周报（2023. 9.12-2023.9.20）姓名：孙瑞阳

每日小结

	周一	周二	周三	周四	周五
早	学 graph	看代码，学 graph	看文献和代码	Graph 代码	Graph 代码
中	Graph 代码	学 graph	看文献和代码	Graph 代码	Graph 代码
晚	Graph 代码	组会		Graph 代码	

注：简单表述当前时间段工作，如看文献 1，整理数据等

科研详情

文献阅读

文献 1

题目：A Survey on Graph Neural Networks and Graph Transformers in Computer Vision

作者：Chaoqi Chen, Yushuang Wu, Qiyuan Dai, Hong-Yu Zhou, Mutian Xu, Sibei Yang, Xiaoguang Han, and Yizhou Yu

出处：Arxiv 2022

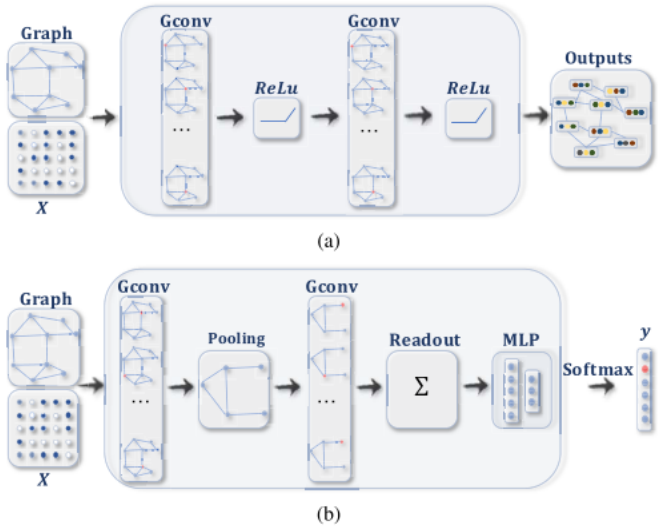
方法：

本综述对计算机视觉中基于图神经网络（包括图 Transformer）的方法和最新进展进行了全面且详细的调研。根据输入数据的模态将图神经网络在计算机视觉中的应用大致划分为五类：自然图像（二维）、视频、视觉+语言、三维数据（例如，点云）以及医学影像。

并提供了一种分类方法，将 GNN 分为四类：RecGNN、ConvGNN、GAE 和 STGNN，并对这几类 GNN 进行了全面的回顾、比较和总结。然后介绍了 GNN 的广泛应用：总结了 GNN 的数据集、开源代码和模型评估。

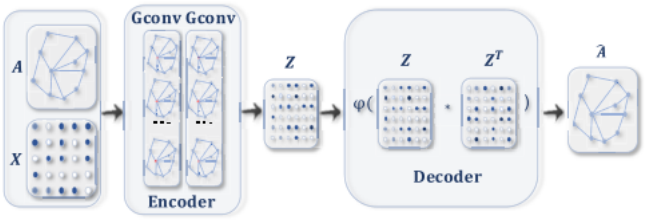
（1）**Recurrent Graph Neural Networks**：GNN 的先驱，其目的是学习**具有循环神经结构的节点表示，RecGNN 假设图中的一个节点不断地与它的邻居交换信息/消息，直到达到稳定的均衡。**

（2）**Convolutional Graph Neural Networks**：ConvGNN 将网格数据的卷积运算推广到 Graph 数据。主要思想：聚合节点 自身的特征 和其邻居的特征 来生成节点 的表示。与 RecGNN 不同，ConvGNN 通过**堆叠多个图卷积层来提取节点表示**，如下所示：



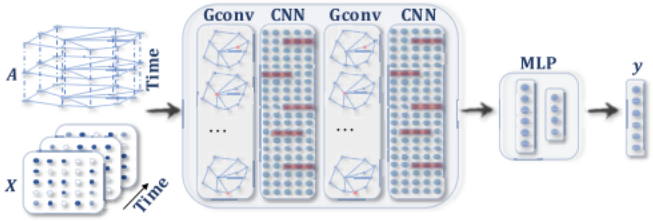
3）**Graph Autoencoders**：图自编码器，**GAE 将节点/图编码到一个潜在的向量空间中，然后从编码信息中重构图数据。**GAE 一般用于学习网络嵌入和图生成：**在网络嵌入方面，GAE 通**

过重构图结构信息(如图邻接矩阵)来学习潜在节点表示。对于图的生成，有些方法是逐步生成图的节点和边，而另一些方法是一次性输出一个图。下图是用于网络嵌入的 GAE：



(c)

（4）**Spatial–Temporal Graph Neural Networks**：时空 GNN，旨在从时空图中学习隐藏模式，如交通速度预测、驾驶员机动预测和人类行为识别。STGNN 的关键思想：**同时考虑空间依赖性和时间依赖性。**目前许多方法将图卷积与 RNN 或 CNN 结合以捕获空间依赖性来建模时间依赖性。下图是用于时空图预测的 STGNN：



(d)

启发：

1. 结合综述在看图的一些基础理论知识，考虑了自己数据集上两种 graph 的建立方式
2. 觉得 Transformer 可以理解为一种图神经网络
3. 找了几个论文里提到的基础算法 demo 跑了一下：deepwalk 和 node2vec

文献2

题目：Multi-Modal Reasoning Graph for Scene-Text Based Fine-Grained Image Classification and Retrieval

作者：Andres Mafla Sounak Dey Ali Furkan Biten Lluis Gomez Dimosthenis Karatzas Computer Vision Center, UAB, Spain

出处：Proceedings of the IEEE/CVF winter conference on applications …, 2021

方法：

作者设计一个完全端到端可训练 pipeline，融合了多模态推理模块，结合文字和视觉特征，且不依赖于集合模型或预先计算的特征。通过图像的文本和视觉特征，本文同时考虑图像的全局信息、局部区分性特征。不仅提取图像的场景文本特征，而且利用图像中的通用目标信息，联合通用目标和场景文本共同推理、分析图像内容。

如图所示，通过 ResNet152 提取图像的全局信息，Faster-RCNN 提取图像中的通用目标特征。之后，将通用目标特征和场景文本实例特征输入图卷积神经网络，推理分析出增强后的特征。将增强后的的特征和图像全局特征一起输入给分类器进行分类。在两个数据集中大大超越了以往最先进的结果，在细粒度分类上超过 5%，在图像检索上超过 10%。

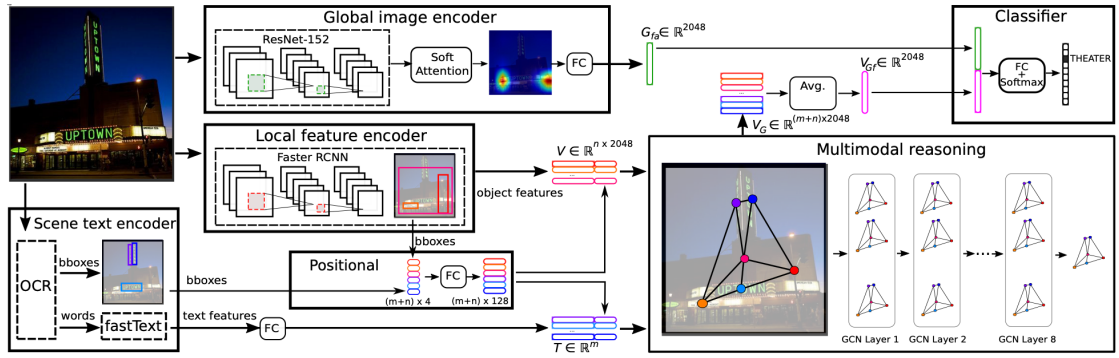


Figure 2. Detailed model architecture. The proposed model combines features of regions of scene text and visual salient objects by employing a graph-based Multi-Modal Reasoning (MMR) module. The MMR module enhances semantic relations between the visual regions and uses the enriched nodes along with features from the Global Encoder to obtain a set of discriminatory signals for fine-grained classification and retrieval.

启发：

在我的工作中，这种将增强特征输入 GNN 的思路也很好，值得借鉴，前面也可以原使用 transformer 进行特征增强提取，通过 GNN 融合多模态特征，代码在看。

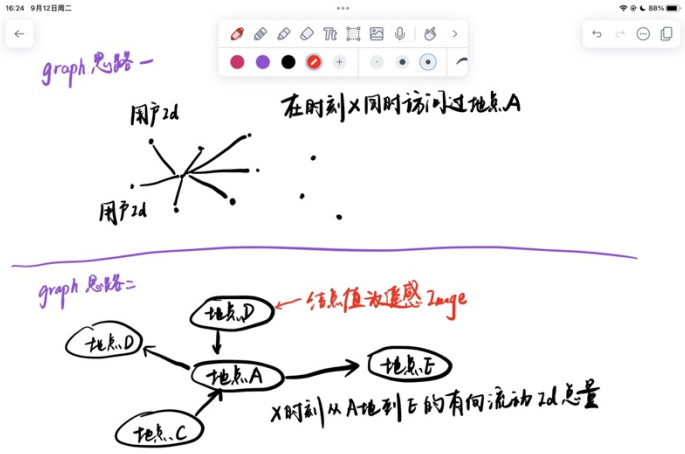
工作进展

1：阅读文献；

2: 跑了几个 GNN 的 demo：deepwalk 和 node2vec 算法的

3:GAT 的论文和代码在找，想看一下 Graph attention 的论文

4:考虑了两种自己数据集的 graph 的建立：目前觉得第二种更好一些，还在想其他的建立方式



5:在自己的数据集上建立思路一的 graph 代码，代码遇到了一些问题，还在解决。

下周计划

1. 在找多模态 GAT 的论文和代码
2. 看 graph 的建立有没有其他思路
3. 修改论文