

面向大尺度遥感图像的地面物体定位方法

孙玉玺¹ 叶允明¹ Yifang Ban² 李旭涛¹

1 哈尔滨工业大学（深圳）

2 瑞典皇家理工学院

摘要：在大尺度遥感图像中检测和检索地面对象，对于智慧城市管理和可持续城市发展非常重要。我们提出了一个新的遥感图像视觉定位问题。视觉定位主要是指通过自然语言表达，在图像中找到目标对象（以边界框的形式标记）。该任务已经在传统计算机视觉中存在。然而，现有的方法主要是针对自然图像而不是遥感图像。与自然图像相比，遥感图像通常是大尺度场景的，并且包含地理空间信息，如经度和纬度。现有的方法忽略了遥感图像的这些特点，从而无法准确地定位到地面目标对象。针对这个问题，我们提出了一种新的视觉定位方法，称为 GeoVG。该方法由语言编码器、图像编码器和融合模块组成。语言编码器用于学习数字地理空间关系并将复杂的文本表达转换为地理空间关系图。图像编码器通过自适应区域注意力机制，学习大尺度遥感图像的视觉特征。融合模块用于融合文本和图像特征以进行视觉定位。我们将 GeoVG 与最先进的方法进行对比。实验结果表明，我们的 GeoVG 方法要优于目前最先进的视觉定位方法。