四元数小波条件扩散模型用于图像超分辨率

Luigi Sigillo, Christian Bianchi, Aurelio Uncini, and Danilo Comminiello
Dept. Information Engineering, Electronics and Telecommunications (DIET), Sapienza University of Rome, Italy
Email: luigi.sigillo@uniroma1.it.

摘要—图像超分辨率是计算机视觉中的一个基本问题，具有广泛的应用范围，从医学成像到卫星分析。从低分辨率输入中重建高分辨率图像是至关重要的，这对于增强下游任务如目标检测和分割非常重要。虽然深度学习显著推进了超分辨率技术的发展，但在实现带有精细细节和逼真纹理的高质量重建方面仍然面临挑战，尤其是在高度放大因子的情况下。最近利用扩散模型的方法展示出有希望的结果，但它们通常难以在感知质量和结构保真度之间取得平衡。在这项工作中，我们引入了一种新的 SR 框架 ResQu，它结合了四元数小波预处理框架与潜在扩散模型，并采用了一个新的四元数小波和时间感知编码器。与以往方法简单地在扩散模型中应用小波变换不同，我们的方法通过利用动态集成于去噪过程各阶段的四元数小波嵌入来增强条件生成过程。此外，我们还利用了基础模型如 Stable Diffusion 的生成先验。在特定领域数据集上的广泛实验表明，我们的方法实现了卓越的超分辨率结果，在许多情况下感知质量和标准评估指标方面都优于现有方法。代码将在修订流程后提供。

***Index Terms*—生成式深度学习，图像超分辨率，扩散模型**

I. 介绍

图像超分辨率 (SR) 是计算机视觉中的一个基石问题，对从医学成像到卫星分析及更多领域的应用具有深远影响 [2], [3]。从低分辨率 (LR) 图像重建高分辨率 (HR) 图像是不仅是一项技术挑战，也是实际需求。确实，LR 图像常常会妨碍诸如目标检测、分割和分类等下游任务的性能 [4]。尽管经过数十年的研究，实现能够保持细粒度细节和逼真纹理的超分辨率，特别是在高放大倍率下，仍然是一个开放性问题。

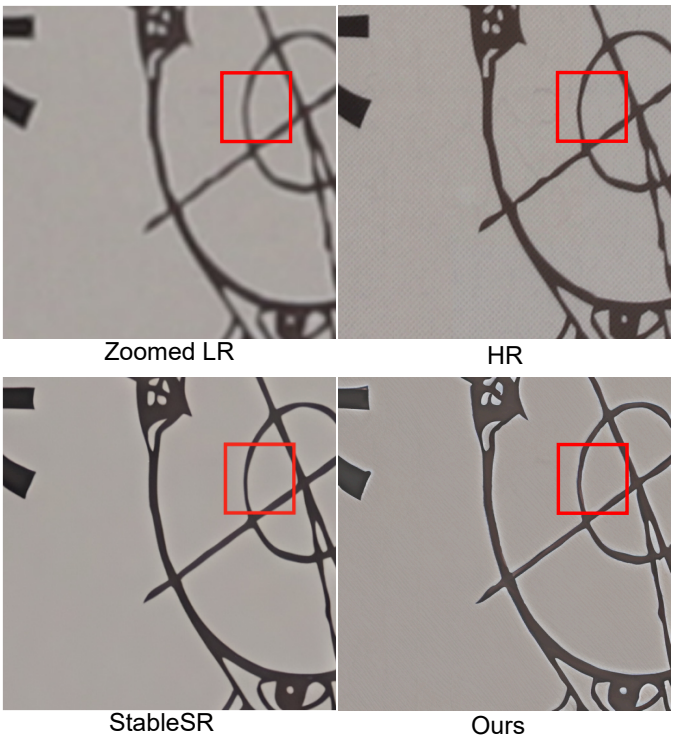

图 1. 对真实世界图像进行定性比较，从 128 像素放大到 512 像素，我们对比了 StableSR [1] 和地面实况，并将我们的结果展示在最后一列。视觉评估突显了图像质量、清晰度以及我们在超分辨率过程中模型所实现的细微细节增强之间的差异。可以明显看出，StableSR 结果的背景是平滑的，而在地面实况中它是粗糙的，类似于我们模型输出的结果。

传统上的超分辨率（SR）方法，如插值、频域变换和基于滤波的技术，已经提供了基础的解决方案，但往往无法生成感知上令人信服的结果。这些方法通常难以恢复高频细节，导致产生伪影和模糊，从而降低图像质量 [5]。随着深度学习的出现，超分辨率（SR）领域经历了一场变革性的转变。卷积神经网络（CNN）、视觉

This work was partly supported by “Ricerca e innovazione nel Lazio - incentivi per i dottorati di innovazione per le imprese e per la PA - L.R. 13/2008” of Regione Lazio, Project “Deep Learning Generativo nel Dominio Ipercomplesso per Applicazioni di Intelligenza Artificiale ad Alta Efficienza Energetica”, under grant number 21027NP000000136, and by the European Union under the Italian National Recovery and Resilience Plan (NRRP) of NextGenerationEU, “Rome Technopole” (CUP B83C22002820006)—Flagship Project 5: “Digital Transition through AESA radar technology, quantum cryptography and quantum communications”.

变换器（ViT），以及最近的生成模型都为 SR 性能设定了新的基准。在这些进步中，扩散模型作为图像生成的一种特别有前景的方法脱颖而出。研究表明，它们具有通过迭代地将噪声精炼成结构化输出 [6]–[9] 来生成连贯图像的无与伦比的能力。在此基础上，在将扩散模型应用于超分辨率任务方面已经取得了许多最近的进步 [10]–[14]。

当代基于扩散的超分辨率研究越来越多地关注探索结合小波技术的协同潜力，利用它们在多尺度分析和特征分解方面的互补优势 [15]。例如，[16] 引入了一个基于小波的扩散框架，通过自适应处理低频和高频分量来弥补生成模型的速度差距。在此基础上，其他研究 [17], [18] 进一步探索了专门针对超分辨率任务的小波扩散方法，为将小波变换与生成模型结合奠定了坚实的基础。这些研究表明，在实现最先进的超分辨率性能方面，多尺度特征提取和自适应处理的重要性。

受该领域近期进展的启发，我们提出了一种新颖的超分辨率框架，它将四元数小波预处理框架 [19] 与潜在扩散模型相结合。与先前的工作简单地结合小波变换和扩散模型不同，我们的方法引入了一个具有四元数小波和时间感知的编码器，这在去噪步骤中显著增强了条件生成过程。该编码器利用动态集成于去噪过程中各个阶段的四元数小波嵌入，从而对超分辨率图像的生成进行细粒度控制。通过利用诸如 Stable Diffusion (SD) [20] 等最先进的文本到图像生成模型学习到的潜在表示，我们的框架不仅推进了现有的超分辨率方法，还建立了一种新的架构范式，将生成先验与超分辨率技术相结合。

四元数小波嵌入捕获了低频近似值和高频细节，在去噪过程中，我们的编码器在多个尺度上对这些进行加权和调节。这种多尺度调节机制与 SD 的生成先验相结合，使我们的框架能够在定量指标中实现优越性能，并且在定性评估中也取得有竞争力的结果。为了验证我们的方法，我们在多样化的数据集上进行了广泛的实验和消融研究，展示了我们方法在实际场景中的鲁棒性和灵活性。

本工作的主要贡献如下：

1) 我们引入了一种集成四元小波特征和生成先验的图像超分辨率潜在扩散模型。
2) 我们提出了一种新颖的四元数小波与时序感知编码器，该编码器在去噪过程中增强了多尺度下的条件生成过程，显著提升了 SD 等预训练基础模型的性能。
3) 我们通过广泛的实验表明，我们的方法在感知质量和定量指标上都优于现有方法。
4) 我们对架构设计进行了消融研究，并且还在特定领域的数据集上进行了研究。

本文组织如下。在第 II节中，我们提供了理论背景，并讨论了 SR 的挑战和应用。在第 III节中，我们介绍了所提出的方法 StableShip-SR，详细描述其架构和关键技术革新。在第 IV节中，我们展示了实验结果并与最先进的方法进行了比较。最后，第 V节通过讨论我们的发现及未来研究的潜在方向来总结本文。

II. 背景

图像超分辨率是计算机视觉中的一个基本任务，旨在从退化的低分辨率（LR）观测中恢复高分辨率（HR）图像。给定一个低分辨率图像 $\tilde{\boldsymbol{x}}$，目标是从中重构一个高分辨率图像 $\boldsymbol{x}$，其中两者之间的关系被建模为：

$$\tilde{\boldsymbol{x}} = (\boldsymbol{x} \otimes k) + n, \tag{1}$$

其中 k 表示退化矩阵，$\otimes$ 表示一种卷积类操作，而 n 是一个噪声项。这一任务本质上是不适定的，因为多个高分辨率解可能对应于相同的低分辨率输入，使得重构高度模糊且具有挑战性 [21]–[24]。

基于深度学习的方法。早期的超分辨率方法主要依赖于传统的技术，如双三次插值和频率域滤波，这些方法假设了简单的退化模型，并在恢复精细细节方面取得了有限的成功。深度学习的到来标志着一个转折点，诸如 SRCNN 的 [22] 模型率先将卷积神经网络应用于超分辨率任务。随后的架构，包括 EDSR [25]、ESRGAN [26] 和 RCAN [27]，展示了卷积神经网络学习低分辨率和高分辨率域之间复杂映射的能力，显著提高了超分辨率图像的感知质量。

卷积神经网络（CNNs）主导了基于深度学习的超分辨率研究的早期格局，直到视觉变压器（ViT）[28] 和 SwinIR [29] 的引入，后者通过利用移位窗口注意力机制来高效处理大规模图像并建模长距离依赖关系。

尽管有这些进展，此类模型通常在全面捕捉全局语义上下文方面表现出局限性，这是在复杂现实世界应用中一个关键的缺点。

生成模型用于超分辨率。近期的进展看到了一个向生成模型（包括 GAN 和扩散模型）转变的趋势，用于超

分辨率任务。基于 GAN 的方法，如 PULSE [30]，利用对抗训练来生成感知上真实的细节。然而，GAN 训练的不稳定性以及产生不自然伪影的趋势仍然是显著挑战。盲超分辨率方法，如 BSRGAN [31] 和 Real-ESRGAN [32]，引入了更准确反映现实世界条件的复杂退化管道。尽管有这些进展，许多超分辨率方法仍然依赖于预定义的退化假设，限制了它们的通用性。

扩散模型 [33] 提供了一个更稳定和可控的图像生成框架。这些模型通过迭代优化噪声输入，在去噪过程中利用数据分布生成高分辨率图像。

在基于扩散的超分辨率（SR）模型中，SR3 [10] 展示了卓越的表现，采用了由 BigGAN [34] 中获取的 G-Blocks 引导的条件扩散过程。然而，SR3 对像素空间扩散的依赖使其计算成本高昂。

潜变量扩散模型（LDMs）[20] 通过将去噪过程转移到压缩后的潜在空间来应对这些挑战，显著降低了计算开销同时保持高分辨率输出。

StableSR [1] 对预训练的文字到图像的潜变量扩散模型 Stable Diffusion (SD) [20] 进行了微调，通过引入时间感知编码器平衡保真度和感知质量。

类似地，DiffBIR [35] 采用两阶段过程，首先重建图像然后使用扩散先验增强细节。

这些方法有效地利用封装在 T2I 模型中的广泛生成先验，该模型基于庞大且多样的数据集进行训练，以解决复杂的现实世界超分辨率挑战。

我们的方法。尽管取得了成功，基于扩散的超分辨率方法仍面临某些限制，包括高昂的计算成本和难以保留特定领域的细节。在这项工作中，我们介绍了一种结合潜在扩散模型与四元数小波的新方法，利用 QUAVE [19]，一个最初为医学成像任务（如分割和重建）设计的预处理框架。QUAVE 利用四元数小波变换将图像分解为低频和高频成分，捕捉数据的丰富多维表示。

我们的方法将 QUAVE 作为潜扩散过程中的条件机制进行集成，使关键的领域特定任务细节得以保留。利用 QUAVE 提取的空间和频域特征，本方法为超分辨率模型提供了更丰富的输入，增强了其在多样化数据集上的泛化能力。此外，与需要从头开始训练的方法不同，我们的方法微调了预训练的 SD [20]，显著降低了计算需求同时保持了最先进的性能。

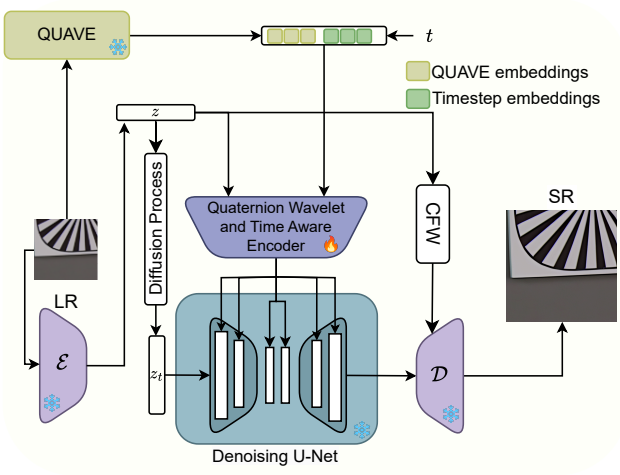

图 2. ResQu 超分辨率框架概述。我们预训练了 QUAVE，并仅训练了我们的编码器，即四元数嵌入和时间感知编码器。框架的其他部分被冻结，这加快了整体训练过程。

III. 提出的 ResQu 方法

在本节中，我们描述了我们的提出方法，该方法利用 SD [20] 的生成先验，并引入了一种新颖的四元数小波和时间感知编码器用于图像超分辨率。通过将四元数小波嵌入和时序调节整合到扩散过程中，我们的方法实现了高保真度的重建同时保留了特定领域的细节。

A. 四元数小波与时态感知编码器

实小波变换。一维离散小波变换（1D-DWT）使用尺度函数 $\phi(t)$ 和小波函数 $\psi(t)$ [36] 来刻画信号 $f(t)$。二维离散小波变换通过空间维度上的张量积扩展了这一点，得到了尺度函数 $\phi(x)\phi(y)$ 和三个分别对应对角线、水平和垂直特征的小波 $\psi(x)\psi(y)$、$\phi(x)\psi(y)$ 和 $\psi(x)\phi(y)$ [37], [38]。这种分解产生了一个低频（LL）和三个高频子带（LH，HL，HH）。然而，离散小波变换缺乏平移不变性和相位信息保存性，使得小波系数对采集变化敏感 [39], [40]。

四元数小波变换。四元数小波变换（QWT）集成了四个四元数小波变换，通过实 DWT 及其沿 x、y 和 xy 轴的三个相应的希尔伯特变换组合而成。该变换将输入图像分解为四个四元数小波子带，总共产生 16 个实子带。QWT 分解遵循与 DWT 类似的结构，生成一个标度函数表示为 ϕ_q 和三个方向的小波：ψ_q^D、ψ_q^V 和 ψ_q^H，分别对应对角线、垂直和水平方向 [41], [42]。

四元数小波嵌入。直接将所有十六个 QWT 子带作为神经模型的输入引入冗余和低效性 [43]。为了解决这

些挑战，我们采用了 QUAVE [19]，这是一种基于学习的方法，仅从 QWT 最具信息量的子带中提取特征，该方法利用了双树四元数小波变换（QWT）[44]。

我们的四元数小波及时感知编码器利用从 QUAVE 获得的四元数小波嵌入来提供多维表示，捕捉对于超分辨率至关重要的低频和高频信息。我们的编码器处理 LR 输入图像以提取增强特征。这些特征通过空间特征变换（SFT）[45] 被注入去噪过程，并且富含来自频率域和空间域的信息。

在扩散过程中，QUAVE 嵌入与时间步长嵌入一起通过编码器，并用于通过 SFT 调节 U-Net 的中间特征图。这确保了模型结合全局和局部细节，提升了重建高分辨率图像的真实性和保真度。

时间感知嵌入。除了四元数小波嵌入之外，编码器还通过嵌入时间步 t 到条件机制中，纳入时间信息。在去噪过程的早期阶段，当潜在表示具有低信噪比并包含大量噪声时，编码器使用从 QUAVE 衍生的增强特征提供强力指导。这确保了图像的结构完整性在生成过程的初始阶段得到保留。这是因为，如在 [46] 中发现，当信噪比增加且潜在表示变得更加精细时，编码器会降低其影响，从而使扩散模型能够专注于细粒度细节。这种自适应行为对于在保持全局结构和增强局部纹理之间取得平衡至关重要。扩散过程已知的噪声调度使编码器能够在每个时间步动态调整其条件强度，从而确保在超分辨率管道中提供最佳指导。

B. 潜扩散过程

条件过程定义为将 QUAVE 和时间步嵌入张量连接成一个统一的向量 $b \in \mathbb{R}^{1024}$，如图 2所示。此向量与输入低分辨率图像 $x \in \mathbb{R}^{3\times H\times W}$ 的潜在表示 $z \in \mathbb{R}^{h\times w\times c}$ 一起处理，其中 $z = \mathcal{E}(x)$ 是通过编码器 $\mathcal{E}$ 获得的。

对于我们的 ResQu，我们利用了来自 [20] 的预训练变分自编码器（VAE），其中编码器 $\mathcal{E}$ 和解码器 $\mathcal{D}$ 保持冻结。此外，我们还采用了他们的时间条件 U-Net 骨干网络用于潜扩散过程。

为了增强条件机制，我们引入了编码器 δ_θ，其设计旨在模仿 U-Net 编码器架构。这种结构能够从多个尺度中提取条件嵌入，并将其随后整合到条件去噪自编码器 ϵ_θ 中，从而促进更精细和分层的特征表示。我们优化以下目标函数：

$$L = \mathbb{E}_{\mathcal{E}(x),y,\epsilon\sim\mathcal{N}(0,1),t,c}\left[\|\epsilon - \epsilon_\theta(z_t, \delta_\theta(c,t,z))\|_2^2\right]. \quad (2)$$

通过整合四元数小波嵌入和时间引导，我们的方法有效地利用了 SD 的生成先验进行船舶图像超分辨率处理。这种结合确保了高保真重建、增强细节保留以及在各种场景中的鲁棒性。

IV. 实验结果

A. 实验设置

所有实验均使用配备 48GB 显存的 NVIDIA RTX A6000 GPU 进行。我们采用了学习率为 5e-05 的 AdamW 优化器。批次大小被限定为 6，以平衡内存效率和稳定收敛。模型训练了 14k 次迭代，从 StableSR [1] 的发布检查点开始。我们将超分辨率从 128x128 提升到 512x512，因此放大倍数为 4。QUAVE 在一个结合数据集上单独进行了预训练，该数据集由 DIV2K [52]、Flickr2K [53] 和 OutdoorSceneTraining (OST) [54] 组成，共进行了 45k 步，确保在多种图像领域中具备强大的特征提取能力。

B. 定量评估

我们在三个基准数据集上评估了模型的性能：DRealSR [50]、RealSR [51] 和 DIV2K，每个数据集都为超分辨率提出了独特的挑战。表 I中总结的数量化结果显示了我们方法的优越性能。我们使用 PSNR 和 SSIM（在 YCbCr 空间的 Y 通道上计算）、FID 和 LPIPS 来评估超分辨率模型，确保全面的像素级准确性和感知质量评估。PSNR 通过测量像素差异量化重建保真度，但它无法与人类感知对齐。SSIM 通过对结构相似性的考虑改进了 PSNR，但仍受限于捕获感知现实性。相反，FID 通过比较学习嵌入空间中的特征分布来评估逼真度，有效地捕捉感知差异。LPIPS 利用深度网络激活提供了更符合人类的相似性衡量标准。这些感知指标更好地反映了图像质量，尤其是在对抗性和生成模型中，传统像素级方法表现不足 [10]。

在 DIV2K [52] 数据集上，我们的模型表现出特别强的性能，在实现更高的 SSIM 和 PSNR 的同时，还具有竞争力的 FID 和 LPIPS 得分。这些结果表明，即使在高分辨率且纹理多样的图像上，我们的方法也能有效保持结构保真度和感知质量。

对于包括真实世界退化的 RealSR [51] 和 DRealSR [50] 数据集，我们的方法展现了强大的泛化能力，实现了更高的 PSNR 和 SSIM，这也表明我们的模型有效保留了细粒度纹理，从而导致感知上更优越的重建。

表 I
与最先进的方法在合成和真实世界基准上的定量比较。每项指标的最佳性能为**粗体**，而第二好的是 下划线的。

	指标	RealSR [47]	BSRGAN [31]	FeMaSR [48]	R-ESRGAN+ [32]	ResShift [49]	LDM [20]	StableSR [1]	**我们的**
DIV2K-验证集	PSNR ↑	24.62	24.58	22.97	24.29	24.53	20.58	23.26	**25.21**
	SSIM ↑	0.5970	0.6269	0.5887	0.6372	0.7323	0.5762	0.5726	**0.645**
	LPIPS ↓	0.5276	0.3351	0.3126	**0.3112**	0.4406	0.3199	0.3114	0.4161
	FID ↓	49.49	44.22	35.87	37.64	49.16	26.47	**24.44**	25.43
真实分辨率	PSNR ↑	25.56	24.70	23.58	24.33	24.79	22.26	23.55	**26.45**
	SSIM ↑	0.7390	**0.7651**	0.7132	0.7456	0.7423	0.6462	0.7080	0.7627
	LPIPS ↓	0.3570	0.2713	0.2937	**0.2524**	0.3134	0.3159	0.3002	0.3215
D 真实 SR	PSNR ↑	27.79	26.18	24.56	25.82	27.87	23.39	24.85	**29.69**
	SSIM ↑	0.8148	0.8028	0.7569	0.8052	0.8056	0.7448	0.7536	**0.8211**
	LPIPS ↓	0.3938	**0.2929**	0.3157	0.2818	0.5408	0.3379	0.3284	0.3332

C. 定性分析

为进一步评估我们方法的有效性，我们在图 3中进行了超分辨率图像的定性比较。视觉结果突显了我们的方法在现有技术上的几项关键优势。

我们的模型在重构复杂的结构细节和精细纹理方面表现出色。基于四元数小波的条件处理有效地保留了锐利边缘，同时减少了基线方法中常见的伪影。这一点在林肯肖像（图 3网格中的倒数第二幅图像）中尤为明显，我们的方法保留了微妙的面部特征，如细纹和发丝纹理，几乎没有过度平滑。与替代方法相比，我们的模型生成了更加自然的表现形式，避免了过度模糊或不自然的锐度。

此外，我们的方法在重建复杂高频纹理（如图 3网格中的第一行所示的猞猁皮毛）方面表现出增强的鲁棒性。传统的超分辨率方法往往引入噪声或无法恢复此类纹理中的自然随机模式，而我们的方法保持了结构一致性，有效地捕捉到了粗细不等的皮毛图案。视觉伪影的减少和纹理真实感的提升使得输出更加悦目。

超分辨率的一个特别具有挑战性的方面在于处理文本元素，因为必须以高保真度恢复细小细节以保持可读性。我们的方法在这一点上显著优于先前的方法，如“Gartner”文本示例（图 3网格中的第五行）所示，在该示例中字母边缘保持清晰，细微的字体特征得以保存且没有混叠或失真。传统方法倾向于引入模糊或边缘伪影，而我们的模型保留了锐度和可读性，突显其在基于文本的图像增强应用中的优势。

总体而言，我们的实验结果验证了所提出的方法在实现最先进的超分辨率性能方面的有效性。平衡结构保真度和感知质量的能力，涵盖了从人像和自然纹理到文本元素等各种内容类型，证明了其在现实世界图像增强任务中的广泛适用性。

D. 消融研究

采样步骤数量的消融研究。为了公平比较，我们将采样步骤的数量设置为 200，遵循 StableSR 的配置 [1]。随后，我们探索了减少这个数量，并观察到我们的模型从较少的采样步骤中受益。图 4描述了这种减少对关键评估指标的影响。我们发现，减少采样步骤的数量在评估指标之间引入了一种权衡。具体来说，我们观察到 PSNR 和 SSIM 增加，这表明与地面真实值的像素级和结构相似性得到了改善。然而，LPIPS 也增加了，这表明感知质量下降了。这表明虽然较少的步骤增强了传统相似度指标方面的保真度，但可能损害了感知现实主义。因此，采样步骤的选择取决于实际应用中期望在像素级准确性和感知质量之间的平衡。

可控特征包裹的消融研究。为了彻底评估不同可控特征包裹（CFW）实现的影响，我们对我们自定义训练的 CFW 模块与 StableSR 引入的一个进行了比较分析。[1] 训练流程生成了专门设计用于增强模块保持真实纹理和自然图像统计能力的合成 LR-HR 对。

表 II展示了两个模型变体在不同指标上的定量结果：我们的模型使用自训练的 CFW 和我们的模型利用由 StableSR 预训练的 CFW 模块。改进尤其在具有复杂几何图案和锐边的区域中显著，这表明我们的训练方法有效地利用了四元数小波特征捕获的结构信息，这一点通过 SSIM 和 PSNR 指标得到了证实。

图 3. LR 输入图像与最新方法和我们提出的模型在 DRealSR [50] 和 RealSR [51] 数据集上生成的 SR 输出的比较。红色边框突出显示了一个放大区域，展示了我们的方法相比现有方法在分辨率和细节保存方面的优越性。

这些结果表明，CFW 实现的选择应由特定的应用需求指导。当结构保真度至关重要时，我们定制训练的 CFW 提供了更优的结果。在图 5中展示了使用不同 CFW 生成输出的视觉比较。

船舶检测数据集的消融研究。为了评估我们模型的泛化能力和其在无需重新训练的情况下应用于特定领域任务的能力，我们在 ShipSpotting 数据集上进行了零样本评估 [13]。该数据集包含在不同条件下捕获的各种海上船舶图像，由于复杂的纹理、精细的结构细节以及显著的比例变化，这些图像提出了独特的超分辨率挑战。

如表 III所示，尽管没有在这个数据集上进行显式训练，我们的模型仍达到了与完全训练方法相当的性能。值得注意的是，ResQu 在保持竞争优势 FID 分数的同时，其 SSIM 和 PSNR 得分与其专门模型非常接近。这展示了它有效推广到未见过领域的的能力，突出了在处理领域特定场景时无需重新训练的强大性。

改进在复杂海事结构的重建中尤为明显，如船舶索

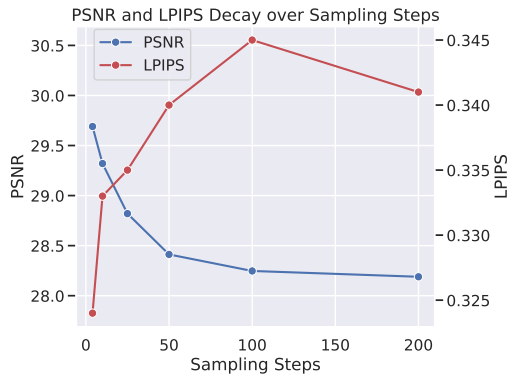

图 4. 采样步骤数量对关键评估指标的影响。减少步骤可以提高效率同时保持性能，突出了我们模型内在的速度优势。

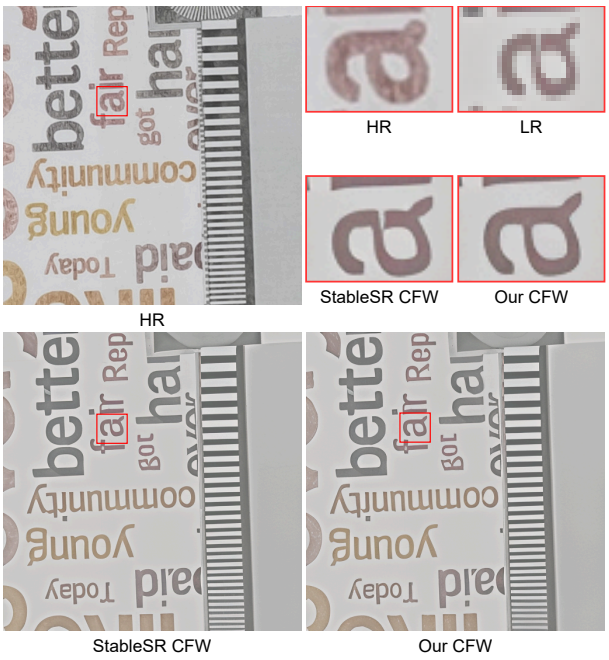

图 5. 可视化比较由改变 CFW 生成的多种超分辨率结果。这些图像展示了我们增强的编码器对 CFW 的影响，产生了更多结构化的细节。

具、天线阵列和甲板设备等传统方法常常难以处理的地方。我们的四元数表示捕捉了重要的相位关系，确保了更锐利的边缘和增强的纹理保真度，这对海事图像至关重要。此外，多尺度条件机制使我们的模型能够持续恢复大尺度结构元素（如船体细节）和细粒度特征（如导航设备和水面纹理）。

这些发现验证了我们方法的通用性，证明它不仅在其主要基准数据集上实现了最先进的结果，还能够有效地扩展到零样本设置下的特定领域应用。

表 II
使用我们 ResQu 的不同版本 CFW 获得的度量。

数据集	CFW	峰值信噪比 ↑	结构相似性指数 ↑	FID ↓
DIV2K	StableSR	23.34	0.562	**24.52**
	Ours	**23.61**	**0.591**	25.42
真实 SR	StableSR	23.26	0.702	**129**
	Ours	**24.98**	**0.714**	134
DrealSR	StableSR	27.94	0.734	151
	Ours	**28.18**	**0.755**	**151**

表 III
在 Shipspotting [13] 数据集上获得的 FID 结果。

	Model	SSIM ↑	PSNR ↑	FID ↓
完整训练	SR3 [10]	0.569	21.07	16.59
	StableSR [1]	0.554	20.29	<u>12.41</u>
	StableShip-SR [13]	0.561	20.51	**11.72**
Zero-Shot	ResQu	**0.611**	**22.01**	14.63

V. 结论

在本文中，我们介绍了 ResQu 这一新颖的超分辨率框架，该框架将四元数小波表示与潜在扩散模型相结合，利用我们的编码器来增强多尺度条件。我们的方法通过有效保持细粒度细节并维持高感知质量，在图像超分辨率领域建立了新的基准。通过对多种数据集进行广泛实验，我们的方法在多个定量和定性指标上始终优于现有技术。值得注意的是，我们的框架在处理复杂纹理和现实世界的退化模式方面表现出鲁棒性，这是图像超分辨率中的一个关键挑战。此外，我们进行了消融研究和跨数据集分析，以评估我们方法的泛化能力。结果突出了四元数小波嵌入在捕捉结构信息和纹理信息方面的有效性。总的来说，这项工作通过引入结合四元数小波特征、生成先验和扩散模型的新架构设计，推动了超分辨率领域的研究进展。

参考文献

[1] J. Wang, Z. Yue, S. Zhou, K. C. Chan, and C. C. Loy, "Exploiting diffusion prior for real-world image super-resolution," *International Journal of Computer Vision*, pp. 1–21, 2024.

[2] D. C. Lepcha, B. Goyal, A. Dogra, and V. Goyal, "Image super-resolution: A comprehensive review, recent trends, challenges and applications," *Information Fusion*, vol. 91, pp. 230–260, 2023.

[3] L. Sigillo, R. Giamba, and D. Comminiello, "Metadata, wavelet, and time aware diffusion models for satellite image super resolution," in *ICLR 2025 Workshop on Machine Learning*

for Remote Sensing (ML4RS), 2025. [Online]. Available: https://ml-for-rs.github.io/iclr2025/camera_ready/papers/19.pdf

[4] J. Kim, J. Oh, and K. M. Lee, "Beyond image super-resolution for image recognition with task-driven perceptual loss," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 2651–2661.

[5] W.-Y. Hsu and P.-W. Jian, "Wavelet pyramid recurrent structure-preserving attention network for single image super-resolution," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 11, pp. 15 772–15 786, 2024.

[6] N. Ruiz, Y. Li, V. Jampani, Y. Pritch, M. Rubinstein, and K. Aberman, "Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2023, pp. 22 500–22 510.

[7] E. Lopez, L. Sigillo, F. Colonnese, M. Panella, and D. Comminiello, "Guess what i think: Streamlined EEG-to-image generation with latent diffusion models," in *ICASSP 2025 - IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2025, pp. 1–5.

[8] Y. Takagi and S. Nishimoto, "High-resolution image reconstruction with latent diffusion models from human brain activity," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2023, pp. 14 453–14 463.

[9] Y. Wang, X. Chen, X. Ma, S. Zhou, Z. Huang, Y. Wang, C. Yang, Y. He, J. Yu, P. Yang *et al.*, "Lavie: High-quality video generation with cascaded latent diffusion models," *International Journal of Computer Vision*, pp. 1–20, 2024.

[10] C. Saharia, J. Ho, W. Chan, T. Salimans, D. J. Fleet, and M. Norouzi, "Image super-resolution via iterative refinement," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 45, no. 4, 2023.

[11] S. Shang, Z. Shan, G. Liu, L. Wang, X. Wang, Z. Zhang, and J. Zhang, "Resdiff: Combining cnn and diffusion model for image super-resolution," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 8, 2024, pp. 8975–8983.

[12] S. Gao, X. Liu, B. Zeng, S. Xu, Y. Li, X. Luo, J. Liu, X. Zhen, and B. Zhang, "Implicit diffusion models for continuous super-resolution," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2023, pp. 10 021–10 030.

[13] L. Sigillo, R. F. Gramaccioni, A. Nicolosi, and D. Comminiello, "Ship in sight: Diffusion models for ship-image super resolution," in *2024 International Joint Conference on Neural Networks (IJCNN)*, 2024, pp. 1–8.

[14] R. Wu, T. Yang, L. Sun, Z. Zhang, S. Li, and L. Zhang, "Seesr: Towards semantics-aware real-world image super-resolution," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2024, pp. 25 456–25 467.

[15] Y. Huang, J. Huang, J. Liu, M. Yan, Y. Dong, J. Lyu, C. Chen, and S. Chen, "Wavedm: Wavelet-based diffusion models for image restoration," *IEEE Transactions on Multimedia*, 2024.

[16] H. Phung, Q. Dao, and A. Tran, "Wavelet diffusion models are fast and scalable image generators," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2023, pp. 10 199–10 208.

[17] L. Aloisi, L. Sigillo, A. Uncini, and D. Comminiello, "A wavelet diffusion gan for image super-resolution," 2024. [Online]. Available: https://arxiv.org/abs/2410.17966

[18] B. B. Moser, S. Frolov, F. Raue, S. Palacio, and A. Dengel, "Waving goodbye to low-res: A diffusion-wavelet approach for image super-resolution," in *2024 International Joint Conference on Neural Networks (IJCNN)*, 2024, pp. 1–8.

[19] L. Sigillo, E. Grassucci, A. Uncini, and D. Comminiello, "Generalizing medical image representations via quaternion wavelet networks," *Neurocomputing*, vol. 638, p. 130195, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925231225008677

[20] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," *IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2021.

[21] T. Dai, J. Cai, Y. Zhang, S.-T. Xia, and L. Zhang, "Second-order attention network for single image super-resolution," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 11 065–11 074.

[22] C. Dong, C. C. Loy, K. He, and X. 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*. Springer, 2014, pp. 184–199.

[23] ——, "Image super-resolution using deep convolutional networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 2, pp. 295–307, 2015.

[24] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14*. Springer, 2016, pp. 391–407.

[25] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. P. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. 105–114, 2016.

[26] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. Change Loy, "ESRGAN: Enhanced super-resolution generative adversarial networks," in *Proceedings of the European Conf. on computer vision (ECCV) workshops*, 2018, pp. 0–0.

[27] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. R. Fu, "Image super-resolution using very deep residual channel attention networks," in *European Conf. on Computer Vision*, 2018.

[28] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," in *Int. Conf. on Learning Representations*, 2021.

[29] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, and R. Timofte, "SwinIR: Image restoration using swin transformer," in *2021 IEEE/CVF Int. Conf. on Computer Vision Workshops (ICCVW)*, 2021.

[30] S. Menon, A. Damian, S. Hu, N. Ravi, and C. Rudin, "Pulse: Self-supervised photo upsampling via latent space exploration of generative models," *IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. 2434–2442, 2020.

[31] K. Zhang, J. Liang, L. Van Gool, and R. Timofte, "Designing a practical degradation model for deep blind image super-resolution," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 4791–4800.

[32] X. Wang, L. Xie, C. Dong, and Y. Shan, "Real-ESRGAN: Training real-world blind super-resolution with pure synthetic data," in *IEEE/CVF Int. Conf. on Computer Vision Workshops (ICCVW)*, 2021.

[33] J. Song, C. Meng, and S. Ermon, "Denoising diffusion implicit models," in *Int. Conf. on Learning Representations*, 2021.

[34] A. Brock, J. Donahue, and K. Simonyan, "Large scale GAN training for high fidelity natural image synthesis," in *Int. Conf. on Learning Representations*, 2019.

[35] X. Lin, J. He, Z. Chen, Z. Lyu, B. Dai, F. Yu, Y. Qiao, W. Ouyang, and C. Dong, "Diffbir: Toward blind image restoration with generative diffusion prior," in *European Conference on Computer Vision*. Springer, 2025, pp. 430–448.

[36] L. A. Barford, R. S. Fazzio, and D. R. Smith, *An introduction to wavelets*. Hewlett Packard, 1992.

[37] M. Vetterli and C. Herley, "Wavelets and filter banks: Theory and design," *IEEE transactions on signal processing*, 1992.

[38] W. L. Chan, H. Choi, and R. Baraniuk, "Quaternion wavelets for image analysis and processing," in *2004 International Conference on Image Processing, 2004. ICIP'04.*, vol. 5. IEEE, 2004, pp. 3057–3060.

[39] I. W. Selesnick, R. G. Baraniuk, and N. C. Kingsbury, "The dual-tree complex wavelet transform," *IEEE signal processing magazine*, vol. 22, no. 6, pp. 123–151, 2005.

[40] W. L. Chan, H. Choi, and R. G. Baraniuk, "Coherent multiscale image processing using dual-tree quaternion wavelets," *IEEE Transactions on Image Processing*, vol. 17, no. 7, pp. 1069–1082, 2008.

[41] Z. Zhancheng, L. Xiaoqing, X. Mengyu, W. Zhiwen, and L. Kai, "Medical image fusion based on quaternion wavelet transform," *Journal of Algorithms & Computational Technology*, vol. 14, 2020.

[42] W. L. Chan, H. Choi, and R. G. Baraniuk, "Coherent multiscale image processing using dual-tree quaternion wavelets," *IEEE Transactions on Image Processing*, vol. 17, no. 7, pp. 1069–1082, 2008.

[43] E. Grassucci, L. Sigillo, A. Uncini, and D. Comminiello, "GROUSE: A task and model agnostic wavelet- driven framework for medical imaging," *IEEE Signal Processing Letters*, vol. 30, pp. 1397–1401, 2023.

[44] W. L. Chan, H. Choi, and R. Baraniuk, "Quaternion wavelets for image analysis and processing," in *2004 International Conference on Image Processing, 2004. ICIP '04.*, vol. 5, 2004, pp. 3057–3060 Vol. 5.

[45] X. Wang, K. Yu, C. Dong, and C. Change Loy, "Recovering realistic texture in image super-resolution by deep spatial feature transform," in *IEEE/CVF Conf. on Computer Vision and Pattern Recognition*, 2018.

[46] J. Choi, J. Lee, C. Shin, S. Kim, H. Kim, and S. Yoon, "Perception prioritized training of diffusion models," in *IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2022.

[47] X. Ji, Y. Cao, Y. Tai, C. Wang, J. Li, and F. Huang, "Real-world super-resolution via kernel estimation and noise injection," in *proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 2020, pp. 466–467.

[48] C. Chen, X. Shi, Y. Qin, X. Li, X. Han, T. Yang, and S. Guo, "Real-world blind super-resolution via feature matching with implicit high-resolution priors," in *Proceedings of the 30th ACM International Conference on Multimedia*, 2022, pp. 1329–1338.

[49] Z. Yue, J. Wang, and C. C. Loy, "Resshift: Efficient diffusion model for image super-resolution by residual shifting," *Advances in Neural Information Processing Systems*, vol. 36, 2024.

[50] P. Wei, Z. Xie, H. Lu, Z. Zhan, Q. Ye, W. Zuo, and L. Lin, "Component divide-and-conquer for real-world image super-resolution," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VIII 16*. Springer, 2020, pp. 101–117.

[51] J. Cai, H. Zeng, H. Yong, Z. Cao, and L. Zhang, "Toward real-world single image super-resolution: A new benchmark and a new model," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 3086–3095.

[52] E. Agustsson and R. Timofte, "Ntire challenge on single image super-resolution: Dataset and study," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2017.

[53] R. Timofte, E. Agustsson, L. Van Gool, M.-H. Yang, and L. Zhang, "Ntire challenge on single image super-resolution: Methods and results," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2017.

[54] X. Wang, K. Yu, C. Dong, and C. C. Loy, "Recovering realistic texture in image super-resolution by deep spatial feature transform," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 606–615.