

面向自动驾驶的 脑启发可信场景分析

Trustworthy Scene Analysis for Autonomous Driving Inspired by **Brain Cognitive**

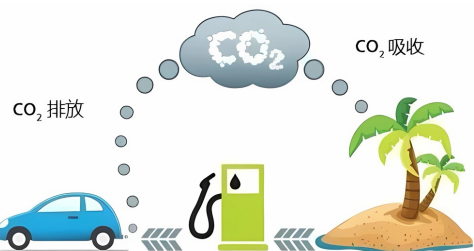
WORKSHOP: 端到端自动驾驶

讲者: 冷佳旭 (Jiaxu Leng)



视觉与学习青年学者研讨会

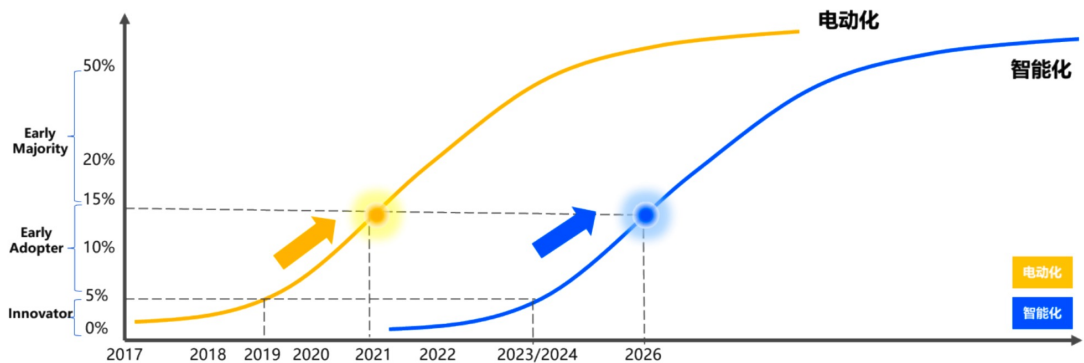
May 7th, 2024 @ VALSE



“碳中和、碳达峰”
双碳国家战略目标



汽车行业“新四化”
(**电动化**、网联化、**智能化**、共享化)



聚焦“33618”现代制造业集群体系“提能级”
重庆深入推进新型工业化
发展新质生产力

重庆市建设世界级智能网联新能源汽车产业集群发展规划

重庆市自动驾驶和车联网创新应用行动计划

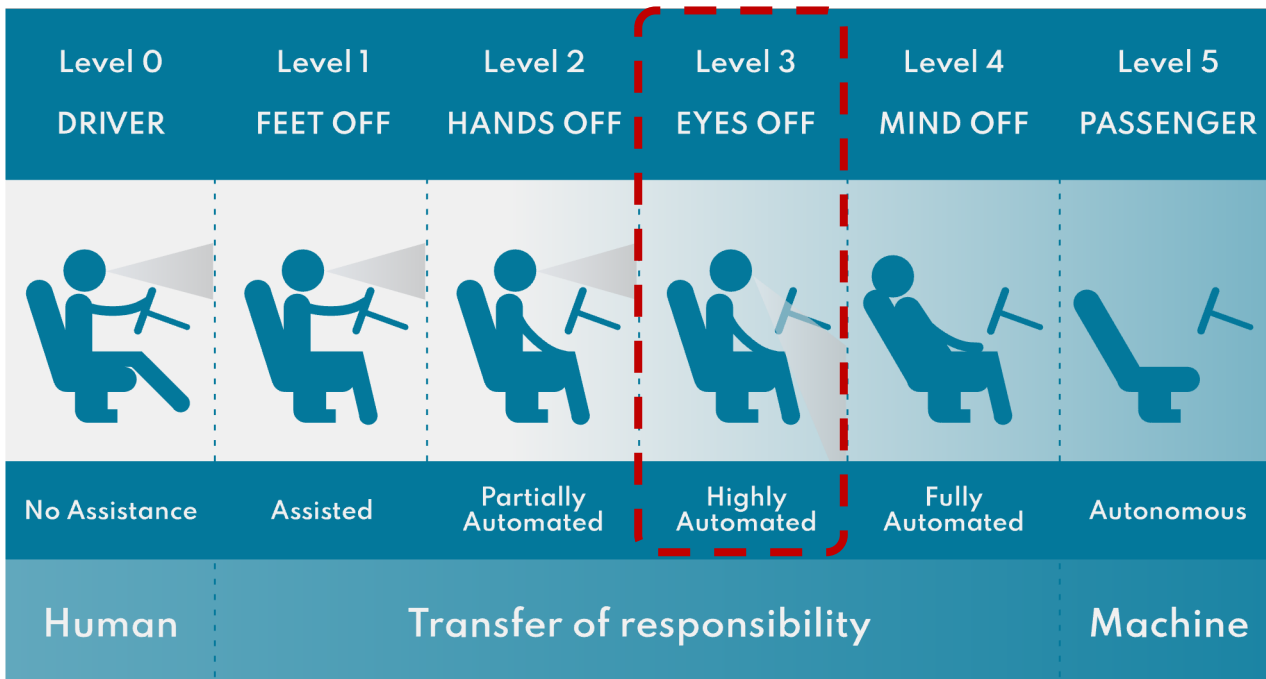


全国首个**自动驾驶**公交车项目落地重庆

个人安全 节省时间 方便出行 降低开支 绿色环保



随着人工智能技术发展及国家地方的高度重视，
汽车产业**智能化**成为电动化后又一个历史性机遇



测试期

- 2022-2023年
- 覆盖一二线城市
- 渗透率 < 5%
- 仅在城市划定范围内试运营

运营政策赋能期

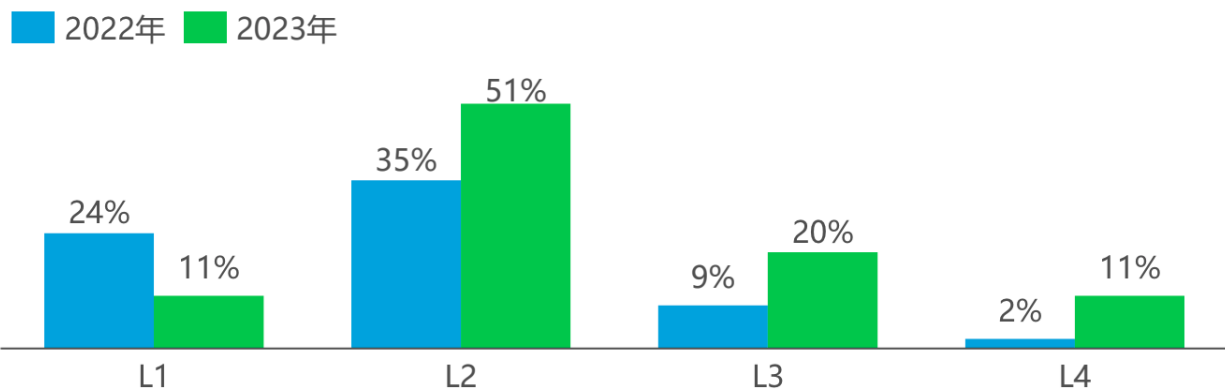
- 2023-2026年
- 覆盖一二线城市
- 渗透率 5-10%+
- 仅在城市划定范围内试运营

驾驶技术成熟期

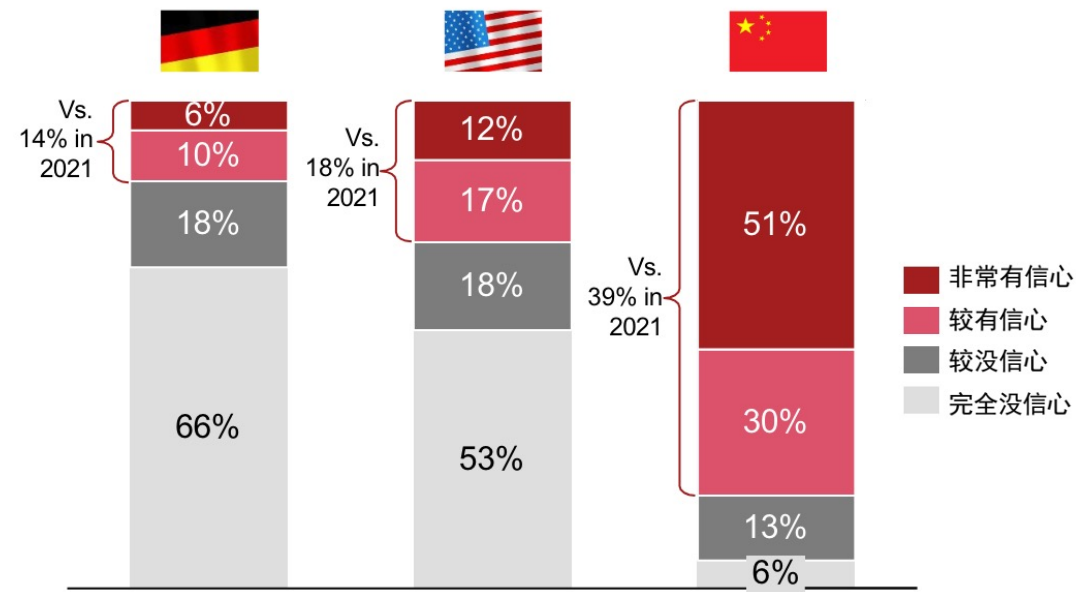
- 2026-2032年
- 一二线城市为主, 低线城市参与
- 渗透率 20-60%
- 扩展到更大的城市区域, 包括市中心

成本效率优势期

- 2032年后
- 全面推广覆盖
- 渗透率 80%+
- 几乎全程覆盖



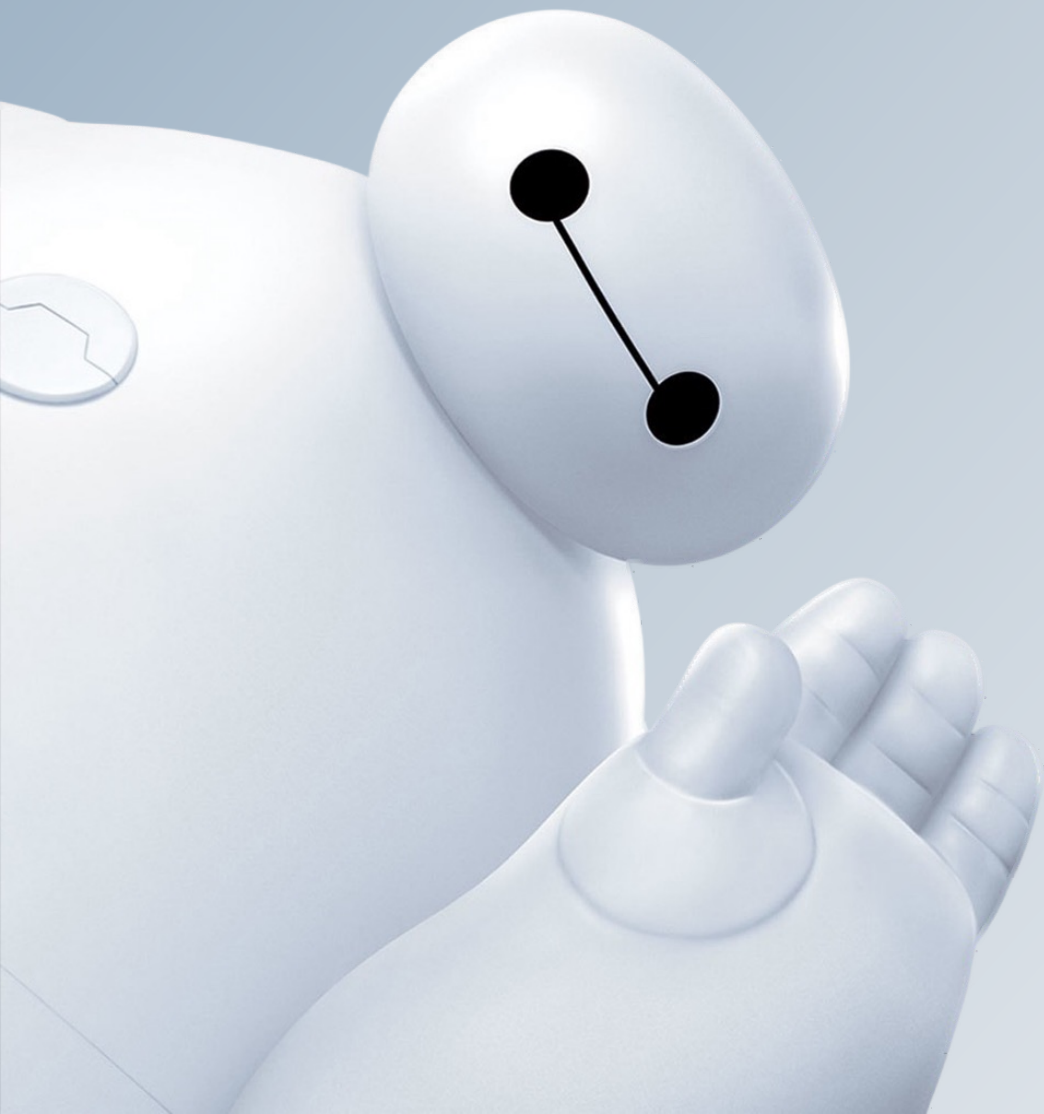
2022-2023年中国在售新车自动驾驶搭载率预测 @共研产业研究院&36氪研究院



2023 数字化汽车报告 @罗兰贝格 @思略特

汇报大纲

- **可信场景分析：以人为本**
- 类脑智能计算：以脑为鉴
- 团队工作简介：类脑计算
- 未来发展趋势：脑机互鉴



01 可信：人工智能发展的必然要求

以人为本

类脑智能计算

团队工作简介

未来发展趋势



孔子

公元前551年-前479年



孟子

约公元前372年-前289年



董仲舒

公元前179年-前104年

“仁、义、礼”

“智”

“信”

个人——人际——社会——理性——人格

道德——伦理——规范——智能——可信赖的人工智能



1637年
勒内·笛卡尔
(René Descartes)

《谈谈方法》

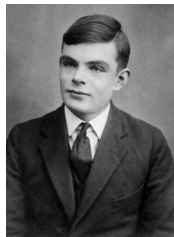
哲学指导



1942年
艾萨克·阿西莫夫
(Isaac Asimov)

“机器人三定律”

道德要求



1950年
艾伦·图灵
(Alan Turing)

图灵测试

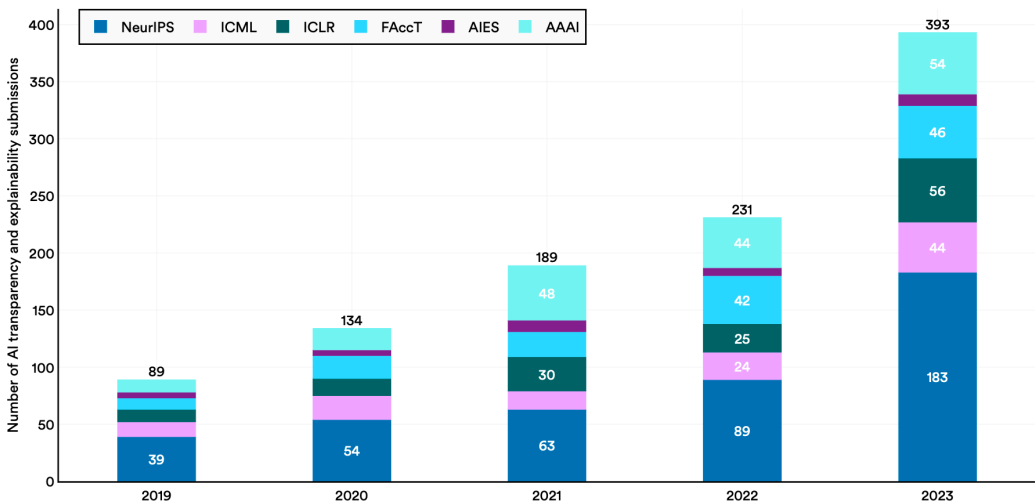
智能评估

对人的要求



对人工智能的要求





学术界对于“可信”问题愈发关注

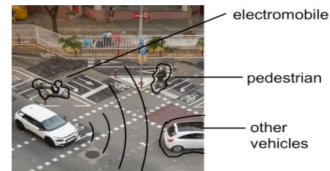


自动驾驶商业化的主要影响因素

可以区分自动化驾驶水平的三个关键指标

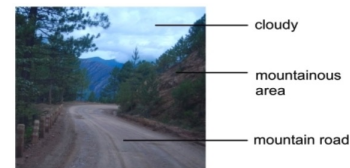
目标和事件检测与反应(OEDR)
Object and Event Detection and Response

由自动驾驶系统或驾驶员负责检测和响应



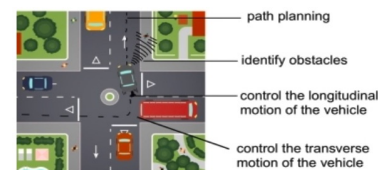
运行设计域(ODD)
Operational Design Domain

规定自动驾驶系统在何种条件下工作

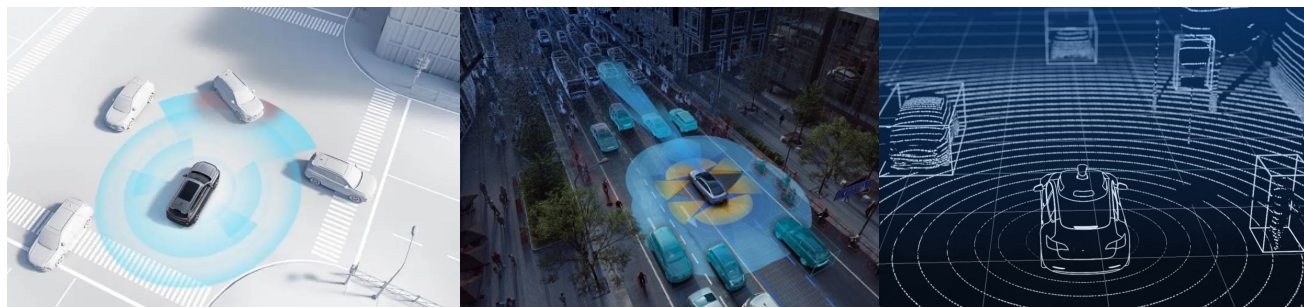


动态驾驶任务(DDT)
Dynamic Driving Task

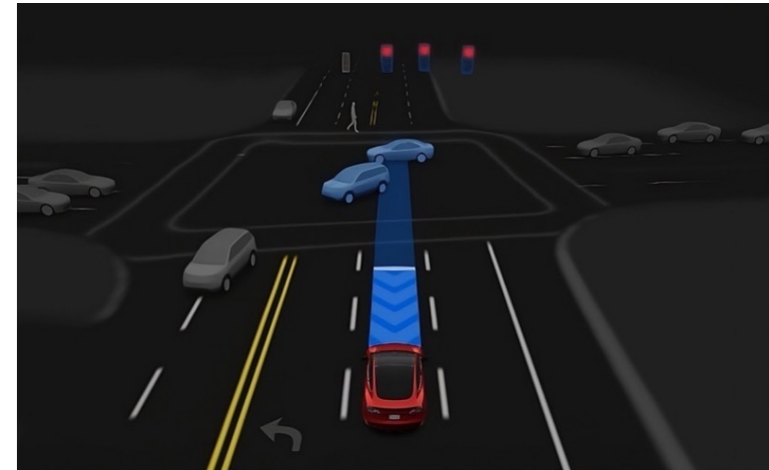
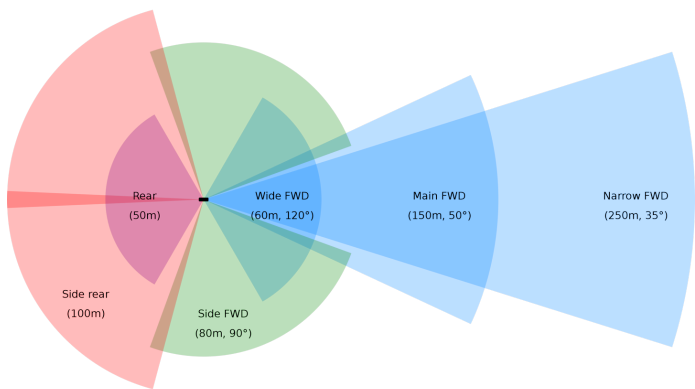
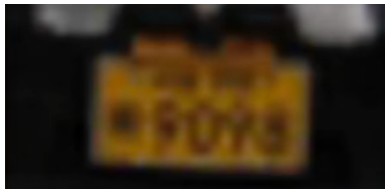
由自动驾驶系统或驾驶员负责执行驾驶任务



@NHTSA(National Highway Traffic Safety Administration)



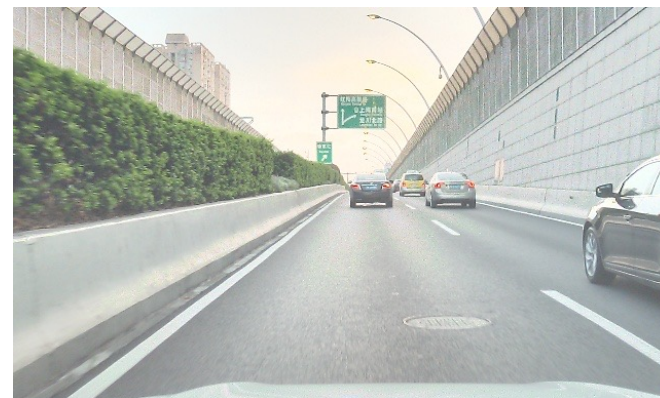
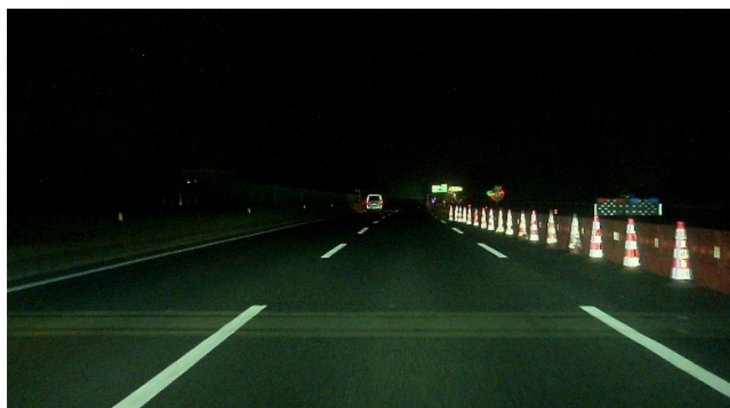
可信场景分析技术是实现智能驾驶感知的核心之一, 对于环境理解的准确与否直接决定了自动驾驶的**安全可靠**程度



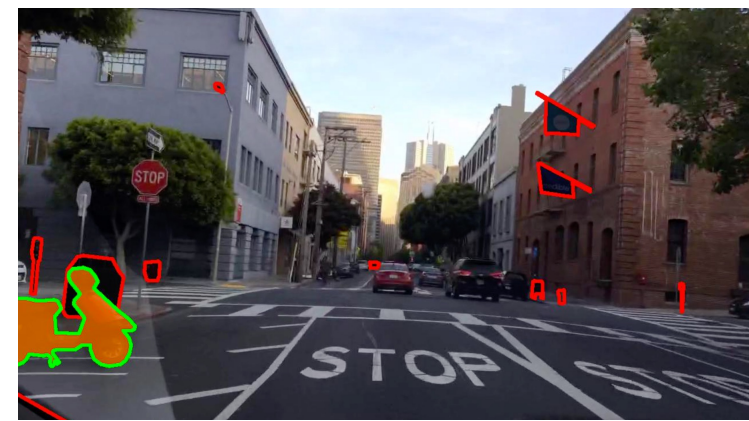
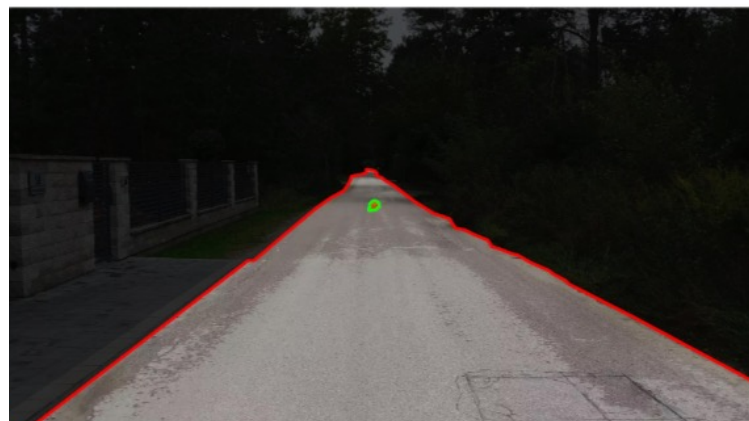
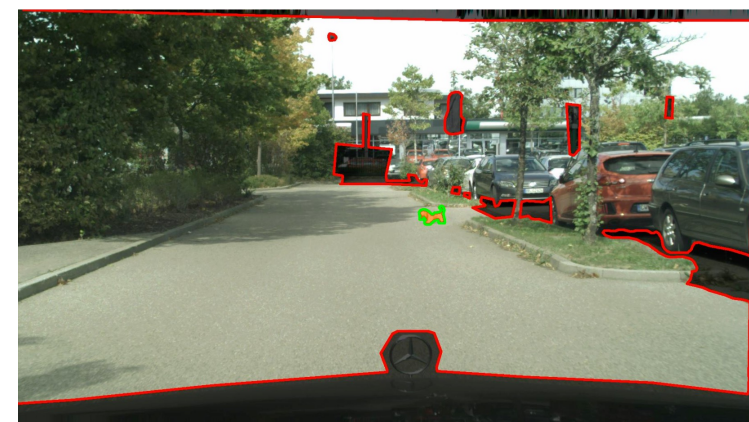
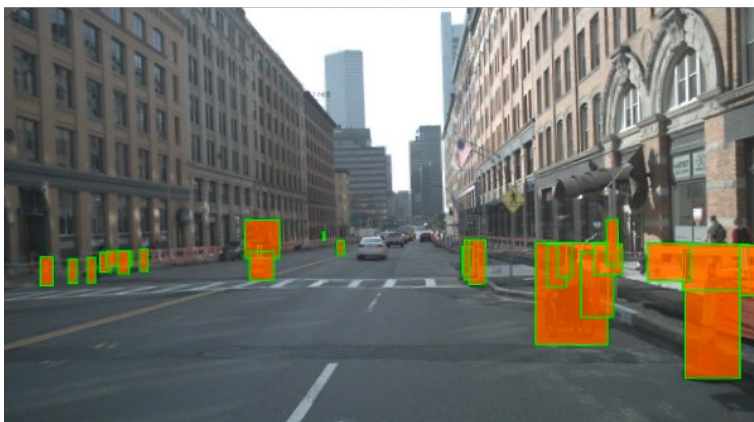
看得清

看得见

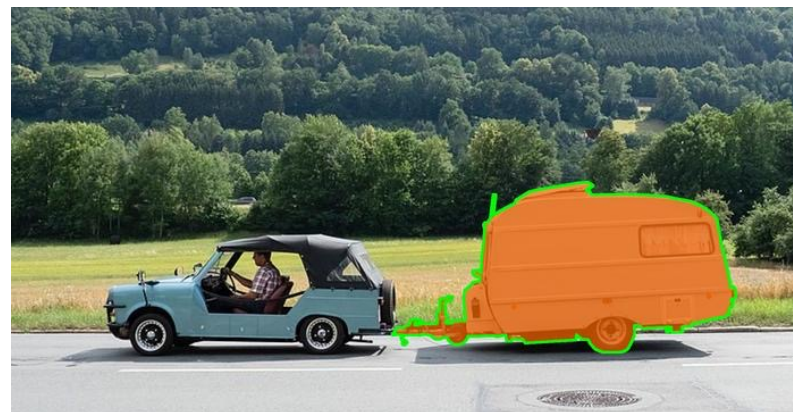
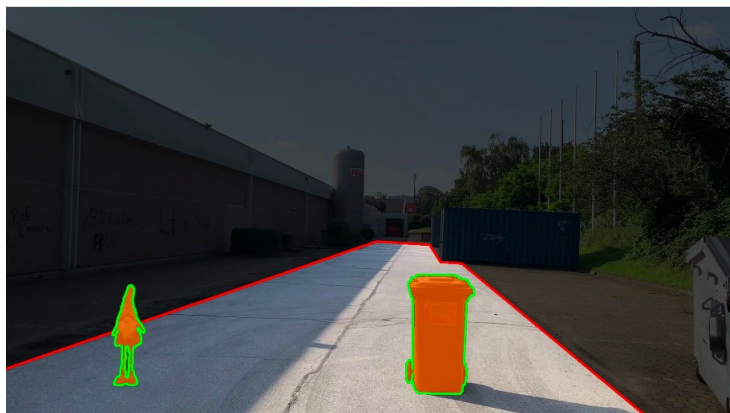
看得懂



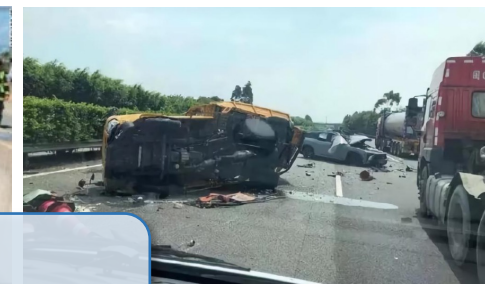
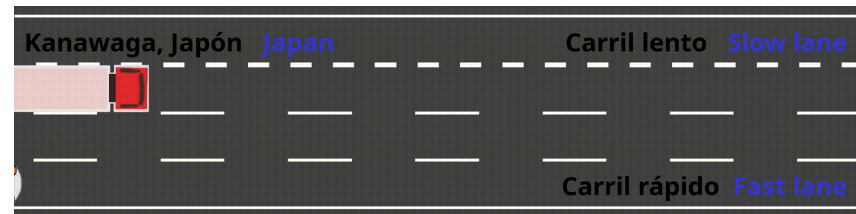
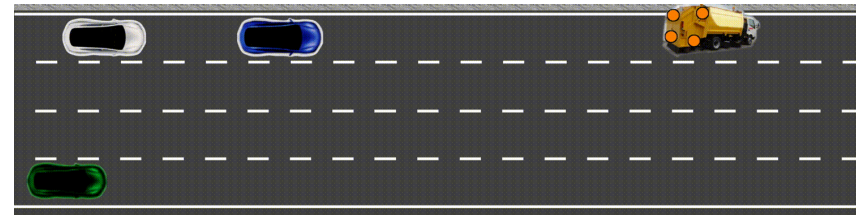
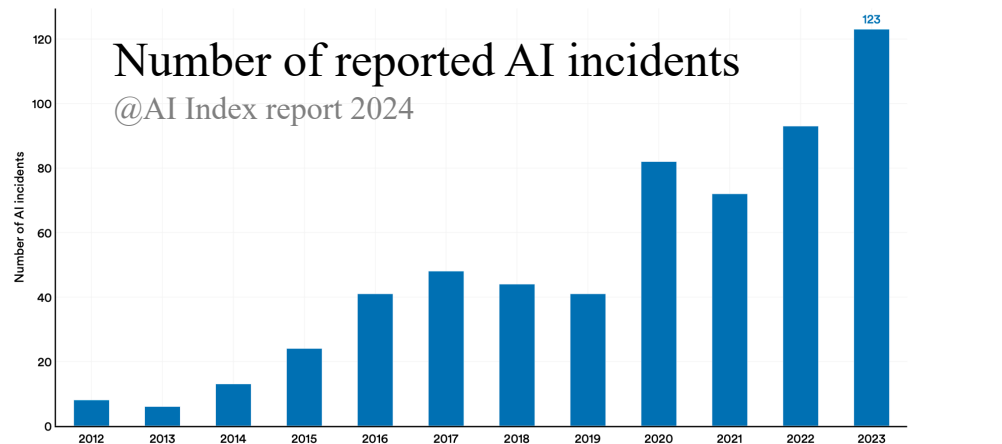
天气、光照变化等因素带来的模糊、噪声等，可能对图像中的关键信息造成干扰，从而给自动驾驶系统带来了更大的挑战，使其难以准确识别路面标志、行人、车辆等关键对象。



地面的钉子、远处的行人、或者交通标志等微小目标的高效捕捉一直是一个具有挑战性的问题。由于自动驾驶系统算力有限，进一步加剧了识别的难度。



对于自动驾驶系统来说，关键在于能够综合多源信息来有效理解场景中的复杂、模糊语义，从而准确地区分正常情况和异常情况，以便做出正确的决策和行动。



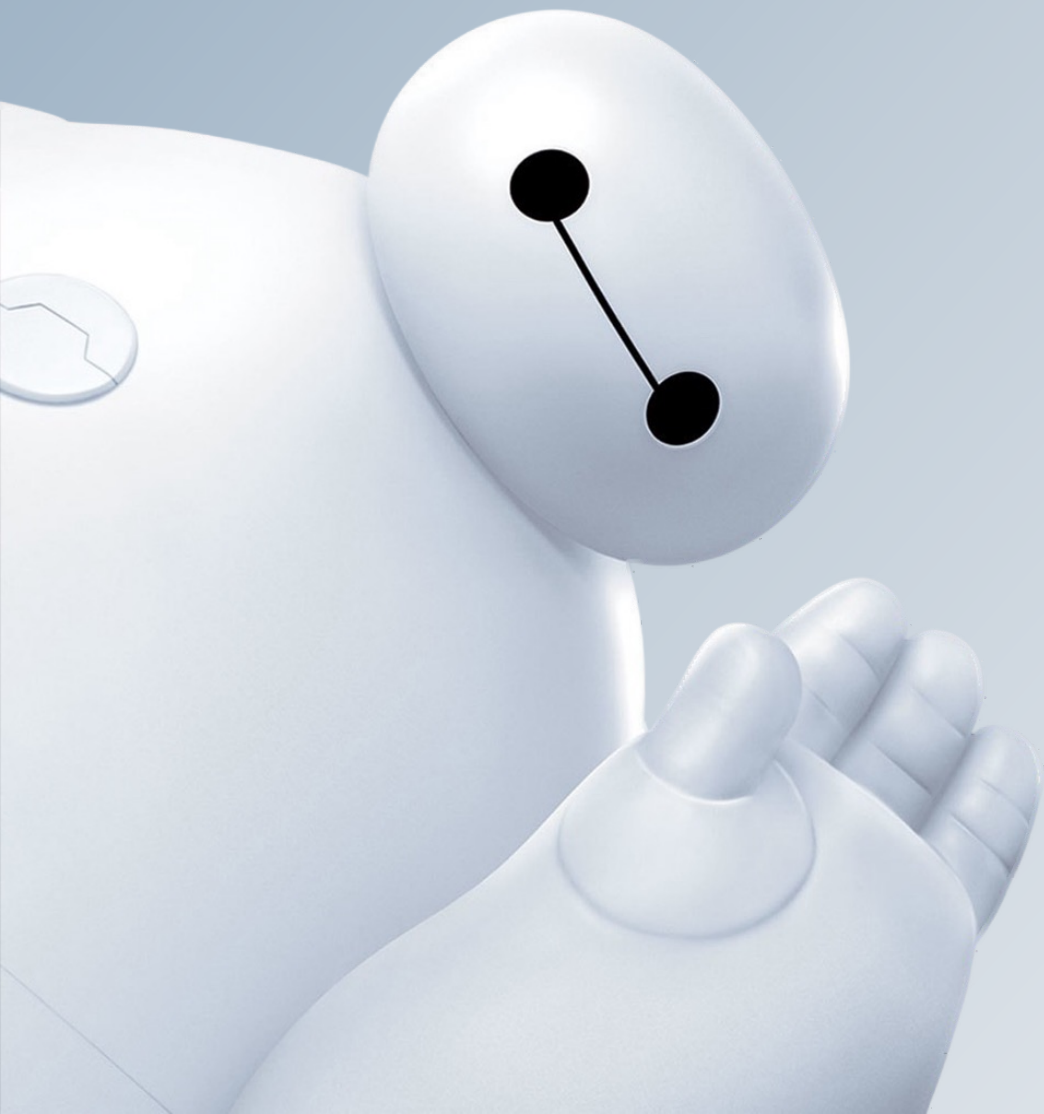
不可信的场景分析方法可能导致无法接受的后果

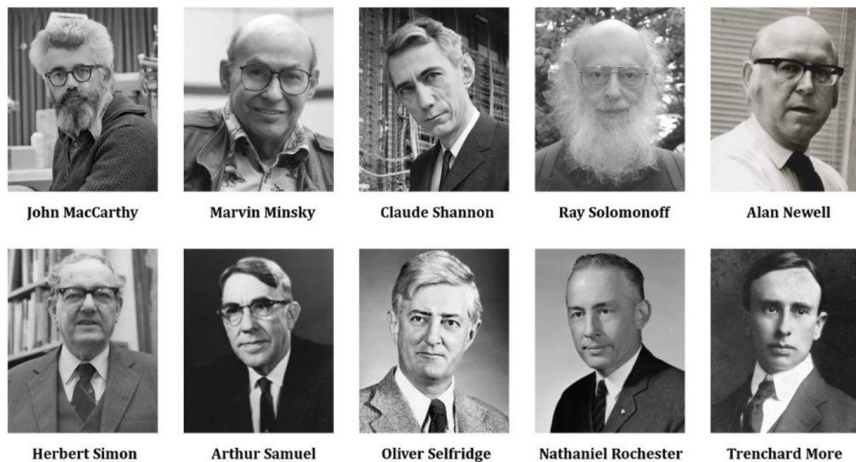


你打车遇到的司机师傅是个陌生人，为什么你会信任他？

汇报大纲

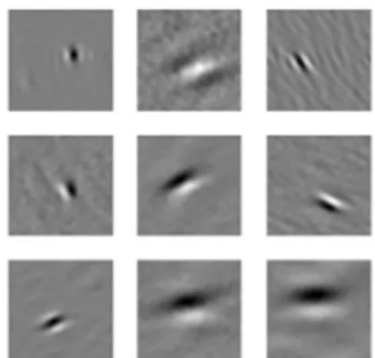
- 可信场景分析：以人为本
- **类脑智能计算：以脑为鉴**
- 团队工作简介：类脑计算
- 未来发展趋势：脑机互鉴



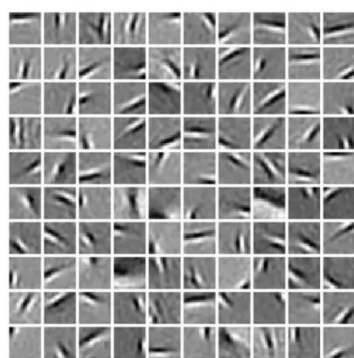


达特茅斯会议，1956

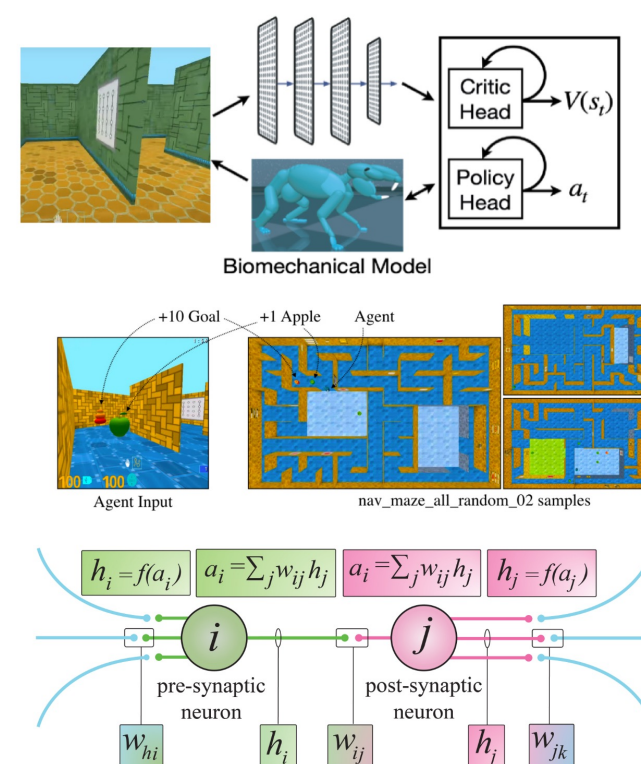
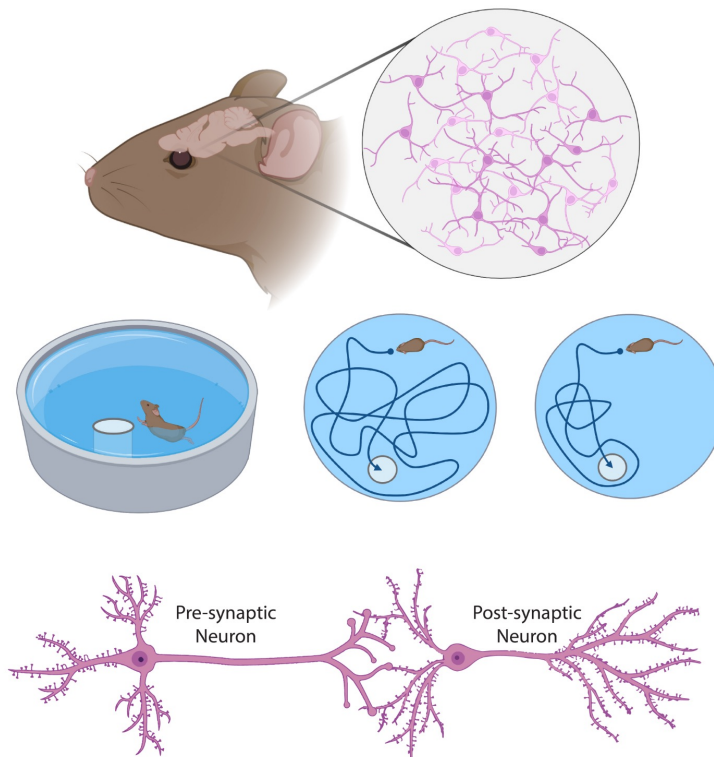
人工智能：让机器像人那样**认知、思考和学习**，
即用计算机来模拟人的智能。



哺乳类初级视觉皮层简单
细胞感受野

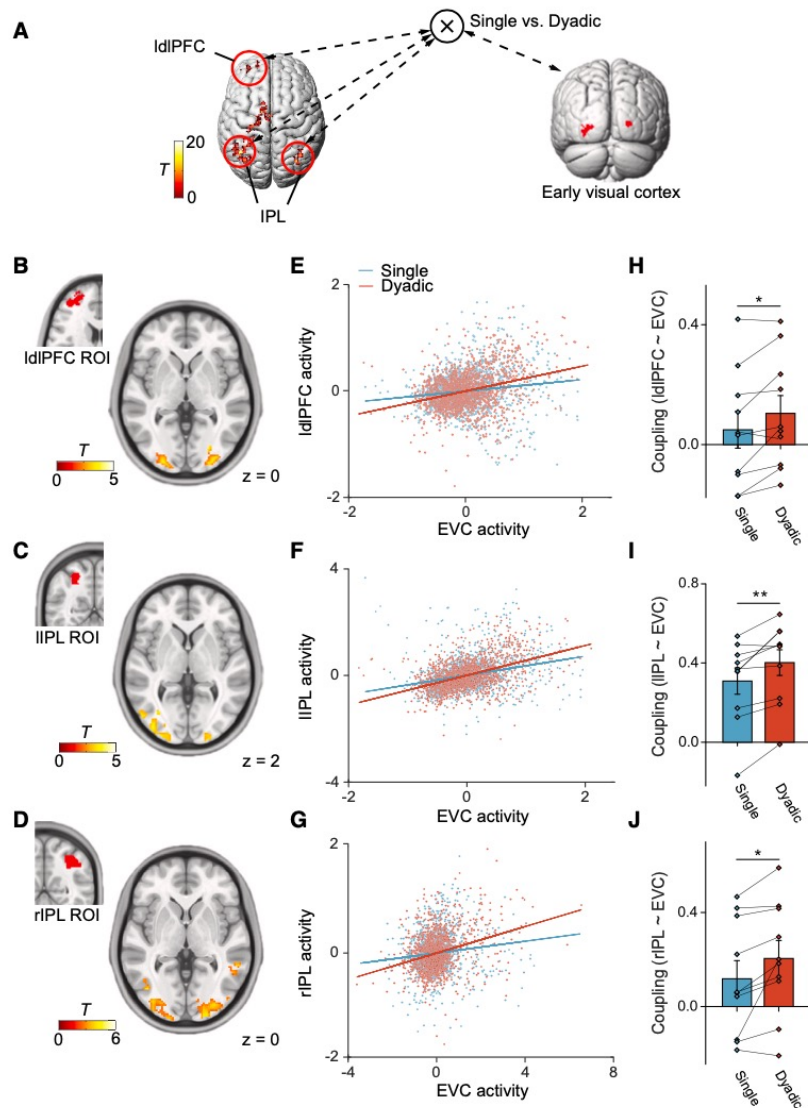


利用稀疏性训练得到
的滤波器



结构形式模仿 → 功能行为模仿 → 学习机制模仿

从信息处理与智能本质角度审视生物脑信息处理，借鉴其原理产生新型
类脑智能计算技术，是实现人工智能创新的重要源泉



与单一训练相比，PPI分析发现在双人训练期间，作为种子区的IdIPFC (B)、IIPLC (C) 和rIPL (D) 与作为目标区的早期视觉皮层之间的功能联系增加。

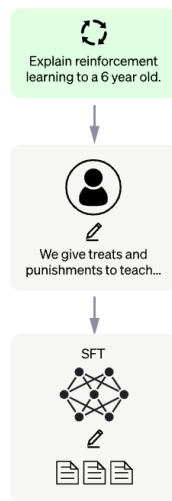
Step 1

Collect demonstration data and train a supervised policy.

A prompt is sampled from our prompt dataset.

A labeler demonstrates the desired output behavior.

This data is used to fine-tune GPT-3.5 with supervised learning.



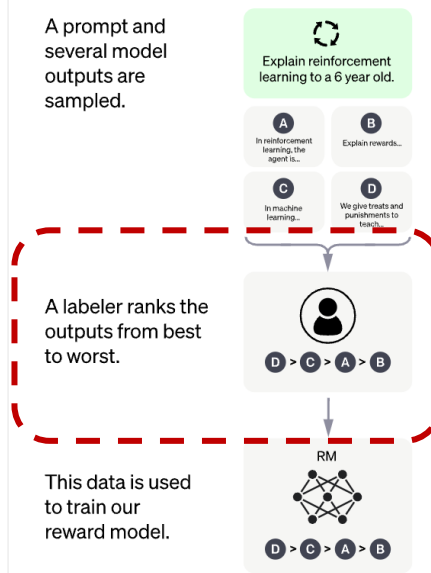
Step 2

Collect comparison data and train a reward model.

A prompt and several model outputs are sampled.

A labeler ranks the outputs from best to worst.

This data is used to train our reward model.



Step 3

Optimize a policy against the reward model using the PPO reinforcement learning algorithm.

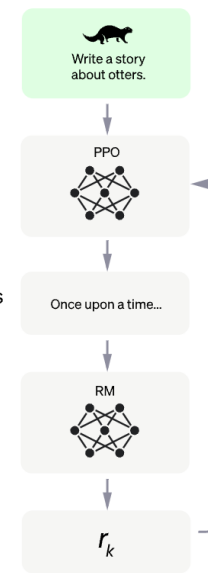
A new prompt is sampled from the dataset.

The PPO model is initialized from the supervised policy.

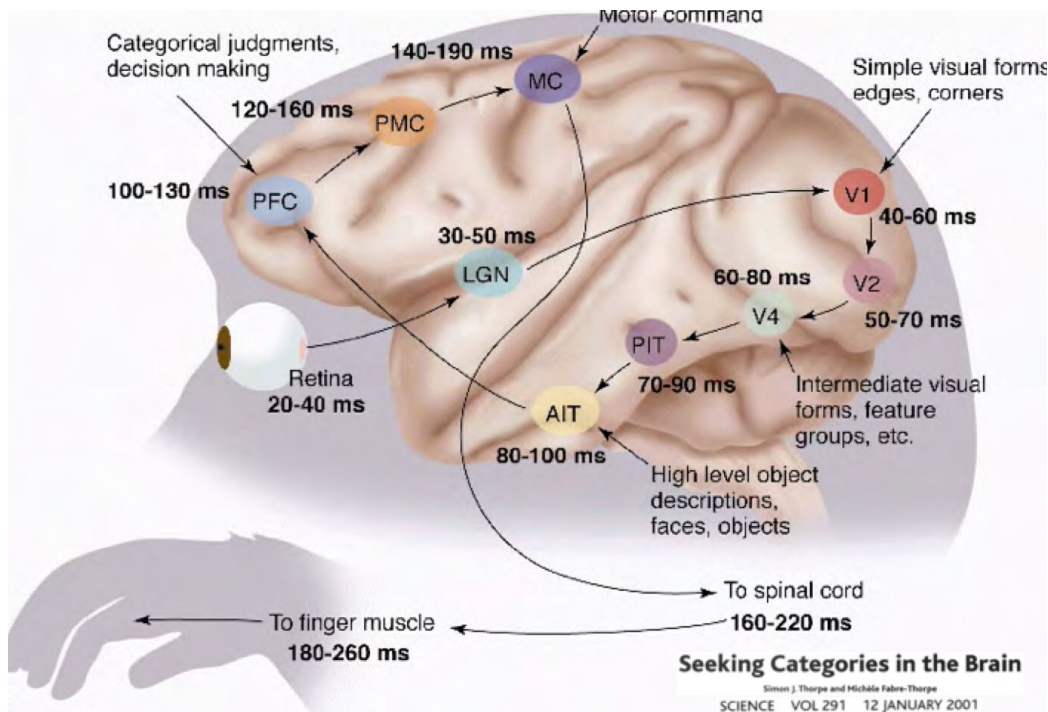
The policy generates an output.

The reward model calculates a reward for the output.

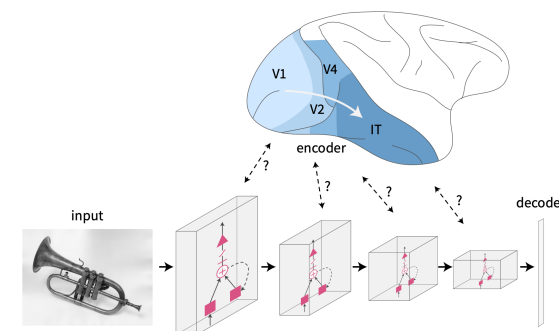
The reward is used to update the policy using PPO.



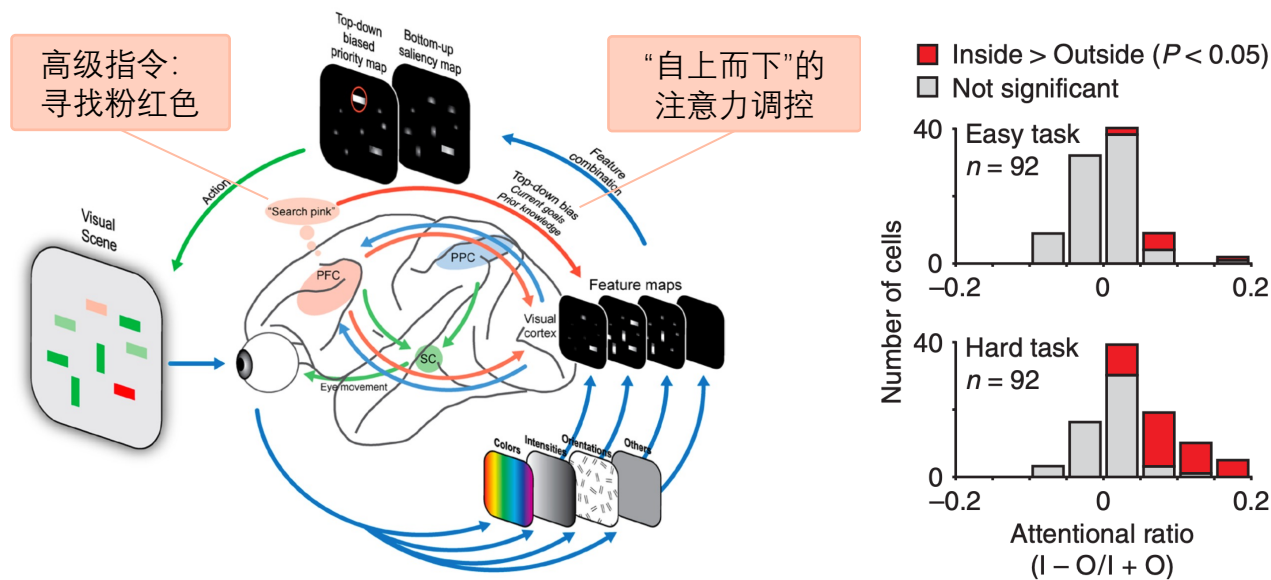
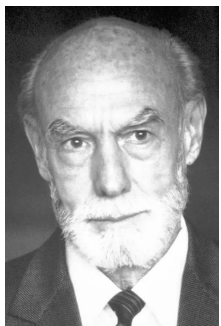
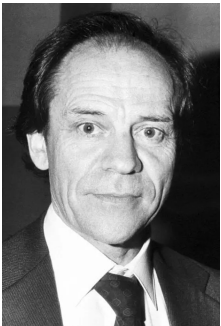
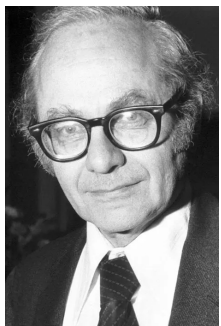
在 chatGPT 中，人类反馈强化学习 (Reinforcement Learning from Human Feedback, RLHF) 使语言模型能够将在文本数据的一般语料库上训练的模型与复杂的人类价值观的模型对齐。



从视觉处理机制可以看出高层的特征是低层特征的组合，从低层到高层的特征表达越来越抽象和概念化。



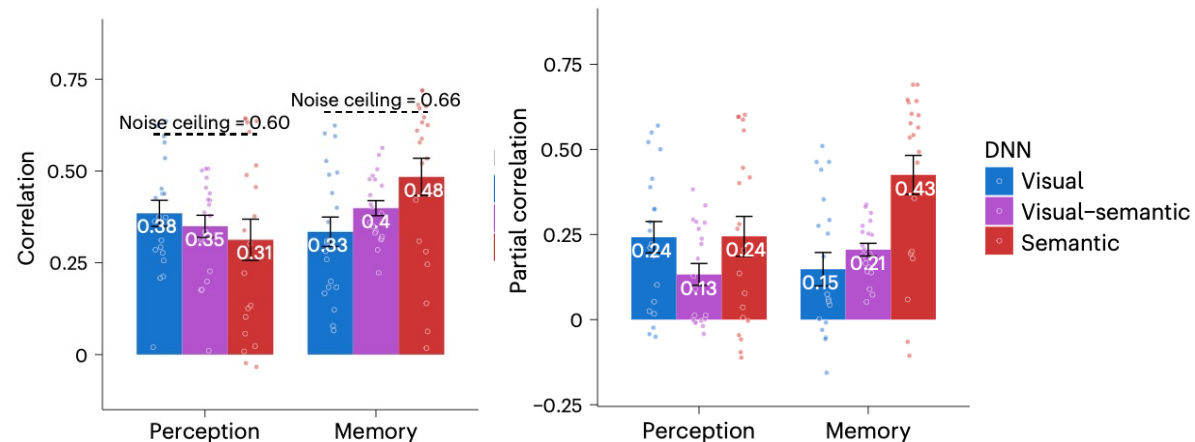
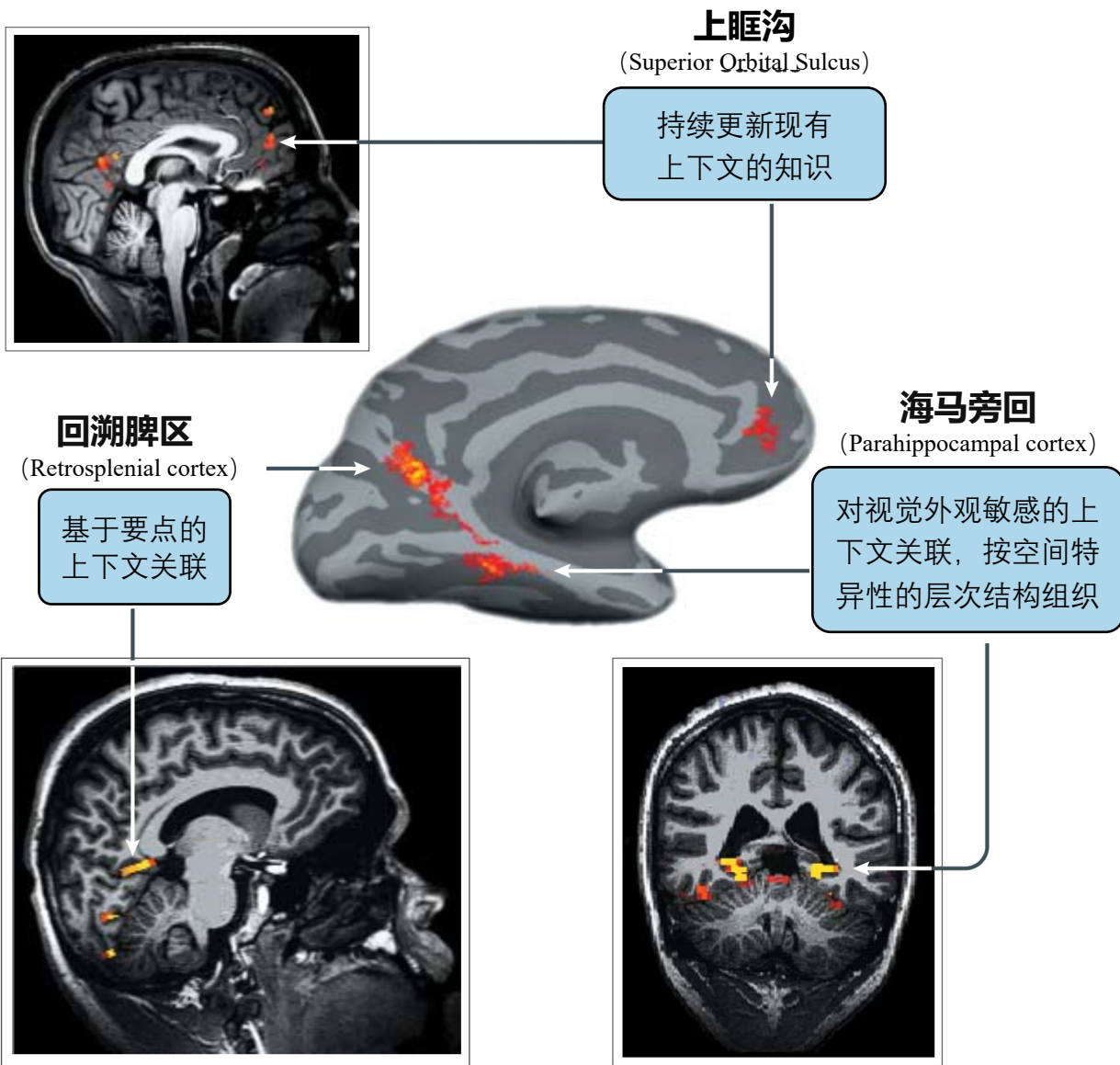
诺贝尔生理学或医学奖 (1981年)
@ David Hubel & Torsten Wiesel & Roger Sperry



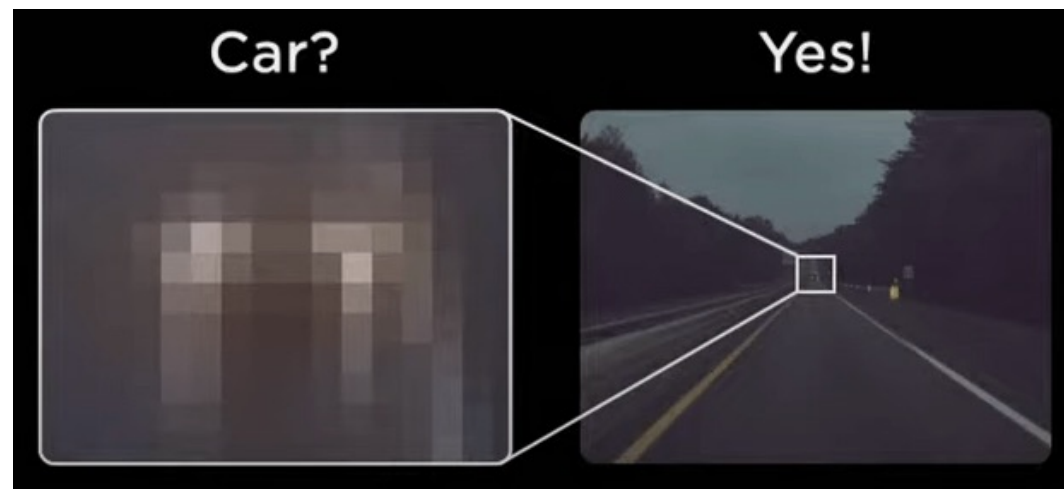
人类在面对**不同难度**事情时，会倾向于选择性地投入**不同精力**，以高效精准地完成识别任务

Katsuki F, Constantinidis C. Bottom-up and top-down attention: different processes and overlapping neural systems. *Neuroscientist* 20:509-21, 2014.

Chen, Yao, et al. "Task difficulty modulates the activity of specific neuronal populations in primary visual cortex." *Nature neuroscience* 11.8 (2008): 974-982.

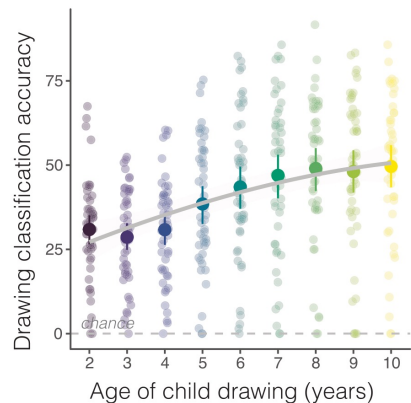
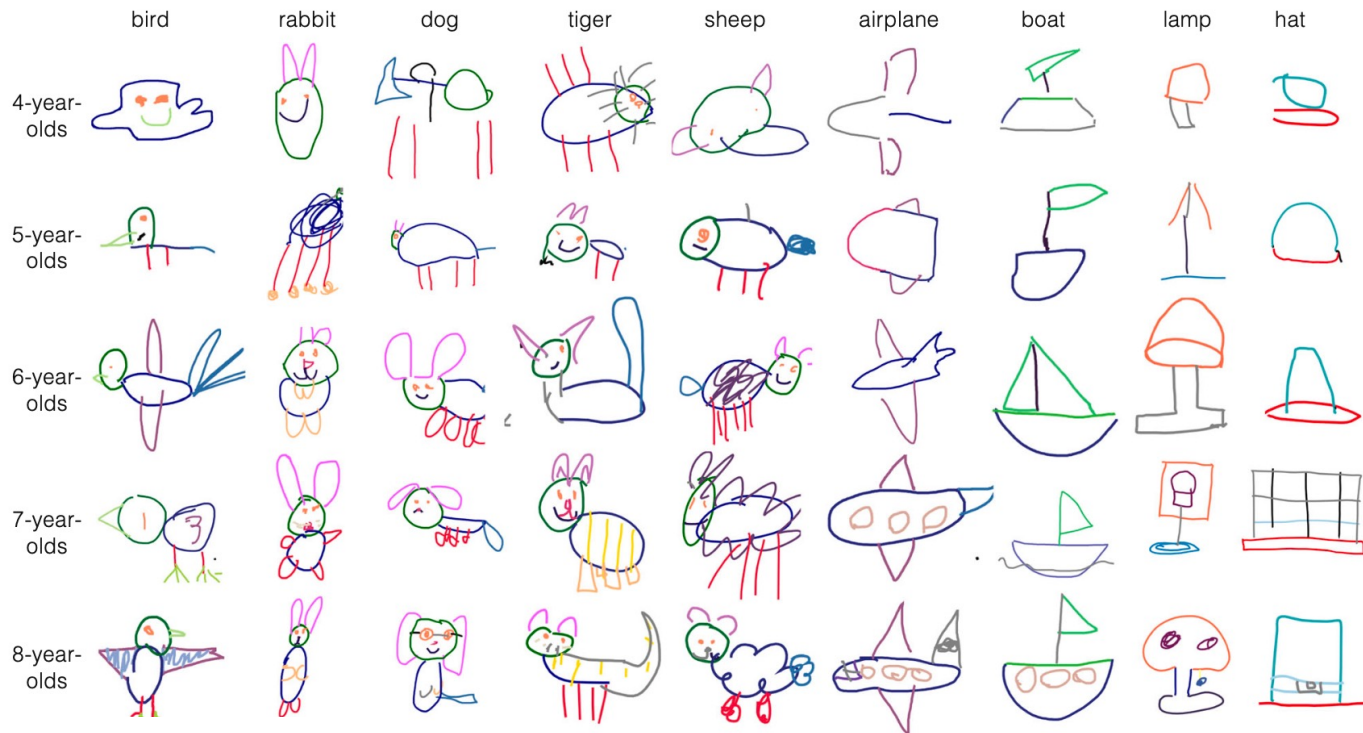


在人类感知中, 基于视觉 (object-VGG)、视觉语义 (CLIP) 和语义 (SGPT) 的深度特征表示与对象及其部件之间的相关性存在差异, **上下文语义信息**在识别物体时发挥着重要作用



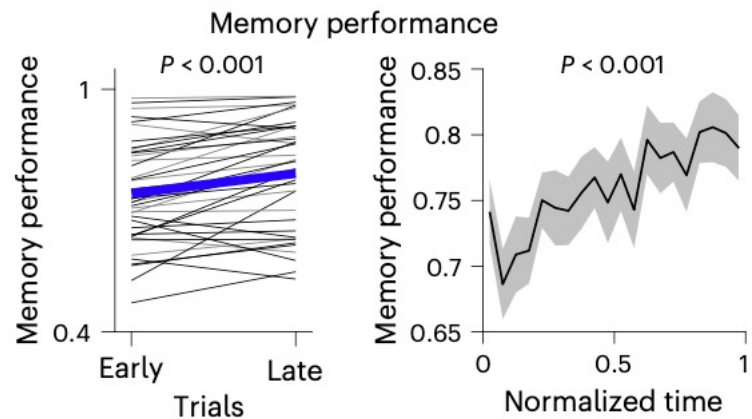
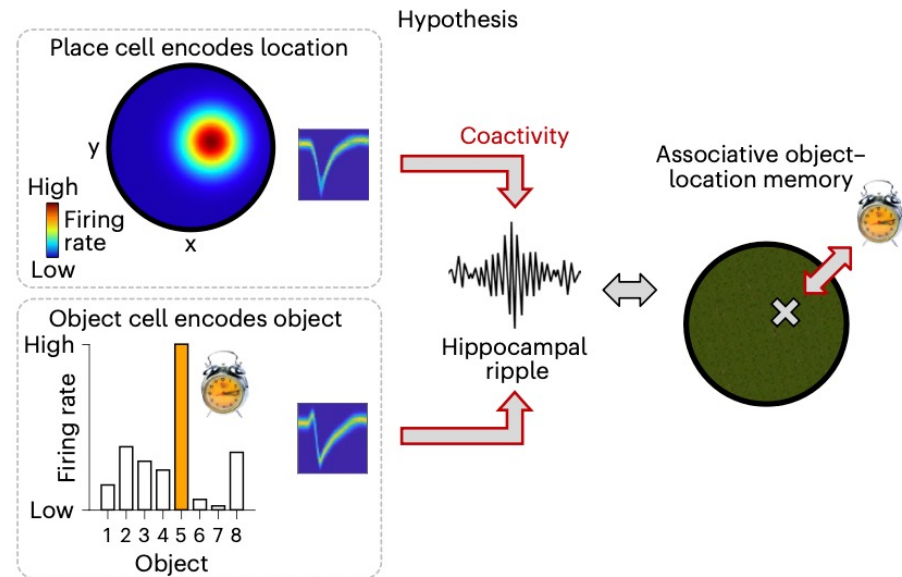
Bar, Moshe. "Visual objects in context." *Nature Reviews Neuroscience* 5.8 (2004): 617-629.

Shoham, Adva, et al. "Using deep neural networks to disentangle visual and semantic information in human perception and memory." *Nature Human Behaviour* (2024): 1-16.



“Children’s drawings of objects reflect not only what they can directly observe, but what they know about these objects.”

儿童对物体的绘画不仅反映了他们可以直接观察到的内容，还反映了他们联想记忆中**对这些物体的理解**。

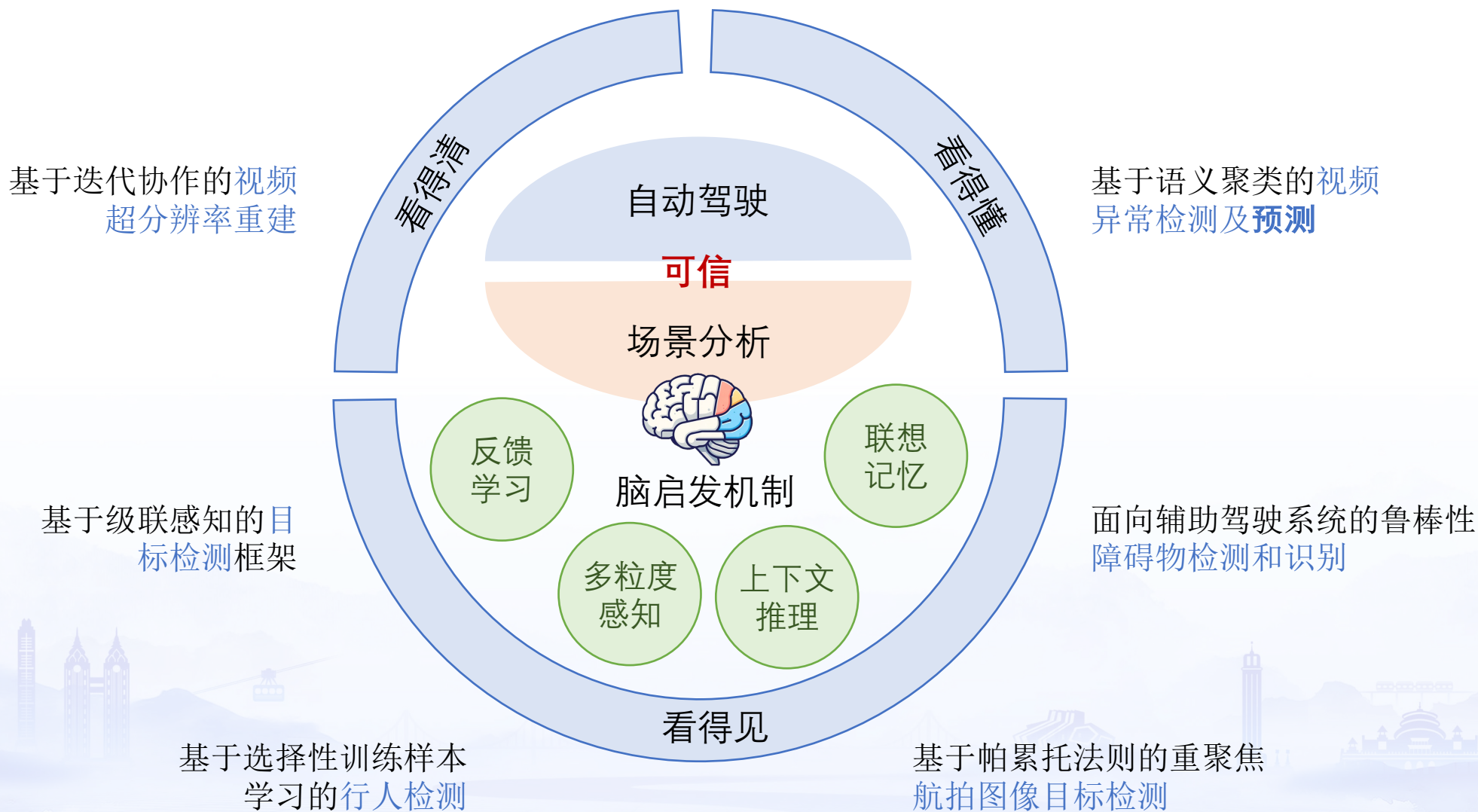


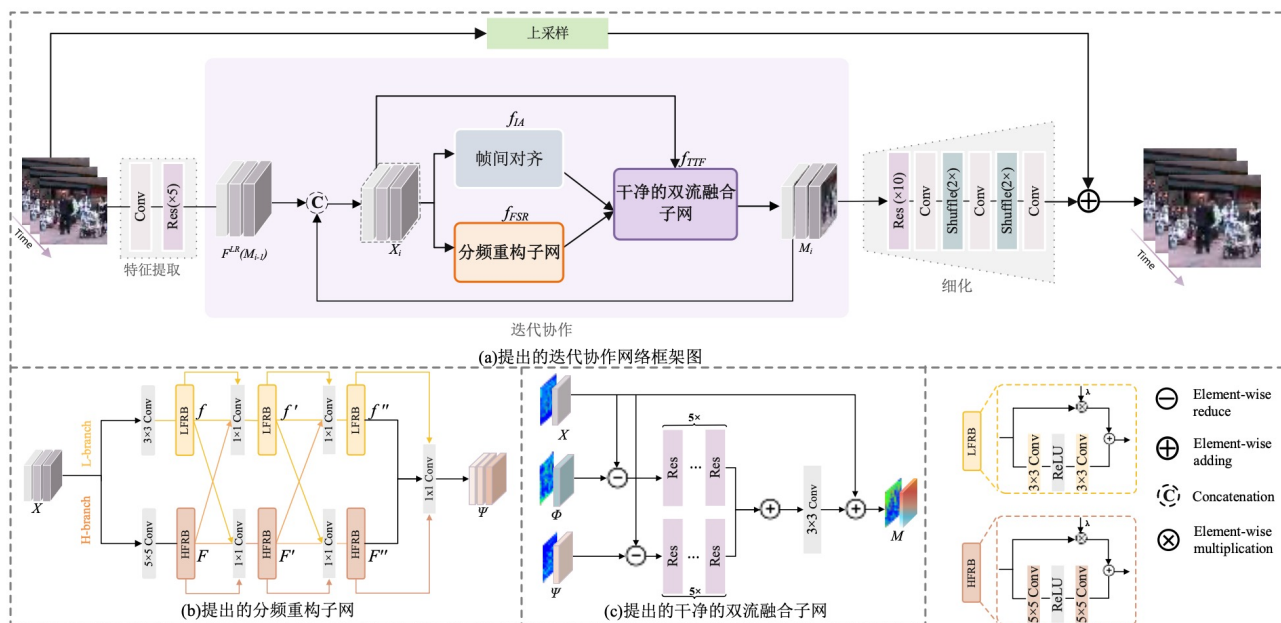
Long, Bria, et al. "Parallel developmental changes in children’s production and recognition of line drawings of visual concepts." *Nature Communications* 15.1 (2024): 1191.

Kunz, Lukas, et al. "Ripple-locked coactivity of stimulus-specific neurons and human associative memory." *Nature Neuroscience* (2024): 1-13.

汇报大纲

- 可信场景分析：以人为本
- 类脑智能计算：以脑为鉴
- **团队工作简介：类脑计算**
- 未来发展趋势：脑机互鉴





针对问题:

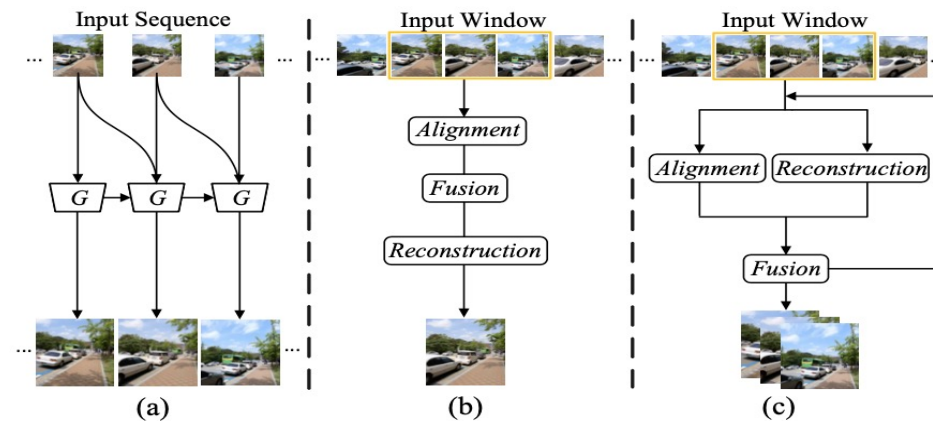
现有大多数方法使用固有网络框架，仅在低分辨率视频帧的特征图上进行对齐操作，导致对齐不准确。这样的错误如果没有处理好，会在后续的融合模块和重构模块中被放大，从而导致模型性能低下。

对齐-重构迭代协作 → 可信视频超分

主要创新:

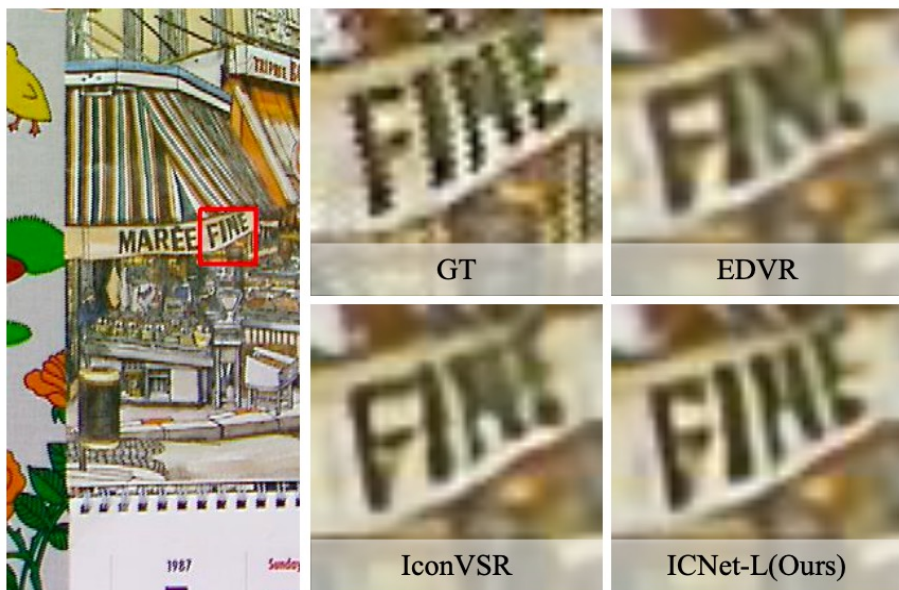
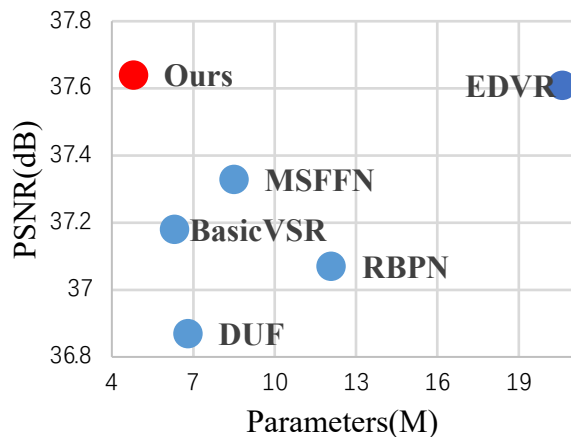
1. 视频超分辨率重建新框架(ICNet):通过对齐操作和重构操作的**迭代协作**来有效的减少以往固有网络框架所带来的错误积累。
2. 分频重构子网络(FSR):通过对输入视频帧的特征图分别学习其高频信息和低频信息，并进行**信息交换**，从而高效利用视频帧的空间信息以达到分辨率增强的目的。
3. 时空残差融合(SRF):通过分别对对齐帧和分辨率增强后的目标帧与低分辨率视频帧的残差特征图进行精细学习和融合，以达到无冗余融合对齐帧和分辨率增强目标帧的目的。

框架对比:



Jiaxu Leng, Jia Wang, Xinbo Gao*. "ICNet: Joint Alignment and Reconstruction via Iterative Collaboration for Video Super-Resolution." In *Proceedings of the 30th ACM International Conference on Multimedia (ACM MM) 2022*.

实验结果：



- 在Vid4数据集上，与现有最先进方法进行了对比实验

Algorithm	Framework	Calendar	City	Folige	Walk	Average
FRVSR	RF-based	-	-	-	-	26.69/0.822
TOFlow	SF-based	22.44/0.7291	26.74/0.7376	26.22/0.757	29.02/0.8776	25.86/0.7626
RBPN	SF-based	23.99/0.807	27.73/0.803	26.22/0.757	30.70/0.909	27.12/0.818
EDVR-L	SF-based	23.98/0.8146	27.98/0.8118	26.33/0.7635	27.32/0.8263	27.32/0.8263
IconVSR	RF-based	24.05/0.8142	27.86/0.8109	26.55/0.7706	31.05/0.9159	27.38/0.8279
ICNet-M	IC	24.14/0.8158	27.81/0.8040	26.41/0.7672	31.10/0.9150	27.37/0.8255
ICNet-L	IC	24.17/0.8165	27.96/0.8133	26.49/0.7698	31.09/0.9150	27.43/0.8287

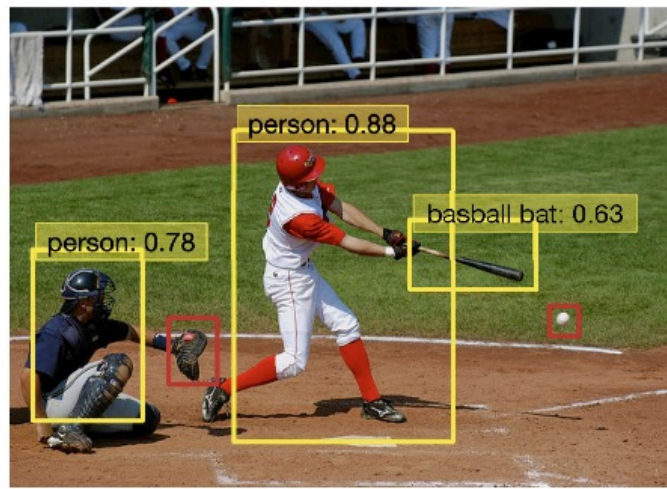
- 在Vimeo-90K-T、REDS4测试集，进行了与现有最先进方法的对比实验

Algorithm	Bicubic	TOFlow	RBPN	EDVR	IconVSR	ICNet-M	ICNet-L
Vimeo-90K-T	31.55/0.866	34.94/0.919	37.16/0.942	37.59/0.9475	37.47/0.9461	37.68/0.9474	37.72/0.9477
REDS4	26.14/0.7292	27.98/0.7990	29.54/0.8538	31.09/0.8800	31.67/0.8948	31.22/0.8825	31.71/0.8963

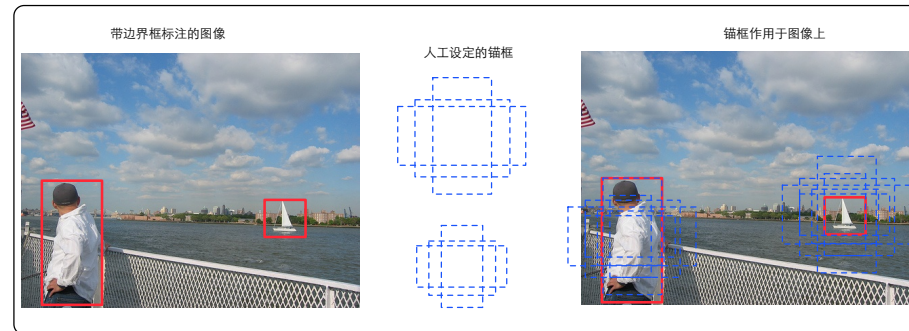
(红色表示最好的，蓝色代表次好)

Jiaxu Leng, Jia Wang, Xinbo Gao*. "ICNet: Joint Alignment and Reconstruction via Iterative Collaboration for Video Super-Resolution." In *Proceedings of the 30th ACM International Conference on Multimedia (ACM MM)* 2022.

MS COCO 数据集上的检测结果示例



针对问题：



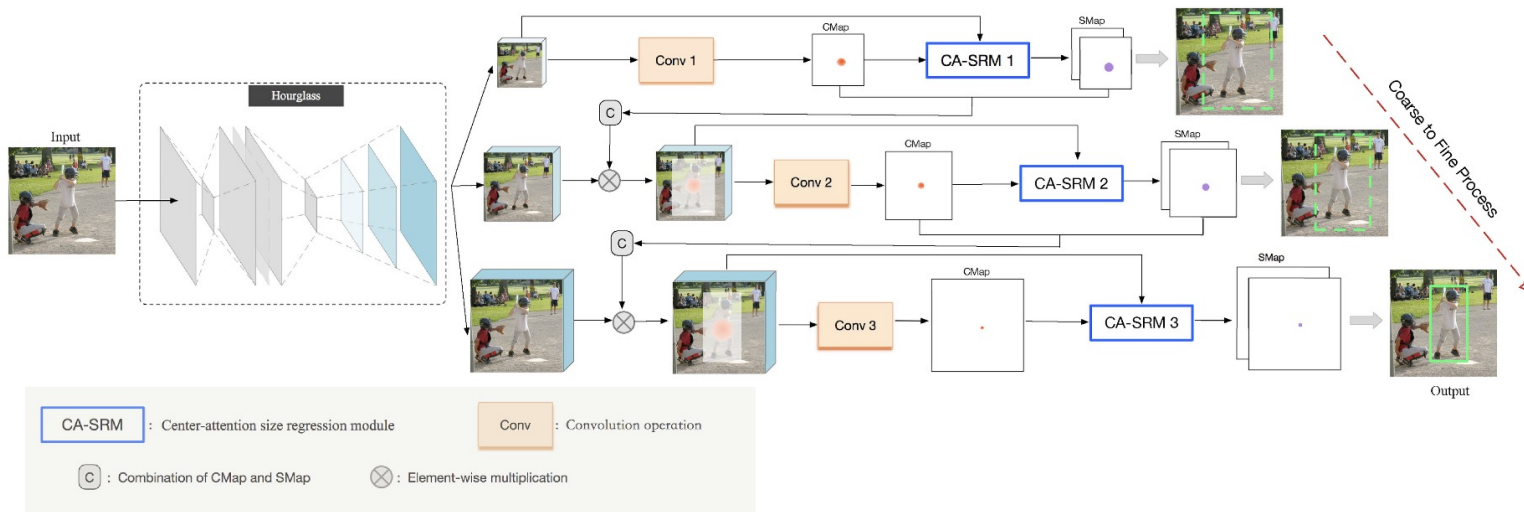
IoU 阈值为 0.7

船的正样本数为：62 << 人的正样本数为：165

	Object Size	Images	Objects
Large object	(96 × 96) to ($\infty \times \infty$)	51%	41%
Medium object	(32 × 32) to (96 × 96)	70%	34%
Small object	(0 × 0) to (32 × 32)	82%	24%

- ◆ 由于中心预测失败，用红色标示的小物体被遗漏；
- ◆ 一些预测的边界框没有紧紧围绕在物体周围，无法准确预测尺寸；
- ◆ **多尺度**问题普遍存在，需要使用层级式的网络结构。

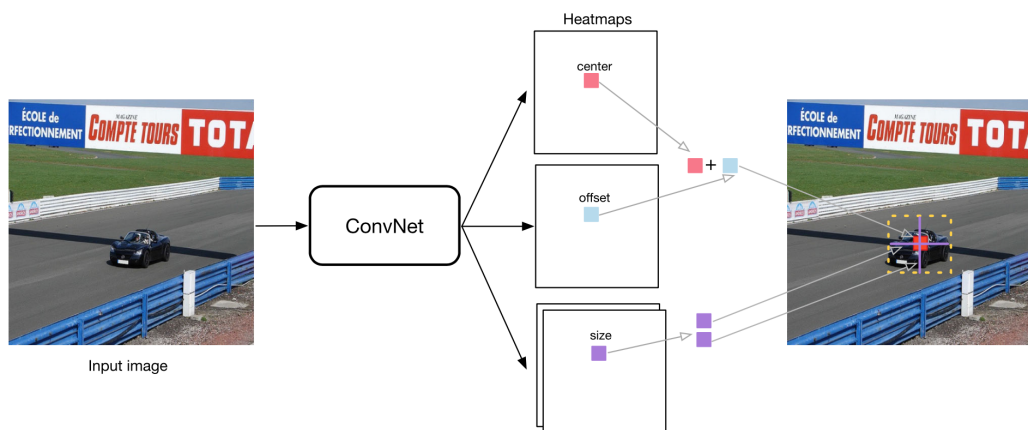
Jiaxu Leng, Ying Liu, Zhihui Wang, Haibo Hu, and Xinbo Gao. "Crossnet: detecting objects as crosses." *IEEE Transactions on Multimedia* (TMM) 2022.



主要创新:

- 设计了一种级联中心预测方法，引入了**从粗到细**的思路，以实现准确的中心定位；
- 设计了一种基于中心注意力的尺寸回归方法，它利用中心关键点的预测结果来改进宽度和高度的回归。

层级中心定位 → **可信**目标检测



中心回归损失

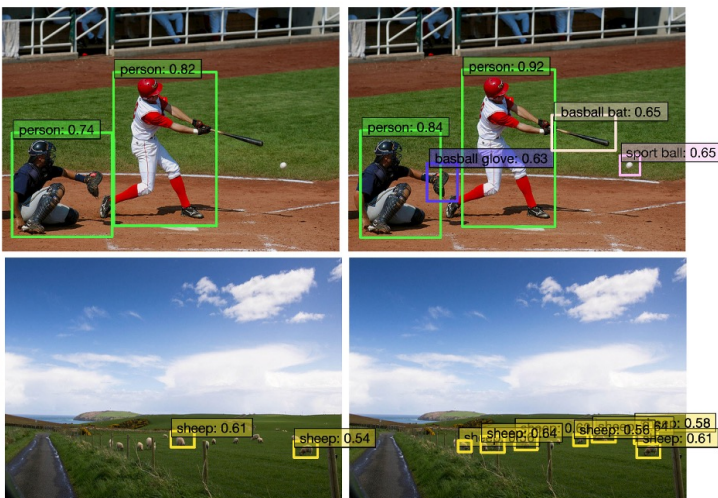
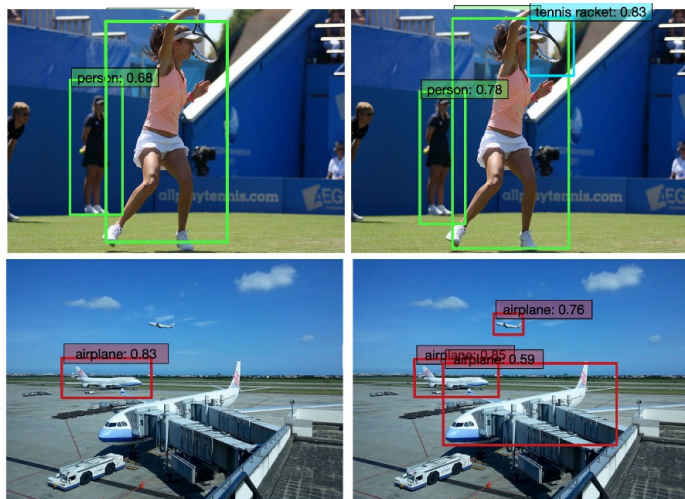
$$L_c^i = -\frac{1}{N} \sum_{xyc} \begin{cases} \frac{3}{i}(1 - \tilde{p}_{xyc})^\mu \log(\tilde{p}_{xyc}) & \text{if } p_{xyc} = 1 \\ (1 - p_{xyc})^\nu (\tilde{p}_{xyc})^\mu \log(1 - \tilde{p}_{xyc}) & \text{otherwise} \end{cases}$$

尺寸损失

$$L_s = \frac{1}{N} \sum_{n=1}^N |\tilde{S}_n - S_n|$$

偏移损失

$$L_o = \frac{1}{N} \sum_{n=1}^N |\tilde{O}_n - O_n|$$



CornerNet

CrossNet(Ours)



Method	FPS	Car			Pedestrian			Cyclist		
		Easy	Moderate	Hard	Easy	Moderate	Hard	Easy	Moderate	Hard
<i>anchor-based approaches:</i>										
Faster R-CNN [7]	0.5	88.97	83.16	72.62	79.97	66.24	61.09	72.40	62.86	54.97
Complexer-YOLO [54]	16.7	91.92	84.16	79.62	42.16	36.45	32.91	66.94	59.78	55.63
SubCNN [55]	0.5	94.26	89.98	79.78	84.88	72.27	66.82	79.36	71.72	62.74
Cascade MS-CNN [3]	4	94.26	91.60	78.84	-	-	-	-	-	-
YOLOv3 [9]	20	95.22	92.42	82.32	83.87	66.54	61.45	79.16	58.57	51.74
TuSimple[56]	0.6	95.12	94.47	86.45	88.87	78.40	73.66	83.68	75.22	65.22
RRC[57]	3.6	95.68	93.40	87.37	85.98	76.61	71.47	86.81	76.81	66.59
<i>3D object detection approaches:</i>										
SMOKE [58]	33.3	93.21	87.51	77.66	-	-	-	-	-	-
Mono3D [59]	0.2	94.52	89.37	79.15	80.30	67.29	62.23	77.19	65.15	57.88
Center3D [60]	20	95.14	85.05	73.06	67.15	48.76	44.05	65.34	40.99	36.50
F-PointNet [61]	5.9	95.85	95.17	85.42	89.83	80.13	75.05	86.86	73.16	65.21
PC-CNN-V2 [62]	2	96.06	95.20	89.37	-	-	-	-	-	-
SegVoxelNet [63]	25	96.00	92.73	87.60	-	-	-	-	-	-
STD [64]	13	96.14	93.22	90.53	68.33	55.04	50.85	83.99	71.63	64.92
HotSpotNet [65]	25	96.21	92.81	89.80	71.43	62.31	59.24	86.06	78.81	71.74
CrossNet w/ DLA-34	20	96.44	93.52	91.23	87.71	76.01	68.85	84.78	72.73	66.12
CrossNet w/ Hourglass-104	5	97.63	94.73	92.53	89.11	78.11	70.21	85.88	74.62	68.31

Jiaxu Leng, Ying Liu, Zihui Wang, Haibo Hu, and Xinbo Gao. "Crossnet: detecting objects as crosses." *IEEE Transactions on Multimedia* (TMM) 2022.

研究动机:

与基于CNN的行人检测器相比，DETRs在行人检测中产生了更多的假阳性。而在拥挤的场景中，传统样本选择方法会选择一些**无法用于训练**的样本。

主要贡献:

- 提出了一种**约束引导**的标签分配方法，以避免在DETRs的训练中选择无法学习的样本；
- 提出了一种**适用性感知的**焦点损失，可以根据每个样本的适用性为每个样本提供**自适应权重**，以实现更好的训练效果。

选择性学习 → **可信行人检测**

方法设计:

● Constraint-guided Label Assignment



$$\begin{aligned} \therefore L1(x_p, x_g) &\leq \alpha w_g \\ L1(y_p, y_g) &\leq \alpha h_g \\ IoU(b, p) &> \beta \\ \therefore C_{cenx} &= 0 \\ C_{ceny} &= 0 \\ C_{pos} &= 0 \end{aligned}$$

- Center of the GT box
- Center of the prediction box
- The GT box
- The Prediction box
- The range of center constraint

● Utilizability-aware Focal Loss

$$L_{UAFL} = -|y - p|^Y (y \log p + (1 - y) \log(1 - p)),$$

当预测等于其交并比（IoU）时，损失是最优的。

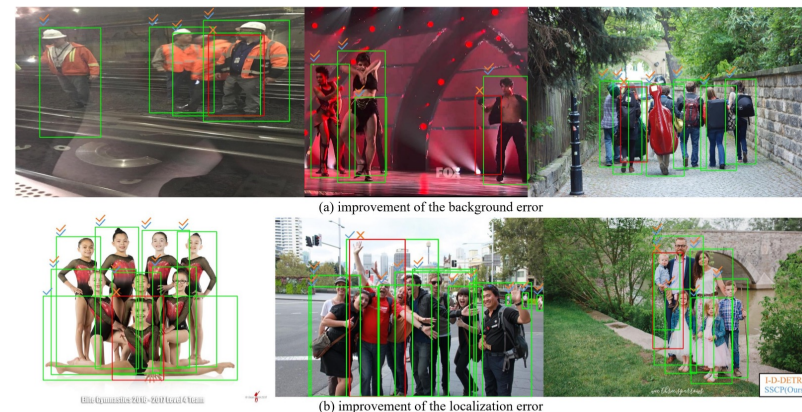
- 优化目标是在训练过程中避免因在拥挤场景中包含少量特征的样本而导致的过度损失；
- 自适应 γ ，可以根据样本自动调节损失权重。

实验结果:

● Crowdhuman 数据集上的性能对比

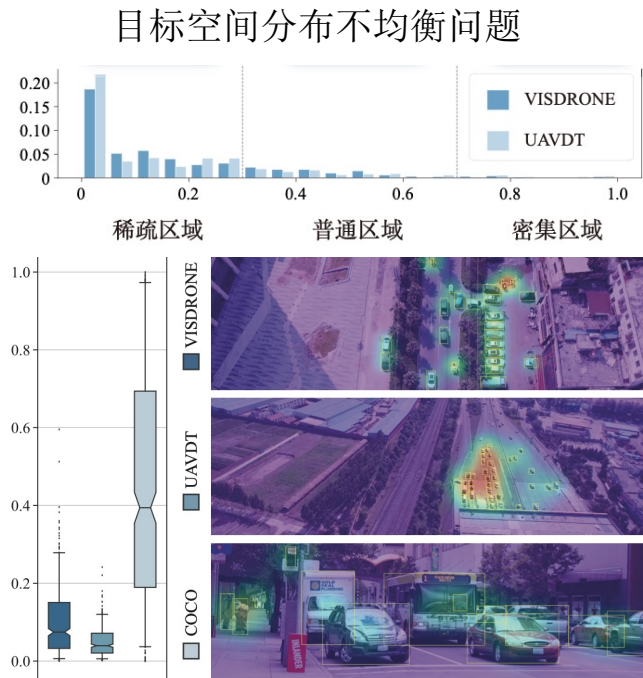
Model	Backbone	MR↓
FPN (CVPR'17)	ResNet-50	42.9
MIP (CVPR'20)	ResNet-50	41.4
R2NMS (CVPR'20)	ResNet-50	43.4
AEVB (CVPR'21)	ResNet-50	40.7
AutoPedestrian (TIP'21)	ResNet-50	40.6
DMSFLN(TITS'21)	VGG-16	43.6
OAF-Net(TITS'22)	HRNet-w32	45.0
D-DETR (ICLR'21)	ResNet-50	44.6
D-DETR+Ours	ResNet-50	42.0
Iter-D-DETR (CVPR'22)	ResNet-50	41.5
Iter-D-DETR (our implementation)	ResNet-50	41.9
Iter-D-DETR+Ours	ResNet-50	39.7

可视化分析:



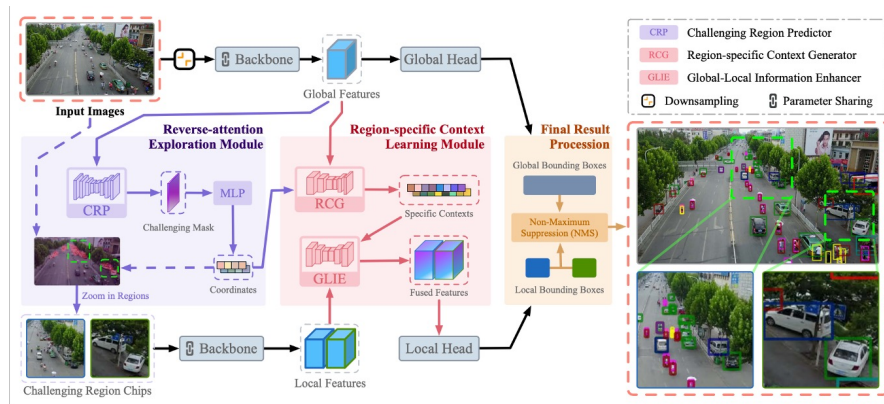
Feng Gao, **Jiaxu Leng***, Ji Gan, and Xinbo Gao*. "Selecting Learnable Training Samples is All DETRs Need in Crowded Pedestrian Detection." In Proceedings of the 31st ACM International Conference on Multimedia (ACM MM) 2023.

研究动机：

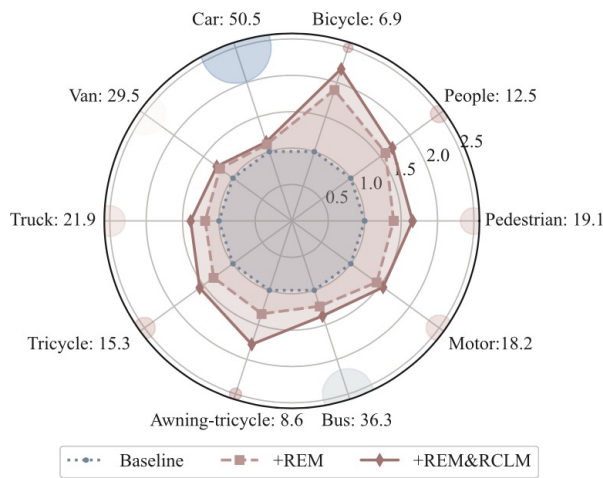


区域差别处理 → 可信目标检测

网络结构：

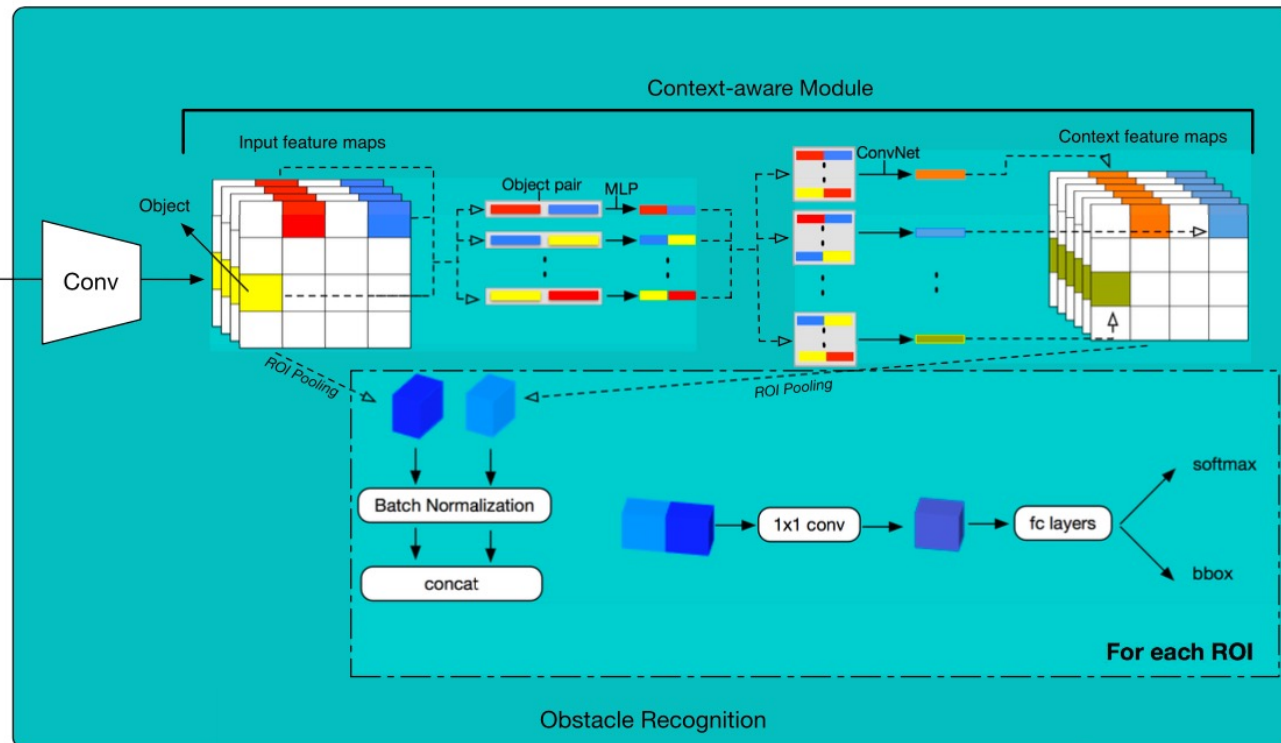
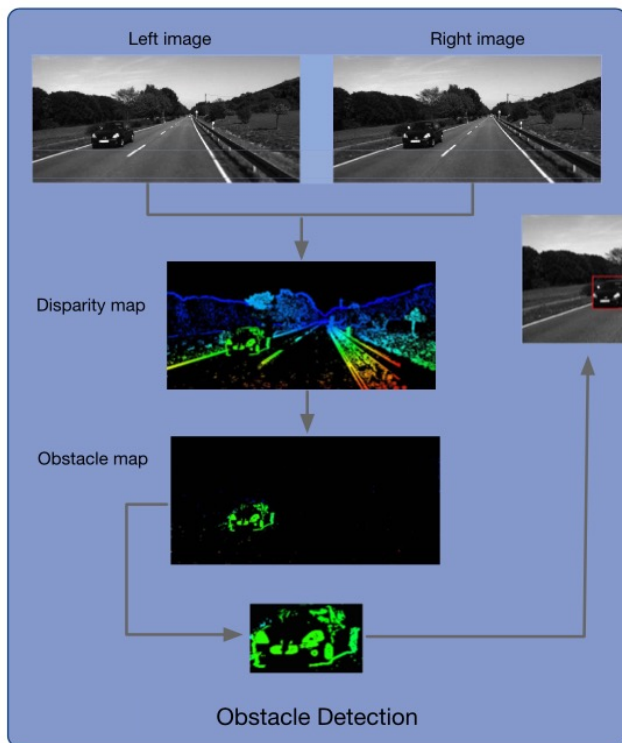


实验结果：



红色标识了图像中预测的困难目标所在区域

- ◆ 提出一种基于帕累托法则的重聚焦检测网络；
- ◆ 通过**反向注意力**引导网络从图像中区分出**挑战性区域**；
- ◆ 并使用**区域独有上下文**辅助检测器更好地识别**困难目标**。



- ◆ 我们的方法旨在检测和识别**道路上的障碍物**，而不是视野中的所有障碍物；
- ◆ 提出了一种特殊的地图，称为现实的U-disparity图，它有助于检测道路上远处的障碍物；
- ◆ 提出了一种**上下文感知模块**，并将其有效地纳入Faster-RCNN的架构中。

道路上下文学习



可信障碍物检测

Jiaxu Leng, Ying Liu, Dawei Du, Tianlin Zhang, and Pei Quan. "Robust obstacle detection and recognition for driver assistance systems." *IEEE transactions on intelligent transportation systems (TITS)* 2020.

实验结果表明，所提出的方法显著提升了识别精度，特别是对小尺寸的和被遮挡的障碍物。

Method	Dataset	Runtime per frame	Car			Pedestrian			Cyclist		
			Moderate	Easy	Hard	Moderate	Easy	Hard	Moderate	Easy	Hard
DPM-VOC+VP[33]	KITTI	8s	66.2%	80.4%	49.8%	44.8%	59.6%	40.3%	31.1%	43.6%	28.2%
LSVM-MDPM-sv[34]	KITTI	10s	57.4%	71.7%	46.5%	39.3%	51.7%	35.9%	29.2%	37.7%	27.5%
Mono3D[30]	KITTI	4.2s	87.8%	90.2%	78.1%	66.7%	77.3%	63.4%	63.8%	75.2%	58.9%
NMRDO[31]	KITTI	0.1s	58.4%	72.6%	47.5%	41.2%	52.2%	36.8%	30.8%	38.4%	28.3%
SubCNN [32]	KITTI	2s	88.8%	90.7%	79.2%	71.3%	83.1%	66.3%	70.7%	77.8%	62.7%
Ours	KITTI	0.2s	89.2%	91.2%	84.8%	83.4%	87.8%	75.1%	79.5%	84.0%	70.2%

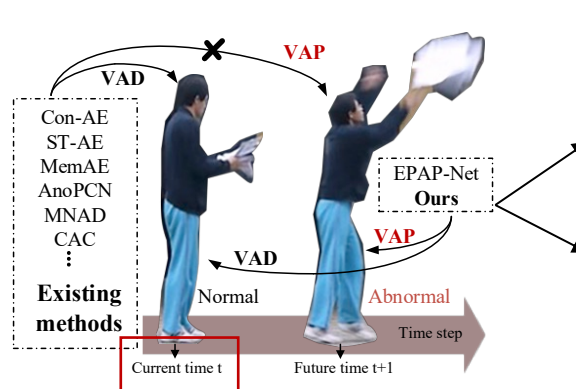


Jiaxu Leng, Ying Liu, Dawei Du, Tianlin Zhang, and Pei Quan. "Robust obstacle detection and recognition for driver assistance systems." *IEEE transactions on intelligent transportation systems (TITS)* 2020.

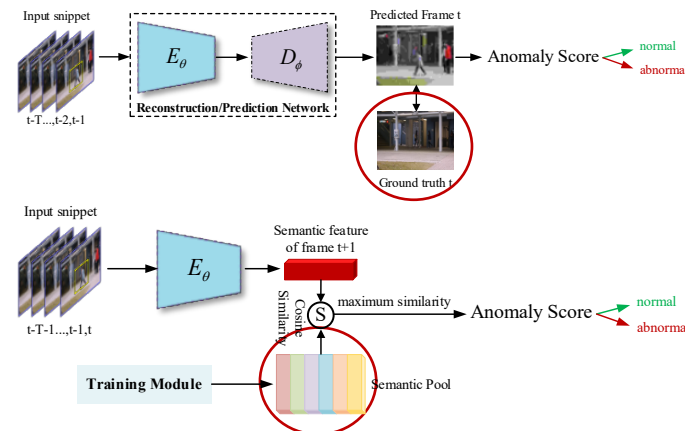
研究动机：



主要工作：



基于预测的视频异常检测流程（现有方法）



基于语义的视频异常预测流程（我们的方法）

- ◆ 由于目前的视频异常检测算法在测试阶段依赖于Ground Truth进行异常判断，所以只能对已经发生的异常进行检测，针对潜在异常预测不到的问题，我们提出了**基于语义聚类的视频异常预测框架**；
- ◆ 我们分别在特征层面和视频帧层面引导模型获得未来帧的语义特征表达，并构建存储正常语义模式的**语义池**，当未来帧的语义模式不属于该语义池时，则将其视为异常视频帧，从而实现**对异常的预测**。

- Song等人在《Science》上的研究表明：人类行为有93%的潜在**可预测性**。
- 从人类认知出发：人类具有**相应的记忆**来判断未来的行为是否符合当前场景的客观规律。

语义特征记忆 ➡ **可信异常预测**

Jiaxu Leng, Mingpi Tan, Xinbo Gao*, Wen Lu, and Zongyi Xu. "Anomaly Warning: Learning and Memorizing Future Semantic Patterns for Unsupervised Ex-ante Potential Anomaly Prediction." In *Proceedings of the 30th ACM International Conference on Multimedia (ACM MM) 2022*.

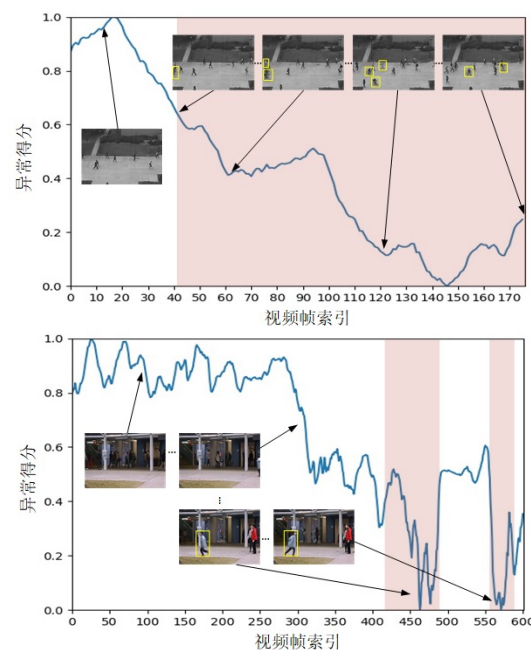
算法AUC值对比：

方法	UCSD Ped2		CUHK Avenue		ShanghaiTech	
	VAD(%)	VAP(%)	VAD(%)	VAP(%)	VAD(%)	VAP(%)
ConvAE ^[4]	85.0	-	80.0	-	60.9	-
ConvLSTM-AE ^[6]	88.1	-	77.0	-	-	-
STAE ^[20]	91.2	-	80.9	-	-	-
MemAE ^[8]	94.1	-	83.3	-	71.2	-
GMFC-VAE ^[7]	92.2	-	83.4	-	-	-
MNAD-R ^[10]	90.2	-	82.8	-	69.8	-
MNAD-R+VT	90.4	-	83.0	-	-	-
CDAE ^[38]	96.5	-	86.0	-	73.3	-
Framepred ^[16]	95.4	-	85.1	-	72.8	-
Framepred+ ^[40]	96.2	-	85.7	-	73.0	-
AnoPCN ^[21]	96.8	-	86.2	-	73.6	-
MNAD-P ^[10]	96.9	-	88.5	-	70.5	-
MNAD-P+VT	97.0	-	88.9	-	-	-
AMMC-Net ^[11]	96.6	-	86.6	-	73.7	-
MONAD ^[70]	97.2	-	86.4	-	70.9	-
STCEN ^[71]	96.9	-	86.6	-	73.8	-
VPC ^[72]	93.6	-	85.4	-	-	-
Baseline ^[10]	94.3	71.3	84.5	67.5	66.8	57.9
EPAP-Net	94.9	85.4	87.1	70.6	70.8	59.9

EPAP-Net 首次提出**异常预测**任务，并在多个数据集上实现了具有竞争力的检测性能

Jiaxu Leng, Mingpi Tan, Xinbo Gao*, Wen Lu, and Zongyi Xu. "Anomaly Warning: Learning and Memorizing Future Semantic Patterns for Unsupervised Ex-ante Potential Anomaly Prediction." In *Proceedings of the 30th ACM International Conference on Multimedia (ACM MM) 2022*.

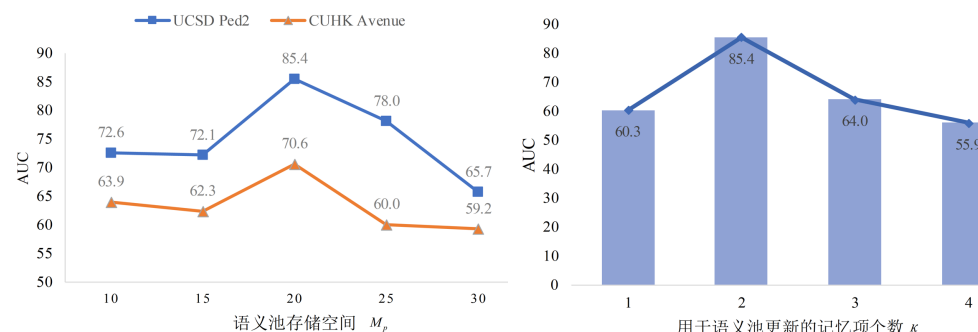
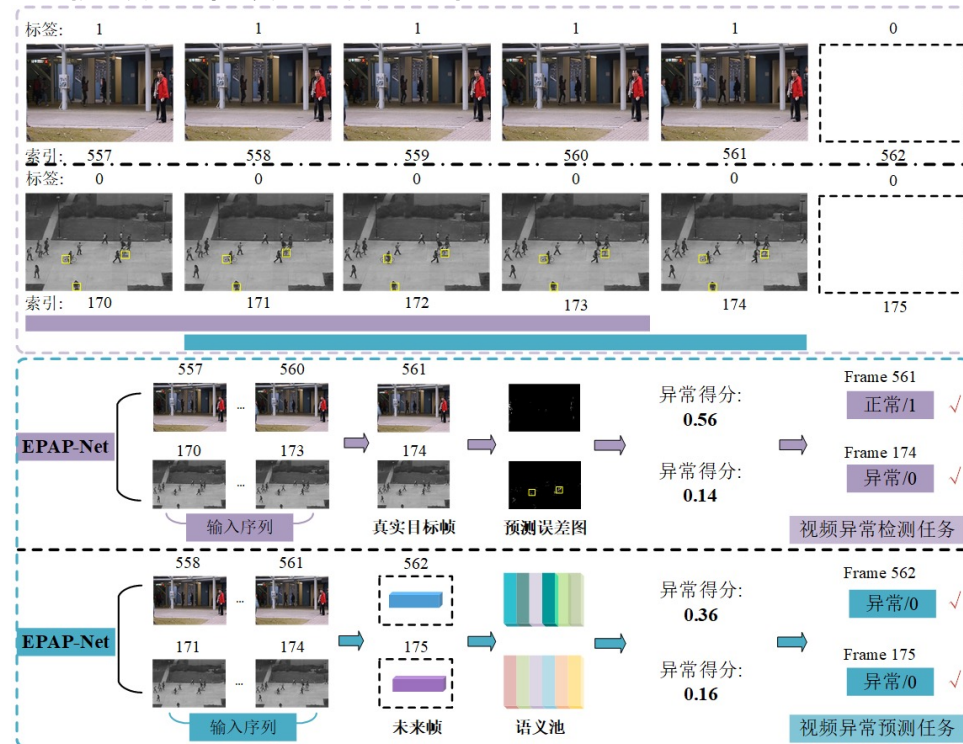
异常得分曲线：



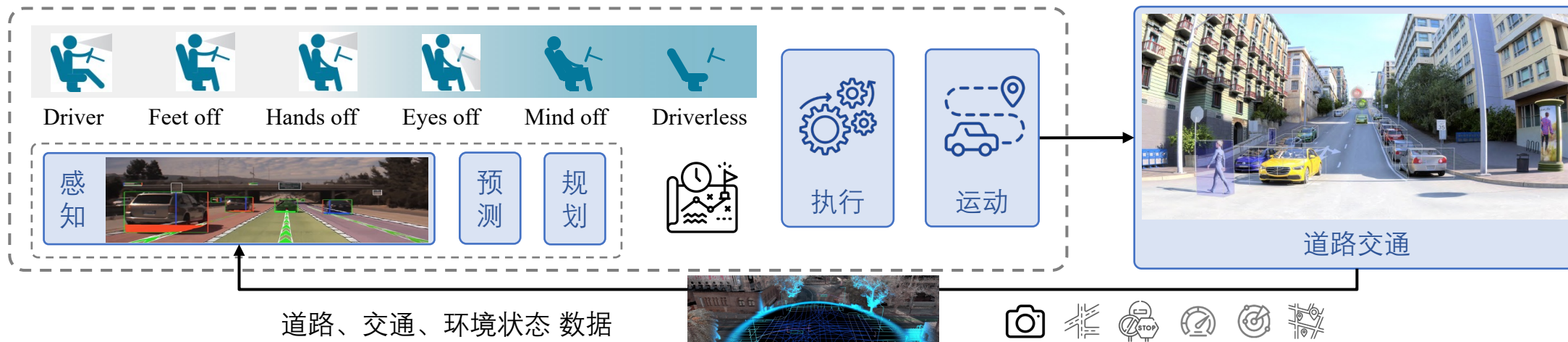
消融实验结果：

模型版本	TSM	CSE	SSLoss	UCSD Ped2
EPAP-Netv0	×	×	×	71.3
EPAP-Netv1	×	✓	×	80.1
EPAP-Netv2	×	×	✓	79.6
EPAP-Netv3	✓	×	✓	83.2
EPAP-Net	×	✓	✓	85.4

视频异常预测的先进性：



面向自动驾驶的脑启发可信场景分析



目的

看清场景

看见要素

看懂内容

挑战

复杂场景

微小目标

模糊语义

方法

反馈学习

多粒度感知

上下文推理

联想记忆

展望

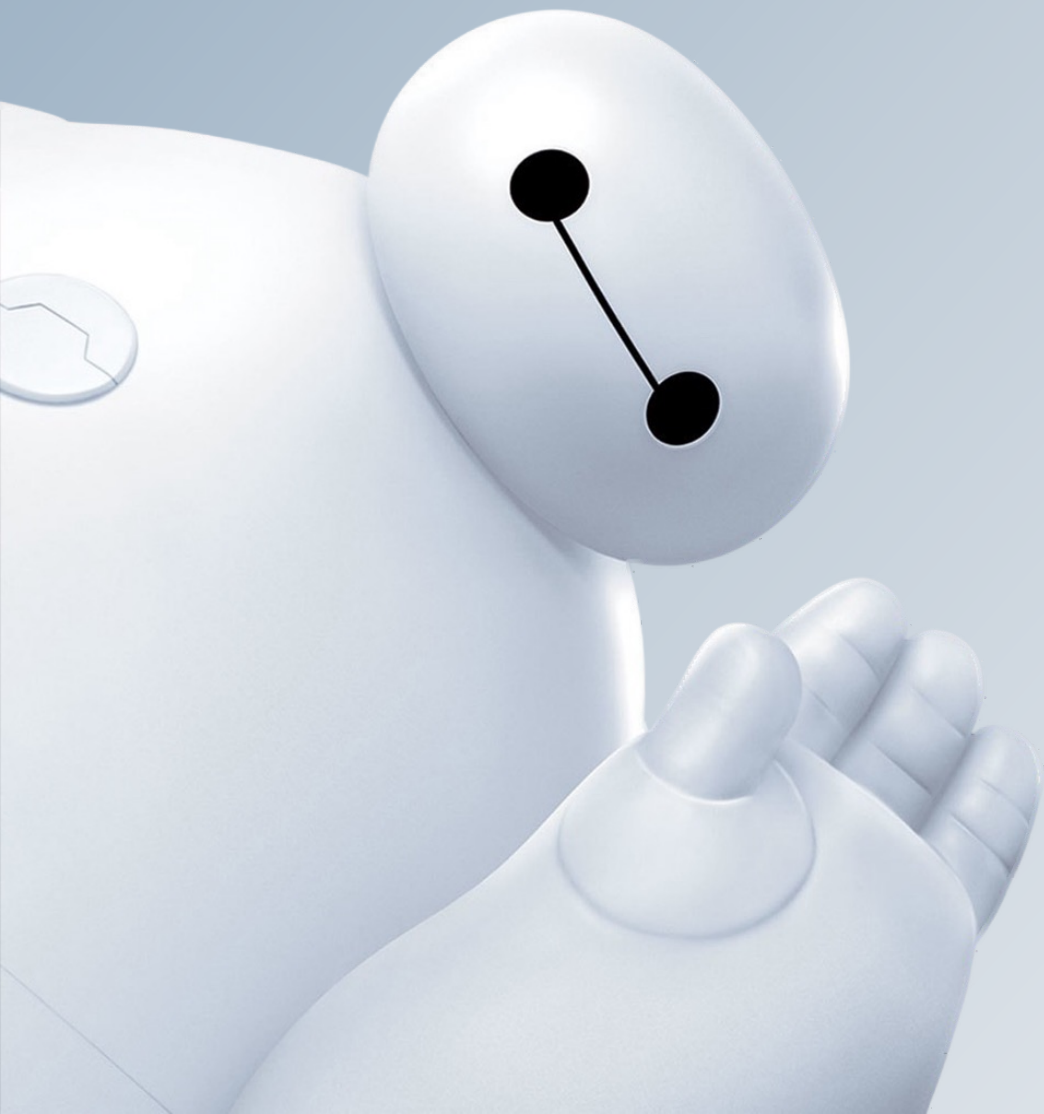
脑启发

脑参与

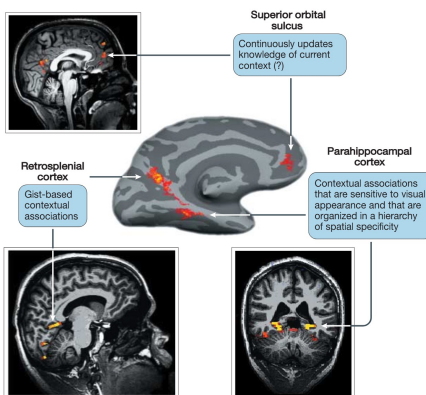
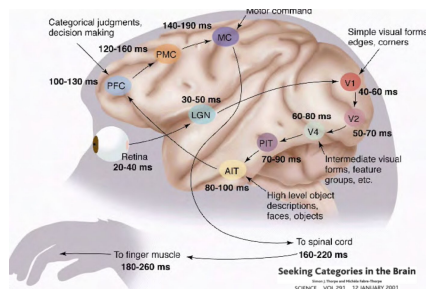
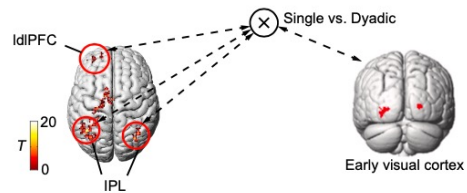
脑认知

汇报大纲

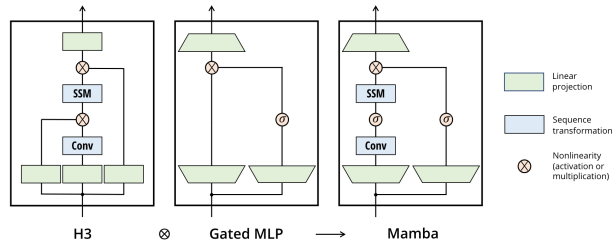
- 可信场景分析：以人为本
- 类脑智能计算：以脑为鉴
- 团队工作简介：类脑计算
- **未来发展趋势：脑机互鉴**



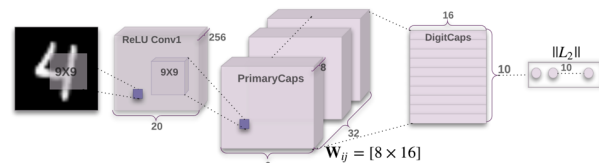
AI from Brain (脑启发)



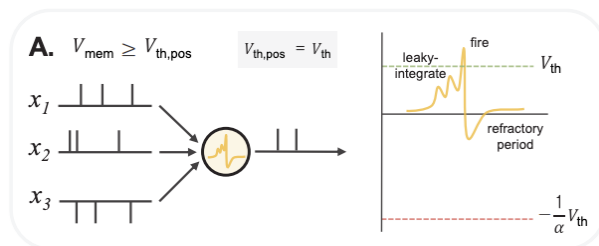
模仿学习机制



Mamba

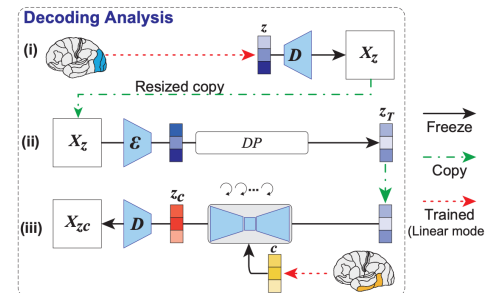


胶囊网络

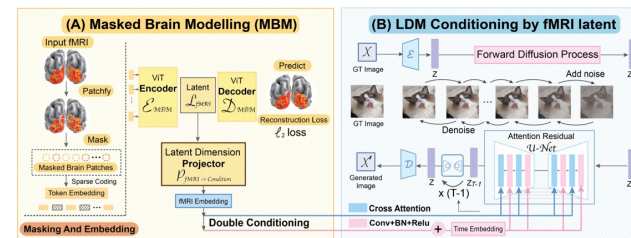


脉冲神经网络

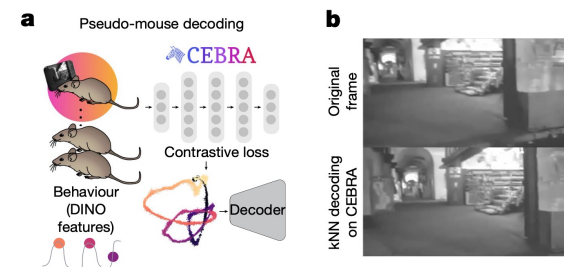
AI with Brain (脑参与)



高清图像重建

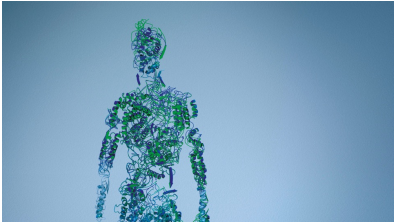


视觉解码

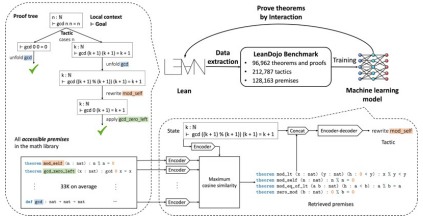


未来帧预测

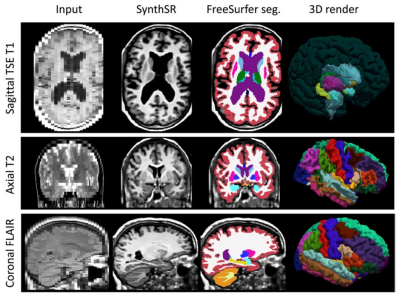
Scientific progress accelerates even further, thanks to AI



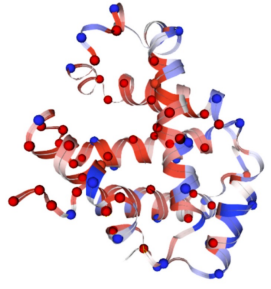
蛋白质解码 (AlphaFold)



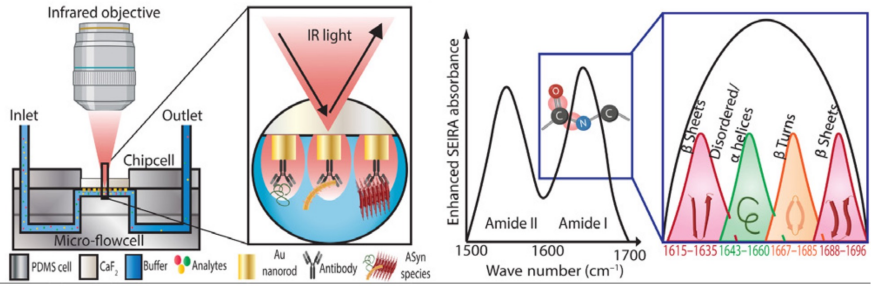
数学定理证明 (LeanDojo)



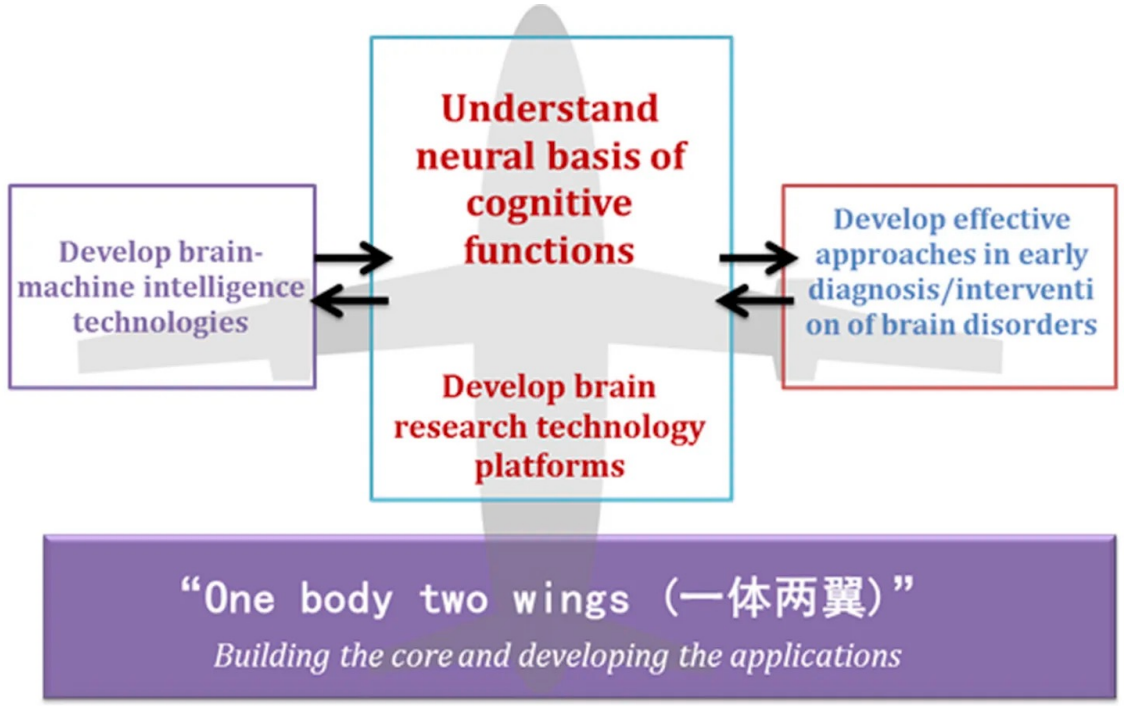
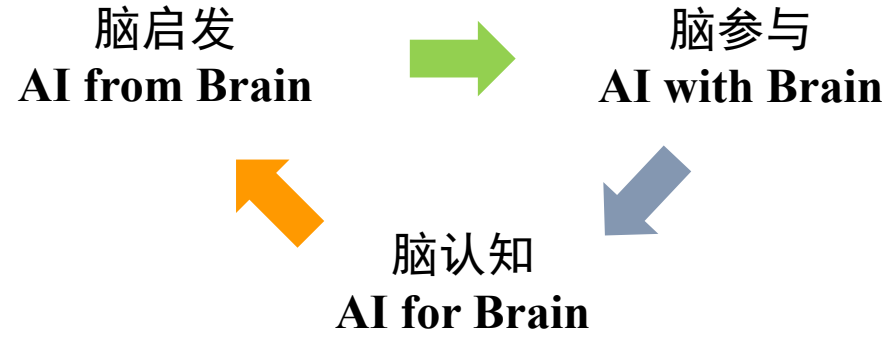
3D脑部渲染 (SynthSR)



突变基因预测 (AlphaMissense)



神经退行性疾病诊断 (Coupled Plasmonic Infrared Sensors)



Thanks for your listening and attention!



Questions, please?



团队微信公众号



个人主页



Google 学术主页

WORKSHOP: 端到端自动驾驶

讲者: 冷佳旭 (Jiaxu Leng)



视觉与学习青年学者研讨会

May 7th, 2024 @ VALSE