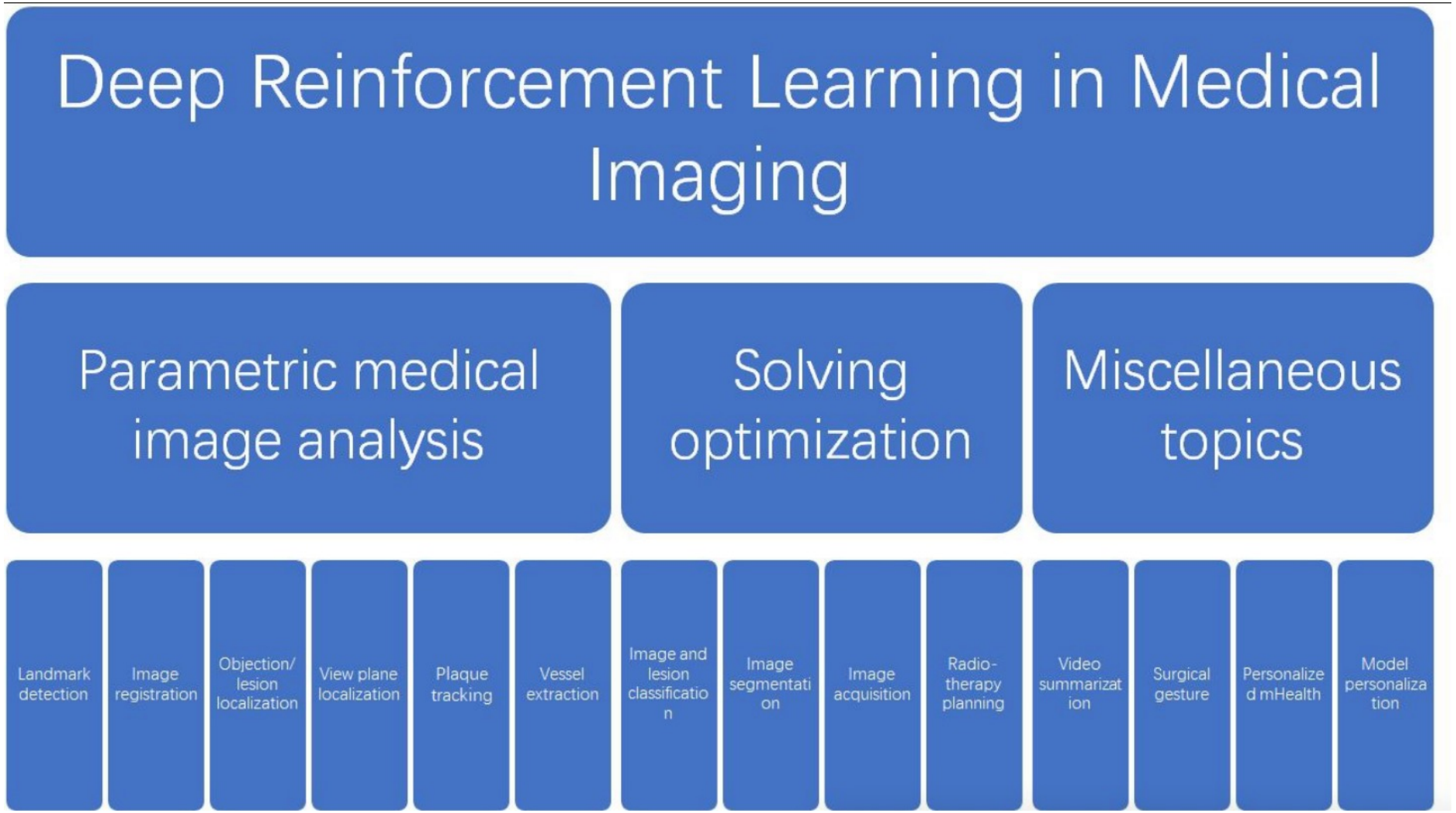
模式识别
Pattern Recognition

李泽榉，复旦大学 生物医学工程与技术创新学院

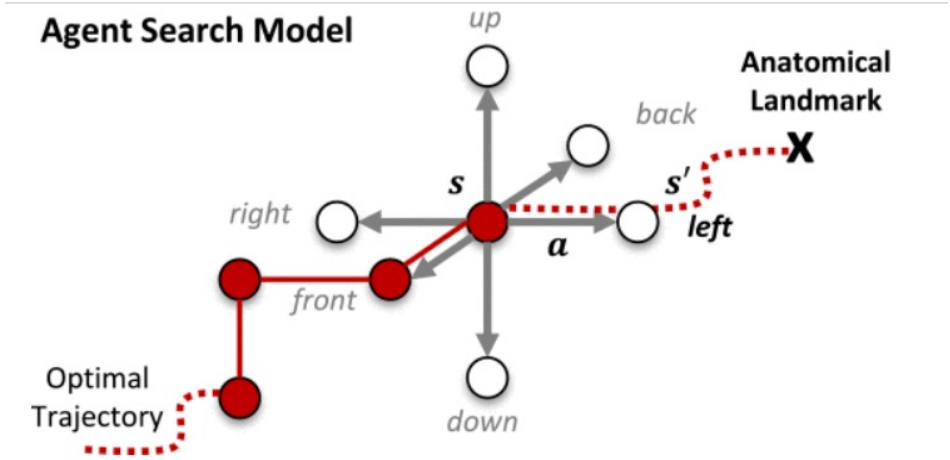
目录

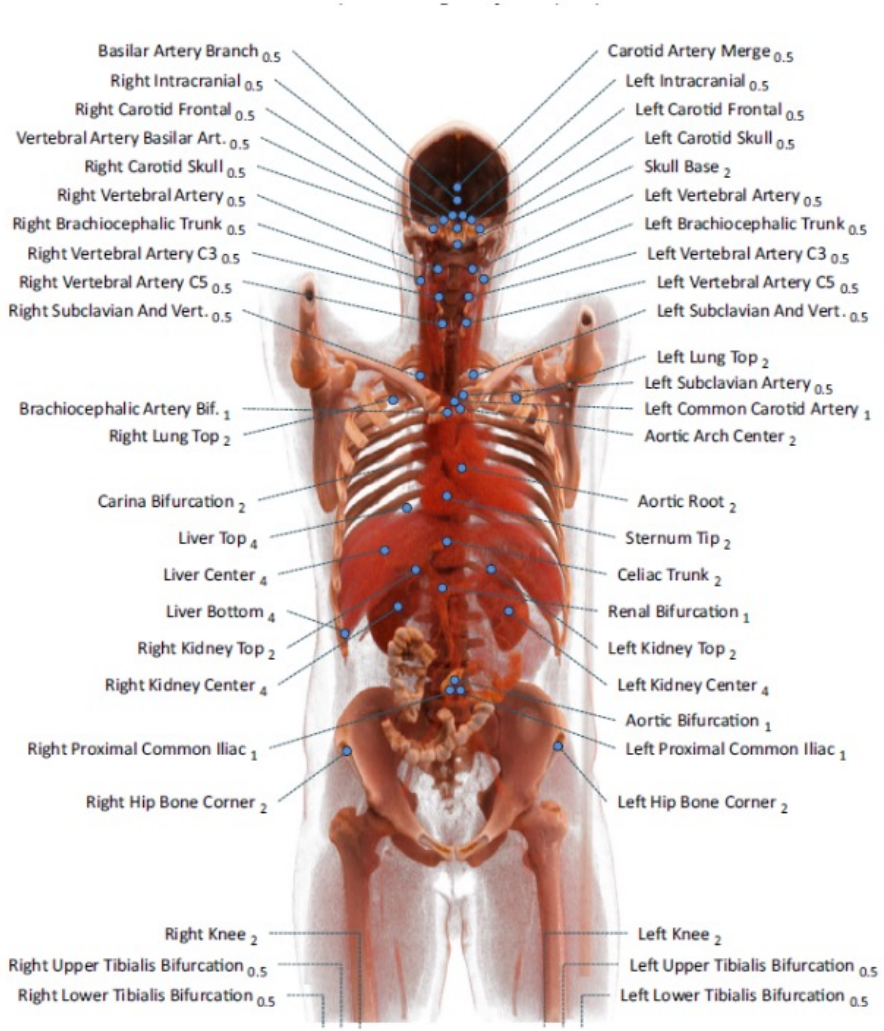
RL在医学图像中的应用



3D-Landmark Detection

- 在CT图像中检测关键点

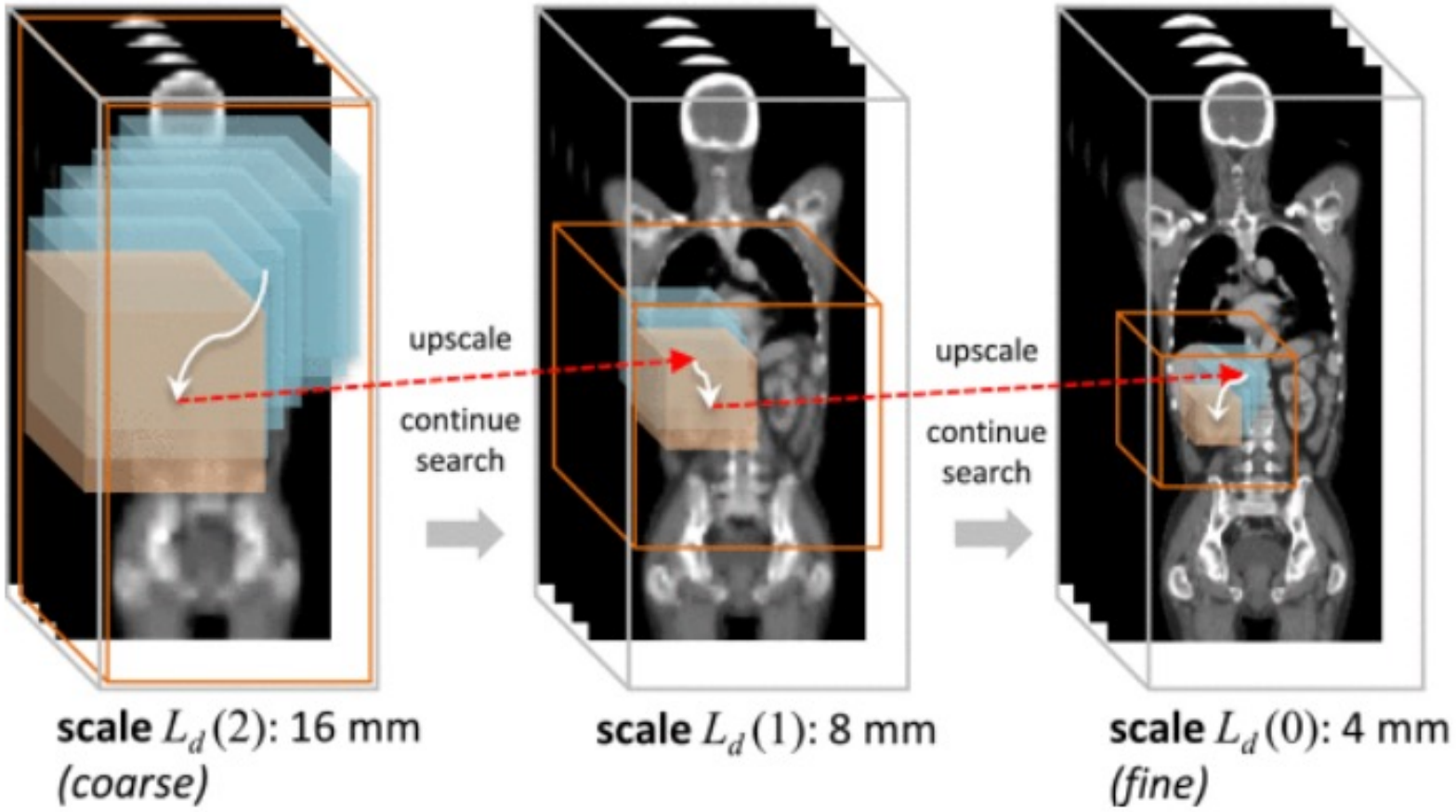




Ghesu F C, Georgescu B, Zheng Y, et al. Multi-scale deep reinforcement learning for real-time 3D-landmark detection in CT scans[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 41(1): 176-189.

3D-Landmark Detection

- 基于多尺度Q-Learning检测关键点



Ghesu F C, Georgescu B, Zheng Y, et al. Multi-scale deep reinforcement learning for real-time 3D-landmark detection in CT scans[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 41(1): 176-189.

3D-Landmark Detection

- 时间快，效果好

	Failed Cases		
Method	**% Failed**	**Median**	**Max**
PBT [3]	4.35%	38.89	141.88
ExtRTrees [10]	8.07%	20.27	460.22
Overfeat [22]	9.31%	35.64	231.29
3D-DL [18]	0.62%	10.17	10.17
SADNN [1]	0%	–	–
Ours	**0%**	–	–
PBT [3]	3.75%	14.30	28.57
ExtRTrees [10]	6.25%	13.43	17.44
Overfeat [22]	3.75%	127.52	242.05
3D-DL [18]	2.50%	262.81	513.81
SADNN [1]	1.25%	12.57	12.57
Ours	**0%**	–	–

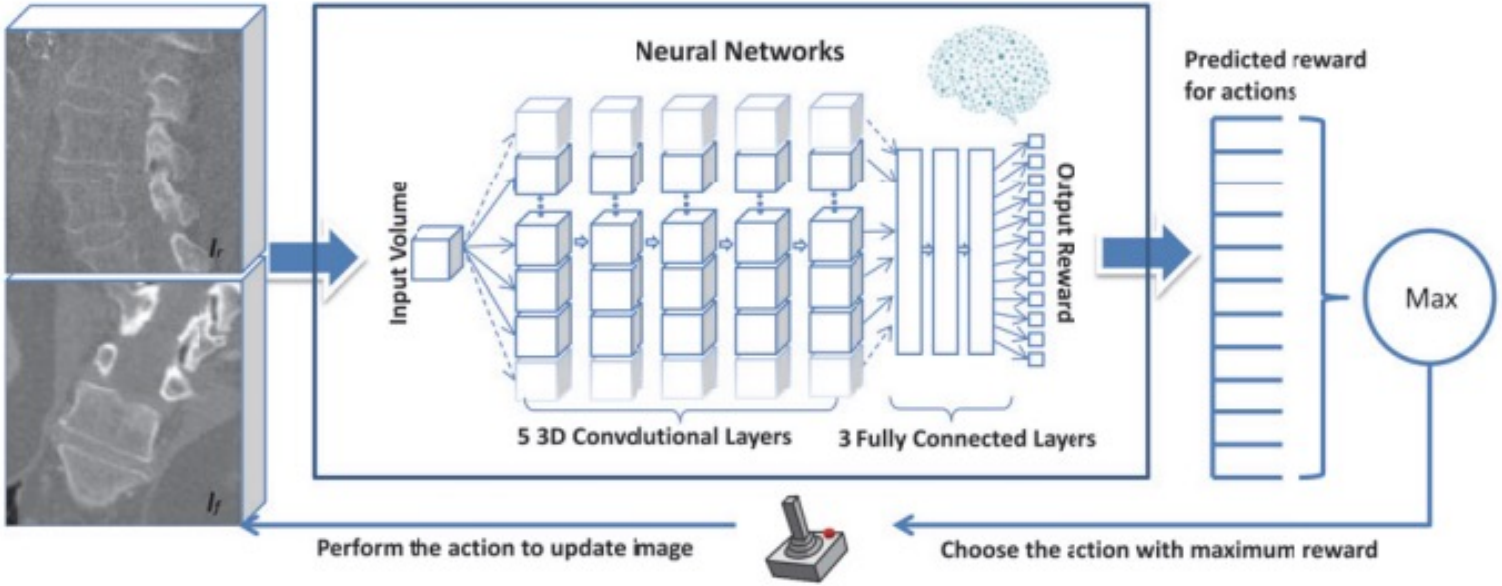
Solution	Dataset Size (Data/Patients)	Accuracy (mm)	Speed (seconds)
Zhan *et al.* [27]	18/18 CT	4.72	4
Fenchel *et al.* [35]	31/31 MR	22.4	20
Criminisi *et al.* [12]	100/– CT	17.60	1
Pauly *et al.* [32]	33/33 MR	14.95	0.8
Cuingnet *et al.* [11]	233/89 CT	10.5	2.8
Donner *et al.* [10]	20/20 CT	5.25	120
Criminisi *et al.* [31]	400/– CT	13.50	4
Chu *et al.* [30]	10/10 CT	1.90[1]	30
Potesil *et al.* [37]	83/83 CT	4.70	N/A
de Vos *et al.* [24]	100/– CT	4.80	10
Ours	**1487/532** CT	**4.19**[2]	**0.061**

[1] Evaluated only on vertebrae localization with strong priors.
[2] With no failures of clinical significance. All other solutions did not provide any information in this respect.

Ghesu F C, Georgescu B, Zheng Y, et al. Multi-scale deep reinforcement learning for real-time 3D-landmark detection in CT scans[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 41(1): 176-189.

Registration

- 利用Q-Learning来做刚性配准

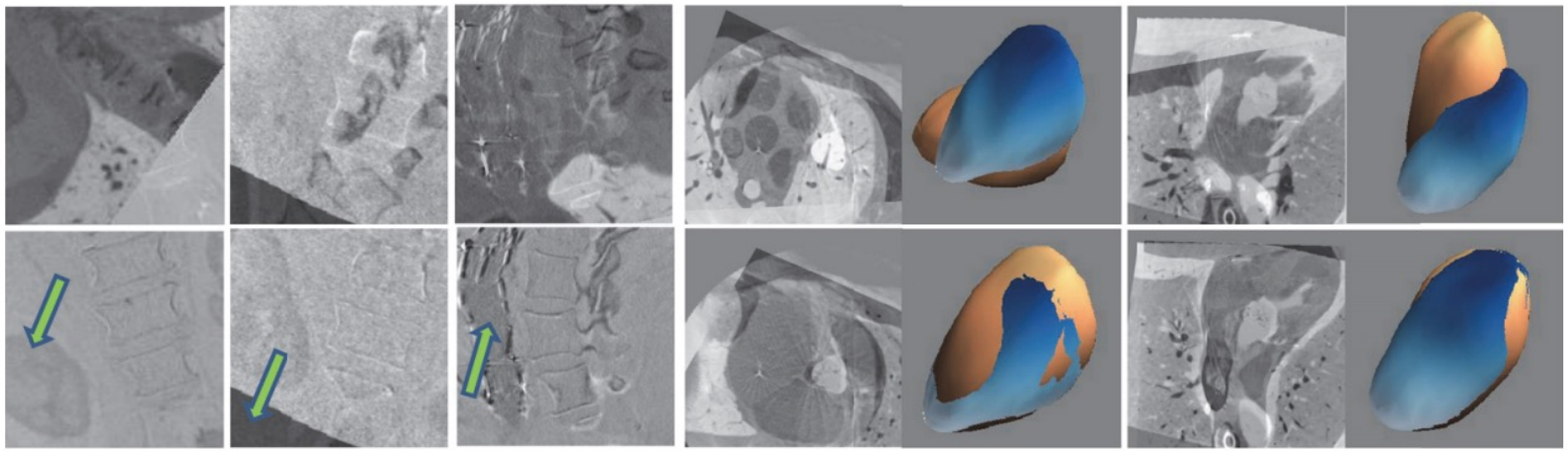


Liao R, Miao S, de Tournemire P, et al. An artificial agent for robust image registration[C]//Proceedings of the AAAI conference on artificial intelligence. 2017, 31(1).

Registration

- 可以更鲁棒地将不同样本配对在一起

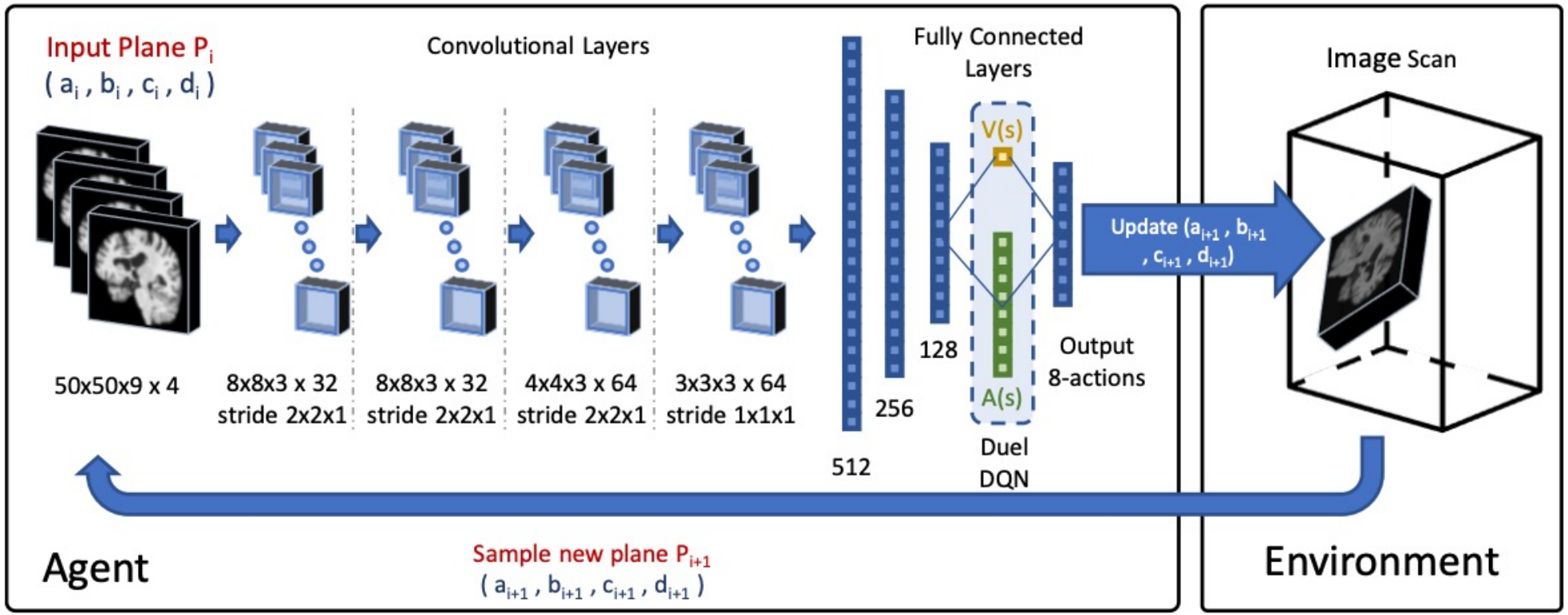
	Spine (E1) (TRE mm)				Heart (E2) (MME mm)			
Methods	**Success**	**10th**	**50th**	**90th**	**Success**	**10th**	**50th**	**90th**
Ground Truth	N/A	0.8	0.9	1.2	N/A	2.1	4.0	5.9
Initial Position	N/A	35.5	73.9	116.2	N/A	9.2	22.8	30.5
ITK(Ibanez et al. 2005)	12%	1.9	77.3	130.4	14%	14.9	34.9	47.6
Quasi-global(Miao et al. 2013)	20%	1.6	60.9	136.2	14%	16.2	35.9	58.7
Semantic registration(Neumann et al. 2015)	24%	3.0	34.9	71.0	72%	7.6	15.3	30.6
Proposed method	92%	1.7	2.5	3.8	100%	3.2	4.8	6.9
Human registration	70%	0.8	1.6	15.8	96%	4.0	6.2	13.4



Liao R, Miao S, de Tournemire P, et al. An artificial agent for robust image registration[C]//Proceedings of the AAAI conference on artificial intelligence. 2017, 31(1).

View Plane Localization

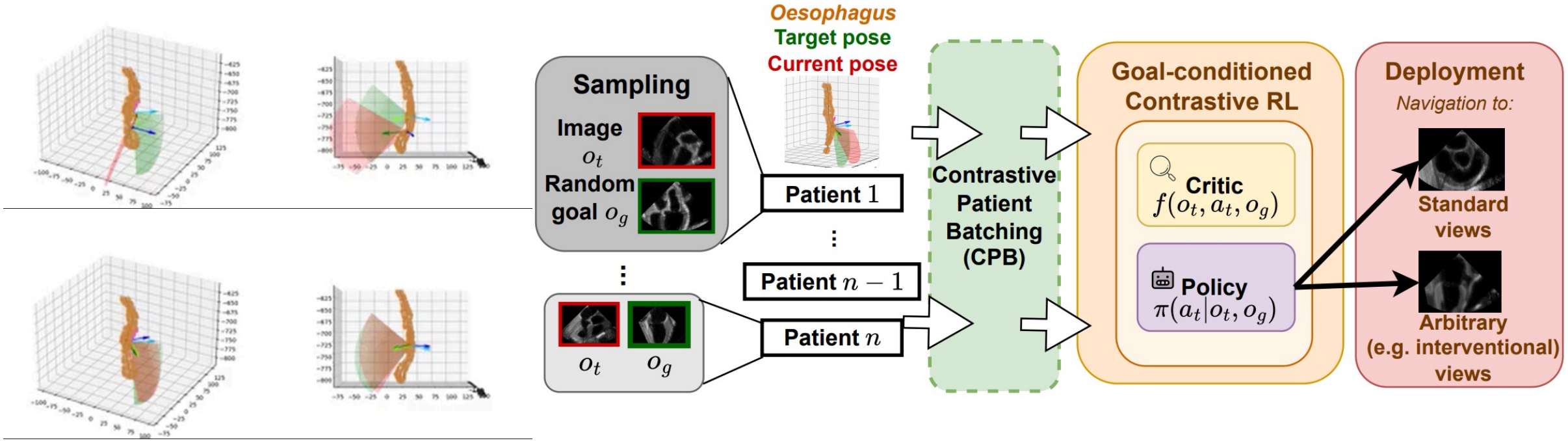
- 找到目标2D平面



Alansary A, Folgoc L L, Vaillant G, et al. Automatic view planning with multi-scale deep reinforcement learning agents[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer International Publishing, 2018: 277-285.

Ultrasound Navigation

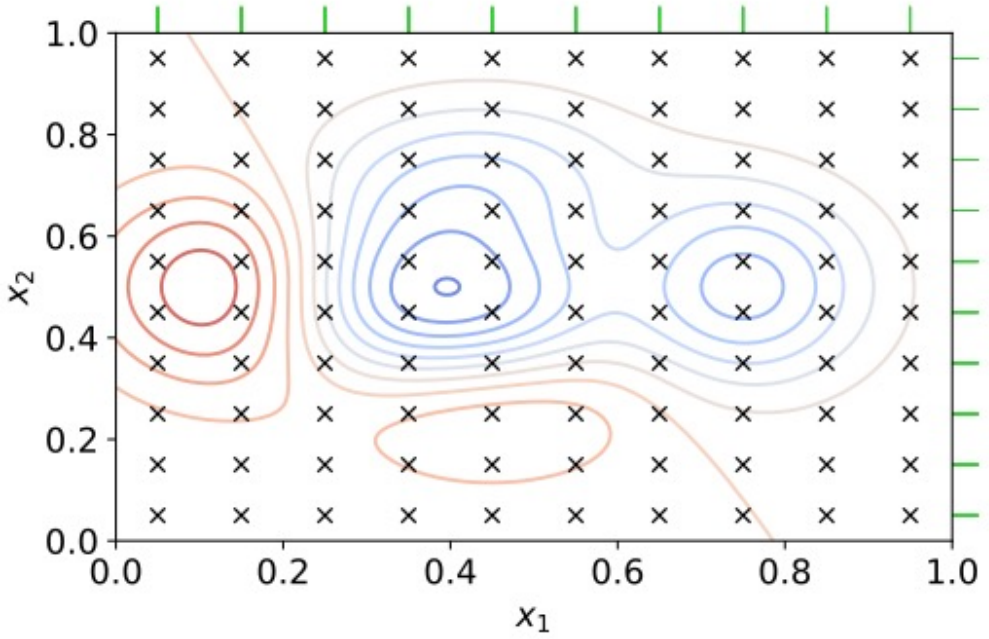
- 找到超声心动图平面

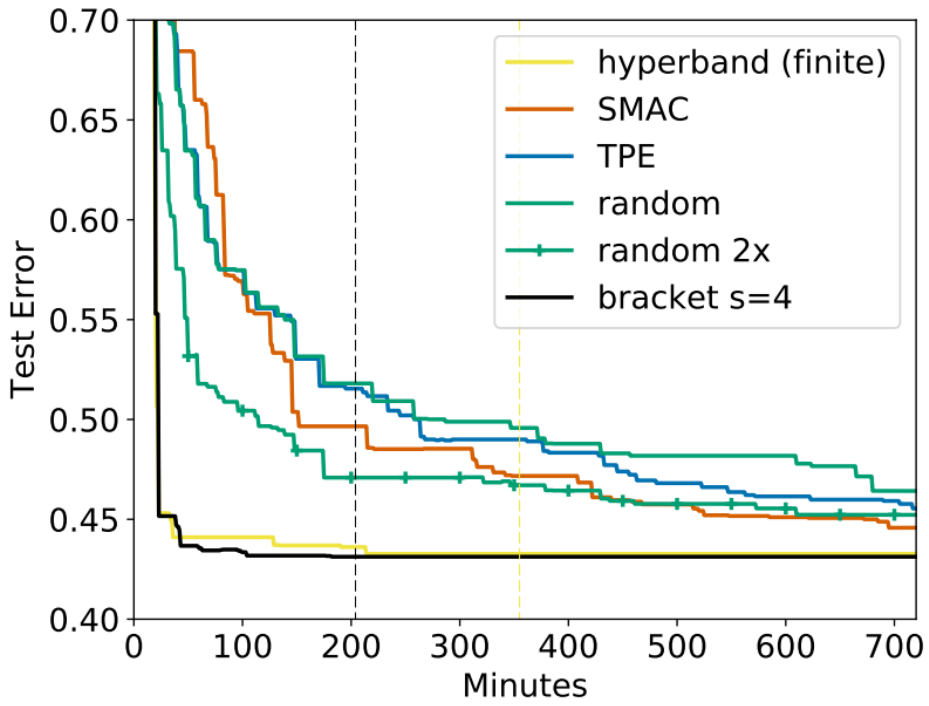


Amadou A A, Singh V, Ghesu F C, et al. Goal-conditioned reinforcement learning for ultrasound navigation guidance[C]//International conference on medical image computing and computer-assisted intervention. Cham: Springer Nature Switzerland, 2024: 319-329.

Hyper-parameter Optimization

- Gradient Descent -> Graduate Student Descent?
- AutoML，可以理解为升级版的Grid Search

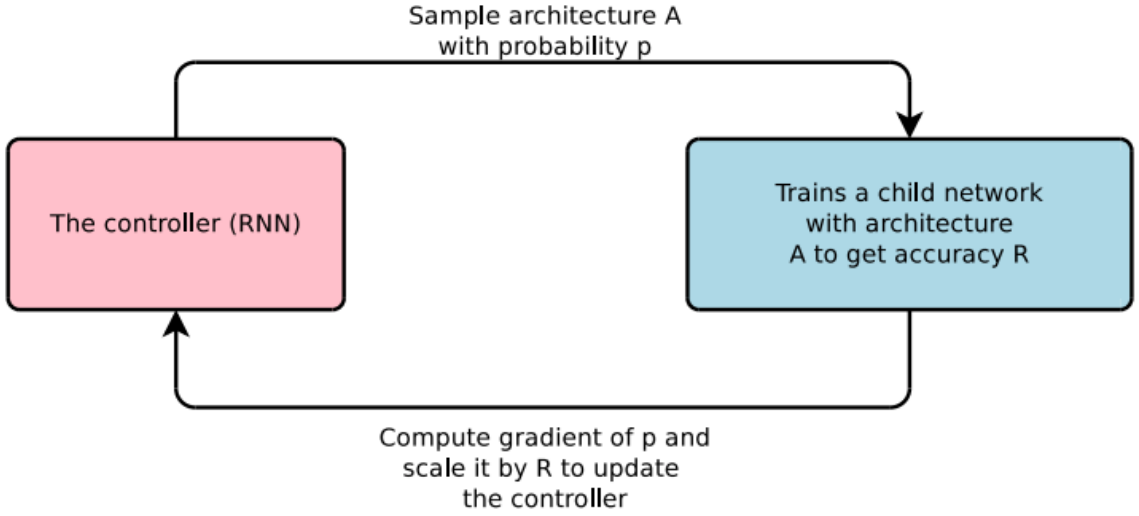


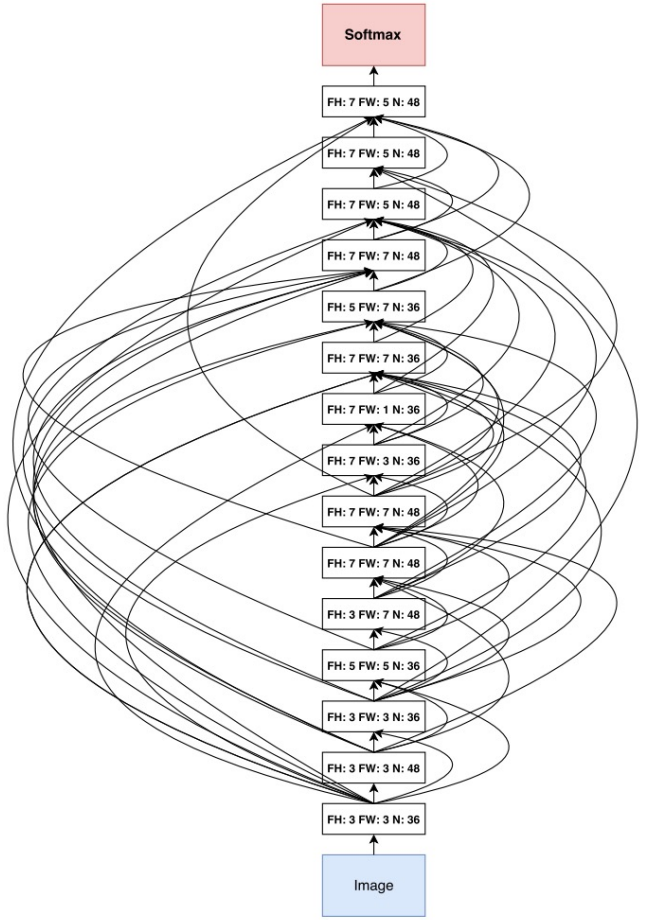


Li L, Jamieson K, DeSalvo G, et al. Hyperband: A novel bandit-based approach to hyperparameter optimization[J]. Journal of Machine Learning Research, 2018, 18(185): 1-52.

Neural Architecture Search

- 自动学习网络结构设计
- 但搜索需要在网络退化和效率之间取平衡





Zoph B, Le Q V. Neural architecture search with reinforcement learning[J]. arXiv preprint arXiv:1611.01578, 2016

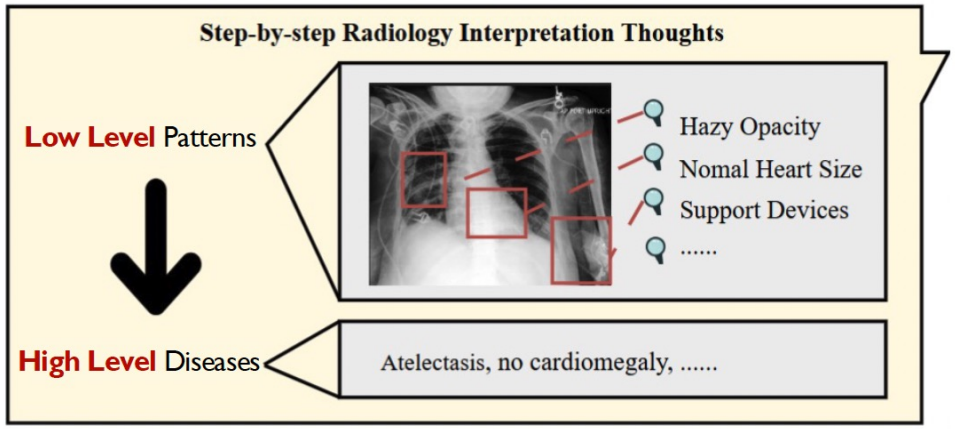
目录

Reasoning in LLMs

现在 黑箱诊断

问题：这张Xray是否存在心肌肥大？
回答：是

未来 有推理过程的诊断



问题：这张Xray是否存在心肌肥大？
回答：是
因为：To determine if the chest X-ray shows cardiomegaly, it's crucial to analyze the size of the heart. The key indicator is the cardiothoracic ratio, which is the ratio of the maximal horizontal cardiac diameter to the maximal horizontal thoracic diameter. In a normal PA chest X-ray, this ratio is expected to be less than 0.50 …

Reasoning in LLMs

- 大语言模型中的推理过程

Chain of Thought (CoT)

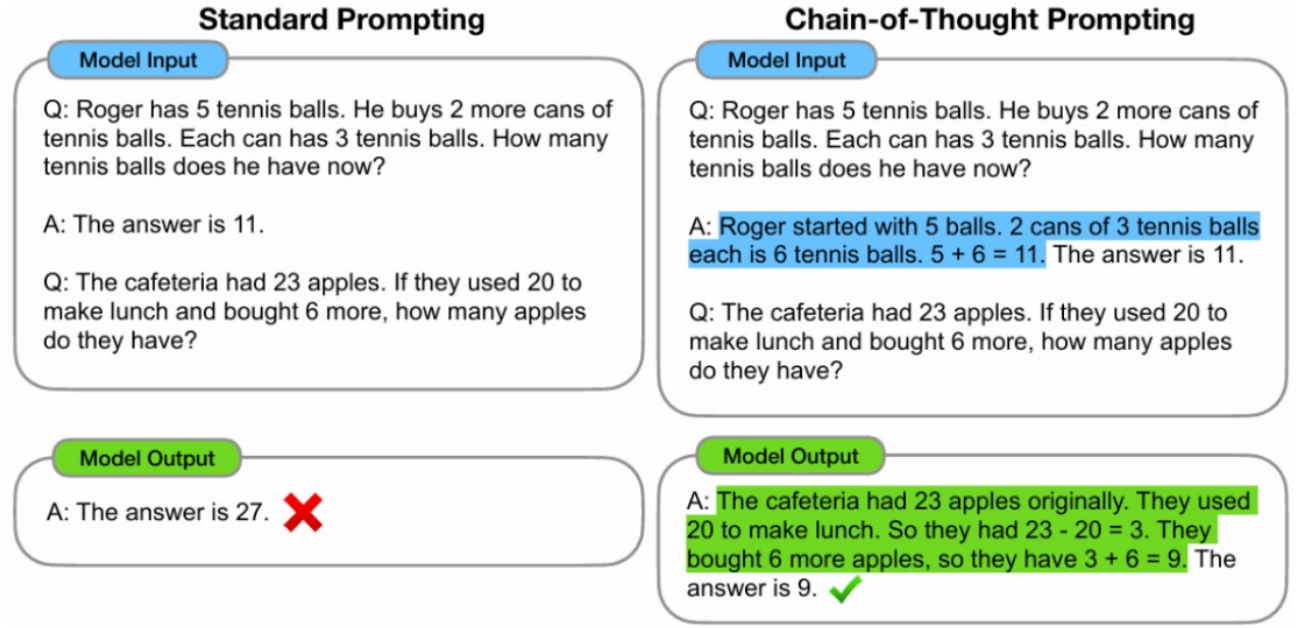
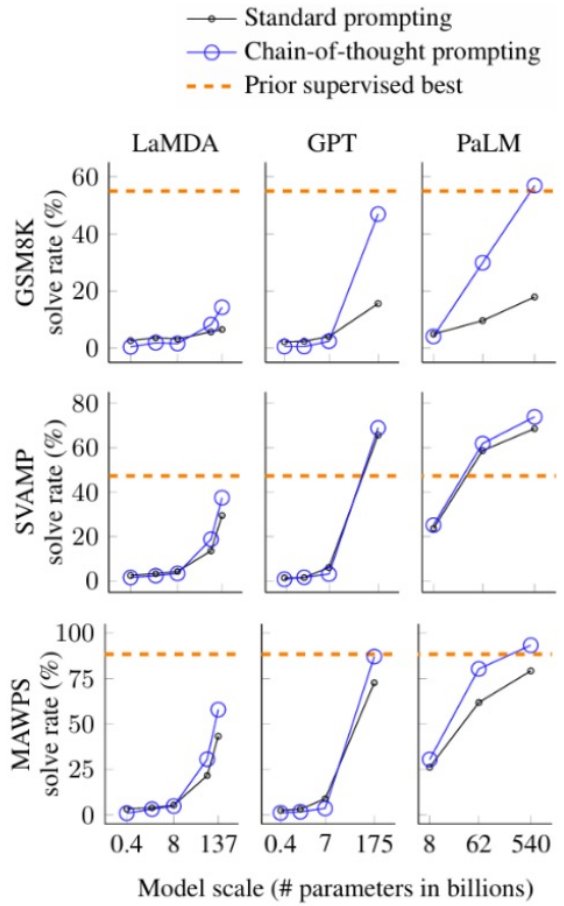


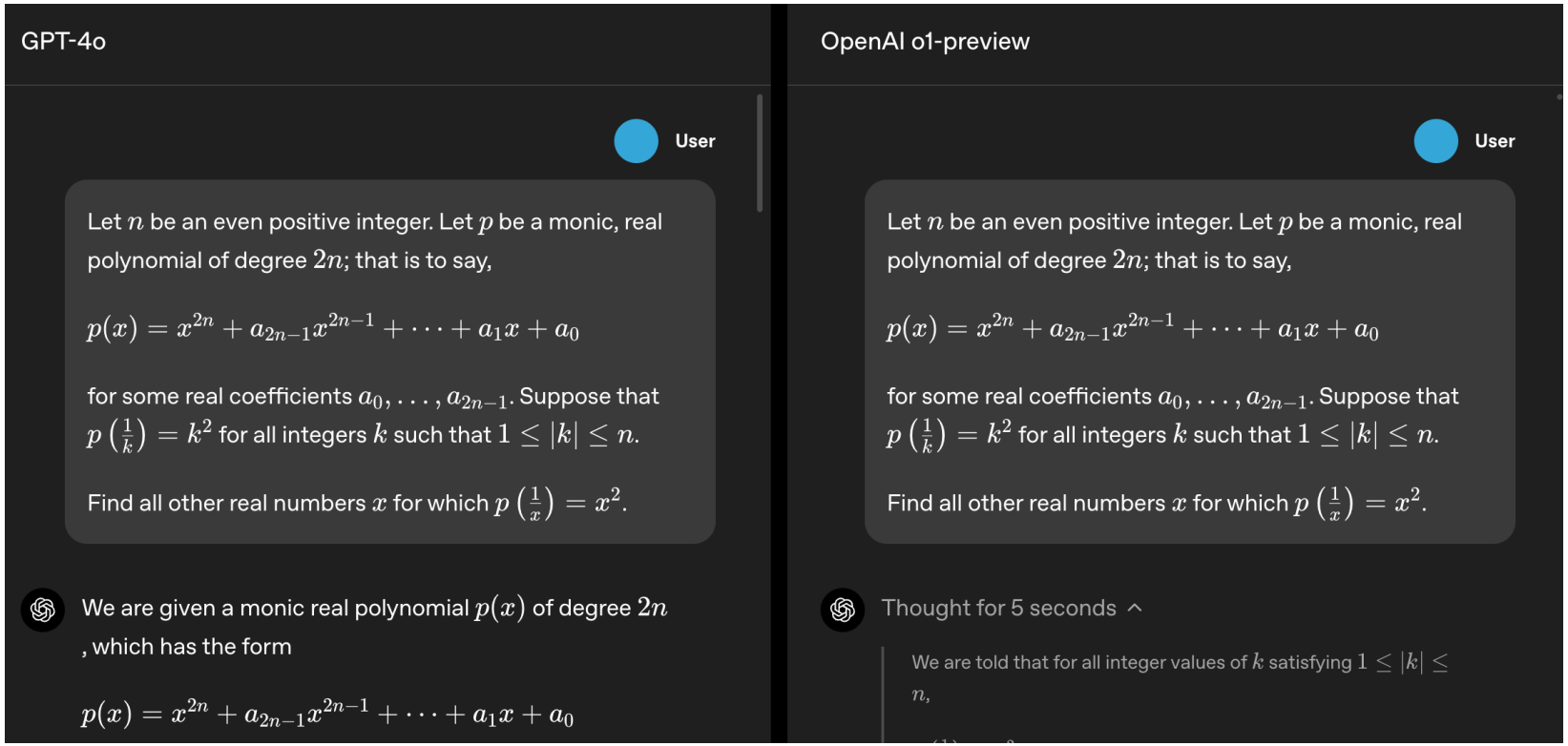
Figure 1: Chain-of-thought prompting enables large language models to tackle complex arithmetic, commonsense, and symbolic reasoning tasks. Chain-of-thought reasoning processes are highlighted.

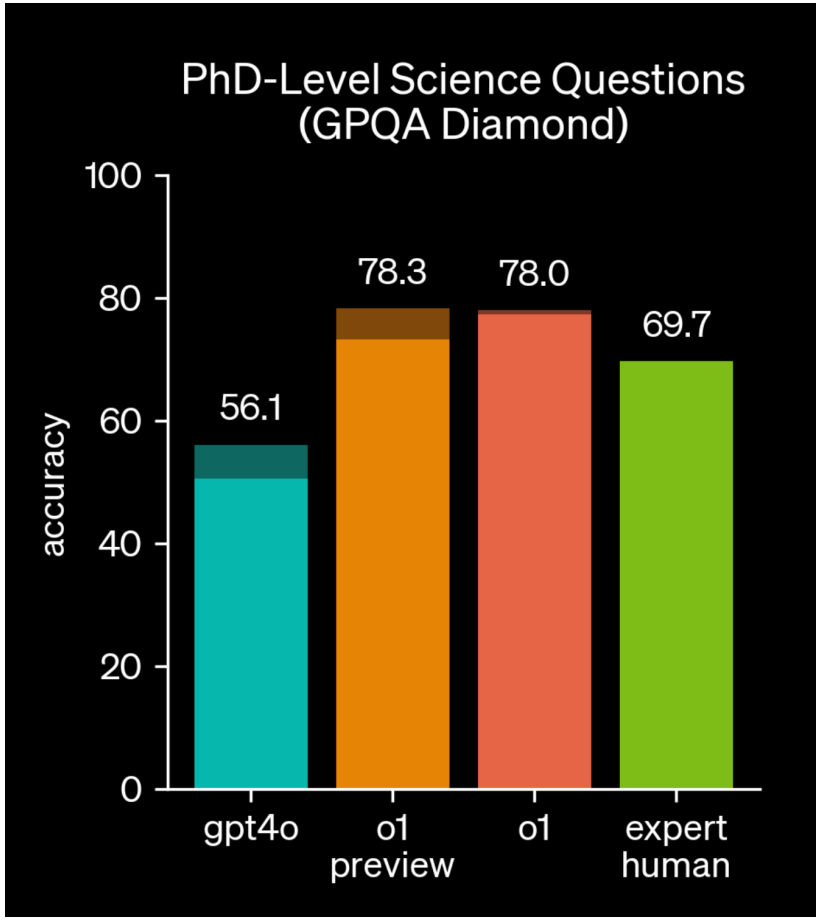


Wei J, Wang X, Schuurmans D, et al. Chain-of-thought prompting elicits reasoning in large language models[J]. Advances in neural information processing systems, 2022, 35: 24824-24837.

Reasoning in LLMs

- ChatGPT o1，新的推理模型

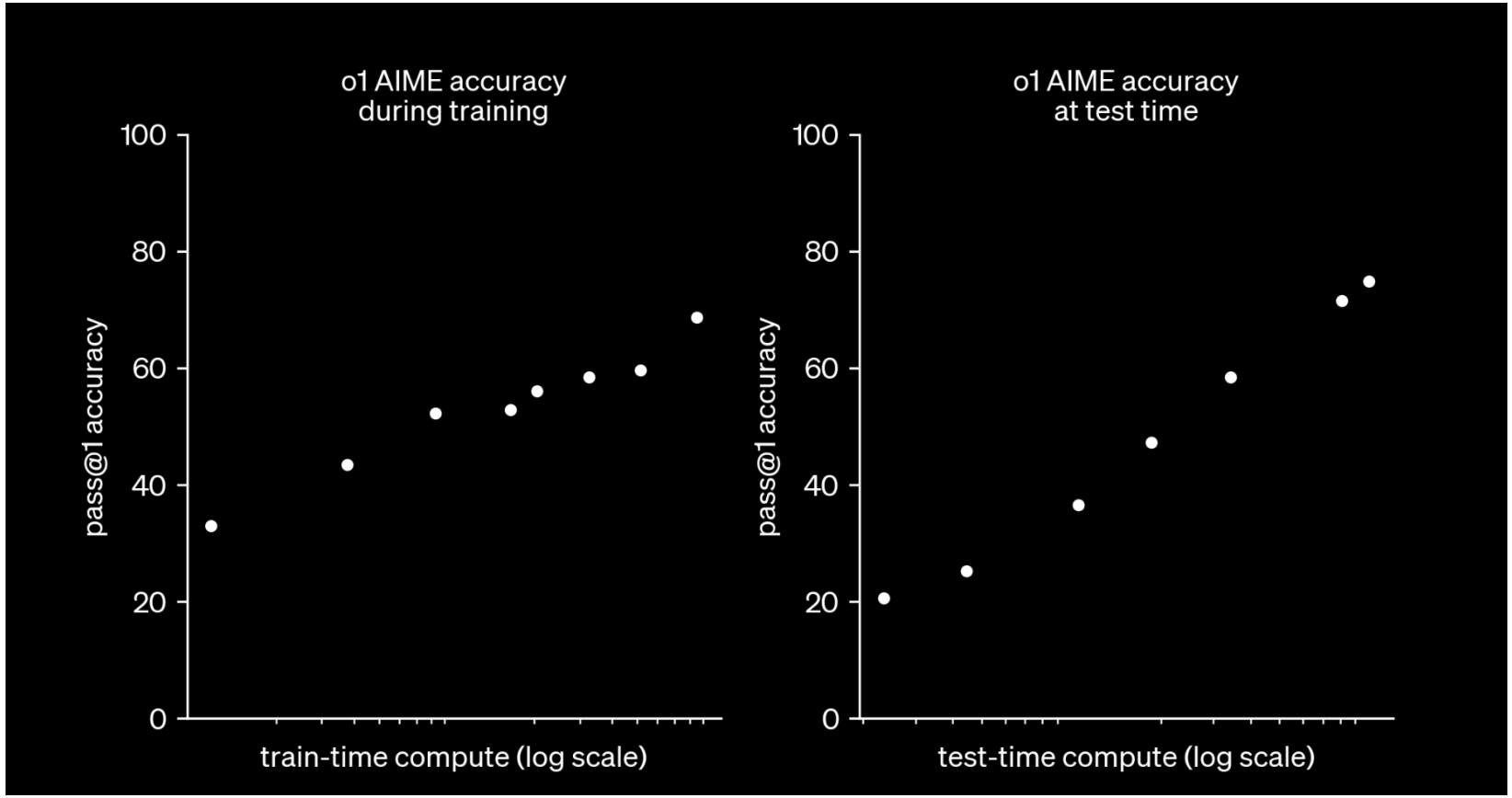




ChatGPT o1

Reasoning in LLMs

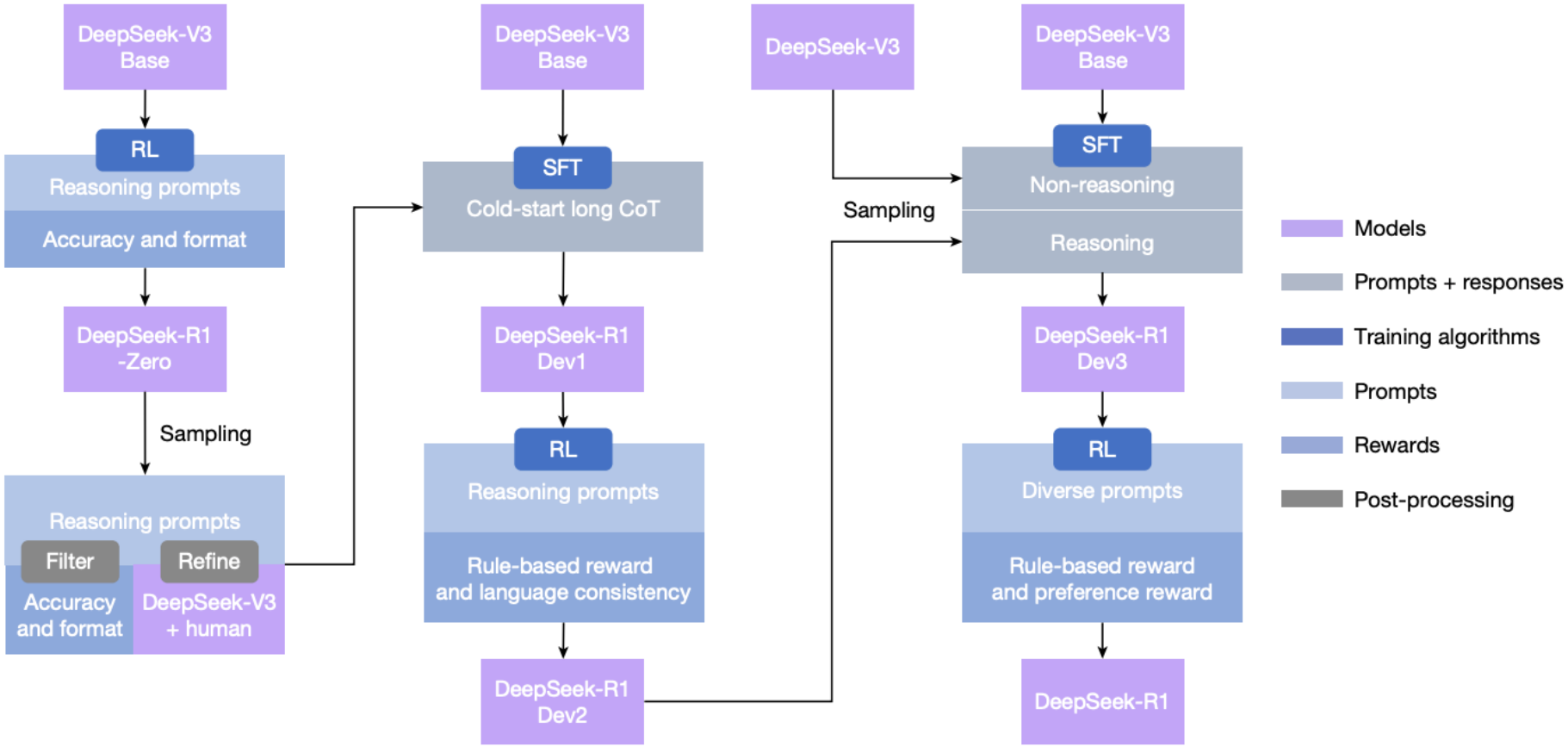
- 训练 （RL） 越长，思考（CoT）越久，效果越好



ChatGPT o1

Reasoning in LLMs

- DeepSeek通过RL来激发推理能力



Guo D, Yang D, Zhang H, et al. Deepseek-r1 incentivizes reasoning in llms through reinforcement learning[J]. Nature, 2025, 645(8081): 633-638.

Reasoning in LLMs

- Guided Region Policy Optimization (GRPO)：SFT过程指导的PPO
- **无需奖励模型**：通过任务自带的可计算反馈（如对错判定、测试用例通过率）直接构建奖励。
- **群体相对优化**：在同一输入下采样多个候选输出，用相对比较（归一化奖励）作为优化信号。

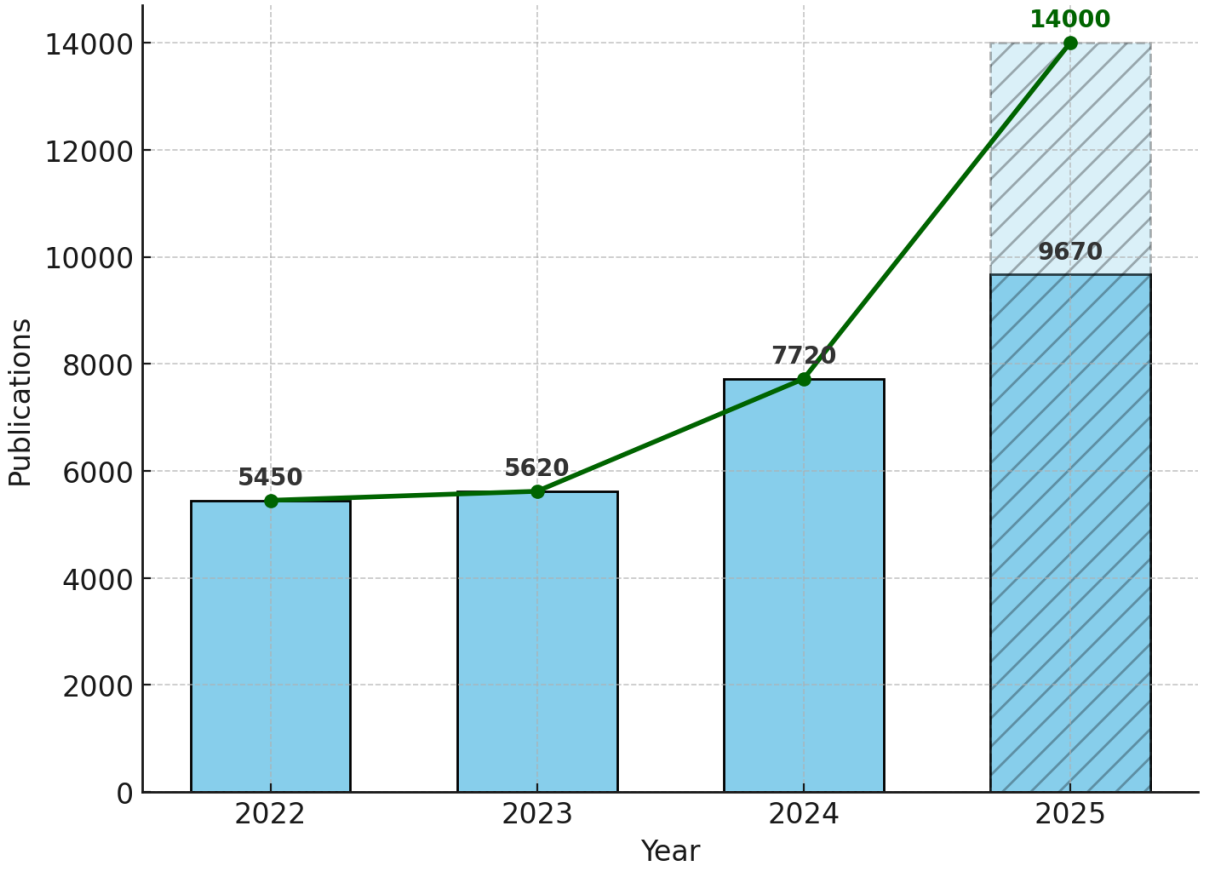
$$\mathcal{J}_{GRPO}(\theta) = \mathbb{E}[q \sim P(Q), \{o_i\}_{i=1}^{G} \sim \pi_{\theta_{old}}(O|q)]$$
$$\frac{1}{G}\sum_{i=1}^{G}\left(\min\left(\frac{\pi_\theta(o_i|q)}{\pi_{\theta_{old}}(o_i|q)}A_i, \text{clip}\left(\frac{\pi_\theta(o_i|q)}{\pi_{\theta_{old}}(o_i|q)}, 1-\varepsilon, 1+\varepsilon\right)A_i\right) - \beta\mathbb{D}_{KL}\left(\pi_\theta||\pi_{ref}\right)\right), \quad (1)$$

PPO 区域奖励

Guo D, Yang D, Zhang H, et al. Deepseek-r1 incentivizes reasoning in llms through reinforcement learning[J]. Nature, 2025, 645(8081): 633-638.

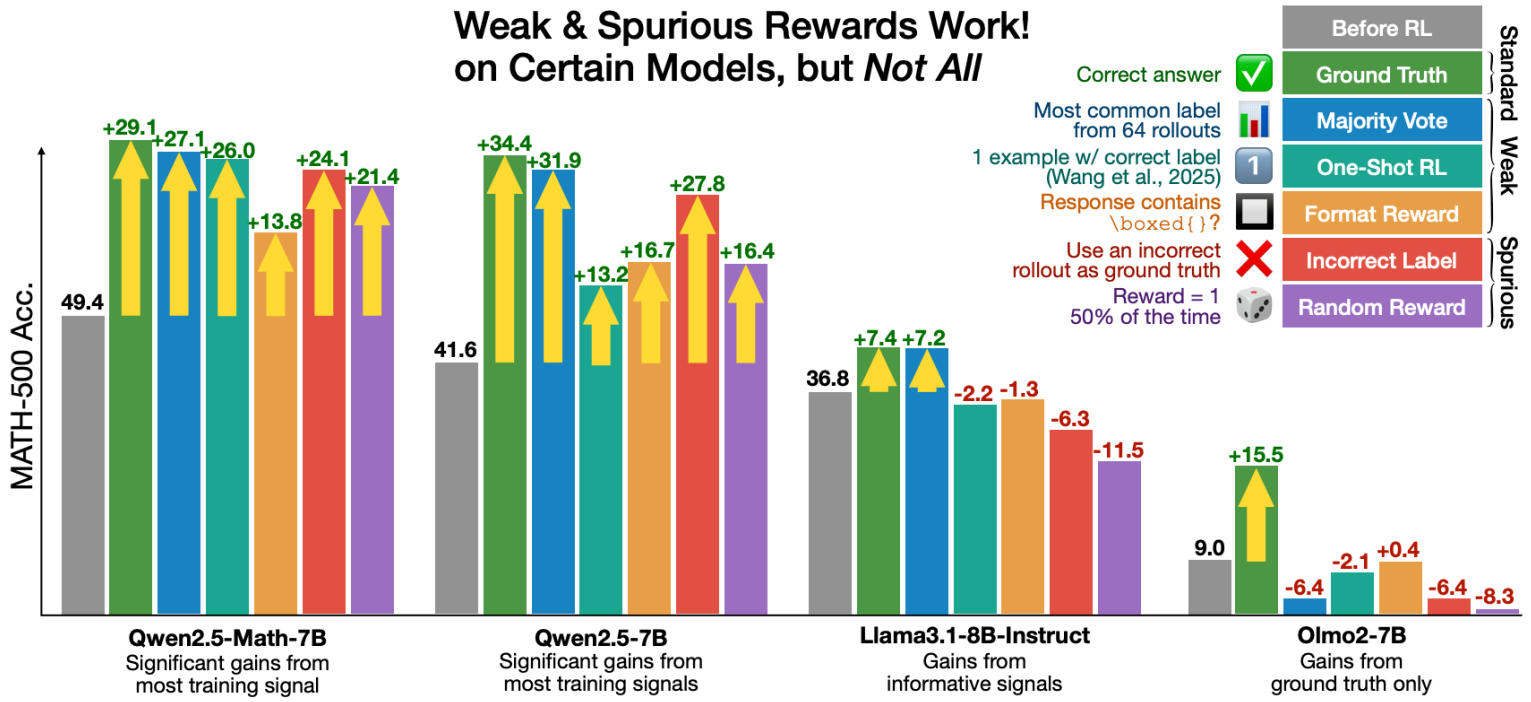
Reasoning in LLMs

- 今年最火主题



Reasoning in LLMs

- 随机的错误奖励也可以带来提升



Shao R, Li S S, Xin R, et al. Spurious rewards: Rethinking training signals in rlvr[J]. arXiv preprint arXiv:2506.10947, 2025.

Reasoning in LLMs

- 无监督的Entropy minimization奖励可以带来提升

		Objectives	Gradients wrt. θ
Finetuning	**EM-FT** §3.1	Min. $\hat{\mathcal{H}}_{\text{tok}}$	$\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{\lvert y^i\rvert}\nabla_\theta \mathcal{H}\left(\pi_\theta(\cdot \mid \boldsymbol{y}^i_{<t})\right)$
RL	**EM-RL-sequence** §3.2	Max. $-\hat{\mathcal{H}}_{\text{traj}}$	$\frac{1}{N}\sum_{i=1}^{N}\left[\sum_{t=1}^{\lvert y^i\rvert}\log \pi_\theta(y^i_t \mid \boldsymbol{y}^i_{<t})\nabla_\theta \log \pi_\theta(\boldsymbol{y}^i)\right)\right]$
RL	**EM-RL-token** §3.2	Max. $-\hat{\mathcal{H}}_{\text{tok}}$	$\frac{1}{N}\sum_{i=1}^{N}\left[-\sum_{t=1}^{\lvert y^i\rvert}\mathcal{H}(\pi_\theta(\cdot \mid \boldsymbol{y}^i_{<t})\nabla_\theta \log \pi_\theta(\boldsymbol{y}^i))\right]$
Inf-Scaling	**EM-INF** §4	Min. $\hat{\mathcal{H}}_{\text{tok}}$	N/A

Agarwal S, Zhang Z, Yuan L, et al. The unreasonable effectiveness of entropy minimization in llm reasoning[J]. arXiv preprint arXiv:2505.15134, 2025.

Reasoning in LLMs

- 无监督的Entropy minimization奖励可以带来提升

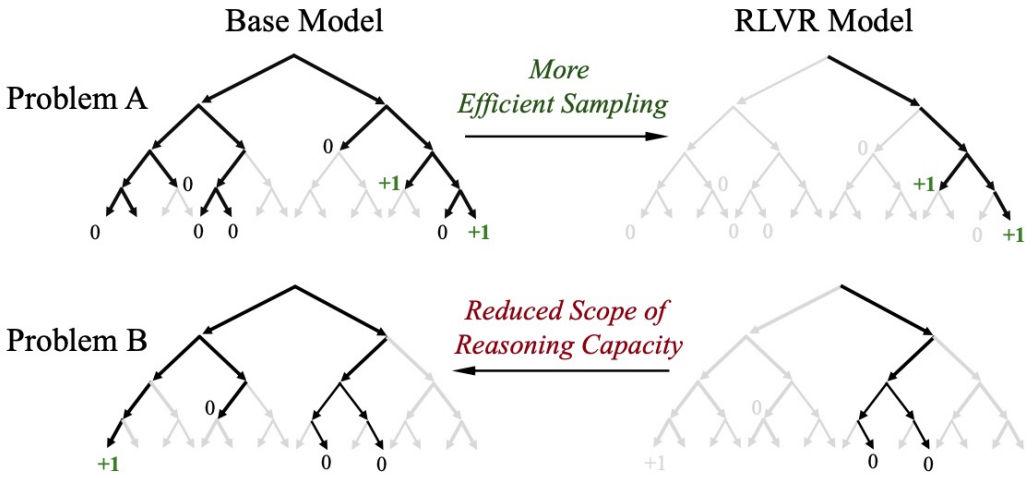
Table 2: Performance comparison of unsupervised finetuning (EM-FT) and various rewarding methods in EM-RL with supervised finetuning and RL. *Italics*, **Bold** indicates performance improvement over GRPO and SC-RL (self-consistency RL), respectively. Dash line ("–") denotes that self-consistency is inapplicable. FLOPs are reported as 10^{17} (§D.4).

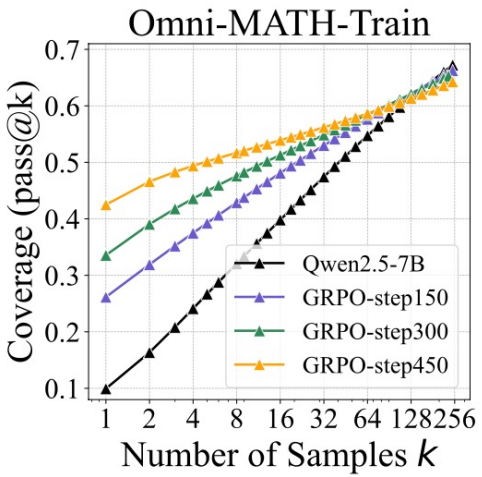
	Math						Coding			
	Math	**AMC**	**AIME**	**Minerva**	**Olymp.**	**Avg.**	**LeetC**	**LiveC**	**Avg.**	**FLOPs**
Qwen2.5-7b	43.8	31.3	15.6	14.7	19.0	24.9	26.1	18.4	22.3	–
Trained using 60K labeled prompts										
w/ SFT N=1	48.2	30.2	10.0	17.6	22.4	25.7	18.3	18.3	18.3	1.0
w/ RLOO N=4	73.0	57.8	23.3	31.2	34.2	43.9	28.3	26.7	27.5	13.1
w/ GRPO N=4	71.8	56.6	21.1	25.0	35.9	42.1	25.0	25.8	25.4	13.1
Our unsupervised methods trained on 60K unlabeled prompts										
EM-FT N=1	67.2	51.8	14.4	*33.1*	*34.4*	40.2	*28.3*	17.2	22.8	1.0
SC-RL N=4	73.2	51.8	15.6	*26.1*	*36.7*	40.7	–	–	–	13.1
EM-RL-SEQUENCE N=4	67.2	**53.0**	**21.1**	***30.9***	35.6	**41.6**	*31.1*	21.7	*26.4*	13.1
EM-RL-TOKEN N=4	70.8	***57.8***	**18.9**	***30.9***	35.9	**42.9**	*29.5*	24.5	*27.0*	13.1

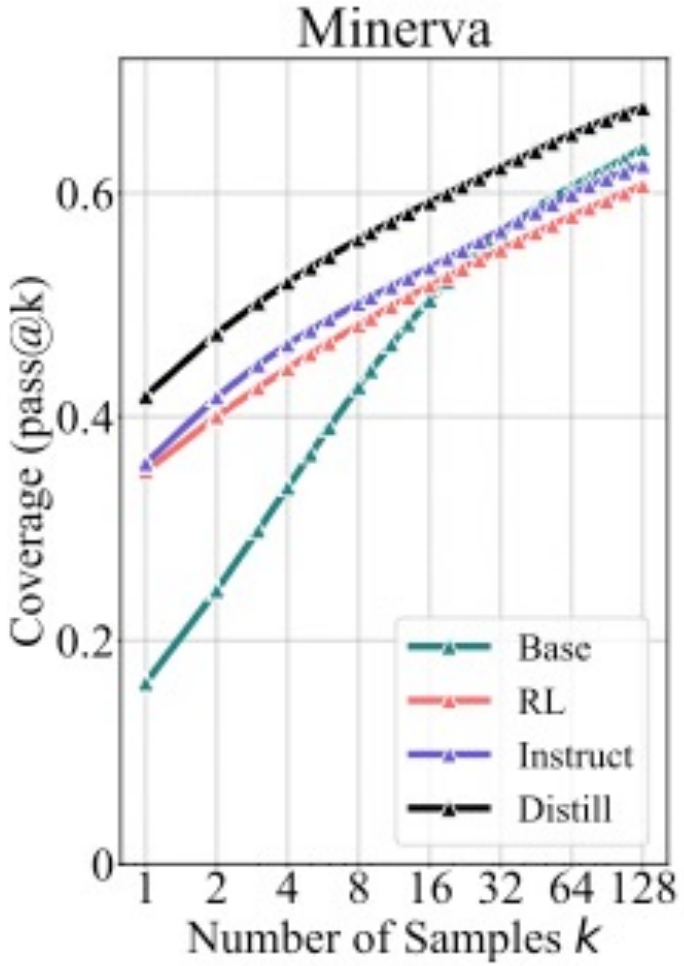
Agarwal S, Zhang Z, Yuan L, et al. The unreasonable effectiveness of entropy minimization in llm reasoning[J]. arXiv preprint arXiv:2505.15134, 2025.

Reasoning in LLMs

- Party is (almost) over
- RL不能产生新的推理路径



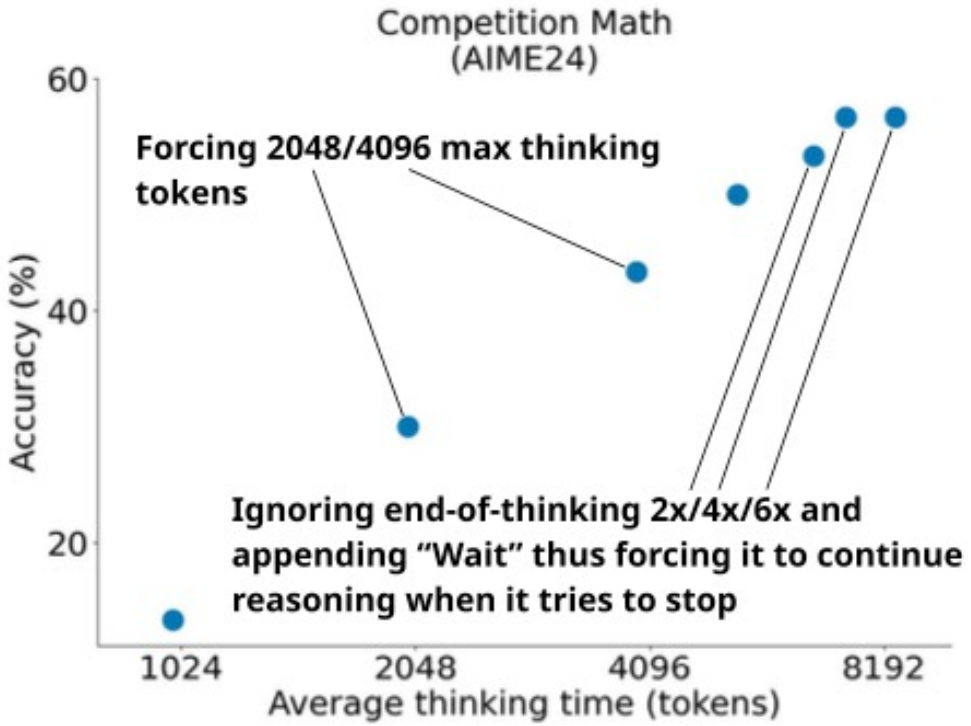




Yue Y, Chen Z, Lu R, et al. Does reinforcement learning really incentivize reasoning capacity in llms beyond the base model?[J]. arXiv preprint arXiv:2504.13837, 2025.

Reasoning in LLMs

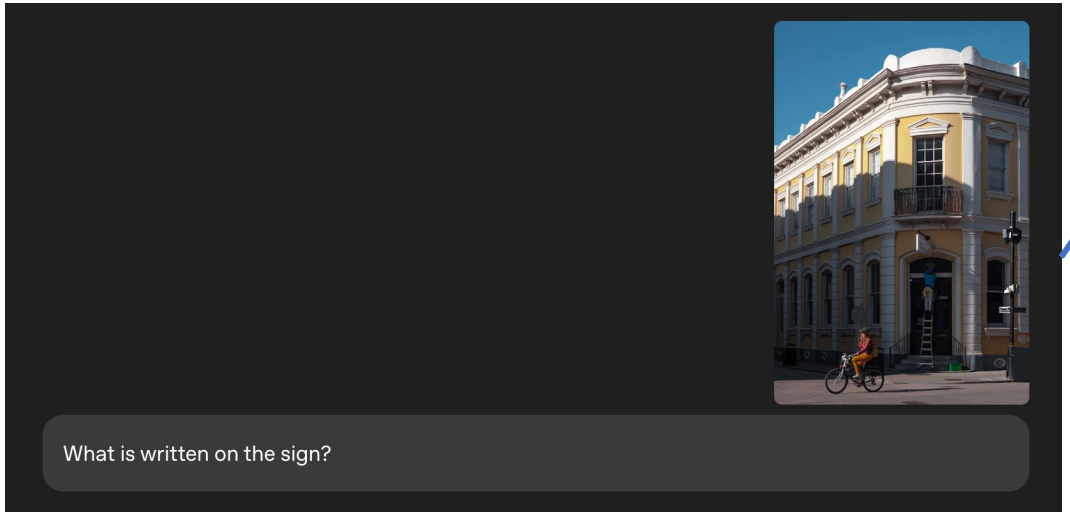
- 仅靠高质量CoT（1k）+SFT（26min 8xH100），也足够有效了
- 只要思考更久就行

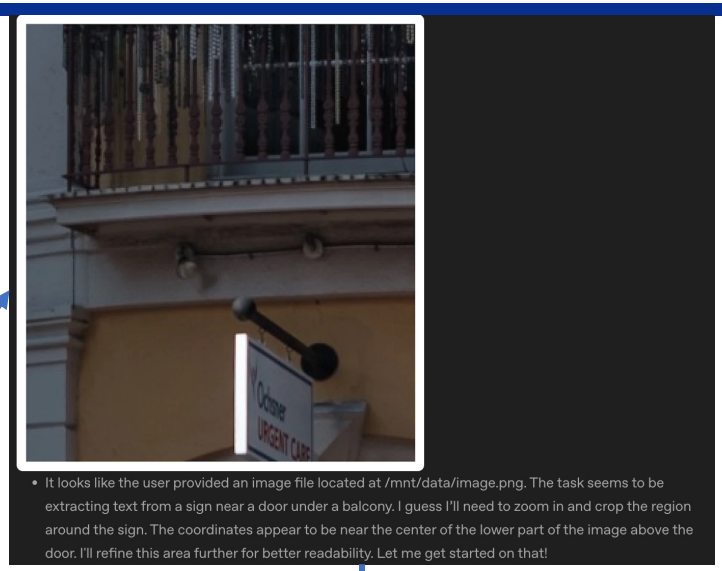


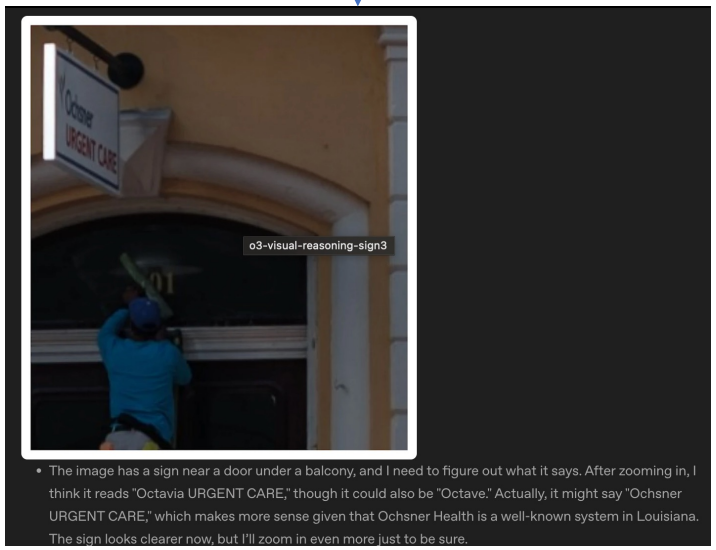
Muennighoff N, Yang Z, Shi W, et al. s1: Simple test-time scaling[J]. arXiv preprint arXiv:2501.19393, 2025.

Reasoning in LLMs

- 前方战场：RL in MLLMs
- Think with Images
- Integrating images directly into CoT



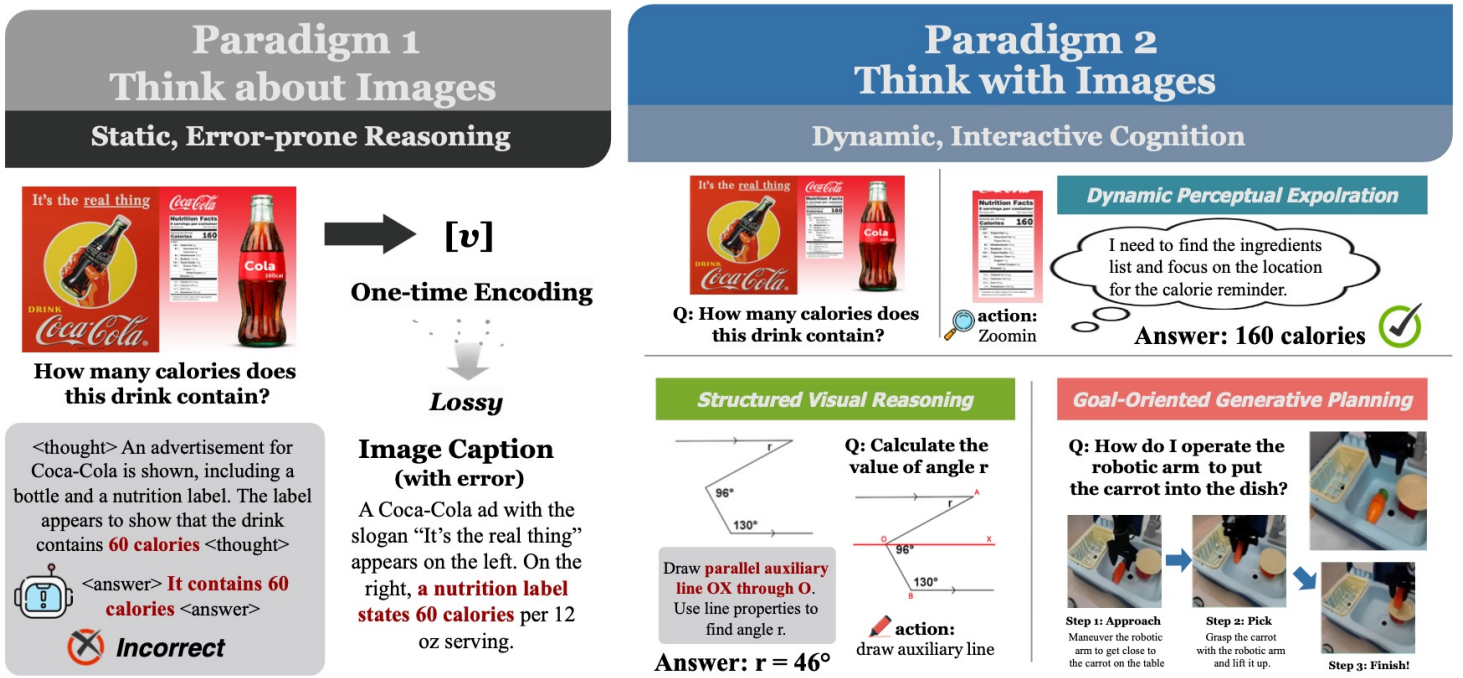




OpenAI o3

Reasoning in LLMs

- 模型调用外部工具（OCR、检测器、缩放等），主动探索图像。
- 个人认为：医学图像可解释诊断的终极形态



Su Z, Xia P, Guo H, et al. Thinking with images for multimodal reasoning: Foundations, methods, and future frontiers[J]. arXiv preprint arXiv:2506.23918, 2025.

小结

- 深度强化学习有一些在医学图像中在上一个时代就有一些应用
- 一是基于Q-learning的目标高效定位
- 二是基于PPO等的高参数搜索
- 最近深度强化学习再次流行，目标是加强LLM的Reasoning，但最近研究表明RL并不能带来新的知识，大部分都是预训练过程中获取
- RL in MLLMs非常适合Medical Imaging，需要工程化强的解决方案