arXiv:2304.00837v1 [cs.CV] 3 Apr 2023

无序不变隐式神经表示

Hao Zhu†、Shaowen Xie†、Zhen Liu†、Fengyi Liu、Qi Zhang、You Zhou、Yi Lin、Zhan Ma*和 Xun Cao

摘要——隐式神经表示（隐式神经表示）将信号的属性表征为相应坐标的函数，已成为求解逆问题的利器。然而，隐式神经表示的表达能力受到网络训练中频谱偏差的限制。本文发现，通过重新排列输入信号的坐标，可以在很大程度上解决此类频率相关问题。为此，我们提出了无序不变隐式神经表示（DINER），即在传统隐式神经表示骨干网络中增设一个哈希表。对于属性直方图相同但排列顺序不同的离散信号，哈希表能够将其坐标投影至相同的分布，使后续的隐式神经表示网络能够更好地对映射信号进行建模，从而显著缓解频谱偏差。此外，DINER 的表达能力由哈希表的宽度决定。不同的宽度对应于属性空间中不同的几何元素，例如，当宽度分别设为1、2 和3时，所对应的几何元素分别为一维曲线、二维曲面和三维曲体。几何元素覆盖的区域越大，表达能力就越强。实验不仅揭示了 DINER 对不同隐式神经表示骨干网络（多层感知机与 SIREN）及多种任务（图像/视频表示、相位恢复、折射率恢复和神经辐射场优化）的泛化能力，还表明其在质量和速度上均优于最先进算法。项目主页：https://ezio77.github.io/DINER-website/

索引词—隐式神经表示、无序不变性、逆问题优化、哈希表。

1 引言

INR [1] 利用神经网络建立输入坐标与信号对应属性之间的映射，无需存储额外样本 [2]，即可实现无需奈奎斯特采样的缩放、插值和外推。将其与可微物理机制相结合，例如光线步进渲染 [3], [4], 、菲涅耳衍射传播 [5] 和偏微分方程 [6], ，隐式神经表示

逐渐成为解决逆问题的通用而锐利的工具，并在各种科学任务中取得了显著进展，例如新视角合成 [7], 自由手持三维超声重建 [8], 强度衍射断层成像 [9] 以及多物理场仿真 [6]。

然而，隐式神经表示的表达能力通常受到其底层网络模型本身的限制。例如，频谱偏差 [10] 通常使隐式神经表示更容易表示低频信号分量 [11]。为了提升隐式神经表示模型的表达能力，先前的探索主要依赖于使用傅里叶基 [1], [3], [11], [12]或小波基 [13], [14]将更多频率基编码到网络中。然而，从理论上讲，函数展开的长度是无限的，而包含更多频率基的更大模型运行速度极慢。

这类问题与输入信号的频谱分布密切相关。信号的频率表示信号属性沿几何坐标的内在顺序变化的快慢。通过适当重排

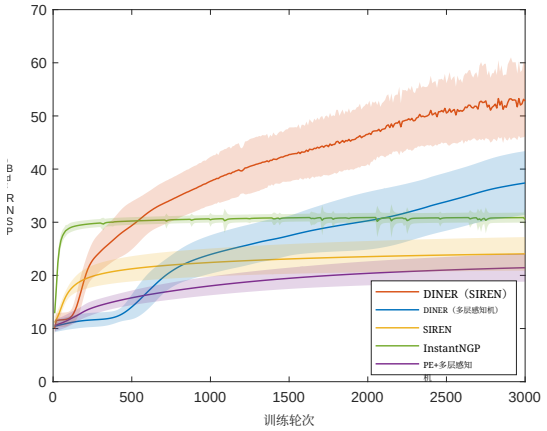


图1. 不同训练轮次下，二维图像拟合中各类隐式神经表示的峰值信噪比。

信号坐标的顺序，我们可以调制其频谱，使其包含更多低频成分，从而使重排后的信号更适合由后续的隐式神经表示建模。因此，我们提出了 DINER：先通过哈希表将输入坐标映射到另一个索引，再将其输入经典的隐式神经表示骨干网络。

我们证明，无论信号属性最初以何种几何排列顺序呈现，联合优化哈希表和网络参数都能保证得到相同的哈希映射信号（图 2(c) 和 (f)），且该信号包含更多低频成分。因此，现有隐式神经表示骨干网络的表达能力和任务性能都得到显著提升。如图 1 和图 2(d)、(g) 所示，在 DINER 框架下，具有浅层窄结构的微型基于多层感知机的隐式神经表示，能够很好地表征排列顺序任意的输入信号。

• 前三位作者贡献相同。
• *H. Zhu、S. Xie、Z. Liu、F. liu、Y.Zhou、Z. Ma* 和 *X. Cao* 任职于中国南京 *210023* 南京大学电子科学与工程学院。电子邮箱：*mazhan@nju.edu.cn*

• *Q. Zhang* 任职于中国深圳 *518054* 的腾讯人工智能实验室。
• *Y. Lin* 任职于中国上海 *200032* 复旦大学中山医院心血管外科。

稿件于*2005*年*4*月*19*日收到；修订于*2005*年*8*月*26*日，*2 015.*

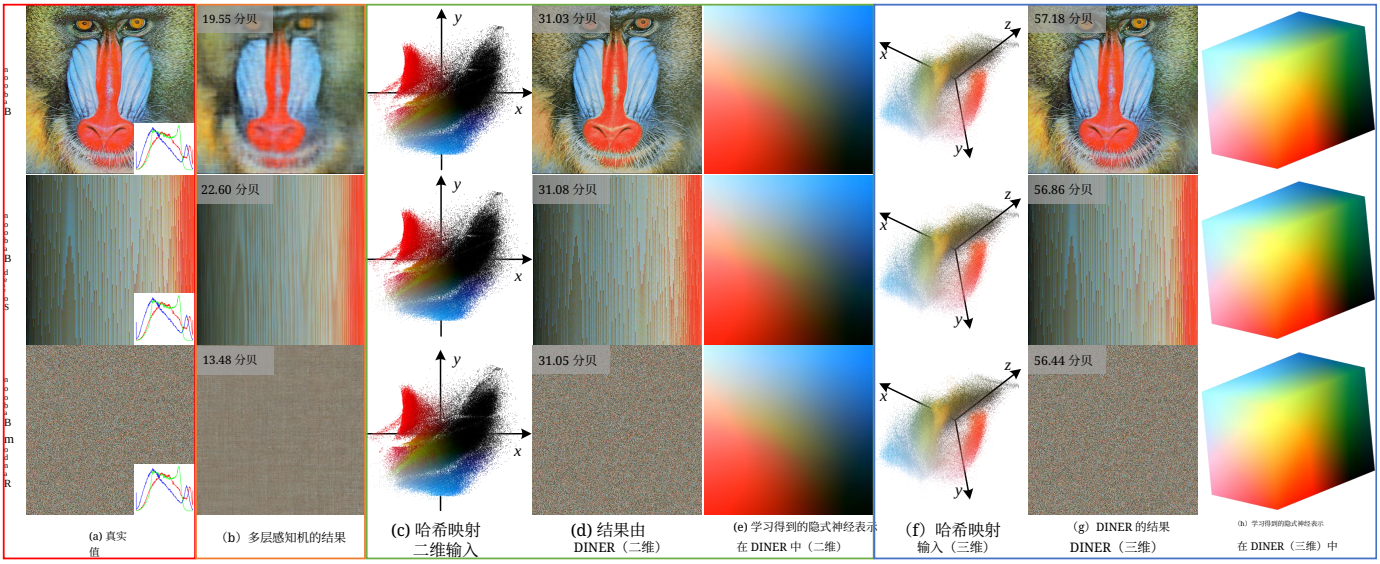


图 2. 现有隐式神经表示与 DINER 在不同排列下表示 Baboon 的比较。从上到下，Baboon 中的像素按不同的几何排列顺序排列，而直方图保持不变（(a) 中右下角的面板）。从左到右，(a) 为真实图像。(b) 展示了采用位置编码（PE+多层感知机） [11] 且宽度为 2×64 的多层感知机结果。(c) 为哈希映射坐标，哈希表宽度为 $L = 2$。(d) 为 DINER 的结果，使用与 (b) 相同宽度的多层感知机和大小为 $L = 2$ 的哈希表。(e) 为 DINER 的学习得到的隐式神经表示结果，其中将坐标输入训练好的多层感知机（哈希表宽度为 $L = 2$；详情见第 4.1 节）。(f)、(g) 和 (h) 的含义与 (c)、(d) 和 (h) 相同，但使用了另一种哈希表，其宽度为 $L = 3$。

使用哈希表以存储空间换取快速计算，其用于映射的可学习参数缓存通常与输入信号大小相近，但计算开销可以忽略不计，因为哈希表推导的反向传播为 $\mathcal{O}(1)$。

我们进一步指出，DINER 的表达能力由哈希表的宽度 L 决定。将 L 设为 1、2 或 3，分别等价于将原始信号的坐标投影到一维曲线、二维弯曲平面和三维弯曲体。DINER 的表达能力会随着哈希表宽度的增加而提升，直至达到信号的秩。

本工作是在 CVPR'23 [15]上发表的初步探索的扩展版本。与会议版本相比，我们对 DINER 的表达能力进行了建模，并发现其由哈希表的宽度决定。为验证这一模型，我们针对不同宽度的哈希表开展了额外实验，并使用新的哈希表宽度设置模型重新进行了会议版本中的全部实验。此外，我们还开展了两项额外任务，即十亿像素表示和神经辐射场优化。

主要贡献总结如下：

1) 所提出的 DINER 大幅提升了现有隐式神经表示模型较弱的表达能力：它引入哈希表，将原始输入的坐标映射到更适合后续隐式神经表示模型表征的形式。

2）所提出的 DINER 为属性直方图相同但排列顺序不同的信号提供一致的映射和表达能力。3）DINER 的表达能力通过参数化函数建模。函数覆盖的区域决定其能力，并且可以通过增加哈希表的宽度显著扩大，直到信号的秩。

4）所提出的 DINER 可推广至多种任务，包括二维图像和三维视频的表示，以及更多视觉任务，例如无透镜成像中的相位恢复、强度衍射层析成像中的三维折射率恢复，以及神经辐射场优化；在这些任务中，其性能均显著优于现有最先进方法。

2 相关工作

2.1 隐式神经表示与逆问题优化

隐式神经表示（有时称为神经场或坐标神经网络）利用神经网络建立坐标与其信号值之间的映射，可对多种信号进行连续且节省内存的建模，例如一维音频 [16], 2D 图像 [11], 3D 形状 [17], 4D 光场 [18] 以及五维辐射场 [3]。对于这些信号，可通过将网络输出与真实值进行比较来直接监督训练精确的隐式神经表示；也可采用间接方式，计算经过可微算子处理后的输出与真实值信号变体之间的损失。因此，隐式神经表示成为解决逆问题的通用工具，因为这类问题的前向过程通常是已知的。隐式神经表示已广泛应用于多个学科中的逆问题优化，例如计算机视觉与图形学 [7], 计算物理学 [6], 临床医学 [8], [19], 生物医学工程 [5], [9],材料科学 [20] 以及流体力学 [21], [22]。

2.2 隐式神经表示中的高频分量编码

大多数隐式神经表示使用多层感知机作为函数逼近器来表征信号的属性。根据逼近理论，多层感知机网络可以逼近

任意函数 [23]。然而，网络训练中存在频谱偏差 [10]，导致隐式神经表示对高频分量的表现不佳。已有多种方法尝试克服这种偏差，可分为两类，即函数展开和参数编码。

函数展开的思路将隐式神经表示的拟合视为使用不同基函数进行函数逼近。Mildenhall 等人 [3] 使用一系列不同频率的正弦 / 余弦 函数对输入坐标进行编码，并在辐射场表示中取得了巨大成功。通过随机傅里叶特征进一步改进了预设频率正弦 / 余弦 函数的策略；事实证明，这种方法在理论和实践中都能有效学习高频分量 [11]。Landgraf 等人 [24] 提出渐进式位置编码，以减轻这种偏差。由于每个层级都会移除占主导地位的低频内容，因此在重建频带较宽的场景时，性能有所提升。Sitzmann 等人 [1] 用周期激活函数（SIREN）取代了经典的 ReLU 激活函数。SIREN 的不同层可以视为逐步增加信号的不同频率支撑范围 [12]。SIREN 适用于表示复杂的自然信号及其导数。除了傅里叶展开之外，Fathony 等人 [13] 还使用多个小波函数的线性组合（多频网络）来表示复杂信号，其中通过调制 Gabor 滤波器中的频率，可以很好地建模高频分量。Lindell 等人 [14] 对多频网络进行了扩展，并提出了带限坐标网络（BACON），其中每个网络层的频率都可以在初始化时指定。Yang 等人 [25] 将多频网络和 BACON 推广为多项式神经场（多项式神经场）。此外，他们还提出了傅里叶多项式神经场，可对信号的不同成分进行操控，并将其应用于纹理迁移和尺度空间插值。与标准多层感知机网络相比，这些方法在表示高频分量方面具有显著优势。然而，这些隐式神经表示的性能受信号自身频率分布的限制，且通常需要更深或更宽的网络架构来提高拟合精度。

从参数编码的角度来看 [26],[27], [28], [29], ，每个输入坐标都通过学习特征进行编码，再将这些特征输入多层感知机进行拟合。Takikawa 等人 [27] 使用稀疏体素八叉树结构划分三维空间，其中每个点由其八个角点的可学习特征向量表示，从而实现高质量有符号距离函数的实时渲染。Martel 等人 [2] 在隐式神经表示训练过程中迭代划分坐标空间（ACORN），每个局部块的编码特征由坐标编码器网络获取，并输入解码器网络以获得信号的属性。ACORN 首次实现了对十亿像素图像的拟合，峰值信噪比接近 40 分贝。Muller 等人 [30] 用多分辨率哈希表取代坐标编码器网络。由于多分辨率哈希表在表征坐标特征方面具有更高的自由度，因此只需使用一个小型网络来映射特征和信号属性值。尽管其收敛速度更快、精度更高

在参数编码方面，两个关键问题仍未得到解答：即，这些特征具有怎样的几何意义？需要多少个特征才足够？

与这些方法相比，所提出的 DINER 中的哈希表明确地将输入坐标投影到另一个空间，换言之，将信号映射为包含更多低频分量的信号。不同的哈希表宽度对应维度不同的映射空间，从而具有不同的表达能力。表达能力会随着哈希表宽度的增加而增强，直至达到信号的秩。因此，与以往方法相比，一个微型网络也能实现很高的精度。

3 隐式神经表示的性能

3.1 隐式神经表示表达能力的背景

根据 Yuce 等人 [12], 的讨论，具有一维输入 x 的隐式神经表示可以建模为函数 $f_\theta(x)$ ，该函数将输入坐标 x 映射为其属性，也就是说，

$$\begin{aligned} \mathbf{z}^0 &= \gamma(x), \\ \mathbf{z}^j &= \rho^j(\mathbf{W}^j\mathbf{z}^{j-1} + \mathbf{b}^j),\ j = 1, ..., J-1, \\ f_\theta(x) &= \mathbf{W}^J\mathbf{z}^{J-1} + \mathbf{b}^J \end{aligned} \tag{1}$$

其中， $\gamma(\cdot)$ 是预处理函数，通常用于在网络中编码更多频率基； $\mathbf{z}^j$ 是隐式神经表示中第 j层的输出； ρ 是激活函数； $\mathbf{W}^j$ 和$\mathbf{b}^j$ 分别是第 j层的权重矩阵和偏置矩阵； J 是隐式神经表示的层数； $\theta = \{\mathbf{W}^j,\mathbf{b}^j\}_1^J$ 表示网络中所有训练参数的集合。

将非线性激活函数 ρ 展开为多项式激活函数后，隐式神经表示能够表示的信号具有如下形式

$$f(x) = \sum_{\omega' \in \mathcal{H}_\Omega} c_{\omega'} \sin(\langle \omega', x \rangle + \phi_{\omega'}), \tag{2}$$

其中， $\mathcal{H}_\Omega$ 是由预处理函数 $\gamma(\cdot)$中选定的频率决定的频率集合 [12], 例如，傅里叶编码 [11], 或 sin 激活 [1]。换言之，隐式神经表示的表达能力仅限于能够用某些谐波的线性组合表示的函数，这些谐波的频率为 $\gamma(\cdot)$ [12]。

3.2 信号的排列顺序决定隐式神经表示的表达能力

根据隐式神经表示的表达能力（公式 2），当隐式神经表示中的编码频率与信号的频率分布一致时，信号就能得到良好学习。然而，将这一结论应用于实践时存在两个问题，

1) 信号的频率分布无法预先确定，尤其是在逆问题中。因此，在设计隐式神经表示的架构时，无法很好地设定合适的频率。2) 由于网络训练中的频谱偏差 [10],，信号中的低频分量会先被学习，而高频分量的收敛速度则极其缓慢 [11], [31], [32], [33]。

我们注意到，如今记录或有待求解的信号大多是离散信号。通过以不同顺序排列元素，可以改变离散信号的频率分布，但需要额外存储排列规则，因此对公式 2 的满足程度也会有所不同。因此，隐式神经表示表征信号的表达能力会随排列顺序的不同而变化。

图2直观地展示了这一现象。图2(a)按不同顺序排列了Baboon[1]图像。原始图像包含丰富的低频、中频和高频信息。按像素强度对 Baboon 进行排序后， y轴上的高频信息几乎消失了。随后，我们按随机顺序排列像素。此时，Baboon 包含大量高频信息。接着，使用 PE+多层感知机（2×64，即每层包含64个神经元的2个隐藏层，采用 ReLU 激活函数）网络，学习这三幅图像中坐标与强度之间的映射（图2(b)）。拟合结果差异显著。对于排序后的图像，PE+多层感知机的表现最好，因为该图像包含最多的低频信息；而对于随机排序的图像，结果最差，因为该图像包含最多的高频信息。总之：

命题1. 信号的不同排列会产生不同的频率分布，从而使隐式神经表示表示信号自身时具有不同的表达能力。

4 无序不变隐式神经表示

4.1 隐式神经表示的哈希映射

给定配对离散信号 $Y = \{(\vec{x}_i,\ \vec{y}_i)\}_{i=1}^{N}$，其中$\vec{x}_i$ 为第 i 个 d_{in}维坐标， $\vec{y}_i$ 为相应的 d_{out}维信号属性。根据上述分析，理想排列规则M^*： $\mathbb{R}^{d_{in}} \to \mathbb{R}^L$ 应满足以下规则，

$$M^* = \arg\min_{M} \sum_{k=1}^{K_M} |\omega_k| \tag{3a}$$

$$\Omega_M \subseteq \mathcal{H}_\Omega,\ \Omega_M = \{\omega_k\}_{k=1}^{K_M}, \tag{3b}$$

其中Ω_M 是按照规则 M映射信号后得到的频率集合， $\mathcal{H}_\Omega$ 是隐式神经表示网络支持的频率集合（公式2）， K_M 是排列后信号中的频率数量， $|\cdot|$ 返回.....的绝对值

。需要注意的是，映射后的坐标具有 L 个维度，这表示要在 L维空间中重新排列点，而非在原始的 d_{in}维空间中排列。采用这种排列方式的信号能够得到良好学习，因为频率设置不当的问题（公式3b）以及
频谱偏差（公式 3a）已纳入考量。

然而，这种策略需要预先了解信号分布，因此只适用于压缩任务。相比之下，它无法优化逆问题，因为无法事先获得待优化的信号分布。在本小节中，我们将详细介绍所提出的 DINER，以解决这一问题。

1. Bab oon 是 MATLAB 中由 MathW 创建的独立图像 orks©。

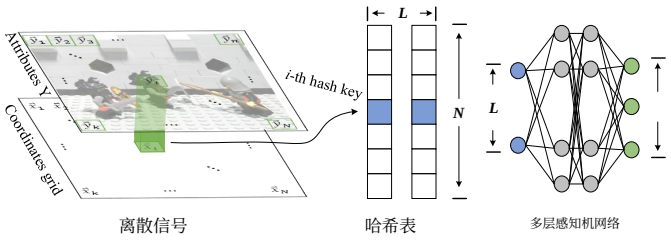


图 3. DINER 流程。

我们专门设计了一个全分辨率哈希表 $\mathcal{HM}$，用于对公式 3 中的映射 M^* 进行建模。“全分辨率”是指哈希表 $\mathcal{HM}$ 的长度 N 设置为与 Y中的元素数量相同。 $\mathcal{HM}$ 的宽度设置为 L （对于不同的信号，L 可以有所变化；有关 L的选择，请参阅第 5 节）。首先，使用输入坐标 $\vec{x}_i$ 查询 $\mathcal{HM}$ 中的第 i 个哈希键$M(\vec{x}_i)$。然后，将映射后的坐标 $M(\vec{x}_i)$ 输入标准多层感知机。 $\mathcal{HM}$ 中的所有参数均设为可学习参数，即， $\mathcal{HM}$ 中的参数和网络参数将在训练过程中联合优化。图 3 展示了上述过程（所用的 Lego Knight 来自 [34]。）

由于哈希表的作用，多层感知机网络实际学习的是映射信号。图 2(c) 和 (f) 分别展示了 Baboon 在哈希表宽度为 $L = 2$ 和3时的映射像素。可以看到，原始网格坐标（图 2(a)）被投影为不规则点（图 2(c) 和 (f)）。我们根据映射坐标的最小值和最大值均匀采样网格，并将其输入训练好的多层感知机，即即图 2(e) 和 (h)。为简便起见，后文将图 2(f) 中的图像称为“学习得到的隐式神经表示”。与 Baboon 相比，学习得到的隐式神经表示存在显著差异：前者更加平滑，并包含许多低频分量（图 2(e) 和 (h)）。因此，使用哈希表可以实现高精度的多层感知机拟合（图 2(d) 和 (g)）。

4.2 无序不变性分析

将隐式神经表示应用于信号表示和逆问题优化任务时，传统隐式神经表示和 DINER 的训练或优化过程可分别建模为公式（4a）和（4b）。

$$\theta^* = \arg\min_{\theta} \mathcal{L}\left(\mathcal{P}\left(\{f_\theta(\vec{x}_i)\}_1^N\right), \mathcal{P}\left(\{\vec{y}_i\}_1^N\right)\right) \tag{4a}$$

$$\begin{aligned}\theta^*, \mathcal{HM}^* &= \arg\min_{\theta,\mathcal{HM}} \mathcal{L}\left(\mathcal{P}\left(\{f'_\theta(\mathcal{HM}(\vec{x}_i))\}_1^N\right), \mathcal{P}\left(\{\vec{y}_i\}_1^N\right)\right) \\ &= \arg\min_{\theta,\mathcal{HM}} \mathcal{L}\left(\mathcal{P}\left(\{f'_\theta(\mathcal{HM}_i)\}_1^N\right), \mathcal{P}\left(\{\vec{y}_i\}_1^N\right)\right),\end{aligned} \tag{4b}$$

其中 $\mathcal{P}$ 是一个物理过程，也是表示任务中的恒等变换， $\mathcal{L}$ 是根据测量值和重建结果计算的损失函数， $\mathcal{HM}_i$ 是哈希表中的第 i个键， f'_θ 的参数设置与 f_θ 相同，只是根据哈希表的宽度 L增加或减少了输入变量的数量。由于哈希索引操作没有

梯度，公式 4b 中的第一个等式可以简化为第二个等式。

可以发现，公式4a中的坐标 $\vec{x}_i$ 与数值 $\vec{y}_i$ 之间的配对关系在公式4b中被打破了。损失函数（公式4b）中只有一个独立变量 $\vec{y}_i$ 。假设 θ 中的参数均以相同的值初始化， $\mathcal{HM}$ 中的所有键在每次实验中都设为同一个值（例如，0），并且每次迭代的批大小设为 N；当对信号 Y采用不同的顺序时，例如， $Y' = \{\vec{x}_j, \vec{y}_i\}^{j=1,\ i=N}_{j=N\ i=1,}$ ，信号 Y和 Y'的 θ与 $\mathcal{HM}$ 训练在每次梯度更新中共享相同的优化进程，因为公式4b中 Y 的所有参数都与 Y'的参数等价。因此，优化的将是相同的 θ^* ，而 Y' 的 $\mathcal{HM}'$ 也是 Y的 $\mathcal{HM}$ 的逆排列。这种等价性并不限于顺序相反的 Y'；实际上，可以很容易地证明，对于顺序任意的 Y，这种等价性也成立。

图2(c)–(h)展示了这种等价性。尽管 Baboon 的排列顺序不同，哈希表仍将它们映射为相同的信号（图2(c)、(e)和图2(f)、(h)），而 DINER 对它们进行优化后得到的峰值信噪比相近（ $L = 2$时分别为31.03、31.08、31.05； $L = 3$ 时分别为57.18、56.86、56.44）[2]。总而言之：

命题2。 *DINER* 具有无序不变性，具有相同属性直方图分布的信号共享一个经过优化且参数值相同的网络。

4.3 讨论

骨干网络。 所提出的 DINER 的骨干网络并不局限于上述标准多层感知机；事实上，也可以将 SIREN 等其他网络结构与哈希表结合，从而获得优于原始结构的性能。更多详情请参阅实验部分。

复杂度。 尽管哈希表的参数数量远多于网络，但训练成本非常低，因为使用批量大小为 1 训练多层感知机时，每次只需更新一个哈希键。因此，训练哈希表的计算复杂度在每次训练迭代中为 $\mathcal{O}(1)$。请注意，当哈希表规模较小时， $\mathcal{O}(1)$ 复杂度才成立。随着哈希表规模增大，由于图形处理器中内存与缓存之间的通信成本上升，训练所需的时间也会增加（更多详情请参见第 6.4 节和第 6.5 节）。

5 DINER 的表达能力

根据上述分析，DINER 能够为属性具有相同直方图分布的信号提供稳定一致的性能。然而，DINER 的表达能力究竟如何仍是一个未解决的问题。本节将对此进行分析。

2.数值上的微小差异源于图形处理器在对具有相同直方图和不同排列顺序的矩阵求和时产生的浮点误差。

5.1 DINER 的参数化函数

重新审视公式 4b，DINER 的性能与信号的排列顺序无关，而取决于哈希表 $\mathcal{HM}$ 和后续网络 f'_θ的表达能力。给定一个哈希表$\mathcal{HM}$，有三个可调变量，即哈希表长度 N、宽度 L 以及 $\mathcal{HM}$中的值。由于第一个变量 N 指的是信号中的点数，因此无法更改。此外，考虑到 $\mathcal{HM}$中的所有值都是可学习的，第三个变量无法在训练过程之前预先设定。因此，*DINER* 的性能取决于宽度 L 的 $\mathcal{HM}$ 以及后续网络 f'_θ的表达能力。

为获得更直观的体验，以下演示和推导将聚焦于学习一幅二维彩色图像（$d_{out} = 3$）。根据隐式神经表示的表达能力（公式 2）以及第 4.1 节中的分析，当 $L = 1$、2 和 3 时，公式 4b 的优化可视为分别追求形状为一维曲线、二维平面和三维体积的学习得到的隐式神经表示，以及一个将原始坐标映射到一维曲线、二维平面和三维体积中坐标的哈希表。根据这些学习得到的隐式神经表示的属性，在 d_{out} 维属性空间中绘制它们（例如，当 $d_{out} = 3$ 时，具有属性 $[y_1, y_2, y_3]^\top$ 的点将绘制在位置 $[y_1, y_2, y_3]^\top$），一维曲线、二维平面或三维体积便会变成曲面版本（下文统一称为“超曲面”）。图 4 展示了学习二维 RGB图像时的一维／二维／三维“超曲面”。从图 4 可以看出，*DINER* 的表达能力问题转化为曲面如何覆盖属性空间中的所有点这一问题。

根据隐式神经表示的表达能力，即公式 2，这些“超曲面”可以使用参数函数显式描述，即

$$\begin{cases} r = f_1(x) = \sum_{\omega' \in \mathcal{H}_\Omega} c^1_{\omega'} \sin(\langle \omega', x \rangle + \phi^1_{\omega'}) \\ g = f_2(x) = \sum_{\omega' \in \mathcal{H}_\Omega} c^2_{\omega'} \sin(\langle \omega', x \rangle + \phi^2_{\omega'}) \\ b = f_3(x) = \sum_{\omega' \in \mathcal{H}_\Omega} c^3_{\omega'} \sin(\langle \omega', x \rangle + \phi^3_{\omega'}) \end{cases} \tag{5}$$

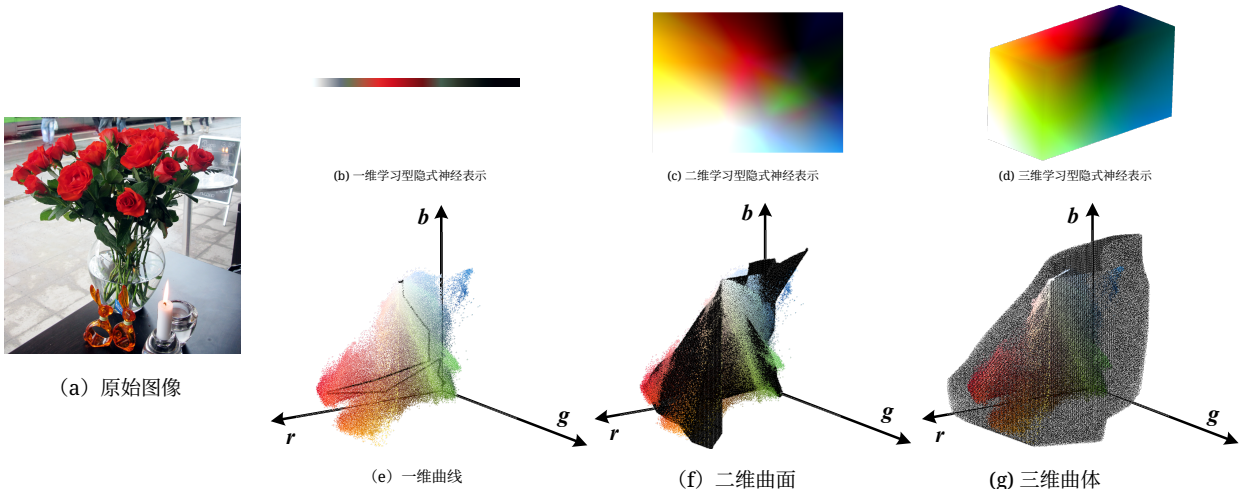
其中， x 是参数变量，指代输入网络的坐标输入，或学习得到的隐式神经表示所在空间中的坐标。 x 的形状分别为 1×1、 2×1 和 3×1，当 $\mathcal{HM}$的宽度分别设置为1、2和3时。同时， ω' 的形状会随 x的形状而变化。根据公式 5，DINER 在 RGB空间中的表达能力由 x的维度，即哈希表的宽度和频率集合 $\mathcal{H}_\Omega$决定。

5.2 宽度与频率

假设公式 5 所能表示的所有可能“超曲面”构成集合 $\mathcal{S}^\omega_L$，其中 L 表示哈希表的宽度， ω 是预处理函数 $\gamma(\cdot)$中的编码频率。可以很容易地证明

$$\mathcal{S}^\omega_1 \subset \mathcal{S}^\omega_2 \subset \mathcal{S}^\omega_3 \tag{6}$$

依次将哈希键中的第2个和第3个值设为0（详情见附录A）。引入



（a）原始图像 （e）一维曲线 （f）二维曲面 (g) 三维曲体

图 4. 分别使用哈希表宽度为1、2 和3的 DINER 学习二维图像的示例。(a) 原始图像。(b)-(d) 哈希表宽度分别为1、2 和3 时学习得到的隐式神经表示。(e)-(g) 学习得到的隐式神经表示在三维 RGB 空间中对应的“超曲面”，其中“超曲面”上的点以黑色标记，原始图像中的点则以其原始颜色标记。

哈希键中的第2个和第3个变量显著提升了 DINER 的表达能力，因为二维曲面能够覆盖比一维曲线大得多的空间，而三维曲体则进一步大幅扩展了空间。相反，在 $\gamma(\cdot)$ 中编码更多频率，只会线性增加 $\mathcal{H}_\Omega$ 的大小和公式 5 的空间。

总之，相较于在预处理函数中增加更多频率，哈希表的宽度起着重要得多的作用 $\gamma(\cdot)$。

5.3 关于宽度的数量

使用 $d_{out} = 3$ 表示二维彩色图像时，以下公式成立：

$$\forall\, i, j \geq 3,\ \mathcal{S}_i^0 = \mathcal{S}_j^0. \tag{7}$$

为了证明这一点，我们引入以下用于表示 RGB 色彩空间的参数化函数，

$$\begin{cases} r = \lambda_1 + 0\lambda_2 + 0\lambda_3 \\ g = 0\lambda_1 + \lambda_2 + 0\lambda_3 \\ b = 0\lambda_1 + 0\lambda_2 + \lambda_3 \end{cases}, \tag{8}$$

其中，$\lambda = [\lambda_1, \lambda_2, \lambda_3]^\top$ 是一个三维参数变量（公式 8 表示的空间稍后标记为 S_{rgb}）。众所周知，这些参数化函数可以用标准多层感知机轻松建模（公式 1），例如，构建一个具有 3 个输入变量和 3 个输出值的 1 层多层感知机，其中 $\gamma(\cdot)$ 中未编码任何频率，同时参数$\mathbf{W}^1$ 和 $\mathbf{b}^1$ 设为

$$\mathbf{W}^1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},\ \mathbf{b}^1 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \tag{9}$$

此外，不应用激活函数。由于这个简单的多层感知机可以表示三维空间中的所有点，并且它属于由三维输入 x 和 $\omega = 0$ 通过公式 5 构造的可能多层感知机集合，因此以下公式成立

$$\mathcal{S}_{rgb} = \mathcal{S}_3^0. \tag{10}$$

由于 $\mathcal{S}_3^0$ 可以覆盖整个RGB空间，因此进一步增加哈希表的宽度以提升DINER的表达能力毫无意义，公式 7 成立。

此外，上述推导并不限于 RGB 彩色图像；实际上，它可以推广到具有 d_{out} 维属性的其他信号，公式 6、7 相应修改为

$$\begin{aligned} &\mathcal{S}_1^\omega \subset \mathcal{S}_2^\omega \subset \ldots \subset \mathcal{S}_{d_{out}}^\omega \\ &\forall\, i \geq d_{out},\ \mathcal{S}_i^\omega = \mathcal{S}_{d_{out}}^\omega \end{aligned}. \tag{11}$$

总而言之，

命题 3. 当哈希表的宽度设为信号属性的维度数时，*DINER* 的学习得到的隐式神经表示覆盖整个属性空间。进一步增加宽度并不能提升 *DINER* 的表达能力。

备注：
上述分析聚焦于线性相关信号，即，属性的任何维度都不能用其他 $d_{out} - 1$ 个维度线性表示。对于线性相关信号，可以很容易证明，当参数化函数（式 5）具有 $\rho(Y)$ 个输入变量时，它能够覆盖整个属性空间，其中 $\rho(Y)$ 指的是信号属性的秩。式 11 修改为

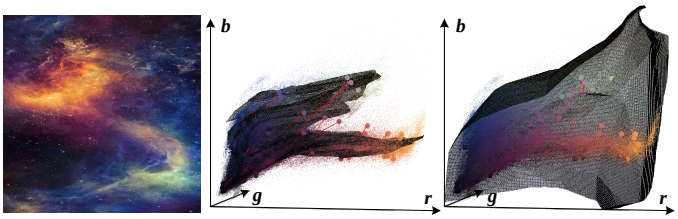
$$\begin{aligned} &\mathcal{S}_1^\omega \subset \mathcal{S}_2^\omega \subset \ldots \subset \mathcal{S}_{\rho(Y)}^\omega \\ &\forall\, i \geq \rho(Y),\ \mathcal{S}_i^\omega = \mathcal{S}_{\rho(Y)}^\omega. \end{aligned} \tag{12}$$

因此，

命题 4. 对于线性相关信号，建议将哈希表的宽度设为属性的秩，而非属性的维度数。

5.4 参数空间中的 DINER与传统隐式神经表示

图 5比较了使用相同多层感知机的参数化 RGB 空间中的 DINER（$L = 2$）与传统隐式神经表示



(a) 原始图像 (b) 传统隐式神经表示 (c) DINER

图 5. DINER与传统隐式神经表示在RGB空间中的比较。(a)为原始图像。(b)、(c)中的离散点是(a)中的像素，其位置由颜色决定。(b)和(c)中的黑色曲面分别表示传统隐式神经表示和DINER学习到的连续函数。(b)和(c)中的红线表示真实值与预测值之间的距离（为便于观察，共有150对像素以较大的点标出。）

网络结构。受输入坐标与属性之间配对关系的限制，传统隐式神经表示致力于构建一个复杂的二维曲面，使其按照原始图像中各点的顺序穿过 RGB 空间中的所有点。因此，必须在参数化函数中编码复杂的频率基，以支持构建这种复杂曲面。相反，哈希表的引入打破了这种配对关系，因此 DINER 中后续的多层感知机网络可以专注于寻找一个尽可能穿过所有点的曲面，而无需考虑连接这些点的顺序。曲面构建完成后，哈希表的优化将专注于寻找曲面上与真实值点距离最小的点。因此，尽管在使用相同多层感知机网络时，传统隐式神经表示与 DINER 共享相同的参数化函数族（公式 5），但哈希表的引入使 DINER 能够找到一个曲面，使真实值与预测值之间的距离小于传统隐式神经表示。

6 实验

本节将重点验证 DINER 的表达能力。我们将在二维图像拟合任务中验证 DINER 的所有特征，并进一步通过表示十亿像素图像和三维视频的任务测试 DINER 的性能。

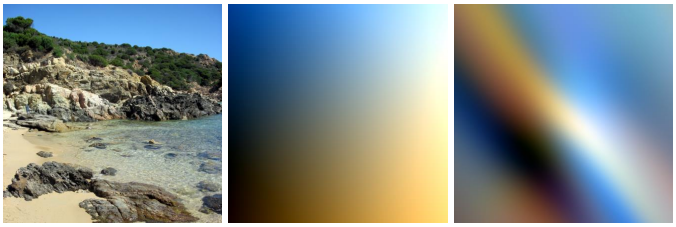
6.1 数据集和算法设置

对于二维图像拟合任务，我们使用来自 TESTIMAGES 数据集 SAMPLING 类别的 30 张高分辨率图像，分辨率为 1200×1200 [35] 。TESTIMAGES 数据集中的每张图像均由专门编写的 Octave/MATLAB 自定义软件脚本生成，以确保每个像素的位置和值都精确无误，并包含丰富的低频、中频和高频信息。对于表示三维视频的任务，

使用了UVG数据集中的 'ReadySetGo' 和 'ShakeNDry' [36] 。

表 1 原始图像与学习得到的隐式神经表示之间频率分布比例的比较。本文对比了两种骨干网络：即多层感知机和 SIREN。

频带				
原始图像	0.4426	0.2484	0.1753	0.1337
学习得到的隐式神经表示（多层感知机）	0.6784	0.1218	0.0973	0.1025
学习得到的隐式神经表示（SIREN）	0.6354	0.1220	0.1149	0.1276



(a) 原始图像 (b) 学习得到的隐式神经表示（多 (c) 学习得到的隐式神经表示（

图 6. 分别比较采用多层感知机和 SIREN 骨干网络的 DINER 中学习得到的隐式神经表示。

6.2 有无哈希表时频率分布的比较

我们在具有相同网络规模的多层感知机和 SIREN 结构上验证了哈希表 2×64 （即，2 层，每层 64 个神经元）。正如第 4.1 节和图 2 所示，哈希表将原始信号映射为包含更多低频成分的信号。表 1 统计了 30 张图像的原始图像与学习得到的隐式神经表示的平均频率分布（哈希表的宽度设置为 2）。映射后，中频和高频信息的比例均有所降低，而低频信息有所增加。

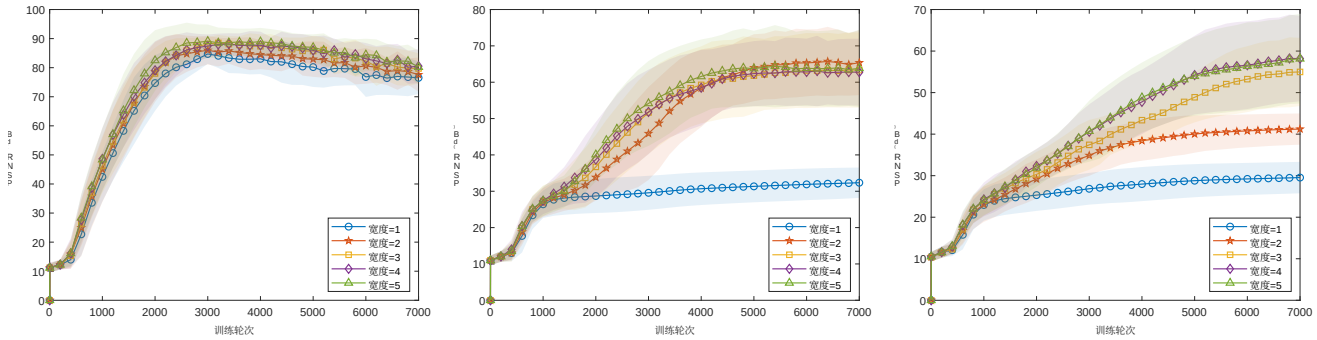
与 SIREN 结构所支持的频率集合 $\mathcal{H}_{\Omega}^{SIREN}$相比（其激活函数中使用的默认频率为 30 ），未使用任何频率编码的多层感知机结构的 $\mathcal{H}_{\Omega}^{MLP}$包含更多低频和更少高频。因此，与采用 SIREN 骨干网络的 DINER 映射图像相比，采用多层感知机骨干网络的 DINER 映射图像包含更多低频信息（ 0.6784 与0.6354，图 6）。

6.3 二维图像拟合中不同哈希表宽度下 DINER 的比较

在本小节中，我们将验证 DINER 中哈希表宽度对不同通道数信号的影响。

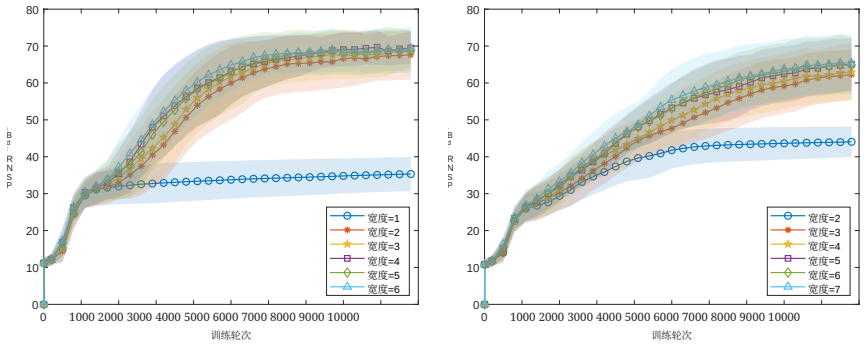
具有线性无关通道的图像。我们进行了三项实验，以验证 DINER 在具有线性无关通道的图像上的性能，即，将 DINER 应用于具有 1、2 和 3 个通道的图像。对于每种类型的数据，我们使用不同宽度的哈希表评估 DINER 的性能，例如，$\{1, 2, 3, 4, 5\}$。

图 7 比较了不同设置下的训练曲线。可以注意到，随着哈希表宽度的增加，峰值信噪比也随之提高，直到宽度达到图像的



(a) $d_{out} = 1$ (b) $d_{out} = 2$ (c) $d_{out} = 3$

图7. 对具有不同线性无关属性的二维图像应用不同宽度哈希表的 DINER 所得到的峰值信噪比。



(a) $d_{out} = 6,\ \rho(Y) = 2$ (b) $d_{out} = 7,\ \rho(Y) = 3$

图8. 对具有不同秩的二维图像应用不同宽度哈希表的 DINER 所得到的峰值信噪比。

通道数；当哈希表宽度大于通道数时，峰值信噪比趋于稳定。

通道线性相关的图像。我们另外进行了两项实验，以验证 DINER 在通道线性相关图像上的性能。在第一项实验中，我们合成了一个包含 30 张 6 通道多光谱图像的数据集，其中前 2 个通道直接复制自 30 张 SAMPLING 图像的 R 和 G 通道，最后 4 个通道则由前 2 个通道线性生成，即此处的 $\rho(Y) = 2$ 。随后，我们将哈希表宽度为 $\{1, 2, 3, 4, 5, 6\}$ 的 DINER 应用于这些图像。在第二项实验中，我们合成了一个包含 30 张 7 通道多光谱图像的数据集，其中前 3 个通道保留自原始 RGB 通道，最后 4 个通道同样由前 3 个通道线性生成，即此处的 $\rho(Y) = 3$ 。我们将哈希表宽度为 $\{2, 3, 4, 5, 6, 7\}$ 的 DINER 应用于这些 7 通道多光谱图像。

图 8(a) 和 (b) 分别展示了这两组实验的训练曲线。随着哈希表宽度增加，峰值信噪比也随之提高，直到分别达到这两个数据集的属性秩。

6.4 二维图像拟合与最先进方法的比较

我们将提出的 DINER 与傅里叶特征位置编码（PE+多层感知机）[11], SIREN [1]以及 Instant-NGP [30]进行比较。需要注意的是，两种骨干网络，即也就是标准的采用 ReLU 激活的多层感知机和采用周期函数激活的 SIREN，均与提出的哈希表结合，以便更好地评估性能。我们调整 InstantNGP 中哈希表的大小，以确保其参数数量与我们的方法相近，例如， 2^{21} 在二维图像拟合任务中，我们的方法使用的哈希表长度为 $1200^2 < 2^{21}$。所提出方法和 InstantNGP 中哈希表的宽度均设为 3，与二维图像的属性数量相同。除此之外，所有 5 种方法均采用相同的损失函数来衡量预测值与真实值之间的差异，其他参数则采用作者设定的默认值。

图 1 展示了不同方法在各训练轮次 [3] 下的峰值信噪比。可以看到，SIREN 和 PE+多层感知机在训练早期快速收敛，最终分别达到约24分贝和 21分贝。相比之下，本文提出的两种方法都比骨干网络具有更高的精度。使用哈希表进行图像拟合后，两个骨干网络的峰值信噪比分别提高了30分贝和 17分贝。此外，尽管 InstantNGP 很快收敛到约30分贝（约 200 个训练轮次），但在最后 2800 个训练轮次中，曲线趋于稳定。采用多层感知机骨干网络的 DINER 比 InstantNGP 高6分贝。

图9展示了训练3000轮次后的定性结果。

3. 尽管基于 DINER 的方法在图 1 中未收敛至最优解，但相较于其他方法已取得显著优势。因此，本文未绘制更多训练轮次的训练曲线，详情请参见图 7(c) 中的黄色曲线。



图9. 训练3000轮后，各种方法在二维图像拟合上的定性比较。可以看出，与 SIREN 和 PE+多层感知机相比，我们的方法能够提供更加清晰的细节，尤其是在钟楼的高频边界处（黄色框）。尽管 InstantNGP 达到了更好的结果，但如红色框中的墙壁和黄色框中的天空所示，由于训练过程中采用了线性插值，其放大结果仍然存在许多噪声。相比之下，我们的方法取得了最佳结果，展现了 DINER 的性能。

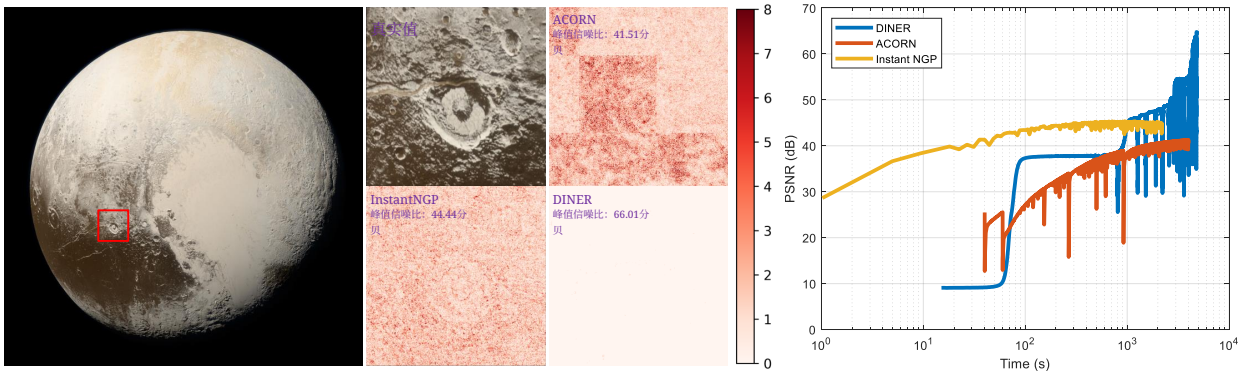


图10. 各种方法表示十亿像素图像“冥王星”的比较。从左到右依次为真实值、不同方法的误差图比较，以及随训练时间变化的峰值信噪比曲线。可以看出，所提出的 DINER 在十亿像素图像表示方面优于其他方法，尤其是在高频区域。此外，ACRON 存在明显的块状伪影，而 InstantNGP 生成的结果较为平滑。然而，我们的方法优于这两种基线方法。此外，与基线方法相比，所提出的方法展现出更强的表达能力。

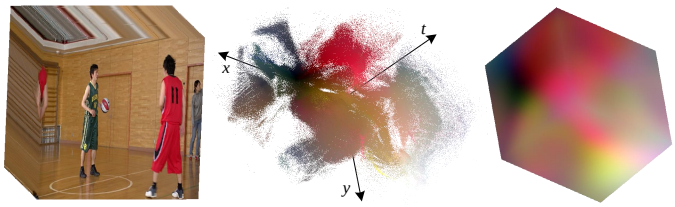
所提出的方法优于 SIREN 和 PE+多层感知机。我们的方法能更清晰地呈现细节，尤其是在高频边界处，例如图9中钟楼（黄色方框）的细节。乍看之下，InstantNGP 拟合的图像与真实值非常相似，但放大后可以看到许多噪声，例如图9中墙壁（红色方框）和天空（黄色方框）中有大量噪点，导致峰值信噪比指标较低。根据第5节和公式6中的分析，由于周期激活函数编码了高频信息，DINER（SIREN）略强于 DINER（多层感知机），因此在最初3000 训练轮次中，DINER（SIREN）的表现优于 DINER（多层感知机）。经过 10000 训练轮次后，这两种方法的性能趋于相近（59.64分贝与 59.53 分贝；由于这些结果不会影响此处的结论，因此未予展示）。

表2列出了5种方法的训练时间。Instant-NGP 使用 tiny-cuda-nn [37], 实现，其他4种方法则使用 PyTorch 实现。5种方法均在 NVIDIA A100 40GB GPU 上训练。哈希表的优化在 SIREN 架构上额外耗时3秒，而与经典的 PE+MLP 架构相比，可减少20秒，这验证了优化

哈希表的复杂度较低。

表2 使用3000个训练轮次拟合二维图像的方法比较。

	DINER（SIREN）	DINER（多层感知机）	SIREN	InstantNGP	PE +MLP
Time	81.1秒	59.1秒	77.6秒	38s	78.8秒



(a) 原始视频 (b) 哈希映射像素 (c) 学习得到的隐式神

图11. 三维视频“Basketball- Pass” [38]上学习得到的隐式神经表示的可视化。(a)和(b)比较了使用和不使用哈希表时的坐标。(c)展示了SIREN经过哈希表映射后的学习结果。

6.5 十亿像素图像表示的最先进方法比较

将坐标映射到十亿像素图像的颜色，可用于测试模型对高频细节的表现，也是隐式神经表示中的一项重要任务。我们比较了

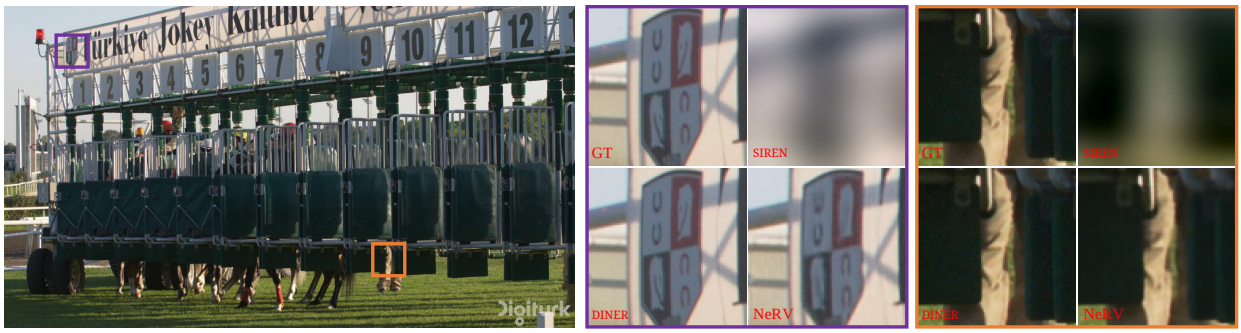


图12. 经过500个训练轮次后，各种方法在三维视频表示上的定性比较。所提出的 DINER 优于所比较的基线方法。具体而言，与高分辨率相比，SIREN 的网络规模较小，因此结果较为平滑、模糊。NeRV 能够恢复高清晰度视频，但会丢失高频细节，例如右侧放大图中的广告牌和马腿细节。

表 3 使用500个训练轮次对三维视频“ReadySetGo”和“S hakeNDry”进行训练的比较。

方法	网络参数	训练时间	PSNR
SIREN	8.77K	706s	21.22 分贝
NeRV	97.24M	7445s	36.08 分贝
DINER	8.77K	1309秒	50.74 分贝

我们使用冥王星图像，将 DINER 与两种近期最先进方法进行比较，即 ACORN [2] 和 InstantNGP [30]（参数设置与上述实验相同）。图10展示了分辨率为 8000×8000 时冥王星图像的对比结果。与最先进方法相比，DINER 能为十亿像素图像提供可靠得多的结果。

6.6 在三维视频表示方面与最先进方法的比较

视频描述由多帧组成的动态三维场景 $I(t,\ x,\ y)$。三维视频的精确表示正成为隐式神经表示领域的一项热门任务 [39], [40], [41], [42] 。我们将所提出的 DINER 与 SIREN [1] 以及最先进的隐式神经表示方法 NeRV [39]进行比较。需要注意的是，这项任务中仅评估了 Hash+SIREN 架构。NeRV 使用其默认参数实现，而 SIREN 和所提出的 Hash+SIREN 使用相同的网络结构，尺寸为 4×64。这三种方法均在 UVG数据集 [36] 中的“ReadySetGo”和“ShakeNDry”视频上进行评估，并训练 500个训练轮次。实验中使用前 30 帧，分辨率为 1920×1080 。

图11(a)和(b)展示了映射前后坐标的映射结果。图 11(c)展示了学习得到的隐式神经表示。值得注意的是，在我们的 DINER 中，三维视频学习得到的隐式神经表示也呈现出低频特性。表3展示了定量比较。所提出的方法在质量和速度上均优于 NeRV，分别提升了14 分贝和 5× 。由于 DINER 使用了较大的哈希表，图形处理器中内存与缓存之间的传输需要更多时间，因此 DINER 的训练时间比 SIREN 更长。图12展示了定性比较。值得注意的是，使用微型网络的 SIREN 无法为“ReadySetGo”数据提供合理的表示，$30 \times 1920 \times 1080$ 像素，因此所有像素都变得平滑。NeRV 的结果优于 SIREN，但会丢失高频细节，例如紫色方框中左上角的字符“U”和右上角的马头红色标志，以及橙色方框中的裤子褶皱。相比之下，所提出的 DINER 对原始视频的映射中仅包含少量高频成分（图11(c)）。因此，上述细节都能得到良好表示。

7 应用

本节将验证 DINER 求解逆问题的性能。我们将开展三个独立任务，即无透镜成像中的二维相位恢复、强度衍射层析成像中的三维折射率恢复，以及神经辐射场优化。请注意，这些任务的哈希表宽度根据第 5 节中的分析设定，因此部分结果可能与会议版本 [15]有所不同。

7.1 无透镜成像中的相位恢复

无透镜成像 [43] 无需任何光学透镜，即可在极近距离观测标本。通过直接记录衍射测量值，这种显微技术能够实现大视场观测，因此已成为一种颇具吸引力的显微技术 [44] ，可用于分析标本的特性。我们以经典的多高度无透镜成像为例，其中 N 测量值 $\{I_z\}_{z=z_1}^{z_N}$是在不同的标本与传感器距离 z下采集的，用于恢复标本的振幅与相位成像。在多高度无透镜成像中，I_z 可建模为对标本的复场 $O(x,\ y)$ 应用菲涅耳传播，即，

$$I_z = |PSF_z * [P(x,y) \cdot O(x,y)]|^2, \tag{13}$$

其中，$P(x,y)$ 是照明图案，PSF_z是标本与传感器之间距离为z 时菲涅耳传播的点扩散函数。

我们使用所提出的方法对 $O(x,\ y)$ 进行建模，采用 SIREN 骨干网络，网络规模为 2×64。损失函数通过比较测量值与将公式 13 应用于网络输出后得到的结果来构建。我们将所提出的方法与当前最先进方法进行比较，即，

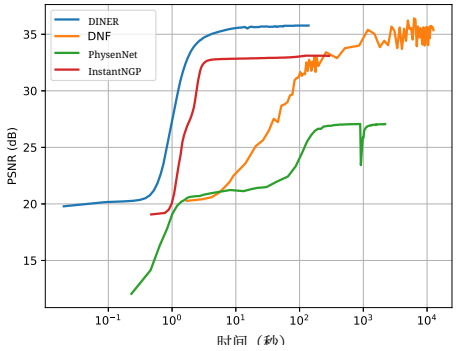


图 13. 无透镜成像真实数据在训练时间内重建测量值的峰值信噪比。

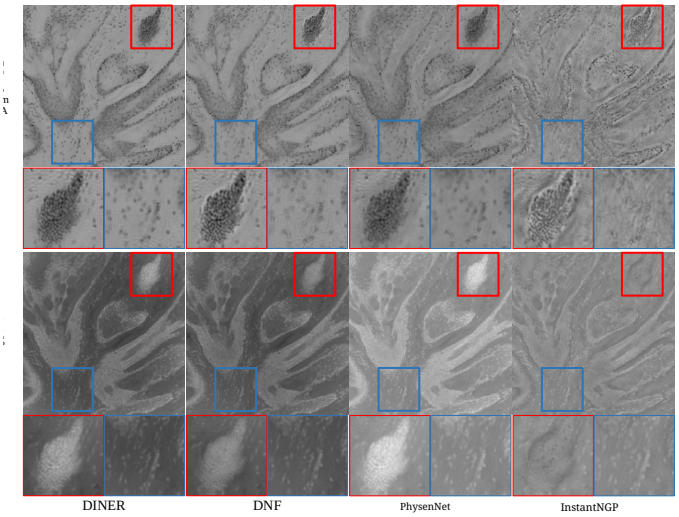


图 14. 动物皮肤切片无透镜成像真实数据的比较。尽管重建性能接近 DINER，DNF 和 PhysenNet 收敛所需的计算时间却是其 100× 倍。InstantNGP 的训练速度很快，但结果中包含严重伪影。相比之下，所提出的 DINER 收敛所需时间更短，并能恢复动物皮肤切片的高质量细节。

衍射神经场（DNF） [5] （骨干网络采用 PE+多层感知机结构）和 PhysenNet [45]，。此外，InstantNGP [30]也应用于无透镜成像模型（采用与 DINER 中类似的哈希表参数设置），并进行比较。请注意，此处仅提供真实测量值上的结果；合成数据上的比较请参阅会议版本。

图 13 展示了动物皮肤切片真实数据的重建测量值峰值信噪比曲线。与 PhysenNet 相比，DINER 的重建测量值具有更高的峰值信噪比（> 7分贝）。此外，DINER 在达到 25 分贝时，收敛速度提升约 100× 。尽管 DNF 能取得与 DINER 相近的结果，但由于使用了复杂的多层感知机（ 8×256 对比 2×64），它需要超过 100× 倍的训练时间才能达到35 分贝。InstantNGP 的收敛速度与 DINER 相近，但后者在峰值信噪比指标上高出3分贝。图 14 展示了不同方法在相同训练轮次（10000次）下的定性结果。DINER 能够还原皮肤样本的精细细节。尽管 DNF 和 PhysenNet 的重建性能与 DINER 相近，但它们的重建

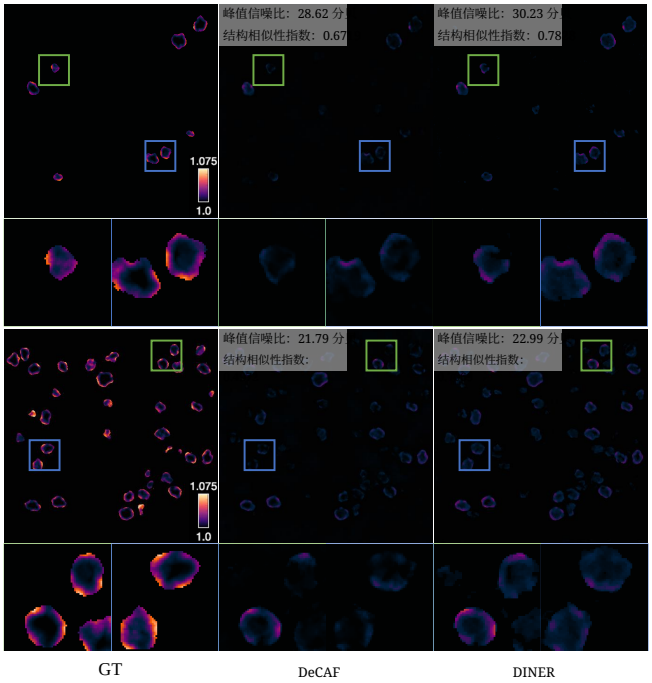


图 15. 三维折射率恢复的比较。DINER 所需训练时间更短，并能重建粒细胞更多的表面细节。

时间约为 100×。InstantNGP 的成像质量相对较低，且存在严重的重建伪影，尽管其收敛速度与 DINER 一样快。如图 14 的放大图所示，InstantNGP 无法准确还原高分辨率图像细节。

7.2 强度衍射层析成像中的三维折射率恢复

三维折射率表征标本内光与物质之间的相互作用。它是无需对标本进行染色或标记即可用于成像的内源性光学对比度来源，在许多领域都发挥着重要作用，例如形态发生、细胞病理生理学和生物化学 [46]。强度衍射层析成像通过多次测量不同角度下从标本散射的光的振幅平方，已成为恢复三维折射率的一种流行技术。

给定标本的三维折射率 $\mathbf{n} = (\mathbf{n}_{re} + j\mathbf{n}_{im})$，其中$\mathbf{n}_{re}$ 和 $\mathbf{n}_{im}$ 分别是标本折射率的实部和虚部。位于 ρ 处的传感器的正向成像过程可建模为

$$I_\rho = \mathbf{A}_\rho \Delta\epsilon, \tag{14}$$

其中，$\mathbf{A}\rho$ 记录了照明条件下的样品-强度映射。 $\Delta\epsilon = \Delta\epsilon_{re} + j\Delta\epsilon_{im}$ 是复介电常数对比，可通过求解获得

$$\begin{aligned} \mathbf{n}_{\mathrm{re}} &= \sqrt{\frac{1}{2}\left(\left(\mathbf{n}_0^2 + \Delta\epsilon_{\mathrm{re}}\right) + \sqrt{\left(\mathbf{n}_0^2 + \Delta\epsilon_{\mathrm{re}}\right)^2 + \Delta\epsilon_{\mathrm{im}}^2}\right)}, \\ \mathbf{n}_{\mathrm{im}} &= \frac{\Delta\epsilon_{\mathrm{im}}}{2 \cdot \mathbf{n}_{\mathrm{re}}} \end{aligned} \tag{15}$$

表4 新视角合成的定量比较。“平均值”误差指标中的数值指均方误差、$\sqrt{1-SSIM}$ 和学习感知图像块相似度（LPIPS） [48]的平均值。

	DINER	NeRF	Plenoxels	InstantNGP	DVGO
PSNR	33.18	31.04	30.99	33.53	33.38
SSIM	0.967	0.953	0.956	0.964	0.968
学习感知图像块相似度	0.033	0.054	0.050	0.038	0.032
平均值	0.067	0.088	0.084	0.072	0.067
时间（秒）	418	10887	305	343	252

其中，$\mathbf{n}_0$ 是背景介质的折射率。

我们使用 DINER 对$(\Delta\epsilon_{re}, \Delta\epsilon_{im})$ 进行建模，网络规模为 2×64 ，并使用与 DeCAF 相同的损失函数。我们将我们的方法与最先进水平进行比较，即即 DeCAF [9] ，它结合了标准多层感知机结构（网络规模为 10×208 ）以及位置编码+和径向编码。

图 15 使用 Fiji 软件 [47] 对比了我们的方法与 DeCAF 在三维粒细胞体模数据上的表现。由于粒细胞体模的真实值是最近才发布的，我们调整了损失函数中不同成分的权重，因此图 15 中的结果图像与会议版本 [15] 不同。为获得上述结果，DeCAF 耗时401 分钟，而 DINER 仅耗时91 分钟。由于哈希表能够以低频方式映射高频信号，我们的结果展示了粒细胞表面更多细节。DeCAF 的结果中，背景使粒细胞的表面边界过度平滑，因为 PE+多层感知机无法准确地建模高频分量。

7.3 神经辐射场优化

神经辐射场（NeRF）极大地推动了新视角合成的发展。给定从不同位置拍摄的一组稀疏图像，新视角合成旨在合成输入集合中不存在的位置所对应的图像。与传统的深度相关方法相比，NeRF 实现了最先进水平的性能，包括显式深度方法、无深度方法和半深度方法例如。NeRF 最成功的经验之一是使用隐式神经表示（INR），将三维空间中光线的属性建模为神经网络。NeRF 将每条光线的位置 $\vec{x}$ 和方向 $\vec{d}$ 作为隐式神经表示的输入，并输出颜色 c和透射率 σ。随后，对隐式神经表示的输出应用光线步进过程，以实现每个像素颜色的渲染。最后，构建渲染颜色 C 与真实值 C 之间的损失函数，以监督隐式神经表示的训练。整个过程可用以下公式表示

$$
\begin{aligned}
&(c_i, \sigma_i) = INR(\vec{x}_i, \vec{d}_i) \\
&\hat{C}(p) = \sum_i T_i(1 - e^{-\sigma_i \delta_i})c_i, \ \ T_i = e^{-\sum_{j<i} \sigma_j \delta_j} \\
&loss = \sum_p ||C(p) - \hat{C}(p)||_2^2,
\end{aligned}
\tag{16}
$$

其中， δ_i 表示沿光线相邻采样点之间的间隔，通常为常数。

由于场景具有复杂的结构和纹理分布，NeRF 使用了大型多层感知机（8×256 网络结构和位置编码中的10 个傅里叶基），因此训练时间较长。DINER 使用规模为 1×128 的小型多层感知机网络，可显著加快 NeRF 的训练。由于 DINER 只能处理离散信号（关于连续信号的讨论，请参阅会议版本或图 18），我们将连续的三维世界划分为包含 160^3 个点的体素网格，并采用球谐（SH）系数 [50], [51]来建模从不同方向观测到的网格点的颜色方差。基于 DINER 的新视角合成建立在直接体素优化（DVGO） [52]的代码之上。我们将基于 DINER 的新视角合成与原始 NeRF 以及三种近期最先进方法即InstantNGP [30], 、Plenoxels [51] 和 DVGO [52]进行比较。

表 4 和图 16 分别展示了缩小版 Blender 数据集 [3] 在 400×400 分辨率下的定量和定性结果。与采用相同表示形式（体素网格和球谐系数）的 Plenoxels 相比，DINER 在优化球谐系数方面的表现优于 Plenoxels 所用的非神经网络方法。尽管球谐系数的表达能力不如连续方式 [53], ，但基于 DINER 的方法能够取得比原始 NeRF 更好的结果，并与 InstantNGP 和 DVGO 的结果相当，这验证了 DINER 具有良好的表达能力。

球谐系数的冗余性。Yu 等人 [50] 建议将球谐系数设置为一个27维向量，以建模视角相关外观。通过分析不同哈希表宽度下的新视角合成表现，我们发现使用这种27维系数存在冗余。

图 17(a) 绘制了不同哈希表宽度下的峰值信噪比。可以注意到，数值在开始时大幅增加，并且当宽度达到约15时趋于稳定。根据第 5.3 节的分析和第 6.3 节的结果，图 17(a) 中曲线的趋势表明球谐系数是秩亏的。为验证这一结论，当哈希表宽度设置为27时，我们对输出的27维球谐系数应用主成分分析（PCA） [49] 。主成分分析（PCA）是一种降维工具，通过在信号的变换域中选取前-K 个线性无关的主成分，使所选前-K 个成分的贡献能够反映信号的线性依赖性 [49]。图 17(b) 绘制了主成分分析（PCA）中前-K 个维度的贡献。可以注意到，图 17(b) 中贡献曲线的趋势与图 17(a) 中的峰值信噪比曲线相似（即，峰值信噪比和贡献值都在开始时增加，并且当哈希表宽度或前-K 个成分达到约15时趋于稳定），这验证了通过分析不同哈希表宽度下的性能（即图 17(a)）所得出的球谐系数冗余性结论。

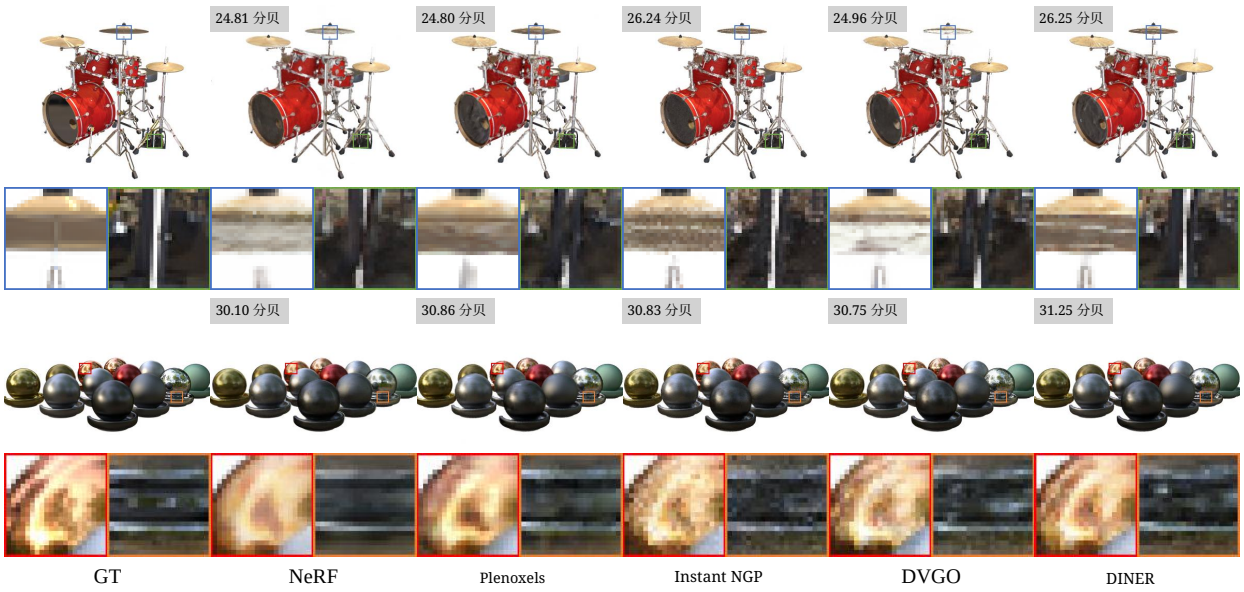
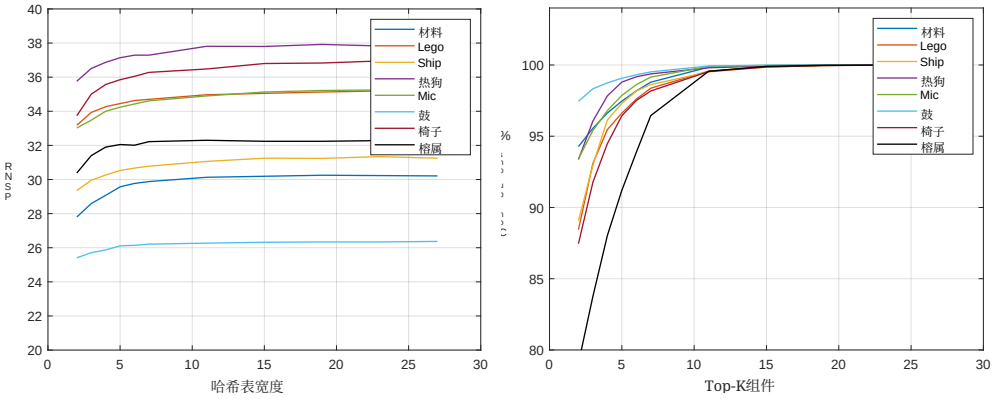
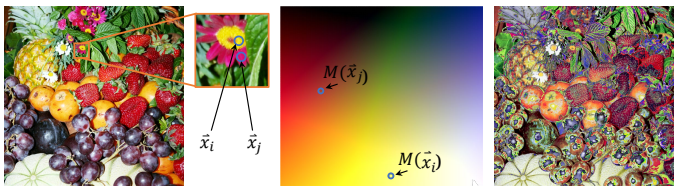


图16. 新视角合成任务中各种方法的定性比较。



(a) 不同哈希表宽度下的峰值信噪比　(b) 前K 个主成分的贡献在球谐系数中

图17. 球谐系数的冗余性。(a) 不同哈希表宽度下新视角合成的峰值信噪比。(b) 以下各项的贡献：前K 个主成分，通过分析球谐系数并使用主成分分析 [49]得到。



（a）原始空间中的相邻像素 原始空间　（b）映射空间中的相邻像素 映射空间　（c）插值结果

图18. 将插值后的哈希键输入DINER的分析。(a) 原始图像中标出了两个相邻坐标 $\vec{x}_i$ 和 $\vec{x}_j$。(b) 映射坐标 $\mathcal{HM}(\vec{x}_i)$ 与 $\mathcal{HM}(\vec{x}_j)$ 之间的距离大于原始空间中的距离。(c) 将插值后的映射坐标输入训练好的多层感知机所得的结果。

7.4 讨论

DINER 是为离散信号设计的。为了查询连续信号中的未见坐标（即，哈希表中没有对应哈希键的坐标），建议应用对网络输出执行后插值操作，而非将插值后的哈希键输入网络（见图 18），例如第 7.3 节的工作将连续辐射场拆分为体素网格，并使用球谐系数对颜色方差进行建模 [51]，而不是直接将未见过的位置和方向坐标输入网络 [3]。

8 结论

在这项工作中，我们提出了 DINER，通过引入额外的哈希表，大幅提升了当前隐式神经表示骨干网络的准确性。我们指出，隐式神经表示表示信号的性能取决于其中元素的排列顺序。所提出的 DINER 可以将输入的离散信号映射为低频信号；只要属性的直方图不变，即使排列顺序发生变化，该低频信号也保持不变。因此，不同

隐式神经表示骨干网络的准确性都可以得到显著提升。此外，我们证明，将哈希表的宽度增加到信号的秩，可以大幅增强 DINER 的表达能力。大量实验验证了所提出的 DINER 在信号拟合和逆问题优化任务中的高准确性与效率。

然而，目前的 DINER 只能处理离散信号。未来，我们将重点研究连续映射方法，而非基于离散哈希表的映射，以拓展其在有符号距离函数 [17] 等连续信号上的优势。

附录A 式（6）的证明

假设参数变量 $x = [x_1, x_2, x_3]^\top$ 是一个三维坐标，那么 $\mathcal{S}_3^\omega$ 中的“超曲面”可以表示为

$$\begin{cases} r = \sum\limits_{\omega' \in \mathcal{H}_\Omega} c^1_{\omega'} \sin(\langle \omega', [x_1, x_2, x_3]^\top \rangle + \phi^1_{\omega'}) \\ g = \sum\limits_{\omega' \in \mathcal{H}_\Omega} c^2_{\omega'} \sin(\langle \omega', [x_1, x_2, x_3]^\top \rangle + \phi^2_{\omega'}) \\ b = \sum\limits_{\omega' \in \mathcal{H}_\Omega} c^3_{\omega'} \sin(\langle \omega', [x_1, x_2, x_3]^\top \rangle + \phi^3_{\omega'}) \end{cases}, \tag{17}$$

其中，$\omega' = [\omega'_1, \omega'_2, \omega'_3]^\top$ 在此处也是一个三维向量。通过设置 $x_3 = 0$，上述函数可简化为以下形式

$$\begin{cases} r = \sum\limits_{\omega' \in \mathcal{H}_\Omega} c^1_{\omega'} \sin(\langle \omega', [x_1, x_2, 0]^\top \rangle + \phi^1_{\omega'}) \\ g = \sum\limits_{\omega' \in \mathcal{H}_\Omega} c^2_{\omega'} \sin(\langle \omega', [x_1, x_2, 0]^\top \rangle + \phi^2_{\omega'}) \\ b = \sum\limits_{\omega' \in \mathcal{H}_\Omega} c^3_{\omega'} \sin(\langle \omega', [x_1, x_2, 0]^\top \rangle + \phi^3_{\omega'}) \end{cases}, \tag{18}$$

它等于 $\mathcal{S}_2^\omega$ 中“超曲面”的参数化函数。因此，

$$\mathcal{S}_2^\omega \subset \mathcal{S}_3^\omega, \tag{19}$$

依此类推，式（6）成立。

参考文献

[1] V. Sitzmann、J. Martel、A. Bergman、D. Lindell 和 G. Wetzstein，“具有周期激活函数的隐式神经表示”，*Advances in Neural Information Processing Systems*，第 33 卷，第 7462–7473 页，2020。[2] J. N. Martel、D. B. Lindell、C. Z. Lin、E. R. Chan、M. Monteiro 和 G. Wetzstein，“ACORN：用于神经场景表示的自适应坐标网络”，*arXiv* 预印本 *arXiv:2105.02788*，2021。[3] B. Mildenhall、P. P. Srinivasan、M. Tancik、J. T. Barron、R. Ramamoorthi 和 R. Ng，“NeRF：将场景表示为用于视图合成的神经辐射场”，载于欧洲计算机视觉会议。施普林格，2020，第 405–421 页。[4] P. Kellnhofer、L. C. Jebe、A. Jones、R. Spicer、K. Pulli 和 G. Wetzstein，“神经光场渲染”，载于*IEEE/CVF*计算机视觉与模式识别会议论文集，2021，第 4287–4297 页。[5] H. Zhu、Z. Liu、Y. Zhou、Z. Ma 和 X. Cao，“DNF：用于无透镜显微成像的衍射神经场”，*Optics Express*，第 30 卷，第 11 期，第 18 168–18 178 页，2022。[6] G. E. Karniadakis、I. G. Kevrekidis、L. Lu、P. Perdikaris、S. Wang 和 L. Yang，“物理信息机器学习”，*Nature Reviews Physics*，第 3 卷，第 6 期，第 422–440 页，2021。[7] A. Tewari、J. Thies、B. Mildenhall、P. Srinivasan、E. Tretschk、W. Yifan、C. Lassner、V. Sitzmann、R. Martin-Brualla、S. Lombardi 等，“神经渲染进展”，载于《Computer Graphics Forum》，第41卷第2期。Wiley Online Library，2022年，第703–735页。[8] C. Shen、H. Zhu、Y. Zhou、Y. Liu、L. Dong、W. Zhao、D. Brady、X. Cao、Z. Ma 和 Y. Lin，“Cardiacfield：面向普遍筛查的计算超声心动图”，Research Square，2023年。[9] R. Liu、Y. Sun、J. Zhu、L. Tian 和 U. S. Kamilov，“利用神经场从离散强度测量中恢复连续三维折射率图”，Nature Machine Intelligence，第4卷第9期，第781–791页，2022年。[10] N. Rahaman、A. Baratin、D. Arpit、F. Draxler、M. Lin、F. Hamprecht、Y. Bengio 和 A. Courville，“论神经网络的频谱偏差”，载于International Conference on Machine Learning。PMLR，2019年，第5301–5310页。[11] M. Tancik、P. Srinivasan、B. Mildenhall、S. Fridovich-Keil、N. Raghavan、U. Singhal、R. Ramamoorthi、J. Barron 和 R. Ng，“傅里叶特征使网络能够学习低维域中的高频函数”，载于Advances in Neural Information Processing Systems，第 33 卷，第7537–7547页，2020年。[12] G. Y¨uce、G. Ortiz-Jim´enez、B. Besbinar 和 P. Frossard，“从结构化字典视角看隐式神经表示”，载于 IEEE/CVF计算机视觉与模式识别会议论文集，2022年，第19 228–19 238页。[13] R. Fathony、A. K. Sahu、D. Willmott 和 J. Z. Kolter，“乘法滤波器网络”，载于 International Conference on Learning Representations，2020年。[14] D. B. Lindell、D. Van Veen、J. J. Park 和 G. Wetzstein，“BACON：用于多尺度场景表示的带限坐标网络”，载于 IEEE/CVF计算机视觉与模式识别会议论文集，2021年。[15] S. Xie、H. Zhu、Z. Liu、Q. Zhang、Y. Zhou、X. Cao 和 Z. Ma，“DINER：无序不变隐式神经表示”，载于IEEE/CVF计算机视觉与模式识别会议论文集，2023年，第1–10页。[16] R. Gao、Z. Si、Y.-Y. Chang、S. Clarke、J. Bohg、L. Fei-Fei、W. Yuan 和 J. Wu，“Objectfolder 2.0：用于仿真到现实迁移的多感官物体数据集”，载于IEEE/CVF计算机视觉与模式识别会议论文集，2022年，第10 598–10 608页。[17] J. J. Park、P. Florence、J. Straub、R. Newcombe 和 S. Lovegrove，“DeepSDF：用于形状表示的连续有符号距离函数学习”，载于 IEEE/CVF计算机视觉与模式识别会议论文集，2019年，第165–174页。[18] V. Sitzmann、S. Rezchikov、B. Freeman、J. Tenenbaum 和 F. Durand，“光场网络：采用单次评估渲染的神经场景表示”，载于 Advances in Neural Information Processing Systems，第34卷，第19 313–19 325页，2021年。[19] P.-H. Yeung、L. Hesse、M. Aliasi、M. Haak、W. Xie、A. I. Namburete 等，“Implicitvol：利用深度隐式表示进行无传感器三维超声重建”，arXiv预印本 arXiv:2109 12108，2021年。。[20] Y. Chen、L. Lu、G. E. Karniadakis 和 L. Dal Negro，“用于纳米光学和超材料逆问题的物理信息神经网络”，Optics Express，第28卷，11 633，2020 8，11 618期，第–页。[21] M. Raissi、A. Yazdani 和 G. E. Karniadakis，“隐式流体力学：从流动可视化中学习速度场和压力场”，Science，第367卷第6481期，第1026–1030页，2020年。[22] B. Reyes、A. A. Howard、P. Perdikaris 和 A. M. Tartakovsky，“学习非牛顿流体的未知物理规律”，Physical Review Fluids，第6卷第7期，第073301页，2021年。[23] M. Leshno、V. Y. Lin、A. Pinkus 和 S. Schocken，“具有非多项式激活函数的多层前馈网络能够逼近任意函数”，Neural Networks，第6卷第6期，第861–867页，1993年。[24] Z. Landgraf、A. S. Hornung 和 R. S. Cabral，“PINS：用于多尺度神经表示的渐进式隐式网络”，载于 International Conference on Machine Learning (ICML)论文集，2022年，第11 969–11 984页。[25] G. Yang、S. Benaim、V. Jampani、K. Genova、J. T. Barron、T. Funkhouser、B. Hariharan 和 S. Belongie，“用于子带分解与操作的多项式神经场”，载于 Advances in Neural Information Processing Systems，第35卷，2022年，第1–15页。[26] L. Liu、J. Gu、K. Zaw Lin、T.-S. Chua 和 C. Theobalt，“神经稀疏体素场”，载于 Advances in Neural Information Processing Systems，第33卷，第15 651–15 663页，2020年。

[27] T. Takikawa、J. Litalien、K. Yin、K. Kreis、C. Loop、D. Nowrouzezahrai、A. Jacobson、M. McGuire 和 S. Fidler，“神经几何细节层次：利用隐式三维形状实现实时渲染”，载于 IEEE/CVF计算机视觉与模式识别会议论文集，2021年，第11 358–11 367页。[28] R. Chabra、J. E. Lenssen、E. Ilg、T. Schmidt、J. Straub、S. Lovegrove 和 R. Newcombe，“深度局部形状：学习用于精细三维重建的局部符号距离函数先验”，载于欧洲计算机视觉会议。施普林格，2020年，第608–625页。[29] C. Jiang、A. Sud、A. Makadia、J. Huang、M. Nießner、T. Funkhouser 等，“用于三维场景的局部隐式网格表示”，载于 IEEE/CVF计算机视觉与模式识别会议论文集，2020年，第6001–6010页。[30] T. M¨uller、A. Evans、C. Schied 和 A. Keller，“采用多分辨率哈希编码的即时神经图形基元”，ACM图形学汇刊（ToG），第41卷第4期，第1–15页，2022年。[31] B. Ronen、D. Jacobs、Y. Kasten 和 S. Kritchman，“神经网络对不同频率学习函数的收敛速率”，载于 Advances in Neural Information Processing Systems，第32卷，2019年。[32] A. Bietti 和 J. Mairal，“论神经切线核的归纳偏置”，载于 Advances in Neural Information Processing Systems，第32卷，2019年。[33] R. Heckel 和 M. Soltanolkotabi，“使用未经训练的神经网络进行压缩感知：梯度下降找到平滑近似”，载于 International Conference on Machine Learning。PMLR，2020年，第4149–4158页。[34] “新的斯坦福光场档案馆”，http://lightfield.stanford.edu/lfs.html。[35] N. Asuni 和 A. Giachetti，“TestImages：用于测试视觉设备和基础图像处理算法的大型档案库”，载于 STAG，2014年，第63–70页。[36] A. Mercat、M. Viitanen 和 J. Vanne，“UVG数据集：用于视频编解码器分析与开发的50/120 fps 4K序列”，载于第11届ACM多媒体系统会议论文集，2020年，第297–302页。[37] T. M¨uller，“Tiny CUDA神经网络框架”，https://github.com/nvlabs/tiny-cuda-nn，2021年。[38] G. J. Sullivan、J.-R. Ohm、W.-J. Han 和 T. Wiegand，“高效视频编码（HEVC）标准概述”，IEEE视频技术电路与系统汇刊，第22卷第12期，第1649–1668页，2012年。[39] H. Chen、B. He、H. Wang、Y. Ren、S. N. Lim 和 A. Shrivastava，“NeRV：视频神经表示”，载于 Advances in Neural Information Processing Systems，第34卷，第21 557–21 568页，2021年。[40] D. Rho、J. Cho、J. H. Ko 和 E. Park，“用于高效视频表示的神经残差流场”，arXiv预印本 arXiv:2201.04329，2022年。[41] S. Kim、S. Yu、J. Lee 和 J. Shin，“具有可学习位置特征的可扩展神经视频表示”，arXiv预印本 arXiv:2210.06823，2022年。[42] Z. Li、M. Wang、H. Pi、K. Xu、J. Mei 和 Y. Liu，“E-NeRV：利用解耦的时空上下文加速视频神经表示”，arXiv预印本 arXiv:2207.08132，2022年。[43] A. Ozcan 和 E. McLeod，“无透镜成像与传感”，《生物医学工程年度评论》，第18卷，第77–102页，2016年。[44] Y. Zhou、X. Hua、Z. Zhang、X. Hu、K. Dixit、J. Zhong、G. Zheng 和 X. Cao，“用于减少无透镜显微成像中高斯噪声的 Wirtinger梯度下降优化”，《工程光学与激光》，第134卷，第106131页，2020年。[45] F. Wang、Y. Bian、H. Wang、M. Lyu、G. Pedrini、W. Osten、G. Barbastathis 和 G. Situ，“利用未经训练的神经网络进行相位成像”，Light-Sci. Appl.，第9卷第1期，第1–7页，2020年。[46] Y. Park、C. Depeursinge 和 G. Popescu，“生物医学中的定量相位成像”，《自然·光子学》，第12卷第10期，第578–589页，2018年。[47] J. Schindelin、I. Arganda-Carreras、E. Frise、V. Kaynig、M. Longair、T. Pietzsch、S. Preibisch、C. Rueden、S. Saalfeld、B. Schmid 等，“Fiji：用于生物图像分析的开源平台”，Nature Methods，第9卷第7期，第676–682页，2012年。[48] J. T. Barron、B. Mildenhall、M. Tancik、P. Hedman、R. Martin-Brualla 和 P. P. Srinivasan，“Mip-NeRF：用于抗锯齿神经辐射场的多尺度表示”，载于 IEEE/CVF 国际计算机视觉大会论文集，2021年，第5855–5864页。[49] C. M. Bishop 和 N. M. Nasrabadi，《模式识别与机器学习》。施普林格，2006年，第4卷第4期。

[50] A. Yu、R. Li、M. Tancik、H. Li、R. Ng 和 A. Kanazawa，“用于神经辐射场实时渲染的 Plenoc-trees”，发表于*IEEE/CVF* 国际计算机视觉大会论文集，2021年，第5752–5761页。[51] S. Fridovich-Keil、A. Yu、M. Tancik、Q. Chen、B. Recht 和 A. Kanazawa，“Plenoxels：无需神经网络的辐射场”，发表于*IEEE/CVF*计算机视觉与模式识别会议论文集，2022，第5501–5510页。[52] C. Sun、M. Sun 和 H.-T. Chen，“直接体素网格优化：辐射场重建的超快速收敛”，发表于*IEEE/CVF*计算机视觉与模式识别会议论文集，2022，第5459–5469页。[53] A. Karnewar、T. Ritschel、O. Wang 和 N. Mitra，“Relu 场：小小非线性，大有可为”，发表于*ACM SIGGRAPH 2022* 会议论文集，2022。



Hao Zhu 是南京大学电子科学与工程学院的副研究员。他于2014年和2020年分别获得西北工业大学的理学学士学位和博士学位。他曾在澳大利亚国立大学担任访问学者。他的研究兴趣包括计算摄影以及逆问题优化。



Shaowen Xie 是南京大学电子科学与技术学院的研究生。他于2021年获得南京大学的理学学士学位。他的研究兴趣包括深度学习和隐式神经表示。



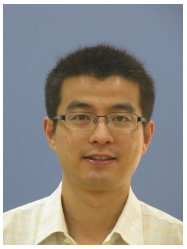
Zhen Liu 是南京大学计算机科学与技术系的研究生，由Xun Cao教授和Yang Yu教授共同指导。他于2021年获得北京理工大学的理学学士学位。他的研究兴趣包括计算摄影和隐式神经表示。



Fengyi Liu 是南京大学电子科学与工程学院的一名研究生。她于2022年获得四川大学理学学士学位。她的研究兴趣包括新视角合成和隐式神经表示。



Qi Zhang 现为腾讯人工智能实验室研究员。他于2021年获得西北工业大学计算机科学学院博士学位。他的研究兴趣包括三维视觉重建、光场成像与处理、多视图几何及其应用。



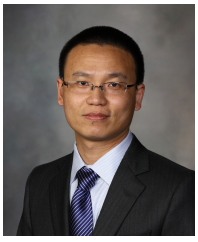
Xun Cao 于2006年获得中国南京南京大学理学学士学位，并于2012年获得中国北京清华大学自动化系博士学位。2008年，他曾在德国亚琛的飞利浦研究院担任访问学者；2009年至2010年，他曾在北京的微软亚洲研究院任职。2010年至2011年，他曾在美国得克萨斯州奥斯汀的德克萨斯大学担任访问学者。他目前是电子科学与工程学院的教授——

究机构为南京大学。他目前的研究兴趣包括计算摄影、基于图像的建模与渲染。



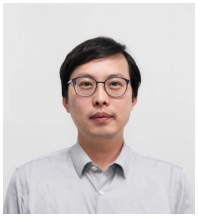
You Zhou 是中国南京大学电子科学与工程学院的副研究员。他分别于2019年1月和2012年7月获得清华大学自动化系博士学位和华东师范大学通信工程系理学学士学位。2016年12月至2017年5月，他曾在康涅狄格大学生物医学工程系担任交流研究员。他的研究方向包括

计算显微术和计算光学。他是 IEEE 会员。



YiLin Yi Lin 医生是复旦大学附属中山医院的心血管外科医生和副教授。他在复旦大学上海医学院接受医学教育，随后在复旦大学附属中山医院完成了普通外科、心血管外科和胸外科住院医师培训。临床培训结束后，他在罗切斯特的梅奥诊所担任一年外科研究员，随后又在罗切斯特的梅奥诊所接受高级心血管外科培训。之后他返回

他于2016年加入复旦大学附属中山医院，现任心血管疾病主治外科医生、副教授。他的主要研究兴趣包括心血管疾病的创新医学成像技术和机器人辅助治疗。



Zhan Ma （SM'19）现任南京大学电子科学与工程学院教师，地址为中国江苏南京，邮编210093。他于2004年和2006年分别获得华中科技大学学士和硕士学位，并于2011年获得纽约大学博士学位。2011年至2014年，他先后在三星美国研究院（达拉斯，得克萨斯州）和Future wei Technologies, Inc.（圣克拉拉，加利福尼亚州）任职。他的研

他的研究重点是基于学习的视频编码和智能相机。他是以下奖项的共同获奖者：2018年 PCM 最佳论文入围奖、2019年 IEEE 广播技术协会最佳论文奖、2020年 IEEE 多媒体信号处理图像压缩大挑战赛最佳表现方案，以及2023年 IEEE WACV 最佳算法论文奖。