

# 图神经网络在计算机视觉任务中的拓展研究

叶明壕

华北电力大学 北京 102206

**摘要:** 随着计算机视觉任务的复杂度日益增加, 传统的卷积神经网络 (CNN) 逐渐面临难以处理的挑战。近年来, 图神经网络 (GNN) 作为一种处理图结构数据的强大工具, 在计算机视觉领域展示了巨大的潜力。图神经网络能够有效地捕捉图像中的空间和上下文关系, 这在图像分类、目标检测、实例分割等任务中具有显著优势。本文探讨了图神经网络在计算机视觉中的应用, 并针对其拓展方向进行了详细分析, 包括跨模态学习、自监督学习以及高效架构设计等。此外, 本文还讨论了图神经网络在视觉任务中的挑战与前景, 为未来的研究方向提供了新的思路。

**关键词:** 图神经网络; 计算机; 视觉任务; 拓展

## 引言

计算机视觉作为人工智能的一个重要分支, 已经在许多领域取得了显著的成就。传统的视觉任务, 如图像分类、目标检测和实例分割, 通常依赖于卷积神经网络 (CNN) 来从图像中提取特征。然而, 随着任务复杂度的提高, CNN 在处理一些具有非欧几里得结构的数据时表现出一定的局限性。这类数据例如图像中的结构化信息、对象间的关系等, 无法通过传统的卷积操作有效捕捉<sup>[1]</sup>。图神经网络 (GNN) 作为一种专门处理图数据的深度学习模型, 其能够在图结构数据中有效地进行信息传播和学习, 因此在计算机视觉任务中展现了广阔的应用前景。图神经网络通过建模图中的节点和边之间的关系, 可以更好地处理复杂的空间和结构化信息<sup>[2]</sup>。这为解决计算机视觉中的多任务和多模态问题提供了新的思路。特别是在图像和视频分析、视觉问答、图像生成等任务中, GNN 能够对图像中的不同对象及其关系进行建模, 从而提高任务的性能和精度。本文旨在深入探讨图神经网络在计算机视觉中的应用现状、挑战以及拓展研究方向, 并提出未来可能的研究路径。

## 1 图神经网络基础

### 1.1 图神经网络概述

图神经网络 (GNN) 是一类专门用于处理图结构数据的深度学习模型。与传统的卷积神经网络 (CNN) 不同, GNN 能够在节点之间传递信息, 从而学习图中节点、边的特征<sup>[3]</sup>。GNN 通过一系列的信息传播和聚合操作, 使得每个节点能够通过其邻居节点的信息来更新自身的表示。根据不同的应

用场景, GNN 模型有多种变种, 包括图卷积网络 (GCN)、图注意力网络 (GAT) 等, 它们在不同的计算机视觉任务中有着广泛应用<sup>[4]</sup>。

### 1.2 图神经网络的核心组件

图神经网络的基本构件包括节点、边和图结构。节点通常表示图中的对象, 边表示对象之间的关系。GNN 通过以下几种主要组件实现其功能:

**消息传递机制:** GNN 的核心是消息传递机制, 节点会通过边接收来自邻居节点的信息, 并根据一定的规则进行信息聚合<sup>[5]</sup>。

**聚合与更新操作:** 聚合操作通常采用加权平均、最大池化等方式将邻居节点的信息合并, 并通过非线性激活函数进行更新。

**图卷积与图注意力:** 图卷积网络 (GCN) 通过固定的邻接矩阵进行图卷积运算, 而图注意力网络 (GAT) 则通过自注意力机制来赋予每条边不同的权重<sup>[6]</sup>。

### 1.3 图神经网络的优化与训练

GNN 的训练过程与其他神经网络类似, 采用反向传播算法进行优化。其损失函数通常根据任务类型 (如分类、回归等) 进行设计, 并通过梯度下降法进行优化。由于图数据的稀疏性和图规模的庞大, GNN 的训练面临着计算效率和内存消耗等挑战<sup>[7]</sup>。近年来, 研究者提出了一些优化方法, 如采样技术、图剪枝、图卷积优化等, 以提高 GNN 的训练效率。

## 2 图神经网络在计算机视觉中的应用

### 2.1 图像分类

图像分类是计算机视觉中的经典任务，其目标是根据输入图像预测该图像属于哪个类别。传统的卷积神经网络（CNN）在此任务中取得了巨大成功，但当图像中包含复杂的空间关系和长距离依赖时，CNN 的性能会受到一定限制。图神经网络能够通过节点和边来表示图像中的不同区域及其关系，从而捕捉到更多的上下文信息。例如，将图像划分为多个区域，每个区域作为一个节点，区域之间的关系则通过图的边来表示。GNN 通过其信息传递机制，有效地提高了图像分类的精度。

### 2.2 目标检测与实例分割

目标检测和实例分割任务要求模型不仅要识别图像中的目标物体，还要精确定位它们的边界。GNN 在这类任务中的应用，通常是通过构建图模型来表示物体之间的空间关系，并通过图神经网络来增强物体检测的准确性<sup>[8]</sup>。例如，可以将图像中的物体区域建模为图的节点，物体之间的相对位置和关系作为边，通过 GNN 实现信息的传播与整合，从而改善检测和分割的效果。

### 2.3 视觉问答（VisualQuestionAnswering,VQA）

视觉问答任务要求模型根据给定的图像和问题，生成一个合理的答案。图神经网络在视觉问答中的应用主要集中在图像和文本信息的多模态融合。通过将图像中的物体和场景关系建模为图，并利用图神经网络对图像进行深度学习，可以为问题解答提供更加丰富的上下文信息。图神经网络能够有效地处理图像中的结构化信息并与文本信息进行结合，提高了 VQA 任务的准确性和鲁棒性。

## 3 图神经网络在计算机视觉中的拓展研究方向

### 3.1 跨模态学习

跨模态学习是指通过联合学习多个模态（如图像、文本、语音等）之间的关系，解决多模态任务的一种方法。在计算机视觉中，跨模态学习的目标是通过结合不同模态的信息来提升任务的表现，尤其是图像与文本的联合建模。例如，视觉语言联合模型可以通过图神经网络（GNN）来实现图像和文本之间的深度融合。在这种模型中，图像的不同区域或物体被建模为图中的节点，而它们之间的关系则通过边来表示。同时，文本信息也可以通过将词语或句子表示为图节点与图像节点进行对接。这种图结构的表示使得不同模态之间

的信息得以高效地共享与传播，尤其是在多模态任务如视觉问答（VQA）和图像字幕生成（ImageCaptioning）中，能够有效提升模型的理解与推理能力。尽管跨模态学习在计算机视觉中取得了一定进展，但仍然面临诸多挑战。不同模态数据的异质性是跨模态学习中的一大难题，例如，图像数据通常是连续的像素信息，而文本则是离散的词汇或句子。如何设计一个高效的模型来有效融合这些不同的数据，并学习到其潜在的语义关联，仍是一个开放的问题。当前研究正在尝试通过图神经网络来解决这个问题，因为 GNN 天生具备处理异构数据和图结构信息的能力，从而为跨模态学习提供了新的可能性。

### 3.2 自监督学习与图神经网络

自监督学习是一种无需依赖人工标注数据的学习方法，通过设计代理任务来让模型从无标签数据中学习有效的特征。在计算机视觉中，自监督学习已经被广泛应用于图像和视频的表示学习，并且在图神经网络（GNN）的框架下也展现出了极大的潜力。通过自监督学习，GNN 可以从图中的节点关系中提取出有用的信息，而无需依赖大量的标注数据。比如，图神经网络可以通过对节点之间的相似性进行建模，来学习图的结构信息，这不仅有助于提高图的表示能力，还能够提升模型在没有标签的情况下的表现。在自监督学习中，图中的结构信息和节点的内在关系常常作为自监督信号，来引导模型学习图数据的潜在模式。例如，图神经网络可以通过预测节点之间的边或推测节点的属性来进行自监督学习，这使得图神经网络能够在无标签数据上进行有效的训练。自监督学习与图神经网络的结合，尤其在视觉任务中，如目标检测和图像分类任务中，提供了一个新的解决方案，有效解决了标注数据不足的问题。尽管自监督学习在图神经网络中的应用潜力巨大，但仍存在一些挑战，如设计合适的代理任务和生成有效的标签。此外，自监督学习模型的泛化能力和稳定性也是未来研究的重点。

### 3.3 更高效的图神经网络架构

图神经网络（GNN）作为一种强大的工具，能够处理复杂的图结构数据，但其计算复杂度和推理效率问题仍然是制约其应用的关键瓶颈。图数据往往稀疏且规模庞大，这使得在训练和推理过程中需要大量的计算资源和存储空间，特别是在处理大型图或高维图数据时，GNN 的计算开销显著增加。为了解决这些问题，研究者们提出了一些高效的 GNN

架构和优化方法。图卷积网络 (GCN) 和图注意力网络 (GAT) 是最常见的图神经网络架构, 然而它们在处理大规模图时依然面临效率瓶颈。为提高效率, 近年来的研究主要集中在图采样、节点剪枝和图池化等技术上。例如, 通过采样技术, GNN 可以在每轮迭代中只选择部分邻居节点进行信息传递, 从而有效减少计算量。节点剪枝则通过去除对模型训练贡献不大的节点或边, 进一步优化了计算效率。此外, 异构图神经网络 (HGNN) 和图池化技术也被提出, 利用图中不同类型的节点和边的特性, 设计出更加高效的模型。这些新架构和优化方法的提出, 为 GNN 在实际应用中的推广提供了重要的理论支持和技术基础。在未来, 随着计算资源的不断发展, 如何进一步优化 GNN 架构, 提升其在大规模数据集上的训练和推理效率, 仍然是一个需要重点解决的问题。

### 3.4 图神经网络的可解释性与鲁棒性

图神经网络 (GNN) 因其强大的表达能力和灵活的结构建模能力, 在计算机视觉和其他领域取得了显著的成果。然而, GNN 的可解释性和鲁棒性问题仍然是当前研究中的一个热点。由于 GNN 通过节点间的信息传递和聚合操作来学习图的结构信息, 其决策过程往往较为复杂, 难以直观地理解和解释。这使得在一些实际应用中, 特别是在医疗、金融等对模型解释要求较高的领域, GNN 的黑箱特性成为了一个重要的挑战。为了解决这个问题, 研究者们提出了多种可解释性方法, 其中包括基于注意力机制的可解释性方法和图分析技术。这些方法通过分析模型在决策过程中的关键节点和边, 帮助我们理解模型的决策依据。例如, 利用图注意力网络 (GAT) 中的注意力机制, 可以揭示哪些节点和边对最终的预测结果有重要贡献, 从而为 GNN 提供一定的可解释性。同时, 如何提高 GNN 在处理噪声数据或遭受对抗攻击时的鲁棒性, 也是当前的研究重点。GNN 通常对输入数据的微小扰动较为敏感, 因此提升其鲁棒性, 避免在对抗攻击下失效, 成为了一个亟待解决的问题。研究者们正在探索通过正则化方法、数据增强和对抗训练等技术来增强 GNN 的鲁棒性, 以提高其在不确定环境中的表现。未来, 如何在保证模型性能的同时, 提升其可解释性和鲁棒性, 将是 GNN 研究的重要方向之一。

## 4 结论

图神经网络 (GNN) 作为一种革命性的深度学习方法,

已经在计算机视觉任务中展示出了显著的优势和广泛的应用前景。本文回顾了 GNN 的基本原理, 并详细探讨了其在图像分类、目标检测、实例分割、视觉问答等任务中的成功应用。同时, 本文还深入分析了 GNN 在计算机视觉领域中的拓展研究方向, 包括跨模态学习、自监督学习、高效架构设计以及可解释性和鲁棒性等问题。尽管图神经网络在处理图数据的能力上取得了显著进展, 但其计算复杂度、模型解释性和鲁棒性等问题依然存在, 这也是未来研究的重点。随着技术的不断发展, 图神经网络将逐步克服这些挑战, 并在更多实际应用中发挥其巨大的潜力。通过不断优化现有方法、提出新的算法和架构, 图神经网络将在计算机视觉和其他领域继续拓展其应用边界。

### 参考文献:

- [1] 仵赛飞, 张渊, 谢迪, 等. 面向视觉目标检测的脉冲神经网络综述: 从生物机制到前沿应用 [J]. 功能材料与器件学报, 2025, 31(05): 364378.
- [2] 周浩艇, 房川凯, 刘稔远, 等. 基于运动视觉与视差协同计算的逼近感知神经网络 [J]. 电子学报, 116[2025]1126.
- [3] 汪认, 崔云龙, 何建英, 等. 基于机器视觉和深度学习神经网络的焊缝质量智能评价技术研究 [J]. 电焊机, 2025, 55(10): 7785.
- [4] 张鸿儒, 吴小蝶, 蔡雄友, 等. 面向视觉识别的神经网络结构化剪枝方法研究 [J]. 科学技术创新, 2025, (19): 135138.
- [5] 刘志阳. 融合视觉-力觉感知与深度神经网络的机器人精密装配力控制算法 [J]. 中国机械, 2025, (24): 2023.
- [6] 李欢, 李亮. 基于卷积神经网络与机器视觉的纸张尘埃度测量系统的设计与应用研究 [J]. 中国造纸, 2025, 44(08): 157163.
- [7] 杨旭涛, 秦进, 胡滨. 生物启发的无人机航拍前景提取视觉神经网络 [J]. 计算机工程与设计, 2025, 46(08): 22892296.
- [8] 白依宁, 杨松. 基于图神经网络的图像情感分析研究 [J]. 智能计算机与应用, 2025, 15(08): 197203.

作者简介: 叶明壕 (2000.06—), 男, 汉族, 福建省三明市, 硕士, 学生, 研究方向: 人工智能, 神经网络。