



ELSEVIER

Contents lists available at ScienceDirect

Engineering

journal homepage: www.elsevier.com/locate/eng



Research
Artificial Intelligence—Review

视频压缩感知技术十年之旅——从理论到实际应用

张志宏^{a,#}, 郑已明^{b,#}, 仇旻^c, 司徒国海^d, David J. Brady^e, 戴琼海^a, 索津莉^{a,f,*}, 袁鑫^{b,*}

^a Department of Automation, Tsinghua University, Beijing 100084, China

^b Research Center for Industries of the Future (RCIF) & School of Engineering, Westlake University, Hangzhou 310030, China

^c Key Laboratory of 3D Micro/Nano Fabrication and Characterization of Zhejiang Province, School of Engineering, Westlake University, Hangzhou 310024, China

^d Shanghai Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Shanghai 201800, China

^e Wyant College of Optical Sciences, University of Arizona, Tucson, AZ 85721, USA

^f Shanghai Artificial Intelligence Laboratory, Shanghai 200232, China

ARTICLE INFO

Article history:

Received 27 January 2024

Revised 25 July 2024

Accepted 6 August 2024

Available online 31 August 2024

关键词

视频压缩感知

计算成像

深度学习

实际应用

摘要

自从第一个编码孔径视频压缩感知(CS)系统被报道以来已经过去了十余年。与传统相机直接采集视频不同,视频压缩感知通过引入高速调制器件,在一个积分时间内对高速场景进行编码并记录成一张二维压缩图像,之后使用重建算法来恢复高动态的目标场景视频。视频压缩感知系统可以从单张曝光图像中重建多帧视频,即采用低速相机来捕捉高速场景,从而有效降低成像系统的数据通量,形成了该技术的核心优势。基于该成像框架,过去十年间,研究者已经利用不同的调制器件搭建了多种视频压缩感知系统,并开发了不同类型的解码算法以重建高质量的视频。解码算法经历了从迭代优化算法向基于深层神经网络推理的重大转变,目前基于深度学习的新兴方法因其优异的推理效率和重建质量而形成主流重建方案,也为视频压缩感知的落地应用提供了可能。鉴于此,本文回顾和总结了过去十年视频压缩感知的进展,对硬件系统和重建算法进行了详细梳理和论述,进一步分析了视频压缩感知在硬件和算法方面的局限性、解决方案和未来的研究方向,以期从事该领域研究的人员提供参考和启发。

© 2024 THE AUTHORS. Published by Elsevier LTD on behalf of Chinese Academy of Engineering and Higher Education Press Limited Company. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. 引言

视觉是人类感知世界进而认知环境最重要的渠道。自20世纪60年代发明数字相机以来[1],互联网和智能手机的普及让人们可以不受时间和地点的限制,用视频记录日常生活,并通过Twitter、TikTok等社交媒体平台分享。顺应这一新兴趋势,对超高清视频(如4K和8K)和高速视频的获取与显示已成为核心需求。在工业领域,监控、自

动驾驶、无人机(UAVs)和机器人等新兴技术也在很大程度上依赖实时、高质量的视频采集和大规模视频数据处理来完成各自的任务。

在过去的几十年里,学界利用电荷耦合器件(CCDs)或互补金属氧化物半导体(CMOSs)以及图像信号处理器(ISPs)构建了经典的成像模式。具体来说,被记录场景反射的光线首先由传感器集成形成原始图像,随后经由ISP处理生成相应的输出。然而,在当前的视频大规模应

* Corresponding authors.

E-mail addresses: jl suo@tsinghua.edu.cn (J. Suo), xyuan@westlake.edu.cn (X. Yuan).

These authors contributed equally to this work.

2095-8099/© 2024 THE AUTHORS. Published by Elsevier LTD on behalf of Chinese Academy of Engineering and Higher Education Press Limited Company. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

英文原文: *Engineering* 2025, 46(3): 172–185

引用本文: Zhihong Zhang, Siming Zheng, Min Qiu, Guohai Situ, David J. Brady, Qionghai Dai, Jinli Suo, Xin Yuan. A Decade Review of Video Compressive Sensing: A Roadmap to Practical Applications. *Engineering*. <https://doi.org/10.1016/j.eng.2024.08.013>

用背景下, 该处理流程不仅依赖于大容量片上存储器来缓存采集的原始图像并执行ISP计算, 还需要高带宽以支持图像传输。各种场景对视频采集和处理的需求不断增加, 给现有的成像框架带来了巨大压力。此外, 在后摩尔时代, 增加片上存储器容量已成为一项挑战, 而且成本高昂。物联网 (IoT) 平台等应用场景也对大规模数据传输提出了新的挑战。因此, 迫切需要一种既能提供高吞吐量, 又能在低带宽限制下运行的新型成像范式。

在这种背景下, 计算成像技术获得了越来越多的关注, 并被视为未来成像发展的前沿方向[2]。其目标是提高成像质量[3-5]、捕捉高维信息[6-8]或优化成像系统的性能[9-11]。过去十年中, 光学工程、集成电路和深度学习领域的重大进展为新型计算成像技术的实际应用铺平了道路。在这些技术中, 视频压缩感知 (CS) [7,12]已成为高通量、低带宽成像领域的代表性方法。这种方法是在成像过程中将场景信息编码为压缩测量数据, 然后通过后处理算法将其重建为原始/目标视频帧。通过采用这种方法, 可对海量视频数据进行压缩采集、存储和传输, 而不会对片上存储器和传输带宽造成显著负担。

视频CS的底层数学原理在于CS理论[13-15]。尽管CS提出已有几十年, 但由于硬件限制和算法性能不足, 其在成像领域的实际应用进展有限。幸运的是, 光学调制设备和基于学习的重建算法的最新进展在很大程度上缓解了这些问题。

在硬件方面, 空间光调制器的设计和生产技术, 如数字微镜器件 (DMDs) 和硅基液晶显示器 (LCoS) 等, 取得了创新突破, 显著提升了刷新速度和空间分辨率。这使得视频CS系统能够满足对高速、高保真视频采集日益增长的需求。此外, 半导体芯片和集成电路的发展催生了能够提供像素级曝光控制能力的特定应用传感器[16-17]。这些传感器无需外部光学调制设备, 使得视频CS有可能直接在紧凑的工业相机上完成[7]。

在算法方面, 传统的基于迭代的优化算法存在重建质量低和时间成本高的问题。大约五年前, 从编码快照重建 $1024 \times 1024 \times 10$ 视频序列可能需要数小时[18-19], 这严重限制了视频CS的实际应用, 尤其是在实时任务中。在过去的几十年里, 深度学习在各种低级和高级计算机视觉 (CV) 任务中迅速超越了传统算法, 包括去噪[20]、去模糊[21]、分类[22]、检测[23]和跟踪[24]。在CS问题的重建算法方面, 研究者开发了基于学习的方法以及学习-优化混合框架, 如即插即用 (PnP) [25-27]和深度展开[28-30], 并在质量、速度和灵活性方面显著改善了重建性能。总体而言, 光学硬件和重建算法的最新进展为视频CS在

实际场景中的应用开辟了一条前景广阔的道路[31]。

作为一种新颖且有前景的成像范式, 视频CS不仅有助于获取符合人类视觉特性的视觉信息, 还为设计更高效的端到端机器视觉框架 (包括信息获取、存储、传输、处理和分析) 提供了新的思路。视频CS的成像机制使其相较于传统成像范式在高信息容量和低带宽占用方面具备显著优势。此外, 压缩数据格式减轻了计算负担, 从而提高了目标检测和路径规划等高级CV任务的处理速度。因为这些特性大大降低了功耗, 对于无人机和物联网边缘传感器等负载受限平台尤其有利。更重要的是, 视频CS可以帮助自动驾驶汽车、无人机和机器人等智能系统执行实时感知和决策任务。研究者为了验证该框架在实际应用中的有效性, 已进行了一些初步尝试[32-35]。

在视频CS发展的这一里程碑时刻, 本文旨在对视频CS的历史进行全面回顾。本文的其余内容安排如下: 在第2节中, 概述了视频CS的基本理论和成像原理; 第3节和第4节分别详细回顾了视频CS的硬件设计和重建算法; 第5节讨论了现有挑战和机遇, 并为视频CS的进一步发展提供了前瞻性路线图; 随后, 第6节总结了视频CS的代表性应用; 最后, 第7节对全文进行总结, 并对视频CS的未来进行了展望。

2. 基本理论与成像原理

从信息论的角度来看, 传统成像系统遵循香农-奈奎斯特采样定理[36], 而视频CS则基于压缩采样或CS定理[13-14]。香农-奈奎斯特采样定理指出, 如果采样率高于信号最高频率的两倍, 则可以从采样中完美恢复原始信号。这条规则使得仅通过硬件升级来提高成像系统吞吐量面临巨大挑战, 尤其是在后摩尔定律时代。然而, CS定理的提出为规避这一问题提供了新的思路。根据该定理, 在满足两个条件的情况下, 可以用远少于香农-奈奎斯特采样定理所要求的样本重建信号, 而且几乎没有信息损失。其中, 第一个条件是原始信号在某个域中应具有稀疏性。音频、图像和视频等大多数自然信号通常都满足这个条件。第二个条件是采样矩阵应与上述稀疏域的基矩阵保持非相干性。在实践中, 可以采用服从高斯分布或伯努利分布的随机采样矩阵来满足这一条件。

尽管CS定理可以在现有传感器带宽的限制下提高成像系统吞吐量, 但由于缺乏有效的采样硬件和高效的重建算法, 在过去几十年中, 其在成像领域并未得到广泛应用。近年来, 随着光学工程、计算成像和深度学习技术的发展, 各种视频CS系统和重建算法相继问世, 极大地推

动了CS定理在日常生活中的实际应用。参考文献[37]专门针对CS理论进行了系统阐述，其内容在参考文献[7]中得到了进一步的详细评述。本节将简要介绍视频CS的通用成像原理和数学公式。对特定视频CS系统和重建算法的详细回顾可参见后续章节。

2.1. 成像原理

视频CS的基本原理如图1所示。为简单起见，我们用离散的高速帧来表示连续的动态场景，这些帧与传感器的输出相吻合。如图所示，在采集过程中，来自场景空间的原始帧首先通过随机编码掩模进行调制，随后在一次曝光内由传感器集成，形成压缩编码测量数据。根据CS定理，可以借助CS重建算法从测量结果中恢复原始高速帧。通过这种方式，视频CS系统就能在保持低带宽的同时，有效提高数据吞吐量。

如上所述，视频CS的硬件系统主要由成像、调制和记录三部分组成。其中，成像和记录部分与传统相机类似。调制在视频CS中起着至关重要的作用，它通过执行压缩采样，在将海量视频帧转换为较少编码测量的过程中，有效减少数据量并缓解传感器的带宽压力。调制设备的刷新率直接决定了视频CS系统的压缩比上限，从而限制了视频恢复的速度。编码掩模的设计同样对最终的重建性能具有显著影响。得益于光学和机械工程的发展，刷新率达到10 kHz或更高的空间光调制器已作为商业产品上市。在实际应用中，为降低实现复杂度并提高对比度，调制过程通常采用源自伯努利分布的 $\{0,1\}$ 随机二值掩模。获取编码测量数据后，可以采用各种CS重建算法从这些压缩测量数据中恢复原始高速视频帧。第4节中将详细介绍各种重建算法。

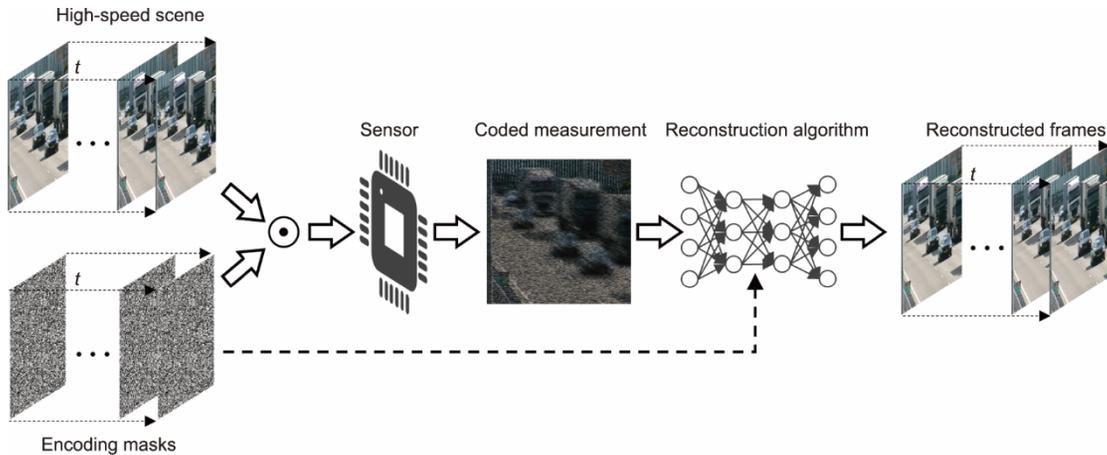


图1. 视频CS的基本原理图。在单次曝光内，多帧高速帧首先通过编码掩模进行调制，然后由传感器集成形成编码测量。给定编码掩模和测量值，原始帧可通过重建算法恢复。 t : 时间。

2.2. 数学公式

根据上述示意图和原理，我们可以用线性方程来建立视频CS的数学模型，具体如下：

$$\mathbf{Y} = \sum_{k=1}^M \mathbf{C}_k \odot \mathbf{X}_k + \mathbf{G} \quad (1)$$

式中，实数矩阵 $\mathbf{X}_k \in \mathbb{R}^{n_x \times n_y}$ 和 $\mathbf{C}_k \in \mathbb{R}^{n_x \times n_y}$ 分别表示第 k 帧 ($k = 1, \dots, M$) 高速视频帧及其对应的编码掩模，每帧均包含 $n_x \times n_y$ 个像素。 M 表示传感器在单次曝光期间所集成的高速视频帧数。 $\mathbf{G} \in \mathbb{R}^{n_x \times n_y}$ 是加性噪声， $\mathbf{Y} \in \mathbb{R}^{n_x \times n_y}$ 表示最终得到的编码测量值。 \odot 表示哈达玛（逐项）积。

2.2.1. 前向模型

从数学角度来看，可以通过向量化将式（1）进一步推导为以下形式：

$$\mathbf{Y} = \mathbf{H}\mathbf{x} + \mathbf{g} \quad (2)$$

式中， $\mathbf{g} = \text{Vec}(\mathbf{G}) \in \mathbb{R}^n$ ， $\mathbf{y} = \text{Vec}(\mathbf{Y}) \in \mathbb{R}^n$ ，且 $n = n_x n_y$ 。相应地，高速视频信号 $\mathbf{x} \in \mathbb{R}^{nM}$ 与采样矩阵 $\mathbf{H} \in \mathbb{R}^{n \times nM}$ 可表示为以下表达式：

$$\mathbf{x} = [\text{Vec}(\mathbf{X}_1)^\top, \dots, \text{Vec}(\mathbf{X}_M)^\top]^\top \quad (3)$$

$$\mathbf{H} = [\text{diag}(\text{Vec}(\mathbf{C}_1)), \dots, \text{diag}(\text{Vec}(\mathbf{C}_M))] \quad (4)$$

从式（4）可以看出，采样矩阵 \mathbf{H} 具有特殊的稀疏结构，它由 M 个对角矩阵串联组成。因此，这里的压缩比等于 $1/M$ 。先前的研究[37]已经证明，即使在 $M > 1$ 的情况下，只要信号结构足够清晰，视频CS的重建误差也是有界的。这种特殊结构使得在某些基于优化的视频CS重建算法中降低计算复杂度成为可能[18–19,38]。

下面，我们总结了参考文献[37]中针对视频CS所提出的理论结果。首先，视频CS与传统CS [如单像素成像 (SPI)] 的前向模型存在差异，因此，为传统CS导出的理论保证并不适用于视频CS。视频CS与SPI [39]的主要区

别体现在以下两个方面:

(1) 在 SPI 中, 感知矩阵为稠密矩阵。感知矩阵的每一行对应调制器施加于场景[一幅二维(2D)静态图像]上的一个图案, 而单像素探测器则捕捉一个测量值(即测量中的一个元素)。

(2) 在视频 CS 中, 式(4)中的感知矩阵为稀疏矩阵, 由对角矩阵串联而成。测量值中的每个元素是经过掩模调制的视频帧中对应元素的加权和。

2.2.2. 理论保证

视频 CS 的理论推导[37]基于应用于压缩感知的信号压缩结果[40]。参考文献[37]提出了一种可压缩信号追踪(CSP)型的优化方法作为基于压缩的重构算法用于视频 CS。假设紧集 $\mathcal{Q} \in \mathbb{R}^{n_x n_y M}$, 配备一个压缩码, 其压缩率等于 r , 并假设该紧集可以通过映射 (f, g) 描述, 其中, f 表示编码映射函数, g 表示解码映射函数。

假设 $\mathbf{x} \in \mathcal{Q}$, 重建信号 $\hat{\mathbf{x}}$ 可通过以下优化问题[式(5)]求解。

$$\hat{\mathbf{x}} = \operatorname{argmin}_{\mathbf{u} \in \mathcal{U}} \|\mathbf{y} - \mathbf{H}\mathbf{u}\|_2^2 \quad (5)$$

式中, \mathcal{U} 表示由 (f, g) 定义的编解码器的码本。换句话说, 给定测量向量 \mathbf{y} , 该优化问题会在所有可压缩信号(即码本中的信号)中, 选择那个通过 \mathbf{H} 采样时最接近观测测量的信号。

参考文献[37]中的以下定理通过将(压缩/解压)编解码器参数、压缩率 r 及对应失真度 δ 与帧数 M 相关联, 阐明了采用 CSP 型优化的视频 CS 恢复性能特征。其中, 帧数 M 决定了压缩采样率和最终整体重建质量。

定理 1. 假设对于所有 $\forall \mathbf{x} \in \mathcal{Q}$, $\|\mathbf{x}\|_\infty \leq \frac{\rho}{2}$; ρ 是一个有上界的正数。且假设速率为 r 的代码在 \mathcal{Q} 上实现了失真 δ 。此外, 假设 \mathbf{C} 中的每个元素都服从标准高斯分布, 即 $C_{i,j,k} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$; 对于所有 $\forall i = 1, \dots, n_x; j = 1, \dots, n_y; k = 1, \dots, M$ 。令 $\hat{\mathbf{x}}$ 表示式(5)中可压缩信号追踪优化的解。假设 $\epsilon > 0$ 是一个自由参数, 使得 $\epsilon \leq \frac{16}{3}$ 。那么

$$\frac{1}{n_x n_y} \|\mathbf{x} - \hat{\mathbf{x}}\|_2^2 \leq M(\delta + \rho^2 \epsilon) \quad (6)$$

其概率大于 $1 - 2^{-n_x n_y M r + 1} e^{-n_x n_y \left(\frac{3\epsilon}{32}\right)^2}$ 。

需要注意的是, 假设信号是有界的, 即 $\|\mathbf{x}\|_\infty \leq \frac{\rho}{2}$ 。这是定理证明的必要条件; 此外, 图像和视频像素值在被相机捕捉后(通过传感器的动态范围)通常是有界的。定理 1 表明, 对于以压缩率 r 和失真 δ 为参数的可压缩信号, 如果该信号通过视频 CS 系统进行压缩捕捉, 那么重建误差

(即估计信号与真实信号之间的误差)很有可能被失真限制。

定理 1 的结果可以应用于任何具有相同前向模型的视频 CS 系统。其核心假设是所需的高维数据具有高度可压缩性, 这对于设计的视频 CS 系统通常是成立的。本质上, 如前所述, 由于视频 CS 的压缩采样率等于 $1/M$, 因此重建误差受信号的固有可压缩性的限制。在此, 我们要指出的是, 定理 1 中推导出的界限并不严格, 该研究方向仍需更多探索。进一步的分析和有待填补的研究空白可见参考文献[7,28]。

3. 视频 CS 硬件

随着光学、电子和机械技术的最新进展, 研究者相继提出了新型光学编码方案和多种视频 CS 系统, 并在空间分辨率、帧率和成像质量等方面取得了显著的性能提升。本节将列出现有的视频 CS 系统设计, 并详细总结其原理、性能和应用。表 1 [9,12,16,41–54]和图 2 [9,12,16,41–44,46,49–58]中也展示了一些代表性的视频 CS 系统。

3.1. 空间光调制

视频 CS 系统中最常用的方法是在光路中引入外部空间光调制器和相应的中继光学元件, 以编码入射场景光线。Llull 等[12]首次利用压电平台驱动的光刻图案掩模的机械平移来生成时变图案, 进行光学编码。他们搭建的视频 CS 系统能够从单个编码快照中重建 10 帧以上(有时甚至超过 100 帧)灰度视频; 后续还将其扩展到捕捉彩色视频方面[41]。Koller 等[42]随后通过优化掩模图案设计和硬件实现, 在空间维度和重建质量方面改进了这项工作。其原型机可以实现 200 万像素分辨率和每秒 743 帧(FPS)的高速视频记录。这些工作表明, 机械调制方案具有高空间分辨率、低系统成本和可灵活扩展等显著优势, 尽管其在实际应用中可能导致系统体积庞大, 且存在一定不稳定性。此外, 压电平台的平移速度和频率响应限制也成为提高压缩比的关键障碍。

另一种方法是利用现成的可编程空间光调制器, 包括 DMDs [46,56,59–62]和 LCoS [43–45], 进行光学编码。与机械平移方案相比, 这些方案具有结构紧凑、稳定、快速和灵活的优点, 但它们成本更高, 且空间分辨率限制在数百万像素级别。

DMD 是一种微型光电机系统, 由数百万个微型反射镜组成的矩形阵列构成。每个镜片安装在一个轭架上, 轭架通过柔性扭转铰链连接到两个支柱。这些铰链可以驱

动镜片旋转 $\pm 12^\circ$ ，从而通过改变反射方向实现对入射光的二进制调制。DMDs的最大刷新率可达10 kHz以上，这使得在视频CS系统中能实现高压缩比。其缺点在于其周期性微镜排列可能引入衍射效应，该效应会导致生成的图像中产生伪影，从而降低成像质量。

LCoS是另一种广泛使用的空间光调制器，它利用液晶的偏振特性实现光调制。更具体地说，LCoS中的每个像素由液晶层、反射层和硅基板组成。在操作过程中，硅基板会调节施加到液晶上的电压，以改变液晶分子的取向。由于液晶材料独特的光学各向异性，液晶的折射率和通过液晶的光的相位会发生相应的变化。因此，通过调节硅基板的输出电压，可以调制入射光的偏振态。反射层用于将

调制后的光反射回光学系统。一般来说，视频CS系统中，会在LCoS的入射端和出口端放置一对正交偏振器，将偏振调制转换为所需的二进制振幅调制。视频CS系统中常用的硅基铁电液晶（FLCoS）的刷新率高达4.5 kHz。尽管这低于DMD，但仍高于光刻掩模的机械平移速度，可以满足典型应用的需求。基于LCoS的视频CS系统的主要问题是其光吞吐量较低，因为前置偏振片会滤除一半自然光。但其优势在于成像质量优于机械平移或DMD方案。

虽然上述系统通常需要使用全局快门相机来同步调制设备和相机，但研究人员最近也探索了使用卷帘快门相机进行视频CS [47–48]。尽管在视频CS系统中直接使用卷帘

表1 代表性视频CS系统总结

Scheme	Implementation	References	Frame dimension (pixels)	Acquisition frame rate (frame·s ⁻¹)	Compressive ratio	Reconstruction frame rate (frame·s ⁻¹)	Data throughput (voxel·s ⁻¹)
Spatial light modulation	Mask translation	[12]	281 × 281	30	14	420	33.2 M
		[41]	512 × 512 × 3	30	22	660	519 M
		[42]	1600 × 1200	74.3	10	743	1.43 G
	LCoS	[43]	1024 × 768	25	8	200	157 M
		[44–45]	1280 × 1024 × 3	55	18	990	3.89 G
		[46]	1280 × 1024	50	50	2500	3.28 G
Active illumination	Rolling shutter	[47–48]	1024 × 1024	17	32	544	570 M
	Projector	[49]	296 × 325	200	5	1000	96.2 M
Pixel-wise coded exposure sensors	Infrared-pulsed	[50]	1440 × 1080	15	14	210	327 M
	One-tap PCE sensor	[16]	127 × 90	5	20	100	1.14 M
	Tow-tap PCE sensor	[51]	128 × 128	10	64	640	10.5 M
Sophisticated encoding schemes	Quasi PCE sensor	[52]	656 × 496	15	16	240	78.1 M
	Dual-view video CS	[53]	650 × 650	50	20	1000	845 M
	Hybrid coded aperture	[9]	3200 × 3200	15	30	450	4.61 G
	Sinusoidal sampling	[54]	2048 × 2048	15.6	128	1996.8	8.38 G

PCE: pixel-wise coded exposure.

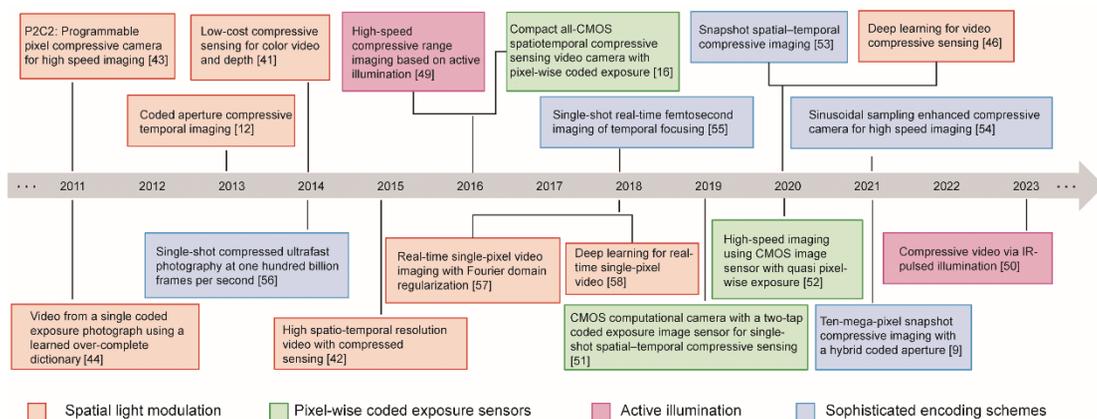


图2. 视频CS系统发展路线图[9,12,16,41–44,46,49–58]。

快门具有挑战性，但参考文献[48]提出了一种混洗解决方案，并已通过DMD验证。要在硬件中实现这一解决方案仍有很长的路要走，但前景可观。另一种解决方案是使用多个具有不同旋转角度的卷帘快门相机来协同捕捉场景[47]。然而，这给硬件设计带来了挑战，并增加了光学系统的复杂性。

通常情况下，会使用2D传感器来捕捉使用上述方法生成的调制图像。然而，在某些特殊应用（如太赫兹成像和红外成像）中，可能难以获取阵列探测器，或成本过高。在这种情况下，单像素相机提供了一种可行的替代方案。尽管SPI最初主要用于压缩图像采集，但由于需要大量测量且单帧重建耗时较长，近年来高速调幅调制策略[63–64]与重建算法[57–58,65]的突破，使得该技术得以应用于视频CS领域。例如，Hahamovich等[63]提出了一种基于旋转掩模进行高速空间调制的快速SPI系统。调制速率达2.4 MHz，空间分辨率为 101×103 像素，可实时捕获72 FPS的动态场景。Kilcullen等[64]通过使用DMD和激光扫描硬件实现的扫描聚合模式，加速SPI，调制速度达14.1 MHz。他们还开发了一种支持并行计算的轻量级算法，实现了 101×103 像素、100 FPS的实时视频重建。其他研究人员，如Higham等[58]和Mur等[65]，分别将卷积自动编码器网络和循环神经网络引入SPI框架。这些技术显著提高了重建质量和速度，特别是在高压缩比的情况下。

总而言之，空间光调制方案常用于利用现有商用传感器实现视频CS系统。然而，它也伴随着某些缺点。一方面，引入空间光调制器增加了系统功耗和成本；同时，也牺牲了光吞吐量，从而导致捕捉测量的信噪比（SNR）降低。另一方面，由于不可避免的光学像差和系统干扰，这些系统在每次数据采集前都需要烦琐而精细的校准，并且重建质量对校准误差非常敏感。然而，在光照不足或平台不稳定等恶劣条件下，校准过程可能无法产生准确的结果。

3.2. 主动照明

视频CS的核心思想是在相机捕捉高速场景之前对其进行调制。虽然上述空间光调制方法以被动方式实现这一原理（即使用自然光，并在成像系统内部调制），但另一种实现方式是通过向目标场景施加结构光来调制出射辐射度。主动照明是一种实现结构光的一种新型方式，尤其适用于室内环境。

为此，Sun等[49]使用投影仪实现了视频CS的主动照明。他们还利用图案比例与物体深度之间的关系实现了高速三维（3D）成像。在这个主动系统中，可以从200 FPS捕捉的测量值中重建1000 FPS的深度图视频。这

种装置的缺点是工作距离短，并且由于使用的是可见光照明，因此对环境光的干扰高度敏感。最近，Guzmán等[50]通过使用红外（IR）脉冲照明调制场景解决了这个问题，其中IR照明的使用使得空间图像（可见光）和时间图像通道能够分离。他们的设置实现了从15 FPS捕捉的压缩测量值中重建210 FPS的视频。值得一提的是，该系统使用了两个额外的测量值来改善重建效果。除了上述系统外，基于发光二极管（LEDs）和DMDs的主动照明也被用于视频和光谱CS的联合处理[66]。

从数学角度看，主动照明与使用其他可编程空间光调制器（如DMD或LCoS等）原理相同，但由于其高度灵活性和低成本，在短期内更容易实现。

3.3. 像素级编码曝光传感器

随着半导体和集成电路技术的不断进步，学界设计出具备像素级集成控制功能的新型CMOS传感器。这些传感器无需额外的外部组件即可实现视频CS，为这项技术未来的批量生产和广泛应用提供了可能性。

Zhang等[16]设计了一种具有像素级编码曝光（PCE）能力的全CMOS芯片，并展示了其在视频CS中的应用。他们构建了一个分辨率为 127×90 像素的原型图像传感器，能够从以5 FPS捕捉的编码测量值中重建100 FPS的视频。随后，Martel等[67]基于SCAMP-5设备实现了视频CS系统，SCAMP-5是一个具有 256×256 像素的可编程传感器处理器。他们在其框架中共同设计了逐像素快门功能和重建算法，并在16倍压缩比下实现了卓越的重建质量。不同的是，Sarhangnejad等[69]和Luo等[51,70]分别设计了一种新型双通道编码曝光像素传感器，该传感器能够在单次曝光期间输出两个互补的编码图像。这种传感器也被称为编码双桶相机[17]。它集成了两个电荷收集桶和一个可写入存储器，用于控制每个像素中哪个桶处于激活状态。通过分配可编程二进制模式来控制激活的电荷桶，该传感器可以实现像素级曝光编码，并在每帧视频输出两个互补的编码快照。与Zhang等[16]提出的单桶范式相比，编码双桶传感器充分利用了所有入射光，因此具有更高的光效率，这有助于视频CS的重建算法[71]。近年来，学界围绕编码双桶传感器开展了一系列研究[72–73]，以提高传感器的空间分辨率、填充因子、调制速度、动态范围、功耗等方面的性能。此外，还提出了其他像素级曝光控制范式[52]。

与外部调制方案相比，具有像素级曝光控制能力的传感器无需中继光学器件即可直接实现视频CS，并且可以实现批量生产。这一特性显著减小了视频CS系统的体积，降低了功耗，并提高了稳定性，从而拓宽了视频CS在不

同领域的应用。然而,值得注意的是,现有的像素编码曝光传感器与成熟的传统全局曝光图像传感器在填充因子、空间尺寸(分辨率)、成像质量等方面仍存在显著差距。探索这些领域是该领域未来的发展重点。

3.4. 精密编码方案

除了使用外部调制器或PCE传感器实现的直接空间调制方案外,还有多种性能更优或功能特殊的CS系统。压缩超快成像技术(CUP)是基于视频CS的代表性高速成像技术之一[55–56]。在CUP中,首先使用DMD对输入图像进行静态伪随机二进制图案的空间编码。然后,使用条纹相机将编码后的图像沿空间轴进行时域色散,然后集成到传感器上形成编码快照测量值。从本质上讲,将使用DMD进行空间编码和使用条纹相机进行时间剪切相结合,实现了视频CS所需的时空调制,这与编码孔径快照光谱成像的核心思想一致[74]。由于条纹相机具有高速切换能力,CUP能够以最高 10^{11} FPS的速率捕捉瞬态事件,从而能够观察激光脉冲的反射和折射等物理现象。近年来,CUP的成像速度和重建质量得到了显著提升[75]。此外,CUP已扩展到捕捉光谱等多维视觉信息方面[59]。

CUP的主要局限在于空间分辨率受限,空间分辨率由条纹相机开口狭缝的宽度决定。其硬件成本高,使得其应用主要限于科学研究。另一个研究领域侧重于通过设计更先进的调制方案,同时实现更好的空间和时间分辨率,以实现工业和日常生活中的更普遍应用。Deng等[54]将通过正弦采样实现的频域调制引入视频CS,以在保持空间分辨率的同时追求更高的压缩比。他们方法的基本原理涉及将多组时空编码测量值映射到傅里叶域中的不同位置,从而提高了捕捉编码快照的信息密度,最终使整体压缩比达到0.01以下。Zhang等[9]则专注于克服当前调制设备带来的空间分辨率限制,研发了新型混合编码孔径快照压缩成像(HCA-SCI)方案,用于视频CS中的时空调制,并使用静态高分辨率光刻掩模和动态低分辨率LCoS的组合实现了一个千万像素的原型系统。与同样有可能实现更高空间分辨率的掩模平移方案相比,HCA-SCI在功耗、体积系数和系统稳定性方面表现出显著优势。

除上述方法外,一些创新性工作还将视频CS扩展到不同维度。例如,Tsai等[76]通过引入光谱色散,将基于机械平移的视频CS系统扩展到以压缩方式捕捉多光谱高速场景。其原型相机能够从单个编码快照中重建15个光谱通道和10个时间帧。Sun等[77]将非对称立体技术引入视频CS,以实现被动高速三维成像,该系统可以从以80 FPS捕捉的测量数据中重建800 FPS的深度图视频。Qiao

等[53]设计了一个双视图视频CS系统,该系统同时编码来自两个不同视角的入射光,并通过相同的传感器将其集成,形成单个编码测量。然后可以从编码快照中重建来自相应视图的两个视频序列[60]。该系统的基本原理是将每个视图调制到正交偏振光上,并移动其编码图像以引入它们之间的横向位移。研究人员构建了原型系统来验证这一设计,并实现了分辨率为 650×650 像素、时间压缩比为20的双视图视频CS。最近,Dou等[78]将视频CS的空间调制扩展到数字全息技术领域,以提高采样速度,Luo等[79]将视频CS的原理引入结构光超分辨率技术。类似的思路也被用于快照叠层成像技术中,以在单次曝光中同时捕捉大视野和高空间分辨率的场景[80]。参考文献[81]对快照压缩成像的不同采样技术进行了比较研究。

4. 视频CS算法

在视频CS中,从通过压缩采样获得的编码测量值中恢复原始高速视频是一个难题。在其求解过程中,方程的数量远少于未知数的数量,因此不存在唯一解。传统的优化算法会探索各种先验知识,如稀疏性,以减少这种问题的不确定性。图像和视频处理领域的大量理论研究和技術为自然场景的结构分析提供了强有力的支持。研究表明,自然图像具有局部平滑性、非局部自相似性和稀疏性等特征。这些先验约束为单曝光压缩成像的计算重建提供了有效的约束条件,从而降低了求解问题的难度。最近,基于深度学习的重建算法已成为主导算法。

目前,重建算法可分为四类,如图3所示。第一类包括采用不同先验的迭代优化算法。基于深度学习的算法分为三种不同场景。为了整合迭代算法和深度去噪算法,研究者提出了采用预训练去噪网络的PnP算法来重建大规模视频。后来,即使在没有训练数据的情况下,也使用了采用未经训练的神经网络进行去噪的自监督重建算法进行重建,但其效果有限[82]。为实现更快推理,有人提出了使用神经网络的端到端重建算法。到目前为止,性能最好的算法是由多个阶段组成的深度展开网络实现的,每个阶段都包括投影和学习先验的神经网络[28]。这些类别总结在图3中,图4展示了包含代表性算法的路线图[12,18–19,29–30,38,43–44,83–93]。

4.1. 传统优化算法

传统优化框架中的稀疏先验包括全变分(TV)、离散余弦变换(DCT)、小波变换、低秩先验、过完备字典和高斯混合模型(GMM)。其中,TV约束[38]基于梯度统

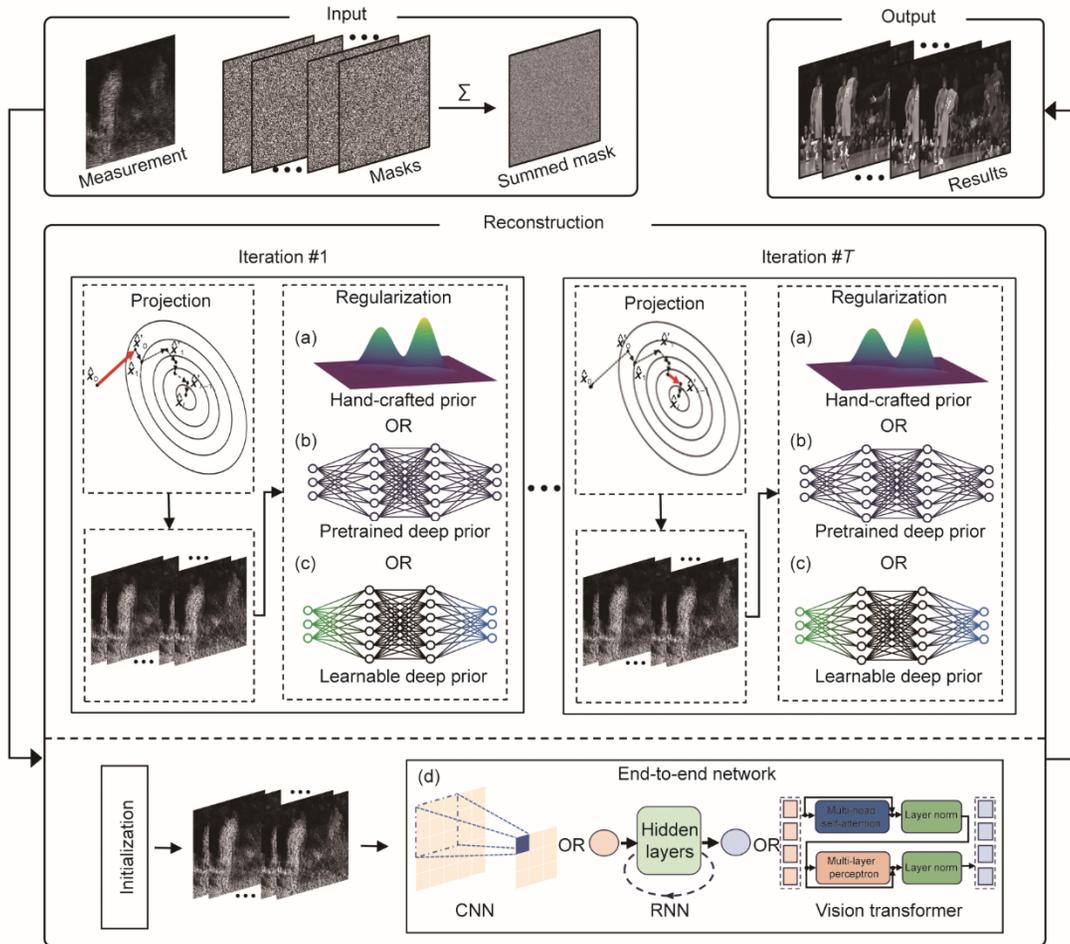


图3. 视频CS算法。目前，大多数重建算法可分为以下四类：(a) 采用不同手工先验的迭代优化算法；(b) 采用预训练去噪网络的PnP算法；(c) 由多个阶段组成的深度展开网络，其中每个阶段包括投影和可学习先验；(d) 使用神经网络的端到端重建算法。 T ：迭代次数；CNN：卷积神经网络；RNN：循环神经网络。

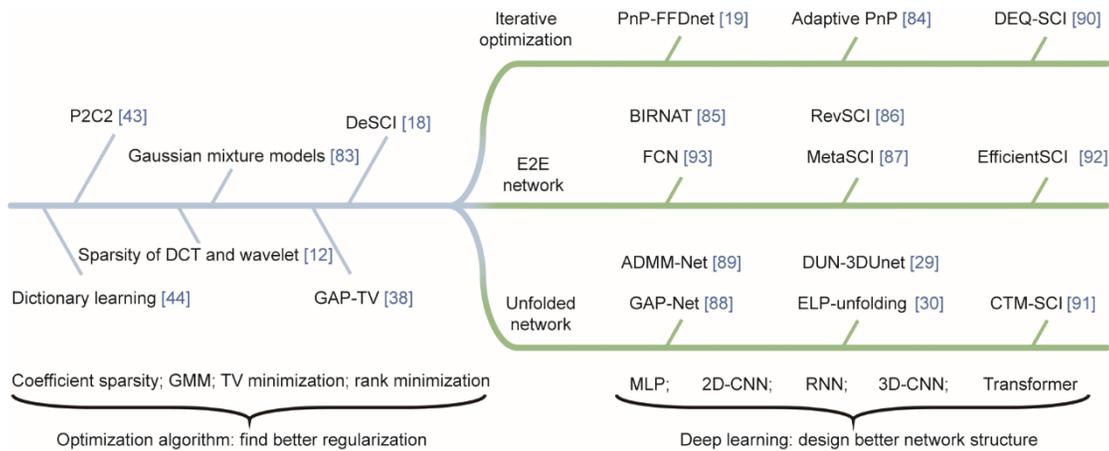


图4. 视频CS算法发展路线图及代表性算法[12,18–19,29–30,38,43–44,83–93]。DCT：离散余弦变换；GMM：高斯混合模型；TV：全变分；GAP：广义交替投影；MLP：多层感知器；ADMM：交替方向乘子法。

计服从拉普拉斯统计。它具有相对较好的噪声鲁棒性和纹理细节保留能力，并且运行速度快，强调局部平滑性。DCT和小波变换属于全局稀疏先验，但它们在某些特定场景中往往无法提供所需的稀疏性。与前三个全局先验相

比，低秩先验是一种非局部先验，它认为自然图像具有非局部自相似性[94–95]。自然图像在不同位置存在相似纹理，并且许多自然图像中的纹理本身也具有规律性。这表明自然图像的信息是冗余的。因此，可以利用图像的冗余

信息来重建和恢复图像或视频。DeSCI算法[18]进一步扩展了非局部自相似模型在视频CS中的应用，通过低秩约束高维高分辨率视频序列不同区域的相似信息，从而提高了重建精度。

过完备字典和GMM都属于局部稀疏先验。Naya等[44–45]使用基于过完备字典的算法重建了动态场景。这些算法从大量视频中学习过完备字典，并将任何给定视频表示为字典中元素的稀疏线性组合。由于字典是从视频数据本身学习而来的，因此它能够捕捉常见的视频特征，并且效果良好。GMM将局部图像块视为一系列高斯分布的混合。每个图像块都是独立、同分布的，并遵循具有特定均值和方差的高斯分布。因此，只要知道这些高斯分布的参数和每个图像块的索引，就可以重建图像。从图像数据库获得的局部稀疏先验可以提取更多信息，因此具有更高的重建质量[83]。其缺点是它只能获得特定训练场景的低维特征结构；对于不同的动态场景，需要重新训练，计算成本较高[96]。

传统优化算法具有高度的灵活性，可以快速适应不同应用场景中的新硬件系统。然而，它们通常需要大量的计算，因此导致重建过程缓慢，无法满足实时性要求，且重建质量较差。

4.2. 基于深度学习的算法

传统的优化算法通常依赖于不同的优化框架和各种正则化先验，通过迭代程序来解决问题。随着深度学习的进步，研究人员已经认识到深度去噪网络可以作为有效的图像先验，在保持传统优化方法灵活性的优势的同时，实现更优性能。因此，深度PnP方法的概念应运而生。

4.2.1. 深度PnP算法

与传统优化算法相比，深度PnP重建方法用深度神经网络取代了传统的先验项，从而实现了更高的推理速度和更好的重建质量。PnP框架最初由Venkatakrishnan等[97]于2013年提出，用于图像重建，尽管当时并未采用深度神经网络作为先验。2020年，针对大规模视频（特别是4K分辨率视频）提出了一种PnP方法[19]，并广泛应用于各种视频尺度。适当的预训练视频和图像去噪网络作为深度先验被整合到广义交替投影（GAP）框架中[38]，在当时无监督方法中取得了最优效果。此外，通过利用GPU，相较于传统的凸优化方法，提高了计算效率。

上述方法利用在其他数据集上预训练的先验模型，因此必须针对特定场景预训练去噪模型。一个网络中的预训练参数可能不匹配实际应用中的不同场景。2022年，Wu

等[84]开发了一种自适应PnP方法，根据特定动态场景自动更新深度去噪先验网络中的参数，从而弥合了预训练网络与实际应用之间的差距。

然而，上述PnP方法仍然存在另一个问题：缺乏训练数据。为了获得高质量的去噪网络，插入的模型需要在大量数据基础上进行训练。然而，在某些特殊应用场景中，生成训练数据可能具有挑战性。未经训练的去噪网络不依赖于特定的训练数据，因此对各类噪声与信号的通用性更强。这增强了它们在处理未知或不同类型数据时的鲁棒性。基于这些优势，Qiao等[82]提出了一种基于未经训练去噪网络的PnP算法，用于快照时域压缩显微成像。

众所周知，PnP算法表现出强大的泛化能力，适用于各种系统。然而，PnP算法仍面临一个显著挑战：无法保证全局收敛。引入去噪算法或正则化项（通常是非线性的）可能导致在某些场景下收敛困难或重建结果不稳定，并且此类算法在实践中需要精细调整超参数。为了解决这一挑战，大量研究通过引入稳定性条件、采用自适应步长等技术，提出改进的PnP算法。尽管如此，重建稳定性问题仍然是一个有待解决的难题。

4.2.2. 端到端深度学习算法

为了解决迭代方法推理速度慢的问题，开发了用于解决视频CS重建任务的端到端网络。其中，Qiao等[46]于2020年提出的端到端卷积神经网络（E2E-CNN）算法采用了U-Net架构，与传统优化方法相比，显著提高了视频CS的重建质量。该算法还受益于神经网络的高推理速度，展现出实时重建能力。然而，它仍然存在局限性：在视频重建过程中，它未能充分利用帧间相关信息，并且随着数据压缩比的增加，性能也会下降。为了应对这些挑战，基于双向循环神经网络（RNNs）的BIRNAT算法显著提高了重建性能，同时最大化了帧间信息利用[85]。此外，RNN结构的可扩展性使其能适应数据压缩比的提升。随着模型复杂度的增加，尽管重建性能有所提高，但训练收敛速度显著下降，同时伴随着GPU资源需求的增加。2021年引入的RevSCI算法有效地克服了这些问题[86]。RevSCI采用了可逆神经网络结构，首次将三维卷积神经网络（3D-CNNs）引入视频CS重建领域。可逆神经网络的使用使得在训练期间可以从后续层的值计算每个网络层的激活值，从而无需存储激活值，并大幅减少了训练期间的GPU资源的需求。3D-CNNs的引入使得网络能够同时考虑单帧内的空间特征和相邻帧之间的帧间相关信息。RevSCI改进了视频CS重建结果，并促进了高压压缩比下的数据重建。

然而，另一个不匹配问题也随之出现，因为训练中的掩模集可能与实际系统中使用的掩模不同。这会导致性能下降，因为深度神经网络在学习过程中为了提高重建质量，会与大量掩模相关信息耦合。若掩模变化，这就需要消耗大量时间再训练才能达到之前的重建质量。因此，在掩模发生变化的情况下，尤其是在实际应用中，快速适应全新系统的能力变得至关重要。Wang等[87]于2021年提出的MetaSCI解决了这一难题，它通过构建一个具有轻量级元调制参数的共享骨干网络来解决这一挑战，可快速适配新掩模，并易于扩展至大规模数据。

4.2.3. 深度展开算法

受交替方向乘法（ADMM）[98]和GAP[99]等优化算法的启发，领域提出了许多深度展开算法[29–30,88–89]用于解决视频CS中的逆问题。这些方法由多个结构相似的模块组成，每个模块代表传统优化算法中的一个迭代步骤。尽管成功地融合了迭代优化算法的优势并实现了端到端训练，但深度展开中的网络模块数量需控制在较少水平，原因有两个：①这些网络应保持简洁以实现实时重建速度；②由于内存限制，训练具有多阶段的深度展开网络具有挑战性。为了解决这个问题，Zhao等[90]于2023年提出了一种基于深度平衡模型（DEQ）的视频CS重建算法。该算法有效地将数据驱动的正则化方法与稳定的迭代收敛相结合，实现了低内存消耗和稳定的重建效果。DEQ在每个迭代层采用相同的变换，类似于以固定内存占用率训练任意深度的网络。这既符合PnP架构，也符合深度展开网络架构，有效地模拟了无限次的迭代步骤。

4.2.4. 简要总结

端到端深度神经网络在很大程度上依赖于丰富的训练数据，并且训练数据与测试数据之间的分布对齐是实现最优性能的关键。尽管在重建质量方面取得了显著进展，但由于深度神经网络的“黑箱”性质，这些算法仍缺乏可解释性。

深度展开算法代表了传统优化方法与神经网络的融合。这些算法涉及将优化过程的每次迭代展开到网络层中。在传统优化中，人工选择的稀疏先验（如梯度域中的TV和信号域中的DCT或小波变换）用于在重建过程中对潜在视频施加约束。然而，不同的稀疏先验对信号施加了不同的约束，选择理想的约束往往具有挑战性。在深度展开算法中，人工选择稀疏先验的过程被在每一层中嵌入的深度神经网络所取代，以自适应地学习相应的约束条件。虽然传统优化算法通常需要数百或数千次迭代，但深

度展开网络利用深度神经网络的学习能力，可以将迭代次数（大约两个数量级）大幅减少到数十次左右。GAP-net是第一个用于视频CS重建的深度展开算法[88]。DUN-3DUnet显著提高了重建性能[29]；它采用3DUnet作为主干网络，并使用密集特征图融合来克服网络内部信息传递的局限性。Li等[100]引入了安德森（Anderson）加速方法，以提升模型的收敛速度。

深度展开算法融合了传统优化技术和深度学习的优势，从而增强了模型的可解释性。通过结合传统优化方法，网络可以更加专注于学习图像和视频的内在特征，在更大程度上实现网络与掩模信息的解耦。这种灵活性旨在确保模型在面对不同系统（如调制掩模的变化）时的适应性。然而，现有算法仍然缺乏灵活性，因为深度展开算法的深度直接决定了优化迭代的次数。为实现更优的重建结果，通常需要更多的阶段，同时伴随着更高的GPU内存需求，这对使用大规模、高压比数据进行的训练构成了巨大挑战。

TwoStage-VCS[101]是一种结构精简的两级深度展开网络，专为视频CS重建任务设计。这种深度展开模型不仅在视频CS重建问题上取得了理想的效果，还表现出对不同调制掩模的高度适应性。它能够在多种掩模与不同尺度下产生令人满意的结果，展现出重建质量的稳定性。基于低分辨率数据集训练的模型可以直接迁移并应用于大规模数据集，从而显著缓解训练资源的计算需求。为应对视频CS的可扩展性挑战，研究者引入了集成学习先验模型[30]。自此之后，大规模数据训练难题已得到有效解决，重建算法的主流框架也从CNNs逐渐过渡到Transformer模型。

STFormer[102]通过在每个子模块中结合空间自注意力分支和时间自注意力分支，利用了时间域和空间域中的相关性。CTM-SCI[91]由3D CNN和3D可扩展分块稠密与扩张稀疏注意力机制组成，能够有效捕捉局部与全局的时空交互。此外，该方法引入了不确定性估计，增强对重建方差高的区域的关注。众所周知，基于Transformer架构的算法通常计算成本高。为应对这一问题，EfficientSCI[92, 103]在单个残差块内构建了分层稠密连接，从而降低模型的计算复杂度。

4.3. 重建网络骨干结构的演变

除了基于深度学习的视频CS中算法的不同结构外，研究网络骨干结构的演变也值得关注，如图3（d）所示。首个用于视频CS的深度神经网络[93]基于深度全连接网络。此后，尽管各种CNN已用于其他图像恢复问题，但据了解，尚未有关于视频CS重建的重要成果报告。U-Net

[104]在其他 CV 任务中广泛使用后不久，便开始被用作视频 CS 的骨干网络[46]。

视频 CS 之所以能够取得成功，主要原因之一就是视频帧之间的冗余，这使得利用相邻帧之间的时序信息变得直观。受此启发，BIRNAT [85]首次将 RNN 引入视频 CS，并进一步扩展到参考文献[105]中的其他视频 CS 系统。值得注意的是，BIRNAT 是第一个在精度方面超越最先进的基于优化的算法 DeSCI [18]的深度学习算法。然而，BIRNAT 的性能纪录仅保持了很短一段时间，很快就被参考文献[91,102]中开发的最新 Transformer 网络打破。

从这种演变流程可以看出，每当出现新的网络架构时，它都可以应用于视频 CS，并提升重建效果。然而，这种趋势只是填补了深度学习算法的精度差距。深度学习在视频 CS 中得到应用的另一个原因是推理速度的考量。

4.4. 从精度到速度

精度是阻碍视频 CS 实际应用的首要障碍。自 BIRNAT 算法提出以来，这个问题在一定程度上得到了解决。如前所述，尽管有些算法（如 DeSCI）可以提供高精度，但基于优化的算法通常速度较慢。另一方面，基于深度学习的算法在训练后（通常需要长达数天或数周的时间），推理速度更快。参考文献[46]使用 U-Net 模型为进行实时（30-FPS）重建提供了可行性。

如图 5 所示，随着研究人员追求更高的精度，模型规模也越来越大。这需要巨大的 GPU 内存，以及漫长的训练时间。为了解决这个问题，RevSCI 算法[86]应运而生，旨在减少重建过程中的内存占用。

4.5. 大模型还是高效模型

受大型语言模型（LLMs）在自然语言处理任务中成

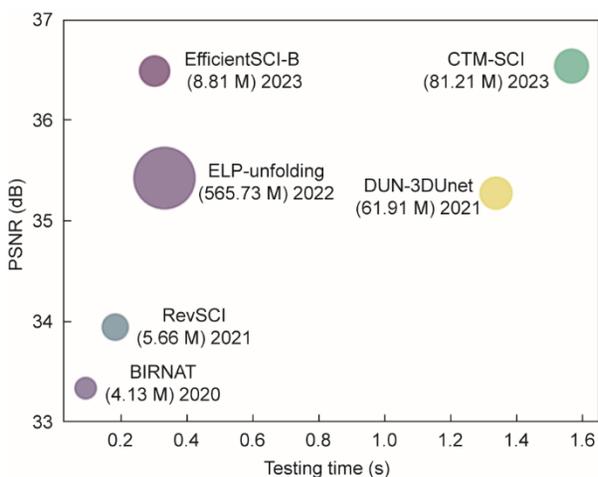


图 5. 基于深度学习的视频 CS 重建算法，绘制了峰值信噪比（PSNR）与测试时间的关系。各算法节点的半径与模型大小成正比。

功的启发，一方面，学界期望训练适用于多种场景（如不同空间尺寸、不同压缩比，甚至不同光照条件）的视频 CS 大模型；另一方面，如果要将视频 CS 推向实际应用，例如，在移动设备上使用，就必须使模型小型化，并缩短测试时的推理时间。这种两难选择导致模型泛化能力与部署或操作效率之间的权衡。虽然普遍认为大型模型是各种视觉任务的发展趋势，但高效模型在移动应用中也备受期待。

为此，研究者提出了 EfficientSCI [92]用于小模型的视频 CS 重建，并取得了当前最优效果。EfficientSCI 也已应用于数字全息显微镜等其他相关领域[78]。

目前广大研究者正在研究轻量级技术，如网络量化，以进一步推进实际应用。令人鼓舞的是，领域已开发出用于光谱压缩成像的二进制网络[106]，在轻量化网络设计方面做出了有价值的尝试。

5. 更多待填补的空白

尽管视频 CS 的硬件系统和重建算法方面已投入大量努力并取得了显著进展，但仍存在一些空白需要填补，以提高其在实际应用中的可用性和性能。

5.1. 硬件与系统

在硬件方面，视频 CS 系统的工作原理是调制和集成 M 个连续帧到一个单幅快照中，从而将重建帧的动态范围降低为原来的 $1/M$ 。视觉上，有限的动态范围限制了亮度和对比度，从而影响了图像细节的辨识度[62]。缓解此问题的一种可行的方法是采用具有更大满阱容量的传感器。此外，采用动态范围增强后处理算法也可提升重建视频的视觉质量。

与传统相机相比，视频 CS 系统还需要考虑用户体验，如系统紧凑性、鲁棒性、功耗以及对焦和变焦能力[107]。目前，大多数视频 CS 原型系统都建立在实验室平台上，并使用演示性的简单案例验证。这些系统由于使用分立光学元件进行系统构建，需额外控制板用于调制和同步，以及需计算机用于数据存储和重建，导致体积庞大、易损坏，难以在实际户外场景中应用。此外，它们通常具有固定的焦距和焦平面，难以改变成像距离和视场。

为了克服或缓解这些缺点并实现商业化生产，需要利用光学工程技术优化系统。此外，半导体和集成电路的未来发展可以促进整个系统的集成，利用成熟的 PCE 图像传感器、ISP 和神经网络处理器（NPU），有望实现“片上”视频 CS 和重建。

5.2. 重建算法

与硬件和系统类似，算法也需要进一步改进，以便将视频CS系统部署到终端用户。随着这些系统在监控、医疗成像的各种应用中得到推广，解决可能阻碍其成功应用的算法挑战至关重要。

重建算法的鲁棒性是视频CS系统性能的基础。这些算法必须能够有效处理不同类型的噪声和变化的光照条件，以确保高质量的视频重建。通过网络传输的视频数据通常会受到各种类型的噪声影响，如高斯噪声、椒盐噪声、量化噪声等，这些噪声会显著降低重建视频的质量。为了解决这些难题，重建算法必须整合先进的降噪技术，以提高重建视频的清晰度和质量。在不同光照条件下拍摄的视频通常在亮度、对比度和色彩平衡方面存在差异。为了确保视频质量的稳定性，重建算法必须具备光照归一化技术和色彩校正能力。这些技术可以调整视频帧的亮度和对比度，并校正色彩失衡，从而提高整体视频质量。

优化低功耗重建网络对于视频CS系统的成本效益部署至关重要，特别是在移动和远程应用场景中。现场可编程门阵列（FPGAs）和专用芯片为降低视频CS系统的功耗提供了切实可行的解决方案。通过在这些平台上实现低功耗重建网络，可以显著降低系统的整体能耗，使其更适合在资源受限的环境中利用[108]。能源效率是低功耗重建网络的一个关键方面，算法需在保持高质量视频重建的同时，通过算法剪枝、量化、迁移学习、低秩近似等技术降低计算复杂度，从而减少能耗。

实时自适应感知对于视频CS系统有效处理动态场景和变化内容至关重要[109]。视频CS系统必须能够实时适应场景动态的变化。这需要开发自适应感知算法，能够根据场景的复杂性和运动内容调整采样率和压缩比。因此系统就可以确保重要的视频内容得到充分捕捉和重建。视频CS系统中使用的算法必须足够灵活，以处理不同的视频内容（即刚性运动和流体运动）、传输条件和用户需求，这需要开发能够适应不同采集条件的自适应算法。这种灵

活性是确保系统在不同应用中的可用性和有效性的关键。

6. 迈向实际应用

视频CS是一种新颖且有前景的成像范式，具有高吞吐量、低带宽以及更低功耗、内存占用、计算需求等优势。图6展示了一个真实视频CS系统捕捉的自由落体场景。该场景的空间分辨率为 800×800 像素，相机帧率为50 FPS。在压缩比40的情况下，可以轻松实现2000 FPS的视频效果。尽管某些现成的高速相机也可以支持如此高的帧率，但它们通常依赖昂贵的高端传感器，或存在记录时长有限的问题。图6对应的重建视频以及更高帧率（8000~18 000 FPS）的更多数据可见附录A中的视频S1和视频S2。

结合LLMs和大型视觉模型（LVMs）等新兴技术，视频CS在人机交互、自动驾驶、无人机和机器人等各种应用领域开辟了新的可能性。这些进步有助于开发更高效的端到端基于视觉的感知、规划、决策和控制框架。

近年来，研究人员已初步尝试将前端视频CS与后端语义CV任务集成，以提高整体效率。这些方法包括在视频CS之后级联特定的动作识别[33]、目标检测[34,110]等特定CV任务。这种方法无需进行视频重建，并可实现编码策略和视觉算法的端到端优化。与基于传统传感器和独立流水线设计的传统方法相比，这些方法的性能有了显著提高，特别是在高速场景中，而传统传感器在这类场景中会出现严重的运动模糊。

这些工作的缺点在于其应用场景有限，以及前端视频采集与后端CV任务的集成不足。考虑到这些缺点，Lu等[108]在网联自动驾驶汽车（CAVs）中视频CS的实际应用方面取得了进展，开发了一种名为EdgeCompression的新型车辆-边缘服务器-云闭环框架。该框架对CAVs在成像系统、视频分析和边缘计算平台方面存在的挑战进行了优化。通过引入视频CS来降低功耗、内存和计算消耗，该框架实



图6. 真实视频CS系统捕捉的自由落体。展示了两个编码测量值（第一列）和一些相应的重建帧（其他列）。系统的采集帧率和压缩比分别为50 FPS和40，最终重建帧率高达2000 FPS。

现了理想的检测精度（与基于重建的方法相当），同时显著提高了处理速度。

类似地，Zhang等[35]通过将视频CS与成熟的CV神经网络相结合，提出了一种更高效的基于视觉的语义检索框架。他们的框架包含两个压缩域网络骨干结构，可以直接从编码测量中提取描述性和区分性的特征。通过对这些特征上的现有计算机视觉网络进行再训练或微调，视频CS范式与现有CV算法兼容，促进了它们的联合开发。此外，该框架借助CV任务侧的反馈机制，强化了任务特定或自适应的视频CS能力。

总而言之，视频CS为更高效的视觉信息获取和处理提供了新的机遇。然而，由于现有硬件、算法和 workflows 等基础设施主要针对传统成像范式，视频CS在实际场景中的广泛应用仍面临重大挑战。尽管如此，随着支持技术和平台的成熟，预计相关技术在未来几年将加速发展，从而使视频CS在视觉领域产生革命性的影响。

7. 结论

研究者开发了多种成像技术来提升人类的视觉感知能力，视频CS是其中的典型代表。一方面，视频CS可以捕捉超快场景，用于发现新的现象；另一方面，也增强了现有相机的感知能力，特别是提升了成像系统的效能。这项技术的推广应用取决于视频CS所能带来的效能提升相比于其他竞争技术（如事件相机）[111]的优势。基于上述考量，本文综述了视频CS十年来的进展，涵盖了硬件系统和重建算法，同时指出了当前存在的研究空白，以期为未来的研究方向提供参考。

致谢

本研究得到国家自然科学基金(61931012、62171258、62088102、62271414)、浙江省杰出青年科学基金(LR23F010001)及西湖大学光电研究院重点项目(2023GD007)的资助。

Compliance with ethics guidelines

Zhihong Zhang, Siming Zheng, Min Qiu, Guohai Situ, David J. Brady, Qionghai Dai, Jinli Suo, and Xin Yuan declare that they have no conflict of interest or financial conflicts to disclose.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.eng.2024.08.013>.

References

- [1] Smith GE. Nobel lecture: the invention and early history of the CCD. *Rev Mod Phys* 2010;82(3):2307–12.
- [2] Mait JN, Euliss GW, Athale RA. Computational imaging. *Adv Opt Photonics* 2018;10(2):409–83.
- [3] Peng YE, Veeraraghavan A, Heidrich W, Wetzstein G. Deep optics: joint design of optics and image recovery algorithms for domain specific cameras. In: *Proceedings of the ACM SIGGRAPH 2020 Courses*; 2020 Aug 17–28; online. New York City: Association for Computing Machinery; 2020. p. 1–133.
- [4] Zhang B, Yuan X, Deng C, Zhang Z, Suo J, Dai Q. End-to-end snapshot compressed super-resolution imaging with deep optics. *Optica* 2022;9(4):451–4.
- [5] Zhang Z, Dong K, Suo J, Dai Q. Deep coded exposure: end-to-end co optimization of flutter shutter and deblurring processing for general motion blur removal. *Photon Res* 2023;11(10):1678–86.
- [6] Baek SH, Ikoma H, Jeon DS, Li Y, Heidrich W, Wetzstein G, et al. Single-shot hyperspectral-depth imaging with learned diffractive optics. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*; 2021 Oct 11–17; online. New York City: IEEE; 2021. p. 2651–60.
- [7] Yuan X, Brady DJ, Katsaggelos AK. Snapshot compressive imaging: theory, algorithms, and applications. *IEEE Signal Process Mag* 2021;38(2):65–88.
- [8] Tang H, Men T, Liu X, Hu Y, Su J, Zuo Y, et al. Single-shot compressed optical field topography. *Light Sci Appl* 2022;11:244.
- [9] Zhang Z, Deng C, Liu Y, Yuan X, Suo J, Dai Q. Ten-mega-pixel snapshot compressive imaging with a hybrid coded aperture. *Photon Res* 2021;9(11):2277–87.
- [10] Luo Y, Zhao Y, Li J, ÇetintasE, Rivenson Y, Jarrahi M, et al. Computational imaging without a computer: seeing through random diffusers at the speed of light. *eLight* 2022;2:4.
- [11] Sinha A, Lee J, Li S, Barbastathis G. Lensless computational imaging through deep learning. *Optica* 2017;4(9):1117–25.
- [12] Lull P, Liao X, Yuan X, Yang J, Kittle D, Carin L, et al. Coded aperture compressive temporal imaging. *Opt Express* 2013;21(9):10526–45.
- [13] Candes EJ, Tao T. Near-optimal signal recovery from random projections: universal encoding strategies? *IEEE Trans Inf Theory* 2006;52(12):5406–25.
- [14] Donoho D. Compressed sensing. *IEEE Trans Inf Theory* 2006;52(4):1289–306.
- [15] Yao H, Dai F, Zhang S, Zhang Y, Tian Q, Xu C. DR2-Net: deep residual reconstruction network for image compressive sensing. *Neurocomputing* 2019;359:483–93.
- [16] Zhang J, Xiong T, Tran T, Chin S, Etienne-Cummings R. Compact all-CMOS spatiotemporal compressive sensing video camera with pixel-wise coded exposure. *Opt Express* 2016;24(8):9013–24.
- [17] Wei M, Sarhangnejad N, Xia Z, Gusev N, Katie N, Genov R, et al. Coded two bucket cameras for computer vision. In: *Proceedings of the Computer Vision ECCV 2018*; 2018 Sep 8–14; Munich, Germany. Berlin: Springer; 2018. p. 54–71.
- [18] Liu Y, Yuan X, Suo J, Brady DJ, Dai Q. Rank minimization for snapshot compressive imaging. *IEEE Trans Pattern Anal Mach Intell* 2019;41(12):2990–3006.
- [19] Yuan X, Liu Y, Suo J, Dai Q. Plug-and-play algorithms for large-scale snapshot compressive imaging. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; 2020 Jun 13–19; Seattle, WA, USA. New York City: IEEE; 2020. p. 1444–54.
- [20] Izadi S, Sutton D, Hamarneh G. Image denoising in the deep learning era. *Artif Intell Rev* 2023;56(7):5929–74.
- [21] Zhang K, Ren W, Luo W, Lai WS, Stenger B, Yang MH, et al. Deep image deblurring: a survey. *Int J Comput Vis* 2022;130(9):2103–30.
- [22] Rawat W, Wang Z. Deep convolutional neural networks for image classification: a comprehensive review. *Neural Comput* 2017;29(9):2352–449.
- [23] Zhu H, Wei H, Li B, Yuan X, Kehtarnavaz N. A review of video object detection: datasets, metrics and methods. *Appl Sci* 2020;10(21):7834.
- [24] Jiao L, Wang D, Bai Y, Chen P, Liu F. Deep learning in visual tracking: a review. *IEEE Trans Neural Netw Learn Syst* 2021;34(9):5497–516.

- [25] Yuan X. Various plug-and-play algorithms with diverse total variation methods for video snapshot compressive imaging. In: Proceedings of the Artificial Intelligence: First CAAI International Conference; 2021 Jun 5–6; Hangzhou, China. Berlin: Springer; 2021. p. 335–46.
- [26] Yuan X, Liu Y, Suo J, Durand F, Dai Q. Plug-and-play algorithms for video snapshot compressive imaging. *IEEE Trans Pattern Anal Mach Intell* 2021; 44(10):7093–111.
- [27] Chen Y, Gui X, Zeng J, Zhao XL, He W. Combining low-rank and deep plug-and-play priors for snapshot compressive imaging. *IEEE Trans Neural Netw Learn Syst*. In press.
- [28] Meng Z, Yuan X, Jalali S. Deep unfolding for snapshot compressive imaging. *Int J Comput Vis* 2023;131(11):2933–58.
- [29] Wu Z, Zhang J, Mou C. Dense deep unfolding network with 3D-CNN prior for snapshot compressive imaging. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV); 2021 Oct 11–17; Montreal, BC, Canada. New York City: IEEE; 2021. p. 4892–901.
- [30] Yang C, Zhang S, Yuan X. Ensemble learning priors driven deep unfolding for scalable video snapshot compressive imaging. In: Proceedings of the Computer Vision-ECCV 2022; 2022 Oct 23–27; Tel Aviv, Israel. Berlin: Springer; 2022. p. 600–18.
- [31] Suo J, Zhang W, Gong J, Yuan X, Brady DJ, Dai Q, et al. Computational imaging and artificial intelligence: the next revolution of mobile vision. *Proc IEEE* 2023;111(12):1607–39.
- [32] Kwan C, Chou B, Yang J, Rangamani A, Tran T, Zhang J, et al. Target tracking and classification using compressive measurements of MWIR and LWIR coded aperture cameras. *JSP* 2019;10(3):73–95.
- [33] Okawara T, Yoshida M, Nagahara H, Yagi Y. Action recognition from a single coded image. In: Proceedings of the 2020 IEEE International Conference on Computational Photography (ICCP); 2020 Apr 24–26; LouisSaint, MO, USA. New York City: IEEE; 2020. p. 1–11.
- [34] Hu C, Huang H, Chen M, Yang S, Chen H. Video object detection from one single image through opto-electronic neural network. *APL Photonics* 2021;6(4):046104.
- [35] Zhang Z, Zhang B, Yuan X, Zheng S, Su X, Suo J, et al. From compressive sampling to compressive tasking: retrieving semantics in compressed domain with low bandwidth. *PhotonIX* 2022;3:19.
- [36] Shannon C. Communication in the presence of noise. *Proc IRE* 1949;37(1):10–21.
- [37] Jalali S, Yuan X. Snapshot compressed sensing: performance bounds and algorithms. *IEEE Trans Inf Theory* 2019;65(12):8005–24.
- [38] Yuan X. Generalized alternating projection based total variation minimization for compressive sensing. In: Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP); 2016 Sep 25–28; Phoenix, AZ, USA. New York City: IEEE; 2016. p. 2539–43.
- [39] Duarte MF, Davenport MA, Takhar D, Laska JN, Sun T, Kelly KF, et al. Single pixel imaging via compressive sampling. *IEEE Signal Process Mag* 2008;25(2):83–91.
- [40] Jalali S, Maleki A. From compression to compressed sensing. *Appl Comput Harmon Anal* 2016;40(2):352–85.
- [41] Yuan X, Llull P, Liao X, Yang J, Brady DJ, Sapiro G, et al. Low-cost compressive sensing for color video and depth. In: Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2014 Jun 23–28; Columbus, OH, USA. New York City: IEEE; 2014. p. 3318–25.
- [42] Koller R, Schmid L, Matsuda N, Niederberger T, Spinoulas L, Cossairt O, et al. High spatio-temporal resolution video with compressed sensing. *Opt Express* 2015;23(12):15992–6007.
- [43] Reddy D, Veeraraghavan A, Chellappa R. P2C2: programmable pixel compressive camera for high speed imaging. In: Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2011 Jun 21–23; Springs, CO, USA. New York City: IEEE; 2011. p. 329–36.
- [44] Hitomi Y, Gu J, Gupta M, Mitsunaga T, Nayar SK. Video from a single coded exposure photograph using a learned over-complete dictionary. In: Proceedings of the 2011 International Conference on Computer Vision (ICCV); 2011 Nov 6–13; Barcelona, Spain. New York City: IEEE; 2011. p. 287–94.
- [45] Liu D, Gu J, Hitomi Y, Gupta M, Mitsunaga T, Nayar S. Efficient space-time sampling with pixel-wise coded exposure for high-speed imaging. *IEEE Trans Pattern Anal Mach Intell* 2014;36(2):248–60.
- [46] Qiao M, Meng Z, Ma J, Yuan X. Deep learning for video compressive sensing. *APL Photonics* 2020;5(3):030801.
- [47] Guzmán F, Meza P, Vera E. Compressive temporal imaging using a rolling shutter camera array. *Opt Express* 2021;29(9):12787–800.
- [48] Vera E, Guzmán F, Diaz N. Shuffled rolling shutter for snapshot temporal imaging. *Opt Express* 2022;30(2):887–901.
- [49] Sun Y, Yuan X, Pang S. High-speed compressive range imaging based on active illumination. *Opt Express* 2016;24(20):22836–46.
- [50] Guzmán F, Skowronek J, Vera E, Brady DJ. Compressive video via IR-pulsed illumination. *Opt Express* 2023;31(23):39201–12.
- [51] Luo Y, Jiang J, Cai M, Mirabbasi S. CMOS computational camera with at-warp coded exposure image sensor for single-shot spatial-temporal compressive sensing. *Opt Express* 2019;27(22):31475–89.
- [52] Yoshida M, Sonoda T, Nagahara H, Endo K, Sugiyama Y, Taniguchi R. High speed imaging using CMOS image sensor with quasi pixel-wise exposure. *IEEE Trans Comput Imaging* 2020;6:463–76.
- [53] Qiao M, Liu X, Yuan X. Snapshot spatial-temporal compressive imaging. *Opt Lett* 2020;45(7):1659–62.
- [54] Deng C, Zhang Y, Mao Y, Fan J, Suo J, Zhang Z, et al. Sinusoidal sampling enhanced compressive camera for high speed imaging. *IEEE Trans Pattern Anal Mach Intell* 2021;43(4):1380–93.
- [55] Liang J, Zhu L, Wang LV. Single-shot real-time femtosecond imaging of temporal focusing. *Light Sci Appl* 2018;7:42.
- [56] Gao L, Liang J, Li C, Wang LV. Single-shot compressed ultrafast photography at one hundred billion frames per second. *Nature* 2014;516(7529):74–7.
- [57] Czajkowski KM, Pastuszczyk A, Kotyn'ski R. Real-time single-pixel video imaging with Fourier domain regularization. *Opt Express* 2018;26(16):20009–22.
- [58] Higham CF, Murray-Smith R, Padgett MJ, Edgar MP. Deep learning for real time single-pixel video. *Sci Rep* 2018;8:2369.
- [59] Wang P, Liang J, Wang LV. Single-shot ultrafast imaging attaining 70 trillion frames per second. *Nat Commun* 2020;11:2091.
- [60] Lu R, Chen B, Liu G, Cheng Z, Qiao M, Yuan X. Dual-view snapshot compressive imaging via optical flow aided recurrent neural network. *Int J Comput Vis* 2021;129(12):3279–98.
- [61] Liu X, Zhu M, Zheng S, Luo R, Wu H, Yuan X. Video snapshot compressive imaging using adaptive progressive coding for high-quality reconstruction under different illumination circumstances. *Opt Lett* 2024;49(1):85–8.
- [62] Wang P, Wang L, Qiao M, Yuan X. Full-resolution and full-dynamic-range coded aperture compressive temporal imaging. *Opt Lett* 2023;48(18):4813–6.
- [63] Hahamovich E, Monin S, Hazan Y, Rosenthal A. Single pixel imaging at megahertz switching rates via cyclic hadamard masks. *Nat Commun* 2021;12:4516.
- [64] Kilcullen P, Ozaki T, Liang J. Compressed ultrahigh-speed single-pixel imaging by swept aggregate patterns. *Nat Commun* 2022;13:7879.
- [65] Mur AL, Peyrin F, Ducros N. Recurrent neural networks for compressive video reconstruction. In: Proceedings of the 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI); 2020 Apr 3–7; Iowa City, IA, USA. New York City: IEEE; 2020. p. 1651–4.
- [66] Ma X, Yuan X, Arce GR. High resolution LED-based snapshot compressive spectral video imaging with deep neural networks. *IEEE Trans Comput Imaging* 2023;9:869–80.
- [67] Martel JNP, Muller LK, Carey SJ, Dudek P, Wetzstein G. Neural sensors: learning pixel exposures for HDR imaging and video compressive sensing with programmable sensors. *IEEE Trans Pattern Anal Mach Intell* 2020;42(7):1642–53.
- [68] Carey SJ, Lopich A, Barr DRW, Wang B, Dudek PA. 100,000 fps vision sensor with embedded 535GOPS/W 256 256 SIMD processor array. In: Proceedings of the 2013 Symposium on VLSI Circuits; 2013 Jun 12–14; Kyoto, Japan. New York City: IEEE; 2013. p. C182–3.
- [69] Sarhangnejad N, Katic N, Xia Z, Wei M, Gusev N, Dutta G, et al. 5.5 Dual-tap pipelined-code-memory coded-exposure-pixel CMOS image sensor for multi exposure single-frame computational imaging. In: Proceedings of the 2019 IEEE International Solid-State Circuits Conference (ISSCC); 2019 Feb 17–21; San Francisco, CA, USA. New York City: IEEE; 2019. p. 102–4.
- [70] Luo Y, Ho D, Mirabbasi S. Exposure-programmable CMOS pixel with selective charge storage and code memory for computational imaging. *IEEE Trans Circuits Syst* 2018;65(5):1555–66.
- [71] Shedligeri P, Anupama S, Mitra K. A unified framework for compressive video recovery from coded exposure techniques. In: Proceedings of the 2021 IEEE Winter Conference on Applications of Computer Vision (WACV); 2021 Jan 3–8; Waikoloa, HI, USA. New York City: IEEE; 2021. p. 1599–608.
- [72] Gulve R, Sarhangnejad N, Dutta G, Sakr M, Nguyen D, Rangel R, et al. A 39,000 subexposures/s CMOS image sensor with dual-tap coded-exposure data memory pixel for adaptive single-shot computational imaging. In: Proceedings of the 2022 IEEE Symposium on VLSI Technology and Circuits; 2022 Jun 12–17; Honolulu, HI, USA. New York City: IEEE; 2022. p. 78–9.
- [73] Gulve R, Rangel R, Barman A, Nguyen D, Wei M, Skar MA, et al. Dual-port CMOS image sensor with regression-based HDR flux-to-digital conversion and 80 ns rapid-update pixel-wise exposure coding. In: Proceedings of the 2023

- IEEE International Solid State Circuits Conference (ISSCC); 2023 Feb 19 23; San Francisco, CA, USA. New York City: IEEE; 2023. p. 104–6.
- [74] Wagadarikar A, John R, Willett R, Brady D. Single disperser design for coded aperture snapshot spectral imaging. *Appl Opt* 2008;47(10):B44–51.
- [75] Qi D, Zhang S, Yang C, He Y, Cao F, Yao J, et al. Single-shot compressed ultrafast photography: a review. *Adv Photonics* 2020;2(1):014003.
- [76] Tsai TH, Llull P, Yuan X, Carin L, Brady DJ. Spectral-temporal compressive imaging. *Opt Lett* 2015;40(17):4054–7.
- [77] Sun Y, Yuan X, Pang S. Compressive high-speed stereo imaging. *Opt Express* 2017;25(15):18182.
- [78] Dou Y, Cao M, Wang X, Liu X, Yuan X. Coded aperture temporal compressive digital holographic microscopy. *Opt Lett* 2023;48(20):5427–30.
- [79] Luo R, Cao M, Liu X, Yuan X. Snapshot compressive structured illumination microscopy. *Opt Lett* 2024;49(2):186–9.
- [80] Chen Z, Zheng S, Wang W, Song J, Yuan X. Temporal structured illumination and vision-transformer enables large field-of-view binary snapshot ptychography. *Opt Express* 2024;32(2):1540–51.
- [81] Hu M, Wu Z, Huang Q, Yuan X, Brady D. Sampling for snapshot compressive imaging. *Intell Comput* 2023;2:0038.
- [82] Qiao M, Liu X, Yuan X. Snapshot temporal compressive microscopy using an iterative algorithm with untrained neural networks. *Opt Lett* 2021;46(8):1888–91.
- [83] Yang J, Yuan X, Liao X, Llull P, Brady DJ, Sapiro G, et al. Video compressive sensing using gaussian mixture models. *IEEE Trans Image Process* 2014;23(11):4863–78.
- [84] Wu Z, Yang C, Su X, Yuan X. Adaptive deep PnP algorithm for video snapshot compressive imaging. *Int J Comput Vis* 2023;131(7):1662–79.
- [85] Cheng Z, Lu R, Wang Z, Zhang H, Chen B, Meng Z, et al. BIRNAT: bidirectional recurrent neural networks with adversarial training for video snapshot compressive imaging. In: *Proceedings of the Computer Vision—ECCV 2020*; 2020 Aug 23–28; Glasgow, UK. Berlin: Springer; 2020. p. 258–75.
- [86] Cheng Z, Chen B, Liu G, Zhang H, Lu R, Wang Z. Memory-efficient network for large-scale video compressive sensing. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; 2021 Jun 20–25; Nashville, TN, USA. New York City: IEEE; 2021. p. 16241–50.
- [87] Wang Z, Zhang H, Cheng Z, Chen B, Yuan X. MetaSCI: scalable and adaptive reconstruction for video compressive sensing. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; 2021 Jun 20–25; Nashville, TN, USA. New York City: IEEE; 2021. p. 2083–92.
- [88] Meng Z, Jalali S, Yuan X. GAP-Net for snapshot compressive imaging. 2020. [arXiv:2012.08364](https://arxiv.org/abs/2012.08364).
- [89] Ma J, Liu XY, Shou Z, Yuan X. Deep tensor ADMM-net for snapshot compressive imaging. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*; 2019 Oct 27–Nov 2; Seoul, Republic of Korea. New York City: IEEE; 2019. p. 10222–31.
- [90] Zhao Y, Zheng S, Yuan X. Deep equilibrium models for snapshot compressive imaging. In: *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*; 2023 Feb 7–14; Washington, DC, USA. Pennsylvania Ave: The Association for the Advancement of Artificial Intelligence; 2023. p. 3642–50.
- [91] Zheng S, Yuan X. Unfolding framework with prior of convolution-transformer mixture and uncertainty estimation for video snapshot compressive imaging. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*; 2023 Oct 2–6; Paris, France. New York City: IEEE; 2023. p. 12738–49.
- [92] Wang L, Cao M, Yuan X. EfficientSCI: densely connected network with space time factorization for large-scale video snapshot compressive imaging. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; 2023 Jun 17–24; Vancouver, BC, Canada. New York City: IEEE; 2023. p. 18477–86.
- [93] Iliadis M, Spinoulas L, Katsaggelos AK. Deep fully-connected networks for video compressive sensing. *Digit Signal Process* 2018;72:9–18.
- [94] Dong W, Shi G, Li X, Ma Y, Huang F. Compressive sensing via nonlocal low rank regularization. *IEEE Trans Image Process* 2014;23(8):3618–32.
- [95] Maggioni M, Boracchi G, Foi A, Egiazarian K. Video denoising, deblocking, and enhancement through separable 4D nonlocal spatiotemporal transforms. *IEEE Trans Image Process* 2012;21(9):3952–66.
- [96] Yang J, Liao X, Yuan X, Llull P, Brady DJ, Sapiro G, et al. Compressive sensing by learning a gaussian mixture model from measurements. *IEEE Trans Image Process* 2015;24(1):106–19.
- [97] Venkatakrishnan SV, Bouman CA, Wohlberg B. Plug-and-play priors for model based reconstruction. In: *Proceedings of the 2013 IEEE Global Conference on Signal and Information Processing*; 2013 Dec 3–5; Austin, TX, USA. New York City: IEEE; 2013. p. 945–8.
- [98] Boyd S, Parikh N, Chu E, Peleato B, Eckstein J. *Distributed optimization and statistical learning via the alternating direction multipliers*. Norwell: Now Foundations and Trends; 2011. method of
- [99] Liao X, Li H, Carin L. Generalized alternating projection for weighted- ℓ_2, ℓ_1 minimization with applications to model-based compressive sensing. *SIAM J Imaging Sci* 2014;7(2):797–823.
- [100] Li Y, Qi M, Wei M, GenovR, Kutulakos KN, Heidrich W, et al. End-to-end video compressive sensing using Anderson-accelerated unrolled networks. In: *Proceedings of the 2020 IEEE International Conference on Computational Photography (ICCP)*; 2020 Apr 24–26; LouisSaint, MO, USA. New York City: IEEE; 2020. p. 1–12.
- [101] Zheng S, Yang X, Yuan X. Two-stage is enough: a concise deep unfolding reconstruction network for flexible video compressive sensing. 2022. [arXiv:2201.05810](https://arxiv.org/abs/2201.05810).
- [102] Wang L, Cao M, Zhong Y, Yuan X. Spatial-temporal transformer for video snapshot compressive imaging. *IEEE Trans Pattern Anal Mach Intell* 2022; 45(7):9072–89.
- [103] Cao M, Wang L, Zhu M, Yuan X. Hybrid CNN-transformer architecture for efficient large-scale video snapshot compressive imaging. *Int J Comput Vis* 2024;132:4521–40.
- [104] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. In: *Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*; 2015 Oct 5–9; Munich, Germany. Berlin: Springer; 2015. p. 234–41.
- [105] Cheng Z, Chen B, Lu R, Wang Z, Zhang H, Meng Z, et al. Recurrent neural networks for snapshot compressive imaging. *IEEE Trans Pattern Anal Mach Intell* 2023;45(2):2264–81.
- [106] Cai Y, Zheng Y, Lin J, Yuan X, Zhang Y, Wang H. Binarized spectral compressive imaging. In: *Proceedings of the Thirty-Seventh Conference on Neural Information Processing Systems (NeurIPS-2023)*; 2023 Dec 10; New Orleans, LA, USA. San Diego: NeurIPS Proceedings; 2023. p. 1–9.
- [107] Wang P, Wang L, Yuan X. Deep optics for video snapshot compressive imaging. In: *Proceedings of the Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*; 2023 Oct 1–6; Paris, France. New York City: IEEE; 2023. p. 10646–56.
- [108] Lu S, Yuan X, Shi W. Edge compression: an integrated framework for compressive imaging processing on CAVs. In: *Proceedings of the 2020 IEEE/ACM Symposium on Edge Computing (SEC)*; 2020 Nov 11–13; JoseSan, CA, USA. New York City: IEEE; 2020. p. 125–38.
- [109] Lu S, Yuan X, Katsaggelos AK, Shi W. Reinforcement learning for adaptive video compressive sensing. *ACM Trans Intell Syst Technol* 2023;14(5):1–21.
- [110] Bethi YRT, Narayanan S, Rangan V, Chakraborty A, Thakur CS. Real-time object detection and localization in compressive sensed video. In: *Proceedings of the 2021 IEEE International Conference on Image Processing (ICIP)*; 2021 Sep 19–22; Anchorage, AK, USA. New York City: IEEE; 2021. p. 1489–93.
- [111] Gallego G, Delbruck T, Orchard G, Bartolozzi C, Taba B, Censi A, et al. Event based vision: a survey. *IEEE Trans Pattern Anal Mach Intell* 2022;44(1):154–80.