arXiv:2602.12255v1 [physics.optics] 12 Feb 2026

用于相干衍射成像中多畴相位恢复的视觉Transformer

Jialun Liu[1], David Yang[2], Ian Robinson[1,2]

[1]伦敦大学学院伦敦纳米技术中心，英国伦敦 WC1E 6BT。[2]布鲁克海文国家实验室凝聚态物理与材料科学系，美国纽约州厄普顿，邮编 11973。

摘要

在强相位区间，布拉格相干衍射成像（BCDI）的相位恢复会迅速变得困难，此时晶体中的畸变超过半个晶格间距。一个重要的特殊情形是相位畴问题，即晶体的各个块体发生位移，并在畴壁处出现突变。本文将超过 $\pm\pi/2$ 定义为强相位；它会产生分裂的布拉格峰和密集的条纹结构，经典迭代求解器对此往往停滞不前，或因初始化不同而得到不同的解。本文提出一种无监督傅里叶视觉Transformer（傅里叶视觉Transformer），直接从测得的二维布拉格衍射强度中求解这一块相位多畴相位恢复问题。傅里叶视觉Transformer通过多尺度傅里叶令牌混合，在全局范围内耦合倒易空间信息；同时，浅层卷积前后端负责局部滤波与重建。我们在大规模合成数据集上验证了该方法，这些数据集包含具有强相位衬度、并受到真实噪声扰动的Voronoi多畴晶体；此外，还在来自一颗 $\mathbf{La_{2-x}Ca_xMnO_4}$ 纳米晶体的实验衍射数据上进行了验证。在所考察的各个条件下，与所比较的方法相比，傅里叶视觉Transformer取得了最低的倒易空间失配（$\boldsymbol{\chi^2}$），并能在畴数增加时保留分畴相位重建。对于实验数据，在使用相同实空间支撑域的情况下，傅里叶视觉Transformer达到与迭代基准 $\boldsymbol{\chi^2}$ 相当的效果，同时提高了对随机初始化的鲁棒性，因此获得低-$\boldsymbol{\chi^2}$重建的成功率高于复值卷积神经网络基线。

关键词：视觉Transformer；布拉格相干衍射成像；相位恢复；多畴

1 引言

布拉格相干衍射成像（BCDI）通过以三维方式揭示单晶的内部结构和晶格畸变，改变了我们表征纳米材料的方式（Miao 等人，1999年；Robinson 和 Harder，2009；Pfeifer 等人，2006）。这项无需透镜的X射线技术可通过快速傅里叶变换迭代施加倒空间和实空间约束，将布拉格峰反演至实空间（Fienup，1982年）。BCDI重建揭示了单个铂纳米催化剂中的边缘应变如何影响甲烷氧化活性（Kim 等人，2018），纳米应变累积如何导致富锂层状氧化物正极的结构退化和电压衰减（Liu 等人，2022年），以及电场如何影响 $BaTiO_3$ 纳米颗粒中的畴壁动力学（Liu 等人，2025）。然而，其核心挑战仍在于相位恢复：探测器只能记录衍射强度，因此缺少实空间重建所需的相位（Sayre，1952；Fienup，1982年）。

仅凭强度恢复相位是一个非凸、病态的逆问题，其难度会随着内部应变和畴复杂性的增加而急剧上升。经典算法包括 Gerchberg–Saxton 误差约减法（Gerchberg 和 Saxton，1972）、混合输入输出法（HIO）（Fienup，1982）、差分映射（Elser，2003）和松弛平均交替反射法（RAAR）（Luke，2005）；这些算法在实空间和倒易空间之间交替，并施加支撑域和测量强度等约束。对于“弱相位”晶体，当相位移低于约 $\pi/2$ 时，这些方法通常可靠，并能在数百次迭代内收敛（Robinson 和 Harder，2009）。一旦相位移超出这一范围，不同相位区域之间的干涉便会导致峰分裂和复杂的条纹图样，这一现象被称为“强相位”问题（Robinson 等人，2020；Wu 等人，2021a；Gao 等人，2021）。中心布拉格极大值可能分裂成多个瓣，在布拉格峰周围形成相互分离的强度极大值和密集条纹，使测得的衍射对内部相位分布极其敏感。在这种强相位多畴状态下，迭代求解器会陷入停滞，或因随机种子不同而收敛到不同解，反映出重建似乎并非唯一（Bates，1984；Ulvestad 等人，2017；Carnis 等人，2019）。这种缓慢且不稳定的收敛限制了布拉格相干衍射成像的常规应用，也使原位或工况实验中的实时反馈难以实现，尤其是在X射线自由电子激光器（XFEL）和现代同步辐射光源上（Clark 等人，2013；Harder，2021）。

在理想条件下，关于相位问题唯一性的理论研究表明，在两个或更多维度上进行过采样，足以保证解在实空间平移、共轭反演、全局相位偏移和菲涅耳传播等平凡对称变换下唯一（Bates，1984；Bruck 和 Sodin，1979；Zhuang 等人，2022年；Huang 等人，2011）。相比之下，一维问题通常存在许多彼此不等价、但具有相同傅里叶幅值的解（Beinert 和 Plonka，2015；Bendory 等人，2017年）。受此启发，我们聚焦于一种二维表述，其中每个合成晶体都被视为三维纳米晶体的投影，从而保留完整三维问题的核心傅里叶相位结构。遵循 Bates（Bates，1984），我们认为二维和三维相位问题是等价的。

我们聚焦于与多畴晶体相关的相位畴极限。在布拉格相干衍射成像中，实空间相位 $\phi(\mathbf{r}) = \mathbf{Q}\cdot\mathbf{u}(\mathbf{r})$ 是晶格位移场 $\mathbf{u}(\mathbf{r})$ 在所测布拉格峰的倒易晶格矢量 $\mathbf{Q}$ 上的投影（Pfeifer 等人，2006；Robinson 和 Harder，2009）。我们将相位畴定义为一个连通的块状区域，其中投影相位近似恒定，并由狭窄畴壁上的尖锐相位不连续与相邻区域分隔。随着畴的数量增加，畴间的相干干涉会产生更强的布拉格峰分裂和更丰富的条纹结构，使相位恢复景观变得更加复杂（Robinson 等人，2020）。这类相位畴结构纹理存在于真实的量子材料中，例如，

布拉格相干衍射成像已用于成像铜氧化物超导体 $La_{1.875}Ba_{0.125}CuO_4$ (Assefa 等人，2020)中低温正交畴的形成和尺度变化，以及磁铁矿 Fe_3O_4 (Dong 等人，2025)中的单斜畴。

实际上，实验数据可能包含噪声，而支撑域或非负性等约束也只能近似满足（Fienup，1987）。对于含噪数据，相位问题不再具有唯一的精确解。相反，得到的是一组近似解，其分散程度同时反映了噪声和算法陷入局部解的影响（BatesandNapier，1972）。这促成了集成策略的发展：运行许多相互独立的重建，然后根据评价指标对其进行筛选或平均。相关示例包括引导式混合输入输出程序（Ulvestad et al.，2017；Chen et al.，2007）以及在 PyNX 中实现的共模（特征解）分析（Lee and Seung，1999），该方法利用自由对数似然指标组合多个布拉格相干衍射成像重建（Favre-Nicolin et al.，2020b，a）。包括傅里叶加权投影在内的其他投影方案（Guizar-Sicairos and Fienup，2008）提高了截断衍射数据处理的稳健性，但仍属于迭代投影算法，并且仍可能对初始化较为敏感。总体而言，多畴晶体的相位恢复仍然缓慢、脆弱且难以实现自动化。

与此同时，深度学习正日益被用于解决相干成像问题。监督式神经网络学习从衍射强度到实空间结构的逆映射，与迭代相位恢复相比，能够实现近乎实时的重建（Cherukara等，2018；Wu等，2021b，2025）。采用U-Net架构的卷积神经网络已应用于光学全息术（Rivenson等，2018）、X射线叠层成像（Cherukara等，2020）和相干X射线衍射成像（CXDI）（Vu等，2025），包括布拉格相干衍射成像（Wu等，2021a，b；Yu等，2024）。训练完成后，这些网络的重建速度可比经典迭代求解器快几个数量级（Cherukara等，2018，2020；Yu等，2024）。监督式方法的主要局限在于，它们只能学习训练集中提供的内容，因此无法保证对训练分布之外的新对象给出正确解；尤其是在实验衍射图样没有真实标签时，这一问题更为突出（Cherukara等，2018；Wu等，2021b；Yao等，2022）。由于在简单条件下，标签生成和验证最为直接，许多已发表的演示和基准测试都侧重于单畴至数畴系统（Wu等，2021b）。无监督和物理信息驱动的变体，例如用于超快布拉格相干衍射成像的AutoPhaseNN和深度复值卷积神经网络，会将正向成像模型直接嵌入网络和损失函数，从而降低对带标签训练数据的依赖（Yao等，2022；Yu等，2024）。这种无监督设置尤其适用于布拉格相干衍射成像，因为实验衍射图样没有可用的真实实空间相位，而且强相位多畴结构会使优化景观高度非凸，并对初始化十分敏感（Wu等，2021b；Yao等，2022）。视觉变换器（ViTs）近期已应用于相干X射线成像，例如叠层成像（Gan等，2024），但尚未被整合进完全无监督、物理信息驱动的布拉格相干衍射成像框架。

在此，我们引入一种简单的块相位域模型，以探索强相位多畴状态，并利用一种具有全局倒易空间令牌混合的无监督傅里叶视觉Transformer，解决由此产生的相位恢复问题。所提出的傅里叶视觉Transformer将卷积特征提取与谱令牌混合相结合，以 $\mathcal{O}(N \log N)$ 复杂度实现全局耦合，与标准Transformer 中完整点积自注意力的 $\mathcal{O}(N^2)$ 计算开销形成对比（Vaswani 等人，2017年）。从无参数傅里叶混合（FNet）到可学习的全局频率滤波器（GFNet）和自适应傅里叶算子（AFNO/FNO），傅里叶令牌混合已在各种模态中得到探索（Lee-Thorp 等人，2022年；Rao 等人，2021；Guibas 等人，2021；Li 等人，2021）。在此，我们将这些思路应用于

通过针对布拉格相干衍射成像（BCDI）图样结构设计的多尺度频谱方案实现过采样布拉格衍射。

我们首先在大规模 Voronoi多畴晶体合成数据集上验证了傅里叶视觉Transformer。在所研究的条件范围内，傅里叶视觉Transformer在所比较的方法中实现了最低的倒易空间失配（χ^2），同时在现实的噪声退化下保留了分畴相位重建。我们进一步将该方法应用于 $La_{2-x}Ca_xMnO_4$ (LCMO) 的实验衍射数据（Robinson 等人，2026年）。在此，傅里叶视觉Transformer优于 C-CNN 基线（Yu 等人，2024年），并实现了与迭代基准相当的 χ^2 值。不同运行之间的 χ^2 分布在不同随机初始化下更为分散，凸显了强相位恢复具有多盆地特性。

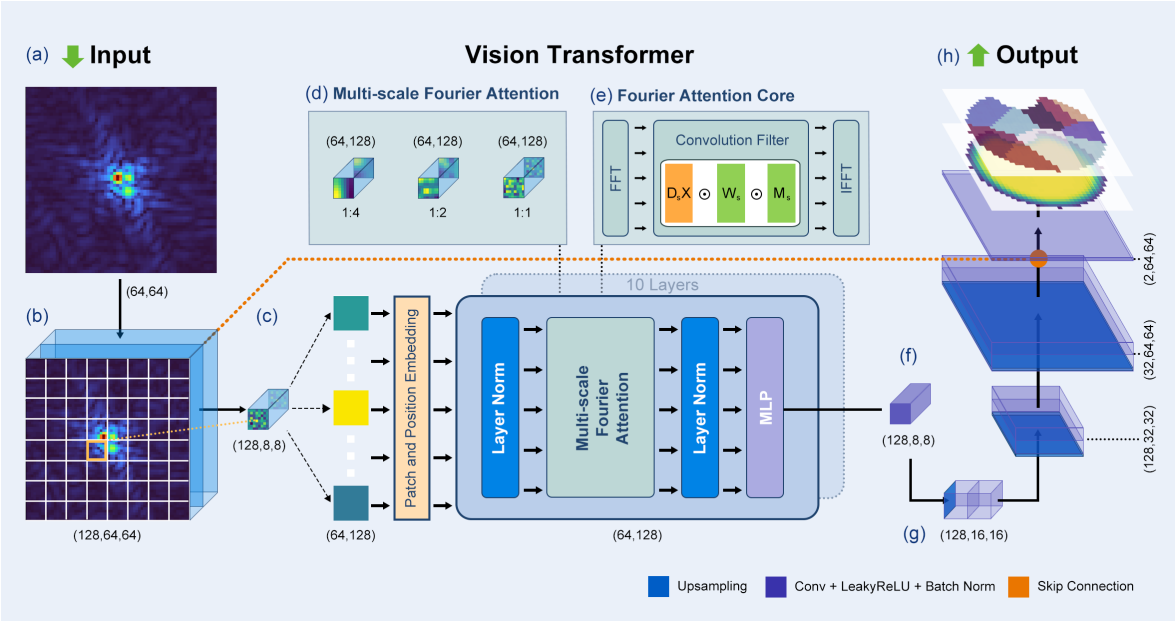


图 1：所提出的用于 BCDI相位恢复 的傅里叶视觉Transformer模型架构。(a) 输入衍射幅值（64×64 像素）经由 (b) 浅层卷积特征提取器处理，生成 128 个通道（$128 \times 64 \times 64$）。(c) 每幅图像被划分为图像块并嵌入为词元（为清晰起见，图中显示 8×8 个图像块；所有实验均使用 16×16 个图像块），形成形状为 (64×128) 的令牌序列。该序列由 (d) 多层视觉Transformer处理，该Transformer采用多尺度傅里叶注意力（在 1:4、1:2 和 1:1 尺度进行频率混合），并结合 (e) 基于FFT的全局卷积模块。(f) Transformer输出被重塑为潜在特征图（$128 \times 8 \times 8$），并由 (g) 带有跳跃连接的卷积上采样路径解码，以重建复晶体场，得到 (h) 64×64 实空间幅度和相位图，且与测得的衍射图样一致。

2 结果

2.1 傅里叶视觉Transformer模型架构

2.1.1 词元化与位置嵌入

如图 1(a) 所示，输入为一个包含 64×64 像素的二维衍射图样幅度，其最大值归一化为 $[0,1]$。在图 (b) 中，一个浅层卷积神经网络在原始 64×64 分辨率下提取特征，并输出一个 128 通道特征图。我们将其保留为全分辨率跳跃映射 S $\in \mathbb{R}^{B\times128\times64\times64}$，其中 B 表示批量大小。图 1(c) 中的图像块嵌入采用单个 4×4 卷积，步幅为 4，作用于 S，将图像划分为不重叠的 4×4 图像块，并将每个图像块投影为嵌入维度为 E=128的词元。由此生成一个 16×16 词元网格，总计包含 N=16×16=256 个词元。我们的数据集中，经过分箱后典型的散斑/条纹间距约为两个像素，因此一个 4×4图像块至少跨越一个完整的条纹周期，使每个词元都能编码衍射图样中具有物理意义的局部片段。按照视觉Transformer的构造方式（Dosovitskiy 等，2021；Vaswani 等人，2017），我们随后将 16×16 网格展平为词元序列，并通过两个 1×1 卷积进行细化，然后添加学习得到的位置嵌入。最终得到的序列 $\mathbf{X} \in \mathbb{R}^{B\times N\times E}$ 被输入第 2.1.2 节所述的傅里叶 Transformer 块堆栈。

2.1.2 多尺度傅里叶注意力

受基于傅里叶变换的 Transformer 和频谱门控机制启发（Lee-Thorp 等人，2022年；Patro 和 Agneeswaran，2023年），我们使用多尺度傅里叶注意力替代点积自注意力（Vaswani 等人，2017年），如图1 （d、e）中的绿色方框所示。词元被重塑为特征图 $\mathbf{X} \in \mathbb{R}^{B\times E\times H\times W}$，其尺寸为 H=W=16，并在三个空间尺度 $s \in \{1,2,4\}$ 上进行处理:

$$\mathbf{Z}_s = \mathcal{U}_s\Big(\mathcal{F}^{-1}\big(\,\mathcal{F}(\mathcal{D}_s\mathbf{X})\;\odot\;\mathbf{W}_s\;\odot\;\mathbf{M}_s\,\big)\Big), \tag{1}$$

此处， $\mathcal{D}_s$ 表示步幅为 s 的平均池化，它将空间尺寸缩减至 $h_s\times w_s$ =（H/s）×（W/s）；$\mathcal{U}_s$ 则将其上采样回（H，W）。算子 $\mathcal{F}$表示逐通道二维快速傅里叶变换，$\odot$表示在频域中的逐元素乘法。实际应用中，我们在分辨率（H，W）、（$H/2$，$W/2$）和（$H/4$，$W/4$）下使用三个混合器，并通过尺度权重 α_s 学习它们的相对重要性，得到 $\mathbf{Z}$ = $\sum_s \alpha_s\mathbf{Z}_s$，其中 $\boldsymbol{\alpha}$ = 为 softmax（$\boldsymbol{\theta}$）。在每个尺度上，我们学习逐通道频率响应 $\mathbf{W}_s \in \mathbb{R}^{E\times h_s\times w_s}$ 和通道共享的频谱门控 $\mathbf{M}_s \in \mathbb{R}^{1\times h_s\times w_s}$。这种显式的多尺度设计使模块既能响应间距较小的条纹，也能响应较宽泛的低频分量。

我们在每个傅里叶变换器块中使用预归一化残差连接:

$$\mathbf{X}' = \mathbf{X} + \mathrm{Mixer}\big(\mathrm{LN}(\mathbf{X})\big), \qquad \mathbf{X}_{\mathrm{out}} = \mathbf{X}' + \mathrm{MLP}\big(\mathrm{LN}(\mathbf{X}')\big), \tag{2}$$

其中 LN 表示层归一化（Ba 等人，2016；Dosovitskiy 等人，2021），Mixer 表示上述定义的多尺度傅里叶频谱混合器，而多层感知机（MLP）是一种隐藏维度为 d_{mlp} 且采用 ReLU 非线性的前馈投影。残差路径有助于改善梯度流，而预归一化则能稳定更深层块堆叠的优化过程。更多细节，包括卷积定理和复杂度分析，见补充说明1。

2.1.3 解码器

经过最后一个 Transformer 块后，将输出序列 $\mathbf{X} \in \mathbb{R}^{B\times N\times E}$（其中 $N{=}16\times 16$ 和 $E{=}128$）重塑为 $E\times 16\times 16$ 特征图，如图1 (f) 所示，并通过小型 CNN 解码器上采样至 64×64（图1 (g)）。在全分辨率下，解码器融合三种输入：（i）上采样后的词元特征；（ii）编码器跳跃映射 $\mathbf{S}$ $\in \mathbb{R}^{B\times 128\times 64\times 64}$，它携带来自 CNN 特征提取器的局部衍射图样上下文；（iii）$\mathbf{S}$ 的频域摘要，该摘要通过计算其逐通道傅里叶幅值、应用 $\log(1+|\mathcal{F}\{\mathbf{S}\}|)$ 压缩动态范围，并按通道进行归一化获得（详情见补充说明 2）。三个分支沿通道维度拼接后，输入 3×3 卷积以混合通道，随后经过一个简短的细化堆栈。这样的融合为解码器提供互补信息：（i）来自 Transformer 的非局部衍射相关性；（ii）来自编码器的精细散斑特征；（iii）学习到的衍射特征的显式频谱摘要。

解码器头输出实值晶体振幅图 $s(x,y)$），用于参数化晶体相位。振幅头采用 Softplus 非线性函数，以确保输出非负；振幅和相位均与固定的实空间支持掩膜 $S(x,y)$ 相乘，符合相干X射线衍射成像（CXDI）的标准支撑约束（Fienup，1982年；Marchesini 等，2003年）。对于实验数据，傅里叶视觉Transformer使用与 MATLAB 迭代重建相同的支撑域（见第 4.4 节），以进行公平比较。在支撑域内，预测振幅通过标量调度函数 $\alpha(t)$ 与简单先验混合，随后进行柔性重缩放，使其 ℓ_2 范数与先验振幅的范数匹配；具体的混合和归一化规则见补充说明 2。对于相位，首先将两个解码器输出归一化至单位圆上，$(\tilde{c},\tilde{s}) = (\hat{c},\hat{s})/\sqrt{\hat{c}^2+\hat{s}^2}$，然后通过 $\phi(x,y) = \mathrm{atan2}(\tilde{s}(x,y)\ \tilde{c}(x,y))$ 恢复相位。——

给定输入衍射图样幅值 $M(\mathbf{q})$，网络因此会在 64×64 网格上输出复数实空间密度 $\rho(x,y)$，其振幅和相位受固定支撑域约束。

2.1.4 损失函数

对于给定的输入衍射图样，网络根据晶体的振幅和相位预测复数实空间密度 $\rho(x,y)$。随后，通过布拉格相干衍射成像前向模型传播该估计，得到预测衍射幅值 $M_{\mathrm{pred}}(\mathbf{q})$（第2.1.3节），并通过混合傅里叶空间损失将其与测量幅值 $M_{\mathrm{meas}}(\mathbf{q})$进行比较。该损失结合了皮尔逊相关系数（PCC）、均方根（RMS）归一化的 χ^2、幂加权的 χ^2 项以及较小的全变分（TV）正则项，并采用随训练轮次变化的权重；随着训练推进，权重会将关注重点从整体图样相关性转向精细尺度上的强度一致性。

对于 χ^2类项，我们使用均方根归一化幅值 $\hat{M}(\mathbf{q}) = M(\mathbf{q})/\sqrt{\langle M(\mathbf{q})^2\rangle_{\mathbf{q}}}$，分别对 M_{meas} 和 M_{pred} 进行处理，以去除任意的整体强度尺度。在训练轮次 t，我们最小化：

$$L_{\mathrm{total}}(t) = w_{\mathrm{PCC}}(t)\,L_{\mathrm{PCC}} + w_{\chi^2}(t)\,L_{\chi^2} + w_{\mathrm{p}\chi^2}(t)\,L_{\mathrm{p}\chi^2} + \lambda_{\mathrm{TV}}(t)\,L_{\mathrm{TV}}, \tag{3}$$

其中，数据保真权重 $w_{\mathrm{PCC}}(t)$、$w_{\chi^2}(t)$、$w_{\mathrm{p}\chi^2}(t)$ 和全变分权重 $\lambda_{\mathrm{TV}}(t)$ 遵循简单的依赖于训练轮次的调度策略。其设计目的是使 PCC 在训练早期占主导地位，功率–χ^2 在训练初期逐步增加，以增强低频，而在后期细化阶段略微偏重标准 χ^2，同时全变分始终作为较小的稳定项。完整的调度定义见补充说明 3。

(i) PCC 损失。为确保衍射图样的整体相似性，我们使用

$$L_{\mathrm{PCC}} = 1 - \rho\big(M_{\mathrm{meas}}, M_{\mathrm{pred}}\big), \tag{4}$$

对于单个图样,

$$\rho\big(M_{\mathrm{meas}}, M_{\mathrm{pred}}\big) = \frac{\sum_{\mathbf{q}} \big(M_{\mathrm{meas}}(\mathbf{q}) - \mu_{\mathrm{meas}}\big)\big(M_{\mathrm{pred}}(\mathbf{q}) - \mu_{\mathrm{pred}}\big)}{\sqrt{\sum_{\mathbf{q}} \big(M_{\mathrm{meas}}(\mathbf{q}) - \mu_{\mathrm{meas}}\big)^2}\,\sqrt{\sum_{\mathbf{q}} \big(M_{\mathrm{pred}}(\mathbf{q}) - \mu_{\mathrm{pred}}\big)^2}}, \tag{5}$$

其中 $\mu_{\mathrm{meas}} = \langle M_{\mathrm{meas}}(\mathbf{q})\rangle_{\mathbf{q}}$, $\mu_{\mathrm{pred}} = \langle M_{\mathrm{pred}}(\mathbf{q})\rangle_{\mathbf{q}}$。

(ii) 均方根归一化 χ^2 损失。为惩罚绝对误差，同时对残余比例因子保持不敏感，我们定义

$$L_{\chi^2} = \frac{\sum_{\mathbf{q}} \big[\hat{M}_{\mathrm{meas}}(\mathbf{q}) - \hat{M}_{\mathrm{pred}}(\mathbf{q})\big]^2}{\sum_{\mathbf{q}} \hat{M}_{\mathrm{meas}}(\mathbf{q})^2}, \tag{6}$$

ˆ ˆ 其中 M_{meas} 和 M_{pred} 是上述定义的均方根归一化幅值。

(iii) 幂χ^2 损失。为进一步强调明亮且具有高q 的条纹，我们加入一个幂加权 χ^2 项。从均方根归一化幅值出发，我们构造测量和预测衍射图样的幂次幅值，ˆ$M^{(p)}_{\mathrm{meas}}(\mathbf{q}) = \big[\,\hat{}M_{\mathrm{meas}}(\mathbf{q})\big]^{p(t)}$ 和 ˆ$M^{(p)}_{\mathrm{pred}}(\mathbf{q}) = \big[\,\hat{}M_{\mathrm{pred}}(\mathbf{q})\big]^{p(t)}$，然后定义

$$L_{\mathrm{p}\chi^2} = \frac{\sum_{\mathbf{q}} \big[\hat{M}^{(p)}_{\mathrm{meas}}(\mathbf{q}) - \hat{M}^{(p)}_{\mathrm{pred}}(\mathbf{q})\big]^2}{\sum_{\mathbf{q}} \big[\hat{M}^{(p)}_{\mathrm{meas}}(\mathbf{q})\big]^2}. \tag{7}$$

指数 $p(t)$ 从训练早期的约 2 平滑衰减至后期的 0.5，因此优化一开始侧重于拟合最明亮的区域，并逐渐增加对较弱条纹的权重。

（iv）全变分幅度正则化。令 $S(x, y) \in \{0, 1\}$ 为二值支撑掩模，$N_S = \sum_{x,y} S(x, y)$ 为支撑像素的数量。我们定义一个各向异性全变分惩罚

$$L_{\mathrm{TV}} = \frac{1}{N_S} \sum_{x,y} S(x, y)\,\big(|\nabla_x A(x, y)| + |\nabla_y A(x, y)|\big), \tag{8}$$

其中 $\nabla_x A(x,\ y) = A(x + 1,\ y) - A(x,\ y)$ 和 $\nabla_y A(x,\ y) = A(x,\ y + 1) - A(x,\ y)$。在实现中，我们先用 $S(x,\ y)$ 对振幅进行掩蔽，再计算有限差分，因此支撑边界处的梯度不会产生影响。

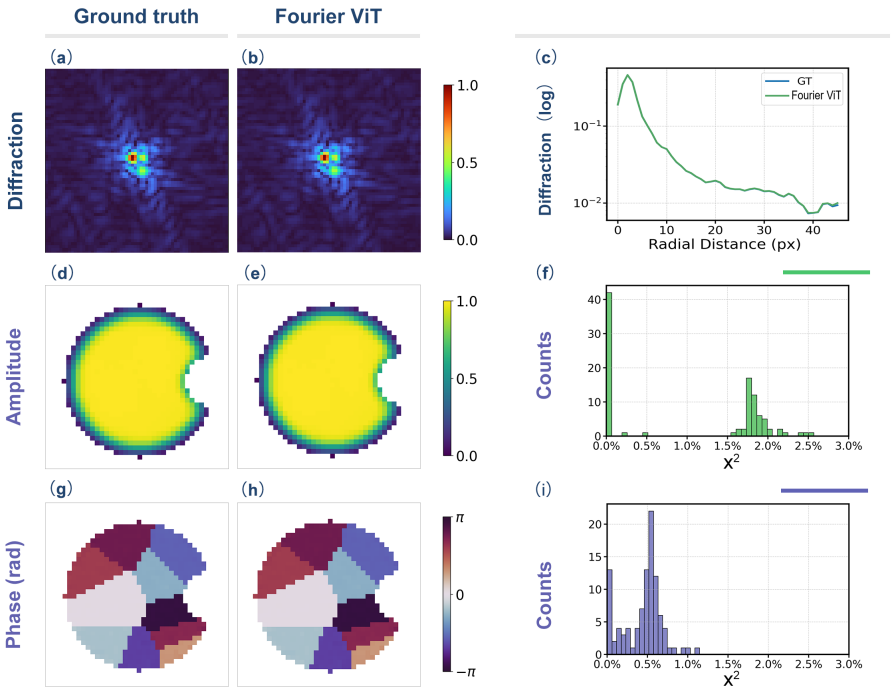


图 2：仅根据衍射对合成 10 域晶体进行傅里叶 ViT 重建。(a、d、g) 真实值衍射幅值（64×64 像素）、晶体振幅和相位（40×40 像素）；(b、e、h) 对应的傅里叶 ViT 重建。(c) 真实值和重建衍射的径向对数强度与 q 剖面对比，显示两者在整个 q范围内吻合。(f、i) 已知晶体振幅（f）和未知振幅（i）的相位恢复中，100 次随机初始化所得衍射 χ^2 的直方图。

2.2 合成数据上的性能

对于合成多域晶体，傅里叶视觉Transformer能够从倒易空间衍射幅值中恢复实空间物体，无需实空间监督或支撑域更新。在 64×64 衍射分辨率下，若已知晶体振幅，该方法在能够实现完美收敛（此处定义为 $\chi^2\leq10^{-5}$）的运行中最多可解析 19 个畴。对于 10 个畴，100 次运行中有 42 次实现完美重建；对于 15 个畴，有 18 次；对于 19 个畴，有 4 次。联合恢复晶体振幅和相位时，收敛更难，但仍然可行：对于 10 域晶体，100 次运行中有 12 次在经过 1000 个训练轮次后达到 $\chi^2\leq10^{-5}$。

为展示物体振幅-相位恢复，我们在图2中展示一个具有代表性的 10 畴案例。图2 (a,b) 对比了真实值与傅里叶视觉Transformer预测的衍射幅值，而图2(c) 展示了衍射图样的径向平均对数强度剖面随动量转移 q （倒易空间中的径向坐标）变化的情况。该剖面在整个 q 范围内均与真实值吻合，包括由强相位对比和锐利畴壁产生的高q 尾部。与弱相位情形中常见的单个中心峰不同，该图样具有多个中心

，因为物体的相位值变化很大。图2 (d,e) 表明，恢复出的晶体振幅与真实值平滑吻合；图2(g,h) 展示了相位解恢复出的实空间畴构型，其边界清晰、位置准确，且没有伪畴。真实值图仅用于定量评估和可视化，因为网络是在没有实空间标签的情况下，基于衍射数据进行无监督训练的。为便于视觉比较，我们将所有重建中中心像素处的相位设为零，以固定全局相位偏移。

图 2（f）展示了仅相位预测情况下 100 次随机初始化的双峰分布。许多运行收敛到具有 $\chi^2 \leq 10^{-5}$ 的全局最小值解，而其余运行则停留在误差更高的局部极小值中，在 $\chi^2 \approx$（1。5–2。5）$\times 10^{-2}$ 处形成第二个簇。这反映了非凸的强相位恢复景观：畴壁位置或拓扑的微小变化，可能会在高q 条纹场中产生结构化失配。在振幅固定的情况下，优化可调整的自由度有限，因此这些解会保留稳定的非零残差。在图2（i）所示的联合振幅-相位恢复中，振幅预测带来的额外自由度减少了完全收敛的运行次数，但使其余结果集中在更窄的较低χ^2范围内。在此情况下，部分不正确的相位图可通过支撑域内的振幅重新加权进行补偿，从而更好地匹配衍射幅值。

这些结果表明，采用傅里叶注意力进行倒易空间建模，可以再现衍射结构，并在没有实空间标签的情况下恢复多畴相位拓扑。除非另有说明，所有傅里叶视觉Transformer模型均采用无监督方式训练，使用第4.2节所述的优化和损失调度方案。

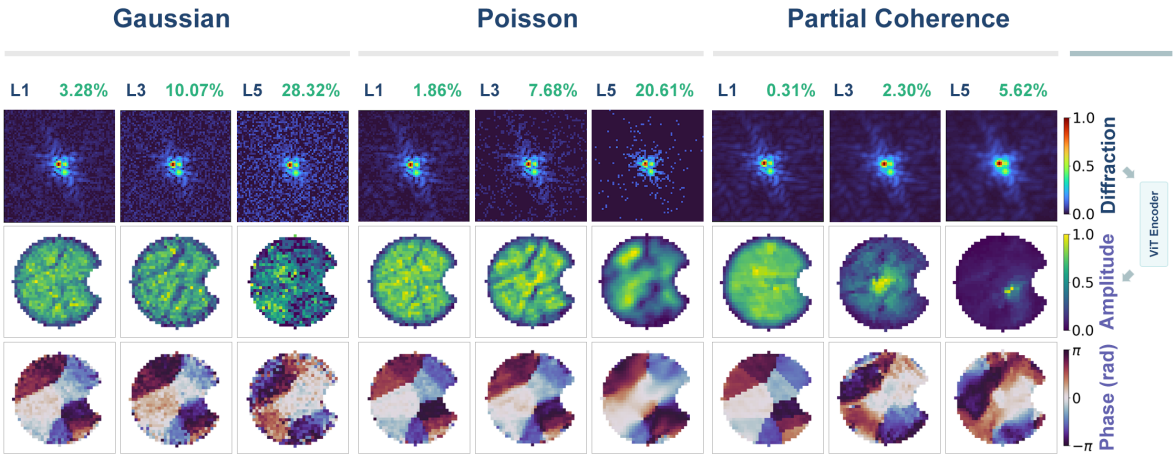


图 3：傅里叶视觉Transformer 的抗噪性。同一衍射图样分别施加三种噪声：高斯噪声、泊松噪声和部分相干性。对于每个模型，L1、L3 和 L5 表示逐渐增强的噪声强度（等级 1、3 和 5）；绿色百分比表示 χ^2_{n}（以 % 计），即相对于干净真值的输入衍射误差。顶部一行显示含噪输入的衍射振幅（64×64像素）。中间和底部两行分别显示放大至 40×40像素的重建晶体振幅和相位；二者均经过支撑域掩膜处理，且相位单位为弧度。

表1：傅里叶视觉Transformer（幅度域）的噪声鲁棒性。所有数值均为最大值归一化的 χ^2 误差（%）。 χ^2_{n} 比较含噪输入与干净真值；$\chi^2_{\mathrm{rec,\ n}}$ 和 $\chi^2_{\mathrm{rec,\ c}}$ 分别比较重建与含噪输入以及干净真值。噪声强度从等级1逐渐增加到等级5。

	高斯噪声			泊松噪声			部分相干性		
噪声	χ^2_{n}	$\chi^2_{\mathrm{rec,n}}$	$\chi^2_{\mathrm{rec,c}}$	χ^2_{n}	$\chi^2_{\mathrm{rec,n}}$	$\chi^2_{\mathrm{rec,c}}$	χ^2_{n}	$\chi^2_{\mathrm{rec,n}}$	$\chi^2_{\mathrm{rec,c}}$
等级1	3.28	1.76	1.54	1.86	1.18	0.92	0.31	0.13	0.25
等级 2	6.55	3.48	2.71	4.15	2.67	1.92	1.50	0.25	1.94
等级 3	10.07	5.16	4.14	7.68	4.69	3.65	2.30	0.23	2.78
等级 4	18.84	9.03	8.21	12.90	8.07	5.82	3.95	0.17	4.66
等级 5	28.32	12.49	14.36	20.61	11.91	10.53	5.62	0.11	6.37

对于同一 10 域晶体的模拟衍射图样和固定支撑域，结合误差缩减（ER）与混合输入输出（HIO）的迭代方法在经过 500 次迭代后，约半数重建的 χ^2 值介于10^{-4} 和 10^{-3} 之间。采用迭代方法时，没有任何一次运行达到我们的“完美”标准 $\chi^2 \leq 10^{-5}$；但将迭代次数增加到 2000 次后，$10^{-4} \leq \chi^2 \leq 10^{-3}$达到该标准的运行比例升至 90%。在采用匹配的初始化和支撑约束的情况下，带有跳跃连接的复数卷积神经网络（C-CNN）取得高得多的 χ^2值，范围为 6×10^{-3}–1.2×10^{-2}。在这三种方法中， χ^2 在不同随机初始化下的结果呈双峰分布；然而，只有傅里叶视觉Transformer达到了最低的χ^2 值（$\chi^2 \leq 10^{-5}$），而 C-CNN 的结果仍局限于误差较高的峰值。

2.3 噪声鲁棒性

为评估模型在真实实验退化情况下的鲁棒性，我们对图 2 中使用的同一无噪声衍射图样加入了三种噪声：用于表征探测器背景的加性高斯噪声；用于表示计数统计的泊松计数噪声；以及入射光束的部分相干性，如图3所示。严格来说，部分相干性与统计噪声模型并不相同，因为它可以得到校正（Clark 等人，2012），但仍值得考察其对收敛的影响，以便将其用于识别。每种噪声模型均按其对 χ^2 的影响设置五种强度，从等级 1 到等级 5。为便于直观比较，图 3 展示了等级 1、3 和 5；等级 2 和 4 见补充图 1。

对于每种噪声类型，表1 报告了三种归一化衍射误差：(i) 含噪输入相对于干净真值的误差， χ^2_{n}；(ii) 重建与含噪输入之间的不匹配程度，$\chi^2_{\mathrm{rec,\ n}}$；以及 (iii) 重建相对于干净真值的误差，$\chi^2_{\mathrm{rec,\ c}}$。在无监督设置下，我们通过最小化 $\chi^2_{\mathrm{rec,\ n}}$，从多次随机初始化中选取最佳运行结果，并报告相应的 $\chi^2_{\mathrm{rec,\ c}}$，作为对干净真值的评估。

对于高斯噪声， χ^2_{n} 会随噪声强度增加，从 3.28% 升至 28.32%。傅里叶视觉Transformer 将相对于无噪声衍射图样的误差降低至 $\chi^2_{\mathrm{rec,\ c}} = 1.54$–14.36%，相较于 χ^2_{n}，误差降低了 49–59%。对于泊松噪声， χ^2_{n} 会从 1.86% 升至 20.61%，而模型将其降低至$\chi^2_{\mathrm{rec,\ c}} = 0.92$–10.53%，降幅同样达到 49–55%。在这两种随机噪声

过程中，重建结果始终比受噪声污染的输入本身更接近无噪声衍射图样，这表明模型实现了真正的去噪，而非仅仅复现噪声统计特征。

图 3 展示了高斯噪声和泊松噪声对实空间图像的影响。在低噪声水平下，衍射条纹仍清晰锐利，重建结果呈现出清晰且能分辨畴结构的晶体相位图。随着噪声增大，像素尺度涨落开始出现，并逐渐破坏相位，最终使畴边界变得模糊或消失。在高斯噪声下，重建晶体的振幅和相位中均出现像素尺度的强度涨落；而在泊松统计下，振幅中出现了与底层畴结构相呼应的晶界对比度。在这两种情况下，傅里叶视觉Transformer都会抑制无规律的散斑和背景涨落，同时保留中心布拉格峰周围具有物理意义的条纹模式，从而即使在条件严重退化时也能实现稳健的相位恢复。与表 1一致，傅里叶视觉Transformer因此起到了滤波器的作用，使 χ^2 中的衍射畴误差相较于含噪输入降低约一半。

对于部分相干条件， χ^2_{n} 随着模糊程度增加，从 0.31% 增至 5.62%。重建结果与模糊测量结果极为接近，$\chi^2_{\mathrm{rec,\ n}} = 0.11$–0.25%，但相对于无噪声衍射图样的误差处于$\chi^2_{\mathrm{rec,\ c}} = 0.25$–6.37% 的范围内。比值 $\chi^2_{\mathrm{rec,\ c}}/\chi^2_{\mathrm{n}}$ 在最弱水平下为 ≈ 0.81，但在模糊程度较高时增至 ≈ 1.13–1.29。衍射条纹逐渐模糊，能量集中在中心布拉格峰附近。重建振幅呈现过度局域化的中心强度（“热点”区域），相位则趋于平滑，这反映了高q 频信息的损失。因此，部分相干性是重建晶体振幅中出现中心热点的一种自然物理成因。这种现象已有文献记载（Vartanyants 和 Robinson，2001），并且在第 2.4 节讨论的实验重建中也有所观察。

2.4 实验验证

作为初步实验验证，我们在来自弱相位、单畴、尺寸为 450 nm 的 $SrTiO_3$ (STO) 纳米晶体的二维衍射图样上测试了傅里叶视觉Transformer；结果见补充图 2。在图 7(a,b) 中，重建结果在各次运行中均恢复了单畴形貌。在 100 次随机初始化中，最终倒易空间失配 χ^2 均处于较窄的范围内（$\sim 0.35\%$–$\sim$0.68%；图7(c)），表明在弱相位区域内能够可靠收敛。这与下文讨论的强相位、多畴情形形成对比；在该情形下，多个局部极小值导致各次运行的 χ^2 值分布范围显著更宽。

随后，我们使用来自强烈畸变的多畴LCMO-500纳米晶体的实验布拉格相干衍射成像图样，考察强相位区域（第 4.3 节）。图 4 比较了采用传统迭代方法（ER/HIO）、傅里叶视觉Transformer 和 C-CNN 基线得到的重建结果，三种方法使用相同的实空间支撑域。对于每种方法，我们报告了 200 次随机初始化中表现最佳的一次运行结果（最低 χ^2）。

对于图中原始测量的衍射幅值。4(a)，迭代重建（图 4(b,e)）恢复了多畴相位结构，但振幅图中出现了孤立的热点。最佳结果达到 $\chi^2 = 0.25\%$，相对于测量幅值的皮尔逊相关系数为 $= 99.77\%$。傅里叶视觉Transformer在相同的单一衍射图样上进行无监督训练，并受相同支撑域约束，得到图 4(c,f) 所示的振幅和相位。与 (b) 相比，傅里叶ViT振幅图 (c) 中的高振幅区域范围更广、紧凑度更低，整体振幅延伸至晶体更大比例的区域；与 (d) 相比，其边缘局部化程度较低，也更可能反映物体的二维投影。在相位图中，恢复出的畴图案 (f) 与迭代解 (e) 一致，畴边界更清晰，相位区域的空间连贯性更强，孤立的高对比度特征也更少。傅里叶视觉Transformer的最佳运行结果达到 $\chi^2 = 0.30\%$，皮尔逊相关系数为 $= 99.79\%$。

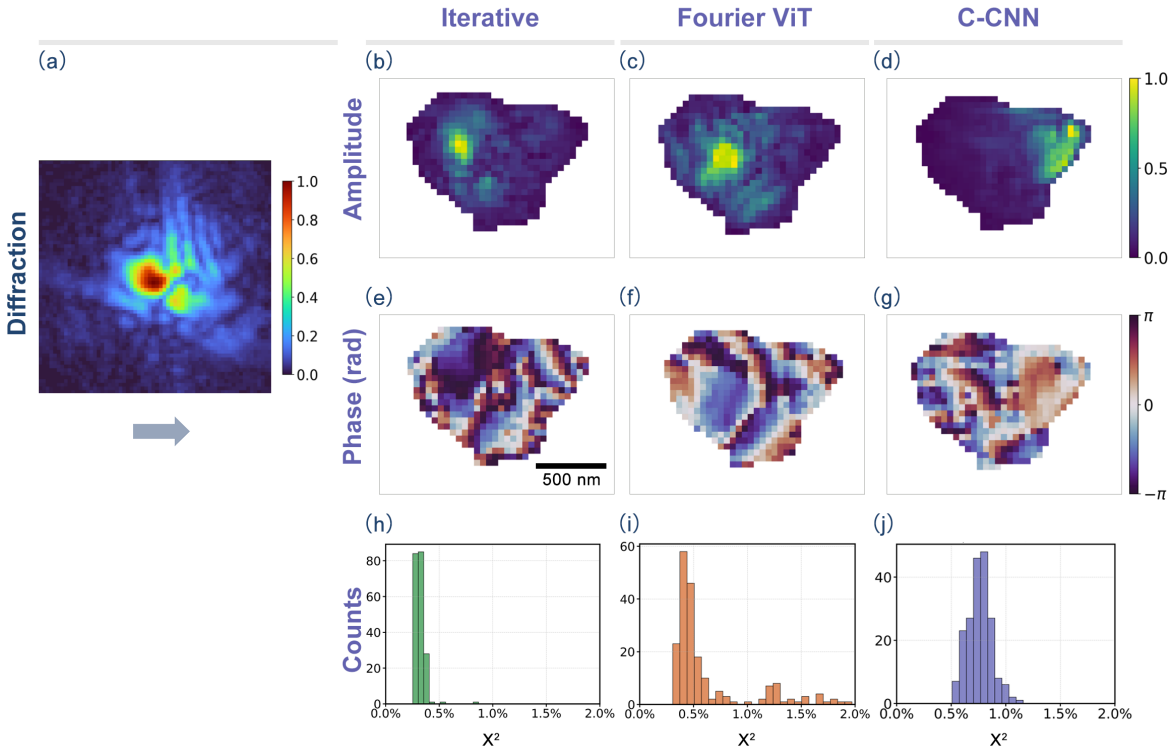


图4：多畴LCMO-500 纳米晶体实验数据相位恢复的对比研究。(a) 测得的布拉格相干衍射成像衍射幅值（64×64 像素）。(b–d) 迭代方法、傅里叶视觉Transformer和C-CNN重建的晶体振幅，以及(e–g)相位。比例尺为500nm。(h–j) 每种方法运行200次、每次使用随机初始相位所得的相应 χ^2 直方图。χ^2 范围介于0.0%和2.0%之间。

相比之下，图 4(d,g) 中 C-CNN 的最佳结果满足支撑约束，但收敛到不同的解：振幅集中在支撑边界附近，相位拓扑也偏离迭代重建和傅里叶 ViT 重建。与此同时，$\chi^2 = 0.50\%$ 较高，这表明在相同前向模型下，其与测量衍射幅值的匹配较差。

为去除晶体振幅中的热点，我们通过对测得的衍射幅值进行高斯反卷积，测试了温和的部分相干校正（Vartanyants和Robinson，2001）。虽然反卷积明显使条纹变得更清晰，但也会放大其他探测器噪声，且重建结果 χ^2 相较于原始数据情况有所增加：迭代方法为 $\chi^2 = 0.61\%$，傅里叶视觉Transformer为 $\chi^2 = 0.57\%$，但热点并未得到显著改善。因此，我们在图4中使用原始测量图样作为主要比较依据。

图4(h–j)中的 χ^2 直方图绘制在同一坐标轴上，总结了每种方法200次独立运行之间的变异性。迭代重建(h)紧密聚集在 $\chi^2 \approx 0.25$–0.87%范围内，表明收敛稳定。傅里叶视觉Transformer(i)取得了相当的最佳情况数值 $\chi^2 = 0.3\%$

但显示出更宽的分布范围（~ 0.3–2.0%），而C-CNN分布(j)则移向更高误差（$\sim$ 0.5–1.1%）。补充视频展示了三种方法逐轮次的演化过程，并突出了不同的优化轨迹。

总体而言，在这一实验数据集上，傅里叶视觉Transformer 对实测衍射的拟合至少与表现最佳的迭代重建同样出色，并且在 χ^2方面优于 C-CNN 基线，同时生成的多畴晶体振幅和相位在相同支撑约束下仍与迭代解一致。

3 讨论

我们开发了一种采用傅里叶注意力的视觉Transformer（傅里叶视觉Transformer），用于布拉格相干衍射成像中的多畴相位恢复。在合成 64×64 衍射图样上，该模型能够在无监督设置下直接从衍射数据中恢复强相位畴结构。对于已知振幅的仅相位预测，模型运行可与目标衍射达到近乎完美的一致性（$\chi^2 \leq 10^{-5}$），同时还能解析多达 19 个畴的复杂畴结构（图 2f）。对于振幅与相位联合预测，一个具有代表性的 10 畴案例在整个 q范围内重现了真实衍射强度，包括高q 尾部（图 2c）；重建晶体相位则恢复了相同的畴布局，且畴壁清晰锐利（图 2h）。在相同衍射图样和相同固定支撑域条件下，ER/HIO 在进行 500 次迭代后，约半数重建达到 $\chi^2 = 10^{-4}$–10^{-3} 的水平；随着迭代次数增加，达到这一范围的运行比例也会提高。复值卷积神经网络（C-CNN）在相同初始设置下仍存在较高的 χ^2 误差（6×10^{-3}–1.2×10^{-2}），且即使训练轮次充足，也无法达到傅里叶视觉Transformer 所实现的 $\chi^2 \leq 10^{-5}$ 区域。在三种方法中，在强相位区域进行随机初始化时，即使约束相同，不同运行也会分化为较低和较高的χ^2结果。

图 3 和表 1 中的鲁棒性测试表明，在高斯噪声和泊松噪声下，各噪声水平的重建结果均比含噪输入更接近无噪声衍射（$\chi^2_{\mathrm{rec,\ c}} < \chi^2_{\mathrm{n}}$），且 $\chi^2_{\mathrm{rec,\ c}}$ 相较于 χ^2_{n}约降低了一半。这表明衍射幅值实现了去噪。相比之下，部分相干性会模糊衍射图样并降低条纹可见度。尽管模型能够极为准确地拟合模糊后的测量结果，但随着模糊加剧，重建的衍射图样会偏离无噪声目标。部分相干性还可能导致中心振幅出现过度局部化的“热点”，此前已有文献记载这一现象（Vartanyants 和 Robinson，2001），且我们在实验重建中也观察到了这一现象（第 2.4节）。因此，当模糊较为显著时，需要显式地纳入相干性模型。三种失真在预测图像中的表现各不相同：高斯噪声会引入局部噪声特征，泊松噪声会产生颗粒状波动，而部分相干性则会模糊条纹，并使畴边界不那么清晰。

随后，我们使用相同的支撑约束和损失函数中的相同基于 FFT 的前向模型，将该方法应用于实测的 LCMO-500 纳米晶体衍射数据，并与 ER/HIO 和 C-CNN 进行比较。如补充视频所示，这三种方法呈现出不同的优化行为。在 ER/HIO 中，负反馈步骤有助于迭代摆脱局部极小值，并达到较低的 χ^2 解。C-CNN 倾向于从平滑的初始场构建相位图，逐渐形成更清晰的畴边界，并在训练过程中逐步增加空间细节；在我们的测试中，它更常收敛到具有边缘局域振幅且相位拓扑不同于其他方法的解。相比之下，傅里叶视觉Transformer通过令牌混合在全局范围内耦合衍射图样特征，并在训练早期生成多畴构型。这减少了在过度平滑中间态上花费的时间，使优化快速进入相关的强相位解类别。

在图 4中，傅里叶视觉Transformer 达到 χ^2，其数值与最佳迭代重建相当（迭代法：$\chi^2 \approx 0.25\%$，傅里叶视觉Transformer：$\chi^2 \approx 0.30\%$），并优于 C-CNN 基线（$\chi^2 \approx 0.50\%$）。在相同的支撑约束下，恢复的畴形貌与 ER/HIO 结果一致。与 ER/HIO 相比，傅里叶视觉 Transformer 的振幅较少集中于紧凑的热点中，相位则呈现出更清晰且空间上更连贯的畴区域。图 χ^2 4 中的重建误差直方图 （h-j）还凸显了强相位区域的一个实际问题。在存在现实噪声且仅幅值数据约束下不可避免地出现模型失配时，许多近简并的极小值都能以几乎同样的效果拟合实测衍射数据。因此，重建结果在倒易空间中可能几乎无法区分，但在实空间中仍可能存在有意义的差异。因此，降低 χ^2 对于确定唯一的畴拓扑是必要的，但并不充分。傅里叶视觉Transformer 结果更广的分布，更自然地应被理解为：全局令牌混合使其能够快速获得多个可接受的强相位解，而不是优化不稳定。实际上，我们报告最优的χ^2 重建结果，并验证在实测强度和施加的约束无法强制确定唯一解时，关键的畴级特征在独立运行之间具有可重复性。

从计算角度看，ER/HIO 的计算主要耗在反复进行 FFT 投影上，通常每次迭代都要进行一次正向 FFT 和一次逆向 FFT （ Fienup，1982 年；Duhamel 和 Vetterli，1990 ）。在无监督学习条件下，FFT前向模型与网络前向/反向传播一同纳入损失函数，适用于 C-CNN 和傅里叶视觉Transformer。在傅里叶视觉Transformer中，傅里叶令牌混合实现了全局耦合，同时避免了点积自注意力的二次复杂度（Vaswani 等人，2017年；Lee-Thorp 等人，2022年）。

总体而言，傅里叶视觉Transformer为强相位、多畴BCDI重建提供了一种实用途径。在近乎理想的条件下，所有方法都能通过增加迭代次数或训练轮次获得提升；但在固定的优化预算下，傅里叶视觉Transformer能更早得到结构化的多畴解，其表现优于 C-CNN 基线，同时在实验数据上的表现与最佳 ER/HIO 重建相当。未来的工作将把显式噪声和部分相干模型纳入前向算子，探索从解集成中进行不确定性感知选择，学习如何从衍射数据推导支撑域，并将该方法扩展至完整的三维实验数据集。

4 方法

4.1 合成数据与噪声

4.1.1 Voronoi多畴晶体

为模拟功能纳米晶体的异质结构，我们构建了由多个具有锐利、笔直边界的相位畴组成的二维晶体。在图 5(a) 中，我们在带有新月形切口的软边缘圆形支撑域内随机放置种子点，以表示颗粒形状，并为每个种子随机指定一个 $[-\pi, \pi]$ 相位。柔和的振幅边缘可抑制傅里叶空间中由方形边界引起的人为高频分量，而新月形切口可打破对称性，有助于避免孪生解。我们将种子之间的最小间距设为 2 像素，以防止聚集。随后，每个种子生成一个 Voronoi 单元，其像素取与之最近的种子的相位，从而形成由笔直边界划分、互不重叠的畴镶嵌。这种基于 Voronoi 的构造模拟了铁电和铁弹性材料中通常观察到的突变畴壁（Setter 等，2006）。

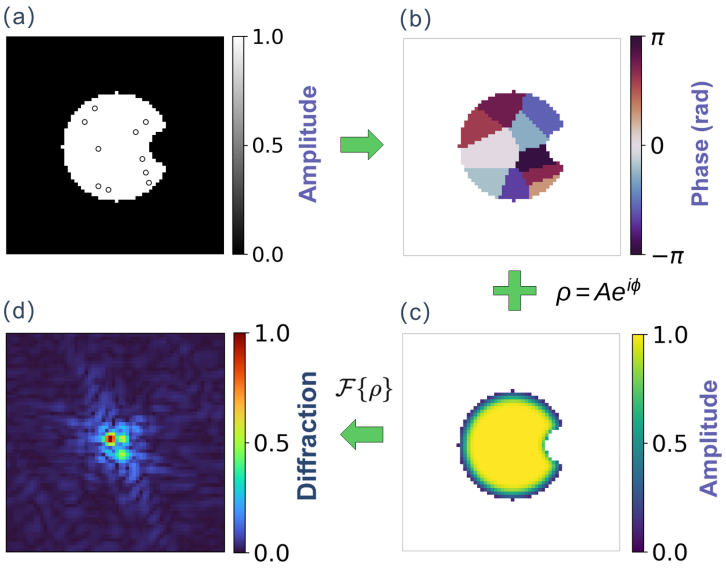


图 5：合成晶体与衍射图样的模拟流程。(a) 在支撑域内随机放置种子。(b) 根据种子生成 Voronoi 相位畴，并将相位包裹至 $[-\pi, \pi]$。(c) 构建晶体振幅掩膜，并归一化至最大值 $[0, 1]$。(d) 通过快速傅里叶变换（FFT）计算得到的对应衍射幅值，并归一化至最大值 $[0, 1]$。所有图像均为 64×64 像素。

在图 5(b) 中，我们将中心畴设为 $\phi = 0$，以消除全局相位偏移。我们要求相邻畴之间的相位差为 $\Delta\phi > 0.5$ 弧度，以确保强相位衬度并避免反演对称情形。图 5(c) 展示了晶体振幅 $A(x,\ y)$，其用于表示有限三维颗粒的二维投影振幅。我们采用平顶核心径向轮廓并进行高斯边缘平滑，再以新月形掩膜进行调制，从而得到受限但边缘柔和的支撑域。最终的实空间物体定义为复密度

$$\rho(x, y) = A(x, y)\, e^{i\phi(x,y)}, \tag{9}$$

其中 $A(x,\ y)$ 固定，$\phi(x,\ y)$ 由 Voronoi 分区确定。

在图 5(d) 中，衍射强度是复密度的傅里叶变换的模平方，

$$I_{\text{pred}}(\mathbf{q}) = \left|\mathcal{F}\{\rho(x, y)\}(\mathbf{q})\right|^2, \tag{10}$$

在包含直径为 32 像素的支撑域的 64×64 网格上，该设置对应倒易空间中 2 的过采样比。在支撑域之外，$\rho(x,\ y)$ 被设为零，因此只有物理上允许的区域会对散射产生贡献。晶体振幅、相位和衍射图样均以 16 位 TIFF 图像存储。总体而言，该流程生成了具有受控数量的域、强相位边界以及逼真的倒易空间结构的合成数据，我们利用这些数据对傅里叶 ViT 相位恢复网络进行基准测试。

4.1.2 噪声测试

我们评估这样一种情况下的鲁棒性：网络预测晶体振幅和相位，同时在衍射图样中加入噪声。从最大值归一化的衍射图样强度$I(q_x,\ q_y) \in [0,1]$出发，我们构建了三类受扰数据：泊松散粒噪声、高斯读出噪声和部分相干性。

对于这两种探测器噪声模型，我们采用曝光量（光子计数）尺度。在这种情况下，传感器链产生的读出噪声可很好地近似为计数值中的加性零均值高斯噪声，而光子散粒噪声则用泊松统计描述（Goodman，2015；Janesick，2001，2007年）。给定从归一化强度到期望计数值的转换因子 $S > 0$，理想无噪声计数值为 $C_0(q_x,q_y) = S\,I(q_x,q_y)$。高斯读出噪声建模为

$$C_g(q_x,q_y) = C_0(q_x,q_y) + \eta(q_x,q_y), \tag{11}$$

其中，随机变量 $\eta(q_x,q_y)$ 在每个像素处独立地从均值为零、方差为 σ_c^2 的正态分布中抽取。泊松散粒噪声建模为

$$C_p(q_x,q_y) \sim \mathrm{Poisson}\big(C_0(q_x,q_y)\big), \tag{12}$$

这反映了光子在每个像素处离散且相互独立的到达过程。泊松函数使用 NumPy 数组库中的泊松随机数生成器生成（Harris 等，2020）。在这两种情况下，我们都通过截断 $C(q_x,q_y) \geq 0.\tilde{}$ 来确保记录的计数值符合物理上的非负性。为进行分析，我们通过 $I(q_x,q_y) = C(q_x,q_y)/S$ 恢复到强度尺度，并使用相应的衍射幅值 $\tilde{M}(q_x,q_y) = \sqrt{\tilde{I}(q_x,q_y)}$ 计算所有报告的图像空间指标以及表1中的重建损失 χ^2_{rec}。

部分相干性被视为衍射强度的退化，而非随机探测器噪声。我们将其建模为一种降低对比度的模糊

$$I_{\mathrm{pc}}(q_x,q_y) = \mathcal{G}_\sigma * I(q_x,q_y), \tag{13}$$

其中 $\mathcal{G}_\sigma$ 将远场互相干函数近似为倒易空间中的高斯函数，因此部分相干性被建模为相干强度的卷积（Williams 等人，2007年；Clark 等人，2012）。对于模拟数据，我们将相干衍射与 $\mathcal{G}_\sigma$ 进行卷积，然后重新归一化 I_{pc}，使其与清晰图样的整体缩放相匹配，从而使重建质量的变化主要反映高q条纹可见度的损失，而非强度缩放。对于实验数据（第 2.4 节），我们使用正则化反卷积反向应用相同的高斯模型，以部分消除模糊，获得更接近相干状态的有效衍射图样。

4.2 Fourier ViT网络训练

傅里叶视觉Transformer处理的是 64×64 最大值归一化的衍射幅值。图像块大小为 4，因此得到 16×16 个词元，并采用嵌入维度为 128 的表示，经 10 个傅里叶变换器块处理。网络使用 AdamW（Loshch并 Hutter，2019）（初始学习率为 5×10^{-4}，权重衰减为 10^{-5}，$\beta_1 = 0.9$，$\beta_2 = 0.999$），并采用余弦退火学习率调度，将学习率降至 10^{-6}。训练在 NVIDIA Quadro RTX 5000 (Max-Q) 上使用自动混合精度。

全文均使用第2.1.4 节的混合损失函数。采用固定的二元支撑域；振幅在支撑域内以高斯先验初始化，并在前 10 个训练轮次中保持固定，之后通过混合调度 $\alpha(t)$（补充说明 2）。丢弃率设为 0，并采用最大梯度范数为 1.0 的梯度裁剪，以稳定优化过程。

对于每种相位畴数量，我们使用不同的随机种子进行100次独立运行，并报告这些运行的统计结果。我们通过最小化预测与测量衍射幅值之间的衍射失配 χ^2，从不同随机初始化中选出最佳运行，并报告相应的重建指标，包括在可用时与干净目标相比的 χ^2。

4.3 布拉格相干衍射成像实验

我们测量了 $La_{0.5}Ca_{0.5}MnO_3$ (LCMO) 晶体的布拉格相干衍射成像数据。LCMO 是一种重要的复杂过渡金属氧化物，在接近室温时表现出庞磁阻（CMR）；磁场的微小变化会引起电阻的大幅变化。我们关注相位畴是否会对 CMR 行为产生影响。LCMO 通过化学合成制备为尺寸约为 300 nm 的纳米晶体，并聚集成团簇（Robinson 等人，2026）。随后，我们采用正硅酸四乙酯（TEOS）键合法将这些团簇附着在硅晶圆基底上，以便进行布拉格相干衍射成像测量（Monteforte 等人，2016）：使用稀释的 TEOS 溶液涂覆干燥粉末，然后将其置于空气中，在炉内煅烧数小时，温度为 $500\ ^\circ\mathrm{C}$。

X 射线测量在先进光子源（APS）的 34-ID-C 光束线上进行（Robinson 等人，2026）。波长为 $\lambda = 0.138\mathrm{nm}$ 的单色 X 射线束经准直后具有相干性，所用狭缝尺寸为 $50\,\mu\mathrm{m}$（垂直方向）$\times\ 20\,\mu$ m（水平方向），距光源 48 m；随后通过 Kirkpatrick–Baez（KB）反射镜聚焦，使光束在精密空气轴承旋转轴中心处的尺寸达到 600 nm $\times$ 800nm， θ。通过旋转 θ，选取了一个纳米晶体，直至 Medipix 面探测器上出现孤立的峰。该探测器的像素尺寸为 $55\,\mu\mathrm{m}$，位于 (110) 布拉格峰对应的布拉格角处，距样品 $D = 2.2\mathrm{m}$。本研究从该三维数据集中选取了一帧，即 LCMO-500。对应图像是三维纳米晶体结构的相位投影；根据傅里叶切片定理，其傅里叶变换幅值平方（加上噪声）应与测得的帧完全匹配。

4.4 迭代方法

我们按照标准的 BCDI 实践（Robinson 和 Harder，2009；Clark 等人，2012），在 MATLAB 中实现了常规的二维迭代相位恢复算法。初始支撑域根据重建网格定义，其中 $s_x = 0.5\,N_x$ 和 $s_y = 0.5\,N_y$。随后，我们采用预定义的切换计划，在误差缩减（ER）算法的20次迭代与混合输入输出（HIO）算法的80次迭代之间交替进行，HIO 松弛参数为 $\beta = 0.9$（Fienup，1982年）。我们使用收缩包裹更新（Marchesini 等人，2003年）对支撑域进行细化，最大值归一化振幅阈值为 $= 0.15$。在 ER 过程中，我们将支撑域外晶体的振幅设为零，并将相位限制在 $[-\pi, \pi]$范围内。LMCO-500的重建使用了300次迭代，并从第200次迭代开始进行收缩包裹以确定支撑域，随后使用该支撑域比较傅里叶视觉Transformer、卷积神经网络和迭代方法的结果。模拟数据的固定支撑域重建最多进行2000次迭代，以统计收敛情况。

4.5 复值卷积神经网络

为了与所提出的傅里叶视觉Transformer进行实验比较，我们按照用于BCDI相位恢复的复值网络方案（Yu 等人，2024年），实现了一个复值卷积神经网络基线。该模型采用 U-Net编码器-解码器架构。在我们的实现中，编码器使用的复数通道宽度为32、64、128、256和512（解码器则为256、128、64和32），后接一个1×1 复数卷积层，输出单个复数通道。整个网络均采用负斜率为0.2的LeakyReLU激活。特征图以实部和虚部通道表示，所有卷积均通过复数运算实现，即按照复数乘法混合两个通道。跳跃连接贯穿编码器和解码器各阶段，以保留精细的空间细节。对于LCMO-500测量，我们将网络输入初始化为两个通道，其中 $\Re\{x\}$ 设为最大值归一化后的实测衍射幅值，$\Im\{x\} = 0$则对应于零初始相位。网络输出复对象 $\rho(x) = \rho_{\Re}(x) + i\rho_{\Im}(x)$，并据此通过 $|\rho|$ 和 arg(ρ)求得振幅和相位。训练和评估遵循与第4.2节相同的流程。

代码可用性

用于训练和评估傅里叶视觉Transformer以及复现所有图表的代码可在 https://github.com/JialunSimonLiu/Fourier-Attention Vision-Transformer获取。

补充说明1：多尺度傅里叶注意力的详细信息

此处定义第 2.1.2 节中介绍的频谱混合器。令 $\mathbf{X} \in \mathbb{R}^{B\times E\times H\times W}$ 表示 Transformer 块内部的 $H{=}W{=}16$ 词元特征图。在尺度 s 下，我们通过平均池化进行下采样，应用逐通道傅里叶门控，然后再上采样恢复：

$$\mathbf{Z}_s = \mathcal{U}_s\Big(\mathcal{F}^{-1}\big(\,\mathcal{F}(\mathcal{D}_s\mathbf{X})\;\odot\;\mathbf{W}_s\;\odot\;\mathbf{M}_s\,\big)\Big),$$

其中，$\mathcal{D}_s$ 是步幅为 s 的平均池化，用于将尺寸降至 $h_s{\times}w_s = (H/s)\times(W/s)$；$\mathcal{U}_s$ 将其上采样回 $(H,\ W)$；$\mathcal{F}$ 是逐通道二维快速傅里叶变换；$\mathbf{W}_s \in \mathbb{R}^{E\times h_s\times w_s}$ 是逐通道学习得到的频率响应；而$\mathbf{M}_s \in \mathbb{R}^{1\times h_s\times w_s}$ 是共享门控。多尺度输出为

$$\mathbf{Z} = \sum_s \alpha_s\,\mathbf{Z}_s, \qquad \boldsymbol{\alpha} = \mathrm{softmax}(\boldsymbol{\theta}),$$

其中，尺度权重 α_s 由 $\boldsymbol{\theta} \in \mathbb{R}^{|\{s\}|}$ 参数化。

根据卷积定理，

$$\mathbf{Z}_s = \mathcal{U}_s\big(\,\mathcal{D}_s\mathbf{X}\;*\;\mathbf{g}_s\,\big), \quad \mathbf{g}_s = \mathcal{F}^{-1}(\mathbf{W}_s\odot\mathbf{M}_s),$$

其中，$*$ 表示在 $h_s{\times}w_s$ 网格上的循环卷积。每个尺度都应用学习得到的核 $\mathbf{g}_s$，将结果上采样至 $(H,\ W)$，并使用权重 α_s 合并各尺度。这样便通过基于快速傅里叶变换的卷积实现了全局令牌混合。在频域中，$\mathbf{W}_s\odot\mathbf{M}_s$ 是作用于傅里叶模态的乘性门控。

令 $N = HW$ 表示词元数量。对于每个 Transformer 块，频谱混合器所需的时间复杂度为$\mathcal{O}(|\{s\}|\,B\,E\,N\log N)$，内存复杂度为 $\mathcal{O}(|\{s\}|\,B\,E\,N)$；相比之下，点积注意力的时间复杂度为 $\mathcal{O}(B\,E\,N^2)$，内存复杂度为 $\mathcal{O}(B\,N^2)$。这沿用了基于傅里叶变换的词元混合（FNet）（Lee-Thorp 等人，2022年）以及傅里叶神经算子中使用的频谱参数化方法（Li 等人，2021），但在我们的模型中，它仅作为紧凑的多尺度门控模块置于每个 Transformer 块内部。

补充说明 2：振幅先验与归一化

解码器在 64×64 网格上预测实空间振幅 $\hat{A}(x, y)$，并按第 2.1.3 节所述预测相位场。在构造复密度之前，我们将 $\hat{A}(x, y)$ 与定义在实空间支撑域 $S(x, y)$ 上的简单先验振幅 $A_{\text{prior}}(x, y)$ 结合，其中 $S(x, y) = 1$ 位于支撑区域内，而 $S(x, y) = 0$ 位于支撑区域外。基于该支撑域，我们构造三种先验之一：(i) 平坦先验，其中 $A_{\text{prior}}(x, y) = S(x, y)$；(ii) 通过对 $S(x, y)$ 反复进行局部平均得到的模糊先验；或 (iii) 以支撑域为中心并由 $S(x, y)$ 截断的高斯先验。

训练期间，我们通过标量权重 $\alpha(t)$ 混合先验振幅和预测振幅：

$$A_{\text{blend}}(x, y; t) = \alpha(t)\, A_{\text{prior}}(x, y) + \big(1 - \alpha(t)\big)\, \hat{A}(x, y).$$

这样，训练早期由基于支撑域的先验引导，后期则主要依赖网络预测。我们采用余弦调度，并设置锁定轮次 t_{lock} 和衰减结束轮次 t_{end}：

$$\alpha(t) = \begin{cases} 1, & t < t_{\text{lock}}, \\ \dfrac{1}{2}\Big(1 + \cos\dfrac{\pi\,(t - t_{\text{lock}})}{t_{\text{end}} - t_{\text{lock}}}\Big), & t_{\text{lock}} \le t < t_{\text{end}}, \\ 0, & t \ge t_{\text{end}}. \end{cases}$$

对于实验单图样运行，我们使用 $t_{\text{lock}} = 10$ 和 $t_{\text{end}} = 50$。对于合成数据集运行，我们采用较慢的调度（$t_{\text{lock}} = 50$, $t_{\text{end}} = 200$），以减少振幅-相位联合恢复早期的振幅漂移；在所有情况下，一旦 $\alpha(t) = 0$，就移除先验。

为稳定重建振幅的整体尺度，我们在支撑域内应用软能量匹配步骤。令 $A_{\text{blend}}(x,\ y; t)$ 表示第 t 轮的混合振幅。我们计算混合振幅和先验振幅在支撑域内的总“能量”，如下所示

$$E_{\text{cur}}(t) = \sqrt{\sum_{x,y} \big(A_{\text{blend}}(x, y; t)\, S(x, y)\big)^2}, \qquad E_{\text{tgt}} = \sqrt{\sum_{x,y} \big(A_{\text{prior}}(x, y)\, S(x, y)\big)^2}.$$

其中，E_{cur} 为当前能量，E_{tgt} 为目标能量。随后我们重新缩放

$$\tilde{A}(x, y; t) = \gamma(t)\, A_{\text{blend}}(x, y; t), \qquad \gamma(t) = \text{clip}\big(E_{\text{tgt}}/E_{\text{cur}}(t),\, \gamma_{\min}, \gamma_{\max}\big),$$

在所有实验中，使用 $\gamma_{\min}$=0.8 和 $\gamma_{\max}$=1.2 。这使支撑区域能量保持接近先验值。在实践中，它能防止全局振幅漂移，同时允许相对于 A_{prior} 出现局部偏差。

补充说明 3：损失权重和指数

混合损失结合了皮尔逊相关系数 (PCC)、标准 χ^2 和功率–χ^2，

$$L_{\text{total}}(t) = w_{\text{PCC}}(t)\, L_{\text{PCC}} + w_{\chi^2}(t)\, L_{\chi^2} + w_{\text{p}\chi^2}(t)\, L_{\text{p}\chi^2},$$

其中训练轮次索引为 $t \geq 0$。我们将原始权重调度设为

$$\begin{aligned} \text{raw}_{\text{PCC}}(t) &= 200\, \exp(-t/15) + 1, \\ \text{raw}_{\chi^2}(t) &= 1, \\ \text{raw}_{\text{p}\chi^2}(t) &= 0.5\, r(t), \end{aligned}$$

其中 $r(t)$是在前 20 个训练轮次内从 0 到 1 的正弦渐变。对于 $t \geq 200$，我们将原始权重重新缩放为

$$\text{raw}_{\text{PCC}}(t) = 0.6\, \text{raw}_{\text{PCC}}(t), \quad \text{raw}_{\text{p}\chi^2}(t) = 0.6\, \text{raw}_{\text{p}\chi^2}(t), \quad \text{raw}_{\chi^2}(t) = 1.1\, \text{raw}_{\chi^2}(t).$$

随后，我们用训练轮次 t 时的原始权重之和进行归一化，得到 $w_{\text{PCC}}(t)$、$w_{\chi^2}(t)$ 和 $w_{\text{p}\chi^2}(t)$。幂–χ^2 项中的指数从 2.0 递减至 0.5，

$$p(t) = p_{\text{end}} + \big(p_{\text{start}} - p_{\text{end}}\big)\, \exp(-\lambda t),$$

其中 $p_{\text{start}} = 2.0$、 $p_{\text{end}} = 0.5$ 和 $\lambda = 0.005$。我们在振幅上加入全变差项，其权重为 $\lambda_{\text{TV}}(t) = 0.01$ （适用于 $t < 200$），此后为 $\lambda_{\text{TV}}(t) = 0$ 。

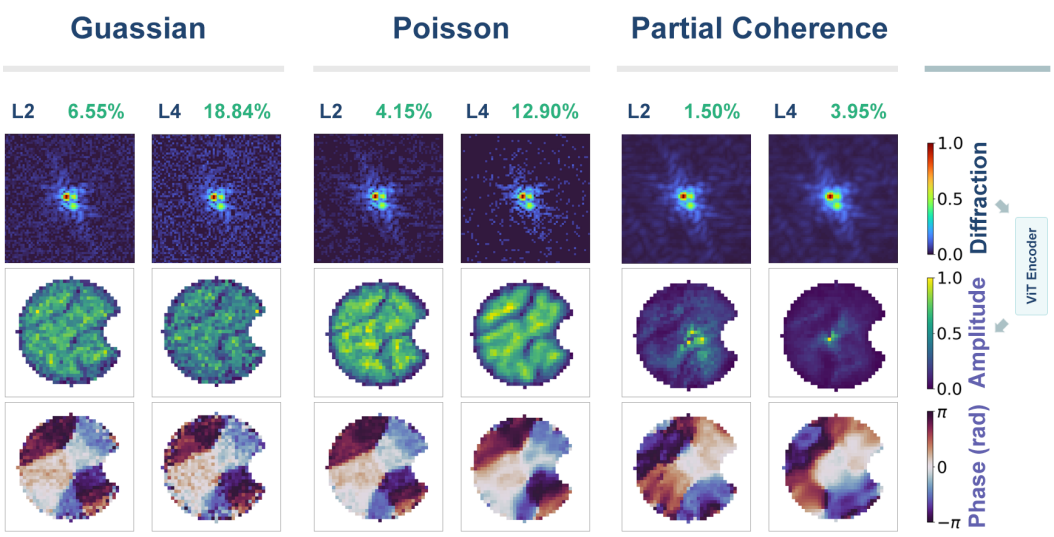


图 6：傅里叶视觉Transformer在等级 2 和 4 下的额外噪声鲁棒性可视化（等级 1、3 和 5 的结果见图 3）。各列分别展示了对同一参考衍射图样施加高斯噪声、泊松噪声和部分相干性的结果。对于每种噪声模型，绿色百分比表示 χ^2_{n}（%），即相对于干净真值的输入衍射误差。顶行显示含噪输入衍射幅度（64×64 像素）。中间行和底行分别显示重建晶体振幅和相位（40×40 像素）；二者均以支撑域为掩膜，相位以弧度表示。

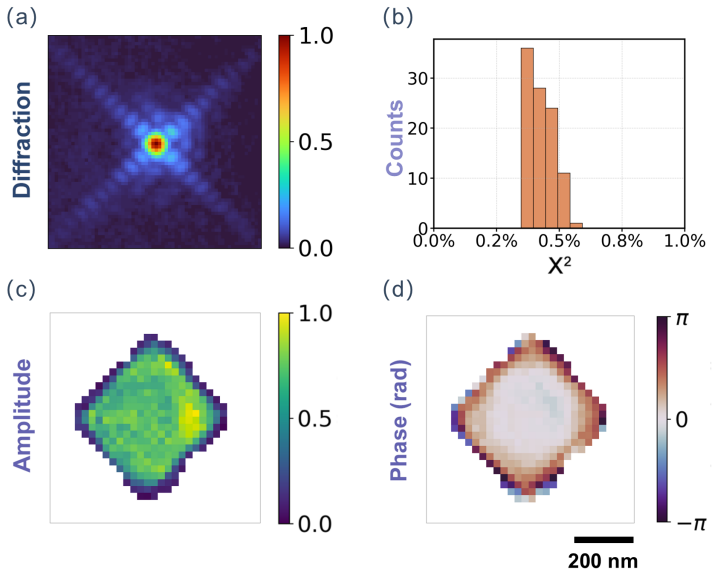


图 7：截面尺寸为 $\sim 460 \times 450\ \mathrm{nm}^2$ 的单畴 $SrTiO_3$（钛酸锶）纳米晶体。(a) 测得的钛酸锶衍射幅值，显示为以布拉格峰为中心的 64×64 像素裁剪区域，并归一化至最大值 $[0, 1]$。(b) 100 次随机初始化后最终倒易空间失配 χ^2 的分布，范围为 $\sim 0.35\%$ 至 $\sim 0.68\%$。(c) 重建的实空间振幅，归一化至最大值 $[0, 1]$。(d) 重建的实空间相位，缠绕至 $[-\pi, \pi]$；比例尺表示 200 nm。

致谢

我们谨向 Xi Yu、Wenxuan Fang、Jack Griffiths、Yue Dong、Ruiyang Duan、Seungjun Lee、Yuewei Lin、Longlong Wu 和 Shinjae Yoo 致谢，感谢他们进行有益的讨论，并感谢 Weibin Peng 在图表设计和展示方面提供的建议。我们还要感谢 Ross Harder、Tadesse Assefa、Yue Cao、Xiaojing Huang 和 Jesse Clark 在 LCMO 测量方面提供的帮助，以及 Josh Turner 和 Arup Kumar Raychaudhuri 提供样品。测量工作在先进光子源（APS）的 34-ID-C 光束线上进行，该光束线由美国能源部科学办公室基础能源科学办公室根据合同编号 DE-AC02-06CH11357 提供支持。34-ID-C 光束线由美国国家科学基金会资助项目 DMR-9724294 建设。UCL 开展的工作得到英国工程与自然科学研究委员会和欧洲研究委员会的支持。布鲁克海文国家实验室开展的工作由美国能源部科学办公室基础能源科学办公室根据合同编号 DE-SC0012704 提供支持。为实现开放获取，作者已对由此产生的任何作者接收稿版本采用知识共享署名（CC BY）许可。

参考文献

Assefa TA、Cao Y、Diao J 等（2020）原型高温超导体 $la_{1.875}ba_{0.125}cuo_4$ 中低温正交晶畴的标度行为。《物理评论B》101:054104。https://doi.org/10.1103/PhysRevB.101.054104，网址 https://link.aps.org/doi/10.1103/PhysRevB. 101.054104 Ba JL、Kiros JR、Hinton GE（2016）层归一化。arXiv 预印本 https://doi.org/10.48550/a rXiv.1607.06450，网址 https://arxiv.org/abs/1607.06450 Bates RHT（1984）局域正图像二维傅里叶相位问题解的唯一性。《计算机视觉、图形学与图像处理》25(2):205–217。https://doi.org/10.1016/0734-189X(84)90103-8 Bates RHT、Napier PJ（1972）干涉测量中相位误差的识别与消除。《皇家天文学会月报》158(4):405–424。https://doi.org/10.1093/mnras/158.4. 405，网址 https://doi.org/10.1093/mnras/158.4.405 Beinert R 、 Plonka G （ 2015 ）由傅里叶幅值进行一维离散相位恢复中的歧义。《傅里叶分析与应用杂志》21(6):1169–1198。https://doi.org/10.1007/s00041-015-9405-2，网址 https://doi.org/10.1007/s00041-015-9405-2 Bendory T、Beinert R、Eldar YC（2017）傅里叶相位恢复：唯一性与算法。载于：Boche H、Caire G、Calderbank R 等（编）《压缩感知及其应用》。《应用与数值调和分析》，施普林格，第 55–91 页，https://doi.org/10.1007/978-3-319-69802-1 2，网址 https://doi.org/10.1007/978-3-319-69802-1 2 Bruck YM、Sodin LG（1979）论图像重建问题的歧义。《光学通讯》30(3):304–308。https://doi.org/10.1016/0030-4018(79)90358-4，网址 https: //doi.org/10.1016/0030-4018(79)90358-4 Carnis J、Gao L、Labat S 等（2019）迈向布拉格相干 X 射线衍射成像中应变的定量测定：重建中的伪影与符号约定。《科学报告》9:17357。https://doi.org/10.1038/s41598-019-53774-2 Chen C、Miao J、Wang CW 等（2007）优化技术在非晶态 X 射线衍射显微术中的应用：引导式混合输入输出法。《物理评论B》76:064113。https: //doi.org/10.1103/PhysRevB.76.064113 Cherukara MJ、Nashed YSG、Harder RJ 等（2018）利用深度生成网络进行实时相干衍射反演。《科学报告》8:16520。https://doi.org/10.1038/s41598-018-34525-1 Cherukara MJ、Zhou T、Nashed YSG 等（2020）人工智能赋能的高分辨率扫描相干衍射成像。《应用物理快报》117(4):044103。https://doi.org/10.1063/5.0013065 Clark JN、Huang X、Harder R 等（2012）高分辨率三维部分相干衍射成像。《自然·通讯》3:993。https://doi.org/10.1038/ncomms1994 Clark JN、Beitra L、Xiong G 等（2013）单个金纳米晶体中晶格动力学的超快三维成像。《科学》341(6141):56–59。https://doi.org/10.1126/science.1236034，网址 https://www.science.org/doi/10.1126/science.1236034 Dong Y、Yang D、Liu J 等（2025）微米级磁铁矿晶体低温畴形成过程中的对称性破缺。《物理学文献》100(8):085937。https://doi.org/10.1088/1402-4 896/adee52，网址 https://doi.org/10.1088/1402-4896/adee52 Dosovitskiy A、Beyer L、Kolesnikov A 等（2021）一幅图像胜过 16x16 个词：用于大规模图像识别的 Transformer。载于：《国际学习表征会议》（ICLR），https://doi.org/10.48550/arXiv.2010.11929

Duhamel P, Vetterli M (1990) 快速傅里叶变换：教程式综述与最新进展。《信号处理》19(4):259–299. https://doi.org/10.1016/0165-1684(90)90158-UElser V (2003) 通过迭代投影进行相位恢复。《美国光学学会会刊A辑》20(1):40–55. https://doi.org/10.1364/JOSAA.20.000040Favre-Nicolin V, Girard G, Leake S, et al (2020a) PyNX：基于算子的相干X射线成像高性能计算工具包。《应用晶体学杂志》53(5):1404–1413.https://doi.org/10.1107/S1600576720010985，URL https://doi.org/10.1107/S1600576720010985Favre-Nicolin V, Leake S, Chushkin Y (2020b) 自由对数似然：相干衍射成像的无偏度量。《科学报告》10:2664. https://doi.org/10.1038/s41598-020-57561-2Fienup JR (1982) 相位恢复算法：比较。《应用光学》21(15):2758–2769. https://doi.org/10.1364/AO.21.002758Fienup JR (1987) 利用支撑约束从傅里叶变换的模重建复值物体。《美国光学学会会刊A辑》4(1):118–123. https://doi.org/10.1364/JOSAA.4.000118，URL https://doi.org/10.1364/JOSAA.4.000118Gan W, ZhaiQ, McCann MT, et al (2024) PtychoDV：基于视觉Transformer的ptychographic图像重建深度展开网络。《IEEE开放信号处理期刊》5:539–547.https://doi.org/10.1109/OJSP.2024.3375276Gao Y, Huang X, Yan H, et al (2021) 通过同时重建多个衍射峰进行布拉格相干衍射成像。《物理评论B》103:014102. https://doi.org/10.1103/PhysRevB.103.014102，URL https://link.aps.org/doi/10.1103/PhysRevB.103.014102Gerchberg RW, Saxton WO (1972) 一种根据图像面和衍射面图像确定相位的实用算法。《Optik》35(2):237–246 Goodman JW (2015)《统计光学》，第2版。约翰·威立父子出版公司，霍博肯，新泽西州 Guibas J, Mardani M, Li Z, et al (2021) 自适应傅里叶神经算子：用于Transformer的高效词元混合器。arXiv预印本 arXiv:211113587 https://doi.org/10.48550/arXiv.2111.13587Guizar-Sicairos M, Fienup JR (2008) 基于傅里叶加权投影的相位恢复。《美国光学学会会刊A辑》25(3):701–709. https://doi.org/10.1364/JOSAA.25.000701Harder R (2021) 实时相干衍射成像中的深度神经网络。IUCrJ 8(Pt 1):1–3.https://doi.org/10.1107/S2052252520016590Harris CR, Millman KJ, van der Walt SJ, et al (2020) 使用NumPy进行数组编程。《自然》585:357–362. https://doi.org/10.1038/s41586-020-2649-2Huang X, Harder R, Xiong G, et al (2011) 三维相干衍射成像中的传播唯一性。《物理评论B》83:224109. https://doi.org/10.1103/PhysRevB.83.224109，URLhttps://link.aps.org/doi/10.1103/PhysRevB.83.224109Janesick JR (2001)《科学电荷耦合器件》。SPIE出版社，贝灵厄姆，华盛顿州 Janesick JR (2007)《光子传递》。SPIE出版社，贝灵厄姆，华盛顿州 Kim D, Chung M, Carnis J, et al (2018) 铂纳米晶体上甲烷氧化活性位点的定位。《自然·通讯》9(1):3422. https://doi.org/10.1038/s41467-018-05464-2Lee DD, Seung HS (1999) 通过非负矩阵分解学习物体的组成部分。《自然》401(6755):788–791. https://doi.org/10.1038/44565Lee-Thorp J, Ainslie J, Eckstein I, et al (2022) Fnet：使用傅里叶变换混合词元。载于：2022年北美计算语言学协会年会论文集：人类语言技术。计算语言学协会，美国西雅图，第4296–4313页，https://doi.org/10.18653/v1/2022.naacl-main.319

Li Z, Kovachki NB, Azizzadenesheli K, et al (2021) 用于参数偏微分方程的傅里叶神经算子。载于：国际学习表征会议，URL https://op enreview.net/forum?id=c8P9NQVtmnO Liu J, Yang D, Suzana AF, et al (2025) 电场驱动的钛酸钡$_3$纳米颗粒畴壁动力学。Physical Review B 111:054101. https://doi.org/10.1103/PhysRevB.111.054101 Liu T, Liu J, Li L, et al (2022) 富锂层状氧化物正极结构退化的起源。《自然》606(7913):305–312. https://doi.org/10.1038/s41586-022-04689-y Loshchilov I, Hutter F (2019) 解耦权重衰减正则化。载于：国际学习表征会议（ICLR），https://doi.org/10.48550/arXiv.1711.05101，URL https: // arxiv.org/abs/1711.05101 Luke DR (2005) 用于衍射成像的松弛平均交替反射法。《逆问题》21(1):37–50. https://doi.org/10.1088/0266-5611/21/1/004 Marchesini S, He H, Chapman HN, et al (2003) 仅根据衍射图样进行X射线图像重建。《物理评论B》68(14):140101. https://doi.org/10.1103/PhysRevB.68.140101 Miao J, Charalambous P, Kirz J, et al (1999) 扩展X射线晶体学方法，以实现对微米级非晶样品的成像。《自然》400(6742):342–344. https: // doi.org/10.1038/22498 Monteforte M, Estandarte AK, Chen B, et al (2016) 一种用于利用相干X射线衍射成像分析细小纳米晶体的新型二氧化硅稳定化方法。《同步辐射杂志》23(4):953–958. https://doi.org/10.1107/S1600577516006408 Patro BN, Agneeswaran VS (2023) 散射视觉变换器：谱混合至关重要。arXiv预印本 https://doi.org/10.48550/arXiv.2311.01310，URL https://arxiv.org/abs/2311.01310 Pfeifer MA, Williams GJ, Vartanyants IA, et al (2006) 纳米晶体内部变形场的三维映射。《自然》442(7098):63–66. https://doi.org/10.1038/nature04867 Rao Y, Zhao W, Zhu Z, et al (2021) 用于图像分类的全局滤波器网络。arXiv预印本 https://doi.org/10.48550/arXiv.2107.00645，URL https://arxiv.org/abs/2107.00645 Rivenson Y, Zhang Y, G¨unaydin H, et al (2018) 使用神经网络中的深度学习进行相位恢复与全息图像重建。Light: Science & Applications 7:17141. https://doi.org/10 .1038/lsa.2017.141 Robinson I, Assefa TA, Cao Y, et al (2020) $La_{2-x}Ba_xCuO_4$正交相的畴结构纹理。《超导与新型磁性杂志》33:99–106. https://doi.org/10 .1007/s10948-019-05252-z Robinson IK, Harder R (2009) 纳米尺度应变的相干X射线衍射成像。《自然·材料》8(4):291–298. https://doi.org/10.1038/nmat2400 Robinson IK, Assefa T, Cao Y, et al (2026) 晶体中的相位畴。准备中的手稿 Sayre D (1952) 香农定理的一些推论。《晶体学报》5:843–843. https://doi.org/10.1107/S0365110X52002276 Setter N, Damjanovic D, Eng L, et al (2006) 铁电薄膜：材料、性质与应用综述。《应用物理杂志》100(5):051606. https://doi.org/10.1063/1.2336999 Ulvestad A, Nashed Y, Beutier G, et al (2017) 在布拉格相干衍射成像中使用引导算法识别缺陷。《科学报告》7:9920. https://doi.org/10.1038/s41598-017-09582-7 Vartanyants IA, Robinson IK (2001) 相干X射线衍射对小晶体成像的部分相干效应。《凝聚态物理杂志》13(47):10593–10611. https://doi. org/10.1088/0953-8984/13/47/305

Vaswani A, Shazeer N, Parmar N, 等（2017）Attention is all you need。载于：神经信息处理系统进展，https://doi.org/10.48550/arXiv.1706.03762，网址https://arxiv.org/abs/1706.03762Vu TS, Ha MQ, Dam HC（2025）Pid3net：一种用于动态现象单次相干X射线衍射成像的深度学习方法。npj计算材料 11:66。https://doi.org/10.1038/s41524-025-01549-xWilliams GJ, Quiney HM, Peele AG, 等（2007）相干衍射成像与部分相干性。Physical Review B 75(10):104102。https://doi.org/10.1103/PhysRevB.75.104102Wu L, Juhas P, Yoo S, 等（2021a）深度神经网络对相位畴的复数成像。IUCrJ 8(1):12–21。 https://doi.org/10.1107/S2052252520013780Wu L, Yoo S, Suzana AF,等（2021b）通过深度卷积神经网络实现三维相干X射线衍射成像。npj计算材料 7:175。https://doi.org/10.1038/s41524-021-00644-zWu L, Yang D, Wang W, 等（2025）利用相干X射线动力学衍射揭示纳米尺度晶体变形。npj计算材料 11:379。https://doi.org/10.1038/s41524-025-01858-1，网址 https://doi.org/10.1038/s41524-025-01858-1 Yao Y, Chan H, Sankaranarayanan S, 等（2022）AutoPhaseNN：用于三维纳米尺度布拉格相干衍射成像的无监督物理感知深度学习。npj计算材料 8:124。https://doi.org/10.1038/s41524-022-00803-wYu X, Wu L, Lin Y, 等（2024）利用深度复值神经网络对外延薄膜进行超快布拉格相干衍射成像。npj计算材料 10:24。https://doi.org/10.1038/s41524-024-01208-7Zhuang Z, Yang D, Hofmann F, 等（2022）利用双重深度图像先验实现实用相位恢复。网址https://arxiv.org/abs/2211.00799，arXiv:2211.00799