

基于图卷积神经网络的小尺寸图像分类算法

尹冠千*

December 11, 2024

摘要

近年来,深度学习技术在图像识别领域取得了显著进展,从经典卷积神经网络(CNN)再到轻量化模型和迁移学习,它们大幅提升了图像分类与目标检测的精度与效率。然而,小尺寸图像分类由于分辨率低、信息量有限、易受噪声干扰还存在一些挑战。针对这一问题,本文探索了图卷积网络(GCN)在小尺寸图像分类任务中的潜力:通过构建图结构模型处理低分辨率图像分类问题。本研究提出了一种基于GCN的小尺寸图像识别算法,通过Fashion MNIST数据集的实验验证了其可行性。实验结果表明,在 7×7 像素的分类任务中GCN的准确率为64.25%。本文的研究表明,GCN在超小尺寸图像分类和视角变形任务中具有广泛的应用前景,为相关领域的研究提供了新的方向。

关键词: 图卷积神经网络; 深度学习; 图像识别

Abstract

In recent years, deep learning technologies have achieved remarkable progress in the field of image recognition. From classical Convolutional Neural Networks (CNNs) to lightweight models and transfer learning, these advancements have significantly improved the accuracy and efficiency of image classification and object detection tasks. However, challenges remain in the classification of small-size images due to their low resolution, limited information, and susceptibility to noise.

To address this issue, this paper explores the potential of Graph Convolutional Networks (GCNs) for tiny-size image classification by leveraging graph-structured models to handle low-resolution image classification tasks. A novel GCN-based algorithm for small-size image recognition is proposed and validated through experiments on the Fashion MNIST dataset. Experimental results demonstrate the feasibility of the approach, achieving an accuracy of 64.25% on a 7×7 pixel classification task.

This study highlights the broad application prospects of GCNs in ultra-small image classification and tasks involving geometric transformations, offering new directions for research in related fields.

Keywords: Graph Convolutional Networks, Deep Learning, Image Recognition

1 引言

近年来,随着深度学习技术的迅猛发展,图像识别领域取得了突破性进展。从经典的卷积神经网络(CNN)到轻量化模型和迁移学习技术,图像分类、目标检测等任务的精度和效率都得到了显著提升。然而,在小尺寸图像识别任务中,由于图像分辨率低、信息量有限、噪声影响显著,传统的CNN方法在特征提取和类别区分方面仍面临诸多挑战。针对这些难点,现有研究尝试通

*邮箱: gqyin@udel.edu

过超分辨率技术、数据增强以及轻量化模型优化等手段来提升小尺寸图像分类性能，但在全局信息建模和复杂关系捕获方面仍然存在局限性^[1]。

在图像识别的早期阶段，传统方法主要依赖于手工设计的特征提取和机器学习算法。这些方法通常包括边缘检测、颜色直方图、纹理分析和形状分析等。这些特征提取方法基于人类设计的规则，将图像转化为特征向量，再通过分类器（如支持向量机（SVM）、k 近邻（KNN）、随机森林等）进行分类。边缘检测方法（如 Sobel、Canny 边缘检测）通过捕捉图像中的轮廓信息来区分不同的图像类别^[2]。尽管这些方法在一定程度上取得了成功，但它们存在明显的局限性。手工设计的特征难以捕捉图像中的复杂模式，且对图像的光照变化、变形和遮挡等情况敏感。此外，这些方法对特征提取和分类器的选择和参数调整要求较高，无法自动适应数据变化。近年来，随着计算能力的提升和大规模标注数据的获取，深度学习，尤其是卷积神经网络（CNN），在服装图像识别领域取得了突破性进展。CNN 通过多层卷积和池化操作，从图像中自动提取多尺度、多层次的特征，极大地提高了图像分类的准确性和鲁棒性^[3]。在图像识别中，CNN 可以自动学习到复杂的纹理、形状和颜色特征，具有很强的泛化能力。例如，著名的 AlexNet、VGG、ResNet 等网络架构在 ImageNet 等大规模数据集上展示了卓越的性能^[4]。这些网络结构通过逐层学习，从底层的边缘和纹理特征到高层的语义信息，逐步提高了图像识别的能力。

图神经网络（Graph Neural Network, GNN）是一类用于处理图结构数据的深度学习模型。图结构数据包括节点、边和节点之间的关系，常见于社交网络、生物网络、交通网络等场景。在图像识别中，GNN 通过将图像像素或超像素作为节点，构建出反映像素关系的图结构，能够更好地捕捉图像中的局部和全局信息^[5]。通过将图像转化为图结构数据，GNN 可以有效地建模各部分之间的关系，捕捉到更丰富的特征信息^[6]。其次，kanyazev^[7] 等人的研究表明 GNN 具有良好的泛化能力。由于它能够学习到数据的内在结构和关系，GNN 在面对新数据时可以更准确地进行预测。GNN 已经被广泛应用于各种领域。在推荐系统中，wu^[8] 等人构建 GNN 构建用户-物品的二分图，基于用户行为和物品特征进行推荐。在生物信息学中，Jha^[9] 等人用 GNN 预测蛋白质相互作用和分子特性分析。在社交网络分析中，GNN 可以用于社区检测、节点分类和连接预测^[10]。GNN 在图像超分辨率任务中可以通过构建图结构来捕捉像素之间的复杂关系，从而更精确地重建高分辨率图像。例如，ESRGAN^[11] 等方法使用 GNN 来增强图像的细节和纹理。

GNN 的一个典型例子是图卷积网络（Graph Convolutional Network, GCN），它通过卷积操作对图进行过滤和特征提取，类似于 CNN 在图像上进行的卷积操作^[12]。另一种常见的 GNN 是图注意力网络（GAT），它使用注意力机制动态地调整不同邻居节点的影响权重^[13]。本研究用到的 GCN 图卷积网络通过在图上定义的卷积操作，有效地从节点及其邻居节点的关系中提取特征，广泛应用于节点分类、图分类、边预测等任务。例如，在在社交网络中预测用户类别或兴趣^[14]，化学领域对分子图分类和预测^[15]。图卷积网络（Graph Convolutional Networks, GCN）因其对非欧几里得空间数据的强大表示能力^[16]，近年来受到广泛关注。通过将小尺寸图像映射到图结构空间，可以有效捕捉像素之间的关系以及局部上下文特征，从而克服传统 CNN 对有限像素信息的依赖性，GCN 在节点关系建模、局部与全局信息融合等方面具有天然优势，为小尺寸图像的特征提取和分类任务提供了新的可能性。

基于此，本文提出一种基于图卷积网络的小尺寸图像识别算法，旨在充分利用 GCN 的拓扑结构建模能力，对小尺寸图像进行更深层次的特征挖掘。新算法通过构建像素或区域关系图，实现对图像全局和局部特征的联合建模，以期在低分辨率图像分类任务中取得更优的性能和更广泛的应用价值。

2 图神经网络原理和数据集

2.1 数据预处理和图数据集的构建

Fashion-MNIST 是一个由 Zalando 的商品图片组成的数据集，包含 60,000 个训练样本和 10,000 个测试样本。每个样本是一个 28×28 的灰度图像，并附有一个属于 10 个类别之一的标签。Zalando Research 团队希望 Fashion-MNIST 能作为原始 MNIST 数据集的直接替代品，用于机器学习算法的基准测试。它与 MNIST 数据集在图像大小以及训练和测试集划分的结构上完全一致^[17]。Fashion MNIST 中的图像更加复杂，具有更高的视觉多样性，包含了更多的纹理和边缘信息。因此，它对分类算法提出了更高的挑战，能够更好地评估算法的实际应用能力。

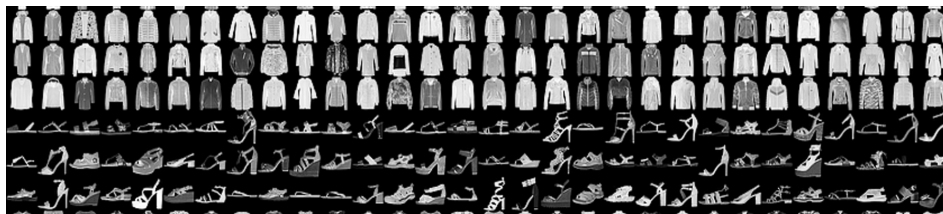


Figure 1: fashionmnist 数据集中的服装图像

本实验基于此 Fashion MNIST 数据集，由于 28×28 像素的图像仍然包含足够的信息和细节，所以本实验使用线性插值法将原本 28×28 像素的图像缩小为 7×7 像素的图像，从而模拟超小尺寸图像分类的实验环境。下图为缩小前后同样的图像的比较。

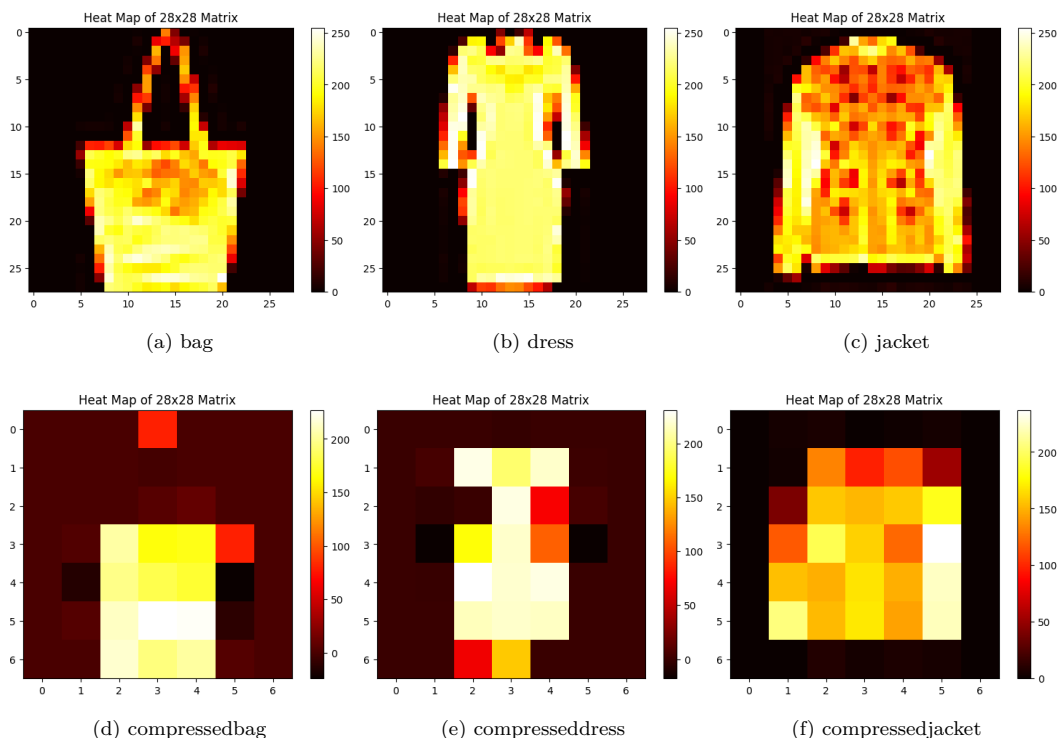


Figure 2: 28×28 的原图像和压缩成 7×7 后的图像

在数据处理中使用 pandas 中的 isna 函数查找缺失值，结果为训练集和测试集中的每个像素

都不存在缺失值。用 `scipy.ndimage` 中的 `zoom` 函数使用线性插值法将图片压缩为 7*7 的图片（例如 figure2 中的 d、e、f 分别为 a、b、c 的包、连衣裙、夹克压缩后的图像）。

由于 Fashion Mnist 为灰度图像，所以对每个像素都只包含一个值，将其标准化使每个像素的值都在 $[0, 1]$ 中，方便后续进行训练。

图 $G = (V, E)$ 由节点集 V 和边集 E 组成, V 是图中节点（顶点）的集合, E 是图中的边的集合, 用来描述节点之间的关系。邻接矩阵 A 是表示节点间连通性的矩阵, 特征矩阵 X 是每个节点的特征向量的集合, 维度为 $N \times F$, 其中 N 是节点数, F 是特征维度。

图卷积神经网络的核心思想是聚合节点及其邻居节点的信息, 通过层级操作学习节点的嵌入表示。公式如下:

$$H^{(l+1)} = \sigma \tilde{A} H^{(l)} W^{(l)}$$

其中:

- $H^{(l)}$: 第 l 层的节点特征表示。
- $\tilde{A} = D^{-\frac{1}{2}}(A + I)D^{-\frac{1}{2}}$: 加入自环并进行归一化的邻接矩阵, 用于避免数值不稳定性。
- $W^{(l)}$: 第 l 层的可训练权重矩阵。
- σ : 非线性激活函数 (如 ReLU)。

因此创建数据集需要构造标签, 节点和边的集合。数据的第 1 列定义为分类的标签, 其取值范围为 $[0, 9]$ 中的整数共十个标签。第 2 到最后一列也就是第 50 列表示对应的图 7*7 个像素的值, 将后 49 列转为 7*7 的矩阵。将新的矩阵转为节点和边邻接矩阵, 其中节点共有 49 个, 每条边都由相邻的两个像素点构成, 其坐标为两个节点的一位坐标表示, 例如位于第一行第一列的像素相关的邻边有 $\{(0, 0), (0, 1), (0, 7), (0, 8)\}$, 位于中间的某个像素出发的邻边则有最多九条。边的总数一共为 361 条, 由 `torch_geometric` 中的 `grid()` 函数生成。

创建好了训练集和测试集并将训练集打乱顺序后, 得到了用来训练的 `train_loader` 的 `DataLoader` 数据以及 `test_loader`。

2.2 模型架构与设计思路

模型结构描述

输入参数:

`in_channels`: 输入特征的维度 (每个节点的特征维度)。

`hidden_channels1`, `hidden_channels2`: 第一层和第二层隐藏特征的维度。

`out_channels`: 输出特征的维度 (通常与任务的类别数相同)。

网络层次:

第一层卷积(conv1): 使用 `GCNConv` 将输入特征从 `in_channels` 转换为 `hidden_channels1`。

第二层卷积 (conv2): 将中间特征从 `hidden_channels1` 转换为 `hidden_channels2`。

第三层卷积 (conv3): 将中间特征从 `hidden_channels2` 转换为 `out_channels`。

全局池化层: 使用 *global_mean_pool* 对每个图的节点特征求均值, 将节点级别的信息聚合为图级别的表示 (用于图分类任务)。

激活函数: 在每层卷积后使用 ReLU 激活函数, 引入非线性能力。

输出层: 使用 *F.log_softmax* 将输出特征转换为概率分布, 方便计算分类损失 (如交叉熵损失)。

前向传播过程

输入数据: 输入数据, 通常包含节点特征矩阵 x 和边索引矩阵 *edge_index*, 以及图的批次信息 *data.batch* (用于处理多个图)。

第一层卷积 (conv1): 利用图的拓扑结构 (由 *edge_index* 描述) 计算节点的新特征。输出特征维度从 *in_channels* 变为 *hidden_channels1*。通过 ReLU 激活引入非线性。

第二层卷积(conv2): 再次利用图的拓扑结构更新节点特征。输出特征维度从 *hidden_channels1* 变为 *hidden_channels2*。通过 ReLU 激活。

第三层卷积 (conv3): 生成每个节点的最终特征。输出特征维度为 *out_channels*。

全局池化: 对每个图的所有节点特征求均值, 获得整个图的表示 (*global_mean_pool*)。

输出层: 使用 *F.log_softmax* 输出每个图的类别概率分布。

2.3 图卷积网络时间复杂度分析

训练时, 单层 GCN 的计算主要分为两个部分:

1. 邻接矩阵和特征矩阵的乘法

$\tilde{A}H^{(l)}$ 时间复杂度: $O(E \cdot F^{(l)})$, 其中 E 是边的数量, $F^{(l)}$ 是节点特征维度。

2. 特征矩阵和权重矩阵的乘法

$(\tilde{A}H^{(l)})W^{(l)}$ 时间复杂度: $O(E \cdot F^{(l)} \cdot F^{(l+1)})$, 其中 $F^{(l+1)}$ 是输出特征维度。

假设 GCN 具有 L 层, 那么其总时间复杂度为:

$$O(L \cdot (E \cdot F^{(l)} + E \cdot F^{(l)} \cdot F^{(l+1)}))$$

3 模型实现

3.1 模型评估和结果分析

模型特点: 1. 利用 GCNConv 处理图的拓扑结构, 聚合节点邻居的信息。

2. 多层设计能够提取更高阶的图特征。

3. 全局池化适用于图分类任务。

这个模型通过多层图卷积对节点的特征进行更新和提取，同时利用全局池化将节点级别的信息整合为图级别表示，最终通过分类层完成图分类任务。该模型可以有效地对 7*7 的服装图像提取节点和相邻节点的边的特征关系，从而有效地对微小图片进行分类。

3.2 实验设计

使用参数 $in_channels = 1$, $hidden_channels1 = 16$, $hidden_channels2 = 32$, $lr = 0.01$, $weight_decay = 1e - 5$ 的参数进行训练，共训练 10 个轮次，在每一个轮次后输出 loss 值以及测试集和测试集的准确率。

设置与当前主流的 CNN（卷积神经网络）的对比实验来对比两种模型的效果。对于 GCN 和 CNN 设置同样的十个轮次，在每个训练轮次输出训练的 loss 值以及用时。在十个轮次结束后输出总用时以及测试集的准确率进行比较。

3.3 实验结果

将训练集和测试集的 batch size 都设为 64，CNN 模型和 GCN 模型的表现如下：

epochs	loss	time
1	1.9971	5.33s
2	1.7574	5.37s
3	1.6182	5.41s
4	1.5801	5.02s
5	1.5654	5.07s
6	1.5569	4.89s
7	1.5475	5.10s
8	1.5390	4.78s
9	1.5289	4.84s
10	1.5208	4.97s
Total Training Time: 50.78s		
Test Accuracy: 47.28%		

(a) GCN

epochs	loss	time
1	1.0032	1.83s
2	0.6801	1.69s
3	0.6288	1.80s
4	0.5969	1.78s
5	0.5746	1.82s
6	0.5583	1.83s
7	0.5441	1.81s
8	0.5323	1.86s
9	0.5222	1.78s
10	0.5145	1.73s
Total Training Time: 17.94s		
Test Accuracy: 82.19%		

(b) CNN

Table 1: GCN 和 CNN 训练结果

3.4 结果分析和优化

在相同的 batch size 设置下，分别使用图卷积神经网络和卷积神经网络对有十个类别 Fashion MNIST 数据集的训练集进行训练。图卷积神经网络在十个轮次的训练后 loss 值降到了 1.52，共用时 50.78 秒，测试集的准确率为 47.28%。虽然对比卷积神经网络来说，卷积神经网络在速度和

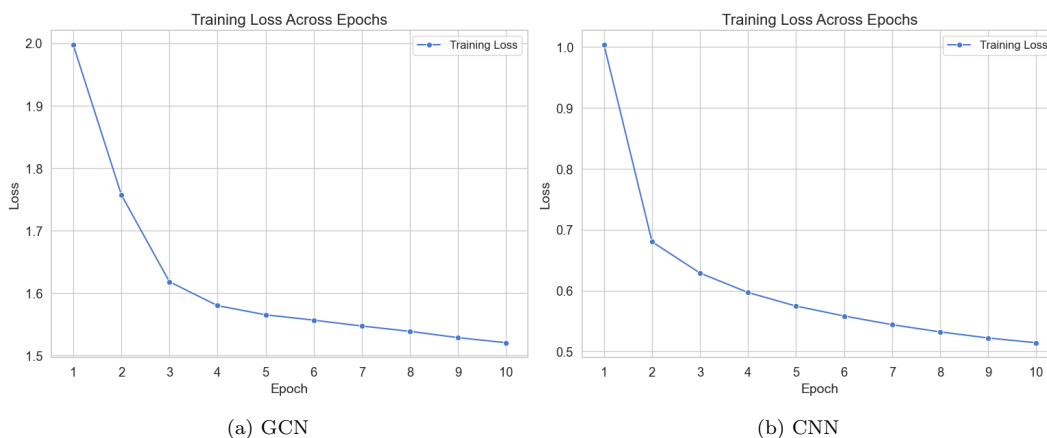


Figure 3: GCN 和 CNN 的 loss 值

准确率上都有很大的优势，但是图卷积神经网络的 loss 值在后半段下降的速率略高于 CNN，这意味着增加训练的轮次可以更有效地提高 GCN 的分类效果。

同时使用相同的参数设置对没有进行压缩的 28×28 的原数据进行训练，十个轮次的训练后测试集的分类准确率为 37.0400% 并且用时 82.20 秒。可以发现图卷积网络确实在更小尺寸的图像训练分类上具有优势，图结构减小不仅减少了计算的时间复杂度，更简化了图结构减少了过拟合的情况。

观察图 2.1GCN 的 loss 随着轮次变化的折线图可以看出，后半段 loss 值下降较为线性，并且 loss 值初始是较高的，所以可以通过调整参数和增加轮次数来提升模型性能。由于训练速度已经比卷积神经网络落后了很多，所以卷积层数不做变化，仅尝试调整特征维度等参数。

最终使用参数 $in_channels = 1$, $hidden_channels1 = 8$, $hidden_channels2 = 32$, $lr = 0.01$, $weight_decay = 1e - 5$ ，共训练 30 个轮次，batch size 仍然保持 64。

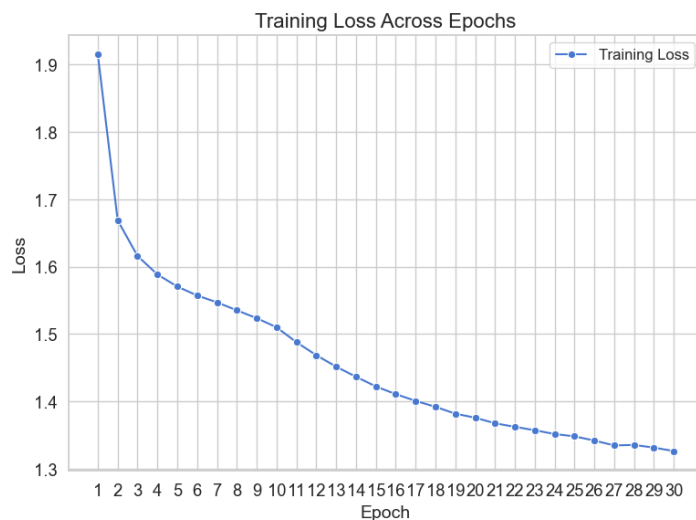


Figure 4: 30 轮次 GCN 的 loss 值

最终结果训练共用时 166.08 秒，在测试集上的分类准确率为 64.25%。

4 小结

4.1 研究总结

虽然基于图卷积神经网络的超小尺寸图片分类模型在目前仍然表现不如卷积神经网络，但是值得注意的是在试验过程中发现，当数据集从原本 Fashion MNIST 数据集原始的 28×28 像素换成 7×7 像素时，由于图像细节的丢失，卷积神经网络的分类准确率会有一定的下降，但是图卷积神经网络反而有一定的提升。由于图结构的特性，在图片的尺寸下降后，图结构本身并不会受到太大的影响。在训练速度上 GCN 也并没有比 CNN 落后一个数量级，最终 64.25% 的分类准确率也更加证明了图卷积神经网络模型在超小尺寸图像分类任务上的可用性。另外，若图形发生了几何变形，图像的图结构并不会产生改变，而 CNN 的固定架构和局部特征提取方式，使其在处理因视角变化导致变形的图像时存在不足，所以理论上若将数据集换成有视角变换的图像，GCN 有可能会比普通的 CNN 模型有更好的效果。

4.2 局限性与不足

本实验所采用的数据集 Fashion MNIST 只包含十种类别的服饰的图像，并且数量也仅为七万张，还未验证 GCN 在更大型的数据集上的效果。在模型的构建和参数选择上也存在调整的空间。图结构也有进一步优化的空间，可以使用数学方法或是其它神经网络模型提炼图像的节点而不是将每个像素都作为图像的节点。

4.3 未来工作

由于 Fashion MNIST 数据集中的服饰图像都为标准正面视角，这给 CNN 的训练带来了一定的便利性，CNN 对平移具有一定的等变性（由卷积和池化操作实现），但对旋转、缩放、倾斜或透视变换等几何变化不具鲁棒性，所以对于有角度不同的物体图像的数据集需要引入空间变换网络或者等变卷积网络来对抗几何变形，这些操作都增加了模型的复杂度，而图结构本身并不会被因为视角变化而产生的几何变化影响。在未来的工作中，可以尝试使用存在视角角度不同的图像来进行训练，相信会有比用 Fashion MNIST 数据集更好的效果。

参考文献

- [1] Chenyi Chen, Ming-Yu Liu, Oncel Tuzel, and Jianxiong Xiao. R-cnn for small object detection. In Shang-Hong Lai, Vincent Lepetit, Ko Nishino, and Yoichi Sato, editors, *Computer Vision – ACCV 2016*, pages 214–230, Cham, 2017. Springer International Publishing.
- [2] Djemel Ziou and Salvatore Tabbone. Edge detection techniques-an overview. */Pattern Recognition and Image Analysis: Advances in Mathematical Theory and Applications*, 8(4):537–559, 1998.
- [3] Jianxin Wu. Introduction to convolutional neural networks. *National Key Lab for Novel Software Technology. Nanjing University. China*, 5(23):495, 2017.
- [4] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012.

- [5] Nurul A Asif, Yeahia Sarker, Ripon K Chakraborty, Michael J Ryan, Md Hafiz Ahamed, Dip K Saha, Faisal R Badal, Sajal K Das, Md Firoz Ali, Sumaya I Moyeen, et al. Graph neural network: A comprehensive review on non-euclidean space. *Ieee Access*, 9:60588–60606, 2021.
- [6] Qi Zhang, Jianlong Chang, Gaofeng Meng, Shibiao Xu, Shiming Xiang, and Chunhong Pan. Learning graph structure via graph convolutional networks. *Pattern Recognition*, 95:308–318, 2019.
- [7] Boris Knyazev, Graham W Taylor, and Mohamed Amer. Understanding attention and generalization in graph neural networks. *Advances in neural information processing systems*, 32, 2019.
- [8] Shiwen Wu, Fei Sun, Wentao Zhang, Xu Xie, and Bin Cui. Graph neural networks in recommender systems: a survey. *ACM Computing Surveys*, 55(5):1–37, 2022.
- [9] Kanchan Jha, Sriparna Saha, and Hiteshi Singh. Prediction of protein–protein interaction using graph neural networks. *Scientific Reports*, 12(1):8360, 2022.
- [10] Xiao Li, Li Sun, Mengjie Ling, and Yan Peng. A survey of graph neural network based recommendation in social networks. *Neurocomputing*, 549:126441, 2023.
- [11] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European conference on computer vision (ECCV) workshops*, pages 0–0, 2018.
- [12] Si Zhang, Hanghang Tong, Jiejun Xu, and Ross Maciejewski. Graph convolutional networks: a comprehensive review. *Computational Social Networks*, 6(1):1–23, 2019.
- [13] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. *arXiv preprint arXiv:1710.10903*, 2017.
- [14] Le Wu, Peijie Sun, Richang Hong, Yanjie Fu, Xiting Wang, and Meng Wang. Socialgcn: An efficient graph convolutional network based model for social recommendation. *arXiv preprint arXiv:1811.02815*, 2018.
- [15] Jiarui Chen, Yain-Whar Si, Chon-Wai Un, and Shirley WI Siu. Chemical toxicity prediction based on semi-supervised learning and graph convolutional neural network. *Journal of cheminformatics*, 13:1–16, 2021.
- [16] Zhiyuan Liu and Jie Zhou. *Introduction to graph neural networks*. Springer Nature, 2022.
- [17] Han Xiao, Kashif Rasul, and Roland Vollgraf. Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. *arXiv preprint arXiv:1708.07747*, 2017.