

## 小白视频抠图神器---基于 OSVOS 算法的机器学习

李子晗 (Li Zihan)

东北大学 (Northeastern University)

e-mail:1339843636@qq.com

**Summary.** *Frame Cutout of a picture (FCoP) is one of the basic operations of video editing. Whether FCoP operation can be completed with high quality has a huge influence on the quality of video editing. At present, the popular video editing software on the market has more complicated operations and cannot complete batch operations of continuous frames. This idea intends to use the OSVOS algorithm to analyze the color environment of a single frame of a video screen to realize the functions of object recognition, background filling, and pixel propagation, so as to perform batch operations on consecutive frames and complete FCoP operation, so that Xiaobai, who loves video editing, can also easily complete related operations.*

Premiere 是当前市面上是一款非常常用的视频编辑软件，利用它可以实现视频的剪辑、设置图片的转场、修饰或创意等操作。不过，该软件的操作非常复杂，普通人不易掌握。尤其是在进行抠像操作的时候，操作者首先需要使用钢笔工具，沿着边缘选中需要抠除的物体。之后使用吸管工具，获取附近相近颜色对抠除部分进行背景填充。在单纯背景下，这种操作较为适用。不过，如果背景较为复杂时，就需要对周边背景进行分析，通过吸取、复制、模糊等操作填充缺失背景，这对操作者的技巧无疑是一种考验。有时候，为了追求视频的质量，操作者还会将每一帧的图片导入至 Photoshop 图片处理软件中进行精细处理，再传输回 Premiere 中。由此可见，目前视频剪辑的门槛是很高的。针对视频剪辑门槛较高的问题，本创意拟利用 OSVOS 算法，通过分析视频画面单帧色彩环境，实现物体识别、背景填充、像素传播功能，从而对连续帧进行批量操作，完成抠图作业，让爱好视频编辑与制作的小白，也能轻松完成抠图作业。

## 一、传统视频制作时的抠图作业



上面是两张对比图，利用 Photoshop 软件抠除了背景中的多余人物（如图中方框所示）。尽管这里移除的只是一个占比非常小的人影，但在移除填充后，背景中的栅栏出现缺失，并呈现出色差。可以推测，如果移除的是画面中的红衣男子，那么不和谐的现象将会更加明显，大大影响了图片的质量。想要获得无违和感的图片，就需要更高水平的剪辑师。所以，在传统视频剪辑软件抠图时，图片的质量直接由剪辑师的技术水平决定。

## 二、机器学习背景下的抠图作业



上述两张图片是使用了机器学习算法抠除了图片正中央的火烈鸟。可以看到，无论是操作的难易程度、还是背景填充的效果，机器学习背景下都是非常优异的。基本通过肉眼无法辨认出图片二中曾经存在过火烈鸟。而该种方式的图片处理效果，仅仅取决于算法的优劣，与使用者没有关系。哪怕是小白，也可以利用本软件进行高精度的抠图操作。

### 1. 物体识别

本设想的核心以 OSVOS 算法为基础，实现对目标中的物体进行分割。大大免去了小白逐帧抠图所带来的技术困难。对于一段视频，制作者本需要逐帧抠除所选择的物体，使用裁剪一点点选中物体边缘部分。而该算法在识别了第一帧的掩码后，可通过像素 RGB 通道值的突变来识别出物体边缘坐标位置，生成一个图像识别的模型，用于后续帧的识别。

OSVOS 算法是一种半监督式的物体分割算法，这意味着，小白只需要在第一帧图像上选择出希望识别的图像物体，然后算法根据这一物体进行标注与学习，进行分割操作，并生成识别模型。

### 2. 背景填充

在分割出待处理的图像后，帧画面中会出现因移除图片所带来的空白，专业的人士会使用相似画面替代，或者从别的帧中寻找固定坐标背景。而这一切同样可以交给计算机去自主学习。

此处采用的算法名为 Deep Flow-Guided Video Inpainting，该算法自动地对后续每一帧的图片进行分割与识别操作。通过对于不同帧中物体的移动与对背景的识别，算法自动学习每一帧中物体所遮挡住的背景，并在智能扣除图像后进行填充。用简单的话来说，机器将会识别未遮挡此处背景的画面，复制到该帧抠图后的空白处。

### 3. 像素传播

完成了一帧图片的处理，对于后续的每一帧都需要采用相似的操作。对于一个业余爱好者来说，肯定是没有这么多时间与精力做这么多复杂的操作。而机器又一次站了出来。

虽然所有帧都是独立处理的，但是可以在时间上连贯和稳定。假设对象不会在一帧和下一个帧之间变化太多。此类方法在整个方面平滑地调整单帧模型，用户可以在精度与速度上进行选择。每帧处理 181ms 可得 71.5 % 的准确率，而每帧处理 7.85s 时可达 79.5 % 的准确率。但是对于遮挡或者剧烈的移动，该算法会造成偏差，并且这个偏差会随着时间的推移而逐渐累积放大。虽然存在误差，但是对于一个普通的爱好者来说，却是绰绰有余了。

我的软件将上述算法封装到这个软件的后端，其前端设计成简约利于操作的界面。用户在进入软件后，首先导入自己希望剪辑的视频。之后，在选择栏输入想要删除或者更改的物体名称。算法将通过图像识别的方法，识别出对应物体。最后，用户点击确认后，软件直接将视频以 RGB 像素流的形式输入到算法中，就可以获得删除该物体后的视频。