《国际计算机视觉杂志》
https://doi.org/10.1007/s11263-022-01708-3

# 用于非均匀运动去模糊的事件引导多图块自监督网络

**Hongguang Zhang · Limeng Zhang · 戴玉超 · 李宏东 · Piotr Koniusz**

arXiv:2302.07689v1 [cs.CV] 14 Feb 2023

收稿日期：2021年9月4日 / 接收日期：2022年10月20日

**摘要** 当代深度学习多尺度去模糊模型存在诸多问题：1）在非均匀模糊图像/视频上的表现不佳；2）仅通过增加模型深度和细化尺度层级无法提升去模糊性能；3）单个 RGB帧包含的运动信息有限，难以用于去模糊；4）以往模型对空间变换和噪声的鲁棒性有限。下面，我们在初步论文 [59]的基础上引入若干机制，以解决上述问题：I）我们提出一种新颖的自监督事件引导深度分层多图块网络（MPN），通过由细到粗的分层局部表示处理模糊图像和视频；II）我们提出一种新颖的堆叠式流水线 StackMPN，以在网络深度增加时提升去模糊性能；III）我们提出一种事件引导架构，利用视频中包含的运动线索来处理视频中的复杂模糊

IV) 我们提出一种新颖的自监督步骤，使模型接触随机变换（旋转、尺度变化），并增强其对高斯噪声的鲁棒性。与当前多尺度方法相比，我们的 MPN 在 GoPro 和 VideoDeblur 数据集上达到了最先进水平，且运行时间更快，为 40×。处理一张 1280 ×720 分辨率的图像仅需 30 毫秒，因此它是首个可实时处理 720p 图像、达到 30 fps 的深度运动去模糊模型。对于 StackMPN，通过增加网络深度，我们在 GoPro 数据集上取得了超过 1.2 dB 的显著提升。利用事件信息和自监督进一步将结果提升至 33.83 dB。

H. Zhang（通讯作者）是中国北京 100141 的 AMS 系统工程研究所助理教授。电子邮件：zhang.hongguang @outlook.com。L. Zhang 是上海交通大学博士生。 ·Y. Dai 是西北工业大学教授。 ·H. Li 是澳大利亚国立大学教授。 ·P. Koniusz 是 Data61/CSIRO 的高级研究科学家，也是澳大利亚国立大学的高级荣誉讲师。电子邮件：piotr.koniusz@data61.csiro.au

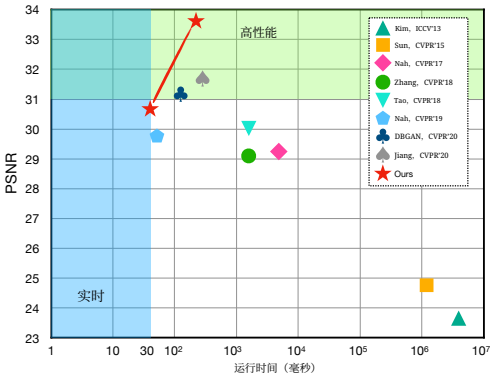



**图1：** GoPro 数据集上最先进的深度图像去模糊方法与我们的方法的峰值信噪比和运行时间对比 [28]。蓝色区域表示实时推理，绿色区域则表示高性能运动去模糊（超过 30 dB）。对于 1280×720图像，我们的方法在 30 fps 下实现了最佳性能。事件引导版本和模型的堆叠变体进一步提升了性能，但运行时间也有所增加。

## 1 引言

非均匀盲图像去模糊旨在去除由相机运动和场景动态造成的不需要的模糊 [28, 49, 32]。在深度学习取得成功之前，传统去模糊方法通常采用各种约束或正则化来近似运动模糊滤波器，这需要进行代价高昂的非凸非线性优化，并且依赖于过度严格的空间均匀模糊核假设，导致复杂模糊模式的去模糊效果不佳。

基于深度卷积神经网络（CNN）的去模糊方法 [21, 43]以端到端方式 [28, 49]学习模糊输入图像与对应清晰图像之间的回归关系。为了利用不同处理层级的去模糊线索，“由粗到细”方案已

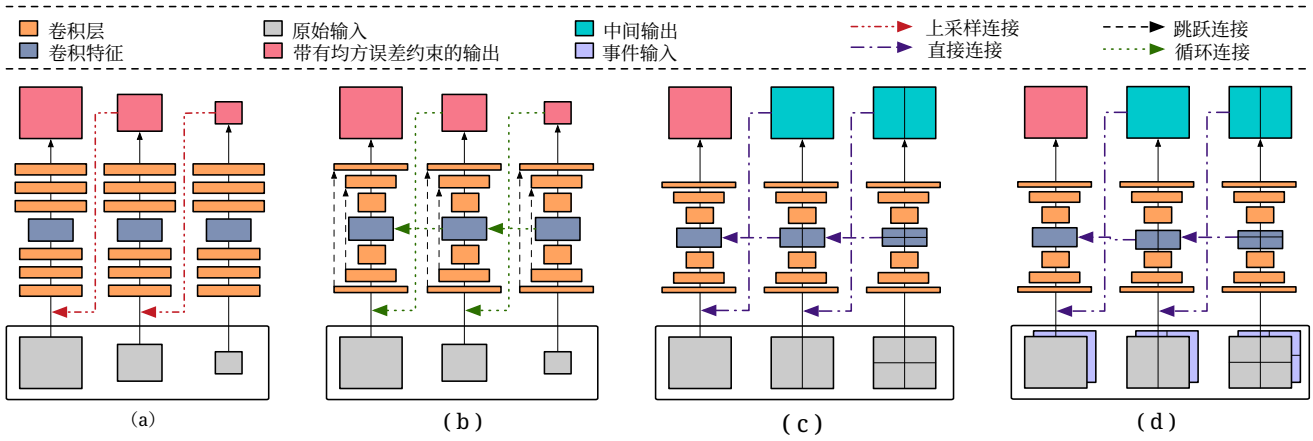



**图2：** 不同网络架构的比较。从左到右：(a) 多尺度 [28], (b) 尺度递归 [49], (c) 我们的分层多图块架构。我们没有采用任何跳跃连接或循环连接，因此模型结构更简洁。(d) 我们的事件引导的多图块网络架构，其中事件表示与原始模糊帧拼接，作为双流输入。建议以彩色查看。

通过多尺度网络架构 [28] 和尺度循环架构 [49]扩展到深度卷积神经网络场景。在“由粗到细”方案下，清晰图像在金字塔的不同分辨率上逐步复原。Nah 等.[28] 展示了卷积神经网络模型从多尺度模糊图像中去除运动模糊的能力，并设计了一种多尺度损失函数来模仿传统的由粗到细方法。沿用类似的流程，Tao 等.[49] 在不同尺度间共享网络权重，以提升训练效果和模型稳定性，因此与 [28]相比实现了高效的去模糊。然而，当前的深度去模糊方法仍面临重大挑战：

– 在由粗到细的多尺度方案下，由于滤波器尺寸较大，大多数网络需要大量训练参数。因此，多尺度和尺度递归方法的运行时间开销很大（见图 1），且难以提升去模糊质量。– 在多尺度方法中，增加低分辨率输入对应的网络深度似乎无法提升去模糊性能 [28]。– 在模糊复杂的情况下，模型无法从 RGB 帧中捕获运动信息，因此无法有效处理视频去模糊。– 训练得到的模型对空间变换和随机噪声的鲁棒性有限，这限制了其在实际应用中的实用性。

本文针对多尺度和尺度递归架构所面临的挑战，提出了一种新型架构，通过分层多图块模型利用不同尺度下的去模糊线索。具体而言，我们提出了一种简单而有效的多层卷积神经网络模型，称为深度多图块网络（MPN），该模型以多图块层次结构作为输入。通过这种方式，来自局部区域去模糊的残差线索可经由类残差连接传递至网络的下一级，该级处理更粗粒度区域。多图块上的特征聚合已被用于图像分类 [23, 14, 26, 20]。例如，[23] 提出了空间金字塔匹配（SPM），将图像划分为由粗到细的网格，并在其中计算特征直方图。在 [20]，中，一种二阶细粒度图像分类模型使用重叠图像块的特征嵌入和位置嵌入进行聚合。Sun等。 [48] 通过卷积神经网络学习逐图块运动模糊核，并通过代价高昂的能量优化实现复原。

我们的网络具有三方面优势：1）由于不同层级的输入具有相同的空间分辨率，我们采用类残差学习，这种方法所需的滤波器尺寸更小，推理速度更快；2）在最精细层级上，可用的图像块相对更多，因此我们采用能够接触到更多训练数据的类 SPM 模型；3）我们的架构鼓励模型从较简单的任务（小图像块）逐步学习到较难的任务（大图像块），这一渐进式学习过程有助于确保不同位置和空间尺寸上的去模糊一致性。

为了克服多尺度和多图块模型中堆叠深度 的局限性，仅通过引入更粗或更细的网格来增加模型深度，并不能提升现有模型的整体去模糊性能。因此，我们提出了 MPN 的全新堆叠版本，通过堆叠多个子模型，其性能能够得到有效且持续的提升。

作为我们初步论文 [59],的扩展，我们提出了事件引导 MPN，以处理快速变化场景中的复杂运动模糊。为此，我们采用动态与主动像素传感器（DAVIS），同时生成灰度主动像素传感器（APS）帧和事件帧，其中物体运动以极高的时间分辨率（1$\mu s$）被捕捉，从而提升模型去除复杂真实世界模糊的能力。

此外，我们注意到，深度去模糊模型对不同类型的变换和扰动（例如随机旋转、尺度变换和高斯噪声）的鲁棒性有限。例如，对输入的模糊图像施加微弱的高斯噪声后，峰值信噪比会骤降至约 20 dB。因此，我们提出一种新颖的自监督鲁棒训练策略，显式对齐输入图像及其增强版本的去模糊输出（增强图像的去模糊输出在对齐前先进行逆增强），从而增强模型的鲁棒性。我们的贡献总结如下：

I. 我们提出了一种端到端卷积神经网络层级模型，通过利用多图块局部到粗尺度操作，在由细到粗的网格上执行去模糊。每个更细层级都以残差方式发挥作用，将其残差图像贡献给更粗层级，从而使网络的每个层级专注于不同的模糊尺度。为进行公平比较，我们在通用测试平台上进行了基线比较（在可能的情况下）。II. 我们发现了当前深度去模糊模型在堆叠深度方面的局限性，并引入了一种能够有效克服这一局限的新型堆叠方法。III. 我们在深度多图块架构中引入事件，以捕获更丰富的运动信息，从而帮助模型对具有复杂模糊和复杂场景的视频进行去模糊。IV. 我们提出利用前置增强任务施加辅助自监督一致性损失，以增强模型对于不同几何变换和光度失真的鲁棒性，从而减少对特定训练姿态的过拟合，进而帮助模型对真实世界图像进行去模糊。

我们的实验清楚地证明了事件引导的类 SPM 模型在非均匀运动去模糊方面的优势。据我们所知，我们的模型是首个采用多图像块方案进行盲运动去模糊的方法，例如，MPN 是首个支持实时（30fps）去模糊 720p图像的模型。我们也首次证明了自监督步骤在深度去模糊场景中的有效性。

## 2 相关工作

下面我们将讨论与图像去模糊相关的工作。传统图像去模糊方法 [4, 16, 57, 24,35, 17, 15, 40] 由于采用空间不变去模糊核，无法去除非均匀运动模糊。此外，其复杂的计算推理会导致处理时间过长，无法满足日益增长的实时去模糊需求。

**深度去模糊**。近年来，卷积神经网络已被应用于非均匀图像去模糊，以时间高效的方式处理复杂的运动模糊 [58, 48, 28, 39, 30, 46]。

Xu 等. [58] 提出了一种反卷积卷积神经网络，通过在给定估计模糊核的情况下恢复清晰图像，在非盲场景中去除模糊。其网络使用可分离卷积核，这些卷积核可分解为一组小型滤波器。Sun等。 [48] 通过学习 30×30 图像块与其对应卷积核之间的回归关系，估计并去除图像中的非均匀运动模糊。该方法采用传统基于能量的优化方案来估计潜在清晰图像。

Su 等人. [46] 提出了一种深度学习框架，用于处理模糊视频序列并累积跨帧信息。该方法不需要空间对齐样本对。Nah等人. [28] 利用多尺度卷积神经网络，以端到端方式从由各种因素导致模糊的图像中恢复清晰图像。他们采用多尺度损失来模拟传统去模糊方法中的由粗到细流程。

循环神经网络（RNN）常用于去模糊，因为它能够按序处理信息。以网络 [62] 为例，它由三个深度卷积神经网络和一个循环神经网络组成。该循环神经网络用作解码器，对第一个卷积神经网络模块提取的特征图进行反卷积。另一个卷积神经网络模块为循环神经网络的每一层学习权重。最后一个卷积神经网络模块重建清晰图像。尺度循环网络（SRN-DeblurNet） [49]使用卷积长短期记忆单元聚合从粗到细不同尺度的特征图。最后，Nah等人。 [29] 提出了一种循环网络，通过现有参数迭代更新隐藏状态。

生成对抗网络（GAN）也被用于去模糊，因为它们擅长保留纹理细节并生成照片级真实感图像。Kupyn等人。 [22] 提出了一种条件生成对抗网络，通过 Wasserstein损失生成高质量的去模糊图像。

Gao 等人提出了一种新颖的选择性参数共享方案，以改进动态去模糊任务。尽管该方法取得了令人印象深刻的结果，但复杂的嵌套连接导致处理运行时间过长，无法满足实时应用的需求。

值得注意的是，近期的一些工作基于我们的多图像块网络 [59]展开。Suin 等人 [47]提出了改进的MPN，能够处理不同空间位置的模糊变化，并自适应地处理测试图像以提升性能。Dipta 等人 [6]提出了一种快速的多图像块架构，用于解决图像去雾任务。

**基于事件的建模**。事件在近期的运动去模糊任务中发挥着重要作用。DAVIS [3] 和 DVS [25] 等事件相机以微秒尺度记录对数强度变化，且几乎不受运动模糊影响，因此能够补偿运动模糊造成的信息损失。事件相机的输出是由事件构成的数据流，每个事件以四元组（$x$，$y$，$t$，$p$）表示，用于编码亮度变化的位置、时间和极性。

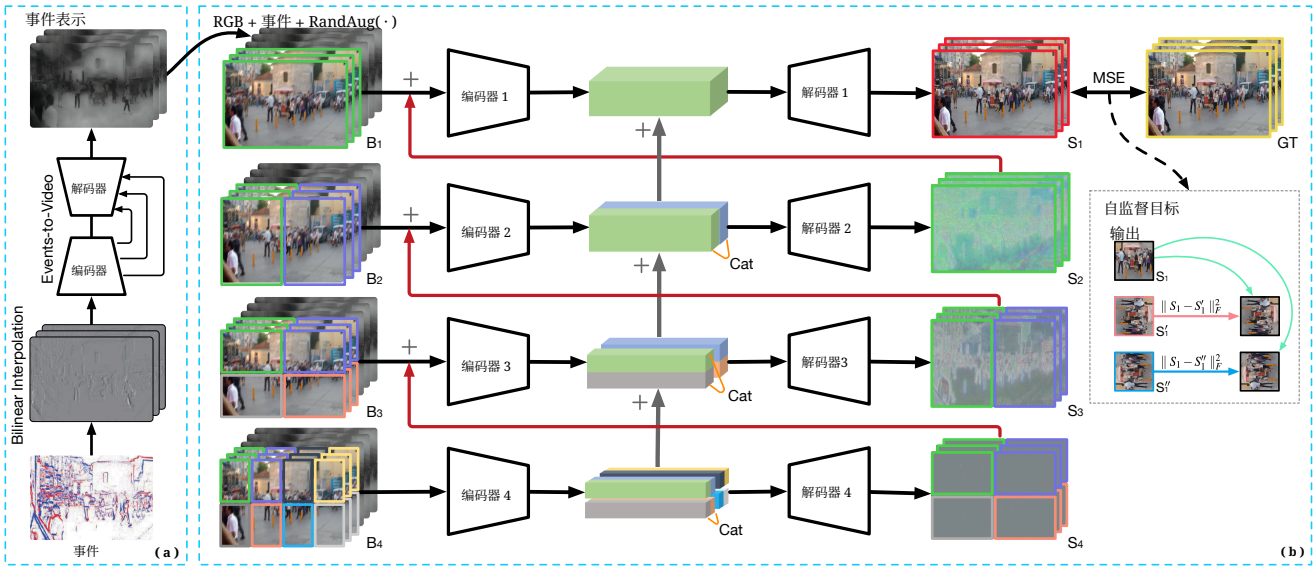



**图 3：** 我们提出的事件引导的多补丁网络（E-MPN）由两部分组成：（a）基于 10 层残差 U-Net 的事件表示生成器；（b）由多级由粗到细分支组成的多补丁去模糊网络。由于各图像块彼此不重叠，可能会产生边界伪影，而我们模型中连续的上层会将其消除。符号 + 是一种类似于残差网络的求和操作。

近期研究探索了如何将事件直接变换为清晰图像，例如，Bardow 等人 [1]通过求解能量最小化问题，利用固定长度的滑动时空窗口同时估计光流和强度图像。Barua等人 [2] 提出学习稀疏基于图像块的字典，以匹配事件图像块与梯度图像块，然后使用所谓的泊松积分重建强度图像。Munda等人 [27] 通过流形正则化恢复强度图像。Rebecq 等人 [37]提出了一种新型循环网络，用于从事件流中重建视频。由于事件是异步的，场景中出现局部强度变化时便会触发事件，因此单个事件可以表征静态场景/纹理，而事件序列则可以表征非常快速的运动。

DAVIS [3]可以同时输出事件和包含静态纹理的有源像素传感器（APS）强度图像。它直接在 APS 帧上累积事件，并刷新事件累积。Scheerlinck等人 [38]提出了一种异步事件驱动互补滤波器，将 APS 帧与事件结合起来进行连续时间强度估计。Pan 等人 [33] 将去模糊任务表述为一个优化问题，并利用双重积分模型求解单变量非凸问题。Jiang 等人 [18] 提出了一种卷积循环网络，在全局和局部尺度上融合视觉特征与时间知识。借助一种新颖的方向性事件滤波模块，该方法提取清晰边缘边界引导，从而提高了重建细节的质量。eSL-Net 模型 [52]构建了一种基于事件的稀疏学习网络，以改进去模糊性能。E-CIR [44] 提出利用事件构建参数基，并引入细化模块，在帧之间传播视觉特征。Wang等人 [55] 提出基于统一事件与帧不确定性模型，利用异步卡尔曼滤波器重建强度图像。然而，使用这些方法重建的图像会因事件噪声的累积而产生伪影。

**自监督学习。** 网络可以通过所谓的前置任务进行训练，例如，预测增强标签，或预测易于获取的自信息，将其作为辅助目标，通过使网络“意识到”辅助任务来提升性能。自监督学习已被用于物体识别 [8, 7, 13, 42, 51],视频表示学习 [9, 41, 10, 50, 54, 53]，以及少样本图像与视频识别 [12, 45, 61, 60]。

两种自监督方法较为常见：i）对比损失；ii）前置任务标签预测。例如，Gidaris 等人. [13] 预测随机图像旋转的标签，Doersch 等人. [7] 预测相对像素位置，Doso- vitskiy等人. [8] 学习区分一组代理类别，而相关方法 [12, 45] 通过预测图像旋转和拼图模式的标签来提升少样本性能。

与以往的自监督流程不同，我们利用自监督来促进不同数据增强下去模糊结果的一致性，从而提升模型对几何变换和光度失真的鲁棒性。本文的自监督策略会对分别在早期和后期阶段应用相同数据增强后获得的特征进行对齐。这与所谓的

知识蒸馏有些相似，知识蒸馏通过鼓励一个信息流向另一个信息流传递知识，从而提升模型的去模糊性能。然而，需要注意的是，这种自监督并非蒸馏流程，即，它没有使用教师网络和学生网络等双网络流。

## 3 方法

本文提出利用多图块层次结构，高效地进行盲运动去模糊。图 3 展示了我们提出的 MPN 网络的整体架构，其中以（1-2-4-8）模型（见第 3.2 节）为例。我们的网络受到由粗到细的空间金字塔匹配 [23], 的启发，该方法已用于场景识别问题 [20]，通过聚合多个图像块来提升性能。与图 [28, 49] 2 所示的多尺度和尺度递归网络模型需要进行高开销推理不同，我们的方法（同样见图 2）采用类残差架构，因此只需使用尺寸较小的滤波器，从而实现快速处理。尽管我们的模型架构非常简单（移除了跳跃连接和循环连接），但效果十分出色。与采用反卷积/上采样连接的 Nah 等人 [28] 不同，我们使用特征图拼接等操作，这得益于我们提出的多图块设置。此外，我们的自监督单元不同于典型的自监督表征：我们通过从去模糊输出中逆转增强操作，约束增强图像与未增强图像之间的去模糊一致性。

### 3.1 编码器-解码器架构

我们的 MPN 网络的每个层级都由一个编码器和一个解码器组成，其架构如图 4 所示。编码器由 15 个卷积层、6 条残差连接和 6 个 ReLU 单元组成。解码器和编码器的层完全相同，区别在于其中两个卷积层被反卷积层取代，以生成图像。

我们的编码器和解码器使用 ∼3.6 MB 参数。由于卷积核尺寸较小且模型具有残差结构，因此去模糊运行速度很快。相比之下，多尺度去模糊网络 [28]有 303.6 MB 参数，因此推理速度较慢。

### 3.2 网络架构

图 3 展示了我们的 MPN 架构，其中我们使用 (1-2-4-8) 模型进行说明。(1-2-4-8) 表示从最粗层级到最细层级的不重叠图像块数量，即，一个垂直在第二层级进行分割，$2\times 2=4$ 在第三层级进行分割，而 $2\times 4=8$ 在第四层级进行分割。对于输出多块残差的第三和第四层级，我们在相邻边界之间施加区域感知的一致性损失。

我们将初始模糊图像输入记为 $\mathbf{B}_1$，而$\mathbf{B}_{i,j}$ 表示第$i$层级中的第$j$个图像块。此外，$\mathscr{F}_i$ 和 $\mathscr{G}_i$分别表示第$i$层级的编码器和解码器，$\mathbf{C}_{i,j}$表示$\mathscr{G}_i$针对$\mathbf{B}_{i,j}$的输出，而，$\mathbf{S}_{i,j}$表示来自 $\mathscr{G}_i$ 的输出图像块。

我们的网络的每个层级都由一对编码器-解码器组成。每个层级的输入是将原始模糊图像输入$\mathbf{B}_1$ 划分为多个不重叠图像块后生成的。较低层级（对应更精细的网格）编码器的输出会与高一层级编码器的输出相加。较低层级（对应更精细的网格）解码器的输出会与传递到编码器输入端的上层输入网格相加，从而使顶层包含在更精细层级中推断出的所有信息。请注意，每个层级的输入和输出图像块数量并不相同，因为我们工作的核心思路是让较低层级专注于局部信息（更精细的网格），以便为较粗的网格（通过拼接卷积特征得到）生成残差信息。

以（1-2-4-8）变体为例。MPN 的去模糊过程从最底层级 4 开始。$\mathbf{B}_1$被切分为 8 个不重叠图像块$\mathbf{B}_{4,j}$，$j=1,\cdots,8$，并输入编码器 $\mathscr{F}_4$，以生成以下卷积特征表示：

$$\mathbf{C}_{4,j}=\mathscr{F}_4(\mathbf{B}_{4,j}),\quad j\in\{1,\cdots,8\}. \tag{1}$$

然后，我们在空间维度上拼接相邻的特征，得到一个新的特征表示$C^*_{4,j}$，其大小与第 3 层的卷积特征表示相同：

$$\mathbf{C}^*_{4,j}=\mathbf{C}_{4,2j-1}\oplus\mathbf{C}_{4,2j},\quad j\in\{1,\cdots,4\}, \tag{2}$$

其中，$\oplus$表示拼接算子。拼接后的特征表示$\mathbf{C}^*_{4,j}$经过编码器 $\mathscr{G}_4$ 处理，生成$\mathbf{S}4_{,j}=\mathscr{G}_4(\mathbf{C}^*_{4,j})$。

接下来，我们上移一级至第 3 层。 $\mathscr{F}_3$的输入由$\mathbf{S}4_{,j}$ 与图像块$\mathbf{B}3_{,j}$相加得到。生成 $\mathscr{F}_3$ 的输出后，我们将$\mathbf{C}^*_{4,j}$加到该输出上：

$$\mathbf{C}_{3,j}=\mathscr{F}_3(\mathbf{B}_{3,j}+\mathbf{S}_{4,j})+\mathbf{C}^*_{4,j},\quad j\in\{1,\cdots,4\}. \tag{3}$$

在第 3 层，我们拼接该层的特征表示，得到$\mathbf{C}^*_{3,j}$，并将其输入 $\mathscr{G}_3$，得到$\mathbf{S}3_{,j}$：

$$\mathbf{C}^*_{3,j}=\mathbf{C}_{3,2j-1}\oplus\mathbf{C}_{3,2j},\quad j\in\{1,2\}, \tag{4}$$

$$\mathbf{S}_{3,j}=\mathscr{G}_3(\mathbf{C}^*_{3,j}),\quad j\in\{1,2\}. \tag{5}$$

请注意，所有层级的特征都沿空间维度拼接：可以将相邻图像块拼接起来，形成一个更大的图像块。

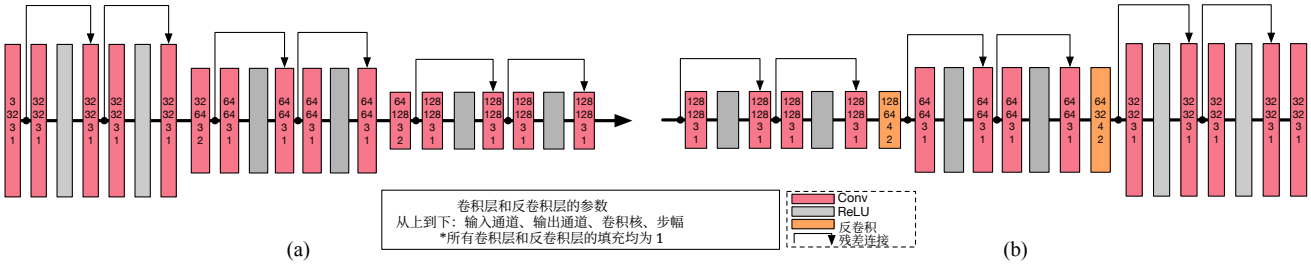



**图 4：** 我们（a）解码器和（b）编码器的架构及层配置。

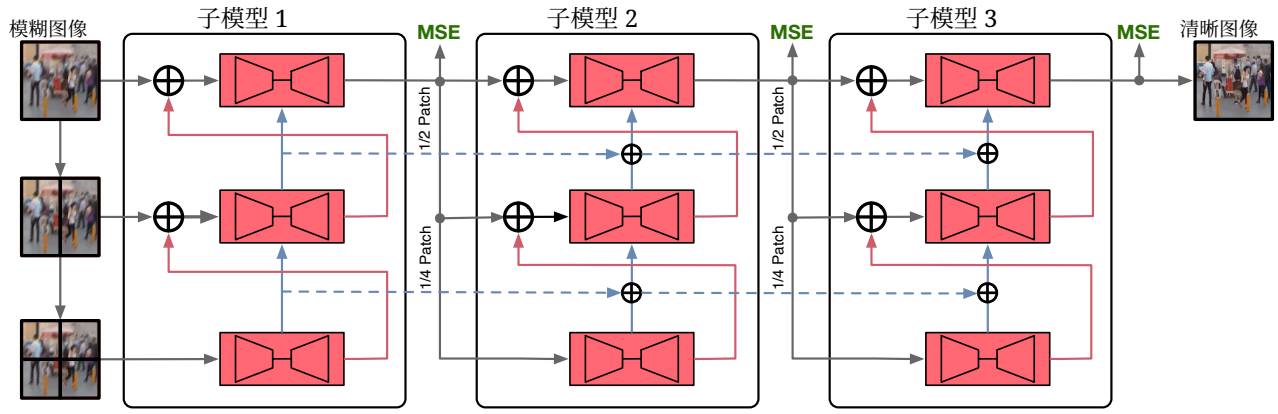



**图 5：** 堆叠式多块网络（StackMPN）的架构，其中多个 MPN 串行堆叠，将去模糊任务分配给多个子模型，并逐级降低训练难度。这种设计使网络深度增加时性能也得到提升。

在第 2 层，我们的网络将两个图像块 $\mathbf{B}_{2,1}$ 和$\mathbf{B}_{2,2}$作为输入。我们更新 $\mathbf{B}2,_j$，使得 $\mathbf{B}2,_j := \mathbf{B}2,_j + \mathbf{S}_{3,j}$并将其传递 $\mathscr{F}_2$:

$$\mathbf{C}_{2,j} = \mathscr{F}_2(\mathbf{B}_{2,j} + \mathbf{S}_{3,j}) + \mathbf{C}^*_{3,j}, \quad j \in \{1,2\}, \tag{6}$$

$$\mathbf{C}^*_2 = \mathbf{C}_{2,1} \oplus \mathbf{C}_{2,2}. \tag{7}$$

第 2 层的残差映射为:

$$\mathbf{S}_2 = \mathscr{G}_2(\mathbf{C}^*_2). \tag{8}$$

在第 1 层，最终的去模糊输出$S_1$ 为:

$$\mathbf{C}_1 = \mathscr{F}_1(\mathbf{B}_1 + \mathbf{S}_2) + \mathbf{C}^*_2, \tag{9}$$

$$\mathbf{S}_1 = \mathscr{G}_1(\mathbf{C}_1). \tag{10}$$

不同于 [28, 49] 在每一层计算均方误差（MSE）损失的方法，我们仅在第 1 层的输出处计算 MSE 损失（该层类似于残差网络）。MPN 的损失函数为:

$$\mathscr{L}_{deblur} = \frac{1}{2}\sum_j \|\mathbf{S}_{1j} - \mathbf{G}_j\|^2_F, \tag{11}$$

其中，$\mathbf{G}^j$ 表示真实清晰图像$j$。得益于分层多图块架构，我们的网络遵循残差学习原则：不同层级$\mathbf{S}^{ij}$ 的中间输出能够捕捉不同尺度的图像统计信息。因此，我们仅在第一个层级计算损失函数。我们研究了采用多层均方误差损失的效果，该损失会促使每个层级的输出接近真实图像。然而，正如预期的那样，采用多尺度均方误差损失并未带来明显的性能提升。

### 3.3 堆叠式多块网络

正如 Nah 等。 [28] 和 Tao 等。 [49],所报告的那样，增加更精细的网络层级无法提升多尺度和尺度递归架构的去模糊性能。对于我们的多图块网络，我们也观察到，将模糊图像划分为越来越小的网格并不能进一步提升去模糊性能。这主要是因为较粗层级能够在训练数据上快速获得较低的经验损失，从而排除最精细层级对其残差的贡献。

在本节中，我们提出了一种新颖的去模糊堆叠范式。我们不再纵向加深网络（向网络模型中添加更精细的层级，这会增加单个工作单元的难度），而是提出横向增加深度（堆叠多个网络模型），让多个 MPN 工作单元横向协同执行去模糊。

图 5展示了我们如何级联 MPN 来提升去模糊性能。这个堆叠模型称为 StackMPN，由多个“自底向上”的 MPN 堆叠而成。请注意，子模型 $i-1$ 的输出与子模型$i$ 的输入相连，这意味着在优化子模型 $i$时，需要用到子模型 $i-1$ 的输出。子模型 $i-1$ 的所有中间特征都会传递给子模型 $i$。通过最小化 StackMPN 目标函数，在每个子模型$i$ 的输出处计算均方误差损失：

$$\mathscr{L}_{deblur}=\frac{1}{2}\sum_{j}\sum_{i=1}^{N}\|\mathbf{S}_{ij}-\mathbf{G}_{j}\|_{F}^{2}, \tag{12}$$

其中 $N$ 表示所使用的子模型数量，$\mathbf{S}^{ij}$ 表示子模型 $i$ 的输出（注意，$\mathbf{S}^{ij}$ 在式 11 和 12 中的定义不同），$\mathbf{G}$ 表示真实清晰图像，而$j$ 遍历图像的一个子集。

我们的实验将表明，这种堆叠网络能够显著受益于网络深度的增加，并相应地提升去模糊性能。尽管我们的堆叠式流水线使用 MPN 单元，但我们认为这些单元具有通用性，也就是说，其他深度去模糊方法也可以采用类似方式进行堆叠，以提升其性能。然而，如果基本单元采用计算开销较大的去模糊模型，总处理时间可能难以接受。得益于快速高效的 MPN 单元，我们可以将堆叠网络的运行时间和大小控制在合理范围内，以满足各种应用需求。

### 3.4 事件引导的多块网络

上述提出的 MPN 模型应用于图像去模糊任务，因此其处理真实且复杂模糊模式的能力不足。为了进一步提升去模糊效果，可以将我们的模型扩展到视频，以便根据时间信息恢复模糊核。

一种简单的扩展 MPN 的方法是沿通道模式拼接多帧，形成输入（随后调整输入和输出通道的数量）。然而，这种模型并未充分利用时间信息：编码器第一层中的逐通道卷积算子无法确保模型学到足够隐式的运动模型。因此，我们提出将事件表示引入 MPN，构建一种混合事件引导去模糊过程，如图所示3。

具体而言，设$T$ 个连续的模糊帧为 $\{\mathbf{B}^{(t)}\}_{t=1}^{T}$，一组事件为 $\{\mathbf{E}^{(t)}\}_{t=1}^{T}$，$\Delta\mathbf{E}^{(t)}$ 表示从时刻$t$ 到$t+\Delta t$的事件信息，其中 $\Delta t$为无穷小时间步长。设 $\{\mathbf{S}^{(t)}\}_{t=1}^{T}$ 为恢复后的清晰帧。受事件相机设计的启发 [33]，我们有：

$$\mathbf{B}^{(f)}=\frac{1}{T}\sum_{t=f-T/2}^{f+T/2}\mathbf{S}^{(t)}, \tag{13}$$

$$\mathbf{S}^{(t)}=\mathbf{S}^{(f)}e^{c\mathbf{E}^{(t)}}, \tag{14}$$

$$\mathbf{E}^{(t)}=\sum_{h=f}^{t}\Delta\mathbf{E}^{(h)}, \tag{15}$$

其中$c$ 是用于确定是否记录事件的阈值，$\mathbf{E}^{(t)}$ 表示从时刻$f$ 到$t$期间事件的总和。

上述公式表明，模糊帧是在曝光时间 $[f-T/2、f+T/2]$期间由潜在清晰帧生成的，而时刻$t$的清晰帧则可以通过对时刻$t-T$的清晰帧与时间步长$T$内的事件进行简单插值来生成。结合公式13、14和15，可得：

$$\mathbf{B}^{(f)}=\frac{1}{T}\sum_{t=f-T/2}^{f+T/2}\mathbf{S}^{(t)}=\frac{\mathbf{S}^{(f)}}{T}\sum_{t=f-T/2}^{f+T/2}e^{c\sum_{h=f}^{t}\Delta\mathbf{E}^{(h)}}, \tag{16}$$

$$\log\mathbf{S}^{(f)}=\log\mathbf{B}^{(f)}-\log\left(\frac{1}{T}\sum_{t=f-T/2}^{f+T/2}e^{c\mathbf{E}^{(t)}}\right). \tag{17}$$

根据上述公式，我们得出清晰帧与原始模糊帧及事件信息相关联。一旦收集到从$f-T/2$到$f+T/2$ 的事件信息，MPN的学习形式即可改写为以下双流变体：

$$\mathbf{S}^{(f)}=\mathrm{MPN}(B^{(f)},\ \mathbf{E}^{(f-T/2:f+T/2)}). \tag{18}$$

我们在真实事件数据集 [33]上评估模型，其中的样本由DAVIS传感器记录。对于非事件数据集，我们采用 [36] 中提出的方法，从RGB帧模拟事件信息。

在将事件流输入训练和评估流水线之前，需要先将其转换为类似图像的事件表示。将事件整合到二维平面上是一种自然选择，而使用空间-时间体素网格对其进行编码可能更为有效。因此，我们采用基于10层残差U-Net构建的Events-to-Video模型 [37]，，以为中心帧生成10个相邻的事件表示，其时间网格为0.1$s$。随后，将这10个事件表示与原始RGB模糊帧拼接，以恢复清晰RGB输出。由于事件以极高的帧率采集，相较于传统相机，我们能够自然地捕捉到更加准确的物体运动。因此，我们的双流去模糊模型有望以端到端方式显著提升合成模糊图像和真实模糊图像的去模糊性能。

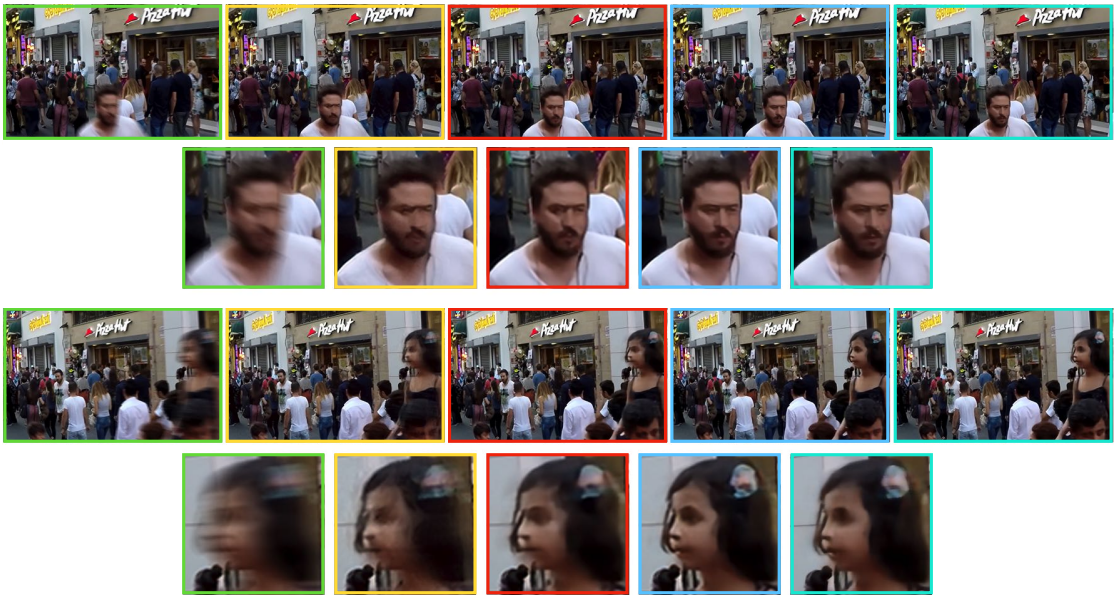


**图 6：** 去模糊结果。从左到右的图像分别展示了原始模糊图像、 [28], [49], 我们的 MPN 和经过 MPN + 自监督处理后的结果。如图所示，我们的方法能够生成最清晰、最逼真的面部细节。

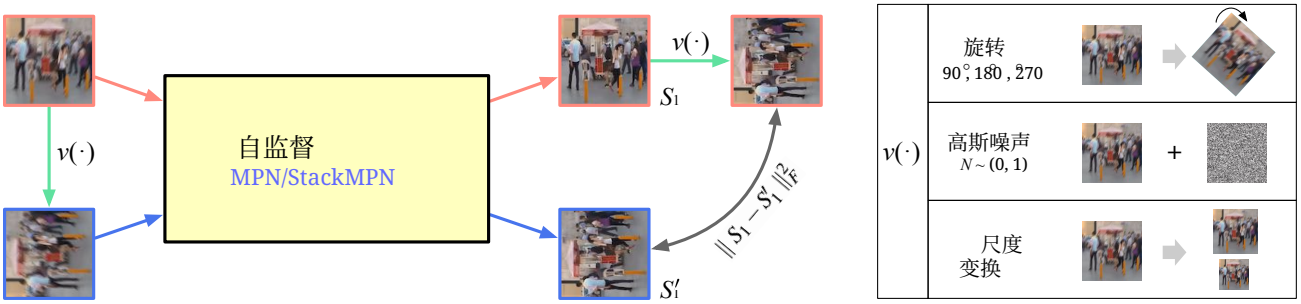



**图 7：** 自监督训练步骤。在原始图像的增强去模糊图像与其增强对应图像的去模糊输出之间应用一致性损失，从而提升模型在训练和推理过程中应对各种变换和噪声的鲁棒性。函数 $v(\cdot)$ 表示选定的数据增强。请注意，自监督损失不会对输出施加真实值约束，从而限制了过拟合。

### 3.5 通过自监督增强 MPN

下面我们介绍自监督，以提升我们MPN方法的性能和鲁棒性。图7 展示了我们流水线的自监督特性，我们将研究旋转、尺度变换和高斯噪声数据增强的影响。

**旋转。** 模糊核的一个自然属性是对旋转具有不变性，这增强了模型在现实场景中的鲁棒性和泛化能力。然而，即使采用了随机旋转增强，先前的方法 [28, 49, 62] 也无法保证具备这一属性。因此，我们提出以自监督的方式促进原始恢复输出与旋转后恢复输出之间的一致性，从而提升 MPN 对旋转的鲁棒性。

令$\mathbf{B}_j$为模糊图像，Deblur(·)为去模糊网络，例如*e.g*，MPN 或 StackMPN；Rot(·)为旋转函数——（随机选择旋转 $90°, 180°$ 或 $270°$）。自监督损失$L_{ss}$ 定义如下：

$$\mathbf{S}_{Nj} = \text{Deblur}(\mathbf{B}_j), \tag{19}$$

$$\mathbf{S}'_{Nj} = \text{Deblur}(v(\mathbf{B}_j)), \tag{20}$$

$$\mathscr{L}_{ss} = \sum_j ||v(\mathbf{S}_{Nj}) - \mathbf{S}'_{Nj}||_F^2, \tag{21}$$

其中，$j$ 遍历图像子集，$N$ 表示堆叠网络的层数，Deblur(·) 是堆叠网络第 $N$ 层的输出，而 $v(\cdot)$ 执行选定的增强操作，例如 *e.g.*,Rot(·)。

通过这种方式，去模糊网络能够接触到多种方向，并捕获旋转不变模糊核。加入基于旋转的自监督损失项后，MPN 的最终目标定义如下：

$$\mathscr{L} = \mathscr{L}_{deblur} + \alpha \mathscr{L}_{ss}, \tag{22}$$

**表1：** 我们模型在 GoPro 数据集上的定量分析 [28]。大小和运行时间以兆字节和毫秒为单位。报告的时间为卷积神经网络的运行时间（不包括将生成的图像写入磁盘所需的时间）。请注意，我们对 StackMPN 和 E-StackMPN 采用了 (1-2-4) 多图块架构。由于 GPU 内存限制和训练时间较长，我们未研究更深层的堆叠网络。

| 模型 | PSNR | SSIM | Size | 运行时间 |
|---|---|---|---|---|
| ♣ Sun等。 [48] | 24.64 | 0.843 | 54.1 | 12000 |
| ♣ Nah等。 [28] | 29.23 | 0.916 | 303.6 | 4300 |
| ♣ Zhang等人。 [62] | 29.19 | 0.931 | 37.1 | 1400 |
| ♣ Tao 等人。[49] | 30.10 | 0.932 | 33.6 | 1600 |
| ♣ Gao 等人。[11] | 30.92 | 0.942 | 2.8 | N/A |
| ♣ Park 等人。[34] | 31.15 | 0.945 | 2.6 | 70 |
| ♦ Nah等人。[29] | 29.97 | 0.895 | N/A | 34.7 |
| ♦ DBGAN [63] | 31.10 | 0.943 | 11.8 | N/A |
| ♠ Pan 等人。[33] | 29.06 | 0.943 | N/A | N/A |
| ♠ eSL-Net [52] | 30.23 | 0.870 | N/A | N/A |
| ♠ Jiang 等人。[18] | 31.79 | 0.949 | N/A | N/A |
| ♠ Pan 等人。[31] | 31.89 | 0.921 | 286 | 6500 |
| ♠ Xiang 等人。[56] | 32.63 | 0.935 | 61.8 | 6000 |
| （@ 表示使用的 RGB帧数量） | | | | |
| ♣ MPN | 30.21 | 0.935 | 21.7 | 17 |
| ♣ StackMPN | 31.16 | 0.945 | 65.1 | 233 |
| ♦ MPN @3 | 30.68 | 0.940 | | |
| ♦ MPN @5 | 30.89 | 0.941 | 21.7 | 17 |
| ♦ MPN @7 | 30.58 | 0.939 | | |
| ♦ StackMPN @5 | 31.63 | **0.951** | 65.1 | 233 |
| ♠ E-MPN | 33.14 | 0.937 | 64.7 | 121 |
| ♠ E-MPN @5 | 33.24 | 0.936 | 64.7 | 121 |
| ♠ E-StackMPN | 33.56 | 0.939 | 118.2 | 338 |
| ♠ E-StackMPN @5 | **33.83** | 0.941 | 118.2 | 338 |

♣ 单图像去模糊； ♦：视频去模糊； ♠：基于事件的去模糊。

其中 $\alpha \geq 0$ 是需要调整的超参数。

**尺度变换。** 模糊核的另一个公认特性是，它们在不同尺度输入下具有一致性。近期研究 [28, 62]利用多尺度输入进行图像去模糊。正如第 2 节所讨论的，与我们的多图块网络相比，这种设计带来的改进并不显著。理想情况下，可以用多尺度架构替换 MPN 的每一层，从而构建一个多尺度多补丁网络，在表 +3 中记作“MPN+MSN”，以便同时捕获来自不同尺度和位置的非均匀模糊核。然而，这种设计在网络参数、训练开销和推理时间方面代价极高，因此并不实用。

相反，我们采用一种与基于旋转的自监督步骤理念相似的自监督步骤，旨在促进不同尺度之间的一致性，并以高效的自监督方式捕获多尺度信息。与基于旋转的自监督不同的是，我们不使用一批旋转图像，而是随机对原始模糊图像进行下采样或上采样，并同时将原始输入和不同尺度的图像馈入 MPN，以获取相应输出。我们遵循公式 21 和 22，但仅将尺度增强函数 Scale(·) 替换 Rot(·)。

**表2：** 分层架构性能消融实验 .

| 模型 | PSNR | SSIM | Size | 运行时间 |
|---|---|---|---|---|
| MPN(1) | 28.70 | 0.9131 | 7.2 | 5 |
| MPN(1-2) | 29.77 | 0.9286 | 14.5 | 9 |
| MPN(1-1-1) | 28.11 | 0.9041 | 21.7 | 12 |
| MPN(1-2-4) | 30.21 | 0.9345 | 21.7 | 17 |
| MPN(1-4-16) | 29.15 | 0.9217 | 21.7 | 92 |
| MPN(1-2-4-8) | **30.25** | **0.9351** | 29.0 | 30 |
| MPN(1-2-4-8-16) | 29.87 | 0.9305 | 36.2 | 101 |

**高斯噪声。** 使去模糊模型具备抵抗从正态分布中抽取的加性像素噪声的鲁棒性，是一项理想属性。以往的工作 [28, 49, 62]，（包括我们的 MPN）都不具备应对高斯噪声的内在鲁棒性。表6 显示，一旦在模糊测试图像中注入数值较小的噪声，推理性能便会大幅下降。以往的工作会随机向模糊图像添加噪声，并将其用于训练。因此，模型需要同时处理去模糊和去噪。

## 4 实验

下面，我们将展示对 MPN 若干变体的实验评估。首先，介绍我们使用的数据集。

### 4.1 数据集

我们在 GoPro 数据集的多个版本 [28] 和 VideoDeblurring 数据集 [46],上训练并评估我们的方法，并对真实模糊图像 [33]进行定性分析，以直观比较各个模型的去模糊能力。最后，我们采集了一些真实的严重模糊图像，其中同时包含相机运动和物体运动，以进一步验证各方法在现实场景中的有效性。

**GoPro** 数据集 [28] 由从 33 段序列中提取的 3214 对模糊图像和清晰图像组成，分辨率为 720×1280。模糊图像通过对数量不等（7–13）的连续潜在帧取平均生成，从而产生不同程度的模糊。为公平比较，我们遵循 [28],中的方案，使用 2103 对图像进行训练，其余 1111 对用于测试。

**VideoDeblurring** 数据集 [46] 包含由多种设备拍摄的视频，例如 iPhone、GoPro 和 Nexus。定量部分包含 71 段视频。每段视频均由 100 帧组成，分辨率为 720×1280 。按照 [46], 中的设置，我们使用 61 段视频进行训练，其余 10 段用于测试。此外，我们还在 VideoDeblurring 数据集上评估使用 GoPro 数据集 [28]训练的模型，以展示我们方法的泛化能力。

**表3：** GoPro 数据集上多尺度和多块方法的基线性能 [28]。MSN 是一种遵循 [28] 的设计、使用我们的编码器和解码器的基线模型（详情请参阅第 4 节中的基线方法）。请注意，MSN(1) 和 MPN(1) 实际上是同一个模型。

| 模型 | PSNR | SSIM | 运行时间 |
|---|---|---|---|
| Nah等人。 [28] | 29.23 | 0.9162 | 4300 |
| MSN(1) | 28.70 | 0.9131 | 4 |
| MPN(1) | | | |
| MSN(2) | 28.82 | 0.9156 | 21 |
| MPN(1-2) | 29.77 | 0.9286 | 9 |
| MSN(3) | 28.97 | 0.9178 | 27 |
| MPN(1-2-4) | 30.21 | 0.9345 | 17 |
| MPN +MSN | 30.34 | 0.9351 | 523 |

## 4.2 评估设置与结果

我们将原始高分辨率 $720 \times 1280$ 像素图像输入 MPN。表 1 报告了峰值信噪比、结构相似性指数、模型大小和运行时间，以便与具有竞争力的先进运动去模糊模型进行深入比较。对于堆叠网络，考虑到运行时间和训练难度，我们在每个模型单元中采用 (1-2-4) 多块架构。

对于堆叠模型，每个子模型的输出都会按层级进行优化，这意味着第一个输出的质量最差，而最后一个输出的性能最佳。图8 展示了 Stack(3)-MPN（将 3 个子模型堆叠在一起）的输出，以说明每个子模型都逐步提升了去模糊效果。

## 4.3 实现细节

我们的所有实验均使用 PyTorch 实现，并在单块 NVIDIA Tesla P100 上进行评估。为了训练 MPN，




**图 8：** Stack(3)-MPN 中不同子模型的输出。从左到右依次为$\mathbf{M}_1$至$\mathbf{M}_3$的输出。结果的清晰度逐层提升。

我们随机裁剪图像至 $256 \times 256$像素大小。随后，我们从裁剪后的图像中提取图像块，并将其输入各个层级。训练时的批次大小设为 6。Adam 优化器 [19] 用于训练我们的模型，共训练 3000 轮。初始学习率设为 0.0001，衰减率设为 0.1。我们将图像归一化至 $[0, 1]$ 并减去 0.5。

**性能**。表1 显示，根据峰值信噪比和结构相似性指数指标，我们提出的 MPN 优于其他竞争方法，这证明了通过利用模型所用的局部信息去除非均匀模糊的优越性。由于 GPU 内存限制，我们训练和评估的最深 MPN 为（1-2-4-8-16）。采用（1-2-4-8）模型时性能最佳，其峰值信噪比和结构相似性指数均高于所有当前最先进的模型。值得注意的是，我们的模型比其他竞争方法更简单，例如，我们不使用循环单元。我们注意到，过小的图像块（小于 1/16 大小）对去除运动模糊没有帮助。

此外，堆叠变体 StackMPN 的峰值信噪比比更浅的 MPN 高出约 1.0 dB。结构相似性指数也呈现出相同的趋势。通过串行堆叠更多 MPN 单元，可以进一步提升 StackMPN 的性能，这与我们的预期一致。

对于视频去模糊，使用多帧输入不会显著影响运行时间，但能将峰值信噪比提升 0.41 dB。然而，由于我们没有研究处理多帧的先进策略，而这超出了本文的重点，因此仅通过使用更多帧数无法持续提升性能。使用 5 个模糊帧可实现最佳性能。

对于事件引导去模糊，我们观察到，图像去模糊和视频去模糊的性能均显著提升。为证明这一点，E-MPN 在 GoPro 数据集上达到了 33。14 dB，比 MPN 高出最多 ~2.9 dB。在视频去模糊任务中也观察到了类似趋势，这与我们的理论分析一致：将事件信息与模糊图像关联起来，作为管线的组合输入，能够捕捉准确的运动信息，帮助模型获得更好的去模糊性能。当我们在管线中用常见的 TV-L1 光流或预训练 FlowNet 的输出替代“事件表示”时，E-StackMPN 的结果从 32.57 dB 降至 32.01 dB/32.07 dB（仍优于 [31, 56]的结果），但不及采用“事件表示”的 E-StackMPN。尽管事件信息可以借助事件相机等额外硬件获取，但我们实验中 GoPro 数据集所用的事件信息完全由 Esim 根据 RGB帧模拟生成 [36], ; Esim 包含的是渲染引擎，而非用于真实值标注的事件相机。如果可以使用事件相机，定量性能

**表4：** 在 VideoDeblurring 数据集上进行的定量分析（峰值信噪比） [46] 针对在 GoPro 数据集上训练的模型。PSDeblur 表示使用 Photoshop CC 2015。为公平比较，我们选择了该方法的“单帧”版本 [46] 。

| 方法 | #1 | #2 | #3 | #4 | #5 | #6 | #7 | #8 | #9 | #10 | 平均值 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| ♣PSDeblur [46] | 24.42 | 28.77 | 25.15 | 27.77 | 22.02 | 25.74 | 26.11 | 19.75 | 26.48 | 24.62 | 25.08 |
| ♣WFA [5] | 25.89 | 32.33 | 28.97 | 28.36 | 23.99 | 31.09 | 28.58 | 24.78 | 31.30 | 28.20 | 28.35 |
| ♣Su等。 [46] | 25.75 | 31.15 | 29.30 | 28.38 | 23.63 | 30.70 | 29.23 | 25.62 | 31.92 | 28.06 | 28.37 |
| ♦Nah等。 [29] | - | - | - | - | - | - | - | - | - | - | 30.80 |
| ♦STFAN [64] | - | - | - | - | - | - | - | - | - | - | 31.24 |
| ♦Xiang等。 [56] | - | - | - | - | - | - | - | - | - | - | 31.68 |
| ♦Pan等。 [31] | - | - | - | - | - | - | - | - | - | - | 31.67 |
| ♣MPN | 29.89 | 33.35 | 31.82 | 31.32 | 26.35 | 32.49 | 30.51 | 27.11 | 34.77 | 30.02 | 30.76 |
| ♣StackMPN | 30.48 | 34.31 | 32.24 | 32.09 | 26.77 | 33.08 | 30.84 | 27.51 | 35.24 | 30.57 | 31.39 |
| ♠E-MPN | 31.32 | 34.73 | 33.21 | 32.68 | 27.85 | 33.81 | 31.83 | 28.46 | 36.09 | 31.45 | 32.14 |
| ♠E-StackMPN | **31.68** | **35.81** | **33.35** | **33.17** | **28.32** | **34.17** | **32.15** | **29.01** | **36.27** | **31.75** | **32.57** |

♣：单图像去模糊；♦：视频去模糊；♠：基于事件的去模糊。

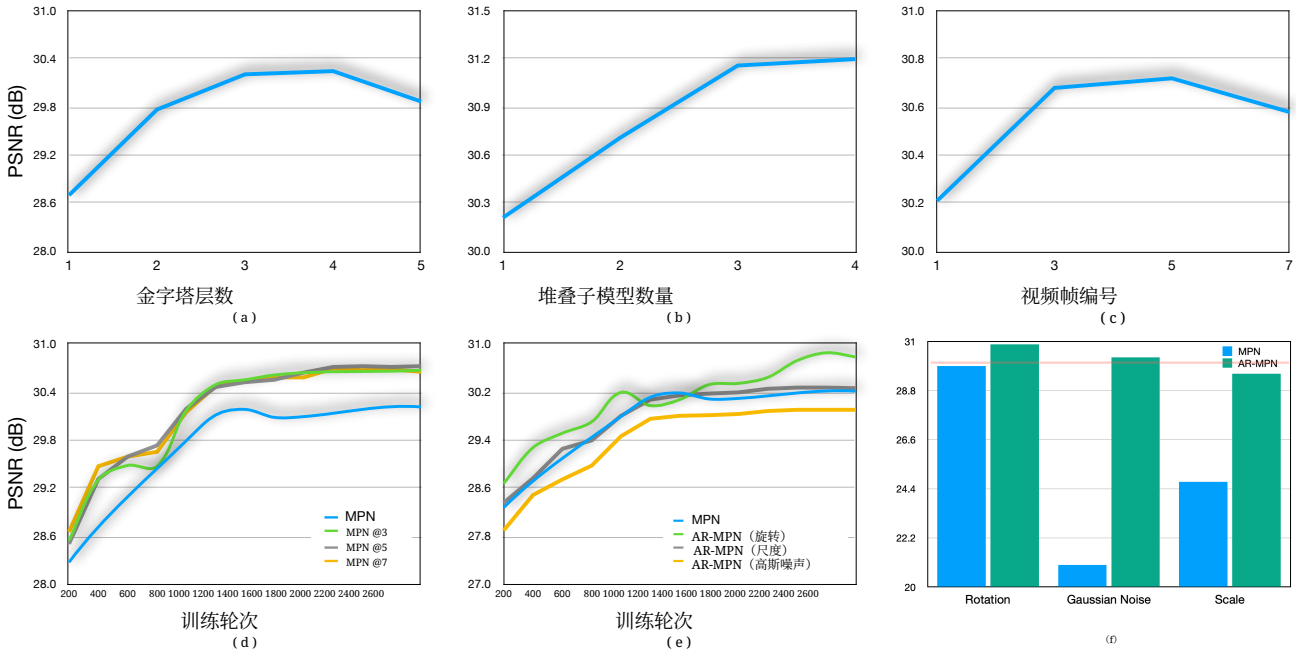



**图 9：** 消融研究。峰值信噪比与（*a*）金字塔层级数、（*b*）堆叠单元数、（*c*）拼接帧数的关系。此外，我们还展示了峰值信噪比与（*d*）MPN 使用的视频帧数（以 @ 标示）的关系，以及（*e*）不同训练轮次数下的数据增强类型。最后，（*f*）比较了单独使用 MPN与采用不同自监督策略的 MPN 的最终性能。

**表 5：** GoPro 上权重共享方案的评估 [28]。

| 模型 | PSNR | SSIM | 大小（兆字节） |
|---|---|---|---|
| MPN(1-2) | 29.77 | 0.9286 | 14.5 |
| MPN(1-2)-WS | 29.22 | 0.9210 | 7.2 |
| MPN(1-2-4) | 30.21 | 0.9343 | 21.7 |
| MPN(1-2-4)-WS | 29.56 | 0.9257 | 7.2 |
| MPN(1-2-4-8) | 30.25 | 0.9351 | 29.0 |
| MPN(1-2-4-8)-WS | 30.04 | 0.9318 | 7.2 |

在这种情况下，由于事件信息的质量最高，我们的 E-MPN 性能应会进一步提升。

使用光流的另一个缺点是，它编码的是位移信息，而非变化信息，等等。事件模型能够应对快速运动

**表6：** 在 GoPro [28] 上，采用不同变换/噪声时，MPN 与采用自监督的 MPN 模型的鲁棒性消融研究。通过在测试图像上随机应用变换或噪声并测量峰值信噪比，评估模型对数据增强的鲁棒性。

| 增强类型 | None | Rot. | 高斯 | 尺度 |
|---|---|---|---|---|
| MPN | 30.24 | 29.89 | 20.98 | 24.71 |
| MPN + 随机增强 | 30.01 | 29.90 | 28.05 | 28.31 |
| MPN + 自监督（旋转） | 30.85 | 30.86 | - | - |
| MPN + 自监督（高斯） | 30.46 | - | 30.31 | - |
| MPN + 自监督（尺度） | 30.35 | - | - | 29.49 |
| MPN + 自监督（混合） | 30.92 | 30.91 | 30.35 | 29.91 |

这是设计使然，而光流算法众所周知在大位移情况下会失效。

GoPro 数据集上的去模糊图像如图所示6、12 和9。具体而言，图6 展示了去模糊后的

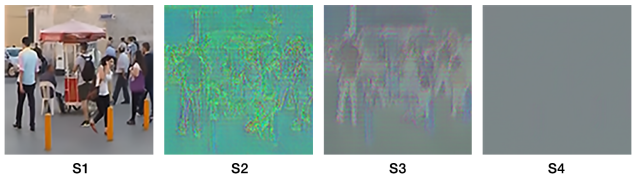



**图10：** 不同 MPN（1-2-4-8）层级的输出$\mathbf{S}_i$。从右到左的图像分别可视化底层$\mathbf{S}_4$到顶层$\mathbf{S}_1$。

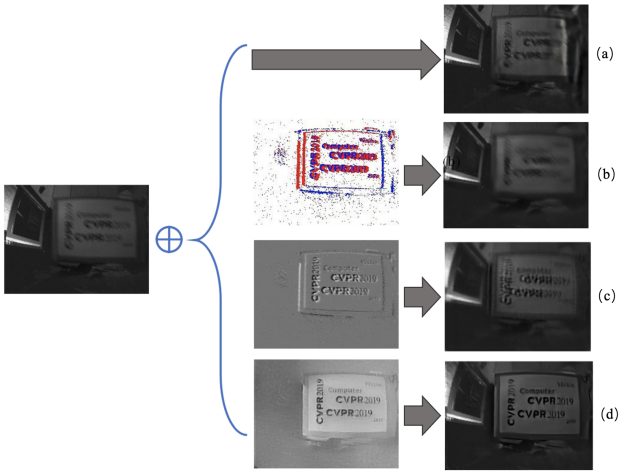



**图11：** 将事件注入我们的 MPN 的不同策略：(a) 直接对原始模糊输入进行去模糊；(b) 对原始模糊输入与沿时间维度累积的事件相结合的输入进行去模糊；(c) 对“原始模糊图像 + 事件体素”进行去模糊；以及 (d) 使用原始模糊图像与 Event-to-Video 表示拼接的去模糊模型。

几种模型在包含严重运动模糊的图像上的去模糊性能。为清晰起见，我们放大了主要物体。图12 展示了不同场景中的精选图像，以说明我们模型的优势：在所有案例中，该模型都能呈现最清晰的细节。此外，我们在图13 中展示了模型对真实模糊图像的去模糊性能，以说明 E-MPN 的优势。在这种情况下，E-MPN 明显优于以往的深度模型。图14 对由相机运动和物体运动造成的真实严重模糊图像进行了性能比较。与基线模型相比，我们的流程实现了更好的去模糊效果。

**运行时间**。除了我们的模型具有优异的峰值信噪比和结构相似性指数外，据我们所知，MPN也是首个能够实时运行的深度去模糊模型。例如，MPN (1-2-4-8)处理一张 720×1280 图像需要30毫秒，这意味着它支持以30帧/秒的速度进行实时720p图像去模糊。然而，I/O操作会带来运行时间开销，因此实时去模糊应用需要在视频采集器和GPU之间进行快速传输，并配备更大的GPU内存和/或SSD驱动器，等等。

以下因素有助于实现较快的运行时间：i）采用较浅的编码器-解码器结构和较小尺寸的卷积滤波器；ii）移除不必要的连接，例如跳跃连接或循环连接；iii）减少不同层级卷积特征之间的上采样/反卷积次数。

**基线比较**。 尽管我们的模型性能远优于多尺度模型 [28],，但这在一定程度上是不公平的比较，因为我们提出的模型与 [28] 其网络架构存在显著差异。与[28], 使用超过 303.6 兆字节参数的模型相比，我们采用了更浅的 CNN编码器和解码器，使模型大小10×更小。因此，我们构建了一个深层多尺度网络（MSN），采用我们的编码器-解码器，其设置遵循 [28]，用于在多块方法与多尺度方法之间进行基线比较（合理性检查）。如表 3所示，MSN 的峰值信噪比低于 [28],，这符合预期，因为我们简化了 CNN 架构。与我们的 MPN 相比，MSN 获得的最佳结果低于 MPN（1-2）模型。由于采用了通用测试平台，MSN 与 MPN 报告的性能能够公平比较多块层级模型与多尺度模型 [28]。

### 4.4 消融研究

我们在图 10中可视化了 MPN 单元的输出，以分析其中间层的贡献。如前文所述，MPN 采用残差设计。因此，与较粗层级相比，较细层级包含的信息更精细，但在视觉上不那么重要。在图 10中，我们展示了 MPN 各层级（1-2-4-8）的输出 $\mathbf{S}_i$。其中，$\mathbf{S}_4$ 所包含的信息最为精细且最为稀疏。随着我们在 MPN 中逐级向上移动，输出变得不那么稀疏、更清晰，且色彩更加丰富。

**各层级之间的权重共享**。 下面，我们研究在网络所有层级的编码器-解码器对之间共享权重，以减少参数数量。表5显示，权重共享会导致性能略有下降，但能显著减少参数数量。

**事件预处理的必要性**。 我们的网络采用所谓的“事件到视频”模拟网络，以改善事件信息与 RGB帧的聚合。为证明其必要性，我们将 E-MPN 与两个近期的深度事件引导去模糊模型进行比较 [52, 44], ，并发现，与其他模型相比，添加前端网络对事件体素进行预处理，能够更有效地提取事件中的运动信息。作为替代方案，需要后端模块来弥补性能损失，但在使用事件的背景下，这类模块并不容易解释。

**注入事件的各种策略**。为了证明在我们的 MPN 流程中使用 Event-to-Video 网络的必要性，




**图12：** GoPro 和 VideoDeblurring 数据集中的模糊图像的去模糊性能。我们裁剪出蓝色和红色边界框所示的区域，并按如下方式获得更小的窗口。第一列包含原始模糊裁剪图像，第二列是 [28], 第三列是 [49]的结果。我们的结果显示在最后一列中。我们认为，在多个不同场景中，我们的模型均取得了最佳的视觉效果。

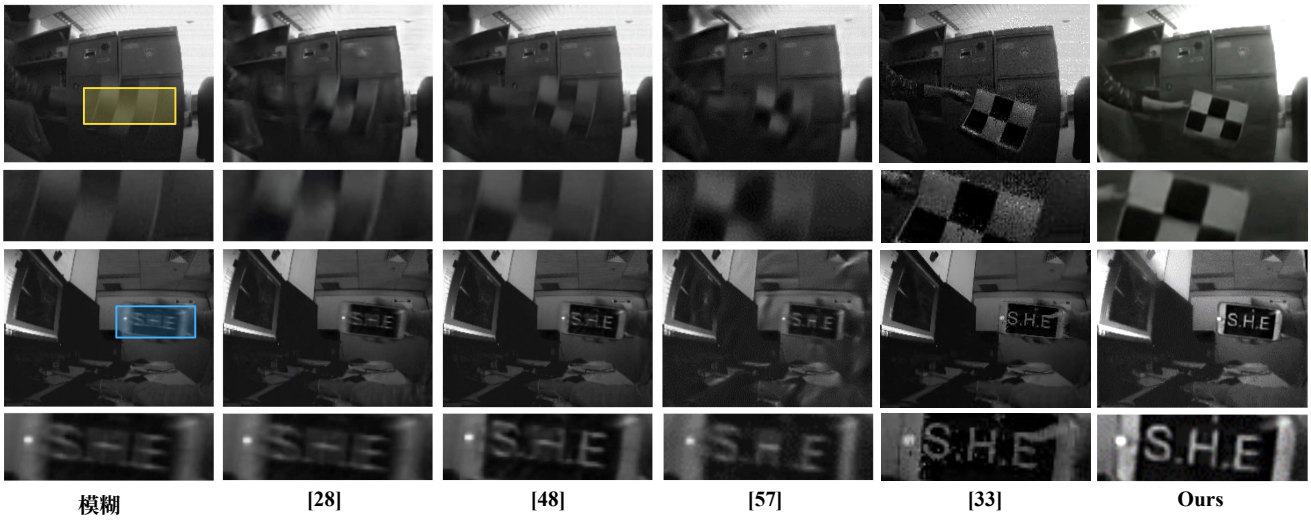



**图13：** 模糊图像去模糊性能的定性比较 [33]。第一列为原始模糊图像，第二列采用方法 [28], 第三列采用方法 [49], 第四列采用方法 [33]。我们的结果显示在最后两列中。前三个深度去模糊模型（从左到右）在处理复杂模糊时效果不佳，而采用事件表示 E-MPN 则能生成清晰的图像。

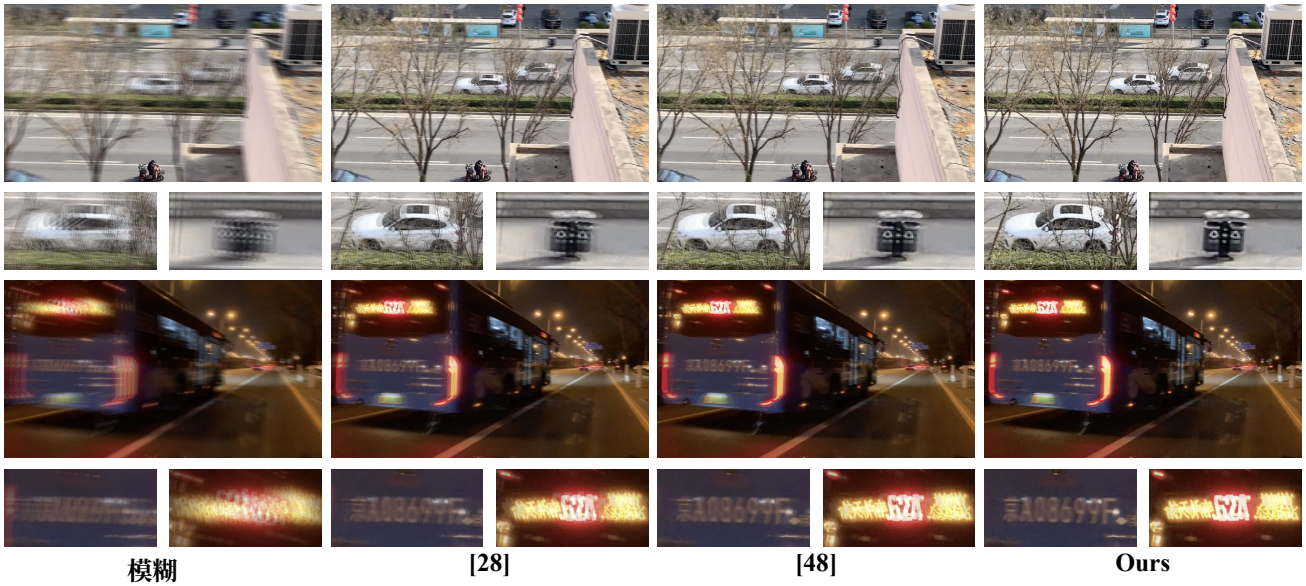



**图14：** 真实模糊图像去模糊性能的定性比较。第一列为原始模糊图像。第二列由方法 [28], 得到，第三列由方法 [49] 得到。最后一列展示了我们的结果。

我们针对不同的事件注入策略进行了消融研究。图 11 显示，最后一种模型（将原始模糊图像与事件到视频表示拼接）在定性比较方面显著优于其他基线。我们认为，这是因为 Event-to-Videos 网络将事件解码为图像域中的底层事件/运动动态表示，而非事件域中的表示。期望去模糊流程能够自行学习解码事件是不合理的，因此我们采用专门的 Event-to-Videos 网络。

**自监督流程**。如表6所示，将我们提出的自监督步骤与 MPN 结合，可提升模型对不同几何变换和光度噪声的鲁棒性。例如，测试时对原始模糊图像添加低强度加性高斯噪声，会使原始模型的去模糊性能降至 20.98 dB，这表明原始模型无法对含噪图像进行去模糊。应用高斯噪声自监督步骤后，去模糊性能达到 30.31 dB。在 MPN 中将旋转作为自监督任务，可使 GoPro 数据集上的峰值信噪比提升约 0.6 dB。

## 5 结论

本文利用多图块模型，而非广泛使用的多尺度和尺度递归架构，解决了极具挑战性的非均匀运动去模糊问题。为此，我们设计了一种端到端的深度多图块层次去模糊网络。与现有的深度去模糊框架相比，我们的模型取得了最先进的性能（以峰值信噪比和结构相似性指数衡量），并且能够以 30 fps 的速度处理 720p 图像。为克服相邻图像块边界之间的差异，我们显式地最小化这些边界之间的 $\ell_2$ 度量，以促进图像块的全局一致性。我们的堆叠变体 StackMPN 相较于较浅的 MPN 和竞争方法进一步提升了结果，同时比后者 ∼4×更快。我们的堆叠架构似乎克服了其他竞争方法所面临的堆叠深度限制。此外，我们提出的新型自监督机制增强了模型应对几何变换和光度噪声的能力。最后，将摄像机事件表示与模糊图像结合使用，使包含复杂模糊模式的帧取得了最大幅度的提升。我们希望本研究能为后续去模糊工作提供若干有价值的见解。

**致谢**。 本研究得到了中国国家自然科学基金（项目编号：62106282）的资助。

代码: https://github.com/HongguangZhang/DMPHN-cvpr19-master。

**数据可用性声明**：本文使用和研究的所有数据集均已公开。

## 参考文献

1. Bardow, P., Davison, A.J., Leutenegger, S.：基于事件相机的光流与强度同步估计。载于：IEEE 计算机视觉与模式识别会议论文集，第 8844 2892 2016–() . Barua, S., Miyatani, Y., Veeraraghavan, A.：基于事件相机的直接人脸检测与视频重建。载于：2016 年 IEEE 冬季计算机视觉应用会议（WACV），第 14 3 9 2016 – . IEEE ( ) . Brandli, C., Berner, R., Yang, M., Liu, S.C., Delbruck, T.：一款 240×180 130 db 3 $\mu$s 延迟全局快门时空视觉传感器。《IEEE 固态电路期刊》**49**(10),2333–2341(20143,4 4 ) . Cho, S., Lee, S.：快速运动去模糊。《ACM 图形学汇刊》**283** 5 5 145 1 145 8 2009 ( ), : – : ( ) . Delbracio, M., Sapiro, G.：通过高效傅里叶聚合进行手持视频去模糊。《IEEE 计算成像汇刊》**1**(4), 270–283 (2015)。数字对象标识符 10.1109/TCI.2015.2501245116. Dipta Das, S., Dutta, S.：用于非均匀图像去雾的快速深度多块分层网络。载于：IEEE/CVF 计算机视觉与模式识别会议研讨会论文集，第 482–483 页（2020）3

7. Doersch, C., Gupta, A., Efros, A.A.：通过上下文预测进行无监督视觉表征学习。载于：IEEE国际计算机视觉会议论文集，第1422–1430页 4 8 2015( )。Dosovitskiy, A., Springenberg, J.T., Riedmiller, M., Brox, T.：利用卷积神经网络进行判别式无监督特征学习。载于：神经信息处理系统进展，第766 4 9 774页 2014– ( )。Fernando, B., Bilen, H., Gavves, E., Gould, S.：利用异类网络进行自监督视频表示学习。载于：IEEE计算机视觉与模式识别会议论文集，第3636 4 10 3645页 2017– ( )。Gan, C., Gong, B., Liu, K., Su, H., Guibas, L.J.：用于自监督视频表征学习的几何引导卷积神经网络。载于：IEEE计算机视觉与模式识别会议论文集，第5589 4 11 5597页 2018– ( )。Gao, H., Tao, X., Shen, X., Jia, J.：采用参数选择性共享和嵌套跳跃连接的动态场景去模糊。载于：IEEE计算机视觉与模式识别会议（CVPR）（2019 3, 9 12）。Gidaris, S., Bursuc, A., Komodakis, N., P´erez, P., Cord, M.：通过自监督提升少样本视觉学习。arXiv预印本 arXiv:1906 4 13 05186 2019. ( )。Gidaris, S., Singh, P., Komodakis, N.：通过预测图像旋转进行无监督表征学习。arXiv预印本 arXiv:1803 4 14 07728 2018. ( )。He, K., Zhang, X., Ren, S., Sun, J.：深度卷积网络中用于视觉识别的空间金字塔池化。载于：欧洲计算机视觉会议论文集，第346 2 15 361页 2014 –。Springer ( )。Hyun Kim, T., Mu Lee, K.：动态场景的广义视频去模糊。载于：IEEE计算机视觉与模式识别会议论文集，第5426 3 16 5434页 2015– ( )。Jia, J.：利用透明度进行单图像运动去模糊。载于：IEEE计算机视觉与模式识别会议论文集，第1–8页。IEEE (2007) 3 17。Jia, J.：均匀运动去模糊的数学模型与实用求解器。(2014 3 18)。Jiang, Z., Zhang, Y., Zou, D., Ren, J., Lv, J., Liu, Y.：基于事件的运动去模糊学习。载于：IEEE/CVF计算机视觉与模式识别会议论文集（CVPR）（2020 4, 9 19）。Kingma, D.P., Ba, J.：Adam：一种随机优化方法。arXiv预印本 arXiv:1412 10 20 6980 2014. ( )。Koniusz, P., Zhang, H., Porikli, F.：深入探究幂归一化。载于：IEEE计算机视觉与模式识别会议论文集，第5774 2, 5 21 5783页 2018– ( )。Krizhevsky, A., Sutskever, I., Hinton, G.E.：利用深度卷积神经网络进行ImageNet分类。载于：神经信息处理系统进展论文集，第1097 1 22 1105页 2012– ( )。Kupyn, O., Budzan, V., Mykhailych, M., Mishkin, D., Matas, J.：DeblurGAN：利用条件对抗网络进行盲运动去模糊。arXiv预印本 arXiv:1711 3 23 07064 2017. ( )。Lazebnik, S., Schmid, C., Ponce, J.：超越特征袋：用于识别自然场景类别的空间金字塔匹配。载于：IEEE计算机视觉与模式识别会议论文集，第2169–2178页。IEEE (2006 2, 5 24)。Levin, A.：利用图像统计信息进行盲运动去模糊。载于：神经信息处理系统进展论文集，第

841 848 2007 3 – ( )。25 128× 128 120 Lichtsteiner, P., Posch, C., Delbruck, T.：一种延迟为15$\mu$秒的异步时域对比视觉传感器。IEEE固态电路期刊43 3 26 2 566 576 2008 ( )， – ( )。Lu, X., Lin, Z., Shen, X., Mech, R., Wang, J.Z.：用于图像风格、美学和质量评估的深度多块聚合网络。载于：IEEE计算机视觉与模式识别会议论文集，第990–998页 (2015) 2

27. Munda, G., Reinbacher, C., Pock, T.：利用流形正则化实时重建事件相机的强度图像。《国际计算机视觉杂志》126(12), 1381–1393 (2018 4 28)。Nah, S., Kim, T.H., Lee, K.M.：用于动态场景去模糊的深度多尺度卷积神经网络。载于：IEEE计算机视觉与模式识别会议论文集，第257–265页 (2017) 1, 2, 3, 5, 6, 8, 9, 10, 11, 12, 13, 14 29。Nah, S., Son, S., Lee, K.M.：利用帧内迭代进行视频去模糊的循环神经网络。载于：IEEE计算机视觉与模式识别会议（CVPR）（2019）3, 9, 11 30。Nimisha, T.M., Singh, A.K., Rajagopalan, A.N.：用于盲去模糊的模糊不变深度学习。载于：IEEE国际计算机视觉会议论文集，第4762 3 31 4770页 2017– ( )。Pan, J., Bai, H., Tang, J.：利用时域清晰度先验进行级联深度视频去模糊。载于：IEEE/CVF计算机视觉与模式识别会议论文集，第3043–3051页 9, 10, 11 32 2020( )。Pan, L., Dai, Y., Liu, M., Porikli, F.：同时进行立体视频去模糊和场景流估计。载于：IEEE计算机视觉与模式识别会议论文集，第6987 1 33 6996页 2017 –。IEEE ( )。Pan, L., Scheerlinck, C., Yu, X., Hartley, R., Liu, M., Dai, Y.：利用事件相机以高帧率让模糊帧重现清晰。载于：IEEE/CVF计算机视觉与模式识别会议论文集，第6820–6829页 (2019) 4, 7, 9, 14 34。Park, D., Kang, D.U., Kim, J., Chun, S.Y.：用于渐进式非均匀单图像去模糊的多时域循环神经网络及增量式时域训练。载于：欧洲计算机视觉会议论文集，第327 9 35 343页 2020 –。Springer ( )。Rajagopalan, A.N., Chellappa, R.：运动去模糊：算法与系统。剑桥大学出版社 (2014 3 36)。Rebecq, H., Gehrig, D., Scaramuzza, D.：Esim：一种开放的事件相机模拟器。载于：机器人学习会议论文集，第969–982页 7, 10 37 2018。PMLR ( )。Rebecq, H., Ranftl, R., Koltun, V., Scaramuzza, D.：从事件到视频：将现代计算机视觉引入事件相机。载于：IEEE/CVF计算机视觉与模式识别会议论文集，第3857 4, 7 38 3866页 2019– ( )。Scheerlinck, C., Barnes, N., Mahony, R.：利用事件相机进行连续时间强度估计。载于：亚洲计算机视觉会议论文集，第308 4 39 324页 2018 –。Springer ( )。Schuler, C., Hirsch, M., Harmeling, S., Scholkopf, B.：学习去模糊。IEEE模式分析与机器智能汇刊 (7), 1439–1451 (2016 3 40)。Sellent, A., Rother, C., Roth, S.：立体视频去模糊。载于：欧洲计算机视觉会议论文集，第558 3 41 575页 2016 –。Springer ( )。Sermanet, P., Lynch, C., Chebotar, Y., Hsu, J., Jang, E., Schaal, S., Levine, S.：时间对比网络：从像素进行自监督学习 (2017 4 42)。Simon, C., Koniusz, P., Harandi, M.：关于为增量学习学习测地路径。载于：IEEE/CVF计算机视觉与模式识别会议论文集（2021 4 43）。Simonyan, K., Zisserman, A.：用于大规模图像识别的超深卷积网络。arXiv预印本 arXiv:1409.1556 (2014 1 44)。Song, C., Huang, Q., Bajaj, C.：E-CIR：增强事件的连续强度恢复。载于：IEEE/CVF计算机视觉与模式识别会议论文集，第7803–7812页 (2022 4, 12 45)。Su, J.C., Maji, S., Hariharan, B.：利用自监督提升少样本学习的监督效果。arXiv预印本 arXiv:1906 4 46 07079 2019. ( )。Su, S., Delbracio, M., Wang, J., Sapiro, G., Heidrich, W., Wang, O.：手持相机的深度视频去模糊。载于：IEEE

计算机视觉与模式识别会议（2017，3、9、11、47）。Suin, M., Purohit, K., Rajagopalan, A.：用于自适应运动去模糊的空间注意力图像块层次网络。载于：IEEE/CVF计算机视觉与模式识别会议论文集，第3606–3615页（2020，3、48）。Sun, J., Cao, W., Xu, Z., Ponce, J.：学习用于非均匀运动模糊去除的卷积神经网络。载于：IEEE计算机视觉与模式识别会议论文集，第769–777页（2015，2、3、9、49）。Tao, X., Gao, H., Shen, X., Wang, J., Jia, J.：用于深度图像去模糊的尺度递归网络。载于：IEEE计算机视觉与模式识别会议论文集，第8174–8182页（2018，1、2、3、5、6、8、9、13、14、50）。Tas, Y., Koniusz, P.：通过核特征映射，在三维人体骨架上进行基于CNN的动作识别和有监督域适应。载于：英国机器视觉会议（BMVC）（2018，4、51）。Tas, Y., Koniusz, P.：基于AUDITED的简易对话系统。载于：英国机器视觉会议（BMVC）（2021，4、52）。Wang, B., He, J., Yu, L., Xia, G.S., Yang, W.：事件增强高质量图像恢复。载于：欧洲计算机视觉会议论文集，第155–171页。Springer（2020，4、9、12、53）。Wang, L., Koniusz, P.：通过统计矩和子空间描述子进行自监督动作识别。载于：ACM国际多媒体会议（ACM MM）（2021）。数字对象标识符：10.1145/3474085.3475572（4、54）。Wang, L., Koniusz, P., Huynh, D.Q.：为基于CNN的动作识别生成幻觉式IDT描述子和I3D光流特征。载于：IEEE国际计算机视觉会议（ICCV）（2019，4、55）。Wang, Z., Ng, Y., Scheerlinck, C., Mahony, R.：用于混合事件相机的异步卡尔曼滤波器。载于：IEEE/CVF国际计算机视觉会议论文集，第448–457页（2021，4、56）。Xiang, X., Wei, H., Pan, J.：利用样本中的清晰度特征进行深度视频去模糊。《IEEE图像处理汇刊》29卷，第8976–8987页（2020，9、10、11、57）。Xu, L., Jia, J.：用于稳健运动去模糊的两阶段卷积核估计。载于：欧洲计算机视觉会议论文集，第157–170页。Springer（2010，3、58）。Xu, L., Ren, J.S., Liu, C., Jia, J.：用于图像反卷积的深度卷积神经网络。载于：神经信息处理系统进展会议论文集，第1790–1798页（2014，3、59）。Zhang, H., Dai, Y., Li, H., Koniusz, P.：用于图像去模糊的深度堆叠层次多块网络。载于：IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集（2019，1、2、3、60）。Zhang, H., Koniusz, P., Jian, S., Li, H., Torr, P.H.S.：重新思考类别关系：绝对—相对有监督与无监督少样本学习。载于：IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集，第9432–9441页（2021，4、61）。Zhang, H., Zhang, L., Qi, X., Li, H., Torr, P., Koniusz, P.：利用置换不变注意力进行少样本动作识别。载于：欧洲计算机视觉会议（ECCV）（2020，4、62）。Zhang, J., Pan, J., Ren, J., Song, Y., Bao, L., Lau, R.W., Yang, M.H.：利用空间变化递归神经网络进行动态场景去模糊。载于：IEEE计算机视觉与模式识别会议论文集，第2521–2529页（2018，3、8、9、63）。Zhang, K., Luo, W., Zhong, Y., Ma, L., Stenger, B., Liu, W., Li, H.：通过真实模糊进行去模糊。载于：IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集（2020，9、64）。Zhou, S., Zhang, J., Pan, J., Xie, H., Zuo, W., Ren, J.：用于视频去模糊的时空滤波器自适应网络。载于：IEEE/CVF国际计算机视觉会议论文集，第2482–2491页（2019，11）