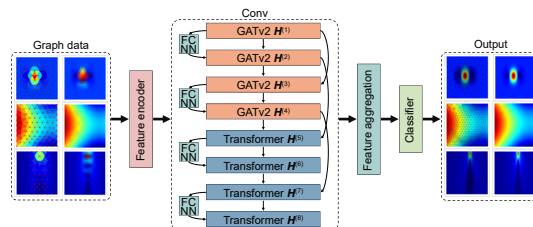


图注意力 Transformer 网络的自适应网格划分

韩 婷, 叶 佳*, 闫连山, 甘宗鑫

西南交通大学信息科学与技术学院, 四川 成都 611756



摘要: 针对有限元分析中自适应网格划分面临的计算精度与效率协同优化难题, 本研究提出基于注意力融合机制的 GTF-Net 框架。该模型创新融合图注意力网络与 Transformer 架构, 通过多头交叉注意力模块实现局部几何特征与全局物理场的动态耦合, 增强对奇异场及复杂边界的表征能力。经光波导传输和贝塞尔函数双案例验证, 相较传统 Scikit-FEM (skFem) 方法, GTF-Net 在保持计算效率优势的同时, 梯度误差标准差分别降低 85.9% 和 23.8%。结果表明, 该框架通过非线性特征映射显著提升网格分布与物理场变化的匹配度, 为解决工程计算中的自适应网格优化问题提供深度学习新方案。

关键词: 自适应网格细化; GATv2-Transformer 融合网络; 有限元分析; 深度学习

中图分类号: TP391.9

文献标志码: A

韩婷, 叶佳, 闫连山, 等. 图注意力 Transformer 网络的自适应网格划分 [J]. 光电工程, 2025, 52(4): 250058

Han T, Ye J, Yan L S, et al. Adaptive mesh partitioning for graph attention Transformer networks[J]. *Opto-Electron Eng*, 2025, 52(4): 250058

Adaptive mesh partitioning for graph attention Transformer networks

Han Ting, Ye Jia*, Yan Lianshan, Gan Zongxin

School of Information Science and Technology, Southwest Jiaotong University, Chengdu, Sichuan 611756, China

Abstract: To address the challenge of balancing the computational accuracy and efficiency in adaptive finite element meshing, this study proposes a GTF-Net model based on the attention fusion mechanism. The model combines the graph attention network with the Transformer architecture, dynamically couples local geometric features with the global physical field through a multi-head cross-attention module, and enhances the representation of singular fields and complex boundaries. The verification of two case studies of waveguide transmission and Bessel equation shows that compared with the traditional Scikit-FEM (skFem) method, GTF-Net improves computational efficiency while reducing the standard deviation of gradient error by 85.9% and 23.8%, respectively. The results show that the model significantly improves the fit between mesh distribution and physical field changes through nonlinear feature mapping, providing a novel deep learning solution for adaptive mesh optimization in engineering calculations.

Keywords: adaptive mesh refinement; GATv2-Transformer fusion network; finite element analysis; deep learning

收稿日期: 2025-02-27; 修回日期: 2025-04-09; 录用日期: 2025-04-10

基金项目: 中国国家重点研发计划 (2022YFB2802701); 国家自然科学基金 (U23A20376, 62075185, 62271422); 四川省杰出青年科学基金 (2024NSFJQ0016)

*通信作者: 叶佳, jiaye@home.swjtu.edu.cn。

版权所有©2025 中国科学院光电技术研究所

1 引言

在现代应用科学研究与工程中, 复杂的几何模型和物理现象往往用偏微分方程 (partial differential equation, PDE)^[1] 建模, 并借助有限元法 (finite element method, FEM)^[2] 或有限体积法 (finite volume method, FVM)^[3] 等数值方法进行求解。该过程可分为预处理、求解和后处理三个阶段。网格划分是预处理的主要任务, 也是整个分析过程的关键任务。若网格尺寸过大, 则无法满足计算精度要求; 反之, 尺寸过小, 单元数量增多, 则求解计算量快速增加^[4]。自适应网格细化 (adaptive mesh refinement, AMR) 可以大大缓解计算精度和计算效率之间的矛盾, 已被应用于固体力学^[5]、流体力学^[6]、航空航天^[7]、声学物理模拟^[8] 等许多领域。由此可以看出, AMR 是提高 PDE 求解精度的重要手段。

目前, AMR 方法主要分为传统方法、基于神经网络 (neural network, NN) 和强化学习 (reinforcement learning, RL) 的扩展方法。传统 AMR 方法主要包括基于误差估计、几何特征和物理量的网格细化策略, 这些方法通过分析数值解的误差^[9]、几何形状的曲率^[10] 或物理量的梯度分布^[11] 来动态调整网格的密度。经过多年的发展, 传统方法在理论和应用上已经成熟, 适用于多种科学计算问题。然而, 误差估计、物理量计算和相关迭代需要大量的计算资源, 增加了整体的计算负担。同时, 对于复杂的多物理场耦合问题, 可能难以兼顾不同物理量的需求, 影响网格划分效果。

近年来, 神经网络的快速发展为 AMR 提供了新的思路, 研究者利用包括人工神经网络 (artificial neural network, ANN)^[12]、自组织映射 (self-organizing maps, SOM)^[13]、递归神经网络 (recurrent neural network,

RNN)^[14]、全连接神经网络 (fully connected neural network, FCNN)^[15] 在内的多种方法, 在理论和实践上证明了神经网络在 AMR 领域中相较于传统 AMR 的显著优势。但上述方法高度依赖于初始条件和参数选择, 其性能对不同的初始条件、几何形状等非常敏感。图神经网络 (graph neural network, GNN) 直接处理图结构数据, 可以自然适应任意拓扑结构的网格, 而不受固定输入、输出格式的限制^[16-17], 因此在 AMR 中具有显著的优势。现有的相关方法还存在一些问题, GMR-Net 无法泛化超出以许多预定范围顶点为特征的几何形状^[18]; GraphMesh 在求解复杂几何形状和偏微分方程方面的表现仍需进一步验证^[19]; Pelissier 等^[20] 提出的方法会产生不稳定状态, 预测质量较差; 3DMeshNet 方法在处理复杂几何形状时仍面临一定的挑战^[21]。在此基础上, 研究人员尝试将 GNN 与 FEM 求解器相结合^[22], 将 NN 与 RL 相结合^[23], 以进一步拓展 AMR 的应用和效率。尽管如此, 现有方法仍然面临一定的挑战, 需要进一步优化数据效率、增强泛化能力、加速训练过程、优化对复杂几何形状的适应性。

本文将 AMR 问题转化为一个节点分类任务, 即将初始网格中的每个网格节点视为一个待分类样本, 结合局部物理量 (如梯度、残差、网格尺寸等) 构建特征向量, 并将节点分类为细化或保持不变, 从而以数据驱动的方式替代传统的启发式自适应策略, 实现更具泛化能力的网格划分方法。这个过程见图 1, 标红节点为需要进行细化的节点。

本文提出了一种基于图神经网络的 AMR 方法, 即 GATv2-Transformer 融合网络 (GTF-Net)。该方法利用图神经网络对网格单元之间的关系进行建模, 其中每个网格节点被视为图中的一个节点, 节点之间的

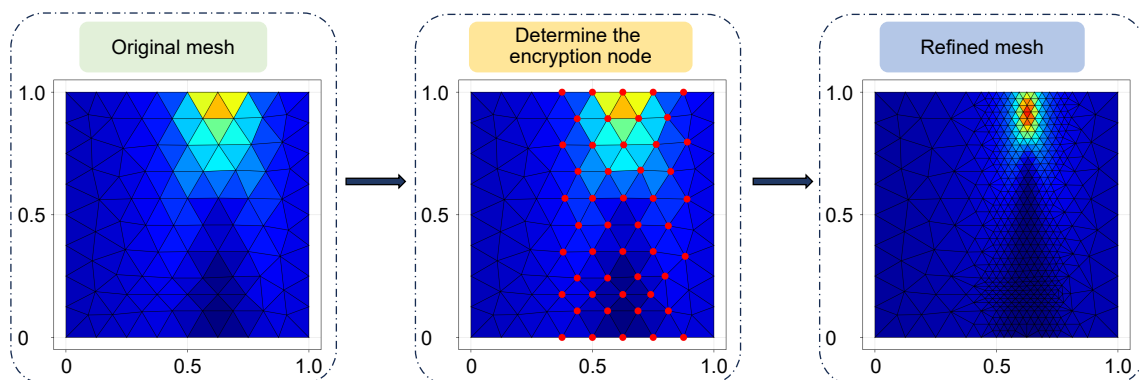


图 1 节点分类问题下自适应网格划分过程

Fig. 1 Process of adaptive mesh refinement under node classification problem

边表示单元之间的拓扑关系。为了增强该方法的泛化能力并减少其对人工标注数据的依赖, 本文结合了基于可控随机函数的图结构数据生成方法, 该方法通过模拟多种实际应用场景生成多样化的训练数据, 在保证数据分布多样性的同时显著减少了真实数据采集和标注的成本。通过多层图卷积网络对局部和全局信息的协同捕捉, GTF-Net 能够有效学习生成数据中的复杂关系模式, 从而实现对不同几何形态和物理特性的自适应调整。

特别地, 本文在网络结构中引入了注意力融合机制, 以进一步增强模型对局部细节与全局结构的综合感知能力^[24-25]。在 GATv2 和 Transformer 模块的结合过程中, 不同层次与尺度的注意力权重通过融合操作动态整合, 赋予模型更强的表达能力和判别能力, 能够引导模型在特征空间中聚焦于误差敏感区域, 提升关键区域的识别与细化精度。因此, 注意力融合不仅提升了 GTF-Net 的稳定性与泛化性, 也为复杂场景下的自适应网格划分提供了更具泛化性的解决方案。

2 准备工作

GNN 是一类深度学习模型, 旨在处理图结构数据^[26]。与对欧几里得数据进行操作的传统神经网络不同, GNN 利用节点和边之间的关系来学习有意义的表示, 其关键思想是通过图的拓扑结构来聚合邻域信息, 学习目标节点的信息表征^[27]。

在有限元计算过程中, 整个计算域 Ω 被划分为多个网格单元 K , 其中包含若干个节点 L , 对于每个节点 L , 定义一个特征向量 $\mathbf{x}$, 表达式为

$$\mathbf{x}(L) = (\mathbf{x}_1(L), \mathbf{x}_2(L), \dots, \mathbf{x}_m(L)) \in \mathbb{R}^m, \quad (1)$$

式中: m 为特征数量。 $\mathbf{x}$ 用来度量该节点的局部性质 (例如数值解的梯度、几何特征、物理量等)。

设置一个映射, 表示为

$$\Phi: \mathbb{R}^m \rightarrow y, \quad (2)$$

式中: y 为一个有限类别集 $\{0, 1\}$; Φ 为分类函数。对每个节点 L , 令

$$I(L) = \Phi(\mathbf{x}(L)). \quad (3)$$

当 $I(L)$ 等于 1 时, 对节点 L 执行“细化”操作; 当 $I(L)$ 等于 0 时, 不执行细化操作。

在上述框架下, AMR 的核心过程即转化为对网格中每个节点 L 进行分类预测, 根据分类结果执行相应的细化/保持操作, 从而得到更新后的自适应网格, 实现将自适应网格的“局部误差判定”等价地表达为

一个分类问题。

在本文提出的框架中, 给定一个由顶点集 V 和边集 E 组成的网格 $G = (V, E)$, 其中顶点特征矩阵 $\mathbf{X} \in \mathbb{R}^{N \times d}$ 表示网格点的几何信息, N 表示节点的数量, d 表示节点的维度, 边特征矩阵 $\mathbf{E} \in \mathbb{R}^{M \times d_e}$ 反应网格的拓扑关系及属性, M 表示边的数量。其目标是通过梯度误差估计驱动自适应网格划分, 从而优化计算精度。梯度误差由计算网格的插值梯度 $\nabla \mathbf{u}_h$ 与真实物理场梯度 $\nabla \mathbf{f}$ 之间的偏差衡量, 定义为

$$grad_{diff} = \int_{\Omega} (\nabla \mathbf{f} - \nabla \mathbf{u}_h) \cdot (\nabla \mathbf{f} - \nabla \mathbf{u}_h) d\Omega. \quad (4)$$

较大的 $grad_{diff}$ 反映了数值解的不准确性, 意味着该区域需要进行细化以提高计算精度。不同于基于手工设置阈值的方法, 本研究采用图神经网络直接学习细化策略, GNN 通过端对端训练预测细化评分 y_i , i 为点的标签, 从而决定网格点的细化程度。具体而言, GNN 通过映射学习细化规则, 表达式为

$$y_i = f_{\theta}(\mathbf{X}, \mathbf{E}, \mathbf{A}), \quad (5)$$

式中: $\mathbf{A}$ 为邻接矩阵; f_{θ} 是 GNN 的非线性映射函数。模型通过梯度误差驱动优化, 使得细化评分 y_i 与 $grad_{diff}$ 保持一致性。最终, 基于 y_i 进行局部预测, 实现高效的自适应网格优化。

数据集生成方面, 本文所用的数据生成算法融合了非结构化网格动态解析与电磁场参数随机化技术, 构建出兼具物理规律性和数据多样性的图结构数据集 (如算法 1 所示)。其中, 随机函数中包含多种电磁场的生成与处理方法, 能够生成多种不同的电磁场配置, 部分函数可见表 1。

算法1 自适应网格训练数据生成方法

输入: 非结构化网格mesh, 生成样本的数量n_samples, 用于节点标记的阈值tolerance, 输出图类型graph_type

输出: 网格列表graph_list, 标签列表label_list

01. 使用随机函数创建初始网格图

02. 初始化graph_list和label_list

03. **FOR** 1: n_samples:

04. 复制基础网格并更新为新的随机函数

05. 计算单元梯度差

06. 初始化节点标签为0

07. **IF** 梯度差>阈值:

08. 标记节点标签为1

09. **END IF**

10. 按照图类型分类, 添加到网格列表

11. **END FOR**

12. 返回 网格列表, 标签列表

表 1 训练数据生成方法中部分随机函数
Table 1 Part of random function in training data generation method

Function name	Description
plane_wave()	Generate two-dimensional plane wave, using cosine function to represent the propagation of electromagnetic waves.
cylindrical_wave()	Generate cylindrical wave, where the wave intensity is related to the radial distance from the source point, described using Bessel function.
random_waveguide_field()	Generate random waveguide field, providing random refractive index and waveguide parameter.
gaussian_beam()	Generate Gaussian beam, calculate the radial distance, and adjust the beam width based on propagation distance.
random_interface_field()	Generate random interface field, selecting random interface types and Gaussian beams as base field.
step_interface()	Define step interface, determining the change in dielectric constant of material at boundary position.
periodic_interface()	Define periodic interface, producing periodic material structure.
multipole_field()	Define multipole field, using Bessel function and angle to describe multipoles of different orders.
evanescent_wave()	Define evanescent wave, primarily decaying in the y-direction.

3 网络框架理论描述

以下部分提供了神经网络框架的基本组成部分, 整体网络框架见图 2。

3.1 多层图注意力网络卷积 (GATv2Conv)

本方法首先将原始节点特征 X 映射到一个维度为 h 的中间空间, 并采用带有激活函数和丢弃机制的前馈网络进行特征变换:

$$H^{(0)} = \text{Dropout}(\sigma(W^0 X)), \quad (6)$$

式中: $W^0 \in \mathbb{R}^{h \times d}$ 是可训练的权重矩阵; $\sigma(\cdot)$ 是非线性激活函数 (例如 ReLU); $\text{Dropout}(\cdot)$ 是随机将元素置零以进行正则化的操作。

随后构建四层堆叠的 GATv2Conv 模块进行基于注意力机制的信息传递。每个 GATv2Conv 层通过动态注意力权重聚合邻居节点信息^[28]。设 $H^{(l)}$ 为第 l 层的节点表示, 第 l 个 GATv2Conv 可以简化为

$$H^{(l)} = \text{GATv2Conv}(G, H^{(l-1)}, E). \quad (7)$$

组件 GATv2 通过可学习的注意力系数 α_{ij} 来实现邻居加权聚合, 表达式为

$$\alpha_{ij} = \text{softmax}_{j \in N(i)}(\alpha^T \text{LeakyReLU}(W \cdot [h_i \| h_j])), \quad (8)$$

式中: $N(i)$ 表示节点 i 的邻居节点集合; $\|$ 表示向量连接; α 和 W 表示可学习参数; h_i 、 h_j 分别表示节点 i 和节点 j 的特征向量。每个 GATv2Conv 层后接层归一化、激活函数及可选的 Dropout 操作。

在实际模型中, 本文提到的方法在层级结构中引入跨层跳跃连接机制, 建立非相邻层间的直接通路, 以实现多层次特征的复用和融合, 从而增强模型的表达能力。

$$\begin{cases} H^{(1)} = \text{Dropout}(\text{Norm}(\sigma(\text{GATv2Conv}(G, H^{(0)}, E)))) \\ H^{(2)} = \text{Norm}(\sigma(\text{GATv2Conv}(G, H^{(1)}, E))) + \text{Skip}(H^{(1)}) \\ H^{(3)} = \text{Dropout}(\text{Norm}(\sigma(\text{GATv2Conv}(G, H^{(2)}, E)))) + H^{(1)} \\ H^{(4)} = \text{Norm}(\sigma(\text{GATv2Conv}(G, H^{(3)}, E))) + \text{Skip}(H^{(3)}) \end{cases}, \quad (9)$$

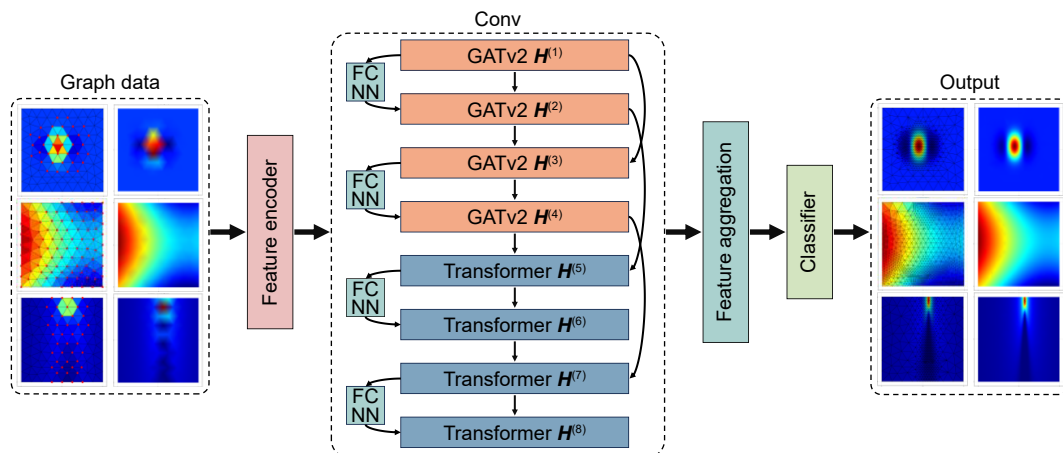


图 2 GTF-Net 结构示意图

Fig. 2 GTF-Net structure diagram

式中: $\text{Norm}(\cdot)$ 为表示对张量或矩阵的归一化操作; $\text{Skip}(\cdot)$ 表示跳跃连接。跳跃连接的引入使得每层的输出可以直接与后续层的输入结合, 实现多层次特征的协同优化。这种结构不仅增强了模型的泛化能力, 还通过局部注意力机制增强了关键特征的提取。

3.2 多层 Transformer 卷积 (TransformerConv)

在 GATv2Conv 层之后, 网络架构依次接入四层 TransformerConv 模块。TransformerConv 是基于多头自注意力机制的图数据变体 (类似于标准 Transformer 但专为图数据设计)^[29]。设 $\mathbf{Q}$ 、 $\mathbf{K}$ 、 $\mathbf{V}$ 为节点嵌入计算的查询、键和值。在包含 H 个头的多头设置中, 每个注意力头的计算过程定义为

$$\text{Attn}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V}, \quad (10)$$

式中: $\sqrt{d_k}$ 是缩放因子, 用于防止点积结果过大导致 softmax 函数的梯度消失。多头注意力合并时,

$$\mathbf{Z} = \parallel_{h=1}^H \text{Attn}(\mathbf{Q}_h, \mathbf{K}_h, \mathbf{V}_h) \left(\mathbf{Z} \in \mathbb{R}^{N \times d_k \cdot H} \right), \quad (11)$$

式中: N 为节点数; d_k 为键的维度。

每个 TransformerConv 层在实现时既考虑图结构的连接关系, 又融合潜在的边特征信息, 但概念上遵循相同的多头注意力机制。注意力计算完成后, 执行包含层归一化、非线性激活函数、Dropout 正则化及残差连接的后处理操作,

$$\begin{cases} \mathbf{H}^{(5)} = \text{Dropout}\left(\text{Norm}\left(\sigma\left(\text{TransformerConv}\left(G, \mathbf{H}^{(4)}, \mathbf{E}\right)\right)\right) + \mathbf{H}^{(2)}\right) \\ \mathbf{H}^{(6)} = \text{Norm}\left(\sigma\left(\text{TransformerConv}\left(G, \mathbf{H}^{(5)}, \mathbf{E}\right)\right) + \text{Skip}\left(\mathbf{H}^{(5)}\right)\right) \\ \mathbf{H}^{(7)} = \text{Dropout}\left(\text{Norm}\left(\sigma\left(\text{TransformerConv}\left(G, \mathbf{H}^{(6)}, \mathbf{E}\right)\right)\right) + \mathbf{H}^{(4)}\right) \\ \mathbf{H}^{(8)} = \text{Norm}\left(\sigma\left(\text{TransformerConv}\left(G, \mathbf{H}^{(7)}, \mathbf{E}\right)\right) + \text{Skip}\left(\mathbf{H}^{(7)}\right)\right) \end{cases} \quad (12)$$

这种多层级联的 TransformerConv 结构通过全局注意力机制动态捕捉图中节点间关系, 有效提取从局部到全局的特征, 进一步增强模型表达能力, 稳定训练过程, 并在一定程度上缓解了梯度消失问题。

3.3 特征聚合与最终分类

在特征聚合阶段, 网络采用跨层连接与均值池化相结合的策略整合多尺度节点表示为

$$\mathbf{H}^{(\text{agg})} = \text{mean}\left(\text{concat}\left(\mathbf{H}^{(1)}, \mathbf{H}^{(2)}, \dots, \mathbf{H}^{(8)}, \text{dim} = 3\right)\right), \quad (13)$$

式中: $\text{mean}(\cdot)$ 表示对输入的向量或矩阵进行均值操作; $\text{concat}(\cdot)$ 表示将多个向量或矩阵按指定维度拼接在一起; dim 表示拼接的维度。进一步通过可学习的参数矩阵优化特征聚合层, 表示为

$$\mathbf{H}^{(\text{agg})} = \text{Dropout}\left(\sigma\left(\mathbf{W}^f \mathbf{H}^{(\text{agg})}\right)\right), \quad (14)$$

式中: $\mathbf{W}^f$ 是可训练的权重矩阵。

最终得到一个集成所有层多尺度节点表示的特征矩阵 $\mathbf{H}^{(\text{agg})}$, 应用小型多层感知器 (MLP) 进一步优化该组合嵌入。接下来, 将聚合后的嵌入输入更深的 MLP 进行分类。设 $\mathbf{Z}_1$ 为最终 MLP 的输出, 表示为

$$\mathbf{Z}_1 = \text{Dropout}\left(\sigma\left(\mathbf{W}^{(\text{MLP2})} \sigma\left(\mathbf{W}^{(\text{MLP1})} \mathbf{H}^{(\text{agg})}\right)\right)\right), \quad (15)$$

式中: $\mathbf{W}^{(\text{MLP1})}$ 和 $\mathbf{W}^{(\text{MLP2})}$ 为可学习权重。

最后, 通过全连接层 $\mathbf{W}^{(\text{out})}$ 生成每个节点的标量得分, 表示为

$$\mathbf{Y} = \sigma'\left(\mathbf{W}^{(\text{out})} \mathbf{Z}_1 + \mathbf{b}^{(\text{out})}\right), \quad (16)$$

式中: $\sigma'(\cdot)$ 可以是二分类的 sigmoid 函数 (或多分类任务的 softmax 函数等); $\mathbf{b}^{(\text{out})}$ 是偏置项。每个节点 i 获得最终预测 $\mathbf{Y}_i$, 完成端到端的架构。

4 结果与分析

本文通过下列的数值计算案例及仿真结果进一步验证前述方法的有效性。

4.1 计算案例与指标

贝塞尔函数: 贝塞尔函数在波动与振动问题中应用广泛, 通常用于描述波的传播、电磁场等现象, 能够很好地描述波从源点向外辐射的行为。第一类贝塞尔函数的积分形式为

$$\mathbf{J}_n(x) = \frac{1}{\pi} \int_0^\pi \cos(n\theta - x \sin \theta) d\theta, \quad (17)$$

式中: n 为阶数。低阶贝塞尔函数通常表现为比较平缓的波动, 适合描述比较简单的波传播模式; 高阶贝塞尔函数通常表现为比较复杂的振荡模式, 适合描述高阶多极子场和复杂的振动模式。具体求解公式

$$f(x) = \mathbf{J}_n(k \cdot r) \cdot \cos(n \cdot \theta), \quad (18)$$

式中: r 为场点 x 与光源位置 x_0 之间的距离; θ 为场点相对于光源的方向角; $\mathbf{J}_n(k \cdot r)$ 为波在径向的传播, k 为波数。随着 $(k \cdot r)$ 的增加, 波模式将表现为波在不同方向上的强度分布。随着 n 阶数的增加, 波的方向性变得更加复杂。

光波导: 模拟光波在矩形波导中的传播, 通过高斯模式分布模拟光的横向模式分布, 结合光的折射率

分布和传播波特特性描述最终的光场分布。

$$E(x, z, t) = n(x) \cdot \exp\left(-\frac{r^2}{w_0^2}\right) \cdot \cos(\beta z - \omega t), \quad (19)$$

式中: $E(x, z, t)$ 为光场的强度分布, 其中 x 是横向坐标, z 是沿着光波传播方向的坐标, t 是时间; $n(x)$ 为波导折射率; w_0 为高斯模腰半径; β 和 ω 为传播常数和角频率。

本文将网格细化问题看作一个节点分类问题, 因此仅通过准确率来评价网络性能是不够准确的。在平均损失和准确率的基础上, 进一步增加了 F1-Score (F_1) 作为新的评价指标, 这在类别不平衡的情况下尤其有用。 F_1 的计算公式为

$$F_1 = \frac{2TP}{2TP + FP + FN}, \quad (20)$$

式中: TP 为模型正确预测为正样本数; FP 为模型错误预测为正样本数 (实际为负样本数); FN 为模型错误预测为负样本数 (实际为正样本数)。 F_1 的取值范围为 $[0, 1]$, 值越接近 1, 模型性能越好。

其他两个指标的计算公式为

$$MSE = \frac{1}{N_{\text{test}}} \sum_{i=1}^{N_{\text{test}}} (y_i - \hat{y}_i)^2, \quad (21)$$

$$Accuracy = \frac{1}{N_{\text{test}}} \sum_{i=1}^{N_{\text{test}}} \mathbb{I}(y_i = \hat{y}_i), \quad (22)$$

式中: MSE 为均方误差; y_i 为第 i 个样本的真实标签; $\hat{y}_i$ 为模型对第 i 个样本的预测值; N_{test} 为测试样本数; $\mathbb{I}(y_i = \hat{y}_i)$ 为指示函数, 当 $y_i = \hat{y}_i$ 时取值为 1, 否则为 0。

另外, 在有限元分析中, 网格质量同样影响数值解的精度和计算稳定性, 本文使用 Gmsh 软件中评估网格质量的三个关键指标对不同网格进行评估^[30]: 最小逆条件数 (SICN) 主要衡量网格的几何条件数, 计算主要涉及单元的雅可比矩阵, 该值趋近于 1 时说明网格变换等距, 数值计算最稳定; Gamma 形状测度, 用于评估网格单元的形状是否接近理想的几何形状, 通常基于单位球体积的归一化形状度量, 其计算方式定义为单元最短边长与单元最长边长的比值, 该值为 1 时, 单元是理想形状; 最小逆梯度误差 (SIGE) 衡量网格单元的梯度误差, 主要用于评估网格的插值误差, 该值为 1 时, 表示网格误差最小。

4.2 消融实验

为了说明 GTF-Net 的融合优势, 本文设计了一组消融实验, 将其与单独使用 TransformerConv 的 TF-Net (Transformer-Net) 和单独使用 GATv2Conv 的 GATv2-Net 进行详细比较, 具体的实验设置见表 2, 且性能评估的关键就是上述三个指标。

从表 3 的数据可以看出, GTF-Net 通过融合

表 2 不同神经网络框架参数设置

Table 2 Parameter setting for different neural network frameworks

Parameter	GTF-Net	TF-Net	GATv2-Net
Feature encoder	Dense	Dense	Dense
Transformer layer	4-layer TransformerConv	8-layer TransformerConv	None
GATv2 layer	4-layer GATv2Conv	None	8-layer GATv2Conv
Skip connections	4 dense skip connections	4 dense skip connections	4 dense skip connections
Layer normalization	8-layer LayerNorm	8-layer LayerNorm	8-layer LayerNorm
Feature aggregation	Dense -> ReLU -> Dropout	Dense -> ReLU -> Dropout	Dense -> ReLU -> Dropout
Classifier	Dense	Dense	Dense
Activation function	ReLU, Sigmoid	ReLU, Sigmoid	ReLU, Sigmoid
Dropout	0.1	0.1	0.1

表 3 GTF-Net 与其他方法的性能比较

Table 3 Performance comparison of GTF-Net with other methods

Network	MSE/%	Accuracy/%	F1-Score/%
GTF-Net	1.52	97.9	96.4
TF-Net	1.86	97.5	95.7
GATv2-Net	3.12	95.9	94.2

Transformer 与 GATv2 的协同架构, 实现了全局特征与局部图结构的高效联合建模。其中, Transformer 层擅长捕捉长距离依赖关系, 确保全局信息的完整性; 而 GATv2 层通过动态调整节点间注意力权重, 增强对局部复杂图模式的适应性。相比 TF-Net 和 GATv2-Net, GTF-Net 的融合框架能够兼顾全局与局部特征, 显著增强模型对图数据中多层次关系的解析能力, 从而在降低预测误差 (如平均损失降低近 50%)、提高分类精度 (准确率达 97.9%) 以及平衡精确率与召回率 (F_1 为 96.4%) 等方面展现出综合优势。

4.3 对比验证

本文使用 GTF-Net 实现了神经网络和训练脚本。批量大小设置为 64, 学习率优化为 0.001, 训练时设置了最大训练轮数和提前停止耐心轮数, 即如果验证集损失在 20 次迭代中没有改善, 则停止训练, 损失函数为均方误差。相应的损失曲线如图 3 所示。

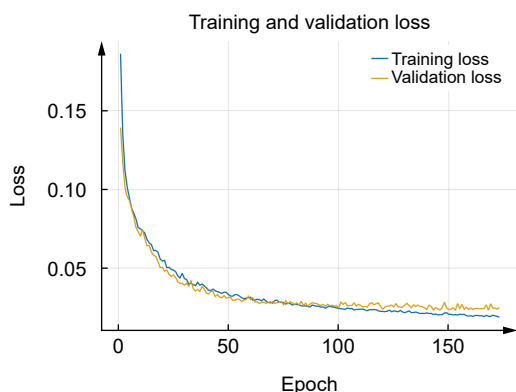


图 3 GTF-Net 的训练和验证损失

Fig. 3 The training and validation loss of GTF-Net

图 4 展示了不同情况下贝塞尔函数的求解及其误差分布。在该实验中, 研究了不同网格方案对贝塞尔函数数值解的影响, 并通过表 4 计算梯度误差标准差以度量其准确性, 计算运行时间以度量其计算效率。

实验对比了初始均匀网格、GTF-Net 网格和 skFem 网格, 后两者采用了不同的自适应网格优化策略, 且网格点数量基本一致。skFem 作为 Python 的开源有限元方法库, 被用于求解偏微分方程, 并在本实验中作为基准方法^[31]。在实验中, 贝塞尔函数的波数 k 被放大 3 倍, 以提高空间频率, 光源位置设定在 (0.5, 0.5), 阶数 $n = 2$ 。结果表明, 相较于 skFem 网格, GTF-Net 网格的梯度误差小了一个数量级, 同时计算时间提升了大约 35.74%。说明 GTF-Net 网格在误差

分布上的离散程度更小, 即误差在网格上的均匀性较高, 同时计算效率更高。

图 5 描述了光在矩形波导中的传播, 波导芯折射率为 1.45, 包层折射率为 1.44, 波导宽度为 $0.2 \mu\text{m}$, 波导中心位置为 (0.5, 0.5)。在波参数中, 波长为 $1.55 \mu\text{m}$, 角频率 ω 为光速除以芯折射率; 传播常数 β 为波数乘以芯折射率。光场分布采用高斯模式, 横模分布用高斯函数描述, 传播特性用余弦函数表示。传播方向沿 x 轴, 时间固定为 $t = 0$ 。

光波导实验虽然同样采用了原始网格、GTF-Net 网格和 skFem 网格进行对比, 但由于光波导方程的解结构具有更强的局部集中性, 其误差分布呈现出不同的特点。在贝塞尔函数实验中, 误差主要分布在高频振荡区域, 而在光波导方程中, 误差集中于光源中心区域的高梯度区域, 且边界区域误差较小。

结合表 4 来看, 原始网格在光波导实验中的表现比贝塞尔实验更差, 因为均匀剖分的网格无法充分解析光波导中心的高频变化, 使得数值解的误差更加集中, 尤其在光源位置附近的梯度误差较大。GTF-Net 网格在两次实验中均通过局部加密优化了网格结构, 但在光波导实验中, 它的效果更明显, 使光波导中心的解析能力得到提升, 同时降低了局部梯度误差。在两个实验中均展现了良好的全局误差控制能力, 但在光波导实验中, 它的优势更为突出, 使整体梯度误差分布更加均匀, 高误差点更少。

同时, 本节以光波导计算为案例, 分析不同网格单元的质量, 参数形式为“平均值 (最小值->最大值)”, 见表 5。

由表 5 可知, 不同网格方案在单元质量分布方面存在一定差异。初始网格的平均单元质量较高, 整体质量分布较为均匀, 但由于节点数量较少, 在计算精度上存在一定的不足。相比之下, skFem 生成的网格在单元质量一致性方面表现优异, 各单元质量值较为均衡, 虽然平均单元质量相对较低, 但这种均匀分布有助于提高计算的稳定性, 减少局部网格过度变形带来的误差累积。

GTF-Net 网格则在提升计算效率的同时, 牺牲了一部分网格质量。从图 6 可知, 相比于初始网格, 部分网格单元的形状质量较低。但这些低质量单元主要分布在网格加密区域的外围, 而加密区域内部的单元质量相对较高, 从而在一定程度上减缓了单元质量下降对计算精度的不利影响。

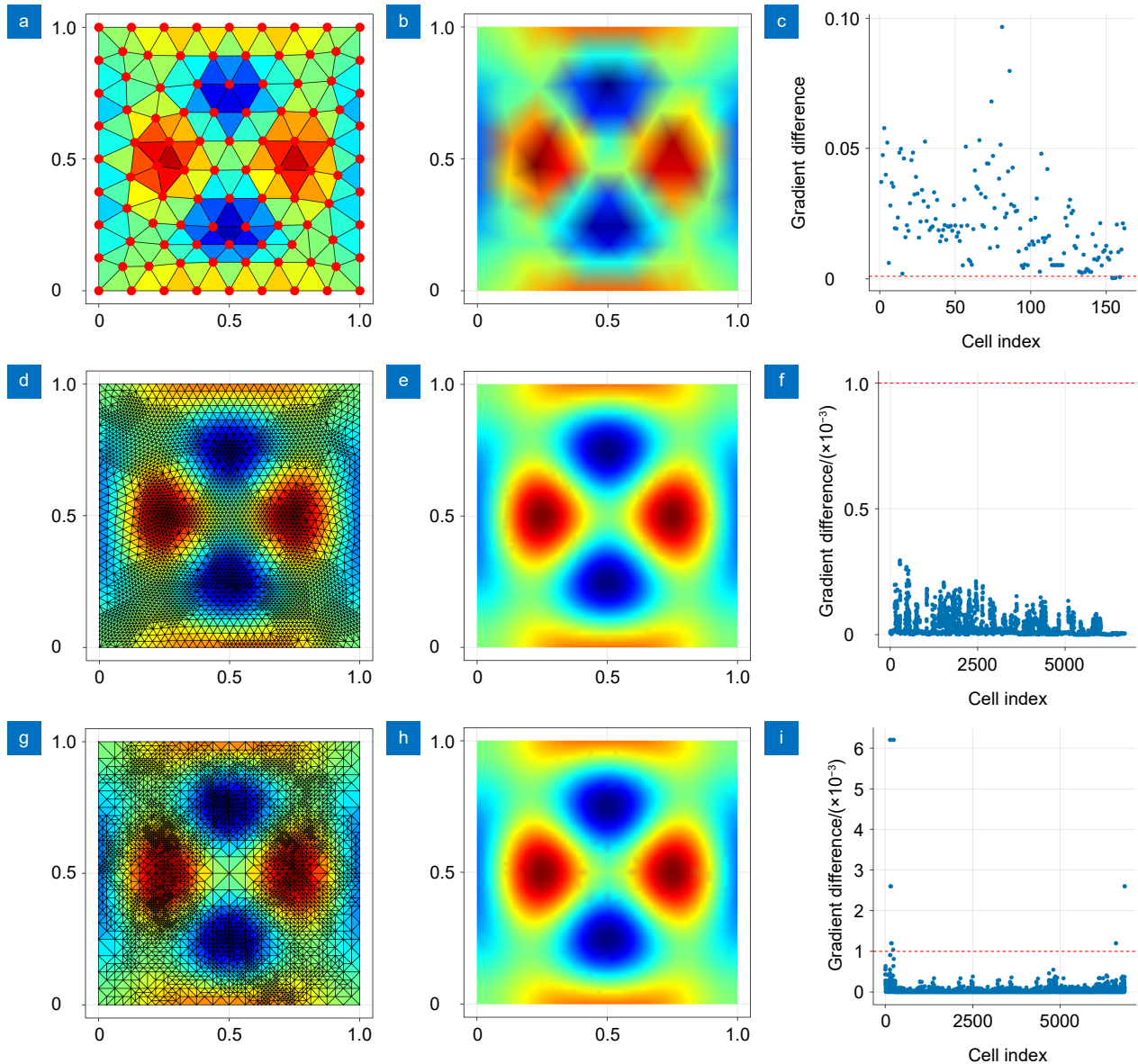


图4 贝塞尔函数。(a) 初始网格; (b) 解场 (初始网格); (c) 梯度误差分布 (初始网格); (d) GTF-Net 网格; (e) 解场 (GTF-Net 网格); (f) 误差梯度分布 (GTF-Net 网格); (g) skFem 网格; (h) 解场 (skFem 网格); (i) 梯度误差分布 (skFem 网格)

Fig. 4 Bessel's equations. (a) Original mesh; (b) Solution field (original mesh); (c) Gradient error distribution (original mesh); (d) GTF-Net mesh; (e) Solution field (GTF-Net mesh); (f) Error gradient distribution (GTF-Net mesh); (g) skFem mesh; (h) Solution field (skFem mesh); (i) Gradient error distribution (skFem mesh)

表4 GTF-Net 网格与其他方法的误差比较

Table 4 Error comparison between GTF-Net Grid and other methods

Mesh	Bessel equations					Original mesh				
	Number of nodes	Number of elements	Standard deviation	Time/s	Ratio/%	Number of nodes	Number of elements	Standard deviation	Time/s	Ratio/%
Original mesh	98	162	0.016	—	—	98	162	0.031	—	—
skFem	3474	6842	2.384×10^{-4}	5.46	—	873	1728	3.379×10^{-5}	6.43	—
GTF-Net	3489	6825	3.353×10^{-5}	3.51	35.74	874	1714	2.573×10^{-5}	2.58	59.91

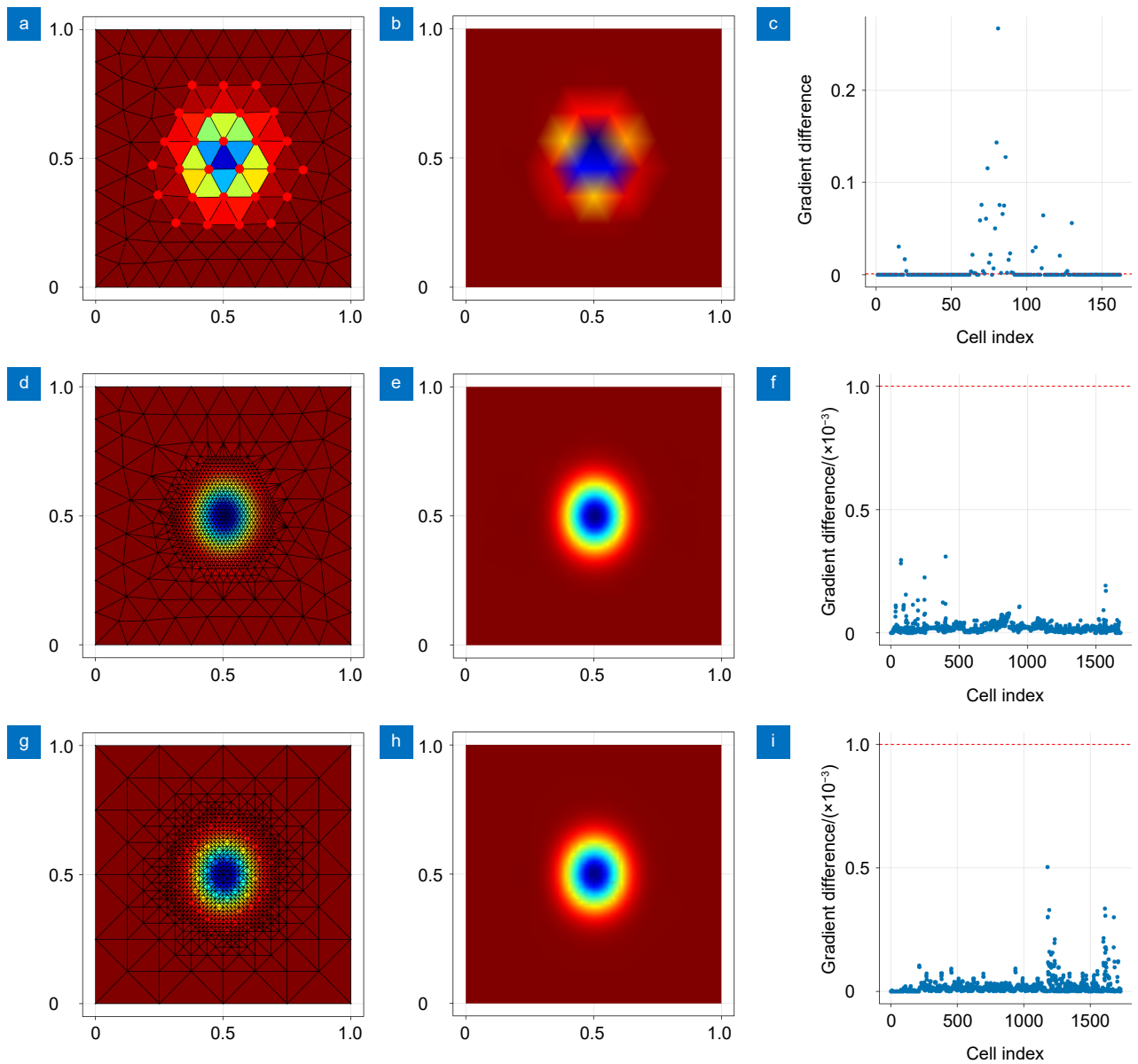


图 5 光波导的求解。(a) 初始网格; (b) 解场 (初始网格); (c) 梯度误差分布 (初始网格); (d) GTF-Net 网格; (e) 解场 (GTF-Net 网格); (f) 误差梯度分布 (GTF-Net 网格); (g) skFem 网格; (h) 解场 (skFem 网格); (i) 梯度误差分布 (skFem 网格)

Fig. 5 Solving optical waveguide. (a) Original mesh; (b) Solution field (original mesh); (c) Gradient error distribution (original mesh); (d) GTF-Net mesh; (e) Solution field (GTF-Net mesh); (f) Error gradient distribution (GTF-Net mesh); (g) skFem mesh; (h) Solution field (skFem mesh); (i) Gradient error distribution (skFem mesh)

表 5 不同网格的单元质量分析

Table 5 Element quality analysis of different meshes

Quality parameter	Original mesh	GTF-Net mesh	skFem mesh
SICN	0.9736 (0.8311->1)	0.9644 (0.362->1)	0.866 (0.866->0.866)
Gamma	0.9685 (0.7812->1)	0.9622 (0.3179->1)	0.8284 (0.8284->0.8284)
SIGE	0.9884 (0.9068->1)	0.9874 (0.6281->1)	0.9292 (0.9292->0.9292)

5 结 论

本文针对有限元求解中自适应网格划分问题，

提出了一种基于注意力融合机制的自适应网格细化框架——GATv2-Transformer 融合网络 (GTF-Net)，该模型将图注意力机制与 Transformer 架构相结合，构

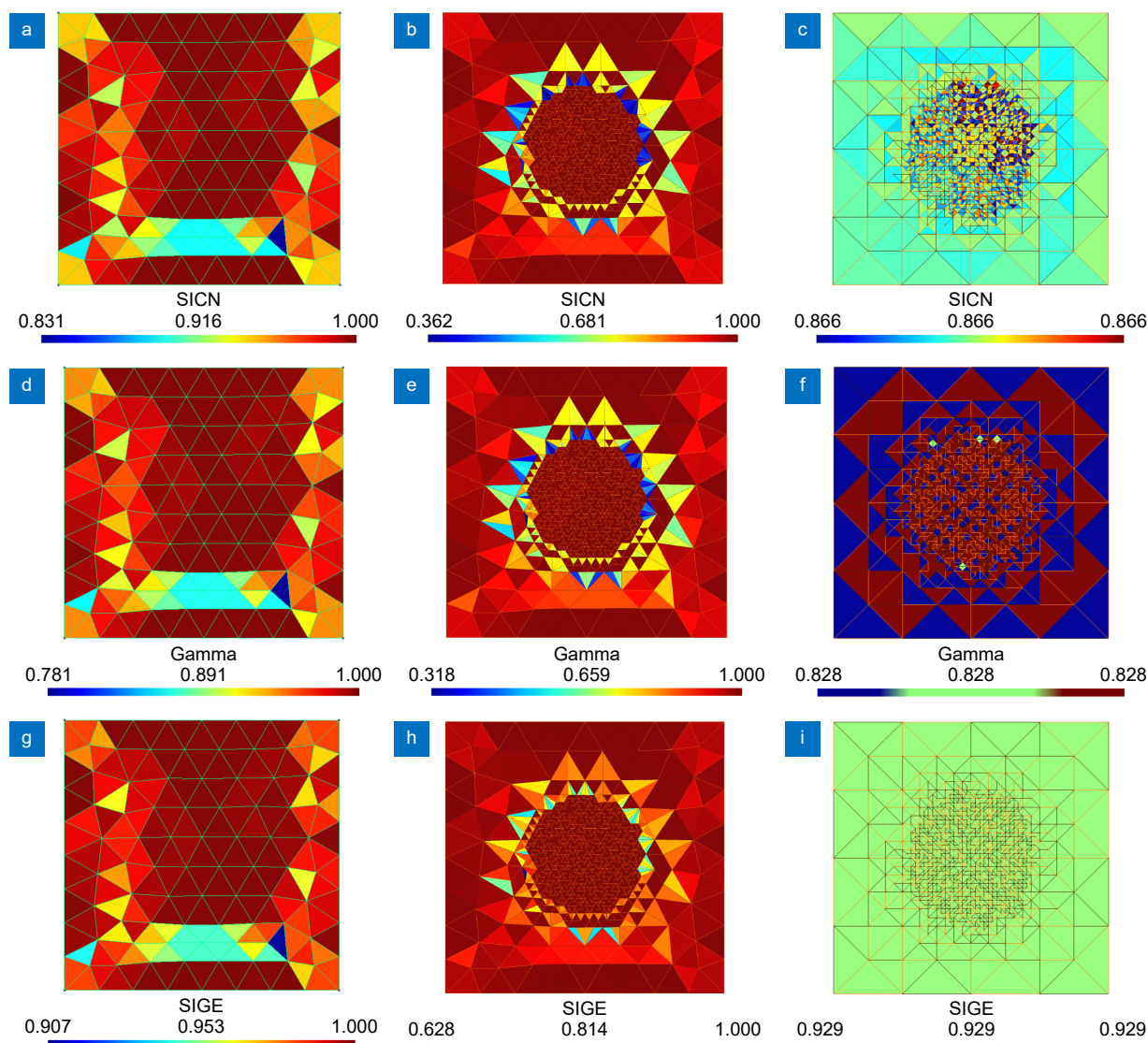


图 6 不同网格单元质量分布图。(a) 初始网格单元 SICN 值分布; (b) GTF-Net 网格单元 SICN 值分布; (c) skFem 网格单元 SICN 值分布; (d) 初始网格单元 Gamma 值分布; (e) GTF-Net 网格单元 Gamma 值分布; (f) skFem 网格单元 Gamma 值分布; (g) 初始网格单元 SIGE 值分布; (h) GTF-Net 网格单元 SIGE 值分布; (i) skFem 网格单元 SIGE 值分布

Fig. 6 Quality distribution of different mesh cells. (a) SICN value distribution of original mesh cells; (b) SICN value distribution of GTF-Net mesh cells; (c) SICN value distribution of skFem mesh cells; (d) Gamma value distribution of original mesh cells; (e) Gamma value distribution of GTF-Net mesh cells; (f) Gamma value distribution of skFem mesh cells; (g) SIGE value distribution of original mesh cells; (h) SIGE value distribution of GTF-Net mesh cells; (i) SIGE value distribution of skFem mesh cells

建了局部几何特征与全局物理场的动态耦合框架, 旨在提高自适应网格划分的质量。实验结果表明, 相较于传统迭代方法, GTF-Net 在贝塞尔函数与光波导案例中的梯度误差标准差分别降低 23.8% 和 85.9%, 同时计算耗时分别降低 35.74% 和 59.91%, 验证了其在高精度与高效率协同优化方面的显著优势。未来的研究方向将重点突破三维非规则域网格生成技术、多物理场耦合建模等核心关键技术, 并结合迁移学习框架, 进一步优化模型的泛化能力与适用范围。

利益冲突: 所有作者声明无利益冲突

参考文献

- [1] Pasciak J E. The mathematical theory of finite element methods (Susanne C. Brenner and L. Ridgway Scott)[J]. *SIAM Rev*, 1995, **37**(3): 472–473.
- [2] Salminen T, Lehtinen K E J, Seppänen A. Application of the finite element method to the multicomponent general dynamic equation of aerosols[J]. *J Aerosol Sci*, 2023, **174**: 106260.
- [3] Barth T, Ohlberger M. Finite volume methods: foundation and analysis[M]//Stein E, de Borst R, Hughes T J R. *Encyclopedia of Computational Mechanics*. Chichester: John Wiley, 2004: 1–60. <https://doi.org/10.1002/0470091355.ecm010>.
- [4] Berger M J, Oliger J. Adaptive mesh refinement for hyperbolic

- partial differential equations[J]. *J Comput Phys*, 1984, **53**(3): 484–512.
- [5] Perera R, Agrawal V. Dynamic and adaptive mesh-based graph neural network framework for simulating displacement and crack fields in phase field models[J]. *Mech Mater*, 2023, **186**: 104789.
- [6] Zhao Y X, Li H R, Zhou H S, et al. A review of graph neural network applications in mechanics-related domains[J]. *Artif Intell Rev*, 2024, **57**(11): 315.
- [7] Economon T D, Palacios F, Copeland S R, et al. SU2: an open-source suite for multiphysics simulation and design[J]. *AIAA J*, 2016, **54**(3): 828–846.
- [8] Marburg S, Nolte B. *Computational Acoustics of Noise Propagation in Fluids-Finite and Boundary Element Methods*[M]. Berlin, Heidelberg: Springer, 2008. <https://doi.org/10.1007/978-3-540-77448-8>.
- [9] Verfürth R. A posteriori error estimation and adaptive mesh-refinement techniques[J]. *J Comput Appl Math*, 1994, **50**(1-3): 67–83.
- [10] Xie L J, Chen J J, Liang Y, et al. Geometry-based adaptive mesh generation for continuous and discrete parametric surfaces[J]. *J Inf Comput Sci*, 2012, **9**(8): 2327–2344.
- [11] Möller M, Kuzmin D. Adaptive mesh refinement for high-resolution finite element schemes[J]. *Int J Numer Methods Fluids*, 2006, **52**(5): 545–569.
- [12] Triantafyllidis D G, Labridis D P. A finite-element mesh generator based on growing neural networks[J]. *IEEE Trans Neural Netw*, 2002, **13**(6): 1482–1496.
- [13] Jilani H, Bahreininejad A, Ahmadi M T. Adaptive finite element mesh triangulation using self-organizing neural networks[J]. *Adv Eng Softw*, 2009, **40**(11): 1097–1103.
- [14] Bohn J, Feischl M. Recurrent neural networks as optimal mesh refinement strategies[J]. *Comput Math Appl*, 2021, **97**: 61–76.
- [15] Zhang Z Y, Wang Y X, Jimack P K, et al. MeshingNet: a new mesh generation method based on deep learning[C]// *Proceedings of the 20th International Conference on Computational Science*, Amsterdam, 2020: 186–198. https://doi.org/10.1007/978-3-030-50420-5_14.
- [16] Wu Z H, Pan S R, Chen F W, et al. A comprehensive survey on graph neural networks[J]. *IEEE Trans Neural Netw Learn Syst*, 2020, **32**(1): 4–24.
- [17] Zhang L Y, Sun H H, Sun Y F, et al. Review of node classification methods based on graph convolutional neural networks[J]. *Comput Sci*, 2024, **51**(4): 95–105. 张丽英, 孙海航, 孙玉发, 等. 基于图卷积神经网络的节点分类方法研究综述[J]. *计算机科学*, 2024, **51**(4): 95–105.
- [18] Kim M, Lee J, Kim J. GMR-Net: GCN-based mesh refinement framework for elliptic PDE problems[J]. *Eng Comput*, 2023, **39**(5): 3721–3737.
- [19] Khan A, Yamada M, Chikane A, et al. GraphMesh: geometrically generalized mesh refinement using GNNs[C]// *Proceedings of the 24th International Conference on Computational Science*, Malaga, 2024: 120–134. https://doi.org/10.1007/978-3-031-63775-9_9.
- [20] Pelissier U, Parret-Fréaud A, Bordeu F, et al. Graph neural networks for mesh generation and adaptation in structural and fluid mechanics[J]. *Mathematics*, 2024, **12**(18): 2933.
- [21] Peng J M, Chen X H, Liu J. 3DMeshNet: a three-dimensional differential neural network for structured mesh generation[J]. *Graph Models*, 2025, **139**: 101257.
- [22] Rowbottom J, Maierhofer G, Deveney T, et al. G-adaptivity: optimised graph-based mesh relocation for finite element methods[Z]. arXiv: 2407.04516, 2024.
- [23] Freymuth N, Dahlinger P, Würth T, et al. Swarm reinforcement learning for adaptive mesh refinement[C]// *Proceedings of the 37th International Conference on Neural Information Processing Systems*, New Orleans, 2024: 3206.
- [24] Liang L M, Dong X, Li R J, et al. Classification algorithm of retinopathy based on attention mechanism and multi feature fusion[J]. *Opto-Electron Eng*, 2023, **50**(1): 220199. 梁礼明, 董信, 李仁杰, 等. 基于注意力机制多特征融合的视网膜病变分级算法[J]. *光电工程*, 2023, **50**(1): 220199.
- [25] Zhang H M, Tian Q Q, Yan D D, et al. GLCrowd: a weakly supervised global-local attention model for congested crowd counting[J]. *Opto-Electron Eng*, 2024, **51**(10): 240174. 张红民, 田钱前, 颜鼎鼎, 等. GLCrowd: 基于全局-局部注意力的弱监督密集场景人群计数模型[J]. *光电工程*, 2024, **51**(10): 240174.
- [26] Scarselli F, Gori M, Tsoi A C, et al. The graph neural network model[J]. *IEEE Trans Neural Netw*, 2009, **20**(1): 61–80.
- [27] Shang Y M, Wu A B, Yuan Y, et al. Graph neural network enhancement based on personalized PageRank high-order neighborhood aggregation[J]. *Comput Eng*, 2024. 商雅名, 吴安彪, 袁野, 等. 基于个性化 PageRank 高阶邻域聚合的图神经网络增强[J]. *计算机工程*, 2024.
- [28] Brody S, Alon U, Yahav E. How attentive are graph attention networks?[C]// *Proceedings of the 10th International Conference on Learning Representations*, 2022: 1–26.
- [29] Yun S, Jeong M, Kim R, et al. Graph transformer networks[C]// *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, 2019: 1073.
- [30] Geuzaine C, Remacle J F. Gmsh: a 3 - D finite element mesh generator with built - in pre - and post - processing facilities[J]. *Int J Numer Methods Eng*, 2009, **79**(11): 1309–1331.
- [31] Gustafsson T, McBain G D. scikit-fem: a Python package for finite element assembly[J]. *J Open Source Softw*, 2020, **5**(52): 2369.

作者简介



韩婷 (1999-), 女, 硕士研究生, 主要从事光子集成设计方面的研究。

E-mail: hanting19991020@163.com



【通信作者】叶佳 (1985-), 男, 博士, 教授, 博士生导师, 主要从事光电子领域相关研究工作, 包括面向毫米波通信的光纤前传技术、基于深度学习的微波信号感知技术、AI 辅助的光电系统智能设计、微波光子链路仿真软件开发等。

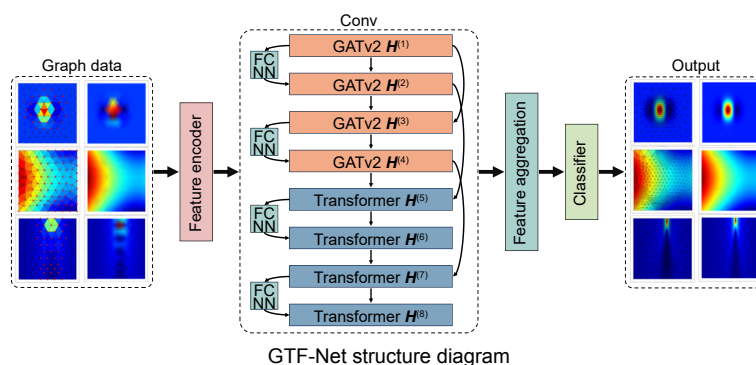
E-mail: jiaye@home.swjtu.edu.cn



扫描二维码, 获取PDF全文

Adaptive mesh partitioning for graph attention Transformer networks

Han Ting, Ye Jia*, Yan Lianshan, Gan Zongxin



Overview: In the fields of engineering and computational science, finite element analysis is a commonly used numerical simulation tool. Its accuracy and efficiency directly affect the reliability and computational cost of engineering design. However, traditional adaptive mesh refinement technology faces certain challenges in the pursuit of high-precision and high-efficiency collaborative optimization. Especially when dealing with engineering problems with singular fields or complex boundary conditions, traditional iterative methods often show the problems such as uneven gradient error distribution and slow convergence. These limitations not only affect the accuracy of the calculation results, but also limit the application of finite element analysis in complex problems. To address the above problems, this study proposes an adaptive mesh partitioning framework based on the attention fusion mechanism, namely GATv2-Transformer fusion network (GTF-Net). This method transforms the mesh partitioning problem into a node classification problem. Each mesh node is regarded as a node in the graph, and the edges between nodes represent the relationship between units. The relationship between mesh units is modeled using graph neural networks, thereby achieving adaptive adjustment of mesh partitioning. Graph neural networks automatically adjust the mesh structure by learning these relationships. The network innovatively combines the graph attention mechanism with the Transformer architecture, and realizes the dynamic coupling of local geometric features and global physical fields through multi-head cross attention modules, effectively improving the representation ability of complex environments. The analytical solution of multiple equations is introduced into the network training, and a multi-task learning objective is constructed to ensure the generalization performance of the model under different physical field characteristics. The typical optical waveguide transmission problem example and the solution results of the first-kind Bessel function show that compared with the traditional skFem method, GTF-Net has improved the calculation speed while reducing the standard deviation of the gradient error by more than 20% (the Bessel function case is reduced by 23.8%, and the optical waveguide case is reduced by 85.9%). The experimental results show that the network significantly improves the matching degree between the grid density distribution and the physical field changes through nonlinear mapping of the feature space, and the method has a certain generalization ability and can adapt to different types of problems and application scenarios. This method provides a new deep learning solution for adaptive finite element analysis in engineering calculations, and also opens up a new technical path for the development of data-driven intelligent finite element analysis.

Han T, Ye J, Yan L S, et al. Adaptive mesh partitioning for graph attention Transformer networks[J]. *Opto-Electron Eng*, 2025, 52(4): 250058; DOI: 10.12086/oee.2025.250058

Foundation item: National Key Research and Development Program of China (2022YFB2802701), National Natural Science Foundation of China (U23A20376, 62075185, 62271422), and Sichuan Science Fund for Distinguished Young Scholars (2024NSFJQ0016)

School of Information Science and Technology, Southwest Jiaotong University, Chengdu, Sichuan 611756, China

* E-mail: jiaye@home.swjtu.edu.cn