卷积神经网络文献阅读报告

于烨泳 [1]

（1. 上海大学 计算机工程与科学学院，上海 200444）

摘要：在过去的几年里，深度学习已经在各种问题上取得了非常好的表现，如视觉识别、语音识别和自然语言处理。在不同类型的深度神经网络中，卷积神经网络得到了最广泛的研究。利用标注数据量的快速增长和图形处理器单元的巨大改进，卷积神经网络的研究已经迅速兴起，并在各种任务上取得了最先进的成果。本文总结了本人阅读的 3 篇关于卷积神经网络的论文，其中包括卷积神经网络的开始、中期发展和最新的研究方向，第二部分总结了与卷积神经网络息息相关的图卷积神经网络，笼统介绍了基于谱域和基于空域的图卷积方法，对比了两种方法的异同点，最后对比了卷积神经网络和图卷积神经网络。

关键词： 卷积神经网络；差分卷积网络；图卷积网络；谱域卷积；空域卷积

Convolutional Neural Network Literature Reading Repor

Yeyong Yu[1]

(1. School of Shang Hai University, Shanghai 200444, China)

Abstract: In the past few years, deep learning has achieved very good performance on various problems, such as visual recognition, speech recognition and natural language processing. Among the different types of deep neural networks, convolutional neural networks have been the most widely studied. Taking advantage of the rapid growth in the amount of labeled data and the tremendous improvements in graphics processor units, research on convolutional neural networks has rapidly emerged and achieved state-of-the-art results on a variety of tasks. This paper summarizes three papers I have read on convolutional neural networks, including the beginning, mid-term development, and latest research directions of convolutional neural networks. The second part summarizes graph convolutional neural networks, which are closely related to convolutional neural networks, introduces spectral domain-based and null domain-based graph convolution methods in general, compares the similarities and differences of the two methods, and finally compares convolutional neural networks and graph convolutional neural networks .

Key words: Convolutional neural network; Differential convolutional network; Graph convolutional network; Spectral domain convolution; Space domain convolution

1 引言

随着移动互联网的发展和各种社交媒体的普及，互联网上的图像数据量迅速增加，但人类无法有效地处理这么多的图像数据。因此，人们期望借助计算机自动进行这些数据处理，以解决大规模的视觉问题。随着人们对图像处理技术的深入了解，对图像的全面理解和对图像目标对象的准确识别变得越来越重要[1]，解决计算机视觉问题的传统方法的成功在很大程度上取决于特征提取过程，而卷积神经网络(Convolutional Neural Networks, CNN)就是广泛使用的特征提取深度学习框架[2]，其灵感来自动物的视觉皮层[3]，小孔成像会把图像翻转呈现在视网膜上，随后经过大脑处理之后得到正向的图像，所以 CNN 在定义时，就有一个翻转的操作。所谓两个函数的卷积，本质上就是先将一个函数翻转，然后进行滑动叠加，如图 1 所示，在连续情况下，叠加指的是对两个函数的乘积求积分，在离散情况下就是加权求和，为简单起见就统一称为叠加，多次滑动得到的一系列叠加值，构成了卷积函数。

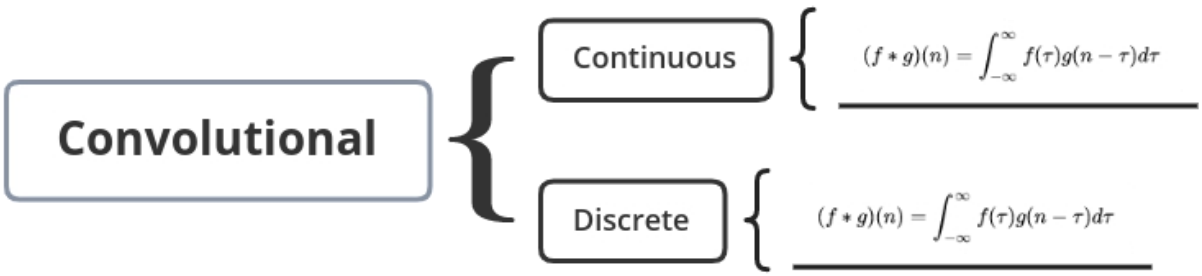

图 1　连续与离散的卷积公式

Fig.1　Continuous and discrete convolution formulas

卷积操作在被应用与图像领域之前，已经在信号分析领域大放异彩。卷积在信号中的定义就是输入信号与原始信号的叠加，之后图片可以用傅里叶变换转换到频域之后，科研人员们就开始思考卷积在图像领域的应用。1980 年的 neocognitron[4]被认为是 CNN 的前身。LeNet 是 LeCun 等人在 1990 年进行的卷积神经网络的开创性工作[5]，后来又对其进行了改进[6]。它是专门为手写数字分类而设计的，并成功地从输入图像中直接识别出视觉模式，而无需任何预处理。但是由于缺乏足够的训练数据和计算能力，这个架构在

复杂的问题上未能表现良好。后来在 2012 年，Krizhevsky 等人[7]提出了一个 CNN 模型，成功地降低了 ILSVRC 比赛的错误率[8]。多年以后，他们的工作已经成为计算机视觉领域中最有影响力的工作之一，并被许多人用来尝试 CNN 架构的变化。

CNN 发展至今，已经和计算机科学有了深度融合，如 CNN 中的翻转操作在图像处理中已经“丢弃”了，因为如果是可训练的卷积核，那么是否翻转是不会影响训练的效果；如果是定义好的算子（如拉普拉斯算子）则可以在定义时就直接做翻转操作，训练时的翻转会浪费大量的计算时间。此次我的文献阅读报告会展示 CNN 发展至今一些经典的论文，其中有 CNN 在不同任务中的独特应用，也有对 CNN 参数优化的论文介绍，最后类比 CNN 的发展历程，简单介绍一下进来火热的图卷积神经网络（Graph Convolutional Network, GCN）的发展现状和经典论文。

2 CNN 文献摘要

2.1 Handwritten Digit Recognition with a Back-Propagation Network[9]

该论文介绍了反向传播网络在手写数字识别中的应用。此方法需要对数据进行最少的预处理，但网络的结构是高度受限的，并且是专门为该任务设计的。网络的输入由独立数字的归一化图像组成，该方法的错误率在当时仅为 1%。

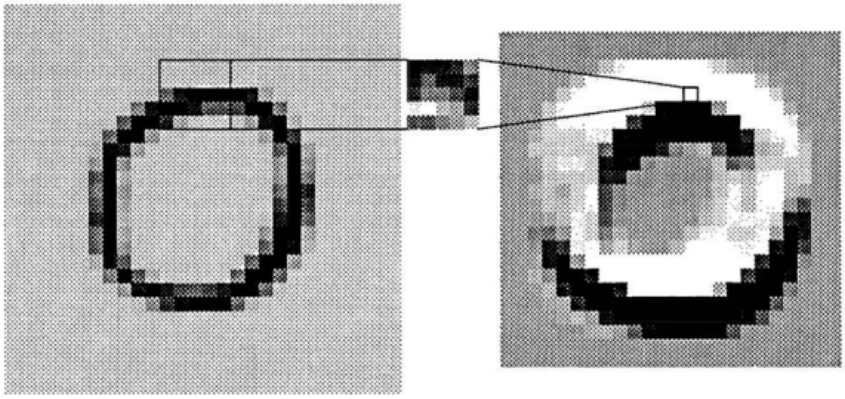

图 2　手写数字局部信息聚合示意图

Fig.2　Handwritten digital local information aggregation schematic

这篇论文可以说是 CNN 在图像领域应用的开山之作，但是当时被没有引入卷积这个概念，那个时代反向传播算法（Back-propagation，BP）被成功地应用于一个大型的、真实世界的任务，并且实验的结果似乎达到了手写数字识别的技术水平是非常震撼的。该网络有许多连接，但自由参数相对较少，网络结构和对权重的限制被设计为将关于任务的几何知识纳入系统。由于它的结构，该网络可以在数据的低水平表示上进行训练，而这些数据的预处理是最少的（相对于精心设计的特征提取）。由于数据的冗余性和对网络的限制，考虑到训练集的大小，学习时间相对较短。缩放特性远远好于人们从较小的人工问题上推断出的反向传播的结果，并且对字母数字字符的初步结果表明，该方法可以直接扩展到更大的任务。当时科研工作人员已经意识到通过共享权重采样聚合的方式可以得到图像的深层语义信息，将这些信息传入全连接层可以得到比原始数据输入更高的准确率，所以从这篇文章开始，拉开了图像领域深度学习和卷积神经网络的序幕。

2.2 AlexNet[10]

在 CNN 的发展历史中，有一篇不得不提的论文——AlexNet，现在的引用量已经达到了 9 万多次，可见其对图像领域的影响程度。在 ImageNet LSVRC-2010 竞赛中，需要将 120 万张高分辨率图像分类到 1000 个不同的类别。在测试数据上，AlexNet 取得了 37.5%的错误率，比之前最好的模型低了 17.0%，大大优于以前的最先进水平。

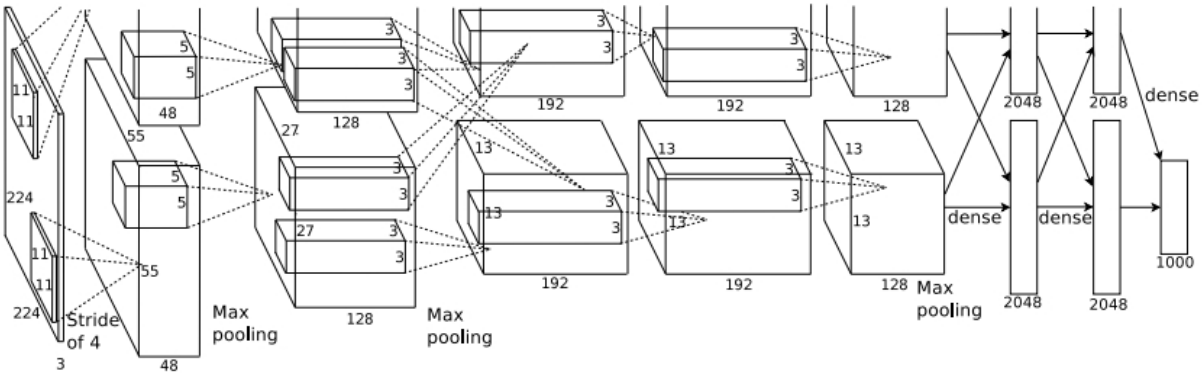

图 3　AlexNet 网络结构图

Fig.3　: An illustration of the architecture of AlexNet

该神经网络有 6000 万个参数和 65 万个神经元，由 5 个卷积层组成，其中一些是 MaxPooling 层，还有 3 个全连接层，最后是 1000 个 softmax 神经元用于分类（神经如图 3 所示）。为了使训练更快，AlexNet 使用了非饱和神经元和一个非常高效的 GPU 实现卷积操作。为了减少全连接层中的过度拟合，AlexNet 采用了 "dropout "的正则化方法，该方法被证明是非常有效的。

AlexNet 采用低采样率把每张图片的分辨率降为 256×256，最重要的是激活函数则是采用了 Relu 函数，有效解决了 Tanh 函数在深层网络中的梯度消失问题，并且采用了 GPU 并行训练、重叠池化、局部响应归一化（针对 Relu 函数），最后还采用了当时比较新颖的 Dropout 正则方法，来减小网络的结构化风险，以此来获得测试集中优秀的表现。

2.3 Pixel Difference Networks for Efficient Edge Detection[11]

该论文主要介绍由 Oulu 大学、国防科技大学、哈工大、西点军校主导的差分卷积（Difference Convolution）工作及其在图像、视频领域中的应用。 相关工作已被 TPAMI, TIP, CVPR'20, ICCV'21 (Oral), IJCAI'21 等顶级期刊会议接收，并斩获两项国际大赛冠亚军（1st Place in the ChaLearn multi-modal face anti-spoofing attack detection challenge with CVPR 2020 和 2nd Place on Action Recognition Track of ECCV 2020 VIPriors Challenges）。边缘检测检测图像中的边缘来确定目标的边界，从而分离感兴趣的目标。近代边缘检测算法起源于一阶导数算子非常经典的有 Sobel[12]，随后出现的二阶算子代表有 Canny[13]，都采用差分信息来表征边缘上下文的突变及细节特征。但是这些基于手工传统算子的模型往往局限于它的浅层表征能力。 随后深度学习的出现给边缘检测注入了新的活力，CNN 通过卷积的深层堆叠，能够有效地捕捉图像的语义特征，在这个阶段 CNN 是作为分类器，用 sigmod 函数来输出一个点是否是边缘点的概率，卷积核扮演了捕捉局部图像模式的作用。 但是原始 CNN 在对卷积核的初始化过程中并没有显式的梯度编码限制，使其在训练过程中很难聚焦对图像梯度信息的提取，从而影响了边缘预测的精度。像素差分网络（Pixel Difference Networks，PiDiNet）则很好的融合了传统的边缘检测算子和深度神经网络。

PiDiNet 结合传统边缘检测算子 LBP，提出了 3 种差分卷积算子（Pixel Difference Convolution, PDC），如图 4 所示。

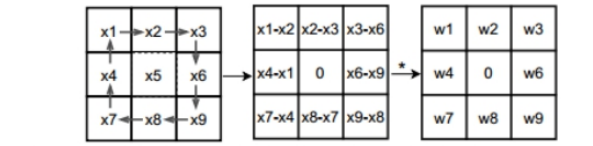

Figure 4. Selection of pixel pairs and convolution in APDC.

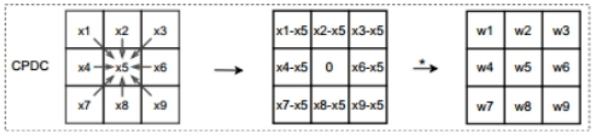

Figure 10. Selection of pixel pairs and convolution in CPDC.

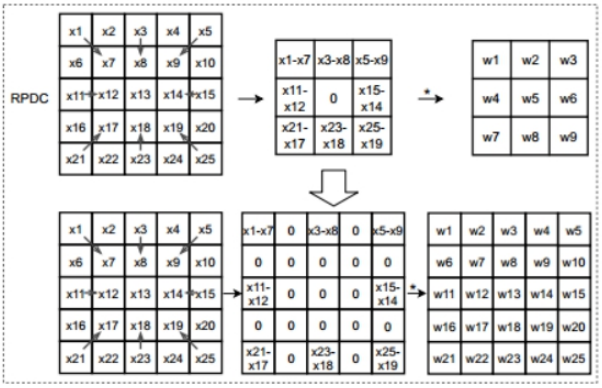

Figure 12. Selection of pixel pairs and convolution in RPDC.

图 4　PDC 算子

Fig.4　: Pixel Difference Convolution operator

PDC 思想就是将传统算子的差分思想带入到深度神经网络中，即在图像进入卷积层之前就进行差分操作来得到浅层的边缘信息，但是在一张千万级像素的图片中做差分的代价是很大的，PiDINet 就开创新的提出直接在卷积核上做差分操作，如图 5 所示，如此在一次卷积操作中就可以节省下上千万次的减法操作，如果递推到整个训练和预测过程中，这个差分思想带来的时间收益是非常巨大的。

$$y = f(\boldsymbol{x}, \boldsymbol{\theta}) = \sum_{i=1}^{k \times k} w_i \cdot x_i, \qquad \text{(vanilla convolution)} \quad (5)$$

$$y = f(\nabla \boldsymbol{x}, \boldsymbol{\theta}) = \sum_{(x_i, x_i') \in \mathcal{P}} w_i \cdot (x_i - x_i'), \qquad \text{(PDC)} \quad (6)$$

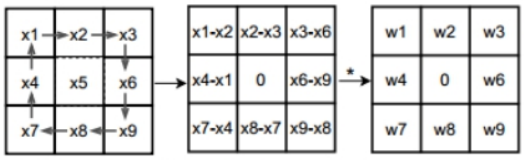

Figure 4. Selection of pixel pairs and convolution in APDC.

$$\begin{aligned} y =& w_1 \cdot (x_1 - x_2) + w_2 \cdot (x_2 - x_3) + w_3 \cdot (x_3 - x_6) \\ &+ w_4 \cdot (x_4 - x_1) + w_6 \cdot (x_6 - x_9) + w_7 \cdot (x_7 - x_4) \\ &+ w_8 \cdot (x_8 - x_7) + w_9 \cdot (x_9 - x_8) \\ =& (w_1 - w_4) \cdot x_1 + (w_2 - w_1) \cdot x_2 + (w_3 - w_2) \cdot x_3 \\ &+ (w_4 - w_7) \cdot x_4 + (w_6 - w_3) \cdot x_6 + (w_7 - w_8) \cdot x_7 \\ &+ (w_8 - w_9) \cdot x_8 + (w_9 - w_6) \cdot x_9 \\ &+ 0 \cdot x_5 \\ =& \hat{w}_1 \cdot x_1 + \hat{w}_2 \cdot x_2 + \hat{w}_3 \cdot x_3 + \ldots = \sum \hat{w}_i \cdot x_i \end{aligned} \quad (8)$$

图 5　PiDiNet 差分卷积核

Fig.5　: PiDiNet Difference Convolution core

并且 PiDiNet 在图像领域复杂网络当道的如今，提出了一种轻量级的图像边缘检测网络框架，如图 6

所示，在运行效率和吞吐量上相比之前的 BDCN[14]、RCF[15]的边缘检测模型有非常大的提升。

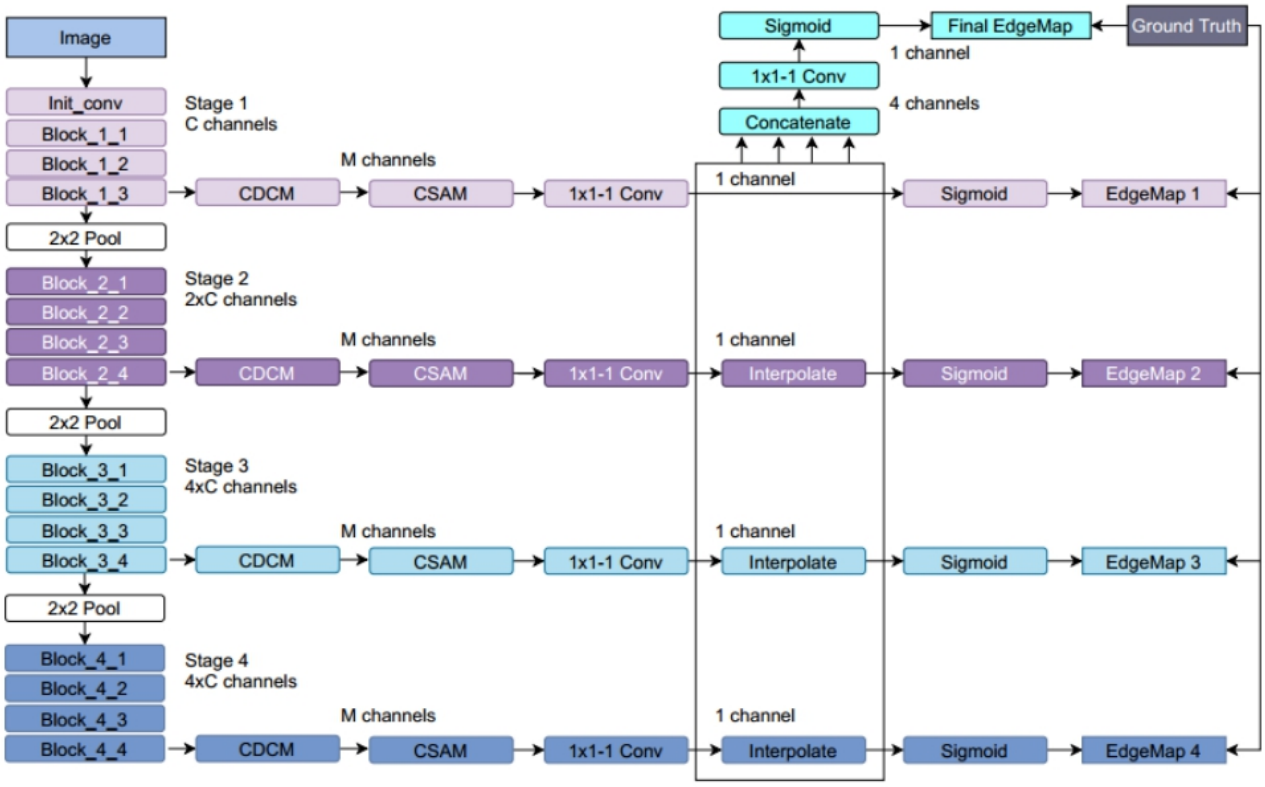

图 6　PiDiNet 网络框架

Fig.6　: PiDiNet architecture

PiDiNet 给出了 CNN 新的方向，其实其算法思想并不复杂（将像素差分转换成卷积核的差分），但是在 CNN 发展至今中，PiDiNet 这种改造卷积核的思想在未来 CNN 的发展历程中会有很重要的作用，并且结合传统的边缘检测算子和轻型的网络架构不禁引起我们的反思，传统基于数学的方法在现在深度学习火热的阶段还是有很高的借鉴意义，复杂的网络结构确实能得到好的结果，但是同时也需要兼顾网络的运行效率，轻量级网络架构是科研人员应该追求的方向。

3 GNN 文献摘要

像上文中提到，CNN 快速发展，并借由其强大的建模能力引起了广泛关注，相比传统方法，卷积神经网络的引入给图像处理带来了很大的提升。但是传统的卷积神经网络只能处理欧式空间数据（如图像、文

本、语音），这些领域的数据具有平移不变性。平移不变形使得我们可以再输入数据空间定义全局共享的卷积核，从而定义卷积神经网络，但是图数据不具有平移不变性，所以卷积操作在非欧式空间数据中遇到了前所未有的挑战。近来由于图数据的普及性，研究人员开始关注如何在图上构造端到端的深度学习模型。借助于卷积神经网络对局部结构的建模能力及图上普遍存在的节点依赖关系，图卷积神经网络称为了其中最活跃最重要的一支。

图卷积神经网络中主要分为 2 类，一类是有严格数学支持的谱域图卷积神经网络，其核心思想是将图结构通过拉普拉斯矩阵和傅里叶变换转换到频域，进行卷积操作后转回空域的图卷积操作；另一类更符合计算机工程的思想，将 GCN 类比 CNN，定义了 GCN 的采样（确定邻居节点）和聚合（与邻居节点聚合信息）两个问题，没有严格的数学证明，但是应用起来会比谱域方法更加自由（如可以应用于有向图的卷积操作）。

3.1 谱域图卷积

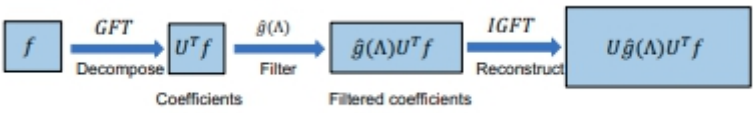

图 7　谱域图卷积流程

Fig.7　: Spectral domain graph convolution process

如图 7 所示，图谱域卷积的思想是调制图信号的频率，使得其中一些频率分量被保留或者放大，而另一些频率的分量被移除或者减少。因此给定一个图信号，首先需要对其进行图的傅里叶变换(Graph Fourier Transform, GFT)，以获得它的图傅里叶系数，然后对这些系数进行调制，再在空域中重构该信号。

2013 年发表的第一代 GCN[16]存在很多缺陷，计算复杂，非局部连接，且在巨型网络中因为参数的限制具有不可行性，所以一直没有得到很好的普及和应用，2017 年发表的第二代 GCN[17]利用了切比雪夫等式来简化 GCN 的计算过程，且只聚合 K 层的邻居节点，最后 GCN 演变为一阶切比雪夫的图卷积。

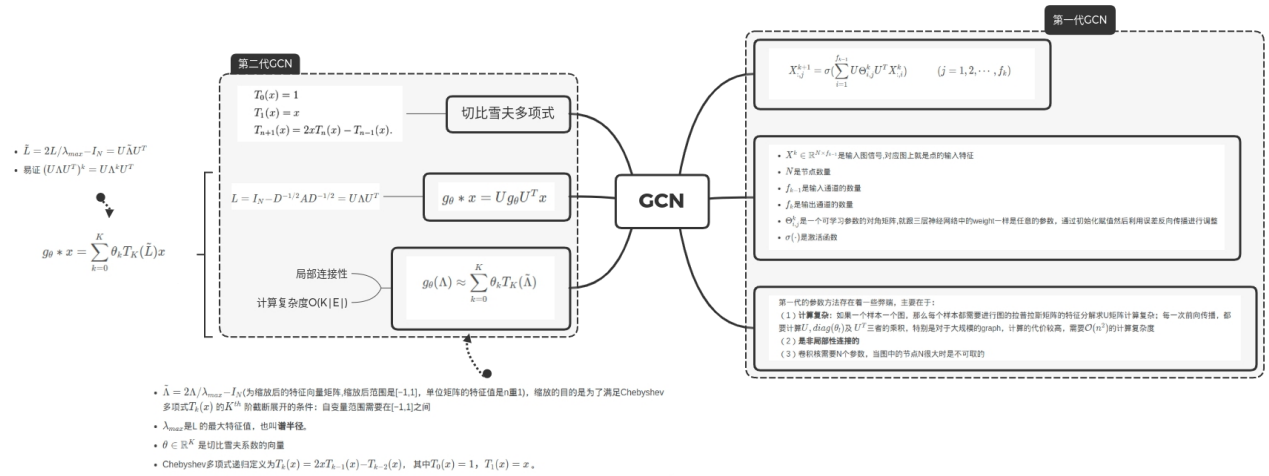

图 8　谱域图卷积发展历程

Fig.8　: Spectral domain graph convolution development

第二代的 GCN 是权值共享的，且具有局部性 Local Connectivity，也就是局部连接的，因为每次聚合的只是一阶邻居，感受野正比于卷积层层数，第一层的节点只包含与直接相邻节点有关的信息，第二层以后，每个节点还包含相邻节点的相邻节点的信息，这样的话，参与运算的信息就会变多。层数越多，感受野越大，参与运算的信息量越充分。也就是说随着卷积层的增加，从远处邻居的信息也会逐渐聚集过来，同时复杂度大大降低，不用再计算拉普拉斯矩阵，特征分解，所以可以很好的普及到各个应用领域。但是基于谱域的 GCN 在正常情况下无法处理有向图，并且巨型图中的扩展性较差。

3.2　空域图卷积

信息传递神经网络(MPNNs)概述了基于空间的卷积神经网络的一般框架[18]。它把图卷积看作一个信息传递过程，信息可以沿着边直接从一个节点传递到另一个节点。空域卷积更符合计算机的思维，其绕开了图谱的理论，无需将信号在空域和频域之间转换，直接在空域上进行定义和操作，更加直观和清晰，且没有了图谱理论的限制，定义更加灵活，并且可以应用在有向图中，但是对比谱域方法，缺失了数学理论的支撑。

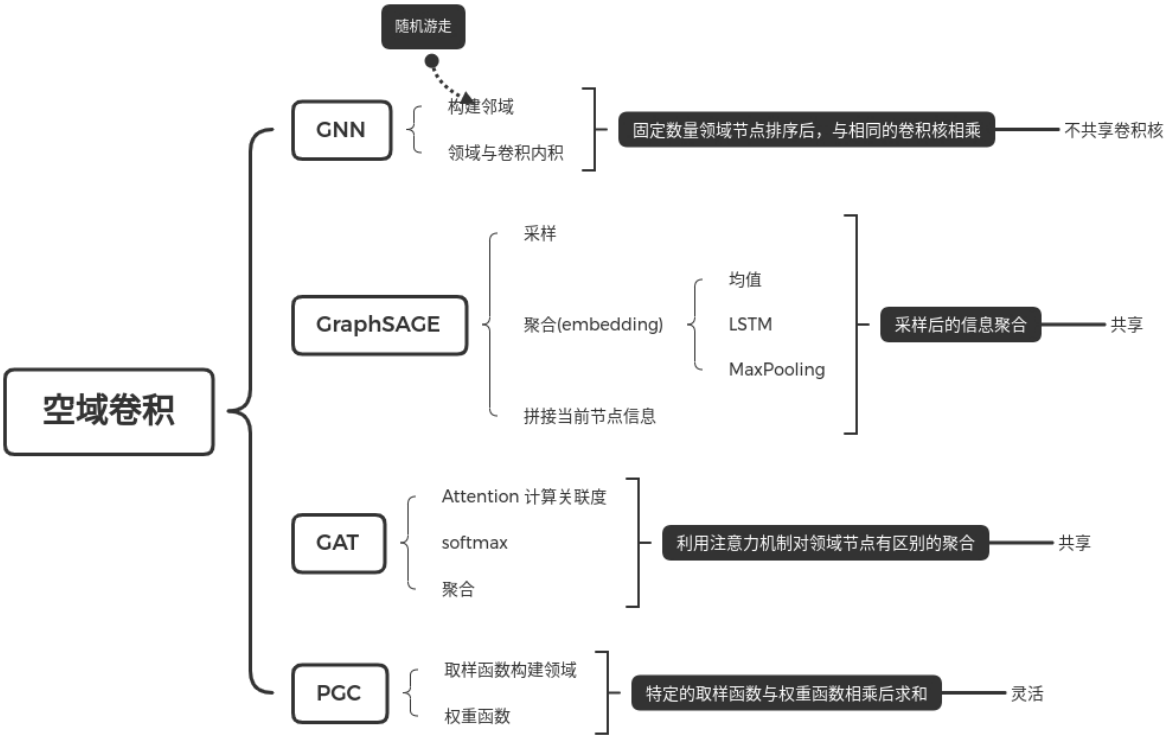

图 9　空域图卷积方法

Fig.9　: Space domain graph convolution method

在图 9 中罗列了现在常用的空域图卷积模型。空域卷积其实仅需要回答两个问题：1）如何确定邻居 ；2）如何与邻居进行聚合。GNN[19]采用随机游走的方式来选取需要聚合的邻居节点，同时邻居节点需要排序，所以图卷积就可以转换成普通的 CNN，但是卷积参数不共享；GraphSAGE[20] 采用均匀采样，所以邻居节点不需要排序，同时也实现了参数共享卷积；GAT[21]则是引入了注意力机制，直接采用一阶近邻节点，使用注意力来修正共享的卷积参数；PGC[24]则是定义了一个抽象的采样和聚合的算法。同时 GNN、GraphSAGE、GAT 不需要训练集和测试集的图结构需要相同，PSG 则是要考虑具体的采样和聚合函数。

空域卷积操作虽然没有严格的数学支持，但是它可以处理有向图和异质图，这是空域卷积的优势，同时空域卷积不需要计算拉普拉斯矩阵和傅里叶变换，所以在卷积时间上也比谱域卷积更有优势。

4 我对 CNN 与 GNN 发展的思考

由于图卷积会使相邻节点的表征更加接近，理论上，如果有无限多的图卷积层，所有节点的表征将收敛到一个点[23]，这种现象与 CNN 类似。这就提出了一个问题：对于学习图数据来说，深层网络结构是否仍然是一个好的策略。

不同于 CNN 可以通过堆叠非常多层数的神经网络来取得更好的模型表现，因为现在图片的数据像素特别多，但是图结构数据大多不会很大，想象一下上亿像素的图片和上亿节点的图结构，明显前者是普遍存在，而后者是很难获取批量的上亿节点的图结构数据的，且邻接矩阵和上亿节点特征的存储也是不可想象的，同时，图结构比图像的集聚系数更大，如课程中提到的小世界网络，在这种网络中大部分的结点不与彼此邻接，但大部分结点可以从任一其他点经少数几步（如 6 跳）就可到达（若将一个小世界网络中的点代表一个人，而连结线代表人与人认识，则这小世界网络可以反映陌生人由彼此共同认识的人而连结的小世界现象），所以普通 GNN 会比 CNN 聚合节点的信息更快。

在 CNN 中可以使用多层 CNN 来捕获图像深层的语义信息，而 GNN 则不行。但是 GNN 的采样、聚合的思想和 CNN 是相同的，两者都是卷积操作，一个在图结构数据，一个在图像数据，所以 GNN 之后的发展历程也会与 CNN 一样，当 GNN 成熟之后也会像 CNN 一样称为一个基础的模型出现在各个模型中，并且多模态模型的出现，让处理不同输入数据的网络结构有了结合的可能性。

致谢 感谢李晓强教授在 10 周内的数字图像处理课程的授课，我在其中受益匪浅，特别是后面课程中的论文分享课程，我不仅学到了我所看的那篇论文方法，还可以听到同学们对各自论文的理解与分享，打开了我的眼界，再次特别致谢李晓强老师的授课！

参考文献：

[1] C. Szegedy, A. Toshev and D. Erhan, Deep Neural Networks for object detection, Advances in Neural Information Processing Systems, 2013, 26: 2553-2561.

[2] Ramachandran R, Rajeev DC, Krishnan SG, P Subathra, Deep learning an overview, IJAER, Volume 10, Issue 10, 2015, Pages

25433-25448.

[3] J. Fan, W. Xu, Y. Wu, and Y. Gong, Human tracking using convolutional neural networks, Neural Networks, IEEE Transactions, 2010.

[4] K. Fukushima, Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position, Biological cybernetics, 1980.

[5] B. B. Le Cun, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard,and L. D. Jackel, Handwritten digit recognition with a backpropagation network, in NIPS. Citeseer, 1990.

[6] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, Gradient-based learningapplied to document recognition, Proceedings of the IEEE, 1998.

[7] Alex Krizhevsky, Sutskever I, and Hinton G.E, Imagenet classification with deep convolutional neural networks. In NIPS, 2012.

[8] A. Berg, J. Deng, and L. Fei-Fei, Large scale visual recognition challenge 2010, www.image-net.org/challenges. 2010.

[9] LeCun Y, Boser B, Denker J, et al. Handwritten digit recognition with a back-propagation network[J]. Advances in neural information processing systems, 1989, 2.

[10] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[J]. Advances in neural information processing systems, 2012, 25: 1097-1105.

[11] Su Z, Liu W, Yu Z, et al. Pixel Difference Networks for Efficient Edge Detection[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 5117-5127.

[12] Duda R O, Hart P E. Pattern classification and scene analysis[M]. New York: Wiley, 1973.

[13] Canny J. A computational approach to edge detection[J]. IEEE Transactions on pattern analysis and machine intelligence, 1986 (6): 679-698.

[14] Jianzhong He, Shiliang Zhang, Ming Yang, Yanhu Shan, and Tiejun Huang. Bi-directional cascade network for perceptual edge detection. In CVPR, pages 3828–3837, 2019. 2, 5, 7, 8

[15] Yun Liu, Ming-Ming Cheng, Xiaowei Hu, Jia-Wang Bian, Le Zhang, Xiang Bai, and Jinhui Tang. Richer convolutional features for edge detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 41(8):1939–1946, 2019.

[16] Y. Li, R. Yu, C. Shahabi, and Y. Liu, "Diffusion convolutional recurrent neural network: Data-driven traffic forecasting," in Proc. ICLR, 2018, pp. 1–16.

[17] Defferrard M, Bresson X, Vandergheynst P. Convolutional neural networks on graphs with fast localized spectral filtering[J]. Advances in neural information processing systems, 2016, 29: 3844-3852.、

[18] Gilmer J, Schoenholz S S, Riley P F, et al. Neural message passing for quantum chemistry[C]//International conference on machine learning. PMLR, 2017: 1263-1272.

[19] Scarselli F, Gori M, Tsoi A C, et al. The graph neural network model[J]. IEEE transactions on neural networks, 2008, 20(1): 61-80.

[20] W. Hamilton, Z. Ying, and J. Leskovec, "Inductive representation learning on large graphs," in Proc. NIPS, 2017, pp. 1024–1034.

[21] Veličković P, Cucurull G, Casanova A, et al. Graph attention networks[J]. arXiv preprint arXiv:1710.10903, 2017.

[22] D. V. Tran, N. Navarin, and A. Sperduti, "On filter size in graph convolutional networks," in Proc. IEEE Symp. Ser. Comput. Intell. (SSCI), Nov. 2018, pp. 1534–1541.

[23] Q. Li, Z. Han, and X.-M. Wu, "Deeper insights into graph convolutional networks for semi-supervised learning," in Proc. AAAI, 2018, pp. 1–8.