

【Café速递】米黎：视觉关系检测概述及 CVPR 投稿经历分享

核心提示：计算机视觉领域的视觉关系检测任务概述， CVPR 投稿经历和相应论文思路，以及科研心得分享。

主持：张硕 摄影：程响 摄像：黄文哲 文字：程响

>>>人物名片

米黎，武汉大学遥感信息工程学院 2018 级硕士研究生，模式识别与智能系统专业，师从陈震中教授。以第一作者/第一学生作者身份在 ISPRS Journal of Photogrammetry and Remote Sensing 和 IEEE Transactions on Geoscience and Remote Sensing 发表论文各 1 篇，以第一作者身份在 CCF-A 类会议 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2020) 发表论文 1 篇，曾获得“华为杯”第十六届中国研究生数学建模比赛一等奖。

>>>报告现场

9 月 18 日晚上 7 点，武汉大学遥感信息工程学院 2018 级硕士研究生米黎做客 GeoScience Café 第 266 期经验分享活动。米黎学姐就自身研究领域、论文思路及投稿经历和线上线下的同学们进行分享，让听众们受益匪浅。（如图 1）



图 1 米黎学姐作精彩报告

视觉关系检测任务概述


视觉关系检测是一个新兴的图像视频理解任务，其目标是为单张图像或一

段视频预测一系列三元组。该三元组通常被结构化地描述为<subject-predicate-object>，即包含某种关系的主体，描述该关系的谓词，以及关系的客体。一般来讲，视觉关系检测任务分为两大步骤：首先进行目标检测，定位感兴趣的对象；然后进行关系预测，对检测出的目标之间进行关系进行理解。

什么是目标对象，视觉关系又有哪几种，米黎学姐从文字和数学两个层面阐述了视觉关系检测任务的定义。目标既可以是某种物体，如电脑、人、交通工具等，还可以指物体的一部分，比如人的手、眼睛等等；关系则分三大类：空间关系、从属关系和语义关系。

Introduction
Visual Relationship Detection (VRD)

Given an image or a video, **Visual Relationship Detection** aims to provide several triplets, shown as <subject-predicate-object>.



- ✓ Input: An image/ A video.
- ✓ Output: Triplets, <subject-predicate-object>.
- ✓ Goals: ① Localize the objects; ② Determine the predicates

4

图2 视觉关系检测任务的文字定义

Introduction
Problem Formulation

$$P(r) = P(p|s, o)P(s|b_s)P(o|b_o)$$

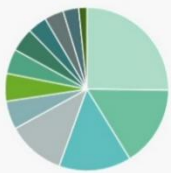
Predicate Predication Object Detection

R Relationship set
 $R = \{r(s, p, o) | s, o \in O, p \in P\}$


O Objects set

P Predicate set

Objects



Relationships



5

图3 视觉关系检测任务的数学定义

为了进一步明确视觉关系检测任务和其他任务的区别与联系，米黎学姐解释

了视觉关系检测的两个“桥梁”作用。

其一，视觉关系检测定义在计算机视觉和自然语言处理交叉领域，是沟通计算机视觉和自然语言处理的“桥梁”之一。视觉关系检测和其他一些如自动生成图像的文本描述的任务一样，同时涉及到对于视觉场景的理解和语言表达，是视觉图像和语言的中间领域。

其二，仅针对图像理解而言，视觉关系检测还是沟通低层理解任务（Low-level Understanding）和高层理解任务（High-level Understanding）的“桥梁”。作为对场景中实体关系的理解，视觉关系检测在目标检测识别任务和视觉语言任务之间起到了一种连接作用。

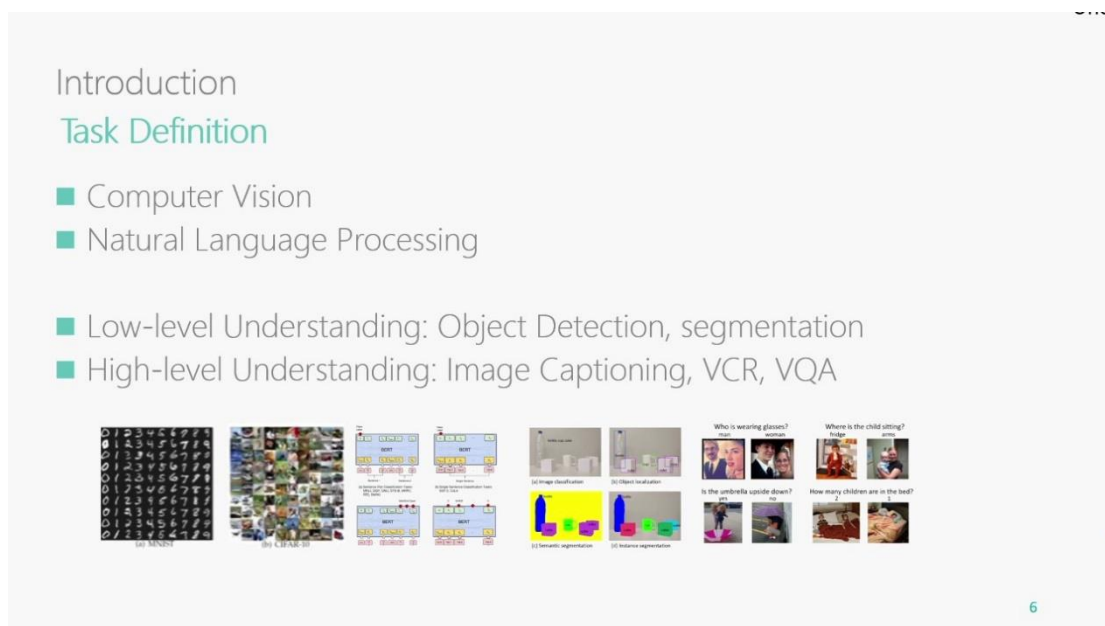


图 4 视觉关系检测任务和其他任务的联系

视觉关系检测领域内的关键问题和解决思路

接着，米黎学姐总结了该领域的 4 点关键性的挑战：

一是搜索空间爆炸。学姐借由领域内的一篇经典文章来引出解决思路。

二是标签不平衡和不完整。学姐指出，大量视觉关系的出现频率呈现长尾分布，同时图片中一些未标注的区域也可能含有目标。

三、四则分别是对关系的建模能力和模型的推理能力。

最后，米黎学姐归纳了 5 种解决思路：

一是基于特征的方法。米黎学姐总结到，早期的方法主要是通过改善特征提取过程来提升模型能力，本质是尽可能融合场景中实体的视觉特征、语义特征和空间特征。

二是基于“排序”(Rank)的方法。该方法的提出其实是为了解决标注不完全的问题。由于标注不完全,没有出现在标注数据中的关系预测结果不能直接被判定为“错误”,即标签并不能绝对化成0/1。该方法引入“排序机制”进行平滑,为正样本赋予“更高的排序”,负样本赋予“更低的排序”,以此弱化标注数据不完整对训练过程的影响。

三是基于“转换”(Translation)的方法。通过运算,将检测从“特征空间”转换到“关系空间”。在关系空间中,关系的发出和承受者都是点,两者间的关系则可以表示为一个向量。

四是基于循环神经网络的方法。米黎学姐分析到,视觉关系三元组就像一句话中的主谓宾一样,因此可以借鉴自然语言处理领域中一些方法来输出关系三元组。

五是基于图的方法。米黎学姐认为,图结构在实体关系表达上具有一定的优势。比如,边的有无可以表示是否存在关系,边的权重又可以表达关系的强弱或者其他特性。这类方法一般以一种特定的方式定义图结构并利用图网络进行关系推理。

论文分享

米黎学姐和大家分享了她被 CVPR2020 接收的文章。学姐通过三个通俗易懂的例子,进一步分析了现有的基于图的视觉关系检测方法存在的问题。

米黎学姐指出,某些关系三元组间经常共同出现在一个视觉场景中,而现有物体层面的图结构无法显式表达这种共同出现的频率关系。比如“汽车在路上”和“人骑自行车”共同出现的频率比“人骑自行车”和“大象在草地上”共同出现的频率高。

同时,仅基于空间邻近性建立的图结构还存在“缺失边”和“冗余边”的问题。米黎学姐又举了相应的两个例子:人放风筝,风筝和人在空间上相隔很远,导致无法在图结构中建立联系;站得很近两个人,一个人身上的衣服可能和另一个人之间可能达到建立“边”的阈值,但是他们实际上没有任何关系。

学姐就以上情况提炼出了问题的本质,同时介绍她解决以上问题的思路:通过建立关系层面的图结构、加入先验信息和引入“注意力机制”。

之后,米黎学姐就实验流程、模型性能评价指标、数据集、消融实验、以及和其他 SOTA 方法的比较等论文方面的细节和同学们展开分享。

投稿经历

米黎学姐先就论文发表周期的问题,谈论了自己的看法。“3 或 6 个月是否可

以看作投稿计算机视觉顶会的周期”，米黎学姐认为，做研究其实是有一定节奏的，初学者应该尽量去探究自己的节奏，而不是被3个月、6个月的周期所限制。

随后，米黎学姐就自身科研经历和大家分享了几点心得。初入科研接手新任务后，有两点很重要，一是充分的现状调研，二是冷静的思考能力。谈及第一点，米黎学姐认为，首先现状调研要充分，如果调研不充分可能与其他研究重复或相似，将造成时间和精力浪费；其次调研的过程中要归纳分类，触类旁通；最后，调研需要持续更新，计算机视觉领域的发展很快，边做还要边追踪新的思路，比如一些顶会的最新成果。

至于第二点，学姐谈到了应该理性对待性能提升在评价计算机视觉方法有效性中的地位。面对一个新的想法，学姐认为至少有两点需要冷静思考，一是思考问题的重要性，二是思考方法的必要性。第一，以问题为导向，思考自己到底要解决什么问题，这个问题是否重要，是否是领域内的一些基础性问题。学姐以分割领域，上下文信息和边缘信息不能兼得为例，解释了立足于领域内的基本问题做研究的思路。第二，方法是否必要。方法的有效性强调对性能的提升，方法的必要性则侧重对问题的解决。

实验设计上，学姐分享了三点：理论和实践相辅，框架和细节相成，以及站在读者的角度。第一，对于理论上的改进，一定要设计实验去证明它，这种证明形式可以是非常多样的，比如可视化中间结果，比如在子任务和小数据集上的实验。学姐认为，哪怕是一些很简单的实验，只要能证明理论，都是必要的。理论必须和实践紧密耦合才会有说服力的；第二，设计实验时首先要有大的框架，再补充一些细节。每个细节的实验可能都对应着想着重阐述的某个点，细节实验越充分，对方法的阐述也就越完善。第三，做实验设计或写论文的过程中，多站在读者的角度思考，考虑易于读者理解的表达方式和读者关切的实验设计。

最后，米黎学姐分享了论文写作过程的四个要点：易懂、准确、连贯、逻辑。论文中的图表力求直观易懂，涉及的名词术语力求专业准确，语言连贯，逻辑清晰。



图5 观众认真听报告



图6 米黎学姐（左二）与部分听众、Geoscience Café 团队成员合影留念

GeoScience Café 以“谈笑间成就梦想”为目标，于每周五晚 7:00 在实验室四楼休闲厅，邀请 1-4 位嘉宾，为大家带来学术报告或经验分享。报告内容包括摄影测量与遥感、地理信息系统、导航与定位服务等研究方向，听众可在报告结束后向嘉宾提问、与嘉宾交流探讨，同时每学期还会举办 2 期人文类讲座和 2 场导师信息分享会。每期报告会根据嘉宾意愿在 B 站开设直播，使不能来到现场的听

众同步参与。报告 PPT 和视频会在征得嘉宾同意的情况下在 qq 群和 B 站上发布。

更多精彩内容（讲座预告、讲座回顾、报告 PPT、报告视频）敬请通过以下方式获取：



QQ群



微信公众号



B站直播