# Joint Patch and Multi-label Learning for Facial Action Unit Detection

[1]Beijing University of Posts and Telecommunications [2]Carnegie Mellon University [3]University of Pittsburgh

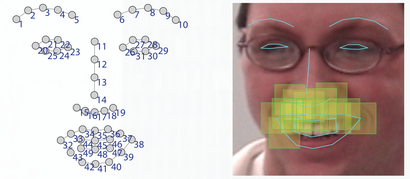
Kaili Zhao[1] Wen-Sheng Chu[2] Fernando De la Torre[2] Jeffery F. Cohn[2,3] Honggang Zhang[1]

## Problem

- Facial Action Unit (AU) detection

- Observations

(1) AUs describe region-based muscle movements; there are local dependencies between features.



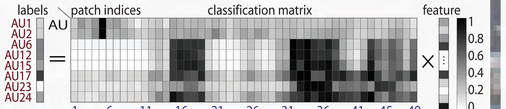
(2) The combinations of AUs produce basic expressions; there are co-occurence and competition between AUs.

- Objectives

(1) Identify discriminative patch subset for each AU (patch learning)
(2) Incorporate AU relations into model learning (multi-label learning)

$$\min_{\mathbf{W}} \underbrace{L(\mathbf{W},\mathcal{D}) + \alpha\Omega(\mathbf{W})}_{\text{patch learning}} + \underbrace{\Psi(\mathbf{W},\mathbf{X})}_{\text{multi-label learning}}$$

$$\text{where } L(\mathbf{W},\mathcal{D}) = \sum_{l=1}^{L}\sum_{i\in\mathcal{I}_l} ln(1+exp(-y_{il}\mathbf{w}_l^\top \mathbf{x}_i))$$

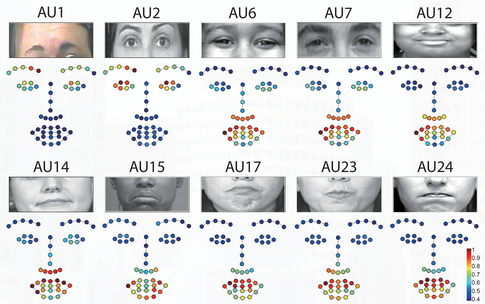



## Patch Learning (PL)

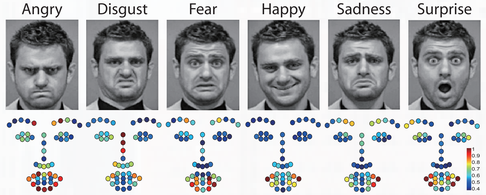
- Model

$$\min_{\mathbf{W}} L(\mathbf{W},\mathbf{D}) + \alpha\Omega(\mathbf{W}), \text{ where } \Omega(\mathbf{W}) = \sum_{l=1}^{L}\sum_{p=1}^{49}\|\mathbf{w}_l^p\|_2$$

- Patch importance learned for AUs

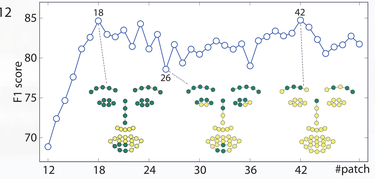



- Patch importance learned for basic expressions



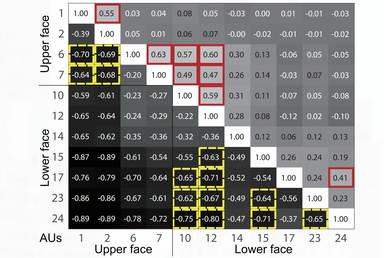

- The number of patches vs performance

▲F1-Norm of AU12 on CK+ dataset




## Multi-label Learning (ML)

- Observed AU relations




- AU relations discovered from >350,000 frames

| AU relations | AU groups |
|---|---|
| Positive correlation | (1,2), (6,7), (6,10), (7,10), (6,12), (7,12), (10,12), (17,24) |
| Negative competition | (1,6), (1,7), (2,6), (2,7), (10,17), (10,23), (10,24), (12,15), (12,17), (12,23), (12,24), , (15,23), (15,24), (23,24) |

- Model

$$\min_{\mathbf{W}} L(\mathbf{W},\mathcal{D}) + \Psi(\mathbf{W},\mathbf{X})$$

$$\Psi(\mathbf{W},\mathbf{X}) = \beta_1 PC(\mathbf{W},\mathbf{X},\mathcal{P}) + \beta_2 NC(\mathbf{W},\mathbf{X},\mathcal{N})$$

▲ Regularizer with positive correlation:

$$PC(\mathbf{W},\mathbf{X},\mathcal{P}) = \frac{1}{2}\sum_{(i,j)\in\mathcal{P}} \gamma_{ij}\|\mathbf{w}_i^\top\mathbf{X} - \mathbf{w}_j^\top\mathbf{X}\|_2^2$$

▲ Regularizer with negative competition [2]:

$$NC(\mathbf{W},\mathbf{X},\mathcal{N}) = \sum_{i=1}^{N}\sum_{n=1}^{|\mathcal{N}|}\Big(\sum_{j\in\mathcal{N}_n}|\mathbf{w}_j^\top\mathbf{x}_i|\Big)^2$$

[1] S. Taheri, Q. Qiu, and R. Chellappa. Structure-preserving sparse decomposition for facial expression analysis. TIP, 23(8):3590–3603, Aug 2014.
[2] Y. Zhou, R. Jin, and S. Hoi. Exclusive lasso for multi-task feature selection. In AISTATS, 2010.

## JPML Optimization

- ADMM

$$\min_{\mathbf{W}_1,\mathbf{W}_2} L(\mathbf{W}_1,\mathcal{D}) + \alpha\Omega(\mathbf{W}_1) + \Psi(\mathbf{W}_2,\mathbf{X}) + \frac{\rho}{2}\|\mathbf{W}_1 - \mathbf{W}_2\|_F^2$$
$$\text{s.t.} \mathbf{W}_1 = \mathbf{W}_2$$

▲ Algorithms

**Algorithm 1** Patch learning (PL)

**Input:** Training data $\mathcal{D}=\{(\mathbf{x}_i,\mathbf{y}_i)\}_{i=1}^N$, ML matrix $\mathbf{W}_2$, Lagrange multiplier of ADMM $\rho$ and U, learning rate $\eta_1$, and penalty parameter $\alpha$.
**Output:** PL matrix $\mathbf{W}_1 \in \mathbb{R}^{D\times L}$ with sparse groups of rows.

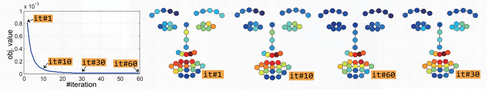
```
for ℓ = 1, ..., L do
    w_ℓ^(0) = 1/L 1_D, v^(0) = 1/L 1_D, a^(0) = 1, t = 0; // Initialization
    while not convergence do
        z^(t) = v^(t) − η_1 (∇L(w_ℓ^(t), D) + u_ℓ^(t) + ρ(w_ℓ^(t) − w_ℓ^(t)));
        for p = 1, ..., 49 do
            w_ℓ^p(t+1) = I(||z^p(t)||_2 > α)(1 − α/||z^p(t)||_2) z^p(t);
            // w_ℓ^p is the p-th patch within the ℓ-th column of W_1
        end for
        a^(t+1) = (1 + sqrt(1 + 4(a^(t))^2))/2;
        v^(t+1) = w_ℓ^(t+1) + ((1 − a^(t))/a^(t+1))(w_ℓ^(t+1) − w_ℓ^(t));
        t = t + 1;
    end while
end for
```

**Algorithm 2** Multi-label learning (ML)

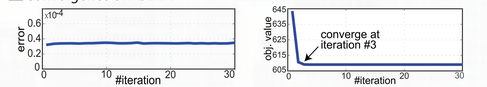
**Input:** Training data $\mathcal{D}=\{(\mathbf{x}_i,\mathbf{y}_i)\}_{i=1}^N$, PL matrix $\mathbf{W}_1$, Lagrange multiplier of ADMM $\rho$ and U, learning rate $\eta_2$, penalty parameter $\beta_2$, and accuracy control parameter $\mu$.
**Output:** ML matrix $\mathbf{W}_2 \in \mathbb{R}^{D\times L}$.

```
W_2^(0) = 1_{D×L}, V^(0) = 1_{D×L}, a^(0) = 1, t = 0; // Init.
while not convergence do
    U^(t) = (1 − a^(0))W_2^(t) + a^(t)V^(t);
    H_t = 0_{L×D};
    for i = 1, ..., N do
        z_i = min(1, max(−1, y_i^(t)^⊤ x_i));
        q_i = z_i U^(t)^⊤ x_i − ||x_i||_2^2;
        H_t = H_t + q_i (x_i x_i^⊤);
    end for
    V^(t+1) = V^(t) − η_2 (H_t^⊤ u + ρ(W_1 − U^(t))) + ∇PC(U^(t));
    W_2^(t+1) = (1 − a^(t))W_2^(t) + a^(t)V^(t+1);
    a^(t+1) = (1 + sqrt(1 + 4(a^(t))^2))/2;
    t = t + 1;
end while
```

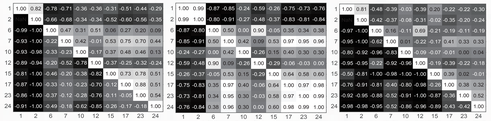
▲ Convergence of patch learning




▲ Convergence of ADMM




(a) Error of JPML (b) Obj. value of JPML



(c) Relation matrix in CK+ (d) it#1 Patch Learning (e)it#30 JPML

★ The difference of correlation coefficent between (c) and (d) is 0.51, and that between (c) and (e) is 0.15

[3] L. Zhong, Q. Liu, P. Yang, B. Liu, J. Huang, and D. N. Metaxas. Learning active facial patches for expression analysis. In CVPR, 2012.
[4] X. Zhang, M. H. Mahoor, S. M. Mavadati, and J. F. Cohn. A lp-norm mtmkl framework for simultaneous detection of multiple facial AUs. In WACV, 2014
[5] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews. The extended Cohn-Kanade dataset (CK+): A complete dataset for action unit and emotion-specified ex- pression. In CVPRW, 2010.
[6] M. A. Sayette et al. Alcohol and group formation a multimodal investigation of the effects of alcohol on social bonding. Psychological science, 2012.
[7] X. Zhang et al. A high-resolution spontaneous 3D dynamic facial expression database. In Automatic Face and Gesture Recognition Workshop, 2013.

## Experiments

- Settings
  - ▲ SIFT descriptor on pre-determined 49 facial landmarks
  - ▲ Leave-one-subject-out protocol
- Metrics

$$\text{F1-Norm} = \frac{2s\cdot R\cdot P}{s\cdot R + P}, \qquad \text{F1-Event} = \frac{2\cdot ER\cdot EP}{ER+EP}$$

- Methods
  - ▲ SP-SVM: Self-defined Patches SVM
  - ▲ SPSD [1]: Structure Preserving Sparse Decomposition
  - ▲ LSVM: Linear SVM
  - ▲ LL1: L1-regularized logistic regression
  - ▲ APL [3]: Active Patch Learning
  - ▲ MT-MKL [4]: Multi-Task Multiple Kernel Learning
  - ▲ JPML: Our method

- CK+ [5]
  - ▲ 593 posed videos from 123 subjects

| | F1-Norm | | | | | | |
|---|---|---|---|---|---|---|---|
| AU | SP-SVM | SPSD | LSVM | LL1 | MT-MKL | ML | APL | JPML |
| 1 | 61.8 | 44.4 | 85.8 | 83.4 | 73.0 | **89.0** | 86.4 | **[90.0]** |
| 2 | 63.9 | 47.9 | 90.9 | 87.4 | 87.8 | **92.7** | 86.6 | **[93.0]** |
| 6 | 61.7 | 34.2 | **75.3** | **[76.2** | 61.9 | 70.7 | 70.5 | 74.2 |
| 7 | 60.0 | 50.8 | **[70.8]** | **70.0** | – | 61.6 | 62.8 | 66.7 |
| 12 | 65.5 | 47.4 | **[80.7]** | 80.0 | 73.3 | 75.2 | 76.6 | **[80.7]** |
| 14 | 66.3 | 59.9 | **67.7** | **67.7** | – | – | **[69.5]** | – |
| 15 | 65.8 | 53.4 | 67.5 | 66.7 | 67.8 | 61.9 | **79.2** | **73.6** |
| 17 | 60.4 | 62.2 | **80.5** | **80.5** | 68.3 | 80.3 | 80.0 | **[83.5]** |
| 23 | 66.2 | 65.0 | 69.3 | 69.8 | – | 69.7 | **83.5** | **74.3** |
| 24 | 68.3 | 65.8 | 71.1 | **71.4** | – | 67.5 | **75.9** | 65.9 |
| *AA.* | 64.0 | 52.3 | **76.0** | 75.3 | – | – | **[77.1]** | – |
| *AR.* | 63.7 | 51.5 | 76.9 | 76.2 | 72.0 | 74.3 | **77.0** | **[78.0]** |

- GFT [6]
  - ▲ 50 2-minute spontaneous videos from 50 participants

| | F1 Norm | | | | | | | F1 Event | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AU | SPSVM | SPSD | LSVM | LL1 | ML | APL | JPML | SPSVM | SPSD | LSVM | LL1 | ML | APL | JPML |
| 1 | 29.9 | 33.0 | 53.0 | 52.0 | **[66.7]** | 44.1 | **58.0** | **17.8** | 12.2 | **[20.6]** | 17.8 | 11.5 | 11.5 | 15.0 |
| 2 | 60.2 | 34.7 | 51.3 | 45.1 | **[64.4]** | 43.6 | **63.2** | **[21.2]** | 12.9 | 19.6 | 16.8 | 12.5 | **16.6** | 15.0 |
| 6 | **[77.2]** | 34.8 | 74.7 | 75.2 | 57.3 | **77.2** | **[79.6]** | 46.6 | 21.6 | 33.2 | 42.3 | 25.5 | **50.3** | **[50.8]** |
| 7 | 56.5 | 40.3 | **72.7** | 70.5 | 67.6 | **[73.6]** | **[73.6]** | 41.3 | 25.3 | 38.2 | 34.4 | 34.3 | **47.9** | **[54.7]** |
| 10 | 74.6 | 41.8 | 75.8 | **77.5** | – | **[78.6]** | – | **45.6** | 30.7 | 41.2 | 37.9 | – | **50.2** | – |
| 12 | 77.1 | 76.2 | 79.2 | 80.2 | 67.1 | **81.3** | **[84.1]** | 47.9 | **48.6** | 47.9 | 48.4 | 15.3 | **[53.6]** | 46.7 |
| 14 | 64.1 | **68.9** | 68.5 | **[70.4]** | – | 66.7 | – | 42.1 | 49.0 | 42.1 | **55.0** | – | **[60.6]** | – |
| 15 | 47.2 | 30.1 | 45.8 | 65.3 | 66.3 | **[67.1]** | **66.2** | 16.4 | 10.6 | **39.1** | **[39.7]** | 17.8 | 18.9 | **37.9** |
| 17 | 51.8 | 32.8 | 47.6 | 46.8 | 67.1 | **[74.5]** | **72.0** | 33.8 | 22.9 | 38.3 | **38.9** | 27.1 | **[48.7]** | 38.8 |
| 23 | 49.7 | 35.9 | 38.8 | 43.5 | **[66.9]** | **63.9** | 60.0 | 25.9 | 18.0 | **35.4** | 28.4 | 28.6 | 35.0 | **[37.6]** |
| 24 | 51.1 | 35.3 | 56.6 | 59.2 | 67.1 | **79.0** | **[79.3]** | 18.7 | 12.9 | **27.3** | 25.0 | 26.7 | 19.2 | **[35.5]** |
| *AA.* | 56.5 | 42.3 | 59.8 | 55.4 | – | **[67.1]** | – | 31.1 | 23.4 | 34.2 | 34.6 | – | **[36.3]** | – |
| *AR.* | 53.6 | 39.4 | 57.0 | 51.3 | 65.6 | **65.9** | **[70.7]** | 38.3 | 19.7 | 32.5 | 32.0 | 22.1 | **32.7** | **[37.0]** |

- BU4D [7]
  - ▲ 48 spontaneous videos from 41 participants

| | F1 Norm | | | | | | | F1 Event | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AU | SPSVM | SPSD | LSVM | LL1 | ML | APL | JPML | SPSVM | SPSD | LSVM | LL1 | ML | APL | JPML |
| 1 | 22.9 | 27.6 | 40.6 | 35.6 | **[58.6]** | **56.0** | 55.5 | 9.5 | 10.6 | 13.0 | 11.7 | 14.9 | **17.1** | **[17.5]** |
| 2 | 15.8 | 15.8 | 32.1 | 24.1 | 56.9 | **60.2** | **62.7]** | 7.6 | 7.6 | 11.5 | 10.3 | 14.0 | **16.0** | **[17.2]** |
| 6 | 45.5 | 54.6 | 59.4 | **75.2** | 62.9 | 75.0 | **[75.7]** | 21.9 | 27.9 | 17.2 | 21.1 | 15.6 | **[32.7]** | **30.0** |
| 7 | 44.1 | 56.0 | 55.7 | **[70.5]** | 66.7 | 64.3 | **66.7** | 22.9 | **30.5** | 20.5 | 23.6 | 17.1 | **[33.7]** | 26.3 |
| 10 | 50.1 | 55.6 | 63.0 | **[74.3]** | – | **72.9** | – | 29.7 | 32.8 | 22.0 | **34.1** | – | **41.0** | – |
| 12 | 46.5 | 54.9 | 62.5 | 82.0 | 67.1 | **[82.3]** | **81.4** | 28.4 | 30.6 | 23.4 | 25.0 | 20.5 | **[41.3]** | **31.6** |
| 14 | 44.2 | 52.7 | 51.5 | 61.2 | – | **66.0** | – | 19.3 | 28.3 | 23.5 | 29.3 | – | **29.8** | – |
| 15 | 13.2 | 40.5 | 49.6 | 56.3 | 66.0 | **[68.4]** | **65.9** | **23.4** | 22.9 | 23.9 | 18.6 | 20.4 | 13.1 | **[30.1]** |
| 17 | 42.3 | 46.9 | 40.3 | 63.4 | 66.7 | **[69.2]** | **65.3** | 19.3 | 21.9 | 21.2 | 25.6 | 20.8 | **[33.5]** | **29.4** |
| 23 | 11.3 | 23.9 | 42.1 | 57.2 | 67.1 | **[68.0]** | **65.2** | **19.4** | 19.6 | **[21.8]** | 19.0 | 20.6 | 16.2 | **27.7** |
| 24 | 7.3 | 47.3 | 21.3 | 69.5 | 66.7 | **[78.1]** | **77.3** | 17.7 | 18.4 | 19.0 | **[23.1]** | 20.4 | 13.2 | **[26.4]** |
| *AA.* | 29.3 | 42.1 | 47.1 | 59.5 | – | **[68.7]** | – | 18.9 | 21.8 | 19.7 | 23.1 | – | **[24.6]** | – |
| *AR.* | 25.3 | 39.4 | 44.8 | 57.7 | 64.3 | **[68.5]** | **68.4** | 15.9 | 19.9 | 19.0 | 21.1 | 18.3 | **22.2** | **[26.2]** |