Visual Dialogue: From VisDial to FlipDial

21821120 Guangzhao Cheng
College of Computer Science and Technology, Zhejiang University
cgz@zju.edu.cn

Abstract

虽然视觉问答（Visual Question Answer ,VQA）向人机交互迈出了重要一步，但它仍然只代表一轮对话。而视觉对话给定一张图片、会话历史以及关于该图像的一个问题，需要机器将问题置于图像中，从历史中推断出背景，并准确地回答该问题。本文首先介绍了自然语言处理领域的对话系统，然后介绍了视觉对话的任务定义、数据集和评估标准。接着介绍和评述了视觉对话的几篇经典论文和最新论文，并尝试和复现了其中的部分实验。

1 对话系统（Dialogue System）

作为人工智能的终极难题之一，一个完整的人机对话系统涉及到的技术极为广泛，例如计算机科学中的语音技术，自然语言处理，机器学习，规划与推理，知识工程，甚至语言学和认知科学中的许多理论在人机对话中都有所应用。

对话系统是自然语言处理（NLP）领域的任务。一般可以将对话系统分为两类：闲聊型对话系统（Non-Task-Oriented Dialogue Systems & Chatbot & Open Domain Dialogue）和任务型对话系统（Task-Oriented Dialogue Systems）。

Chatbot 的典型例子如 Microsoft Xiaoice，它不以完成某项具体的任务为目的，对话轮数和有意义的回答是主要评估标准。对于每轮对话的回复，有基于生成的模型和基于检索的模型。

任务型对话系统的典型例子如阿里小蜜，它的目的是帮助用户完成一个特定的任务，如订餐、订票等。任务驱动的多轮对话不是一个简单的自然语言理解加信息检索的过程，而是一个决策过程，需要机器在对话过程中不断根据当前的状态决策下一步应该采取的最优动作。问答系统和任务驱动的多轮对话最根本的区别在于系统是否需要维护一个用户目标状态的表示和是否需要一个决策过程来完成任务。

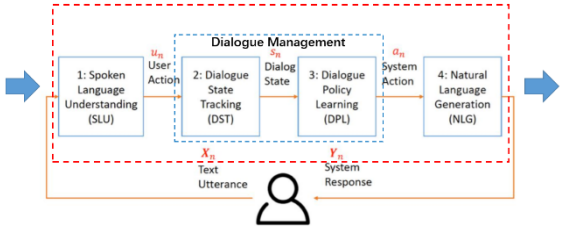


图 1　任务型对话系统

任务型对话系统主演包括 3 个子模块，自然语言理解（Spoken Language Understanding）、对话管理（Dialog Management）和自然语言生成（Natural Language Generation）。其中最重要的是对话管理，其又包括对话状态追踪（Dialogue State Tracking）和对话策略（Dialog Policy）。最近，随着 Seq-2-Seq 方法的兴起，这几者之间的界限在变得模糊。

2 视觉对话（Visual Dialogue）

2.1 任务定义

[1]首先提出了 Visual Dialogue 这个任务。Visual Dialogue 要求 agent 与人类维持一个与图像内容相关的用自然语言交谈的对话。给出图像，对话历史和一个关于该图像的问题，agent 必须必须将问题置于图像中，从历史推断背景，并准确回答该问题。其展示了第一个视觉对话机器人！[1]

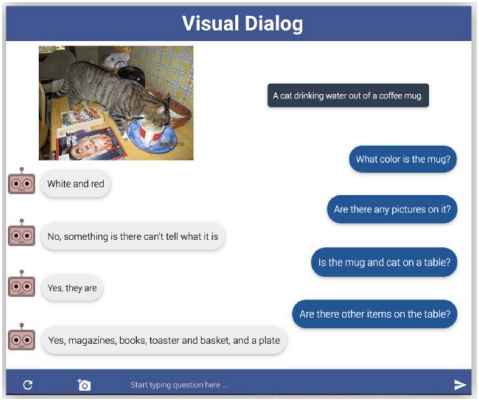


图 2　视觉对话示意图

任务定义：给定一个图像 *I*，一个由一系列问答对组成的对话历史 S(Q1: ‘How many people are in wheelchairs?’, A1: ‘Two’, Q2: ‘What are their genders?’,

[1] https://visualdialog.org/

A2: ‘One male and one female’),以及一个自然语言跟进问题 Q(Q3: ‘Which one is holding a racket?’)，机器的任务是用自由形式的自然语言回答问题(A3: ‘The woman’)。 注意：提问者是看不到图像。

2.2 数据集

[1]同时还提出了一种全新的双人聊天数据集收集协议并开发了 VisDial 数据集[2]，数据集的版本有 v0.5、v0.9 和 v1.0（v1.0 数据集格式示例见附录 1）。该数据的图像来自于 Common Objects in Context (COCO)[[2]]。他们在 AMT 上雇用了 2 名工作人员，以便实时聊天，其中一人 questioner 只看到描述图像的一行文字（来自 COCO 的标题）但是看不到图像；answerer 可以看到图像和标题，其任务是回答 questioner 提出的问题。与 VQA[3]不同，答案不限于简短或简洁，而是鼓励其尽可能自然地用自然对话来问答。

2.3 评估

对话系统的另一个基本挑战是如何评估，[[1]]也为 Visual Dialog 提出了一种基于检索的评估协议，其不是评估下游任务[7]或全面评估整个对话（如无目标聊天[8]），而是评估每一轮的个人反应。在每一轮要求 agent 对一组候选答案进行排序。

3 相关论文评述

3.1 Visual Dialog (CVPR 2017 Spotlight)

[1]首先提出了 Visual Dialogue 这个任务和数据集收集协议并开发了 VisDial 数据集。这篇文章还对数据集进行了多种分析和可视化展示，以区别 VQA：（1）视觉启动偏差（Visual Priming Bias），受试者在询问有关它的问题时会看到一个图像，这导致了其会产生特殊偏见[4][5][6]，而在数据集收集时，questioner 看不到图像则会避免这一点；（2）VisDial 中的答案更长且更具描述性；（3）答案类型更丰富；（4）区分二元问题与二元答案。针对其数据集，这篇文章也提出了不同与 NLP 领域对话的评估标准。

这篇文章展示了第一视觉聊天机器人。同时这篇文章还为 Visual Dialog 引入了一系列神经网络模型（Encoder-Decoder Models），在这个模型中，Encoder 的输入是 (I,H,Q_t)，其中 I 是 answerer 可以看到图像，H 是对话历史，$H=(C,(Q_1,A_1),\ldots,(Q_{t-1},A_{t-1}))$，$Q_t$ 是问题和 A_t 是 100 个候选答案，$A_t=\{A_t^1,\ldots,A_t^{100}\}$。Decoder 把向量空间转化成输出。其中 Encoder 包括 3 种模型（Late Fusion, Hierarchical Recurrent Encoder 和 Memory Network），Decoder 包括 2 种模型（Generative 和 Discriminative），并进行了大量的组合实验，本次报告对这些实验做了尝试和分析（参见本报告 4.1 节）。

3.2 Learning Cooperative Visual Dialog Agents with Deep Reinforcement Learning (ICCV 2017 Oral)

[9]为 VQA 和视觉对话引入了第一个目标驱动的培训。其在两个代理 Q-BOT 和 A-BOT 之间构建一个合作的“图像猜测”游戏，它们在自然语言对话中进行通信，以便 Q-BOT 可以从一系列图像中选择一个看不见的图像。

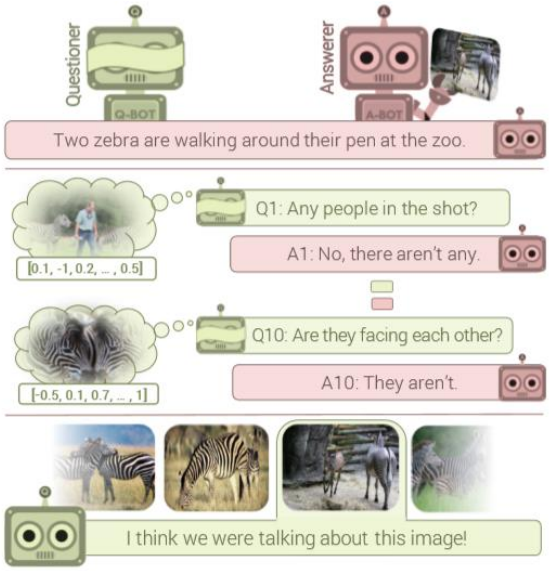


图 3 Cooperative Visual Dialog Agents with DRL

这篇文章进行了两个实验结果。首先，作为纯粹 RL（从头开始）的“健全性检查”演示，在合成世界中展示结果，其中代理人以未接触的词汇进行交流，即没有预先指定含义的符号（X，Y，Z）。结果发现两个机器人发明了他们自己的通信协议，并开始使用某些符号来询问/回答某些视觉属性（形状/颜色/样式）。因此，其证明了“视觉”对话代理之间没有人为监督的基础语言和沟通的出现。

其次，这篇文章在 VisDial 数据集上进行了大规模的实物图像实验，先使用监督对话数据进行预训练，然后用强化学习微调。结果显示“RL+微调”明显优于 SL 代理。有趣的是，RL Q-BOT 学会提出 A-BOT 擅长的问题，最终导致更多信息对话和更好的团队。

在 NLP 领域，先用监督学习进行训练，然后用强化学习进行微调是最新最流行的方法，也是目前最好的方法。但是和 NLP 领域的对话系统不同的是，完全使用 RL 会使得两个机器人发明了他们自己的通信协议。

3.3 Evaluating Visual Conversational Agents via Cooperative Human-AI Games (HCOMP 2017)

[14]首先在 NLP 领域的对话系统提出了“Human-

[2] https://visualdialog.org/data

in-the-loop”。开发会话代理的一个重要方面是让机器人能够通过与人交流来改进，并从中汲取错误。大多数研究都侧重于从固定的标记数据培训集学习，而不是以在线方式与对话伙伴进行交互。但[14]在强化学习环境中探索这个方向，其中机器人通过教师在其生成的反应之后给出的反馈来改进其问答能力。

[10]也探索了 Human-AI 中人与机器合作的问题。目前大多数 AI 的进展通常是孤立地测量的，并没有人参与循环中。这篇文章通过设计了一个合作游戏（GuessWhich）来探究 Human-AI 的表现。AI 是一个视觉对话代理（称之为 ALICE）提供了一个人类看不见的图像。在对图像进行简要描述之后，人们向 ALICE 提出关于该秘密图像的问题，以从固定的图像池中识别它。通过在使用 ALICE 进行固定数量的对话轮次之后人类正确识别秘密图像所需的猜测次数来测量人类 ALICE 团队的表现。

其在监督学习和强化学习的版本的游戏中评估 Human-AI 团队的表现。实验结果表明，尽管在之前的工作中报道了 SL 和 RL 之间存在显着差异，但当与人类伙伴配对时，SL 和 RL 之间的性能没有显着差异。这表明虽然自我对话和 RL 是追求建立更好的视觉对话代理的有趣方向，但 AI-AI 与 Human-AI 的评估之间似乎存在分歧，前者的进展似乎无法预测进展后者。这是一个指导未来研究的重要发现。

3.4 Visual Coreference Resolution in Visual Dialog using Neural Module Networks (ECCV 2018)

[11]提出了 Visual Dialogue 中视觉指代消解（Visual Coreference Resolution）的问题。指代消解是自然语言处理的一大任务之一，它是信息抽取不可或缺的组成部分。在信息抽取中，由于用户关心的事件和实体间语义关系往往散布于文本的不同位置，其中涉及到的实体通常可以有多种不同的表达方式，例如某个语义关系中的实体可能是以代词形式出现的，为了更准确且没有遗漏地从文本中抽取相关信息，必须要对文章中的指代现象进行消解。指代消解不但在信息抽取中起着重要的作用，而且在机器翻译、文本摘要和问答系统等应用中也极为关键。

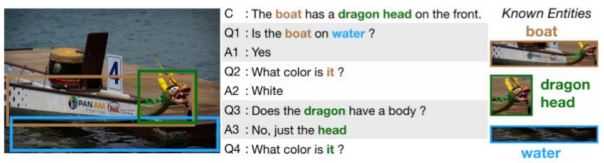


图 4　视觉指代消解

视觉指代消解涉及确定哪些单词（通常是名词短语和代词）共同引用图像中的同一实体/对象实例。这是至关重要的，特别是对于代词（例如“it”），因为对话代理必须首先将它链接到先前的共同参考（例如，“boat”），然后才能依赖于共同参与“boat”的视觉基础推断代词“it”。视觉对话先前的工作中,或者通过历史记录中的存储网络，或者通过整个问题的粗略水平（而不是明确地在词组级别的粒度）来解决这个问题。

在这项工作中，[11]提出了一个用于视觉对话的神经模块网络架构，通过引入两个新模块（Refer 模块和 Exclude 模块）在一个单词级别上形成明确的，基础的共参考分辨率。通过实现接近完美的精度，以及在真实图像上的视觉对话数据集 VisDial 和 MNIST Dialog 上展示了该模型的有效性。

该模型首先识别当前问题中涉及图像中实体（通常是对象和属性）的相关单词或短语。该模型还预测到目前为止在对话框中是否已提及其中的每一个。接下来，如果这些是新颖的实体（在对话历史中看不到），则在继续之前将它们本地化在图像中，并且对于看到的实体，模型预测对话历史中的（第一）相关共参照，并检索其相应的视觉基础。 因此，随着对话进展的轮次，模型收集独特的实体及其相应的视觉基础，并使用此参考池来解决后续问题中的任何共指关系。

3.5 FlipDial: A Generative Model for Two-Way Visual Dialogue (CVPR 2018 Oral)[3]

[12]提出了 FLIPDIAL，这是一种视觉对话的生成模型，同时在视觉基础对话中扮演两个参与者的角色。给定图像形式的上下文和总结图像内容的相关标题，FLIPDIAL 学习回答问题并提出问题，能够产生多种对话（问答对），这些对话是多样的，并且与图片相关。

图 5　FlipDial

要做到这一点，FLIPDIAL 依赖于一个简单但令人惊讶的强大理念：它使用卷积神经网络（CNN）直接编码整个对话，隐式捕获对话上下文，以及条件 VAE 来学习生成模型。使用生成的答案，FLIPDIAL 在 VisDial 数据集上的顺序回答任务（1VD）中的平均等级优于 5 个点，超过了最先进的模型。

而且该论文是第一个将这种范式扩展到完全双向视觉对话（2VD）的地方，其模型能够根据视觉输入顺序生成问题和答案，为此其提出了一套新的评估指标

[3] http://www.robots.ox.ac.uk/~daniela/research/flipdial/

和指标。

3.6 Visual Dialogue without Vision or Dialogue

[13]描述了视觉对话探索中的一些怪癖和缺点：一个连续的问答环节，其中问题和相应的答案通过给定的视觉刺激相关联。为此，这篇文章开发了一种基于典型相关分析（Canonical Correlation Analysis，CCA）的简单方法，该方法在标准数据集上实现了平均等级（MR）的近乎最先进的性能。与计算和时间密集的当前复杂和过度参数化架构形成鲜明对比，该方法忽略了视觉刺激，忽略了对话的顺序，不需要渐变，使用现成的特征提取器，至少有一个参数数量级减少，几乎没有时间学习。其认为这些结果表明当前的视觉对话方法存在问题，并进行分析以突出隐含的数据集偏差和过度约束的评估指标的影响。

4 实验

本报告所进行的实验主要采用的数据集是 v0.9 和 v1.0，其格式参见本报告附录 1，详细介绍参见 2.2 节。

4.1 Visual Dialog (CVPR 2017)实验

[1]中提出了为 Visual Dialogue 引入了一系列神经网络模型（Encoder-Decoder Models），包括 3 种 Encoder 模型（Late Fusion, Hierarchical Recurrent Encoder 和 Memory Network）和 2 种 Decoder 模型（Generative 和 Discriminative）。

在实验中，使用[encoder]-[input]-[decoder]的方式对模型进行命名。但是本次实验把 Q、I、H 全部输入且在 v0.9 和 v1.0 两个数据集上进行实验。

表 1　Visual Dialog (CVPR 2017)实验

模型	SMRR		R@1		R@5		R@10	
	v0.9	v1.0	v0.9	v1.0	v0.9	v1.0	v0.9	v1.0
LF-QIH-G	0.520	0.581	41.82	43.83	61.79	70.23	67.60	71.23
HRE-QIH-G	0.523	0.578	42.29	44.15	62.15	71.34	67.90	70.23
MN-QIH-G	0.525	0.582	42.29	44.25	62.76	72.23	68.89	72.13
LF-QIH-D	0.580	0.610	43.73	46.31	74.68	**78.23**	84.07	87.14
HRE-QIH-D	0.584	**0.621**	44.25	47.12	74.58	78.12	84.22	86.59
MN-QIH-D	**0.596**	0.620	**45.55**	**47.33**	**76.12**	77.98	**85.27**	**88.28**

实验结果分析：（1）v1.0 数据集要好于 v0.9 数据集，这也是扩大数据集的必然结果；（2）所有判别模型都明显优于一般模型，这是预期的，因为反思模型可以调整到答案选项中的偏差；（3）最好的生成和判别模型是 MN-QIH-G 和 MN-QIH-D。上述实验结果和结论基本和论文描述的一致。

4.2 Visual Coreference Resolution (ECCV 2018)实验

原文在 MNIST Dialog Dataset 和 VisDial v0.9 Dataset 两个数据集上进行了实验。本次实验则在 VisDial v0.9 和 VisDial v1.0 上进行实验。

表 2　Visual Coreference Resolution (ECCV 2018)实验

模型	SMRR		R@1		R@5		R@10	
	v0.9	v1.0	v0.9	v1.0	v0.9	v1.0	v0.9	v1.0
LF-QIH-D	0.554	0.623	40.95	44.05	72.43	74.88	82.23	84.26
HRE-QIH-D	0.540	0.617	39.43	42.15	70.34	73.18	81.50	83.22
MN-QIH-D	0.552	0.629	40.03	43.95	71.25	73.98	83.30	84.29
NMN	0.587	0.703	44.15	48.25	76.88	78.02	86.88	88.22
CorefNMN	**0.612**	**0.725**	**48.34**	**49.35**	**78.89**	**79.88**	**88.06**	91.02

实验结果分析：CorefNMN 在 VisDial v0.9 和 v1.0 两个数据集都优于先前工作，同时通过构造更具可解释性，真实性和一致性。

5 总结与展望

笔者目前的研究方向是 NLP 领域的对话系统，通过此次报告的文献综述和实验，发现 NLP 领域的对话系统和视觉对话的发展有些相似的趋势，一些在 NLP 领域流行的方法，在 Visual Dialog 中也有同样的运用，结合本次报告，做出如下总结与展望。

（1）在 NLP 领域的对话系统中，目前都会增加人类评估。正如我们之前所说，如何评估是对话系统的一个基本挑战。加入人类评估是一种策略。

（2）用监督训练学习，然后用强化学习进行微调是目前流行也是效果最优的方法。

（3）但当与人类作为一方去参与对话的时候，SL 和 RL 之间的性能没有显着差异。**这表明虽然自我对话和 RL 是追求建立更好的视觉对话代理的有趣方向，但 AI-AI 与 Human-AI 的评估之间似乎存在分歧，前者的进展似乎无法预测进展后者。但实际上我们的最终目标是后者，也就是人类与机器进行对话。**

（4）VQA 可以看成一轮的对话，本文介绍的 Visual Dialog 也只是把对话历史加进去，并不是像 NLP 领域的对话系统那样进行真正多轮的对话。**多轮对话的重点是决策问题，把图像加入真正的多论对话，这也是为未来发展的一个趋势。**

（5）在对话中引入 KB 等知识库可以大大增加 agnet 的智能程度。为了扩充 KB，还可以**引入 Lifelong Learning（终身学习）来动态扩充 KB。这是笔者预测的另一个趋势和方向。**

参考文献(References):

[1] Das A, Kottur S, Gupta K, et al. Visual dialog[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 326-335.

[2] Lin T Y, Maire M, Belongie S, et al. Microsoft coco: Common objects in context[C]//European conference on computer vision. Springer, Cham, 2014: 740-755.

[3] Antol S, Agrawal A, Lu J, et al. Vqa: Visual question answering[C]//Proceedings of the IEEE international conference on computer vision. 2015: 2425-2433.

[4] Agrawal A, Batra D, Parikh D. Analyzing the behavior of visual question answering models[J]. arXiv preprint arXiv:1606.07356, 2016.

[5] Goyal Y, Khot T, Summers-Stay D, et al. Making the V in VQA matter: Elevating the role of image understanding in Visual Question Answering[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 6904-6913.

[6] Zhang P, Goyal Y, Summers-Stay D, et al. Yin and yang: Balancing and answering binary visual questions[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016

[7] Bordes A, Boureau Y L, Weston J. Learning end-to-end goal-oriented dialog[J]. arXiv preprint arXiv:1605.07683, 2016.

[8] Amazon. Alexa. http://alexa.amazon.com/.

[9] Das A, Kottur S, Moura J M F, et al. Learning cooperative visual dialog agents with deep reinforcement learning[C]//Proceedings of the IEEE International Conference on Computer Vision. 2017: 2951-2960.

[10] Chattopadhyay P, Yadav D, Prabhu V, et al. Evaluating visual conversational agents via cooperative human-ai games[C]//Fifth AAAI Conference on Human Computation and Crowdsourcing. 2017.

[11] Kottur S, Moura J M F, Parikh D, et al. Visual coreference resolution in visual dialog using neural module networks[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018: 153-169.

[12] Massiceti D, Siddharth N, Dokania P K, et al. FlipDial: A generative model for two-way visual dialogue[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 6097-6105.

[13] Massiceti D, Dokania P K, Siddharth N, et al. Visual Dialogue without Vision or Dialogue[J]. arXiv preprint arXiv:1812.06417, 2018.

[14] Li J, Miller A H, Chopra S, et al. Dialogue learning with human-in-the-loop[J]. arXiv preprint arXiv:1611.09823, 2016.

附录(Appendix):

1 数据集格式

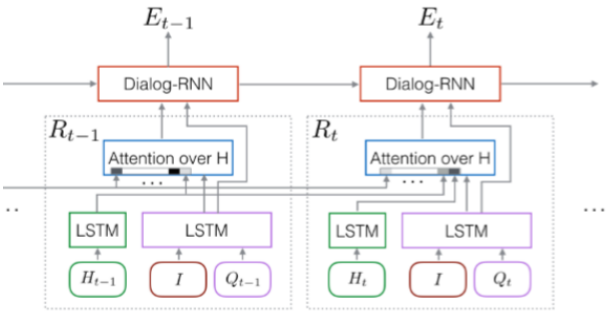
v1.0 版本：

```
{
  'data': {
    'questions': [
      'does it have a doorknob',
      'do you see a fence around the bear',
      ...
    ],
    'answers': [
      'no, there is just green field in foreground',
      'countryside house',
      ...
    ],
    'dialogs': [
      {
        'image_id': <image id>,
        'caption': <image caption>,
        'dialog': [
          {
            'question': <index of question in `data.questions` list>,
            'answer': <index of answer in `data.answers` list>,
            'answer_options': <100 candidate answer indices from `data.answers`>,
            'gt_index': <index of `answer` in `answer_options`>
          },
          ... (10 rounds of dialog)
        ]
      },
      ...
    ]
  },
```

v0.9 版本：

```
{
  'data': {
    'questions': [
      'does it have a doorknob',
      'do you see a fence around the bear',
      ...
    ],
    'answers': [
      'no, there is just green field in foreground',
      'countryside house',
      ...
    ],
    'dialogs': [
      {
        'image_id': <COCO image id>,
        'caption': <image caption from COCO>,
        'dialog': [
          {
    'question': <index of question in `data.questions` list>,
    'answer': <index of answer in `data.answers` list>,
    'answer_options': <100 candidate answer indices from `data.answers`>,
    'gt_index': <index of `answer` in `answer_options`>
          },
          ... (10 rounds of dialog)
        ]
      },
      ...
    ]
  },
  'split': <COCO split>,
  'version': '0.9'
}': <VisDial split>,
  'version': '1.0'
}
```

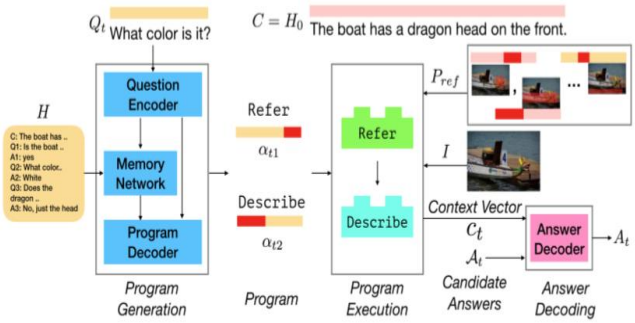
2 HRE 模型示意图



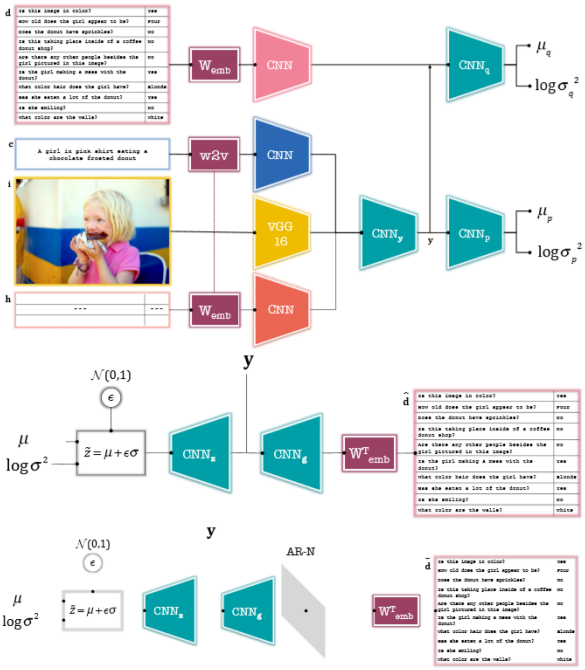
3 Visual Dialog 实验原文结果

	Model	MRR	R@1	R@5	R@10	Mean
Baseline	Answer prior	0.3735	23.55	48.52	53.23	26.50
	NN-Q	0.4570	35.93	54.07	60.26	18.93
	NN-QI	0.4274	33.13	50.83	58.69	19.62
Generative	LF-Q-G	0.5048	39.78	60.58	66.33	17.89
	LF-QH-G	0.5055	39.73	60.86	66.68	17.78
	LF-QI-G	0.5204	42.04	61.65	67.66	16.84
	LF-QIH-G	0.5199	41.83	61.78	67.59	17.07
	HRE-QH-G	0.5102	40.15	61.59	67.36	17.47
	HRE-QIH-G	0.5237	**42.29**	62.18	67.92	17.07
	HREA-QIH-G	0.5242	42.28	62.33	68.17	**16.79**
	MN-QH-G	0.5115	40.42	61.57	67.44	17.74
	MN-QIH-G	**0.5259**	**42.29**	**62.85**	**68.88**	17.06
Discriminative	LF-Q-D	0.5508	41.24	70.45	79.83	7.08
	LF-QH-D	0.5578	41.75	71.45	80.94	6.74
	LF-QI-D	0.5759	43.33	74.27	83.68	5.87
	LF-QIH-D	0.5807	43.82	74.68	84.07	5.78
	HRE-QH-D	0.5695	42.70	73.25	82.97	6.11
	HRE-QIH-D	0.5846	44.67	74.50	84.22	5.72
	HREA-QIH-D	0.5868	44.82	74.81	84.36	5.66
	MN-QH-D	0.5849	44.03	75.26	84.49	5.68
	MN-QIH-D	**0.5965**	**45.55**	**76.22**	**85.37**	**5.46**
VQA	SAN1-QI-D	0.5764	43.44	74.26	83.72	5.88
	HieCoAtt-QI-D	0.5788	43.51	74.49	83.96	5.84

4 Refer 模块和 Exclude 模块



5 FlipDial 模型示意图



6 Visual Coreference Resolution 实验原文结果

Model	MRR	R@1	R@5	R@10	Mean
MN-QIH-D [13]	0.597	45.55	76.22	85.37	5.46
HCIAE-D-MLE [30]	0.614	47.73	77.50	86.35	5.15
AMEM+SEQ-QI [41]	0.623	48.53	78.66	87.43	4.86
NMN[20]	0.616	48.24	77.54	86.75	4.98
CorefNMN\Mem	0.618	48.56	77.76	86.95	4.92
CorefNMN\$\mathcal{L}_C^{aux}$	**0.636**	**50.49**	79.56	88.30	4.60
CorefNMN\Mem\$\mathcal{L}_C^{aux}$	0.617	48.47	77.54	86.77	4.99
CorefNMN	**0.636**	50.24	**79.81**	**88.51**	**4.53**
CorefNMN (ResNet-152)	0.641	50.92	80.18	88.81	4.45

Table 3: Retrieval performance on the validation set of VisDial v0.9 [13] (discriminative models) using VGG [42] features (except last row). Higher the better for mean reciprocal rank (MRR) and recall@k (R@1, R@5, R@10), while lower the better for mean rank. Our CorefNMN model outperforms all other models across all metrics.