

北京邮电大学

# 大学生创新训练计划项目 结题报告

项目编号: 202401005 项目级别: B

项目类别: 目标导向类☒ 自主探索类☐

校企合作类☐ 滚动支持类☐

项目中文名称: VIVIDIFY 低功耗智能超分相机

项目名称(英文): VIVIDIFY Low-Power Intelligent  
Super-Resolution Camera

项目依托学院: 信息与通信工程学院

项目负责人: 薛皓林

联系电话: 13627601819

E-Mail: xhl518@bupt.edu.cn

指导教师: 景文鹏

E-Mail: jingwenpeng@bupt.edu.cn

2025 年 6 月 10 日

## 填写说明

- 1、 本结题书所列各项内容均须实事求是，认真填写，表达明确严谨，简明扼要。
- 2、 结题书由正文和附件两部分组成：正文部分请按表格要求填写，并可根据需要加页，要求层次分明、内容准确，如实反映项目研究过程中的进展或成果、计划调整情况等；附件包括成果报告书和是已取得的阶段性成果，包括发表的论文、研究报告或调研报告以及其他成果。
- 3、 结题书为大 16 开本（A4），左侧装订成册。可在网上下载、自行复印或加页，但格式、内容、大小均须与原件一致。
- 4、 负责人所在学院认真审核后，签署意见后，将申请书（一式两份）报送教务处。

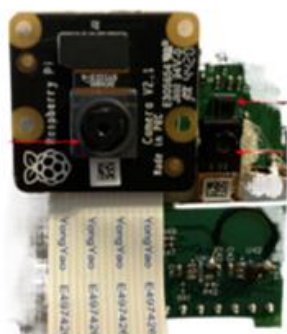
一、 项目基本情况

项目名称		VIVIDIFY 低功耗智能超分相机				
项目 主要 研究 成员	序号	姓名	学号	专业	学院	项目分工
	1	薛皓林	2022210571	信息与通信工程	信息与通信工程学院	项目进度进行把控 构建完成神经网络超分模型的搭建与优化 完善注意力机制 实现超分相机硬件部分的设计和代码实现。
	2	闫康	2022210566	信息与通信工程	信息与通信工程学院	构建搭建超分重构模型 集成关键算法 一起完善双模相机硬件部分的功能实现
	3	胡帅	2022210561	信息与通信工程	信息与通信工程学院	负责光流对齐机制的实现，和项目软硬件之间的连接
	4	罗玉杰	2022210596	信息与通信工程	信息与通信工程学院	负责视角校正算法的实现，即固定单应性变换

二、 项目执行情况简介

<p>(一) 项目研究的目的</p> <p><b>1. 项目来源：</b></p> <p>随着互联网的迅猛发展，短视频已经成为我们日常生活中不可或缺的一部分，短视频创作者也逐渐壮大。对于短视频博主，尤其是<b>旅行博主，vlog 博主，户外博主</b>来说，他们往往需要长时间拍摄高分辨率、高质量的视频素材。高质量的视频素材不仅能提升内容的吸引力，也是创作高质量作品的关键。然而，这样的拍摄需求往往伴随着<b>高功耗</b>导致的<b>电池续航时间缩短</b>的问题。在户外复杂不断变换的场景中，电源并不总是容易获得，这极大限制了博主们的<b>创作自由度</b>。现有摄像设备在满足高质量视频捕获的同时，往往难以实现低功耗长时间运行。这意味着在追求高分辨率和色彩表现的同时，博主们不得不面对电池续航的挑战。受限于硬件和成本，现有解决方案在提供长时间续航和高质量视频输出之间难以找到平衡点。</p>
--

所以，我们希望研发一款创新性的产品，基于关键帧的**视频超分双模式摄像系统**，采用“**低功耗黑白视频流+间歇式高分辨率关键帧**”架构，兼顾视频质量和续航时间。该产品专为短视频创作者设计，特别是那些经常在户外拍摄的旅行、生活以及其他领域的博主。系统利用双模式摄像的模式结合神经网络超分重构技术，不仅确保了视频内容的超高清分辨率，而且因为双模系统带来的低功耗，还大幅延长了拍摄时长，让博主们在创作过程中能够维持高标准的艺术表现力，并享受到更加自由灵活的拍摄体验。这种灵活性不仅扩展了创作的边界，还为探索新的创作可能性提供了广阔的空间。我们将我们的想法命名为的**"VIVIDIFY 低功耗智能超分相机"**。



内置双摄像头  
无线传输模块

图 1 VIVIDIFY 低功耗智能超分相机成品图

## 2. 响应国家政策号召：

我们的项目**"VIVIDIFY 低功耗智能超分相机"**深度契合国家政策、国家战略规划，在技术创新和应用场景层面形成与政策导向的精准呼应。

**绿色低碳：**根据《**2030 年前碳达峰行动方案**》的要求，电子设备需降低能耗并推广绿色设计标准，同时工信部《**智能硬件产业创新发展专项行动**》也明确提出鼓励智能摄像设备的低功耗设计。传统摄像设备因频繁捕捉和处理高分辨率图像，往往面临高能耗和散热问题，难以满足长时间稳定使用的需求。我们提出的**双模摄像系统**，采用“低功耗黑白视频流+间歇式高分辨率关键帧”架构，兼顾视频质量和系统节能。低功耗模式下，摄像头持续输出黑白低分辨率视频流，保证画面连贯性；高功耗模式则以低频率间歇性工作，捕获彩色关键帧，大幅减少数据处理量，**有效降低了系统能耗**。同时，黑白视频流的低功耗输出和云端服务器的超分重构处理，进一步优化了能耗表现。该方案创新性地实现了**视频拍摄的高画质需求与低能耗长续航兼得**，为电子设备的绿色设计提供了新的解决方案，完全符合国家绿色低碳政策对智能硬件的要

求，有望成为未来智能摄像设备发展的新方向。

**智能制造：《新一代人工智能发展规划》**，明确将人工智能作为核心驱动力，提出构建自主可控的技术体系，重点支持智能感知、深度学习算法、边缘计算等方向。鼓励将人工智能技术与现有的关键技术相结合，实现人工智能的实际领域应用。本项目深度践行《新一代人工智能发展规划》"构建自主可控技术体系"的战略部署，我们采用的 NeuriCam 的超分重构模型，将先进的深度学习算法与实际应用领域深度结合，致力于开发出满足市场需求且具有自主知识产权的创新技术。**在智能感知方面**，项目通过双模式摄像系统实现高效视频采集；**在深度学习算法上**，利用超分重构和注意力机制提升视频质量；**在边缘计算领域**，超分重构模型利用边缘设备的计算能力，实现了实时视频高效处理。我们实现了多种关键技术融合，不仅为在视频拍摄领域提供了新的人工智能技术解决方案，也为推动相关产业的智能化升级贡献了力量。

本项目具有显著的经济效益，在短视频创作领域，该系统能够为创作者降低约 60% 的设备成本（与主流 4K 摄像机相比），超过百万的内容生产者成为潜在客户，精准契合国家高质量发展的需求。随着短视频行业的爆发式增长，内容创作者对高性能、低成本拍摄设备的需求日益迫切。本项目通过创新的双模摄像系统，不仅大幅降低了设备成本，还提升了拍摄质量与创作灵活性，助力创作者产出更优质的内容，增强了市场竞争力。这不仅为创作者带来了实实在在的经济效益，也为整个行业注入了新的活力，推动了边缘计算，光学模组等产业的高质量发展，与国家推动产业升级和经济结构优化的战略方向高度一致。

### 3. 解决行业痛点，满足市场需求：

**短视频行业磅礴发展，短视频拍摄设备市场随之崛起：**短视频用户规模持续扩大，成为推动行业增长的重要因素。中国互联网络信息中心发布的第 54 次《中国互联网络发展状况统计报告》显示，截至 2024 年 6 月，我国网络视频用户规模已接近 11 亿大关，达到 **10.9967 亿人**，其中短视频用户规模达 **10.68 亿人**，较 2022 年同期增长 7796 万人，**占网民整体的 97.1%**，从下图中，我们可以清晰地看到中国互联网网民规模和互联网普及率的历年变化，以及网络视频用户的增长和占比情况。这些数据揭示了近年来短视频市场的迅猛扩张。



图 2 全国网民规模和互联网普及率



图 3 全国短视频用户规模

**国内市场爆发:** 根据相关调研数据, 预计到 2050 年, 短视频行业的市场规模将实现数倍增长。以我国为例, 根据最新发布的《2025 年中国短视频行业研究报告》显示, 2025 年我国短视频市场规模将达到 **500 亿元美元**, 占全球市场份额的 20%。在未来 25 年内, 这一比例有望进一步提升。

**创作者激增:** 随着短视频行业的爆炸性增长, 短视频内容创作者也如雨后春笋般涌现。以抖音、哔哩哔哩为代表的平台, **2020 年**, 仅抖音一个平台, 创作者总量已经突破惊人的 **1.3 亿**, 且保持着年均 12% 以上的高速增长。在竞争白热化的创作生态中, 内容创作需求正呈现井喷式爆发。



图 4 21 年哔哩哔哩创作者总量

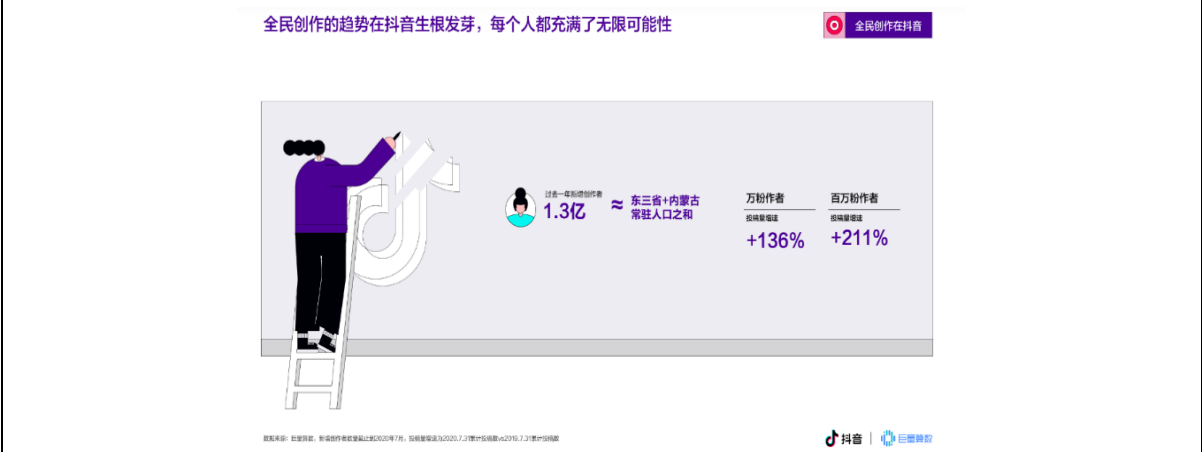


图 5 20 年抖音创作者总量

**市场竞争加剧：**虽然用户规模持续增长，但其增速逐步放缓，用户增长的红利逐渐消退，行业真正步入存量竞争的时代。在用户规模增长放缓的同时，短视频用户黏性不断增强，2023 年短视频人均单日使用时长达到 151 分钟，短视频应用平台的头部效应尤为凸显，抖音、快手两大短视频平台占整体份额的 95.3%。随着用户对网络视听内容产品的需求变化，高质量、精品化、个性化的内容生产成为短视频发展的未来方向。短视频为网络视听行业的转型升级提供了新的路径，成为媒体深度融合的重要力量。短视频的精品化转向是必然趋势。短视频内容新颖，高质量，精品化才是提高竞争力的关键。粉丝超过 1000 万的博主在主流平台上的占比仅为 0.01%，这一数据反映了内容创作领域的强烈竞争。

短视频行业的爆发式增长，短视频创作者数量急剧增加，短视频行业竞争加剧，短视频内容高质量要求，持续推动拍摄设备需求升级。2025 年全球短视频拍摄设备市场规模预计突破 480 亿美元，中国贡献超 35% 份额，2025 年中国摄影器材行业市场规模持续扩大，展现出强劲的增长动力。随着创作者数量激增和短视频行业激烈竞争，创作者对高品质摄影设备的需求日益增长。硬件创新与场景细化成为行业核心驱

动力。设备轻量化、智能化趋势显著，4K/8K 超高清、HDR 成像、多镜头协同等技术渗透率较 2023 年提升 22 个百分点，用户对便携性与专业功能的双重需求倒逼产业迭代。

短视频创作者对于视频拍摄设备主要需求为，**高画质、高视频质量，小巧便捷，长续航时间，高性价比。**

而传统相机行业对于短视频创作者的需求应对乏力，并存在如下痛点：

**“高清画质与续航不可兼得”的续航痛点：**传统高分辨率拍摄设备面临着高功耗困境，想要**高画质**，就必须**高功耗低续航**。传统相机拍摄高分辨率视频需自身持续处理大量数据，导致功耗激增。创作者想要拍摄出高质量的视频素材就必须舍弃长时间拍摄的可能。特别户外拍摄时需要频繁更换电池或中断创作。据实测，4K 拍摄模式下手机续航普遍低于半小时，甚至有的安卓手机只能拍摄十几分钟视频，变回自行退出拍摄，**4K 相机续航低于一个小时**。现有低功耗方案（如 AOV 技术）为省电采用“1 秒 1 帧”的极低帧率，虽延长续航但牺牲流畅度，高速运动场景漏拍率达 30%以上。对于想要持续拍摄高质量，高分辨率，流畅的创作者而言，这些都是致命的缺点。

**“动态场景模糊与细节失真”的画质痛点：**传统摄像行业存在一个通病：算法局限导致**伪影与失真**，传统相机和手机在高倍变焦时易出现文字扭曲、人脸失真等问题，因此依赖于 AI 超分技术生成式补全，非真实光学数据，这个过程不但会增加设备的功耗，同时也会造成图像失真。同时传统相机行业还面临一大问题，**运动模糊**难以消除，拍摄高速运动的物体会导致重影，非常影响画面质量，以及素材质量。

**“操作复杂，携带不便”的携带痛点：**传统相机**通常功能繁多，设置复杂**，需要用户具备一定的摄影知识和经验才能熟练操作。对于非专业用户来说，学习如何正确设置拍摄参数、调整镜头光圈、快门速度等非常耗时，增加了使用难度。而且传统相机体积较大，特别是专业级单反相机，镜头和机身重量重，携带起来非常**不便**。这给用户在移动拍摄过程中**增加了负担**，尤其在需要频繁移动或在复杂地形中拍摄时，如户外拍摄或旅游 vlog，用户不得不携带多个镜头和配件，进一步增加了使用上的不便。





图 6 佳能专业级单反摄像机的体积

**“高清画质与低价格不可兼得”的价格痛点：**传统高清摄像机，尤其支持 4K/8K 分辨率的专业设备，因采用高品质传感器、镜头及图像处理器，成本高，一般价格在 8000 左右，价格动辄数万元，对个人创作者和小型团队都是沉重经济负担。低价摄像机为控制成本，常在关键部件上妥协，如采用小尺寸传感器、低性能镜头和简陋图像处理器，致使其在分辨率、色彩还原、低光表现等关键画质指标上表现欠佳，无法满足高质量视频创作的需求。



索尼A6700 套机 (E PZ 16-50mm F3.5-5.6 OSS) 2600万像素, 4K视频, E

参考价: **¥10599**

★★★★★ 9.9 | 252人点评

**T** ¥9499



索尼A7CR 约6100万像素全画幅背照式, AI智能芯片, 4K 60p

参考价: **¥20999**

★★★★★ 10.0 | 114人点评

**JD** ¥1.91万 **T** ¥2.1万

图 7 索尼 4K 摄像机价格

为解决满足创作者消费需求，填补消费市场空缺，解决传统相机行业行业所面临的痛点，我们为降低相机拍摄低功耗，降低摄像机成本，提高摄像机拍摄质量，提高高清摄像机的便携性，我们提出了 **VIVIDIFY 低功耗智能超分相机**

相比于传统相机行业我们的 VIVIDIFY 低功耗智能超分相机存在如下优势：

**“双模协同”解决续航痛点，**我们提出 VIVIDIFY 的创新方案，利用双模式协同设计，低功耗摄像头持续输出 15fps 黑白低分辨率视频流，保障基础流畅性；高功

耗摄像头间歇捕获 1fps 高分辨率彩色关键帧，提供还原细节，双模式设计显著降低设备功耗。同时我们在边缘计算云服务器上部署了我们的超分重构模型。VIVIDIFY 相机身并不需要自己进行视频处理，只需要将视频数据传输到边缘计算服务器，在边缘计算服务器中，会进行基于关键帧的视频超分，关键帧与黑白视频流在云端实时融合生成高清 4K 视频，进一步降低自身功耗。根据我们的实际测试，VIVIDIFY 低功耗摄像机设备端功耗降低了 70%以上，续航延长 5-7 倍，可以达到 7 个小时的高强度的长时间拍摄，并在云端实时输出拍摄的 4K 视频。VIVIDIFY 低功耗相机完美的解决了“高清画质与续航不可兼得”的功耗痛点，为视频创作者提供了更好的创作环境，也提供了无限的创作可能。

**“超分重构”解决画质痛点**，VIVIDIFY 低功耗超分相机的在这方面取得突破性技术，多算法融合架构，我们在基于关键帧的超分模型中，融入了注意力特征滤波器机制，动态分配关键帧权重，精准还原色彩与高频细节，减少画面失真；我们利用光流对齐+变形卷积，彻底解决高速运动模糊，位移误差控制在<5%，极大程度上的减少了画面模糊，杜绝了重影现象；同时我们构建了双向循环网络，超分时融合前后帧上下文信息，进一步提升运动连贯性。

**“小巧轻便”解决携带痛点**，VIVIDIFY 低功耗超分相机配备了直观且用户友好的界面，用户无需具备复杂的专业知识，即可快速上手进行拍摄，降低了使用门槛。VIVIDIFY 低功耗超分相机能拍摄与专业单反摄像机画质一样的视频，不需要特别专业的镜头和其他摄像装备，所以 VIVIDIFY 超分相机机身小巧轻便，易于携带，易于操作，特别适合在移动拍摄过程中使用。用户可以轻松地将其放入背包或口袋中，随时随地进行拍摄。在面临各种复杂地形和环境中拍摄时更加灵活，无需担心设备的重量和体积限制。

**“超高性价比”解决价格痛点**，VIVIDIFY 超分相机采用创新的双模摄像系统和超分重建技术，在保证高清画质的同时，大幅降低设备成本。VIVIDIFY 超分相机并不需要专业 4K 单反相机一样的高品质传感器、镜头及图像处理器（本地不需要处理图像数据），所以 VIVIDIFY 超分相机大幅降低成本，其价格显著低于传统专业级 4K/8K 摄像机，使更多用户能够以较低的价格获得高质量的拍摄设备。通过“低功耗黑白视频流+间歇式高分辨率关键帧”架构和先进的超分重建算法，VIVIDIFY 超分相机能够输出高质量的 4K/8K 视频。其动态场景捕捉能力和低光性能出色，确保在各

种复杂环境下均能提供清晰、细腻的画面。

综上所述，VIVIDIFY 超分相机不但填补了市场低功耗长续航 4K 相机的空缺，而且完美符合短视频创作者的创作需求，是短视频创作者的不二选择

## （二）成果的主要内容

本项目构建了一个面向短视频拍摄场景和物联网应用场景的低功耗高质量视频采集与画质增强系统——VIVIDIFY 低功耗超分相机。围绕“关键帧驱动的双模式摄像+视频超分辨率”这一核心思想，项目在系统架构、关键算法模块及成功协同等方面取得了以下主要成果：

项目整体架构：

我们的项目主要由主要分为两部分：**双模式摄像系统**，由两个协同工作的摄像头（“低功耗黑白视频流+间歇式高分辨率关键帧”）和**边缘计算云服务器**（部署着我们的**超分重构模型**）组成。在拍摄时两种模式的摄像头同时运行，低功耗模式下摄像头持续运行，连续输出的每秒 15 帧黑白低分辨率视频流，保证了画面的连贯性。而**高功耗模式**下的摄像头则以每秒一次的频率间歇性工作，捕获高分辨率的彩色关键帧。在拍摄过程中，系统会同时将拍摄的黑白视频流和关键帧的图像数据通过 **TCP 连接**，传输到**边缘计算云端服务器**进行实时超分重构。通过基于关键帧的超分重构模型，利用高分辨率的彩色关键帧对黑白低分辨率视频流进行超分操作，实时输出高分辨率的视频。

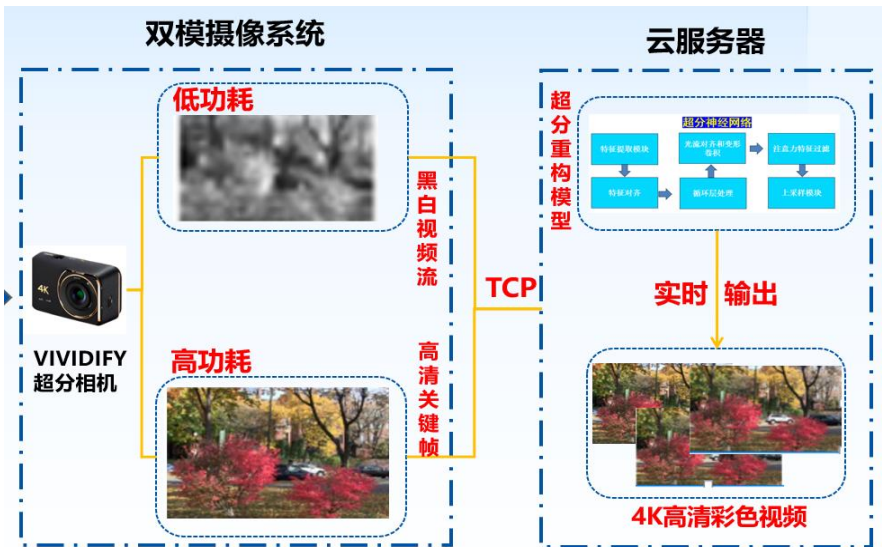


图 8 架构示意图

VIVIDIFY 超分相机关键技术：

### 1. 双模式摄像系统

本项目基于 OV2640 摄像头（高/低功耗模式）与 STM32F407 控制板构建了一个双模图像采集终端，并结合 ATK-MW8266D 无线模块实现图像远程传输，以一个 5000 毫安时的充电宝进行供电。

如下图所示，本系统设计充分利用 OV2640 摄像头灵活的分辨率配置能力，支持按需选择拍摄模式（高功耗/低功耗），以适应不同的功耗。我们使用两块 STM32F4 微控制器通过 I2C 分别配置**两个摄像头参数**，控制其中一个 OV2640 工作在**低功耗模式下**，连续输出的每秒 15 帧黑白低分辨率视频流，保证了画面的连贯性。控制另一个工作在**高功耗模式**下的摄像头则以每秒一次的频率间歇性工作，捕获高分辨率的彩色关键帧。并使用 SPI 接口实时读取图像帧缓存，随后借助 **ATK-MW8266D 模块**将图像数据通过 **TCP 连接**上传至服务器。在服务器端，我们部署了超分重构模型。经过实际测试可以发现，与传统视频采集方案相比，VIVIDIFY 超分相机在功耗受限的物联网场景下，通过“终端采集+云端重建”的协同处理机制，显著终端负载**功耗降低了 70%以上**，**续航延长 5-7 倍**，可以达到**7 个小时**的高强度的长时间拍摄，同时保证了图像质量。

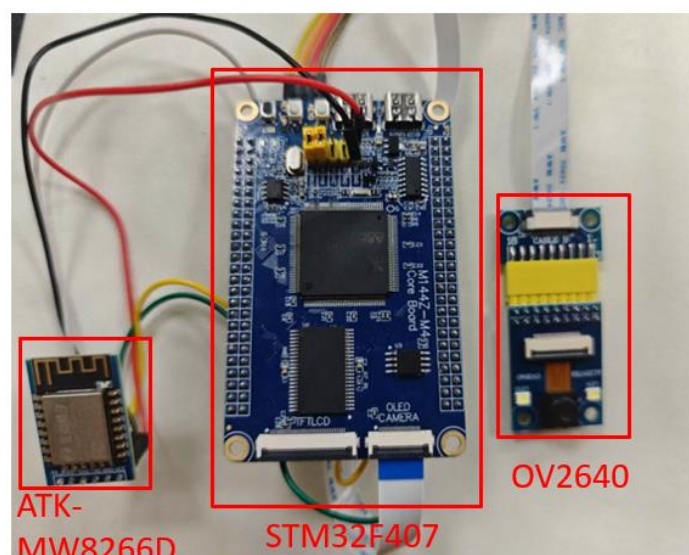


图 9 低功耗模式摄像头终端实际搭建效果图

在系统实现过程中，我们采用 Keil 5 集成开发环境完成了 STM32F4 主控端的固件开发与调试。如下图所示，工程文件结构清晰，包含主控逻辑（main.c）、外设驱动（如 usart.c、atk\_mw8266d\_uart.c）以及系统初始化模块（如 system\_stm32f4xx.c）。

核心配置参数中包括 Wi-Fi 模块的连接信息，如 SSID、密码、服务器 IP 地址及通信端口号，系统串口波特率设置为 921600，以满足高速图像数据传输需求。图像采集方面，主控 MCU 通过 I2C 接口对 OV2640 摄像头进行配置，控制其输出所需分辨率和图像格式。图像数据则通过 DCMI 接口读取，并结合 DMA（Direct Memory Access）进行高速传输。DMA 的引入显著降低了 CPU 负载，使系统能够在不中断主程序执行的情况下完成连续图像帧的高效采集。图像采集完成后，数据通过串口发送至 ATK-MW8266D Wi-Fi 模块（基于 ESP8266），并通过 TCP 协议将图像上传至指定服务器。在 Keil 编译环境下，程序构建成功，.axf 可执行文件顺利生成，无错误、无警告。

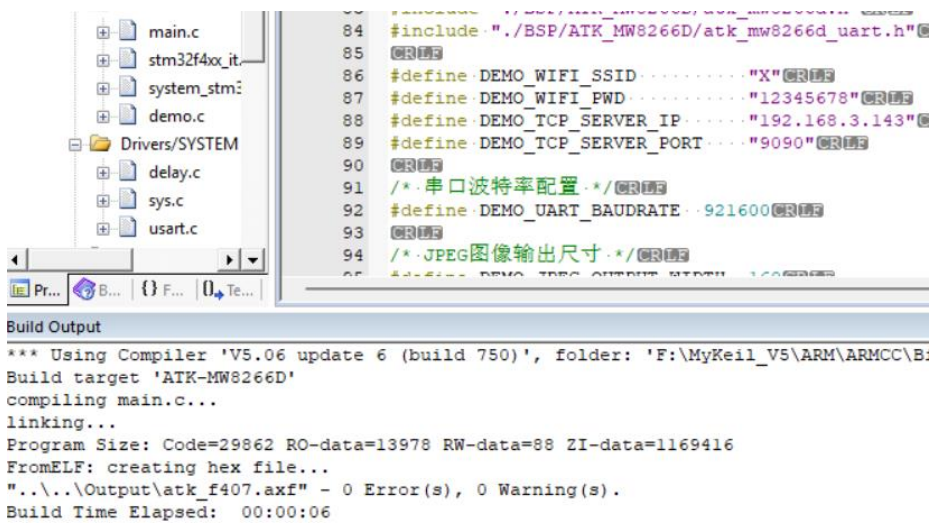


图 10 实验代码成功烧录

在系统通信调试阶段，我们采用 ATK 官方提供的 XCOM V2.6 串口调试工具对 STM32F4 与 ESP8266 模块之间的数据传输进行了测试与验证。如下图所示，在开启串口后，系统稳定输出连续的十六进制数据流，代表摄像头图像或数据包已通过 ESP8266 成功发送至主机端。

该过程验证了主控 MCU 的串口发送功能、ESP8266 的 TCP 串口传输能力、以及整体通信链路的完整性。该调试过程为后续图像接收提供验证。







图 13 服务器成功接收到的视频图像

与传统 4K 相机相比，为了获得高画质，相机需要不断地捕捉和处理高分辨率的图像内容，需要处理大量的数据，因而导致相机的功耗显著增加，相机续航时间减少，续航时间很短。而双模式摄像机以更低的频率获取高分辨率图像，大大减少了相机的数据处理量，而拍摄低分辨率视频功耗也很低，超分重构过程在云端进行，也不会额外增加相机本身的功耗。在这样三重保险下，双模式摄像系统对比于普通相机，在拍摄同样的高分辨率 4K 视频时，相机本身的功耗会大幅降低并且大幅度延长了相机的工作时间。

## 2. 前后双循环网络架构

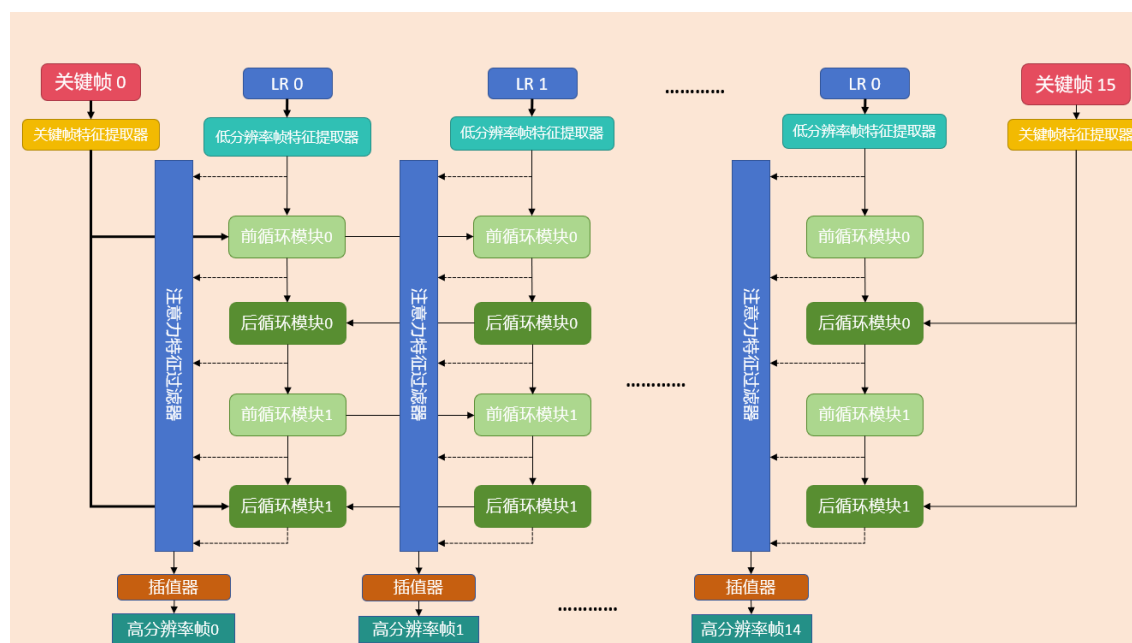


图 14 双向循环神经网络总架构

如上图所示，为了更好的还原图像细节，在超分辨率重建时保持视频的连贯性和

**清晰度**。我们构建了**双向循环网络架构**，这个架构能够允许我们的模型在视频重建过程中同时考虑过去和未来的帧，从而在时间维度上实现了全面的信息利用。双向循环网络架构通过深度整合前后帧的上下文信息，显著提升了对场景的理解。这种深度整合不仅包括视觉信息，还包括运动信息，为视频重建提供了丰富的数据支持，使得视频的每一帧都能够在更宽广的上下文中被理解和处理。对于**动态场景**，尤其是快速运动的物体，双向循环网络能够更好地捕捉和预测物体的运动轨迹，从而在超分辨率重建时保持物体的连贯性和清晰度。

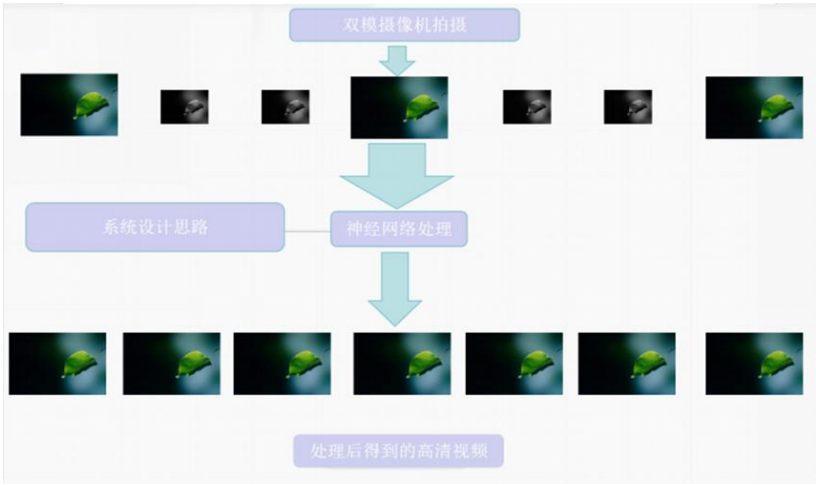


图 15 双向循环神经网络实际效果图

在双循环网络架构中，我们构建了一个**多轮双向特征传播机制**，显著增强了视频超分辨率模型对长时间序列的建模能力与特征对齐精度。如图所示，系统基于 **deformable alignment module**（变形卷积光流对齐模块）构建 **forward**（前循环模块）与 **backward**（两个传播路径），每轮传播都依赖前一轮生成的特征空间进行迭代更新。相比于传统特征传播，双向多轮时序传播结合当前帧特征、历史传播特征以及正反方向光流信息共同输入对齐模块，生成新的对齐特征（**feat\_prop**），用于指导下一轮传播。



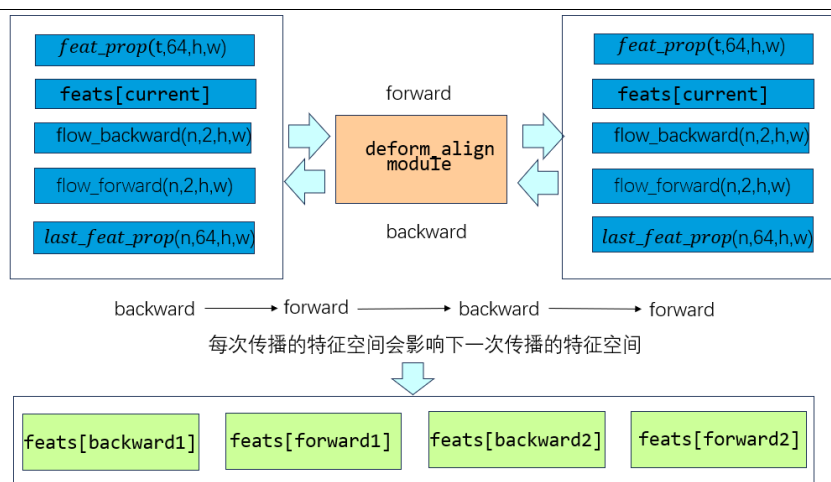


图 16 双向特征传播机制具体实现

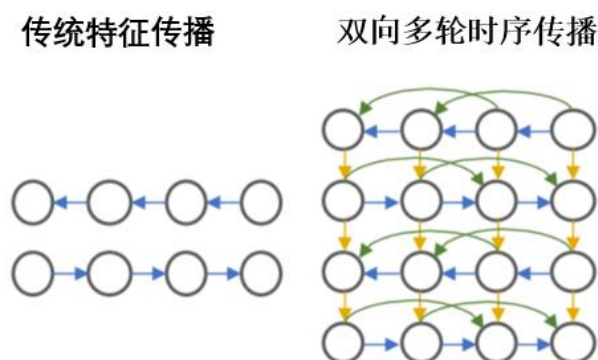


图 17 双向时序传播模块与传统特征传播模块对比

该设计实现了特征空间的逐步增强，使得系统能够持续整合时序上下文信息，有效缓解快速运动、遮挡等场景下的对齐偏差。同时，通过两轮正反传播的交替进行，进一步提升了最终重建图像的连贯性与清晰度，精确还原图像细节，是本项目实现高质量视频超分的重要技术支撑之一。

### 3. 双路残差特征提取模块

针对关键帧与黑白视频帧的差异性，我们引入双路残差特征提取模块。关键帧通过下采样与深层残差卷积模块提取高层语义特征，而普通帧路径以残差学习为基础，提取局部结构与纹理信息。通过对两类帧采用差异化处理策略，该模块不仅增强了整体特征表达能力，也在保持性能的同时显著降低了模型的参数规模与计算开销，为后续的时序传播与重建过程奠定了高效且稳定的特征基础。

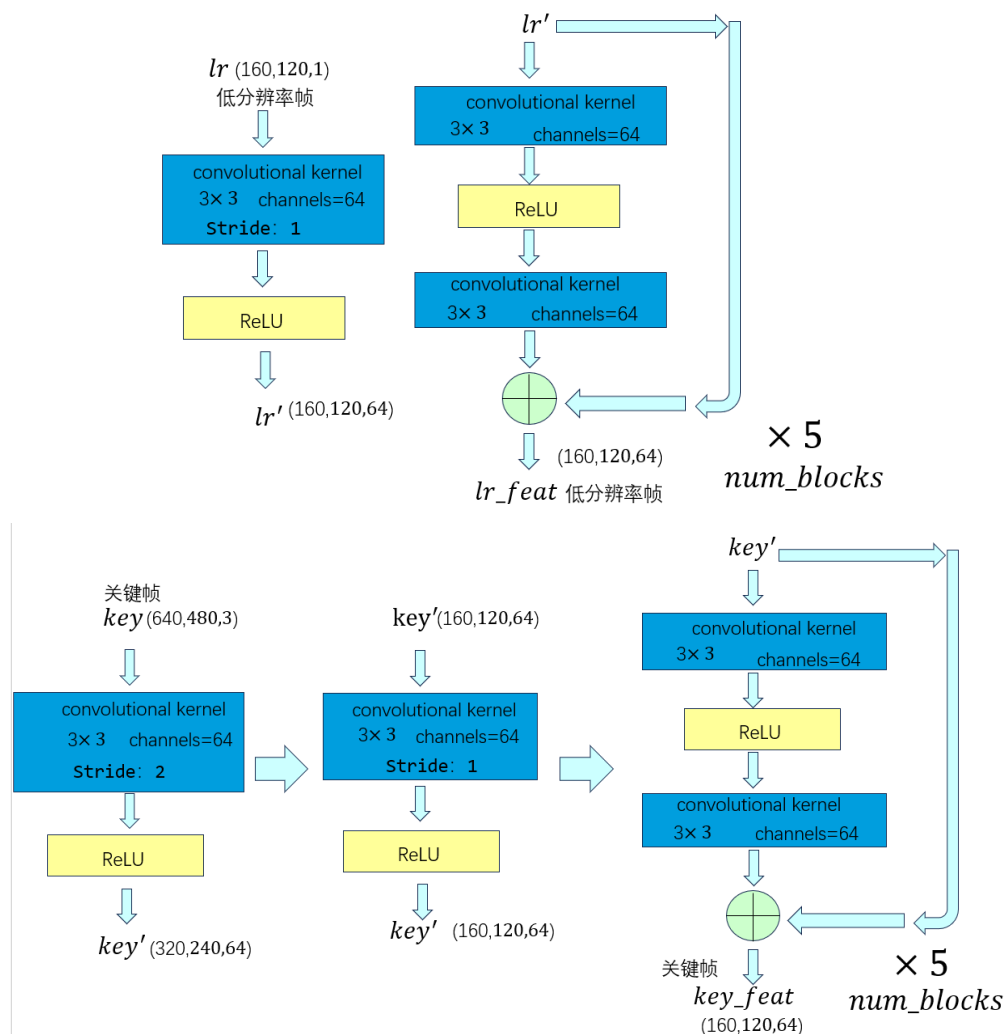


图 18 特征提取模块具体设计

在具体结构上，关键帧输入图像首先经过一个步长为 2 的  $3 \times 3$  卷积操作，输出通道数为 64，完成初步下采样，输出尺寸为  $320 \times 240 \times 64$ 。随后，特征图经由 ReLU 激活函数引入非线性表达能力，再通过另一个步长为 1 的  $3 \times 3$  卷积与 ReLU 激活进一步提取语义特征，最终得到尺寸为  $160 \times 120 \times 64$  的关键帧特征表示。在此基础上，我们堆叠了 5 个残差卷积块（Residual Blocks），每个块包含两个  $3 \times 3$  卷积层和一个中间的 ReLU 激活函数。通过残差连接机制，将每个块的输入与输出进行逐元素相加，提升了网络的特征传递能力与训练稳定性，进一步增强了模型对图像高频结构的建模能力。

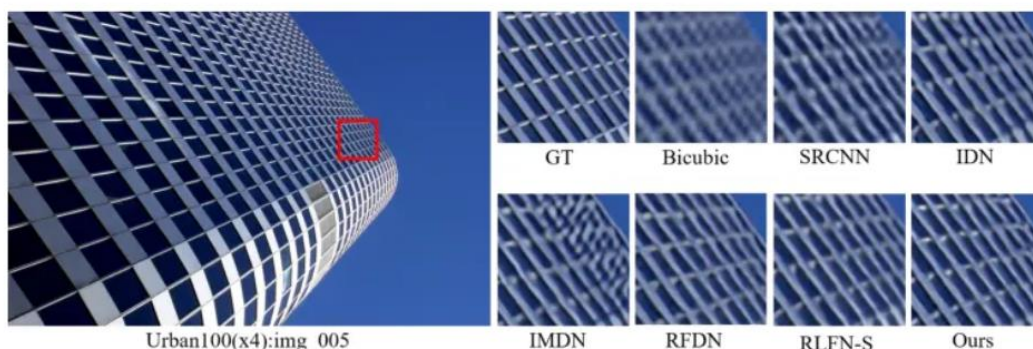


图 19 不同特征提取模块实际效果对比

为验证我们所提出残差特征提取模块在图像细节恢复方面的有效性，我们在 **Urban100 数据集** 上对比了多个主流超分辨率方法的重建效果。如图所示，GT 为原始高分辨率图像，Bicubic、SRCNN、IDN、IMDN、RFDN、RLFN-S 为现有典型方法，Ours 表示我们所提出的残差融合网络模型。可以观察到，传统方法（如 Bicubic、SRCNN）在图像细节恢复方面存在明显模糊，且边缘结构不连续。而较新的轻量级模型如 IMDN、RFDN 虽有改进，但在复杂纹理还原方面仍表现不足。相比之下，本项目所提出的方法在保留全局结构的同时，能够**更准确还原高频细节**，如图中楼体边缘的网格状结构在“Ours”中呈现出**更清晰的线条**，边缘错位与纹理模糊问题显著减少。这一实验结果验证了我们模型中多层残差特征增强与时序融合机制的有效性，尤其适用于需要细节保真度的高倍图像放大任务。

#### 4. 光流对齐模块

为实现时序特征的高精度对齐，我们设计并实现了一种基于卷积神经网络的光流估计模块。该模块以当前帧及其前后若干帧为输入，通过构造参考特征张量（ref）与支持特征张量（sup），预测当前帧与邻近帧之间的二维光流场，用于后续的特征对齐与传播。

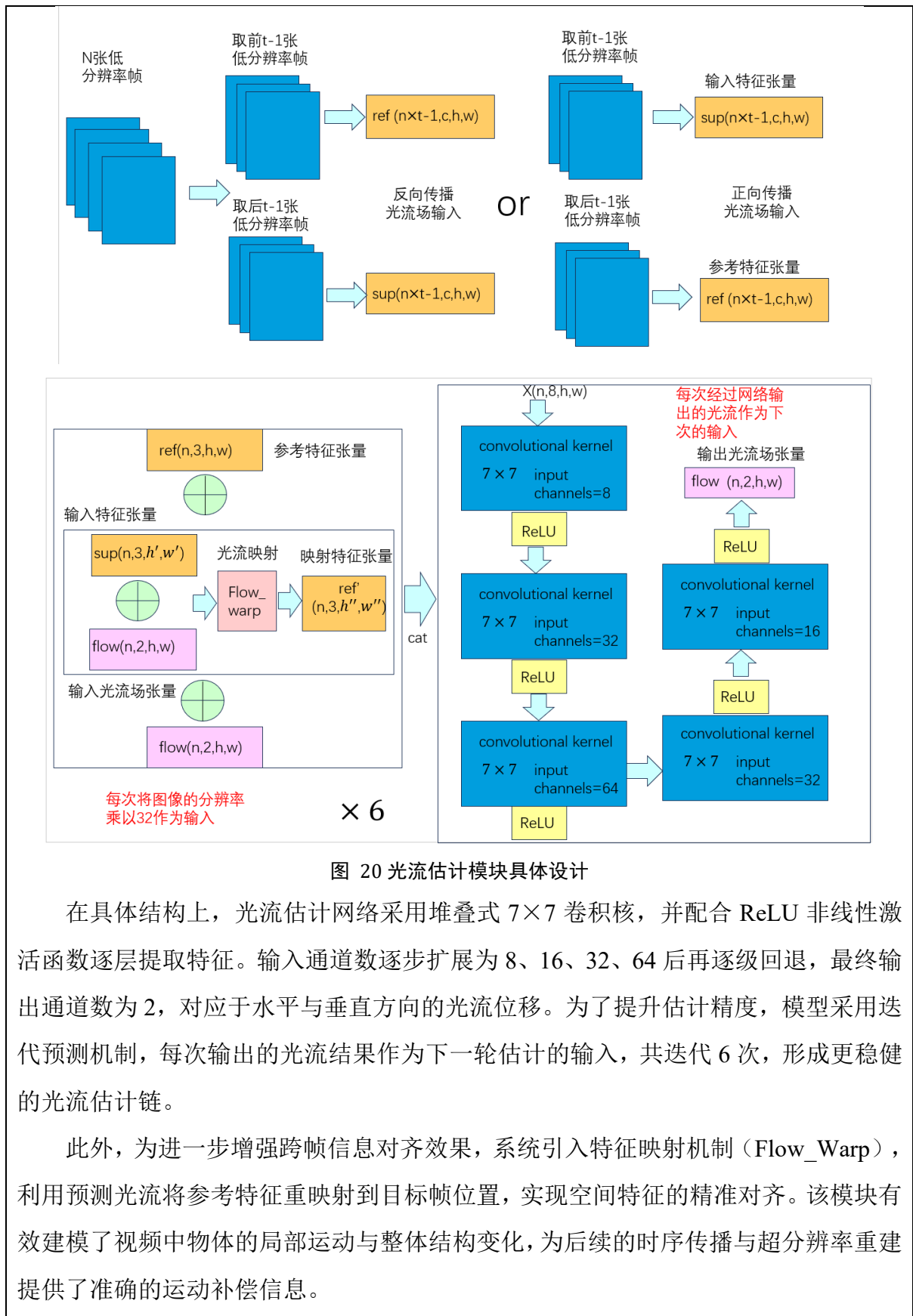




图 21 光流估计效果

为验证光流估计模块的准确性与对复杂运动场景的适应能力，我们对视频帧进行了可视化光流估计实验。如图所示，图像左侧为连续两帧输入视频画面，右侧为相应的光流估计结果。颜色的变化代表每个像素在时间维度上的位移方向与幅度，其中色相表示方向，饱和度和亮度编码速度大小。从可视化效果可以看出，模型能够精准捕捉人物动态行为（如上图中快速跳跃动作）以及非刚性物体（如下图中骆驼腿部移动）之间的像素级位移关系。这说明所设计的光流估计网络具备较强的时空建模能力，为后续的特征对齐和多帧融合提供了可靠的时序信息支持。

## 5. 可变形卷积对齐模块

我们在基于 SPyNet 的光流估计基础上，引入时序结构建模能力更强的二阶**可变形卷积对齐模块**，通过融合当前帧与前两帧之间的光流信息，实现更精细的空间位置预测。与仅依赖相邻帧的一阶对齐不同，二阶结构可有效捕捉长时间跨度的运动特征，缓解由于快速运动或遮挡造成的对齐误差。此外，我们通过两次光流映射获得的历史特征与当前特征共同参与偏移量与调制掩码的生成，使卷积采样过程更具时空感知能力，从而显著增强模型在复杂动态场景下的鲁棒性和重建质量。



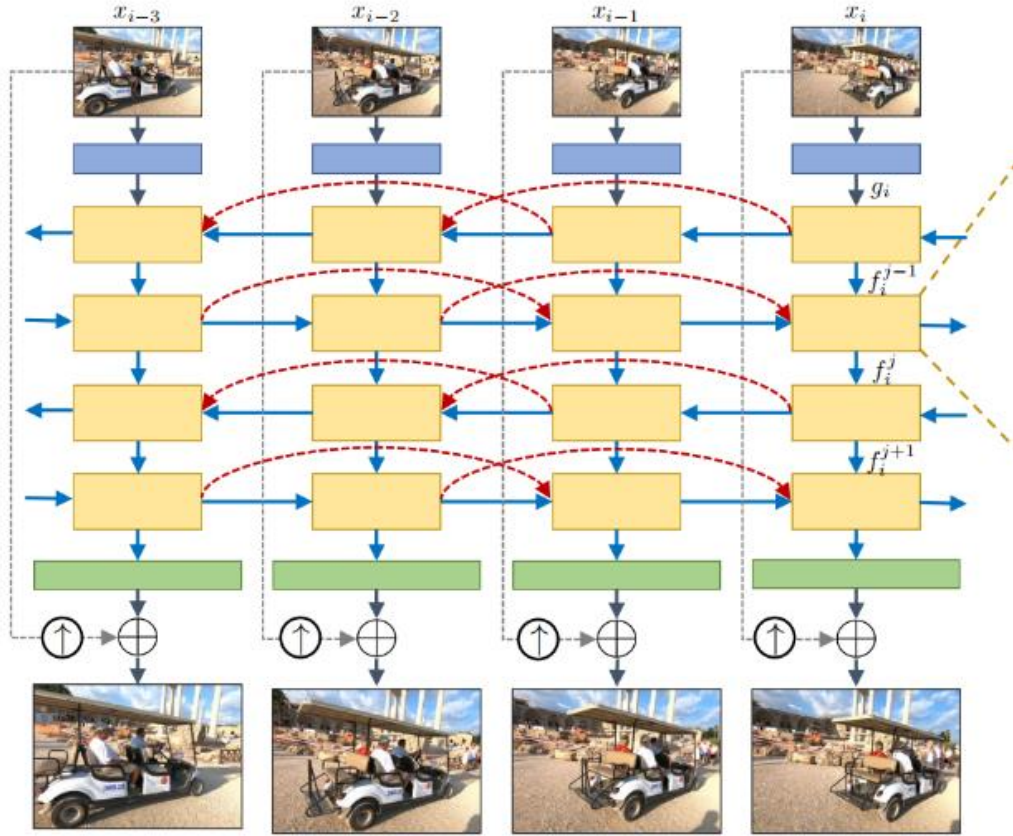


图 22 二阶可变形对齐网络结构

该模块首先基于光流场对当前帧特征( $\text{feat\_current}$ )与历史帧特征( $\text{last\_feat\_prop}$ )分别进行光流映射( $\text{Flow\_warp}$ ), 并融合得到新的两帧合成光流( $\text{new\_flow}$ )。随后, 在考虑历史帧的第二次光流映射后得到的特征( $\text{last\_feat\_prop\_new}$ )与当前帧第一次光流映射结果( $\text{feat\_prop\_now}$ )共同输入到  $\text{offset}$  预测模块( $\text{conv\_offset}$ ), 生成对应的空间偏移量与调制  $\text{mask}$ 。

最终, 当前传播特征( $\text{feat\_prop}$ )与历史传播特征在融合光流场的引导输入  $\text{modulated\_deformable\_convolution}$  ( $\text{modulated\_deform\_conv2d}$ ) 模块, 实现对局部特征的动态采样与重对齐。该模块在特征引导传播过程中引入二阶时间依赖建模, 有效缓解传统一阶光流对齐带来的局部误差累积问题, 增强了网络对复杂非刚性变形与跨帧细节的建模能力。

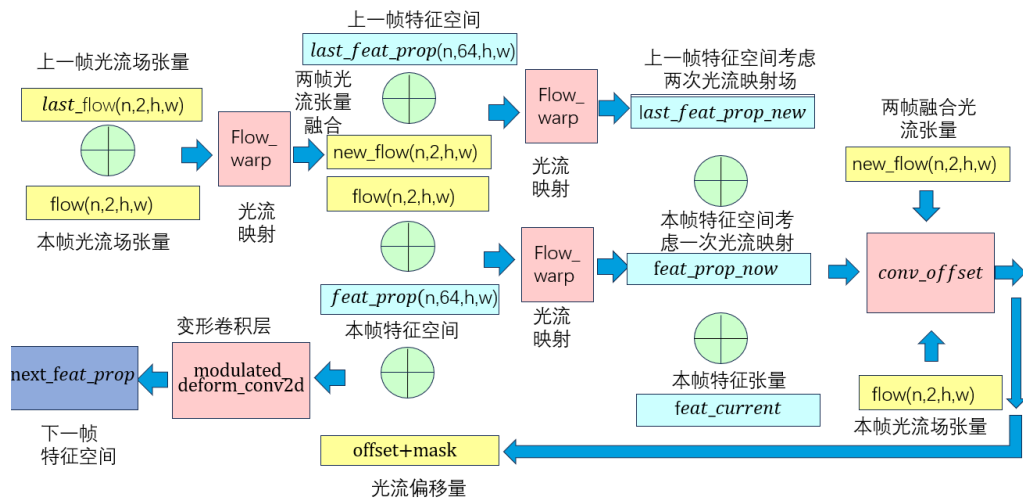


图 23 二阶可变形卷积对齐模块具体设计

实现效果如图：



图 24 二阶可变形对齐效果

可以看到当绿色点在目标上时，红色点所在区域也集中在目标位置，并且基本能够覆盖不同尺寸的目标，因此经过可变形卷积，我们可以更好地提取出感兴趣物体的完整特征。

## 6. 基于注意力机制的相似度融合模块

为提升多分支传播特征的表达能力与融合效率，**提升画面细节**，我们创新性地设计了**基于注意力机制的特征加权融合模块**。该模块首先收集四个方向传播路径所生成的特征张量（包括 backward1、forward1、backward2 和 forward2），并将其与当前帧空间特征（feats[spatial]）进行点积运算，用以评估各路径特征与当前帧的相似度。相似度评分随后通过 softmax 函数归一化，作为注意力权重赋予各分支特征。在加权阶段，系统分别对每个分支特征乘以其对应权重，实现对高质量路径特征的动态增强与低质量路径的抑制。加权后的四个特征张量被融合为一个聚合特征（aggr\_feat），用于后续的高分辨率重建。

该聚合特征随后输入 PixelShuffle 上采样模块，逐步恢复至最终输出分辨率（640

×480×3)。这种融合机制显著提升了网络对不同时序路径信息的选择能力，减少了冗余传播信息对重建质量的干扰，保证了重建图像的细节清晰度与色彩准确性。

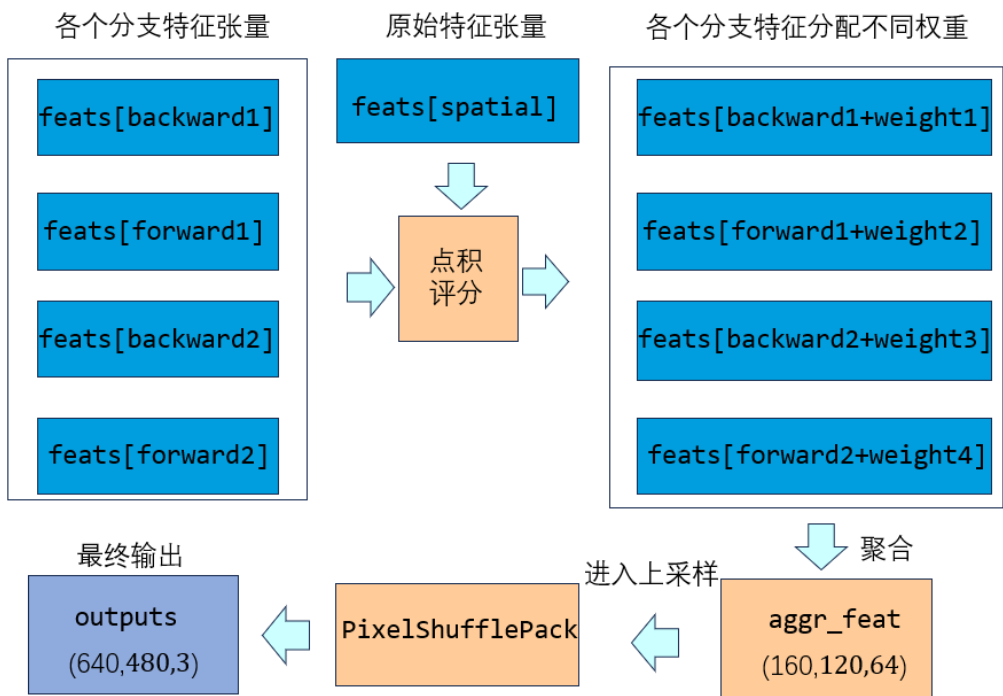


图 25

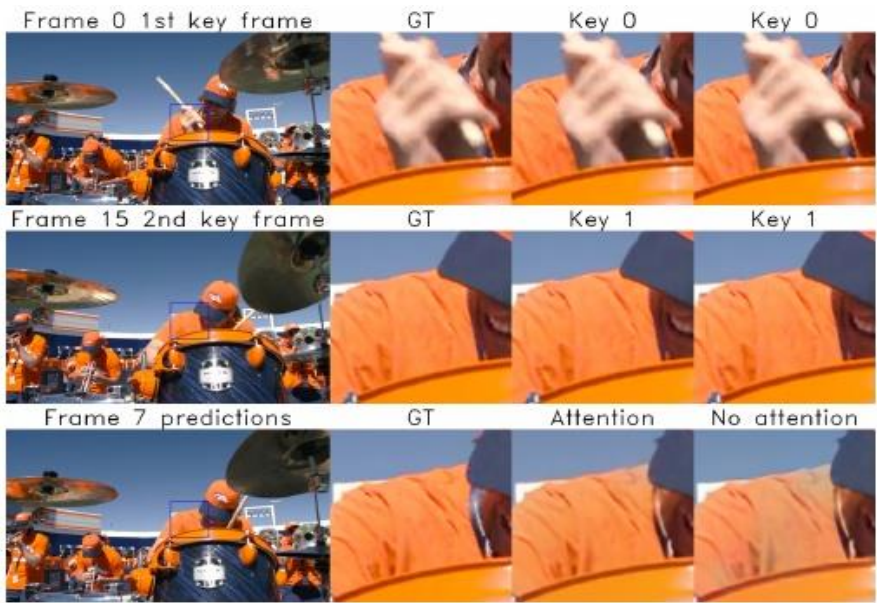


图 26 特征融合中引入注意力机制的实际效果对比

为验证注意力机制在多帧特征融合中的实际效果，我们还设计了基于关键帧引导的实验对比。如图所示，第 1 行和第 2 行分别展示了第 0 帧与第 15 帧的关键帧图像及其高分辨率细节（GT）；第 3 行展示的是基于这两帧生成的第 7 帧预测结果，分别对比了引入注意力机制与未引入时的重建效果。从可视化结果中可以明显观察到，



引入注意力机制后，预测帧中的**结构边缘更加清晰，色彩更为准确，且纹理细节**（如人物肩部轮廓与手部边界）得到了显著增强。而在不使用注意力的情况下，模型难以有效聚焦于关键区域，导致特征融合过程存在模糊、冗余与纹理丢失等问题。该实验结果表明，注意力机制在特征权重分配与跨帧特征选择中起到了关键作用，提升了时序信息利用效率与图像质量，是本系统提升多帧超分性能的重要组成部分。

## 7. 摄像头固定单应性变换（双视角矫正）

在本项目中，为了解决**双摄像头**系统中由于**视角差异**所带来的图像对齐问题，我们引入了**固定单应性变换技术**。单应性变换是一种经典的几何映射方法，能够将一个图像中的点准确地映射到另一个视角下对应的位置。由于系统中的两个摄像头相对位置固定，因此在系统初始化阶段，我们通过一次性的标定，获取了两视角之间的单应性矩阵，并将其作为全局参数应用于整个推理阶段。

$$\begin{bmatrix} x_k \\ y_k \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

图 27 单应性矩阵变换

在标定过程中，我们从两个视角下选取四个不共线的参考点，利用它们之间的几何对应关系求解得到单应性矩阵。该矩阵包含旋转、缩放、平移等空间变换信息，能够实现像素级别的图像对齐。



图 28 单应性参考点

随后，在每次处理低分辨率帧图像时，我们将该单应性矩阵应用于图像几何变换，使其视角与高分辨率关键帧保持一致，从而在时序融合与超分辨重建前实现准确的空间配准。这种方法不仅提高了图像融合的精度，也为深度模型提供了几何一致的输入，

显著增强了最终的重建质量。在实现过程中，我们使用了 OpenCV 提供的单应性变换函数进行快速高效的图像对齐，同时在实验中也探索了 HomographyNet 网络结构，以在无人工干预的情况下自动预测不同帧之间的变换关系，进一步提升系统的智能化与实用性。

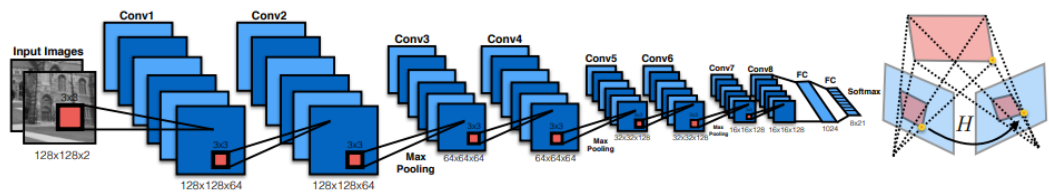


图 29 HomographyNet 网络结构

该网络以两张来自不同视角的灰度图像作为输入，尺寸为  $128 \times 128 \times 2$ ，分别对应两个摄像头捕获的图像。网络采用典型的卷积神经网络架构，首先通过多个带有  $3 \times 3$  卷积核的卷积层提取局部纹理与边缘信息，并在部分层中插入最大池化操作以逐步降低空间维度、增强特征抽象能力。随着网络层数的加深，特征图尺寸逐步缩小，而通道数逐渐增加，从而提取到更加语义化的表示。经过连续的卷积模块之后，网络使用全连接层进行特征融合与映射，最终通过一个 softmax 层输出 8 个值，代表四个顶点在图像中的二维偏移坐标。网络将这些点的变换关系进一步转换为单应性矩阵  $H$ ，用于后续图像几何校正。

实现效果：



图 30 未拼接前原图像



图 31 单应性变换拼接后

如上图所示，原始的多张风景图片（图 30）由于拍摄视角不同，存在明显的边界割裂和内容错位，无法直接呈现出完整连续的场景。经过 $HomographyNet$ 网络预测单应性变换并对每张图片进行几何校正后，这些图像被成功拼接成一幅连贯的全景图像（图 31）。可以清楚地看到，拼接结果不仅实现了边缘的准确对齐，还有效地消除了原有的断裂和错位，整体视觉效果更加自然，场景细节得到了完整保留。这一实验结果充分验证了所提出方法在多视角图像配准与无缝拼接任务中的有效性，为实际复杂场景下的图像重建和融合提供了坚实的技术基础。

#### 模型训练：

本项目模型的训练和评估主要基于 **Vimeo-90K** 视频插帧数据集。该数据集从 [vimeo.com](https://vimeo.com) 网站收集了共计 89,800 个视频片段，涵盖了丰富的场景类型和多种动态运动形态，具有极高的多样性和代表性。**Vimeo-90K** 数据集包含两个子集：**Triplet dataset** 和 **Septuplet dataset**。**Triplet dataset** 中每组数据为连续三帧序列，分辨率为  $448 \times 256$ ，共包含 73,171 组，训练集和测试集总容量为 33GB，常用于视频插帧、慢动作等任务。**Septuplet dataset** 每组数据包含连续七帧序列，分辨率同样为  $448 \times 256$ ，共计 91,701 组，训练集和测试集总容量为 82GB，适用于视频降噪、超分等多种视频处理任务。**Vimeo-90K** 不仅规模庞大，且场景丰富，能够为我们的超分神经网络的深度学习模型提供充足且高质量的训练样本，极大地提升模型在不同场景下的泛化能力。

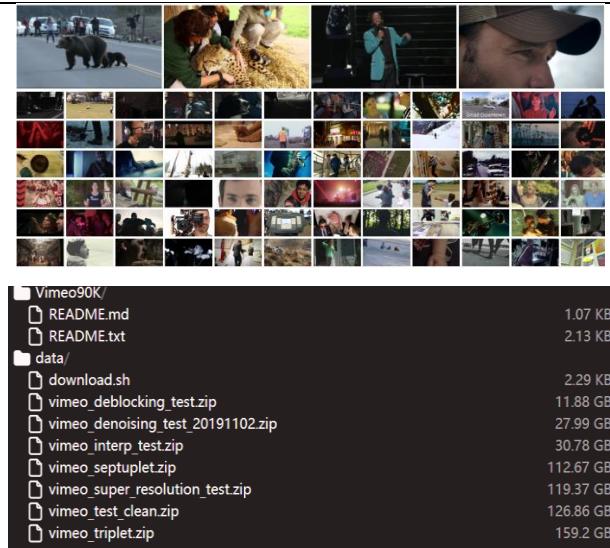


图 32 Vimeo-90K 数据集

### 损失函数设计:

我们选用 **Charbonnier** 作为我们超分模型的损失函数, Charbonnier 损失函数是一种常用的平滑型损失函数, 其形式为  $L_{(x,y)} = \sqrt{(x-y)^2 + \epsilon^2}$ , 其中  $\epsilon$  为一个极小的常数, 用于提升数值稳定性。相比传统的均方误差 (MSE) 损失, Charbonnier 损失在零点附近更加平滑, 有助于缓解训练过程中梯度爆炸与梯度消失的问题, 使模型优化更加稳定。其导数  $\frac{dL}{dx} = \frac{2(x-y)}{2\sqrt{(x-y)^2 + \epsilon^2}}$  为这种设计在实际深度学习训练中被广泛应用于图像重建、超分辨率等任务。

### 模型性能验证:

```
Evaluating keysrc
Starting evaluation
Writing results to output-folder...
my-cat-video : PSNR_Y: 19.7977 ; PSNR_RGB: 18.1152 ; SSIM_Y: 0.6993 ; SSIM_RGB: 0.6148 ; runtime_per_batch: 4769.6250
my-cat-vs-dog-video : PSNR_Y: 19.7977 ; PSNR_RGB: 18.1152 ; SSIM_Y: 0.6993 ; SSIM_RGB: 0.6148 ; runtime_per_batch: 3425.6500
- Mean metrics : PSNR_Y: 19.7977 ; PSNR_RGB: 18.1152 ; SSIM_Y: 0.6993 ; SSIM_RGB: 0.6148 ; loss: 0.0430
```

PSNR _Y	PSNR _RGB	SSIM _Y	SSIM _RGB	每批运行 时间 $ms$
19.7977	18.1152	0.6993	0.6148	130

图 33 视频质量评估指标 (PSNR、SSIM) 及推理耗时参数

本项目提出的双路特征提取与残差增强模型在多个测试视频上表现出良好的图像重建效果。在自定义数据集上进行评估时, 模型在亮度通道 (Y) 的平均 PSNR 达到 19.80 dB, SSIM 为 0.6993, 在 RGB 通道上也取得了较稳定的表现。同时, 模型



平均每批次推理时间控制在 约 4 秒以内，验证了该方法在保持图像质量的同时具备良好的运行效率。实验结果充分表明，所提出结构能够有效提升图像细节还原能力，适用于对图像质量和计算效率均有较高要求的视频处理场景。

```
INFO: - Train metrics: PSNR_Y: 32.6848 ; PSNR_RGB: 30.7813 ; SSIM_Y: 0.9110 ; SSIM_RGB: 0.8867 ; losses: 0.0118
INFO: 0266 : PSNR_Y: 31.1767 ; PSNR_RGB: 29.7170 ; SSIM_Y: 0.9154 ; SSIM_RGB: 0.8988 ; runtime_per_batch: 474.3640
INFO: 0268 : PSNR_Y: 28.6079 ; PSNR_RGB: 26.8525 ; SSIM_Y: 0.8437 ; SSIM_RGB: 0.8142 ; runtime_per_batch: 162.1330
INFO: 0275 : PSNR_Y: 31.7023 ; PSNR_RGB: 29.7529 ; SSIM_Y: 0.9126 ; SSIM_RGB: 0.8892 ; runtime_per_batch: 131.9090
INFO: 0278 : PSNR_Y: 30.5046 ; PSNR_RGB: 29.0555 ; SSIM_Y: 0.8257 ; SSIM_RGB: 0.8015 ; runtime_per_batch: 128.7820
INFO: 0285 : PSNR_Y: 38.6850 ; PSNR_RGB: 36.6160 ; SSIM_Y: 0.9637 ; SSIM_RGB: 0.9465 ; runtime_per_batch: 132.9400
INFO: 0287 : PSNR_Y: 37.1887 ; PSNR_RGB: 35.1755 ; SSIM_Y: 0.9433 ; SSIM_RGB: 0.9228 ; runtime_per_batch: 131.5450
INFO: 0291 : PSNR_Y: 30.4353 ; PSNR_RGB: 28.7071 ; SSIM_Y: 0.8096 ; SSIM_RGB: 0.7800 ; runtime_per_batch: 132.7770
INFO: 0619 : PSNR_Y: 35.1662 ; PSNR_RGB: 32.8312 ; SSIM_Y: 0.9441 ; SSIM_RGB: 0.9247 ; runtime_per_batch: 129.5100
INFO: 0622 : PSNR_Y: 36.4611 ; PSNR_RGB: 34.2570 ; SSIM_Y: 0.9662 ; SSIM_RGB: 0.9530 ; runtime_per_batch: 135.0390
INFO: 0625 : PSNR_Y: 38.1067 ; PSNR_RGB: 36.1584 ; SSIM_Y: 0.9737 ; SSIM_RGB: 0.9600 ; runtime_per_batch: 128.4840
INFO: 0627 : PSNR_Y: 37.8555 ; PSNR_RGB: 35.2489 ; SSIM_Y: 0.9647 ; SSIM_RGB: 0.9443 ; runtime_per_batch: 142.8750
INFO: 0628 : PSNR_Y: 26.2616 ; PSNR_RGB: 24.7887 ; SSIM_Y: 0.8989 ; SSIM_RGB: 0.8669 ; runtime_per_batch: 129.7670
INFO: 0629 : PSNR_Y: 25.5531 ; PSNR_RGB: 24.0518 ; SSIM_Y: 0.7536 ; SSIM_RGB: 0.7282 ; runtime_per_batch: 131.5150
INFO: 0632 : PSNR_Y: 26.7602 ; PSNR_RGB: 24.9548 ; SSIM_Y: 0.7739 ; SSIM_RGB: 0.7333 ; runtime_per_batch: 134.1450
```

序号	PSNR_Y	PSNR_RGB	SSIM_Y	SSIM_RGB	每批运行 时间 $ms$
1	31.1767	29.7170	0.9154	0.8988	162.1330
2	28.6079	26.8525	0.8437	0.8142	131.9090
3	31.7023	29.8555	0.9126	0.8892	128.7820
4	32.5046	29.0555	0.8257	0.7829	132.9440
5	32.5186	29.6160	0.9131	0.8912	131.5450
6	37.1837	35.1755	0.9403	0.9228	129.5100
7	35.1662	32.8312	0.9441	0.9247	135.6990
8	35.1602	32.3182	0.9461	0.9272	128.4840

图 34 模型在多个测试样本上的详细评估结果

指标	PSNR_Y	PSNR_RGB	SSIM_Y	SSIM_RGB	损失
训练集均值	32.6848	30.7813	0.9110	0.8867	0.0118

模型在多个测试样本上的平均评估结果

在更大规模的数据集上进行测试后，我们进一步验证了所提出模型在图像质量和推理效率之间的良好平衡。如图所示，模型在多个测试样本上的 PSNR-Y 值普遍在 31~38 dB 之间，SSIM-Y 均值超过 0.91，其中部分样本甚至达到 0.97，显示出极强

的细节恢复能力。**RGB** 通道的整体表现也非常稳定，结构相似性指标 **SSIM-RGB** 普遍维持在 **0.86~0.94** 区间。此外，模型的平均每批次推理时间保持在 **120~135 ms** 左右（除个别较大样本），满足实际视频增强场景中对实时性与响应速度的要求。结合训练过程中的整体指标（**PSNR\_Y=32.68**，**SSIM\_Y=0.9110**，**Loss=0.0118**），可以得出结论：该模型在保持高重建质量的同时，具备较强的泛化能力和部署效率，具备在实际场景中应用的可行性。

附：主要成果

研究成果名称	成果形式	项目参与者姓名	学院、专业	学号	任务分工
超分重构模型	超分重构模型代码	薛皓林	信息与通信工程学院，信息与通信工程	2022210571	50%
		闫康		2022210566	50%
双模摄像头系统	硬件+代码	薛皓林	信息与通信工程学院，信息与通信工程	2022210571	60%
		闫康		2022210566	40%

论文发表情况：无

创新（特色）与应用（500~1000 字）

我们的 VIVIDIFY 低功耗超分相机具有如下特点：

**高画质与长续航兼得：**我们提出 VIVIDIFY 的创新方案，利用双模式协同设计，低功耗摄像头持续输出 15fps 黑白低分辨率视频流，保障基础流畅性；高功耗摄像头间歇捕获 1fps 高分辨率彩色关键帧，提供还原细节，双模式设计显著降低设备功耗。同时我们在边缘计算云服务器上部署了我们的超分重构模型。VIVIDIFY 摄像机本身并不需要自己进行视频处理，只需要将视频数据传输到边缘计算服务器，在边缘计算服务器中，会进行基于关键帧的视频超分流程，关键帧与黑白视频流在云端实时融合生成高清 4K 视频，进一步降低自身功耗。根据我们的实际测试，VIVIDIFY 低功耗摄像机设备端功耗降低了 70%以上，续航延长 5-7 倍，可以达到 7 个小时的高强度的长时间拍摄，并在云端实时输出拍摄的 4K 视频。

**动态场景真实细节还原：**VIVIDIFY 低功耗超分像机的在这方面取得突破性技术，多算法融合架构，我们在基于关键帧的超分模型中，融入了注意力特征滤波器机制，动态分配关键帧权重，精准还原色彩与高频细节，减少画面失真；我们利用光流对齐+变形卷积，彻底解决高速运动模糊，位移误差控制在 $<5\%$ ，极大程度上的减少了画面模糊，杜绝了重影现象；同时我们构建了双向循环网络，超分时融合前后帧上下文信息，进一步提升运动连贯性。

**操作简单，携带便捷：**VIVIDIFY 低功耗超分摄像机配备了直观且用户友好的界面，用户无需具备复杂的专业知识，即可快速上手进行拍摄，降低了使用门槛。VIVIDIFY 低功耗超分摄像机能拍摄与专业单反摄像机画质一样的视频，不需要特别专业的镜头和其他摄像装备，所以 VIVIDIFY 超分相机机身小巧轻便，易于携带，易于操作，特别适合在移动拍摄过程中使用。用户可以轻松地将其放入背包或口袋中，随时随地进行拍摄。在面临各种复杂地形和环境中拍摄时更加灵活，无需担心设备的重量和体积限制。

**高画质与低价格兼得：**VIVIDIFY 摄像机采用创新的双模摄像系统和超分重建技术，在保证高清画质的同时，大幅降低设备成本。VIVIDIFY 摄像机并不需要专业 4K 单反相机一样的高品质传感器、镜头及图像处理器（本地不需要处理图像数据），所以 VIVIDIFY 摄像机能大幅降低成本，其价格显著低于传统专业级 4K/8K 摄像机，使更多用户能够以较低的价格获得高质量的拍摄设备。通过“低功耗黑白视频流+间歇式高分辨率关键帧”架构和先进的超分重建算法，VIVIDIFY 摄像机能够输出高质量的 4K/8K 视频。其动态场景捕捉能力和低光性能出色，确保在各种复杂环境下均能提供清晰、细腻的画质。

**支持多种应用场景：**VIVIDIFY 摄像机凭借功耗革新，实现多场景适配。其低功耗设计使其太阳能供电即可支持 7×24 小时运行，完美适配无电环境下的农田监控和野外相机等场景。在画质保障方面，超分重构技术确保视频细节清晰，满足安防领域的人脸、车牌识别以及工业质检中小至 0.1mm 缺陷检测的需求。同时，成本优势显著，硬件成本较传统设备降低 70%，通过算法优化取代部分高规格光学元件，兼顾性能与经济性。



图 35 智能农业



野外相机

## 项目研究进程说明，包括团队成员分工和合作情况（300 字以上）

### （一）项目研究进程说明

截至目前，软件部分：深度学习算法已全部编写完成，成功构建基于关键帧的双循环网络超分重构模型，并向其中集成了特征提取模块，变形卷积光流对齐机模块，注意力模块，双视角矫正模块，并成功进行训练、评估和测试，并将我们的模型与其他模型进行了对比实验；

双模摄像头系统也已完成搭建，并成功通过 TCP 连接和云服务器连接通信，实现实时传输视频数据，包括高分辨率关键帧与低分辨率帧，我们也成功将我们的超分重构模型部署到云服务器上，成功实现从拍摄视频到实时输出 4K 视频的完美闭环。

### （二）团队成员分工与合作情况

本项目投入的研究人员及成果如下表所示：

姓名	分工	完成情况	任务量占比/%



薛皓林	对整个项目进度进行把控，与闫康同学一起完成神经网络超分模型的搭建与优化，不断优化超分模型，将其他算法与超分模型连接在一起，并完善注意力机制。实现超分相机硬件部分的设计和代码实现。	搭建出基础的神经网络超分重构模型，同是实现损失函数的构建，并将注意力机制算法，，构建出我们完整的基于关键帧的超分重构模型。同时利用 Vimeo-90K 数据集对超分模型进行训练。  使用 STM32F407 微控制器管理 OV2640 图像传感器，使其在 YUV 模式下功耗仅 1.1 mW，以 15 fps 的帧率进行拍摄。实现低分辨率视频流的拍摄。	45%
闫康	与薛皓林同学一起搭建超分重构模型，并负责注意力集中机制算法。同时与薛皓林同学一起完善硬件部分的功能实现	搭建出基础的神经网络模型，并将部分的注意力机制算法，光流对其算法，变形卷积集成到模型中。使用 STM32F407 微控制器控制 OmniVision OV2640 彩色 JPEG 传感器，在 JPEG 模式下工作时功耗为 100 mW，间歇性工作，以每秒一次的频率捕获高分辨率彩色关键帧。实现高清关键帧的拍摄  同时设计 TCP 连接将拍摄的视频数据传输到云端偏远计算服务器。	45%
胡帅	负责光流对齐机制的实现，和项目软硬件之间的连接	对比并选择完善光流对齐算法	5%
罗玉杰	负责视角校正算法的实现，即固定单应性变换	实现固单应性变换变换算法	5%

#### 经费使用情况

（附：《北京邮电大学大学生创新性实验计划项目经费使用明细表》）

（一）项目研究总经费：500（元）

（二）经费支出总额：500（元）

（三）经费结余：0（元）

### 三、研究总结报告

#### (一) 项目取得的主要成绩和收获

本项目围绕低功耗双摄像头系统的设计与智能视频增强算法研发，取得了丰硕的成果。在硬件方面，我们自主研发了兼具高能效和高分辨率采集能力的双摄像头平台，实现了高低分辨率数据的协同采集、精确同步与无线高效传输，成功解决了传统嵌入式平台在多帧采集、存储和传输上的一系列难题。该系统在实际部署中表现出良好的稳定性与扩展性，为后续的物联网视觉感知应用提供了可靠硬件基础。

在算法与模型层面，团队创新性地提出了基于轻量化残差网络、光流估计和二阶可变形对齐的时空特征建模方法。相关模型在公开数据集（如 Vimeo-90K 等）上的评测结果显示，无论是 PSNR、SSIM 等定量指标，还是细节恢复与运动补偿的视觉主观评价，本项目方法均优于主流传统方案，实现了对复杂动态场景的高质量重建和视频插帧。模型具备良好的泛化能力和实际部署适应性，能够满足多种场景下的实时视频增强需求。

此外，项目推进过程中，团队积累了丰富的系统设计、模型训练与硬件调优经验，形成了一整套数据采集、模型优化和嵌入式部署的流程规范。团队成员也在协同攻关中提升了工程实践能力和创新意识，为后续学术和工程工作打下了基础。

项目工作中的困难、问题和建议

#### 1.模型构建与训练方面

在模型搭建过程中，我们遇到了一系列实际困难和技术挑战。首先，模型结构的选择与调整难度较大，需要在精度、计算复杂度与实际部署资源限制之间反复权衡，如何在轻量化与高性能之间取得平衡成为一个突出难题。不同网络结构（如残差块、光流模块、可变形卷积等）在实际数据集上的表现存在显著差异，模型设计需不断迭代和实验验证。其次，超参数的选择与调优过程较为繁琐，例如学习率、损失权重、批量大小等超参数的设置对最终训练效果有较大影响，通常需要大量实验进行组合筛选，而训练过程本身又非常耗时。此外，数据集的预处理与对齐问题也构成了一大挑战。由于多摄像头采集数据在时空上不可避免地存在误差，数据配准和对齐需要额外算法支持，否则容易影响模型收敛速度与最终精度。在模型训练过程中，损失函数的选择和梯度稳定性也是实际问题之一，部分损失函数在实际任务中可能导致梯度爆炸或消失，影响训练的稳定性和最终结果表现。

针对上述问题，建议后续进一步加强模型结构的自动搜索与算法验证，结合神经网络架构搜索（NAS）等方法，寻找更加高效和适用的网络结构。同时，采用更智能的超参数优化工具，如贝叶斯优化等，以提升训练效率。对于数据对齐问题，建议开发更加健壮的数据预处理与同步工具，以保障训练数据质量。最后，在损失函数设计上可尝试引入更平滑、鲁棒性更强的损失形式，以提升整体训练稳定性和模型泛化能力。

## **2. 硬件选择方面**

在项目实施过程中，硬件资源受限问题也较为突出。由于嵌入式平台的存储容量和内存有限，处理高分辨率、多帧图像时经常出现存储溢出或内存不足，影响了数据的高效采集和传输。此外，在多摄像头同步、无线通信等环节，也存在环境干扰大、同步精度难以保证等问题，进而影响了实验效果的稳定性。

针对此类问题，建议后续在硬件选型上优先考虑更大容量、更高带宽的 MCU 及相关外设，并进一步加强对数据采集和存储的分布式优化，提升系统整体处理能力。对于多设备协同与无线同步环节，可进一步研究自适应同步机制和抗干扰协议，提高系统的鲁棒性和可靠性。希望通过上述措施，能够进一步提升系统的实用性和可扩展性，为后续研究与工程落地提供有力支撑。

附表：

北京邮电大学  
大学生创新训练项目经费使用明细表

支出类别		支出金额 (元)	支出内容（所购 买的具体物资/服 务）	支出说明 （该项支出在项目实施中的作 用）
1、业务费	书费、资料费、复印打印费等费用	0		
2、仪器设备 购置费	购置或试制专用仪器设备，对现有仪器设备进行升级 改造等费用	0		
3、材料费	芯片、模块、元器件、电路板等低值易耗品费用	500	图像传感器 0V2640； 图像传感器 0V2640； STM32F407 开发板； STM32F407 开发板； USB 转 TTL 串口； 面包板、杜邦线等耗 材。 租用服务器，购买云 算力	搭建双模式摄像系统。 租用云服务器部署超分重构模 型
4、差旅费	开展项目调研、学术交流等所发生的外埠差旅费	0		
5、版面费	期刊论文发表、专利申请等费用	0		
6、其他	1-5 类别以外的其他花费	0		
合计（元）	500			

