基于海杂波的有效波高估计谱点 Transformer

Yi Zhou、Li Wang、Hang Su、Tian Wang

arXiv:2601.04581v1 [eess.SP] 8 Jan 2026

摘要—本文提出了一种基于Transformer的方法（SPT），利用稀疏谱点估计有效波高（H_s）。根据经验观察，只有少数功率较强的谱点会对波浪能量产生贡献，所提出的SPT通过多维特征表征，有效整合海洋表面波的几何和谱特征，以估计 H_s。实验揭示了一个有趣的现象：SPT学到的特征与物理色散关系高度吻合，所选点的贡献分数图集中分布在色散曲线上。与处理图像序列和完整频谱的传统视觉网络相比，SPT在 H_s 回归任务中表现更优，同时消耗的计算资源显著更少。在消费级GPU上，SPT仅需不到4分钟即可完成针对1080个海杂波图像序列的回归模型训练，展现出降低雷达波浪测量系统部署成本的潜力。SPT的开源实现将发布于https://github.com/joeyee/spt

I. 引言

有效波高（H_s）是描述海况的重要统计参数，其准确估算一直是物理海洋学和遥感领域关注的焦点。传统的波浪测量技术，如波浪浮标和声学多普勒流速剖面仪（ADCP），虽然精度很高，但只能提供单点时间序列数据。这些方法无法捕捉波浪的空间分布，且维护成本高，因此不适合在大范围区域广泛部署。

相比之下，X波段海洋雷达作为一种主动微波遥感设备，具有全天候运行、米级空间分辨率、秒级时间分辨率以及最远可达数公里的广阔覆盖范围等优势。尽管其最初是为船舶导航和避碰而设计的，但雷达接收到的海杂波信号包含丰富的海洋表面动力学信息。自20世纪80年代以来，利用X波段雷达图像序列反演波浪参数（如波高、周期、方向和流速），已成为海洋遥感的一个重要分支，并催生了 WaMoS II 等商业系统 [1]。

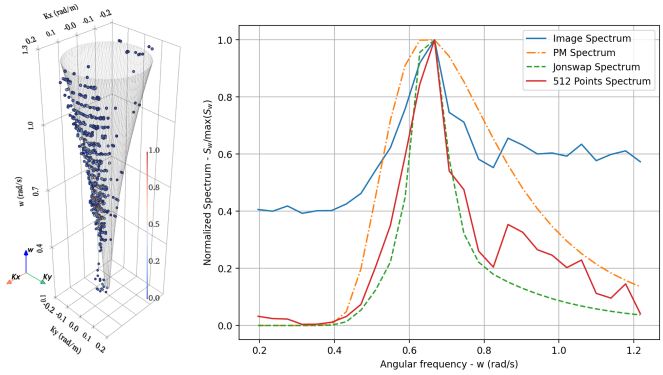
Yi Zhou 任职于中国大连大连海事大学电子信息工程系，邮编 116026。电子邮箱：(yi.zhou@dlmu.edu.cn)。Li Wang 任职于中国上海 SVA通信技术有限公司软件开发部，邮编 200233。电子邮箱：(wang li apple@163.com)。Hang Su 任职于中国北京清华大学计算机科学与技术系，邮编 100084。电子邮箱：(suhangss@mail.tsinghua.edu.cn)。Tian Wang 任职于中国北京航空航天大学人工智能学院，邮编 100083。电子邮箱：(wangtian@buaa.edu.cn)。

然而，根据雷达图像的灰度强度推导海浪波高颇具挑战。这不仅是一个非线性且病态的数学反演问题，还受到多种复杂因素的影响，例如雷达成像机制（倾斜调制、阴影效应、水动力调制）、环境条件（风速、降雨干扰）和硬件参数（天线高度、磁控管老化）。近年来，计算能力的提升以及人工智能的融合改变了这一领域的研究范式。研究方法已从早期的信噪比（SNR） [2], [3]转向未经校准的几何方法 [4], [5], ，如今则发展为数据驱动的深度学习方法 [6], [7]。

传统方法（信噪比和几何方法）都依赖专家进行特征工程。然而，雷达图像包含更丰富的信息，能够反映纹理和形状的演变，而这些方法无法充分利用这些信息。随着深度学习的发展，基于通用视觉模型的端到端反演方法——卷积神经网络（CNNs）[8]和视觉Transformer [9] ——已成为替代方案。这些方法通过在大型非海洋视觉数据集上预训练模型，并在小型海洋杂波图像或序列数据集上进行微调，来利用迁移学习。尽管这些方法具有优势，但也存在一些缺点：大量预训练数据主要由非海洋视觉图像构成，这会使网络参数偏离海洋特定特征，降低模型对与波浪相关视觉线索的敏感度，并增加在有限海洋数据上微调时过拟合的风险。此外，通用视觉模型的高计算需求和漫长的预训练周期，使其难以适应动态环境中的实时更新。

我们能否利用海杂波中波浪能量的先验知识，设计一个用于有效波高回归的高效网络？在 [10], [3],中，海杂波频谱中的阈值被用于区分波浪能量与背景噪声，结果显示两者之间的最大差异达两个数量级。受此观察启发，我们提出收集位于功率最大值百分之一范围内的前 N 个谱点，作为波浪能量的代表。

图1 显示，对于一个具有$H_s = 6.3$米、峰值角频率为 $\omega_p = 0.65$弧度/秒的海杂波序列，大多数排名前 512 的谱点都位于色散壳表面附近。在最大值归一化频谱图（图1 b）中，图像频谱表示对海杂波图像序列进行三维快速傅里叶变换（3D-FFT）后得到的振幅。从图像频谱中提取排名前 512 的谱点



(a) (b)

图 1. (a) 波数-频率域中排名前 512 的海杂波谱点及其底层色散壳。(b) 稀疏点的能量分布与经验 PM 和 JONSWAP 波谱的比较。

，再将它们在各频率区间内的功率累积，整合为一个频谱。这些点的能量分布与经验 Pierson–Moskowitz（PM）[11] 和 JONSWAP 谱 [12],吻合良好，表明稀疏的高能谱点与波谱之间存在很强的内在联系。

基于这一发现，我们提出一种谱点 Transformer（SPT）网络，利用高值谱点的能量大小和空间分布（波数-频率坐标）回归有效波高。通过利用波谱特性的先验知识，与传统卷积神经网络 [8] 或视觉 Transformer（ViT） [9], 相比，SPT 显著减少了输入数据量，提升了训练效率。

本研究的贡献总结如下：1) SPT 利用海杂波的稀疏谱点，通过深度学习回归有效波高 H_s，减少对经验调制的依赖，并提高回归精度。据我们所知，本研究提出了首个直接利用频谱中的稀疏点训练、用于反演有效波高的神经网络。值得注意的是，实验表明，SPT 学到的特征与物理色散关系一致。

2）实验结果表明，与最先进的视觉回归模型相比，SPT 的回归精度更高，同时所需的计算资源显著更少。例如，处理 1080 个海杂波序列（每个序列包含 64 帧、分辨率为 128×128 的图像）仅占用 1098 Mb 图形内存，并可在单个 RTX5090 GPU 上用 4 分钟完成 300 个训练轮次。这种轻量高效的回归网络在降低部署于前端雷达设备上的海浪反演系统成本方面具有巨大潜力。

II. 相关工作

X波段海洋雷达发射电磁波（波长约为 3 cm），主要通过布拉格共振与海面相互作用。雷达图像中观察到的明暗条纹并不直接对应波峰和波谷，而是由长重力波调制较短细小波纹所致 [13]。NietoBorge 及其同事建立了标准的三维谱分析流程，用于在波数-频率域中分离海浪信号与噪声，成为该领域的基石 [14], [2], [3]。尽管该方法已广泛应用于 WaMoS II 等商业系统中 [1], ，但其对外部校准数据的强依赖，以及在特定环境条件下失效，仍是关键局限。相比之下，我们提出的 SPT方法利用海杂波图像中高能谱点信息。通过构建基于点云回归的 Transformer网络，我们建立了稀疏谱点与 H_s 之间的数据驱动映射关系。这种方法降低了对经验数值公式的依赖，同时提高了有效波高估计的准确性。

传统的基于深度学习的方法通常引入先进的视觉分类模型，以根据海杂波图像估计有效波高。文献中，[15], [7], [16], [17], [6], 随着计算能力的不断提升，VGG[8], ResNet[18],ViT[9] 分别或联合用于有效波高估计。然而，将这些通用视觉模型直接应用于海杂波图像存在若干缺点：不仅浪费计算资源，还可能引入在非海洋数据集上训练所导致的误差。此外，大量模型参数限制了在雷达端点灵活更新权重的能力。相比之下，我们的 SPT 方法基于这样一项先验知识：具有强强度的稀疏谱点与波浪的能量指标存在内在关联 H_s。基于这一原理构建用于回归的 Transformer 网络，可以有效减少图像序列中的冗余信息，并减轻计算负担。因此，即使采用轻量级 Transformer 架构，我们的方法也能取得优于通用视觉网络的性能。

PointNet [19] 和 PointNet++[20]奠定了从点云中提取全局和层次化特征的基础范式。PCT（点云变换器）[21] 引入 Transformer，反映了近期转向采用注意力机制以实现更优全局-局部上下文建模的趋势。PCT方法将法向参数、局部曲率等几何特征以及三维坐标作为输入特征。尽管这些点是离散分布的，但它们呈现出连续的几何相关性。因此，在为PCT设计嵌入模块时，通常使用卷积来关联邻近点并扩展特征维度。鉴于样本的坐标值和特征幅值在相同数值范围内保持一致，归一化时通常采用批归一化（BatchNorm）。对于基于谱点的网络，我们将每个点的谱能量 p 和空间位置都作为输入。关键挑战在于处理谱点之间的空间独立性，以及它们能量级的显著差异。为稀疏的高能量谱点设计SPT网络时，我们采用多层感知机（MLP）嵌入和精细的归一化，以提升回归性能。

III. 方法

本节介绍所提出方法的方法，重点阐述预处理步骤和SPT网络架构。图2 展示了系统框架概览。

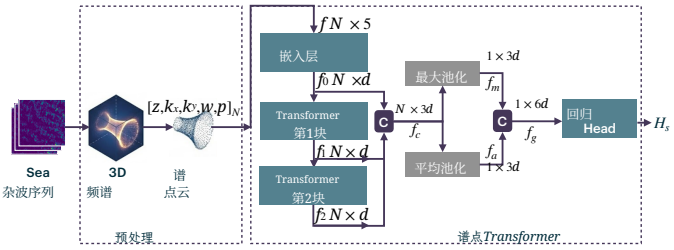


图 2. 谱点 Transformer（SPT）的框架。

A. 预处理

输入由在不同海况下采集的海杂波图像序列组成。为了提取与有效波高估计最相关的特征，我们执行以下预处理步骤，以获得频谱点云。

首先，对原始海杂波图像序列进行三维快速傅里叶变换。其次，在得到的功率频谱中，选取振幅最强的前N个谱点。第三，将相应的波数-频率坐标（k_x， k_y， ω）、振幅值 p以及杂波区域索引 z作为点云表示的初始特征，其中杂波区域索引的取值范围为 $\{1,2,3\}$:

$f = [z, k_x, k_y, \omega, p]_N$。 f 的维度为 $N \times 5$。

这种设计利用了谱能量集中在色散曲面上的特点。由于分辨率限制，色散曲面由离散的谱点构成（如图 1所示）。根据波浪诱导能量约比背景噪声高两个数量级这一观察[10], 我们从功率最大值与其百分之一之间提取 N 个谱点，以表征波浪能量。预处理后，原始图像序列频谱数据仅保留约 0.05%，压缩比约为 2048 倍。这显著降低了后续模型训练的计算需求。

B. SPT网络架构

SPT网络由三个主要组件构成：嵌入模块、堆叠的 Transformer 块和回归头。每个组件都针对频谱点云在不同条件下面临的特定挑战。

1) 嵌入模块：在输入特征 f 中，杂波区域索引 z、波数坐标（k_x， k_y）和角频率 ω 并非以相等的几何间隔变化。标准 Transformer [22]中使用的传统正弦/余弦位置嵌入不适用于处理这种不等间距。因此，谱点的位置信息和振幅作为初始词元特征同等重要。通过省略显式的位置嵌入，SPT 得以自主学习并着重关注隐层特征表示中具有判别性的谱坐标。

所选谱点按强度排序，因此在频谱空间中是不连续的。使用一维卷积嵌入（例如， [21]）会整合错误的邻域，并通过虚假的邻接关系削弱孤立的强点。因此，我们采用两层多层感知机，将词元特征嵌入到更高维度。多层感知机的第一层表示如下:

$$\mathrm{Layer}(f) = \mathrm{SiLU}\left(\mathrm{LayerNorm}\left(\mathrm{Linear}(f)\right)\right). \tag{1}$$

其中，SiLU 表示 Sigmoid 加权线性单元（SiLU）[23],; 层归一化针对每个样本独立地计算特征维度上的统计量，表示为 [22] ； 线性层表示用于改变特征维度的矩阵乘法。设 f， f_0 分别表示嵌入模块的输入和输出特征，则整体表达式可写为:

$$f_0 = \mathrm{Layer}_2\left(\mathrm{Layer}_1(f)\right). \tag{2}$$

其中， $f_0 \in R^{N\times d}$、 d 表示嵌入维度。内部层1 首先将维度 f 从 5 增加到$d/2$，然后外部层2 进一步将中间维度增加到 d。

2) Transformer 块：每个 Transformer 块由两个子层组成：自注意力模块和一个包含两层多层感知机的前馈网络（FFN）。针对不同海况下各谱点功率动态范围较大的挑战（高海况与低海况之间的振幅差异可达几个数量级），本文将查询-键归一化（QK 归一化） [24]引入块设计，以提升模型性能。

自注意力机制是 Transformer 架构的核心。它通过根据相关性为序列中的其他词元分配不同权重的注意力分数，使模型能够与这些词元建立联系。令$\mathbf{X} \in \mathbb{R}^{N\times d}$ 表示输入矩阵，其中 N 为序列长度， d 为嵌入维度。它将输入投影到查询、键和值空间。

定义投影矩阵$\mathbf{W}_Q$， $\mathbf{W}_K$， $\mathbf{W}_V \in \mathbb{R}^{d\times d}$ 如下:

$$\mathbf{Q} = \mathbf{X}\mathbf{W}_Q, \quad \mathbf{K} = \mathbf{X}\mathbf{W}_K, \quad \mathbf{V} = \mathbf{X}\mathbf{W}_V, \tag{3}$$

其中$\mathbf{Q}$， $\mathbf{K}$， $\mathbf{V} \in \mathbb{R}^{N\times d}$。为提高计算效率，这三项操作通常合并为一次矩阵乘法:

$$\mathrm{Proj}(\mathbf{X}) = [\mathbf{Q}, \mathbf{K}, \mathbf{V}] = \mathbf{X}\cdot[\mathbf{W}_Q, \mathbf{W}_K, \mathbf{W}_V]_{N\times 3d}. \tag{4}$$

给定查询、键和值矩阵，标准的缩放点积注意力定义如下:

$$\mathrm{Sa}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \mathrm{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d}}\right)\mathbf{V}. \tag{5}$$

当$\mathbf{Q}$ 与$\mathbf{K}$的点积过大时，softmax 操作会产生高度尖锐的概率分布（趋近于独热分布），导致梯度消失和训练不稳定。为减轻高能量

特征对反向传播期间梯度稳定性的影响，我们采用 QK 归一化方案 [24] 。具体而言，在计算$\mathbf{Q}$ 与$\mathbf{K}$的内积之前，我们对 [25] 每个行向量$\mathbf{q}_i \in \mathbb{R}^d$（属于$\mathbf{Q}$）进行均方根归一化（RMSNorm）：

$$\hat{\mathbf{q}}_i = \frac{\mathbf{q}_i}{\sqrt{\frac{1}{d}\sum_{j=1}^{d} q_{ij}^2}}. \tag{6}$$

ˆ ˆ对所有 N 行进行归一化后，我们得到$\mathbf{Q}$。同样地，计算得到$\mathbf{K}$ˆ和$\mathbf{V}$ 。记作：

$$[\hat{\mathbf{Q}}, \hat{\mathbf{K}}, \hat{\mathbf{V}}] = \mathrm{RMSNorm}([\mathbf{Q}, \mathbf{K}, \mathbf{Q}]). \tag{7}$$

代入（5）可得注意力模块的输出：

$$\mathrm{Attn}(\mathbf{X}) = \mathrm{Sa}\left(\mathrm{RMSNorm}\left(\mathrm{Proj}\left(\mathbf{X}\right)\right)\right). \tag{8}$$

值得注意的是，Transformer 模型采用多头注意力，将特征划分为多个子空间，以便在每个子空间内执行并行注意力运算。所有头的输出会进行拼接，以增强特征多样性的学习能力。

最后，对注意力和多层感知机（MLP）的输入均应用前置归一化（层归一化），并采用残差连接 [26]。在此，两层多层感知机（MLP）首先将 token 的特征维度增加到 $4d$，然后按照传统方式投影回原始 d，即 [22]。对于输入特征 $f_{i-1} \in \mathbb{R}^{N\times d}$ 到第 i 个 Transformer 块，其输出特征f_i 计算如下：

$$\begin{cases} f_i' &= f_{i-1} + \mathrm{Attn}\left(\mathrm{LayerNorm}(f_{i-1})\right), \\ f_i &= f_i' + \mathrm{MLP}\left(\mathrm{LayerNorm}(f_i')\right). \end{cases} \tag{9}$$

如图 2所示，两个 Transformer 块依次堆叠，其输出分别记为 f_1 和 f_2。将它们与嵌入模块的输出 f_0 拼接，形成中间变量 $f_c \in \mathbb{R}^{N\times 3d}$。最大池化和平均池化运算分别计算每个特征维度上 N个数据点的最大值和均值，得到大小为 f_m和 f_a的特征 $1\times 3d$。将它们拼接后得到最终的全局特征 $f_g \in \mathbb{R}^{1\times 6d}$。为提高模型容量，可以堆叠更多块。假设使用 M 个块，拼接后的特征 f_c 维度为 $\mathbb{R}^{N\times(M+1)d}$，并得到 $f_g \in \mathbb{R}^{1\times 2(M+1)d}$。

3）回归头：回归头采用与嵌入模块类似的两层多层感知机。第一个线性层将全局特征 f_g 的维度减半。经过层归一化和 SiLU 激活函数后，第二个线性层将ˆ半维映射为一个标量，作为 H_s的估计值。其表达式如下：

$$\hat{H}_s = \mathrm{Linear}\left(\mathrm{SiLU}\left(\mathrm{LayerNorm}\left(\mathrm{Linear.}(f_g)\right)\right)\right) \tag{10}$$

网络使用 Huber 损失进行训练[27]:

$$L(x) = \begin{cases} \frac{1}{2}x^2, & |x| < 1, \\ |x| - \frac{1}{2}, & \text{otherwise}, \end{cases} \tag{11}$$

ˆ其中，$x = H_s - H_s$ 表示真实高度与估计高度之间的误差。这里，Huber 损失对较大误差采用线性惩罚，以取代均方误差（MSE）准则
异常值，同时保留二次惩罚对小误差进行平滑优化的优势。

IV. 实验

A. 实验设置与雷达参数



图 3. 从雷达位置观察到的现场实验环境。白色虚线圆圈标示测波浮标，矩形表示三个观测海域（z_1，z_2，z_3），右侧显示了相应的原始海杂波图像。

本实验所使用的有效波高数据采集自东海嵊山馒头山岛附近的一台测波浮标（图 3）。 H_s 更新间隔为 30 分钟，测量精度为$\pm(0.1 + H_s \times 10\%)$ 米，当地水深约为 39 米。用于观测的海洋雷达配备了一根 2 米长的天线，安装在距海平面约 50 米高的山顶上，以每转 2.5 秒的速度旋转并采集回波。该雷达的波束宽度为 1 度，距离分辨率为 7.5 米。

海洋雷达在浮标周围的三个方向——北、东和南——捕获了平面坐标图像（128x128），并将其保存为图像序列，每个序列最多包含64帧。这些图像的观测距离范围约为距雷达位置1.5 km至2.5 km。数据采集于2021年7月至10月进行，最终得到一个包含1620个图像序列的数据集。在此期间，实验还遇到了极端天气条件，风速超过20 m/s。

该海杂波数据集的有效波高范围为0.5至6.3米。其中，80%的波高低于2.5 m，2%高于4 m。由于在1、2和3区同时采集了三组海杂波，因此在每个记录时刻，按2:1的比例从这三个区域随机选取训练数据和测试数据（得到1080个训练样本和540个测试样本）。这种抽样方案确保训练集和测试集中的波高分布相同。

B. 模型比较

本研究比较了基于视觉变换器（ViT）的 SWHFormer [6]，架构 [9] ，该架构从单幅海杂波图像回归估计有效波高。由于采集图像的分辨率为 128 ×128，而 SWHFormer 是在 ViT 预训练模型的基础上，针对 384 ×384距离-方位图像进行微调的，因此为公平起见，测试了两个版本：SWHFormerI

和 SWHFormerII。SWHFormerI 采用与 [6], 相同的参数设置和预训练权重，并要求先通过预处理将输入图像上采样至 384 ×384 。相比之下，SWHFormerII 虽然采用相同的架构，但使用 64 ×128 ×128 谱幅值数据从头开始训练，并将数据输入 ViT 模型。与使用全部谱点的 SWHFormerII 进行比较，可以评估基于稀疏频率点的 SPT 是否能够与使用完整频谱数据的一般视觉模型相抗衡。此外，本研究还比较了 3D-VGG [7], ，这是一种用于从图像序列回归有效波高的三维卷积网络。3D-VGG 使用经典的 VGG 卷积神经网络模块 [8]学习时空卷积核，以捕捉时频信息与波高之间的关系，从而检验不同架构的有效性。对于稀疏谱点数据，我们比较了朴素点云Transformer（NPT） [22] 和 SPCT（简单点云Transformer） [21], 以验证选择点云特征并采用适配模块的必要性。在此，NPT 将每个谱点的功率 p 作为一个标记，并将区域向量和坐标向量（ $[z, k\ \ x, k\ \ y, w]$）分别嵌入额外的位置嵌入模块中，同时使用类别标记学习用于 H_s 回归的代表性特征。SPCT 使用与 SPT 相同的输入特征，但采用卷积嵌入、批归一化和偏移注意力模块。

用于测试不同模型性能的指标包括参数数量、每个样本的浮点运算次数（FLOPS）、估计值与真实波高之间的皮尔逊相关系数（CC），以及低海况（H_s，范围为 0.5–2.5 米）、中海况（2.5–4.0 米）、高海况（ >4 米）和总体估计的均方根误差（RMSE） H_s 。

在表I中，SPT 在多项关键指标上表现出色，包括浮点运算次数、相关系数和总体均方根误差。结果表明，

- SPT 在模型复杂度与性能之间实现了最佳平衡，参数数量（0.70 M）和浮点运算次数（1.04 G）均最少。
- 尽管 3D-VGG 的参数数量少于 SPT，但由于采用了三维卷积，其计算成本显著更高。
- SPT 的表现优于使用全部频谱数据的 SWHFormerII，这表明稀疏谱点足以有效估计有效波高。
- SPCT 和 SPT 的性能优于 NPT。这表明，应将位置信息视为初始特征，以增强频谱中能量分布的判别能力。
- 与 SPCT 相比，SPT 的 MLP嵌入以及采用 QK 归一化的自注意力机制更适合频谱中的点云结构。

实验结果验证了我们所提出的方法能够从稀疏谱数据中捕获关键特征，同时保持计算效率。

C. SPT 特性分析

本节讨论 SPT 模型的两个关键特征：训练效率，以及所学习特征与色散关系的对齐。

(1)**训练效率**:

表II分别基于图像序列、完整频谱和稀疏谱点，对 3D-VGG、SWHFormerII 和 SPT 的每批次计算性能进行了比较。评估在配备 32 GB 图形内存的 RTX 5090 GPU 上进行。

由于 1620 序列数据集规模庞大（约 26 GB），必须以批次为单位将序列图像或完整三维频谱加载到主机内存和 GPU 显存中。在这三种方法中，3D-VGG 的计算负载最高，其批次大小受 GPU 显存容量限制。本研究中，3D-VGG 的批次大小设为 32，此时 GPU 显存利用率几乎达到饱和。为公平比较，另外两种方法的批次大小也设为 32。

3D-VGG 和 SWHFormerII 都需要分批从硬盘加载海杂波数据，将数据从硬盘传输到主机内存需要 190 毫秒以上。尽管三维卷积的计算成本约为 SWHFormerII 所用轻量级视觉变换器的四倍，但其总体回归精度并未超过 SWHFormerII。

SPT 使用稀疏频率点数据，显著缩短了从硬盘加载数据所需的时间。在预处理过程中，SPT 仅选取最强的 0.05% 谱点用于有效波高估计，这是其计算效率的关键。由于所有稀疏数据都可以一次性加载到 GPU 显存中，因此批次处理期间硬盘、主机内存和 GPU 显存之间的通信开销降至最低，从而提高了 GPU 利用率，并节省了大量训练时间。对于 SPT，主要的数据传输发生在 GPU 显存缓存与 CUDA核心之间，耗时仅约 5 毫秒。

SPT 的单批次训练时间显著更短，分别比 3D-VGG 和 SWHFormerII 快约 44× 和 13×。在 GPU 利用效率方面，3D-VGG 由于计算与数据加载并行进行，能够保持较高的 GPU 活跃度；然而，其内存带宽使用率达到 98%，接近上限。对于 SWHFormerII，数据加载时间远超前向传播和反向传播所需时间，导致 GPU 长时间处于空闲状态。其 6% 的 GPU 利用率表明，通过改进数据预处理和缓存策略仍有加速空间。对于 SPT，GPU 显存内的数据传输时间与前向传播时间相当，因此总体 GPU 计算效率达到 70%。

当训练轮次总数设为 300、训练数据包含 34 个批次时，SPT 可在 4 分钟内完成训练。其内存带宽使用率仅为 3%，这表明 SPT 算法能够扩展到规模大十倍的数据集，同时仍可在单个 GPU 上完成训练。结果表明，SPT 以极低的内存开销实现了出色的计算效率，因此适用于实时应用。

(2)**物理色散对齐**:

图4 展示了 SPT 学习到的特征与物理色散关系的对齐情况。左列显示

表 I 不同模型的性能比较

模型	输入特征	参数 (M)↓	浮点运算次数 (G)↓	CC ↑	均方根误差(m)↓ 总体	0.5–2.5 米	2.5–4.0 米	> 4 米
SWHFormerI[6]	单幅图像	86.4	49.7	0.85	0.51	0.47	0.69	1.63
SWHFormerII	三维频谱	47.9	3.40	**0.97**	0.24	0.22	**0.31**	0.49
3D-VGG[7]	图像序列	**0.43**	94.6	**0.97**	0.25	0.19	0.33	0.73
NPT[22]	谱点	0.70	1.04	0.71	0.79	0.76	0.61	1.75
SPCT[21]	谱点	0.64	1.54	0.96	0.25	0.20	0.36	0.66
SPT（本文方法）	谱点	0.70	**1.04**	**0.97**	**0.23**	**0.17**	0.34	**0.48**

参数、浮点运算次数和均方根误差的值越低越好，而相关系数的值越高越好。粗体数字表示每列中的最佳性能。

表 II 一个批次的图形处理器训练配置

模型	加载至内存（毫秒）↓	加载至图形处理器（毫秒）↓	前向传播（毫秒）↓	反向传播（毫秒）↓	耗时 (毫秒)↓	GPU 利用率 ↑	内存利用率↓
3D-VGG	195	18.2	138	488	839	**74%**	98%
SWHFormerII	198	19.5	5.81	23.7	247	6%	8%
SPT	**0**	**5.01**	**4.18**	**9.62**	**18.8**	70%	**3%**

不同有效波高（H_s）下波数-频率谱的 w-k_y 平面。实线表示基频和谐波色散关系，虚线表示频率混叠效应。右列显示全局特征 f_a 与每个选定谱点的局部特征之间的余弦相似度。相似度得分越高，对回归的影响越大。

图 4中，高海况（H_s=6.3 米）沿色散曲线呈现能量集中，这一点也反映在 SPT 学到的特征中。相似度较高的点也沿色散曲线密集聚集。随着有效波高降低、波周期缩短，谱干扰变得更加明显。在最终判别中起决定作用的谱点变得更加稀疏，但其高相似度位置仍与色散曲线保持一致。以最后一个子图（有效波高为 2.1 米）为例，特征相似度分数较高的点位于角频率 $w = 0.8$ 和 0.5 弧度/秒处，分别对应 7.9 秒和 12.6 秒的波周期。7.9 秒的周期与 7.8 秒的浮标观测值十分接近，而较长的 12.6 秒周期则表明，SPT 学会利用较长周期的波浪纹理，进一步区分低海况数据之间的细微差异。

D. 消融研究

为确定最佳配置，我们从四个关键方面开展了消融研究：输入特征、注意力机制、查询-键归一化和模型可扩展性。数值结果见表III。

主要发现包括：

- **输入特征**：纳入区域标识符（z）和坐标（k_x，k_y，w），利用空间上下文和波数坐标提升模型性能。根据利用海杂波反演波高的经验，最大功率谱点的位置 w_p 以及波数空间中其他高功率点的分布模式与波浪能量相关。此外，不同方向采集的雷达回波表现出不同的功率水平，导致同一时刻不同区域的海杂波能量存在差异（这一现象也曾在 [28] 中讨论）。将这两项先验知识纳入特征表示，有助于模型取得更好的训练效果。

- **注意力机制**：与 [21] 中使用的偏移注意力相比，残差注意力表现更优，因为它能更好地处理稀疏谱数据，并提升数值稳定性。这主要有两个原因：(1) PCT中的偏移注意力采用一维卷积进行维度扩展。卷积能够有效提取具有连续几何分布的稠密点云局部邻域中的判别性特征（例如飞机或汽车的点云），但往往会抑制稀疏谱点数据中独立高值点的特征。(2) PCT使用批次归一化（BatchNorm）来保证数值稳定性。由于该数据集中台风期间采集的高海况样本数量有限，每个训练批次中的高海况数据都远少于中、低海况样本。Batch Norm中的平均操作会严重削弱高海况特征在波高估计中的判别能力。因此，SPT采用基于多层感知机的维度扩展、标准自注意力和层归一化，从而提升了模型的总体性能。

- **QK归一化**：应用均方根归一化可稳定训练，尤其是在样本量有限的高海况下。考虑到功率 p

表 III SPT 超参数的消融研究结果。

超参数 参数	设置 选项	CC ↑	均方根误差（米）↓ 总体	0.5–2.5 米	2.5–4.0 米	> 4 米
输入特征	$[w, p]$	0.93	0.34	0.26	0.45	1.14
输入特征	$[k_x, k_y, w]$	0.96	0.24	0.18	0.39	0.80
输入特征	$[z, k_x, k_y, w, p]$	**0.97**	0.23	0.17	0.34	0.48
注意力类型	偏移量	0.96	0.24	0.17	0.36	0.78
注意力类型	残差	**0.97**	0.23	0.17	0.34	0.48
QK 归一化	No	0.96	0.24	0.18	0.34	0.70
QK 归一化	Yes	**0.97**	0.23	0.17	0.34	0.48
[N,注意力头 d, 块]	[256, 4, 64, 1]	0.95	0.27	0.19	0.36	1.04
[N,注意力头 d, 块]	[512, 8, 128, 2]	0.97	0.23	0.17	0.34	0.48
[N,注意力头 d, 块]	[1024, 16, 256, 4]	**0.97**	**0.22**	**0.17**	**0.31**	**0.43**

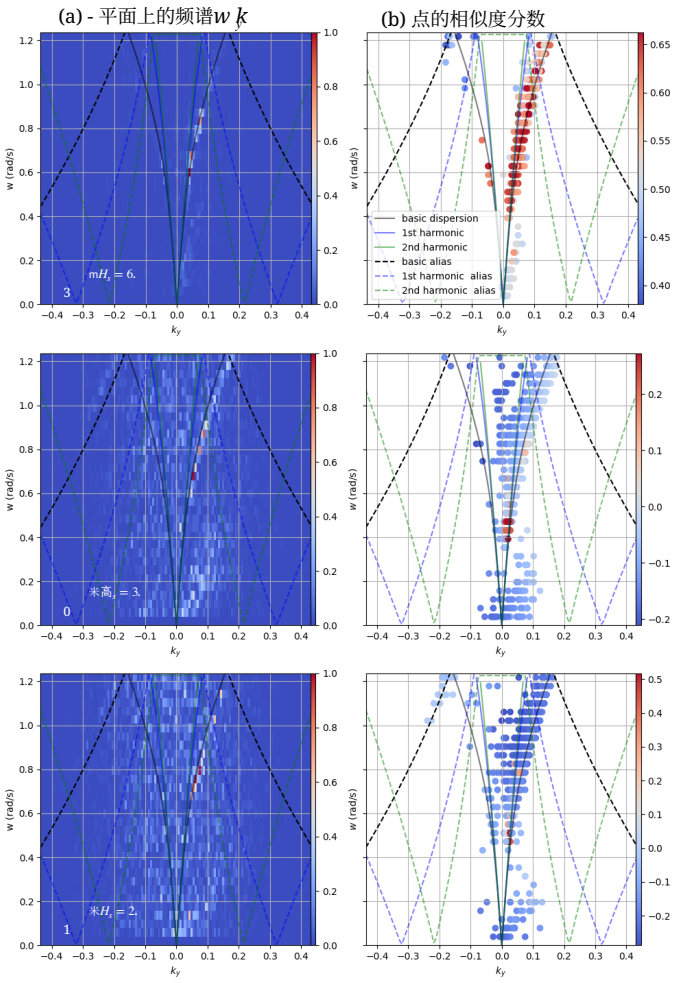


图 4. SPT 特征与物理色散关系的对齐。左列展示了不同 H_s 值对应的 w-k_y 平面。实线表示色散关系，虚线表示混叠效应。右列突出显示特征相似度分数，强度越高表示对回归的影响越大。

同一频谱中的点可相差最多两个数量级，而不同海况之间的差异更大，因此在计算查询和键的内积之前引入均方根归一化

注意力机制有助于稳定数值范围，防止反向传播梯度出现剧烈波动，从而使模型更好地收敛。

- **模型缩放**：提升模型容量（增大 N、注意力头数量和嵌入维度 d 以及 Transformer 块数量）可改善性能，这与 Transformer 的缩放定律一致 [29]。

这些见解有助于选择最优配置，从而有效估计波高。

V. 结论

本研究为基于稀疏谱数据估计 H_s 提供了一个高效框架，并深入探讨了 Transformer 模型在学习谱判别方面的应用。所提出的方法不仅与物理色散关系高度吻合，也展现出在实时波浪监测系统中的实用价值。然而，本研究仍存在局限。当前数据集规模较小、多样性不足，尤其是近岸岛屿站点的数据有限。未来的研究方向包括收集更多样化的训练数据，以进一步探索该方法在不同环境条件下的适用性，并将模型扩展到其他海洋学参数，例如流速和流向。

VI. 致谢

作者感谢何志伟（SVA通信技术有限公司）为海杂波数据集的准备以及实验期间的技术讨论所提供的帮助。

参考文献

[1] K. G. Hessner、J. C. Nieto-Borge 和 P. S. Bell，欧洲航海雷达测量：*WaMoS II* 作为海况、海流和水深测量传感器的应用。多德雷赫特：施普林格·荷兰，2008，435 446 1, 2 第 – 页。J. C. Nieto-Borge、G. R. Rodr´ıguez、K. Hessner 和 P. I. Gonz´alez，“用于表面波分析的海洋雷达图像反演”，《大气与海洋技术杂志》，第 21 卷，2004 年，第 8 期，第 1291–1300 页。[3] [2] 1, 2 J. C. Nieto-Borge 、 K. Hessner 、 P. Jarabo-Amores 和 D. de la Mata-Moya，“通过信噪比分析从 X 波段海洋雷达图像时间序列估计海浪高度”，《*IET*雷达、声呐与导航》，第 2 卷，第 1 期，第 35–41 页，2008 年。1，2

[4] R. Gangeskar，“一种利用 X 波段雷达海面图像中的遮蔽效应估计波高的算法”，IEEE 地球科学与遥感汇刊，第 52 卷，第 6 期，第 3373–3381 页，2014年。1[5] W. Navarro、M. Streßer、R. Carrasco、J. Seemann 和 J. Horstmann，“ 一种无需校准、利用 X 波段海洋雷达图像时间序列估计有效波高的方法”，IEEE 国际地球科学与遥感研讨会，第

6907 6910, 2022 1 – 页。

[6] Z. Yang 和 W. Huang，“Swhformer：一种用于从航海雷达图像估计有效波高的视觉变换器”，IEEE 地球科学与遥感汇刊，第 62 卷，第 1–13 页，2024 年。1、2、4、5、6[7] J.-W. Kwon、W.-D. Chang 和 Y.-J. Yang，“利用三维卷积深度学习从 X 波段海洋雷达图像预测有效波高”，PLoS ONE，第 18 卷，第 10 期，文章编号 e0292884，2023 年。1、2、[8] K. Simonyan 和 A. Zisserman，“用于大规模图像识别的超深卷积网络”，载于国际学习表征会议（ICLR），

2015 1, 2, 5 。A.

[9] Dosovitskiy、L. Beyer、A. Kolesnikov、D. Weissenborn、X. Zhai、T. Unterthiner、M. Dehghani、M. Minderer、G. Heigold、S. Gelly、J. Uszkoreit 和 N. Houlsby，“一张图像胜过 16×16 个词：用于大规模图像识别的 Transformer”，载于国际学习表征会议，

2021 1, 2, 4 。C. Senet、J. Seemann 和 F.

[10] Ziemer，“根据海面图像序列确定近表层流速”，IEEE 地球科学与遥感汇刊，第 39 卷，第 3 期，第 492–505 页，2001 年。1、3[11] W. J. Pierson 和 L. Moskowitz，“基于 S. A. Kitaigorodskii 相似理论提出的充分发展风浪谱形式”，《地球物理研究杂志》，第 69 卷，1964 年，第 24 期，第 5181–5190 页。[12] K. Hasselmann、T. Barnett、E. Bouws、H. Carlson、D. Cartwright、K. Enke、J. Ewing、H. Gienapp、D. Hasselmann、P. Kruseman 等，“北海联合波浪项目（JONSWAP）期间的风浪增长和涌浪衰减测量”，Erg¨anzungsheft zur Deutschen Hy-drographischen Zeitschrift，A辑，第

12, 1 95, 1973 2 卷，第 – 页。

[13] K. D. Ward、R. J. A. Tough 和 S. Watts，《海杂波：散射、K分布与雷达性能》。工程技术学会，

2006 2 。I. R. Young、W.

[14] Rosenthal 和 F. Ziemer，“用于确定海浪方向性和表层洋流的海洋雷达图像三维分析”，《地球物理研究杂志：海洋》，第

90, 1, 1049 1059, 1985 2 卷，C期，第 – 页。

[15] W. Duan、K. Yang、L. Huang 和 X. Ma，“利用深度卷积神经网络对合成X波段雷达海杂波图像进行波浪遥感的数值研究”，《遥感》，第 12 卷，第 7 期，2020 年。2[16]S. Zuo、D. Wang、X. Wang、L. Suo、S. Liu、Y. Zhao 和 D. Liu，“利用卷积神经网络从 X 波段海洋雷达图像估计波高和周期”，《海洋科学与工程杂志》，第 12 卷，第 2 期，2024 年。 [在线]。可访问：https://www.mdpi.com/ 2077 1312 12 2 311 2 - / /

[17] / V. Rezvov 、 M. Krinitskiy 、 A. Gavrikov 、 V. Golikov 、 M. Borisov 、 A. Suslov 和 N. Tilinina，“通过在合成X波段雷达海杂波图像上进行预训练，改进有效波高的数据驱动估计”，《海洋科学前沿》，第 11 卷，文章编号 1363135，2024 年，第 2 页。[18] K. He、X. Zhang、S. Ren 和 J. Sun，“用于图像识别的深度残差学习”，载于 IEEE计算机视觉与模式识别会议，

2016, 770 778 2 第 – 页。R. Q. Charles、H. Su、M.

[19] Kaichun 和 L. J. Guibas，“PointNet：用于三维分类和分割的点集深度学习”，载于 2017 年 IEEE计算机视觉与模式识别会议（CVPR），

2017, 77 85 2 第 – 页。C. R. Qi、L. Yi、H. Su 和 L. J. Guibas，“

[20] PointNet++：度量空间中点集的深度层次特征学习”，载于国际神经信息处理系统会议论文集，

2017, 5105 5114 2 第 – 页。M.-H. Guo、J.-X.

[21] Cai、Z.-N. Liu、T.-J. Mu 和 S.-M. Hu，“PCT：点云变换器”，《计算视觉媒体》，第 7 卷，第 2 期，第 187–199 页，2021 年。2、3、5、6[22] A. Vaswani、N. Shazeer、N. Parmar、J. Uszkoreit、L. Jones、A. N. Gomez、Ł. Kaiser 和 I. Polosukhin，“注意力就是你所需的一切”，载于神经信息处理系统进展（NeurIPS），2017 年，第 5998–6008 页。3

, 4, 5, 6

[23] S. Elfwing、E. Uchibe 和 K. Doya，“用于强化学习中神经网络函数逼近的 Sigmoid 加权线性单元”，*NeuralNetworks*，第

107, 3 11, 2018 3 页 – 。A. Henry、P. R. Dachapally、S.

[24] Pawar 和 Y. Chen，“面向 Transformer 的查询-键归一化”，载于计算语言学协会年会论文集：*EMNLP* 研究成果

2020, 2020, 4246 4253 3, 4 页 – 。B. Zhang 和

[25] R. Sennrich，“ 均方根层归一化 ”，载于神经信息处理系统进展 *(NeurIPS)*，2019 年，页

12 381 12 392 4 – 。K. He、X. Zhang、S. Ren 和 J. Sun，“用于图像识别的深度

[26] 残差学习”，载于*IEEE*计算机视觉与模式识别会议论文集，

2016, 770 778 4 页 – 。P. J. Huber，“位置参数的稳健估

[27] 计”，《数理统计年刊》，第 35, 1, 73 101, 1964 4

[28] 卷，第。B. Lund、C. O. Collins、H. C. Graber、E. Terrill 和 T. H. C. Herbers，“海洋雷达海浪反演对距离和方位角的依赖性”，*Ocean Dynamics*，第

64, 7, 999 1018, 2014 6 卷，第。J. Kaplan、S.

[29] McCandlish、T. Henighan、T. B. Brown、B. Chess、R. Child、S. Gray、A. Radford、J. Wu 和 D. Amodei，“神经语言模型的缩放定律”，*arXiv* 预印本 *arXiv:2001.08361*，2020年。[在线]。可获取于：https://arxiv.org/abs/2001.08361 7

地点
照片
HERE

Yi Zhou 于2003年和2006年分别获得中国大连海事大学电子工程学士学位和理学硕士学位。2012年，他获得中国上海交通大学与法国特鲁瓦技术大学联合授予的信号处理联合博士学位。自2013年起，他一直担任大连海事大学电子信息工程系讲师。2021年冬季，他在清华大学TSAIL团队担任访问研究员，为期三个月。他的研究兴趣包括海洋雷达信号处理和计算机视觉。

请在此处
放置照片

Li Wang 于1984年出生于中国四川宜宾。2019年，他获得中国清华大学博士学位。目前，他是SVA通信技术有限公司的高级工程师。他的研究兴趣主要包括自适应雷达信号处理和雷达波形设计。

地点
照片
HERE

Hang Su （IEEE会员）现任中国北京清华大学计算机科学与技术系副教授。他于2014年获得上海交通大学博士学位，并于2011年至2013年在卡内基梅隆大学担任访问学者。他的研究兴趣包括对抗机器学习和鲁棒计算机视觉。他曾担任NeurIPS领域主席和AAAI 2022研讨会联合主席。他曾获MICCAI 2012颁发的“青年研究者奖”，以及“最佳论文AVSS 2012 的“最佳论文奖”，以及 ICME 2018 的“白金最佳论文奖”。



Tian Wang 于2010年获得中国西安交通大学理学硕士学位，并于2014年获得法国特鲁瓦技术大学博士学位。他现任北京航空航天大学人工智能学院教授。其研究兴趣包括计算机视觉和模式识别。