

Beyond Face Recognition: Deep Face Analysis from a Micro Perspective

韩 琥

中科院计算所

2020.7.28



中国科学院计算技术研究所
Institute of Computing Technology, Chinese Academy of Sciences



Outline

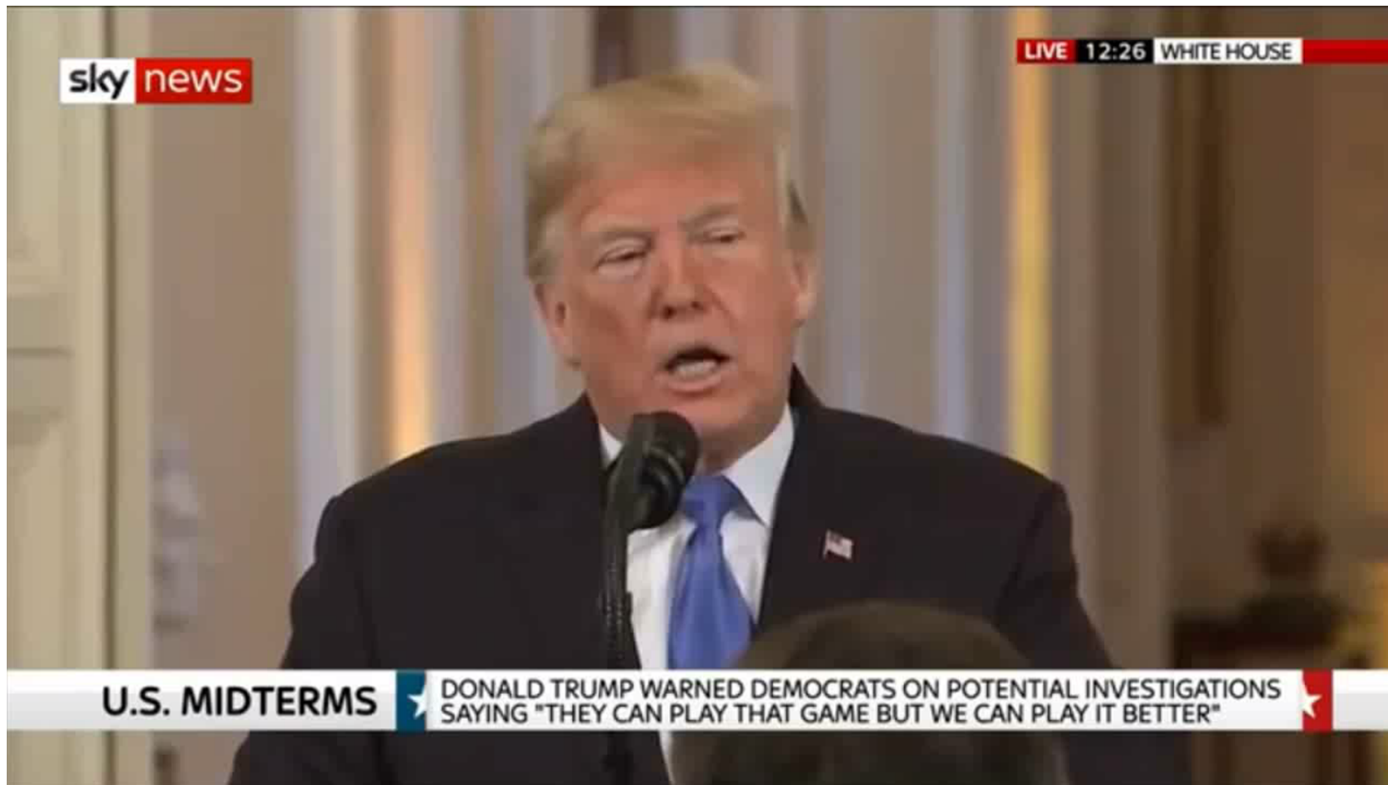
- ❖ What is the focus of micro-perspective face analysis
- ❖ Knowledge and data driven modeling
- ❖ Disentangled representation learning
- ❖ Summary



What can we get from a face image/video?

中科院计算所

Institute of Computing Technology, Chinese Academy of Sciences



Identity

Trump

Attribute

Old, White, Male

Expression/Emotion

Anger

Action Unit

4+5+7+23

Physiological signals

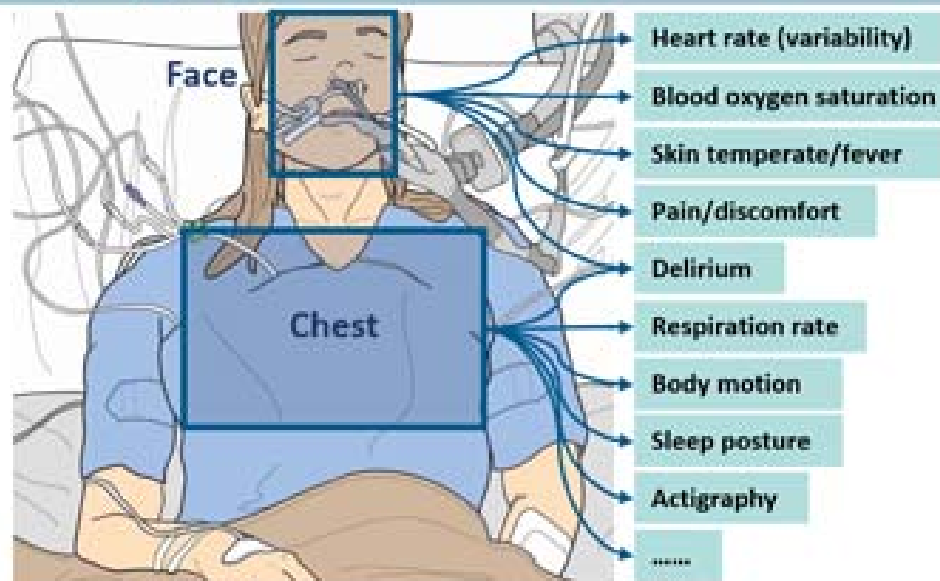
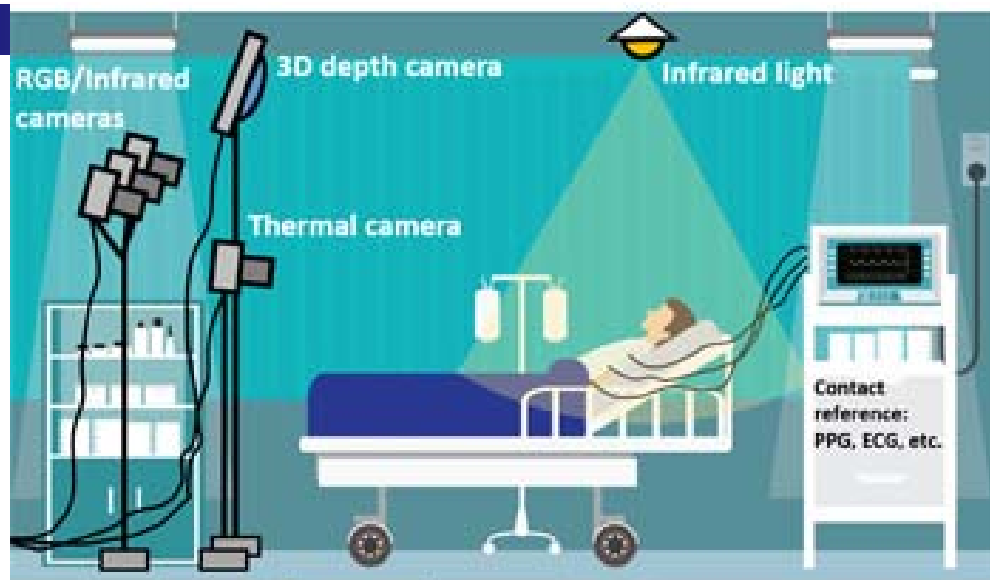
Heart rate, respiratory rate, ...



Vision based physiological signal measurement

中科院计算所

Institute of Computing Technology, Chinese Academy of Sciences



心率 (成人60-100)、心率变异性(逐次心跳周期差异的变化情况, 它含有神经体液因素对心血管系统调节的信息, 从而判断其对心血管等疾病的病情及预防, 可能是预测心脏性猝死和心律失常性事件的一个有价值的指标)

血氧饱和度(血液中被氧结合的氧合血红蛋白占全部可结合的血红蛋白的百分比, 是呼吸循环的重要生理参数, 正常95%以上)

呼吸频率(成人12~20次)

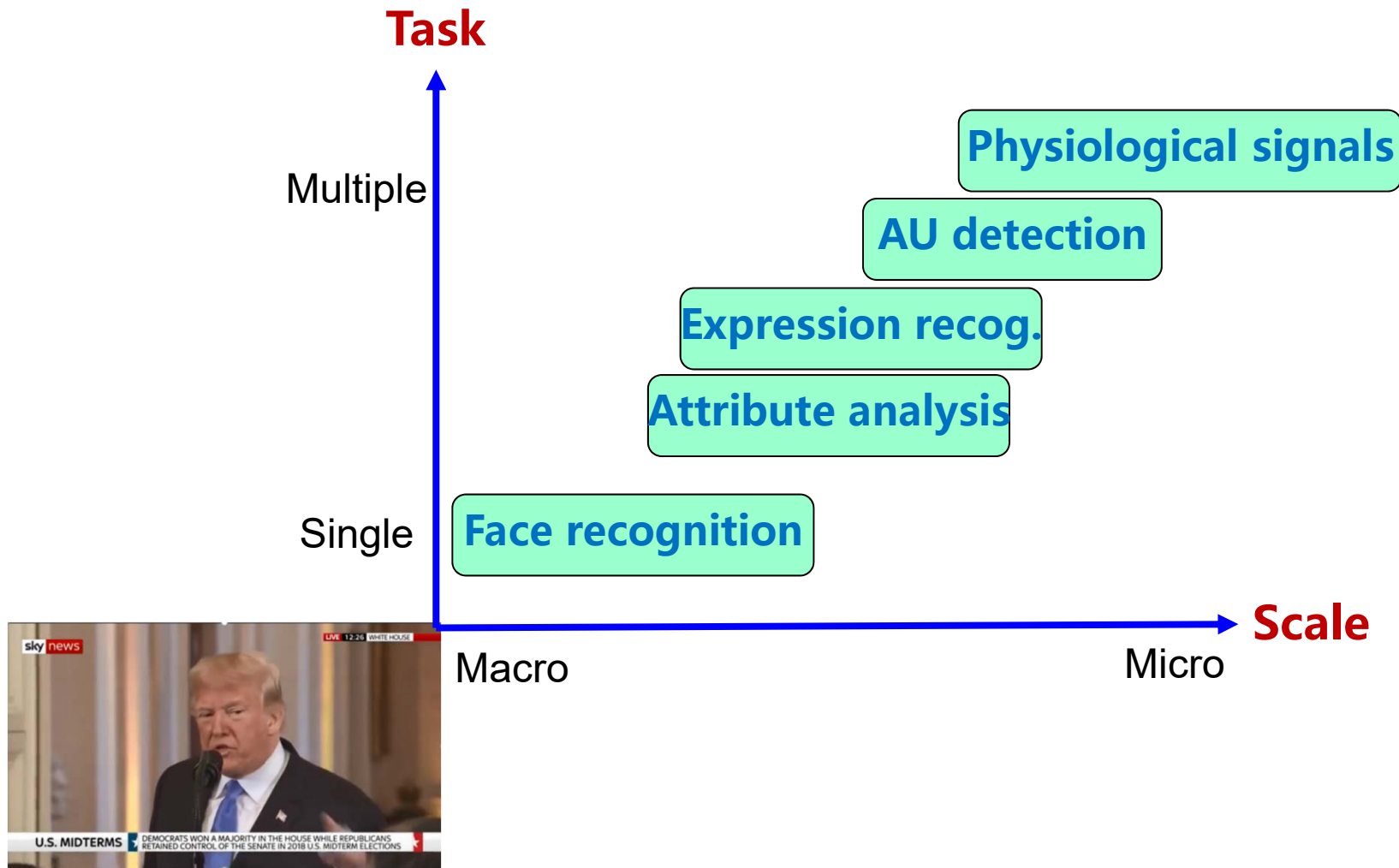
体温、疼痛...



Trend of face analysis

中科院计算所

Institute of Computing Technology, Chinese Academy of Sciences



Deep face analysis from macro to micro perspectives!



Why is micro perspective analysis important?

中科院计算所

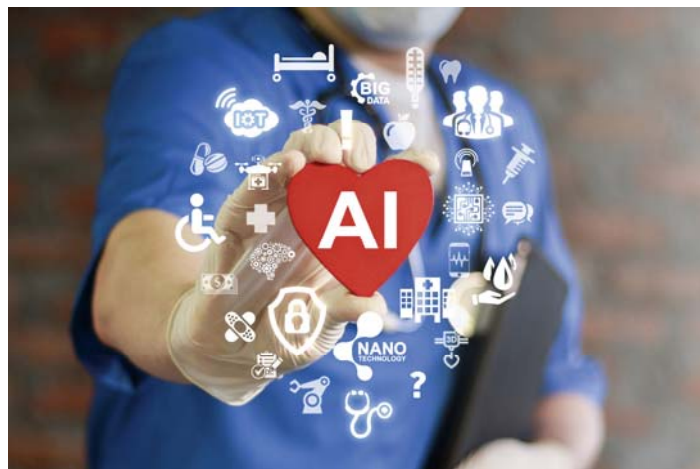
Institute of Computing Technology, Chinese Academy of Sciences



NHCI



HMC

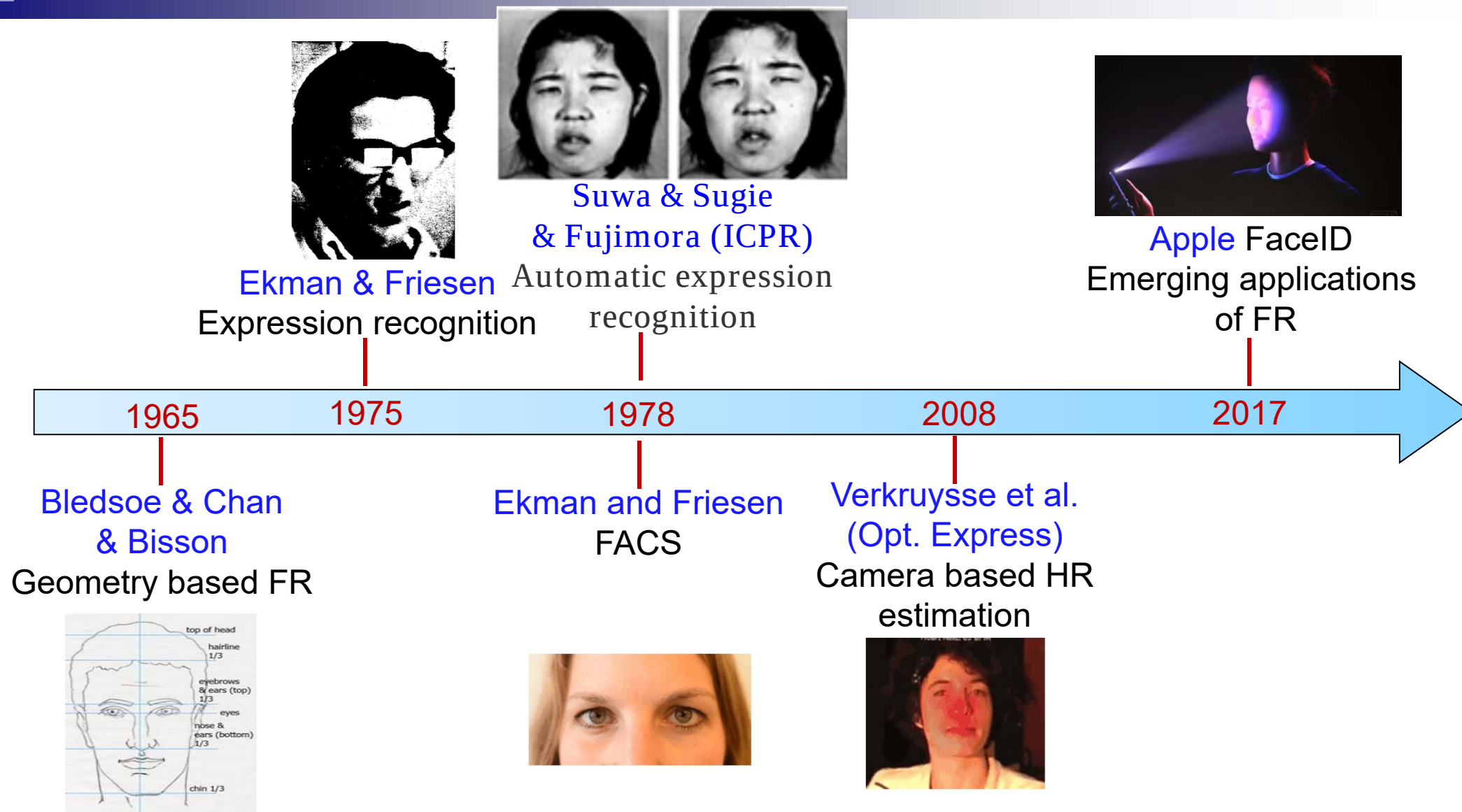


Intelligent healthcare



ADAS

Milestones of face analysis in micro perspective

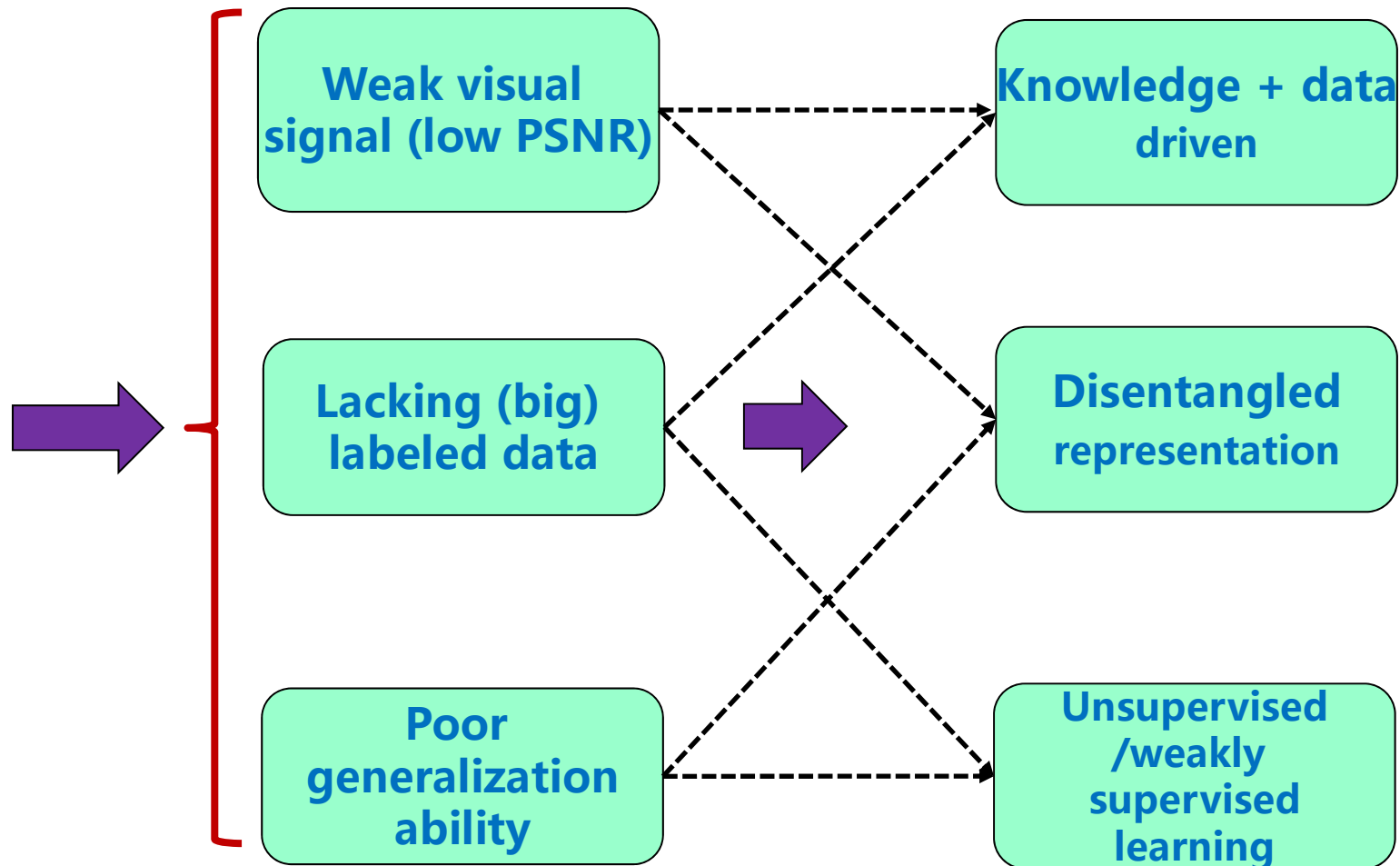




Micro perspective
face analysis

Challenges

Approaches

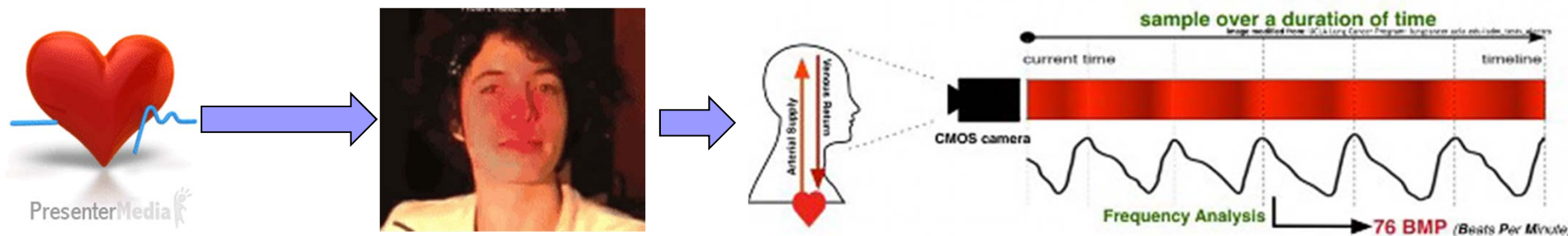




Outline

- ❖ What is the focus of micro-perspective face analysis
- ❖ Knowledge and data driven modeling [TIP2020, FG2019 Best Poster]
 - Weak visual signal, insufficient labeled data
- ❖ Disentangled representation learning [CVPR2019, CVPR2020, ECCV2020]
 - Weak visual signal, poor generalization ability
- ❖ Summary

- ❖ What kind of knowledge do we have in rPPG HR estimation?
 - A review of the physical imaging process w.r.t. rPPG



Periodic heartbeat $\rightarrow$ periodic blood volume changes $\rightarrow$ periodic light absorption by blood $\rightarrow$ periodic color changes in video

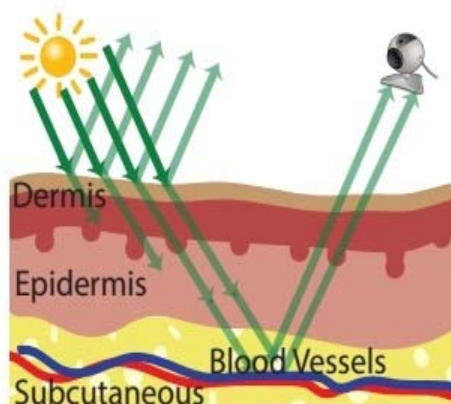
Knowledge/prior: There is indeed a periodic signal contained in the raw face video



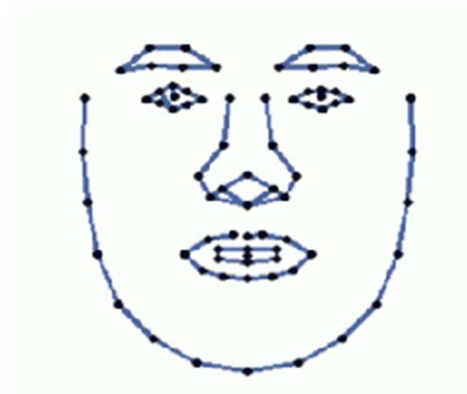
Well-known knowledge/prior in different tasks

中科院计算所

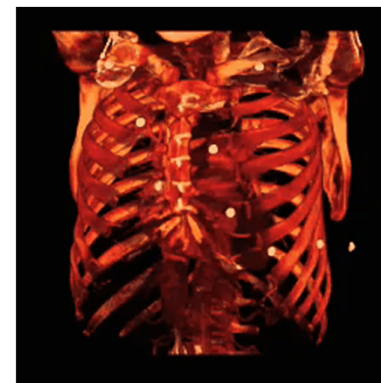
Institute of Computing Technology, Chinese Academy of Sciences



Imaging prior



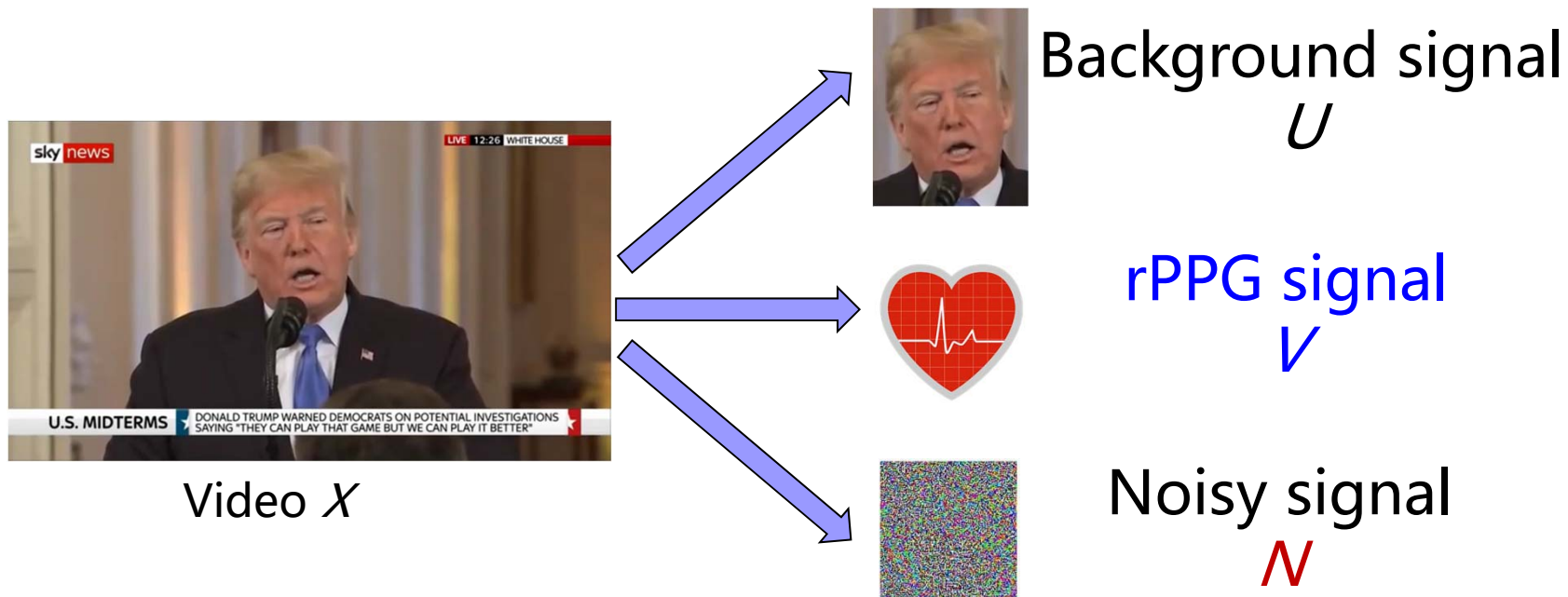
Shape prior



Structural prior

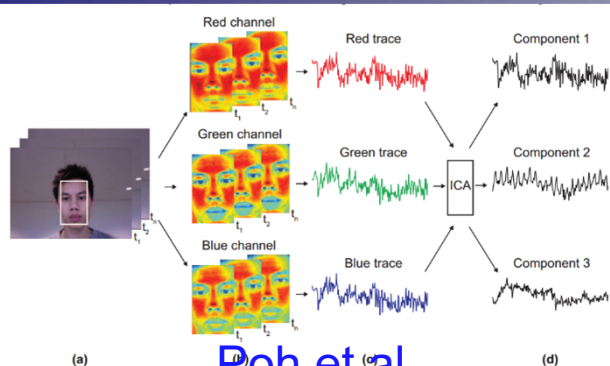


rPPG based HR estimation

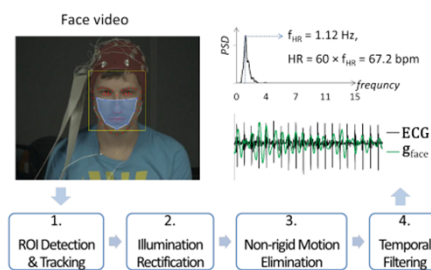


How to leverage prior knowledge to extract or enhance V ?

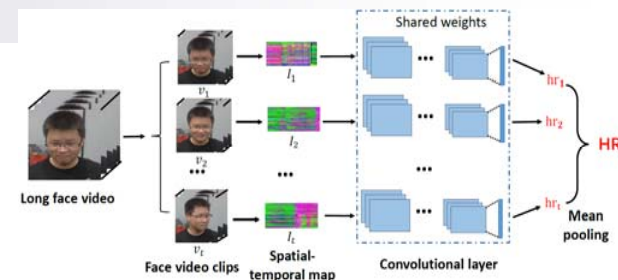
Milestone of rPPG based HR estimation



Poh et al.
Opt. Express 2010/2011
(ICA)



Li et al.
CVPR2014
(light rectification & motion estimation)

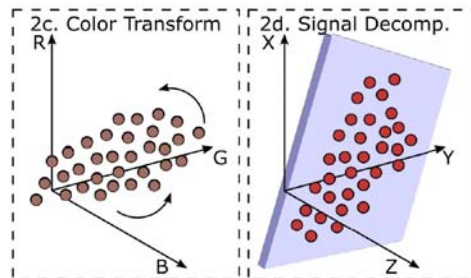


Niu et al.
ACCV2018/TIP19
(knowledge + data driven)

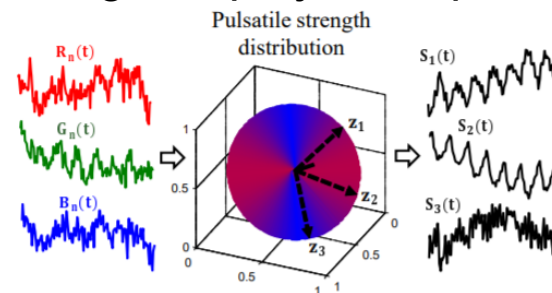


Verkruyse et al.
Opt. Express 2008
First work on camera
based HR est.
(spatial avg. & filtering)

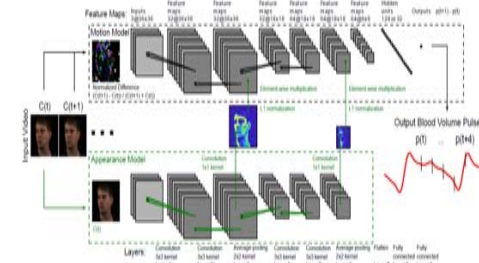
Haan et al.
TBME 2013
(chrominance)



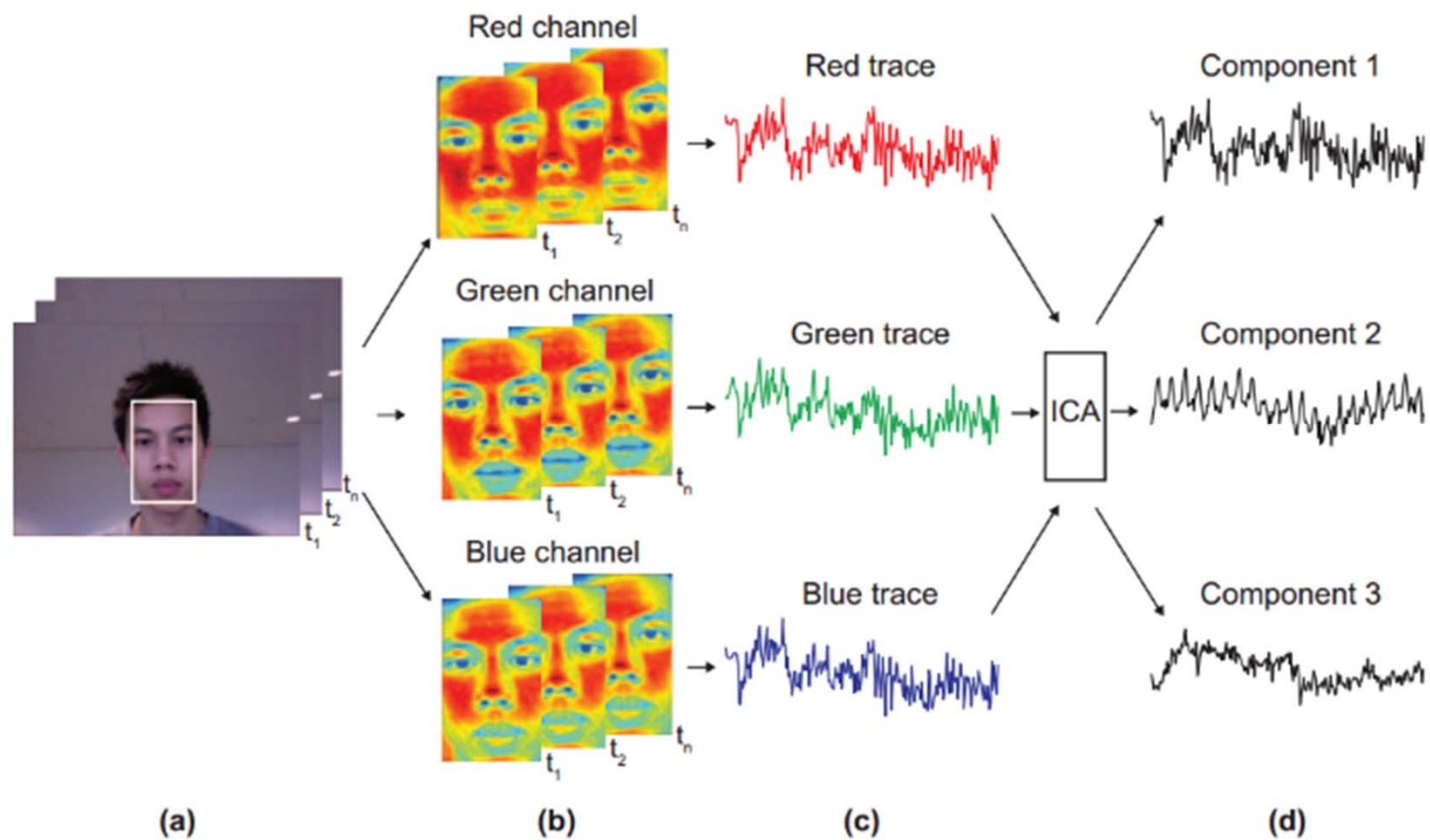
Wang et al.
TBME 2017
(orthogonal projection plane)



Chen et al.
ECCV 2018
(DeepPhys)



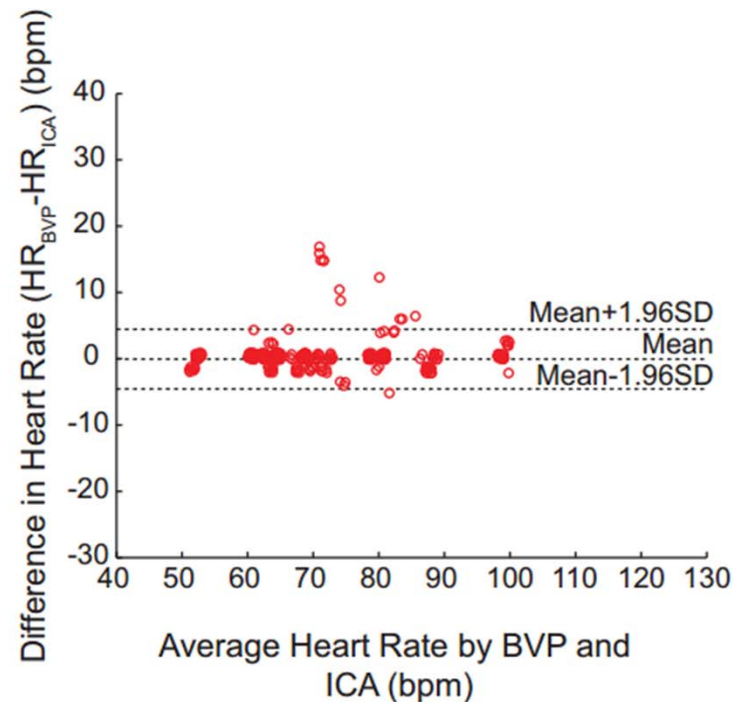
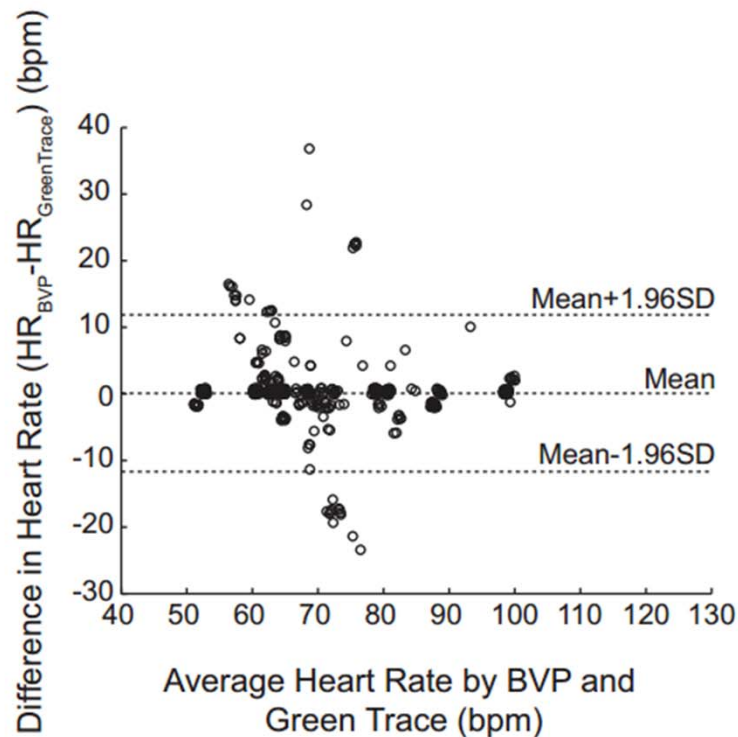
Review: ICA based HR estimation



M.-Z. Poh, D. J. McDuff, and R. W. Picard, "Non-contact, automated cardiac pulse measurements using video imaging and blind source separation." Opt. Express, vol. 18, no. 10, pp. 10 762–10 774, 2010.

M.-Z. Poh, D. J. McDuff, and R. W. Picard, "Advancements in noncontact, multiparameter physiological measurements using a webcam," IEEE Trans. Biomed. Eng., vol. 58, no. 1, pp. 7–11, 2011.

- ❖ Better performance than using green channel

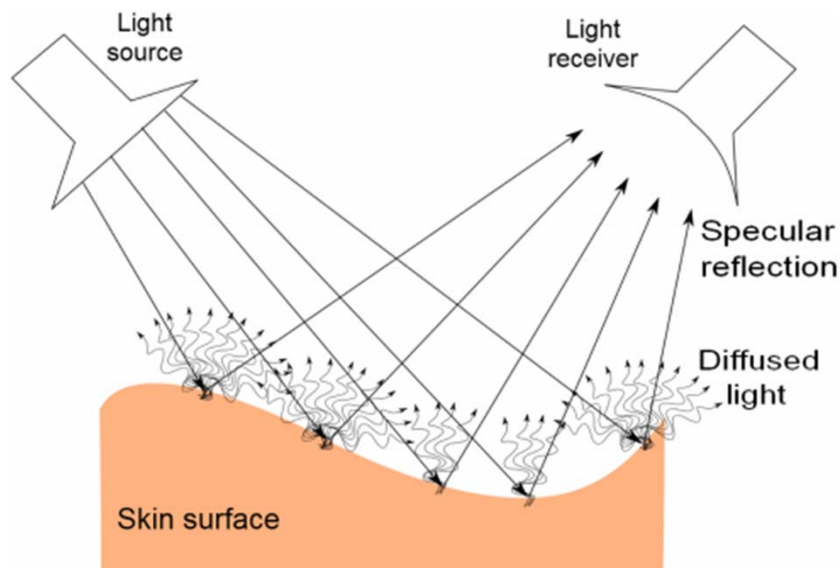


M.-Z. Poh, D. J. McDuff, and R. W. Picard, "Non-contact, automated cardiac pulse measurements using video imaging and blind source separation." Opt. Express, vol. 18, no. 10, pp. 10 762–10 774, 2010.

M.-Z. Poh, D. J. McDuff, and R. W. Picard, "Advancements in noncontact, multiparameter physiological measurements using a webcam," IEEE Trans. Biomed. Eng., vol. 58, no. 1, pp. 7–11, 2011.

Review: Chrominance-based method

- ❖ Assume the influence of motion to different channels is equal

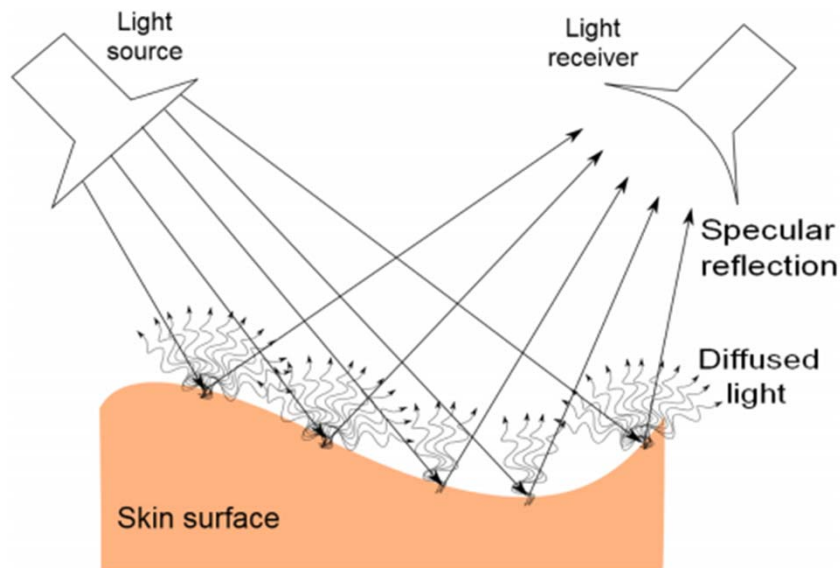


$$C_i = I_{Ci}(\rho_{Cdc} + \rho_{Ci} + s_i)$$

$$C \in \{R, G, B\}$$

Review: Chrominance-based method

- ❖ Assume the influence of motion to different channels is equal



$$C_i = I_{Ci}(\rho_{Cdc} + \rho_{Ci} + s_i)$$

$$C \in \{R, G, B\}$$

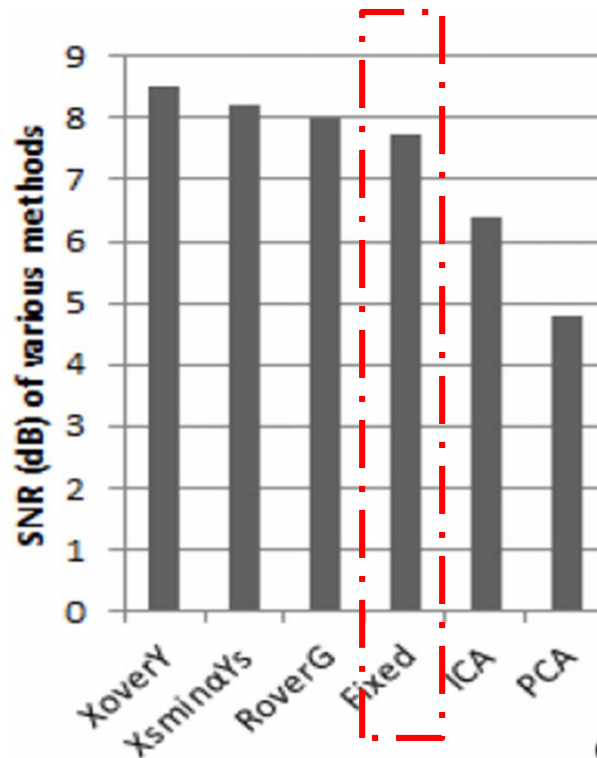


$$\begin{aligned} X &= R - G \\ Y &= 0.5R + 0.5G - B \end{aligned} \quad \left. \vphantom{\begin{aligned} X &= R - G \\ Y &= 0.5R + 0.5G - B \end{aligned}} \right\} \text{Two orthogonal signals}$$

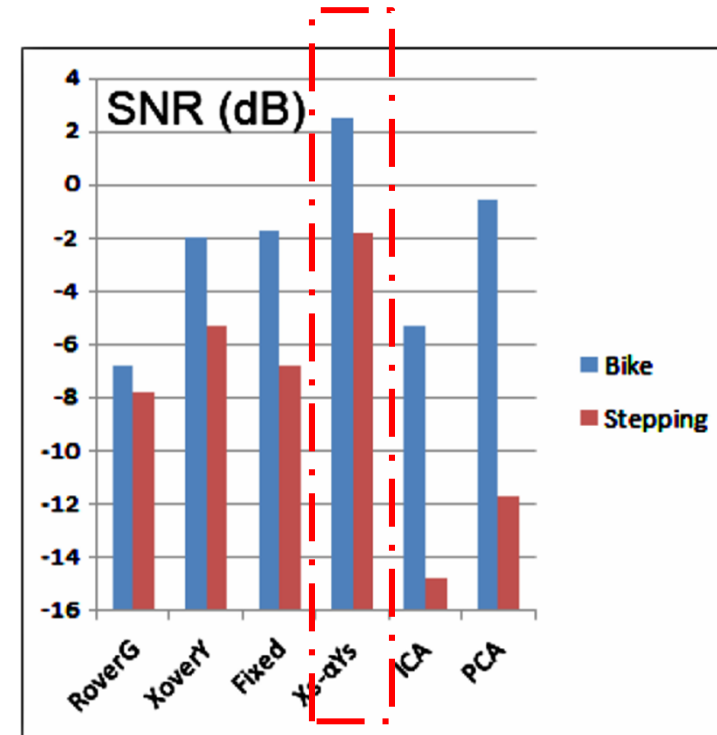


$$S = \frac{X_n}{Y_n} - 1.$$

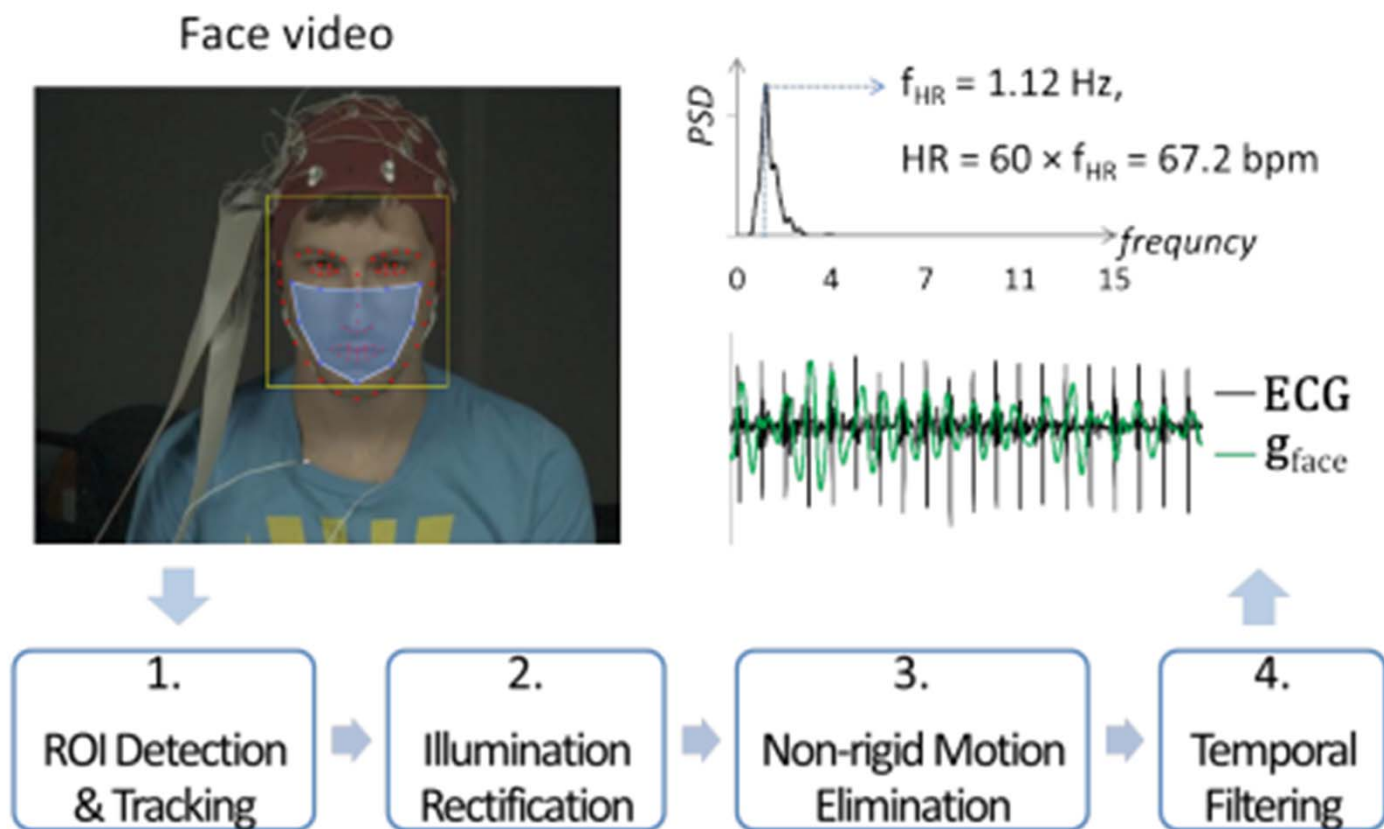
❖ Result



SNR of stationary scenario

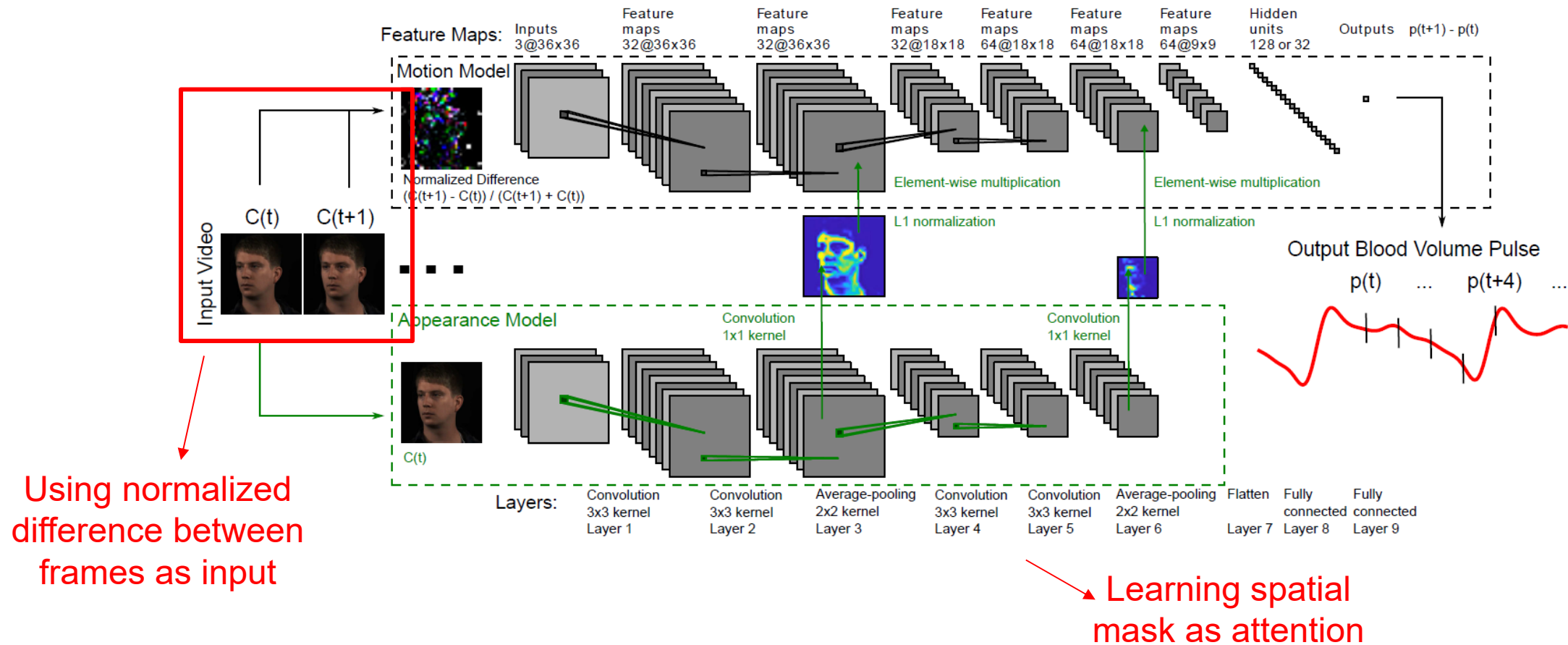


SNR of exercise scenario



❖ Best results on MAHNOB-HCI in 2014

Method	$M_e(SD_e)$ (bpm)	$RMSE$ (bpm)	M_{eRate}	r
Poh2010	-8.95(24.3)	25.9	25.0%	0.08
Kwon2012	-7.96(23.8)	25.1	23.6%	0.09
Poh2011	2.04(13.5)	13.6	13.2%	0.36*
Balakrishnan2013	-14.4(15.2)	21.0	20.7%	0.11
Ours				
Step 1+4	-3.53(8.62)	9.31	8.03%	0.69*
Step 1+2+4	-3.46(7.36)	8.13	7.02%	0.79*
All steps	-3.30(6.88)	7.62	6.87%	0.81*





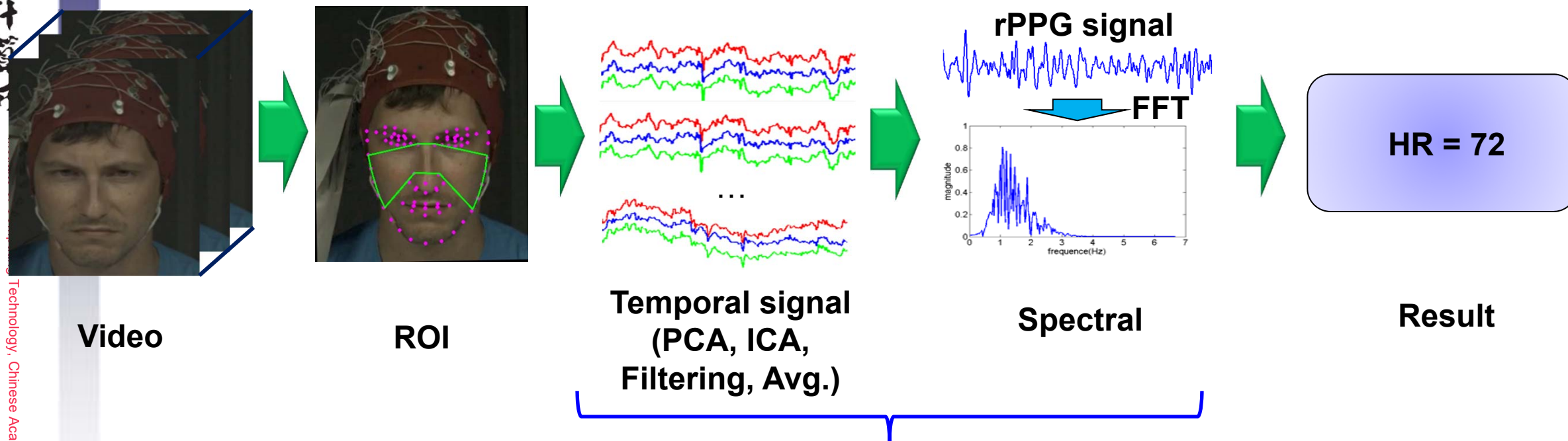
Review: CNN+Attention

HEART RATE															
Methods	Mean Absolute Error / BPM								Signal-To-Noise Ratio / dB						
	1	2	3	4	5	6	Avg.		1	2	3	4	5	6	Avg.
Estepp et al. [10]	3.48	3.95	3.80	6.55	11.8	13.4	7.16		6.06	4.82	3.77	-0.10	-4.72	-9.63	0.03
McDuff et al. [19]	1.17	1.70	1.70	4.00	5.22	11.8	4.29		10.9	9.55	6.69	3.08	0.08	-6.93	3.90
Balakrishnan et al. [3]	4.99	5.16	12.7	17.4	18.7	14.2	12.2		-1.08	-0.34	-8.83	-12.6	-14.2	-12.1	-8.19
De Haan et al. [13]	4.53	4.59	4.35	4.84	6.89	10.3	5.92		1.72	1.38	3.97	3.63	2.02	-2.47	1.71
Wang et al. [38]	1.50	1.53	1.50	1.84	2.05	6.11	2.42		6.84	6.21	4.80	2.97	0.77	-4.33	2.88
Tulyakov et al. [34]	1.76	2.14	14.9	19.0	15.7	22.0	12.6		4.32	2.29	-11.8	-14.3	-12.3	-15.3	-7.85
Ours: Part. Dep.															
Motion-only CNN	1.17	1.29	1.29	1.66	2.04	2.95	1.73		10.2	9.28	7.02	4.18	1.95	-1.00	5.28
Stacked CNN	1.18	1.30	1.26	1.51	1.82	2.62	1.61		10.7	9.60	8.28	5.59	2.84	0.29	6.22
CAN	1.17	1.26	1.16	1.61	2.04	1.78	1.50		10.9	9.66	8.20	5.69	3.57	1.33	6.55
Ours: Part. Ind.															
Motion-only CNN	1.17	1.41	1.33	1.91	3.57	6.41	2.63		9.60	8.62	6.25	2.73	-0.15	-4.02	3.84
Stacked CNN	1.12	1.66	1.30	1.65	2.33	6.25	2.38		9.26	8.18	7.27	3.81	1.14	-2.84	4.47
CAN	1.16	1.45	1.23	1.72	2.42	5.59	2.26		9.52	8.82	7.36	3.90	1.10	-2.72	4.66
CAN (task 1)	1.16	3.01	3.43	4.85	7.92	13.9	5.70		9.52	5.74	2.01	-2.13	-5.77	-9.70	-0.06
CAN (task 2)	1.12	1.45	1.67	3.35	7.51	12.9	4.66		10.1	8.82	4.25	-0.90	-5.33	-8.96	1.33
CAN (task 3)	1.18	1.41	1.23	1.78	2.73	9.23	2.93		9.75	8.79	7.36	4.33	0.55	-5.88	4.15
CAN (task 4)	1.13	1.57	1.24	1.72	2.79	8.98	2.91		9.95	8.72	7.06	3.90	0.37	-5.68	4.05
CAN (task 5)	1.16	1.38	1.30	1.54	2.42	7.05	2.48		9.86	8.98	7.68	4.40	1.10	-4.15	4.65
CAN (task 6)	1.14	1.32	1.22	1.47	2.17	5.59	2.15		10.4	9.62	8.21	5.15	1.83	-2.72	5.41
CAN (all tasks)	1.13	1.34	2.45	1.64	1.83	6.32	2.45		9.88	7.21	1.46	8.48	4.11	-3.23	4.65
BREATH. RATE															
Methods	Mean Absolute Error / BPM								Signal-To-Noise Ratio / dB						
	1	2	3	4	5	6	Avg.		1	2	3	4	5	6	Avg.
Tarassenko et al. [31]	2.51	2.53	3.19	4.85	4.22	4.78	3.68		-1.29	-1.82	-6.32	-8.55	-8.79	-10.6	-6.22
Ours: Part. Dep.															
Motion-only CNN	2.03	2.47	3.21	3.04	3.11	4.27	3.02		-0.33	-1.91	-5.28	-4.83	-5.33	-9.64	-4.55
Stacked CNN	1.74	2.27	2.98	2.79	3.03	5.33	3.02		1.84	-0.93	-6.31	-5.18	-5.70	-11.2	-4.58
CAN	1.70	2.19	3.24	3.05	3.06	3.96	2.86		2.73	-0.02	-4.39	-4.47	-4.36	-7.97	-3.08
Ours: Part. Ind.															
Motion-only CNN	1.70	2.31	4.09	4.85	4.60	4.06	3.60		0.75	-0.17	-6.03	-9.19	-9.05	-9.06	-5.46
Stacked CNN	2.00	2.10	5.67	5.55	6.34	5.76	4.57		0.19	-0.25	-12.0	-11.5	-12.7	-13.0	-8.20
CAN	1.28	1.64	4.15	4.37	3.77	4.37	3.26		4.45	2.96	-5.05	-6.72	-6.70	-8.93	-3.33
CAN (task 1)	1.28	1.72	6.34	7.28	6.28	4.01	4.48		4.45	2.37	-10.0	-13.5	-13.9	-9.02	-6.60
CAN (task 2)	1.21	1.64	5.73	5.65	4.92	3.82	3.83		4.39	2.96	-7.97	-12.3	-11.7	-8.54	-5.52
CAN (task 3)	1.62	1.71	4.15	4.57	4.16	3.56	3.30		3.38	2.74	-5.05	-7.05	-6.71	-6.23	-3.15
CAN (task 4)	1.74	1.85	3.80	4.37	4.60	3.42	3.30		2.69	2.56	-5.10	-6.72	-7.86	-6.63	-3.51
CAN (task 5)	1.65	1.74	4.37	4.45	3.77	3.37	3.22		2.90	3.18	-6.20	-7.02	-6.70	-5.69	-3.25
CAN (task 6)	2.06	1.76	5.89	5.92	5.21	4.37	4.20		1.11	1.54	-9.72	-11.9	-10.1	-8.93	-6.33
CAN (all tasks)	1.54	4.69	3.71	2.14	5.27	3.33	3.45		3.09	-6.45	-7.09	0.21	-10.7	-7.36	-4.71

Internal dataset

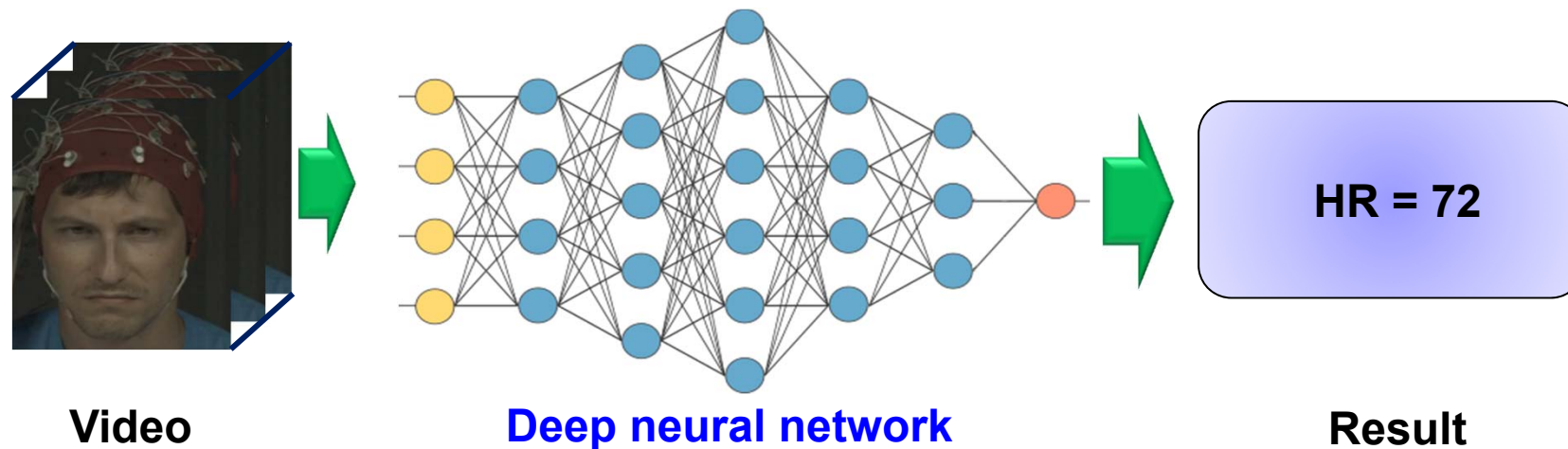
Motivation of our approach

❖ Review of traditional rPPG pipeline



Can we learn a more informative representation instead of using the hand-crafted features?

Motivation of our approach



Datasets in this area are usually very small; may easily cause overfitting

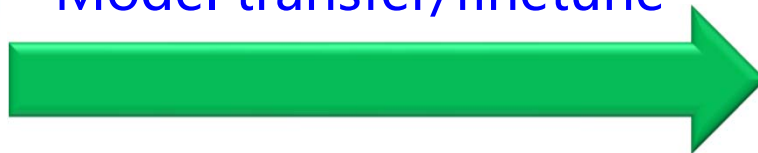
Motivation of our approach

- ❖ General to specific fintuning may not work under such big domain gap



ImageNet
Pre-train

Model transfer/fintune

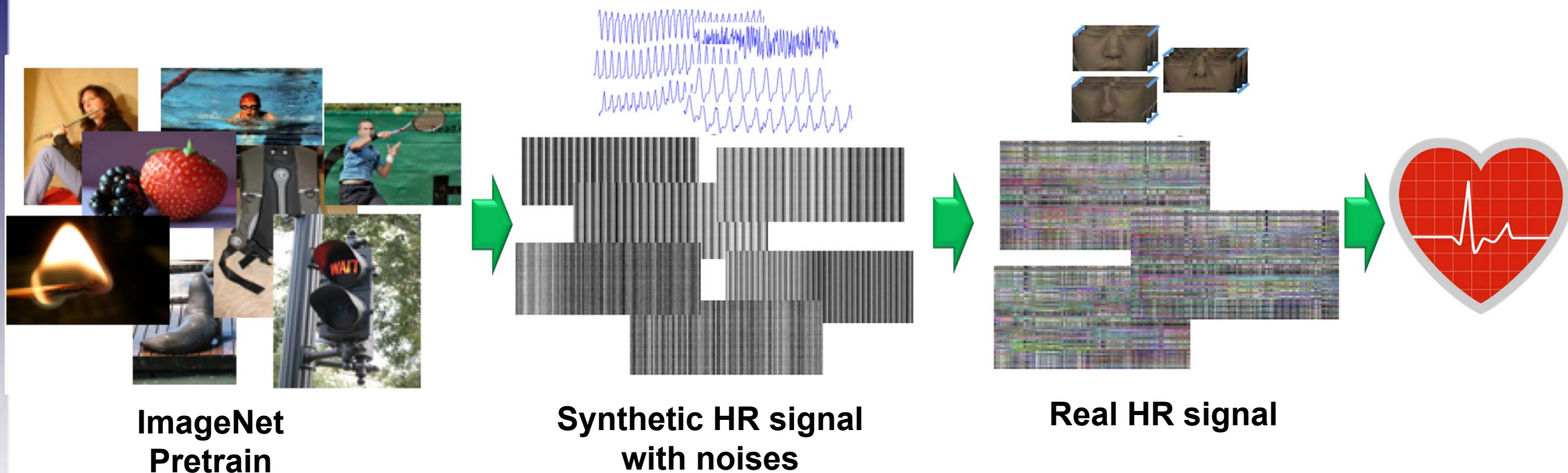


rPPG based HR est.



Our approach: knowledge and data driven

- ❖ Reduce the domain gap using synthetic HR signal



RythmNet v1.0 (2017-2018)

Xuesong Niu, et al., SynRhythm: Learning a Deep Heart Rate Estimator from General to Specific. [ICPR, 2018](#).



Our approach: knowledge and data driven

❖ Result: MAHNOB-HCI (27sub., 20vid/sub.)

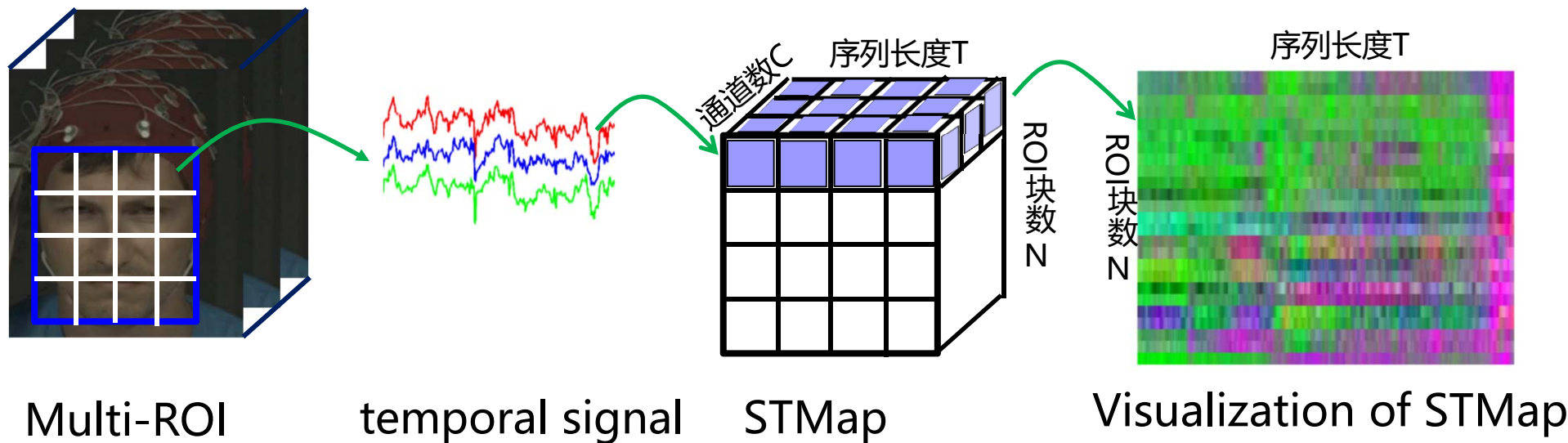
Method	HR_{me} (bpm)	HR_{sd} (bpm)	HR_{rmse} (bpm)	HR_{mer}
Poh2010 [1]	-8.95	24.3	25.9	25.0%
Poh2011 [3]	2.04	13.5	13.6	13.2%
Balakrishnan2013 [2]	-14.4	15.2	21.0	20.7%
Li2014 [5]	-3.30	6.88	7.62	6.87%
Haan2013 [4]	-2.89	13.67	10.7	12.9%
Niu2017 [7]	-0.38	10.81	8.72	11.5%
Tulyakov2016 [6]	3.19	5.81	6.23	5.93%
Hus2014 [11]	-0.20	11.32	11.31	12.8%
Proposed method	0.30	4.48	4.49	4.37%

RythmNet v1.0 (2017-2018)

Xuesong Niu, et al., SynRhythm: Learning a Deep Heart Rate Estimator from General to Specific. [ICPR](#), 2018.

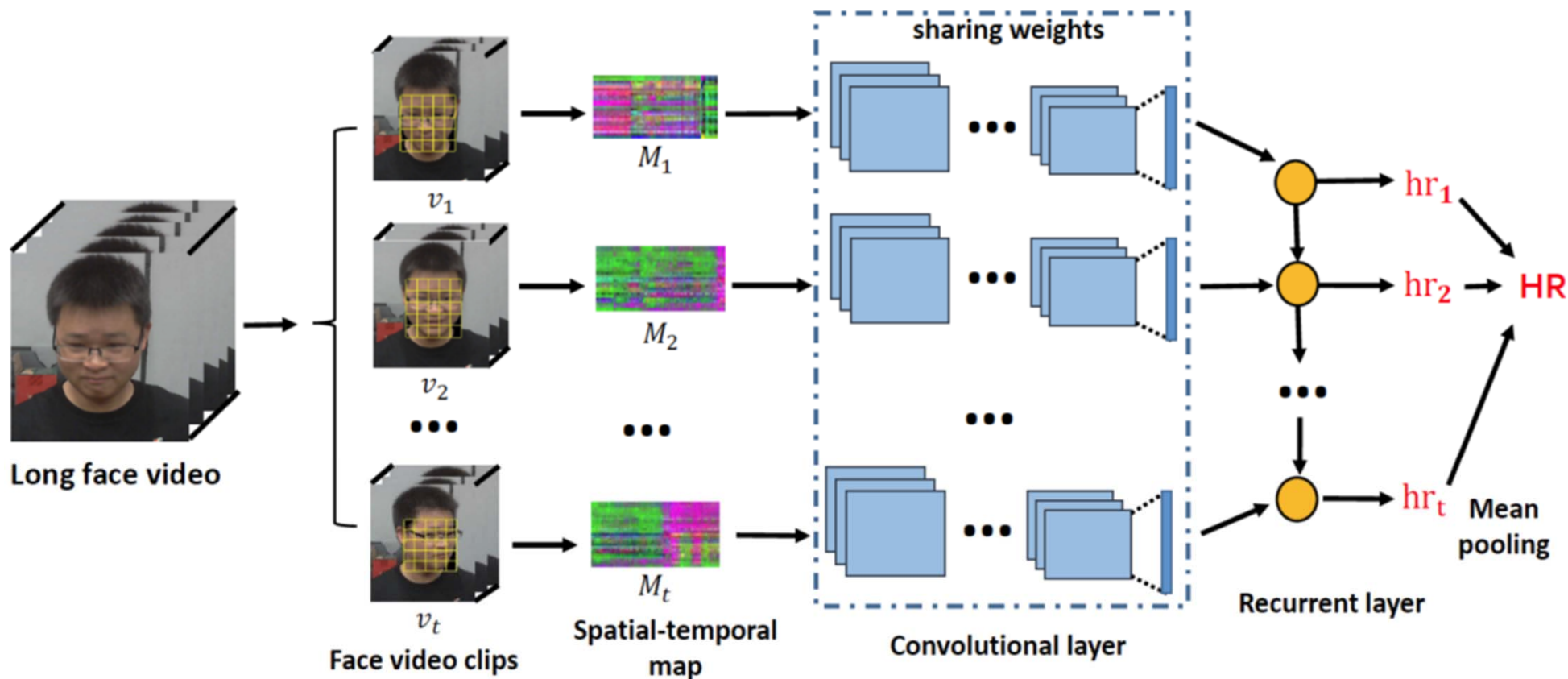
Our approach: knowledge and data driven

- ❖ **Spatial-temporal map based on imaging prior :**
Suppressing the background and noisy signals while enhancing the rPPG signal



Our approach: knowledge and data driven

- ❖ Learn representation with CNN, model temporal relationship with RNN



Xuesong Niu, Shiguang Shan, Hu Han, and Xilin Chen. RhythmNet: End-to-end Heart Rate Estimation from Face via Spatial-temporal Representation. *IEEE Trans. Image Process.*, 2020.



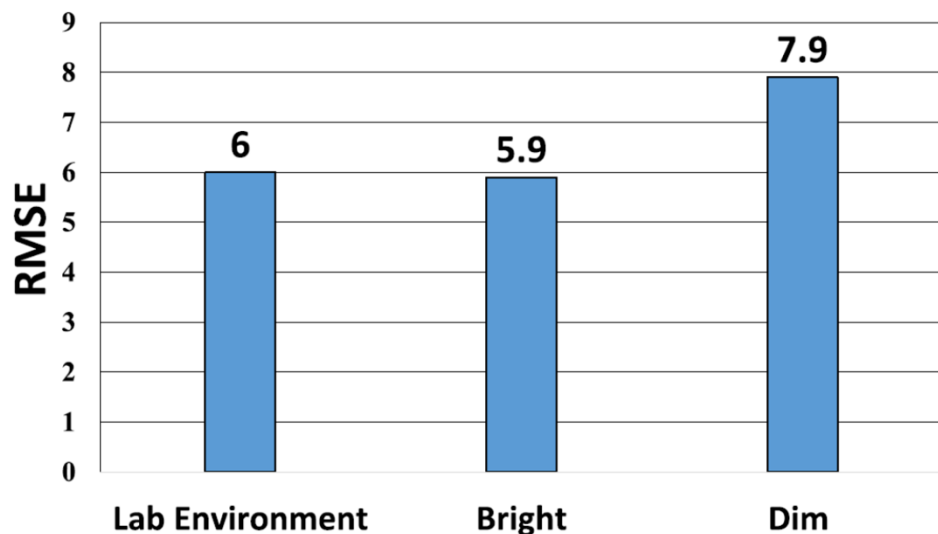
Our approach: knowledge and data driven

❖ Result on VIPL-HR (unconstrained)

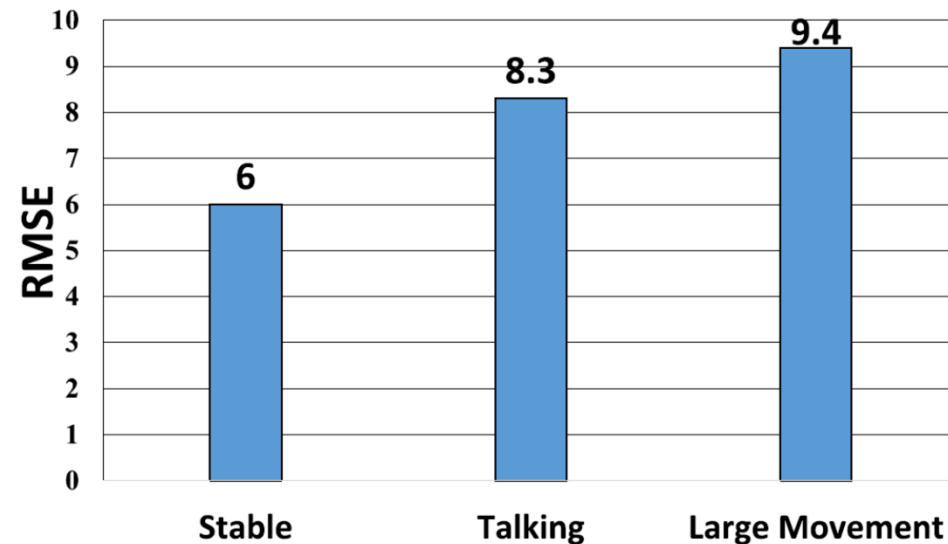
Method	<i>Mean</i> (bpm)	<i>Std</i> (bpm)	<i>MAE</i> (bpm)	<i>RMSE</i> (bpm)	<i>MER</i>	<i>r</i>
Tulyakov2016 [6]	10.8	18.0	15.9	21.0	26.7%	0.11
POS [7]	7.87	15.3	11.5	17.2	18.5%	0.30
Haan2013 [4]	7.63	15.1	11.4	16.9	17.8%	0.28
I3D [37]	1.37	15.9	12.0	15.9	15.6%	0.07
DeepPhy [27]	-2.60	13.6	11.0	13.8	13.6%	0.11
RhythmNet w/o GRU	1.02	8.88	5.79	8.94	7.38%	0.73
RhythmNet	0.73	8.11	5.30	8.14	6.71%	0.76

Xuesong Niu, Shiguang Shan, Hu Han, and Xilin Chen. RhythmNet: End-to-end Heart Rate Estimation from Face via Spatial-temporal Representation. *IEEE Trans. Image Process.*, 2020.

❖ Results under different conditions



Error under different illuminations



Error under different movements

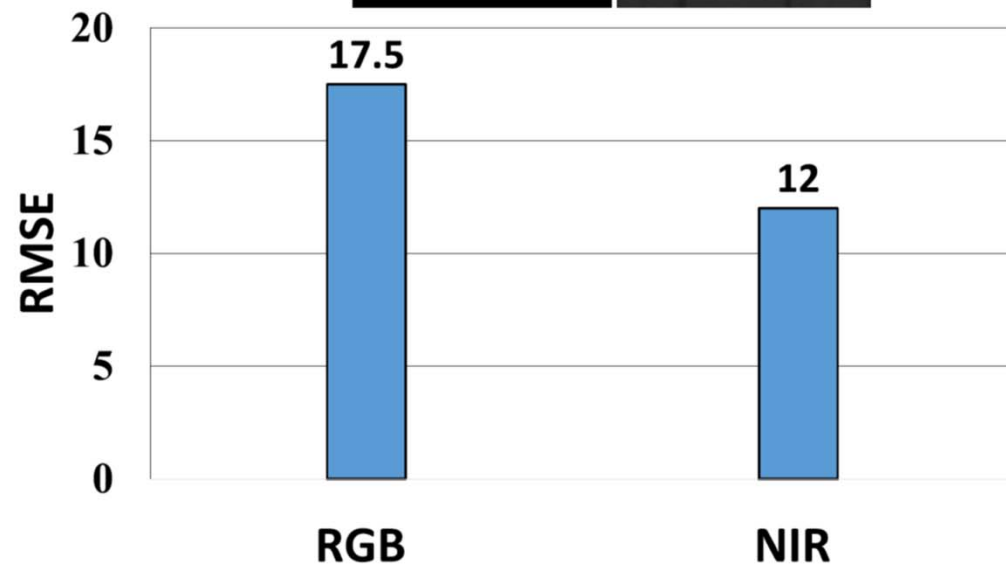


Our approach: knowledge and data driven

❖ Results under different conditions

Datatype	Mean (bpm)	Std (bpm)	MAE (bpm)	RMSE (bpm)	MER	r
Compressed	0.34	8.76	5.94	8.77	7.54%	0.58
Uncompressed	0.08	8.69	5.81	8.69	7.24%	0.59

Compressed video (x10, MJPG) vs. raw video



RGB vs. NIR under dark illumination

Xuesong Niu, Shiguang Shan, Hu Han, and Xilin Chen. RhythmNet: End-to-end Heart Rate Estimation from Face via Spatial-temporal Representation. *IEEE Trans. Image Process.*, 2020.



Our approach: knowledge and data driven

❖ Cross-database testing on MAHNOB-HCI

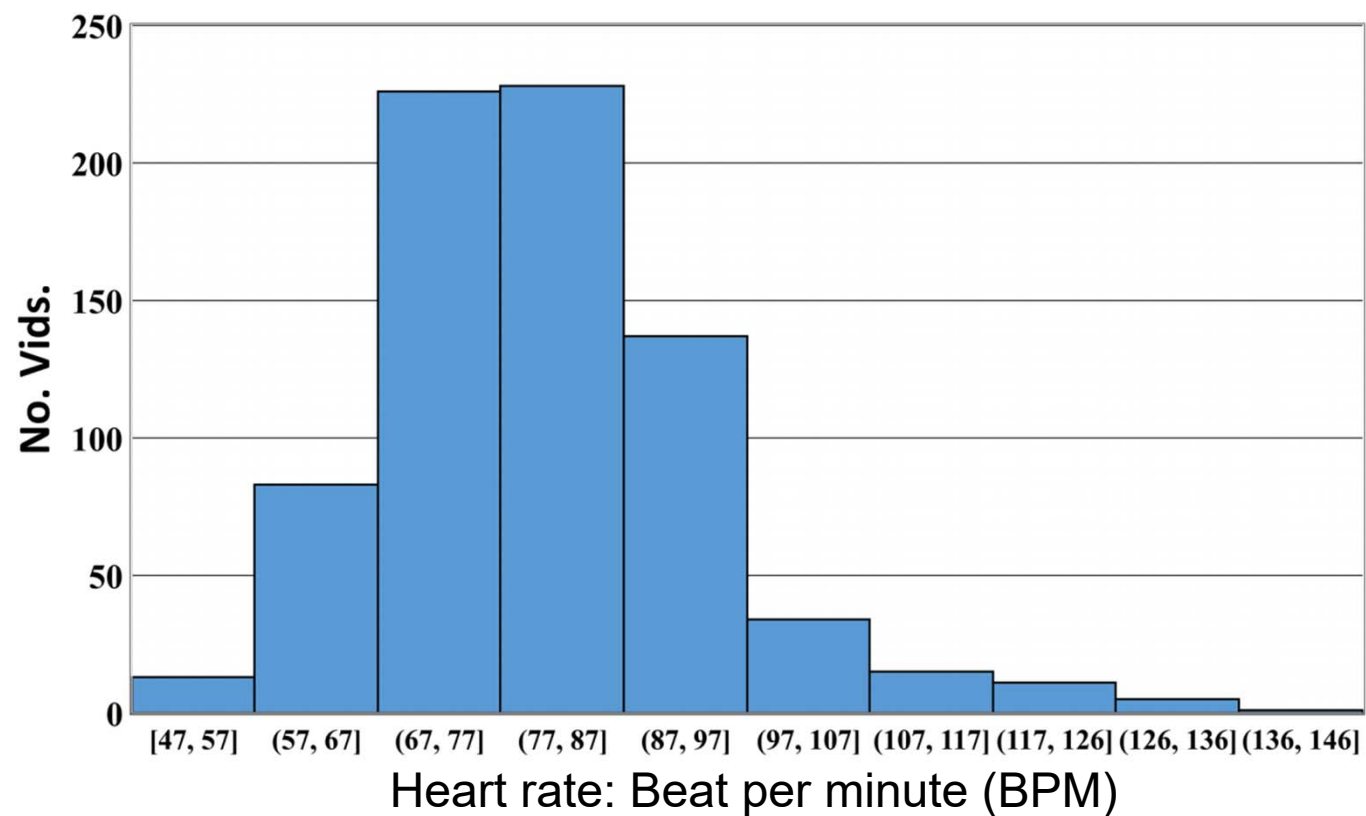
Method	<i>Mean</i> (bpm)	<i>Std</i> (bpm)	<i>RMSE</i> (bpm)	<i>MER</i>	<i>r</i>
Poh2010 [1]	-8.95	24.3	25.9	25.0%	0.08
Poh2011 [2]	2.04	13.5	13.6	13.2%	0.36
Balakrishnan2013 [3]	-14.4	15.2	21.0	20.7%	0.11
Li2014 [5]	-3.30	6.88	7.62	6.87%	0.81
Haan2013 [4]	-2.89	13.67	10.7	12.9%	0.82
Tulyakov2016 [6]	3.19	5.81	6.23	5.93%	0.83
RhythmNet(WithinDB)	0.41	3.98	4.00	4.18%	0.87
RhythmNet(CrossDB)	-5.66	6.06	8.28	8.00%	0.64
RhythmNet(Fine-tuned)	0.43	3.97	3.99	4.06%	0.87

Xuesong Niu, Shiguang Shan, Hu Han, and Xilin Chen. RhythmNet: End-to-end Heart Rate Estimation from Face via Spatial-temporal Representation. *IEEE Trans. Image Process.*, 2020.

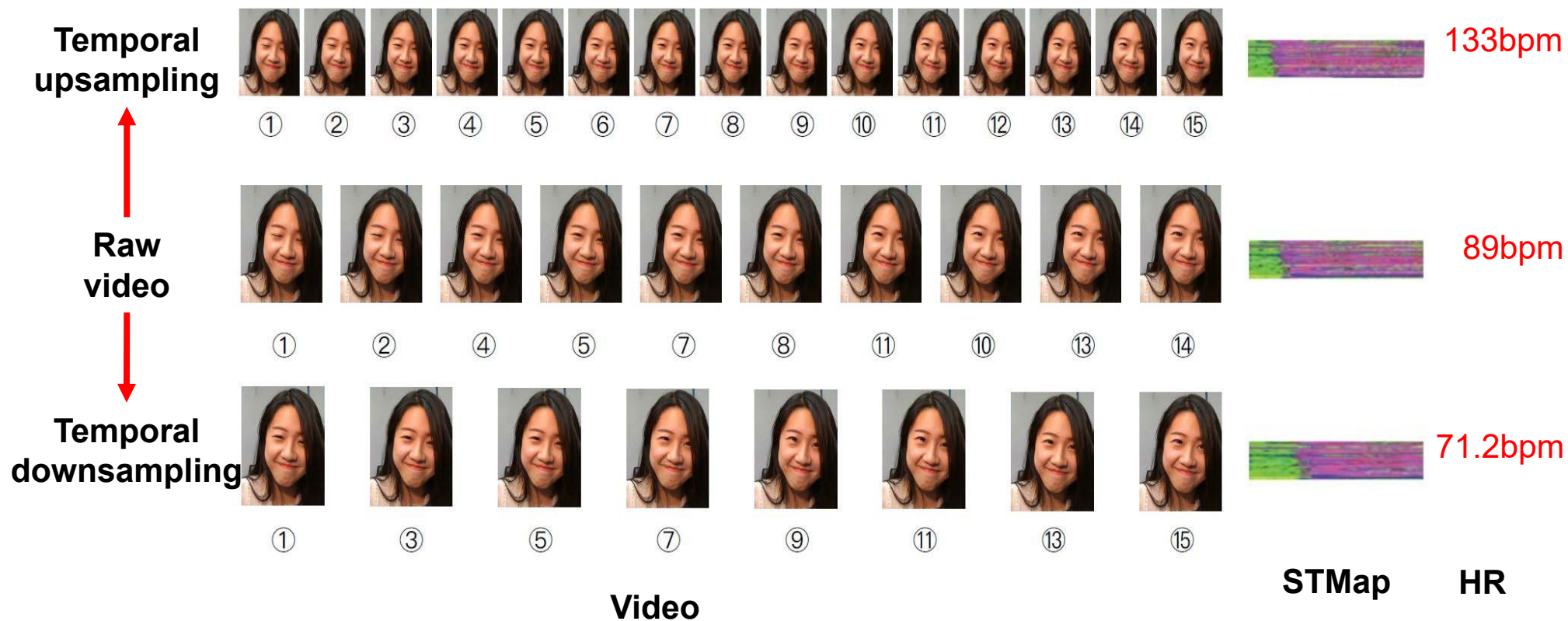


Imbalanced data

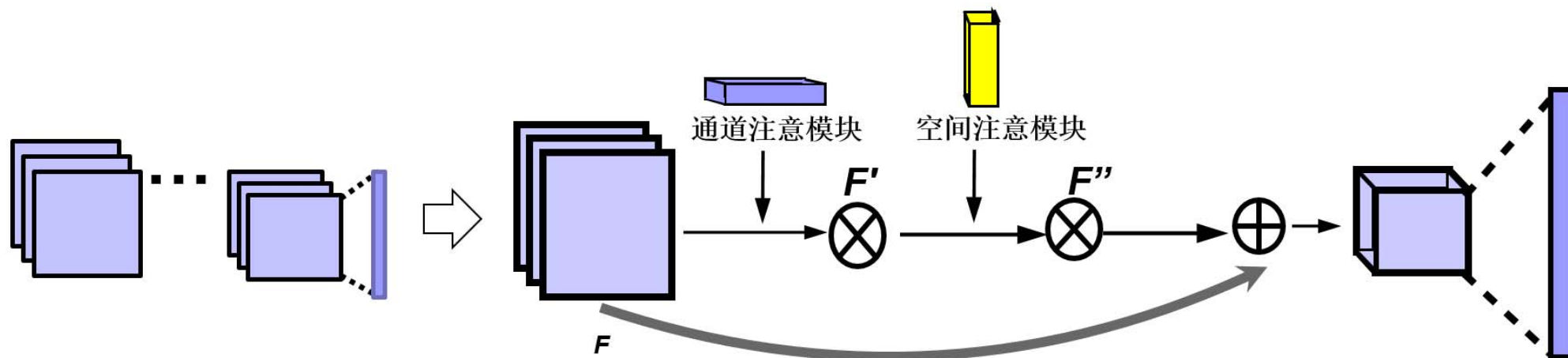
❖ The ground-truth HR distribution in VIPL-HR



❖ Temporal data augmentation



❖ Channel and spatial attention





❖ VIPL-HR

Method	$HR_{std}(bpm)$	$HR_{std}(bpm)$	$HR_{mae}(bpm)$	$HR_{rmse}(bpm)$	HR_{mer}	r
Haan13	7.63	15.1	11.4	16.9	17.8%	0.28
Tulyakov16	10.8	18.0	15.9	21.0	26.7%	0.11
Wang17	7.87	15.3	11.5	17.2	18.5%	0.30
RhythmNet v1	1.02	8.88	5.79	8.94	7.38%	0.73
RhythmNet v1 + DA	-0.08	8.14	5.58	8.14	6.91%	0.63
RhythmNet v1 + DA +Attention	-0.16	7.99	5.40	7.99	6.70%	0.66



Database collection

- ❖ VIPL-HR: Unconstrained, 100+ subjects, 3000 + videos, RGB-D modality, diverse illumination, movement, sensor, etc.

Movement



Lighting



Camera



(2018)

Download VIPL-HR



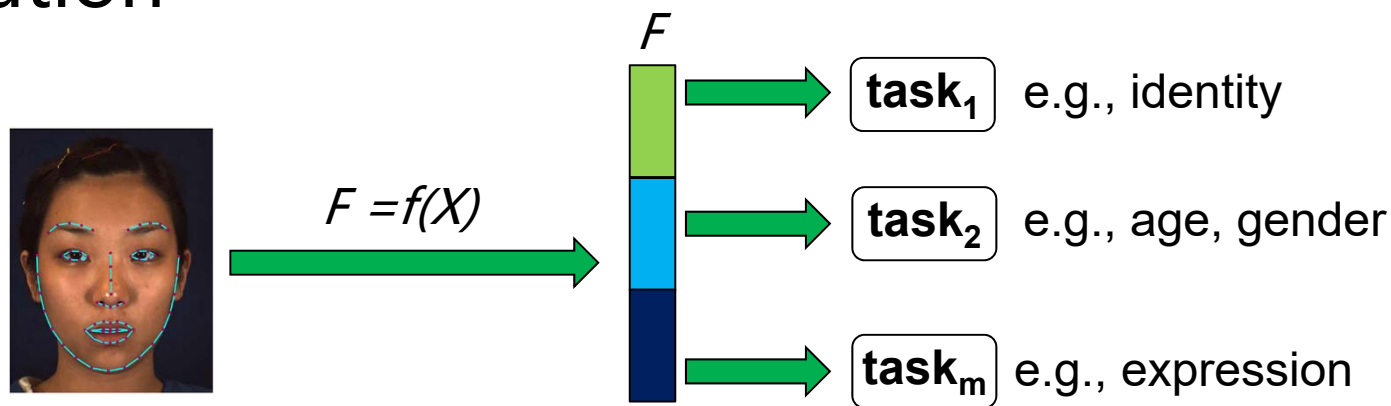
http://vipl.ict.ac.cn/view_database.php?id=15



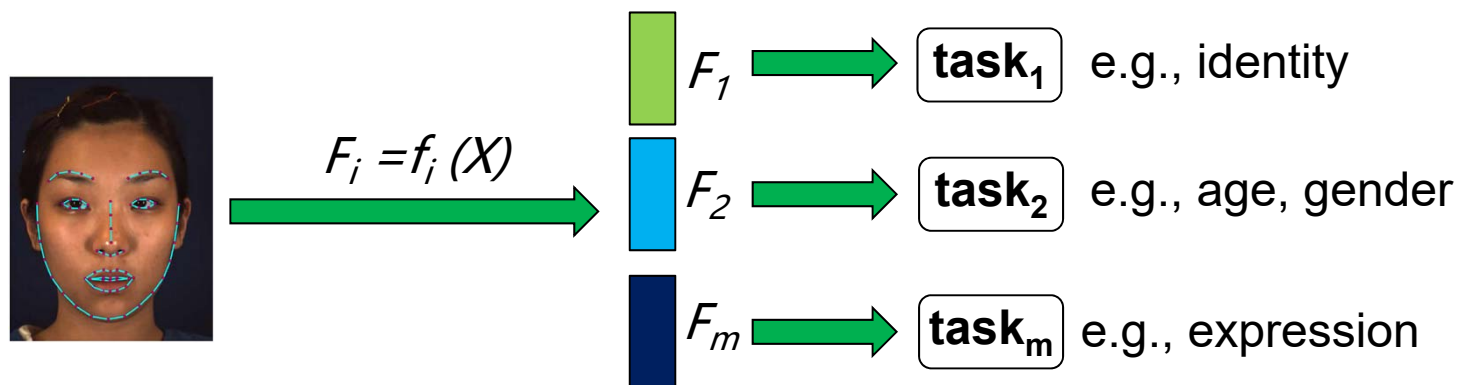
Outline

- ❖ What is the focus of micro-perspective face analysis
- ❖ Knowledge and data driven modeling [TIP2020, FG2019, ICPR2018]
 - Weak visual signal, insufficient labeled data
- ❖ Disentangled representation learning [CVPR2019, CVPR2020, ECCV2020]
 - Weak visual signal, poor generalization ability
- ❖ Summary

❖ Motivation



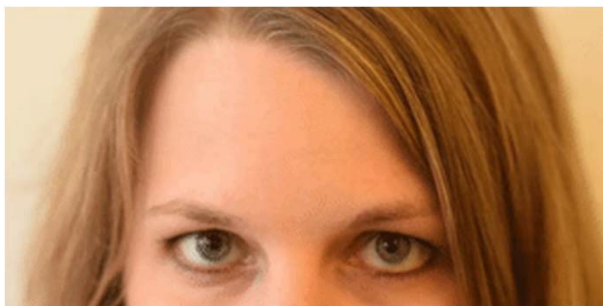
“大而全” vs “小而专”





Facial Action Unit (AU) recognition

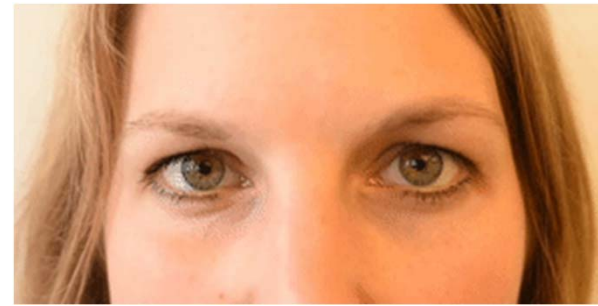
- ❖ AUs intensity can be minor and their intensities can be subject dependent



AU1



AU5

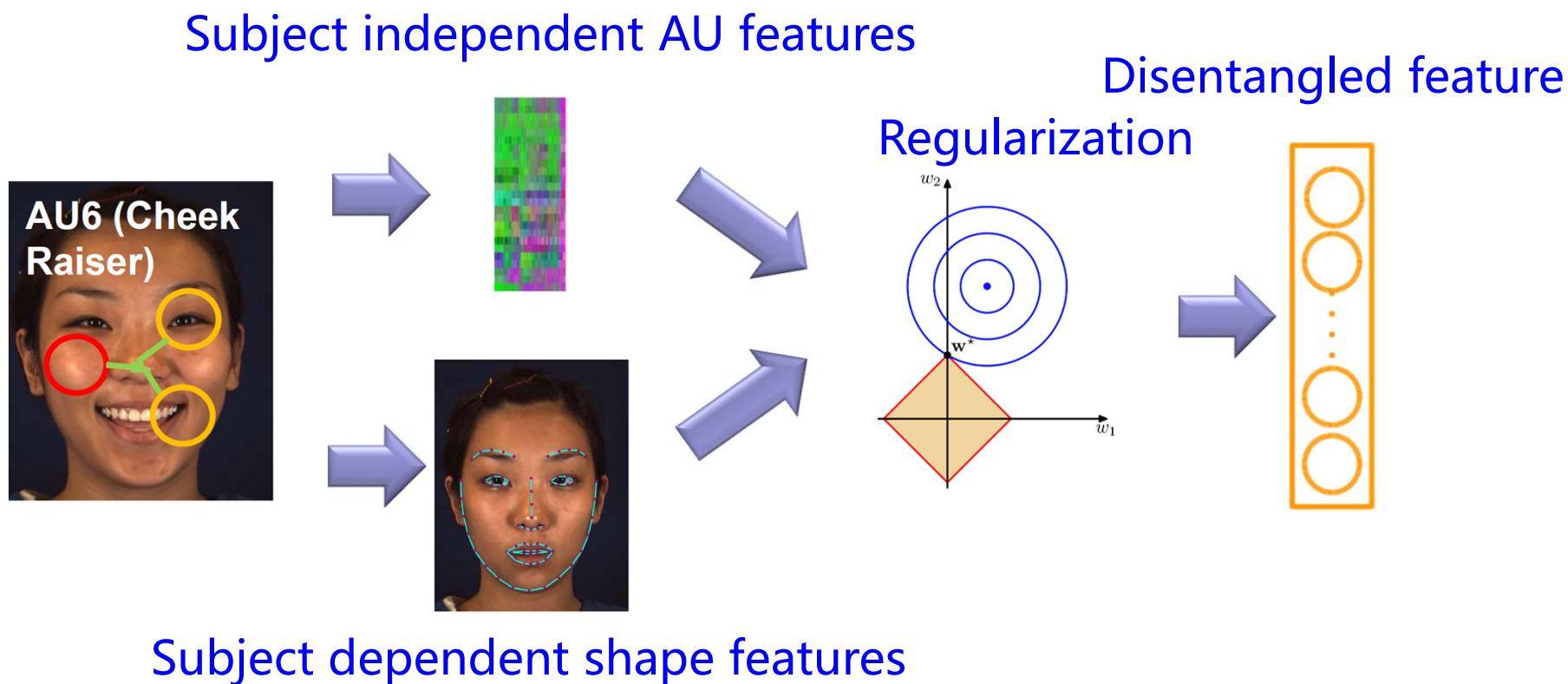


AU7

How to learn a subject-independent representation for AU?

Disentangled representation learning for AU recognition

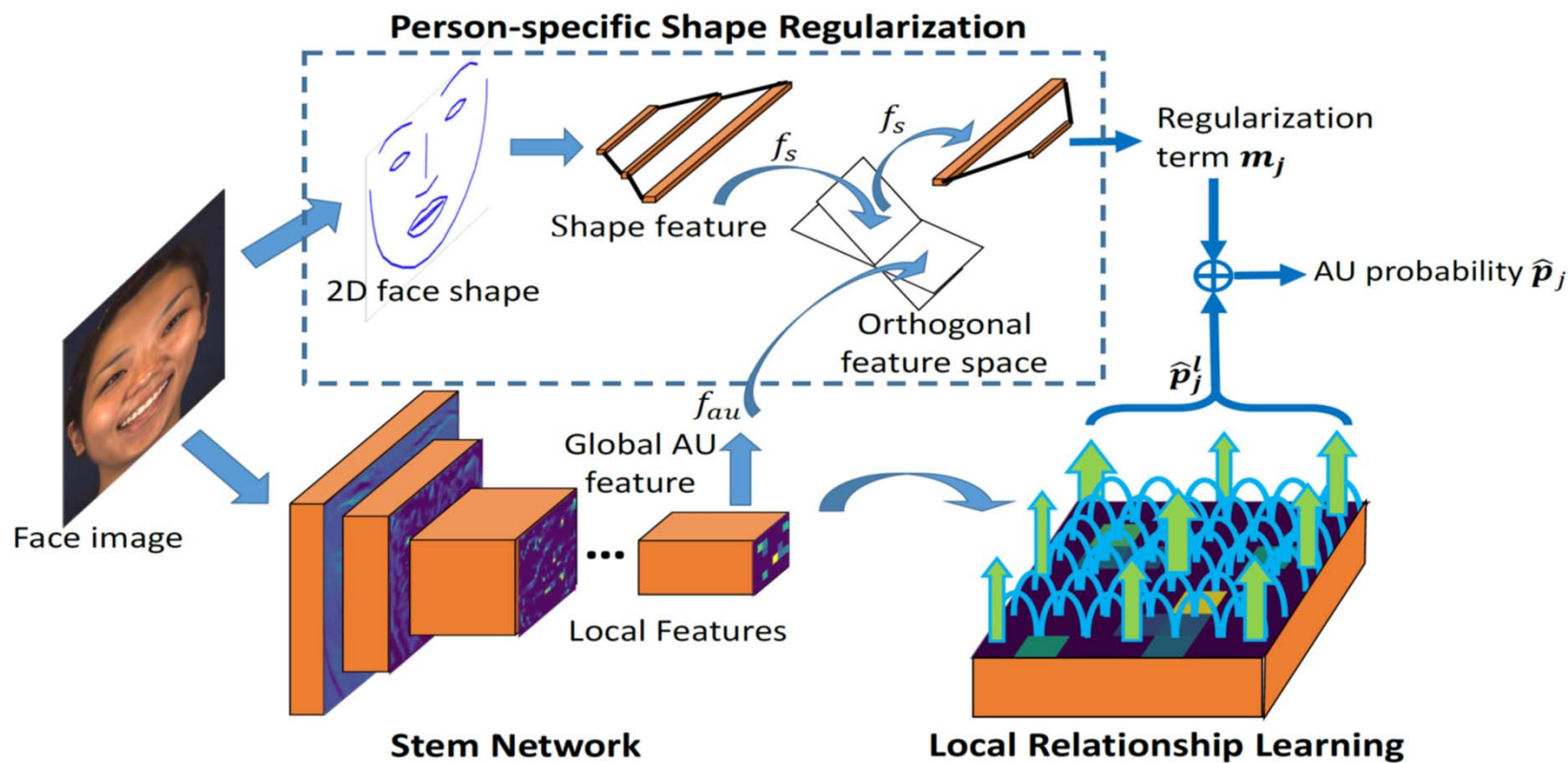
❖ Disentangled representation learning via regularization



X. Niu, H. Han, S. Yang, Y. Huang, and S. Shan. Local Relationship Learning with Person-specific Regularization for Facial Action Unit Detection. *IEEE CVPR*, 2019.

Disentangled representation learning for AU recognition

❖ Network



X. Niu, H. Han, S. Yang, Y. Huang, and S. Shan. Local Relationship Learning with Person-specific Regularization for Facial Action Unit Detection. [IEEE CVPR, 2019.](#)



Disentangled representation learning for AU recognition

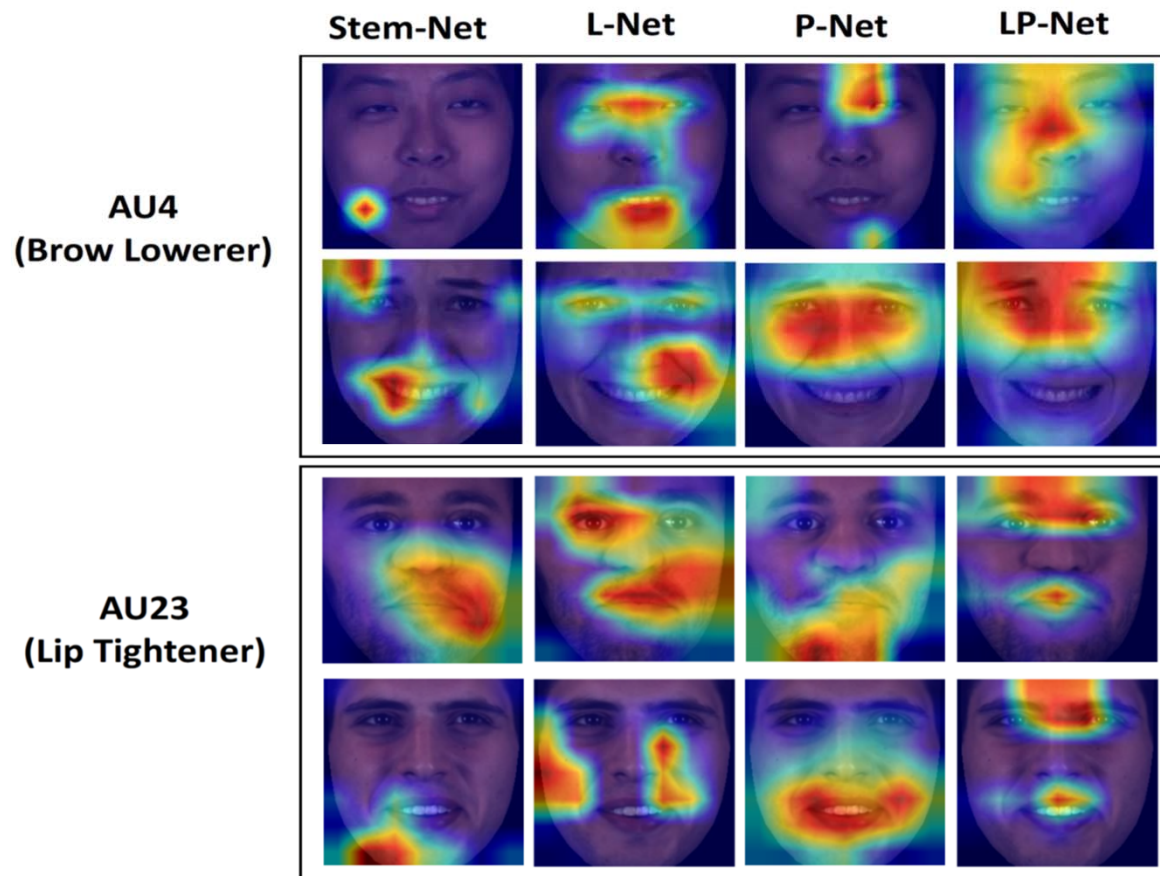
❖ Results on DISFA

Method	AU1	AU2	AU4	AU6	AU9	AU12	AU25	AU26	Avg.
LSVM [11]	10.8	10.0	21.8	15.7	11.5	70.4	12.0	22.1	21.8
DRML [47]	17.3	17.7	37.4	29.0	10.7	37.7	38.5	20.1	26.7
APL [48]	11.4	12.0	30.1	12.4	10.1	65.9	21.4	26.9	23.8
ROI [24]	41.5	26.4	66.4	[50.7]	8.5	89.3	88.9	15.6	48.5
JAA-Net [35]	[43.7]	[46.2]	56.0	41.4	44.7	69.6	88.3	[58.4]	[56.0]
DSIN [8]	[42.4]	[39.0]	[68.4]	28.6	[46.8]	[70.8]	[90.4]	42.2	53.6
LP-Net	29.9	24.7	[72.7]	[46.8]	[49.6]	[72.9]	[93.8]	[65.0]	[56.9]

X. Niu, H. Han, S. Yang, Y. Huang, and S. Shan. Local Relationship Learning with Person-specific Regularization for Facial Action Unit Detection. [IEEE CVPR](#), 2019.

Disentangled representation learning for AU recognition

❖ Visualization

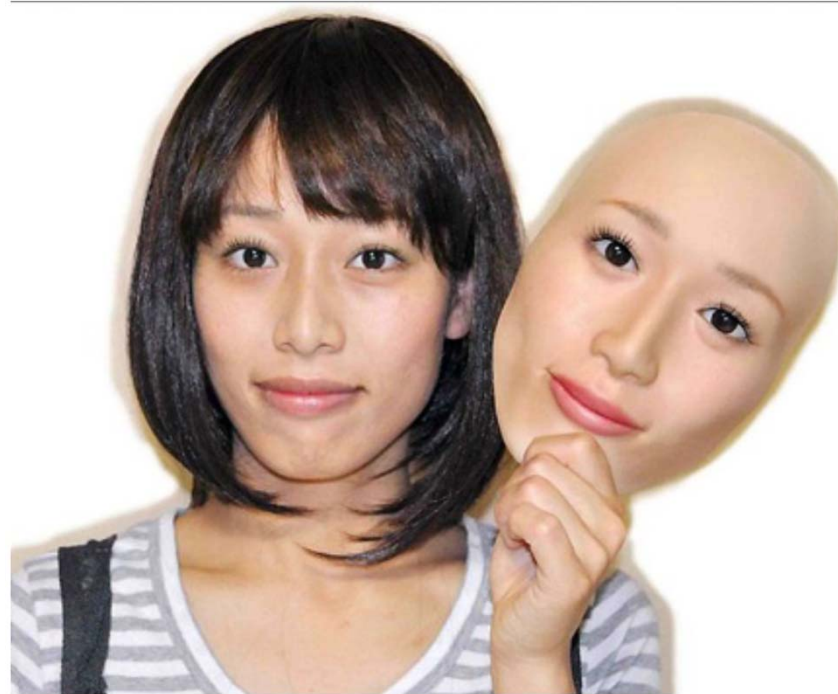


X. Niu, H. Han, S. Yang, Y. Huang, and S. Shan. Local Relationship Learning with Person-specific Regularization for Facial Action Unit Detection. *IEEE CVPR*, 2019.



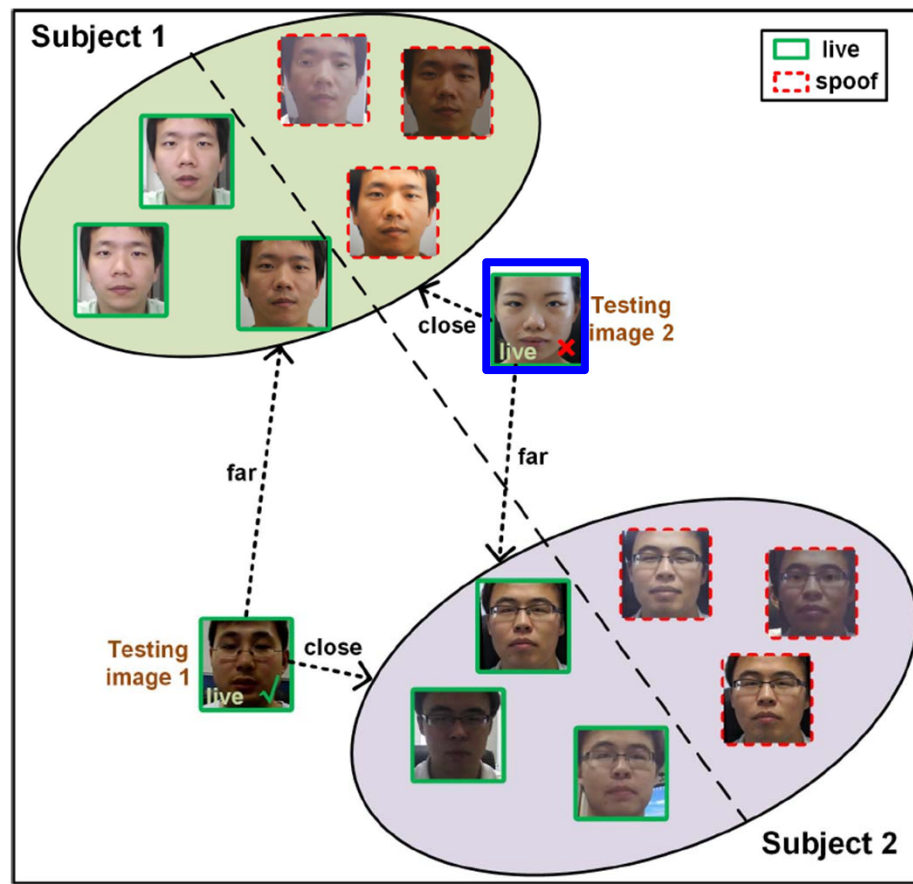
Face anti-spoofing(presentation attack detection)

❖ Is this a live face?

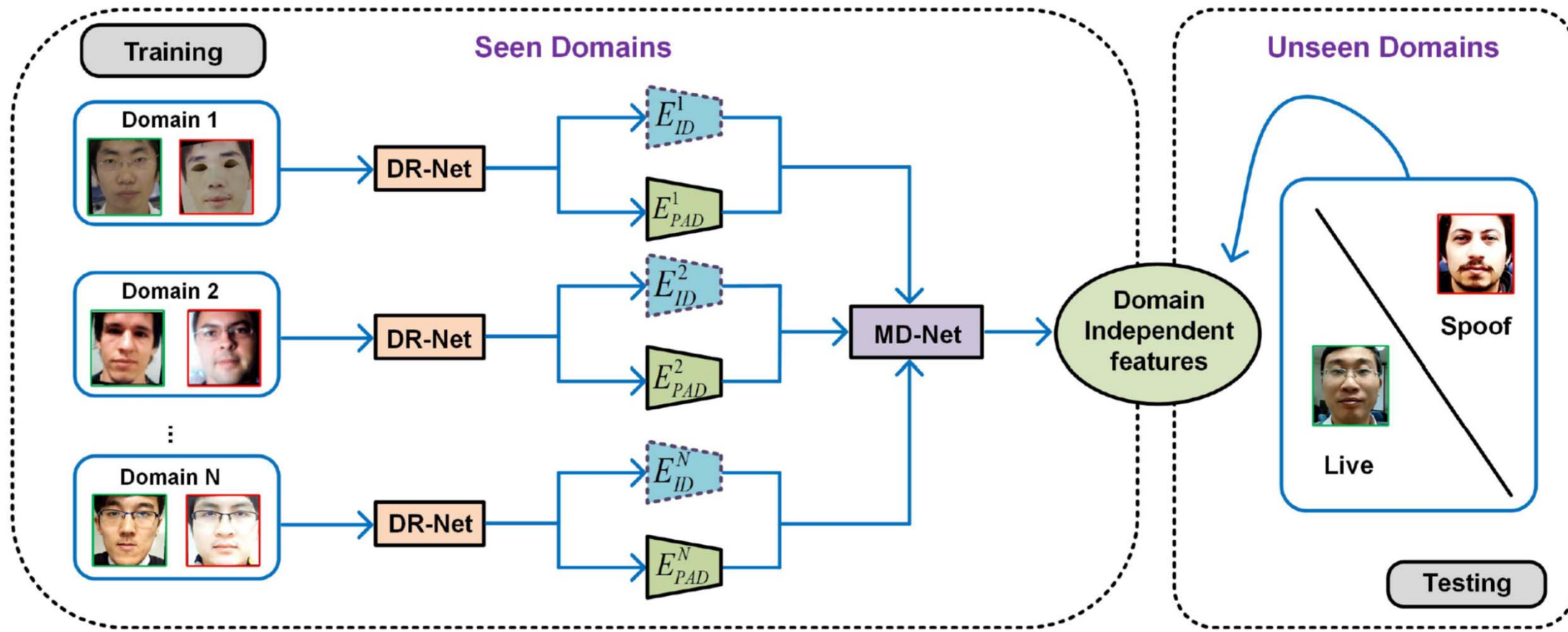


The differences between live and spoof faces can be minor!

- ❖ While the error rate for intra-database testing can be close zero, the error rate for cross-database testing can be high
 - How to learn a domain-independent representation for PAD?

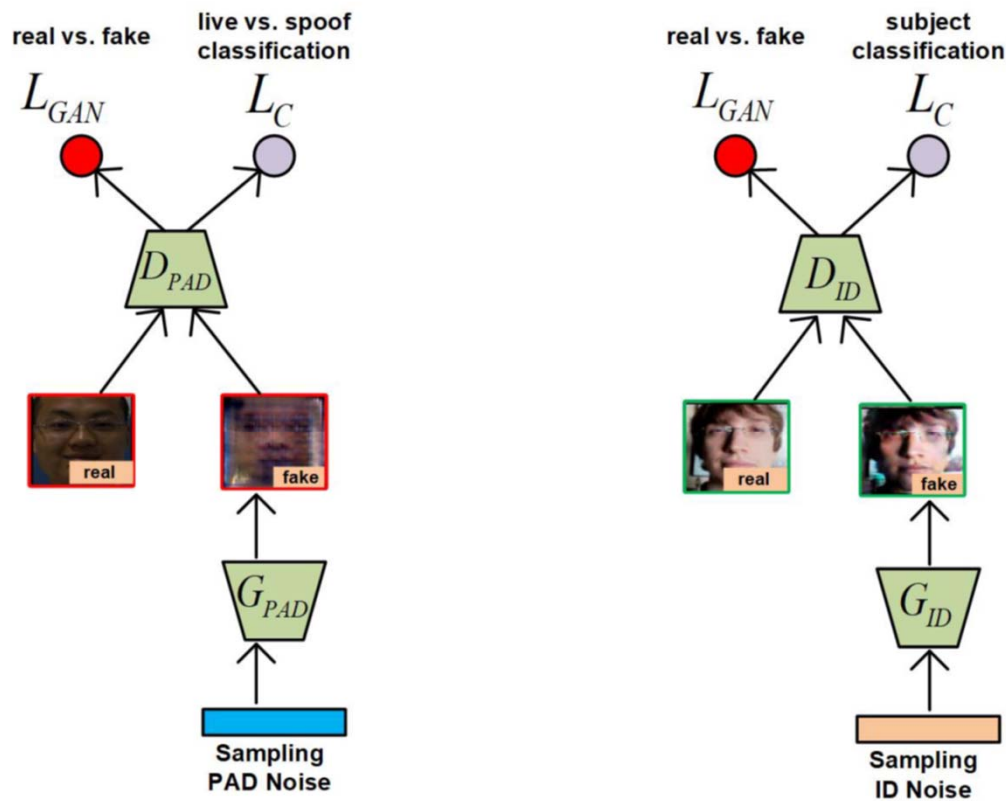


❖ Multi-domain Disentangled Representation Learning



G. Wang, H. Han, S. Shan, and X. Chen. Cross-domain Face Presentation Attack Detection via Multi-domain Disentangled Representation Learning. [IEEE CVPR](#), 2020.

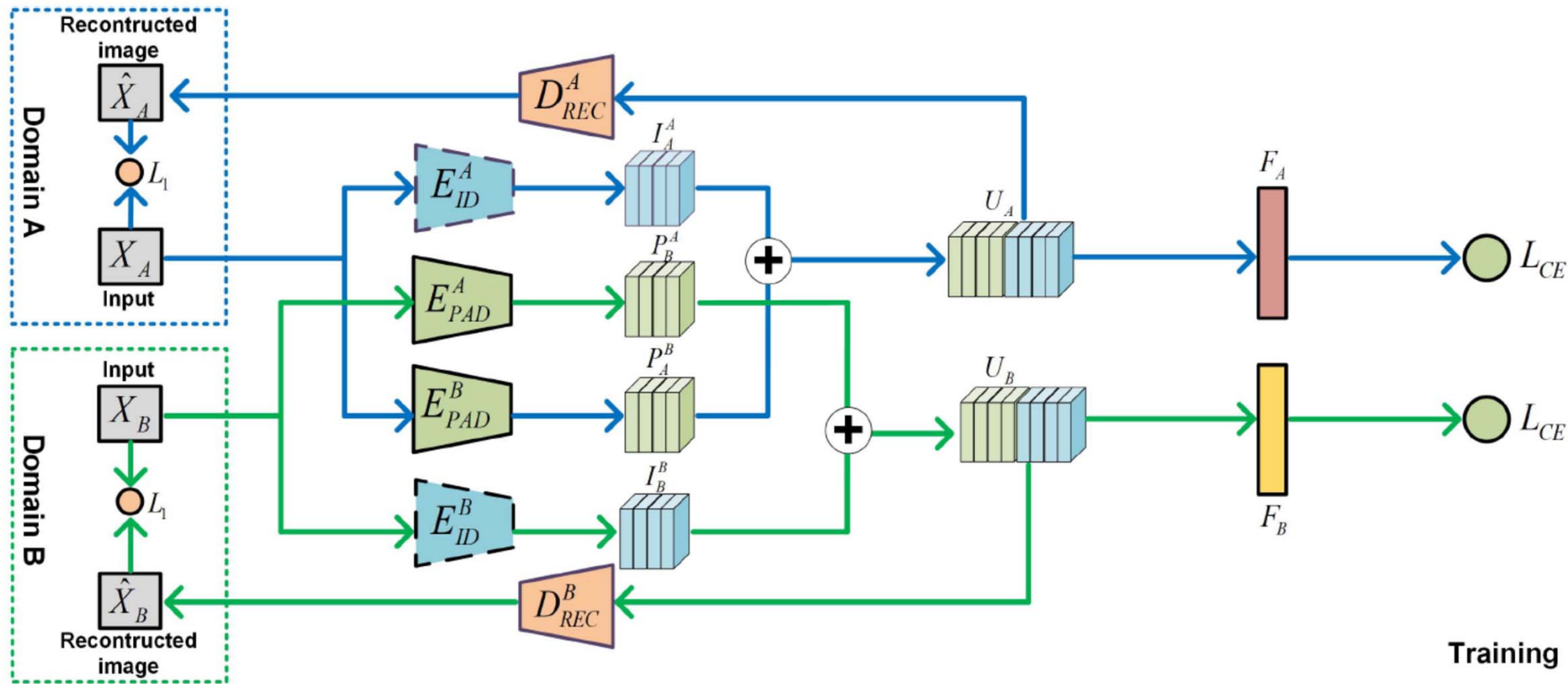
❖ Two types of encoders for representation learning



(a) Encoder for anti-spoof feature (b) Encoder for ID feature

G. Wang, H. Han, S. Shan, and X. Chen. Cross-domain Face Presentation Attack Detection via Multi-domain Disentangled Representation Learning. [IEEE CVPR, 2020.](#)

❖ Self-supervision



G. Wang, H. Han, S. Shan, and X. Chen. Cross-domain Face Presentation Attack Detection via Multi-domain Disentangled Representation Learning. [IEEE CVPR](#), 2020.



Challenge of face anti-spoofing

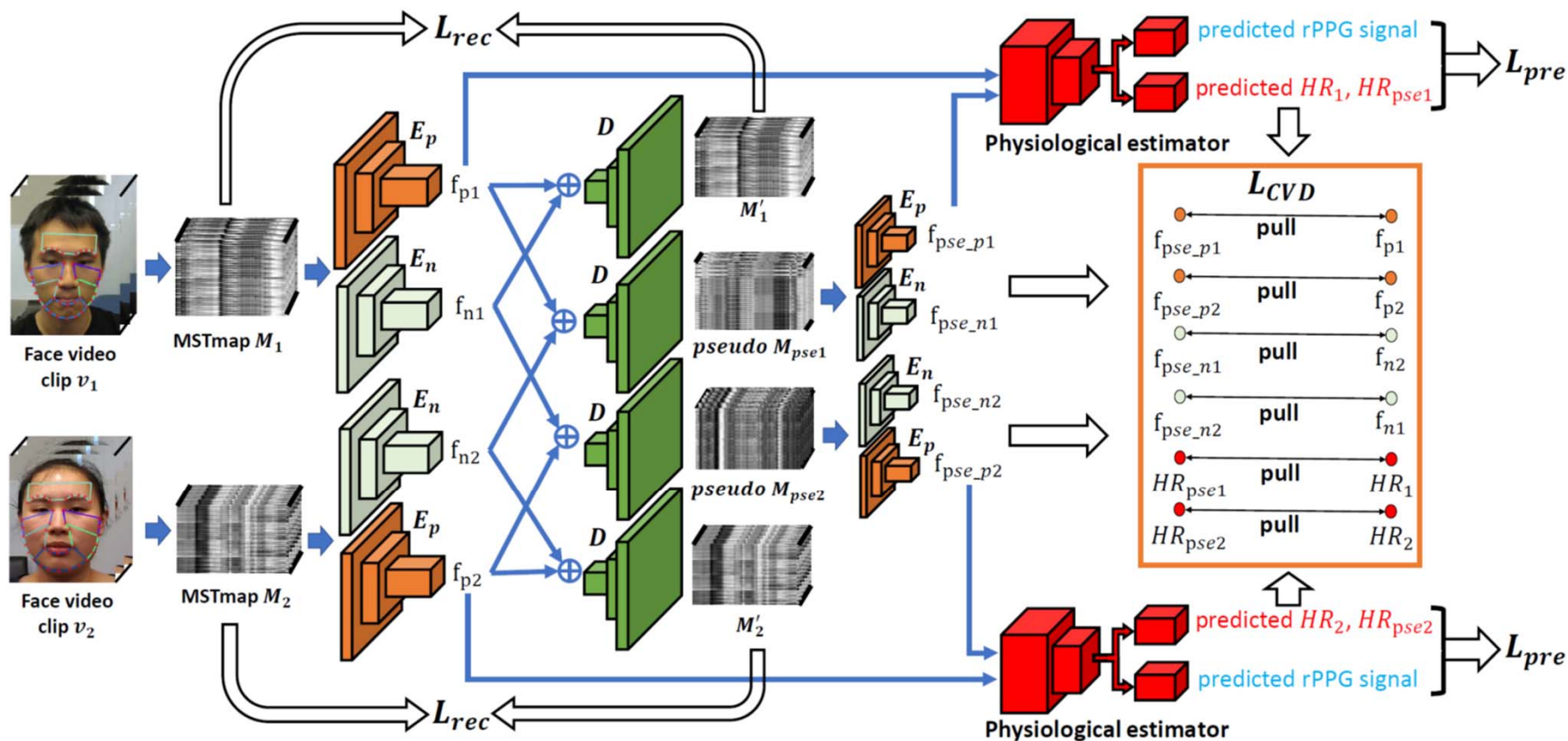
❖ Cross-database testing results

Method	[O, C, I]→M		[O, M, I]→C		[O, C, M]→I		[I, C, M]→O	
	HTER(%)	AUC(%)	HTER(%)	AUC(%)	HTER(%)	AUC(%)	HTER(%)	AUC(%)
MS_LBP [32]	29.76	78.50	54.28	44.98	50.30	51.64	50.29	49.31
CNN [50]	29.25	82.87	34.88	71.95	34.47	65.88	29.61	77.54
IDA [48]	66.67	27.86	55.17	39.05	28.35	78.25	54.20	44.59
CT [5]	28.09	78.47	30.58	76.89	40.40	62.78	63.59	32.71
LBPTOP [11]	36.90	70.80	42.60	61.05	49.45	49.54	53.15	44.09
Aux(Depth only) [30]	22.72	85.88	33.52	73.15	29.14	71.69	30.17	66.61
Aux(All) [30]	-	-	28.4	-	27.6	-	-	-
MMD-AAE [26]	27.08	83.19	44.59	58.29	31.58	75.18	40.98	63.08
MADDG [38]	17.69	88.06	24.5	84.51	22.19	84.99	27.98	80.02
Proposed approach	17.02	90.10	19.68	87.43	20.87	86.72	25.02	81.47

G. Wang, H. Han, S. Shan, and X. Chen. Cross-domain Face Presentation Attack Detection via Multi-domain Disentangled Representation Learning. [IEEE CVPR](#), 2020.

Disentangled Representation Learning for HR Estimation

❖ Cross-verified feature learning



Xuesong Niu, Zitong Yu, Hu Han, Xiaobai Li, Shiguang Shan, Guoying Zhao. Video-based Remote Physiological Measurement via Cross-verified Feature Disentangling. *ECCV*, 2020. (Oral, 2%)



Disentangled Representation Learning for HR Estimation

❖ Result on VIPL-HR

Method	Std↓ (bpm)	MAE↓ (bpm)	RMSE↓ (bpm)	r ↑
SAMC [19]	18.0	15.9	21.0	0.11
POS [22]	15.3	11.5	17.2	0.30
CHROM [3]	15.1	11.4	16.9	0.28
I3D [1]	15.9	12.0	15.9	0.07
DeepPhy [2]	13.6	11.0	13.8	0.11
RhythmNet [13]	8.11	5.30	8.14	0.76
Proposed	7.92	5.02	7.97	0.79

Xuesong Niu, Zitong Yu, Hu Han, Xiaobai Li, Shiguang Shan, Guoying Zhao. Video-based Remote Physiological Measurement via Cross-verified Feature Disentangling. [ECCV, 2020. \(Oral, 2%\)](#)



Disentangled Representation Learning for HR Estimation

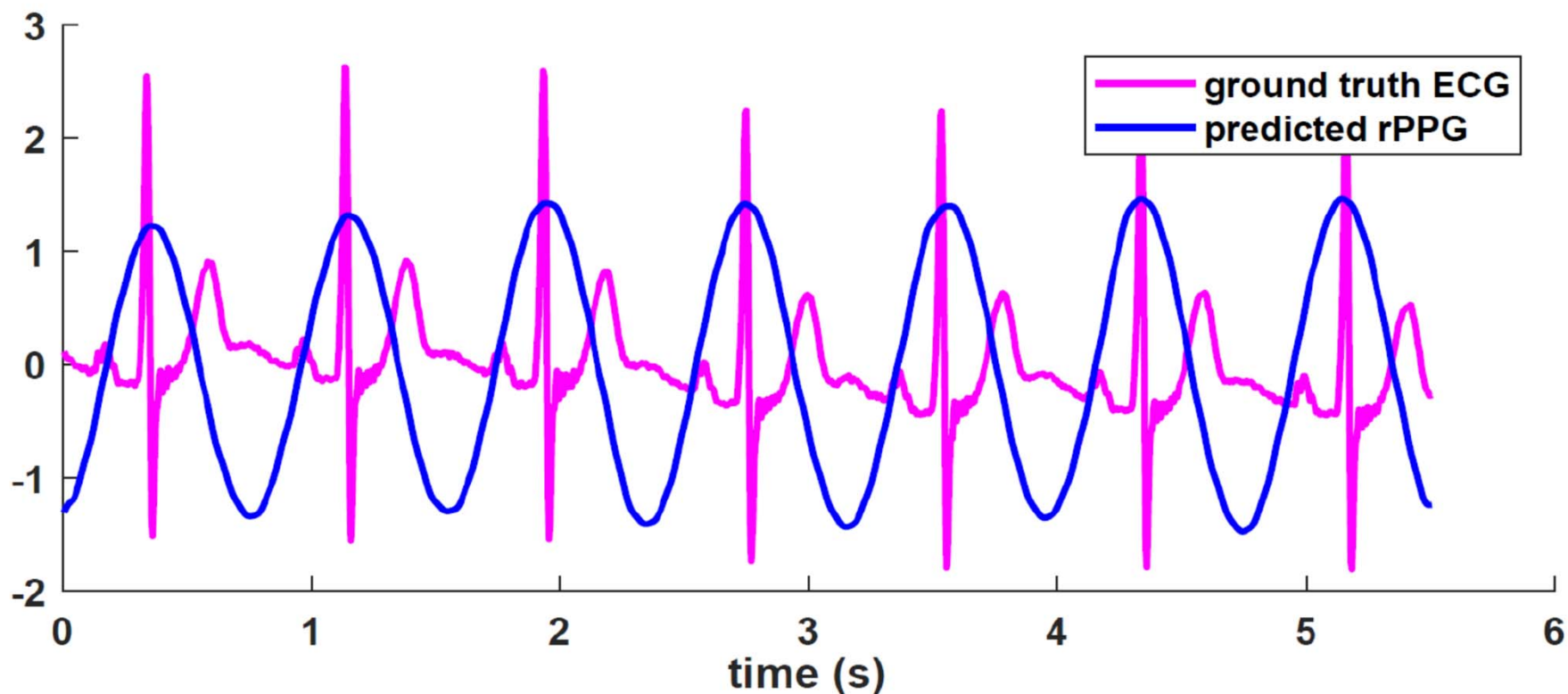
❖ Train on VIPL-HR, test on MMSE-HR

Method	Std↓ (bpm)	RMSE↓ (bpm)	r ↑
Li2014 [9]	20.02	19.95	0.38
CHROM [3]	14.08	13.97	0.55
SAMC [19]	12.24	11.37	0.71
RhythmNet [13]	6.98	7.33	0.78
Proposed	6.06	6.04	0.84

Xuesong Niu, Zitong Yu, Hu Han, Xiaobai Li, Shiguang Shan, Guoying Zhao. Video-based Remote Physiological Measurement via Cross-verified Feature Disentangling. [ECCV, 2020](#). (Oral, 2%)

Disentangled Representation Learning for HR Estimation

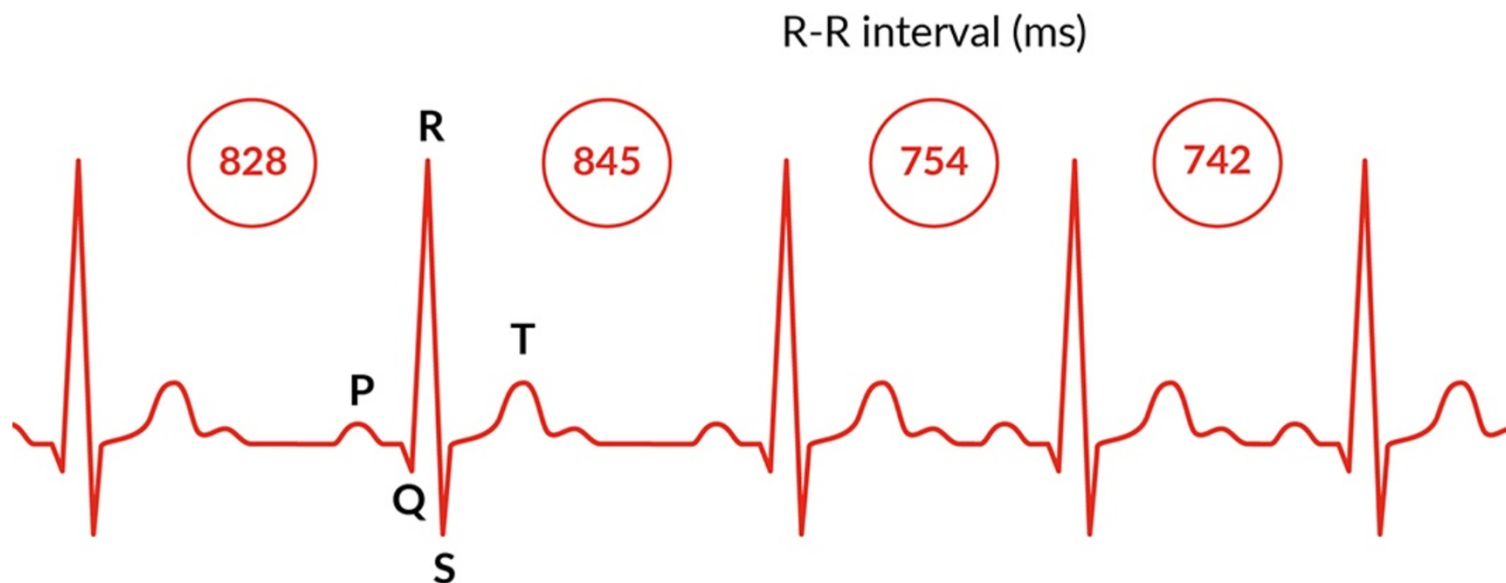
❖ Visualization



Xuesong Niu, Zitong Yu, Hu Han, Xiaobai Li, Shiguang Shan, Guoying Zhao. Video-based Remote Physiological Measurement via Cross-verified Feature Disentangling. [ECCV, 2020](#). (Oral, 2%)

Disentangled Representation Learning for HR Estimation

- ❖ Heart rate variability (HRV): Variation in the time interval between heartbeats
 - Sudden cardiac death have been found to have **lower HRV** than healthy individuals



Xuesong Niu, Zitong Yu, Hu Han, Xiaobai Li, Shiguang Shan, Guoying Zhao. Video-based Remote Physiological Measurement via Cross-verified Feature Disentangling. [ECCV, 2020](#). (Oral, 2%)



Disentangled Representation Learning for HR Estimation

❖ HRV and RF measurement on OBF

Method	RF(Hz)			LF(u.n)			HF(u.n)			LF/HF		
	Std	RMSE	r	Std	RMSE	r	Std	RMSE	r	Std	RMSE	r
ROI _{green} [8]	0.078	0.084	0.321	0.22	0.24	0.573	0.22	0.24	0.573	0.819	0.832	0.571
CHROM [3]	0.081	0.081	0.224	0.199	0.206	0.524	0.199	0.206	0.524	0.83	0.863	0.459
POS [22]	0.07	0.07	0.44	0.155	0.158	0.727	0.155	0.158	0.727	0.663	0.679	0.687
rPPGNet[25]	0.064	0.064	0.53	0.133	0.135	0.804	0.133	0.135	0.804	0.58	0.589	0.773
Proposed	0.058	0.058	0.606	0.09	0.09	0.914	0.09	0.09	0.914	0.453	0.453	0.877

Xuesong Niu, Zitong Yu, Hu Han, Xiaobai Li, Shiguang Shan, Guoying Zhao. Video-based Remote Physiological Measurement via Cross-verified Feature Disentangling. [ECCV, 2020. \(Oral, 2%\)](#)



Outline

- ❖ What is the focus of micro-perspective face analysis
- ❖ Knowledge and data driven modeling [TIP2020, FG2019, ICPR2018]
 - Weak visual signal, insufficient labeled data
- ❖ Disentangled representation learning [CVPR2019, CVPR2020, ECCV2020]
 - Weak visual signal, poor generalization ability
- ❖ Summary



Summary

- ❖ Professional domain knowledge/prior is becoming more important when we need to take a close look at face from a micro perspective
 - Medical image analysis
- ❖ It is necessary to leverage unlabeled/inexactly/inaccurately labeled data to learn a better representation
- ❖ Multi-modality/cues can be helpful for micro perspective (face) analysis



Related papers

1. Guoqing Wang, Hu Han, Shiguang Shan, and Xilin Chen. Unsupervised Adversarial Domain Adaptation for Cross-domain Face Presentation Attack Detection, IEEE Transactions on Information Forensics and Security (T-IFS), Jun. 2020, DOI: 10.1109/TIFS.2020.3002390.
2. Han Li, Hu Han, Zeju Li, Lei Wang, Zhe Wu, Jingjing Lu, and S. Kevin Zhou. High-Resolution Chest X-ray Bone Suppression Using Unpaired CT Structural Priors, IEEE Transactions on Medical Imaging (T-MI), Apr. 2020, DOI: 10.1109/TMI.2020.2986242.
3. Xuesong Niu, Shiguang Shan, Hu Han, and Xilin Chen. RhythmNet: End-to-end Heart Rate Estimation from Face via Spatial-temporal Representation, IEEE Transactions on Image Processing (T-IP), vol. 29. no. 1, pp. 2409-2423, Dec. 2020.
4. Xuesong Niu, Zitong Yu, Hu Han, Xiaobai Li, Shiguang Shan, Guoying Zhao. Video-based Remote Physiological Measurement via Cross-verified Feature Disentangling. To appear in Proc. 16th European Conference on Computer Vision (ECCV), pp. 1-16, Glasgow, United Kingdom, Aug. 23-28, 2020. (Oral, 2%)
5. Guoqing Wang, Hu Han, Shiguang Shan, and Xilin Chen. Cross-domain Face Presentation Attack Detection via Multi-domain Disentangled Representation Learning. in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 6678-6687, Seattle, Washington, USA, Jun. 14-19, 2020.
6. Zeju Li, Han Li, Hu Han, Gonglei Shi, Jiannan Wang, and S. Kevin Zhou. Encoding CT Anatomy Knowledge for Unpaired Chest X-ray Image Decomposition, in Proc. International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI), pp. 275-283, Shenzhen, China, Oct 13-17, 2019.
7. Xuesong Niu, Hu Han, Songfan Yang, Yan Huang, and Shiguang Shan. Local Relationship Learning with Person-specific Regularization for Facial Action Unit Detection. in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 11917-11926, Long Beach, CA, USA, Jun. 18-20, 2019.
8. Xuesong Niu, Hu Han, Shiguang Shan, and Xilin Chen. Multi-label Co-regularization for Semi-supervised Facial Action Unit Recognition, in Proc. the 33rd Annual Conference on Neural Information Processing Systems (NeurIPS), pp. 909-919, Vancouver, Canada, Dec. 8-14, 2019.



Thank you!