Facial Expression Recognition Using Spatial-temporal Semantic Graph Network

Jinzhao Zhou, Xingming Zhang, Yang Liu
Department of Computer Science and Engineering,
South China University of Technology, China

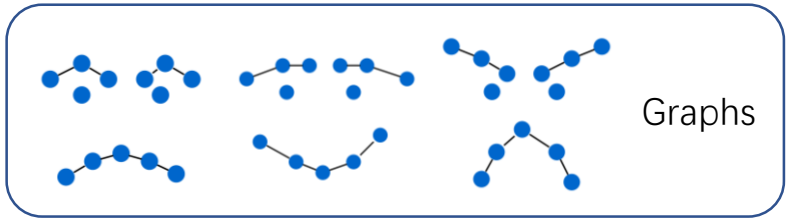
Xiangyuan Lan
Department of Computer Science,
Hong Kong Baptist University, HK, China

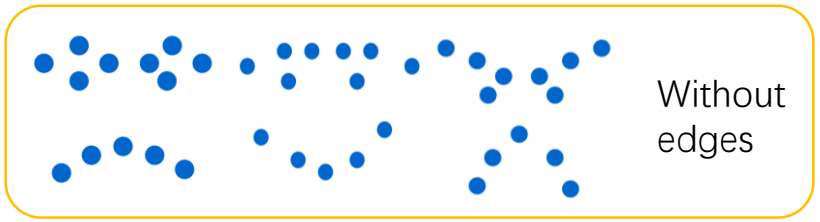
presenter: Jinzhao Zhou (Charles)

1. Overview
2. Review & Intuition
3. Method
4. Experiments Settings & Results
5. Future Work

Intuition & Review of recent FER methods

Cartoon Expressions





Problems to be solved
- Node Selection
- Edge Selection
- Representation

Image Classification

Appearance-based
- Holistic Image
- Local Patch

Geometry-based
- Hand-crafted
- Network fusion

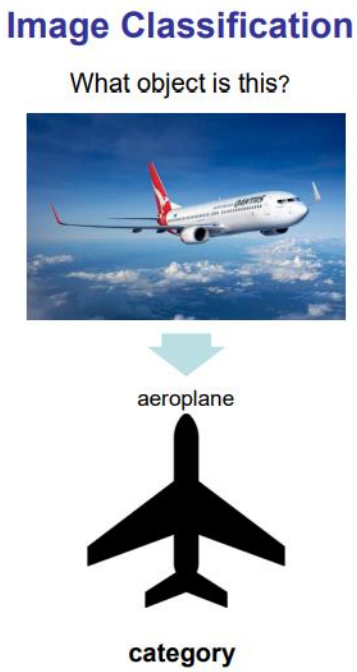
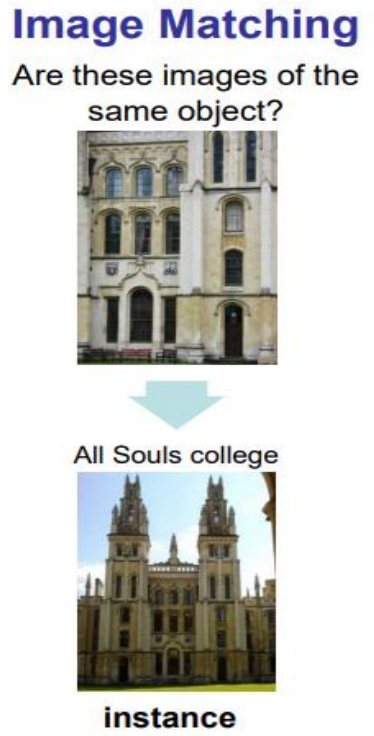


Image Matching

Elastic graph matching
- Hand-crafted
- Network fusion

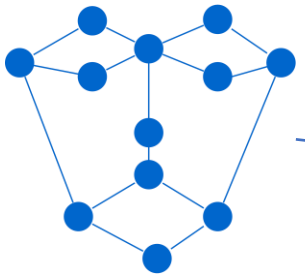
Feature similarity

...



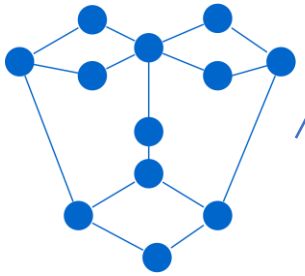
Method Overview

Abstraction

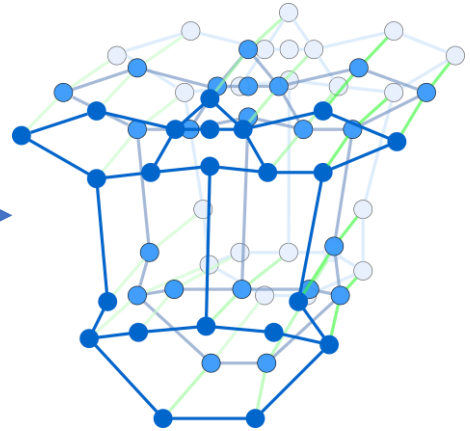


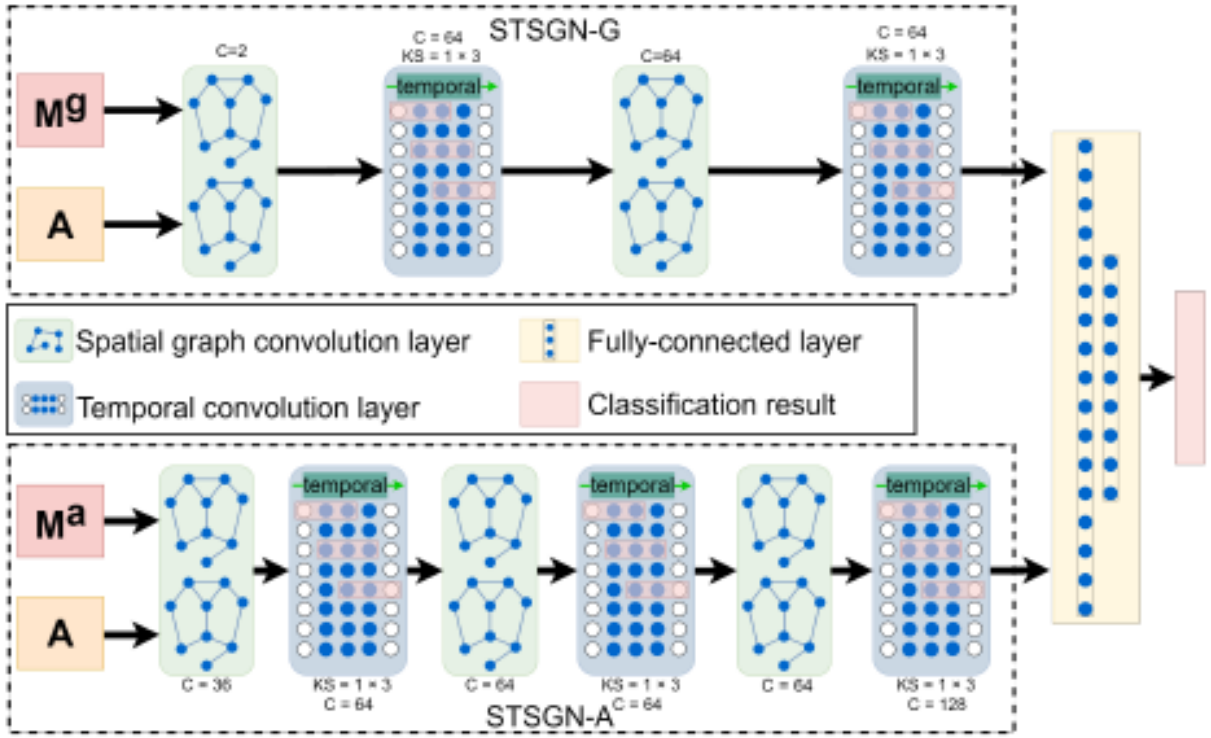
... ...

Abstraction



Temporal extension





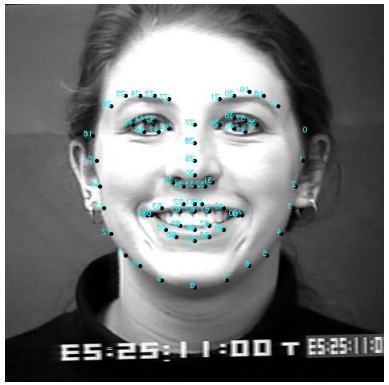
Benefits:

- The input dimensionality is greatly reduced
- The structural information is mostly retained

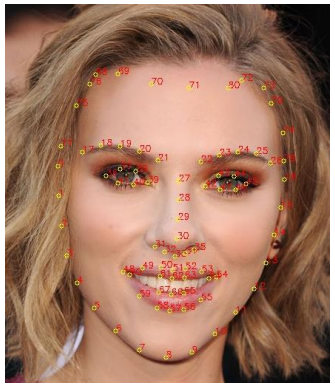
Proposed Method

1. Landmark Location

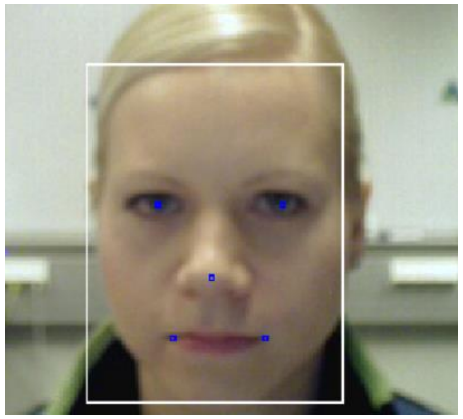
Dlib 68 landmarks



Dlib 81 landmarks

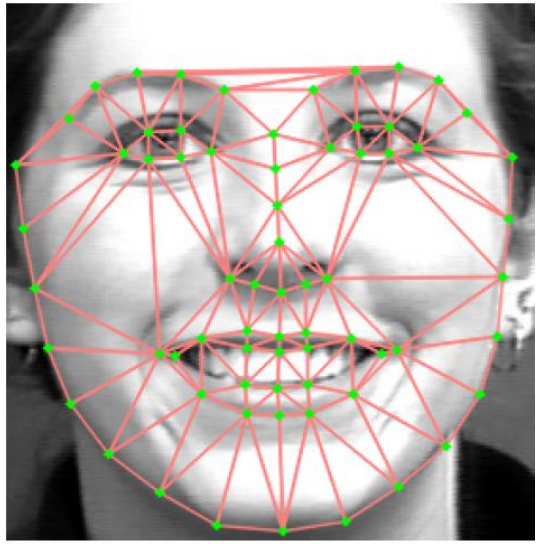


mtcnn 5 landmarks

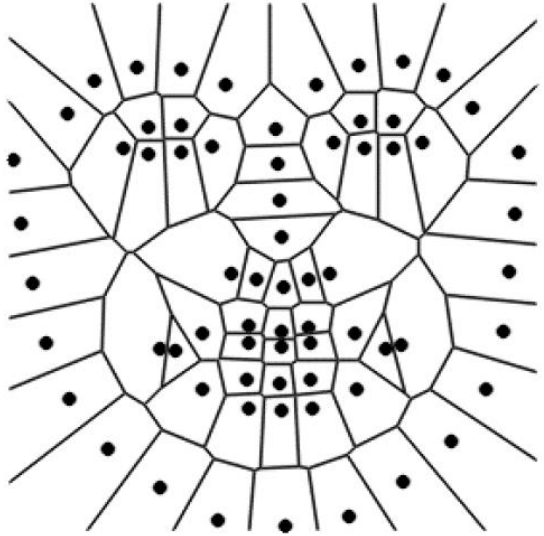


2. Graph Generation

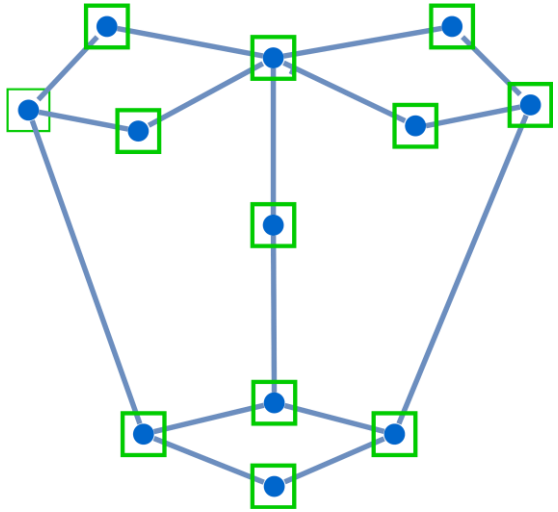
Delaunay Triangulation



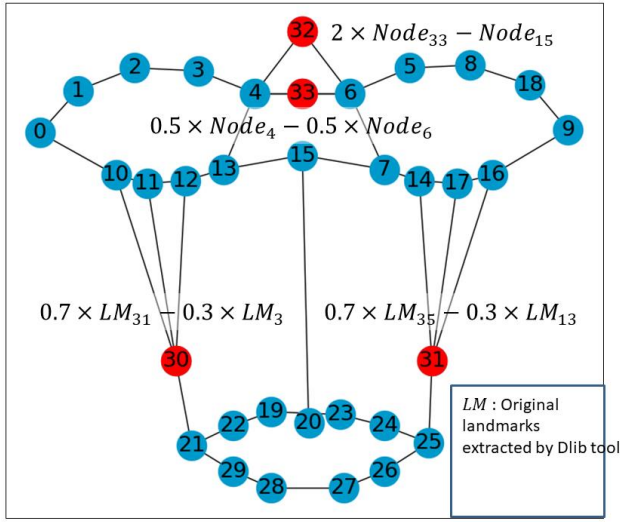
Voronoi Diagram

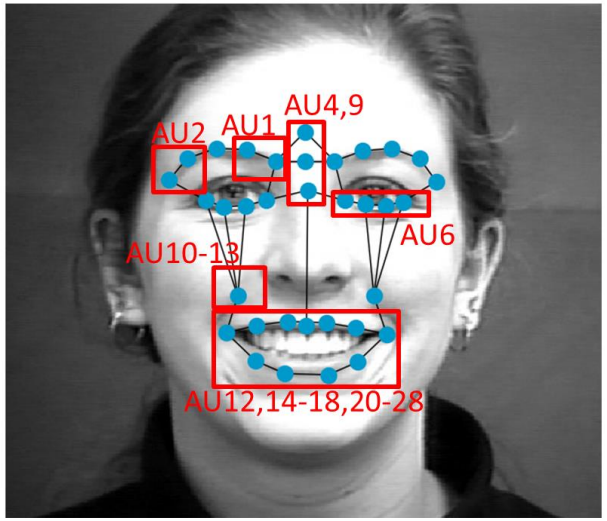


Landmarks and links



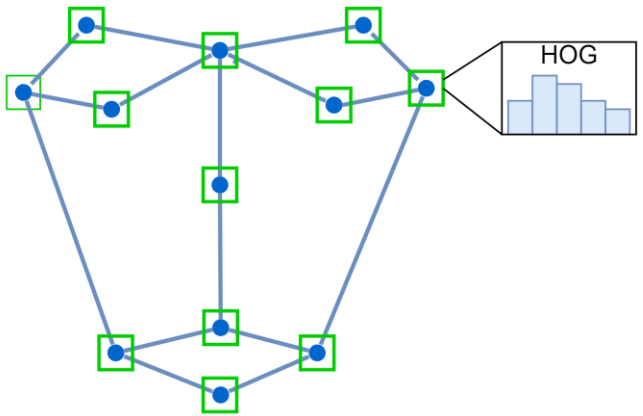
3. Landmark Selection based on psychology

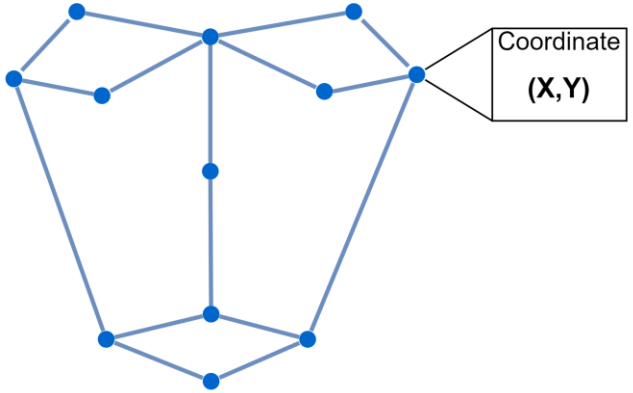




AU	Description
1	Inner Brow Raiser
2	Outer Brow Raiser
4	Brow Lowerer
5	Upper Lid Raiser
6	Cheek Raiser
7	Lid Tightener
9	Nose Wrinkler
10	Upper Lip Raiser
11	Nasolabial Deepener
12	Lip Corner Puller
13	Cheek Puffer
14	Dimpler
15	Lip Corner Depressor
16	Lower Lip Depressor
17	Chin Raiser
18	Lip Puckerer
20	Lip stretcher
22	Lip Funneler
23	Lip Tightener
24	Lip Pressor
25	Lips part**
26	Jaw Drop
57	Head forward
58	Head back
61	Eyes turn left
62	Eyes turn right
63	Eyes up
64	Eyes down
27	Mouth Stretch
28	Lip Suck
41	Lid droop**
42	Slit
43	Eyes Closed
44	Squint
45	Blink
46	Wink
51	Head turn left
52	Head turn right
53	Head up
54	Head down
55	Head tilt left
56	Head tilt right

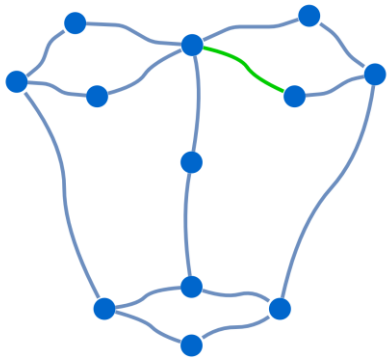
4. Node representation





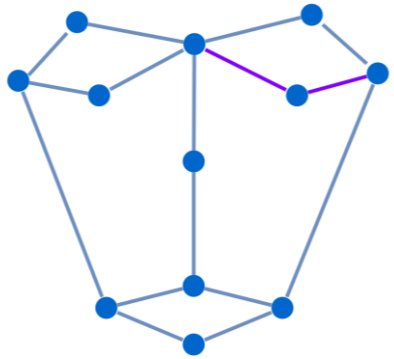
Expression	AU	
Happiness	6 ,12	Cheek Raiser, Lip Corner Puller
Sadness	1, 4,15	Inner Brow Raiser, Brow Lowerer, Lip Corner Depressor
Surprise	1,2 ,5 ,26	Inner Brow Raiser, Outer Brow Raiser, Upper Lid Raiser, Jaw Drop
Fear	1,2,4,5,7,20 ,26	Inner Brow Raiser, Outer Brow Raiser, Brow Lowerer, Upper Lid Raiser, Lid Tightener, Lip Stretcher, Jaw Drop
Anger	4, 5,7, 23	Brow Lowerer, Upper Lid Raiser, Lid Tightener, Lip Tightener
Disgust	9 ,15,16	Nose Wrinkler, Lip Corner Depressor, Lower Lip Depressor
Contempt	12,14	Lip Corner Puller, Dimple

4. Edge representation



Hop distance

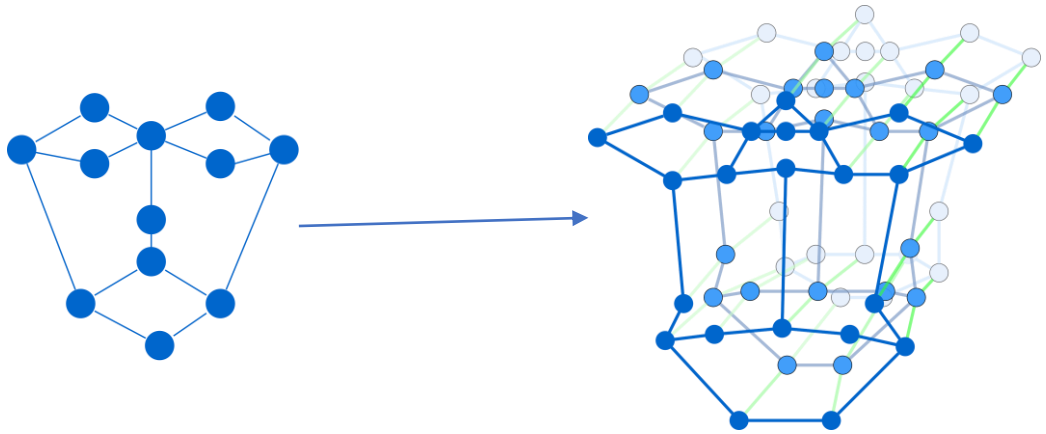
$$e_{ij} = \begin{cases} d(v_i, v_j), & \text{if } d(v_i, v_j) \le D \\ 0, & \text{otherwise} \end{cases}$$



Euclid distance

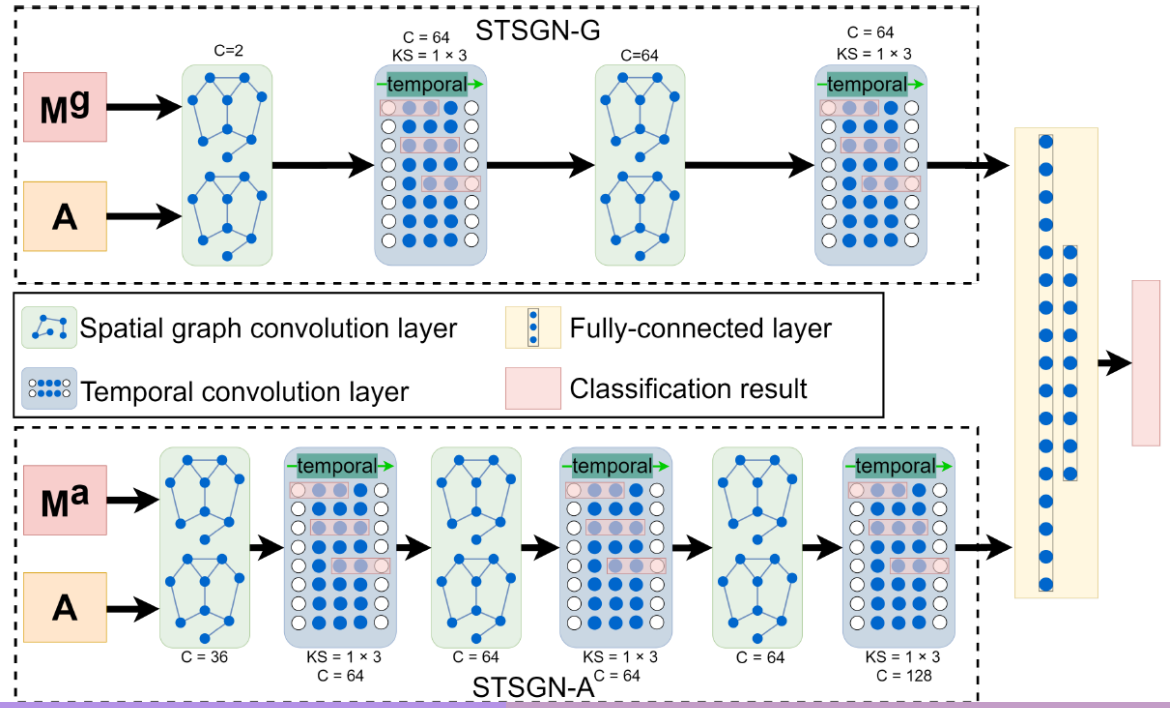
$$e_{ij} = \begin{cases} \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}, & \text{if } (i,j) \in S \\ 0, & \text{otherwise} \end{cases}$$

5. Network architecture



6. Network architecture

$$f_{out} = \sum M \times \Lambda^{-\frac{1}{2}} A \Lambda^{-\frac{1}{2}} \times f_{in} W$$



Visualization

Table 1. Comparison on different edge's strategy

Edge Strategy	CK+	Oulu-CASIA
Euclidean	98.92	83.09
Hop	98.63	87.23

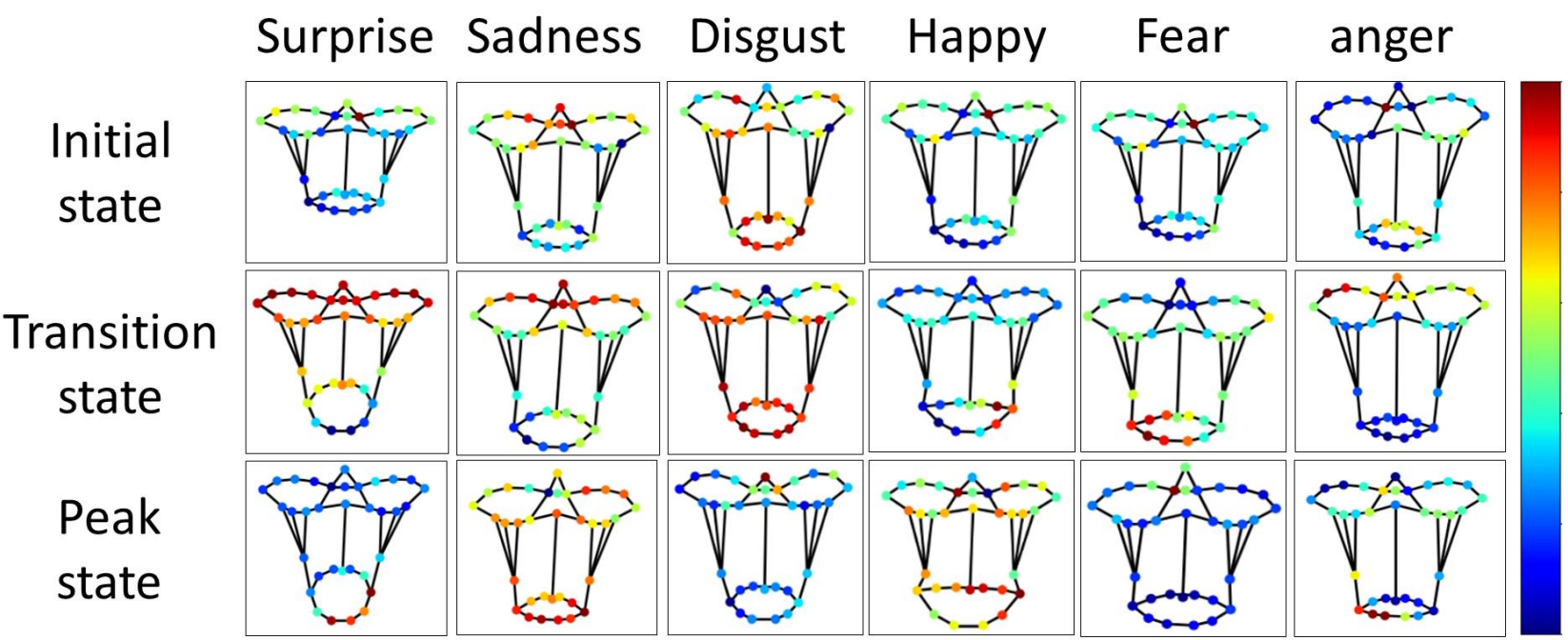


Table 2. Comparison on CK+ and Oulu Dataset

Method	CK+	Oulu-CASIA
STM-ExpLet	94.19	74.59
LOMo	95.10	82.10
DTAGN(Joint)	97.25	81.46
DAUGN	97.67	84.28
PHRNN-MSCNN	97.78	86.25
RCFN	97.84	86.94
STSGN-G	90.24	56.82
STSGN-A	95.71	64.14
STSGN	**98.63**	**87.23**

[1] The bolder one means best.

Conclusion

- spatial-temporal facial graph
- spatial-temporal graph convolutional network

Future work

- Better topology (dynamic graph)
- Better network