



CCF 学生领航计划

第二期：英文学术论文写作规范与日常积累

于静

中科院信息工程研究所

2022.06.29@CCF SPP

主要内容

- 1/ 学术研究^与论文写作（第一期）
- 2/ 学术论文之写作思路（第一期）
- 3/ 学术论文之图示绘制（第二期）
- 4/ 学术论文之英文规范（第二期）
- 5/ 学术研究之日常积累（第二期）
- 6/ 本讲总结与心得分享（第二期）

为何从“写作思路”说到“规范与积累”？

规范：英文怎么写规范、简洁、清晰？

(第二期 *How?*)

写法：怎么写？怎么改？怎么高效？

(第二期 *How?*)

思路：写哪些？什么逻辑？什么内容？

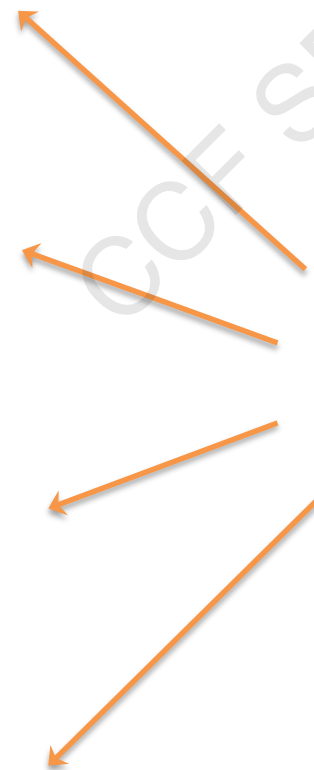
(第一期 *What?*)

价值观：为何做科研？为何选这个问题？

(第一期 *Why?*)

积累

(第二期 *How?*)



学术论文 之图示绘制



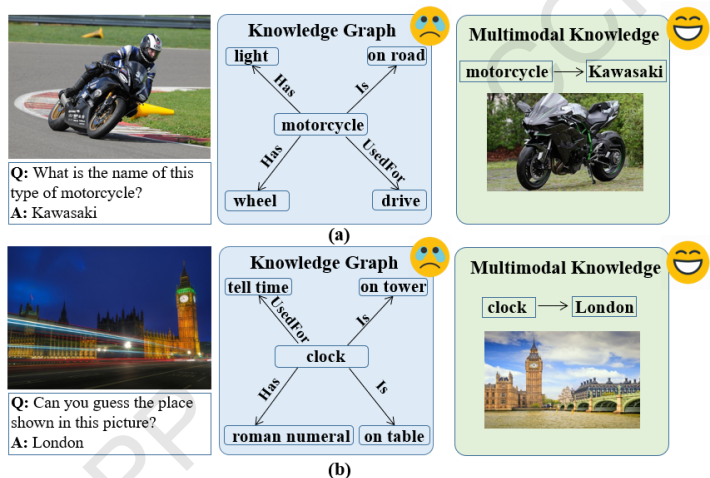
从论文图示与论文写作的关系说起

论文引言

科学问题 是否探索本质？为何重要？

写作逻辑 背景-现状-问题-方法如何环环相扣？

表达方式 如何抓住重点？如何深入浅出？



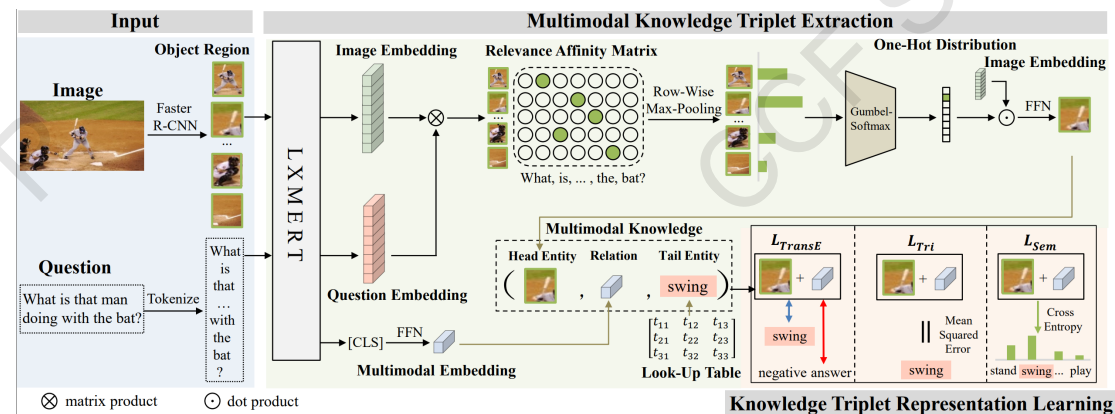
研究动机图

论文模型

解决方法 是否专注核心挑战？如何解决？

写作逻辑 方法如何分解？各自解决哪些问题？

表达方式 如何准确清晰？如何减少返工？



模型框架图

让我们一起来复盘



MuKEA: Multimodal Knowledge Extraction and Accumulation for Knowledge-based Visual Question Answering

CVPR 2022

Paper: <https://arxiv.org/abs/2203.09138>

Code: <https://github.com/AndersonStra/MuKEA>

Yang Ding , **Jing Yu*** , Bang Liu , Yue Hu, Mingxin Cui , Qi Wu



中国科学院 信息工程研究所
INSTITUTE OF INFORMATION ENGINEERING, CAS



中国科学院大学
University of Chinese Academy of Sciences

Université
de Montréal



THE UNIVERSITY
of ADELAIDE

直观的记录研究动机



Q: What is the name of this type of motorcycle?

A: Kawasaki

摩托车有本田，哈雷，雅马哈.....，语言很难说清每种摩托车的特点

解决的问题是？为什么重要？



这种摩托车是Kawasaki



这种摩托车是哈雷

多个例子如何展示？

箭头啥意思？如何表达知识？

一个图片怎么得到多条知识？

主要问题：缺乏核心问题说明！缺乏专业语言、符号表达！

从数据集里挑选样本，覆盖不同场景，进行对比，说明研究动机



Q: Is this a room for a boy or girl?

A: girl

textual knowledge

- (toy, at location, baby)
- (doll, is a, toy)
- (bed, is, in room)
- (man, is, on bed)
- (woman, is, on bed)

文本知识只有三元组?

visual knowledge

(room → girl)



视觉知识如何界定?

例子有无歧义?

罗列三元组太不直观!



Q: What is the name of this type of motorcycle?

A: kawasaki

textual knowledge

- (motorcycle, has, light)
- (motorcycle, has, wheel)
- (motorcycle, is, on road)
- (seat, is on, motorcycle)
- (motorcycle, used for, drive)

visual knowledge

(motorcycle → kawasaki)



左右不对齐! 浪费空间! 不美观!

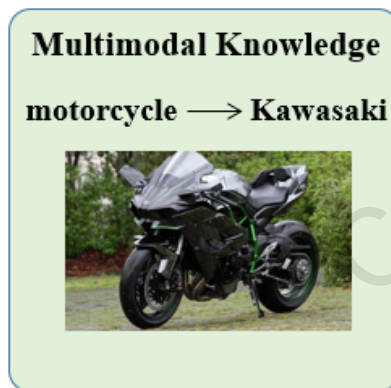
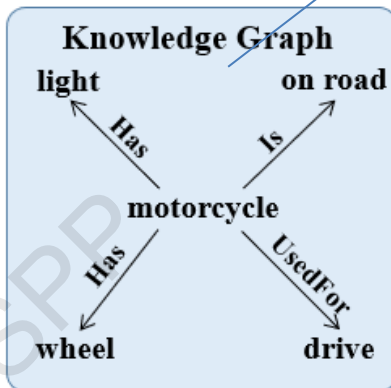
主要问题: 展示形式不直观! 所解决问题难理解!

改进知识图谱表现形式，直观对比问题；更换例子；降低歧义

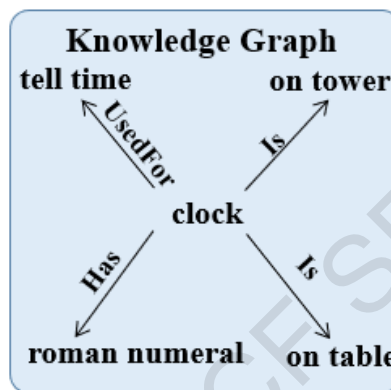
文字多，三元组实体、关系不突出



Q: What is the name of this type of motorcycle?
A: Kawasaki



Q: Can you guess the place shown in this picture?
A: London



只看动机图，不明确本文研究内容！

主要问题：表达的信息需要更充分、传递信息需要更高效！

最终版本!

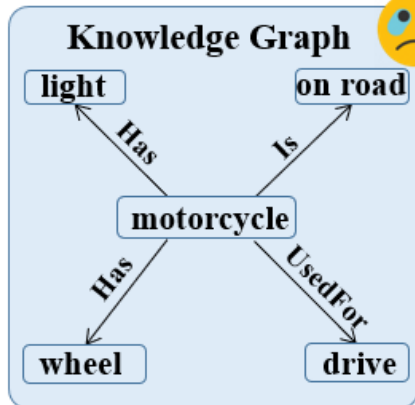
明确本文研究内容!

与正文问题提出对应!

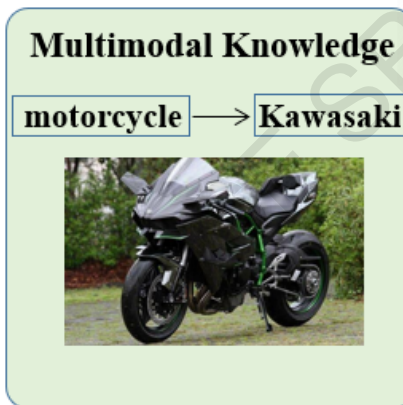
cate and multimodal knowledge, which is essential for human to tackle complex problems. Considering the question in Figure 1(a), the agent needs visual knowledge of motorcycle appearance in each brand to identify the given motorcycle, but the knowledge graph lacks of such instantiated information. Besides object understanding, implicit visual knowledge in mind mostly dominate over the rigid facts when humans are asked for simple scene discrimination like the question 'Can you guess the place?' in Figure 1(b). How to represent and accumulate the complex multimodal knowledge in the VQA scenario while maintaining the advantages of traditional knowledge graph in explainable reasoning is an essential but less studied problem.



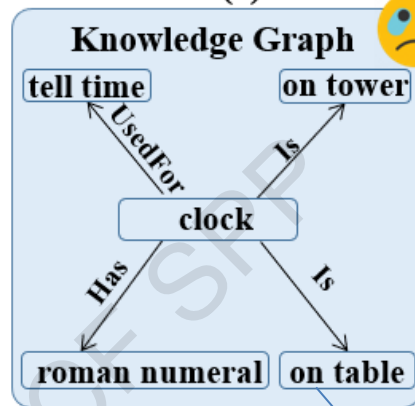
Q: What is the name of this type of motorcycle?
A: Kawasaki



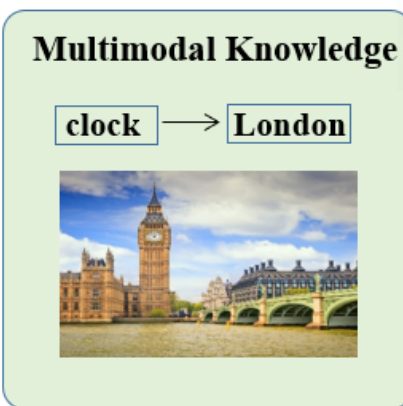
(a)



Q: Can you guess the place shown in this picture?
A: London

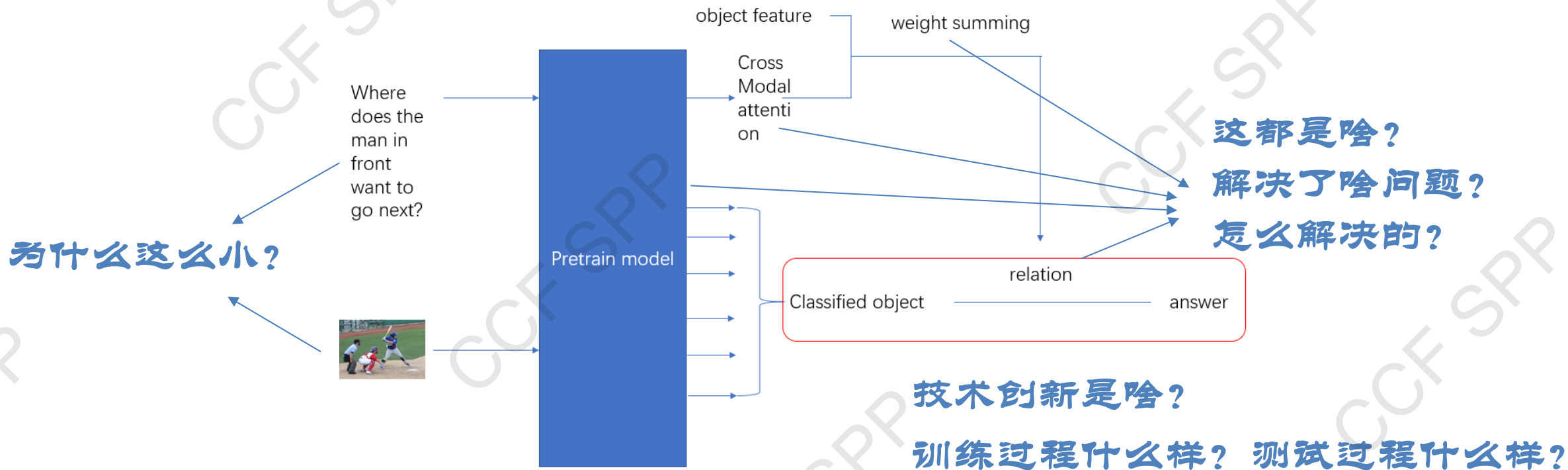


(b)



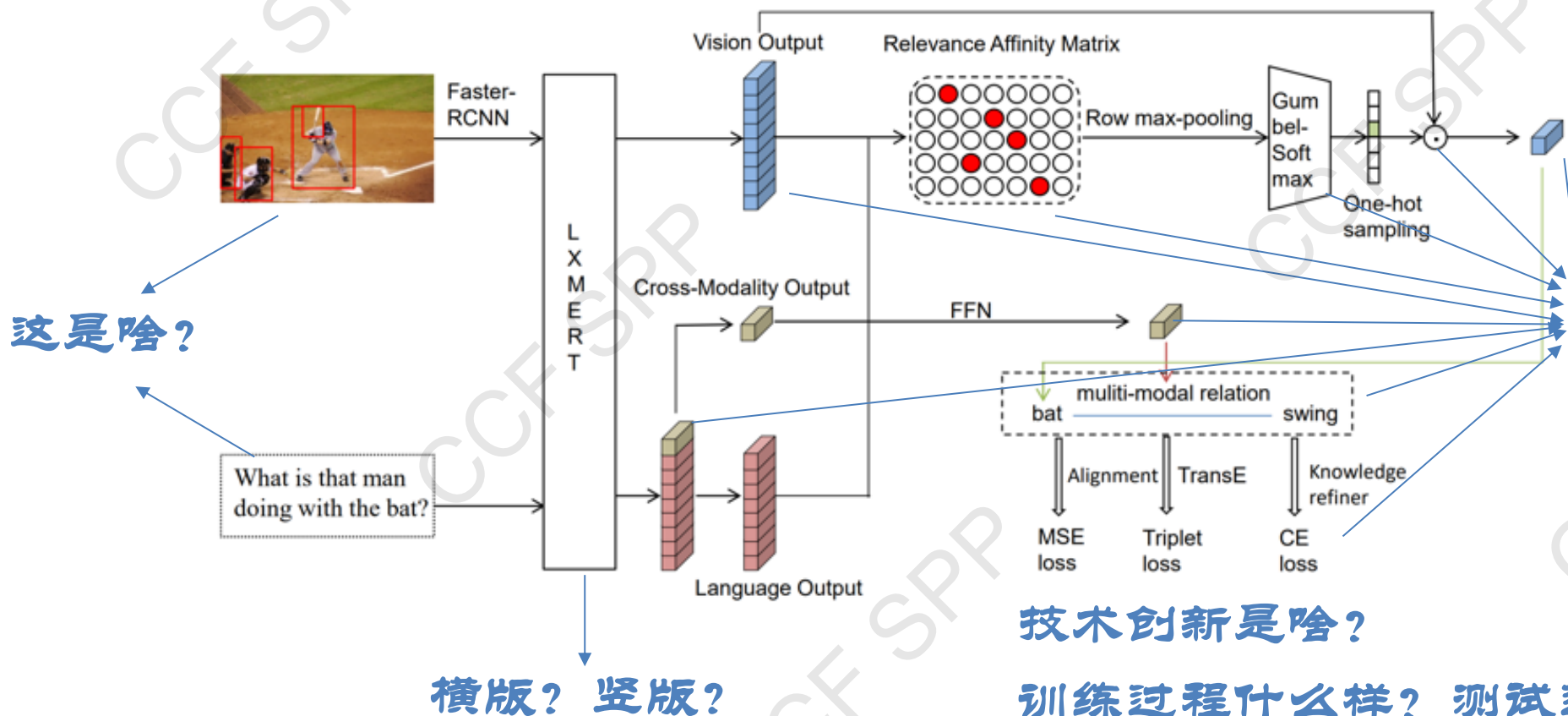
增加文字框，突出三元组实体、关系

最初版本：直观的记录模型输入、输出



主要问题：过程不明确，模块边界不清晰，如何解决没体现！

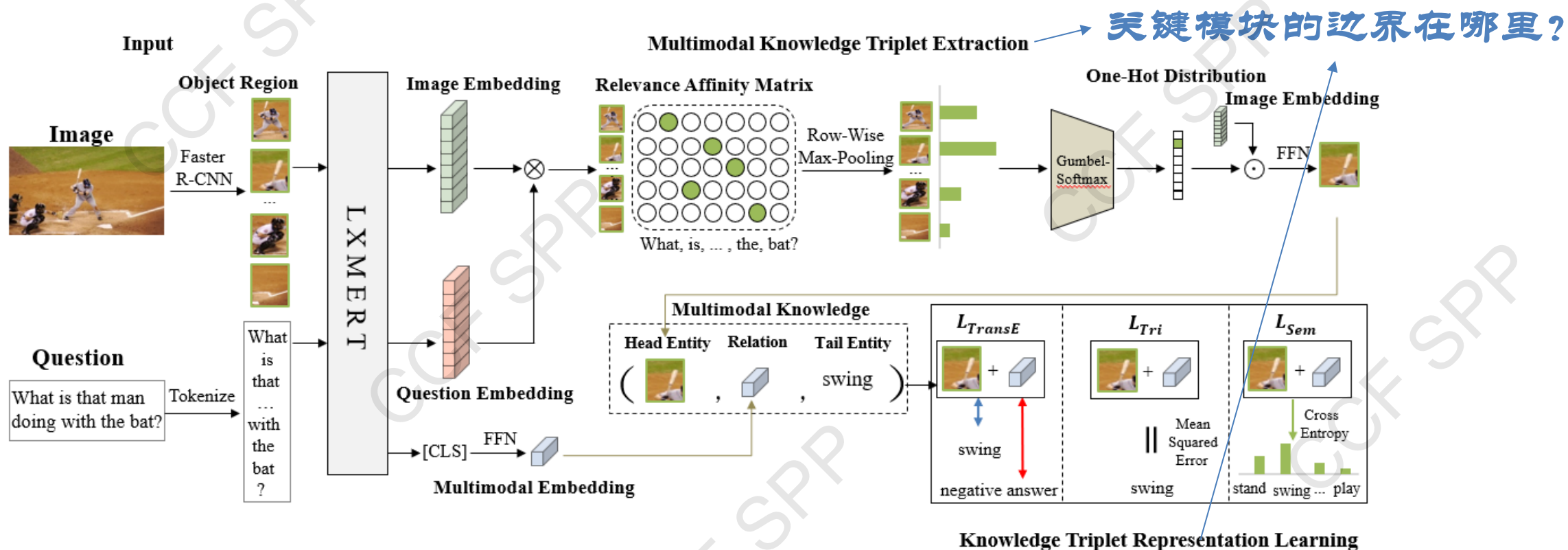
初步刻画了模型的完整实现过程



主要问题：模块边界不清晰，如何解决没体现！

模型框架图 V3

明确定义所有过程目标、变量名称，细化每个阶段实现，避免线条交叉

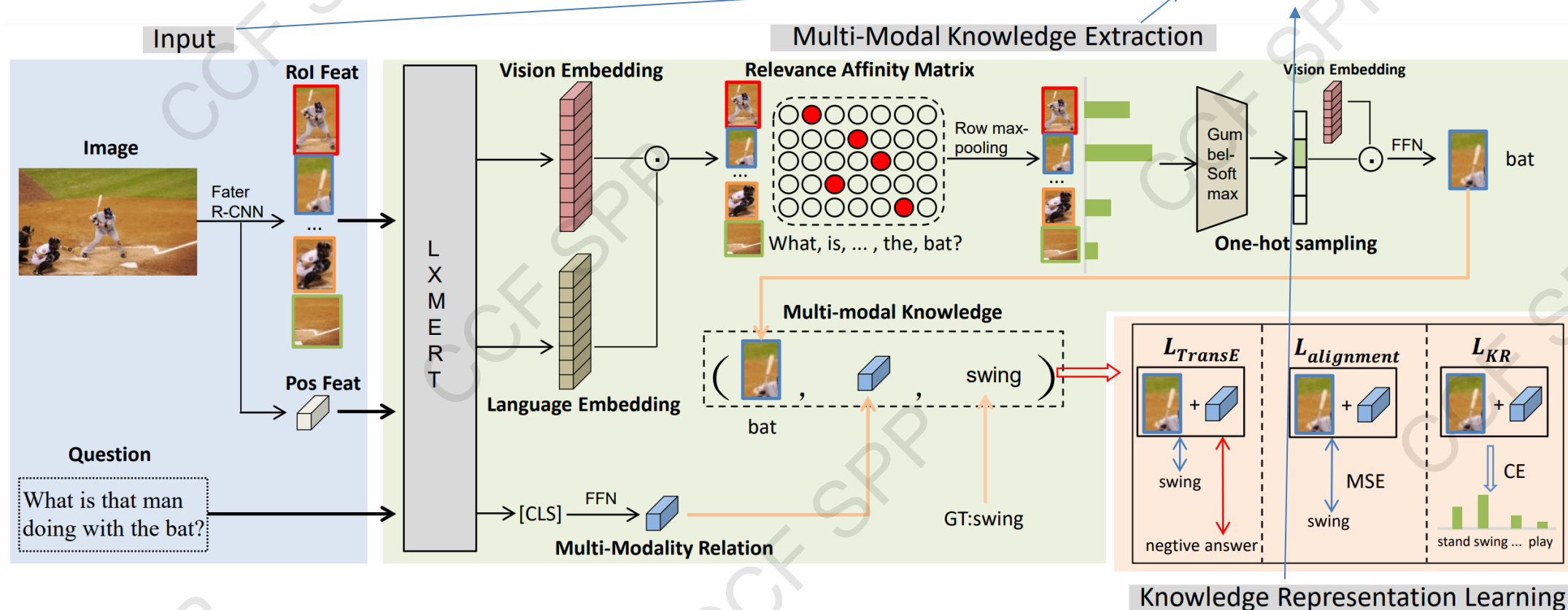


主要问题：模块边界不清晰，算法逻辑不明确！

模型框架图 V4

每个模块用不同颜色区分，模块标题突出

排版不整齐！标题不明显！



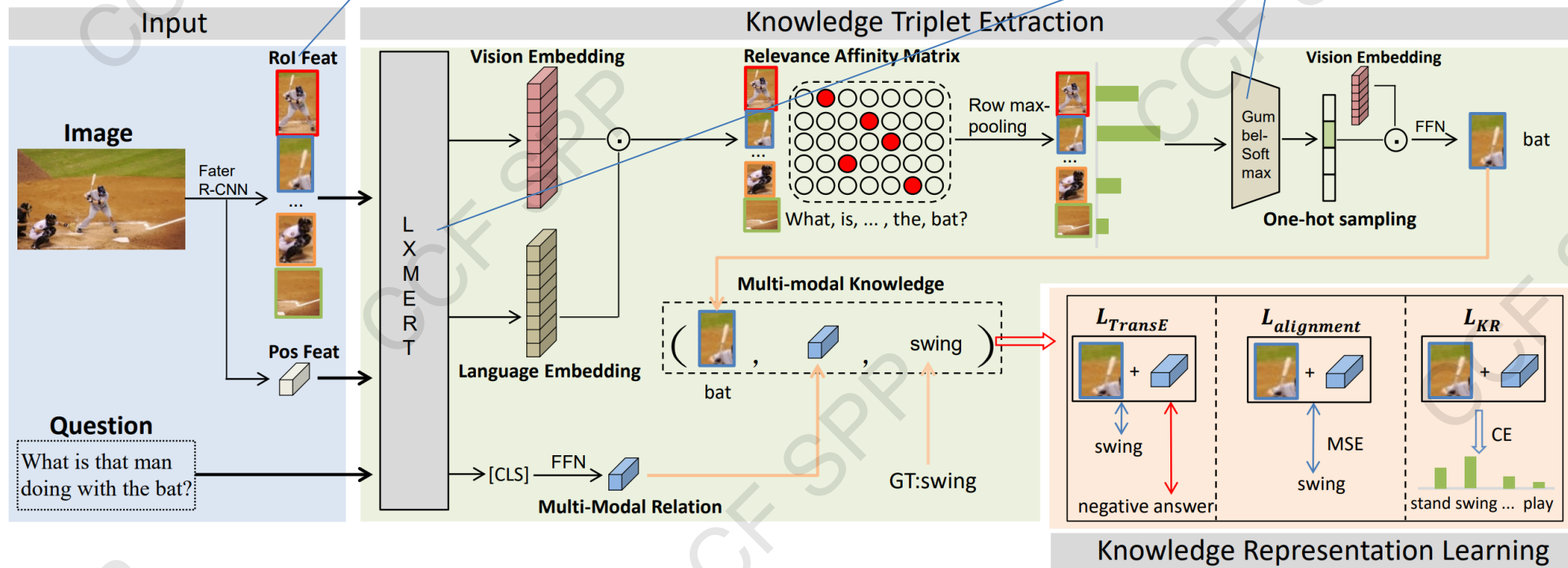
主要问题：不直观！布局问题！

模型框架图 V5

标题整体布局合理、一致

字体不一致

排版不规范



主要问题：看着乱！字体、排版问题！

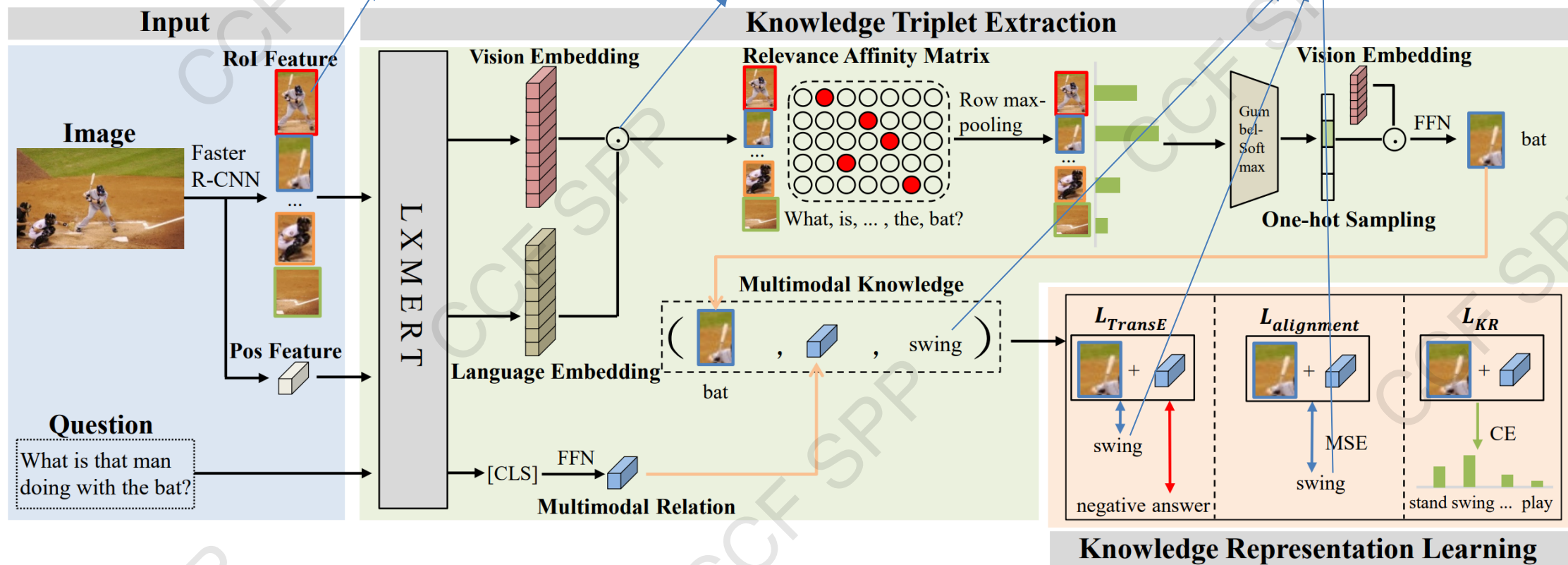
模型框架图 V6

字体、字号一致，排版符合常识

图片大小不统一

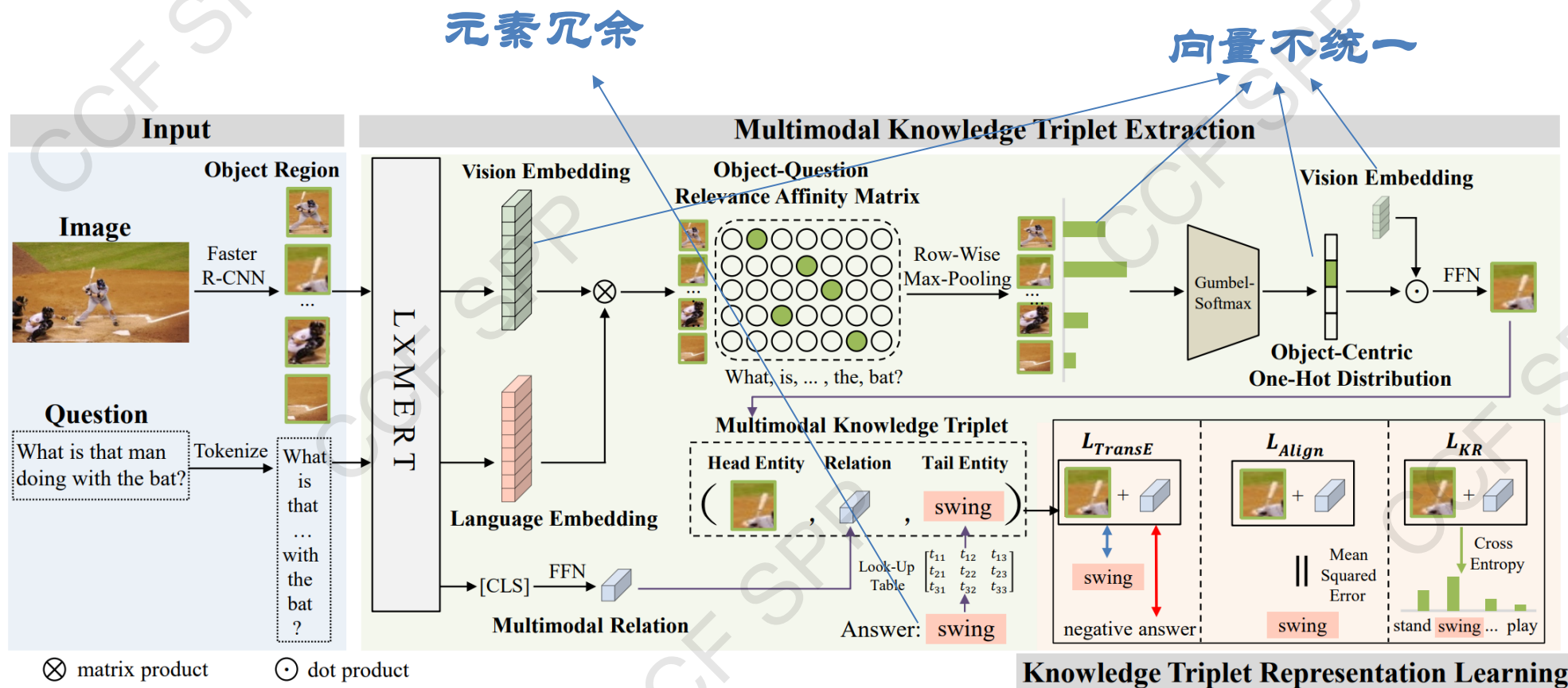
重要符号没有注释

同样的符号背景不一致



主要问题：颜色乱！文字多！线条多！不直观！

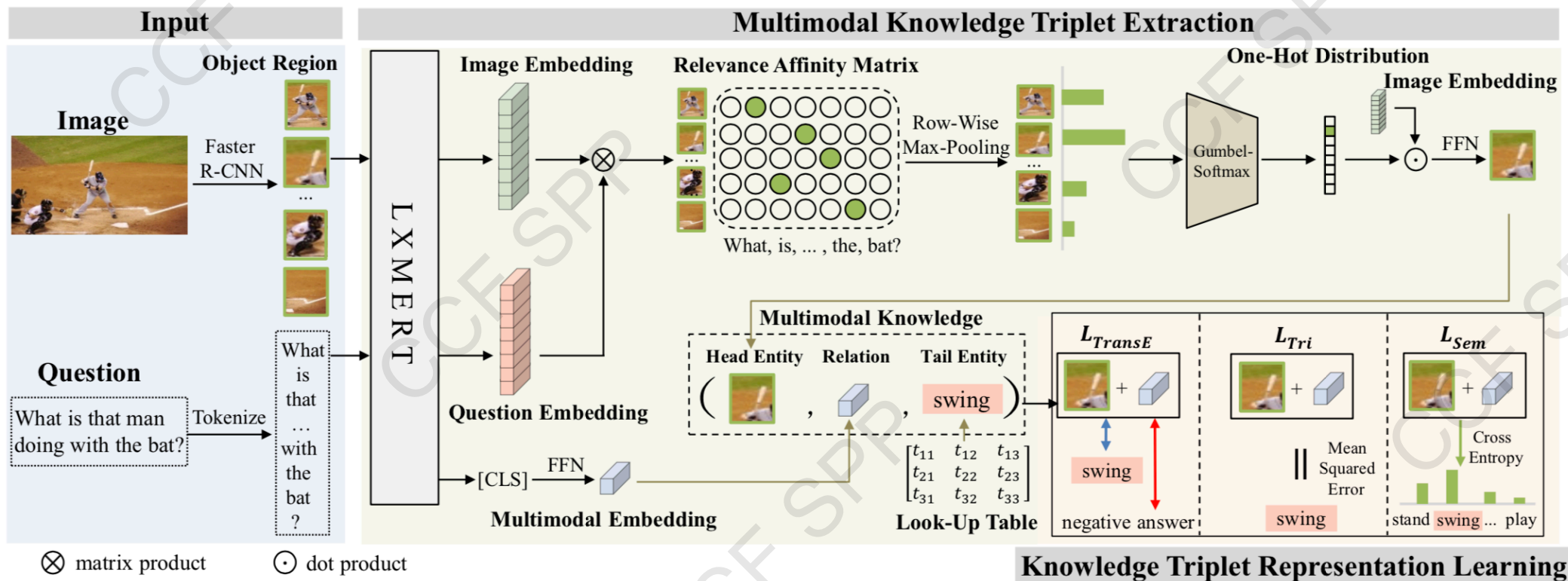
相同内容表现形式一致，符号有注释



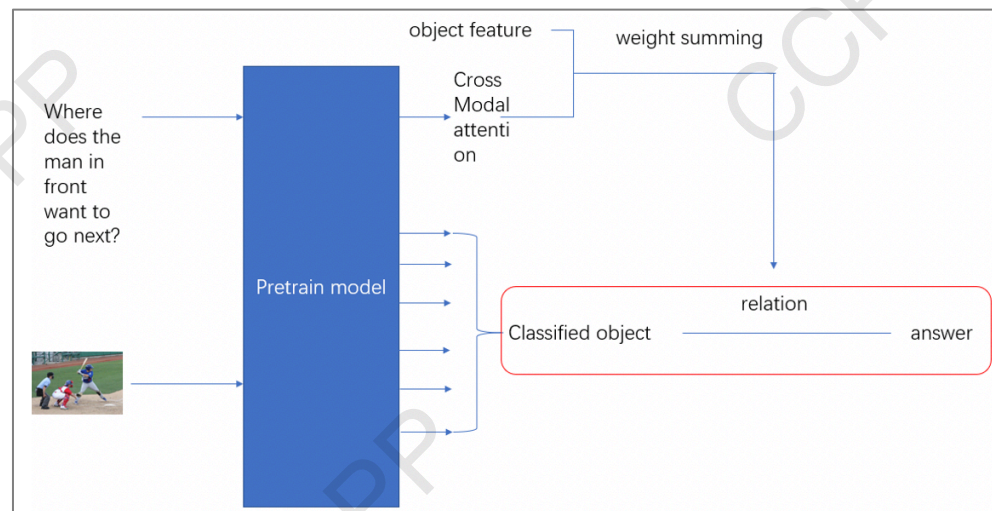
主要问题：相同内容前后不对应！信息冗余！

模型框架图 V8

最终版本!

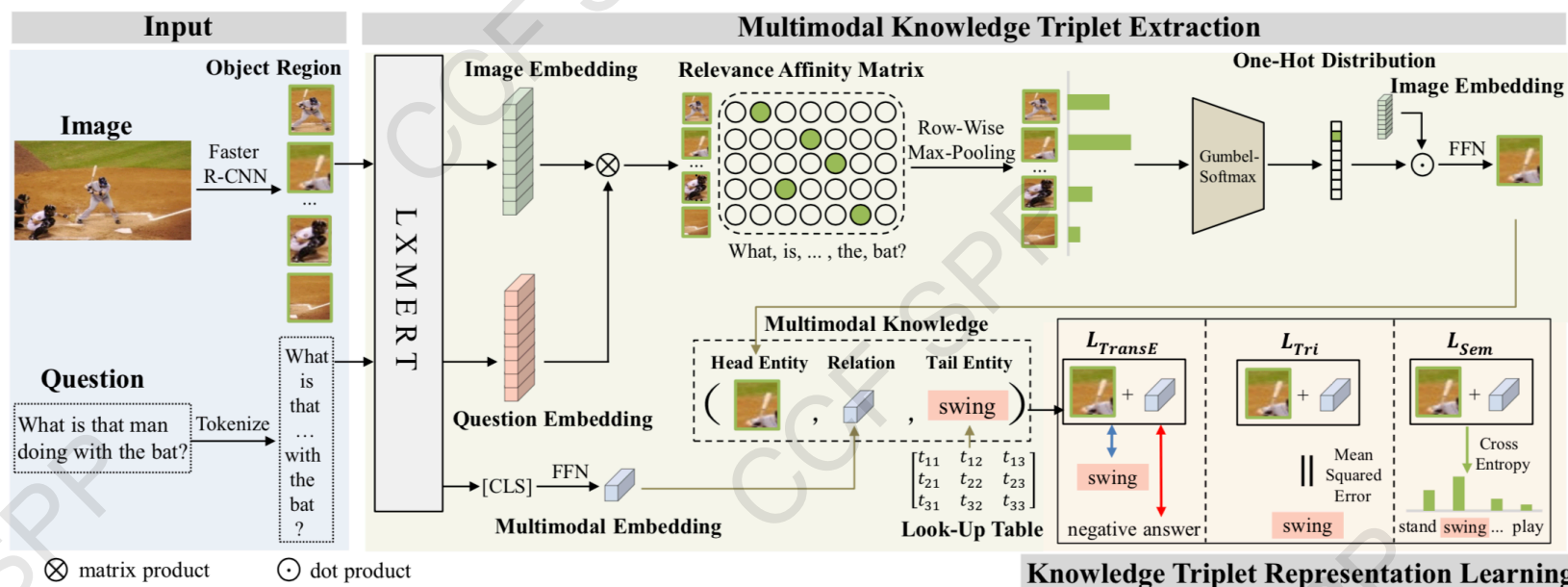


模型框架图 从0到1



10月
最初版本!

20+ 版本



11月
最终版本!

ET-BERT: A Contextualized Datagram Representation with Pre-training Transformers for Encrypted Traffic Classification

WWW 2022

Paper: <https://dl.acm.org/doi/pdf/10.1145/3485447.3512217>

Code: <https://github.com/linwhitehat/ET-BERT>

Xinjie Lin, Gang Xiong, Gaopeng Gou, Zhen Li, Junzheng Shi, **Jing Yu***

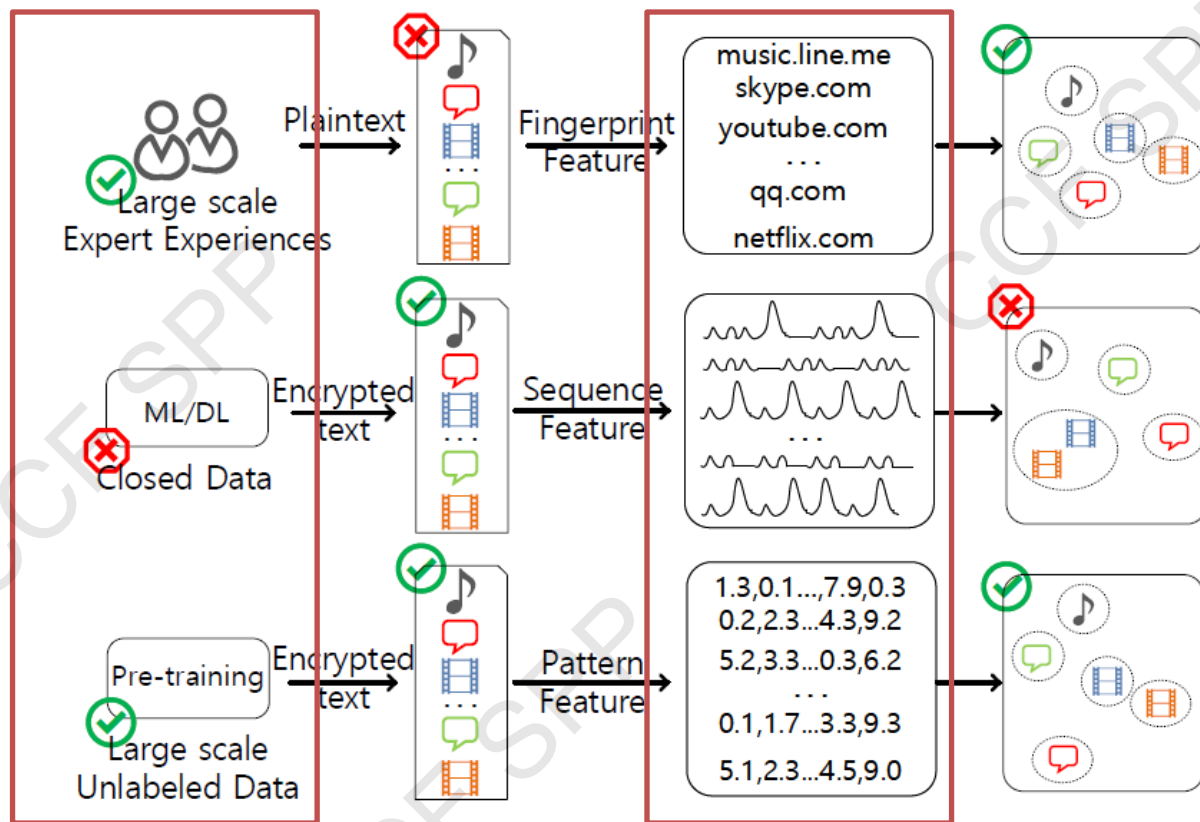


中国科学院 信息工程研究所
INSTITUTE OF INFORMATION ENGINEERING, CAS



中国科学院大学
University of Chinese Academy of Sciences

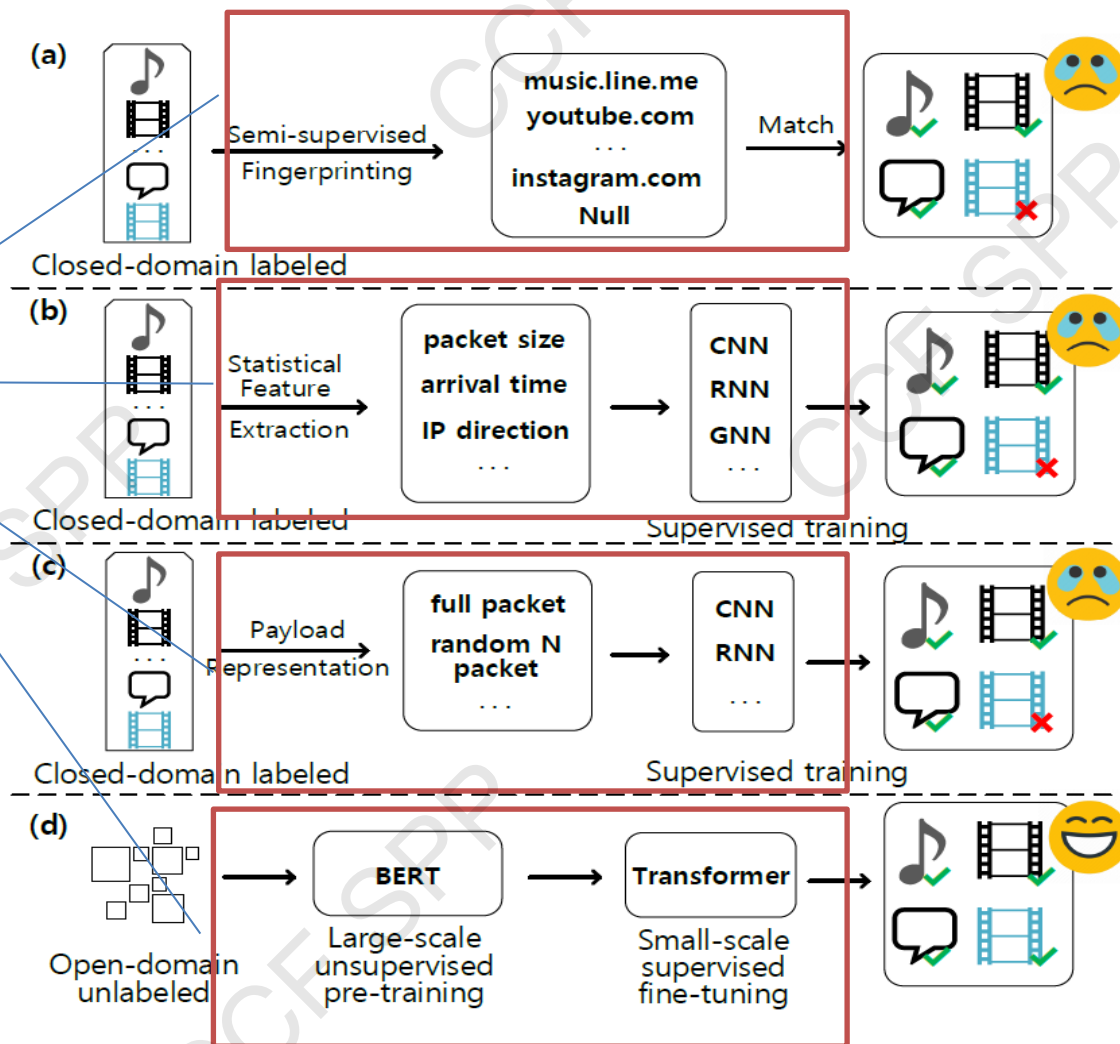
方法差异性不明确



技术发展脉络不清晰

主要问题：不同框架刻画不准确，研究动机不明确！

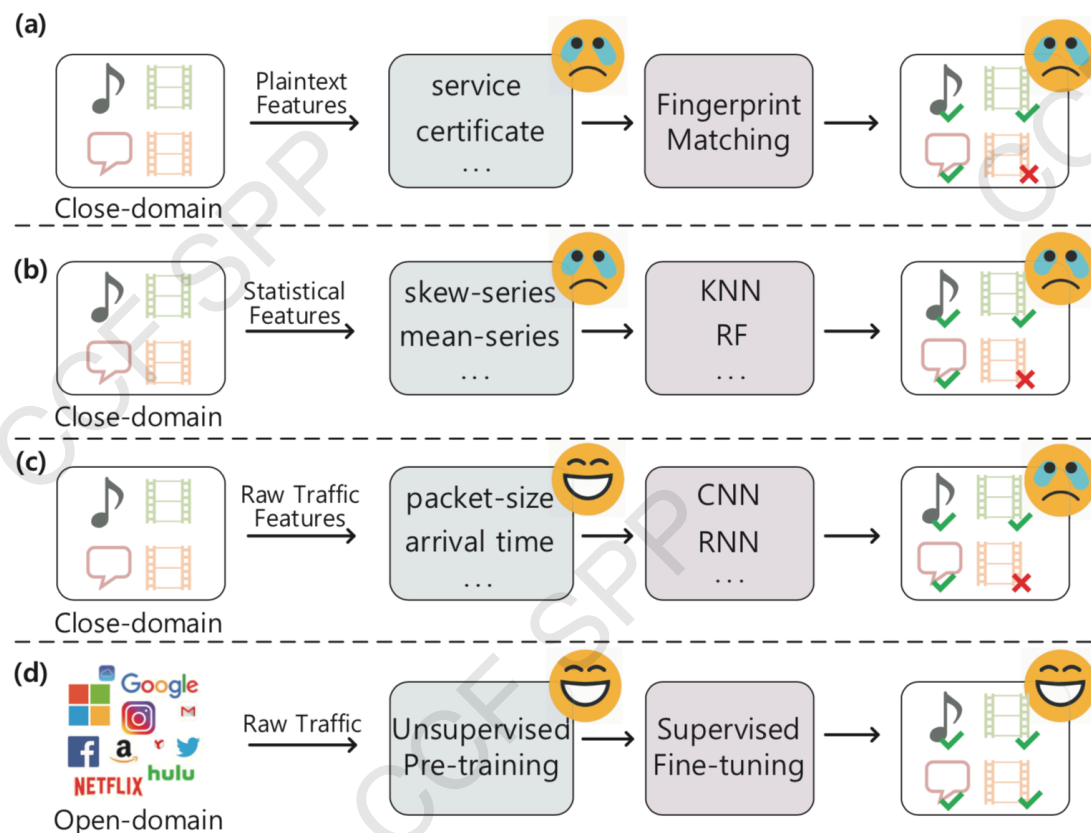
不同框架差异模糊



主要问题：不同框架能力边界刻画不清晰，创新性难反应！

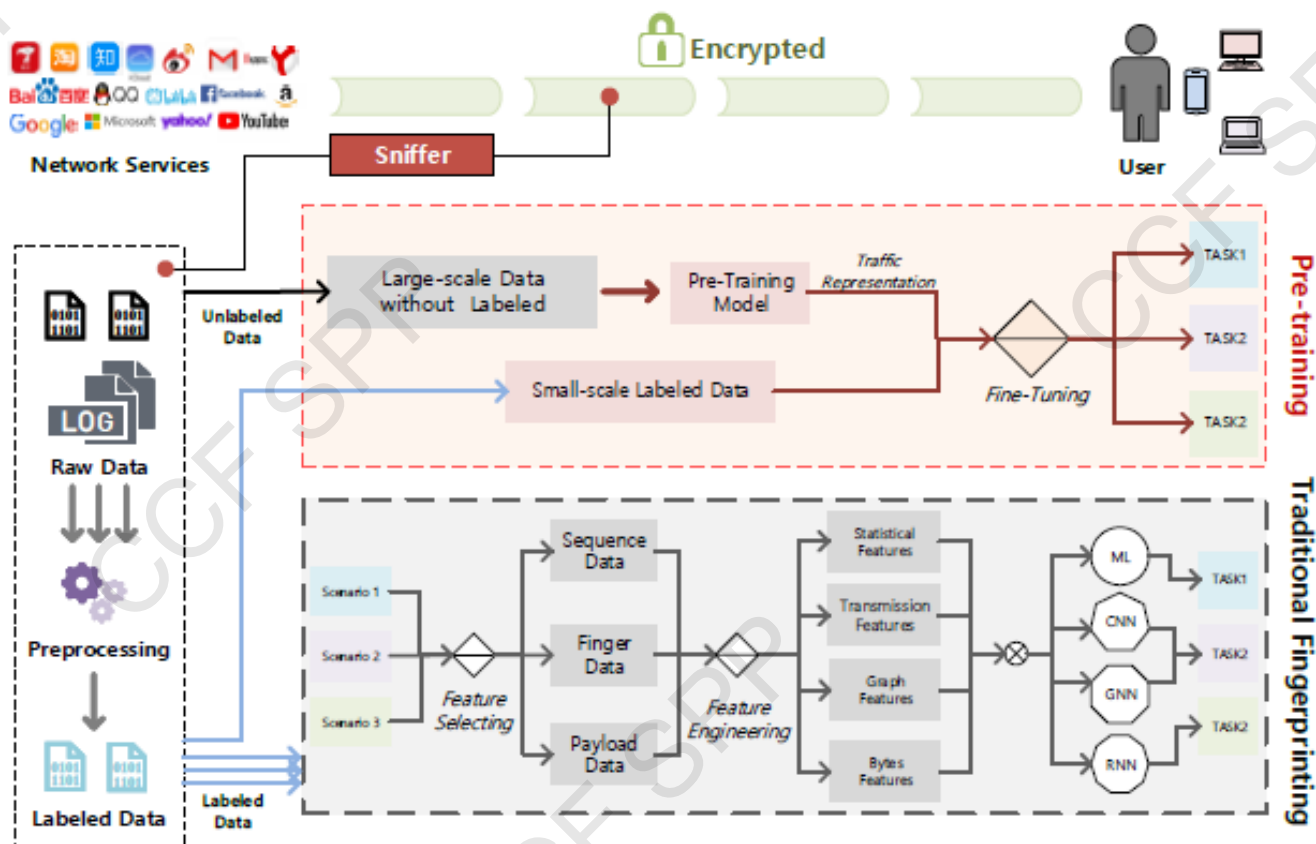
最终版本!

统一模块化，凸出不同框架在同一处理环节的差异



模型框架图 V1

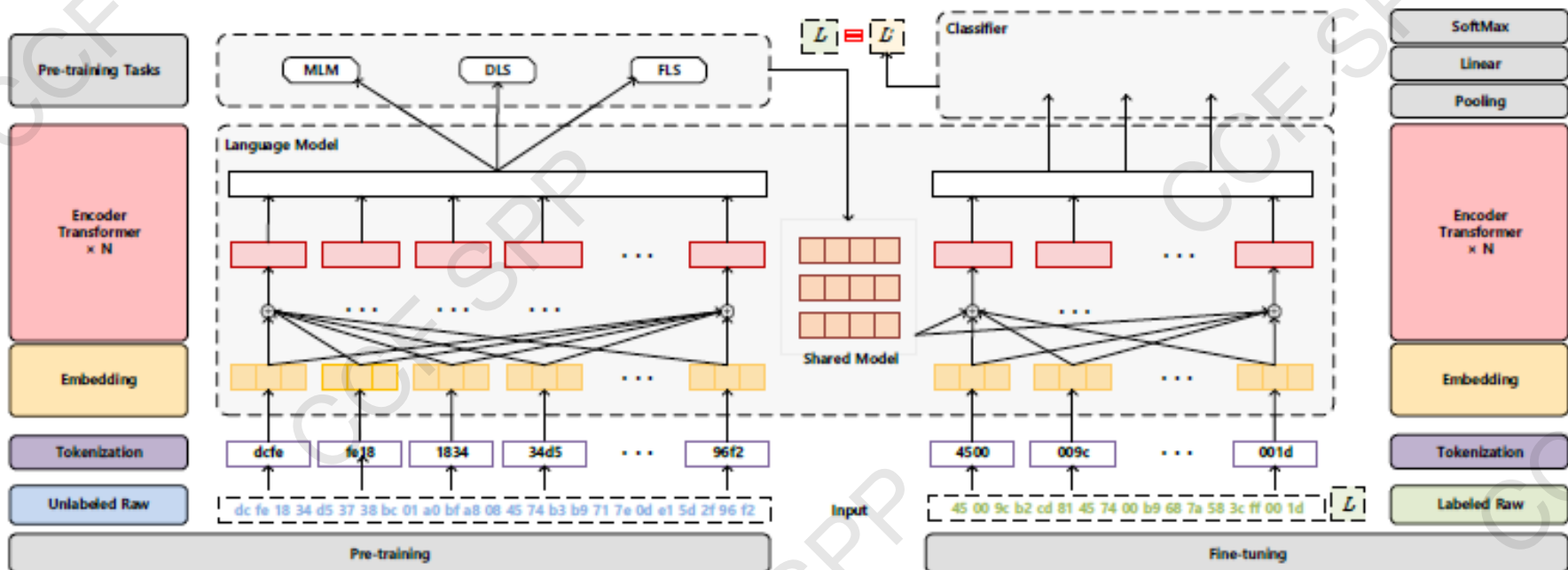
最初版本



主要问题：太多细节，大量数据预处理，甚至方法对比，毫无重点！

模型框架图 V2

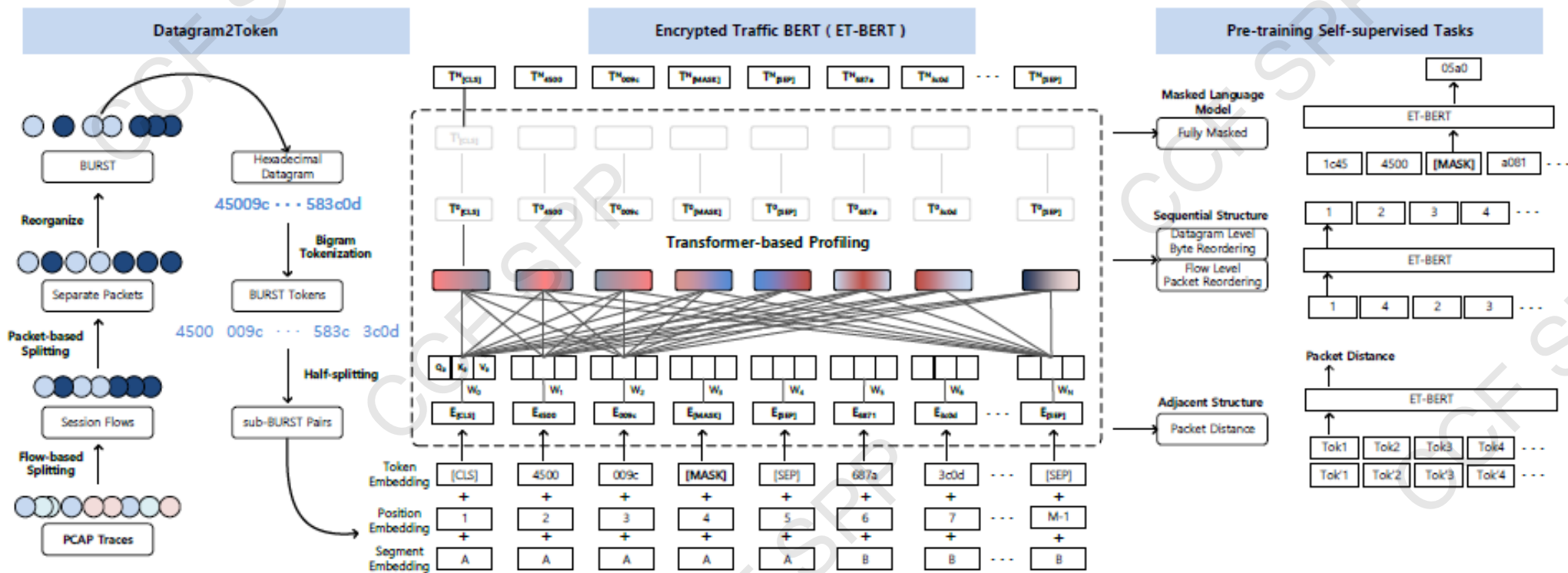
明确输入、输出、主要过程



主要问题：模块冗余，毫无说明，过程逻辑不清晰！

模型框架图 V3

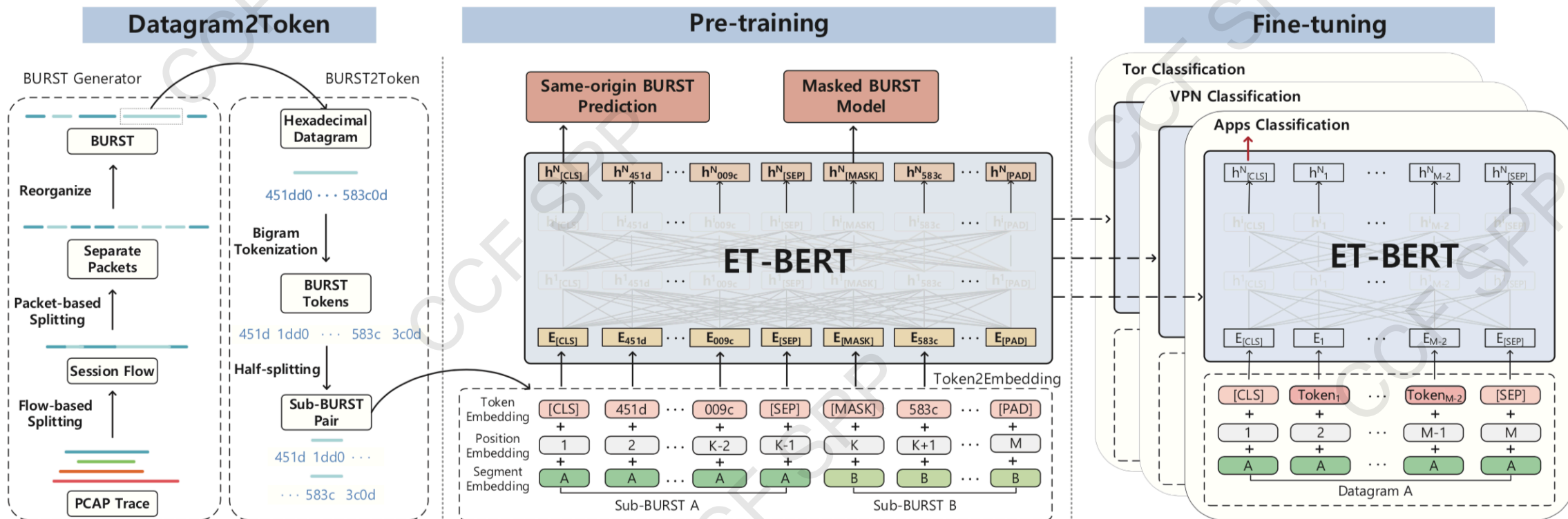
划分主要模块，明确模块功能



主要问题：技术创新不明确，训练、微调过程不清晰！

模型框架图 V4

20+ times later... 最终版本!



论文示意图——小结

研究动机图

- ☀ 明确所解决的问题
- ☀ 通过鲜明对比展示
- ☀ 例子准确覆盖全面

模型框架图

- ☀ 先明确输入、输出、关键过程
- ☀ 划分模块，确定模块解决的问题
- ☀ 精准语言描述每个过程、变量、符号
- ☀ 相同语义内容表现形式一致，避免冗余

学术论文 之英文规范



英语写作规范——八项注意

1. 精简的表达方式

2. 严谨的叙述逻辑

不太容易

3. 专业的学术用语

4. 规范的符号使用

5. 标准的学术术语

很容易

6. 客观的图表绘制

7. 正确的文献引用

8. 坚守的学术道德

英语写作规范（1）——精简的表达方式



论文只是一个载体，是为了向同行们宣告你的科研发现，是科学领域交流的重要工具。所以，在科研论文写作时，一定要谨记于心的就是：**用最简单的话表达最明白的意思！**

——施一公

- 表达：一句话3个从句
- 问题：混淆方法和问题
- 无重点！有语病！

- 表达：一句话44个单词
- 问题：加入太多细节
- 无重点！有语病！

Motivated by the above idea, we propose a novel framework for classifying encrypted traffic in this paper, encrypted datagram representation by pre-training (ET-BERT), for learning generic features in large-scale unlabeled encrypted traffic(Figure 1(c)). We define a special structure, BURST, to depict a traffic flow and represent it with the bi-gram language model for highlighting the structural features of the traffic transmission. To learn application-specific generic features, the proposed framework consists of a two-part model: pre-training and fine-tuning. Specifically, the high-dimensional representations of dependencies that bridges the gap between different datagram bytes are provided by pre-training in large-scale unlabeled encrypted traffic, and the generic representations of specific encryption scenarios are supported in fine-tuning with small labeled traffic through reusing the pre-training results.

More insightfully, we propose a relation-aware pre-training model for encrypted traffic to adaptively capture a generic representation of traffic in multiple encryption scenarios. It can be demonstrated by two procedures. First, the Masked Burst Model(MBM) procedure captures the correlation between different payload bytes from unmasked contexts. Then, the Same-origin BURST Prediction(SBP) procedure captures same-origin evidence between different classes of sub-BURST pairs.

最初版本

In this paper, we propose a novel pre-training model for classifying encrypted traffic, called Encrypted Traffic Bidirectional Encoder Representations from Transformer (ET-BERT). It aims to learn generic traffic representations from large-scale unlabeled encrypted traffic (Figure 1(d)). We first propose a raw traffic representation model to transform the datagram to language-like tokens for pre-training. Each traffic flow is presented by a transmission-guided structure, denoted as BURST, which serves as the input. The proposed framework consists of two stages: pre-training and fine-tuning. Specifically, the pre-training network with Transformer structure obtains datagram-level generic traffic representations by self-supervised learning on large-scale unlabeled encrypted traffic. Thereinto, we propose two novel pre-training tasks to learn the traffic-specific patterns: the Masked BURST Model (MBM) task captures the correlated relationship between different datagram bytes in the same BURST and represent them by their context; the Same-origin BURST Prediction (SBP) task models the transmission relationships of preceding and subsequent BURST. Then, ET-BERT incorporates with the specific classification task and fine-tune the parameters with small number of task-specific labeled data.

最终版本

- 一句话命名方法及问题
- 一句话突出方法目标
- 一句话介绍方法创新

- 一句话表达核心方法思路

Tips: (1) 一句话只表达一个意思！ (2) 减少中文翻译英文！ (3) 避免重复表达！

英语写作规范（2）——严谨的叙述逻辑

假设审稿人还有一个小时审稿 *deadline*，刚打开你的论文，读到…

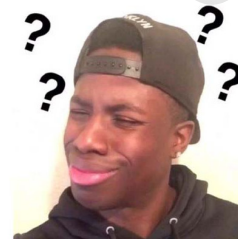
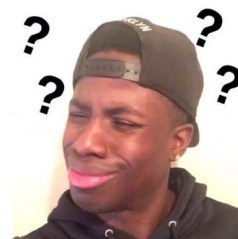
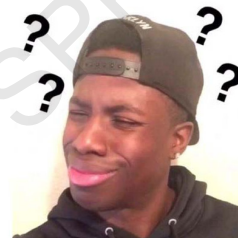
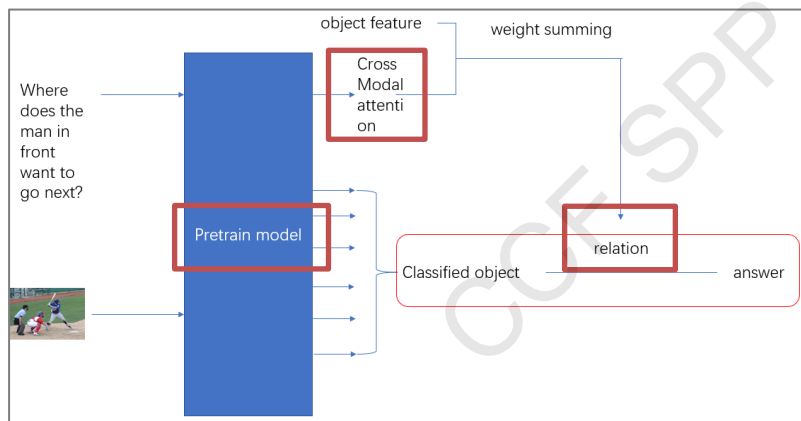
这么一句话

Now, we could use **CoTNet model** to predict the result of that dataset.

这么一个公式

$$\alpha_i = \text{softmax}(\boxed{w_a^T} \tanh(\boxed{W_1 v_i} + \boxed{W_2 q}))$$

这么一个图



英语写作规范（2）——严谨的叙述逻辑

Tips: (1) 在术语使用前定义解释!

3. Our Approach

In this section, we first provide a brief review of the conventional self-attention widely adopted in vision backbones. Next, a novel Transformer-style building block, named Contextual Transformer (CoT), is introduced for image representation learning. This design goes beyond conventional self-attention mechanism by additionally exploiting the contextual information among input keys to facilitate self-attention learning, and finally improves the representational properties of deep networks. After replacing 3×3 convolutions with CoT block across the whole deep architecture, two kinds of Contextual Transformer Networks, i.e., CoTNet and CoTNeXt deriving from ResNet [22] and ResNeXt [53], respectively, are further elaborated.



英语写作规范（2）——严谨的叙述逻辑

Tips: （2）给出公式后集中对定义及公式中的符号进行解释

$$\alpha_i = \text{softmax}(w_a^T \tanh(\mathbf{W}_1 \mathbf{v}_i + \mathbf{W}_2 \mathbf{q})) \quad (1)$$

where $\mathbf{W}_1, \mathbf{W}_2$ and w_a (as well as $\mathbf{W}_3, \dots, \mathbf{W}_{12}$, w_b , w_c mentioned below) are learned parameters. $\mathbf{q}$ is question embedding encoded by the last hidden state of LSTM.

$$\beta_{ji} = \text{softmax}(w_b^T \tanh(\mathbf{W}_3 \mathbf{v}'_j + \mathbf{W}_4 \mathbf{q}')) \quad (2)$$

where $\mathbf{v}'_j = \mathbf{W}_5[\mathbf{v}_j, \mathbf{r}_{ji}]$, $\mathbf{q}' = \mathbf{W}_6[\mathbf{v}_i, \mathbf{q}]$ and $[\cdot, \cdot]$ denotes concatenation operation.

$$m_j^{(t+1)} = \mathbf{W}_{11}[m_j^{(t)}, c_j^{nei}, h^{(t)}] \quad (11)$$

$$c_j^{nei} = \sum_{k \in \mathcal{N}_j} \mathbf{W}_{12}[m_k^{(t)}, r_{jk}] \quad (12)$$

where $\mathcal{N}_i$ represents a set of 1-hop neighboring nodes regarding the memory entity m_j and c_j^{nei} is the contextual memory representation. Then the updated memory is served as the new knowledge memory used in the next reasoning step.

英语写作规范 (2) —— 严谨的叙述逻辑

Tips: (3) 有清晰的段落结构, 段落/章节之间有过度!

Abstract

Fact-based Visual Question Answering (FVQA) requires external knowledge beyond visible content to answer questions about an image, which is challenging but indispensable to achieve general VQA. One limitation of existing FVQA solutions is that they jointly embed all kinds of information without fine-grained selection, which introduces unexpected noises for reasoning the final answer. How to capture the question-oriented and information-complementary evidence remains a key challenge to solve the problem. In this paper, we depict an image by a multi-modal heterogeneous graph, which contains multiple layers of information corresponding to the visual, semantic and factual features. On top of the multi-layer graph representations, we propose a modality-aware heterogeneous graph convolutional network to capture evidence from different layers that is most relevant to the given question. Specifically, the intra-modal graph convolution selects evidence from each modality and cross-modal graph convolution aggregates relevant information across different modalities. By stacking this process multiple times, our model performs iterative reasoning and predicts the optimal answer by analyzing all question-oriented evidence. We achieve a new state-of-the-art performance on the FVQA task and demonstrate the effectiveness and interpretability of our model with extensive experiments. The code is available at <https://github.com/astro-zihao/mucko>.

1 Introduction

Visual question answering (VQA) [Antol et al., 2015] is an attractive research direction aiming to jointly analyze multi-modal content from images and natural language. Equipped with the capacities of grounding, reasoning and translating, a VQA agent is expected to answer a question in natural language based on an image. Recent works [Cadene et al., 2019;

*Equal contribution.

†Corresponding author.

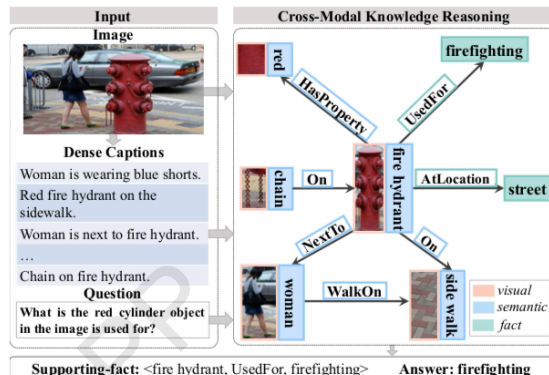


Figure 1: An illustration of our motivation. We represent an image by multi-layer graphs and cross-modal knowledge reasoning is conducted on the graphs to infer the optimal answer.

Li et al., 2019b; Ben-Younes et al., 2019] have achieved great success in the VQA problems that are answerable by solely referring to the visible content of the image. However, such kinds of models are incapable of answering questions which require external knowledge beyond what is in the image. Considering the question in Figure 1, the agent not only needs to visually localize ‘the red cylinder’, but also to semantically recognize it as ‘fire hydrant’ and connects the knowledge that ‘fire hydrant is used for firefighting’. Therefore,

• 为了缓解上述问题, ***提出

To advocate research in this direction, [Wang et al., 2018] introduces the ‘Fact-based’ VQA (FVQA) task for answering questions by joint analysis of the image and the knowledge base of facts. The typical solutions for FVQA build a fact graph with fact triplets filtered by the visual concepts in the image and select one entity in the graph as the answer. Existing works [Wang et al., 2017; Wang et al., 2018] parse the question as keywords and retrieve the supporting-entity only by keyword matching. This kind of approaches is vulnerable when the question does not exactly mention the visual concepts (e.g. synonyms and homographs) or the mentioned information is not captured in the fact graph (e.g. the visual

attribute ‘red’ in Figure 1 may be falsely omitted). To resolve these problems, [Narasimhan et al., 2018] introduces visual information into the fact graph and infers the answer by implicit graph reasoning under the guidance of the question. However, they provide the whole visual information equally to each graph node by concatenation of the image, question and entity embeddings. Actually, only part of the visual content are relevant to the question and a certain entity. Moreover, the fact graph here is still homogeneous since each node is represented by a fixed form of image-question-entity embedding, which limits the model’s flexibility of adaptively capturing evidence from different modalities.

In this work, we depict an image as a multi-modal heterogeneous graph, which contains multiple layers of information corresponding to different modalities. The proposed model is focused on **Multi-Layer Cross-Modal Knowledge Reasoning** and we name it as **Mucko** for short. Specifically, we encode an image by three layers of graphs, where the object appearance and their relationships are kept in the *visual layer*, the high-level abstraction for bridging the gaps between visual and factual information is provided in the *semantic layer*, and the corresponding knowledge of facts are supported in the *fact layer*. We propose a modality-aware heterogeneous graph convolutional network to adaptively collect complementary evidence in the multi-layer graphs. It can be performed by two procedures. First, the Intra-Modal Knowledge Selection procedure collects question-oriented information from each graph layer under the guidance of question; Then, the Cross-Modal Knowledge Reasoning procedure captures complementary evidence across different layers.

The main contributions of this paper are summarized as follows: (1) We comprehensively depict an image by a heterogeneous graph containing multiple layers of information based on visual, semantic and knowledge modalities. We consider these three modalities jointly and achieve significant improvement over state-of-the-art solutions. (2) We propose a modality-aware heterogeneous graph convolutional network to capture question-oriented evidence from different modalities. Especially, we leverage an attention operation in each convolution layer to select the most relevant evidence for the given question, and the convolution operation is responsible for adaptive feature aggregation. (3) We demonstrate good interpretability of our approach and provide case study in deep insights. Our model automatically tells which modality (visual, semantic or factual) and entity have more contributions to answer the question through visualization of attention weights and gate values.

• 然而, 他们方法只...
• 但事实上, 需要...

• 本工作中, 我们提出...

• 具体来说, 我们首先...

• 在上面构建的表征基础上, 我们进行...

• 然后, 实现...

英语写作规范（2）——严谨的叙述逻辑

Tips: (4) 图表文字清晰直接表达内容，与图注、表注、正文一致！

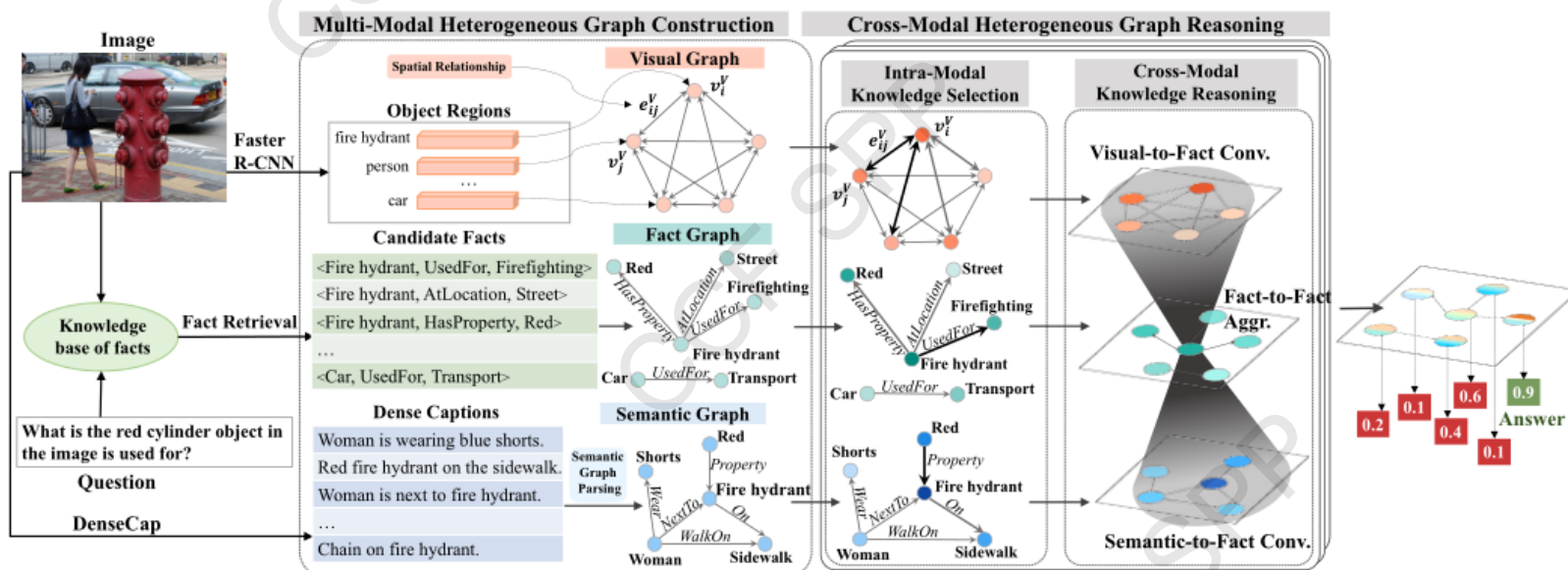


Figure 2: An overview of our model. The model contains two modules: Multi-modal Heterogeneous Graph Construction aims to depict an image by multiple layers of graphs and Cross-modal Heterogeneous Graph Reasoning supports intra-modal and cross-modal evidence selection.

Method	Overall Accuracy	
	top-1	top-3
LSTM-Question+Image+Pre-VQA	24.98	40.40
Hie-Question+Image+Pre-VQA	43.14	59.44
FVQA (top-3-QQmapping)	56.91	64.65
FVQA (Ensemble)	58.76	-
Straight to the Facts (STTF)	62.20	75.60
Reading Comprehension	62.96	70.08
Out of the Box (OB)	69.35	80.25
Human	77.99	-
Mucko	73.06	85.94

Table 1: State-of-the-art comparison on FVQA dataset.

Method		Overall Accuracy	
		top-1	top-3
Mucko (full model)		73.06	85.94
1	w/o Semantic Graph	71.28	82.76
2	w/o Visual Graph	69.12	78.05
3	w/o Semantic Graph & Visual Graph	20.43	29.10
4	S-to-F Concat.	67.82	76.65
5	V-to-F Concat.	69.93	80.12
6	V-to-F Concat. & S-to-F Concat.	70.68	82.04
7	w/o relationships	72.10	83.75

Table 2: Ablation study of key components of Mucko.

英语写作规范 (3) ——专业的学术用语

你的审稿人可能是...



当他打开你的论文，读到...

- This is **no doubt the first** and **the robust** model to achieve **the best** performance in VQA.
- **As we know**, the channel gain varies much more slower than the channel phase, **and we** have thousands of ways to prove it.
- His work is **good enough** to provide essential multimodal information for reasoning.

太绝对!

太口语!

太主观!

英语写作规范 (3) ——专业的学术用语

Tips: 阅读英文论文写作规范，告别懒惰！最容易提升！

导师“防秃”指南！

The Most Common Habits from more than 200 English Papers written by Graduate Chinese Engineering Students

By Felicia Britzman

This paper presents some of the most common Chinese-English habits observed from over two hundred English technical papers by Chinese writers. The habits are explained and in most cases, example text from an actual paper is given along with preferred text. An attempt is made to explain how to correct and prevent such mistakes. In some cases a possible explanation of why the habit occurs is also given. This paper can serve as an individual guide to editing technical papers especially when a native English-speaking editor is unavailable.

Introduction

Most Chinese universities require their doctoral and master candidates in technical and scientific fields to publish at least one English paper in an international journal as a degree requirement. However, many factors make this task difficult to accomplish. First, previous English studies may not have focused enough on writing, let alone technical writing. Current studies may not include English, causing the writers English fluency level to decline. Second, most writers have never lived in an English-speaking country. Third, due to the special aspects of technical writing, even native English-speaking engineering students have a technical writing course as part of their study.

☀ *a, an, the*

☀ *which/that*

☀ *such as/etc.*

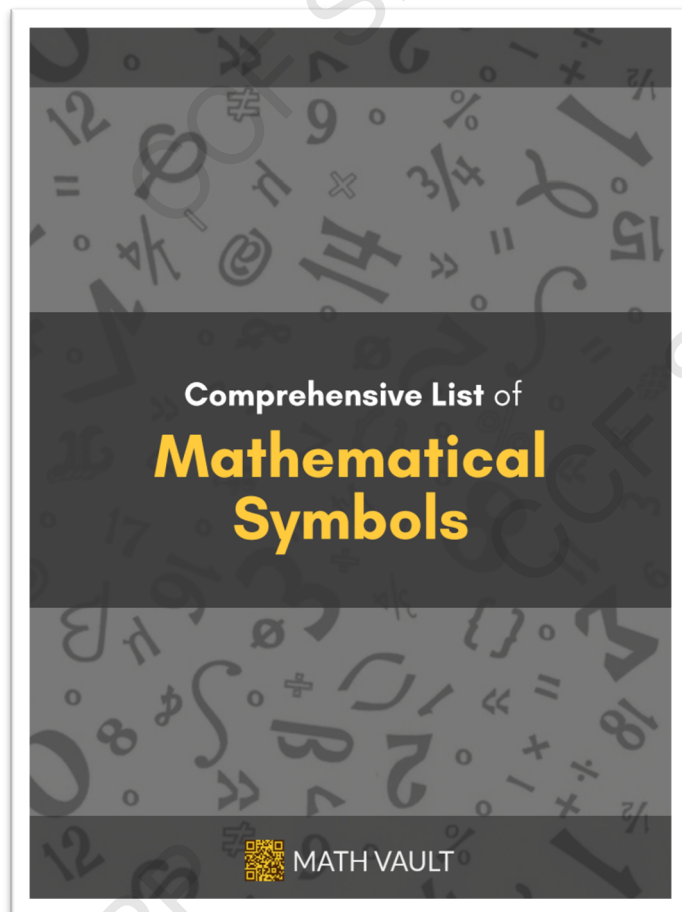
☀ *numbers and equations*

☀ ...

<http://papertalks.org/p/resources/Academic/EnglishWritingSkills/MostCommonEnglishWritingHabitsOfChinese.pdf>

英语写作规范（4）——规范的符号使用

Tips: （1）阅读英文论文符号公式定义规范，别偷懒😊 （2）学习本领域经典论文！



Comprehensive List of Mathematical Symbols

For the corresponding web guides, see [Mathematical Symbols](#).

Table of Contents

1	Constant	3
1.1	Key Mathematical Numbers	3
1.2	Key Mathematical Sets	4
1.3	Key Mathematical Infinities	5
1.4	Other Key Mathematical Objects	6
2	Variables	6
2.1	Variables for Numbers	6
2.2	Variables in Geometry	7
2.3	Variables in Calculus	7
2.4	Variables in Linear Algebra	8
2.5	Variables in Set Theory and Logic	8
2.6	Variables in Probability and Statistics	9
3	Delimiters	10
3.1	Common Delimiters	10
3.2	Other Delimiters	10
4	Operators	11

4.2	Function-related Operators	12
4.3	Elementary Functions	12
4.4	Algebra-related Operators	13
4.5	Geometry-related Operators	14
4.6	Logic-related Operators	15
4.7	Set-related Operators	16
4.8	Vector-related Operators	16
4.9	Matrix-related Operators	17
4.10	Probability-related Operators	18
4.11	Statistics-related Operators	18
4.12	Key Probability Functions and Distributions	19
4.13	Calculus-related Operators	20
5	Relational Symbols	21
5.1	Equality-based Relational Symbols	21
5.2	Comparison-based Relational Symbols	21
5.3	Number-related Relational Symbols	22
5.4	Geometry-related Relational Symbols	22
5.5	Set-related Relational Symbols	22
5.6	Logic-related Relational Symbols	23
5.7	Probability-related Relational Symbols	23
5.8	Calculus-related Relational Symbols	24
6	Notational Symbols	24
6.1	Common Notational Symbols	24
6.2	Notational Symbols in Geometry and Trigonometry	25
6.3	Notational Symbols in Calculus	25
6.4	Notational Symbols in Probability and Statistics	26
7	Additional Resources	26

Math Vault, <https://mathvault.ca/wp-content/uploads/Comprehensive-List-of-Mathematical-Symbols.pdf>

英语写作规范（5）——标准的学术术语

人工智能技术发展迅速，学术术语内涵和外延不断变化！

学术论文需要精准把握学术术语边界（共识）！

CCF计算机术语工委及 CCFpedia

术语工委及术语平台介绍：

计算机术语审定委员会(Committee on Terminology)主要职能是为收集、翻译、释义、审定和推荐计算机新词，并在CCF平台上宣传推广。这对厘清学科体系，开展科学研究，并将科学和知识在全社会广泛传播，都具有十分重要的意义。

术语众包平台CCFpedia的建设和持续优化，可以有效推进中国计算机术语的收集、审定、规范和传播工作，同时又能起到各领域规范化标准定制的推广作用。

新版的CCFpedia计算机术语平台(<http://term.ccf.org.cn>)将术语的编辑运营与浏览使用进行了整合，摒弃老版中跨平台操作的繁琐步骤，在界面可观性上进行了升级，让用户能够简单方便地查阅术语信息。同时，新版平台中引入知识图谱的方式对所有术语数据进行组织，通过图谱多层关联的形式升级了术语浏览的应用形态。



术语内涵及外延介绍

开篇导语：

本期发布术语热词：多模态表示学习 (Multimodal Representation Learning)。在大数据环境和新基建背景下，人机交互过程中数据对象的日益丰富和变化，呈现多模态的特点。其中，“多模态”可以直观地理解为数据不同的多媒体形式，也可以作为一个更加细粒度的概念。多模态表示学习旨在通过利用不同多模态数据之间的互补性，剔除模态冗余性。从而将多模态数据语义表征为实值向量...

多模态表示学习 (Multimodal Representation Learning)

作者：王萌（东南大学）

InfoBox：

中文名：多模态表示学习

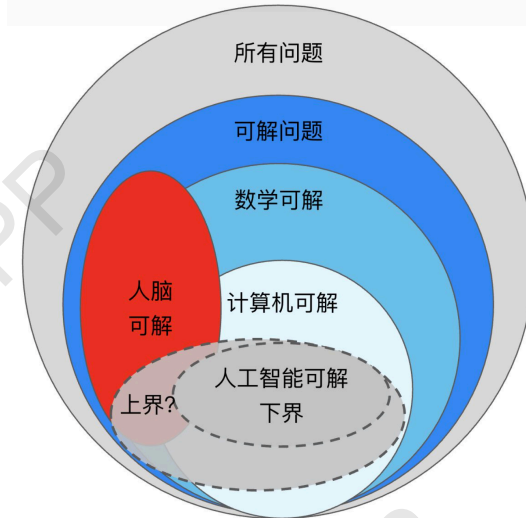
外文名：Multimodal Representation Learning

学科：机器学习

实质：表示学习

简介：

在大数据环境和新基建背景下，人机交互过程中数据对象的日益丰富和变化，呈现多模态的特点。其中，“多模态”可以直观地理解为数据不同的多媒体形式，也可以作为一个更加细粒度的概念，关键点在于为数据是否具有异构性，例如，对于某个歌手，互联网上可以找到他的照片和歌曲视频，同时也有相关的文本信息（百科，新闻等）以及具体的歌曲音频。这四种数据代表了图片、视频、文本、语音，可以被理解为该



参考文献/经典工作

参考文献：

- [1] Ngiam J, Khosla A, Kim M, Nam J, Lee H, Ng A Y. Multimodal deep learning [C]. Proceedings of the 28th International Conference on Machine Learning, 2011: 689-696.
- [2] Wang D, Cui P, Ou M, Zhu W. Learning compact hash codes for multimodal representations using orthogonal deep structure [J]. IEEE Transactions on Multimedia, 2015, 17(9): 1404-1416.
- [3] Peng Y-x, Zhu W-w, Zhao Y, Xu C-s, Huang Q-m, Lu H-q, Zheng Q-h, Huang T-j, Gao W. Cross-media analysis and reasoning: advances and directions [J]. Frontiers of Information Technology & Electronic Engineering, 2017, 18(1): 44-57.
- [4] Wu Z, Zheng C, Cai Y, Chen J, Leung H-f, Li Q. Multimodal Representation with Embedded Visual Guiding Objects for Named Entity Recognition in Social Media Posts [C]. Proceedings of the 28th ACM International Conference on Multimedia, 2020: 1038-1046.
- [5] Mouselly-Sergieh H, Botschen T, Gurevych I, Roth S. A multimodal translation-based approach for knowledge graph representation learning [C]. Proceedings of the Seventh Joint Conference on Lexical and Computational Semantics, 2018: 225-234.
- [6] Xie R, Liu Z, Luan H, Sun M. Image-embodied knowledge representation learning [C]. Proceedings of the 26th International Joint Conference on Artificial Intelligence, 2017: 3140-3146.
- [7] Chen L, Li Z, Wang Y, Xu T, Wang Z, Chen E. MMEA: Entity Alignment for Multi-modal Knowledge Graph [C]. Proceedings of the International Conference on Knowledge Science, Engineering and Management, Springer, 2020: 134-147.
- [8] Guo H, Tang J, Zeng W, Zhao X, Liu L. Multi-modal Entity Alignment in Hyperbolic Space [J]. Neurocomputing, 2021, 461(1): 598-607.

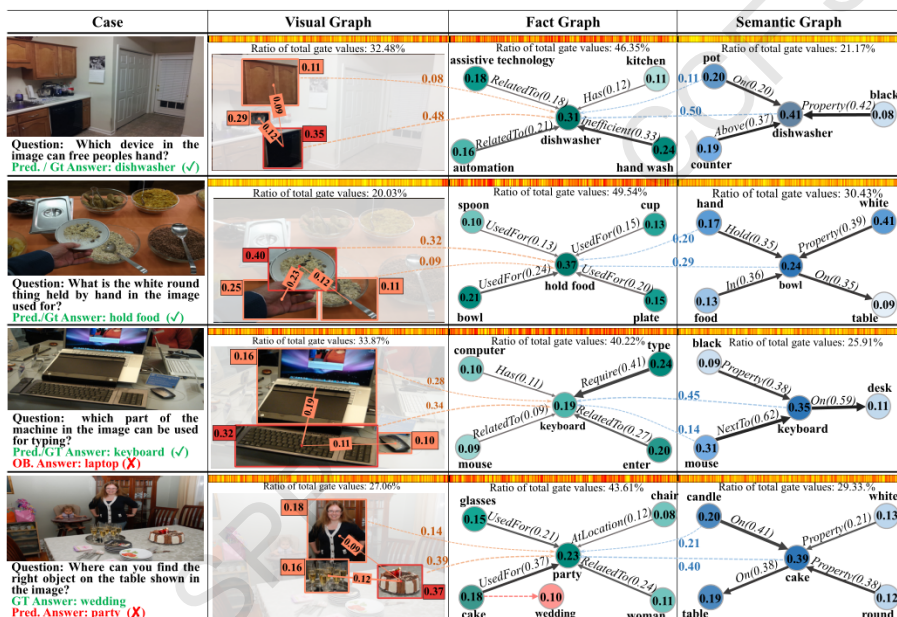
英语写作规范（6）——客观的图表绘制



是否真实、准确地展示和反映实验结果？

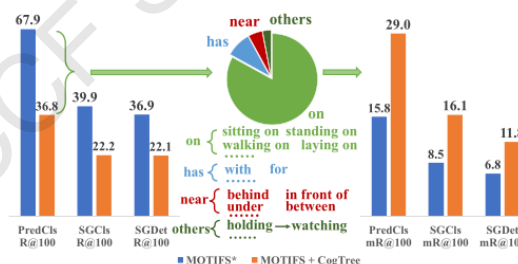
- ◆ 只展示自己方法的 *best case*? 其他方法的 *worst case*?
- ◆ 只有定性分析，缺乏定量分析？
- ◆ 只有一次结果，缺乏重复验证？

Tips: (1) 展示探究本质的全面结果 (2) 多视角量化分析结果提升原因 (3) 模型细节明确，提供代码



Method	Overall Accuracy	
	top-1	top-3
LSTM-Question+Image+Pre-VQA	24.98	40.40
Hie-Question+Image+Pre-VQA	43.14	59.44
FVQA (top-3-QQmapping)	56.91	64.65
FVQA (Ensemble)	58.76	-
Straight to the Facts (STTF)	62.20	75.60
Reading Comprehension	62.96	70.08
Out of the Box (OB)	69.35	80.25
Human	77.99	-
Mucko	73.06	85.94

Table 1: State-of-the-art comparison on FVOA dataset.



Abstract

Knowledge-based visual question answering requires the ability of associating external knowledge for open-ended cross-modal scene understanding. One limitation of existing solutions is that they capture relevant knowledge from text-only knowledge bases, which merely contain facts expressed by first-order predicates or language descriptions while lacking complex but indispensable multimodal knowledge for visual understanding. How to construct vision-relevant and explainable multimodal knowledge for the VQA scenario has been less studied. In this paper, we propose MuKEA to represent multimodal knowledge by an explicit triplet to correlate visual objects and fact answers with implicit relations. To bridge the heterogeneous gap, we propose three objective losses to learn the triplet representations from complementary views: embedding structure, topological relation and semantic space. By adopting a pre-training and fine-tuning learning strategy, both basic and domain-specific multimodal knowledge are progressively accumulated for answer prediction. We outperform the state-of-the-art by 3.35% and 6.08% respectively on two challenging knowledge-required datasets: OK-VQA and KRQA. Experimental results prove the complementary benefits of the multimodal knowledge with existing knowledge bases and the advantages of our end-to-end framework over the existing pipeline methods. We make our code available at <https://anonymous.4open.science/r/MuKEA-87D7>.

英语写作规范 (7) —— 正确的文献引用

终于！终于！写完啦！



- [1] Khaled Al-Naami, Swarup Chandra, Ahmad Mustafa, and et al. 2016. Adaptive Encrypted Traffic Fingerprinting with Bi-Directional Dependence. In *Proceedings of the 32nd Annual Conference on Computer Security Applications (ACSAC '16)*. Association for Computing Machinery, 177–188.
- [2] Emily Alsentzer, John Murphy, and et al. 2019. Publicly Available Clinical BERT Embeddings. In *Proceedings of the 2nd Clinical Natural Language Processing Workshop*. Association for Computational Linguistics, Minneapolis, Minnesota, USA, 72–78. <https://doi.org/10.18653/v1/W19-1909>
- [6] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and et al. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Association for Computational Linguistics, Minneapolis, Minnesota.
- [7] Ali Doganaksoy, Baris Ege, Onur Koçak, and Fatih Sulak. 2010. Cryptographic Randomness Testing of Block Ciphers and Hash Functions. *IACR Cryptol. ePrint Arch.* (2010), 564.

Google Scholar 的格式就这样！
某篇A类论文的格式就这样！
信息就是找不着！



- [9] Gerard Draper-Gil, Arash Habibi Lashkari, Mohammad Saiful Islam Mamun, and Ali A. Ghorbani. 2016. Characterization of Encrypted and VPN Traffic using Time-related Features. In *Proceedings of the 2nd International Conference on Information Systems Security and Privacy - ICISSP, INSTICC, SciTePress*, 407–414.
- [43] Zhengyan Zhang, Xu Han, Zhiyuan Liu, Xin Jiang, Maosong Sun, and Qun Liu. 2019. ERNIE: Enhanced Language Representation with Informative Entities. In *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers*. Association for Computational Linguistics, 1441–1451.
- [44] Zhe Zhao, Hui Chen, Jinbin Zhang, and et al. 2019. UER: An Open-Source Toolkit for Pre-training Models. *EMNLP-IJCNLP 2019* (2019), 241.
- [33] Meng Shen, Jinpeng Zhang, Liehuang Zhu, and et al. 2021. Accurate Decentralized Application Identification via Encrypted Traffic Analysis Using Graph Neural Networks. *IEEE Trans. Inf. Forensics Secur.* 16 (2021), 2367–2380.

英语写作规范（7）——正确的文献引用

Tips: (1) 遵守会议/期刊标准模板 (2) 快速、准确导入文献信息 (3) 认真、多次检查文献格式

期刊/会议标准

HOW TO USE THE IEEEtran BIB_T_EX STYLE

How to Use the IEEEtran BIB_T_EX Style

Michael Shell, Member, IEEE

Abstract—This article describes how to use the IEEEtran.bst BIB_T_EX style file to produce bibliographies that conform to the standards of the publications of the Institute of Electrical and Electronics Engineers (IEEE).

Index Terms—bibliography, BIB_T_EX, IEEE, L_AT_EX, paper, references, style, template, typesetting.

I. INTRODUCTION

THE IEEEtran.bst BIB_T_EX style file described in this document can be used with BIB_T_EX to produce L_AT_EX bibliographies of high quality that are suitable for use in IEEE publications. Other potential applications include thesis and academic work, especially when such work is in the area of electrical and/or computer engineering.

This document applies to version 1.12 and later of the IEEEtran BIB_T_EX style. Prior versions do not have all of the features described here. IEEEtran.bst will display the version number on the user's console during execution. The most recent version of this package can be obtained on CTAN [1] and may also be mirrored at various places within IEEE's website [2]. Additional support may be found at the IEEEtran homepage [3].

It is assumed that the reader has a basic understanding of the operation and use of BIB_T_EX. Documentation for the use of BIB_T_EX includes the user's guide [4] as well as supplementary information such as a comprehensive tutorial [5], FAQs [6], [7], and a guide using practical examples [8]. The large collection of sample bibliographies and string definitions at the T_EX User Group Bibliography Archive may also be of help [9]. General support for BIB_T_EX related questions can be obtained in the usenet newsgroup comp.text.tex.

Note that the references section of this document is used for two purposes: (1) to provide information where additional information can be found; and (2) to provide examples of references created using the IEEEtran BIB_T_EX style. The first few citations above fall into the first category, while the vast

may not carry the non-IEEE compliant (natbib compatible, etc.) variants. See the CTAN site [1] for the complete set of files.

IEEEtran.bst_HOWTO.pdf: This documentation.

IEEEtran.bst: The standard IEEEtran BIB_T_EX style file (unsorted, i.e., references will appear in the order in which they are cited). Recommended for work that is to be submitted to the IEEE.

IEEEtranS.bst: The IEEEtran BIB_T_EX style file, but with additional sorting code, similar to that of plainnat.bst, which sorts the entries based on the names of the authors, editors, organizations, etc. Some IEEE conferences/publications may allow/use sorted bibliographies, but the vast majority are unsorted.

IEEEtranN.bst: Like IEEEtranS.bst, but based on plainnat.bst and is compatible with Patrick W. Daly's natbib package [10]. Not for normal IEEE use.

IEEEtranSN.bst: Sorting version of IEEEtranN.bst. Not for normal IEEE use.

IEEEexample.bib: A BIB_T_EX database that contains the references shown in the references section of this document. Users can copy the entries therein to serve as starting templates. The entries also have comments which may be of additional help.

IEEEfull.bib: A file that contains a comprehensive set of BIB_T_EX string definitions for the full names of IEEE journals and magazines. Because IEEE's bibliography style uses abbreviated journal names, this file's intended use is for specialized or non-IEEE related work.

IEEEabbrv.bib: Same as above, but contains the abbreviated form of the journal and magazine names. Recommended for work that is to be submitted to the IEEE.

BIB_T_EX .bst files can be accessed system-wide when they are placed in the

DBLP参考文献信息

computer science bibliography

Search dblp

powered by Complete Search, courtesy of Hannah Bast, University of Freiburg

Publication search results

found 17 matches

2021

Yicong Hong, Qi Wu, Yuankai Qi, Cristian Rodriguez Opazo, Stephen Gould: A Recurrent Vision-and-Language BERT for Navigation. CVPR 2021: 1643-1653

Zizheng Pan, Yicong Hong, Ming-Hsuan Yang, Anton van den Hengel, Qi Wu: to Know-Where: An Object-and-Room Informed Sequential BERT for Indoor Vision-Navigation. ICCV 2021: 1635-1644

Yuankai Qi, Yan Huang, Qi Wu, Liang Wang, Tieniu Tan: ~View Enhanced Model for Vision and Language Navigation. ACM Multimedia 2021: 5101-

Yicong Hong, Qi Wu, Lianhua Yang, Dong An, Shuanglin Sima, Lians Wang:

BibTeX record conf/cvpr/Hong0QOG21

download as .bib file

```
#inproceedings{DBLP:conf/cvpr/Hong0QOG21,
  author = {Yicong Hong and
    Qi Wu and
    Yuankai Qi and
    Cristian Rodriguez Opazo and
    Stephen Gould},
  title = {(V2L) (BERT): (A) Recurrent Vision-and-Language (BERT) for Navigation},
  booktitle = {(IEEE) Conference on Computer Vision and Pattern Recognition, (CVPR) 2021, virtual, June 19-25, 2021},
  pages = {(1643--1653)},
  publisher = {Computer Vision Foundation / (IEEE)},
  year = {2021},
  url = {https://openaccess.thecvf.com/content/CVPR2021/html/Hong_V2L_BERT_A_Recurrent_Vision-and-Language_BERT_for_},
  timestamp = {Mon, 30 Aug 2021 17:00:27 +0200},
  biburl = {https://dblp.org/rec/conf/cvpr/Hong0QOG21.bib},
  bibsource = {dblp computer science bibliography, https://dblp.org}}
```

一键完善论文参考文献信息:

<https://github.com/yuchenlin/rebiber>

将不完整的 arXiv 链接转换成最终发表的链接。

```
@article{lin2020birds,
  title={Birds have four legs?! NumerSense: Probing Numerical Commonsense Knowledge of Pre-tr.},
  author={Lin, Bill Yuchen and Lee, Seyeon and Khanna, Rahul and Ren, Xiang},
  journal={arXiv preprint arXiv:2005.00683},
  year={2020}}
```

An example normalized output entry with the official information:

```
@inproceedings{lin2020birds,
  title = "{B}irds have four legs?! {N}umer{S}ense: {P}robing {N}umerical {C}ommonsense {K}nowledge",
  author = "Lin, Bill Yuchen and
    Lee, Seyeon and
    Khanna, Rahul and
    Ren, Xiang",
  booktitle = "Proceedings of the 2020 Conference on Empirical Methods in Natural Language Process.",
  month = nov,
  year = "2020",
  address = "Online",
  publisher = "Association for Computational Linguistics",
  url = "https://www.aclweb.org/anthology/2020.emnlp-main.557",
  doi = "10.18653/v1/2020.emnlp-main.557",
  pages = "6862--6868",
}
```

使用方法:

```
pip install -e git+https://github.com/yuchenlin/rebiber.git
rebiber -i /path/to/input.bib -o /path/to/output.bib
```

其中 input.bib 为待完善的参考文献，output.bib 为工具输出的补全的参考文献引用，可直接放到 LaTeX 中。

英语写作规范（8）——坚守的学术道德

一篇抄十篇，CVPR Oral被指大量抄袭，大会最后一天曝光，合著者忙甩锅

机器之心 2022-06-25 12:57 发表于北京

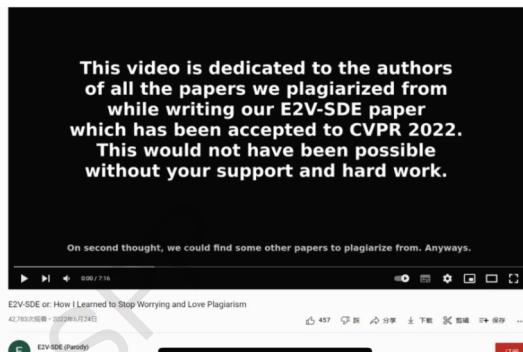
机器之心报道

机器之心编辑部

围观群众吐槽，作者忙着分锅。

人工智能顶会 CVPR 2022 的最后一天，一篇重点论文意外火了，但不是以正经的方式。

这天，YouTube 上出现了一个新的账号，它的目的非常单纯，就是为了指证一篇 CVPR 2022 Oral 论文涉嫌抄袭。



如何看待 ICCV21 接收的某港科大学生为一作的论文被指抄袭 ICML21 发表的论文？

正在发生

ICCV21接收论文m-RevNet: Deep Reversible Neural Networks with Momentum被指出与ICML21接收论文Momentum residual neural networks在核心思路、实验和图表上有多处雷同，疑似抄袭

ICML21论文作者的声明和详细的抄袭证据：

<https://michaelsdr.github.io/momentumnet/plagiarism/>

michaelsdr.github.io/momentumnet/plagiarism/

TOP大学派了

对自己一辈子声誉的影响

对合作者的影响

对所在学校/单位的影响

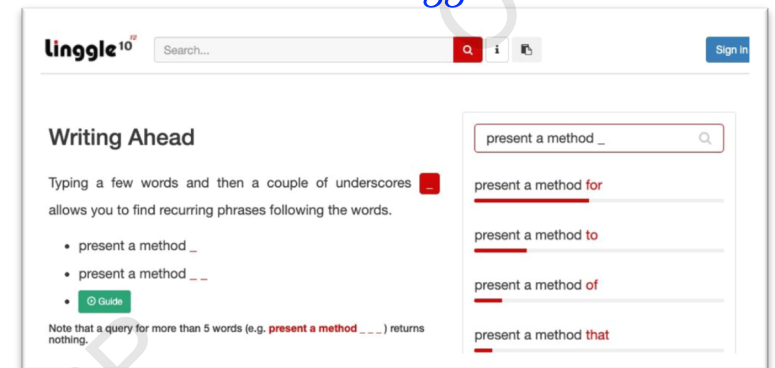
...

严守学术道德，
切记抄袭！

日常积累常见表达

- I. 研究概述常用句型
1. Previous research has shown that ...
2. In most studies of ... attention has been given to ...
3. The previous work on ... has indicated that it is ...
4. ... models have been proven useful in ...
5. There have been a few studies highlighting ...
6. Great concern have arisen ... due to the increasing number of ...
7. Most ... conduct ... testing on ... to monitor ... performance.
8. However, the problems exist in ...
9. However, there appears to have ...
10. Because of ... is impossible.
- II. 研究范围 常用句型
1. With the aim of ...
2. This paper is intended to ...
3. This paper aims at providing ...
4. The primary goal of this research is ...
5. The overall objective of his study is ...
6. The author(s) made this study in order to find ...

英语搭配工具 Linggle



学术研究 之日常积累



学术研究之日常积累

日常 **5L** 积累:

☀ Paper List

☀ Idea List

☀ Math List

☀ English List

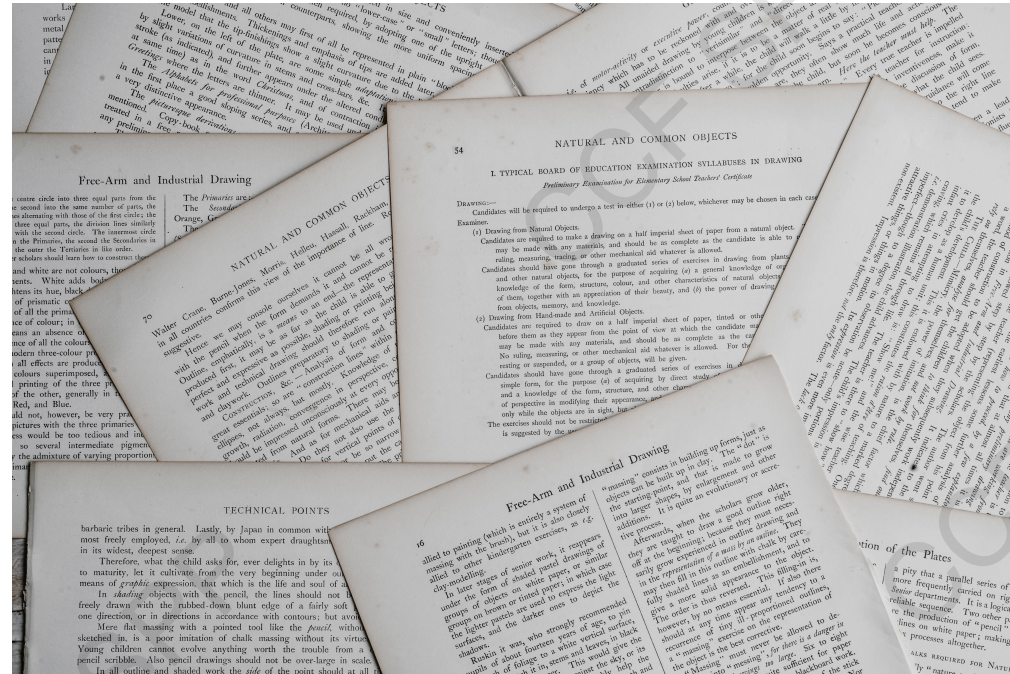
☀ Code List



欲穷千里目，更上一层楼

Paper List

- 如何**述**论文？
- 如何**读**论文？
- 如何**讲**论文？
- 如何**整理** & **总结**论文？



Paper List——如何找论文？

入门阶段

盲目搜索 ✕

- 刷顶会期刊
- 刷 arxiv
- 刷学术自媒体等

精准定位 ✓

- 讲座，报告
- 会议 Workshop
- 会议 Tutorial
- 综述论文

入门阶段从有影响力的经典工作开始阅读！

Paper List——如何找论文？

终于找到了一篇自己想要看的论文

— 从领域一角到整个领域

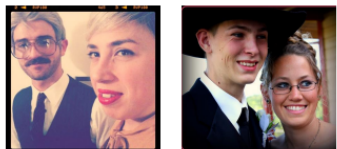
VQA 1.0 (2015)



What color are her eyes?
What is the mustache made of?

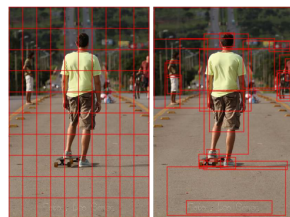
VQA 2.0 (2017)

Who is wearing glasses?
man woman



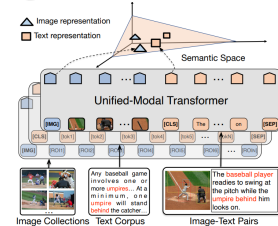
基于知识的VQA
FVQA (2017)

Bottom-up (2018)

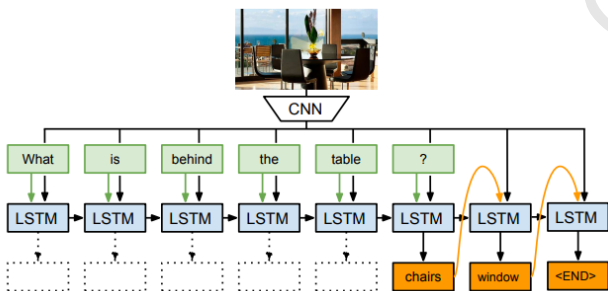


VQA中的计数问题
Learning to count
(2018)

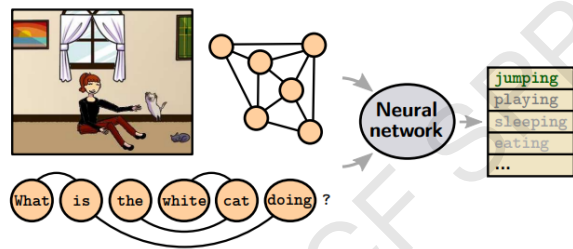
UNIMO (2021)



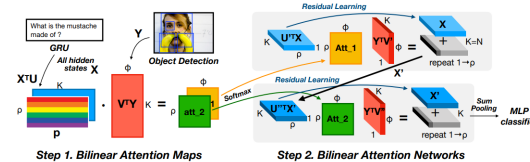
主线



Ask Your Neurons (2015)



Graph VQA (2017)



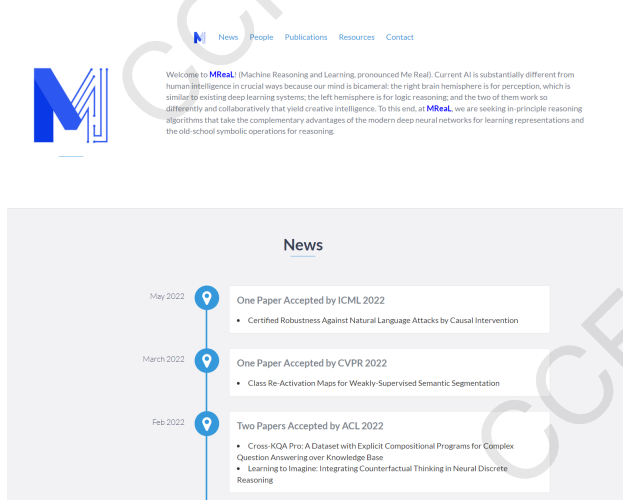
BAN (2018)

VQA可解释性
interpretable-VQA (2021)

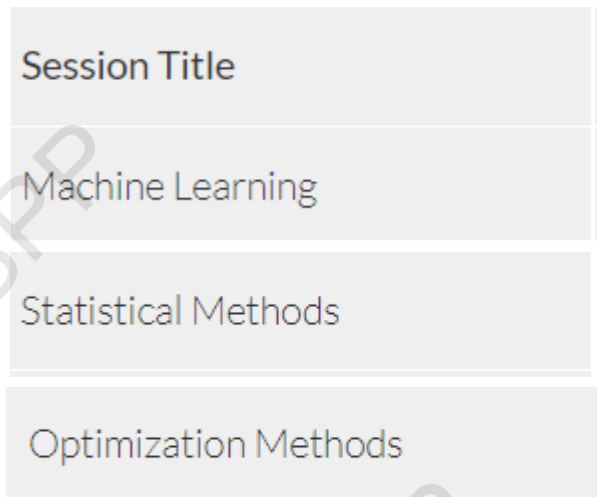
VQA中的偏置问题
CF-VQA (2021)

Paper List——如何找论文？

终于找到了一篇自己想要看的论文 — 从冰山一角到整个冰山



Follow研究团队



会议track

2. Related Work

Single-purpose vision-language models. Over the last decade, specialized and effective approaches have been developed for vision-language tasks, including image captioning [10, 21, 25, 33, 47, 54], phrase grounding [37, 38, 43], referring expression comprehension [22, 34], visual question answering (VQA) [2, 11, 16, 48, 51, 55], visual dialog [7], and text-to-image generation [6]. Advances that have pushed the performance envelope include cross-model transformer architectures [45], powerful self-supervised [3, 8, 28] and multitask [41] language models, pretrained visual representations from object and attribute detectors [1, 57] or text conditioned detectors [20], and large-scale image/video-text [19, 27, 39, 56] pretraining.

Related work

Paper List——如何找论文？

终于找到了一系列自己想要看的论文

— 是否值得花时间去

- 是否是**顶会顶刊**
- Incremental的工作还是**novel**的工作
- 对自己的研究是否会有**启发**



Paper List——如何读论文？

读之前——积极的提问

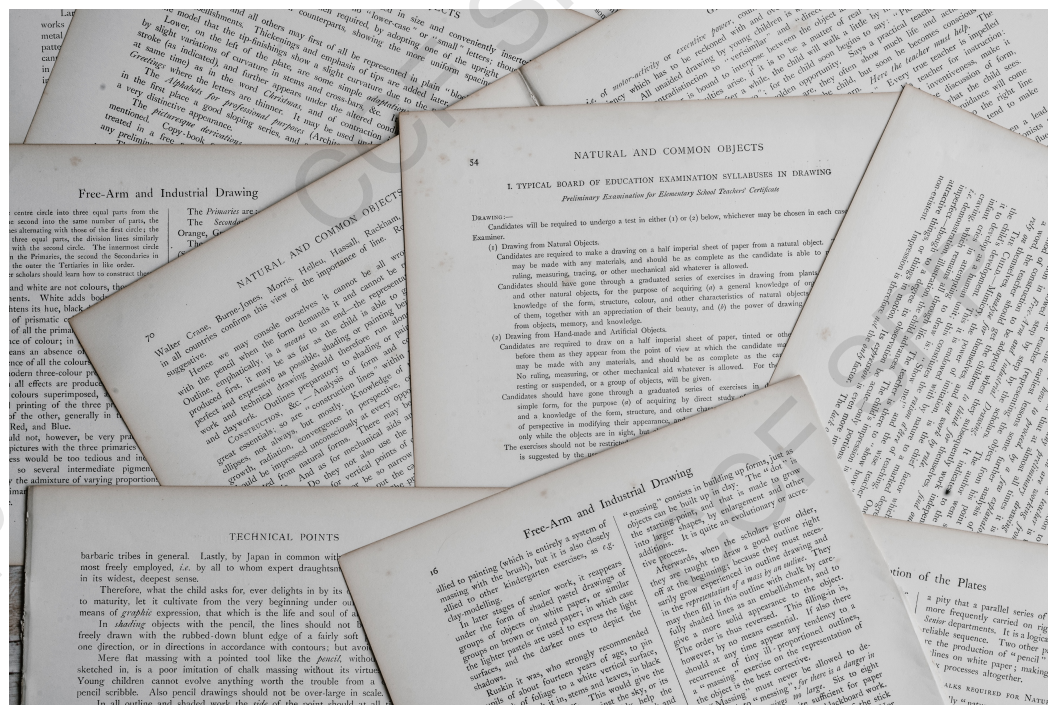
☀ 文章是关于**什么领域**的？

☀ **解决什么问题？为什么要解决此问题？**

☀ **创新点在哪里？为何巧妙？**

☀ **如何验证并得出结论的？**

☀ 对“我”的研究有何**启发**？



Paper List——如何读论文？

读之中——批判性思考

- ☀ 论文是否**正确、真正地**解决了问题？
- ☀ 论文所用的方法是否具有**局限性**？
- ☀ 论文的**实验能否证明观点** or 方法的有效性？
- ☀ 所读论文**没有解决的问题**，“我”能解决吗？
- ☀ 能采用比论文中**更简单的方法**解决问题吗？



Paper List——如何读论文？

读之中——循序渐进阅读！

- ☀ **第一遍：阅读标题、摘要和文中图表**
- ☀ **第二遍：阅读引言、结论、关键信息，结合图表快速扫视其他内容**
- ☀ **第三遍：整体阅读论文，可跳过陌生复杂的技术、数学公式**
- ☀ **第四遍：增强对数学、技术和未知术语了解**
- ☀ **为对领域深入研究，还可再多读几遍**

Paper List——如何读论文？

读之中——系统记录！

☀️ 记录文章的主体内容

☀️ 记录有启发的地方

☀️ 记录有问题的地方

1、本文提出的这种「泛化性评估」，有何意义???

- 当「Text 数据缺失」时，模型的性能如何？
感觉没有意义呀!!!
- 了解「预训练模型」可行的机理
(「勉强强强」算是个启发吧)
- 现实世界中，模态缺失，是会发生的
由于「隐私」或「安全限制」，部分模态的数据会得不到
(感觉有点扯)
 - 社交网络：不可得到「位置信息」，如果「用户拒绝共享」
 - 健康应用：不会得到「全部记录」，如果「病人不愿意分享」
- 基于上述原因，研究 Transformer 模型面对「模态确实数据」时，能否有强鲁棒性，很重要
(感觉这里的 modal 指的范围更广!!!)

问题 & 启发

标题及出处

- 标题：Understanding the Role of Individual Units in a Deep Neural Network
理解单个单元在深度神经网络中的作用
([ˌɪndrɪˈvɪdʒuəl])([ˈjuːnɪts])
- 出处：2020 PNAS, 美国国家科学院院刊 (Proceedings of the National Academy of Sciences of the United States of America)
 - link 链接：论文链接
 - demo 主页：demo
 - github 链接：<https://github.com/davidbau/dissect>
 - 作者：Bolei Zhou
 - 研究团队：香港中文大学信息工程系，香港
 - 其他链接：2017 年，周博磊对 CNN 学习深度特征的可解释的 paper：链接
Network Dissection: Quantifying Interpretability of Deep Visual Representations
 - 补充材料：<https://www.pnas.org/content/suppl/2020/08/31/1907375117.DCSupplemental>

标题 & 出处

Contribution

- 第一篇工作：探索「多模态 Transformer」对「模态缺失数据」的泛化性
- 重要的实验发现
 - 「多模态 Transformer」对「缺失模态数据」敏感
 - 「最优的模态融合策略」是「数据集依赖的」
 - 不存在「统一的最优策略」
- 通过「多任务优化」，提升模型的鲁棒性
且提出了一种「可微分算法」，来获取「最佳融合策略」
- 在「三个数据集」上，做了大量的「实验证明」

贡献

Paper List——如何读论文？

读之后——思考文章对接下来的工作有何帮助！

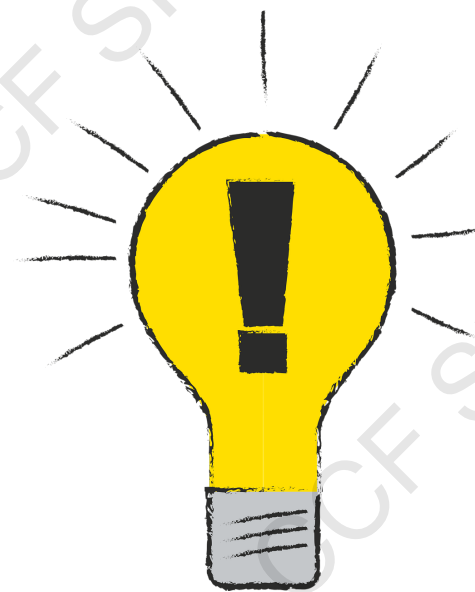
☀ “我”错过了哪些**相关论文**？

☀ 这篇论文值得**关注**吗？对我的工作**有何帮助**？

☀ 这个领域的**领头人**有哪些？哪些**团队**值得**关注**？

☀ 如果我遇到作者，“我”会问什么问题？

☀ **下一步**可以做什么？



Paper List —— 如何讲论文

报告前

☀ 确定值得讲的论文

- 领域顶会顶刊
- 契合研究方向
- 能够带来启发
- 研究具有体系

☀ 确定讲论文的受众

- 组内
- 公众
- 大同行
- 小同行

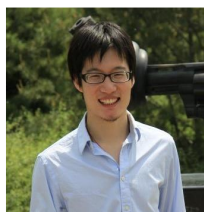


Paper List —— 如何讲论文

报告前

☀ 论文作者团队调研

- 团队研究方向
- 团队发展历史
- 团队代表成果



Peng CUI

崔 鹏



Associate Professor (*Tenured*)

Lab of Media and Network
Department of Computer Science and Technology
Tsinghua University

Address: Room 9-316, East Main Building, Tsinghua University, Beijing 100084, P.R.China
Tel: +86-10-6279 0810
Fax: +86-10-6277 1138
Email: cui_p@tsinghua.edu.cn

[[News](#) | [Biography](#) | [Research Interests](#) | [Publications](#) | [Honors/Awards](#) | [Professional Experiences](#)]

Attention: We have several openings for Postdoctoral Fellowships on causal learning and related topics.

News

- We have kicked off the [NICO Challenge](#), a contest for OOD image recognition.
- Together with [Susan Athey](#), we published a [perspective paper on stable learning](#) in *Nature Machine Intelligence*.
- I gave an [Annual Progress Report on Out-of-Distribution Generalization](#) in VALSE2021.
- I gave a tutorial on [Stable Learning and its Causal Implication](#) in VALSE2021.
- We published a survey paper on out-of-distribution generalization [here](#), accompanied with a [website](#) to maintain the paper list for OOD generalization.

Biography

Peng Cui is an Associate Professor with tenure in Tsinghua University. He got his PhD degree from Tsinghua University in 2010. His research interests include causally-regularized machine learning, network representation learning, and social dynamics modeling. He has published more than 100 papers in prestigious conferences and journals in data mining and multimedia. His recent research won the IEEE Multimedia Best Department Paper Award, SIGKDD 2016 Best Paper Finalist, ICDM 2015 Best Student Paper Award, SIGKDD 2014 Best Paper Finalist, IEEE ICME 2014 Best Paper Award, ACM MM12 Grand Challenge Multimodal Award, and MM13 Best Paper Award. He is PC co-chair of IJCAI2019 and MM2020, SPC or area chair of ICML, KDD, WWW, UCAI, AAAI, etc., and Associate Editors of IEEE TKDE, IEEE TBD, ACM TIST, and ACM TOMM etc. He received ACM China Rising Star Award in 2015, and CCF-IEEE CS Young Scientist Award in 2018. He is now a Distinguished Member of ACM and CCF, and a Senior Member of IEEE.

Selected Publications

☀ 确定论文讲解结构

- 演讲逻辑结构: **动机**、**思路**、**结果**、**不足**
- 清晰图文展示
- 演讲文稿练习



Paper List —— 如何讲论文

报告中 & 报告后

- 以“**理解**”为前提
- 以“**问题**”为核心
- 以“**反馈**”为目的

组会 PPT/onenote 笔记讲解

技能习得的三个阶段

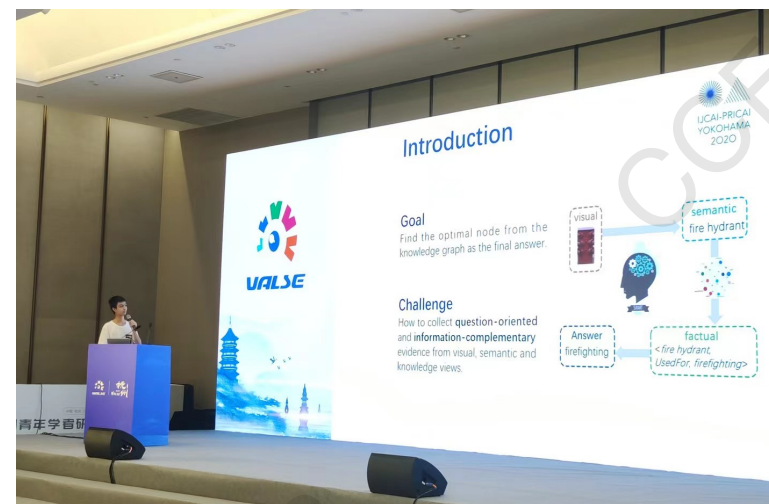
- 认知阶段 (第8章的解决问题思路)
 - 定义执行任务的算子
 - 记录陈述性事实 (如: 倒车档是向上向左)
- 关联阶段
 - 纠正错误的理解 (如: 英语语法的过去式并非都是加 ed)
 - 各算子间的联系被强化 (如: 思考换挡的停顿时间减少)
- 自主阶段
 - 整个过程变得**自动化**和**迅速**;
 - 大脑的中枢认知会逐渐退出执行任务

随着对专业技能的掌握, 大脑需要耗费的能量会减少

Poster 讲解



大会报告讲解



Paper List——文献整理

- 按**内容**整理：将论文按照不同的主题进行分类
- 按**时间**整理：年份 + 来源 + 题目 + 内容简介

1992-Learning factorial codes by predictability minimization..pdf
2006-science-Reducing the dimensio...neural networks-AutoEncoder 模型.pdf
2008-ICML-Extracting and Composin...toencoders-对 AutoEncoder 加噪声.pdf
2011-ICML-Contractive Auto-Encoder...variance During Feature Extraction.pdf
2013-Bengio-Representation Learning A Review and New Perspectives.pdf
2014-ICLR-Auto-Encoding Variational Bayes-VAE 模型.pdf
2014-ICML-Learning to Disentangle F...Variation with Manifold Interaction.pdf
2014-ICML-Stochastic Backpropagati...ference in Deep Generative Models.pdf
2014-NIPS-Generative Adversarial Nets-GAN 模型.pdf
2015-Adversarial Autoencoders-AAE 模型.pdf
2016-arxiv-Early Visual Concept Learni...with Unsupervised Deep Learning.pdf
2016-NIPS-Disentangling factors of v...l training-将 VAE 和 GAN 结合 的模型.pdf
2016-NIPS-InfoGAN- Interpretable Re...tive Adversarial Nets-InfoGAN 模型.pdf
2017-ICLR-beta-VAE Learning Basic V...riational Framework-betaVAE 模型.pdf
2017-NIPS-Learning Disentangled Re...upervised Deep Generative Models.pdf
2018-arxiv-Recent Advances in Autoe...er-Based Representation Learning.pdf
2018-arxiv-Towards a Definition of Disentangled Representations.pdf
2018-arxiv-Understanding disentangling in β -VAE.pdf
2018-arxiv-Variational Inference- A Review for Statisticians.pdf
2018-ICLR-A Framework for the Quan...tangled Representations-评判标准.pdf
2018-ICLR-Variational Inference of Di...nabeled Observations-DIPVAE 模型.pdf
2018-ICML-Disentangling by Factorising-factorVAE 模型.pdf

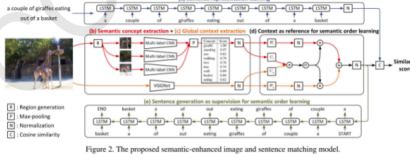
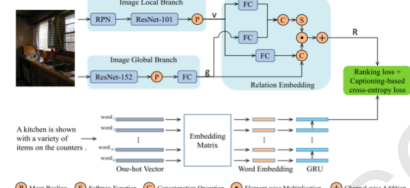
按时间整理

基础	应用
综述	VQA
NLP相关	Visual Reas...
多模态表示	image capti...
NLP预训练	video quest...
多模态融合	visual dialog
多模态预训练	image text...
元学习	referring ex...
多模态对齐	text2img
生成式学习	visual relati...
生成对抗网	Voice-Visu...
注意力相关	video capti...
自监督学习	Language...
对抗攻击	多模态机器...
图神经网络...	虚假新闻检测
终身学习	Video mom...
零样本学习	speech
可解释学习	

按内容整理

Paper List——文献整理

- 可以将领域内的**经典模型**方法记录在 Excel 表格中，便于查找。

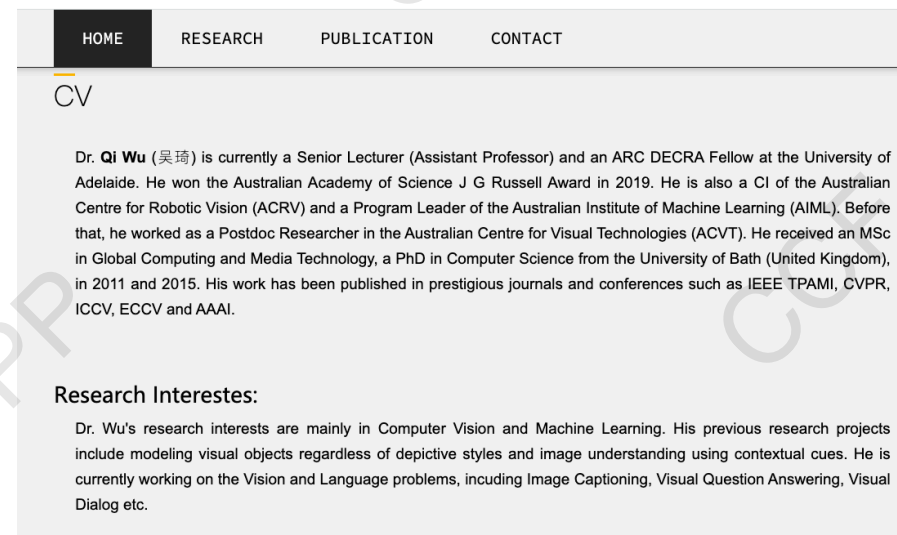
	A	B	C	D	E	F	G	H	I	J	L	M	N	O	P	Q	R	S
1	年份	模型类型	模型名称	代码链接	模型	主要思路	是否局部	图像输入	文本输入	输出	MSCOCO							
2											sentence retrieval			image retrieval			sentence r	
3											R@1	R@5	R@10	R@1	R@5	R@10	R@1	R@5
4											这里用的测试集都是1k而不是5k的，同时融合的方法会做5折交							
5	2017	共享空间	VSE++ 《VSE++: Improving Visual-Semantic Embeddings with Hard Negatives	https://github.com/fartashf/vsepp	$\ell_{MH}(i, c) = \max_c [\alpha + s(i, c') - s(i, c)]_+ + \max_{c'} [\alpha + s(i', c) - s(i, c)]_+.$	由sum hinge转为max hinge，关注最难负样本是从整个训练集中选，而不是从batch中选。	无	random crop+fine-tuning resnet-152 CNN2048维特	300维词向量 Bi-GRU 到1024维	mini-batch : 128	64.6	90	95.7	52	84.3	92	52.9	8
6	2018	共享空间	SCO			模型主要是学习 semantic concept 和 semantic order，在现有词汇中抽取有意义的词并在当前配对文本描述的监督下做分类。之后 global 和	有	在测试阶段用了一种方法提取特征	global 特征时用 10crop 做平均得到	300维词向量 (max_len=50)	1024维	69.9	92.9	97.5	56.7	87.5	94.8	55.5
7	2021	共享空间	CRGN 《Deep Relation Embedding for Cross-Modal Retrieval》	https://github.com/zyfsa/CRGN		射到公共区域方法，主要思想是图像 global presentation 过 attention 机制融合 (gate) region 域特征，之后本图像映射到公共区域，在 ranking hard loss 和图像的文	在训练的第二步骤使用了 Faster RCNN 提取了 36 个区域	ResNet-152 CNN	初始词向量: mini-Word2vec (T=4)	1024维 mini-batch : 128	73.8	95.6	98.5	60.1	88.9	94.5	70.5	9

Idea List

- Incremental idea or breakthrough ?
- 短期 or 长期?
- 个人 or 团队?



CCF 数字图书馆，大家关注的热点问题



大牛的研究路线

Idea List

- 及时进行交流、复盘
- 记录**知识点**、**遇到的问题**、**好的想法**、**论文阅读计划**、**实验计划**等

组会复盘模板

- 需要了解的知识点 & 技术方面：概率图、图卷积神经网络、贝叶斯定理
- 产生的问题：找到参数 θ 使得样本 X 的联合概率密度最大有何意义？
- 思想方面：多分析数据，多看这个领域的较好方法；发现数据中的问题，从方法中吸取思想
- 需要养成的好习惯：每天坚持锻炼 30 分钟
- 有趣的东西：学会用 LaTeX 画图
- 要写的博客：hexo 博客搭建技巧
- 计划内容
 - **论文计划**：2013 Bengio 的经典 paperVAE 及 β -VAE 的 paper
 - **实验计划**：VAE 的实验跑通 + 看明白代码
 - **生活计划**：周末羽毛球
 - 老师任务：形成 VAE 的理论 report
- 其他：参加 IJCAI 大会

Math List

- 积累**基础的**数学知识
- 对模型中重要的**数学公式**进行**推导**
- 积累领域常用的**符号、公式**

为何特征向量的模为 1 后，损失值就不会出现 NAN、INF 了？ and 超参 τ 的有何作用？

- 将 $f(x)$ 、 $f(x^+)$ 和 $f(x^-)$ **向量的模归一化到 1**后

$$f(x) \times f(x^+) = |f(x)| \times |f(x^+)| \times \cos\theta = 1 \times 1 \times \cos\theta = \cos\theta$$

即 $f(x) \times f(x^+)$ 的范围在 $[-1, 1]$

同理， $f(x) \times f(x^-)$ 的范围也在 $[-1, 1]$

- $e^{f(x)^T f(x^+)}$ 范围在 $[\frac{1}{e}, e]$

- $\frac{e^{f(x)^T f(x^+)}}{e^{f(x)^T f(x^+)} + \sum_{i=1}^N e^{f(x)^T f(x^-)}}$ 的范围是 $[\frac{1}{(N+1)e^2}, \frac{e^2}{N+1}]$

(最大值：分子取最大值 e ，分母取最小值 $\frac{N+1}{e}$)

(最小值：分子取最小值 $\frac{1}{e}$ ，分母取最大值 $(N+1)e$)

- 即 $-\log \frac{e^{f(x)^T f(x^+)}}{e^{f(x)^T f(x^+)} + \sum_{i=1}^N e^{f(x)^T f(x^-)}}$ 的范围是 $[-\log \frac{e^2}{N+1}, -\log \frac{1}{(N+1)e^2}]$ ，即 $[-2\log(e) + \log(N+1), 2\log(e) + \log(N+1)]$

- 即又 N 的值有限，固 Infonce 的 Loss 值在有限范围内

- 加上超参 τ 后， $f(x) \times f(x^+) / \tau$ 的范围在 $[-\frac{1}{\tau}, \frac{1}{\tau}]$

- $\mathcal{L}_y = -\log \frac{\exp(y \cdot y' / \tau)}{\sum_{i=0}^N \exp(y \cdot y' / \tau)}$ 的范围是 $[-\log \frac{e^{\frac{2}{\tau}}}{N+1}, -\log \frac{1}{(N+1)e^{\frac{2}{\tau}}}]$ ，即 $[-\frac{2}{\tau}\log(e) + \log(N+1), \frac{2}{\tau}\log(e) + \log(N+1)]$

当 τ 小于 1 时，Loss 的整体范围更广

数学知识总结示例

English List

- 积累**专业知识相关的英语词汇**
- 积累论文中的**好句**、**好词**

好句

2019年8月29日 星期四 下午3:34

- Note that the use of fasterRCNN and Skip-Gram for generic representation extraction might be not optimal, but we empirically find they can already achieve satisfactory performance. (对于自己论文不是特别关注的, 做的不够充分的部分的解释方法)
- Although similar ideas of gated feature fusion have been previously explored (Arevalo et al. 2017), directly applying them to the fusion of metrics is infeasible. (对于自己的论文采用的与别人论文相似的方法, 只做了一点点优化的说法)
- As the first work for end-to-end spoken question answering (SQA), our model gets the results close to the performance of cascading ASR and QA models. Although the room for improvement exists, it is a good first step towards end-to-end SQA. (对于自己的论文没有优于SOTA方法的说法)

英语总结示例

Code List

- 如何**入门深度学习**?
- 如何**进阶**?
- 如何进行**代码积累**?

Microsoft
CNTK



Caffe

Caffe2

PYTORCH

Chainer

K Keras

TensorFlow

theano

dy/net

mxnet

GLUON

Code List – 如何入门

- 学习常用框架
- 实践检验学习效果

比赛名称	比赛链接
Titanic	https://www.kaggle.com/competitions/titanic
House Prices	https://www.kaggle.com/competitions/house-prices-advanced-regression-techniques
Digit Recognizer	https://www.kaggle.com/competitions/digit-recognizer
Dogs vs Cats	https://www.kaggle.com/competitions/dogs-vs-cats-redux-kernels-edition
Leaf Classification	https://www.kaggle.com/competitions/leaf-classification

深度学习入门比赛推荐

工具	发布者	支持语言	支持系统
TensorFlow(工程)	Google	Python/C++/Java/Go	Linux/MacOS/Android/iOS/Windows
Caffe	UC Berkeley	Python/C++/Matlab	Linux/Mac OS/Windows
CNTK	Microsoft	Python/C++/BrainScript	Linux/Windows
MXNet	DMLC	Python/C++/Matlab/ Julia/Go/R/Scala	Linux/MacOS/Windows/Android/iOS
PyTorch(学术)	Facebook	C/Lua/Python	Linux/MacOS/Windows/Android/iOS
Theano	蒙特利尔大学	Python	Linux/Mac OS/Windows
Neon	Intel	Python	Linux
PaddlePadde(工程)	百度	Python	Linux/Windows

深度学习常用框架

Code List – 如何进阶

- 学习领域常用代码库
- 复现 & 精读经典论文代码

```
class ResNetVLBERTForPretrainingMultitask(Module):
    def __init__(self, config):

        super(ResNetVLBERTForPretrainingMultitask, self).__init__(config)

        # 使用Fast-RCNN作为图像特征提取器
        self.image_feature_extractor = FastRCNN(config,
                                                average_pool=True,
                                                final_dim=config.NETWORK.IMAGE_FINAL_DIM,
                                                enable_cnn_reg_loss=False)

        # 对象语言嵌入, 得到一个embedding
        self.object_linguistic_embeddings = nn.Embedding(1, config.NETWORK.VLBERT.hidden_size) # 固定词向量的大小
        # 改

        # 如果没有预装图像或者使用原始的像素则生成1个2048维度的词嵌入, 得到图像的embedding
        if config.NETWORK.IMAGE_FEAT_PRECOMPUTED or (not config.NETWORK.MASK_RAW_PIXELS):
            self.object_mask_visual_embedding = nn.Embedding(1, 2048)

        # 看看是否使用MVRC的LOSS
        if config.NETWORK.WITH_MVRC_LOSS:
            self.object_mask_word_embedding = nn.Embedding(1, config.NETWORK.VLBERT.hidden_size)

        # 使用辅助文字视觉嵌入, 这个嵌入是使用文字信息作为图像信息的推导进行的
        self.aux_text_visual_embedding = nn.Embedding(1, config.NETWORK.VLBERT.hidden_size)

        # image_feature_bn_eval - 图像特征评估
        self.image_feature_bn_eval = config.NETWORK.IMAGE_FROZEN_BN
```

精读经典工作代码

代码库名称	领域	链接
Huggingface	NLP / Multimodal	https://huggingface.co/
fastText	NLP	https://github.com/facebookresearch/fastText
Detectron	CV	https://github.com/facebookresearch/detectron
Detectron2	CV	https://github.com/facebookresearch/detectron2
open-mmlab	CV	https://github.com/open-mmlab

领域常用代码库

Code List – 如何积累

- 代码规范积累
- 代码复用积累
- 常用函数积累

常用函数
Numpy
Pytorch
Tensorflow

Pytorch

2020年12月2日 10:29

[torch.nn.Embedding](#)

`torch.nn.Embedding`是一个简单的存储固定大小的词典的嵌入向量的查找表，意思就是说：给一个编号，嵌入层就能返回这个编号对应的嵌入向量，嵌入向量反映了各个编号代表的符号之间的语义关系。

输入为一个编号列表，输出为对应的符号嵌入向量列表

要理解这个首先得弄清楚什么是词嵌入（Word Embedding）

Word Embedding翻译过来的意思就是词嵌入，通俗来讲就是将文字转换为一串数字。因为数字是计算机更容易识别的一种表达形式。

我们词嵌入的过程，就相当于是在给计算机制造出一本字典的过程。计算机可以通过这个字典来间接地识别文字。

词嵌入向量的意思也可以理解成：**词在神经网络中的向量表示**。

而`torch.nn.Embedding`的用处就是

1. 作为一个简单的查找表（lookup table），存储固定字典和大小词嵌入。
2. `torch.nn.Embedding`通常用于存储单词嵌入并使用索引检索它们(类似数组)。模块的输入是一个索引列表，输出是相应的词嵌入。

Embedding.weight（Tensor）-形状模块（num_embeddings, Embedding_dim）的可学习权重，初始化自（0，1）。

```
def get_label_embedding(embedding_class_dict, label_num_cnt, embedding_list, label_list, shape_list, word_option):
    """
    :param label_num_cnt:(dict) label : the number of embedding have been select for this label
    :param embedding_list: embedding have been select for all label
    :param label_list: One-to-one label corresponding to embedding list
    :param word_option: Whether to deal with word embeddings
    :param shape_list: The shape of word and ROI embedding to show
    :return: label_num_cnt, embedding_list, label_list, shape_list
    """
    for key, value in tqdm(embedding_class_dict.items()):
        if args.random_sample_option == True:
            sample_key = random.sample(embedding_class_dict.keys(), 1)
            sample_key = sample_key[0]
            embedding_class_list = embedding_class_dict[sample_key]
        else:
```

注意代码的命名，注释规范

```
def get_label_embedding(embedding_class_dict, label_num_cnt, embedding_list, label_list, shape_list, word_option):...
def get_data():...
def plot_embedding(data, label, shape_list, RGB_list, title):...
def get_savepath():...
def show_class_color(label_unqie_list, RGB_list, save_path):...
def get_class_color_dict():...
def get_class_color_dict_inline():...
```

统一函数接口，形成自己的代码库，便于复用代码

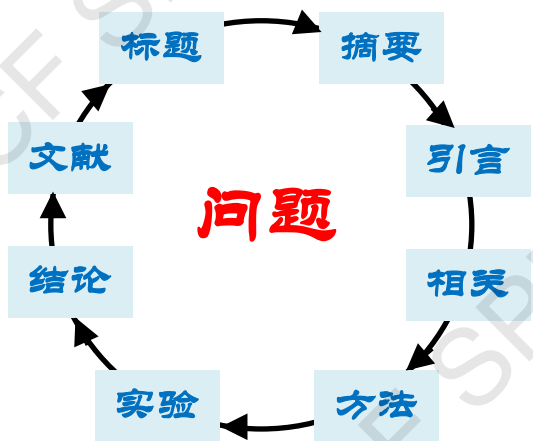
整理深度学习框架的常用函数

本讲总结 与心得分享

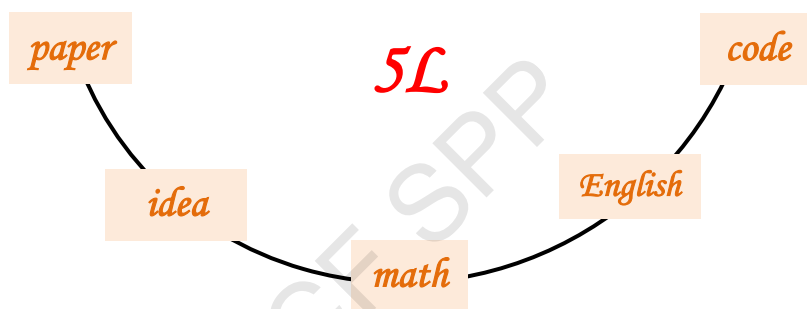
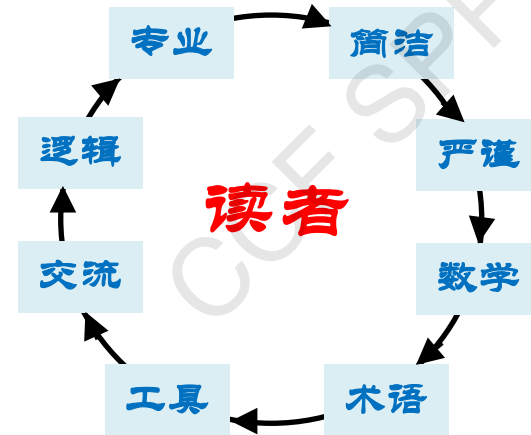


“快乐之力”——透视本质的科研认知能力

☀ 写作思路



☀ 英文规范



☀ 日常积累

有道无术，术尚可求，有术无道，止于术。

“引路之力”——循序渐进的科研规划能力

天花板

☀ *Learning by Book* (10%)

- ☞ 科研、写作书籍
- ☞ CCF领航计划报告

☀ *Learning by Example* (20%)

- ☞ 团队内学术研究资料
- ☞ 其他学者成果、分享
- ☞ 共同参与学术会议、其他学术活动

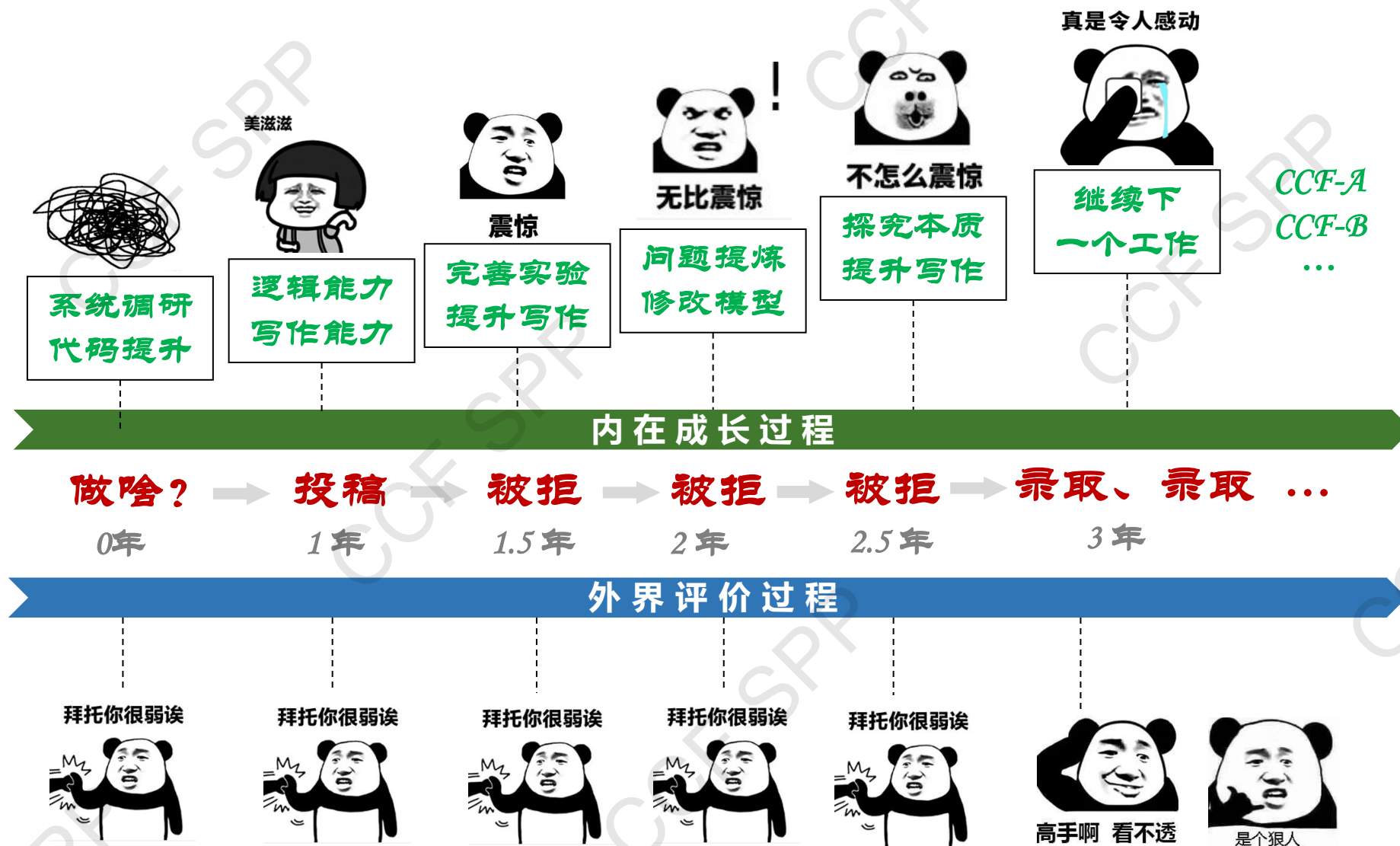
☀ *Learning by Doing* (70%)

- ☞ 了解领域发展历史，熟悉领域前沿
- ☞ 坚持领域内深耕，见证本领域发展



底线

“成长之力”——以“我”为中心的进度把控能力



具备自己的进度条，你永远可以相信你自己！

“王者之力”——面向未来的核心竞争力

ACE

Aesthetic

审美力

Creative

创造力

Empathic

同理心

参考资料

参考书籍

- ☀ Booth, Wayne C., et al. *The craft of research*. University of Chicago press, 2003.
- ☀ Petre, Marian, and Gordon Rugg. *The Unwritten Rules Of Phd Research*. McGraw-Hill Education (UK), 2010.
- ☀ Macrina, Francis L. *Scientific Integrity*. ASM Press, 2005.
- ☀ 张军平, *高质量读研: 教你如何写论文、做科研*, 2022.
- ☀ 凌晓峰, 杨强, *学术研究, 你的成功之道*, 2012.

参考课程

- ☀ [董彬, 北京大学, 北京国际数学研究中心, 《研究型学习》](#)

参考资料——如何找论文

参考资料

- ☀ [科研新手该如何找到合适的文献？\(哈尔滨工业大学 王威 教授\)](#)

实用工具

- ☀ Paperwithcode: 快速了解领域的数据集上的 SOTA 结果、相关代码
- ☀ [CCF优秀博士学位论文【历届论文】](#)
- ☀ 轻量级定制化每日arxiv论文筛选工具 [Erutan-pku/harxiv: A highlight tool for reading ArXiv papers \(github.com\)](#)

参考资料——如何读论文

参考视频

- ☀ [吴恩达，斯坦福大学，如何有效阅读论文并了解一个领域.](#)
- ☀ [尹芝南，南开大学，生命科学学院，如何高效正确的阅读文献.](#)
- ☀ [沈向洋，前微软全球执行副总裁，You are how you read.](#)

实用工具

- ☀ 论文笔记分享平台：[readpaper](#)
- ☀ 论文文献管理工具：[Zotero](#)、[Mendeley](#)

参考资料——如何讲论文

参考资料

- ☀ 王晋东. *ResearchGo*. 微软亚研, 2021.
- ☀ 黎家. *如何做好一场精彩的学术报告*. 兰州大学, 2020.

实用工具

- ☀ Powerpoint: 绘制演讲文稿, 设置排练
- ☀ Xmind: 梳理演讲逻辑结构的思维导图

参考资料——如何积累代码

参考资料

- ☀ 《深度学习》中国科学院大学，徐俊刚

实用工具

- ☀ Kaggle: 全球最大的数据科学家汇聚的平台，适合深度学习新手入门到进阶
- ☀ Huggingface: 对Pytorch友好的深度学习代码库，有诸多预训练模型及其checkpoints。



让未来的你成为你的灯塔，而不是KPI

于静

邮箱: yujing02@iie.ac.cn

研究组主页: <https://mmlab-iie.github.io/>

知乎专栏: https://www.zhihu.com/column/c_1284803871596797952

研究组主页



知乎专栏



中国科学院 信息工程研究所
INSTITUTE OF INFORMATION ENGINEERING, CAS



中国科学院大学
University of Chinese Academy of Sciences