



北京大学

图神经网络课程报告

具有残差连接 (ResNet) 和稠密连接 (DenseNet) 结构的
图神经网络在节点分类问题中的尝试

小组成员：

李瑞庆 2001213502

王楠哲 2001111734

韩鑫明 2001111718

赵越 2001111713

摘要

本文针对引文网络的半监督节点分类问题,进行了较为深入的文献调研工作,在调研过程中发现,随着层数的加深,在图神经网络的训练过程中可能会存在梯度消失、过渡平滑等问题。而且在调研过程中也发现,已有一些学者尝试将传统的卷积网络中的结构引入到图卷积中,以提高其性能。针对这些问题以及从文献中获得的启发,我们尝试将传统卷积网络中的一些技巧引入到图卷积网络中,如残差连接网络(ResNet)、稠密连接网络(DenseNet)等。基于此,我们提出将 ResGCN 和 DenseGCN 应用到节点分类问题中,并将测试结果与几种经典的图卷积网络的 baseline 进行了对比。结果发现,过多网络层数对半监督节点分类问题是有害的,使用残差连接和稠密连接这两种连接方式能够一定程度上抑制这种效应,在一些模型中取得了比 baseline 更好地准确率。

关键词: 半监督节点分类 图卷积网络 残差连接网络 稠密连接网络

目录

摘要.....	1
第一章 问题背景.....	3
1.1 引文网络中的节点分类问题	3
1.2 图卷积神经网络	4
1.2.1 谱图卷积.....	4
1.2.2 ChebNet.....	4
1.2.3 GCN(ChebNet 的一阶近似).....	5
1.3 图小波神经网络(Graph Wavelet Neural Network)	6
1.4 图注意力网络(Graph Attention Networks)	6
第二章 文献调研.....	8
2.1 具有跨层连接的 ARMA 滤波	8
2.2 分布式循环反馈网络(DFNets)	9
2.3 基于图结构数据的 U-net (Graph U-net)	10
第三章 方法改进.....	12
第四章 实验	14
4.1 模型	14
4.2 数据	14
4.3 结果	15
第五章 总结	16
参考文献	17

第一章 问题背景

1.1 引文网络中的节点分类问题

节点分类任务，是图神经网络研究中广泛采用的任务之一，该任务目标是通过给定图中某些节点对应的类别，对没有标签的节点的类别进行预测，也被称为半监督节点分类任务^[1-3]，如图 1 所示。

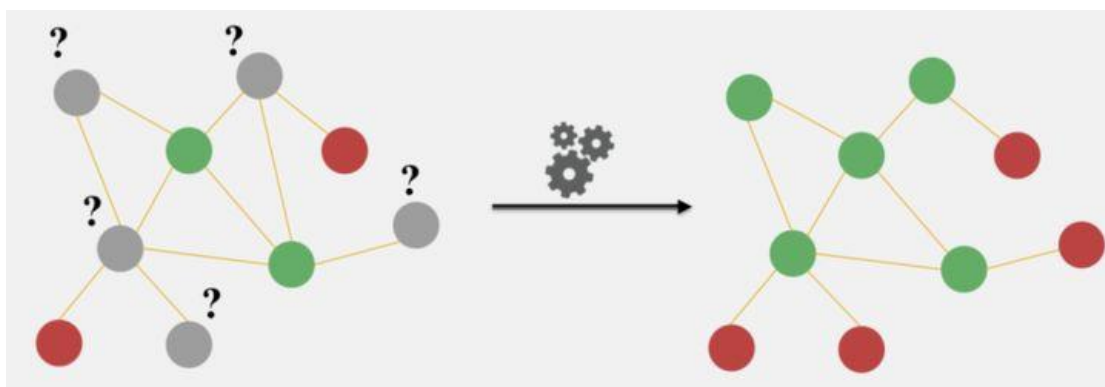


图 1 节点分类任务图示

而引文网络，顾名思义就是由论文和它们之间的引用关系构成的网络，具有天然的图结构，比较常用的数据集有 Cora、Citeseer 等等，数据集的组成结构如表 1 所示，Nodes 代表数据集中的论文数量，features 是每篇论文的特征，数据集中有一个包含多个单词的词汇表，去除了出现次数小于 10 的词，论文的属性由一串二进制构成，用 0 和 1 表示该论文有无某个词汇。

表 1 引文网络常用数据集组成结构

	Cora	Citeseer	Pubmed
Task	Transductive	Transductive	Transductive
# Nodes	2708 (1 graph)	3327 (1 graph)	19717 (1 graph)
# Edges	5429	4732	44338
# Features/Node	1433	3703	500
# Classes	7	6	3
# Training Nodes	140	120	60

# Validation Nodes	500	500	500
# Test Nodes	1000	1000	1000

因此，在引文网络上进行节点分类，其实可以看作是节点分类任务的一个应用，而图神经网络可以利用节点自身的特征，以及其相邻节点和边的特征进行特征传播，通过多次重复传播，可以聚合来自更大范围的邻居节点的信息，在节点分类任务上取得了不错的表现。

1.2 图卷积神经网络

传统的 CNN 在特征提取方面有着很好的效果，但是固定大小的卷积核限制了 CNN 仅适合处理 Euclidean Structure 的数据，而不适合处理 Non Euclidean Structure 的数据，这是因为离散卷积在 Non Euclidean Structure 数据上无法保持平移不变性，也就是对于 Non Euclidean Structure 数据如拓扑图，每个顶点的相邻顶点数目可能不同，因此无法用一个固定大小的卷积核进行特征提取。目前的研究主要在两个方向给出了这个问题的解决方案，一个方向是在空间域定义图卷积操作，另一个方向是在谱域定义图卷积操作，而 GCN 是基于谱域图卷积操作对图数据进行特征提取的，下面简要介绍一下谱图卷积和 GCN 的原理。

1.2.1 谱图卷积

谱图卷积的主要思想是通过图的傅里叶变换，将图的卷积操作变换到谱域做乘法，再通过傅里叶变换的逆变换变换回图域，完成谱图卷积操作。谱图卷积主要是通过图的拉普拉斯矩阵的特征值和特征向量来研究图的性质，其公式如(1)式所示：

$$g_{\theta} \star x = U g_{\theta} U^{\top} x \quad (1)$$

其中， x 为图节点的特征向量， g_{θ} 为卷积核， U 为图的拉普拉斯矩阵 L 经过特征分解得到的特征向量矩阵。

1.2.2 ChebNet

谱图卷积的提出，解决了非欧结构数据上的卷积操作定义问题，但是仍存在一些不足之处，如图卷积核是全局的且参数量大，图的拉普拉斯矩阵特征值分解

操作计算复杂度高等。为了解决这些问题，Hammond 等^[4]提出的 ChebNet 采用了 k 阶迭代定义的切比雪夫多项式对 g_θ 进行多项式近似，公式如(2)式所示：

$$g_{\theta'}(\Lambda) \approx \sum_{k=0}^K \theta'_k T_k(\tilde{\Lambda}) \quad (2)$$

$$\tilde{\Lambda} = \frac{2}{\lambda_{\max}} \Lambda - I_N \quad (3)$$

其中， Λ 是图的拉普拉斯矩阵分解得到的特征值对角矩阵， I_N 是单位矩阵，经过切比雪夫多项式近似后，图的卷积公式如式(4)所示：

$$g_{\theta'} \star x \approx \sum_{k=0}^K \theta'_k T_k(\tilde{L})x \quad (4)$$

$$\tilde{L} = \frac{2}{\lambda_{\max}} L - I_N \quad (5)$$

此时卷积操作不再需要对拉普拉斯矩阵进行特征分解，大大降低了计算复杂度，同时也关注到了 k 阶邻居的信息，不再是全图的卷积，降低了参数量。

1.2.3 GCN(ChebNet 的一阶近似)

根据 ChebNet 的改进思路，Kipf 和 Welling^[5]采用一阶切比雪夫近似，提出了 GCN，使图卷积变的更加简单快速，在 ChebNet 的基础上，令多项式的阶数 $K=1$ ， $\lambda_{\max}=2$ ，实现了一个卷积核更小，参数量更少，计算更快的网络，并通过重归一化等优化手段，得到的最终形式如式(6)所示：

$$Z = \tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} X \Theta \quad (6)$$

与 ChebNet 相比，GCN 的复杂度和参数量更小，而且可以通过叠加多个 GCN 层扩大卷积的感受野，灵活性很高，更加易于训练，因此，GCN 在图神经网络的许多任务上都取得了出色的成绩。多层 GCN 网络结构示意图如图 2 所示：

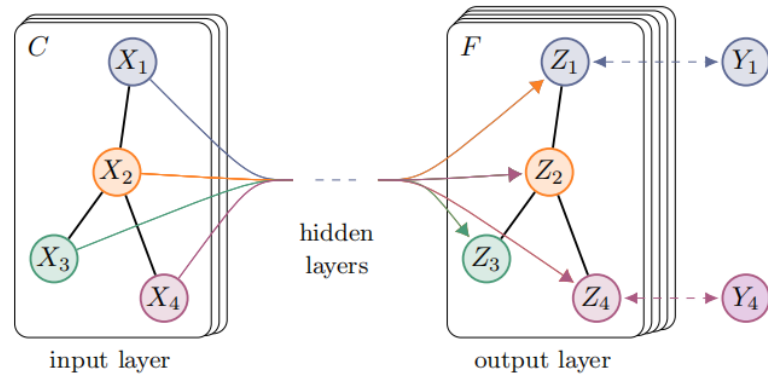


图 2 多层 GCN 网络结构示意图

1.3 图小波神经网络(Graph Wavelet Neural Network)

Xu 等^[6]提出的 GWNN 是 GCN 的一种变体，通过引入图信号上的小波变换来代替傅里叶变换，重新定义了图的卷积操作，卷积公式如式(7)所示：

$$\mathbf{x} *_G \mathbf{y} = \psi_s \left(\left(\psi_s^{-1} \mathbf{y} \right) \odot \left(\psi_s^{-1} \mathbf{x} \right) \right) \quad (7)$$

其中， ψ_s 是一组小波基，可以通过快速算法获得，无需进行拉普拉斯矩阵特征分解；而且小波基的稀疏性高，图小波变换比图傅里叶变换计算效率更高。

图小波神经网络通过引入图信号上的小波变换来改造 GCN，将卷积计算变换到空域局部邻域内，将数据表征从欧式空间转化到双曲空间，不仅能获得更好的层次化表示，同时能大大节约空间维度。

1.4 图注意力网络(Graph Attention Networks)

Velickovic 等^[7]于 2018 年提出了引入注意力机制的 GAT。与 GCN 与 GWNN 的思路不同，GAT 将视角放在空间域，通过引入注意力机制进行图的特征提取。

GAT 的原理是，针对每一个输入特征，根据一个 attention 机制计算节点与其邻居的 attention score，然后根据 attention score 对节点的所有邻居的特征进行加权求和，并通过线性变换和激活函数得到最后的输出特征。为了更好地训练网络和提取特征，引入 mask 机制，对每个节点只计算与其一阶邻居的 score 并加权求和；还引入了多头注意力机制，即对多个 attention 的结果进行 concat 操作。引入多头注意力机制的 GAT 部分结构如图 3 所示：

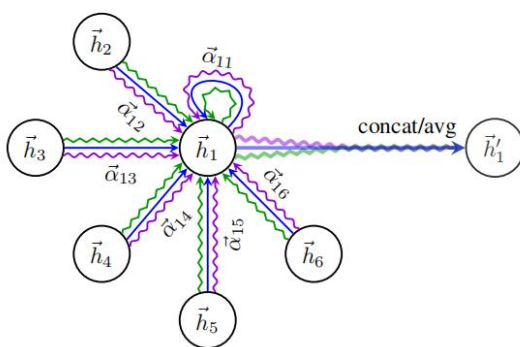


图 3 多头注意力机制

相比于 GCN 这类基于谱域图卷积的图神经网络，GAT 能够更好地处理有向图和动态图，不仅能处理 *transductive* 的任务，还能处理 *inductive* 的任务，也就是训练和测试时可以使用不同的图结构。而且 GAT 通过引入注意力机制，能够为节点的邻居分配不同的权重，更好地提取图的特征信息。

第二章 文献调研

针对基于图神经网络的半监督节点分类问题,我们也进行了广泛的文献调研工作。除了前文中提到的 ChebNet, GCN, GWNN 和 GAT 等等,还有一系列对于滤波形式进行改进的方法,如 Liao 等^[8]提出的基于 Lanczos 滤波的 LanczosNet, Levie 等^[9]提出的基于 Cayley 滤波的 CayleyNets 等等。但是我们本篇报告更多关注的是网络结构的优化对于图卷积性能的提升,尤其是对于传统 CNN 结构的迁移,故在这里我们重点介绍几篇对于图卷积网络结构做出一些改进的工作。

2.1 具有跨层连接的 ARMA 滤波

Bianchi^[10]等人于 2019 年提出了一种基于自回归移动平均(ARMA)滤波的图卷积神经网络,该方法与多项式方法相比,具有更强的鲁棒性和更灵活的图频响应。不同于 Cayley 滤波的方法,他们开发了一阶 ARMA 滤波,可以用一阶递归来进行近似:

$$\bar{\mathbf{X}}^{(t+1)} = a\mathbf{M}\bar{\mathbf{X}}^{(t)} + b\mathbf{X} \quad (8)$$

其中, a 和 b 是滤波的参数, $\bar{\mathbf{X}}^{(0)} = \mathbf{X}$, 并且 $\mathbf{M} = \frac{1}{2}(\lambda_{\max} - \lambda_{\min})\mathbf{I} - \mathbf{L}$ 。因此,

频率响应可以定义为:

$$h(\tilde{\lambda}) = \frac{r}{\tilde{\lambda} - p} \quad (9)$$

其中, $\tilde{\lambda} = (\lambda_{\max} - \lambda_{\min}) / 2\lambda$, $r = -b/a$, $p = 1/a$ 。作者提出的基于 ARMA 滤波的图卷积网络形式如下:

$$\bar{\mathbf{X}}^{(t+1)} = \sigma\left(\mathbf{L}\bar{\mathbf{X}}^{(t)}\mathbf{W}^{(t)} + \mathbf{X}\mathbf{V}^{(t)}\right) \quad (10)$$

其中, $\mathbf{W}^{(t)} \in \mathbb{R}^{F_{\text{out}}^t \times F_{\text{out}}^{t+1}}$, $\mathbf{V}^{(t)} \in \mathbb{R}^{F_{\text{in}} \times F_{\text{out}}^{t+1}}$, 均为网络训练的参数, \mathbf{X} 为初始的结点特征, 修正的拉普拉斯矩阵 $\mathbf{L} = \mathbf{I} - \mathbf{L}$ 。多个 ARMA 滤波也可以并行应用, 然而, k 个并行 ARMA 滤波的内存复杂度是单个 ARMA 图滤波的 k 倍。网络结构如图 4 所示。值得注意的是, 在所示的网络结构中, 存在一个“跳跃连接”, 也就是在图上进行相邻节点之间的局部交换来执行卷积操作之后, 还会将最初始的

节点信息考虑进来，这有效地避免了多个卷积层叠加之后，由于过渡平滑导致的信息丢失。

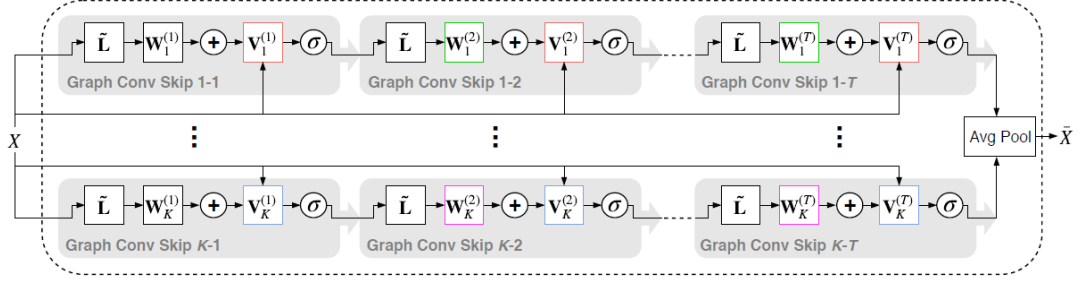


图 4 ARMA 图卷积层

2.2 分布式循环反馈网络(DFNets)

Wijesinghe^[11]等人提出了一种新的谱卷积神经网络模型，即分布式循环反馈网络(DFNets)。该模型引入了具有较强鲁棒性的谱图滤波器，称为循环反馈滤波，可以在节点上实现更好的局部化，同时也具有较快的收敛速度，并能满足线性存储的要求。

文中提出的循环反馈滤波也属于一种 ARMA 滤波，其滤波形式如下：

$$\bar{x}^{(t)} = -\sum_{j=1}^p \psi_j \tilde{L}^j \bar{x}^{(t-1)} + \sum_{j=0}^q \phi_j \tilde{L}^j x \quad (11)$$

其中， p 和 q 分别指的是后馈和前馈的度， ψ 和 ϕ 是两个复数因子的向量， $\bar{x}^{(0)} = x$ ，

$$\tilde{L} = \hat{L} - \left(\frac{\hat{\lambda}_{\max}}{2} \right) I, \quad \hat{L} = I - \hat{D}^{-1/2} \hat{A} \hat{D}^{-1/2}, \quad \hat{A} = A + I, \quad \hat{D}_{ii} = \sum_j \hat{A}_{ij}, \quad \hat{\lambda}_{\max} \text{ 是 } \hat{L} \text{ 最大的特征值。}$$

循环反馈滤波器的频率响应可定义为：

$$h(\lambda_i) = \frac{\sum_{j=0}^q \phi_j \lambda_i^j}{1 + \sum_{j=1}^p \psi_j \lambda_i^j} \quad (12)$$

基于循环反馈滤波，作者提出了相应的图卷积神经网络，称之为分布式循环反馈网络(DFNets)，图卷积网络的前向传播格式如下：

$$\bar{X}^{(t+1)} = \sigma \left(\mathbf{P} \bar{X}^{(t)} \theta_1^{(t)} + \mathbf{Q} X \theta_2^{(t)} + \mu(\theta_1^{(t)}; \theta_2^{(t)}) + b \right) \quad (13)$$

其中, $\mathbf{P} = -\sum_{j=1}^p \psi_j \tilde{L}^j$, $\mathbf{Q} = \sum_{j=0}^q \phi_j \tilde{L}^j$ 。在分布式循环反馈网络(DFNets)中, 可以将多个循环反馈滤波卷积层叠加起来, 以提取越来越抽象的特征, 同时为了避免过度平滑, 作者还引入了 DenseNet 结构的跨层连接。

2.3 基于图结构数据的 U-net (Graph U-net)

Gao 和 Ji^[12]提出了基于图结构数据的图池化(gPool)操作和图去池化(gUnpool)操作, 并与传统的 CNN 类似, 基于此构建了针对图结构数据的 U-net。gPool 层根据节点在可训练投影向量上的标量投影值, 自适应地选择一些节点形成较小的图, 其具体操作过程如图 5 所示。

在这一层, 可以使用一个可训练的投影向量 \mathbf{p} 。通过将所有节点特征投影到一维向量, 由于希望在 \mathbf{p} 方向上投影时能保留尽可能多的原始图上的信息, 可以选择 \mathbf{p} 上标量投影值最大的节点来形成一个新的图。对于保留下来的节点, 对其在邻接矩阵中的信息也要进行保留。最后, 可以通过一个 Sigmoid 函数对投影的向量中的元素进行操作, 实现门控的作用, 来控制信息流。

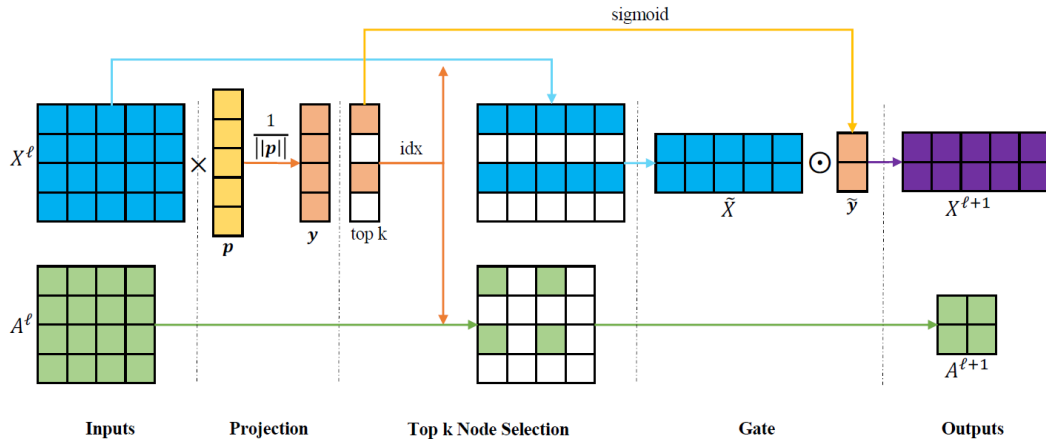


图 5 图池化(gPool)层操作图示

作者也进一步提出 gUnpool 层作为 gPool 层的逆操作。gUnpool 层利用对应 gPool 层中所选节点的位置信息, 将图恢复到原来的结构。gUnpool 层的操作如图 6 所示。在这一层, 可以记录在对应的 gPool 层中选择的节点的位置, 并使用这个信息将节点放置到它们在图中的原始位置, 在重构原始图结构过程中, 对于重构出来的节点使用的是空的特征向量(empty feature vectors)。

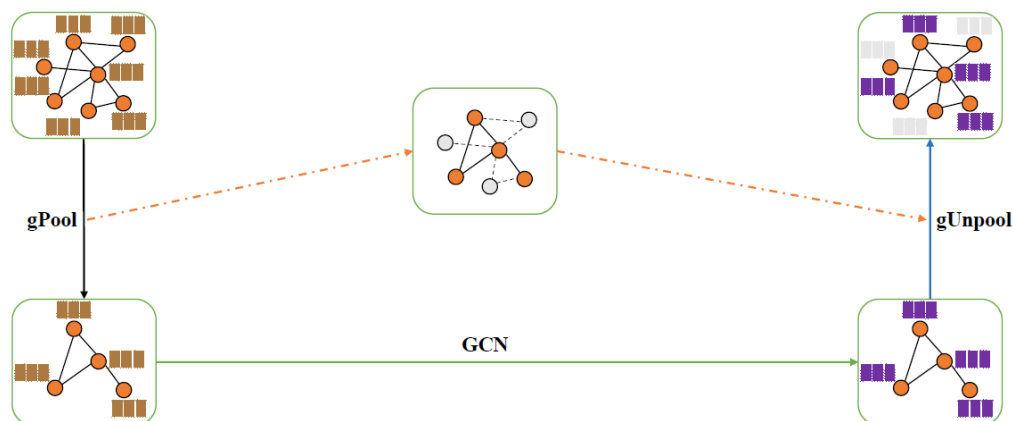


图 6 图去池化(gUnpool)操作图示

有了图池化(gPool)操作和图去池化(gUnpool)操作，作者便仿照传统 CNN 结构中的 U-net 形式，用这两个操作构建了针对图结构数据的 U-net，称为图 U-net (Graph U-net)，其结构如图 7 所示。基于此结构，图 U-net (Graph U-net)可用于节点分类和图分类等任务。

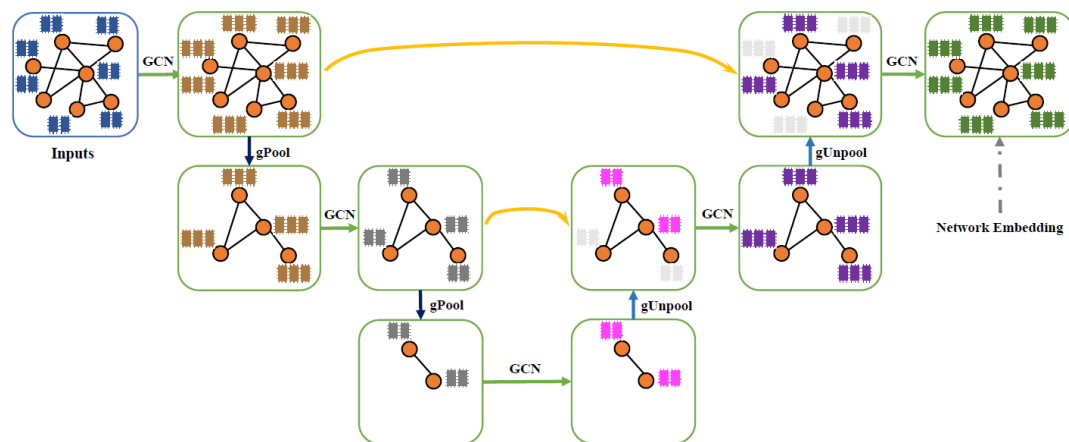


图 7 图 U-net (Graph U-net)结构图示

通过以上文献的介绍可以看出，我们的调研的文献多是将传统的卷积神经网络(CNN)中，一些已经取得很好效果的结构借鉴到图卷积网络中，来实现更好的效果。这也启发了我们，可以尝试将 CNN 中的残差连接(ResNet)、稠密连接(DenseNet)的构架引入到图卷积网络中来，来探索能否实现对节点分类更好的效果。

第三章 方法改进

卷积神经网络(CNNs)在许多领域都取得了不错的应用，这得益于当深层的 CNN 可以有效的得到训练。尽管卷积神经网络有其优点，CNNs 不能有效地解决非欧几里德数据的问题。为了克服这一挑战，图卷积网络(GCNs)构建图来表示非欧几里德数据，借用 CNN 的概念，并将其应用于训练中。但由于梯度消失，过平滑等问题，它们通常仅限于非常浅的模型。因此，大多数最先进的 GCN 模型不深于 3 或 4 层。

对于过平滑问题，常见的解决策略有两个角度，深度拓展和广度扩展。对于深度拓展来说，常用的方法有残差神经网络以及强化自连接等，对于广度拓展来说，常用的方法为扩大感受野等。在本文中，我们选择深度拓展的角度，通过借用 CNN 的相关概念来优化这一点，特别是残差卷积和稠密卷积，并将它们适应 GCN 体系结构。

Li 等^[13]在 2019 年中使用该思路解决 3D 点云分割的相关问题，其贡献主要有以下两点。第一，提出了三种不同的图卷积神经网络结构，分别为残差卷积，稠密卷积以及膨胀卷积。第二，使用点云分割平台，展示了这些不同的网络结构对深度 GCNs 训练稳定性和性能的影响。

因此，本文尝试使用残差图卷积神经网络以及稠密图卷积神经网络作为模型，随后应用于我们目标的节点分类任务之中，并与节点分类的传统方法在分类准确率上进行对比，本节的其余部分将简单介绍 ResGCN 以及 DenseGCN 表达式以及网络结构示意图。

对于传统的 GCN，计算公式如下：

$$\begin{aligned}\mathcal{G}_{l+1} &= \mathcal{F}(\mathcal{G}_l, \mathcal{W}_l) \\ &= \text{Update} \left(\text{Aggregate} \left(\mathcal{G}_l, \mathcal{W}_l^{\text{agg}} \right), \mathcal{W}_l^{\text{update}} \right)\end{aligned}\quad (14)$$

然而传统的 GCN 随着层数的增加，模型的效果会因为过平滑的等因素逐渐变差，因此需要将高维特征对底层特征进行一定的融合，受到 ResNet 的启发把 GCN 进行了以下改进，并把它称为 ResGCN：

$$\begin{aligned}\mathcal{G}_{l+1} &= \mathcal{H}(\mathcal{G}_l, \mathcal{W}_l) \\ &= \mathcal{F}(\mathcal{G}_l, \mathcal{W}_l) + \mathcal{G}_l = \mathcal{G}_{l+1}^{\text{res}} + \mathcal{G}_l\end{aligned}\quad (15)$$

DenseNet 的提出是为了利用各层之间的紧密连通性，这改善了网络中的信息流，实现了各层之间特征的高效重用。受 DenseNet 的启发，将类似的想法应用到 GCNs 中，以便利用来自不同 GCN 层的信息流。具体做法如下：

$$\begin{aligned}
 \mathcal{G}_{l+1} &= \mathcal{H}(\mathcal{G}_l, \mathcal{W}_l) \\
 &= \mathcal{T}(\mathcal{F}(\mathcal{G}_l, \mathcal{W}_l), \mathcal{G}_l) \\
 &= \mathcal{T}(\mathcal{F}(\mathcal{G}_l, \mathcal{W}_l), \dots, \mathcal{F}(\mathcal{G}_0, \mathcal{W}_0), \mathcal{G}_0)
 \end{aligned} \tag{16}$$

为了更好的理解其网络结构，对比传统的 GCN，ResGCN，DenseGCN 的网络结构如图 8：

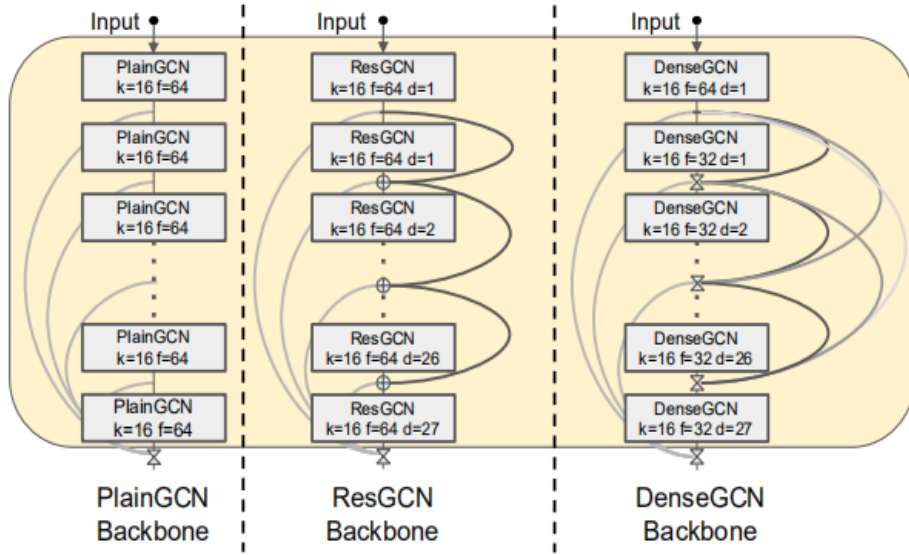


图 8 网络结构对比

随后为本文的实验部分，我们将分别对比 GCN，ResGCN，DenseGCN 这三种网络结构在不同数量的隐层下分类的准确率。

第四章 实验

4.1 模型

我们首先在传统 GCN 网络的基础上增加了网络深度，测试模型效果。进一步，我们分别将 DenseNet 和 ResNet 的技巧应用到传统的 GCN 上，包括深层 GCN，以此对比模型的效果。下面图 9 是模型的代码截图，具体代码见附件，或链接：https://github.com/sixingdeguo/GNN_project.git。

```
def forward(self, x, adj):
    x1 = F.relu(self.gc1(x, adj))
    x11 = F.dropout(x1, self.dropout1, training=self.training)
    x2 = torch.cat((x11, x), 1)
    x2 = F.dropout(x2, self.dropout2, training=self.training)
    x22 = F.relu(self.gc2(x2, adj))
    x3 = torch.cat((x22, x11, x), 1)
    x3 = F.dropout(x3, self.dropout3, training=self.training)
    x33 = F.relu(self.gc3(x3, adj))
    x4 = torch.cat((x33, x22, x11, x), 1)
    x4 = F.dropout(x4, self.dropout4, training=self.training)
    x44 = F.relu(self.gc4(x4, adj))
    x5 = torch.cat((x44, x33, x22, x11, x), 1)
    x5 = F.dropout(x5, self.dropout5, training=self.training)
    x55 = F.relu(self.gc5(x5, adj))
    return F.log_softmax(x55, dim=1)
```

图 9 DenseGCN 模型代码

4.2 数据

我们使用了三个文献引用网络数据集：cora、citeseer、pubmed 进行实验。文献间的引用视作无向边，三个数据集描述性统计数据如下：

表 2 数据集描述性统计

Dataset	Type	Nodes	Edges	Classes	Features	Label rate
Citeseer	Citation network	3,327	4,732	6	3,703	0:036
Cora	Citation network	2,708	5,429	7	1,433	0:052
Pubmed	Citation network	19,717	44,338	3	500	0:003

4.3 结果

我们通过网格搜索模型参数，选择每个模型最优的参数组合，下表展示了各模型在各数据集上的最优正确率：

表 3 模型测试结果

模型	Cora	Citeseer	Pubmed
GCN	0.853	0.647	0.780
DenseGCN	0.845	0.664	0.777
ResGCN	0.836	0.659	0.788
GCN-3	0.826	0.625	0.761
DenseGCN-3	0.850	0.639	0.783
ResGCN-3	0.837	0.662	0.778
GCN-4	0.793	0.61	0.771
DenseGCN-4	0.854	0.649	0.774
ResGCN-4	0.833	0.647	0.782
GCN-5	0.744	0.548	0.772
DenseGCN-5	0.837	0.626	0.750
ResGCN-5	0.830	0.642	0.767
GCN-10	0.309	0.373	0.655
DenseGCN-10	0.851	0.580	0.777
ResGCN-10	0.834	0.502	0.669
GCN-15	0.309	0.335	0.550
DenseGCN-15	0.849	0.449	0.695
ResGCN-15	0.833	0.391	0.622
GCN-20	0.309	0.338	0.480
DenseGCN-20	0.847	0.365	0.671
ResGCN-20	0.834	0.240	0.486
GCN-25	0.309	0.327	0.639
DenseGCN-25	0.851	0.313	0.413
ResGCN-25	0.834	0.213	0.661

注：模型后的数字为该模型的卷积层数

整体上来看，过多网络层数对半监督节点分类问题是有害的，过深的卷积层导致节点较近邻居信息被平滑，从而降低准确率。但使用残差连接网络(ResNet)和稠密连接网络(DenseNet)这两种连接方式能够一定程度上抑制这种效应，其中DenseGCN的抑制效果是最明显的。在三个不同的数据集上，我们的设计网络结构都比传统GCN网络的准确率有细微提升。

第五章 总结

针对半监督节点分类问题，我们进行了较为深入的文献调研工作，在调研过程中发现，随着层数的加深，在图神经网络的训练过程中可能会存在梯度消失、过渡平滑等问题。通过调研图卷积网络 GCN 的相关文献，我们发现许多文献将传统 CNN 中创造出的结构迁移到 GCN 上，并取得了不错的效果。

针对这些问题以及从文献中获得的启发，我们尝试将传统卷积网络中的一些技巧引入到图卷积网络中，如残差连接网络(ResNet)、稠密连接网络(DenseNet)等。基于此，我们提出将 ResGCN 和 DenseGCN 这两种结构迁移至 GCN，并将其应用到节点分类问题中，并将测试结果与几种经典的图卷积网络的 baseline 进行了对比。结果发现，发现其整体上来看，过多网络层数对半监督节点分类问题是有害的，使用残差连接和稠密连接这两种连接方式能够一定程度上抑制这种效应，在一些模型中取得了比 baseline 更好地准确率。

在未来的工作中，我们也会继续尝试将残差连接网络(ResNet)、稠密连接网络(DenseNet)与 GWNN、GAT 等图卷积方式做一些结合，探索残差连接网络(ResNet)、稠密连接网络(DenseNet)两种结构对 GWNN 和 GAT 性能的影响。

参考文献

- [1] Belkin M, Niyogi P, Sindhwani V. Manifold Regularization: A Geometric Framework for Learning from Labeled and Unlabeled Examples. *Journal of Machine Learning Research*. 2006;7(85):2399-2434.
- [2] Zhu X, Ghahramani Z, Lafferty J. Semi-supervised learning using Gaussian fields and harmonic functions. *International Conference on Machine Learning* 2003:912-919.
- [3] Zhou D, Bousquet O, Lal TN, et al. Learning with Local and Global Consistency. *Neural Information Processing Systems* 2003:321-328.
- [4] Hammond DK, Vandergheynst P, Gribonval R. Wavelets on graphs via spectral graph theory. *Applied and Computational Harmonic Analysis*. 2011;30(2):129-150.
- [5] Kipf TN, Welling M. Semi-Supervised Classification with Graph Convolutional Networks. *International Conference on Learning Representations* 2016.
- [6] Xu B, Shen H, Cao Q, et al. Graph Wavelet Neural Network. *International Conference on Learning Representations* 2019.
- [7] Velickovic P, Cucurull G, Casanova A, et al. Graph Attention Networks. *International Conference on Learning Representations* 2018.
- [8] Liao R, Zhao Z, Urtasun R, et al. LanczosNet: Multi-Scale Deep Graph Convolutional Networks. *International Conference on Learning Representations* 2019.
- [9] Levie R, Monti F, Bresson X, et al. CayleyNets: Graph Convolutional Neural Networks With Complex Rational Spectral Filters. *IEEE Transactions on Signal Processing*. 2019;67(1):97-109.
- [10] Bianchi FM, Grattarola D, Alippi C, et al. Graph neural networks with convolutional arma filters. *arXiv preprint arXiv: 190101343*. 2019.
- [11] Wijesinghe W, Wang Q. DFNets: Spectral CNNs for graphs with feedback-looped filters. *Advances in Neural Information Processing Systems*. 2019;32:6009-6020.
- [12] Gao H, Ji S. Graph U-Nets. *International Conference on Machine Learning* 2019:2083-2092.
- [13] Li G, Muller M, Thabet A, et al. DeepGCNs: Can GCNs Go As Deep As CNNs? *International Conference on Computer Vision* 2019:9267-9276.