Asynchronous Events-based Panoptic Segmentation using Graph Mixer Neural Network

作者：Sanket Kachole, Yusra Alkendi, Fariborz Baghaei Naeini, Dimitrios Makris, Yahya Zweiri

期刊：CVPRW

年份：2023

1、问题背景/研究动机

本文研究的主要问题：**机器人环境（物体抓取）中基于事件的全景分割问题**

作者认为，现有的图像分割方法，其局限性在于**忽略了基于事件数据的高时间分辨率**。

- 全景分割

全景分割（Panoptic Segmentation）是一种计算机视觉任务，涉及**同时分割和识别图像中的所有对象**，同时将**前景物体**和**背景区域**进行像素级别的分割和标记。

- 是**实例分割**与**语义分割**的融合。
  - **实例分割**：识别单个对象并为其分配唯一标签
  - **语义分割**：为图像中的每个像素分配特定的语义类别标签
- 全景分割有两类，即thing和stuff。
  - Stuff：不可数的区域，eg. 天空、人行道和地面
  - Thing：所有可数的物体，eg. 汽车、人等
- 通过给每一个目标赋予不同的颜色，使其与其他目标区分开来。
- Pros：
  - 提供更丰富的场景理解，使计算机能够更好地感知和理解图像中的的不同对象和背景。
  - 通过实现对特定目标的分析而无需检查图像的整个区域，**减少计算时间**，最大限度地**减少了对某些目标的漏检**，确定了图像或视频中不同区域的边缘显著性。

- **全景分割的常见挑战：**
  - 由于杂乱的场景，物体的几何和外观变化，遮挡，运动模糊和传统相机的低时间分辨率，高延迟会导致处理传感器数据的延迟，从而导致执行任务的响应时间变慢和准确性降低。

2、解决方法

（1）提出了**基于事件的全景分割的图混合神经网络**（Graph Mixer Neural Network，GMNN），利用事件数据的异步性质和时空相关性，构建3D图形来对事件进行建模，并使用混合和采样模块对邻近事件的子图进行处理，实现分割任务。

（2）关键的技术贡献是图神经网络架构中的协同上下文混合(CCM)层，该层能够同时将来自多个邻近事件集的特征进行混合。CCM层在四个最近邻层级上并行地传播时空相关性，从而更好地捕捉运动动态并提高分割性能。

（3）在基于事件的分割（Event-based Segmentation，ESD）数据集上进行评估，该数据集包括5个图像退化，包括遮挡、模糊、亮度、轨迹、尺度变化和已知以及未知对象的分割。

3、具体实现

- **事件表示**

连续的事件流在数学上表示为由i个事件位置、时间戳和极性组成的元组序列

$$(x_1, y_1, t_1, z_1), (x_2, y_2, t_2, z_2), ..., (x_n, y_n, t_n, z_n)$$

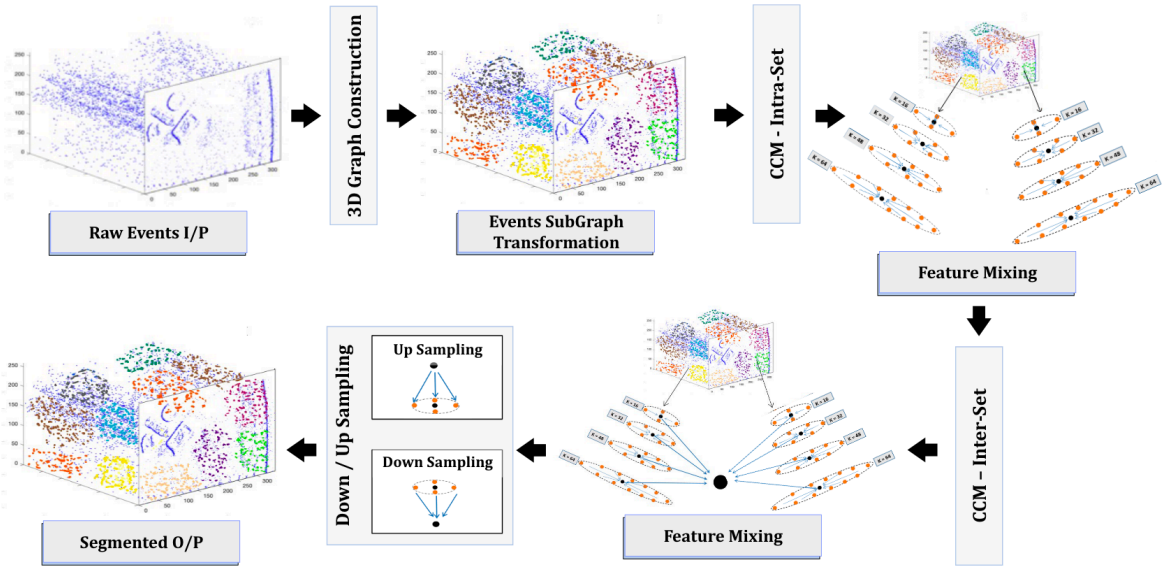
- **图神经网络**

GNN的核心思想是将图中的节点和边作为输入，通过多层神经网络的计算，得到每个节点的表示向量，从而实现对整个图的分类、聚类、预测等任务。

- **k近邻(k-Nearest Neighbor，kNN)方法**

通常通过考虑邻近性对事件数据进行局部处理

- 图混合神经网络



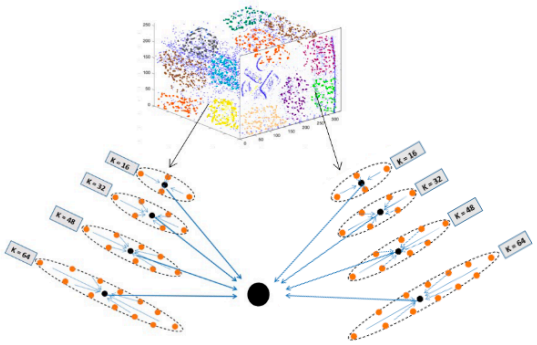
- 三维事件图形的构建

基于事件数据的异步性质和时空相关性。

首先，将事件数据转化为3D图形，其中每个节点表示一个事件，节点特征包括事件的位置、时间等信息。然后，通过k近邻方法将每个节点与其k个最近邻节点连接起来，构建图形的拓扑结构（得到的时空邻域称为子图，每个子图有k+1个节点）。

由此，可以**将事件之间的时空关系编码到图形**中。

- 协同上下文混合层（CCM）



一种**在不同集合之间传播事件特征**的方法。

CCM层的设计旨在同时在四个最近邻层级上传播时空相关性，以更好地捕捉运动动态。具体而言，CCM方法在每个最近邻层级上应用空间金字塔块，通过计算节点之

间的相对位置编码和特征向量的加权和，将事件特征在多个最近邻层级上并行地传播，从而有效地进行**特征混合**。

- **Transition down block**

过渡向下块是GMNN架构的一个组件，负责**对3D事件图中的图节点进行下采样**。该块的目的是**减少图的基数**。

在Transition Down block中，使用kNN算法对图形进行下采样。原始图中的每个节点都连接到其k个最近的相邻节点，形成一个简化图。还原系数由所要求的下采样率决定。例如，如果原始图有N个节点，并且要求下采样系数为4，则过渡下行模块会产生一个带有N/4节点的新图。

该模块中的下采样实现了**图节点的卷积**，并有助于识别事件之间的时空相关性。

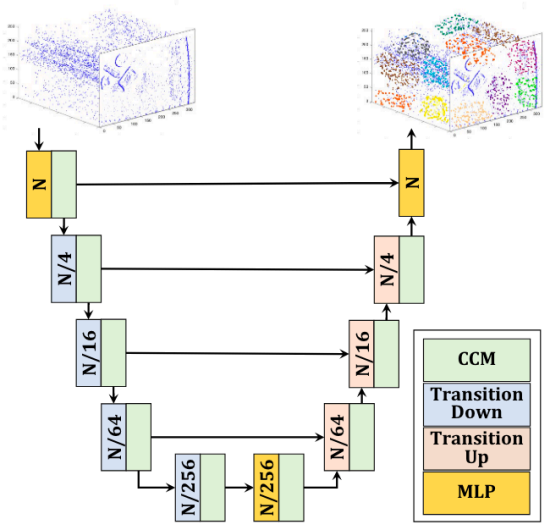
- Transition up block

负责**对3D事件图中的图节点进行上采样**。该块用于GMNN的解码器部分，用于在Transition Down块中进行下采样后**重建原始图形大小**。

在Transition Up块中，对简化图的图节点进行上采样，以匹配原始图的大小。这是通过使用在Transition Down块中计算的索引映射来实现的。逆索引映射用于将减少的图节点映射回原始图形中的相应节点。通过执行反向映射，在上采样期间保留节点之间的空间关系。

Transition Up块通过**图节点的上采样和保持事件数据的空间一致性**，确保在不同下采样级别捕获的信息被有效集成，并用于最终的分割任务。

- 网络架构



GMNN架构有四个组件：MLP块，下采样，上采样和Mixer块，它们构成编码器和解码器。

在编码器中，3D事件图先通过MLP层，然后通过4个下采样块，每个下采样块将节点数减少4个。Mixer模块使用CCM方法并行**传播特征**。

然后将编码器的输出传入解码器，解码器开始是一个MLP层，然后是4个上采样块。Mixer模块使用CCM方法进行**特征扩展**。

该架构采用**深度金字塔式结构**，通过**逐级下采样节点来获取全局特征**。与传统的图神经网络(Graph Neural Network，GNN)方法不同，GMNN从每个输入图中构建子图，并在Mixer层和采样模块中并行处理。该方法有利于**识别事件之间的时空相关性**，有效捕捉运动动态。

4、实验对比

文中使用ESD数据集来评估所提出的方法在**平均交并比(mIoU)、像素精度**以及**计算效率**方面的表现。下图是在ESD数据集的样本上测试GMNN时获得的分割结果。

- ESD数据集

ESD数据集包括17186张标注图像和177条标注事件流，使用安装在机械臂末端的Davies346传感器捕获。它具有相机运动、手臂速度、光照条件和杂乱场景等的变化。该数据集对6个类别的15个对象类进行了实例级标注。训练集(ESD-1)由10个已知物体的13984张图像组成，测试集(ESD-2)由5个不在ESD - 1中的未知物体的3202张图像组成。

- 评价指标

像素精度和**mIoU**被用来评估全景分割的性能。

- 像素精度

计算图像中被正确分类的像素的百分比。为了使像素精度适应基于事件的视觉数据，计算了预测事件的每个对象计数与真实事件的比值，然后计算所有对象的平均

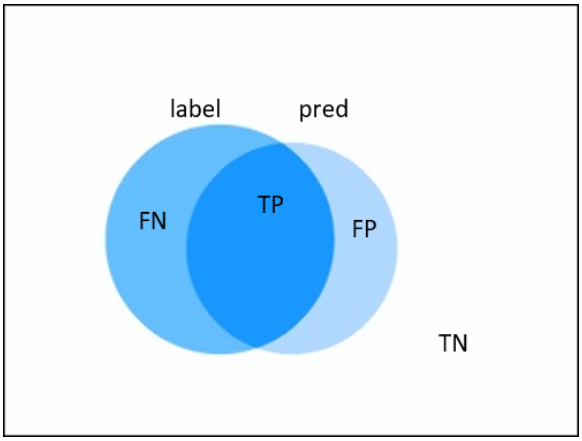
精度。该方法提供了一种基于事件数据评估目标检测模型的方法，并考虑了事件的稀疏性。

- **mIoU**

  也称为Jaccard指数，表示平均交并比，即数据集上每一个类别的IoU值的平均。

$$mIoU = \frac{1}{C}\sum_{i}^{C}\frac{\sum_{i}^{N}\delta(d_{i,c},1)\delta(d_{i,c},d'_{i,c})}{max(1,\delta(d_{i,c},1)+\delta(d'_{i,c},1)}$$

  交并比（IoU）：



  图中，左边的圆表示某个类的真实标签，右边的圆表示其预测结果，中间的TP部分即为真实值和预测值的交，FN+FP+TP部分即为真实值和预测值的并，IoU就是交与并的比值。

- **定量评价**

  文中对ESD数据集的每个子任务进行了评估，包括整个数据集的**物体数量、光照条件、运动方向、相机速度和物体大小的变化**。进一步，在**未知数据集上**进行类似的评估，以了解模型在**未知物体分割上的准确性**。

Exp 1: **varying clutter objects**, Bright light, 62cm height, Rotational motion, 0.15 m/s speed					
Method	2 Obj	4 Obj	6 Obj	8 Obj	10 Obj
EV-SegNet [3]	82%	73%	67%	54%	51%
ESS [28]	86%	76%	68%	64%	60%
GTNN [2]	89%	86%	84%	77%	71%
GMNN (ours)	**97%**	**96%**	**91%**	**89%**	**87%**

Exp 2: 6 Objects, **varying lighting conditions**, 62cm height, Rotational Motion, 0.15 m/s speed.		
Method	Bright Light	Low light
EV-SegNet [3]	76%	75%
ESS [28]	79%	78%
GTNN [2]	81%	79%
GMNN (ours)	**95%**	**94%**

Exp 3: 6 Objects, Bright Light, 62cm height, **Varying directions of motion**, 0.15 m/s speed.			
Method	Linear	Rotational	Partial Rotational
EV-SegNet [3]	65%	73%	69%
ESS [28]	68%	78%	74%
GTNN [2]	75%	89%	78%
GMNN (ours)	**84%**	**93%**	**90%**

Exp 4: 6 Objects, Bright Light, 62cm height, Rotational motion, **Varying speed**.			
Method	0.15 m/s	0.3 m/s	0.1 m/s
EV-SegNet [3]	69%	60%	56%
ESS [28]	72%	63%	59%
GTNN [2]	75%	71%	63%
GMNN (ours)	**93%**	**91%**	**87%**

Exp 5: 6 Objects, Bright Light, **Varying camera height**, Rotational motion, Varying speed.		
Method	62 cm	82 cm
EV-SegNet [3]	76%	74%
ESS [28]	82%	75%
GTNN [2]	85%	83%
GMNN (ours)	**97%**	**93%**

• **第1个实验**使用了具有**不同物体数量的测试集的子集**，包括2，4，6，8和10个对象，更多的物体意味着更多的遮挡和更有挑战性的场景。相比之下，GMNN模型仅下降了10 %，并且在任何数量的场景对象上都优于其他方法。

• **第2个实验**在**两种光照条件**下进行测试，GMNN对不同的光照条件具有鲁棒性，因为它在两种光照条件下都达到了最高的准确率，即在明亮和黑暗条件下分别达到了95%和94 %。

• **第3个实验**，机械臂的**运动方向变化为线性、旋转或部分旋转**。相比之下，GMNN在旋转运动中获得了最高的平均精度分数93 %，对于部分旋转运动和直线运动分别下降到90 %和84 %。

• **第4个实验**，其末端执行器的速度是变化的。GMNN模型在0.15 m/s时具有93 %的最高精度，在1m/s时下降到91 %。高速条件下CCM混合层的清晰影响支持了轮廓处信息的恢复。

• 第5个实验，为了理解模型的尺度不变性，令平台与相机之间的距离变化为62 cm和82 cm。表中可以看出，摄像机和物体距离对所有模型的精度影响最小，GMNN保持了其优越性。

Methods	Known Obj		Unknown Obj	
	mIoU %	Acc %	mIoU %	Acc %
EV-SegNet [3]	7.73	76.98	5.29	53.31
ESS [28]	8.92	81.59	7.01	67.29
GTNN [2]	74.24	87.53	58.70	81.30
GMNN (ours)	**78.32**	**96.91**	**66.05**	**89.91**

• 该表比较了EVSegNet、ESS、GTNN和GMNN 4种方法在已知对象数据集和ESD-2未知对象数据集上的性能。可以看出，GMNN方法获得了最高的mIoU和准确率，优于其他所有架构。

• 实验结果证明了将异步事件的图结构用于分割任务的合理性。

- 模型尺寸

Methods	Parameters
EV-SegNet [3]	22M
ESS [28]	17M
GTNN [2]	5.3M
GMNN (ours)	**3.9M**

比较GMNN与其他3种先进的全景分割方法：EV-SegNet、ESS和GTNN的参数量。GMNN使用的参数数量最少，为390万，表明GMNN在模型大小和训练时间方面可能更有效和可扩展。

- 计算时间

Model	Sequential-mode $\mu \pm \sigma$ (sec)	Batch-mode μ (sec)
GTNN	$9.63 \times 10^{-2} \pm 1.93 \times 10^{-4}$	13.52×10^{-4}
GMNN	$\mathbf{4.06 \times 10^{-3} \pm 2.05 \times 10^{-4}}$	$\mathbf{10.27 \times 10^{-5}}$

对40个事件图进行分析，每个事件图持续时间为10ms，分两种操作模式，即顺序模式和批处理模式。PyTorch中的顺序模型将事件图依次进行处理，而批处理模式将事件图作为单个批处理。结果表明，在两种工作模式下，GMNN在计算时间方面均优于GTNN。这对于在保持传感器**高时间分辨率**的同时**并行处理批处理事件**是必要的。

- 消融实验

为了评估CCM方法在分割任务中的有效性，消融实验主要关注CCM中最近邻层级的数量和特征混合的数量对模型性能的影响。

- 最近邻层级数量

	kNN in Each Layer of CCM							
	L1	**L2**	**L3**	**L4**	**L5**	**L6**	**L7**	**ACC%**
1	16	-	-	-	-	-	-	81.23
2	16	32	-	-	-	-	-	89.04
3	16	32	48	-	-	-	-	92.50
4	16	32	48	64	-	-	-	**96.91**
5	16	32	48	64	80	-	-	96.05
6	16	32	48	64	80	96	-	95.37
7	16	32	48	64	80	96	112	93.75

通过**改变CCM中的kNN层数**来研究其对模型准确率的影响。结果表明，增加kNN数通常会提高模型的准确性。具体而言，**增加kNN层数可以更好地捕捉事件之间的时空关系**，从而提高模型的分割性能。

- 特征混合的数量

通过改变**CCM中特征混合的并行数量**来研究其对模型准确性的影响。结果显示，增加特征混合的并行数量可以进一步提高模型的准确性。因为**增加并行数量可以增加特征之间的交互和信息传递**，从而更好地捕捉事件数据的时空相关性。

5、局限与总结

- 局限性

• 该方法的性能受相机捕获事件数据的速度的影响。更高的相机速度会产生更多事件，而较低的速度会产生更少的事件。这种依赖性可能会限制该方法在不同相机设置中的通用性。

• **内存限制**：该方法在给定的最大节点数的每个时间间隔内构造图形。采用这种方法来平衡内存使用和准确率。然而，它可能导致早期事件中重要信息的丢失，并影响准确确定时空关系的能力。

• **CCM层中的过度平滑**：研究发现，增加CCM层中kNN层数最初可以提高准确率，但过多的特征会导致过度平滑和准确率降低。找到kNN的最佳数量对于避免这个问题至关重要。

- 启发

• GMNN架构利用图神经网络来处理基于事件的数据并执行分割任务。该架构结合了下采样和上采样块以及CCM技术，以有效处理数据的空间和时间相关性。

• CCM能够有效融合来自图层次结构多个级别的上下文信息，提高模型在事件数据中捕获时空关系的能力。

• 未来的研究可以探索所提出的方法**在具有不同机器人、传感器和环境的现实世界场景中的泛化能力**，并结合其他传感器，如深度传感器或热像仪，以改善模型的低光性能。

6、总结

- 本研究提出了图混合神经网络架构（GMNN），用于处理机器人物体抓取任务中的异步事件数据的全景分割问题。
- 提出了加入协同上下文混合(CCM)层，该层能够实现由多组邻域事件生成的事件特征的并行混合。

- GMNN架构在多个层次上组合相邻事件，产生并行的特征学习表示。编码器执行下采样操作，而解码器对事件执行上采样操作，从而得到有效的用于机器人抓取的全景分割模型。
- 所提出的模型在不同条件下的ESD数据集上进行实验，证明了引入的CCM方法对遮挡、低光照、小物体、高速和直线运动等挑战的鲁棒性。
- 利用了GNNs和混频器技术，使得预测时间比现有的最先进的方法更短。