文献阅读笔记

题目	Secrets of Event-Based Optical Flow（基于事件的光流的秘密）		
作者	Shintaro Shiba	文献来源	2022 ECCV
研究价值	**目前挑战**：CM 过拟合问题，已经通过改变目标函数来解决（从对比度到平均时间戳图像的能量），但是这种损失难以解释且训练难以收敛。 **本文价值**：克服了以上问题，并填补了空白--------合理地扩展 CM 以准确地估计密集光流；基于运动补偿的先验无监督学习方法在没有平均时间戳图像损失的情况下成功估计了光流。		
研究问题	解决 CM（对比度最大化框架）过拟合问题，扩展 CM 用于密集光流估计。 **我的理解**：上周文献阅读 CVPR2018 提出的通过 CM 找到与 event 对齐的点轨迹，event 对齐→IWE≈原始灰度图的梯度=微分方程，求解这个线性方程就可以得到图像。本文完全可以用 CM 来进行密集光流估计，即通过 f 找到对齐点轨迹。		
想法动机	1、如何设计目标函数以防止过拟合--------多参考聚焦损失函数可提高精度并阻止过拟合 2、如何扭曲事件以更好地处理遮挡--------一个原则性的时间感知流（采用多个时刻来减少过拟合） 3、如何改进事件的收敛--------对原始事件采用多尺度方法（以提高收敛性，避免局部最优）		
具体方法	本文针对密集光流估计，区别于（CVPR2018）采用的为光流场，即 $$x'_k = x_k - (t_k - t_{ref})v(x_k)$$ 其中 $\theta = v(x_k)$，密集光流估计产生光流场，该光流场包含了每个像素的运动矢量。 **1、多参考聚焦损失函数** 本文主要采用 tile-based 方法，将图像分成小块，每个小块中心定义 flow。 不同点：①采用 3 个参考时间（$t_1(\min)$, $t_{mid}(\frac{t_1+t_{N_e}}{2})$, $t_{N_e}(\max)$ ）来计算 f ②使用 IWE 梯度来测量事件对齐（之前是采用均方差） IWE 的平方梯度函数： $$G(\theta;t_{ref}) = \frac{1}{\lvert\Omega\rvert}\int_\Omega \left\lVert\nabla I(x;t_{ref})\right\rVert^2 dx$$ 所提出的多参考聚焦目标函数变成在多个参考时间处的 IWE 的 G 函数的平均值，通过零流量基线进行归一化： $$f(\theta) = (G(\theta;t_1) + 2G(\theta;t_{mid}) + G(\theta;t_{N_e}))/4G(0;-)$$ $f<1$，即 flow 比零流量基线还差，$f>1$，即产生的 IWE 图像更清晰。 **我的理解**：首先，分成小块可以降低复杂性，其次在每个小块中心定义 flow，会提升准确度。针对微小变化、纹理和边缘等局部特征的任务中，		

梯度通常比方差提供更高的精度；其次，梯度比方差更容易收敛；零流量基线的归一化，主要是比较两者的差异性。

2、时间感知流

①用 $v(x_k,t_k)$ 代替 $v(x_k)$（需要考虑事件的时空性，即不是每个像素触发时间都相同）

②受特征线法的启示，假设 flow 沿着它的流线是恒定的，即 $v(x(t),t)=const$，以 t_{mid} 作为边界，得到相应的偏微分方程（PDEs）：

$$\frac{\partial v}{\partial x}\frac{dx}{dt}+\frac{\partial v}{\partial t}=0$$

通过求解 PDEs 得出的 v 进行 wrap：

$$x_k^{'}=x_k+(t_k-t_{ref})\hat{v}(x_k,t_k)$$

两种方法来求解偏微分方程，一种是迎风差分，另一种是适合 Burgers 项的保守方案（使用 Buegers 求解器）

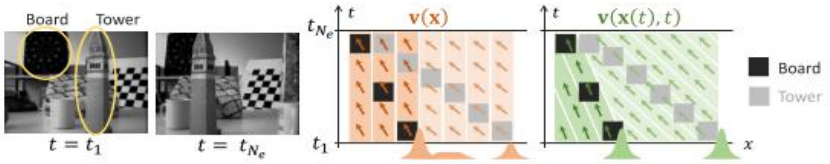


Fig. 3. *Time-aware Flow.* Traditional flow (4), inherited from frame-based approaches, assumes per-pixel constant flow $\mathbf{v}(\mathbf{x})$ = const, which cannot handle occlusions properly. The proposed space-time flow assumes constancy along streamlines, $\mathbf{v}(\mathbf{x}(t),t)$ = const, which allows us to handle occlusions more accurately. (See results in Fig. 8)

我的理解：如图所示，分别对 board 和 tower 的流线做为一个特征线，得到两个 v，以此来 wrap，能够有效的处理遮挡问题。

3、多尺度方法

我们 tile-based 方法与多尺度方法结合起来以避免局部最优；我们以从粗到细的方式应用基于图块的 CM（分辨率尺度 $N_l=5$）。

①使用双线性插值在任意两个尺度之间放大。

②对于后续集合 Ei+1 的初始化，可以用最精细的流 Ei 下采样到较粗糙的尺度 Ei+1；对于更精细的尺度，初始化为来自 Ei+1 的较粗糙尺度的上采样流和来自 Ei 的相同尺度流的平均值。

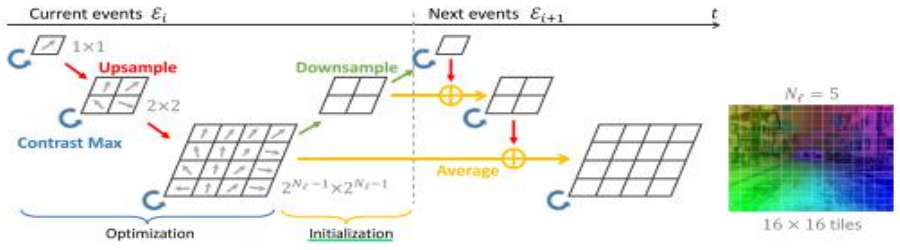


Fig. 4. *Multi-scale Approach* using tiles (rectangles) and raw events.

	复合目标：为了鼓励流的额外平滑性，添加正则化器 R（θ），flow 作为复合目标问题的解为： $$\theta^* = \arg\min_{\theta}(1/f(\theta) + \lambda R(\theta))$$ 其中 $\lambda > 0$ 为权重，正则化器选择 TV。
实验对比	数据集：主要在 MVSEC 数据集上评估我们的方法，还评估了提供地面实况流的数据集：DSEC； 评估指标：AEE（平均端点误差）、AEE 超过 3 像素误差的像素所占的百分比（百分比度量）和 FWL。 参数：使用 $N_l = 5$ 分辨率尺度，$\lambda = 0.0025$，以及最多 20 次迭代的 Newton-CG 优化算法。 **1、MVSEC 数据集** ①在三个室内序列和一个室外序列（列）上比较不同的方法（行） ②根据它们所需数据进行分类：监督学习（SL）需要地面实况流；半监督学习（SSL）使用灰度图像进行监督；无监督学习（USL）仅使用事件；基于模型（MB）不需要训练数据。Ours（有无时间感知；迎风差分；Burgers 求解器）

		indoor_flying1		indoor_flying2		indoor_flying3		outdoor_day1	
	$dt = 1$	AEE ↓	%Out ↓	AEE ↓	%Out ↓	AEE ↓	%Out ↓	AEE ↓	%Out ↓
SL	EV-FlowNet-EST [16]	0.97	0.91	1.38	8.20	1.43	6.47	–	–
SL	EV-FlowNet+ [46]	0.56	1.00	0.66	1.00	0.59	1.00	0.68	0.99
SL	E-RAFT [18]	–	–	–	–	–	–	**0.24**	1.70
SSL	EV-FlowNet (original) [57]	1.03	2.20	1.72	15.10	1.53	11.90	0.49	0.20
SSL	Spike-FlowNet [25]	0.84	–	1.28	–	1.11	–	0.49	–
SSL	Ziluo et al. [10]	0.57	0.10	0.79	1.60	0.72	1.30	0.42	**0.00**
USL	EV-FlowNet [58]	0.58	**0.00**	1.02	4.00	0.87	3.00	0.32	**0.00**
USL	EV-FlowNet (retrained) [34]	0.79	1.20	1.40	10.90	1.18	7.40	0.92	5.40
USL	FireFlowNet [34]	0.97	2.60	1.67	15.30	1.43	11.00	1.06	6.60
USL	ConvGRU-EV-FlowNet [21]	0.60	0.51	1.17	8.06	0.93	5.64	0.47	0.25
MB	Nagata et al. [30]	0.62	–	0.93	–	0.84	–	0.77	–
MB	Akolkar et al. [1]	1.52	–	1.59	–	1.89	–	2.75	–
MB	Brebion et al. [5]	0.52	0.10	0.98	5.50	0.71	2.10	0.53	0.20
MB	Ours (w/o time aware)	**0.42**	0.09	**0.60**	**0.59**	**0.50**	0.29	0.30	0.11
MB	Ours (Upwind)	**0.42**	0.10	**0.60**	**0.59**	**0.50**	**0.28**	0.30	0.10
MB	Ours (Burgers')	**0.42**	0.10	**0.60**	**0.59**	**0.50**	**0.28**	0.30	0.10
	$dt = 4$								
SSL	EV-FlowNet (original) [57]	2.25	24.70	4.05	45.30	3.45	39.70	1.23	7.30
SSL	Spike-FlowNet [25]	2.24	–	3.83	–	3.18	–	1.09	–
SSL	Ziluo et al. [10]	1.77	14.70	2.52	**26.10**	2.23	22.10	**0.99**	**3.90**
USL	EV-FlowNet [58]	2.18	24.20	3.85	46.80	3.18	47.80	1.30	9.70
USL	ConvGRU-EV-FlowNet [21]	2.16	21.51	3.90	40.72	3.00	29.60	1.69	12.50
MB	Ours (w/o time aware)	**1.68**	**12.79**	**2.49**	26.31	**2.06**	**18.93**	1.25	9.19
MB	Ours (Upwind)	1.69	12.83	**2.49**	26.37	**2.06**	19.02	1.25	9.23
MB	Ours (Burgers')	1.69	12.95	**2.49**	26.35	**2.06**	19.03	1.25	9.21

粗体是所有方法中最好的;下划线是第二好的

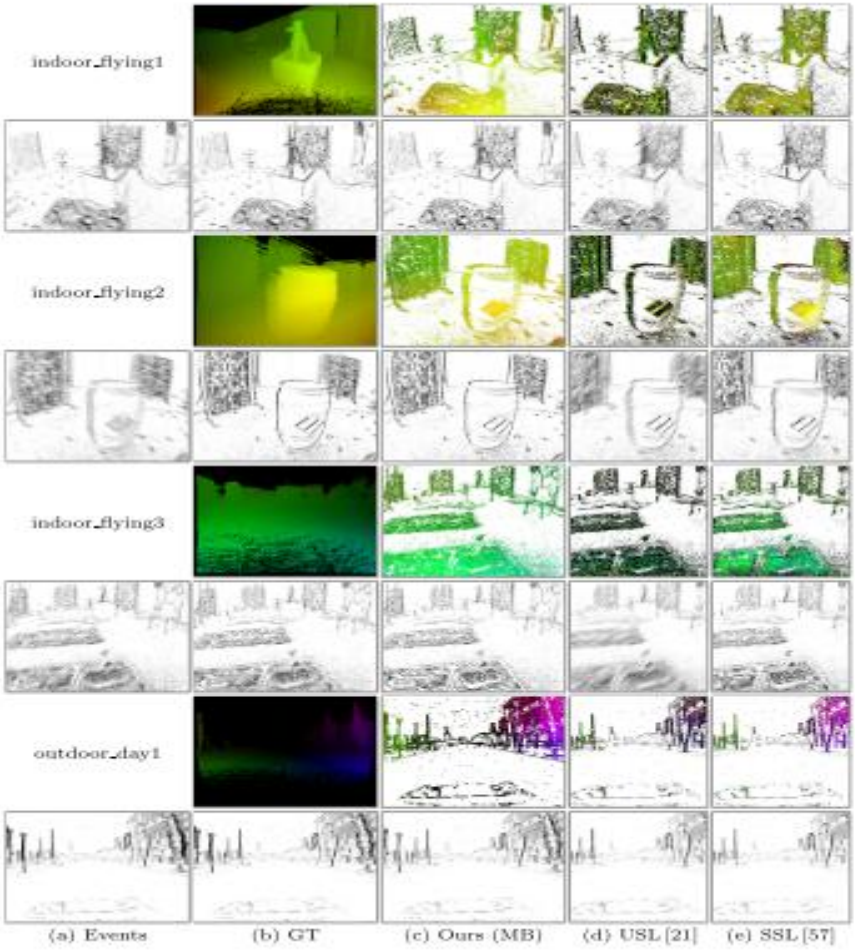
结果定量分析：

①可以看出本文的方法在所有室内序列中是最好的；并且在室外序列的 USL 和 MB 中是最好的。

②dt = 4 时的误差大约是 dt = 1 时的误差的四倍（根据时间间隔来说，确实是合理的）；

③我们的三个方法间没有显著差异，这是因为 MVSEC 数据集不包括大像

素位移或遮挡。



结果定性分析：

①本文的方法提供了比基线更清晰的 IWE，没有过拟合，估计的流量类似于地面实况；

②地面实况（GT）在整个图像平面上不可用，例如在 LiDAR 范围、FOV 或空间采样未覆盖的像素中。在室外序列中，来自 LiDAR 和相机运动的 GT 不能为独立移动对象（IMO）提供正确的流。

2、DSEC 数据集

Table 2. Results on DSEC test sequences [18]. For the calculation of FWL, we use events within 100ms. More sequences are provided in the supplementary material.

	thun_01_a			thun_01_b			zurich_city_15_a		
	AEE ↓	%Out ↓	FWL ↑	AEE ↓	%Out ↓	FWL ↑	AEE ↓	%Out ↓	FWL ↑
E-RAFT [18]	**0.65**	**1.87**	1.20	**0.58**	**1.52**	1.18	**0.59**	**1.30**	1.34
Ours	2.12	17.68	**1.24**	2.48	23.56	**1.24**	2.35	20.99	**1.41**

结果定量分析：

①正如预期的那样，这种监督学习方法在准确性方面优于本文的方法，因为（i）它具有额外的训练信息（GT 标签），以及（ii）它使用与评估

中使用的相同类型的 GT 信号进行训练。

②然而，本文的方法提供了有竞争力的结果，并且在 FWL 方面更好

Fig. 6. *DSEC results* on the interlaken_00_b test sequence (no GT available). Since GT is missing at IMOs and points outside the LiDAR's FOV, the supervised method [18] may provide inaccurate predictions around IMOs and road points close to the camera, whereas our method produces sharp edges. For visualization, we use 1M events.

结果定性分析：

①可以明显看出本文方法产生的 IWE 更清晰；

②由于 GT 在 IMO 和 LiDAR FOV 之外的点处缺失，因此 SL 可能会在 IMO 和靠近相机的道路点周围提供不准确的预测。

3、多参考焦距损失的影响

将单参考聚焦损失函数和本文多参考焦距损失进行比较

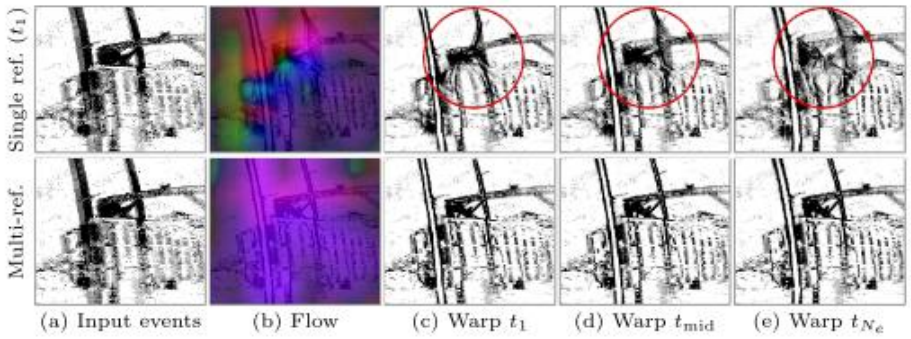


Fig. 7. *Effect of the multi-reference focus loss.*

结果分析：

①单参考聚焦损失函数过拟合到唯一时间 t1，会在其他时间（tmid 和 tNe）产生模糊的 IWE。相反，我们提出的多参考焦点损失，有利于在任何参考时间产生清晰 IWE 的流场。

②在 flow 中也很明显：来自单一参考损失的流动是不规则的，相比之下，来自多参考损耗的 flow 是相当规则的。

4、时间感知流的影响

在 MVSEC、DSEC 和 ECD 数据集上进行了将 GT 与本文三个方法进行比较。

Table 3. FWL (IWE sharpness) results on MVSEC, DSEC, and ECD. Higher is better.

	MVSEC ($dt = 4$)				ECD	DSEC	
	indoor1	indoor2	indoor3	outdoor1	slider_depth	thun_00_a	zurich_city_07_a
Ground truth	1.09	1.20	1.12	1.07	–	1.01	1.04
Ours: w/o time aware	**1.17**	**1.30**	**1.23**	**1.11**	1.88	1.39	1.57
Ours: Upwind	**1.17**	**1.30**	**1.23**	**1.11**	1.92	1.40	1.60
Ours: Burgers'	**1.17**	**1.30**	**1.23**	**1.11**	**1.93**	**1.42**	**1.63**

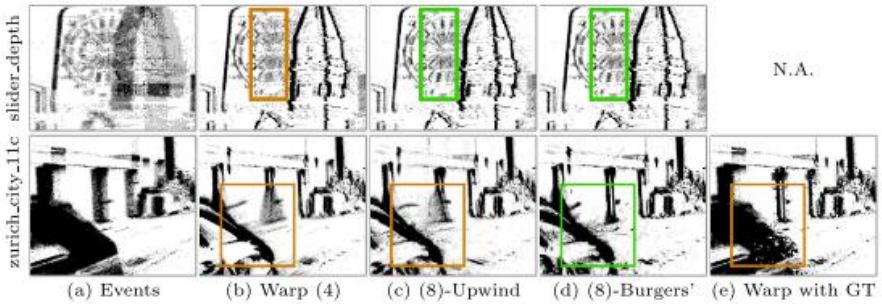


Fig. 8. *Time-aware flow.* Comparison between 3 versions of our method: Burgers', upwind, and no time-aware (4). At occlusions (dartboard in slider_depth [29] and garage door in DSEC [17]), upwind and Burgers' produce sharper IWEs. Due to the smoothness of the flow conferred by the tile-based approach, some small regions are still blurry.

结果分析：

①从表格分析，这三种方法都提供了比地面真相更清晰的 IWE，且在 ECD 和 DSEC 序列上最为明显，因为存在遮挡和较大的运动。

②从图像看出，在遮挡情况下（滑板深度的镖靶和 DSEC 中的车库门），逆风和 Burgers 产生更清晰的 IWE；

5、深度神经网络（DNN）的应用

所提出的秘密不仅适用于基于模型的方法，也适用于无监督学习方法。我们以无监督的方式训练 EV-FlowNet ，使用 θ^* 作为数据保真度项，Charbonnier 损失作为正则化器。为了确保泛化，我们在室内序列上训练我们的网络，并在室外序列上进行测试。

Table 4. Results of unsupervised learning on MVSEC's outdoor_day1 sequence.

	$dt = 1$			$dt = 4$		
	AEE ↓	%Out ↓	FWL ↑	AEE ↓	%Out ↓	FWL ↑
EV-FlowNet [58]	**0.32**	**0.00**	–	**1.30**	**9.70**	–
EV-FlowNet (retrained) [34]	0.92	5.40	–	–	–	–
ConvGRU-EV-FlowNet [21]	0.47	0.25	0.94	1.69	12.50	0.94
Our EV-FlowNet using (9)	0.36	0.09	**0.96**	1.49	11.72	**1.11**

结果分析：

①表 4 显示了与非监督学习方法的定量比较。在现有的方法中，我们的模型精度仅次于 EV-FlowNet 最好的清晰度 FWL。

②EV-FlowNet 是在室外第二天序列上训练的，这是与测试序列相似的驾驶序列，因此，根据训练数据的选择，EV-FlowNet 存在过度匹配，而我们的不是。

实验结论	本文扩展了 CM 框架来估计密集的光流,提出了原则性的解决方案来克服过拟合，遮挡和收敛问题。 综合实验表明，①在 MVSEC 室内基准测试、室外序列中的无监督和基于模型的方法中达到了最好的准确率；②在 DSEC 光流基准测试中也具有竞争力。③我们的方法提供了最清晰的 IWE，并暴露了基准数据的局限性（在 LiDAR 范围、FOV 或空间采样未覆盖的像素中；来自 LiDAR 和相机运动的 GT 不能为独立移动对象（IMO）提供正确的流）。④本文的方法可以应用到无监督的深度学习中，产生显着的效果。
不足	1、我们的方法基于亮度恒定性假设。因此，它很难从不是由于运动引起的事件（例如由闪烁的灯光引起的事件）中估计流量。 2、我们的方法可能会受到孔径问题的影响。如果图块变得更小，或者没有适当的正则化或初始化，flow 能会出错。 3、在事件较少的区域，例如均匀亮度区域和表观运动较小的区域，也很难估计光流。
收获启示	1、当处理不是由物体运动引起的事件时，光流估计可能会更具挑战性。通过结合背景建模、运动稳定性检测、光度不变假设、变分方法、先验知识和多传感器融合等技术，可以尝试提高光流估计的准确性。 2、优化亮度低的环境中的光流估计：可以使用适当的传感器（比如红外传感器）；也可以使用光照不变特征可以提高光流估计的稳定性；也可以用本文的对比度最大化对图像进行预处理，来增强图像质量，改善光流估计的准确性。