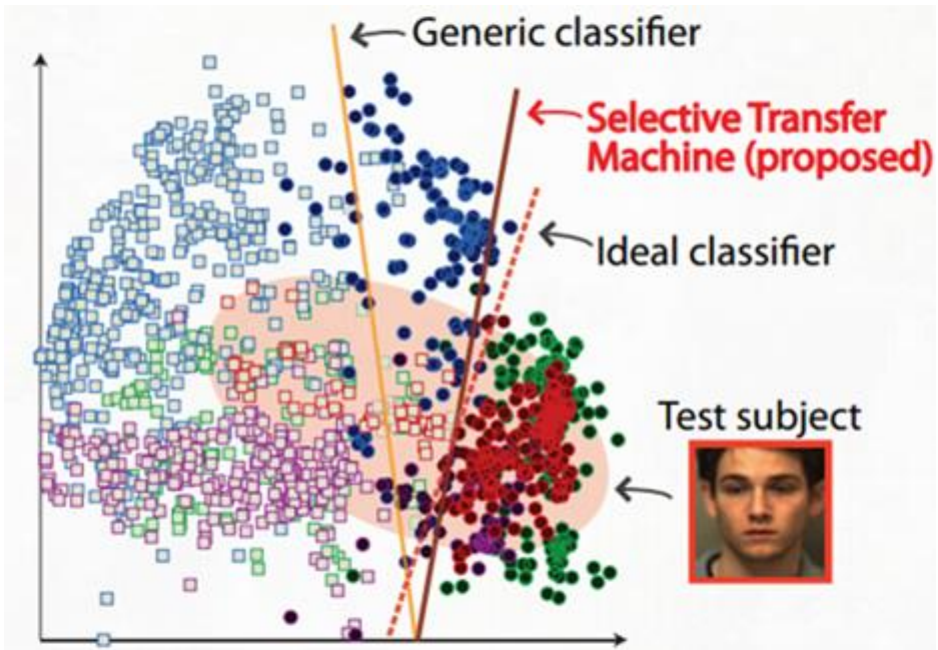
Personalized Facial Expression Analysis

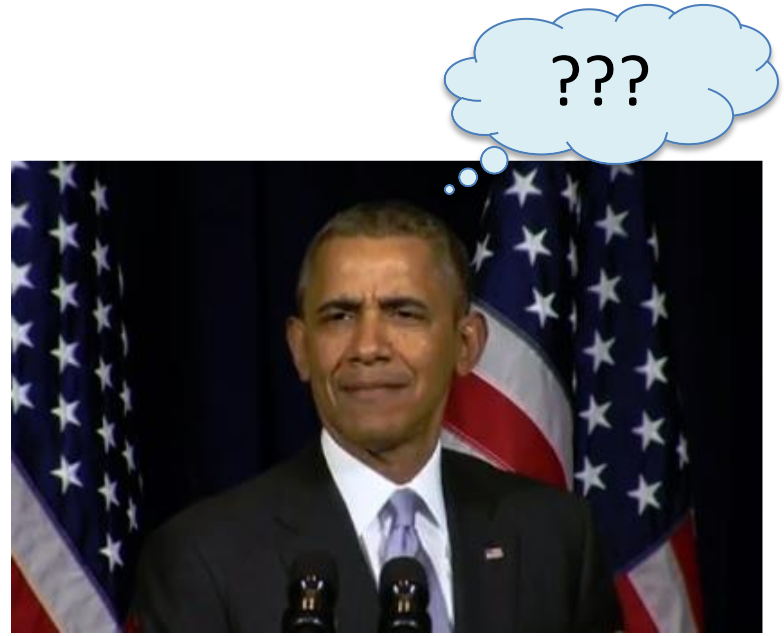


Wen-Sheng Chu
September 9, 2015

Committee:
Fernando De la Torre
Jeffrey F. Cohn
Abhinav Gupta
Xuehan Xiong

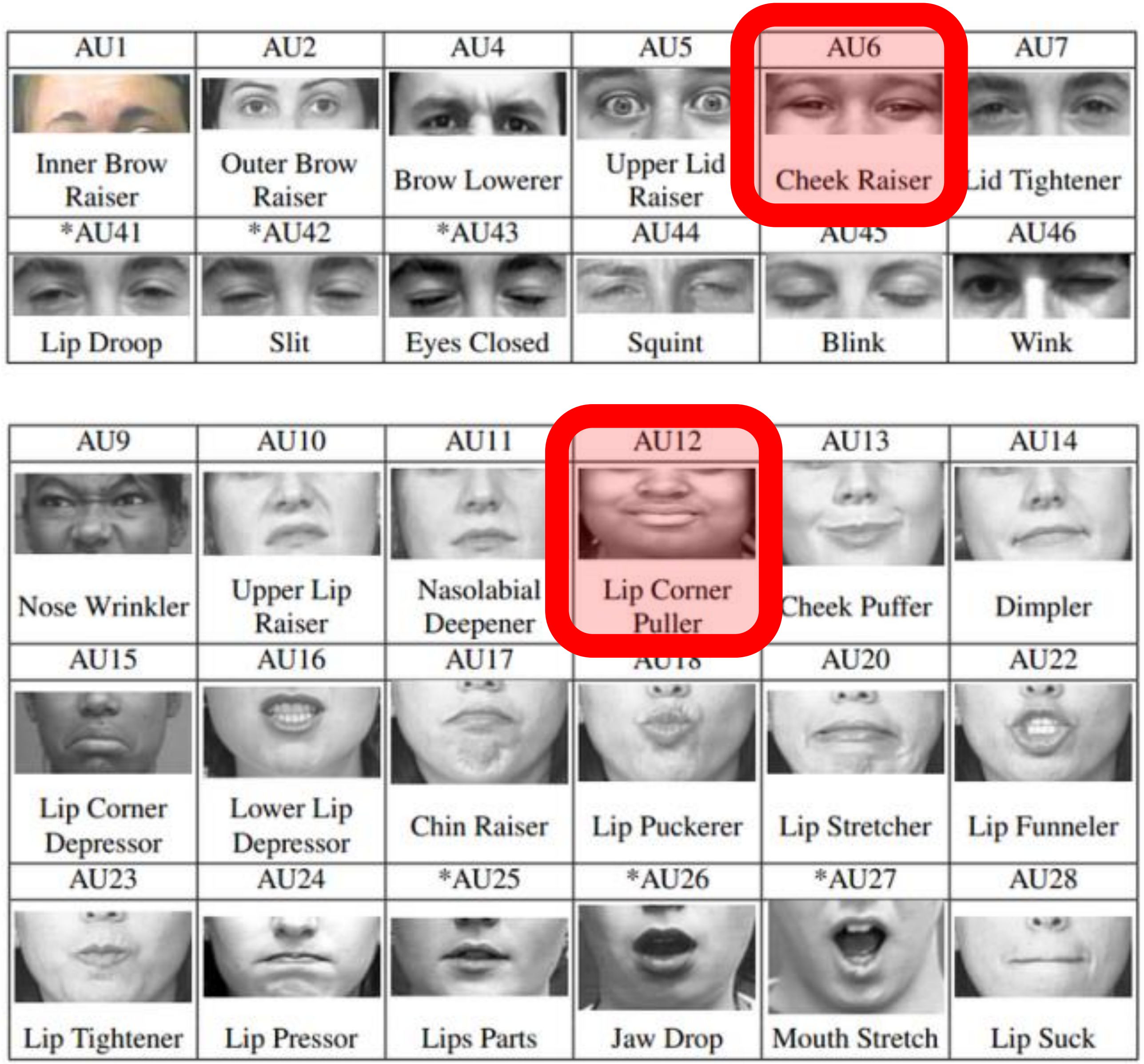
Carnegie Mellon

(Facial) Actions speak louder than words



Facial Action Units (AU)

AU 6+12



Facial Action Coding System, Ekman and Friesman, 1977.

Applications using AUs

Drowsy driver detection

Human-robot interaction

Pain assessment

Animation

Marketing

NVISO

EMOTIENT

affectiva®

Noldus

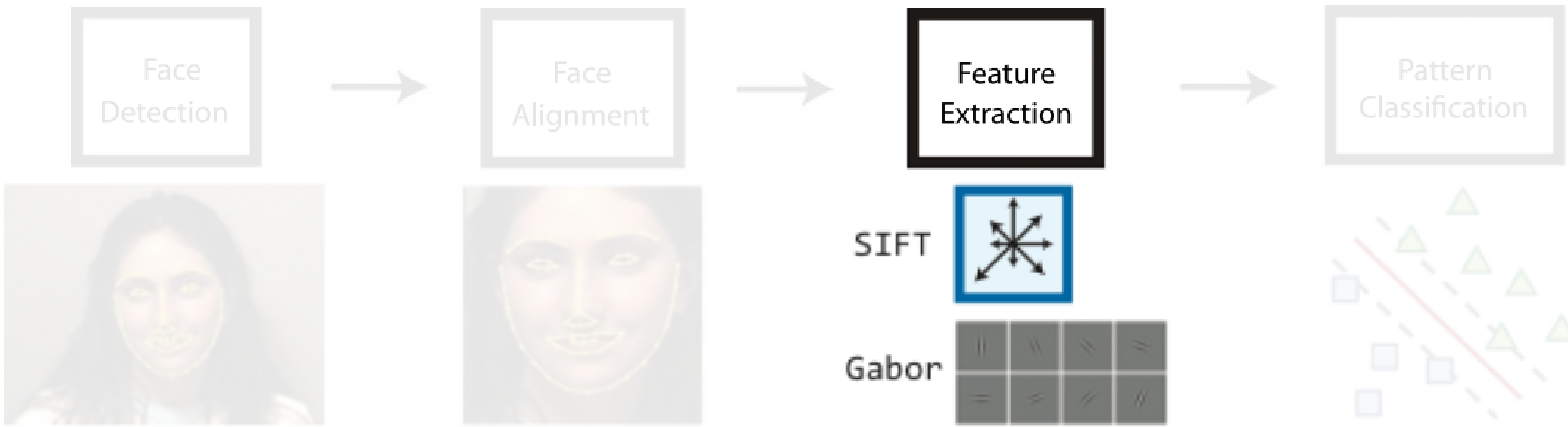
real eyes

Outline

1. Introduction
   – Related work
   – Challenges
2. Our method
   – Objective function
   – Optimization algorithm
3. Experiments
   – Within-subject
   – Cross-subject
   – Cross-dataset
4. Conclusion

INTRODUCTION

Automatic AU coding pipeline

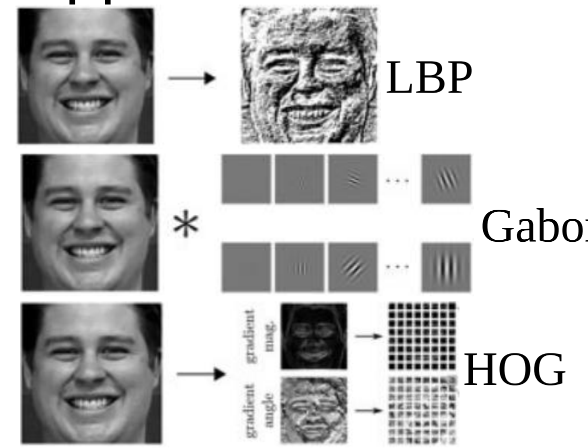


Related work: **features**

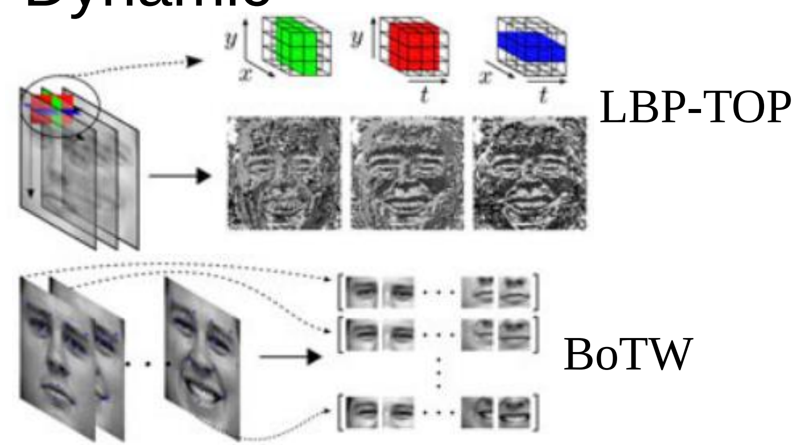
Type	Feature	Year	Reference
Geometric	Shape model parametrization	2012	[57]
	Geometry of facial components	2010	[80]
	Landmark locations	2006	[45]
Appearance	Active facial patches	2012	[79]
	SIFT/DAISY	2011	[81]
	Discrete Cosine Transform (DCT)	2011	[27]
	Local Phase Quantization (LPQ)	2011	[35]
	Local Binary Patterns (LBP)	2009	[58], [66]
	Hist. of Oriented Gradient (HOG)	2009	[48]
	Gabor	2006	[4], [41]
	Raw pixels	2000	[37]
Dynamic	Longitudinal expression atlases	2012	[33]
	Gabor motion energy	2010	[72]
	Bag of Temporal Words (BoTW)	2010	[60]
	Volume LBP (LBP-TOP)	2007	[78]
	Optical flow	2005	[32]
Fusion	Multiple feature kernels	2012	[57]

Geometric

Appearance



Dynamic

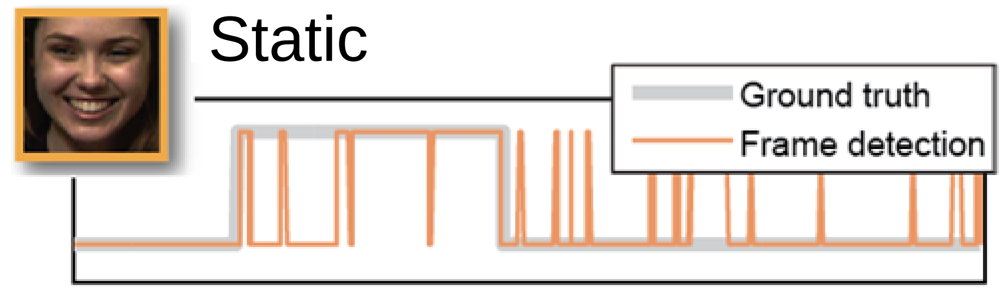


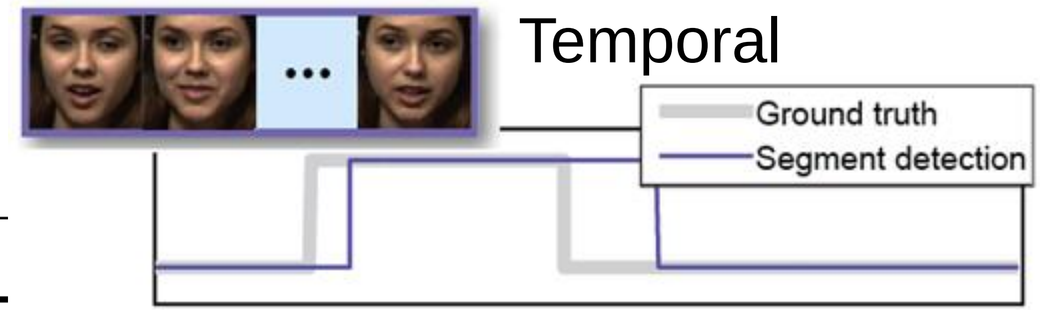
[Image credits: Evangelos Sariyanidi et al, TPAMI'15]

Automatic AU coding pipeline

Related work: **classifiers**

Type	Classifier	Year	Reference
Static	Deep Networks	2013	[42]
	Support Vector Machine (SVM)	2007	[45]
	AdaBoost	2005	[4]
	Neural Network (NN)	2005	[38]
Temporal	Conditional Random Field (CRF)	2009	[9]
	Gaussian process	2009	[13]
	Dynamic Bayesian Network (DBN)	2007	[65], [70]
	Isomap embedding	2006	[10]
Hybrid	Cascade of Tasks (CoT)	2013	[20]





Task 1: Static Task 2: Temporal Task 3: Transition

Challenges in automatic AU coding

1. Generalization to previously unseen subjects
2. Unbalanced data - One or more classes are rare compared to others
3. Sample selection bias - Training data do not represent the distribution of testing scenario
4. Change in data distributions (covariate shift / concept drift) - The marginal distributions of observations change across subjects

Difficulties in generalizing to unseen subjects

- Behavior
- Facial morphology (face shape, texture, etc)
- Recording environments
- Ethnic/racial background
- Age/development level
- Etc.

Julia Roberts

Mona Lisa

Challenges in automatic AU coding

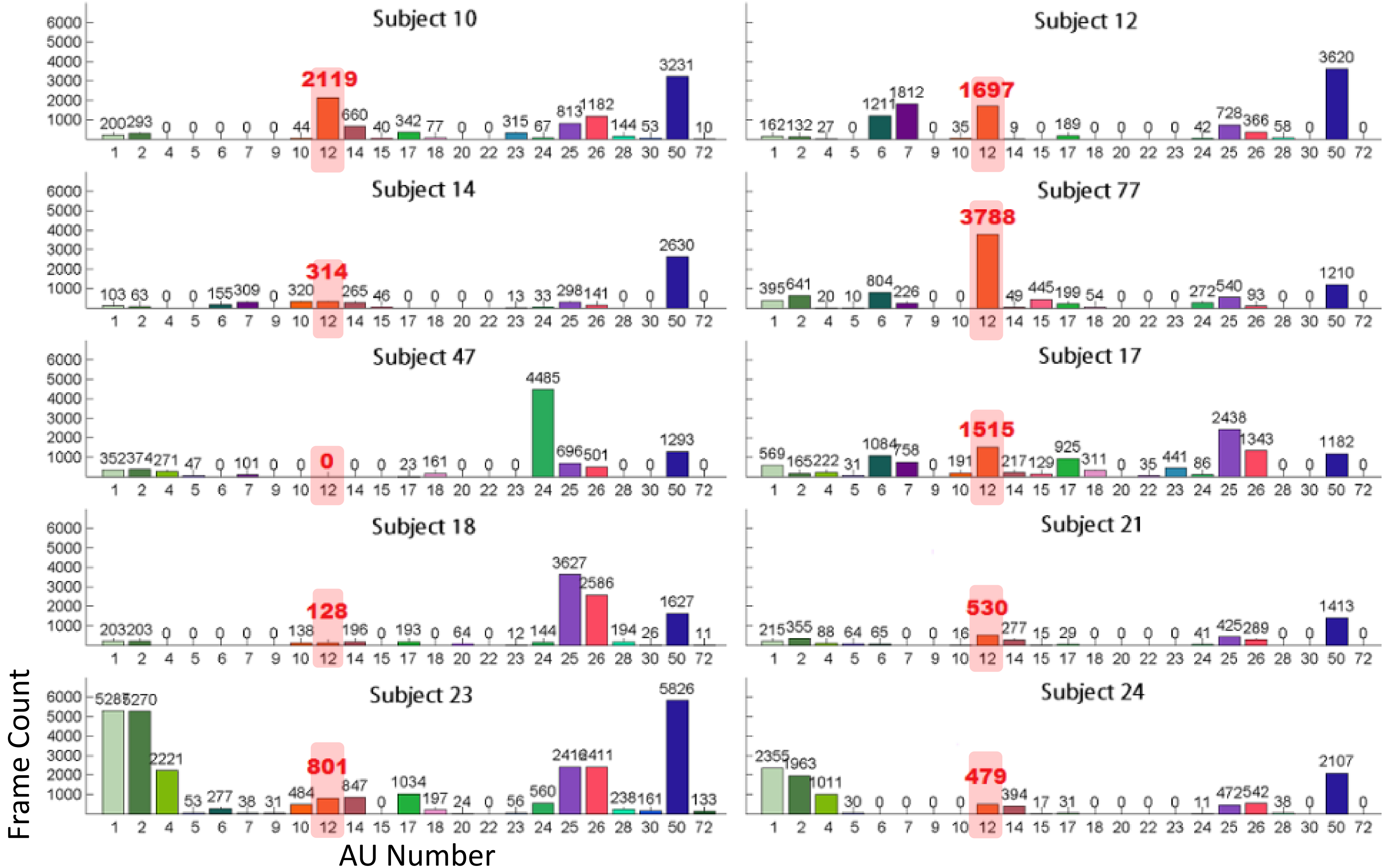
1. Generalization to previously unseen subjects

2. Unbalanced data - One or more classes are rare compared to others

3. Sample selection bias - Training data do not represent the distribution of testing scenario

4. Change in data distributions (covariate shift / concept drift) - The marginal distributions of observations change across subjects

Frame occurrence



Challenges in automatic AU coding

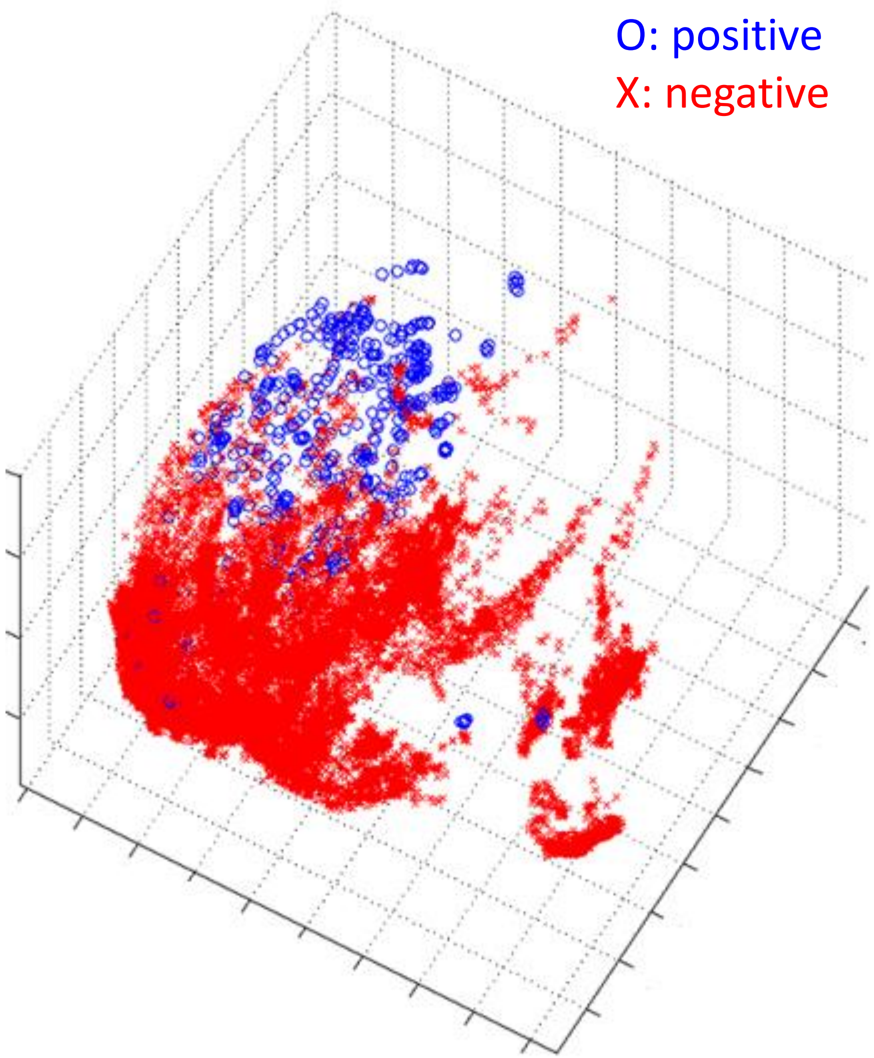
1. Generalization to previously unseen subjects

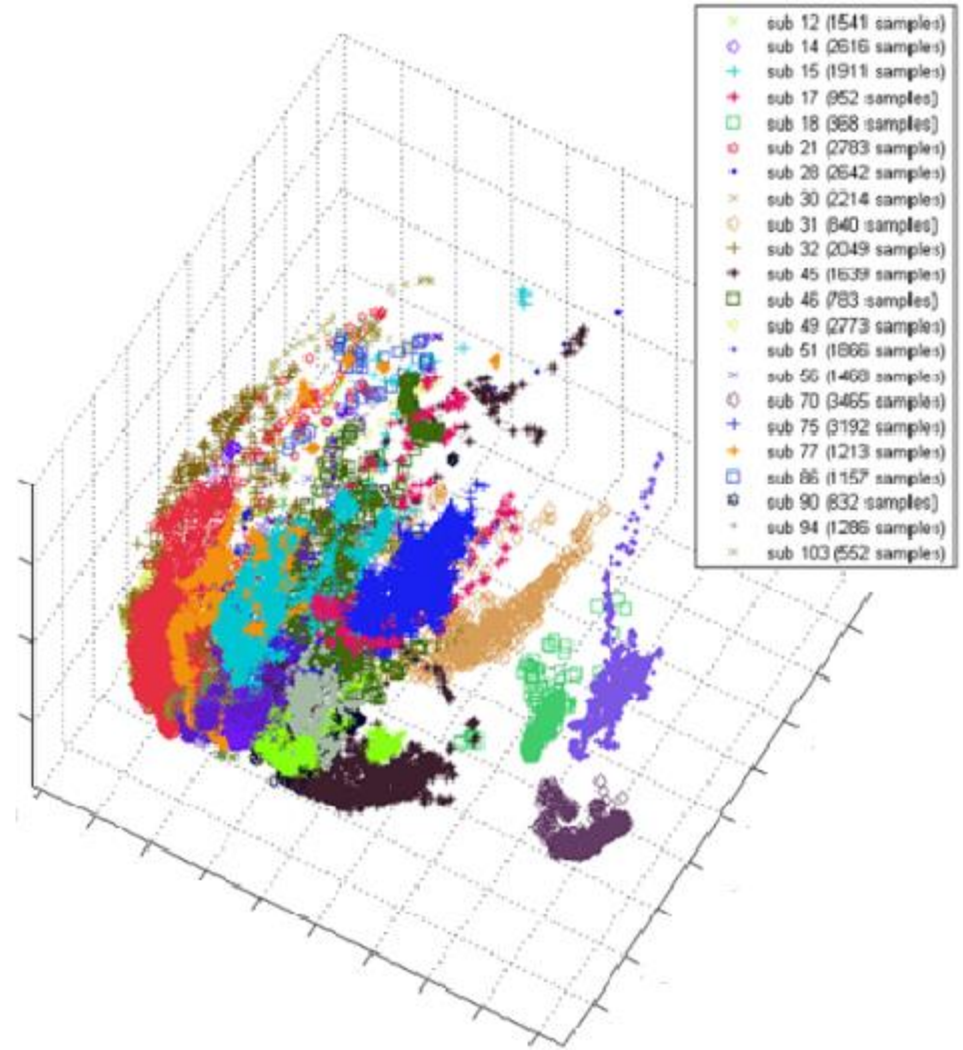
2. Unbalanced data - One or more classes are rare compared to others

3. Sample selection bias - Training data do not represent the distribution of testing scenario

4. Change in data distributions (covariate shift / concept drift) - The marginal distributions of observations change across subjects

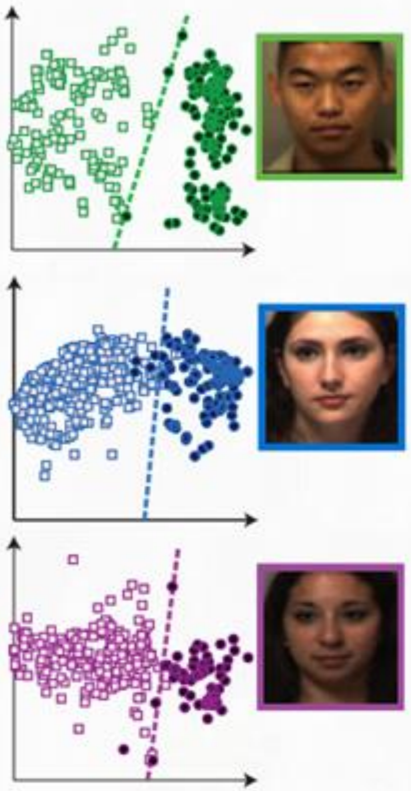
Data distribution





Existence of ideal classifiers

△ Train and test on the same subject



	S012	S014	S015	S017	S018	S021	S028	S030	S031	S032	S045	S046	S047	S049	S051	S056	S057	S060	S070	S071	S075	S077	S079	S080	S086	S090	S094	S103
S012	1.00	0.89	0.81	0.60	0.92	0.76	0.83	0.83	0.86	0.96	0.79	0.89	0.79	0.85	0.38	0.80	0.92	0.95	0.85	0.32	0.80	0.93	0.92	0.79	0.92	0.88	0.88	0.23
S014	0.96	1.00	0.96	0.94	0.94	0.95	0.95	0.95	0.95	0.96	0.96	0.97	0.95	0.96	0.03	0.96	0.96	0.96	0.95	0.65	0.95	0.71	0.96	0.95	0.98	0.96	0.96	0.93
S015	0.74	0.77	0.99	0.72	0.86	0.86	0.76	0.80	0.72	0.79	0.70	0.75	0.70	0.53	0.39	0.75	0.80	0.77	0.72	0.59	0.71	0.85	0.82	0.70	0.61	0.66	0.83	0.85
S017	0.80	0.87	0.43	0.99	0.89	0.76	0.84	0.86	0.37	0.87	0.80	0.89	0.75	0.56	0.42	0.76	0.87	0.83	0.84	0.29	0.75	0.88	0.86	0.75	0.81	0.71	0.84	0.24
S018	0.98	0.98	0.93	0.01	1.00	0.99	0.37	0.98	0.97	0.99	0.20	0.87	0.98	0.97	0.25	0.98	0.99	0.99	0.93	0.02	0.98	0.99	0.46	0.98	0.97	0.99	0.99	0.14
S021	0.46	0.83	0.92	0.44	0.90	1.00	0.07	0.93	0.92	0.93	0.93	0.93	0.92	0.90	0.47	0.94	0.94	0.28	0.79	0.08	0.92	0.77	0.85	0.92	0.91	0.38	0.64	0.71
S028	0.96	0.96	0.95	0.46	0.93	0.93	1.00	0.93	0.94	0.83	0.70	0.85	0.93	0.94	0.05	0.93	0.93	0.94	0.93	0.50	0.93	0.93	0.94	0.93	0.72	0.93	0.95	0.78
S030	0.90	0.86	0.86	0.48	0.75	0.80	0.75	0.99	0.62	0.86	0.76	0.86	0.69	0.81	0.35	0.83	0.83	0.87	0.70	0.47	0.71	0.67	0.93	0.69	0.72	0.83	0.81	0.74
S031	0.95	0.96	0.97	0.37	0.92	0.91	0.94	0.97	1.00	0.96	0.96	0.96	0.96	0.86	0.23	0.97	0.95	0.94	0.95	0.84	0.96	0.96	0.97	0.96	0.90	0.93	0.92	0.39
S032	0.80	0.68	0.76	0.74	0.84	0.81	0.81	0.83	0.78	0.99	0.80	0.81	0.77	0.79	0.37	0.79	0.78	0.79	0.63	0.21	0.78	0.86	0.81	0.77	0.77	0.74	0.85	0.76
S045	0.94	0.95	0.85	0.53	0.94	0.94	0.74	0.93	0.93	0.93	1.00	0.92	0.93	0.95	0.42	0.95	0.95	0.97	0.92	0.14	0.93	0.96	0.94	0.93	0.96	0.89	0.95	0.39
S046	0.63	0.76	0.66	0.66	0.70	0.64	0.65	0.71	0.63	0.67	0.65	0.99	0.63	0.35	0.46	0.84	0.67	0.71	0.47	0.76	0.63	0.86	0.66	0.63	0.50	0.80	0.79	0.80
S047	1.00	1.00	0.99	0.35	1.00	0.99	1.00	1.00	1.00	1.00	1.00	0.97	1.00	0.21	0.15	1.00	0.99	1.00	0.99	0.89	1.00	0.99	1.00	1.00	0.09	0.99	0.99	0.93
S049	0.40	0.76	0.87	0.65	0.67	0.71	0.56	0.86	0.61	0.88	0.84	0.83	0.84	1.00	0.37	0.82	0.86	0.79	0.41	0.28	0.84	0.70	0.81	0.84	0.85	0.66	0.57	0.87
S051	0.86	0.86	0.86	0.19	0.86	0.86	0.42	0.86	0.86	0.86	0.77	0.86	0.86	0.55	1.00	0.86	0.86	0.86	0.83	0.17	0.86	0.86	0.86	0.86	0.86	0.45	0.86	0.16
S056	0.77	0.91	0.91	0.76	0.84	0.90	0.46	0.92	0.88	0.91	0.92	0.91	0.90	0.90	0.10	1.00	0.91	0.86	0.89	0.21	0.90	0.63	0.85	0.90	0.65	0.56	0.66	0.54
S057	0.91	0.93	0.93	0.63	0.94	0.90	0.94	0.94	0.90	0.93	0.89	0.94	0.89	0.87	0.25	0.94	1.00	0.92	0.88	0.45	0.89	0.84	0.94	0.89	0.82	0.91	0.81	0.80
S060	0.94	0.88	0.86	0.44	0.68	0.73	0.78	0.87	0.63	0.67	0.81	0.69	0.59	0.45	0.39	0.62	0.64	1.00	0.80	0.79	0.68	0.89	0.73	0.59	0.63	0.87	0.73	0.88
S070	0.92	0.92	0.92	0.07	0.92	0.92	0.85	0.92	0.92	0.92	0.27	0.92	0.92	0.91	0.10	0.92	0.92	0.93	1.00	0.22	0.92	0.91	0.92	0.92	0.92	0.88	0.92	0.07
S071	0.88	0.89	0.82	0.81	0.84	0.87	0.85	0.86	0.77	0.84	0.84	0.86	0.83	0.84	0.15	0.84	0.86	0.88	0.39	1.00	0.84	0.89	0.84	0.83	0.86	0.85	0.83	0.85
S075	0.70	0.82	0.91	0.34	0.83	0.95	0.95	0.78	0.97	0.93	0.94	0.98	0.98	0.37	0.88	0.98	0.95	0.80	0.23	0.01	1.00	0.95	0.81	0.98	0.32	0.56	0.35	0.00
S077	0.88	0.88	0.66	0.70	0.81	0.65	0.79	0.84	0.32	0.57	0.35	0.66	0.32	0.77	0.59	0.62	0.47	0.79	0.39	0.66	0.32	1.00	0.87	0.32	0.84	0.73	0.81	0.67
S079	0.59	0.83	0.81	0.32	0.87	0.59	0.91	0.85	0.85	0.85	0.85	0.80	0.85	0.53	0.33	0.66	0.60	0.64	0.83	0.80	0.85	0.80	0.99	0.85	0.32	0.41	0.79	0.44
S080	0.98	0.80	0.99	0.43	0.97	0.96	0.99	0.98	0.92	0.99	0.99	0.95	1.00	0.96	0.80	0.94	0.99	0.98	0.05	0.95	1.00	0.72	0.72	1.00	0.20	0.56	0.97	0.93
S086	0.62	0.85	0.85	0.72	0.80	0.75	0.75	0.77	0.65	0.79	0.70	0.71	0.67	0.84	0.57	0.67	0.67	0.85	0.72	0.47	0.67	0.87	0.83	0.67	0.99	0.83	0.86	0.73
S090	0.92	0.93	0.89	0.90	0.92	0.89	0.26	0.89	0.88	0.87	0.89	0.91	0.89	0.93	0.14	0.89	0.91	0.91	0.71	0.18	0.89	0.92	0.94	0.89	0.85	1.00	0.91	0.81
S094	0.94	0.94	0.95	0.76	0.56	0.86	0.94	0.94	0.95	0.94	0.94	0.95	0.94	0.94	0.21	0.94	0.95	0.94	0.94	0.17	0.94	0.95	0.94	0.94	0.85	0.95	1.00	0.32
S103	0.92	0.90	0.92	0.74	0.90	0.88	0.81	0.93	0.84	0.90	0.89	0.91	0.86	0.86	0.13	0.86	0.88	0.93	0.75	0.42	0.87	0.93	0.93	0.86	0.87	0.90	0.93	1.00

Challenges in automatic AU coding

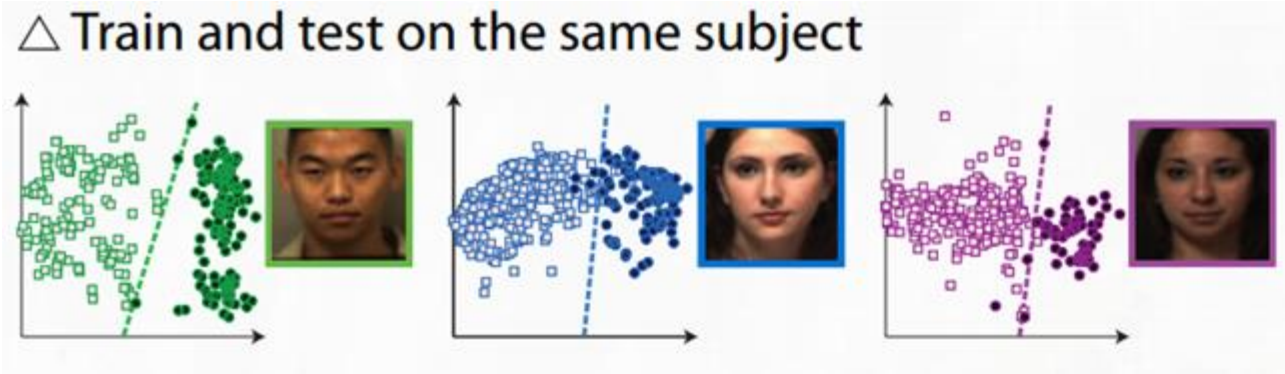
1. Generalization to previously unseen subjects

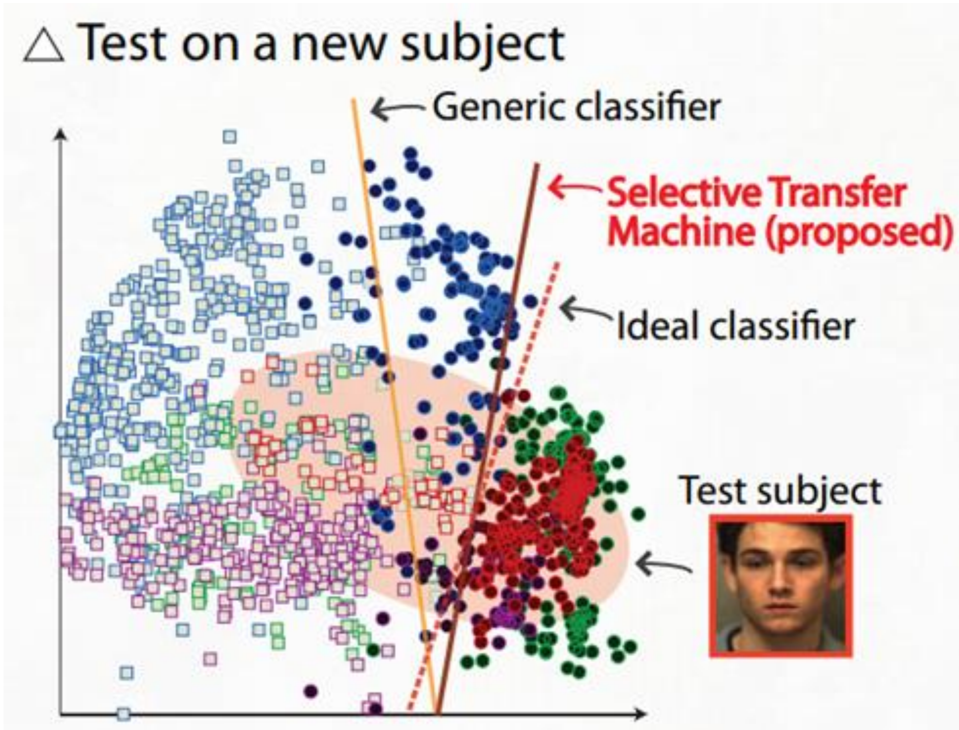
2. Unbalanced data **-** One or more classes are rare compared to others

3. Sample selection bias - Training data do not represent the distribution of testing scenario

4. Change in data distributions (covariate shift / concept drift) - The marginal distributions of observations change across subjects

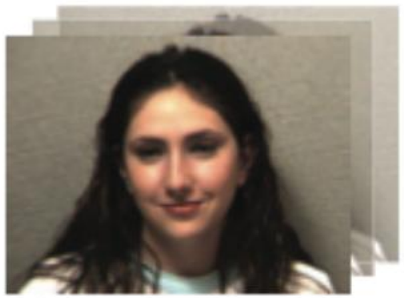
Personalized facial expression analysis





OUR METHOD

STM formulation



$$\mathcal{D}^{\mathrm{tr}} = \{\mathbf{x}_i, y_i\}_{i=1}^{n_{\mathrm{tr}}}, y_i \in \{+1, -1\}$$

$$\mathbf{X}^{\mathrm{tr}} = [\mathbf{x}_1, \ldots, \mathbf{x}_{n_{\mathrm{tr}}}]$$

Unconstrained optimization problem:

$$\min_{f,\mathbf{s}} \underbrace{R_f(\mathcal{D}^{\mathrm{tr}}, \mathbf{s})}_{\text{Maximizes margin of penalized SVM}} + \overbrace{\lambda \Omega_{\mathbf{s}}(\mathbf{X}^{\mathrm{tr}}, \mathbf{X}^{\mathrm{te}})}^{\text{Minimize distribution mismatch}}$$

Goal (1): **maximize penalized SVM margin**

$$\min_{f,\mathbf{s}} \boxed{R_f(\mathcal{D}^{\mathrm{tr}}, \mathbf{s})} + \lambda\Omega_{\mathbf{s}}(\mathbf{X}^{\mathrm{tr}}, \mathbf{X}^{\mathrm{te}})$$

Linear: $$R_{\mathbf{w}}(\mathcal{D}^{\mathrm{tr}}, \mathbf{s}) = \underbrace{\frac{1}{2}\|\mathbf{w}\|^2}_{\text{margin}} + \underbrace{C\sum_{i=1}^{n_{\mathrm{tr}}} s_i L^p(y_i, \mathbf{w}^\top \mathbf{x}_i)}_{\text{penalized loss}}$$

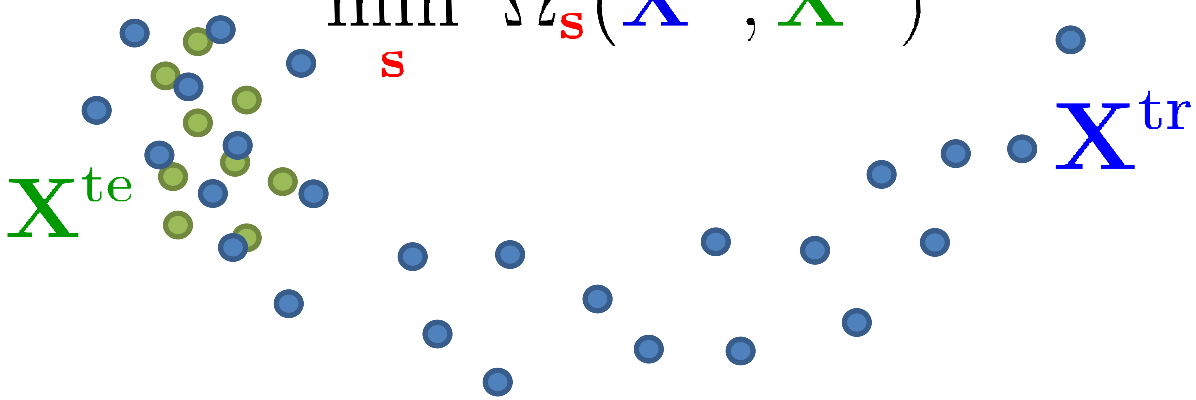
Non-linear: $$R_{\boldsymbol{\beta}}^{\mathrm{nonlin}}(\mathcal{D}^{\mathrm{tr}}, \mathbf{s}) = \frac{1}{2}\boldsymbol{\beta}^\top \mathbf{K}\boldsymbol{\beta} + C\sum_{i=1}^{n_{\mathrm{tr}}} s_i L^p(y_i, \mathbf{K}_i^\top \boldsymbol{\beta})$$

Goal (2): **minimize distribution mismatch**

$$\min_{f,\mathbf{s}} \; R_f(\mathcal{D}^{\mathrm{tr}}, \mathbf{s}) + \lambda \Omega_{\mathbf{s}}(\mathbf{X}^{\mathrm{tr}}, \mathbf{X}^{\mathrm{te}})$$

KMM*: $\Omega_{\mathbf{s}}(\mathbf{X}^{\mathrm{tr}}, \mathbf{X}^{\mathrm{te}}) = \left\| \frac{1}{n_{\mathrm{tr}}} \sum_{i=1}^{n_{\mathrm{tr}}} s_i \varphi(\mathbf{x}_i^{\mathrm{tr}}) - \frac{1}{n_{\mathrm{te}}} \sum_{j=1}^{n_{\mathrm{te}}} \varphi(\mathbf{x}_j^{\mathrm{te}}) \right\|_{\mathcal{H}}^2$

$$\min_{\mathbf{s}} \; \Omega_{\mathbf{s}}(\mathbf{X}^{\mathrm{tr}}, \mathbf{X}^{\mathrm{te}})$$



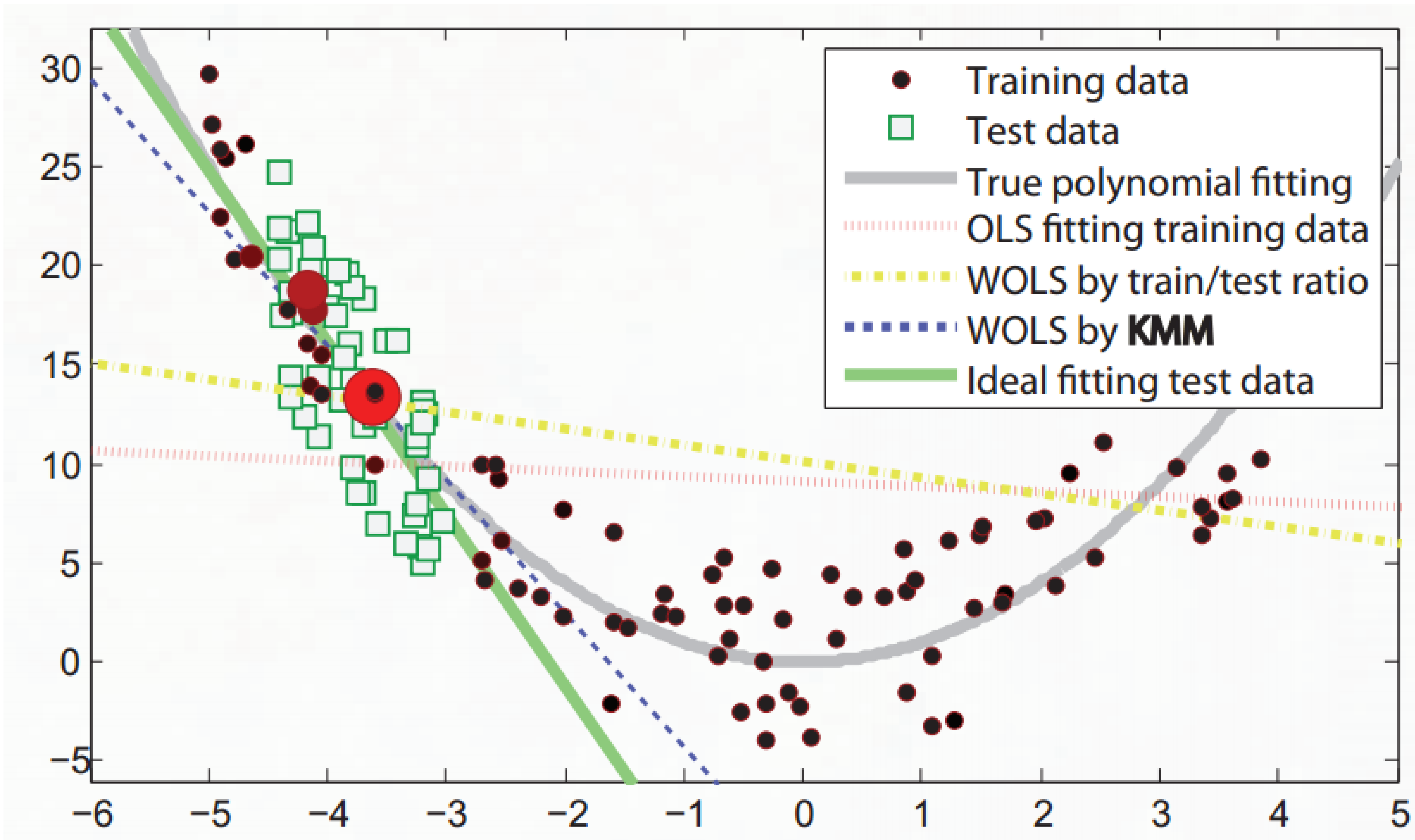
$$\Rightarrow \quad \min_{\mathbf{s}} \frac{1}{2} \mathbf{s}^\top \mathbf{K} \mathbf{s} - \kappa^\top \mathbf{s}$$

$$K_{ij} := k(x_i^{\mathrm{tr}}, x_j^{\mathrm{tr}})$$

$$\kappa_i := \frac{n_{\mathrm{tr}}}{n_{\mathrm{te}}} \sum_{j=1}^{n_{\mathrm{te}}} k(x_i^{\mathrm{tr}}, x_j^{\mathrm{te}})$$

* "Covariate shift by kernel mean matching", Dataset shift in machine learning, 2009.

Goal (2): a synthetic example



STM is a biconvex problem

$$\min_{f,\mathbf{s}} \; R_f(\mathcal{D}^{\mathrm{tr}}, \mathbf{s}) + \lambda \Omega_{\mathbf{s}}(\mathbf{X}^{\mathrm{tr}}, \mathbf{X}^{\mathrm{te}})$$

1. Biconvexity: The objective and constraint set are biconvex.
2. Boundedness: R_f is bounded due to its quadratic form and non-negati $\Omega_{\mathbf{s}}$ s; is bounded because K is PSD.

Guaranteed convergence using Alternate Convex Search:

1. Convergence in objective value
2. Convergence in variables

"Biconvex sets and optimization with biconvex functions: a survey and extensions," Mathematical Methods of Operations Research, vol. 66, no. 3, pp. 373–407, 2007.

Solve for *f*: **a primal solver**

$$R_{\mathbf{w}}(\mathcal{D}^{\mathrm{tr}}, \mathbf{s}) = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n_{\mathrm{tr}}} s_i L^p(y_i, \mathbf{w}^\top \mathbf{x}_i)$$

$$R_{\boldsymbol{\beta}}^{\mathrm{nonlin}}(\mathcal{D}^{\mathrm{tr}}, \mathbf{s}) = \frac{1}{2}\boldsymbol{\beta}^\top \mathbf{K}\boldsymbol{\beta} + C\sum_{i=1}^{n_{\mathrm{tr}}} s_i L^p(y_i, \mathbf{K}_i^\top \boldsymbol{\beta})$$

▲ Minimize over w

Sample weights

$$\nabla_{\mathrm{lin}} = \mathbf{w} + 2C\sum_{i\in sv} s_i(\mathbf{w}^\top \mathbf{x}_i - y_i)\mathbf{x}_i$$

$$H_{\mathrm{lin}} = \mathbf{I}_d + 2C\sum_{i\in sv} s_i \mathbf{x}_i \mathbf{x}_i^\top$$

$$\nabla_{\mathrm{nonlin}} = \mathbf{K}\boldsymbol{\beta} + 2C\mathbf{K}\mathbf{S}\mathbf{I}^0(\mathbf{K}\boldsymbol{\beta} - \mathbf{y})$$

$$H_{\mathrm{nonlin}} = \mathbf{K} + 2C\mathbf{K}\mathbf{S}\mathbf{I}^0\mathbf{K}$$

➔With the gradient and Hessian, one can use standard Newton's method with quadratic convergence rate.

Solve for s: **weight refinement**

▲ Minimize over s

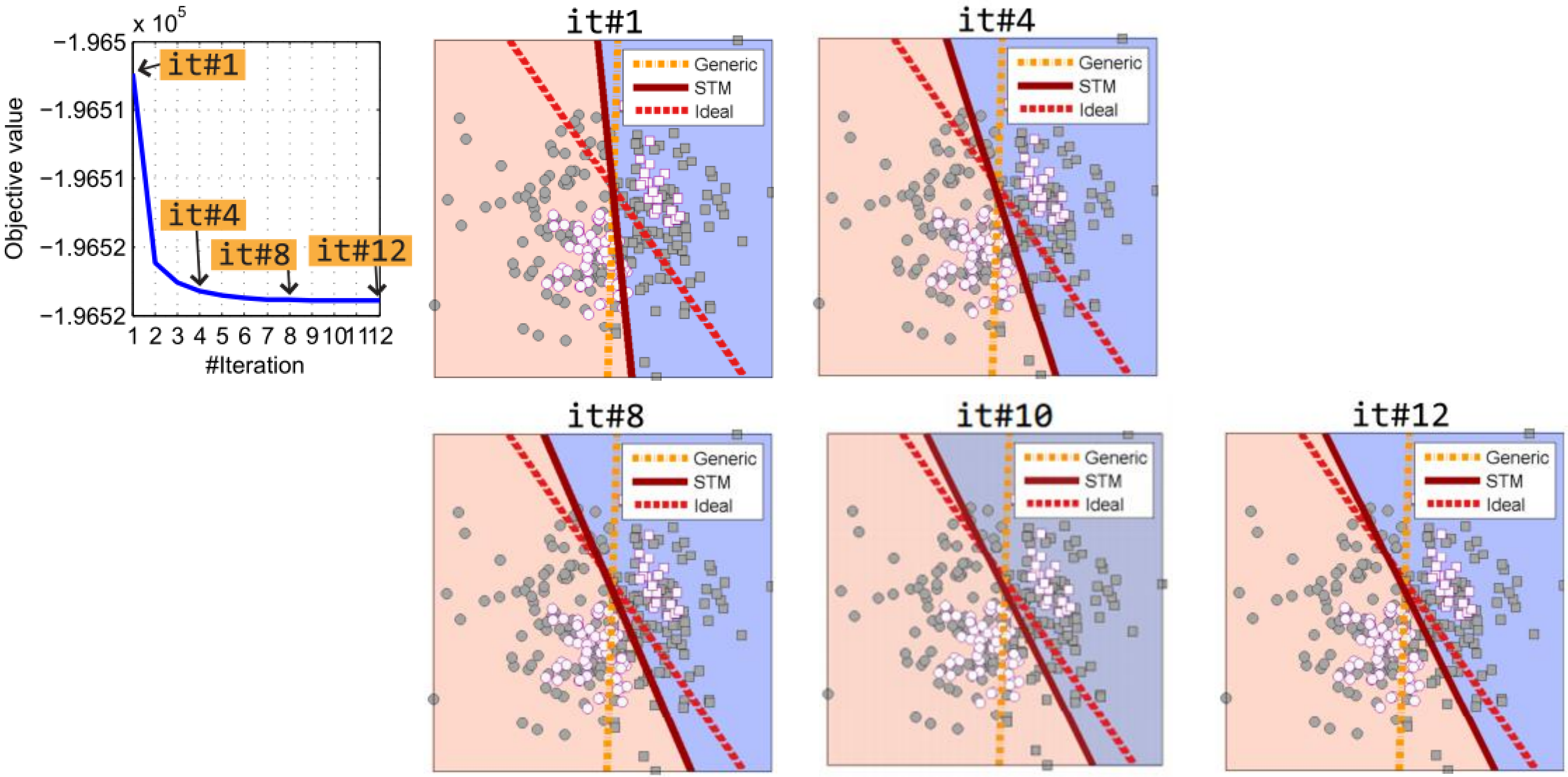
Correcting sample weights
with training loss

$$\min_{\mathbf{s}} \quad \frac{1}{2}\mathbf{s}^\top \mathbf{K}\mathbf{s} + \underbrace{\left(\frac{C}{\lambda}\boldsymbol{\ell}_p - \boldsymbol{\kappa}\right)}^\top \mathbf{s}$$

$$\text{s.t.} \quad 0 \le s_i \le B, n_{\mathrm{tr}}(1-\epsilon) \le \sum^{n} s_i \le n_{\mathrm{tr}}(1+\epsilon).$$

➔This weight-refinement is crucial for introducing the label information into sample re-weighting.

Synthetic example



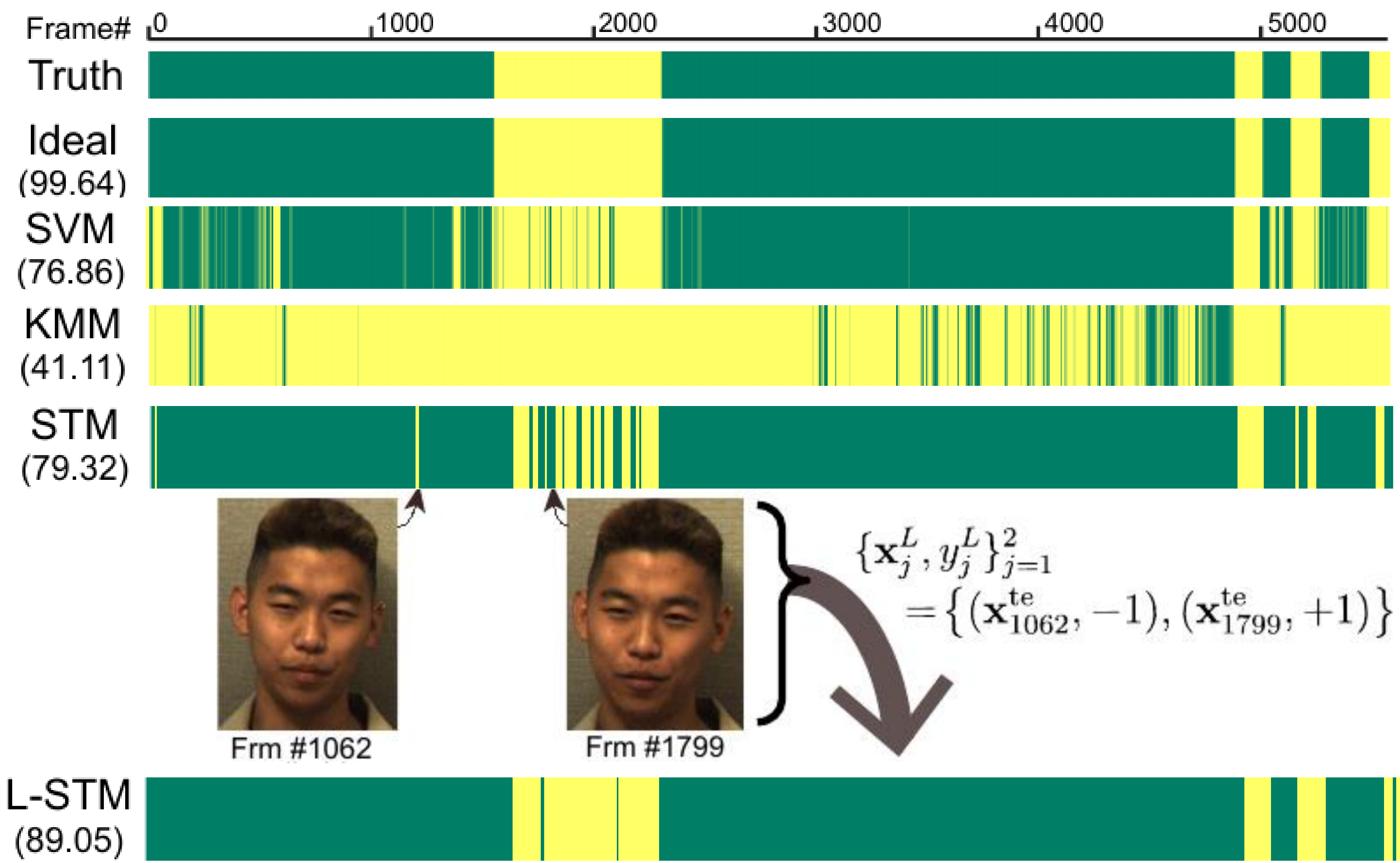
An Extension: **STM with labeled target data (L-STM)**

- In some cases, one have partially labeled data from the target subject $\mathcal{D}^L = \{\mathbf{x}_j^L, y_j^L\}_{j=1}^{L}$

$$\min_{f,\mathbf{s}} \underbrace{R_f(\mathcal{D}^{\text{tr}}, \mathbf{s}) + \lambda \Omega_{\mathbf{s}}(\mathbf{X}^{\text{tr}}, \mathbf{X}^{\text{te}})}_{\text{Original STM}} + \underbrace{\boxed{\lambda_L \Omega_L(\mathcal{D}^L)}}_{\text{Additional regularization}}$$

$$\Omega_L(\mathcal{D}^L) = \sum_{j=1}^{n_L} L^p(y_j^L, f(\mathbf{x}_j^L))$$

An example using L-STM



EXPERIMENTS

Datasets

CK+

GEMEP-FERA

RU-FACS

GFT

Datasets	#Subjects	#Videos	#Frms/vid	Content	AU annotation	Emotion annotation
CK+ [44]	123	593	$\sim$20	Neutral$\rightarrow$peak	Per video	Per video
GEMEP-FERA [67]	7	87	20$\sim$60	Acting	Frame-by-frame	Per video
RU-FACS [4]	34	34	5000$\sim$8000	Interview	Frame-by-frame	–
GFT [56]	720	720	$\sim$60,000	Multi-person social interaction	Frame-by-frame	–

Settings

- Features
  - SIFT descriptors on 49 facial landmarks + PCA (98% energy)
- AU selection
  - 8 most commonly observed AUs across the datasets
- Evaluation
  - Area Under the ROC Curve (AUC) and F1 score
- Data split & validation
  - Leave-one-subject-out protocol, cross-validation
- Three scenarios
  - (1) Within-subject, (2) cross-subject, and (3) cross-dataset

Competitive methods

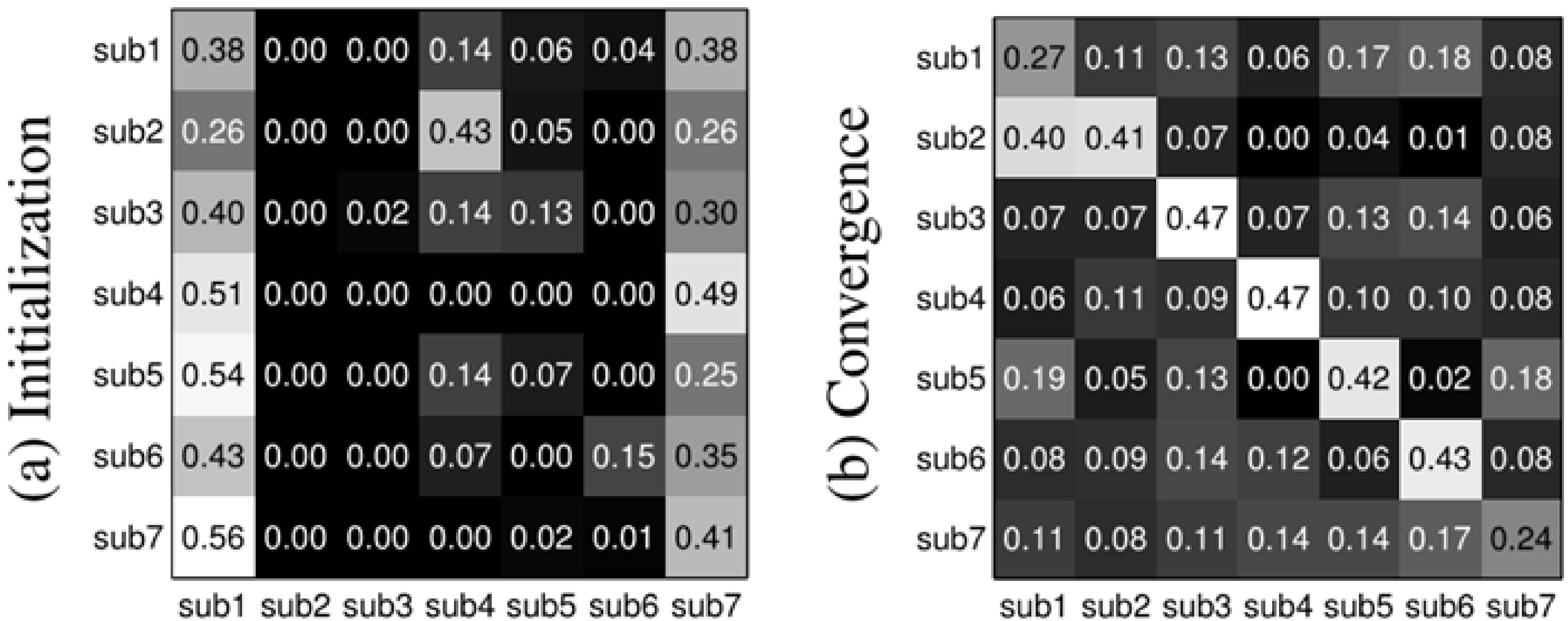
1. Generic classifier:
   - Linear SVM
2. Semi-supervised learning:
   - Transductive SVM (T-SVM)
3. Transductive transfer learning:
   - Kernel Mean Matching (KMM)
   - Domain Adaptation SVM (DA-SVM)
   - Subspace Alignment (SA)
4. Multiple source domain adaptation:
   - Frustratingly easy domain adaptation (FR)
   - Domain Adaptation Machine (DAM)

(1) Within-subject AU Detection: **Comparison to Person-Specific (PS) Classifiers**

- Two definitions
  - PS_1: train/test are separate data of the same subject
  - PS_2: training subjects include test subject (same protocol as in FERA 2012)
- GEMEP-FERA

	AUC			$F1$ Score		
AU	PS_1-SVM	PS_2-SVM	STM	PS_1-SVM	PS_2-SVM	STM
1	48.0	72.4	**79.2**	45.0	54.8	**61.9**
2	46.5	71.1	**80.2**	45.9	55.7	**64.3**
4	62.6	61.9	**66.5**	46.6	40.7	**60.4**
6	70.3	80.0	**86.4**	60.2	69.7	**78.5**
7	47.5	54.3	**72.4**	49.4	55.3	**58.4**
12	65.7	74.0	**72.3**	69.5	70.4	**72.6**
15	41.4	64.0	**70.5**	44.5	49.0	**56.0**
17	32.6	**70.3**	61.7	25.0	**40.3**	36.3
Avg	51.8	68.5	**73.6**	48.3	54.5	**61.0**

(1) Within-subject AU Detection:
Selection Ability of STM



(2) Cross-subject AU detection: **CK+**

- Neutral expression → Peak expression
- 123 subjects
- 597 videos, ~20 frames/video

	AUC					$F1$ Score				
AU	SVM	KMM	T-SVM	DA-SVM	STM	SVM	KMM	T-SVM	DA-SVM	STM
1	79.8	68.9	69.9	72.6	**88.9**	61.1	44.9	56.8	57.7	**62.2**
2	**90.8**	73.5	69.3	71.0	87.5	73.5	50.8	59.8	64.3	**76.2**
4	74.8	62.2	63.4	69.9	**81.1**	62.7	52.3	51.9	57.7	**69.1**
6	89.7	87.7	60.5	**94.7**	94.0	75.5	70.1	47.8	68.2	**79.6**
7	82.1	68.2	55.7	61.4	**91.6**	59.6	47.0	43.8	53.1	**79.1**
12	88.1	89.5	76.0	**95.5**	92.8	76.7	74.5	59.6	59.0	**77.2**
15	93.5	66.8	49.9	94.1	**98.2**	75.3	44.4	40.4	76.9	**84.8**
17	90.3	66.6	73.1	94.7	**96.0**	76.0	53.2	61.7	81.4	**84.3**
Avg	86.1	72.9	64.7	81.7	**91.3**	70.0	54.7	52.7	64.8	**76.6**

(2) Cross-subject AU detection: **GEMEP-FERA**

- Professional actors acting pre-selected expressions
- 7 subjects
- 87 videos, 20~60 frames/video

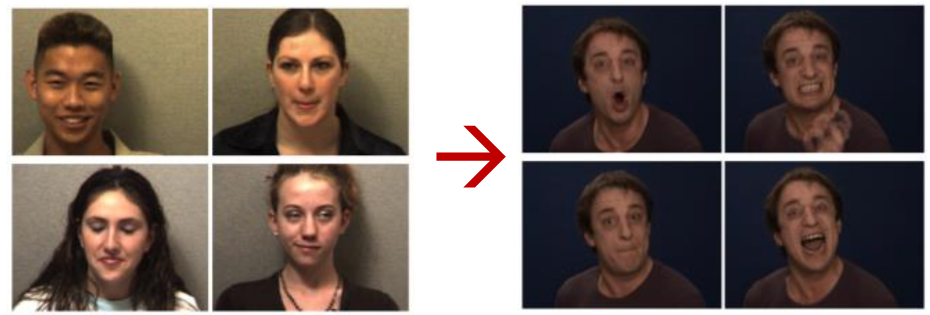
(a)	AUC								*F*1 Score							
AU	SVM	KMM	T-SVM	DA-SVM	SA (N|S)	FR	DAM	STM	SVM	KMM	T-SVM	DA-SVM	SA (N|S)	FR	DAM	STM
1	71.5	43.3	72.2	83.3	21.5|53.1	**84.7**	80.6	84.3	56.5	48.5	60.3	59.1	30.3| 0.0	6.9	36.2	**68.1**
2	73.9	51.0	74.3	76.8	20.2|57.3	76.7	**78.1**	73.3	56.9	50.2	58.5	57.1	27.9| 0.6	9.4	31.7	**65.5**
4	58.5	53.5	42.8	**66.6**	12.3|52.2	62.5	58.6	60.0	43.5	39.8	36.9	**46.3**	16.1| 0.1	2.0	5.8	43.3
6	80.4	60.2	81.1	**91.1**	14.7|52.6	82.0	83.2	87.7	63.7	58.7	63.8	**72.7**	20.5| 0.8	10.3	49.7	71.6
7	66.9	59.4	70.8	**76.9**	17.8|48.5	78.7	77.2	75.4	63.1	63.5	63.7	**68.3**	27.8| 0.0	17.7	34.0	66.2
12	77.7	58.8	74.8	74.5	25.3|53.4	84.2	**85.8**	84.7	79.1	68.4	77.6	75.5	49.7|25.1	76.4	74.5	**82.1**
15	55.5	58.7	67.2	67.5	12.6|52.4	37.6	75.2	**67.8**	33.4	35.2	35.2	**41.3**	9.4| 0.1	4.0	6.5	39.3
17	59.8	51.8	63.8	66.5	7.4|43.6	**70.6**	70.3	63.3	32.0	27.8	36.2	**42.0**	9.1| 0.2	12.9	19.8	35.9
Av.	68.0	54.6	68.4	75.4	16.5|51.7	72.1	**76.1**	74.5	53.5	49.0	54.0	57.8	23.9| 3.4	17.5	32.3	**59.0**

(2) Cross-subject AU detection: **RU-FACS**

- Video-recorded interviews of varying ethnicity
- 29 subjects
- 29 videos, 5000~7000 frames/vid

(b)	AUC								$F1$ Score							
AU	SVM	KMM	T-SVM	DA-SVM	SA (N|S)	FR	DAM	STM	SVM	KMM	T-SVM	DA-SVM	SA (N|S)	FR	DAM	STM
1	72.0	74.0	72.0	77.0	41.2|82.0	80.9	82.6	**83.9**	40.8	37.7	37.4	35.5	20.9|24.2	0.4	11.3	**55.3**
2	66.6	58.6	71.1	76.5	38.2|81.4	82.3	81.2	**82.4**	35.7	32.2	36.2	34.1	18.6|21.8	2.0	17.0	**52.6**
4	74.8	62.2	50.0	76.4	24.5|71.1	62.3	51.3	**82.4**	25.2	14.5	11.2	**35.3**	5.7| 5.8	0.0	2.9	30.4
6	89.1	88.8	61.6	60.3	46.2|78.3	80.1	81.2	**93.1**	58.3	39.2	33.1	42.9	23.2|19.2	1.5	20.9	**72.4**
12	86.7	87.0	86.7	84.4	55.9|86.1	91.8	**93.1**	92.3	61.9	63.0	62.6	71.4	37.5|38.6	10.8	36.6	**72.3**
14	71.8	67.8	74.4	70.4	38.0|78.5	78.2	79.5	**87.4**	31.3	25.8	25.8	40.9	16.5|15.7	0.0	5.7	**51.0**
15	72.5	68.8	73.5	58.1	37.7|79.2	79.9	71.8	**86.1**	32.3	29.5	32.3	34.9	10.1| 8.8	0.0	3.2	**45.4**
17	78.5	76.7	79.5	75.7	55.8|89.9	90.2	**93.9**	89.6	39.5	35.6	44.0	46.5	21.9|17.2	0.3	22.9	**55.3**
Av.	76.5	72.3	71.1	72.3	42.2|80.8	80.7	79.3	**86.3**	40.6	37.3	40.6	42.7	19.3|18.9	1.9	15.1	**54.3**

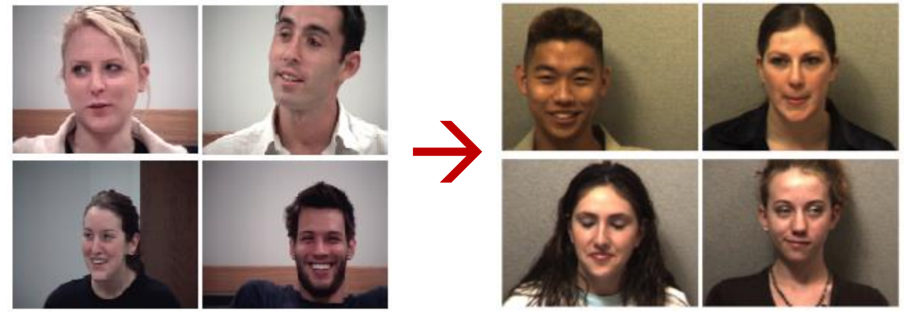
(3) Cross-dataset: RU-FACS→GEMEP-FERA



(a)	AUC					F1 Score				
AU	SVM	KMM	T-SVM	DA-SVM	STM	SVM	KMM	T-SVM	DA-SVM	STM
1	44.7	48.8	43.7	56.9	**63.2**	46.3	46.4	41.8	46.1	**50.4**
2	52.8	70.5	52.1	52.3	**74.0**	47.4	54.2	38.6	45.4	**54.6**
4	52.7	55.4	54.2	52.7	**58.6**	57.1	57.1	40.2	42.9	**57.4**
6	73.5	55.2	77.1	79.9	**83.4**	60.7	55.2	52.8	56.3	**72.7**
12	56.8	60.1	70.9	76.1	**78.1**	67.7	67.7	63.5	62.6	**71.5**
15	55.1	52.1	59.3	**60.2**	58.6	31.5	32.8	29.7	26.4	**41.1**
17	44.3	41.1	39.1	46.2	**52.7**	27.3	27.1	24.3	24.6	**31.4**
Av.	54.3	54.8	56.6	60.6	**66.9**	48.3	48.6	41.6	43.5	**54.2**

(3) Cross-dataset AU detection: GFT→RU-FACS

- 3-person social interaction
- 50 subjects
- 50 videos, ~6000 frames/vid

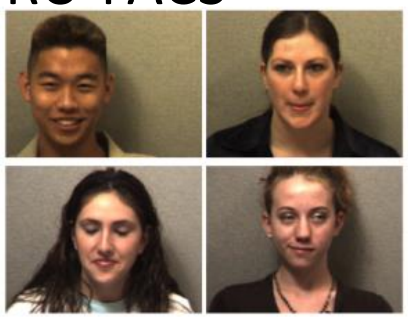


(b)	AUC					F1 Score				
AU	SVM	KMM	T-SVM	DA-SVM	STM	SVM	KMM	T-SVM	DA-SVM	STM
1	45.8	63.6	70.3	71.2	**73.7**	23.7	29.8	26.6	31.8	**38.6**
2	46.4	62.8	68.5	68.2	**71.7**	21.3	25.4	19.4	**32.1**	30.2
4	56.9	60.1	59.1	47.2	**61.7**	18.3	24.5	20.7	19.4	**28.5**
6	65.5	73.9	81.5	74.1	**93.3**	42.2	46.8	30.4	38.7	**61.4**
12	65.3	72.1	76.3	80.9	**90.3**	43.2	47.6	45.8	56.8	**62.2**
14	57.2	54.8	53.7	70.2	**72.2**	25.8	23.8	25.9	29.7	**36.2**
15	56.9	61.8	64.2	65.5	**80.4**	23.7	30.3	28.2	29.9	**37.8**
17	52.4	54.5	64.8	**72.6**	**72.6**	30.8	31.5	32.3	38.9	**39.5**
Av.	55.8	62.9	67.3	68.7	**77.0**	28.6	32.5	28.7	34.7	**41.8**

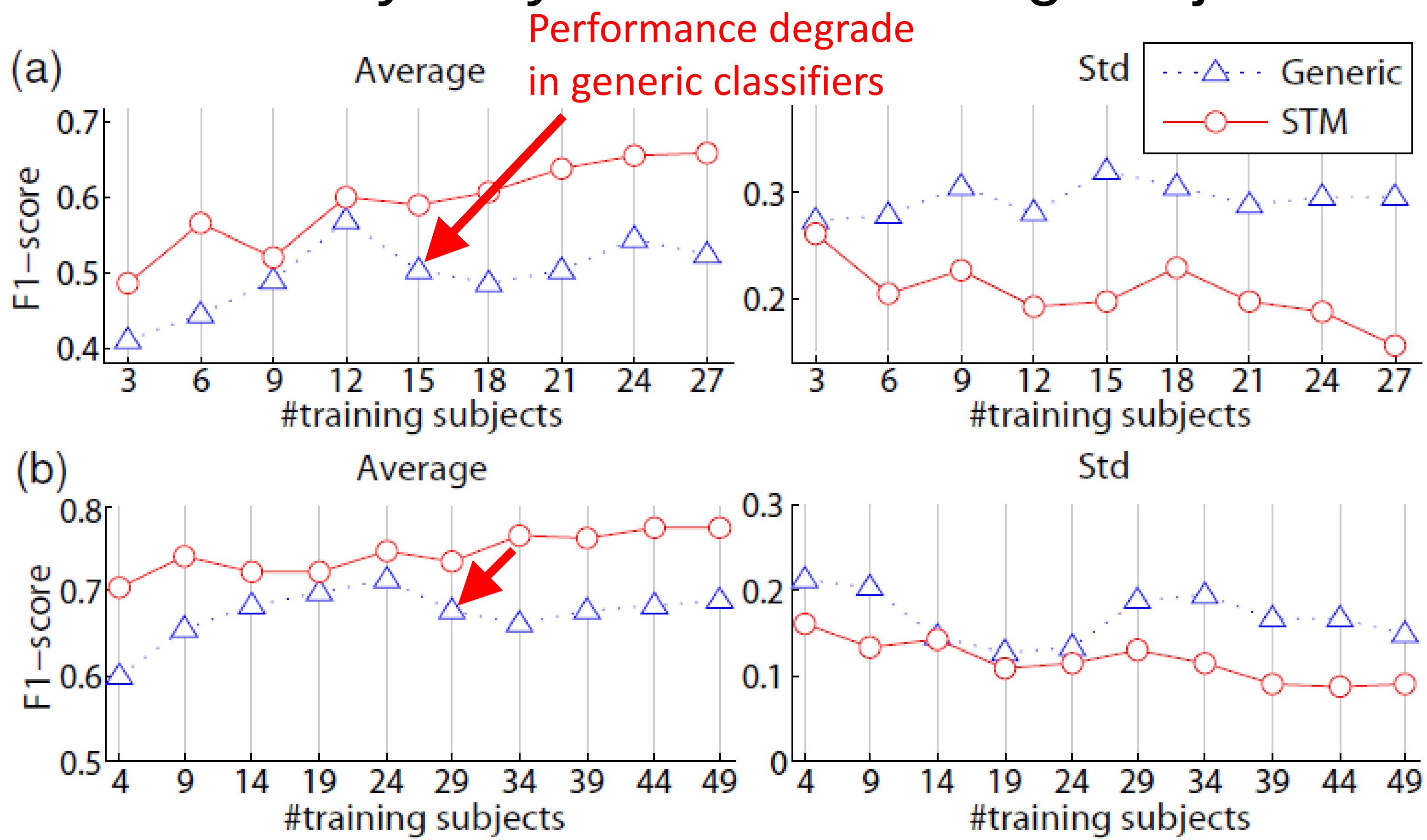
Analysis on domain size

- STM works better due to a judicious selection on training samples.
- How does the selection ability vary across #training subjects?

RU-FACS



GFT



CONCLUSION

Automatic AU detection is still an open problem

2015

- E. Sariyanidi, H. Gunes, and A. Cavallaro, "Automatic analysis of facial affect: A survey of registration, representation, and recognition", *TPAMI*, 2015.
- A. Martinez and S. Du, "A model of the perception of facial expressions of emotion by humans: Research overview and perspectives," *JMLR*, 2012.
- F. De la Torre and J. F. Cohn, "Facial expression analysis," *Visual Analysis of Humans: Looking at People*, 2011.
- Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang. "A survey of affect recognition methods: Audio, visual, and spontaneous expressions," *TPAMI*, 2009.
- Y. Tian, T. Kanade, and J. F. Cohn. "Facial Expression Analysis," *Handbook of Face Recognition*, Springer, 2008.
- M. Pantic and M. S. Bartlett, "Machine analysis of facial expressions," Face recognition, 2007.
- B. Fasel and J. Luettin. "Automatic facial expression analysis: a survey," *Pattern Recognition*, 2003.

2003

Summary

- Based on observation of individual differences, we present Selective Transfer Machine for personalized facial expression analysis.
- We show that STM is biconvex, and can be solved using an alternating algorithm with a primal solver.
- We introduced L-STM, an extension that exhibited significant improvement with few labeled test data.

- Experiments on three scenarios reveal two messages:
  1. Training samples do not matter equally.
  2. Extending variety of training subjects improves performance.

Future Work

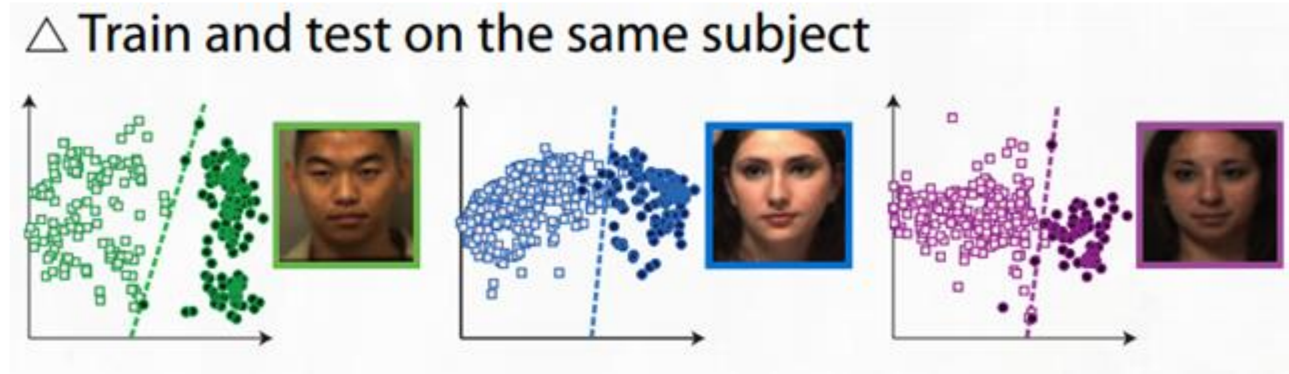
Limitations of STM:

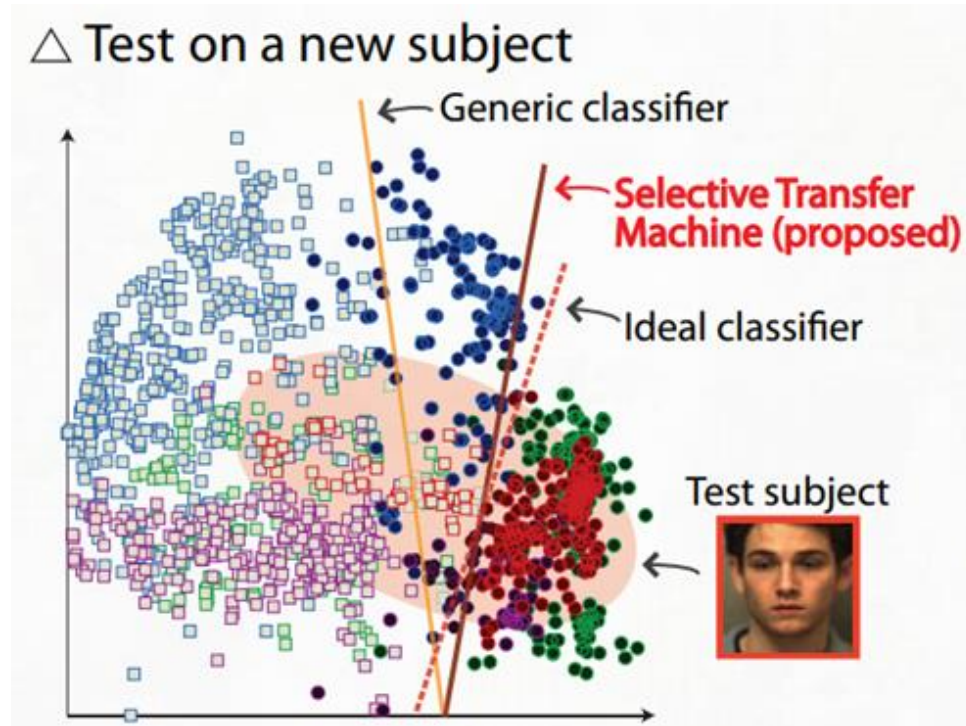
- Similar to all transductive learning methods, STM needs to be trained for each test subject.
- Training is potentially slow due to the QP for solving s.

Extensions:

- Incorporate STM with other classifiers / regressors
- Apply STM to other fields where object-specific issues are involved, e.g., object or activity recognition

QUESTIONS?


△ Train and test on the same subject

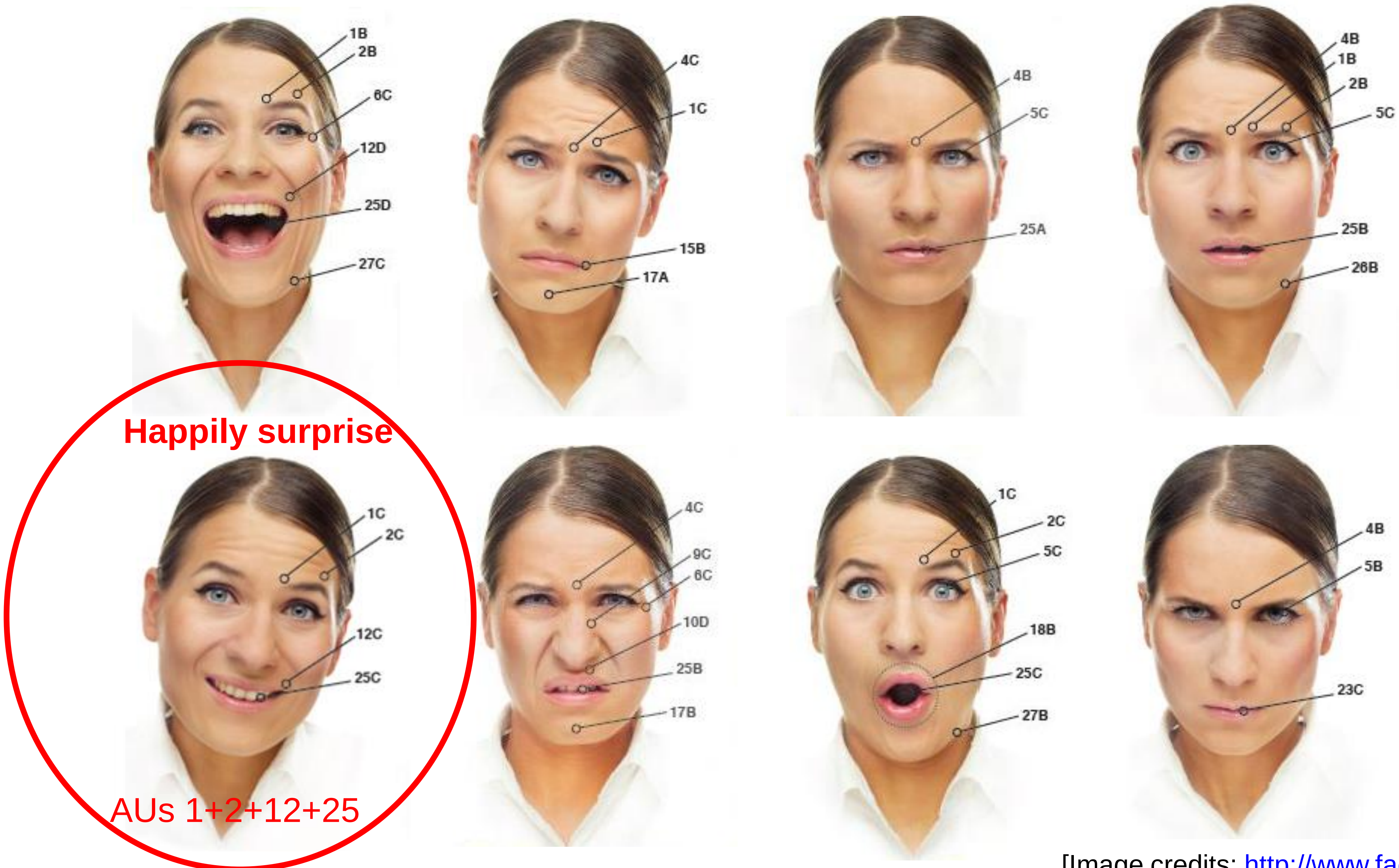

△ Test on a new subject
Generic classifier
Selective Transfer Machine (proposed)
Ideal classifier
Test subject

Connection to Transductive Transfer Learning

Methods	Importance re-weight	Weight refine	Convexity	Labeled target data
SVM-KNN [79]	×	×	–	×
T-SVM [17]	×	×	non-convex	×
KMM [31]	✓	×	convex	×
DA-SVM [6]	×	✓	non-convex	×
DT-MKL [21]	×	×	jointly convex	optional
DAM [22]	×	×	convex	optional
STM (proposed)	✓	✓	bi-convex	optional

[79] "Discriminative NN classification for visual category recognition," In CVPR 2006.
[17] "Large-scale transductive SVMs," JMLR 2006.
[31] "Covariate shift by kernel mean matching," Dataset Shift in Machine Learning, 2009.
[6] "A DASVM classification technique and a circular validation strategy," In TPAMI 2010.
[21] "Domain transfer multiple kernel learning," in TPAMI 2012.
[22] "Domain adaptation from multiple sources: A domain-dep. reg. approach," in TNNLS 2012.

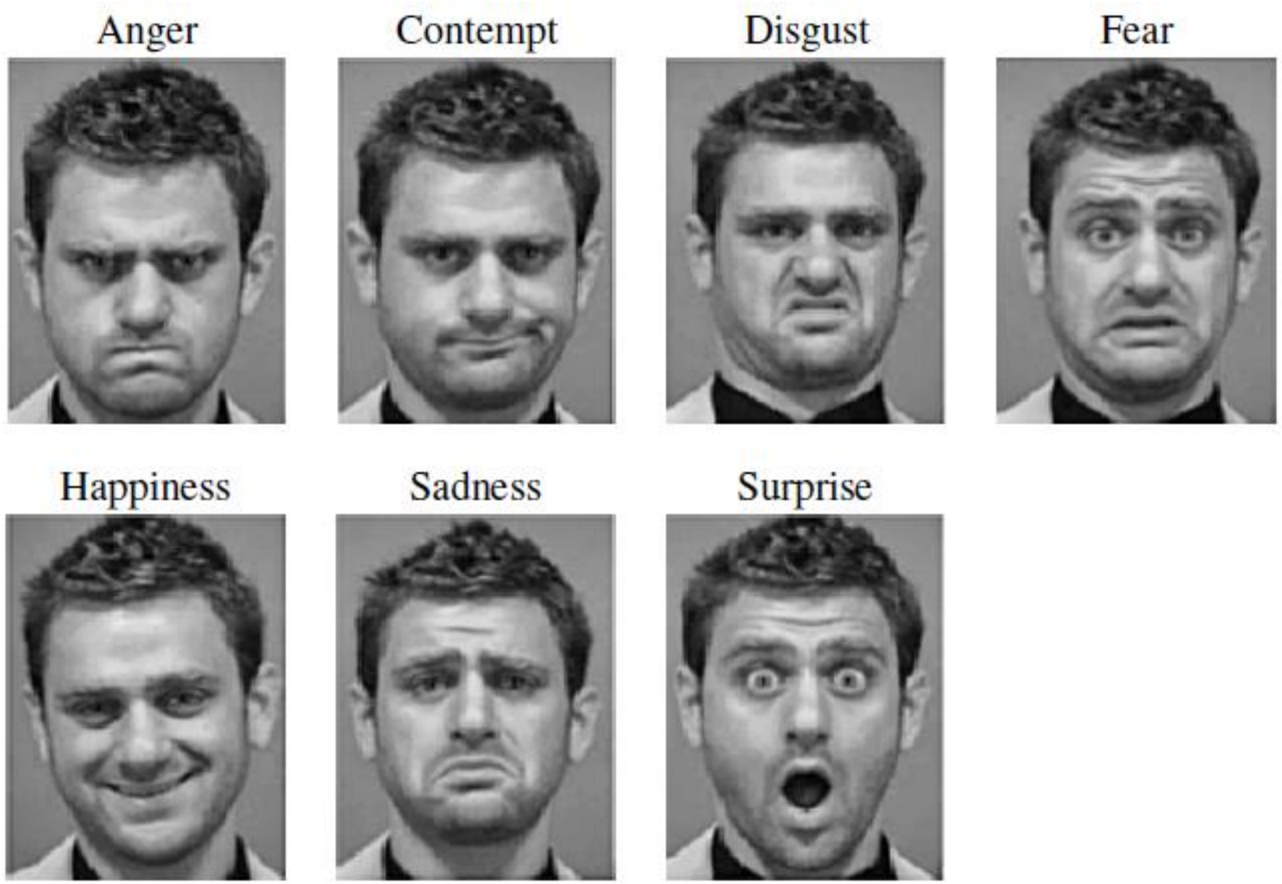
Descriptive power of AUs



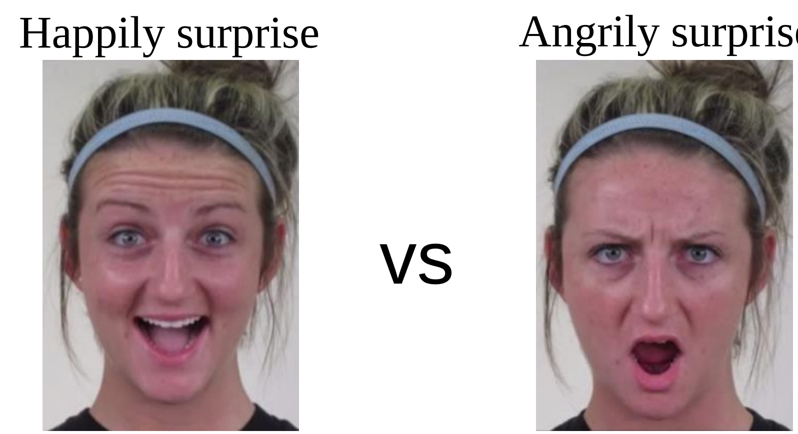
[Image credits: http://www.facscodinggroup.com/]

Universal Facial Expressions

- Standard emotion categories

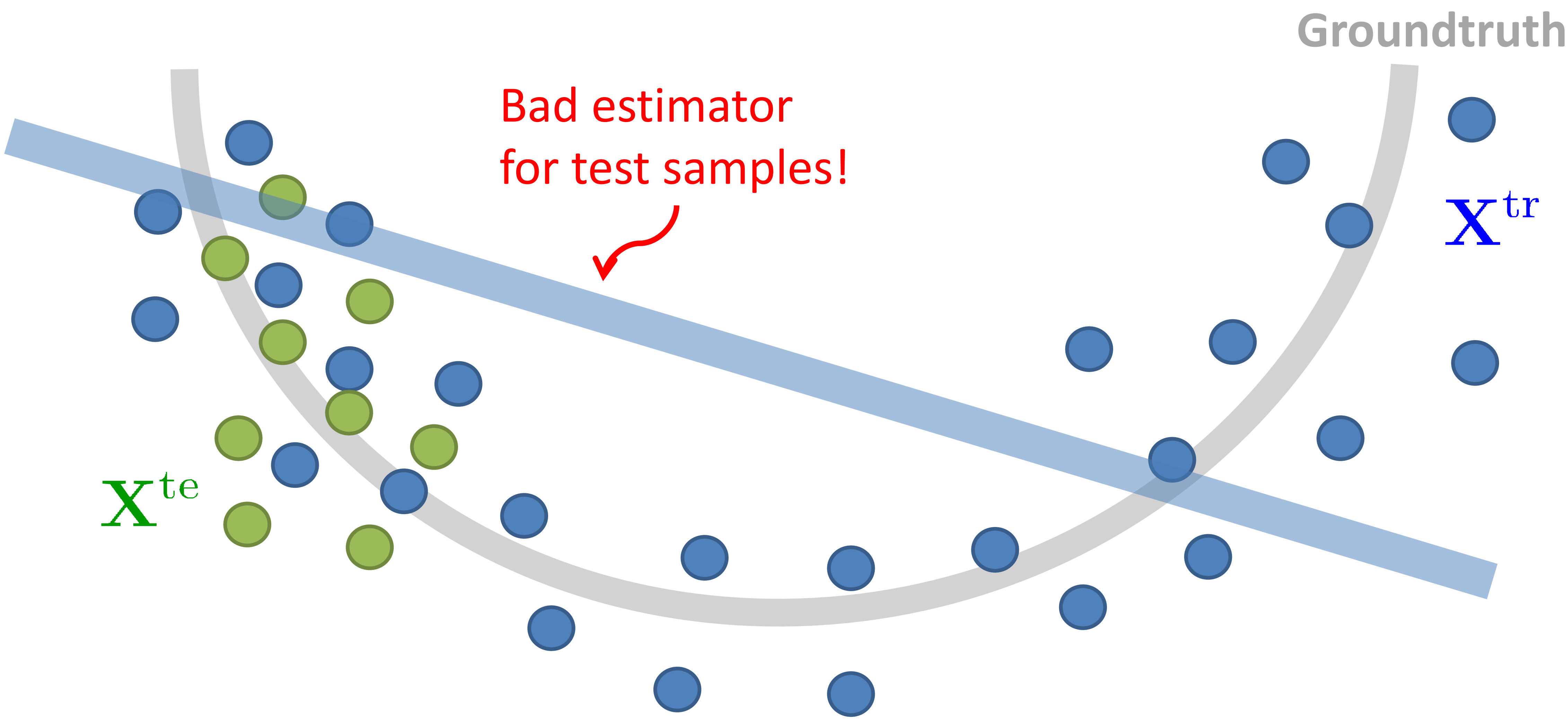


- Compound emotion categories



[Image credits: Jeffrey F. Cohn and Alexi Martinez]

Goal (2): **Minimize distribution mismatch**



Goal (2): **Minimize distribution mismatch**