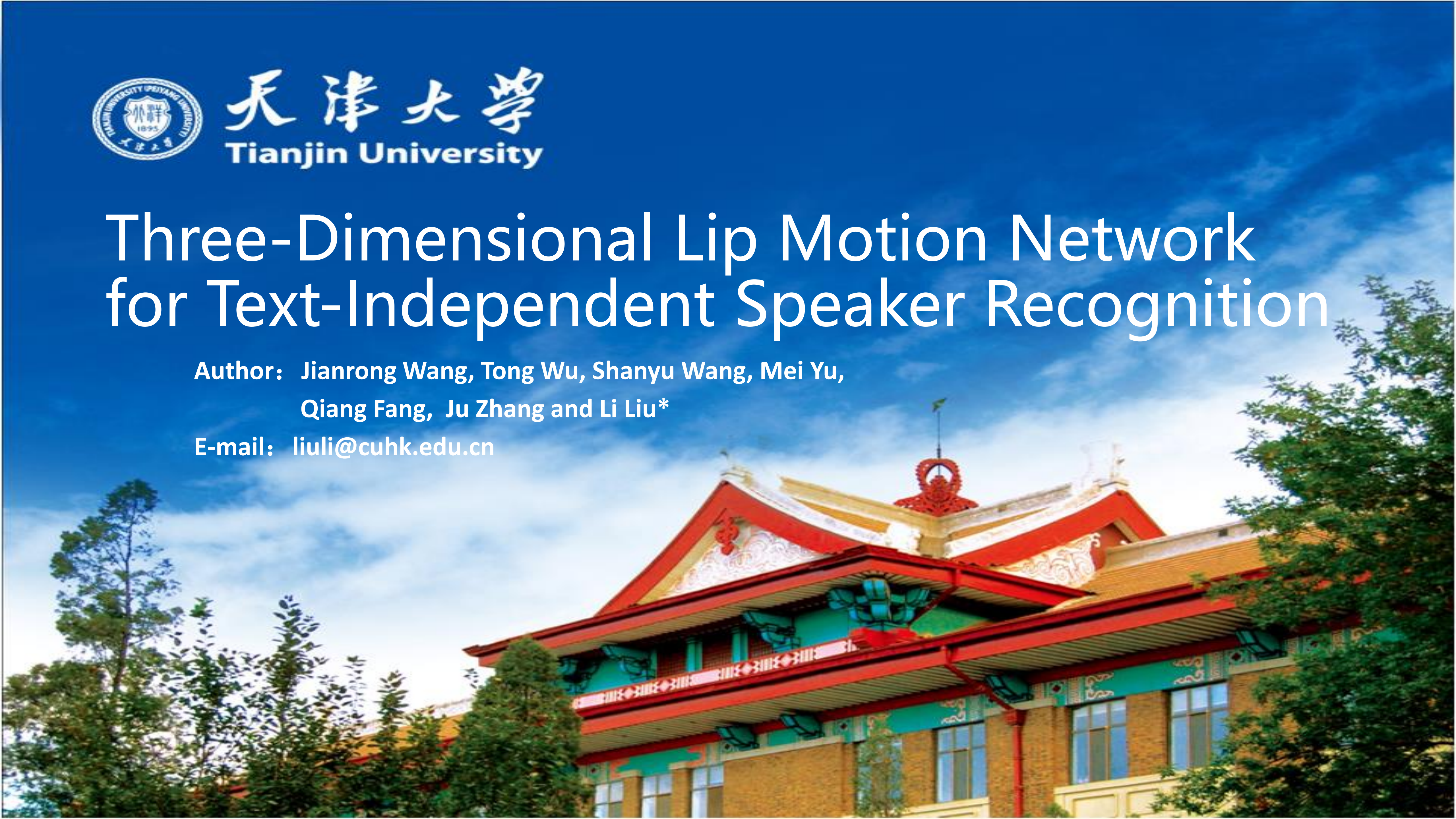


天津大学
Tianjin University

Three-Dimensional Lip Motion Network for Text-Independent Speaker Recognition

Author：Jianrong Wang, Tong Wu, Shanyu Wang, Mei Yu,
Qiang Fang, Ju Zhang and Li Liu*

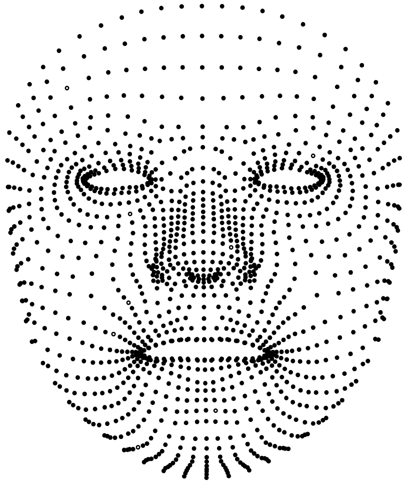
E-mail：liuli@cuhk.edu.cn

Introduction

- Lip motion can be used as a new kind of biometrics in speaker recognition. Lots of works used 2D lip images.
- However, 2D lip easily suffers from face orientations.
- To this end, we present a novel end-to-end 3D lip motion Network (3LMNet) by utilizing the sentence-level 3D lip motion (S3DLM) to recognize speakers.

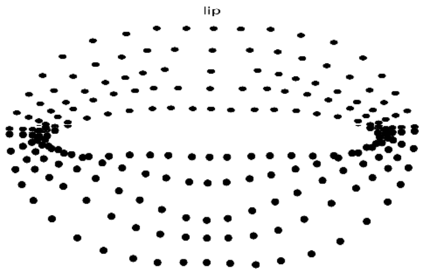
Method： S3DLM

- 200 lip points selected from the 1347 facial landmarks.
- 28 frames in each sentence represent the motion of lip.



1347 facial landmarks
in LSD-AV dataset

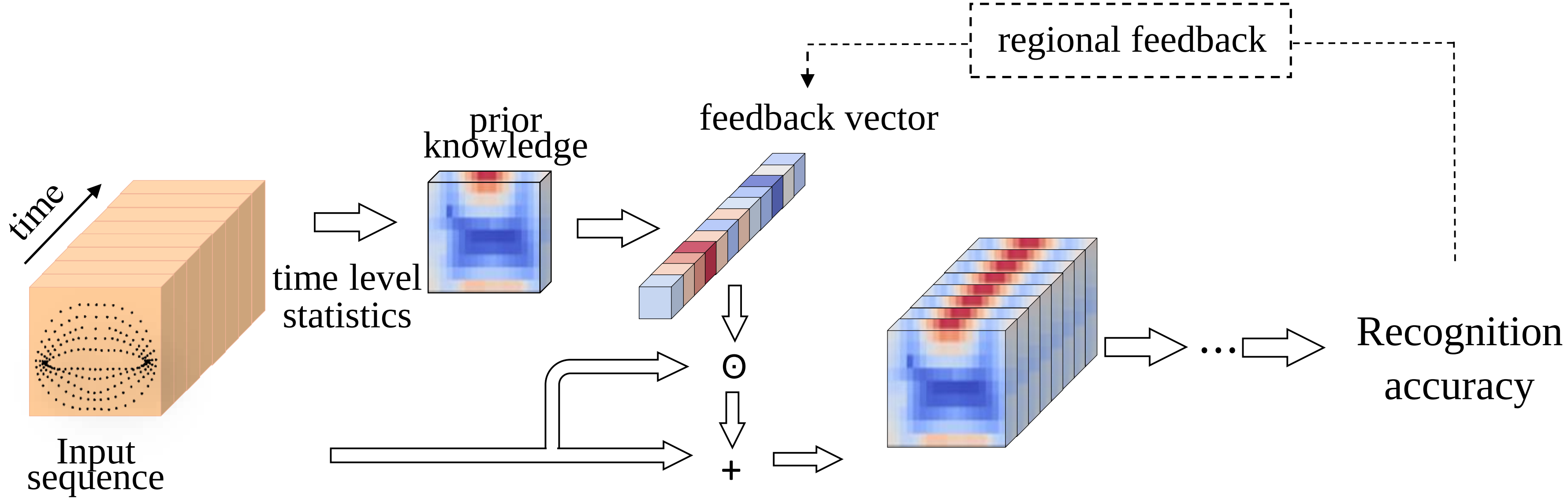
select



200 lip points

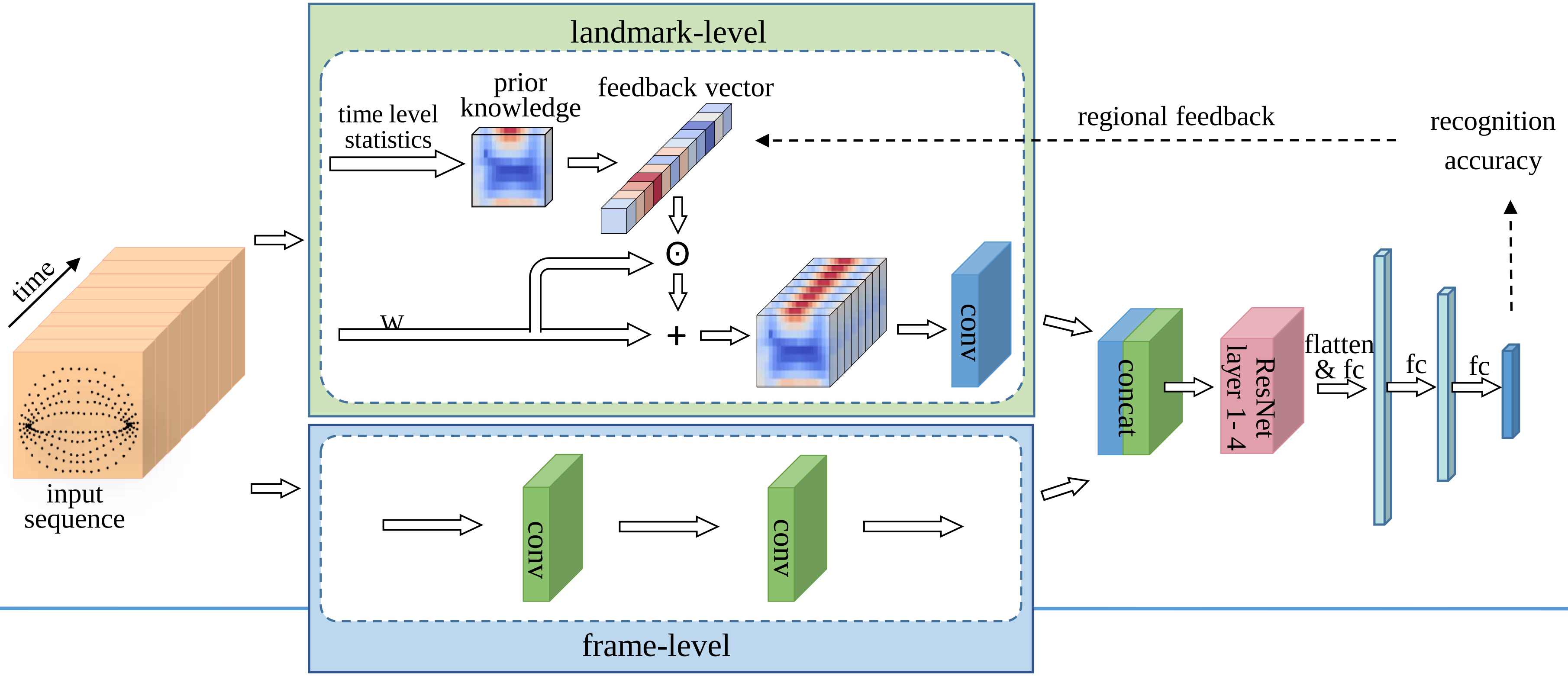
Method： RFM & prior knowledge

- RFM and prior knowledge of the lip motion is proposed to screen out key identifying information in lip motion.



Method：3LMNet

- The network learns landmark-level features, frame-level features of S3DLM sequences.



Experiment：

- Discussion of the relationship between text and lip motion

COMPARISON BETWEEN TEXT-BASED LIP MOTION IN THE LSD-AV DATASET.

Text-based sample	1-20	21-40	41-60	61-80	81-100	101-120	121-140	std
D_t	0.0067	-0.0048	0.0024	0.0038	-0.0017	0.0060	0.0004	**0.0046**

COMPARISON BETWEEN SPEAKER-BASED LIP MOTION IN THE LSD-AV DATASET.

Speaker-based sample	1-10	11-20	21-30	31-40	41-50	51-60	61-68	std
D_s	-0.0385	0.0293	-0.0096	0.0738	0.0187	-0.0465	-0.0272	**0.0431**

Experiment：

- Performance of the S3DLM sequences

TEXT-INDEPENDENT SPEAKER RECOGNITION ACCURACY COMPARISON OF SUPERIOR S3DLM SEQUENCE AND 2D LANDMARKS IN DIFFERENT NETWORKS.

Model	S3DLM	2D landmarks
LSTM	82.46%	76.23%
VGG-16	91.00%	87.10%
ResNet-34	93.50%	88.47%

Experiment:

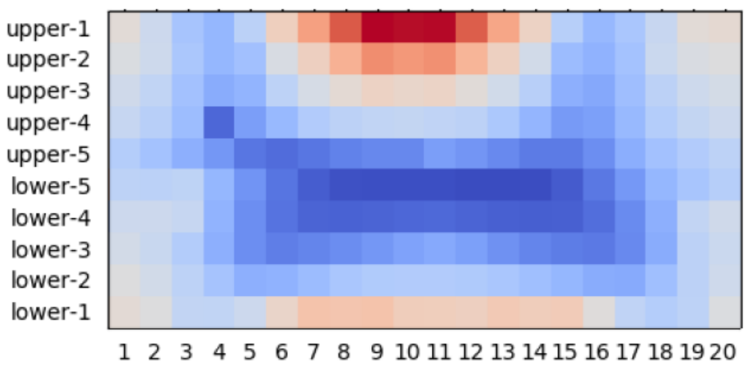
- Performance of the 3LMNet

ABLATION STUDY OF THE PROPOSED 3LMNET.

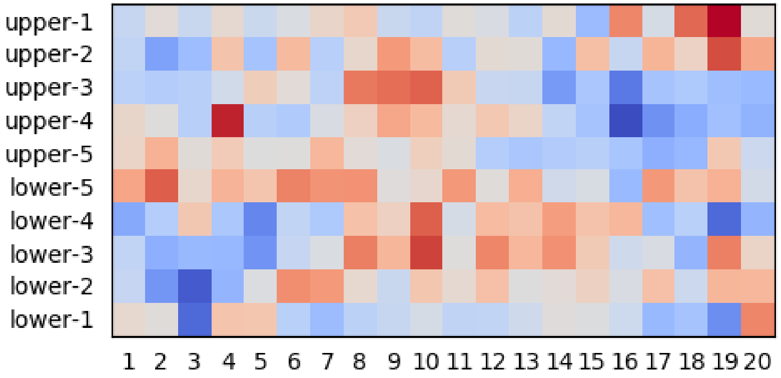
Model	Text-independent	Text-dependent
Lai et al.	—	92.61%
Liao et al.	—	97.11%
3LMNet-RFM-prior	93.91%	98.38%
3LMNet+RFM-prior	94.94%	98.73%
3LMNet+RFM+prioropp	91.94%	97.73%
3LMNet+RFM+prior	**95.22%**	**99.10%**

Experiment:

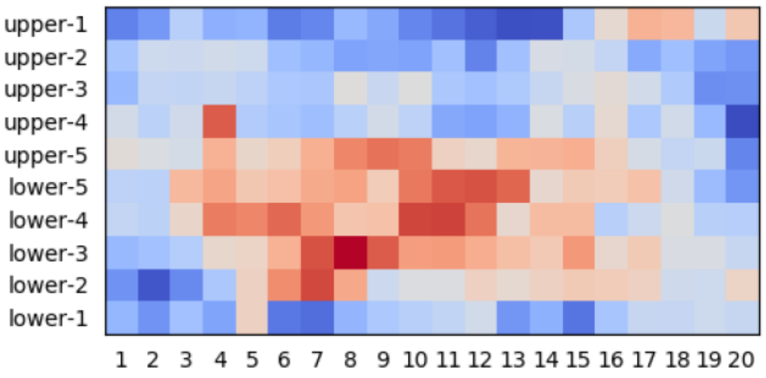
- Effect of RFM and the prior knowledge of lip motion



(a) Original lip points fluctuation



(b) Regional feedback visualization of lip points



(c) Regional feedback visualization with prior knowledge of lip points

THANKS！