Generalizing Event-Based Motion Deblurring in Real-World Scenarios

简介

目前基于事件的运动去模糊假设了在相同的输入空间分辨率和特定的模糊分布，这在实际应用中受到限制，为了解决这一限制，本文提出了一种尺度感知网络，开发了一个两阶段的自监督学习方案以及引入了由多尺度模糊帧和事件组成的真实世界数据集，以促进基于事件的去模糊研究。

一、去模糊方法

1.1 基于帧的方法

通常假设特定的运动模式，在具有复杂非均匀运动的现实场景中面临挑战，且由于模糊图像中的运动模糊和纹理擦除问题，基于帧的方法很难从严重模糊的帧中提取精确的运动并恢复准确的潜在图像。

1.2 基于事件的方法

事件的微秒级低延迟使动态场景几乎可以连续观察，并减轻了模糊帧中的运动模糊，另外，事件流记录的亮度变化对应于高对比度边缘，补偿运动模糊所擦除的强度纹理。但是目前基于事件的去模糊方法的性能通常局限于训练数据的分布，在现实场景中存在局限。

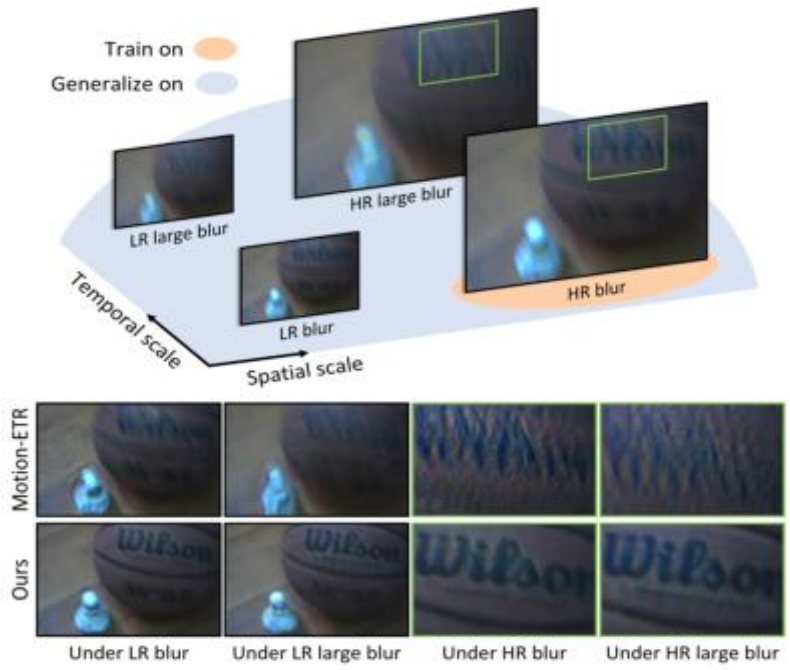
1.2.1 时间限制

先前方法在固定的曝光时间范围内合成或收集模糊帧进行训练，假设了具有特定模糊分布的运动模糊。然而，现实世界的运动模糊经常在高度动态的场景中，这导致预训练模型的性能下降。

1.2.2 空间限制

先前方法采用相同空间分辨率的帧和事件作为输入，忽略了实践中基于帧的摄像机通常比基于事件的摄像机具有更大的空间分辨率。

1.3 本文方法



图一 网络结构

设计了一个尺度感知网络(SAN)来从单个 HR 模糊帧及其并发 LR 事件中提取高帧率的 HR 序列。实现了一个多尺度特征融合(MSFF)模块，以空间连续的方式表示帧和事件特征，从而允许灵活设置输入空间分辨率。

时间维度上，提出了一种曝光引导事件表示(Exposure-Guided Event Representation, EGER)，可以在不需要修改模型或重新训练的情况下任意选择目标潜在图像。为了适应现实世界的数据分布，进一步提出了一种两阶段自监督学习框架。在第一阶段，利用模糊度的相关性对隐图像的恢复亮度和结构进行有效的监督。然后，采用自蒸馏策略对运动模糊的去模糊性能进行泛化，以处理不同时空尺度的运动模糊。

自蒸馏是一种知识蒸馏的变体，其中一个模型（通常被称为学生模型）通过学习另一个模型(通常被称为教师模型)的知识来提高性能。用于训练和改进深度神经网络模型的方法，基于模型自身的输出进行知识迁移和模型改进。

SAN 用 LR 事件去模糊 HR 帧，同时实现输入空间分辨率的灵活设置。

二、 具体方法

2.1 公式化

$$I(t) = \frac{B_T}{E^{\uparrow}(t,\mathcal{T})}, \tag{1}$$

$E^{\uparrow}(t,\mathcal{T})$表示$E(t,\mathcal{T})$的上采样版本，以匹配 B_T 的空间分辨率，其中$E(t,\mathcal{T}) = \frac{1}{T}\int_{f\in\mathcal{T}} \exp\left(c\int_t^f e(s)ds\right)df$ 。B_T 表示曝光时间为 T 时拍摄的模糊画面。

$$B_T = \frac{E^{\uparrow}(t,\mathcal{T})}{E^{\uparrow}(t,\tilde{\mathcal{T}})} B_{\tilde{T}} \tag{2}$$

该式将较模糊的帧$B_{\tilde{T}}$转换为较模糊的潜在图像 B_T。之后将 SAN 近似为一般函数：

$$\mathrm{L} = \frac{\mathrm{E}^{\uparrow}(\mathrm{t},\hat{\mathcal{T}})}{\mathrm{E}^{\uparrow}(\mathrm{t},\tilde{\mathcal{T}})} \mathrm{B}_{\tilde{\mathrm{T}}} \approx \mathrm{SAN}(\hat{\mathcal{T}}; \mathrm{B}_{\tilde{\mathrm{T}}}, \mathcal{E}_{\tilde{\mathcal{T}}}) \tag{3}$$

其中$\hat{\tau} \subset \tilde{\tau}$，控制目标潜像 L 的输出时间尺度，因此可以通过设置不同的$\hat{T}$来恢复图像。

上采样，在卷积神经网络中，由于输入图像通过卷积神经网络提取特征后，输出的尺寸往往会变小，而有时需要将图像恢复到原来的尺寸以便进行进一步的计算，这个使图像由小分辨率映射到大分辨率的操作叫做上采样。

2.2 曝光引导事件表示(EGER)

(a) Event representation

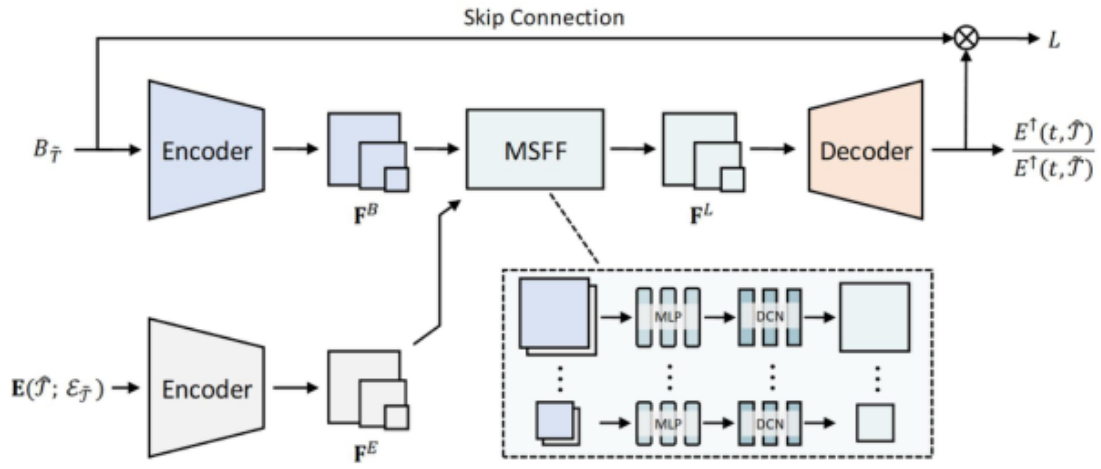
该模块可以理解为，给定事件流$\mathcal{E}_{\tilde{\tau}}$，将$\bar{T}$均匀划分为 N 个时间仓，并生成三个 2N × H × W 的事件张量 E_1， E_2 和 E_3，其中 2，H，W 表示事件极性，高度和宽度。三个张量累加了[ts,te]时间段分割的事件，然后串联形成 EGER。

[ts,te]为目标曝光周期，进行曝光是为了改进亮度和对比度。

该模块作用是使得输入事件流可以根据不同的$\bar{T}$表示$\mathbf{E}(\mathcal{T}; \mathcal{E}_{\tilde{\tau}})$，从而 SAN 能够确定输出的时间尺度，并从相同的输入中恢复模糊图像。

2.3 SAN 及多尺度特征融合(MSFF)

2.3.1 SAN



(b) Scale-aware network

该结构可以理解为，将模糊帧和事件通过两个 Encoder 网络提取多尺度模糊 $F^B=F^Bi$ 和事件特征 $F^E=F^Ei$。i 表示第 i 个尺度。之后通过 MSFF 模块，从而生成潜在图像特征，然后将生成特征 F^L 通过 Decoder 网络传递，恢复潜像 I。

2.3.2 MSFF 原理

该模块结构如上，在该模块中使用模糊特征提供亮度参考并指导事件特征的上采样。即采用 MLP 从跨模态局部特征中预测融合特征值，然后将粗融合的特征通过 DCN 进行细化，生成最终的潜在图像特征 F^L。此外，由于坐标连续 MSFF 可以灵活设置输入空间尺度。

MSFF 能够有效结合多个尺度上提取的特征并融合在一起，使系统能够全面的理解场景从而使得 SAN 网络可以灵活设置输入空间分辨率。

跨模态主要关注如何通过一个模态的信息来理解另一个模态的信息。由于基于帧的相机和基于事件的相机之间的跨传感器差距，需要从跨模态局部特征中预测融合特征值。

2.4 两阶段自监督学习

2.4.1 第一阶段

利用模糊度的相关性约束潜在图像的恢复亮度和结构，具体可以理解为，通过平均 B_T 的 M 相邻模糊帧来合成更模糊的图像$B_{\tilde{T}}$，通过公式(4)让网络学习从$B_{\tilde{T}}$中恢复 B_T。

恢复其中结构即对公式(1)每个 I(t)精确估计$E^{\uparrow}(t,\mathcal{T})$，该功能由公式(5)实现。然后通过设置$\hat{\mathcal{T}}=[t,t]$能够恢复锐利潜在图像，这里由于是 blur2sharp 的转换于

是将$E^{\uparrow}(t,\hat{\mathcal{T}})$假设为 1。从而从公式(5)转换到公式(6)。

最后通过公式(7)约束恢复后的 I(t)的结构。由于$\mathrm{SAN}^{E}(\mathcal{T};B_{\tilde{\mathcal{T}}},\mathcal{E}_{\tilde{\mathcal{T}}})$在式(4)受到约束从而提供了强大的自监督避免崩溃，保证结构恢复。通过 L_{BC} 及 L_{SC}，在保证清晰潜像的亮度和结构的前提下，有效实现了运动去模糊。

通过不断训练学习，网络对于模糊图像的处理的能力就会越来越强。

$$\mathcal{L}_{BC} = \| B_T - \mathrm{SAN}(\mathcal{T};B_{\tilde{\mathcal{T}}},\mathcal{E}_{\tilde{\mathcal{T}}}) \|_1 \tag{4}$$

$$\frac{E^{\uparrow}(t,\hat{\mathcal{T}})}{E^{\uparrow}(t,\tilde{\mathcal{T}})} \approx \mathrm{SAN}^{E}(\hat{\mathcal{T}};B_{\tilde{\mathcal{T}}},\mathcal{E}_{\tilde{\mathcal{T}}}) \tag{5}$$

$$\frac{1}{E^{\uparrow}(t,\tilde{\mathcal{T}})} \approx \mathrm{SAN}^{E}([t,t];B_{\tilde{\mathcal{T}}},\mathcal{E}_{\tilde{\tau}}). \tag{6}$$

$$\mathcal{L}_{SC} = \left\| \mathrm{SAN}^{E}(\mathcal{T};B_{\tilde{\mathcal{T}}},\mathcal{E}_{\tilde{\mathcal{T}}}) - \frac{\mathrm{SAN}^{E}([t,t];B_{\tilde{\mathcal{T}}},\mathcal{E}_{\tilde{\mathcal{T}}})}{\mathrm{SAN}^{E}([t,t];B_{\mathcal{T}},\mathcal{E}_{\tau})} \right\|_1 \tag{7}$$

2.4.2 第二阶段

训练目的是为了在时间和空间两个维度上推广 SAN 的去模糊性能。

时间尺度

对于时间泛化提出了自蒸馏损失，如式(8)

$$\mathcal{L}_{TG} = \| \overline{\mathrm{SAN}}([t,t];B_T,\mathcal{E}_{\mathcal{T}}) - \mathrm{SAN}([t,t];B_{\tilde{\mathcal{T}}},\mathcal{E}_{\tilde{\mathcal{T}}}) \|_1 \tag{8}$$

其中$\overline{\mathrm{SAN}}$表示使用 L_{BC} 和 L_{SC} 预训练的固定教师模型，SAN 表示从$\overline{\mathrm{SAN}}$加载并继续训练的学生网络。由于$\overline{\mathrm{SAN}}$可以从较不模糊的帧 B_T 恢复恢复相对可靠的潜在图像，因此将$\overline{\mathrm{SAN}}$的输出作为伪地真图像，并教 SAN 对较模糊的帧$B_{\tilde{\mathcal{T}}}$进行去模糊处理，从而提高了 SAN 的去模糊能力，并推广了 SAN 处理不同时间尺度运动模糊的性能。

自蒸馏使用不同的时间尺度生成目标标签，使得模型在学习时同时考虑到不同时间尺度的信息，提高模型的泛化能力，从而可以适应不同时间尺度下的输入数据。

空间尺度

SAN 学习在固定空间比例上用 LR 事件去模糊 HR 帧，为了更好的推广空间域的去模糊性能，在任意空间比例中进行随机下采样，从而传播到不同的输入空间尺度。如式(9)

$$\mathcal{L}_{SG} = \| \overline{\mathrm{SAN}}^{\downarrow}([t,t];B_T,\mathcal{E}_{\mathcal{T}}) - \mathrm{SAN}([t,t];B^{\downarrow}_{\tilde{\mathcal{T}}},\mathcal{E}_{\tilde{\mathcal{T}}}) \|_1 \tag{9}$$

这里下采样是为了增大感受野，使得网络能够学到更加全局的信息。

使用 LR 事件模糊 HR 帧是为了在高分辨率图像上模拟低分辨率事件的影响，达到仿真真实场景的模糊效果，降低过分辨率的感知细节，加强模型泛化能力的作用，从而更好的推广空间域的去模糊性能。

三、 实验

3.1 数据库

3.1.1 Ev-REDS

合成数据集，在不同的空间尺度上进行了评价。

3.1.2 HS-ERGB

包含相同空间分辨率的清晰视频和真实事件，因此我们使用它来评估不同时间尺度的运动模糊。

3.1.3 MS-RBD

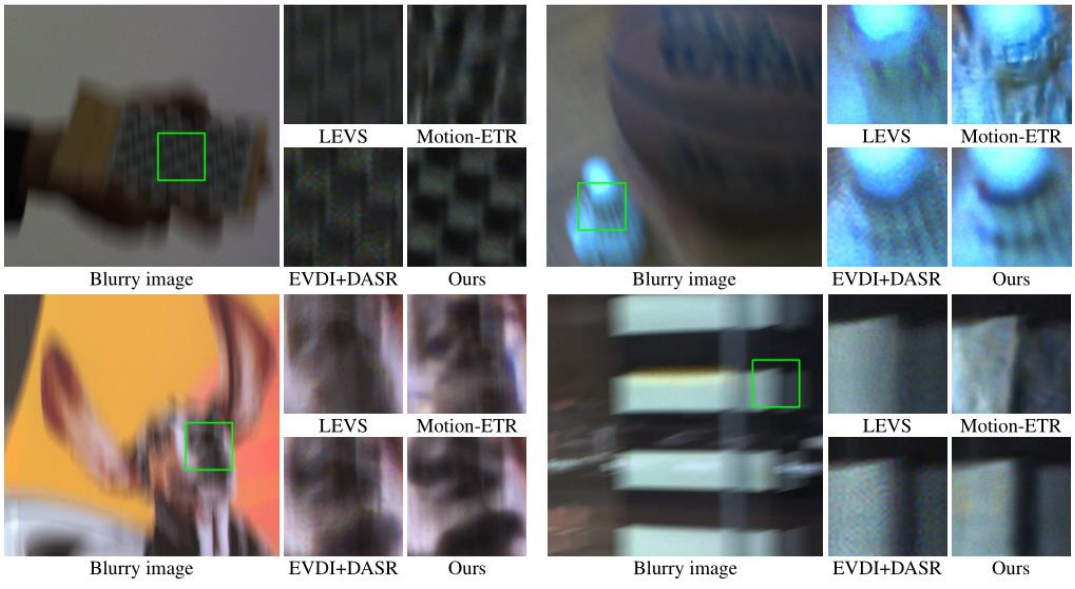
使用 FLIR Blackfly S global shutter RGB 及 DA VIS346 相机构建的多尺度真实世界模糊数据集。

3.2 对比及假设

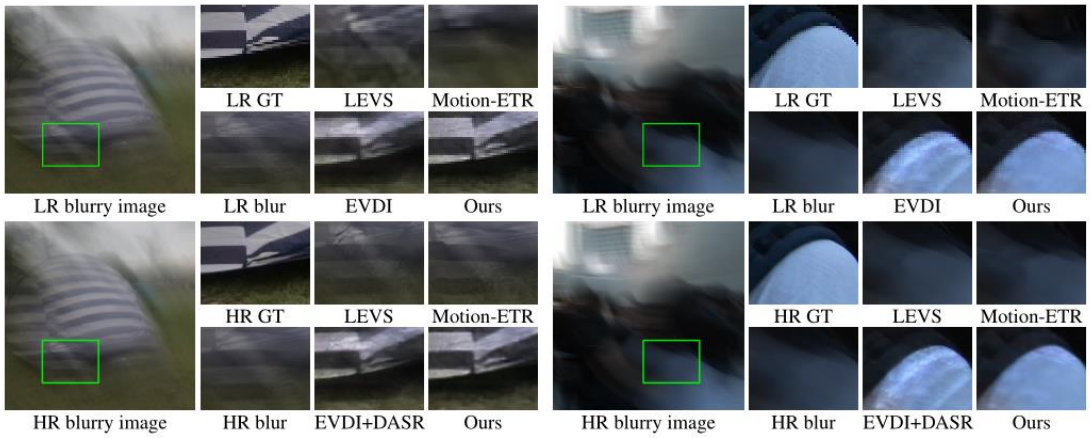
基于帧的算法 LEVS、Motion-ETR 和基于事件的方法 EDI、eSL-net、RED 和 EVDI。假设在没有真实图像的真实场景。

3.3 结果对比

3.3.1 不同空间尺度上对比



图二 在 MS-RBD 数据集上对真实 HR 框架和 LR 事件定性比较

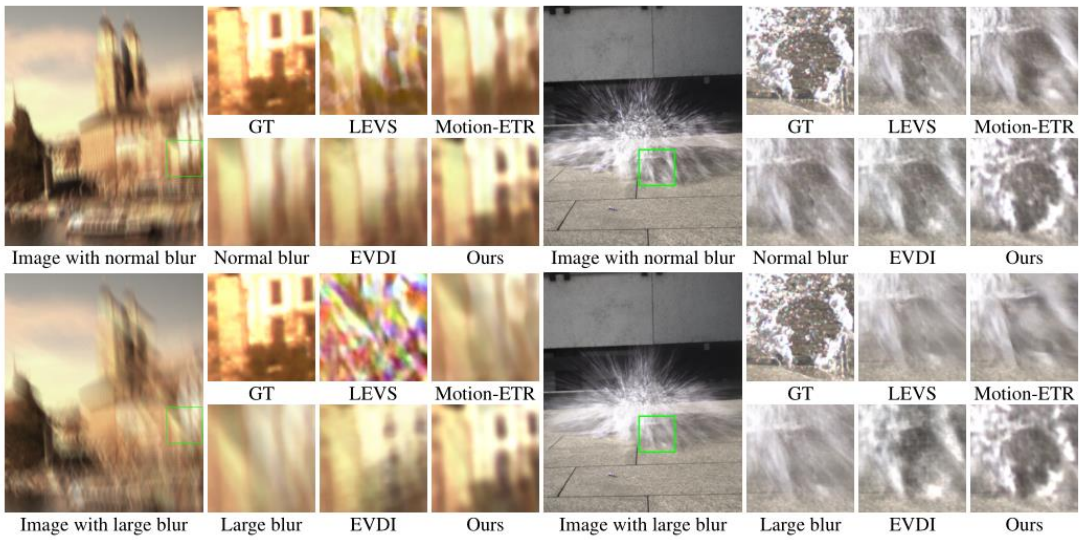


图三 在 Ev-REDS 数据集上不同空间尺度上的定性比较

由图二可以看出基于事件的算法由于只接收相同分辨率的模糊帧和事件导致信息丢失，限制性能。因此需要超分辨率技术来恢复 HR 结果，由图三这种级联会导致去模糊及超分辨率误差传播到后续阶段。

空间尺度 $R(B_T, \varepsilon_T)$为帧对事件的空间分辨率，例如 $R(B_T,\varepsilon_T)=4$ 表示帧 B_T 的分辨率是事件 ε_T 的 4 倍。

3.3.2 相同空间尺度不同时间尺度上对比



图四 HS-ERGB 数据集上，正常模糊和大模糊的定性比较

对于不同时间尺度的情况，以往方法受到训练数据模糊分布的限制，遇到较大模糊时，性能下降明显。而我们的方法由于时间泛化技术可以在正常和大模糊情况下恢复目标场景的可靠潜在图像。

综上，我们的方法不仅能够灵活设置输入空间分辨率，而且在处理不同时间尺度的运动模糊方面表现出良好的性能，有利于在现实场景中的应用。

四、总结

总体来说，作者提出了一个尺度感知网络，从而允许灵活设置输入空间分辨率，且能从不同时间尺度中学习。设置了两阶段自监督学习框架，在时间和空间维度上推广去模糊性能。发布了包含高分辨率模糊帧和低分辨率事件的真实世界数据集以促进在真实场景中评估去模糊性能。