

11.07-11.13

Learning to Estimate Two Dense Depths from LiDAR and Event Data

Learning to Estimate Two Dense Depths from LiDAR and Event Data

Vincent Brebion,
Julien Moreau,
Franck Davoine
<https://vbrebion.github.io/ALED/>

GitHub - heudiasyc/ALED: Code for the "Learning to Estimate Two Dense Depths from LiDAR and Event Data" article

Code for the "Learning to Estimate Two Dense Depths from LiDAR and Event Data" article - GitHub - heudiasyc/ALED: Code for the "Learning to Estimate Two Dense Depths from LiDAR and Event Data" article

<https://github.com/heudiasyc/ALED>

heudiasyc/ALED

Code for the "Learning to Estimate Two Dense Depths from LiDAR and Event Data" article

Contributor 1 Issues 0 Stars 3 Forks 0

Vincent Brebion, Julien Moreau, and Franck Davoine
Heudiasyc (Heuristics and Diagnosis of Complex Systems) Laboratory, CNRS,
Université de technologie de Compiègne (UTC), 60319 Compiègne Cedex, France

简介

将激光雷达（LiDAR）数据和事件数据进行融合。两个部分：

1. LiDAR深度数据的稠密化：使用事件数据作为引导来增加LiDAR数据的密度。解决激光雷达的稀疏性缺点，同时保持metric scale。
2. 事件-深度关联：将深度信息与每个事件相关联。这意味着每个事件都可以在3D中进行投影，然后在另一个视觉传感器中进行2D反投影。在完全校准和同步的设置中，这个过程可以用于将事件数据与RGB图像叠加。

允许将事件数据中的每个事件点（在时间和空间上的事件）投影到三维空间（3D），从而确定它们在现实世界中的位置。然后，这些事件点可以被反投影回到另一个视觉传感器的二维平面（2D），这使得可以将事件数据与该视觉传感器中的图像数据进行关联和匹配。这种过程有助于将事件数据与视觉图像进行对应，以实现更丰富的感知和分析。

这里作者说明并不是任务一包含任务二，认为事件对应的深度可以从任务一得到的稠密化深度图中获取。每个事件不仅仅代表深度的变化，还可能代表光照的变化。作者考虑事件发生前和事件发生后的深度图，考虑事件对深度的影响。**也就是说事件发生不一定就对应深度变化**

提出了一种基于学习的融合方法，用于估计来自事件和稀疏 LiDAR 数据的密集深度图对。

Related Work

1. LiDAR Densification

LiDAR 传感器仅产生稀疏点云，要通过稀疏数据来获取密集的深度图，目前有学习的方法，图像处理方法以及辅助模式方法，通常可以使用RGB相机，还有最近的事件相机。

2.Fusion of Events and Other Modalities

大多数研究都集中在事件和帧的融合上，这要归功于提供两种模式的传感器，如 DAVIS 相机,这也是前面诸多文中提到的事件与图像融合方法。

事件与lidar融合的方法目前的方向有，校准，点云增强事件和LiDAR稀疏致密化。

3.Depth Estimations with Events

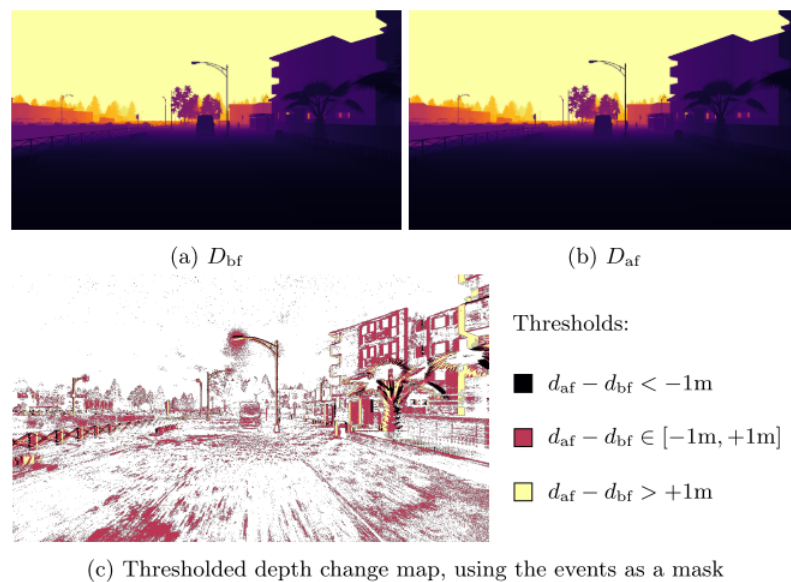
[可以阅读提到的论文](#)

Depth Change Map: Two Depths per Event

通过计算深度变化图 ($D_{af}-D_{bf}$)，(D_{af} 是之前的深度图， D_{bf} 为之后)

以区分不同情况，包括像素位于深度变化不大的区域、像素原本位于物体的边缘现在位于更远的物体上以及像素位于较远物体上现在位于较近物体的边缘（判断像素对应深度差 $d_{af}-d_{bf}$ 符号即可）。

深度变化图提供的深度差信息对于事件处理非常有用，可以帮助区分真实对象与伪影（如阴影或噪声）。



来自作者做的SLED 数据集，通过设定合适的深度差阈值（三个颜色方块），可以对这三种情况进行分类，如道路纹理、路灯光晕、天空中的噪声事件。

Method

1.The ALED Network

提出了一个全卷积循环网络Fully Convolutional Recurrent Network来估计异步 LiDAR 和事件数据的密集深度，称为 ALED 网络。

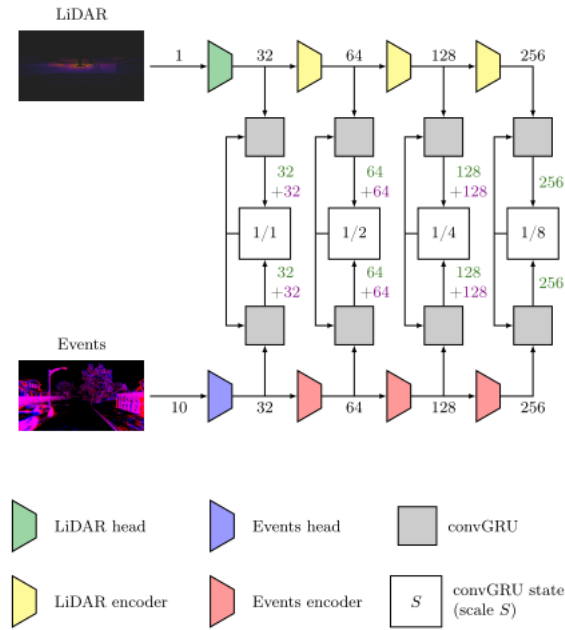


Fig. 2: The encoder part of the network.

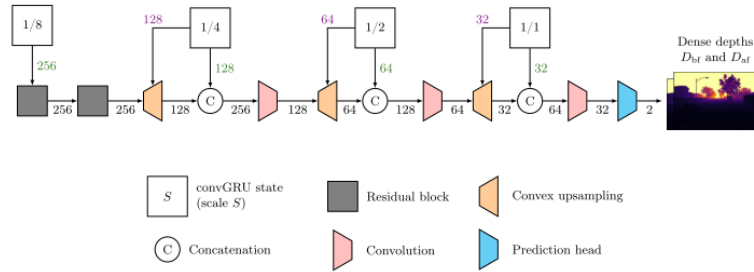


Fig. 3: The decoder part of the network.

实现：

部分	描述
编码器	<ul style="list-style-type: none"> - 负责融合异步事件和LiDAR特征，分为LiDAR编码器和事件编码器。 - 输入独立传递，经过编码头生成32通道的第一个特征图。 - 使用卷积编码器计算1/2、1/4和1/8尺度的特征图，通道数每次翻倍。 - 每个尺度的特征图输入卷积门控循环单元（convGRU）块，更新其状态。 - LiDAR和事件编码器共享状态，可以异步更新。
解码器	<ul style="list-style-type: none"> - convGRU状态在最低尺度经过两个残差块。 - 对于每个后续尺度，使用凸上采样对前一尺度的解码特征图进行上采样。 - convGRU设计使其状态的前半部分引导凸上采样。 - 上采样的解码特征和当前尺度的状态通过连接和卷积进行融合。 - 使用预测头生成两个最终深度图，以二通道张量形式输出，具有与事件输入相同的完整分辨率。

实现细节	- 编码头使用5的卷积核尺寸，LiDAR和事件编码器使用5的卷积核尺寸和步长2。
	- convGRU和残差块使用3的卷积核尺寸。
	- 凸上采样块中的卷积使用5的卷积核尺寸，连接后的卷积使用1的卷积核尺寸。
	- 预测层使用1的卷积核尺寸，卷积后使用PReLU激活函数。
	- ResNet编码器中使用Pan等人提出的实例归一化。
	- 网络总共包含2600万可训练参数。

2.Data Representation

Event Volume V for input events $\{e_i = (x_i, y_i, p_i, t_i)\}_{i=1}^N$ is described as:

$$V_{t,p,y,x} = \sum_{e_i, x_i=x, y_i=y, p_i=p} \max(0, 1 - |t - t_i^*|) \quad (1)$$

$$t_i^* = (B - 1) \frac{t_i - t_0}{t_N - t_0} \quad (2)$$

where (x, y) is the position of the event, t its timestamp, and p its polarity.

V 是对事件求和，将极性分为负极性和正极性，并使用了 5 个 "bins"。所以，实验中设置了 $B = 5$ 个 时间仓，并将负极性和正极性的沿着第一个维度进行连接，得到了一个形状为 $(10, H, W)$ 的张量。类比**时空体素**

通过对位于相同空间位置 (x, y) 、具有相同极性 p 的事件，根据它们发生的确切时间与当前时间的关系进行求和，生成离散化的事件体积。

将LiDAR测量的深度信息与事件相机的图像数据关联起来，以提供更丰富的深度信息给神经网络进行学习。

1. **LiDAR点云投影**：为了将LiDAR数据与事件相关联，需要将LiDAR点云投影到事件相机的图像平面上。
2. **深度图生成**：通过将LiDAR点投影到事件相机上，可以在图像平面上得到对应的像素位置。对于每个像素位置，选择其上的一个或多个LiDAR点，将这些LiDAR点的最小深度作为该像素的深度值。
3. **深度值的归一化**：生成的深度图中的深度值通常根据LiDAR的最大量程进行归一化，以便将深度值映射到区间 $[0, 1]$ 。使网络更轻松地学习深度信息，因为深度值变化范围被限制在标准范围内。
4. **处理没有LiDAR点的像素**：如果某些像素位置没有对应的LiDAR点，可以将这些位置的深度值设置为0。这是因为没有LiDAR点的地方通常表示相机视野内没有物体，因此深度可以设为0。

3.Loss functions

1. Pixel-wise L1 Loss:

- 主要监督损失应用于“before”和“after”深度图，用于衡量估计深度图和ground truth之间的差异。
- L1 loss定义为两个深度图差异的绝对值之和。

$$\mathcal{L}_{pw} = \sum_{x,y} \|D(x,y) - \hat{D}(x,y)\|$$

2. Multiscale Gradient Matching Loss(多尺度梯度匹配损失):

- 为了解决由于仅使用L1 loss而导致的深度图模糊和非平滑的问题，引入了多尺度梯度匹配损失。
- 该损失在“before”和“after”深度图上应用，通过比较多个尺度的梯度来提高深度不连续性的显著性，并平滑均匀区域。
- 多尺度梯度匹配损失定义了对应于不同尺度的梯度之间的欧氏距离。

$$\mathcal{L}_{\text{msg}} = \sum_{h \in \{1,2,4,8,16\}} \sum_{x,y} \left\| \mathbf{g}[D](x,y,h) - \mathbf{g}[\hat{D}](x,y,h) \right\|$$

$$\mathbf{g}[f](x,y,h) = (f(x+h,y) - f(x,y); f(x,y+h) - f(x,y))^T$$

3. Total loss L :

- 总体损失由L和多尺度梯度匹配损失组成。
- α 是多尺度梯度匹配损失的权重参数，用于平衡两种损失的贡献。
- 在实验中，观察到在训练的早期阶段，给予多尺度梯度匹配损失过大的权重会使网络无法得到正确的深度估计，因此在初始训练的第一个时期，将 α 设置为0.1，以强制网络主要使用L loss，随后的时期将 α 设置为1。

$$\mathcal{L} = \sum_{t=0}^T \sum_{\text{bf,af}} (\mathcal{L}_{\text{pw}}^t + \alpha \mathcal{L}_{\text{msg}}^t)$$

这样的损失设计有助于提高深度估计的质量，确保深度图在不同尺度上都能够准确且清晰地反映场景的深度结构。

The SLED Dataset

Table 1: Detailed content of our SLED dataset containing 160 sequences of 10 seconds each.

Map	Set	Environment	Features	Night seq.	Day seq.
Town01	Test	Town	Small buildings, bridges, distant forests and mountains, palm trees	4	16
Town02	Train	Town	Small buildings, plazas, forest road	4	16
Town03	Test	City	Tall and small buildings, roundabouts, tunnel, aerial railway	4	16
Town04	Val.	Town	Small buildings, highway, parking, lake, forests and mountains	4	16
Town05	Train	City	Tall buildings, parking, aerial beltway and railway	4	16
Town06	Train	Suburban	Small buildings, U-turns, distant hills	4	16
Town07	Train	Countryside	Barns, grain silos, fields, mountain road	4	16
Town10HD	Train	City	Buildings, monuments and sculptures, playgrounds, seaside	4	16

包含事件、LiDAR点云和深度的密集GT的数据集。

由于MVSEC数据集在低分辨率相机方面表现良好，但其GT是通过累积LiDAR传感器的点云构建的，可能在存在移动物体时引入误差。因此，为了获得完美同步和传感器校准的数据，以及完美的GT，作者使用CARLA模拟器生成了一个名为SLED（Synthetic LiDAR Events Depths）的数据集。

SLED数据集的特点：

- 由CARLA模拟器生成，包含160个10秒的序列，总计超过20分钟的数据。
- 记录在Town01到Town07和Town10HD上，每个map有20个序列，每个序列从不同的地理位置开始，以在数据集中涵盖多样的环境。
- 每个序列包含一个1280×720的事件相机、一个40通道LiDAR和一个与事件相机完美对齐的1280×720深度相机。
- 事件数据和深度图以200Hz的频率记录，LiDAR配置为以10Hz运行，RGB图像（1280×720）以30Hz的频率提供，与事件传感器对齐。
- LiDAR传感器配置最大范围为200米，模拟中还添加了AI控制的车辆和行人，以及不同时间、天气和云量的条件。

作者还通过CARLA模拟器的配置使事件相机使用线性强度刻度而不是默认的对数刻度，以使模拟器生成的事件更加真实。

Evaluation

Map	Cutoff	Dense depths errors				Sparse depths errors				Depth change map errors	
		On D_{bf}		On D_{af}		On D_{bf}		On D_{af}		Absolute error	Correctly classified events (with a threshold of $\pm 1m$)
		Raw	Rel.	Raw	Rel.	NN	ALED _s	NN	ALED _s		
Town01	10m	1.24m	20.99%	1.37m	23.60%	1.32m	1.46m	2.24m	1.79m	2.11m	90.27%
	20m	2.08m	23.06%	2.27m	25.48%	1.51m	1.84m	2.53m	2.15m	3.18m	85.07%
	30m	2.72m	23.76%	2.92m	26.03%	1.71m	2.37m	2.83m	2.67m	3.88m	81.68%
	100m	4.25m	24.01%	4.51m	26.07%	2.40m	3.48m	3.91m	3.95m	5.12m	77.48%
	200m	4.53m	17.20%	4.81m	18.66%	7.86m	5.44m	9.76m	6.23m	7.36m	75.54%
Town03	10m	2.00m	28.91%	2.09m	30.11%	0.47m	0.56m	0.67m	0.66m	1.14m	93.70%
	20m	2.85m	29.91%	2.97m	31.15%	0.64m	0.75m	1.12m	0.87m	2.54m	87.16%
	30m	3.33m	29.10%	3.45m	30.24%	0.92m	1.11m	1.61m	1.26m	3.23m	83.71%
	100m	4.60m	27.37%	4.77m	28.42%	1.88m	2.55m	3.17m	2.88m	4.47m	78.50%
	200m	4.86m	21.50%	5.03m	22.33%	4.43m	3.60m	5.93m	4.10m	6.20m	77.23%

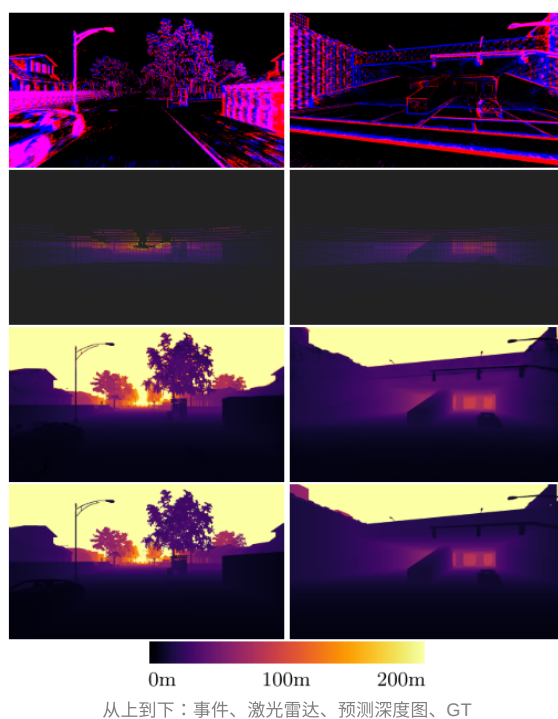
1.Dense Depths

On the SLED Dataset

在测试集上的数值结果显示在表中的“Dense depths errors”列中。

误差分析：LiDAR传感器的垂直覆盖范围较小，导致无法准确估计近地物体或建筑物顶部的深度，网络在这些区域的深度估计可能不准确,总体还是较好的，而对于完整的200米截止点，ALED能够在整个图像上准确估计深度。

定性结果展示网络使用事件作为LiDAR无法到达的区域的导向来估计整个图像的准确深度的能力。



On the MVSEC Dataset

ALED在MVSEC上训练和评估，不同截止深度下，ALED在准确性和鲁棒性上均超越了现有方法。

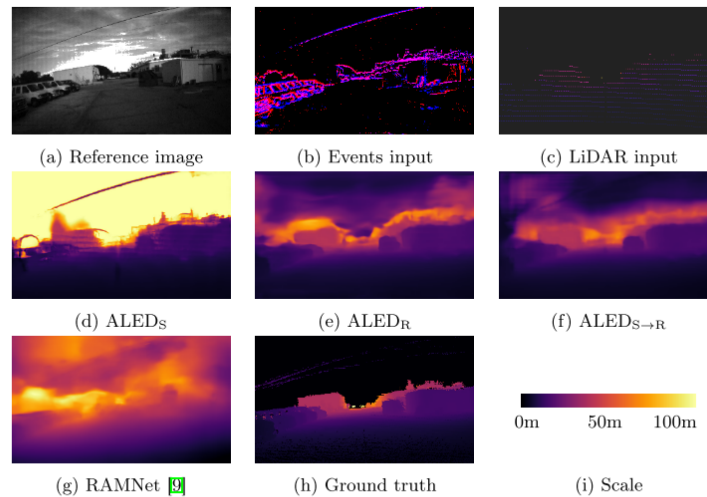


Fig. 5: Qualitative results on “Outdoor day 1” from the MVSEC dataset with real data.

R: 仅在模拟中进行训练，使用提出的 SLED 数据集。S: 仅在真实数据上进行训练，使用 MVSEC 数据集。S → R: 首先在 SLED 数据集上进行训练，然后在 MVSEC 数据集上进行微调。

S能够对天空区域进行准确预测，但由于缺乏微调，对地面物体的估计存在困难。MVSEC数据集缺乏天空区域的GT，效果差。S → R不平整性相较于R较小。与RAMNet的结果相比，物体边界更加突出，更接近于GT。

2.Associating a Depth to Each Event

研究的目的是不仅估计图像中每个像素的深度，还要将每个事件与LiDAR数据关联，以便进行更深入的3D分析。

比较了两种方法，一种是使用最近邻（NN）方法，即将每个事件关联到最近的LiDAR点的深度值；另一种是使用提出的ALED网络，该网络可以学习事件与LiDAR深度的更复杂关联，ALED网络能够更准确地学习事件和LiDAR深度之间的复杂关联。

3.Depth Difference

作者进行了两种评估：

1. 深度变化图的平均误差（Daf - Dbf）
2. 使用深度变化图进行事件分类：作者设定了一个深度变化的阈值（1米），并评估了通过深度变化图对事件进行分类的准确率。

在这些评估中，尽管深度变化图存在显著的绝对误差，但观察到仍然能够正确分类事件。在Town01上，成功率超过了75%，在Town03上超过了77%。

ALED网络并没有专门在深度变化图和分类任务上进行训练。因此作者认为通过设计一个专门用于这些任务的网络可能会带来更大的改进。

深度变化图不仅仅是评估网络性能的工具，还提供了一种了解网络对场景中深度变化敏感程度的方式。