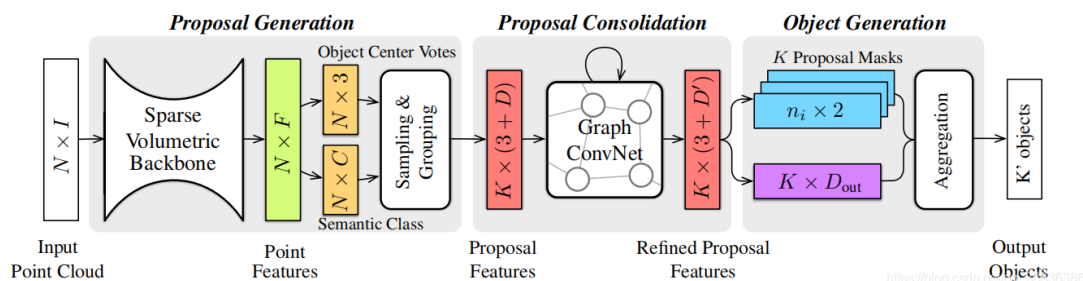


GNN 计算机视觉应用

1、3D-MPA: Multi Proposal Aggregation for 3D Semantic Instance Segmentation

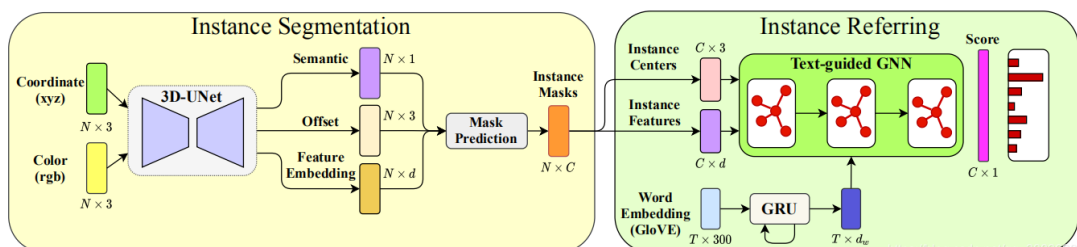
3D 点云语义实例分割

将 proposal 表示为 proposal 图中的节点，然后输入到 GCN 中，来挖掘 proposal 间的高阶相互作用



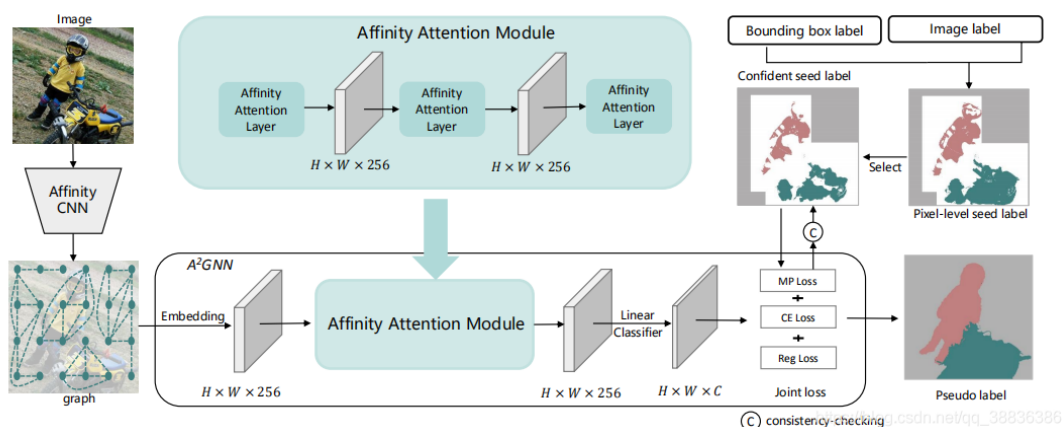
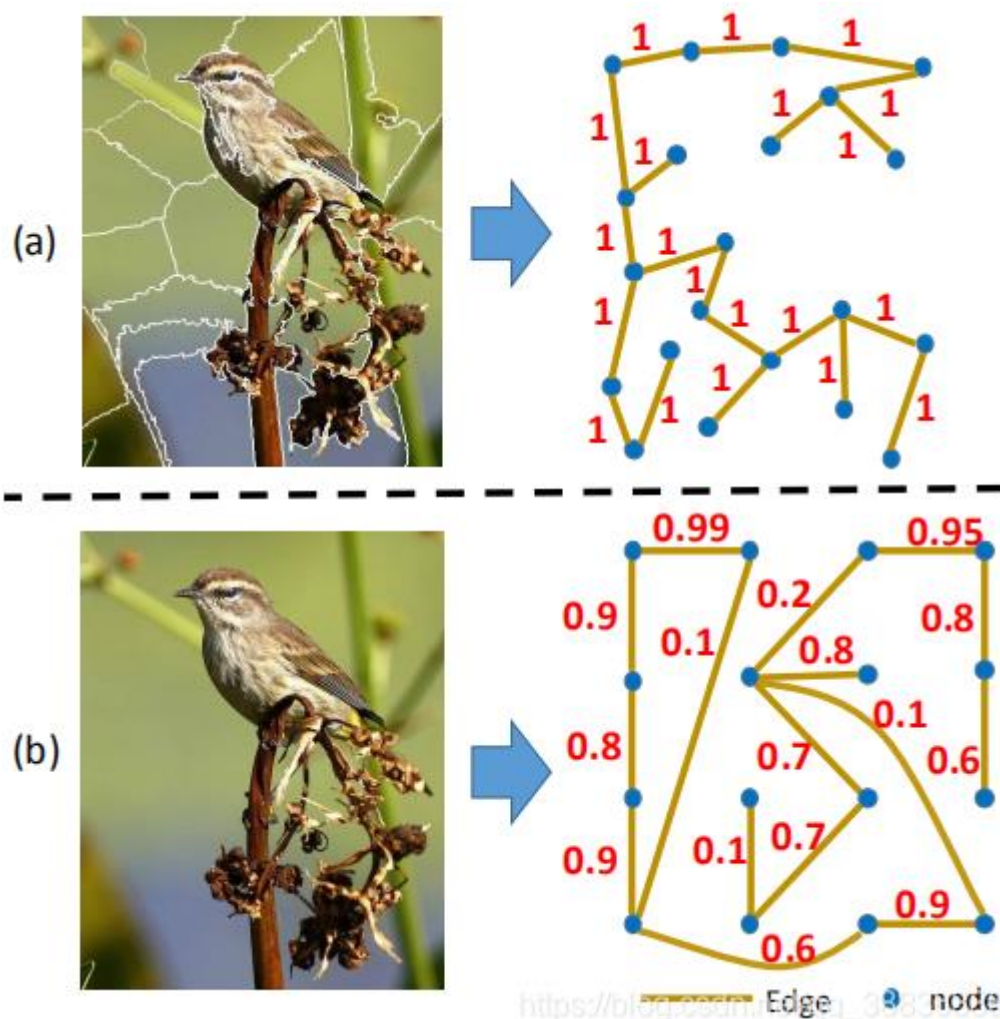
2、Text-Guided Graph Neural Networks for Referring 3D Instance Segmentation

摘要：本文提出了一项称为参考三维实例分割的新任务，目的是在给定一个查询句子的三维场景中分割出目标实例。之前关于场景理解的工作已经在 nlp 的指导下探索了视觉基础，但重点主要局限于图像和视频。我们提出了一种文本引导的图神经网络 TGNN，用于参考点云上的三维实例分割。给定一个查询句子和三维场景的点云，我们的方法学习提取每个点的特征，并预测一个偏移，将每个点移动到它的对象中心。基于点特征和偏移量，我们对点进行聚类，生成候选对象的融合特征坐标。所得到的集群被建模为图神经网络中的节点，以学习包含每个候选对象的关系结构表示。GNN 层利用每个对象的特征及其与邻居的关系来生曾输入句子表达式的注意力热图。最后，使用注意力热图来引导来自邻域节点的信息聚合



3、Affinity Attention Graph Neural Network for Weakly Supervised Semantic Segmentation

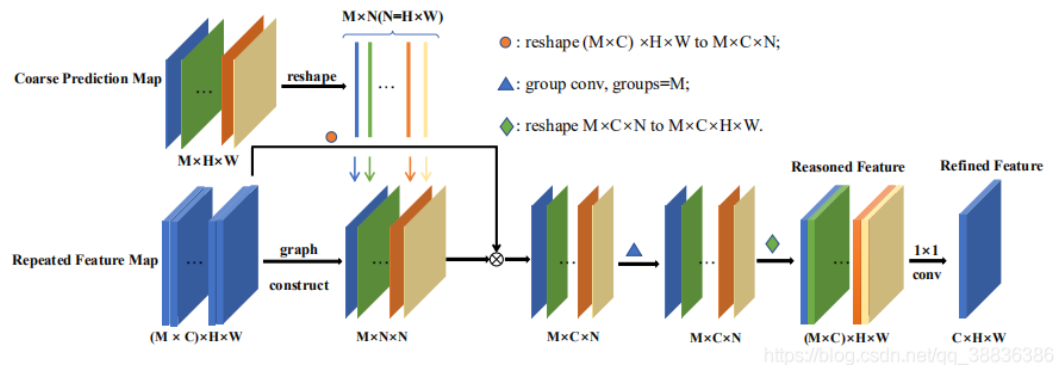
摘要：弱监督语义分割因其人工注释成本低而收到广泛关注。在本文中，我们的目标是处理边界框监督的语义分割，即使用边界框注释作为监督来训练精确的语义分割模型。为此，我们提出亲和图注意力网络 A2GNN，根据之前的实践，我们首先生成伪语义感知种子，然后根据我们提出的亲和卷积神经网络将这些种子形成语义图然后将所构建的图输入到 A2GNN 种，其中设计了一个亲和注意力层来从 soft graph edge 获取长短距离信息，以准确将语义标签从种子传播到未标记的像素中，然而，为了保证种子的精确度，我们只为 A2GNN 采用了有限数量的可信像素种子标签，这可能会导致对训练的监督不足，为了缓解这一问题，我们进一步引入了一种新的损失函数和一种一致性检查机制利用边界框约束，以便为模型优化提供更可靠的指导。



4、Class-wise Dynamic Graph Convolution for Semantic Segmentation

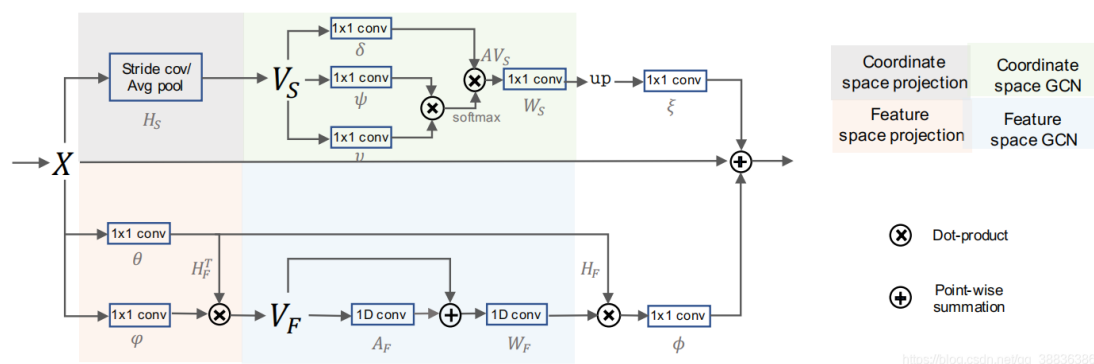
摘要：最近一些工作在语义分割方面取得了巨大的进展，通过利用局部或全局方式的上下文信息，以及扩展的卷积、pyramid pooling 或者自注意力机制。为了避免在以前的工作中出现潜在的误导性上下文信息聚合，我们提出了一个面向类的动态图卷积模块来自适应地传播信息图推理在同一类中的像素之间执行，基于所提出的 CDGC 模块，我们进一步引入了类动态图卷积网络 CDGCNet，它由 CDGC 模块和基本的分割网络主要部分组成，形成了一个粗到细的范式。具体地说，CDGC 模块将粗分割结果作为 class mask 来提取节点特征进行图

构建，并对构造图进行动态图卷积，学习特征聚合和权重分配。然后将细化特征和原始特征进行融合，得到最终的预测



5、Dual Graph Convolutional Network for Semantic Segmentation

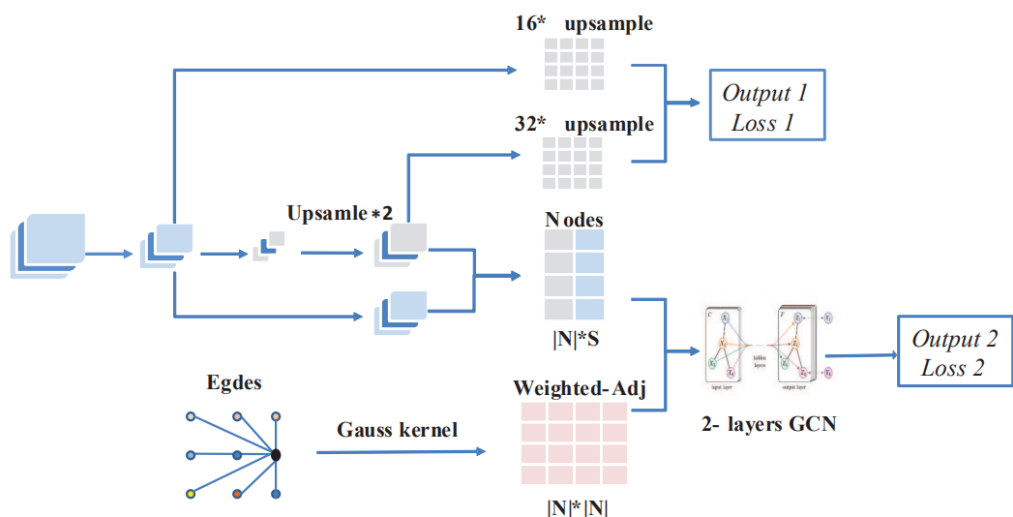
摘要：利用远程上下文信息是语义分割等像素级预测任务的关键，与之前使用多尺度特征融合或扩张卷积的工作相比，我们提出了一种新的图卷积网络来解决这个问题。我们的对偶图卷积网络通过在一个框架中建模两个正交图来建模输入特征的全局上下文信息。第一个组件建模图像中像素之间的空间关系，而第二个组件则沿着网络特征图的通道维度建模相互依赖关系。这是通过将特征投影到一个新的低维空间中来有效实现的，在重新投影到原始空间之前，所有的两对相互作用都可以被建模



6、Graph-FCN for image semantic segmentation

发表在 CVPR2020

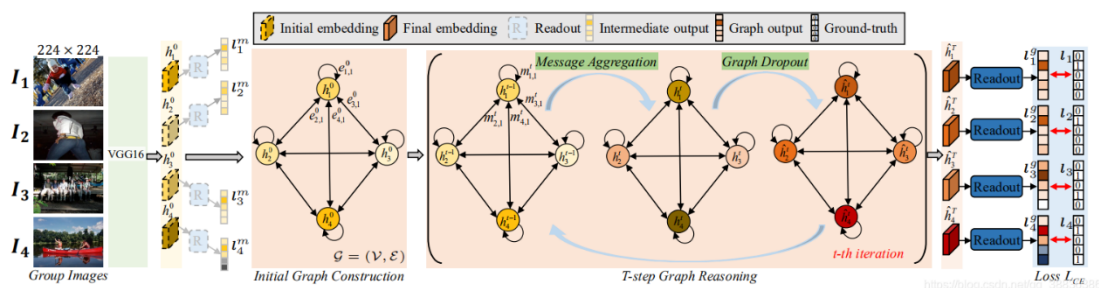
摘要：深度学习的语义分割在图像像素分类方面取得了很大的进展。然而，在深度学习的高级特征提取中，通常会忽略局部位置信息，这对图像语义分割具有重要意义。为了避免这个问题，我们提出一个名为 Graph-FCN 的全卷积网络初始化图模型，用于图像分割。首先通过卷积网络将图像网络数据扩展到图结构数据中，将语义分割问题转化为图节点分类问题。然后，我们应用图卷积网络来解决这个图节点的分类问题。据我们所知，这是我们第一次将图卷积网络用于图像语义分割。与原始的 FCN 模型相比，我们的方法在 VOC 数据集上的平均交集(miOU)上取得了竞争性的性能（提高了约 1.34%）。



https://blog.csdn.net/qq_38836386

7、Group-Wise Semantic Mining for Weakly Supervised Semantic Segmentation

摘要:由于深度学习渴望数据的特性, 获得足够的真实标签来监督训练深度视觉模型一直是一个瓶颈。比如语义分割, 它需要像素级的注释。这项工作解决了弱监督语义分割, 其目标时弥合图像级注释和像素级分割之间的差距。我们将 WSSS 表示为一种新的图式学习任务, 明确地建模图像种的语义依赖, 以估计更可靠的真实标签, 可用于训练更准确的分割模型。特别地, 我们设计了一个图神经网络, 其中输入图像表示为图节点, 一对图像之间的潜在关系具有一种有效的协同注意机制。此外, 为了防止模型过度关注公共语义, 我们进一步提出了一个 graph dropout 层, 鼓励模型学习更准确和完整的对象。整个网络课通过迭代消息传递机制进行端到端训练, 从而在图像上传播交互信息, 以逐步提高性能

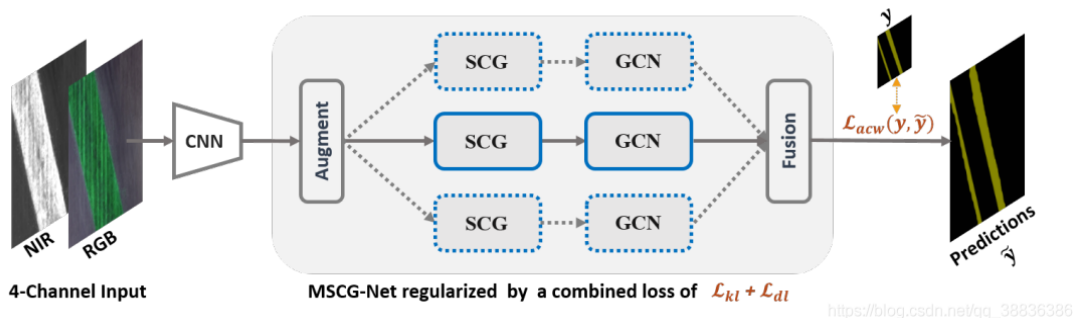


8、LoGAN: Latent Graph Co-Attention Network for Weakly-Supervised Video Moment Retrieval

弱监督的视频实时检索的隐式图协同注意力网络

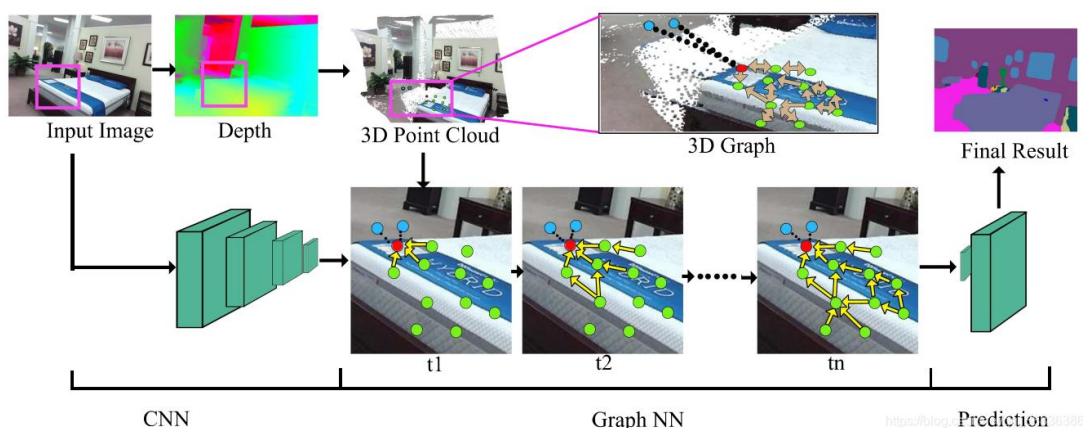
9、Multi-view Self-Constructing Graph Convolutional Networks with Adaptive Class Weighting Loss for Semantic Segmentation

摘要: 我们提出了一种新的语义分割架构, 称为多视图自构造图卷积网络语义分割。基于最近所提出的自构造图 SCG 模块, 它利用可学习的潜在变量直接输入特征自构建底层图, 而不依赖于手动构建的先验知识图, 我们利用多个视图来显式地利用图像中的旋转不变性我们进一步提出了一个自适应的类加权损失来解决类的不平衡问题我们在农业视觉挑战数据集上演示了所提出的方法的有效性和灵活性, 与相关的基于纯 CNN 的工作相比, 我们的模型取得了非常有竞争力的结果(0.547mIOU)。代码可在: github.com/samleoqh/MSCG-Net 上获得



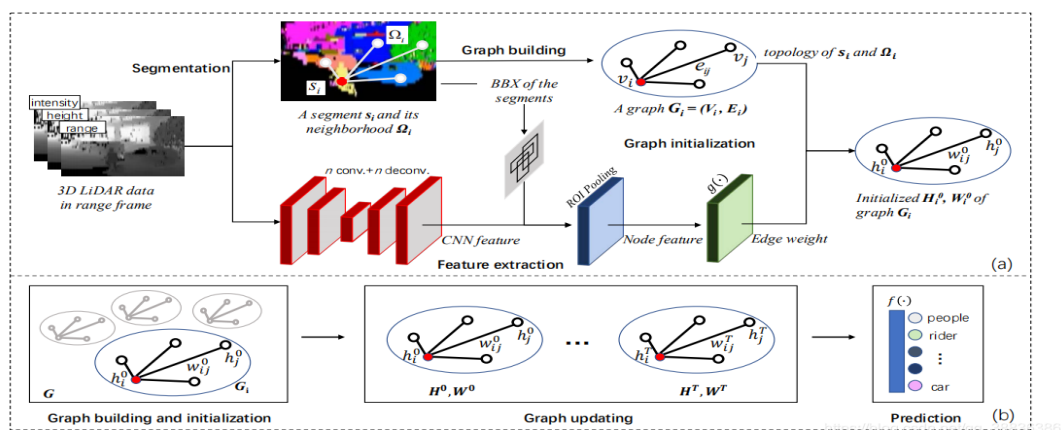
10、3D Graph Neural Networks for RGBD Semantic Segmentation

摘要：RGBD 语义分割需要对二维外观和三维几何信息的联合推理，**在本文中，我们提出了一个三维图神经网络。它早三维点云之上构建了一个 k-近邻图，图中的每个节点对应于一组点，并与由一元 CNN 从二维图像中提取的外观特征初始化的隐藏表示向量相关联。依赖于循环函数，每个节点根据当前状态和来自邻居的传入消息动态更新其隐向量表示。**该传播模型展开了一定的时间步长，最终的每个节点表示用于预测每个像素的语义类



11、Scene Context Based Semantic Segmentation for 3D LiDAR Data in Dynamic Scene

摘要：我们提出了一种基于图神经网络的方法来结合场景上下文进行三维激光雷达数据的语义分割，该问题被定义为构建一个图来表示中心段及其邻域的拓扑，然后推断分割标签。图的节点由范围图像上的片段生成，既适用于稀疏点云，又适合密集点云。评估中心节点及其邻域相关性的边缘权值由神经网络自动编码，因此相邻节点的数量不再是一个敏感参数。设计了一个包括语义生成、图生成、边权估计、节点更新和节点预测系统。

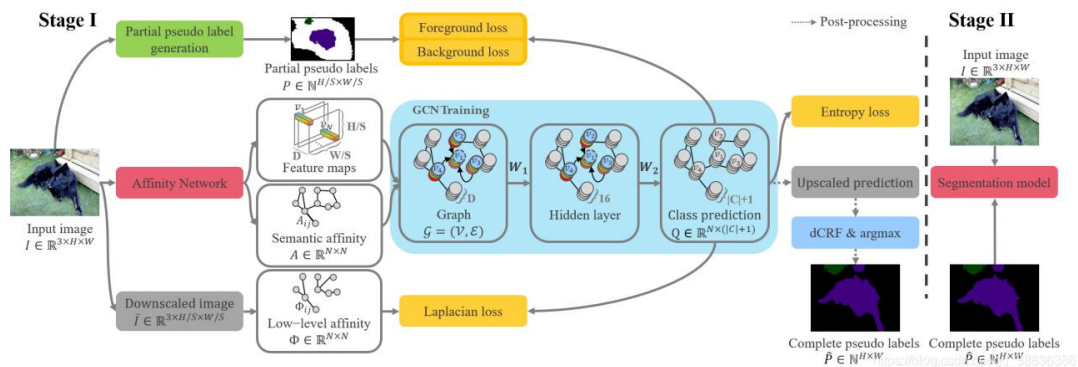


12、Spatially Invariant Unsupervised 3D Object Segmentation with Graph Neural Networks

摘要：在本文中，我们解决了一个没有 RGB 信息的点云的无监督三维对象分割问题，特别地，我们提出了一个框架，SPAIR3D，来建模一个点云作为一个空间混合模型，并通过变分自编码器联合学习三维中的多对象表示和分割。受 SPAIR 的启发，我们采用了一种对象规范方案，它描述了每个对象对于其本地相关网络单元的位置，而不是整个点云。为了模拟点云上的空间混合模型，我们推到了张角似然，它自然适合变分训练管道。我们进一步设计了一个新的空间不变图神经网络，以在我们的 VAE 中生成不同数据的三维作为编码器

13、WEAKLY-SUPERVISED IMAGE SEMANTIC SEGMENTATION USING GRAPH CONVOLUTIONAL NETWORKS

摘要：本工作解决了基于图像级类标签的弱监督图像语义分割，完成这项任务的一种常见方法是使用随机行走机制来传播类激活映射的激活分数，以获得完整的伪标签，以全监督的方式训练语义分割网络。然而，随机游走的前馈性质并没有对由此产生的完整伪标签的质量进行正则化。为了克服这个问题，我们提出了一个基于图卷积网络的特征传播框架。我们将完整伪标签的生成表示为半监督学习任务，并通过反向传播拉普拉斯量和熵正则化损失，分别为每一副训练图像学习两次 GCN



14、Scale-Aware Graph Neural Network for Few-Shot Semantic Segmentation

摘要：小样本语义分割的目标是分割看不见的类对象，因为来自同一类的很少有密集注释的支持图像。现有的 FSS 方法通过使用支持原型或者直接依赖启发式多尺度特征融合来查找查询对象。然而，它们未能充分利用支持查询图像对之间的多尺度特征的高阶外观关系，从而导致查询对象的定位不准确。为了解决上述挑战，我们提出了一种端到端 FSS 的跨尺度关系，通过推理端到端尺度感知图神经网络 SAGNN。具体的说，首先以支持诱导的多尺度查询特征作为节点来构建尺度感知图，同时将每条边建模为其连接节点的成对交互。通过渐进式消息传递图，SAGNN 能够捕获跨尺度关系并克服对象的变化（如外观、尺度和位置）从而可以学习更精确的节点嵌入。这反过来又使它能够预测更准确的前景对象。此外，为了充分利用查询图像跨尺度的位置关系，提出了一种新的自节点协作机制来丰富当前节点，赋予 SAGNN 感知相同对象不同分辨率的能力。

15. Class-wise Dynamic Graph Convolution for Semantic Segmentation (ECCV 2020)

将 graph convolution 设计为一个模块提升分割效果，核心在于只在类内建立 graph，以及关注 hard samples

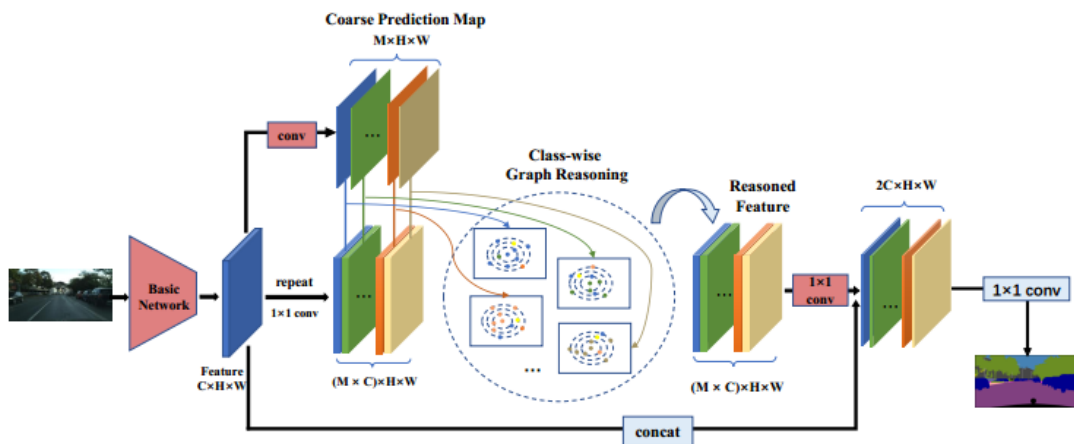


Fig. 2. An Overview of the Class-wise Dynamic Graph Convolution Network. Given an input image, we first feed it into the basic segmentation network to get the high-level feature map and the corresponding coarse segmentation result. Then the CDGC module is applied to preform graph reasoning along nodes of the feature map, producing a refined feature which is subsequently fused with the original feature to get the final refined segmentation result. Specially, in the class-wise graph reasoning part, different colors of lines and dots denote different classes of pixels, under the guidance of coarse prediction map, most positive pixels are sampled while also harvesting few hard pixels in different colors from the target color.

16. Vision GNN: An Image is Worth Graph of Nodes

华为诺亚实验室的研究员发现图神经网络（GNN）也能做视觉骨干网络。将图像表示为图结构，通过简洁高效的适配，提出一种新型视觉网络架构 ViG，表现优于传统的卷积网络和 Transformer。在 ImageNet 图像识别任务，ViG 在相似计算量情况下 Top-1 正确率达 82.1%，高于 ResNet 和 Swin Transformer。

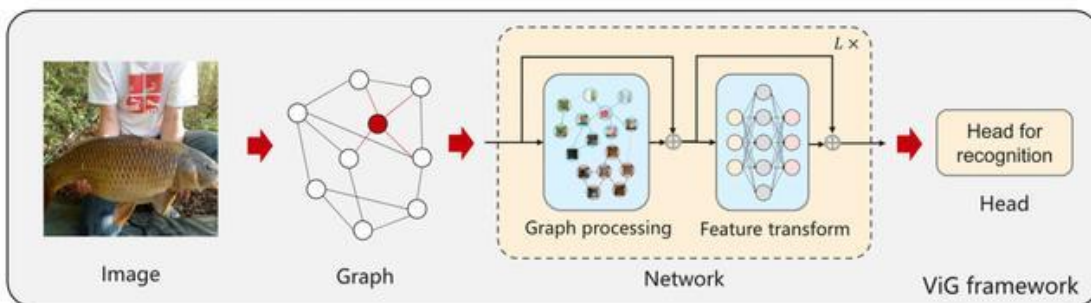


Figure 2: The framework of the proposed ViG model.

Table 5: Results of Pyramid ViG and other pyramid networks on ImageNet. ♠ CNN, 🟠 MLP, 🍈 Transformer, ★ GNN.

Model	Resolution	Params (M)	FLOPs (B)	Top-1	Top-5
♠ ResNet-18 [16, 56]	224×224	12	1.8	70.6	89.7
♠ ResNet-50 [16, 56]	224×224	25.6	4.1	79.8	95.0
♠ ResNet-152 [16, 56]	224×224	60.2	11.5	81.8	95.9
♠ BoTNet-T3 [44]	224×224	33.5	7.3	81.7	-
♠ BoTNet-T3 [44]	224×224	54.7	10.9	82.8	-
♠ BoTNet-T3 [44]	256×256	75.1	19.3	83.5	-
🍈 PVT-Tiny [54]	224×224	13.2	1.9	75.1	-
🍈 PVT-Small [54]	224×224	24.5	3.8	79.8	-
🍈 PVT-Medium [54]	224×224	44.2	6.7	81.2	-
🍈 PVT-Large [54]	224×224	61.4	9.8	81.7	-
🍈 CvT-13 [57]	224×224	20	4.5	81.6	-
🍈 CvT-21 [57]	224×224	32	7.1	82.5	-
🍈 CvT-21 [57]	384×384	32	24.9	83.3	-
🍈 Swin-T [33]	224×224	29	4.5	81.3	95.5
🍈 Swin-S [33]	224×224	50	8.7	83.0	96.2
🍈 Swin-B [33]	224×224	88	15.4	83.5	96.5
🟠 CycleMLP-B2 [4]	224×224	27	3.9	81.6	-
🟠 CycleMLP-B3 [4]	224×224	38	6.9	82.4	-
🟠 CycleMLP-B4 [4]	224×224	52	10.1	83.0	-
🟠 Poolformer-S12 [64]	224×224	12	2.0	77.2	93.5
🟠 Poolformer-S36 [64]	224×224	31	5.2	81.4	95.5
🟠 Poolformer-M48 [64]	224×224	73	11.9	82.5	96.0
★ Pyramid ViG-Ti (ours)	224×224	10.7	1.7	78.2	94.2
★ Pyramid ViG-S (ours)	224×224	27.3	4.6	82.1	96.0
★ Pyramid ViG-M (ours)	224×224	51.7	8.9	83.1	96.4
★ Pyramid ViG-B (ours)	224×224	92.6	16.8	83.7	96.5

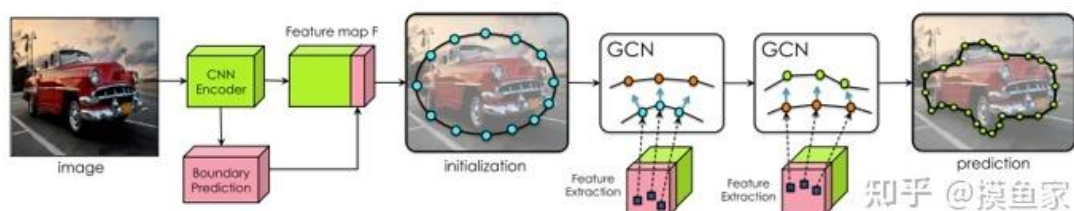
17. Fast Interactive Object Annotation With Curve-GCN

这篇文章的任务是基于 GCN 做交互式标注，将 mask 轮廓用边缘来表示，边缘就是 graph。

之前也提到了，在图像中使用 graph 结构，关键就是 graph 怎么构建。这篇文章中初始化节点的连接关系就是一个椭圆形状，节点的特征由 CNN 提供。将构建好的初始化图给到 GCN，GCN 的作用是预测每个节点的坐标值，即每个 node 最终有两个坐标，代表其在图像中的位置，这些点连接起来的形状就是物体的轮廓。

• Method—Architecture

- 使用椭圆作为初始化轮廓
- CNN提取的特征作为顶点的特征
- 将构建的初始化图输入GCN预测每个顶点的坐标

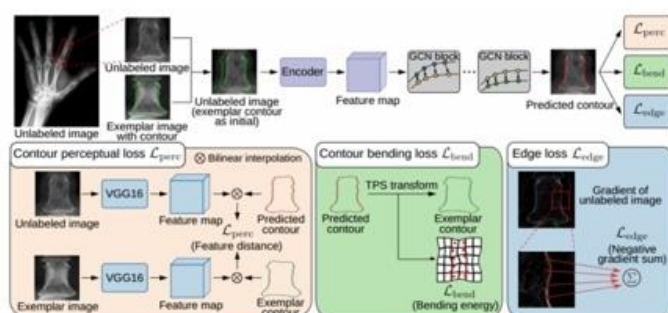


18. Learning to Segment Anatomical Structures Accurately from One Exemplar

这篇文章是做医学图像的分割，具体来说是 One shot learning，即只有一张有标注的 label。实际做法和上一篇文章很相似，先给一个初始 graph，CNN 提取特征作为节点特征，唯一的 label 作为初始轮廓。GCN 来预测每个点相对初始轮廓的偏移量，本质也是坐标回归。

Method

- 单一标注作为初始轮廓
- CNN提取特征作为顶点特征
- 多层GCN，输出为每个顶点的偏移量



Yuhang Lu et al, "Learning to Segment Anatomical Structures Accurately from One Exemplar," MICCAI 2020.

19. View-volume Network for Semantic Scene Completion from a Single Depth Image

这篇文章的任务是利用深度图来进行 2D 图像的语义分割。

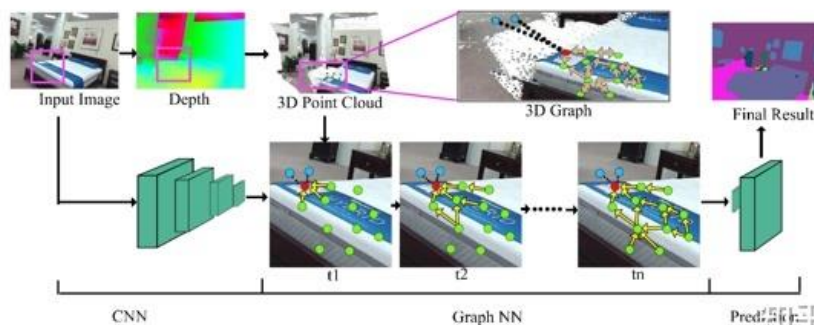
Method

图的构建

- 顶点及特征：使用深度信息得到3D点云，每个点对应一维CNN提取的特征
- 连接关系：每个点，通过最近邻搜索找到最近的64个点连接

更新和输出

- 将初始图送入GCN，最每个节点的最最终输出是像素的类别



Y. Guo and X. Tong, "View-volume network for semantic scene completion from a single depth image," IJCAI 2018.