

MST++: 用于高效光谱重建的多阶段光谱维度Transformer

摘要

目前，光谱重建（SR）领域的主要方法集中于设计更深或更宽的卷积神经网络（CNNs），以实现从RGB图像到其高光谱图像（HSI）的端到端映射。虽然这些基于CNN的方法在恢复性能方面表现出色，但在捕捉长程依赖性和自相似性先验方面存在一定限制。为解决这一问题，本文提出了一种新颖的基于Transformer的方法，即多阶段光谱维度Transformer（MST++），以实现高效的光谱重建。具体而言，采用了基于HSI的光谱维度多头自注意力（S-MSA），该方法在空间上稀疏而在光谱上具有自相似性，构建了基本单元，即光谱维度自注意力块（SAB）。然后，通过SAB构建了单阶段光谱维度Transformer（SST），利用U型结构提取多分辨率的上下文信息。最终，MST++由多个SST级联，逐渐提高重建质量，从粗到细。综合实验证明，相较于其他CNN算法，MST++表现出更好的效果。

关键词：光谱重建；深度学习；Transformer

1 引言

高光谱成像记录了真实世界场景在窄波段中的光谱，其中每个波段捕捉特定光谱波长的信息。与普通的RGB图像相比，HSI具有更多的光谱波段，可以存储更丰富的信息并勾勒出所捕获场景的更多细节。由于这个优势，HSI具有广泛的应用，如医学图像处理 [4]，遥感 [14]，目标跟踪 [12]等。尽管如此，拥有丰富光谱信息的HSI是耗时的，需要使用光谱仪沿着空间或光谱维度扫描场景。这一限制阻碍了HSI的应用范围，特别是在动态或实时场景中。解决这个问题的一种方法是开发快照压缩成像（SCI）系统和计算重建算法 [19]，从2D测量到3D HSI立方体。然而，这些方法依赖于昂贵的硬件设备。为了降低成本，提出了光谱重建（SR）算法，用于从给定的RGB图像重建HSI，这可以通过RGB相机轻松获取。传统的SR方法主要基于稀疏编码或相对较浅的学习模型。尽管如此，这些基于模型的方法在表示能力和泛化能力方面存在局限性。

近年来，随着深度学习的发展，SR取得了显著的进展。深度卷积神经网络（CNNs）已应用于学习从RGB图像到HSI立方体的端到端映射函数。尽管取得了令人印象深刻的性能，但这些基于CNN的方法在捕捉长程依赖性和光谱间的自相似性方面存在局限性。近年来，自然语言处理（NLP）模型Transformer [18]已应用于计算机视觉并取得了巨大成功。Transformer中的多头自注意力（MSA）机制在建模长程依赖性和非局部自相似性方面优于CNN，可以缓解基于CNN的SR算法的限制。然而，直接使用标准Transformer进行SR会遇到两个主要问题。一

方面，标准全局MSA的计算复杂度与空间维度的平方成正比，这是一个可能负担不起的巨大负担。另一方面，基于局部窗口的MSA在位置特定窗口内受限的接受域内受限。

为了解决上述问题，提出了第一个基于Transformer的框架，即用于高效光谱重建的多阶段光谱维度Transformer（MST++），用于从RGB图像高效地进行光谱重建。可以注意到HSI信号在空间上是稀疏的，而在光谱上是自相似的。基于这一特性，采用了光谱维度多头自注意力机制（S-MSA）来构建基本单元，即光谱维度自注意力块（SAB）。S-MSA将每个光谱特征图视为一个token，以计算沿光谱维度的自注意力。其次，通过使用SAB构建了单阶段光谱维度Transformer（SST），并利用U型结构提取多分辨率的光谱上下文信息，这对于HSI恢复至关重要。最后，MST++由多个SST级联，采用多阶段学习方案，逐渐从粗到细地提高重建质量，显著提升了性能，并为SR提出了一个新的框架MST++，这是探索Transformer在这项任务中的潜力的第一次尝试。

2 相关工作

2.1 高光谱图像采集

传统的高光谱成像系统通常采用光谱仪沿空间或光谱维度扫描场景。三种主要类型的扫描仪，包括扫帚扫描仪、推帚扫描仪和波段顺序扫描仪，经常用于捕捉高光谱图像。几十年来，这些扫描仪已广泛用于检测、遥感、医学成像和环境监测。例如，推帚扫描仪和扫帚扫描仪已经在卫星传感器中被用于摄影测量和遥感[5]。然而，扫描过程通常需要很长时间，这使其不适用于测量动态场景。此外，成像设备通常在物理上太大，无法插入便携式平台。为了解决这些限制，研究人员开发了SCI系统[10]来捕捉高光谱图像，其中3D高光谱立方体被压缩成单个2D测量[81]。在这些SCI系统中，编码孔径快照光谱成像（CASSI）[15]脱颖而出，成为一个有前途的研究方向。然而，迄今为止，SCI系统在消费级别的使用仍然价格昂贵。即使是“低成本”的SCI系统通常也在1万美元到10万美元之间。因此，SR主题具有重要的研究和实际价值。

2.2 基于RGB的光谱重建

传统的超分辨率（SR）方法[1]主要基于手工制作的高光谱先验。例如，Paramar等人[16]提出了一种用于HSI重建的数据稀疏扩展方法。Arad等人[2]提出了一种稀疏编码方法，创建了一个HSI信号及其RGB投影的字典。Aeschbacher等人[1]建议使用来自特定光谱先验的相对较浅的学习模型来实现光谱超分辨率。然而，这些基于模型的方法受到了有限的表示能力和差劲的泛化能力的影响。最近，受到深度学习在自然图像恢复中取得的巨大成功的启发，卷积神经网络（CNNs）已经被利用来学习从RGB到HSI的基本映射函数。例如，Xiong等人[22]提出了一个统一的HSCNN框架，用于从RGB图像和压缩测量中重建HSI。Shi等人[17]使用调整后的残差块构建了一个用于SR的深度残差网络HSCNN-R。Zhang等人[23]定制了一个像素感知的深度函数混合网络，用于建模RGB到HSI的映射。然而，这些基于CNN的SR方法取得了令人印象深刻的结果，但在捕捉非局部自相似性和远程相互依赖性方面显示出局限性。

2.3 视觉Transformer

NLP模型Transformer [18]被提出用于机器翻译。近年来，它已被引入计算机视觉，并因其在捕捉空间区域之间的长程相关性方面的优势而受到广泛关注。在高级视觉中，Transformer已被广泛应用于图像分类 [3]、目标检测 [11]、语义分割 [9]、人体姿态估计 [8]等领域。此外，视觉Transformer还被用于低级视觉 [7]。例如，Cai等人 [7]提出了第一个基于Transformer的端到端框架MST，用于从压缩测量中重建HSI。Lin等人 [6]将HSI稀疏性嵌入到Transformer中，建立了一个粗到细的学习方案，用于光谱压缩成像。先前的工作Uformer [20]采用了由Swin Transformer [13]块构建的U形结构，用于自然图像恢复。然而，据我们所知，Transformer在光谱超分辨率中的潜力尚未被探索。本研究旨在填补这一研究空白。

3 本文方法

3.1 网络架构

如图 1所示：(a) 描绘了提出的多阶段光谱Transformer (MST++)，它由 N_s 个单阶段光谱Transformer (SSTs) 级联而成。我们的MST++以RGB图像为输入，重建其HSI对应物。为了简化训练过程，利用了长身份映射。图 1 (b) 显示了U形SST，包括一个编码器、一个瓶颈和一个解码器。嵌入和映射块是单一的conv3×3层。编码器中的特征图依次经过下采样操作（步幅为conv4×4层）、 N_1 个光谱关注块 (SABs)、一个下采样操作和 N_2 个SABs。瓶颈由N3个SABs组成。解码器采用对称结构。上采样操作是一个步幅为deconv2×2的层。为了避免下采样中的信息丢失，编码器和解码器之间使用了跳跃连接。图 1 (c) 说明了SAB的组件，即前馈网络 (如图 1 (d) 所示)、光谱多头自注意 (S-MSA) 和两个层归一化。S-MSA的详细信息见图 1 (e)。

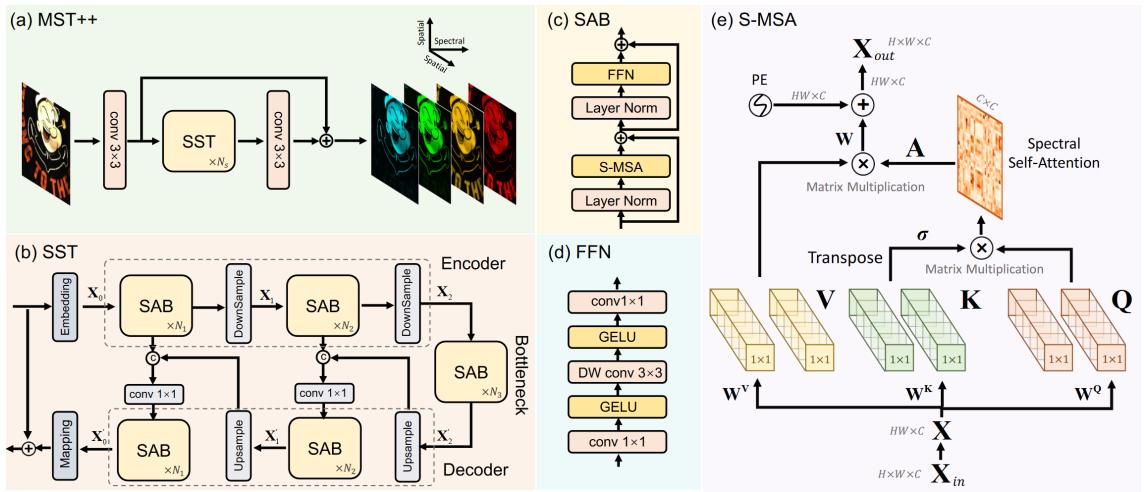


图 1. 网络整体架构

3.2 光谱维度的多头自注意力机制

假设 $X_{in} \in R^{H \times W \times C}$ 是 S-MSA 的输入，被重塑为 token $X \in R^{HW \times C}$ 。然后 X 线性投影为查询 $Q \in R^{HWC}$ ，键 $K \in R^{HW \times C}$ ，和值 $V \in R^{HW \times C}$ ：

$$Q = XW^Q, K = XW^K, V = XW^V \quad (1)$$

其中， W^Q, W^K 和 $W^V \in R^{C \times C}$ 是可学习的参数；为简化起见，省略了偏差。随后，分别沿着光谱通道维度将 Q、K 和 V 分割成 N 个头部： $Q = [Q_1, \dots, Q_N]$, $K = [K_1, \dots, K_N]$, $V = [V_1, \dots, V_N]$ 。每个头部的维度是 $d_h = \frac{C}{N}$ 。请注意，图 1 描述了 $N = 1$ 的情况，为简化起见省略了一些细节。与原始的多头自注意力 (MSA) 不同，S-MSA 将每个光谱表示视为一个令牌，并计算 $head_j$ 的自注意力：

$$A_j = softmax(\sigma_j K_j^T Q_j), head_j = V_j A_j \quad (2)$$

其中， K_j^T 表示 K_j 的转置矩阵。由于光谱密度相对于波长变化较大，使用一个可学习的参数 $\sigma_j \in R^1$ ，通过重新加权 $head_j$ 内的矩阵乘法 $K_j^T Q_j$ 来调整自注意力 A_j 。随后，N 个头部的输出被级联起来经历一个线性投影，然后与位置嵌入相加：

$$S - MSA(X) = (Contact_{j=1}^N(head_j))W + f_p(V) \quad (3)$$

其中， $W \in R^{C \times C}$ 是可学习的参数， $f_p()$ 是用于生成位置嵌入的函数。它由两个深度可分离的 conv3x3 层、一个 GELU 激活和 reshape 操作组成。HSI 沿着光谱维度按波长排序。因此，利用这个嵌入来编码不同光谱通道的位置信息。最后，重新整形方程(3)的结果，得到输出特征图 $X_{out} \in R^{H \times W \times C}$ 。

接下来介绍 Transformer 中 MSA 的一般范式，然后分析原始 Transformer 中的空间感知 MSA 和采用的 S-MSA 的计算复杂性。

3.3 与原始变压器的对比

3.3.1 MSA 的一般范式

将输入令牌表示为 $X \in R^{n \times C}$ ，其中 n 是待定的。在空间维度 MSA 中，n 表示 token 的数量。在 S-MSA 中，n 表示令牌的维度。首先将 X 线性投影到查询 $Q \in R^{n \times C}$ ，键 $K \in R^{n \times C}$ ，和值 $V \in R^{n \times C}$ ：

$$Q = XW^Q, K = XW^K, V = XW^V \quad (4)$$

其中， W^Q, W^K 和 $W^V \in R^{C \times C}$ 是可学习的参数；为简化起见，省略了偏差。随后，分别沿着光谱通道维度将 Q、K 和 V 分割成 N 个头部： $Q = [Q_1, \dots, Q_N]$, $K = [K_1, \dots, K_N]$, $V = [V_1, \dots, V_N]$ ，每个头部的维度是 $d_h = \frac{C}{N}$ 。然后，MSA 计算每个头部的自注意力：

$$head_j = MSA(Q_j, K_j, V_j) \quad (5)$$

随后，沿着光谱维度级联 N 个头的输出，并经过线性投影生成输出特征图 $X_{out} \in R^{n \times C}$ ：

$$X_{out} = (Contact_{j=1}^N(head_j))W \quad (6)$$

其中， $W \in R^{C \times C}$ 是可学习的参数。请注意，为了简化起见，省略了一些其他内容，如位置嵌入。因为只比较原始空间感知 MSA 和 S-MSA 之间的主要区别，即方程(5)的具体表达形式。

3.3.2 空间MSA

空间感知MSA将沿着光谱维度的像素向量视为一个令牌，然后计算每个头部的自注意力。因此，方程(5)可以具体规定为：

$$head_j = A_j V_j, \quad A_j = \text{softmax}\left(\frac{Q_j K_j}{\sqrt{d_h}}\right) \quad (7)$$

方程(7)需要计算N次。因此，空间感知MSA的计算复杂度为：

$$O(Spatial - MSA) = N(n^2 d_h + n^2 d_h) = 2n^2 C \quad (8)$$

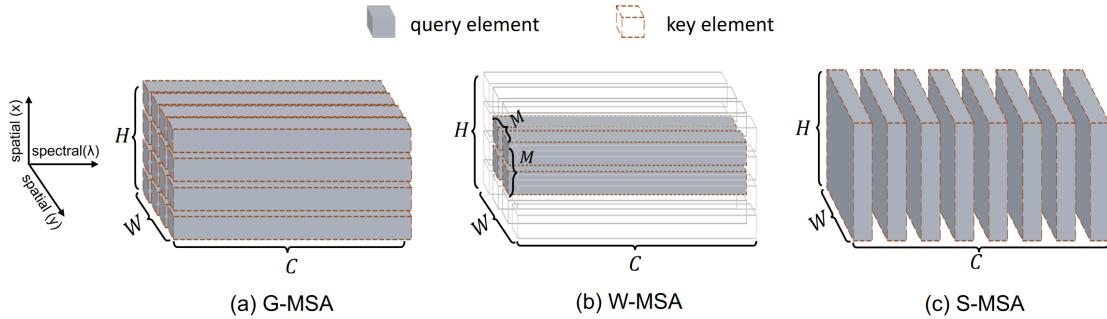


图 2. 不同MSA的关系图

空间感知MSA主要分为两类：全局MSA 和基于局部窗口的MSA [11]。现在分析这两种MSA。如图 2 (a) 所示，全局MSA对所有token进行采样，作为键和查询元素，然后计算自注意力。因此，令牌的数量n（键或查询元素）等于HW。然后，根据方程(8)，全局MSA的计算复杂度是：

$$O(Global MSA) = 2(HW)^2 C \quad (9)$$

这对于输入特征图的空间大小是二次的。全局MSA具有非常大的感受野，但其计算成本不可忽视，有时难以承受。同时，对冗余键元素的采样可能很容易导致过度平滑的结果，甚至出现不收敛的问题。为了降低计算成本，研究人员提出了基于局部窗口的MSA。

如图 2 (b) 所示，W-MSA首先将特征图分割为大小为 M^2 的非重叠窗口，并采样每个窗口内的所有令牌来计算自注意力。因此，令牌的数量n等于 M^2 ，W-MSA为所有窗口执行 $\frac{HW}{M^2}$ 次。因此，计算复杂度是：

$$O(W - MSA) = \frac{HW}{M^2} (2(M^2)^2 C) = 2M^2 HWC \quad (10)$$

这对于空间大小 (HW) 是线性的。W-MSA具有低计算成本的优势，但在特定位置的窗口内具有有限的感受野。因此，可能会忽略一些高度相关的非局部令牌。原始的空间感知MSA旨在捕捉空间区域的长程依赖关系。然而，HSI表示在空间上是稀疏的，而在光谱上是相似且相关的。捕捉空间上的相互作用可能不如建模谱上的相关性划算。基于这一HSI特性，采用了S-MSA。

3.3.3 S-MSA

如图 2 (c) 所示, S-MSA 将每个光谱特征图视为一个 token, 并沿着光谱维度计算自注意力。然后, 方程(5)可以具体规定为:

$$A_j = \text{softmax}(\sigma_j K_j^T Q_j), \text{head}_j = V_j A_j \quad (11)$$

其中, K_j^T 表示 K_j 的转置矩阵。注意到, 与波长相比, 光谱密度变化很大。因此, 利用一个可学习的参数 $\sigma_j \in R^1$, 通过重新加权 head_j 内的矩阵乘法 $K_j^T Q_j$ 来调整自注意力 A_j 。由于 S-MSA 将整个特征图视为一个 token, 每个 token 的维度 n 等于 HW 。方程 (11) 需要计算 N 次。因此, S-MSA 的复杂性是:

$$O(S-MSA) = N(d_h^2 n + d_h^2 n) = \frac{2HWC^2}{N} \quad (12)$$

W-MSA 和 S-MSA 的计算复杂性对于空间大小 (HW) 是线性的, 比全局 MSA (对于 HW 是二次的) 要便宜得多。然而, S-MSA 将每个光谱特征视为一个令牌。在计算自注意力 A_j 时, S-MSA 将其视为全局空间位置。因此, S-MSA 的感受野是全局的, 而不限于位置特定的窗口。此外, S-MSA 沿着光谱维度计算自注意力, 基于 HSI 的特性, 与空间感知 MSA 相比更适用于 HSI 重建。因此, S-MSA 被认为比全局 MSA 和 W-MSA 更具成本效益。。S-MSA 具有全局感受野, 模拟光谱方面的自相似性, 并且计算成本是线性的。

4 复现细节

由于原文采用沿着光谱维度的 S-MSA, 能显著降低计算的复杂度, 但是也导致了 token 数量少以及单个序列过长的问题, 此时自注意力计算量将远超其他部分, 成为模型的瓶颈, 针对这种长序列的 Transformer 主要存在以下的问题: 在自注意力计算过程中, 非线性函数 softmax 的存在使得计算复杂度始终无法降到输入序列长阻碍了矩阵连乘中运算顺序的长度的平方以下。第二, 矩阵乘法计算规模较大参与连乘计算的矩阵规模与输入序列长度成量很大, 且在自注意力计算中的占比较高正比, 导致这部分矩阵乘法的计算。第三, 计算数据存在冗余, 阵中具体参与乘加计算的元素数值差异很大这部分不重要的数据导致存在大量冗余的乘数值较小的元素对结果影响很小加运算。将 softmax 函数用 $s_i = \frac{e^i}{\sum_j e^j}$ 的形式展开后可以得到每行自注意力矩阵 A_i 的计算方式:

$$A_i = \frac{\sum_{j=1}^N e^{\frac{Q_i k_j^T}{\sqrt{u}}} v_j}{\sum_{j=1}^N e^{\frac{Q_i k_j^T}{\sqrt{u}}}} \quad (13)$$

其中 Q_i 和 K_j 分别表示阵的第 i 行和 K 阵的第 j 行。通过将注意力模块中的 softmax 函数替换为 ReLU 函数, 并对其进行加权调整, 这样能降低激活函数的计算复杂度, 同时能够对自注意力部分进行线性化, 从而能够调整 Q, K, V 的矩阵相乘的乘法顺序, 由于模型在自注意力模块的复杂度较高, 减少此处的计算量后也能从一定程度上加快训练速度, 参考 Nystromformer [21] 通过平均池化的方法, 对形状为 $N \times d$ 的 Q 和 K 矩阵按行进行聚类, 对每 $\lceil \frac{N}{k} \rceil$ 个行向量求一次平均值, 得到维度为 $R^{k \times d}$ 的矩阵和 \bar{Q} 和 \bar{R} , 然后参考 Nystrom 方法按如下方式计算自注意力:

$$A_i = \frac{\text{ReLU}(Q_i) \sum_{j=1}^N \frac{1}{j} \text{ReLU}(K_j^T) V_j}{\text{ReLU}(Q_i) \sum_{j=1}^N \frac{1}{j} \text{ReLU}(K_j^T)} \quad (14)$$

4.1 实验环境搭建

本实验使用PyCharm作为主要的开发工具。项目根目录下有一个 requirements.txt 文件，其中包含了所有依赖项的列表。使用 pip install -r requirements.txt 命令来安装这些依赖性即可。

4.2 界面分析与使用说明

项目结构如图 3 所示，将训练、验证、测试、预测的代码分四个文件夹存放，此外 visualization 文件夹存放可视化的代码，用于生成 RGB 图像重构高光谱图像的结果。

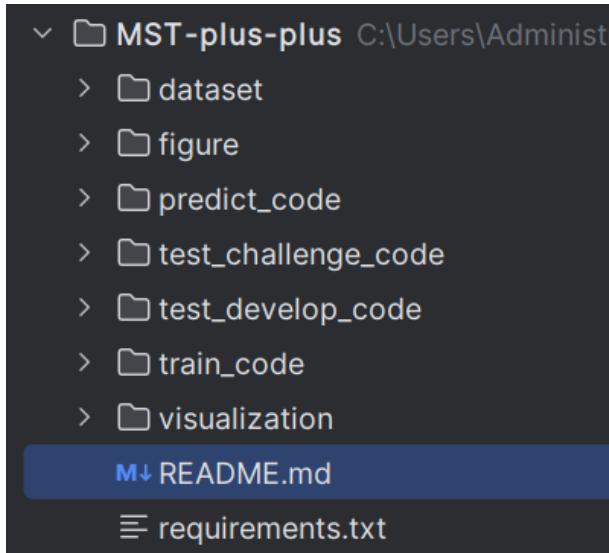


图 3. 项目结构示意图

4.3 创新点

替换 Softmax 函数：通过将注意力模块中的 Softmax 函数替换为 ReLU 函数，以及对其进行加权调整，实现对自注意力部分的线性化。这种替换能够调整 Q、K、V 的矩阵相乘的乘法顺序，从而降低计算复杂度。这有助于减少模型在自注意力模块的计算负担，提高训练速度。
Nystrom 方法的引入：参考 Nystrom 方法，通过平均池化对 Q 和 K 矩阵按行进行聚类，从而降低了输入序列长度，减少了计算规模。该方法能够通过求取每一小块的平均值，得到维度较小的矩阵，然后应用于自注意力计算。
减少计算数据冗余：注意到存在具体参与乘加计算的元素数值差异很大，通过减少计算数据中的冗余，特别是那些对结果影响较小的元素，进一步提高计算效率。

5 实验结果分析

使用原始论文的方法重构的高光谱图像，从中提取 480nm、520nm、580nm、660nm 的单波段图像，并用对应波段的颜色值生成伪彩图像，与改进后的结果对比如图 4 所示，从视觉上两者在重构效果上并无明显差距，通过计算相关系数，如图 5 所示，图像的相关性较高，但在 410nm 波段有较大误差，此外 480nm 左右也有一定误差，可以从结果图中看出。此外，模型的训练时间，改进方案大致是 MST++ 原始时间的 0.95 倍，有一定的提升。

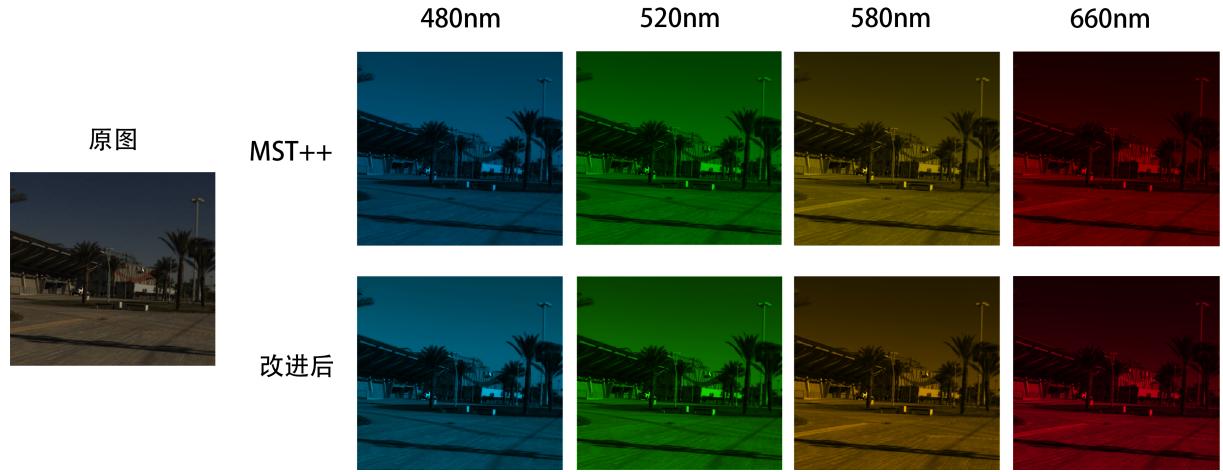


图 4. 相关系数对比

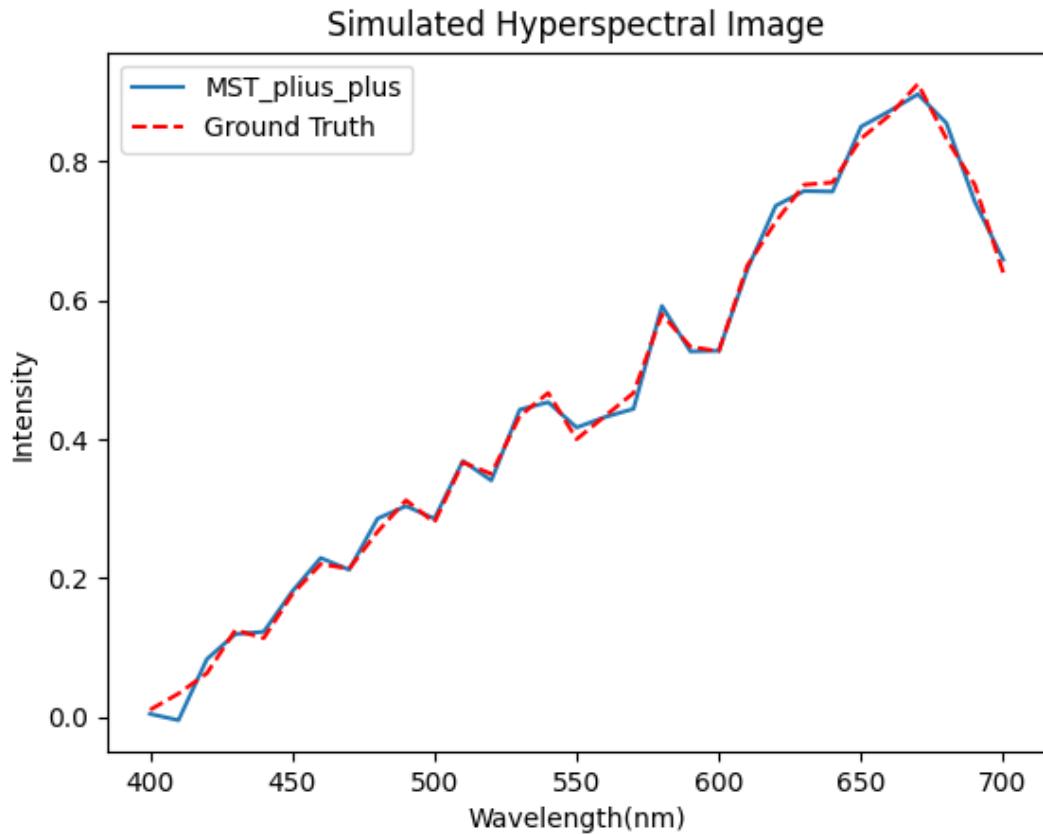


图 5. 实验结果示意

6 总结与展望

本篇文章主要是创新性的使用Transformer进行光谱重构任务，并采用了光谱维度多头自注意力机制（S-MSA）来构建基本单元，即光谱维度自注意力块（SAB）。S-MSA将每个光谱特征图视为一个token，以计算沿光谱维度的自注意力，大大降低了模型的复杂度。但是由于S-MSA使用沿光谱维度的自注意力机制，每个序列过长，本文通过替换Softmax函数、引

入 Nystrom 方法以及对矩阵乘法进行调整，思考了长序列带来的挑战，在保证能达到SOTA的情况下提升了一定训练的速度，并且还解决了在训练早期的梯度不稳定的问题。为提高模型的适用性和性能提供了有效的方法。

参考文献

- [1] Jonas Aeschbacher, Jiqing Wu, and Radu Timofte. In defense of shallow learned spectral reconstruction from rgb images. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 471–479, 2017.
- [2] Boaz Arad and Ohad Ben-Shahar. Sparse recovery of hyperspectral signal from natural rgb images. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part VII 14*, pages 19–34. Springer, 2016.
- [3] Anurag Arnab, Mostafa Dehghani, Georg Heigold, Chen Sun, Mario Lučić, and Cordelia Schmid. Vivit: A video vision transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 6836–6846, 2021.
- [4] V Backman, Michael B Wallace, LT Perelman, JT Arendt, R Gurjar, MG Müller, Q Zhang, G Zonios, E Kline, T McGillican, et al. Detection of preinvasive cancer cells. *Nature*, 406(6791):35–36, 2000.
- [5] Michael Breuer and Jörg Albertz. Geometric correction of airborne whiskbroom scanner imagery using hybrid auxiliary data. *International Archives of Photogrammetry and Remote Sensing*, 33(B3/1; PART 3):93–100, 2000.
- [6] Yuanhao Cai, Jing Lin, Xiaowan Hu, Haoqian Wang, Xin Yuan, Yulun Zhang, Radu Timofte, and Luc Van Gool. Coarse-to-fine sparse transformer for hyperspectral image reconstruction. In *European Conference on Computer Vision*, pages 686–704. Springer, 2022.
- [7] Yuanhao Cai, Jing Lin, Xiaowan Hu, Haoqian Wang, Xin Yuan, Yulun Zhang, Radu Timofte, and Luc Van Gool. Mask-guided spectral-wise transformer for efficient hyperspectral image reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17502–17511, 2022.
- [8] Yuanhao Cai, Zhicheng Wang, Zhengxiong Luo, Binyi Yin, Angang Du, Haoqian Wang, Xiangyu Zhang, Xinyu Zhou, Erjin Zhou, and Jian Sun. Learning delicate local representations for multi-person pose estimation. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16*, pages 455–472. Springer, 2020.
- [9] Hu Cao, Yueyue Wang, Joy Chen, Dongsheng Jiang, Xiaopeng Zhang, Qi Tian, and Manning Wang. Swin-unet: Unet-like pure transformer for medical image segmentation. In *European conference on computer vision*, pages 205–218. Springer, 2022.

- [10] Xun Cao, Tao Yue, Xing Lin, Stephen Lin, Xin Yuan, Qionghai Dai, Lawrence Carin, and David J Brady. Computational snapshot multispectral cameras: Toward dynamic capture of the spectral world. *IEEE Signal Processing Magazine*, 33(5):95–108, 2016.
- [11] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-to-end object detection with transformers. In *European conference on computer vision*, pages 213–229. Springer, 2020.
- [12] Min H Kim, Todd Alan Harvey, David S Kittle, Holly Rushmeier, Julie Dorsey, Richard O Prum, and David J Brady. 3d imaging spectroscopy for measuring hyperspectral patterns on solid objects. *ACM Transactions on Graphics (TOG)*, 31(4):1–11, 2012.
- [13] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 10012–10022, 2021.
- [14] Farid Melgani and Lorenzo Bruzzone. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on geoscience and remote sensing*, 42(8):1778–1790, 2004.
- [15] Ziyi Meng, Jiawei Ma, and Xin Yuan. End-to-end low cost compressive spectral imaging with spatial-spectral self-attention. In *European conference on computer vision*, pages 187–204. Springer, 2020.
- [16] Manu Parmar, Steven Lansel, and Brian A Wandell. Spatio-spectral reconstruction of the multi-spectral datacube using sparse recovery. In *2008 15th IEEE International Conference on Image Processing*, pages 473–476. IEEE, 2008.
- [17] Zhan Shi, Chang Chen, Zhiwei Xiong, Dong Liu, and Feng Wu. Hscnn+: Advanced cnn-based hyperspectral recovery from rgb images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 939–947, 2018.
- [18] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.
- [19] Lizhi Wang, Chen Sun, Maoqing Zhang, Ying Fu, and Hua Huang. Dnu: Deep non-local unrolling for computational spectral imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1661–1671, 2020.
- [20] Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. Uformer: A general u-shaped transformer for image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 17683–17693, 2022.

- [21] Yunyang Xiong, Zhanpeng Zeng, Rudrasis Chakraborty, Mingxing Tan, Glenn Fung, Yin Li, and Vikas Singh. Nyströmformer: A nyström-based algorithm for approximating self-attention. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 14138–14148, 2021.
- [22] Zhiwei Xiong, Zhan Shi, Huiqun Li, Lizhi Wang, Dong Liu, and Feng Wu. Hscnn: Cnn-based hyperspectral image recovery from spectrally undersampled projections. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 518–525, 2017.
- [23] Lei Zhang, Zhiqiang Lang, Peng Wang, Wei Wei, Shengcai Liao, Ling Shao, and Yanning Zhang. Pixel-aware deep function-mixture network for spectral super-resolution. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 12821–12828, 2020.