Architect
SACC
2022 中国系统架构师大会
SYSTEM ARCHITECT CONFERENCE CHINA 2022
激发架构性能 点亮业务活力
云上会议 网络直播 | 2022年10月27-29日
ITT168.com
ChinaUnix.net
ITPUB

多模态认知计算介绍及实践

智精灵 CTO 蔡龙军（牧己）

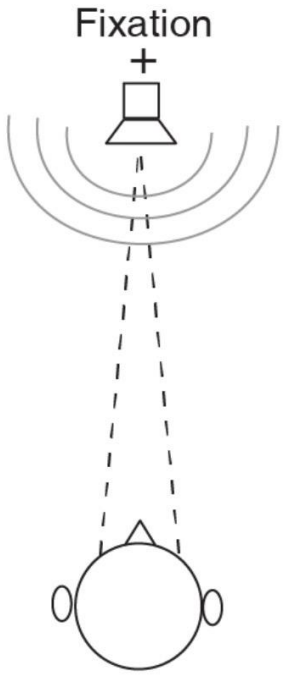
Agenda

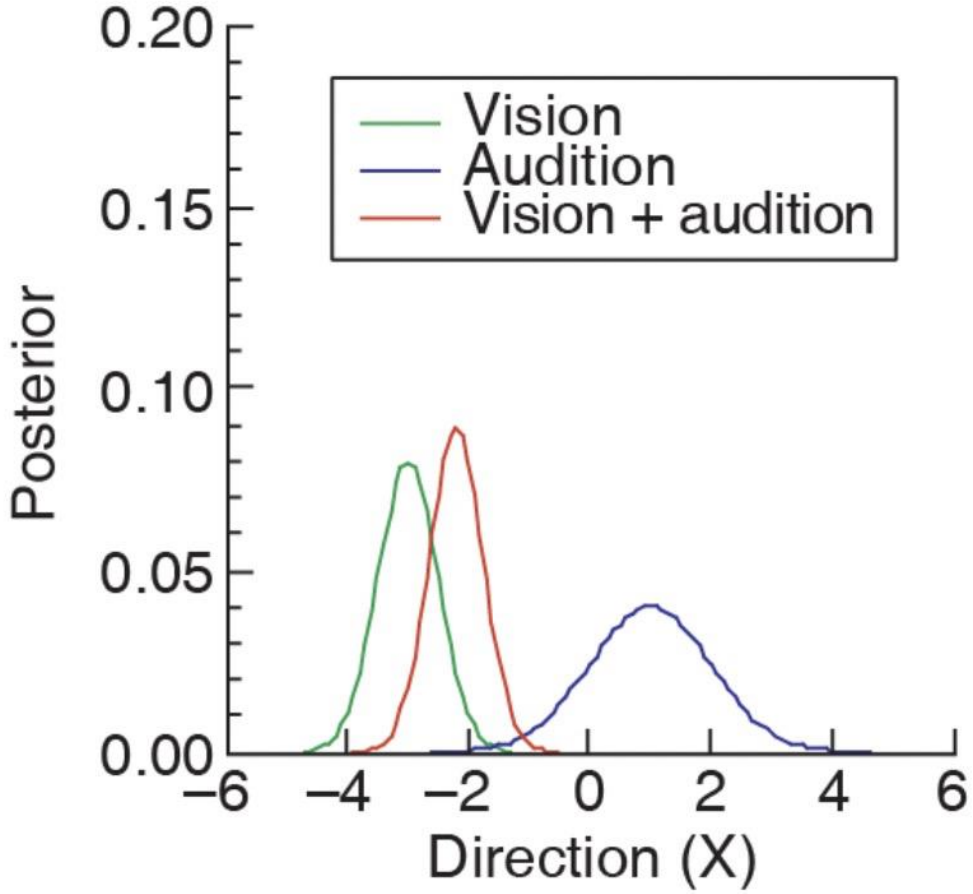
- 认知能力基本定义
- 人类大脑多模态信息整合机制
  - 贝叶斯大脑
  - 神经编码机制
  - 信息整合与分离机制
- 多模态认知计算
  - 多模态认知计算定义
  - 基本框架及AI进展介绍
- 类脑SNN与ANN融合新趋势
  - 现有融合方式介绍
- 结论

AI认知能力

Cognitive Computing breaks the traditional boundary between neuroscience and computer science, and paves the way for machines that will have reasoning abilities analogous to a human brain.

大脑信息整合（贝叶斯脑）





$$p(s \mid x_V, x_A) \propto p(x_V \mid s)p(x_A \mid s)p(s)$$

$$p(x_m \mid s) = N[x_m \mid s, \sigma_m^2], (m = V, A)$$

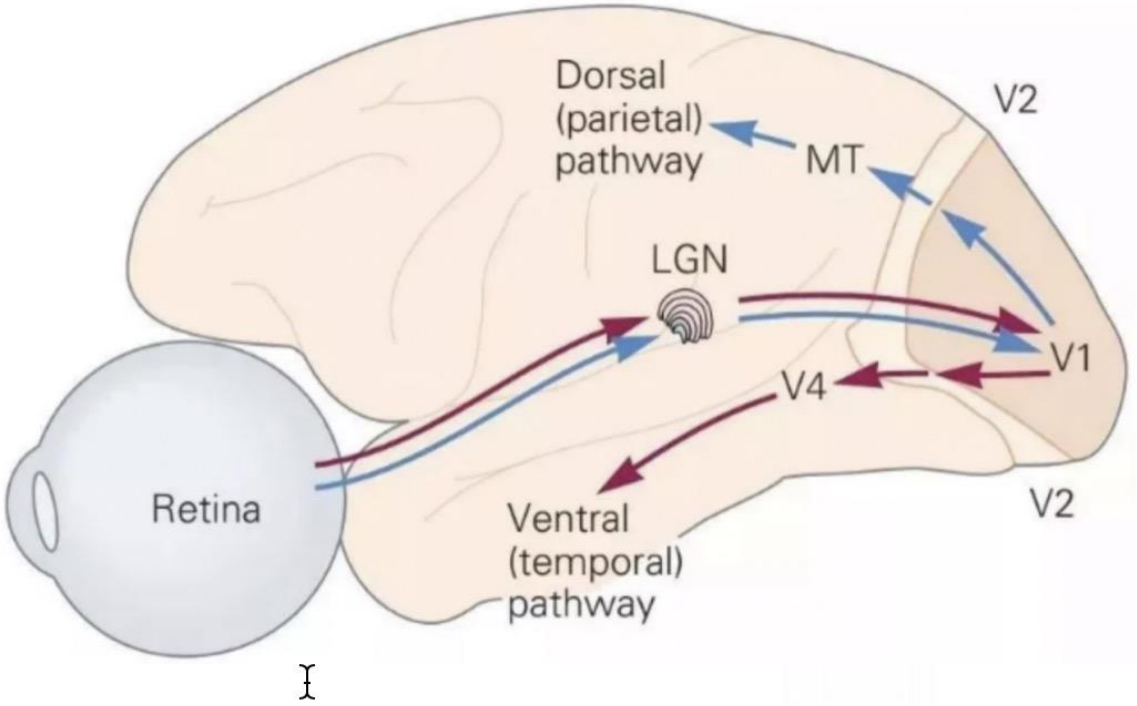
$$V(\hat{s})^{-1} = \sigma_V^{-2} + \sigma_A^{-2} \quad (3)$$

$$\hat{s} = V(\hat{s})(\sigma_V^{-2} x_V + \sigma_A^{-2} x_A) \quad (4)$$

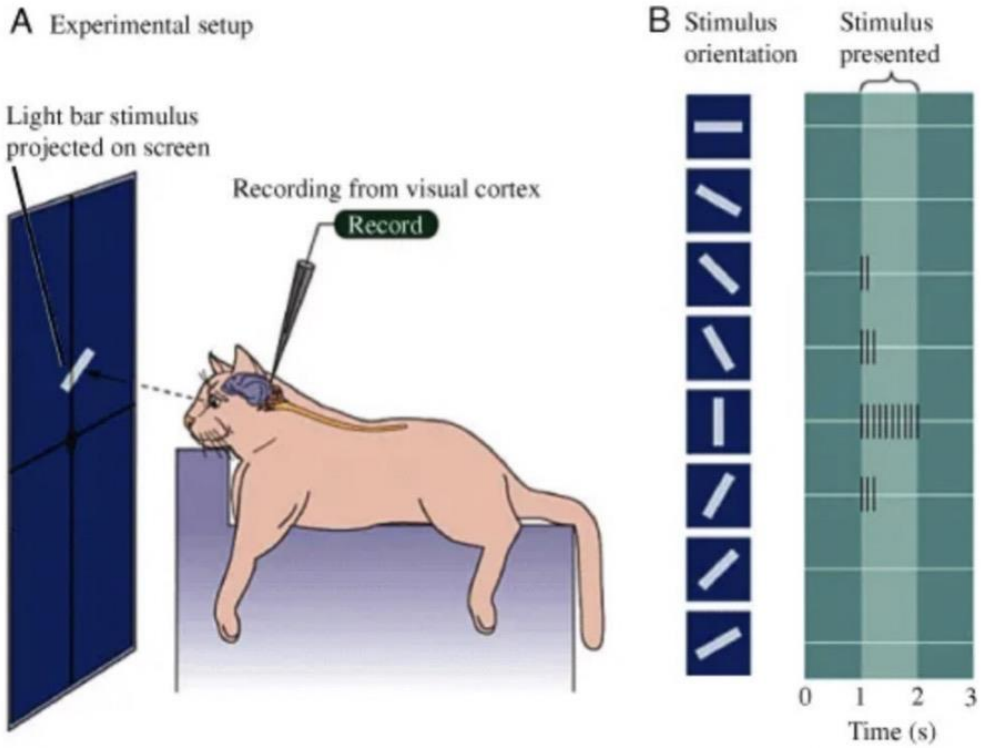
$$\sigma_C^2 = \frac{\sigma_A^2 \sigma_B^2}{\sigma_A^2 + \sigma_B^2} \qquad \sigma_C < \sigma_A < \sigma_B$$

A. Pouget, J. M. Beck, W. J. Ma, and P. E. Latham, “Probabilistic brains: knowns and unknowns.,” Nat. Neurosci., vol. 16, no. 9, pp. 1170–8, 2013.
https://mp.weixin.qq.com/s/cAq5R972li7R2260nViFMg

神经元编码--V1视觉系统具有方位选择性

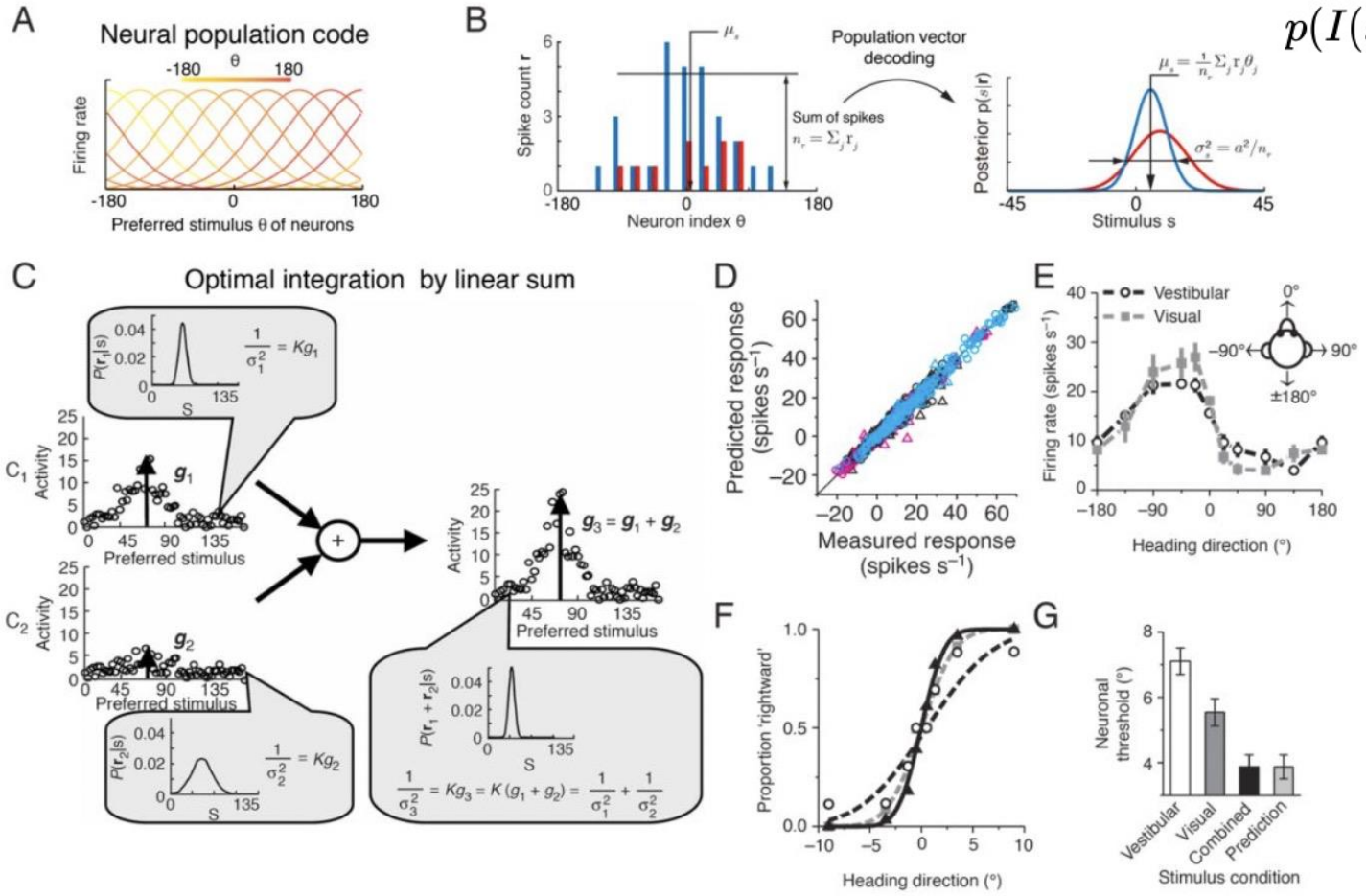


人类视觉系统



Hubel-Weisel实验

神经元编码--多模态整合机制验证



$$p(I(x)|s) = \text{Poisson}(\lambda(s)),\ \lambda(s) = \exp[(x-s)^2/2a^2]$$

$$p(O|s) \propto p(I|s) = \mathcal{N}(s|s^o, \Lambda^{-1}),$$
$$s^o = \int xI\mathrm{d}x / \int I\mathrm{d}x,\ \ \Lambda = a^{-2}\int I\mathrm{d}x,$$

$$p(s|O) \propto p(O|s)p(s) = \mathcal{N}(s|\kappa, \Omega^{-1}),$$
$$\kappa = \Omega^{-1}(\Lambda s^o + L\mu),$$
$$\Omega = \Lambda + L,$$

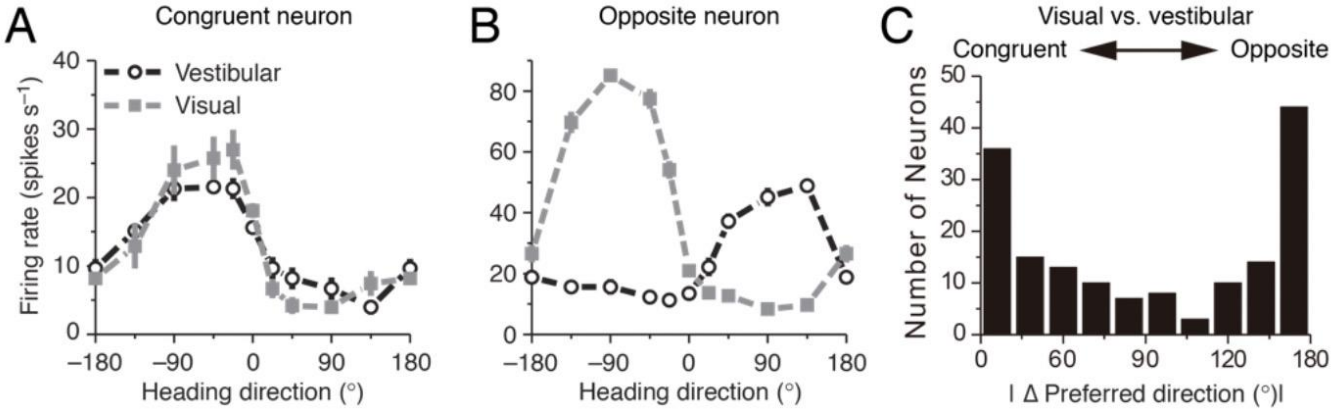
$$f_i(s) = R * e^{-(s-\theta_i)^2/2a^2}$$

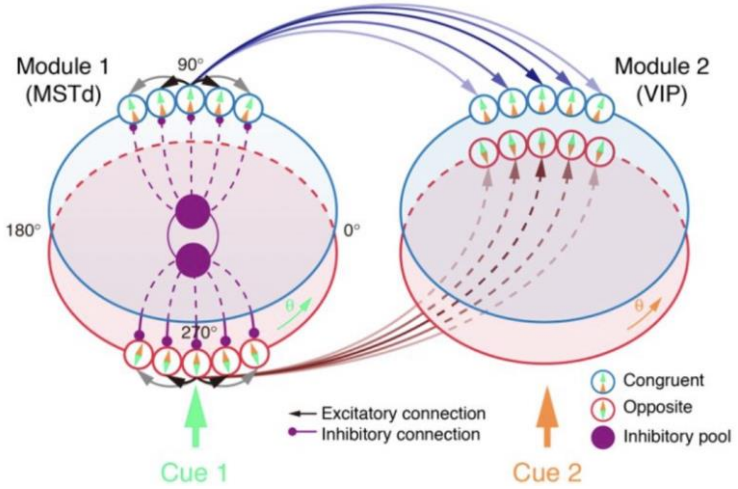
$$p(r \mid s) = \prod_{i=1}^{N} p(r_i \mid s) = \prod_{i=1}^{N} f_i(s)^{r_i}/r_i! * e^{-f_i(s)}$$

$$\hat{s} = \sum_{i=1}^{N} r_i\theta_i / \sum_{i=1}^{N} r_i, \quad V(\hat{s}) = (\sum_{i=1}^{N} r_i/a^2)^{-1}$$

https://mp.weixin.qq.com/s/cAq5R972li7R2260nViFMg

信息整合与分离机制





$$p_d(s_1 \mid x_1, x_2) \propto \frac{p(x_1 \mid s_1)}{p(x_2 \mid s_1)}$$

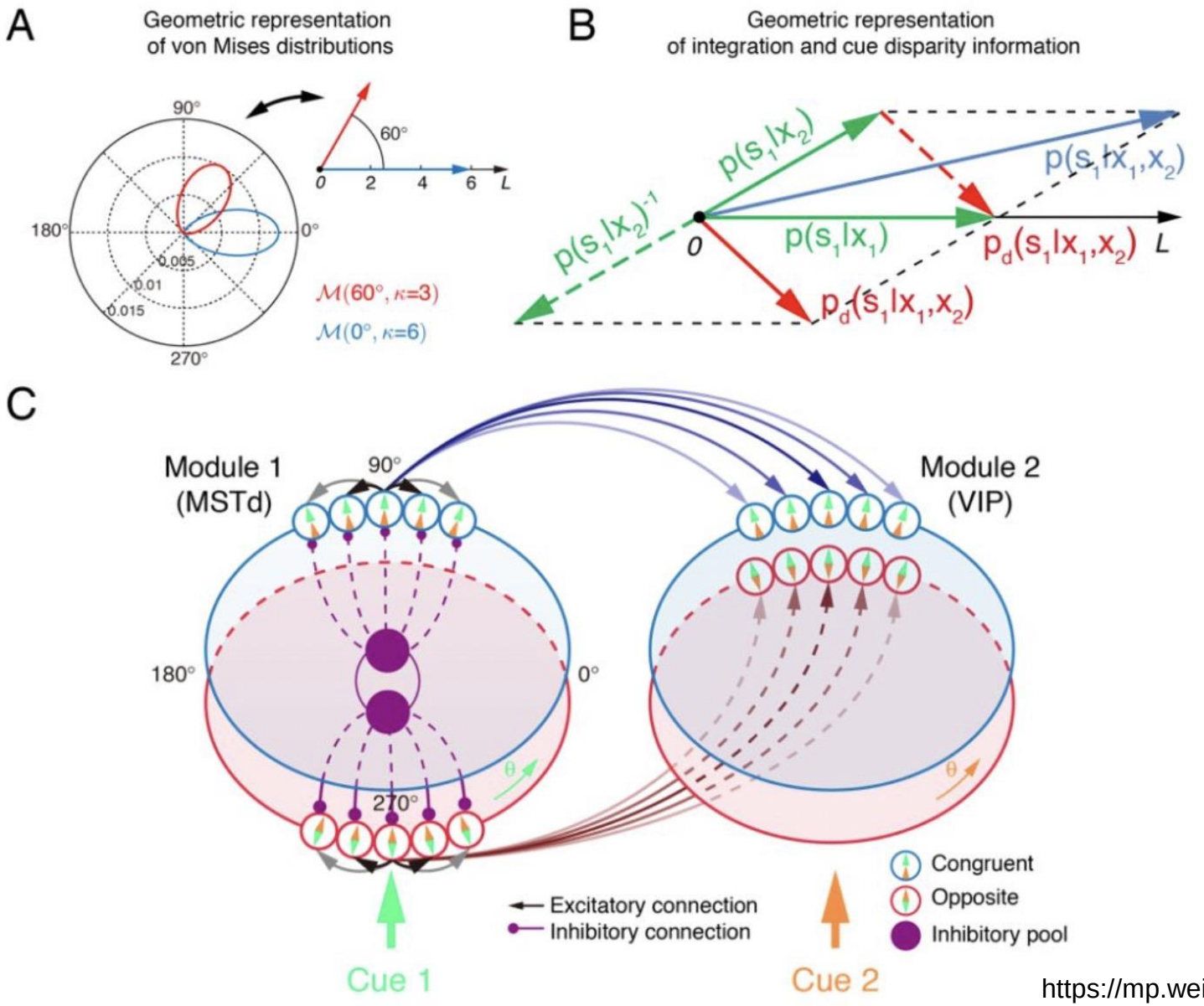
$$p_d(s_1 \mid x_1, x_2) \propto p(x_1 \mid s_1)p(x_2 + \pi \mid s_1)$$

Von Mises：

$$p(x_m \mid s_1) \propto exp[\kappa_m cos(x_m - s_1)], (m = 1, 2)$$

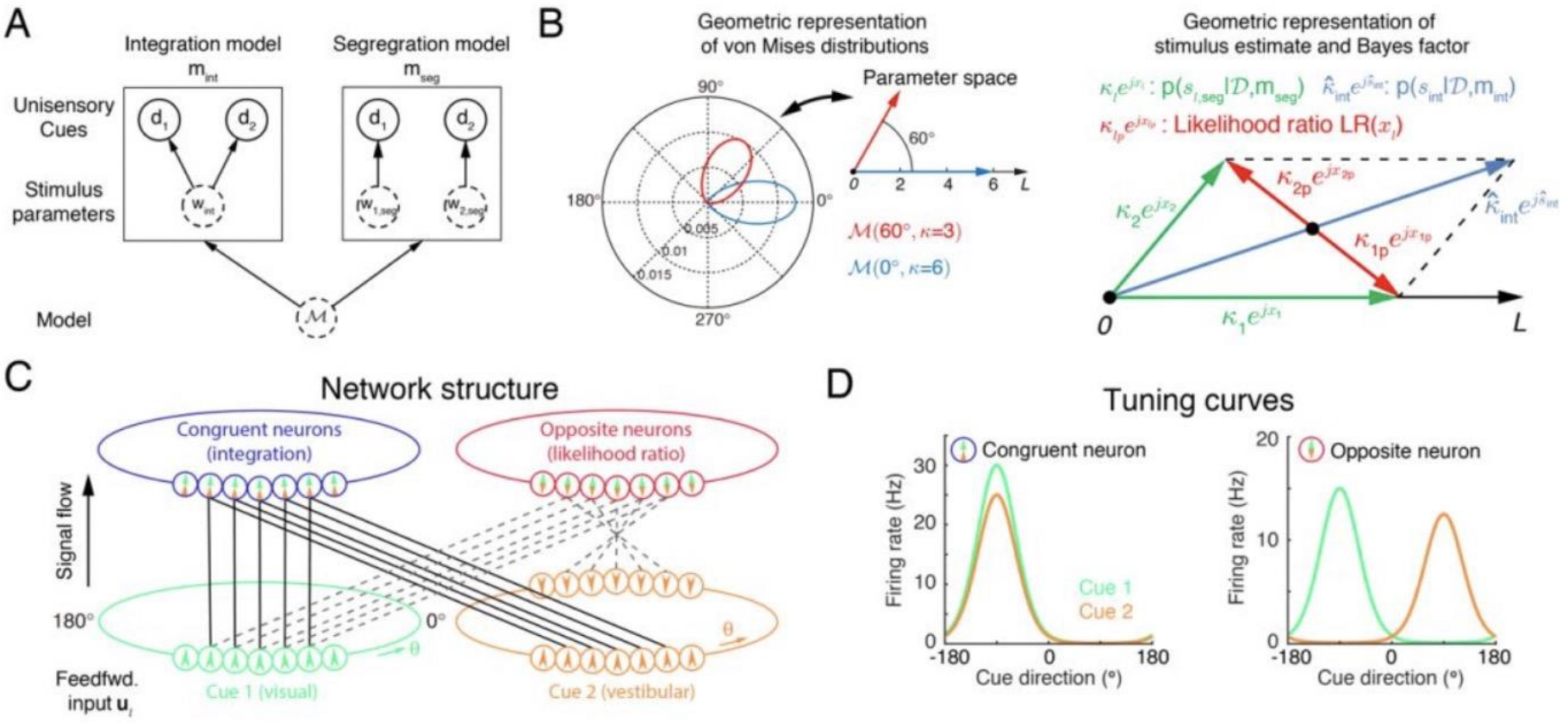
https://mp.weixin.qq.com/s/cAq5R972li7R2260nViFMg

信息整合与分离机制



https://mp.weixin.qq.com/s/cAq5R972li7R2260nViFMg

信息整合与分离因果推断



$$p(m_{int}|d) = [1 + \frac{p(d|m_{seg})}{p(d|m_{int})}]^{-1}$$

https://mp.weixin.qq.com/s/cAq5R972li7R2260nViFMg

多模态认知计算

熵

$$h(x) = -\log p(x)$$

信息量

$$h(x,T) = -\log p(x|_T) \quad X \in \mathbb{R}^{m\times s\times t}$$

信息量最大化

$$K^* = \max_{||A||_1=1} \left\| A \odot \sum_{i=1}^{m} I_i \right\|$$

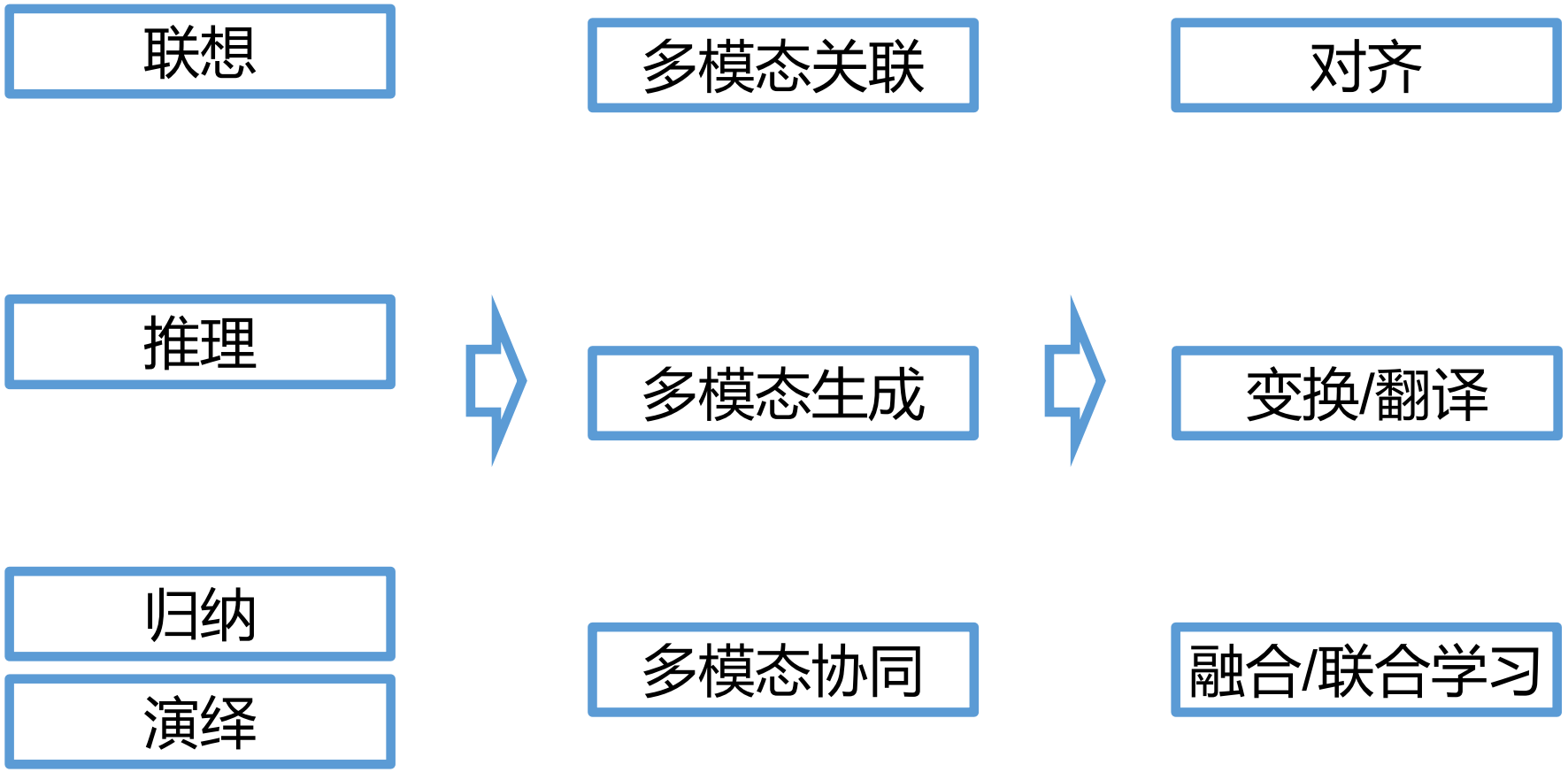
信容

$$\rho = \frac{K^*}{D} = \frac{\max_{||A||_1=1} \left\| A \odot \sum_{i=1}^{m} I_i \right\|}{D}$$

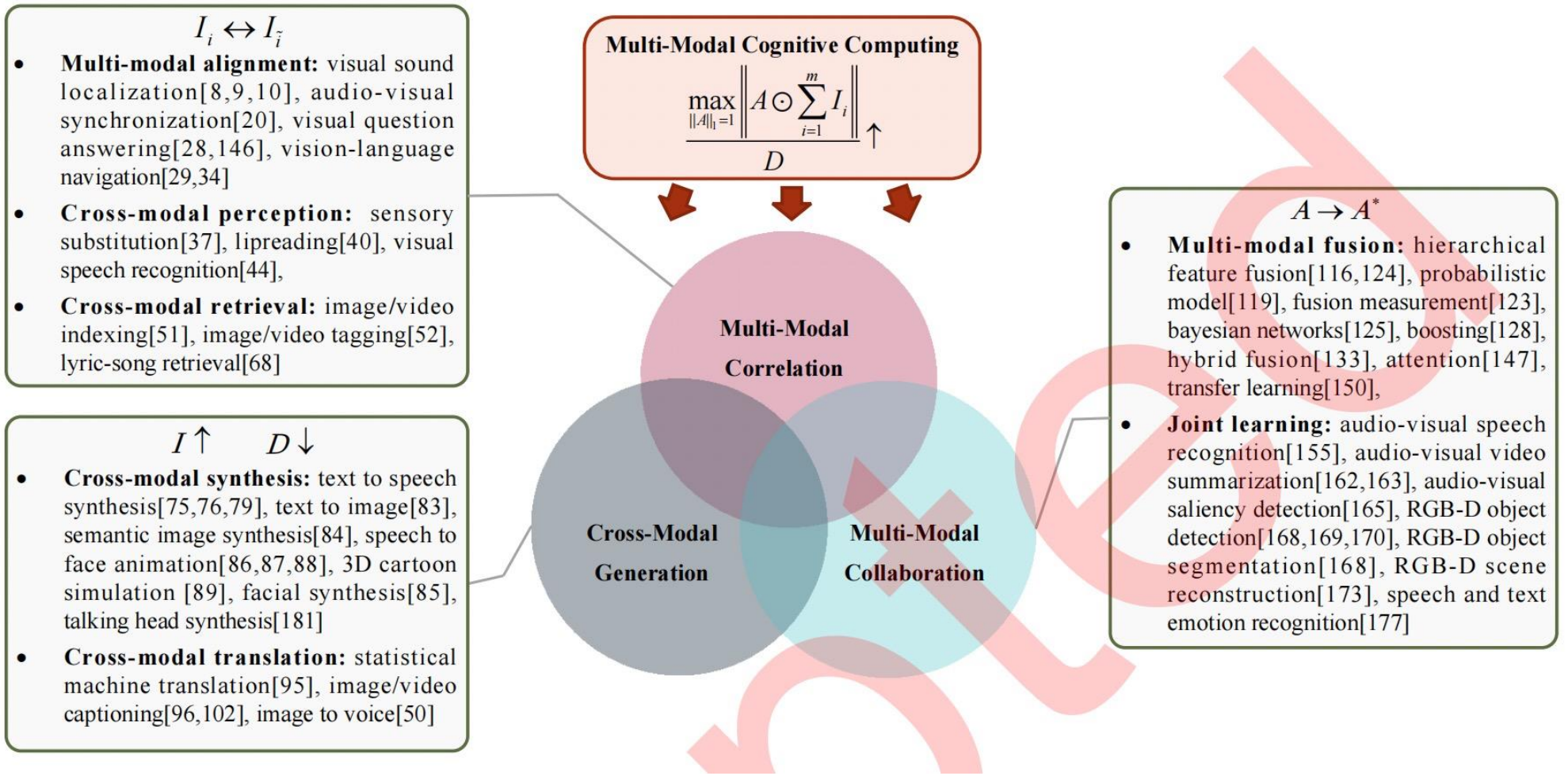
信容（Information Capacity）:从单位数据获取最大信息量的能力

Xuelong Li. Multi-Modal Cognitive Computing (in Chinese). Sci Sin Inform

认知能力与多模态认知计算间关系

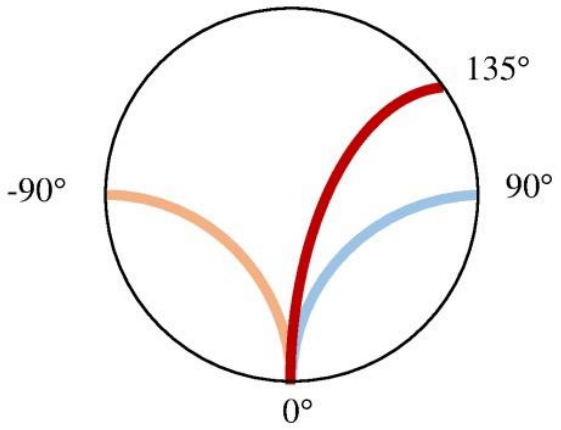


多模态认知计算架构关系图

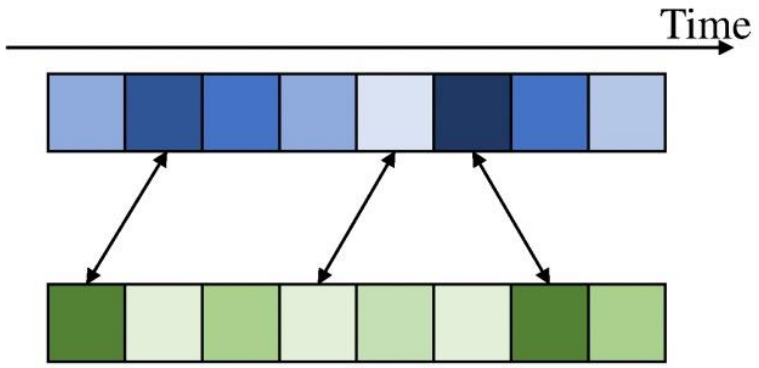


Xuelong Li. Multi-Modal Cognitive Computing (in Chinese). Sci Sin Inform

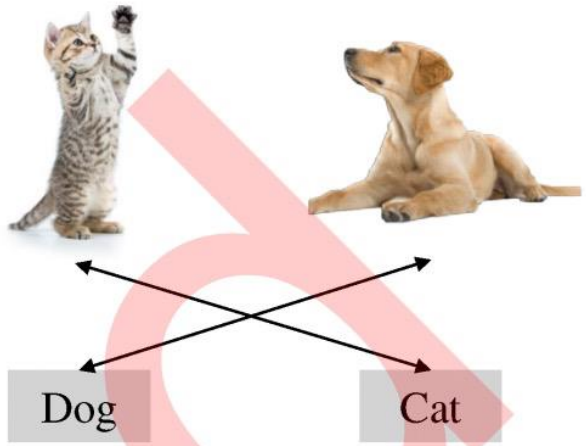
多模态对齐



Spatial alignment



Temporal alignment

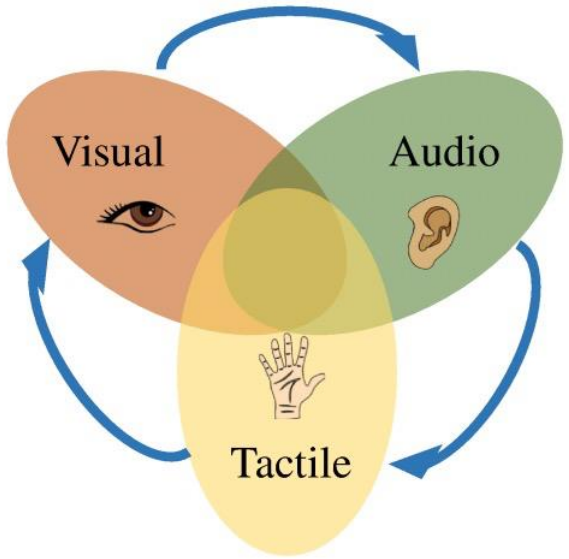


Semantic alignment

图 2　多模态对齐示意图.

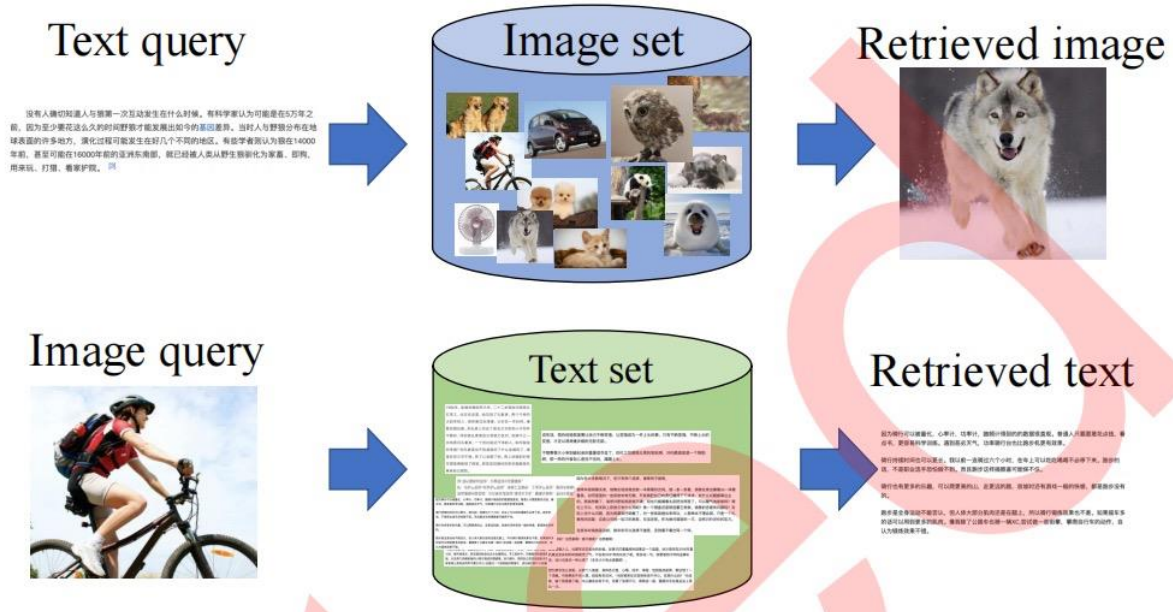
Xuelong Li. Multi-Modal Cognitive Computing (in Chinese). Sci Sin Inform

多模态感知与检索



Multi-modal perception

(a)



Multi-modal retrieval

(b)

图 3 (a) 多模态感知与 (b) 多模态检索示意图.

Xuelong Li. Multi-Modal Cognitive Computing (in Chinese). Sci Sin Inform

多模态生成

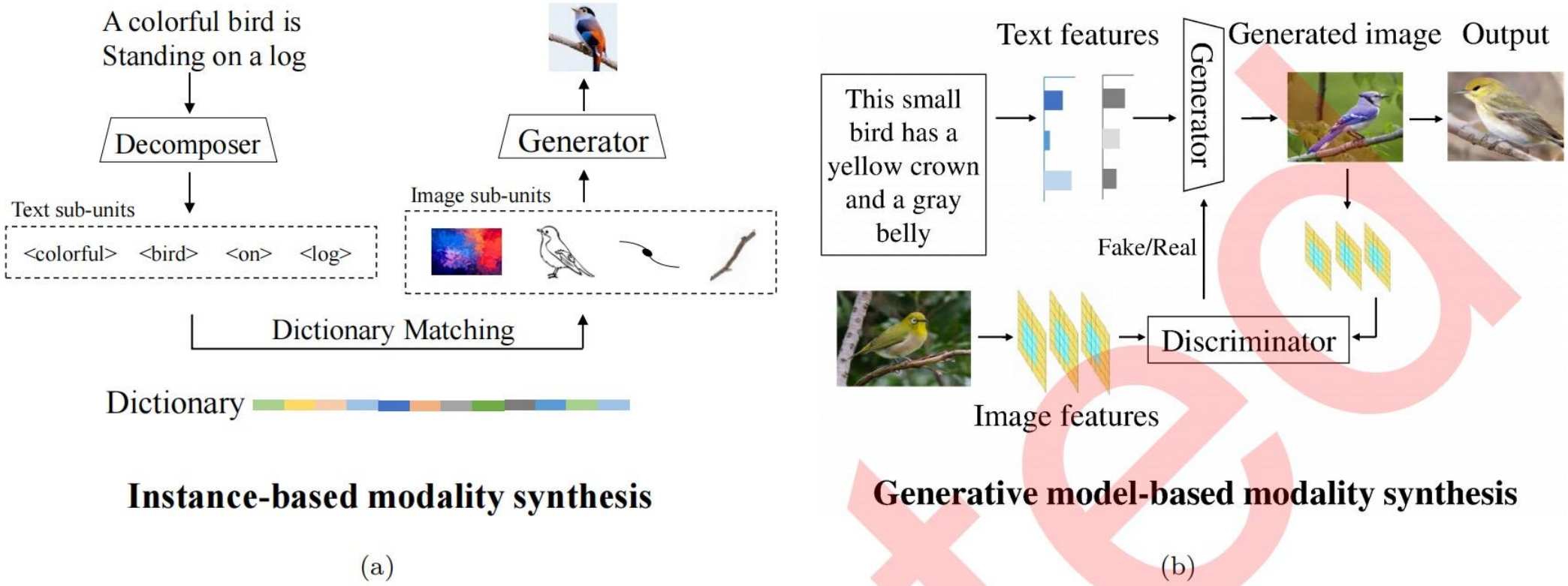


图 4　典型的两种跨模态合成方式.

Xuelong Li. Multi-Modal Cognitive Computing (in Chinese). Sci Sin Inform

多模态融合

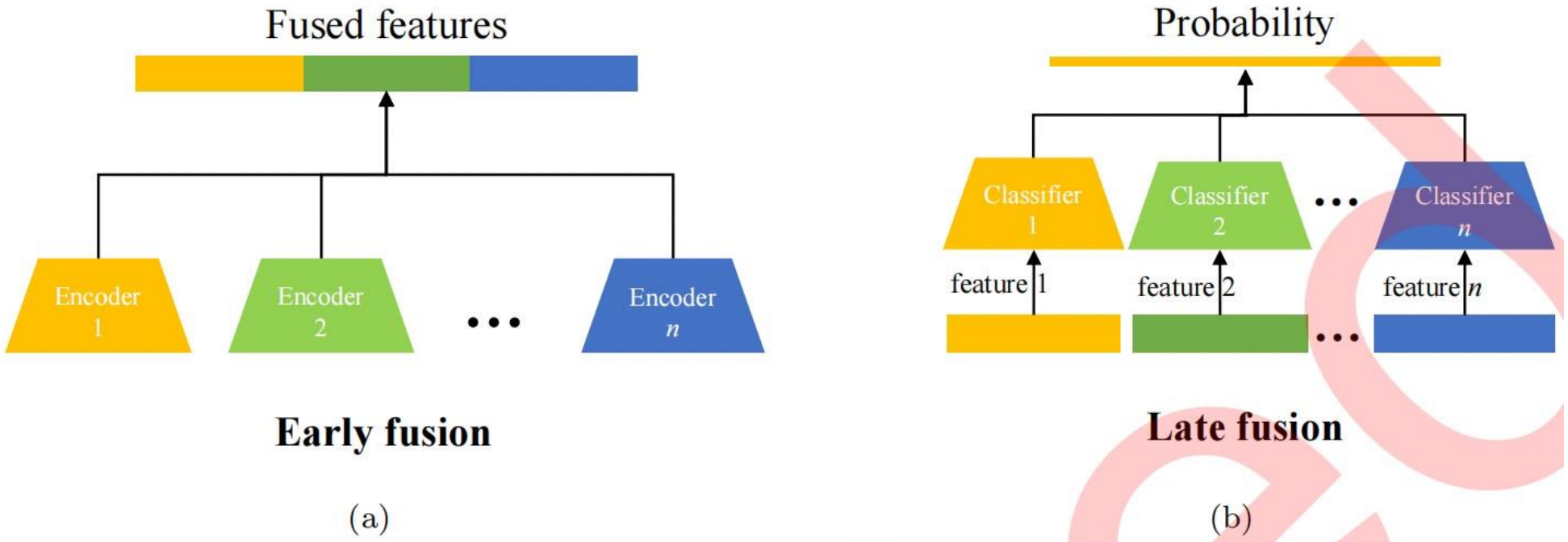


图 6　多模态融合中的 (a) 前期融合与 (b) 后期融合示意图.

Xuelong Li. Multi-Modal Cognitive Computing (in Chinese). Sci Sin Inform

多模态协同--联合学习

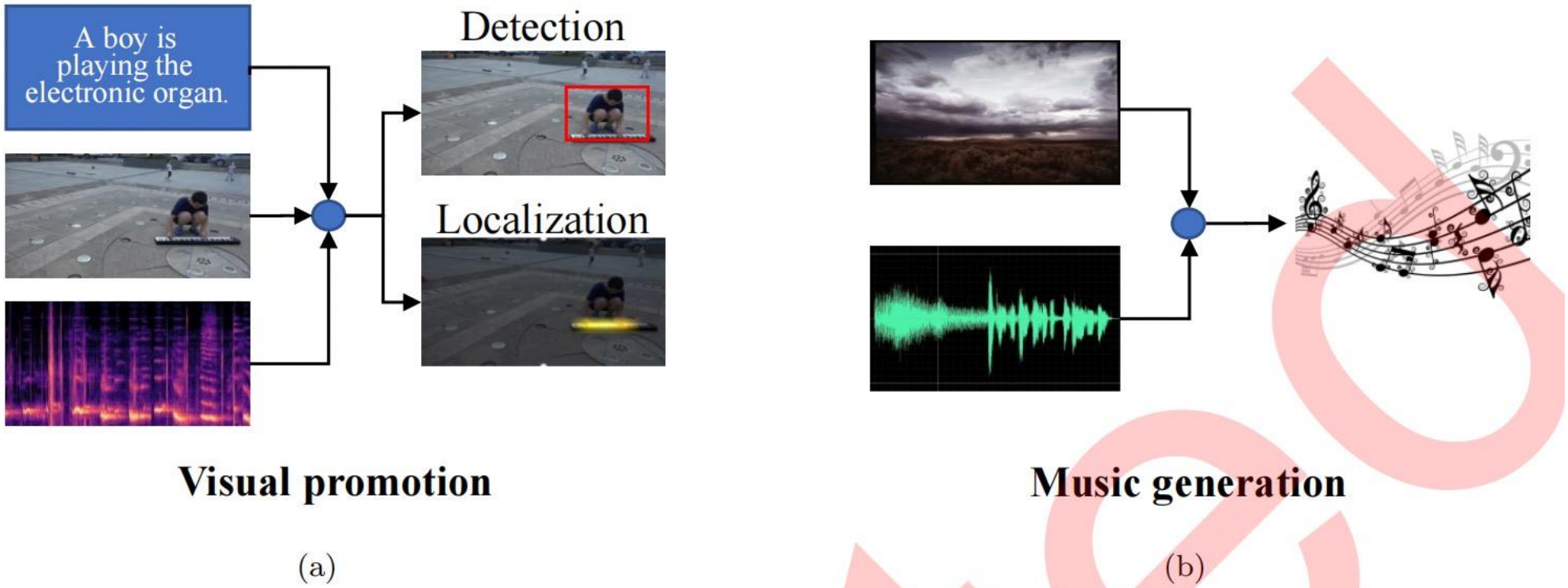
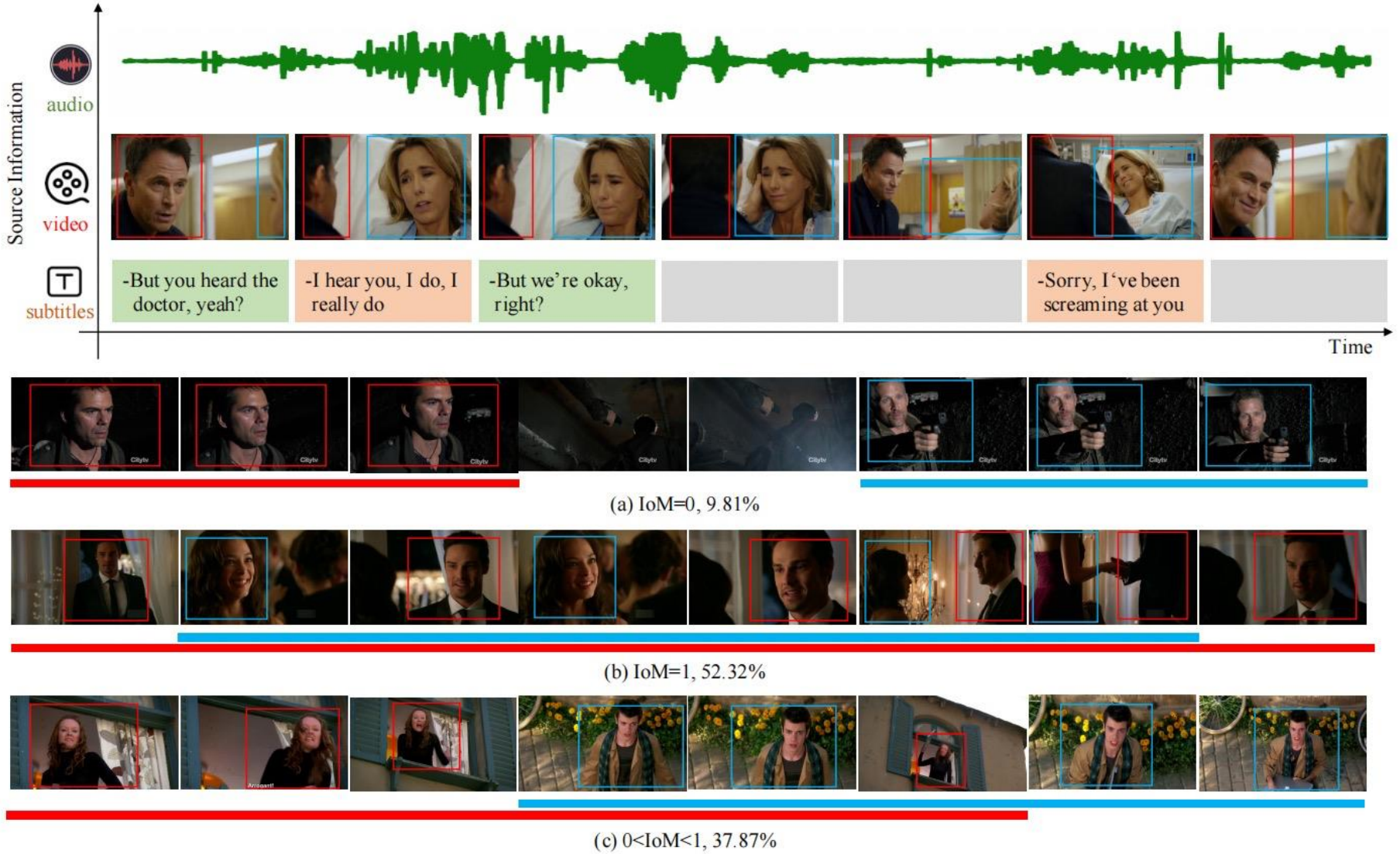


图 7　多模态联合学习中的 (a) 模态性能提升与 (b) 模态创新应用示例.

Xuelong Li. Multi-Modal Cognitive Computing (in Chinese). Sci Sin Inform

基于多模态协同的认知计算：

Pairwise Emotional Relationship Recognition



Xun Gao; Yin Zhao; Jie Zhang; Longjun Cai ; Pairwise Emotional Relationship Recognition i n Drama Videos: Dataset and Benchmark

多模态协同的认知计算

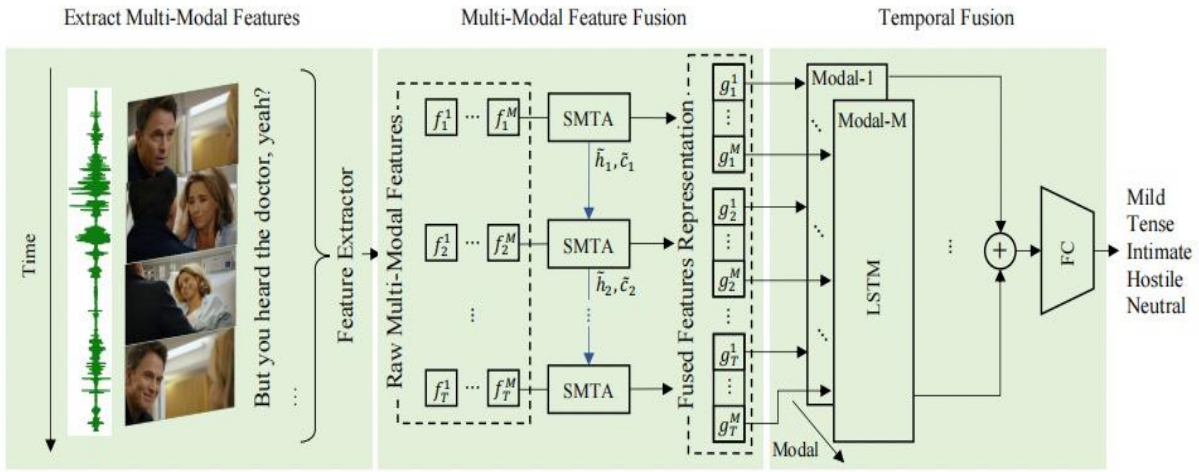
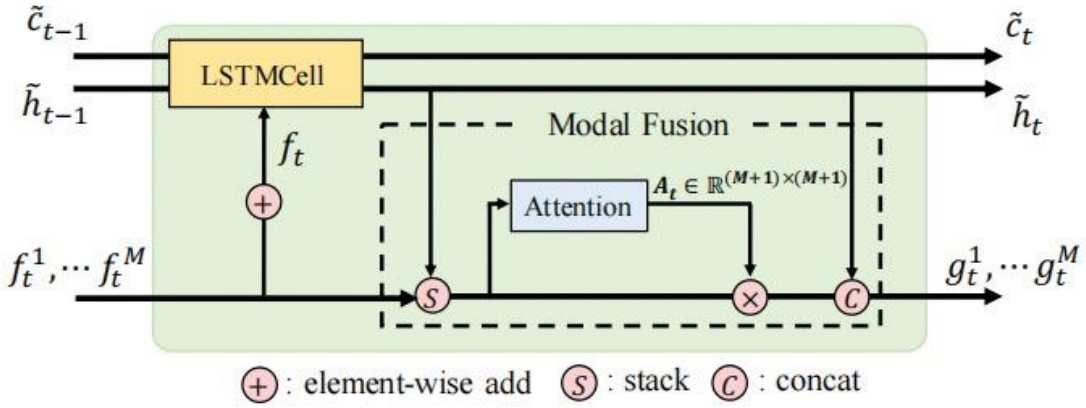


Figure 4: Overall framework of our proposed method. The structure of STMA is shown in Fig. 5.



Attention-based Multi-Modal Fusion

$$\min \quad \mathcal{L}(\hat{y}, y) + \sum_{m=1}^{M} \mathcal{L}(\hat{y}^m, y)$$

$$\hat{y} = softmax(FC(\sum_{m} h_T^m))$$

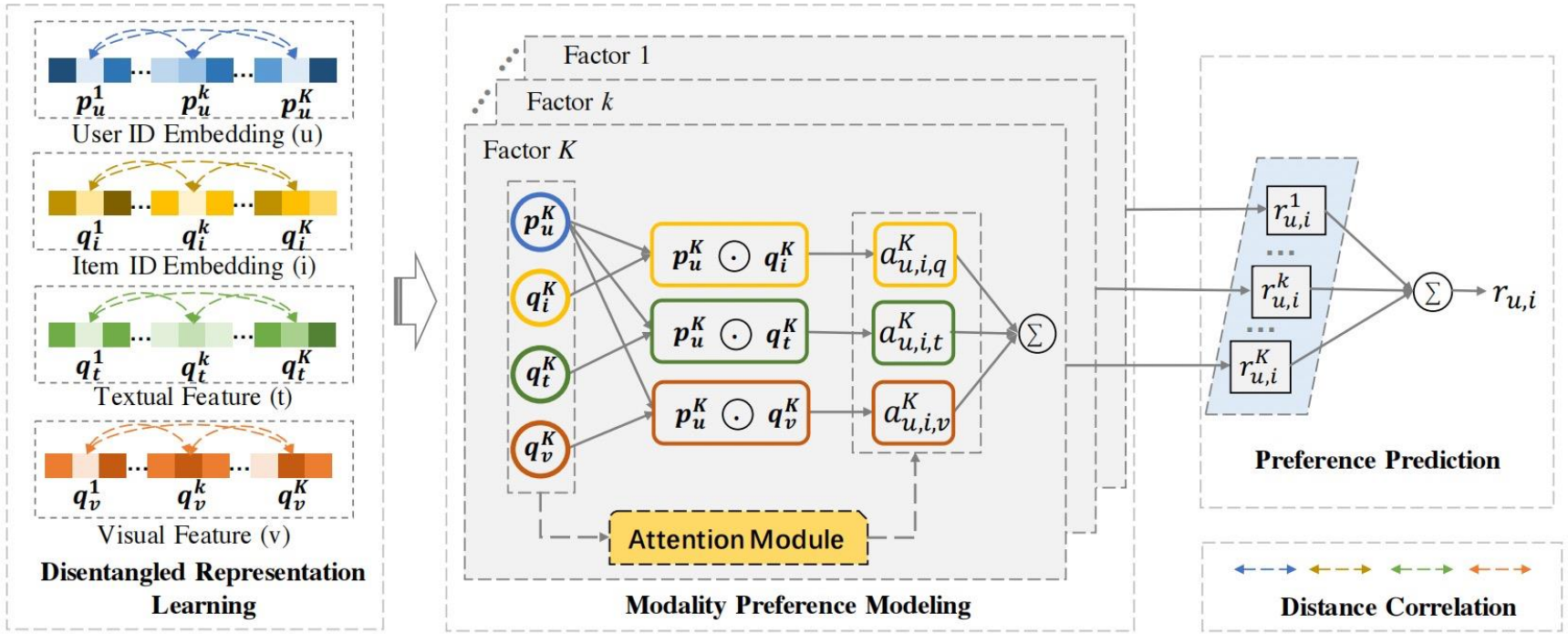
$$\hat{y}^m = softmax(FC(h_T^m))$$

Joint Learning

Method	5 Category		3 Category	
	Micro-F1	Macro-F1	Micro-F1	Macro-F1
LSTM	63.80	44.43	64.98	56.23
Dot-product [63]	64.35	45.80	66.55	58.89
Non-local [62]	63.72	47.16	67.49	59.74
Multi-head [60]	64.11	47.15	67.30	60.20
Transformer [60]	64.82	48.90	68.59	58.98
$SMTA_{Dot\text{-}product}$	65.45	47.92	69.18	60.20
$SMTA_{Non\text{-}local}$	65.33	48.00	69.66	60.80
$SMTA_{Multi\text{-}head}$	65.60	47.71	68.95	61.35
$SMTA_{Transformer}$	**66.55**	**49.81**	**70.05**	**61.73**

Xun Gao; Yin Zhao; Jie Zhang; Longjun Cai ; Pairwise Emotional Relationship Recognition i n Drama Videos: Dataset and Benchmark

Disentangled Representation Learning

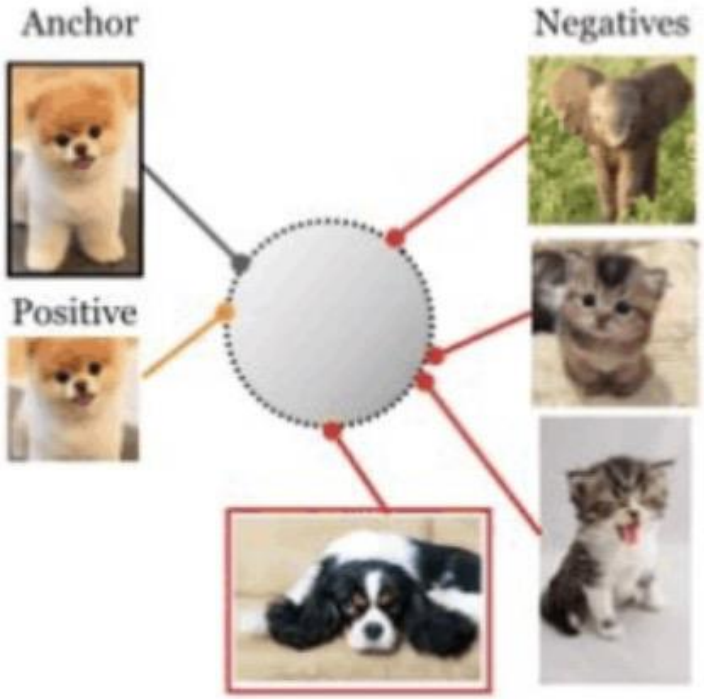


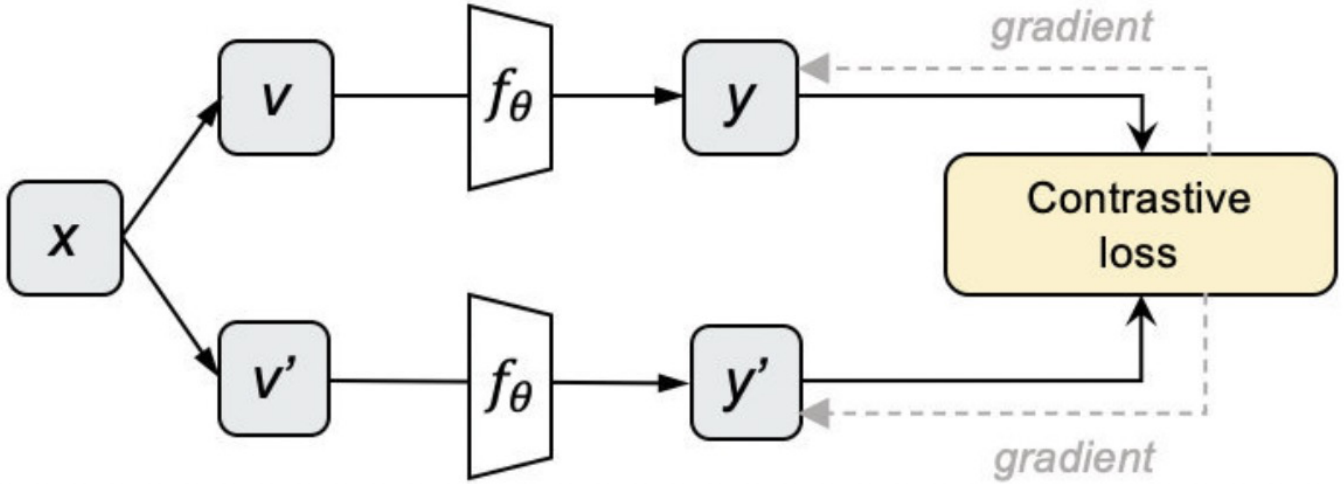
$$L_y = \sum_{k=1}^{K} \sum_{k'=k+1}^{K} dCor\left(\boldsymbol{y}^k, \boldsymbol{y}^{k'}\right)$$

$$dCor\left(\boldsymbol{y}^k, \boldsymbol{y}^{k'}\right) = \frac{dCov\left(\boldsymbol{y}^k, \boldsymbol{y}^{k'}\right)}{\sqrt{dVar\left(\boldsymbol{y}^k\right) dVar\left(\boldsymbol{y}^{k'}\right)}},$$

Fan Liu, Zhiyong Cheng, Huilin Chen, Anan Liu, Liqiang Nie ; Disentangled Multimodal Representation Learning for Recommendation

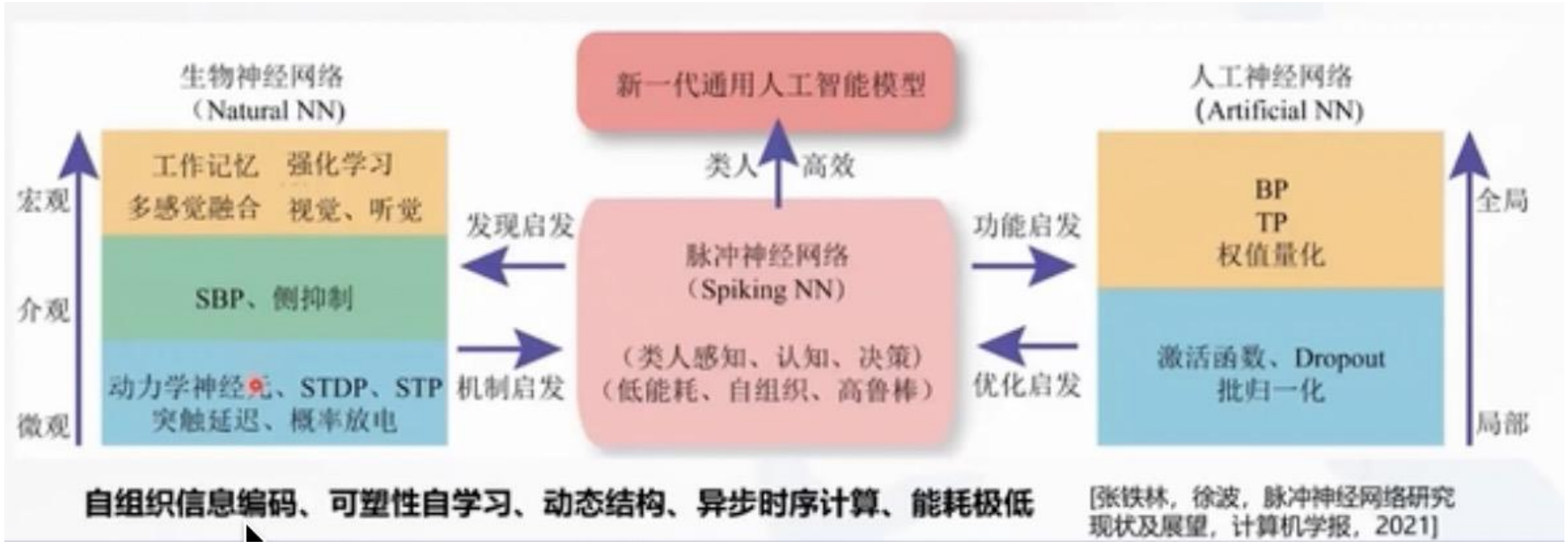
数据量大，标注不足--自监督对比学习



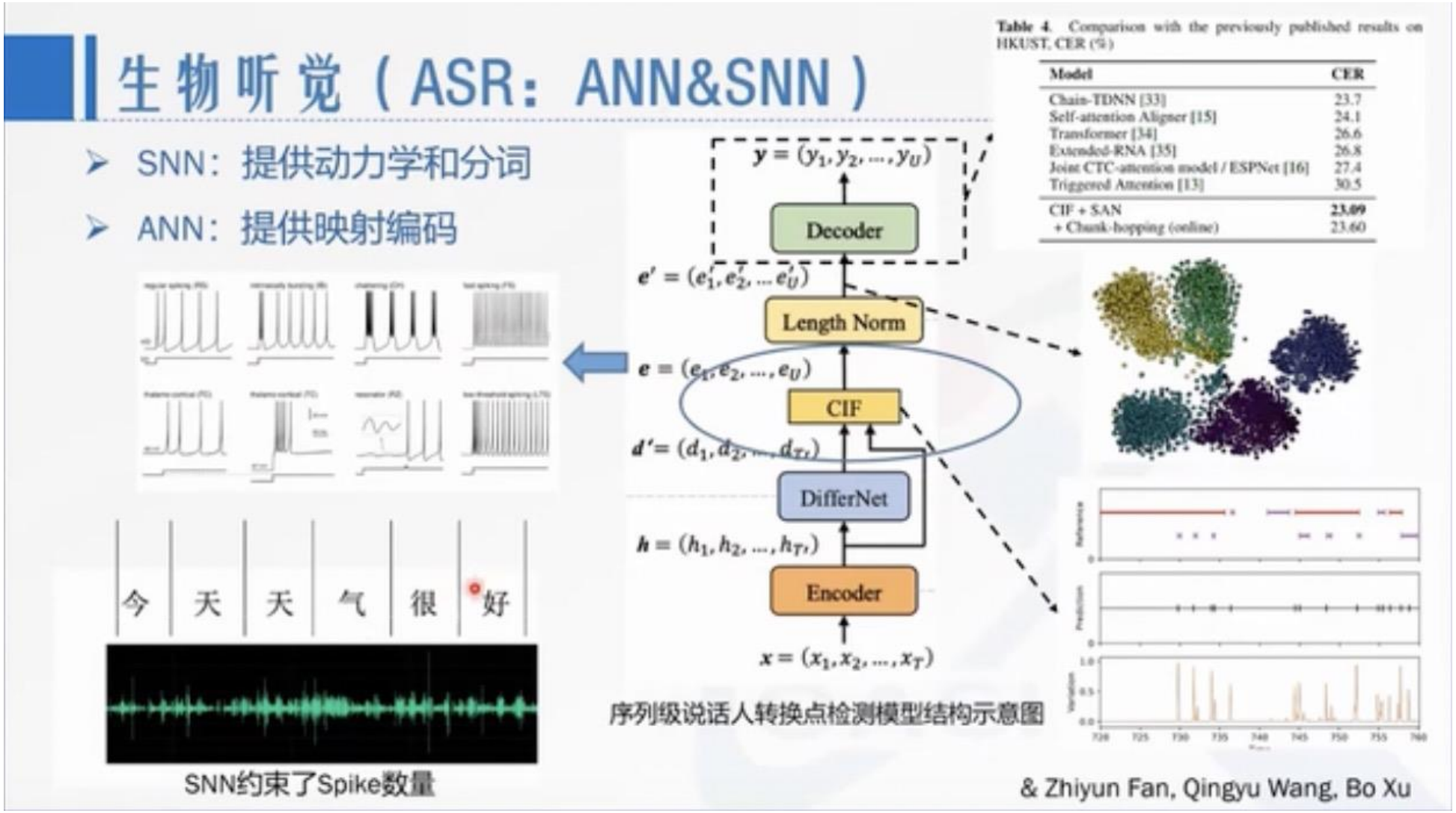


$$\mathcal{L}_y = -log\frac{exp(sim(y,y')/\tau)}{\sum_{i=0}^{N} exp(sim(y,y')/\tau)}$$

多模态认知计算新趋势：类脑SNN与ANN融合

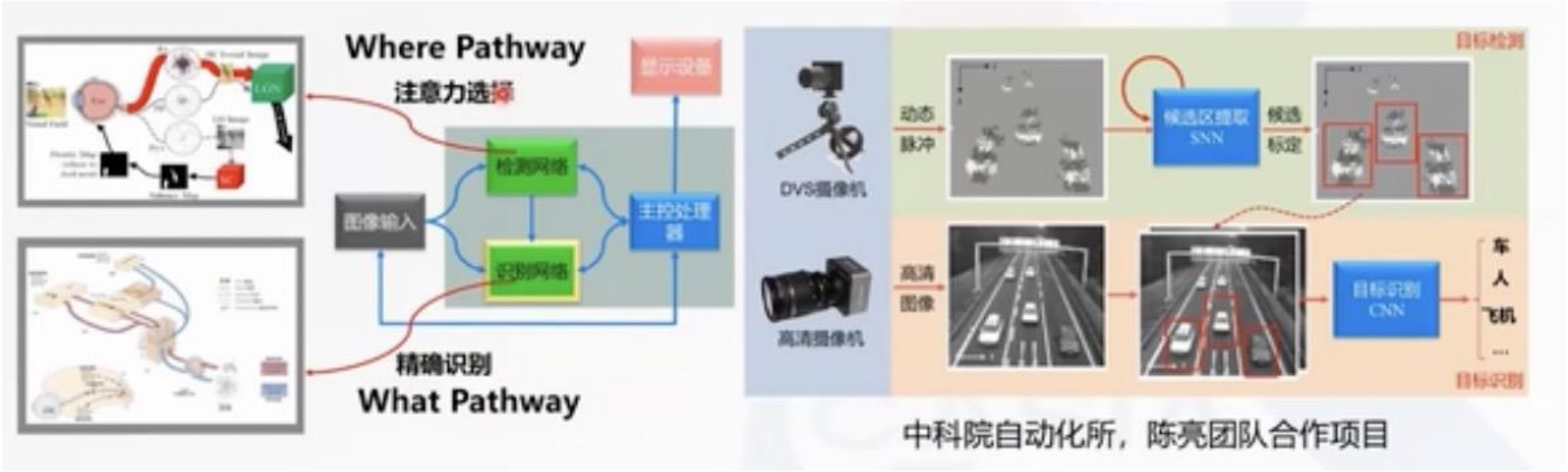


融合方式一： 结构混合融合显著提升特异性



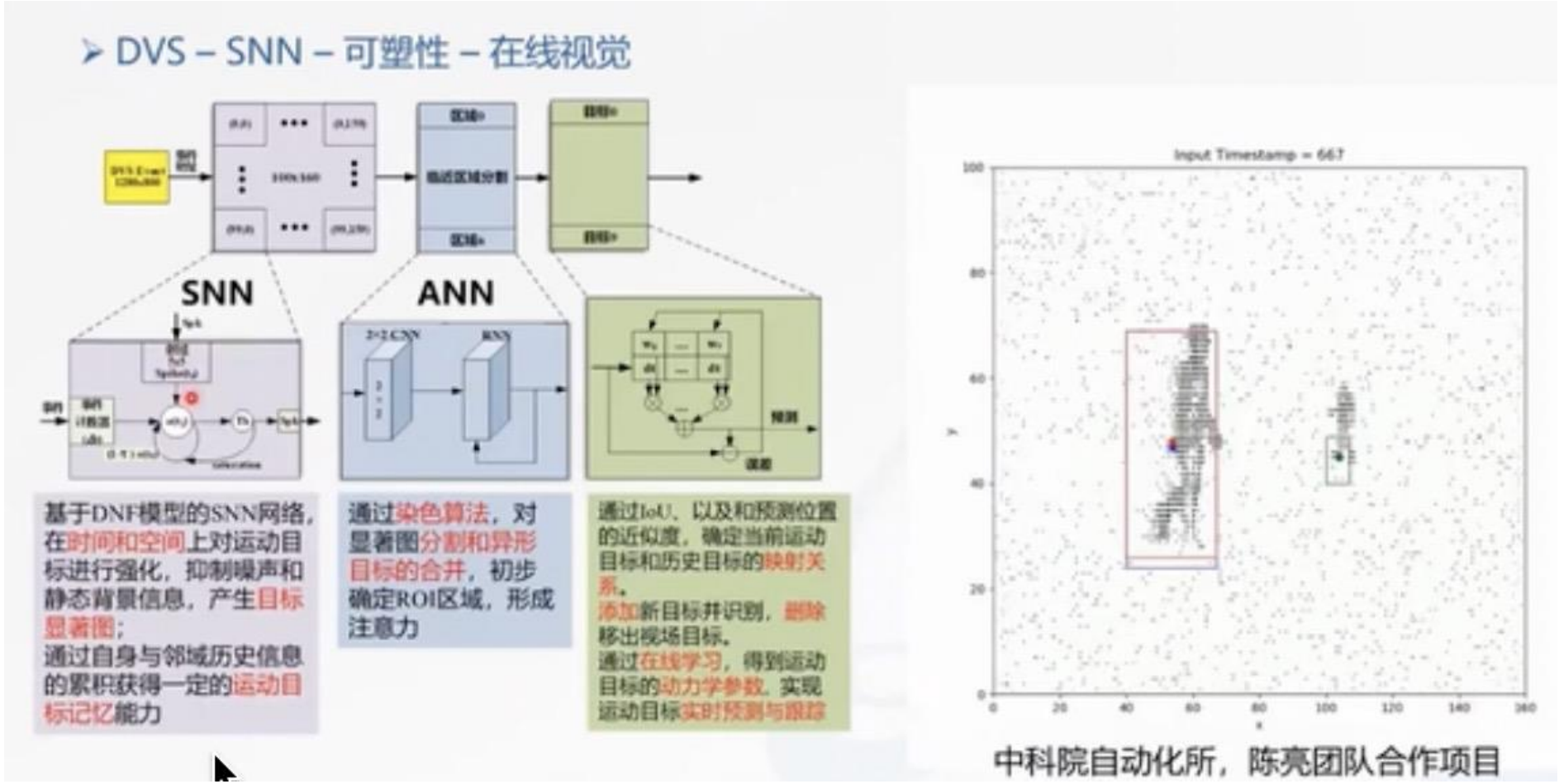
https://www.bilibili.com/video/BV12A4y197mW/

融合方式二： 串型混合融合提升模型效率



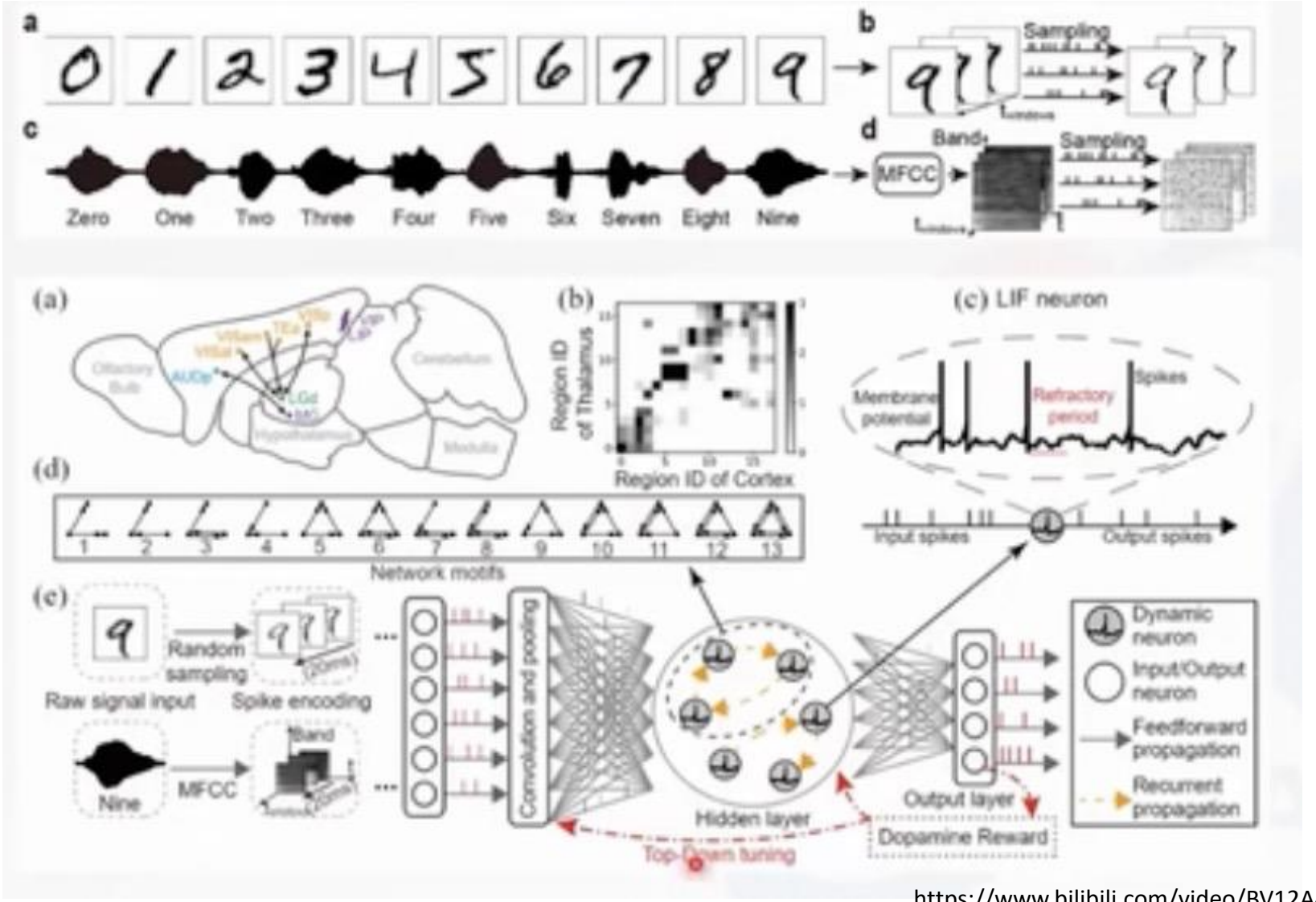
https://www.bilibili.com/video/BV12A4y197mW/

串型混合模型



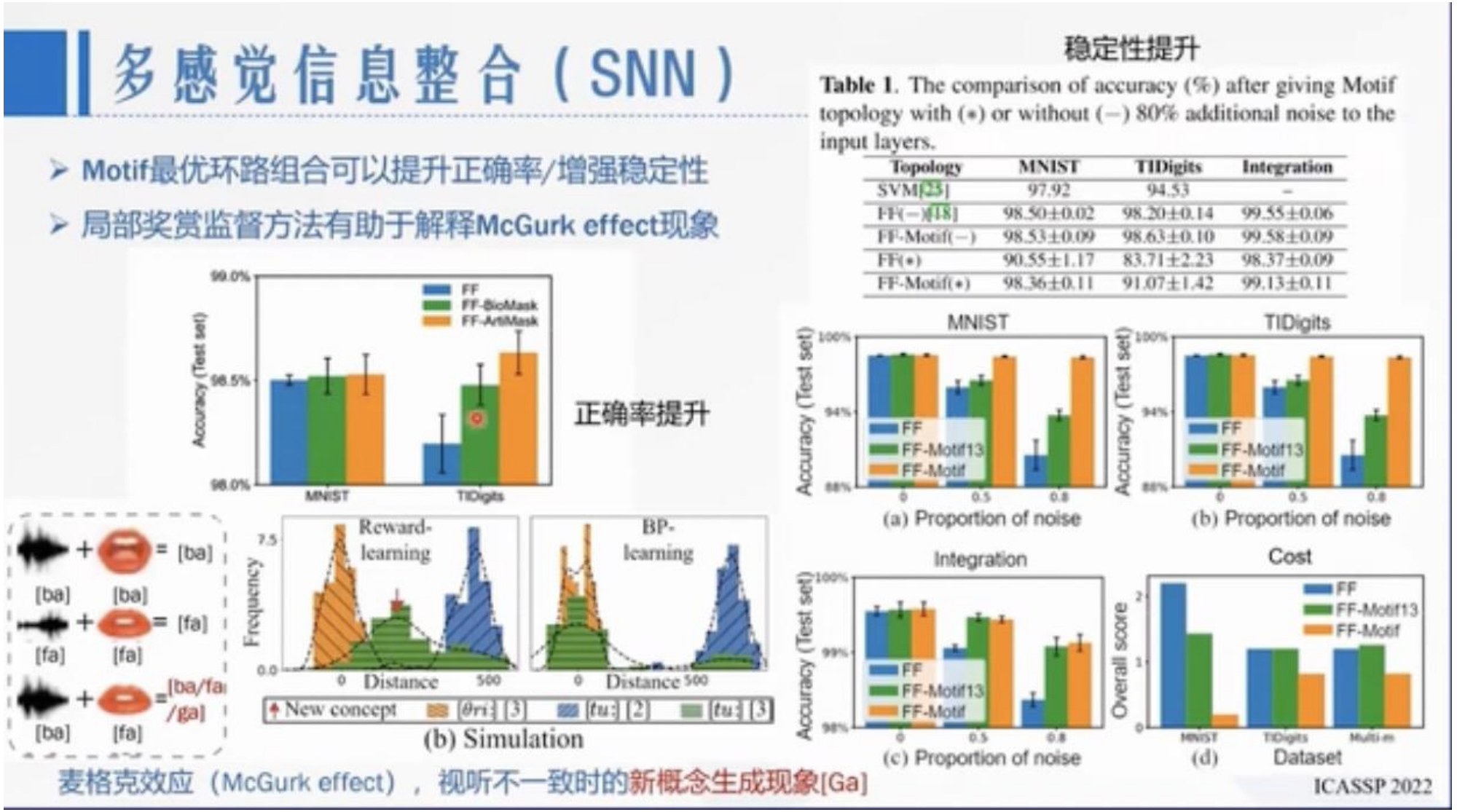
https://www.bilibili.com/video/BV12A4y197mW/

多模态整合： 通过Motif整合，提升新概念识别能力



https://www.bilibili.com/video/BV12A4y197mW/

Motif整合机制的新“特异性”发现



https://www.bilibili.com/video/BV12A4y197mW/

总结

- 认知科学对人类大脑多模态信息整合机制研究表明多模态整合有利于提升准确性，核心问题是如何进行整合分离对齐
- 多模态认知计算理论框架确定了以信息量为核心的新标准，而提高信容是认知计算的核心
- 围绕联想、推理、归纳/演绎为核心的多模态认知计算主要围绕多模态关联、生成、协同来展开，在实践中取得了不错的结果
- 类脑模型和传统AI模型给多模态认知计算带来了新的趋势，提高准确度和效率
- 对于机制层面的深挖对认知计算有更为深远的未来价值

THANKS