



西北大学
NORTHWEST UNIVERSITY

第十一届“挑战杯”大学生课外学术科技作品竞赛

参赛作品

作品名称：老影焕新记——经典灰度影视作品的超分辨率着色

作者姓名：刘梦琪、鄢博达、朱建鹏、李敬铮、万琦玥、杨雨欣、宋佩玲、张波

项目类别：

☐ 个人项目 ☒ 集体项目

学生类别：

☐ 本科 B ☒ 硕士 M

作品类别：

- ☐ 1. 自然科学类学术论文
☐ 2. 哲学社会科学类社会调查报告和学术论文
☒ 3. 科技发明制作 A 类
☐ 4. 科技发明制作 B 类

所属领域：（作品类别为 1 和 2 的参赛作品请勾选）

- | | | |
|--------------------------------|-------------------------------|-------------------------------|
| <input type="checkbox"/> 机械与控制 | <input type="checkbox"/> 信息技术 | <input type="checkbox"/> 数理 |
| <input type="checkbox"/> 生命科学 | <input type="checkbox"/> 能源化工 | |
| <input type="checkbox"/> 发展成就 | <input type="checkbox"/> 文明文化 | <input type="checkbox"/> 美丽中国 |
| <input type="checkbox"/> 民生福祉 | <input type="checkbox"/> 中国之治 | <input type="checkbox"/> 战疫行动 |
| <input type="checkbox"/> 哲学 | <input type="checkbox"/> 经济 | <input type="checkbox"/> 社会 |
| <input type="checkbox"/> 法律 | <input type="checkbox"/> 教育 | <input type="checkbox"/> 管理 |

作品原创性承诺书

本人刘梦琪郑重承诺：

在参加西北大学第十一届“挑战杯”大学生课外学术科技作品竞赛中，本人所提交的参赛作品《老影焕新记——经典灰度影视作品的超分辨率着色》系本人及团队成员在老师指导下，进行深入研究或调研后取得的原创性成果，未有抄袭、剽窃等侵犯他人知识产权的行为。本人代表项目团队做出上述承诺，如有不实，愿按有关规定接受处理，并承担相应责任。

承诺人：_____

2023 年 3 月 15 日

摘要

影像时代电影的跨媒介资源转化，让原本抽象的主题发生了形象化的转变，其对于人类精神的教育具有春风化雨的作用。一些经典的灰度影视作品或具有深远的教育意义，或带给人们视觉和精神上的享受，广受人们的青睐。相比于灰度影像，人眼对彩色影像更敏感，彩色影像本身价值更高，且视觉效果更好。翻新经典老旧影视作品不仅能够更好地发挥经典灰度老电影的历史价值和商业价值，而且对于影视作品的数字化保护、优秀传统文化的传承与传播具有重要的现实意义。

本研究旨在基于深度学习方法，围绕经典灰度影视作品的超分辨率着色展开研究，实现经典灰度影视作品上色及分辨率的提高，使其焕发新的生机，为大众喜闻乐见。视频的超分辨率着色技术旨在为单张或序列灰度视频添加颜色的同时提高分辨率，即实现灰度视频上色及清晰化。目前，超分辨率上色方式主要是利用视频处理软件对灰度影视作品进行人工上色和提高分辨率，其存在对操作人员的专业技能要求很高、工作量大、成片速度慢，上色和超分辨率结果严重依赖操作人员的主观判断、且相邻视频帧之间的色彩连续性不足等弊病。针对上述问题，我们设计了一个新的基于深度学习方法，在参考图像引导和时空特征补偿约束下的端到端视频超分辨率着色网络框架，该框架统一了语义对齐、颜色传播和光流补偿模块，通过语义对齐与颜色传播将参考图像的颜色传播至灰度视频，使其上色；光流补偿网络则利用相邻视频帧之间的短时连续和内容相似性特征为当前帧着色（当前待着色帧）提供有效丰富的细节信息；融合上色帧及细节信息可获得具有时间连续性和高感知质量的彩色视频。基于上述超分辨率着色方案，我们实现了经典灰度电影《林海雪原》和《智取威虎山》，以及纪录片《开国大典》的上色及清晰度的提高。

灰度老电影在影视发展史中具有重要的历史地位、研究价值及现实教育意义，本项技术成功的将经典灰度影视作品旧貌换新颜的实验结果表明，我们的技术路线可行，并可以推广应用于其他老旧影视作品的超分辨率着色，实现影视文化遗产的数字化保护、传承与弘扬，满足广大人民群众对优秀传统文化的需求，同时

也能增强国家软实力和国际影响力。

关键词：视频着色，超分辨率重建，灰度影视作品，数字化保护与传播

Abstract

The cross-media resource transformation of films in the image era has transformed the original abstract themes into figurative ones, and its effect on the education of human spirit is spring and rain. Some classic grayscale film and television works are widely popular because they either have profound educational significance or bring people visual and spiritual enjoyment. Compared with grayscale images, the human eye is more sensitive to color images, which have higher value and better visual effects. Refurbishing classic old film and TV works not only can better utilize the historical and commercial value of classic grayscale old films, but also has important practical significance for digital preservation, inheritance and dissemination of film and TV works.

Based on deep learning methods, this study aims to focus on the super-resolution coloring of classic grayscale film and television works to achieve the improvement of coloring and resolution of classic grayscale film and television works, so that they can be revitalized and enjoyed by the public. The purpose of super-resolution video coloring is to add color to single or sequential grayscale videos while increasing the resolution, i.e., to achieve colorization and clarity of grayscale videos. At present, the super-resolution coloring method mainly uses video processing software to manually color and increase the resolution of grayscale video works, which has the drawbacks of requiring high professional and technical skills of operators, large workload, slow completion speed, serious dependence on the subjective judgment of operators for coloring and super-resolution results, and insufficient color continuity between adjacent video frames. To address the above problems, we design a new end-to-end video super-resolution coloring network framework based on deep learning method with the constraints of reference image guidance and spatio-temporal feature compensation, which unifies semantic alignment, color propagation and optical flow compensation modules, and propagates the color of reference image to grayscale video for coloring through semantic alignment and color propagation; the optical flow compensation network uses the short-time continuity and content similarity between adjacent video

frames to The optical flow compensation network uses the short time continuity and content similarity features between adjacent video frames to provide effective and rich detail information for the current frame coloring (current frame to be colored); fusing the colored frames and detail information can obtain color video with time continuity and high perceptual quality. Based on the above super-resolution coloring scheme, we have achieved the colorization and clarity improvement of the classic grayscale movies *Forest and Snow* and *Wisdom of the Mountain*, as well as the documentary *The Founding Ceremony*.

Grayscale old films have an important historical position, research value and practical educational significance in the history of film and television development. The experimental results of this technology successfully changing the old appearance of classic gray-scale film and television works show that our technical route is feasible, and can be popularized and applied to the super-resolution coloring of other old film and television works, so as to realize the digital protection, inheritance and promotion of film and television cultural heritage.

KEY WORDS: Video Colorization, Video Super-resolution Reconstruction, Film And Television Works, Digital Protection And Communication

目 录

摘要	I
Abstract	I
第 1 章 研究背景与意义	1
第 2 章 国内外研究现状	3
2.1 着色技术研究现状	3
2.2 超分辨率技术研究现状	4
第 3 章 经典灰度影视作品超分辨率着色方案	6
3.1 灰度影视作品的超分辨率着色的难点与创新	6
3.2 灰度影视作品超分辨率着色方案	7
3.3 着色网络	7
3.3.1 语义对齐模块	8
3.3.2 颜色传播模块	9
3.3.3 光流补偿模块	9
3.4 超分辨率网络	10
第 4 章 灰度影视作品超分辨率着色系统软件设计	12
4.1 软件体系结构	12
4.2 PC 前端用户界面设计	12
4.3 PC 后端设计	13
第 5 章 灰度影视作品超分辨率着色系统测试与分析	14
5.1 测试环境	14
5.2 视频数据集构建	14
5.2.1 着色训练集	14
5.2.2 超分训练集	14
5.2.3 参考视频选择	15
5.2.4 测试数据	15
5.3 测试结果的定性分析	15
5.3.1 《开国大典》结果及分析	15
5.3.2 《林海雪原》结果及分析	16

5.3.3 《智取威虎山》结果及分析	20
5.4 测试结果的定量分析	21
第6章 总结与展望	25
参考文献	26

第1章 研究背景与意义

人类借助感官收集环境信息，形成对客观世界的认知，进而探索和改造世界。感官中的视觉获取的视频或图像信息最为直观且丰富，因此视频和图像作为视觉信息传播的媒介，在通讯交流和科技发展中占据着举足轻重的位置。大数据时代，机器设备亦能依赖海量的图像和视频进行视觉感知，进而代替人类进行信息筛选与理解分析，例如人脸识别、无人驾驶和智能安防等。随着深度学习理论及相关技术的发展，高质量的视频和图像数据对于促进互联网技术的发展至关重要。

图像利用空间像素对信息进行记录和存储。视频则由连续运动的图像构成，其蕴含的信息不仅包括空间维度，还包括时间维度。在数字图像处理技术^[1]中，往往要求通过提高分辨率获得高品质的视频或图像，其为后续进一步的信息提取、理解、识别与分析奠定良好的基础。图像分辨率是指图像每单位面积包含的像素点数量，像素数越多表示图像中包含的信息内容越丰富。相比于低分辨率视频/图像，高分辨率（High Resolution, HR）视频/图像所携带的细节信息更加丰富，成像更加精细。实践中，视频/图像的质量极容易受到多方面因素的影响，而导致信息丢失和分辨率降低。造成视频/图像分辨率质量降低的因素主要有来自于成像设备精度的限制，以及数据处理或存贮过程中的误差与损耗等。提高视频/图像分辨率的主要手段分为硬件技术和软件技术两种。硬件技术是指通过增加成像装置的传感器密度、构造高精度成像芯片以及升级光学组件等方式获取更精细的成像结果。但由于制造工艺以及硬件成本的限制，以硬件方式提高分辨率的上升空间有限。软件技术是指对劣化的视频/图像采用信号处理的方式进行分辨率的提升。该方法可以打破硬件技术的局限性，平衡计算成本与硬件成本，同时具有操作简便，效果显著，高效快捷等特点。基于软件手段提高图像/视频分辨率的技术也称为超分辨率(Super-Resolution, SR)重建。

颜色是视频/图像中最重要的视觉信息要素之一。人眼对彩色图像更敏感，人眼只能识别出十几种视频/图像的灰度等级，但却可以分辨出成千上万种色彩，彩色视频/图像对于人眼的影响要远远大于灰度视频/图像。此外，灰度视频/图像对于信息特征的表达能力有限，而彩色视频/图像可以全面的表征信息特征。若将灰度视频/图像进行着色处理，则有利于视频/图像特征的表达和提取，使人眼更快、更全面的捕捉到视频/图像本身蕴含的内容信息，以及对视频/图像的准确理解，使视

频/图像充分发挥其具有价值。图像/视频的着色(colorization)技术^[2]旨在为序列或单张的灰度视频/图像添加颜色，是对黑白或灰度视频/图像进行彩色化^[3]的过程。着色技术最早始于人工为黑白图像上色，随着着色技术的发展其在数字图像处理、遥感卫星、工业生产以及医学影像处理等领域具有重要的研究价值和应用需求。由于着色问题的多样性和病态性，使得感知质量成为衡量着色生成图像质量的一个重要标准。目前，如何将黑白或灰度视频/图像生成内容与语义相符合，且感知质量高的彩色结果这一课题备受相关研究人员的关注。

历史的车轮滚滚向前，留下了许多珍贵的影视作品，其不仅是珍贵的文化遗产，有的还具有重要的历史和教育意义。同时党的二十大报告指出，弘扬传统文化、保护文化遗产是全面建设社会主义现代化国家的重要内容之一，还提出了一系列具体措施和目标，如“推动中华优秀传统文化创造性转化、创新性发展”。遗憾的是受限于当时的技术，很多珍贵的影像都是黑白或灰度的，且分辨率不高。研究将珍贵的黑白影像着色为彩色影像，并且提高其分辨率，使其整旧如新不仅能够大大提升影像的观赏性，更重要的是提高了影像本身的历史和商业等价值，实现影视文化遗产的数字化保护与弘扬。传统的手工上色处理方式使用后期视频处理软件对影像帧进行处理和上色，该模式对操作人员的专业技术水平和主观判断有很强的依赖性，且完成一整部黑白或灰度影像作品的上色需逐帧进行手动上色，工作量巨大，成片速度慢，往往无法满足当前影视工业对于黑白老影片上色的巨大需求。

第2章 国内外研究现状

2.1 着色技术研究现状

视频/图像着色技术^[4]在图像处理领域有着非常重要的研究和应用价值，着色问题涉及到多个方面的算法研究，包括计算机视觉^[5]、灰度图像处理^[6]、图像颜色空间变换^[7]等。由于对同一个视频/图像来说，不同的用户有不一样的着色感官，即着色的结果可能是多种多样的，因此视频/图像着色具有一定的主观性。目前，常见的图像着色方法可以分为基于局部颜色着色方法^[8]和基于颜色传递着色方法^[9]。

基于局部颜色扩展方法的基本步骤是首先在被着色图像的某一块区域中输入一定量的颜色；然后根据某算法，将输入的颜色扩展并蔓延到整幅待着色图像。Levin^[10]等人提出了改进上述步骤的方法，其假设灰度值相近的像素着色后的颜色相近作为前提，以待着色图像上的输入颜色为条件，在 YUV 颜色空间^[11]着色。该方法将着色问题当成对整幅图像优化问题来处理。这类方法只适用于处理小尺度的图像着色，其着色处理的速度会随着图像长和宽的加大而变慢。Qu^[12]和 Luan^[13]提出在图像的着色过程中加入图像纹理特征的相似性，对待着色图像中的纹理相似的区域，可以适量的减少输入，此类算法可以大大减小计算复杂度，但是需要人为交互^[14]。基于局部颜色着色的方法是以用户标注的颜色为导向，都需要用户的人为干预，如在目标图像上进行区域颜色标注，或者将图像分割成小的图像单元块，这样着色的结果才能更符合预期，更接近真实彩色图像的水准，也更容易被用户接受。基于局部颜色着色方法的弊端在于，由于每个用户对于图像的理解不同，对图像的审美观也不相同，因此在局部区域标注的颜色不同，进而导致最后的着色结果可能截然不同。同时，如果人工标注的颜色不准确，则着色结果将与预期相背，甚至出现颜色错误、混乱的结果，尤其是在亮度变化较为明显的区域，如果错将颜色标注在了边缘区域，则颜色扩散的结果将使着色失败。

为了消除人为因素在图像着色过程中的干预和影响，Welsh^[15]等人最早提出了基于颜色传递的图像着色方法。其与基于局部颜色扩展不同，此类方法需要一幅彩色的参考图像，将参考图像中的颜色信息传递到待着色图像上。这就需要在彩色参考图像和待着色图像中寻找相似部分，实现将参考图像的颜色传递给待着色图像。Welsh 等人在图像的亮度通道上寻找相似的像素，将相似像素对应的色度进

行迁移，获得与参考图像颜色类似的着色结果。但其存在具有同样亮度的像素只能有一种颜色，这极大的缩小了可着色的颜色种类，着色结果极大的会受到参考图像的限制。

相比于局部颜色扩展的方法，基于颜色传递的方法基本不需要人机交互，自动化程度较高，只要寻找到合适的参考图像，选取合适的匹配特征，即可得到比较理想的着色结果，并且着色过程会根据用户给出的情感词不同得到不同的着色结果。

相比于图像着色，视频着色的相关研究较少。视频着色需要在满足生成真实且符合语义的上色结果的同时，还需要满足视频着色的时间一致性。现有的视频着色方法可以分为三类：第一类方法是用时间滤波器对着色框架进行后处理^[16, 17]，这种方法通常会把颜色过滤掉。第二类方法是通过显式计算光流^[18-20]将颜色涂鸦传播到其他帧，但存在某一帧中的特定涂鸦未必适用于其他帧，往往造成着色的混乱。第三类方法是以某一个彩色帧为例，并按顺序对后续帧着色。传统的视频着色方法依赖于手工制作的低级特征来寻找时间对应，而最新的视频着色研究趋势则是使用深度神经网络以数据驱动的方式来学习时间传播^[21, 22]，其可以获得更好的着色质量，但存在如果传播的颜色在某帧上失效，则会造成颜色传播错误累积。

2.2 超分辨率技术研究现状

超分辨率（Super-Resolution, SR）技术的目的是根据若干观测到的低分辨率（Low Resolution, LR）图像/视频构建包含更多细节信息的相应的高分辨率（High Resolution, HR）图像/视频。解决 SR 问题的关键在于如何利用 LR 图像/视频提供的局部空间特征以及额外的先验信息预测 HR 图像/视频中缺失的纹理和细节。SR 算法可以分为两类：单幅图像 SR 方法（Single-Image Super-Resolution, SISR）和视频 SR 方法（Video Super-Resolution, VSR）。SISR 算法利用单幅图像的空间自相似性重建 HR 图像缺失的高频细节。目前，SISR 方法经过几十年的研究取得了良好的性能。由于视频数据计算复杂度高，且视频中存在目标运动和背景变化等产生的闪烁伪影，VSR 质量提高仍然是一个难以较好解决的问题。

视频是由多幅在同一场景或多个场景下对目标进行观测的图像序列构成。因此相比于单幅图像，相邻视频帧在时域和空域包含丰富的相关性信息。VSR 可根据视频序列具有的时间依赖性和空间相似性，从 LR 视频帧重建对应的 HR 结果。

根据对帧间信息的利用方式可将 VSR 分为两种类型：显式帧对齐的 VSR 和隐式帧对齐的 VSR。显示运动估计补偿的 VSR：此类方式是指 VSR 模型中存在独立的运动估计和补偿模块用于帧间信息的提取及融合，其利用提取的运动信息使相邻帧与目标帧明确对齐。显示帧对齐的 VSR 通常由两个阶段组成：运动估计补偿和 SR 重建。2021 年，Chan 等人^[23]提出了 BasicVSR 框架，通过帧间信息双向传播充分捕获视频序列的时序关系，并利用光流信息进行帧间特征匹配，实现高质量的 VSR 重建。隐式运动估计补偿的 VSR：此类方法不依赖于独立的运动估计补偿模块，而是通过网络自主学习空间特征或时间与空间的融合特征来引入帧间相关信息，从而利用帧间的强关联性实现高质量的 VSR。Isobe 等人^[24]将 RNN 结构与残差学习思想相结合，提出了一种用于长视频序列 VSR 的隐式状态结构递归网络 RRN，生成了时序一致且特征丰富的 VSR 结果。

第3章 经典灰度影视作品超分辨率着色方案

党的二十大报告提出了“两个一百年”奋斗目标和中华民族伟大复兴中国梦。利用数字技术对传统文化进行保护和传播，既能夯实中华民族共同精神家园、增强民族自信心和自豪感，又能推动经济社会发展、科技创新和文明交流互鉴。利用数字技术对传统文化进行保护和传播，既能满足广大人民群众对优秀传统文化的需求，又能促进传统文化与时俱进、创新发展，同时也能增强国家软实力和国际影响力。

本着翻新经典灰度影视作品，使其整旧如新和发挥他们的历史价值和商业价值，以及实现影视作品的数字化保护、传承与传播为目标，我们基于深度学习方法，围绕经典灰度影视作品的超分辨率着色展开研究。

3.1 灰度影视作品的超分辨率着色的难点与创新

经典灰度影视作品存在没有颜色，以及分辨率低的特点，对其超分辨率着色具有三个方面的难点和挑战：一是视频维度相对较高，如何保持视频具有时空一致性的同时增加高频细节信息；二是着色结果的病态性质，以及颜色传播过程中出现在特定帧上失效，造成错误的颜色扩散累积、非局部空间对应经常导致的实例之间的颜色泄漏，使得对算法方案或参考图像的要求高；三是保证系统与用户的可交互性，以及降低计算成本。

鉴于传统的着色方法基于低级特征，依赖手工寻找时间对应的制作方式上色，无论是效率还是质量均难以保证。我们拟采用深度神经网络用数据驱动的方式来实现灰度影视作品的超分辨率着色，以实现灰度影视作品的上色及分辨率的提高。为了避免颜色传播过程中出现在特定帧上失效，造成错误的颜色扩散累积，以及非局部空间对应经常导致的实例之间的颜色泄漏，或使用多个对象的平均颜色着色问题。考虑到可以利用语义特征计算目标帧和参考帧之间的对应关系^[25, 26]，以及通过关注参考帧的颜色对目标帧进行着色，可达到减少颜色传播过程中累积预测失误^[27, 28]，我们设计了一个新的引入参考图像和时空特征补偿的端到端视频着色网络框架，该框架结合超分辨率网络可获得具有时间连续性，感知质量高的彩色视频。该研究的创新性主要有：

(1) 设计了一个引入参考图像和时空特征补偿的端到端视频着色网络框架，该框架统一了语义对齐、颜色传播和光流补偿模块，通过语义对齐与颜色传播将参考颜色传播到灰度视频中；光流补偿网络则利用相邻视频帧之间的短时连续和内容相似性特征为当前帧着色提供有效丰富的细节信息。

(2) 设计了一个光流估计补偿结合多特征鉴别 GAN 的视频超分辨率网络框架，该网络使光流估计补偿网络和多特征鉴别 GAN 在一个框架里端到端训练，以确保视觉敏感特征边缘纹理清晰，颜色逼真，获得令人愉悦的视觉感受。

(3) 设计并实现了灰度影视作品超分辨率着色系统，目前该系统集成了三部上色且清晰度提高的经典灰度影视作品：电影《林海雪原》和《智取威虎山》，以及纪录片《开国大典》。经典灰度影视作品旧貌换新颜的实验结果表明，本项技术有益于灰度老旧影视作品的数字化保护、传承与传播。

3.2 灰度影视作品超分辨率着色方案

针对此次设计总共提出了两种可行的方案。

方案一：

先超分再着色，将黑白视频先经过视频超分网络进行超分得到高分辨率视频，再对结果进行着色，得到最终的彩色高分辨率视频。

方案二：

先着色再超分，先将黑白视频经视频着色网络进行着色，再对着色的结果进行超分得到最终的彩色高分辨率视频。

比较上述两种方案，方案一与方案二在整体结构上大致相同，但在效率方面发现方法二的效率高于方法一，这是由于在视频着色模块中，相较于高分辨率的输入，网络的计算量要高于低分辨率的输入，不仅会使得模型速率降低，而且最终效果也不如方案二。

综上所述，本次设计选取方案二来完成经典灰度影视作品的超分辨率着色。

3.3 着色网络

本研究拟采用的网络框架如图所示，网络主要由三部分组成：语义对齐模块、颜色传播模块和光流补偿模块。待上色的灰度图像和彩色参考图像首先输入对应网络中，通过 VGG-19 网络提取多层特征图，然后输入到对应网络中得到扭曲结果及基于注意力的对应度置信图，将其当前帧灰度图像一同输入到自动着色网络

中；着色网络拟采用 U-net 结构，同 U-net 网络结构不同的是它由两个向下采样层和八个残差块组成，避免了基于 U-net 架构的瓶颈问题，过多的下采样层会对输入数据产生稀疏的不利影响，而残差块结构减少了向下采样的需要，并且允许在训练中跳过层，对应网络的输入通过着色网络会优化扭曲的结果 W_{ab} ，使其与连续的帧具有时间一致性，同时对 S 中对应度较低的像素，即输入图像与参考图像不对应的像素位置根据其语义、纹理等信息进一步着色，提高空间上的一致性，最终得到一个优化的结果 H_t 。第三部分是光流补偿网络，将着色后的前一帧和当前帧灰度图像输入光流估计网络中，计算出前一帧灰度图像到当前帧灰度图像的正向光流 $Flow$ ，通过该光流对彩色前一帧图像进行扭曲，扭曲后的结果同着色网络的输出结合后得到着色的最终结果 X 。

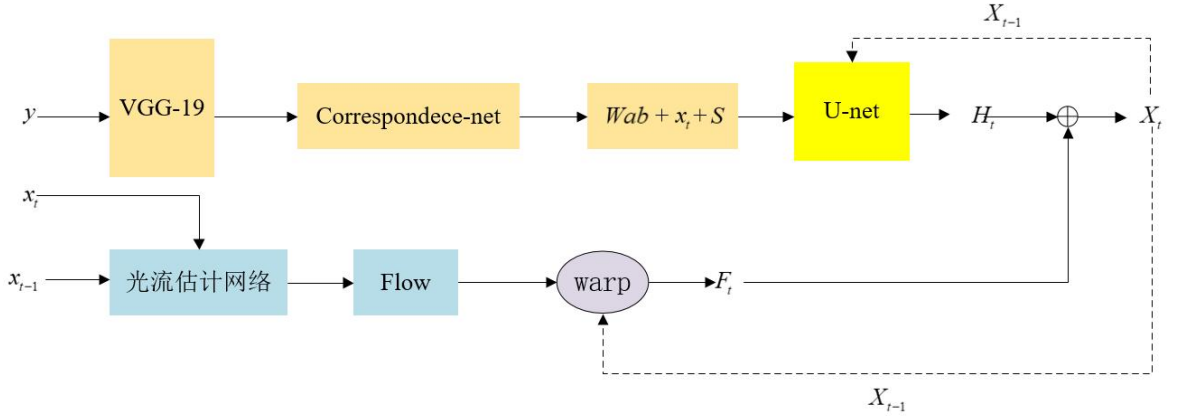


图 3.1 视频着色网络框架

3.3.1 语义对齐模块

语义对齐模块的结构如图所示，它的作用在于将参考图片的颜色迁移到目标图片上。当前帧灰度图像 x_t 和参考图像输入 VGG-19 网络中提取特征，得到多层特征图连接后输入到几个残差块中以更好地利用来自不同层的特征，然后将残差块的输出重新整形为两个特征向量 F_x , F_y ，形式上计算一个相关矩阵 $M(i,j)$ 表示输入的灰度图像同参考图像之间特征的成对相似性，其元素代表了两者分别在 i 位置和 j 位置的相似性，然后近似选择相似度得分最高的像素将参考颜色 y 向 x 扭曲，得到加权颜色矢量 W_{ab} 以及匹配置信度图 S （代表了对 x 中每个位置 i 取样参考颜色的可靠性）用以指导下一步的着色，两者同灰度图像 x 以及上一帧彩色生成图像一起输入到着色网络中。

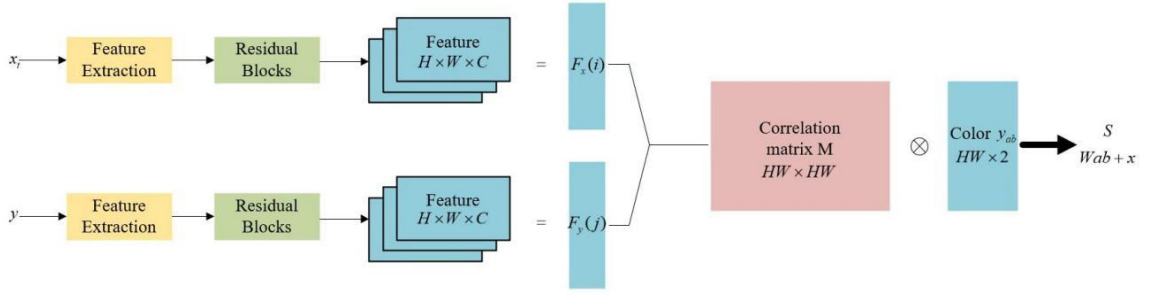


图 3.2 语义对齐模块结构

3.3.2 颜色传播模块

语义对齐模块的输出同待着色的灰度图像一起作为颜色传播模块的输入，它是一个经典的 CGan 的结构，由 2 层下采样和 8 个残差块以及 2 层上采样组成，其结构取自于 U-net 的网络结构，但由于 U-net 对输入进行多次下采样提取特征后又进行上采样，这会导致数据的丢失，特别是对于稀疏的输入数据影响更大，即信息瓶颈问题。为解决该问题，在网络中引入了残差块结构，它不需要下采样来提取特征，而是允许在训练中跳过层，同时避免梯度消失问题。输入的扭曲颜色 $W_{ab} + x_t$ 作为条件输入到网络中，在置信图 S 的指导下进行进一步着色，同时彩色的前一帧作为额外的条件也输入到网络中，引入时间信息，保持时间一致性， S 相当于一个注意力矩阵，着色子网络将着重于 S 中得分较低的区域，即同参考图片相似度不高的区域根据其语义信息和纹理特征着色，优化扭曲颜色，提高空间一致性，最终得到一个优化的扭曲颜色结果 H_t 。

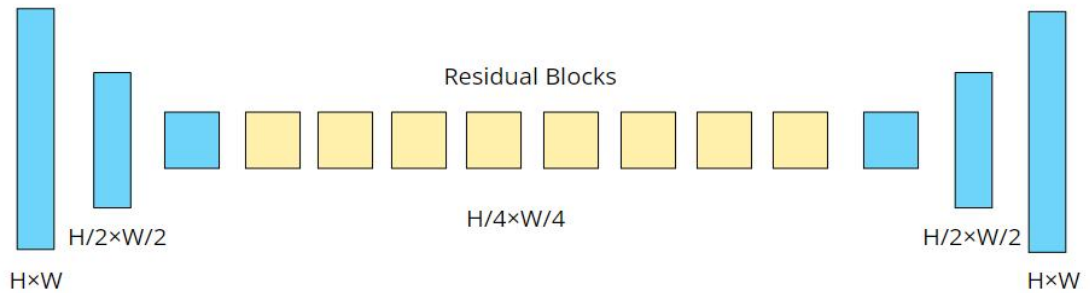


图 3.3 颜色传播模块结构

3.3.3 光流补偿模块

光流补偿网络采用 Flownet 的架构，输入当前帧和前一帧的灰度图像，直接将

两者叠加在一起，通过编码-解码的结构得到光流图，在解码部分通过一系列只有卷积层的网络，在解码部分对每层的反卷积层，不仅输入前一层的输出，同时输入前一层计算的低尺度的光流和对应的编码模块中的特征层，这样不仅可以获得深层的抽象信息，同时还可以获得浅层的具象信息，以弥补因特征空间尺度的缩小而损失的信息，提高光流的准确度。最终网络输出前一帧到当前帧的正向光流 Flow，用该 Flow 对彩色的前一帧结果进行扭曲，得到扭曲结果 C_t ，然后同 H_t 两者结合得到最终的彩色输出视频帧。

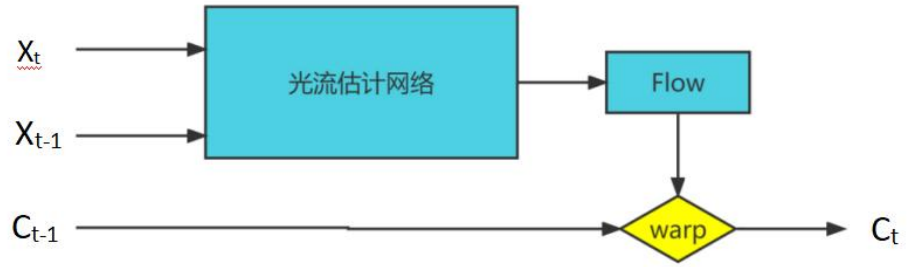


图 3.4 光流补偿模块

3.4 超分辨率网络

视频超分网络框架如图 3.5 所示，网络由光流估计补偿网络和多特征鉴别 GAN 构成，多特征鉴别器包括像素鉴别器、边缘鉴别器和纹理鉴别器。

视频超分网络的训练过程为：首先将视频 LR 帧 I_{t-1}^{LR} 、 I_t^{LR} 和 I_{t+1}^{LR} 输入光流估计补偿网络分别估计相邻两帧 I_{t-1}^{LR} 和 I_{t+1}^{LR} 与目标帧 I_t^{LR} 之间的光流 $F_{t \rightarrow t-1}^{LR}$ 和 $F_{t \rightarrow t+1}^{LR}$ ，并计算用光流 $F_{t \rightarrow t-1}^{LR}$ 和 $F_{t \rightarrow t+1}^{LR}$ 补偿后的补偿帧 $I_{t-1}^{LR'}$ 和 $I_{t+1}^{LR'}$ ；其次在深度维度上堆叠 $I_{t-1}^{LR'}$ 、 I_t^{LR} 和 $I_{t+1}^{LR'}$ 获得融合特征 F_{fused} ，并将 F_{fused} 输入生成器生成 SR 结果 I_t^{SR} ；最后生成器 G 与像素鉴别器 D_P 、边缘鉴别器 D_E 和纹理鉴别器 D_T 对抗鉴别 SR 帧 I_t^{SR} 与 HR 帧 I_t^{HR} ，促使 I_t^{SR} 和 I_t^{HR} 的整体像素值、边缘和纹理趋于一致。测试过程是将 LR 待超分帧和相邻两帧输入到训练好的视频超分网络，获取生成器输出的 SR 结果。

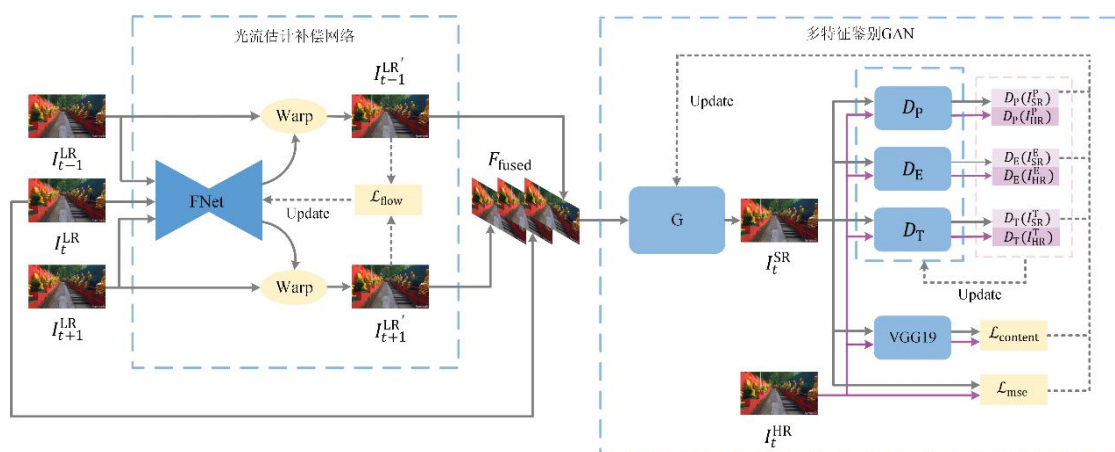


图 3.5 视频超分网络

第4章 灰度影视作品超分辨率着色系统软件设计

4.1 软件体系结构

根据系统论证和系统所用算法的原理分析，可以进行软件的设计，首先根据需求画出功能模块图，再进行所需环境的配置，前后端的设计。用户界面保存用户的操作、待着色图片和参考图片的选择及可视化结果等的一系列操作。后台算法对于输入图片进行特征提取，然后分别通过语义对齐模块、颜色传播模块、光流补偿模块及超分网络进行视频着色和视频超分操作，最终输出高分辨率的彩色视频结果。

4.2 PC 前端用户界面设计

如图所示是我们设计的软件界面，它的功能是对用户选择的灰度视频进行着色，界面上包含四个按钮：选择输入（黑白影视）、选择参考、开始、退出。



图 4.1 用户界面

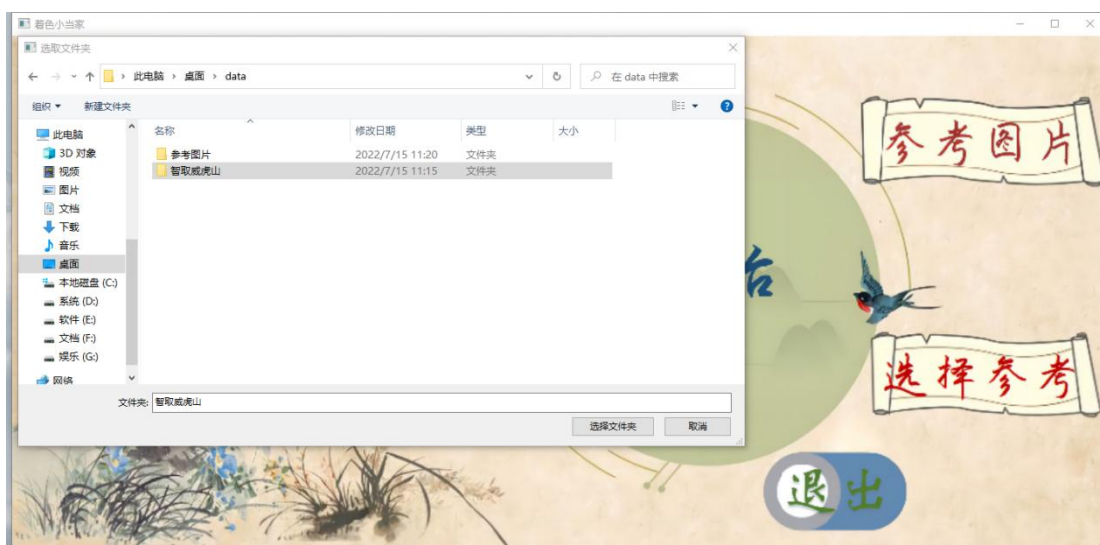


图 4.2 视频片段选取

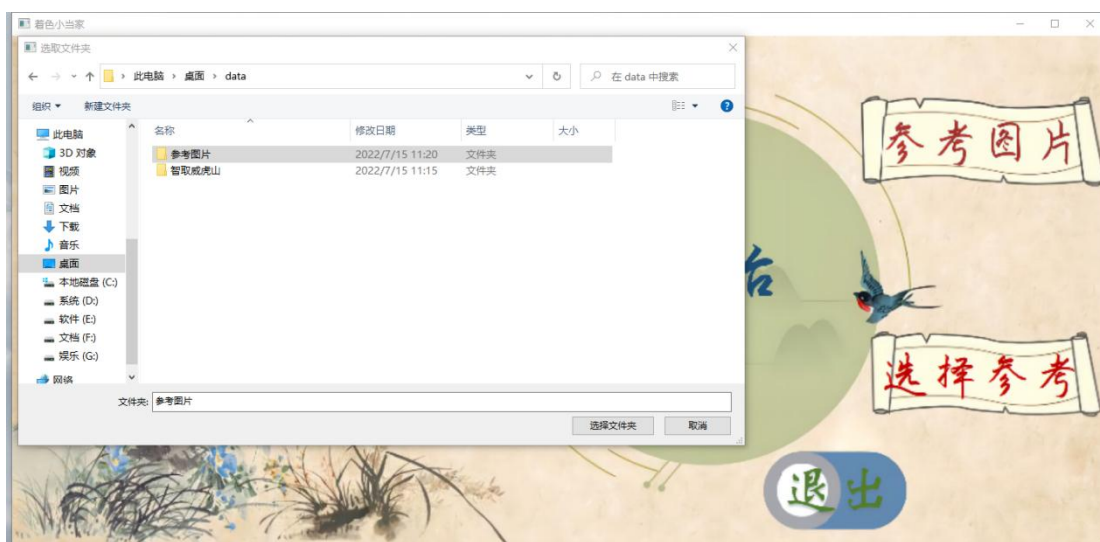


图 4.3 参考图片选取

4.3 PC 后端设计

后端计算处理主要包括着色模块，前端使用 PYQt5，配合 Python 实现控制逻辑。算法实现则主要通过着色网络和超分网络进行超分并着色。

软件开发与测试运行环境为 windows 10 操作系统，使用语言为 Python，编程环境为 Pycharm 和 PYQT5。

第 5 章 灰度影视作品超分辨率着色系统测试与分析

5.1 测试环境

本系统的测试环境如下表所示。

表 5.1 系统测试环境表

类型	名称	版本
系统	Windows	Windows 10 专业版
CPU	Intel	R7-4800H
GPU	Nvidia	NVIDIA RTX 2060
内存	Samsung	16G
硬盘	SanDisk	SDSSDH3 1T

5.2 视频数据集构建

5.2.1 着色训练集

为了覆盖广泛的场景，我们使用多个数据集进行训练。首先，我们从 Videostock 收集了 452 段视频，主要包含动物和景观。此外，我们还自己搜集了 248 个肖像视频，共 700 段视频用于训练。对于每个视频，我们通过查询 ImageNet 数据集中相应类中最相似的五幅图像来提供参考候选图像。

5.2.2 超分训练集

我们从互联网上下载了 110 段包括自然风景、人、动物和建筑等的 1080P HD (1080×1920) 视频，这些视频涵盖了复杂的场景转换和多样的运动形式。从每段视频中提取连续的 32 帧，总共 3520 帧构成训练样本。利用 Matlab 的 `imresize` 函数对原始视频帧进行两倍下采样获得 HR 视频帧 (540×960)，将 HR 视频帧四倍下采样得到相应的 LR 视频帧 (135×240)，成对的 HR/LR 视频帧构成训练集。

我们在网络中收集了众多经典的黑白电影，如《林海雪原》，《魂断蓝桥》等

等。这些电影通常是以胶片为存储媒介，之后进行数字化处理，录制下来的。中间过程会不断添加噪声进入影片，使得影片质量越来越差，高频细节出现丢失。

5.2.3 参考视频选择

着色网络对黑白影像进行着色时可以使用参考图片，参考图片越贴近黑白图像的内容，着色的效果越合理。本次我们对于林海雪原的参考图片选择的是电影《智取威虎山》以及已有的修复彩色版《林海雪原》。参考视频真实还原了那个年代的时装与环境配色，已经在 CCTC-6 频道播出过，收到不错的反响。我们把黑白版的林海雪原按场景镜头进行切分，每切换一次镜头，进行一次裁剪。在每个小部分用对应的彩色修复版截图进行参考。

5.2.4 测试数据

在测试时，选用互联网中收集的经典灰度影视作品。我们找到了《锦上添花》（1962），《湖上的斗争》（1955），《地道战》（1965），《英雄儿女》等等珍贵影像资料。这些电影中凝聚着那个时代珍贵的精神品格。

为检验我们方法的实用性，我们用训练后的网络进行测试。测试数据选择的是经典黑白电影——《林海雪原》（1960）、纪录片《开国大典》以及《智取威虎山》褪色后的版本，我们大约选取了一小时的电影片段，并将其分解为约 91361 帧来进行着色。并对着色结果进行超分辨率操作，得到高分辨率的彩色化电影片段。

5.3 测试结果的定性分析

5.3.1 《开国大典》结果及分析

我们对开国大典的珍贵视频资料进行着色处理。



(a) 参考图像



(b) 灰度序列

(c) 超分着色序列

图 5.1 纪录片《开国大典》超分辨率着色结果

选定参考图像，对黑白视频进行着色处理，得到上色后的视频。图 5.1 是纪录片《开国大典》着色效果的演示。可以看到，图像中衣物的颜色是一致的，体现了本文方法的时间一致性很好。如果参考图像更贴近于被修复的黑白图像，颜色的恢复会更加合理。

5.3.2 《林海雪原》结果及分析

对深入人心，脍炙人口的《林海雪原》进行着色，参考图像来源于央视播出过的《林海雪原》彩色修复版。更有针对性的参考帧，可以让网络对相近的特征进行更准确的匹配。我们把电影分割为若干片段，每个片段不超过 100 帧。当场景发生变化时，进行切分。对彩色版电影进行和黑白电影完全一致的切分。每一段黑白视频着色时使用一张参考图像，该参考图像取自对应的彩色视频片段。

老影新色：经典灰度影视作品的超分辨率着色



(a) 参考 1



(b) 参考 2



(c) 灰度帧 1



(d) 着色结果 1



灰度帧 2



(f) 着色结果 2

(e)

图 5.2 林海雪原修复结果部分展示 1



(a) 灰度序列

(b) 着色结果

图 5.3 林海雪原修复结果部分展示 2



(a) 参考图片



(b) 灰度帧

(c) 着色结果

图 5.4 林海雪原修复结果部分展示 3



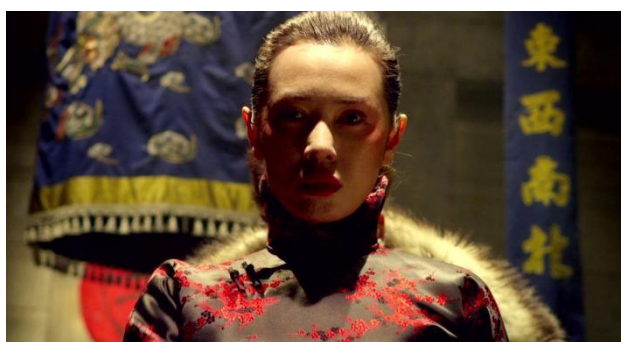
图 5.5 电影《林海雪原》超分辨着色成果截图

如图 5.2、5.3、5.4 和 5.5 所示展示了我们的测试结果，我们在实验中选用了大量的老电影视频来测试网络的性能，通过对《林海雪原》约十万张的灰度帧进

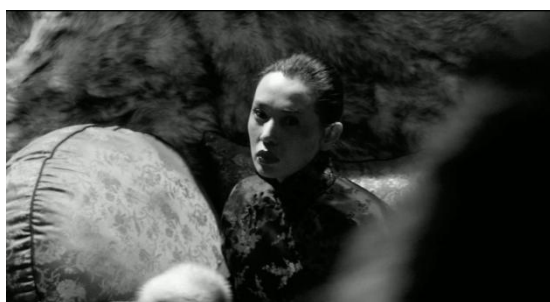
行超分着色，我们生成了一个多小时的彩色视频。通过观察分析，人物衣物的毛绒感恢复的很好，色调与参考图像保持一致。人物之间的颜色区别鲜明。背景物品颜色着色合理。图中伪影较少，同一场景内，相同物品颜色没有发生突变。室外自然景观的着色合理准确。总的来说，修复效果很好，可以带来良好的观影体验。

5.3.3 《智取威虎山》结果及分析

我们的方法在图像颜色和每个部分边界的准确性方面表现良好，从图 5.6 可以看出，我们的着色结果中的每个元素（如衣服、人脸等）都有一个清晰的边界，并在区域块内实现一致性，即衣服部分具有相同的颜色。色调与参考图像保持一致。人物之间的颜色区别鲜明，背景物品颜色着色合理。图中伪影较少，同一场景内，相同物品颜色没有发生突变。



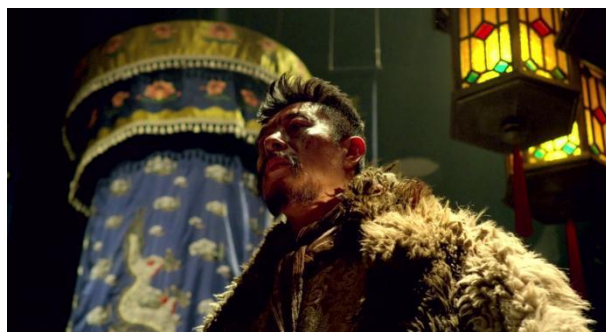
(a) 参考图片



(b)灰度序列



(c)着色结果



(d) 参考图片



(e) 灰度序列

(f) 着色结果

图 5.6 《智取威虎山》着色结果

5.4 测试结果的定量分析

图 5.7 给出了我们的方法和竞争视频着色方法 TCVC 在测试数据集上的着色对比结果。通过观察分析，我们的方法在图像颜色和每个部分边界的准确性方面表现良好，而 TCVC 存在图像着色不准确和衣服上颜色不连续块状等问题。从图 5.7 的第三行可以看出，我们的着色结果中的每个元素（如衣服、人脸等）都有一个清晰的边界，并在区域块内实现一致性，即衣服部分具有相同的颜色。相较于对比方法，图中伪影较少，室外自然景观的着色合理准确。



(a) 灰度帧 1

(b) 灰度帧 2

(c) 灰度帧 3



图 5.7 我们方法和 TCVC 方法着色结果对比

对于视频帧的着色，我们使用峰值信噪比（PSNR）和结构相似性（SSIM）来评估像素级生成的图像，而 Fréchet 初始距离（FID）评估语义级的图像分布和颜色。由表 5-2 可知，本文方法与竞争的视频着色方法 TCVC 和 DeepRemaster 相比，具有更高的 PSNR 和 SSIM 值，具有较低的 FID 值。相比于对比方法失真指标和感知指标，以及生成图像的真实性更好。

表 5.2 着色定量比较结果

Method	PSNR (DB) ↑	SSIM ↑	FID ↓
TCVC	17.383	0.856	21.595
DeepRemaster	24.234	0.871	18.580
Ours	27.051	0.892	15.860

图 5.8 给出了我们的方法在着色后的《智取威虎山》上的 4 倍 VSR 重建的结果。通过观察分析，本文方法可以恢复出低质量视频中的更多信息，如视频中的 人物和背景（如树枝）的清晰度显著提高。我们的方法可以生成合理且真实的细节纹理，并能够有效地克服视频运动伪影，重建出包含更多高频细节、边缘和纹理完整清晰、色彩真实，视觉感知更加愉悦的视频。

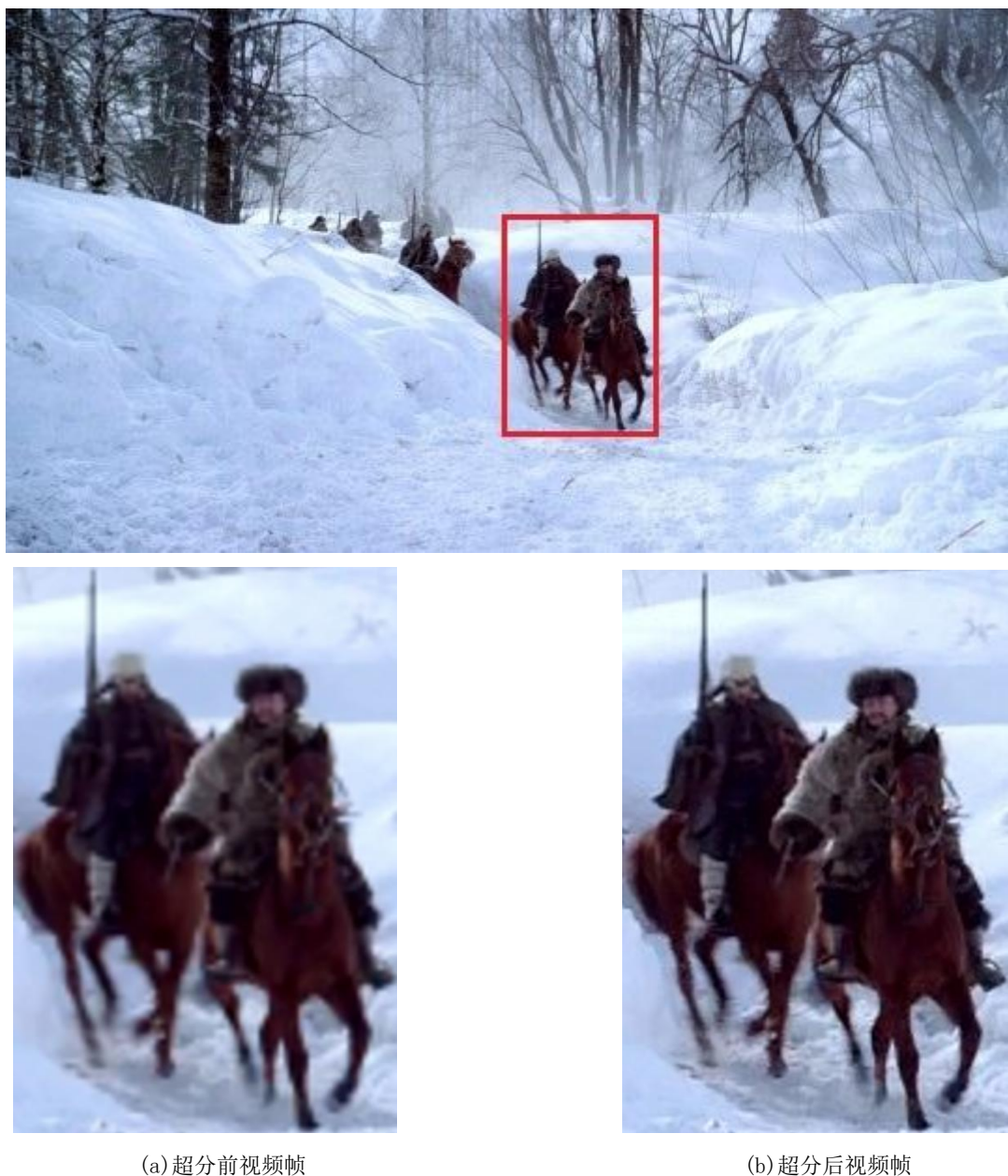


图 5.8 《智取威虎山》4 倍超分前后对比结果

表 5.3 和表 5.4 分别列出了本文方法以及对比方法在公共数据集 Vid 上 4 倍 VSR 重建的 PI、NIQE 和 PSNR 值。PI 和 NIQE 值越低表示视觉感知质量越好，PSNR 值越高表示失真度越低。由表 5-3 和表 5-4 可知，本文方法与竞争的 VSR 方法 VSRnet，VESPCN，TDVSR 和 SOF-VSR 相比，具有低的 PI 值。本文方法与 SISR 方法 EnhanceNet 和 SRGAN 相比，具有高的 PSNR 值。本文方法相比于 VSR 方法感知指标更好；相比于 SISR 方法失真指标更好。

表 5.3 测试数据集上四倍 SR 重建定量比较结果

评价指标	VSRnet	VESPCN	TDVSR	EnhanceNet	SRGAN	SOF-VSR	Ours
PI	6.181	6.233	5.937	2.942	2.511	5.812	3.817
NIQE	6.560	6.516	6.397	3.299	2.417	6.460	4.281
PSNR	24.840	25.350	25.360	22.752	23.019	26.010	24.970

表 5.4 测试数据集上四倍 SR 重建定量比较结果

评价指标	EnhanceNet	SRGAN	SOF-VSR	MC-PETGAN
PI	2.743	2.577	5.512	3.766
NIQE	3.263	2.702	5.582	4.068
PSNR	27.362	28.175	30.742	29.511

第6章 总结与展望

本文提出一种基于深度学习方法灰度影视作品的超分辨率着色方案，并实现了电影《林海雪原》和《智取威虎山》，以及纪录片《开国大典》的超分辨率着色。该方案是一个在参考图像引导和时空特征补偿约束下的端到端视频超分辨率着色网络框架，其统一了语义对齐、颜色传播和光流补偿模块，通过语义对齐与颜色传播将参考图像的颜色传播至灰度视频，使其上色；光流补偿网络则利用相邻视频帧之间的短时连续和内容相似性特征为当前帧着色（当前待着色帧）提供有效丰富的细节信息；融合上色帧及细节信息可获得具有时间连续性和高感知质量的彩色视频。之后，我们对着色视频超分辨率重建，提高视频清晰度。

经典灰度电影《林海雪原》和《智取威虎山》，以及纪录片《开国大典》的成功超分辨率着色结果表明，我们的方法不仅可以生成合理且真实的细节纹理，而且有效地克服了视频运动伪影，确保局部信息的过度自然流畅。此外，色彩逼真，人物之间的颜色区别鲜明，且色调与参考图像保持一致。背景物品颜色着色合理，同一场景内，相同物品颜色没有发生突变。测试结果证明我们的技术路线可行，且可以推广应用于其他老旧灰度影视作品的超分辨率着色，可实现影视文化遗产的数字化保护、弘扬与传承，促进传统文化与时俱进、创新发展，同时也能增强国家软实力和国际影响力。

参考文献

- [1] 李红俊, 韩冀皖. 数字图像处理技术及其应用[J]. 计算机测量与控制, 2002, 10(9): 620-622.
- [2] 李玉润. 灰色图像着色方法研究及实现[D]. 云南大学, 2013.
- [3] Li H, Chen C, Fang S, et al. Brain MR image segmentation using NAMS in pseudo-color[J]. Computer Assisted Surgery, 2017, 22(sup1): 170.
- [4] 金舟. 图像着色关键技术分析及其应用[D]. 天津大学, 2011.
- [5] 郑南宁. 计算机视觉与模式识别[M]. 国防工业出版社, 1998.
- [6] 王红梅, 张伟, 张西学. 灰度图像变换处理技术的研究及开发[J]. 科技创新导报, 2008(29): 244-245.
- [7] 许宝升. 彩色 CRT 色彩特性及色彩空间转换方法的研究[D]. 西安理工大学, 2005.
- [8] 滕升华. 基于颜色扩展的图像彩色化技术研究进展[J]. 信息技术, 2013(12): 75-79.
- [9] 陈小娥. 基于色彩直方图匹配的颜色传递算法研究[J]. 武夷学院学报, 2017, 36(9): 44-48.
- [10] Levin A, Lischinski D, Weiss Y. Colorization using optimization[C]//W/ACM Transactions on Graphics (TOG). ACM, 2004, 23(3): 689-694.
- [11] Wang L, Shi S, Jin W, et al. Color fusion algorithm for visible and infrared images based on color transfer in YUV color space[C]//MIPPR 2007: Multispectral Image Processing. International Society for Optics and Photonics, 2007, 6787: 67870S.
- [12] Qu Y, Wong T T, Heng P A. Manga colorization[C]//W/ACM Transactions on Graphics (TOG). ACM, 2006, 25(3): 1214-1220.
- [13] Luan Q, Wen F, Cohen-Or D, et al. Natural image colorization[C]//Proceedings of the 18th Eurographics conference on Rendering Techniques. Eurographics Association, 2007: 309-320.
- [14] Kolling A, Walker P, Chakraborty N, et al. Human Interaction With Robot Swarms: A Survey[J]. IEEE Transactions on Human-Machine Systems, 2016, 46(1): 9-26.
- [15] Welsh T, Ashikhmin M, Mueller K. Transferring color to greyscale images[C]//W/ACM Transactions on Graphics (TOG). ACM, 2002, 21(3): 277-280.
- [16] Bonneel N, Tompkin J, Sunkavalli K, et al. Blind video temporal consistency[J]. ACM Transactions on Graphics (TOG), 2015, 34(6): 1-9.
- [17] Lai W S, Huang J B, Wang O, et al. Learning blind video temporal consistency[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 170-185.
- [18] Sheng B, Sun H, Magnor M, et al. Video colorization using parallel optimization in feature space[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2013, 24(3): 407-417.
- [19] Doğan P, Aydın T O, Stefanoski N, et al. Key-frame based spatiotemporal scribble propagation[C]//Proceedings of the Eurographics Workshop on Intelligent Cinematography and Editing. 2015: 13-20.
- [20] Paul S, Bhattacharya S, Gupta S. Spatiotemporal colorization of video using 3d steerable pyramids[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2016, 27(8):

1605-1619.

[21]Jampani V, Gadde R, Gehler P V. Video propagation networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 451-461.

[22]Vondrick C, Shrivastava A, Fathi A, et al. Tracking emerges by colorizing videos[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 391-408.

[23]Chan K C, Wang X, Yu K, et al. BasicVSR: The search for essential components in video super-resolution and beyond[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 4947-4956.

[24]Isobe T, Zhu F, Jia X, et al. Revisiting Temporal Modeling for Video Super-resolution[J]. arXiv e-prints, 2020: arXiv: 2008.05765.

[25] Iizuka S, Simo-Serra E. Deepre-master: temporal source-reference attention networks for comprehensive video enhancement[J]. ACM Transactions on Graphics (TOG), 2019, 38(6): 1-13.

[26] Zhang B, He M, Liao J, et al. Deep exemplar-based video colorization[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 8052-8061.

[27] Jampani V, Gadde R, Gehler P V. Video propagation networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 451-461.

[28]Liu S, Zhong G, De Mello S, et al. Switchable temporal propagation network[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018: 87-102.

附录 1

1.1 经典灰度影视作品的超分着色系统的社会价值

党的十八大以来，以习近平同志为核心的党中央把文艺工作摆在重要位置，习近平总书记主持召开文艺工作座谈会，党中央印发《关于繁荣发展社会主义文艺的意见》、国务院办公厅印发《关于支持戏曲传承发展的若干政策》等文件，推动我国文艺事业进入新的发展阶段。老电影的色彩修复也成为文艺工作蓬勃发展的重要组成部分。如：在 2021 年 10 月我国首部黑白转彩色电影《永不消逝的电波》重映。由此可以看出，老电影的修复工作具有重要的社会价值与意义，新中国成立初期，受限于技术条件很多优秀的电影作品只能拍摄成黑白。但经典的故事永远让人难忘，彩色画面具有更加强了老电影在细节处的表达，同时增强了老电影的传播性，便于观众读懂传统的历史与现实意义。通过我们的研究希望让经典电影焕发出新的生命力，让理想永不褪色。

党的二十大报告提出了“两个一百年”奋斗目标和中华民族伟大复兴中国梦。利用数字技术对传统文化进行保护和传播，既能夯实中华民族共同精神家园、增强民族自信心和自豪感，又能推动经济社会发展、科技创新和文明交流互鉴。利用数字技术对传统文化进行保护和传播，既能满足广大人民群众对优秀传统文化的需求，又能促进传统文化与时俱进、创新发展，同时也能增强国家软实力和国际影响力。

附录 2

2.1 证明材料

我们的项目在第十七届研电赛中荣获西北赛区二等奖，证明该项目具有一定的科学性、先进性与实用性。

