从事件到视频：递归重建的数学、证据与测量边界

01 / 把视觉接口与直接测量分开

原文 pp.1–3；本册重新组织问题与计量接口

事件相机输出的是像素位置、时刻与极性，不是连续灰度照片。E2VID 的工程意义在于把这种异步数据变成传统视觉算法可以消费的图像序列，并用递归状态保存跨窗的信息。它解决的是“怎样形成有用的中间表示”，不是证明每个重建像素都具有辐射标定的真值。对具身导航，稳定的图像特征可能已足够；对实验流体的粒子质心或亮度定量分析，则需要另立测量误差预算。

本包固定底本为 arXiv:1906.07165v1，2019-06-15，26 页，包含正文、参考文献和补充附录 A–F。作者为 Rebecq、Ranftl、Koltun 与 Scaramuzza；TPAMI 论文的 DOI 另列在元数据，不能把本次下载的预印本说成期刊排版终稿。官方页面只有 v1，且标示默认非独占分发许可，因此本册采用独立推导、工程映射与证据分析，而不沿原文逐段替换。

阅读时保持三种时间量：传感器事件的时间戳分辨率、所请求视频帧之间的物理间隔、计算机生成一帧的耗时。它们既不相等，也不能互相推出。全册的英文段落是原创对应论述，原文完整内容仍在随包英文 PDF。前八章建立模型与实现，九至十二章检验证据与边界，最后两章连接 PTV、EDI 和具身任务。

接口中可以给每项输出附加 provenance：measured 表示直接记录的事件或外部标定，estimated 表示由模型输出的位置/图像，inferred 表示补足覆盖空洞的场值。三者采用同一坐标系，却不能共享同一置信度来源。一个可追溯的样本至少包含原始数据文件、窗起止时刻、模型/后处理版本和消费结果；这样才能把下游失效重新定位到输入与假设。

对象 / Object	单位或形状	能直接解释什么
原始事件	(x, y, t, p)	阈值跨越及其时刻
重建图像	[H,W], usually [0,1]	模型给出的图像表示
粒子位置	px 或标定后的 m	需要独立定位验收
输出与运行时间	s / frame; s / inference	采样节奏与算力负担

> The reconstruction is an interface, not a radiometric certificate. A recurrent image representation can make existing vision systems useful on events, while particle metrology requires an additional localization and calibration argument. This companion fixes the 26-page arXiv v1 rather than silently substituting the journal layout. It separates sensor timestamp resolution, physical output spacing and computation latency throughout. The English passages are original analysis; the untouched source PDF remains the complete source text. A compact provenance field separates measured events, model estimates and field values inferred beyond coverage. They may share coordinates but not an uncertainty source. Record the input file, physical window, model and display versions with the consumer output. This makes a downstream failure traceable to the observation and assumptions that produced it.

02 / 对数积分与不可恢复的初值

原文 p.3 §3；本册独立可识别性推导

设线性光强为 I，记录的对数亮度为 ℓ=log I。理想事件触发式只约束相邻触发之间的增量：极性 p 对应 ±C 的阈值跨越。固定像素从参考时刻 t_0 累计事件，得到 $\ell(t)=\ell(t_0)+C\Sigma p$。每个像素都仍带有未知初始亮度，且实际观测还包括尚未跨过下一阈值的残差。即使计数完全准确，也不能由差分观测单独确定绝对图像。

给任意静态空间函数 b(x,y)，将整段对数亮度改成 ℓ+b，理想事件序列不变。等价地，线性光强乘以任意正的静态空间增益，仍产生相同的时间对比事件。这是观察算子的零空间，不是加深网络就能消掉的噪声。网络可以利用自然图像统计选择一个合理解，却无法把这项选择变成新增传感器观测。

真实阈值依赖像素、极性和工作条件，漏事件及热像素还会破坏理想累积。若每次事件的阈值有独立、零均值误差 σC，计数 n 次后仅阈值误差的标准差也按 √n 增长；有固定偏差则可按 n 增长。这是解释积分漂移的简化误差模型，不是本文传感器的实测分布。无事件像素仅表示变化没有产生被记录的阈值跨越，不能证明信号严格恒定。

一个独立的无事件例子说明量化边界：从恰好触发后的参考亮度出发，若 C=0.2，强度增加不到 exp(0.2)−1≈22.14% 仍可能没有下一次正事件。这不是器件保证的死区百分比，而是理想模型在这一参考条件下的计算。若读者把“无事件”编码成高置信静态像素，误差会系统集中到弱对比和慢运动区域。

$$\ell(t)-\ell(t_0)=C\sum_{t_0<t_i\le t}p_i+r(t)-r(t_0)$$

$$\ell'(x,y,t)=\ell(x,y,t)+b(x,y)\quad\Longrightarrow\quad\mathcal{E}[\ell']=\mathcal{E}[\ell]$$

> At each pixel, ideal events determine log-intensity differences, leaving an unknown initial value and a quantization residual. Adding a static spatial offset to log intensity preserves the event stream. A learned prior selects among compatible appearances; it does not observe the missing offset. Independent threshold errors accumulate differently from systematic bias. The square-root and linear growth laws here are elementary error-budget models, not measured noise statistics of the original experiment. With ideal threshold 0.2 and a reference exactly at the last trigger, an increase below about 22.14% can remain below the next positive threshold. This is a model calculation, not a sensor specification. Treating every empty pixel as a high-confidence constant-intensity measurement therefore creates a predictable failure mode in slow or low-contrast regions.

工程动作：保留原始事件和固定曝光的 APS 亮度锚；把图像先验补出的区域与直接观测区域区分。

03 / 五个时间箱：离散化保留了什么

原文 p.3 §3.2、式(1)；本册离散算子分析

原文把每一窗事件映射为五个时间箱，事件极性按时间距离线性分配给最近两个箱。同一像素的空间位置不插值；时间索引从零到 B−1。窗内归一化只用于网络输入坐标，物理起止时间仍必须保存在元数据。把所有窗都缩放为相同长度，并不表示它们具有相同的真实持续时间。

线性时间核在合法范围内具有分割统一性：每个事件分配到相邻箱的权重之和为一，因此所有箱的和等于带符号计数。该等式是体素实现的直接单元测试。但体素并不保留每个事件的精确时间；同一箱对附近不同事件配置可能给出相同张量。正负极性在同一像素和时间邻域也可能抵消。网络无法恢复被编码过程严格丢弃的独立信息。

独立可算例：B=5，某像素在归一化时间 0、0.25、0.625、1 上出现极性 +1、−1、+1、+1，得到 [1,−1,0.5,0.5,1]，总和为 2，等于事件极性和。原文随后对非零体素做均值与标准差归一化，这有助于分布稳定，却不再保持计数的绝对幅度。计数守恒应在归一化前测试；ΔT=0 的同时刻事件要有显式分箱约定，不能除零。

实现时还要明确事件时戳的原点与数值类型。若以很大的绝对微秒时戳先转 float32、再做差，小窗中的微秒差可能先被舍入掉。先在 int64 或 float64 中减去窗起点，再换算单位与归一化，才能保留所需间隔。总极性和检查发现不了所有时间错位，因此应再检查首末事件恰落在箱 0 与箱 B−1，以及中间事件的两箱权重。

$$\tau_i = (B-1)\frac{t_i - t_0}{t_{N-1} - t_0}, \qquad E_{bxy} = \sum_{i:(x_i,y_i)=(x,y)} p_i [1 - |b - \tau_i|]_+$$

$$\sum_{b=0}^{B-1} E_{bxy} = \sum_{i:(x_i,y_i)=(x,y)} p_i$$

```
events [N,4] -> physical window metadata
  -> time splat -> tensor [5,H,W]
  -> normalize nonzero entries -> network input
retain original timestamps for geometry and velocity
```

> Temporal splatting preserves the signed event sum before input standardization, provided the time kernel is correctly clipped at the boundary. It does not preserve all timestamps or all opposing events. Physical duration belongs in metadata even though the tensor uses normalized time. The worked five-bin example is a local CPU calculation. Test count conservation before the paper's nonzero-value normalization, and handle zero-duration packets explicitly rather than letting a division by zero enter the network. Subtract the packet origin before reducing timestamp precision. Large absolute microsecond timestamps converted to float32 can lose the small differences that the temporal bins need. Signed-count conservation alone will not detect every time-order error. Test endpoint placement and the exact two-bin weights as separate invariants, with an explicit time unit.

04 / 固定计数与固定时长不是同一相机

原文 pp.3、6–8、13；本册独立吞吐与采样分析

固定事件数 N 的窗在活跃场景中更短，在低对比、静止或暗区更长。因此它在“每次计算接收多少事件”上稳定，却在物理采样频率上变化。固定时长 τ 的窗给出稳定时钟，但其计数、信息密度和噪声比例随场景变化。原文训练使用固定计数窗，并评估不同时间窗的泛化；这不等价于网络显式理解真实 ΔT。

若事件率近似 R，非重叠窗的平均持续时间约 N/R。高帧率展示还可以用长度 N、步进 D 的重叠滑窗：新输出的物理间隔约 D/R，而每一张图仍覆盖 N/R 的事件上下文。独立算例中 N=10,000、D=1,000、R=1 M event/s，则上下文约 10 ms，输出间隔约 1 ms；称其为 1 kHz 输出不能推成每帧只依赖 1 ms 的观测。

重叠窗还有状态问题：把重叠原始数据当成互不重复的新输入连续更新递归状态，会使相同事件多次贡献。工程实现必须明确：采用匹配训练协议的重叠重建，还是从检查点重放、单独分支求目标时刻图像。对于流体计量，保存每帧真实参考时刻和有效时间支撑，比给视频写一个常数 FPS 更重要。帧间位移除以错误的时间差，会直接把速度单位算错。

实时查询还受到因果性的限制：以某时刻为中心、使用前后事件的窗，必须等未来半窗到齐后才能输出。若报告时间戳是中心时刻，数据龄期天然至少包含这半窗；若报告末时刻，图像语义就相应改变。离线评估可以对称取窗，在线控制不能把尚未到来的事件当成零延迟输入。

$$\Delta T_{\mathrm{context}} \simeq N/R, \qquad \Delta T_{\mathrm{output}} \simeq D/R$$

$$\widehat{\boldsymbol{v}} = \frac{\widehat{\boldsymbol{x}}_{k+1} - \widehat{\boldsymbol{x}}_k}{t_{k+1} - t_k}$$

独立算例，R=1 M/s	N	D	上下文 / 输出间隔
非重叠	10,000	10,000	10 ms / 10 ms
重叠查询	10,000	1,000	10 ms / 1 ms
低事件率 0.1 M/s	10,000	1,000	100 ms / 10 ms

> Count-based packets adapt their physical duration to event activity. Time-based packets preserve a clock but not information density. With overlapping packets, output spacing is controlled by the hop D, while the context duration depends on N. These are distinct temporal supports. A recurrent implementation must specify whether repeated events are replayed, branched or counted again. Velocity estimation needs actual frame timestamps, not an assumed constant video frame rate. A centered window uses future events and cannot be produced causally until that future support arrives. If the output timestamp denotes the center, its age includes at least half the window. An end-stamped window has different semantics. Offline symmetric reconstruction is valid, but it is not a zero-latency online observation.

05 / 递归 U-Net 的形状与数据契约

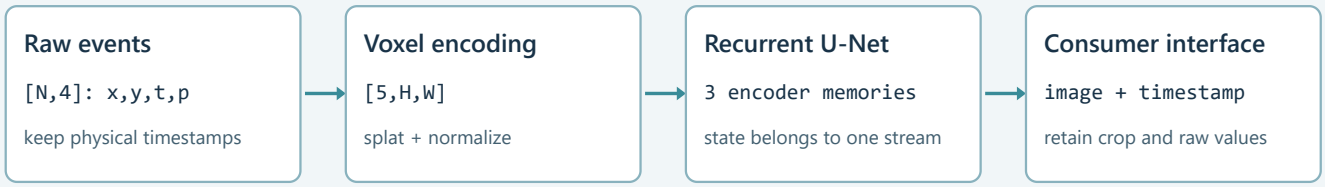
原文 pp.4–5 §3.4/Fig.4，p.17 Appendix B；本册张量展开

默认网络是头层、三个递归编码器、两个残差块、三个解码器和单通道预测层。头层通道数 32，每次编码加倍，最深层达到 256。对称跳接使用逐元素相加，不是把通道拼接；输出经 sigmoid。编码器先做步长 2 的卷积，再用 ConvLSTM 更新状态；解码器采用双线性上采样与卷积。图中画的少层示例不能误当默认三层配置。

为了具体检查形状，设输入采用 [batch,5,184,240]。184 是将实际 180 行按 8 的倍数做边界填充的一种工程选择，本册不把该填充说成论文唯一规定。三个编码网格依次为 92×120、46×60、23×30；解码网格逆序恢复，并在输出裁回原始有效区域。若直接在 180 行上连用取整、再无条件把双倍上采样与跳接相加，就可能出现 45 与 46 行的错配。

每个编码尺度持有自己的隐藏/记忆状态，分辨率、batch 或序列来源改变时，旧状态不能直接复用。相机 A 的状态接到相机 B，不是在线记忆而是跨样本污染。落地接口应同时携带图像、时戳、有效裁剪、状态代号和传感器标识；不要只返回一张看起来正常的 PNG。全卷积意味着空间尺寸可变，并不取消状态尺寸与边界处理的约束。

阶段	通道	空间形状（独立填充例）
头层	32	184 × 240
编码 1 / 2 / 3	64 / 128 / 256	92×120 / 46×60 / 23×30
解码 1 / 2 / 3	128 / 64 / 32	46×60 / 92×120 / 184×240
预测与裁剪	1	184×240 -> 180×240



本册原创接口图 / Original interface diagram. 箭头只占节点间空隙；隐藏状态单列，避免把历史状态误作新传感观测。

> The default channels are 32, 64, 128 and 256 with three recurrent encoder scales and additive skip connections. A padded 184 by 240 example makes the down/up-sampling contract explicit. The padding choice is an engineering example, not a reported experimental setting. Recurrent states are tied to a sequence and spatial shape. A usable API carries timestamps, crop transforms and state ownership together with the image.

06 / 记忆门并不保证无限记忆

原文 pp.4、13–14；本册 ConvLSTM 机制与状态测试

ConvLSTM 把普通 LSTM 的矩阵乘换成局部卷积，输入门、遗忘门与输出门控制空间记忆。为避免与原文把整体状态写作 c 的记号混淆，本册把隐藏特征记为 h、记忆单元记为 m。记忆更新由保留旧信息与写入新信息两项组成；遗忘门接近 1 时更能跨步传播，接近 0 时旧信息迅速衰减。门值由网络学得，不是明确的物理守恒律。

原文把训练序列展开为 40 步。这里的步是事件窗，不是 40 ms，也不是模型只能记住 40 次更新。反向传播截断影响训练中可优化的依赖长度；推理时状态仍可连续更新，但超出训练分布的长期稳定性需要另外验证。固定计数输入又使“一次状态更新”对应的物理时间随事件率变化，所以记忆长度应同时用更新次数和秒表示。

原文用空事件张量反复驱动网络，观察到初期维持图像、随后逐渐衰减，而且高对比区域保留更久。这是关于学得记忆的实验证据，不证明现实中无事件时也必须按同样频率调用网络。合适的 CPU 接口测试是：重复读取同一状态不应修改它；重置后输出应可重复；并行流不能共享可变状态；丢包和时间跳变应能触发明确的重置或不确定度标记。

用一个不含输入的新信息的标量例可看清更新次数与秒的差别：$m_k=0.99m_{k-1}$ 时，100 次更新后只剩约 0.366 倍。若每步 10 ms，这发生在 1 s；若每步 100 ms，则是 10 s。真实门是空间可变且依赖输入的，不能用这一例拟合其全部行为；它只是证明同一离散记忆因子并不定义固定的物理时间常数。

$$i_k, f_k, o_k = \sigma(W * [x_k, h_{k-1}] + b)$$

$$m_k = f_k \odot m_{k-1} + i_k \odot \tanh(W_m * [x_k, h_{k-1}] + b_m)$$

$$h_k = o_k \odot \tanh(m_k)$$

> A ConvLSTM gate is a learned dynamical mechanism, not a conservation constraint. Forty unrolled training windows are neither forty milliseconds nor a hard inference-memory limit. Because the physical duration of a packet varies, memory should be described in both updates and time. The paper's empty-tensor experiment reveals decay of learned appearance. State isolation, explicit resets and deterministic replay are practical invariants to check before interpreting a recurrent output as a measurement. For the toy recurrence m[k]=0.99 m[k−1], one hundred updates retain about 0.366 of the state. That corresponds to one second with 10 ms steps but ten seconds with 100 ms steps. Real gates depend on space and input; the example merely shows why a discrete update count cannot define a fixed physical memory duration.

07 / 合成训练：随机化的是哪种偏差

原文 pp.3–4 §3.3/3.6，p.17 Appendix C；本册分布推理

模拟数据的优势是可以同时得到事件和无运动模糊的图像目标，并在 HDR 或高速条件下避免把传统相机的缺陷当成标签。原作把 COCO 图像映射到平面上，再模拟相机运动，以 ESIM 生成事件。1000 条、每条 2 s 的序列按 950/50 分为训练和验证。这能提供外观多样性，但运动机制主要是平面引起的全局单应变化；独立粒子、遮挡和散射介质不因此自动得到覆盖。

原文随机采样正负对比阈值，给出的均值 0.18、标准差 0.03 用于模拟场景层面的阈值随机化，不能直接称为真实器件所有像素独立同分布的标定结果。其作用是让网络不依赖一种固定事件幅度。旋转、翻转、128×128 裁剪和非零体素归一化增加外观/尺度鲁棒性，却不能替代传感器带宽、暗电流、热像素和时间偏差的真实建模。

转向实验流体时，优先随机化会改变粒子测量的因素：点扩散函数、粒径与亮度、成对粒子重叠、出平面运动、非均匀阈值、照明闪烁和事件丢失。标签应保留无噪声粒子中心与物理速度，而不是只保留自然图像。一个合理的验证切分按流场、粒子轨迹和相机实例隔离，不能把同一轨迹的相邻窗口随机分到训练与测试；后者会让记忆和背景泄漏掩盖域外误差。

仿真也需要保留相机前向模型，而不是仅在重建图像上加噪。先对无噪光强做显示 gamma、再模拟事件，会改变对数差与阈值含义；正确顺序取决于传感器实际响应。训练标签、模拟事件与评估图像应共享同一辐射约定，同时把可视化映射留在最后。只有这样，域差异才能定位到模型缺项，而不是被无意的亮度变换掩盖。

训练要素	原作设置或事实	迁移时要补什么
时序	1000 × 2 s；950/50	独立动力学与长暂停
阈值	mean 0.18; std 0.03	逐像素/极性/照度标定
运动	平面图像与相机运动	独立粒子及遮挡
目标	自然图像重建	位置、速度与可信掩膜

> Simulation supplies synchronized clean targets when ordinary frames are poor references. The original planar-image motion and threshold randomization explain a useful training route, not coverage of all fluid or sensor phenomena. For particle metrology, simulate the point-spread function, overlap, lighting and sensor defects while retaining latent particle centers. Split by trajectory, scene and sensor rather than by adjacent windows. Otherwise temporal leakage can make an apparently robust representation fail on genuinely new dynamics. Noise should enter a declared sensor forward model, not only the final display. Applying an arbitrary display gamma before event simulation changes the log increments and threshold interpretation. Keep targets, simulated events and evaluation images in compatible radiometric conventions, and apply display mapping last. Otherwise a domain gap can be caused by an undocumented transform rather than missing physics.

08 / 感知与时间损失优化的目标

原文 p.4 §3.5，p.13 Tables 6–7；本册损失解析

原作以 LPIPS 的多层图像特征距离训练重建，而不是仅按像素平方误差拟合。这样可以奖励自然图像中有用的形状和纹理，避免均值化产生的过度平滑。但感知特征对小平移、局部对比和某些细节变化不敏感，恰好可能与亚像素测量的需求相反。损失的成功首先要按它所优化的目标解释，不能把“自然”换成“位置无偏”。

时间项用光流把上一重建图像扭到当前时刻，再计算加权绝对误差。遮挡权重根据真值图像扭曲残差衰减，α=50；时间损失权重为 5，并从序列后部开始计入以留出初始化时间。训练需要光流，部署推理不需要这个监督光流。这一区分影响运行依赖：不能把训练标签提供的信息误写成实时传感输入。

扭曲一致性也有退化方向：一段几乎恒定的图像可能在某些运动/纹理条件下误差较小，却没有保留真实粒子。实际 PTV 验证应并列评估事件重投影残差、位置误差、轨迹连续性以及静止背景幻觉。若增加粒子中心损失，可以针对计量目标约束偏差，但要保留对无标签位置的未知状态，而不是让网络把每个模糊局部都自信地解释为粒子。

另一独立算例是平移一个高斯亮斑。亚像素位移 δ 会直接使质心移动 δ，但一维高斯的平方误差能量在小位移下只按 δ² 增长；整幅图再除以大量背景像素后，这个变化可能非常小。因此高 PSNR 与有用的几何误差阈值之间不存在不经假设的换算。评价应在原始像素尺度报告定位误差，而不是仅把小颗粒裁图放大后目测。

$$\mathcal{L}_R = d_{\mathrm{LPIPS}}(\hat{I}_k, I_k)$$

$$\mathcal{L}_{TC} = \left\| M_k \odot \left(\hat{I}_k - W_{k-1}^{k}(\hat{I}_{k-1}) \right) \right\|_1$$

$$M_k = \exp\left[-50 \left\| I_k - W_{k-1}^{k}(I_{k-1}) \right\|_2^2 \right]$$

> LPIPS encourages perceptual structure; it does not certify subpixel position. The temporal objective compares motion-warped predictions and downweights likely occlusions using clean target images. Those flow maps are training supervision, not required inference input. Temporal smoothness alone can favor uninformative solutions. A metrology-oriented extension needs independent center and trajectory errors alongside image and event consistency, with explicit handling of unlabeled or occluded regions. A shifted Gaussian spot moves its centroid by the shift itself, while its small-shift squared image error grows quadratically. Dividing by a large background area further dilutes the global MSE. A favorable PSNR thus does not supply a localization tolerance without additional assumptions. Evaluate center errors in the original pixel coordinate system, not by inspecting enlarged crops alone.

09 / 复盘消融：平均进步与反例

原文 p.7 Table 1，p.13 Tables 6–8；本册差值计算

原文消融给出清晰的功能分工：加时间损失后时间误差从 2.08 降至 1.43，按该表四舍五入值计算下降 31.25%；移除递归连接则时间误差达到 4.00，恢复递归后的下降为 64.25%。这些比较支持对应结构在作者测试分布中的作用。它们不是本机重新训练的结果，也不能解释为任意相机和任意目标都获得相同比例改进。

Table 1 的平均 SSIM 从最佳基线均值 0.50 到 0.62，按显示值为 24% 相对增加；但 dynamic_6dof 上本法 MSE=0.14、MR=0.05，SSIM=0.46、MR=0.52，明显存在单序列退步。把平均优势写成“所有指标都领先”会抹掉这个反例。LPIPS 又会给出不同排序，说明指标本来就在强调不同性质。

Table 8 比较 τ=50 ms 与 5 ms，本文 SSIM 从 0.62 到 0.58，按两位小数计算下降约 6.45%。原文用“≤6%”描述，可能基于未舍入值；本册保留显示值与算式，不把差异藏起来。迁移验收应预注册整图与局部目标指标，在相同事件、相同时戳、相同有效掩膜上配对比较，并给出序列级分布，而非只挑一帧最漂亮的图。

统计汇总还需区分逐序列等权平均与逐帧加权平均。前者使短序列和长序列权重相同，后者让帧数多的场景主导；若两种场景难度不同，排序可以变化。本册引用作者表的显示均值，不重构未公开逐帧分布。新实验应同时保留逐样本 CSV 和序列标识，避免只有一个无法回溯的平均分数。

作者消融 / Ablation	MSE ↓	SSIM ↑	Temporal error ↓
无时间损失	0.07	0.59	2.08
完整网络	0.05	0.62	1.43
无递归	0.08	0.54	4.00
5 ms 测试窗	0.06	0.58	见 Table 8（未列此项）

> The ablations support distinct roles for recurrence and temporal supervision, but aggregate gains include individual regressions. Dynamic 6-DoF is a concrete counterexample to universal metric superiority. Rounded SSIM values of 0.62 and 0.58 imply a 6.45% relative decrease, while the paper describes at most 6%; unrounded values may explain the discrepancy. Preserve that distinction. None of these published training results is a local rerun. A mean over sequences weights short and long recordings equally; a mean over frames does not. Either can be appropriate, but they answer different questions and may rank methods differently. The numbers here are the displayed published aggregates. A new experiment should retain per-sample values and sequence IDs so aggregation remains inspectable.

CPU 复算：(2.08−1.43)/2.08 = 31.25%；(4−1.43)/4 = 64.25%；(0.62−0.58)/0.62 = 6.4516%。

10 / 五千帧每秒不等于实时五千次推理

原文 pp.6–8 Figs.7–8，pp.11–12 Table 5；本册系统预算

高速示例展示的是可从细时间事件中合成的物理视频帧率。它可以离线生成，并可依赖重叠时间窗；不能由此推出每秒可以在线执行五千次完整网络。原文 Table 5 对 10,000 事件一批报告本法 5.53 ms，硬件是 RTX 2080 Ti；HF 在 i9-9900K CPU 上是 0.70 ms，附滤波为 1.45 ms，MR 在 GPU 上为 0.84 ms。不同算法、预处理和硬件各不相同，不能把这张表当严格同设备的单核竞速。

仅把 5.53 ms 倒数换成 180.8 次/s，是不含事件成包、传输、排队和输出后处理的理想上界；它并不是本文额外测得的持续吞吐。若完整批处理服务率低于到包率，队列会增长，即使单次平均耗时看起来不大，控制器也会使用越来越旧的状态。实际闭环关心端到端数据龄期及尾延迟，而不仅是算子均值。

本轮只做 CPU 文献解析、代数验证与 Edge 印刷排版，没有加载网络权重或运行 GPU。要复现系统时，记录物理窗末时刻、到达时刻、推理开始/结束、控制消费时刻；分别统计 p50/p99、积压与丢包策略。在安全控制里允许丢旧帧有时比无限排队更合理，但必须让状态更新与被跳过的事件保持一致，不能简单删帧而仍声称事件全部被递归消费。

若 10 ms 上下文以中心时刻标记，再经历 5 ms 服务时间，则在无传输与排队时数据龄期也约为 10 ms：其中 5 ms 等待未来半窗，5 ms 用于处理。这个独立系统算例说明为何控制预算要从物理观测时刻开始，而非从函数调用开始。异步系统还要记录哪一帧最终被控制器消费，单独的推理计时日志不足以重建闭环延迟。

$$T_{\mathrm{age}} = t_{\mathrm{control}} - t_{\mathrm{physical}}, \qquad \rho = \lambda_{\mathrm{packets}}\,\mathbb{E}[T_{\mathrm{service}}]$$

作者报告设置	每批 10k 事件耗时	设备
E2VID	5.53 ms	RTX 2080 Ti GPU
HF / HF + filter	0.70 / 1.45 ms	i9-9900K CPU
MR	0.84 ms	RTX 2080 Ti GPU
本轮手册	未测神经网络推理	仅 CPU 文献与数学验证

> A high physical output frame rate and a high real-time inference rate are different capabilities. The published 5.53 ms batch inference uses a GPU, while HF's CPU time and MR's GPU time have different implementation contracts. The reciprocal 180.8 calls/s is only arithmetic, not an observed end-to-end throughput. For a controller, timestamp age, queue growth and p99 latency determine usability. This production task performs no model inference or GPU work. A centered 10 ms context followed by a 5 ms service already yields roughly 10 ms data age without transfer or queueing. Half comes from waiting for future context. This independent budget example explains why timing should start at the physical reference, not at the function call. Log the frame actually consumed by the controller, not just completed inference calls.

11 / 图像桥接何时胜过专用事件特征

原文 pp.9–11 §5.4/Tables 3–4，p.23 Appendices E–F；本册任务分析

图像接口能复用已有分类网络、预训练视觉特征与 VIO 后端。这是系统收益，而不只是单张重建质量的变化：原本难以直接输入标准图像网络的事件，被转换为通用表示。迁移时需要评估完整消费链，而不是只比较事件前端的参数量。已有图像数据的先验与工程成熟度是实质资源，同时也可能带入不属于传感器观测的语义偏差。

原作分类并非每一数据集都最好：N-MNIST 上 98.3% 低于 HATS 99.1%，N-CARS 上 91.0% 高于 90.4%，N-Caltech101 上 86.6% 对 70.0%。这些结果发生在各自的训练、预处理和划分下。补充材料记录了窗口、分类器和训练方法；不能把三种任务混成相同难度或把 N-Caltech101 增益移植为粒子追踪增益。

VIO 的 Table 4 报告序列中位平移误差 0.15 m，对比仅事件加 IMU 的 UltimateSLAM 0.27 m；按显示值下降约 44.4%。但 shapes_6dof 本法 1.09 m，而对方 0.09 m，原文解释为特征稀少造成 VINS-Mono 初始化困难。系统误差应该连同消费者失效一起报告。对于机器人，重建图像能否稳定初始化、保持相机时序与标定，通常比重建的视觉锐度更直接决定闭环效果。

像素精度转成几何误差还依赖相机尺度。小角近似下 $\delta\theta \approx \delta u / f$；独立算例取 f=500 px、δu=0.1 px，则角误差约 0.0002 rad。在 5 m 距离上对应约 1 mm 横向误差，尚未加入深度与外参误差。因而同样的重建定位指标，在近距离抓取和远距离导航中对应不同的物理影响。

原作任务	E2VID 路线	比较对象	解释
N-MNIST accuracy	98.3%	HATS 99.1%	本法退步
N-CARS accuracy	91.0%	HATS 90.4%	小幅提升
N-Caltech101 accuracy	86.6%	HATS 70.0%	迁移特征收益
VIO median error	0.15 m	U.SLAM E+I 0.27 m	保留初始化反例

> The image bridge benefits from mature pretrained features and downstream systems, but its success remains task-dependent. N-MNIST is a published counterexample; N-Caltech101 benefits substantially. VIO's median error improves, yet a sparse-feature sequence fails to initialize reliably. Keep front-end appearance, downstream initialization and final pose error separate. A particle tracker requires its own evidence rather than inheriting the classification or VIO gains. Pixel error becomes geometric error through camera calibration. With focal length 500 pixels, a 0.1-pixel error is about 0.0002 radians; at five meters that alone corresponds to about one millimeter laterally. This is an illustrative first-order calculation. Depth and extrinsic errors add further terms, so identical image metrics do not imply identical physical accuracy across tasks.

12 / HDR、颜色与辐射一致性

原文 p.4 §3.7、pp.8–10 §§5.2–5.3；本册标定与误差传播

事件触发反映对数变化，因而可以在宽亮度范围内保持有用的动态信息；重建图像显示 HDR 场景细节，是利用这种传感能力的接口。可是输出 sigmoid 的 [0,1] 与随后采用的分位数归一化都不定义绝对辐亮度。原文用 1% 与 99% 分位数做稳健拉伸再裁剪，目的是让图像更容易观看。把该 PNG 当线性光强输入另一个物理反演模型会引入未记录的增益和截断。

对于单粒子质心，统一的正比例增益在无截断、背景扣除正确时不会改变强度加权中心；加性背景、局部对比变换、饱和裁剪和逐窗分位数漂移则可能改变中心及阈值分割。对整个视场使用同一拉伸也可能让少数高亮热像素影响弱粒子的可见性。应保存网络原始浮点输出、后处理参数和显示版本，而不是仅保存最终可视化。

彩色事件的通道由像素滤色阵列分配。原作通过各颜色子集重建和颜色合成展示可行性，无须重新训练单色网络；这并不提供同时共址、独立三通道的原始观测。通道间插值和不同事件密度会影响彩边与空间细节。流体实验若只需单色散射粒子，应先选与照明谱段一致的观测和标定；若要做浓度或温度计量，则必须单独建立响应曲线和通道串扰模型。

背景扣除是上述增益不变性的关键条件。如果在粒子裁片上加一个常数背景而不减去，分子会额外加入背景乘坐标和，分母也会增大，质心被拉向裁片的几何中心。改动裁片边界即可改变这个偏差。工程上应固定裁片规则、背景估计与饱和掩膜，并把这些参数与网络版本一同冻结再比较候选。

$$I_{\mathrm{display}} = \mathrm{clip}_{[0,1]}\left(\frac{\hat{I} - q_{0.01}}{q_{0.99} - q_{0.01}}\right)$$

$$\bar{x} = \frac{\sum_i x_i (I_i - b_i)}{\sum_i (I_i - b_i)}$$

> High dynamic range visibility is different from calibrated radiometry. Percentile stretching and clipping make a useful display but alter the intensity model. Preserve the raw floating-point reconstruction and the display transform. A centroid is invariant to a global gain only under explicit background and clipping assumptions. Color reconstruction from a filter mosaic also contains spatial interpolation; it is not three co-located measurements. Quantitative spectroscopy or concentration estimation needs a separate response calibration. Unremoved constant background pulls an intensity centroid toward the geometric center of its crop. Changing the crop can change that bias even when the underlying particle has not moved. Freeze crop rules, background estimation and saturation masks alongside the network version. Gain invariance is useful only under the assumptions that make its algebra valid.

13 / 接入 EDI、光流与实验粒子测速

本册独立工程映射；关联库内 06/12/19/25/28，不冒充原作实测

E2VID 与 EDI 应按观测结构选用：E2VID 只用事件和递归先验生成图像；EDI 额外利用模糊帧提供曝光积分锚。只用事件时绝对亮度存在自由度；加入曝光帧可以约束它，但仍依赖阈值、曝光时序与辐射线性假设。把一个模型的重建结果当另一个模型的“实测帧”，会形成依赖同一先验的循环验证，不能获得独立证据。

粒子定位需要局部误差。设背景扣除后的真实粒子强度为 I，总强度 S；重建误差为 δI。对质心一阶展开，位置扰动与 Σ(x−x̄)δI/S 成正比。总 MSE 或 PSNR 不包含误差的这种空间一阶矩，因此两幅图具有相同 MSE 时，粒子中心偏差可以完全不同。误差如果主要落在背景，整图指标很敏感而中心不变；弱粒子一侧的小偏差则可能明显移动中心。

从二维追踪段到流场也有下一层逆问题：没有追踪覆盖的区域，即使施加无散约束也不能自动变成直接测量。可用 E2VID 提供候选观测，再以原始事件一致性、独立 APS/高速相机或合成真值检查位置；以 PTV 覆盖掩膜约束数据项，再把物理投影结果标为场推断。对具身控制则保留未知空间，不用美观纹理填补碰撞安全距离。优先改测量链和掩膜契约，再改网络规模。

数据一致性必须覆盖实际前向观测，而非仅让两张网络输出互相吻合。对事件重生成作检查仍保留静态亮度自由度，因此它可以排除部分时间变化错误，却不能证明绝对纹理恢复。把前向残差低作为必要证据，再用独立位置或曝光锚补足未可识别部分，比简单叠加更多一致性损失更有判别力。

$$\delta\bar{x} \simeq \frac{1}{S}\sum_i (x_i - \bar{x})\delta I_i, \qquad S = \sum_i I_i$$

$$\mathrm{PSNR} = 10\log_{10}(I_{\max}^2/\,\mathrm{MSE})$$

> Choose E2VID or EDI according to the available observations, not image sharpness alone. A reconstruction fed back as a supposed measurement is not independent validation. The first-order centroid error depends on the spatial moment of intensity error, which global MSE does not encode. A divergence-free field inferred beyond tracked particles is still an inference. Preserve raw events, coverage masks and independent references when connecting reconstruction to fluid metrology or safety-critical embodied perception. Forward event consistency rejects some incorrect temporal changes but still leaves the static brightness gauge. Agreement between two network outputs is weaker than agreement with an actual observation. Use the forward residual as necessary evidence, then add independent positions or exposure anchors for the unidentifiable component. More consistency terms do not automatically supply missing information.

本页提出的是工程研究路径；没有把本文的图像、分类或 VIO 数字写成 PTV/具身实验结果。

14 / 可执行阅读与复现验收清单

原文 pp.14–26 含书目及附录；本册独立 CPU 验证与术语

先复现输入算子，再接网络：用小事件集合验证体素计数和时间端点；记录每窗 ΔT、N 与 hop；在每个编码尺度检查状态尺寸；固定裁剪变换后才把像素坐标变成相机光线。其次冻结模型与后处理，分别对图像质量、粒子位置或下游任务作配对评估。最后再测端到端延迟和资源。这样可以定位失败发生在传感、编码、先验、几何还是控制消费端。

随包 math_checks.py 在 CPU 上验证时间体素的计数守恒、静态对数偏置的观测等价性、质心对纯增益的不变性和一阶扰动式，并复算文中消融百分比。这些是代数与实现检查，不包含预训练模型推理、重新训练或传感器数据重测。原始论文正文、图表及补充材料的逐页文本清单保存在 source_map.json，完整英文原文可直接回查。

研究优先级建议：针对粒子轨迹保真，先比较“事件直接追踪、E2VID 图像追踪、EDI 图像追踪”三条固定后端路线，按遮挡、低对比和陌生粒子纹理分层。若目标是具身导航，先验证初始化成功率、数据龄期与故障恢复。只有当这些误差定位到表征瓶颈，再考虑任务感知重训。不要把神经图像、物理投影场和直接粒子测量共用一个未经区分的真值标签。

推荐保存的最小结果记录为：sensor_id、window_start/end、reference_time、N、hop、model_id、state_id、crop_transform、raw_output 与 validity_mask。后续控制或测速程序按 reference_time 计算时间差，按 crop_transform 恢复原坐标，按 validity_mask 决定是否生成测量。可视化文件独立命名，避免用户把经过对比拉伸的版本误当原始数值。

English term	本册固定中文	使用注意
event voxel grid	事件体素网格	时间归一化不改变物理时戳
recurrent state	递归状态	逐序列、逐尺度隔离
perceptual loss	感知损失	不是定位误差
radiometric calibration	辐射标定	独立于显示拉伸
identifiability	可识别性	先验选择不等于新增观测

> Reproduce the observation and encoding contract before judging the network. The included CPU tests establish small algebraic invariants and rounded-number calculations, not model or sensor performance. For fluid experiments, compare direct-event, E2VID-image and EDI-image trackers with a fixed backend and stratified independent references. For embodied navigation, prioritize initialization, data age and recovery. The next experiment should target the observed bottleneck rather than automatically enlarging the model. A minimal record carries sensor and state IDs, physical window and reference times, event count and hop, model version, crop transform, raw output and validity mask. Consumers use those fields for time differences and geometry. Store display images separately so a contrast-stretched visualization cannot silently become the numerical input to a measurement pipeline.

原文与补充材料：arxiv.org/abs/1906.07165v1。代码与权重另按其许可处理；本包不含权重。