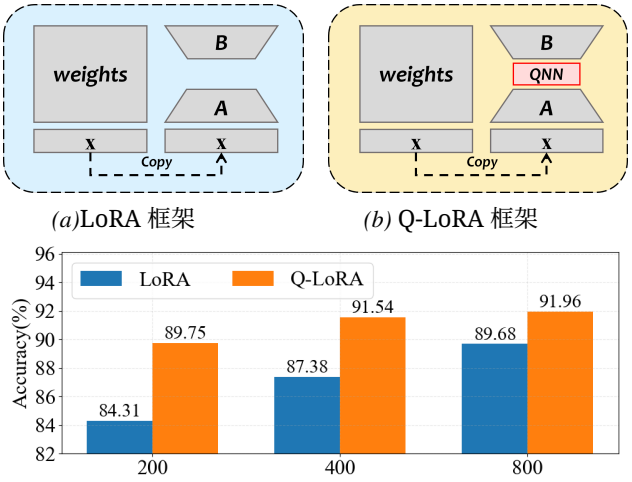
arXiv:2603.02281v1 [cs.LG] 2 Mar 2026

通过相位结构化重参数化实现少样本 AIGC 检测的量子启发式微调

Kaiyang Xing [1] **Han Fang** [2] **Zhaoyun Chen** [3] **Zhonghui Li** [1] **Yang Yang** [1 3] **Weiming Zhang** [4] **Guoping Guo** [4]

摘要

近期研究表明，量子神经网络（QNNs）在少样本场景中具有良好的泛化能力。为将这一优势扩展到大规模任务，我们提出 Q-LoRA，这是一种量子增强型微调方案，将轻量级量子神经网络集成到低秩适配（LoRA）适配器中。在人工智能生成内容（AIGC）检测任务中，Q-LoRA 在少样本设置下持续优于标准 LoRA。我们分析了这一性能提升的来源，并从量子神经网络中识别出两种可能的结构归纳偏置：（i）相位感知表示，通过正交的幅度–相位分量编码更丰富的信息；（ii）范数约束变换，借助内在的正交性稳定优化过程。然而，由于量子模拟，Q-LoRA 会带来不可忽视的额外开销。受此分析启发，我们进一步提出 H-LoRA，这是一种完全经典的变体，在 LoRA 适配器中应用希尔伯特变换，以保留类似的相位结构和约束。少样本 AIGC 检测实验表明，Q-LoRA 和 H-LoRA 的准确率均比标准 LoRA 高出 5% 以上；在该任务中，H-LoRA 以显著更低的成本实现了相当的准确率。



*(a)*LoRA 框架 *(b)* Q-LoRA 框架

(c) 不同训练样本数量下 LoRA 和 Q-LoRA 模型的检测准确率

图 *1.* 将 LoRA 和 Q-LoRA 应用于 CLIP 模型的方法，以及在不同训练样本下用于人工智能生成图像检测任务的检测准确率。

1. 引言

量子机器学习凭借量子系统独特的计算特性展现出巨大潜力。在其主要模型中，理论研究和实证结果均表明，量子神经网络（QNN）在少样本分类中具有较强的泛化能力场景（Caro 等人，2022）。这种优势通常归因于希尔伯特空间的几何结构以及量子操作的范数保持特性，两者共同限制了模型的有效容量，并减少了低数据量设置下的过拟合。然而，这些发现仍仅限于玩具规模的架构。这引出了一个自然的问题：**这种具有样本效率的归纳偏置能否扩展到大模型和现实世界任务?**

为探究量子泛化优势能否扩展至现实世界任务，我们提出了 Q-LoRA，这是一种受量子启发的微调方案，将轻量级 QNN 集成到低秩适配（LoRA）（Hu 等人，2022）框架中。LoRA 是一种广泛用于适配大型预训练模型的参数高效方法。具体而言，我们在 CLIP（Rad- ford 等人，2021）之上实现 Q-LoRA，冻结骨干网络，并将 QNN 注入 LoRA 瓶颈层。随后训练一个浅层分类头，以检测人工智能生成内容（AIGC）。

如图 1a 和 1b所示，架构上的修改微乎其微。然而，正如图1c所示，Q-LoRA

[1]安徽大学，中国合肥 [2]新加坡国立大学，新加坡，新加坡 [3]合肥综合性国家科学中心人工智能研究院，中国合肥 [4]中国科学技术大学，中国合肥。通信作者：Han Fang <fanghan@nus.edu.sg>，Yang Yang <skyyang@ahu.edu.cn>。

预印本。*2026*年*3*月*4*日。

在不同少样本规模下始终优于标准 LoRA，尤其在低数据场景中优势显著。这些结果表明，此前仅在小规模模型中观察到的量子神经网络泛化优势，可通过量子启发式设计有效迁移至大规模微调任务。

这一性能提升促使我们进一步探究其根本原因。然而，在实践中部署 Q-LoRA 会暴露出一个主要局限：量子模拟带来了巨大的计算开销。例如，在 200 个样本的训练设置下，Q-LoRA 每个训练周期需要超过 30 分钟，而标准 LoRA 只需几秒钟。这种显著的效率差距令人质疑基于量子神经网络的方法在可扩展微调中的实用性。

为了解决这一局限，同时保留 Q-LoRA 的优势，我们转而探寻更有依据的解释：究竟是什么使量子神经网络具有更强的泛化能力？我们推测，这些优势并非源自量子力学本身，而是来自量子神经网络的结构属性，这些属性能够引导特征提取与变换。具体而言，我们识别出两种关键的归纳偏置：（i）感知相位的表示，即信息编码在相互正交的幅度–相位分量中；（ii）范数约束变换，通过将更新限制在正交子空间内，促进更平滑的优化。这些洞见表明，我们或许能够在完全经典的设置中模拟 Q-LoRA 的优势，从而避免量子模拟的成本，同时保留其性能提升。

受上述分析启发，我们提出 H-LoRA，这是一种完全基于经典计算的变体，无需依赖量子计算，同时保留了量子神经网络的核心归纳偏置。具体而言，H-LoRA 在 LoRA 内对特征流应用希尔伯特变换，生成解析信号，将实部和虚部耦合为正交的振幅—相位表示。这一操作自然地涵盖了前文所指出的两项结构优势：（i）通过加入正交分量丰富特征，引入感知相位的编码；（ii）通过维持固定的正交耦合，施加范数约束变换，从而规范优化几何。实验表明，在少样本人工智能生成内容检测中，H-LoRA 的表现始终与 Q-LoRA 持平或更优，同时降低了时间消耗。

本文的核心贡献如下：

- 我们通过引入 Q-LoRA 对量子神经网络在大规模少样本任务中的泛化优势进行了实证验证。Q-LoRA 将轻量级量子神经网络集成到 LoRA微调框架中，并在人工智能生成内容检测任务上取得了稳定提升。

- 我们深入分析了这一优势的来源，并提炼出两种结构归纳偏置：相位感知编码和范数约束变换。这些偏置启发我们借助希尔伯特变换构建经典代理模型 H-LoRA。

- 大量实验表明，H-LoRA 与 Q-LoRA 均优于 LoRA。H-LoRA 在降低时间消耗的同时，性能达到或超过 Q-LoRA。这也验证了相位结构化归纳偏置对大模型少样本适应的有效性。

2. 相关工作

2.1. 低秩适配

大型预训练模型的快速扩展推动了对参数高效微调方法的广泛研究，这些方法旨在接近全量微调的性能，同时大幅降低计算与存储成本。在这些方法中，LoRA 基于任务特定更新通常位于低秩子空间这一观察，冻结骨干网络并优化低秩更新矩阵。该设计显著减少了可训练参数的数量，并且通常能以极小的推理开销达到甚至超越全量微调的性能。

尽管标准 LoRA 效果显著，但其仍可能存在参数利用不充分和泛化能力有限的问题，尤其是在低数据场景中。若干扩展方法通过改进秩分配或参数效率来应对这些局限，例如 AdaLoRA 会根据参数重要性动态重新分配秩预算（Zhang 等人，2023年），以及 NoRM，该方法会从已训练的 LoRA 适配器中移除冗余组件（Jiang 等人，2025）。最近，Q-LoRA 通过量子—经典混合架构将量子启发的归纳偏置引入 LoRA 框架，从而在少样本设置中实现了更好的泛化（Kong 等人，2025）。

2.2. 量子—经典混合神经网络

为探索量子计算在机器学习中的潜力并克服当前量子硬件的局限，研究人员提出了量子-经典混合模型。其核心思想是，将对量子特性敏感或能从高维希尔伯特空间处理中受益的部分计算任务交由量子单元或模拟器处理，其余部分则由高效的经典计算机完成。将量子电路作为神经网络中的可训练层进行集成的混合架构，在医疗保健、人工智能安全和图像处理等领域展现出潜力（Gordienko 等人，2025）。QNN 与经典网络的常见集成方法可分为两类：

一项具有代表性的研究（Dutta 等人，2025）将参数化量子电路作为一层融入经典全连接神经网络，在复杂回归和预测任务中表现优于纯经典模型。类似方法，如（Pandey 与 Rudra）和（Lin 等人，2025）中的方法，将 QNN 集成为深度伪造音频检测的中间层，取得了卓越的性能。本文提出的 Q-LoRA 采用了相同的方法。

另一类工作将量子神经网络用作最终决策层，置于用于特征提取与融合的经典网络之后。量子神经网络随后根据经典网络的输出给出最终预测。例如，（Xu etal.，2026）将量子神经网络用作分类头，在人工智能生成图像检测中取得了优于经典算法的性能。

2.3. 人工智能生成内容检测

随着生成模型的快速发展，合成内容与真实数据越来越难以区分，因而更容易在虚假信息传播和身份冒用等高风险场景中被滥用。人工智能生成内容检测旨在判断图像、音频及其他媒体是否由生成模型合成或篡改。该问题通常被表述为二元分类任务，其关键目标是捕捉生成过程留下的残留痕迹。对于图像，检测器通常关注异常纹理、变换域（Zhanget al.，2025；Karageorgiou et al.，2025）以及颜色分布（Jia etal.，2025）；对于音频，检测器通常依据微频率变化（Kumari et al.，2025）和韵律模式等线索进行判别。

3. 预备知识

3.1. LoRA实现

LoRA 的核心思想是通过低秩分解实现参数高效更新。具体而言，LoRA 不会更新整个权重矩阵 $W_0 \in \mathbb{R}^{(d\times k)}$，而是将其冻结，并注入一对可训练低秩矩阵 $B \in \mathbb{R}^{(d\times r)}$ 和 $A \in \mathbb{R}^{(r\times k)}$，其中秩为$r \ll min(d,\ k)$。有效的权重更新表示为 $\Delta W = B \cdot A$。输入特征 x 首先被投影到一个低维子空间：

$$y = xB^{\mathsf{T}}. \tag{1}$$

随后通过以下方式将其投影回原始空间：

$$z = yA^{\mathsf{T}}. \tag{2}$$

最后，将LoRA分支输出按因子 α/r缩放，并加到冻结骨干网络输出上：

$$h = W_0x + (\alpha/r)\cdot z. \tag{3}$$

这种方法在保持模型表达能力的同时，显著减少了可训练参数的数量，因此适用于大规模适配场景。

3.2. 希尔伯特变换

设 $s(t) \in \mathbb{R}$ 表示一个实值信号，其中 t表示连续索引，例如时间（一维信号）或经过适当向量化后的空间坐标。在神经网络中，$s(t)$可视为沿特定维度的实值特征激活。希尔伯特变换（HT）是一种线性算子 $\mathcal{H}[\cdot]$，它将一个实值信号映射为另一个实值信号，并使每个频率分量产生 90° 相移。其形式化定义为卷积：

$$\mathcal{H}[s(t)] = \frac{1}{\pi}\,\text{P.V.}\int_{-\infty}^{\infty}\frac{s(\tau)}{t-\tau}\,d\tau, \tag{4}$$

其中，P.V. 表示柯西主值。在频域中，希尔伯特变换有一种等价且直观的解释：它保留 $s(t)$的幅度谱，同时将正频率的相位偏移 $-\frac{\pi}{2}$，将负频率的相位偏移 $+\frac{\pi}{2}$。利用希尔伯特变换，可以构造解析信号：

$$s_a(t) = s(t) + j\cdot\mathcal{H}[s(t)], \tag{5}$$

它将原始实值信号嵌入复值表示中。该解析信号可进行显式的幅相分解：

$$A(t) = |s_a(t)|, \quad \Phi(t) = \arg(s_a(t)), \tag{6}$$

其中，$A(t)$表示瞬时幅度，$\Phi(t)$表示瞬时相位。重要的是，$s_a(t)$的实部和虚部构成一对正交对：它们彼此正交，共同编码关于底层信号的互补信息。

在一维信号处理中，希尔伯特变换（HT）被广泛用于非平稳信号的包络检测和基于相位的分析，并应用于语音、雷达和故障诊断。近期研究利用基于 HT 的相位表示来检测合成或篡改音频，证明了其在法证场景中的有效性（Shi 等人，2025；Unoki 等人，2024年）。在二维图像处理中，HT 及其扩展被应用于空间特征图，以实现稳健的边缘检测、纹理提取和相位一致性分析；已知该分析对光照变化不敏感（Yang 等人，2018a；b）。基于 HT 的表示在用于三维形状重建的干涉条纹分析中同样具有基础性作用，凸显了其从复杂视觉模式中捕获细粒度结构信息的能力（Guerra-Rosas 等人，2023年）。

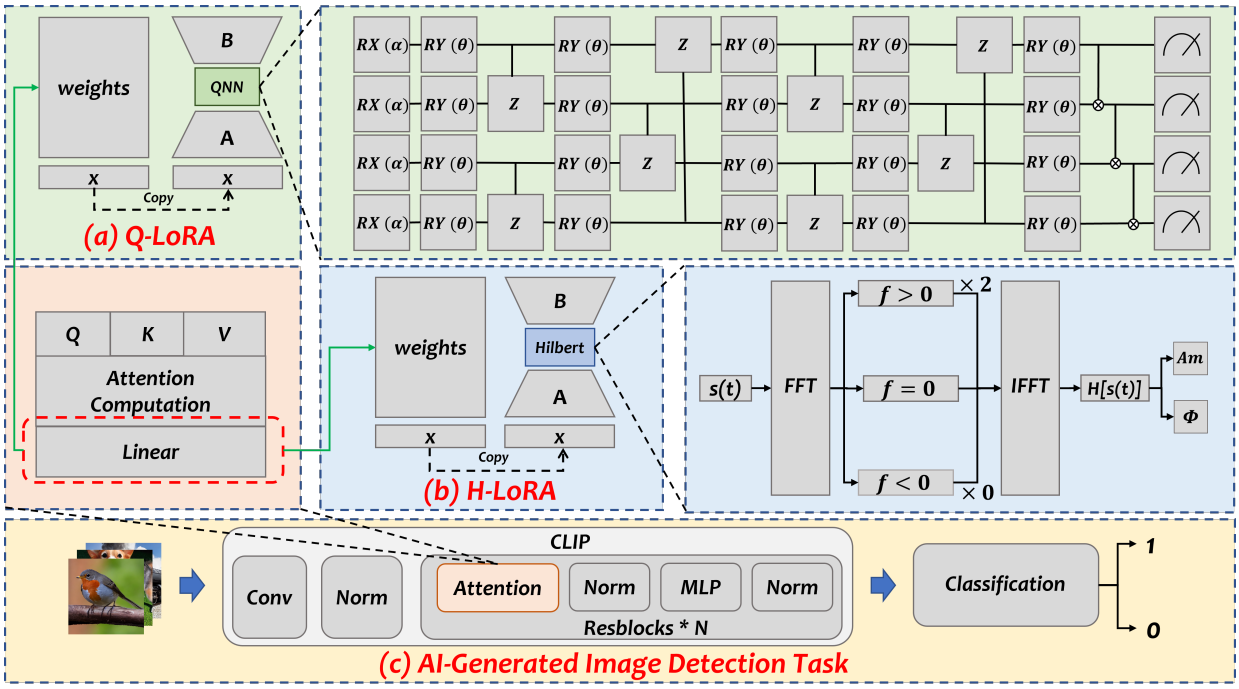


图 2. 人工智能生成图像检测框架：(a) Q-LoRA：在 CLIP 特征之后连接一个量子神经网络，通过量子态编码和处理提取高维特征，并通过旁路 LoRA 模块将其与骨干网络特征融合；(b) H-LoRA：对 CLIP 提取的特征进行希尔伯特变换，以分离并增强信号的振幅和相位特征，从而丰富模型表征，最后通过旁路 LoRA 模块与骨干网络特征融合，其中 FFT 和 IFFT 分别表示快速傅里叶变换和快速傅里叶逆变换；(c) 整体框架：展示了用于检测人工智能生成图像的系统流程，该流程基于 CLIP 视觉编码器，并整合了多种高效微调适配器。

4. 方法

4.1. 概述

为探究量子神经网络在少样本设置中的经验优势是否源于其结构归纳偏置，即相位感知表示和范数约束变换，我们采用少样本 AIGC 检测作为代表性基准。

如图 2所示，我们的框架以冻结的 CLIP 视觉编码器为基础进行特征提取。AIGC 检测任务被表述为一个二元分类问题。为在数据有限的情况下实现高效适配，我们采用 LoRA 模块形式的参数高效适配器。

具体而言，输入图像首先由 CLIP 编码为语义特征。LoRA 适配器被插入选定的 Transformer 层，仅学习低秩更新。在 LoRA 瓶颈层中，我们引入两种用于结构化归纳增强的替代设计：**Q-LoRA** 和 **H-LoRA**。Q-LoRA 通过轻量级量子神经网络增强低秩子空间，并借助量子电路注入感知相位且受范数约束的变换。H-LoRA 保留相同的集成策略，但将量子模块替换为经典希尔伯特变换流程，明确分离特征空间中的振幅和相位分量。随后，这些增强特征被投影回原始空间，并与骨干网络激活融合，再由轻量级分类器进行处理。训练期间仅更新适配器和分类器参数。

4.2. Q-LoRA

Q-LoRA 将量子神经网络模块集成到 LoRA 适配器结构中，形成量子-经典混合微调方案。该量子神经网络采用四量子比特架构，如图2所示。输入特征首先通过 RX 门编码为量子态。随后是两个纠缠层，每一层由以下部分组成：

- 单量子比特 RY 旋转，
- 将受控 Z（CZ）门作用于量子比特对（1–2、3–4）和（2–3、4–1），以实现循环纠缠。

最后，再应用一层单量子比特 RY 旋转，随后连接一串 CNOT 门，从而在测量所有量子比特之前引入全局纠缠

使用泡利-Z算符生成量子增强特征向量。

该架构能够实现：

• **相位感知表示**：信息同时编码在正交的幅度和相位分量中。

• **范数约束变换**：量子门具有酉性，可对优化轨迹进行正则化。

因此，Q-LoRA 通过模块化纠缠设计，在保持参数效率的同时，提升了特征表达能力和学习稳定性。

4.3. H-LoRA

我们认为，量子神经网络的核心优势可能源于两种关键的结构归纳偏置：（i）相位感知表示，将信息编码在正交的幅度–相位分量中，以增强特征空间的表达能力；（ii）范数约束结构化变换，通过保几何变换对优化过程进行正则化。为此，我们提出 H-LoRA，这是一种经典的基于希尔伯特变换的模块，通过解析信号构造来模拟受量子启发的信息编码。

H-LoRA层作为即插即用模块嵌入微调流程中。它处理来自上游层的低秩特征投影，并按以下步骤处理信号：

1) 低秩投影。 首先通过低秩矩阵 B 对输入特征 x 进行投影，得到低维表示 x_l:

$$x_l = Bx. \tag{7}$$

2) 解析信号构造。 为了模拟形式为 $\psi(t) = A(t)e^{j\phi(t)}$的量子波函数，我们通过希尔伯特变换构造 x_l 的解析信号：

$$x_a[n] = x_l[n] + j \cdot \mathcal{H}(x_l[n]), \tag{8}$$

其中， $\mathcal{H}(x_l[n])$为离散希尔伯特变换。由此得到的信号可进行幅相分解：

$$A[n] = |x_a[n]|, \quad \Phi[n] = \arg(x_a[n]). \tag{9}$$

这两个正交分量表示信号的能量和相位动力学，类似于量子相位–振幅表示。

3)相位感知增强。 我们将原始投影特征 x_l、振幅 $A[n]$ 和相位 $\Phi[n]$ 组合，形成增强特征：

$$x_{\text{enhanced}} = x_l + A[n] + \Phi[n]. \tag{10}$$

这种组合对实部、振幅和相位信息之间的相互作用进行建模，模拟量子系统中的复合测量过程。

4) 范数约束结构变换。 尽管H-LoRA不采用酉矩阵，但由于 x_l、 $A[n]$和 $\Phi[n]$之间的几何耦合，其增强特征位于受约束子空间中。具体而言， $A[n]$和 $\Phi[n]$ 由 x_l 推导而来，因此具有有界范数和相关性：

$$\|x_{\text{enhanced}}\| \leq \|x_l\| + \|A[n]\| + \|\Phi[n]\|.$$

这意味着一个范数约束变换空间，它会产生类似于满足 $U^\dagger U = I$ 的量子酉算子的正则化效果。

5）重新投影与融合。最后，使用第二个低秩矩阵 A 将增强特征投影回原始特征维度，并由可学习的缩放因子 α 调节：

$$\Delta x = \alpha \cdot A x_{\text{enhanced}}, \tag{11}$$

$$x_{\text{output}} = x + \Delta x. \tag{12}$$

由此完成了具有希尔伯特空间量子类比归纳偏置的 LoRA 式调整。

5. 实验

5.1. 设置

所有实验均在同一服务器平台上进行，该平台配备一颗 Intel(R) Xeon(R) Platinum 8369B CPU@2.90GHz 中央处理器，以及一块配有 24GB 显存的 NVIDIA GeForce RTX 4090 GPU 作为图形处理器。

图像检测任务。我们使用 CLIP 作为骨干网络。除非另有说明，我们将 LoRA 适配器注入最后 6 个 Transformer 块，秩为 $r = 4$ 和 $\alpha = 1$，并使用单个线性层作为分类器头。我们最多训练 20 轮，批量大小为 16，学习率为 5×10^{-4}。输入分辨率为 224×224。数据来自 AIGCDetectionBenchMark：训练时，我们从 StableDiffusion v1.4 子集中随机采样。测试时，我们从每个生成器类别中随机选取 100 张图像：Midjourney、Wukong、VQDM、Stable Diffusion v1.4 (SD1.4)、Stable Diffusion v1.5 (SD1.5)、Glide 和 ADM。我们考察 200、400 和

800 个样本的少样本训练规模，其中每个数字表示训练样本总数，包括真实图像和人工智能生成图像。为评估数据稀缺情况下的稳定性，我们使用种子 $\{0, 1, ..., 9\}$ 进行 10 次独立试验。每次试验中，少样本训练集都重新随机采样，并报告各次试验的平均值。

音频检测任务。此外，我们在 ASVspoof 2019 LA（Todisco 等人，2019）上进行少样本音频伪造检测，以考察跨模态泛化能力。Whisper（Radford 等人，2023年）用作骨干网络。我们将 LoRA 适配器注入最后 3 个编码器层，秩为$r = 4$，且 $\alpha = 1$，并使用单个线性层作为分类器头。我们训练最多 20 轮，批量大小为 16，学习率为 2×10^{-4}。音频重采样至 16 kHz，并裁剪或填充至 4 秒。由于我们关注少样本场景，因此通过随机抽样构建少样本划分：从完整样本池（超过 2.5 万个片段）中随机选取训练集和测试集，其中 1,500 个片段用于测试。与图像实验一样，我们使用种子 $\{0, 1, ..., 9\}$ 重复进行 10 次试验，并在每次试验中重新抽取少样本训练子集。

指标。 在少样本图像检测任务中，我们报告ROC曲线下面积（AUC）、准确率（ACC）、精确率（PR）和召回率（RE）；在少样本音频检测任务中，我们报告AUC、F1分数（F1）、准确率（ACC）和等错误率（EER）。

5.2. 少样本图像检测任务的结果

定量分析。 本节通过实验评估这些方法在人工智能生成图像检测的不同少样本设置下的表现。具体而言，我们采用（i）仅微调分类头的 CLIP，（ii）标准 LoRA，（iii）Q-LoRA，以及（iv）我们的经典模拟方法 H-LoRA。

表1 总结了在 SD1.4 上训练、并在少样本设置下于多种生成器上测试时的跨生成器泛化能力。在样本数最受限的情况下（200 个样本），Q-LoRA 和 H-LoRA 的表现始终优于标准 LoRA；Q-LoRA 的平均准确率达到 89.75%（比 LoRA 高 5.44%），H-LoRA 达到 89.94%（比 LoRA 高 5.63%）。随着训练样本数增加到 400 和 800，各方法的表现均有所提升，彼此间的差距也逐渐缩小。

在大多数测试生成器上，Q-LoRA 和 H-LoRA 的改进并非由某一个类别单独驱动，而是在与训练分布差异较大的生成器上更为显著；在这些生成器上，标准 LoRA 更容易过拟合 SD1.4 特有的伪影。表 1 中的未加权平均值表明，在不同训练规模和指标下，H-LoRA 的表现始终与 Q-LoRA 相当，而且两者均优于标准 LoRA。无阈值的曲线下面积和依赖阈值的准确率同时提升，表明这些改进反映了真实内容与生成内容的表征更易区分，而非校准效应。这些结果表明，H-LoRA 通过纯经典实现复现了 Q-LoRA 的大部分优势。

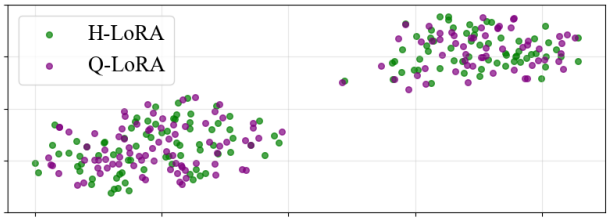
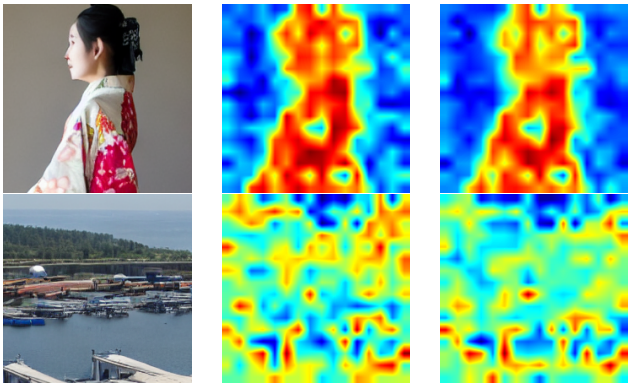


图*3.* Q-LoRA 和 H-LoRA 的 t-SNE 可视化



(a) 原始图像 *(b)* Q-LoRA *(c)* H-LoRA

图*4.* Q-LoRA 和 H-LoRA 的注意力可视化

可视化分析。 除了汇总指标之外，我们还展示了 H-LoRA 和 Q-LoRA 提取特征相似性的可视化结果。图3 展示了学习特征的 t-SNE 投影，其中 H-LoRA 和 Q-LoRA 生成的嵌入在投影空间中高度重叠，这表明在 t-SNE 投影下，两种方法产生的特征分布总体相似。图4 进一步展示了代表性样例的注意力图；H-LoRA 突显的空间模式与 Q-LoRA 的高度一致，说明两种适配方法关注的区域相近，且这些区域对检测具有信息价值。总体而言，这些可视化结果补充了定量结果，表明所提出的经典替代方法能够在很大程度上复现与量子启发式变体相关的关键表征特征。

5.3. 少样本音频检测任务的结果

除了图像检测之外，我们还进行了生成音频检测实验。我们比较了

表*1.* 人工智能生成图像检测在不同少样本设置下的性能比较。对于每个指标，同一训练样本量内的最佳和次佳结果分别以粗体和下划线标出。

模型	指标	CLIP			LoRA			Q-LoRA			H-LoRA		
样本		200	400	800	200	400	800	200	400	800	200	400	800
Midjourney	曲线下面积(%)	98.38	98.29	98.38	94.02	95.78	97.59	**98.62**	**98.74**	**98.68**	96.31	97.80	97.91
	准确率(%)	**93.40**	**92.85**	**91.85**	82.60	80.55	78.70	90.40	86.15	81.75	86.10	81.90	84.85
	精确率(%)	**93.86**	**90.96**	**88.56**	77.80	73.42	71.37	87.62	79.93	73.45	80.37	75.08	77.95
	召回率(%)	93.00	95.50	96.50	93.60	97.40	98.20	96.00	**98.60**	**99.50**	**96.40**	98.40	98.70
WUKONG	曲线下面积(%)	97.31	99.00	99.27	97.22	99.34	99.85	**99.43**	99.83	**99.97**	99.18	**99.84**	99.94
	准确率(%)	92.70	94.95	95.90	91.00	95.30	97.80	95.00	97.90	**99.12**	**95.80**	**98.40**	98.35
	精确率(%)	93.05	94.73	95.56	88.33	93.62	96.81	94.86	97.60	**98.76**	**95.02**	**97.86**	98.15
	召回率(%)	92.30	95.30	96.30	94.90	97.50	98.90	95.40	98.30	**99.50**	**96.80**	**99.00**	98.60
SD1.4	曲线下面积(%)	98.51	99.38	99.75	98.10	99.56	99.93	99.67	**99.99**	**99.99**	**99.67**	99.84	99.92
	准确率(%)	94.40	96.75	97.60	93.25	96.60	98.60	96.50	**99.10**	**99.75**	**97.20**	98.75	98.70
	精确率(%)	96.28	98.26	98.89	93.35	95.91	99.10	97.48	99.01	**99.75**	**97.93**	**99.30**	99.51
	召回率(%)	92.40	95.20	96.30	93.50	97.60	98.10	95.60	**99.20**	**99.70**	**96.50**	98.20	97.90
SD1.5	曲线下面积(%)	98.74	99.32	99.72	98.61	99.54	99.86	**99.80**	**99.98**	99.97	99.53	99.93	**99.97**
	准确率(%)	94.70	96.35	97.25	93.55	96.75	98.05	**97.00**	**99.15**	**99.37**	96.90	98.65	98.80
	精确率(%)	96.59	97.65	97.90	93.54	96.46	98.25	**98.80**	**99.41**	**99.75**	98.20	99.19	99.60
	召回率(%)	92.70	95.00	96.60	93.80	97.40	97.90	95.20	**98.90**	**99.00**	**95.60**	98.10	98.00
GLIDE	曲线下面积(%)	94.35	97.09	98.09	92.15	96.78	99.14	**97.60**	**99.41**	**99.64**	97.59	98.99	99.48
	准确率(%)	85.20	89.00	90.15	80.50	86.65	91.30	89.70	**92.50**	**94.87**	**90.75**	91.95	93.35
	精确率(%)	81.56	85.32	86.05	74.92	81.10	86.42	87.18	88.39	**91.27**	**87.03**	**87.92**	89.42
	召回率(%)	91.30	94.60	96.10	92.70	97.50	98.60	93.60	98.60	**99.50**	**96.40**	**99.30**	99.00
VQDM	曲线下面积(%)	91.75	95.24	96.67	91.96	96.20	98.65	95.90	**99.16**	99.38	**96.61**	98.65	**99.92**
	准确率(%)	81.50	84.75	86.65	81.15	84.95	88.80	85.10	90.00	90.75	**87.75**	**91.50**	**91.95**
	精确率(%)	76.15	78.84	86.91	76.31	79.23	82.96	79.56	84.76	85.15	**82.73**	**87.17**	**88.12**
	召回率(%)	91.90	95.20	96.10	93.20	96.50	98.20	95.60	98.30	**99.00**	**96.40**	**98.30**	97.70
ADM	曲线下面积(%)	87.22	91.91	93.81	80.03	87.71	94.96	**91.33**	**96.80**	**97.52**	90.96	95.34	95.42
	准确率(%)	73.45	77.45	**78.90**	68.15	70.85	74.50	74.50	76.00	78.12	**75.05**	**80.35**	78.75
	精确率(%)	67.28	70.11	**71.38**	62.50	64.22	67.16	67.82	68.64	69.80	**67.97**	**73.76**	71.09
	召回率(%)	91.60	96.10	96.80	94.40	97.00	98.10	95.20	**98.90**	**99.30**	**96.70**	98.40	99.00
平均值	曲线下面积(%		97.17	97.96	93.15	96.42	98.57	**97.48**	**99.13**	**99.31**	97.12	98.63	98.74
	准确率(%)	87.91	90.30	91.19	84.31	87.38	89.68	89.75	91.54	91.96	**89.94**	**91.64**	**92.11**
	精确率(%)	86.40	87.98	88.64	80.96	83.42	86.01	**87.61**	88.25	88.28	87.03	**88.61**	**89.12**
	召回率(%)	92.17	95.27	96.39	93.76	97.27	98.10	95.29	**98.69**	**99.38**	**96.40**	98.53	98.44

仅微调分类头、(ii) 标准 LoRA，以及 (iii) 我们的经典模拟方法 H-LoRA，在不同少样本设置下进行音频伪造检测。如表2所示，在随机少样本划分下，H-LoRA 在 50、100 和 200 个样本的设置中，始终优于采用线性分类器探测的 Whisper 和标准 LoRA。其提升不仅体现在无阈值指标曲线下面积（AUC）达到最佳值（99.43%），也体现在阈值相关指标达到最高值，例如 F1分数（96.73%）和准确率（96.77%）；同时，EER 更低（0.0333）。EER 是 ASVspoof 风格反欺骗评估中的一项主要指标。此外，在 50 个样本的设置下，H-LoRA 的准确率达到 90.69%，显著高于 LoRA 的 72.99%。这些趋势表明，所提出的相位结构化重参数化在图像领域之外提供了有用的归纳偏置。

表*2.* 音频伪造检测中不同样本量下的算法性能

方法	样本	曲线下面积(%)	F1分数(%)	准确率(%)	EER
WHISPER	50	90.85	72.54	72.99	0.1616
	100	93.71	87.42	88.19	0.1260
	200	95.82	89.93	90.69	0.0957
LoRA	50	92.86	78.89	80.34	0.1358
	100	97.45	89.39	90.40	0.0711
	200	99.04	95.54	95.67	0.0431
H-LORA	50	96.52	88.32	89.33	0.0871
	100	98.57	94.37	94.58	0.0548
	200	**99.43**	**96.73**	**96.77**	**0.0333**

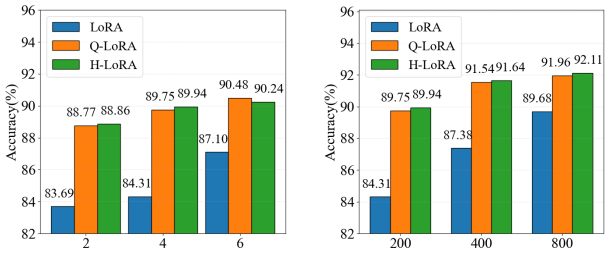
5.4. 效率与实用性分析

表 *3.* H-LoRA 与 Q-LoRA 的效率比较。

方法	H-LoRA	Q-LoRA
推理时间（秒）	**0.09**	65.68
训练时间（周期/秒）	**4.07**	2088.34
额外参数	**0**	24

在本节中，我们比较了 H-LoRA 与 Q-LoRA 在图像检测任务中的效率。如表 3所示，尽管检测性能相当，H-LoRA 在训练与推理效率方面均展现出压倒性优势。具体而言，H-LoRA 的推理仅需 0.09 秒，每个训练周期仅需 4.07 秒，而 Q-LoRA 的推理耗时 65.68 秒，每个周期超过 2,000 秒。此外，与 LoRA 相比，H-LoRA 不引入额外可训练参数，而 Q-LoRA 则增加了 24 个参数。这些结果清楚地表明，H-LoRA 有效保持了标准 LoRA 的效率，同时避免了显式量子神经网络所带来的巨大时间开销。

5.5. 消融实验



(a) 秩　　*(b)* 样本

图*5.* LoRA、Q-LoRA 和 H-LoRA 在不同秩和训练样本数量下的准确率

我们进行了消融实验，以更好地评估哪些因素促成了 H-LoRA 的性能提升。

秩的影响。 我们改变适配器秩 $r \in \{2, 4, 6\}$，同时保持训练方案的其余部分不变，并测试检测准确率。如图 5a所示，提高秩可以提升所有方法的性能，这反映出更高秩的适配具有更强的表达能力。不过，在所有秩设置下，H-LoRA 和 Q-LoRA 的表现始终优于标准 LoRA。在 $r = 2$处，相对优势尤为明显；这表明，即使参数预算受到严格限制，所提出的结构化变换仍然有益。

训练样本量的影响。 图中的消融结果5b 与第5.2节的观察结果一致：Q-LoRA 和 H-LoRA 在性能方面取得了最大提升，尤其是在数据最少的情形下；随着训练样本量增加，这一差距逐渐缩小，表明当可用监督信号更多时，对强归纳偏置的依赖会降低。

表*4.* LoRA 中不同类型中间层的性能

方法	H-LORA	线性	双曲正切	Sigmoid	SILU
曲线下面积(%)	**97.12**	96.32	96.51	96.89	97.01
准确率(%)	**89.94**	87.79	88.30	87.19	88.36
精确率(%)	**87.03**	85.57	85.70	84.20	85.91
召回率(%)	**96.40**	96.44	95.14	95.87	95.59

非线性的影响。 为检验性能提升是否仅仅源于在低秩特征上增加非线性，我们将 LoRA 路径中的中间变换替换为常见的激活函数或线性层，并对所得变体进行评估。表4 显示，尽管引入了额外的表达组件，这些通用的非线性或线性插入在准确率方面仍无法达到 H-LoRA 的性能，所有变体的数值均低于 89%。此外，表 5 显示，增加中间线性层的数量会带来递减收益：性能趋于在约 89.5% 附近达到平台期，且准确率不会随着层数增加而单调提升。综合来看，这些结果支持这样一种观点：性能提升并非仅仅归因于增加了非线性，而是与所提出的相位感知表示和范数约束结构化变换更为密切相关。

表 *5*。 LoRA 中间线性层数量的性能

方法 H-LORA		1	2	3	4	5
曲线下面积(%)	97.12	96.32	97.39	97.52	97.58	**97.61**
准确率(%)	**89.94**	87.79	88.90	89.60	89.10	89.30
精确率(%)	**87.03**	85.57	86.04	86.25	85.77	86.68
召回率(%)	96.40	96.44	96.41	**97.19**	97.17	96.23

6. 结论

在本研究中，我们提出了 H-LoRA，这是一种完全基于经典方法、但受量子启发的少样本学习微调方法。基于 Q-LoRA 的实证成功，我们分析了其性能背后的结构归纳偏置，具体而言，即相位感知表示和受范数约束的变换，并证明这些特性可以通过希尔伯特变换以经典方式复现。将这一发现融入 LoRA 适配器后，H-LoRA 在避免高昂模拟成本的同时，保留了 Q-LoRA 的泛化优势。少样本 AIGC 检测实验结果表明，H-LoRA 的准确率与 Q-LoRA 相当或更高，同时效率显著提升。

致谢

本研究部分受到中国国家自然科学基金（项目编号：62272003）、量子科学与技术国家科技重大专项（项目编号：2021ZD0302300）以及安徽省科技重大专项（项目编号：202423s06050001）的资助。

影响声明

本文旨在推动机器学习领域的发展。我们的工作可能对社会产生许多影响，但我们认为没有哪项影响需要在此特别强调。

参考文献

Caro, M. C., Huang, H.-Y., Cerezo, M., Sharma, K., Sornborger, A., Cincio, L., and Coles, P. J. 基于少量训练数据的量子机器学习泛化能力。《自然-通讯》, 13(1):4919, 2022.

Dutta, S. S., Sandeep, S., 等。混合量子神经网络：利用着装量子电路，通过地震数据融合提升海啸预测能力。《*EPJ*量子技术》, 12(1):4, 2025.

Gordienko, Y., Trochun, Y., Taran, V., Khmelnytskyi, A.与 Stirenko, S. HNN-QCn：以量子变换作为数据增强技术的多骨干网络混合神经网络。人工智能，6(2):36，2025.

´Guerra-Rosas, E., L´opez-Avila, L. F., Garza-Flores, E.,´Vidales-Basurto, C. A.与 Alvarez-Borrego, J. 利用人工智能方法，通过径向傅里叶–梅林变换与希尔伯特变换签名对皮肤病变图像进行分类。《应用科学》，13(20):11425，2023年.

Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W.等. LoRA：大语言模型的低秩适配。国际学习表征会议，1(2):3，2022.

Jia, Z., Huang, C., Zhu, Y., Fei, H., Duan, X., Yuan, Z., Deng, Y., Zhang, J., Zhang, J. 和 Zhou, J. 秘密藏于颜色之中：通过颜色分布分析增强人工智能生成图像检测。载于计算机视觉与模式识别会议论文集，第13445–13454 页，2025。

Jiang, S., Liao, Y., Zhang, Y., Wang, Y., 和 Wang, Y. 保留多数类以降噪的微调方法。发表于第十三届国际学习表征会议，2025。

Karageorgiou, D., Papadopoulos, S., Kompatsiaris, I., 和 Gavves, E. 基于谱学习的任意分辨率人工智能生成图像检测。发表于计算机视觉与模式识别会议论文集，第 18706– 18717 页，2025。

Kong, X., Li, L., Chen, Z., Xue, C., Xu, X., Liu, H., Wu, Y., Fang, Y., Fang, H., Chen, K. 等。量子增强的大语言模型高效微调。*arXiv* 预印本 *arXiv:2503.12790*，2025。

Kumari, K., Abbasihafshejani, M., Pegoraro, A., Rieger, P., Arshi, K., Jadliwala, M. 和 Sadeghi, A.-R. Voicer-adar：使用微频率和组合分析进行语音深度伪造检测。载于*NDSS*，2025。

Lin, C.-H. A.、Liu, C.-Y.、Chen, S. Y.-C. 和 Chen, K.-C. 基于量子训练卷积神经网络的深度伪造音频检测。载于*2025*年 *IEEE*声学、语音与信号处理国际会议研讨会（*ICASSPW*），第1–5页。IEEE，2025年。Pandey, A. 和 Rudra, B. 用于深度伪造音频检测的经典量子混合学习模型框架。载于第*11*届信息系统安全与隐私国际会议论文集，第2卷，第231–239页。Radford, A.、Kim, J. W.、Hallacy, C.、Ramesh, A.、Goh, G.、Agarwal, S.、Sastry, G.、Askell, A.、Mishkin, P.、Clark, J. 等。通过自然语言监督学习可迁移视觉模型。载于国际机器学习会议，第8748–8763页。PmLR，2021年。Radford, A.、Kim, J. W.、Xu, T.、Brockman, G.、McLeavey, C. 和 Sutskever, I. 通过大规模弱监督实现鲁棒语音识别。载于国际机器学习会议，第28492–28518页。P MLR，2023年。Shi, C.、Qiu, X.、Wu, M.、Niu, X.、Li, X. 和 Abdullahi, S. M. 结合递减卷积核神经网络与频谱图融合的音频复制移动篡改检测。《复杂与智能系统》，11(9):382，2025。Todisco, M.、Wang, X.、Vestman, V.、Sahidullah, M.、Delgado, H.、Nautsch, A.、Yamagishi, J.、Evans, N.、Kinnunen, T. 和 Lee, K. A. ASVspoof 2019：欺骗音频和虚假音频检测的未来方向。*arXiv*预印本 *arXiv:1904.05441*，2019年。Unoki, M.、Li, K.、Chaiwongyen, A.、Nguyen, Q.-H. 和 Zaman, K. 深度伪造语音检测：从与听觉感知相关的声学特征到深度神经网络的方法。*IEICE*信息与系统学报，2024年。Xu, J. C.、Fang, H.、Yang, Y. 等。基于经典量子混合神经网络的人工智能生成图像检测算法。中国科学：信息科学，69(1):112501，2026。

Yang, Y.、Kou, K. I. 和 Zou, C. 基于修正单分量信号微分相位一致性的边缘检测方法。多维系统与信号处理，29(1):339–359，2018年a。

Yang, Z.、Zhang, Q.、Zhou, F. 和 Yang, L. 分段平稳信号的希尔伯特谱分析及其在纹理分类中的应用。*Digital Signal Processing*，82：1–10，2018年b。

Zhang, H.、He, Q.、Bi, X.、Li, W.、Liu, B. 和 Xiao, B. 通过变分信息瓶颈网络实现通用人工智能生成图像检测。载于计算机视觉与模式识别会议论文集，第23828–23837页，2025。Zhang, Q.、Chen, M.、Bukharin, A.、Karampatziakis, N.、He, P.、Cheng, Y.、Chen, W. 和 Zhao, T. AdaloRA：面向参数高效微调的自适应预算分配。*arXiv* 预印本 *arXiv:2303.10512*，2023年。

A. 离散信号的希尔伯特变换

给定一个长度为 N的离散信号 $x(n)$，其中 $n = 0$，1，2，.....，$N - 1$，对其 N点离散傅里叶变换（DFT）进行计算，可得到其频域表示 $X(k)$，其中 $k = 0$，1，2，.....，$N - 1$:

$$X(k) = \sum_{n=0}^{N-1} x(n) \cdot \exp\left(-j \cdot \frac{2\pi kn}{N}\right)$$

接下来，将 $X(k)$ 的负频率分量（即 $k > \frac{N}{2}$ 的情况）的幅值取反，以获得对称的频域表示 $Y(k)$:

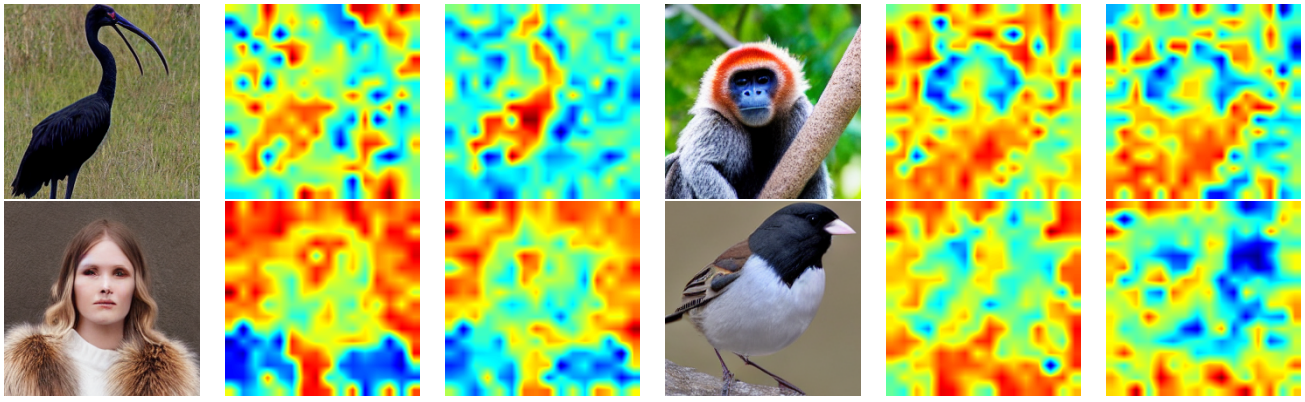
$$Y(k) = \begin{cases} X(k) & \text{for } k = 0, 1, \ldots, \frac{N}{2}, \\ -X(k) & \text{for } k = \frac{N}{2} + 1, \frac{N}{2} + 2, \ldots, N - 1. \end{cases}$$

最后，对对称的频域表示 $Y(k)$ 应用 N 点离散傅里叶变换的逆变换（IDFT），以获得希尔伯特变换后的离散信号 $h(n)$:

$$h(n) = \frac{1}{N} \sum_{k=0}^{N-1} Y(k) \cdot \exp\left(j \cdot \frac{2\pi kn}{N}\right)$$

由此得到原始离散信号 $x(n)$ 的希尔伯特变换。

B. 更多注意力可视化

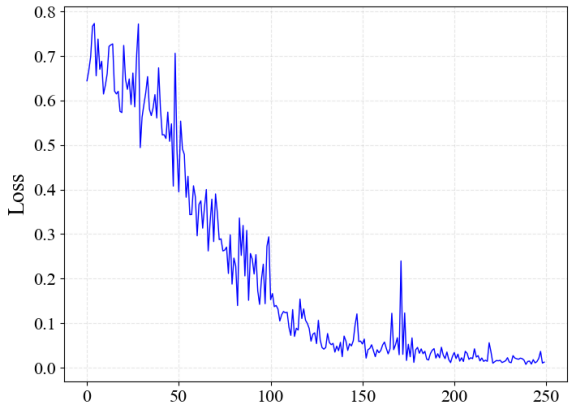


原始图像　Q-LoRA　H-LoRA　原始图像　Q-LoRA　H-LoRA

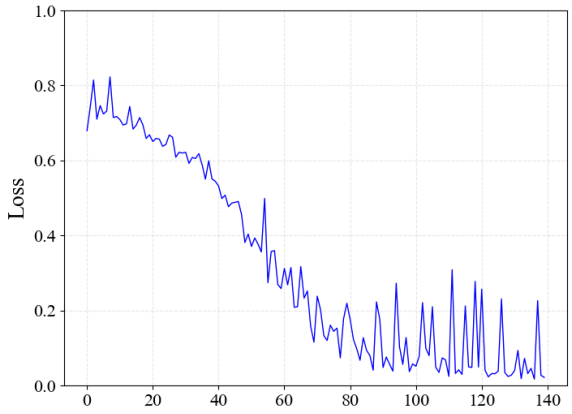
我们进一步进行了额外的可视化实验，以展示 H-LoRA 与 Q-LoRA 在特征捕获方面的相似性。左侧图像为原始图像，右侧图像为对应的注意力热图。可以看出，H-LoRA 和 Q-LoRA 都高度相似地关注图像中的关键区域（例如物体的显著部分）。这表明两种方法在特征捕获方面表现一致，进一步支持了两种可能的结构归纳偏置。

C. H-LoRA 的训练损失曲线

我们展示了图像实验和音频实验的训练损失曲线。如左图所示，图像实验收敛迅速，表明该模型易于训练。类似地，右图展示的音频实验也呈现出较为平稳且快速的收敛，进一步验证了 H-LoRA 在不同领域中易于训练。这些结果表明，即使在低数据场景下，H-LoRA 也能高效地学习任务。



(a) 图像检测任务中 H-LoRA 的训练损失



(b) 音频检测任务中 H-LoRA 的训练损失