

# Ev-TTA: Test-Time Adaptation for Event-Based Object Recognition

作者：Junho Kim, Inwoo Hwang, Young Min Kim

期刊：CVPR

年份：2022

## 1、问题背景/研究动机

本文研究的主要问题：解决目标识别中不同条件下事件数据存在差异的问题

### - 现有方法存在的问题：

1、事件相机可以在快速运动或剧烈光照变化的场景中提供数据，但在极端条件中获得的事件通常是嘈杂的，缺乏视觉特征。

- 暗光/高速运动时事件的分布特性和正常情况存在差异

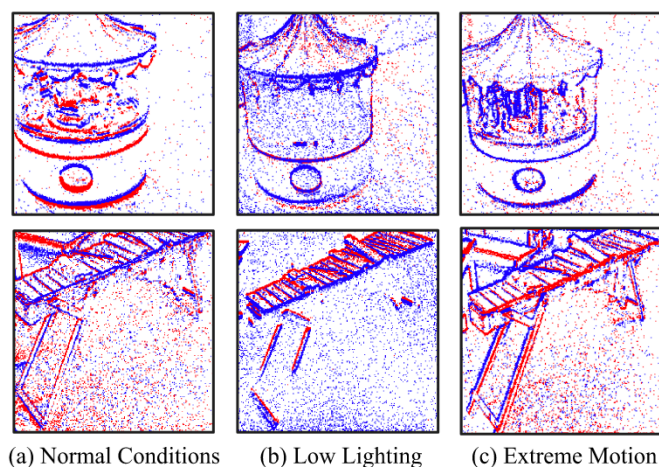


图1：各种环境条件下的事件可视化

(a) 正常情况；(b) 低照度情况；(c) 极端运动情况

图1显示，在正常照明和运动的事件与低照明或极端相机运动的事件之间存在明显的视觉对比。正、负性事件分别用蓝色和红色表示。低照明(b)中的事件表现出噪声爆发，大量的噪声事件从一个极性触发；与正常情况(a)相比，极端运动(c)中沿边缘触发的事件更密集。

2、现有方法在极端条件中性能下降，存在明显的域偏移。

- **域偏移(Domain Shift)**: 在机器学习中，训练数据和测试数据之间存在分布差异的情况。当模型在一个域中训练得到良好的性能，但在另一个域中表现不佳时，就出现了域偏移。

#### - 测试时间自适应 (Test-Time Adaptation, TTA)

- TTA是一种在测试阶段进行模型自适应的技术。

- 在机器学习中，通常使用训练数据对模型进行训练，然后使用测试数据对模型进行评估。然而，由于测试数据可能与训练数据存在差异，模型在测试阶段的性能可能会下降。为了解决这个问题，可以使用TTA来对模型进行自适应。

- TTA的目的是，通过在测试阶段对输入数据进行适应性变换或调整，使模型更好地适应测试数据的特点，提高模型的泛化能力和鲁棒性。

- TTA并不改变模型的基本结构或参数，而是在测试阶段对输入数据进行调整。

- TTA可以看作是一种在测试阶段对模型进行优化的技术，可以提高模型在实际应用中的性能和适应性。

- Pros: 有助于解决数据漂移、光照条件变化、意外噪声和其他可能影响现实场景中模型性能的因素等问题，适合用于测试数据可能偏离训练数据分布的实际应用。

## 2、解决方法

(1) 提出Ev-TTA，一种基于事件的物体识别的测试时间适应算法，用于解决基于事件的物体识别中的域偏移问题。

(2) 提出了一种利用空间一致性的条件去噪算法，利用事件两个极性之间的空间相关性来处理极端光照下的噪声，其中不同极性的事件显示出不同的噪声分布。

(3) 将Ev - TTA扩展到基于事件的回归任务。

(4) 使用广泛的事件表示来综合评估Ev-TTA在基于事件的物体识别中的性能。

### 3、具体实现

Ev - TTA将在源域上训练的预训练事件分类器适应于测量发生显著变化的目标域。定义**源域**为用于**训练的原始外部条件**，**目标域**为用于**测试的新条件**。例如，分类器可以使用正常光照下捕获的数据进行训练，然后在低光照下的数据上进行测试。

Ev - TTA由两个关键组件组成，利用了事件数据在时空域的显著特点：

- 1) 测试时间适应策略加强了对时间相邻流预测的一致性，新的损失函数联合最小化相邻事件片段对之间的差异，同时选择性地最小化预测的熵。
- 2) 提出移除在空间上缺乏相反极性的相邻事件。基于在极端光照下，事件流中的严重噪声在一个极性上产生。

#### - 测试时间适应策略

这部分的目标：旨在**加强相邻事件序列之间时间域的一致性**。对于在测试阶段使预先训练的模型适应新的、看不见的环境至关重要。

为了实现时间一致性，Ev-TTA从输入事件流中提取相同长度的随机切片，并利用两个损失函数（预测相似性损失和选择性熵损失）来实现时间一致性。

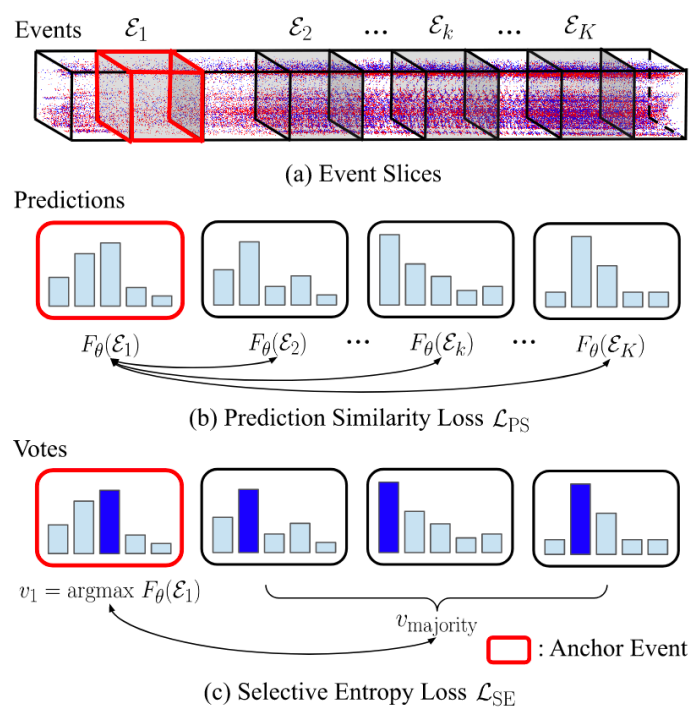


图2：(a) 事件切片；(b) 预测相似性损失；(c) 选择性熵损失

### • 预测相似性损失

由于广泛的两两比较会导致计算量的平方增加，因此使用**第一个事件片作为锚点**，利用**对称KL散度**来预测相似性损失。

$$\mathcal{L}_{\text{PS}} = \frac{1}{2} \sum_{k=2}^K S_{\text{KL}}(F_{\theta}(\mathcal{E}_1), F_{\theta}(\mathcal{E}_k)).$$

- 该损失**最小化**了第一个事件切片和其余事件切片的**预测之间的差异**。
- 使得时间相邻事件的**预测标签分布相似**，确保了**时间一致性**

### • 选择性熵损失

提出仅当预测与其他事件片一致时，有选择地最小化第一个事件切片的预测熵。

- 一致性通过检查预测的类标签是否与时间上相邻的事件一致来确定。
- 具体来说，每个事件片**对概率最大的类标签进行投票**，当投票一致时，有选择地**最小化锚的预测熵**。

$$\mathcal{L}_{\text{SE}} = \begin{cases} H(F_{\theta}(\mathcal{E}_1)) & \text{if consistent} \\ 0 & \text{if inconsistent.} \end{cases}$$

- 这种损失函数通过**减少预测的不确定性**来帮助进一步加强时间一致性。

### • 优化

给定总的训练损失函数，**限制优化只在预训练分类器的批归一化层上操作**。

- 当**目标域数据稀缺**时，改变整个参数集可能会使模型偏离从预训练中获得的基本先验；**只优化批归一化层**时，不同的标签可以更好地分离。

### • 对回归的拓展

除了分类任务外，Ev-TTA还可以扩展到**回归任务**。回归任务涉及**预测连续值**。例如，预测自动驾驶的转向角度，回归器旨在**预测转向角的均值和方差**。为了使Ev-TTA适应回归任务，作者对算法中使用的损失函数进行了三个修改：

- 首先用**高斯分布的KL散度**代替对称KL散度

$$S_{\text{KL}}(F_{\theta}(\mathcal{E}_1), F_{\theta}(\mathcal{E}_k)) = \frac{\sigma_1^4 + \sigma_k^4 + (\sigma_1^2 + \sigma_k^2)(\mu_1 - \mu_k)^2}{2\sigma_1^2\sigma_k^2}. \quad (4)$$

- 用**高斯分布的熵**对选择性熵损失中的熵进行修正

$$H(F_{\theta}(\mathcal{E}_1)) = \log \sigma_1 \sqrt{2\pi e}.$$

- 调整连续网络输出的一致性标准。如果锚事件预测的方差在相邻预测的一定方差范围内，则锚事件被认为是一致的

## - 条件去噪

该部分的目标：在**低光照条件**下使用基于事件空间一致性的准则来**有条件地去除噪声**。

弱光条件显著恶化了基于事件的视觉算法，在弱光下，有效的事件信号会变小，暗电流会触发大量的噪声。且从图1可以看出，事件流中的**严重噪声在一个极性上产生**。极端光照条件下的强噪声超出了以往方法所能处理的范围。

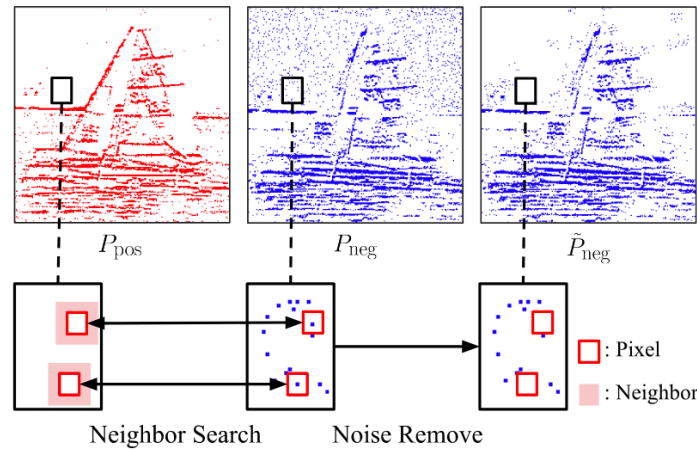


图3：条件去噪示意图

该算法在双通道事件表示上运行，其中一个通道包含积极事件，另一个通道包含消极事件。该算法识别单个极性中占主导地位的噪声突发，当**包含事件的像素缺少相**

反极性的空间邻域时，则消除噪声。去噪操作只有在正负事件比例存在较大不平衡的情况下才会发生。

- 确定噪声突发的存在

算法利用不受低光照条件影响的源域的统计信息对目标域的事件比进行变换

$$T(R) = \frac{\mu_{\text{neg}} R - \mu_{\text{pos}}}{\sqrt{\sigma_{\text{pos}}^2 R^2 - 2\rho\sigma_{\text{pos}}\sigma_{\text{neg}}R + \sigma_{\text{neg}}^2 R^2}},$$

如果用上式变换后的比值服从标准高斯分布，我们可以假设目标域不受噪声的影响。

如果该批次包含显著的极性不平衡，则使用条件去噪算法。修改后的通道代替原来的通道，形成新的锚事件表示 $\bar{P} = \{p_{\text{pos}}, \bar{p}_{\text{neg}}\}$ ，并随后计算损失。

## 4、实验对比

在这一部分中，对Ev-TTA的各个方面进行了实验。

- 1) 使用广泛的事件表示来综合评估Ev-TTA在基于事件的物体识别中的性能。

- 使用了六种事件表示：二进制事件图像、事件直方图、时间戳图像、时间曲面、排序时间曲面和DiST。

- 2) 验证Ev-TTA的各个关键组成部分的重要性。

### - 数据集

使用N-ImageNet数据集对Ev-TTA进行评估，以系统地评估在大范围变化下的鲁棒性增强。

N-ImageNet是一个基于事件的物体识别数据集，由原始训练集和在不同相机运动和光照变化下记录的9个变体组成。

### - Ev-TTA在基于事件的物体识别中的性能

本文使用原始N-ImageNet数据集训练具有6种事件表示的分类器，并在N-ImageNet变体上对分类器进行评估，与4种基线方法进行比较。



Change	None	Trajectory					Brightness				Average
Validation Dataset	Orig.	1	2	3	4	5	6	7	8	9	All
No Adaptation	46.76	43.32	33.78	39.56	24.78	36.16	21.52	30.31	36.60	34.91	33.44
Mumtaz <i>et al.</i> [22]	-	46.27	46.04	46.35	43.27	44.61	25.59	35.23	45.73	45.48	42.07
URIE [35]	-	42.04	41.45	42.48	38.66	40.43	17.59	29.63	41.77	41.45	37.28
SENTRY [28]	-	46.63	46.51	46.45	42.11	44.44	21.92	34.78	45.53	45.13	41.50
Tent [40]	-	43.86	44.96	44.82	41.55	42.81	26.47	34.87	44.10	44.00	40.83
Ev-TTA	-	<b>47.99</b>	<b>47.38</b>	<b>47.47</b>	<b>44.54</b>	<b>46.28</b>	<b>29.46</b>	<b>38.44</b>	<b>47.45</b>	<b>46.90</b>	<b>43.99</b>
No Adaptation (Max)	-	45.17	36.58	42.28	26.57	38.70	24.39	32.76	38.99	37.37	35.87
Ev-TTA (Min)	-	<b>45.50</b>	<b>46.46</b>	<b>46.58</b>	<b>43.48</b>	<b>43.87</b>	<b>27.28</b>	<b>37.06</b>	<b>46.72</b>	<b>46.12</b>	<b>42.91</b>
Ev-TTA (Online)	-	44.77	44.80	45.05	41.77	43.12	26.43	35.42	44.42	44.22	41.11

表1：在N - ImageNet及其变体上的鲁棒性评估结果

表1显示了在六种表示中平均的分类精度。数据集中的某些变化引起的大的域偏移会导致性能急剧下降而无法适应。Ev-TTA优于所有其他基线，并成功地将预训练的分类器适应于新的、不可见的环境。

此外，通过比较表1中每个N - ImageNet变体的事件表征中准确率最高的“*No Adapt (Max)*”和每个变体准确率最低的“*Ev-TTA (Min)*”，验证了Ev-TTA的性能增强是通用的，所有测试的事件表示都显示出很大的改进。

## - Ev-TTA在真实世界的性能

这一部分在**Prophesee Megapixel数据集**验证Ev - TTA在不受控制的外部真实世界中的性能。该数据集包含了真实世界记录的对象标签，按天进行记录，包含5个物体标签，从中选取3个(汽车、卡车、公交车)进行实验。

Dataset	Source	Day 1	Day 2	Day 3	Day 4	Day 5
None	77.30	70.47	78.53	74.88	71.36	83.37
Tent [40]	-	73.60	80.81	75.71	74.74	87.37
Ev-TTA	-	<b>74.83</b>	<b>82.77</b>	<b>77.15</b>	<b>74.76</b>	<b>88.38</b>

表2：在Prophesee Megapixel数据集上的评估结果

表2显示，在所有测试的记录中，Ev - TTA优于Tent。与Tent的简单熵最小化相比，Ev - TTA利用事件的时间特性增加了额外的损失函数，获得了更好的性能。

实验结果验证了Ev - TTA在包含事件相机的实际场景中的适用性。

### - 仿真与现实差距

本文利用事件相机模拟器，通过在ImageNet图像周围移动一个虚拟事件相机创建了N-ImageNet的合成版本，命名为**SimN-ImageNet**。

通过将Ev-TTA应用于SimN-ImageNet中的预训练模型，并观察在N - ImageNet验证集中的性能变化，评估了Ev-TTA对Sim2Real的适应性。

Representation	Sim	None	Tent [40]	Ev-TTA
Timestamp Image [25]	53.53	31.36	38.96	<b>40.66</b>
Binary Event Image [7]	54.63	26.62	38.67	<b>40.94</b>
Event Histogram [19]	44.44	21.97	30.2	<b>34.87</b>

表3: Sim2Real gap的评估结果

表3显示了三种测试表示的结果，即时间戳图像，二进制事件图像和事件直方图。实验证明，Ev - TTA在所有情况下都表现出最高的精度，有效降低了Sim2Real差距带来的性能损失。

### - 回归任务

使用DDD17数据集，评估自适应策略在转向角预测的回归任务中的性能。该数据集包含大约12小时的带注释的驾驶记录，在各种外部条件下获取。

Scene Type	City (Source)	Freeway	City	Town	City
Time	Day (Source)	Evening	Night	Day	Day
None	25.48	6.15	16.09	32.01	43.02
Tent [40]	-	6.52	15.65	30.94	41.66
Ev-TTA	-	<b>5.84</b>	<b>15.45</b>	<b>30.65</b>	<b>41.44</b>

表4: 使用DDD17数据集对转向角预测的评估结果

表4显示了在转向角预测的回归任务的评估结果，其中测量了地面真实转向角的RMSE。在所有测试场景中，Ev-TTA都优于Tent，验证了Ev-TTA可以扩展到回归任务中，并降低预测误差。但与分类任务相比，性能提升并不显著。



## - 消融实验

这一部分对Ev-TTA的各个组成部分进行了消融实验。实验在N - ImageNet数据集的6和7变体中进行，这些是N - ImageNet变体中最具挑战性的部分，因为是在低光照条件下记录的，包含大量的噪声。

Method	Validation 6	Validation 7
Tent [40]	21.16	30.02
Tent + $\mathcal{L}_{PS}$	26.51	35.83
Tent + $\mathcal{L}_{PS}$ + $\mathcal{L}_{SE}$	26.82	36.87
Tent + $\mathcal{L}_{SE}$ (SENTRY [28])	20.13	33.92
Tent + $\mathcal{L}_{SE}$ (Ignore Inconsistency)	27.13	36.69
Tent + $\mathcal{L}_{PS}$ + $\mathcal{L}_{SE}$ + CD (Ev-TTA)	<b>29.20</b>	<b>38.45</b>

表5: Ev-TTA关键组成部分的消融研究

首先验证预测相似性损失、选择性熵损失和条件去噪的效果。

如表5所示，通过在Tent上添加预测相似性损失（第2行）以及选择性熵损失在性能增益（第3行），性能得到了很大的提升。相比于SENTRY最大化不一致样本的熵（第4行），简单地忽略此类样本是更有效的（第5行）。

最后，通过对比表5的第3行和第6行，可以看出，条件去噪对低光照条件下普遍存在的噪声影响有显著的性能提升。

## 5、总结

- 提出了一种基于事件的物体识别测试时间自适应算法Ev-TTA。
- 采用迁移的方法进行Test-Time Adaptation策略，加强相邻事件序列之间时间域的一致性。
- 提出了一种条件去噪算法，利用事件的空间一致性来消除低光条件下的噪声。
- 通过对损失函数进行修改，将Ev-TTA扩展到回归任务中。
- 在N-ImageNet数据集上进行了实验，证明了该方法在增强性能上的有效性与广泛的适用性。且该方法轻量，可以通过简单的掩码实现，适用于实际应用。

## 6、启发

- **极端条件事件的分布特性和正常情况存在差异可能的原因**：由于事件生成的阈值是动态的，在低光照情况下该阈值较低，此时噪声更容易触发事件，导致噪声水平偏高；同时物体边缘对比度不如正常光照，造成物体边缘产生的事件不清晰。
- **测试时间适应**：允许预先训练的模型在测试阶段适应新的、看不见的环境。这种方法减轻由于**测试数据可能与训练数据存在差异**带来的影响，适合用于**测试数据可能偏离训练数据分布**的实际应用。
- **事件的时空特性**：利用**事件的时空特性和极性之间的空间相关性**，有助于有效处理极端环境带来的噪声

## 7、补充（不完整）

### - 低分辨率事件可能的影响：

- 1) 关键特征丢失，降低了识别和检测算法的准确性；
- 2) 将噪声或伪事件误认为是真实事件；
- 3) 时空特性的缺失，低分辨率事件在连续帧之间的位置变化难以确定

### - 可能的应对方法：

- 提升低分辨率事件的质量
- **增强算法的鲁棒性**，使其能够处理低分辨率情况下的噪声和不确定性。（eg.

Test-Time Adaptation策略）

- 融合多模态信息弥补低分辨率事件
- 数据增强，通过模拟低分辨率条件来增强模型对低分辨率图像的适应性