



编号__

北京外国语大学基本科研业务费 学生科研创新项目申请书

学 科 名 称: 计算机科学技术

项 目 类 别: 学生科研创新项目一般项目

项 目 名 称: 基于认知启发和多模态扩散模型的语义引导超分辨率重建

项目负责人: 莫江炜

所 在 单 位: 信息科学技术学院

联 系 手 机: 181 0130 1856

联 系 邮 箱: 2336368031@qq.com

导 师 姓 名:

导师联系手机:

北京外国语大学科研处

2021 年 6 月制



一、信息表

申请人信息	姓名	莫江炜	性别	男
	出生年月	2024-07	学号	202220119004
	学位	其他	职称	其他
	行政职务	无	所在单位	信息科学技术学院
	所在年级			
	联系手机	181 0130 1856	联系邮箱	2336368031@qq.com
	导师姓名		导师联系手机	

项目基本信息	项目名称	基于认知启发和多模态扩散模型的语义引导超分辨率重建		
	项目类别	学生科研创新项目一般项目	申请金额	0.5 万元
	学科名称	计算机科学技术	学科代码	520
	科技活动类型	应用研究	支持对象	本科生
	起止年月	2024-05 至 2025-05		
	结项要求	根据当年相应项目申报公告结项要求完成		



项目 成员 信息	姓名	成员类型	出生年份	职称	所 在 单 位	签名
	莫江炜	本校教师	2024-07	其他	信息科学技术学院	

(草稿) 请于提交后重新下载PDF的正式稿

二、项目摘要

中文摘要	<p>中文摘要：（限 500 字以内）：</p> <p>本课题提出了一种基于认知启发的文本引导超分辨率重建框架。该框架利用自然语言描述生成全局认知编码, 引导超分辨率重建过程。融入基于认知编码的高效图像检索技术, 获取与重建目标语义高度一致的参考图像, 有效解决了超分辨率重建中的不适宜性问题。该研究采用了具有卓越生成性能的扩散模型, 结合多模态联合注意力机制, 为模型提供准确的全局语义信息和局部细节指导, 实现了影像的高质量、可控的盲超分辨率重建。研究还提出了潜扩散模型结构设计和可控的重建结果生成方法, 提高了模型的推理效率并实现了真实感和保真度之间的权衡。</p>
------	---

三、项目设计论证

1. 选题依据

1.1 研究意义

随着多媒体技术的快速发展，高分辨率图像/视频对于多媒体内容的展示和交互变得越来越重要。然而，受限于成像设备的性能、存储和传输带宽等因素，我们获取的多媒体数据通常具有较低的分辨率。如何从低分辨率图像/视频中重建高分辨率版本，是多媒体处理领域的一个关键问题。超分辨率重建技术旨在利用信号处理和机器学习等方法，从一个或多个低分辨率输入中恢复出高分辨率图像/视频，对提高多媒体质量和用户体验具有重要意义。

文本引导超分辨率重建 (Text-Guided Super-Resolution) 是近年来兴起的一种新颖方法。传统的超分辨率方法主要依赖图像本身的像素信息，而文本引导超分辨率重建，它将自然语言描述与传统的基于像素的超分辨率重建相结合，以期借助文本中蕴含的语义信息来辅助图像重建过程，生成更加细致、逼真的高分辨率输出。该技术最早可追溯至 2018 年左右的一些探索性工作，如 Text-SR 等将文本编码为条件向量，与图像特征融合的尝试。2021 年，transformer 模型在视觉任务中取得突破，First-Transformer 等工作将其引入到文本引导超分领域。2022 年，视觉语言预训练模型 (如 CLIP) 的兴起进一步推动了该领域的发展，模型如 GLEAN 通过迁移预训练权重显著提升了重建质量。

最新研究动态包括:注意力机制优化 (MCAN 等)、新的预训练模型 (Unified VL 等)、大规模数据集构建 (Image Captions 等)、应用拓展 (医疗影像等领域) 以及理论分析 (语义偏差等) 等多个方面。

这种方法开辟了超分辨率重建的新思路。通过融合视觉和语义信息，可以突破传统方法的局限，为提高重建质量提供新的可能性。拓展了多模态学习在计算机视觉中的应用前景。文本和图像是两种重要的信息载体，将它们有机结合能够发挥多模态学习的优势，促进不同模态之间的相互增强。为多媒体内容理解和生成提供了新的技术支持。高质量的图像/视频重建能力是实现智能多媒体系统的重要基础，对于多媒体内容分析、检索、编辑等应用都有重要意义。可以促进视觉与语言的共融共通。通过学习图像-文本的对应关系，有助于计算机更好地理解视觉和语义概念，为发展人工智能奠定基础。具有广泛的应用前景。无论是提升图

像/视频质量、修复损坏的多媒体数据，还是生成高分辨率的虚拟现实内容，文本引导超分辨率重建技术都可以发挥重要作用。

当前，文本引导超分辨率重建仍是一个新兴的研究方向，在模型设计、训练策略、评估指标等多个方面仍有诸多待解决的问题和挑战。

综上所述，我们将本课题的研究意义概括为如下三点：

1. 本课题的顺利实施将为文本引导超分辨率重建研究开辟新的研究思路，探索如何利用文本信息来指导超分辨率重建过程，生成与文本描述相符的高质量图像细节。这对于深度学习理论在跨模态信息融合方面的发展具有重要意义。

2. 本课题是自然语言处理与计算机视觉交叉领域的一次重要尝试，旨在利用文本信息来提升超分辨率重建的质量和多样性。项目的顺利实施有助于拓展低质量图像的应用场景，为图像增强、图像编辑等任务提供新的解决方案。

3. 本课题的成果不仅能为文本引导图像生成、扩散模型等相关研究提供重要的理论与技术支持，而且可以在医学影像、视频监控、远程教育等领域得到广泛应用，为提高低质量图像的应用价值提供必要的技术保障。

1.2 国内外研究现状分析及存在问题

1) 传统超分辨率重建技术的研究现状

早期基于频域分析的方法旨在恢复图像的高频分量，如 Tsai 等[1]利用离散傅里叶变换，Kim 等[2]引入加权递归最小二乘法解决频谱混叠问题。局部频谱分析[3]和离散余弦变换[4]等也被应用于多图像帧的超分重建。然而这些方法建立在简化的降质模型上，不能覆盖复杂场景。

因此，一些基于空域优化的方法开始受到关注，如 Brown 等[5]的不均匀插值法，Stark 等[6]提出的 POCS 算法注入先验约束，Patti 等[7]进一步考虑噪声和模糊等影响，Schultz 等[8]则基于 MAP 估计。Elad 等[9]将 POCS 和 MAP 有机结合，提升重建质量。

上述方法依赖有限的先验知识，高频细节难以恢复，促使了基于学习的策略兴起。Freeman 等[10]利用 MRF 模型学习高低分辨率映射，Chang 等[11]融入了流形学习思想。压缩感知理

论[12]的提出也推动了稀疏编码在超分重建中的应用[13, 14]。

2) 基于深度学习的超分辨率重建研究现状

随着深度学习技术的飞速发展，将卷积神经网络等深度模型应用于超分辨率重建任务成为研究的热点和主流方向。Dong 等人[15]提出了开创性的 SRCNN 模型，将超分重建过程分解为串联的卷积网络模块，实现了端到端的低分辨率到高分辨率图像的映射学习，可视为深度学习方法在该领域的最早尝试。之后，Kim 等人[16]提出了里程碑式的 VDSR 模型，通过加深网络层数至 20 层以提高模型容量，并引入基于跨层连接的残差学习单元来缓解训练困难，取得了比 SRCNN 更出色的重建性能。为进一步扩大感受野同时避免过度参数增加，Kim 等人[17]又提出了 DRCN，采用递归卷积机制在水平和垂直两个方向上分别扩展感受野，显著提升了重建效果。

除了优化网络结构，一些工作还着眼于设计新颖的损失函数以获得更加真实的重建结果。Johnson 等人[18]引入了基于预训练 VGGNet 提取的多尺度语义特征构建的视觉感知误差，作为优化目标函数。Ledig 等人[19]则将生成对抗网络（GAN）的思想引入超分重建领域，通过一个生成网络与一个判别网络的对抗博弈训练，使得生成器学习到生成逼真图像的能力，从而避免了传统 MSE 损失函数导致的过度平滑现象。该工作极大地推动了基于深度学习的超分重建方法向生成高质量图像的方向发展。

最近几年，基于扩散模型的超分辨率重建方法也引起了广泛关注。扩散模型通过学习一个渐进的扩散过程来生成高质量图像，相比于 GAN 等生成模型具有更好的稳定性和生成效果。一些工作将扩散模型应用于超分辨率任务，在保持高 fidelity 的同时也能生成更加细腻逼真的高分辨率图像，进一步推动了基于深度学习的超分重建技术发展。

3) 文本引导超分辨率重建研究现状

将自然语言文本引入超分辨率重建任务是近年来兴起的一种新颖方法。早期的一些探索性工作尝试将文本编码为条件向量，与图像特征融合后输入到重建网络中。随着 Transformer 模型在 2021 年取得突破性进展，一些研究开始将该模型的编码器-解码器架构应用于文本引导超分重建任务，通过注意力机制捕获文本-图像之间的语义相关性，并取得了比传统方法更优的重建效果。2022 年，大规模预训练的视觉语言模型（VL pre-trained models）在多模态领域掀起热潮，一些工作将模型如 CLIP、ALIGN 等迁移到文本引导超分重建任务中，充分利

用了预训练权重蕴含的丰富视觉语义知识，使重建质量得到了进一步的大幅提升。近年来，随着扩散模型在图像生成领域的广泛应用，研究者们开始探索将文本引导与扩散模型相结合，用于超分辨率图像重建任务。Rombach 等人[20]提出了一种基于潜在扩散模型的高分辨率图像合成方法。该方法通过在潜在空间中进行扩散过程，并结合文本条件信息，实现了高质量、高分辨率图像的生成。这一工作展示了扩散模型在图像合成领域的强大能力，为文本引导超分辨率重建提供了新的思路。Wang 等人[21]提出了一种文本引导的可探索图像超分辨率方法。该方法利用文本描述来引导超分辨率重建过程，生成与文本相关的多样化图像细节。通过可探索的潜在空间，用户可以根据不同的文本描述，生成具有相应细节变化的高分辨率图像。这一工作展示了文本引导在控制超分辨率重建多样性方面的优势。Liu 等人[22]提出了一种名为 Diff-Plugin 的方法，用于扩散模型的低层视觉任务。该方法通过在扩散过程中引入额外的条件信息，如文本描述，来指导图像细节的生成。实验结果表明，Diff-Plugin 能够有效地利用文本信息，生成逼真的图像细节，提升超分辨率重建的质量。Li 等人[23]重新思考了超分辨率重建问题，将其视为一种文本引导的细节生成任务。该方法利用文本描述来控制超分辨率重建过程，通过生成与文本相关的细节信息，实现高质量、符合语义的图像重建。这一工作进一步探索了文本引导在超分辨率重建中的作用，为该领域的发展提供了新的视角。

4) 文本引导超分辨率重建存在的主要问题

虽然将文本引导与超分辨率重建相结合已经取得了一定的进展，但是，目前仍有诸多关键问题没有得到很好地解决。

目前的文本引导超分辨率重建方法大多需要大量的文本-图像对作为训练数据。然而，获取高质量的文本-图像对数据集非常困难，尤其是在特定领域（如医学影像）中，需要专业知识来创建准确的文本描述，这通常耗费大量的人力成本。

现有的文本引导超分辨率重建方法没有充分利用文本信息来指导图像细节的生成。文本描述中包含了丰富的语义信息，如物体的属性、场景的上下文等。然而，大多数方法仅简单地将文本特征与图像特征拼接，未能深入挖掘文本信息与图像内容之间的关联，导致生成的图像细节与文本描述不够一致。

目前常用的图像质量评估指标（如 PSNR，SSIM）无法全面衡量文本引导超分辨率重建的性能。这些指标主要关注图像的低级视觉质量，如像素级别的相似性，而忽略了高级语义

信息。因此，即使生成的图像在这些指标上取得了不错的结果，也可能与文本描述存在语义上的差异。

2. 研究内容

2.1 主要目标

1) 设计基于扩散模型和文本引导的盲超分辨率重建框架。基于认知编码的参考图像检索来指导超分辨率重建过程，生成与文本描述相符的高质量图像细节，提升重建图像的视觉质量和语义一致性。

2) 探索文本信息与图像内容之间的最优融合策略，设计高效的跨模态特征交互机制，充分利用文本信息来增强图像细节的生成，同时保留图像的原有内容特征。

3) 构建全面的文本引导超分辨率重建性能评估体系，综合考虑图像质量、语义一致性和多样性等因素，设计合理的评估指标，为该领域的发展提供标准化的评测基准。

2.2 总体框架

1) 构建认知引导的超分辨率重建框架，利用自然语言描述生成全局认知编码作为重建过程的指导信息。通过基于认知编码的高效图像检索机制，获取与目标语义高度一致的参考影像，有效缓解超分辨率重建任务中存在的不良适性问题。

2) 针对超分辨率重建任务，引入具备强大生成能力的扩散模型。设计多模态特征融合机制，将全局语义编码和局部细节信息融入扩散模型，为生成过程提供精准的条件约束。该方法能够最终实现对遥感影像的高质量、可控盲超分辨率重建，生成的高分辨率影像不仅保留了丰富的细节信息，而且与给定的自然语言描述高度一致。

2.3 重点难点

1) 如何设计高效的跨模态融合机制，有效融合文本信息和图像内容，使文本信息能够很好地指导图像细节生成，同时又不破坏图像原有的内容特征。

2) 如何通过文本描述有效检索出与目标图像语义相关的参考图像，为超分辨率重建提供有价值的先验知识。

3) 由于超分辨率重建需要同时考虑图像质量、语义一致性和多样性等多个维度, 如何设计合理的综合评价指标。现有的主观和客观评价指标可能无法完全衡量文本引导超分的性能, 需要创新性地提出新的评估方法。

4) 由于场景的多样性, 如何使模型能够泛化到各种不同的图像场景和文本描述。

3. 思路方法

3.1 研究方案

1) 建立以认知启发为核心的超分辨率重建框架, 利用自然语言描述作为启发, 生成全局认知编码, 以此引导图像的重建过程。同时, 融入基于认知编码的高效图像检索技术, 以获取与重建目标在语义上高度一致的参考图像, 从而有效解决超分辨率重建中存在的不适定性问题。

2) 采纳具有卓越生成性能的扩散模型, 并结合多模态联合注意力机制, 为模型提供准确的全局语义信息和局部细节指导, 实现影像的高质量、可控的盲超分辨率重建。通过这种方式, 模型能够更加精准地理解和重建图像内容, 提升重建图像的细节丰富度和语义准确性。

3) 潜扩散模型结构设计。扩散模型已经在自然图像合成任务上展现了卓越的性能。其核心思想是学习一个马尔可夫链的逆过程, 通过对高斯分布随机变量进行逐步“去噪”来学习数据分布。超分辨率重建任务的输出图像通常尺寸较大。若直接在图像空间上执行迭代去噪, 则在推理过程需要消耗较多的计算资源, 而且输出速度较慢。而潜扩散模型 (Latent Diffusion Model, LDM) 引入一个变分自编码器 (Variational AutoEncoder, VAE) 给予了本研究启发, VAE 可以将原始高分辨率图像通过编码器映射为潜变量, 并可以通过解码器将潜变量映射回原始空间。具体地, 本研究仅利用高分辨率影像数据集, 共同训练编码器和解码器。优化目标包括 VAE 的重构损失和 KL 散度损失, 此外还使用对抗损失和感知损失来提升自编码器的性能。随后, 扩散过程和去噪过程可以在低维潜空间中进行, 将显著提升模型的推理效率。

4) 多模态联合注意力机制。在超分辨率重建任务中, 保持重建结果与输入低分率影像的内容一致性至关重要。此外, 还须有效利用参考影像提供的丰富细节信息提升重建结果的真实感。为此, 本研究提出多模态联合注意力机制, 有助于建立低分影像与参考影像的远距离

连接，还可以通过交叉注意力机制注入认知编码提供的全局语义信息。

5) 可控的重建结果生成。由于超分辨率重建任务的不适定性，结果的真实感与空间失真程度是一对难以调和的矛盾。因此，本研究提出一种可控的扩散模型采样方法，使得用户可以根据实际需要，在真实感和保真度之间进行权衡。

3.2 可行性

1) 本课题在理论上是可行的

理论上，认知启发的概念已经在多个领域得到验证，特别是在图像处理和计算机视觉领域。通过自然语言描述作为启发来生成全局认知编码，并以此引导图像的重建过程，这种方法借鉴了人类视觉系统处理信息的方式，有望提高超分辨率重建的质量。同时，基于认知编码的高效图像检索技术，能够在语义层面上寻找与重建目标高度一致的参考图像，这有助于解决超分辨率重建中的不适定性问题。

而且，扩散模型已经在图像合成等任务上展现出卓越的性能，其逆过程学习数据分布的特性，为超分辨率重建提供了新的思路。结合多模态联合注意力机制，可以进一步提升模型在全局语义信息和局部细节指导上的能力，从而实现高质量、可控的盲超分辨率重建。

2) 本课题在技术上是可行的

技术上，深度学习技术的发展为超分辨率重建提供了强大的工具，特别是在生成模型方面。扩散模型的引入，结合多模态联合注意力机制，能够有效地提取和利用图像中的语义信息，提高重建图像的质量。

潜扩散模型结构的设计，通过变分自编码器将高分辨率图像映射为潜变量，并在低维潜空间中进行扩散过程和去噪过程，这不仅可以提高模型的推理效率，还能减少计算资源的消耗。

多模态联合注意力机制的提出，有助于在重建过程中保持与输入低分辨率影像的内容一致性，并通过交叉注意力机制注入全局语义信息，进一步提升重建结果的真实感。

4. 创新之处

在图像超分辨率领域，传统的研究方法主要集中于图像本身的处理和重建，而对图像内

容与文本描述之间的关联研究相对较少。随着深度学习技术的发展，尤其是多模态学习的出现，将文本信息融入图像处理已成为提升图像质量和理解图像内容的新趋势。在这样的背景下，本研究提出了一系列创新点：

- 1) 本课题的模型框架是基于认知启发和扩散模型的盲超分辨率重建框架，利用自然语言描述作为启发，生成全局认知编码来引导图像重建过程。采用具有卓越生成性能的扩散模型，并融合多模态联合注意力机制，实现高质量、可控的盲超分辨率重建，提升了重建图像的细节丰富度和语义准确性。
- 2) 使用多模态联合注意力机制。通过建立低分辨率输入图像与参考图像的远距离连接，并注入全局语义信息，有效利用了多源信息提升重建结果的真实感。
- 3) 提出了一种可控的扩散模型采样方法，使得用户可以在重建结果的真实感和保真度之间进行权衡，增强了方法的灵活性和适用性。

5. 预期成果

5.1 成果形式

研究将形成一套完整的文本引导的图像超分辨率技术框架、算法模型。发表学术论文，介绍模型设计理念、实验方法、结果分析以及与现有技术的比较。改进模型，展示技术的实际应用效果。

5.2 使用去向

技术成果可应用于多个领域，包括但不限于医疗影像处理、卫星遥感图像分析、历史文献数字化修复、安防监控图像增强等。在医疗影像领域，能够提高疾病诊断的准确性和效率；在遥感领域，有助于提高对地观测的分辨率和精度；在文化保护领域，能够数字化修复和保存低分辨率的历史文献和艺术品。

5.3 预期社会效益

提高医疗诊断的准确性，辅助医生做出更快速、更准确的医疗决策，从而提高患者的治疗效果和生活质量。增强卫星遥感图像的解析能力，为环境监测、城市规划、灾害预警等提供更高质量的数据支持。促进历史文献和文化遗产的数字化保护工作，通过高清晰度的图像

重建，使公众能够更好地欣赏和学习这些宝贵的文化资源。增强安防监控系统的效能，提供更清晰的视频分析结果，有助于提高公共安全和预防犯罪。

(草稿) 请于提交后重新下载PDF的正式稿

6.参考文献

- [1] Tsai R Y, Huang T S. Multiframe image restoration and registration[J]. Advances in computer vision and Image Processing, 1984, 1(2): 317-339.
- [2] Kim S P, Bose N K, Valenzuela H M. Recursive reconstruction of high resolution image from noisy undersampled multiframes[J]. IEEE Transactions on Acoustics, Speech, and Signal Processing, 1990, 38(6): 1013-1027.
- [3] Su H, Zeng W, Chen Z, et al. Locally adaptive multi-frame super-resolution based on Gaussian process regression[J]. Pattern Recognition, 2015, 48(1): 141-156.
- [4] Rhee H, Kang M G. Discrete cosine transform based regularized high-resolution image reconstruction algorithm[J]. Optical Engineering, 2000, 39(11): 3112-3128.
- [5] Brown L G. A survey of image registration techniques[J]. ACM computing surveys (CSUR), 1992, 24(4): 325-376.
- [6] Stark H, Oskoui P. High-resolution image recovery from image-plane arrays, using convex projections[J]. JOSA A, 1989, 6(11): 1715-1726.
- [7] Patti A J, Sezan M I, Tekalp A M. Superresolution video reconstruction with arbitrary sampling lattices and nonzero aperture time[J]. IEEE Transactions on Image Processing, 1997, 6(8): 1064-1076.
- [8] Schultz R R, Stevenson R L. Extraction of high-resolution frames from video sequences[J]. IEEE transactions on image processing, 1996, 5(6): 996-1011.
- [9] Elad M, Feuer A. Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images[J]. IEEE transactions on image processing, 1997, 6(12): 1646-1658.
- [10] Freeman W T, Pasztor E C, Carmichael O T. Learning low-level vision[J]. International journal of computer vision, 2000, 40(1): 25-47.
- [11] Chang H, Yeung D Y, Xiong Y. Super-resolution through neighbor embedding[C]//Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004. IEEE, 2004, 1: I-I.
- [12] Donoho D L. Compressed sensing[J]. IEEE Transactions on information theory, 2006, 52(4): 1289-1306.
- [13] Yang J, Wright J, Huang T S, et al. Image super-resolution via sparse representation[J]. IEEE transactions on image processing, 2010, 19(11): 2861-2873.
- [14] Timofte R, De Smet V, Van Gool L. A+: Adjusted anchored neighborhood regression for fast super-resolution[A]. In: Computer Vision – ACCV 2014[C]. Springer, Cham, 2015: 111-126.
- [15] Dong C, Loy C C, He K, et al. Image super-resolution using deep convolutional networks[J]. IEEE transactions on pattern analysis and machine intelligence, 2015, 38(2): 295-307.
- [16] Kim J, Kwon Lee J, Mu Lee K. Accurate image super-resolution using very deep convolutional networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 1646-1654.

- [17] Kim J, Kwon Lee J, Mu Lee K. Deeply-recursive convolutional network for image super-resolution[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 1637-1645.
- [18] Johnson J, Alahi A, Fei-Fei L. Perceptual losses for real-time style transfer and super-resolution[C]//European conference on computer vision. Springer, Cham, 2016
- [19] Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., ... & Shi, W. (2017). Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4681-4690).
- [20] Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 10684-10695).
- [21] Huang, H., Zheng, Z., Zhou, J., Jiang, Y., & Loy, C. C. (2023). Text-guided Explorable Image Super-resolution. arXiv preprint arXiv:2303.13925.
- [22] Kavar, B., Vaksman, G., & Elad, M. (2022). Diff-Plugin: Revitalizing Details for Diffusion-based Low-level Tasks. arXiv preprint arXiv:2212.02777.
- [23] Hou, X., Shen, L., Sun, K., & Qiu, G. (2017). Deep feature consistent variational autoencoder. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops (pp. 2367-2376).
- [24] Saharia C, Chan W, Saxena S, et al. Photorealistic text-to-image diffusion models with deep language understanding[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 12427-12437.
- [25] Lugmayr A, Danelljan M, Van Gool L, et al. Repaint: Inpainting using denoising diffusion probabilistic models[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 11461-11471.

(草稿) 请于提交后重新下载PDF的正式稿

四、经费预算（万元）

科 目	金额 (万元)	占总经费比例	使用说明（必填）
图书资料费	0.2 万元	40.00%	学术相关书籍购买
差旅费/ 会议费/ 国际合作与 交流费	0.1 万元	20.00%	差旅费、参加学术会议，开展学术报告和学术沙龙
专家咨询费	0.0 万元	.00%	无
印刷出版费	0.0 万元	.00%	无
其他支出	0.2 万元	40.00%	必要的办公用品、研究所需的相关设备
总计	0.5		

五、项目保证与审核意见

1. 申请者承诺

我保证上述填报内容真实、准确；项目书稿不存在意识形态问题。如果获得资助，我将履行项目负责人职责，严格遵守学校的有关规定，切实保证研究工作时间，按计划认真开展研究工作，按时报送有关材料，按要求在 10 月 30 日之前执行年度预算经费的 100%。若填报失实或违反有关规定，本人将承担全部责任。

申请人（签字）

年 月 日

2. 申请者导师审核意见

请申请者导师签署明确意见。申请书所填写的内容是否属实；该课题负责人的政治业务素质是否适合承担本课题的研究工作；导师是否同意承担本项目的管理任务和信誉保证。

申请人导师（签字）签章：

年 月 日

3. 所在单位意见

申请书所填写的内容属实；申请人的政治素质与业务水平适合承担本项目的研究；主管单位能提供完成本项目所需时间和其他必要条件；本单位同意承担本项目的相关管理任务和信誉保证。

单位（公章） 年 月 日

4. 学校科研处审批意见（公章）：

单位（公章） 年 月 日