



面向视觉问答的推理方法研究及应用

答辩人： 刘进

指导老师：周风余 教授

学无止境
气有浩然



山东大学
SHANDONG UNIVERSITY

目录

CONTENTS

- 1. 研究背景**
- 2. 研究内容**
- 3. 技术方案**
- 4. 课题创新点**
- 5. 计划安排**



课题研究背景



➤ 课题来源

□ 国家重点研发计划项目

项目名称：服务机器人云服务平台

项目编号：2017YFB1302400

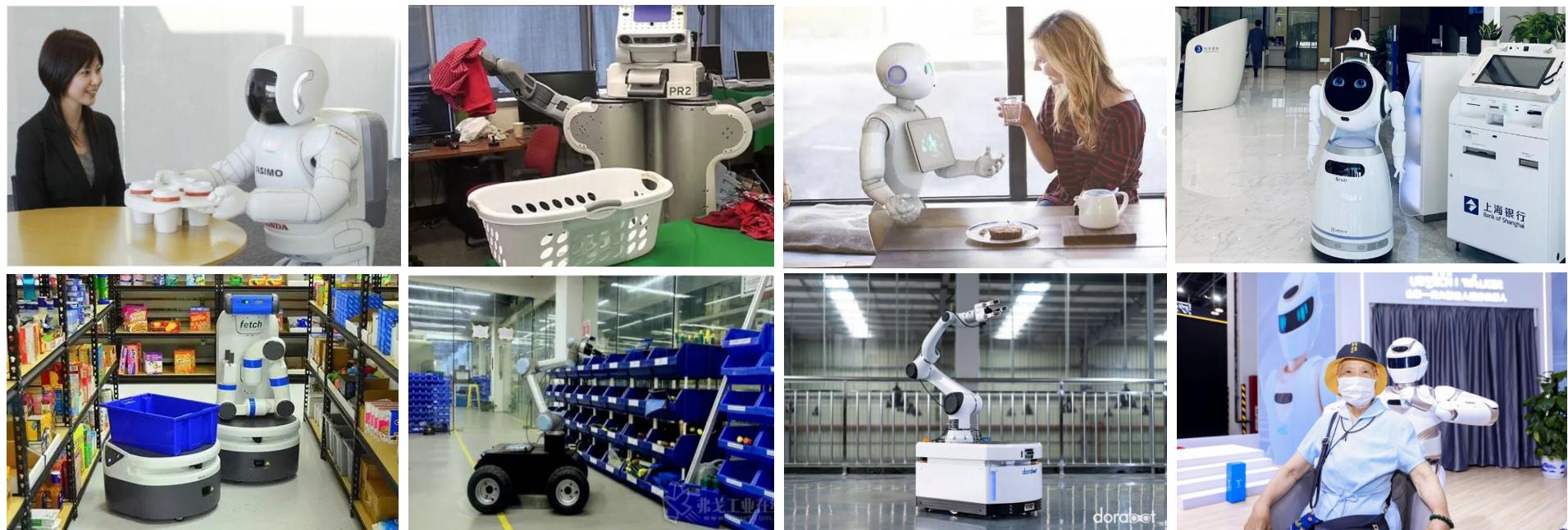
□ 济南市“新高校20条”资助项目

项目名称：机器人超融合云服务平台实用化关键技术研究

项目编号：2021GXRC079

➤ 研究背景

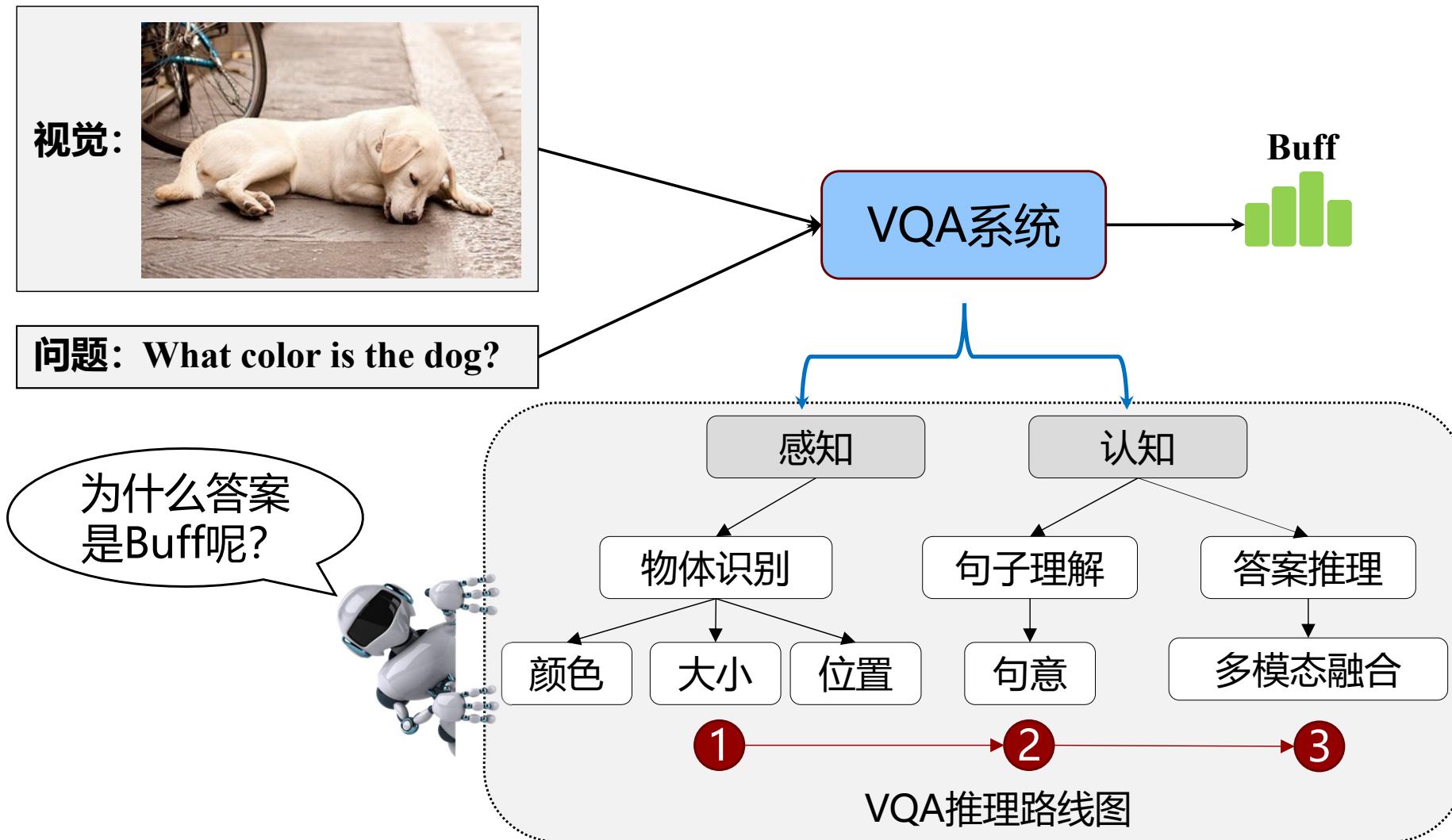
- 融合视觉、语言及音频等的**多模态数据**进行跨模态**推理**逐渐成为交互软件的研发主流，可以充分减少单一模态带来的信息不足；
- 在多模态数据交互中，基于**视觉信息和文本信息的视觉问答**成为目前的一个热点话题；
- 基于**视觉问答**的智能交互已经开始应用于各种类型服务机器人，**服务性能取决于其推理水平**。



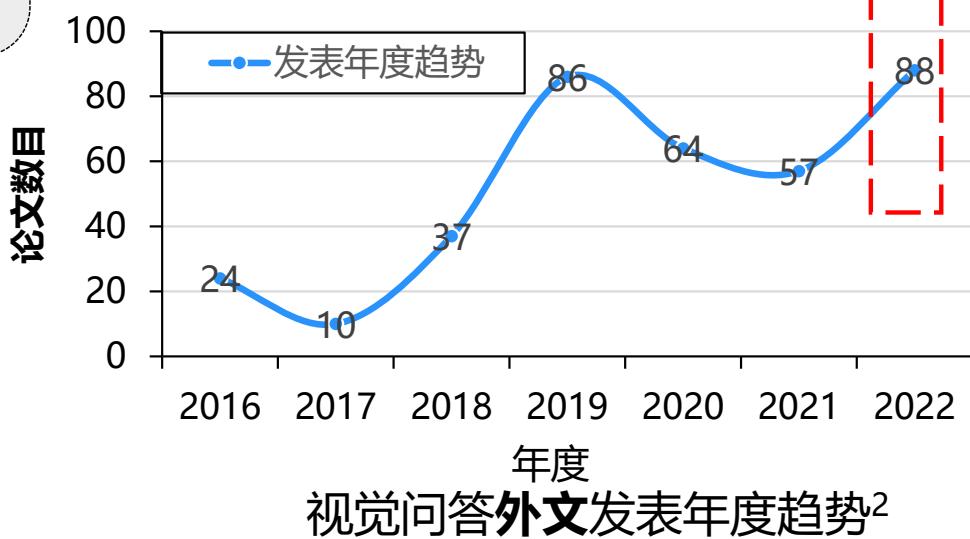
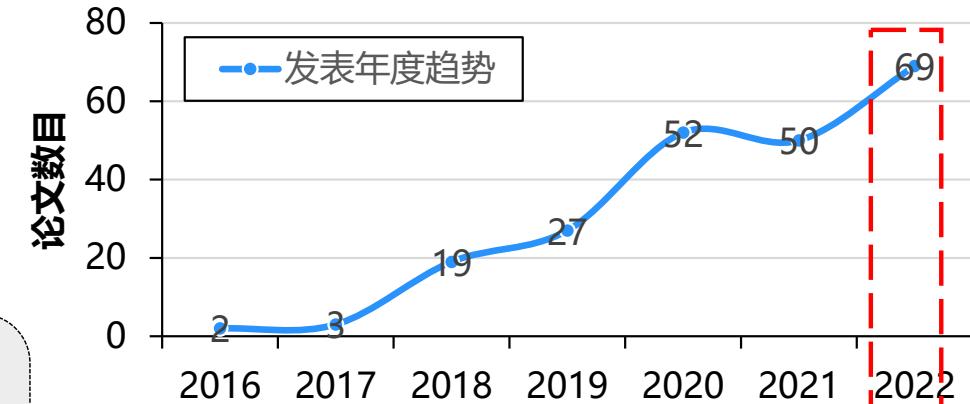
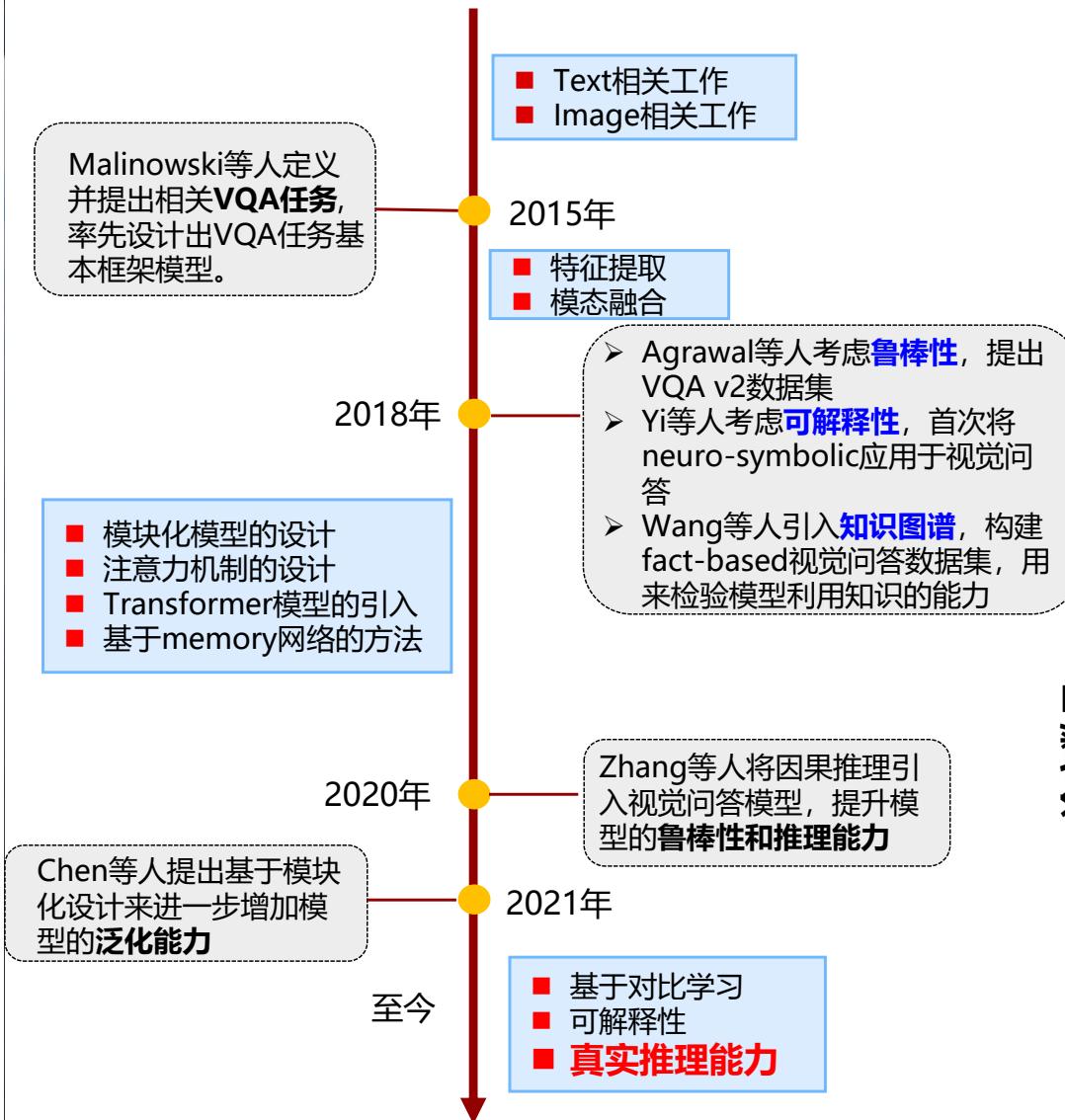


➤ 研究任务-视觉问答

视觉问答(Visual question answering, VQA): 旨在让机器人像人一样根据获取的**视觉信息**和**问题信息(感知)**来对图片的内容、问题的含义和意图以及相关的常识有一定的理解，并通过**推理(认知)**提供合理的答案。



国内外研究现状



机遇与挑战：相关研究增速迅猛，但缺乏真实的推理能力，过度依赖于融合的信息特征



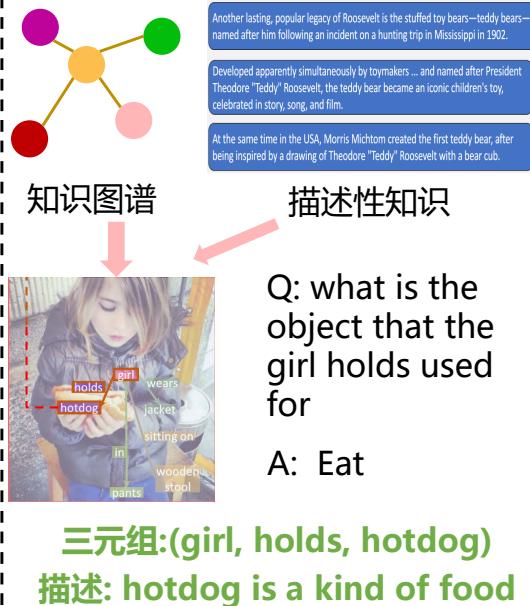
➤ 存在问题难点

- 难点一：如何设计有效的提升策略融入问答模型，提升模型的鲁棒性？ → (鲁棒性)
对比采样是通过获取一定的问题或图像负样本，需要具备一定的**有效性**，同时将视觉特征增强和文本特征增强策略作为提升模型性能的关键途径，**但不能影响视觉问答模型的执行过程。**
- 难点二：如何构建外部知识数据库，可供问答系统进行知识扩展？ → (知识性)
由于不同领域场景分布不均，导致各情景下问答模型的表现不一，需要考虑如何**引入外部知识库**，并根据检索的话题进行**知识匹配、融合**，以供上层话题规划使用。
- 难点三：如何设计有效的融合推理机制，以保证推理的可解释性？ → (可解释性)
在视觉问答系统中，需要考虑推理过程的**可解释性**，充分发挥视觉感知模块和语义理解模块的性能，实现推理步骤的可追踪性。
- 难点四：如何构建广泛的程序块，可供推理过程进行泛化扩展？ → (组合性)
由于难点三中研究的推理步骤需要人为定义，形式单一，且执行函数较为复杂，需要考虑从**概念和元概念的角度抽象**的程序执行的策略，简化执行函数，同时提升推理的泛化组合能力。

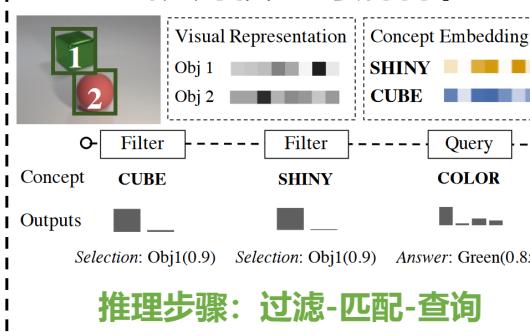
课题研究内容

➤ 主要研究框架

加强模型利用知识的能力



提升模型可解释性



(一) 基于对比采样和特征融合的视觉问答方法研究



(二) 基于知识增强的视觉问答方法研究



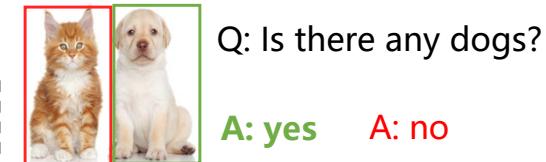
(三) 基于neuro-symbolic的视觉问答方法研究



(四) 基于meta-module的视觉问答方法研究

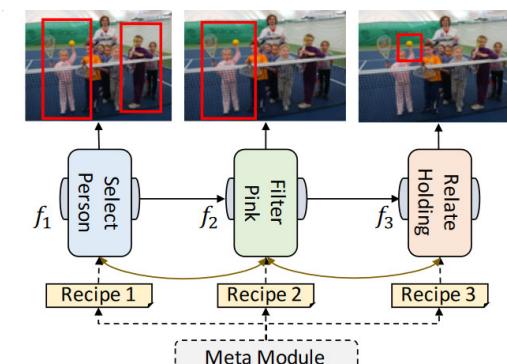


解决模型鲁棒性



正确答案&正确grounding
错误的答案&错误的grounding

实现模型泛化组合能力



在机器人应用落地



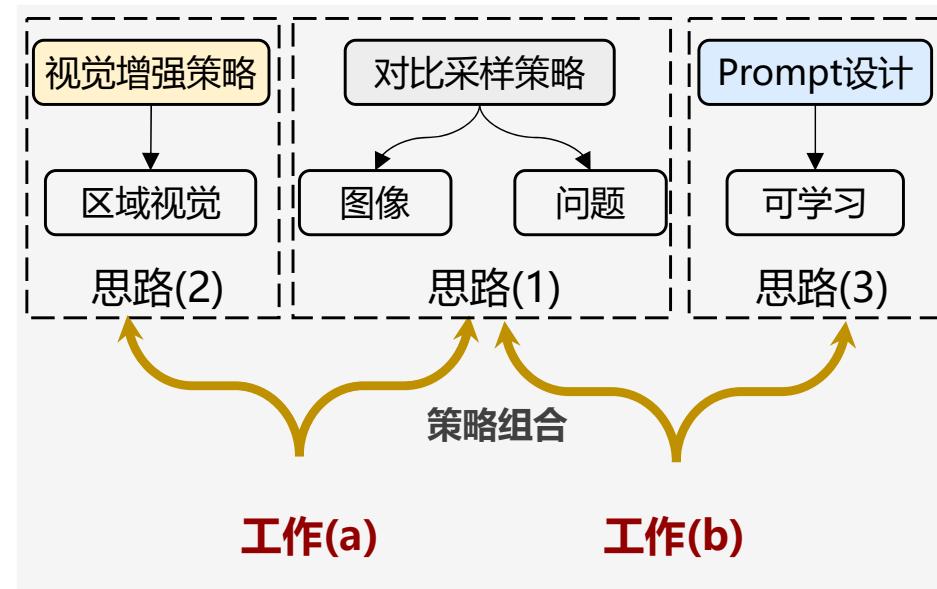
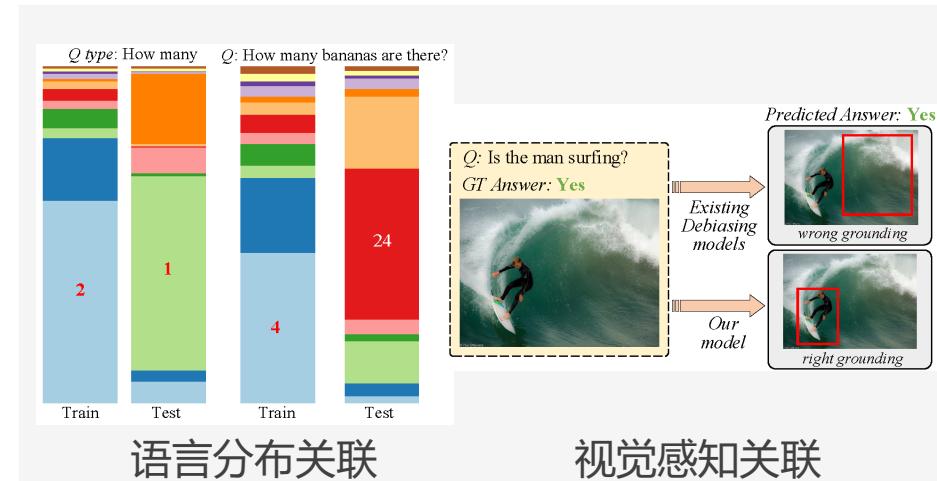
课题技术方案

➤ 研究内容一：基于对比采样和特征融合的视觉问答方法

研究目的：为了保证模型的鲁棒性，研究在不同偏差验证数据集下，对模型预测的答案进行消偏，即**消除语言分布关联**和**视觉感知关联**等偏差保证模型可以**预测正确的答案**，同时能给出**正确的视觉理解区域**

研究思路：

- 1) 采用**视觉增强策略**，基于POS tag的关键单词选取，并将其融合到视觉信息中得到相应的显著视觉区域，从而保证模型获取更加有利于推理的视觉信息；
- 2) 基于**对比采样策略**，构建图像和问题的正负样本，实现多种模态信息的均衡使用，**缓解由于样本分布不均匀带来的统计偏差问题**。
- 3) 此外，进一步研究设计一种不定长度且可融合到训练过程的**Prompt的设计策略**，挖掘句子的深层语义知识，并增强对问题语义的理解。



➤ 工作(a)——特征增强下的二次消偏

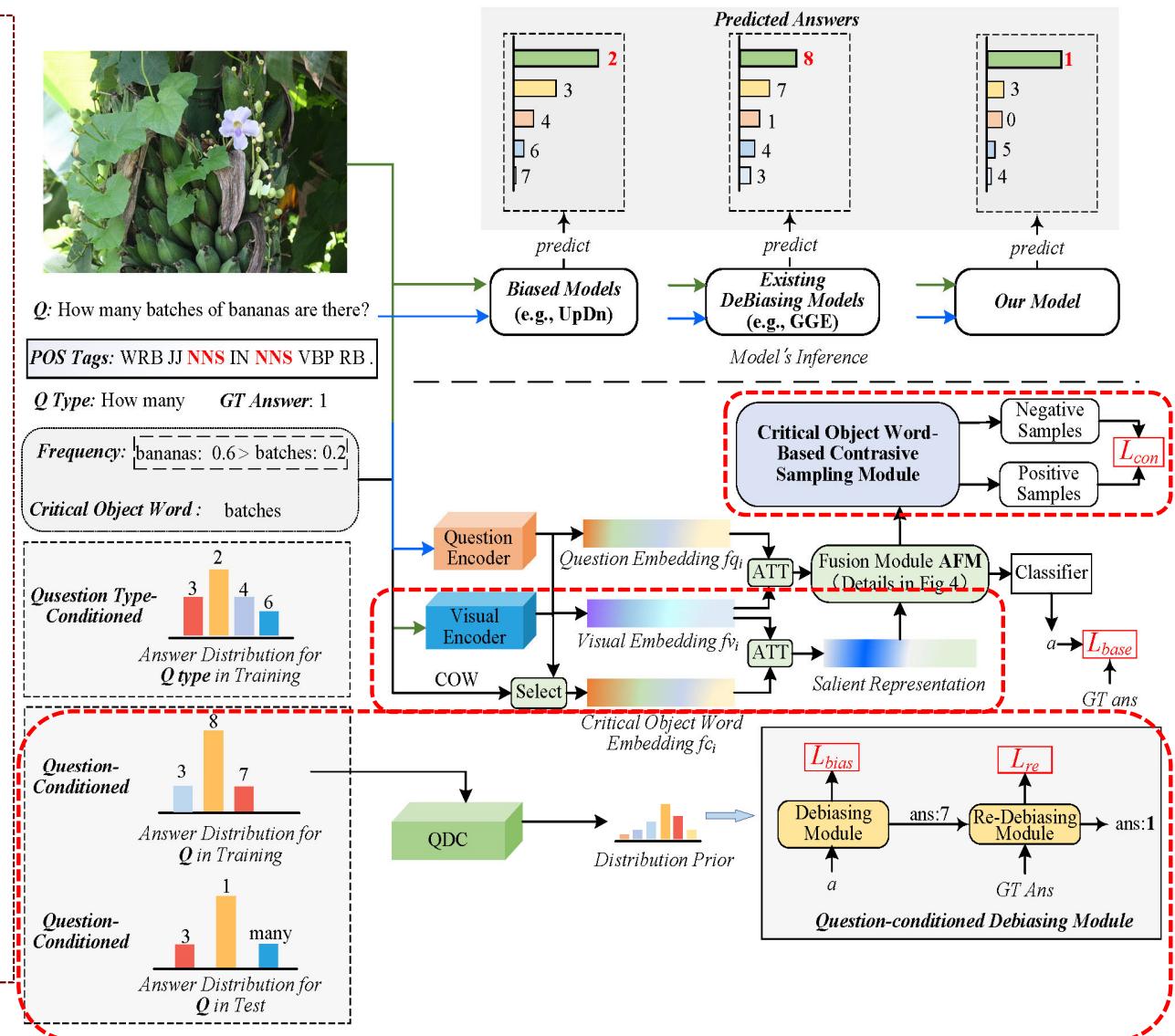
实施策略：

Step1：采用**POS标注**获取问题关键词，并进一步利用其获取相应的**视觉显著信息**；

Step2：结合**对比学习采样**，获取的**视觉显著信息**和原始图片信息的正负样本，**提升模型对视觉信息的利用**；

Step3：从问题-答案关联的虚假偏差统计角度出发，提出了**二次消偏**的机制，在训练过程中构建真实答案和预测答案的伪标签，并用以模型的纠偏，**防止模型陷入预测其他频率较高的虚假关联中**；

Step4：最后基于**VQA v2、VQA-CP v2、VQA-CP v1等开源基准数据集**来验证所提算法的可行性和鲁棒性。



➤ 工作(a)实验结果

Categories	Model	Venue	VQA-CP v2 test (%)				VQA v2 val (%)			
			All	Yes/No	Num	Other	All	Yes/No	Num	Other
基于融合策略	Question-only [8]	CVPR'18	15.95	35.09	11.63	7.11	43.01	67.95	30.97	27.20
	SAN [33]	CVPR'16	24.96	38.35	11.14	21.74	52.41	70.06	39.28	47.84
	UpDn [†] [7]	CVPR'18	39.74	42.27	11.93	46.05	63.48	81.18	42.14	55.66
	HAN [51]	ECCV'18	28.65	52.25	13.79	20.33	-	-	-	-
	S-MRL [13]	CVPR'18	38.46	42.85	12.81	43.20	63.10	-	-	-
	GVQA [8]	CVPR'18	31.30	57.99	13.68	22.14	48.24	72.03	31.17	34.65
	LXMERT [52]	EMNLP'19	46.23	42.84	18.19	55.51	-	-	-	-
基于数据集调整策略	SBS [11]	TMM'21	42.21	53.51	11.12	44.82	63.84	81.81	43.03	55.67
	Unshuffling [53]	ICCV'21	42.39	47.72	14.43	47.27	61.08	78.32	42.16	52.81
	LRS [10]	TIP'22	49.45	72.36	10.93	48.02	62.20	78.80	41.60	54.40
基于ensemble策略	AdvReg. [15]	NIPS'18	41.17	65.49	15.48	35.48	62.75	79.84	42.35	55.16
	LM [54]	EMNLP'19	48.78	70.37	14.24	46.42	63.26	81.16	42.22	55.22
	LMH [54]	EMNLP'19	52.73	72.95	31.90	47.79	56.35	65.06	37.63	54.69
	RUBi [13]	NIPS'19	45.42	63.03	11.91	44.33	58.19	63.04	41.00	54.43
	GRL. [55]	NAACL'19	42.33	59.74	14.78	40.76	-	-	-	-
	DLP [16]	AAAI'20	48.87	70.99	18.72	45.57	57.96	76.82	39.33	48.54
	LPF [17]	SIGIR'21	55.34	88.61	23.78	46.57	55.01	64.87	37.45	52.08
	GGE [12]	ICCV'21	57.32	87.04	27.75	49.59	59.11	73.27	39.99	54.39
基于标注特征	HINT [20]	ICCV'19	46.73	67.27	10.61	45.88	63.38	81.18	42.99	55.56
	SCR [19]	NIPS'19	49.45	72.36	10.93	48.02	-	-	-	-
	AttAlign [20]	ICCV'19	39.37	43.02	11.89	45.00	63.24	80.99	42.55	55.22
基于对比学习或者因果推理	CSS [†] [22]	CVPR'20	41.16	43.96	12.78	47.48	-	-	-	-
	SSL-VQA [25]	IJCAI'20	57.59	86.53	29.87	50.03	63.73	-	-	-
	CF-VQA [24]	CVPR'21	49.74	74.81	18.46	45.19	63.75	82.15	44.29	54.86
	ECD [23]	WACV'22	41.78	42.74	14.89	48.66	-	-	-	-
	AVM [44]	NCA'22	40.08	42.51	12.36	46.41	63.87	81.43	43.82	55.82
DQF (Ours)			60.47	89.76	42.49	50.10	65.23	82.60	44.32	57.56

结论：我们提出的模型在两份数据集上的所有评价指标中均为最优，其中整体性评价比现有的最优对比模型在VQA-CP v2上高~3%，在VQA v2上高~2%。实验证明了我们提出的模型可以保持很好的鲁棒性。

➤ 工作(a)可视化

模型可以提供准确的视觉信息

Q: how many **people** are in the picture?



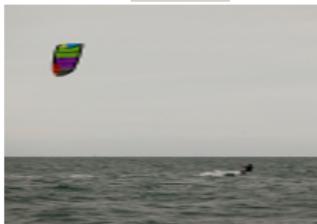
GT: 4

Q: what are there a **lot** of being pictured?



GT: Signs

Q: is the person in the water **surfing**?



GT: yes

Q: what colors are the **bananas**?



GT: Yellow

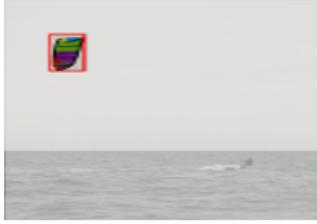
模型可以很好的克服语言偏差，保证鲁棒性



UpDn: 2



UpDn: Signs



UpDn: no



UpDn: Green



GGE: 4



GGE: Signs



GGE: no



GGE: Green



Ours: 4



Ours: Signs



Ours: yes



Ours: Yellow

模型具备很好的鲁棒性，准确的回答问题，提供准确的视觉原因

模型可以解决视觉偏差



➤ 工作(a)总结

该部分创新点：

- 创新性地引入了基于**POS tagging**的特征融合策略，增强显著视觉特征，提升视觉信息利用能力；
- 采用**基于对比采样**的方式扩充图像样本和显著区域样本的空间；
- 创造性地提出了二次消偏模型来解决问题-答案的关联分布，即对**答案预测消偏及预测答案和真实答案所构成的伪标签消偏**；
- 相关研究成果已经归纳整理成**论文投稿至TIP¹**。

工作--1： Jin Liu, Fengyu Zhou, et al. Question-Conditioned Debiasing with Focal Visual Context Fusion for Visual Question Answering. *ieee transactions on image processing.* (中科院分区：一区 (TOP)) (已投稿, under review)

➤ 工作(b) 问题类型引导采样

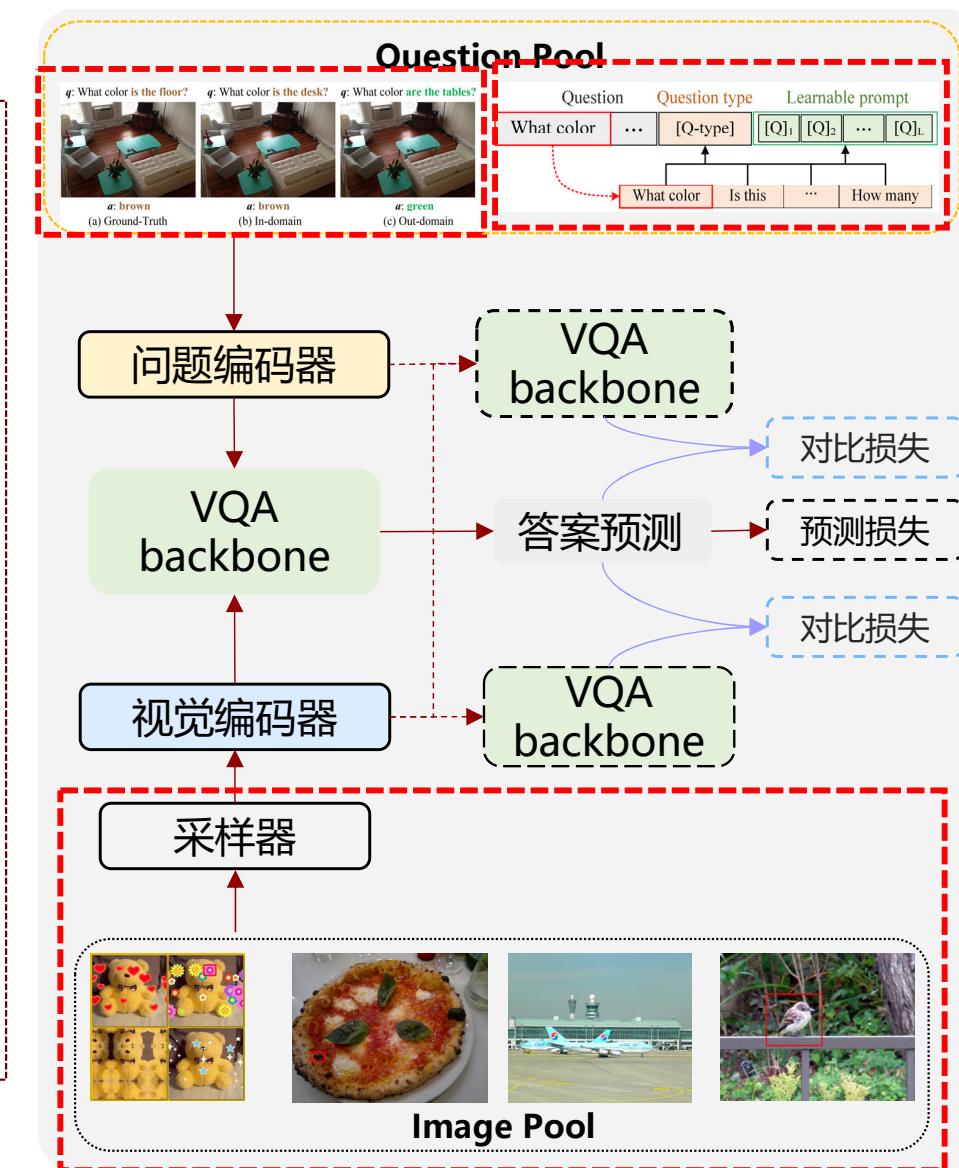
实施策略：

Step1: 从问题类型和答案统计关联角度出发，设计**基于问题类型的对比采样机制**，从in-domain(答案与正样本一致)和out-domain(答案与正样本不一致)；

Step2: 基于**dropout**机制对图片进行采样，减少构建的图片负样本的数目，防止模型过度使用图片信息；

Step3: 引入**不同长度的可学习Prompt**，融合到原始的问题中，以便增强对问题的理解；

Step4: 最后**基于VQA v2、VQA-CP v2等开源基准数据集**来验证所提算法的**可行性和鲁棒性**。



➤ 工作(b)实验结果

基于融合策略



Category	Model	Base	Venue	VQA-CP v2 test (%)				VQA v2 val (%)			
				All	Yes/No	Num	Other	All	Yes/No	Num	Other
I	UpDn [†] (Anderson et al., 2018)	-	CVPR'18	40.63	41.27	13.63	47.70	64.21	81.72	43.54	56.36
	GVQA(Agrawal et al., 2018)	-	CVPR'18	31.30	57.99	13.68	22.14	48.24	72.03	31.17	34.65
	LXMERT(Tan and Bansal, 2019)	-	EMNLP'19	42.84	18.91	55.51	46.23	-	-	-	-

基于标注特征



II	SCR(Wu and Mooney, 2019)	UpDn	NIPS'19	49.45	72.36	10.93	48.02	62.20	78.80	41.60	54.40
	CSS [†] (Chen et al., 2020)	UpDn	CVPR'20	41.16	43.96	12.78	47.48	59.21	72.97	40.00	55.13
	HINT(Selvaraju et al., 2019)	UpDn	ICCV'19	47.50	67.21	10.67	46.80	63.38	81.18	42.14	55.66

基于ensemble策略



III	LMH(Clark et al., 2019)	UpDn	EMNLP'19	52.73	72.95	31.90	47.79	56.35	65.06	37.63	54.69
	DLR(Jing et al., 2020)	UpDn	AAAI'20	48.87	70.99	18.72	45.57	57.96	76.82	39.33	48.54
	SSL-VQA(Zhu et al., 2020)	UpDn	IJCAI'20	57.59	86.53	29.87	50.03	63.73	-	-	-
	LPF(Liang et al., 2021)	UpDn	SIGIR'21	55.34	88.61	23.78	46.57	55.01	64.87	37.45	52.08
	GGE(Han et al., 2021)	UpDn	ICCV'21	57.32	87.04	27.75	49.59	59.11	73.27	39.99	54.39
	D-VQA [†] (Wen et al., 2021)	UpDn	NIPS'21	60.64	88.51	46.75	49.85	64.04	81.32	43.65	56.30
	LRSV(Guo et al., 2021)	UpDn	TIP'22	47.09	68.42	21.71	42.88	55.50	64.22	39.61	53.09

基于对比学习或者因果推理



IV	AdvReg.(Ramakrishnan et al., 2018)	SAN	NIPS'18	33.29	56.65	15.22	26.02	52.31	69.98	39.33	47.63
	RUBi(Cadene et al., 2019)	S-MRL	NIPS'19	47.11	68.65	20.28	43.18	61.16	-	-	-
	ECD(Kolling et al., 2022)	LMH	WACV'22	59.92	83.23	52.59	49.71	-	-	-	-
III	Ours	UpDn	-	61.95	89.50	52.44	50.12	65.26	82.38	44.77	57.67

结论：我们的模型在整体性评价上取得了最优的结果，在VQA-CP v2上高~2%，在VQA v2上高~1.5%。实验证明了我们提出的模型保持了很好的鲁棒性，同时维持了很好的in-domain的性能

➤ 工作(b)消融实验

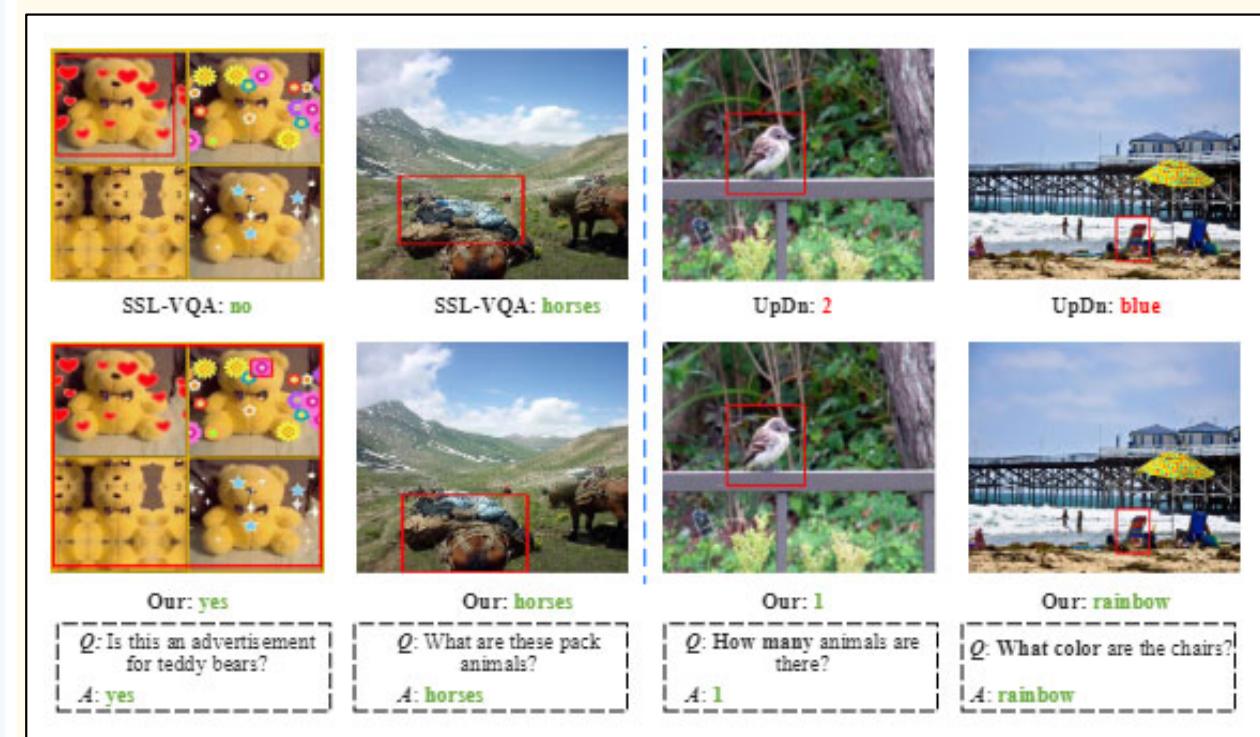
NSR	NSQ	Prompt	All	Yes/No	Num	Other
✓			40.63	41.27	13.63	47.70
✓	✓		58.41	88.54	32.68	49.68
✓	✓	✓	59.83	88.89	40.36	49.48
✓	✓	✓	61.95	89.50	52.44	50.12

Models	All	Yes/No	Num	Other
Baseline	40.63	41.27	13.63	47.70
+ SSL-VQA(Zhu et al., 2020)	57.59	86.53	29.87	50.03
+ NSR (Ours)	58.41	88.54	32.68	49.68

Models	All	Yes/No	Num	Other	$\Delta \uparrow$
Baseline	40.63	41.27	13.63	47.70	-
+ ECD(Kolling et al. (2022))	40.69	41.71	13.41	47.63	+0.06
+ CSS(Chen et al. (2020))	40.05	42.16	12.30	46.56	-0.58
+ D-VQA(Wen et al. (2021))	41.40	43.33	12.97	48.19	+0.77
+ NSQ (ours)	44.12	50.92	13.27	49.02	+3.49
+ NSQ (random)	41.00	42.81	13.29	47.44	+0.37

Models	All	Yes/No	Num	Other
baseline†	59.83	88.89	40.36	49.48
[Q]+context+type	61.03	89.99	46.77	49.76
[Q]+type+context	61.95	89.50	52.44	50.12
context+[Q]+context+type	60.62	89.86	48.58	48.61
[Q]+answer+context	60.85	89.52	48.52	49.21

➤ 工作(b)可视化



结论：

- ①模型中设计的各个模块均对模型性能起到一定的促进作用；
- ②相比于之前的算法模型，我们提出的创新点极大的提升了模型的精度；
- ③从可视化结果可以看出来，模型可以很好的处理视觉偏差和语义偏差，准确的回答问题，保证了很好的鲁棒性。



➤ 工作(b)总结

该部分创新点：

- 创新性地提出了**基于问题类型的采样机制**构建问题负样本策略；
- 创新性地将**dropout**机制引入到图像采样策略中，减少对视觉信息的过度依赖，平衡模态信息；
- 创造性地设计了一种**可学习的Prompt机制**，提升模型对问题的理解能力，并将其引入到训练和测试过程；
- 相关研究成果已经归纳整理成**论文投稿至IPM¹**。

工作--2： Jin Liu, Fengyu Zhou, et al. Be Flexible! Learn to Sample and Prompt for Robust Visual Question Answering. **information processing and management.** (中科院分区：一区 (TOP)) (已投稿, under review)

➤ 研究内容二：基于知识增强的视觉问答方法

研究目的：为了提升提升模型的推理能力，通过引入知识图谱（结构化）或描述性文本知识（非结构化），使模型可以在新场景下迅速检索到相关知识

研究思路：

- 1) 通过引入**层次化知识表示算法**对检索出来的知识图谱内容进行特征化表示；
- 2) 借助**特征学习和关系建模算法**，降低对检索的子图的质量要求，保证在小样本环境下，**模型依然保持足够的建模能力**；
- 3) 拟充分融合多种来源的数据知识，在保证模型鲁棒性的基础上，**提升VQA模型的性能**，增加模型对新场景的适应能力。



➤ 工作(a)——层次化空间建模下的知识表征

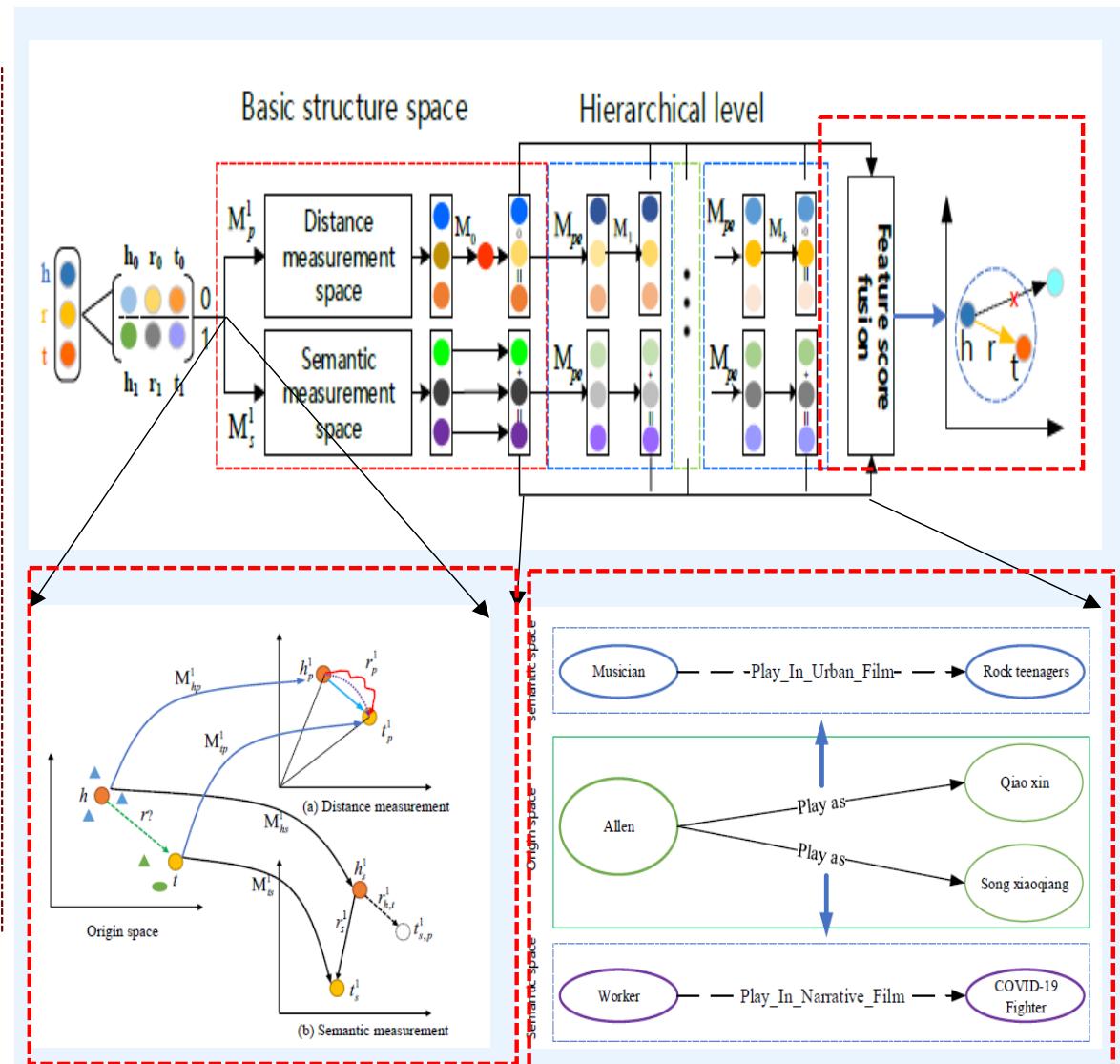
实施策略：

Step1：提出了双空间建模的策略，将三元组中的头尾实体映射至不同的距离空间和语义空间；

Step2：引入分层机制，获取模型在不同层次的特征；

Step3：引入调整系数，将获取底层、高层的语义和距离度量的特征信息进行均衡融合；

Step4：最后基于WN18RR、FB15K-237等开源基准数据集来验证所提算法的可行性和鲁棒性。

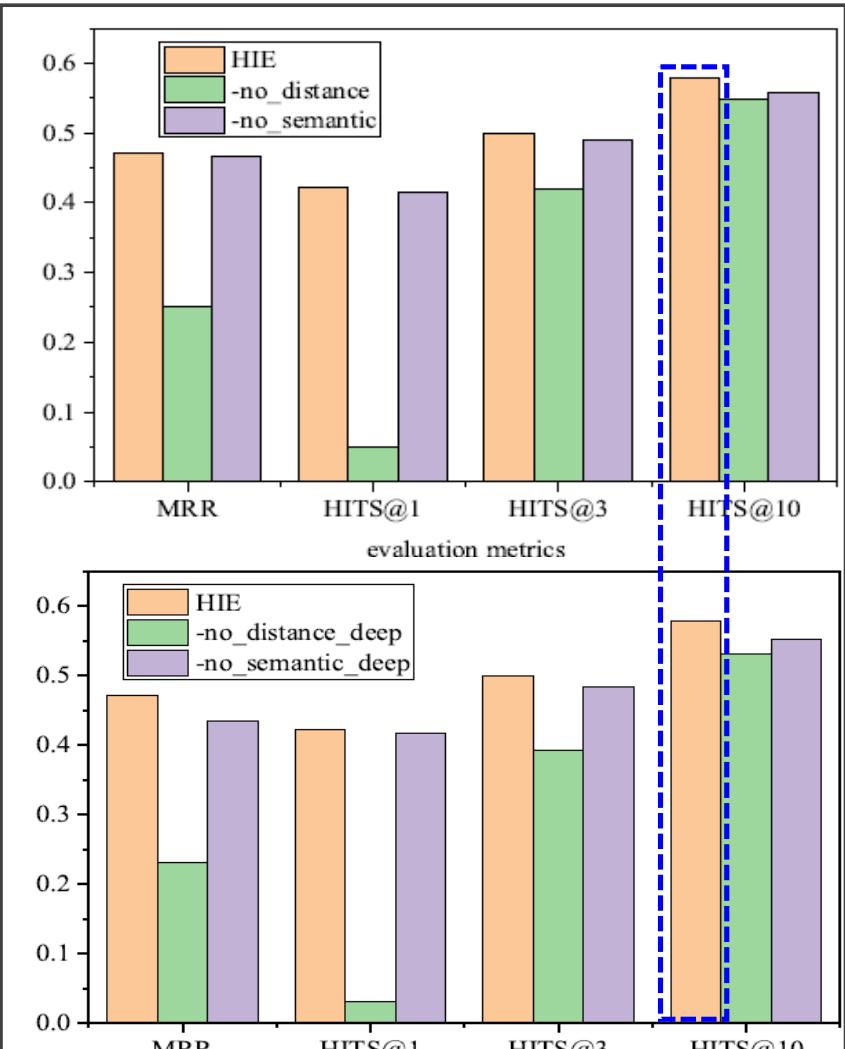


➤ 工作(a)实验结果

Methods	WN18					YAGO3-10									
	MRR↑		MR↓		Hits@↑			MRR↑		MR↓		Hits@↑			
			1	3	10			1	3	10			1	3	10
TransE ^[*]	0.454	-	0.089	0.823	0.934			0.238	-	0.212	0.361	0.447			
TransR ^[*]	0.605	-	0.335	0.876	0.940			0.256	-	0.223	0.356	0.478			
TransD(unif)	-	229	-	-	0.925			-	-	-	-	-			
RotatE	0.949	309	0.944	0.952	0.959			0.495	1767	0.402	0.550	0.670			
DistMult ^[•]	0.822	902	0.728	0.914	0.936			0.340	5926	0.237	0.379	0.540			
ComplEx ^[•]	0.941	-	0.936	0.945	0.947			0.355	6351	0.258	0.399	0.547			
M-DCN	0.950	-	0.946	0.954	0.958			0.505	-	0.423	0.587	0.682			
InteractE	-	-	-	-	-			0.541	2375	0.462	-	0.687			
ConvE	0.942	504	0.955	0.947	0.935			0.523	2792	0.448	0.564	0.658			
R-GCN	0.819	-	0.697	0.929	0.964			-	-	-	-	-			
HIE	0.930	131	0.913	0.954	0.970			0.542	2042	0.452	0.593	0.695			

Methods	WN18RR					FB15k-237									
	MRR↑		MR↓		Hits@↑			MRR↑		MR↓		Hits@↑			
			1	3	10			1	3	10			1	3	10
TransE ^[•]	0.226	-	-	-	0.501			0.294	357	-	-	0.465			
MuRP	0.477	-	0.438	0.489	0.555			0.324	-	0.235	0.356	0.506			
RotatE	0.476	3340	0.428	0.492	0.571			0.338	177	0.241	0.375	0.533			
DistMult	0.430	5110	0.390	0.440	0.490			0.241	254	0.155	0.263	0.419			
TorusE	0.452	-	0.422	0.464	0.512			0.305	-	0.219	0.337	0.485			
ConvKB ^[•]	0.220	2741	-	-	0.508			0.302	196	-	-	0.483			
KBGAN	0.215	-	-	-	0.469			0.277	-	-	-	0.458			
R-GCN	-	-	-	-	-			0.249	-	0.151	0.264	0.417			
HIE	0.480	2821	0.430	0.499	0.580			0.346	215	0.255	0.380	0.523			

➤ 工作(a)消融实验



结论：我们的模型在图谱表示建模上**具有较强的能力**，可以实现对知识的准确表示；此外，**层次距离及语义学习策略可以有效的提升模型的性能**。



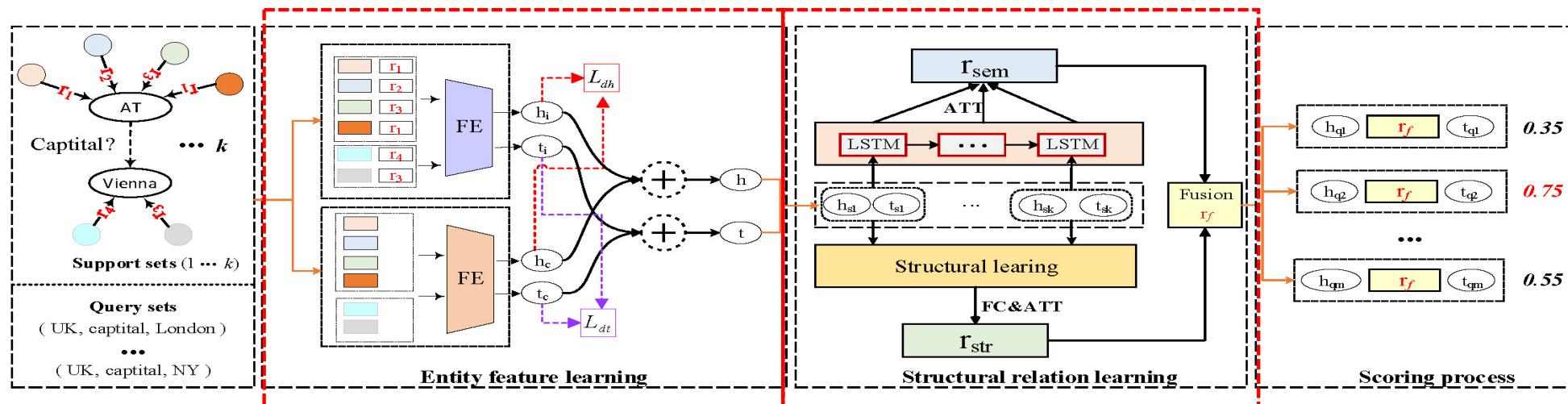
➤ 工作(a)总结

该部分创新点：

- 创新性地引入了双空间的建模机制，即语义空间和距离度量空间，在保证三元组语义的基础上，同时保证了其结构化信息；
- 创新性地引入了层次化建模机制，采用多层特征提取机制来不断获取高维语义和空间状态信息，从而保证三元组信息的完备性；
- 相关研究成果已经归纳整理成论文投稿至Big Data¹。

工作--3：Jin Liu, Fengyu Zhou, et al. Joint embedding in Hierarchical distance and semantic representation learning for link prediction. Big data. (中科院分区：二区) (已投稿, under review)

➤ 工作(b)——小样本场景下知识表征



实施策略：

Step1: 在知识表示的基础上，引入了**实体表征学习策略**，借助**双分支结构**提取不同的邻居特征，并依据相似度的多样性约束函数，保证两个分支不但可以**学习到互补的信息**，而且可以**充分利用稀疏邻居的特征**；

Step2: 在建模的实体表示的基础上，引入**结构化关系建模机制**，在语义表示学习的基础上，融入结构化信息，保证分数评价过程中的关系一致性；

Step3: 探究**不同结构化建模策略**，验证一致性建模的必要性，同时选择最优的结构化建模方案；

Step4: 最后基于**Wiki-One、NELL-One等开源基准数据集**来验证所提算法的**可行性和鲁棒性**。

➤ 工作(b)实验结果

Category	Model	NELL-One						Wiki-One					
		1-shot		3-shot		5-shot		1-shot		3-shot		5-shot	
		MRR	Hits@10	MRR	Hits@10	MRR	Hits@10	MRR	Hits@10	MRR	Hits@10	MRR	Hits@10
Traditional KGE-based	TransE	0.093	0.192	0.193	0.320	0.174	0.313	0.035	0.052	0.111	0.176	0.133	0.187
	RESCAL	0.140	0.229	0.223	0.383	-	-	0.072	0.082	0.081	0.126	-	-
	DistMult	0.102	0.192	0.231	0.375	0.200	0.311	0.035	0.052	0.112	0.195	0.071	0.151
	ComplEx	0.131	0.223	0.185	0.273	0.184	0.297	0.069	0.121	0.106	0.145	0.080	0.181
Metric matching-based	GMatching(TransE)	0.171	0.255	0.279	0.464	-	-	0.219	0.328	0.171	0.324	-	-
	GMatching(DistMult)	0.171	0.301	-	-	-	-	0.222	0.340	-	-	-	-
	FSRL	0.256	0.435	0.318	0.507	0.153	0.319	0.128	0.301	0.241	0.406	0.158	0.287
	FAAN	0.194	0.331	-	-	0.279	0.428	0.272	0.387	-	-	0.341	0.463
Meta relational learning-based	MetaR(In-Train)	0.250	0.401	-	-	0.261	0.437	0.193	0.280	-	-	0.221	0.302
	MetaR(Pre-Train)	0.164	0.334	-	-	0.209	0.355	0.344	0.404	-	-	0.342	0.463
	FKGC	0.307	0.483	0.322	0.510	0.344	0.517	0.301	0.416	0.331	0.435	0.351	0.446
	Ours	0.331	0.531	0.346	0.553	0.374	0.566	0.343	0.460	0.341	0.463	0.346	0.472

➤ 工作(b)消融结果

提升~5%

提升~3%

scoring	modeling	TransE($\ h+r-t\ $)				RotatE($\ h \circ r-t\ $)				SE($\ h+E-t\ $)			
		MRR	Hits@10	Hits@5	Hits@1	MRR	Hits@10	Hits@5	Hits@1	MRR	Hits@10	Hits@5	Hits@1
	TransE	0.374	0.566	0.468	0.282	0.239	0.443	0.350	0.147	0.313	0.505	0.412	0.210
	RotatE	0.369	0.552	0.458	0.279	0.250	0.445	0.351	0.149	0.317	0.511	0.421	0.214
	SE	0.276	0.516	0.389	0.172	0.191	0.395	0.299	0.101	0.341	0.532	0.425	0.252

DL	F1	F2	F1	MRR	Hits@10	Hits@5	Hits@1
✓	✓	✓		0.374	0.566	0.468	0.282
✓	✓		✓	0.348	0.534	0.446	0.262
✓	✓			0.343	0.550	0.460	0.240
✓				0.324	0.527	0.412	0.229
	✓			0.284	0.500	0.374	0.200

SR	TR	MRR	Hits@10	Hits@5	Hits@1
✓	✓	0.374	0.566	0.468	0.282
	✓	0.345	0.556	0.447	0.242
✓		0.356	0.550	0.455	0.255

结论：我们的模型在小样本知识图谱上具有很强的知识表达能力，同时引入实体表征学习机制和结构化关系建模策略提升了模型的对长尾三元组的建模能力，此外，也验证了一致性建模对模型的性能起到关键作用。



➤ 工作(b)总结

该部分创新点：

- 创新性地引入了实体表征学习策略，获取稀疏邻居的特征，提升建模中中心节点实体的特征表达能力；
- 创新性的引入了结构化关系建模，实现了三元组建模过程中结构化和语义关系的一致性；
- 探索了多种结构化建模和融合策略，验证了一致性建模的必要性。
- 相关研究成果已经归纳整理成论文投稿至TKDE¹。

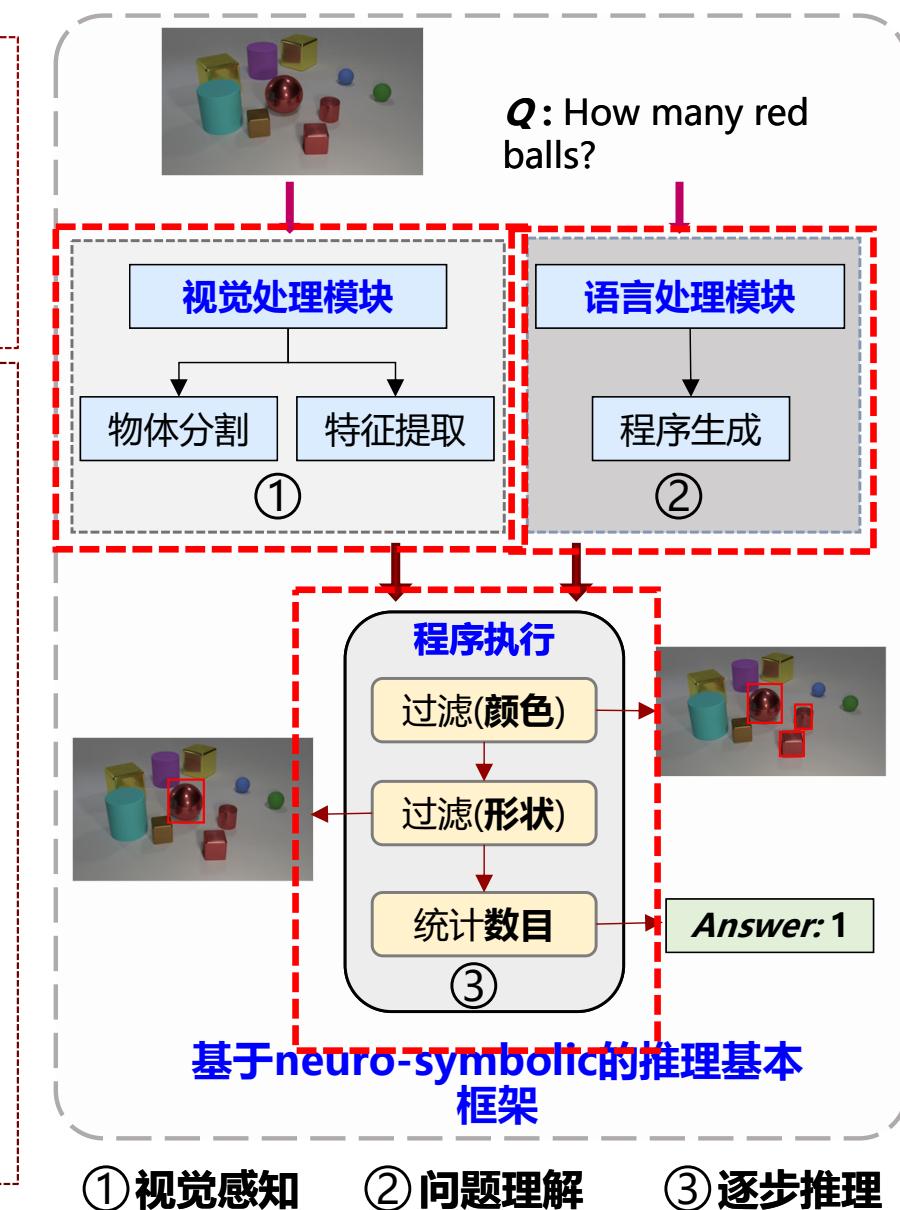
工作--4： Jin Liu, Fengyu Zhou, et al. Complete Feature Learning and Consistent Relation Modeling for Few-shot Knowledge Graph Completion. *ieee transactions on knowledge and data engineering.* (中科院分区：二区) (已投稿, under review)

➤ 研究内容三：基于neuro-symbolic的视觉问答方法

研究目的：为了保证**推理过程的可解释性和可溯源性**，在保证推理过程鲁棒性的基础上，**解耦视觉编码(①)和问题理解(②)**，通过相应的推理步骤设计，**实现模型的逐步推理(③)**，

研究思路：

- 1) 拟结合neuro-symbolic方法，将问题端的符号表示和视觉信息特征进行对齐，充分**解耦感知和推理**的两个过程；
- 2) 拟**引入课程学习的概念，借助研究内容二学习到的知识**，从简单的物体属性(如颜色、形状等)开始学习，逐步实现复杂的推理逻辑；
- 3) 拟**引入弱监督的训练策略**，通过设计极少量的推理步骤来引导整体推理策略，实现推理步骤的快速、准确生成。



① 视觉感知

② 问题理解

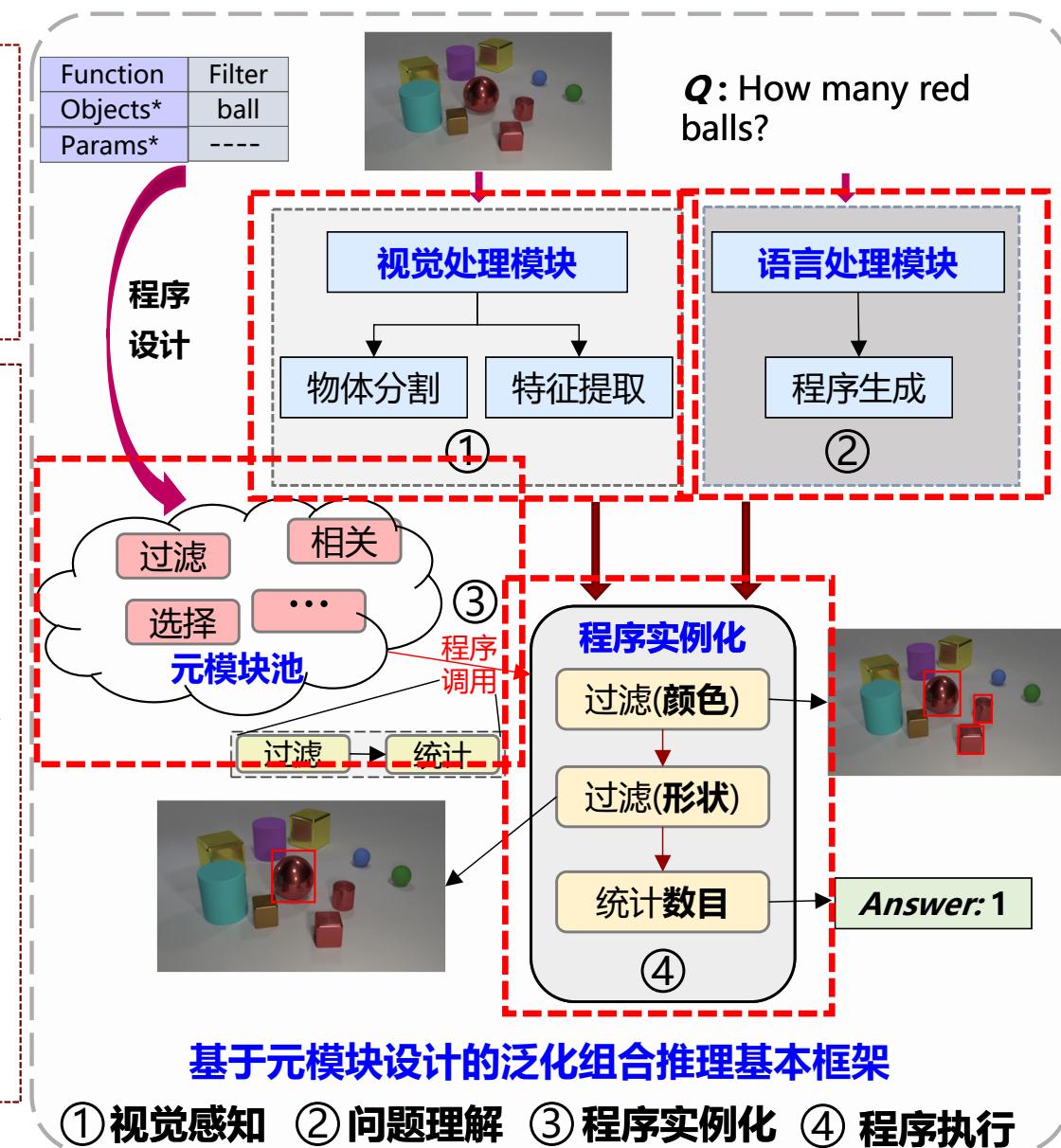
③ 逐步推理

➤ 研究内容四：基于meta-module的视觉问答方法

研究目的：为了提升模型的泛化组合性，通过抽取概念和元概念设计更为高层的meta-module，在可解释性的基础上，引入由粗到细的步骤分配策略，自由实例化程序。

研究思路：

- 1) 拟通过引入蒸馏策略，借助教师-学生模型将更多的symbolic知识传递到程序实例化模块，使模型具备一定的**模块泛化性**；
- 2) 拟研究**基于物体级别的概念图构建**，**减少对更加细粒度和准确的物体识别模型的依赖**；
- 3) 拟引入预训练语言模型，**借助研究内容二的知识**，对组合的步骤进行评价，**增加程序实例化的可靠性**。



基于元模块设计的泛化组合推理基本框架

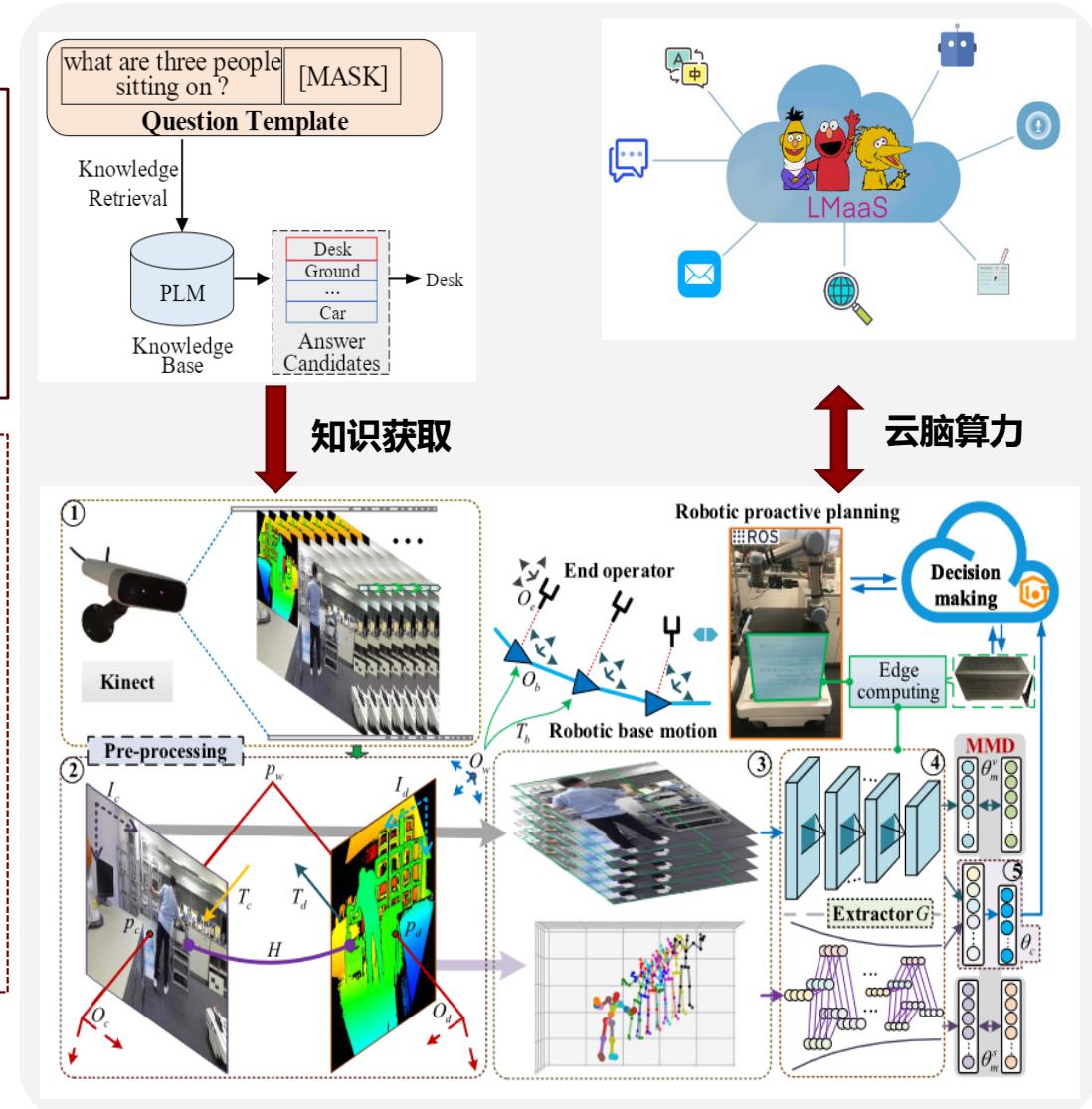
① 视觉感知 ② 问题理解 ③ 程序实例化 ④ 程序执行

➤ 在服务机器人上的应用（模型落地）

研究目的：借助智慧云脑提供的算力，将具备**推理能力的视觉问答模型**部署到**服务机器人端**，验证机器人在多种实验环境下的推理性问答能力。

研究方案：

- 1) 拟在图像的视觉问答模型基础上，研究**基于视频的推理模型**，提升机器人端的场景理解速度和处理速度；
- 2) 拟将**云脑中的知识**引入到**服务机器人执行动作**的过程中，提升机器人在不同场景下的任务执行能力；



课题创新点

— 学无止境 气有浩然 —



➤ 创新点

➤ 创新点一：提出了基于对比采样和特征融合的视觉问答模型方法。

研究通过引入对比采样，扩充负样本，并借助视觉特征和问题prompt的设计来提升模型的预测性能，保证模型可以预测正确的答案，同时能给出正确的视觉理解区域，**构建具备极强鲁棒性的视觉问答系统。**

➤ 创新点二：提出了基于知识增强的视觉问答模型方法。

研究通过引入知识图谱（结构化）或描述性文本知识（非结构化），**使模型可以在新场景下迅速检索到相关知识**，丰富模型的推理过程和推理能力。

➤ 创新点三：提出了基于neuro-symbolic机制的视觉问答模型方法。

研究进一步解耦视觉编码和问题编码过程，充分发挥现有视觉模型和文本处理模型的优势，通过引入课程学习等机制，**在实现推理过程可解释的基础上**，实现更为复杂的推理逻辑。

➤ 创新点四：提出了基于meta-module设计的视觉问答模型方法。

研究通过抽取概念和元概念设计更为高层的meta-module，设计由粗到细的步骤分配策略，并借助视觉概念图，**来提升模型的泛化组合性。**

课题计划安排



➤ 计划安排

时间节点	预期任务
2021年09月——2022年03月	查阅对话系统相关的文献，确定智能视觉问答系统的主要 研究趋势和研究内容 ；
2022年04月——2022年10月	从视觉问答的推理机制入手，研究实现推理的算法 模型进展 ；
2022年11月——2023年04月	研究模型 鲁棒性 问题；
2023年05月——2023年12月	研究 知识融合性 ；
2024年01月——2024年07月	研究 可解释性和组合泛化性 ；
2024年07月——2024年10月	将视觉问答系统 部署至服务机器人 ，进行 实验验证 ；
2024年11月——2025年06月	整理论文，毕业 答辩 。



已取得成果总结

申请发明专利1项：

序号	专利名	作者	状态
P1	基于动态向量混合遗传算法的云机器人服务选择方法及系统	周风余, 刘进, 等.	二次审查

竞赛及获奖：

序号	竞赛名	名次
1	“华为杯”第十八届中国研究生数学建模竞赛	国家二等奖
2	山东省第七届“互联网+	山东省银奖
3	山东省第八届“互联网+	校级金奖
4	山东大学研究生一等学业奖学金	校级一等

共撰写文章6篇，投稿SCI文章4篇，待投稿A类会议1篇；已接受B类会议1篇：

序号	题目	期刊	作者	收录情况
C1	Syntax Controlled Knowledge Graph-to-Text Generation with Order and Semantic Consistency	NAACL2022	Jin Liu, Fengyu Zhou, el.al.	EI会议检索
C2	Separation Grounding with Knowledge-Guided for Video Question Answering	CVPR2023	Jin Liu, Fengyu Zhou, el.al.	暂无
J1	Joint embedding in Hierarchical distance and semantic representation learning for link prediction	Big Data	Jin Liu, Fengyu Zhou, el.al.	Under review
J2	Complete Feature Learning and Consistent Relation Modeling for Few-shot Knowledge Graph Completion	TKDE	Jin Liu, Fengyu Zhou, el.al.	Under review
J3	Be Flexible! Learn to Sample and Prompt for Robust Visual Question Answering	Information Processing and Management	Jin Liu, Fengyu Zhou, el.al.	Under review
J4	Question-Conditioned Debiasing with Focal Visual Context Fusion for Visual Question Answering	TIP	Jin Liu, Fengyu Zhou, el.al.	Under review