# 摘要

视频分割旨在为视频中的目标逐帧生成像素级掩码，从而构建精细的目标表征，支撑后续的视觉感知与推理任务，是计算机视觉领域的重要研究问题之一。视频分割不仅可以为动作识别、视频摘要和目标跟踪等计算机视觉与人工智能研究提供关键先验信息，也是视频编辑、自动驾驶和机器人感知等应用的基础之一，兼具重要的科学研究意义与应用价值。

近年来，得益于深度学习与计算机视觉技术的快速发展，视频分割研究已取得显著进展。然而，复杂开放场景下采集的视频中，同类物体外观相似难区分、语义类别开放难判别、分割任务定义各异难统一，导致现有视频分割方法在鲁棒性、准确性、适应性上存在明显缺陷。首先，同一视频往往包含多个外观相似的同类物体，增加了对特定目标进行跨帧持续精准分割的难度。其次，有限的分割训练数据集限制了视频分割模型对开放语义的判别能力，难以实现开放词汇分割。最后，不同视频分割任务定义和优化目标各异，导致难以设计跨任务适用的视频分割方法。

视频分割模型往往逐帧对视频进行处理以得到分割掩码序列。因此，其可以利用已知的分割结果对当前帧的分割进行辅助。由于已知的掩码序列包含了很强的物体、语义和任务先验，从已知掩码序列中建模和提取有益的时空上下文信息成为解决上述难点问题的有效途径。本文以“时空上下文建模”为核心思路，从三个层面对时空上下文信息进行建模和提取，分别提升了视频物体分割、开放词汇分割和跨任务分割性能。本文的主要贡献如下：

第一，针对同类物体外观相似难区分问题，本文提出了基于物体上下文聚合的视频物体分割方法。该方法将已知掩码序列中目标物体的特征与标签信息建模为物体上下文，利用物体上下文提升对该目标物体的跨帧持续分割精度。首先，物体特征聚合模块在帧间进行由粗到精的多尺度匹配以建立鲁棒的帧间对应关系，实现对目标物体特征的精准聚合。其次，物体标签聚合模块利用非对称的标签聚合策略对邻近帧和首帧内的物体标签信息进行处理，实现了噪声抑制和可靠的物体标签信息聚合。实验表明该方法所提取的物体上下文能够有效提升视频分割模型对外观相似物体的判别能力，在多个视频物体分割数据集上取得了同期领先的性能。

第二，针对语义类别开放难判别问题，本文提出了基于语义上下文增强的开放词汇视频分割方法。该方法将对目标词汇具有判别力的时空特征建模为语义上下文，利用语义上下文增强掩码区域的视觉特征，使其与预训练语言-图像对比模型输出的文本特征对齐，提升视频分割模型对开放词汇语义的判别能力。首先，空间适配模块在保

持原始特征分布的前提下增强空间判别力。其次，语义上下文增强模块通过挖掘对目标词汇有判别力的时空区域，使掩码特征与预训练特征对齐。在多个开放词汇分割任务上的实验表明，该方法仅使用单一训练集即可在多个测试集上取得优于同期同类方法的性能。

第三，针对分割任务定义各异难统一问题，本文提出了基于任务上下文挖掘的跨任务视频分割方法。该方法以特定任务模型输出的掩码序列为基础，将掩码序列内含的任务先验建模为任务上下文，利用任务上下文指导分割基座模型完成特定的分割任务。该方法首先利用特定任务模型生成初始掩码，进而引入一种静态-动态结合的质量评估策略对分割质量进行评估，最终挖掘出可靠的初始掩码来提取任务上下文。实验结果表明该方法具有跨任务适应性，在隐藏物体视频分割、医学视频分割和视频物体分割等多种视频分割任务中，相比同期方法取得了显著的性能提升。

第四，为验证所提方法在复杂真实场景中的有效性，本文在大规模视频监控场景中开展了关键目标自动跟踪与分割验证。为应对复杂监控场景中目标密集且外观相似、已有算法鲁棒性差等问题，构建的关键目标自动跟踪与分割系统集成了本文提出的三种方法。在大规模公安监控视频上的应用结果表明，本文所提的方法能够有效提升复杂场景中关键目标的持续跟踪与分割精度，验证了方法的有效性。

综上所述，本文以“时空上下文建模”为核心思路，通过从已知的视频分割结果中提取物体、语义和任务先验，提升了视频物体分割、开放词汇分割、跨任务分割三类视频分割任务的性能。本文的研究也为视频分割方法的应用落地及其他视觉任务研究提供了有价值的参考与启示。

**关键词：**视频分割，时空上下文，视频物体分割，开放词汇视频分割，跨任务视频分割