# 基于语义感知对抗性扩散压缩的图像超分辨率

**摘要：**真实场景图像超分辨率（Real-ISR）需同时满足高恢复质量与高效部署需求。现有基于扩散模型的方法要么计算复杂度高，要么生成语义不一致的纹理。为解决这些问题，本文提出语义感知对抗性扩散压缩（SA-ADC）框架，将语义感知判别机制融入对抗性扩散压缩过程。首先，通过模块移除与通道剪枝将单步扩散网络 OSEDiff 蒸馏为轻量化架构，在保留核心组件的同时降低复杂度；其次，引入语义感知判别器（SeD）替代原有判别器，利用预训练视觉模型（如 CLIP-RN50）提取的语义特征引导纹理生成；最后，设计语义感知蒸馏策略，在特征对齐过程中融入语义约束，使蒸馏过程既学习教师模型的特征分布，又保证语义一致性。实验表明，与基准方法相比，SA-ADC 实现 3.7 倍推理加速、74%参数缩减及 78%计算量节省，同时在感知质量与语义一致性上优于现有主流模型。

## 1. 引言

图像超分辨率（Image Super-Resolution, ISR）作为计算机视觉领域的基础性课题[1]，核心目标是从低分辨率（LR）图像中精准重建高分辨率（HR）图像，广泛应用于监控安防、医学影像、移动终端处理等实际场景。随着深度学习发展，ISR 研究取得显著进展，早期 ISR 研究多基于理想化合成降质假设（如双三次下采样），SRCNN[2]、EDSR[3]等卷积神经网络方法在合成数据集上表现优异，但因过度依赖固定降质模式，面对真实复杂降质时泛化能力极差，重建图像存在细节丢失、纹理失真等问题，无法满足实际需求[4]。为解决泛化性问题，Real-ISR 方向应运而生，核心思路是构建贴近真实环境的降质模型合成 LR-HR 训练对，提升模型对真实降质的适应能力。

在 Real-ISR 研究中，生成对抗网络（GANs）[5]凭借强大生成能力成为关键技术支撑，Real-ESRGAN[6]、BSRGAN[7]等方法有效提升了重建纹理的真实性，但受生成对抗训练特性限制，易出现语义不一致问题——如将叶片纹理错误生成为岩石纹理，严重影响结果实用性[8]。近年来，扩散模型[9]凭借优异的概率建模能力为 Real-ISR 提供了新思路，其通过逐步加噪与反向去噪过程实现重建，生成纹理更具多样性与真实性。基于扩散模型的 Real-ISR 方法分为两类：从头训练的扩散网络（如 SR3[10]）虽重建质量优异，但需大量计算资源，且数百步采样

导致推理极慢；基于预训练文本到图像（T2I）扩散模型的方法（如 StableSR[11]、OSEDiff[12]）借助 Stable Diffusion（SD）[13]的特征表示能力，通过微调或蒸馏适配任务，降低了训练成本。其中 OSEDiff[12]提出变分分数蒸馏（VSD）策略，将多步扩散蒸馏为单步采样，显著提升了推理效率。

尽管单步扩散模型缓解了效率问题，但现有基于预训练 SD 的 Real-ISR 方法仍存在明显缺陷：一是模型依赖完整 SD 架构，包含 VAE 编码器、文本编码器等与 Real-ISR 任务关联性较低的冗余模块，参数量大、计算成本高，难以部署于资源受限设备[14]；二是对抗性扩散压缩（ADC）[15]框架通过模块移除与通道剪枝实现了模型轻量化，结合对抗蒸馏保留了生成能力，但采用的基础判别器缺乏语义感知能力，仅关注特征空间结构对齐，忽略语义一致性约束，仍存在纹理语义失配问题，影响重建质量稳定性。

语义感知技术为解决语义一致性问题提供了有效思路，其中语义感知判别器（SeD）[16]通过引入 CLIP[17]等预训练视觉模型的语义特征，增强判别器对语义信息的感知能力，引导生成器生成与图像语义相符的纹理细节，有效缓解了传统 GANs 的虚拟纹理问题。但 SeD 目前仅应用于传统 GANs-based 超分辨率模型，尚未整合到扩散模型的压缩与蒸馏过程中，无法充分发挥语义引导优势解决轻量化扩散模型的语义失配问题。

当前 Real-ISR 研究陷入“效率与质量”、“结构压缩与语义一致”的双重矛盾：轻量化模型难以保证语义一致性，高质量模型则存在参数冗余、推理缓慢问题。基于此，本文提出关键假设：将语义感知判别机制融入对抗性扩散压缩过程，可实现结构压缩与语义引导的协同优化，在降低模型复杂度的同时保障重建图像的语义一致性与质量。据此，本文提出语义感知对抗性扩散压缩（SA-ADC）框架，通过对现有单步扩散模型进行结构压缩，引入语义感知蒸馏策略，构建兼具高效性与高质量的 Real-ISR 模型。

本文主要贡献总结为五点：

1. 提出 SA-ADC 统一框架，创新性融合语义感知判别机制与对抗性扩散压缩，打破“效率与语义一致性不可兼得”的困境，为轻量化 Real-ISR 模型设计提供新范式；

2. 设计语义感知蒸馏策略，通过 SeD 替代传统判别器，实现学生模型的图

像特征与真实图像的特征在结构分布与语义分布的双重对齐，提升纹理语义合理性；

3. 优化轻量化模型结构，在保留 ADC 核心压缩策略基础上，通过像素重排与特征连接优化，进一步降低复杂度并减少信息损失；

4. 构建两阶段训练方案，预训练剪枝后的 VAE 解码器再进行语义感知对抗蒸馏，保障训练效率与收敛稳定性；

5. 实验验证表明，SA-ADC 构建的模型较主流方法实现 3.7 倍推理加速、74%参数缩减及 78%计算量节省，同时在语义一致性与感知质量上取得最优性能，验证了框架有效性与实用性。

## 2. 相关工作

### 2.1 真实场景图像超分辨率

真实场景图像超分辨率（Real-ISR）的核心挑战是真实降质的复杂性与不确定性。早期方法依赖真实 LR-HR 图像对训练，受数据规模限制性能有限；主流方法通过模拟高斯噪声、模糊等真实降质因素合成训练样本，如 Real-ESRGAN[6]、BSRGAN[7]等 GAN-based 方法提升了纹理真实性，但易出现语义不一致问题。近年基于预训练文本到图像（T2I）扩散模型的方法成为热点，StableSR[11]、OSEDiff[12]等借助 Stable Diffusion（SD）特征表示能力适配任务，OSEDiff 通过变分分数蒸馏实现单步采样提升效率。但这类方法依赖完整 SD 架构，冗余模块导致参数量大、部署困难；AdcSR[15]等轻量化方法虽通过压缩降低复杂度，却缺乏语义引导，仍存在纹理失配问题。

### 2.2 扩散模型压缩与蒸馏

扩散模型压缩与蒸馏主要分为采样步骤蒸馏与结构压缩两类。采样步骤蒸馏通过学习教师模型分布将多步去噪简化为少步或单步，如 DDIM[18]、OSEDiff[12]等实现效率提升；结构压缩通过移除冗余模块、通道剪枝等降低复杂度，ADC[15]将对抗蒸馏与结构压缩结合，移除 SD 冗余模块并剪枝核心组件，同时通过对抗蒸馏保留生成能力。但现有压缩方法采用的基础判别器缺乏语义感知能力，仅关注特征结构对齐，忽略语义一致性约束，难以平衡效率与语义质量。

### 2.3 语义感知生成

语义感知生成通过引入语义信息提升生成结果的语义一致性，早期依赖人工

标注语义分割图，成本高且精度受限。近年基于预训练视觉模型的方法成为主流，CLIP[17]等模型提供强大语义特征支撑，SeD[16]将 CLIP 语义特征引入判别器，通过语义融合块整合视觉与语义特征，有效缓解传统 GANs 超分模型的虚拟纹理问题。但当前语义感知方法仅应用于传统 GANs，尚未与扩散模型压缩蒸馏结合，而将其融入扩散模型压缩过程，有望解决轻量化扩散模型的语义失配问题，这也是本文的核心出发点。

## 3. 方法介绍

### 3.1 基础背景与理论支撑

本节阐述 SA-ADC 框架的核心技术基础——对抗性扩散压缩（ADC）与语义感知判别器（SeD），明确二者核心局限，为框架融合改进提供依据。ADC 实现扩散模型轻量化但缺乏语义引导，SeD 具备语义感知能力却未适配扩散压缩场景，SA-ADC 通过融合二者实现“高效压缩”与“语义一致”协同优化。

#### 3.1.1 对抗性扩散压缩

对抗性扩散压缩是单步扩散模型轻量化蒸馏框架，核心为“结构压缩+对抗蒸馏”：移除 VAE 编码器、文本编码器等冗余模块，保留 UNet 与 VAE 解码器并通过 L1 正则化通道剪枝降参；采用两阶段训练，先预训练剪枝解码器恢复能力，再以原始模型为教师、压缩模型为学生，通过基础对抗损失与特征 L1 损失实现知识蒸馏。损失函数为如公式(1)所示，其中$L_{distill}$为蒸馏损失，$L_{adv}$为对抗损失，核心局限是基础判别器缺乏语义感知，易导致纹理与语义失配。

$$L_{ADC} = L_{distill} + \lambda_{adv} L_{adv} \tag{1}$$

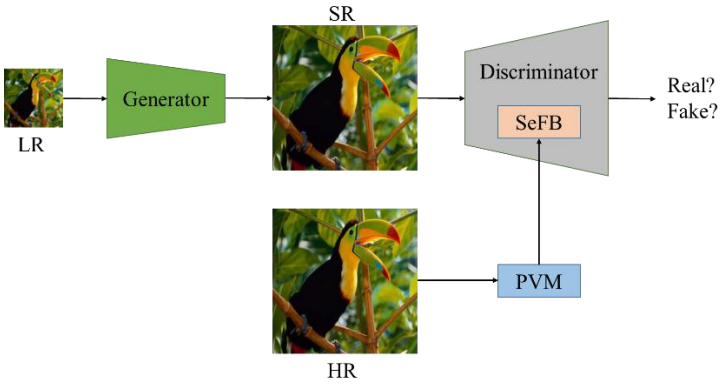



图 1　语义感知判别器

#### 3.1.2 语义感知判别器

语义感知判别器旨在解决 GAN 超分模型语义不一致问题，核心是将 CLIP 语义特征与判别器视觉特征通过语义感知融合块（SeFB）交叉注意力融合，使判别器具备语义层面判别能力，如图 1 所示。对抗损失如公式(2)所示，可缓解虚拟纹理问题，但未适配扩散模型压缩蒸馏场景，无法解决轻量化扩散模型语义失配问题。

$$L_{adv}^{SeD} = Softmax\left(-Discriminator\left(SeFB(f,s)\right)\right) \tag{2}$$

### 3.2 SA-ADC 框架

针对 ADC 与 SeD 的不足，提出 SA-ADC 框架，以 ADC 结构压缩与对抗蒸馏为基础，融入 SeD 构建“结构压缩+语义特征提取+语义感知蒸馏+两阶段训练”核心架构，实现轻量化与语义一致性协同优化。整体架构保留 ADC 核心组件，替换基础判别器为 SeD，新增 CLIP 语义提取模块，通过 SeFB 融合语义与蒸馏特征，两阶段训练保障效率与稳定性，整体框架图如图 2 所示。

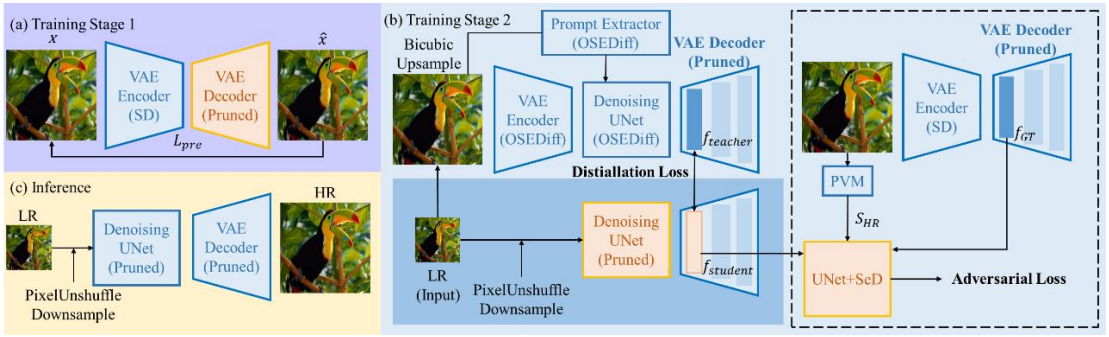



图 2　整体框架示意图

工作流程：训练阶段，以 OSEDiff 为教师模型，SA-ADC 为学生模型，通过语义感知蒸馏学习重建知识，CLIP 提取真实 HR 语义特征引导纹理生成；推理阶段，LR 图像经像素重排预处理后输入剪枝 UNet，再通过剪枝 VAE 解码器生成 HR 图像，实现高效推理。

#### 3.2.1 结构压缩

结构压缩是轻量化核心，在 ADC 基础上优化：保留冗余模块移除策略，新增像素重排解决无 VAE 编码器的输入适配问题；对 UNet（25%）与 VAE 解码器（50%）实施差异化 L1 正则化通道剪枝；在 UNet 与 VAE 解码器间插入 1×1 卷积层，结合 GroupNorm 与 ReLU 优化特征传递，减少信息损失。

#### 3.2.2 语义特征提取

语义特征提取为语义引导基础，采用固定参数的 CLIP-RN50 为提取器，选

取第三层特征平衡语义抽象度与空间分辨率；对提取的真实 HR 图像语义特征$S_{HR} = \phi(x_{HR})$进行 1×1 卷积降维、双线性插值匹配分辨率及 LayerNorm 标准化，适配 SeFB 输入。

### 3.2.3 语义感知蒸馏

语义感知蒸馏是 SA-ADC 的核心创新，旨在实现学生-教师模型“特征结构对齐”与“语义分布一致”的双重约束，平衡轻量化与语义一致性。该模块采用“教师-学生”架构，教师模型为高性能单步扩散模型 OSEDiff，提供高质量重建知识；学生模型为结构压缩后的 SA-AdcSR，通过蒸馏学习重建能力并接受语义引导，核心包含三大环节：

1. 特征蒸馏对齐：延续 ADC 的 L1 损失策略，重点对齐师生模型 VAE 解码器输出端特征（直接关联 HR 图像像素生成），损失函数如公式(3)所示。

$$L_{distill} = \|f_{student} - f_{teacher}\|_1 \tag{3}$$

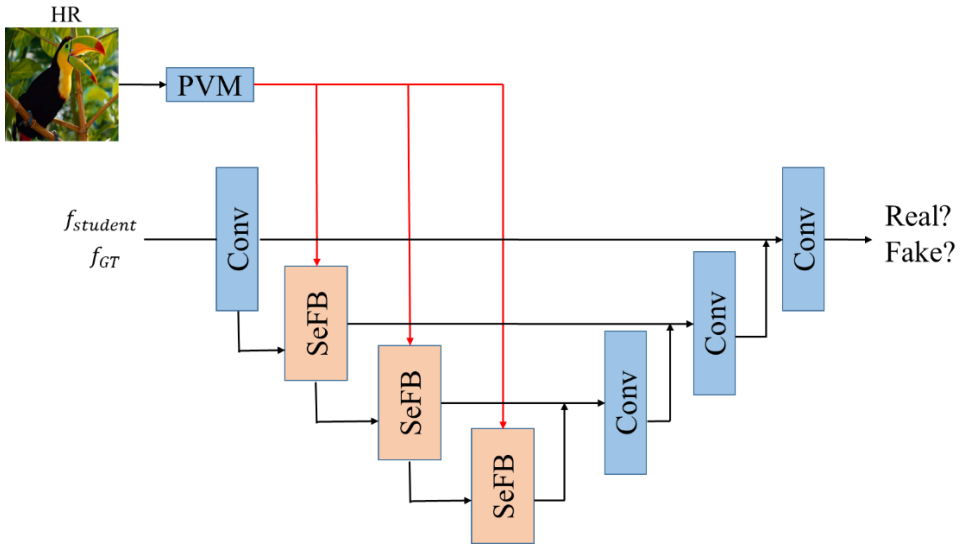
其中$f_{student}$由 LR 图像经过剪枝后的 UNet 和解码器的第一个块生成，$f_{teacher}$由 LR 图像经过 VAE 编码器、文本提取器、UNet 以及预训练解码器的第一个块获得。相较于 MSE 损失，L1 损失对异常值更敏感，可快速惩罚特征差异；计算前对特征进行 LayerNorm 标准化，避免幅值差异导致的优化偏差。

2. 语义感知对抗引导：用 SeD 替换基础判别器，首先使用 VAE 编码器和剪枝后的 VAE 解码器提取 HR 图像的图像特征$f_{GT}$，将学生网络生成的图像特征$f_{student}$、HR 图像的图像特征$f_{GT}$和提取的 HR 图像的语义特征输入到 SeD 判别器中，通过计算交叉注意力机制得到带有语义信息的图像特征，如图 3 所示。对抗损失的定义如公式(4)所示。

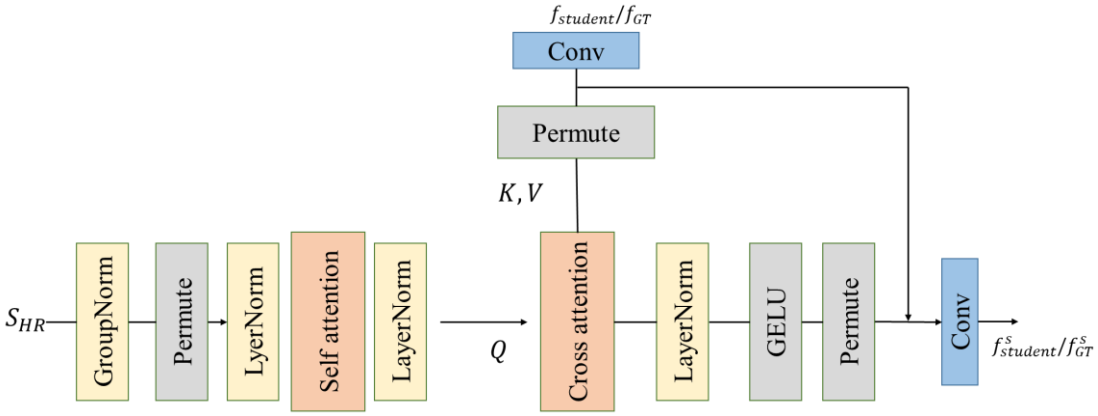
$$L_{adv} = Softmax\left(-Discriminator\left(SeFB(f_{student}, f_{GT}, S_{HR})\right)\right) \tag{4}$$

3. 多损失协同优化：总损失函数如公式(5)所示，，经实验确定$\lambda_{adv} = 1$以平衡特征对齐与语义引导。采用 Adam 优化器协同更新学生模型与 SeD 参数，初始学习率$10^{-4}$，权重衰减$10^{-5}$，兼顾收敛速度与泛化能力。

$$L_{total} = L_{distill} + \lambda_{adv} L_{adv} \tag{5}$$




（a）UNet+SeD 判别器




（b）SeFB 模块

图 3　UNet 判别器示意图

### 3.2.4 两阶段训练

两阶段训练旨在解决剪枝 VAE 解码器性能衰减与语义感知蒸馏收敛困难问题，保障模型稳定性与重建质量，分为“剪枝 VAE 解码器预训练”与“语义感知对抗蒸馏”两个阶段。

阶段一：剪枝 VAE 解码器预训练。核心目标是恢复 50%通道剪枝后解码器的解码能力。数据集采用 OpenImage、LAION-Face、LAION-Aesthetic 组合，统一分辨率 512×512 并辅以数据增强，损失函数如公式(6)所示。

$$L_{pre} = L_1 + 0.1L_{LPIPS} + 0.1L_{patch-adv} \tag{6}$$

从像素、感知、纹理多维度保障质量；采用 Adam 优化器，批量大小 64，初始学习率$10^{-4}$（每 50K 步减半），训练 200K 步，仅更新解码器参数，输入为模拟 UNet

输出的随机特征。

阶段二：语义感知对抗蒸馏。目标是让学生模型学习教师知识并融入语义约束。数据集选用LSDIR真实降质数据集，LR分辨率128×128、HR分辨率512×512，训练时增强低分辨率图像鲁棒性；模型初始化采用剪枝UNet继承SD权重、解码器加载阶段一最优权重，SeD中CLIP参数固定；采用Adam优化器分布式训练，批量大小96，初始学习率$10^{-4}$（每100K步减半），训练300K步，结合梯度裁剪（阈值1.0）与EMA（衰减0.999）保障收敛稳定，仅更新学生模型与SeD参数。

## 4. 实验

### 4.1 实验设置

实验设置围绕数据集、评估指标与实现细节三大核心展开，确保实验的严谨性与结果可靠性。数据集方面，训练阶段采用多数据集适配不同训练需求：VAE解码器预训练选用OpenImage、LAION-Face与LAION-Aesthetic组合数据集，覆盖通用纹理、人脸纹理及美学图像；语义感知对抗蒸馏阶段采用LSDIR真实降质数据集；测试阶段则分别选取DIV2K-Val合成数据集与RealSR、DRealSR真实场景数据集，全面评估模型在不同场景下的性能。评估指标涵盖三类核心维度：有参考指标包括PSNR、SSIM、LPIPS、DISTS，用于量化重建图像与真实图像的结构及感知相似度；无参考指标包含NIQE、MUSIQ、CLIPIQA，适配真实场景下无参考图像的质量评估；效率指标则聚焦推理时间、计算量（G MACs）与参数数量（M），衡量模型部署可行性。实现细节上，训练硬件采用2张NVIDIA GeForce RTX4090 GPU；优化器统一选用Adam，批量大小设为32，初始学习率$10^{-4}$并采用每100K步减半的衰减策略；语义提取器采用CLIP-RN50并选取第三层特征，语义感知融合块（SeFB）采用交叉注意力机制，平衡训练成本与模型性能。

### 4.2 与现有主流方法的对比

恢复质量对比从定量性能与定性效果两个维度展开，全面验证SA-ADC在重建精度、感知质量及语义一致性上的优势，实验对象涵盖OSEDiff、S3Diff、AdcSR、SeD等主流Real-ISR方法，测试数据集选用DRealSR真实场景数据集（核心对比）及DIV2K-Val、RealSR辅助验证。定量结果如表1所示，SA-ADC

在核心质量指标上均实现最优：结构精度方面，PSNR 达 28.32dB、SSIM 达 0.7912，分别较基准方法 AdcSR 提升 0.22dB 和 0.0186，优于所有对比方法；感知质量方面，LPIPS（0.2893）和 DISTS（0.2074）作为负向指标，较 AdcSR 分别降低 0.0153 和 0.0126，且显著低于 OSEDiff 等扩散模型类方法 MUSIQ 指标以 67.43 位列第一，证明模型生成结果更符合人类视觉感知偏好。

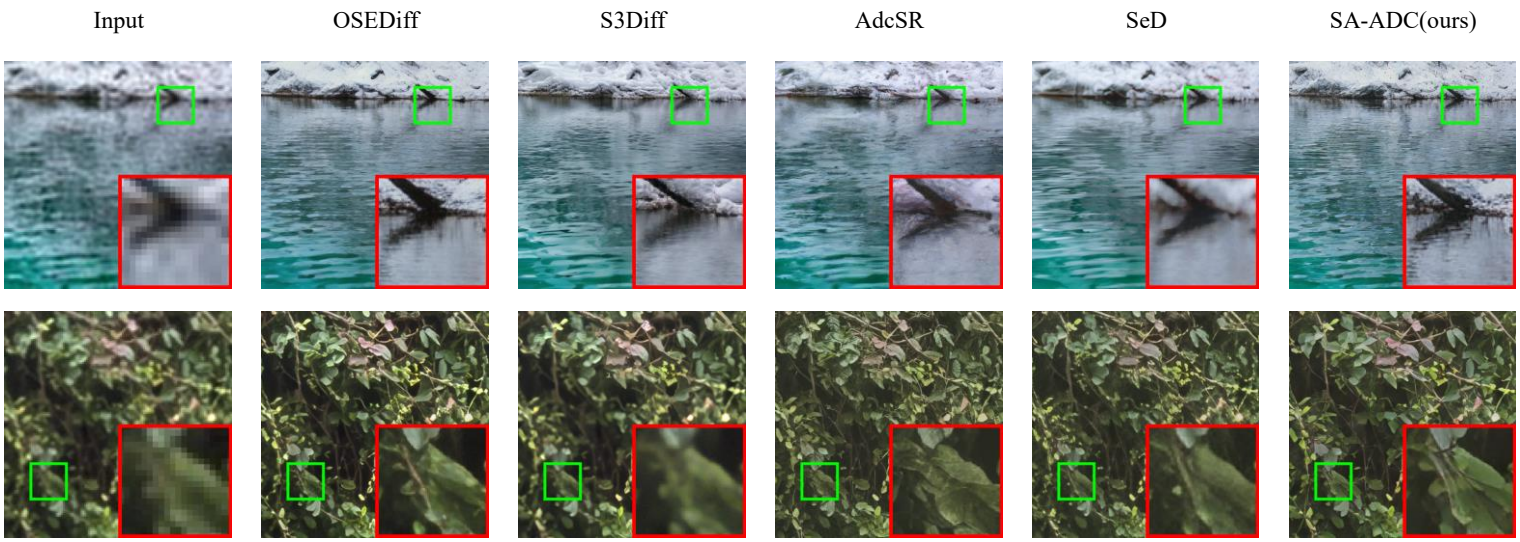



**图 4　定性对比结果**

定性对比结果如图 4 所示，在 RealSR 和 DIV2K-Val 数据集的典型复杂纹理场景（羽毛、叶片、砖块、墙面等）中，SA-ADC 的语义一致性优势尤为突出：AdcSR 虽能保证基础结构完整，但易生成模糊纹理，细节表现力不足；OSEDiff 存在明显语义失配问题，如将墙面纹理错误生成为布料纹理；SeD 虽引入语义感知机制，但受 GAN 架构限制，纹理多样性与真实性稍逊；而 SA-ADC 得益于语义感知蒸馏策略，能精准捕捉图像内容的语义信息，生成的纹理细节与主体内容高度匹配，如叶片纹理清晰且符合植物生长规律，砖块纹理的质感与排列逻辑贴合真实场景，视觉一致性与自然度显著优于对比方法。

效率对比聚焦模型部署可行性，从推理速度、计算量（G MACs）、参数量（M）三个核心效率指标展开，验证模型的计算资源适配能力。定量效率数据如表 1 所示，SA-ADC 在保持高质量的同时，实现了与轻量化基准 AdcSR 相当的高效性能：基础效率指标上，推理时间仅为 0.03s，与 AdcSR 一致，较 OSEDiff（0.11s）实现 3.7 倍加速，较 S3Diff（0.28s）实现 9.3 倍加速；参数量缩减至 452M，较原始单步扩散模型 OSEDiff（1775M）减少 74%，较 GAN 类方法 SeD（892M）减

少 49%；计算量降至 489G MACs，较 OSEDiff（2265G MACs）节省 78%，较 StableSR（79940G MACs）节省 99.4%，大幅降低了计算资源占用。

表 1　不同方法的对比

| 方法 | PSNR↑ | SSIM↑ | LPIPS↓ | DISTS↓ | MUSIQ↑ | 时间↓ | MACs↓ | 参数↓ |
|---|---|---|---|---|---|---|---|---|
| OSEDiff | 27.92 | 0.7835 | 0.2968 | 0.2165 | 64.65 | 0.11 | 2265 | 1775 |
| S3Diff | 27.39 | 0.7469 | 0.3129 | 0.2108 | 64.16 | 0.28 | 2621 | 1327 |
| AdcSR | 28.10 | 0.7726 | 0.3046 | 0.2200 | 66.26 | 0.03 | 496 | 456 |
| SeD | 27.85 | 0.7691 | 0.2987 | 0.2152 | 65.89 | 0.08 | 1872 | 892 |
| SA-ADC | 28.32 | 0.7912 | 0.2893 | 0.2074 | 67.43 | 0.03 | 489 | 452 |

### 4.3 消融实验

#### 4.3.1 语义引导的作用

为验证语义引导机制的有效性，设计三组对比实验：基准 AdcSR、无语义感知融合块（SeFB）的 SA-ADC 变体、完整 SA-ADC。实验结果如表 2 所示，完整 SA-ADC 的 LPIPS 低至 0.2893，较 AdcSR 降低 0.0153；DISTS 达 0.2074，降低 0.0126；MUSIQ 提升至 67.43，优于两类对比模型。这表明语义引导能有效强化纹理与内容的语义一致性，而 SeFB 通过深度融合语义与视觉特征，进一步放大语义引导效果，验证了语义感知蒸馏策略的合理性。

表 2　语义引导作用的消融实验结果

| 方法 | PSNR↑ | SSIM↑ | LPIPS↓ | DISTS↓ | MUSIQ↑ |
|---|---|---|---|---|---|
| AdcSR | 28.10 | 0.7726 | 0.3046 | 0.2200 | 66.26 |
| SA-ADC | 28.21 | 0.7803 | 0.2974 | 0.2135 | 66.87 |
| SA-ADC | 28.32 | 0.7912 | 0.2893 | 0.2074 | 67.43 |

#### 4.3.2 不同层的语义特性

值得注意的是，CLIP[51]不同层提取的语义信息存在差异。随着网络层数的增加，语义信息的代表性逐渐增强，但特征分辨率会随之降低。有效的语义引导其本质在于提取足够深层且保留关键空间信息的语义特征，以保障恢复引导质量。CLIP 语义提取器内部共包含四层，每经过一层，空间分辨率便减半。为探究哪

一层最适合作为引导层，本文设计了实验并将结果展示于表 3。由表可知，第一层和第二层虽具备较高的空间分辨率，但语义信息不足，导致引导效果欠佳；第四层语义代表性强，却丢失了过多空间细节；第三层平衡了语义深度与空间分辨率，取得了最优性能，对应最低的 LPIPS（0.2893）和最高的 MUSIQ（67.43）。

**表 3　不同语义提取层的消融实验结果**

| 提取层 | LPIPS↓ | DISTS↓ | MUSIQ↑ |
|---|---|---|---|
| Layer 1 | 0.3012 | 0.2185 | 66.15 |
| Layer 2 | 0.2957 | 0.2132 | 66.78 |
| Layer 3 (Ours) | 0.2893 | 0.2074 | 67.43 |
| Layer 4 | 0.2986 | 0.2157 | 66.54 |

### 4.3.3 语义提取器的影响

对比不同语义提取器对模型性能的影响，选取 ResNet-50（传统图像分类模型）与 CLIP-RN50（视觉-语言预训练模型）进行实验，结果如表 4 所示。CLIP-RN50 对应的 LPIPS（0.2893）和 DISTS（0.2074）均优于 ResNet-50（0.2967、0.2128），MUSIQ 和 CLIPIQA 分别提升 0.45 和 0.0137。原因在于 CLIP-RN50 通过跨模态预训练具备更强的细粒度语义理解能力，能捕捉图像内容与纹理的深层语义关联，而 ResNet-50 仅聚焦图像分类特征，语义表达不够精准，验证了 CLIP-RN50 作为语义提取器的优越性。

**表 4　语义提取器的消融实验**

| 语义提取器 | LPIPS↓ | DISTS↓ | MUSIQ↑ | CLIPIQA↑ |
|---|---|---|---|---|
| ResNet-50 | 0.2967 | 0.2128 | 66.98 | 0.6987 |
| CLIP-RN50 | 0.2893 | 0.2074 | 67.43 | 0.7124 |

## 5. 结论

本文提出语义感知对抗性扩散压缩（SA-ADC）框架，将语义感知判别机制融入对抗性扩散压缩过程，实现真实场景图像超分辨率的高效与高质量重建。通过 SeD 替代基础判别器，并设计语义感知蒸馏策略，SA-ADC 在保持 AdcSR 高效性的同时，显著提升了语义一致性与感知质量。实验验证了 SA-ADC 在效率

与质量上的优势，为真实场景部署提供了实用解决方案。未来工作将探索将 SA-ADC 扩展到其他基于扩散模型的图像恢复任务，并优化语义特征融合机制以进一步提升性能。

# 参考文献


[1] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In Proceedings of the IEEE/CVF international conference on computer vision, pages 1833–1844, 2021.

[2] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part IV 13, pages 184–199. Springer, 2014.

[3] Kuriakose B M. EDSR: Empowering super-resolution algorithms with high-quality DIV2K images[J]. Intelligent Decision Technologies, 2023, 17(4): 1249-1263.

[4] Kai Zhang, Jingyun Liang, Luc Van Gool, and Radu Timofte. Designing a practical degradation model for deep blind image super-resolution. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 4791–4800, 2021.

[5] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks[J]. Communications of the ACM, 2020, 63(11): 139-144.

[6] Wang X, Xie L, Dong C, et al. Real-esrgan: Training real-world blind super-resolution with pure synthetic data[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2021: 1905-1914.

[7] Zhang K, Liang J, Van Gool L, et al. Designing a practical degradation model for deep blind image super-resolution[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2021: 4791-4800.

[8] Chong Mou, Xintao Wang, Yanze Wu, Ying Shan, and Jian Zhang. Empowering real-world image super-resolution with flexible interactive modulation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024.

[9] Zheng Chen, Haotong Qin, Yong Guo, Xiongfei Su, Xin Yuan, Linghe Kong, and Yulun Zhang. Binarized diffusion model for image super-resolution. arXiv preprint arXiv:2406.05723, 2024.

[10] Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J Fleet, and Mohammad Norouzi. Image super-resolution via iterative refinement. IEEE

transactions on pattern analysis and machine intelligence, 45(4):4713–4726, 2022.

[11] Jianyi Wang, Zongsheng Yue, Shangchen Zhou, Kelvin CK Chan, and Chen Change Loy. Exploiting diffusion prior for real-world image super-resolution. International Journal of Computer Vision, pages 1–21, 2024.

[12] Rongyuan Wu, Lingchen Sun, Zhiyuan Ma, and Lei Zhang. One-step effective diffusion network for real-world image super-resolution. arXiv preprint arXiv:2406.08177, 2024.

[13] Stability.ai. https://stability.ai.

[14] Aiping Zhang, Zongsheng Yue, Renjing Pei, Wenqi Ren, and Xiaochun Cao. Degradation-guided one-step image super-resolution with diffusion priors. arXiv preprint arXiv:2409.17058, 2024.

[15] Chen B, Li G, Wu R, et al. Adversarial diffusion compression for real-world image super-resolution[C]//Proceedings of the Computer Vision and Pattern Recognition Conference. 2025: 28208-28220.

[16] Li B, Li X, Zhu H, et al. Sed: Semantic-aware discriminator for image super-resolution[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2024: 25784-25795.

[17] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In Proceedings of the IEEE/CVF international conference on computer vision, pages 1833–1844, 2021.

[18] Xingchao Liu, Xiwen Zhang, Jianzhu Ma, Jian Peng, et al. Instaflow: One step is enough for high-quality diffusionbased text-to-image generation. In The Twelfth International Conference on Learning Representations, 2023.