

# 基于视觉变换器的高精度光子晶体能带预测

陈冠桦<sup>1)</sup> 姚俊驰<sup>1)</sup> 朱惠芳<sup>2)</sup> 智婷<sup>2)</sup> 汪金<sup>2)</sup>  
薛俊俊<sup>2)</sup> 陈琳<sup>2)</sup> 陶涛<sup>3)</sup> 陶志阔<sup>2)†</sup>

1) (南京邮电大学通信与信息工程学院, 南京 210003)

2) (南京邮电大学电子与光学工程学院、柔性电子(未来技术)学院, 南京 210003)

3) (南京大学电子科学与工程学院, 南京 210023)

(2025年6月17日收到; 2025年7月11日收到修改稿)

光子晶体因其独特的能带结构在光子学领域具有重要应用前景, 而准确预测其能带结构对于光子器件的设计与优化也至关重要. 鉴于此, 本文应用视觉变换器 (vision transformer, ViT) 模型, 探索高效、准确的光子晶体能带结构预测方法. 首先, 通过传统数值仿真方法得到光子晶体的能带结构数据, 构建了训练和测试数据集; 其次, 利用数据集对 ViT 模型进行训练, 训练过程中模型展现出良好的学习能力, 损失函数值持续下降, 最低可至  $4.42 \times 10^{-6}$ ; 最终, 测试结果表明, ViT 模型预测平均均方误差 (MSE) 低至  $3.46 \times 10^{-5}$ , 决定系数 ( $R^2$ ) 达到 0.9996, 表明 ViT 模型具有极高的预测精度和良好的泛化能力. 研究表明, ViT 模型能够有效预测光子晶体的能带结构, 为光子晶体相关研究和应用提供一种新的高效预测工具, 有望推动光子器件设计的进一步发展.

**关键词:** 光子晶体, 能带结构, 预测, 视觉变换器

**PACS:** 02.60.Cb, 02.70.-c, 42.70.Qs, 87.10.Kn

**CSTR:** 32037.14.aps.74.20250778

**DOI:** 10.7498/aps.74.20250778

## 1 引言

随着计算能力的飞速发展以及人工智能领域的蓬勃兴起<sup>[1,2]</sup>, 众多科学领域正经历着变革性的变化. 光子晶体<sup>[3-5]</sup> 凭借其独特的能力, 即通过周期性排列的介质结构来操控光子并形成光子带隙, 已成为现代光学和光电子学的基石. 光子晶体能带结构图, 通过在特定频率范围内操控光的传播, 为解锁诸如光通信<sup>[6]</sup>、光子集成电路<sup>[7]</sup>、光学传感器<sup>[8]</sup> 以及先进光子器件<sup>[9-14]</sup> 等高性能应用领域的关键所在. 然而, 光子晶体能带结构的精确设计与调控仍面临巨大挑战, 这主要是因为传统的能带结构计算方法存在计算复杂度高以及处理时间长等问题. 这些限制因素阻碍了实际应用中所需的快速设计

与优化进程.

近年来, 人工智能的迅速发展, 尤其是神经网络为解决复杂的物理问题引入了一种强大的数据驱动方法<sup>[15-18]</sup>. 神经网络<sup>[19-22]</sup> 凭借其能够通过大量的数据学习自动提取输入参数与输出结果之间复杂映射关系的能力, 为快速且准确的预测提供了一种有前景的替代方案. 在光子晶体领域, 利用神经网络预测光子晶体能带结构已受到越来越多的关注<sup>[23,24]</sup>. 然而, 这一领域仍处于起步阶段, 目前仍在努力解决与预测精度、模型泛化能力以及实际应用效果相关的挑战. 尽管已有研究在利用神经网络预测光子晶体能带结构方面取得了一定进展, 但现有方法在预测精度和模型泛化能力方面仍存在不足, 且在实际应用中的有效性仍需进一步验证. 因此, 有必要开展更多的研究工作, 以提高预测精

† 通信作者. E-mail: zktao@njupt.edu.cn

度、增强模型泛化能力,并探索其在实际应用中的潜力。

本文旨在应用一种新型的神经网络模型,即视觉变换器 (vision transformer, ViT),用于快速且准确预测光子晶体能带结构.研究过程中,利用传统的有限元仿真生成数据集,训练 ViT 模型以提高预测精度和泛化能力,通过将该模型的性能与传统方法进行比较,并探索其在结构优化中的潜在应用,为光子晶体的应用提供有价值的见解.本研究不仅有望推动光子晶体能带结构预测技术的发展,还为光子晶体在高性能光子器件中的应用提供了新的思路和方法。

## 2 数据集准备

本研究首先利用 COMSOL 软件,采用平面波展开法 (plane-wave expansion, PWE) 结合有限元法 (finite element method, FEM),计算光子晶体的能带结构.为了训练能够预测能带结构的模型,需要生成足够数量的代表性体积单元 (representative volume element, RVE) 样本,并获取其对应的光子晶体能带结构数据.本文采用二维光子晶体作为研究对象.光子晶体的基本单元由空气介质和介质柱组成,空气介质与介质柱通过边界曲线分隔.边界曲线采用基圆半径的三阶傅里叶展开形式<sup>[25]</sup>,其参数方程定义如下:

$$\begin{cases} x = f_{\Omega}(\theta) \cdot \cos \theta, \\ y = f_{\Omega}(\theta) \cdot \sin \theta, \\ f(\theta) = r_0 [1 + c_1 \cos(n_1 \theta) + c_2 \cos(n_2 \theta) \\ + c_3 \cos(n_3 \theta)], \end{cases} \quad (1)$$

其中,  $x$  和  $y$  是边界曲线上某点的水平和垂直坐标;  $r_0$  是基圆的半径;  $f_{\Omega}(\theta)$  是以  $\theta$  为自变量的曲线函数,  $\theta$  的取值范围为  $[0, 2\pi]$ .  $c_1$ ,  $c_2$  和  $c_3$  是傅里叶系数,用于控制边界曲线形状的变化;  $n_1$ ,  $n_2$  和  $n_3$  是傅里叶频率参数,用于确定曲线的周期性变化.通过随机选择  $c_1$ ,  $c_2$ ,  $c_3$ ,  $n_1$ ,  $n_2$ ,  $n_3$  和  $r_0$  的值,构建了 4800 个不同的基本单元,并利用 COMSOL 软件得到了 4800 个对应的光子晶体能带结构.每个能带结构包含 8 条色散曲线,每条色散曲线取 31 个数据点,共 248 个特征频率.随后,选取其中 4500 个光子晶体能带数据作为训练数据集,300 组光子晶体能带数据作为测试数据集。

图 1(a) 展示了具有随机生成几何参数的 4 种不同光子晶体单元结构.关键参数从预定义的范围随机采样:傅里叶系数  $c_1$ ,  $c_2$ ,  $c_3$  分别在  $[0.1, 0.3]$ ,  $[0.05, 0.2]$  和  $[0.02, 0.1]$  范围内变化;傅里叶频率参数  $n_1$ ,  $n_2$ ,  $n_3$  取 1—30 之间的整数值;而基圆半径  $r_0$  在  $[50, 100]$  nm 范围内均匀分布.这种随机化的参数空间确保了本系统研究中对可能的光子晶体结构的多样化表示.图 1(b) 给出了某一种光子晶体能带结构以及在不可约布里渊区顶点处的高对称点示意图。

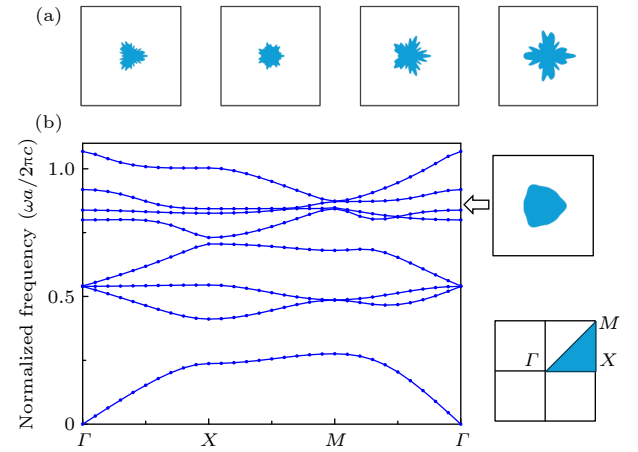


图 1 (a) 随机生成的 4 个基本单元示意图; (b) 某一种光子晶体能带结构以及在不可约布里渊区顶点处的高对称点  
Fig. 1. (a) Four constructed randomly unit cells; (b) the photonic band diagram of one proposed structure and high symmetric point at the vertex of irreducible Brillouin region.

## 3 ViT 模型建立

图 2 展示了分别使用传统方法和神经网络方法对光子晶体能带结构进行计算的过程.传统方法利用平面波展开法和有限元方法进行求解,通常需要将问题从实空间转换到倒空间进行求解.这一过程的必要性在于,光子晶体的能带结构本质上描述的是光在周期性介质中的传播特性,而这些特性在倒空间中能够更加直观和有效地进行分析.具体而言,倒空间中的波矢是描述光在周期性介质中传播的关键参数.在计算能带结构时,需要在布里渊区内扫描不同的波矢,通过数值方法计算每个波矢对应的本征频率.这种方法虽然物理意义明确,能够精确地计算出每个波矢点的能带结构,但其计算成本较高,尤其是对于复杂的几何结构和高维参数空间,计算过程耗时且资源密集。

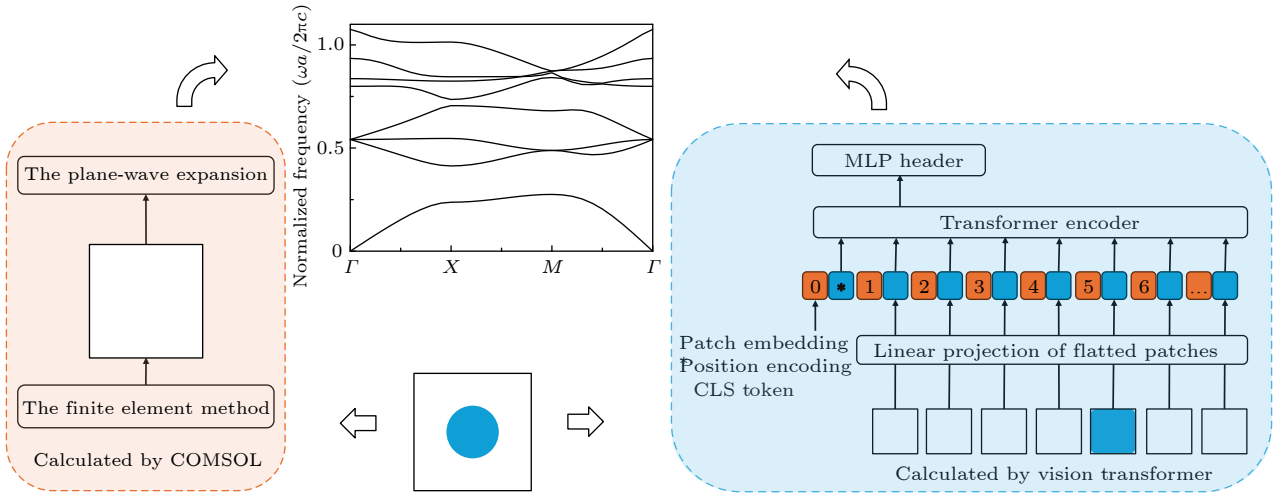


图 2 COMSOL 计算光子晶体能带结构以及利用 ViT 模型进行预测的框架图

Fig. 2. Framework of the photonic band diagram calculation by COMSOL and the prediction by ViT model.

相比之下,神经网络方法则采用了一种截然不同的端到端预测思路.神经网络,尤其是深度学习模型,能够直接从输入的几何参数预测出能带结构.这种方法的核心在于数据驱动的学习,而非基于物理模型的解析或数值求解.神经网络通过大量的训练数据学习输入与输出之间的复杂映射关系.这些数据通常通过传统的数值方法预先生成.一旦模型训练完成,就可以直接用于预测新的几何参数对应的能带结构,而无需进行复杂的数值计算.神经网络的这种端到端预测能力主要得益于其强大的特征提取能力和泛化能力.神经网络能够自动提取输入数据中的局部和全局特征,并学习到这些特征与能带结构之间的复杂非线性关系.此外,训练好的神经网络在预测新的能带结构时速度极快,比传统方法快几个数量级,且具有良好的泛化能力,即使在训练数据范围之外的参数空间中也能保持较高的预测精度.

本研究选择基于 ViT 架构构建深度学习模型来预测光子晶体的能带结构,该选择基于 ViT 模型的独特机制及其在处理类似任务时展现出的潜在优势. ViT 模型的核心在于其多头自注意力机制,该机制允许模型动态地评估图像块之间的相互作用,而不仅仅是局部邻域信息.这种能力对于光子晶体的能带结构预测尤为重要,因为光子晶体能带结构不仅依赖于单个单元格的局部几何特征,还依赖于这些单元格之间的相互作用.光子晶体的周期性排列和单元格之间的相互作用对光的传播和带隙形成机制有着显著影响. ViT 模型的自注意力

机制能够捕捉这些复杂的相互作用,这使其在理论上非常适合处理光子晶体的能带结构预测任务.

位置编码的引入进一步增强了 ViT 模型对空间信息的处理能力.光子晶体具有明显的周期性特征,其能带结构的形成与这种周期性紧密相关.位置编码使得模型能够明确地识别和利用这些周期性特征,从而更准确地预测光子晶体能带结构.这种对空间信息的明确保留,使得 ViT 模型在处理光子晶体的图像数据时,能够更好地理解光子能带结构的形成机制.

ViT 模型的机制特别适合光子能带结构的预测任务,因为其能同时处理局部和全局信息.在光子晶体中,局部几何特征对能带结构的形成起着关键作用,但这些局部特征之间的全局排列同样重要. ViT 模型通过自注意力机制和位置编码,能够有效地捕捉这些特征之间的复杂关系,从而更准确地预测光子晶体能带结构.这种能力不仅克服了传统卷积神经网络 (CNN) 由于局部感受野限制而难以捕捉长程相互作用的问题,还能够更好地适应光子晶体的物理特性,为光子晶体能带结构的预测提供了一种更为有效的方法.

在本研究中, ViT 模型的关键超参数配置如下:图像块大小为  $16 \times 16$  像素, Transformer 层数为 12 层,注意力头数为 12 个,嵌入维度为 768, MLP 维度为 3072, Dropout 率为 0.1,嵌入 Dropout 率为 0.1. 这些超参数适用于多种视觉任务,并在预训练阶段经过优化以实现高效的特征提取和学习.

## 4 模型训练与结果分析

### 4.1 数据预处理

在模型训练前, 数据预处理是确保模型性能的关键环节. 本文对训练数据集进行了标准化处理, 以保障数据的一致性与模型的稳定性. 将图像尺寸统一调整为  $224 \times 224$  像素, 以契合预训练模型的输入要求; 同时, 对图像数据实施归一化处理, 旨在加速模型收敛进程并增强数值计算的稳定性. 训练集方面, 选用 4500 组数据进行模型训练, 这不仅确保了数据的丰富多样性, 也为模型的泛化能力奠定了基础.

### 4.2 模型训练与测试

在完成数据预处理后, 本文采用基于 ViT 架构的定制深度学习模型. 训练过程中, 以均方误差 (MSE)<sup>[26]</sup> 作为损失函数, 并结合 Adam 优化器进行参数优化. 在每个训练批次 (批量大小为 8) 中, 模型执行前向传播以生成预测结果, 随后基于预测值与实际值之间的 MSE 损失评估性能, 并通过反向传播更新参数. 整个训练过程共计 100 个周期. 在每个周期结束时, 通过测试集上的平均 MSE 损

失监控模型的泛化能力, 并根据需要对超参数进行微调:

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (f_{\text{predicted},i} - f_{\text{simulated},i})^2. \quad (2)$$

此外, 本研究引入了决定系数  $R^2$  来全面评估模型的性能. 这一指标反映了预测值与实际值之间的相关性, 其值越接近 1, 表明模型的拟合效果越佳.

$$R^2 = 1 - \frac{\sum_{i=1}^N (f_{\text{predicted},i} - f_{\text{simulated},i})^2}{\sum_{i=1}^N (f_{\text{simulated},i} - \overline{f_{\text{simulated}}})^2}. \quad (3)$$

为进一步验证预测精度, 本研究从 300 个数据集中选取 3 个不同单元格的测试数据集. 图 3(a), (c), (e) 呈现了 COMSOL 仿真与 ViT 预测的不同单元格的特征频率对比, 插图展示了相应的单元格结构. 图 3(b), (d), (f) 展示了 COMSOL 仿真与 ViT 预测的不同单元格的光子带结构图对比, 分别对应于图 3(a), (c), (e) 中的单元格. 结果表明, 这 3 种情况下的  $R^2$  值均能达到 0.999 以上, ViT 预测结果在趋势和数值上与 COMSOL 得到的结果高度一致.

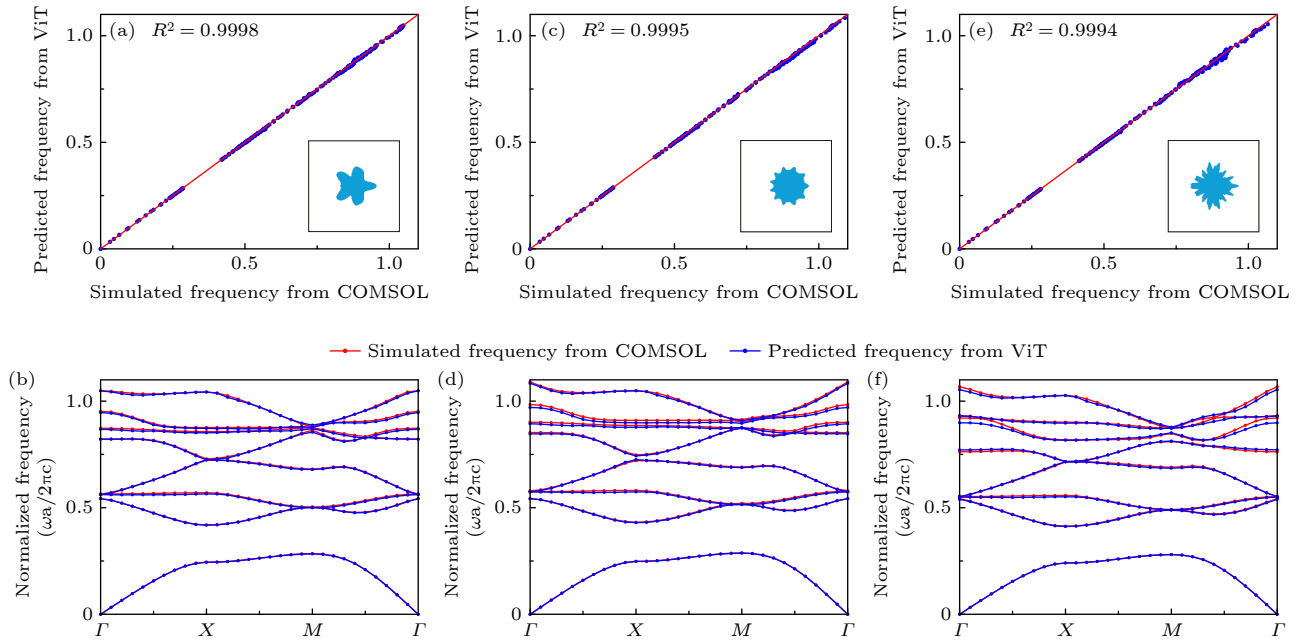


图 3 (a), (c), (e) COMSOL 仿真与 ViT 预测的不同单元格的特征频率对比, 插图展示了相应单元格结构; (b), (d), (f) COMSOL 仿真与 ViT 预测的不同单元格的的光子带结构图对比, 分别对应于 (a), (c), (e) 中的单元格

Fig. 3. (a), (c), (e) Comparison of eigen frequencies simulated by COMSOL versus predicted by ViT for different unit cells shown in the inset; (b), (d), (f) comparison of photonic band diagrams simulated by COMSOL versus predicted by ViT for different unit cells shown in panels (a), (c), (e), respectively.

为进一步探究预测特征, 本研究对全部 300 个测试数据集中的 74400 个特征频率进行分析. 在此基础上, 对绝对误差的分布进行统计分析, 其计算式为

$$\text{Absolute error} = |f_{\text{predicted},i} - f_{\text{simulated},i}|. \quad (4)$$

该指标能够全面反映各模型在预测本征频率时的准确性和可靠性. 图 4(a) 呈现了 ViT 模型在训练过程中的损失函数收敛情况, 图 4(b) 展示了利用 ViT 模型预测的频率与通过 COMSOL 仿真得到的频率之间的绝对误差分布, 以及对应的  $R^2$  值分布. 从图中可以观察到, ViT 模型的最小均方误差 (MMSE) 低至  $4.46 \times 10^{-6}$ , 而总  $R^2$  值高达 0.9996. 此外, 每个单元格的  $R^2$  值大多集中在 0.999425—0.999825 之间, 且大部分绝对误差分布在 0—0.01 内, 这充分表明 ViT 模型在 300 个测试数据集上展现出了卓越的预测精度.

进一步, 通过计算每个能带的 MSE 和  $R^2$  值来评估预测精度. 图 4(c) 展示了不同能带指数对应的 MSE 和  $R^2$  值, 其中蓝色线条代表各个能带的  $R^2$  值, 红色线条对应各能带的 MSE 值. 可以看出, 随着能带指数的增大, MSE 和  $R^2$  值呈现出一定的波动. 然而总体而言, MSE 值保持在相对较低的水平, 而  $R^2$  值则始终接近于 1, 这表明该模型在不同能带间维持了较高的预测精度. 具体而言, 模型对第一能带的预测精度最高, 而随着能带指数的增大, 预测精度出现了轻微的变化, 这表明模型对特定能带的预测能力存在差异. 尽管如此, 模型的整体预测性能依然稳健, 突显了其在不同能带间的适应性和可靠性.

ViT 模型在训练性能方面展现出了显著的优势, 该模型仅需较少的训练周期, 便能达到 MSE

损失值  $4.42 \times 10^{-6}$  的水平. 这一结果不仅体现了训练效率的大幅提升, 更凸显了其在预测方面的卓越性能, 充分证明了其在处理复杂图像数据任务时的高效率与优越性. 在计算效率方面, ViT 模型也展现出了卓越的性能, 平均预测时间为 0.066 s, 远低于传统计算方式. 传统计算方式受到光子晶体具体结构的影响, 对于复杂的结构, 传统算法需要划分更多的网络以确保计算精度, 从而导致计算时间大幅增加, 对于本文所采用的硬件能力, 通常在 50—70 s 之间. ViT 模型不仅在预测精度上达到了高水平, 其高效的计算能力也为光子晶体能带结构的快速设计和优化提供了强有力的支持.

另外, 本研究还引入了卷积神经网络 (CNN) 模型, 以与 ViT 模型进行对比分析. 浅层 CNN 模型的超参数配置如下: 输入图像尺寸为  $224 \times 224$  像素, 批量大小为 8. 模型包含一个卷积层, 其卷积核大小为  $3 \times 3$ , 输出通道数为 8, 步长为 1, 并应用了填充操作. 其后是一个最大池化层, 池化核大小为  $2 \times 2$ , 步长为 2. 全连接层部分包含两个层级, 第 1 层全连接层的输入维度为  $8 \times 112 \times 112$ , 输出维度为 64; 第 2 层全连接层的输入维度为 64, 输出维度为 248, 对应于光子晶体能带结构的特征频率. 模型采用 ReLU 激活函数, MSE 作为损失函数, 并使用 Adam 优化器进行参数优化, 学习率设置为 0.001. 模型共训练 100 个周期. 图 5(a) 展示了浅层 CNN 模型在训练周期中的损失收敛情况, 而图 5(b) 则呈现了利用浅层 CNN 模型预测与 COMSOL 仿真频率之间的绝对误差分布以及对应的  $R^2$  值分布. 从图中可以看出, 浅层 CNN 模型的 MMSE 计算结果为  $6.51 \times 10^{-3}$ , 总  $R^2$  值为 0.9303, 低于 ViT 模型预测的精确度.

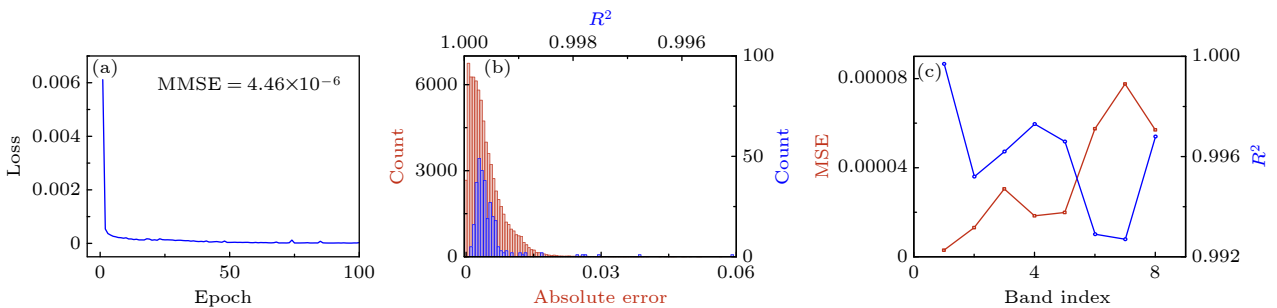


图 4 (a) ViT 模型在训练周期的损失收敛情况; (b) 使用 ViT 预测与仿真之间的绝对误差分布以及  $R^2$  的分布; (c) 不同能带对应的 MSE 和  $R^2$

Fig. 4. (a) ViT model's loss convergence over training epochs; (b) distribution of absolute errors between predicted and simulated frequencies using ViT, along with the distribution of  $R^2$  values; (c) MSE and  $R^2$  values for different band index.

同时,为进一步验证 ViT 在光子晶体能带结构的预测能力,引入深层 CNN 对比分析其预测效果. 本文采用 12 层结构的 CNN 架构,输入图像尺寸为  $224 \times 224$  像素,批量大小为 8. 模型包含 12 个卷积层,每个卷积层使用  $3 \times 3$  的卷积核,输出通道数从 64 开始逐渐增至 512. 每两个卷积层后跟一个  $2 \times 2$  的最大池化层. 全连接层部分包含两个 4096 维的全连接层,最后是一个 248 维的输出层,对应于光子晶体能带结构的特征频率. 模型采用 ReLU 激活函数, MSE 作为损失函数,并使用 Adam 优化器进行参数优化,学习率设置为 0.001,模型共训练 100 个周期. 从图 5 和图 6 可以看到,不管是浅层还是深层 CNN 模型,其收敛速度和预测精度均不如 ViT 模型. 具体来说, CNN 模型的 MMSE 为  $6.51 \times 10^{-3}$ ,而 ViT 模型的 MMSE 为  $4.42 \times 10^{-6}$ . 此外, CNN 模型的总  $R^2$  值为 0.9303,远低于 ViT 模型的 0.9996.

研究表明, ViT 模型凭借其自注意力机制和位

置编码,在能带结构预测方面表现出色. 自注意力机制通过学习得到的查询 (query)、键 (key) 和值 (value) 投影,动态地模拟所有图像块对之间的相互作用,能够同时捕捉局部特征和全局模式. 多头注意力机制则进一步拓展了这一能力,通过并行处理多种相互作用模式,增强了模型的表现力. 位置编码则在保持注意力机制灵活性的同时,明确保留了空间关系信息. 与具有固定几何约束的 CNN 模型不同, ViT 模型能够根据输入内容自适应地调整空间相关性的权重. 这种架构通过共享权重投影实现了参数效率的提升,同时保留了学习复杂空间依赖关系的能力,即使在训练数据有限的情况下,也能够有效地进行泛化. 内容依赖的注意力机制与明确的位置编码相结合,为分析局部和全局特征共同决定目标属性的周期性结构提供了一个理想的框架,这表明 ViT 模型在光子能带结构预测方面相较于 CNN 等其他模型能够展现出更优异的表达能力.

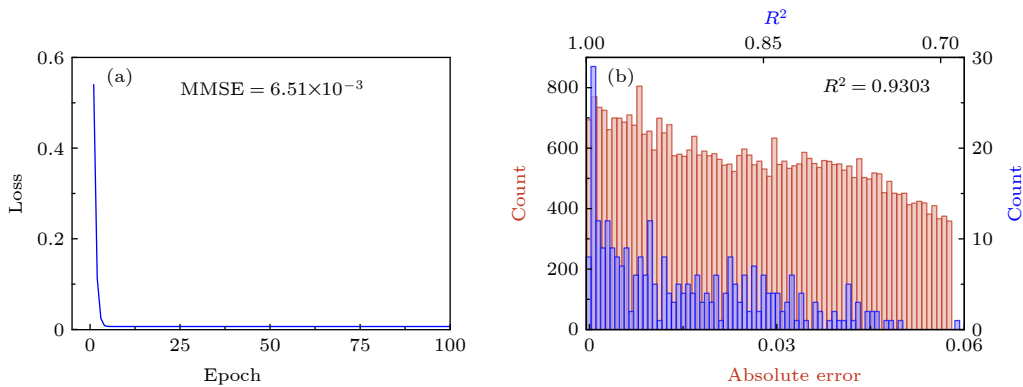


图 5 (a) 浅层 CNN 模型随训练迭代的损失收敛; (b) 使用浅层 CNN 预测频率与模拟频率的绝对误差分布以及  $R^2$  值的分布  
Fig. 5. (a) Shallow CNN model's loss convergence over training epochs; (b) distribution of absolute errors between predicted and simulated frequencies using Shallow CNN, along with the distribution of  $R^2$  values.

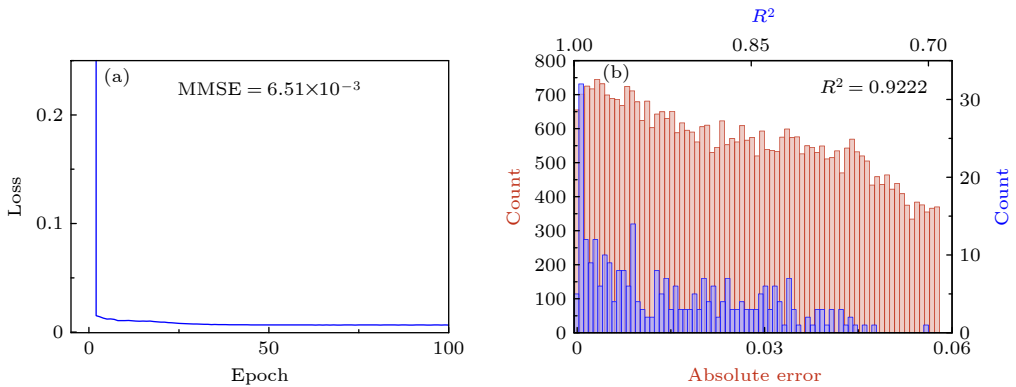


图 6 (a) 深层 CNN 模型随训练迭代的损失收敛; (b) 使用深层 CNN 预测频率与模拟频率的绝对误差分布以及  $R^2$  值的分布  
Fig. 6. (a) Deep CNN model's loss convergence over training epochs; (b) distribution of absolute errors between predicted and simulated frequencies using deep CNN, along with the distribution of  $R^2$  values.

在光子晶体的能带结构预测中, 这种机制与光子晶体的物理特性紧密相关. 光子晶体的能带结构主要受单元胞的局部几何特征 (如边界曲线形状和填充因子) 及其周期排列的影响. ViT 模型的自注意力机制能够捕捉到这些局部特征和全局排列之间的复杂相互作用. 此外, 位置编码的引入使得模型能够更好地保留空间信息, 这对于周期性结构的分析尤为重要, 因为其确保了模型能够识别和利用光子晶体的周期性特征.

基于本研究所采用的硬件配置, ViT 模型在训练阶段呈现出较长的时间需求 (为 105950 s), 相较于深层 CNN 模型的 13441 s 和 MLP 模型的 8651 s, 这一现象主要是由于 ViT 模型的架构复杂性以及对大规模参数优化的需求所导致. ViT 模型在测试阶段的单次预测时间仅需 0.066 s, 虽然略高于 CNN 模型的 0.041 s 和 MLP 模型的 0.009 s, 但考虑到 ViT 在预测精度和泛化能力方面的显著优势, 这种差异在实际应用中的影响相对较小. 进一步彰显了 ViT 模型在预测精度以及与仿真数据拟合度方面的显著优势. 同时, 对于基于不同晶格结构、不同材料 (介电常数) 的光子晶体, 其需要考虑的参数集将更多, 补充材料图 S1 ([online](#)) 给出了增大介电常数参数集后的能带结构预测结果. 450 组数据用于训练, 并预留了 30 组数据用于测试, 结果表明, ViT 模型在光子晶体能带结构预测任务上的准确性仍很高. 进一步, 应用 ViT 模型也可以对基于光子晶体能带结构的其他物理量进行有效预测, 如群速度、态密度、损耗等, 补充材料图 S2 ([online](#)) 给出了群速度的预测结果, 其 MMSE 为  $9.60 \times 10^{-3}$ , 大部分绝对误差集中在较小范围内, 且  $R^2$  值为 0.9671, 展示较好预测精度.

## 5 结 论

本文应用 ViT 模型高效、准确地得到了光子晶体能带结构. 一方面, 借助传统仿真软件 COMSOL 收集光子晶体能带结构的训练与测试数据集; 另一方面, 训练基于 ViT 架构的深度学习模型以预测光子晶体能带结构图. 在训练过程中, 该模型 MSE 损失值降至  $4.42 \times 10^{-6}$ , 充分证明了其能够高效地从数据中提取关键信息并进行学习. 在测试过程中, 平均 MSE 损失值仅为  $3.46 \times 10^{-5}$ ,

且  $R^2$  值高达 0.9996, 展现出卓越的预测精度和泛化能力. 本研究不仅为光子晶体的设计提供了一种高效、准确的新方法, 还为人工智能技术在光学领域的深度应用奠定了坚实基础, 有望推动光子晶体技术的进一步发展与创新.

## 参考文献

- [1] Heaton J, Goodfellow I, Bengio Y, Courville A 2018 *Genet. Program. Evolution Mach.* **19** 305
- [2] LeCun Y, Bengio Y, Hinton G 2015 *Nature* **521** 436
- [3] Yablonovitch E 1995 *Phys. Rev. Lett.* **58** 2059
- [4] John S 1987 *Phys. Rev. Lett.* **58** 2486
- [5] Joannopoulos J D, Meade R D, Winn J N 2008 *Photonic Crystals: Molding the Flow of Light* (Princeton NJ: Princeton Univ. Press)
- [6] Nyachionjeka K, Tarus H, Langat K 2020 *Sci. Afr.* **9** e00511
- [7] Bogaerts W, Pérez D, Capmany J, Miller D A B, Poon J, Englund D, Morichetti F, Melloni A 2020 *Nature* **586** 207
- [8] Fallahi V, Kordrostami Z, Hosseini M 2024 *Sci. Rep.* **14** 2001
- [9] Fu Y L, Hu X Y, Gong Q H 2013 *Phys. Lett. A* **377** 329
- [10] Safinezhad A, Babaei Ghoushti H, Shirri M, Rezaei M H 2021 *Opt. Quant. Electron.* **53** 259
- [11] Giden I H, Mahariq I 2024 *Opt. Quant. Electron.* **56** 170
- [12] Tavares S C da C, Sousa F B de, Oliveira L A de, Sousa F M de, Miranda I R S, Costa M B C 2024 *Opt. Quant. Electron.* **56** 622
- [13] Sathyadevaki R, Raja A S, Sundar D S 2017 *Photon. Netw. Commun.* **33** 77
- [14] Bazian M 2021 *Photon. Netw. Commun.* **41** 57
- [15] Liu Y, Zhao T L, Ju W W, Shi S Q 2017 *J. Materiomics* **3** 159
- [16] Ma W, Liu Z C, Kudyshev Z A, Boltasseva A, Cai W S, Liu Y M 2021 *Nat. Photonics* **15** 77
- [17] Christensen T, Loh C, Picek S, Jakobović D, Jing L, Fisher S, Ceperic V, Joannopoulos J D, Soljačić M 2020 *Nanophotonics* **9** 4183
- [18] Ferreira A da S, Silveira G N M, Figueroa H E H 2018 *SBFoton International Optics and Photonics Conference Campinas, Brazil, October 8–10, 2019* p1
- [19] He K, Zhang X, Ren S, Sun J 2016 *IEEE Conference on Computer Vision and Pattern Recognition Las Vegas, NV, USA, June 27–30, 2016* p770
- [20] Girshick R 2015 *IEEE International Conference on Computer Vision Santiago, Chile, December 7–13, 2015* p1440
- [21] Shen D G, Wu G R, Suk H 2017 *Annu. Rev. Biomed. Eng.* **19** 221
- [22] Vinyals O, Toshev A, Bengio S, Erhan D 2017 *IEEE Trans. Pattern Anal. Mach. Intell.* **39** 652
- [23] Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, Uszkoreit J, Houlsby N 2020 *arXiv: 2010.11929 [cs.CV]*
- [24] Lecun Y, Bottou L, Bengio Y, Haffner P 1998 *Proc. IEEE* **86** 2278
- [25] Li Y L, Yin G H, Yan G W, Yao S 2025 *Mech. Syst. Sig. Process.* **224** 111975
- [26] Michelucci U, Venturini F 2021 *Mach. Learn. Knowl. Extr.* **3** 357

# High-precision prediction of photonic band structure based on vision transformer model

CHEN Guanhua<sup>1)</sup> YAO Junchi<sup>1)</sup> ZHU Huifang<sup>2)</sup> ZHI Ting<sup>2)</sup> WANG Jin<sup>2)</sup>  
XUE Junjun<sup>2)</sup> CHEN Lin<sup>2)</sup> TAO Tao<sup>3)</sup> TAO Zhikuo<sup>2)†</sup>

1) (*School of Communications and Information Engineering, Nanjing University of Post and Telecommunication, Nanjing 210003, China*)

2) (*College of Electronic and Optical Engineering & College of Flexible Electronics, Nanjing University of Post and Telecommunication, Nanjing 210003, China*)

3) (*School of Electronic Science Engineering, Nanjing University, Nanjing 210023, China*)

( Received 17 June 2025; revised manuscript received 11 July 2025 )

## Abstract

Photonic crystals have received widespread attention in the field of photonics due to their unique band structures, which can manipulate the propagation of light through periodic dielectric arrangements. Accurate prediction of these band structures is crucial for designing and optimizing photonic devices. However, traditional numerical simulation methods, such as plane wave expansion and finite element methods, are often limited by high computational complexity and long processing times. In this study, we explore the application of the vision transformer (ViT) model to predicting the band structures of photonic crystals efficiently and accurately. To further validate the superiority of the ViT model, we also conduct experiments by using CNN and MLP models on the same scale for band structure prediction. We first generate a dataset of photonic band structures by using traditional numerical simulations and then train the ViT model on this dataset. The ViT model demonstrates excellent learning capabilities, with the loss function value decreasing to as low as  $4.42 \times 10^{-6}$  during training. The test results show that the average mean squared (MSE) error of the ViT model predictions is  $3.46 \times 10^{-5}$ , and the coefficient of determination ( $R^2$ ) reaches 0.9996, indicating high prediction accuracy and good generalization capability. In contrast, the CNN and MLP models, despite being trained on the same dataset and having the same computational resource allocation, show higher MSE values and lower  $R^2$  scores. This highlights the superior performance of the ViT model in predicting the band structures of photonic crystals. Our study shows that the ViT model can effectively predict the band structures of photonic crystals, providing a new and efficient prediction tool for relevant research and applications. This work is expected to advance the development of photonic device design by offering a rapid and accurate alternative to traditional methods.

**Keywords:** photonic crystal, band structure, prediction, vision transformer

**PACS:** 02.60.Cb, 02.70.-c, 42.70.Qs, 87.10.Kn

**DOI:** [10.7498/aps.74.20250778](https://doi.org/10.7498/aps.74.20250778)

**CSTR:** [32037.14.aps.74.20250778](https://cstr.cn/32037.14.aps.74.20250778)

† Corresponding author. E-mail: [zktao@njupt.edu.cn](mailto:zktao@njupt.edu.cn)



## 基于视觉变换器的高精度光子晶体能带预测

陈冠桦 姚俊驰 朱惠芳 智婷 汪金 薛俊俊 陈琳 陶涛 陶志阔

### High-precision prediction of photonic band structure based on vision transformer model

CHEN Guanhua YAO Junchi ZHU Huifang ZHI Ting WANG Jin XUE Junjun CHEN Lin TAO Tao  
TAO Zhikuo

引用信息 Citation: *Acta Physica Sinica*, 74, 190203 (2025) DOI: 10.7498/aps.74.20250778

CSTR: 32037.14.aps.74.20250778

在线阅读 View online: <https://doi.org/10.7498/aps.74.20250778>

当期内容 View table of contents: <http://wulixb.iphy.ac.cn>

## 您可能感兴趣的其他文章

### Articles you may be interested in

基于时延光子储备池计算的混沌激光短期预测

Short-time prediction of chaotic laser using time-delayed photonic reservoir computing

物理学报. 2021, 70(15): 154209 <https://doi.org/10.7498/aps.70.20210355>

一种基于束缚态的可调等离子体光子晶体窄带滤波器

A tunable narrow-band plasma photonic crystal filter based on bound state

物理学报. 2021, 70(17): 175201 <https://doi.org/10.7498/aps.70.20210241>

Li(Na)AuS体系拓扑绝缘体材料的能带结构

Band structure of topological insulator Li(Na)AuS

物理学报. 2021, 70(2): 027101 <https://doi.org/10.7498/aps.70.20200885>

高O/Ga比的 $a\text{-GaO}_x$ 能带结构及其金属-半导体界面电子传输机制

Band structure of high O/Ga ratio  $a\text{-GaO}_x$  and its electron transport mechanism at metal-semiconductor interface

物理学报. 2025, 74(13): 137103 <https://doi.org/10.7498/aps.74.20241571>

二硫化钼的电子能带结构和低温输运实验进展

Experimental research progress of electronic band structure and low temperature transport based on molybdenum disulfide

物理学报. 2022, 71(12): 127306 <https://doi.org/10.7498/aps.71.20220015>

基于二维六方氮化硼材料的光子晶体非对称传输异质结构设计

Design of asymmetric transmission of photonic crystal heterostructure based on two-dimensional hexagonal boron nitride material

物理学报. 2021, 70(2): 028501 <https://doi.org/10.7498/aps.70.20200741>