

摘要

场景图像作为众多图像内容的一种，对应人们生活环境的写照。随着网络的发展和社交平台的兴起，越来越多的人通过移动设备记录周围生活的环境和旅途中的风景。通过理解场景图像内容，从而实现图像的分类整理、语义标注、地标检测、索引排序等便成为了当下互联网产业亟待解决的任务。同时，机器人技术（如无人驾驶技术）的进步，也要求计算机对场景图像的理解更加智能和高效。

本文针对场景图像，特别是室外场景图像的建模与理解展开研究。场景建模作为计算机视觉领域的基础研究方向之一，近年来得到了广泛的研究。纵观已有方法，大多数将场景建模等价于分类问题：基于图像底层特征对不同场景的类别进行描述，而忽略了场景的结构信息。认知心理学的研究表明，结构信息是人类进行场景理解的重要线索之一，不仅有助于人们快速识别场景（20 毫秒），还有助于人们识别场景中的物体。因此，本文将从场景结构的角度出发，对场景图像的建模进行深入探讨，并提出了两种新颖的场景建模方式。

I. 多尺度场景语义分割模型

图像理解的精度和效率往往与图像尺度有关。例如识别场景中的山峰、河流，可使用分辨率较低的图像，而识别场景中的汽车、行人，则需使用高分辨率的图像。因此，本文首先提出一种多尺度模型，用于弱监督模式下的场景语义分割。语义分割包含图像分割和目标识别两个任务，通过将图像分割成小块，并识别出每个图像块的语义信息（如天空，道路），最终得到一幅具有逐像素语义标注的图像。而弱监督模式的语义分割，表示训练图像的语义标注只描述场景中存在哪些物体，而不指出标注所对应的具体图像区域。不同于传统的语义分割方法，本文提出的多尺度语义分割模型可以在不同尺度下对图像进行分割和理解，并挖掘不同尺度间场景的上下文信息。此外，为了刻画场景结构的时空相关性，本文在多尺度模型中还引入两项先验知识：（1）物体空间位置先验，用于描述物体可能出现在场景中的位置，如“天空”经常出现在图像顶端；（2）物体相对位置先验，用于描述物体间的相对位置关系，如“汽车”总会出现在“道路”之上。最后，本文提出一种自顶向下和自底向上的迭代方法对多尺度语义分割模型进行学习。在公共数据集的实验结果表明，本文提出的多尺度语义分割模型不仅优于所有现有的弱监督语义分割算法，还优于某些流行的全监督语义分割算法。

II. 层次化空间拼贴模型

人类视觉系统中由简到繁的认知机制，启发本文采用层次化方法对场景图像进行

描述。同时，由于场景图像的拍摄角度、景深的不同，以及场景中物体的形状差异、位置差异等，场景结构通常会非常丰富多变，层次化模型以其强大的组合和重构能力可以为复杂场景结构提供一个有效的表达。因此，本文提出层次化空间拼贴模型（Hierarchical Space Tiling，简称 HST），通过简单几何基元（如矩形、正方形）对场景结构空间进行量化，通过几何基元的组合实现场景结构的解析，通过模型的学习获得特定场景类的结构统计特性。此外，本文在层次化空间拼贴模型中引入场景属性，以此对场景图像的语义信息进行刻画。通过建立关联矩阵，本文将场景属性与场景结构联系在一起。层次化空间拼贴模型采用弱监督的学习模式：即给定训练图像及语义标签，其中语义标签只说明场景中存在哪些属性，通过模型的学习将场景属性定位到图像的具体区域中。实验阶段，本文首先采用率失真曲线对模型的表达力进行定量分析，结果表明相较于传统场景建模方法，层次化空间拼贴模型能够更加简洁高效地对场景结构进行建模。接着，本文将层次化空间拼贴模型应用于多种实际问题，如场景分类、场景属性识别、属性定位、场景语义分割，在公共数据集的实验结果表明，本文提出的模型不仅能够更加准确地完成场景理解任务，还具有更高效的处理效率。

关键词：场景结构，场景属性，层次化空间拼贴模型，多尺度模型，弱监督学习