

### 三、项目设计论证

本表参照以下提纲撰写，要求逻辑清晰，主题突出，层次分明，内容翔实，排版清晰，**6000**字以内。

1. [选题依据] 国内外相关研究的学术史梳理及研究动态；本课题相对于已有研究的独到学术价值和应用价值等。
2. [研究内容] 本课题的研究对象、总体框架、重点难点、主要目标等。
3. [思路方法] 本课题研究的基本思路、具体研究方法、研究计划及其可行性等。
4. [创新之处] 在学术思想、学术观点、研究方法等方面的特色和创新。
5. [预期成果] 成果形式、使用去向及预期社会效益等。
6. [参考文献] 开展本课题研究的主要中外参考文献。

#### 1. 选题依据

##### 1.1. 研究意义

图像恢复旨在从受损观测中重建高质量清晰图像，是计算机视觉领域的基础性研究课题。在安防监控、自动驾驶、医学成像、遥感观测等实际应用中，图像采集与传输过程不可避免地受噪声、模糊、雾霾、雨线、低光照、压缩伪影等多种退化因素干扰，严重制约后续视觉分析任务的精度与可靠性。

现有深度学习图像恢复方法面临的核心瓶颈在于**退化空间异质性**的表征缺失。真实场景中，同一图像的不同区域往往同时受到类型迥异、程度各异的退化影响，而现有方法普遍采用全图统一的处理策略，无法根据区域级退化特征进行差异化恢复。具体而言，单任务专用模型<sup>[1-2]</sup>针对特定退化类型优化，面对多退化共存场景时泛化能力不足；All-in-One 统一模型<sup>[3-5]</sup>虽能处理多种退化类型，但受限于训练分布，对未见退化组合的适应性较差；现有智能体化方法虽引入动态决策能力，但普遍将整幅图像视为单一实体，缺乏对退化空间分布的精细化感知，导致在异质退化场景下出现局部过度处理或处理不足的问题。

预研实验定量验证了上述瓶颈的实际影响。在复杂混合退化测试集上，代表性模型的峰值信噪比（Peak Signal-to-Noise Ratio, PSNR）较单一退化场景下降 1.5–3.0 dB，学习感知图像块相似度（Learned Perceptual Image Patch Similarity, LPIPS）指标出现系统性退化，且推理时延显著增加。实验结果表明，现有方法的理论假设——即退化分布均匀或可通过全局特征统一描述——与真实场景存在显著偏差，这一偏差构成了制约图像恢复技术实用化部署的关键障碍。

针对上述理论瓶颈，本项目研究**多智能体协同的空间感知图像恢复方法**，构建“空间退化感知—区域协同决策—规划轨迹蒸馏”的层级化框架。该框架将恢复过程从传统的“全图单策略”范式转变为“区域差异化处理”范式：首先通过空间退化图显式建模区域级退化分布；继而由协调智能体进行全局任务分解，各区域子智能体针对特定退化特征执行定制化恢复策略；最终通过规划轨迹蒸馏实现大模型能力向轻量级部署模型的有效迁移，形成可感知、可决策、可迭代、可部署的新型恢复范式。

本项目的研究意义可归纳为三个层面。

**学术层面**, 本项目推动图像恢复从“全局统一表征”向“区域级动态表征”演进, 建立退化空间异质性的显式建模理论。通过引入空间退化图与多智能体协同机制, 丰富复杂场景图像恢复的认知框架, 为后续研究提供新的理论视角与分析工具。

**方法层面**, 本项目促进空间感知、多智能体协同与知识蒸馏的深度融合, 形成兼顾恢复精度与推理效率的轻量级方法体系。通过轨迹蒸馏机制突破大模型依赖导致的部署瓶颈, 实现从研究原型到实用系统的技术跨越。

**应用层面**, 本项目为安防巡检、智能交通、遥感解译等场景提供高可靠视觉基础能力, 支撑复杂环境下的智能感知系统部署, 具有明确的工程价值与社会效益。

## 1.2. 国内外研究现状

图像恢复领域近十年经历了从卷积神经网络到 Transformer、从确定性回归到扩散生成模型的技术演进, 并逐步向统一化 All-in-One (AiO) 框架、基于智能体的动态决策范式跃迁, 每一次范式转换均以解决前驱方法的特定局限为直接动因。“退化空间异质性”——真实场景图像不同区域同时受到类型迥异、程度各异的退化影响——是贯穿各范式演进历程、迄今尚未被充分应对的核心技术问题。以下从五个维度系统梳理国内外研究现状, 聚焦各范式的关键方法、技术局限及其与本项目研究动机的内在联系。图 1 给出了传统全图统一恢复与本项目拟研究的空间感知分区恢复的框架对照示意。

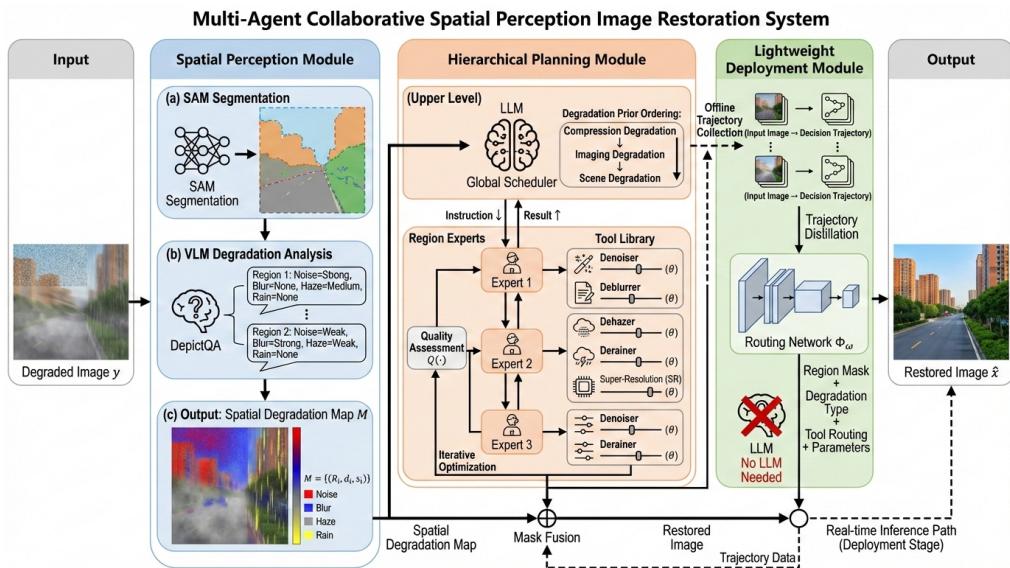


图 1: 传统全图统一恢复与空间感知分区恢复的框架对照示意

### (一) 基于 Transformer 与扩散模型的图像恢复基础方法

图像恢复领域的基础方法主要围绕 Transformer 架构与扩散生成模型两条技术路线展开, 分别代表了确定性回归与生成式建模两类范式。标准自注意力机制的计算复杂度与图像像素数量的平方成正比, 对高分辨率恢复任务代价高昂, 这一计算瓶颈驱动了一系列高效 Transformer 架构的系统性探索。SwinIR<sup>[1]</sup>通过残差 Swin Transformer 块与移位窗口多头自注意力将计算复杂度降至线性, 在图像去噪、超分辨率与 JPEG 压缩伪影消除等基准任务

上较此前最优方法分别提升 0.14–0.31 dB PSNR，成为基于 Transformer 图像恢复研究的标准参照基线。SwinIR 的核心思想在于：通过在固定空间窗口内计算局部自注意力，并利用窗口移位实现跨窗口信息交互，在保持局部感受野精细建模的同时实现全局上下文的间接建模。Restormer<sup>[2]</sup>以“转置注意力”重新定义了 Transformer 的计算路径：提出多深度卷积头转置注意力与门控深度卷积前馈网络，沿通道维度而非空间维度计算注意力，以线性空间复杂度保持了真正意义上的全局感受野，在 SIDD 去噪基准上较前一最优方法提升 0.13–0.52 dB，在去雨、去模糊等任务上建立了当时的最优基准。

在计算效率的进一步提升方面，研究者从频域计算、状态空间模型等角度探索了传统自注意力的替代方案。Wang 等人<sup>[6]</sup>提出 Uformer，构建层次化 U 型 Transformer 架构，结合编码器-解码器跳跃连接与可学习的多尺度恢复调制器，使模型在去噪、去运动模糊等多个任务中同时保持高性能。FFTformer<sup>[7]</sup>将传统空间域自注意力替换为傅里叶域操作，以更低计算代价实现等效的全局感受野，在图像去模糊任务上尤为有效。基于 Mamba 选择性状态空间模型的 VmambaIR<sup>[8]</sup>通过多方向选择性状态空间扫描以严格线性复杂度建模长程依赖，与 Transformer 相比在内存消耗和推理速度上均有显著优势，为超高分辨率图像的高效恢复开辟了新路径。

扩散概率模型通过迭代去噪过程建模复杂数据分布，为图像恢复提供了强大的生成先验能力，在感知质量评估指标上取得了显著优于确定性方法的成果。Palette<sup>[22]</sup>将条件扩散模型系统化应用于图像间转换任务，证明扩散方法能够生成多样化高质量输出——特别是在严重退化导致信息大量缺失的场景下，生成先验可以“合理补全”丢失细节，而确定性方法在此场景下往往产生过度平滑的结果。DiffBIR<sup>[24]</sup>开创了利用大型扩散模型先验的两阶段范式：第一阶段由专用退化消除网络去除主要退化成分，第二阶段借助预训练 Stable Diffusion 模型的生成先验对中间图像进行高质量细节增强，在真实世界盲图像超分辨率和人脸恢复任务上取得了卓越的感知质量。尽管感知质量出众，扩散模型面临迭代采样的高计算成本制约，单次恢复通常需数十次至数百次神经网络前向传播。Liu 等人<sup>[23]</sup>提出残差去噪扩散模型，将扩散过程重新定义为在洁净图像与退化图像残差上操作，仅需对变化量建模而非完整图像，将所需采样步数显著压缩至少数步骤，为扩散模型实用化部署奠定了基础。

上述基于 Transformer 的确定性方法与基于扩散的生成式方法在各自适用场景均达到了较高性能水平，但其共同局限在于对整幅图像采用统一的恢复策略，完全未能利用退化的空间分布信息。当退化区域与固定窗口边界不对齐时，强制切割会在边界引入分割伪影；面对多退化共存的真实场景时，统一处理必然导致局部区域“过度处理”或“处理不足”。这一“空间感知”能力的缺失，驱动了统一化 All-in-One 方法与基于智能体的动态决策范式的发展。

## （二）统一化 All-in-One（AiO）图像恢复方法

AiO 恢复的核心挑战在于使单一模型实现对多种退化类型的判别性表征与差异化处理，同时避免对已习得恢复能力的“灾难性遗忘”。AirNet<sup>[3]</sup>通过对比式降质编码器（CBDE）以无监督方式构建退化判别表征，利用不同退化类型样本之间的对比损失驱动特征空间的有序组织，实现了在去噪、去雾、去雨三类任务上无需显式退化标签的自适应处理，开创了基于对比学习的 AiO 恢复研究范式。Hu 等人<sup>[12]</sup>进一步将对比学习扩展至图像、特征、块三个

层级，构建多级协同对比框架，在提升退化判别精度的同时增强了对未见退化强度的泛化能力。

提示学习范式的引入是 AiO 恢复方法论的重大进展，其核心思想是将退化类型信息以“提示向量”的形式显式注入网络，指导统一骨干网络做出任务差异化的响应。PromptIR<sup>[4]</sup>通过可学习提示向量对 Transformer 主干进行软条件化（soft-conditioning），在去噪、去雾、去雨、去模糊四类任务内实现统一恢复，验证了连续提示空间能够自然表征不同退化之间的过渡状态。后续多项工作从不同角度扩展和深化了提示范式：Ma 等人<sup>[13]</sup>的 ProRes 从退化先验知识中显式构建可解释的视觉提示，增强了提示向量的语义可解释性；Wu 等人<sup>[14]</sup>提出输入自适应的动态提示生成策略，使提示内容随输入图像的退化特征自动调整；Wu 等人<sup>[15]</sup>在 FrePrompter 中直接从输入图像的频谱特征中提取物理有据的退化提示，赋予提示向量明确的频域物理解释；Wu 等人<sup>[16]</sup>通过对比提示学习解决不同退化类型间提示表征高度相关（冗余）的问题，在多个 AiO 基准上建立了新的最优结果。

在语言引导与架构创新方向，InstructIR<sup>[5]</sup>首次将自然语言理解引入图像恢复，以用户自然语言指令控制恢复过程，可同时处理七类任务，实现了“以语言为接口”的新型人机协同恢复模式，也预示着图像恢复与大语言模型交互的整合趋势。Tang 等人<sup>[17]</sup>提出 RamIR，将 Mamba 状态空间架构与链式思维（Chain-of-Thought）推理相结合，通过显式推理-行动分解处理复杂混合退化，将语言推理能力引入轻量化图像恢复框架。Zamfir 等人<sup>[18]</sup>提出 MoCE-IR 混合专家架构，通过动态稀疏专家激活将计算资源按退化处理难度自适应分配，使简单退化由少数专家高效处理、复杂退化调用更多专家协同完成，在多个基准上达到最优性能。Wu 等人<sup>[19]</sup>在“多样中的和谐”框架中，通过动态生成退化感知卷积核，使单一网络能为每个输入图像定制滤波响应，无需显式退化类型切换。Dudhane 等人<sup>[20]</sup>采用动态预训练策略，以课程学习方式将模型渐进暴露于从简单到复杂的退化场景，显著提升了模型对未见退化组合的泛化能力。

然而，Jiang 等人<sup>[21]</sup>在系统综述中明确指出 AiO 方法存在三项难以克服的持续局限：第一，性能在遇到训练集外的新型退化时显著下降，泛化能力受训练集退化类型的强约束；第二，全图单一处理流程从根本上忽视了退化的空间异质性，对整幅图像施加相同的恢复强度和策略，在退化分布不均匀的真实场景中必然导致局部区域处理不当；第三，单次前向推理的静态处理模式无法基于中间恢复质量进行迭代细化，一旦恢复策略选择有误便无从自我修正。这三项局限共同构成了智能体化方法兴起的直接动因，也明确界定了本项目研究的问题空间。

### （三）基于智能体的图像恢复方法

基于智能体的方法代表着图像恢复领域的新兴范式，将恢复过程重新定义为具有感知-规划-执行-评估闭环的智能决策问题：视觉语言模型（Visual Language Model, VLM）或大语言模型（Large Language Model, LLM）作为协调决策者，动态分析图像退化、规划工具调用序列、评估中间恢复结果质量，并根据评估反馈迭代调整策略。与前述所有方法相比，这一范式的根本突破在于引入了动态推理和迭代细化机制，使恢复过程能够适应每张图像的具体退化特征，而非套用预先设计的固定策略。

AgenticIR<sup>[26]</sup>是基于智能体图像恢复领域的开创性工作，系统化建立了五阶段动态规划

框架：（1）感知阶段——VLM 全面分析输入图像的退化类型、严重程度及空间分布；（2）调度阶段——基于感知结果规划专用恢复工具的调用序列；（3）执行阶段——按计划调用图像恢复工具库中的对应工具完成局部处理；（4）反思阶段——VLM 评估中间恢复图像的质量，判断是否已满足恢复目标；（5）重调度阶段——若质量不足则修正计划并触发新一轮迭代。该框架支持动态组合任意专用工具序列，并通过深度优先搜索在解空间中探索优质方案，与此前所有方法相比对复杂退化的应对能力有了根本性突破；然而，深度优先搜索策略在最坏情况下导致指数级计算代价，是实时部署的主要障碍。该框架进一步引入按退化复杂度三分（简单、中等、复杂）的自适应系统与经验记忆模块，积累和复用成功的恢复决策路径，使系统能够在部署过程中随经验积累持续自我改进，而无需周期性重训练。

在多智能体协同与效率提升方面，MAIR<sup>[27]</sup>引入三阶段退化先验（压缩失真→成像退化→场景退化）构建多智能体协作架构，利用先验知识大幅压缩工具选择的搜索空间，并对独立子任务进行并行化处理，在与 AgenticIR 相当的恢复质量下将推理效率提升 44%，显著缩小了基于智能体方法与实时部署之间的效率差距。RestoreAgent<sup>[28]</sup>通过在大规模恢复任务-决策对上端到端微调多模态 LLM，将多步推理压缩为单次前向传播，以牺牲部分动态规划灵活性为代价换取推理速度的大幅提升；Q-Agent<sup>[29]</sup>以图像质量评估模型预测的预期质量提升作为贪婪选择准则，在每一步选择最大化质量提升的工具，将规划复杂度从指数级降至相对于工具数量的线性复杂度，显著提升了大规模工具库场景下的实用部署可行性。

然而，上述所有基于智能体的系统均将整幅图像视为单一实体处理，制定并执行全局统一的恢复策略，未能在决策框架中纳入退化的空间分布信息。在真实复杂场景中，图像前景受运动模糊影响、背景弥漫雾霾、暗部阴影区域充斥噪声，同时还可能横贯整体的雨纹——以统一全图策略处理此类图像，必然在部分区域造成“过度处理”、在部分区域造成“处理不足”，导致明显的局部视觉质量损失。这一“空间感知”能力的缺失，构成了本项目直接的研究起点。

#### （四）空间感知工具基础与视觉退化分析

实现空间感知图像恢复所需的关键工具链已经逐步成熟，为本项目的技术路线提供了坚实的基础设施支撑，也表明当前是开展此类系统整合研究的最佳时机。

在图像空间分解方面，Kirillov 等人<sup>[30]</sup>提出的 SAM (Segment Anything Model) 以超过 10 亿个掩码的大规模开放图像数据集训练，提供了强大的零样本语义区域分割能力，可在无需特定退化类型先验的情况下将图像分解为具有语义边界的空间区域，为区域级退化感知提供了基础工具。SAM 采用简洁的接口设计（点、框、文本等多种提示格式），可以便捷地嵌入智能体系统的工具调用流程中。然而，标准 SAM 直接应用于退化图像时存在分割质量下降的问题：退化导致的模糊边界、噪声纹理和对比度损失会干扰 SAM 的特征提取，造成过分割、分割漂移或边界不准确等现象，在高噪声或严重模糊的图像上尤为明显。针对这一问题，Chen 等人<sup>[31]</sup>提出 RobustSAM，在保持与 SAM 相同灵活接口的基础上，通过专项训练在噪声、模糊、低对比度等各类退化条件下保持稳健分割性能，使分割驱动的空间分解能够在真实退化图像上可靠运行，为智能体系统在退化场景下的区域感知提供了保障。Ren 等人<sup>[32]</sup>提出 Grounded SAM，将开放词汇目标检测与 SAM 精细分割相结合，支持通过任意自然语言提示精确定位特定语义区域，例如“受雾霾影响的天空区域”或“强噪声干扰的阴

影区域”。这种语言引导的精细空间定位能力使协调智能体能够通过自然语言表达对特定退化区域的感知结果，直接连接 VLM 的语言推理输出与分割工具的区域定位输入，构建了完整的语言引导空间感知链路。

在视觉退化描述与质量评估方面，You 等人<sup>[33]</sup>提出 DepictQA 系列工作，借助视觉语言模型提供对图像质量属性的详细自然语言描述，内容包括具体退化类型（噪声、模糊、雾霾等）、退化严重程度（轻微、中等、严重），以及受各类退化影响的空间区域，其表达粒度远超传统无参考质量评估方法的单一质量分值，为智能体感知阶段提供了语义丰富、空间精确的退化信息输入。Wu 等人<sup>[34]</sup>构建了 Q-Instruct——专门面向低级视觉感知任务的大规模指令遵循数据集，包含丰富的退化描述、质量比较和感知分析等指令类型；在此数据集上微调的视觉语言模型在退化类型识别和细粒度质量描述方面显著优于通用 VLM，为构建区域级退化感知的 VLM 提供了数据构建方法论参考。Wu 等人<sup>[35]</sup>建立的 Q-Bench 基准系统化覆盖低级视觉感知、图像质量描述和比较评估三类核心任务，提供了视觉语言模型在图像质量感知场景下能力评估的标准化协议，为本项目选择合适的基础 VLM 感知组件提供了方法参照。

上述空间分解工具（SAM、RobustSAM、Grounded SAM）与视觉退化分析工具（DepictQA、Q-Instruct、Q-Bench）的汇聚，在技术层面已为本项目所设想的空间感知多智能体恢复框架提供了完整的组件基础：协调智能体调用 Grounded SAM 将图像分解为退化语义区域；经 Q-Instruct 微调的 VLM 对各空间区域独立进行退化类型识别与严重程度评估；区域专用子智能体针对各区域的具体退化特征施加定制化恢复策略；各区域恢复结果经融合模块无缝整合为全局一致的高质量输出。然而，协调上述工具与多智能体协同恢复策略的系统性框架在方法论层面尚属空白——现有基于智能体的方法均未将空间感知能力纳入其感知-规划-执行框架，这正是本项目拟填补的核心研究缺口。

## （五）现有研究的局限性

综合上述分析，现有研究存在以下三个层面的理论局限，构成本项目的直接研究动因。

### （1）退化表征层面：空间异质性建模不足

现有方法对退化的表征停留在图像级或全局特征级，缺乏对“退化类型—严重程度—空间位置”耦合关系的区域级建模能力。Transformer 类方法虽具备全局上下文建模能力，但其窗口划分策略与退化区域边界往往不对齐，导致在窗口边界引入分割伪影；AiO 方法的提示向量虽能区分不同退化类型，但无法定位退化在空间上的具体分布；基于智能体的方法虽引入动态决策，但感知阶段缺乏对空间区域的精细化分析能力。这一表征粒度的缺失，使得现有方法在面对异质退化时难以实现区域级的差异化处理。

### （2）恢复决策层面：缺乏工具链与参数的联合优化机制

现有基于智能体的方法在决策层面存在两大瓶颈：一是工具选择策略多基于贪婪搜索或经验规则，缺乏对“工具链组合—超参数配置—区域特征”的联合优化能力；二是规划与执行之间存在语义鸿沟，协调智能体生成的抽象计划难以精确映射到具体工具的参数配置。这导致恢复策略的选择往往次优，且无法根据区域级退化特征进行自适应调整。

### （3）部署效率层面：大模型依赖导致的实时性瓶颈

现有智能体化方法普遍依赖 VLM/LLM 进行感知与规划，虽然增强了决策灵活性，但

大模型的推理时延和算力消耗显著制约了实时部署可行性。如何在保持空间感知与动态决策能力的前提下，通过知识蒸馏、轨迹压缩等手段降低推理开销，是从研究原型走向实用部署的关键挑战。

### 1.3. 本课题独到的学术价值和应用价值

#### (一) 学术价值:

- (1) 提出空间退化图的区域级表征方法，将退化建模从图像级细化为区域级，为异质退化场景的分析与处理提供新的技术途径；
- (2) 构建多智能体协同的恢复框架，通过任务分解与分工协作实现差异化处理，为复杂场景图像恢复提供可扩展的架构参考；
- (3) 探索规划轨迹蒸馏机制，在保持恢复质量的同时降低模型部署成本，为高效图像恢复系统的实现提供可行方案。

#### (二) 应用价值:

- (1) 提升安防监控与自动驾驶场景下的图像可用性，改善夜间及恶劣天气条件下的视觉感知能力；
- (2) 改善遥感图像的地物边界与纹理细节质量，辅助变化检测与精细化解译任务；
- (3) 为医学成像、工业视觉质检等相关领域提供可迁移的技术框架，支撑低质量影像的质量增强需求。

## 2. 研究目标、总体框架及重点难点

### 2.1. 研究目标

面向复杂退化场景下图像恢复的精细化与高效化需求，本项目拟实现以下可评价目标：

- 1) **感知：**构建空间感知退化分析机制，实现区域级退化识别与量化。在公开复杂退化基准上，相较全图级识别基线，区域级退化识别准确率显著提升。对应输出物：空间退化图构建算法、区域级标注与评测脚本、阶段论文成果。
- 2) **决策：**构建层级多智能体协同机制，实现区域级恢复规划与连续参数自适应优化。在混合退化恢复任务中，相较静态恢复管线获得更高恢复质量与更优效率平衡。对应输出物：层级智能体恢复系统原型、工具链调度策略、发明专利/论文成果。
- 3) **部署：**构建轨迹蒸馏驱动的轻量路由模型，实现高质量恢复策略的快速推理与边缘部署。在保持主要质量指标可接受损失范围内显著降低部署时延与算力需求。对应输出物：轻量路由网络模型、部署测试报告、可复现实验配置。

### 2.2. 总体框架

本项目总体框架如图 2所示，由“空间感知模块—层级规划模块—轻量部署模块”三大核心模块组成。空间感知模块负责将退化图像分割为语义区域并生成空间退化图；层级规划模块基于全局调度器与区域专家的协同架构，对各区域独立规划恢复策略并迭代优化；轻量部署模块通过轨迹蒸馏将多智能体的复杂决策知识压缩为轻量路由网络。三个模块层层递进，形成“感知—决策—执行—蒸馏”的闭环流程。

其中，系统输入为退化图像  $y$ ；模块一输出空间退化图  $M$ ；模块二输出区域修复结果及全图恢复结果  $\hat{x}$ ；模块三输出轻量推理模型  $\Phi_\omega$  及对应调度参数。三模块在训练阶段通过轨迹数据闭环耦合，在部署阶段形成“轻量网络 + 工具库”的高效执行路径。

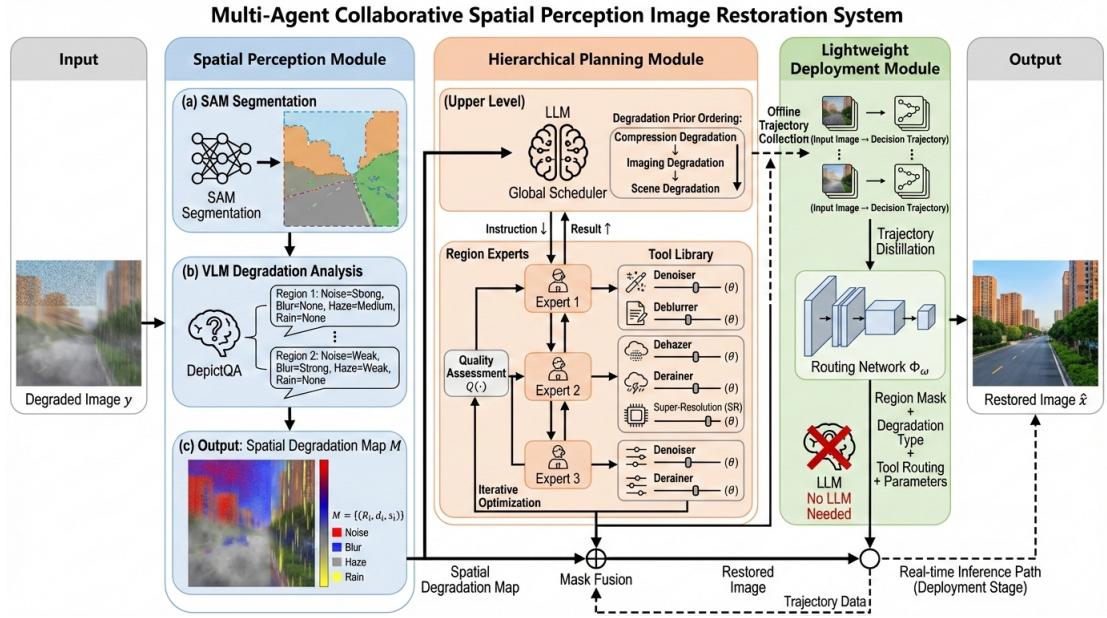


图 2: 本项目总体研究框架

### 2.3. 研究内容

面向上述目标，本项目拟开展以下三个方面研究。

#### (一) 空间感知退化分析

现有方法主要基于全图级退化判断，难以刻画复杂场景的空间异质退化。针对这一问题，本研究将构建区域级退化感知机制，实现“退化类型—严重程度—空间位置”的联合建模。

本研究首先利用零样本分割模型 SAM<sup>[30]</sup>将退化图像划分为若干语义一致的区域。SAM 基于超过 10 亿掩码的预训练，具备良好的零样本泛化能力，但直接应用于退化图像时存在分割质量下降问题。为此，本研究将引入 RobustSAM<sup>[31]</sup>作为备选方案，其在噪声、模糊、低对比度等退化条件下保持稳健分割性能。

对每个分割区域，将其裁剪图像输入经 Q-Instruct 数据集微调的视觉语言模型，采用链式思维推理逐类判断退化是否存在及其程度，输出退化类型判别与对应严重程度估计。为提升分析精度，引入空间上下文聚合机制，通过图注意力机制聚合邻域退化信息，修正孤立区域的误判。汇总所有区域分析结果，构建空间退化图，为后续智能体规划提供结构化输入。

#### (二) 层级多智能体协同规划

单一静态恢复流程难以兼顾复杂退化下的质量与效率。针对这一问题，本研究将构建“

全局调度器 + 区域专家”的两层智能体架构，实现工具链选择与连续参数优化的一体化决策。

全局调度器依据退化先验（场景退化 → 成像退化 → 压缩退化的逆序处理原则<sup>[27]</sup>），为各区域分配处理优先级并制定整体修复策略。该先验基于物理成像过程的逆向逻辑：压缩退化（JPEG 伪影）发生在图像编码阶段，应在最后处理；成像退化（噪声、模糊）源于采集设备；场景退化（雾霾、雨线）来自环境因素，应优先处理以还原场景辐射。

每个区域由专属的区域专家负责处理。区域专家从预构建的工具库中选择合适的恢复工具序列，并为每个工具设定连续参数（如去噪强度、超分辨率尺度）。工具库包含去噪器、去模糊器、去雾器、去雨器、超分辨率模型、JPEG 伪影消除器等基础恢复模块。区域专家通过迭代方式优化工具参数，直至修复结果满足质量要求。

全局调度器通过区域掩码加权融合各区域修复结果，合成最终恢复图像，并对全图进行整体质量评估与一致性检查，确保跨区域边界处的平滑过渡。

### （三）基于轨迹蒸馏的轻量路由网络

多智能体高质量推理依赖 LLM/VLM 在线调用，导致时延和算力成本偏高，不利于边缘部署与实时应用。针对这一问题，本研究将构建轨迹蒸馏机制，将复杂多步推理知识压缩为可部署的轻量决策网络。

利用上述多智能体系统在大规模合成退化数据集上离线运行，收集决策轨迹数据。每条轨迹记录输入图像、空间退化图、各区域最优工具序列与参数配置。基于该数据集训练轻量级路由网络，使其直接根据退化图像预测区域划分、退化类型及工具调度方案。

路由网络采用编码器—多任务解码器架构。编码器提取图像特征；解码器分为四个分支头，分别输出区域掩码、退化类型概率、工具选择概率和连续参数值。通过多任务联合训练，使网络同时学习分割、分类、路由与参数回归四项任务。

部署时仅需轻量路由网络与恢复工具库，无需依赖 LLM/VLM 在线推理，可实现端侧实时推理。为缓解蒸馏偏差问题，采用高质量轨迹重加权与困难样本再训练策略。

## 2.4. 重点难点

本项目的特点难点包括：

1) **难点 1（对应目标 1）：**如何在零样本条件下准确分割退化区域并量化退化程度。退化导致的模糊边界、噪声纹理和对比度损失会干扰分割质量，造成过分割、分割漂移或边界不准确。拟解决路径：通过“语义分割 + 区域级退化推理 + 一致性约束”联合优化空间退化图质量；引入 RobustSAM 作为备选分割方案；设置区域置信度门控与候选掩码回退机制，降低分割误差向后续模块的传播。

2) **难点 2（对应目标 2）：**如何在空间维度上协调多智能体决策，避免区域边界不一致与策略冲突。不同区域专家的独立决策可能导致边界处处理痕迹明显，且工具链调用顺序会影响最终效果。拟解决路径：通过全局调度器的优先级约束实现跨区域协同；引入区域掩码渐变融合机制，在边界处实现平滑过渡；通过质量反馈闭环与全局一致性检查发现并修正区域间冲突。

3) **难点 3（对应目标 3）：**如何将复杂多步推理知识有效压缩到轻量网络中并保持性能

稳定。多智能体的迭代优化策略涉及复杂的状态转移逻辑，轻量网络容量有限可能导致知识蒸馏不充分。拟解决路径：采用高质量轨迹筛选，仅使用多智能体收敛后的最优决策作为监督信号；引入困难样本重训练机制，针对蒸馏偏差较大的样本进行迭代优化；设置性能降级阈值，当蒸馏损失过大时回退到轻量网络 + 有限 LLM 调用的混合模式。

以上难点分别对应“感知—决策—部署”三类核心目标，上述解决路径已在技术路线中预留备选方案与回退机制，确保项目可执行性。

### 3. 思路方法

本节围绕“如何做成”展开，按照技术流程、实施计划与可行性支撑三层推进，重点回答模型如何训练、如何推理、如何规避风险。

#### 3.1. 具体研究方法

本项目的整体研究思路为“感知  $\rightarrow$  规划  $\rightarrow$  执行  $\rightarrow$  蒸馏”。首先构建空间感知模块实现区域级退化分析，然后设计层级智能体进行协同规划与执行，最后通过轨迹蒸馏实现轻量化部署。三个研究内容层层递进、相互依托。具体方法流程如图 3 所示。

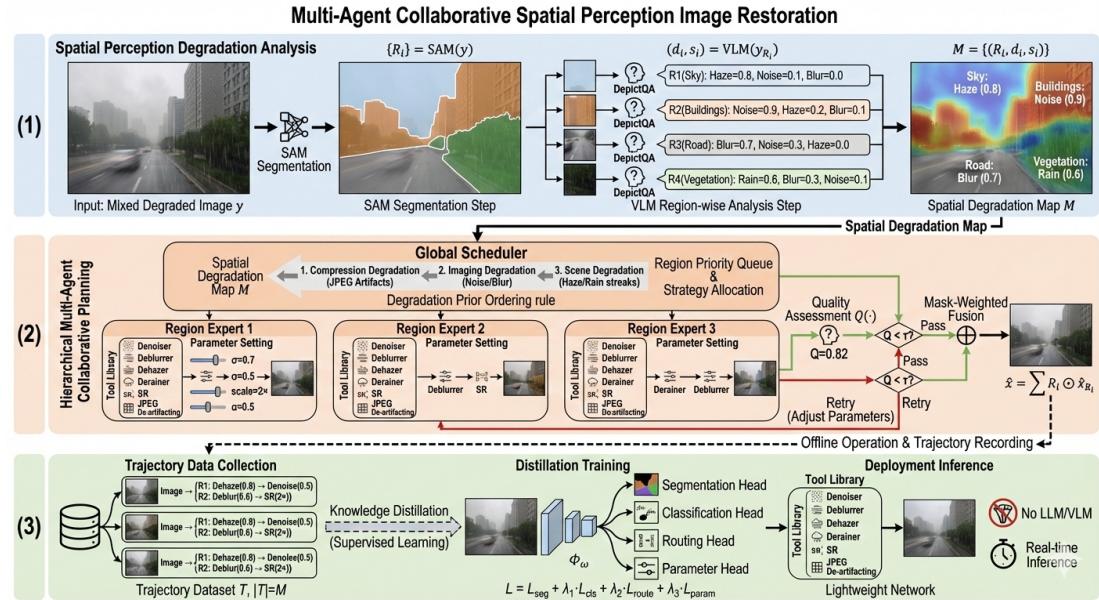


图 3: 具体研究方法流程

#### (1) 空间感知退化分析

给定退化图像  $y \in \mathbb{R}^{H \times W \times 3}$ ，本模块旨在生成空间退化图  $M$ ，为后续智能体规划提供结构化输入。

**步骤一：语义区域分割。**利用零样本分割模型 SAM<sup>[30]</sup>将图像划分为  $N$  个语义一致的区域：

$$\{R_i\}_{i=1}^N = \text{SAM}(y), \quad \bigcup_{i=1}^N R_i = \Omega, \quad R_i \cap R_j = \emptyset \quad (i \neq j) \quad (1)$$

其中  $\Omega$  为图像像素域。SAM 基于超过 10 亿掩码的预训练，具备良好的零样本泛化能力。然而，直接应用于退化图像时，模糊边界、噪声纹理和对比度损失会导致分割质量下降。为此，本研究引入置信度门控机制：

$$R_i^{\text{final}} = \begin{cases} R_i, & \text{if } \mathcal{C}(R_i) > \tau_{\text{seg}} \\ R_i^{\text{robust}}, & \text{otherwise} \end{cases} \quad (2)$$

其中  $\mathcal{C}(R_i)$  为 SAM 对区域  $R_i$  的分割置信度， $\tau_{\text{seg}}$  为预设阈值。当置信度低于阈值时，启用 RobustSAM<sup>[31]</sup>重新分割，其在噪声、模糊、低对比度等退化条件下保持稳健性能。

**步骤二：区域级退化分析。**对每个分割区域  $R_i$ ，提取其裁剪图像  $\mathbf{y}_{R_i} = \mathbf{y} \odot \mathbf{M}_i$ ，其中  $\mathbf{M}_i \in \{0, 1\}^{H \times W}$  为区域掩码。将  $\mathbf{y}_{R_i}$  输入经 Q-Instruct 数据集微调的视觉语言模型  $V_\theta$ ，采用链式思维推理逐类判断退化：

$$\mathbf{d}_i^{(c)}, s_i^{(c)} = V_\theta(\text{Prompt}_c, \mathbf{y}_{R_i}), \quad c = 1, \dots, C \quad (3)$$

其中  $\text{Prompt}_c$  为针对第  $c$  类退化的查询提示。为提升分析精度，引入空间上下文聚合：

$$\mathbf{h}_i = \text{ContextAgg}(\{V_\theta(\mathbf{y}_{R_j})\}_{j \in \mathcal{N}(i)}) \quad (4)$$

其中  $\mathcal{N}(i)$  表示区域  $R_i$  的空间邻域集合，通过图注意力机制聚合邻域退化信息，修正孤立区域的误判。

**步骤三：空间退化图构建。**汇总所有区域分析结果，构建结构化空间退化图：

$$\mathcal{M} = \{(R_i, \mathbf{d}_i, \mathbf{s}_i, \mathbf{h}_i)\}_{i=1}^N \quad (5)$$

该图显式编码了退化的空间分布、类型判别与严重程度估计，可直接作为后续决策模块的状态描述。

**验证方案：**在 SIDD 去噪、GoPro 去模糊、NH-HAZE 去雾等公开基准上，采用分类准确率 (Acc)、平均精度 (mAP) 和 F1 分数评估退化识别性能。对比方案包括：(1) 全图级 DepictQA 分析基线；(2) SAM 分割 + 独立区域分析 (无上下文聚合)；(3) 本研究完整方法。通过消融实验量化各组件 (RobustSAM 回退、上下文聚合) 对最终恢复质量 (PSNR/SSIM) 的贡献。

## (2) 层级多智能体协同规划

基于空间退化图  $\mathcal{M}$ ，本研究设计全局调度器  $G_\phi$  与区域专家  $\{E_i\}_{i=1}^N$  的协同架构。全局调度器负责跨区域策略协调，区域专家负责区域内恢复决策。

**步骤一：全局调度与优先级分配。**全局调度器依据退化先验 (场景退化 → 成像退化 → 压缩退化的逆序处理原则<sup>[27]</sup>)，为各区域分配处理优先级。定义退化优先级函数：

$$\mathcal{P}(R_i) = \sum_{c=1}^C w_c \cdot d_i^{(c)} \cdot s_i^{(c)} \quad (6)$$

其中  $w_c$  为第  $c$  类退化的先验权重（场景退化  $w_c > 1$ , 成像退化  $w_c = 1$ , 压缩退化  $0 < w_c < 1$ ）。全局调度器按  $\mathcal{P}(R_i)$  降序排列区域处理顺序，并生成全局策略向量：

$$\mathbf{g} = G_\phi(\mathcal{M}) = \text{Softmax}(\text{MLP}(\{\mathbf{h}_i\}_{i=1}^N)) \quad (7)$$

其中  $\mathbf{g} \in \mathbb{R}^N$  编码各区域的处理优先级权重。

**步骤二：区域专家工具链选择与参数优化。**对于区域  $R_i$ , 区域专家  $E_i$  从预构建的工具库  $\mathcal{F} = \{f_1, f_2, \dots, f_L\}$  中选择恢复工具序列  $\pi_i = (f_{\sigma_1}, \dots, f_{\sigma_{K_i}})$ , 并为每个工具设定连续参数  $\theta_i = (\theta_{\sigma_1}, \dots, \theta_{\sigma_{K_i}})$ 。区域修复结果通过函数复合计算：

$$\hat{\mathbf{x}}_{R_i} = (f_{\sigma_{K_i}} \circ \dots \circ f_{\sigma_1})(\mathbf{y}_{R_i}; \theta_i) \quad (8)$$

工具链选择采用策略梯度方法优化。定义策略网络  $\pi_\psi(\cdot | \mathbf{h}_i, \mathbf{g})$  输出工具选择概率分布，通过最大化期望恢复质量进行训练：

$$\max_{\psi} \mathbb{E}_{\pi_i \sim \pi_\psi} [Q((f_{\sigma_{K_i}} \circ \dots \circ f_{\sigma_1})(\mathbf{y}_{R_i}))] - \beta \cdot \text{Length}(\pi_i) \quad (9)$$

其中  $Q(\cdot)$  为质量评估函数,  $\beta$  为工具链长度惩罚系数, 鼓励简洁高效的处理流程。

**步骤三：连续参数自适应优化。**对于选定的工具链  $\pi_i$ , 区域专家通过迭代优化确定最优参数：

$$\theta_i^* = \arg \max_{\theta_i} Q(\hat{\mathbf{x}}_{R_i}(\theta_i)), \quad \text{s.t.} \quad Q(\hat{\mathbf{x}}_{R_i}(\theta_i)) \geq \tau \quad (10)$$

其中  $\tau$  为质量阈值。采用贝叶斯优化或梯度上升法求解, 迭代更新直至收敛或达到最大迭代次数  $T_{\max}$ 。

**步骤四：全局融合与一致性约束。**全局调度器通过区域掩码加权融合各区域结果：

$$\hat{\mathbf{x}} = \sum_{i=1}^N \mathbf{M}_i \odot \hat{\mathbf{x}}_{R_i} + \lambda_{\text{smooth}} \cdot \nabla^2 \hat{\mathbf{x}} \quad (11)$$

其中  $\lambda_{\text{smooth}}$  为平滑正则系数, 通过拉普拉斯算子约束边界一致性。全局质量评估通过一致性损失实现跨区域协调：

$$\mathcal{L}_{\text{consist}} = \sum_{(i,j) \in \mathcal{E}} \left\| \hat{\mathbf{x}}_{R_i}^{\text{border}} - \hat{\mathbf{x}}_{R_j}^{\text{border}} \right\|_2^2 \quad (12)$$

其中  $\mathcal{E}$  为相邻区域边界的像素集合。

**验证方案：**在混合退化合成数据集（通过 SIDD、GoPro、NH-HAZE 等组合生成）与真实测试样本上, 与 AirNet、PromptIR 等 All-in-One 静态方法以及 AgenticIR、MAIR 等智能体方法进行全面对比。评估指标包括：(1) 恢复质量：PSNR、SSIM、LPIPS；(2) 推理效率：FPS、平均工具调用次数；(3) 决策质量：工具链选择准确率、参数优化收敛速度。通过消融实验验证层级架构各组件（全局调度、区域专家、质量反馈、一致性约束）的贡献度。

### (3) 基于轨迹蒸馏的轻量路由网络

**步骤一：决策轨迹数据集构建。**利用上述多智能体系统在大规模合成退化数据集  $\mathcal{D}_{\text{syn}} = \{(\mathbf{y}^{(j)}, \mathbf{x}^{(j)})\}_{j=1}^M$  上离线运行，收集决策轨迹数据集：

$$\mathcal{T} = \left\{ \left( \mathbf{y}^{(j)}, \mathcal{M}^{(j)}, \left\{ \pi_i^{*,(j)}, \boldsymbol{\theta}_i^{*,(j)}, \mathbf{M}_i^{(j)} \right\}_{i=1}^{N_j} \right) \right\}_{j=1}^M \quad (13)$$

其中每条轨迹记录输入图像  $\mathbf{y}^{(j)}$ 、空间退化图  $\mathcal{M}^{(j)}$ 、各区域最优工具序列  $\pi_i^*$ 、最优参数  $\boldsymbol{\theta}_i^*$  以及区域掩码  $\mathbf{M}_i$ 。

为提升轨迹质量，引入结果过滤机制：仅保留最终恢复质量  $Q(\hat{\mathbf{x}}^{(j)}) \geq \tau_{\text{high}}$  的轨迹参与蒸馏。同时，采用轨迹多样性采样确保退化类型、区域数量、工具链长度的分布均衡。

**步骤二：轻量路由网络架构设计。**设计轻量级路由网络  $\Phi_\omega$ ，采用编码器—多任务解码器架构：

编码器采用轻量化骨干网络（如 MobileNet-V3 或 EfficientNet-B0）提取多尺度特征金字塔  $\{\mathbf{F}_l\}_{l=1}^L$ ，其中  $\mathbf{F}_l \in \mathbb{R}^{H/2^l \times W/2^l \times C_l}$ 。

解码器分为四个并行分支头：

(a) **分割头  $\Phi_\omega^{\text{seg}}$** ：输出区域掩码概率图  $\hat{\mathbf{P}}^{\text{mask}} \in \mathbb{R}^{H \times W \times N_{\text{max}}}$ ，通过 Softmax 获取区域归属概率：

$$\hat{\mathbf{M}}_i = \mathbb{1} \left[ \arg \max_k \hat{P}^{\text{mask}}(u, v, k) = i \right] \quad (14)$$

(b) **分类头  $\Phi_\omega^{\text{cls}}$** ：对每个区域输出退化类型概率分布：

$$\hat{\mathbf{d}}_i = \text{Sigmoid} \left( \text{GAP} \left( \mathbf{F}_{\text{roi}}^{(i)} \right) \right) \in [0, 1]^C \quad (15)$$

其中 GAP 为全局平均池化， $\mathbf{F}_{\text{roi}}^{(i)}$  为区域  $R_i$  对应的 RoI 特征。

(c) **路由头  $\Phi_\omega^{\text{route}}$** ：输出工具选择概率矩阵：

$$\hat{\mathbf{P}}_i^{\text{route}} = \text{Softmax} \left( \text{MLP} \left( \mathbf{F}_{\text{roi}}^{(i)} \oplus \hat{\mathbf{d}}_i \right) \right) \in \mathbb{R}^L \quad (16)$$

其中  $\oplus$  表示特征拼接。

(d) **参数头  $\Phi_\omega^{\text{param}}$** ：对每个选中工具输出连续参数值：

$$\hat{\boldsymbol{\theta}}_i = \text{Sigmoid} \left( \text{MLP} \left( \mathbf{F}_{\text{roi}}^{(i)} \right) \right) \cdot \boldsymbol{\theta}_{\text{max}} \quad (17)$$

其中  $\boldsymbol{\theta}_{\text{max}}$  为各参数的最大允许值。

**步骤三：多任务联合蒸馏训练。**基于轨迹数据集进行监督学习，联合优化四个任务：

$$\mathcal{L} = \mathcal{L}_{\text{seg}} + \lambda_1 \mathcal{L}_{\text{cls}} + \lambda_2 \mathcal{L}_{\text{route}} + \lambda_3 \mathcal{L}_{\text{param}} + \lambda_4 \mathcal{L}_{\text{kd}} \quad (18)$$

各损失项定义如下：

$$\mathcal{L}_{\text{seg}} = \frac{1}{M} \sum_{j=1}^M \sum_{i=1}^{N_j} \text{Dice} \left( \hat{\mathbf{M}}_i^{(j)}, \mathbf{M}_i^{*,(j)} \right) \quad (19)$$

$$\mathcal{L}_{\text{cls}} = \frac{1}{M} \sum_{j=1}^M \sum_{i=1}^{N_j} \text{BCE} \left( \hat{\mathbf{d}}_i^{(j)}, \mathbf{d}_i^{*,(j)} \right) \quad (20)$$

$$\mathcal{L}_{\text{route}} = \frac{1}{M} \sum_{j=1}^M \sum_{i=1}^{N_j} \sum_{k=1}^{K_i} \text{CE} \left( \hat{\mathbf{P}}_{i,k}^{\text{route}}, f_{\sigma_k}^{*(j)} \right) \quad (21)$$

$$\mathcal{L}_{\text{param}} = \frac{1}{M} \sum_{j=1}^M \sum_{i=1}^{N_j} \sum_{k=1}^{K_i} \left\| \hat{\boldsymbol{\theta}}_{i,k}^{(j)} - \boldsymbol{\theta}_{i,k}^{*,(j)} \right\|_2^2 \quad (22)$$

其中  $\lambda_1, \lambda_2, \lambda_3, \lambda_4$  为任务平衡系数。知识蒸馏损失  $\mathcal{L}_{\text{kd}}$  用于保留多智能体的“软”决策信息：

$$\mathcal{L}_{\text{kd}} = \text{KL} (\text{Softmax}(\hat{\mathbf{z}}/T) \| \text{Softmax}(\mathbf{z}^*/T)) \quad (23)$$

其中  $\hat{\mathbf{z}}$  和  $\mathbf{z}^*$  分别为路由网络和多智能体的 logits 输出， $T$  为温度系数。

**步骤四：困难样本再训练与部署优化。**针对蒸馏偏差问题，实施两阶段训练策略：

第一阶段：在全量轨迹数据上训练至收敛。

第二阶段：识别困难样本集  $\mathcal{T}_{\text{hard}} = \{(\mathbf{y}^{(j)}, \cdot) \in \mathcal{T} : Q(\hat{\mathbf{x}}^{(j)}) < Q(\mathbf{x}_{\text{agent}}^{(j)}) - \delta\}$ ，即路由网络与多智能体性能差距超过阈值  $\delta$  的样本。在该子集上进行困难样本聚焦训练，损失函数加权：

$$\mathcal{L}_{\text{hard}} = \sum_{(\mathbf{y}^{(j)}, \cdot) \in \mathcal{T}_{\text{hard}}} w_j \cdot \mathcal{L}^{(j)}, \quad w_j \propto Q(\mathbf{x}_{\text{agent}}^{(j)}) - Q(\hat{\mathbf{x}}^{(j)}) \quad (24)$$

部署阶段，轻量路由网络直接根据输入图像预测区域划分、退化类型、工具调度与参数配置，无需 LLM/VLM 在线推理，实现端到端实时恢复：

$$\left\{ \hat{\mathbf{M}}_i, \hat{\mathbf{d}}_i, \hat{\pi}_i, \hat{\boldsymbol{\theta}}_i \right\}_{i=1}^{\hat{N}} = \Phi_{\omega}(\mathbf{y}) \quad (25)$$

**验证方案：**在测试集上对比蒸馏后的轻量路由网络与原多智能体系统的性能差距，要求 PSNR 损失  $\leq 5\%$ 、SSIM 损失  $\leq 3\%$ 。测试部署时延 (ms/图像)、模型参数量 (MB)、计算量 (FLOPs)，验证在 Jetson Nano 或同等边缘设备上的实时推理可行性 (目标： $\geq 10$  FPS @ 720p)。通过消融实验验证：(1) 多任务联合训练 vs 单任务独立训练；(2) 困难样本再训练 vs 单阶段训练；(3) 知识蒸馏损失 vs 硬标签监督。

**训练与推理区别：**训练阶段使用离线轨迹监督并联合优化多任务损失；推理阶段仅执行前向路由预测与工具调用，避免大模型在线推理开销。

**风险与回退策略：**针对分割误差传播问题，设置区域置信度门控与候选掩码回退；针对规划误差累积问题，引入全局一致性约束与早停机制；针对蒸馏偏差问题，采用高质量轨迹重加权与困难样本再训练。

### 3.2. 研究计划

研究计划按“阶段目标—里程碑—可交付件—验收标准”组织，分两年六个阶段推进：

### **第一阶段（第 1-4 月）：空间感知模块基础构建**

阶段目标：完成空间感知退化分析的基础架构，实现 SAM 分割与 VLM 退化识别的初步联调。

里程碑：建立 SAM 分割流程，完成 DepictQA 等 VLM 模型的区域级退化识别适配；形成空间退化图的初步生成能力。

可交付件：空间感知模块原型代码、基础评测脚本、阶段技术文档。

验收标准：在单一退化类型（去噪/去模糊/去雾）的公开基准上，区域级退化识别准确率优于全图级基线；分割置信度门控机制有效降低边界错误率。

### **第二阶段（第 5-8 月）：空间感知模块优化与评测**

阶段目标：完善空间感知模块，引入上下文聚合与 RobustSAM 回退机制，建立完整评测体系。

里程碑：完成上下文聚合机制的图注意力实现；在混合退化场景下验证空间退化图的质量；建立多基准联合评测协议。

可交付件：优化的空间感知模块、区域级退化识别评测基准、第一篇论文初稿（方法部分）。

验收标准：在 SIDD、GoPro、NH-HAZE 等多基准上，退化类型识别  $mAP \geq 0.85$ ；空间上下文聚合显著降低孤立区域误判率（相对降幅  $\geq 15\%$ ）。

### **第三阶段（第 9-12 月）：层级多智能体系统基础构建**

阶段目标：完成层级多智能体协同架构搭建，实现全局调度器与区域专家的基础功能。

里程碑：实现基于退化先验的全局优先级调度；完成区域专家的工具链选择与参数优化基础版本；建立质量反馈闭环。

可交付件：层级智能体系统原型、混合退化合成数据集、发明专利交底书（技术方案）。

验收标准：在合成混合退化数据上，系统能够正确识别不同区域的退化类型并分配处理优先级；工具链选择准确率达到预设阈值（ $\geq 80\%$ ）。

### **第四阶段（第 13-16 月）：层级多智能体系统优化与验证**

阶段目标：完善层级协同机制，引入一致性约束与早停机制，开展全面对比验证。

里程碑：完成全局一致性约束的融合优化；实现早停与参数自适应的鲁棒版本；与 AirNet、PromptIR、AgenticIR 等方法完成系统性对比。

可交付件：完善的层级智能体系统、对比实验报告、第一篇论文投稿、专利申请文件。

验收标准：在混合退化任务上相较 AirNet、PromptIR 等静态管线，PSNR 提升  $\geq 0.5dB$  或推理效率提升  $\geq 20\%$ ；与 AgenticIR、MAIR 等智能体方法相比在特定场景（空间非均匀退化）具有竞争优势。

### **第五阶段（第 17-20 月）：轨迹蒸馏与轻量网络构建**

阶段目标：构建多智能体决策轨迹数据集，完成轻量路由网络的基础训练。

里程碑：构建规模不少于 10000 条的高质量决策轨迹数据集；完成轻量路由网络的多任务联合蒸馏；实现基础版本的端到端推理。

可交付件：轨迹数据集、轻量路由网络基础模型、蒸馏训练框架、第二篇论文初稿。

验收标准：轻量路由网络在主要测试集上的 PSNR、SSIM 达到多智能体系统的 90% 以

上；端到端推理时延较原系统降低 $\geq 40\%$ 。

#### 第六阶段（第 21-24 月）：轻量部署优化与系统集成

阶段目标：完成困难样本再训练与部署优化，实现系统集成与边缘设备验证。

里程碑：完成困难样本聚焦训练，缩小与多智能体系统的性能差距；在边缘设备（Jetson Nano 或同等级平台）完成部署测试；形成完整可复现的系统与文档。

可交付件：优化后的轻量路由网络、边缘部署测试报告、第二篇论文投稿、系统开源代码与完整实验配置。

验收标准：轻量路由网络与多智能体系统的性能差距控制在可接受范围（PSNR 损失 $\leq 5\%$ 、SSIM 损失 $\leq 3\%$ ）；边缘设备实时推理达到 $\geq 10$  FPS @ 720p；系统具备完整的感知—规划—执行—蒸馏闭环能力。

时间映射关系：研究内容 1（空间感知）对应第一至二阶段，研究内容 2（层级多智能体）对应第三至四阶段，研究内容 3（轨迹蒸馏）对应第五至六阶段。阶段间依赖：第一、二阶段输出空间退化图  $\mathcal{M}$  作为第三、四阶段输入；第四阶段运行产生轨迹数据作为第五、六阶段输入；最终系统整合全部组件形成完整解决方案。

### 3.3. 可行性分析

#### （1）理论可行性

多智能体协同与大模型规划机制已在复杂任务求解中证明有效<sup>[26-29]</sup>；空间感知分析所依赖的 SAM<sup>[30]</sup>与 DepictQA<sup>[33]</sup>具备成熟的预训练基础；轨迹蒸馏作为知识压缩手段已在多任务学习中展现稳定性。上述理论基础与本项目“感知—规划—蒸馏”技术路线一致。

#### （2）技术可行性

在团队基础方面，已形成图像恢复与连续比例因子建模研究积累<sup>[36-38]</sup>；在平台条件方面，具备预训练模型接入、训练评测与可视化分析环境；在数据条件方面，具备构建合成退化与真实测试样本的能力。针对潜在瓶颈（算力开销、数据噪声、模型稳定性），已预设轻量路由替代、数据清洗与困难样本重训等缓释策略。

综上，本项目在理论与工程两个层面均具备实施条件，可支撑既定目标按计划推进。

### 4. 创新之处

**理论创新：**提出“空间退化状态驱动的恢复决策”框架，将图像恢复问题从全图单步映射扩展为区域状态感知下的多步决策问题。

传统图像恢复方法将整幅图像视为单一退化实体，采用统一的恢复函数  $\hat{\mathbf{x}} = f(\mathbf{y})$  进行端到端映射。这一范式忽视了真实场景中退化的空间异质性——同一张图像的不同区域可能同时受到类型迥异、程度各异的退化影响。本项目提出空间退化图  $\mathcal{M} = \{(R_i, \mathbf{d}_i, \mathbf{s}_i)\}$  作为决策先验，将恢复问题重新定义为状态条件策略函数  $\hat{\mathbf{x}}_{R_i} = \pi(\mathbf{y}_{R_i} | \mathcal{M})$ ，使恢复策略能够根据区域退化状态动态调整。相对已有方法的增量边界在于：由“统一恢复函数”转向“状态条件策略函数”，为复杂退化场景下的精细化恢复提供理论支撑。

**方法创新：**提出全局调度器与区域专家协同的层级多智能体机制，实现工具链选择与连续参数优化一体化。

现有基于智能体的恢复方法（如 AgenticIR、MAIR）虽引入动态决策能力，但均采用整图级统一规划，未能利用空间分布信息。本项目提出“全局调度器+区域专家”的两层协同架构：全局层依据退化先验（场景退化→成像退化→压缩退化的逆序处理原则）分配处理优先级；区域层执行工具链选择与连续参数优化，并通过质量反馈迭代更新。该机制突破了离散工具选择与连续参数调节的分离局限，实现了“离散路由+连续调参”的联合优化。相对已有方法的增量边界在于：从全图统一决策扩展到区域级差异化决策，从单一任务选择扩展到工具链序列规划与参数自适应。

**技术创新：**提出基于轨迹蒸馏的轻量路由部署方案，将高质量多步推理能力压缩为可实时执行的轻量模型。

基于 LLM/VLM 的智能体系统虽具备强大的规划能力，但在线推理成本高、部署门槛高，难以满足边缘实时应用需求。本项目提出离线轨迹收集与蒸馏训练相结合的技术路线：首先利用多智能体系统在合成数据上离线运行，收集“输入图像→最优决策”的轨迹数据；然后训练轻量级路由网络直接预测区域分割、退化类型、工具调度与参数配置，实现端到端推理。部署时无需依赖大模型在线调用，仅需轻量网络与恢复工具库即可完成实时恢复。相对已有方法的增量边界在于：从“依赖大模型在线推理”转向“离线蒸馏后边缘高效推理”，在保持恢复质量的同时显著降低部署时延与算力开销。

## 5. 预期成果

本项目将构建多智能体协同的空间感知图像恢复系统，在理论与方法上形成“空间退化图表示—层级协同规划—轨迹蒸馏部署”的完整技术链。

预期成果与目标对应关系如下：

(1) 对应目标 1（感知）：形成空间退化分析算法、评测脚本与实验数据资产，完成阶段论文投稿。

(2) 对应目标 2（决策）：形成层级多智能体恢复原型系统与调度策略库，申请发明专利。

(3) 对应目标 3（部署）：形成轻量路由部署模型与端侧测试报告，完成系统化验证。

总体成果计划为：在重要期刊或会议发表论文 2–4 篇，申请发明专利 1 项，培养硕士研究生 1–2 名。

## 6. 参考文献

- [1] LIANG J, et al. SwinIR: Image restoration using swin transformer[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. 2021: 1833-1844.
- [2] ZAMIR S W, et al. Restormer: Efficient transformer for high-resolution image restoration[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 5728-5739.
- [3] LI B, et al. All-in-one image restoration for unknown corruption[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 17452-17462.

- [4] POTLAPALLI V, et al. PromptIR: Prompting for all-in-one image restoration[C]//Advances in Neural Information Processing Systems. 2023.
- [5] CONDE M V, GEIGLE G, TIMOFTE R. High-quality image restoration following human instructions[C]//Proceedings of the European Conference on Computer Vision. 2024: 453-470.
- [6] WANG Z, et al. Uformer: A General U-Shaped Transformer for Image Restoration[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022. DOI: 10.1109/cvpr52688.2022.01716.
- [7] KONG L, et al. Efficient Frequency Domain-based Transformers for High-Quality Image Deblurring[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023. DOI: 10.1109/cvpr52729.2023.00570.
- [8] SHI Y, et al. VmambaIR: Visual State Space Model for Image Restoration[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2025. DOI: 10.1109/tcsvt.2025.3530090.
- [9] WU G, et al. DSwinIR: Rethinking Window-based Attention for Image Restoration[J]. arXiv preprint arXiv:2504.04869, 2025. DOI: 10.48550/arxiv.2504.04869.
- [10] ZHANG D, ZHOU F. Self-Supervised Image Denoising for Real-World Images With Context-Aware Transformer[J]. IEEE Access, 2023. DOI: 10.1109/access.2023.3243829.
- [11] CHEN X, et al. Hybrid CNN-Transformer Feature Fusion for Single Image Deraining[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2023. DOI: 10.1609/aaai.v37i1.25111.
- [12] HU B, et al. Collaborative Semantic Contrastive for All-in-one Image Restoration[J]. Engineering Applications of Artificial Intelligence, 2025. DOI: 10.1016/j.engappai.2025.110017.
- [13] MA J, et al. ProRes: Exploring Degradation-aware Visual Prompt for Universal Image Restoration[J]. arXiv preprint arXiv:2306.13653, 2023. DOI: 10.48550/arxiv.2306.13653.
- [14] WU G, et al. Learning Dynamic Prompts for All-in-One Image Restoration[J]. IEEE Transactions on Image Processing, 2025. DOI: 10.1109/tip.2025.3567205.
- [15] WU Z, et al. FrePrompter: Frequency Self-Prompt for All-in-One Image Restoration[J]. Pattern Recognition, 2024. DOI: 10.1016/j.patcog.2024.111223.
- [16] WU G, et al. Beyond Degradation Redundancy: Contrastive Prompt Learning for All-in-One Image Restoration[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2025. DOI: 10.1109/tpami.2025.3642852.
- [17] TANG A H, WU Y, ZHANG Y. RamIR: Reasoning and Action Prompting with Mamba for All-in-One Image Restoration[J]. Applied Intelligence, 2025. DOI: 10.1007/s10489-024-06226-y.

- [18] ZAMFIR E, et al. Complexity Experts are Task-Discriminative Learners for Any Image Restoration[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2025. DOI: 10.1109/cvpr52734.2025.01190.
- [19] WU G, et al. Harmony in Diversity: Improving All-in-One Image Restoration via Multi-Task Collaboration[C]//Proceedings of the ACM International Conference on Multimedia. 2024. DOI: 10.1145/3664647.3680762.
- [20] DUDHANE A, et al. Dynamic Pre-training: Towards Efficient and Scalable All-in-One Image Restoration[J]. arXiv preprint arXiv:2404.02154, 2024. DOI: 10.48550/arxiv.2404.02154.
- [21] JIANG J, et al. A Survey on All-in-One Image Restoration: Taxonomy, Evaluation and Future Trends[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2025. DOI: 10.1109/tpami.2025.3598132.
- [22] SAHARIA C, et al. Palette: Image-to-Image Diffusion Models[C]//ACM SIGGRAPH 2022 Conference Proceedings. 2022. DOI: 10.1145/3528233.3530757.
- [23] LIU J, et al. Residual Denoising Diffusion Models[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024. DOI: 10.1109/cvpr52733.2024.00268.
- [24] LIN X, et al. DiffBIR: Toward Blind Image Restoration with Generative Diffusion Prior[C] //Proceedings of the European Conference on Computer Vision. 2024. DOI: 10.1007/978-3-031-73202-7\_25.
- [25] YANG T, et al. Pixel-Aware Stable Diffusion for Realistic Image Super-Resolution and Personalized Stylization[C]//Proceedings of the European Conference on Computer Vision. 2024. DOI: 10.1007/978-3-031-73247-8\_5.
- [26] ZHU K, et al. AgenticIR: An intelligent agentic system for complex image restoration problems[C]//Proceedings of the International Conference on Learning Representations. 2025.
- [27] JIANG X, et al. Multi-agent image restoration[J]. arXiv preprint arXiv:2503.09403, 2025.
- [28] CHEN H, et al. RestoreAgent: Autonomous image restoration agent via multimodal large language models[J]. arXiv preprint arXiv:2407.18035, 2024.
- [29] ZHOU Y, et al. Q-Agent: Quality-driven chain-of-thought image restoration agent through robust multimodal large language model[J]. arXiv preprint arXiv:2504.07148, 2025.
- [30] KIRILLOV A, et al. Segment anything[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023: 4015-4026.
- [31] CHEN W T, et al. RobustSAM: Segment Anything Robustly on Degraded Images[J]. arXiv preprint arXiv:2406.09627, 2024.
- [32] REN T, LIU S, ZENG A, et al. Grounded SAM: Assembling Open-World Models for Diverse Visual Tasks[J]. arXiv preprint arXiv:2401.14159, 2024.

- [33] YOU Z, et al. DepictQA-Wild: Enhancing descriptive image quality assessment with a large-scale multi-modal dataset[J]. IEEE Transactions on Image Processing, 2025.
- [34] WU H, ZHANG Z, ZHANG E, et al. Q-Instruct: Improving Low-Level Visual Abilities for Multi-Modality Foundation Models[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024. DOI: 10.1109/cvpr52733.2024.02408.
- [35] WU H, ZHANG Z, ZHANG E, et al. Q-Bench: A Benchmark for General-Purpose Foundation Models on Low-level Vision[J]. arXiv preprint arXiv:2309.14181, 2023.
- [36] WU H, NI N, ZHANG L. Learning dynamic scale awareness and global implicit functions for continuous-scale super-resolution of remote sensing images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 1-15.
- [37] 吴瀚霖, 李宛谕, 张立保. 跨尺度耦合的连续比例因子图像超分辨率[J]. 中国图象图形学报, 2022, 27(5): 1604-1615.
- [38] WU H, ZHANG L, MA J. Remote sensing image super-resolution via saliency-guided feed-back GANs[J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 60: 1-16.

## 四、前期研究基础

本表参照以下提纲撰写，**1000**字以内。

1. 课题组近五年来已有的相关研究成果（负责人和参加者分开填写）。
2. 课题负责人近五年来曾完成哪些重要研究课题，科研成果的社会评价（引用、转载、获奖及被采纳情况）。
3. 为本课题研究已作的前期准备工作（已收集的数据，进行的调查研究等）。
4. 本课题与已立项项目、博士论文（博士后出站报告）的联系与区别。

### 1. 课题负责人近五年来已有的相关研究成果

课题负责人近五年持续围绕计算机视觉与遥感能力处理开展研究，研究方向主要包括：图像恢复、连续比例因子超分辨率、视觉语言多模态学习、基于扩散模型的生成式方法、遥感变化分析等。

根据个人主页公开成果，近年已形成较为系统的研究积累。在论文方面，相关成果发表于 IEEE Transactions on Geoscience and Remote Sensing、IEEE Geoscience and Remote Sensing Letters、Remote Sensing、ICASSP、IGARSS 等期刊与会议，研究主题覆盖：盲超分辨率重建、连续比例因子超分辨率、跨模态检索、变化描述与变化问答等。

代表性成果可归纳为三类，具体如下。

**一、连续比例因子超分辨率建模问题：**固定比例模型训练/存储成本高，且对非整数倍率泛化不足。方法：构建连续尺度感知建模框架，引入跨尺度特征耦合与动态上采样机制。贡献：实现单模型覆盖连续倍率重建并兼顾质量与效率。结果：形成高水平论文成果并沉淀可复用模型组件。与本项目关联：为“区域工具参数连续优化”提供直接方法学基础。

**二、复杂退化过程下的生成式恢复问题：**真实场景退化分布复杂，传统单一退化假设鲁棒性不足。方法：结合生成式建模与退化过程表征，提升模型对复杂退化的适应能力。贡献：增强复杂场景中的恢复稳定性与视觉一致性。结果：形成复杂退化恢复方向的连续研究成果。与本项目关联：为“空间退化图驱动的策略规划”提供可迁移技术储备。

**三、遥感多模态理解与指令驱动分析问题：**仅依赖视觉特征难以支撑高层语义决策。方法：开展跨模态检索、变化描述与变化问答研究，探索指令驱动视觉分析流程。贡献：形成视觉—语言联合推理能力与任务编排经验。结果：建立多模态任务原型与评测流程。与本项目关联：为“多智能体协同推理与反馈”提供实现基础。

科研成果社会评价方面，相关工作已在领域主流期刊/会议发表并被持续关注，研究路线具有明确延续性。上述积累与本课题拟解决的“复杂退化场景下的智能恢复决策”问题具有直接衔接关系。

### 2. 课题负责人近五年的项目经历

近五年主持/参与项目清单如下（来源于个人主页公开信息）：

(1) 2025.01–2027.12，国家自然科学基金青年科学基金项目：开放场景下认知启发的遥

感影像超分辨率重建方法研究，主持。

(2) 2024.06–2027.06，北京外国语大学学术青年创新团队项目：生成式大语言模型的核心价值观对齐研究，参与。

(3) 2024.06–2027.06，北京外国语大学项目：基于状态空间扩散模型的遥感影像变化描述方法研究，参与。

(4) 2023.01–2025.12，国家自然科学基金面上项目：演进学习框架下协同感知显著性引导的弱标注遥感影像语义分割方法研究，参与。

(5) 2022.09–2025.09，北京外国语大学项目：自适应学习框架下显著性引导的遥感影像超分辨率重建方法研究，主持。

上述项目对本课题形成的能力支撑包括：1) **数据与实验支撑**：沉淀了多类遥感图像处理任务的数据处理与评测流程；2) **方法与实现支撑**：积累了超分辨率、生成式建模与多模态分析的实现经验；3) **协同与组织支撑**：形成了跨任务协同研究与阶段成果产出的组织基础。

### 3. 前期准备工作

本节从“已完成事项 + 产出物 + 可验证状态”三个维度说明前期准备，并与本课题研究内容 1/2/3 建立对应关系。

#### 3.1. 理论与技术储备

(1) 已形成图像恢复与超分辨率方法学储备（对应研究内容 1/2）。产出物：模型设计与评测方案。可验证状态：已有多篇相关成果与可复现流程。

(2) 已形成多模态遥感理解储备（对应研究内容 2）。产出物：跨模态任务原型与分析脚本。可验证状态：已完成多类任务实验闭环。

(3) 已形成生成式建模储备（对应研究内容 1/3）。产出物：生成式建模实验管线。可验证状态：已具备复杂退化建模与对比验证基础。

#### 3.2. 数据与实验环境构建

(1) 已建立遥感图像恢复与多模态任务实验环境。产出物：统一训练/验证/消融流程。可验证状态：可稳定复现主流方法。

(2) 已完成多类任务基线复现与评测流程搭建。产出物：基线模型与对比评测脚本。可验证状态：可生成标准化评测结果，为后续方案提供可靠对照。

(3) 已具备公开预训练模型与开源代码接入能力。产出物：可复用模块接入模板。可验证状态：可快速开展新模块原型验证。

(4) 阶段完成度：目前已完成基础环境与基线体系建设，正在推进面向多智能体决策的专项实验管线建设。

#### 3.3. 关键技术预研

(1) **预研问题**：连续比例因子恢复中的效率与质量平衡。技术路径：轻量化模型与动态参数策略。初步结果：已验证该路径可行。后续计划：纳入区域级参数调优机制。

(2) 预研问题：复杂场景下指令驱动决策稳定性。技术路径：视觉语言联合分析与反馈闭环。初步结果：已验证指令驱动分析有效。后续计划：扩展至多智能体交互流程。

(3) 预研问题：复杂退化建模与部署效率矛盾。技术路径：生成式建模 + 轨迹蒸馏。初步结果：已验证潜空间建模鲁棒性。后续计划：推进轻量模型蒸馏与边缘部署测试。

(4) 风险识别与改进策略：针对预研中可能出现的训练不稳定、跨场景泛化不足等问题，已制定困难样本增强、分阶段训练与多指标联合评估策略。

#### 4. 本课题与已立项项目的联系与区别

已立项项目主要包括：国家自然科学基金青年项目“开放场景下认知启发的遥感影像超分辨率重建方法研究”、新教师科研启动项目“自适应学习框架下显著性引导的遥感影像超分辨率重建方法研究”。

(1) 联系：本课题与既有项目均聚焦遥感图像增强与理解，均强调利用前沿人工智能方法提升模型在复杂场景下的性能与泛化能力。

(2) 区别：针对国家自然科学基金青年项目：核心：强调“认知启发”与开放场景推理机制。区别：本课题不以认知机制建模为主线，而以“空间退化状态驱动的层级多智能体决策”作为核心增量。针对新教师科研启动项目：核心：强调“显著性引导”的局部增强策略。区别：本课题从局部增强扩展为“区域感知—工具路由—参数优化—蒸馏部署”的全流程系统方案。

本课题充分继承既有项目在超分辨率、生成式建模与多模态理解方面的积累，在此基础上推进跨模块协同与轻量化部署，具有明确连续性。本课题与既有项目在研究对象、技术路线、评价指标和预期输出上均有清晰区分：从单模型性能提升转向系统级决策能力提升，重点考察复杂退化场景下的综合质量—效率指标。