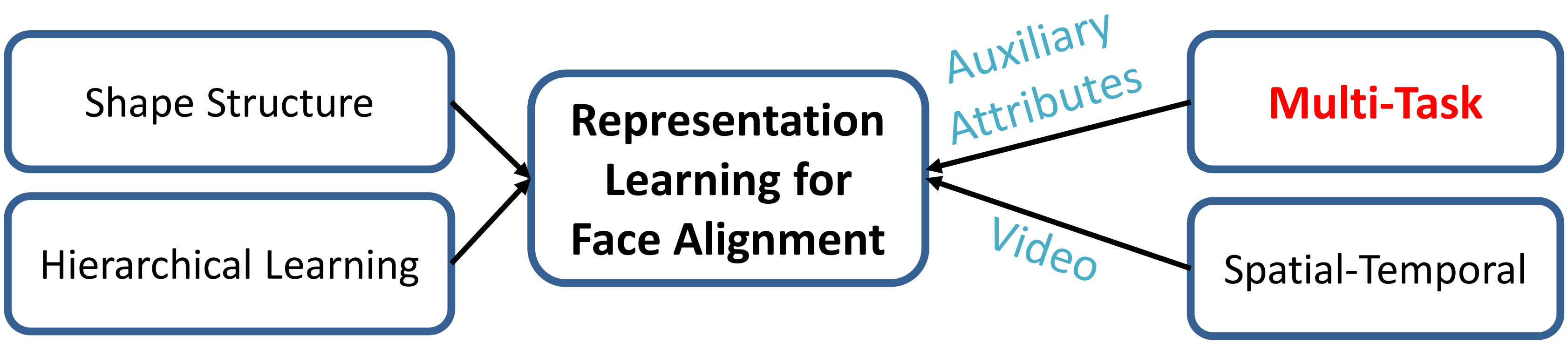
- Spatial-Temporal
  - spatial appearance
  - temporal consistency

Spatial Representation

Temporal Representation

Spatial-Temporal Representation

Summary



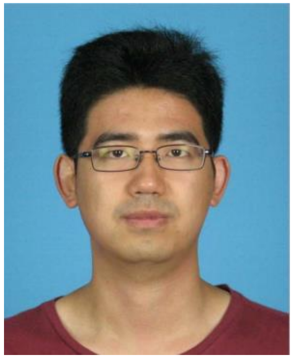
- Shape Structure
  - correlation for neighboring landmarks (locality)
  - holistic shape constraint
- Hierarchical Learning
  - nonlinear relationship
  - end-to-end optimization
- Spatial-Temporal
  - spatial appearance
  - temporal consistency
- Multi-Task
  - Learning with correlated tasks (facial attributes)
  - shared representation

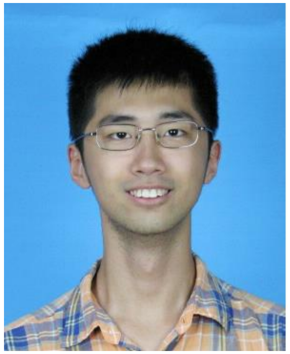
Coffee Break

IEEE FG 2018 Tutorial on

Representation Learning for Face Alignment and Recognition

Tutors: Hao Liu, Yueqi Duan, and Jiwen Lu





URL: http://ivg.au.tsinghua.edu.cn/FG18_tutorial/FG18_face.pdf

Outline

- Part 1: Introduction
- Part 2: Representation Learning for Face Alignment

------------------- Short Break: 30 minutes --------------------

- Part 3: Representation Learning for Face Recognition
- Part 4: Discussions

Part 3: Representation Learning for Face Recognition

Applications of Face Recognition

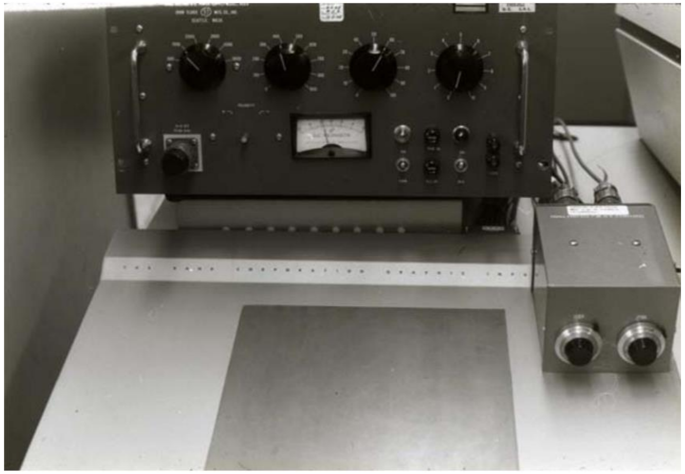
- Information security
  - Access security
  - Data privacy
  - User authentication
- Access management
  - Secure access authentication
  - Permission based systems
- Biometrics
  - Person identification
  - Automated identity verification
- Law enforcement
  - Video surveillance
  - Suspect identification
  - Suspect tracking
  - Forensic reconstruction
- Personal security
  - Home video surveillance
  - Expression interpretation
- Entertainment

History of Face Recognition

1950 — *Psychology*

History of Face Recognition

Woodrow Wilson Bledsoe



RAND TABLET

1960 — *Man-machine facial recognition*
(record the coordinate locations of facial features)

1950 — *Psychology*

History of Face Recognition

DEFENSE TECHNICAL INFORMATION CENTER

Select Search | Keywords | Advanced Search

Accession Number : AD0713252

Title : VISUAL IDENTIFICATION OF PEOPLE BY COMPUTER.

Descriptive Note : Doctoral thesis,

Corporate Author : STANFORD UNIV CALIF DEPT OF COMPUTER SCIENCE

Personal Author(s) : Kelly,Michael David

Report Date : AUG 1970

Pagination or Media Count : 249

Abstract : The thesis describes a computer program which performs a complex picture processing task. The task is to choose, from a collection of pictures of people taken by a TV camera, those pictures that depict the same person. The primary purpose of this research has been directed toward the development of new techniques for picture processing. (Author)

Descriptors : (*HUMANS, IDENTIFICATION), (*ARTIFICIAL INTELLIGENCE, PATTERN RECOGNITION), (*PICTURES, PROCESSING), COMPUTER PROGRAMMING, SELECTION, MATHEMATICAL MODELS, ANATOMICAL MODELS, OPTICAL SCANNING, THESES

Subject Categories : COMPUTER PROGRAMMING AND SOFTWARE
PRINTING AND GRAPHIC ARTS
BIONICS

京都大学学術情報リポジトリ KURENAI 紅 Kyoto University Research Information Repository

京都大学 KYOTO UNIVERSITY

Title	Picture Processing System by Computer Complex and Recognition of Human Faces(Abstract_)
Author(s)	Kanade, Takeo
Citation	Kyoto University ( )
Issue Date	1974-05-23
URL	https://doi.org/10.14989/doctor.k1486
Right	
Type	Thesis or Dissertation
Textversion	author

1970 — *Geometric parameters*

(distances and angles between landmarks, without human intervention)

1960 — *Man-machine facial recognition*

1950 — *Psychology*

History of Face Recognition

Low-dimensional procedure for the characterization of human faces

L. Sirovich and M. Kirby

Division of Applied Mathemetics, Brown University, Providence, Rhode Island 02912

Received August 25, 1986; accepted November 10, 1986

A method is presented for the representation of (pictures of) faces. Within a specified framework the representation is ideal. This results in the characterization of a face, to within an error bound, by a relatively low-dimensional vector. The method is illustrated in detail by the use of an ensemble of pictures taken for this purpose.

1980 — *Geometric measurement, PCA, ANN*

(the first mention to eigenfaces and "deep" faces)

1970 — *Geometric parameters*

1960 — *Man-machine facial recognition*

1950 — *Psychology*

History of Face Recognition

426

PRACTICAL FACE RECOGNITION AND VERIFICATION WITH WISARD.

T. J. STONHAM

1. INTRODUCTION

WISARD (Wilkie, Aleksander, and Stonham's Recognition Device) is a general purpose pattern recognition machine with a special semi-parallel structure unlike that of conventional single instruction single data computers. The machine is self-adapting. It does not require

1980 — *Geometric measurement, PCA, ANN*

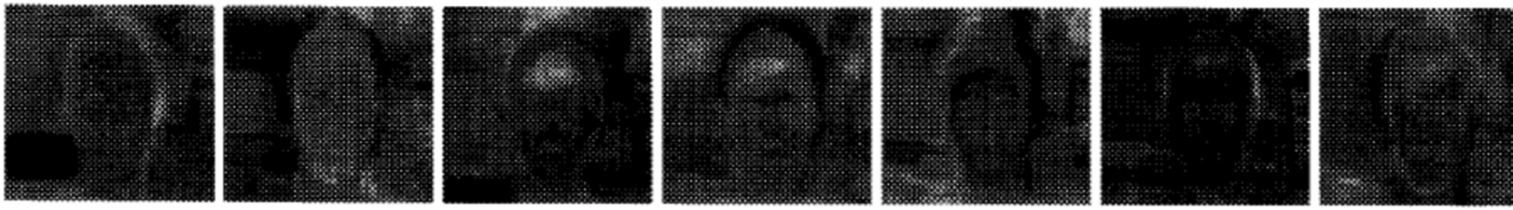
(the first mention to eigenfaces and "deep" faces)

1970 — *Geometric parameters*

1960 — *Man-machine facial recognition*

1950 — *Psychology*

History of Face Recognition



Eigenfaces, 1991

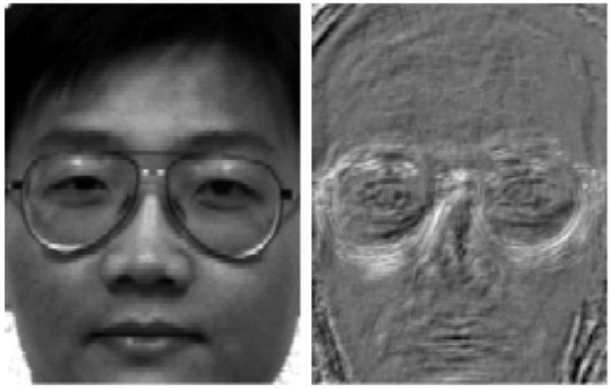
1990 — *Holistic subspace learning, graph matching (Eigenfaces, Fisherfaces)*

1980 — *Geometric measurement, PCA, ANN*

1970 — *Geometric parameters*

1960 — *Man-machine facial recognition*

1950 — *Psychology*



Fisherfaces, 1997

History of Face Recognition

1990 — *Holistic subspace learning, graph matching (Eigenfaces, Fisherfaces)* — *MIT, UMIST, Yale, AT&T, AR (~10-120)*

1980 — *Geometric measurement, PCA, ANN*

1970 — *Geometric parameters* — *Kanade (20)*

1960 — *Man-machine facial recognition*

1950 — *Psychology*

(number of identities)

History of Face Recognition

2000	*Hand-crafted features (Gabor, LBP, SIFT, HOG; Laplacianfaces)*	*FERET, LFW, YTC, PaSC, Caltech (~100-10K)*
1990	*Holistic subspace learning, graph matching*	*MIT, UMIST, Yale, AT&T, AR (~10-120)*
1980	*Geometric measurement, PCA, ANN*	
1970	*Geometric parameters*	*Kanade (20)*
1960	*Man-machine facial recognition*	
1950	*Psychology*	

(number of identities)

History of Face Recognition

		YTF , Multi-PIE, IJB-A, CASIA, VGG, MS-Celeb, Megaface (~300-600K)
2010	*Local feature learning, deep learning (LE, CBFD, CNN)*	
2000	*Hand-crafted features*	*FERET, LFW, YTC, PaSC, Caltech (~100-10K)*
1990	*Holistic subspace learning, graph matching*	*MIT, UMIST, Yale, AT&T, AR (~10-120)*
1980	*Geometric measurement, PCA, ANN*	
1970	*Geometric parameters*	*Kanade (20)*
1960	*Man-machine facial recognition*	
1950	*Psychology*	
		(number of identities)

History of Face Recognition

Year	Methods	Datasets (number of identities)
2010	*Local feature learning, deep learning*	*YTF , Multi-PIE, IJB-A, CASIA, VGG, MS-Celeb, Megaface (~300-600K)*
2000	*Hand-crafted features*	*FERET, LFW, YTC, PaSC, Caltech (~100-10K)*
1990	*Holistic subspace learning, graph matching*	*MIT, UMIST, Yale, AT&T, AR (~10-120)*
1980	*Geometric measurement, PCA, ANN*	
1970	*Geometric parameters*	*Kanade (20)*
1960	*Man-machine facial recognition*	
1950	*Psychology*	

(number of identities)

History of Face Recognition

- 2010 — *Local feature learning, deep learning*
- 2000 — *Hand-crafted features*
- 1990 — *Holistic subspace learning, graph matching*
- 1980 — *Geometric measurement, PCA, ANN*
- 1970 — *Geometric parameters*
- 1960 — *Man-machine facial recognition*
- 1950 — *Psychology*

- Deep Learning
- Local Features
  - Learning-based
  - Hand-crafted
- Holistic Subspace
- Geometry

History of Face Recognition

2010 — *Local feature learning, deep learning*

2000 — *Hand-crafted features*

1990 — *Holistic subspace learning, graph matching*

1980 — *Geometric measurement, PCA, ANN*

1970 — *Geometric parameters*

1960 — *Man-machine facial recognition*

1950 — *Psychology*

- Deep Learning
- Local Features
  - Learning-based
  - Hand-crafted
- Holistic Subspace
- Geometry

History of Face Recognition

2010	*Local feature learning, deep learning*
2000	*Hand-crafted features*
1990	*Holistic subspace learning, graph matching*
1980	*Geometric measurement, PCA, ANN*
1970	*Geometric parameters*
1960	*Man-machine facial recognition*
1950	*Psychology*

- Deep Learning
- Local Features
  - Learning-based
  - Hand-crafted
- Holistic Subspace
- Geometry

History of Face Recognition

2010	*Local feature learning, deep learning*
2000	*Hand-crafted features*
1990	*Holistic subspace learning, graph matching*
1980	*Geometric measurement, PCA, ANN*
1970	*Geometric parameters*
1960	*Man-machine facial recognition*
1950	*Psychology*

- Deep Learning
- Local Features
  - Learning-based
  - Hand-crafted
- Holistic Subspace
- Geometry

History of Face Recognition

- 2010 — *Local feature learning, deep learning*
- 2000 — *Hand-crafted features*
- 1990 — *Holistic subspace learning, graph matching*
- 1980 — *Geometric measurement, PCA, ANN*
- 1970 — *Geometric parameters*
- 1960 — *Man-machine facial recognition*
- 1950 — *Psychology*

- Deep Learning
- Local Features
  - Learning-based
  - Hand-crafted
- Holistic Subspace
- Geometry

History of Face Recognition

2010	*Local feature learning, deep learning*
2000	*Hand-crafted features*
1990	*Holistic subspace learning, graph matching*
1980	*Geometric measurement, PCA, ANN*
1970	*Geometric parameters*
1960	*Man-machine facial recognition*
1950	*Psychology*

- Deep Learning
- Local Features
  - Learning-based
  - Hand-crafted
- Holistic Subspace
- Geometry

History of Face Recognition

- Deep Learning
- Local Features
  - Learning-based
  - Hand-crafted
- Holistic Subspace
- Geometry

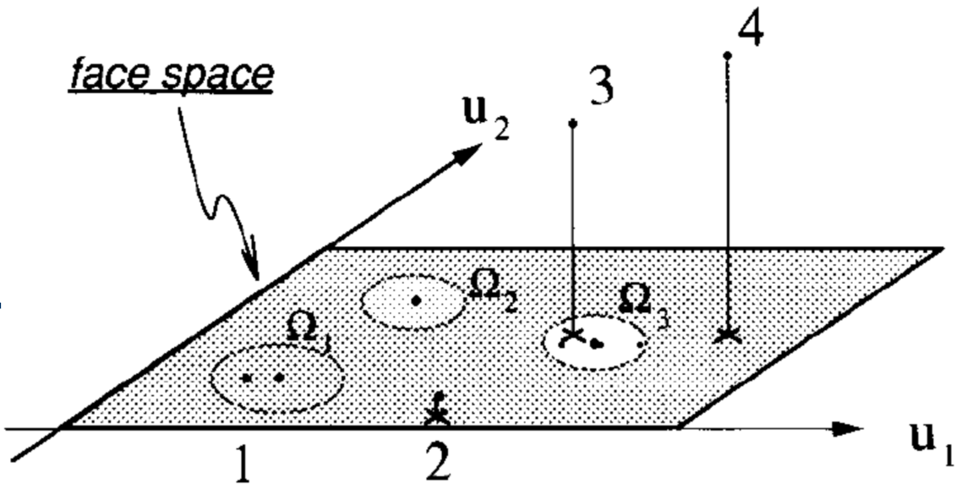
Representation Learning

3.1 Holistic Representation Learning

Eigenfaces

Based on PCA

- A face space best encodes the variation
- The face space is spanned by eigenfaces
- The weights form a vector that describe the contribution of each eigenface in representing the input face image

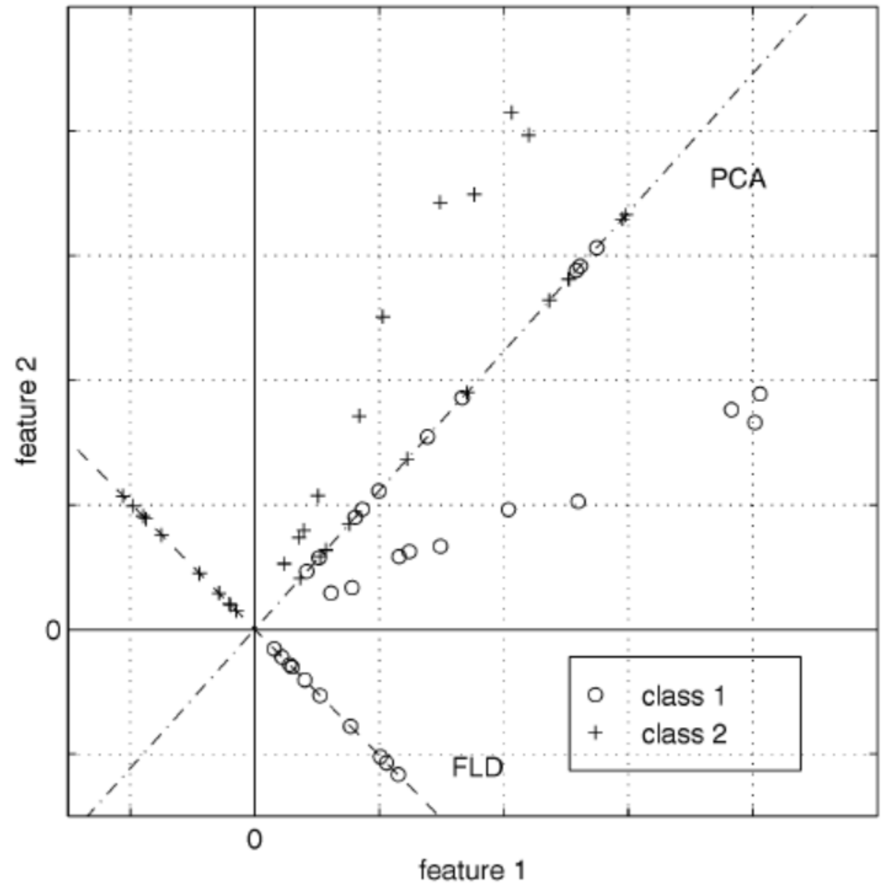


[Turk and Pentland, CVPR'91]

Fisherfaces

Based on Fisher's Linear Discriminant

- Insensitive to large variations in lighting and facial expressions
- Maximizes the ratio of between-class scatter to that of within-class scatter

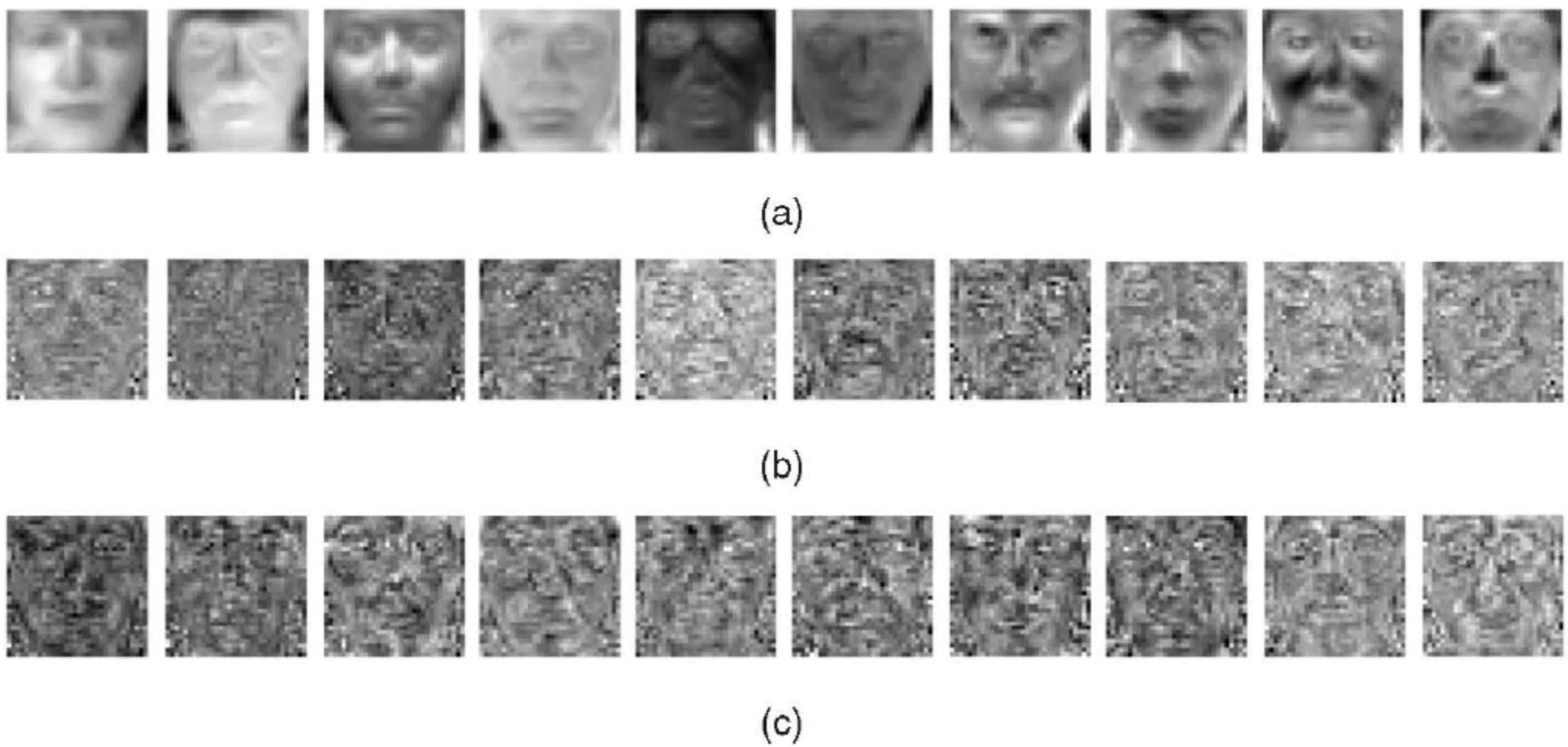


[Belhumeur et al., TPAMI'97]

Laplacianfaces

Based on Locality Preserving Projections

- LPP finds an embedding that preserves local information



(a) Eigenfaces (b) Fisherfaces (c) Laplacianfaces

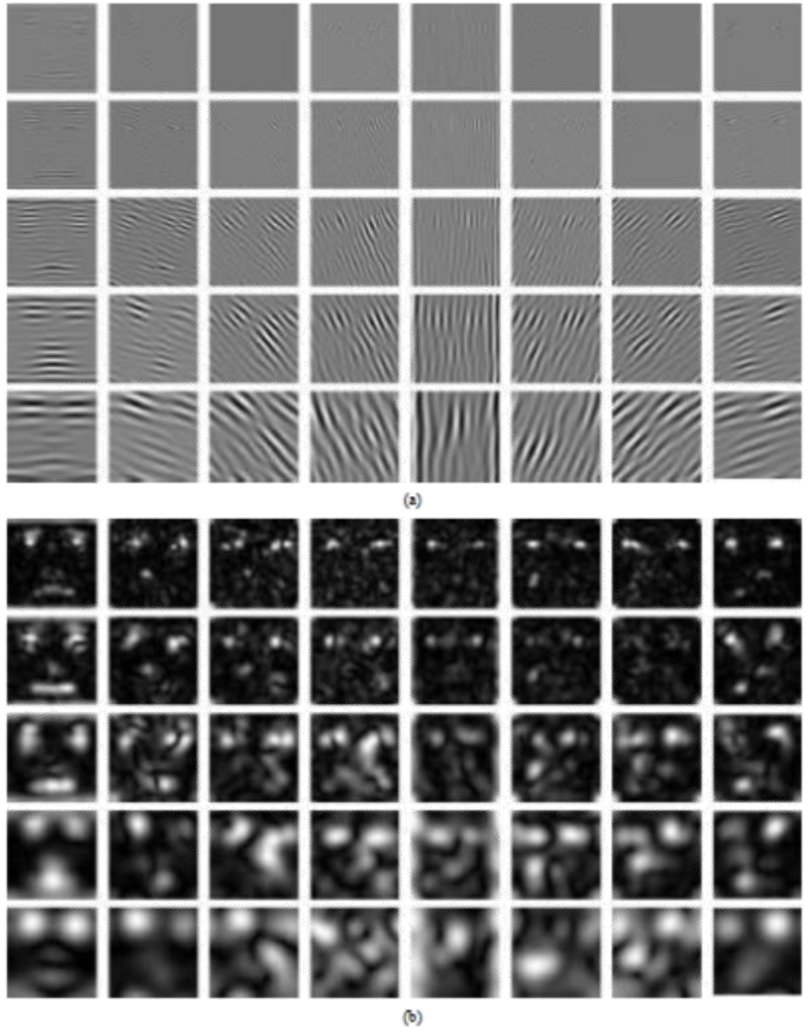
[He et al., TPAMI'05]

3.2 Hand-crafted Representation

Gabor Wavelet Representation

Gabor-Fisher Classifier (GFC)

- Augmented Gabor feature vector
- Gabor-Fisher classifier for multi-class classification

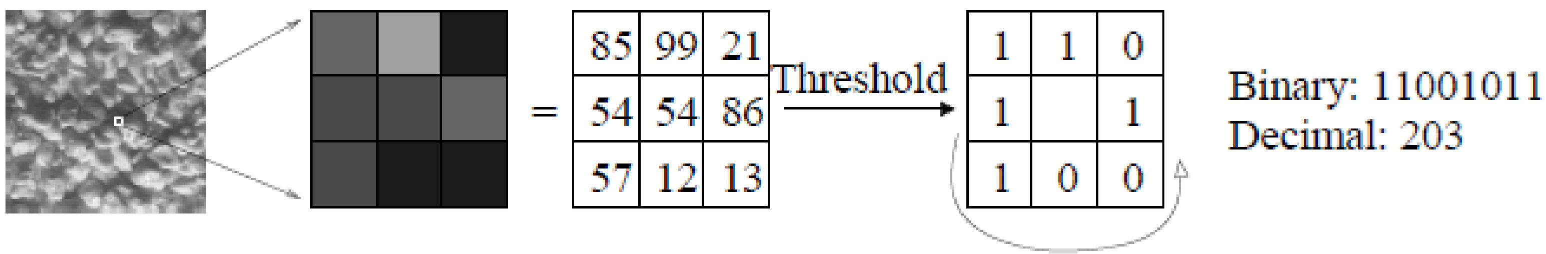


[Liu and Wechsler, TIP'02]

Local Binary Pattern

LBP

- Varying scales
- Uniform patterns



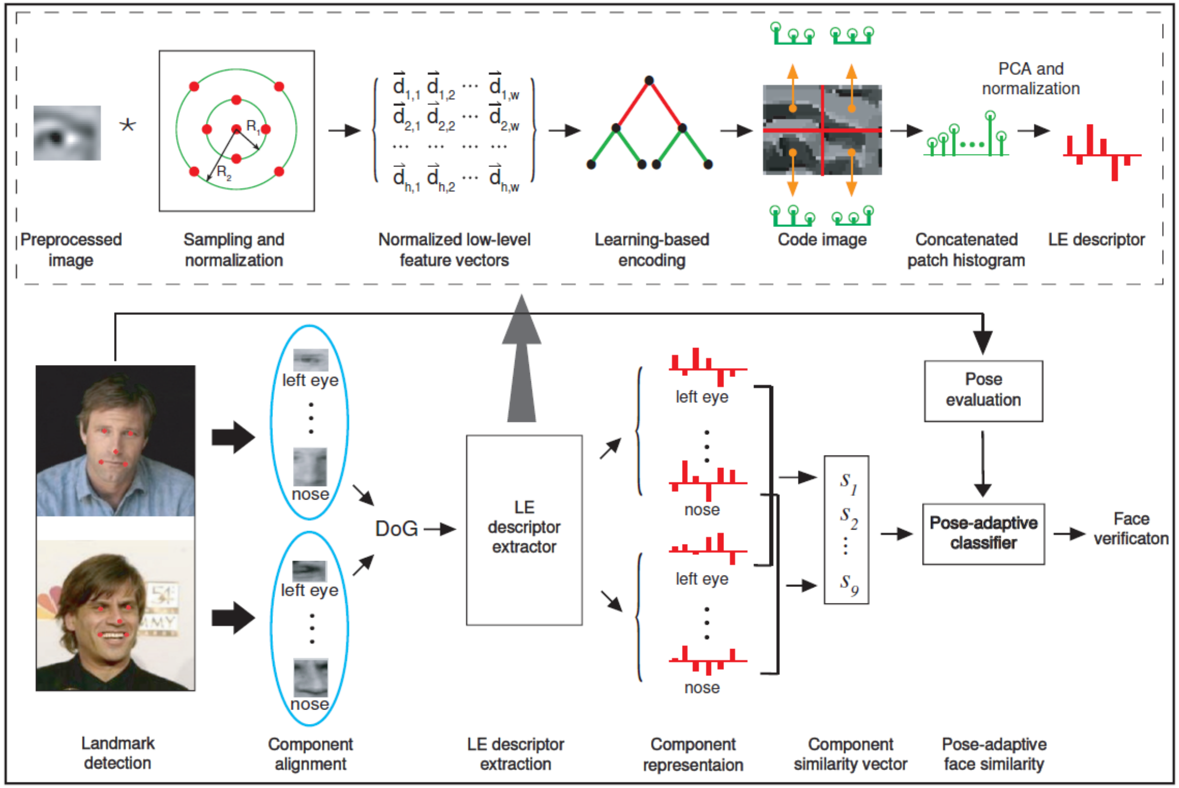
[Ojala et al., TPAMI'02; Ahonen et al., TPAMI'06]

3.3 Local Representation Learning

Learning-based Descriptor

LE

- Learning to encode the local microstructures of the face into a set of discrete codes in an unsupervised manner



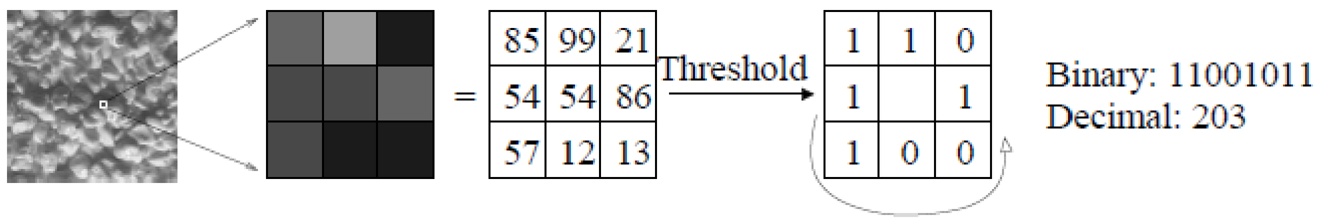
[Cao et al., CVPR'10]

Local Binary Representation Learning

Motivation

- Local Binary Pattern:

  Local Patch → Threshold → Local Binary Representation



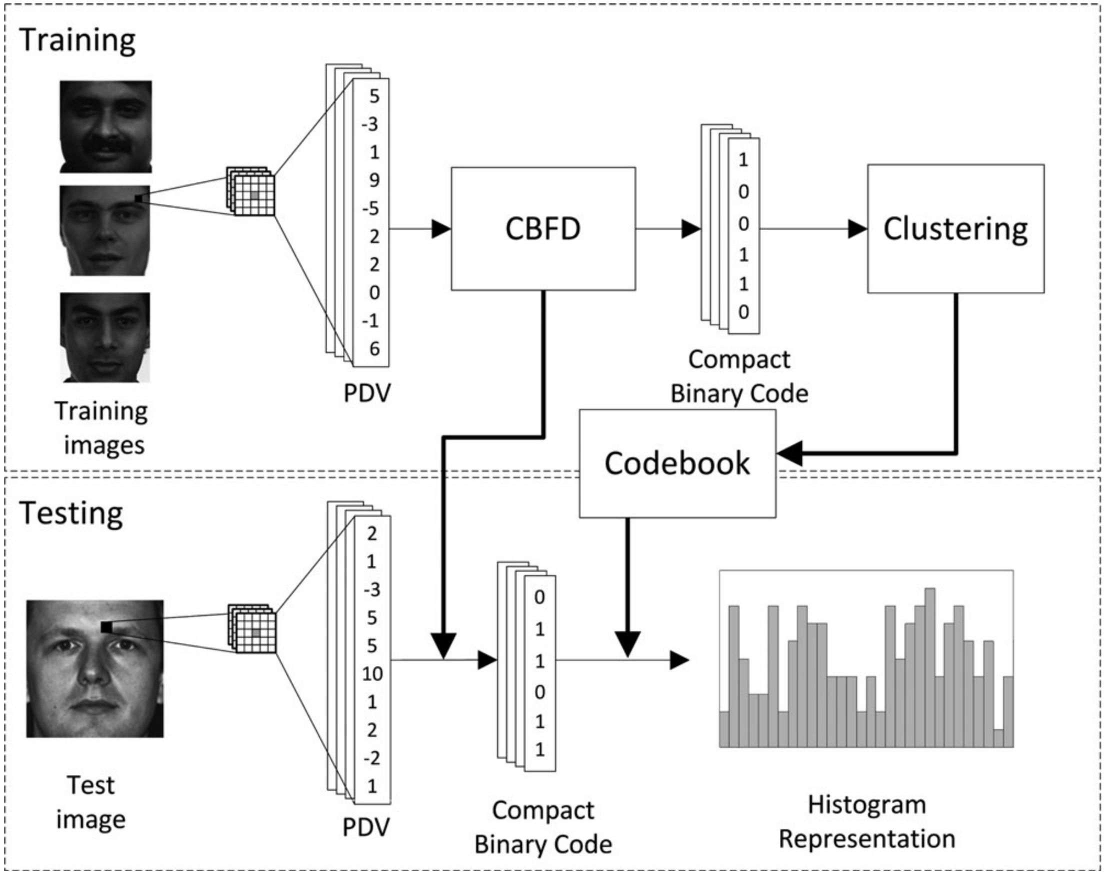
- Local Binary Representation Learning:

  Local Patch → Mapping → Local Binary Representation

From hand-crafted pattern designing to data-dependent representation learning.

Local Binary Representation Learning

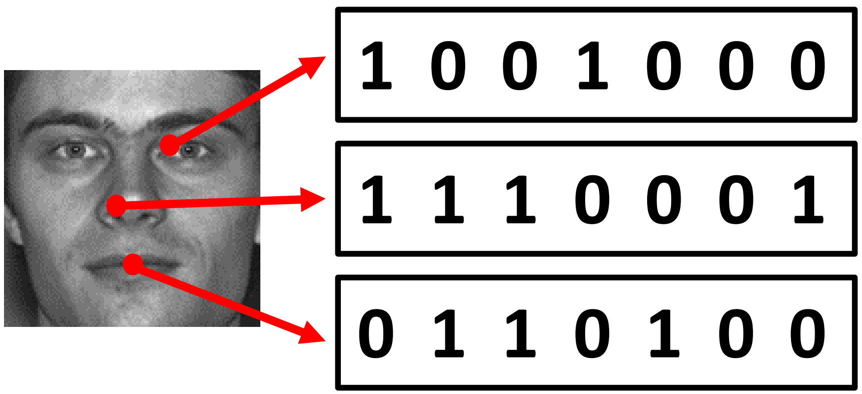
Basic idea: Learn a projection matrix to map local patches to binary features.



Jiwen Lu, Venice Erin Liong, Xiuzhuang Zhou, and Jie Zhou, Learning Compact Binary Face Descriptor for Face Recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 37, no. 10, pp. 2041-2056, 2015.

Local Binary Representation Learning

Compact Binary Face Descriptor (CBFD)

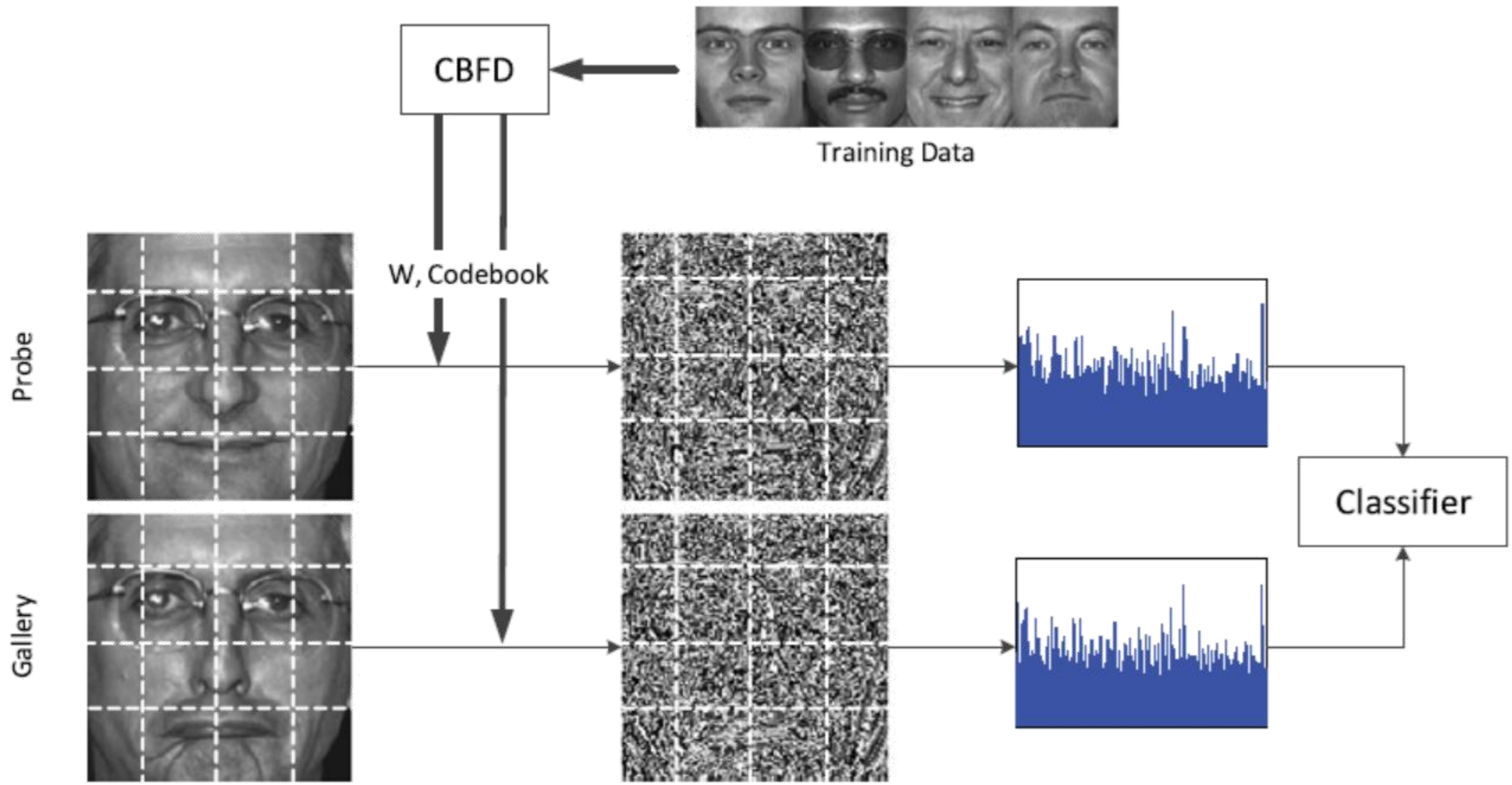


$$\begin{aligned}\min_{w_k} J(w_k) &= J_1(w_k) + \lambda_1 J_2(w_k) + \lambda_2 J_3(w_k) \\ &= -\sum_{n=1}^{N} \|b_{nk} - \mu_k\|^2 \\ &\quad + \lambda_1 \sum_{n=1}^{N} \|(b_{nk} - 0.5) - w_k^T x_n\|^2 \\ &\quad + \lambda_2 \|\sum_{n=1}^{N} (b_{nk} - 0.5)\|^2,\end{aligned}$$

Jiwen Lu, Venice Erin Liong, Xiuzhuang Zhou, and Jie Zhou, Learning Compact Binary Face Descriptor for Face Recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 37, no. 10, pp. 2041-2056, 2015.

Local Binary Representation Learning

Face Representation



Jiwen Lu, Venice Erin Liong, Xiuzhuang Zhou, and Jie Zhou, Learning Compact Binary Face Descriptor for Face Recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI),* vol. 37, no. 10, pp. 2041-2056, 2015.

Experimental Results

Face Recognition

- Dataset---FERET face database

Method	fb	fc	dup1	dup2
LBP [26]	93.0	51.0	61.0	50.0
LBP+WPCA [26]	98.5	84.0	79.4	70.0
LGBP [70]	94.0	97.0	68.0	53.0
LGBP+WPCA [70]	98.1	99.0	83.8	85.0
LVP [41]	97.0	70.0	66.0	50.0
LGT [30]	97.0	90.0	71.0	67.0
HGGP [69]	97.6	98.9	77.7	76.1
HOG [42]	90.0	74.0	54.0	46.6
DT-LBP [39]	99.0	100.0	84.0	80.0
LDP [68]	94.0	83.0	62.0	53.0
GV-LBP-TOP [31]	98.4	99.0	82.0	81.6
DLBP [40]	99.0	99.0	86.0	85.0
GV-LBP [31]	98.1	98.5	80.9	81.2
LQP+WPCA [23]	99.8	94.3	85.5	78.6
POEM [59]	97.0	95.0	77.6	76.2
POEM+WPCA [59]	99.6	99.5	88.8	85.0
s-POEM+WPCA [58]	99.4	100.0	91.7	90.2
DFD [32]	99.2	98.5	85.0	82.9
DFD+WPCA [32]	99.4	100.0	91.8	92.3
CBFD	98.2	100.0	86.1	85.5
CBFD+WPCA	99.8	100.0	93.5	93.2

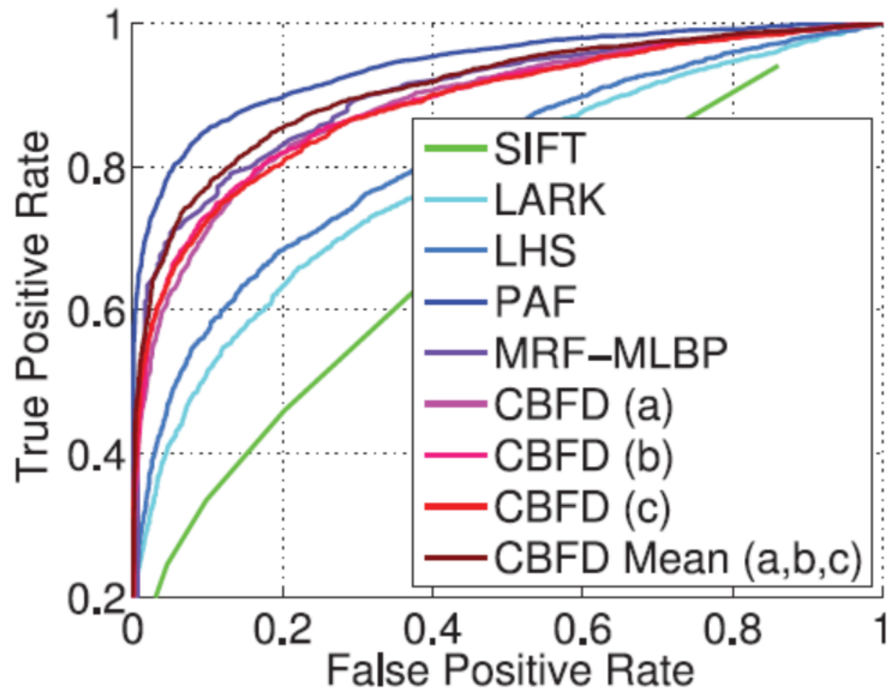
Jiwen Lu, Venice Erin Liong, Xiuzhuang Zhou, and Jie Zhou, Learning Compact Binary Face Descriptor for Face Recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 37, no. 10, pp. 2041-2056, 2015.

Experimental Results

Face Recognition

- Dataset---LFW face database

Method	AUC
LBP [57]	75.47
SIFT [57]	54.07
LARK [48]	78.30
LHS [51]	81.07
PAF [67]	94.05
MRF-MLBP [2]	89.94
CBFD (a)	82.32
CBFD+WPCA (a)	88.75
CBFD+WPCA (b)	88.89
CBFD+WPCA (c)	88.65
CBFD+WPCA (mean: a, b, c)	**90.91**



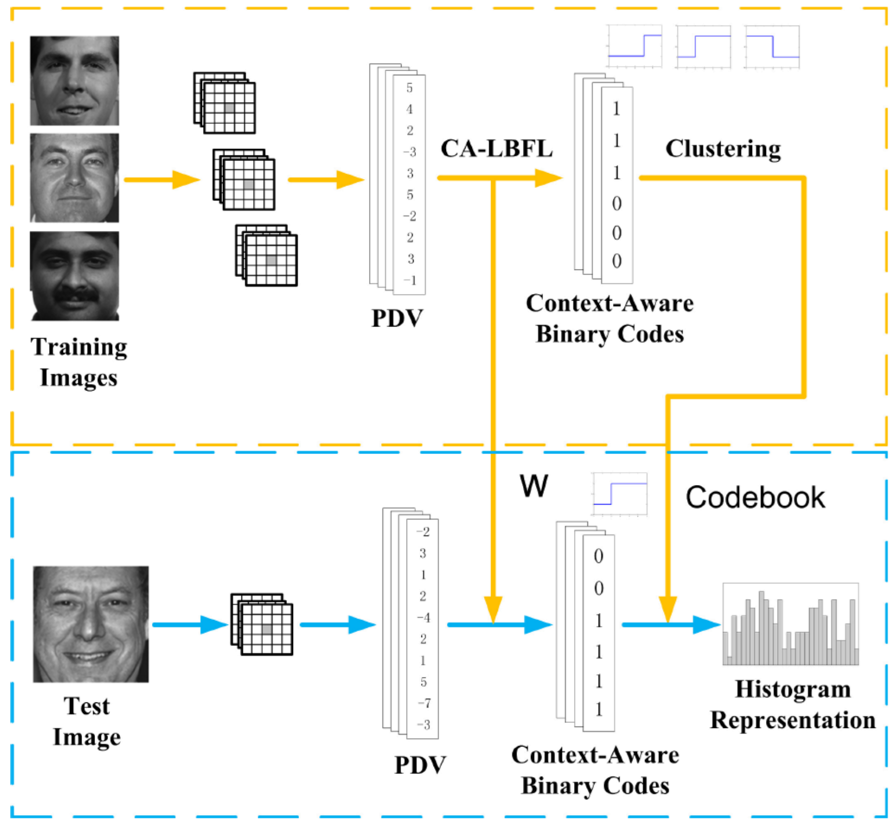
Jiwen Lu, Venice Erin Liong, Xiuzhuang Zhou, and Jie Zhou, Learning Compact Binary Face Descriptor for Face Recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI),* vol. 37, no. 10, pp. 2041-2056, 2015.

Local Binary Representation Learning

Context-Aware Local Binary Feature Learning (CA-LBFL)

- Contextual information is widely exploited in various tasks as prior knowledge

- Learning context-aware features to enhance the robustness

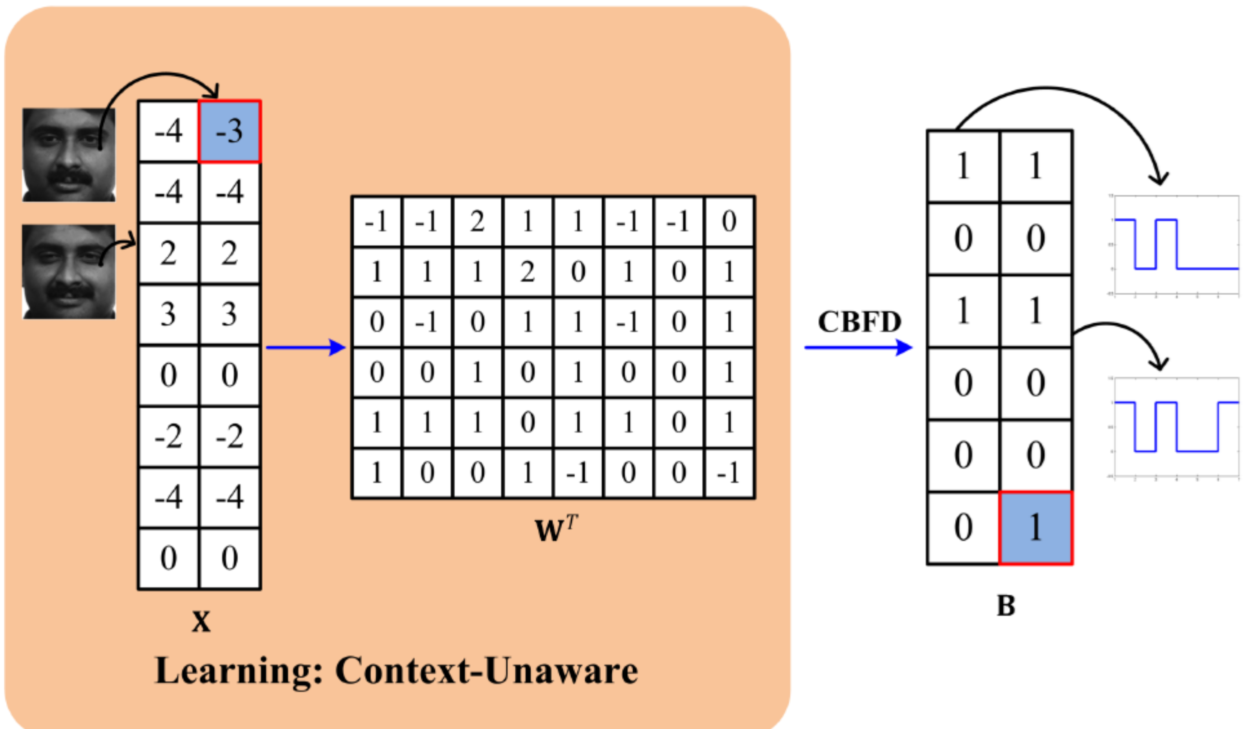


1 0 …… 1 1 0 …… 0 0

Yueqi Duan, **Jiwen Lu**, Jianjiang Feng, and Jie Zhou, Context-Aware Local Binary Feature Learning for Face Recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 40, no. 5, pp. 1139-1153, 2018.

Local Binary Representation Learning

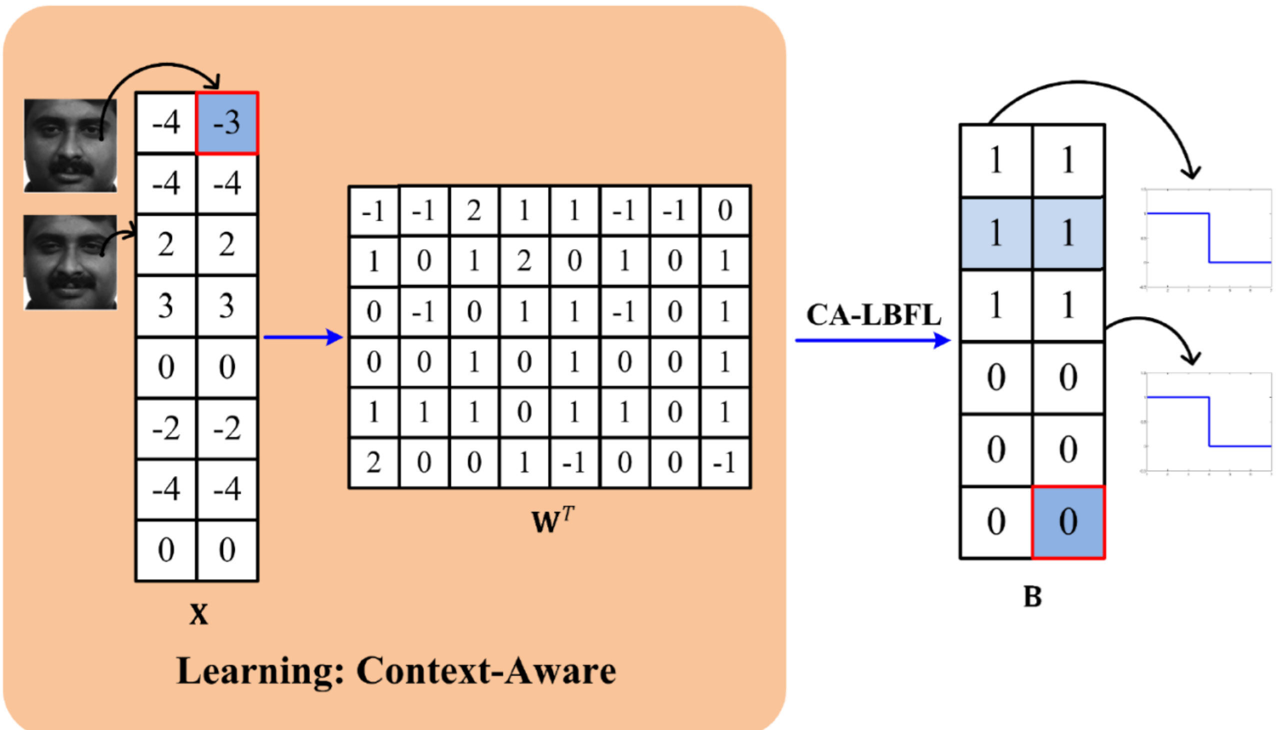
Reducing the number of 0/1 shifts



Yueqi Duan, **Jiwen Lu**, Jianjiang Feng, and Jie Zhou, Context-Aware Local Binary Feature Learning for Face Recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 40, no. 5, pp. 1139-1153, 2018.

Local Binary Representation Learning

Reducing the number of 0/1 shifts



Yueqi Duan, **Jiwen Lu**, Jianjiang Feng, and Jie Zhou, Context-Aware Local Binary Feature Learning for Face Recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 40, no. 5, pp. 1139-1153, 2018.

Experimental Results

Face Recognition

- Dataset---FERET face database

Method	fb	fc	dup1	dup2
LBP [1]	93.0	51.0	61.0	50.0
LGBP [12]	94.0	97.0	68.0	53.0
LGT [6]	97.0	90.0	71.0	67.0
HGGP [11]	97.6	98.9	77.7	76.1
HOG [72]	90.0	74.0	54.0	46.6
LDP [10]	94.0	83.0	62.0	53.0
GV-LBP-TOP [7]	98.4	99.0	82.0	81.6
GV-LBP [7]	98.1	98.5	80.9	81.2
LQP [34]	99.8	94.3	85.5	78.6
POEM [9]	97.0	95.0	77.6	76.2
s-POEM [73]	99.4	100.0	91.7	90.2
DFD [8]	99.4	100.0	91.8	92.3
CBFD [13]	99.8	100.0	93.5	93.2
CA-LBFL (R = 2)	98.5	99.5	91.2	89.3
CA-LBFL (R = 3)	99.8	100.0	94.9	94.5
CA-LBFL (R = 4)	99.8	100.0	95.2	94.9
CA-LBMFL (R = 3)	99.8	100.0	95.3	95.3

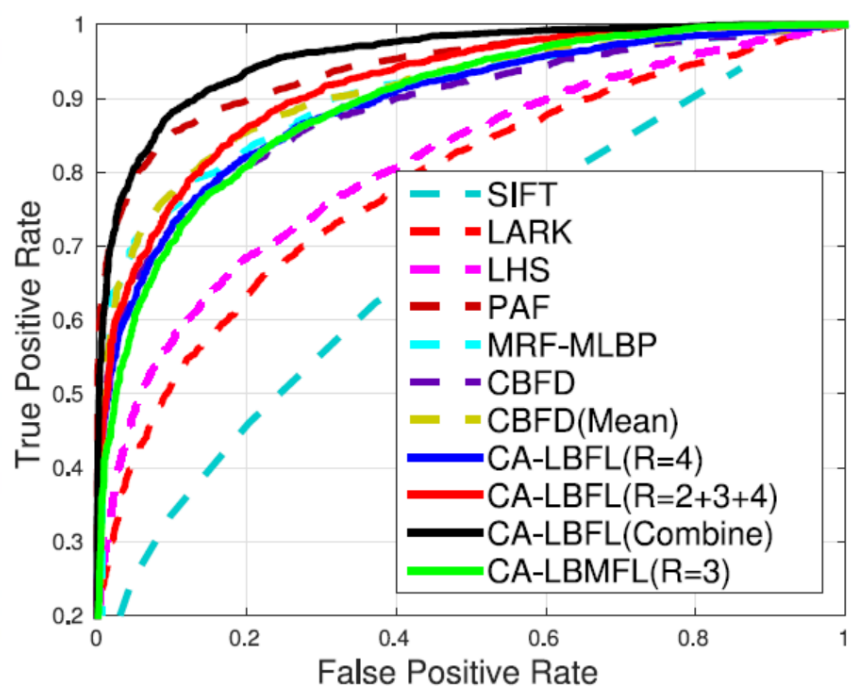
Yueqi Duan, **Jiwen Lu**, Jianjiang Feng, and Jie Zhou, Context-Aware Local Binary Feature Learning for Face Recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 40, no. 5, pp. 1139-1153, 2018.

Experimental Results

Face Recognition

- Dataset---LFW face database

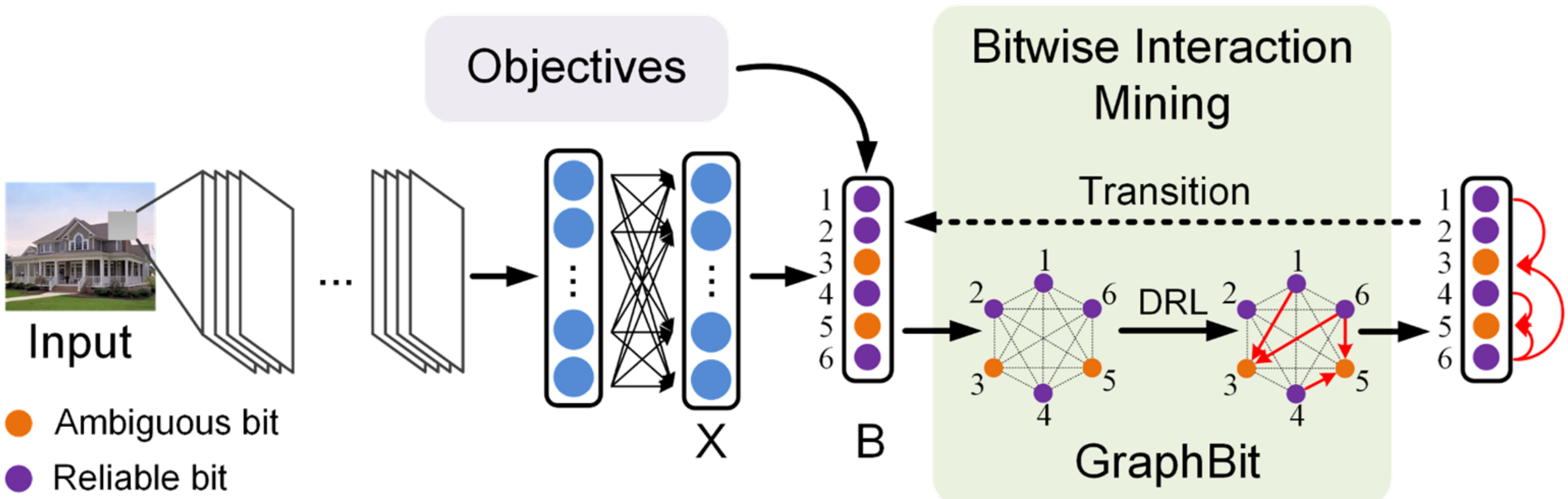
Method	VR	AUC
LBP [1]	69.45	75.47
SIFT [16]	64.10	54.07
LARK [56]	72.23	78.30
POEM [9]	75.22	-
LHS [57]	73.40	81.07
MRF-MLBP [58]	80.08	89.94
PEM (LBP) [59]	81.10	-
PEM (SIFT) [59]	81.38	-
DFD [8]	84.02	-
High-dim LBP [60]	84.08	-
PAF [61]	-	94.05
CBFD [13]	-	88.65
CA-LBFL ($R = 2$)	81.50	86.44
CA-LBFL ($R = 3$)	82.97	88.92
CA-LBFL ($R = 4$)	83.30	89.24
CA-LBFL ($R = 2 + 3 + 4$)	84.72	91.66
CA-LBFL (combine)	86.57	95.67
CA-LBMFL ($R = 3$)	83.22	89.26



Yueqi Duan, **Jiwen Lu**, Jianjiang Feng, and Jie Zhou, Context-Aware Local Binary Feature Learning for Face Recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 40, no. 5, pp. 1139-1153, 2018.

Local Binary Representation Learning

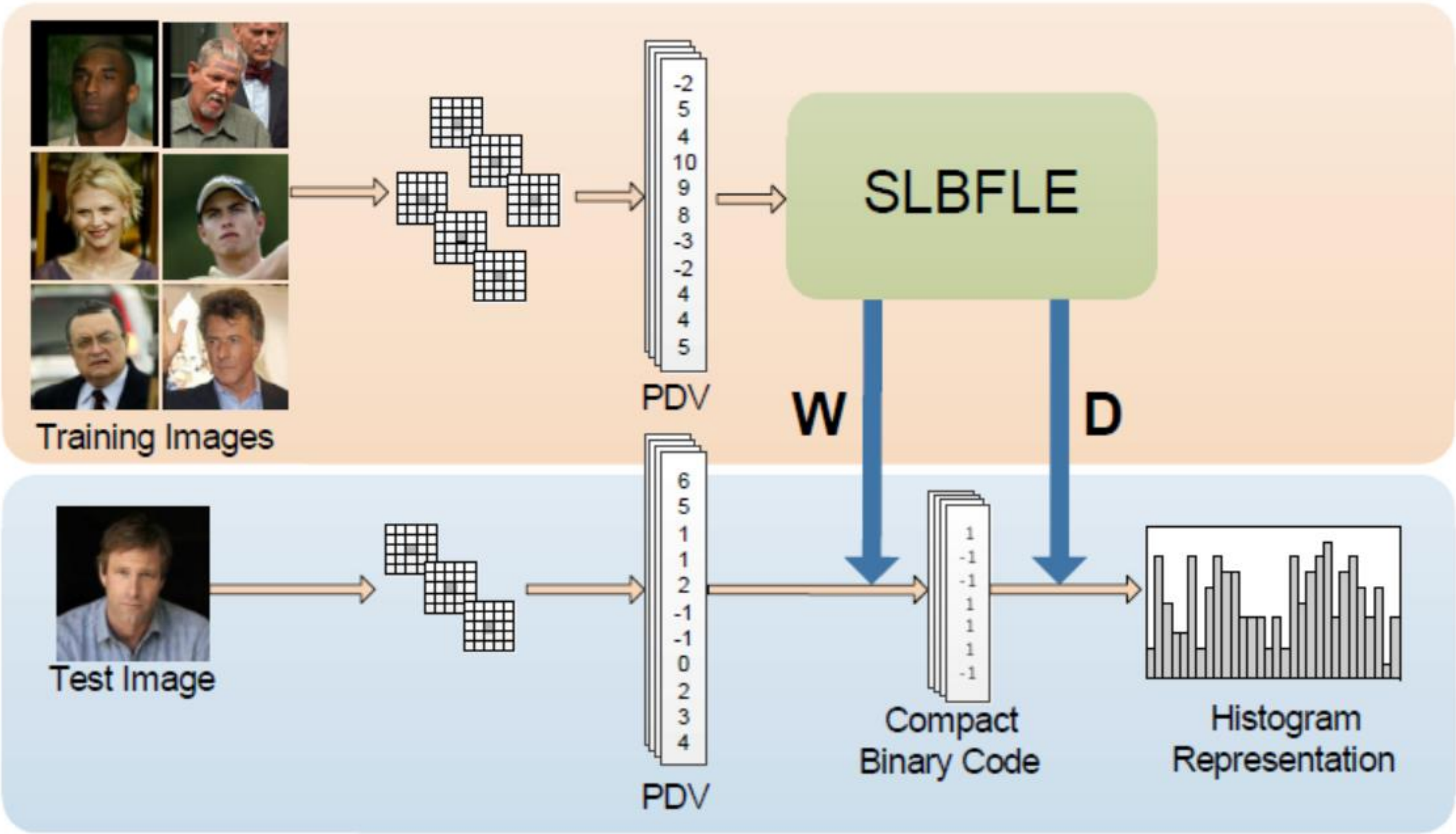
GraphBit: Bitwise Interaction Mining via Deep Reinforcement Learning



Yueqi Duan, Ziwei Wang, **Jiwen Lu**, Xudong Lin, and Jie Zhou, GraphBit: Bitwise Interaction Mining via Deep Reinforcement Learning, *IEEE CVPR*, 2018.

Local Binary Representation Learning

Simultaneous Local Binary Feature Learning and Encoding (SLBFLE)



Jiwen Lu, Venice Erin Liong, and Jie Zhou, Simultaneous Local Binary Feature Learning and Encoding for Homogeneous and Heterogeneous Face Recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2018, accepted.

Local Binary Representation Learning

Simultaneous Local Binary Feature Learning and Encoding (SLBFLE)

$$
\begin{aligned}
\min_{\mathbf{w},\mathbf{D},\alpha} J &= J_1 + \lambda_1 J_2 \\
&= \sum_{n=1}^{N} \left(\|\mathrm{b}_n - \mathbf{D}\alpha_n\|^2 + \gamma\|\alpha_n\|_1 \right) \\
&+ \lambda_1 \sum_{n=1}^{N} \sum_{k=1}^{K} \|\mathrm{b}_{nk} - \mathbf{w}_k^\top \mathbf{x}_n\|^2 \\
\text{subject to} \quad & \|\sum_{n=1}^{N} \mathrm{b}_{nk}\|^2 = 0, \quad \forall\, k \\
& \mathrm{b}_n \mathrm{b}_n^\top = \mathbf{I}^{k\times k}, \quad \forall\, n
\end{aligned}
$$

Jiwen Lu, Venice Erin Liong, and Jie Zhou, Simultaneous Local Binary Feature Learning and Encoding for Homogeneous and Heterogeneous Face Recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2018, accepted.

Experimental Results

Face Recognition

- Dataset---FERET face database

Method	fb	fc	dup1	dup2
LBP [1]	93.0	51.0	61.0	50.0
LGBP [86]	94.0	97.0	68.0	53.0
HGGP [83]	97.6	98.9	77.7	76.1
LDP [82]	94.0	83.0	62.0	53.0
GV-LBP-TOP [35]	98.4	99.0	82.0	81.6
GV-LBP [35]	98.1	98.5	80.9	81.2
LQP [24]	99.8	94.3	85.5	78.6
POEM [68]	97.0	95.0	77.6	76.2
s-POEM [66]	99.4	**100.0**	91.7	90.2
DFD [36]	99.4	**100.0**	91.8	92.3
CBFD [45]	99.8	**100.0**	93.5	**93.2**
SLBFLE (R=2)	99.7	99.7	89.9	80.0
SLBFLE (R=3)	**99.9**	**100.0**	94.5	90.9
SLBFLE (R=4)	**99.9**	**100.0**	**95.2**	**92.7**

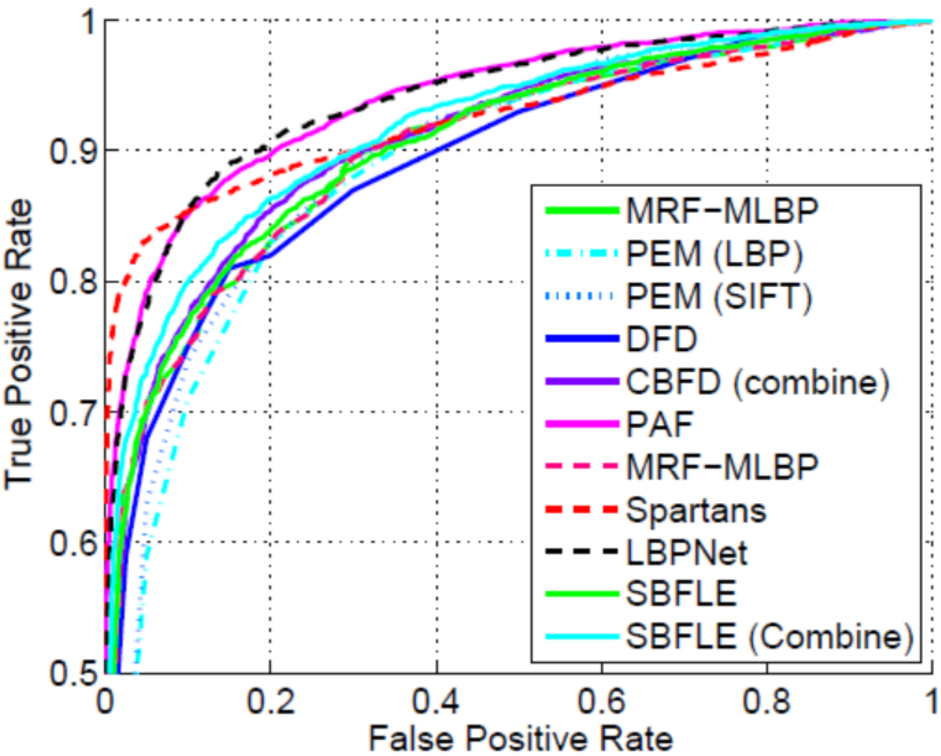
Jiwen Lu, Venice Erin Liong, and Jie Zhou, Simultaneous Local Binary Feature Learning and Encoding for Homogeneous and Heterogeneous Face Recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2018, accepted.

Experimental Results

Face Recognition

- Dataset---LFW face database

Method	VR	AUC
LBP [65]	69.45	75.47
SIFT [65]	64.10	54.07
LARK [55]	72.23	78.30
POEM [67]	75.22	-
LHS [59]	73.40	81.07
MRF-MLBP [2]	80.08	89.94
PEM (LBP) [38]	81.10	-
PEM (SIFT) [38]	81.38	-
DFD [36]	84.02	-
CBFD (combine) [45]	-	90.91
High-dim LBP [10]	84.08	-
PAF [79]	**87.77**	94.05
MRF-Fusion-CSKDA [2]	-	**98.94**
Spartans [28]	-	94.24
LBPNet [77]	-	94.04
SLBFLE (R=2)	82.02	88.95
SLBFLE (R=3)	84.08	90.46
SLBFLE (R=4)	84.18	90.53
SLBFLE (R=2+3+4)	**85.62**	**92.00**



Jiwen Lu, Venice Erin Liong, and Jie Zhou, Simultaneous Local Binary Feature Learning and Encoding for Homogeneous and Heterogeneous Face Recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2018, accepted.

3.4 Deep Representation Learning

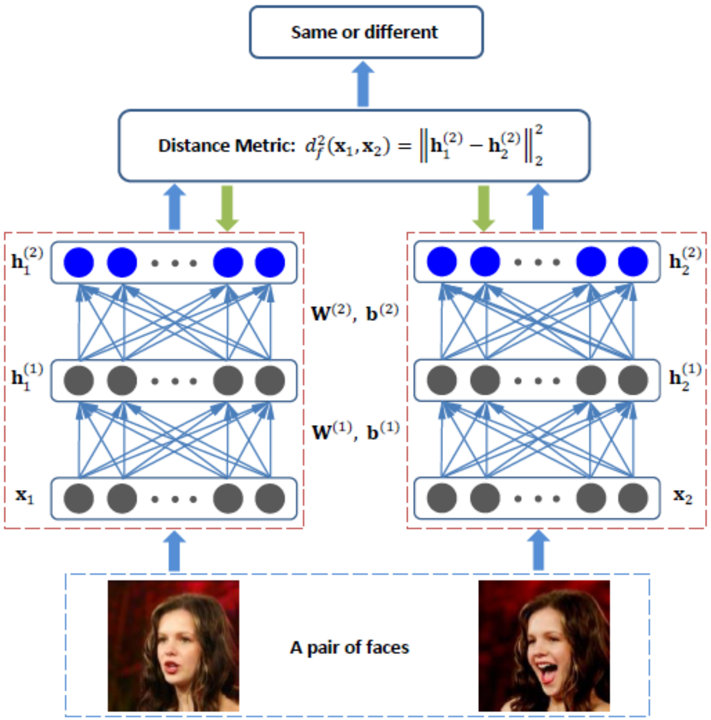
Representative Deep Learning Methods

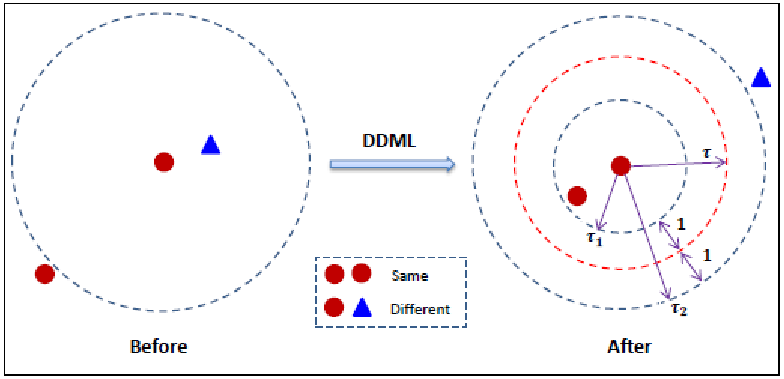
- DDML (CVPR’14, TIP’17)
- DeepFace (CVPR’14)
- DeepID/DeepID2/DeepID2+/DeepID3 (CVPR’14, NIPS’14, CVPR’15, arXiv’15)
- FaceNet (CVPR’15)
- VGG Face (BMVC’15)
- Center Face (ECCV’16)
- Large-Margin Face (ICML’16)
- SphereFace (CVPR’17)
- Range Face (ICCV’17)

Deep Representation Learning

Discriminative Deep Metric Learning (DDML)

- Contrastive loss





$$f(\mathbf{x}) = \mathbf{h}^{(M)} = s\big(\mathbf{W}^{(M)}\mathbf{h}^{(M-1)} + \mathbf{b}^{(M)}\big) \in \mathbb{R}^{p^{(M)}}$$

$$\ell_{ij}\left(\tau - d_f^2(\mathbf{x}_i, \mathbf{x}_j)\right) > 1.$$

$$\begin{aligned}
\arg\min_f \; J &= J_1 + J_2 \\
&= \frac{1}{2}\sum_{i,j} g\Big(1 - \ell_{ij}\left(\tau - d_f^2(\mathbf{x}_i, \mathbf{x}_j)\right)\Big) \\
&+ \frac{\lambda}{2}\sum_{m=1}^{M}\left(\|\mathbf{W}^{(m)}\|_F^2 + \|\mathbf{b}^{(m)}\|_2^2\right)
\end{aligned}$$

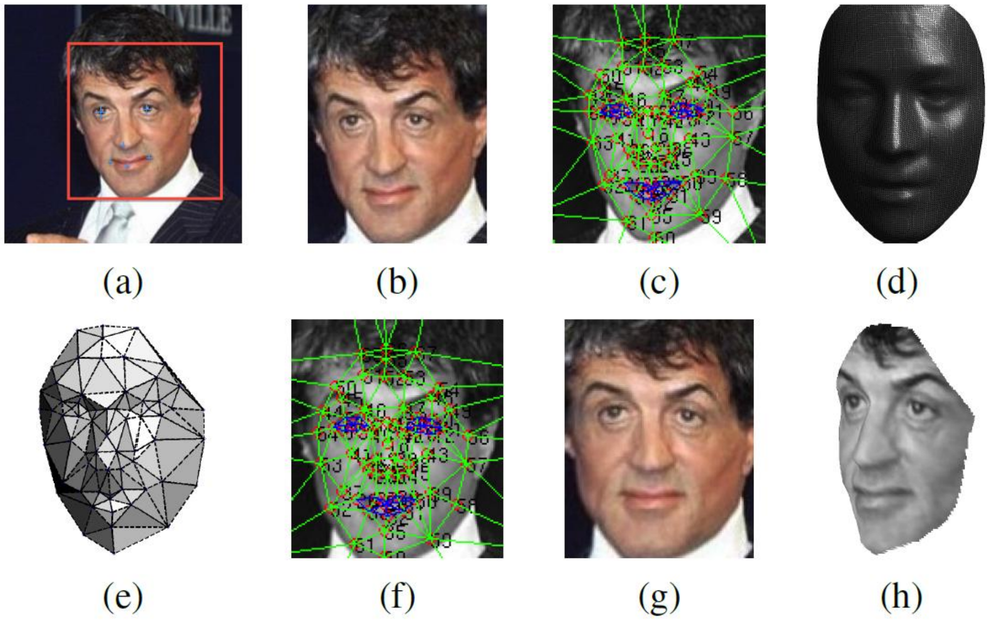
Junlin Hu, **Jiwen Lu**, and Yap-Peng Tan, Discriminative deep metric learning for face verification in the wild, *IEEE CVPR*, pp. 1875-1882, 2014.

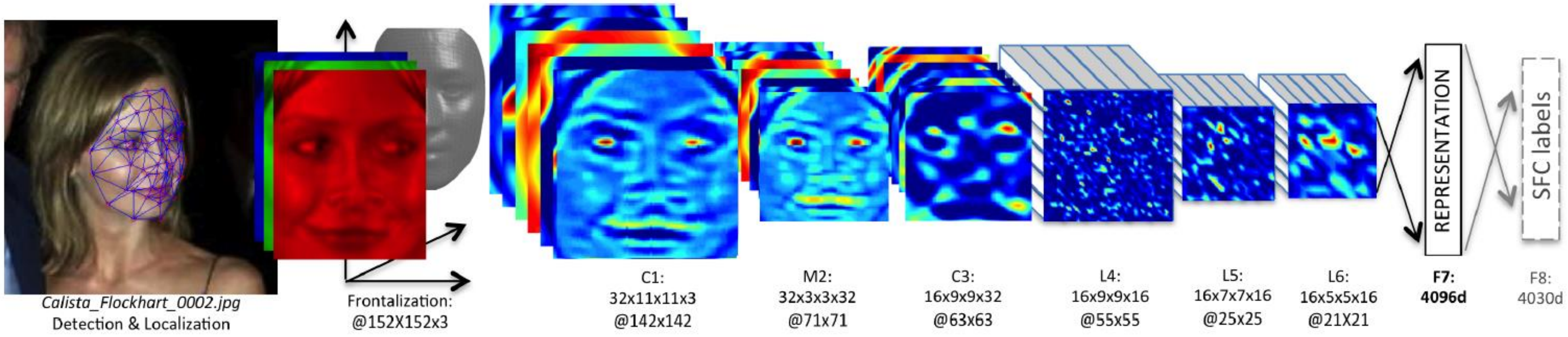
Jiwen Lu, Junlin Hu, and Yap-Peng Tan, Discriminative deep metric learning for face and kinship verification, *IEEE Transactions on Image Processing (TIP)*, vol. 26, no. 9, pp. 4269-4282, 2017.

Deep Representation Learning

DeepFace

- Softmax loss
- 3D face alignment



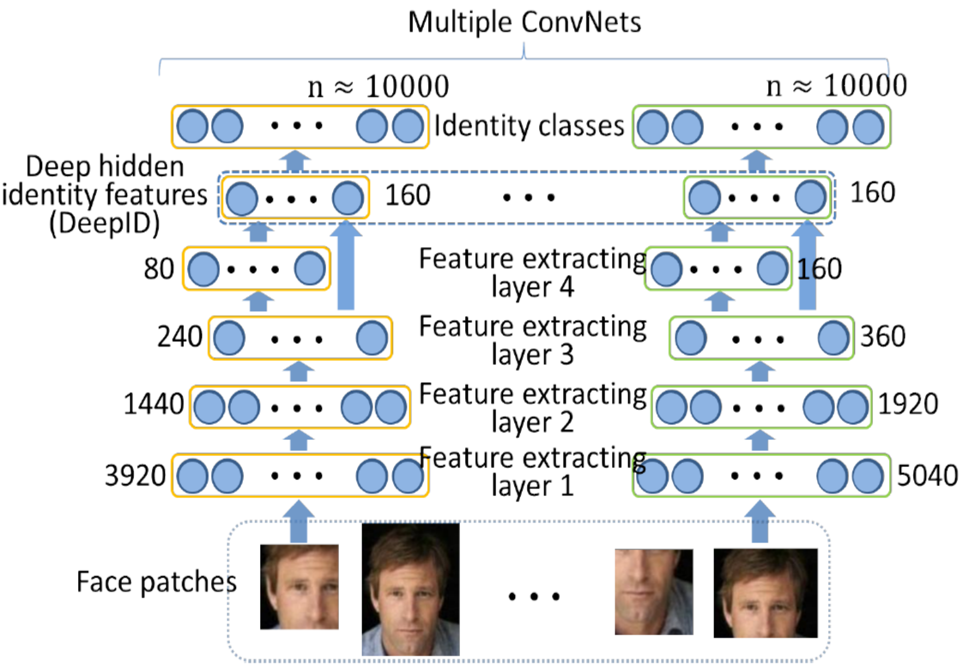


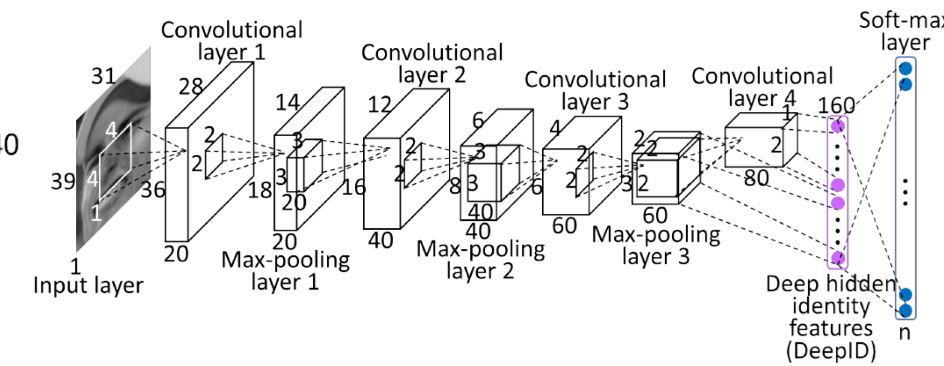
[Taigman et al., CVPR'14]

Deep Representation Learning

DeepID

- Softmax (+ Contrastive) loss
- Multiple CNNs: 60 ConvNets





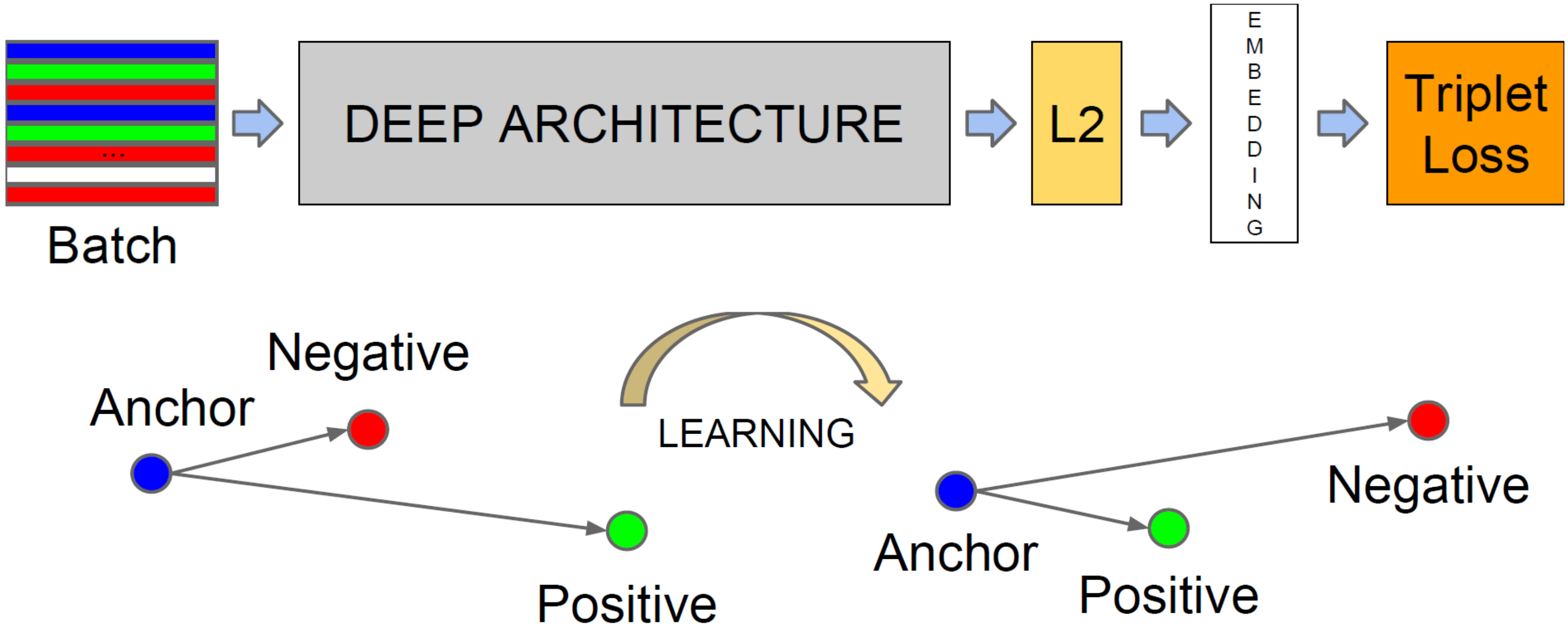
[Sun et al., CVPR'14]

Deep Representation Learning

FaceNet

- Triplet loss
- Semi-hard negative mining

$$\sum_{i}^{N}\left[\|f(x_i^a)-f(x_i^p)\|_2^2-\|f(x_i^a)-f(x_i^n)\|_2^2+\alpha\right]_+$$



[Schroff et al., CVPR'15]

Deep Representation Learning

VGG Face

- Softmax loss
- Dataset collection: 2.6M images with 2,622 identities
- "Very deep"

layer	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
type	input	conv	relu	conv	relu	mpool	conv	relu	conv	relu	mpool	conv	relu	conv	relu	conv	relu	mpool	conv
name	–	conv1_1	relu1_1	conv1_2	relu1_2	pool1	conv2_1	relu2_1	conv2_2	relu2_2	pool2	conv3_1	relu3_1	conv3_2	relu3_2	conv3_3	relu3_3	pool3	conv4_1
support	–	3	1	3	1	2	3	1	3	1	2	3	1	3	1	3	1	2	3
filt dim	–	3	–	64	–	–	64	–	128	–	–	128	–	256	–	256	–	–	256
num filts	–	64	–	64	–	–	128	–	128	–	–	256	–	256	–	256	–	–	512
stride	–	1	1	1	1	2	1	1	1	1	2	1	1	1	1	1	1	2	1
pad	–	1	0	1	0	0	1	0	1	0	0	1	0	1	0	1	0	0	1

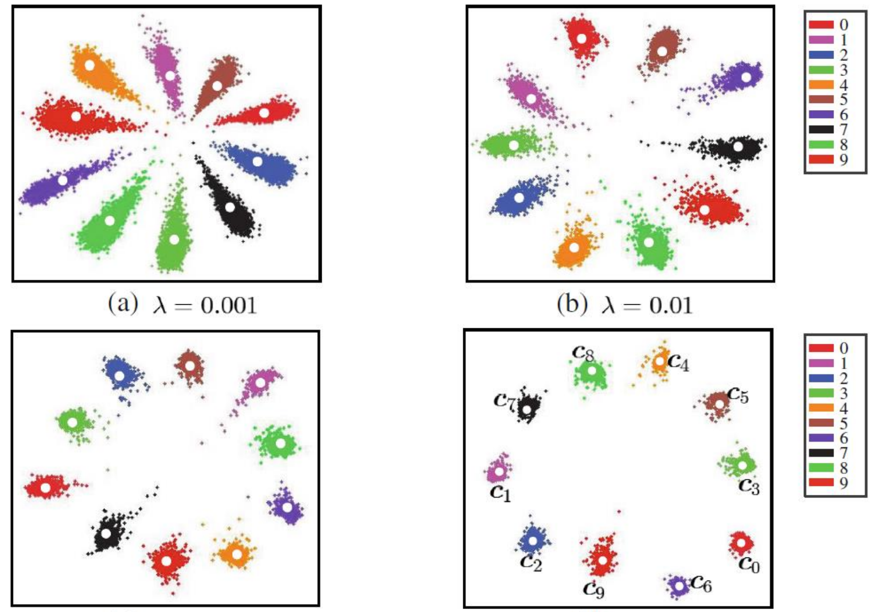
layer	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37
type	relu	conv	relu	conv	relu	mpool	conv	relu	conv	relu	conv	relu	mpool	conv	relu	conv	relu	conv	softmx
name	relu4_1	conv4_2	relu4_2	conv4_3	relu4_3	pool4	conv5_1	relu5_1	conv5_2	relu5_2	conv5_3	relu5_3	pool5	fc6	relu6	fc7	relu7	fc8	prob
support	1	3	1	3	1	2	3	1	3	1	3	1	2	7	1	1	1	1	1
filt dim	–	512	–	512	–	–	512	–	512	–	512	–	–	512	–	4096	–	4096	–
num filts	–	512	–	512	–	–	512	–	512	–	512	–	–	4096	–	4096	–	2622	–
stride	1	1	1	1	1	2	1	1	1	1	1	1	2	1	1	1	1	1	1
pad	0	1	0	1	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0

[Parkhi et al., BMVC'15]

Deep Representation Learning

Center Face

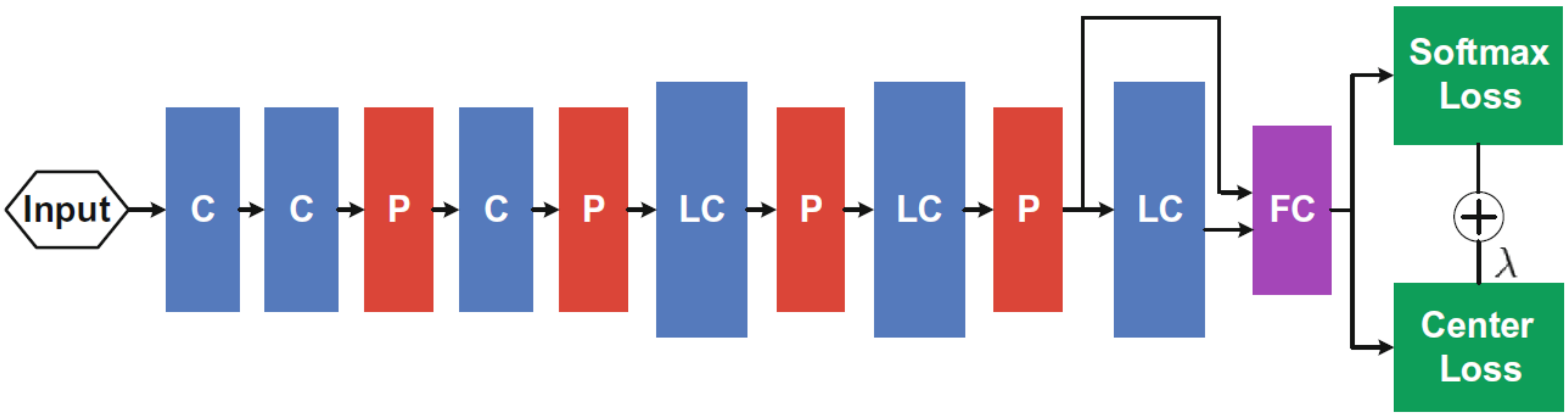
- Softmax loss + Center loss
- Intra-class variations



(a) $\lambda = 0.001$ (b) $\lambda = 0.01$ (c) $\lambda = 0.1$ (d) $\lambda = 1$

$$\mathcal{L} = \mathcal{L}_S + \lambda \mathcal{L}_C$$

$$= -\sum_{i=1}^{m} \log \frac{e^{W_{y_i}^T \boldsymbol{x}_i + b_{y_i}}}{\sum_{j=1}^{n} e^{W_j^T \boldsymbol{x}_i + b_j}} + \frac{\lambda}{2} \sum_{i=1}^{m} \|\boldsymbol{x}_i - \boldsymbol{c}_{y_i}\|_2^2$$

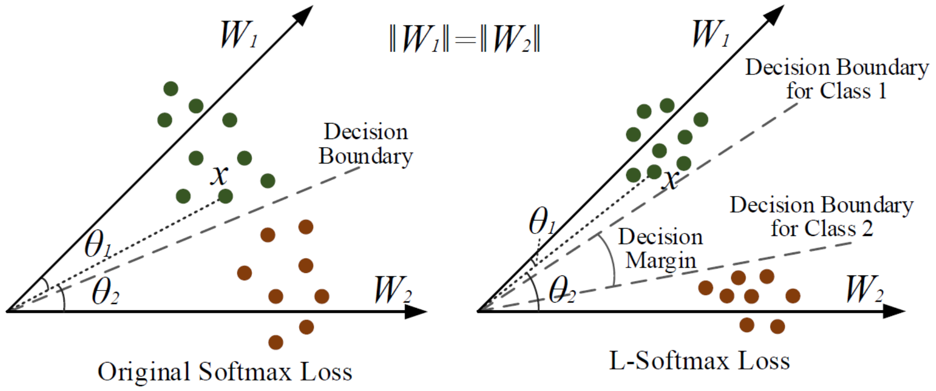


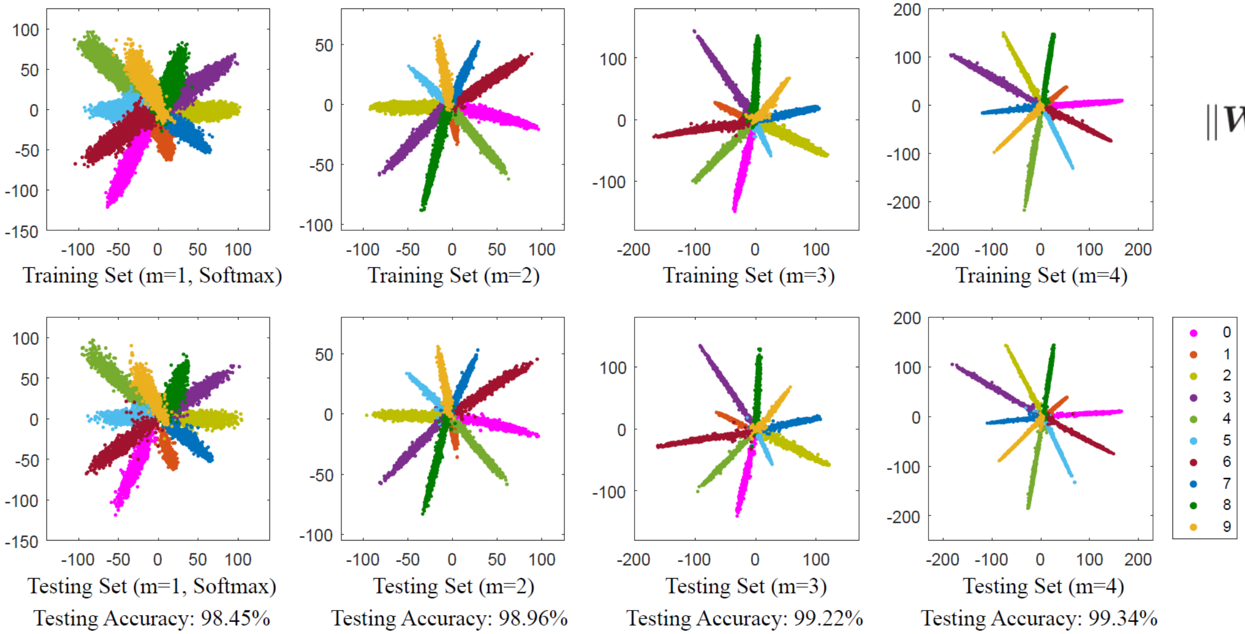
[Wen et al., ECCV'16]

Deep Representation Learning

Large-Margin Face

- L-Softmax loss
- Potentially larger angular separability





$$\|\boldsymbol{W}_1\|\|\boldsymbol{x}\|\cos(\theta_1) \geq \|\boldsymbol{W}_1\|\|\boldsymbol{x}\|\cos(m\theta_1) > \|\boldsymbol{W}_2\|\|\boldsymbol{x}\|\cos(\theta_2).$$

[Liu et al., ICML'16]

Deep Representation Learning

SphereFace

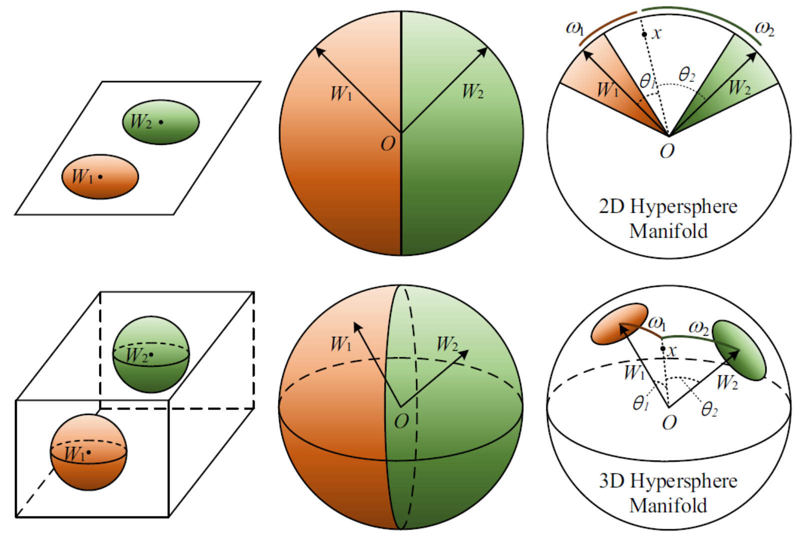
- A-Softmax loss

Modified Softmax loss:

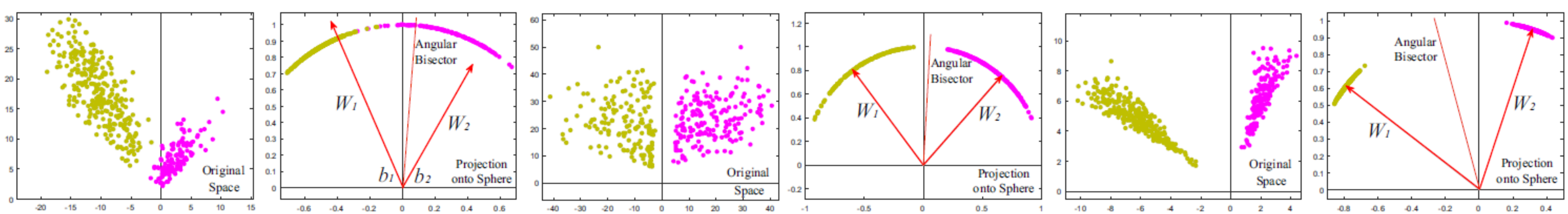
$$L_{\text{modified}} = \frac{1}{N}\sum_i -\log\left(\frac{e^{\|\boldsymbol{x}_i\|\cos(\theta_{y_i,i})}}{\sum_j e^{\|\boldsymbol{x}_i\|\cos(\theta_{j,i})}}\right)$$

A-Softmax loss:

$$L_{\text{ang}} = \frac{1}{N}\sum_i -\log\left(\frac{e^{\|\boldsymbol{x}_i\|\psi(\theta_{y_i,i})}}{e^{\|\boldsymbol{x}_i\|\psi(\theta_{y_i,i})} + \sum_{j\neq y_i} e^{\|\boldsymbol{x}_i\|\cos(\theta_{j,i})}}\right)$$



Euclidean Margin Loss | Modified Softmax Loss | A-Softmax Loss (m≥2)



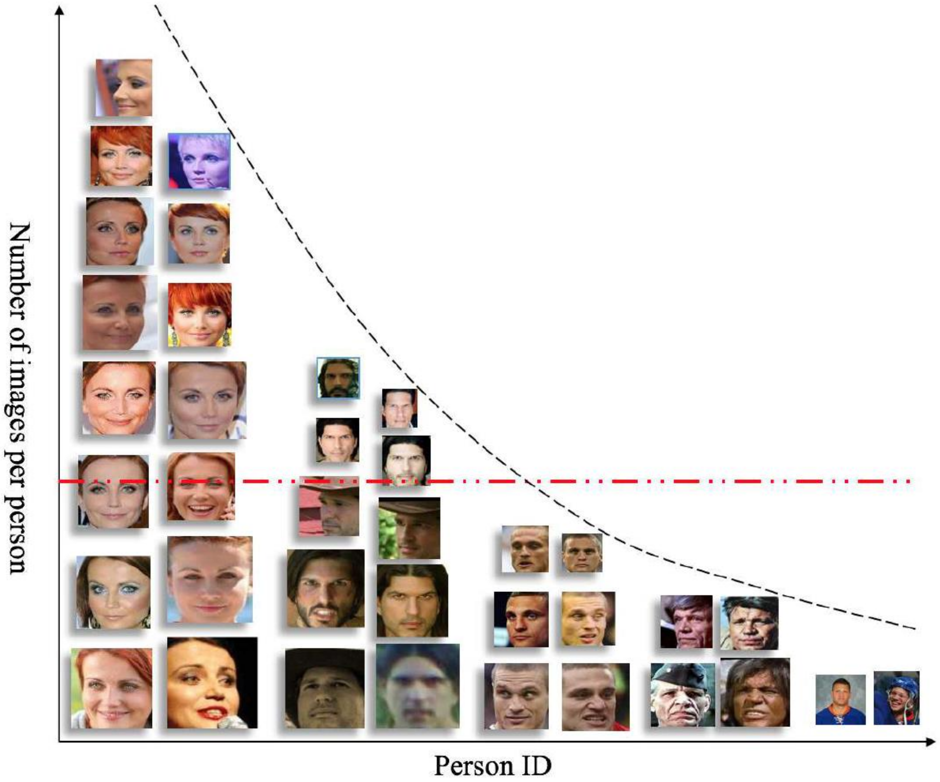
(a) Original Softmax Loss (b) Original Softmax Loss (c) Modified Softmax Loss (d) Modified Softmax Loss (e) A-Softmax Loss (f) A-Softmax Loss

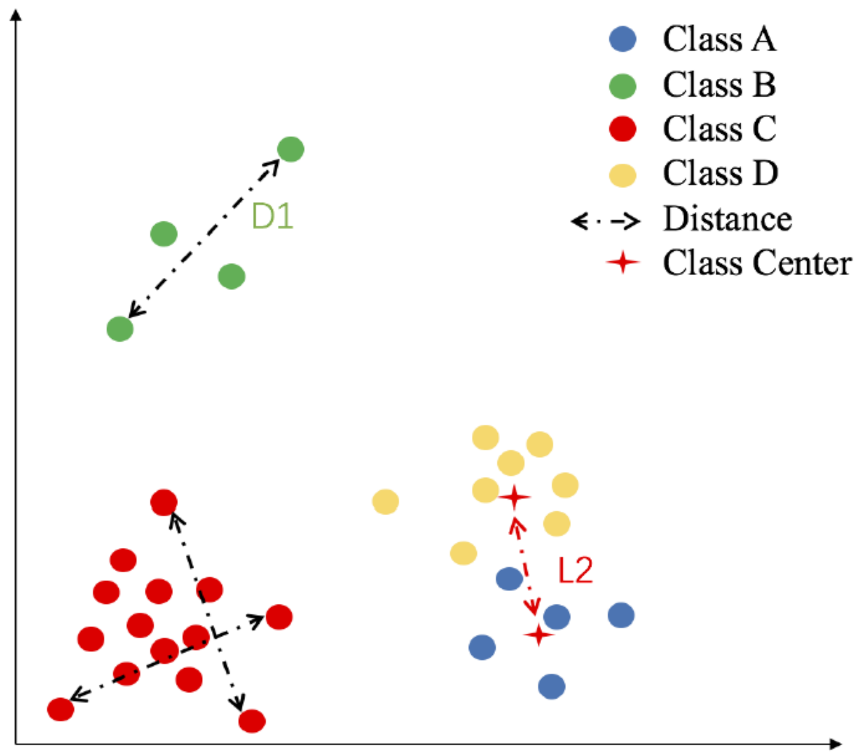
[Liu et al., CVPR'17]

Deep Representation Learning

Range Face

- Range loss
- Effectively utilizing the tailed data in training process





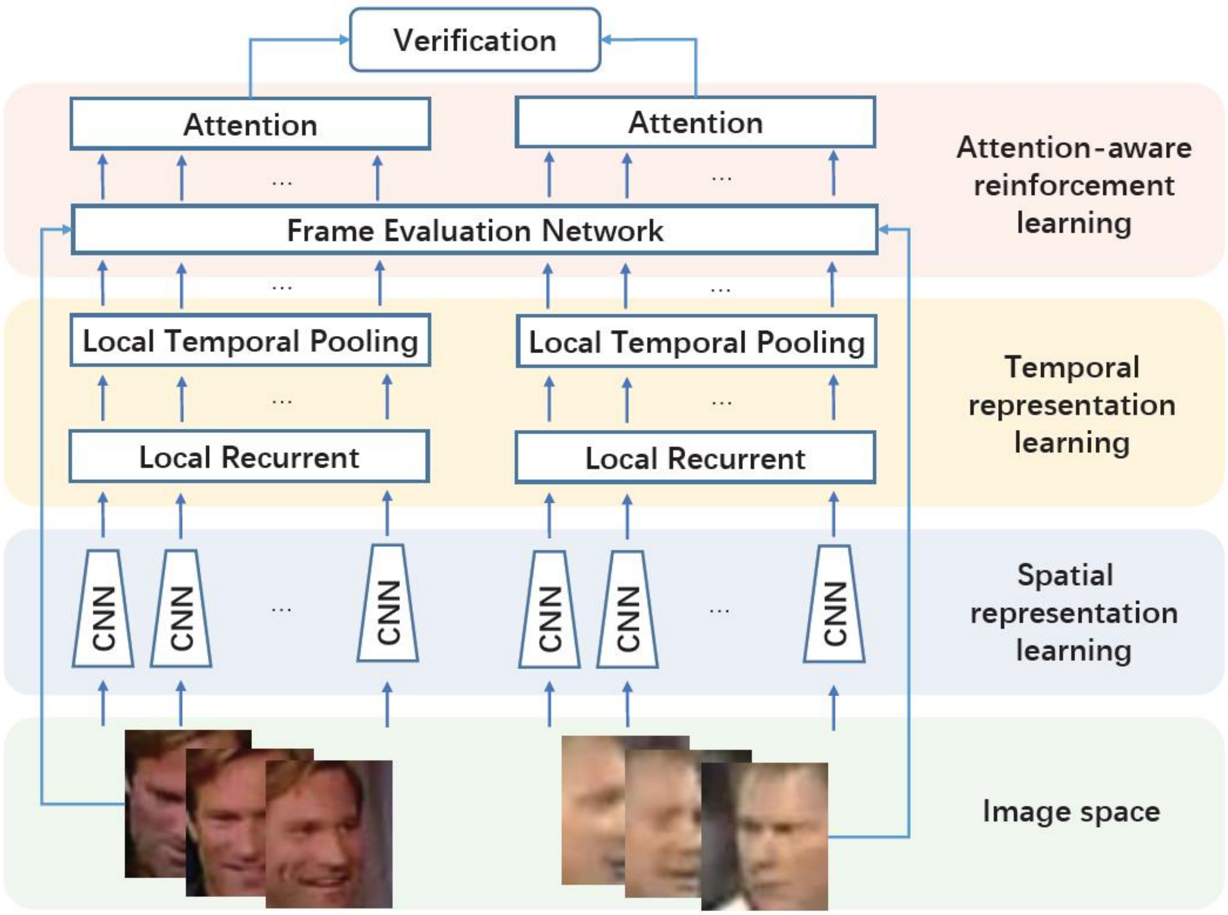
[Zhang et al., ICCV'17]

3.5 More Face Recognition Tasks

Video-based Representation Learning

Attention-aware Deep Reinforcement Learning (ADRL)

- Attention frames selection

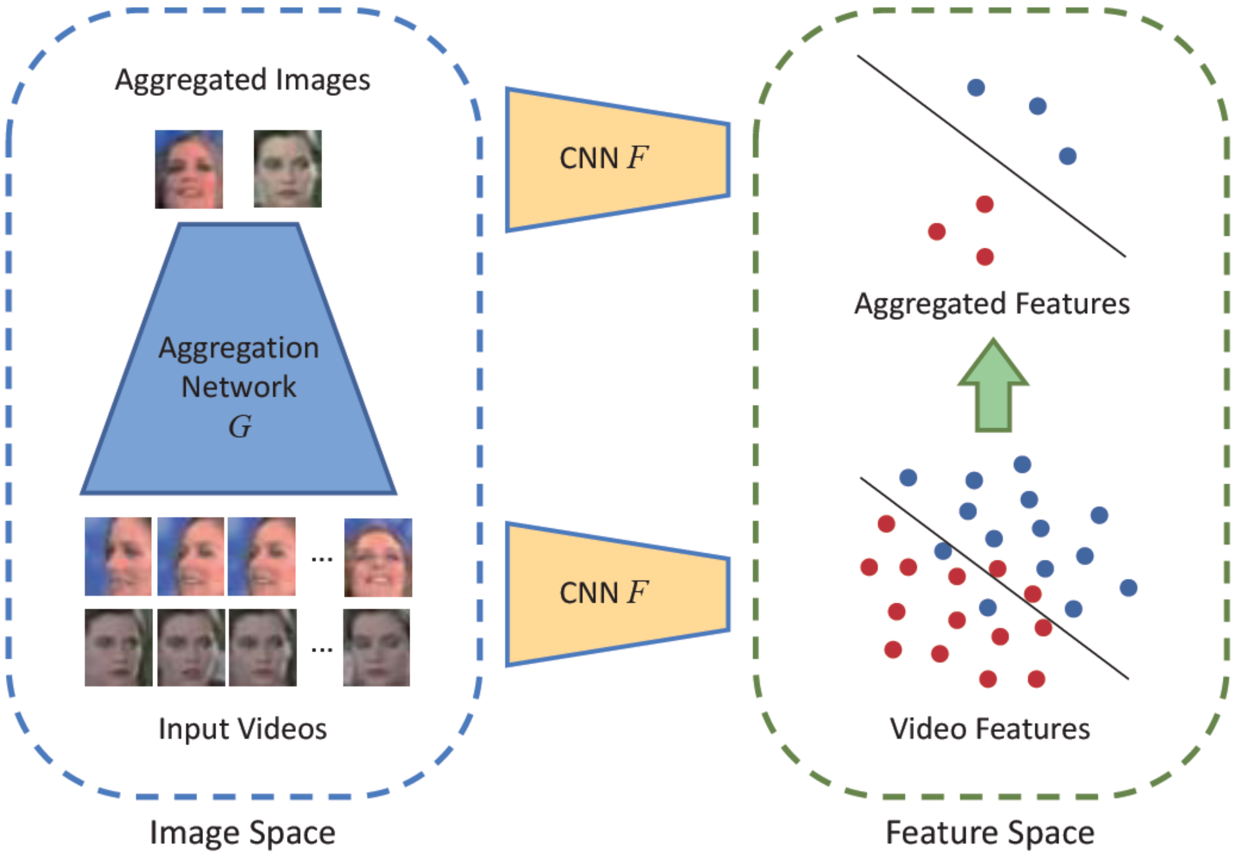


Yongming Rao, **Jiwen Lu**, and Jie Zhou, Attention-aware deep reinforcement learning for video face recognition, *IEEE ICCV*, pp. 3731-3740, 2017.

Video-based Representation Learning

Discriminative Aggregation Network (DAN)

- Aggregated images generation



Yongming Rao, Ji Lin, **Jiwen Lu**, and Jie Zhou, Learning discriminative aggregation network for video-based face recognition, *IEEE ICCV*, pp. 3781-3790, 2017.

Cross-Modal Representation Learning

Motivation

- Cross-modal face matching suffer from large intra-class variations
  - CASIA NIR-VIS 2.0

  - CUFSF

Cross-Modal Representation Learning

Basic idea:

- Modality-invariant feature extraction
- Image synthesis
- Common space projection
  - CDFE [Lin and Tang, ECCV'06]
  - CCA [Yi et al., ICB'07]
  - CSR [Lei and Li, CVPR'09]
  - CMML [Mignon and Jurie, ACCV'12]
  - MvDA [Kan et al., TPAMI'16]
  - MvML [Hu et al., TPAMI'18]

Kinship Verification

- KinFaceW-I: 500 kinship image face pairs

Kinship Verification

- KinFaceW-II:1000 kinship image face pairs

Kinship Verification

- Baseline Results

Feature	F-S	F-D	M-S	M-D	Mean
LBP	62.7	60.2	54.4	61.4	59.7
LE	66.1	59.1	58.9	68.0	63.0
SIFT	65.5	59.0	55.5	55.4	58.8
TPLBP	56.3	60.5	56.0	62.2	58.7

Correct verification accuracy on the KinFaceW-I dataset.

Feature	F-S	F-D	M-S	M-D	Mean
LBP	64.0	63.5	62.8	63.0	63.3
LE	69.8	66.1	72.8	72.0	69.9
SIFT	60.0	56.9	54.8	55.4	56.8
TPLBP	64.4	60.6	60.8	62.9	62.2

Correct verification accuracy on the KinFaceW-II dataset.

Jiwen Lu, Junlin Hu, Xiuzhuang Zhou, Yuanyuan Shang, Yap-Peng Tan, and Gang Wang, Neighborhood repulsed metric learning for kinship verification, *IEEE CVPR*, pp. 2594-2601, 2012.

Jiwen Lu, Xiuzhuang Zhou, Yap-Peng Tan, Yuanyuan Shang, and Jie Zhou, Neighborhood repulsed metric learning for kinship verification, *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 36, no. 2, pp. 331-345, 2014.

Webpage

- www.kinfacew.com

KinFaceW

Home Datasets Protocol Download Results References Contact Changes

Home

Welcome to Kinship Face in the Wild (**KinFaceW**), a database of face images collected for studying the problem of kinship verification from unconstrained face images. There are many potential applications for kinship verification such as family album organization, genealogical research, missing family members search, and social media analysis.

The aim of kinship verification is to determine whether there is a kin relation between a pair of given face images. The kinship is defined as a relationship between two persons who are biologically related with overlapping genes. Hence, there are four representative types of kin relations: Father-Son (F-S), Father-Daughter (F-D), Mother-Son (M-S) and Mother-Daughter (M-D), respectively.

News!

Sep-22-2014: The detailed information of The Kinship Verification in the Wild Evaluation can be found here, which is organized as part of FG2015.

Media Coverage

NewScientist Tech

search New Scientist Go»

Home News In-Depth Articles Blogs Opinion TV Galleries Topic Guides Last Word Subscribe Dating

SPACE TECH ENVIRONMENT HEALTH LIFE PHYSICS&MATH SCIENCE IN SOCIETY

Home | Tech | Science in Society | News

Facial recognition software spots family resemblance

› Updated 18:06 13 December 2011 by Kate McAlpine
› Magazine issue 2842. Subscribe and save

PRINT SEND SHARE

FACIAL recognition software that's as good as people at spotting family resemblances could help to reunite lost family members - or help the likes of Facebook work out which of your friends are blood relatives.

We intuitively recognise family resemblance through features like shared eye colour or chin contours, but computers have a hard time making such links between photos of different people.

Jiwen Lu of Nanyang Technological University in Singapore and his colleagues at Capital Normal University in Beijing, China, trained a piece of software to determine whether or not a pair of photos shows a parent and child. To do this, the team used a database of public figures and their parents or children - such as French president Nicolas Sarkozy and his son Jean - and fed the program 320 pairs each of parent-child matches and mismatches. The program analyses pictures one pixel at a time and looks for trends in the surrounding pixels.

The software then compared the difference between a test pair of photos with pairs of photos in its database. If the differences between the photos were similar to those between parent-child pairs, the images were declared a kinship match. In tests using 160 pairs - 80 parent-child matches and 80 mismatches - the system had a success rate of 68 per cent. The work was presented last week at the Association for Computing Machinery's Multimedia conference in Scottsdale, Arizona.

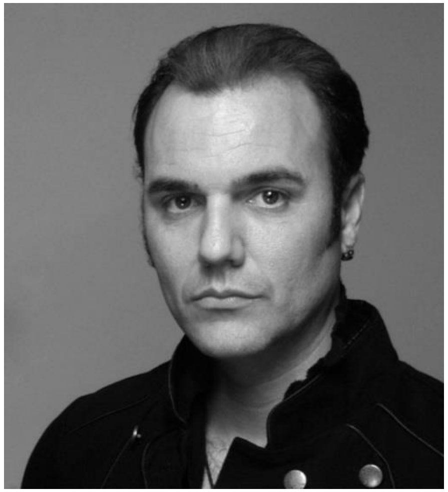
Unlike some previous kinship-recognition programs, Lu's system can deal with variations in pose, expression and illumination. But because it simply compares groups of pixels, it doesn't reveal anything about which facial characteristics might be the best indicators of family ties.

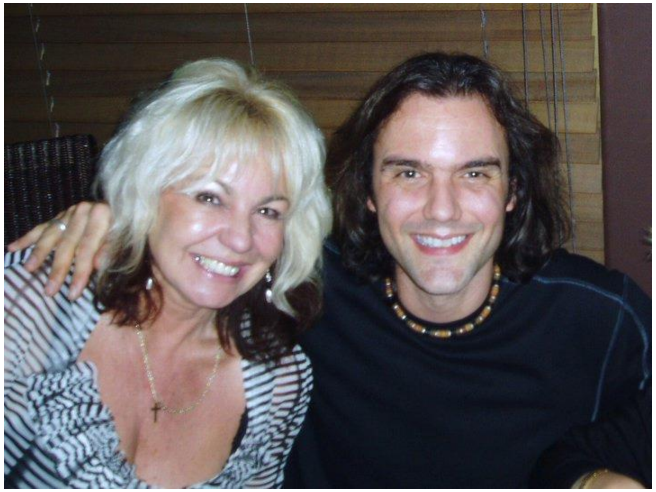
Computers could one day connect you to long-lost relatives just by looking at your photo *(Image: Jeff R Clow/Getty)*

Social Impact

- Successfully identify the father-son kinship of the king and prince in the Netherlands.

- Successfully help the adoptee in UK to kinship verification

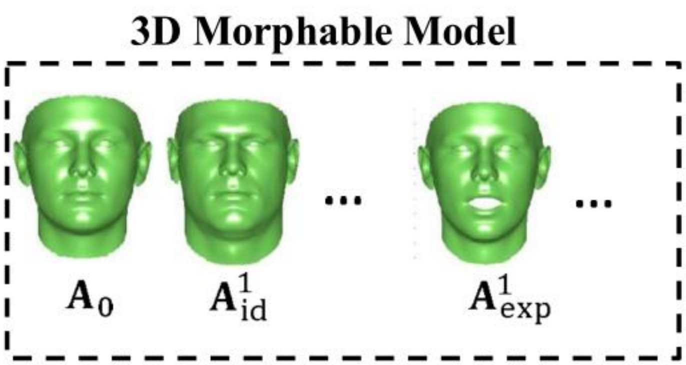


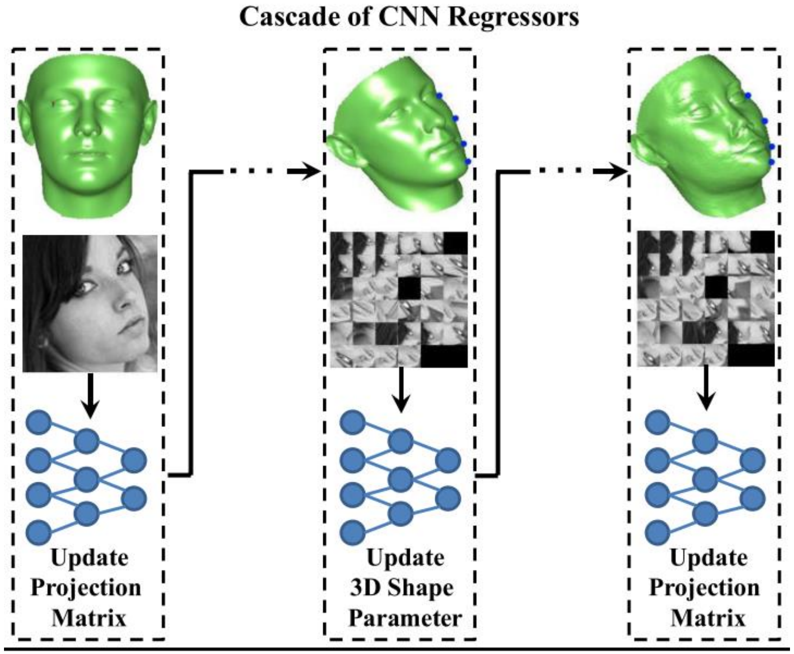


Part 4: Open Questions and Discussions

Face Alignment

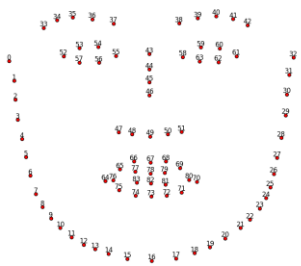
- Frontal face alignment: **Solved**
  - structural and hierarchical reorientation
  - coarse-to-fine shape refinement
- Large-pose
  - Profile/self-occlusion
  - 2D-3D face fitting

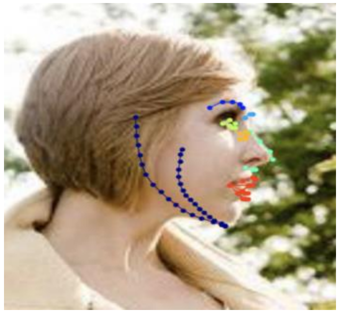


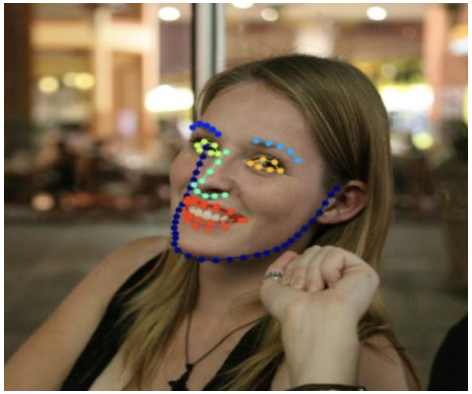


Face Alignment

- Frontal face alignment: **Solved**
  - structural and hierarchical reorientation
  - coarse-to-fine shape refinement
- Large-pose
  - Profile/self-occlusion
  - 2D-3D face fitting
- 3D face tracking
  - Low-resolution
  - Facial motion
  - Pose changes







Face Recognition

Scalability: large-scale face recognition

Robustness: partial faces, large poses, illuminations, expressions, makeups, noisy/missing labels

Efficiency: equipped on mobile devices

Anti-Spoofing: attack and defense

Unsupervised Settings: with no/less training labels

3D Face Recognition: exploitation of 3D data

References

[1] A. Asthana, S. Zafeiriou, S. Cheng, and M. Pantic, Robust discriminative response map fitting with constrained local models, *IEEE CVPR*, pp. 3444–3451, 2013.

[2] T. Baltrusaitis, P. Robinson, and L. Morency, Openface: An open source facial behavior analysis toolkit. *IEEE WACV*, pp. 1–10, 2016.

[3] P. N. Belhumeur, D. W. Jacobs, D. J. Kriegman, and N. Kumar, Localizing parts of faces using a consensus of exemplars. *IEEE CVPR*, pp. 545–552, 2011.

[4] J. Amin, and X. Liu. Large-Pose Face Alignment via CNN-Based Dense 3D Model Fitting. *IEEE CVPR*, pp. 4188-4196, 2016.

[5] A. Bulat and G. Tzimiropoulos. How far are we from solving the 2d and 3d face alignment problem (and a dataset of 230,000 3d facial landmarks). *IEEE ICCV*, Oct 2017.

[6] X. P. Burgos-Artizzu, P. Perona and P. Dolla ́r. Robust face landmark estimation under occlusion. *IEEE ICCV*, pp. 1513–1520, 2013.

[7] X. Cao, Y. Wei, F. Wen, and J. Sun. Face alignment by explicit shape regression. *IEEE CVPR*, pp. 2887–2894, 2012.

[8] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 23(6):681–685, 2001.

[9] G. Trigeorgis, P. Snape, M. A. Nicolaou, E. Antonakos and S. Zafeiriou. Mnemonic descent method: A recurrent process applied for end-to-end face alignment. *IEEE CVPR*, pp. 4177–4187, 2016.

[10] G. Tzimiropoulos. Project-out cascaded regression with an application to face alignment. *In CVPR*, pp. 3659–3667, 2015.

References

[11] G. Tzimiropoulos and M. Pantic. Gauss-newton deformable part models for face alignment in-the-wild. *IEEE CVPR*, pp. 1851–1858, 2014.

[12] Y. Wu, C. Gou, and Q. Ji. Simultaneous facial landmark detection, pose and deformation estimation under facial occlusion. *IEEE CVPR*, July 2017.

[13] Y. Wu and Q. Ji. Discriminative deep face shape model for facial point detection. *International Journal on Computer Vision (IJCV)*, 113(1):37–53, 2015.

[14] Y. Wu and Q. Ji. Robust facial landmark detection under significant head poses and occlusion. *IEEE CVPR*, pp. 3658–3666, 2015.

[15] S. Xiao, J. Feng, J. Xing, H. Lai, S. Yan, and A. A. Kassim. Robust facial landmark detection via recurrent attentive-refinement networks. *Springer ECCV*, pp. 57–72, 2016.

[16] X. Xiong and F. D. la Torre. Supervised descent method and its applications to face alignment. In CVPR, pp. 532–539, 2013.

[17] X. Xiong and F. D. la Torre. Global supervised descent method. *IEEE CVPR*, pp. 2664–2673, 2015.

[18] J. Zhang, S. Shan, M. Kan, and X. Chen. Coarse-to-fine auto-encoder networks (CFAN) for real-time face alignment. *Springer ECCV*, pp. 1–16,2014.

[19] Z. Zhang, P. Luo, C. C. Loy, and X. Tang. Learning deep representation for face alignment with auxiliary attributes. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 38(5):918–930,2016.

[20] S. Zhu, C. Li, C. C. Loy, and X. Tang. Face alignment by coarse-to-fine shape searching. *IEEE CVPR*, pp. 4998–5006, 2015.

References

[21] S. Zhu, C. Li, C. C. Loy, and X. Tang. Unconstrained face alignment via cascaded compositional learning. *IEEE CVPR*, pp. 3409–3417, 2016.

[22] X. Zhu and D. Ramanan. Face detection, pose estimation, and landmark localization in the wild. *IEEE CVPR*, pp. 2879–2886, 2012.

[23] H. Liu, J. Lu, J. Feng and J. Zhou. Two-stream transformer networks for video-based face alignment. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, accepted, 2017.

[24] H. Liu, J. Lu, J. Feng and J. Zhou. Learning deep sharable and structural detectors for face alignment. *IEEE Transactions on Image Processing (TIP)*, 26(4):1666-1678, 2017.

[25] W. Zhao, R. Chellappa, J. Phillips, and A. Rosenfeld, Face recognition: A literature survey, *ACM Computing Surveys*, vol. 35, no. 4, pp. 399-458, 2003.

[26] M. Turk and A. Pentland, Face recognition using Eigenfaces, *IEEE CVPR*, pp. 586-591, 1991.

[27] P. Belhumeur, J. Hespanha, and D. Kriegman, Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection, *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 19, no. 7, pp. 711-720, 1997.

[28] X. He, S. Yan, Y. Hu, P. Niyogi, and H. Zhang, Face recognition using Laplacianfaces, *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 27, no. 3, pp. 328-340, 2005.

[29] C. Liu and H. Wechsler, Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition, *IEEE Transactions on Image Processing (TIP)*, vol. 11, no. 4, pp. 467-476, 2002.

References

[30] T. Ojala, M. Pietikainen, and T. Maenpaa, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 24, no. 7, pp. 971-987, 2002.

[31] T. Ahonen, A. Hadid, and M. Pietikainen, Face description with local binary patterns: Application to face recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 28, no. 12, pp. 2037-2041, 2006.

[32] Z. Cao, Q. Yin, X. Tang, and J. Sun, Face recognition with learning-based descriptor, *IEEE CVPR*, pp. 2707-2714, 2010.

[33] J. Lu, V. Liong, X. Zhou, and J. Zhou, Learning Compact Binary Face Descriptor for Face Recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 37, no. 10, pp. 2041-2056, 2015.

[34] Y. Duan, J. Lu, J. Feng, and J. Zhou, Context-Aware Local Binary Feature Learning for Face Recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 40, no. 5, pp. 1139-1153, 2018.

[35] J. Lu, V. Liong, and J. Zhou, Simultaneous Local Binary Feature Learning and Encoding for Homogeneous and Heterogeneous Face Recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2018, accepted.

[36] J. Hu, J. Lu, and Y. Tan, Discriminative deep metric learning for face verification in the wild, *IEEE CVPR*, pp. 1875-1882, 2014.

[37] J. Lu, J. Hu, and Y. Tan, Discriminative deep metric learning for face and kinship verification, *IEEE Transactions on Image Processing (TIP)*, vol. 26, no. 9, pp. 4269-4282, 2017.

References

[38] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, Deepface: Closing the gap to human-level performance in face verification, *IEEE CVPR*, pp. 1701-1708, 2014.

[39] Y. Sun, X. Wang, and X. Tang, Deep learning face representation from predicting 10,000 classes, *IEEE CVPR*, pp. 1891-1898, 2014.

[40] Y. Sun, Y. Chen, X. Wang, and X. Tang, Deep learning face representation by joint identification-verification, *NIPS*, pp. 1988-1996, 2014.

[41] Y. Sun, X. Wang, and X. Tang, Deeply learned face representations are sparse, selective, and robust, *IEEE CVPR*, pp. 2892-2900, 2015.

[42] Y. Sun, D. Liang, X. Wang, and X. Tang, Deepid3: Face recognition with very deep neural networks, *arXiv preprint arXiv:1502.00873*, 2015.

[43] F. Schroff, D. Kalenichenko, and J. Philbin, Facenet: A unified embedding for face recognition and clustering, *IEEE CVPR*, pp. 815-823, 2015.

[44] O. Parkhi, A. Vedaldi, and A. Zisserman, Deep face recognition, BMVC, 2015.

[45] Y. Wen, K. Zhang, Z. Li, and Y. Qiao, A discriminative feature learning approach for deep face recognition, *Springer ECCV*, pp. 499-515, 2016.

[46] W. Liu, Y. Wen, Z. Yu, and M. Yang, Large-margin softmax loss for convolutional neural networks, *ICML*, pp. 507-516, 2016.

[47] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song, Sphereface: Deep hypersphere embedding for face recognition, *IEEE CVPR*, pp. 6738-6746, 2017.

[48] X. Zhang, Z. Fang, Y. Wen, Z. Li, and Y. Qiao, Range loss for deep face recognition with long-tailed training data, *IEEE ICCV*, pp. 5419-5428, 2017.