

每日小结

| | 周一 | 周二 | 周三 | 周四 | 周五 |
|---|---------------------|-------------|-------------------------------|-------------|-------------|
| 早 | Response 撰写 | 阅读文献 | 修改论文 | Response 撰写 | 修改论文 |
| 中 | 阅读文献， 补实验时间复杂度 | Response 撰写 | 修改论文， Response 撰写， 阅读文献 | Response 撰写 | Response 撰写 |
| 晚 | 阅读文献 Response 撰写 | 修改论文 | 组会 | 组会 | |

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: MMoT: Mixture-of-Modality-Tokens Transformer for Composed Multimodal Conditional Image Synthesis

作者: Jianbin Zheng • Daqing Liu • Chaoyue Wang • Minghui Hu • Zuopeng Yang • Changxing Ding • Dacheng Tao

出处: IJCV 2024

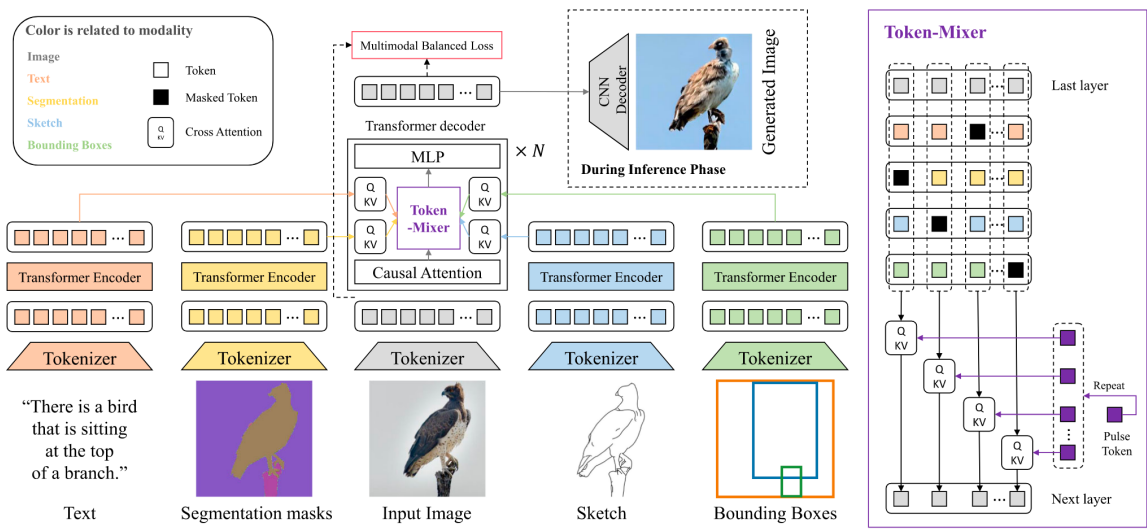
方法:

为了解决模态协调问题，提出了 MMoT (Mixture-of- Modality- Tokens Transformer) 来充分利用模态之间的协同性。现有的方法需要一个“整合信号”才能在多种模态条件下产生高质量的图像，而 MMoT 缓解了这一苛刻的要求，只需要一个“互补信号”。输入可以由多个互补模态(例如, 文本、草图、分割掩码和边界框)组成，MMoT 可以利用所有这些输入产生合理的结果组合多模态条件图像合成，它允许用户输入各种不完全互补的信号特征。

具体来说，MMoT 使用多个编码器对模态内交互进行建模。然后，采用模式特异性交叉注意力机制注入多模态条件信息。融入到解码器中。最后，关键模块多级 token 混合器通过掩码交叉注意力机制自适应融合多模态条件化信息。

为了解决模态不平衡问题，提出了一种多模态平衡损失，在训练阶段自适应控制每个模态的优化，以及在采样阶段的多模态采样引导，以控制不同模态的影响，并将散度图引入采样过程。

International Journal of Computer Vision



给定一幅图像和多个模态，包括文本、分割掩膜、草图和边界 box，将其权标化转换成具有不同权标化 rs 的离散标记，然后与模式特异性编码器进行模态内交互；B 使用模式特异性交叉注意力向解码器中注入多模态条件信息；C 通过多级 token 混合器自适应地融合条件信号。使用多模态平衡损失和散度驱动的多模态指导来训练 MMoT

多级 token 混合器：在交叉注意力后，提出了多级 token 混合器来融合包含与特定模态相关的条件信息的模态令牌。在 token - mixer 中引入一个特殊的[PULSE] token P，自适应地估计每个模态 token 的组合权重(即,注意力分数)，并将其与掩蔽交叉注意力机制进行融合。



启发：

1. 用多级token的思想很不错，可以借鉴（作为后面工作的思路）
2. 有融合loss，在看loss代码

文献2

题目：Graph Convolutional Networks with Adaptive Neighborhood Awareness

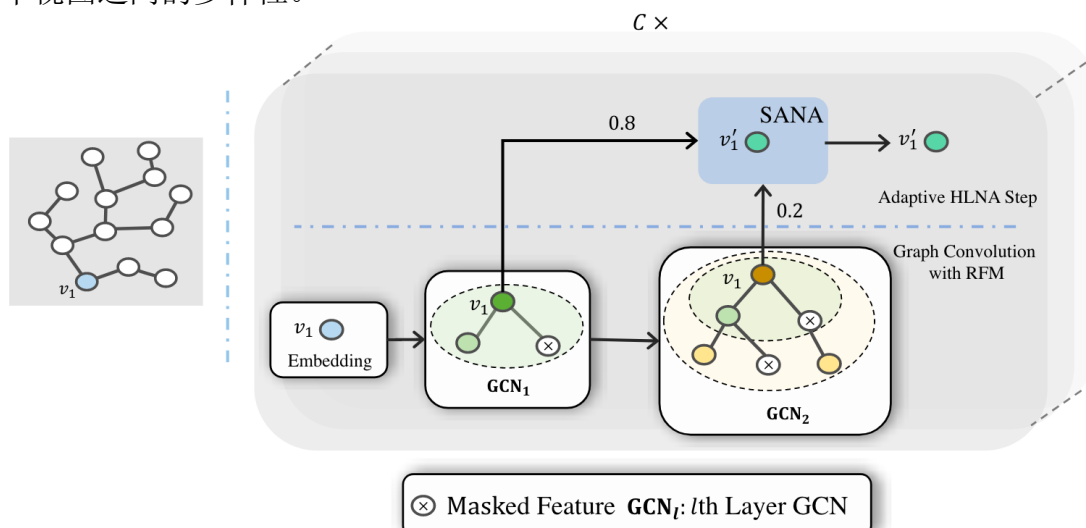
作者：Mingjian Guang, Chungang Yan, Yuhua Xu, Junli Wang, and Changjun Jiang

出处：TPAMI 2024

方法：

现有方法通常仅从节点或跳级别执行邻域感知步骤，这导致缺乏从全局和局部角度学习节点邻域信息的能力。此外，大多数方法从单个视图学习节点的邻域信息，忽略了多个视图的重要性。在这项工作中，提出 MVANA 来实现 GCN 的自适应邻域感知。主要步骤如下。

1. 首先，构建了一个多通道 HNA 编码器，它从多个视图和节点跳级别感知邻域，并将非欧几里德空间中的图编码为欧几里德空间中的节点表示张量。
2. 接下来，利用读出函数来压缩图节点并获得图表示。
3. 最后，使用交叉熵损失来缩小预测值和真实值之间的距离。我们还采用多视图损失函数来控制多个视图之间的多样性。



1. 多通道 HNA 编码器：采用 HNA 和多视图学习来提高 GCN 的编码能力。编码器由三个模块组成：带有 RFM 的图卷积、自适应 HLNA 步骤和多通道结构。

- RFM 的图卷积：一些 NLNA 方法通常采用节点采样技术来改变邻居的分布，并利用特定的图卷积算子来学习新邻居分布的表示。利用 RFM 扰乱邻域分布，作为节点采样的替代方案。这种方法可以被视为一种更温和的节点采样形式，因为它不会破坏底层的图结构。此外，节点特征代表了节点的基本信息，改变这些特征本质上会改变节点信息，从而影响图卷积层对邻域信息的聚合。

- **自适应 HLNA 步骤:** 在本模块中, 提出了一种用于 HLNA 的软自适应邻域感知 (SANA) 方法。主要思想是遍历每个跳级表示 H_l 并计算该表示的 0 到 1 之间的选择因子。具有较大选择因子的表示很重要。当选择因子等于 0 时, 其对应的跳级表示被丢弃。
- **多通道结构:** 以上公式仅使用一个通道。还采用多通道结构从多个角度感知邻域。具体来说, 对于每个通道 c , 计算第 c 个通道编码器的输出 Z_c 如下:

$$Z_c = \alpha H_c^L + (1 - \alpha) M_c^L$$

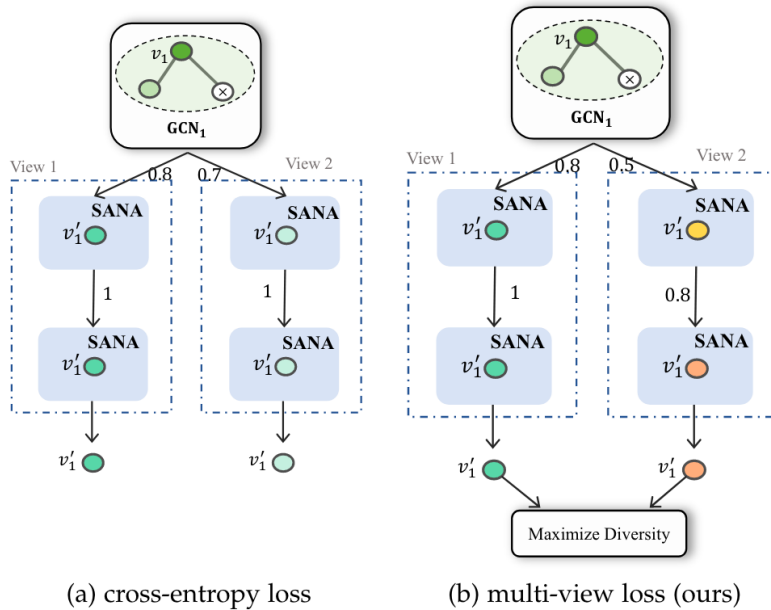
其中 α 是一个超参数, 用于权衡 RFM 和自适应图卷积的重要性海航网络步骤。当 $\alpha = 1$ 时, 编码器可以退化为没有 SANA 的网络。

2. 接下来, 采用读出函数将节点表示张量 Z 压缩为图表示, 即 $\text{fre} : Z \in \mathbb{R}^{C \times n \times d} \rightarrow \mathbb{R}^d$ 。由于 2D 卷积和池化算子在提取 3D 张量方面表现出了出色的性能, 因此我们使用它们来提取和压缩 Z 。不幸的是, Z 的维度对于不同的图而言是不同的, 因为图具有不同数量的节点, 即 n 是一个不可知的值。由于输入维度不确定, 我们无法计算 2D 卷积算子的输出维度并定义后续全连接层的输入维度。因此, 我们利用之前的工作[10], 利用注意力机制[70]、[71], 在输入到 2D 卷积算子之前将 Z 映射到固定维度张量。

3. **最后, 交叉熵和多视角 loss:** 仅使用交叉熵 loss 不是最理想的, 因为无法控制多个视图之间的多样性。例如, 这个损失可能会学习如图 3 (a) 所示的模型, 并且两个视图的邻域感知过程可能非常相似。为了缓解上述问题, 额外添加了**多视图损失**来监督和控制多个视图之间的多样性, 如图 3 (b) 所示。损失函数受到惩罚, 并鼓励多个通道的输出更加多样化。具体来说, 我们将多视图损失定义如下:

$$\mathcal{L}_{\text{view}} = \frac{1}{C} \sum_{i=1}^C \text{Div}^2(Z^i)$$

其中 Z_i 是编码器对 G_i 的输出, $\text{Div}(\cdot)$ 是衡量不同视图之间多样性的函数。 Z_i 的多通道输出矩阵。因为对于同一个输入图, 当各个通道的参数不一致时, 各个通道的输出会有很大差异。因此, 假设输出矩阵之间的相似度越低, 信道多样性越高。可以通过测量通道输出之间的相似性来提供有关通道参数多样性的反馈。皮尔逊相关系数是衡量相似度的有效方法。



| Baselines | PROTEINS | DD | NCI1 | NCI109 | Mutagenicity | REDDIT-MULTI-12K |
|-------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
| GCN | 73.22±3.98 | 71.82±6.86 | 69.46±2.81 | 68.36±1.99 | 79.92±2.00 | 40.62±1.87 |
| Set2Set | 73.40±4.21 | 70.98±4.78 | 72.75±3.11 | 69.30±4.03 | 80.13±1.99 | 40.79±2.29 |
| JK-Net | 73.22±3.98 | 71.82±6.86 | 69.46±2.81 | 68.36±1.99 | 79.92±2.00 | 40.62±1.87 |
| APGCN | 73.94±3.25 | 78.74±5.60 | 74.22±4.96 | 69.58±7.59 | 80.69±1.50 | 44.67±1.50 |
| GraphSAGE | 72.86±4.66 | 72.41±4.10 | 73.02±3.26 | 72.01±1.69 | 79.99±2.72 | 39.26±2.03 |
| GAT | 72.24±4.10 | 71.31±4.89 | 70.22±2.35 | 65.66±4.83 | 78.81±3.53 | 40.37±1.79 |
| SortPool | 74.66±5.08 | 67.47±6.23 | 70.58±3.68 | 68.87±2.38 | 75.81±2.43 | 40.83±1.98 |
| DiffPool | 73.62±5.44 | 76.24±5.19 | 71.03±3.04 | 70.04±2.80 | 79.71±2.31 | 42.36±2.13 |
| EdgePool | 73.04±4.64 | 72.33±4.48 | 73.94±2.88 | 72.23±1.87 | 79.92±1.72 | 41.52±2.16 |
| gPool | 73.14±4.55 | 75.56±4.74 | 70.92±3.31 | 71.51±2.86 | 75.91±3.60 | 42.09±2.07 |
| SAGPool | 74.84±4.92 | 77.35±4.31 | 75.16±2.58 | 73.69±2.70 | 78.40±1.76 | 41.27±2.35 |
| HGP-SL | 74.57±3.78 | 78.61±5.86 | 75.40±3.90 | 74.05±2.32 | 79.76±1.90 | 42.25±1.97 |
| ASAP | 74.66±4.48 | 77.42±5.12 | 72.60±2.02 | 71.29±1.21 | 77.96±1.65 | 41.53±2.09 |
| GMT | 75.09±5.90 | 78.72±5.90 | 74.21±1.88 | 71.38±2.03 | 80.26±2.20 | 42.49±2.18 |
| DropGIN | 74.39±5.09 | OOM | 77.59±1.71 | 77.08±1.73 | 80.77±1.53 | OOM |
| GraphMAE | 76.30±6.1 | 77.67±2.40 | 80.40±0.30 | 74.33±1.07 | 78.23±1.72 | 41.22±1.48 |
| MVAGC | 75.64±2.58 | 78.34±4.29 | 76.23±2.39 | 74.97±2.46 | 80.18±1.65 | 41.37±2.36 |
| MVGRL | 75.12±4.58 | OOM | 80.17±2.67 | 77.73±1.32 | 80.07±1.27 | OOM |
| MAC | 76.28±3.55 | 79.13±4.70 | 77.62±1.16 | 75.84±1.86 | 80.33±1.49 | 42.67±2.23 |
| ours | 77.26±3.77 | 80.90±3.72 | 81.44±2.30 | 79.06±1.48 | 83.08±1.80 | 49.68±1.52 |

启发:

1.我目前用的是lstm提取后的特征直接作为边的权重，其实可以按照这个论文的思路，用lstm作为GCN的聚合函数，然后用类似的自适应邻域选择方法，格外再加入多视角loss（考虑邻域相似性）

本周工作

1. 阅读文献;
2. 修改论文，写response（response在润色，还差2条意见没回复完，以及还差增加最新相关工作综述）
3. 时间复杂度有几个很早跑过对比模型没有了，在重新整理补跑

下周计划

1. 投递论文response
2. 准备组会ppt（目前打算讲之前那个nature的船舶GPS的工作，待定）