



研究生学习工作周报

院 系	人工智能学院
-----	--------

专 业	电子信息
-----	------

姓 名	余依函
-----	-----

学 号	231226006052
-----	--------------

导 师	周静 张俊驰
-----	--------

周报日期	2023 年 10 月 14 日
------	------------------

摘要

1. R2VOS 和 PTP 的结合使用探索
2. 论文阅读

目录

摘要	I
第一章 学习工作总结	1
第二章 第七周总结及第六周学习计划	3
参考	4

第一章 学习工作总结

1.1 R2VOS 和 PTP 的结合使用探索

本周对 PTP 的基本架构进行了一个了解，同时 PTP 可以应用于不同的视觉语言下游任务的预训练中从而减轻成本加快速度。用于新 pretext task 的损失函数如下所示。

$$\mathcal{L}_{\text{PTP}}(\theta) = -\mathbb{E}_{\mathbf{y} \sim D} \left[\sum_{t=1}^T \log P_{\theta}(\mathbf{y}_t | \mathbf{y}_{<t}) \right]$$

目前还在看 PTP 的实验部分和 R2VOS 的模型部分。

1.2 Two-shot Video Object Segmentation 论文阅读

本周对 Two-shot Video Object Segmentation 论文阅读进行阅读，此文章主要研究的是提高 VOS 的效率，本文提出了一个在同一个数据集上训练两次的方法，通过生成伪标签来避免过多帧注释的大量数据量，同时通过不断生成置信度更高的伪标签的方式更新伪标签库。将此方法和 PTP 进行一个比较讨论双方的优劣。

Two-shot Video Object Segmentation

Meta Data

Title	Two-shot Video Object Segmentation
Journal	
Authors	Kun Yan; Xiao Li; Fangyun Wei; Jinglu Wang; Chenbin Zhang; Ping Wang; Yan Lu
Pub. date	2023-03-21
期刊标签	
DOI	10.48550/arXiv.2303.12078
附件	Yan et al 2023 Two-shot Video Object Segmentation.pdf

研究背景 & 基础 & 解决方法

研究背景

密集注释的视频上训练昂贵耗时

基础

基本思想是在训练期间为未标记的帧生成伪标签，并结合标记数据和伪标记数据来优化模型。

解决方法

1. 以半监督的方式在稀疏注释的视频上预训练 VOS 模型，第一帧始终是有标签的帧。
2. 采用预训练的 VOS 模型为所有未标记的帧生成伪标签，随后将其存储在伪标签库中。
3. 标记数据和伪标记数据上重新训练 VOS 模型，对第一帧没有任何限制。

“Previous Methods (Densely Annotated Videos)” (Yan 等, 2023, p. 1)

图 1.1 部分论文阅读笔记

第二章 第八周总结及第九周学习计划

序号	上周任务	完成情况及备注	本周任务
1	R2VOS 与 PTP 是否可以结合使用	完成，尝试把损失函数直接加入到 R2VOS 的损失函数中去	R2VOS 代码继续阅读
2	Two-shot Video Object Segmentation	完成	阅读一篇论文
3			
备注：			

参考

1. <https://doi.org/10.48550/arXiv.2303.12078>