 



高斯时间感知网络(GATN)用于动作定位

# 摘要

*在视频理解中，暂时(temporally)定位视频中的动作是一个基本的挑战。大多数现有的方法往往从图像对象检测中得到启发，并扩展了SSD和更快的R-CNN等先进技术，以在ID序列中动作的暂时定位(temporal locations of action)。然而，由于预先确定的时间尺度(temporal scalse)的设计,结果可能存在鲁棒性(robustness)，这忽略了动作的时间结构，限制了检测复杂变化动作的实用性。在本文中，我们通过提出引入高斯核(Gasssian Kenel)来动态优化每个动作建议的时间尺度来解决这个问题。具体来说，我们提出了高斯时间感知网络(GTAN)-一种新的架构，它新颖的将时间结构的利用集成到一个单阶段的动作定位框架中。从技术上讲，GATN通过学习一组高斯核建模时间结构我们在未剪辑的长视频中解决时间动作定位问题。这一点很重要，因为真实应用程序中的视频通常不受约束，包含多个动作实例以及背景场景或其他活动的视频内容。为了解决这个具有挑战性的问题，我们通过三个基于片段的3D ConvNet，利用深度网络在时间动作定位中的有效性：（1）提议网络识别可能包含动作的长视频中的候选片段；（2） 分类网络学习一对多动作分类模型，作为定位网络的初始化；以及（3）定位网络微调学习的分类网络以定位每个动作实例。我们提出了一个新的损失函数的定位网络显式考虑时间重叠，并实现高的时间定位精度。最后，在预测过程中只使用建议网络和定位网络。在两个大型基准测试中，我们的方法与其他最先进的系统相比取得了显著的优越性能：在Mexaction 2上，mAP从1.7%增加到7.4%，在THUMOS 2014上从15.0%增加到19.0%。*

# 1. 介绍