项目编号：

**大学生科技创新项目**

**申 报 书**

**创新项目名称：基于XXXXX的**

**视频编辑系统**

**创新项目负责人：XXX**

**学校名称：杭州电子科技大学**

**申报日期：2020年12月01日**

**项目类别：个人项目□ 团队项目■**

**浙江省大学生科技创新活动计划暨新苗人才计划实施办公室 制**

填写说明

一、申报书要按照要求，逐项认真填写，填写内容必须实事求是，表达明确严谨。

二、格式要求：申报书中各项内容以Word文档格式填写，表格中的字体为小四号仿宋体，1.5倍行距；表格空间不足的，可以扩展或另附纸张；均用A4纸双面打印，于左侧装订成册。

三、申报书由所在学校领导审查、签署意见并加盖公章后，一式1份（原件），报送浙江省大学生科技创新活动计划暨新苗人才计划实施办公室。

1. **项目简介**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 项  目  概  况 | 项目名称 | | | 基于xxxx的xxxx系统 | | | | | | | | |
| 项目性质 | | | （）基础研究 （√）应用研究 | | | | | | | | |
| 项目来源 | | | （√）自主立题 （）教师指导选题 | | | | | | | | |
| 起止时间 | | | 自 2021 年 1 月 至 2022 年 5 月 | | | | | | | | |
| 项目状况 | | 1、研发阶段√ 2、中试阶段 3、批量（规模）生产（选项打√） | | | | | | | | | | |
| 申请人 | 姓名 |  | 性别 | | 男 | 出生年月 | 1999.01 | | | 入学年份 | | 2019 |
| 所在院系 | 计算机学院 | 学号 | |  | 联系电话 |  | | | 电子信箱 | |  |
| 项 目 组  主要成员 | | 姓名 | 性别 | | 年龄 | 学院 | 专业 | | 学号 | | 具体分工 | |
|  | 男 | |  | 计算机学院 | 计算机科学与技术 | | 191050045 | | 系统设计与实现 | |
|  | 男 | |  | 计算机学院 | 计算机技术 | | 192050146 | | 系统测试 | |
|  |  | |  |  |  | |  | |  | |
|  |  | |  |  |  | |  | |  | |
| 项  目  指  导  教  师 | | 姓名 | | |  | | 性 别 |  | | 出生年月 | |  |
| 主要研究方向 | | | 机器学习、计算机视觉、数据挖掘 | | | | | | | |
| 近三年获奖成果：国家级\_\_等奖\_\_\_项，省部级\_\_等奖\_\_\_项 | | | | | | | | | | |
| 近三年科研经费\_\_\_\_\_万元，年均\_\_\_\_\_\_万元 | | | | | | | | | | |
| 项  目  主  要  内  容  简  介 | | 在互联网+和大数据时代，人们拥有便捷的视频采集设备，使得视频数据无时不刻以指数级迅猛增长，而对视频进行编辑逐渐成为日常生活中的一部分。现有的视频编辑系统大多智能化程度较低、功能不完善、操作繁琐、耗费过多的人力物力。为此，本项目研发一套智能化程度高、功能较完善、操作简便的视频编辑系统。该系统以生成式目标分割为核心技术，提出低维嵌入相似度匹配方法获取优越的目标分割性能。具体来说，该技术以残差孪生网络为基础模型，同时设计相似度匹配模块并将其与生成式外观模块融合，最后通过上采样获得目标对应的掩膜。本系统能自动将目标从视频中分离，可根据不同的需求对视频进行智能化编辑，从而减少大量人力开销，且人工干预少、响应速度快。项目研发的视频编辑系统可广泛应用于电影制作、节目采编、视频会议、动画创作、AR(增强现实)游戏、短视频生成、网络直播等实际场景中。 | | | | | | | | | | |

**二、项目的研究目的及意义**

|  |
| --- |
| 1、申请项目的必要性、目的及意义  1.1项目必要性  随着智能终端设备与社交网络的迅速发展，每时每刻都将产生数以万计的视频。大量的各类视频资源丰富了我国民众的日常生活，已经成为人们文化休闲活动中不可或缺的部分。2017年文化部发布了《关于推动数字文化产业创新发展的指导意见》，指出要鼓励全民创意、创作联动的新方式，鼓励网络视频、网络音乐、网络文学等数字文化产品的创作。\*\*\*\*\*\*\*\*\*\*\*\*\*\*。可靠的视频编辑系统有利于文创人员更好地无障碍进行作品的构思和创意落地，对推动文化产业的发展具有重要作用和应用价值。  然而，传统的视频编辑系统有着各种缺点与不足：1）所需操作较为繁琐，对用户有较高的专业技能要求；2）编辑时间长，需要用户逐帧逐像素地对目标进行编辑操作；3）机械劳动多，编辑过程对相似帧进行重复操作；4）人力成本高，招聘经验丰富的视频编辑工作者需要提供过上万月薪。一套智能化程度高、功能较完善、操作简便的视频编辑系统十分必要，为此本研发项目将针对上述困难进行技术攻关。例如，本项目研发的视频编辑系统能让用户仅需单次交互即可将目标从视频中分离出来，减少重复机械劳动；同时单次交互可以降低对用户专业技能的要求，减少企业人力成本。除此之外，该系统能以较快的响应速度处理交互信息，大大减少编辑时间。综上所述，该项目拥有广阔的市场前景，可广泛应用于短视频生成、电影制作、节目采编、视频会议、动画创作、AR(增强现实)游戏、网络直播等实际场景中。  1.2研究目标  项目拟研发一套智能化程度高、功能较完善、操作简便的视频编辑系统。该系统以视频目标分割为核心技术，主要实现视频中前景目标与背景的分离，仅需用户单次交互来指定前景目标，即可自动将其从视频中分割出来。在此核心功能的基础之上，本系统可广泛应用至影视动画制作的视频合成领域。当得到分离的前景物体后，用户可以对其进行一系列相关的后续处理，并且将其合成至其他的视频中；用户在使用本系统将获得简便、快捷、智能、高效的体验。  1.3研究意义  XXX |
| 2、项目的背景、主要内容、技术水平及应用范围  2.1 项目背景  在移动互联网时代，信息的传播变得更加快速、扁平化和平民化。传统媒体作为信息主要提供者的时代逐渐远去，取而代之的是普通互联网用户作为信息的提供者，每时每刻分享自己的见闻。视频也随网络基础设施的升级逐步走向互联网舞台中央。由于生活节奏的加快，互联网用户浏览信息趋向快餐化，对信息的即时性要求很高。截至2019年6月，中国网络视频用户规模达7.59亿人，较2018年底增长3391万人，占网民整体的88.8%[11]，例如知名短视频平台抖音拥有日活跃用户超过3.2亿[12]。如此庞大的原创视频产业催生了对视频编辑的需求，视频编辑行业对高质量的视频处理往往以万元/分钟计算价格。但使用传统视频编辑系统需要较高的专业水平，普通用户的学习成本高且编辑时间长，因此用户迫切亟需一套智能化程度高、功能较完善且操作简便的视频编辑系统。  图1现有编辑系统的专业门槛高(左1)；操作复杂(左2)；消耗时间长(右1)。  2.2 主要内容  项目主要目标是研发XXX系统，该系统能够XXX，如对视频中某个目标进行追踪，对运动的目标进行像素级挖取，将某个视频中的目标移动到另一个视频中等，用户还可以在此基础上对视频做进一步的剪辑。项目重点研究实现xxx。  2.2.1 项目创新点  1) 在系统核心模块中设计带有xxxx。骨架网络结构图如下：  图2 模型的XXXX图  系统通过利用这些信息，能够对于目标边界的处理更加准确、精细，从而大大减少人工后期修改的时间。如图3所示，XXX。  图3 电影xxxx分割  技术路线：该模型中xxxx。  2) 提出并设计xxxx，从而提高对前景目标运动跟踪的能力。  模块结构图4如下：  图4 相似度匹配模块结构图  系统xxxx下图是一组关于该模块的实验结果：    未添加相似度模块时的分割结果  添加相似度模块后的分割结果  正确分割结果  图5 添加相似度匹配模块后的结果对比  图5中的第一行为原始视频帧，第二行为正确的分割结果， 第三行为未使用相似度匹配模块的分割结果，第四行为使用后的结果。xxxxx。  技术路线：新设计的xxxx。  2.2.2 项目实施思路  1）数据集准备和数据预处理  前主流的视频目标分割数据集主要有三个：DAVIS2016[4]，DAVIS2017[5]和YouTubeVOS[6]，xxxx。  2） 模型训练和多前景目标处理  数据集预处理完成后，xxx。  3） 模型测试与评估  视频目标分割领域最常用的两个模型测试指标为J和F。xxxx。  4）交互功能实现  xxxxx。  5）视频编辑系统设计  xxx操作。  2.3 项目技术水平  当前主流的视频目标分割技术主要通过传播和匹配两种手段来完成，传播的方法多使用前一帧已经得到的掩膜来对当前帧进行预测，如Masktrack[7]网络；而基于匹配的方法大多更关注当前帧与给定第一帧的相似程度，如OSVOS[8]网络。而在目前表现最好的一些模型中，xxxx。  本项目使用了Deeplab[3]中的模块作为网络的Backbone，Deeplab[3]是目前在语义分割领域内使用最广泛、效果最好的模型之一，这一创新有望大大提升特征提取的效果。在视频目标分割领域中的传播和匹配方法都与相似度学习有着紧密的联系，项目研发的系统核心模块使用xxxxx，符合xxx主流趋势，有望获得很好的分割效果，从而提升xxxxxxx。  2.4 应用范围  项目研发的视频编辑系统可广泛应用于电影制作、节目采编、视频会议、动画创作、AR(增强现实)游戏、短视频生成、网络直播等实际场景中。xxxx除此之外，如图8所示，在视频会议场景下，该系统能够快速地将会议人员的背景去除，从而避免隐私问题甚至安全问题。  图8 电影后期制作时，系统自动将任务从场景中分离出来，加载至新场景中    图9 视频会议时，系统自动将会议者从背景中分离，然后更换指定背景 |
| 3、实施该项目所具备的基础、优势和风险  3.1 项目基础  项目团队均由计算机专业的同学组成，具有专业领域的知识，能够进行探索和研究。 团队人员分工明确，各有所长，富有创新意识，能积极地参与到研究工作中来。 项目由经验丰富的专业老师指导，实验室资源丰富，为研究提供了充分的硬件支持。  项目组成员介绍：  \*\*\*\*：熟悉机器学习、神经网络方面相关知识， 能使用 Pytorch 等深度学习框架进行开发。英文文献阅读能力强。在校期间获得校一等奖学金三次，省政府奖学金两次、三好学生两次，在浙江省数学竞赛中获一等奖。阅读并实现了多篇视频目标分割相关论文算法和模型，对该领域有一定的认识和理解。  项目组成员在项目前期调研时，在Pytorch平台复现了多篇发表于CVPR、ECCV等顶级会议上的视频目标分割有关论文，对该领域有比较充分地了解，能理解并掌握目前该任务的核心技术和方法。图9给出了使用AGAME模型AGAME[1]在DAVIS2017数据集上进行复现的一组实验结果。从图中可以看到\*\*\*\*。  图10 使用AGAME模型在DAVI S2017检验集上的实验结果  3.2 项目优势  1）技术优势  语义分割是一个与目标分割紧密联系的领域，在项目中使用在语义分割领域中广泛使用、效果最好的Deeplab[3]网络作为模型的前端进行特征提取，输出的特征在提取图片深层信息的同时，还保留了较大的特征尺寸，对于相似度学习十分有利。通过Backbone的修改，项目所设计的模型可以学习得到表征能力更强的深度特征，有望取得更优的分割效果。  视频目标分割领域中长期基于传播和匹配两种手段来设计模型，在所设计的模型中同时考虑了这两种方法来进行相似度学习，并使用在目标跟踪领域内效果良好的互相关等运算来学习相连帧之间的局部相关性。这种方法和传统的内积等方法相比，不仅运算速度更快，还能更好地考虑到相连帧之间的运动变化特性，有望提高目标分割性能，从而增加视频编辑系统的实用性。  2）市场优势  互联网时代，随着科学技术的不断进步，越来越多由科技所驱动的新兴工具受到用户的青睐，在2019年Apple APP Store 编辑推荐的优质应用中，有不少都是以人工智能技术作为底层来开发的，如帮助用户进行高度自动化的图片修正和优化的图片处理APP咔嗒， 使用AI智能推荐用户音乐的APP Mubert等。可见，基于高新技术的产品有着良好的发展势头，当今市面上基于AI进行视频处理的产品还较少，本项目研发的视频编辑系统有着广阔的市场前景。  3）政策优势  互联网时代，科学技术的不断进步推动着其他各行各业的发展，随着文化强国战略的实施，科学技术和文化产业的结合越来越受到社会各界的关注。近年来，政府各部门多次出台了有关扶持文化产业发展和推动科技文化融合的政策，这些政策为文化创新提供了一片沃土，而先进的技术就是最好的肥料。在有关政策的扶持之下，相信本项目开发出的系统能够为用户提供一个容易上手、方便快捷的视频创作平台。  3.3 项目风险  项目主要基于视频目标分割技术来实现基本功能，而用到的深度神经网络比较复杂，包含多个模块，涉及参数数目庞大，训练时间也会较长，调参优化的过程可能存在一定的困难，从而导致模型达不到理想效果。但考虑到项目是基于当前效果最好的视频目标分割技术，输出模型效果极差的可能性较小。再者，由于实验室提供的良好硬件支持，大大减少了训练时间，为项目的顺利完成提供了保障。总体来说，尽管存在一些优化上的不确定性，但项目风险仍然比较小。 |
| 4、项目计划目标  1）设计一个端到端训练的视频目标分割模型，能够通过输入的视频和给出的第一帧掩膜输出后续各帧的掩膜。  2） 训练完成的视频目标分割模型能够得到state-of-the-art的测试结果，其指标J的值在70左右，指标F的值在75左右，指标J&F的值在72左右。  3） 通过一个图像目标分割模型，让用户能够通过简单的输入标记来得到视频第一帧的掩膜。  4） 设计一个视频编辑系统，其主要功能为让用户能够选择自己的视频，并通过简单的界面交互指引用户输入一些标记，得到第一帧的掩膜，之后据此得到视频的完整分割结果返回给用户。  5） 在视频编辑系统中实现视频编辑软件的一些基本功能，如像素级的视频目标挖取，以及两个不同视频的前景目标和后景的合成等。  参考文献：  [1] Joakim Johnander,Martin Danelljan, Emil Brissman, Fahad Shahbaz Khan, Michael Felsberg. A Generative Appearance Model for End-to-end Video Object Segmentation . [C].Proceedings of the IEEE conference on computer vision and pattern recognition. 2019: 8953-8962.  [2] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C].Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.  [3] Chen L C, Papandreou G, Schroff F, et al. Rethinking atrous convolution for semantic image segmentation[J]. arXiv preprint arXiv:1706.05587, 2017.  [4] Perazzi F, Pont-Tuset J, McWilliams B, et al. A benchmark dataset and evaluation methodology for video object segmentation[C].Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 724-732.  [5] Pont-Tuset J, Perazzi F, Caelles S, et al. The 2017 davis challenge on video object segmentation[J]. arXiv preprint arXiv:1704.00675, 2017.  [6] Xu N, Yang L, Fan Y, et al. Youtube-vos: Sequence-to-sequence video object segmentation[C]. Proceedings of the European Conference on Computer Vision (ECCV). 2018: 585-601.  [7] Perazzi F, Khoreva A, Benenson R, et al. Learning Video Object Segmentation from Static Images[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition., 2017: 3491-3500.  [8] Caelles S, Maninis K, Ponttuset J, et al. One-Shot Video Object Segmentation[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 5320-5329.  [9]Jonathon Luiten, Paul Voigtlaender, Bastian. PReMVOS: Proposal-generation, Refinement and Merging for Video Object Segmentation. Proceedings of Asian Conference on Computer Vision. 2018: 565-580.  [10] Voigtlaender P, Chai Y, Schroff F, et al. FEELVOS: Fast End-To-End Embedding Learning for Video Object Segmentation[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019: 9481-9490.  [11] 温梦华. 字节跳动：旗下产品总月活跃用户数超过15亿 抖音日活跃用户超3.2亿.[2019.7.9]. http://finance.ifeng.com/c/7oAdsm0ukJ5.  [12] 卢扬,杨雅. 中国网络视频用户规模达7.59亿 占网民总数近九成.[2019.12.18].http://finance.ifeng.com/c/7sVKFMP4pzp. |

**三、预期成果、知识产权形成及经济、社会效益分析**

|  |
| --- |
| 1. 项目的预期成果及知识产权归属情况   1）核心及以上期刊论文发表 1 篇；  2）受理发明专利申请 1 项；  3）软件著作权 1 项；  4）移动端APP平台 1款；  5）项目总结报告1份； |
| 1. 项目的市场前景分析   当前市场中的视频编辑软件存在以下不足和缺点：  \*\*\*\*\*\*\*  本项目所研发的视频编辑系统能够\*\*\*\*\*。   1. 项目的盈利能力分析及财务预算   项目盈利能力分析：  根据“前程无忧”和“中华英才网”等求职网站显示，短视频编辑员在北京、上海、广州、深证等地的工资为6k-10k不等，动画设计师的工资为5k-15k不等，而特效设计师更需要8k-15k不等。这样的高素质人才往往由于视频编辑技术的局限导致时间耗费在重复机械的视频目标分割上，这是对人力的巨大浪费。而企业引进所完成的产品可以大大减少人力的浪费，将人力投入到更具有创造性的环节中。因而，视频相关企业与视频行业从业者对视频编辑系统有很大的需求，但是传统视频编辑应用存在处理功能单薄、智能化程度低、操作复杂、耗费时间长等问题，不能很好地满足实际需求。从这个角度来看，项目所研发 的视频编辑系统具有很大的盈利空间。  项目财务预算：  项目总预算共计5000元，各类支出项目及用途具体见表1。  表1 财务预算表   |  |  |  |  | | --- | --- | --- | --- | | 序号 | 支出项目 | 预算金额 | 支出用途 | | 1 | 图书资料版面费 | 700元 | 图书资料购买、打印、复印及论文版面费。 | | 2 | 调研差旅费 | 1000元 | 参加相关学术会议注册、交通、住宿费用。 | | 3 | 实验材料费 | 1000元 | 移动硬盘、路由器等相关实验材料费。 | | 4 | 专利申请费 | 2000元 | 国家发明专利的申请费用。 | | 5 | 软著申请费 | 300元 | 软件著作申请费用。 |  1. 项目的社会效益分析   本项目中设计的视频编辑系统可以大大简化视频编辑操作，减少编辑过程中的人力浪费与时间成本，从而减轻企业负担，促进经济发展。通过使用智能化的系统，将高素质人才从重复机械的劳动中解放出来，转向更有创造性的劳动中来，为人们带来更多的优秀作品。智能化的视频编辑将大大降低视频处理的准入门槛，令普通用户也能进行视频创作，从而鼓励全民文化创作。 |

**四、项目实施进度方案**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | 时间段 | 实施进度安排 | | 2021.01-04 | 项目前期调研，对国内外最新成果进行深入研究和讨论 | | 2021.05-08 | 研究学习项目相关知识并编写项目demo | | 2021.09-12 | 完成项目代码编写调试，训练模型并分析模型效果 | | 2022.01-02 | 进一步讨论研究，对模型架构进行适当调整  根据训练结果分析调整模型超参数 | | 2022.03-05 | 撰写项目结题报告  撰写视频目标分割相关论文并申请相关专利 | |

**五、项目组承诺**

|  |
| --- |
| **承 诺 书**  以上所填内容真实可靠，本项目组承诺：该项目立项后，将严格遵守有关规定、遵守本申报书和预算表中规定的条款和内容，保证按计划进度完成项目任务。  项目组全体成员（签章）：    年　　月　　日 |

**六、学校审核意见：**

|  |
| --- |
| 负责单位（公章）：  年 月 日 |

**七、专家组审核意见**

|  |
| --- |
| 专家组组长签章：  年 月 日 |

**八、实施办公室审核意见**

|  |
| --- |
| 公章：  年 月 日 |