课程报告：视频流中物品的零样本交互式自动化消除

项目简要介绍

本次课程项目旨在

深入探索视频对象去除与修复的自动化方法，以显著提升处理效率和修复质量。我们采用了一系列创新性技术，其中包括自动分割技术 SAM、掩膜传播技术 AOT 以及视频修复技术 E2FGVI，以构建一个综合性解决方案。

该方案专注于高效地从视频流中消除特定对象，实现修复并保持连贯性。通过在这些技术之间巧妙地结合，我们的目标是实现一个快速、自动化且高质量的视频对象处理流程。

项目标题：

视频流中物品的零样本交互式自动化消除-SAM、AOT 和 E2FGVI 的综合应用。

硬件平台：

Python：3.8 (Ubuntu 20.04)

CUDA：11.8

GPU：NVIDIA V100-32GB (1 个)

CPU：6 vCPU Intel(R) Xeon(R) Gold 6130 CPU @ 2.10GHz

内存：25GB

搭建指令：

在开始项目前，请按照以下指令搭建所需的开发环境和依赖项：

创建并激活虚拟环境：

conda create -n ikunnet python=3.10

conda activate ikunnet

安装 Segment Anything 和其他必要库：

pip install git+https://github.com/facebookresearch/segment-anything.git

pip install opencv-python pycocotools Pillow scikit-image tqdm

安装 PyTorch 和相关库：

conda install pytorch torchvision torchaudio pytorch-cuda=11.7 -c pytorch -c nvidia

克隆并安装 Pytorch-Correlation-extension：

git clone https://github.com/ClementPinard/Pytorch-Correlation-extension.git

cd Pytorch-Correlation-extension

python setup.py install

cd -

安装 mmcv-full：

pip install mmcv-full -f https://download.openmmlab.com/mmcv/dist/cuda117/torch2.0/index.html

创建权重文件夹并下载预训练模型：

mkdir weights

cd weights

wget https://dl.fbaipublicfiles.com/segment_anything/sam_vit_h_4b8939.pth

手动下载预训练模型：

https://drive.google.com/file/d/10wGdKSUOie0XmCr8SQ2A2FeDe-mfn5w3/view

https://drive.google.com/file/d/1tNJMTJ2gmWdIXJoHVi5-H504uImUiJW9/view

以上指令能够搭建所需的开发环境，并获取必要的预训练模型和库。完成上述步骤后，可以在搭建好的环境中顺利运行项目代码。

成员工作及贡献比及考勤

在本项目中，我们团队成员分工如下：

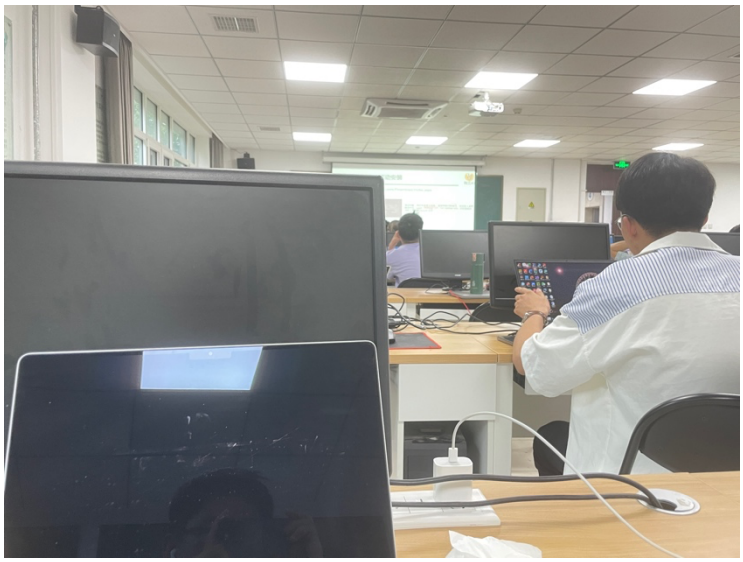
1. 黄键楠：负责项目的技术方案设计和实现的大部分工作，占比 50%。黄键楠承担了项目中的主要研究和技术实现，涵盖了大部分的创意和代码编写。此外，黄键楠还负责了 AOT 模型和 E2FGVI 模型的研究和集成，为项目的整体成功做出了重要贡献。

第一次考勤黄键楠的出勤如下图所示：

第二次考勤：黄键楠位于 401 教室第五排从右边数第三个位置。

2. 王德阳：负责项目的想法和文档编写，以及部分代码的实现，占比 50%。王德阳在项目中提供了部分想法和思路，为项目推进助力。此外，王德阳负责了项目的文档编写工作，确保了项目的整体文档质量和可理解性。王德阳还参与了 SAM 模型的研究和实现，为项目的技术深度提供了支持。

第一次考勤王德阳的出勤如下图所示：



第二次考勤：王德阳 401 外第八排从左往右第 7 个位置。

我们项目团队成员分工明确，每个成员在 SAM、AOT 和 E2FGVI 等技术的研究、实现和优化方面都做出了重要贡献。团队的合作精神和协同工作是项目取得成功的关键。通过充分的讨论和合作，我们能够将不同的技术模块有机地结合在一起，成功实现了对象去除和修复处理。

项目背景

在当今数字技术不断创新的时代，图像和视频处理领域正经历多项突破性的技术进展。自动分割、掩膜传播和视频修复等技术在数字媒体处理领域引起了广泛关注，为影视制作创作和内容控制提供了新的可能性。近期，综艺节目《奔跑吧》中使用数字技术移除特定对象的行为引起了广泛关注，这个创新方式揭示了数字图像处理方面相关技术的巨大潜力。

综艺节目《奔跑吧》中的数字处理实例，为我们提供了一种创新的思路，即利用自动分割、掩膜传播和视频修复等技术，从视频中移除特定的对象，实现内容的修复和再创作。这一实例展示了数字媒体技术在实际娱乐制作中的应用潜力，同时也激发了对于这些技术更广泛应用的兴趣。

通过将这些技术相互结合，我们的项目旨在探索如何在短视频中实现对象的去除和修复，从而为数字媒体处理领域带来更多创新和可能性。这也为未来数字媒体制作提供了新的思路，同时也为技术在艺术和娱乐领域的应用带来了新的启示。

项目目的

提升处理效率，保持连续性和一致性： 我们的项目旨在提高对象去除和修复的处理效率，同时保持视频的连续性和一致性。通过自动化处理流程，我们能够在短时间内完成复杂的去除和修复任务，减少了人工处理的时间和成本。

提高修复质量，探索创作可能性： 我们的目标不仅仅是快速处理，还追求高质量的修复结果。通过综合应用 SAM、AOT 和 E2FGVI 等技术，我们能够实现更加真实和自然的修复效果，为数字媒体内容的创作提供更多可能性。

探索数字媒体处理领域的创新应用： 我们的项目不仅限于一种应用场景，而是在数字媒体处理领域探索了新的可能性。这种综合应用的方法可以在影视制作、广告、数字艺术等领域中发挥重要作用，为创作者和制作者提供更多创新手段。

原理分析

SAM（Automatic Segmentation with SAM）：

SAM 是 META 最新的 SOTA 视觉分割大模型，其在包含 1100 万张授权图像上的 10 亿多个掩码上进行训练。该模型的设计和训练具有可迁移性，因此它可以在新的图像分布和任务中进行很好的无训练迁移。大量任务中的实验结果表示它的零样本性能令人印象深刻，通常可与之前的完全监督结果相媲美，甚至更胜一筹。

AOT（Mask Propagation）：

掩膜传播技术用于将一个图像或视频帧中的掩膜（标记出需要处理的区域）传播到其他帧中，以确保在多个帧之间保持对象位置和形态的一致性。这在视频处理中尤其重要，因为去除一个对象可能涉及到多个连续帧的修复。AOT 是一种高效的掩膜传播算法。

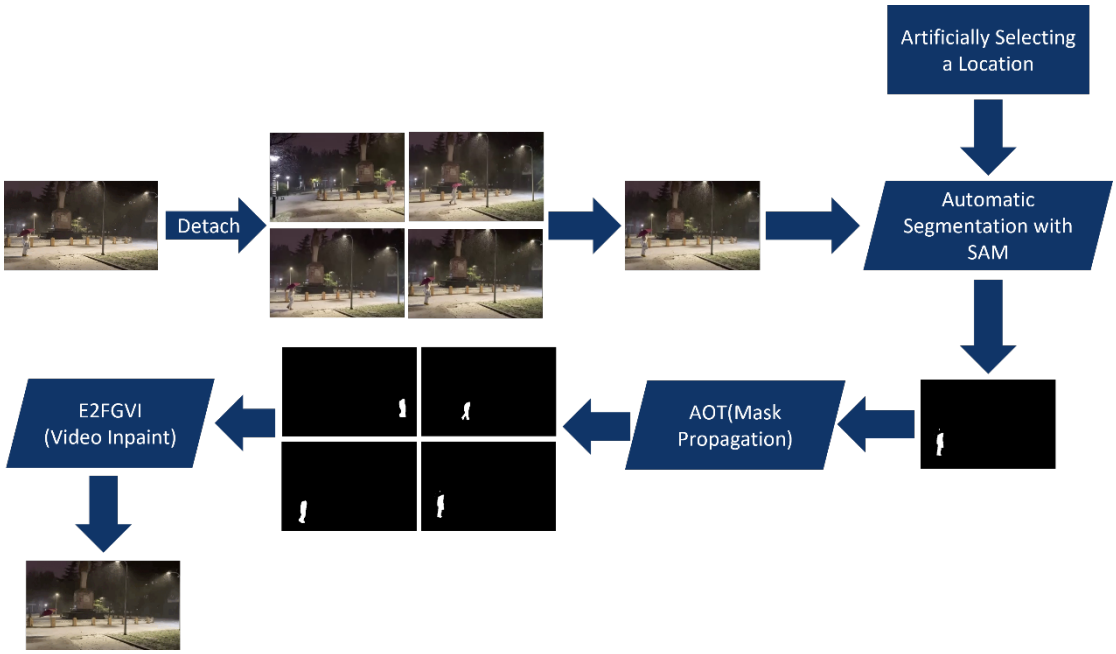
E2FGVI（Video Inpaint）：

视频绘制的目标是用空间和时间上都一致的内容填补给定视频序列中的缺失区域，如去除不想要的物体和视频修复。E2FGVI 采用光流引导的方式，提出了端到端的流程引导视频内画框架。这三个模块与以往基于流量的方法的三个阶段相对应，但可以共同优化，从而实现更高效、更有效的绘制过程。

项目流程

我们的项目程序大致如下：

1. 将视频分解为多个视频帧。
2. 获取第一帧图像以及外部指定的坐标。
3. 使用 SAM 获得第一帧掩膜。
4. 利用 AOT 在每一帧之间传播掩膜，以保持对象一致性。
5. 使用 E2FGVI 进行视频修复，将被去除对象的背景修复得更加真实自然。



项目优势

交互式用户指定消除对象： 传统的自动化对象去除方法往往缺乏用户参与，而我们的项目允许用户通过交互方式指定需要去除的对象。这不仅增加了用户的参与度和控制感，还保证了最终修复结果符合用户的预期。

快速处理： 通过综合应用 SAM、AOT 和 E2FGVI，我们实现了短时间内完成对象的去除和修复。这种高效率处理方式在现代数字媒体制作中尤为重要，能够节省制作时间和资源。

端到端处理： 我们的项目实现了从自动分割到修复的端到端处理流程。这确保了各个处理步骤之间的一致性和连贯性，避免了在不同模块之间进行繁琐的数据转换和调整。

不需要训练零样本推理： 传统的对象去除方法通常需要在新任务上进行重新训练，耗费时间和资源。而我们的项目利用了 SAM 的零样本迁移能力，避免了重新训练的过程，极大地提高了处理效率。

封装性良好： 项目将复杂的技术模块整合成一个封装完善的处理流程。这使得用户可以轻松地应用该方法，无需深入了解每个技术细节，从而降低了使用门槛。

可扩展性和应用前景

我们的项目不仅仅停留在综艺节目中的特定案例，而是具有更加广泛的应用前景，能够为多个领域带来创新解决方案。

此外，随着技术的不断发展，我们有望结合更多先进技术，进一步扩展这种方法的应用领域，如虚拟现实、增强现实等。随着这些技术的融合，我们的方法将在更多领域创造更多可能性，为各行各业的创新注入源源不断的活力。

数据集和挑战

为了训练和测试我们的方法，我们采用了多种数据集，这些数据集涵盖了不同场景、光照条件和对象种类的视频素材。然而，与之相对应的是一些挑战，例如：

复杂背景下的掩膜传播： 在复杂背景情况下，掩膜的传播可能会遇到困难，导致对象的去除和修复效果不如预期。针对这一挑战，我们需要进一步研究和优化掩膜传播算法，以应对不同背景条件下的情况。

对象与背景的自然融合： 修复后的对象需要与周围的背景自然融合，以实现更真实的效果。这需要在技术层面上加强对象与背景的无缝融合能力，以免修复后的区域显得突兀或不协调。

这些挑战鼓励我们继续深入研究和创新，以不断提升我们的方法在各种情况下的适应性和效果。

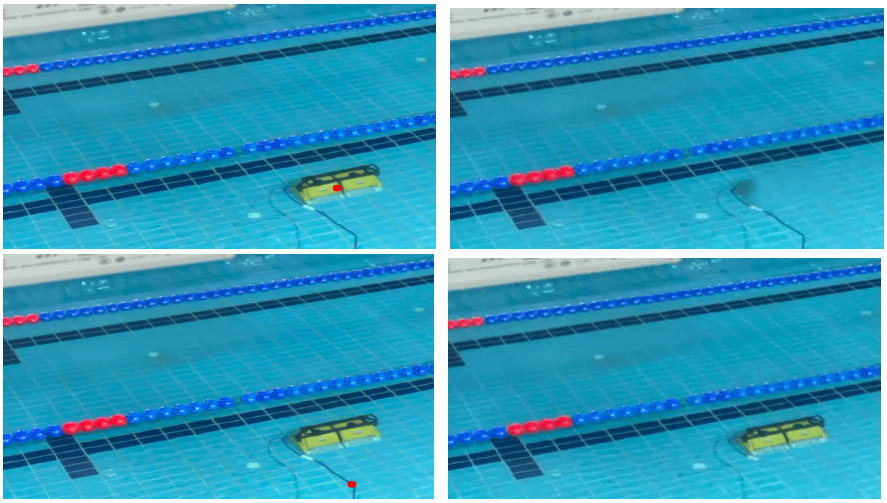
实验结果

我们进行了一系列实验，使用不同类型的视频素材，并进行了对象的去除和修复。实验结果显示，通过 SAM、AOT 和 E2FGVI 的综合应用，我们能够在保持视频连贯性和一致性的同时，高效地完成对象去除和修复，达到预期的效果。

例子 1：

例子 2：

例子 3：



结论

通过本项目，我们综合应用了 SAM、AOT 和 E2FGVI 等技术，实现了在视频流中对物品进行零样本交互式自动化消除和修复。该项目不仅在数字媒体处理领域探索了新的可能性，也为未来数字媒体制作提供了新的思路，同时还为技术在艺术和娱乐领域的应用带来了新的启示。

致谢

我们衷心感谢以下开源项目，为本项目的成功实现提供了重要的技术支持和帮助：

AOT Benchmark: 我们在项目中使用了 AOT 算法，在这个项目中得到了关键的参考和指导。项目链接：GitHub - yoxu515/aot-benchmark: An efficient modular implementation of Associating Objects with Transformers for Video Object Segmentation in PyTorch

E2FGVI: 项目中我们也借鉴了 E2FGVI 技术，该项目为我们的视频修复步骤提供了重要的框架。项目链接：GitHub - MCG-NKU/E2FGVI: Official code for "Towards An End-to-End Framework for Flow-Guided Video Inpainting" (CVPR2022)

Segment Anything: 我们在项目中应用了 Segment Anything 技术，用于视频中对象的自动分割。该项目对我们的分割步骤起到了关键作用。项目链接：GitHub - facebookresearch/segment-

anything: The repository provides code for running inference with the SegmentAnything Model (SAM), links for downloading the trained model checkpoints, and example notebooks that show how to use the model.

感谢上述项目的贡献者们为开源社区做出的努力，以及他们的工作对我们项目的启发和帮助。在这些优秀的开源资源的支持下，我们的项目得以更加顺利地实现。谢谢！