EvTurb: 事件相机引导的湍流去除

Yixing Liu[1,2†] Minggui Teng[1,2†] Yifei Xia[1,2] Peiqi Duan[1,2] Boxin Shi[1,2*]

[1] State Key Laboratory of Multimedia Information Processing, School of Computer Science, Peking University
[2] National Engineering Research Center of Visual Technology, School of Computer Science, Peking University
luiginixy@stu.pku.edu.cn { minggui_teng, yfxia, duanqi0001, shiboxin } @pku.edu.cn

Abstract

大气湍流通过引入模糊和几何倾斜失真降低了图像质量，对后续的计算机视觉任务提出了重大挑战。现有的单图像和多帧方法由于湍流引发的失真组成复杂性而在解决这个高度病态的问题上举步维艰。为了解决这一问题，我们提出了 EvTurb，一个事件引导的湍流去除框架，利用高速事件流来解耦模糊和倾斜效应。EvTurb 通过建模基于事件的湍流形成来解耦模糊和倾斜效应，具体而言，这通过一个新颖的两步事件引导网络实现：首先利用事件积分来减少粗略输出的模糊。接着，采用从原始事件流中得出的方差图来消除精细输出的倾斜失真。此外，我们介绍了 TurbEvent，这是第一个具有多样湍流场景的真实捕获数据集。实验结果表明，EvTurb 在保持计算效率的同时，超越了最先进的方法。

1. 介绍

大气湍流指的是小尺度的不规则空气运动，通常由热量引起，通过光路中折射率的空间变化来降低图像质量。这些变化会引入复杂的模糊和几何倾斜效果（Figure 1 的顶部），在固定曝光时间内累积。虽然最近提出了一些方法 [15, 36] 用于单张图像的大气湍流去除，但由于该问题的极度病态性质，它们的效果仍然有限（Figure 1 (d)）。

为了解决这一挑战，多帧方法 [40, 41] 利用图像序列中的时间信息，有助于缓解问题的不适定性质（e.g.，利用幸运帧现象 [7]）。虽然这些方法通常比单图像方法表现更好 [15, 36]，但基于帧的方法需要大量数据，并且由于物体运动而更加复杂。计算摄影方法 [24, 32] 需要额外的硬件以实现较少湍流的输出。例如，引入窄带滤波器 [32] 通过与传统 RGB 成像相结合来减轻湍流。然而，由于传统相机帧速率的限制，倾斜和模糊效果只能部分减少，并在拍摄的图像中仍然存在。

事件相机 [22]，一种神经形态传感器，当像素辐射变化超过预设阈值时，会生成连续的事件流。这种能力

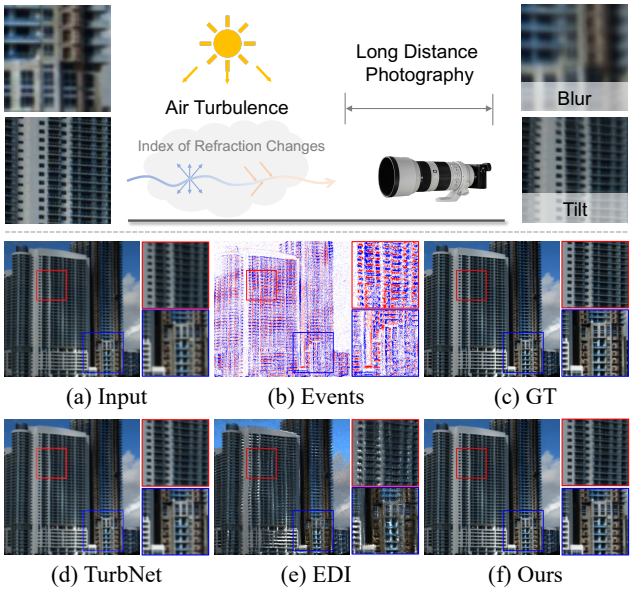


Figure 1. 由于大气湍流的影响，远距离摄影捕捉的湍流图像往往会出现模糊和几何畸变（上图）。给定一幅湍流图像 (a) 及其曝光期间记录的对应事件 (b)，我们的目标是恢复一幅清晰的图像 (c)。单图像方法 TurbNet [15] 未能完全校正畸变 (d)，而基于事件的方法 EDI [18] 则引入了伪影 (e)。相比之下，我们的 EvTurb 方法产生了更加清晰且畸变较少的图像 (f)。

使其能够检测微秒级精度的变化，支持从事件流生成高速视频等应用 [21, 27–29]。这一特性进一步促进了动态场景的高速观测，这促使我们探索：事件相机能否捕捉湍流场的变化，并通过高速事件信号有效地对模糊和倾斜效应进行单独建模？

在本文中，我们提出了 EvTurb 框架，这是一种事件相机引导的湍流去除方法，旨在通过利用高速事件流减轻单幅图像中的大气湍流效应。事件相机捕捉湍流场的变化，在曝光时间内记录这些变化的强度和频率。但是，现有的基于事件的去模糊模型仅考虑模糊效果，无法直接应用于具有结合倾斜效果的这种场景，还会引入伪影（Figure 1 (e)）。为了解决这个问题，我们制定了一个基于事件的图像湍流形成模型，其中模糊

† Contributed equally to this work as first authors
* Corresponding author
* Project page: https://github.com/ChipsAhoyM/EvTurb

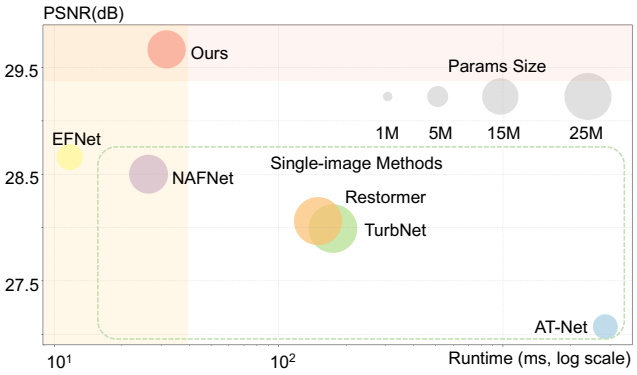


Figure 2. PSNR、运行时间和模型大小的比较。PSNR 是在 TurbEvent 数据集上评估的，运行时间是在 NVIDIA RTX 3090 GPU 上测量的。通过对基于事件的湍流进行物理建模并利用高速信息，我们的方法在具有竞争力的运行时间和紧凑的模型大小下实现了最高的 PSNR。

算子来源于事件记录的强度变化，而倾斜算子则由事件触发频率来表征。事件提供了湍流场变化的高保真度表示，允许模糊和倾斜效应的分开处理，从而减少问题的不确定性。基于这个模型，我们提出了一种两步事件引导的湍流去除网络，利用事件的时间信息。首先，代表湍流场变化强度的事件积分与 RGB 图像融合，生成主要缓解模糊的粗略结果。接下来，从原始事件流计算出的方差地图捕捉这些变化的频率，并与粗略结果集成以去除倾斜失真，产生精致的最终输出（Figure 1 (f)）。总体来说，我们的贡献包括：

- 通过解耦模糊和倾斜效应，在高速事件数据与湍流扭曲图像之间建立物理连接；
- 提出一个两步框架，该框架整合了来自事件数据和 RGB 图像的湍流场变化，以分别解决模糊和倾斜失真问题；并
- 提供首个实际捕捉的 TurbEvent 基准数据集，具有多样的湍流模式用于全面评估。

在我们新收集的 TurbEvent 数据集上的定量和定性结果表明，我们的方法在恢复无湍流的图像方面超过了现有的技术，减少了失真。如 Figure 2 所示，我们的方法在保持较短的运行时间的同时，PSNR 提高了超过 1 dB。

2. 相关工作

对于湍流缓解的研究，长久以来依赖传统方法，例如对于具有有限视野的图像进行去卷积，并假设空间不变的湍流 [19]，以及在处理空间变化的湍流时选择幸运帧 [7, 30]。最近，深度学习方法已经出现，可以分为单帧和多帧方法。单帧策略采用包括卷积神经网络 (CNN) [37]、Transformers [15]、生成对抗网络 (GANs) [11, 12, 16, 20]、以及用于湍流校正的去噪扩散概率模型 [17]，同时采用将 RGB 与窄带图像相结合的方法 [32]。然而，这些技术常常忽视关键的时间信息。相比之下，多帧方法有效利用时间数据，使用基于 transformer 的架构 [34, 41] 或循环模型 [40]。一些方法使用特殊技巧来促进这个过程，例如统计先验 [34]、神经表示 [2, 10] 或在特殊领域 (如 e.g.，行星图像 [33])。尽管效果显著，它们需要大量的 10 到 12 湍流帧数据，并面临帧间湍流变化和物体运动相关的挑战。

基于事件的图像去模糊利用运动事件和模糊图像之间的内在联系来重建清晰的细节。该领域的一个基本模型是基于事件的双积分（EDI）模型 [18]，其在对数域中表示事件流和清晰/模糊图像。为了提高对抗噪声的鲁棒性和增强细节，后续引入了基于 CNN 的方法 [3, 44]。最近，transformer 架构 [25, 26] 已被利用，使注意力机制可以更好地弥合事件数据和图像之间的领域差距，从而实现更精确的重建。此外，已经探索了事件表示 [27]，以更好地融合图像和事件模态。为了解决图像和事件数据领域差距的挑战，采用了跨模态校准策略来解决空间和时间对齐问题 [35]。

与传统相机不同，事件相机直接记录亮度变化，使其非常适合捕捉湍流变化。虽然 Boehrer et al. [1] 使用事件来重建高速视频并应用基于视频的去湍流方法，但它们没有明确建模事件与湍流图像之间的关系。因此，我们在像素级将事件流中的运动线索整合到去湍流框架中，以更有效地利用连续事件数据中的丰富时间信息进行去湍流。

3. 方法

本节中，我们首先在 Section 3.1 中介绍事件摄像头和基于事件的湍流形成的概念。接下来，我们在 Section 3.2 中展示一种两步策略，通过事件流中编码的湍流信息引导重建无湍流图像。然后我们在 Section 3.3 中介绍我们的 EvTurb 框架。在 Section 3.4 中描述了 TurbEvent 数据集的策展过程，接下来是在 Section 3.5 中的实现细节。

3.1. 公式

通过事件生成清晰图像。 事件相机 [22] 旨在检测对数域中的辐射变化。当强度变化超过预设阈值 c 时，会生成一个事件信号，表示为四元组，记作 (x, y, t, p)、i.e.、

$$\log(\boldsymbol{I}^{t}_{x,y}) - \log(\boldsymbol{I}^{t_0}_{x,y}) = p \cdot c, \tag{1}$$

，其中 $\boldsymbol{I}^{t}_{x,y}$ 和 $\boldsymbol{I}^{t_0}_{x,y}$ 分别表示 (x, y) 在时刻 t 和 t_0 的像素强度值。极性 $p \in \{1, -1\}$ 表示强度是增加还是减少。由于 Equation (1) 是独立应用于每个像素，因此像素索引在此省略。

给定两个连续的无湍流清晰图像 $\boldsymbol{I}^0$ 和 $\boldsymbol{I}^1$，以及在它们之间触发的相应的 N_e 事件 $\{e_k\}^{N_e}$，我们可以使用事件积分将这两个帧连接起来：

$$\boldsymbol{I}^1 = \boldsymbol{I}^0 \cdot \exp(\int_{t_0}^{t_1} c_k \cdot p_k \, \mathrm{d}k), \tag{2}$$

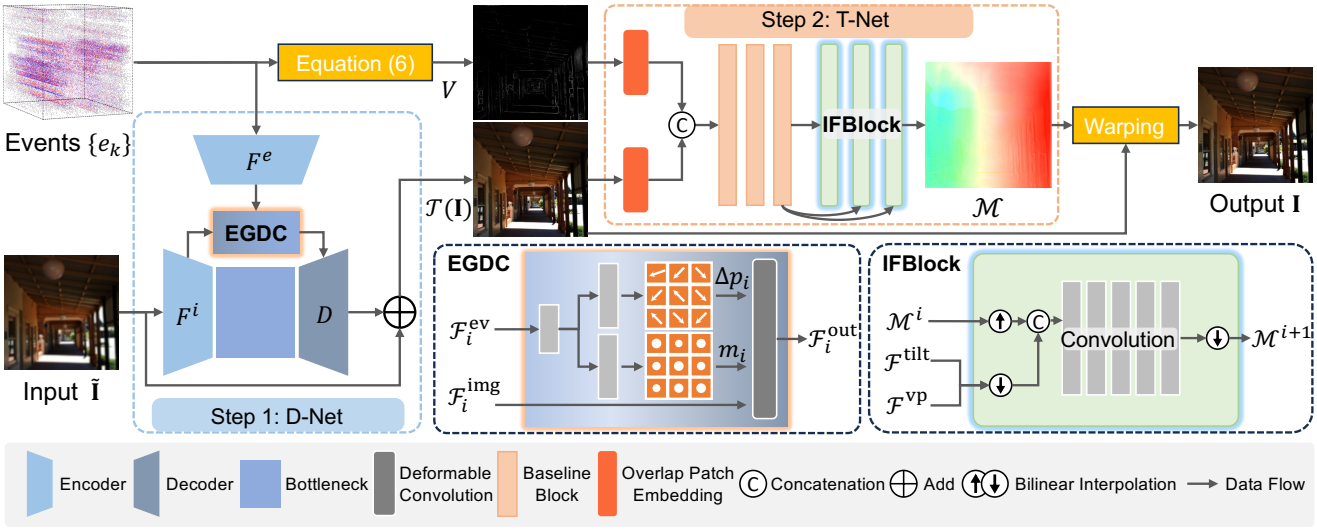


Figure 3. 我们的 EvTurb 方法的整体框架。我们首先使用一个多尺度的架构 D-Net 进行去模糊，它从两种模态估计事件积分：RGB 图像 $\tilde{\boldsymbol{I}}$ 和事件流 $\{e_k\}$ 。在解码过程中应用事件引导的可变形卷积（EGDC）块以增强数据融合。在获得无模糊图像 $\mathcal{T}(\boldsymbol{I})$ 后，T-Net 在计算的方差图 $\boldsymbol{V}$ 的引导下预测倾斜流 $\mathcal{M}$ 。然后使用 IFBlock 在较低分辨率下迭代地细化流。最终，估计的流 $\mathcal{M}$ 被用来扭曲 $\mathcal{T}(\boldsymbol{I})$ ，生成输出 $\boldsymbol{I}$ 。

，其中 c_k 表示空间-时间变体阈值，这依赖于场景条件 [8] 。

一个被称作 $\tilde{\boldsymbol{I}}$ 的湍流图像可以通过经典的分步波传播方程生成。为简单起见，此过程可以实现为一个向前方程 [15] ，将湍流畸变近似为两个退化算子的组合：

$$\tilde{\boldsymbol{I}} = \mathcal{B} \circ \mathcal{T}(\boldsymbol{I}) + \boldsymbol{N}, \tag{3}$$

，其中 $\mathcal{B}$ 表示空间变化模糊，$\mathcal{T}$ 表示几何像素位移（通常称为倾斜）。Figure 1 中展示了一个例子。$\boldsymbol{N}$ 是附加噪声信号，$\circ$ 表示功能组合。如果只有单个湍流图像可用，则模糊和倾斜操作是分不开的，因为这些畸变在曝光时间内是结合在一起的。

如 Equation (2) 所示，潜在图像可以有效地与事件流联系起来。在大气湍流的情境下，由于空气折射主要引发倾斜失真，高速潜在图像尽管不那么模糊，但倾斜失真。进一步，模糊效果可以主要视为潜在帧的平均值。通过将 Equation (2) 与 Equation (3) 结合，我们推导出以下方程：

$$\begin{aligned}\tilde{\boldsymbol{I}} &\approx \mathcal{T}(\boldsymbol{I}) \cdot \frac{1}{T} \int_T \left(\exp\left(\int_t c_k \cdot p_k \, \mathrm{d}k \right) \right) \mathrm{d}t + \boldsymbol{N} \\ &= \mathcal{T}(\boldsymbol{I}) \cdot \boldsymbol{E} + \boldsymbol{N}. \end{aligned} \tag{4}$$

模糊算子替换为 $\boldsymbol{E}$ ，它表示在曝光期间 T 发生的事件的双积分。需要注意的是，我们的模糊算子与标准运动模糊不同。利用高速事件相机捕获散焦内核的能力 [14] ，我们的模糊算子整合了局部散焦效应，这可以通过基于事件的引导进一步优化。

3.2. 基于事件的湍流消除

去除模糊。 形成模型 Equation (4) 表明湍流失真是构成性的，包括倾斜和模糊。通过利用事件流中捕获的高速湍流场信息，可以将模糊算子公式化为事件的双重积分，并随即从湍流失真的图像中取消掉。认知到噪声的存在，我们进一步将 Equation (4) 转化为一个最小二乘优化问题：

$$\mathcal{T}(\boldsymbol{I})^* = \underset{\mathcal{T}(\boldsymbol{I})}{\operatorname{argmin}} \| \tilde{\boldsymbol{I}} - \mathcal{T}(\boldsymbol{I}) \cdot \boldsymbol{E} \|^2. \tag{5}$$

解决这个目标便于初步恢复那些虽然较少模糊但显示倾斜失真的图像 $\mathcal{T}(\boldsymbol{I})^*$ 。

由于事件积分也受到倾斜畸变的影响，因此有必要探索事件流中的时间信息，以进一步增强倾斜去除。受之前工作启发，我们认识到方差图可以指示与倾斜强度相关的像素不确定性。我们从原始事件流中构建方差图，捕捉湍流场中的变化频率，并指示倾斜畸变的不确定性。由于事件记录了潜在帧的强度值变化，我们将整个潜在帧序列的方差定义为方差图。形式上，我们按如下方式计算方差图：其中表示归一化操作，表示累积事件。

利用方差图 $\boldsymbol{V}$ 作为指导，可以在最大后验估计下制定倾斜流 $\mathcal{M}$ ：

$$\mathcal{M}^* = \underset{\mathcal{M}}{\operatorname{argmax}} \, \mathbb{P}(\mathcal{M} \circ \mathcal{T}(\boldsymbol{I})^* \mid \boldsymbol{V}), \tag{6}$$

，其中 $\mathbb{P}(\cdot)$ 表示没有倾斜效应的自然图像分布，$\circ$ 表示图像扭曲函数。通过使用自然图像先验来估计最佳倾斜流 $\mathcal{M}^*$ ，我们可以有效地将其应用于初始图像以进行倾斜校正。

3.3. EvTurb 框架

D-Net。 优化 Equation (5) 的关键组件是获得准确的事件积分。由于事件相机的阈值表现出时空变化 [8]，直接集成原始事件数据会在恢复的图像中引入伪影。相反，我们建议以数据驱动的方式估计这一积分，将 Equation (5) 转化为回归问题，i.e. :

$$\mathcal{T}(\boldsymbol{I}) = D(F^i(\tilde{\boldsymbol{I}}), F^e(\{e_k\})), \tag{7}$$

其中 F^i 和 F^e 分别是用于编码图像和事件的隐函数，而 D 是用于解码和融合两种模态信息的隐函数。

为了增强从事件数据和 RGB 图像中提取特征，我们提出了一种双分支网络 D-Net，专门设计用于有效融合这两种模态。先前的研究 [3, 44] 已验证 U-Net 架构在恢复清晰图像细节方面非常有效，因为它能在各种尺度上提取特征。因此，我们采用多尺度 U-Net 结构，以在不同尺度上融合两种模态，并在两个编码器中使用密集块以有效地重用提取的特征。

为了进一步改善局部特征的恢复，我们引入了一个事件引导的可变形卷积（EGDC）模块，将可变形卷积引入跳跃连接中。事件数据的极性作为偏移估计的重要指标，而累积的强度则突出显示了重要区域，有效地作为一个掩码。具体而言，在某些尺度下的事件特征用于计算采样点的偏移和掩码图，如下所示：

$$\Delta p_i = C_p(B(\mathcal{F}_i^{\text{ev}})), \quad m_i = C_m(B(\mathcal{F}_i^{\text{ev}})), \tag{8}$$

，其中，Δp_i 、m_i 和 $\mathcal{F}_i^{\text{ev}}$ 分别表示偏移图、掩码图和第 i 个事件特征。这里，B 表示瓶颈卷积，而 C_p 和 C_m 表示两个独立的卷积层。使用这些计算出的偏移和掩码，可以计算可变形卷积为：

$$\mathcal{F}_i^{\text{out}} = \text{Deform}(\mathcal{F}_i^{\text{img}}, \Delta p_i, m_i). \tag{9}$$

，其中 Deform 代表可变形卷积 [5] 。通过采用这种事件引导的可变形卷积，我们有效地解决了显著局部特征的问题，并在特征空间内对齐梯度。

在估计事件积分后，我们通过全局连接将其与湍流图像融合，有效地产生去模糊的输出。具体来说，全局连接使模型能够利用事件估计与湍流图像特征之间的长程依赖关系，从而有效地减轻模糊。

尽管从 D-Net 获得了粗糙的结果，倾斜效应仍然存在。因此，在 Equation (5) 中用于重建图像的一个重要步骤是拟合无倾斜的自然图像先验。受先前解决空间变化位移工作 [9, 45] 的启发，我们提出了一种基于流的网络，T-Net，用于估计倾斜流 $\mathcal{M}$ ，该流将失真图像扭曲为无倾斜图像。并且 Equation (6) 可以被表述为：

$$\mathcal{M} = f_{\text{VT}}(\mathcal{T}(\boldsymbol{I}), \boldsymbol{V}), \tag{10}$$

，其中 f_{VT} 是由 T-Net 表示的隐函数，专为倾斜流估计设计，$\boldsymbol{V}$ 是从 ?? 导出的计算方差图，代表了湍流引入的不确定性。

在设计 T-Net 时，我们采用了重叠补丁嵌入，这种方法擅长提取上下文和边界信息。这种方法确保了补丁之间的平滑过渡并减少了最终输出中的伪影。随后，我们采用一系列卷积和基线块 [4] 来有效地融合来自两种模态的特征。下一步是从融合后的潜在特征中确定一个倾斜流，该流可以准确地将失真的图像向真实情况进行扭曲。为此，我们采用 IFBlocks [9] 进行流估计，迭代地改进并接近正确的倾斜流。形式上，这种迭代改进可以表示为

$$\mathcal{M}^{i+1} = \text{IF}(\mathcal{F}^{\text{tilt}}, \mathcal{F}^{\text{vp}}, \mathcal{M}^i), \tag{11}$$

其中 $\mathcal{F}^{\text{tilt}}$ 和 $\mathcal{F}^{\text{vp}}$ 分别表示从去模糊粗略图像和方差图中提取的特征，而 IF 代表 IFBlock。如 Figure 3 所示，我们在较低分辨率级别计算倾斜流，以确保倾斜流的平滑性。通过将最终估计的流应用于从 D-Net 获得的粗略结果，我们可以重建无湍流的图像。

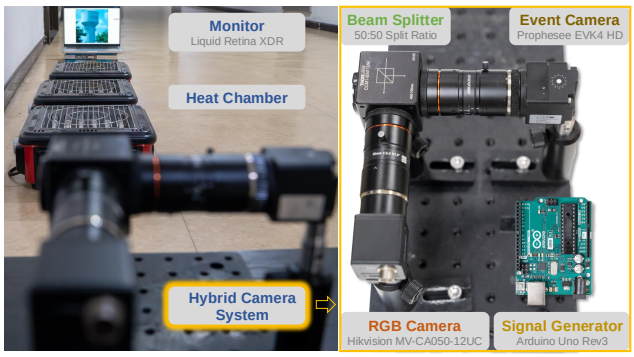


Figure 4. 我们构建了一个混合相机系统（a），以研究事件形成与图像湍流之间的关系。此外，我们实施了一个带有热室的显示-相机系统（b），用于 Turbevent 数据集的组织。

3.4. TurbEvent 数据集

由于缺乏封闭形式的公式化，将湍流模拟为连续过程具有挑战性。这种困难也体现在为模拟事件流生成高帧率的视频上。受 EventNFS 数据集启发 [6] ，我们设计了一个显示-相机系统，以捕获湍流图像、配对的无湍流图像以及对应的事件流。

我们的设置包括一个 Liquid Retina XDR 显示器（分辨率：3024×1964 ，120 Hz）和一个混合摄像系统，如 Figure 4 所示。该混合系统由两台摄像机组成：一台机器视觉摄像机（HIKVISION MV-CA050-12UC）和一台事件摄像机（PROPHESEE GEN4.0），两者都配备了 50 毫米 F/ 1.8 镜头，并通过分光器对齐。为了产生空气湍流，我们使用多个热室，将混合摄像系统置于距离显示器大约五米处。空间校准使用棋盘格进行，而时间同步通过信号发生器实现。

我们选择了来自 ADE20K 数据集的图像 [42, 43] 。该数据集涵盖多个类别，包括户外和室内场景，并提供了用于稳健分析的细粒度细节。视频以 10 FPS 播放，以避免掉帧并最大程度地减少显示刷新带来的影响。为了确保数据的完整性，录制过程中尽量减少外部光源的影响。最终，我们成功获得了 6318 对湍流和无湍流图像及其对应的事件流，具有 592×592 分辨率。

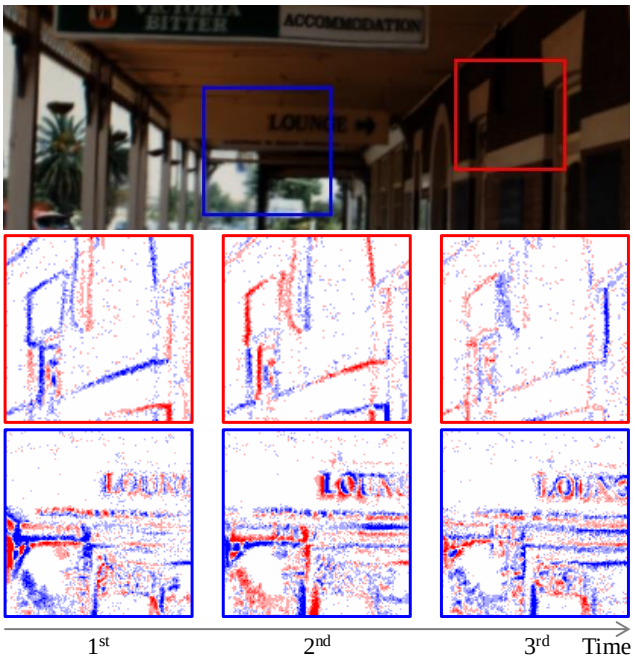


Figure 5. 上方是我们 TurbEvent 湍流视频帧的可视化，下方是该湍流视频的三段重复事件录制。这三段事件片段中的湍流模式各不相同，表现出随机性。

我们将这个基于事件的湍流数据集称为“TurbEvent”数据集。[1]

在 Figure 5 中展示了一个例子，描绘了来自 ADE20K 数据集 [42, 43] 的一帧及其相应的事件补丁，这些补丁是使用我们的混合相机系统在连续的时间戳下捕获的。我们在相同的场景下连续记录了三个事件流。尽管捕获条件固定，三个事件补丁由于湍流的随机性而表现出不同的外观。一些边缘信号丢失，而其他事件在三次捕获中由于湍流模式的变化和随机性在同一位置表现出极性反转。通过捕获大量数据，我们能够在 TurbEvent 数据集中收集多样的湍流模式。这种内在的随机性使得从高速湍流视频模拟中生成逼真的模拟事件变得具有挑战性，加强了我们捕获真实世界数据集的决策。

3.5. 实现细节

损失函数。 我们将均方误差（MSE）损失和感知损失 [39] 的加权组合用于 D-Net 和 T-Net 的训练：

$$\mathcal{L} = \lambda_1 \mathcal{L}_2(I_G, I_O) + \lambda_2 \mathcal{L}_{\text{perc}}(I_G, I_O), \qquad (12)$$

其中，I_G 和 I_O 分别表示真实图像和生成的图像。我们经实验设置 $\lambda_1 = 1.0$ 和 $\lambda_2 = 0.02$,并使用 VGG16 [23] 作为感知损失的预训练模型。

我们使用 PyTorch 框架实现了我们提出的方法，并在 NVIDIA RTX 3090 GPU 上进行所有实验。通过同时训练 D-Net 和 T-Net，共训练 150 个周期，我们采用端到端策略。我们采用 Adam 优化器，初始学习率为 1×10^{-4} ，并应用余弦退火学习率调度器来逐步调整学习率。为了提高鲁棒性和泛化能力，我们在数据增强中加入了随机裁剪。

[1]数据集配置可以在补充材料中找到。

Table 1. 在 TurbEvent 数据集上的定量比较。“事件”列指定方法是否需要事件（是 [✓] 或否 [✗]）。↑（↓）表示越高（越低）越好。

	Event	PSNR ↑	SSIM ↑	LPIPS ↓
ATNet	✗	27.07	0.7491	0.3416
TurbNet	✗	27.99	0.7872	0.3418
N-DIR	✗	23.19	0.6941	0.4560
NAFNet	✗	28.50	0.8190	0.3218
Restormer	✗	28.06	0.7995	0.2163
Restormer *	✓	28.63	0.8114	0.2122
EFNet	✓	28.66	0.8095	0.2140
Ours	✓	29.67	0.8405	0.2010

4. 实验

对于 TurbEvent 数据集中训练和测试数据的划分，我们首先使用图像标签将数据分类为多个组。在每个组内，我们随机将数据按 9 : 1 比例分为训练和测试子集。该策略确保训练和测试数据集之间具有相似的分布。最终，我们得到 5665 对用于训练和 653 对用于测试。

我们将我们的方法与三种基于图像的湍流去除方法、两种通用图像恢复方法和一种基于事件的去模糊方法进行比较：AT-Net [36] ，TurbNet [15] ，N-DIR [13] ，NAFNet [4] ，Restormer [38] ，以及 EFNet [25] 。此外，对于通用图像恢复方法 Restormer，我们也将事件数据与图像帧一起输入到网络中，记作 Restormer * 。所有监督的方法 [4, 15, 25, 36, 38] 都在我们的 TurbEvent 数据集上重新训练，以进行公平比较。恢复图像的质量通过三个常用的指标进行评估：PSNR、SSIM 和 LPIPS。

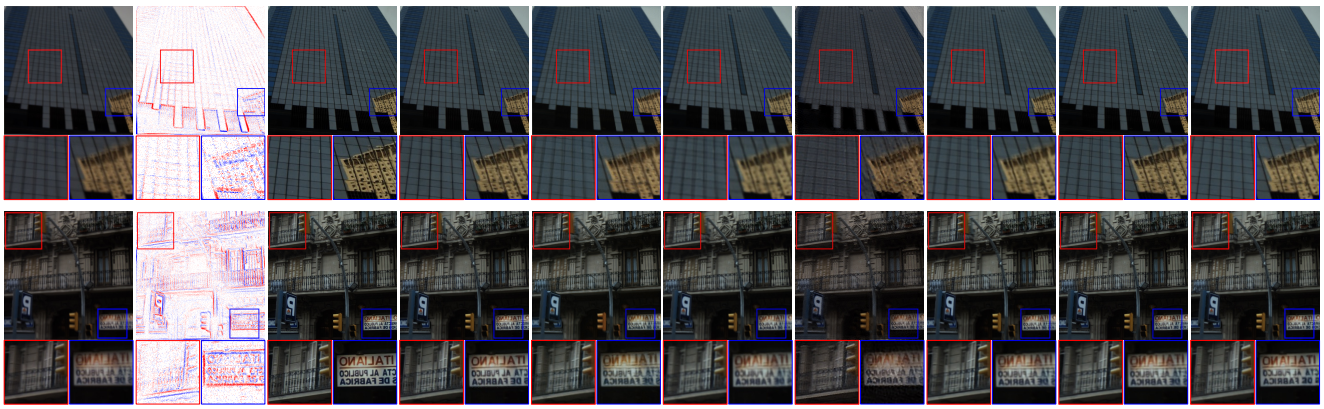
定量结果展示在 Table 1 中，定性比较展示在 Figure 6 中。正如所示，我们的方法在所有三个指标中始终获得最高的性能。在 PSNR 方面，事件引导的方法显著超过所有其他基于图像的方法。

对于视觉质量，我们的方法相比于单图像技术，能够有效减少图像模糊。此外，我们的方法能够比基于事件的去模糊方法 EFNet 重建出更少伪影且细节更清晰的图像。对于 Restormer * ，简单的连接策略仅带来性能上的轻微提升。值得注意的是，如 Figure 6 的第一行所示，只有我们的方法有效地解决了像素位移问题。

4.1. 对实际采集数据的评估

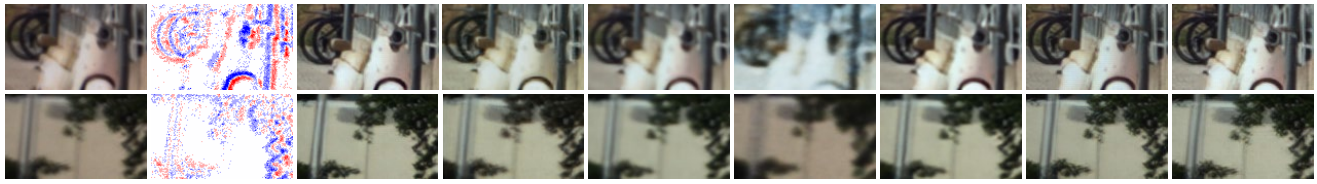
我们使用一个混合摄像系统 [2] 拍摄了真实世界的湍流图像，结果如 Figure 7 所示。正如所示，我们的方法

[2]详细配置可以在补充材料中找到。



(a) Inputs (b) Events (c) GT (d) Ours (e) AT-Net (f) TurbNet (g) N-DIR (h) NAFNet (i) Restormer (j) EFNet

Figure 6. 在 TurbEvent 数据集上的视觉质量比较。(a) 输入图像。(b) 事件。(c) 真值。(d) ∼ (j) 我们的方法的结果，AT-Net [36] ，TurbNet [15] ，N-DIR [13] ，NAFNet [4] ，Restormer [38] ，和 EFNet [25] 。每张图片下方提供了特写视图。更多结果见补充材料。



(a) Inputs (b) Events (c) Ours (d) AT-Net (e) TurbNet (f) N-DIR (g) NAFNet (h) Restormer (i) EFNet

Figure 7. 实拍数据的视觉质量比较。(a) 输入图像。(b) 事件。(c) ∼ (i) 我们的结果、AT-Net [36] 、TurbNet [15] 、N-DIR [13] 、NAFNet [4] 、Restormer [38] 和 EFNet [25] 。附加结果在补充材料中提供。

(a) Inputs (b) Deblur (c) Deblur + Detilt

Figure 8. 展示了 D-Net 和 T-Net 有效性的一例。给定一个湍流图像（a），D-Net 有效去除了模糊（b）。当与 T-Net 结合时，我们的方法可以纠正倾斜失真（c）。

在真实世界数据上始终优于其他最先进的方法，能够恢复更多细节，且展现出卓越的泛化能力。

为了展示专为去模糊和去倾斜设计的 D-Net 和 T-Net 的有效性，我们展示了仅使用 D-Net 获得的结果。如 Figure 8 所示，D-Net 有效去除了模糊（绿框）。当与 T-Net 结合时，我们的方法解决了倾斜问题（红框）。D-Net 利用具有全局连接的残差学习，这一技术在图像去模糊任务中已被广泛应用。相反，T-Net 估计了一个倾斜流场，以准确建模几何像素偏移。

为了全面评估复原图像的质量，我们评估 EvTurb 在基于图像的语义分割中的性能提升。原始的 ADE20K 数据集 [42, 43] 为语义分割提供了真实标签。我们选择 InternImage-XL [31] 作为基准模型，该模型利用可变形卷积实现自适应空间聚合和有效感受野。

Table 2. 对来自我们的 TurbEvent 数据集的复原图像进行语义分割任务的定量比较。

	Event	mIoU ↑	Dice ↑	FWIoU ↑
Turb		0.923	0.959	0.931
Ground truth		0.941	0.969	0.948
AT-Net	✗	0.922	0.958	0.932
TurbNet	✗	0.926	0.961	0.936
N-DIR	✗	0.910	0.951	0.921
NAFNet	✗	0.923	0.959	0.936
Restormer	✗	0.928	0.961	0.937
EFNet	✓	0.930	0.963	0.939
Ours	✓	0.936	0.966	0.943

三种常用的评估指标是：平均交并比（mIoU）、Dice 系数和频率加权交并比（FWIoU）。在 Table 2 中展示的定量结果表明，与其他方法相比，我们的方法取得了最高的改进，接近于在真值图像上的表现。在 Figure 9 中的视觉比较显示，我们的方法能够实现更准确的图像分割。[3]

[3]请注意，我们的训练和测试划分与原始的 ADE20K 不同，因此绝对值不能直接与原始数据集的结果进行比较。但在复原和湍流图像之间的差异显示了我们的方法提升了下游性能。

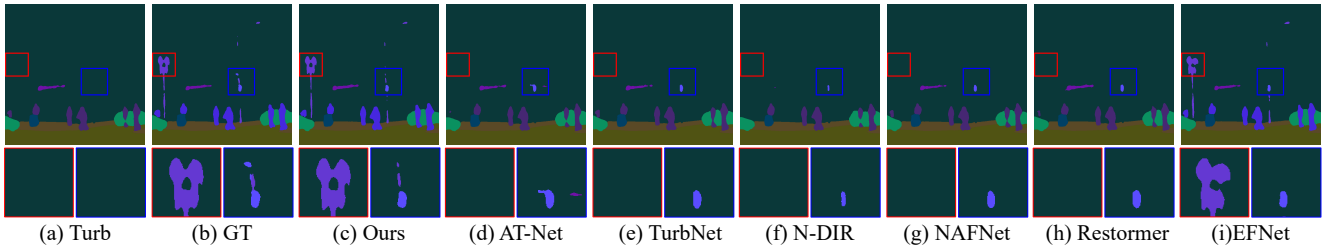


Figure 9. 涡流事件数据集上语义分割的视觉质量比较。(a) 和 (b) 分别展示了来自湍流图像和真实图像的分割结果。(c) ∼ (i) 展示了使用我们的方法、AT-Net、TurbNet、N-DIR、NAFNet、Restormer 和 EFNet 复原的图像的分割结果。每幅图像下方提供了放大视图。额外的结果提供在补充材料中。

我们进行了四项消融研究，以评估我们 EvTurb 方法中每个模块的贡献，如 ?? 所总结。具体来说，我们首先移除了 EGDC 块（表示为 “W/o EGDC”）以研究它们在两种模态之间数据融合中的作用。接下来，我们移除了 T-Net，以评估其在减轻倾斜失真方面的有效性（表示为 “W/o T-Net”）。然后，我们排除了 T-Net，以直接评估其在倾斜校正上的具体影响（表示为 “W/o Variance”）。最后，我们移除了基于流的架构以直接恢复图像，以确定其在精炼中的作用（表示为 “W/o Flow”）。结果表明，我们的完整模型实现了更好的整体性能。没有 EGDC 块的情况下，该模型在多尺度图像事件数据集成方面表现挣扎。没有 T-Net 模块，结果表现出显著的倾斜退化。

在本文中，我们提出了 EvTurb，这是一种利用高速事件流去除大气湍流的新框架。通过基于事件记录的强度变化建模模糊算子，以及通过事件触发频率建模倾斜算子，我们有效地解耦了模糊和倾斜失真。我们进一步通过一个结构化的两步框架进行湍流去除。在我们真实捕获的 TurbEvent 数据集上的评估表明，我们的方法显著优于其他方法，同时保持了竞争性的运行效率。

局限性。 我们方法的一个限制在于事件相机和帧相机之间的潜在误差。尽管我们采用精确校准和光束分离光学设置来缓解这个问题，但仍可能存在轻微的误差，可能影响数据融合的准确性。此外，我们的实现假设两台相机具有相同的分辨率，这简化了处理，但忽略了在现实应用中常见的跨分辨率差异。此外，当事件数据与 RGB 图像融合时，事件数据的单色特性引入了领域差异。

References

[1] Nicolas Boehrer, Robert PJ Nieuwenhuizen, and Judith Dijk. Turbulence mitigation in imagery including moving objects from a static event camera. Optical Engineering, 60(5):053101–053101, 2021.

[2] Haoming Cai, Jingxi Chen, Brandon Feng, Weiyun Jiang, Mingyang Xie, Kevin Zhang, Cornelia Fermuller, Yiannis Aloimonos, Ashok Veeraraghavan, and Chris Metzler. Temporally consistent atmospheric turbulence mitigation with neural representations. In Adv. of Neural Information Processing Systems, 2024.

[3] Haoyu Chen, Minggui Teng, Boxin Shi, Yizhou Wang, and Tiejun Huang. A residual learning approach to deblur and generate high frame rate video with an event camera. IEEE Transactions on Multimedia, 25:5826–5839, 2023.

[4] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. Proc. of European Conference on Computer Vision, 2022.

[5] Jifeng Dai, Haozhi Qi, Yuwen Xiong, Yi Li, Guodong Zhang, Han Hu, and Yichen Wei. Deformable convolutional networks. In Proc. of International Conference on Computer Vision, 2017.

[6] Peiqi Duan, Zihao W. Wang, Xinyu Zhou, Yi Ma, and Boxin Shi. Eventzoom: Learning to denoise and super resolve neuromorphic events. In Proc. of Computer Vision and Pattern Recognition, 2021.

[7] David L Fried. Probability of getting a lucky short-exposure image through turbulence. Journal of the Optical Society of America, 68(12):1651–1658, 1978.

[8] Yuhuang Hu, Shih-Chii Liu, and Tobi Delbruck. v2e: From video frames to realistic dvs events. In Proc. of Computer Vision and Pattern Recognition, 2021.

[9] Zhewei Huang, Tianyuan Zhang, Wen Heng, Boxin Shi, and Shuchang Zhou. Real-time intermediate flow estimation for video frame interpolation. In Proc. of European Conference on Computer Vision, 2022.

[10] Weiyun Jiang, Vivek Boominathan, and Ashok Veeraraghavan. Nert: Implicit neural representations for unsupervised atmospheric turbulence mitigation. In Proc. of Computer Vision and Pattern Recognition Workshops, 2023.

[11] Darui Jin, Ying Chen, Yi Lu, Junzhang Chen, Peng Wang, Zichao Liu, Sheng Guo, and Xiangzhi Bai. Neutralizing the impact of atmospheric turbulence on complex scene imaging via deep learning. Nature Machine Intelligence, 3(10):876–884, 2021.

[12] Chun Pong Lau, Carlos D. Castillo, and Rama Chellappa. AtfaceGAN: Single face semantic aware image restoration and recognition from atmospheric turbulence. IEEE Transactions on Biometrics, Behavior, and Identity Science, 3(2):240–251, 2021.

[13] Nianyi Li, Simron Thapa, Cameron Whyte, Albert W. Reed, Suren Jayasuriya, and Jinwei Ye. Unsupervised non-rigid image distortion removal via grid deformation. In Proc. of International Conference on Computer Vision, 2021.

[14] Hanyue Lou, Minggui Teng, Yixin Yang, and Boxin Shi. All-in-focus imaging from event focal stack. In Proc. of Computer Vision and Pattern Recognition, 2023.

[15] Zhiyuan Mao, Ajay Jaiswal, Zhangyang Wang, and Stanley H. Chan. Single frame atmospheric turbulence mitigation: A benchmark study and a new physics-inspired transformer model. In Proc. of European Conference on Computer Vision, 2022.

[16] Kangfu Mei and Vishal M. Patel. LTT-GAN: Looking through turbulence by inverting GANs. IEEE Journal of Selected Topics in Signal Processing, 17(3):587–598, 2023.

[17] Nithin Gopalakrishnan Nair, Kangfu Mei, and Vishal M Patel. AT-DDPM: Restoring faces degraded by atmospheric turbulence using denoising diffusion probabilistic models. In Proc. of Winter Conference on Applications of Computer Vision, 2023.

[18] Liyuan Pan, Cedric Scheerlinck, Xin Yu, Richard Hartley, Miaomiao Liu, and Yuchao Dai. Bringing a blurry frame alive at high frame-rate with an event camera. In Proc. of Computer Vision and Pattern Recognition, 2019.

[19] J. Primot, G. Rousset, and J. C. Fontanella. Deconvolution from wave-front sensing: a new technique for compensating turbulence-degraded images. Journal of the Optical Society of America, 7:1598–1608, 1990.

[20] Shyam Nandan Rai and C. V. Jawahar. Removing atmospheric turbulence via deep adversarial learning. IEEE Transactions on Image Processing, 31:2633–2646, 2022.

[21] Henri Rebecq, René Ranftl, Vladlen Koltun, and Davide Scaramuzza. High speed and high dynamic range video with an event camera. IEEE Transactions on Pattern Analysis and Machine Intelligence, 43(6): 1964–1980, 2019.

[22] Teresa Serrano-Gotarredona and Bernabé Linares-Barranco. A 128 × 128 1.5% contrast sensitivity 0.9% fpn 3 µs latency 4 mw asynchronous frame-free dynamic vision sensor using transimpedance preamplifiers. IEEE Journal of Solid-State Circuits, 48(3):827–838, 2013.

[23] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In Proc. of International Conference on Learning Representations, 2015.

[24] Michael J Steinbock. Implementation of branch-point-tolerant wavefront reconstructor for strong turbulence compensation. Master' s thesis (Air Force Institute of Technology), 2012.

[25] Lei Sun, Christos Sakaridis, Jingyun Liang, Qi Jiang, Kailun Yang, Peng Sun, Yaozu Ye, Kaiwei Wang, and Luc Van Gool. Event-based fusion for motion deblurring with cross-modal attention. In Proc. of European Conference on Computer Vision, 2022.

[26] Lei Sun, Christos Sakaridis, Jingyun Liang, Peng Sun, Jiezhang Cao, Kai Zhang, Qi Jiang, Kaiwei Wang, and Luc Van Gool. Event-based frame interpolation with ad-hoc deblurring. In Proc. of Computer Vision and Pattern Recognition, 2023.

[27] Minggui Teng, Chu Zhou, Hanyue Lou, and Boxin Shi. NEST: Neural event stack for event-based image enhancement. In Proc. of European Conference on Computer Vision, 2022.

[28] Stepan Tulyakov, Daniel Gehrig, Stamatios Georgoulis, Julius Erbach, Mathias Gehrig, Yuanyou Li, and Davide Scaramuzza. Time Lens: Event-based video frame interpolation. In Proc. of Computer Vision and Pattern Recognition, 2021.

[29] Stepan Tulyakov, Alfredo Bochicchio, Daniel Gehrig, Stamatios Georgoulis, Yuanyou Li, and Davide Scaramuzza. Time Lens++: Event-based frame interpolation with parametric non-linear flow and multi-scale fusion. In Proc. of Computer Vision and Pattern Recognition, 2022.

[30] Mikhail A. Vorontsov and Gary W. Carhart. Anisoplanatic imaging through turbulent media: image recovery by local information fusion from a set of short-exposure images. Journal of the Optical Society of America, 18(6):1312–1324, 2001.

[31] Wenhai Wang, Jifeng Dai, Zhe Chen, Zhenhang Huang, Zhiqi Li, Xizhou Zhu, Xiaowei Hu, Tong Lu, Lewei Lu, Hongsheng Li, et al. InternImage: Exploring large-scale vision foundation models with deformable convolutions. In Proc. of Computer Vision and Pattern Recognition, 2023.

[32] Yifei Xia, Chu Zhou, Chengxuan Zhu, Minggui Teng, Chao Xu, and Boxin Shi. NB-GTR: Narrow-band guided turbulence removal. In Proc. of Computer Vision and Pattern Recognition, 2024.

[33] Yifei Xia, Chu Zhou, Chengxuan Zhu, Chao Xu, and Boxin Shi. PlaNet: Learning to mitigate atmospheric turbulence in planetary images. In Proc. of AAAI Conference on Artificial Intelligence, 2025.

[34] Shengqi Xu, Run Sun, Yi Chang, Shuning Cao, Xueyao Xiao, and Luxin Yan. Long-range turbulence mitigation: A large-scale dataset and a coarse-to-fine framework. In Proc. of European Conference on Computer Vision, 2024.

[35] Wen Yang, Jinjian Wu, Jupo Ma, Leida Li, and Guangming Shi. Motion deblurring via spatial-temporal collaboration of frames and events. In Proc. of AAAI Conference on Artificial Intelligence, 2024.

[36] Rajeev Yasarla and Vishal M. Patel. Learning to restore images degraded by atmospheric turbulence using uncertainty. In Proc. of International Conference on Image Processing, 2021.

[37] Rajeev Yasarla and Vishal M. Patel. CNN-based restoration of a single face image degraded by atmo-

spheric turbulence. IEEE Transactions on Biometrics, Behavior, and Identity Science, 4(2):222–233, 2022.

[38] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In Proc. of Computer Vision and Pattern Recognition, 2022.

[39] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In Proc. of Computer Vision and Pattern Recognition, 2018.

[40] Xingguang Zhang, Nicholas Chimitt, Yiheng Chi, Zhiyuan Mao, and Stanley H Chan. Spatio-temporal turbulence mitigation: A translational perspective. In Proc. of Computer Vision and Pattern Recognition, 2024.

[41] Xingguang Zhang, Zhiyuan Mao, Nicholas Chimitt, and Stanley H. Chan. Imaging through the atmosphere using turbulence mitigation transformer. IEEE Transactions on Computational Imaging, 10:115–128, 2024.

[42] Bolei Zhou, Hang Zhao, Xavier Puig, Sanja Fidler, Adela Barriuso, and Antonio Torralba. Scene parsing through ADE20K dataset. In Proc. of Computer Vision and Pattern Recognition, 2017.

[43] Bolei Zhou, Hang Zhao, Xavier Puig, Tete Xiao, Sanja Fidler, Adela Barriuso, and Antonio Torralba. Semantic understanding of scenes through the ADE20K dataset. International Journal of Computer Vision, 127(3):302–321, 2019.

[44] Chu Zhou, Minggui Teng, Jin Han, Jinxiu Liang, Chao Xu, Gang Cao, and Boxin Shi. Deblurring low-light images with events. International Journal of Computer Vision, 131(5):1284–1298, 2023.

[45] Xinyu Zhou, Peiqi Duan, Yi Ma, and Boxin Shi. EvUnroll: Neuromorphic events based rolling shutter image correction. In Proc. of Computer Vision and Pattern Recognition, 2022.