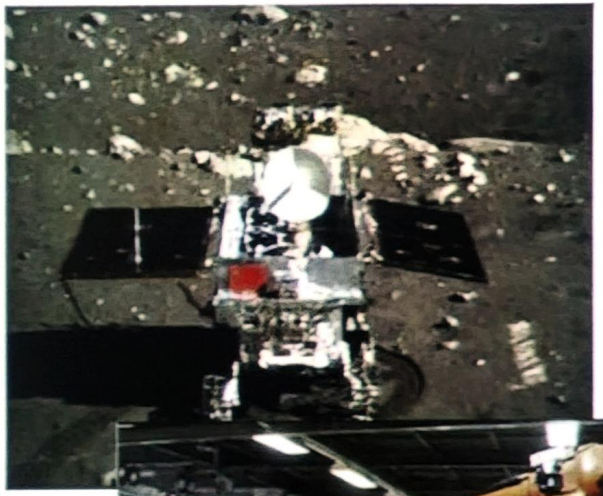
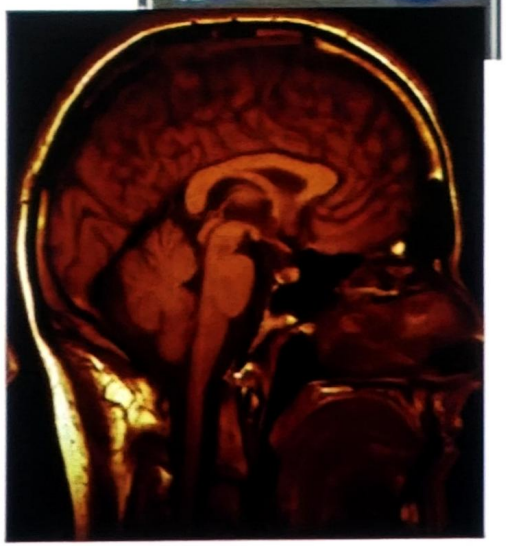
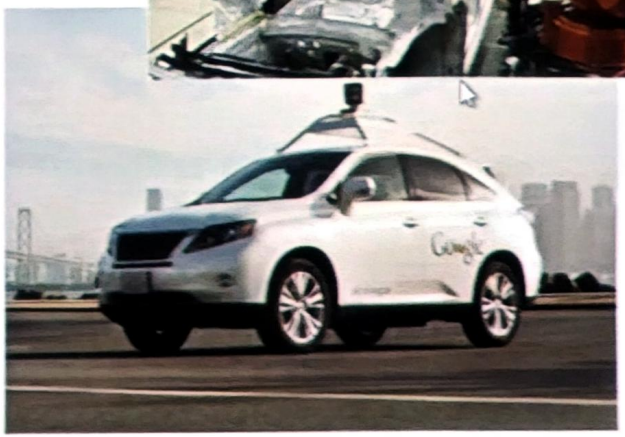
广泛的应用前景

- 通信娱乐
- 平安城市
- 智能制造
- 人机交互
- 智慧医疗
- 航天国防
- ……







表情分析

■ 表情分析流程

表情（静态、动态）视觉学习

表情（AU）识别器

高兴、厌恶、生气.....

■ 表情分类表示

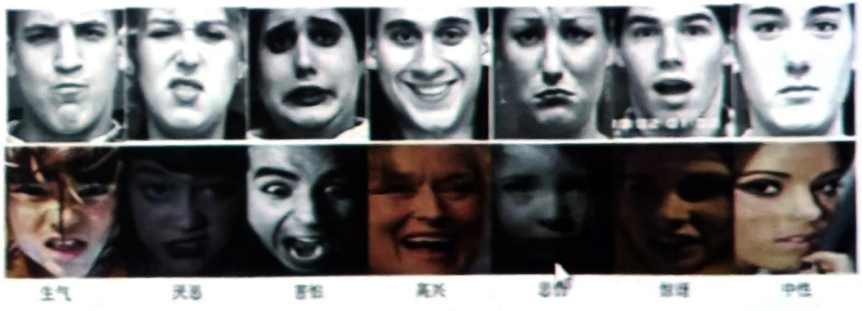
- Ekman分类理论 (1972)

  负面情绪: Anger, Disgust, Fear, Sadness

  正面情绪: Joy(Happiness), Surprise

- Ekman和Wallac AU编码系统1978)

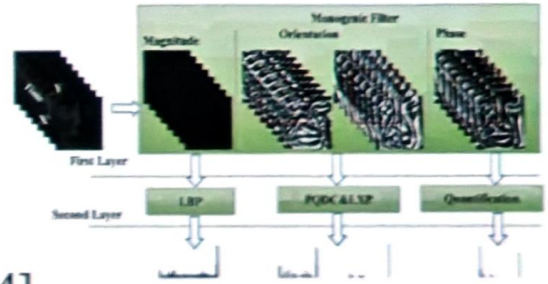
  面部运动单元编码系统（Action Unit）



表情特征研究发展

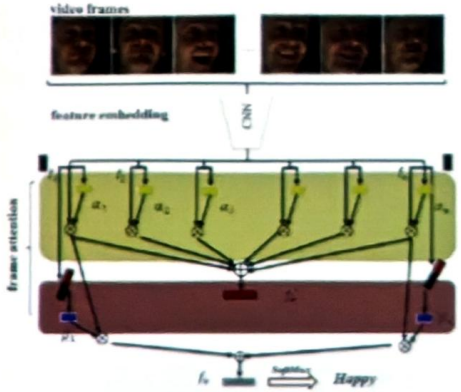
■ 前深度学习时代：手工设计特征

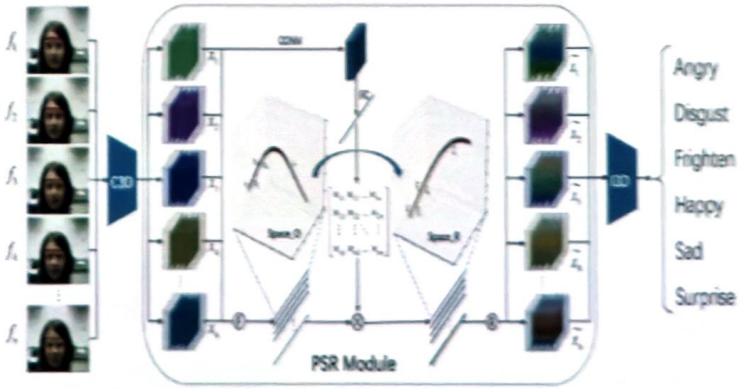
- **LBP-TOP** (D. Huang, C. Shan et al., T-CYB'11)
- **ST-LBP** [X. Huang, Q. He et al., ICMI'14]
- **HOG-TOP** [J. Chen, Z. Chen et al., ICMI'14]
- **Spatial-temporal Manifold, STM** [Liu et al., CVPR'14]

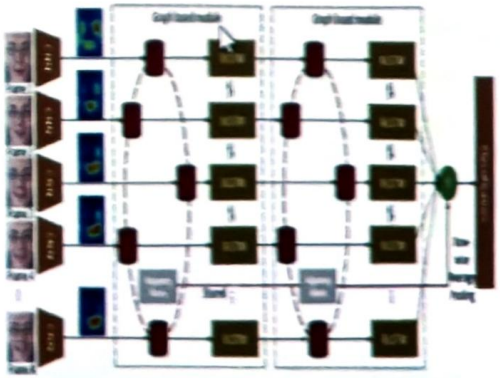


■ 深度学习时代：表情深度特征学习

- **CNN-based**
  [Meng et al., ICIP'19]
  [Lee et al., FG'19]
- **3D-CNN-based**
  [Wang et al., T-AC'20]
  [Jiang et al., ACM MM'20]
- **CNN-RNN(LSTM/GRU)-based**
  [Kuo et al., CVPR'18]
  [Liu et al., ICPR'20]



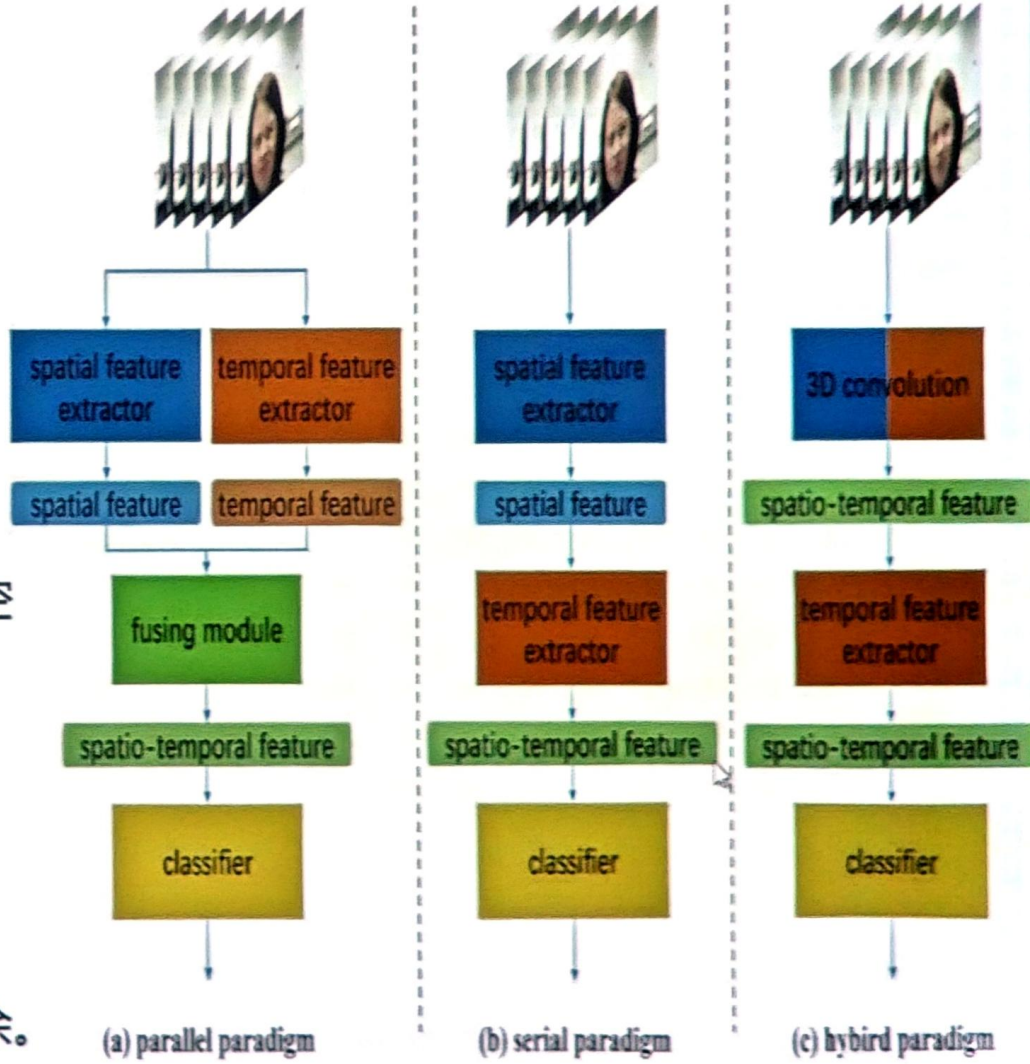




1、相空间驱动时空表情特征学习 (IEEE T-AC 2020)

■ **关键问题：时空动态特征抽取**

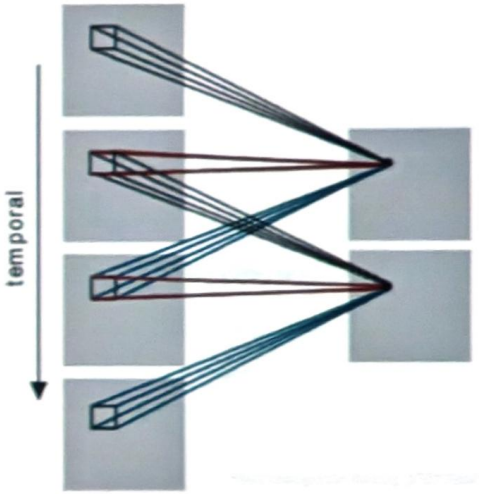
- **并行模式**：如双流网络，割裂地提取时空特征忽略了两者的互补信息；
- **串行模式**：如CNN + LSTM，忽略了不同时刻空间纹理对应关系；
- **混合模式**：如C3D + LSTM， C3D和LSTM 均能提取时序特征，存在信息冗余。



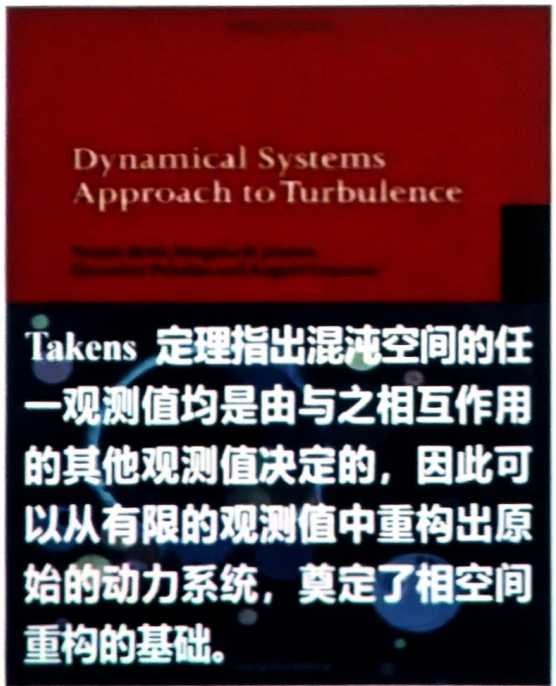
1、相空间驱动时空表情特征学习 (IEEE T-AC 2020)

■ **三维卷积存在的问题：**能同步提取时空特征，但全局时序信息缺失;

$$V_{ij}^{xyz} = b_{ij} + \sum_{m} \sum_{p=0}^{P_i-1} \sum_{q=0}^{Q_i-1} \sum_{r=0}^{R_i-1} w_{ijm}^{pqr} v_{(i-1)m}^{(x+p)(y+q)(z+r)}$$



■ **研究动机：将相空间重构思想应用到时空特征学习**



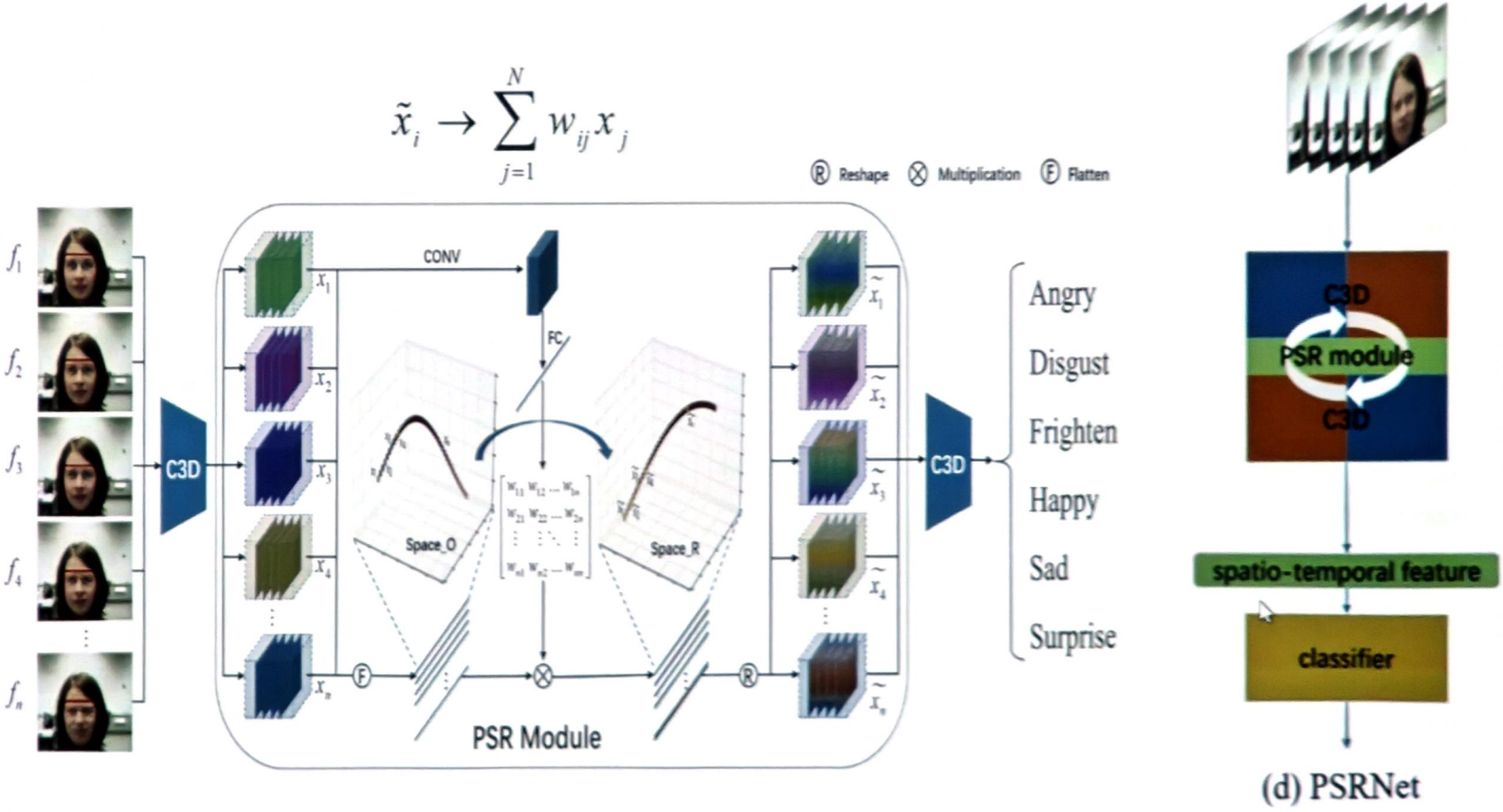
- 坐标延迟法通过衡量观测值关系计算延时参数和维度参数重构相空间;

$$\tilde{X} = f(X,\tau,d) = \begin{bmatrix} x_1 & x_{1+\tau} & \cdots & x_{1+(d-1)\tau} \\ x_2 & x_{2+\tau} & \cdots & x_{2+(d-1)\tau} \\ \vdots & \vdots & \ddots & \vdots \\ x_n & x_{n+\tau} & \cdots & x_{n+(d-1)\tau} \end{bmatrix}$$

1、相空间驱动时空表情特征学习 (IEEE T-AC 2020)

- 受相空间重构启发，提出了相空间驱动时空表情特征学习的方法，改进相空间重构模块，学习互相关矩阵刻画建立观测值关系，重构时空特征。

$$\tilde{x}_i \rightarrow \sum_{j=1}^{N} w_{ij} x_j$$



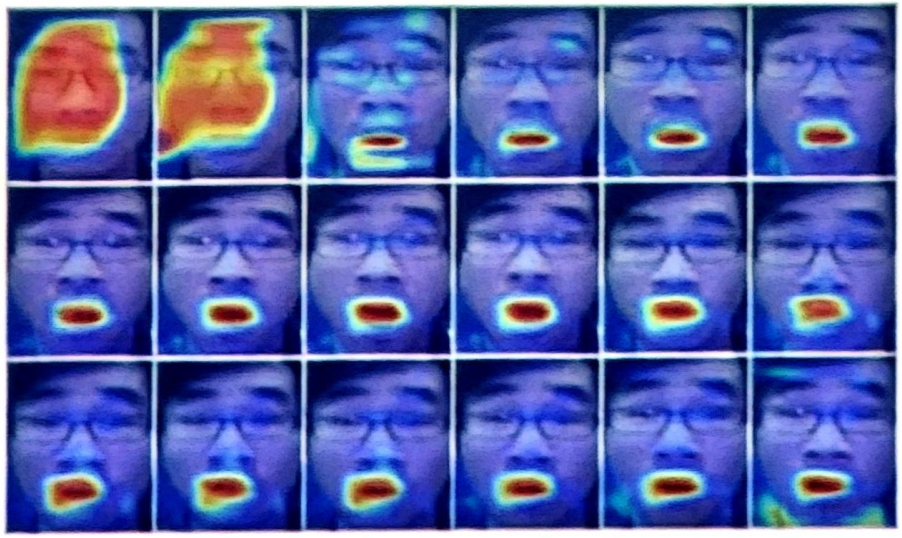
1、相空间驱动时空表情特征学习 (IEEE T-AC 2020)

PSRNet与基线方法的准确率和模型参数对比

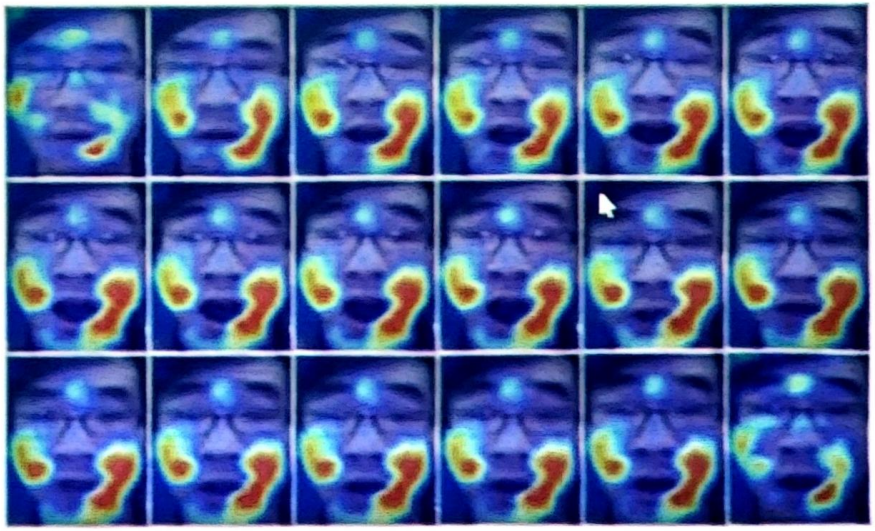
Dataset	Method	Model Parameters(M)	Flops(M)	Accuracy(%)
CK+	C3D	17.9	0.28	95.34
	C3D+LSTM	24.5	504.84	96.13
	PSRNet	18.1	9.66	98.67
Oulu	C3D	18.0	0.56	72.51
	C3D+LSTM	27.8	600	82.9
	PSRNet	18.1	30.95	92.50
MMI	C3D	18.4	0.56	73.81
	C3D+LSTM	41.9	745.15	78.05
	PSRNet	18.5	84.99	85.23

PSRNet 增加了可忽略不计的模型参数，显著提升了模型的识别准确率。

PSRNet与基线方法的可视化对比



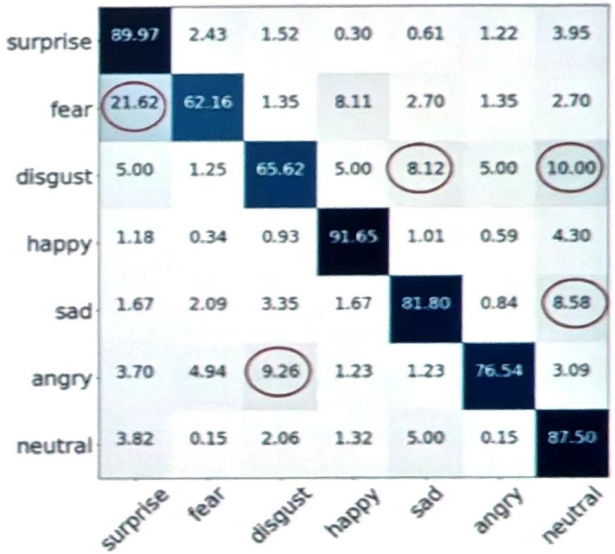
(a) C3D Grad-CAM



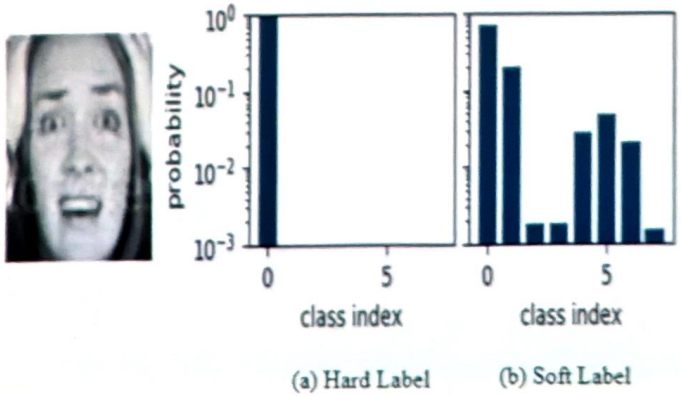
(b) PSRNet Grad-CAM

2、基于反馈的表情软标签协同学习(IEEE T-AC 2022)

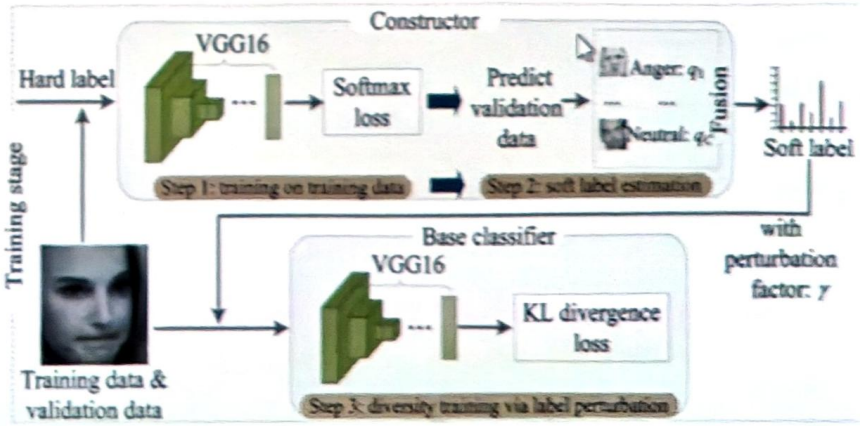
■ **关键问题：标签相似性问题**



- **使用软标签代替单一的标签可以刻画表情的相似性；**

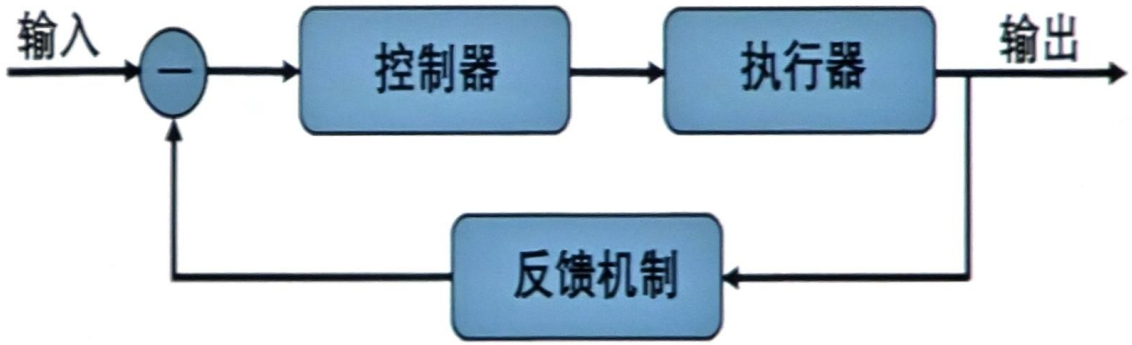


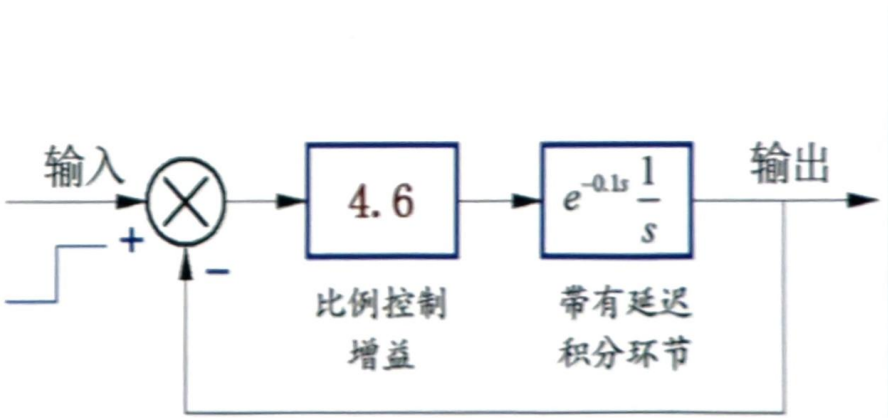
- **常规的软标签学习模型与表情识别模型互相独立，无法实现两者的共同优化。**

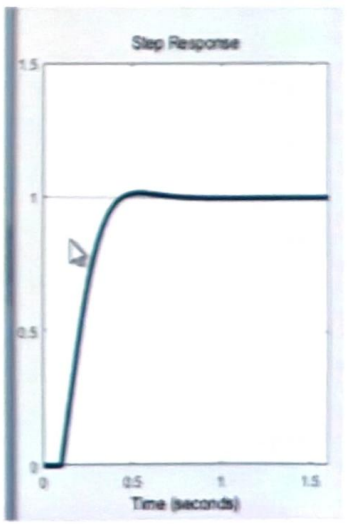


2、基于反馈的表情软标签协同学习(IEEE T-AC 2022)

- 研究动机：将闭环控制理论应用于表情识别任务

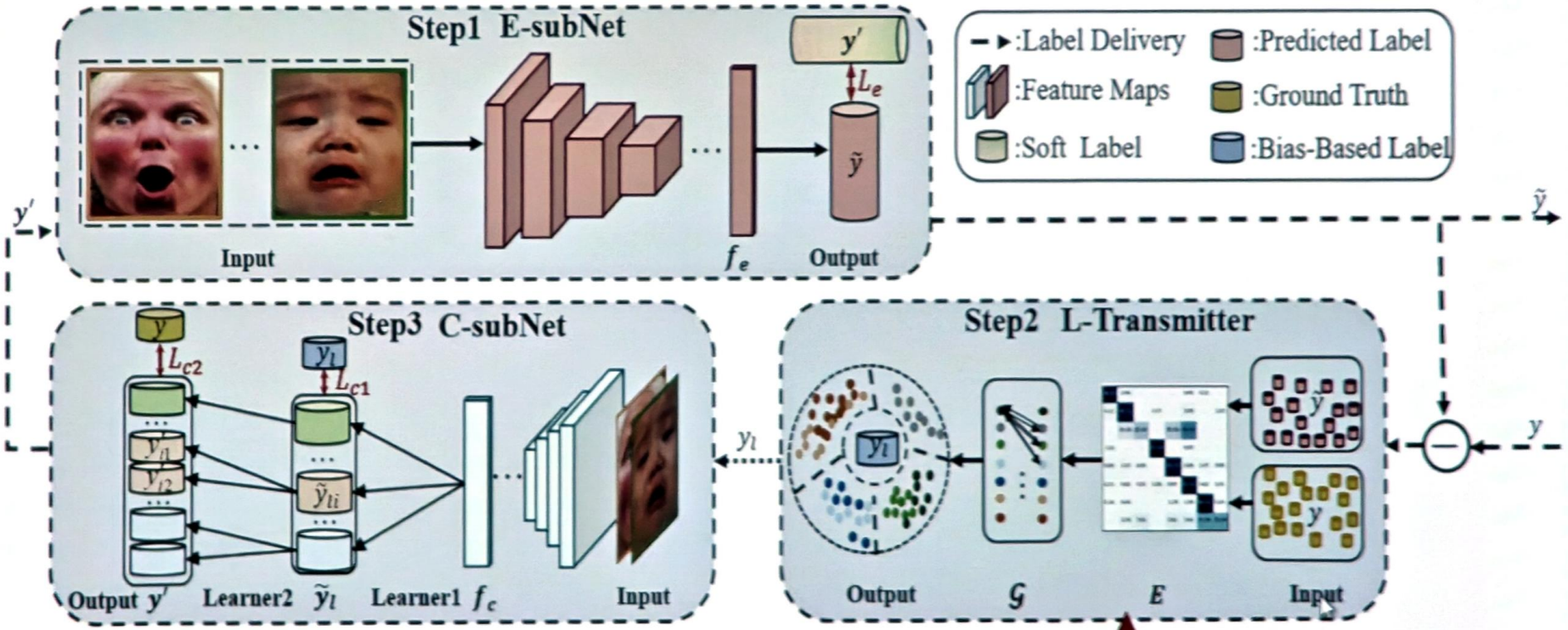






2、基于反馈的表情软标签协同学习(IEEE T-AC 2022)

执行器子网络：表情识别系统



控制器：根据反馈调节目标

传送器：分析模型预测结果，反馈容易误识别的表情，把反馈信息转换为可控制信息

S. Wang，H. Shuai and Q. Liu, Bias-Based Soft Label Learning for Facial Expression Recognition, IEEE Transactions on Affective Computing. 2022.

2、基于反馈的表情软标签协同学习(IEEE T-AC 2022)

EC-Net与经典方法的识别准确率对比

- AffectNet-8类表情

Method	Acc.
DMEU(Res-50)[11]	63.11
DAN [51]	63.45
EfficientFace[36]	63.7
KTN [52]	63.97
EC-Net	**64.45**

- AffectNet-7类表情

Method	Acc.
DAN [51]	59.41
RAN[2]	59.50
EfficientFace[36]	59.89
SCN[28]	60.23
EC-Net	**60.24**

- CAERS-7类表情

Method	Acc.
CAER-Net-S[49]	73.51
MobileNet-V2[56]	79.23
DAN [51]	84.48
ResNet18[8]	85.28
Res2Net-50[57]	85.35
EfficientFace	85.87
EC-Net	**88.01**

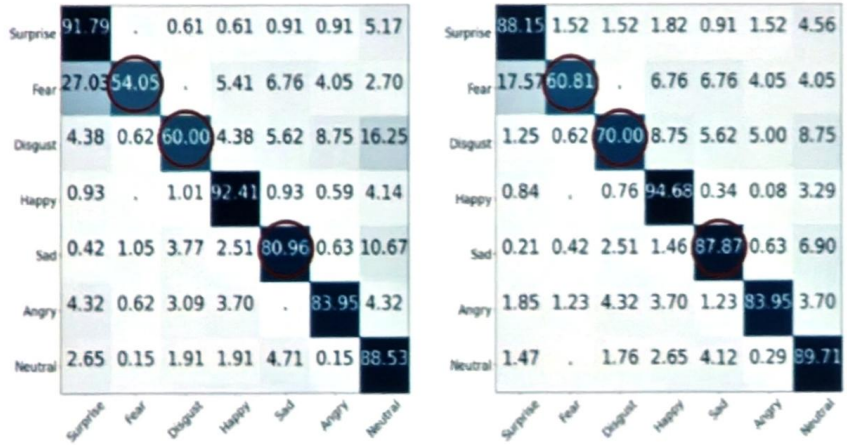
- EC-Net 有无闭环反馈机制的识别准确率对比

	ResNet18		ARM	
	w/o CLF	w/ CLF	w/o CLF	w/ CLF
RAF-DB	85.69	88.14	87.58	89.41
AffectNet-8	54.81	59.19	56.24	60.24
AffectNet-7	62.76	63.5	63.68	64.45
CAER-S	84.62	87.19	82.14	88.01
SFEW	41.28	58.26	50.23	61.01

- **EC-Net 在现有数据集上均取得了较好的结果；**
- **提出的反馈机制在五个公开数据集上的识别率平均提升了5.04%。**

2、基于反馈的表情软标签协同学习(IEEE T-AC 2022)

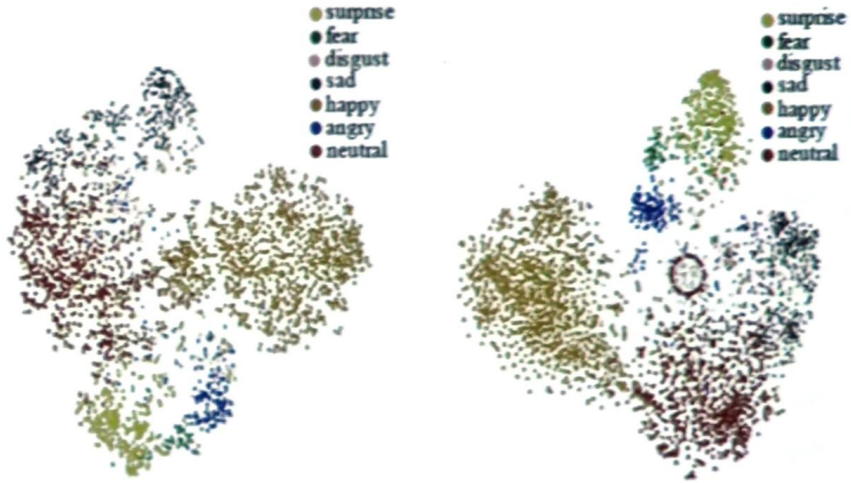
- 各类表情准确率对比



(a) Confusion matrix of the baseline method.

(b) Confusion matrix of EC-Net.

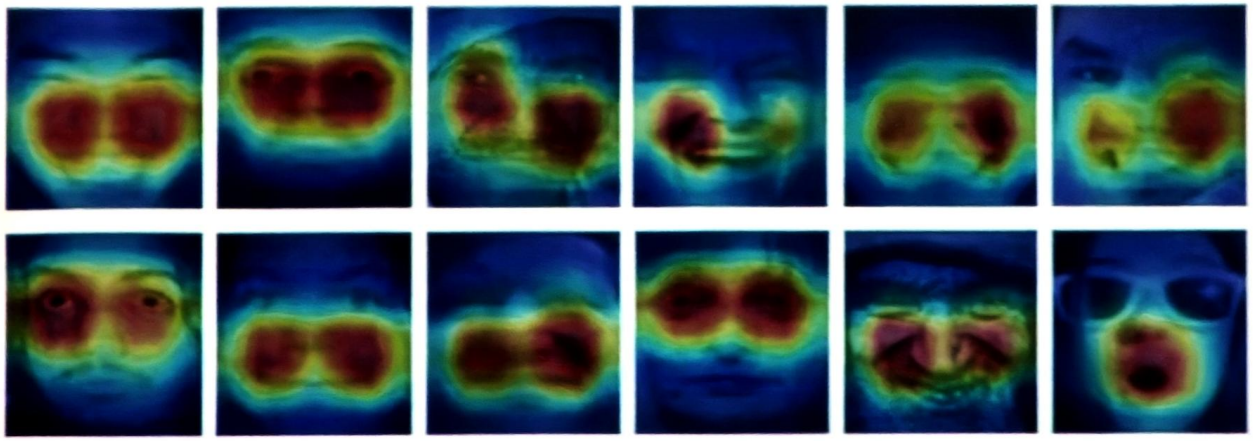
- 各类表情特征分布对比



(a) Baseline.

(b) EC-Net.

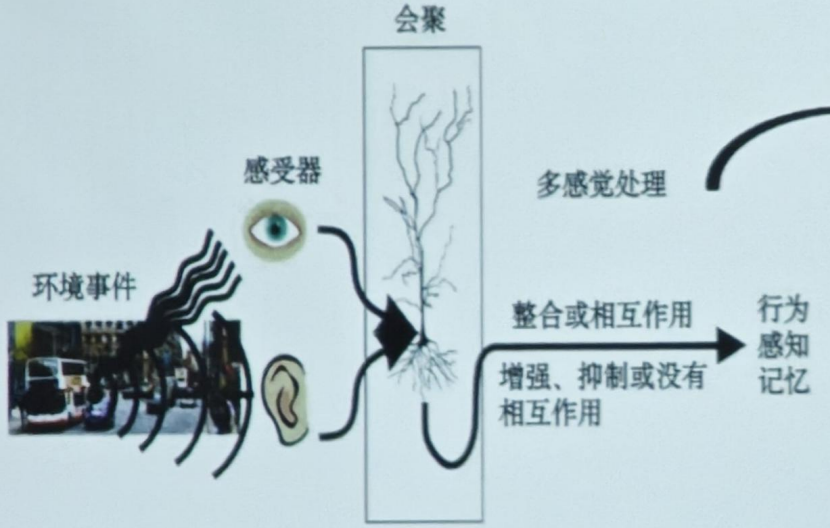
- 显著区域的可视化



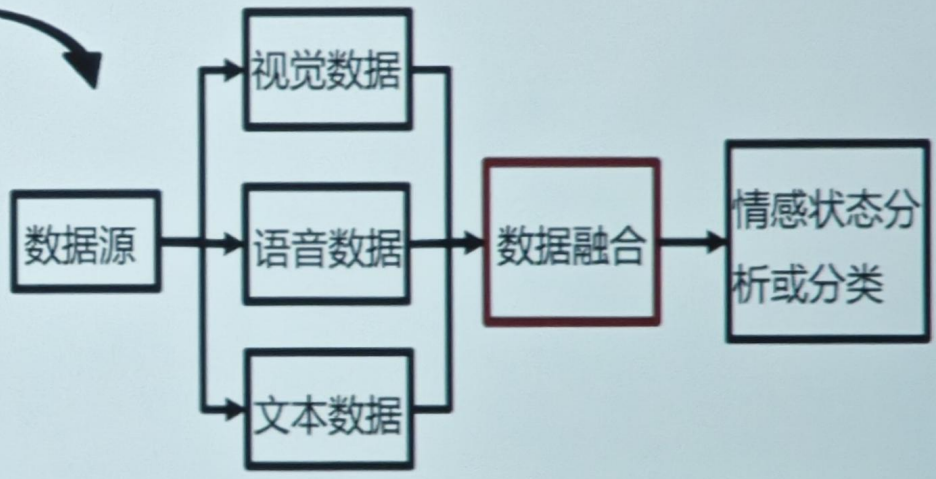
EC-Net对准确率较低的类别 Fear、Disgust 和Sad 提升效果显著。

多模态情感分析

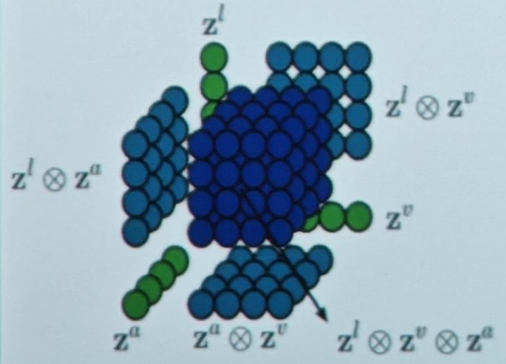
计算认知神经科学中的多感觉整合



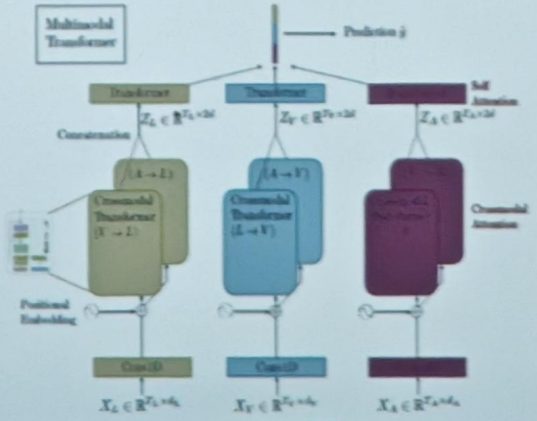
多模态情感分析流程



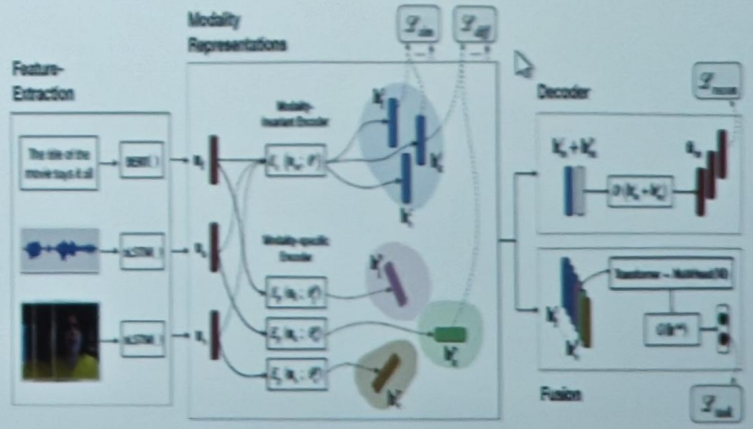
- **特征级的模态融合方法：**基于张量（TFN), 基于门单元 （MFN), 基于transformer (MulT) 等
- **跨模态共性差异性解耦 + 特征级模态融合**: 基于对抗（MISA、FDMER), 基于模态迁移 （CRNet、TCSP）等



TFN



MulT



MISA

3、基于多情感智能体协同的情感计算(Under review)

研究动机

- 多智能体协同控制通过定义智能体间的协同方式和协同目标，利用行为-奖励机制，不断调整各智能体的策略至最优。
- 个体释放多个行为信号表达情感状态的时候，也是多个模态信号的协同表达，如表情、语调和语言内容等。每个模态的信号变化可以看成一个情感智能体，利用多智能体协同来实现多模态情感的分析。

无人机 灯光秀

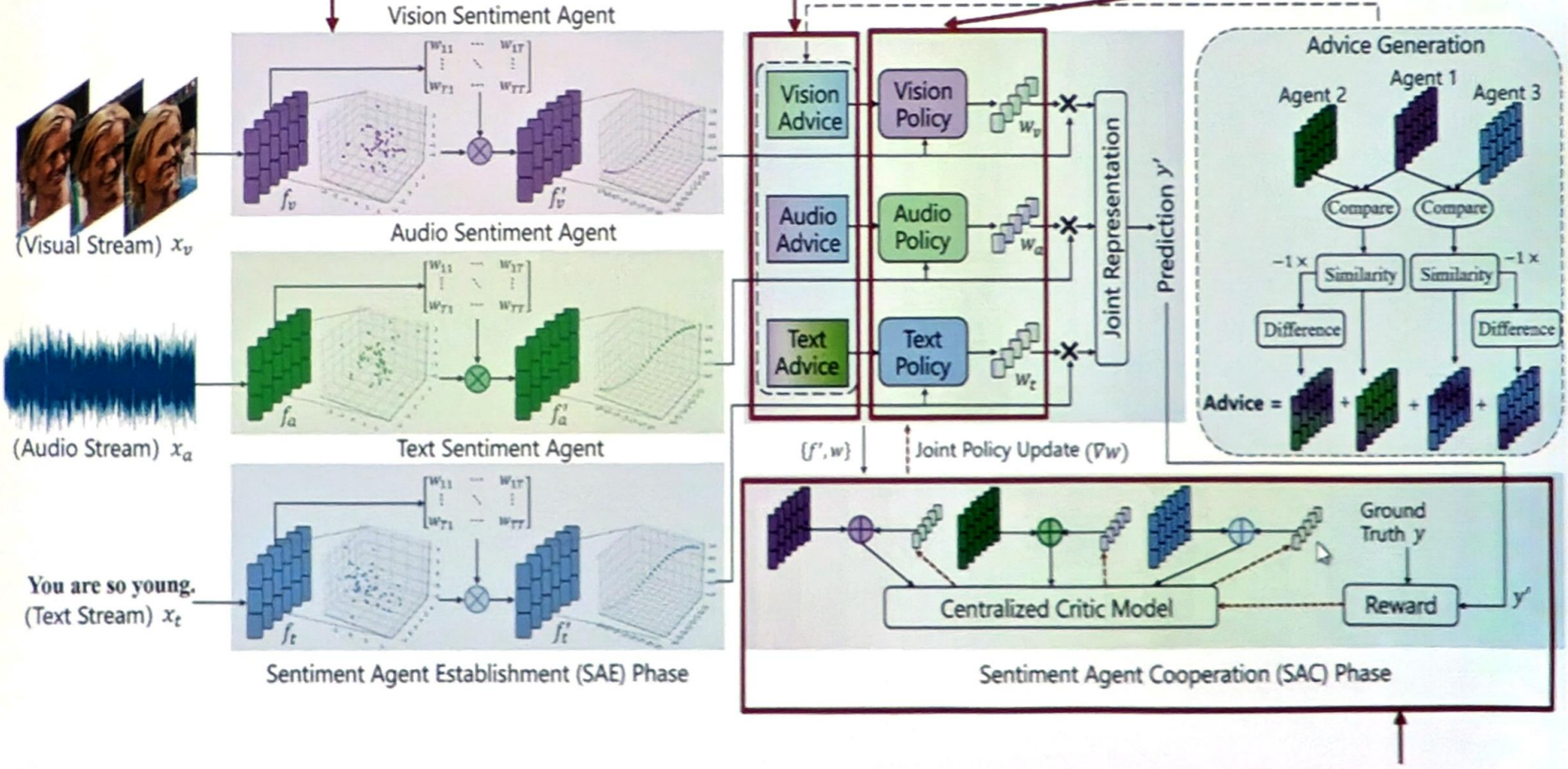
多智能体协同为多模态情感分析任务中动态协同属性的学习提供了理论基础。

3、基于多情感智能体协同的情感计算(Under review)

1.构建多个情感智能体，加强单模态的情感变化属性，作为智能体观测值；

2.情感智能体对比各自观测模态和其它模态的共性和差异性，为其它智能体生成建议，作为信息交流通道；

3.每个情感智能体根据自身的观测和其它智能体的综合建议，采取行为，决定协同的情感属性；



4.将学习到的协同属性应用于下游任务，根据量表或情绪类别计算奖励值,指导各智能体联合的策略调整；

3、基于多情感智能体协同的情感计算(Under review)

■ 实验一：多模态情感分析（MSA）

- MOSI 和 MOSEI 数据集是由 CMU 在线采集的视频，包含视觉，语音和文本三个模态，标注了极度消极到极度积极的情感状态。

	Dataset: MOSI					Dataset: MOSEI				
	Acc7 (↑)	Acc2 (↑)	F1 (↑)	MAE (↓)	Corr (↑)	Acc7 (↑)	Acc2 (↑)	F1 (↑)	MAE (↓)	Corr (↑)
TFN [19]	44.7	82.6	82.6	0.761	0.789	51.8	84.5	84.5	0.622	0.781
LMF [20]	45.1	84.0	84.0	0.742	0.785	51.2	84.2	84.3	0.612	0.779
MFN [24]	44.1	83.5	83.5	0.759	0.786	52.6	84.8	84.8	0.607	0.771
MulT [56]	41.5	83.7	83.7	0.767	0.799	50.7	84.7	84.6	0.625	0.775
GFN [23]	47.0	84.3	84.3	0.736	0.790	51.8	85.0	85.0	0.611	0.774
ICCN [61]	39.0	83.0	83.0	0.860	0.710	51.6	[illegible]	[illegible]	[illegible]	0.713
MAG [53]	42.9	83.5	83.5	0.790	0.769	51.9	[illegible]	[illegible]	0.602	0.778
MISA [29]	42.3	83.4	83.6	0.783	0.761	52.2	[illegible]	[illegible]	0.555	0.756
TFR-Net [63]	42.6	84.0	83.9	0.787	0.788	51.7	[illegible]	[illegible]	[illegible]	0.781
HyCon [64]	46.6	85.2	85.1	0.713	0.790	52.8	85.4	85.6	0.601	0.776
MCL [55]	49.2	86.1	86.1	0.713	0.793	53.3	[illegible]	[illegible]	[illegible]	0.791
MHE [66]	41.5	83.6	83.5	0.801	0.722	52.5	84.2	84.0	0.577	0.712
CMHFM [67]	37.0	81.0	81.3	0.912	0.677	52.6	[illegible]	[illegible]	[illegible]	0.731
CRNet [27]	47.4	86.4	86.4	0.712	0.797	53.8	86.2	86.1	0.541	0.771
TMBL [30]	36.3	83.8	84.3	0.867	0.762	52.4	[illegible]	[illegible]	0.545	0.766
DTN [31]	48.1	86.2	86.2	0.714	0.807	52.5	[illegible]	[illegible]	0.579	0.788
Co-SA(add)	**49.8**	**87.2**	**87.0**	**0.688**	0.812	**54.3**	**86.8**	**86.7**	**0.535**	**0.791**
Co-SA(concatenate)	49.5	86.6	86.5	**0.688**	**0.814**	54.0	86.4	86.4	**0.535**	0.790

与常用方法相比，基于情感智能体的方法（Co-SA）在MSA任务上，多个指标均取得了最好的结果。

3、基于多情感智能体协同的情感计算(Under review)

■ 实验一：多模态情感分析（MSA）

- MOSI 和 MOSEI 数据集是由 CMU 在线采集的视频，包含视觉，语音和文本三个模态，标注了极度消极到极度积极的情感状态。

	Dataset: MOSI					Dataset: MOSEI				
	Acc7 (↑)	Acc2 (↑)	F1 (↑)	MAE (↓)	Corr (↑)	Acc7 (↑)	Acc2 (↑)	F1 (↑)	MAE (↓)	Corr (↑)
TFN [19]	44.7	82.6	82.6	0.761	0.789	51.8	84.5		0.622	0.781
LMF [20]	45.1	84.0	84.0	0.742	0.785	51.2	84.2		0.612	0.779
MFN [24]	44.1	83.5	83.5	0.759	0.786					0.771
MulT [56]	41.5	83.7	83.7	0.767	0.799					0.775
GFN [23]	47.0	84.3	84.3	0.736	0.790					0.774
ICCN [61]	39.0	83.0	83.0	0.860	0.710					0.713
MAG [53]	42.9	83.5	83.5	0.790	0.769					0.778
MISA [29]	42.3	83.4	83.6	0.783	0.761					0.756
TFR-Net [63]	42.6	84.0	83.9	0.787	0.788					
HyCon [64]	46.6	85.2	85.1	0.713	0.790					
MCL [55]	49.2	86.1	86.1	0.713	0.793					0.791
MHE [66]	41.5	83.6	83.5	0.801	0.722	52.5				0.712
CMHFM [67]	37.0	81.0	81.3	0.912	0.677	52.6				0.731
CRNet [27]	47.4	86.4	86.4	0.712	0.797	53.8				0.771
TMBL [30]	36.3	83.8	84.3	0.867	0.762	52.4		85.9		0.766
DTN [31]	48.1	86.2	86.2	0.714	0.807	52.5	86.3	86.3	0.579	0.788
Co-SA(add)	**49.8**	**87.2**	**87.0**	**0.688**	0.812	**54.3**	**86.8**	**86.7**	**0.535**	**0.791**
Co-SA(concatenate)	49.5	86.6	86.5	**0.688**	**0.814**	54.0	86.4	86.4	**0.535**	0.790

与常用方法相比，基于情感智能体的方法（Co-SA）在MSA任务上，多个指标均取得了最好的结果。

3、基于多情感智能体协同的情感计算(Under review)

■ 实验二：多模态情绪识别（MER)

- IEMOCAP 是由南加州大学在实验室环境下采集的对话数据，包含视觉，语音和文本三种模态，被标注了中立，高兴，悲伤，愤怒，惊讶，恐惧，厌恶，挫败，兴奋和其他等情感类别。一般识别高兴，悲伤，愤怒和中性四种情感。

	Happy		Sad		Angry		Neutral		Average	
	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1
MFN [24]	86.5	84.0	83.5	82.1	85.0	83.7	69.6	69.2	81.2	79.8
Graph-MFN [24]	86.8	84.2	83.8	83.0	85.8	85.5	69.4	68.9	81.5	80.4
RAVEN [62]	87.3	85.8	83.4	83.1	87.3	86.7	69.7	69.3	81.9	81.2
LMF [20]	86.9	82.3	85.4	84.7	87.1	86.8	71.6	71.4	82.8	81.3
MulT [56]	87.4	84.1	84.2	83.1	88.0	87.5	69.9	68.4	82.4	80.8
HFFN [65]	86.8	82.1	84.4	84.5	86.6	85.8	69.6	69.3	81.9	80.4
MISA [29]	86.1	80.8	82.3	79.1	84.1	83.8	69.0	67.7	80.4	77.8
TCM-LSTM [57]	87.2	84.8	84.4	84.9	89.0	88.6	71.3	71.2	83.0	82.4
HyCon [64]	88.0	85.5	86.2	85.9	89.4	89.2	70.4	70.5	83.5	82.8
MCL [55]	88.8	86.8	86.6	86.6	90.3	90.3	71.6	71.4	84.3	83.8
UniMF [68]	83.4	85.3	82.9	84.0	84.0	83.2	69.5	70.0	80.0	80.6
TLRF [69]	84.2	86.1	84.9	85.0	87.9	87.7	72.9	73.1	82.5	83.0
TMBL [30]	86.4	85.1	82.8	83.0	86.0	86.3	69.9	69.5	81.3	81.0
Co-SA(add)	88.0	86.5	88.1	87.5	90.4	90.4	73.7	73.5	85.1	84.5
Co-SA(concatenate)	88.4	86.4	87.2	87.2	89.8	89.6	73.5	73.3	84.7	84.1

The optimal results are bolded and underlined, and the sub-optimal results are underlined.

3、基于多情感智能体协同的情感计算(Under review)

■ 实验三：多模态抑郁症检测（MDA）

- CMDC 是采集自78个个体的多模态抑郁数据集。每个被试者回答12个问题，并根据 PHQ9 量表自评抑郁得分。自评得分低于 9 为正常人，否则为抑郁症患者。

	MAE(↓)	RMSE(↓)	Pearson(↑)	Precision(↑)	Recall(↑)	F1(↑)
Bi-LSTM [70]	4.55	5.67	0.68	0.87	0.89	0.88
MulT [56]	4.32	5.61	0.72	0.97	0.85	0.91
MISA [29]	2.47	3.43	0.88	**1.00**	0.93	0.96
TMBL [30]	3.09	3.98	0.90	**1.00**	0.93	0.96
Co-SA(add)	2.27	3.19	0.92	**1.00**	**1.00**	**1.00**
Co-SA(concatenate)	**2.04**	**2.77**	**0.94**	**1.00**	**1.00**	**1.00**

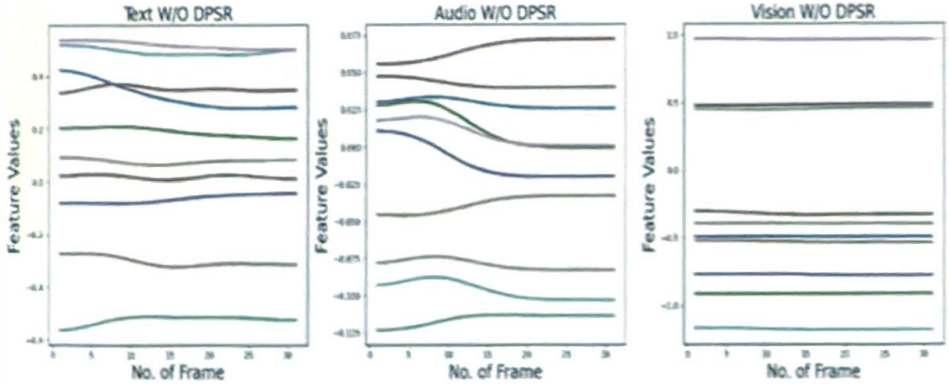
The optimal results are bolded and underlined, and the sub-optimal results are underlined.

与常用方法相比，基于情感智能体的方法（Co-SA）能准确区分正常人和抑郁症患者。

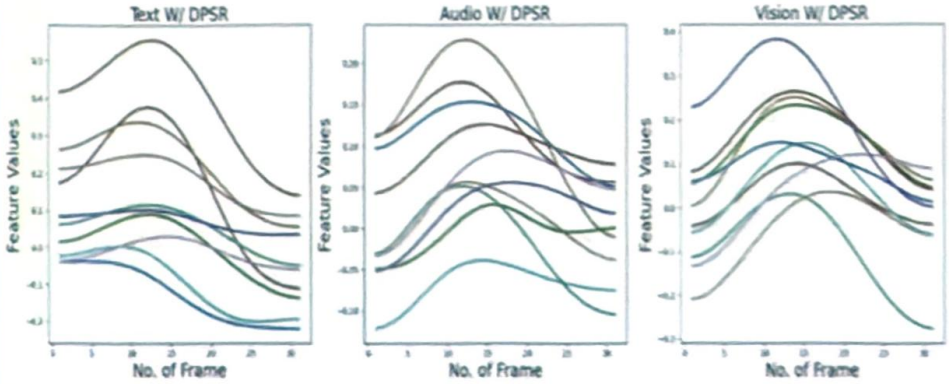
3、基于多情感智能体协同的情感计算(Under review)

■可视化分析

- 情感智能体构建阶段，成功加强了单个模态中的情感变化趋势；

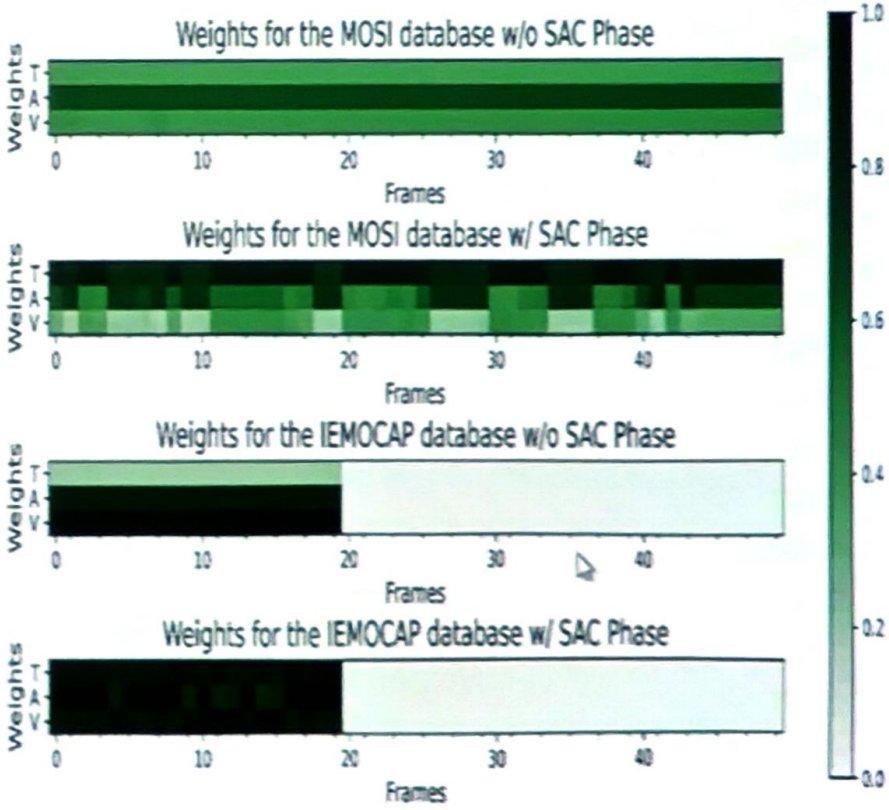


(a) 原始特征的运动趋势



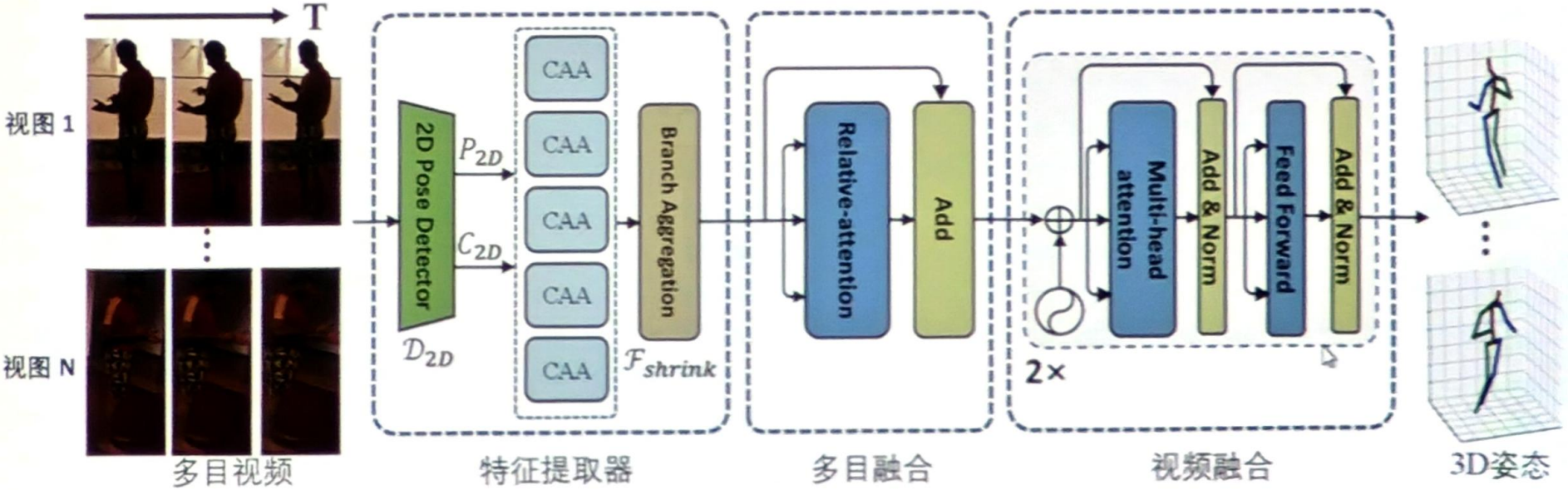
(b) 时序重构后特征的运动趋势

- 情感智能体协同阶段，成功提取每个模态的共性和互补性特征，揭示模态随时间变化的重要性；



4、自适应多视角时空Transformer3D姿态估计(T-PAMI 2023)

- 提出了一个能够兼容从图像到视频，从单目到多目的3D人体姿态估计方法。并且在多目场景下不依赖相机参数，能泛化到和训练集不同的场景中。
- 在multi-view fusing transformer中，设计了relative attention机制来学习不同视角间的相互关系，并且融合重构各个视角的特征。
- 设计Random block mask，任意丢弃视角和时序，提高模型的泛化能力。



H. Shuai, L. Wu and Q. Liu, "Adaptive Multi-view and Temporal Fusing Transformer for 3D Human Pose Estimation," in IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022

4、自适应多视角时空Transformer3D姿态估计(T-PAMI 2023)

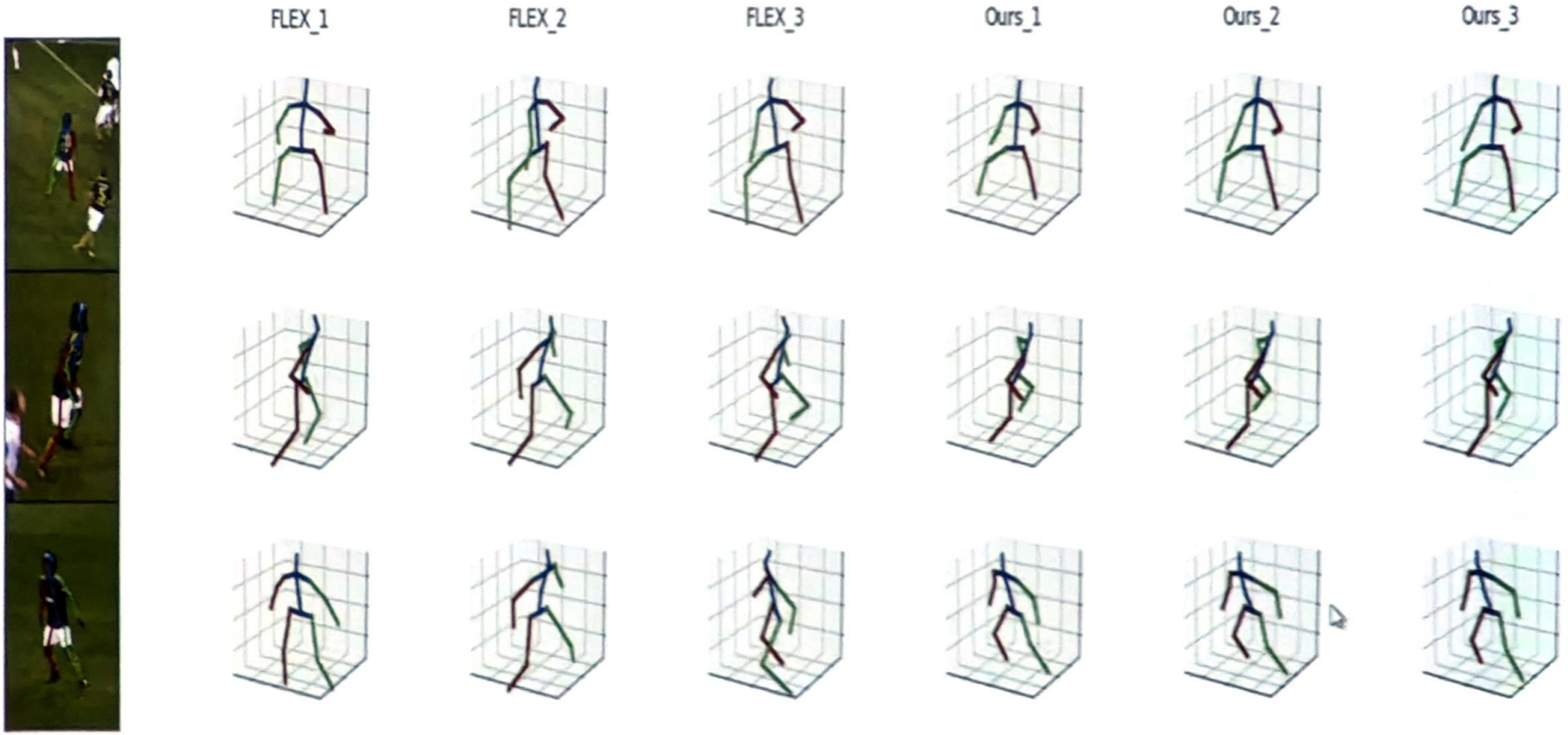
Human36M数据集（室内3D姿态数据集，360万帧）上性能对比

	Dir.	Disc.	Eat.	Greet	Phone	Photo	Pose	Purch.	Sit.	SitD.	Smoke	Wait	WalkD.	Walk	WalkT.	Avg
Monocular methods																
Pavllo et al. [16]	45.2	46.7	43.3	45.6	48.1	55.1	44.6	44.3	57.3	65.8	47.1	44.0	49.0	32.8	33.9	46.8
Chen et al. [58]	43.8	48.6	49.1	49.8	57.6	61.5	45.9	48.3	62.0	73.4	54.8	50.6	56.0	43.4	45.5	52.7
Liu et al. [59]	41.8	44.8	41.1	44.9	47.4	54.1	43.4	42.2	56.2	63.6	45.3	43.5	45.3	31.3	32.2	45.1
Wang et al. [18]	40.2	42.5	42.6	41.1	46.7	56.7	41.4	42.3	56.2	60.4	46.3	42.2	46.2	31.7	31.0	44.5
Zeng et al. [60]	46.6	47.1	43.9	41.6	45.8	49.6	46.5	40.0	53.4	61.1	46.1	42.6	43.1	31.5	32.6	44.8
Multi-view methods with camera parameters																
Pavlakos et al. [61]	41.2	49.2	42.8	43.4	55.6	46.9	40.3	63.7	97.6	119	52.1	42.7	51.9	41.8	39.4	56.9
Qiu et al. [24]	24.0	26.7	23.2	24.3	24.8	22.8	24.1	28.6	32.1	26.9	31.0	25.6	25.0	28.0	24.4	26.2
Iskakov et al. [22]	19.9	20.0	18.9	18.5	20.5	19.4	18.4	22.1	22.5	28.7	21.2	20.8	19.7	22.1	20.2	20.8
He et al.(IMU) [21]	25.7	27.7	23.7	24.8	26.9	31.4	24.9	26.5	28.8	31.7	28.2	26.4	23.6	28.3	23.5	26.9
Zhang et al. [23]	17.8	19.5	17.6	20.7	19.3	16.8	18.9	20.2	25.7	20.1	19.2	20.5	17.2	20.5	17.3	19.5
Zhang et al. [62]	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	21.7
Remeli et al. [63]	27.3	32.1	25.0	26.5	29.3	35.4	28.8	31.6	36.4	31.7	31.2	29.9	26.9	33.7	30.4	30.2
MTF-Transformer+ (CPN,$T = 1$)	23.8	26.0	23.9	25.0	28.2	29.7	23.6	25.5	30.1	37.3	26.6	24.5	27.4	23.1	23.4	26.5
MTF-Transformer+ (CPN,$T = 27$)	23.4	25.2	23.1	24.4	27.4	28.5	22.8	25.2	28.7	36.2	25.9	23.6	26.6	22.6	22.7	25.8
Multi-view methods without camera parameters																
Huang et al. [26]	26.8	32.0	25.6	52.1	33.3	42.3	25.8	25.9	40.5	76.6	39.1	54.5	35.9	25.1	24.2	37.5
Gordon et al. [46] (based on Iskakov et al. [22])	**23.1**	28.8	26.8	28.1	31.6	37.1	25.7	31.4	36.5	39.6	35.0	29.5	35.6	26.8	26.4	30.9
Gordon et al. [46](CPN,$T = 27$)	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	31.7
MTF-Transformer (CPN,$T = 1$)	24.2	26.4	26.1	25.6	29.4	29.7	25.1	25.4	32.4	37.4	27.1	25.4	29.5	23.8	24.4	27.5
MTF-Transformer (CPN,$T = 27$)	**23.1**	**25.4**	**24.7**	**24.5**	**27.9**	**28.3**	**23.9**	**24.6**	**30.7**	**35.7**	**25.8**	**24.2**	**28.4**	**22.8**	**23.1**	**26.2**
Gordon et al. [46](GT, $T = 27$)	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	22.9
MTF-Transformer (GT,$T = 27$)	15.5	17.1	13.7	15.5	14.0	16.2	15.8	16.5	15.8	16.1	14.5	14.5	16.9	14.3	13.7	15.3

H. Shuai, L. Wu and Q. Liu, "Adaptive Multi-view and Temporal Fusing Transformer for 3D Human Pose Estimation," in IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022

4、自适应多视角时空Transformer3D姿态估计(T-PAMI 2023)

KTH Multiview Football II上可视化结果（在Human36M训练）



H. Shuai, L. Wu and Q. Liu, "Adaptive Multi-view and Temporal Fusing Transformer for 3D Human Pose Estimation," in IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022

汇报提纲

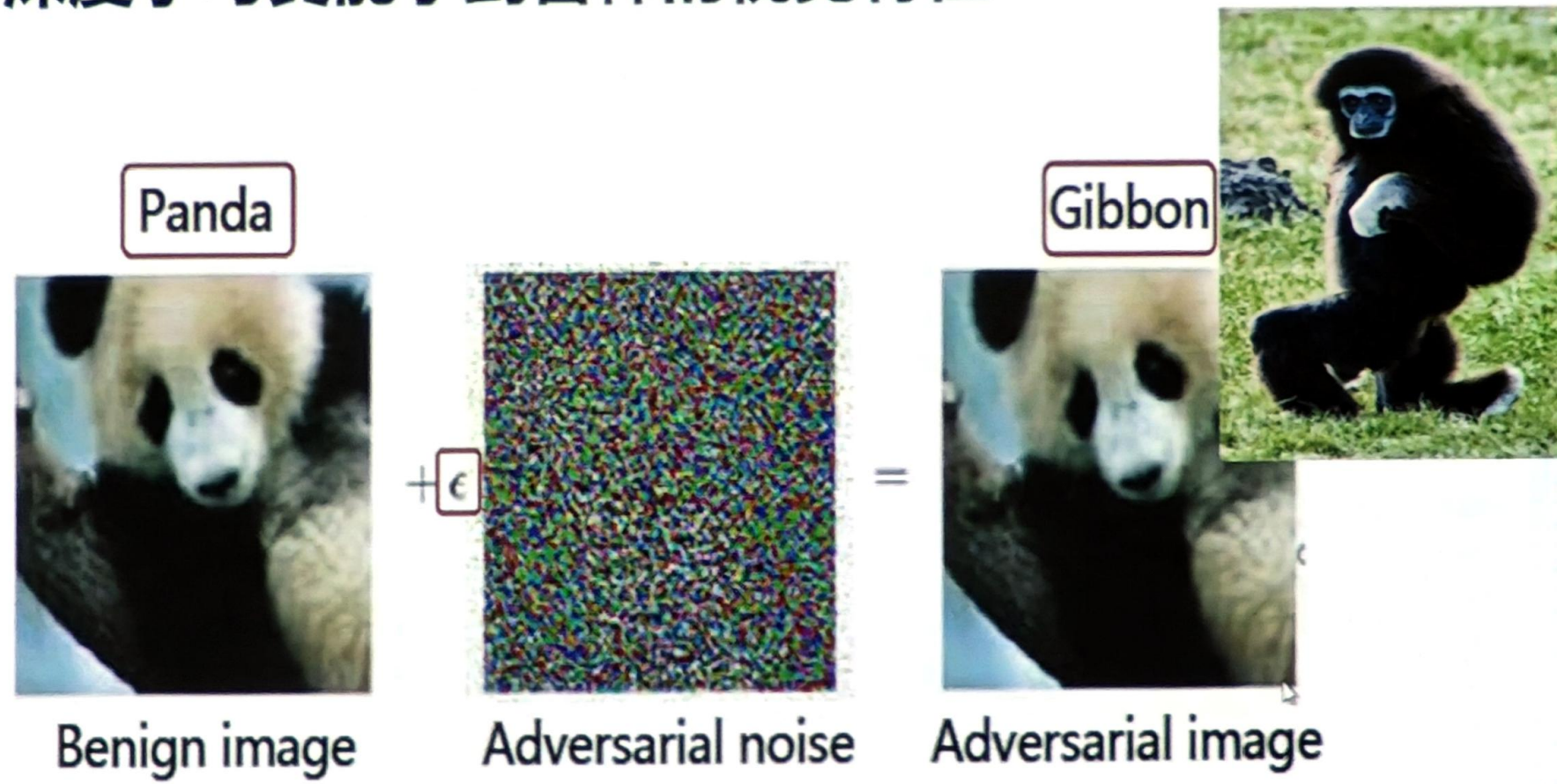
1 背景介绍

2 研究工作

3 总结思考

问题与思考

- 深度学习真能学到鲁棒的视觉特征?



------Sample from ICLR 2015

问题与思考

- 深度学习真的和人眼视觉学习一样吗?